



**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Οπτικοποίηση Δεδομένων και Εξόρυξη Γνώσης –
Συγκριτική Αξιολόγηση Διαθέσιμων Τεχνολογιών

Πολυξένη Ορκοπούλου

Επιβλέπων καθηγητής: Τζήμας Γιάννης

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή
Πάτρα, 12/10/21

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Τζήμας Ιωάννης, Καθηγητής
2. Τσακνάκης Ιωάννης, Αναπληρωτής Καθηγητής
3. Κίτσος Παρασκευάς, Αναπληρωτής Καθηγητής

Αφιέρωση

*Στα παιδιά μου,
Τους ανεκτίμητους θησαυρούς της ζωής μου..*

Ευχαριστίες

Στον καθηγητή μου Γιάννη Τζήμα,

Για την απίστευτη υπομονή του! 😊

Στην οικογένειά μου για όλη την υποστήριξή τους αυτό το διάστημα!

Περιεχόμενα

Αφιέρωση.....	2
Ευχαριστίες.....	3
Περιεχόμενα	4
Λίστα Εικόνων.....	7
Πρόλογος	12
Περίληψη	13
Abstract.....	15
1 Γιατί Κάνουμε Οπτικοποίηση;.....	16
2 Εγκέφαλος και Οπτική Αντίληψη	23
2.1 Προσοχή στις Τρισδιάστατες Απεικονίσεις.....	30
2.2 Το «Μεγαλύτερο Πακέτο»/Ένα Καλό Περιτύλιγμα.....	33
3 Γενικός Οδηγός Τύπων Πλοκής.....	45
3.1 Ποσότητες (Amounts)	45
3.2 Κατανομές (Distributions)	45
3.3 Αναλογίες (Proportions).....	46
3.4 Σχέσεις X-Y (X-Y Relationships)	48
3.5 Γεωχωρικά Δεδομένα (Geospatial Data)	49
3.6 Αβεβαιότητα (Uncertainty)	49
4 Αρχή Επίλυσης Γενικών Οπτικών Προβλημάτων... ..	51
4.1 Πρόβλημα 1. Αναλογία Δεδομένων – Μελανιού	51
4.2 Πρόβλημα 2. Αιώρηση	52
4.3 Πρόβλημα 3. Πλέγμα	53

4.4	Πρόβλημα 4. Σύνθετες Οπτικοποιήσεις	56
4.5	Πρόβλημα 5. Γενίκευση Οπτικοποιήσεων	58
4.6	Πρόβλημα 6. Οπτική Διάκριση.....	59
5	Τι Χρειάζεται να Γνωρίζουμε για τη Σωστή Επιλογή Λογισμικού	63
5.1	Βασικές Αρχές	63
5.2	Οι Πιο Συνηθισμένοι Τύποι Εικόνας	66
5.3	Διαδραστικότητα.....	67
6	Big Data.....	72
7	Χρώμα.....	76
7.1	Για να Διαφοροποιεί τα Δεδομένα Μεταξύ τους	76
7.2	Για να Αντιπροσωπεύει Τιμές Δεδομένων.....	77
7.3	Για να Τονίζει.....	80
7.4	Κοινά Χρωματικά Λάθη.....	81
7.4.1	Κωδικοποίηση Πολλών Διαφορετικών Στοιχείων με Διαφορετικά Χρώματα	81
7.4.2	Μη Χρήση Σωστού Μεγέθους για τα Χρωματιστά Στοιχεία	85
7.4.3	Η Μη Αντιμέτωπιση Χρωμάτων που Αναμειγνύονται και Υπερκαλύπτονται.....	87
8	Αισθητική σε Τίτλους, Λεζάντες, Υπομνήματα, Ετικέτες Αξόνων, Γραμμών και Πίνακες! ...	93
8.1	Τίτλοι και Υπομνήματα	93
8.2	Πίνακες.....	101
8.3	Αποφυγή Χρήσης Γραμμών για Απεικόνιση Διαγραμμάτων.....	102
9	Εργαλεία Οπτικοποίησης Δεδομένων.....	109
9.1	Εργαλεία για Μη Επαγγελματίες	113
9.1.1	Διαγράμματα και Γραφήματα	113
9.1.2	Infographics.....	127

9.1.3	Χάρτες	130
9.1.4	Γραφήματα Δικτύου (Network graphs)	131
9.1.5	Μαθηματικά Γραφήματα	132
9.2	Εργαλεία για Προγραμματιστές.....	133
9.2.1	Διαγράμματα και Γραφήματα	133
9.2.2	Χάρτες	142
9.2.3	Γραφήματα Δικτύου	146
9.2.4	Χρηματοοικονομικά Διαγράμματα.....	147
Επίλογος.....		148
10	Βιβλιογραφία	149

Λίστα Εικόνων

Εικόνα 1-1: [1] Πολλαπλές οπτικοποιήσεις για καλύτερη κατανόηση	18
Εικόνα 2-1: [3] [4] Αρχές Gestalt.....	23
Εικόνα 2-2: [3] Αρχές Gestalt	24
Εικόνα 2-3: [1] Αναλογία δεδομένων/μελανιού	24
Εικόνα 2-4: [1] Χρήση επισημάνσεων	25
Εικόνα 2-5: [1] Εξομάλυνση, ομαδοποίηση άκρων.....	26
Εικόνα 2-6: [1] Ψευδαισθήσεις	27
Εικόνα 2-7: [1] Παράμετροι δύσκολοι να ερμηνευτούν πίσω σε νούμερα (με αστερίσκο)	28
Εικόνα 2-8: [2] Έμφαση στους άξονες	28
Εικόνα 2-9: [1] Στοιχεία με διαφορετικό σημείο εκκίνησης	29
Εικόνα 2-10: [1] Ένα παράδειγμα εξαίρεσης προβληματικής παραμέτρου	29
Εικόνα 2.1-11: [2] Τρισδιάστατες απεικονίσεις με λανθασμένες εντυπώσεις	31
Εικόνα 2.1-12: [2] Τρισδιάστατες απεικονίσεις αρκετής νοητικής προσπάθειας	31
Εικόνα 2.1-13: [2] Παράδειγμα λύσεων για προβληματικές τρισδιάστατες απεικονίσεις	32
Εικόνα 2.1-14: [2] Παράδειγμα προβληματικής τρισδιάστατης απεικόνισης.....	32
Εικόνα 2.1-15: [2] Παραδείγματα χρήσης τρισδιάστατων απεικονήσεων, καλή πρακτική	33
Εικόνα 2.2-16: [1] Πίνακας ελέγχου	37
Εικόνα 2.2-17: [1] Αρχικές πτυχές δεδομένων με διαφορετικό επίπεδο λεπτομέρειας - παράδειγμα	39
Εικόνα 2.2-18: [1] Τελικό αποτέλεσμα παραδείγματος.....	40
Εικόνα 2.2-19: [1] Παράδειγμα εύκολα αναγνωρίσιμο	42
Εικόνα 2.2-20: [1] Χειροπιαστό παράδειγμα	42
Εικόνα 2.2-21: [5] Παράδειγμα εύκολα αναγνωρίσιμο και χειροπιαστό	43
Εικόνα 2.2-22: [6] Παράδειγμα εύκολα αναγνωρίσιμο και χειροπιαστό	44
Εικόνα 3.1-1: [2] Γενικοί τύποι πλοκής Ποσοτήτων	45
Εικόνα 3.2-2: [2] Γενικοί τύποι πλοκής Κατανομών	46
Εικόνα 3.3-3: [2] Γενικός τύπος πλοκής Αναλογιών	47
Εικόνα 3.4-4: [2] Γενικός τύπος πλοκής Χ-Υ Σχέσεων.....	48

Εικόνα 3.5-5: [2] Γενικός τύπος πλοκής Γεωχωρικών Δεδομένων	49
Εικόνα 3.6-6: [2] Γενικός τύπος πλοκής Αβεβαιότητας.....	50
Εικόνα 4.1-1: [2] Παράδειγμα προβλήματος αναλογίας δεδομένων-μελανιού	51
Εικόνα 4.1-2: [2] Παράδειγμα λύσης προβλήματος αναλογίας δεδομένων-μελανιού.....	51
Εικόνα 4.2-3: [2] Παράδειγμα προβλήματος-λύσης μείωσης μελανιού σε μικρά πολλαπλάσια	52
Εικόνα 4.3-4: [2] Παράδειγμα προβλήματος-λύσης χρήσης πλέγματος	53
Εικόνα 4.3-5: [2] Παράδειγμα προβλήματος χρήσης πλέγματος σε ράβδους	54
Εικόνα 4.3-6: [2] Παράδειγμα μικρού αριθμού σημείων δεδομένων με προσδιορισμό σε αυτά	55
Εικόνα 4.3-7: [2] Παράδειγμα προβλήματος-λύσης μεγάλου αριθμού σημείων δεδομένων συγκριτικής γραμμής $x = y$	55
Εικόνα 4.4-8: [2] Παράδειγμα προβλήματος-λύσης πολύπλοκων οπτικοποιήσεων.....	56
Εικόνα 4.4-9: [2] Παράδειγμα λύσης για εξαιρετικά σύνθετη οπτικοποίηση	57
Εικόνα 4.5-10: [2] Παραδείγματα γενικευμένων γραφημάτων.....	58
Εικόνα 4.5-11: [2] Ισοτυπικό διάγραμμα (isotype plot).....	59
Εικόνα 4.6-12: [2] Παράδειγμα 1 κοινής οπτικής γλώσσας αλλά όχι όμοιας σχεδίασης.....	60
Εικόνα 4.6-13: [2] Παράδειγμα 2 κοινής οπτικής γλώσσας αλλά όχι όμοιας σχεδίασης.....	60
Εικόνα 5.1-1: [2] Αναπαραγωγή πανομοιότυπων οπτικοποιήσεων	63
Εικόνα 5.1-2: [2] Τροποποίηση της εμφάνισης της οπτικοποίησης κατά την έξοδο.....	64
Εικόνα 5.3-3: [7] Παράδειγμα διαδραστικότητας πολλών τύπων	69
Εικόνα 5.3-4: [8] Παράδειγμα διαδραστικότητας, κινούμενο	70
Εικόνα 5.3-5: [9] Παράδειγμα διαδραστικότητας μεγάλων δεδομένων (big data).....	71
Εικόνα 6-1: [10] Παράδειγμα ζωντανών ροών δεδομένων - Χάρτης Αστραπών (live big data)..	74
Εικόνα 6-2: [11] Παράδειγμα ζωντανών ροών δεδομένων - Χάρτης παρακολούθησης πτήσεων (live big data)	75
Εικόνα 7.1-1: [2] Χρήση χρώματος για τη διαφοροποίηση των δεδομένων μεταξύ τους	76
Εικόνα 7.2-2: [2] Χρήση χρώματος για αντιπροσώπευση τιμών δεδομένων	78
Εικόνα 7.2-3: [2] Χρήση χρώματος για αντιπροσώπευση τιμών δεδομένων	78

Εικόνα 7.2-4: [2] Παράδειγμα αποκλίνουσας κλίμακας χρώματος (πάνω αριστερά) και, χρήση χρώματος για τονισμό	79
Εικόνα 7.3-5: [2] Παράδειγμα χρήσης ενός χρώματος για τονισμό	80
Εικόνα 7.4.1-6: [2] Κωδικοποίηση διαφορετικών στοιχείων με διαφορετικά χρώματα.....	82
Εικόνα 7.4.1-7: [2] Παραδείγματα μη μονοτονικών κλιμάκων χρώματος.....	83
Εικόνα 7.4.1-8: [2] Χρωματικές ανεπάρκειες στην όραση (CVD).....	84
Εικόνα 7.4.2-9: [2] Προσοχή στη λανθασμένη χρήση χρωματικών στοιχείων	86
Εικόνα 7.4.2-10: [2] Παραδείγματα μεταβλητών ως χρώμα	87
Εικόνα 7.4.3-11: [2] Ανάμειξη και υπερκάλυψη χρωμάτων	88
Εικόνα 7.4.3-12: [2] Χρήση σχημάτων ως σύμβολα, πιο σαφή σε άτομα με CVD.....	89
Εικόνα 7.4.3-13: [2] Πρόβλημα υπερκάλυψης.....	89
Εικόνα 7.4.3-14: [2] Λύση προβλήματος υπερκάλυψης μεγάλων (πυκνών) δεδομένων με χρήση ιστογράμματος.....	90
Εικόνα 7.4.3-15: [2] Πίσω στο αρχικό πρόβλημα.....	91
Εικόνα 7.4.3-16: [2] Λύση προβλήματος υπερκάλυψης με χρήση γραμμών περιγράμματος	91
Εικόνα 7.4.3-17: [2] Λύση προβλήματος υπερκάλυψης με χρήση μικρών πολλαπλών διαγραμμάτων	92
Εικόνα 8.1-1: [2] Τίτλοι ενσωματωμένοι και μη.....	93
Εικόνα 8.1-2: [1] Σύντομοι τίτλοι, ισχυρό αντίκτυπο	94
Εικόνα 8.1-3: [2] Υπερβολή και λάθη στις αριθμητικές μεταβλητές	95
Εικόνα 8.1-4: [2] Τίτλοι, αριθμητικές μεταβλητές και αυτονόητες παραλήψεις	95
Εικόνα 8.1-5: [2] Μεγέθη ετικετών, σωστή χρήση και ισορροπίες.....	96
Εικόνα 8.1-6: [2] Παράλειψη αξόνων σε ραβδογράμματα και διαγράμματα μωσαϊκού	97
Εικόνα 8.1-7: [2] Παράλειψη αξόνων σε πίτες και δεντροχάρτες	97
Εικόνα 8.1-8: [2] Αντιληπτή σειρά δεδομένων και υπομνήματος	98
Εικόνα 8.1-9: [2] Αποφυγή υπομνημάτων με μέτρο	99
Εικόνα 8.1-10: [2] Κωδικοποίηση μεμονωμένης μεταβλητής σε πολλαπλά αισθητικά στοιχεία με χρήση στοιχείου και όχι υπομνήματος	100
Εικόνα 8.2-11: [2] Λεζάντες και πίνακες.....	102

Εικόνα 8.3-12: [2] Αποφυγή γραμμών στην απεικόνιση με χρήση συμπαγών/ημιδιάφανων έγχρωμων σχημάτων	103
Εικόνα 8.3-13: [2] Αποφυγή γραμμών στην απεικόνιση.....	103
Εικόνα 8.3-15: [2] Παρερμηνεία διαγραμμάτων λόγω άξονα	104
Εικόνα 8.3-15: [2] Σωστή χρήση αξόνων για σωστή ερμηνεία διαγραμμάτων	104
Εικόνα 8.3-16: [2] Σωστή χρήση αξόνων για εμφάνιση διαφορών μεταξύ των συνθηκών	104
Εικόνα 8.3-17: [2] Έγκυρη χρήση ράβδων για αναλογίες σε μεγάλη κλίμακα	106
Εικόνα 8.3-18: [2] Παράδειγμα γραμμικής/λογαριθμικής κλίμακας.....	106
Εικόνα 8.3-19: [2] Αποφυγή χρήσης ράβδων σε λογαριθμική κλίμακα, χρήση κουκίδων.....	106
Εικόνα 8.3-20: [2] Χρήση κλίμακας τετραγωνικής ρίζας.....	107
Εικόνα 8.3-21: [2] Αντίληψη σχετικά με περιοχές, δεντροχάρτης.....	108
Εικόνα 8.3-22: Αντίληψη σχετικά με αποστάσεις και περιοχές, ραβδογραφήματα και πίτες.	108
Εικόνα 9.1.1.1-1 [13] Στιγμιότυπο οθόνης από την ιστοσελίδα του RAWGraphs	113
Εικόνα 9.1.1.1-2: [14] Στιγμιότυπο οθόνης από την ιστοσελίδα του ChartBlocks.....	114
Εικόνα 9.1.1.2-3: [15] Στιγμιότυπο οθόνης από την ιστοσελίδα του Tableau	115
Εικόνα 9.1.1.2-4: [16] Στιγμιότυπο οθόνης από την ιστοσελίδα του Power BI	118
Εικόνα 9.1.1.2-5: [17] Στιγμιότυπο οθόνης από την ιστοσελίδα του QlikView	120
Εικόνα 9.1.1.2-6: [18] Στιγμιότυπο οθόνης από την ιστοσελίδα του FineReport.....	121
Εικόνα 9.1.1.2-7: [19] Στιγμιότυπο οθόνης από την ιστοσελίδα του Datawrapper	122
Εικόνα 9.1.1.2-8: [20] Στιγμιότυπο οθόνης από την ιστοσελίδα του Visme.....	124
Εικόνα 9.1.1.3-9: [21] Στιγμιότυπο οθόνης από την ιστοσελίδα του Grow.....	126
Εικόνα 9.1.2-10: [22] Στιγμιότυπο οθόνης από την ιστοσελίδα του Infogram	127
Εικόνα 9.1.2-11: [23] Στιγμιότυπο οθόνης από την ιστοσελίδα του Visual	129
Εικόνα 9.1.3-12: [24] Στιγμιότυπο οθόνης από την ιστοσελίδα του InstantAtlas	130
Εικόνα 9.1.4-13: [25] Στιγμιότυπο οθόνης από την ιστοσελίδα του Gephi	131
Εικόνα 9.1.5-14: [26] Στιγμιότυπο οθόνης από την ιστοσελίδα του Wolfram Alpha	132
Εικόνα 9.2.1.1-15: [27] Στιγμιότυπο οθόνης από την ιστοσελίδα του ECharts 	133
Εικόνα 9.2.1.1-16: [28] Στιγμιότυπο οθόνης από την ιστοσελίδα του D3.js.....	134
Εικόνα 9.2.1.1-17: [29] Στιγμιότυπο οθόνης από την ιστοσελίδα του Plot.ly.	135

Εικόνα 9.2.1.1-18: [30] Στιγμιότυπο οθόνης από την ιστοσελίδα του Chart.js	136
Εικόνα 9.2.1.1-19 [31] Στιγμιότυπο οθόνης από την ιστοσελίδα του Google Charts.....	137
Εικόνα 9.2.1.1-20: [32] Στιγμιότυπο οθόνης από την ιστοσελίδα του Chartist.js	138
Εικόνα 9.2.1.2-21: [33] Στιγμιότυπο οθόνης από την ιστοσελίδα του Highcharts	139
Εικόνα 9.2.1.2-22: [34] Στιγμιότυπο οθόνης από την ιστοσελίδα του FusionCharts.....	140
Εικόνα 9.2.1.2-23: [35] Στιγμιότυπο οθόνης από την ιστοσελίδα του ZingChart	141
Εικόνα 9.2.2-24: [36] Στιγμιότυπο οθόνης από την ιστοσελίδα του Leaflet.....	142
Εικόνα 9.2.2-25: [37] Στιγμιότυπο οθόνης από την ιστοσελίδα του OpenLayers	144
Εικόνα 9.2.2-26: [38] Στιγμιότυπο οθόνης από την ιστοσελίδα του CARTO:	145
Εικόνα 9.2.3-27: [39] Στιγμιότυπο οθόνης από την ιστοσελίδα του Sigma	146
Εικόνα 9.2.4-28: [40]Στιγμιότυπο οθόνης από την ιστοσελίδα του dygraphs.....	147

Πρόλογος

Ο συνολικός όγκος των δεδομένων και πληροφοριών που παράγονται και καταναλώνονται παγκοσμίως, αυξάνεται με εκπληκτικό ρυθμό δημιουργώντας έτσι μια αυξανόμενη ανάγκη να γίνουν πιο εύπεπτα, κατανοητά και εφαρμόσιμα και σε αυτό στοχεύει η οπτικοποίηση δεδομένων. Γραφικές παραστάσεις μυριάδων σημείων δεδομένων, δίνουν τη δυνατότητα για βαθύτερη αναγνώριση σημαντικών αριθμών, τάσεων και μετρήσεων. Πίνακες ελέγχου (dashboards) και γραφικά με την κατάλληλη ανά περίπτωση μορφή βοηθούν, εκτός από τους επιστήμονες δεδομένων και τα τμήματα πληροφορικής, επιχειρηματίες, διαφημιστές αλλά και οργανισμούς. Η οπτικοποίηση δεδομένων θεωρείται πλέον ένα από τα πιο κρίσιμα εργαλεία για την επιτυχία τόσο απλών, όσο και σύνθετων δομών. Η δυσκολία να διαχειριστούν, να ερμηνεύσουν ή/και να οπτικοποιήσουν τους τεράστιους όγκους δεδομένων, η πολυπλοκότητα των διαθέσιμων πληροφοριών, αλλά και η εξαγωγή αυτών ώστε να είναι σχετικές και κρίσιμες, είναι ο βασικός λόγος της οπτικοποίησης.

Περίληψη

Η οπτικοποίηση παίζει σημαντικό ρόλο στην καλύτερη κατανόηση και διαχείριση της πληροφορίας, καθώς: α) ο ανθρώπινος εγκέφαλος λειτουργεί με αυτό τον τρόπο, β) βοηθά τους αναλυτές να ανακαλύψουν λάθη ή προβλήματα που τυχόν υπάρχουν στα δεδομένα ώστε να τα διορθώσουν, γ) τους βοηθά να κατανοήσουν σε βάθος τα δεδομένα ώστε να μπορέσουν να εξάγουν σωστά συμπεράσματα, και δ) ενδιαφερόμαστε να τραβήξουμε την προσοχή των αναγνωστών μας ώστε να τους περάσουμε το κατάλληλο μήνυμα. Έτσι, οι οπτικοποιήσεις μας πρέπει να είναι κατανοητές και προσβάσιμες από όλους. Πρέπει να στοχεύουμε σε δημιουργικές και αξιομνημόνευτες απεικονίσεις, με σωστό έλεγχο περιπλοκότητας, συνέπεια δίχως όμως να επαναλαμβανόμαστε. Σημαντικό ρόλο παίζει και το περιτύλιγμα, το μεγαλύτερο/γενικότερο πακέτο. Η κατηγοριοποίηση των τύπων πλοκής και η αντιμετώπιση οπτικών προβλημάτων που κάποια από αυτά έχουν αλλά συνήθως περνάνε απαρατήρητα, μπορεί να βοηθήσει στη σωστή γενικότερη χρήση τους και να κάνει τη διαφορά. Η επιλογή του σωστού εργαλείου για τις ιδιαίτερες ανάγκες μας, μπορεί να είναι πρόκληση. Η μορφή του τελικού αποτελέσματος είτε εικόνα, είτε διαδραστικό, επηρεάζει σε αυτό. Το ίδιο ισχύει και για τη διαφορετική προσέγγιση των (ζωντανών ή όχι) big data. Το χρώμα είναι σημαντικός παράγοντας στην οπτικοποίηση καθώς μπορεί να χρησιμοποιηθεί για να διαφοροποιεί τα δεδομένα μεταξύ τους, να αντιπροσωπεύει τιμές δεδομένων και να τονίζει. Η αντιμετώπιση συνηθισμένων λαθών χρήσης του, είναι επίσης σημαντική για το τελικό αποτέλεσμα. Η αισθητική επηρεάζεται επίσης από τίτλους, λεζάντες, υπομνήματα, άξονες, γραμμές και πίνακες. Η μορφή της πληροφορίας καθώς και η διαθέσιμη πηγή της, μπορεί να παίζει σημαντικό ρόλο. Χρειάζεται μια καλή κατανόηση των λύσεων που προσφέρονται όσον αφορά αφενός τον τρόπο λειτουργίας τους και την ικανοποίηση των αναγκών μας, και αφετέρου το διαθέσιμο προϋπολογισμό μας. Σε αυτό απάντηση δίνει το διαδίκτυο μέσω της διαρκώς επικαιροποιημένης πληροφορόφορησης σχετικά με εργαλεία και λύσεις στο χώρο της οπτικοποίησης.

Λέξεις κλειδιά: οπτικοποίηση δεδομένων, γραφήματα, διαγράμματα, απεικόνιση δεδομένων, τύποι πλοκής, χρώμα, μεγαλύτερο πακέτο, αφήγηση ιστορίας, αντίληψη, αναλογία μελανιού/δεδομένων, εργαλεία οπτικοποίησης δεδομένων, πίνακες ελέγχου, πολύπλοκες

οπτικοποιήσεις, αναπαραγωγή οπτικοποιήσεων, διαδραστικότητα, επισημάνσεις, γραμμές πλέγματος, κλίμακες χρωμάτων, μεγάλα / ζωντανά μεγάλα δεδομένα, μεγέθη χρωματικών στοιχείων, υπερκάλυψη, τίτλοι, λεζάντες, υπομνήματα, ετικέτες, άξονες, πίνακες, προβλήματα χρωματικής όρασης (CVD)

Abstract

Visualization plays an important role in better understanding and managing the information, since: a) the human brain works this way, b) it helps analysts discover errors or problems that may exist in the data to correct them, c) helps them to understand the data in depth so that they can draw the right conclusions, and d) we are interested in capturing the readers' attention in order to pass the appropriate message to them. Thus, our visualizations must be understandable and accessible to all. We must aim for creative and memorable visualizations, with proper control of complexity, and consistency but without repetition. The “bigger package”, also plays an important role. Categorizing plot types, and dealing with visual problems that some of them have but usually go unnoticed, can help in their proper general use and make a difference. Finding the right tool for our specific needs may be a challenge. The form of the final result, either image or interactive, affects our choice of tool. The same goes for (live or not) big data. Color is an important factor in visualization as it can be used to differentiate data from each other, to represent data values and to emphasize. Dealing with common color mistakes is also important for the end result. Aesthetics are also influenced by the use of titles, captions, legends, axes, lines and tables. The format of the given information, as well as its available source, can play an important role. We need a good understanding of the solutions offered in terms of how well they work for us and satisfy our needs, and the available budget. The internet has a solution to this through the constantly updated information on tools and solutions in the field of visualization.

Keywords: data visualization, graphs, diagrams, plot types, color, bigger package, storytelling, perception, ink / data ratio, data visualization tools, dashboards, complex visualizations, visualization reproduction , interactivity, labels, grid lines, color scales, big / live big data, color element sizes, overplot, titles, captions, legends, labels, axes, tables, color vision deficiencies (CVD)

1 Γιατί Κάνουμε Οπτικοποίηση;

Ένας από τους βασικούς λόγους που κάνουμε οπτικοποίηση δεδομένων, είναι για να μπορέσουμε να τραβήξουμε την προσοχή των πολυάσχολων αναγνωστών μας ώστε **να μπορέσουμε να τους περάσουμε το μήνυμα που μας ενδιαφέρει εύκολα, γρήγορα και αξιωματικώς**. Για να το καταφέρουμε αυτό, το πρώτο που πρέπει να κάνουμε είναι να σχεδιάσουμε ένα σωστό διάγραμμα φτιαγμένο έτσι, ώστε να είναι καθαρό, πιστικό και οπτικά ελκυστικό., τόσο, ώστε να το θυμάται ο αναγνώστης και να ενεργεί επάνω στα ευρήματά μας. Και για να έχει αυτό το αντίκτυπο που ψάχνουμε, πρέπει πρώτα να απεικονιστεί.

Η οπτικοποίηση δεδομένων αποτελείται από δύο πράγματα: την τέχνη και την επιστήμη. Η πρόκληση σε αυτό είναι, να συνδυάσουμε αυτά τα δύο μαζί έτσι ώστε να δημιουργείται κάτι αισθητικά ευχάριστο που να μεταφέρει όμως με ακρίβεια τα δεδομένα χωρίς αυτά να παραπλανούν ή να παραμορφώνονται.

Οι άνθρωποι πιστεύουν ότι γνωρίζουν τι είναι οπτικοποίηση δεδομένων, επειδή μπορούν να δημιουργήσουν προεπιλεγμένα γραφήματα (όπως Bar Charts, κτλ) από κάποιο λογισμικό που χρησιμοποιούν, χωρίς να πολυσκεφτούν ούτε το σχεδιασμό, αλλά ούτε και το μήνυμα που μεταφέρουν. Αυτό που χρειάζεται, είναι κριτική σκέψη και έμπνευση.

Ένας καλός τρόπος έμπνευσης ή εκμάθησης είναι μελετώντας άλλα πετυχημένα διαγράμματα με ποιο τρόπο ξεχωρίζουν και πως περνάνε το μήνυμα που θέλουν. Τέτοια παραδείγματα βρίσκονται παντού, στις κινούμενες εικόνες (animations), σε διαδραστικές ιστοσελίδες (interactive animations), ακόμα και σε στατικές σελίδες στον τύπο. Με το να δώσουμε στους αναγνώστες έναν τίτλο, δεν έχουν λόγο να μας πιστέψουν. Πρέπει να υποστηρίζεται από δεδομένα (που να δοθούν έτσι, ώστε το μήνυμά μας να γίνει πιο πιστευτό). Θα μπορούσε να γίνει με άλλους τρόπους, όπως να το κάνουμε με λέξεις, αλλά θα ήταν κάπως βαρετό και επαναλαμβανόμενο, ή με ένα πίνακα, αλλά θα ήταν αρκετά μεγάλος και λίγοι θα έδιναν σημασία. Κάνοντάς το με μια απεικόνιση όμως, θα ήταν συμπαγές, προσβάσιμο σε όλους, εντυπωσιακό ώστε να τραβήξει τα βλέμματα και το μήνυμα θα μεταδοθεί γρήγορα εάν γίνει με το σωστό τρόπο (δηλαδή εάν είναι επιστημονικό, αντικειμενικό, έχει ενδιαφέρον αλλά είναι και οπτικά ελκυστικό).

Μια εικόνα είναι πιο ισχυρή από αριθμούς, ο εγκέφαλός μας είναι φτιαγμένος με αυτό τον τρόπο. Ο εγκέφαλός μπορεί να δει και να κατανοήσει διαφορές, πρότυπα και χωρικές σχέσεις τόσο

αβίαστα, ώστε η συνειδητή μας σκέψη να είναι ελεύθερη για πιο σημαντικά πράγματα, όπως στη δική μας περίπτωση, το αν θα πιστέψουν αυτό που θέλουμε να τους πούμε. Για παράδειγμα, το πέρασμα στο απέναντι πεζοδρόμιο, απαιτεί υπολογισμό της απόστασης των αυτοκινήτων που περνάνε, της ταχύτητας που αυτά έχουν και άλλων μικροστοιχείων που υπάρχουν εκείνη τη στιγμή τριγύρω, έτσι ώστε να μπορέσουμε να αποφασίσουμε αν είναι ασφαλές να περάσουμε ή όχι, και αυτό γίνεται σε κλάσματα δευτερολέπτου.

Μερικές φορές το μοτίβο μπορεί να είναι πολύ δύσκολο για να δειχθεί με έναν πίνακα, αλλά αρκετά σαφές αν δειχθεί με μια απεικόνιση, χρησιμοποιώντας τον πίνακα για να το υποστηρίξει για όσους ενδιαφέρονται να μάθουν περισσότερα. Η τοποθέτηση πληροφοριών σε επίπεδα, βοηθά όλων των ειδών τους αναγνώστες ως εξής:

- **η επικεφαλίδα:** για όσους έχουν περιορισμένο χρόνο και θέλουν να μάθουν την ουσία γρήγορα, εν συντομία και με εμφανή τρόπο χωρίς πολύ ψάξιμο
- **η οπτικοποίηση:** για τους ειδικούς που καταλαβαίνουν τα μοτίβα
- **οι πίνακες:** για όσους είναι σκεπτικιστές και θέλουν να δουν πώς καταλήξαμε σε αυτά τα συμπεράσματα

Οι οπτικοποιήσεις είναι επίσης πολύ σημαντικές και για τους ίδιους τους αναλυτές. Μια καλή πρακτική που χρησιμοποιούν, είναι να δημιουργούν πολλές οπτικοποιήσεις ώστε να μπορέσουν να κατανοήσουν τα δεδομένα τους σε βάθος και να εντοπίσουν έτσι σφάλματα και ανωμαλίες (Εικόνα 1-1, δεξιά). Αυτό είναι εξαιρετικά σημαντικό, ειδικά στα αρχικά στάδια της ανάλυσης των δεδομένων. Έχοντας μόνο τα στατιστικά στοιχεία, καταλαβαίνουμε συνήθως τη γενική συμπεριφορά, αλλά καμιά φορά χάνουμε τη λεπτομέρεια. Μπορούμε να το καταλάβουμε καλύτερα αυτό, κοιτώντας το παράδειγμα των τεσσάρων τεχνητών συνόλων δεδομένων που δημιουργήθηκε από τον στατιστικό Frank Anscombe στην Εικόνα 1-1(αριστερά), παρατηρώντας ότι παρόλο που έχουν τις ίδιες τυπικές αποκλίσεις, τα ίδια μέσα και τις ίδιες συσχετίσεις, είναι τελείως διαφορετικά μεταξύ τους. Αν δεν είχαν απεικονιστεί, θα κατέληγε κανείς στο συμπέρασμα ότι είναι πανομοιότυπα.

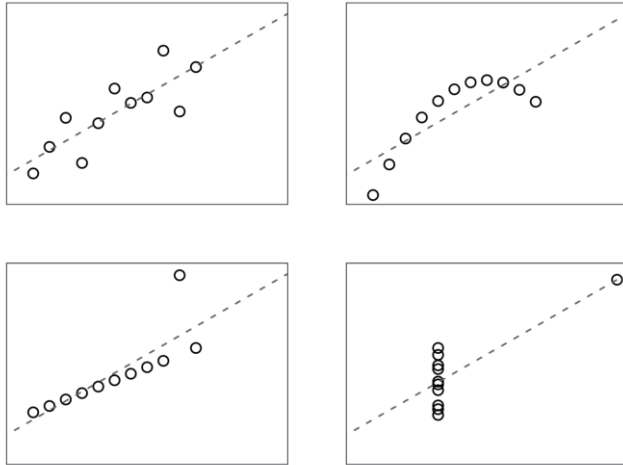


Figure 1.3 Anscombe's quartet: each plot shows a dataset with exactly the same statistics, masking great differences.

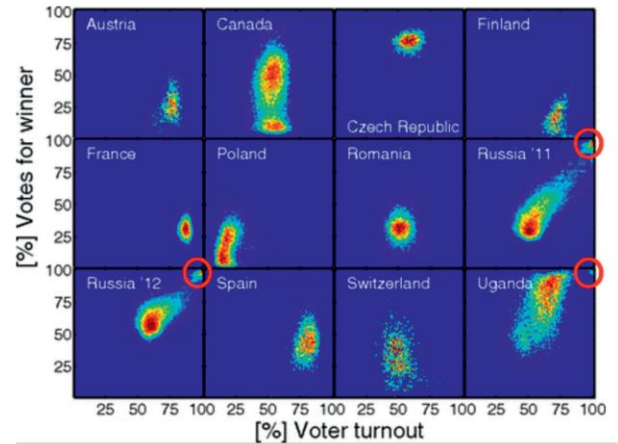


Figure 1.4 Voter turnout and winner support. Possible irregularities are circled in red. From “Statistical detection of systematic election irregularities” by Peter Klimek, Yuri Yegorov, Rudolf Hanel, and Stefan Thurner. Reproduced under open access publication license from the *Proceedings of the National Academy of Sciences*.

Εικόνα 1-1: [1] Πολλαπλές οπτικοποιήσεις για καλύτερη κατανόηση

Οι στατιστικοί Andrew Gelman και Antony Unwin θεωρούν ότι οι λόγοι για να γίνει οπτικοποίηση των δεδομένων είναι οι έξι παρακάτω:

1. για μια δώσουμε μια επισκόπηση και να δείξουμε τη γενική εικόνα
2. για να δείξουμε την κλίμακα και την πολυπλοκότητα των δεδομένων
3. για εξερεύνηση των δεδομένων
4. για την ανακοίνωση των ευρημάτων/αποτελεσμάτων
5. για να πούμε μια ιστορία
6. για να προσελκύσουμε την προσοχή και να κεντρίσουμε το ενδιαφέρον του αναγνώστη

Όταν έχουμε κάποιο εύρημα/διαπίστωση να δείξουμε ή μια σύσταση να κάνουμε, αυτό πρέπει να γίνει λέγοντας μια ιστορία. Και εάν υπάρχει η ανησυχία ότι δεν θα κοιτάζουν τη δουλειά μας, τότε πρέπει να τραβήξουμε την προσοχή τους. Εάν το καταφέρουμε αυτό δίνοντας το σε επίπεδα όπως προαναφέρθηκε, τους επιτρέπουμε να διεισδύσουν όσο βαθιά θέλουν οι ίδιοι (αυτό μπορεί να γίνει και με τις διαδραστικές απεικονίσεις όπου διαφορετικές ανάγκες αναγνωστών μπορούν να καλυφθούν σε ένα μόνο σημείο).

Τα infographics είναι εντυπωσιακά διαγράμματα που μεταδίδουν/μεταφέρουν το μήνυμα, είναι πολύ δημοφιλή στο χώρο της διαφήμισης και πείθουν ως επιστημονικές, αξιόπιστες πληροφορίες,

όμως δεν είναι το ίδιο με την οπτικοποίηση δεδομένων. Τα infographics συνήθως μεταδίδουν/μεταφέρουν κάποια νούμερα, δεν χρησιμοποιούν οπτικές παρουσιάσεις για να επιτρέψουν στους αναγνώστες να κάνουν τις δικές τους συγκρίσεις.

Όταν οι απεικονίσεις είναι μπερδεμένες ή δυσνόητες, τις περισσότερες φορές μπορεί να γίνει η οπτικοποίηση τους με διαφορετικό/καλύτερο τρόπο. Αυτός ο τρόπος όμως, δεν είναι πάντα προφανής και θα πρέπει σε αυτές τις περιπτώσεις να ζυγίσουμε τα πλεονεκτήματα με τα μειονεκτήματα και να διαλέξουμε την καλύτερη λύση. Ερωτήσεις που μπορούν να μας βοηθήσουν προς αυτήν την κατεύθυνση, είναι οι παρακάτω:

- Πόσες διαφορετικές παρατηρήσεις και μεταβλητές χρειάζεται να δείξουμε;
- Θα δείξουμε τα μεμονωμένα δεδομένα, κάποια σύνολά αυτών, ή κάποια στατιστικά τους στοιχεία;
- Οι μεταβλητές είναι ονομαστικές (nominal), κανονικές (ordinal) ή κλίμακες;
- Θέλουμε οι αναγνώστες να επικεντρωθούν σε κάποια σύγκριση, ή σε κάποια αλλαγή στα δεδομένα;
- Σε ποιες μορφές αρχείων (formats) είναι πιθανό να είναι εξοικειωμένοι οι αναγνώστες μας;

Όταν μιλάμε για οπτικοποίηση, δεν μπορούμε να πούμε ότι υπάρχει απλή απάντηση για το ποια είναι η «καλύτερη» απεικόνιση για να οπτικοποιήσουμε τα δεδομένα μας. Αυτό συμβαίνει γιατί αυτή εξαρτάται από πολλά πράγματα, όπως: από το περιεχόμενο, από το τι γνωρίζουν ήδη οι αναγνώστες, από το πόσο επιστημονικά καταρτισμένοι είναι, από το ποιες μορφές θεωρούνται συνηθισμένες στον συγκεκριμένο τομέα/πεδίο στον οποίο εργαζόμαστε, από το μέσο που χρησιμοποιούμε, από το προσωπικό μας στυλ που σχεδιάζουμε και ούτω καθεξής. Και αυτό είναι που το κάνει εν μέρη δύσκολο, αλλά ταυτόχρονα και συναρπαστικό. Τα ίδια δεδομένα οπτικοποιημένα από δύο διαφορετικούς ανθρώπους μπορούν να φτιάξουν δύο τελείως διαφορετικές απεικονίσεις, εξίσου χρήσιμες και έγκυρες.

Ας πούμε για παράδειγμα, ότι έχουμε μια συλλογή ποσοστών. Αυτό από μόνο του δεν μας κατευθύνει για το ποια οπτικοποίηση να χρησιμοποιήσουμε. Πρέπει να έχουμε μια σαφή ιδέα για το ποιο είναι το μήνυμα που θέλουμε να περάσουμε (καθώς είναι πολύ εύκολο να δημιουργηθεί σύγχυση εάν προσπαθήσουμε να στριμώξουμε/ συμπεριλάβουμε πολλά μηνύματα μέσα), ποια

μέρη των δεδομένων αποδεικνύουν τον ισχυρισμό μας και ποια άλλα κομμάτια πρέπει να εμφανιστούν για αντίθεση/πλαισίωμα.

Υπάρχουν κάποιοι απλοί κανόνες και αρχές για το σχεδιασμό καλών απεικονίσεων: ενίσχυση του μηνύματος, μη ενοχλητικά στοιχεία, χαρακτηριστικά που δεν αποσπούν την προσοχή, μη τρανταχτά χρώματα κτλ.

Επιπρόσθετα, βοηθά το να δίνουμε προσοχή στις μικρές λεπτομέρειες καθώς επίσης και στις οπτικές επιλογές που έχουν κάνει άλλοι άνθρωποι σε αντίστοιχες καταστάσεις. Το εάν θα τις υιοθετήσουμε ή όχι, το αποφασίζουμε αργότερα.

Υπάρχουν αρκετοί ιστότοποι που δείχνουν την τυπολογία απεικονίσεων και σε τι χρησιμεύει ο καθένας, αλλά ο καλύτερος τρόπος είναι να αντιμετωπίσουμε κάθε πρόβλημα οπτικοποίησης ξεχωριστά, σκεπτόμενοι τι πρέπει να δείξουμε και πώς να το μορφοποιήσουμε.

Σημαντικό επίσης είναι, εάν γνωρίζουμε πως να το δημιουργήσουμε ή αν πρέπει να προσλάβουμε κάποιον.

Στην περίπτωση που δεν μπορούμε οι ίδιοι, η πρόσληψη του κατάλληλου ανθρώπου που γνωρίζει πως να το κάνει, συνήθως είναι η καλύτερη λύση. Στην περίπτωση που προσλάβουμε κάποιον/α, θα πρέπει να μπορούμε να ακούμε και να αποδεχόμαστε αυτά που μας λέει και να μην πιέζουμε/αναγκάζουμε να υλοποιηθεί η δική μας σκέψη. Στην περίπτωση που θέλουμε να κάνουμε οι ίδιοι απεικονίσεις για άλλους, τότε μερικές συμβουλές που μπορεί να βοηθήσουν είναι:

- Θα πρέπει να μάθουμε τη μορφή των δεδομένων (πως είναι τα δεδομένα) πριν ξεκινήσουμε. Θα χρειαστούν ανάλυση; Εάν ναι, μπορούμε να το κάνουμε οι ίδιοι ή θα πρέπει να προσλάβουμε κάποιον να το κάνει για εμάς; Τα δεδομένα αυτά, είναι από την αρχή που θα μας δοθούν σωστά επισημασμένα και ορισμένα ή/και κωδικοποιημένα με συνέπεια; Αυτά, θα πρέπει να τα ζητήσουμε πριν δώσουμε προσφορά και αν χρειάζεται να υπογράψουμε ακόμα και συμφωνητικό εμπιστευτικότητας. Τα παραπάνω, δείχνουν σοβαρότητα και διάθεση να βοηθήσουμε τον εκάστοτε οργανισμό.
- Θα πρέπει να μας έχουν δοθεί επίσης τα χρονοδιαγράμματα, ο προϋπολογισμός, οι προδιαγραφές και οι όποιες απαιτήσεις εμπιστευτικότητας συμφωνούνται, γραπτώς. Οποιαδήποτε πολιτική επικοινωνίας ή κατευθυντήριες γραμμές για την επωνυμία του

οργανισμού που μας προσλαμβάνει (branding guidelines), θα πρέπει να το έχουμε σε αντίγραφο και να βεβαιωθούμε ότι θα το τηρήσουν.

- Θα πρέπει να προτείνουμε ένα πρόγραμμα δοκιμών από χρήστες (user-testing) και ανατροφοδότησης και να υπάρξει σχετική συμφωνία για τις ώρες συνάντησης που θα γίνει αυτό. Το να εντοπίσουμε τα προβλήματα είναι μέρος της διαδικασίας σχεδιασμού και μπορούμε αυτές τις συναντήσεις να τις χρησιμοποιήσουμε για να έρθουμε σε μια συμφωνία πώς θα πρέπει αυτά να αντιμετωπιστούν.
- Σε όλη τη διάρκεια της σχέσης μας με τους πελάτες μας, θα πρέπει να θυμόμαστε ότι η σχέση αυτή είναι επαγγελματική. Θέλουν να δουν εξαιρετική δουλειά από εμάς και να τους πούμε την άποψή μας για το αν κάτι είναι καλή ή κακή ιδέα. Ίσως κάποιες φορές υπάρξει πίεση για μια κακή ιδέα από τη μεριά τους, μέχρι όμως να πειστούν για το αντίθετο. Δουλειά μας είναι να δώσουμε άμεσες/καθαρές απαντήσεις. Δεν πρέπει να μπούμε στον πειρασμό να συμβιβαστούμε απλά και μόνο για να εξασφαλίσουμε τη δουλειά. Πρέπει να είμαστε προετοιμασμένοι να αποχωρήσουμε από μια συμφωνία η οποία είναι πιθανό να μας προκαλέσει πονοκεφάλους. Με παρόμοιο τρόπο, μπορεί οι πελάτες μας να ζητήσουν να «πειράζουμε» τα στοιχεία λίγο ώστε να ευνοηθεί η δική τους οπτική/άποψη. Θα πρέπει να τους εξηγήσουμε και να τους βοηθήσουμε να καταλάβουν γιατί κάτι τέτοιο δεν είναι καλή ιδέα. Πέρα από το ότι με αυτό τον τρόπο παραπλανούμε τους πιθανούς αναγνώστες, ίσως και να τους κάνουμε κακό, η ζημιά που θα προκληθεί στην φήμη και των δύο, δεν αξίζει τον κόπο.

Επίσης σχετικά με τις ανάγκες για δημιουργία γραφικών για τρίτους, θα χρειαστεί και εξοικείωση με πράγματα όπως:

- Ένα λογισμικό για στατιστικά και επιστήμη δεδομένων το οποίο μπορεί να προγραμματίζεται, όπως το Strata, ή μια γλώσσα προγραμματισμού που εστιάζει στα δεδομένα, όπως η R ή η Python.
- Κάποιο λογισμικό γρήγορης απεικόνισης/οπτικοποίησης, όπως το Tableau
- Επεξεργασία SVG αρχείων σε επεξεργαστή κειμένου όπως επίσης και σε κάποιο γραφικό περιβάλλον όπως Inkscape ή Illustrator.
- Σχεδίαση με το χέρι

- Σχετικές βιβλιοθήκες Javascript για οπτικοποίηση δεδομένων (αυτές αλλάζουν συνέχεια, αλλά η D3 και το Leaflet είναι μια επιλογή . Μια αναζήτηση στο διαδίκτυο είναι απαραίτητη ώστε να δούμε τη στιγμή που μας ενδιαφέρει ποιες είναι ιδανικές για τις ανάγκες μας. (Ενδεικτικά, κάποια τέτοια παραδείγματα αναζήτησης: <https://blog.logrocket.com/top-javascript-data-visualization-libraries-2021/> , <https://www.monterail.com/blog/javascript-libraries-data-visualization> καθώς και πολλά άλλα.)
- Ένα online εργαλείο χαρτογράφησης όπως το Mapbox
- Ένα ευέλικτο εργαλείο για Big data και Fast data όπως το Spark

Κατά τη διαδικασία της απεικόνισης, μπορεί να χρειαστεί να πετάξουμε πολλά σκίτσα πάνω στον πειραματισμό μας και στην προσπάθειά μας να βρούμε τρόπους να γίνει αυτή πιο ελκυστική και ουσιαστική, μέχρι να καταλήξουμε σε κάποιες καλές επιλογές. Στη συνέχεια, τα υλοποιούμε με το κατάλληλο λογισμικό και προχωράμε σε user-testing ώστε να επιλεγθεί η τελική μορφή (γιατί πρέπει να αποφύγουμε την παγίδα ότι επειδή έχει νόημα για εμάς, θα έχει και για τους άλλους).

Θα πρέπει να βελτιώσουμε το μήνυμα (να μην χρησιμοποιούμε ενοχλητικά χρώματα, μη ισορροπημένα οπτικά στοιχεία, ή χαρακτηριστικά που αποσπούν την προσοχή, και θα πρέπει να αποφεύγουμε τη χρήση τεχνικών όρων που μπερδεύουν για επισήμανση (labeling)).

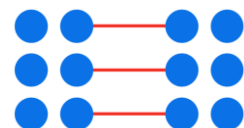
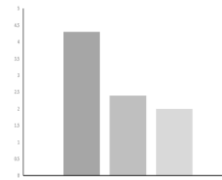
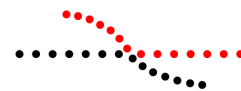
Επίσης, τα σχήματα πρέπει να είναι μπορούν να αναπαραχθούν αυτόματα, έτσι ώστε με οποιαδήποτε ενημέρωση δεδομένων να μπορούν να αναδημιουργούνται, ανεξάρτητα από το ποιος ή πώς το δημιούργησε αρχικά. Με αυτόν τον τρόπο μειώνουμε την πιθανότητα να κάνουμε λάθη (όπως για παράδειγμα, λόγω του ότι δεν θυμόμαστε ποια ήταν η τελευταία έκδοση για να γίνει η νέα αλλαγή).

2 Εγκέφαλος και Οπτική Αντίληψη

Από τα πολύ παλιά χρόνια, ο ανθρώπινος εγκέφαλος επεξεργάζεται οπτικές πληροφορίες για την εύρεση τροφής ή/και για την αποφυγή από αρπακτικά. Γενικά, έχουμε την τάση να παρατηρούμε μόνο το ευρύτερο περίγραμμα του περιβάλλοντός μας, εκτός από ένα ή δύο πράγματα που τραβούν την προσοχή μας. Γι' αυτό το λόγο, θα πρέπει να αποφεύγουμε τον οπτικό θόρυβο που βάζει πιο πολύ δουλειά στο μυαλό των αναγνωστών για επεξεργασία, το οποίο δημιουργεί μεγαλύτερη πιθανότητα: 1) εγκατάλειψης της προσπάθειά τους ή 2) να ερμηνεύσουν λάθος το μήνυμά μας. Η συμβουλή του στατιστικού Edward Tufte για την αναλογία δεδομένων/μελανιού, είναι κάτι που πρέπει να λαμβάνουμε υπόψιν μας, αλλά δεν σημαίνει ότι το απόλυτο ελάχιστο είναι πάντα η καλύτερη λύση. Η αναφορά - για παράδειγμα - από πού προήλθαν τα δεδομένα, είναι ζωτικής σημασίας.

Οι καλές οπτικοποιήσεις δεδομένων βασίζονται στις αρχές της Gestalt:

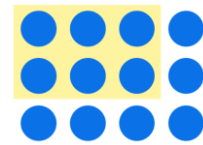
- **Εγγύτητα** (είναι η αντίληψη ότι στοιχεία που βρίσκονται κοντά μεταξύ τους θεωρούνται ότι ανήκουν στην ίδια ομάδα αντικειμένων),
- **Ομοιότητα** (είναι η αντίληψη ότι στοιχεία που μοιάζουν μεταξύ τους στο χρώμα, το μέγεθος, τα χαρακτηριστικά τους ότι ανήκουν στην ίδια ομάδα αντικειμένων),
- **Συνέχεια** (κάνοντας χρήση γραμμών μπορούμε να τραβήξουμε και να καθοδηγήσουμε τον αναγνώστη προς τα σημαντικά στοιχεία μέσα σε ένα γράφημα),
- **Κλείσιμο** (δοθείσας της ευκαιρίας το μυαλό μας θα επιλέγει την απλότητα, συμπληρώνοντας έτσι τα κενά ή συνδέοντας τις τελείες για να ολοκληρώσει σχήματα ή εικόνες που υπονοούνται. Αυτός είναι και ο λόγος που απαιτούνται μόνο δύο άξονες για να καθοριστεί ο χώρος στον οποίο εμφανίζονται τα δεδομένα και όχι πλήρες κλείσιμο με περίγραμμα),
- **Συνδεσιμότητα** (είναι η τάση να ομαδοποιούμε στοιχεία όταν αυτά συνδέονται μεταξύ τους. Στην οπτικοποίηση δεδομένων, το πιο προφανές θα ήταν, δείκτες ή γραμμές κοντινές μεταξύ τους και



Εικόνα 2-1: [3] [4] Αρχές Gestalt

διαχωρισμός τους από άλλα πιο μακρινά, κωδικοποίηση ορισμένων μεταβλητών ως τοποθεσία ή μήκος, αλλά και αναγνώριση μιας ομάδας στοιχείων από το ξεχωριστό της χρώμα, και σχήμα (αν μιλάμε για σχήματα), ή από το μοτίβο και πάχος τους (αν μιλάμε για γραμμές), το να περικλείονται σε σκιασμένες περιοχές, ακόμη και σχεδιασμός λεπτών γραμμών που να συνδέουν αντικείμενα που μας ενδιαφέρουν όποτε είναι δυνατόν χωρίς βέβαια να δημιουργούν οπτικό θόρυβο ή υπερφόρτωση του αναγνώστη. Θα πρέπει να υπάρχει μόνο μία συνδεδεμένη ομάδα σε μια οπτικοποίηση, το να υπάρχουν περισσότερες μπορούν να οδηγήσουν σε αποτυχία. Εάν υπάρχουν περισσότερα μηνύματα από ένα, θα πρέπει να σκεφτούμε τη χρήση πολλών οπτικοποιήσεων ή ενός διαδραστικού (interactive).

- **Περίφραξη** (είναι η αντίληψη ότι αντικείμενα σε μια περιοχή που διαφέρει από τα υπόλοιπα όσα βλέπουμε είναι ξεχωριστά)



Μερικές βασικές αρχές που πρέπει να έχουμε υπόψη μας είναι:

Εικόνα 2-2: [3] Αρχές Gestalt

1. να μεταφέρουμε το **μήνυμα γρήγορα και εύκολα**, κάνοντας ξεκάθαρο το τί μήνυμα θέλουμε να περάσουμε, χωρίς περιθώρια για παρανοήσεις (οπτικός θόρυβος)
2. να παρουσιάζουμε **γραφήματα που μπορούν να κατανοήσουν**
3. η **αναλογία δεδομένων/μελανιού** να προσπαθούμε να είναι **όσο πιο υψηλή γίνεται**

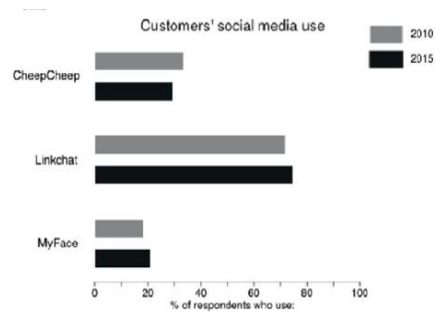
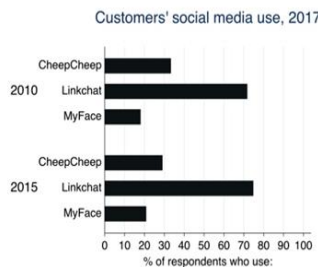
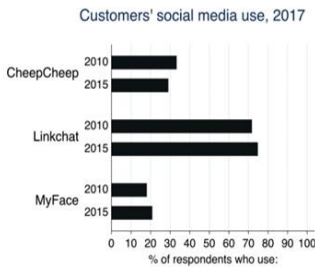


Figure 4.6 Clustered bar charts comparing three binary variables over time; notice how the small changes are easier to see in the left chart, and easier than in the stacked version above.

Figure 7.1 A version of Figure 4.6 with a better data-to-ink ratio.

Εικόνα 2-3: [1] Αναλογία δεδομένων/μελανιού

4. η **περιοχή** που εμφανίζονται τα δεδομένα να είναι **ελεύθερη από θόρυβο**
5. Γενικός κανόνας: **Όσο λιγότερα, τόσο καλύτερα**. Η απλοποίηση των γραφημάτων όσο το δυνατόν περισσότερο με μόνο τα σημαντικά σημεία να παραμένουν, βοηθά στην αποφυγή της απόσπασης της προσοχής. Μια καλή απεικόνιση δεν χρειάζεται εξηγήσεις, παρόλ’

αυτά στις περισσότερες περιπτώσεις κάποια εξήγηση και καθοδήγηση είναι απαραίτητη. Για πολύ περίπλοκες οπτικοποιήσεις η χρήση ενός κειμένου όπως "πώς να διαβάσετε αυτό το γράφημα", βοηθά τον αναγνώστη στο να το καταλάβει.

6. Το να καταλαβαίνει κάποιος μια οπτικοποίηση και το να την θυμάται, είναι δύο τελείως διαφορετικά πράγματα. Κάποιες έρευνες έχουν δείξει ότι ο/η **οπτικός/η θόρυβος/ακαταστασία μπορεί να βοηθήσει στην απομνημόνευση**, αλλά θα πρέπει να γίνεται με πολύ προσοχή.



Figure 7.2 Examples of pre-attentive features.

Εικόνα 2-4: [1] Χρήση επισημάνσεων

7. Να χρησιμοποιούμε **προειδοποιητικά μηνύματα** χωρίς όμως να προσθέτουμε θόρυβο στο γράφημα. Αυτό γίνεται με τη χρήση ελάχιστων διακριτικών λειτουργιών όπως επισημάνσεις με χρώμα, μέγεθος, σκίαση κτλ.

Έννοιες που θα χρειαστεί να γνωρίζουμε:

Τρεμούλιασμα (Jittering): παίρνει αντικείμενα που συνυπάρχουν στο ίδιο σημείο και μπερδεύουν για το πόσα είναι σε αυτό, και τα μετακινεί κατά μικρές τυχαίες ποσότητες έτσι ώστε να είναι πιο ξεκάθαρος ο αριθμός τους.

Εξομάλυνση (ας πούμε το αντίθετο του jittering): πολλές πληροφορίες συνοψίζονται σε μία πιο απλή εντύπωση (όπως στην Εικόνα 2-5 αριστερά), όπου μια καμπύλη σε ένα διάγραμμα πολλαπλών γραμμών, δείχνει μια σύνοψη του μοτίβου τους. Ουσιαστικά συμβιβάζομαστε, κάμπτοντας τη φυσική γραμμή των δεδομένων, με στόχο/σκοπό να βελτιώσουμε την κατανόηση τους από τον αναγνώστη.

Ομαδοποίηση άκρων (Edge bundling): γραμμές που συνδέουν σημεία, τραβιούνται τεχνητά, αν ξεκινούν και τελειώνουν σε παρόμοια μέρη, για να μειωθεί το φαινόμενο των «μακαρονιών» (Εικόνα 2-5 δεξιά).

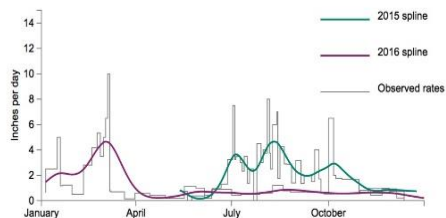


Figure 7.3 Observed rates of bird seed consumption in my garden, and smoothed lines through the data using splines. robertgrantstats.co.uk/dataviz/birdfeeders

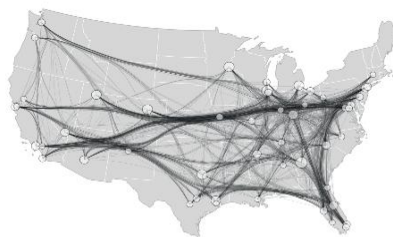


Figure 13.10 A network showing airline routes within the United States, using edge bundling and semi-transparency to clarify the patterns and avoid the spaghetti effect. For example, flights from San Diego to Houston initially curve north so that they can join with other flights from the west coast; this is not what happens in reality. By Sophie Eagle, reproduced with permission and under GPL-3 license.

Εικόνα 2-5: [1] Εξομάλυνση, ομαδοποίηση άκρων

Ημιδιαφάνεια ή αδιαφάνεια: όταν τα στοιχεία (markers) καλύπτει το ένα το άλλο, δημιουργούν ή όχι, ένα πιο σκοτεινό αποτέλεσμα σε σύγκριση με τα υπόλοιπα στοιχεία, επιτρέποντας έτσι την απορρόφηση περισσότερων πληροφοριών.

Όλα τα παραπάνω, γίνονται αντιληπτά ως ερεθίσματα από τον εγκέφαλο, όμως δεν έχουν όλα τα ερεθίσματα το ίδιο αποτέλεσμα. Ο ψυχολόγος Stanley Smith Stevens, παρατήρησε ότι:

- ο διπλασιασμός του μήκους ενός στοιχείου, γίνεται αντιληπτός σωστά, δηλαδή καταλαβαίνουμε ότι πρόκειται για τη διπλάσια ποσότητα ενός αρχικού μεγέθους.
- ο διπλασιασμός μιας περιοχής γίνεται αντιληπτός σαν αύξηση κατά 1,6 φορές παραπάνω από το αρχικό μέγεθος (υποτιμείται δηλαδή), ενώ
- ο διπλασιασμός της ερυθρότητας ενός χρώματος, γίνεται αντιληπτός ως 3,2 φορές μεγαλύτερος από το αρχικό μέγεθος (υπερεκτιμείται).

Αυτός είναι και ο λόγος για τον οποίο οι ειδικοί οπτικοποίησης δεδομένων δεν τους αρέσει/προτιμούν να κωδικοποιούν πράγματα με χρώμα ή με περιοχή εκτός αν είναι τακτικά (σε σειρά) (ordinal), ή αν δεν μας πειράζει εάν κάποιος τα θεωρήσει ως τέτοια.

Κάτι άλλο που πρέπει να έχουμε κατά νου όταν δημιουργούμε μία οπτικοποίηση, είναι ότι δεν έχουν όλα μία καθολική σημασία και ότι ο αναγνώστης δεν καταλαβαίνει πάντα την κωδικοποίηση που βασίζεται πάνω σε πολιτισμική κωδικοποίηση. Το κόκκινο χρώμα για παράδειγμα, σε ορισμένα μέρη θεωρείται επικίνδυνο, ενώ σε άλλα, θεωρείται πολλά υποσχόμενο. Η ανάγνωση ωστόσο του οριζόντιου χρονολογικού άξονα σε ένα διάγραμμα οπτικοποίησης από αριστερά προς τα δεξιά, θεωρείται καθολική, ακόμη και σε αλφάβητα τα οποία ξεκινούν από δεξιά

προς τα αριστερά όπως τα αραβικά. Επίσης, υπάρχουν μορφές απεικόνισης που θεωρούνται από μόνες τους ενδείξεις ερμηνείας των δεδομένων με συγκεκριμένο τρόπο. Για παράδειγμα, συνδέσεις μεταξύ σημείων δεδομένων με το ίδιο στυλ που χρησιμοποιούμε για απεικόνιση ενός χάρτη του μετρό, δεν έχουν για όλους νόημα εκτός και αν απευθύνονται αποκλειστικά σε κατοίκους πόλεων ή χημικούς. Μπορεί τέτοιες μορφές να είναι δημιουργικές και διασκεδαστικές, αλλά αν δεν γίνουν σωστά, μπορεί να θεωρηθούν από κάποιους ότι δεν καταλάβαμε ή δεν κατανοήσαμε αυτό που πήγαμε να μιμηθούμε. Εδώ για παράδειγμα, αντιπροσωπεύουν κάτι οι αποστάσεις μεταξύ των στάσεων του μετρό; Υπάρχει κάποια περιοδικότητα;

Το πως αντιλαμβάνεται κάποιος μια απεικόνιση, μπορεί να επηρεαστεί από πράγματα όπως:

- κομμάτια χρώματος: μπορεί να γίνουν αφορμή να φαίνονται οι γραμμές κεκλιμένες ενώ δεν είναι (Εικόνα 2-6 αριστερά)
- κυματιστές γραμμές: να επηρεάσουν την αντίληψή μας και να εμφανίζονται ψηλότερες ή πλατύτερες (Εικόνα 2-6 αριστερά)
- εσωτερικοί κύκλοι: ανάλογα με τα αντικείμενα του περιβάλλοντός τους, μπορεί να δώσουν μια αίσθηση διαφορετικού μεγέθους (Εικόνα 2-6 κέντρο)
- γραμμές εντός ή εκτός σκιασμένων περιοχών: μπορεί να γίνουν αντιληπτές ως λυγισμένες (Εικόνα 2-6 δεξιά).

Η επιρροή της αντίληψής μας σε τέτοιο επίπεδο είναι πολύ σημαντική, καθώς το μέγεθος ή το χρώμα ενός στοιχείου της απεικόνισης (όπως για παράδειγμα σε κάποιο bubble plot) θα μπορούσε να αντιπροσωπεύει κάποια μεταβλητή, και χρησιμοποιώντας τα, μπορεί να επηρεάσει λανθασμένα την αντίληψή μας.

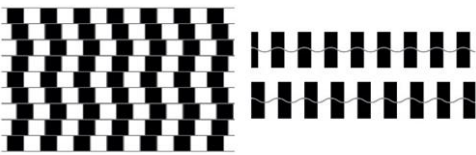


Figure 7.5 The café wall illusion (left), where all lines are actually straight and either vertical or horizontal. A related illusion by Akiyoshi Kitaoka (right), where the two gray waves are identical in height. Left image by Wikipedia user "Fibonacci" - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=1788689>. Right image by Akiyoshi Kitaoka, used with permission.



Figure 7.6 The Ebbinghaus illusion: the darker gray circles are the same size.

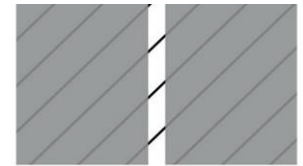


Figure 7.7 An illusion by Akiyoshi Kitaoka: the diagonal lines are all straight and parallel, used with permission.

Εικόνα 2-6: [1] Ψευδαισθήσεις

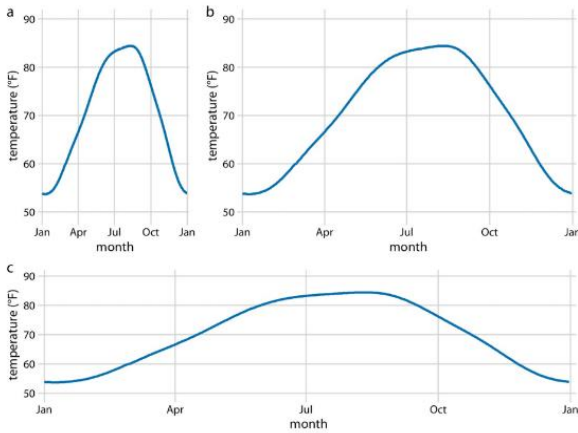


Figure 3-2. Daily temperature normals for Houston, TX. Temperature is mapped to the y axis and day of the year to the x axis. Parts (a), (b), and (c) show the same figure in different aspect ratios. All three parts are valid visualizations of the temperature data. Data source: NOAA.

Εικόνα 2-8: [2] Έμφαση στους άξονες

Μερικές φορές, ανάλογα με την ιστορία που θέλουμε να πούμε, μπορεί να στοχεύουμε οι σημαντικές διαφορές να είναι αρκετά αισθητές. Ένας τρόπος για να γίνει αυτό είναι να αλλάξουμε την αναλογία που τους δείχνουμε.

Δηλαδή να τεντώσουμε ή να συμπιέσουμε τους άξονες (το οποίο μπορεί να γίνει εάν και εφόσον οι άξονες είναι με διαφορετικές μονάδες μέτρησης). Όταν συμπιέζεται ο άξονας δημιουργώντας έτσι στενότερο και ψηλότερο γράφημα, δίνουμε έμφαση στις αλλαγές που γίνονται στον άξονα των y, ενώ όταν τεντώνεται ο άξονας (έχοντας φαρδύτερο και κοντύτερο γράφημα) δίνουμε έμφαση στον άξονα των x (Εικόνα 2-8).

Από έρευνα που έχει γίνει για την οπτική αντίληψη του ανθρώπου, έχει αποδειχθεί ότι ορισμένες παράμετροι είναι πιο δύσκολο να ερμηνευτούν πάλι πίσω σε νούμερα από ότι κάποιες άλλες. Παράμετροι που παρουσιάζουν τέτοιου είδους προβλήματα φαίνονται στην Εικόνα 2-7 με αστερίσκο. Οι περιοχές (Areas) για παράδειγμα, θεωρούνται ότι είναι αρκετά παρόμοιες μεταξύ τους, οι όγκοι (Volumes) ότι είναι δύσκολο να κρίνουμε το ακριβές μέγεθος του κάθε αντικειμένου (αλλά θεωρούνται καλοί για μήκη ή για να αντιπροσωπεύουν τακτικά (σε σειρά) (ordinal) δεδομένα, κ.λπ.

TABLE 2.1 Visual parameters

Parameter	Ideal data type	
Position	Quantitative	
Length	Quantitative	
Angle	Quantitative*	
Area	Quantitative*	
Volume	Quantitative*	
Color hue	Nominal or ordinal*	
Color saturation	Ordinal	
Marker shape	Nominal*	
Features	Quantitative*	
Line width	Ordinal*	

Εικόνα 2-7: [1] Παράμετροι δύσκολοι να ερμηνευτούν πίσω σε νούμερα (με αστερίσκο)

Επίσης, είναι πολύ πιο εύκολο να καταλάβουμε τα στοιχεία όταν αυτά ξεκινούν από το ίδιο σημείο, ενώ γίνεται αρκετά δύσκολο όταν αυτά ξεκινούν από διαφορετικά σημεία, ή έχουν διαφορετικές γωνίες, ή πλάτοι (Εικόνα 2-9). Αυτός είναι και ο λόγος για τον οποίο τα διαγράμματα πίτας (pie charts), τα ακτινωτά ιστογράμματα (radial histograms), τα διαγράμματα ηλιοφάνειας (sunburst diagrams) είναι δυσκολότερο να μεταφραστούν πίσω σε νούμερα, παρόλο που είναι εντυπωσιακά. Πολλοί αναγνώστες αν όχι οι περισσότεροι, βγάζουν συμπεράσματα με βάση την πρώτη τους εντύπωση, οπότε είναι πολύ σημαντικό να τους μεταδώσουμε το σωστό μήνυμα. Φυσικά, όπως συμβαίνει με όλα τα πράγματα, έτσι και εδώ υπάρχουν

εξαιρέσεις (όπως με τον όγκο (volume) για παράδειγμα, που όταν απεικονίζεται στην Εικόνα 2-10 δεν πέφτει στην παγίδα που αναφέραμε).

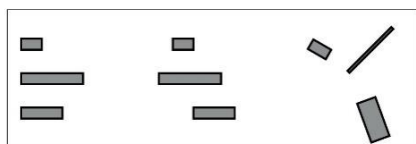


Figure 2.7 To let the reader translate a parameter back into numbers, keep everything else constant. The lengths of the three bars on the left can be compared, but less so for the three in the center, where the baseline position has been shifted, or on the right where the width and angle changes too.

Εικόνα 2-9: [1] Στοιχεία με διαφορετικό σημείο εκκίνησης

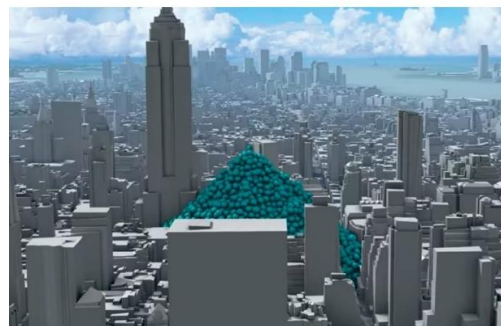


Figure 16.5 Volume of carbon dioxide emitted in New York City in one hour, visualized as a heap of one-ton balls smothering Midtown. A still from a video at youtu.be/DtqSIplGXOA, by Carbon Visuals (www.realworldvisuals.com), reproduced with permission.

Εικόνα 2-10: [1] Ένα παράδειγμα εξαιρέσης προβληματικής παραμέτρου

Όταν θέλουμε να συγκρίνουμε τιμές, θα πρέπει να ξεχωρίζουν και να βρίσκονται κοντά η μία στην άλλη. Κάποιες από αυτές τις τιμές θα τις δούμε και αργότερα.

Όταν το μήνυμά μας περιλαμβάνει ειδικές μεταβλητές ή τιμές, η αρχή της διαστρωμάτωσης (layering principle) είναι καλύτερη επιλογή: να διαχωριστούν οπτικά κάνοντας ένα γράφημα με τίτλο και όλες οι υπόλοιπες λεπτομέρειες αλλού. Η δημιουργία καινοτόμων οπτικοποιήσεων πάντα αξίζει τον κόπο, αρκεί να ληφθούν υπόψη τα πολύτιμα δευτερόλεπτα που μπορεί να χρειαστεί ο αναγνώστης για να μάθει πώς να διαβάσει την απεικόνιση, δεδομένου ότι τις περισσότερες φορές, ο πρωταρχικός μας στόχος είναι «η επικοινωνία».

2.1 Προσοχή στις Τρισδιάστατες Απεικονίσεις...

Αν και τα τρισδιάστατα γραφικά είναι πολύ δημοφιλή, τις περισσότερες φορές έχουν διακοσμητικό χαρακτήρα χωρίς να μεταφέρουν τα αληθινά δεδομένα, με αποτέλεσμα να χρησιμοποιούνται εσφαλμένα και γι' αυτό πρέπει να αποφεύγονται.

Η προβολή τρισδιάστατων αντικειμένων σε ένα δισδιάστατο γράφημα, αλλάζει την αντίληψή μας για τα δεδομένα, παρόλο που το οπτικό μας σύστημα προσπαθεί να τα διορθώσει. Ούτε η χρήση κλίμακας σε πραγματική θέση (x, y, z) δεν μπορεί να βοηθήσει σε αυτό. Τέτοιου είδους παραδείγματα είναι:

- Τρισδιάστατα διαγράμματα πίτας (3d pie charts): καθώς, αλλάζοντας τη γωνία της πίτας, μας δίνεται η λανθασμένη εντύπωση ότι αλλάζει και το μέγεθος της κάθε φέτας (Εικόνα 2.1-11 αριστερά)
- Τρισδιάστατες ράβδοι (3d bars): καθώς, αλλάζοντας τη γωνία των ράβδων μας δίνεται λανθασμένη εντύπωση για το μέγεθος τους. Στην Εικόνα 2.1-11 (δεξιά) για παράδειγμα, ο αριθμός των επιβατών στην πρώτη θέση (1st Class) ήταν 322, αλλά στο γράφημα φαίνεται ότι το ύψος της πρώτης ράβδου είναι μικρότερο από 300.
- Τρισδιάστατα διαγράμματα διασποράς (3d scatterplots): καθώς, ακόμα και αν τα δούμε από διαφορετικές οπτικές γωνίες, και πάλι χρειάζεται πολλή νοητική προσπάθεια για να μπορέσουμε να τα ερμηνεύσουμε, μιας και είναι δύσκολο να καταλάβουμε ποια στοιχεία είναι πιο μπροστά από τα άλλα. Και δεν μπορούμε να καθορίσουμε με μοναδικό τρόπο πού τοποθετούνται/υπάρχουν αυτά τα σημεία στον τρισδιάστατο χώρο, με αποτέλεσμα να κάνοντάς τα αναξιόπιστα και εξαρτόμενα από την προβολή τους από διαφορετικές οπτικές γωνίες (Εικόνα 2.1-12)

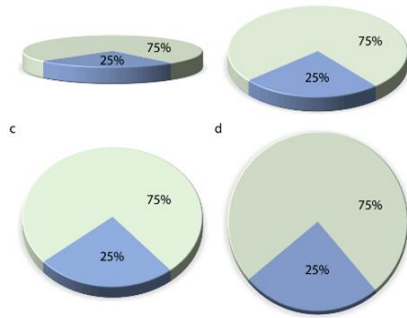


Figure 26-1. The same 3D pie chart shown from four different angles. Rotating a pie into the third dimension makes pie slices in the front appear larger than they really are and pie slices in the back appear smaller. Here, in parts (a), (b), and (c), the blue slice corresponding to 25% of the data visually occupies more than 25% of the area representing the pie. Only part (d) is an accurate representation of the data.

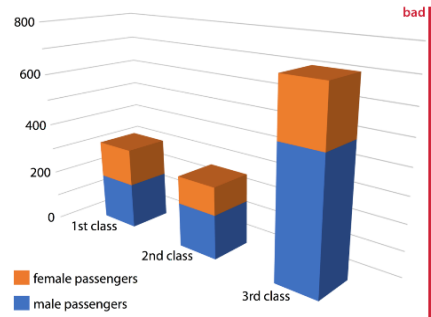


Figure 26-2. Numbers of female and male passengers on the Titanic traveling in 1st, 2nd, and 3rd class, shown as a 3D stacked bar plot. The total numbers of passengers in 1st, 2nd, and 3rd class are 322, 279, and 711, respectively (see Figure 6-10). Yet in this plot, the 1st class bar appears to represent fewer than 300 passengers, the 3rd class bar appears to represent fewer than 700 passengers, and the 2nd class bar seems to be closer to 210 passengers than the actual 279 passengers. Furthermore, the 3rd class bar visually dominates the figure and makes the number of passengers in 3rd class appear larger than it actually is.

Εικόνα 2.1-11: [2] Τρισδιάστατες απεικονίσεις με λανθασμένες εντυπώσεις

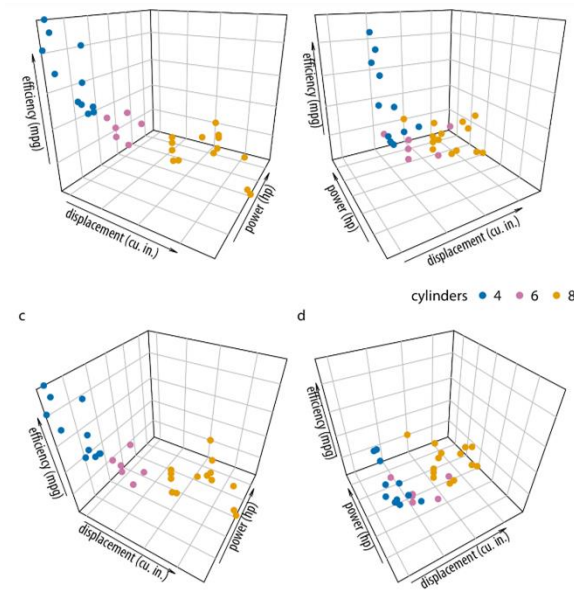


Figure 26-3. Fuel efficiency versus displacement and power for 32 cars (1973–74 models). Each dot represents one car, and the dot color represents the number of cylinders of the car. The four panels (a)–(d) show exactly the same data but use different perspectives. Data source: Motor Trend, 1974.

Εικόνα 2.1-12: [2] Τρισδιάστατες απεικονίσεις αρκετής νοητικής προσπάθειας

Σε τέτοιες περιπτώσεις, η χρήση τρίτης διάστασης γενικά δεν είναι απαραίτητη. Θα μπορούσε αυτή να αντικατασταθεί από κλίμακα χρώματος, μεγέθους ή σχήματος. Για παράδειγμα, να δείξουμε τα δύο στοιχεία και να αντιστοιχίσουμε την τρίτη μεταβλητή με το μέγεθος, όπως στην Εικόνα 2.1-13 (αριστερά), ή, να το ξανασχεδιάσουμε (δεύτερη φορά) αλλά τώρα να χρησιμοποιήσουμε την άλλη μεταβλητή που μας ενδιαφέρει χωριστά (Εικόνα 2.1-13, δεξιά).

Θα μπορούσαμε επίσης να χρησιμοποιήσουμε 3d ράβδους (3d bars) αντί για διαγράμματα διασποράς (scatterplots) για να δείξουμε αυτήν την τρίτη διάσταση, αλλά είναι **δύσκολο** να

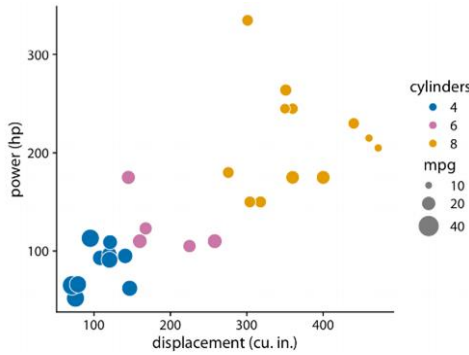


Figure 26-6. Power versus displacement for 32 cars, with fuel efficiency represented by dot size. Data source: Motor Trend, 1974.

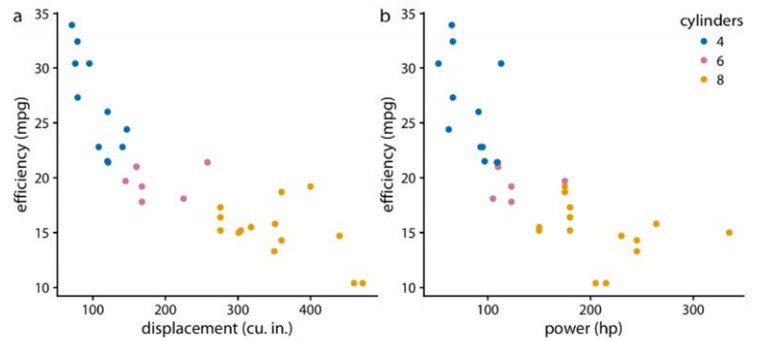


Figure 26-5. Fuel efficiency versus displacement (a) and power (b) for 32 cars. Data source: Motor Trend, 1974.

Εικόνα 2.1-13: [2] Παράδειγμα λύσεων για προβληματικές τρισδιάστατες απεικονίσεις

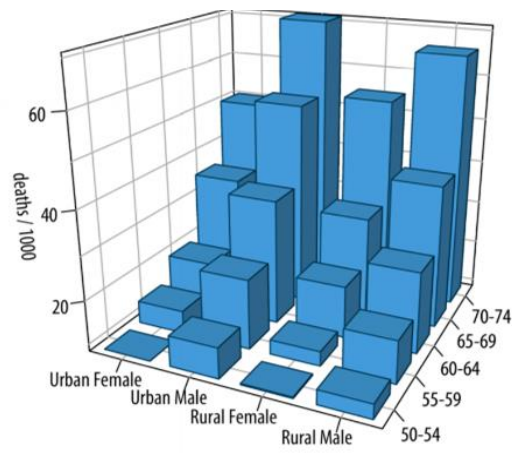


Figure 26-7. Mortality rates in Virginia in 1940, visualized as a 3D bar plot. Mortality rates are shown for four groups of people (urban and rural females and males) and five age categories (50–54, 55–59, 60–64, 65–69, 70–74), and they are reported in units of deaths per 1,000 persons. This figure is labeled as “bad” because the 3D perspective makes the plot difficult to read. Data source: [Molyneaux, Gilliam, and Florant 1947].

Εικόνα 2.1-14: [2] Παράδειγμα προβληματικής τρισδιάστατης απεικόνισης

κρίνουμε πόσο ψηλές είναι οι μεμονωμένες ράβδοι και να μπορέσουμε να κάνουμε άμεσες συγκρίσεις (για παράδειγμα στην Εικόνα 2.1-14, δε φαίνεται ξεκάθαρα ποιο ήταν πιο ψηλό, το ποσοστό θνησιμότητας των αστικών γυναικών στην ηλικιακή ομάδα 65-69 ετών ή των ανδρών της πόλης στην ηλικιακή ομάδα 60-64).

Λύση: είναι καλύτερα να χρησιμοποιήσουμε μικρά πολλαπλάσια, μιας και είναι πιο καθαρά και ευκολότερα στο να τα ερμηνεύσει/κατανοήσει κανείς.

Ωστόσο, αυτά τα προβλήματα γίνονται λιγότερο σημαντικά εάν η απεικόνισή μας:

- είναι διαδραστική και μπορεί ο χρήστης να την περιστρέψει, ή εάν εμφανίζεται ως VR ή οπτικοποίηση επαυξημένης πραγματικότητας (augmented reality visualization), οπότε μπορεί να προβληθεί από πολλές γωνίες
- μπορεί να εμφανιστεί ως μια αργά περιστρεφόμενη κινούμενη εικόνα, έτσι ώστε ο χρήστης να αντιληφθεί πού στον τρισδιάστατο χώρο υπάρχουν τα διαφορετικά στοιχεία.

Μπορούμε επίσης να χρησιμοποιήσουμε τρισδιάστατες απεικονίσεις, όταν θέλουμε να δείξουμε αληθινά τρισδιάστατα αντικείμενα ή/και δεδομένα πάνω τους, για παράδειγμα: η Εικόνα 2.1-15 (αριστερά): δείχνει το τοπογραφικό ανάγλυφο ενός ορεινού νησιού, ενώ στην Εικόνα 2.1-15 (δεξιά) έχει οπτικοποιηθεί η εξελικτική αλληλουχία διατήρησης μιας πρωτεΐνης χαρτογραφημένης στη δομή της), αν και παρόλ' αυτά, θα ήταν ευκολότερο να ερμηνευτούν εάν τα παρουσιάζαμε ως περιστρεφόμενα κινούμενα σχέδια (animations) στο internet ή σε παρουσιάσεις.



Figure 26-9. Relief of the Island of Corsica in the Mediterranean Sea. Data source: Copernicus Land Monitoring Service.

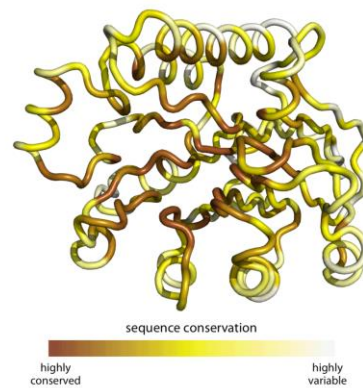


Figure 26-10. Patterns of evolutionary variation in a protein. The colored tube represents the backbone of the protein Exonuclease III from the bacterium *Escherichia coli*. The coloring indicates the evolutionary conservation of the individual sites in this protein, with dark coloring indicating conserved amino acids and light coloring indicating variable amino acids. Data source: [Marcos and Echave 2015].

Εικόνα 2.1-15: [2] Παραδείγματα χρήσης τρισδιάστατων απεικονίσεων, καλή πρακτική

2.2 Το «Μεγαλύτερο Πακέτο»/Ένα Καλό Περιτύλιγμα

(η σχεδίαση παίζει ρόλο)

Μια πολύ καλή οπτικοποίηση χρειάζεται ένα πολύ καλό περιτύλιγμα. Δεν είναι κάτι που πρέπει να αφήνουμε για το τέλος. Χρειάζεται σκέψη και δοκιμές με χρήστες (user-testing)

Όταν έχουμε ένα μήνυμα που θέλουμε να μεταφέρουμε, θα πρέπει να παρουσιαστεί ως μια ουσιαστική και συναρπαστική ιστορία. Οι ιστορίες παίζουν σημαντικό ρόλο στο συλλογισμό και

τη μνήμη. Εάν δεν τους παρέχουμε εμείς κάποια, σημαίνει ότι θα δημιουργήσουν μία οι ίδιοι οι αναγνώστες, που στο καλύτερο πιθανό σενάριο, μπορεί να είναι κοντά στην άποψή μας. Επομένως, προσπαθούμε να χρησιμοποιήσουμε παρατηρήσεις, γεγονότα (facts) ή γεγονότα (events) (αληθινά ή επινοημένα) και να τα παρουσιάσουμε με τέτοια συγκεκριμένη σειρά, ώστε να δημιουργηθεί συναισθηματική αντίδραση σε αυτά. Ο στόχος μας; Να πούμε μια ιστορία χρησιμοποιώντας γεγονότα και λογική συλλογιστική (logical reasoning), ώστε να τους κάνουμε να ενδιαφερθούν και να ενθουσιαστούν.

Υπάρχουν κάποια τυπικά μοτίβα για την αφήγηση ιστοριών, τα οποία έχουν απήχηση στον τρόπο σκέψης μας: Το μοτίβο «Opening–Challenge–Action–Resolution», το μοτίβο «Lead–Development–Resolution», που επίσης ονομάζεται Lead-Development», το μοντέλο «Action-Background – Development –Climax– Ending». Έχοντας μόνο την οπτικοποίηση, σπάνια θα καταφέρουμε να πούμε μια ιστορία, αφού δεν μπορεί να τα απεικονίσει όλα ταυτόχρονα. Αυτός είναι και ο λόγος για τον οποίο χρειαζόμαστε συνήθως περισσότερες οπτικοποιήσεις. Είναι πιθανό, το μήνυμά μας να μπορεί να εμφανιστεί με ένα μόνο σχήμα, όμως ανάλογα με το εάν μας επιτρέπει το μέσο (δηλαδή, το σύντομο χρονικό διάστημα προσοχής μέσω των μέσων κοινωνικής δικτύωσης), θα μπορούσαμε να παρουσιάσουμε το ίδιο μήνυμα με περισσότερες απεικονίσεις, έτσι ώστε να έχουμε ισχυρότερη επίδραση, όπως σε μια παρουσίαση. Θα μπορούσαμε για παράδειγμα, να εμφανίσουμε κάποιο υπόβαθρο, μετά το σχήμα που δημιουργεί την πρόκληση, και τέλος, το σχήμα που παρέχει την ανάλυση.

Υπάρχει η εσφαλμένη αντίληψη από ανθρώπους της επιστήμης δεδομένων και της στατιστικής, ότι η σχεδιαστική σκέψη είναι ένα «πράγμα» και όχι μια διαδικασία, η οποία χαρακτηρίζεται από μινιμαλισμό και γραμματοσειρές Helvetica. Αν και αυτό είναι ένα καλό σημείο εκκίνησης, δεν υπάρχει η έννοια μίας καλύτερης οπτικοποίησης ως η μοναδική σωστή και χρειάζεται πειραματισμός, δοκιμές χρηστών και βελτιστοποιήσεις μέχρι να φτάσουμε στο καλύτερο δυνατό αποτέλεσμα.

Ας δούμε μερικές ιδέες σχεδιασμού που μπορούν να ενισχύσουν τον αντίκτυπο της δουλειάς μας:

- Η γραφή σε επιστημονική μορφή «στόχοι-μέθοδοι-αποτελέσματα-συμπεράσματα» δεν βοηθάει, ή τουλάχιστον θα μπορούσαμε να βάλουμε πρώτα τα συμπεράσματα. Μια άλλη πιθανή μορφή που μπορούμε να χρησιμοποιήσουμε (η οποία προέρχεται από τον Simon

Sinek) είναι, αντί να δείξουμε τι βρήκαμε, πώς το κάναμε και γιατί αυτό έχει σημασία, να το αντιστρέψουμε στο «γιατί, πώς, τι».

- Το μήνυμα θα πρέπει να είναι όσο το δυνατόν συντομότερο, με προσεκτικά επιλεγμένους όρους και με σειρά εισαγωγής των γεγονότων. Στη συνέχεια, θα πρέπει να χρησιμοποιήσουμε αυτήν τη δομή για την οπτικοποίησή μας. Θα πρέπει όμως να έχουμε κατά νου, ότι διαφορετικοί αναγνώστες θέλουν διαφορετικά πράγματα από μια απεικόνιση, και το πακέτο με το οποίο φτάνει σ' αυτούς η ιστορία μας, θα πρέπει να είναι το άνοιγμα για να το εξερευνήσουν περαιτέρω. Μπορούμε να το κάνουμε αυτό βάζοντας τις πληροφορίες μας σε στρώματα με τέτοιο τρόπο, ώστε να μπορέσουν να φτάσουν στο επίπεδο που τους ενδιαφέρει.

Προτεινόμενες ερωτήσεις που μπορεί να βοηθήσουν είναι:

- Ποια μπορεί να είναι αυτά τα στρώματα;
- Ποια σειρά θα θέλουν να συναντήσουν οι αναγνώστες μας;
- Ποιες λέξεις ή σύμβολα θα κατανοούσαν ως σημάνσεις σε διαφορετικά επίπεδα;

Ο σχεδιαστής οπτικοποιήσεων Andy Kirk προτείνει ως βοηθητικές ερωτήσεις τις:

- Ποιος είναι ο λόγος ύπαρξης της οπτικοποίησης;
- Για ποιους τη δημιουργούμε και πόσο καλά ορισμένες είναι οι απαιτήσεις τους;
- Ποια λειτουργία επιδιώκει να εκπληρώσει το πακέτο;
- Ποιος είναι ο πιθανός τόνος της σχεδίασης;

Επίσης:

- Η απεικόνιση προορίζεται να είναι σοβαρή, εντυπωσιακή, συγκλονιστική, όμορφη, αστεία;
- Πώς μπορούμε να κρατήσουμε την ίδια κωδικοποίηση αλλά να τροποποιήσουμε τη μορφή για να επιτύχουμε αυτόν τον τόνο;
- Τι μπορεί να εννοούμε με το «ομορφιά στην οπτικοποίηση δεδομένων»;

Μια όμορφη οπτικοποίηση θα πρέπει να είναι καινοτόμα (novel), ενημερωτική, αποτελεσματική και αισθητική (Noah Iliinsky).

Αν θέλουμε να δημιουργήσουμε μια ομάδα για την κατασκευή τέτοιου πακέτου που να περιβάλλει την οπτικοποίησή μας, θα πρέπει να έχουμε σε αυτήν: έναν σχεδιαστή, έναν επιστήμονα δεδομένων (data scientist), έναν ειδικό για το συγκεκριμένο θέμα και έναν προγραμματιστή ιστού (web developer). Όλοι αυτοί θα πρέπει να συνεργάζονται παρά τους διαφορετικούς κόσμους από τους οποίους μπορεί να προέρχονται, αποτελεσματικά, μέχρι τη στιγμή της δημοσίευσής της. Θα πρέπει να καταλαβαίνει ο ένας τα κίνητρα του άλλου και να διασφαλίσουν ότι όλα τα μέλη ως ένα σημείο του έργου μπορεί να είναι υπερήφανα μετέπειτα.

Ο στόχος ενός καλού πακέτου σχεδιασμού συμπεριλαμβανομένης και της οπτικοποίησης δεδομένων, είναι να δώσει στους αναγνώστες πρόσβαση σε πληροφορίες με τρόπους, τόπους και περιόδους που δεν θα μπορούσαν διαφορετικά. Οι σχεδιαστές Stephanie Posavec και Georgia Lupi, τονίζουν την αξία του ανθρωπισμού των δεδομένων (data humanism): δηλαδή τη χρήση των δεδομένων για να εμπλουτίσουμε τη ζωή μας και όχι για να αυξήσουμε την αποτελεσματικότητα με κόστος την ευτυχία. Κάπου σε αυτό το φάσμα θα βρούμε τους τελικούς στόχους των αναγνωστών μας. Όταν προσδιορίσουμε το επίπεδο στατιστικής παιδείας στο αναγνωστικό κοινό μας, μπορούμε να κάνουμε τις κατάλληλες απεικονίσεις και κείμενο, έχοντας αυτούς ως στόχο.

Ο έλεγχος από χρήστες (user testing) είναι επίσης απαραίτητος, ειδικά όταν πολλές πληροφορίες συγκεντρώνονται σε ένα πακέτο. Με αυτόν τον τρόπο μπορούμε να συλλέξουμε σκέψεις και ιδέες σχετικά με την ευκολία χρήσης, την προσβασιμότητα και την απόλαυση και να δούμε εάν τα χρώματα, το συνοδευτικό κείμενο και οι επισημάνσεις (annotations) συνεργάζονται όμορφα. Οι γνωστοί τύποι/μορφές (formats) έχουν δύναμη, αλλά πολλοί άνθρωποι έλκονται από την καινοτομία, κυρίως από περιέργεια για το κάτι καινούργιο. Αυτό είναι μια χαρά, αρκεί να είμαστε αρκετά προσεκτικοί και να γίνει πρώτα δοκιμή από χρήστες (user testing).

Αποτελεσματική οπτικοποίηση, είναι αυτή που λειτουργεί καλά για τους αναγνώστες για τους οποίους προορίζεται.

Για να το πετύχουμε αυτό, θα πρέπει να σκεφτούμε ποιος χρειάζεται να συμμετέχει για να μας δώσει αυτή την ανατροφοδότηση (feedback), καθώς και για να βεβαιωθούμε ότι ο επιθυμητός αντίκτυπος επιτυγχάνεται. Το να έχουμε ένα μέτρο κατανόησης και ανάκλησης των πληροφοριών μεταξύ των αναγνωστών μας θα ήταν πολύ χρήσιμο, εάν φυσικά το επιτρέπει ο χρόνος και οι πόροι μας (όπως με συνέντευξη και ενημέρωση των αναγνωστών).

Οι πίνακες ελέγχου (dashboards) είναι συλλογές οπτικοποιήσεων που είναι συνδεδεμένες μεταξύ τους και παρουσιάζουν συνοψισμένη την δραστηριότητα η οποία ενημερώνεται συχνά/συνεχώς, όλη σε ένα μέρος. Είναι πολύ δημοφιλείς ως κομμάτι της επιχειρηματικής νοημοσύνης (business intelligence). Όμως, η κατοχή πολλής διαφορετικής πληροφορίας σε ένα μέρος, μπορεί να οδηγήσει σε υπερφόρτωση πληροφοριών. Μπορούμε να απλοποιήσουμε την πληροφορία χρησιμοποιώντας διαδραστικότητα (interactivity) και μικρά πολλαπλάσια (small multiples). Η τοποθέτηση των λεπτομερειών σε επίπεδα για διαφορετικούς αναγνώστες είναι επίσης πολύτιμη.

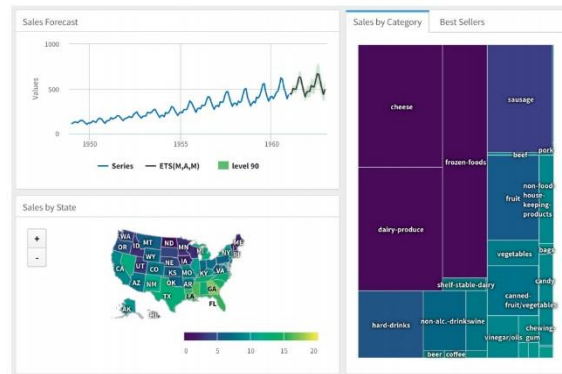


Figure 16.1 Example of a dashboard tracking retail sales data, created by Joshua Kunst and published by RStudio. Reproduced with permission.

Εικόνα 2.2-16: [1] Πίνακας ελέγχου

Στην περίπτωση μοντέλων μεγάλης υπολογιστικής ισχύος (computer-intensive models) όπως νευρωνικά δίκτυα που χρειάζονται πολύ χρόνο ώστε να ενημερώσουν τα νέα δεδομένα, θα μπορούσαμε να δημιουργήσουμε και να απεικονίσουμε γρήγορα προβλέψεις για αυτά από προηγούμενα μοντέλα, βάζοντας τις κατάλληλες προειδοποιήσεις στον πίνακα ελέγχου.

Προσοχή: Τα μάτια του αναγνώστη θα ψάξουν να βρουν από πού να ξεκινήσουν και αν φανεί μπερδεμένο ή κουραστικό, θα οδηγήσει σε μια «κακή» πρώτη εντύπωση. Δεν πρέπει να υποθέτουμε ότι μπορούν να επεξεργαστούν γρήγορα πολύπλοκες απεικονίσεις και ότι κατανοούν τις βασικές τάσεις και σχέσεις που τους παρουσιάζονται. Πρέπει πάντα να έχουμε στο νου μας, ότι είναι πολύ απασχολημένοι και δεν μπορούν να ξοδεύσουν πολύ χρόνο προσπαθώντας να αποκρυπτογραφήσουν μια αινιγματική απεικόνιση.

Για να αποφευχθεί αυτό, απαιτείται καθοδήγηση, χρησιμοποιώντας πράγματα όπως:

- μεγάλους τίτλους κειμένου για κάθε κομμάτι/μέρος,
- βέλη που να οδηγούν από τη μία απεικόνιση στην άλλη,
- ξεχωριστός οδηγός "πώς να διαβάσετε αυτόν τον πίνακα ελέγχου"

- σχολιασμοί ως μέσο καθοδήγησης. Κάποιοι πιστεύουν ότι μια καλή απεικόνιση δεν θα έπρεπε να απαιτεί καμία επεξήγηση, ούτε καν ένα υπόμνημα ή τίτλο. Αλλά αυτό το θεωρητικό ιδεώδες, δε θα έπρεπε να παρεμποδίζει τους αναγνώστες να έχουν πρόσβαση και να απολαμβάνουν μια οπτική διεπαφή στα δεδομένα
- κάθε οπτικοποίηση να συνδέεται οπτικά με το υπόλοιπο πακέτο
- να έχει εύκολα αναγνωρίσιμα στοιχεία για να τραβήξουν την προσοχή
- προειδοποιητικά στοιχεία, εάν υπάρχουν στο κείμενο, να ταιριάζουν με αυτά της απεικόνισης όπως, ένα επισημασμένο (highlighted) σημείο σε ένα γράφημα να χρησιμοποιεί ένα χρώμα που να ταιριάζει με αυτό ενός κομματιού κειμένου παράλληλα).
- Για διαδραστικούς πίνακες ελέγχου (interactive dashboards), ο αναγνώστης θα μπορούσε να καθοδηγείται μέσω μιας απεικόνισης κάθε φορά, ενώ οι υπόλοιπες ξεθωριάσουν στο παρασκήνιο.
- Το χρώμα, οι γραμματοσειρές και άλλα στυλ όπως μοτίβα γραμμών και σχήματα ένδειξης (marker shapes) θα πρέπει να είναι συνεπή καθ' όλη τη διάρκεια, ώστε να τους βοηθήσουν να εστιάσουν στη σωστή πληροφορία. Οι οπτικοποιήσεις που δημιουργούνται για οργανισμό που έχει λογότυπο, ή χρησιμοποιεί συγκεκριμένο συνδυασμό χρωμάτων ή υπάρχουν κατευθυντήριες οδηγίες για την επωνυμία, θα πρέπει να λαμβάνονται υπόψη και να ταιριάζουν με αυτές. Όταν εξετάζουμε το χρώμα που θα χρησιμοποιήσουμε, θα πρέπει επίσης να λάβουμε υπόψη τους αναγνώστες με αχρωματοψία ή διάφορα άλλα προβλήματα όρασης. Υπάρχουν διάφορα εργαλεία στο διαδίκτυο για την προσομοίωση του τρόπου εμφάνισης της παλέτας χρωμάτων. Εργαλεία που προτείνουν άλλα χρώματα που συμπληρώνουν το χρώμα αρεσκείας μας, και που μας δίνουν πληροφορίες για το πώς το αντιλαμβάνονται άνθρωποι με προβλήματα όρασης. Τέτοια εργαλεία είναι: <https://www.colorhexa.com/> , <https://colorbrewer2.org> , <https://coolors.co> , <https://color.adobe.com/create/color-wheel>, <https://material.io/resources/color/>, <https://www.designspiration.com/>, <http://www.0to255.com/>, <https://medialab.sciencespo.fr/en/tools/i-want-hue/>
- Τύποι (formats) όπως Μετρητές (Gauges) που δείχνουν μεμονωμένες τιμές και μοιάζουν με ταχύμετρα, ή Καντράν (Dials) σε κάποιο επιστημονικό όργανο, αν και είναι τυπικά στοιχεία του ταμπλό και μπορεί να είναι διασκεδαστικά, είναι αρκετά φτωχά όσον αφορά

τη μετατροπή τους πίσω σε αριθμούς από τον αναγνώστη και τη σύγκριση μεταξύ τους και πρέπει να χρησιμοποιούνται με προσοχή.

Ας δούμε ένα τέτοιο παράδειγμα που χρησιμοποιεί αυτές τις ιδέες για να δημιουργήσει μια αφίσα (ή φυλλάδιο) αναφορικά με δεδομένα καθυστερήσεων τρένου, έχοντας ως αφητηρία ορισμένες απεικονίσεις διαφορετικών πτυχών των δεδομένων και διαφορετικών επιπέδων λεπτομέρειας:

Στην Εικόνα 2.2-17 (πάνω αριστερά) φαίνεται το αρχικό, λεπτομερές διάγραμμα γραμμών, μια γρήγορη επισκόπησή του φαίνεται στην Εικόνα 2.2-17 (πάνω δεξιά) χρησιμοποιώντας εξομαλυμένες γραμμές για το γράφημα (κέντρο) (smoothed splines), και η Εικόνα 2.2-17 (κάτω) επικεντρώνεται στο «ποια εποχή του χρόνου ήταν η χειρότερη» για κάθε χρόνο.

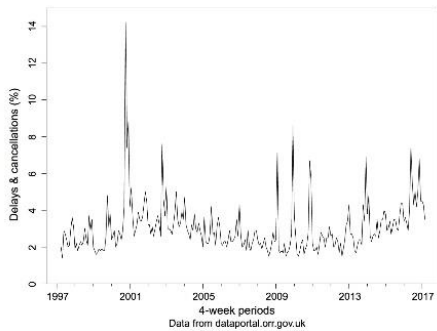


Figure 2.2 Train delays: a line chart

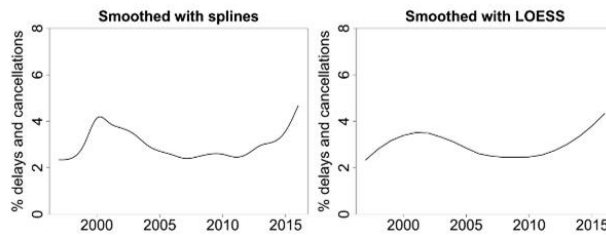


Figure 5.7 Smoothed data to summarize change in train delays over 20 years, using splines (left) and LOESS (right).

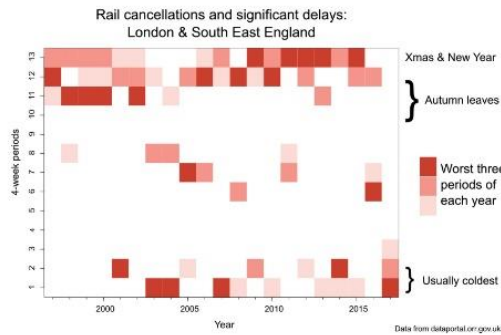


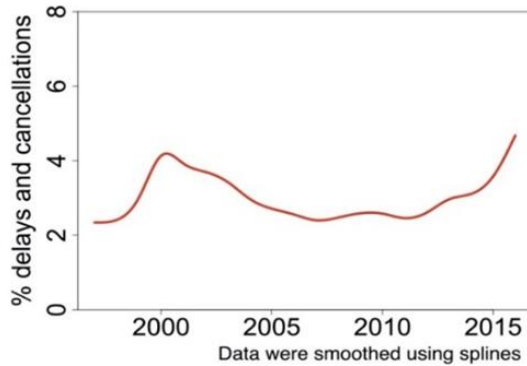
Figure 2.6 Train delays: year-period grid, with the worst 4-week periods in each year colored

Εικόνα 2.2-17: [1] Αρχικές πτυχές δεδομένων με διαφορετικό επίπεδο λεπτομέρειας - παράδειγμα

Συνδυάζοντας τα παραπάνω ο Robert Grant, δημιούργησε την Εικόνα 2.2-18. Ο αναγνώστης καθοδηγείται από το ανώτερο επίπεδο λεπτομέρειας προς τα κάτω, χωρίς να μπερδευτούν για το σημείο εκκίνησης και μπορούν να εγκαταλείψουν σε όποιο σημείο θεωρούν ότι τους ταιριάζει/καλύπτει.

Leaves on the line

Do Autumn leaves delay trains in SE England?

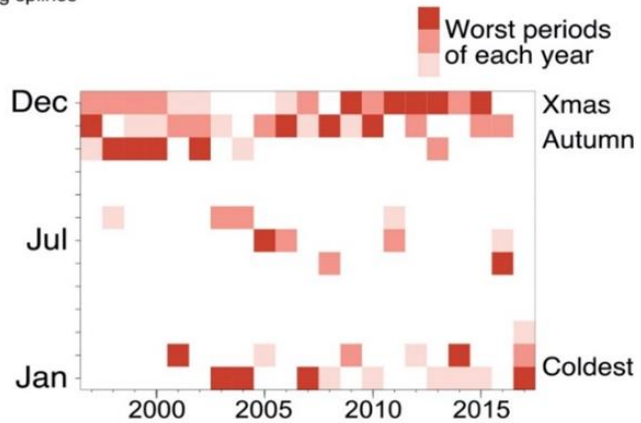


Delays and cancellations are recorded in 4-week periods.

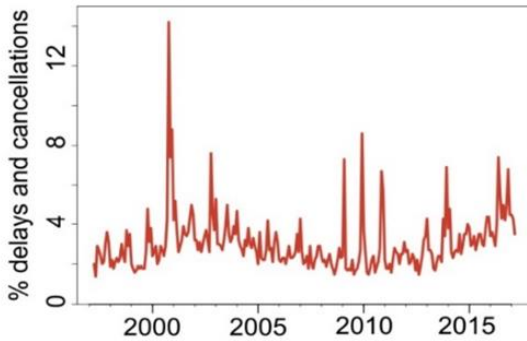
The long-term trend was that performance improved, then deteriorated again.

Government subsidy peaked in 2006, and was reduced each year since then.

The worst delays have shifted from Autumn to Winter.
Leaves are no longer a problem.



Short-term problems (spikes below) in context



- 1994-97: Privatisation
- 1999: Paddington crash
- 2000: Hatfield crash
- 2000: Potters Bar crash; speed limits
- 2002: New infrastructure company
- 2009-11: Unusually snowy winters
- 2017: Southern Rail strikes

Data from dataportal.orr.gov.uk

Figure 16.2 The train delay data visualizations could be combined in a poster like this

Εικόνα 2.2-18: [1] Τελικό αποτέλεσμα παραδείγματος

Αναλυτικά: Αυτό, γίνεται εδώ τοποθετώντας εικόνες μαζί με κάποιο συνοδευτικό κείμενο σε τρεις γραμμές. Χρησιμοποιήθηκε μεγάλη επικεφαλίδα, ώστε το σημείο εκκίνησης είναι πολύ σαφές. Διατηρείται ζωντανό αλλά χωρίς να είναι μονότονο, αλλάζοντας τα γραφήματα από αριστερά σε δεξιά και πίσω. Θα ήταν δελεαστικό να φτιαχτεί η εικόνα με το φύλλο περισσότερο, αλλά η διατήρησή του ως απλό γραφικό, ελαχιστοποιεί τη φασαρία και το κρατά μακριά από τις περιοχές πλοκής των διαγραμμάτων. Το να είχε χρησιμοποιηθεί ως φόντο σίγουρα δεν θα ήταν καλή λύση, ανεξάρτητα από το πόσο ξεθωριασμένο μπορεί να γινόταν, και η προσθήκη περισσότερων διακοσμητικών στοιχείων απλά θα πρόσθετε θόρυβο. Το να έμοιαζε με πίνακα ανακοινώσεων σιδηροδρομικού σταθμού για παράδειγμα, θα ευνοούσε σαφώς το στυλ από την ουσία. Δόθηκε ιδιαίτερη προσοχή στο σχολιασμό ώστε να μην προστεθεί υπερβολικά πολύς, αλλά ούτε και λίγος (με γεγονότα που είναι σχετικά και εξηγούν μερικές από τις κορυφές του γραφήματος, αλλά αποφασίστηκε ότι αυτά θα υπερφόρτωναν την οπτικοποίηση εάν σχολιάζονταν και αποφεύχθηκαν).

Οι δοκιμές από χρήστες (user testing) είναι σημαντικές, καθώς εξυπηρετούμε τον αναγνώστη και όχι τις δικές μας σχεδιαστικές ιδέες ή περιέργειά μας. Η χρήση της γραμματοσειράς Helvetica, ειδικά στην επικεφαλίδα, θυμίζει τις βρετανικές σιδηροδρομικές πινακίδες. Ένας τέτοιος σχεδιασμός που βασίζεται στα συμφραζόμενα, μπορεί να τραβήξει το ενδιαφέρον των αναγνωστών, αν γίνει ως μια ελαφριά πινελιά που δεν παρεμβαίνει. Για χρωματισμό, χρησιμοποιήθηκε το χρώμα του φύλλου, για να συνδέσει οπτικά όλο το πακέτο. Τέλος, αναφέρθηκε ο ιστότοπος της πηγής, όπως πρέπει πάντα να γίνεται.

Υπάρχει η τάση από άτομα που ασχολούνται με τη στατιστική ή άτομα με υπόβαθρο στην επιστήμη δεδομένων, να χρησιμοποιούν κλασικές μορφές (όπως ραβδογραφήματα, γραφήματα διασποράς (scatter plots) ή γραμμικά γραφήματα (line charts)) για την οπτικοποίησή τους. Σε ένα πακέτο όμως, θα μπορούσαμε να επιτύχουμε καλύτερη εμπλοκή και ανάκληση αν χρησιμοποιούσαμε σχεδιαστική σκέψη. Το θέμα του γενικότερου πλαισίου (context) μπορεί επίσης να βοηθήσει προς αυτή την κατεύθυνση. Για παράδειγμα, για να δειχθούν οι αποστάσεις μετακίνησης στην Ατλάντα της Τζόρτζια, γίνεται χρήση χάρτη με δακτυλίους απόστασης και ραβδογράμματα που το καθιστά πιο ενημερωτικό και πιο αναγνωρίσιμο από τους αναγνώστες (Εικόνα 2.2-19).

Η καθηγήτρια απεικόνισης και σχεδίασης δεδομένων Isabelle Meirelles προτείνει τη χρήση μεταφορών σαν ένα καλό τρόπο για να κάνουμε τους λιγότερο στατιστικά εγγράμματους αναγνώστες, να κατανοήσουν καλύτερα τα δεδομένα και να γίνουν πιο χειροπιαστά για αυτούς. Κάποια τέτοια παραδείγματα είναι:

Το βίντεο εκπομπών άνθρακα της Νέας Υόρκης (<https://www.realworldvisuals.com/cv-projects/new-yorks-carbon-emissions>), στιγμιότυπό του φαίνεται στην Εικόνα 2.2-20, που εμφανίζεται ως ένας τεράστιος σωρός, σχεδιασμένος έτσι ώστε να μπορεί να προσεγγίσει το «άτομο στο δρόμο» (person on the street) (<https://www.realworldvisuals.com/climate-projects>).

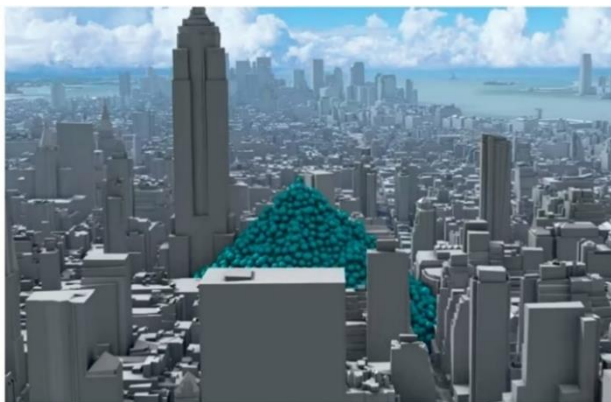


Figure 16.5 Volume of carbon dioxide emitted in New York City in one hour, visualized as a heap of one-ton balls smothering Midtown. A still from a video at youtu.be/DtqSlplGXOA, by Carbon Visuals (www.realworldvisuals.com), reproduced with permission.

Εικόνα 2.2-20: [1] Χειροπιαστό παράδειγμα

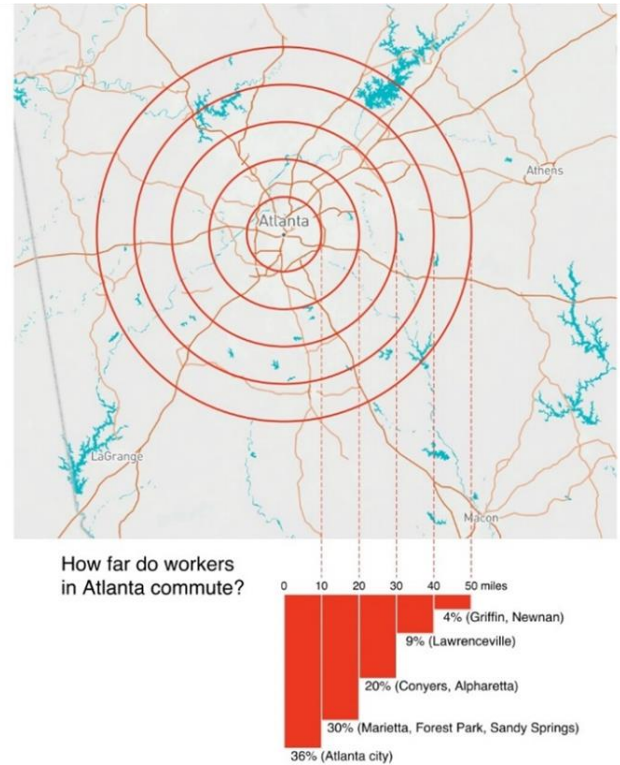


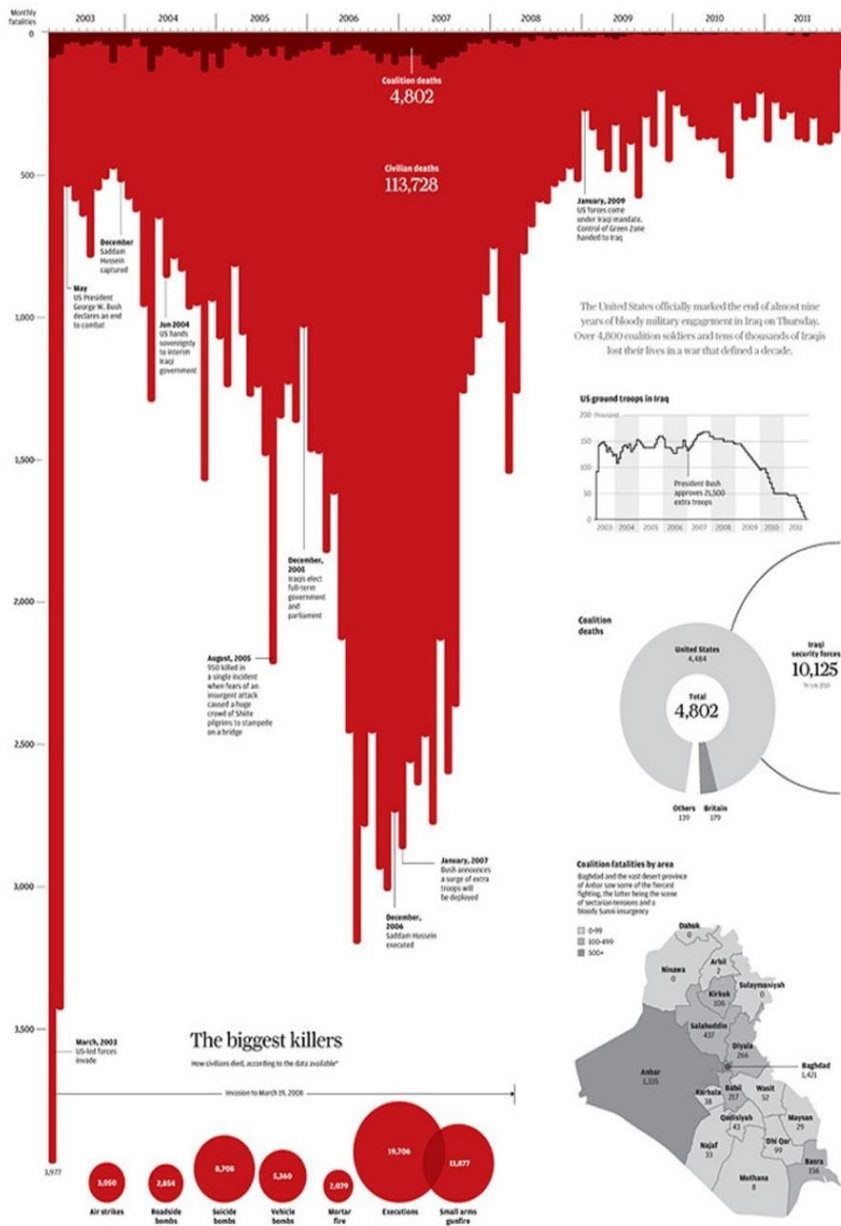
Figure 1.10 A first digital draft of a contextualized bar chart of commuting distances in Atlanta, Georgia. I decided to leave extra space on the map outside the 50-mile radius, so that the rings and the bar chart were visually spaced apart, linked just by the dashed vertical lines. Empty space can be valuable. This is ready to be given to some commuters for feedback, to check that it is clear and understood. Data from *Statistics: Unlocking the Power of Data* by the Locks, map image copyright Mapbox, map data copyright OpenStreetMap contributors.

Εικόνα 2.2-19: [1] Παράδειγμα εύκολα αναγνωρίσιμο

Το ραβδόγραμμα αριστερά (Εικόνα 2.2-21), χρησιμοποιείται για την ανάδειξη των θυμάτων στο Ιράκ με την πάροδο του χρόνου, αλλά το αναποδογύρισμά του και η χρήση κόκκινου χρώματος υποδηλώνει το αίμα που στάζει.

Του Simon Scarr για τη South China Morning Post (<https://www.scmp.com/infographics/article/1284683/iraqs-bloody-toll>).

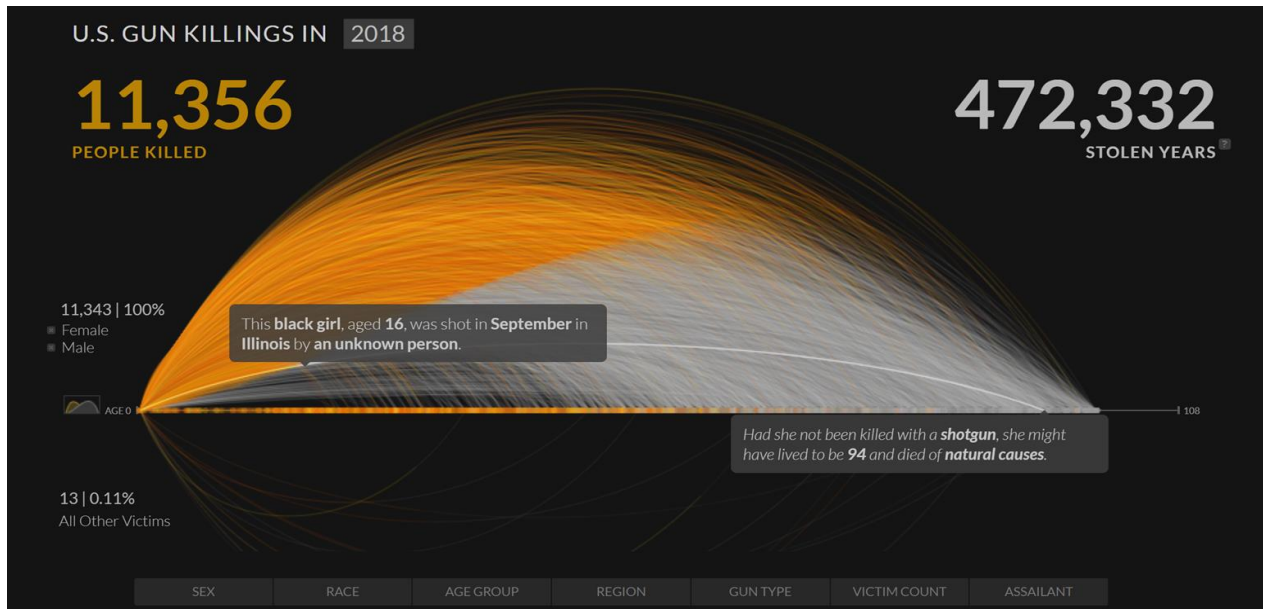
Iraq's bloody toll



Εικόνα 2.2-21: [5] Παράδειγμα εύκολα αναγνωρίσιμο και χειροπιαστό

Διακοπτόμενες τροχιές (Εικόνα 2.2-22) χρησιμοποιούνται για να δείξουν δολοφονίες με όπλα στις ΗΠΑ σε μια διαδικτυακή διαδραστική κινούμενη εικόνα (animation).

Από το διαδικτυακό πρακτορείο Periscope (<https://guns.periscope.com/>)



Εικόνα 2.2-22: [6] Παράδειγμα εύκολα αναγνωρίσιμο και χειροπιαστό

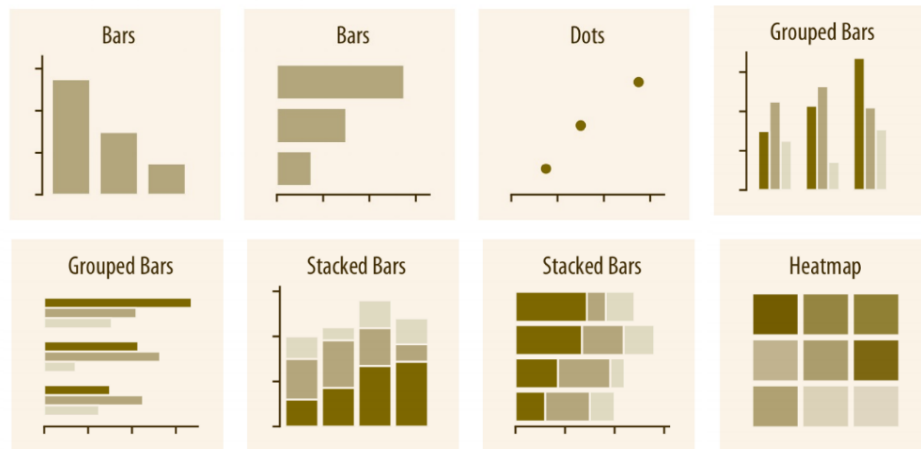
3 Γενικός Οδηγός Τύπων Πλοκής

Ας δούμε τους διαφορετικούς τύπους πλοκής ως γενικό οδηγό.

3.1 Ποσότητες (Amounts)

όταν ενδιαφερόμαστε να δείξουμε το μέγεθος των ποσοτικών τιμών.

Οι **ράβδοι** (Bars) είναι οι πιο χρησιμοποιούμενοι, αλλά και οι **κουκκίδες** (dots) εξυπηρετούν τον ίδιο σκοπό. Οι ράβδοι θα μπορούσαν επίσης να χρησιμοποιηθούν για να δείξουν ποσότητα για περισσότερα από δύο σύνολα κατηγοριών είτε **ομαδοποιημένες** (Grouped Bars), είτε **στοιβαγμένες μπάρες** (Stacked Bars) ή για την εμφάνιση ποσότητας ανά χρώμα με το **χάρτη θερμότητας** (Heatmap).



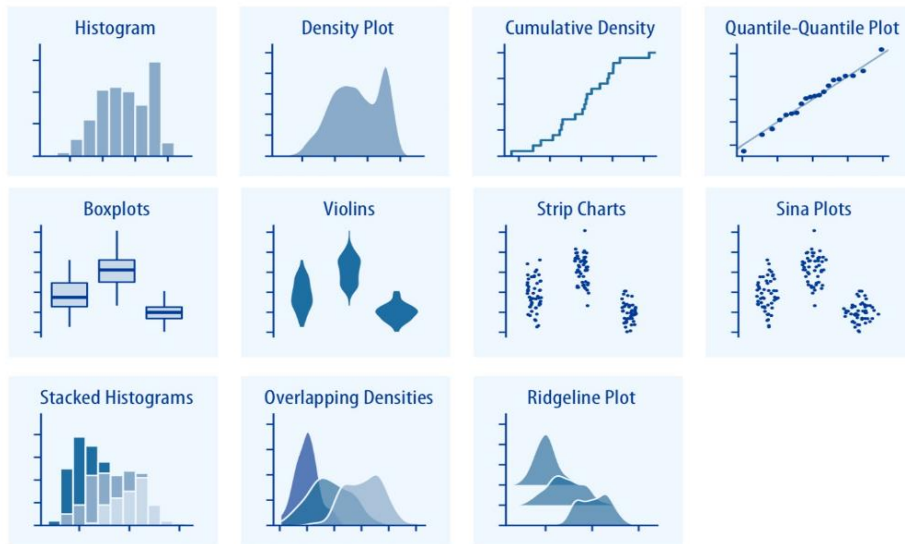
Εικόνα 3.1-1: [2] Γενικοί τύποι πλοκής Ποσοτήτων

3.2 Κατανομές (Distributions)

όταν ενδιαφερόμαστε να κατανοήσουμε πώς μια συγκεκριμένη μεταβλητή κατανέμεται σε ένα σύνολο δεδομένων.

Τα **ιστογράμματα** (Histograms), για την απεικόνιση μιας μεμονωμένης κατανομής, και τα **διαγράμματα πυκνότητας** (Density plots), για την απεικόνιση πολλών κατανομών ταυτόχρονα, με την κατανομή να σχεδιάζεται ως συνεχής καμπύλη, είναι τα πιο χρησιμοποιούμενα αν και, απαιτούν αυθαίρετες επιλογές παραμέτρων και μπορεί να είναι παραπλανητικά. Οι **αθροιστικές πυκνότητες** (Cumulative densities) και τα **διαγράμματα Quantile-Quantile** (q-q),

χρησιμοποιούνται για πιο ακριβή αναπαράσταση των δεδομένων (δεν απαιτούν αυθαίρετες επιλογές παραμέτρων και εμφανίζουν όλα τα δεδομένα ταυτόχρονα, αλλά συχνά, είναι δύσκολο να ερμηνευτούν). **Boxplots, Violin plots, Strip charts** και **Sina plots**, χρησιμοποιούνται για την οπτικοποίηση πολλαπλών κατανομών ταυτόχρονα ή/και εάν μας ενδιαφέρουν οι συνολικές μετατοπίσεις (overall shifts) μεταξύ των κατανομών. **Στοιβαγμένα ιστογράμματα** (Stacked histograms) και **αλληλεπικαλυπτόμενες πυκνότητες** (Overlapping densities), για τη σύγκριση σε μεγαλύτερο βάθος μικρότερου αριθμού κατανομών, δηλαδή πολλαπλές κατανομές για να οπτικοποιηθούν ταυτόχρονα (αν και τα στοιβαγμένα ιστογράμματα μπορεί να είναι δύσκολο να ερμηνευτούν και πρέπει να αποφευχθούν). Τα **Ridgeline διαγράμματα** (plots), χρησιμοποιούνται για την απεικόνιση πολύ μεγάλου αριθμού κατανομών ή των αλλαγών τους με την πάροδο του χρόνου και μπορούν επίσης να είναι μια χρήσιμη εναλλακτική λύση στα **διαγράμματα βιολιού** (Violin plots).



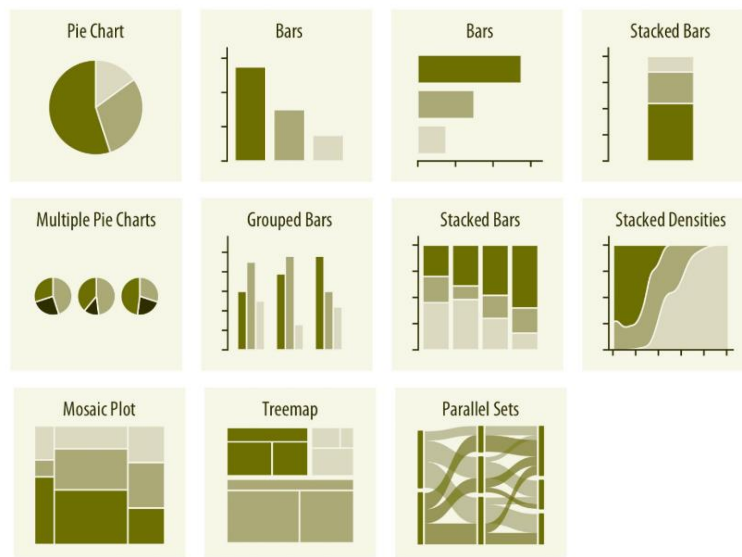
Εικόνα 3.2-2: [2] Γενικοί τύποι πλοκής Κατανομών

3.3 Αναλογίες (Proportions)

όταν θέλουμε να δείξουμε τα μεμονωμένα κομμάτια μιας οντότητας σπασμένα, αλλά ως μέρος ενός συνόλου (αναλογίες).

Αυτά που έχουν χρησιμοποιηθεί περισσότερο, είναι τα **διαγράμματα πίτας** (pie chart), τα οποία δίνουν έμφαση στο ότι κάθε κομμάτι προσθέτει στο σύνολο, είναι όμως καλύτερα να

αποφεύγονται για πολλαπλά σύνολα αναλογιών ή μεταβολές τους σε όλες τις συνθήκες. Οι αναλογίες μπορούν επίσης να απεικονιστούν από **γειτονικές μπάρες** (Side-by-side bars), με τις οποίες γίνεται ευκολότερα η σύγκριση μεταξύ τεμαχίων για έναν λογικό αριθμό συνθηκών (αλλά δεν μπορεί να απεικονίσει το μέγεθος κάθε μέρους/τεμαχίου σε σχέση με το σύνολο), **στοιβαγμένες ράβδοι** (Stacked Bars) χρήσιμες για τη σύγκριση πολλαπλών συνόλων αναλογιών εάν υπάρχουν μόνο δύο ράβδοι σε κάθε στοίβα (αλλά δεν διευκολύνει τη σύγκριση για περισσότερες, λόγω των διαφορετικών γραμμών βάσης), και **στοιβαγμένες πυκνότητες** (Stacked Densities) για αναλογίες που αλλάζουν κατά μήκος μιας συνεχούς μεταβλητής.



Εικόνα 3.3-3: [2] Γενικός τύπος πλοκής Αναλογιών

φωλιασμένες αναλογίες [nested proportions]: κάθε πρόσθετη κατηγορηματική μεταβλητή που προστίθεται, δημιουργεί μια λεπτότερη υποδιαίρεση φωλιασμένων δεδομένων εντός των προηγούμενων αναλογιών

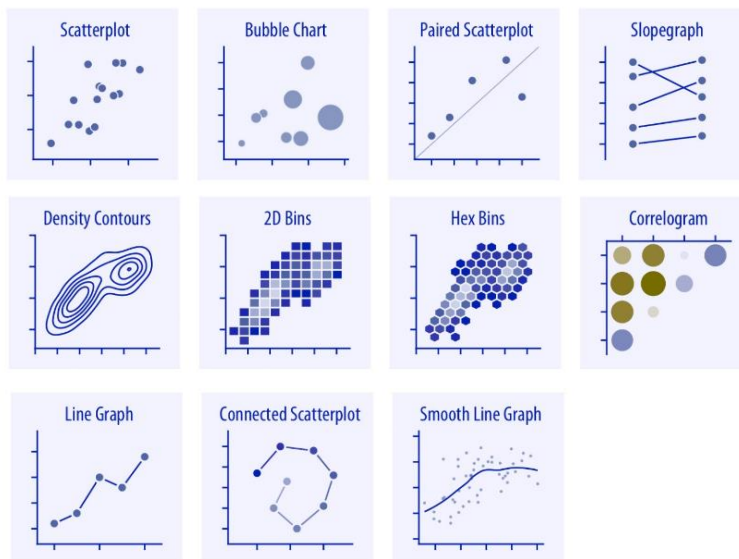
Για τη διάσπαση ενός συνόλου δεδομένων από πολλαπλές κατηγορηματικές μεταβλητές ταυτόχρονα μπορεί να χρησιμοποιηθεί: **γράφημα μωσαϊκού** (Mosaic plots) το οποίο προϋποθέτει ότι κάθε κατηγορηματική τιμή σε αυτό, καλύπτει όλες τις παρατηρήσεις μέσα στο σύνολο δεδομένων, **δέντρα/χάρτες** (Treemaps) τα οποία λειτουργούν καλά ακόμη και αν οι υποδιαίρεσεις μιας ομάδας είναι εντελώς διαφορετικές από τις υποδιαίρεσεις μιας άλλης, και, **παράλληλα**

σύνολα (Parallel Sets) τα οποία είναι καλύτερα από τα γραφήματα μωσαϊκού ή τα παράλληλα σύνολα εάν υπάρχουν περισσότερες από δύο μεταβλητές ομαδοποίησης

3.4 Σχέσεις X-Y (X-Y Relationships)

για σύνολα δεδομένων που περιέχουν δύο ή περισσότερες ποσοτικές μεταβλητές και μας ενδιαφέρει η σχέση μεταξύ αυτών των μεταβλητών.

Τα **διαγράμματα διασποράς** (scatterplots) είναι αυτά που χρησιμοποιούνται περισσότερο για αποτύπωση της σχέσης δύο ποσοτικών μεταβλητών ταυτόχρονα. Για την αποτύπωση τριών, θα μπορούσαμε να χρησιμοποιήσουμε ένα **γράφημα φυσαλίδων** (Bubble chart), ενώ για αποτύπωση περισσότερων από τρία, ένα **διάγραμμα διασποράς Matrix** (Scatterplot Matrix) ή ένα **Correlogram**. **Ζευγαρωμένα δεδομένα** (Paired data) για όταν οι μεταβλητές μετρούνται με τις ίδιες μονάδες, προσθέτοντας μια $x = y$ γραμμή. Αυτά, μπορούν επίσης να εμφανιστούν ως **διάγραμμα κλίσης** (Slopegraph) το οποίο χρησιμοποιεί ευθείες γραμμές για να συνδέσει τα ζευγαρωμένα σημεία. Contour Lines, 2D Bins και Hex bin, ίσως είναι μια καλή λύση για όταν έχουμε μεγάλο αριθμό σημείων και υπάρχουν προβλήματα επικάλυψης, ή, για μικρότερο αριθμό σημείων αλλά που υπάρχουν προβλήματα επικάλυψης. **Γραφήματα γραμμών** (Line Graphs), για όταν έχουμε μια μεταβλητή που έχει σημεία των δεδομένων σε χρονική σειρά, όπως ο χρόνος, ή μια μεταβλητή που επιβάλλει μια σειρά στα δεδομένα. Για όταν έχουμε δύο τέτοιες μεταβλητές,



Εικόνα 3.4-4: [2] Γενικός τύπος πλοκής X-Y Σχέσεων

μπορούμε να χρησιμοποιήσουμε το **συνδεδεμένο διάγραμμα διασποράς** (Connected Scatterplot) ενώ, οι **ομαλοποιημένες γραμμές** (Smooth lines) μπορούν να χρησιμοποιηθούν για να δείξουν την τάση που έχει ένα μεγαλύτερο σύνολο δεδομένων.

3.5 Γεωχωρικά Δεδομένα (Geospatial Data)

για την απεικόνιση γεωχωρικών δεδομένων στο κατάλληλο γεωχωρικό τους πλαίσιο (δηλαδή, γίνεται σύνδεση των πληροφοριών, με τις αντίστοιχες θέσεις στον φυσικό κόσμο ή με διάγραμμα που μοιάζει με χάρτη).

Το πιο συχνά χρησιμοποιημένο είναι ο **Χάρτης** (Map) που είναι εύκολα κατανοητός από τους αναγνώστες, αν και, δύσκολος να σχεδιαστεί. Παίρνει συντεταγμένες του πλανήτη και τις προβάλλει σε μια επίπεδη επιφάνεια, έτσι ώστε σχήματα και αποστάσεις στον πλανήτη να αντιπροσωπεύονται περίπου με σχήματα και αποστάσεις στην 2D αναπαράσταση. Το **Choropleth** μπορεί να χρησιμοποιηθεί για τις περιπτώσεις που θέλουμε να δείξουμε επιπρόσθετα τιμές δεδομένων χρωματίζοντας τις διαφορετικές περιοχές σύμφωνα με τα δεδομένα. Το **Χαρτογράφημα** (Cartograms), για τις φορές που η παραμόρφωση των διαφορετικών περιοχών δεν έχει σημασία και μπορεί να βοηθήσει να αναδειχθεί κάποια άλλη ποσότητα, ή ο **Χάρτης θερμότητας χαρτογράματος** (Cartogram Heatmap) για όταν θέλουμε να αναπαραστήσουμε κάθε περιοχή απλοποιημένη σε ισοδύναμου μεγέθους τετράγωνα.

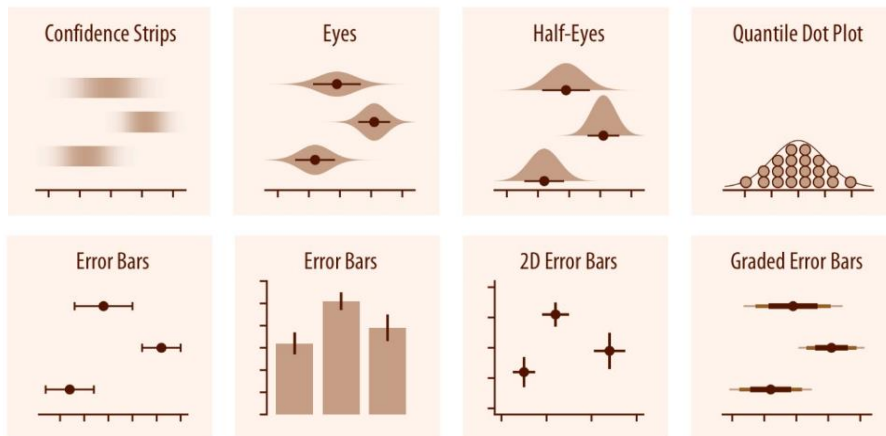


Εικόνα 3.5-5: [2] Γενικός τύπος πλοκής Γεωχωρικών Δεδομένων

3.6 Αβεβαιότητα (Uncertainty)

όταν τα δεδομένα μας περιλαμβάνουν κάποια αβεβαιότητα μέσα τους και θέλουμε να τη δείξουμε, όπως και με ποιον τρόπο

Τα πιο χρησιμοποιημένα είναι οι **μπάρες σφαλμάτων** (Error Bars), για να δείξουμε (με διάφορους τρόπους) το εύρος αβεβαιότητας της τιμής ή της μέτρησής της, και οι **λωρίδες αυτοπεποίθησης** (Confidence Strips), για λεπτομερέστερη απεικόνιση καθώς και για να δείξουμε την πραγματική αυτοπεποίθηση [the actual confidence] ή μεταγενέστερες κατανομές (και τα δύο είναι ακριβή, αλλά απαιτούνται ορισμένες γνώσεις για τη σωστή ερμηνεία της). **Διαβαθμισμένες μπάρες σφαλμάτων** (Graded Error Bars), για ταυτόχρονη εμφάνιση πολλών ευρών σφαλμάτων όπου το καθένα αντιστοιχεί σε διαφορετικό βαθμό εμπιστοσύνης. **Eye and Half-eyes**, είναι συνδυασμοί μπαρών σφάλματος με προσεγγίσεις για την οπτικοποίηση των κατανομών (βιολιά [violins] και κορυφογραμμές [ridgelines]) επομένως δείχνουν ακριβή εύρη για κάποια επίπεδα αυτοπεποίθησης καθώς και τη συνολική κατανομή αβεβαιότητας. Οι **ποσοστιαίες κουκίδες** (quantile dot plot) μπορεί να χρησιμεύσουν ως εναλλακτική απεικόνιση μιας κατανομής αβεβαιότητας, αν και παρόλο που δεν είναι τόσο ακριβής λόγω του ότι δείχνει την κατανομή σε διακριτές μονάδες, μπορεί να είναι πιο ευανάγνωστη από τη συνεχή κατανομή που φαίνεται από μια γραφική παράσταση βιολιού ή κορυφογραμμής.



Εικόνα 3.6-6: [2] Γενικός τύπος πλοκής Αβεβαιότητας

4 Αρχή Επίλυσης Γενικών Οπτικών Προβλημάτων...

Η «μείωση μελανιού» της ποσότητας των στοιχείων που δεν αφορούν τα ίδια τα δεδομένα σε μία οπτικοποίηση (όπως είναι για παράδειγμα: οι άξονες του διαγράμματος, οι ετικέτες τους, οι τίτλοι των αξόνων, τα σημεία που έχουν πάνω οι άξονες, το υπόμνημα, κτλ), πολλές φορές μπορεί να οδηγήσει σε πιο κομψή και λιγότερο ακατάστατη οπτικοποίηση. Όμως θέλει προσοχή να μην το παρακάνουμε, καθώς μειώνοντας τα στοιχεία που αποτελούν το γενικότερο πλαίσιο περιεχομένου /συμφραζόμενα (όπως είναι τα σημεία σε ένα scatterplot, οι ράβδοι σε ένα ραβδόγραμμα, οι σκιές σε ένα χάρτη θερμότητας κτλ), μπορούν να δημιουργήσουν σύγχυση στις οπτικοποιήσεις και να γίνουν δύσκολες να ερμηνευτούν.

4.1 Πρόβλημα 1. Αναλογία Δεδομένων - Μελανιού

Η μεγιστοποίηση της αναλογίας δεδομένων -μελανιού εντός λογικών ορίων, αφαιρεί τη σύγχυση και την ακαταστασία και προσπαθεί να αποδώσει καθαρότερες και κομψότερες οπτικοποιήσεις (εάν δεν το παρακάνουμε). Για παράδειγμα, στην Εικόνα 4.1-1 αριστερά: υπάρχει «πολύ μελάνι»

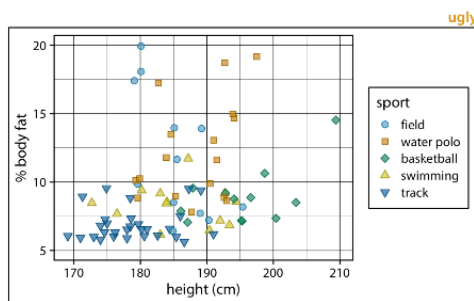


Figure 23-1. Percent body fat versus height in professional male Australian athletes. Each point represents one athlete. This figure devotes way too much ink to non-data. There are unnecessary frames around the entire figure, around the plot panel, and around the legend. The coordinate grid is very prominent, and its presence draws attention away from the data points. Data source: [Telford and Cunningham 1991].

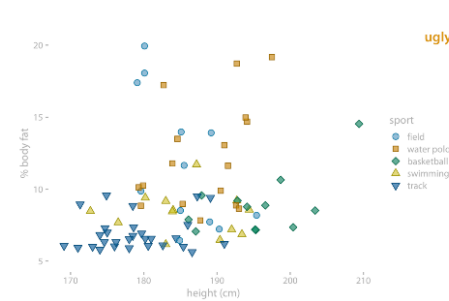


Figure 23-3. Percent body fat versus height in professional male Australian athletes. In this example, the concept of removing non-data ink has been taken too far. The axis tick labels and title are too faint and are barely visible. The data points seem to float in space. The points in the legend are not sufficiently set off from the data points, and the casual observer might think they are part of the data. Data source: [Telford and Cunningham 1991].

Εικόνα 4.1-1: [2] Παράδειγμα προβλήματος αναλογίας δεδομένων-μελανιού

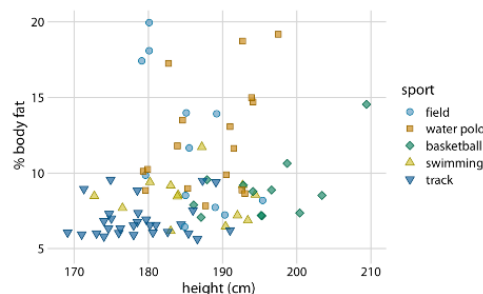


Figure 23-2. Percent body fat versus height in professional male Australian athletes. This figure is a cleaned-up version of Figure 23-1. Unnecessary frames have been removed, minor grid lines have been removed, and major grid lines have been drawn in light gray to stand back relative to the data points. Data source: [Telford and Cunningham 1991].

Εικόνα 4.1-2: [2] Παράδειγμα λύσης προβλήματος αναλογίας δεδομένων-μελανιού

σε διάφορα στοιχεία που δεν αφορούν τα ίδια τα δεδομένα, στην Εικόνα 4.1-1 (δεξιά): η απαλοιφή των στοιχείων που δεν αφορούν δεδομένα έγινε σε πολύ μεγάλο βαθμό με αποτέλεσμα να προκαλεί σύγχυση η τελική οπτικοποίηση, ενώ στην Εικόνα 4.1-2: βλέπουμε μια μινιμαλιστική έκδοση όπου πολλά στοιχεία που δεν αφορούν δεδομένα έχουν απαλειφτεί και διατηρείται μια καλή ισορροπία.

4.2 Πρόβλημα 2. Αιώρηση

Μειώνοντας πολύ το μελάνι στα υπόλοιπα στοιχεία πλην των δεδομένων, δίνεται η αίσθηση ότι η απεικόνιση επιπλέει και δεν υπάρχει σαφή αναφορά σε τίποτα, με αποτέλεσμα αυτό να αποτελεί μεγάλο πρόβλημα για τα διαγράμματα μικρών πολλαπλάσιων (small multiples plots), όπως για παράδειγμα στην Εικόνα 4.2-3 αριστερά: οι ράβδοι δεν έχουν μια βασική γραμμή και έτσι, δεν μπορούμε να δούμε καθαρά σε ποιο μέρος ανήκει ποια όψη.

Λύση: μπορούμε να χρησιμοποιήσουμε ανοιχτό γκρι χρώμα για φόντο, με λεπτές οριζόντιες γραμμές ως πλέγμα για το κάθε ένα από αυτά, όπως στην Εικόνα 4.2-3 δεξιά. Τα πλέγματα (grids) βοηθούν τον αναγνώστη να διακρίνει συγκεκριμένες τιμές και να τις συγκρίνει με άλλες από διαφορετικά μέρη του διαγράμματος. Επίσης, βοηθάει το διάγραμμα να δείχνει ως μια ενιαία οπτική οντότητα και εμποδίζει την εμφάνιση της διαγράμματος ως ένα λευκό πλαίσιο που περιβάλλεται από σκούρο κείμενο.

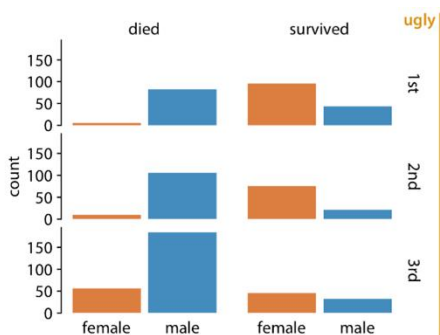


Figure 23-5. Survival of passengers on the Titanic, broken down by gender and class. This small multiples plot is too minimalistic. The individual facets are not framed, so it's difficult to see which part of the figure belongs to which facet. Further, the individual bars are not anchored to a baseline, and they seem to float. Data source: Encyclopedia Titanica.

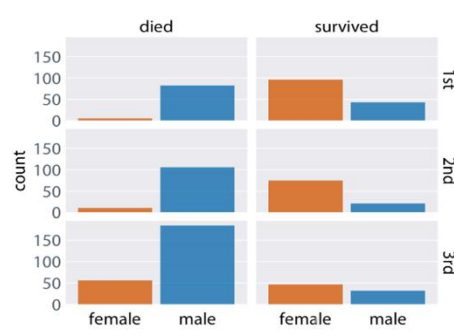


Figure 23-6. Survival of passengers on the Titanic, broken down by gender and class. This is an improved version of Figure 23-5. The gray background in each facet clearly delineates the six groupings (survived or died in 1st, 2nd, or 3rd class) that make up this plot. Thin horizontal lines in the background provide a reference for the bar heights and facilitate comparison of bar heights among facets. Alternatively, we could put a frame around each individual plot panel and use gray bars to highlight the grouping variables (see Figure 21-1). Data source: Encyclopedia Titanica.

Εικόνα 4.2-3: [2] Παράδειγμα προβλήματος-λύσης μείωσης μελανιού σε μικρά πολλαπλάσια

4.3 Πρόβλημα 3. Πλέγμα

Η χρήση ή όχι του πλέγματος, χρειάζεται προσεκτική σκέψη, για το πώς πρέπει να τοποθετηθεί στο διάγραμμα και πόσο πυκνό θα πρέπει να γίνει. Αυτό είναι σημαντικό, καθώς το πλέγμα προσθέτει οπτικό θόρυβο και μειώνει την προσοχή των αναγνωστών από τα πραγματικά δεδομένα, ειδικά εάν αυτά είναι πολύ πυκνά τοποθετημένα, ή εάν είναι πολύ περισσότερο ορατά από ότι πρέπει, ή έχουν μεγάλες και μικρές γραμμές.

Λύση: η ένταση χρώματος του σχήματος στο προσκήνιο, η τυπογραφία των κειμένων που βρίσκονται γύρω από το σχήμα και τα χρώματα τα οποία χρησιμοποιούνται, πρέπει επίσης να ταιριάζουν με το γκρι φόντο που θέλουμε να βάλουμε, οπότε θα πρέπει να το προσαρμόσουμε αναλόγως ώστε ο τόνος του γκρι να είναι τέτοιος που να μην χαλάει την αισθητική της οπτικοποίησης.

Εικόνα 4.3-4, αναλυτική επεξήγηση: Αφαιρώντας το γκρι φόντο από το διάγραμμα της εικόνας αριστερά και από το υπόμνημα, και προσθέτοντας ένα μινιμαλιστικό πλέγμα οριζόντιων γραμμών που τοποθετούνται μόνο στα κύρια σημάδια του άξονα y, μιας και αυτές οι τιμές y μας ενδιαφέρουν και αφαιρώντας έτσι και τη γραμμή του άξονα y, βοηθά πολύ στην οπτικοποίηση του διαγράμματος (δεξιά).

Οι γραμμές πλέγματος (Gridlines) τείνουν να είναι πιο χρήσιμες, όταν τοποθετούνται κάθετα στη μεταβλητή που μας ενδιαφέρει (αλλά να είμαστε συνετοί σχετικά με αυτές).

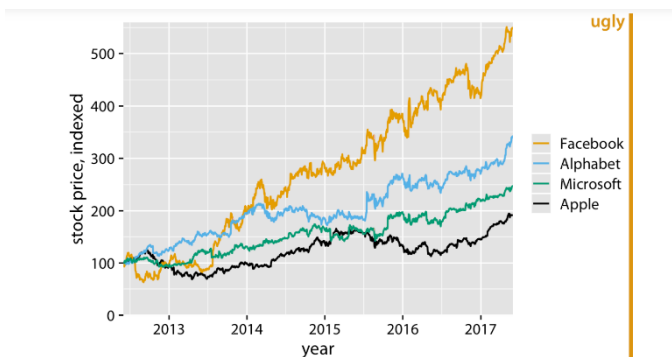


Figure 23-7. Stock price over time for four major tech companies. The stock price for each company has been normalized to equal 100 in June 2012. This figure mimics the ggplot2 default look, with white major and minor grid lines on a gray background. In this particular example, I think the grid lines overpower the data lines, and the result is a figure that is not well balanced and that doesn't place sufficient emphasis on the data. Data source: Yahoo! Finance.

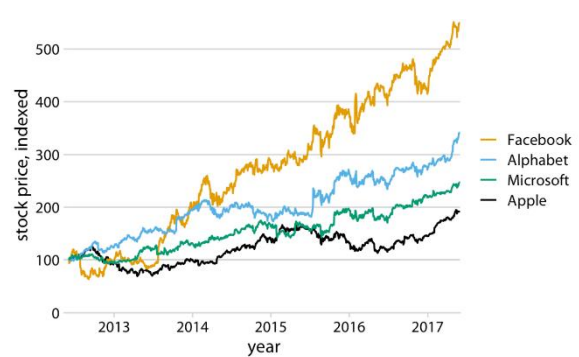


Figure 23-10. Indexed stock price over time for four major tech companies. Adding thin horizontal lines at all major y-axis ticks provides a better set of reference points than just the one horizontal line of Figure 23-9. This design also removes the need for prominent x- and y-axis lines, since the evenly spaced horizontal lines create a visual frame for the plot panel. Data source: Yahoo! Finance.

Εικόνα 4.3-4: [2] Παράδειγμα προβλήματος-λύσης χρήσης πλέγματος

Το πλέγμα και των δύο αξόνων στο φόντο, είναι το πιο κατάλληλο συνήθως για τα διαγράμματα διασποράς, όπου οι γραμμές των αξόνων σε γενικές γραμμές δεν χρειάζονται αφού δεν υπάρχει άξονας ενδιαφέροντος.

Τις περισσότερες φορές, ένα ελαφρύ πλέγμα σε λευκό φόντο εξυπηρετεί μια χαρά.

Η τοποθέτηση λευκών γραμμών πάνω από τις ράβδους/μπάρες, δίνει την εντύπωση ότι τις χωρίζουμε σε τμήματα και οι απόψεις δίστανται σχετικά με αυτές (Εικόνα 4.3-5): από τη μία, ότι βοηθά τον αναγνώστη να αντιληφθεί το μήκος των ράβδων, και από την άλλη, ότι δεν σχηματίζεται οπτική ενότητα στις μπάρες. Επίσης, επειδή οι αριθμοί με τα ποσοστά δεν χωρούν ακριβώς στα τελευταία κομμάτια των ράβδων (εφόσον δεν έχουν όλα τα τελευταία τμήματα τα ίδια μήκη), αυτό επιμηκύνει τις ράβδους μιας και προστίθενται δίπλα (βλέπε Εικόνα 4.3-5) και πρέπει να αποφεύγεται όταν αυτό είναι δυνατό.

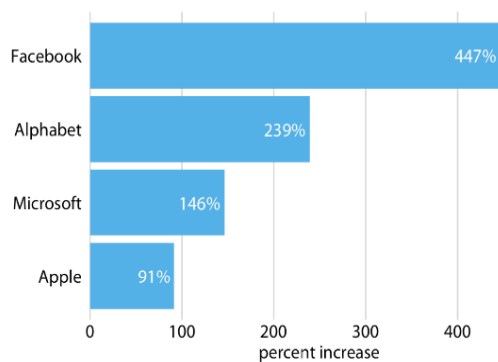


Figure 23-11. Percent increase in stock price from June 2012 to June 2017, for four major tech companies. Because the bars run horizontally, vertical grid lines are appropriate here. Data source: Yahoo! Finance.

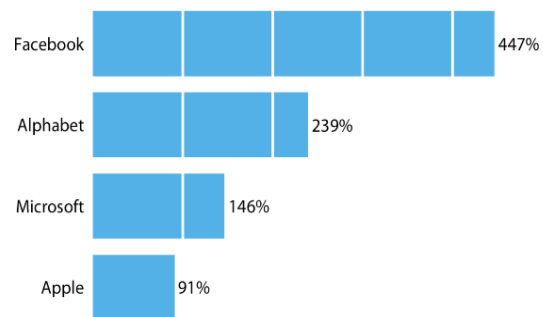


Figure 23-12. Percent increase in stock price from June 2012 to June 2017, for four major tech companies. White grid lines on top of bars can help the reader perceive the relative lengths of the bars. At the same time, they can also create the perception that the bars are falling apart. Data source: Yahoo! Finance.

Εικόνα 4.3-5: [2] Παράδειγμα προβλήματος χρήσης πλέγματος σε ράβδους

Για μεγάλο αριθμό σημείων δεδομένων όπου η σχετική σύγκριση είναι η γραμμή $x = y$, όπως σε διαγράμματα διασποράς ζευγαρωμένων δεδομένων, η σχεδίαση μιας διαγώνιας γραμμής αντί για ένα πλέγμα, εξυπηρετεί καλύτερα στο να δούμε αμέσως τον σκοπό του γραφήματος (Εικόνα 4.3-7). Εάν δεν φαίνεται αρκετά καλά, το να κάνουμε πιο έντονη τη γραμμή και κάνοντας τα σημεία μεγαλύτερα ή πιο σκούρα ίσως βοηθήσει.

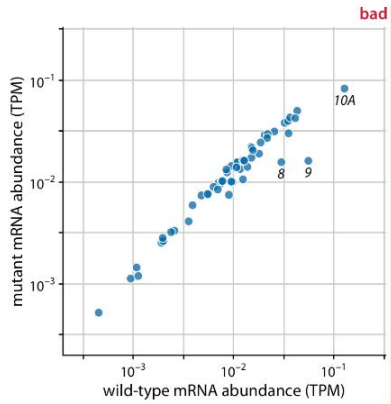


Figure 23-14. Gene expression levels in a mutant bacteriophage T7 relative to wild type. By plotting this dataset against a background grid instead of a diagonal line, we are obscuring which genes are higher or lower in the mutant than in the wild-type bacteriophage. Data source: [Paff et al. 2018].

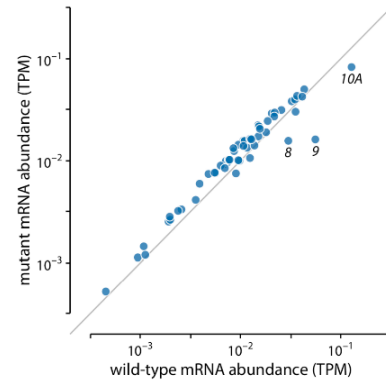


Figure 23-13. Gene expression levels in a mutant bacteriophage T7 relative to wild type. Gene expression levels are measured by mRNA abundances, in transcripts per million (TPM). Each dot corresponds to one gene. In the mutant bacteriophage T7, the promoter in front of gene 9 was deleted, and this resulted in reduced mRNA abundances of gene 9 as well as the neighboring genes 8 and 10A (highlighted). Data source: [Paff et al. 2018].

Εικόνα 4.3-7: [2] Παράδειγμα προβλήματος-λύσης μεγάλου αριθμού σημείων δεδομένων συγκριτικής γραμμής $x = y$

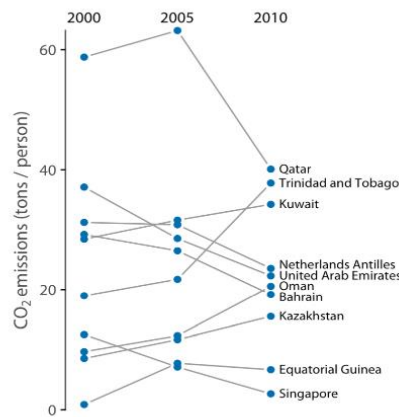


Figure 12-13. CO₂ emissions per person in 2000, 2005, and 2010, for the 10 countries with the largest difference between the years 2000 and 2010. Data source: Carbon Dioxide Information Analysis Center.

Εικόνα 4.3-6: [2] Παράδειγμα μικρού αριθμού σημείων δεδομένων με προσδιορισμό σε αυτά

Για μικρό αριθμό σημείων δεδομένων όπου μας ενδιαφέρει να προσδιορίσουμε κάθε μεμονωμένο σημείο, το διάγραμμα κλίσης (Slopegraph) είναι η καλύτερη επιλογή (Εικόνα 4.3-6).

Τα διαγράμματα κλίσης έχουν το πλεονέκτημα (σε σχέση με τα διαγράμματα διασποράς) ότι μπορούν να χρησιμοποιηθούν για σύγκριση περισσότερων από δύο μετρήσεων τη φορά.

4.4 Πρόβλημα 4. Σύνθετες Οπτικοποιήσεις

Η δημιουργία σύνθετων/πολύπλοκων εξελιγμένων οπτικοποιήσεων δεδομένων, αν και μπορεί να είναι πολύ εντυπωσιακές, κάνουν τον αναγνώστη να μπερδεύεται σε σχέση με το ποιο είναι το μήνυμα.

Λύση: μπορούμε να χρησιμοποιήσουμε απλά γραφικά που μεταφέρουν σαφώς το μήνυμα.

Παράδειγμα: εάν στην Εικόνα 4.4-8(α), το μήνυμά μας είναι να δείξουμε ότι η American και η Delta έχουν τις μικρότερες καθυστερήσεις άφιξης, η οπτικοποίηση του σε ένα απλό ραβδογράφημα, το μεταφέρει πολύ καλύτερα (Εικόνα 4.4-8(β)). Και αν θέλουμε να δείξουμε ότι σε αυτές τις αεροπορικές εταιρείες οι μικρές καθυστερήσεις οφείλονται στο ότι δεν γίνονται συχνές πτήσεις έξω από την περιοχή της Νέας Υόρκης, θα μπορούσαμε να το δείξουμε με ένα δεύτερο γράφημα που να περιλαμβάνει τους κύριους αερομεταφορείς της περιοχής (Εικόνα 4.4-8(γ)). Και στα δύο γραφήματα, απορρίπτουμε τη μεταβλητή απόστασης της Εικόνα 4.4-8(α).

Τα απλά και σαφή διαγράμματα είναι πάντα καλύτερα από τα περίπλοκα και μπερδεμένα, ακόμα κι αν είναι εντυπωσιακά. Με το να δείχνουμε πάρα πολλές πληροφορίες ταυτόχρονα, μπορεί να καταλήξουμε να μην δείχνουμε τίποτα.

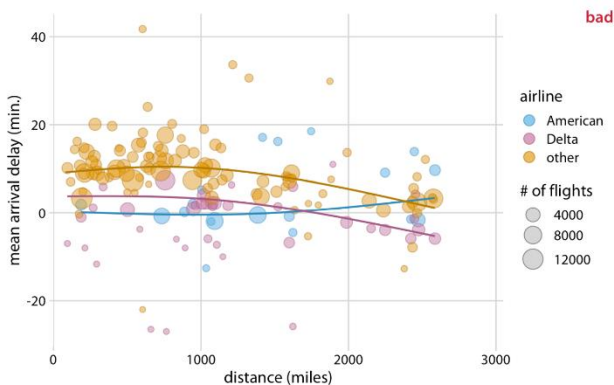


Figure 29-3. Mean arrival delay versus distance from New York City. Each point represents one destination, and the size of each point represents the number of flights from one of the three major New York City airports (Newark, JFK, or LaGuardia) to that destination in 2013. Negative delays imply that the flight arrived early. Solid lines represent the mean trends between arrival delay and distance. Delta has consistently lower arrival delays than other airlines, regardless of distance traveled. American has among the lowest delays, on average, for short distances, but has among the highest delays for longer distances traveled. This figure is labeled as "bad" because it is overly complex. Most readers will find it confusing and will not intuitively grasp what it is the figure is showing. Data source: US Dept. of Transportation, Bureau of Transportation Statistics.

α

Εικόνα 4.4-8: [2] Παράδειγμα προβλήματος-λύσης πολύπλοκων οπτικοποιήσεων

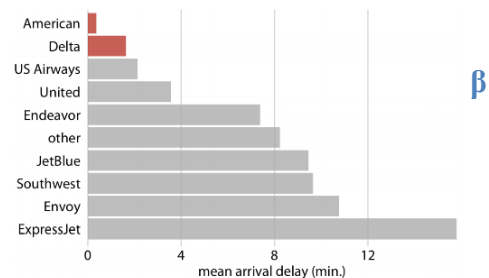


Figure 29-4. Mean arrival delay for flights out of the New York City area in 2013, by airline. American and Delta have the lowest mean arrival delays of all airlines flying out of the New York City area. Data source: US Dept. of Transportation, Bureau of Transportation Statistics.

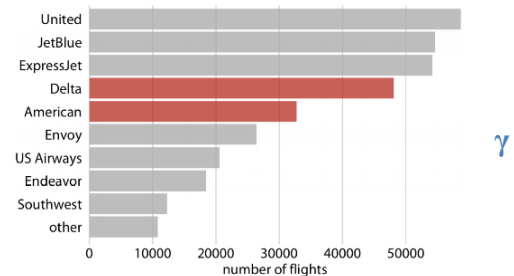


Figure 29-5. Number of flights out of the New York City area in 2013, by airline. Delta and American are the fourth and fifth largest carriers by flights out of the New York City area. Data source: US Dept. of Transportation, Bureau of Transportation Statistics.

Όταν θέλουμε να δείξουμε μια εξαιρετικά σύνθετη απεικόνιση, θα πρέπει πάντα να ξεκινάμε με ένα πιο εύπεπτο πρώτο διάγραμμα (αυτό είναι ιδιαίτερα χρήσιμο για τα μικρά πολλαπλάσια). Έχοντας δει και αφομοιώσει πρώτα το ένα σχήμα, είναι πολύ πιο εύκολο μετά να επεξεργαστούμε τις ίδιες πληροφορίες για περισσότερα

Παράδειγμα: έχοντας δει και κατανοήσει το πρώτο σχήμα της Εικόνα 4.4-9 επάνω, είναι πολύ πιο εύκολο να το καταλάβουμε/χωνέψουμε μετά για ολόκληρο το σύνολο (πιο σύνθετο και πολύπλοκο διάγραμμα) στο συγκεκριμένο: 10 αεροπορικές εταιρείες και 3 αεροδρόμια ταυτόχρονα (σχήματα Εικόνα 4.4-9 κάτω).

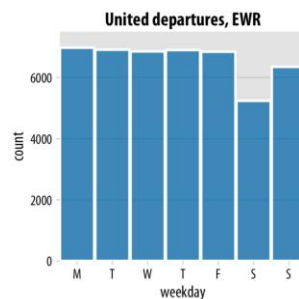


Figure 29-6. United Airlines departures out of Newark Airport (EWR) in 2013, by weekday. Most weekdays show approximately the same number of departures, but there are fewer departures on weekends. Data source: US Dept. of Transportation, Bureau of Transportation Statistics.

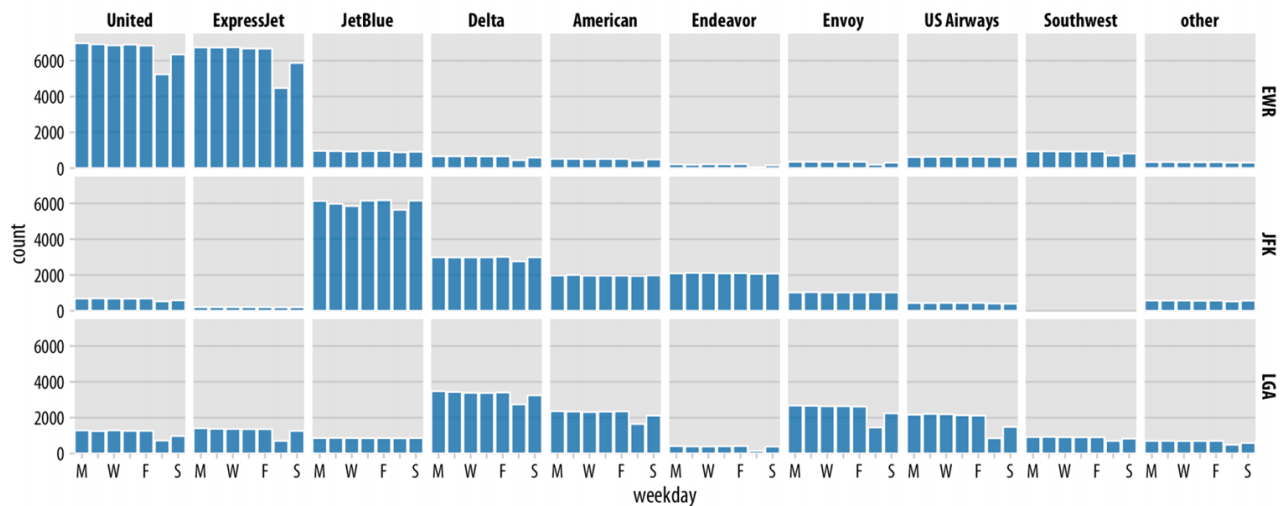


Figure 29-7. Departures out of airports in the New York City area in 2013, broken down by airline, airport, and weekday. United Airlines and ExpressJet make up most of the departures out of Newark Airport (EWR); JetBlue, Delta, American, and Endeavor make up most of the departures out of JFK; and Delta, American, Envoy, and US Airways make up most of the departures out of LaGuardia (LGA). Most but not all airlines have fewer departures on weekends than during the workweek. Data source: US Dept. of Transportation, Bureau of Transportation Statistics.

Εικόνα 4.4-9: [2] Παράδειγμα λύσης για εξαιρετικά σύνθετη οπτικοποίηση

Σε όλες τις επιλογές μας πρέπει να λάβουμε υπόψη και την οπτική επωνυμία και ταυτότητα (visual branding and identity). Πολλοί έχουν ένα αναγνωρίσιμο εταιρικό στυλ και η χρήση του σκιασμένου φόντου με μια συγκεκριμένη επιλογή πλέγματος φόντου μπορεί να βοηθήσει στη δημιουργία αυτής της μοναδικής οπτικής ταυτότητας.

4.5 Πρόβλημα 5. Γενίκευση Οπτικοποιήσεων

Παρόλο που οι απλές και καθαρές οπτικοποιήσεις αποτρέπουν τους αντιπερισπασμούς, είναι ευανάγνωστες και διατηρούν τους αναγνώστες μας συγκεντρωμένους στα πιο σημαντικά σημεία που θέλουμε να αναδείξουμε, καταλήγουν να φαίνονται πολύ γενικευμένες/μονότονες, χωρίς να έχουν χαρακτηριστικά που να ξεχωρίζουν ώστε να είναι αξιομνημόνευτες.

Παράδειγμα: κοιτάζοντας τα δύο σχήματα της Εικόνα 4.5-10, παρόλο που δεν έχουν τίποτα κοινό (εκτός από ότι είναι γραφήματα ράβδων), είναι οπτικά παρόμοια, χωρίς στοιχεία που να μας βοηθούν να αντιληφθούμε το θέμα τους καλύτερα, έτσι κανένα δεν είναι ιδιαίτερα αξιομνημόνευτο.

Λύση: Θα πρέπει να διατηρείται μια ισορροπία μεταξύ των «αξιομνημόνευτο αλλά μπερδεμένο» και «ξεκάθαρο αλλά μονότονο και βαρετό». Φυσικά, ρόλο παίζει και το σε ποιον απευθυνόμαστε (όπως, για τεχνικό επιστημονικό περιοδικό, ή για ευρείας ανάγνωσης εφημερίδα, ή blog). Η έρευνα στην ανθρώπινη αντίληψη έχει δείξει ότι η μοναδικότητα των σχημάτων κάνει τις οπτικοποιήσεις πιο αξιομνημόνευτες αν αυτή δεν εμποδίζει τον αναγνώστη να έχει μια γρήγορη και σαφή επισκόπηση των πληροφοριών.

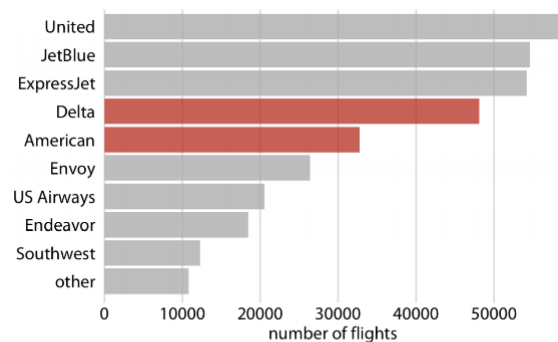


Figure 29-5. Number of flights out of the New York City area in 2013, by airline. Delta and American are the fourth and fifth largest carriers by flights out of the New York City area. Data source: US Dept. of Transportation, Bureau of Transportation Statistics.

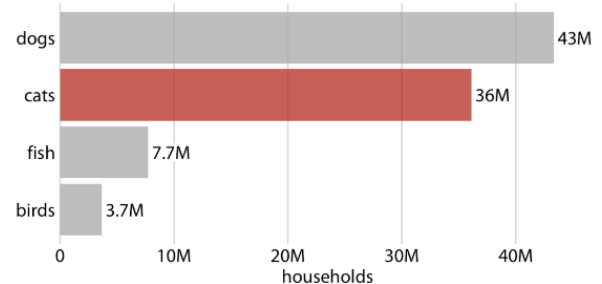


Figure 29-8. Number of households having one or more of the most popular pets: dogs, cats, fish, or birds. This bar graph is perfectly clear but not necessarily particularly memorable. The “cats” column has been highlighted solely to create visual similarity with Figure 29-5. Data source: 2012 US Pet Ownership & Demographics Sourcebook, American Veterinary Medical Association.

Εικόνα 4.5-10: [2] Παραδείγματα γενικευμένων γραφημάτων

Γενικά, ο στόχος μας πρέπει να είναι **καθαρές και αξιομνημόνευτες οπτικοποιήσεις**.

Θα μπορούσαμε να προσθέσουμε κάποια οπτικά στοιχεία που να αντικατοπτρίζουν χαρακτηριστικά των δεδομένων (όπως, σχέδια - εικονογράμματα (pictograms) πραγμάτων/αντικειμένων) στα οποία αναφέρεται το σύνολο δεδομένων. Στην Εικόνα 4.5-11, φαίνεται ένα ισοτυπικό διάγραμμα (isotype plot) όπου εικόνες χρησιμοποιήθηκαν για να αναπαραστήσουν τα ίδια τα δεδομένα με τη μορφή επαναλαμβανόμενων εικόνων κρατώντας τη λογική των ράβδων.

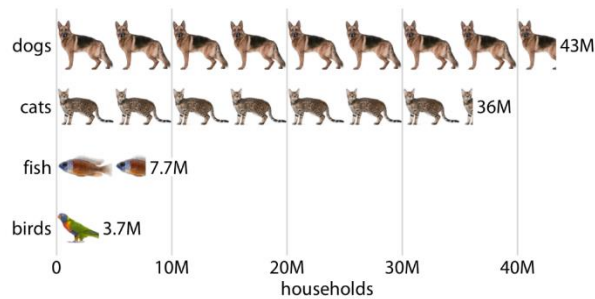


Figure 29-9. Number of households having one or more of the most popular pets, shown as an isotype plot. Each complete animal represents 5 million households that have that kind of pet. Data source: 2012 US Pet Ownership & Demographics Sourcebook, American Veterinary Medical Association.

Εικόνα 4.5-11: [2] Ισοτυπικό διάγραμμα (isotype plot)

4.6 Πρόβλημα 6. Οπτική Διάκριση

Το να χρησιμοποιείται μια σταθερή οπτική γλώσσα για όλα τα διαφορετικά μέρη μιας μεγαλύτερης απεικόνισης, ισχύει εξίσου και για τα διαγράμματα. Για να μπορέσουν να γίνουν μέρος μιας μεγαλύτερης ιστορίας, πρέπει να σχεδιαστούν έτσι ώστε να μοιάζουν ότι ανήκουν και αυτά στην ίδια ιστορία, χωρίς όμως να είναι ακριβώς τα ίδια (Εικόνα 4.6-12 και Εικόνα 4.6-13, αριστερά σχήματα), αντιθέτως! Σχήματα που περιγράφουν διαφορετικές αναλύσεις θα πρέπει να είναι οπτικά διακριτά, έτσι ώστε οι αναγνώστες να μπορούν εύκολα να αναγνωρίσουν πού τελειώνει μια ανάλυση και πού αρχίζει η επόμενη. Αυτό γίνεται με το να χρησιμοποιούμε διαφορετικές προσεγγίσεις οπτικοποίησης για τα διαφορετικά τμήματα/κομμάτια της μεγαλύτερης ιστορίας μας, έχοντας όμως κοινή οπτική γλώσσα (Εικόνα 4.6-12 και Εικόνα 4.6-13, δεξιά σχήματα).

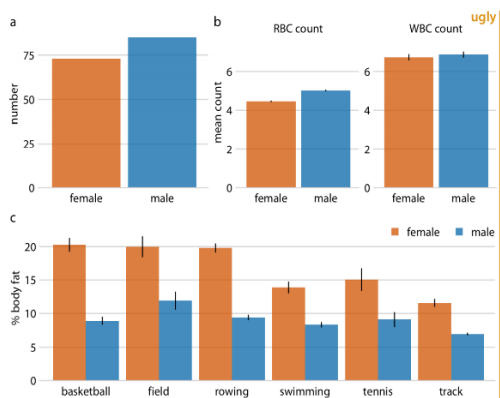


Figure 29-10. Physiology and body composition of male and female athletes. Error bars indicate the standard error of the mean. This figure is overly repetitive. It shows the same data as Figure 21-8 and it uses a consistent visual language, but all the subfigures use the same type of visualization (bar plots). This makes it difficult for the reader to process that parts (a), (b), and (c) show entirely different results. Data source: [Telford and Cunningham 1991].

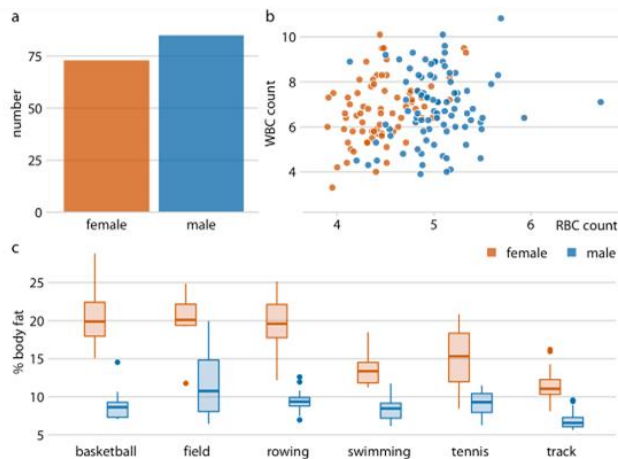


Figure 21-8. Physiology and body composition of male and female athletes. This figure shows the exact same data as Figure 21-7, but now uses a consistent visual language. Data for female athletes is always shown to the left of the corresponding data for male athletes, and genders are consistently color-coded throughout all elements of the figure. Data source: [Telford and Cunningham 1991].

Εικόνα 4.6-12: [2] Παράδειγμα 1 κοινής οπτικής γλώσσας αλλά όχι όμοιας σχεδίασης

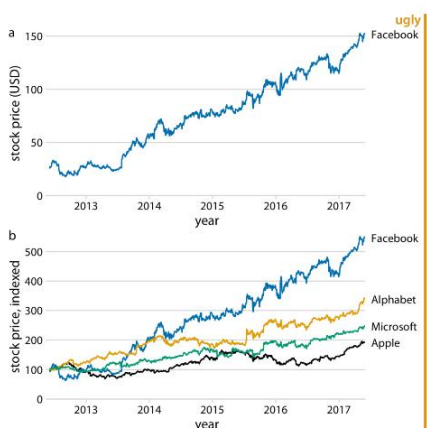


Figure 29-11. Growth of Facebook stock price over a five-year interval and comparison with other tech stocks. (a) The Facebook stock price rose from around \$25/share in mid-2012 to \$150/share in mid-2017. (b) The prices of other large tech companies did not rise comparably over the same time period. Prices have been indexed to 100 on June 1, 2012 to allow for easy comparison. This figure is labeled as “ugly” because parts (a) and (b) are repetitive. Data source: Yahoo! Finance.

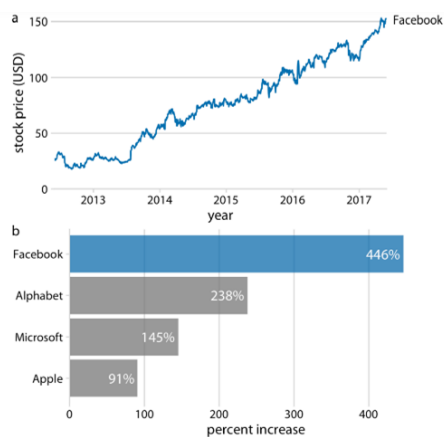


Figure 29-12. Growth of Facebook stock price over a five-year interval and comparison with other tech stocks. (a) The Facebook stock price rose from around \$25/share in mid-2012 to \$150/share in mid-2017, an increase of almost 450%. (b) The prices of other large tech companies did not rise comparably over the same time period. Price increases ranged from around 90% to almost 240%. Data source: Yahoo! Finance.

Εικόνα 4.6-13: [2] Παράδειγμα 2 κοινής οπτικής γλώσσας αλλά όχι όμοιας σχεδίασης

Έτσι, για τη δημιουργία μιας παρουσίασης, μιας αναφοράς ή μιας έκθεσης που έχουμε να ετοιμάσουμε για την οπτικοποίησή μας, πρέπει να στοχεύουμε σε διαφορετικούς τύπους διαγραμμάτων για κάθε ξεχωριστή ανάλυση μέσα σε αυτή, χρησιμοποιώντας όμως την ίδια οπτική γλώσσα, ώστε να φαίνονται κομμάτι του ίδιου συνόλου.

Ένας γενικός κανόνας που θα μπορούσαμε να χρησιμοποιήσουμε για την αφήγηση μιας ιστορίας, είναι:

- Να στοχεύουμε για την εμφάνιση όσο το δυνατό πιο ακατέργαστων δεδομένων και,
- για τα διαγράμματα που ακολουθούν τα ακατέργαστα δεδομένα, την εμφάνιση ολοένα και περισσότερων παραγόμενων ποσοτήτων (ποσοστιαίες αυξήσεις, μέσοι όροι κτλ). Αυτά, είναι πολύ χρήσιμα για τη σύνοψη βασικών τάσεων σε μεγάλα και πολύπλοκα σύνολα δεδομένων, όμως, καθώς έχουν παραχθεί από τα ακατέργαστα δεδομένα, είναι λιγότερο διαισθητικά. Με αποτέλεσμα, εάν εμφανιστούν πριν από τα ακατέργαστα δεδομένα, οι αναγνώστες θα δυσκολευτούν να τα παρακολουθήσουν. Από την άλλη μεριά, εάν προσπαθήσουμε να δείξουμε όλες τις τάσεις των ακατέργαστων δεδομένων, το αποτέλεσμα θα είναι, παραπάνω σχήματα από όσα χρειάζονται και που είναι πολύ πιθανό να επαναλαμβάνονται.

Ο αριθμός των διαγραμμάτων/σχημάτων που χρειάζονται ώστε να ειπωθεί μια ιστορία, εξαρτάται από το μέσο για το οποίο αυτή προορίζεται. Έτσι:

- Για tweets και μικρές αναρτήσεις σε blogs, ένα σχήμα είναι επαρκές.
- Για επιστημονικές εργασίες, η χρήση τριών έως έξι σχημάτων είναι συνήθως καλή αναλογία. Εάν θέλουμε να χρησιμοποιήσουμε περισσότερα, θα πρέπει να τα μεταφέρουμε σε ένα παράρτημα ή σαν μια συμπληρωματική ενότητα, ώστε να μην κουράσουμε τους αναγνώστες παρουσιάζοντας υπερβολικό αριθμό, παρόμοιων συνήθως, μορφών σχημάτων.
- Σε κάποιες άλλες περιπτώσεις μπορεί να είναι απαραίτητος και μεγαλύτερος αριθμός σχημάτων, αλλά συνήθως αφορά την αφήγηση πολλαπλών ιστοριών ή μιας γενικής ιστορίας με δευτερεύοντες ιστορίες τριών έως έξι σχημάτων. Ας πούμε για παράδειγμα, αν θέλουμε να ετοιμάσουμε μια επιστημονική παρουσίαση διάρκειας μίας ώρας, θα πρέπει να στοχεύσουμε σε τρεις ιστορίες. Για ένα βιβλίο όμως, μία ιστορία ανά κεφάλαιο/ενότητα ίσως αρκεί.

Γενικά πρέπει να υπάρχει μια ισορροπία μεταξύ της υπερφόρτωσης ενός σχήματος με τα στοιχεία μελάνης χωρίς δεδομένα (non-data ink elements) και χωρίς αυτά. Έχοντας πάρα πολλά τέτοια στοιχεία ή πολύ λίγα, οδηγούμαστε σε κακή απεικόνιση. Θα πρέπει να βρούμε την κατάλληλη ισορροπία, ώστε στα σημεία των δεδομένων να δίνεται έμφαση, ενώ παράλληλα να υπάρχουν

αρκετά στοιχεία μελάνης χωρίς δεδομένα («πλαίσιο» -context) ώστε να φαίνεται και τι αυτά αντιπροσωπεύουν/σημαίνουν.

5 Τι Χρειάζεται να Γνωρίζουμε για τη Σωστή Επιλογή Λογισμικού

Καλύτερο λογισμικό για την απεικόνιση δεδομένων θεωρείται αυτό που μας επιτρέπει να φτιάξουμε τα διαγράμματα που χρειαζόμαστε. Αν αξίζει τον κόπο ή όχι η επένδυση στην εκμάθηση ενός νέου εργαλείου, θα φανεί μόνο αφού περάσει η περίοδος προσαρμογής.

5.1 Βασικές Αρχές

Κάποιες βασικές αρχές για να αξιολογήσουμε τα πλεονεκτήματα από τις διαφορετικές προσεγγίσεις για την παραγωγή των οπτικοποιήσεων είναι:

- **αν μπορούν να αναπαραχθούν** οι οπτικοποιήσεις μας (τα ευρήματά μας παραμένουν αμετάβλητα αν άλλοι επιστήμονες πραγματοποιήσουν την ίδια μελέτη); (Εικόνα 5.1-1 δεξιά)
- **πόσο εύκολο είναι να «εξερευνήσουμε» τα δεδομένα γρήγορα.**
εξερεύνηση δεδομένων (data exploration): η δοκιμή διαφορετικών τύπων απεικονίσεων, μετασχηματισμών δεδομένων, υποσυνόλων δεδομένων, ώστε να κατανοήσουμε τα βασικά χαρακτηριστικά και να δούμε τα πιο σημαντικά, αποτελεσματικά και γρήγορα, που υπό άλλες συνθήκες θα μπορούσαν να αγνοηθούν.
- **κατά πόσο η οπτική εμφάνιση της εξόδου που παράγεται μπορεί να τροποποιηθεί** (Εικόνα 5.1-2)

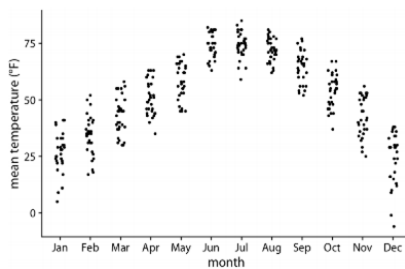


Figure 9-7. Mean daily temperatures in Lincoln, NE, visualized as strip charts. The points have been jittered along the x axis to better show the density of points at each temperature value. Data source: Weather Underground.

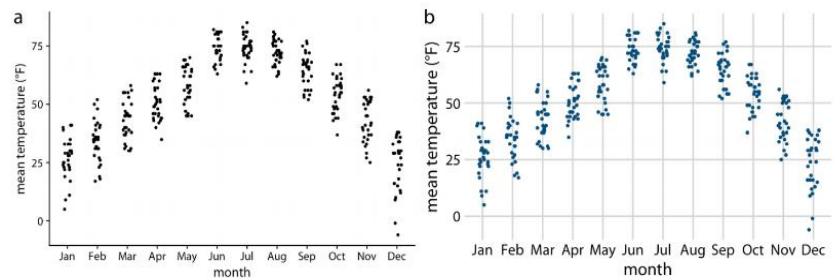


Figure 28-1. Repeat and reproduction of a figure. Part (a) is a repeat of Figure 9-7. The two figures are identical down to the random jitter that was applied to each point. By contrast, part (b) is a reproduction but not a repeat. In particular, the jitter in part (b) differs from the jitter in part (a) or in Figure 9-7. Data source: Weather Underground.

Εικόνα 5.1-1: [2] Αναπαραγωγή πανομοιότυπων οπτικοποιήσεων

Έχοντας κατανοήσει τα δεδομένα μας και γνωρίζοντας ποια πλευρά τους θέλουμε να δείξουμε στους αναγνώστες μας, πρέπει να ετοιμάσουμε το διάγραμμα σε υψηλή ποιότητα ώστε να είναι έτοιμο για οποιαδήποτε χρήση - εκτύπωση, παρουσίαση, ανάρτηση στο Διαδίκτυο κτλ (παρουσίαση δεδομένων -data presentation).

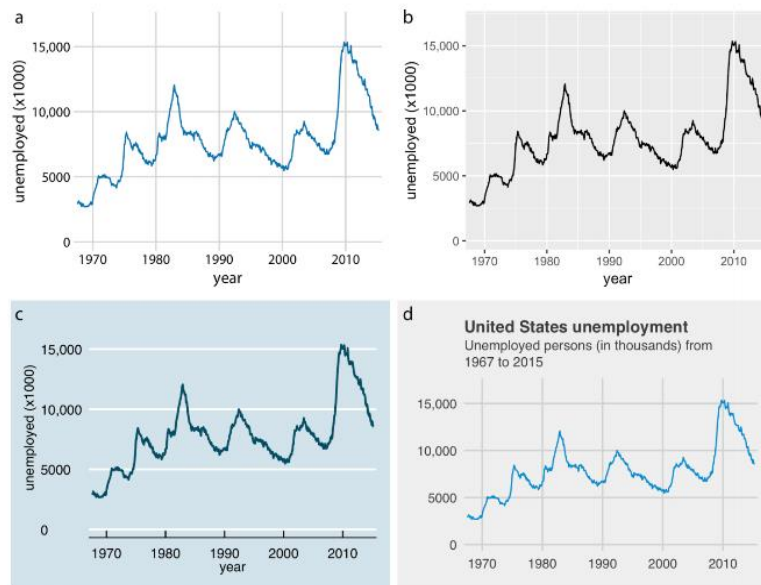


Figure 28-3. Number of unemployed persons in the US from 1970 to 2015. The same figure is displayed using four different ggplot2 themes: (a) the default theme for this book; (b) the default theme of ggplot2, the plotting software I have used to make all the figures in this book; (c) a theme that mimics visualizations shown in the Economist; (d) a theme that mimics visualizations shown by FiveThirtyEight. FiveThirtyEight often foregoes axis labels in favor of plot titles and subtitles, and therefore I have adjusted the figure accordingly. Data source: US Bureau of Labor Statistics.

Εικόνα 5.1-2: [2] Τροποποίηση της εμφάνισης της οπτικοποίησης κατά την έξοδο

Αυτό, μπορούμε να το κάνουμε είτε:

- χρησιμοποιώντας το ίδιο πρόγραμμα που χρησιμοποιήσαμε για την ολοκλήρωση της εξερεύνησης
- αλλάζοντας πλατφόρμα για να έχουμε καλύτερο έλεγχο του τελικού αποτελέσματος
- βλέποντας ένα προσχέδιο από το πρόγραμμά μας, και δημιουργώντας το από την αρχή με κάποιο πρόγραμμα χειρισμού εικόνας
- να το ξαναφτιάξουμε χειροκίνητα από την αρχή (αν και πρέπει να αποφεύγεται ειδικά για ρουτίνες ανάλυσης δεδομένων ή επιστημονικές δημοσιεύσεις, καθώς, σπάνια φτιάχνουμε ένα σχήμα μόνο μία φορά. Μπορεί για παράδειγμα να χρειαστεί να ξανακάνουμε το ίδιο πείραμα με λίγο διαφορετικές συνθήκες)

Ένα καλό λογισμικό οπτικοποίησης, θα πρέπει να κρατά ξεχωριστά το περιεχόμενο (το σύνολο δεδομένων, τους μετασχηματισμούς/μετατροπές, τις αντιστοιχίσεις, τα εύρη των αξόνων, τον τύπο του διαγράμματος) και το σχεδιασμό (χρώματα προσκηνίου-παρασκήνιου, μέγεθος γραμματοσειράς-οικογένεια γραμματοσειράς, σχήματα συμβόλων, πλέγμα, υπομνήματα, σημεία και τίτλοι του άξονα, τίτλος του διαγράμματος), ώστε να μπορούμε να τροποποιήσουμε τον σχεδιασμό ανεξάρτητα από το περιεχόμενο (Εικόνα 5.1-2). Κάτι αντίστοιχο δηλαδή, με την αρχή διαχωρισμού που ισχύει για τις εκδόσεις βιβλίων, όπου το περιεχόμενο δίνεται από συγγραφείς και ο σχεδιασμός από σχεδιαστές. Ένα τέτοιο λογισμικό που επιτυγχάνει τέτοιου είδους διαχωρισμό μέσω θεμάτων (themes) είναι για παράδειγμα το ggplot2.

Πολλές φορές, καταλήγουμε σε μια τροποποίηση της τελευταίας στιγμής, με αποτέλεσμα να χρειαστεί επανασχεδίαση όλων των σχημάτων. Τότε, κάποιες φορές λαμβάνεται η απόφαση, να μην επαναλάβουμε την ανάλυση από την αρχή ή να μην ξανασχεδιάσουμε τα σχήματα, είτε λόγω της προσπάθειας που απαιτείται, είτε επειδή τα άτομα που δημιούργησαν τα αρχικά διαγράμματα δεν είναι πλέον διαθέσιμα. Και στα δύο σενάρια, η μη αναπαραγωγικότητα (nonreproducibility) παρεμβαίνει στην παραγωγή σωστών και αποδοτικών αποτελεσμάτων. Όταν επιλέγουμε ένα πρόγραμμα απεικόνισης, πρέπει επίσης να λάβουμε υπόψη σε ποιο βαθμό μπορούμε να τροποποιήσουμε το σχέδιό μας.

Έτσι λοιπόν όταν θέλουμε να επιλέξουμε λογισμικό οπτικοποίησης θα πρέπει να έχουμε στο νου μας το πόσο εύκολα μπορεί να αναπαραχθεί η οπτικοποίηση, πόσο εύκολα μπορεί αυτή να ανανεωθεί όταν κάνουμε αλλαγές στο σύνολο δεδομένων, πόσο εύκολα μπορούμε να δοκιμάσουμε τύπο οπτικοποίησης για το ίδιο σύνολο δεδομένων και κατά πόσο μπορούμε να πειράζουμε την οπτικοποίηση χωριστά από το σύνολο των δεδομένων μας. Προγράμματα που δεν παρακολουθούν κάθε αλλαγή που κάνουμε θέλουν προσοχή, καθώς κάνουν δύσκολη την επανάληψη της οπτικοποίησης. Είναι καλύτερο να δημιουργούμε τυποποιημένα σχήματα γράφοντας κώδικα που τα παράγει από τα ακατέργαστα δεδομένα, ώστε να μπορούν εύκολα να επαναληφθούν από οποιονδήποτε έχει πρόσβαση στον κώδικα και τις βιβλιοθήκες που χρησιμοποιούνται (όπως Εικόνα 5.1-1). Εάν οι προγραμματιστικές μας ικανότητες δεν μας το επιτρέπουν, τότε όταν χρησιμοποιούμε διαφορετικά εργαλεία απεικόνισης ή κάνουμε τροποποιήσεις διαδραστικά, θα πρέπει να παρακολουθούμε όλα τα δεδομένα και τις οπτικές

τροποποιήσεις που εφαρμόστηκαν, κρατώντας προσεκτικές σημειώσεις, έτσι ώστε όλες οι εργασίες να μπορούν να αναπαραχθούν πάλι.

5.2 Οι Πιο Συνηθισμένοι Τύποι Εικόνας

(για στατικά γραφήματα):

- **PDF:** για αρχεία έτοιμα για δημοσίευση υψηλής ποιότητας
- **PNG:** για online παρουσίαση και
- **JPEG:** ως τελική λύση για οτιδήποτε άλλο (καθώς εξαρτάται από την ανάλυση).

Κάθε ένα έχει τα δικά του πλεονεκτήματα και μειονεκτήματα και η επιλογή του σωστού τύπου μπορεί να μας γλυτώσει από πολύ πονοκέφαλο.

Οι **Bitmap** εικόνες (JPEG, PNG): είναι πάντα οι ίδιες ανεξάρτητα από το πρόγραμμα ή τον υπολογιστή, αλλά επηρεάζεται από οποιαδήποτε αλλαγή μεγέθους. Υπάρχουν δύο κατηγορίες συμπίεσης (compression):

- **Χωρίς απώλειες (Lossless):** παραμένει όσο πιο πανομοιότυπο γίνεται με την αρχική εικόνα, είναι καλύτερη για μεγάλες περιοχές του ίδιου χρώματος. Τέτοια μορφή είναι το PNG
- **Με απώλειες (Lossy):** έχει ως αποτέλεσμα μικρότερου μεγέθους αρχεία αλλά επηρεάζει την ποιότητα. Καλύτερο για φωτογραφικές εικόνες χωρίς πολλές αιχμηρές άκρες όπως σχέδια με γραμμές και κείμενο. Τέτοια μορφή είναι το JPG.

Οι **Vector** εικόνες - **διανυσματικές** (PDF): πάντα δείχνουν το ίδιο ανεξάρτητα από την κλίμακα, αλλά μπορεί να υπάρχουν διαφορές στον τρόπο απόδοσής τους σε δύο διαφορετικά προγράμματα ή υπολογιστές. Λύση: να γίνει outlining ή να ενσωματωθούν όλες τις γραμματοσειρές στο αρχείο pdf.

Η αποθήκευση των οπτικοποιήσεων θα πρέπει να γίνεται σε μορφή που να διατηρεί τη μέγιστη ανάλυση, ακρίβεια και ευελιξία. Εάν χρειάζεται σε οποιοδήποτε άλλη μορφή, μπορούμε να το μετατρέψουμε οποιαδήποτε στιγμή στην πορεία από αυτή τη μορφή. Καλό είναι να αποφεύγουμε

τις συμπίεσεις με απώλειες (Lossy) ή εάν δεν μπορούμε να τις αποφύγουμε, να τις συμπίεσουμε όσο το δυνατόν λιγότερο.

5.3 Διαδραστικότητα

Οι απεικονίσεις που φτιάχνονται για τον ιστό (web browser visualizations) έχουν το πλεονέκτημα της διαδραστικότητας (όπως το ποντίκι να πηγαίνει πάνω από κάποιο σημείο-δεδομένο και να δείχνει περισσότερες πληροφορίες ή όταν γίνεται κλικ σε αυτό αντίστοιχα), καθώς επίσης, τη δυνατότητα συνεχούς ενημέρωσης τους όποτε κρίνεται ότι είναι απαραίτητο. Η JavaScript είναι πολύ χρήσιμη σε όλα αυτά, χρησιμοποιώντας ευέλικτες βιβλιοθήκες όπως η D3 για την οποιαδήποτε αλλαγή στη σελίδα που σχετίζεται με δεδομένα (<https://d3js.org/>) και το Leaflet για τη διαδραστική χαρτογράφηση (<https://leafletjs.com/>). Αναλυτικά κάποια εργαλεία θα τα δούμε αργότερα.

Στα πλεονεκτήματα των απεικονίσεων για το διαδίκτυο, εκτός από το ότι μπορούμε να παρουσιάζουμε διαφορετικά επίπεδα λεπτομέρειας για διαφορετικούς χρήστες, μπορούμε για τις περιπτώσεις στις οποίες θέλουμε να συμπεριλάβουμε πολλές οπτικές γωνίες των δεδομένων, να ανεβάσουμε τα δεδομένα μας και να αφήσουμε την επιλογή στον ίδιο το χρήστη να δει αυτό που τον ενδιαφέρει, πράγμα που μπορεί να οδηγήσει στην καλύτερη κατανόηση των δεδομένων. Μια ενδιαφέρουσα στροφή για την καλύτερη κατανόηση του τρόπου με τον οποίο λειτουργούν οι μέθοδοι μηχανικής μάθησης, είναι η methodviz [100].

Πράγματα που πρέπει να λάβουμε υπόψη:

- ανεξάρτητα από το πόσο περίπλοκη είναι η εργασία, πρέπει να εκτελείται πολύ γρήγορα στα προγράμματα περιήγησης
- διαφορετικά προγράμματα περιήγησης έχουν διαφορετικές δυνατότητες, επομένως χρειάζεται πάντα έλεγχος συμβατότητας (και ταχύτητας) των προγραμμάτων περιήγησης (θα πρέπει επίσης να λάβουμε υπόψιν ότι οι ενημερώσεις των προγραμμάτων περιήγησης δεν γίνονται από όλους)
- πρέπει να αποκρίνεται σωστά, ώστε να φαίνεται καλά από όλες τις συσκευές, κινητά, tablet κτλ (responsiveness)

- έλεγχος εάν επηρεάζεται ουσιαστικά από το εύρος ζώνης διαδικτύου (internet bandwidth)
- εκτός από δημόσια προβολή, μπορούν να δημιουργηθούν και για τοπική (local) μόνο προβολή. Για ιδιωτική προβολή / περιορισμένη πρόσβαση της πληροφορίας και θέματα εμπιστευτικότητας, θα πρέπει να χρησιμοποιηθεί κάποιος προστατευμένος διακομιστής (server).
- Επίσης, περαιτέρω προβλήματα εμπιστευτικότητας για client-side διαδικτυακές απεικονίσεις, καθώς
 - τα δεδομένα εφόσον μεταφορτώνονται στον web browser μπορούν, χωρίς ιδιαίτερα πολλές γνώσεις, να τα κατεβάσουν οι χρήστες,
 - η μεταφορά των δεδομένων σε διαφορετική νομική δικαιοδοσία που εξαρτάται από τη φυσική τοποθεσία του διακομιστή)
 - Τα αρχεία JavaScript πρέπει να είναι για όλες τις χρήσεις, εσωτερικά, χωρίς σχολιασμούς αναγνώσιμους από τον άνθρωπο και να «ασχημύνουν» (σκόπιμη δηλαδή μετατροπή τους σε δυσνόητο κώδικα ώστε να είναι ασαφές το νόημα του προγράμματος/κώδικα), έτσι ώστε να αποφευχθεί η εκμετάλλευση του κώδικα JavaScript από άλλους.

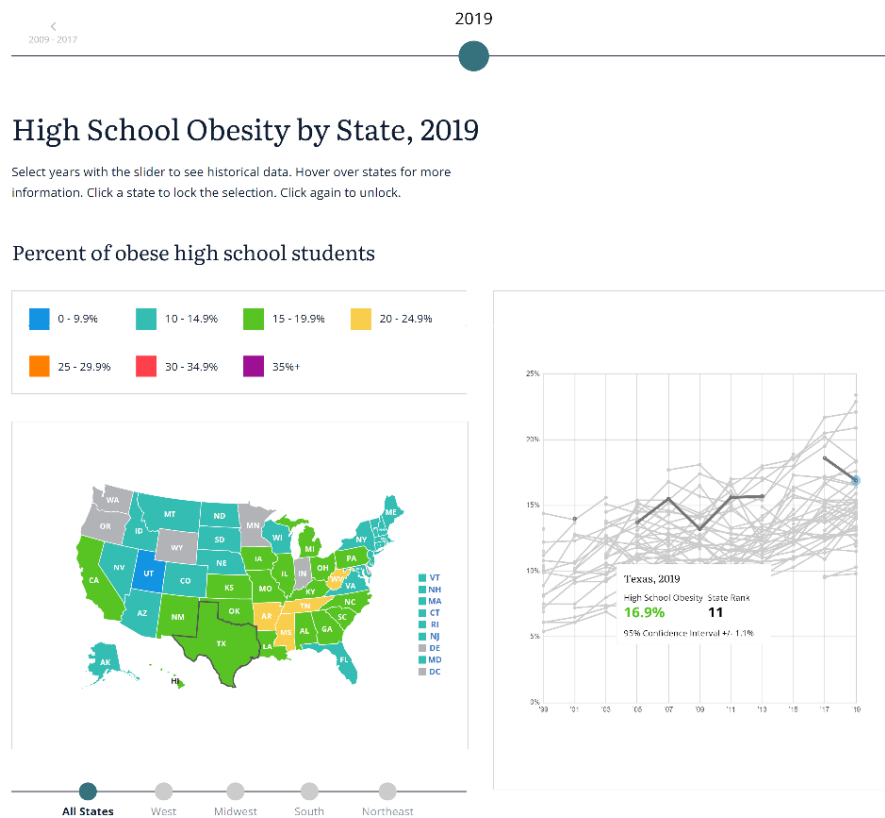
Συχνοί τρόποι αλληλεπίδρασης με μια διαδραστική απεικόνιση και τους πίνακές της είναι:

- τοποθέτηση το δείκτη του ποντικιού πάνω από το σημείο ενδιαφέροντος: για εμφάνιση επεξήγησης
- πάτημα του ποντικιού: για εμφάνιση πρόσθετων πληροφοριών, αντικατάσταση του περιεχομένου, εμβάθυνση της πληροφορίας “drilling down”, click & drag για μεγέθυνση, χρήση βελών για να προχωρήσουμε κ.λπ.
- τσίμπημα ή διεύρυνση των δακτύλων (Pinch or widen fingers): για μεγέθυνση σε οθόνες αφής
- κύλιση ή σύρσιμο προς τα κάτω (Scroll or swipe down): για μετάβαση στο επόμενο επίπεδο λεπτομερειών, "αφήγηση της ιστορίας"
- κύλιση ροδέλας: για μεγέθυνση σε χάρτη, ή περιστροφή ενός αντικειμένου όπως, μια εικονική σφαίρα
- sliders: για έλεγχο κάποιων όψεων της απεικόνισης
- εναλλαγή (toggle): για εμφάνιση/απόκρυψη κάποιων πτυχών

- εμφύχωση: όπως, εισαγωγή, εμφάνιση διαφορετικών αποτελεσμάτων κ.λπ.
- κινούμενο περιεχόμενο (animation) μέχρις ότου ο χρήστης να αλληλοεπιδράσει με αυτό

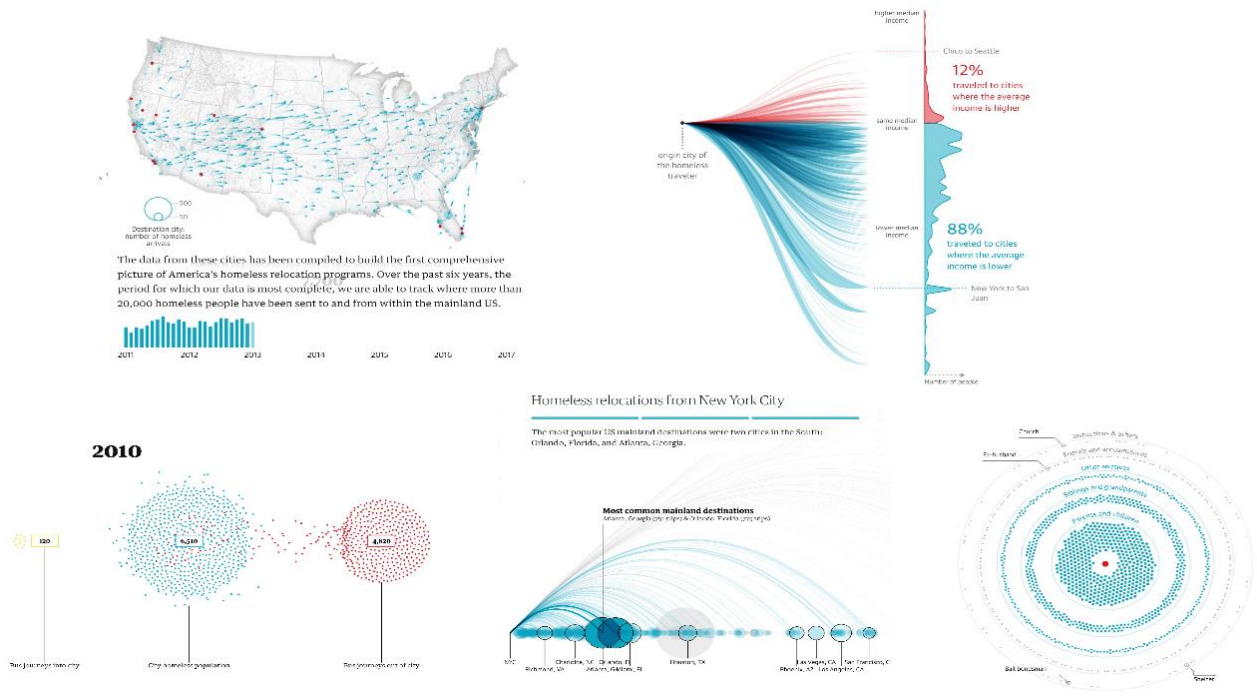
Τέτοια διαδραστικά παραδείγματα είναι:

- Ποσοστά παχυσαρκίας και δεδομένα τάσεων – Η κατάσταση της παιδικής παχυσαρκίας (<https://stateofchildhoodobesity.org/data/>): χρησιμοποίηση πολλών τύπων διαδραστικότητας (Εικόνα 5.3-3)



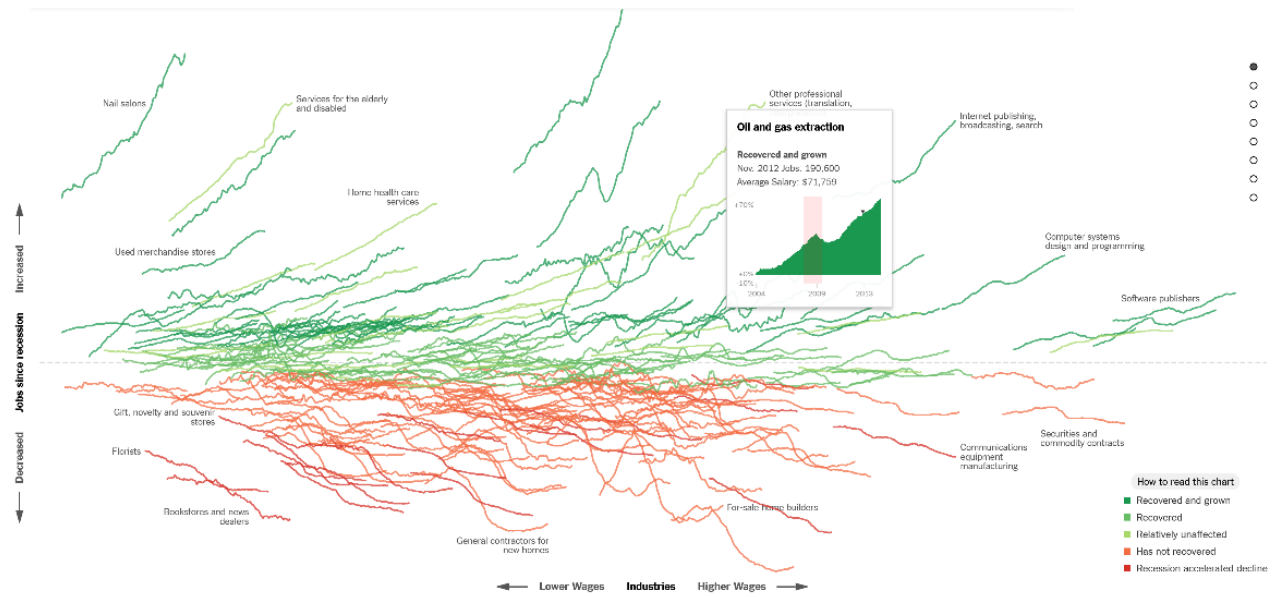
Εικόνα 5.3-3: [7] Παράδειγμα διαδραστικότητας πολλών τύπων

- Bussed out: πώς η Αμερική μετακινεί χιλιάδες άστεγους σε όλη τη χώρα | US news | The Guardian (<https://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study>): διαδραστικό, κινούμενο, εξαιρετικό! (Εικόνα 5.3-4)



Εικόνα 5.3-4: [8] Παράδειγμα διαδραστικότητας, κινούμενο

- Πως η ύφεση αναδιαμόρφωσε την οικονομία, σε 255 διαγράμματα (<https://www.nytimes.com/interactive/2014/06/05/upshot/how-the-recession-reshaped-the-economy-in-255-charts.html>): ένα καινοτόμο παράδειγμα για το πόσο καλά μπορεί να χρησιμοποιηθεί η διαδραστικότητα για να «στρώσει» τεράστιες ποσότητες δεδομένων! (Εικόνα 5.3-5)



Εικόνα 5.3-5: [9] Παράδειγμα διαδραστικότητας μεγάλων δεδομένων (big data)

6 Big Data

Η διαφορά των μεγάλων δεδομένων (big data) από τα υπόλοιπα σύνολα δεδομένων που είδαμε, έγκειται στο γεγονός είτε ότι αυτά είναι πάρα πολλά, είτε ότι είναι πολύ γρήγορα για να γίνει επεξεργασία σε πραγματικό χρόνο (όγκος-ταχύτητα, volume-velocity). Για να λειτουργήσει καλά μια τέτοιου είδους ανάλυση, πρέπει να είμαστε σίγουροι ότι τα μοτίβα (patterns) και τα σφάλματα εντοπίζονται και γίνονται κατανοητά. Η αβεβαιότητα ή οι περιοδικές επιδράσεις (periodical effects) είναι σημαντικές επίσης.

Ένα καλό παράδειγμα τέτοιου συνόλου δεδομένων που μπορούμε να κατεβάσουμε (το οποίο δεν είναι αρκετά μεγάλο για να ζορίσει έναν πολύ καλό υπολογιστή, αλλά είναι αρκετά μεγάλο για να πειραματιστούμε), με συνεχείς, κατηγορηματικές, γεωγραφικές και μεταβλητές χρονικών σειρών, είναι το FOILing NYC's Taxi Trip Data, (https://chriswhong.com/open-data/foil_nyc_taxi/) του Chris Whong (A Day in the Life (<http://chriswhong.github.io/nyctaxi/>)).

Δεδομένου ότι δεν μπορούμε να τα χρησιμοποιήσουμε όπως είναι σε μια σχεδίαση/ένα διάγραμμα καθώς όπως είπαμε είναι πολύ μεγάλα, θα πρέπει να χρησιμοποιήσουμε αντί αυτού, στατιστικά. Ο τρόπος υπολογισμού τους, είναι αυτή τη στιγμή ένα σημαντικό θέμα καθώς, νέο λογισμικό και επιλογές υπολογιστικού νέφους (cloud computing options) βγαίνουν συνεχώς.

Ο Robert Grant συνοψίζει τις γενικές επιλογές για οπτικοποίηση των στατιστικών στις παρακάτω (αναλυτικότερες πληροφορίες στο κεφάλαιό του “big data”):

- Αρχικά αποφασίζουμε ποιες μεταβλητές θέλουμε να δείξουμε και στη συνέχεια ποιες κατηγορίες αυτών. Έπειτα μετά μετράμε πόσες παρατηρήσεις πέφτουν σε αυτές τις κατηγορίες (binning). Έτσι εργαζόμαστε πλέον με τους συνδυασμούς των κατηγοριών και όχι με την κάθε μία ξεχωριστή παρατήρηση.
- Έχοντας κάποια στατιστική ειδικευση, βρίσκουμε βασικά/θεμελιώδη βήματα για τους υπολογισμούς μας τα οποία μπορούν να προστεθούν σε άλλα τμήματα δεδομένων για να μας δώσουν αποτελέσματα. Παίρνοντας αυτά τα αποτελέσματα, τα προσθέτουμε μαζί με των υπολοίπων. Έπειτα, υπολογίζουμε τα αποτελέσματα αυτών των βασικών βημάτων και όχι των μεμονωμένων τιμών.
- Τέλος παίρνουμε ένα τυχαίο δείγμα από τα δεδομένα μας και κάνουμε μια κανονική ανάλυση των δεδομένων σε αυτό.

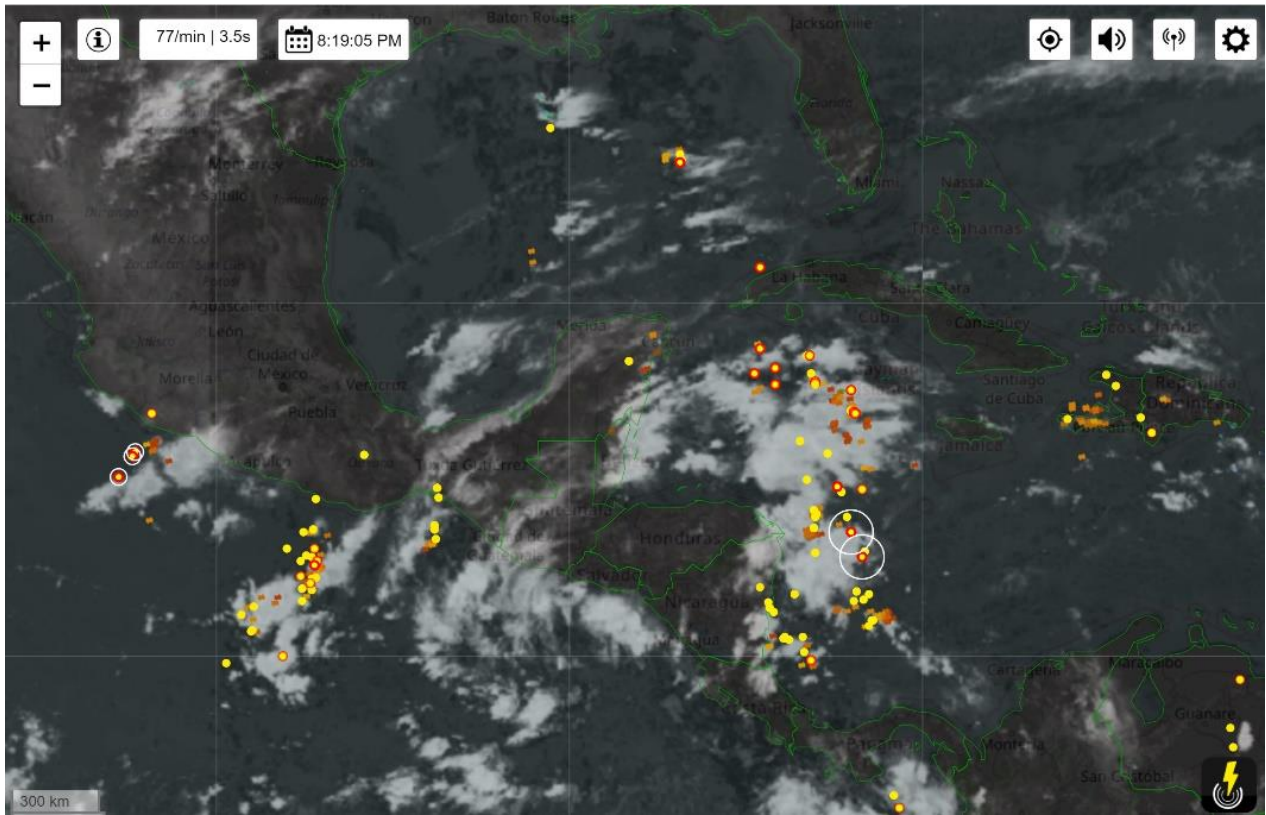
Υπάρχει η δυνατότητα επίσης της ομαλοποίησης είτε των βασικών/θεμελιωδών βημάτων, είτε των στατιστικών.

Τα τυχαία δείγματα που αναφέραμε είναι καλά όταν χρειαζόμαστε γρήγορες οπτικοποιήσεις για να βρούμε σφάλματα στα δεδομένα μας και μόνο για δική μας χρήση. Αν θέλουμε να βρούμε λάθη ή σπάνια μοτίβα, τότε πρέπει να γίνει επεξεργασία όλων των δεδομένων αλλιώς κάποια μπορεί να χαθούν. Μια εναλλακτική είναι, να δείξουμε τα αποτελέσματα της ανάλυσης και να δείξουμε και ένα τυχαίο δείγμα των δεδομένων, αρκεί να φαίνεται ότι είναι δείγμα.

Πολλοί οργανισμοί λειτουργούν βάσει ζωντανών ροών δεδομένων/πληροφοριών (live big data). Σε αυτά, η απεικόνιση συνήθως εμφανίζει ένα κινούμενο μέσο όρο. Το αποτέλεσμα μπορεί να παρέχεται από ένα φυλλομετρητή (web browser), αλλά θα πρέπει, η καθυστέρηση μεταξύ του γεγονότος που συμβαίνει και της καθυστέρησης στην εμφάνιση της σελίδας (latency) να μειωθεί όσο το δυνατό περισσότερο ώστε να γίνεται η ανανέωσή της όσο το δυνατόν πιο κοντά στον πραγματικό χρόνο.

Παραδείγματα ιστοτόπων που ασχολούνται με ζωντανά δεδομένα σε μεγάλη κλίμακα είναι:

- Χάρτης Αστραπών σε πραγματικό χρόνο (<http://www.lightningmaps.org/>) (Εικόνα



Εικόνα 6-1: [10] Παράδειγμα ζωντανών ροών δεδομένων - Χάρτης Αστραπών (live big data)

6-1Εικόνα 5.3-3)

- Χάρτης παρακολούθησης πτήσεων σε πραγματικό χρόνο (<http://www.flightradar24.com/>) (Εικόνα 6-2)



Εικόνα 6-2: [11] Παράδειγμα ζωντανών ροών δεδομένων - Χάρτης παρακολούθησης πτήσεων (live big data)

7 Χρώμα

Οι τρεις θεμελιώδεις/βασικές χρήσεις του χρώματος στην οπτικοποίηση των δεδομένων είναι:

1. να διαφοροποιεί τα δεδομένα μεταξύ τους
2. να αντιπροσωπεύει τιμές δεδομένων
3. να τονίζει

Μια πιο αναλυτική περιγραφή για τις τρεις αυτές αρχές είναι:

7.1 Για να Διαφοροποιεί τα Δεδομένα Μεταξύ τους

Για διακριτικά αντικείμενα ή ομάδες που δεν έχουν μια συγκεκριμένη σειρά (όπως, διαφορετικές χώρες/πόλεις σε χάρτη).

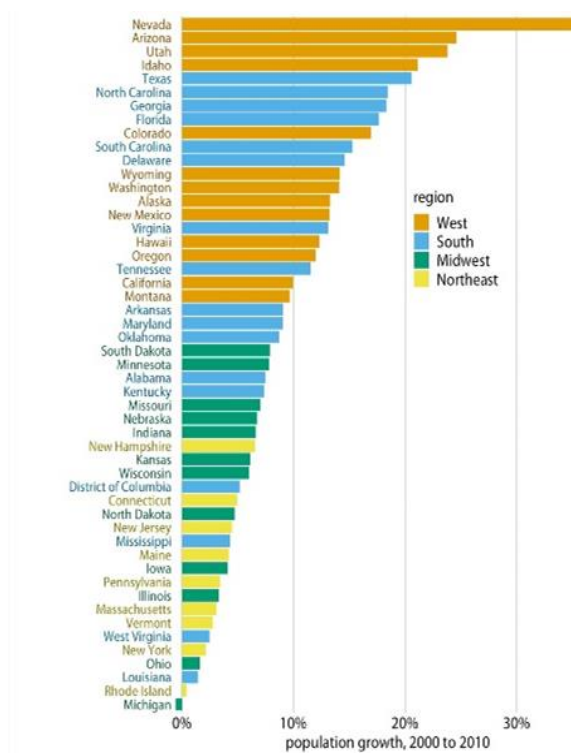


Figure 4-2. Population growth in the US from 2000 to 2010. States in the West and South have seen the largest increases, whereas states in the Midwest and Northeast have seen much smaller increases (or even, in the case of Michigan, a decrease). Data source: US Census Bureau.



Figure 4-1. Example qualitative color scales. The Okabe Ito scale is the default scale used throughout this book [Okabe and Ito 2008]. The ColorBrewer Dark2 scale is provided by the ColorBrewer project [Brewer 2017]. The ggplot2 hue scale is the default qualitative scale in the widely used plotting software ggplot2.

Εικόνα 7.1-1: [2] Χρήση χρώματος για τη διαφοροποίηση των δεδομένων μεταξύ τους

Τα χρώματα σε αυτήν την περίπτωση (ποιοτική κλίμακα χρωμάτων- qualitative color scale), θα πρέπει να ακολουθούν τους παρακάτω κανόνες:

- να είναι πεπερασμένα σύνολα ευδιάκριτων, αλλά ισοδύναμων χρωμάτων
- κανένα χρώμα να μην επικρατεί
- να μην δίνουν την εντύπωση μιας ακολουθίας

Στην Εικόνα 7.1-1αριστερά, βλέπουμε τη χρήση τέτοιου είδους χρωμάτων, όπου οι πολιτείες χρωματίζονται ανά γεωγραφική περιοχή. Ο χρωματισμός εδώ έχει την έννοια ότι πολιτείες της ίδιας περιοχής έχουν παρόμοια αύξηση πληθυσμού.

Υπάρχουν πολλές κατάλληλες χρωματικά ποιοτικές κλίμακες διαθέσιμες στο διαδίκτυο, όπως το ColorBrewer (Εικόνα 7.1-1δεξιά) όπου συμπεριλαμβάνονται και αρκετά ανοιχτά και αρκετά σκούρα χρώματα (<https://colorbrewer2.org/>)

7.2 Για να Αντιπροσωπεύει Τιμές Δεδομένων

Για ποσοτικά δεδομένα που έχουν διαδοχική κλίμακα χρώματος (όπως θερμοκρασία ή ταχύτητα).

Τα χρώματα σε αυτήν την περίπτωση (διαδοχική κλίμακα χρωμάτων- sequential color scale), θα πρέπει να ακολουθούν τους παρακάτω κανόνες:

- η ακολουθία χρωμάτων θα πρέπει να δείχνει ξεκάθαρα ποιες τιμές από αυτές είναι μεγαλύτερες και ποιες μικρότερες και πόσο μακρινές είναι η μία από την άλλη
- η ακολουθία πρέπει να διαφέρει ομοιόμορφα σε όλο το εύρος της

Οι διαδοχικές κλίμακες χρωμάτων μπορούν να είναι βασισμένες μόνο σε μία απόχρωση (single hue), όπως για παράδειγμα από σκούρο μπλε σε ανοιχτό μπλε, ή σε πολλές αποχρώσεις (multihue), όπως από σκούρο κόκκινο σε ανοιχτό κίτρινο (Εικόνα 7.2-3 αριστερά). Τα multihue, έχουν την τάση να μιμούνται τις ακολουθίες της φύσης (π.χ. μπλε σε ανοιχτό κίτρινο. Το αντίθετο φαίνεται αφύσικο, για παράδειγμα σκούρο κίτρινο σε ανοιχτό μπλε).



Figure 4-3. Example sequential color scales. The ColorBrewer Blues scale is a monochromatic scale that varies from dark to light blue. The Heat and Viridis scales are multihue scales that vary from dark red to light yellow and from dark blue via green to light yellow, respectively.

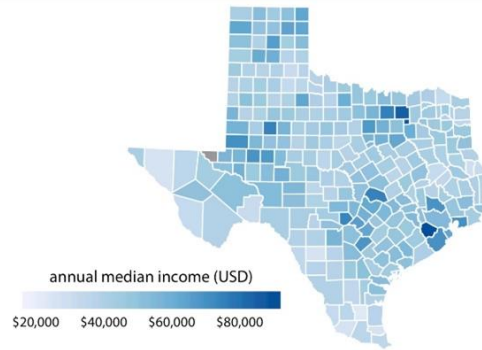


Figure 4-4. Median annual income in Texas counties. The highest median incomes are seen in major Texas metropolitan areas, in particular near Houston and Dallas. No median income estimate is available for Loving County in West Texas, and therefore that county is shown in gray. Data source: 2015 Five-Year American Community Survey.

Εικόνα 7.2-3: [2] Χρήση χρώματος για αντιπροσώπευση τιμών δεδομένων



Figure 4-5. Example diverging color scales. Diverging scales can be thought of as two sequential scales stitched together at a common midpoint color. Common color choices for diverging scales include brown to greenish blue, pink to yellow-green, and blue to red.

Εικόνα 7.2-2: [2] Χρήση χρώματος για αντιπροσώπευση τιμών δεδομένων

Οι διαδοχικές κλίμακες είναι χρήσιμες για:

- Choropleths (είναι χάρτες γεωγραφικών περιοχών, στους οποίους φαίνονται οι διακυμάνσεις των δεδομένων, χρωματίζοντας τις τιμές δεδομένων πάνω σε αυτό (Εικόνα 7.2-3 δεξιά)
- να δείχνουμε την απόκλιση των τιμών των δεδομένων προς μία από τις δύο κατευθύνσεις σε σχέση με ένα ουδέτερο μεσαίο σημείο, όπως επίσης και το πόσο μακριά αποκλίνει προς

οποιαδήποτε κατεύθυνση από αυτό το ουδέτερο σημείο (πχ, θετικοί-αρνητικοί αριθμοί). Αποκλίνουσα κλίμακα χρώματος- diverging color scale: είναι σαν δύο διαδοχικές κλίμακες ενωμένες σε ένα κοινό - συνήθως ανοιχτόχρωμο - σημείο στο κέντρο, και πιο σκοτεινό προς τα έξω και στις δύο κατευθύνσεις (Εικόνα 7.2-2).

Παράδειγμα με αποκλίνουσα χρήση χρώματος: Στην Εικόνα 7.2-4 πάνω αριστερά μπορούμε να δούμε τις κομητείες (με τα ποσοστά πλειοψηφίας ή μειονότητες πάνω ή κάτω από 50% λευκών και μη λευκών), και σε ποιες περιοχές βρίσκονται, σε περίπου ίσες αναλογίες.

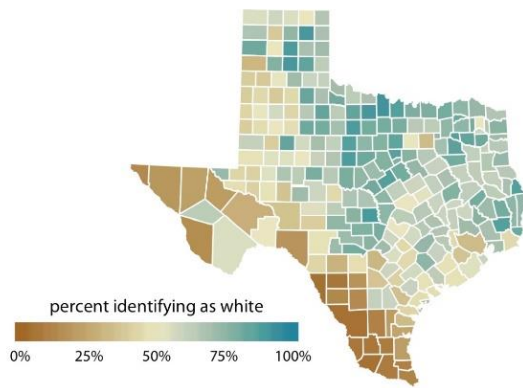


Figure 4-6. Percentage of people identifying as white in Texas counties. Whites are in the majority in North and East Texas but not in South or West Texas. Data source: 2010 US Decennial Census.

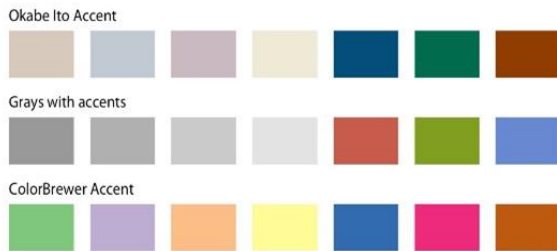


Figure 4-7. Example accent color scales, each with four base colors and three accent colors. Accent color scales can be derived in several different ways: (top) we can take an existing color scale (e.g., the Okabe Ito scale, Figure 4-1) and lighten and/or partially desaturate some colors while darkening others; (middle) we can take gray values and pair them with colors; (bottom) we can use an existing accent color scale (e.g., the one from the ColorBrewer project).

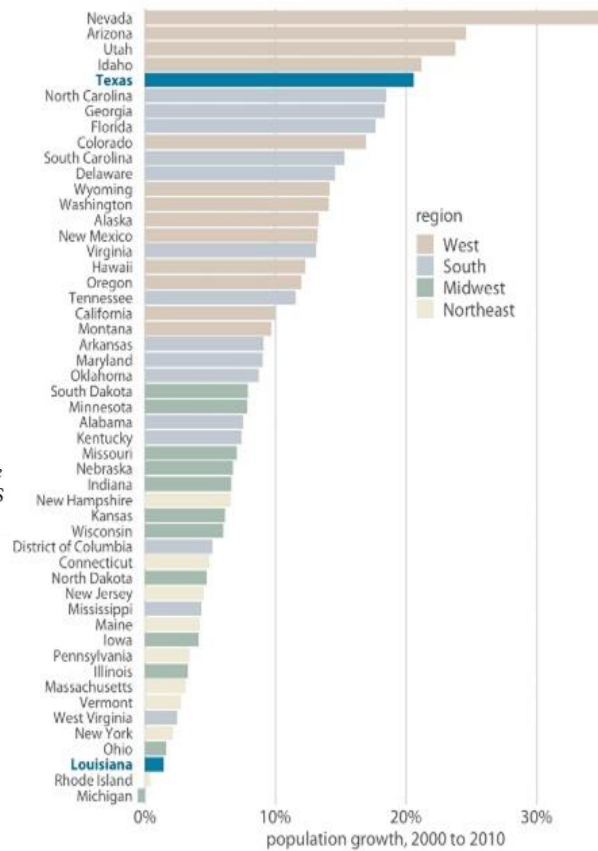


Figure 4-8. From 2000 to 2010, the two neighboring southern states, Texas and Louisiana, experienced among the highest and lowest population growth across the US. Data source: US Census Bureau.

Εικόνα 7.2-4: [2] Παράδειγμα αποκλίνουσας κλίμακας χρώματος (πάνω αριστερά) και, χρήση χρώματος για τονισμό

7.3 Για να Τονίζει

Για να τονίζουμε συγκεκριμένα στοιχεία των δεδομένων που μεταφέρουν σημαντικές πληροφορίες, ώστε να ενισχύσουμε την ιστορία μας δίνοντας έμφαση σε αυτά, χρησιμοποιώντας έντονα/ζωηρά χρώματα ώστε να ξεχωρίζουν από τα υπόλοιπα χρώματα του σχήματος (Εικόνα 7.2-4).

Τα χρώματα σε αυτήν την περίπτωση (χρωματικές κλίμακες τονισμού- accent color scales), θα πρέπει να ακολουθούν τον κανόνα:

- τα βασικά χρώματα δεν πρέπει να ανταγωνίζονται για την προσοχή μας με τα χρώματα τονισμού.
- θα πρέπει να περιέχει κάποια υποτονικά χρώματα καθώς και ισχυρότερα, πιο έντονα, σκούρα και/η πιο κορεσμένα χρώματα

Παράδειγμα: Στην Εικόνα 7.2-4 δεξιά, το μήνυμα ότι οι δύο αυτές γειτονικές πολιτείες έχουν πολύ διαφορετική ταχύτητα ανάπτυξης φαίνεται σαφώς από μια τέτοια χρήση χρωμάτων

Ένας εύκολος τρόπος για να το εφαρμόσουμε είναι: αφαιρώντας όλα τα άλλα χρώματα και αφήνοντας μόνο το χρώμα τονισμού (Εικόνα 7.3-5)

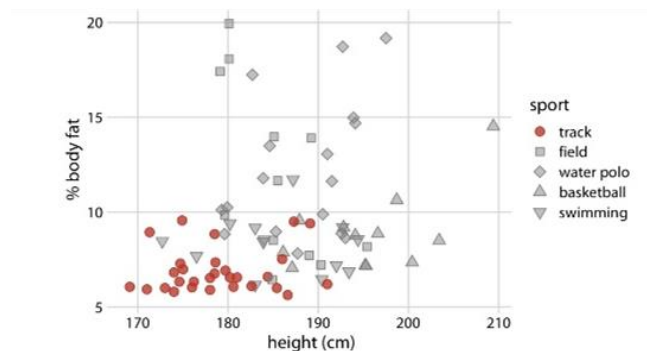


Figure 4-9. Track athletes are among the shortest and leanest of male professional athletes participating in popular sports. Data source: [Telford and Cunningham 1991].

Εικόνα 7.3-5: [2] Παράδειγμα χρήσης ενός χρώματος για τονισμό

Μπορούμε να χρησιμοποιήσουμε το χρώμα για την ποσοτικοποίηση των κωδικοποιήσεών μας, είτε ως Κόκκινο, Πράσινο, Μπλε (RGB), είτε ως Χροιά (Hue) για βασικό χρώμα, είτε ως Κορεσμό (Saturation) και Τιμή (Value) για φωτεινότητα και σκοτεινότητα (HSV).

Ο ανθρώπινος εγκέφαλος δεν μπορεί να δει αποστάσεις μεταξύ των χρωμάτων, αλλά το να είναι κάτι πιο ανοιχτό ή πιο σκοτεινό με μια ματιά, μπορεί μερικές φορές να είναι το μόνο που χρειαζόμαστε.

7.4 Κοινά Χρωματικά Λάθη

Παρόλο που το χρώμα μπορεί να είναι ένα εξαιρετικό εργαλείο για την ενίσχυση των οπτικοποιήσεών μας, η κακή επιλογή χρωμάτων μπορεί να τις καταστρέψει.

Ας δούμε μερικά τέτοια λάθη:

7.4.1 Κωδικοποίηση Πολλών Διαφορετικών Στοιχείων με Διαφορετικά Χρώματα

δίνοντας έτσι στο χρώμα μια δουλειά, η οποία είναι πολύ μεγάλη για να μπορέσει να τη χειριστεί.

Πρόβλημα: ο ανθρώπινος εγκέφαλος χρειάζεται πολλή προσπάθεια για να προσπαθήσει να κάνει το ταίριασμα των χρωμάτων, και πολλές φορές κάποια από αυτά είναι αρκετά παραπλήσια, με αποτέλεσμα να δημιουργούνται και άλλα προβλήματα.

ΚΑΝΟΝΑΣ: το χρώμα πρέπει να ενισχύει το γράφημα, διευκολύνοντας έτσι την ανάγνωσή του και όχι να την παρεμποδίζει προσθέτοντας οπτικό θόρυβο.

Λύση: Μια ασφαλής επιλογή είναι η χρήση τριών έως πέντε διαφορετικών χρωμάτων για την οπτικοποίηση μας.

Στην Εικόνα 7.4.1-6, βλέπουμε τη χρήση διαφορετικών χρωμάτων για κάθε πολιτεία των ΗΠΑ σε μια προσπάθεια αναγνώρισης καθεμίας από αυτές, που όμως δημιουργεί παρά λύνει προβλήματα.

Λύση: να ομαδοποιηθούν σε χρωματικές περιοχές και να προσδιοριστούν μεμονωμένες πολιτείες ενδιαφέροντος (ή ένα αντιπροσωπευτικό υποσύνολό τους που θέλουμε να ξεχωρίσει). Αυτό μπορεί να γίνει με τη χρήση άμεσης επισήμανσης κειμένου (direct text labeling) για περισσότερα από οκτώ κατηγορικά στοιχεία (categorical items) (Εικόνα 7.4.1-6 δεξιά). Μια άλλη λύση θα ήταν η παροχή των συγκεκριμένων πληροφοριών με τη μορφή πίνακα.

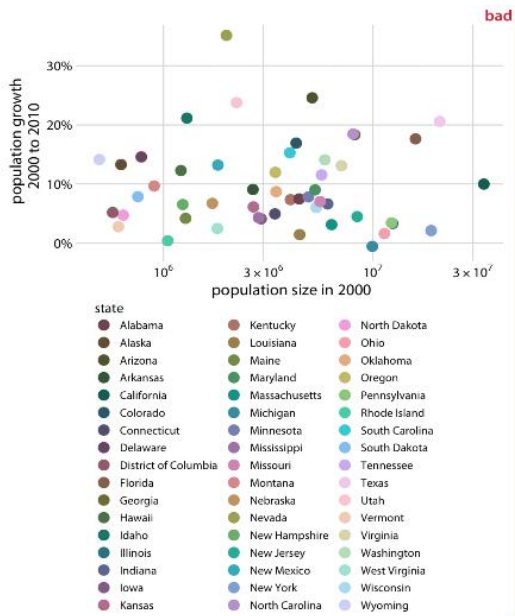


Figure 19-1. Population growth from 2000 to 2010 versus population size in 2000, for all 50 US states and the District of Columbia. Every state is marked in a different color. Because there are so many states, it is very difficult to match the colors in the legend to the dots in the scatterplot. Data source: US Census Bureau.

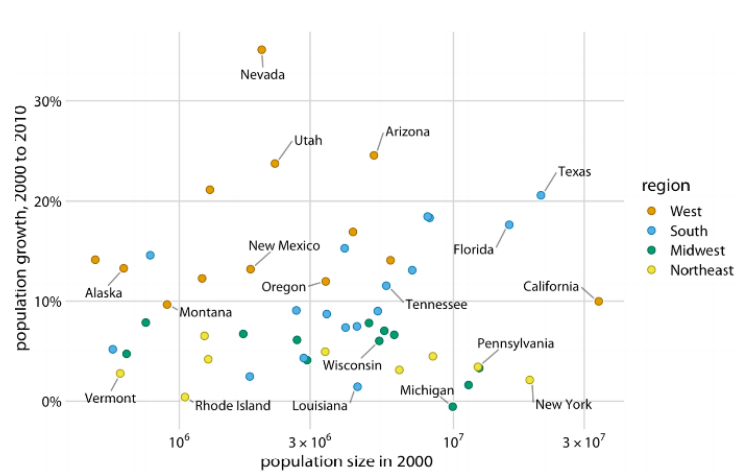


Figure 19-2. Population growth from 2000 to 2010 versus population size in 2000. In contrast to Figure 19-1, I have now colored states by region and have directly labeled a subset of states. The majority of states have been left unlabeled to prevent overcrowding in the figure. Data source: US Census Bureau.

Εικόνα 7.4.1-6: [2] Κωδικοποίηση διαφορετικών στοιχείων με διαφορετικά χρώματα

Στην Εικόνα 7.4.1-7 αριστερά, βλέπουμε μια προσπάθεια χρωματισμού του διαγράμματος, που σκοπό έχει να δείξει την αύξηση του πληθυσμού στις ΗΠΑ. Παρόλο που τα χρώματα φαίνονται οπτικά ελκυστικά, είναι χωρίς σαφή σκοπό, δεν δίνουν καμία εικόνα για τα δεδομένα, είναι τόσο κορεσμένα και έντονα που δυσκολεύει πολύ την ανάγνωση των ονομάτων χωρίς να τραβιέται η προσοχή μας στην έγχρωμη δεξιά περιοχή του διαγράμματος. Επίσης είναι δύσκολη η σύγκριση των τελικών σημείων (endpoints) των ράβδων με τις υποκείμενες διαγραμμίσεις. Τέτοιου είδους χρήση χρώματος πρέπει να αποφεύγεται.

Προσπαθώντας να χρησιμοποιήσουμε μη μονοτονικές κλίμακες χρώματος για τιμές δεδομένων: Όπως είδαμε ήδη, τα χρώματα πρέπει να υποδεικνύουν σαφώς τις μεγαλύτερες και/ή μικρότερες τιμές καθώς επίσης και να απεικονίζουν τις αντίστοιχες διαφορές (ή αποστάσεις) μεταξύ των τιμών δεδομένων. Έχοντας χρώματα που το παραβιάζουν αυτό, όπως είναι η κλίμακα του ουράνιου τόξου (Εικόνα 7.4.1-7 δεξιά) προκαλούμε σύγχυση στους αναγνώστες καθώς:

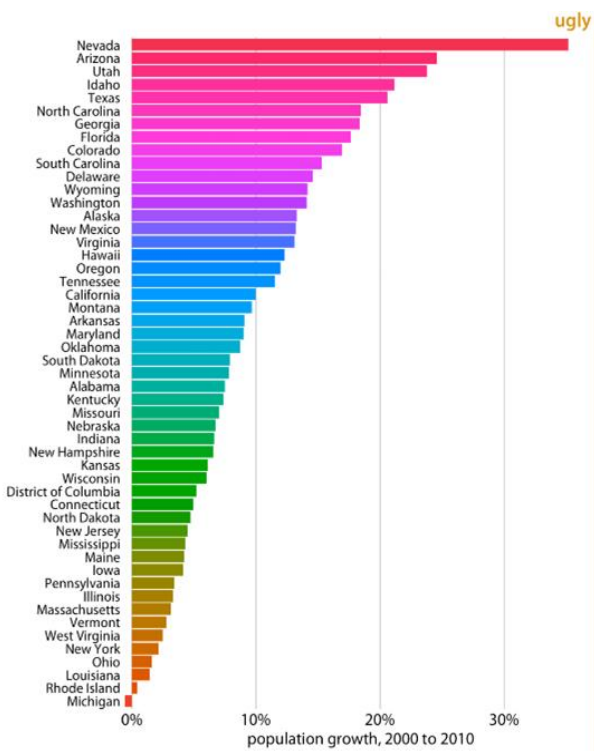


Figure 19-3. Population growth in the US from 2000 to 2010. The rainbow coloring of states serves no purpose and is distracting. Furthermore, the colors are overly saturated. Data source: US Census Bureau.



Figure 19-4. The rainbow color scale is highly nonmonotonic. This becomes apparent when the colors are converted to gray values. From left to right, the scale goes from moderately dark to light to very dark and back to moderately dark. In addition, the changes in lightness are nonuniform. The lightest part of the scale (corresponding to the colors yellow, light green, and cyan) takes up almost a third of the entire scale, while the darkest part (corresponding to dark blue) is concentrated in a narrow region of the scale.

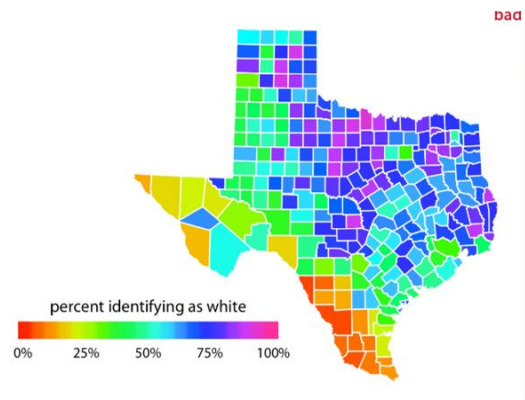


Figure 19-5. Percentage of people identifying as white in Texas counties. The rainbow color scale is not an appropriate scale to visualize continuous data values, because it tends to place emphasis on arbitrary features of the data. Here, it emphasizes counties in which approximately 75% of the population identify as white. Data source: 2010 US Decennial Census.

Εικόνα 7.4.1-7: [2] Παραδείγματα μη μονοτονικών κλιμάκων χρώματος

- το να έχουμε χρώματα στην αρχή και στο τέλος τα οποία είναι παραπλήσια μεταξύ τους, δεν γίνονται αντιληπτά ως τιμές με μέγιστη απόσταση (Εικόνα 7.4.1-7 δεξιά), ή
- το να έχουμε μια υψηλή μη μονοτονική κλίμακα, με αποτέλεσμα τα χρώματα να ποικίλλουν στις εναλλαγές τους (άλλα με πολύ γρήγορο και άλλα πολύ αργό ρυθμό), έχει ως αποτέλεσμα να τείνουν να αποκρύψουν χαρακτηριστικά ή/και να επισημάνουν χαρακτηριστικά των δεδομένων. Η έλλειψη μονοτονικότητας στην κλίμακα γίνεται εύκολα κατανοητή εάν τη μετατρέψουμε σε τιμές στην κλίμακα του γκρι (Εικόνα 7.4.1-7 πάνω δεξιά).

- το να είναι πολύ κορεσμένο (saturated), με αποτέλεσμα, όταν το κοιτάζουμε για μεγάλο χρονικό διάστημα, να γίνεται άβολο/κουραστικό για τα μάτια.

Μη λαμβάνοντας υπόψη τους αναγνώστες με χρωματικές ανεπάρκειες στην όραση (color-vision deficiencies, CVD): (για παράδειγμα, αναγνώστες που δεν μπορούν να διακρίνουν το κόκκινο-πράσινο ή μπλε-κίτρινο)

Από τις τρεις χρωματικές κλίμακες που έχουμε δει (διαδοχικές, αποκλίνουσες, ποιοτικές-sequential, diverging, qualitative), οι διαδοχικές κλίμακες δεν δημιουργούν πρόβλημα στα άτομα με CVD εφόσον το χρώμα παρουσιάζεται ως συνεχής κλίση από σκοτεινό σε ανοιχτό (Εικόνα 7.4.1-8 πάνω).

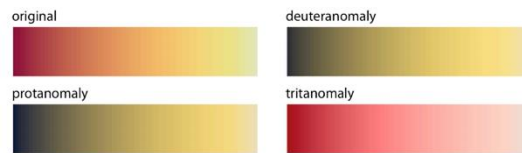


Figure 19-6. Color-vision deficiency simulation of the sequential color scale Heat, which runs from dark red to light yellow. From left to right and top to bottom, we see the original scale and the scale as seen under deuteranomaly, protanomaly, and tritanomaly simulations. Even though the specific colors look different under the three types of CVD, in each case we can see a clear gradient from dark to light. Therefore, this color scale is safe to use for viewers with CVD.

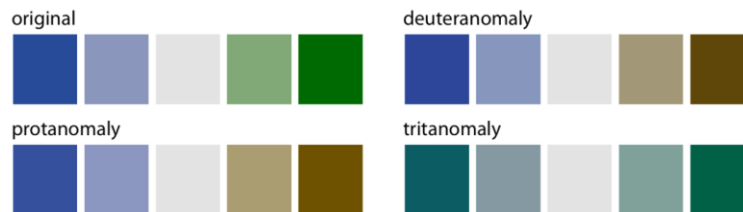


Figure 19-8. A blue-green contrast becomes indistinguishable under blue-yellow CVD (tritanomaly).



Figure 19-10. Qualitative color palette for all color-vision deficiencies [Okabe and Ito 2008]. The alphanumeric codes represent the colors in RGB space, encoded as hexadecimals. In many plot libraries and image manipulation programs, you can just enter these codes directly. If your software does not take hexadecimals directly, you can also use the values in Table 19-1.

Εικόνα 7.4.1-8: [2] Χρωματικές ανεπάρκειες στην όραση (CVD)

Οι αποκλίνουσες κλίμακες, ωστόσο, μπορεί να είναι δυσδιάκριτες σε αυτούς, όμως με μικρές τροποποιήσεις, μπορούν να μετατραπούν σε ασφαλή χρώματα και με χαρακτήρα. Στην Εικόνα 7.4.1-8 (κέντρο) μπορούμε να δούμε ένα παράδειγμα τέτοιων προβληματικών χρωμάτων: το μπλε

και το πράσινο από τα αρχικά χρώματα, μοιάζουν σχεδόν το ίδιο σε άτομα με τριτανομαλία-tritanomaly).

Οι πιο περίπλοκες, είναι οι ποιοτικές κλίμακες, καθώς όλα τα χρώματα που χρησιμοποιούνται πρέπει να ξεχωρίζουν μεταξύ τους υπό όλες τις μορφές του CVD. Σε αυτή την πρόκληση απαντούν τα χρώματα της Εικόνα 7.4.1-8 (κάτω) που λειτουργούν σχεδόν για οποιοδήποτε σενάριο (αφού όπως είπαμε δεν θα πρέπει να γίνεται χρήση περισσότερων των οκτώ διαφορετικών στοιχείων με χρωματική κωδικοποίηση σε μια απεικόνιση ούτως ή άλλως).

Υπάρχουν πολλές χρωματικές κλίμακες ασφαλείς για άτομα με CVD στο διαδίκτυο, αλλά η χρήση μιας τέτοιας κλίμακας από μόνη της δεν σημαίνει απαραίτητα ότι το άτομο με CVD μπορεί να το αποκωδικοποιήσει. Το μέγεθος των χρωματισμένων στοιχείων έχει επίσης σημασία.

7.4.2 Μη Χρήση Σωστού Μεγέθους για τα Χρωματιστά Στοιχεία

Τα χρώματα που χρησιμοποιούνται για γραφικά στοιχεία σε μικρά μεγέθη ή λεπτές γραμμές είναι πιο δύσκολο να ξεχωρίσουν από τα χρώματα σε μεγάλες επιφάνειες ή παχιές γραμμές και αυτό το πρόβλημα, επιδεινώνεται για άτομα με CVD (Εικόνα 7.4.2-9).

Λύσεις: α) η χρησιμοποίηση του χρώματος για να βελτιώσει την οπτική εμφάνιση του σχήματος χωρίς να βασίζεται όμως αποκλειστικά στο χρώμα για να μεταφέρει βασικές πληροφορίες

πλεονάζουσα κωδικοποίηση - redundant coding: είναι η χρήση χρώματος για την ενίσχυση της οπτικής εμφάνισης ενός σχήματος χωρίς αυτό να βασίζεται αποκλειστικά στο χρώμα για τη μετάδοση της σημαντικής πληροφορίας, χρήση δηλαδή πολλαπλών διαφορετικών αισθητικών διαστάσεων)

και β) η χρήση προσομοιωτή ώστε να γίνει δοκιμή του σχήματος για CVD, να μην βασιζόμαστε μόνο σε συγκεκριμένες χρωματικές κλίμακες (υπάρχουν πολλές διαδικτυακές υπηρεσίες και εφαρμογές για υπολογιστές που επιτρέπουν αυθαίρετα σχήματα να ελέγχονται για προσομοίωση CVD όπως έχουμε δει).

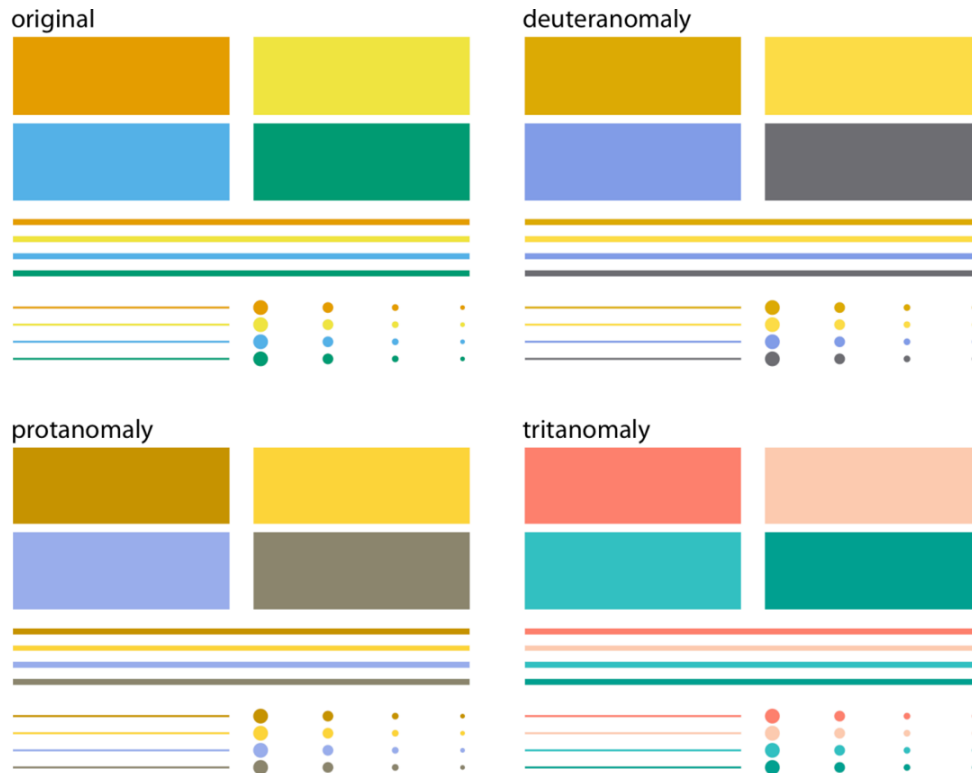


Figure 19-11. Colored elements become difficult to distinguish at small sizes. The top-left panel (labeled “original”) shows four rectangles, four thick lines, four thin lines, and four groups of points, all colored in the same four colors. We can see that the colors become more difficult to distinguish the smaller or thinner the visual elements are. This problem becomes exacerbated in the CVD simulations, where the colors are already more difficult to distinguish even for the large graphical elements.

Εικόνα 7.4.2-9: [2] Προσοχή στη λανθασμένη χρήση χρωματικών στοιχείων

Έτσι, εάν η βασική μεταβλητή ενδιαφέροντος εμφανίζεται ως χρώμα, θα πρέπει να εμφανιστεί σε αρκετά μεγάλες χρωματισμένες περιοχές των χρωμάτων για να μεταφέρει τις σημαντικές πληροφορίες

Παραδείγματα: α) στην Εικόνα 7.4.2-10 πάνω αριστερά, εάν επιλέγονταν τετράγωνα αντί για γραμμές και αυτά χρωματίζονταν βάση της μέσης θερμοκρασίας, θα έμοιαζε με αυτό της εικόνας Εικόνα 7.4.2-10 πάνω δεξιά. β) στην Εικόνα 7.4.2-10 κάτω, φαίνονται οι διαφορετικοί τύποι κλίμακας: θέση, χρώμα, μέγεθος, σχήμα

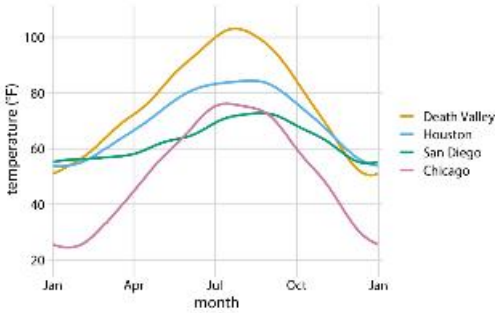


Figure 2-3. Daily temperature normals for four selected locations in the US. Temperature is mapped to the y axis, day of the year to the x axis, and location to line color. Data source: NOAA.

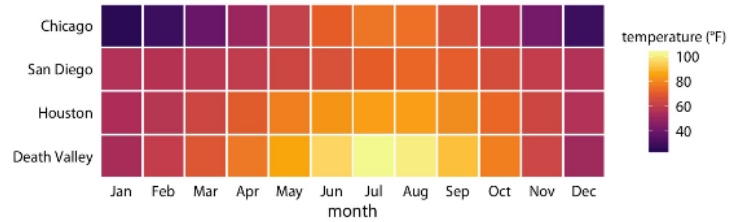


Figure 2-4. Monthly normal mean temperatures for four locations in the US. Data source: NOAA.

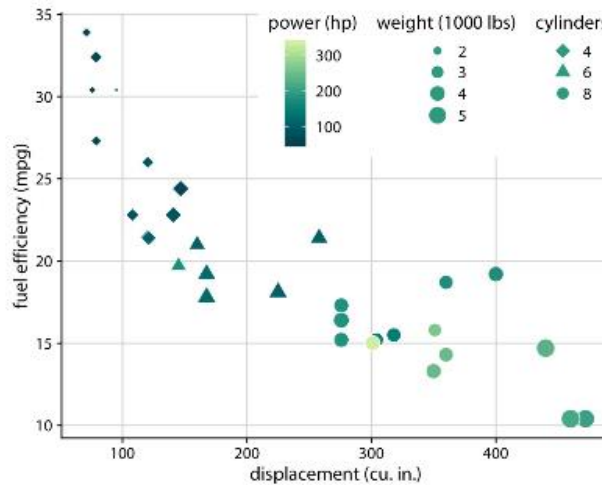


Figure 2-5. Fuel efficiency versus displacement, for 32 cars (1973–74 models). This figure uses five separate scales to represent data: (i) the x axis (displacement); (ii) the y axis (fuel efficiency); (iii) the color of the data points (power); (iv) the size of the data points (weight); and (v) the shape of the data points (number of cylinders). Four of the five variables displayed (displacement, fuel efficiency, power, and weight) are numerical continuous. The remaining one (number of cylinders) can be considered to be either numerical discrete or qualitative ordered. Data source: Motor Trend, 1974.

Εικόνα 7.4.2-10: [2] Παραδείγματα μεταβλητών ως χρώμα

7.4.3 Η Μη Αντιμετώπιση Χρωμάτων που Αναμειγνύονται και Υπερκαλύπτονται

Συχνά, πολλές ομάδες δεδομένων που παρουσιάζονται μόνο με χρώμα, μπλέκονται μεταξύ τους, καθιστώντας δύσκολο το διαχωρισμό του ποια σημεία ανήκουν σε ποιες ομάδες χρωμάτων (όπως για παράδειγμα στο διάγραμμα διασποράς της Εικόνα 7.4.3-11 πάνω αριστερά. Και το πρόβλημα γίνεται πολύ μεγαλύτερο εάν εκτυπωθεί σε κλίμακα του γκρι (Εικόνα 7.4.3-11 δεξιά).

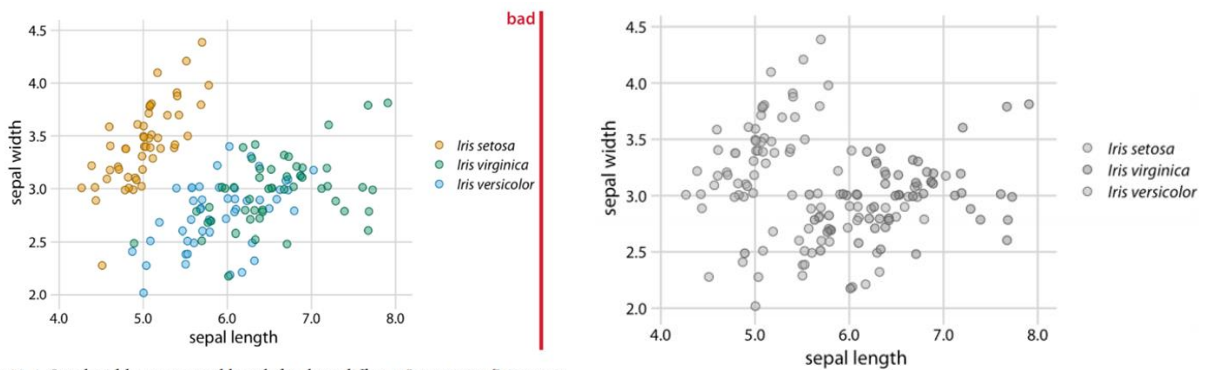


Figure 20-1. Sepal width versus sepal length for three different Iris species (Iris setosa, Iris virginica, and Iris versicolor). Each point represents the measurements for one plant sample. A small amount of jitter has been applied to all point positions to prevent overplotting. The figure is labeled “bad” because the virginica points in green and the versicolor points in blue are difficult to distinguish from each other. Data source: [Fisher 1936].

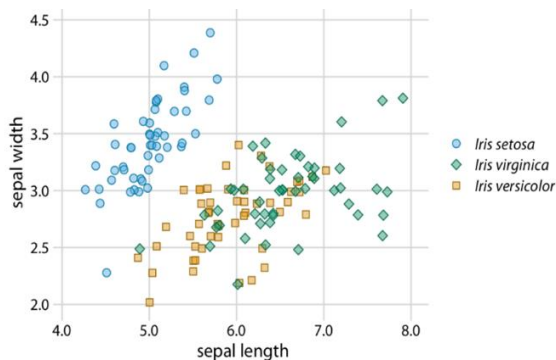


Figure 20-3. Sepal width versus sepal length for three different Iris species. Compared to Figure 20-1, we have swapped the colors for Iris setosa and Iris versicolor and we have given each Iris species its own point shape. Data source: [Fisher 1936].

Εικόνα 7.4.3-11: [2] Ανάμειξη και υπερκάλυψη χρωμάτων

Λύσεις: α) να γίνει ανταλλαγή των χρησιμοποιούμενων χρωμάτων, ώστε χρώματα που είναι παραπλήσια να έχουν μεταξύ τους κάποια απόσταση (πχ Εικόνα 7.4.3-11 κάτω) και, β) να γίνει χρήση διαφορετικών σχημάτων ως σύμβολα της κάθε ομάδας, καθιστώντας τα έτσι πιο σαφή και φιλικά προς άτομα με CVD (Εικόνα 7.4.3-11 κάτω και Εικόνα 7.4.3-12). Αυτό λειτουργεί καλά για τα διαγράμματα διασποράς, αλλά όχι τόσο καλά για διαγράμματα με διακεκομμένες ή διάστικτες γραμμές –(dashed or dotted lines)- εκτός και αν είναι απόλυτες ευθείες ή ελαφρώς καμπυλωμένες.

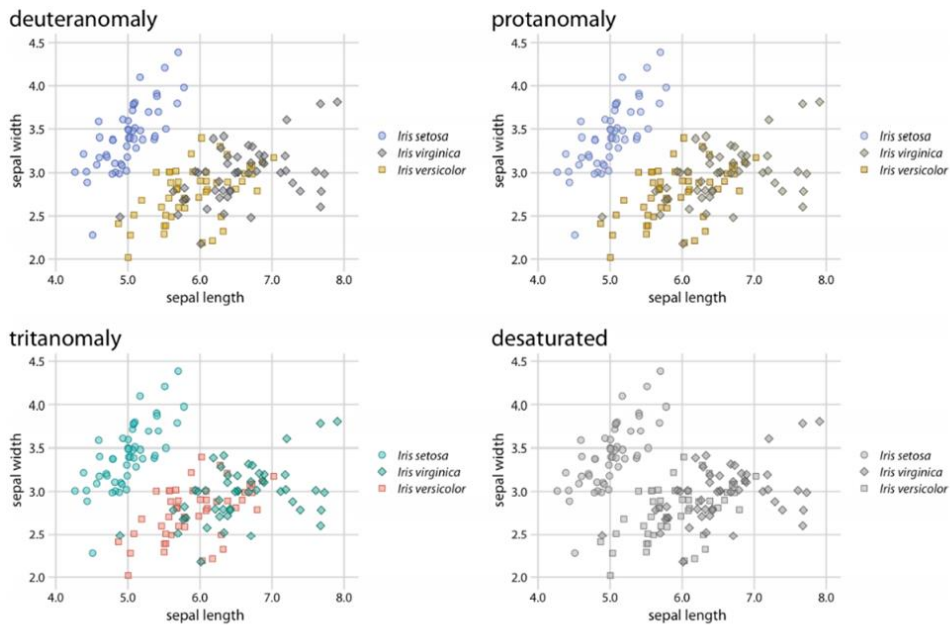


Figure 20-4. Color-vision deficiency simulation of Figure 20-3. Because of the use of different point shapes, even the fully desaturated grayscale version of the figure is legible. Data source: [Fisher 1936].

Εικόνα 7.4.3-12: [2] Χρήση σχημάτων ως σύμβολα, πιο σαφή σε άτομα με CVD

Μερικές φορές όμως η χρήση ημιδιαφάνειας (semi-transparency) δεν είναι αρκετή, είτε επειδή αυτές οι διαφορές στο χρώμα δεν είναι αυτονόητες (σχετικά με το ποιος είναι ο λόγος που είναι τόσο σκούρες), είτε επειδή υπάρχουν περισσότερες από μία ακριβώς ίδιες τιμές η μία πάνω στην άλλη που φαίνονται στο διάγραμμα ως μία, παραπλανώντας έτσι τον αναγνώστη να καταλήξει σε λάθος συμπεράσματα: υπερβολική πλοκή - overplotting (Εικόνα 7.4.3-13 αριστερά).

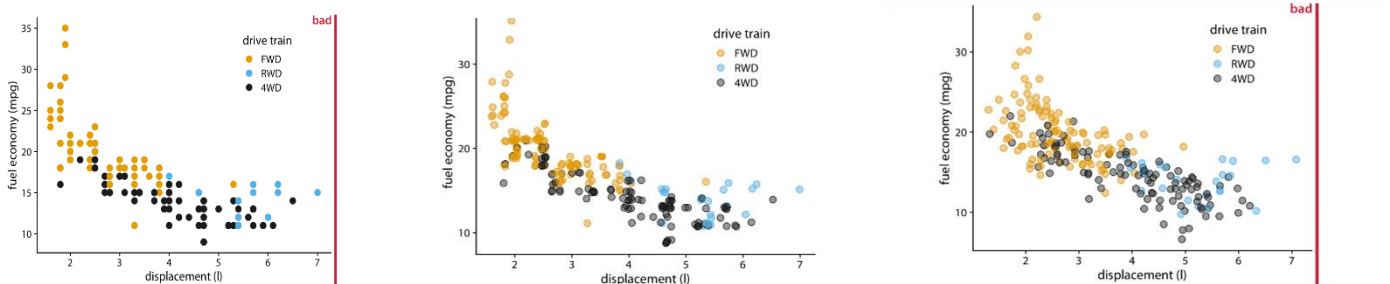


Figure 18-1. City fuel economy versus engine displacement, for popular cars released between 1999 and 2008. Each point represents one car. The point color encodes the drive train: front-wheel drive (FWD), rear-wheel drive (RWD), or four-wheel drive (4WD). The figure is labeled "bad" because many points are plotted on top of others and obscure them. Data source: US Environmental Protection Agency (EPA), <https://fuelconomy.gov>.

Figure 18-3. City fuel economy versus engine displacement. By adding a small amount of jitter to each point, we can increase the visibility of the overplotted points without substantially distorting the message of the plot. Data source: EPA.

Figure 18-4. City fuel economy versus engine displacement. By adding too much jitter to the points, we have created a visualization that does not accurately reflect the underlying dataset. Data source: EPA.

Εικόνα 7.4.3-13: [2] Πρόβλημα υπερκάλυψης

Λύση: να εφαρμοστεί μια μικρή ποσότητα «τρεμοπαίγματος» (jitter) σε αυτά τα σημεία, έτσι ώστε να μπορέσει να λειτουργήσει σωστά η ημιδιαφάνεια (Εικόνα 7.4.3-13 κέντρο). Εφαρμόζοντας πάρα πολύ τρεμοπαίγμα όμως, παραποιεί τα δεδομένα και κάνει το αποτέλεσμα να βγαίνει παραπλανητικό (Εικόνα 7.4.3-13 δεξιά).

Άλλες φορές, τα δεδομένα μπορούν να γίνουν τόσο μεγάλα που ούτε η ημιδιαφάνεια, ούτε το jittering είναι αρκετά για να λύσουν το πρόβλημα, αποκρύπτοντας με αυτόν τον τρόπο χρήσιμες/σημαντικές πληροφορίες. Για παράδειγμα, στην Εικόνα 7.4.3-14 αριστερά, οι κουκκίδες που αντιπροσωπεύουν καθεμία τους και από μία πτήση, είναι τόσο πυκνές που δεν επιτρέπει στον αναγνώστη να δει αν αναχωρούν στην ώρα τους ή με καθυστέρηση, ας πούμε 50 λεπτών ή και περισσότερο, και οι πτήσεις με καθυστέρηση άνω των 400 λεπτών μετά βίας φαίνονται λόγω της ημιδιαφάνειας των κουκκίδων.

Λύση: α) μπορεί να χρησιμοποιηθεί ένα δισδιάστατο ιστόγραμμα (2d-histogram) με το να γίνει συσσώρευση των δεδομένων σε δύο διαστάσεις (αυτό γίνεται με τη διαίρεση του x-y χώρου του διαγράμματος σε μικρά ορθογώνια ή εξάγωνα, μετρώντας πόσα σημεία δεδομένων εμπίπτουν στο καθένα από αυτά και χρησιμοποιώντας στη συνέχεια ημιδιαφάνεια πάνω σε αυτά για να λειτουργήσει σωστά (Εικόνα 7.4.3-14 δεξιά).

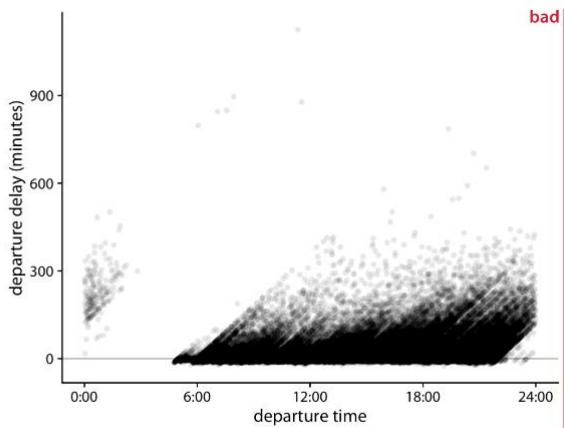


Figure 18-5. Departure delay in minutes versus flight departure time, for all flights departing Newark Airport (EWR) in 2013. Each dot represents one departure. Data source: US Dept. of Transportation, Bureau of Transportation Statistics.

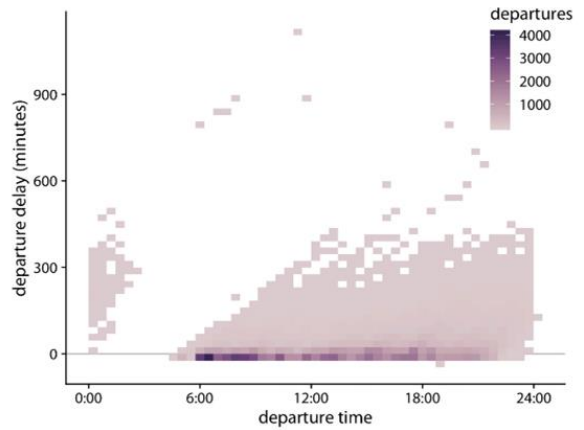


Figure 18-6. Departure delay in minutes versus the flight departure time. Each colored rectangle represents all flights departing at that time with that departure delay. Coloring represents the number of flights represented by that rectangle. Data source: US Dept. of Transportation, Bureau of Transportation Statistics.

Εικόνα 7.4.3-14: [2] Λύση προβλήματος υπερκάλυψης μεγάλων (πυκνών) δεδομένων με χρήση ιστογράμματος

β) ένας άλλος τρόπος να λυθεί αυτό το πρόβλημα είναι με τη χρήση γραμμών περιγράμματος-contour lines (αντί για ορθογώνια ή εξάγωνα όπως παραπάνω), για να δείξουμε περιοχές με διαφορετικές πυκνότητες σημείων (point densities). Αυτό, λειτουργεί πολύ καλά για περιπτώσεις που οι πυκνότητες σημείων αλλάζουν αργά και στις δύο διαστάσεις (ας πούμε για παράδειγμα, ότι υπήρχε πρόβλημα overplotting στην Εικόνα 7.4.3-16 αριστερά, αντιμετωπίζοντας το με τη χρήση της συγκεκριμένης μεθόδου, προκύπτει η Εικόνα 7.4.3-16 κέντρο).

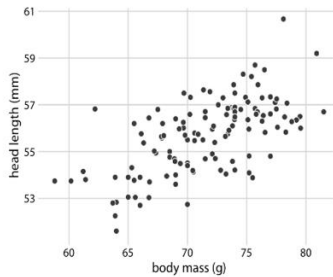


Figure 12-1. Head length (measured from the tip of the bill to the back of the head, in mm) versus body mass (in grams), for 123 blue jays. Each dot corresponds to one bird. There is a moderate tendency for heavier birds to have longer heads. Data source: Keith Tarvin, Oberlin College.

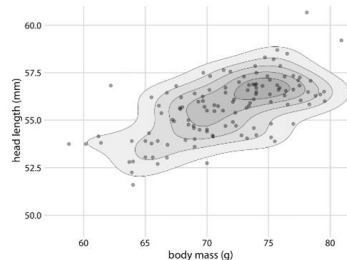


Figure 18-9. Head length versus body mass for 123 blue jays. This figure is nearly identical to Figure 18-8, but now the areas enclosed by the contour lines are shaded with increasingly darker shades of gray. This shading creates a stronger visual impression of increasing point density toward the center of the point cloud. Data source: Keith Tarvin, Oberlin College.

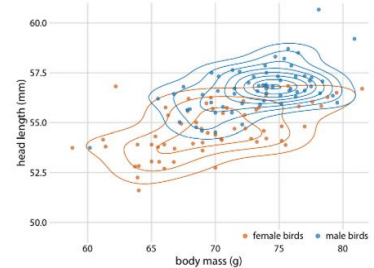


Figure 18-10. Head length versus body mass for 123 blue jays. As in Figure 12-2, we can also indicate the birds' sex by color when drawing contour lines. This figure highlights how the point distribution is different for male and female birds. In particular, male birds are more densely clustered in one region of the plot area whereas female birds are more spread out. Data source: Keith Tarvin, Oberlin College.

Εικόνα 7.4.3-16: [2] Λύση προβλήματος υπερκάλυψης με χρήση γραμμών περιγράμματος

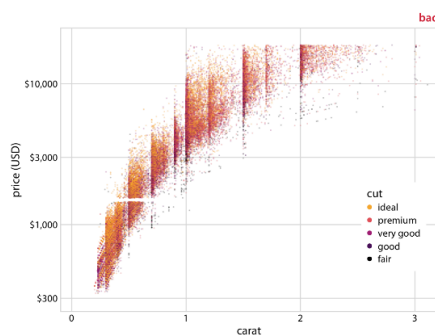


Figure 18-11. Price of diamonds versus their carat value, for 53,940 individual diamonds. Each diamond's cut is indicated by color. The plot is labeled as "bad" because the extensive overplotting makes it impossible to discern any patterns among the different diamond cuts. Data source: Hadley Wickham, ggplot2.

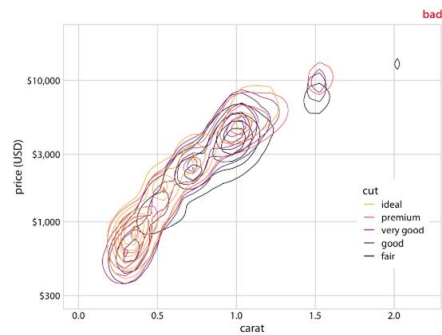


Figure 18-12. Price of diamonds versus their carat value. As in Figure 18-11, but now individual points have been replaced by contour lines. The resulting plot is still labeled "bad," because the contour lines all lie on top of each other. Neither the point distribution for individual cuts nor the overall point distribution can be discerned. Data source: Hadley Wickham, ggplot2.

Εικόνα 7.4.3-15: [2] Πίσω στο αρχικό πρόβλημα

Και αν θέλαμε να δείξουμε ταυτόχρονα πολλαπλά σύνολα γραμμών περιγράμματος για κατανομές πολλών σύννεφων από σημεία (point clouds), θα μπορούσαμε να χρησιμοποιήσουμε ξεχωριστά χρώματα. Ωστόσο, αυτό λειτουργεί μόνο για έναν πολύ μικρό αριθμό σαφώς διαχωρισμένων ομάδων. (όπως για παράδειγμα της Εικόνα 7.4.3-15 δεξιά).

Εάν τώρα υπάρχουν πάρα πολλά τέτοια περιγράμματα που επικαλύπτονται με έντονα χρώματα (Εικόνα 7.4.3-15 δεξιά), τότε ουσιαστικά επιστρέφουμε στο πρόβλημα που υπήρχε εξ αρχής, κάνοντας το σχέδιο περιγράμματος όχι καλύτερο από το αρχικό διάγραμμα διασποράς (εδώ, της Εικόνα 7.4.3-15 αριστερά).

Λύση: Να σχεδιαστούν οι γραμμές περιγράμματος κάθε διαφορετικής χρωματικής κατηγορίας (των ομάδων που φαίνονται στο υπόμνημα) ξεχωριστά, σε δικό τους διάγραμμα το καθένα (Εικόνα 7.4.3-17).

Αυτό, επιτρέπει την οπτική σύγκριση μεταξύ των ομάδων, κοιτώντας πού βρίσκονται οι γραμμές περιγράμματος σε σχέση με τις γραμμές πλέγματος (grid lines). Παρόμοιο αποτέλεσμα θα μπορούσε να είχε επιτευχθεί με τη σχεδίαση ημιδιάφανων μεμονωμένων σημείων αντί για γραμμές περιγράμματος σε κάθε πλαίσιο.

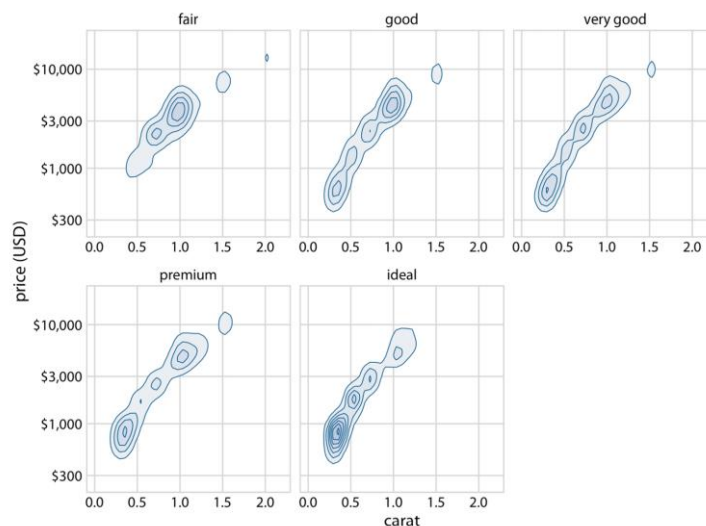


Figure 18-13. Price of diamonds versus their carat value. Here, we have taken the density contours from Figure 18-12 and drawn them separately for each cut. We can now see that better cuts (very good, premium, ideal) tend to have lower carat values than the poorer cuts (fair, good) but command a higher price per carat. Data source: Hadley Wickham, ggplot2.

Εικόνα 7.4.3-17: [2] Λύση προβλήματος υπερκάλυψης με χρήση μικρών πολλαπλών διαγραμμάτων

8 Αισθητική σε Τίτλους, Λεζάντες, Υπομνήματα, Ετικέτες Αξόνων, Γραμμών και Πίνακες!

Μια οπτικοποίηση δεδομένων πρέπει να είναι αισθητικά ευχάριστη, αλλά επίσης, να μεταφέρει πληροφορίες και να δηλώνει μια άποψη (make a point). Για να το επιτύχουμε αυτό, τοποθετούμε τα δεδομένα σε ένα γενικότερο πλαίσιο και παρέχουμε σωστά συνοδευτικούς τίτλους, λεζάντες και άλλους σχολιασμούς για να είναι όσο το δυνατό πιο εύκολα κατανοητά και ξεκάθαρα (αρκεί να λαμβάνεται πάντα υπόψη η αναλογία μελανιού-δεδομένων-ink:data ratio). Σε ορισμένες περιπτώσεις, απαιτείται και η εμφάνιση δεδομένων ως συνοδευτικός πίνακας.

8.1 Τίτλοι και Υπομνήματα

Κάθε διάγραμμα χρειάζεται έναν τίτλο για να μεταφέρει με ακρίβεια στον αναγνώστη περί τίνος πρόκειται, ποιο είναι το νόημα. Ένας τίτλος μπορεί είτε να ενσωματωθεί στο διάγραμμα (Εικόνα 8.1-1 αριστερά) είτε να μπει ως πρώτο στοιχείο της λεζάντας του (Εικόνα 8.1-1 δεξιά). Και οι δύο επιλογές έχουν διαφορετικές περιοχές εφαρμογής:

- οι ενσωματωμένοι τίτλοι είναι πιο κατάλληλοι για αυτόνομα γραφήματα (infographics), μέσα κοινωνικής δικτύωσης ή ιστοσελίδες χωρίς συνοδευτικό κείμενο λεζάντας, ενώ,

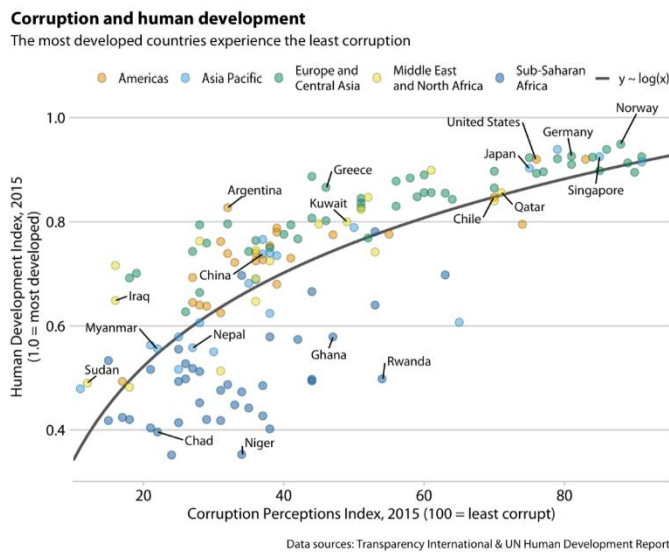


Figure 22-2. Infographic version of Figure 22-1. The title, subtitle, and data source statements have been incorporated into the figure. This figure could be posted on the web as is or otherwise used without a separate caption block.

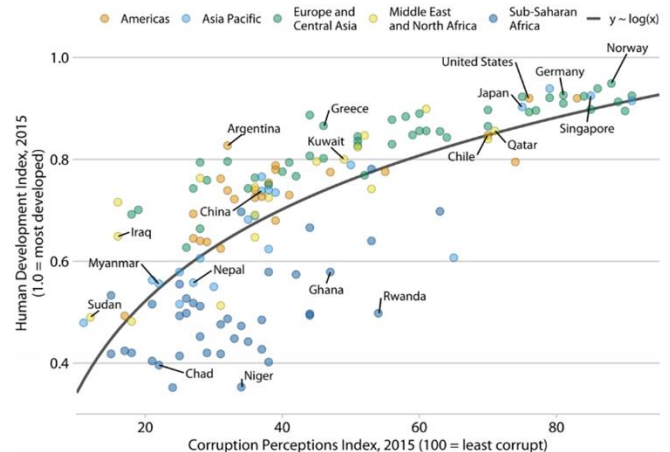


Figure 22-1. Corruption and human development: the most developed countries experience the least corruption. Original figure concept: [The Economist online 2011]. Data sources: Transparency International & UN Human Development Report.

Εικόνα 8.1-1: [2] Τίτλοι ενσωματωμένοι και μη

- οι τίτλοι στις λεζάντες ταιριάζουν καλύτερα για συμβατικές διατάξεις βιβλίων και άρθρων.

Ένας τίτλος δεν χρειάζεται να είναι μια πλήρης πρόταση. Μια σύντομη πρόταση, μπορεί να λειτουργήσει εξίσου καλά.

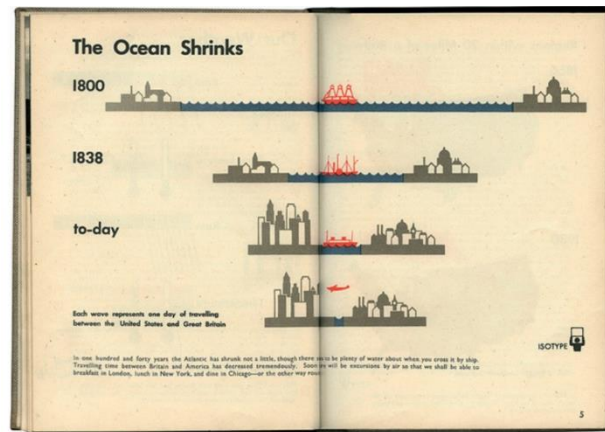


Figure 2.9 “The ocean shrinks,” a visualization by Otto and Marie Neurath of the Isotype Institute. Public domain.

Εικόνα 8.1-2: [1] Σύντομοι τίτλοι, ισχυρό αντίκτυπο

Η επιλογή του τίτλου χρειάζεται προσοχή, μιας και πρέπει να έχει ένα ισχυρό πρώτο αντίκτυπο στον αναγνώστη.

Στην Εικόνα 8.1-2, η αλλαγή του τίτλου “Ο ωκεανός συρρικνώνεται” σε “Χρόνοι ταξιδιού με διαφορετικές μορφές μεταφοράς”, δεν θα είχε τον ίδιο αντίκτυπο. Ή στην Εικόνα 8.1-1, η αλλαγή του σε κάτι πιο σύντομο όπως «Οι πιο ανεπτυγμένες χώρες είναι οι λιγότερο διεφθαρμένες», θα λειτουργούσε μια χαρά.

Τίτλοι, εκτός από το κύριο σχήμα, χρειάζονται επίσης και για τους άξονες και τα υπομνήματα, όπως και επισημάνσεις, ώστε να εξηγούν ποιες είναι οι εμφανιζόμενες τιμές δεδομένων και πώς αυτές χαρτογραφούνται.

Για αριθμητικές μεταβλητές όπως η μάζα σώματος, το μήκος και το μέγεθος, οι σχετικοί τίτλοι δεν θα πρέπει να αναφέρουν μόνο τις μεταβλητές, αλλά και τις μονάδες με τις οποίες μετρούνται (Εικόνα 8.1-4 αριστερά, δείχνει ένα σχήμα με κατάλληλους τίτλους και υπομνήματα). Συχνά, οι τίτλοι αξόνων και υπομνημάτων, παραλείπονται σε περιπτώσεις που οι ετικέτες είναι αυτονόητες, όπως η χρήση του τίτλου “φύλο” για το χρωματικό γένος “θηλυκό” και “αρσενικό” στην Εικόνα 8.1-4 δεξιά ή ο τίτλος “έτη” για τις χρονολογίες στην Εικόνα 8.1-3 αριστερά. Αυτό όμως πρέπει

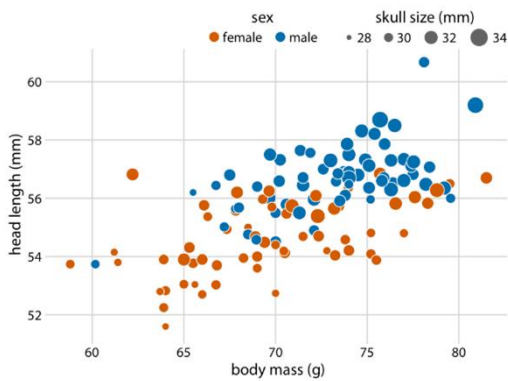


Figure 22-3. Head length versus body mass for 123 blue jays. The birds' sex is indicated by color, and the birds' skull size by symbol size. Head length measurements include the length of the bill while skull size measurements do not. Data source: Keith Tarvin, Oberlin College.

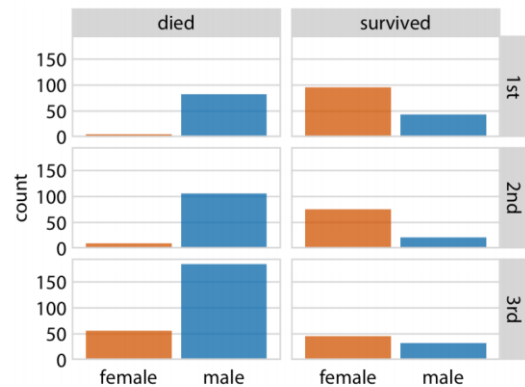


Figure 21-1. Breakdown of passengers on the Titanic by gender, survival, and class in which they traveled (1st, 2nd, or 3rd). Data source: Encyclopedia Titanica.

Εικόνα 8.1-4: [2] Τίτλοι, αριθμητικές μεταβλητές και αυτονόητες παραλήψεις

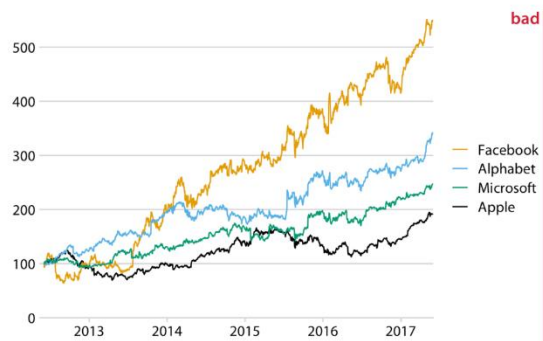


Figure 22-5. Stock price over time for four major tech companies. The stock price for each company has been normalized to equal 100 in June 2012. This variant of Figure 22-4 has been labeled as "bad" because the y axis now does not have a title either, and what the values shown along the y axis represent is not immediately obvious from the context. Data source: Yahoo! Finance.

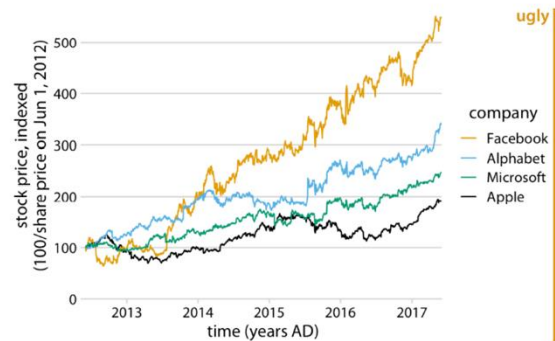


Figure 22-6. Stock price over time for four major tech companies. The stock price for each company has been normalized to equal 100 in June 2012. This variant of Figure 22-4 has been labeled as "ugly" because it is labeled excessively. In particular, providing a unit ("years AD") for the values along the x axis is awkward and unnecessary. Data source: Yahoo! Finance.

Εικόνα 8.1-3: [2] Υπερβολή και λάθη στις αριθμητικές μεταβλητές

να γίνεται με προσοχή, καθώς οι αναγνώστες δεν πρέπει να μαντεύουν τι εννοούμε (όπως στην Εικόνα 8.1-3 αριστερά, τι αντιπροσωπεύουν οι τιμές στον άξονα y), αλλά ούτε και να το παρακάνουμε με τις επισημάνσεις (όπως στην Εικόνα 8.1-3 δεξιά που η χρήση των ετικετών «εταιρεία» ή «χρόνος» δεν χρειάζεται).

Το μέγεθος των ετικετών αξόνων (axis labels), των ετικετών σημείων άξονα (axis tick labels) και άλλων σχολιασμών του διαγράμματός μας, είναι επίσης ένας σημαντικός παράγοντας, αλλά συχνά παραβλέπεται.

Στην Εικόνα 8.1-5 πάνω αριστερά, τα κείμενα παντού είναι πολύ μικρά σε όλα τα στοιχεία γενικά, στις ετικέτες αξόνων, στις ετικέτες των σημείων του άξονα όπως και στις ετικέτες του

υπομνήματος, ενώ η Εικόνα 8.1-5 πάνω δεξιά είναι καλά ισορροπημένη, το κείμενο είναι ευανάγνωστο και ταιριάζει με το συνολικό μέγεθος του σχήματος. Εάν το σχήμα μας προορίζεται για μικρά μεγέθη, μπορεί να χρειάζεται αυτά να γίνουν μεγαλύτερα, έτσι ώστε μαζί με τα άλλα στοιχεία του διαγράμματος να ταιριάζουν καλά μεταξύ τους και να διαβάζονται εύκολα για μικρά μεγέθη. Στην Εικόνα 8.1-5 κάτω αριστερά, τα σημεία/στοιχεία (points) είναι πολύ μικρά, ενώ στην Εικόνα 8.1-5 κάτω δεξιά, είναι πολύ καλά ισορροπημένα.

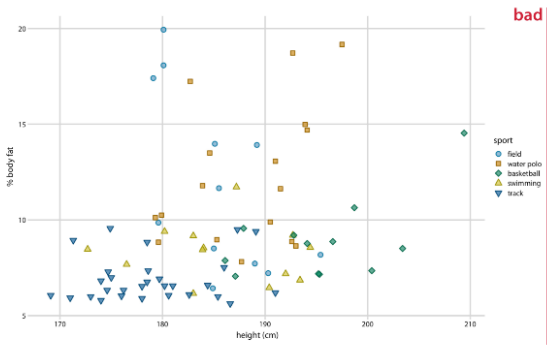


Figure 24-1. Percent body fat versus height in professional male Australian athletes. (Each point represents one athlete.) This figure suffers from the common affliction that the text elements are way too small and are barely legible. Data source: [Telford and Cunningham 1991].

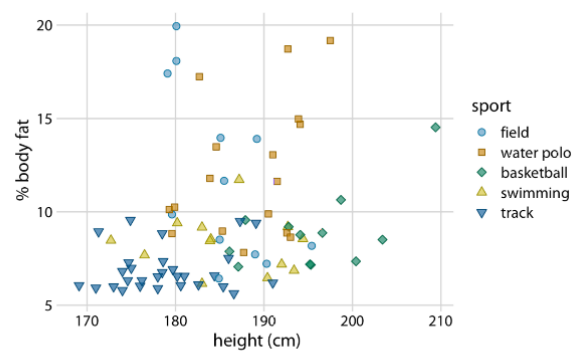


Figure 24-3. Percent body fat versus height in male athletes. All figure elements are appropriately scaled. Data source: [Telford and Cunningham 1991].

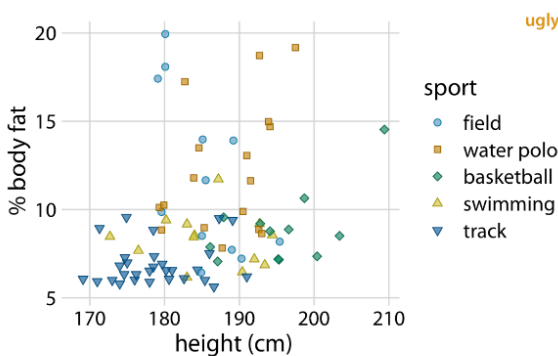


Figure 24-4. Percent body fat versus height in male athletes. The text elements are fairly large, and their size may be appropriate if the figure is meant to be reproduced at a very small scale. However, the figure overall is not balanced; the points are too small relative to the text elements. Data source: [Telford and Cunningham 1991].

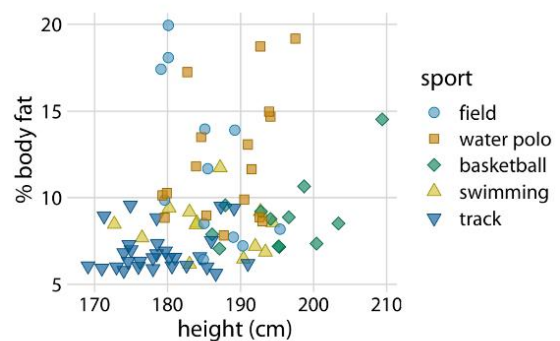


Figure 24-5. Percent body fat versus height in male athletes. All figure elements are sized such that the figure is balanced and can be reproduced at a small scale. Data source: [Telford and Cunningham 1991].

Εικόνα 8.1-5: [2] Μεγέθη ετικετών, σωστή χρήση και ισορροπίες

Μια καλή τακτική για να καταλαβαίνουμε εάν το σχήμα μας συνολικά (με όλα τα στοιχεία του) έχει το σωστό μέγεθος, είναι να παρατηρήσουμε μια μικρότερου μεγέθους έκδοσή του.

Σε ορισμένες περιπτώσεις, μπορούμε ακόμη και να παραλείψουμε ολόκληρο τον άξονα. Αυτό δίνει το μήνυμα στον αναγνώστη ότι τα ποιοτικά χαρακτηριστικά του γραφήματος είναι πιο σημαντικά από τις συγκεκριμένες τιμές των δεδομένων. Τέτοιες περιπτώσεις συναντάμε σε διαγράμματα όπως:

- γραφήματα πίτας και δεντροχάρτες (treemaps), τα οποία συνήθως δεν έχουν σαφείς άξονες (Εικόνα 8.1-7)
- ραβδογράμματα και διαγράμματα μωσαϊκού (mosaic plots), όπου ο ένας ή και οι δύο άξονες μπορούν να παραλειφθούν εάν το νόημα της γραφικής παράστασης είναι σαφές χρησιμοποιώντας κάποιον άλλο τρόπο (Εικόνα 8.1-6).

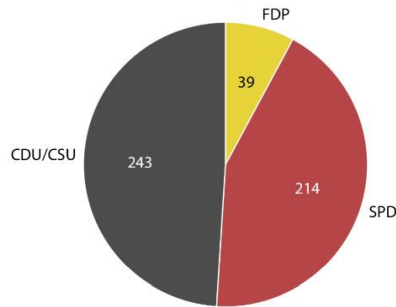


Figure 10-1. Party composition of the eighth German Bundestag, 1976–1980, visualized as a pie chart. This visualization highlights that the ruling coalition of SPD and FDP had a small majority over the opposition CDU/CSU. Data source: Wikipedia.

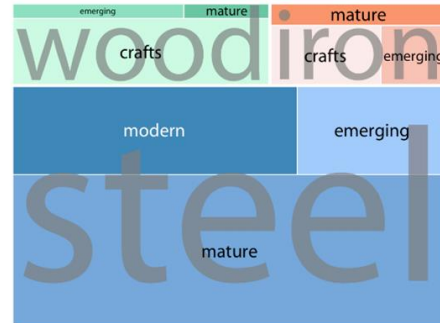


Figure 11-4. Breakdown of bridges in Pittsburgh by construction material (steel, wood, iron) and by era of construction (crafts, emerging, mature, modern), shown as a tree-map. The area of each rectangle is proportional to the number of bridges of that type. Data source: Yoram Reich and Steven J. Fenves.

Εικόνα 8.1-7: [2] Παράλειψη αξόνων σε πίτες και δεντροχάρτες

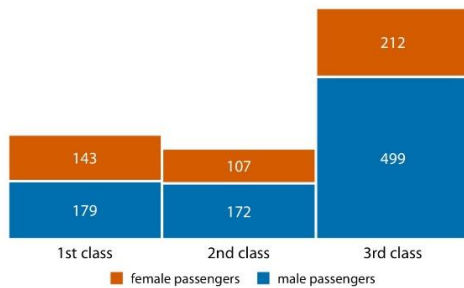


Figure 6-10. Numbers of female and male passengers on the Titanic traveling in 1st, 2nd, and 3rd class. Data source: Encyclopedia Titanica.

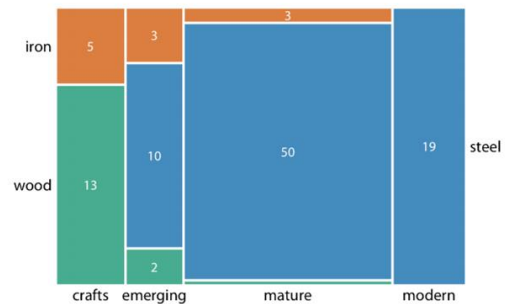


Figure 11-3. Breakdown of bridges in Pittsburgh by construction material (steel, wood, iron) and by era of construction (crafts, emerging, mature, modern), shown as a mosaic plot. The widths of each rectangle are proportional to the number of bridges constructed in that era, and the heights are proportional to the number of bridges constructed from that material. Numbers represent the counts of bridges within each category. Data source: Yoram Reich and Steven J. Fenves.

Εικόνα 8.1-6: [2] Παράλειψη αξόνων σε ραβδογράμματα και διαγράμματα μωσαϊκού

Για περιπτώσεις που υπάρχει οπτική σειρά στα δεδομένα μας, πρέπει να βεβαιωθούμε ότι η ίδια υπάρχει και στο υπόμνημα. Το να μη ληφθεί αυτή υπόψιν, δημιουργεί μια σημαντική ποσότητα διανοητικής προσπάθειας στο να προσπαθήσει κανείς να το καταλάβει. Αυτό συμβαίνει συνήθως επειδή δεν δίνουμε την απαραίτητη προσοχή στην οπτική σειρά των υπομνημάτων που δημιουργούνται αυτόματα από το λογισμικό που χρησιμοποιούμε.

Στην Εικόνα 8.1-8 πάνω αριστερά, χρησιμοποιήθηκαν φιλικές για άτομα με CVD γραμμές, για την εμφάνιση των αλλαγών των τιμών στην πάροδο του χρόνου. Η αντιληπτή σειρά των γραμμών δεδομένων στο σχήμα, διαφέρει από τη σειρά των εταιρειών που υπάρχει στο υπόμνημα, προσθέτοντας έτσι μεγαλύτερη νοητική προσπάθεια ώστε να γίνει το μεταξύ τους ταίριασμα.

Λύση: Να αναδιαταχθούν χειροκίνητα οι καταχωρήσεις των εταιρειών στο υπόμνημα, ώστε να ταιριάζουν με την αντιληπτή σειρά τους στο σχήμα, καθιστώντας έτσι τα δεδομένα πιο χρήσιμα και προφανή Εικόνα 8.1-8 πάνω δεξιά, ακόμη και για αναγνώστες με έλλειψη χρωματικής όρασης (Εικόνα 8.1-8 κάτω).

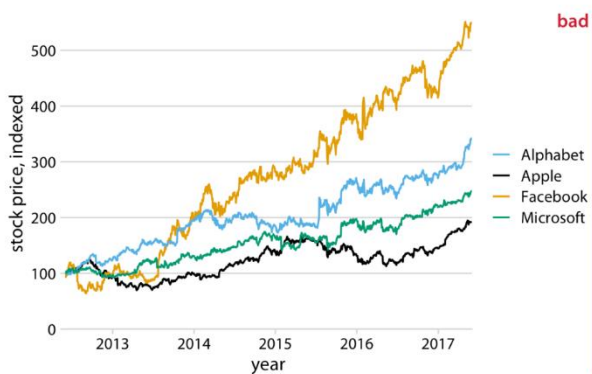


Figure 20-5. Stock price over time for four major tech companies. The stock price for each company has been normalized to equal 100 in June 2012. This figure is labeled a “bad” because it takes considerable mental energy to match the company names in the legend to the data curves. Data source: Yahoo! Finance.

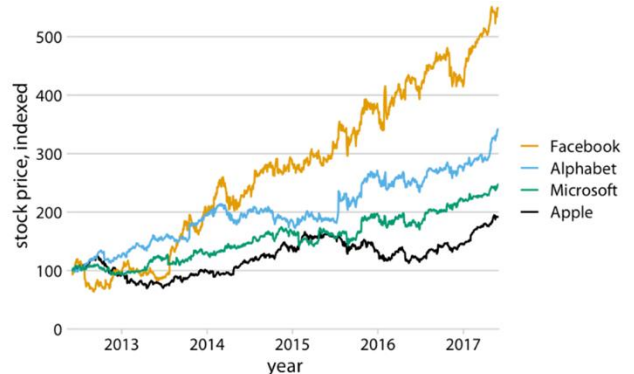


Figure 20-6. Stock price over time for four major tech companies. Compared to Figure 20-5, the entries in the legend have now been ordered such that they match the perceived visual order of the data lines, with Facebook the highest and Apple the lowest. Data source: Yahoo! Finance.

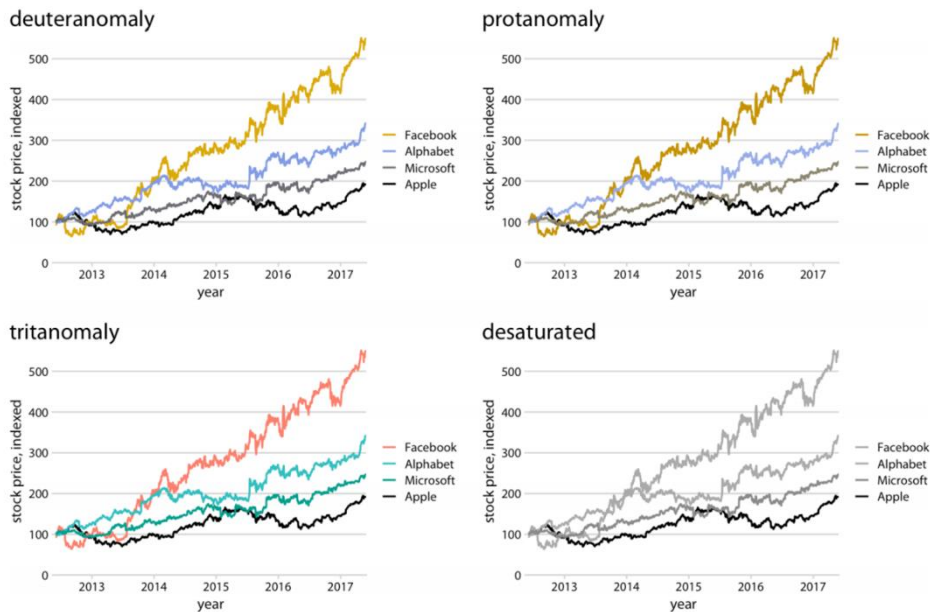


Figure 20-7. Color-vision deficiency simulation of Figure 20-6. Data source: Yahoo! Finance.

Εικόνα 8.1-8: [2] Αντιληπτή σειρά δεδομένων και υπομνήματος

Αν και τα υπομνήματα είναι πολύ χρήσιμα, προσθέτουν νοητική επιβάρυνση στον αναγνώστη. Γι' αυτό λοιπόν, είναι καλύτερο να σχεδιάζουμε το σχήμα μας με τέτοιο τρόπο ώστε να μην χρειάζονται υπομνήματα, αλλά να είναι εξίσου προφανές τι αντιπροσωπεύουν τα διάφορα γραφικά στοιχεία σε αυτό.

Λύση: μπορούμε να βάλουμε ετικέτες απευθείας στα στοιχεία με το κατάλληλο κείμενο (Εικόνα 8.1-9 πάνω αριστερά) ή με χρήση άλλων οπτικών στοιχείων που χρησιμεύουν ως οδηγός για το υπόλοιπο σχήμα. Όπως για παράδειγμα στην Εικόνα 8.1-9 πάνω δεξιά η χρήση κύκλων ώστε να οριστούν οι ομάδες που θέλουμε να βάλουμε την άμεση επισήμανση (ετικέτες) στην πλοκή μας. Συνήθως βοηθά και ο χρωματισμός των ετικετών κειμένου όταν γίνεται με τα ίδια χρώματα που χρησιμοποιούμε στα δεδομένα (Εικόνα 8.1-9)



Figure 20-8. Stock price over time for four major tech companies. The stock price for each company has been normalized to equal 100 in June 2012. Data source: Yahoo! Finance.

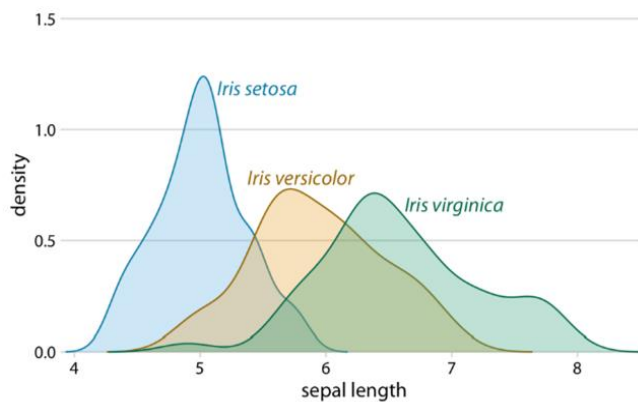


Figure 20-10. Density estimates of the sepal lengths of three different Iris species. Each density estimate is directly labeled with the respective species name. Data source: [Fisher 1936].

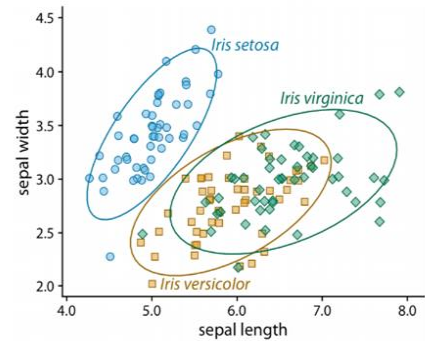


Figure 20-9. Sepal width versus sepal length for three different Iris species. The points representing different Iris species have been directly labeled with colored ellipses and text labels. Compared to Figure 20-3, I have removed the background grid here because the figure was becoming too busy. Data source: [Fisher 1936].

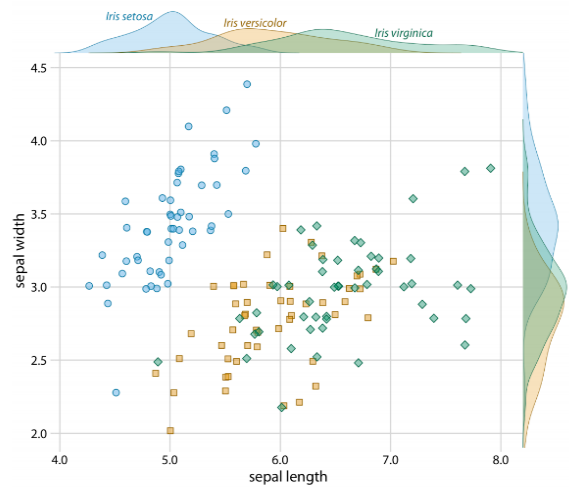


Figure 20-11. Sepal width versus sepal length for three different Iris species, with marginal density estimates of each variable for each species. Data source: [Fisher 1936].

Εικόνα 8.1-9: [2] Αποφυγή υπομνημάτων με μέτρο

Στην εκτύπωση όμως, εάν τα κείμενα είναι πολύ ανοιχτόχρωμα, γίνονται δυσανάγνωστα.

Λύσεις: α) να χρησιμοποιούνται δύο αποχρώσεις κάθε χρώματος, το πιο ανοιχτό για τις περιοχές γεμίματος και το πιο σκούρο για συνοδευτικά κείμενα, γραμμές και περιγράμματα, β) να χρησιμοποιηθεί ένα διάγραμμα πυκνότητας που χρησιμεύει ως αντικατάσταση του υπομνήματος, τοποθετώντας το στο περιθώριο ενός διαγράμματος διασποράς (Εικόνα 8.1-9 κάτω δεξιά), καθιστώντας το σε ορισμένες περιπτώσεις λιγότερο ακατάστατο από την άμεση χρήση επισήμανσης (Εικόνα 8.1-9 πάνω δεξιά).

Για την κωδικοποίηση μιας μεμονωμένης μεταβλητής σε πολλαπλά αισθητικά στοιχεία, θα μπορούσε να γίνει χρήση κάποιου στοιχείου για τη μεταφορά όλης της αντιστοίχισης με την πρώτη ματιά αν αυτό είναι δυνατό, αντί να χρησιμοποιηθούν πολλά ξεχωριστά υπομνήματα για να το πετύχουμε.

Στην Εικόνα 8.1-10, η μεταβλητή εμφανίζεται τόσο ως τοποθεσία, όσο και ως γραμμή χρώματος κατά μήκος του άξονα x).

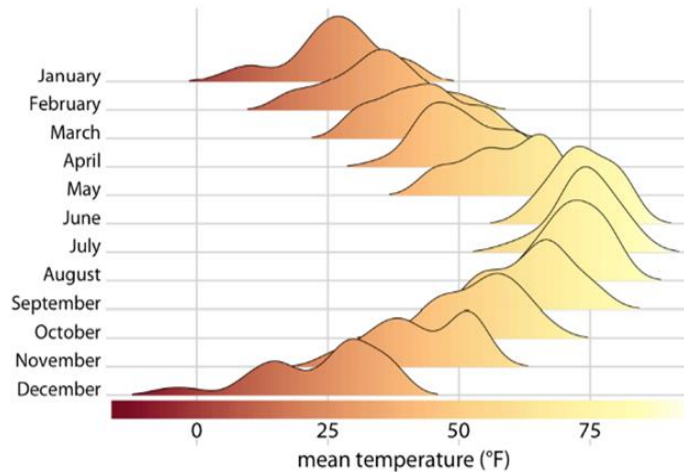


Figure 20-12. Temperatures in Lincoln, NE, in 2016. This figure is a variation of Figure 9-9. Temperature is now shown both by location along the x axis and by color, and a color bar along the x axis visualizes the scale that converts temperatures into colors. Data source: Weather Underground.

Εικόνα 8.1-10: [2] Κωδικοποίηση μεμονωμένης μεταβλητής σε πολλαπλά αισθητικά στοιχεία με χρήση στοιχείου και όχι υπομνήματος

8.2 Πίνακες

Οι πίνακες παίζουν σημαντικό ρόλο στην οπτικοποίηση δεδομένων, αλλά τα προτεινόμενα που δίνουν τα προεπιλεγμένα προγράμματα λογισμικού δεν είναι πάντα τα καλύτερα. Λίγα πράγματα που πρέπει να θυμόμαστε:

- να αποφεύγονται οι κάθετες γραμμές.
- να αποφεύγονται οι οριζόντιες γραμμές μεταξύ των σειρών των δεδομένων: η χρήση των γραμμών ως διαχωριστικών μεταξύ της γραμμής τίτλου και πρώτης γραμμής των δεδομένων, ή ως πλαίσιο για ολόκληρο τον πίνακα, αποτελούν εξαίρεση. Η πρόθεση να βοηθήσουμε οπτικά το μάτι να διαχωρίσει τις σειρές του πίνακα, βοηθά μόνο εάν οι γραμμές είναι φαρδιές και αραιές, αλλά αυτό έχει το κόστος της οπτικής ακαταστασίας. Μια καλή λύση για αυτό, είναι η εναλλαγή πιο ανοιχτόχρωμης και σκουρόχρωμης σκίασης των γραμμών που τείνουν να λειτουργούν καλά χωρίς να δημιουργούν πολύ οπτική ακαταστασία (clutter).
- να γίνεται αριστερή στοίχιση των στηλών κειμένου
- να γίνεται δεξιά στοίχιση στις στήλες με αριθμούς και να χρησιμοποιείται πάντα ο ίδιος αριθμός δεκαδικών ψηφίων σε όλη την έκταση του πίνακα.
- να γίνεται στοίχιση των στηλών που περιέχουν μεμονωμένους χαρακτήρες, στο κέντρο.
- η στοίχιση των επικεφαλίδων να γίνεται μαζί με τα δεδομένα τους, π.χ. για στήλες κειμένου, η στοίχιση της επικεφαλίδας να είναι αριστερά, για αριθμούς δεξιά κτλ.

Τέλος, η λεζάντα ενός πίνακα πρέπει να βρίσκεται πάνω από αυτόν, επειδή οι αναγνώστες τείνουν να τον επεξεργάζονται ως κείμενο, δηλαδή, από πάνω προς τα κάτω (Εικόνα 8.2-11).

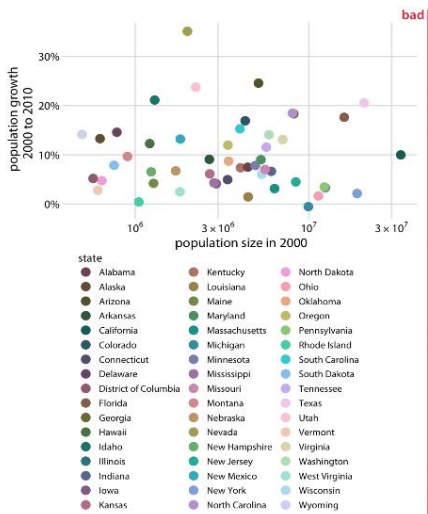


Figure 19-1. Population growth from 2000 to 2010 versus population size in 2000, for 50 US states and the District of Columbia. Every state is marked in a different color. Because there are so many states, it is very difficult to match the colors in the legend to the dots in the scatterplot. Data source: US Census Bureau.

a **ugly**

Rank	Title	Amount
1	<i>Star Wars: The Last Jedi</i>	\$71,565,498
2	<i>Jumanji: Welcome to the Jungle</i>	\$36,169,328
3	<i>Pitch Perfect 3</i>	\$19,928,525
4	<i>The Greatest Showman</i>	\$8,805,843
5	<i>Ferdinand</i>	\$7,316,746

b **ugly**

Rank	Title	Amount
1	<i>Star Wars: The Last Jedi</i>	\$71,565,498
2	<i>Jumanji: Welcome to the Jungle</i>	\$36,169,328
3	<i>Pitch Perfect 3</i>	\$19,928,525
4	<i>The Greatest Showman</i>	\$8,805,843
5	<i>Ferdinand</i>	\$7,316,746

c

Rank	Title	Amount
1	<i>Star Wars: The Last Jedi</i>	\$71,565,498
2	<i>Jumanji: Welcome to the Jungle</i>	\$36,169,328
3	<i>Pitch Perfect 3</i>	\$19,928,525
4	<i>The Greatest Showman</i>	\$8,805,843
5	<i>Ferdinand</i>	\$7,316,746

d

Rank	Title	Amount
1	<i>Star Wars: The Last Jedi</i>	\$71,565,498
2	<i>Jumanji: Welcome to the Jungle</i>	\$36,169,328
3	<i>Pitch Perfect 3</i>	\$19,928,525
4	<i>The Greatest Showman</i>	\$8,805,843
5	<i>Ferdinand</i>	\$7,316,746

Figure 22-7. Examples of poorly and appropriately formatted tables, using the data from Table 6-1 in Chapter 6. (a) This table violates numerous conventions of proper table formatting, including using vertical lines, using horizontal lines between data rows, and using centered data columns. (b) This table suffers from all problems of (a), and also creates visual noise by alternating between very dark and very light rows. Also, the table header is not strongly visually separated from the table body. (c) This is an appropriately formatted table with a minimal design. (d) Colors can be used effectively to group data into rows, but the color differences should be subtle. The table header can be set off by using a stronger color. Data source: *Box Office Mojo*. Used with permission.

Εικόνα 8.2-11: [2] Λεζάντες και πίνακες

8.3 Αποφυγή Χρήσης Γραμμών για Απεικόνιση Διαγραμμάτων

Η χρήση γραμμών για την απεικόνιση των δεδομένων (Εικόνα 8.3-13) θα πρέπει να αποφεύγεται όσο το δυνατόν περισσότερο σε ένα διάγραμμα, καθώς:

- δεν γίνονται εύκολα αντιληπτά ως σταθερά αντικείμενα
- δημιουργούν οπτικές ψευδαισθήσεις και οπτικά αντικείμενα (visual artifacts)
- οι ποσότητες δεν γίνονται εύκολα αντιληπτές,
- προκαλούν σύγχυση σχετικά με το ποια είναι η περιοχή ενδιαφέροντος
- μας αποσπούν από τον κύριο σκοπό τους,
- δεν δημιουργούν ισχυρά όρια για τις κλειστές/περιφραγμένες περιοχές (είτε διακεκομμένες, είτε όχι),
- δημιουργούν οπτικό θόρυβο και

- οι τύποι των σχημάτων που είναι ανοιχτοί δεν διαχωρίζονται έντονα ο ένας από τον άλλο.

Λύση: Η χρήση συμπαγών ή ημιδιάφανων έγχρωμων σχημάτων. Η χρήση χρώματος, εκτός από το να λύνει τέτοια θέματα, δίνει επίσης και μια πιο ξεκάθαρη και ευχάριστη αίσθηση συνολικά (Εικόνα 8.3-12).

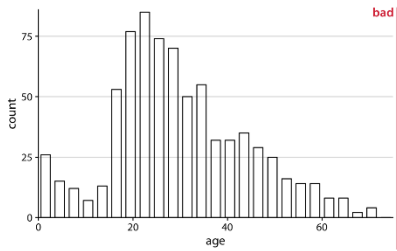


Figure 25-1. Histogram of the ages of Titanic passengers, drawn with empty bars. The empty bars create a confusing visual pattern. In the center of the histogram, it is difficult to tell which parts are inside of bars and which parts are outside. Data source: Encyclopedia Titanica.

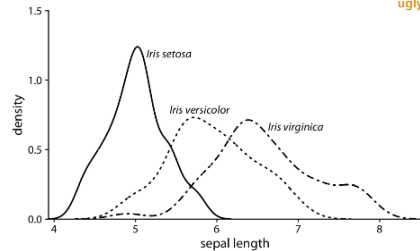


Figure 25-3. Density estimates of the sepal lengths of three different Iris species. The broken line styles used for Iris versicolor and Iris virginica detract from the perception that the areas under the curves are distinct from the areas above them. Data source: [Fisher 1936].

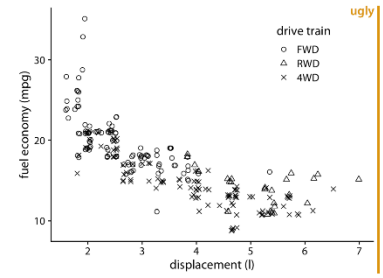


Figure 25-6. City fuel economy versus engine displacement, for cars with front-wheel drive (FWD), rear-wheel drive (RWD), and all-wheel drive (4WD). The different point styles, all black-and-white line-drawn symbols, create substantial visual noise and make it difficult to read the figure. Data source: US Environmental Protection Agency (EPA), <https://fuelconomy.gov>.

Εικόνα 8.3-13: [2] Αποφυγή γραμμών στην απεικόνιση

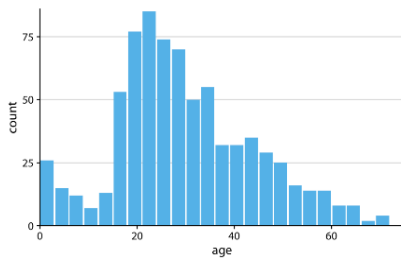


Figure 25-2. Histogram of the ages of Titanic passengers. This is the same histogram in Figure 25-1, now drawn with filled bars. The shape of the age distribution is much more easily discernible in this variation of the figure. Data source: Encyclopedia Titanica.

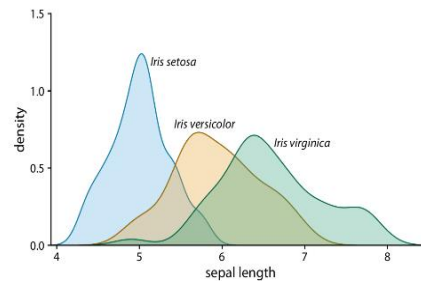


Figure 25-5. Density estimates of the sepal lengths of three different Iris species, shown as partially transparent shaded areas. The shading helps us perceive the three density curves as three distinct objects. Data source: [Fisher 1936].

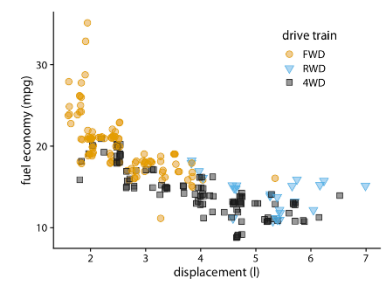


Figure 25-7. City fuel economy versus engine displacement. By using both different colors and different solid shapes for the different drive-train variants, this figure visually separates the drive-train variants while remaining reproducible in grayscale if needed. Data source: EPA.

Εικόνα 8.3-12: [2] Αποφυγή γραμμών στην απεικόνιση με χρήση συμπαγών/ημιδιάφανων έγχρωμων σχημάτων

Αρχή της αναλογικής μελάνης (Principle of Proportional Ink): «Όταν μια σκιασμένη περιοχή χρησιμοποιείται για να αναπαραστήσει μια αριθμητική τιμή, η περιοχή αυτής της σκιασμένης περιοχής θα πρέπει να είναι ευθέως ανάλογη με την αντίστοιχη τιμή».

Στην Εικόνα 8.3-15: τα διαγράμματα είναι εύκολο να παρερμηνευτούν, καθώς ο άξονας των y δεν ξεκινά από το μηδέν και η απόσταση/διαφορά φαίνεται μεγαλύτερη από ό,τι πραγματικά είναι, δίνοντας έτσι τη λανθασμένη εντύπωση ότι η βασική ποσότητα είναι το ύψος και όχι οι τιμές των τελικών σημείων σε σχέση με τον άξονα y. Για να δειχθούν σωστά, θα έπρεπε να έμοιαζαν με τα αντίστοιχα διαγράμματα της Εικόνα 8.3-15. Η σκίαση δίνει έμφαση στην απόσταση από τη θέση του άξονα των x έως τις συγκεκριμένες τιμές y που εμφανίζονται.

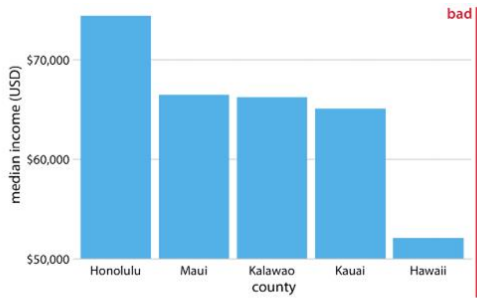


Figure 17-1. Median income in the five counties of the state of Hawaii. This figure is misleading, because the y-axis scale starts at \$50,000 instead of \$0. As a result, the bar heights are not proportional to the values shown, and the income differential between the county of Hawaii and the other four counties appears much bigger than it actually is. Data source: 2015 Five-Year American Community Survey.

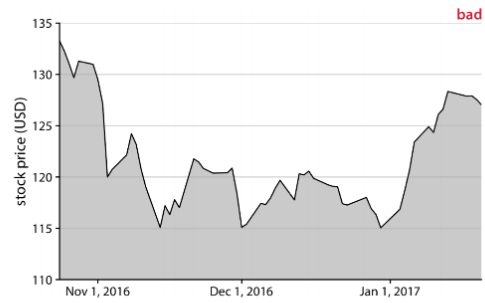


Figure 17-3. Stock price of Facebook (FB) from Oct. 22, 2016 to Jan. 21, 2017. This figure seems to imply that the FB stock price collapsed around Nov. 1, 2016. However, this is misleading, because the y axis starts at \$110 instead of \$0. Data source: Yahoo! Finance.

Εικόνα 8.3-15: [2] Παρερμηνεία διαγραμμάτων λόγω άξονα

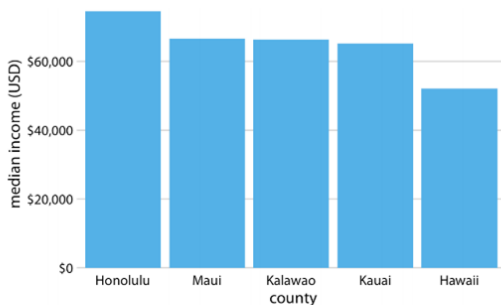


Figure 17-2. Median income in the five counties of the state of Hawaii. Here, the y-axis scale starts at \$0 and therefore the relative magnitudes of the median incomes in the five counties are accurately shown. Data source: 2015 Five-Year American Community Survey.

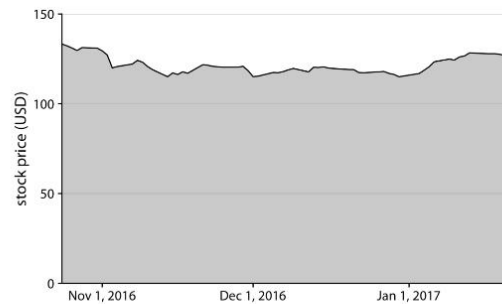


Figure 17-4. Stock price of Facebook (FB) from Oct. 22, 2016 to Jan. 21, 2017. By showing the stock price on a y scale from \$0 to \$150, this figure more accurately relays the magnitude of the FB price drop around Nov. 1, 2016. Data source: Yahoo! Finance.

Εικόνα 8.3-15: [2] Σωστή χρήση αξόνων για σωστή ερμηνεία διαγραμμάτων

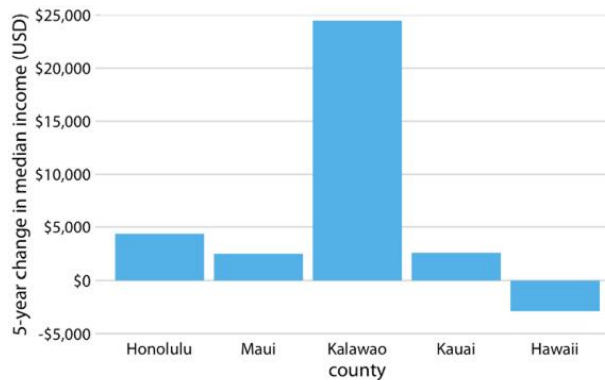


Figure 17-5. Change in median income in Hawaiian counties from 2010 to 2015. Data source: 2010 and 2015 Five-Year American Community Surveys.

Εικόνα 8.3-16: [2] Σωστή χρήση αξόνων για εμφάνιση διαφορών μεταξύ των συνθηκών

Οι ράβδοι (Bars) μεταφέρουν λάθος πληροφορίες/μήνυμα εάν οι άξονές τους δεν ξεκινούν από το μηδέν. Είτε θα πρέπει να εμφανίζονται οι σωστές αναλογίες, ή, θα πρέπει να χρησιμοποιούνται τα στοιχεία ως σημεία κουκκίδας για να εμφανίζεται το μήνυμα. Τα ίδια προβλήματα υπάρχουν

συχνά όταν απεικονίζουμε χρονοσειρές. Οι ράβδοι ή οι σκιασμένες περιοχές μπορούν να χρησιμοποιηθούν για να δείξουν τις διαφορές μεταξύ των συνθηκών, αρκεί μόνο να αναφέρονται ρητά (Εικόνα 8.3-16).

Στην οπτικοποίηση σε μια γραμμική κλίμακα, ρόλο παίζει επίσης το αν θέλουμε να απεικονίσουμε ποσά ή αναλογίες. Οι αναλογίες (Ratios) γενικά εμφανίζονται σε λογαριθμική κλίμακα (log scale). Η λογαριθμική κλίμακα προτιμάται επίσης για σύνολα δεδομένων που περιέχουν αριθμούς με πολύ διαφορετικά μεγέθη (magnitudes) στις τιμές.

Μια λογαριθμική κλίμακα μπορεί να θεωρηθεί ως ο πολλαπλασιασμός με μια σταθερή τιμή για τιμές μεγαλύτερες από 1, και ως διαίρεση με μια σταθερή τιμή για τιμές μικρότερες του 1. Δεν είναι δυνατή η εμφάνιση μηδενικών τιμών (0) σε αυτή την κλίμακα. Εδώ, το μέσο είναι το «1», αντί για το «0» που είναι σε γραμμική κλίμακα, επομένως, εάν το σύνολο δεδομένων περιέχει δεδομένα μηδενικών (0) τιμών, αυτά δεν μπορούν να εμφανιστούν.

Παράδειγμα γραμμικής/λογαριθμικής κλίμακας: Στην Εικόνα 8.3-18, και τα δύο σχήματα, σκοπό έχουν να δείξουν τις κομητείες ανά πληθυσμό από τις περισσότερες έως τις λιγότερο πυκνοκατοικημένες, με τον πληθυσμό στο Harris να είναι 4.092.459 (ο μεγαλύτερος πληθυσμός) και στο Loving να είναι 82 (ο μικρότερος). Ανάλογα με την επιλογή μας (γραμμική ή λογαριθμική κλίμακα), ο αναγνώστης λαμβάνει το μήνυμα διαφορετικά.

Η προσπάθεια οπτικοποίησης των ποσοτήτων σε μια λογαριθμική κλίμακα, δημιουργεί διάφορα προβλήματα και παραπλανά τον αναγνώστη, καθώς οι τιμές δεδομένων δεν γίνεται να ξεκινούν από το 0 (όπως ήδη έχουμε αναφέρει). Και αυτό χειροτερεύει αν χρησιμοποιούμε ράβδους (bars) για να το κάνουμε (Εικόνα 8.3-19 αριστερά).

Μια λύση θα ήταν η αντικατάστασή τους με κουκκίδες (dots), έτσι ώστε: να αποφευχθεί η οπτική αντίληψη του μεγέθους (magnitude), ο άξονας να μην ξεκινά από το μηδέν και η τοποθέτηση των ονομάτων να γίνεται ακριβώς δίπλα από τις κουκκίδες. Έτσι έχουμε μια πιο ξεκάθαρη κατανόηση και αποφεύγουμε την οπτική αντίληψη του μεγέθους που μεταφέρει η απόσταση από τον άξονα (Εικόνα 8.3-19 δεξιά).

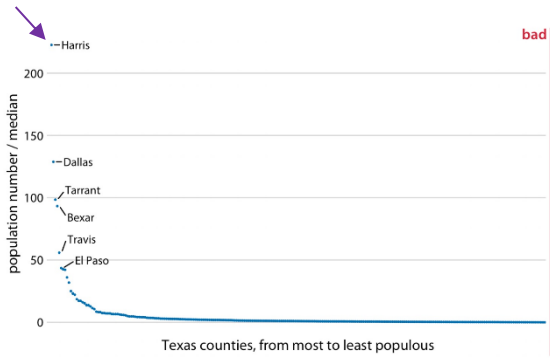


Figure 3-6. Population sizes of Texas counties relative to their median value. By displaying a ratio on a linear scale, we have overemphasized ratios > 1 and have obscured ratios < 1 . As a general rule, ratios should not be displayed on a linear scale. Data source: 2010 US Decennial Census.

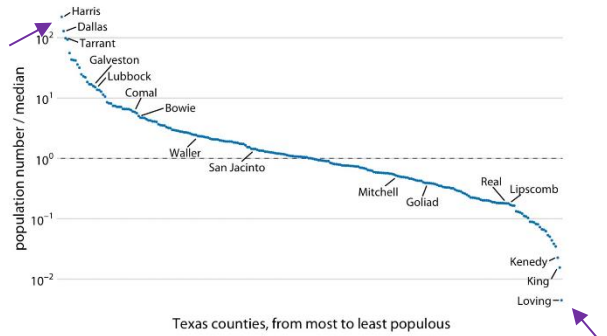


Figure 3-5. Population numbers of Texas counties relative to their median value. Select counties are highlighted by name. The dashed line indicates a ratio of 1, corresponding to a county with median population number. The most populous counties have approximately 100 times more inhabitants than the median county, and the least populous counties have approximately 100 times fewer inhabitants than the median county. Data source: 2010 US Decennial Census.

Εικόνα 8.3-18: [2] Παράδειγμα γραμμικής/λογαριθμικής κλίμακας

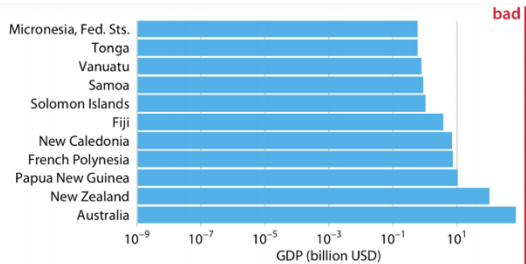


Figure 17-8. GDP in 2007 of countries in Oceania. The lengths of the bars do not accurately reflect the data values shown, since bars start at the arbitrary value of 10^{-9} billion USD. Data source: Gapminder.

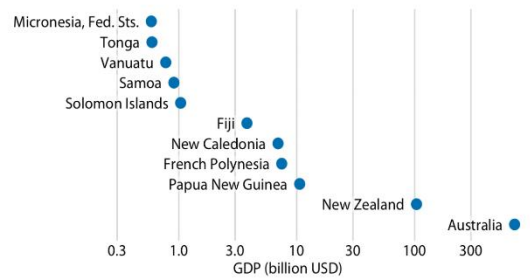


Figure 17-9. GDP in 2007 of countries in Oceania. Data source: Gapminder.

Εικόνα 8.3-19: [2] Αποφυγή χρήσης ράβδων σε λογαριθμική κλίμακα, χρήση κουκίδων

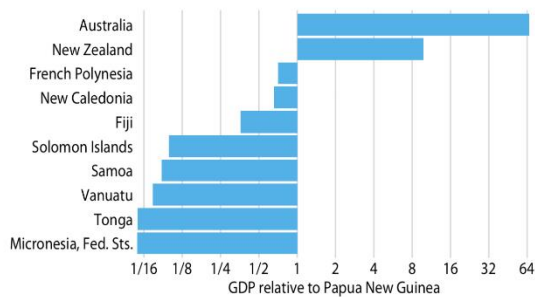


Figure 17-10. GDP in 2007 of countries in Oceania, relative to the GDP of Papua New Guinea. Data source: Gapminder.

Εικόνα 8.3-17: [2] Έγκυρη χρήση ράβδων για αναλογίες σε μεγάλη κλίμακα

Ωστόσο, η χρήση ράβδων για την οπτικοποίηση αναλογιών (ratios) σε μεγάλη κλίμακα μπορεί να αποτελέσει μια έγκυρη επιλογή, αρκεί να ξεκινούν από το «1» αντί για το «0» (Εικόνα 8.3-17).

Όταν υπάρχουν τιμές «0» στο σύνολο δεδομένων μας με τόσο μεγάλη διαφορά σε μέγεθος τιμών, μερικές φορές χρησιμοποιείται η κλίμακα τετραγωνικής ρίζας (square-root scale), αν και με αυτή υπάρχουν διάφορα προβλήματα.

[Προβλήματα όπως: τα βήματα των μονάδων (unit step) εξαρτώνται από την τιμή με την οποία ξεκινάμε. Επίσης, δεν είναι ξεκάθαρο το πως να τοποθετήσουμε με τον καλύτερο τρόπο τα σημεία στους άξονες (axis ticks)].

Ένα παράδειγμα χρήσης αυτής της κλίμακας είναι οι γεωγραφικές περιοχές, ώστε να φανεί η γραμμική έκταση της περιοχής, αν θέλαμε για παράδειγμα να μάθουμε πόσο χρόνο θα χρειαζόταν για να διασχίσουμε μια περιοχή.

Στην Εικόνα 8.3-20, ο σχετικός χρόνος οδήγησης σε όλες τις περιοχές αντιπροσωπεύεται καλύτερα από το σχήμα με την κλίμακα τετραγωνικής ρίζας παρά με τη γραμμική κλίμακα.

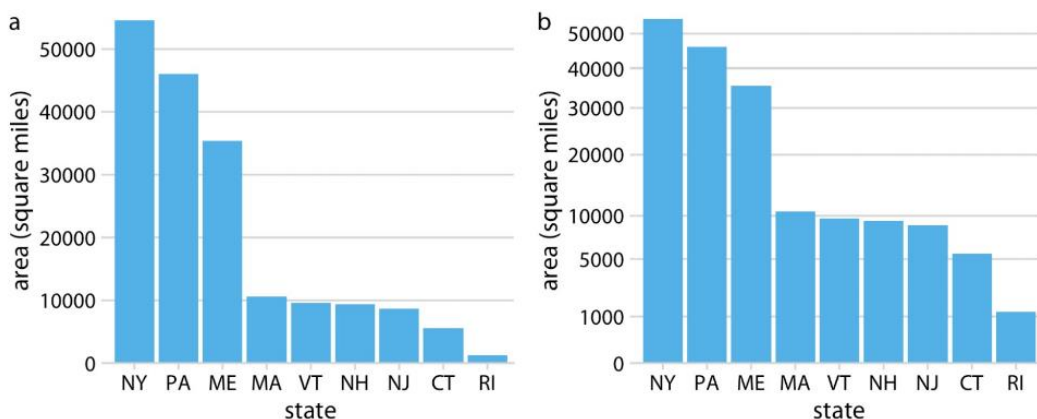


Figure 3-8. Areas of northeastern US states. (a) Areas shown on a linear scale. (b) Areas shown on a square-root scale. Data source: Google.

Εικόνα 8.3-20: [2] Χρήση κλίμακας τετραγωνικής ρίζας

Οι απεικονίσεις όπως γραφήματα πίτας αντιπροσωπεύουν τιμές δεδομένων ανά περιοχή, χωρίς αντίστοιχη χαρτογράφηση θέσης. Θα μπορούσαμε να πούμε ότι τεχνικά, χαρτογραφούνται σε γωνίες, όμως συνήθως δεν κρίνουμε τις γωνίες του γραφήματος πίτας.

Παρόλο που ένα γράφημα πίτας ικανοποιεί την αρχή της αναλογικής μελάνης, η αντίληψή μας για το εμβαδόν σε ένα γράφημα πίτας και σε ένα ραβδόγραμμα, δεν είναι η ίδια. Στα γραφήματα πίτας,

η κυρίαρχη οπτική ιδιότητα που βλέπουμε είναι το εμβαδόν της κάθε φέτας, ενώ στα ραβδογράμματα είναι η απόσταση κάθε ράβδου.

Η ανθρώπινη αντίληψη κρίνει πρωτίστως την απόσταση και όχι την περιοχή, και γι' αυτό η απόσταση του στοιχείου των ράβδων γίνεται αντιληπτή με μεγαλύτερη ακρίβεια από ότι η περιοχή μιας φέτας της πίτας (Εικόνα 8.3-22). Το ίδιο ισχύει και για έναν δεντροχάρτη (treemap), εξάλλου, μοιάζει με την τετραγωνισμένη έκδοση ενός γραφήματος πίτας (Εικόνα 8.3-21).

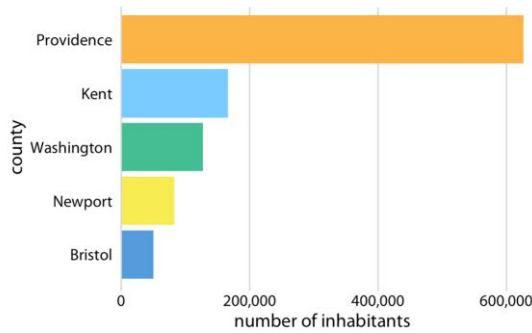


Figure 17-12. Number of inhabitants in Rhode Island counties, shown as bars. The length of each bar is proportional to the number of inhabitants in the respective county. Data source: 2010 US Decennial Census.

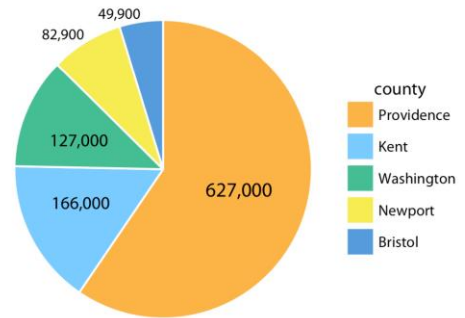


Figure 17-11. Number of inhabitants in Rhode Island counties, shown as a pie chart. Both the angle and the area of each pie wedge are proportional to the number of inhabitants in the respective county. Data source: 2010 US Decennial Census.

Εικόνα 8.3-22: Αντίληψη σχετικά με αποστάσεις και περιοχές, ραβδογραφήματα και πίτες

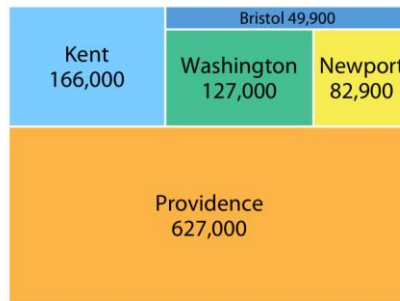


Figure 17-13. Number of inhabitants in Rhode Island counties, shown as a treemap. The area of each rectangle is proportional to the number of inhabitants in the respective county. Data source: 2010 US Decennial Census.

Εικόνα 8.3-21: [2] Αντίληψη σχετικά με περιοχές, δεντροχάρτης

9 Εργαλεία Οπτικοποίησης Δεδομένων

Όπως έχουμε ήδη αναφέρει, η επιλογή των σωστών εργαλείων που θα χρησιμοποιήσουμε για την οπτικοποίηση μας είναι πολύ σημαντική.

Σύμφωνα με την ηλεκτρονική υπηρεσία δεδομένων [Statistica](#) [12], ο συνολικός όγκος δεδομένων και πληροφοριών παγκοσμίως έχει αυξηθεί από περίπου 2 zettabyte το 2010, σε 74 το 2021 και υπολογίζεται πως μέχρι το 2024, ο αριθμός θα φτάσει τα 149 zettabyte. Ο όγκος των δεδομένων που παράγονται και καταναλώνονται από οργανισμούς, αυξάνεται με εκπληκτικό ρυθμό δημιουργώντας έτσι μια αυξανόμενη ανάγκη να γίνουν πιο εύπεπτα, κατανοητά και εφαρμόσιμα και σε αυτό στοχεύει το λογισμικό οπτικοποίησης δεδομένων. Γραφικές παραστάσεις μυριάδων σημείων δεδομένων, δίνουν τη δυνατότητα για βαθύτερη αναγνώριση σημαντικών αριθμών, τάσεων, μετρήσεων και βασικών δεικτών απόδοσης (KPI). Πίνακες ελέγχου (dashboards) και γραφικά με την κατάλληλη ανά περίπτωση μορφή (από πίτες, ραβδογράμματα, διαγράμματα διασποράς και πολικής περιοχής (polar area diagrams), θερμικοί χάρτες (heat maps), χρονοδιαγράμματα (timelines), γραφήματα δακτυλίου (ring charts), γραφήματα μήτρας (matrix charts), σύννεφα λέξεων (word clouds) και πολλοί άλλοι τύποι αναπαραστάσεων, εκτός από τους επιστήμονες δεδομένων και τα τμήματα πληροφορικής, βοηθούν επίσης επιχειρηματίες, διαφημιστές αλλά και οργανισμούς στο να κατανοήσουν καλύτερα ποικίλα σενάρια όπως το συναίσθημα, οι πωλήσεις σε πραγματικό χρόνο, οι τάσεις της υγειονομικής περίθαλψης, η συμπεριφορά των πελατών, οι στόχοι του τμήματος, η έρευνα αγοράς κτλ. Η οπτικοποίηση δεδομένων θεωρείται πλέον ένα από τα πιο κρίσιμα εργαλεία για την επιτυχία τόσο απλών, όσο και σύνθετων εταιρικών δομών.

Φυσικά, διαφορετικές λύσεις οπτικοποίησης δεδομένων προσεγγίζουν διαφορετικά μια εργασία όπως είδαμε. Ορισμένες κλίνουν προς συμβατικές λειτουργίες επιχειρηματικής ευφυΐας (BI), ενώ άλλες συνδέονται σε ζωντανά δεδομένα. Παράγοντες όπως ο τεράστιος όγκος και η πολυπλοκότητα των διαθέσιμων πληροφοριών κάνουν την επιλογή του σωστού εργαλείου μια πρόκληση.

Η μορφή της πληροφορίας καθώς και η διαθέσιμη πηγή μπορεί να παίζει σημαντικό ρόλο. Εσωτερικά δεδομένα που προέρχονται από τους ίδιους ή τον οργανισμό τους, ή εξωτερικά δεδομένα όπως αυτά που υπάρχουν διαθέσιμα στο διαδίκτυο για λήψη, ή ειδικά εξωτερικά

δεδομένα που ενδέχεται να μην μπορούν να ληφθούν απευθείας. Τέτοια είδους παραδείγματα, είναι οι λεπτομέρειες των προϊόντων από την Amazon ή tweets από το Twitter, ποικιλόμορφα δεδομένα τα οποία μπορεί (εάν δεν υπάρχει χρόνος για απόξεση) και μια υπηρεσία απόξεσης δεδομένων (data scraping service) να καθαρίσει ώστε να μπορέσουμε να τα δούμε σωστά και να τα αναλύσουμε, καθώς επίσης και ανάγκη να τα παρουσιάσουμε με ουσιαστικό τρόπο.

Αυτή είναι και η δυσκολία των επιχειρήσεων, να διαχειριστούν, να ερμηνεύσουν ή/και να οπτικοποιήσουν τους τεράστιους όγκους δεδομένων, αλλά και να εξάγουν πληροφορίες που να είναι σχετικές και κρίσιμες με αυτά. Χρειάζεται μια καλή κατανόηση των λύσεων που προσφέρονται: πως αυτές λειτουργούν (κάποιες πχ, ενσωματώνουν μηχανική μάθηση και τεχνητή νοημοσύνη για πιο προηγμένες λειτουργίες και εμπάθυνση γνώσης), ή εάν ικανοποιούν τις ανάγκες μας και το διαθέσιμο προϋπολογισμό μας. Για τη σωστή επιλογή λύσης ελέγχουμε παράγοντες όπως:

- **Τι είδους οπτικοποίηση χρειαζόμαστε;** Μπορεί να δημιουργήσει ελκυστικές και χρήσιμες αναφορές; Διαφορετικές εφαρμογές οπτικοποίησης δεδομένων παρέχουν διαφορετικούς τρόπους συγκέντρωσης και προβολής δεδομένων.
- **Σε ποιους τύπους πλατφόρμας λειτουργεί το λογισμικό και ποιες συσκευές υποστηρίζει;** Ορισμένες λύσεις βασίζονται στο σύννεφο (cloud-based) ενώ άλλες σε επιτραπέζιους υπολογιστές, ακόμη και σε κινητές συσκευές. Ορισμένα μοντέλα εσωτερικής εγκατάστασης έχουν εφαρμογές που εκτελούνται μόνο σε Windows, πράγμα που μπορεί να δημιουργήσει πρόβλημα εάν χρησιμοποιηθούν Mac.
- **Το πακέτο υποστηρίζει επαρκώς τις απαιτήσεις απόδοσης του οργανισμού μας;** Ορισμένες εφαρμογές αντιμετωπίζουν δυσκολίες με εξαιρετικά μεγάλα αρχεία και δεν έχουν καλή απόδοση. Εάν για παράδειγμα, η μηχανή απόδοσης (rendering engine) δεν μπορεί να υποστηρίξει την ταχύτητα που απαιτείται για τις ιστοσελίδες και τους πίνακες εργαλείων σε πραγματικό χρόνο (real-time dashboards), πιθανά να υπάρχει πρόβλημα.
- **Η εφαρμογή ενσωματώνεται με τις ροές εργασίας μας;** Η ευελιξία και η επεκτασιμότητα είναι συχνά ζωτικής σημασίας. Ίσως χρειάζονται αλλαγές προτύπων (templates), εισόδων (inputs) ή κριτηρίων, ακόμα και μέσω άλλων προγραμμάτων που συνδέονται μέσω API. Θα πρέπει το πρόγραμμα να μπορεί να υποστηρίξει αυτές τις αλλαγές.

- **Χρειάζονται ειδικές γνώσεις; Πώς φαίνεται η υποστήριξη προμηθευτή;** Μια εφαρμογή μπορεί να παράγει εντυπωσιακές απεικονίσεις. Ωστόσο, η δημιουργία μιας απεικόνισης μπορεί να είναι εξαιρετικά δύσκολη, οπότε θα πρέπει να βεβαιωθούμε ότι προσφέρεται σταθερή τεκμηρίωση και υποστήριξη (συμπεριλαμβανομένων βίντεο και οδηγιών) όπως και τηλεφωνική 24×7 σε περίπτωση κολλήματος.
- **Πόσο κοστίζει το πακέτο;** Θα πρέπει να εξετάσουμε προσεκτικά τις επιλογές μας, μιας και οι δωρεάν λύσεις μπορεί να μην μας παρέχουν πάντα τις δυνατότητες που χρειαζόμαστε. Κάποιες λύσεις προσφέρουν κλιμακωτές τιμές για επιχειρηματικές επιλογές.
- **Τι προστασίες ασφάλειας προσφέρει;** Θα πρέπει να βεβαιωθούμε ότι η πλατφόρμα παρέχει επαρκή προστασία για την πρόσβαση, την ασφάλεια και την κοινή χρήση δεδομένων (sharing data).

Κάποια από τα εργαλεία με εξαιρετικές δυνατότητες για οπτικοποίηση δεδομένων και ανάλυση για φορητές συσκευές θεωρούνται τα: Tableau Public, Microsoft Power BI, Google Data Studio, Openheatmap, DataWrapper και Leaflet .

- το Power BI μπορεί να οδηγήσει τις αναφορές μας στο επόμενο επίπεδο και με μια ματιά να ελέγχουμε τα πιο κρίσιμα δεδομένα μας.
- το Google Data Studio δίνει στους χρήστες την επιλογή να εργαστούν με πολλαπλές προβολές δεδομένων.
- με το Openheatmap, μπορούμε να μοιραστούμε την ιστορία μας με άλλους.
- το Google Data Studio παρέχει μια εξαιρετική εισαγωγή στην ανάλυση δεδομένων, αλλά
- μπορεί να ξεκινήσει κανείς και με το Leaflet, εάν έχει εξειδίκευση στην κωδικοποίηση.
- το Tableau (έκδοση επί πληρωμή) και το Microsoft BI μπορεί να δημιουργήσει πλούσιες, πλήρως εξοπλισμένες παρουσιάσεις για συναδέλφους. Ορισμένα εργαλεία όπως το Tableau έχουν μεγάλη γκάμα γραφημάτων/διαγραμμάτων.
- άλλα, είναι γνωστά για τη δημιουργία γραφημάτων όπως το Infogram.
- και άλλα κερδίζουν δημοτικότητα λόγω των διαδραστικών γραφημάτων όπως το Gephi.

Ακολουθεί μια λίστα γνωστών εργαλείων οπτικοποίησης δεδομένων, με πλεονεκτήματα και μειονεκτήματα για να δούμε ποιο ταιριάζει καλύτερα στις ανάγκες μας. Έχουν χωριστεί σε δύο

κατηγορίες: εργαλεία που δεν απαιτούν προγραμματισμό και εργαλεία μόνο για προγραμματιστές. Επίσης, έχουν συμπεριληφθεί (όπου υπήρχαν), εισαγωγικά link εκμάθησης λογισμικού, χαρακτηριστικών ή gallery, ως μια πιο ολοκληρωμένη εικόνα της κάθε λύσης. Μια λίστα λοιπόν πιθανών λύσεων είναι και η παρακάτω:

Εργαλεία

- **Για μη επαγγελματίες**
 - **Διαγράμματα και Γραφήματα**
 - Δωρεάν (RawGraphs, ChartBlocks)
 - Εμπορικά – Για Ιδιώτες ή Επιχειρήσεις ((Tableau, Power BI(Microsoft), QlikView, FineReport, Datawrapper, Visme))
 - Εμπορικά – Μόνο για Επιχειρήσεις (Grow))
 - **Infographics** (Infogram, Visual))
 - **Χάρτες** (Instant Atlas))
 - **Γραφήματα Δικτύου** (Gephi)
 - **Μαθηματικά Γραφήματα** (Wolfram Alpha))
- **Για προγραμματιστές**
 - **Διαγράμματα και Γραφήματα**
 - Δωρεάν (ECharts, D3.js, Plot.ly, Charts.js, Google charts, Chartlist.js))
 - Εμπορικά (High Charts, Fusion Charts, Zing Chart))
 - **Χάρτες** (Leaflet, Openlayers, Kartograph, Carto))
 - **Γραφήματα Δικτύου** (sigma)
 - **Οικονομικά Γραφήματα** (dygraphs))

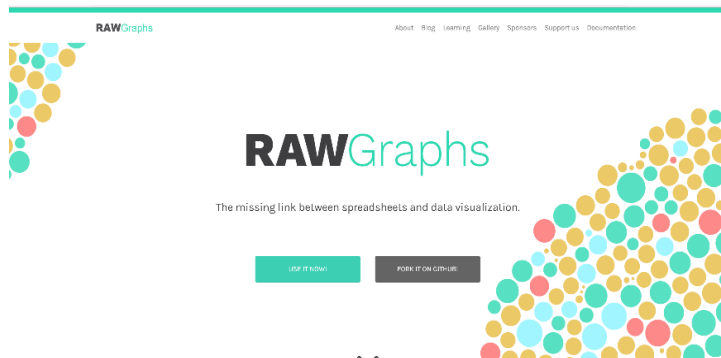
Πιο αναλυτικά:

9.1 Εργαλεία για Μη Επαγγελματίες

9.1.1 Διαγράμματα και Γραφήματα

9.1.1.1 Δωρεάν

1. **RAWGraphs**: στοχεύει στην παροχή της διασύνδεσης που λείπει μεταξύ εφαρμογών υπολογιστικών φύλλων (π.χ. Microsoft Excel, Google Spreadsheets, Apple Numbers κτλ) και διανυσματικών γραφικών επεξεργασίας (π.χ. Adobe Illustrator και Sketch).



Εικόνα 9.1.1.1-1 [13] Στιγμιότυπο οθόνης από την ιστοσελίδα του RAWGraphs

(+): *Πλεονεκτήματα:*

- δωρεάν εργαλείο ιστού ανοιχτού κώδικα (open source web tool) και πλαίσιο οπτικοποίησης δεδομένων (data visualization framework)
- εκτός από τα γνωστά, παρέχει και διαγράμματα που δεν είναι εύκολο να παραχθούν από άλλα εργαλεία.
- Δυνατότητα εξερεύνησης δεδομένων και άμεσης οπτικής ανατροφοδότησης
- ασφάλεια των δεδομένων, εφόσον η επεξεργασία γίνεται μόνο από το πρόγραμμα περιήγησης ιστού (web browser) και δε γίνονται λειτουργίες server-side.
- απλή εισαγωγή δεδομένων, προσαρμογή των γραφημάτων και εξαγωγή τους ως διανυσματικές (SVG) ή μη-διανυσματικές (raster) (PNG) εικόνες

(-): *Μειονεκτήματα:*

- Δε δίνεται δυνατότητα απευθείας διαμοίρασης στα μέσα κοινωνικής δικτύωσης

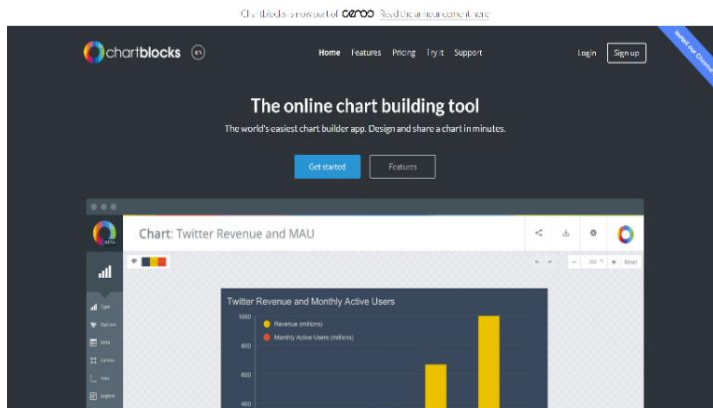
Σύνδεσμος: <https://rawgraphs.io/>

Δείγμα (εισαγωγικό βίντεο): https://www.youtube.com/watch?v=ITQnmlY_x4E

Gallery: <https://rawgraphs.io/gallery>

Κόστος: δωρεάν

2. [ChartBlocks](#) : εξαιρετικό λογισμικό οπτικοποίησης δεδομένων. Θεωρείται κατάλληλο για



επιχειρήσεις στους κλάδους της υγειονομικής περίθαλψης, των ταξιδιών, των ασφαλίσεων και του λιανικού εμπορίου κτλ

Εικόνα 9.1.1.1-2: [14] Στιγμιότυπο οθόνης από την ιστοσελίδα του ChartBlocks

(+): *Πλεονεκτήματα:*

- απλό διαδικτυακό εργαλείο δημιουργίας γραφημάτων με βήμα-βήμα οδηγό για την εισαγωγή δεδομένων και του σχεδιασμού γραφημάτων
- δεν απαιτείται κώδικας για τη δημιουργία γραφημάτων
- αντλεί δεδομένα από σχεδόν οποιαδήποτε πηγή μέσα σε λίγα δευτερόλεπτα
- δυνατότητα διαμοίρασης των γραφημάτων στα μέσα κοινωνικής
- υπάρχει μεγάλο εύρος γραφημάτων για το ίδιο αρχείο δεδομένων
- τα οπτικά στοιχεία είναι συμβατά με τις περισσότερες συσκευές και μεγέθη οθόνης
- μπορεί να γίνει και εξαγωγή τους ως επεξεργάσιμα διανυσματικά γραφικά (.eps, .pdf, .svg) ή να ενσωματωθούν σε ιστότοπους

Σύνδεσμος: <https://www.chartblocks.com/>

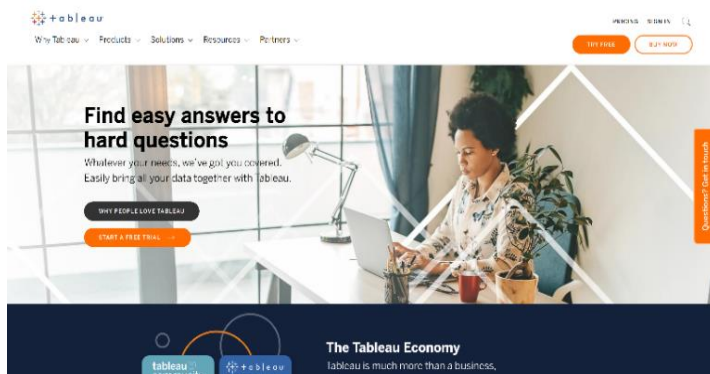
Δείγμα (εισαγωγικό βίντεο): <https://www.youtube.com/watch?v=1HPm8fIGb5w&t=26s>

Κόστος: η δημιουργία έως και 50 διαγράμματα είναι δωρεάν, πέρα από την οποία υπάρχουν πακέτα επί πληρωμή (επαγγελματικό και elite).

9.1.1.2 Εμπορικά – Για ιδιώτες ή επιχειρήσεις

Κάποια εργαλεία, παρέχουν διαφορετικά προγράμματα για μεμονωμένα άτομα , μικρές ομάδες και οργανισμούς επί πληρωμή, όπως έξτρα δυνατότητες/χαρακτηριστικά και τεχνική υποστήριξη.

3. **Tableau**: Το Tableau είναι το πιο διάσημο εργαλείο οπτικοποίησης και ανάλυσης



Εικόνα 9.1.1.2-3: [15] Στιγμιότυπο οθόνης από την ιστοσελίδα του Tableau

δεδομένων και κατατάσσεται από την Gartner και πολλούς άλλους σταθερά ως το κορυφαίο διαδραστικό εργαλείο και ηγέτης της αγοράς όταν πρόκειται να δείξει πώς η οπτικοποίηση μπορεί να βοηθήσει στη γενική διαχείριση της παραγωγικότητας και είναι μια ευέλικτη και ισχυρή

λύση που παρέχει επιχειρηματική ευφυΐα και ισχυρή γκάμα εργαλείων, ολοκληρωμένη απεικόνιση και cloud παρουσία με μεγάλα σύνολα δεδομένων. Θεωρείται κατάλληλο για επιχειρήσεις που θέλουν να διαμορφώσουν την αυτοματοποιημένη οπτικοποίηση δεδομένων. Το Tableau έχει αναδειχθεί ως ένα από τα πιο αποτελεσματικά και ταχύτερα προηγμένα εργαλεία Business Intelligence και οπτικοποίησης δεδομένων. Συνολικά είναι ένα ισχυρό εργαλείο ανάλυσης για την παραγωγή κάθε είδους γραφημάτων και γραφικών παραστάσεων, τα οποία καθιστούν ευκολότερη την επικοινωνία πληροφοριών και απαντήσεων σε ερωτήσεις με διαφορετικά επίπεδα ενδιαφερόμενων μερών.

(+): *Πλεονεκτήματα*

- γρήγορη, πολύ ισχυρή, ασφαλής και ευέλικτη πλατφόρμα ανάλυσης δεδομένων
- καλά σχεδιασμένη και εξαιρετικά διαδραστική διεπαφή (GUI) με drag and drop στοιχεία, και διαδραστικές απεικονίσεις και πίνακες ελέγχου
- ισχυρές δυνατότητες κοινής χρήσης οπτικοποιήσεων
- εύχρηστο, χωρίς να απαιτείται εμπειρία προγραμματισμού

- ενσωματωμένες δυνατότητες ανάλυσης σε πραγματικό χρόνο και υποστήριξη για φορητές συσκευές και επιτραπέζιους υπολογιστές για την ανάμιξη δεδομένων από οπουδήποτε. Υποστήριξη για διάφορους τύπους αναλύσεων δεδομένων, όπως ανάλυση μάρκετινγκ, πληροφορικής, ασφάλισης και κρατικών δεδομένων. Εύκολη ενσωμάτωση με άλλα εργαλεία ανάλυσης δεδομένων. Διαθεσιμότητα επεκτάσεων JavaScript και API για τη δημιουργία προσαρμοσμένων αναφορών με βιβλιοθήκες γραφημάτων JavaScript
- συνδέεται με εκατοντάδες πηγές και υπηρεσιών δεδομένων (από υπολογιστικά φύλλα και συμβατικές βάσεις δεδομένων έως αποθετήρια δεδομένων Hadoop και cloud) και υποστηρίζει όλες τις κύριες μορφές δεδομένων.
- εμφάνιση δεδομένων με πολλές μορφές, διαθέσιμη on-prem ή στο cloud.
- οι χρήστες μπορούν να οπτικοποιήσουν και να μοιραστούν δεδομένα με ένα ξεχωριστό εργαλείο γνωστό ως Tableau Public, παρέχοντας κάποιες δυνατότητες δωρεάν.

(-): Μειονεκτήματα

- Ακριβό (με εξαίρεση το Tableau Public που είναι δωρεάν)
- Μικτές κριτικές σχετικά με την υποστήριξη πελατών
- Μπορεί να απαιτείται εκπαίδευση για τη χρήση του πλήρους συνόλου λειτουργιών και δυνατοτήτων στην πλατφόρμα
- Δύσκολο να γίνει προσαρμογή (difficult to customize)
- Η Public έκδοση δεν επιτρέπει τη διατήρηση ιδιωτικών αναλύσεων δεδομένων

Tableau Public: Είναι το απλούστερο και καλύτερο εργαλείο οπτικοποίησης δεδομένων επειδή επιτρέπει στους χρήστες να δημιουργούν εκπληκτικές απεικονίσεις χωρίς να χρειάζεται να γνωρίζουν προγραμματισμό. Το Tableau Public είναι η ίδια έκδοση του Tableau, αλλά χωρίς τη συνδρομή, καθώς δεν παρέχει το απόρρητο που λαμβάνουμε με την πληρωμένη έκδοση. Ό,τι δημιουργείται με αυτό το εργαλείο είναι διαθέσιμο για δημόσιο έλεγχο και είναι αυτόματα διαθέσιμο στον Ιστό. Σε σύγκριση με το Microsoft

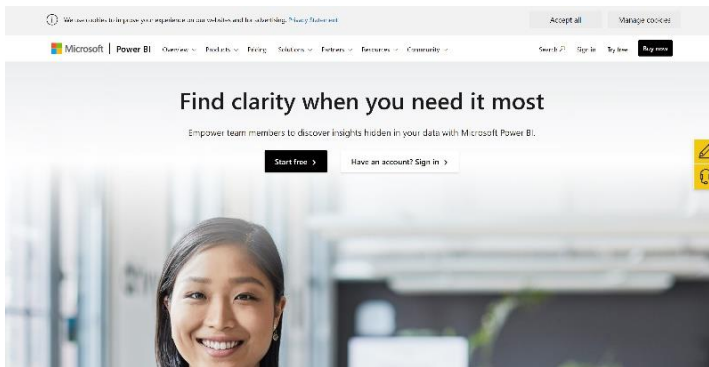
Excel, το Tableau είναι ένα εξαιρετικό εργαλείο για τη δημιουργία οπτικοποιήσεων και την αφήγηση διαδραστικών ιστοριών στον Ιστό και είναι ιδανικός βοηθός για κάθε είδους επιχείρηση που συλλέγει δεδομένα για να βοηθήσει στη λήψη επιχειρηματικών αποφάσεων.

Σύνδεσμος: <https://www.tableau.com/>

Δείγμα (*intro βίντεο*): <https://www.tableau.com/learn/get-started/creator#content-209707>

Κόστος: είναι 70 \$/μήνα/χρήστη για το λογισμικό Tableau Creator, ενώ υπάρχει δυνατότητα και για άλλες [λύσεις](#).

4. **Power BI**: Το Power BI είναι μια σουίτα εργαλείων επιχειρηματικής ανάλυσης της



Εικόνα 9.1.1.2-4: [16] Στιγμιότυπο οθόνης από την ιστοσελίδα του Power BI

Microsoft επομένως καλή ενσωμάτωση με το Office και οικεία εργαλεία και διαισθητικές λειτουργίες που διευκολύνουν τη δημιουργία αναφορών, τη συνεργασία σε πίνακες ελέγχου. Είναι παρόμοιο με το εργαλείο Visualization Data BI του Excel

αλλά με πολύ περισσότερες δυνατότητες (χρήση πολλαπλών πηγών δεδομένων, λογική τιμή και μπορεί να χρησιμοποιηθεί ως ξεχωριστό εργαλείο BI). Μοναδικό καθώς επιτρέπει στους χρήστες να κάνουν προετοιμασία δεδομένων, ανακάλυψη δεδομένων και πίνακες ελέγχου με το ίδιο εργαλείο σχεδίασης. Η πλατφόρμα ενσωματώνεται με το Excel και το Office 365 και έχει μια πολύ ενεργή κοινότητα χρηστών που επεκτείνει τις δυνατότητες του εργαλείου

(+): *Πλεονεκτήματα:*

- απλότητα χρήσης, δημιουργία και διαμοίραση εύχρηστων αναφορών και εξαγωγή κάθε είδους πληροφορίας (πχ εργαλείο αναζήτησης). Επιτρέπει στους χρήστες να εισάγουν με επιτυχία βασικά δεδομένα, δημιουργώντας αναφορά με οπτικοποίηση
- μεμονωμένοι υπάλληλοι και ομάδες μπορούν εύκολα να επεξεργάζονται, να προβάλλουν, να προσθέτουν σχόλια και να διαχειρίζονται εκδόσεις πινάκων εργαλείων και αναφορών. Επιτρέπει την ανάλυση δεδομένων και τη λήψη επιχειρηματικών αποφάσεων, και επιτρέπει τη διαμοίραση των δεδομένων με ασφάλεια δημοσιεύοντάς τα στην υπηρεσία Power BI με ένα κλικ
- δημιουργία και τροποποίηση εντυπωσιακών διαδραστικών απεικονίσεων δεδομένων. Συνδυασμός δεδομένων από πολλαπλές πηγές (Excel, Salesforce, Google Analytics, Social Analytics, ακόμα και συσκευές ΙLT). Επιτρέπει την εισαγωγή αρχείων, φακέλων και βάσεων δεδομένων και την προβολή τους οπουδήποτε (οπτικοποίηση και διαμοίραση δεδομένων στο γραφείο ή εν κινήσει)

- δωρεάν για μεμονωμένους χρήστες

(-): Μειονεκτήματα:

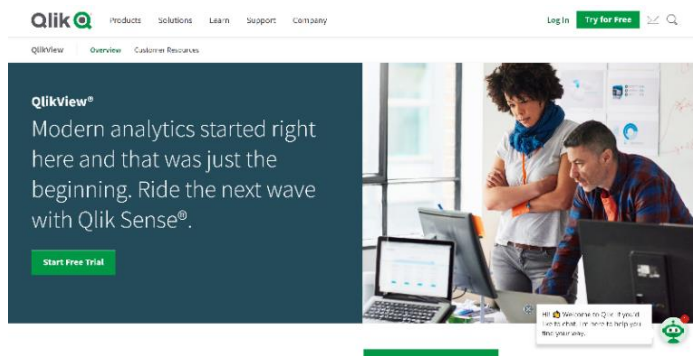
- δεν υπάρχει τρόπος να ενσωματωθεί με υπάρχοντα συστήματα.

Σύνδεσμος: <https://powerbi.microsoft.com/>

Δείγματα (*intro βίντεο*): <https://powerbi.microsoft.com/en-us/demo/> , (best practices): https://www.youtube.com/watch?v=2QMhm_5HVyo&t=62s

Κόστος: δωρεάν για μεμονωμένους χρήστες. Επίσης, δίνονται λύσεις όπως η Pro με χρέωση 9,9 \$ και premium έκδοση

5. **QlikView:** Το εργαλείο QlikView είναι ένα εργαλείο επιχειρηματικής ευφυΐας που εστιάζει



Εικόνα 9.1.1.2-5: [17] Στιγμιότυπο οθόνης από την ιστοσελίδα του QlikView

κυρίως σε επιχειρηματικούς χρήστες σε οργανισμούς και οι χρήστες μπορούν εύκολα να αναλύσουν τα δεδομένα τους, να τα κατανοήσουν και να χρησιμοποιήσουν αναλύσεις και δυνατότητες εταιρικής αναφοράς του QlikView για να υποστηρίξουν τη λήψη αποφάσεων. Παρέχει επίσης μια προσωπική έκδοση, ώστε μεμονωμένοι χρήστες να μπορούν να απολαμβάνουν τις ισχυρές δυνατότητες του. Μπορείτε απλά να πληκτρολογήσετε τις λέξεις-κλειδιά που θέλετε να αναζητήσετε μέσα στο σύνολο δεδομένων και το QlikView μπορεί να σας βοηθήσει να βρείτε απροσδόκητες πληροφορίες και συσχετίσεις δεδομένων.

Σύνδεσμος: <https://www.qlik.com/us/products/qlikview>

Δείγμα (intro βίντεο): <https://www.youtube.com/watch?v=j9721wS0UYE>

Κόστος: η Qlik Sense® Business λύση στα 30 \$ ανά χρήστη το μήνα (τιμολογείται ετησίως) και υπάρχουν και άλλες [λύσεις](#).

6. **FineReport**: Το FineReport είναι λογισμικό για αναφορές και πινάκων ελέγχου με



Εικόνα 9.1.1.2-6: [18] Στιγμιότυπο οθόνης από την ιστοσελίδα του FineReport

εκπληκτικά διαδραστικά εφέ οπτικοποίησης. Παρέχει εντυπωσιακά γραφήματα HTML5 (self-developed HTML5 charts), τα οποία μπορούν να εμφανιστούν ομαλά σε οποιονδήποτε ιστότοπο ή ιστοσελίδα, με εντυπωσιακά 3D και δυναμικά εφέ. Η

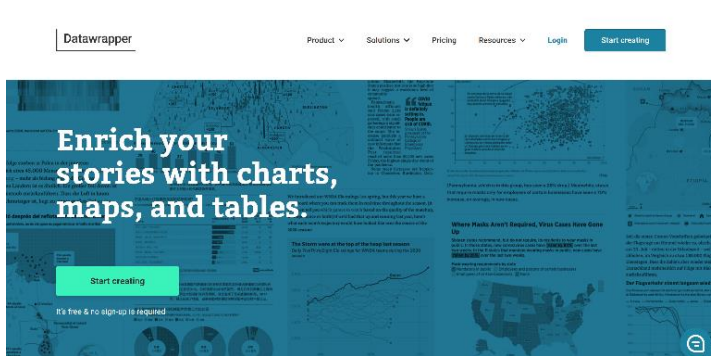
οπτικοποίηση προσαρμόζεται σε οποιοδήποτε μέγεθος οθόνης, από τηλεοράσεις και μεγάλες οθόνες έως φορητές συσκευές. Εύκολες λειτουργίες drag-and-drop επιτυγχάνουν όλα τα εφέ.

Σύνδεσμος: <https://www.finereport.com/>

Δείγμα (intro βίντεο): <https://www.youtube.com/watch?v=Xd9HV03LDAg&t=3s>

Κόστος: το FineReport είναι ΔΩΡΕΑΝ για τους μεμονωμένους χρήστες (αριθμός ταυτόχρονων χρηστών αυτής της έκδοσης μέχρι 2). Επίσης υπάρχει και εμπορική έκδοση.

7. [Datawrapper](#): Το [Datawrapper](#) είναι ένα εύχρηστο διαδικτυακό εργαλείο για τη δημιουργία



Εικόνα 9.1.1.2-7: [19] Στιγμιότυπο οθόνης από την ιστοσελίδα του [Datawrapper](#)

πλήρως αποκριτικών (responsive) γραφημάτων ή χαρτών που δημιουργήθηκε με σκοπό να υποστηρίξει τη δημοσιογραφία η οποία βασίζεται σε δεδομένα (δημιουργικούς συγγραφείς, bloggers, δημοσιογράφους), για

την παραγωγή εξαιρετικών γραφικών ώστε να συνοδεύουν τα άρθρα τους και να απεικονίσουν καλύτερα το περιεχόμενό τους. Επίσης είναι κατάλληλο και για άλλους επαγγελματίες που αναζητούν παρόμοιες πληροφορίες δεδομένων, φοιτητές και μικρές επιχειρήσεις που αναζητούν βασική οπτικοποίηση δεδομένων. Επιτρέπει την αντιγραφή δεδομένων από Excel ή Φύλλων Google ή το ανέβασμα αρχείων CSV ή τη σύνδεση με μια διεύθυνση URL για ζωντανή ενημέρωση

(+) *Πλεονεκτήματα:*

- υπάρχει και δωρεάν έκδοση χωρίς να απαιτείται εγγραφή (εξαγωγή γραφικών σε μορφή PNG)
- καλή απόκριση γραφημάτων και χαρτών ακολουθώντας απλή διαδικασία, με υψηλή ανάλυση ακόμα και για εκτύπωση.
- ωραία διεπαφή χρήστη, διαδραστικό, χωρίς να απαιτείται κώδικας, με ενσωματωμένο ελεγκτή αχρωματοψίας
- μεγάλη ποικιλία προτύπων (templates) ή θεμάτων (themes). Εύκολη δημιουργία διαφορετικών τύπων (ντόνατ, πίτας, χαρτογράμματα (cartograms), χάρτες συμβόλων (symbol maps), απλούς χάρτες εντοπισμού, σκιασμένους ανάγλυφους χάρτες, περίπου χίλιοι προφορωμένοι χάρτες, πίνακες για δημιουργία γραφικών, εξερεύνηση ακατέργαστων δεδομένων για συνδυασμό πολλών τύπων γραφικών μαζί κτλ). Με δυνατότητα προσαρμογής
- πολλά άρθρα για τη δημιουργία καλύτερων γραφημάτων καθώς και διαδικτυακά μαθήματα εκμάθησης

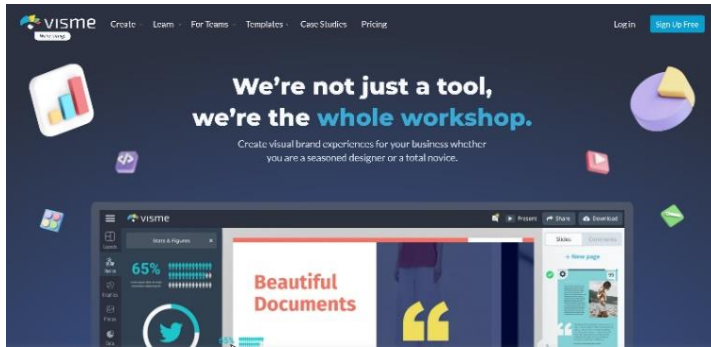
- μπορεί να ενσωματωθεί με υπάρχον CMS και διαθέτει ισχυρό API. Ειδικά σχεδιασμένο για οπτικοποίηση δεδομένων newsroom (newsroom Data Visualization).
- δημιουργία γραφημάτων και χαρτών μέσα σε λίγα λεπτά, με αυτόματη αποθήκευση όταν ολοκληρωθούν οι αλλαγές. Εύκολη επεξεργασία και σχολιασμός

Σύνδεσμος: <https://www.datawrapper.de/>

Δείγμα (*demo*): δωρεάν χρήση του πακέτου (link στην αρχική τους σελίδα, χωρίς απαραίτητη σύνδεση)

Κόστος: Η δωρεάν έκδοση είναι κατάλληλη για μικρούς ιστότοπους και επιτρέπει 1 χρήστη, 1 χάρτη εντοπισμού καθώς και 10.000 προβολές γραφήματος ανά μήνα (με μια διακριτική αναφορά στο κάτω μέρος "Δημιουργήθηκε με το Datawrapper" και εξαγωγή σε μορφή png). Επίσης υπάρχουν [πακέτα](#) επί πληρωμή για πρόσθετους χρήστες και πιο προηγμένες δυνατότητες

8. [Visme](#): Το [Visme](#) είναι ένα εργαλείο οπτικής επικοινωνίας all-in-one που επιτρέπει στους



Εικόνα 9.1.1.2-8: [20] Στιγμιότυπο οθόνης από την ιστοσελίδα του Visme

χρήστες να δημιουργούν παρουσιάσεις, infographics, αναφορές, φυλλάδια, email, ηλεκτρονικά βιβλία, ενσωματωμένα βίντεο, κινούμενες εικόνες και γραφικά μέσω κοινωνικής δικτύωσης και πολλά άλλα. Επίσης τα γραφήματα μπορούν εύκολα να

κοινοποιηθούν, να ληφθούν ή να ενσωματωθούν.

(+): Πλεονεκτήματα:

- παρέχει έναν τεράστιο όγκο φωτογραφιών, εικονιδίων, προτύπων (templates) και γραμματοσειρών στους χρήστες για παρουσιάσεις, γραφικά, διαγράμματα και αναφορές. Διαθέτει περισσότερους από 40 διαφορετικούς τύπους διαγραμμάτων (ραβδογράφημα, διάγραμμα πίτας, διάγραμμα ροής, ιστογράμματα και πολλά άλλα, καθώς και ενσωμάτωση ζωντανών δεδομένων (live data integration). Εύχρηστο εργαλείο για δημιουργία γραφημάτων.
- ενσωματώνεται με τις περισσότερες μεγάλες εφαρμογές και πηγές δεδομένων. Ενσωματώνεται με Slack, YouTube, Vimeo, Dropbox, Google Drive, SurveyMonkey, Mailchimp, Google Maps και πολλά άλλα προϊόντα και υπηρεσίες
- δημιουργία γραφικών μέσω κοινωνικής δικτύωσης (π.χ. Instagram, LinkedIn, YouTube) με δυναμικές εικόνες και δεδομένα σε πραγματικό χρόνο δημιουργούνται και σε 3 βήματα
- προβολή και παρουσίαση περιεχομένου οποιαδήποτε στιγμή από οπουδήποτε και σε οποιαδήποτε συσκευή
- προσφέρει εξαιρετικά μαθήματα και άλλο εκπαιδευτικό υλικό
- παρέχει ισχυρά χαρακτηριστικά συνεργασίας
- πολύ καλή υποστήριξη πελατών

(-): *Μειονεκτήματα:*

- Ορισμένοι χρήστες παραπονιούνται ότι τα ίδια γραφικά εμφανίζονται συχνά σε διαφορετικές εταιρείες και ιστότοπους
- Μπορεί να είναι δύσκολο να μάθεις το πρόγραμμα
- Η διεπαφή μπορεί να είναι αργή μερικές φορές
- Κάποια παράπονα για σφάλματα
- Μόνο πιο ακριβά προγράμματα έχουν βασικές ρυθμίσεις απορρήτου

Σύνδεσμος: <https://www.visme.co/>

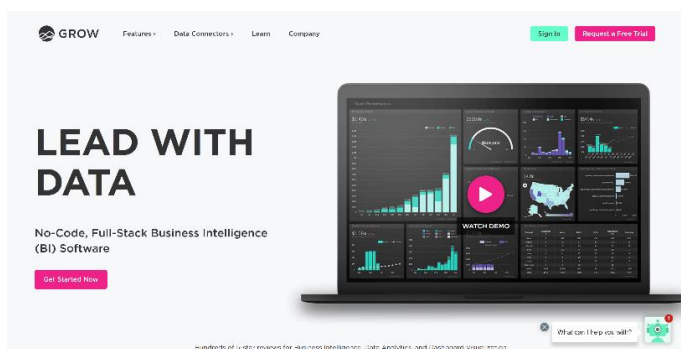
Δείγμα (εισαγωγικό βίντεο): https://www.youtube.com/watch?v=DyZz_jyKn9k

Gallery: <https://www.visme.co/templates/>

Κόστος: προσφέρει μια βασική δωρεάν υπηρεσία () και κλιμακωτά [προγράμματα](#). Επίσης, παρέχει εκπτώσεις εκπαίδευσης και εκπτώσεις μη κερδοσκοπικού χαρακτήρα.

9.1.1.3 Εμπορικά - Μόνο για επιχειρήσεις:

9. [Grow](#): Το Grow είναι ένα εργαλείο επιχειρηματικής ευφυΐας μόνο για επαγγελματίες



Εικόνα 9.1.1.3-9: [21] Στιγμιότυπο οθόνης από την ιστοσελίδα του Grow

χρήστες. Με το Grow, όλοι σε μια εταιρεία μπορούν να παρακολουθούν σημαντικά δεδομένα και να δημιουργούν τους δικούς τους ειδικούς πίνακες ελέγχου, για να λαμβάνουν τεκμηριωμένες αποφάσεις, υποστηρίζοντας περισσότερες από 150 [ενσωματώσεις](#). Η Grow

ισχυρίζεται ότι μπορεί, σε σχέση με τους ανταγωνιστές, να υλοποιήσει 8 φορές πιο γρήγορα με μια απρόσκοπτη διαδικασία.

(+): *Πλεονεκτήματα:*

- υποστηρίζονται περισσότερες από 300 προκατασκευασμένες αναφορές και ενημερώσεις δεδομένων σε πραγματικό χρόνο
- προσφέρει ένα πλήρες εργαλείο επιχειρηματικής ευφυΐας και οπτικοποίησης δεδομένων, χωρίς κώδικα.
- διαθέτει δυνατότητες ενοποίησης δεδομένων που επιτρέπουν στους χρήστες να συνδέονται, να αποθηκεύουν και να συνδυάζουν δεδομένα από εκατοντάδες πηγές δεδομένων.
- παρέχει τη δυνατότητα για συνδυασμό και μετασχηματισμό διαφορετικών πηγών δεδομένων, ώστε να μπορούν να φιλτράρονται, να τεμαχίζονται και να είναι εύκολο να εξερευνηθούν διαφορετικές οπτικοποιήσεις. Ο ενσωματωμένος εξερευνητής δεδομένων καθορίζει τον τρόπο πλοήγησης στα δεδομένα μέσω διαγραμμάτων και γραφημάτων που εμφανίζονται σε μετρήσεις και πίνακες ελέγχου.

Σύνδεσμος: <https://www.grow.com/>

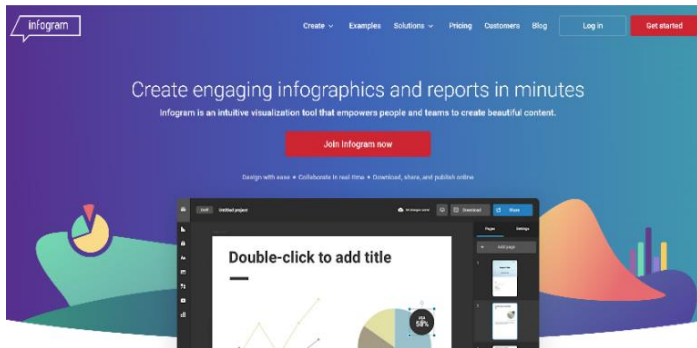
Δείγμα (intro βίντεο): στην αρχική σελίδα

Κόστος: δεν αναφέρεται (υπάρχει «ζητήστε μια δωρεάν δοκιμή»)

9.1.2 Infographics

Είναι πολύ βολικό και εύκολο να δημιουργούμε *infographics* με εργαλεία οπτικοποίησης και να μπορούμε να δημιουργούμε και δικά μας *infographics* κοινωνικών δικτύων

10. [Infogram](#): Το [Infogram](#) είναι πολύ ισχυρό και δημοφιλές εργαλείο που επιτρέπει τη



Εικόνα 9.1.2-10: [22] Στιγμιότυπο οθόνης από την ιστοσελίδα του Infogram

δημιουργία *infographics*, γραφημάτων, και χαρτών διαθέσιμα για λήψη, κοινή χρήση, ή δημοσίευση στο διαδίκτυο. Είναι ένα εργαλείο μάρκετινγκ και πολυμέσων που βασίζεται στο σύννεφο.

(+): *Πλεονεκτήματα:*

- Διαισθητική και εύκολη στη χρήση διεπαφή για χρήστες διαφορετικών επιπέδων δεξιοτήτων
- προσφέρει μια μεγάλη και ποικίλη συλλογή προτύπων, διαδραστικών γραφημάτων, χαρτών και κινούμενων εικόνων, αναφορών, πινάκων ελέγχου, στοιχείων για τα μέσα κοινωνικής δικτύωσης όπως Facebook, LinkedIn και Pinterest. Διαθέτει προσαρμόσιμα πρότυπα
- υποστηρίζει περισσότερους από 35 τύπους διαδραστικών μορφών οπτικοποίησης δεδομένων
- παρέχει drag-and-drop διεπαφή, συνεργασία σε πραγματικό χρόνο (real-time collaboration)
- Ενσωματώνεται καλά με το Google Drive, το OneDrive και το Dropbox
- Ισχυρά χαρακτηριστικά συνεργασίας για ομάδες (μπορούν να οργανωθούν σε γκρουπ και να οριστούν δικαιώματα ανά ρόλο)
- διαδραστικό περιεχόμενο όπως: ρύθμιση αντικειμένων για μεγέθυνση (zoom), αναπήδηση(bounce), αναστροφή(flip) και εξασθένιση (fade), καθώς και

βελτιωμένη διαδραστικότητα (tooltips, tabs), υπομνημάτων με δυνατότητα κλικ και διασύνδεσης.

(-): *Μειονεκτήματα:*

- Ορισμένοι χρήστες αναφέρουν κάποια σφάλματα και κολλήματα (bugs and crashes)
- Δεν υπάρχει δυνατότητα εργασίας σε έργα εκτός σύνδεσης
- Το δωρεάν πρόγραμμα δεν επιτρέπει προσαρμογές και λήψεις αρχείων σε συστήματα και συσκευές, μόνο δημοσίευση στο δίκτυο (αρκετά περιορισμένο). Πιο προηγμένα χαρακτηριστικά και σχέδια μπορεί να είναι ακριβά

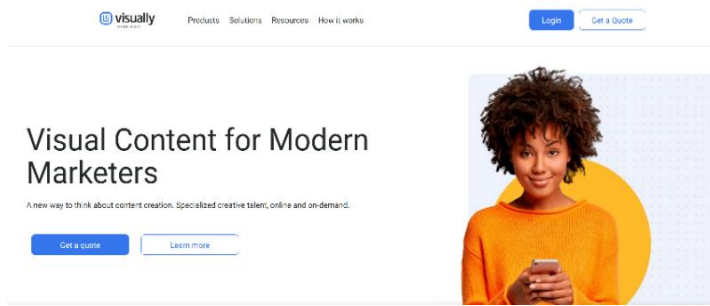
Σύνδεση: <https://infogram.com/>

Δείγμα (εισαγωγικό βίντεο): <https://www.youtube.com/watch?v=Wgg0My-rZnc>

Παραδείγματα: <https://infogram.com/examples>

Κόστος: Πέρα από τη βασική δωρεάν έκδοση, υπάρχουν άλλα τέσσερα [πακέτα](#) για δημιουργούς, μικρομεσαίες επιχειρήσεις και μεγάλες επιχειρήσεις

11. [Visual](#).: Με το Visual.ly, μπορείτε εύκολα να δημιουργήσετε infographics, βίντεο,



Εικόνα 9.1.2-11: [23] Στιγμιότυπο οθόνης από την ιστοσελίδα του Visual

αναφορές και ebook για τις καμπάνιες μάρκετινγκ. Υπάρχουν πολλά εμπνευσμένα και ενημερωτικά infographics που ανεβάζουν οι χρήστες στο Visual.ly και μπορείτε να δημιουργήσετε αυτόματα τα δικά σας infographics

κοινωνικού δικτύου μέσω του Visual.ly, το οποίο είναι πολύ δημοφιλές μεταξύ αυτών των ενεργών χρηστών κοινωνικών δικτύων.

Σύνδεσμος: <https://visual.ly/>

Παραδείγματα: <https://visual.ly/m/design-portfolio/>

Κόστος: κατόπιν επικοινωνίας

9.1.3 Χάρτες

Οι αναλυτές μπορούν να χρησιμοποιήσουν το λογισμικό ως υπηρεσία (SaaS) για να δημιουργήσουν γεωγραφικές αναφορές χωρίς να γράψουν κώδικα.

12. [InstantAtlas](#): Το InstantAtlas είναι ένα SaaS που παρέχει οπτικές αναφορές με



Εικόνα 9.1.3-12: [24] Στιγμιότυπο οθόνης από την ιστοσελίδα του InstantAtlas

επαγγελματικές υπηρεσίες υποστήριξης. Επιτρέπει σε αναλυτές πληροφοριών και ερευνητές να δημιουργούν διαδραστικές δυναμικές αναφορές χαρτών που συνδυάζουν στατιστικά και δεδομένα χαρτών για τη βελτιστοποίηση της

οπτικοποίησης δεδομένων.

Σύνδεσμος: <https://www.instantatlas.com/>

Δείγμα: <https://www.youtube.com/watch?v=GkAdB33-2HA>

Παραδείγματα:

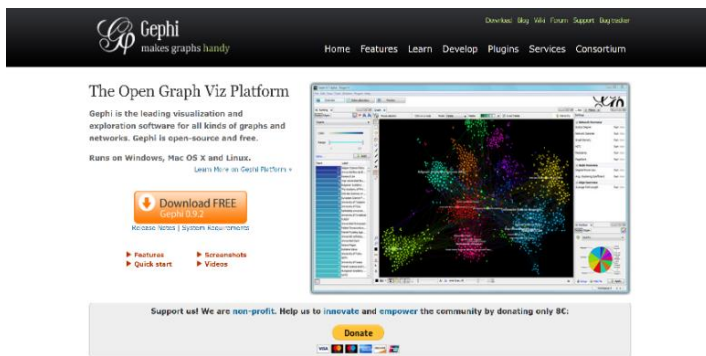
<https://bureau.maps.arcgis.com/apps/instant/filtergallery/index.html?appid=7faf589933a44e09b3c047f6aaf378bb>

Κόστος: κατόπιν επικοινωνίας

9.1.4 Γραφήματα Δικτύου (Network graphs)

Κατά την οπτικοποίηση γραφημάτων κοινωνικών δικτύων, θα πρέπει να επιλέγεται ειδικό εργαλείο οπτικοποίησης δεδομένων για τη δημιουργία περίπλοκων κόμβων και φύλλων (*nodes and leaves*).

13. [Gephi](#): Το Gephi είναι κορυφαίο λογισμικό οπτικοποίησης και εξερεύνησης για όλα τα είδη



Εικόνα 9.1.4-13: [25] Στιγμιότυπο οθόνης από την ιστοσελίδα του Gephi

γραφημάτων και δικτύων και επεξεργάζεται σχεσιακά δεδομένα (για παράδειγμα, στα μέσα κοινωνικής δικτύωσης, ποιος ακολουθεί ποιον. στις εκλογές ποιος ψήφισε ποιον σε έναν οργανισμό, ποιος έχει σχέση συνεργασίας με ποιον κτλ). Είναι ένα εργαλείο για

αναλυτές δεδομένων και επιστήμονες που θέλουν να εξερευνήσουν και να κατανοήσουν γραφήματα. Ο χρήστης αλληλεπιδρά με την αναπαράσταση, χειρίζεται τις δομές, τα σχήματα και τα χρώματα για να αποκαλύψει κρυμμένα μοτίβα. Ο στόχος είναι να βοηθηθούν οι αναλυτές δεδομένων να κάνουν υποθέσεις, να ανακαλύψουν διαισθητικά μοτίβα, να απομονώσουν ιδιαιτερότητες δομών ή σφάλματα κατά την προμήθεια δεδομένων. Ένα συμπληρωματικό εργαλείο των παραδοσιακών στατιστικών, ένα λογισμικό για διερευνητική σήμανση δεδομένων (Exploratory Data Analysis). Το Gephi είναι ανοιχτού κώδικα και λειτουργεί σε Windows, Mac OS X και Linux. (επεξηγηματική ανάλυση δεδομένων, ανάλυση συνδέσμων, ανάλυση κοινωνικών δικτύων, ανάλυση λογικού δικτύου, οπτικοποίηση σε πραγματικό χρόνο, δυναμικό φιλτράρισμα, επεκτάσιμο)

Σύνδεσμος: <https://gephi.org/>

Δείγμα 1(download): link στην αρχική τους σελίδα για κατέβασμα του λογισμικού

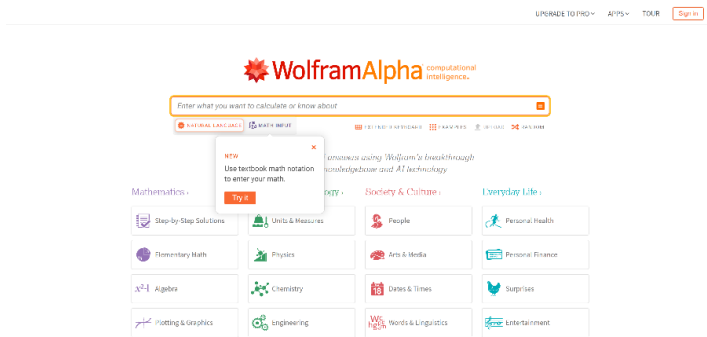
Δείγμα 2(intro βίντεο): <https://www.youtube.com/watch?v=TFBkAO1MjnU>

Κόστος: Δωρεάν

9.1.5 Μαθηματικά Γραφήματα

Τα μαθηματικά γραφήματα χρησιμοποιούνται ευρέως στην εκπαίδευση και μπορεί κανείς εύκολα να τα δημιουργήσει στο διαδίκτυο.

14. [Wolfram|Alpha](#): Το Wolfram|Alpha είναι γνωστό ως μηχανή υπολογιστικής γνώσης - κάνει



Εικόνα 9.1.5-14: [26] Στιγμιότυπο οθόνης από την ιστοσελίδα του Wolfram|Alpha

δυναμικούς υπολογισμούς και επιστρέφει το σχετικό οπτικό γράφημα. Το Wolfram|Alpha βασίζεται στο Mathematica και οι υποκείμενες λειτουργίες και η επεξεργασία δεδομένων γίνονται από το Mathematica που εκτελείται στο παρασκήνιο.

Το Mathematica υποστηρίζει γεωμετρικούς, αριθμητικούς και συμβολικούς υπολογισμούς και διαθέτει ισχυρές γραφικές απεικονίσεις επιστημονικών και μαθηματικών γραφικών. Επομένως, το Wolfram|Alpha μπορεί να απαντήσει σε μια μεγάλη ποικιλία μαθηματικών προβλημάτων και να παρουσιάσει τις απαντήσεις στον χρήστη σε καθαρά και όμορφα γραφήματα. Μπορείτε επίσης να κάνετε αναβάθμιση σε Wolfram|Alpha Pro και να ανεβάσετε τα δεδομένα και τις εικόνες σας για ανάλυση.

Σύνδεσμος: <http://www.wolframalpha.com/>

Δείγμα (εισαγωγικό βίντεο): <https://www.youtube.com/watch?v=HzFG-DHWEA&list=PLcB6CHyaK7mWocok3bi4g5B5RBId0Y30X&index=2>

Κόστος: εκτός από τη δωρεάν έκδοση, υπάρχουν και πακέτα για διαφορετικές χρήσεις

9.2 Εργαλεία για Προγραμματιστές

Οι προγραμματιστές ενσωματώνουν δεδομένα και μετατρέπουν αυτά τα δεδομένα σε όμορφα και περίπλοκα διαγράμματα και γραφήματα χρησιμοποιώντας βιβλιοθήκες διαγραμμάτων. Υπάρχουν εκατοντάδες βιβλιοθήκες που είναι διαθέσιμες στο διαδίκτυο και οι περισσότερες από αυτές υλοποιούνται σε JavaScript για προβολή στον ιστό και σε φορητές συσκευές

9.2.1 Διαγράμματα και Γραφήματα

9.2.1.1 Δωρεάν

15. [ECharts](#): Το Echarts, συντομογραφία του “Enterprise Charts”, είναι μια βιβλιοθήκη



οπτικοποίησης ανοιχτού κώδικα που υλοποιείται σε JavaScript. Διαθέτει πολλές καινοτόμες λειτουργίες που βελτιώνουν σημαντικά την εμπειρία χρήστη, δίνοντας στους χρήστες τη δυνατότητα εξόρυξης και ενσωμάτωσης δεδομένων.

Εικόνα 9.2.1.1-15: [27] Στιγμιότυπο οθόνης από την ιστοσελίδα του ECharts|

(+): Πλεονεκτήματα:

- επανυπολογισμός δεδομένων μέσω drag-and-drop
- προβολή δεδομένων, ευέλικτοι τύποι γραφημάτων
- ισχυρή μηχανή απόδοσης
- επαγγελματική Ανάλυση Δεδομένων (φιλτράρισμα, ομαδοποίηση και παλινδρόμηση, για να βοηθήσετε στην ανάλυση της πολυδιάστατης ανάλυσης των ίδιων δεδομένων)
- κομψή οπτική σχεδίαση, προσβάσιμο
- ενεργή κοινότητα

Σύνδεσμος: <https://echarts.apache.org/>

Παραδείγματα: <https://echarts.apache.org/examples/en/index.html>

Κόστος: Δωρεάν

16. [D3.js](#): Το [D3.js](#) είναι μια ευέλικτη και ισχυρή βιβλιοθήκη οπτικοποίησης δεδομένων που



Εικόνα 9.2.1.1-16: [28] Στιγμιότυπο οθόνης από την ιστοσελίδα του D3.js

βασίζεται σε Javascript, ανοιχτού κώδικα, για χειρισμό εγγράφων βασισμένα σε δεδομένα, χρησιμοποιώντας HTML, CSS και SVG. Είναι κατάλληλο για επιχειρήσεις που θέλουν να δημιουργήσουν ιστότοπους και εφαρμογές οι

οποίες βασίζονται σε δεδομένα. Θεωρείται το κορυφαίο JavaScript πλαίσιο οπτικοποίησης (visualization framework). Μπορεί η εκμάθηση του D3.js να είναι περίπλοκη, αλλά είναι πραγματικά ισχυρό και ευέλικτο και αξίζει τον κόπο σε βάθος χρόνου. Το D3.js δεν μπορεί να εμφανίσει γραφήματα σε χαμηλότερες εκδόσεις προγραμμάτων περιήγησης.

(+): *Πλεονεκτήματα:*

- προσφέρει εξαιρετική ευελιξία, εκθέτοντας τις πλήρεις δυνατότητες των προτύπων ιστού όπως HTML, SVG και CSS.
- Με ελάχιστη επιβάρυνση, το D3 είναι εξαιρετικά γρήγορο, υποστηρίζοντας μεγάλα σύνολα δεδομένων και δυναμικές συμπεριφορές για αλληλεπίδραση και κινούμενα σχέδια
- Προσφέρει σπάνιες λειτουργίες όπως δημιουργία καμπύλης (curve-generation), συσχέτιση δεδομένων με κάποιο στοιχείο HTML και δυνατότητα επαναχρησιμοποίησης κώδικα.
- Υποστήριξη για χειρισμό Document Object Model (DOM).
- Δεν απαιτεί τη φόρτωση ολόκληρης της βιβλιοθήκης κάθε φορά

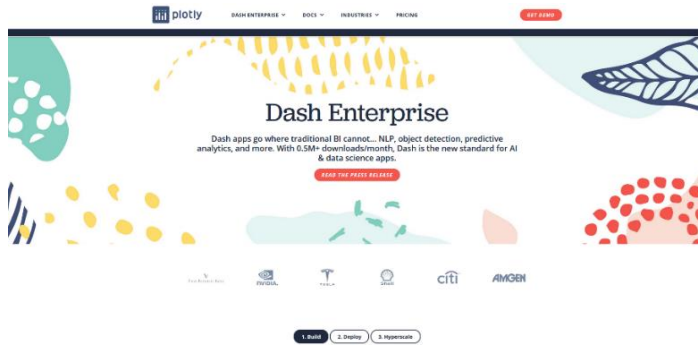
Σύνδεσμος: <https://d3js.org/>

Δείγμα (τυχαίο από youtube): <https://www.youtube.com/watch?v=bp2GF8XcJdY>

Παραδείγματα: <https://observablehq.com/@d3/gallery>

Κόστος: Είναι λογισμικό ανοιχτού κώδικα με [άδεια BSD](#)

17. [Plot.ly](https://plot.ly): Το [Plotly](https://plot.ly) είναι ένα πολύ γνωστό και ισχυρό πλαίσιο οπτικοποίησης δεδομένων το



Εικόνα 9.2.1.1-17: [29] Στιγμιότυπο οθόνης από την ιστοσελίδα του [Plot.ly](https://plot.ly).

οποίο εκτός από γραφήματα, προσφέρει επίσης λεπτομερείς αναφορές επιχειρηματικής ευφυΐας για παρακολούθηση δεδομένων. Κατάλληλο για επιχειρήσεις που απαιτούν τεχνητή νοημοσύνη (AI) και επιστήμη δεδομένων μεγάλης

κλίμακας (data science on a large scale).

(+): *Πλεονεκτήματα:*

- δημιουργία όμορφων διαδραστικών γραφημάτων και χαρτών, ευέλικτη διεπαφή.
- δημιουργία ορισμένων γραφημάτων που δεν υπάρχουν στα περισσότερα πακέτα, όπως γραφήματα περιγράμματος γραφήματα (contour plots), γραφήματα κηροπήγια (candlestick charts) και τρισδιάστατα γραφήματα (3D charts)
- δυνατότητα για εξαγωγή αναφορών για εξωτερική κοινή χρήση
- διατηρεί τις ταχύτερα αναπτυσσόμενες βιβλιοθήκες οπτικοποίησης ανοιχτού κώδικα για R, Python και JavaScript. Παρέχει επίσης σύγχρονες εφαρμογές ανάλυσης σε Python για την επιχείρηση
- πολλαπλή ανάπτυξη — επιλογές που καταλήγουν σε λύσεις με πολλούς τρόπους
- προσθήκη πολλών χρηστών για συνεργασία ως ομάδα και βελτίωση της ροής εργασίας
- υποστήριξη για προσαρμοσμένους πίνακες ελέγχου, σωστές αναφορές και μοντελοποίηση προσομοίωσης
- υποστήριξη για ευρέως χρησιμοποιούμενες γλώσσες όπως Python, Node και R

Σύνδεσμος: <https://plot.ly/>

Δείγμα (τυχαίο από youtube): <https://www.youtube.com/watch?v=GGL6U0k8WYA>

Κόστος: διαθέτει δωρεάν πρόγραμμα (βασικές δυνατότητες), ενώ για τις προηγμένες δυνατότητες [κατόπιν επικοινωνίας](#) βάσει των απαιτήσεων

18. [Chart.js](#): Το Chart.js είναι μια βιβλιοθήκη γραφημάτων JavaScript ανοιχτού κώδικα για σχεδιαστές και



Εικόνα 9.2.1.1-18: [30] Στιγμιότυπο οθόνης από την ιστοσελίδα του Chart.js

προγραμματιστές, η οποία παρέχει 8 προσαρμόσιμες δυναμικές απεικονίσεις για τα δεδομένα. Η πιο χαρακτηριστική ιδιότητα του chart.js είναι ότι μπορεί να σχεδιάσει υπέροχα γραφήματα με απόκριση

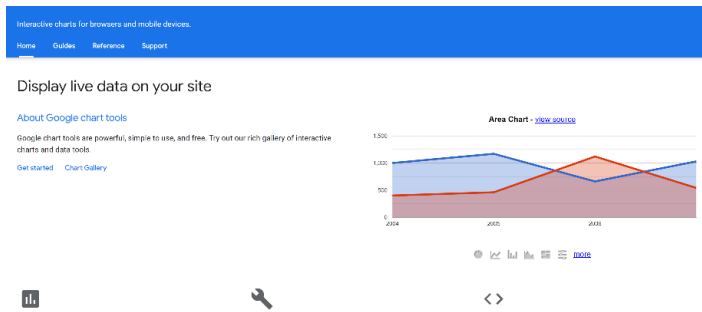
χρησιμοποιώντας τον καμβά HTML5. Το Chart.js επιτρέπει το συνδυασμό διαφορετικών τύπων γραφημάτων και στη συνέχεια τη σχεδίαση δεδομένων ημερομηνίας, λογάριθμου ή προσαρμοσμένης κλίμακας δεδομένων σε αυτό. Μπορούν επίσης να εφαρμοστούν κινούμενα σχέδια εκτός καρέ (out-of-frame animation) κατά την αλλαγή δεδομένων ή την ενημέρωση χρωμάτων.

Σύνδεσμος: <http://www.chartjs.org/>

Δείγματα: <https://www.chartjs.org/docs/latest/samples/bar/vertical.html>

Κόστος: [ανοιχτού κώδικα](#)

19. [Google Charts](#): είναι ένα δημοφιλές δωρεάν λογισμικό για οπτικοποίηση δεδομένων. Το



[Google Charts](#) είναι κατάλληλο για οργανισμούς που θέλουν να δημιουργήσουν γραφήματα από πίνακες δεδομένων και φύλλα Google.

Εικόνα 9.2.1.1-19 [31] Στιγμιότυπο οθόνης από την ιστοσελίδα του Google Charts

(+): *Πλεονεκτήματα*

- λειτουργεί σε HTML5 και SVG, και επιτρέπει τη δημιουργία διαδραστικών γραφημάτων και περισσότερου ελέγχου σε αυτά. Τα ίδια εργαλεία γραφημάτων τα χρησιμοποιεί και η ίδια η Google και τα παρέχει στους προγραμματιστές εντελώς δωρεάν με εγγυημένη συμβατότητα τριών ετών.
- δημιουργία, διαχείριση και αναφορές σε δεδομένα με πίνακες, κινούμενα σχέδια και ενσωμάτωση φύλλων Google
- εύκολη ενσωμάτωση με πρόσθετα εργαλεία και προϊόντα της Google
- ενσωμάτωση γραφημάτων σε ιστοτόπους και κανάλια κοινωνικών μέσων για εμφάνιση δεδομένων σε πραγματικό χρόνο
- διαδραστικός πίνακας ελέγχου και μεγάλη συλλογή με πολλά πρότυπα.
Προσαρμοσσιμο

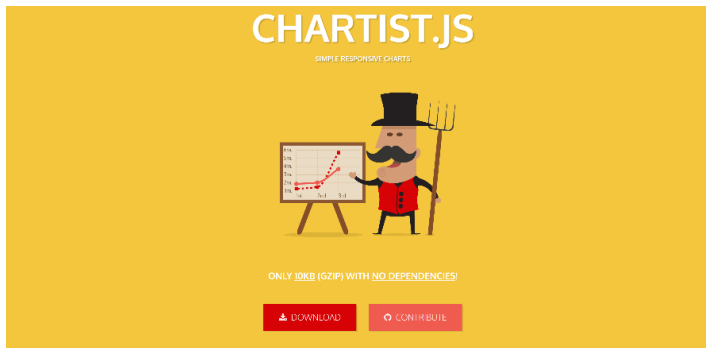
Σύνδεσμος: <https://developers.google.com/chart/>

Δείγμα (quick start): https://developers.google.com/chart/interactive/docs/quick_start?hl=en

Gallery: <https://developers.google.com/chart/interactive/docs/gallery>

Κόστος: Δωρεάν ([όροι](#))

20. [Chartist.js](#): Το Chartist.js είναι το προϊόν μιας κοινότητας που απογοητεύτηκε από τις



Εικόνα 9.2.1.1-20: [32] Στιγμιότυπο οθόνης από την ιστοσελίδα του Chartist.js

ικανότητες που παρέχονται από άλλες βιβλιοθήκες χαρτογράφησης. Το Chartist.js είναι ανοιχτού κώδικα, απλό στη χρήση και μπορεί να χρησιμοποιηθεί για τη δημιουργία γραφημάτων υψηλής προσαρμογής με

απόκριση (highly customizable responsive charts).

Σύνδεσμος: <https://gionkunz.github.io/chartist-js/>

Δείγμα (getting started): <https://gionkunz.github.io/chartist-js/getting-started.html>

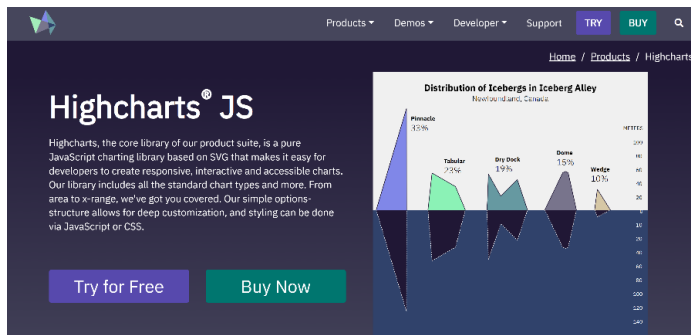
Παραδείγματα: <https://gionkunz.github.io/chartist-js/examples.html#simple-line-chart>

Κόστος: Δωρεάν ([download](#))

9.2.1.2 Εμπορικά

Για την εμφάνιση πολλών περίπλοκων γραφημάτων για επαγγελματικούς σκοπούς με τεχνική υποστήριξη, ίσως χρειαστεί η αγορά αδειών με ισχυρές βιβλιοθήκες χαρτογράφησης.

21. [Highcharts](#): Το Highcharts είναι μια βιβλιοθήκη γραφημάτων ανοιχτού κώδικα, γραμμένη



σε JavaScript και TypeScript η οποία διευκολύνει την προσθήκη διαδραστικών αποκριτικών (responsive) και προσβάσιμων γραφημάτων σε ιστότοπους ή διαδικτυακές εφαρμογές. Το

Εικόνα 9.2.1.2-21: [33] Στιγμιότυπο οθόνης από την ιστοσελίδα του Highcharts

Highcharts είναι πιο συμβατό από το D3.js και μπορεί να

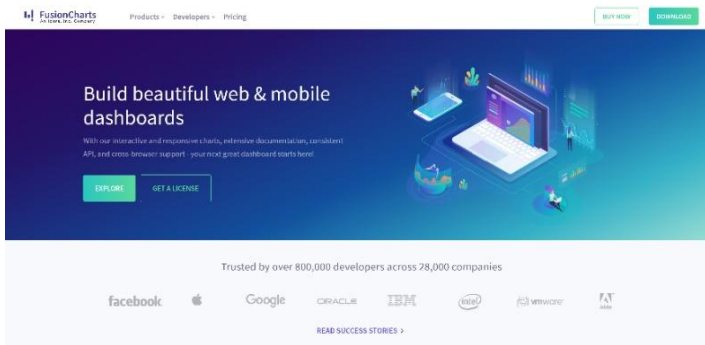
χρησιμοποιηθεί σε όλες τις κινητές συσκευές και τα προγράμματα περιήγησης υπολογιστή, χρησιμοποιώντας SVG σε σύγχρονα προγράμματα περιήγησης και VML σε χαμηλότερες εκδόσεις του προγράμματος περιήγησης IE για τη σχεδίαση γραφικών.

Σύνδεσμος: <https://www.highcharts.com/products/highcharts/>

Δείγμα: <https://www.highcharts.com/demo/line-basic>

Κόστος: Δωρεάν για προσωπική εκμάθηση, προσωπικούς ιστότοπους και μη εμπορική χρήση. Για εμπορική και κυβερνητική χρήση, intranet ή έργο (project) η αγορά άδειας μαζί με υποστήριξη premium είναι απαραίτητη (<https://shop.highsoft.com/>).

22. [FusionCharts](#): Το [FusionCharts](#) είναι μια πλατφόρμα οπτικοποίησης δεδομένων που



Εικόνα 9.2.1.2-22: [34] Στιγμιότυπο οθόνης από την ιστοσελίδα του FusionCharts

βασίζεται σε JavaScript και διαθέτει μια μεγάλη βιβλιοθήκη με 90 γραφήματα και πάνω από 1000 χάρτες. Παρέχει υπηρεσίες με χρέωση με μια ισχυρή βιβλιοθήκη γραφημάτων JavaScript και άμεση τεχνική υποστήριξη. Το FusionCharts μπορεί να μειώσει το βάρος των

προγραμματιστών δημιουργώντας υπέροχα γραφήματα και χάρτες μέσω διαδραστικών απεικονίσεων δεδομένων και οι ενσωματώσεις του είναι επίσης διαθέσιμες για όλες τις δημοφιλείς στοίβες τεχνολογίας (tech stacks). Το FusionCharts θεωρείται κατάλληλο για μεγάλους οργανισμούς που πρέπει να δημιουργήσουν πίνακες ελέγχου για έργα ή προϊόντα. Επιτρέπει την ανάκτηση ακατέργαστων δεδομένων από μια σειρά βάσεων δεδομένων και τη μετατροπή τους σε διορατικές αναφορές και διαδραστικούς πίνακες ελέγχου.

(+): *Πλεονεκτήματα*

- Λύση πλούσια σε δυνατότητες με διαδραστικές επιλογές, πολλούς συνδυασμούς χρωμάτων και δεδομένα με δυνατότητα κλικ (clickable data)
- Μεγάλη γκάμα γραμματοσειρών και διατάξεων (layouts) για τη δημιουργία προσαρμοσμένων πινάκων εργαλείων
- Μοναδικά χαρακτηριστικά, όπως συμβατότητα μεταξύ προγραμμάτων περιήγησης, σχεσιακή εμφάνιση, συνεπές API, οπτική ανακάλυψη και άλλα.

Σύνδεσμος: <https://www.fusioncharts.com/>

Gallery: <https://www.fusioncharts.com/charts>

Κόστος: δυνατότητα [δωρεάν δοκιμαστικής λήψης](#) καθώς επίσης [επιλογή διαφορετικών πακέτων](#), με το βασικό πακέτο να κοστίζει 499\$ το χρόνο.

23. [ZingChart](#): Το ZingChart είναι μια βιβλιοθήκη γραφημάτων που υλοποιείται σε JavaScript



Εικόνα 9.2.1.2-23: [35] Στιγμιότυπο οθόνης από την ιστοσελίδα του ZingChart

κυρίως για SaaS και Enterprise. Παρέχει γραφήματα μεγάλων δεδομένων που μπορούν να αποδώσουν 100.000 σημεία δεδομένων σε λιγότερο από 1 δευτερόλεπτο. Υποστηρίζονται επίσης κινούμενα, αποκριτικά (Responsive) και διαδραστικά γραφήματα που κλιμακώνονται σε

οποιοδήποτε μέγεθος συσκευής.

Σύνδεσμος: <https://www.zingchart.com/>

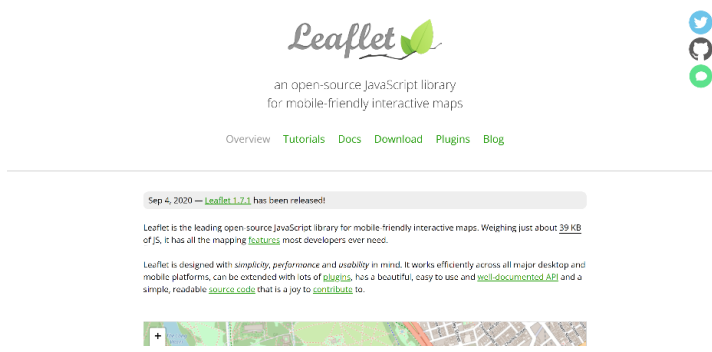
Gallery: <https://www.zingchart.com/gallery>

Κόστος: δυνατότητα δωρεάν λήψης με υποχρεωτικό υδατογράφημα όμως στο αποτέλεσμα, καθώς και πολλές επιλογές πακέτων, ξεκινώντας από 88\$ το χρόνο για το πακέτο «ιστοτόποι»

9.2.2 Χάρτες

Όταν οι προγραμματιστές θέλουν να εμφανίζουν διαδραστικούς χάρτες στους ιστότοπους, οι βιβλιοθήκες χαρτών σε JavaScript είναι απαραίτητες

24. [Leaflet](#): Το Leaflet είναι η κορυφαία σύγχρονη βιβλιοθήκη JavaScript ανοιχτού κώδικα για



Εικόνα 9.2.2-24: [36] Στιγμιότυπο οθόνης από την ιστοσελίδα του Leaflet

• διαδραστικούς χάρτες φιλικούς προς κινητά και ελαφρύ (στα 39Kb). Αυτό το δωρεάν εργαλείο οπτικοποίησης δεδομένων επιτρέπει στον χρήστη να δημιουργεί πλούσιες, διαδραστικές και φιλικές προς κινητά οπτικοποιήσεις

δεδομένων μεγάλου όγκου με 100.000 + σημεία, χρησιμοποιώντας απλώς τη δύναμη ενός προγράμματος περιήγησης. Το εργαλείο χρησιμοποιείται από χιλιάδες (συμπεριλαμβανομένων το NPR, η Washington Post και η Boston Globe).

(+) Πλεονεκτήματα:

- είναι απλό και εύχρηστο και λειτουργεί αποτελεσματικά σε όλες τις μεγάλες πλατφόρμες επιτραπέζιων και φορητών υπολογιστών (απευθείας)
- υπάρχει τεράστιος αριθμός από πρόσθετα (plugins) (πράγμα όχι και τόσο συνηθισμένο σε άλλα εργαλεία οπτικοποίησης δεδομένων) το οποίο συνοδεύεται από ένα εύχρηστο και καλά τεκμηριωμένο API.
- διαδραστικοί, responsive χάρτες. Εμφάνιση χαρτών σε μη σφαιρικές προβολές Mercator
- ως αρχείο HTML μπορεί να ερμηνευτεί από ένα ευρύ φάσμα προγραμμάτων περιήγησης
- δημιουργία θεματικών χαρτών και απεικονίσεις σε στυλ infographic
- Λειτουργίες προσαρμογής

(-) *Μειονεκτήματα:*

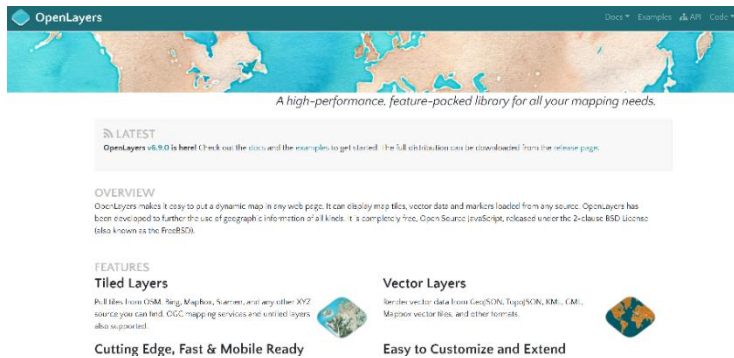
- χρειάζεται βασικές γνώσεις JavaScript, HTML και CSS, κατάλληλο για εξοικειωμένους χρήστες με τη γραμμή εντολών ή τη διεπαφή προγραμματισμού εφαρμογών (API) καθώς, τους αλγόριθμους οπτικοποίησης δεδομένων τους εισάγει ο χρήστης.

Σύνδεσμος: <https://leafletjs.com/>

Παραδείγματα: <https://leafletjs.com/examples.html>

Κόστος: Δωρεάν ([download](#))

25. [OpenLayers](#): Το OpenLayers είναι μια ανοιχτού κώδικα βιβλιοθήκη JavaScript από την



πλευρά του χρήστη (client-side JavaScript library) για τη δημιουργία διαδραστικών χαρτών Ιστού (web maps) οι οποίοι μπορούν να προβληθούν σχεδόν σε οποιοδήποτε πρόγραμμα περιήγησης (browser). Το OpenLayers δεν απαιτεί

Εικόνα 9.2.2-25: [37] Στιγμιότυπο οθόνης από την ιστοσελίδα του OpenLayers

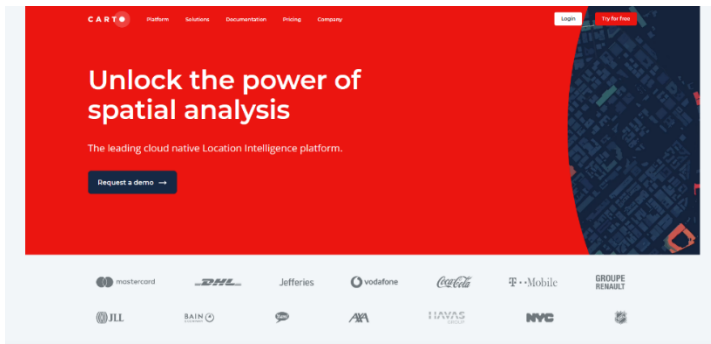
ειδικό λογισμικό από την πλευρά του διακομιστή (server-side) ή οποιαδήποτε διαμόρφωση (configuration) και μπορεί να χρησιμοποιηθεί απευθείας χωρίς τοπική λήψη. Ως μία από τις πιο ευρέως χρησιμοποιούμενες μηχανές χαρτών στον κλάδο, το OpenLayers έχει υιοθετηθεί από μεγάλους πωλητές (vendors) GIS και την πλειοψηφία των προγραμματιστών Web GIS .

Σύνδεσμος: <http://openlayers.org/>

Παραδείγματα: <https://openlayers.org/en/latest/examples/>

Κόστος: Δωρεάν ([download](#))

26. **CARTO:** Το CARTO (πρώην [CartoDB](#)) είναι μια ισχυρή και διαισθητική πλατφόρμα για



Εικόνα 9.2.2-26: [38] Στιγμιότυπο οθόνης από την ιστοσελίδα του CARTO:

αυτόματη ανακάλυψη και ανάλυση δεδομένων. Με το CARTO, μπορούμε να ανεβάσουμε τα γεωχωρικά μας δεδομένα (geospatial data) και να τα οπτικοποιήσουμε σε ένα σύνολο δεδομένων (dataset) ή σε έναν διαδραστικό χάρτη. Το

CARTO μπορεί να εγκατασταθεί στον δικό μας διακομιστή.

Σύνδεσμος: <https://github.com/CartoDB/cartodb>

Κόστος: [ανοιχτού κώδικα](#). Προσφέρονται [υπηρεσίες και λογισμικό επί πληρωμή](#).

9.2.3 Γραφήματα Δικτύου

Το *Gerhi*, το λογισμικό για την οπτικοποίηση κοινωνικών δικτύων (for visualizing social networks), είναι πολύ δημοφιλές μεταξύ όσων ασχολούνται με σχεσιακά δεδομένα (relational data) και μπορούμε επίσης να βρούμε βιβλιοθήκες γραφημάτων JavaScript για προβολή δικτύου (for network display).

27. [Sigma](#): Το Sigma είναι μια βιβλιοθήκη JavaScript διαδραστικής οπτικοποίησης,



Εικόνα 9.2.3-27: [39] Στιγμιότυπο οθόνης από την ιστοσελίδα του Sigma

αφιερωμένη στη σχεδίαση γραφημάτων. Το Sigma στοχεύει στο να βοηθήσει στην εμφάνιση δικτύων (networks) σε ιστοσελίδες, από απλές διαδραστικές δημοσιεύσεις των δικτύων έως πλούσιες εφαρμογές Ιστού που διαθέτουν δυναμική

εξερεύνηση δικτύου και επομένως είναι πολύ χρήσιμο στην οπτικοποίηση δικτύων μεγάλων δεδομένων. Η εξαγωγή (export) ενός γραφήματος από το Gerhi, μπορεί να εμφανιστεί στις ιστοσελίδες χρησιμοποιώντας το Sigma.

Σύνδεσμος: <http://sigmajs.org/>

Δείγμα (quick start): <https://www.sigmajs.org/#quickstart>

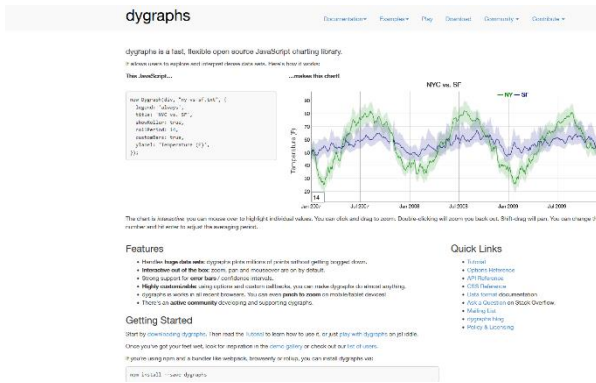
Παραδείγματα: <https://www.sigmajs.org/#usecases>

Κόστος: [MIT άδεια](#)

9.2.4 Χρηματοοικονομικά Διαγράμματα

Οι πληροφορίες μετοχών στο διαδίκτυο υποτίθεται ότι εμφανίζονται σε ειδικές βιβλιοθήκες γραφημάτων που υποστηρίζουν χρονοσειρές και πυκνά δεδομένα (*time series and dense data*).

28. [dygraphs](http://dygraphs.com/): το dygraphs είναι μια γρήγορη, ευέλικτη βιβλιοθήκη γραφημάτων JavaScript



Εικόνα 9.2.4-28: [40]Στιγμιότυπο οθόνης από την ιστοσελίδα του dygraphs

ανοιχτού κώδικα που χρησιμοποιείται κυρίως για γραφήματα μετοχών και χρηματοδότησης (stock and financing charts). Παράγει διαδραστικά και διανυσματικά γραφήματα χρονοσειρών, επιτρέποντας στους χρήστες να εξερευνούν και να ερμηνεύουν

πυκνά σύνολα δεδομένων. Το dygraphs είναι εξαιρετικά διαδραστικό με προεπιλεγμένες ετικέτες μετακίνησης του ποντικιού (mouse-over tags), μεγένθυσης και μετατόπιση (pan), ενώ έχει επίσης και άλλες χρήσιμες διαδραστικές λειτουργίες, όπως συγχρονισμό και επιλογή εύρους (range selector).

Σύνδεσμος: <http://dygraphs.com/>

Παραδείγματα: <https://dygraphs.com/gallery/>

Κόστος: [MIT άδεια](https://creativecommons.org/licenses/by/4.0/)

Επίλογος

Όπως είδαμε λοιπόν, υπάρχει πληθώρα λύσεων για όλες τις ανάγκες μας και πολλά άλλα φοβερά εργαλεία στο διαδίκτυο που δεν αναφέρθηκαν εδώ αλλά αξίζει να εξεταστούν. Όμως, είτε άτομο είτε οργανισμός, αρχικά πρέπει να κατανοηθούν οι απαιτήσεις και ο προϋπολογισμός του, προτού αναλύσει κανείς τα δεδομένα. Το πιο σημαντικό είναι να γίνει η αρχή με κάποιο από τα εργαλεία ώστε όταν ξεκινήσει κανείς να τα μαθαίνει και να μπορέσει να ξεχωρίσει ποιο είναι το καταλληλότερο για τον ίδιο. Τα παραπάνω λογισμικά οπτικοποίησης δεδομένων, μπορεί να συμβάλουν σε μεγάλο βαθμό στο να κάνει κανείς τις [παρουσιάσεις του έξυπνες](#) και διαδραστικές, να προσελκύσει το κοινό του, καθώς και να μπορέσει να αποκτήσει πολύτιμες πληροφορίες, είτε είναι μέλος μικρής ή μεγάλης επιχείρησης.

Τα εργαλεία οπτικοποίησης δεδομένων καθιστούν απλή και εύκολη την παροχή επιχειρηματικών πληροφοριών/γνώσης από δεδομένα, με τη χρήση γραφημάτων για την κοινοποίησή τους στους ενδιαφερόμενους. Η οπτικοποίηση είναι χρήσιμη για μια ολοκληρωμένη σειρά διαδικασιών, από απλές αναφορές, έως παροχή πληροφοριών από μεγάλα δεδομένα και μπορεί να χρησιμοποιηθεί με οποιαδήποτε μορφή, καθεμία από τις οποίες μπορεί να επωφεληθεί από στοχευμένα γραφήματα και διαγράμματα για την παροχή πληροφοριών οπτικά. Το κλειδί για την οπτικοποίηση είναι η απλότητα στην παρουσίαση, ανεξάρτητα από το πώς δημιουργήθηκε.

Οι απεικονίσεις πρέπει να είναι κατανοητές και προσβάσιμες από όλους. Πρέπει να είναι όσο το δυνατό πιο δημιουργικές και αξιομνημόνευτες, με σωστό έλεγχο περιπλοκότητας, συνέπεια αλλά χωρίς να επαναλαμβανόμαστε.

Η αλληλεπιδραστικότητα, η κίνηση, η δημιουργικότητα βοηθάει στην καλύτερη κατανόηση του μηνύματος που θέλουμε να μεταδώσουμε. Στατικές, για έντυπη εκτύπωση, ή όχι, πρέπει «να έχουν φωνή» και να περνάνε το σωστό μήνυμα.

10 Βιβλιογραφία

- [1]. DATA VISUALIZATION: Charts, Maps, and Interactive Graphics - Robert Grant – CRC Press – ISBN: 978-1-138-70760-3
- [2]. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. Claus O. Wilke - O'Reilly Media - ISBN: 978-1-492-03108-6 (τα περισσότερα σχήματα έγιναν με R και ggplot2 και τα οπτικά μπορούν να παραχθούν με λίγες γραμμές κώδικα - πηγές του βιβλίου στο GitHub repository <https://github.com/clauswilke/dataviz>).

Άλλες πηγές ήταν sites όπως:

- [3]. Gestalt's principles for data visualization (<https://www.topcoder.com/blog/gestalt-principles-for-data-visualization/>, as seen on 20/08/21)
- [4]. Gestalt principles of visual perception: cognitive psychology for UX (<https://www.usertesting.com/blog/gestalt-principles>, as seen at 06/11/2021)
- [5]. <https://www.scmp.com/infographics/article/1284683/iraqs-bloody-toll>
- [6]. <https://guns.periscopic.com/>
- [7]. <https://stateofchildhoodobesity.org/data/>
- [8]. <https://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study>
- [9]. <https://www.nytimes.com/interactive/2014/06/05/upshot/how-the-recession-reshaped-the-economy-in-255-charts.html>
- [10]. <http://www.lightningmaps.org/>
- [11]. <http://www.flightradar24.com/>
- [12]. <https://www.statista.com/statistics/871513/worldwide-data-created/>
- [13]. <https://rawgraphs.io/>
- [14]. <https://www.chartblocks.com/>
- [15]. <https://www.tableau.com/>
- [16]. <https://powerbi.microsoft.com/>
- [17]. <https://www.qlik.com/us/products/qlikview>
- [18]. http://www.finereport.com/en/?utm_source=Octoparse&utm_medium=media&utm_term=30dvtools&utm_content=30dvtools

- [19]. <https://www.datawrapper.de/>
- [20]. <https://www.visme.co/>
- [21]. <https://www.grow.com/>
- [22]. <https://infogram.com/>
- [23]. <https://visual.ly/>
- [24]. <https://www.instantatlas.com/>
- [25]. <https://gephi.org/>
- [26]. <http://www.wolframalpha.com/>
- [27]. <https://echarts.apache.org/>
- [28]. <https://d3js.org/>
- [29]. <https://plot.ly/>
- [30]. <http://www.chartjs.org/>
- [31]. <https://developers.google.com/chart/>
- [32]. <https://gionkunz.github.io/chartist-js/>
- [33]. <https://www.highcharts.com/products/highcharts/>
- [34]. <https://www.fusioncharts.com/>
- [35]. <https://www.zingchart.com/>
- [36]. <https://leafletjs.com/>
- [37]. <http://openlayers.org/>
- [38]. <https://carto.com/>
- [39]. <http://sigma.js.org/>
- [40]. <http://dygraphs.com/>
- [41]. <https://www.octoparse.com/blog/top-30-data-visualization-tools-in-2021#> (as seen at 11/06/2021)
- [42]. <https://techresearchonline.com/blog/best-data-visualization-tools-2021/#> (as seen at 11/06/2021)
- [43]. <https://www.techradar.com/best/best-data-visualization-tools> (as seen at 11/06/2021)
- [44]. <https://solutionsreview.com/business-intelligence/the-best-data-visualization-tools-and-software/> (as seen at 11/06/2021)
- [45]. <https://piktochart.com/blog/data-visualization-tools/> (Published on August 10, 2021 November 12, 2021 Updated on November 12, 2021)

[46]. <https://www.datamation.com/big-data/top-data-visualization-tools-for-2021/> (Published on June 4, 2021)

Παράθεμα:

[100]:: Παραδείγματα της νέας αυτής στροφής στην οπτικοποίηση (methoviz) είναι:

Για την καλύτερη κατανόηση του πώς ταιριάζουν τα splines στα δεδομένα, δημιουργήθηκε μία σελίδα από τον Paul Lambert (στατιστικολόγος και ιατρικός ερευνητής), στην οποία μπορεί να γίνει παραμετροποίηση κάποιων μεταβλητών (<https://pclambert.net/interactivegraphs/>).

Οι Ian Johnson, Martin Wattenberg και Fernanda Viégas της Google δημιούργησαν μια σελίδα σχετικά με τη μέθοδο t-SNE για την οπτικοποίηση συστάδων (clusters) σε δεδομένα πολλαπλών μεταβλητών (<https://distill.pub/2016/misread-tsne/>).

Παραδείγματα αναλυτικών και γραφικών εργαλείων με JavaScript για μεγαλύτερη ελευθερία στον χρήστη και διαθεσιμότητας ακόμα και μέσω κινητού:

Εξήγηση της μεθόδου Bayes από τον Rasmus Bååth, για σύγκριση μέσω των μεταξύ δύο ομάδων μη συγκρίσιμων δεδομένων στο διαδίκτυο, ελευθερία να δοκιμάσουμε τα δικά μας δεδομένα και όλο αυτό χωρίς τη χρήση ειδικού λογισμικού καθώς τρέχει στον browser του χρήστη (client-side) (http://www.sumsar.net/best_online/). [Εναλλακτικά, θα μπορούσε να γίνει εγκατάσταση λογισμικού στον server το οποίο να παρέχει τα αποτελέσματα του χρήστη στον browser (server-side). Αυτό εκτός από τον έλεγχο του διακομιστή, χρειάζεται και προγραμματιστική τεχνογνωσία. Επίσης, υπάρχουν απλά εργαλεία που μπορεί κανείς να το επιτύχει, όπως το πακέτο R Shiny]

Μια ποικιλία αναλύσεων χρησιμοποιώντας υπολογιστικές προσομοιώσεις και τυχαιοποιημένες μεθόδους (lock5stat.com/statkey) που φτιάχτηκε από τους Locks, οικογένεια καθηγητών στατιστικής.