



**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Οπτικοποίηση Δεδομένων με Στόχο την Εξόρυξη
Γνώσης

Τριανταφύλλου Θοδωρής

Επιβλέπων καθηγητής: Τζήμας Γιάννης

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή

Πάτρα, Ημερομηνία

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Ονοματεπώνυμο, Υπογραφή
2. Ονοματεπώνυμο, Υπογραφή
3. Ονοματεπώνυμο, Υπογραφή

Αφιέρωση

Στους γονείς μου.

Ευχαριστίες

Ευχαριστώ τους γονείς μου για την στήριξή τους όλα αυτά τα χρόνια και τον καθηγητή μου κ.Τζήμα Ιωάννη για την υπομονή του.

Περιεχόμενα

Αφιέρωση.....	2
Ευχαριστίες.....	3
Περιεχόμενα	4
1 Εισαγωγή.....	10
2 Εξόρυξη δεδομένων.....	12
2.1 Ο όρος	12
2.2 Ο στόχος.....	12
2.3 Ιστορία και Εξέλιξη.....	13
2.4 Διαδικασία	13
2.5 Προ-επεξεργασία	14
2.6 Τεχνικές.....	15
2.6.1 Γραμμική παλινδρόμηση	15
2.6.2 Λογιστική παλινδρόμηση.....	17
2.6.3 Δέντρο απόφασης.....	18
2.6.4 Μηχανές υποστήριξης	20
2.6.5 Naïve Bayes	22
2.6.6 Κ πλησιέστεροι γείτονες	23
2.6.7 K-Means Clustering	25
2.6.8 Τυχαίο Δάσος	26
2.6.9 Μείωση των διαστάσεων	28
2.6.10 Βελτίωση κλίσης	29
2.7 Επικύρωση αποτελέσματος.....	31
2.8 Εφαρμογές	32
2.8.1 Ιατρική.....	32
2.8.2 Οικονομία	33
2.8.3 Τηλεπικοινωνία.....	34
3 Οπτικοποίηση Δεδομένων.....	35
3.1 Επισκόπηση.....	35
3.2 Χαρακτηριστικά των αποτελεσματικών γραφικών απεικονίσεων.....	36

3.3	Ποσοτικά μηνύματα.....	38
3.4	Οπτική αντίληψη και οπτικοποίηση δεδομένων.....	39
3.5	Ιστορία της οπτικοποίησης δεδομένων.....	40
3.6	Ορολογία.....	42
3.7	Παραδείγματα διαγραμμάτων που χρησιμοποιούνται για την απεικόνιση δεδομένων	43
3.7.1	Επιχειρήσεις και οικονομία	43
3.7.2	Γραφήματα μηχανικής και τεχνολογίας	49
3.7.3	Υγεία και ευεξία	54
3.7.4	Ιστορία	61
3.7.5	Γραφήματα μάρκετινγκ.....	66
3.7.6	Μετεωρολογία και περιβάλλον	72
3.7.7	Πολιτική επιστήμη και κοινωνιολογία.....	77
3.7.8	Επιστήμη	82
4	Η οπτικοποίηση	88
4.1	Οι τεχνολογίες.....	89
4.1.1	HTML	89
4.1.2	CSS.....	90
4.1.3	Javascript.....	91
4.1.4	D3.js(data driven documents).....	91
4.2	Η ιστορία.....	92
4.3	Η υλοποίηση	92
4.3.1	CSS.....	92
4.3.2	Data.....	94
4.3.3	Javascript.....	95
5	Βιβλιογραφία	105

Περίληψη

Εκπληκτικά, το 90% των δεδομένων στον κόσμο σήμερα δημιουργήθηκε μόνο τα τελευταία δύο χρόνια. Με την αύξηση των κινητών συσκευών, των δικτύων κοινωνικών μέσων ενημέρωσης και την κοινή χρήση ψηφιακών φωτογραφιών και βίντεο, συνεχίζουμε να αυξάνουμε τα δεδομένα του κόσμου με εκπληκτικό ρυθμό. Η εξόρυξη δεδομένων είναι ο τρόπος με τον οποίο οι επιχειρηματίες μπορούν να διερευνήσουν αυτά τα δεδομένα, να κάνουν ενημερωτικές ανακαλύψεις και να θέσουν αυτές τις πληροφορίες σε λειτουργία σε καθημερινές επιχειρήσεις. Όταν δημιουργούμε μια οπτικοποίηση δεδομένων, αποφασίζουμε ποια δεδομένα χρειαζόμαστε, συλλέγουμε, προετοιμάζουμε και καθαρίζουμε για χρήση, και στη συνέχεια κάνουμε την προβολή και παρουσιάζουμε το τελικό αποτέλεσμα. Σκοπός της πτυχιακής αυτής είναι να παρουσιάσει αυτή τη μεθοδολογία. Παρουσιάζουμε την έρευνα μας στον τομέα της εξόρυξης και οπτικοποίησης δεδομένων χρησιμοποιώντας ένα σύνολο δεδομένων και προσπαθούμε να δείξουμε πως τα πέντε μεγαλύτερα κρυπτονομίσματα έχουν αλλάξει τιμή στο πέρασμα του χρόνου.

Abstract

Surprisingly, 90% of the data in the world today has only been created in the last two years. With the rise of mobile devices, social media and the sharing of digital photos and videos, we continue to grow the world's data at a remarkable rate. Data mining is the way that entrepreneurs can explore this data, make informational discoveries and put this information into daily operations. When we create a data visualization, we decide what data we need, collect, prepare and clean for use, and then view and present the end result. The purpose of this thesis is to present this methodology. We present our research in data mining and visualization using a dataset and try to show how the five largest cryptocurrencies have changed in value over time.

1 Εισαγωγή

Κατακλυζόμαστε από δεδομένα. Η ποσότητα των δεδομένων στον κόσμο και στη ζωή μας φαίνεται να αυξάνεται συνεχώς και δεν υπάρχει κανένα τέλος. Όλοι οι υπολογιστές καθιστούν πάρα πολύ εύκολο να σώσουμε πράγματα που προηγουμένως θα πετούσαμε. Οι φθηνοί δίσκοι και η ηλεκτρονική αποθήκευση καθιστούν πολύ εύκολο να αναβληθούν οι αποφάσεις σχετικά με το τι πρέπει να γίνει με αυτά τα πράγματα-απλά παίρνουμε περισσότερη μνήμη και τα κρατάμε όλα. Τα πανταχού παρόντα ηλεκτρονικά καταγράφουν τις αποφάσεις μας, τις επιλογές μας στο σούπερ μάρκετ, τις οικονομικές συνήθειες, τα έσοδά μας και τα βήματα. Κάνουμε swiρε το δρόμο μας μέσα από τον κόσμο, κάθε swiρε ένα αρχείο σε μια βάση δεδομένων. Ο Παγκόσμιος Ιστός (WWW) μας συγκλονίζει με πληροφορίες ενώ, κάθε επιλογή που κάνουμε καταγράφεται. Και όλα αυτά είναι απλώς προσωπικές επιλογές - έχουν αμέτρητους ομολόγους στον κόσμο του εμπορίου και της βιομηχανίας. Θα μπορούσαμε όλοι να καταθέσουμε στο αυξανόμενο χάσμα μεταξύ της παραγωγής δεδομένων και της κατανόησής της. Δεδομένου ότι ο όγκος των δεδομένων αυξάνεται αναπόφευκτα, το ποσοστό που καταλαβαίνουν οι άνθρωποι μειώνεται ανησυχητικά. Το ψέμα που κρύβεται σε όλα αυτά τα δεδομένα είναι πληροφορίες - δυνητικά χρήσιμες πληροφορίες - που σπάνια γίνονται σαφείς ή εκμεταλλεύονται.

Οι άνθρωποι αναζητούν πρότυπα στα δεδομένα από τότε που άρχισε η ανθρώπινη ζωή. Οι κυνηγοί αναζητούν μοτίβα στη μεταναστευτική συμπεριφορά των ζώων, οι αγρότες αναζητούν πρότυπα στην ανάπτυξη των καλλιεργειών, οι πολιτικοί αναζητούν πρότυπα στη γνώμη των ψηφοφόρων και οι εραστές αναζητούν πρότυπα στις απαντήσεις των συνεργατών τους. Η δουλειά ενός επιστήμονα (όπως ένα μωρό) είναι να κατανοήσει τα δεδομένα, να ανακαλύψει τα πρότυπα που διέπουν τον τρόπο με τον οποίο λειτουργεί ο φυσικός κόσμος και να τα ενσωματώνει σε θεωρίες που μπορούν να χρησιμοποιηθούν για να προβλέψουν τι θα συμβεί σε νέες καταστάσεις. Η δουλειά του επιχειρηματία είναι να εντοπίσει ευκαιρίες - δηλαδή, μοτίβα συμπεριφοράς που μπορούν να μετατραπούν σε μια κερδοφόρα επιχείρηση - και να τα αξιοποιήσουν. Στην εξόρυξη δεδομένων, τα δεδομένα αποθηκεύονται ηλεκτρονικά και η αναζήτηση είναι αυτοματοποιημένη - ή τουλάχιστον επαυξημένη - από υπολογιστή. Ακόμα και αυτό δεν είναι ιδιαίτερα νέο. Οι οικονομολόγοι, οι στατιστικοί, οι προβλέψεις και οι μηχανικοί επικοινωνίας έχουν από καιρό εργαστεί με την ιδέα ότι τα πρότυπα στα δεδομένα μπορούν να αναζητηθούν αυτόματα, ταυτοποιήθηκαν, επικυρώθηκαν και χρησιμοποιήθηκαν για πρόβλεψη. Αυτό που είναι νέο είναι η εκπληκτική αύξηση των ευκαιριών για την εύρεση σχεδίων στα δεδομένα. Η αχαλίνωτη ανάπτυξη βάσεων δεδομένων τα τελευταία χρόνια, οι βάσεις δεδομένων για τέτοιες καθημερινές δραστηριότητες όπως οι επιλογές των πελατών, φέρνουν τα δεδομένα εξόρυξης στο προσκήνιο των νέων τεχνολογιών των επιχειρήσεων. Έχει υπολογιστεί ότι ο όγκος των

δεδομένων που αποθηκεύονται στις βάσεις δεδομένων παγκοσμίως διπλασιάζεται κάθε 20 μήνες και παρόλο που θα ήταν σίγουρα δύσκολο να δικαιολογηθεί αυτός ο αριθμός σε οποιαδήποτε ποσοτική έννοια, μπορούμε όλοι να συνδέσουμε ποιοτικά τον ρυθμό ανάπτυξης. Καθώς οι πλημμύρες των δεδομένων διογκώνονται και οι μηχανές που μπορούν να αναλάβουν την αναζήτηση καθίστανται συνηθισμένες, αυξάνονται οι δυνατότητες εξόρυξης δεδομένων. Καθώς ο κόσμος μεγαλώνει με πολυπλοκότητα, μας συντρίβει με τα δεδομένα που παράγει, η εξόρυξη δεδομένων γίνεται η μόνη μας ελπίδα για την αποσαφήνιση κρυφών προτύπων. Τα έξυπνα αναλυμένα δεδομένα είναι ένας πολύτιμος πόρος. Μπορεί να οδηγήσει σε νέες ιδέες και, σε εμπορικά περιβάλλοντα, σε ανταγωνιστικά πλεονεκτήματα. Η εξόρυξη δεδομένων αφορά την επίλυση προβλημάτων με την ανάλυση δεδομένων που υπάρχουν ήδη σε βάσεις δεδομένων.

Στην σημερινή άκρως ανταγωνιστική, πελατοκεντρική και οικονομικά προσανατολισμένη στις υπηρεσίες οικονομία, τα δεδομένα είναι η πρώτη ύλη που τροφοδοτεί την ανάπτυξη των επιχειρήσεων - αν μόνο μπορεί να εξορύσσεται. Η εξόρυξη δεδομένων ορίζεται ως η διαδικασία ανακαλύψεως προτύπων στα δεδομένα. Η διαδικασία πρέπει να είναι αυτόματη ή (πιο συνηθισμένη) ημιαυτόματη. Τα πρότυπα που ανακαλύφθηκαν πρέπει να έχουν νόημα στο ότι οδηγούν σε κάποιο πλεονέκτημα, συνήθως οικονομικό. Τα δεδομένα είναι πάντοτε παρόντα σε σημαντικές ποσότητες. Και πώς εκφράζονται τα μοτίβα; Τα χρήσιμα μοτίβα μας επιτρέπουν να κάνουμε μη προκλητικές προβλέψεις σε νέα δεδομένα. Υπάρχουν δύο άκρα για την έκφραση ενός μοτίβου: ως ένα μαύρο κουτί των οποίων τα σπλάχνα είναι πραγματικά ακατανόητα και ως ένα διαφανές κιβώτιο του οποίου η κατασκευή αποκαλύπτει τη δομή του σχεδίου. Και οι δύο, υποθέτουμε, κάνουν καλές προβλέψεις. Η διαφορά είναι αν τα σχέδια που εξορύσσονται αντιπροσωπεύονται ή όχι από μια δομή που μπορεί να εξεταστεί, να αιτιολογηθεί και να χρησιμοποιηθεί για να ενημερωθούν οι μελλοντικές αποφάσεις. Αυτά τα μοτίβα ονομάζουμε διαρθρωτικά, επειδή συλλαμβάνουν τη δομή της απόφασης με σαφή τρόπο. Με άλλα λόγια, βοηθούν να εξηγήσουμε κάτι σχετικά με τα δεδομένα.

2 Εξόρυξη δεδομένων

Εξόρυξη δεδομένων (ή ανακάλυψη γνώσης από βάσεις δεδομένων) (αγγλ. *data mining*) είναι η εξεύρεση μιας (ενδιαφέρουσας, αυτονόητης, μη προφανούς και πιθανόν χρήσιμης) πληροφορίας ή προτύπων από μεγάλες βάσεις δεδομένων με χρήση αλγορίθμων ομαδοποίησης ή κατηγοριοποίησης και των αρχών της στατιστικής, της τεχνητής νοημοσύνης, της μηχανικής μάθησης και των συστημάτων βάσεων δεδομένων. Στόχος της εξόρυξης δεδομένων είναι η πληροφορία που θα εξαχθεί και τα πρότυπα που θα προκύψουν να έχουν δομή κατανοητή προς τον άνθρωπο έτσι ώστε να τον βοηθήσουν να πάρει τις κατάλληλες αποφάσεις.

2.1 Ο όρος

Ο όρος εξόρυξη δεδομένων είναι μία έννοια που συνήθως παραπέμπει σε κάθε είδος φόρμας με μεγάλη ποσότητα δεδομένων ή επεξεργασία δεδομένων (συλλογή, εξαγωγή δεδομένων, *warehouse*, ανάλυση δεδομένων και στατιστικής) αλλά επίσης γενικεύεται σε κάθε είδος συστήματος υποστήριξης αποφάσεων συμπεριλαμβανομένου της τεχνητής νοημοσύνης, της εκμάθησης μηχανής και της επιχειρηματικής ευφυΐας. Στην ορθή χρήση του όρου η λέξη κλειδί είναι η ανακάλυψη, που ορίζεται ως η ανίχνευση κάτι καινούριου.

2.2 Ο στόχος

Ο πραγματικός στόχος της εξόρυξης δεδομένων είναι η αυτόματη ή ημιαυτόματη ανάλυση μεγάλων ποσοτήτων δεδομένων για την εξαγωγή κάποιου ενδιαφέροντος προτύπου που ήταν άγνωστο μέχρι εκείνη τη στιγμή, όπως ομάδες από εγγραφές δεδομένων (συσταδοποίηση), ασυνήθιστες εγγραφές (*anomaly detection*) και εξαρτήσεις (κανόνες συσχετίσεων). Αυτό συνήθως συμπεριλαμβάνει τη χρήση βάσης δεδομένων όπως χωρικά ευρετήρια. Αυτά τα πρότυπα ύστερα μπορούν να θεωρηθούν ως μία περιγραφή των δεδομένων εισαγωγής και να χρησιμοποιηθούν για περαιτέρω ανάλυση ή για παράδειγμα στην εκμάθηση μηχανής και στην προγνωστική ανάλυση. Για παράδειγμα, η εξόρυξη δεδομένων θα μπορούσε να προσδιορίσει πολλαπλά σύνολα στα δεδομένα, τα οποία μπορούν να χρησιμοποιηθούν μετά για να εξασφαλίσουν περισσότερο ακριβή αποτελέσματα από ένα σύστημα υποστήριξης αποφάσεων. Παρότι η συλλογή δεδομένων και η προετοιμασία δεδομένων, αλλά και η ερμηνεία των αποτελεσμάτων και εκθέσεων δεν αποτελούν μέρος της εξόρυξης δεδομένων, παρ' όλα αυτά ανήκουν στην ανακάλυψη γνώσης από βάσεις δεδομένων σαν κάποια επιπρόσθετα βήματα.

Άλλοι σχετικοί όροι της εξόρυξης δεδομένων είναι οι *data dredging*, *data fishing* και *data snooping*, που αναφέρονται στην χρήση μεθόδων της εξόρυξης δεδομένων για να πάρουν δείγματα από μεγαλύτερη συλλογή δεδομένων που είναι (ή μπορεί να είναι) πολύ μικρά για αξιόπιστα στατιστικά συμπεράσματα

που έγιναν σχετικά με τη εγκυρότητα των προτύπων που ανακαλύφθηκαν. Αυτές οι μέθοδοι, επίσης, μπορούν να χρησιμοποιηθούν για την δημιουργία νέων υποθέσεων προς εξέταση έναντι μεγαλύτερων συλλογών δεδομένων.

2.3 Ιστορία και Εξέλιξη

Η χειροκίνητη εξαγωγή προτύπων από δεδομένα συμβαίνει εδώ και αιώνες. Οι πρώτες μέθοδοι για τον προσδιορισμό προτύπων ήταν αυτές της θεωρίας Bayes και της ανάλυσης της παλινδρόμησης. Ο πολλαπλασιασμός, η ευρεία διαθεσιμότητα και η εξέλιξη της τεχνολογίας υπολογιστών έχουν αυξήσει τον όγκο των συγκεντρωμένων δεδομένων και την ζήτηση για αποδοτικούς και αποτελεσματικούς χειρισμούς. Καθώς οι συλλογές δεδομένων αυξήθηκαν τόσο σε όγκο όσο και σε πολυπλοκότητα, η χειρωνακτική ανάλυση των δεδομένων έχει αντικατασταθεί από την αυτόματη επεξεργασία δεδομένων. Σε αυτό συνέβαλαν άλλες ανακαλύψεις της επιστήμης των υπολογιστών, όπως τα νευρωνικά δίκτυα, η συσταδοποίηση, οι γενετικοί αλγόριθμοι (1950), τα δέντρα απόφασης (1960) και η μηχανή υποστήριξης διανυσμάτων(1990). Η εξόρυξη δεδομένων είναι η διαδικασία εφαρμογής αυτών των μεθόδων στα δεδομένα με σκοπό την αποκάλυψη άγνωστων προτύπων σε μεγάλα σύνολα δεδομένων. Αυτό γεφυρώνει το χάσμα της εφαρμοσμένης στατιστικής και της τεχνητής νοημοσύνης (τα οποία συνήθως παρέχουν το μαθηματικό υπόβαθρο) με την διαχείριση βάσης δεδομένων κάνοντας χρήση του τρόπου με τον οποίο αποθηκεύονται και κατατάσσονται στη βάση δεδομένων για να εκτελέσουν την θεωρία και τους διαθέσιμους αλγόριθμους περισσότερο αποτελεσματικά, επιτρέποντας σε τέτοιες μεθόδους να εφαρμόζονται σε μεγάλα σύνολα δεδομένων.

2.4 Διαδικασία

Η διαδικασία ανακάλυψης γνώσης από βάσεις δεδομένων(KDD) συνήθως ορίζεται από τα εξής στάδια:

Συλλογή

Προεπεξεργασία

Μετασχηματισμός

Εξόρυξη δεδομένων

Ερμηνεία/Αξιολόγηση.

Υπάρχουν όμως κι άλλες παραλλαγές για τον ορισμό των σταδίων αυτών σύμφωνα και με το Cross Industry Standard Process for Data Mining (CRISP-DM) όπου τα στάδια έχουν ως εξής:

Κατανόηση Θέματος

Κατανόηση δεδομένων

Προετοιμασία δεδομένων

Μοντελοποίηση

Αξιολόγηση

Ανάπτυξη ή απλοποιημένη διαδικασία όπως

Προ-επεξεργασία

Εξόρυξη δεδομένων

Επικύρωση αποτελέσματος

2.5 Προ-επεξεργασία

Πριν την εφαρμογή των αλγορίθμων εξόρυξης δεδομένων, το ερευνώμενο σύνολο δεδομένων πρέπει να συναρμολογείται. Καθώς η εξόρυξη δεδομένων μπορεί να αποκαλύψει μόνο τα πρότυπα που πράγματι εμφανίζονται στα δεδομένα, το σύνολο δεδομένων που ερευνούμε, πρέπει να είναι αρκετά μεγάλο για να περιέχει αυτά τα πρότυπα παραμένοντας να εξορυχθεί σε ένα αποδεκτό χρονικό διάστημα. Μία συνηθισμένη πηγή για δεδομένα είναι η data mart ή η data warehouse. Η προεπεξεργασία είναι απαραίτητη για την ανάλυση πολυπαραγοντικών συνόλων δεδομένων πριν την εξόρυξη δεδομένων.

Έτσι το ερευνώμενο σύνολο καθαρίζεται. Το καθάρισμα δεδομένων διαγράφει τις παρατηρήσεις που περιέχουν θόρυβο και αυτές με ελλιπή ή ελλείποντα δεδομένα.

2.6 Τεχνικές

Η εξόρυξη δεδομένων περιλαμβάνει κάποιες από τις ακόλουθες διατάξεις:

2.6.1 Γραμμική παλινδρόμηση

Χρησιμοποιείται για την εκτίμηση των πραγματικών αξιών (κόστος κατοικιών, αριθμός κλήσεων, συνολικές πωλήσεις κ.λπ.) βάσει συνεχών μεταβλητών.

Εδώ, δημιουργούμε σχέση μεταξύ ανεξάρτητων και εξαρτημένων μεταβλητών με την τοποθέτηση μιας καλύτερης γραμμής.

Αυτή η καλύτερη γραμμή fit είναι γνωστή ως γραμμή παλινδρόμησης και αντιπροσωπεύεται από μια γραμμική εξίσωση $Y = a * X + b$.

Αυτοί οι συντελεστές a και b προκύπτουν με βάση την ελαχιστοποίηση του αθροίσματος της τετραγωνικής διαφοράς απόστασης μεταξύ σημείων δεδομένων και γραμμής παλινδρόμησης.

Η γραμμική παλινδρόμηση είναι κυρίως δύο τύπων:

1. Απλή γραμμική παλινδρόμηση
2. Πολλαπλή γραμμική παλινδρόμηση.

Η απλή γραμμική παλινδρόμηση χαρακτηρίζεται από μία ανεξάρτητη μεταβλητή. Και, η πολλαπλή γραμμική παλινδρόμηση (όπως υποδηλώνει το όνομα) χαρακτηρίζεται από πολλαπλές (περισσότερες από 1) ανεξάρτητες μεταβλητές

Python Code:

```
#import Library
```

```
#import other necessary libraries like pandas, numpy
```

```
From sklearn import linear_model
```

```
#Load Train and Test datasets
```

```
#Identify feature and response variable(s) and values must be numeric and numpy array
```



```

x_train=input_variables_values_training_datasets

y_train=target_variables_values_training_datasets

x_test=input_variables_values_test_datasets

#Create linear regression object

Linear=linear_model.LinearRegression()

#Train the model using the training sets and check score

Linear.fit(x_train, y_train)

Linear.score(x_train,y_train)

#Equation coefficient and Intercept

Print('Cefficient',linear.coef_)

Print('Intercept:',linear.intercept_)

#Predict Output

Predicted=linear.predict(x_test)

```

R code:

```

#Load Train and Test datasets

#Identifyfeatureand response variable(s) and values must be numeric and numpy array

X_train<-input_variables_values_training_datasets

Y_train<-target_variables_values_training_datsets

X_test<-input_variables_values_test_datasets

X<-cbind(x_train,y_train)

#Train the model using the training sets and check score

```

```
Linear<-lm(y_train~.,data=x)
```

```
Summary(linear)
```

```
#Predicted Output
```

```
Predicted=predict(linear,x_test)
```

2.6.2 Λογιστική παλινδρόμηση

Μην συγχέεται με το όνομά της! Είναι μια ταξινόμηση, δεν είναι αλγόριθμος παλινδρόμησης.

Χρησιμοποιείται για την εκτίμηση διακριτών τιμών (δυαδικές τιμές όπως 0/1, ναι / όχι, true / false) με βάση ένα δεδομένο σύνολο ανεξάρτητων μεταβλητών.

Με απλά λόγια, προβλέπει την πιθανότητα εμφάνισης ενός συμβάντος με την τοποθέτηση δεδομένων σε μια λειτουργία logit. Ως εκ τούτου, είναι επίσης γνωστή ως παλινδρόμηση logit. Δεδομένου ότι, προβλέπει την πιθανότητα, οι τιμές εξόδου του είναι μεταξύ 0 και 1 (όπως αναμένεται).

Η σιγμοειδής συνάρτηση δίνεται από

$$Y = mx + c$$

$$S(x) = 1 / (1 + e^{-(\gamma)})$$

Είναι σχήματος S και οριοθετημένη λειτουργία. Καλείται επίσης ως λειτουργία squashing, η οποία χαρτογραφεί ολόκληρο τον πραγματικό άξονα προς το πεπερασμένο διάστημα.

Συνήθως, οι προβλέψεις στο πρόβλημα ταξινόμησης είναι τιμές πιθανότητας. Έτσι, δεν θέλουμε το μοντέλο μας να προβλέψει την τιμή πιθανότητας να είναι κάτω από 0 ή παραπάνω από 1. Η λειτουργία Σιγμοειδούς βοηθά να επιτευχθεί αυτό, οι τιμές κάτω από το 0.5 είναι κλάση 0 και οι τιμές πάνω και ίσες με 0,5 είναι κλάσης 1.

Python code:

```
#Import Library
```

```
From sklearn.linear_model import LogisticRegression
```

```
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset
```

```
#Create logistic regression object
```

```
Model = LogisticRegression()

#Train the model using the training sets and check score

Model.fit(X,y)

Model.score(X,y)

#Equation coefficient and intercept_)

#Predict Output

Predicted = model.predict(x_test)
```

R code:

```
#Load Train and Test datasets

#Identify feature and response variable(s) and values must be numeric and numpy arrays

x_train<- input_variables_values_training_datasets

y_train<- target_variables_values_training_datasets

x_test<- input_variables_values_test_datasets

x <- cbind(x_train, y_train)

#Train the model using the training sets and check score

Logistic <- glm(y_train ~ ., data = x,family='binomial')

Summary(logistic)

#Predict Output

Predicted=predict(logistic,x_test)
```

2.6.3 Δέντρο απόφασης

Είναι ένας τύπος αλγόριθμου εποπτευόμενης μάθησης που χρησιμοποιείται κυρίως για προβλήματα ταξινόμησης.

Παραδόξως, λειτουργεί τόσο για κατηγορηματικές όσο και για συνεχείς εξαρτώμενες μεταβλητές. Σε αυτόν τον αλγόριθμο, χωρίζουμε τον πληθυσμό σε δύο ή περισσότερα ομοιογενή σύνολα.

Αυτό γίνεται με βάση τις πιο σημαντικές ιδιότητες / ανεξάρτητες μεταβλητές, ώστε να γίνονται όσο το δυνατόν διακριτές ομάδες.

Ο καλύτερος τρόπος για να κατανοήσετε πώς δουλεύει το δέντρο αποφάσεων είναι να παίξετε το Jezzball - ένα κλασικό παιχνίδι από τη Microsoft. Ουσιαστικά, έχετε ένα δωμάτιο με κινούμενους τοίχους και πρέπει να δημιουργήσετε τοίχους έτσι ώστε η μέγιστη περιοχή να καθαρίζεται χωρίς τις μπάλες.

Έτσι, κάθε φορά που χωρίζετε το δωμάτιο με έναν τοίχο, προσπαθείτε να δημιουργήσετε 2 διαφορετικούς πληθυσμούς στο ίδιο δωμάτιο. Τα δέντρα αποφάσεων λειτουργούν με πολύ παρόμοιο τρόπο, διαιρώντας έναν πληθυσμό σε όσο το δυνατόν διαφορετικές ομάδες.

Για να χωρίσει τον πληθυσμό σε διαφορετικές ετερογενείς ομάδες, χρησιμοποιεί διάφορες τεχνικές όπως η Gini, η Κέρδος Πληροφορίας, η Chi-square, η εντροπία.

Python Code:

```
#Import Library
```

```
#Import other necessary libraries like pandas, numpy...
```

```
From sklearn import tree
```

```
#Assumed you have , X (predictor) and Y (target) for training data set and x_test (predictor) of test_dataset
```

```
#Create tree object
```

```
Model = tree .DecisionTreeClassifier(criterion='gini') # for classification, here you can change the algorithm as gini or entropy(information gain) by default it is gini
```

```
#model = tree.DecisionTreeRegressor() for regression
```

```
#Train the model using the training sets and check score
```

```
Model.fit(X,y)
```

```
Model.score(X,y)
```

```
#Predict Output
```

```
Predicted=model.predict(x_test)
```

R code:

```
#Load Train and Test datasets
```

```
#Identify feature and response variable(s) and values must be numeric and numpy arrays
```

```
X_train<-input_variables_values_training_datasets
```

```
Y_train<-target_variables_values_training_datasets
```

```
X_test<-input_variables_values_test_datasets
```

```
Library(rpart)
```

```
X<-cbind(x_train,y_train)
```

```
#grow tree
```

```
Fit<-rpart(y_train ~ ., data=x,method="class")
```

```
Summary(fit)
```

```
#Predict Output
```

```
Predicted=predict(fit,x_test)
```

2.6.4 Μηχανές υποστήριξης

Μια μηχανή υποστήριξης (Support Vector Machine) είναι ένας διακριτικός ταξινομητής που τυπικά ορίζεται από ένα διαχωριστικό hyperplane.

Με άλλα λόγια, δεδομένου ότι τα δεδομένα κατάρτισης είναι επισημασμένα (εποπτευόμενη μάθηση), ο αλγόριθμος εξάγει ένα βέλτιστο hyperplane που κατηγοριοποιεί νέα παραδείγματα. Σε δισδιάστατο χώρο αυτό το hyperplane είναι μια γραμμή που διαιρεί ένα plane σε δύο μέρη όπου κάθε τάξη βρίσκεται σε κάθε πλευρά.

Python Code:

```
#Import Library
```

```
From sklearn import svm
```

```
#Assumed you have, X (predictor and Y(target) for training data set and x_test(predictor) of test_dataset
```

```
#Create SVM classification object
```

```
Model = svm.svc() #there is various option associated with it, this is simple for classification. You can refer link, for more detail
```

```
#Train the model using the training sets and check score
```

```
Model.fit(X,y)
```

```
Model.score(X,y)
```

```
#Predict Output
```

```
Predicted=model.predict(x_test)
```

R code:

```
#Load Train and Test datasets
```

```
#Identify feature and response variable(s) and values must be numeric and numpy arrays
```

```
X_train<-input_variables_values_training_datasets
```

```
Y_train<-target_variables_values_training_datasets
```

```
X_test<-input_variables_values_test_datasets
```

```
Library(e1071)
```

```
X<-cbind(x_train,y_train)
```

```
#Fitting model
```

```
Fit<-svm(y_train ~ ., data = x)
```

Summary(fit)

#Predict Output

Predicted= predict(fit,x_test)

2.6.5 Naïve Bayes

Πρόκειται για μια τεχνική ταξινόμησης βασισμένη στο θεώρημα του Bayes με την παραδοχή της ανεξαρτησίας μεταξύ των προγνωστικών.

Με απλά λόγια, ένας ταξινομητής Naïve Bayes υποθέτει ότι η παρουσία ενός συγκεκριμένου χαρακτηριστικού σε μια κατηγορία δεν σχετίζεται με την παρουσία οποιουδήποτε άλλου χαρακτηριστικού. Για παράδειγμα, ένας καρπός μπορεί να θεωρηθεί μήλο αν είναι κόκκινο, στρογγυλό και περίπου 3 ίντσες σε διάμετρο.

Ακόμη και αν αυτά τα χαρακτηριστικά εξαρτώνται το ένα από το άλλο ή από την ύπαρξη των άλλων χαρακτηριστικών, ένας αφελής ταξινομητής Bayes θα θεωρούσε ότι όλες αυτές οι ιδιότητες συμβάλλουν ανεξάρτητα στην πιθανότητα ότι αυτός ο καρπός είναι μήλο.

Το Naïve Bayesian μοντέλο είναι εύκολο να κατασκευαστεί και ιδιαίτερα χρήσιμο για πολύ μεγάλα σύνολα δεδομένων.

Μαζί με την απλότητα, το Naïve Bayes είναι γνωστό ότι ξεπερνάει ακόμη και τις πολύ εξελιγμένες μεθόδους ταξινόμησης.

Python Code:

```
#Import Library
```

```
From sklearn.naive_bayes import GaussianNB
```

```
#Assumed you have, X(predictor) and Y(target) for training data set and x_test(predictor) of test_dataset
```

```
#Create SVM classification object model = GaussianNB() # there is other distribution for multinomial classes like Bernoulli Naïve Bayes, Refer link
```

```
#Train the model using the training sets and check score
```

```
Model.fit(X,y)
```

```
#Predict Output
```

```
Predicted= model.predict(x_test)
```

R code:

```
#Load Train and Test datasets
```

```
#Identify feature and response variable(s) and values must be numeric and numpy arrays
```

```
X_train<-input_variables_values_training_datasets
```

```
Y_train<-target_variables_values_training_datasets
```

```
x_test<-input_variables_values_test_datasets
```

```
library(e1071)
```

```
x<-cbind(x_train,y_train)
```

```
#Fitting model
```

```
Fit<-naïveBayes(y_train ~ ., data = x)
```

```
Summary(fit)
```

```
#Predict Output
```

```
Predicted = predict(fit,x_test)
```

2.6.6 Κ πλησιέστεροι γείτονες

Μπορεί να χρησιμοποιηθεί τόσο για προβλήματα ταξινόμησης όσο και για παλινδρόμηση.

Ωστόσο, χρησιμοποιείται ευρύτερα σε προβλήματα ταξινόμησης στον κλάδο. Κ πλησιέστεροι γείτονες είναι ένας απλός αλγόριθμος που αποθηκεύει όλες τις διαθέσιμες περιπτώσεις και ταξινομεί νέες περιπτώσεις με πλειοψηφία των γειτόνων του.

Η περίπτωση που ανατίθεται στην κλάση είναι πιο συνηθισμένη στους Κ πλησιέστερους γείτονές της που μετρούνται με μια συνάρτηση απόστασης.

Αυτές οι λειτουργίες απόστασης μπορούν να είναι η απόσταση Euclidean, Manhattan, Minkowski και Hamming. Οι πρώτες τρεις λειτουργίες χρησιμοποιούνται για τη συνεχή λειτουργία και η τέταρτη (Hamming) για τις κατηγορικές μεταβλητές. Αν $K = 1$, τότε η περίπτωση απλά ανατίθεται στην κατηγορία του πλησιέστερου γείτονα.

Κατά καιρούς, επιλέγοντας το K , είναι μια πρόκληση κατά την εκτέλεση μοντέλων kNN.

Python Code:

```
#Import Library
```

```
From sklearn.neighbors import KNeighborsClassifier
```

```
#Assumed you have, X(predictor) and Y(target) for training data set and x_test (predictor) of test_dataset
```

```
#CreateKNeighbors classifier object model
```

```
KNeighborsClassifier(n_neighbors=6) # default values for n_neighbors is 5
```

```
#Traing the model usingi the training sets and check score
```

```
Model.fit(X,y)
```

```
#Predict Output
```

```
Predicted=model.predict(x_test)
```

R Code:

```
#Load Train and Test datasets
```

```
#Identify feature and response variable(s) and values must be numeric and numpy arrays
```

```
X_train<-input_variables_values_training_datasets
```

```
Y_train<-target_variables_values_training_datasets
```

```
X_test<-input_variables_values_test_datasets
```

```

Library(knn)

X<-cbind(x_train,Y_train)

#Fitting model

Fit<-knn(y_train ~ ., data = x,k = 5)

Summary(fit)

#Predict Output

Predicted = predict(fit,x_test)

```

2.6.7 K-Means Clustering

Πρόκειται για έναν τύπο μη επιτηρούμενου αλγορίθμου ο οποίος επιλύει το πρόβλημα της ομαδοποίησης.

Η διαδικασία ακολουθεί έναν απλό και εύκολο τρόπο ταξινόμησης ενός δεδομένου συνόλου δεδομένων μέσω ενός ορισμένου αριθμού συμπλεγμάτων (υποθέστε k clusters).

Τα σημεία δεδομένων μέσα σε μια ομάδα είναι ομοιογενή και ετερογενή σε ομάδες ομοτίμων.

Πώς το K-means μορφοποιεί cluster:

1. Το K-means παίρνει k αριθμό σημείων για κάθε ομάδα γνωστά ως centroids.
2. Κάθε σημείο δεδομένων σχηματίζει ένα cluster με τα πλησιέστερα κεντροειδή δηλαδή ομάδες k.
3. Βρίσκει το centroid κάθε συμπλέγματος με βάση τα υπάρχοντα μέλη του συμπλέγματος. Εδώ έχουμε νέα κεντροειδή.
4. Δεδομένου ότι έχουμε νέα κεντροειδή, επαναλάβετε τα βήματα 2 και 3. Βρείτε την πλησιέστερη απόσταση για κάθε σημείο δεδομένων από νέα κεντροειδή και συνδέστε τα με τα νέα k-clusters. Επαναλάβετε αυτή τη διαδικασία έως ότου εμφανιστεί η σύγκλιση, δηλαδή τα κεντροειδή δεν αλλάζουν.

Python Code:

```
#Import Library
```

```

From sklearn.cluster import KMeans

#Assumed you have, X(attributes) for training data set and x_test(attributes) of test_dataset

#CreateKNeighbors classifier object model

K_means = KMeans(n_clusters=3, random_state=0)

#Train the model using the training sets and check score

Model.fit(X)

#Predict Output

Predicted = model.predict(x_test)

```

R Code:

```

#Load Train and Test datasets

#Identify feature and response variable(s) and values must be numeric and numpy arrays

X_train<-input_variables_values_training_datasets

Y_train<-target_variables_vlaues_training_datasets

X_test<-input_variables_values_test_datasets

Library(cluster)

Fit<-kmeans(X,3) #5 cluster solution

```

2.6.8 Τυχαίο Δάσος

Το τυχαίο δάσος είναι όρος που χρησιμοποιείται ως εμπορικό σήμα για ένα σύνολο δέντρων αποφάσεων.

Στο Τυχαίο Δάσος, έχουμε συλλογή δέντρων αποφάσεων (γνωστών ως "Δάσος"). Για να ταξινομήσετε ένα νέο αντικείμενο με βάση τα χαρακτηριστικά, κάθε δέντρο δίνει μια ταξινόμηση και λέμε ότι τα δέντρα ψήφισαν για αυτή την τάξη.

Το δάσος επιλέγει την ταξινόμηση με τις περισσότερες ψήφους (πάνω από όλα τα δέντρα στο δάσος).

Python Code:

```
#Import Library
```

```
From sklearn.ensemble import Random ForestClassifier
```

```
#Assumed you have, X(predictor) and ( target) for training data set and x_test(predictor) of test_dataset
```

```
#Create Random Forest object
```

```
Model=RandomForestClassifier()
```

```
#Train the model usin the training sets and check score
```

```
Model.fit(X,y)
```

```
#Predict Output
```

```
Predicted = model.predict(x_test)
```

R Code

```
#Load Train Tets datasets
```

```
#Identifyfeteature and response variables(s) and values must be numeric and numpy arrays
```

```
X_train<-inputvariables_values_training_datasets
```

```
Y_train<-target_varibales_values_training_datasets
```

```
X_test<-input_variables_values_test_datasets
```

```
Lbrary(randomForest)
```

```
X<-bind(x_train<y_trian)
```

```
#Fitting model
```

```
Fit<-randomForest(Species ~ ., x, ntree=500)
```

```
Summary(fit)
```

```
#Predict Output
```

```
Predicted=predict(fit,x_test)
```

2.6.9 Μείωση των διαστάσεων

Ως επιστήμονας δεδομένων, τα δεδομένα που προσφέρουμε επίσης αποτελούνται από πολλά χαρακτηριστικά, αυτό ακούγεται καλό για την οικοδόμηση ενός καλού ισχυρού μοντέλου, αλλά υπάρχει μια πρόκληση.

Πώς εντοπίσατε σημαντικές μεταβλητές από 1000 ή 2000;

Σε τέτοιες περιπτώσεις, ο αλγόριθμος μείωσης των διαστάσεων μας βοηθά μαζί με διάφορους άλλους αλγόριθμους, όπως Δέντρο Απόφασης, Random Forest, PCA, Ανάλυση Παράγοντα, Προσδιορισμός βασισμένος σε μήτρα συσχέτισης, missing value ratio και άλλοι.

Python code:

```
#Import Library
```

```
From sklearnimport decomposition
```

```
#Assumed you have training and test data set as train and test
```

```
#Cretae PCA object pca= decomposition,PCA(n_components=k) #default value of k=min (n_sample,  
n_features)
```

```
#For Factor analysis
```

```
#fa=decomposition.FactorAnalysis()
```

```
#Reduced the dimension of training dataset using PCA
```

```
Train_reduced=pca.fit_transform(train)
```

```
#Reduced the dimensionof test dataset
```

```
Test_reduced = pca.transform(test)
```

```
#For more detail on this, please refer this link.
```

R Code:

```
#Load Train and Test datasets
```

```
#Identify feature and response variable(s) and values must be numeric and numpy arrays.
```

```
X_train<- input_variables_values_training_datasets
```

```
Y_train<- target_variables_values_training_datasets
```

```
X_test<-input_variables_values_test_datasets
```

```
Library(stats)
```

```
Pca<-princomp(train, cor = TRUE)
```

```
Train_reduced<-predict(pca,train)
```

```
Test_reduced<-predict(pca,test)
```

2.6.10 Βελτίωση κλίσης

Εδώ θα συζητήσουμε 2 μεθόδους:

GBM

Το GBM είναι ένας αλγόριθμος ενίσχυσης που χρησιμοποιείται όταν ασχολούμαστε με πολλά δεδομένα για να κάνουμε μια πρόβλεψη με υψηλή δύναμη πρόβλεψης.

Η ενίσχυση είναι στην πραγματικότητα ένα σύνολο αλγορίθμων μάθησης που συνδυάζει την πρόβλεψη αρκετών εκτιμητών βάσης προκειμένου να βελτιωθεί η ευρωστία σε έναν ενιαίο εκτιμητή. Συνδυάζει πολλαπλούς αδύνατους ή μέσους προγνωστικούς παράγοντες για να βγάλει ισχυρό προγνωστικό παράγοντα.

XGBoost

Ένας άλλος κλασικός αλγόριθμος ενίσχυσης κλίσης που είναι γνωστός ως η αποφασιστική επιλογή ανάμεσα στη νίκη και την ήττα σε κάποιους διαγωνισμούς Kaggle.

Το XGBoost έχει μια εξαιρετικά υψηλή προγνωστική δύναμη που το καθιστά την καλύτερη επιλογή για την ακρίβεια στα γεγονότα, καθώς διαθέτει τόσο το γραμμικό μοντέλο όσο και τον αλγόριθμο εκμάθησης δέντρων, κάνοντας τον αλγόριθμο σχεδόν 10 φορές πιο γρήγορο από τις υπάρχουσες τεχνικές ενίσχυσης κλίσης.

Η υποστήριξη περιλαμβάνει διάφορες αντικειμενικές λειτουργίες, όπως παλινδρόμηση, ταξινόμηση και κατάταξη

Python Code:

```
Import Library from sklearn.ensemble import GradientBoostingClassifier

#Assumed you have, X(predictor) and Y(target)for training dat set and x_test(predictor) of test_dataset

#Create Gradient Boosting Classifier object

Model=Gradient Boosting Classifier(n_estimators=100, learning_rate=1.0, max_depth=1,
random_state=0)

#Train the model using the training sets and check score

Model.fit(X,y)

#Predict Output

Predicted=model.predict(x_test)
```

R Code:

```
#Load Train and Test datasets

#Identify feature and response variables and values must be numeric and numpy arrays

X_train<-input_variables_values_training_datasets

Y_train<-target_variables_values_training_datasets

X_test<-input_variables_values_test_datasets
```

```

Library(caret)

X<-cbind(x_train,y_train)

#Fitting model

fitControl<-trainControl(method="repeatedcv",number=4,repates=4)

fit<-train(y ~ ., data = x, method="gbm, trControl = fitControl, verbose = FALSE)

predicted=predict(fit,x_test,type="prob")[,2]

```

2.7 Επικύρωση αποτελέσματος

Το τελικό βήμα της ανακάλυψης γνώσης από δεδομένα είναι η επικύρωση των προτύπων που εξήχθησαν από τους αλγορίθμους της εξόρυξης δεδομένων που απευθύνονται σε ευρύτερο σύνολο δεδομένων. Δεν είναι όλα τα πρότυπα που βρέθηκαν απαραίτητα έγκυρα. Είναι συνηθισμένο για τους αλγορίθμους της εξόρυξης δεδομένων να βρίσκουν πρότυπα στο σύνολο εκπαίδευσης, τα οποία δεν υπάρχουν στο γενικό σύνολο δεδομένων. Αυτό καλείται υπερφόρτωση(overfitting). Για να ξεπεραστεί αυτό, στην εκτίμηση χρησιμοποιείται ένα δοκιμαστικό σύνολο δεδομένων στο οποίο δεν έχουν εφαρμοστεί οι αλγόριθμοι της εξόρυξης δεδομένων. Τα πρότυπα, που έχουν προκύψει, εφαρμόζονται σε αυτό το δοκιμαστικό σύνολο και το προκύπτον αποτέλεσμα συγκρίνεται με το επιθυμητό. Για παράδειγμα, ένας αλγόριθμος της εξόρυξης δεδομένων που ξεχωρίζει τα ανεπιθύμητα μηνύματα με τα "επιθυμητά" θα εφαρμοζόταν σε ένα σύνολο εκπαίδευσης από δείγματα ηλεκτρονικών μηνυμάτων. Μόλις εφαρμοζόταν, τα εξαχθείσα πρότυπα θα εφαρμόζονταν στο δοκιμαστικό σύνολο μηνυμάτων στο οποίο δεν είχε εφαρμοστεί πριν. Η ευστοχία αυτών των προτύπων μπορεί τώρα να μετρηθεί από τα πόσα μηνύματα έχουν καταταχθεί-ταξινομηθεί σωστά. Ένας αριθμός από στατιστικές μεθόδους μπορεί να χρησιμοποιηθεί για την αξιολόγηση του αλγορίθμου, όπως το ROC curves.

Αν τα πρότυπα δεν ανταποκρίνονται με τα επιθυμητά κριτήρια, τότε είναι απαραίτητο να εκτιμηθεί ξανά και να αλλαχθεί η προ-επεξεργασία και η εξόρυξη δεδομένων. Στην αντίθετη περίπτωση που ανταποκρίνονται με τα επιθυμητά κριτήρια, το τελικό στάδιο είναι να ερμηνευτούν τα πρότυπα και να τα μετατρέψουμε σε γνώση.

2.8 Εφαρμογές

2.8.1 Ιατρική

Τα τελευταία χρόνια, η εξόρυξη δεδομένων χρησιμοποιείται ευρέως στους τομείς της ιατρικής, όπως η βιοϊατρική, το DNA, η γενετική και η φαρμακευτική. Στον τομέα της γενετικής, ο σκοπός είναι να κατανοήσουμε την χαρτογράφηση της σχέσης μεταξύ της μεταβολής των ακολουθιών του ανθρώπινου DNA και την προδιάθεση στην αρρώστια. Η εξόρυξη δεδομένων είναι ένα σημαντικό εργαλείο που μπορεί να βοηθήσει στην βελτίωση της διάγνωσης, της πρόληψης και της θεραπείας των ασθενειών.

Εξαιτίας της αύξησης των βιοϊατρικών ερευνών, η μεγάλη κλίμακα γονιδιακών προτύπων και λειτουργιών πρέπει να εξετασθεί. Τα εργαλεία της εξόρυξης δεδομένων μπορούν να βοηθήσουν σε μεγάλο βαθμό για να μελετήσουμε την σύσταση του DNA και να βρούμε ποικίλα πρότυπα και λειτουργίες αυτού.

Ένας από τους κύριους στόχους που σχετίζεται με την ανάλυση δεδομένων του DNA είναι η σύγκριση ποικίλων ακολουθιών και η αναζήτηση ομοιοτήτων μεταξύ των δεδομένων του DNA. Η σύγκριση κυρίως περιλαμβάνει την γονιδιακή ακολουθία υγιών και βλαβερών ιστών για να βρει την διαφορά ανάμεσα σε αυτούς τους δύο τύπους. Αυτό μπορεί να επιτευχθεί ανακτώντας τις τάξεις υγιών αλλά και βλαβερών γονιδιακών ακολουθιών και μετά βρίσκοντας τις συχνά εμφανιζόμενες μορφές των δύο τάξεων. Αυτή η ανάλυση βοηθάει στο να βρίσκουμε τις ομοιότητες και τις διαφορές στις γενετικές ακολουθίες.

Στην βιοϊατρική, ερευνάται αν οι περισσότερες ασθένειες προκαλούνται από ένα συνδυασμό των γονιδίων. Η μέθοδος της συσχέτισης χρησιμοποιείται για να καθορίσει την συνύπαρξη ομάδων των γονιδίων και επίσης μπορούμε να εξετάσουμε την αλληλεπίδραση και την σχέση μεταξύ των γονιδίων.

Τα εργαλεία της οπτικοποίησης παίζουν επίσης ένα σημαντικό ρόλο στην εξόρυξη δεδομένων στην βιοϊατρική. Τα εργαλεία αυτά μπορούν να παρουσιάσουν πολύπλοκες δομές γονιδίων σε γράφους, δένδρα και αλυσίδες. Η οπτική παρουσίαση βοηθάει στην καλύτερη κατανόηση αυτών των δομών για ανακάλυψη γνώσης και εξερεύνηση των δεδομένων.

Υπάρχουν διάφοροι συνδυασμοί γονιδίων που συμβάλλουν στις ασθένειες, αλλά αυτά τα γονίδια ενεργοποιούνται σε διαφορετικά επίπεδα. Η ανάλυση μονοπατιού (path analysis) χρησιμοποιείται για να συνδέει διαφορετικά γονίδια με διαφορετικά στάδια κατά την εξέλιξη της ασθένειας. Η ανάλυση μονοπατιού διαδραματίζει ένα σπουδαίο ρόλο στην γενετική.

2.8.2 Οικονομία

Άλλος τομέας που εφαρμόζεται η εξόρυξη δεδομένων είναι η οικονομία. Τα οικονομικά δεδομένα κυρίως συλλέγονται από τράπεζες και από άλλους οικονομικούς οργανισμούς. Τα δεδομένα αυτά συνήθως είναι αξιόπιστα, ολοκληρωμένα και έχουν υψηλή ποιότητα και απαιτούν συστηματική μέθοδο για την ανάλυση αυτών. Η συνεισφορά της εξόρυξης δεδομένων στην επιστήμη της οικονομίας συναντάται στην συλλογή και κατανόηση των δεδομένων, στην βελτίωση δεδομένων (data refinement), στην δημιουργία και εκτίμηση ενός μοντέλου και στην ανάπτυξη αυτού. Η σωστή ανάλυση των οικονομικών δεδομένων μας διευκολύνει στο να παίρνουμε καλύτερες αποφάσεις ενεργώντας σύμφωνα με την ανάλυση της αγοράς. Τα εργαλεία και οι τεχνικές της εξόρυξης δεδομένων βοηθούν στο να αναλύσουμε τα οικονομικά δεδομένα με τους παρακάτω τρόπους:

Τα δεδομένα που συλλέγονται από διάφορα οικονομικά ιδρυτήματα, όπως οι τράπεζες, συγκεντρώνονται αρχικά στην αποθήκευση δεδομένων (data warehouse). Οι τεχνικές της πολυδιάστατης ανάλυσης δεδομένων χρησιμοποιούνται για την ανάλυση τέτοιων δεδομένων που συλλέγονται στην αποθήκη δεδομένων για τις γενικές ιδιότητές του.

Μία άλλη εφαρμογή της εξόρυξης δεδομένων σχετίζεται με την πρόβλεψη αποπληρωμής δανείου και πολιτικές πίστωσης του πελάτη. Μέθοδοι της εξόρυξης όπως η επιλογή χαρακτηριστικών (feature selection) βοηθάει στην ταυτοποίηση ποικίλων χαρακτηριστικών όπως το επίπεδο εισοδήματος του πελάτη, την εξόφληση ανάλογα με τα έσοδα, την πιστωτική του ιστορία κτλ. Με την επεξεργασία αυτών των χαρακτηριστικών, η τράπεζα μπορεί να αποφασίσει για τις πολιτικές δανειοδότησης βάσει των σχετικά χαμηλών κινδύνων. Οι τεχνικές της συσταδοποίησης και της ταξινόμησης βοηθούν τα οικονομικά ιδρυτήματα να ομαδοποιούν διάφορους πελάτες που έχουν κοινά χαρακτηριστικά. Η αποτελεσματική συσταδοποίηση και οι μέθοδοι φιλτραρίσματος βοηθούν τις τράπεζες να ταυτοποιούν μία ομάδα πελατών, να συσχετίζουν ένα νέο πελάτη με την παρούσα ομάδα και να τους παρέχουν κοινά οφέλη.

Τα εργαλεία της εξόρυξης δεδομένων βοηθούν τα οικονομικά ιδρυτήματα να αναγνωρίζουν τις απάτες και τα εγκλήματα από παραποιημένα δεδομένα από τις διάφορες βάσεις δεδομένων και από το ιστορικό συναλλαγών που έγιναν από τους πελάτες. Οι τεχνικές οπτικοποίησης βοηθούν στην παρουσίαση δεδομένων με διαφορετικές μορφές, όπως γράφοι που βασίζονται σε συγκεκριμένα γνωρίσματα. Προβάλλοντας τα δεδομένα από διάφορες οπτικές γωνίες, η τράπεζα δύναται να διακρίνει τους πελάτες που έχουν επιχειρήσει παράνομες πράξεις και μετά μια λεπτομερή έρευνα αυτών των ύποπτων περιπτώσεων βοηθάει στην εξιχνίαση των απατών και των εγκλημάτων.

2.8.3 Τηλεπικοινωνία

Η τηλεπικοινωνιακή βιομηχανία αναπτύσσεται πολύ γρήγορα όπως και η τεχνολογία. Αυτές τις μέρες οι τηλεπικοινωνιακές υπηρεσίες έχουν επεκταθεί από τοπικές και μεγάλης απόστασης τηλεπικοινωνίες, στην χρήση φαξ, συσκευές τηλεειδοποίησης, κινητό τηλέφωνο, και ηλεκτρονικό ταχυδρομείο. Εξαιτίας των εξελίξεων στις τηλεπικοινωνιακές τεχνολογίες και για να δουλέψουν αποτελεσματικά αυτές οι τεχνολογίες, οι τεχνικές της εξόρυξης δεδομένων ενσωματώνονται σε αυτές τις τεχνολογίες για να παράγουν αποδοτικά αποτελέσματα. Η εξόρυξη δεδομένων βοηθάει στην διάκριση τηλεπικοινωνιακών προτύπων, καταπολέμησης παράνομων δραστηριοτήτων, και επίσης βοηθάει στην καλύτερη χρήση των πόρων και στη βελτίωση της ποιότητας των υπηρεσιών. Η εξόρυξη δεδομένων βελτιώνει τις τηλεπικοινωνιακές υπηρεσίες με τους εξής τρόπους:

Τα τηλεπικοινωνιακά δεδομένα που συλλέγονται, περιλαμβάνουν τον τύπο κλήσης, την τοποθεσία του καλούντος και του κληθέντος, τον χρόνο κλήσης, την διάρκεια κλήσης κλπ. Η πολυδιάστατη ανάλυση βοηθά στον προσδιορισμό και στην σύγκριση του φορτίου του συστήματος, κίνηση δεδομένων, και κέρδος κλπ. Η ανάλυση μπορεί να δείξει διαγράμματα και γράφους των πόρων του συστήματος, του προορισμού κλπ κάνοντας χρήση των εργαλείων οπτικοποίησης της εξόρυξης δεδομένων. Τέτοια εργαλεία όπως η συσχετισμένη οπτικοποίηση και η συσταδοποίηση παρέχουν χρήσιμες υπηρεσίες στην ανάλυση των δεδομένων τηλεπικοινωνίας.

Το κυρίως πρόβλημα που αντιμετωπίστηκε από την βιομηχανία τηλεπικοινωνιών είναι οι παράνομες δραστηριότητες. Αυτές οι δραστηριότητες μπορεί να έχουν να κάνουν με σκόπιμες κλήσεις κατά την ώρα αιχμής, περιοδικές κλήσεις κ.α. με αποτέλεσμα να επιδρούν αρνητικά στην επίδοση του δικτύου επικοινωνιών. Μέθοδοι όπως η συσταδοποίηση και η ανάλυση ακραίων τιμών, συνεισφέρει στην ανίχνευση παράνομων προτύπων βελτιώνοντας την αποτελεσματικότητα των υπηρεσιών τηλεπικοινωνίας.

Εκμεταλλεούμενοι τα εργαλεία της εξόρυξης δεδομένων είναι δυνατή η δημιουργία προφίλ των πελατών και ο εντοπισμός βλαβών στο δίκτυο.

Τέλος, η ανάλυση συσχετιζόμενων και ακολουθιακών προτύπων ενθαρρύνει την προώθηση νέων και ποικίλων υπηρεσιών τηλεπικοινωνίας.

3 Οπτικοποίηση Δεδομένων

Η οπτικοποίηση των δεδομένων είναι η γραφική αναπαράσταση των δεδομένων. Περιλαμβάνει την παραγωγή εικόνων που επικοινωνούν τις σχέσεις μεταξύ των αντιπροσωπευτικών δεδομένων στους θεατές των εικόνων. Αυτή η επικοινωνία επιτυγχάνεται με τη χρήση μιας συστηματικής χαρτογράφησης μεταξύ των γραφικών σημείων και των τιμών των δεδομένων στη δημιουργία της απεικόνισης. Αυτή η χαρτογράφηση καθορίζει τον τρόπο με τον οποίο οι τιμές των δεδομένων θα εκπροσωπούνται οπτικά, προσδιορίζοντας τον τρόπο και σε ποιο βαθμό μια ιδιότητα ενός γραφικού σημείου, όπως το μέγεθος ή το χρώμα, θα αλλάξει ώστε να αντικατοπτρίζει την αλλαγή στην τιμή ενός στοιχείου.

Για την επικοινωνία των πληροφοριών με σαφήνεια και αποτελεσματικότητα, η οπτικοποίηση δεδομένων χρησιμοποιεί στατιστικά γραφικά, γραφικά, γραφικά πληροφοριών και άλλα εργαλεία. Τα αριθμητικά δεδομένα μπορούν να κωδικοποιούνται με κουκίδες, γραμμές ή ράβδους, για να επικοινωνούν οπτικά ένα ποσοτικό μήνυμα. Η αποτελεσματική απεικόνιση βοηθά τους χρήστες να αναλύουν και να αιτιολογούν τα δεδομένα και τα αποδεικτικά στοιχεία. Κάνει πολύπλοκα δεδομένα πιο προσιτά, κατανοητά και χρήσιμα. Οι χρήστες μπορούν να έχουν ιδιαίτερα αναλυτικά καθήκοντα, όπως να κάνουν συγκρίσεις ή να κατανοήσουν την αιτιότητα, και η αρχή σχεδιασμού του γραφικού (δηλ., Παρουσιάζει συγκρίσεις ή δείχνει αιτιότητα) ακολουθεί την εργασία. Οι πίνακες χρησιμοποιούνται γενικά όπου οι χρήστες θα αναζητήσουν μια συγκεκριμένη μέτρηση, ενώ διαγράμματα διαφόρων τύπων χρησιμοποιούνται για την εμφάνιση μοτίβων ή σχέσεων στα δεδομένα για μία ή περισσότερες μεταβλητές.

Η απεικόνιση των δεδομένων είναι τόσο τέχνη όσο και επιστήμη. Θεωρείται ως κλάδος περιγραφικών στατιστικών από ορισμένους, αλλά και ως εργαλείο ανάπτυξης θεμελιωμένων θεωριών από άλλους. Αυξημένα ποσά δεδομένων που δημιουργούνται από τη δραστηριότητα του Διαδικτύου και ένας αυξανόμενος αριθμός αισθητήρων στο περιβάλλον αναφέρονται ως "μεγάλα δεδομένα" ή Διαδίκτυο των πραγμάτων. Η επεξεργασία, η ανάλυση και η επικοινωνία αυτών των δεδομένων παρουσιάζουν δεοντολογικές και αναλυτικές προκλήσεις για την απεικόνιση δεδομένων. Το πεδίο της επιστήμης των δεδομένων και των επαγγελματιών που καλούνται επιστήμονες δεδομένων βοηθά στην αντιμετώπιση αυτής της πρόκλησης

3.1 Επισκόπηση

Η οπτικοποίηση δεδομένων αναφέρεται στις τεχνικές που χρησιμοποιούνται για την επικοινωνία δεδομένων ή πληροφοριών με την κωδικοποίησή τους ως οπτικά αντικείμενα (π.χ. σημεία, γραμμές ή γραμμές) που περιέχονται στα γραφικά. Ο στόχος είναι η επικοινωνία των πληροφοριών με σαφήνεια και

αποτελεσματικότητα στους χρήστες. Είναι ένα από τα βήματα στην ανάλυση δεδομένων ή την επιστήμη των δεδομένων. Σύμφωνα με τον Friedman (2008), ο «κύριος στόχος της απεικόνισης δεδομένων είναι η επικοινωνία των πληροφοριών με σαφήνεια και αποτελεσματικότητα μέσω γραφικών μέσων», δεν σημαίνει ότι η οπτικοποίηση δεδομένων πρέπει να φαίνεται βαρετή να είναι λειτουργική ή εξαιρετικά περίπλοκη για να φανεί όμορφη. , τόσο η αισθητική μορφή όσο και η λειτουργικότητα πρέπει να συμβαδίζουν, παρέχοντας πληροφορίες σχετικά με ένα αρκετά αραιό και πολύπλοκο σύνολο δεδομένων με την επικοινωνία των βασικών πτυχών της με έναν πιο διαισθητικό τρόπο. Ωστόσο, οι σχεδιαστές συχνά αποτυγχάνουν να επιτύχουν μια ισορροπία μεταξύ μορφής και λειτουργίας, δημιουργώντας πανέμορφα οπτικοποιήσεις δεδομένων που δεν εξυπηρετούν τον κύριο σκοπό τους - να κοινοποιούν πληροφορίες " .

Πράγματι, οι Fernanda Viegas και Martin M. Wattenberg πρότειναν ότι μια ιδανική οπτικοποίηση δεν πρέπει μόνο να επικοινωνεί με σαφήνεια, αλλά να διεγείρει τη συμμετοχή και την προσοχή των θεατών.

Η απεικόνιση δεδομένων σχετίζεται στενά με τα γραφικά πληροφοριών, την απεικόνιση των πληροφοριών, την επιστημονική απεικόνιση, την διερευνητική ανάλυση δεδομένων και τα στατιστικά γραφικά. Στη νέα χιλιετία, η απεικόνιση δεδομένων έχει καταστεί ενεργός τομέας έρευνας, διδασκαλίας και ανάπτυξης. Σύμφωνα με τους Post et al. (2002), έχει ενώσει επιστημονική και οπτική απεικόνιση.

3.2 Χαρακτηριστικά των αποτελεσματικών γραφικών απεικονίσεων

Ο καθηγητής Edward Tufte εξήγησε ότι οι χρήστες των οθονών πληροφοριών εκτελούν συγκεκριμένα αναλυτικά καθήκοντα όπως η πραγματοποίηση συγκρίσεων. Η αρχή σχεδιασμού του γραφικού πληροφοριών θα πρέπει να υποστηρίζει την αναλυτική εργασία. Όπως δείχνουν οι Γουίλιαμ Κλίβελαντ και Ρόμπερτ Μακίλ, διάφορα γραφικά στοιχεία το επιτελούν περισσότερο ή λιγότερο αποτελεσματικά. Για παράδειγμα, τα διαγράμματα τελεία και τα γραφήματα υπερβαίνουν τα διαγράμματα πίτας.

Στο βιβλίο του 1983 "Η οπτική απεικόνιση των ποσοτικών πληροφοριών", ο Edward Tufte ορίζει τις γραφικές απεικονίσεις και τις αρχές για την αποτελεσματική γραφική απεικόνιση στο ακόλουθο απόσπασμα: "Η αριστεία στα στατιστικά γραφικά αποτελείται από σύνθετες ιδέες που επικοινωνούν με σαφήνεια, ακρίβεια και αποδοτικότητα.

- εμφάνιση των δεδομένων
- παρακινεί τον θεατή να σκεφτεί την ουσία και όχι τη μεθοδολογία, τον γραφικό σχεδιασμό, την τεχνολογία της γραφικής παραγωγής ή κάτι άλλο
- να αποφευχθεί η στρέβλωση του τι πρέπει να πει τα δεδομένα
- παρουσιάστε πολλούς αριθμούς σε ένα μικρό χώρο
- καθιστούν συνεκτικά τα μεγάλα σύνολα δεδομένων
- ενθαρρύνετε το μάτι να συγκρίνει διαφορετικά στοιχεία
- αποκαλύπτουν τα δεδομένα σε διάφορα επίπεδα λεπτομέρειας, από μια ευρεία επισκόπηση έως την ωραία δομή
- εξυπηρετούν έναν εύλογα σαφή σκοπό: περιγραφή, εξερεύνηση, χαρτογράφηση ή διακόσμηση
- εξυπηρετούν έναν εύλογα σαφή σκοπό: περιγραφή, εξερεύνηση, χαρτογράφηση ή διακόσμηση
- να είναι στενά ενσωματωμένη στις στατιστικές και λεκτικές περιγραφές ενός συνόλου δεδομένων

Τα γραφικά αποκαλύπτουν δεδομένα. Πράγματι, τα γραφικά μπορούν να είναι πιο ακριβή και αποκαλυπτικά από τους συμβατικούς στατιστικούς υπολογισμούς. "

Για παράδειγμα, το διάγραμμα Minard δείχνει τις απώλειες που υπέστη ο στρατός του Ναπολέοντα κατά την περίοδο 1812-1813. Περιγράφονται έξι μεταβλητές: το μέγεθος του στρατού, η θέση του σε μια δισδιάστατη επιφάνεια (x και y), ο χρόνος, η κατεύθυνση της κίνησης και η θερμοκρασία. Το πλάτος γραμμής απεικονίζει μια σύγκριση (το μέγεθος του στρατού σε χρονικά σημεία) ενώ ο άξονας θερμοκρασίας υποδηλώνει μια αιτία της αλλαγής του μεγέθους του στρατού. Αυτή η πολυπαραγοντική απεικόνιση σε μια δισδιάστατη επιφάνεια αναφέρει μια ιστορία που μπορεί να αντιληφθεί αμέσως, ενώ αναγνωρίζει τα δεδομένα πηγής για να δημιουργήσει αξιοπιστία. Ο Tufte έγραψε το 1983 ότι: "Ίσως να είναι το καλύτερο στατιστικό γραφικό που έχει σχεδιαστεί ποτέ".

Η μη εφαρμογή αυτών των αρχών μπορεί να οδηγήσει σε παραπλανητικά γραφήματα, τα οποία διαστρεβλώνουν το μήνυμα ή στηρίζουν ένα εσφαλμένο συμπέρασμα. Σύμφωνα με τον Tufte, το chartjunk αναφέρεται στην ξένη εσωτερική διακόσμηση του γραφικού που δεν ενισχύει το μήνυμα, ή σε δωρεάν τρισδιάστατα ή προοπτικά αποτελέσματα. Διαχωρίζοντας άσκοπα το επεξηγηματικό κλειδί από την ίδια την εικόνα, απαιτώντας από το μάτι να ταξιδεύει εμπρός και πίσω από την εικόνα στο κλειδί, είναι μια μορφή "διοικητικών συντριμμάτων". Ο λόγος των "δεδομένων με μελάνη" θα πρέπει να μεγιστοποιηθεί, διαγράφοντας μελάνη χωρίς δεδομένα όπου αυτό είναι εφικτό.

Το Γραφείο Προϋπολογισμού του Κογκρέσου περιέγραψε διάφορες βέλτιστες πρακτικές για γραφικές απεικονίσεις σε παρουσίαση τον Ιούνιο του 2014. Αυτές περιλαμβάνουν: α) Γνωρίζοντας το ακροατήριό σας, β) Σχεδιασμός γραφικών που μπορούν να ανήκουν μόνο εκτός του πλαισίου της έκθεσης, και γ) Σχεδιασμός γραφικών που μεταδίδουν τα βασικά μηνύματα στην αναφορά.

3.3 Ποσοτικά μηνύματα

Ο συντάκτης Stephen Few περιγράφει οκτώ τύπους ποσοτικών μηνυμάτων που οι χρήστες ενδέχεται να επιχειρήσουν να κατανοήσουν ή να επικοινωνήσουν από ένα σύνολο δεδομένων και τα σχετικά γραφήματα που χρησιμοποιούνται για να βοηθήσουν στην επικοινωνία του μηνύματος:

1. Χρονολογικές σειρές: Μια ενιαία μεταβλητή καταγράφεται σε μια χρονική περίοδο, όπως το ποσοστό ανεργίας σε μια περίοδο 10 ετών. Μπορεί να χρησιμοποιηθεί ένα γράφημα γραμμής για να αποδειχθεί η τάση.
2. Κατάταξη: Οι κατηγοριοποιημένες υποδιαίρεσεις κατατάσσονται σε αύξουσα ή φθίνουσα σειρά, όπως η κατάταξη των επιδόσεων των πωλήσεων (το μέτρο) από τους πωλητές (η κατηγορία, με κάθε πωλητή κατηγοριοποιημένη υποδιαίρεση) κατά τη διάρκεια μιας μόνο περιόδου. Ένα διάγραμμα ράβδων μπορεί να χρησιμοποιηθεί για να δείξει τη σύγκριση μεταξύ των ατόμων πωλήσεων.
3. Μέρος σε ολόκληρο: Οι κατηγοριοποιημένες υποδιαίρεσεις μετρώνται ως αναλογία στο σύνολο (δηλ. Ένα ποσοστό από το 100%). Ένα διάγραμμα πίτας ή ένα διάγραμμα ράβδων μπορεί να δείξει τη σύγκριση των λόγων, όπως το μερίδιο αγοράς που αντιπροσωπεύουν οι ανταγωνιστές σε μια αγορά.
4. Απόκλιση: Οι κατηγοριοποιημένες υποδιαίρεσεις συγκρίνονται με μια αναφορά, όπως μια σύγκριση των πραγματικών έναντι των δαπανών του προϋπολογισμού για διάφορα τμήματα μιας επιχείρησης για μια δεδομένη χρονική περίοδο. Ένα διάγραμμα ράβδων μπορεί να δείξει σύγκριση του πραγματικού με το ποσό αναφοράς.
5. Κατανομή συχνότητας: Δείχνει τον αριθμό των παρατηρήσεων μιας συγκεκριμένης μεταβλητής για δεδομένο διάστημα, όπως ο αριθμός των ετών κατά τα οποία η χρηματιστηριακή απόδοση είναι μεταξύ διαστημάτων όπως 0-10%, 11-20%, κλπ. Ένα ιστόγραμμα, ένας τύπος διαγράμματος ράβδων, μπορεί να χρησιμοποιηθεί για την ανάλυση αυτή. Μια boxplot βοηθά στην απεικόνιση βασικών στατιστικών στοιχείων σχετικά με τη διανομή, όπως διάμεσος, τεταρτημόρια, ακραίες τιμές, κ.λπ.
6. Συσχέτιση: Σύγκριση μεταξύ παρατηρήσεων που παριστάνονται από δύο μεταβλητές (X, Y) για να προσδιοριστεί εάν τείνουν να κινούνται στις ίδιες ή αντίθετες κατευθύνσεις. Για παράδειγμα, σχεδιάζουμε την ανεργία (X) και τον πληθωρισμό (Y) για ένα δείγμα μηνών. Για αυτό το μήνυμα χρησιμοποιείται συνήθως μια γραφική παράσταση σκέδασης.

7. Ονομαστική σύγκριση: Συγκρίνοντας κατηγορηματικές υποδιαίρεσεις σε καμία συγκεκριμένη σειρά, όπως ο όγκος των πωλήσεων ανά κωδικό προϊόντος. Για τη σύγκριση αυτή μπορεί να χρησιμοποιηθεί ένα διάγραμμα ράβδων.

8. Γεωγραφική ή γεωχωρική: Σύγκριση μιας μεταβλητής σε ένα χάρτη ή διάταξη, όπως το ποσοστό ανεργίας ανά κράτος ή ο αριθμός των ατόμων στους διάφορους ορόφους ενός κτιρίου. Ένα χαρτογράφημα είναι ένα τυπικό γραφικό που χρησιμοποιείται.

Οι αναλυτές που εξετάζουν ένα σύνολο δεδομένων ενδέχεται να εξετάσουν εάν ορισμένα ή όλα τα παραπάνω μηνύματα και τύποι γραφικών ισχύουν για το έργο και το κοινό τους. Η διαδικασία δοκιμής και σφάλματος για τον εντοπισμό σημαντικών σχέσεων και μηνυμάτων στα δεδομένα αποτελεί μέρος της διερευνητικής ανάλυσης δεδομένων

3.4 Οπτική αντίληψη και οπτικοποίηση δεδομένων

Ένας άνθρωπος μπορεί να διακρίνει εύκολα τις διαφορές στο μήκος, το σχήμα, τον προσανατολισμό και το χρώμα (απόχρωση) χωρίς σημαντική προσπάθεια επεξεργασίας. Αυτά αναφέρονται ως "προ-προσεκτικά χαρακτηριστικά". Για παράδειγμα, μπορεί να απαιτηθεί σημαντικός χρόνος και προσπάθεια ("προσεκτική επεξεργασία") για τον προσδιορισμό του αριθμού των ψηφίων "5" που εμφανίζονται σε μια σειρά αριθμών. αλλά εάν το ψηφίο αυτό έχει διαφορετικό μέγεθος, προσανατολισμό ή χρώμα, οι περιπτώσεις του ψηφίου μπορούν να σημειωθούν γρήγορα μέσω της προ-προσεκτικής επεξεργασίας.

Τα αποτελεσματικά γραφικά εκμεταλλεύονται την προ-προσεκτική επεξεργασία και τα χαρακτηριστικά και τη σχετική ισχύ αυτών των χαρακτηριστικών. Για παράδειγμα, επειδή ο άνθρωπος μπορεί να επεξεργάζεται ευκολότερα τις διαφορές στο μήκος της γραμμής από την επιφάνεια, μπορεί να είναι πιο αποτελεσματικό να χρησιμοποιείτε ένα διάγραμμα ράβδων (το οποίο εκμεταλλεύεται το μήκος της γραμμής για να δείξει σύγκριση) και όχι πίνακες πίτας).

Ανθρώπινη αντίληψη / γνώση και οπτικοποίηση δεδομένων

Σχεδόν όλες οι οπτικοποιήσεις δεδομένων δημιουργούνται για ανθρώπινη κατανάλωση. Η γνώση της ανθρώπινης αντίληψης και της γνώσης είναι απαραίτητη κατά το σχεδιασμό διαισθητικών απεικονίσεων. Η γνώση αναφέρεται στις διαδικασίες σε ανθρώπους όπως η αντίληψη, η προσοχή, η μάθηση, η μνήμη, η σκέψη, η σύλληψη ιδεών, η ανάγνωση και η επίλυση προβλημάτων. Η ανθρώπινη οπτική επεξεργασία είναι αποτελεσματική στην αντίληψη αλλαγών και στην πραγματοποίηση συγκρίσεων μεταξύ των ποσοτήτων, των μεγεθών, των σχημάτων και των παραλλαγών της ελαφρότητας. Όταν οι ιδιότητες των συμβολικών δεδομένων χαρτογραφούνται στις οπτικές ιδιότητες, ο άνθρωπος μπορεί να αναζητήσει

αποτελεσματικά μεγάλες ποσότητες δεδομένων. Εκτιμάται ότι τα 2/3 των νευρώνων του εγκεφάλου μπορούν να συμμετέχουν στην οπτική επεξεργασία. Η σωστή απεικόνιση παρέχει μια διαφορετική προσέγγιση για να δείξει πιθανές συνδέσεις, σχέσεις κ.λπ. οι οποίες δεν είναι τόσο προφανείς σε μη οπτικοποιημένα ποσοτικά δεδομένα. Η απεικόνιση μπορεί να γίνει μέσω έρευνας.

3.5 Ιστορία της οπτικοποίησης δεδομένων

Δεν υπάρχει πλήρης «ιστορία» απεικόνισης δεδομένων. Δεν υπάρχουν λογαριασμοί που να καλύπτουν ολόκληρη την εξέλιξη της οπτικής σκέψης και της οπτικής αναπαράστασης των δεδομένων και να συνδυάζουν τις συνεισφορές των διαφόρων κλάδων. Ο Michael Friendly και ο Daniel J Denis του Πανεπιστημίου της Υόρκης ασχολούνται με ένα έργο που επιχειρεί να προσφέρει μια ολοκληρωμένη ιστορία απεικόνισης. Σε αντίθεση με τη γενική πεποίθηση, η απεικόνιση δεδομένων δεν είναι μια σύγχρονη εξέλιξη. Τα αστρικά δεδομένα ή οι πληροφορίες όπως η θέση των αστεριών απεικονίστηκαν στους τοίχους των σπηλαίων (όπως αυτές που βρέθηκαν στο σπήλαιο του Lascaux στη νότια Γαλλία) από την εποχή του Πλειστόκαινου. Τα φυσικά αντικείμενα όπως τα μεσοποταμιακά πήλινα μάρκες (5500 π.Χ.), το Inca quipus (2600 π.Χ.) και τα διαγράμματα ραβδίων των νήσων Μάρσαλ (n.d.) μπορούν επίσης να θεωρηθούν ως απεικονιστικές ποσοτικές πληροφορίες.

Η πρώτη τεκμηριωμένη απεικόνιση δεδομένων μπορεί να εντοπιστεί πίσω στο 1160 π.Χ. με χάρτη του Τορίνου Papyrus, ο οποίος απεικονίζει με ακρίβεια την κατανομή των γεωλογικών πόρων και παρέχει πληροφορίες για την εξόρυξη αυτών των πόρων. Τέτοιοι χάρτες μπορούν να κατηγοριοποιηθούν ως Θεματική Χαρτογραφία, η οποία είναι ένας τύπος απεικόνισης δεδομένων που παρουσιάζει και κοινοποιεί συγκεκριμένα δεδομένα και πληροφορίες μέσω μιας γεωγραφικής απεικόνισης σχεδιασμένης να δείχνει ένα συγκεκριμένο θέμα που συνδέεται με μια συγκεκριμένη γεωγραφική περιοχή. Οι πρώτες τεκμηριωμένες μορφές απεικόνισης δεδομένων ήταν διάφοροι θεματικοί χάρτες από διαφορετικούς πολιτισμούς και ιδεογράμματα και ιερογλυφικά που παρείχαν και επέτρεψαν την ερμηνεία των πληροφοριών που απεικονίζονται. Για παράδειγμα, τα Γραμμικά Β δίσκια των Μυκηνών έδωσαν μια απεικόνιση πληροφοριών σχετικά με τις εποχές της εποχής της Ύστερης Εποχής του Χαλκού στη Μεσόγειο. Η ιδέα των συντεταγμένων χρησιμοποιήθηκε από τους αρχαίους Αιγύπτιους επιθεωρητές για την τοποθέτηση πόλεων, οι γήινες και ουράνιες θέσεις βρισκόταν από κάτι παρόμοιο με το γεωγραφικό πλάτος και το γεωγραφικό μήκος τουλάχιστον κατά το 200 π.Χ. και η προβολή του χάρτη μιας σφαιρικής γης σε γεωγραφικό πλάτος και γεωγραφικό μήκος από τον Κλαύδιο Πτολεμαίο [c.85-c. 165] στην Αλεξάνδρεια θα χρησίμευε ως πρότυπα αναφοράς μέχρι τον 14ο αιώνα.

Η εφεύρεση χαρτιού και περγαμινών επέτρεψε την περαιτέρω ανάπτυξη απεικονίσεων σε όλη την ιστορία. Το σχήμα δείχνει ένα γράφημα από τον 10ο, ενδεχομένως τον 11ο αιώνα, που προορίζεται να είναι μια απεικόνιση του πλανητικού κινήματος, που χρησιμοποιείται σε ένα παράρτημα ενός εγχειριδίου στα μοναστικά σχολεία. Το γράφημα προφανώς προοριζόταν να αντιπροσωπεύει μια γραφική παράσταση των κλίσεων των πλανητικών τροχιών ως συνάρτηση του χρόνου. Για το σκοπό αυτό η ζώνη του ζωδιακού κύκλου απεικονίστηκε σε ένα επίπεδο με οριζόντια γραμμή χωρισμένη σε τριάντα μέρη ως χρονικό ή διαμήκη άξονα. Ο κάθετος άξονας προσδιορίζει το πλάτος του ζωδιακού κύκλου. Η οριζόντια κλίμακα φαίνεται να έχει επιλεγεί για κάθε πλανήτη ξεχωριστά για τις περιόδους που δεν μπορούν να συμφιλωθούν. Το συνοδευτικό κείμενο αναφέρεται μόνο στα πλάτη. Οι καμπύλες προφανώς δεν σχετίζονται εγκαίρως.

Πλανητικές κινήσεις

Μέχρι τον 16ο αιώνα, αναπτύχθηκαν καλά τεχνικές και όργανα για ακριβή παρατήρηση και μέτρηση φυσικών μεγεθών και γεωγραφική και ουράνια θέση (για παράδειγμα, ένα "τεταρτημόριο τοίχου" που κατασκευάστηκε από τον Tycho Brahe [1546-1601] το παρατηρητήριο του). Ιδιαίτερα σημαντική ήταν η ανάπτυξη τριγωνισμού και άλλων μεθόδων για τον ακριβή καθορισμό των θέσεων χαρτογράφησης.

Ο γάλλος φιλόσοφος και μαθηματικός René Descartes και Pierre de Fermat ανέπτυξαν αναλυτική γεωμετρία και δισδιάστατο σύστημα συντεταγμένων που επηρέασαν σε μεγάλο βαθμό τις πρακτικές μεθόδους εμφάνισης και υπολογισμού των αξιών. Οι εργασίες Fermat και Blaise Pascal σχετικά με τις στατιστικές και τη θεωρία των πιθανοτήτων έθεσαν τις βάσεις για αυτό που τώρα θεωρούμε ως δεδομένα. Σύμφωνα με το Ίδρυμα Σχεδίασης της Αλληλεπίδρασης, αυτές οι εξελίξεις επέτρεψαν και βοήθησαν τη William Playfair, που είδε την δυνατότητα γραφικής επικοινωνίας των ποσοτικών δεδομένων, να παράγει και να αναπτύσσει γραφικές μεθόδους στατιστικών.

Playfair TimeSeries

Στο δεύτερο μισό του 20ού αιώνα, ο Jacques Bertin χρησιμοποίησε ποσοτικά γραφήματα για να αναπαριστά πληροφορίες "δισαισθητικά, καθαρά, με ακρίβεια και αποτελεσματικά".

Ο John Tukey και ο Edward Tufte έσπρωξαν τα όρια της απεικόνισης δεδομένων. Ο Tukey με τη νέα στατιστική του προσέγγιση για διερευνητική ανάλυση δεδομένων και ο Tufte με το βιβλίο του "Η οπτική απεικόνιση των ποσοτικών πληροφοριών" άνοιξε το δρόμο για τη βελτίωση των τεχνικών οπτικοποίησης δεδομένων για περισσότερους από τους στατιστικολόγους. Με την εξέλιξη της τεχνολογίας ήρθε η εξέλιξη της οπτικοποίησης δεδομένων? ξεκινώντας με απεικονίσεις με το χέρι και εξελισσόμενες σε πιο

τεχνικές εφαρμογές - συμπεριλαμβανομένων διαδραστικών σχεδίων που οδηγούν στην απεικόνιση του λογισμικού.

Προγράμματα όπως το SAS, SOFA, R, Minitab, Cornerstone και περισσότερα επιτρέπουν την απεικόνιση δεδομένων στον τομέα των στατιστικών. Άλλες εφαρμογές οπτικοποίησης δεδομένων, πιο εστιασμένες και μοναδικές για τα άτομα, οι γλώσσες προγραμματισμού όπως το D3, το Python και το JavaScript συμβάλλουν στη δυνατότητα απεικόνισης των ποσοτικών δεδομένων. Τα ιδιωτικά σχολεία έχουν επίσης αναπτύξει προγράμματα για να καλύψουν τη ζήτηση για οπτικοποίηση μαθησιακών δεδομένων και τις σχετικές βιβλιοθήκες προγραμματισμού, συμπεριλαμβανομένων δωρεάν προγραμμάτων όπως το The Incubator δεδομένων ή τα προγράμματα που πληρώνονται όπως η Γενική Συνέλευση.

Ξεκινώντας από το Συμπόσιο "Data to Discovery" το 2013, το ArtCenter College of Design, Caltech και JPL στην Pasadena διοργάνωσαν ένα ετήσιο πρόγραμμα για την απεικόνιση διαδραστικών δεδομένων. Το πρόγραμμα ρωτά: Πώς μπορεί η διαδραστική απεικόνιση δεδομένων να βοηθήσει τους επιστήμονες και τους μηχανικούς να διερευνήσουν καλύτερα τα δεδομένα τους; Πώς μπορεί η υπολογιστική, ο σχεδιασμός και ο σχεδιασμός να συμβάλλουν στη μεγιστοποίηση των αποτελεσμάτων της έρευνας; Ποιες μεθοδολογίες είναι πιο αποτελεσματικές για την αξιοποίηση της γνώσης από αυτούς τους τομείς; Με την κωδικοποίηση σχεσιακών πληροφοριών με τα κατάλληλα οπτικά και διαδραστικά χαρακτηριστικά για να βοηθήσουν στην ανάκριση και τελικά να αποκτήσουν νέα γνώση των δεδομένων, το πρόγραμμα αναπτύσσει νέες διεπιστημονικές προσεγγίσεις σε πολύπλοκα επιστημονικά προβλήματα, αξιοποιώντας τη σκέψη σχεδιασμού και τις τελευταίες μεθόδους υπολογιστικής, και τρισδιάστατα γραφικά.

3.6 Ορολογία

Η απεικόνιση δεδομένων περιλαμβάνει συγκεκριμένη ορολογία, μερικές από τις οποίες προέρχονται από στατιστικές. Για παράδειγμα, ο συντάκτης Stephen Few ορίζει δύο τύπους δεδομένων, τα οποία χρησιμοποιούνται σε συνδυασμό για να υποστηρίξουν μια σημαντική ανάλυση ή οπτικοποίηση:

- Κατηγορία: Ετικέτες κειμένου που περιγράφουν τη φύση των δεδομένων, όπως "Όνομα" ή "Ηλικία". Ο όρος αυτός καλύπτει επίσης ποιοτικά (μη αριθμητικά) δεδομένα.
- Ποσοτικά: Αριθμητικά μέτρα, όπως "25" για να αντιπροσωπεύουν την ηλικία σε χρόνια.

Δύο κύριοι τύποι προβολών πληροφοριών είναι πίνακες και γραφήματα.

- Ένας πίνακας περιέχει ποσοτικά δεδομένα οργανωμένα σε σειρές και στήλες με κατηγοριοποιημένες ετικέτες. Χρησιμοποιείται κυρίως για την αναζήτηση συγκεκριμένων τιμών. Στο παραπάνω παράδειγμα, ο πίνακας μπορεί να έχει κατηγοριοποιημένες ετικέτες στηλών που αντιπροσωπεύουν το όνομα (ποιοτική μεταβλητή) και την ηλικία (μια ποσοτική μεταβλητή), με κάθε σειρά δεδομένων να αντιπροσωπεύει ένα άτομο (την πειραματική μονάδα ή την υποδιαίρεση κατηγορίας).
- Ένα γράφημα χρησιμοποιείται κυρίως για την εμφάνιση σχέσεων μεταξύ δεδομένων και εικόνων που κωδικοποιούνται ως οπτικά αντικείμενα (π.χ. γραμμές, γραμμές ή σημεία). Οι αριθμητικές τιμές εμφανίζονται σε μια περιοχή που οριοθετείται από έναν ή περισσότερους άξονες. Αυτοί οι άξονες παρέχουν κλίμακες (ποσοτικές και κατηγορικές) που χρησιμοποιούνται για την επισήμανση και την εκχώρηση αξιών στα οπτικά αντικείμενα. Πολλά γραφήματα αναφέρονται επίσης ως γραφήματα

Οι Erppler και Lengler έχουν αναπτύξει τον "Περιοδικό πίνακα μεθόδων απεικόνισης", ένα διαδραστικό χάρτη που παρουσιάζει διάφορες μεθόδους οπτικοποίησης δεδομένων. Περιλαμβάνει έξι τύπους μεθόδων απεικόνισης δεδομένων: δεδομένα, πληροφορίες, έννοια, στρατηγική, μεταφορά και ένωση.

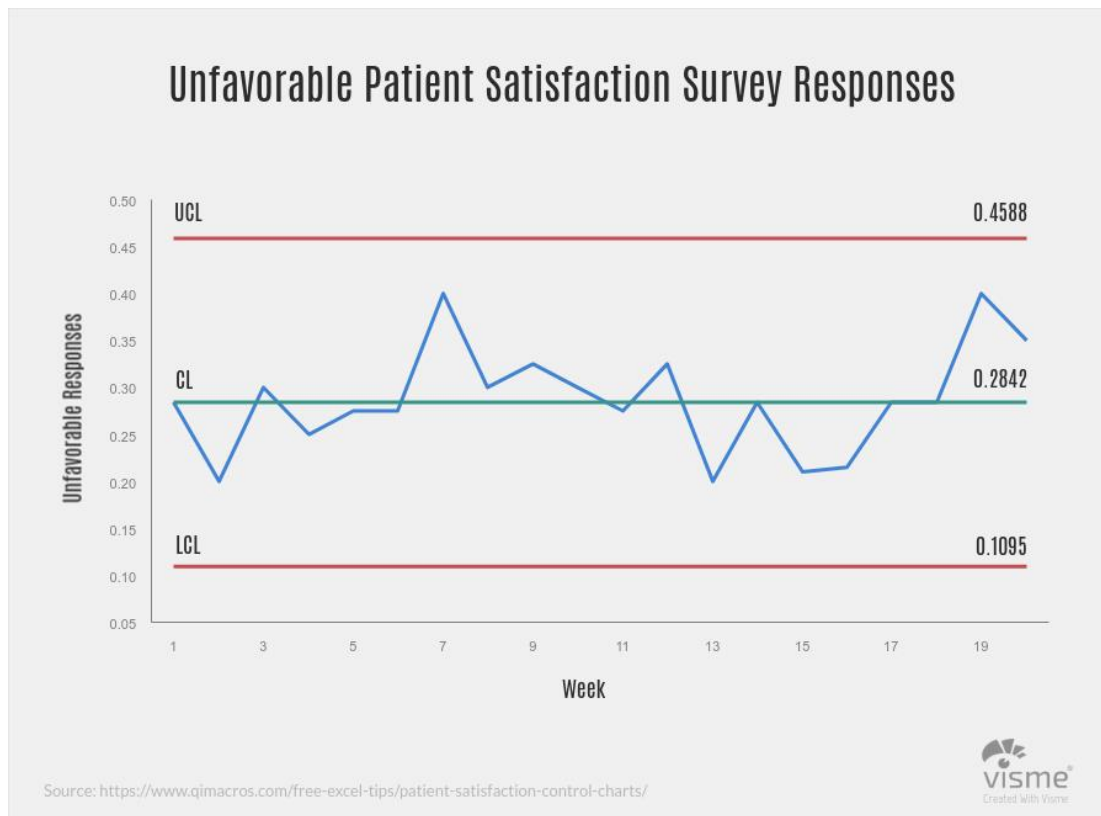
3.7 Παραδείγματα διαγραμμάτων που χρησιμοποιούνται για την απεικόνιση δεδομένων

3.7.1 Επιχειρήσεις και οικονομία

Control Charts

Γνωστό και ως process-behavior chart, το control chart μας βοηθάει να καθορίσουμε αν το data set εντάσσεται σε κάποιον μέσο ή προκαθορισμένο εύρος ελέγχου. Συχνά χρησιμοποιούμενο σε διαδικασίες ελέγχου ποιότητας, ένα τυπικό διάγραμμα ελέγχου αποτελείται από σημεία σχεδιασμένα σε δύο άξονες, αναπαριστώντας δείγματα μετρήσεων.

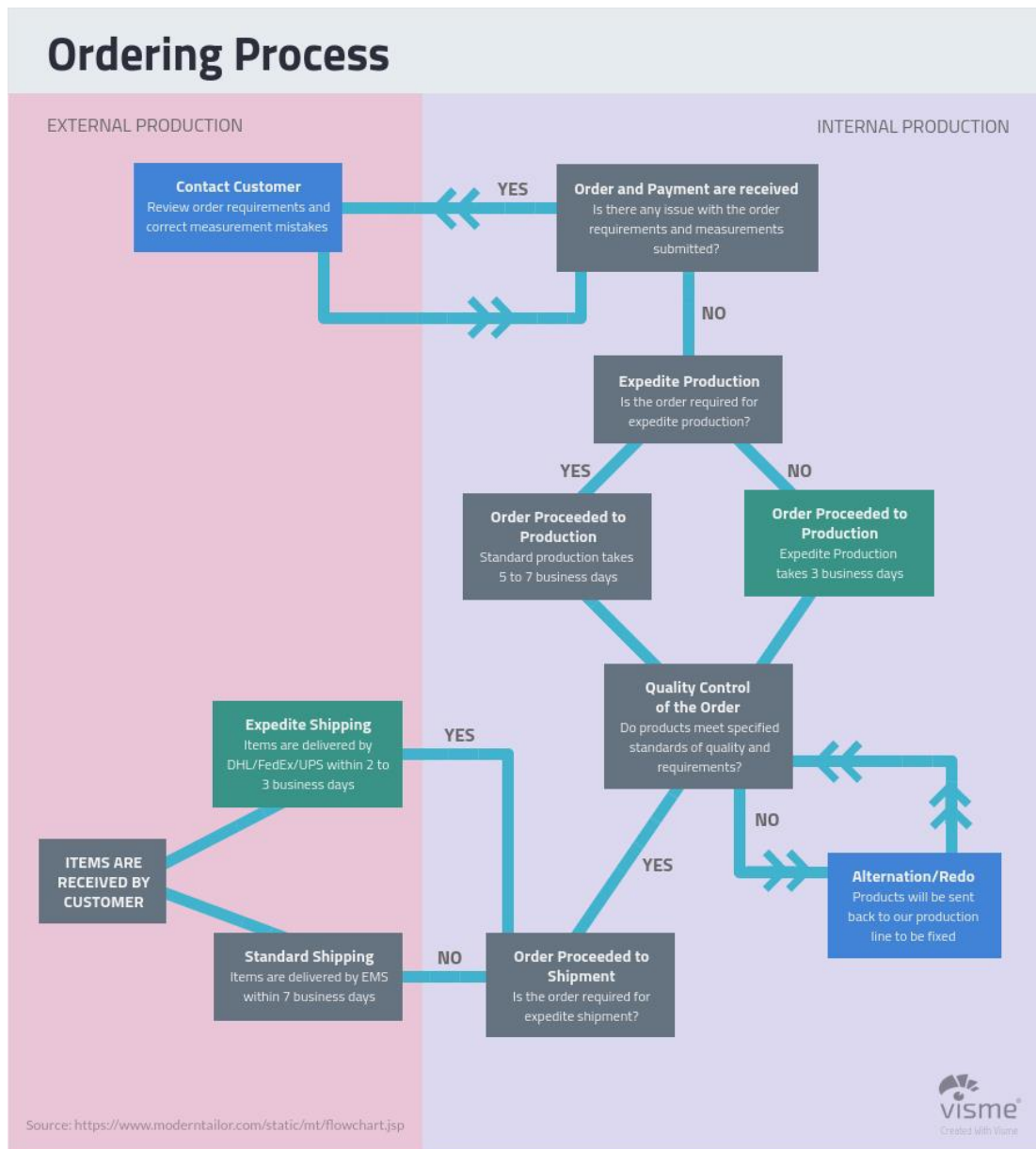
Ο μέσος του κάθε σημείου και μία ευθεία γραμμή διαμέσου του γράφου στο μέσο υπολογίζονται. Μετά υπολογίζετε η τυπική απόκλιση χρησιμοποιώντας όλα τα δείγματα. Στο τέλος καθορίζονται το άνω και κάτω όρια ελέγχου και διαγράφονται για να αντικατοπτρίζουν τα σημεία στα οποία η απόκλιση ξεπερνά την αναμενόμενη.



Flow Charts

Πολλές φορές στις επιχειρήσεις - καθώς και σε άλλες βιομηχανίες - μια διαδικασία πρέπει να είναι διάγραμμα. Ένα διάγραμμα ροής επιτρέπει τη διεξαγωγή μιας διαδικασίας από το ένα στο άλλο, από την αρχή μέχρι το τέλος, με σκοπό την ανάλυση, το σχεδιασμό, την τεκμηρίωση ή τη διαχείριση του. Τα διαγράμματα μπορούν ακόμη να διαθέτουν πολλαπλές αρχές και τελειώματα, με αμέτρητα μονοπάτια μεταξύ τους.

Ενώ ένα απλό διάγραμμα ροής μπορεί σίγουρα να τεκμηριώνει μια βασική διαδικασία από το Α στο Β έως το C, τα διαγράμματα χρησιμοποιούνται συχνότερα για να απεικονίσουν πιο σύνθετες ακολουθίες με πολλαπλές αποφάσεις ή συνθήκες στη διαδρομή. Κάθε φορά που πληρούται μια συνθήκη, το γράφημα διαγράφει τις διάφορες επιλογές και στη συνέχεια η διαδρομή συνεχίζεται μετά από κάθε επιλογή.



Gantt charts

Τα γραφήματα Gantt είναι ειδικοί τύποι γραφημάτων γραμμών που χρησιμοποιούνται για τη διαγράμμιση έργων και χρονοδιαγραμμάτων. Η χρήση χρωματισμένων ράβδων ποικίλου μήκους αντικατοπτρίζει όχι μόνο τις ημερομηνίες έναρξης και λήξης του έργου, αλλά και σημαντικά γεγονότα, καθήκοντα, ορόσημα και χρονικά πλαίσια.

Τα σύγχρονα γραφήματα Gantt μπορούν επίσης να απεικονίσουν τις σχέσεις εξάρτησης των δραστηριοτήτων. Εάν η συμπλήρωση της εργασίας της ομάδας 3, για παράδειγμα, εξαρτάται από την

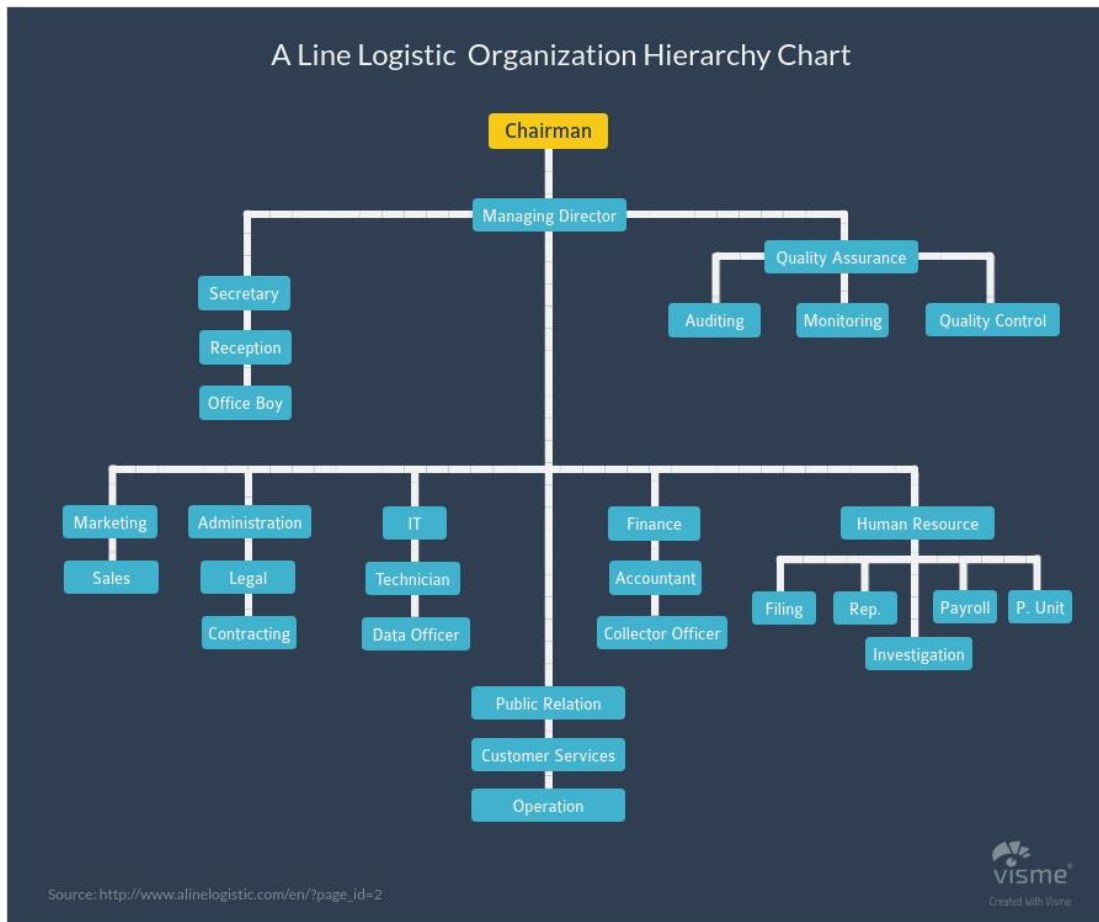
προηγούμενη ολοκλήρωση της εργασίας Β από την ομάδα 2, ο πίνακας δεν μπορεί μόνο να αντικατοπτρίζει αυτή τη σχέση, αλλά τις προγραμματισμένες ημερομηνίες και τις προθεσμίες για κάθε μία.



Διαγράμματα ιεραρχίας

Παρόμοια σε εμφάνιση με ένα διάγραμμα ροής, ένα ιεραρχικό διάγραμμα, επίσης γνωστό ως οργανόγραμμα, απεικονίζει τη δομή μιας οργάνωσης καθώς και τις σχέσεις μέσα σε αυτήν. Ένα τυπικό οργανικό σχήμα της εταιρείας, για παράδειγμα, απαριθμεί τον διευθύνοντα σύμβουλο στην κορυφή, ακολουθούμενο από πρόεδροι, αντιπρόεδροι, διευθυντές κ.ο.κ.

Ένα οργανόγραμμα μπορεί να απεικονίσει την αλυσίδα διοίκησης από οποιονδήποτε υπάλληλο σε όλη τη διαδρομή προς την κορυφή. Τα ιεραρχικά διαγράμματα χρησιμοποιούνται παρομοίως για να αντιπροσωπεύουν τα γενεαλογικά χαρακτηριστικά, τις επιστημονικές ταξινομήσεις, τα δημογραφικά στοιχεία και κάθε σύνολο δεδομένων με παρόμοια ανάλυση.



Stock charts

Ένα από τα πιο ζωτικής σημασίας από όλα τα οικονομικά γραφήματα, οι μετοχικοί χάρτες βοηθούν τους επενδυτές να παρακολουθούν τις αγορές για να καθορίσουν τα κέρδη και τις απώλειες, καθώς και να λαμβάνουν αποφάσεις αγοράς και πώλησης. Παρόλο που μια ποικιλία γραφημάτων χρησιμοποιείται για να αντιπροσωπεύει τις αλλαγές της αγοράς, το πιο συνηθισμένο είναι πιθανό το βασικό γράφημα γραμμής να μετατρέπεται σε ιστόγραμμα.

Οι γραμμές παρακολουθούν απλώς τις αλλαγές στην αξία ενός συγκεκριμένου αποθέματος ή της συνολικής αγοράς για μια χρονική περίοδο. Πολλαπλά αποθέματα μπορούν να παρακολουθούνται και να συγκρίνονται ταυτόχρονα μετασηματίζοντας το γράφημα γραμμής σε ένα διάγραμμα στοιβαγμένων περιοχών ή απλά χρησιμοποιώντας πολλαπλές γραμμές διαφόρων χρωμάτων.



Charts καταρράκτη

Ιδιαίτερα χρήσιμοι στη λογιστική και στην ποιοτική ανάλυση, τα διαγράμματα καταρράκτη δείχνουν πώς επηρεάζεται θετικά και αρνητικά μια αρχική τιμή από διάφορους παράγοντες. Για παράδειγμα, ένα διάγραμμα καταρράκτη θα μπορούσε να ανακοινώσει με σαφήνεια και αποτελεσματικότητα τον τρόπο με τον οποίο ένα αρχικό υπόλοιπο αλλάζει κάθε μήνα κατά τη διάρκεια ενός έτους.

Επειδή συχνά εμφανίζονται σαν να ρέουν ράβδοι σε όλο το γράφημα, τα διαγράμματα καταρράκτη μερικές φορές αναφέρονται ως πλωτά τούβλα ή διαγράμματα Mario.

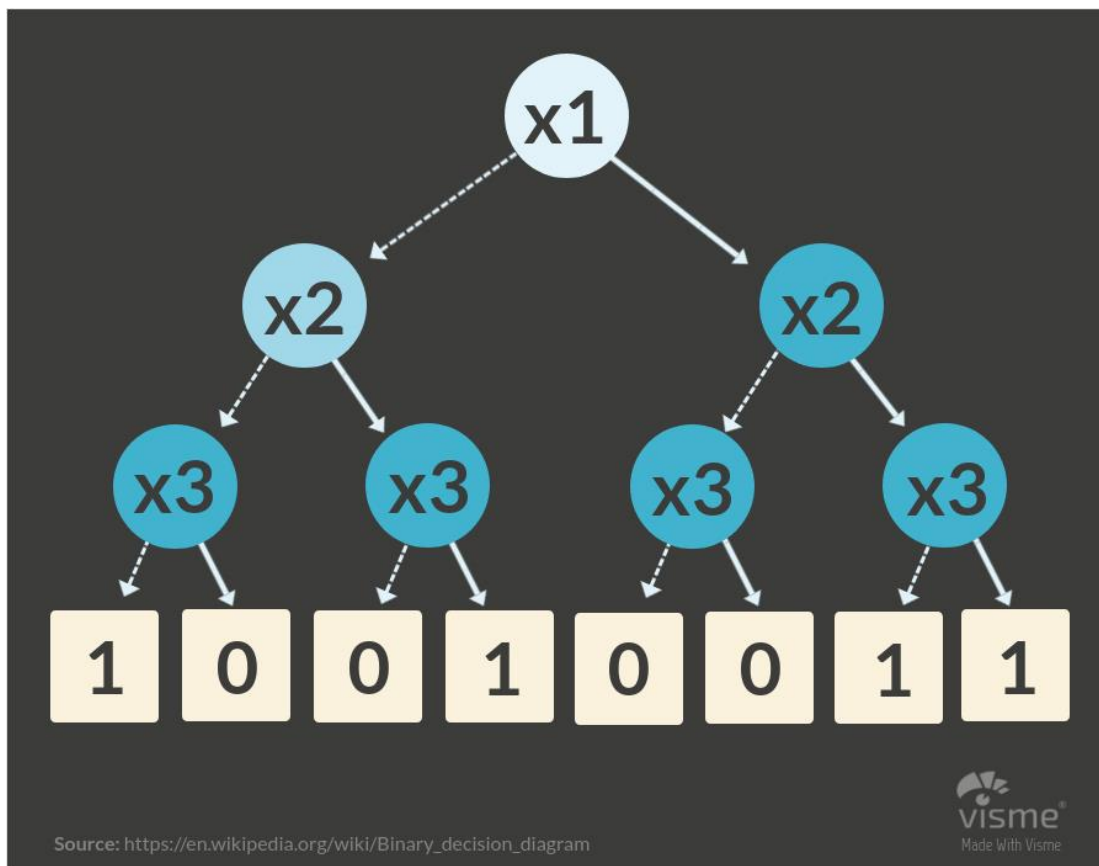


3.7.2 Γραφήματα μηχανικής και τεχνολογίας

Διαδικά διαγράμματα αποφάσεων

Μια δυαδική απόφαση είναι μια επιλογή μεταξύ δύο εναλλακτικών λύσεων, έτσι ένα διάγραμμα δυαδικής απόφασης απεικονίζει το μονοπάτι από τη μία απόφαση στην άλλη. Στην επιστήμη των υπολογιστών, οι δυαδικές αποφάσεις αποτελούν τον τύπο δεδομένων Boolean, στον οποίο δύο τιμές συνδέονται με διαφορετικές ενέργειες μέσα σε μια ροή διεργασιών.

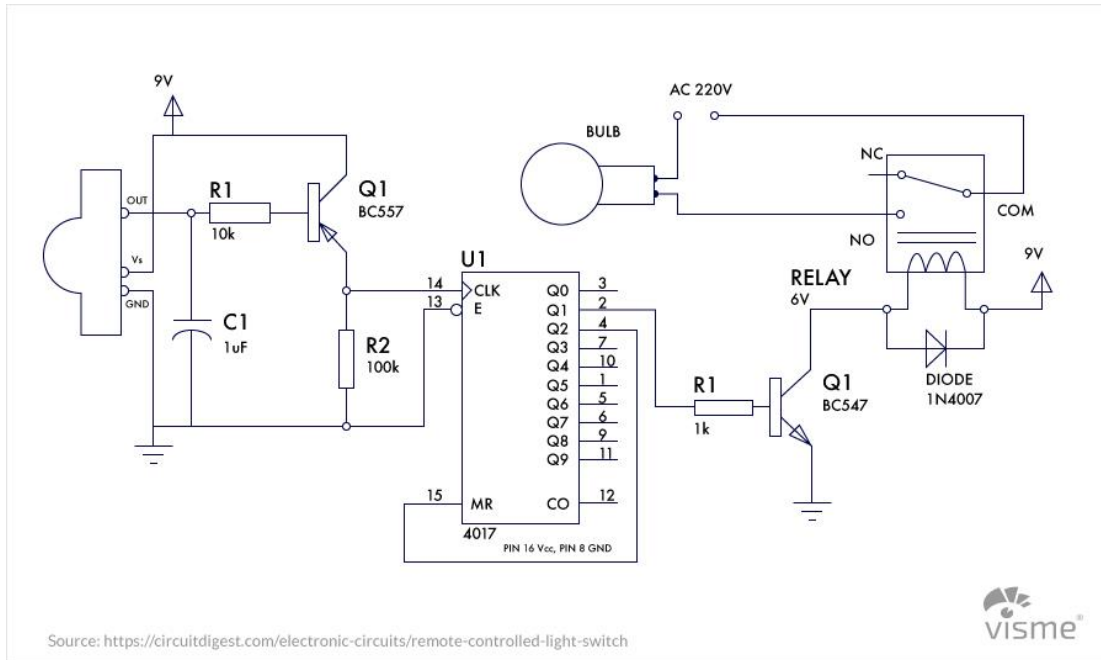
Εκτός από την επιστήμη των υπολογιστών, ένα διάγραμμα δυαδικών αποφάσεων μπορεί ακόμα να χρησιμοποιηθεί για να απεικονίσει οποιαδήποτε διαδικασία με την οποία οι ενέργειες βασίζονται σε μια απόφαση μεταξύ δύο αξιών, εάν οι συνθήκες είναι ναι ή όχι, αληθινή ή ψευδής, 1 ή 0 ή οποιαδήποτε άλλη αντίθετη επιλογή. Τελικά, η διαδρομή που ακολουθεί θα απεικονίσει τον τρόπο ροής της διαδικασίας, από την αρχή μέχρι το τέλος.



Διαγράμματα κυκλώματος

Όπως υπονοεί το όνομά του, ένα διάγραμμα κυκλωμάτων είναι μια οπτική αναπαράσταση ενός ηλεκτρικού κυκλώματος. Χρησιμοποιώντας απλά σχήματα και εικόνες, το διάγραμμα απεικονίζει τα εξαρτήματα και τις διασυνδέσεις ενός κυκλώματος, από την αρχή μέχρι το τέλος. Ενώ οι διαδρομές και οι συνδέσεις είναι ακριβείς, το διάγραμμα δεν αντιπροσωπεύει απαραίτητα αναλογική χωρική κατασκευή του κυκλώματος.

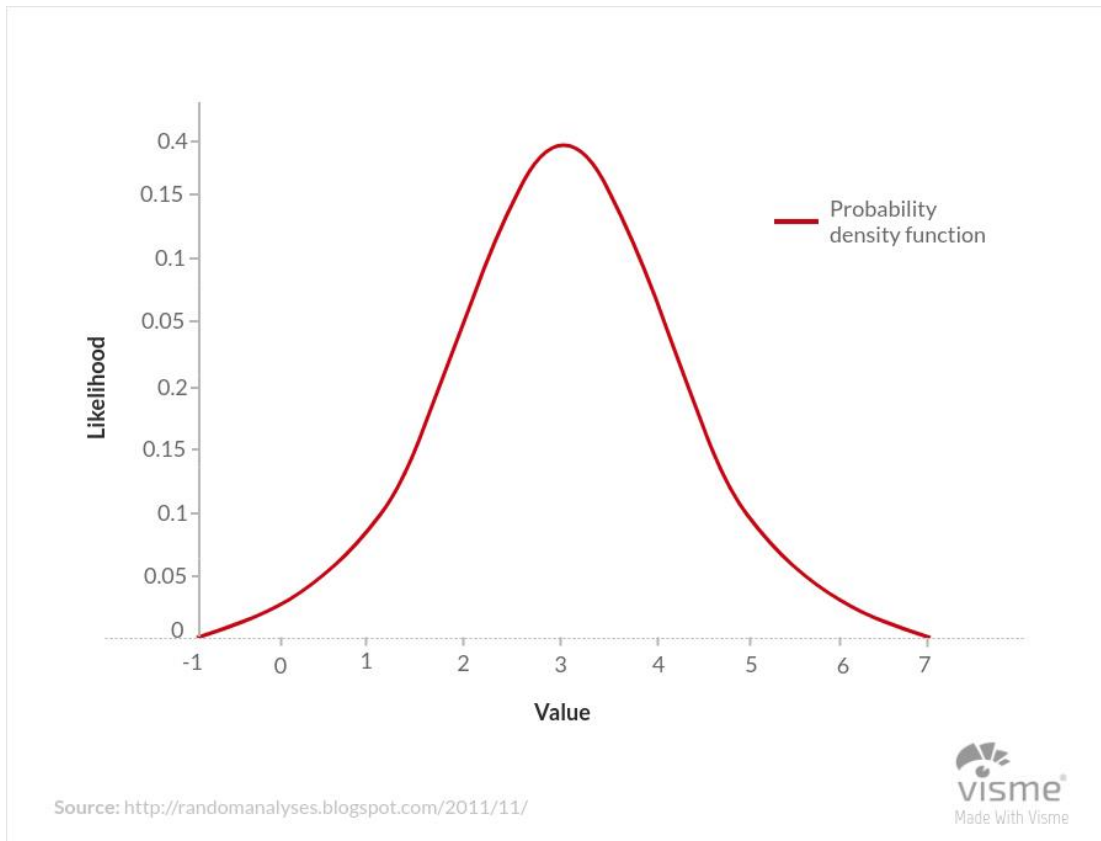
Στην επιστήμη των υπολογιστών, τα διαγράμματα κυκλωμάτων είναι χρήσιμα στην απεικόνιση δεδομένων που σχετίζονται με το υλικό και το λογισμικό. Τα γραφικά όχι μόνο εικονογραφούν τις διαδρομές ενός κυκλώματος με την κυριολεκτική έννοια αλλά συνδέονται επίσης στενά με το προαναφερθέν διάγραμμα δυαδικών αποφάσεων - και οι δύο χρησιμοποιούνται για τη ροή των διαδικασιών προγραμματισμού διαγραμμάτων.



Γραφήματα λειτουργίας

Οι μαθηματικοί, οι μηχανικοί και οι στατιστικοί πρέπει συχνά να καθορίσουν την αξία μιας εξίσωσης με τη γραφική παράσταση του αποτελέσματός της. Το γράφημα μιας συνάρτησης είναι το σύνολο όλων των σημείων των οποίων οι συντεταγμένες ικανοποιούν την εξίσωση.

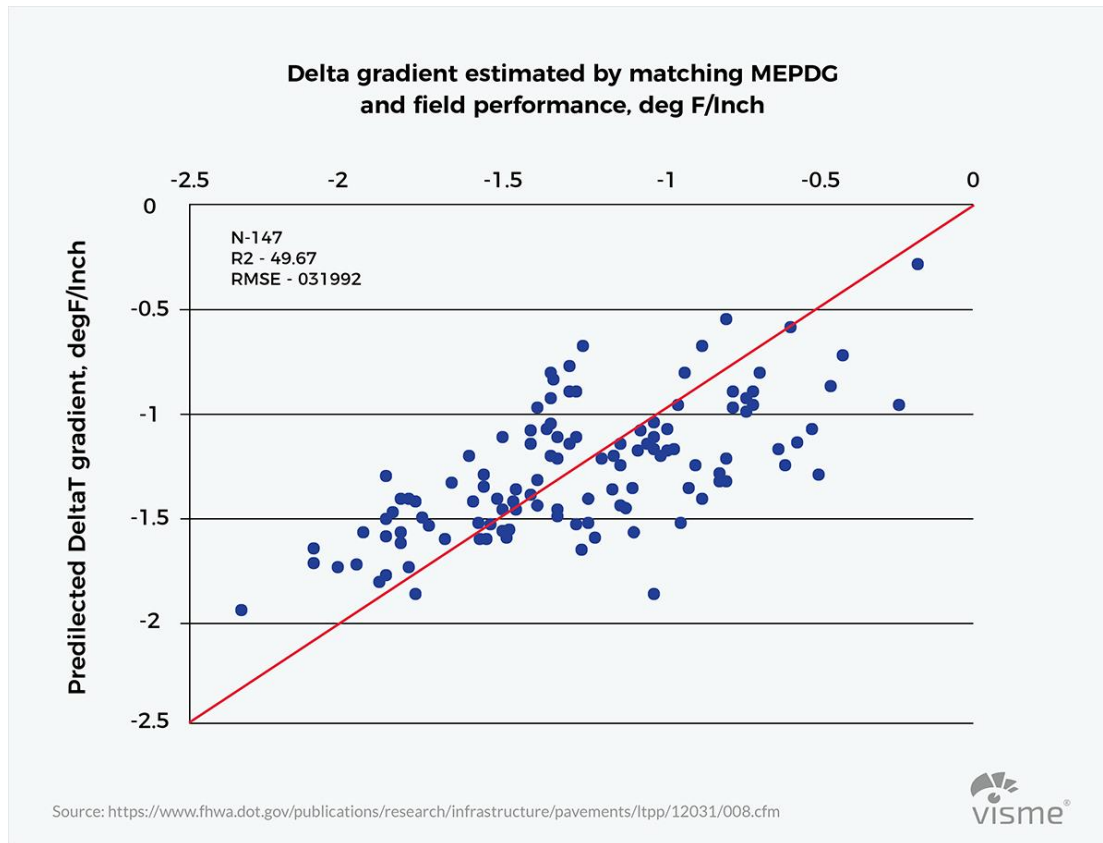
Επομένως, η συνάρτηση μιας εξίσωσης με μεταβλητές των x και y θα σχεδιαστεί σε ένα γράφημα με άξονα x και y . Ομοίως, μια εξίσωση που περιλάμβανε επίσης μια μεταβλητή του z θα χρειαζόταν να σχεδιαστεί σε ένα τρισδιάστατο γράφημα με έναν τρίτο άξονα. Γραφήματα λειτουργιών κοινών σχημάτων συνδέονται οπτικά με τους αντίστοιχους αλγεβρικούς τύπους.



Γραφήματα διασποράς

Επίσης γνωστό ως scattergram, το γράφημα αποτελείται από δύο άξονες, το καθένα αντιπροσωπεύει ένα σύνολο δεδομένων. Για παράδειγμα, ένας άξονας μπορεί να αντιπροσωπεύει τον αριθμό μιλίων που οδηγείται από ένα όχημα, ενώ ο δεύτερος άξονας εμφανίζει τους συνολικούς γαλόνια του χρησιμοποιούμενου αερίου.

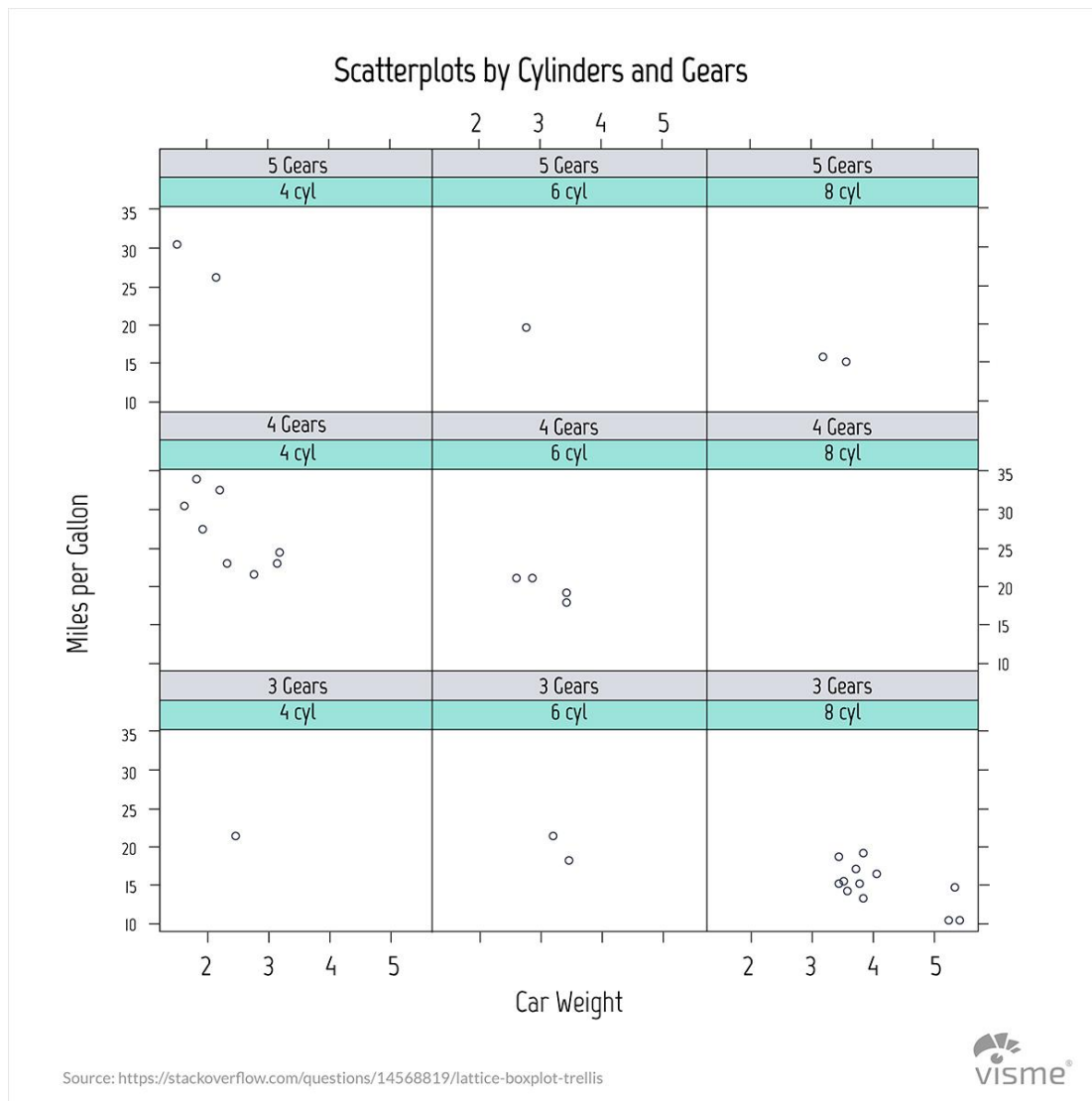
Για κάθε όχημα που υποβλήθηκε σε δειγματοληψία, ο μέσος όρος των μιλίων ανά γαλόνι αντιπροσωπεύεται από μια τελεία που έχει γραφτεί στο γράφημα. Μόλις σχεδιαστούν πολλαπλές κουκίδες, οι τάσεις μπορούν να εντοπιστούν και τα δείγματα μπορούν να συγκριθούν, ανάλογα με τον αριθμό των χρωμάτων που εμφανίζονται στο διάγραμμα.



Trellis plots

Μερικές φορές ένας στατιστικολόγος θα πρέπει να συγκρίνει περισσότερα σύνολα δεδομένων από ότι μπορεί να αναπαρασταθεί από ένα μόνο γράφημα. Τι γίνεται αν, για παράδειγμα, ένα γράφημα πρέπει να συγκρίνει όχι μόνο τα μίλια που κινούνται και τα γαλόνια που χρησιμοποιούνται, αλλά και τον αριθμό των γραναζιών και κυλίνδρων που περιέχονται σε κάθε δείγμα οχήματος;

Μια γραφική παράσταση, που ονομάζεται επίσης γραφική παράσταση ή γραφική παράσταση, μπορεί να εμφανίσει και να συγκρίνει όλες αυτές τις μεταβλητές. Παρόλο που το παραπάνω παράδειγμα χρησιμοποιεί μια σειρά διάγραμμα διασποράς, τα γραφήματα πέργκολας εμφανίζονται συνήθως σε σειρές γραφημάτων γραμμών ή γραμμών.



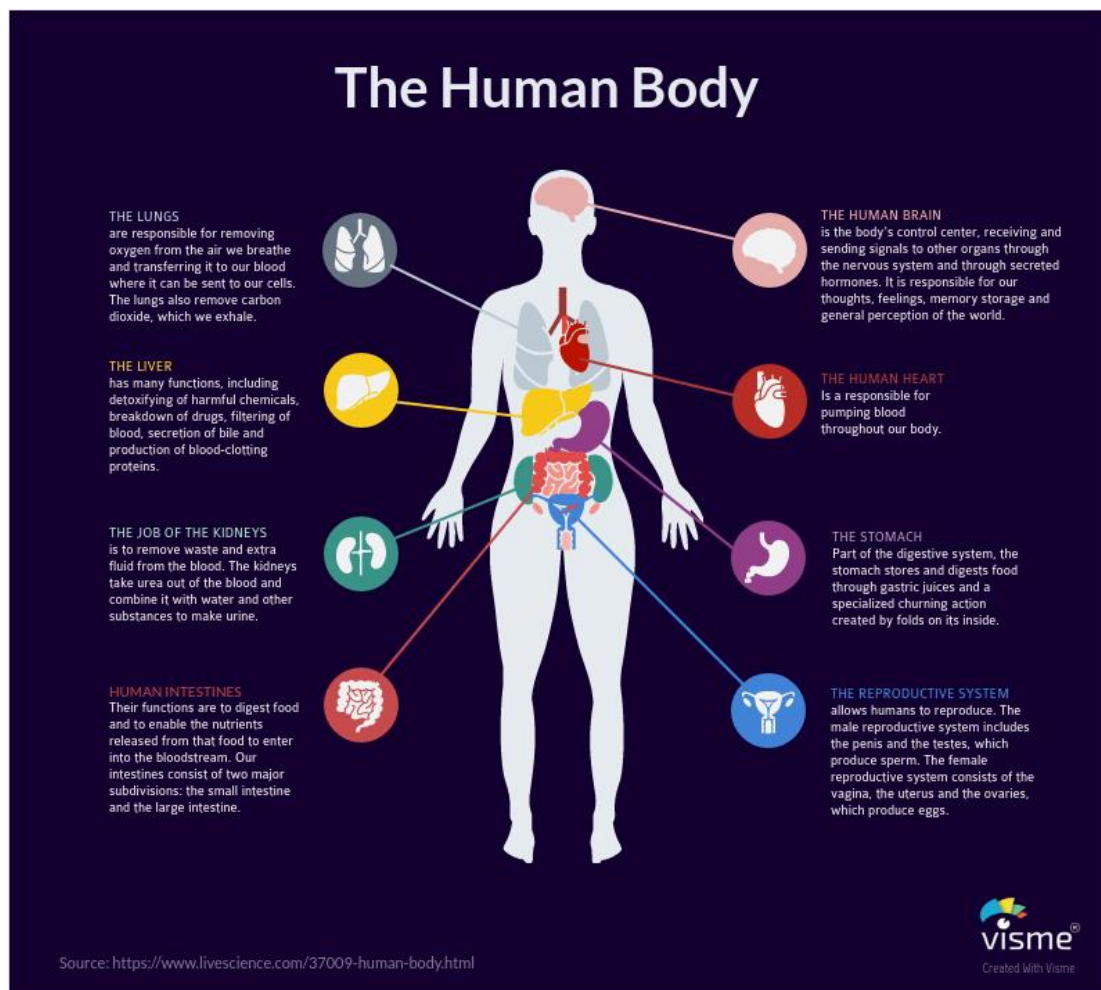
3.7.3 Υγεία και ευεξία

Ανατομικά διαγράμματα

Τα ιατρικά διαγράμματα χρησιμοποιούνται συχνά για την απεικόνιση της ανατομίας, των θεραπειών ή της παθολογίας της νόσου, προκειμένου να εξηγηθούν οι θεραπείες για ασθενείς και για άλλους χωρίς εκτεταμένο βιοϊατρικό υπόβαθρο. Είναι πολύ πιο εύκολο να διαγράψετε το πώς τα διάφορα νεύρα επηρεάζουν διαφορετικά σωματικά συστήματα από το να προσπαθήσουν και να εξηγήσουν προφορικά τα μονοπάτια.

Ενώ τα ιατρικά διαγράμματα θεωρούνται συνδυασμός επιστήμης και τέχνης, μπορούν να είναι εξίσου τεχνικά με οποιοδήποτε άλλο ποσοτικό γράφημα. Και ανεξάρτητα από το πόσο λεπτομερές είναι το σχέδιο, τα ανατομικά διαγράμματα σχεδιάζονται για να παρουσιάζουν σαφώς και αποτελεσματικά τα

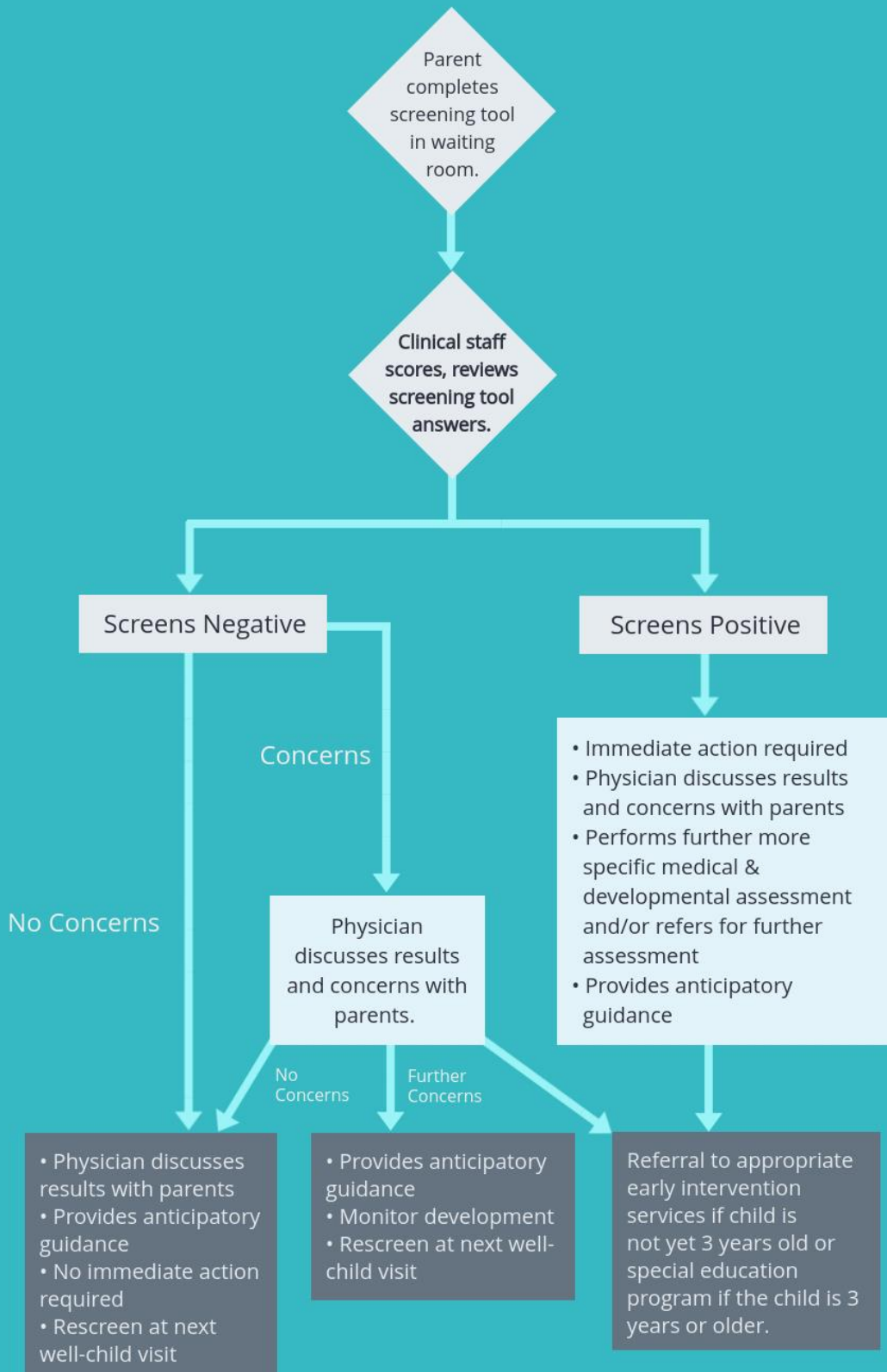
δεδομένα. Και όπως συμβαίνει με ένα πολύπλοκο διάγραμμα περιγράμματος, τα διαγράμματα επικεντρώνονται σε βασικές πληροφορίες, ακόμα κι αν έχουν επιλεγεί από ογκώδη ποσά ιατρικών ή επιστημονικών δεδομένων.



Διαγράμματα ροής

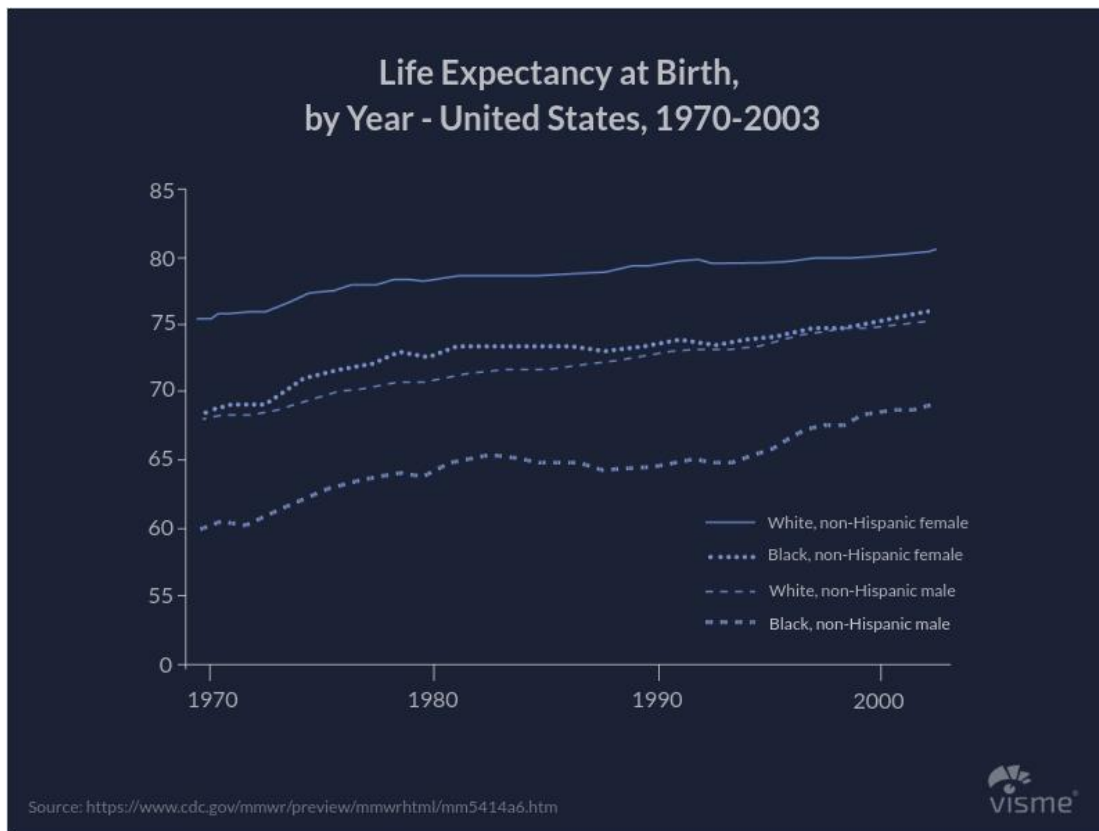
Ακολουθώντας τη σωστή διαδικασία είναι πιθανότατα πιο σημαντική στην ιατρική απ' ό, τι σε οποιοδήποτε άλλο πεδίο. Άλλωστε, αν ο χειρουργός ξεχάσει ένα βήμα, ίσως να αιμορραγείτε πολύ καλά μέχρι να κοιμηθείτε. Οι χάρτες ροής χρησιμοποιούνται συχνά από νοσοκομεία, κλινικές και άλλες ιατρικές εγκαταστάσεις για να διασφαλιστεί η ομοιόμορφη παρακολούθηση των σωστών διαδικασιών.

Pediatric Developmental Screening Flowchart



Γραφήματα πολλαπλών γραμμών

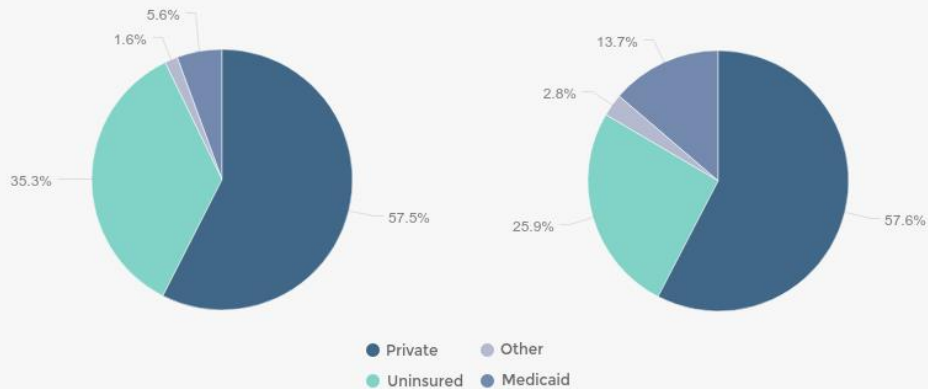
Ακριβώς όπως τα σπάνια απομονωμένα ιατρικά συμπτώματα, ούτε η ανάλυση των βιομετρικών δεδομένων. Μετά από όλα, σπάνια ένα στατιστικό χρώμα ζωγραφίζει ολόκληρη την ιατρική εικόνα. Τα γραφήματα γραμμής μπορούν να αντικατοπτρίζουν πολλά σύνολα δεδομένων με γραμμές διαφορετικών μοτίβων ή χρωμάτων. Για παράδειγμα, ένα γράφημα πολλαπλών γραμμών μπορεί να απεικονίσει αλλαγές στο προσδόκιμο ζωής όχι μόνο του πληθυσμού εν γένει, αλλά και για κάθε φύλο και πολλαπλά φυλετικά υπόβαθρα.



Γραφήματα πολλαπλών πιτών

Ακριβώς όπως στις περιπτώσεις γραφημάτων πίτας πολλαπλών επιπέδων, σελιδοποιημένα γραμμικά γράμματα και γραφήματα πέργκολας, τα γραφικά πολλαπλών πινακίδων ζωγραφίζουν ένα πιο λεπτομερές πορτρέτο του συνόλου δεδομένων που απεικονίζει. Ενώ ένα ενιαίο διάγραμμα πίτας μπορεί να δείξει ποιο τμήμα του συνολικού πληθυσμού έχει μια συγκεκριμένη κατάσταση, ένα γράφημα πολλαπλών καρυδιών μπορεί να σπάσει αυτά τα στατιστικά στοιχεία για να απεικονίσει όχι μόνο το ποσοστό των ανδρών και το ποσοστό των γυναικών, αλλά και τον τρόπο σύγκρισης των δύο ομάδων ο ένας τον ΑΛΛΟΝ.

Distribution of health insurance status for adults aged 20-29 years, by sex: United States, 2008



Source: <https://www.cdc.gov/nchs/products/databriefs/db29.htm>

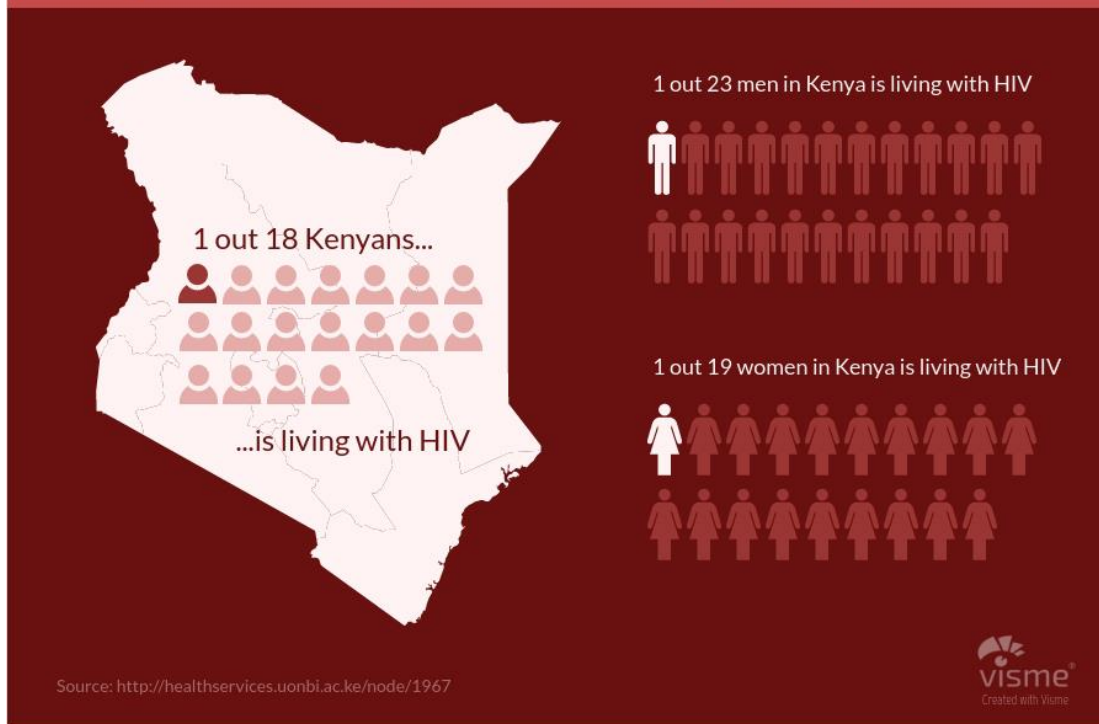


Pictographs

Σε ένα εικονογράφημα ή ένα εικονόγραμμα χρησιμοποιούνται εικόνες και σύμβολα για την απεικόνιση δεδομένων. Για παράδειγμα, ένα βασικό εικονογράφημα μπορεί να χρησιμοποιήσει μια εικόνα του ήλιου για να σηματοδοτήσει κάθε μέρα δίκαιου καιρού σε ένα μήνα και ένα σύννεφο βροχής για να συμβολίσει κάθε θυελλώδη μέρα.

Επειδή οι εικόνες είναι γνωστό ότι κατέχουν περισσότερη συναισθηματική ισχύ από τα ακατέργαστα δεδομένα, συχνά χρησιμοποιούνται εικονογράμματα για την παρουσίαση ιατρικών δεδομένων. Μια απεικόνιση ότι οι αποχρώσεις πέντε από τα 20 σύμβολα προσώπων που αντιπροσωπεύουν ποσοστό θνητότητας 20 τοις εκατό φέρνουν ένα πιο ισχυρό μήνυμα, για παράδειγμα, από μια ράβδο, γραμμή ή πίτα που απεικονίζει τα ίδια δεδομένα.

HIV prevalence among women and men aged 15-64 years in 2012



Γραφήματα διασποράς

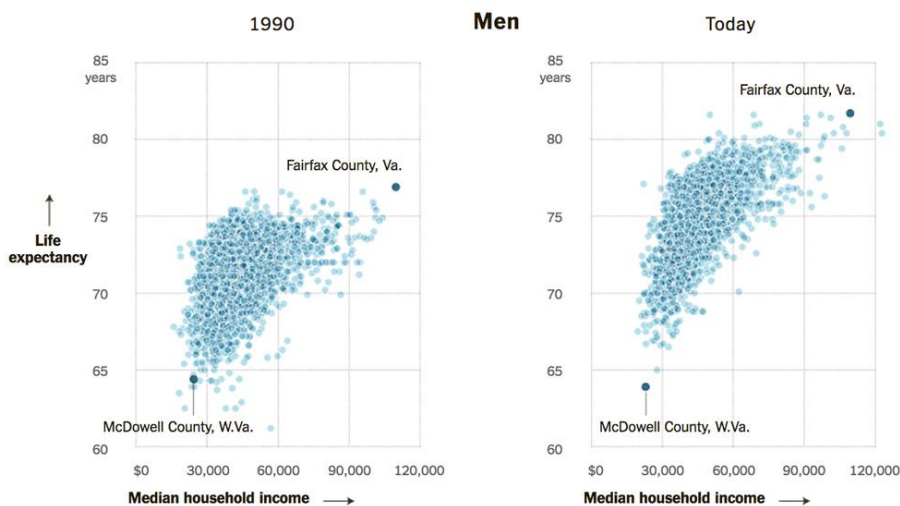
Μπορεί να είναι δύσκολο να αναπαραστήσουμε γραφικά σύνολα ιατρικών δεδομένων που αποτελούνται από εκατοντάδες ή περισσότερους ασθενείς, όπως συμβαίνει στις περισσότερες ιατρικές μελέτες. Αλλά μια γραφική παράσταση διασποράς επιτρέπει την αναπαράσταση κάθε υποκείμενου, που σχεδιάζεται στο γράφημα σύμφωνα με τις μεταβλητές στους δύο άξονες του διαγράμματος.

Το μοτίβο που σχηματίζεται από τις γραφικές παραστάσεις μπορεί να καθορίσει με σαφήνεια τις τάσεις στα δεδομένα. Αναλύοντας μια γραφική παράσταση σκέδασης, για παράδειγμα, ένας ερευνητής θα μπορούσε εύκολα να εντοπίσει μια συσχέτιση μεταξύ του μεγαλύτερου προσδόκιμου ζωής και του υψηλότερου εισοδήματος των νοικοκυριών.

Where Income Is Higher, Life Spans Are Longer

As incomes have diverged between the country's richest counties, like Fairfax County, Va., and its poorest ones, like McDowell County, W.Va., so have the life expectancies of their residents.

• Every U.S. county is represented by a dot.



Stacked bar graphs

Οι διαγραμματισμένες γραφικές παραστάσεις δεν είναι χρήσιμες μόνο για την απεικόνιση τμημάτων ενός συνόλου. Μπορούν επίσης να χρησιμοποιηθούν για την εμφάνιση πρόσθετων μεταβλητών. Ενώ ένα βασικό γράφημα ράβδων μπορεί να αντιπροσωπεύει το τμήμα του πληθυσμού που έχει ταξινομηθεί ως υπερβολικό βάρος κατά τη διάρκεια μιας καθορισμένης χρονικής περιόδου, ένα στοιβαγμένο γράφημα ράβδων μπορεί επίσης να παρακολουθεί πόσο μεγάλο μέρος του συνόλου είναι παχύσαρκοι.

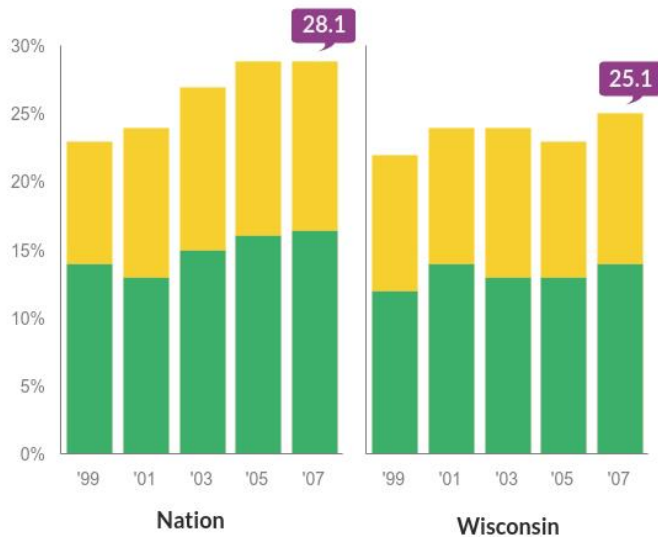
Teens getting heavier

Trends in overweight high school students from 1997-2007 show Wisconsin teens aren't as heavy as teen nationally, but the numbers are heading upward.

Overweight
Obese



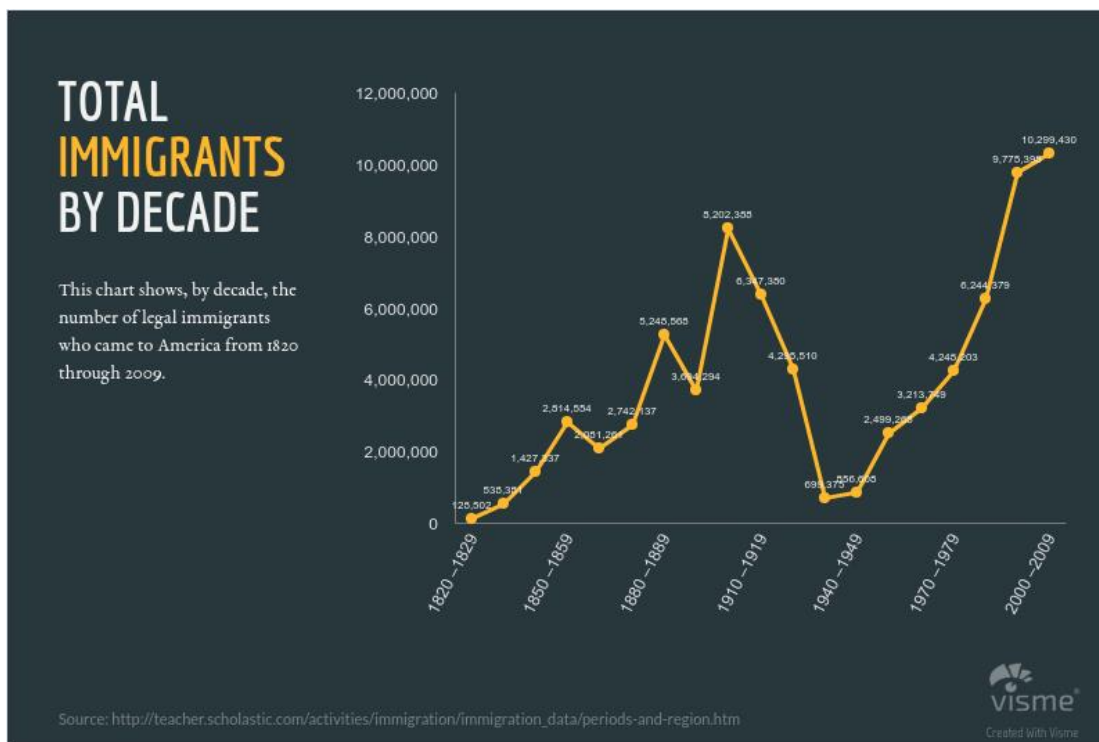
Source: http://host.madison.com/child-obesity-bar-graph/image_18c9b190-a762-11df-ae7a-001cc4c03286.html



3.7.4 Ιστορία

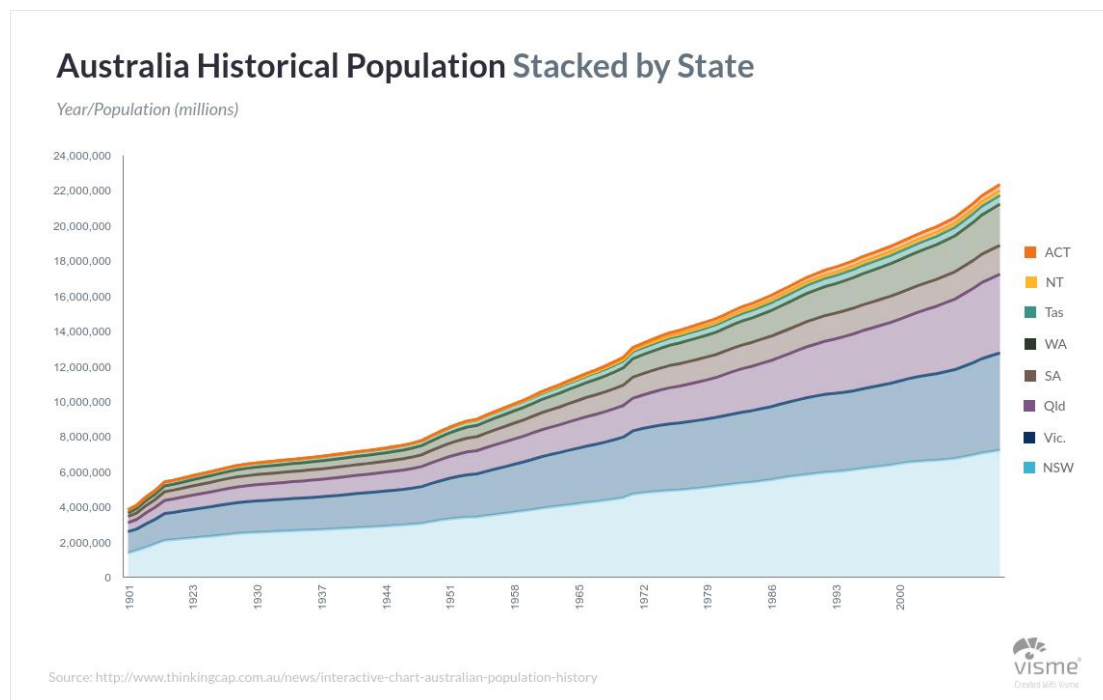
Line graphs

Εάν ένα χρονοδιάγραμμα είναι μια μορφή γραφήματος, τότε έχει νόημα μόνο ότι οι ιστορικοί συχνά το χρησιμοποιούν στην εμφάνιση άλλων δεδομένων. Σχεδιάζοντας τα επίπεδα μετανάστευσης ενάντια σε ένα χρονοδιάγραμμα, το ιστόγραμμα που προκύπτει απεικονίζει τις τάσεις του πληθυσμού για έναν αιώνα ή περισσότερο με ένα βασικό γράφημα γραμμής.



Stacked Area charts

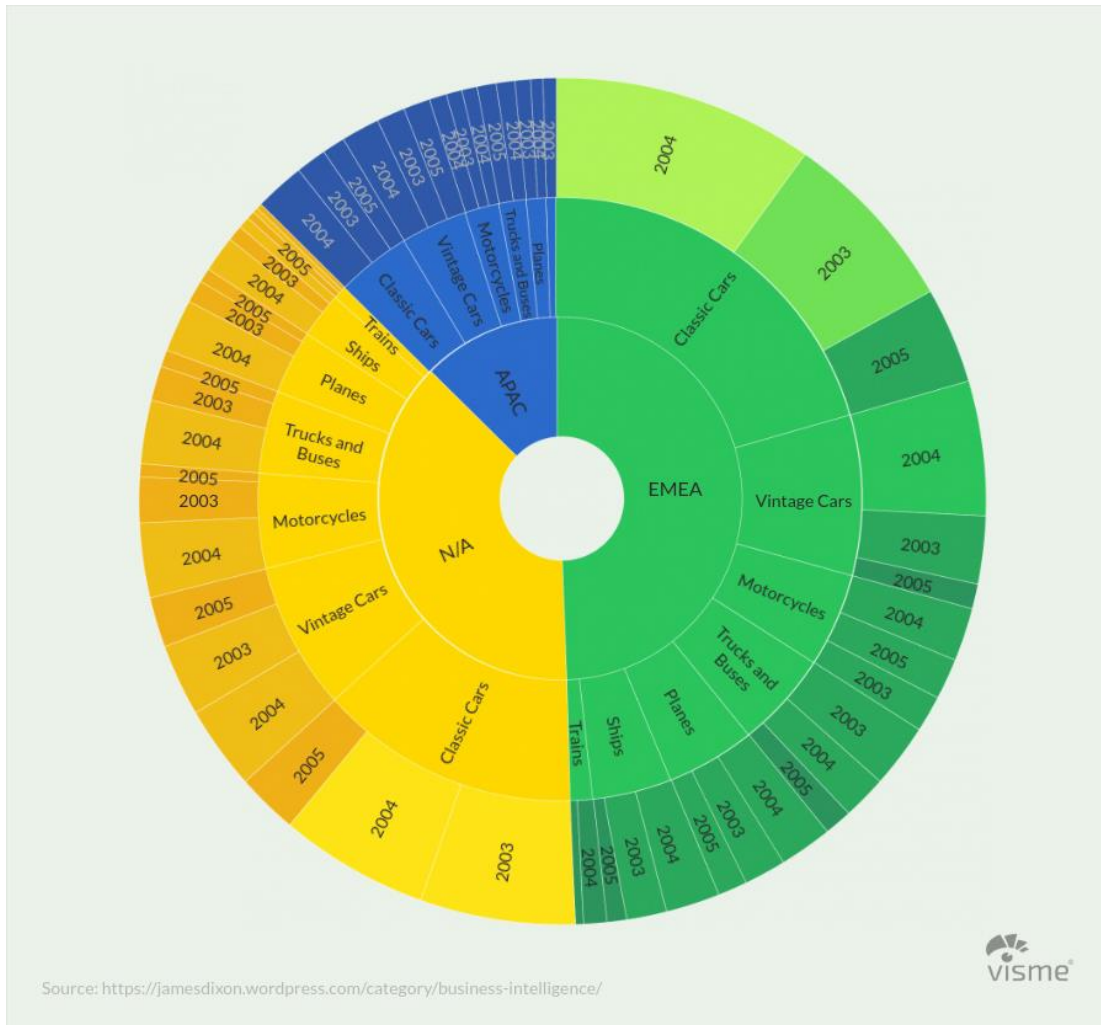
Τα διαγράμματα στοιβαγμένων περιοχών χρησιμοποιούνται συχνά για την αλλαγή διαγραμμάτων πολλαπλών μεταβλητών διαχρονικά. Πολλές γραμμές μπορούν να σχεδιαστούν, για παράδειγμα, για την παρακολούθηση των μεταβολών του πληθυσμού διάφορων κρατών κατά τη διάρκεια του χρόνου. Η περιοχή κάτω από κάθε γραμμή μπορεί να έχει διαφορετική απόχρωση για να αντιπροσωπεύει την κατάσταση που υποδηλώνει, καταλήγοντας σε ένα γράφημα που αντιπροσωπεύει σαφώς τις τάσεις του πληθυσμού, ενώ παράλληλα εμφανίζει τα δεδομένα κάθε κράτους από τα λιγότερο έως τα πολυπληθέστερα.



Sunburst charts

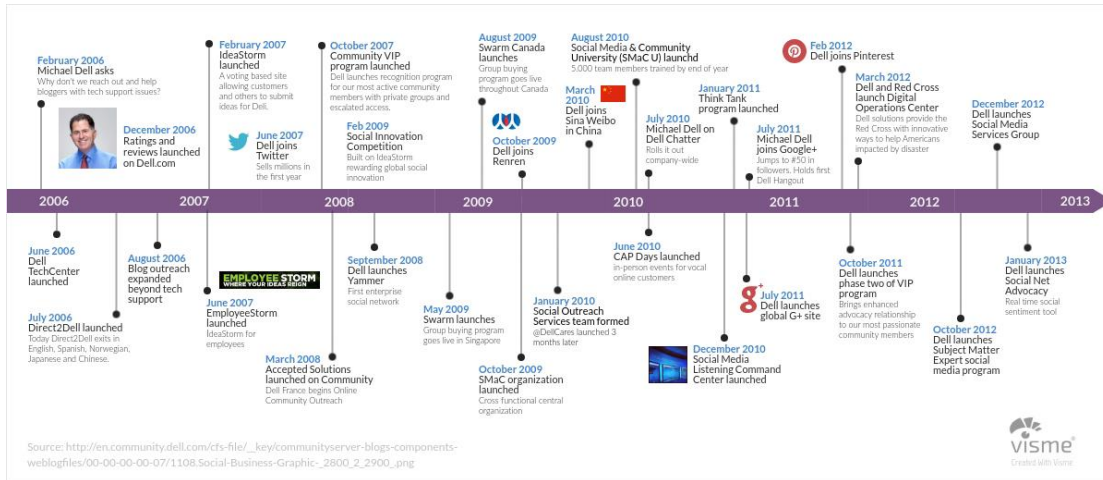
Ένας τύπος διαγράμματος πίτας πολλαπλών επιπέδων, ένας διάγραμμα ηλιοθεραπείας χρησιμοποιείται για την απεικόνιση των ιεραρχικών δεδομένων χρησιμοποιώντας ομόκεντρους κύκλους. Κάθε δακτύλιος του "sunburst" αντιπροσωπεύει ένα επίπεδο στην ιεραρχία, με τον ριζικό κόμβο που αντιπροσωπεύει ο κεντρικός κύκλος, και η ιεραρχία κινείται προς τα έξω. Ενώ ένα διάγραμμα ηλιοθεραπείας μπορεί να χρησιμοποιηθεί για να απεικονίσει μια γνωστή ή εταιρική ιεραρχία, μπορεί επίσης να σπάσει τα δεδομένα κάτω από χρονικές περιόδους, δημιουργώντας μια ιστορική ιεραρχία.

Διάφοροι κλάδοι μιας οργάνωσης μπορούν να εκπροσωπούνται από καθορισμένες αποχρώσεις, με διαφορετικά επίπεδα συχνά λαμβάνοντας ποικίλες αποχρώσεις της ίδιας οικογένειας χρωμάτων. Οι δακτύλιοι μπορούν επίσης να χωριστούν περαιτέρω ώστε να αντιπροσωπεύουν πολλαπλές διαιρέσεις εντός του ίδιου οργανωτικού επιπέδου. Στην πραγματικότητα, ένας παραδοσιακός, σύνθετος τροχός χρωμάτων, όπως αυτός που χρησιμοποιείται από τα καταστήματα χρωμάτων, είναι μια μορφή διάγραμμα ηλιοθεραπείας.



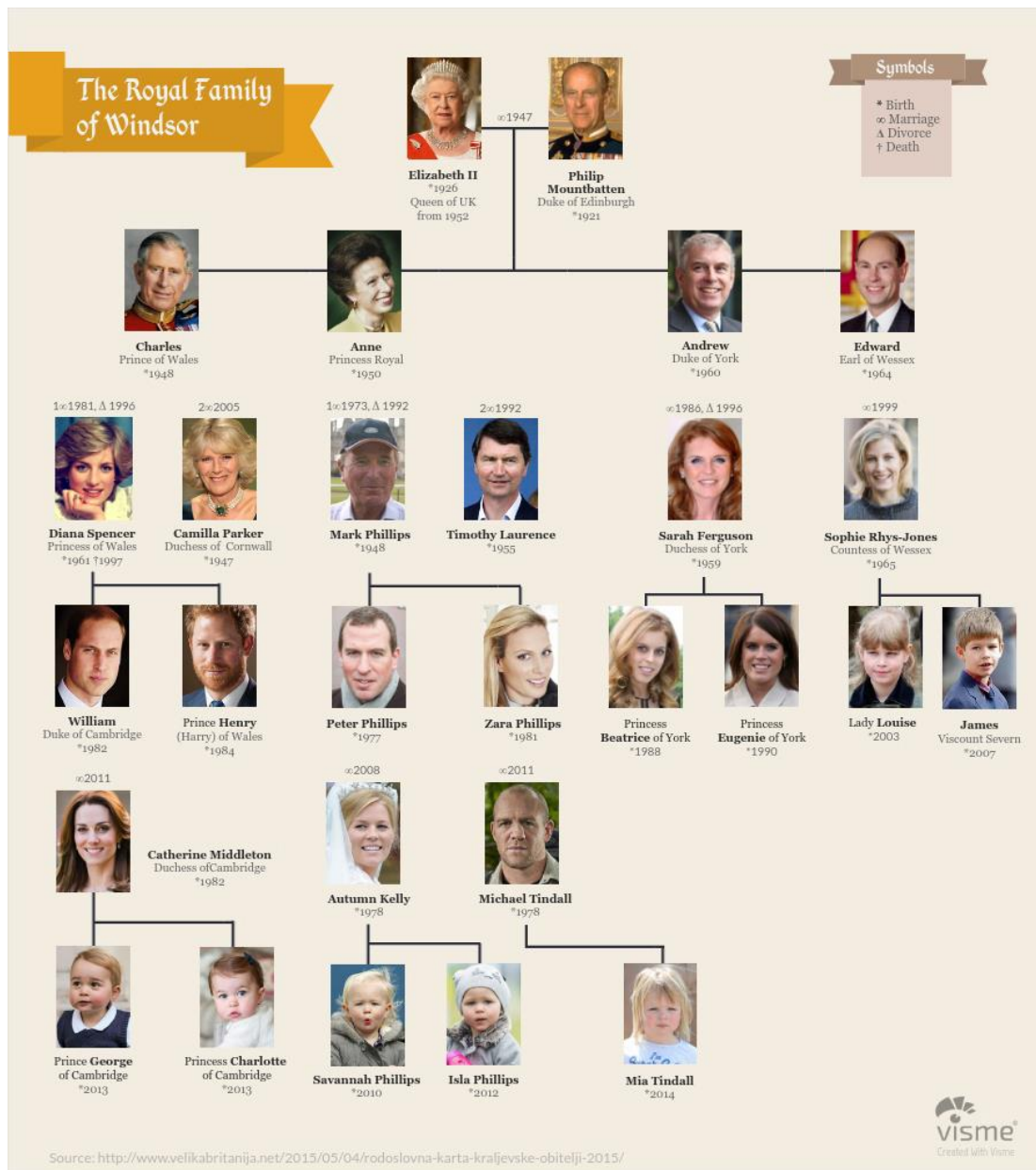
Timelines

Ενδεχομένως η πιο αυτονόητη απεικόνιση δεδομένων, ένα χρονοδιάγραμμα παρακολουθεί δεδομένα σε μια χρονική περίοδο. Σημαντικές ημερομηνίες και γεγονότα επισημαίνονται στο σημείο στο οποίο εμφανίζονται σε χρονολογική κλίμακα. Τα χρονοδιαγράμματα μπορούν να χρησιμοποιηθούν μόνοι τους ή σε συνδυασμό με άλλες απεικονίσεις.



Tree diagrams

Μια μορφή ιεραρχικού διαγράμματος, ένα γενεαλογικό δέντρο, απεικονίζει τη δομή μιας οικογένειας. Μπορεί είτε να ξεκινήσει με έναν πρόγονο, στη συνέχεια να διαγράψει τους απογόνους του, τα αδέρφια τους, τους γάμους και τα παιδιά τους κ.ο.κ. Ένα γενεαλογικό διάγραμμα, από την άλλη, ξεκινά με ένα άτομο και καταγράφει τη γενεαλογία τους, από γονείς σε παππούδες και ούτω καθεξής.



3.7.5 Γραφήματα μάρκετινγκ

Column bar graphs

Ο απλούστερος και πιο απλός τρόπος σύγκρισης διαφόρων κατηγοριών είναι συχνά η κλασική στήλη με βάση τη στήλη. Το γενικά αναγνωρισμένο γράφημα διαθέτει μια σειρά από ράβδους διαφόρων μηκών. Ένας άξονας γραμμικής γραφικής παράστασης παρουσιάζει τις κατηγορίες που συγκρίνονται, ενώ ο άλλος άξονας αντιπροσωπεύει την τιμή του καθενός. Το μήκος κάθε ράβδου είναι ανάλογο με την αξία που αντιπροσωπεύει. Για παράδειγμα, τα \$ 4 θα μπορούσαν να αντιπροσωπεύονται από μια ορθογώνια

ράβδο μήκους τεσσάρων μονάδων, ενώ τα \$ 5 θα ισοδυναμούσαν με μια ράβδο μήκους πέντε μονάδων. Με μια γρήγορη ματιά, οι ακροατές μάθουν ακριβώς πώς τα διάφορα στοιχεία μεγεθύνονται το ένα πάνω στο άλλο.

Τα γραφήματα μπαρ λειτουργούν εξαιρετικά για την οπτική παρουσίαση σχεδόν οποιουδήποτε τύπου δεδομένων, αλλά κατέχουν ιδιαίτερη δύναμη στη βιομηχανία μάρκετινγκ. Τα διαγράμματα χρησιμοποιούνται συνήθως για την παρουσίαση οικονομικών προβλέψεων και αποτελεσμάτων και τα γραφήματα είναι ιδανικά για τη σύγκριση κάθε είδους αριθμητικής τιμής, συμπεριλαμβανομένων των μεγεθών των ομάδων, των αποθεμάτων, των αξιολογήσεων και των απαντήσεων της έρευνας.



Line graphs

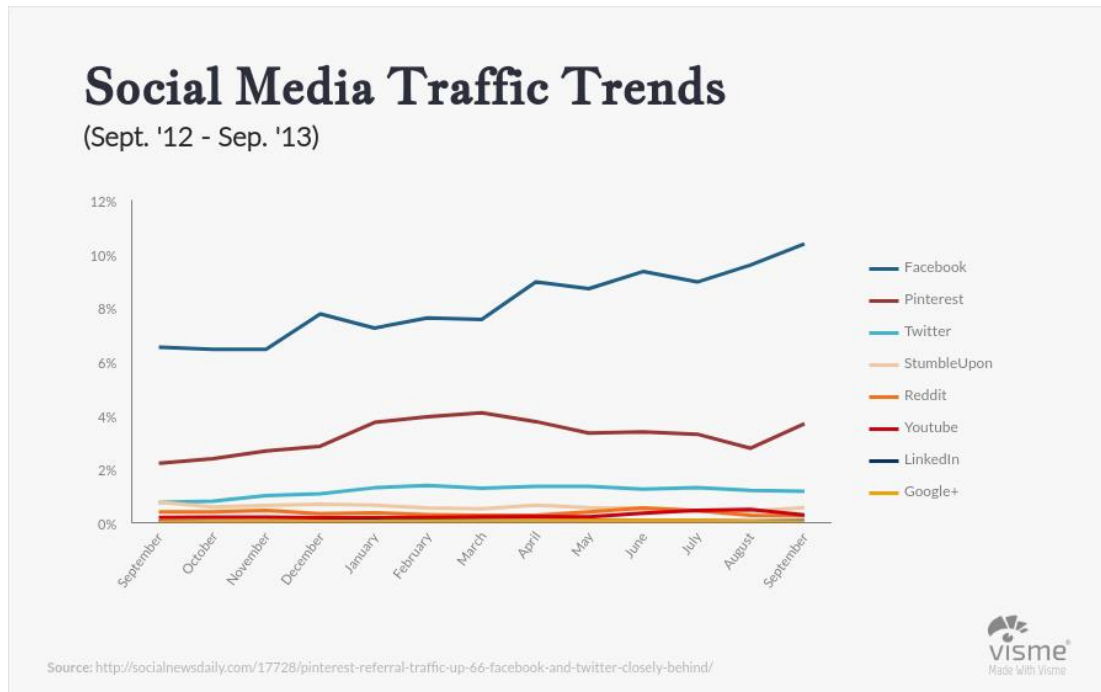
Τα γραφήματα γραμμής ή τα γραμμικά γράμματα είναι ισχυρά οπτικά εργαλεία που απεικονίζουν τις τάσεις των δεδομένων σε μια χρονική περίοδο ή μια συγκεκριμένη συσχέτιση. Για παράδειγμα, ένας άξονας του γραφήματος μπορεί να αντιπροσωπεύει μια μεταβλητή τιμή, ενώ ο άλλος άξονας εμφανίζει συχνά μια χρονική γραμμή.

Κάθε τιμή είναι γραφική παράσταση στο γράφημα, τότε τα σημεία συνδέονται για να εμφανίσουν μια τάση σε σύγκριση με το χρονικό διάστημα σύγκρισης. Πολλές τάσεις μπορούν να συγκριθούν με τη χάραξη γραμμών διαφόρων χρωμάτων ή μοτίβων.

Για παράδειγμα, η δημοτικότητα διαφόρων δικτύων κοινωνικής δικτύωσης μέσα σε ένα χρόνο μπορεί να συγκριθεί οπτικά με ευκολία μέσω της χρήσης ενός γραφήματος γραμμής. Απλώς σχεδιάστε τη βάση

χρηστών κάθε εταιρίας για κάθε μήνα του διαστήματος των 12 μηνών, στη συνέχεια συνδέστε τις κουκκίδες με μια γραμμή καθορισμένου χρώματος.

Το κοινό θα αναγνωρίσει γρήγορα ποια κοινωνικά δίκτυα είναι τα περισσότερα και λιγότερο επιτυχημένα, καθώς και τα οποία αντιμετωπίζουν ανάπτυξη ή απώλεια.



Mosaic or mekko charts

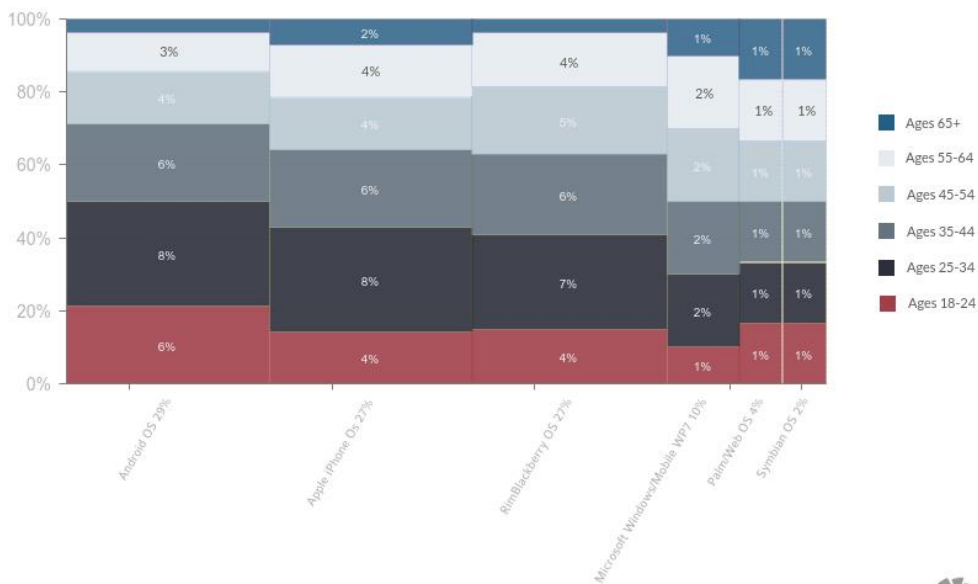
Οι βασικοί χάρτες γραμμών, γραμμών και πίτας είναι εξαιρετικά εργαλεία για τη σύγκριση μιας ή δύο μεταβλητών σε λίγες κατηγορίες, αλλά τι συμβαίνει όταν πρέπει να συγκρίνετε πολλαπλές μεταβλητές ή πολλαπλές κατηγορίες ταυτόχρονα; Τι γίνεται αν όλες αυτές οι μεταβλητές δεν είναι ακόμη αριθμητικές; Μια γραφική απεικόνιση μωσαϊκού - ή Mekko - μπορεί να είναι η καλύτερη επιλογή.

Ίσως ένας αναλυτής της αγοράς, για παράδειγμα, θέλει να συγκρίνει περισσότερο από το μέγεθος των διαφόρων αγορών κινητής τηλεφωνίας. Τι γίνεται αν, αντίθετα, πρέπει να συγκρίνει το μέγεθος των βάσεων χρηστών, καθώς και τις ηλικιακές ομάδες σε κάθε ομάδα; Ένα διάγραμμα μωσαϊκού θα επέτρεπε στον εν λόγω έμπορο να απεικονίσει όλες τις μεταβλητές με σαφή και απλό τρόπο.

Στο παραπάνω παράδειγμα, ένας άξονας του διαγράμματος αντιπροσωπεύει τις κατηγορίες που συγκρίνονται - κατασκευαστές κινητών τηλεφώνων - ενώ ο άλλος άξονας παραθέτει διάφορες ηλικιακές κλίμακες. Το μέγεθος και το χρώμα κάθε διατομής του διαγράμματος αντιστοιχεί με το τμήμα της αγοράς που αντιπροσωπεύει.

Age and operating system share-smartphones

Nov '10-Jan '11, postpaid mobile subscribers, n=14,701



Source: The Nielsen Company

visme
Made With Visme

Pie charts

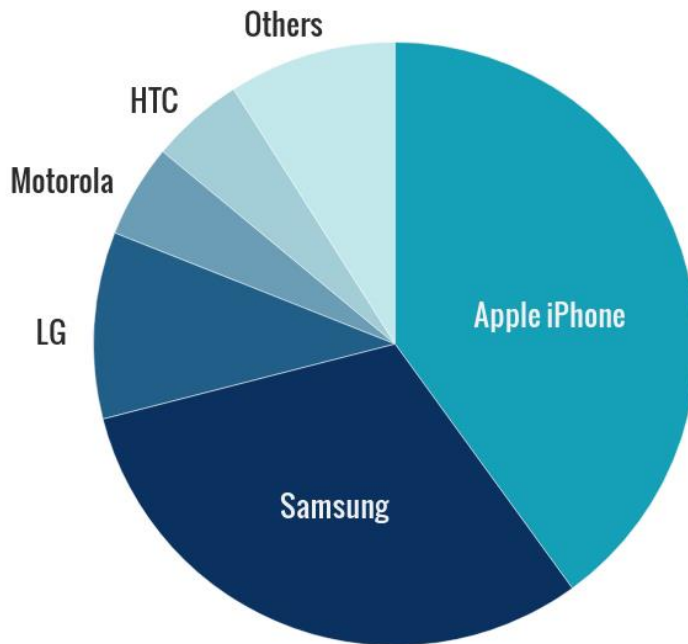
Τα διαγράμματα πίτας είναι το απλούστερο και πιο αποτελεσματικό εργαλείο για τη σύγκριση μερών ενός συνόλου. Για παράδειγμα, ένα διάγραμμα πίτας μπορεί να συγκρίνει γρήγορα και αποτελεσματικά τις διάφορες κατανομές του προϋπολογισμού, τα τμήματα του πληθυσμού ή τις απαντήσεις σε ερωτήσεις έρευνας αγοράς.

Οι σχεδιαστές περιεχομένου μάρκετινγκ συχνά βασίζονται σε διαγράμματα πίτας για να συγκρίνουν το μέγεθος των τμημάτων της αγοράς. Για παράδειγμα, ένα απλό γράφημα πίτας μπορεί να δείξει με σαφήνεια πώς οι πιο δημοφιλείς κατασκευαστές κινητών τηλεφώνων συγκρίνουν με βάση τα μεγέθη των βάσεων χρηστών τους.

Το κοινό μπορεί γρήγορα να καταλάβει ότι η Apple και η Samsung κατέχουν σχεδόν το 75% της αγοράς κινητής τηλεφωνίας, με την Apple ελαφρώς μπροστά. Το μήνυμα αυτό μπορεί να σταλεί χωρίς να εκτυπώνει ένα μοναδικό αριθμητικό ψηφίο.

Smartphone Brand Market Share

Smartphone Owners in U.S.
Broadband Households



© Parks Associates

visme[®]
Made With Visme

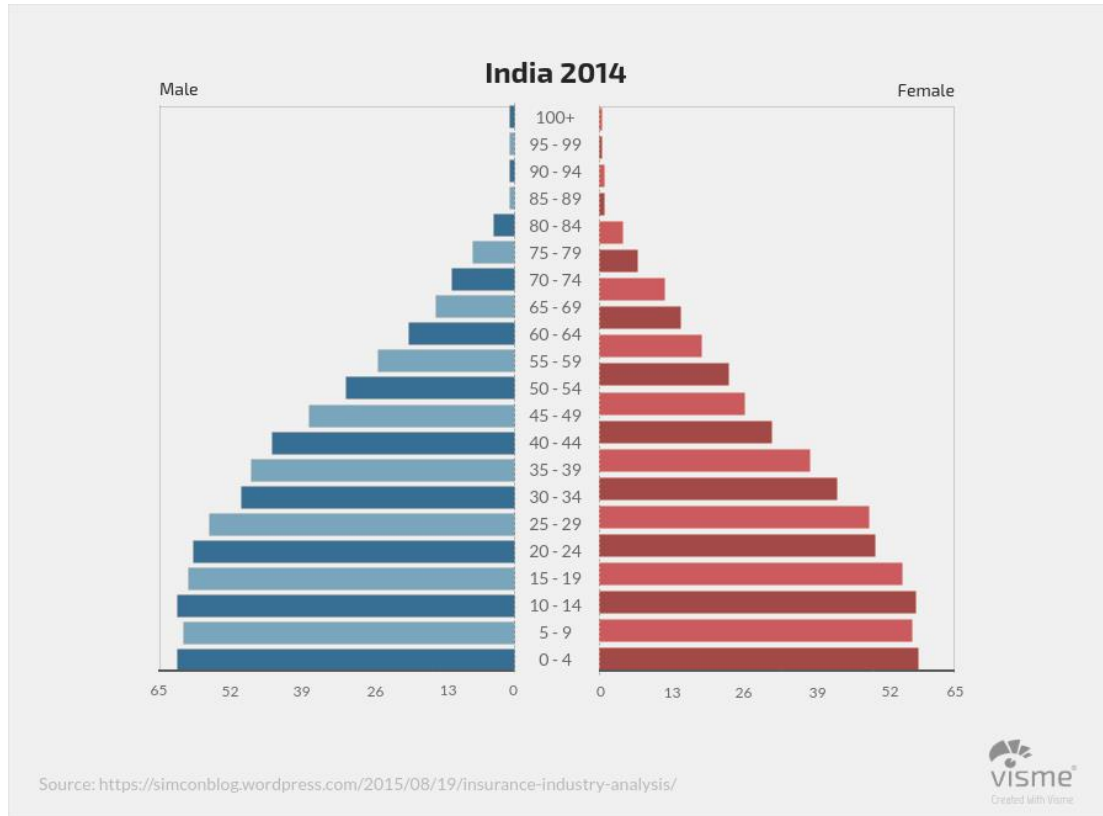
Population Pyramids

Τα τμήματα της αγοράς συχνά διαιρούνται με βάση την ηλικία και το φύλο και μια πληθυσμιακή πυραμίδα είναι μια ιδανική οπτική απεικόνιση των δύο ομάδων. Το γράφημα παίρνει κλασικά το σχήμα μιας πυραμίδας όταν ο πληθυσμός είναι υγιής και αναπτυσσόμενος - οι μεγαλύτερες ομάδες είναι οι νεώτερες και κάθε φύλο εξασθενεί κάπως εξίσου με τον πληθυσμό που μένει, αφήνοντας τις μικρότερες ομάδες στην κορυφή του γραφήματος.

Μια πληθυσμιακή πυραμίδα που ξεφεύγει από το κλασικό της σχήμα μπορεί να υποδεικνύει μια παρατυπία σε έναν πληθυσμό σε μια συγκεκριμένη περίοδο, όπως ένας λιμός ή μια οικονομική άνθηση που οδήγησε σε αύξηση των θανάτων ή των γεννήσεων.

Φυσικά, οι πληθυσμιακές πυραμίδες δεν χρησιμοποιούνται πάντα για να συγκρίνουν τους πληθυσμούς κατά ηλικία και επομένως δεν παίρνουν πάντοτε το ομώνυμο σχήμα του γράφου. Ένας έμπορος, για

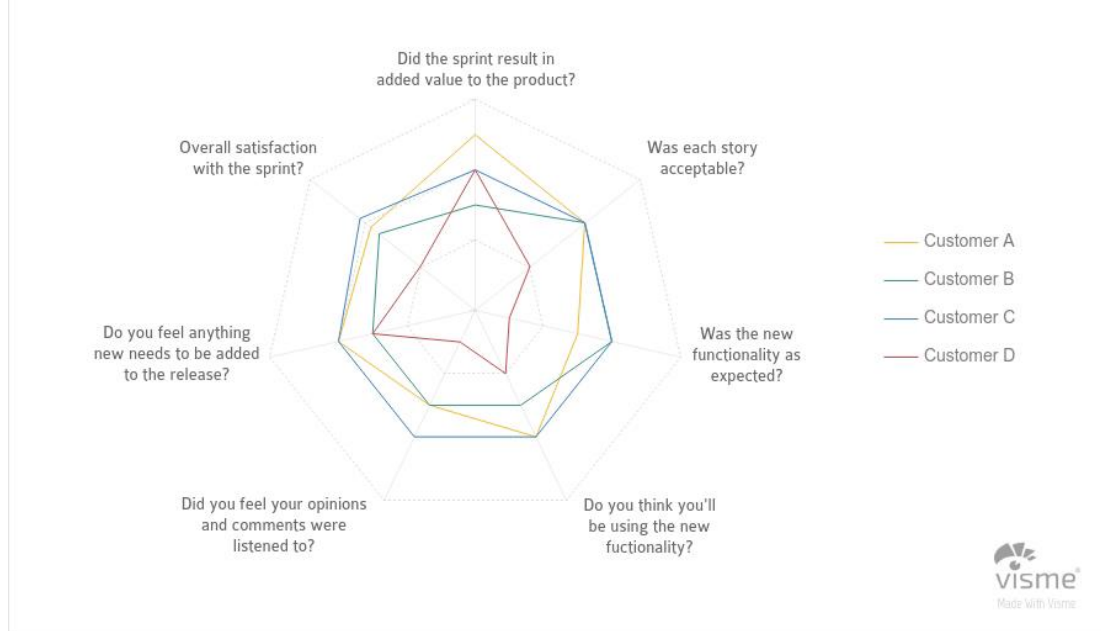
παράδειγμα, μπορεί να χρησιμοποιήσει το σχέδιο για να συγκρίνει έναν πληθυσμό με το εισόδημα, το βάρος ή το IQ, όπου οι μικρότερες ομάδες θα είναι συχνά στο πάνω και στο κάτω μέρος. Ανεξάρτητα, το γράφημα απεικονίζει με σαφήνεια τις τάσεις του πληθυσμού, ενώ συγκρίνει τα μεγέθη δύο συναφών ομάδων.



Spider charts

Όταν ένας στατιστικός πρέπει να συγκρίνει οπτικά τρεις ή περισσότερες ποσοτικές μεταβλητές, αυτός ή αυτή μπορεί να επιλέξει να χρησιμοποιήσει ένα διάγραμμα ραντάρ, γνωστό και ως διάγραμμα αστεριού ή αστέρα. Το διάγραμμα αποτελείται συνήθως από μια σειρά ακτίνων, κάθε μία από τις οποίες αντιπροσωπεύει μια διαφορετική κατηγορία, που εκτοξεύεται από ένα κεντρικό σημείο όπως ακτίνες. Το μήκος κάθε "ομιλίας" είναι ανάλογο με την τιμή που συγκρίνεται. Για κάθε κατηγορία, οι ακτίνες συνδέονται στη συνέχεια με μια γραμμή καθορισμένου σχεδίου ή χρώματος, σχηματίζοντας σχήμα άστρου με σημεία ίσα με τον αριθμό των κατηγοριών. Το αποτέλεσμα είναι μια γραφική παράσταση που μπορεί να αποκαλύψει τις τάσεις και να συγκρίνει τις κατηγορίες ταυτόχρονα.

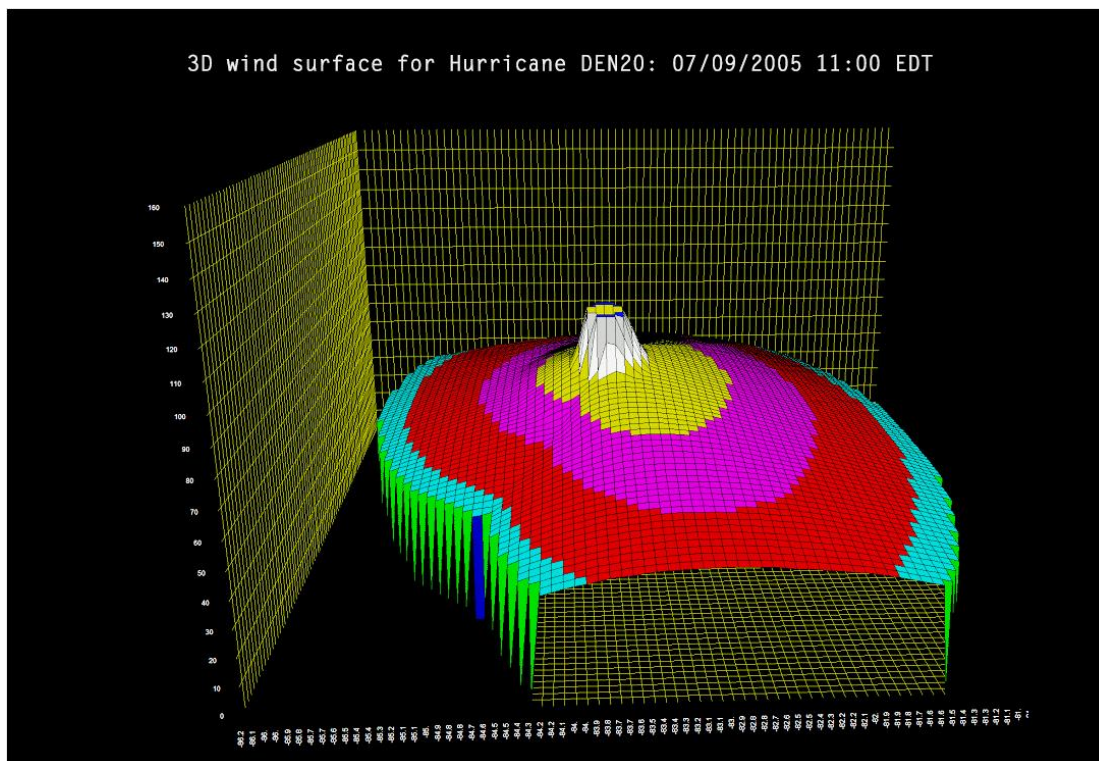
Customer Satisfaction Survey-Sprint Review



3.7.6 Μετεωρολογία και περιβάλλον

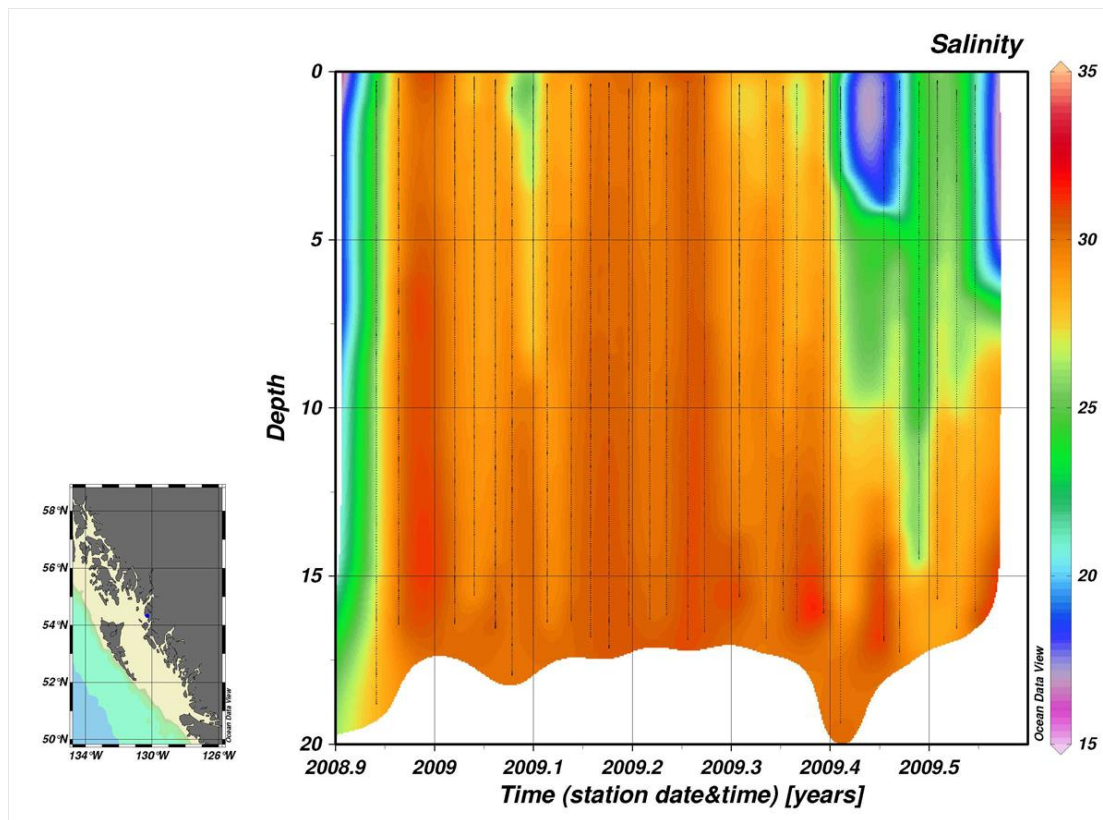
3D graphs

Η τεχνολογία επιτρέπει πλέον στους στατιστικούς να εμφανίζουν πολυδιάστατα σύνολα δεδομένων σε πραγματική μορφή. Τα τρισδιάστατα γραφήματα που δημιουργήθηκαν με εξειδικευμένο λογισμικό απεικονίζουν τη σχέση μεταξύ τριών μεταβλητών που σχεδιάστηκαν σε τρεις άξονες. Ένας μετεωρολόγος μπορεί, για παράδειγμα, να γράψει το αιολικό πεδίο ενός τυφώνα.



Contour Plots

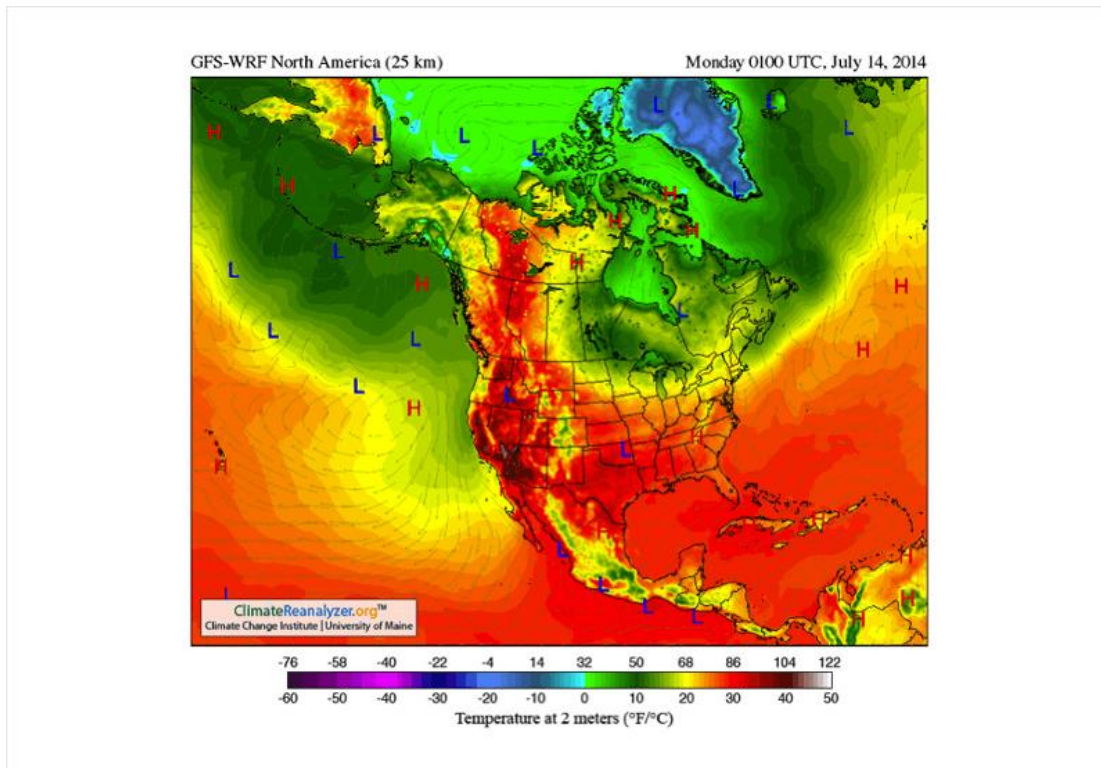
Τα περιγράμματα περιγράμματος επιτρέπουν την ανάλυση τριών μεταβλητών σε δισδιάστατη μορφή. Αντί να σχεδιάζει δεδομένα σε δύο κύριους άξονες, το γράφημα παρουσιάζει μια τρίτη τιμή που ονομάζεται σκίαση ή χρώμα. Ακριβώς όπως ένας τοπογραφικός χάρτης ορίζει γεωγραφικό μήκος, πλάτος και ανύψωση σε ένα δισδιάστατο σχέδιο, ένα γράφημα περιγράμματος απεικονίζει τιμές των x , y και z . Με ένα γράφημα περιγράμματος, για παράδειγμα, ένας κλιματολόγος μπορεί όχι μόνο να σχεδιάσει την αλατότητα του ωκεανού σε διαφορετικές ημερομηνίες, αλλά και την αλατότητα του σε διάφορα βάθη σε αυτές τις ημερομηνίες.



Heat Maps

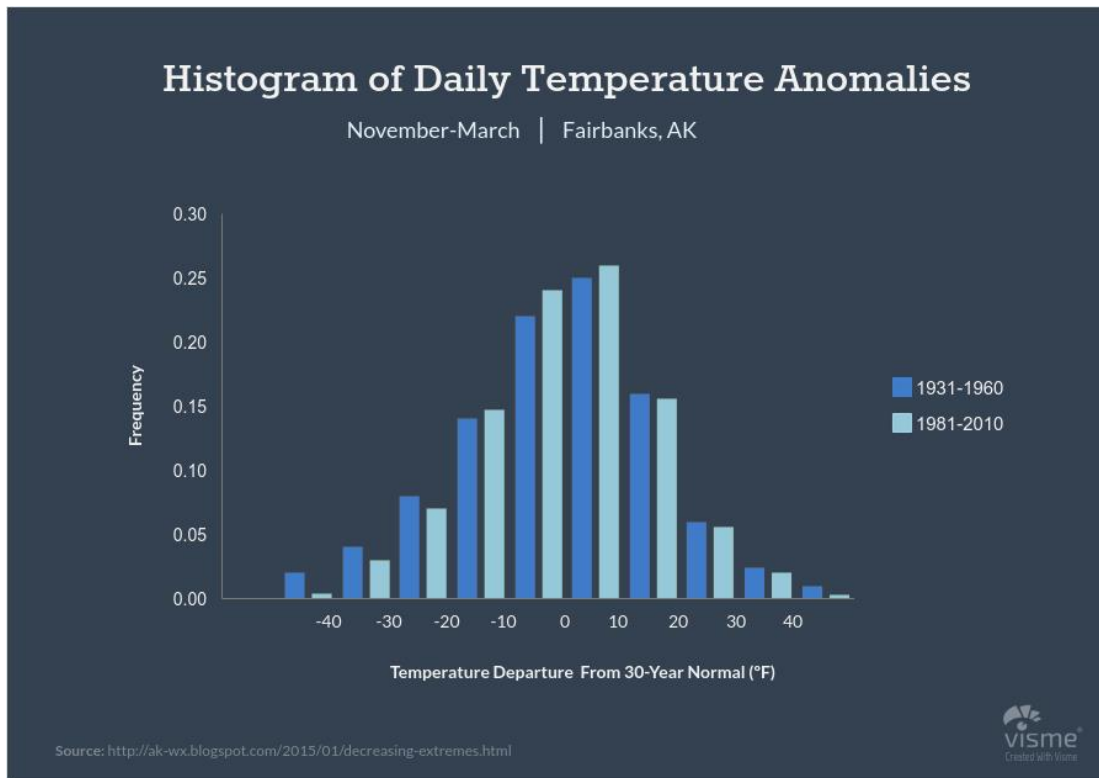
Ένας τύπος γραφήματος περιγράμματος, ένας χάρτης θερμότητας επισημαίνει συγκεκριμένα διαφορετικές θερμοκρασίες σε διαφορετικά γεωγραφικά σημεία. Ενώ οι δύο άξονες του γραφήματος είναι γεωγραφικό πλάτος και γεωγραφικό μήκος, η τρίτη μεταβλητή - θερμοκρασία - αντιπροσωπεύεται από ένα φάσμα χρώματος.

Ενώ οι πιο συχνά χρησιμοποιούμενες για την απεικόνιση του καιρού, οι χάρτες θερμότητας μπορούν επίσης να αντιπροσωπεύουν την κυκλοφορία ιστού, τους οικονομικούς δείκτες και σχεδόν οποιαδήποτε άλλα τρισδιάστατα δεδομένα.



Histograms

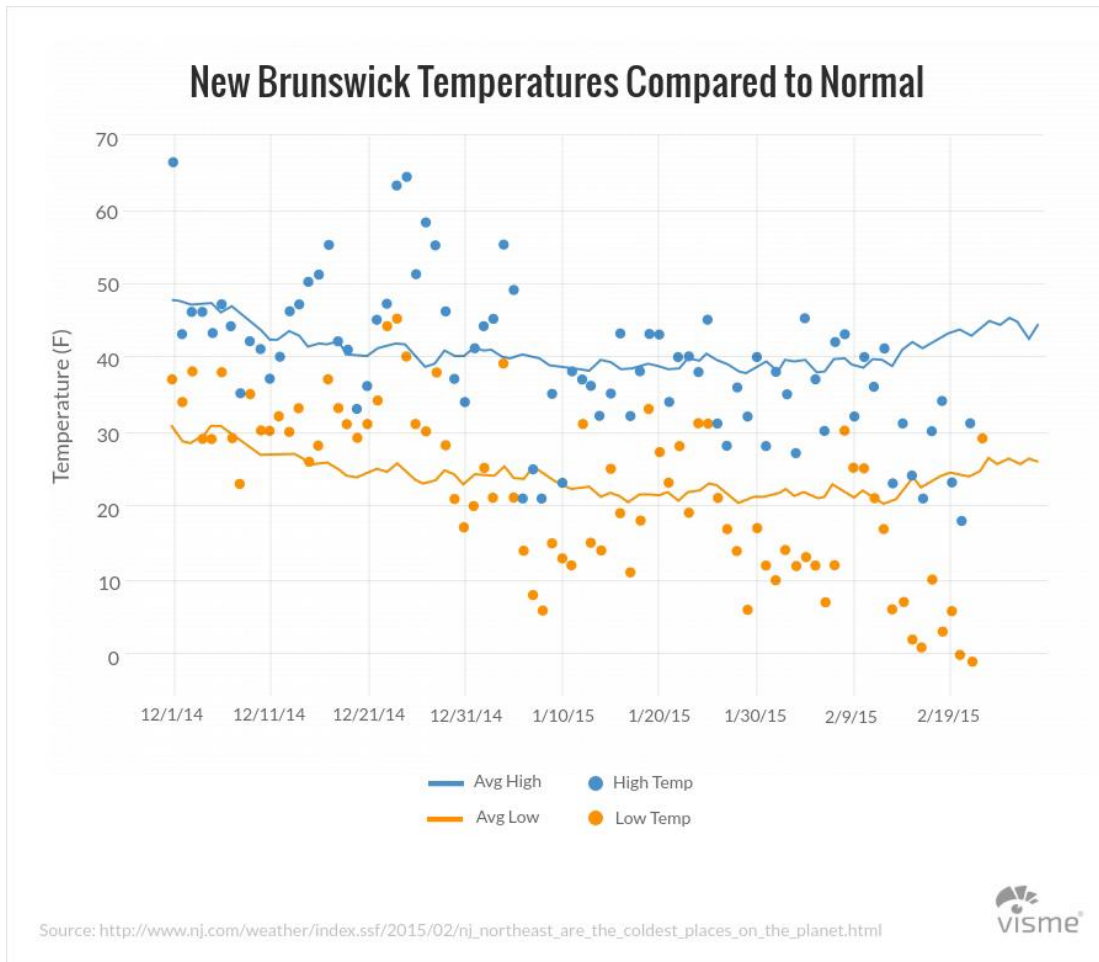
Εξ ορισμού, ένα ιστόγραμμα είναι ένας ειδικός τύπος γραφήματος που παρουσιάζει αριθμητικά δεδομένα και τη διανομή του. Όπως υποδηλώνει και το όνομά της, η κατανομή συχνά απεικονίζεται διαχρονικά, αλλά τα δεδομένα θα μπορούσαν επίσης να σχεδιαστούν με βάση οποιαδήποτε χρονολογική κλίμακα, τέτοια θερμοκρασία, ανύψωση ή νομισματική αξία. Ενώ τα ιστογράμματα είναι συνήθως μια μορφή γραφήματος ράβδων, η έννοια μπορεί επίσης να εφαρμοστεί σε γραφήματα γραμμών και άλλα σχέδια που βασίζονται σε σχεδίαση δύο αξόνων.



Scatter-line combo

Συνδυάζοντας ένα γραμμικό γράφημα με μια γραφική παράσταση σκέδασης, οι μετεωρολόγοι και άλλοι στατιστικολόγοι μπορούν να απεικονίσουν τη σχέση μεταξύ δύο συνόλων δεδομένων. Για παράδειγμα, οι υψηλές και χαμηλές θερμοκρασίες κάθε ημέρας σε ένα μήνα μπορούν να εμφανιστούν σε μια γραφική παράσταση σκέδασης, τότε μπορεί να προστεθεί ένα γραμμικό γράφημα για να σχεδιαστούν οι ιστορικές μέσες υψηλές και χαμηλές θερμοκρασίες κατά την ίδια περίοδο.

Το προκύπτον γράφημα συνδυασμού εμφανίζει με σαφήνεια τον τρόπο σύγκρισης του εύρους θερμοκρασίας κάθε μέρα με τον ιστορικό μέσο όρο και υποδεικνύει ακόμη και τον τρόπο με τον οποίο αυτές οι μετρήσεις εμφανίζουν τάση κατά την εξεταζόμενη χρονική περίοδο.



3.7.7 Πολιτική επιστήμη και κοινωνιολογία

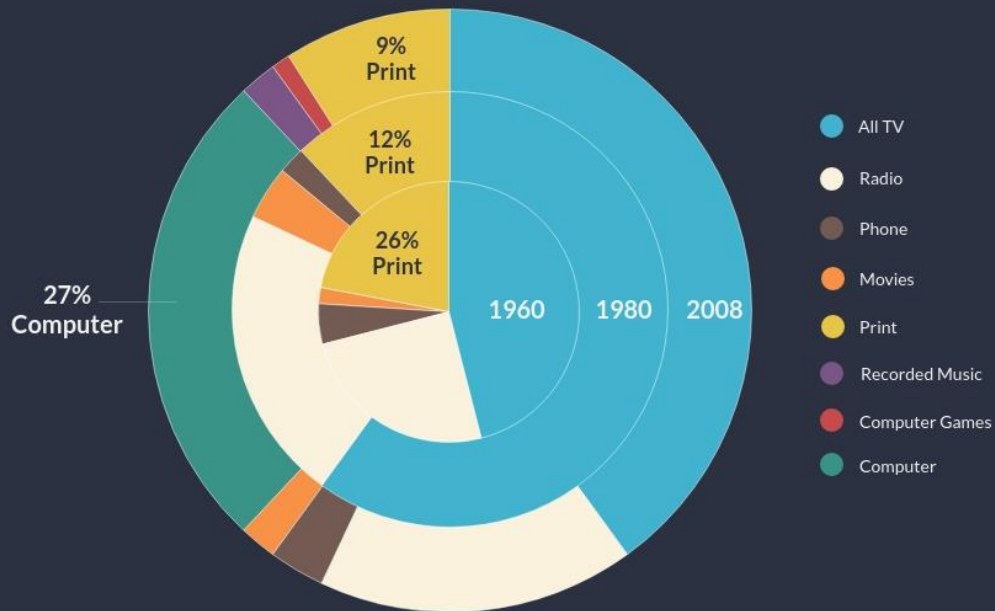
Multi-level Pie charts

Πολύ συχνά ένας σχεδιαστής βρίσκει τον εαυτό του με περισσότερα σύνολα δεδομένων από ότι μπορεί να παρουσιαστεί σε ένα ενιαίο πρότυπο γράφημα. Ευτυχώς, στην περίπτωση ενός διαγράμματος πίτας, μπορούν να παρουσιαστούν πολλαπλά στρώματα δεδομένων χωρίς την ανάγκη πολλαπλών εικόνων ή σχεδίου πέργκολας.

Ένα διάγραμμα πίτας πολλαπλών επιπέδων, για παράδειγμα, αποτελείται από επίπεδα, κάθε στρώμα αντιπροσωπεύει ξεχωριστό σύνολο δεδομένων. Έτσι, ενώ θα χρειαστούν τρία παραδοσιακά γράμματα πίτας για να επεξηγηθούν οι διάφορες πηγές καταγεγραμμένων λέξεων για τρεις διαφορετικές δεκαετίες, ένα γράφημα πίτας πολλαπλών επιπέδων δεν μπορεί να πάρει μόνο τη θέση και των τριών, αλλά προσφέρει επίσης μια σαφέστερη οπτική σύγκριση των αποτελεσμάτων κάθε έτους .

Evolution of Reading

Fraction of words from different sources



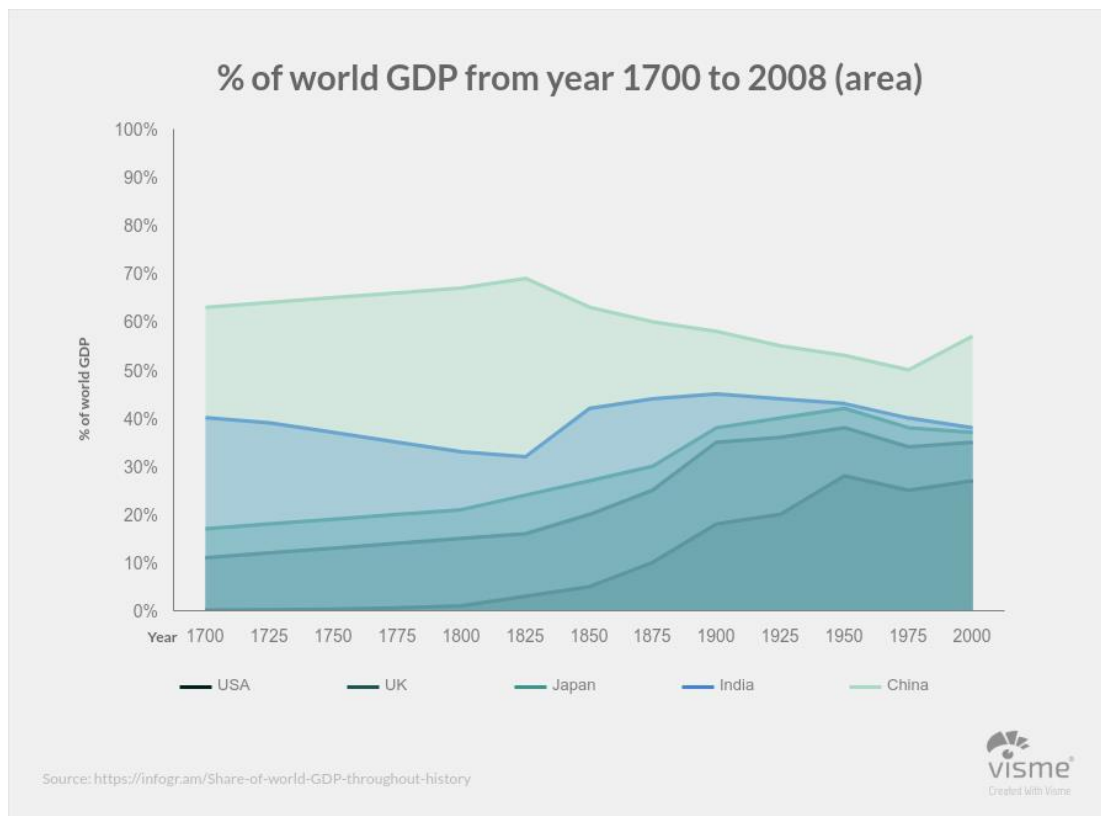
Source: <http://www.calit2.net/newsroom/release.php?id=1630>

visme
Created With Visme

Stacked Area Charts

Τα διαγράμματα στοιβαζόμενων περιοχών είναι ιδανικά για τη σύγκριση τιμών που συνήθως απαιτούν γραφή πολλαπλών γραμμών. Κάθε γραμμή αντιπροσωπεύει μια διαφορετική κατηγορία και η περιοχή κάτω από κάθε γραμμή είναι γενικά σκιασμένη σε ένα καθορισμένο χρώμα, ώστε κάθε σύνολο δεδομένων να μπορεί εύκολα να συγκριθεί.

Για παράδειγμα, ένα διάγραμμα περιοχής με έναν άξονα που αντιπροσωπεύει μια αριθμητική τιμή και έναν άλλο άξονα που χρησιμεύει ως χρονική γραμμή, τα δεδομένα για διάφορες κατηγορίες με την πάροδο του χρόνου μπορούν να παρακολουθούνται και να συγκρίνονται με ένα μόνο γραφικό.



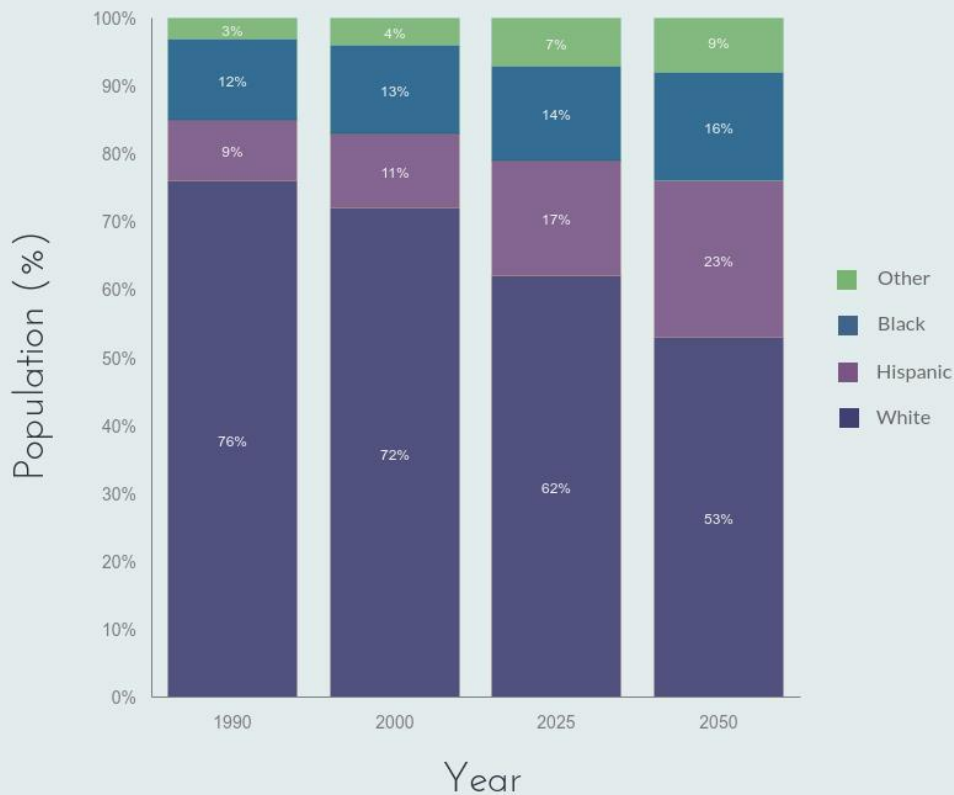
Stacked Bar Graphs

Όταν μελετάτε ομάδες ανθρώπων, είναι κοινό να συγκρίνετε πολλαπλές μεταβλητές ταυτόχρονα.

Εξάλλου, είναι πολύ πιο χρήσιμο να εξεταστεί το φυλετικό υπόβαθρο, οι ηλικίες και το φύλο εκτός από τον συνολικό πληθυσμό. Μια γραφική στήλη που στοιβάζεται συνδυάζει στοιχεία από το παραδοσιακό γράφημα ράβδων και το γράφημα πίτας για να επικοινωνούν τα σύνολα, τις τάσεις και τις αναλογίες σε μία μόνο απεικόνιση.

Για παράδειγμα, αντί να απεικονίζει απλώς τις μεταβολές στον παγκόσμιο πληθυσμό με την πάροδο του χρόνου με ένα παραδοσιακό γράφημα στήλης στήλης, ένα στοιχειογραμμισμένο γράφημα μπορεί επίσης να αντιπροσωπεύει τη φυλετική σύνθεση του συνολικού πληθυσμού κατά τη διάρκεια κάθε έτους και τον τρόπο με τον οποίο οι αναλογίες αυτές άλλαξαν κατά την ίδια περίοδο.

U.S. Population by Race



Source: <http://www.originlab.com/doc/Tutorials/StackColumn-Labels>

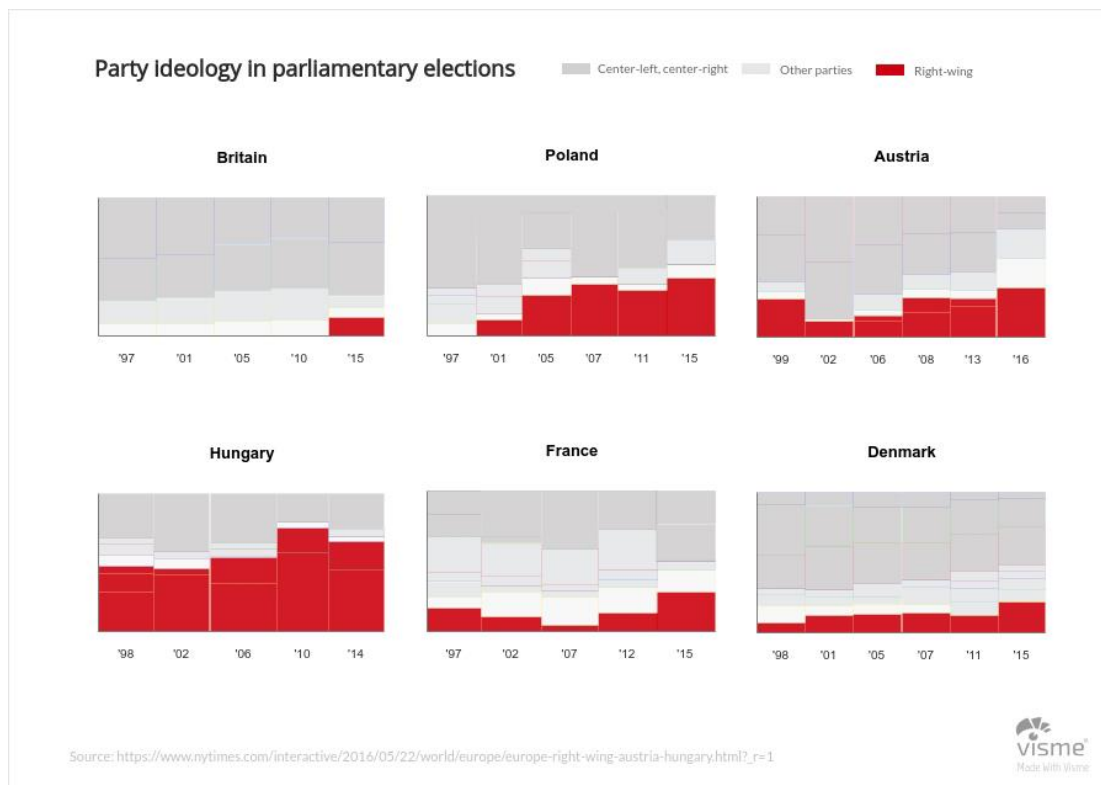


Trellis Bar Graphs

Κατά την παρουσίαση δεδομένων με τρεις μεταβλητές, ένας σχεδιαστής μπορεί να προσπαθήσει να δημιουργήσει ένα τρισδιάστατο γράφημα ράβδων, αλλά η προσθήκη ενός επιπλέον άξονα μπορεί μερικές φορές να φαίνεται γεμάτη και ασαφής, ειδικά σε έντυπη μορφή. Αντίθετα, μπορούν να παρουσιαστούν και άλλες μεταβλητές σε μορφή πέργκολας ή πλέγματος.

Συνδυάζοντας μια σειρά γραμμικών γραφικών παραστάσεων σε ένα αρθρωτό σχεδιασμό, μπορούν να συγκριθούν εύκολα και άλλα σύνολα δεδομένων. Για παράδειγμα, μια ενιαία γραφική παράσταση μπορεί να απεικονίσει την πολιτική κατανομή των εθνικών εκλογών της Πολωνίας για περίοδο πέντε ετών. Αλλά

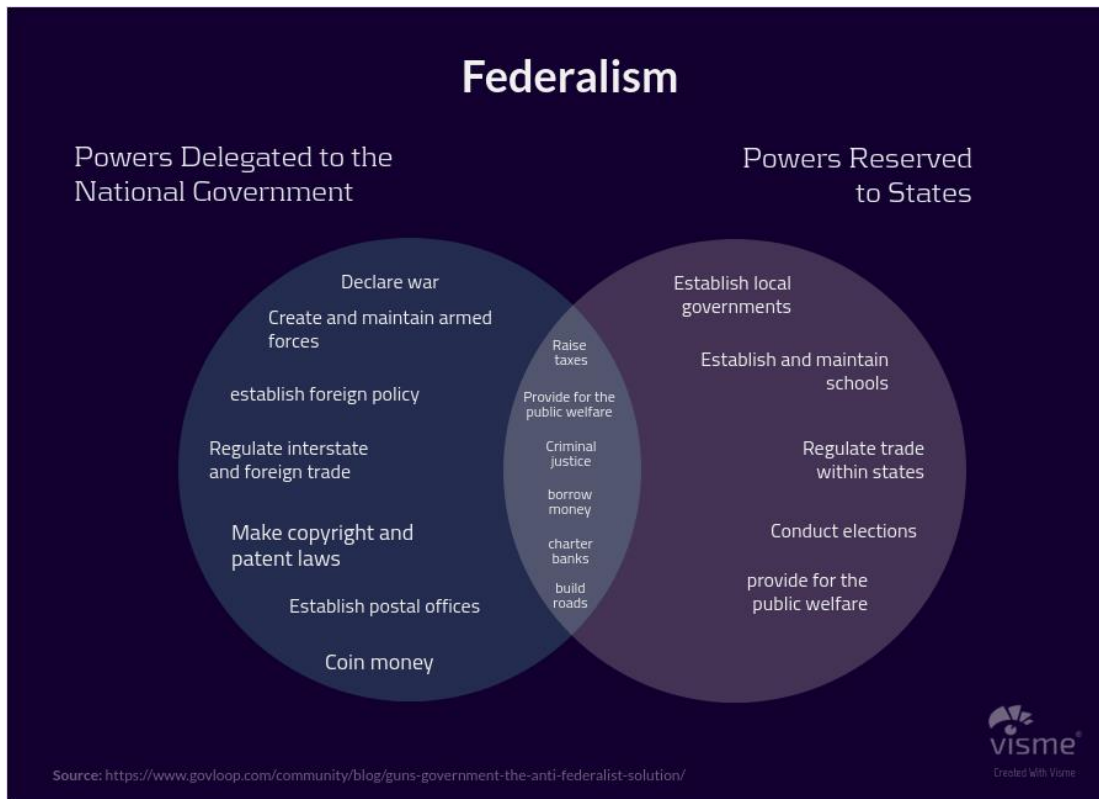
ένα γράφημα ράβδου πλήρης θα μπορούσε να απεικονίσει το ίδιο σύνολο δεδομένων για 16 ευρωπαϊκά έθνη.



Venn Diagrams

Το κλασικό διάγραμμα Venn, επίσης γνωστό ως λογικό διάγραμμα, απεικονίζει όλες τις πιθανές λογικές σχέσεις μεταξύ μιας καθορισμένης συλλογής συνόλων. Για παράδειγμα, η αλληλεπικάλυψη δύο κύκλων - ένας που εκπροσωπεί την εθνική κυβέρνηση και την άλλη κρατική κυβέρνηση - αντιπροσωπεύει οπτικά τις δυνάμεις που διατηρούν και τα δύο όργανα.

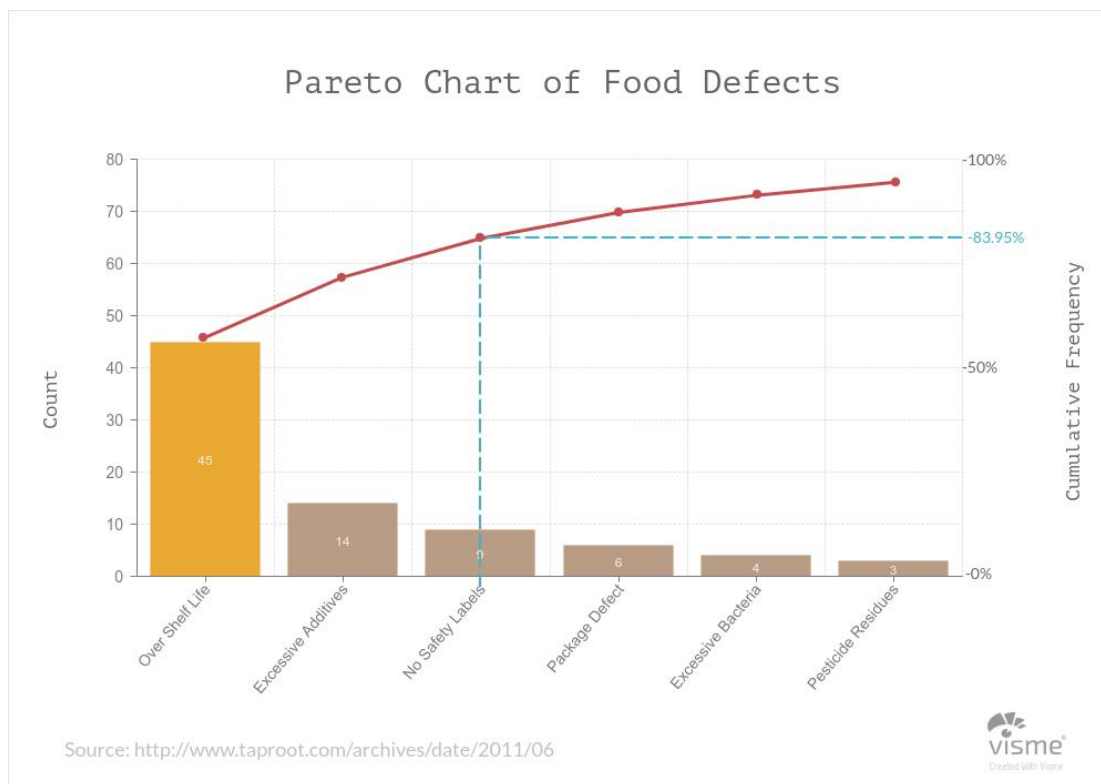
Οι περισσότεροι κύκλοι που χρησιμοποιούνται, τα πιο λογικά συμπεράσματα που μπορούν να αναπαρασταθούν από την επικάλυψή τους. Το συνδυασμένο σύνολο όλων των δεδομένων στο διάγραμμα είναι γνωστό ως ένωση, ενώ οι περιοχές που επικαλύπτονται ονομάζονται διασταυρώσεις. Ένα διάγραμμα Venn στο οποίο το σχετικό μέγεθος και η έκταση κάθε σχήματος είναι ανάλογο με το μέγεθος της ομάδας που αντιπροσωπεύει είναι γνωστό ως διάγραμμα Venn ανάλογο προς την έκταση ή κλιμάκωση.



3.7.8 Επιστήμη

Pareto Charts

Μερικές φορές ένα βασικό γράφημα δεν εμφανίζει αρκετές πληροφορίες για να βγάλει το απαραίτητο συμπέρασμα. Ένα γράφημα Pareto συνδυάζει ένα γράφημα ράβδων με ένα γραμμικό γράφημα για να απεικονίσει όχι μόνο τις ατομικές τιμές των κατηγοριών αλλά και το αθροιστικό σύνολο του συνόλου. Τα διαγράμματα Pareto έχουν σχεδιαστεί για να τονίζουν το σημαντικότερο από μια σειρά παραγόντων. Για παράδειγμα, σε ένα γράφημα Pareto που παρακολουθεί τον τύπο και τη συχνότητα των ελαττωμάτων των τροφίμων, οι ράβδοι απεικονίζουν τις συνολικές εμφανίσεις κάθε είδους ελαττωμάτων - όπως αναφέρεται σε έναν από τους άξονες των διαγραμμάτων, ενώ η γραμμή καταγράφει τη σωρευτική συχνότητα όλων των κατηγοριών, από τις περισσότερες έως τις λιγότερο διαδεδομένες. Το αποτέλεσμα είναι ένα γράφημα που αντικατοπτρίζει σαφώς τα πιο συνηθισμένα ελαττώματα τροφίμων και ποιο ποσοστό αντιπροσωπεύει το καθένα.



Radar Charts

Ένα διάγραμμα ραντάρ, επίσης κοινώς αναφερόμενο ως διάγραμμα αράχνης ή διάγραμμα αστέρα, εμφανίζει σύνολα δεδομένων που αποτελούνται από τρεις ή περισσότερες μεταβλητές σε ένα διδιάστατο γραφικό. Η ποσοτική τιμή κάθε μεταβλητής αντανακλάται σε έναν άξονα που αρχίζει συνήθως στο κεντρικό σημείο του χάρτη.

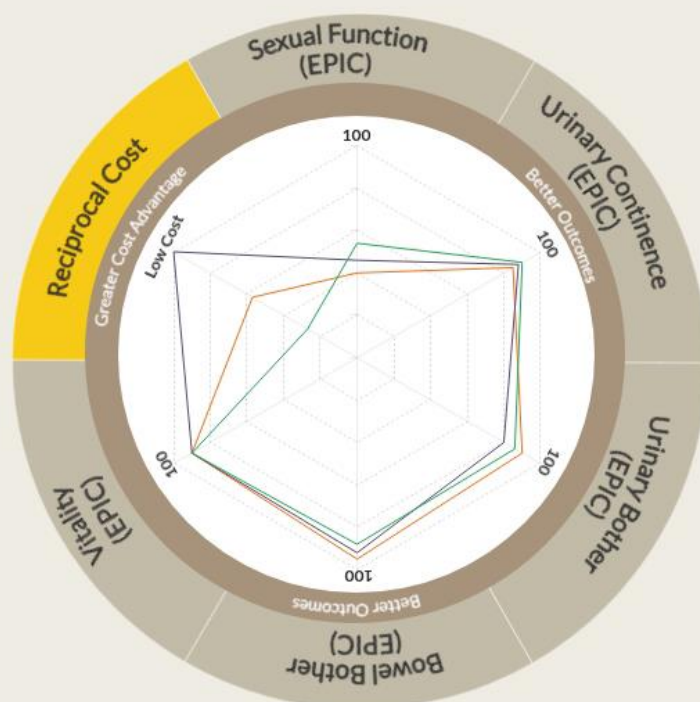
Καθώς οι μεταβλητές κάθε στοιχείου έχουν χαρτογραφηθεί, μια γραμμή συνδέει τα σημεία σε κάθε άξονα, σχηματίζοντας ένα ακανόνιστο πολύγωνο που μπορεί ή δεν μπορεί να μοιάζει με ένα αστέρι ή ιστό αράχνης. Πολλαπλά σύνολα δεδομένων μπορούν να συγκριθούν σε ένα μόνο γράφημα ραντάρ, αντιπροσωπεύοντας το καθένα με διαφορετικό χρώμα, προσδιορισμένο από ετικέτες ή σε συνοδευτικό κλειδί.

Ένα διάγραμμα ραντάρ μπορεί, για παράδειγμα, να συγκρίνει και να απεικονίσει σαφώς το κόστος και τα αποτελέσματα διαφόρων ιατρικών διαδικασιών, καθώς σχετίζονται με πολλαπλές συνθήκες - όλα σε ένα ενιαίο γραφικό.

Radar Chart Displays Patient Value (Outcome & Cost)

Comparing Three Types of Cancer Treatment

Brachytherapy (LDR) Proton (IMPT) Robotic Prostatectomy



Source: <http://catalyst.nejm.org/my-favorite-slide-health-care-value-on-the-radar/>

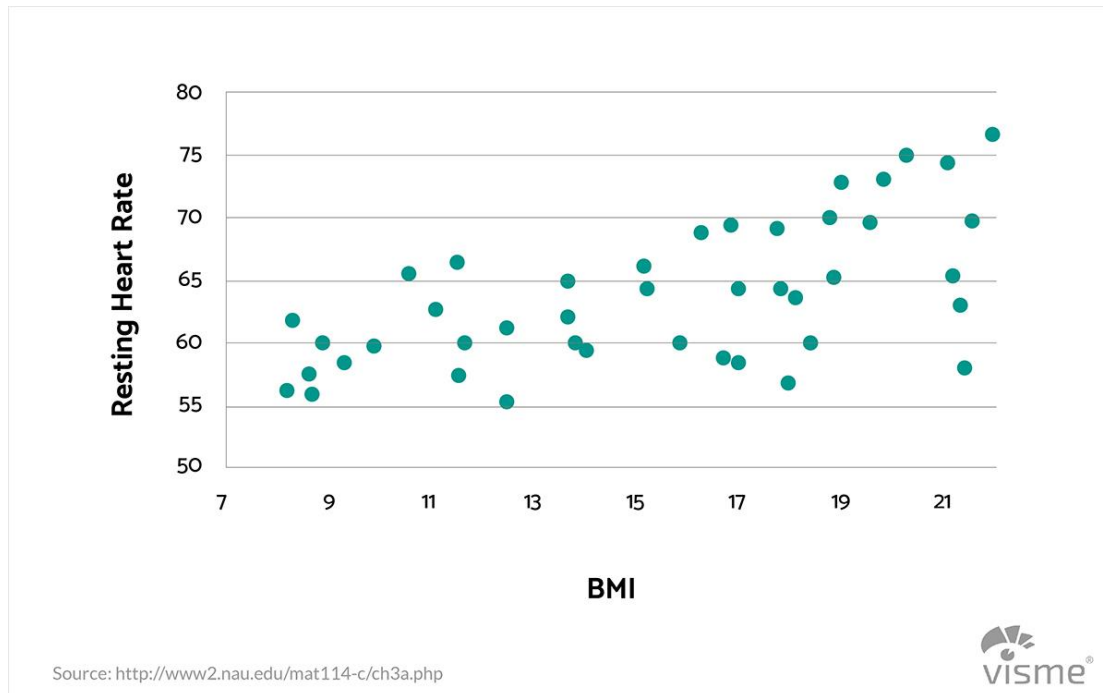
visme
Created with Visme

Stattergrams

Τα scattergrams, επίσης γνωστά ως διαγράμματα scatter, είναι γραφήματα που δείχνουν τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών. Τα οικόπεδα χρησιμοποιούν μαθηματικές συντεταγμένες για να αντιπροσωπεύουν δύο μεταβλητές ενός συνόλου δεδομένων.

Τα δεδομένα εμφανίζονται σε ένα scattergram ως συλλογή σημείων, το καθένα αντιπροσωπεύει τις μεταβλητές τιμών που απεικονίζονται σε έναν οριζόντιο και κατακόρυφο άξονα. Αν τα σημεία έχουν χρωματική κωδικοποίηση, μια πρόσθετη μεταβλητή μπορεί να αναπαριστάται σε ένα μόνο γράφημα. Σχεδιάζοντας ορισμένα σύνολα δεδομένων, οι επιστήμονες μπορούν να ανακαλύψουν τάσεις τις οποίες δεν θα γνωρίζουν διαφορετικά. Για παράδειγμα, ένα scattergram μπορεί να επιτρέψει σε έναν γιατρό να

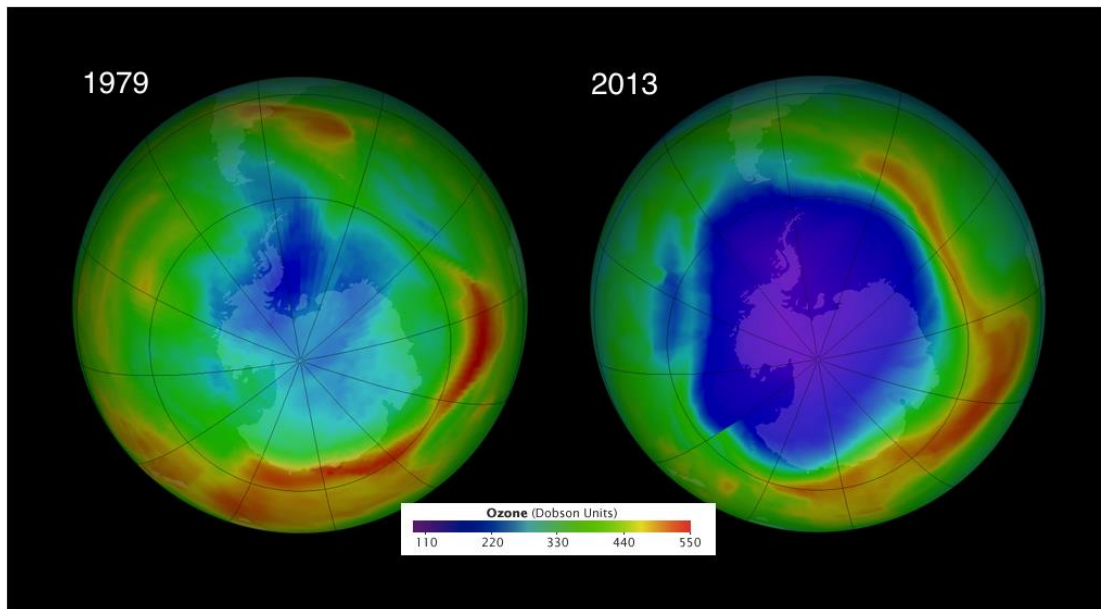
σχεδιάσει τα ποσοστά καρδιάς ανάπαυσης των ασθενών έναντι των αριθμών του δείκτη μάζας σώματος. Το προκύπτον γράφημα αποκαλύπτει ότι ένας υψηλότερος καρδιακός ρυθμός συσχετίζεται με υψηλότερο ΔΜΣ.



Spherical Contour Graphs

Η κατανομή των πλανητικών συνθηκών σε ένα βασικό γράφημα δύο αξόνων μπορεί να δημιουργήσει πρόβλημα. Η Γη, μετά από όλα, είναι μια σφαίρα. Αντ' αυτού, τα δεδομένα μπορούν να γραφούν σε ένα πεδίο τριών αξόνων χρησιμοποιώντας μεταβλητές των x , y και z . Η προκύπτουσα πλοκή, εάν ολοκληρωθεί, θα έχει τη μορφή σφαίρας.

Μια σφαιρική πλοκή μπορεί, για παράδειγμα, να αποκαλύψει τις παγκόσμιες θερμοκρασίες ή τις τάσεις βροχόπτωσης, εκχωρώντας κάθε περιοχή τιμών με ένα συγκεκριμένο χρώμα και σχεδιάζοντας τα δεδομένα με σημεία της αντίστοιχης απόχρωσης.

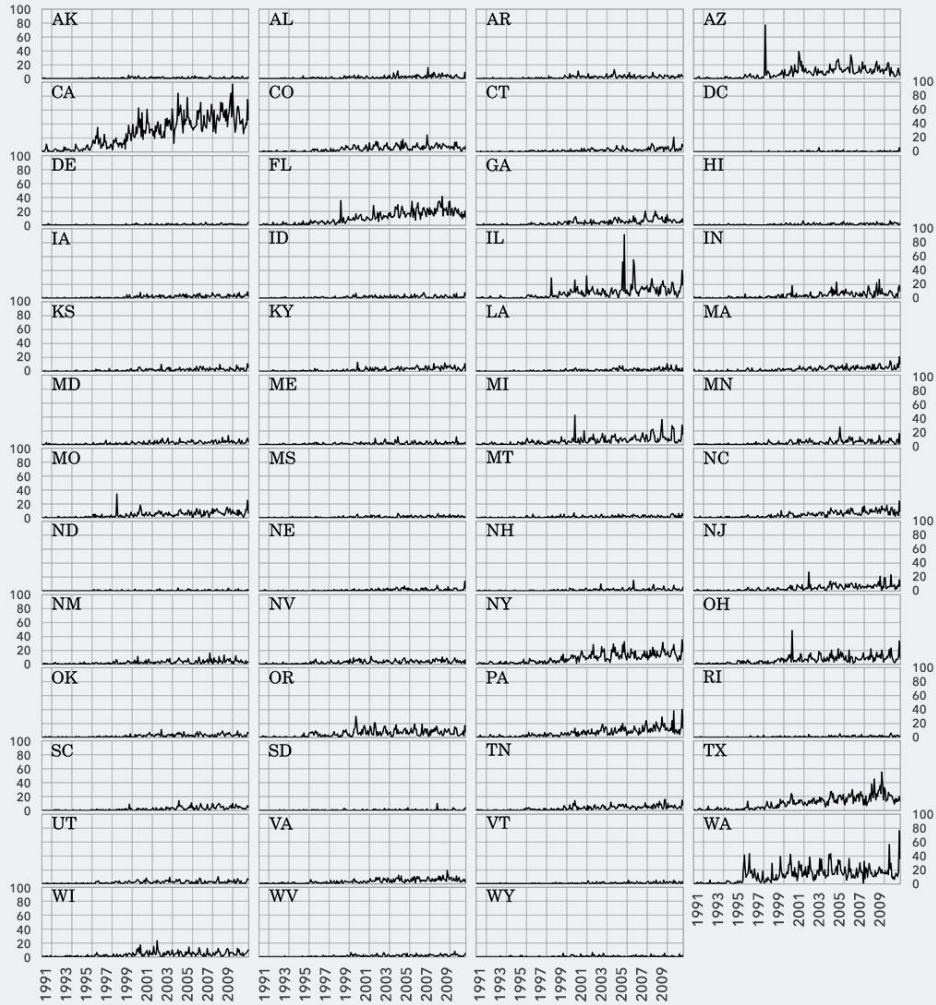


Trellis line Graphs

Τα γραφήματα Trellis επιτρέπουν στους επιστήμονες να εξετάζουν σύνθετα σύνολα δεδομένων πολλαπλών μεταβλητών, συγκρίνοντας ταυτόχρονα περισσότερες πληροφορίες. Ενώ ένα γράφημα γραμμής μπορεί να απεικονίσει μηνιαίες παρατηρήσεις UFO στο Τενεσί κατά τη διάρκεια μιας περιόδου 18 ετών, ένα γράφημα γραμμής πέργκολας θα εμφανίσει τα ίδια δεδομένα για όλες τις 50 πολιτείες σε ένα μόνο γραφικό.

Ένα γράφημα γραμμής πέργκολας βασίζεται στην ίδια αρχή με το απλούστερο αντίγραφο του, σχεδιάζοντας τις τάσεις σε ένα σύνολο δεδομένων που αποτελείται από δύο μεταβλητές - τον αριθμό των παρατηρήσεων UFO και ημερομηνιών - μέσω της χρήσης σημείων σύνδεσης σε δύο άξονες. Αλλά με το συνδυασμό των γραφημάτων πολλαπλών γραμμών σε αρθρωτή μορφή, αντιπροσωπεύεται μια πρόσθετη μεταβλητή - θέση.

Monthly UFO Sightings by U.S. State January 1990 through August 2010



Source: <http://slendermeans.org/ml4h-ch1-p5.html>

4 Η οπτικοποίηση

Bitcoin ▼

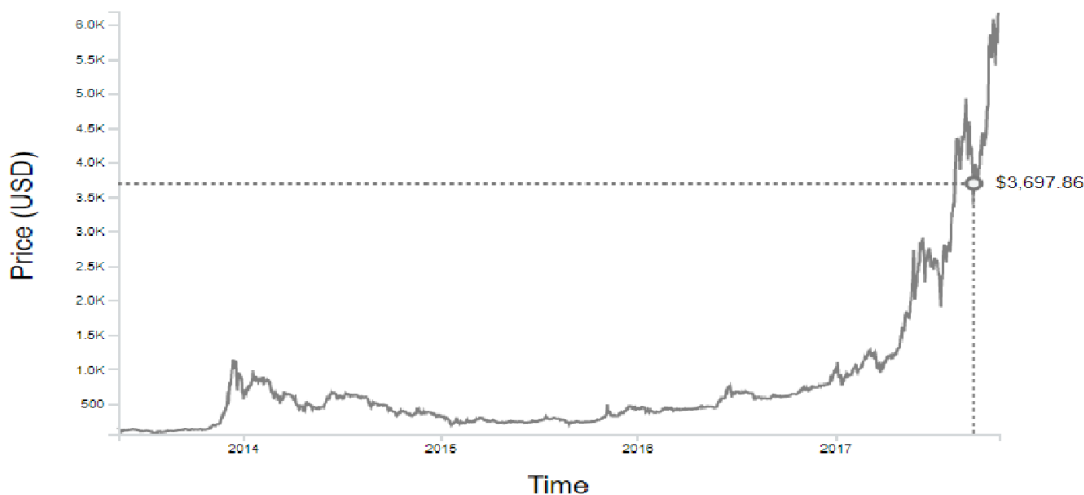
- Bitcoin
- Ethereum
- Bitcoin Cash
- Litecoin
- Ripple

Price in dollars ▼

- Price in dollars
- Market capitalization
- 24 hour trading volume

π

Date: 12/05/2013 - 31/10/2017



4.1 Οι τεχνολογίες

4.1.1 HTML

Η **HTML** (αρχικοποίηση του αγγλικού **HyperText Markup Language**, ελλ. Γλώσσα Σήμανσης Υπερκειμένου) είναι η κύρια γλώσσα σήμανσης για τις ιστοσελίδες, και τα στοιχεία της είναι τα βασικά δομικά στοιχεία των ιστοσελίδων.

Η HTML γράφεται υπό μορφή στοιχείων `H_TML` τα οποία αποτελούνται από *ετικέτες* (tags), οι οποίες περικλείονται μέσα σε σύμβολα «μεγαλύτερο από» και «μικρότερο από» (για παράδειγμα `<html>`), μέσα στο περιεχόμενο της ιστοσελίδας. Οι ετικέτες HTML συνήθως λειτουργούν ανά ζεύγη (για παράδειγμα `<h1>` και `</h1>`), με την πρώτη να ονομάζεται *ετικέτα έναρξης* και τη δεύτερη *ετικέτα λήξης* (ή σε άλλες περιπτώσεις *ετικέτα ανοίγματος* και *ετικέτα κλεισίματος*).

αντίστοιχα). Ανάμεσα στις ετικέτες, οι σχεδιαστές ιστοσελίδων μπορούν να τοποθετήσουν κείμενο, πίνακες, εικόνες κλπ.

Ο σκοπός ενός web browser είναι να διαβάσει τα έγγραφα HTML και να τα συνθέσει σε σελίδες που μπορεί κανείς να διαβάσει ή να ακούσει. Ο browser δεν εμφανίζει τις ετικέτες HTML, αλλά τις χρησιμοποιεί για να παρουσιάσει το περιεχόμενο της σελίδας.

Τα στοιχεία της HTML χρησιμοποιούνται για να κτίσουν όλους του ιστότοπους. Η HTML επιτρέπει την ενσωμάτωση εικόνων και άλλων αντικειμένων μέσα στη σελίδα, και μπορεί να χρησιμοποιηθεί για να εμφανίσει διαδραστικές φόρμες. Παρέχει τις μεθόδους δημιουργίας δομημένων εγγράφων (δηλαδή εγγράφων που αποτελούνται από το περιεχόμενο που μεταφέρουν και από τον κώδικα μορφοποίησης του περιεχομένου) καθορίζοντας δομικά σημαντικά στοιχεία για το κείμενο, όπως κεφαλίδες, παραγράφους, λίστες, συνδέσμους, παραθέσεις και άλλα. Μπορούν επίσης να ενσωματώνονται σενάρια εντολών σε γλώσσες όπως η JavaScript, τα οποία επηρεάζουν τη συμπεριφορά των ιστοσελίδων HTML και από στατικές τις κάνουν διαδραστικές.

Οι Web browsers μπορούν επίσης να αναφέρονται σε στυλ μορφοποίησης CSS για να ορίζουν την εμφάνιση και τη διάταξη του κειμένου και του υπόλοιπου υλικού. Ο οργανισμός W3C, ο οποίος δημιουργεί και συντηρεί τα πρότυπα για την HTML και τα CSS, ενθαρρύνει τη χρήση των CSS αντί διαφόρων στοιχείων της HTML για σκοπούς παρουσίασης του περιεχομένου.

4.1.2 CSS

Η CSS (*Cascading Style Sheets* – διαδοχικά φύλλα ύφους ή επάλληλα φύλλα ύφους) είναι μια γλώσσα υπολογιστή που ανήκει στην κατηγορία των γλωσσών φύλλων ύφους που χρησιμοποιείται για τον έλεγχο της εμφάνισης ενός εγγράφου που έχει γραφτεί με μια γλώσσα σήμανσης.

Χρησιμοποιείται δηλαδή για τον έλεγχο της εμφάνισης ενός εγγράφου που γράφτηκε στις γλώσσες HTML και XHTML, δηλαδή για τον έλεγχο της εμφάνισης μιας ιστοσελίδας και γενικότερα ενός ιστοτόπου. Η CSS είναι μια γλώσσα υπολογιστή προορισμένη να αναπτύσσει στυλιστικά μια ιστοσελίδα δηλαδή να διαμορφώνει περισσότερα χαρακτηριστικά, χρώματα,

στοίχιση και δίνει περισσότερες δυνατότητες σε σχέση με την html. Για μια όμορφη και καλοσχεδιασμένη ιστοσελίδα η χρήση της CSS κρίνεται ως απαραίτητη.

4.1.3 Javascript

Η **JavaScript (JS)** είναι διερμηνευμένη γλώσσα προγραμματισμού για ηλεκτρονικούς υπολογιστές. Αρχικά αποτέλεσε μέρος της υλοποίησης των φυλλομετρητών Ιστού, ώστε τα σενάρια από την πλευρά του πελάτη (client-side scripts) να μπορούν να επικοινωνούν με τον χρήστη, να ανταλλάσσουν δεδομένα ασύγχρονα και να αλλάζουν δυναμικά το περιεχόμενο του εγγράφου που εμφανίζεται.

Η JavaScript είναι μια γλώσσα σεναρίων που βασίζεται στα πρωτότυπα (prototype-based), είναι δυναμική, με ασθενείς τύπους και έχει συναρτήσεις ως αντικείμενα πρώτης τάξης. Η σύνταξή της είναι επηρεασμένη από τη C. Η JavaScript αντιγράφει πολλά ονόματα και συμβάσεις ονοματοδοσίας από τη Java, αλλά γενικά οι δύο αυτές γλώσσες δε σχετίζονται και έχουν πολύ διαφορετική σημασιολογία. Οι βασικές αρχές σχεδιασμού της JavaScript προέρχονται από τις γλώσσες προγραμματισμού Self και Scheme. Είναι γλώσσα βασισμένη σε διαφορετικά προγραμματιστικά παραδείγματα (multi-paradigm), υποστηρίζοντας αντικειμενοστρεφές, προστακτικό και συναρτησιακό στυλ προγραμματισμού.

Η JavaScript χρησιμοποιείται και σε εφαρμογές εκτός ιστοσελίδων — τέτοια παραδείγματα είναι τα έγγραφα PDF, οι εξειδικευμένοι φυλλομετρητές (site-specific browsers) και οι μικρές εφαρμογές της επιφάνειας εργασίας (desktop widgets). Οι νεότερες εικονικές μηχανές και πλαίσια ανάπτυξης για JavaScript (όπως το Node.js) έχουν επίσης κάνει τη JavaScript πιο δημοφιλή για την ανάπτυξη εφαρμογών Ιστού στην πλευρά του διακομιστή (server-side).

Το πρότυπο της γλώσσας κατά τον οργανισμό τυποποίησης ECMA ονομάζεται **ECMAScript**.

4.1.4 D3.js(data driven documents)

Το D3.js (επίσης γνωστό ως D3, σύντομο για Data Driven Documents) είναι μια βιβλιοθήκη JavaScript για την παραγωγή δυναμικών, διαδραστικών απεικονίσεων δεδομένων σε προγράμματα περιήγησης

ιστού. Χρησιμοποιεί τα ευρέως εφαρμοζόμενα πρότυπα Scalable Vector Graphics (SVG), HTML5 και CSS (Cascading Style Sheets). Είναι ο διάδοχος του προηγούμενου πλαισίου Protovis. Σε αντίθεση με πολλές άλλες βιβλιοθήκες, το D3.js επιτρέπει τον εξαιρετικό έλεγχο του τελικού οπτικού αποτελέσματος. Η ανάπτυξή της σημειώθηκε το 2011, καθώς η έκδοση 2.0.0 κυκλοφόρησε τον Αύγουστο του 2011. Το D3.js χρησιμοποιείται σε εκατοντάδες χιλιάδες ιστότοπους. Ορισμένες δημοφιλείς χρήσεις περιλαμβάνουν τη δημιουργία διαδραστικών γραφικών για δικτυακούς τόπους ειδήσεων σε απευθείας σύνδεση, πίνακες πληροφοριών για την προβολή δεδομένων και την παραγωγή χαρτών από δεδομένα δημιουργίας χάρτη GIS. Επιπλέον, ο εξαγώνιμος χαρακτήρας του SVG επιτρέπει τη χρήση γραφικών που δημιουργούνται από το D3 σε έντυπες εκδόσεις.

4.2 Η ιστορία

Δημιουργώντας μία οπτικοποίηση θέλω να δείξω πως τα πέντε μεγαλύτερα κρυπτονομίσματα έχουν αλλάξει τιμή στο πέρασμα του χρόνου. Έχω βρει ιστορικά δεδομένα και τα έχω μετατρέψει σε μία εύχρηστη μορφή. Για κάθε νόμισμα έχω δεδομένα για κάθε μέρα που ανταλλάσετε δείχνοντας την τιμή του σε δολάρια, τον όγκο συναλλαγών της προηγούμενης ημέρας και την συνολική κεφαλαιοποίηση της αγοράς γράφημα επιτρέπει επίσης στον χρήστη να περιορίζει αυτόματα ένα συγκεκριμένο εύρος ημερομηνιών με το γράφημα αυτόματα προσαρμόζοντας έτσι ώστε να καταλαμβάνει ολόκληρη την οθόνη.

4.3 Η υλοποίηση

4.3.1 CSS

```
#chart-area svg {  
  margin-left: auto;  
  margin-right: auto;  
  display: block;  
}  
  
#logo {  
  height:50px;  
}
```

```
.navbar-brand {  
  height: 60px;  
  padding: 5px 0px;  
}
```

```
.axis {  
  font: 10px sans-serif;  
}
```

```
.axis path,  
.axis line {  
  fill: none;  
  stroke: #D4D8DA;  
  stroke-width: 2px;  
  shape-rendering: crispEdges;  
}
```

```
.line {  
  fill: none;  
  stroke-width: 2px;  
}
```

```
.overlay {  
  fill: none;  
  pointer-events: all;  
}
```

```
.focus circle {
```

```
fill: #F1F3F3;
stroke: #777;
stroke-width: 3px;
}
```

```
.hover-line {
stroke: #777;
stroke-width: 2px;
stroke-dasharray: 3,3;
}
```

```
#selections .col-md-4 {
padding-top: 10px;
padding-bottom: 10px;
}
```

4.3.2 Data

```
{
  "bitcoin": [
    {
      "24h_vol": null,
      "date": "12/5/2013",
      "market_cap": null,
      "price_usd": null
    },
    {
      "24h_vol": null,
      "date": "13/5/2013",
      "market_cap": null,
```

```

    "price_usd": null
  },
  {
    "24h_vol": "0",
    "date": "14/5/2013",
    "market_cap": "1500517590",
    "price_usd": "135.3"
  },
  {
    "24h_vol": "0",
    "date": "15/5/2013",
    "market_cap": "1575032004",
    "price_usd": "141.96"
  },
  {
    "24h_vol": "0",
    "date": "16/5/2013",
    "market_cap": "1501657492",
    "price_usd": "135.3"
  },
  }, {...}
]
}

```

4.3.3 Javascript

```
var margin = { left:80, right:100, top:50, bottom:100 },
```

```
    height = 500 - margin.top - margin.bottom,
```

```
    width = 800 - margin.left - margin.right;
```

```
var svg = d3.select("#chart-area")
```



```

.append("svg")
.attr("width", width + margin.left + margin.right)
.attr("height", height + margin.top + margin.bottom)
var g = svg.append("g")
    .attr("transform", "translate(" + margin.left +
        ", " + margin.top + ")");

var t = function(){ return d3.transition().duration(1000); }

var parseTime = d3.timeParse("%d/%m/%Y");
var formatTime = d3.timeFormat("%d/%m/%Y");
var bisectDate = d3.bisector(function(d) { return d.date; }).left;

// Add the line for the first time
g.append("path")
    .attr("class", "line")
    .attr("fill", "none")
    .attr("stroke", "grey")
    .attr("stroke-width", "3px");

// Labels
var xLabel = g.append("text")
    .attr("class", "x axisLabel")
    .attr("y", height + 50)
    .attr("x", width / 2)
    .attr("font-size", "20px")
    .attr("text-anchor", "middle")
    .text("Time");
var yLabel = g.append("text")

```

```

.attr("class", "y axisLabel")
.attr("transform", "rotate(-90)")
.attr("y", -60)
.attr("x", -170)
.attr("font-size", "20px")
.attr("text-anchor", "middle")
.text("Price (USD)")

// Scales
var x = d3.scaleTime().range([0, width]);
var y = d3.scaleLinear().range([height, 0]);

// X-axis
var xAxisCall = d3.axisBottom()
  .ticks(4);
var xAxis = g.append("g")
  .attr("class", "x axis")
  .attr("transform", "translate(0," + height + ")");

// Y-axis
var yAxisCall = d3.axisLeft()
var yAxis = g.append("g")
  .attr("class", "y axis");

// Event listeners
$("#coin-select").on("change", update)
$("#var-select").on("change", update)

// Add jQuery UI slider

```

```

$("#date-slider").slider({
  range: true,
  max: parseTime("31/10/2017").getTime(),
  min: parseTime("12/5/2013").getTime(),
  step: 86400000, // One day
  values: [parseTime("12/5/2013").getTime(), parseTime("31/10/2017").getTime()],
  slide: function(event, ui){
    $("#dateLabel1").text(formatTime(new Date(ui.values[0])));
    $("#dateLabel2").text(formatTime(new Date(ui.values[1])));
    update();
  }
});

```

```

d3.json("data/coins.json").then(function(data){
  // console.log(data);

  // Prepare and clean data
  filteredData = {};
  for (var coin in data) {
    if (!data.hasOwnProperty(coin)) {
      continue;
    }
    filteredData[coin] = data[coin].filter(function(d){
      return !(d["price_usd"] == null)
    })
    filteredData[coin].forEach(function(d){
      d["price_usd"] = +d["price_usd"];
      d["24h_vol"] = +d["24h_vol"];
      d["market_cap"] = +d["market_cap"];
    });
  }
});

```

```

        d["date"] = parseTime(d["date"])
    });
}

// Run the visualization for the first time
update();
})

function update() {
    // Filter data based on selections
    var coin = $("#coin-select").val(),
        yValue = $("#var-select").val(),
        sliderValues = $("#date-slider").slider("values");
    var dataTimeFiltered = filteredData[coin].filter(function(d){
        return ((d.date >= sliderValues[0]) && (d.date <= sliderValues[1]))
    });

    // Update scales
    x.domain(d3.extent(dataTimeFiltered, function(d){ return d.date; }));
    y.domain([d3.min(dataTimeFiltered, function(d){ return d[yValue]; }) / 1.005,
        d3.max(dataTimeFiltered, function(d){ return d[yValue]; }) * 1.005]);

    // Fix for format values
    var formatSi = d3.format(".2s");
    function formatAbbreviation(x) {
        var s = formatSi(x);
        switch (s[s.length - 1]) {
            case "G": return s.slice(0, -1) + "B";
            case "k": return s.slice(0, -1) + "K";
        }
    }
}

```

```

    }
    return s;
}

// Update axes
xAxisCall.scale(x);
xAxis.transition(t()).call(xAxisCall);
yAxisCall.scale(y);
yAxis.transition(t()).call(yAxisCall.tickFormat(formatAbbreviation));

// Clear old tooltips
d3.select(".focus").remove();
d3.select(".overlay").remove();

// Tooltip code
var focus = g.append("g")
    .attr("class", "focus")
    .style("display", "none");
focus.append("line")
    .attr("class", "x-hover-line hover-line")
    .attr("y1", 0)
    .attr("y2", height);
focus.append("line")
    .attr("class", "y-hover-line hover-line")
    .attr("x1", 0)
    .attr("x2", width);
focus.append("circle")
    .attr("r", 5);
focus.append("text")

```

```

.attr("x", 15)
.attr("dy", ".31em");
svg.append("rect")
  .attr("transform", "translate(" + margin.left + "," + margin.top + ")")
  .attr("class", "overlay")
  .attr("width", width)
  .attr("height", height)
  .on("mouseover", function() { focus.style("display", null); })
  .on("mouseout", function() { focus.style("display", "none"); })
  .on("mousemove", mousemove);

function mousemove() {
  var x0 = x.invert(d3.mouse(this)[0]),
      i = bisectDate(dataTimeFiltered, x0, 1),
      d0 = dataTimeFiltered[i - 1],
      d1 = dataTimeFiltered[i],
      d = (d1 && d0) ? (x0 - d0.date > d1.date - x0 ? d1 : d0) : 0;
  focus.attr("transform", "translate(" + x(d.date) + "," + y(d[yValue]) + ")");
  focus.select("text").text(function() { return d3.format("$")(d[yValue].toFixed(2)); });
  focus.select(".x-hover-line").attr("y2", height - y(d[yValue]));
  focus.select(".y-hover-line").attr("x2", -x(d.date));
}

// Path generator
line = d3.line()
  .x(function(d){ return x(d.date); })
  .y(function(d){ return y(d[yValue]); });

// Update our line path

```

```

g.select(".line")
    .transition(t)
    .attr("d", line(dataTimeFiltered));

// Update y-axis label
var newText = (yValue === "price_usd") ? "Price (USD)" :
    ((yValue === "market_cap") ? "Market Capitalization (USD)" : "24 Hour Trading Volume (USD)")
yLabel.text(newText);
}

```

4.3.2 HTML

```

<!doctype html>
<html>
  <head>
    <meta charset="utf-8">
    <meta name="description" content="">
    <title>CoinStats</title>
    <!-- Bootstrap Core CSS -->
    <link rel="stylesheet" href="css/bootstrap.min.css">
    <!-- jQuery UI CSS -->
    <link rel="stylesheet" href="css/jquery-ui.min.css">
    <link rel="stylesheet" href="css/jquery-ui.structure.min.css">
    <link rel="stylesheet" href="css/jquery-ui.theme.min.css">
    <!-- Custom CSS -->
    <link rel="stylesheet" href="css/style.css">
  </head>

  <body>
    <nav class="navbar navbar-default">

```

```

<div class="container">
  <a class="navbar-brand" href="#"></a>
</div>
</nav>

<!-- Bootstrap grid setup -->
<div class="container">
  <div id="selections" class="row">
    <div class="col-md-4">
      <div id="slider-div">
        <label>Date: <span id="dateLabel1">12/05/2013</span> - <span
id="dateLabel2">31/10/2017</span></label>
        <div id="date-slider"></div>
      </div>
    </div>
    <div class="col-md-4">
      <select id="coin-select" class="form-control">
        <option selected value="bitcoin">Bitcoin</option>
        <option value="ethereum">Ethereum</option>
        <option value="bitcoin_cash">Bitcoin Cash</option>
        <option value="litecoin">Litecoin</option>
        <option value="ripple">Ripple</option>
      </select>
    </div>
    <div class="col-md-4">
      <select id="var-select" class="form-control">
        <option selected value="price_usd">Price in dollars</option>
        <option value="market_cap">Market capitalization</option>
        <option value="24h_vol">24 hour trading volume</option>
      </select>
    </div>
  </div>

```



```
        </select>
    </div>
</div>
<div class="row">
    <div id="chart-area"></div>
</div>
</div>

<!-- External JS libraries -->
<script src="js/jquery.min.js"></script>
<script src="js/bootstrap.min.js"></script>
<script src="js/jquery-ui.min.js"></script>
<script src="js/d3.min.js"></script>
<!-- Custom JS -->
<script src="js/main.js"></script>

</body>
</html>
```

5 Βιβλιογραφία

Ελένη Γολέμη.,(2010).Κρυπτογραφία & Εξόρυξη Δεδομένων. Ανακτήθηκε στις 16 Ιουλίου από

<http://nemertes.lis.upatras.gr/jspui/bitstream/10889/4791/1/ergasia-golemie.pdf>

Kantardzic, Mehmed (2003). Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons. ISBN 0-471-22852-4. OCLC 50055336.

Fayyad, Usama (1996). From Data Mining to Knowledge Discovery in Databases (PDF). Ανακτήθηκε στις 16 Ιουλίου 2012

Simmi Bagga., Dr. G.N. Singh., (2012).Applications of Data Mining.Ανακτήθηκε στις 19 Απριλίου ,2012 από <http://www.ijset.com/images/P5.pdf> Αρχειοθετήθηκε 2016-11-23 στο Wayback Machine.

Γούλου Ζωή.,(2010). Εφαρμογή μεθόδων εξόρυξης δεδομένων στη διαχείριση πελατειακών σχέσεων.

Ανακτήθηκε στις 18 Ιουλίου από

<http://dspace.lib.uom.gr/bitstream/2159/14808/6/GoulouZoiMsc2012.pdf>

Data Science App by Aduch Technologies

Visme.co

"Stephen Few-Perceptual Edge-Selecting the Right Graph for Your Message-2004" (PDF). Archived (PDF) from the original on 2014-10-05. Retrieved 2014-09-08.

Manuela Aparicio and Carlos J. Costa (November 2014). "Data visualization". Communication Design Quarterly Review. 3 (1): 7–11. doi:10.1145/2721882.2721883.

Nikos Bikaks (2018) "Big Data Visualization Tools" Encyclopedia of Big Data Technologies, Springer 2018.

Press, Gil. "A Very Short History Of Data Science". Archived from the original on 2017-09-11. Retrieved 2017-08-26.

Vitaly Friedman (2008) "Data Visualization and Infographics" Archived 2008-07-22 at the Wayback Machine in: *Graphics*, Monday Inspiration, January 14th, 2008.

Fernanda Viegas and Martin Wattenberg (April 19, 2011). "How To Make Data Look Sexy". CNN.com. Archived from the original on May 6, 2011. Retrieved May 7, 2017.

Frits H. Post, Gregory M. Nielson and Georges-Pierre Bonneau (2002). *Data Visualization: The State of the Art*. Research paper TU delft, 2002. Archived 2009-10-07 at the Wayback Machine

Tukey, John (1977). *Exploratory Data Analysis*. Addison-Wesley. ISBN 0-201-07616-0.

techatstate (7 August 2013). "Tech@State: Data Visualization - Keynote by Dr Edward Tufte" Archived from the original on 29 March 2017. Retrieved 29 November 2016 – via YouTube.

"Archived copy" (PDF). Archived (PDF) from the original on 2018-10-20. Retrieved 2018-10-20.

Tufte, Edward (1983). *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press. ISBN 0-9613921-4-2. Archived from the original on 2013-01-14. Retrieved 2019-08-10.

"Telling Visual Stories About Data - Congressional Budget Office". www.cbo.gov. Archived from the original on 2014-12-04. Retrieved 2014-11-27.

"Stephen Few-Perceptual Edge-Graph Selection Matrix" (PDF). Archived (PDF) from the original on 2014-10-05. Retrieved 2014-09-08.

"Steven Few-Tapping the Power of Visual Perception-September 2004" (PDF). Archived (PDF) from the original on 2014-10-05. Retrieved 2014-10-08.

"Data Visualization for Human Perception". The Interaction Design Foundation. Archived from the original on 2015-11-23. Retrieved 2015-11-23.

"Visualization" (PDF). SFU. SFU lecture. Archived from the original (PDF) on 2016-01-22. Retrieved 2015-11-22.

Friendly, Michael. "A Brief History of Data Visualization". Springer-Verlag. Archived from the original on 1 December 2017. Retrieved 19 November 2017.

Whitehouse, D. (9 August 2000). "Ice Age star map discovered". BBC News. Archived from the original on 6 January 2018. Retrieved 20 January 2018.

Dragicevic, Pierre; Jansen, Yvonne (2012). "List of Physical Visualizations and Related Artefacts". Archived from the original on 2018-01-13. Retrieved 2018-01-12.

Jansen, Yvonne; Dragicevic, Pierre; Isenberg, Petra; Alexander, Jason; Karnik, Abhijit; Kildal, Johan; Subramanian, Sriram; Hornbaek, Kasper (2015). "Opportunities and challenges for data physicalization".

Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems: 3227–3236. Archived from the original on 2018-01-13. Retrieved 2018-01-12.

Friendly, Michael (2001). "Milestones in the history of thematic cartography, statistical graphics, and data visualization". Archived from the original on 2014-04-14. Retrieved 2017-11-19.

*Funkhouser, Howard Gray (Jan 1936). "A Note on a Tenth Century Graph". *Osiris*. 1: 260–262. doi:10.1086/368425. JSTOR 301609.*

Friendly, Michael (2006). "A Brief History of Data Visualization" (PDF). York University. Springer-Verlag. Archived (PDF) from the original on 2016-05-08. Retrieved 2015-11-22.

*"NY gets new boot camp for data scientists: It's free but harder to get into than Harvard". *Venture Beat*. Archived from the original on 2016-02-15. Retrieved 2016-02-21.*

"Steven Few-Selecting the Right Graph for Your Message-September 2004" (PDF). Archived (PDF) from the original on 2014-10-05. Retrieved 2014-09-08.

Lengler, Ralph; Eppler, Martin. J "Periodic Table of Visualization Methods". www.visual-literacy.org. Archived from the original on 16 March 2013. Retrieved 15 March 2013.

Flanagan, 2006, σ. 1

«ECMAScript Language Overview» (PDF). 23 Οκτωβρίου 2007. σελ. 4. Αρχειοθετήθηκε από το πρωτότυπο (PDF) στις 26 Μαρτίου 2009. Ανακτήθηκε στις 3 Μαΐου 2009.

«ECMAScript Language Specification» (PDF). Αρχειοθετήθηκε από το πρωτότυπο (PDF) στις 12 Απριλίου 2015. Ανακτήθηκε στις 7 Φεβρουαρίου 2013.

Douglas Crockford on Functional JavaScript >2:49< **blinkx** (flv) (αρχειοθετήθηκε από το πρωτότυπο 2009-09-23) (ανακτήθηκε 2013-02-07)

The Little JavaScripter shows the relationship with Scheme in more detail.

Πρότυπο ECMA-262

"d3 Releases". [Github.com](https://github.com).

"For Protovis Users", [Mboostock.github.com](https://github.com), retrieved August 18, 2012

Viau, Christophe (June 26, 2012), "What's behind our Business Infographics Designer? D3.js of course", Datameer's blog, retrieved August 18, 2012

Myatt, Glenn J.; Johnson, Wayne P. (September 2011), "5.10 Further reading", Making Sense of Data III: A Practical Guide to Designing Interactive Data Visualizations, Hoboken, New Jersey: John Wiley & Sons, p. A-2, ISBN 978-0-470-53649-0, retrieved January 23, 2013

"Release Notes", D3.js, retrieved August 22, 2012

<https://www.wappalyzer.com/technologies/d3>

Ian H. Witten, Eibe Frank, Mark A. Hall "Data Mining – Practical Machine Learning Tools and Techniques".

Meta S. Brown, "Data mining for dummies".

Trina Chiasson, Dyanna Gregory, "Data + Design".