

Από:

“Διεπαφές Μέσω Web για μη Ομοιογενή Σύνολα Δεδομένων σε Πραγματικό Χρόνο” Θανάσης Γιάκας – Πανεπιστήμιο Θεσσαλίας

2 Διεπαφές για οπτικοποίηση δεδομένων

2.1 Web Interfaces

Η σχεδίαση διεπαφών αποτελεί υποκατηγορία του επιστημονικού πεδίου που ονομάζεται επικοινωνία ανθρώπου – μηχανής (HCI Human-computer Interaction). Η επικοινωνία ανθρώπου μηχανής είναι η μελέτη και ο σχεδιασμός του τρόπου με τον οποίον άνθρωποι και μηχανές δουλεύουν μαζί έτσι ώστε οι ανάγκες του ανθρώπου να ικανοποιούνται με τον πιο αποδοτικό τρόπο, με βάση το τι θέλουν και τι περιμένουν οι άνθρωποι, τους φυσικούς περιορισμούς, τις ικανότητες που κατέχει ο άνθρωπος και το τι βρίσκει ελκυστικό και απολαυστικό.

Η διεπαφή χρήστη αποτελεί κομμάτι ενός λογισμικού με το οποίο κάποιος μπορεί να αντιληφθεί με κάποια από τις αισθήσεις του (συνήθως με την ακοή ή την όραση) όπως επίσης να κατανοήσει κάτι ή να το κατευθύνει.

Μία εξαιρετικά σχεδιασμένη διεπαφή, είναι αυτή που επιτρέπει τον χρήστη να εστιάσει στην πληροφορία καθεαυτή και όχι στους μηχανισμούς που παρουσιάζουν την πληροφορία. Με άλλα λόγια μια πολύ καλά σχεδιασμένη διεπαφή είναι αυτή που είναι σχεδόν "αόρατη" στον χρήστη, καθώς αυτός δεν μπορεί να την αντιληφθεί. Ο χρήστης άλλωστε επισκέπτεται έναν ιστότοπο ή χρησιμοποιεί μια διαδικτυακή υπηρεσία για να πετύχει έναν σκοπό π.χ. να κάνει μία αγορά ή να μοιραστεί κάτι με τους φίλους του και όχι να παίξει με την διεπαφή δοκιμάζοντας όλες τις δυνατότητες της.

Αν μια διεπαφή μπερδεύει τον χρήστη και δεν ικανοποιεί τις ανάγκες του, ο χρήστης δυσκολεύεται να πραγματοποιήσει τις εργασίες και οδηγείται σε επιπρόσθετα λάθη, κάτι που μπορεί να έχει ως αποτέλεσμα την οριστική εγκατάλειψη του συστήματος από την χρήστη και την καλλιέργεια αισθημάτων απογοήτευσης και άγχους.

Μια διεπαφή στο web (web interface) έχει ως στόχο την διαδραστικότητα ανάμεσα σε έναν χρήστη και σε ένα λογισμικό το οποίο τρέχει σε έναν Web server. Η διεπαφή χρήστη είναι ο φυλλομετρητής και η σελίδα ή η εφαρμογή η οποία έχει κατέβει και έχει αποδοθεί σε αυτόν.

Αξίζει εδώ να αναφέρουμε ότι ένα γραφικό σύστημα αποτελείται από αντικείμενα και πράξεις. Σε μια διεπαφή τα αντικείμενα είναι οι ανεξάρτητες μονάδες που μπορεί να δει ένας άνθρωπος στην οθόνη.

2.1.1 Εμπειρία χρήστη (User Experience)

Ο όρος εμπειρία χρήστη (User Experience, αρχικά UX) συμπεριλαμβάνει τα συναισθήματα κάποιου ανθρώπου που αναπτύχθηκαν κατά την χρήση ενός προϊόντος, ενός συστήματος ή μιας υπηρεσίας. Τονίζει τις πειραματικές, τις συναισθηματικές, τις με νόημα και αξία εκείνες όψεις της επικοινωνίας ανθρώπου και μηχανής και της κυριότητας προϊόντων. Επιπρόσθετα, συμπεριλαμβάνει τις αντιλήψεις του ανθρώπου για πρακτικά θέματα όπως χρησιμότητα, ευκολία χρήσης και αποδοτικότητα συστήματος. Η εμπειρία χρήστη είναι υποκειμενική μιας και έχει να κάνει με προσωπική αντίληψη και δυναμική αφού συνεχώς αλλάζει στο πέρασμα του χρόνου, εξαιτίας των συνεχών πάσης φύσεως αλλαγών του περιεχομένου αλλά και των νέων τεχνολογικών ανακαλύψεων.

Πολλοί παράγοντες μπορούν να επηρεάσουν την συνολική εμπειρία χρήστη. Για να συγκεκριμενοποιήσουμε κάπως αυτήν την πληθώρα, οι παράγοντες αυτοί έχουν ομαδοποιηθεί σε τρεις κύριες κατηγορίες:

- κατάσταση χρήστη και προηγούμενη εμπειρία,
- ιδιότητες συστήματος,
- εννοιολογικό πλαίσιο χρήσης (κατάσταση).

Η μελέτη της συμπεριφοράς του τυπικού χρήστη, του περιβάλλοντος και της μεταξύ τους αλληλεπίδρασης μας βοηθά στην σχεδίαση του συστήματος.

2.2 Σύνολα δεδομένων

Μιλήσαμε πριν για οπτικοποίηση δεδομένων, χωρίς όμως να δώσουμε τον ακριβή ορισμό του τι είναι τα δεδομένα. Ως δεδομένα, χαρακτηρίζεται ένα μη αξιολογημένο σύνολο διακριτών στοιχείων, μιας παρατήρησης-αναφοράς, μια συλλογή που αποτυπώνει "τιμές" επί αντικειμένων, προσώπων, γεγονότων κλπ. Στην ουσία τα δεδομένα είναι καταγραφές από παρατηρήσεις μας για τον έξω κόσμο, ή προϊόντα κάποιου συστήματος προσομοίωσης και μελέτης. Ενδεχομένως κάποιοι από τους τύπους δεδομένων που παρατίθενται εμπεριέχουν ήδη εμφανείς πληροφορίες, όχι όμως την πληροφορία στο επίπεδο που συνθέτει μια αξιολόγησή τους.

Μια συλλογή δεδομένων ονομάζεται σύνολο δεδομένων (data set ή dataset). Συνήθως ένα σύνολο δεδομένων αντιστοιχεί σε έναν πίνακα μιας βάσης δεδομένων ή ενός στατιστικού πίνακα δεδομένων, όπου κάθε στήλη αναπαριστά μια συγκεκριμένη μεταβλητή και κάθε γραμμή ένα συγκεκριμένο μέλος του συνόλου δεδομένων.

Πολλές φορές αναγκαζόμαστε να συγκεντρώσουμε δεδομένα από διάφορες πηγές. Όταν τα δεδομένα μας δεν έχουν την ίδια μορφή τότε έχουμε να κάνουμε με μη ομοιογενή σύνολα δεδομένων. Για παράδειγμα, όπως θα δούμε και στην συνέχεια, για να δημιουργήσουμε ένα παρατηρητήριο ερευνητικών δραστηριοτήτων για τα μέλη ΔΕΠ του Πανεπιστημίου Θεσσαλίας θα πρέπει να συγκεντρώσουμε δεδομένα από διάφορες μηχανές αναζήτησης επιστημονικών δημοσιεύσεων, ώστε τα αποτελέσματα που θα προκύψουν να είναι αξιόπιστα. Στην περίπτωση αυτή, το σύνολο δεδομένων που παίρνουμε μετά από κάποια ερώτηση μας σε μια Διεπαφή Προγραμματισμού Εφαρμογών (API) της μίας μηχανής θα είναι διαφορετικά δομημένο σε σχέση με αυτό κάποιας άλλης μηχανής. Σε αυτή την

περίπτωση έχουμε να κάνουμε με ανομοιογενή σύνολα δεδομένων τα οποία πρέπει να τα επεξεργαστούμε κατάλληλα ώστε να μπορέσουμε να τα ενοποιήσουμε.

2.2.1 Τύπος δομής των δεδομένων

Γραμμική δομή: Τα δεδομένα μπορούν να οργανωθούν σε πίνακες. Για παράδειγμα, πηγαίος κώδικας προγράμματος, αλφαβητική λίστα, χρονολογικά ταξινομημένα αντικείμενα, καταταγμένα αποτελέσματα αναζήτησης.

Χρονική δομή: Δομή που αποτελείται από δεδομένα που μεταβάλλονται στον χρόνο.

Χωρική δομή (Spatial): Τα δεδομένα σχετίζονται με κάτι που αφορά την γεωγραφική τους θέση.

Ιεραρχική δομή: Τα δεδομένα είναι ιεραρχικά όπως μια γενεαλογία και έχουν την μορφή δέντρου όπως για παράδειγμα, ένα σύστημα αρχείων ή ένα σύστημα κατηγοριοποίησης σε μια βιβλιοθήκη.

Δίκτυα: Τα δεδομένα αποτελούνται από οντότητες που σχετίζονται μεταξύ τους. Αποτελούν γενικές δομές γραφημάτων, όπως τα γραφήματα υπερμέσων με κόμβους και συνδέσμους, σημασιολογικά δίκτυα, πλέγματα (webs) και άλλα.

Πολυδιάστατη: Μια πολυδιάστατη δομή εμπεριέχει χαρακτηριστικά μεταδεδομένων όπως τύπος, μέγεθος, εκδότης, ημερομηνία τροποποίησης κτλ. Αντικείμενα με N χαρακτηριστικά γίνονται σημεία στον N -διάστατο χώρο.

Χώρος γνωρισμάτων (Feature spaces): Από την Ανάκτηση Πληροφορίας (Information Retrieval - IR), ένα διάνυσμα γνωρισμάτων αναπαριστά κάθε αντικείμενο της συλλογής μας. Ο Χώρος Γνωρισμάτων προβάλλεται πάνω σε δύο ή τρεις οπτικές διαστάσεις. Τα αντικείμενα συνήθως τοποθετούνται στον χώρο δυναμικά με κάποια τεχνική κατευθυνόμενη από δυνάμεις (βλ. ενότητα 2-4-5).

2.2.2 Δεδομένα πολλών διαστάσεων

Οι οπτικοποιήσεις των δεδομένων χαρακτηρίζονται από τον αριθμό των διαστάσεων που αυτές αναπαριστούν. Μία διάσταση σε μια οπτικοποίηση είναι ένας διακριτός τύπος πληροφορίας που κωδικοποιείται οπτικά στο γράφημα. Για παράδειγμα, ένα γράφημα μπορεί να αναπαριστά τις τιμές των μετοχών μιας εταιρίας για όλες τις μέρες μιας εβδομάδας. Το γράφημα αυτό έχει δύο διαστάσεις: την τιμή και τον χρόνο. Αν τώρα στο ίδιο γράφημα προσθέσουμε και άλλες εταιρίες τότε το γράφημα μας θα αποκτήσει μία τρίτη διάσταση. Ο αριθμός των διαστάσεων ενός γραφήματος μπορεί να περιγραφεί ως το επίπεδο της πολυπλοκότητας της οπτικοποίησης. Η πολυπλοκότητα μιας οπτικοποίησης, λοιπόν, αυξάνει όσο μεγαλώνει ο αριθμός των διαστάσεων και όχι όσο αυξάνει ο αριθμός των δεδομένων της κάθε διάστασης.

Κάθε διάσταση ενός συνόλου δεδομένων αντιστοιχεί σε κάποια συγκεκριμένη οπτική ιδιότητα. Στην ενότητα 2.5 θα εξετάσουμε λεπτομερώς τις οπτικές ιδιότητες.

2.3 Οπτική αναπαράσταση πληροφορίας

Τα τελευταία χρόνια οι μηχανές αυξάνουν με γρήγορους ρυθμούς την χωρητικότητά τους. Για τον λόγο αυτό, μπορούμε να δημιουργήσουμε (μέσω μετρήσεων, δειγματοληψίας, προσομοίωσης) και να αποθηκεύσουμε τεράστιες ποσότητες δεδομένων. Τα δεδομένα μας μπορούμε επίσης να τα αντλήσουμε από το διαδίκτυο στο οποίο παρατηρούμε καταιγισμό δεδομένων. Για να διαπιστώσουμε την αφθονία των δεδομένων στο διαδίκτυο αρκεί να σκεφτούμε ότι ένα μεγάλο μέρος της ζωής μας το περνάμε online. Σχεδόν οποιαδήποτε δραστηριότητά μας συμπυκνώνεται σε bits και στέλνεται σε όλον τον κόσμο με την ταχύτητα του φωτός. Για παράδειγμα στα κοινωνικά δίκτυα παράγεται από τις δραστηριότητες των χρηστών τους ένας τεράστιος όγκος δεδομένων ανά δευτερόλεπτο. Ας σκεφτούμε επιπλέον την περίπτωση των κινητών συσκευών οι οποίες μπορούν ανά πάσα στιγμή να συγκεντρώνουν καταγραφές που αφορούν την καθημερινότητα του χρήστη, όπως η τοποθεσία στην οποία βρίσκεται, η μουσική που ακούει, οι συνομιλίες του, η ταχύτητα με την οποία οδηγεί κ.ο.κ. Στην εικόνα 2-1 βλέπουμε κάποιες από τις τεράστιες ποσότητες δεδομένων που παράγουν οι άνθρωποι καθημερινά. Αυτή η "έκρηξη των δεδομένων" όμως, έχει πολλές φορές ως συνέπεια να μην μπορούμε να συσχετίσουμε τα δεδομένα μας με τον αρχικό μας στόχο για τον οποίο ξεκινήσαμε να τα συγκεντρώνουμε. Όλη αυτή η κατάσταση γεννάει το ερώτημα: "Πως μπορούμε να κατανοήσουμε τόσα πολλά δεδομένα ;".

Η οπτικοποίηση των δεδομένων επωφελείται από τις τεράστιες δυνατότητες του ανθρώπινου οπτικού συστήματος το οποίο λειτουργεί ως δίαυλος επικοινωνίας υψηλού εύρους ζώνης μεταξύ των αισθητήρων όρασης και του ανθρώπινου εγκεφάλου. Επιπρόσθετα χρησιμοποιεί το πλεονέκτημα του "λογισμικού" του εγκεφάλου μας να αναγνωρίζει πρότυπα και να μεταδίδει σχέσεις και νοήματα. Τέλος αξίζει να τονίσουμε ότι ενδεχομένως να χρησιμοποιηθούν και άλλα ιδιαίτερα επιθυμητά πλεονεκτήματα του ανθρώπινου νου που εν μέρει τουλάχιστον οφείλονται στην όραση όπως η πλαστικότητα της μνήμης και η δυνατότητα υποσυνείδητα **απώλειας μνήμης** (Christopher G. Healey, James. T. Enns, 2011). Μια οπτικοποίηση μπορεί να εμπνεύσει νέες ερωτήσεις και περαιτέρω ανακαλύψεις. Μπορεί επίσης να βοηθήσει στο να εντοπίσουμε υποπροβλήματα, τάσεις, ακραίες τιμές, συγκεκριμένα σημεία δεδομένων που βρίσκονται μέσα σε ένα μεγάλο σύνολο δεδομένων.

Η οπτικοποίηση δεδομένων δεν αφορά απλά τον μετασχηματισμό ενός πίνακα δεδομένων σε οπτικοποιημένη μορφή. Με την αποδοτική οπτικοποίηση των δεδομένων επιτυγχάνουμε την αναπαράσταση ακόμα και των πιο σύνθετων μεγάλων και πολύπλοκων συνόλων δεδομένων σε κατανοητή μορφή.

Ας σκεφτούμε ότι ένα σύνολο δεδομένων με M εγγραφές έχει $2m$ υποσύνολα δεδομένων. Κάθε ένα από αυτά τα σύνολα ενδέχεται να είναι αυτό που εμείς αναζητούμε. Με μία καλή απεικόνιση δεδομένων, οι δικές μας ικανότητες αναγνώρισης προτύπων, μπορούν όχι μόνο να ταξινομήσουν τα δεδομένα μέσα από αυτήν την συνδυαστική έκρηξη, αλλά μπορούν ακόμη να εξάγουν εννοήσεις (Insights) μέσα από τα οπτικά πρότυπα.

2.3.1 Αιτιολογική και διερευνητική οπτικοποίηση

Σύμφωνα με την λίστα του Tufte (Tufte, Edward R., 2001) όλες οι οπτικοποιήσεις δεδομένων πρέπει να παρουσιάζουν τα δεδομένα, να είναι ακριβείς χωρίς να αλλοιώνουν τα δεδομένα, να κάνουν μεγάλα

σύνολα δεδομένων συνεκτικά, να εξυπηρετούν έναν σαφή στόχο, να αποκαλύπτουν τα δεδομένα σε διαφορετικά επίπεδα (γενικά ή λεπτομερή) και να ενθαρρύνουν τον χρήστη.

Υπάρχουν δύο ειδών οπτικοποιήσεις δεδομένων η αιτιολογική και η διερευνητική:

Η αιτιολογική οπτικοποίηση των δεδομένων (explanatory data viz), η οποία χρησιμοποιείται για να μεταδώσει δεδομένα ή μια γνώμη από τον σχεδιαστή στον αναγνώστη. Συνήθως υπάρχει μία συγκεκριμένη ιστορία ή πληροφορία την οποία ο σχεδιαστής έχει σκοπό να μεταδώσει σε κάποιον άλλο. Για τον σκοπό αυτό ο σχεδιαστής θα πρέπει να αποφασίσει από πριν ποια πληροφορία είναι χρήσιμη και ποια είναι άσχετη. Όλη αυτή η διαδικασία συλλογής εστιασμένων δεδομένων στηρίζει την ιστορία που θέλει ο σχεδιαστής να περιγράψει. Η αιτιολογική οπτικοποίηση δεδομένων είναι κομμάτι της φάσης αναπαράστασης των δεδομένων

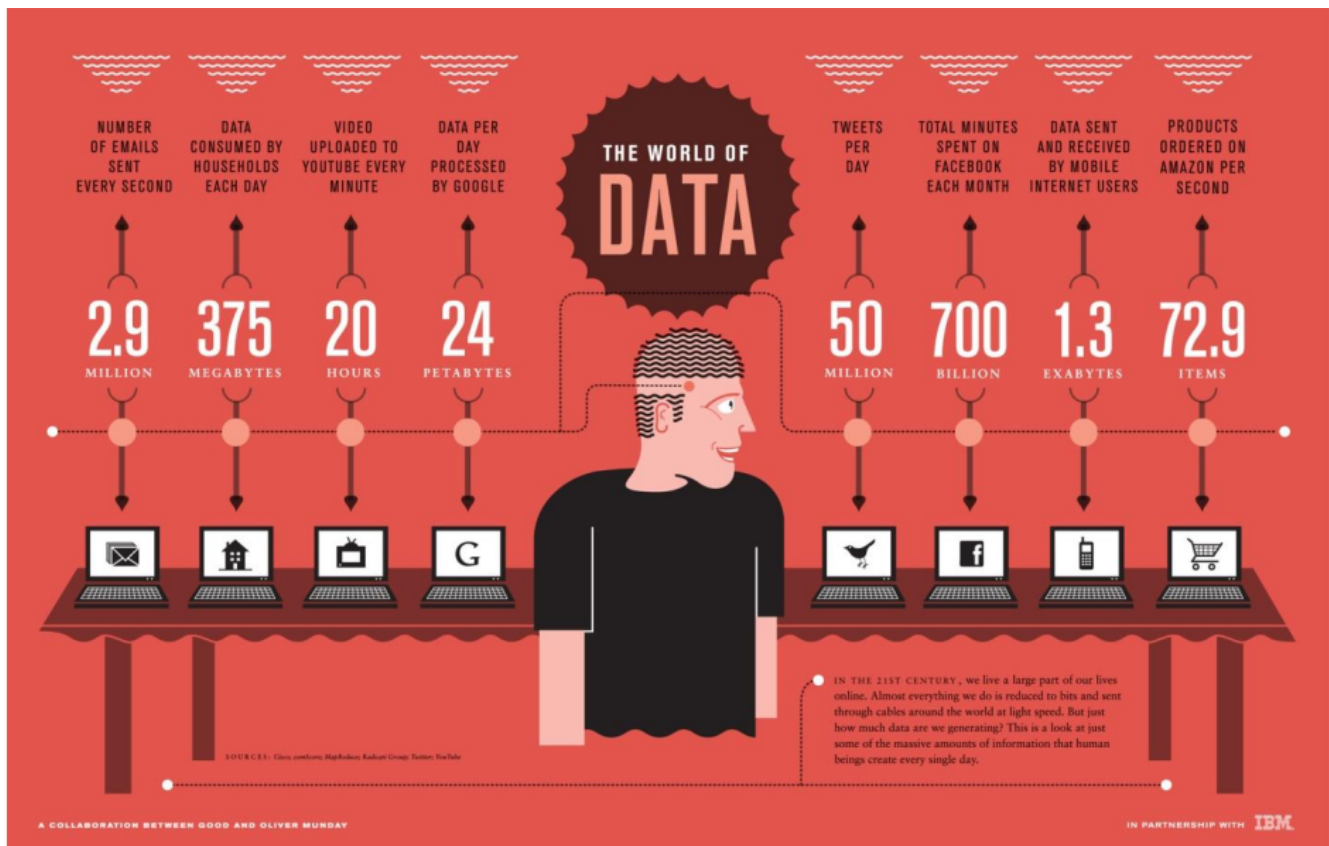
Η διερευνητική οπτικοποίηση (exploratory data viz) των δεδομένων χρησιμοποιείται από τον ίδιο τον σχεδιαστή για να ανακαλύψει μοτίβα, τάσεις ή υποπροβλήματα σε ένα σύνολο δεδομένων. Συνήθως δεν υπάρχει από πριν κάποια ήδη γνωστή ιστορία και ο σχεδιαστής δεν είναι σίγουρος για τα συμπεράσματα που θα προκύψουν από τα δεδομένα. Για το λόγο αυτό τα οπτικοποιεί προκειμένου να ανακαλύψει κάποια ενδιαφέρουσα καμπύλη, γραμμή, τάση ή κάποιες μη ομαλές ακραίες τιμές. Με τον τρόπο αυτό ο σχεδιαστής ανακαλύπτει την ιστορία που τα ίδια τα δεδομένα έχουν να πουν. Η διερευνητική οπτικοποίηση δεδομένων είναι κομμάτι της φάσης ανάλυσης των δεδομένων.

2.3.2 Γραφήματα πληροφοριών (Info Graphics)

Ο όρος γράφημα πληροφοριών (Infographic) αναφέρεται σε κάθε οπτική αναπαράσταση δεδομένων η οποία σχεδιάζεται χειροκίνητα για κάποια συγκεκριμένα δεδομένα. Ωφείλει να είναι εξαιρετικά καλαίσθητη με πλούσιο περιεχόμενο που προσελκύει το μάτι και κρατά το ενδιαφέρον και είναι σχετικά φτωχή ως προς την ποσότητα των δεδομένων.

Στην ουσία τα γραφήματα πληροφοριών είναι εικονογραφήματα όπου η αναπαράσταση των δεδομένων σχεδιάζεται χειρωνακτικά ή με κάποιο λογισμικό όπως το Adobe Illustrator.

Πολλές φορές γίνεται σύγχυση των όρων οπτικοποίηση δεδομένων και γραφήματος πληροφοριών. Από τον παραπάνω ορισμό συγκρίνοντας τα βλέπουμε ότι η οπτικοποίηση της πληροφορίας αν και αρχικά σχεδιάζεται από τον άνθρωπο, στην συνέχεια σχεδιάζεται αλγοριθμικά με κάποιο λογισμικό οπτικοποίησης πληροφορίας ή σχεδίασης γραφημάτων. Για τον λόγο αυτό μια οπτικοποίηση δεδομένων μπορεί να επαναχρησιμοποιηθεί ως έχει για περισσότερα, νέα ή διαφορετικά δεδομένα. Από την άλλη, οι οπτικοποιήσεις δεδομένων αν και δείχνουν μεγάλο όγκο δεδομένων, δεν είναι τόσο καλαίσθητες όσο τα γραφήματα πληροφοριών.



Εικόνα 2-1 Infographic που απεικονίζει την "έκρηξη των δεδομένων". Πηγή: <http://www.ibmbigdatahub.com/blogs>

2.4 Γραφήματα

2.4.1 Γραμμικά δεδομένα

Στα μαθηματικά, ένα γράφημα είναι μια αφηρημένη αναπαράσταση ενός συνόλου στοιχείων, όπου μερικά ζευγάρια στοιχείων συνδέονται μεταξύ τους με δεσμούς. Τα διασυνδεδεμένα στοιχεία αναπαριστώνται με μαθηματικές έννοιες και ονομάζονται κορυφές ενώ οι δεσμοί που συνδέουν τα ζευγάρια των κορυφών ονομάζονται ακμές. Συνήθως, ένας γράφος απεικονίζεται σε διαγραμματική μορφή ως ένα σύνολο κουκκίδων για τις κορυφές, ενωμένα μεταξύ τους με γραμμές ή καμπύλες για τις ακμές.

Τα γραφήματα είναι πολύ σημαντικά για την οπτική αναπαράσταση των δεδομένων, καθώς μοντελοποιούν πολλά σημαντικά προβλήματα.

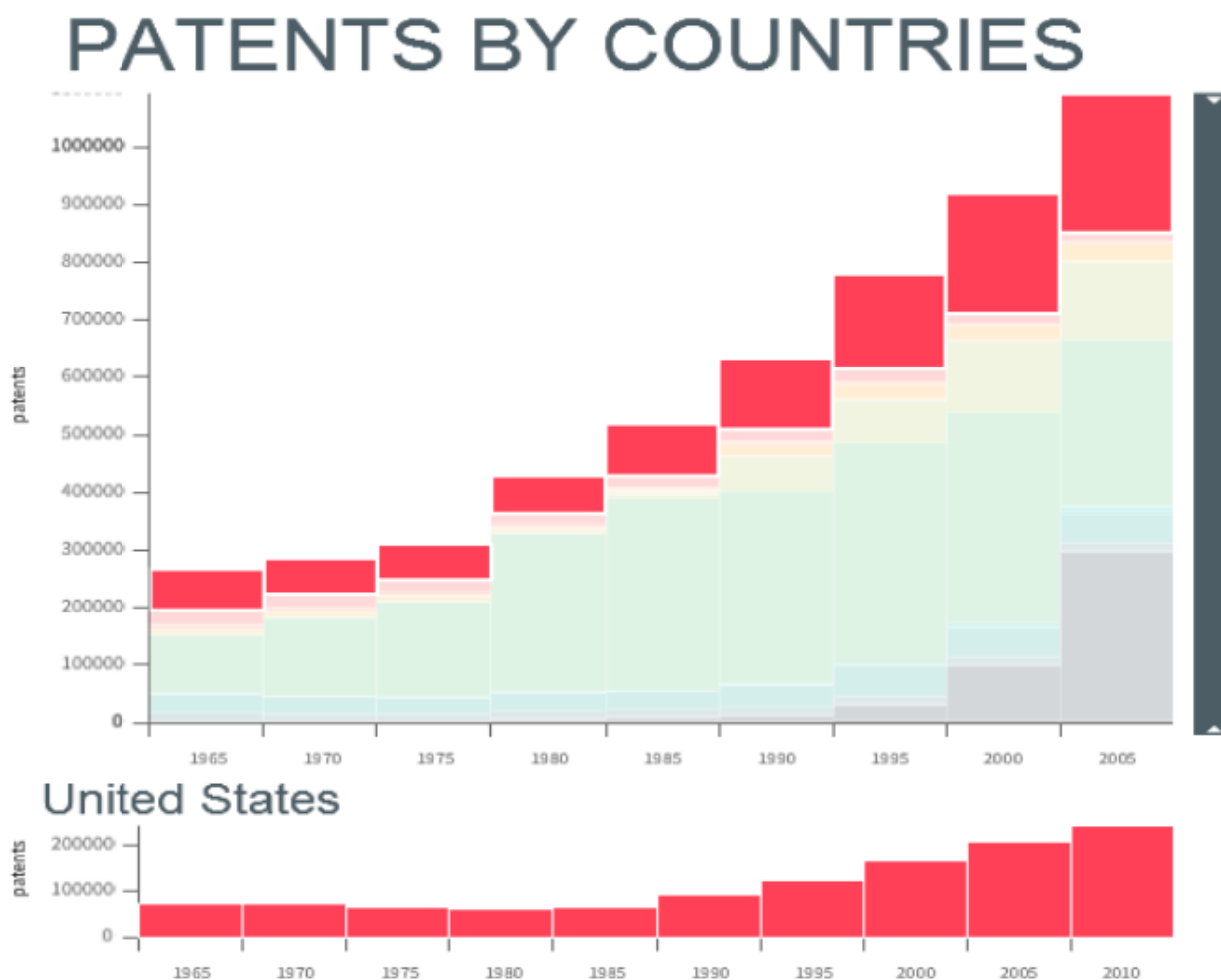
Τα γραμμικά δεδομένα μπορούν να οπτικοποιηθούν με χρήση των διαγραμμάτων που υπάρχουν και στο Excel: διάγραμμα με σημεία, διάγραμμα με γραμμές (line chart), διάγραμμα πίτας (pie chart), διάγραμμα με ράβδους (bar chart). Πρέπει να προσέξουμε ότι, όσον αφορά τα διαγράμματα πίτας, ενώ αυτά είναι αρκετά δημοφιλή, πολλοί είναι αυτοί που προτείνουν να αποφεύγεται η χρήση τους, αφού η αντίληψη του ανθρώπου για τις γωνίες σε δισδιάστατες περιοχές ως ποσοτικό μέτρο, είναι περιορισμένη. Στο Excel έχουμε επίσης την επιλογή να κάνουμε τα γραφήματα μας να φαίνονται τρισδιάστατα (π.χ. 3D bars, 3D pie). Κάτι τέτοιο όμως θα πρέπει να αποφεύγεται (εκτός αν πρέπει

αναγκαστικά με τον τρόπο αυτό να προσθέσουμε μία ακόμα διάσταση) γιατί οι συγκρίσεις στον τρισδιάστατο χώρο είναι πολύ πιο δύσκολες.

Στην συνέχεια θα παρουσιάσουμε κάποια άλλα είδη γραφημάτων που απεικονίζουν γραμμικά δεδομένα και χρησιμοποιούνται συχνά όταν έχουμε περισσότερες από δύο διαστάσεις.

Γράφημα σωρευμένων ράβδων (Stacked Bar Chart)

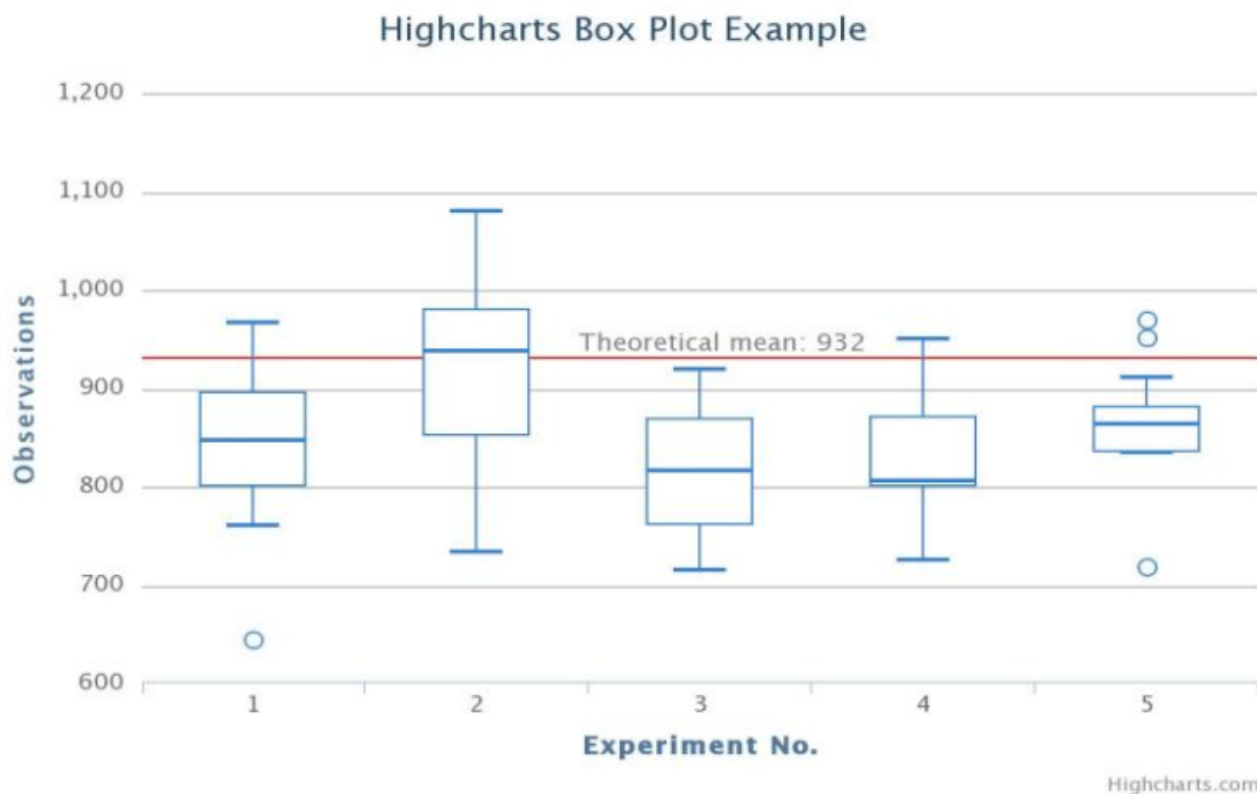
Αναμφίβολα, είναι πιο δύσκολες οι συγκρίσεις σε ένα γράφημα σωρευμένων ράβδων από ότι σε ένα γράφημα με ράβδους τοποθετημένες η μία δίπλα στην άλλη. Ωστόσο όταν έχουμε μια παραπάνω μεταβλητή και θέλουμε να οπτικοποιήσουμε τόσο ολόκληρο το μέτρο όσο και τα επί μέρους στοιχεία του, η χρήση του γραφήματος σωρευμένων ράβδων είναι η κατάλληλη επιλογή.



Εικόνα 2-2 Γράφημα σωρευμένων ράβδων για την απεικόνιση του αριθμού των πατεντών στην πορεία του χρόνου ανά χώρα. Η αναλογία των πατεντών που αντιστοιχούν σε κάθε χώρα φαίνεται από την διαφοροποίηση του χρώματος σε κάθε ράβδο.
Πηγή: [Quadrigram](#)

Γράφημα με κουτιά (Box Plot)

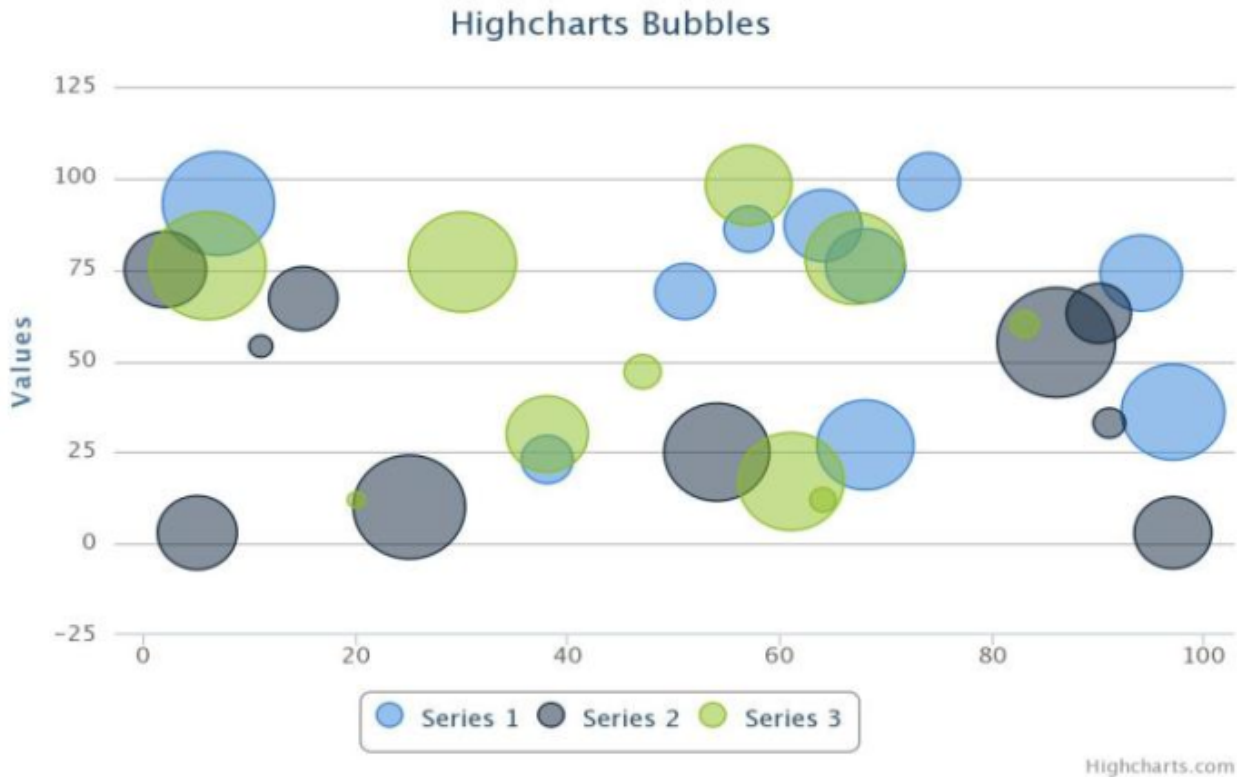
Το κουτί είναι παρόμοιο με την ράβδο με την μόνη διαφορά ότι και οι δύο άκρες του κουτιού χρησιμοποιούνται για να ορίσουν μια τιμή στον γράφο. Κάθε κουτί αναπαριστά μια κατανομή ενός συνόλου τιμών: το κάτω μέρος αναπαριστά την ελάχιστη τιμή, το πάνω την μέγιστη και το μήκος του, το πεδίο τιμών. Μπορούμε επιπλέον να χρησιμοποιήσουμε μία γραμμή για να δείξουμε το κέντρο της κατανομής που συνήθως είναι ο μέσος όρος. Χρησιμοποιούνται σε γραφήματα πολλαπλών κατανομών.



Εικόνα 2-3 Στο γράφημα με κουτιά μπορούμε να διακρίνουμε εύκολα το πεδίο τιμών αλλά και τον μέσο όρο που αντιστοιχεί σε κάθε τιμή της κατηγορικής μεταβλητής. Πηγή: [Highcharts.com](https://www.highcharts.com).

Διάγραμμα σημείων (Scatterplot)

Ιδανικά για αναζήτηση συσχετίσεων μεταξύ δύο ποσοτικών μεταβλητών ή για οπτικοποίηση δεδομένων που μεταβάλλονται κατά μήκος δύο διαστάσεων. Είναι δυνατό να αναπαραστήσουν τρεις ακόμα και τέσσερις διαστάσεις κωδικοποιώντας σημεία δεδομένων ως φουσαλίδες (Διάγραμμα Φουσαλίδας – Bubble Chart), πίτες ή σωρευμένες ράβδους και χρωματίζοντας τα ανάλογα αν υπάρχει κάποια κατηγορική μεταβλητή. Συνήθως, η εξαρτημένη μεταβλητή σχεδιάζεται στον κάθετο άξονα.



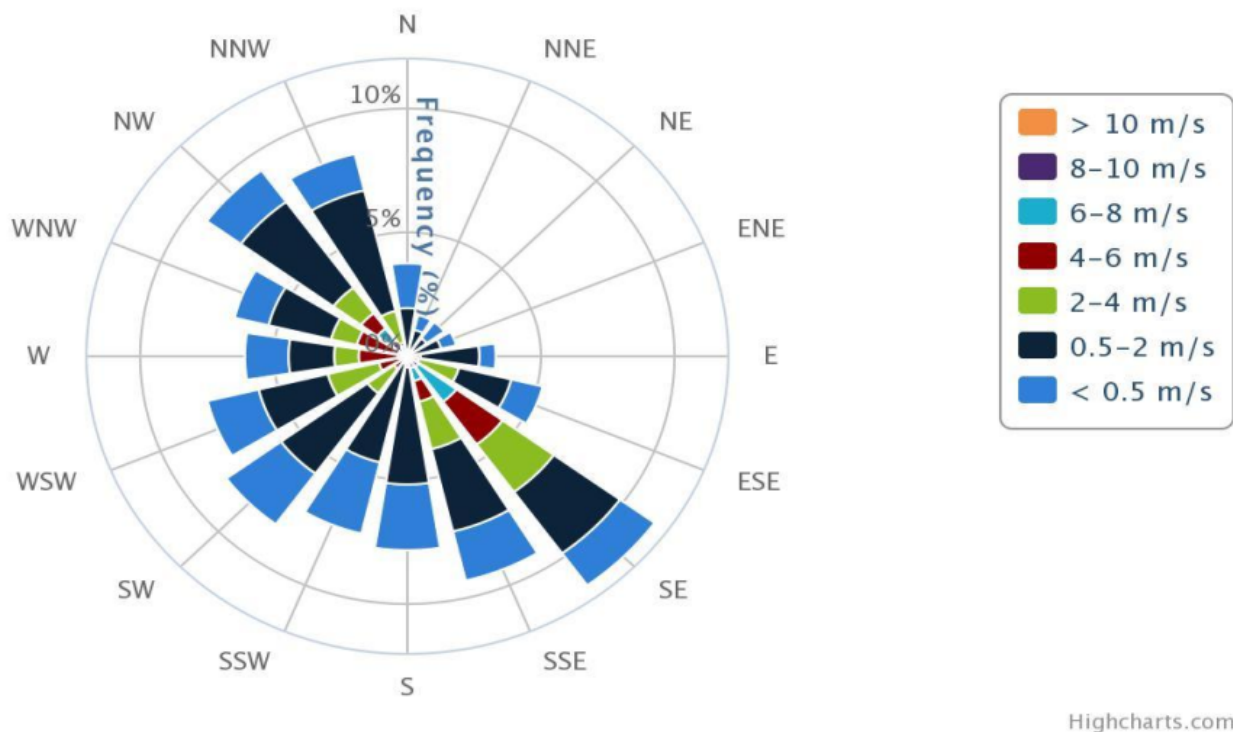
Εικόνα 2-4 Το διάγραμμα φουσαλίδας είναι ουσιαστικά ένα διάγραμμα σημείων με μία επιπλέον διάσταση που αναπαρίσταται από το εμβαδόν της κάθε φουσαλίδας. Πηγή: Highcharts.com.

Wind Rose Plot

Χρησιμοποιείται κατά κόρον από τους μετεωρολόγους για να αποδώσει μια λακωνική όψη του πως η ταχύτητα και η κατεύθυνση του αέρα κατανομονται σε μία συγκεκριμένη περιοχή. Χρησιμοποιούν ένα πολικό σύστημα συντεταγμένων σε μορφή πλέγματος και με τον τρόπο αυτό η συχνότητα των ανέμων μετά από μία μεγάλη χρονική περίοδο σχεδιάζεται πάνω στην διεύθυνση του ανέμου. Μπορούμε να κατηγοριοποιήσουμε τους ανέμους ανάλογα με το είδος τους (π.χ. την ταχύτητα τους) χρωματίζοντας ανάλογα προς την κάθε κατεύθυνση. Τελικά έχουμε τρεις διαστάσεις: κατεύθυνση, εύρος ταχύτητας, συχνότητα.

Wind rose for South Shore Met Station, Oregon

Source: or.water.usgs.gov



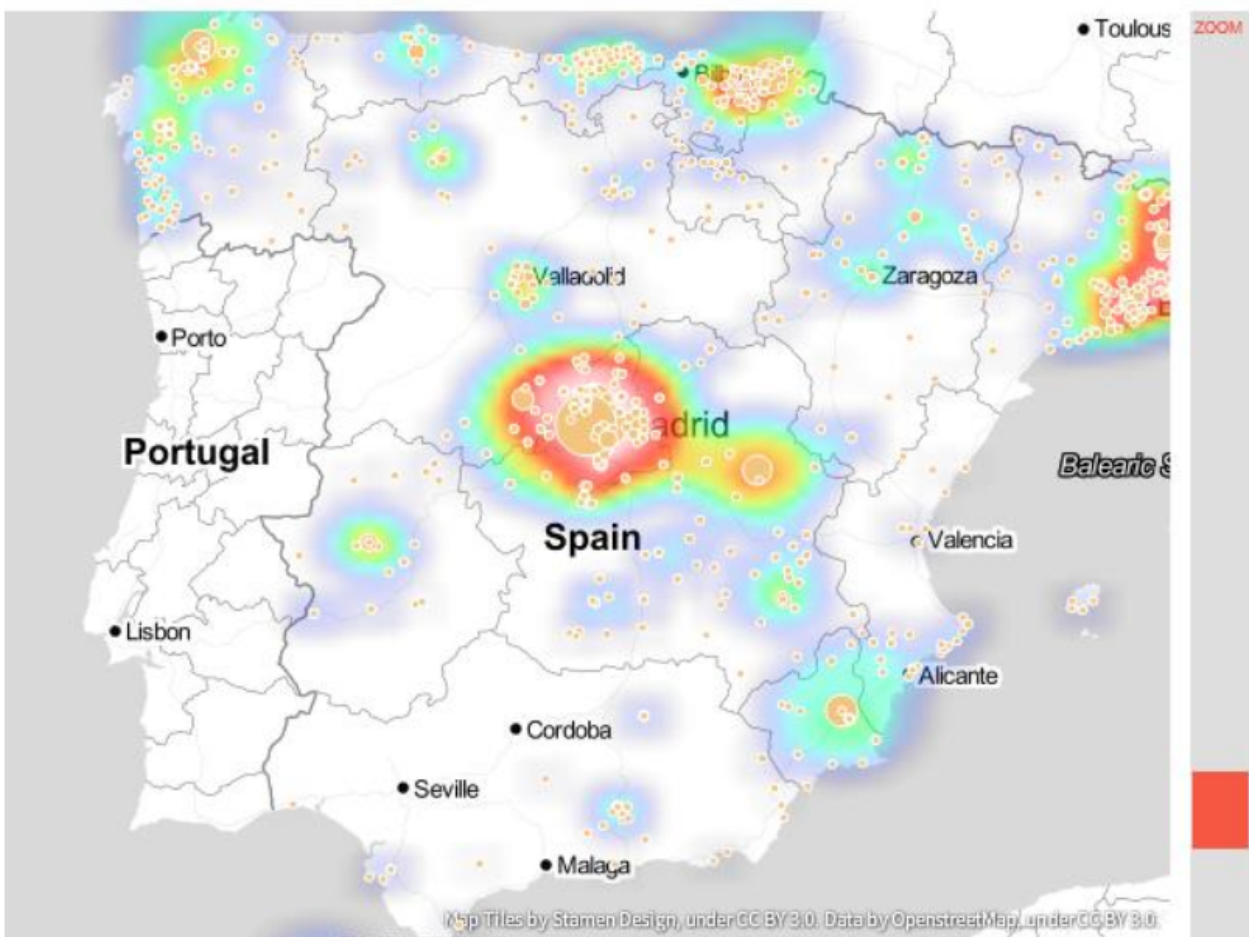
Εικόνα 2-5 Wind Rose για απεικόνιση των ανέμων στην πολιτεία του Oregon. Πηγή: Highcharts.com.

2.4.2 Χωρικά δεδομένα (Spatial Data)

Χάρτες Θερμότητας (Heat Maps)

Οι χάρτες θερμότητας είναι τρισδιάστατα γραφήματα περιοχής τα οποία χρησιμοποιούν το χρώμα και την φωτεινότητα για να υποδείξουν τιμές μεγάλων συνόλων δεδομένων. Οι δύο χωρικές διαστάσεις αναπαρίστανται από το γεωγραφικό μήκος και πλάτος, ενώ η τρίτη διάσταση χρησιμοποιείται για να δώσει την ένταση ενός σημείου δεδομένων σε σχετική σύγκριση με την μέγιστη τιμή του συνόλου δεδομένων. Η ένταση κωδικοποιείται με το χρώμα, συνήθως χρησιμοποιούμε θερμά χρώματα και τιμές μεγάλης έντασης και ψυχρά χρώματα για μικρής έντασης τιμές. Με τον τρόπο αυτό αναδεικνύουμε περιοχές με μεγάλες αλλαγές, τιμές εκτός εύρους ή άλλα ενδιαφέροντα χαρακτηριστικά. Καλή τακτική είναι η χρήση χρωμάτων με αυξημένη αντίθεση, ώστε να ξεχωρίσουμε τις ενδιαφέρουσες από τις κανονικές τιμές.

GEOGRAPHIC SALES DISTRIBUTION IN SPAIN



Εικόνα 2-6 Heat Map απεικόνισης κατανομής εκπτώσεων στην Ισπανία. Τα θερμά χρώματα συμβολίζουν υψηλές πυκνότητες. Πηγή: Quadrigram.

2.4.3 Δεδομένα δικτύου

Ένα δίκτυο μπορεί να αναπαρασταθεί ως γράφος μαθηματικά με τον πίνακα γειτνίασης. Ο πίνακας γειτνίασης ενός γράφου G με N κόμβους είναι ένας $N \times N$ πίνακας A , όπου τα στοιχεία του a_{ij} είναι ίσα με 1 αν ο κόμβος i είναι γειτονικός με τον κόμβο j , διαφορετικά ισούται με 0. Το σύνολο των ιδιοτιμών αυτού του πίνακα ονομάζεται φάσμα γραφήματος και χρησιμοποιείται στον προσδιορισμό του χώρου στον οποίον θα ενσωματωθεί ο γράφος, ως ένα σύνολο κορυφών και ακμών.

Για να σχεδιάσουμε έναν γράφο εισάγουμε τα δεδομένα μας, ορίζουμε τις οντότητες που θα αναπαρασταθούν ως κόμβοι και τις μεταξύ τους σχέσεις ως ακμές. Κατόπιν, εφαρμόζουμε κάποιον αλγόριθμο που θα αναδιατάξει την θέση των κόμβων ώστε οι ακμές να είναι περίπου ίδιου μήκους και να υπάρχουν όσο το δυνατόν λιγότερες διασταυρώσεις ακμών.

Τεχνικές κατευθυνόμενες με δυνάμεις (Force Directed Techniques)

Ο σχεδιασμός γράφων με αλγορίθμους κατευθυνόμενους με δυνάμεις είναι κλάσεις αλγορίθμων για σχεδίαση γράφων με καλαίσθητο και ευχάριστο τρόπο. Οι εν λόγω τεχνικές δανείζονται μια αναλογία που υπάρχει στην φύση, όπου οι κορυφές είναι σώματα με μάζα που έλκονται ή απωθούνται μεταξύ τους εξαιτίας μιας ελαστικότητας ή ενός ηλεκτρικού φορτίου. Οι δυνάμεις εφαρμόζονται σε ένα σύνολο κόμβων ή ακμών ανάλογα με την σχετική τους θέση. Στην συνέχεια, χρησιμοποιούμε αυτές τις δυνάμεις είτε για να προσομοιώσουμε την κίνηση των ακμών και των κορυφών είτε για να ελαχιστοποιήσουμε την ενέργεια τους.

Ο βελτιστοποιημένος γράφος είναι αυτός στον οποίο οι "φυσικές" αυτές δυνάμεις βρίσκονται σε ισορροπία. Σε αυτή τη κατάσταση οι δυνάμεις που εφαρμόζονται σε κάθε επανάληψη δεν αλλάζουν την σχετική θέση των κόμβων. Η ισότητα που αποτυπώνει αυτήν την αναλογία είναι η εξής:

$$Q(X|A, B) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \phi(d_{ij}(X)) - \sum_{i=1}^n \sum_{j=1}^n b_{ij} \psi(d_{ij}(X)),$$

όπου ο $n \times p$ πίνακας X περιέχει τις συντεταγμένες n κορυφών στον R_p και το $d_{ij}(X)$ υποδηλώνει τις αποστάσεις μεταξύ των σημείων με συντεταγμένες x_i και x_j . Τα βάρη a_{ij} αναφέρονται σε εκείνα του πίνακα γειτνίασης A του γράφου G (αν $a_{ij} = \{0,1\}$ μη βεβαρημένος γράφος), ενώ τα βάρη b_{ij} μπορεί να προέρχονται είτε από τον πίνακα γειτνίασης είτε από κάποιον εξωτερικό περιορισμό. Τελικά οι συναρτήσεις $\phi(\cdot)$ και $\psi(\cdot)$ είναι μετασχηματισμοί των οποίων ο ρόλος είναι να επιβάλουν μερικές καλαίσθητες θεωρήσεις στην διάταξη μας. Για παράδειγμα η κυρτή συνάρτηση ϕ θα ενισχύσει τις μεγάλες αποστάσεις αποδίδοντας ακόμα μεγαλύτερες και ως εκ τούτου θα ανιχνεύσει μοναδικά χαρακτηριστικά στα δεδομένα, ενώ ένας κοίλος μετασχηματισμός θα μετριάσει το φαινόμενο των απομονωμένων κορυφών.

Μέτρα κεντρικότητας κόμβου

Στην θεωρία των γράφων και στην ανάλυση δικτύων για να μετρήσουμε πόσο κεντρικός, δηλαδή πόσο σχετικά σημαντικός, είναι ένας κόμβος σε ένα δίκτυο μπορούμε να χρησιμοποιήσουμε κάποιο μέτρο κεντρικότητας για τους κόμβους του δικτύου μας. Για παράδειγμα θα χρησιμοποιήσουμε κάποιο μέτρο κεντρικότητας για να δούμε πόσο μπορεί κάποιο μέλος ενός κοινωνικού δικτύου να επηρεάσει τα άλλα μέλη ή κατά πόσο είναι early adopter, πόσο σημαντικό είναι ένα δωμάτιο μέσα σε ένα σπίτι, πόσο καλά χρησιμοποιείται ένας δρόμος κάποιου συγκοινωνιακού δικτύου. Στην συνέχεια θα εξετάσουμε κάποια από τα μέτρα κεντρικότητας τόσο από την μαθηματική όσο και από την κοινωνιολογική τους σκοπιά.

Απλή συνδεσιμότητα (Degree Centrality)

Η απλή συνδεσιμότητα ορίζεται ως ο αριθμός των συνδέσεων που ξεκινάνε από (out) ή καταλήγουν (in) στον κόμβο μας. Αλλιώς ονομάζεται βαθμός. Όταν το γράφημα μας είναι κατευθυνόμενο μπορούμε να εξετάσουμε χωριστά τον εσωτερικό βαθμό (In Degree) και τον εξωτερικό βαθμό κάποιου κόμβου. (Out Degree).

Μπορούμε να ερμηνεύσουμε την απλή συνδεσιμότητα ως το πόσο άμεσο ρίσκο έχει κάποιος κόμβος να "πάσει" αυτό που κυκλοφορεί στο δίκτυο μας (π.χ. Ένας ιός ή μια πληροφορία). Στην περίπτωση των κατευθυνόμενων γράφων όταν οι συνδέσεις σχετίζονται με κάτι θετικό όπως για παράδειγμα μια σύνδεση φιλίας ή συνεργασίας, ο αριθμός των εισερχόμενων συνδέσεων μπορεί να ερμηνευθεί ως μέτρο δημοτικότητας (Wikipedia) και δείχνει το πρεστίτζ που έχει κάποιος κόμβος δηλαδή το κατά πόσο η οντότητα που αναπαρίσταται από τον κόμβο προκαλεί θαυμασμό, επιρροή ή εμπιστοσύνη. Από την άλλη, ο αριθμός των εξερχόμενων συνδέσεων μπορεί να ερμηνευθεί ως μέτρο της κοινωνικότητας του κόμβου.

Κεντρικότητα εγγύτητας (Closeness centrality)

Για να βρούμε την απόσταση δύο κόμβων σε έναν γράφο βρίσκουμε το μήκος του συντομότερου μονοπατιού. Το πόσο απομακρυσμένος είναι ένας κόμβος το βρίσκουμε αθροίζοντας τις αποστάσεις του από όλους τους υπόλοιπους κόμβους. Το αντίστροφο της παραπάνω διαδικασίας μας δίνει την κεντρικότητα εγγύτητας. Έτσι όσο πιο κεντρικός θα είναι ένας κόμβος, τόσο χαμηλότερη θα είναι η συνολική του απόσταση από τους υπόλοιπους κόμβους. Μπορεί επίσης να θεωρηθεί ότι η κεντρικότητα εγγύτητας είναι ένας τρόπος μέτρησης του πόσο θα διαρκέσει η διάδοση της πληροφορίας από τον προς εξέταση κόμβο σε όλους τους άλλους, αν αυτή γίνει ακολουθιακά.

Μία ιδιότητα της κεντρικότητας εγγύτητας είναι ότι τείνει να αποδώσει υψηλό σκορ σε κόμβους που βρίσκονται κοντά στο κέντρο των τοπικών κοινωνιών (local clusters). Κοινωνιολογικά, μέλη μιας τοπικής κοινωνικής ομάδας με υψηλή κεντρικότητα εγγύτητας είναι αυτά τα μέλη που ασκούν μεγάλη επιρροή στα υπόλοιπα άτομα της κοινωνικής αυτής ομάδας. Συχνά, τα άτομα αυτά είναι δημόσιες φιγούρες σε ολόκληρο το δίκτυο, δέχονται τον σεβασμό της κοινωνικής ομάδας. Σε κάθε περίπτωση, "καταλαμβάνουν" τα συντομότερα μονοπάτια διάδοσης πληροφορίας.



Εικόνα 2-7 Οι κόμβοι που βρίσκονται στο κέντρο των τοπικών κοινωνιών έχουν υψηλότερο closeness centrality. Πηγή: activatenetworks.net.

Εκκεντρότητα Διαμεσότητας (Betweenness Centrality)

Η διαμεσότητα προτάθηκε από τον κοινωνιολόγο Linton Freeman (Freeman, Linton, 2004). Η διαμεσότητα ενός κόμβου ισούται με τον αριθμό των συντομότερων μονοπατιών από όλες τις κορυφές σε όλες τις υπόλοιπες που περνάνε από τον προς εξέταση κόμβο. Η διαμεσότητα είναι ακόμα πιο χρήσιμο μέτρο από την απλή συνδεσιμότητα (connectivity) όσον αφορά την μέτρηση τόσο του φορτίου όσο και της σπουδαιότητας του κόμβου. Η διαμεσότητα σε σχέση με την απλή συνδεσιμότητα είναι πιο γενικό και όχι τοπικό μέτρο.

Για να υπολογίσουμε την εκκεντρότητα διαμεσότητας πρέπει να ξεκινήσουμε βρίσκοντας όλα τα συντομότερα μονοπάτια μεταξύ όλων των δυνατών ζευγαριών κόμβων στο δίκτυο μας. Κατόπιν μετράμε πόσα από αυτά περνάνε από κάθε κόμβο. Οι αριθμοί αυτοί είναι η εκκεντρότητα διαμεσότητας του αντίστοιχου κόμβου.

Αν υποθέσουμε ότι κάθε κόμβος αναπαριστά ένα άτομο, το αποτέλεσμα του παραπάνω υπολογισμού είναι η εύρεση των ατόμων που είναι απαραίτητοι αγωγοί για την πληροφορία που διασχίζει διάφορα μέρη του δικτύου. Αυτοί συνήθως διαφέρουν από αυτούς με υψηλή κεντρικότητα εγγύτητας. Άτομα με υψηλή εκκεντρότητα διαμεσότητας δεν έχουν το κοντινότερο μέσο μονοπάτι σε οποιονδήποτε άλλον κόμβο, αλλά έχουν τον μεγαλύτερο αριθμό συντομότερων μονοπατιών που πρέπει να περάσουν από αυτούς.

Σε ένα κοινωνικό δίκτυο κόμβοι με υψηλή εκκεντρότητα διαμεσότητας εμφανίζονται συχνά στις τομές των πιο πυκνά συνδεδεμένων κοινωνιών. Όπως βλέπουμε στην εικόνα 2.8 όλοι είναι τοποθετημένοι για να πάρουν ρόλο διαμεσολαβητή κατά μήκος των συστάδων, με την έννοια ότι οι διαμεσολαβητές συνδέουν άτομα τα οποία διαφορετικά θα ήταν ασύνδετα μεταξύ τους, αλλά πλέον επωφελοούνται από την ανταλλαγή της πληροφορίας.

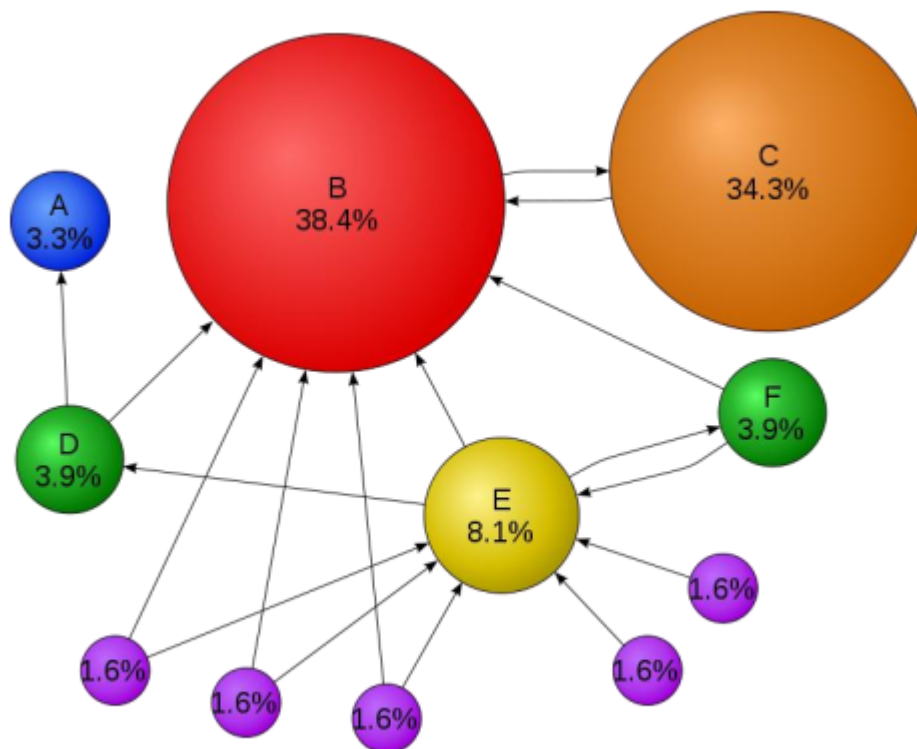
Στην πράξη άτομα με υψηλή εκκεντρότητα διαμεσότητας είναι συχνά κρίσιμα για συνεργασίες μεταξύ διαφορετικών τμημάτων και για να διατηρούν την διασπορά ενός προϊόντος σε όλο το δίκτυο. Εξαιτίας της θέσης τους είναι διαμεσολαβητές πληροφορίας και συνεργασίας. Η μόνη διαφορά τους με τους πραγματικούς μεσίτες είναι ότι οι δεύτεροι έχουν συνήθως ένα δημόσιο προφίλ σαν μέρος της εργασίας τους, ενώ οι πρώτοι συχνά παραβλέπονται. Αυτό συμβαίνει γιατί δεν είναι κεντρικοί σε καμία κοινωνική ομάδα, αντιθέτως κινούνται στην περιφέρεια πολλών κοινωνικών ομάδων σε καθεμία από τις οποίες προκαλούν εμπιστοσύνη και θαυμασμό.



Εικόνα 2-8 Οι κόμβοι που συμπεριφέρονται ως γέφυρες μεταξύ των συστάδων του δικτύου έχουν υψηλότερη διαμεσότητα.
Πηγή: activatenetworks.net.

Κεντρικότητα Page Rank

Ο αλγόριθμος page rank, (Wikipedia) ο οποίος χρησιμοποιήθηκε από την μηχανή αναζήτησης της Google, αποδίδει ένα βάρος σε κάθε στοιχείο ενός υπερσυνδεδεμένου (hyperlinked) συνόλου εγγράφων, όπως για παράδειγμα το World Wide Web, με σκοπό την μέτρηση της σχετικής σημαντικότητας μέσα στο σύνολο. Ο αλγόριθμος μπορεί να εφαρμοσθεί σε κάθε γράφο, για μέτρηση της φασματικής κεντρικότητας αλλά χρησιμοποιείται ευρέως από τις μηχανές αναζήτησης για την κατάταξη των ιστοσελίδων στα αποτελέσματα αναζήτησης για κάθε δεδομένη επερώτηση. Κάθε υπερσύνδεσμος σε μια σελίδα μετριέται ως ψήφος υποστήριξης. Το page rank μιας σελίδας ορίζεται αναδρομικά και εξαρτάται από τον αριθμό των σελίδων που αναφέρονται μέσω υπερσυνδέσμου σε αυτή αλλά και από την τιμή του page rank των σελίδων αυτών.



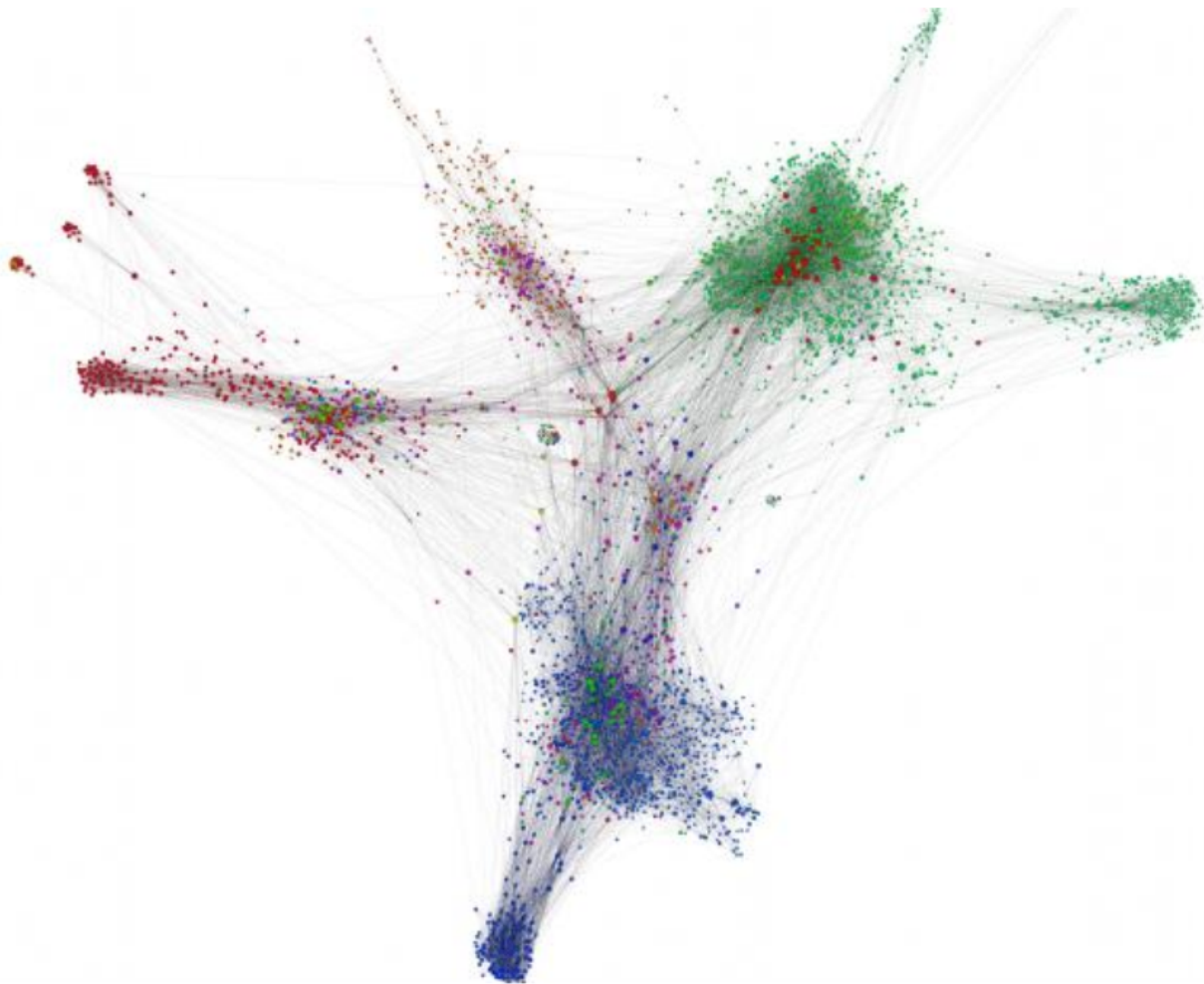
Εικόνα 2-9 Παρατηρούμε ότι η σελίδα C έχει υψηλότερο pagerank από την σελίδα E, παρόλο που οι υπερσύνδεσμοι προς την E είναι λιγότεροι από αυτούς προς την C. Αυτό συμβαίνει γιατί ο μοναδικός σύνδεσμος προς την C έρχεται από τον κόμβο B ο οποίος έχει μεγάλο pagerank. Πηγή: Wikipedia.

Διάρθρωση (Modularity)

Η διάρθρωση μετρά την δομή ενός γράφου. Πιο συγκεκριμένα, υπολογίζει την δύναμη της κάθε υποδιαίρεσης ενός γράφου σε κοινωνίες ή συστάδες (clusters). Δίκτυα με υψηλό modularity έχουν πυκνές συνδέσεις μεταξύ των κόμβων που ανήκουν σε μια συστάδα και αραιές συνδέσεις μεταξύ των κόμβων που ανήκουν σε διαφορετικές συστάδες.

Ο αλγόριθμος υπολογισμού του modularity χρησιμοποιείται για να ανακαλύψουμε την κοινωνική δομή των δικτύων. Μειονέκτημα του αλγορίθμου είναι ότι δεν μπορεί να εντοπίσει μικρές ομάδες, λόγω

περιορισμού στην ανάλυση του. Βιολογικά δίκτυα όπως το ανθρώπινο μυαλό παρουσιάζουν πολύ υψηλό modularity.



Εικόνα 2-10 Σε έναν γράφο οι οντότητες συμβολίζονται ως κόμβοι και οι μεταξύ τους σχέσεις ως σύνδεσμοι. Στην εικόνα μας, οι κόμβοι χρωματίζονται με βάση την ομάδα στην οποία ανήκουν. Πηγή: Gephi.org.

2.4.4 Ιεραρχικά δεδομένα

Ένα δέντρο στη Θεωρία γράφων, είναι ένας μη κατευθυνόμενος Γράφος, στον οποίο οποιοσδήποτε δύο κορυφές συνδέονται με ένα και μόνο απλό μονοπάτι. Με άλλα λόγια κάθε συνεκτικός γράφος χωρίς κύκλους είναι ένα δέντρο.

Ιεραρχικά Δένδρα

Κατευθυνόμενα δένδρα με ένα σύνολο κόμβων-φύλλων (κόμβοι μοναδιαίου βαθμού) που αναπαριστούν ένα σύνολο αντικειμένων και ένα σύνολο κόμβων-γονέων που αναπαριστούν σχέσεις επί των αντικειμένων. Σε ένα ιεραρχικό δένδρο κάθε κόμβος, εκτός της ρίζας του δέντρου, έχει ακριβώς έναν γονέα.

Υπερβολικά Δένδρα (Hyper Trees)

Εμπνευσμένα από την υπερβολική γεωμετρία, απεικονίζουν ιεραρχικά δεδομένα, ενώ ο αριθμός των κόμβων σε κάθε επίπεδο μπορεί να αυξάνει εκθετικά. Χρησιμοποιούν τον Υπερβολικό Χώρο, ο οποίος ουσιαστικά έχει περισσότερο μέρος από τον Ευκλείδειο. Για παράδειγμα, μια γραμμική αύξηση μιας ακτίνας ενός κύκλου στον Ευκλείδειο χώρο αυξάνει την περιφέρεια γραμμικά, ενώ μια ανάλογη γραμμική αύξηση στον Υπερβολικό χώρο θα την αύξανε εκθετικά.



Εικόνα 2-11 Circle Limit IV. Ο M.C. Escher γύρω στο 1956 ασχολήθηκε με τις Υπερβολικές Ψηφιοθετήσεις, που είναι κανονικά πλακίδια στο Υπερβολικό επίπεδο. Για να παρουσιάσει την εργασία του δημιούργησε το 1959 τέσσερις ξυλογραφίες με την ονομασία Circle Limit. Πηγή: Wikipedia.

Σε ένα υπερβολικό δένδρο οι κόμβοι στους οποίους εστιάζουμε τοποθετούνται στο κέντρο και καταλαμβάνουν περισσότερο χώρο, ενώ οι κόμβοι εκτός εστίασης βρίσκονται συμπιεσμένοι κοντά στα όρια του χώρου. Αν εστιάσουμε σε κάποιον άλλον κόμβο, τον φέρνουμε με την σειρά του αυτόν στο κέντρο και γύρω τα παιδιά του, ενώ οι κόμβοι που δεν μας ενδιαφέρουν συμπιέζονται καταλαμβάνοντας πολύ λίγο χώρο.

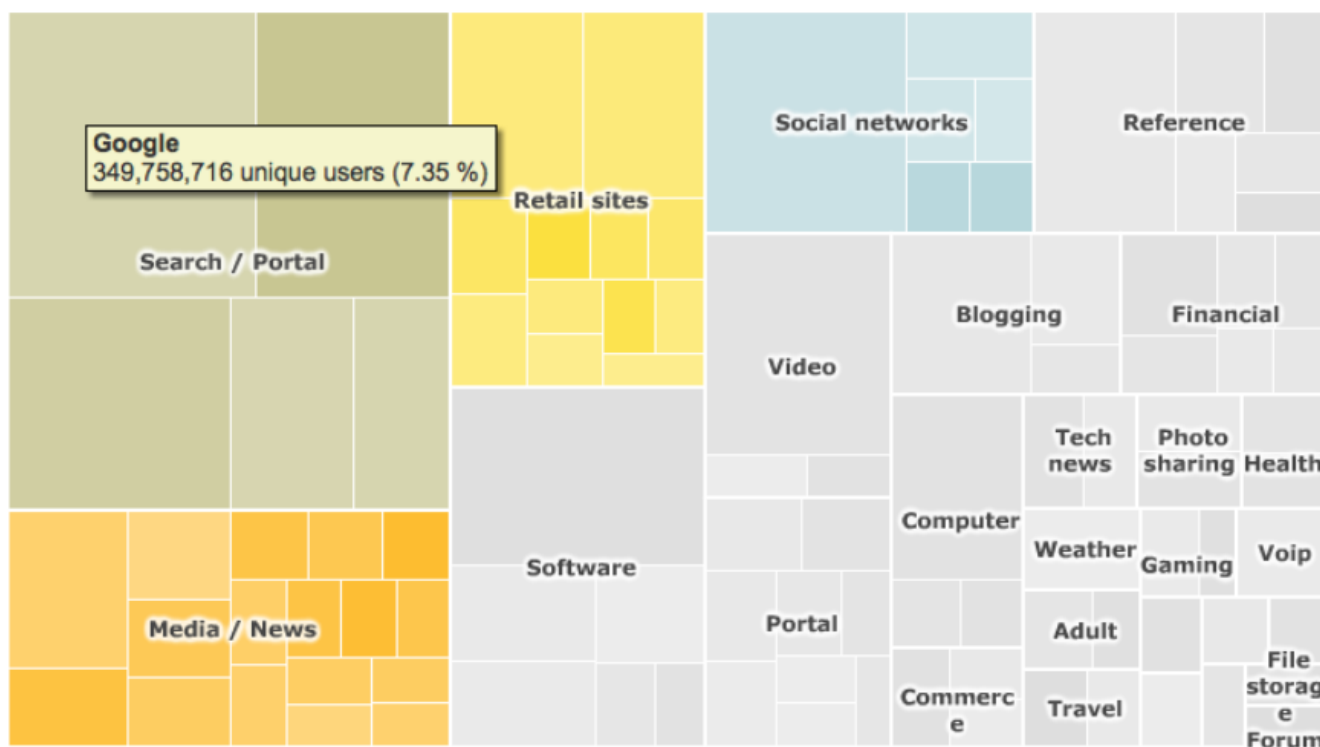
Τα υπερβολικά δένδρα κάνουν εύκολη την πλοήγηση σε μεγάλα δένδρα ακόμα και με εκατοντάδες οι χιλιάδες κόμβους. Με την περιστροφή και το "τράβηγμα" της απεικόνισης στον Καρτεσιανό χώρο μπορούμε εύκολα να αναζητήσουμε και να επιλέξουμε τους κόμβους που μας ενδιαφέρουν.

Δενδρικοί χάρτες (Treemaps)

Οι δενδρικοί χάρτες (treemaps) επινοήθηκαν από τον Ben Shneiderman (Ben Shneiderman-University of Maryland, 2009) σαν μια λύση στο πρόβλημα της αναπαράστασης ανάλογων τιμών και ιεραρχικών σχέσεων την ίδια στιγμή. Είναι ιδανικά στο αναπαριστούν μια πληθώρα ιεραρχικά φωλιασμένων τιμών δεδομένων.

Στους δενδρικούς χάρτες κάθε τερματικός κόμβος αναπαρίσταται ως ένα τετράγωνο. Η κύρια ιδέα είναι να διαμερίσουμε τον διαθέσιμο χώρο αναδρομικά με τον ίδιο τρόπο που ένα δένδρο διαμερίζει τα δεδομένα. Αυτό σημαίνει ότι οι δενδρικοί χάρτες είναι αναπαραστάσεις μοντέλου οδηγούμενες από τα δεδομένα.

Ο τετράγωνος χώρος του δενδρικού χάρτη ανταποκρίνεται σε ολόκληρο το σύνολο δεδομένων. Στο πρώτο βήμα ο χώρος διαμερίζεται οριζόντια σύμφωνα με τις αναλογίες των περιπτώσεων που περνάνε σε κάθε κόμβο παιδί. Στο επόμενο βήμα κάθε διαμέριση χωρίζεται κάθετα ανταποκρινόμενη στις αναλογίες των παιδιών της. Η διαδικασία αυτή επαναλαμβάνεται αναδρομικά με εναλλαγή κάθε φορά της κατεύθυνσης διαμέρισης από οριζόντια σε κάθετη και το αντίστροφο, έως ότου φτάσουμε στους τερματικούς κόμβους.



Εικόνα 2-12 Οπτικοποίηση του BBC με την κατανομή των 100 site με την μεγαλύτερη επισκεψιμότητα ανά θεματική κατηγορία (με βάση το χρώμα). Σε ένα διαδραστικό περιβάλλον ο χρήστης έχει την επιλογή να τοποθετήσει τον κέρσορα του ποντικιού πάνω από ένα τετράγωνο για να δει το όνομα του ιστότοπου και τον αριθμό των επισκέψεων. Πηγή: BBC.

Στο τελικό αποτέλεσμα κάθε τετράγωνο ανταποκρίνεται σε έναν τερματικό κόμβο. Η περιοχή που καλύπτει κάθε τετράγωνο είναι ανάλογη με τον αριθμό των περιπτώσεων που αντιστοιχούν σε αυτόν τον κόμβο.

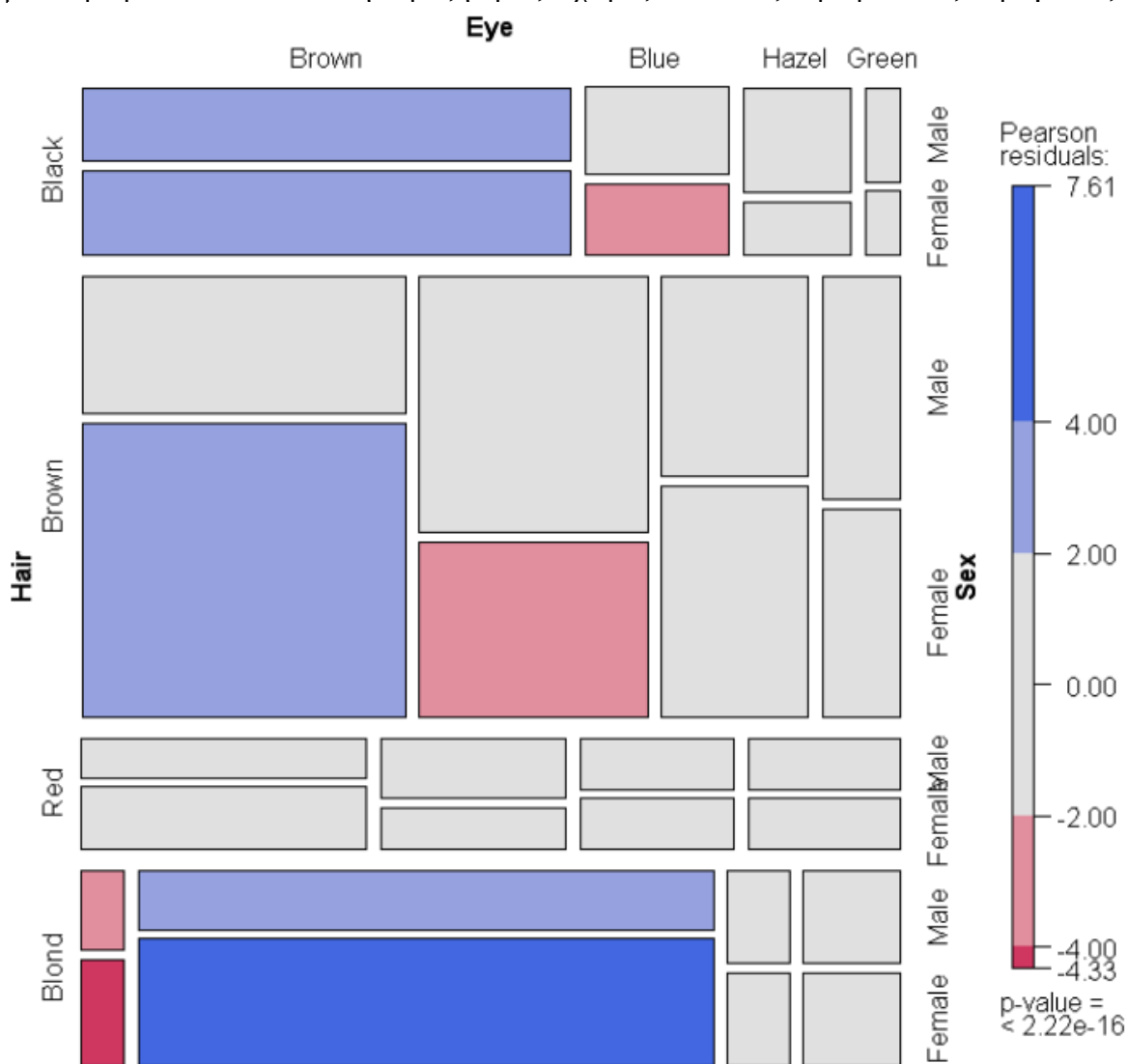
Κατά την σχεδίαση μπορούμε να προσαρμόσουμε τα κενά μεταξύ των διαμερίσεων έτσι ώστε να αναδείξουμε το βάθος στο οποίο συνέβη η διαμέριση, δίνοντας μεγαλύτερα κενά σε όσες διαχωρίσεις έγιναν κοντινότερα της ρίζας του δέντρου.

2.4.5 Δεδομένα πολλών διαστάσεων

Μωσαϊκό διάγραμμα (Mosaic Plot)

Το μωσαϊκό διάγραμμα στην ουσία είναι ένας δενδρικός χάρτης ενός δένδρου αποσύνθεσης, δηλαδή ενός δένδρου στο οποίο οι διαχωρίσεις με το ίδιο βάθος χρησιμοποιούν την ίδια κατηγορική μεταβλητή διαχώρισης και έχουν τόσα παιδιά όσα και οι κατηγορίες των δεδομένων.

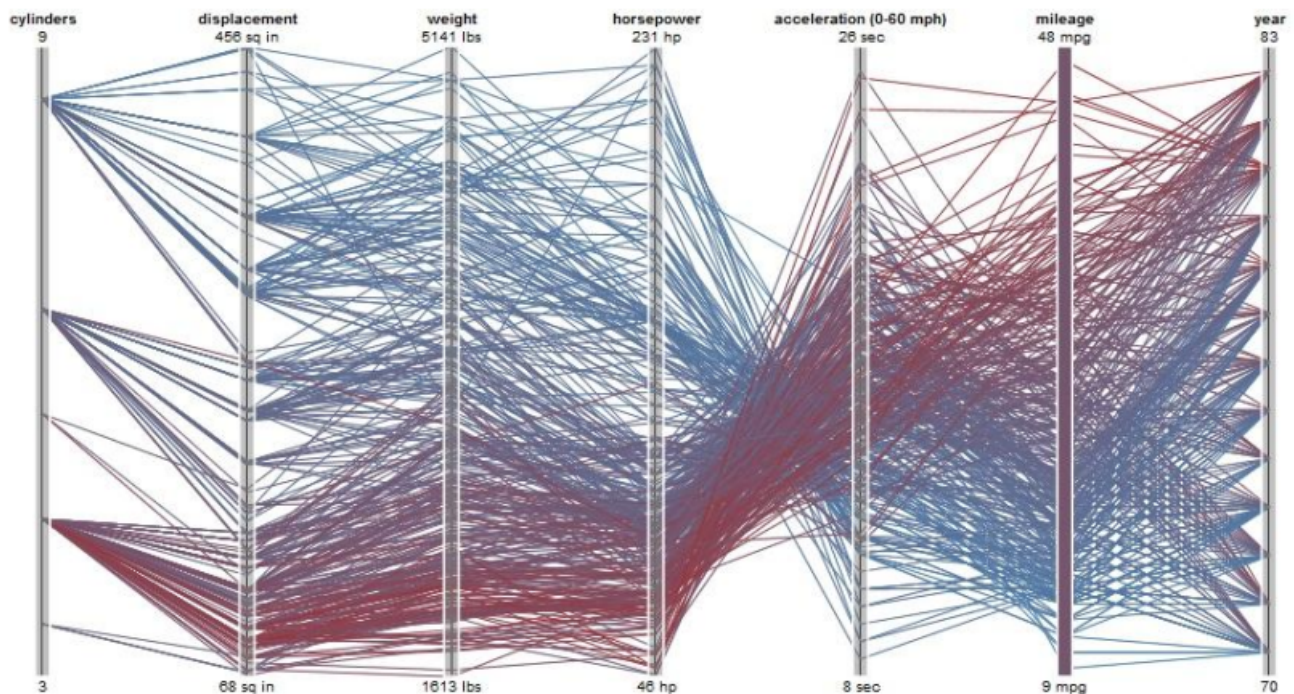
Το μωσαϊκό διάγραμμα αποτελείται από ομάδες τετράγωνων πλακιδίων. Κάθε πλακίδιο αντιστοιχεί σε ένα κελί κάποιου πίνακα συνάφειας (contingency table). Η περιοχή του είναι ανάλογη του μεγέθους του κελιού. Το σχήμα και η τοποθεσία του καθορίζονται κατά την διαδικασία της κατασκευής του. Τα μωσαϊκά διαγράμματα δεν περιορίζονται στις δύο διαστάσεις μπορούν να επεκταθούν σε έναν αυθαίρετο αριθμό διαστάσεων. Στην πράξη όμως ο χώρος είναι ένας περιοριστικός παράγοντας.



Εικόνα 2-13 Στο Mosaic plot της εικόνας τα κουτιά διαχωρίζονται ανάλογα με τις τιμές των τεσσάρων μεταβλητών: hair, sex, eye color, pearson residuals. Πηγή: <http://statmath.wu.ac.at/projects/vcd/>.

Παράλληλες συντεταγμένες (Paraller coordinates)

Κάθε παράλληλη συντεταγμένη αναπαριστά μία μεταβλητή. Παίρνοντας την τομή κάθε γραμμής με κάθε παράλληλη συντεταγμένη βλέπουμε τις τιμές της υπό εξερεύνηση οντότητας για κάθε μεταβλητή. Με τις παράλληλες συντεταγμένες η αναζήτηση για σχέσεις πολλαπλών μεταβλητών σε σύνολα δεδομένων πολλαπλών διαστάσεων μετασχηματίζεται σε ένα δυσδιάστατο πρόβλημα αναγνώρισης προτύπων.



Εικόνα 2-14 Κάθε παράλληλη συντεταγμένη αναπαριστά ένα μέτρο επίδοσης κάποιου αυτοκινήτου, ενώ κάθε γραμμή ένα συγκεκριμένο αυτοκίνητο. Σε ένα διαδραστικό περιβάλλον μπορούμε να φιλτράρουμε το πλήθος γραμμών προς εμφάνιση, με βάση το ποια αυτοκίνητα μας ενδιαφέρουν, για να κάνουμε ευκολότερα τις συγκρίσεις.

Πηγή: <http://mbostock.github.io/protovis/ex/cars-full.html>.

2.4.6 Επιλογές επιστημονικών σχεδιασμών στην οπτικοποίηση δεδομένων

Ο τύπος των μεταβλητών προς σχεδίαση επηρεάζει σημαντικά την επιλογή του γραφήματος. Για παράδειγμα για συνεχείς μεταβλητές προτιμότερα είναι τα γραφήματα με γραμμές, τα ιστογράμματα και τα γραφήματα με κουτιά, ενώ για κατηγορικές μεταβλητές τα γραφήματα με ράβδους. Το αν τα δεδομένα μας θα είναι συναθροισμένα ή μετασχηματισμένα εξαρτάται από την κατανομή τους και από το ποιοι είναι οι στόχοι της οπτικοποίησης.

Ο σχεδιασμός γραφημάτων πολλών μεταβλητών είναι πιο πολύπλοκος. Οι κύριες αποφάσεις που πρέπει να παρθούν είναι η μορφή της απεικόνισης και η επιλογή και η διάταξη των μεταβλητών.

Γενικά μια εξαρτημένη μεταβλητή πρέπει να σχεδιάζεται τελευταία. Το πρώτο βήμα είναι η επιλογή της κατάλληλης οπτικοποίησης. Η επόμενη πρόκληση είναι η σχεδίαση των οπτικών στοιχείων της απεικόνισης με τρόπο που το μήνυμα μας θα δηλώνεται με σαφήνεια, χωρίς να αποσπά την προσοχή σε ασήμαντα στοιχεία, αλλά αναδεικνύοντας ότι είναι το πιο σημαντικό.

Σύμφωνα με τον Edward Tufte, οι πίνακες και τα γραφήματα πρέπει να σχεδιάζονται με την χρήση δύο τύπων μελάνης: μελάνι δεδομένων (data ink) και μελάνι μη δεδομένων (non-data ink). Η αναλογία του μελανιού που χρησιμοποιείται για την απεικόνιση των δεδομένων προς το συνολικό μελάνι που εμφανίζεται στο περιβάλλον πρέπει να είναι υψηλή. Με άλλα λόγια, το μελάνι που χρησιμοποιείται για οτιδήποτε εκτός από δεδομένα θα πρέπει να μειωθεί στο ελάχιστο.

Ο Edward Tufte αναφέρθηκε επίσης και στην πυκνότητα των δεδομένων. Οι αριθμοί που εμφανίζονται σε ένα γράφημα μπορούν να οργανωθούν σε έναν πίνακα. Η πυκνότητα των δεδομένων ορίζεται ως η αναλογία του αριθμού των στοιχείων του παραπάνω πίνακα, ως προς τον χώρο που πιάνει το γράφημα. Στόχος μας είναι να μεγιστοποιήσουμε την πυκνότητα των δεδομένων, ώστε να δώσουμε ένα εννοιολογικό πλαίσιο και αξιοπιστία στα στατιστικά στοιχεία. Όταν ο χρήστης βλέπει ένα γράφημα με μικρή ποσότητα πληροφορίας μπορεί να υποπτευθεί ότι κάτι είναι κρυμμένο ή έχει παραληφθεί. Επίσης όταν έχουμε μεγάλη πυκνότητα δεδομένων μπορούμε να συγκρίνουμε διάφορα μέρη των δεδομένων παρουσιάζοντας μεγάλη ποσότητα πληροφορίας που μπορεί να γίνει αντιληπτή με μια ματιά. Όσο βέβαια η ποσότητα της πληροφορίας μεγαλώνει, τα μέτρα πρέπει να μικρύνουν, π.χ. να μικρύνουν οι τελείες σε ένα διάγραμμα σημείων ή να λεπτύνουν οι γραμμές σε ένα γράφημα χρόνου. Επίσης πρέπει να μεγαλώσει η αναλογία data-ink/no data ink. Εκτός όμως από το να αυξήσουμε την πληροφορία σε ένα γράφημα μπορούμε επίσης να μειώσουμε την περιοχή του γραφήματος (χωρίς να χαθεί πληροφορία ή να αλλοιωθεί η αναγνωσιμότητα) ώστε να αυξηθεί η πυκνότητα των δεδομένων.

2.4.7 Κλίμακες

Μια κλίμακα είναι ένας τύπος συνάρτησης που αντιστοιχίζει σύνολα μεταβλητών σε διαστάσεις. Η επιλογή μιας κατατοπιστικής διάταξης έχει να κάνει με τον ορισμό της κλίμακας του άξονα κάποιας κατηγορικής μεταβλητής. Αυτό εξαρτάται από το τι αναπαριστούν οι κατηγορίες και τα σχετικά τους μεγέθη. Για μια συνεχή μεταβλητή ο ορισμός της κλίμακας γίνεται ακόμα πιο δύσκολος.

Στο "Grammar of Graphics" ο Wilkinson αναφέρεται (Leland Wilkinson, 2005) σε κάποιες ιδιότητες που μία καλή κλίμακα θα πρέπει να κατέχει: απλότητα, σταθμισμένη διασπορά, κάλυψη, ύπαρξη του μηδενός. Πρέπει να λαμβάνουμε υπόψη το τι αναπαριστούν τα δεδομένα και πως αυτά συλλέγονται. Σε κάθε περίπτωση ο χρήστης θα πρέπει να ελέγχει το αποτέλεσμα και να το τροποποιεί την κλίμακα αν αυτό κρίνεται σκόπιμο.

2.4.8 Επικεφαλίδες, λεζάντες, σχόλια

Ιδανικά, οι επικεφαλίδες πρέπει να εξηγούν πλήρως το γράφημα και να δηλώνουν την πηγή προέλευσης των δεδομένων. Οι λεζάντες περιγράφουν το πως τα σύμβολα ή τα χρώματα αντιστοιχίζονται με τις ομάδες των δεδομένων. Ο Tufte υποστήριξε ότι πρέπει να συμπεριλαμβάνονται απευθείας στο γράφημα και όχι να υπάρχουν ξεχωριστά, ώστε να μην χρειάζεται το μάτι του αναγνώστη να μετακινηθεί μπρος ή πίσω. Τα σχόλια χρησιμοποιούνται για να τονίσουν κάποια συγκεκριμένα χαρακτηριστικά του γραφήματος. Λόγω έλλειψης χώρου δεν μπορεί να είναι μεγάλα σε

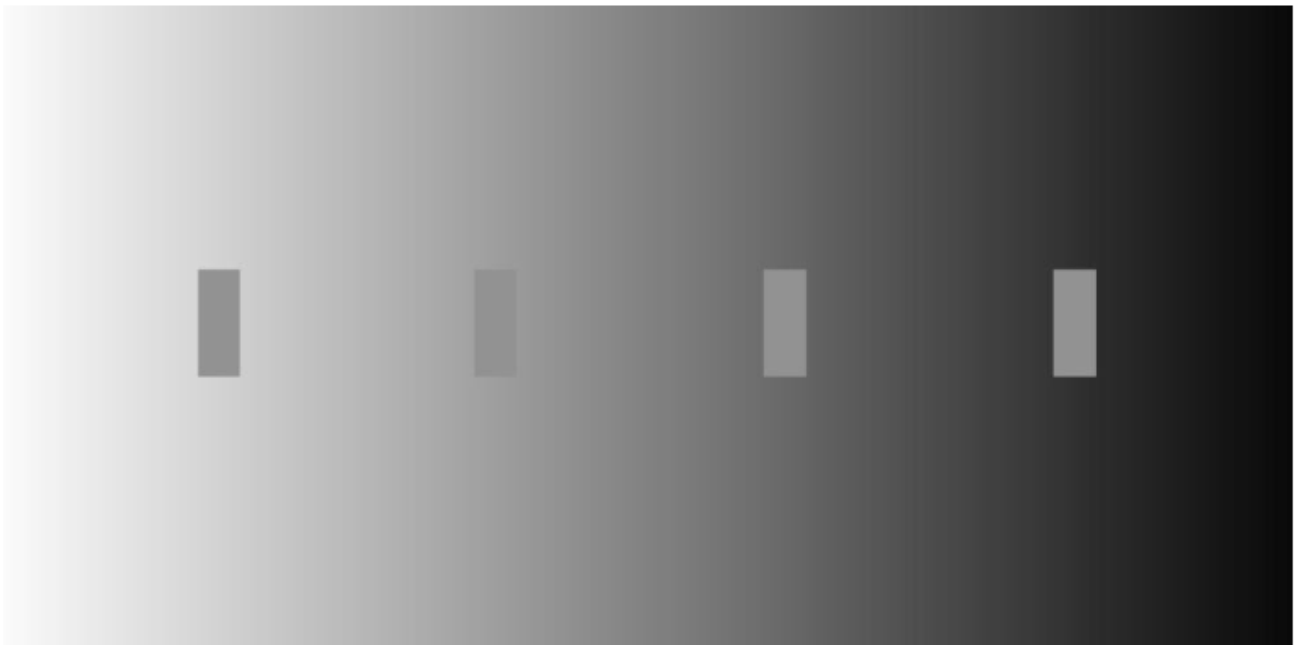
έκταση. Μπορούν να χρησιμοποιηθούν για να προσδιορίσουμε γεγονότα σε κάποιο διάγραμμα χρόνου ή για να τραβήξουμε την προσοχή σε συγκεκριμένα σημεία κάποιου διαγράμματος σημείων.

2.5 Το ανθρώπινο οπτικό σύστημα

Για να μπορέσουμε να παρουσιάσουμε την πληροφορία οπτικά με αποδοτικό τρόπο θα πρέπει πρώτα να κατανοήσουμε κάπως την οπτική αντίληψη του ανθρώπου. Το ανθρώπινο οπτικό σύστημα είναι μία αναζήτηση μοτίβων μεγάλης δύναμης και λεπτότητας. Το μάτι και ο οπτικός φλοιός του εγκεφάλου σχηματίζουν έναν μαζικά παράλληλο επεξεργαστή που παρέχει το κανάλι με το υψηλότερο εύρος ζώνης στα ανθρώπινα κέντρα γνώσης. Αν κατανοήσουμε τον τρόπο με τον οποίο λειτουργεί η αντίληψή μας τότε η γνώση μας θα μπορέσει να μεταφραστεί σε κανόνες απεικόνισης πληροφορίας. Αν δεν υπακούσουμε στους κανόνες αυτούς τα δεδομένα μας θα είναι μη κατανοητά ή παραπλανητικά.

Αυτό που βλέπουμε δεν είναι απλά μια καταγραφή του τι υπάρχει στο περιβάλλον μας. Η όραση είναι μια ενεργή διαδικασία που συμπεριλαμβάνει διερμηνείες από τον εγκέφαλο μας πάνω σε δεδομένα που έχουν ανιχνευτεί από τα μάτια μας σε μια προσπάθεια να βγάλει νόημα από τα συμφραζόμενα. Ο εγκέφαλος μας αντιλαμβάνεται περισσότερο οπτικές διαφορές, παρά απόλυτες τιμές.

Τα μάτια μας ανιχνεύουν το φως που ανακλάται από μια επιφάνεια ενός αντικειμένου. Αυτό που αντιλαμβανόμαστε ως αντικείμενο έχει δημιουργηθεί ως μια σύνθεση πολλαπλών οπτικών ιδιοτήτων, τις οποίες μπορούμε να διαχωρίσουμε. Οι ιδιότητες αυτές είναι μοναδικά χαρακτηριστικά του φωτός, για τις οποίες τα μάτια μας είναι συντονισμένα να ανιχνεύουν. Κάποια από αυτά τα χαρακτηριστικά είναι η θέση στον δυσδιάστατο χώρο, το μέγεθος, το μήκος, η περιοχή, το σχήμα, το χρώμα, ο προσανατολισμός.

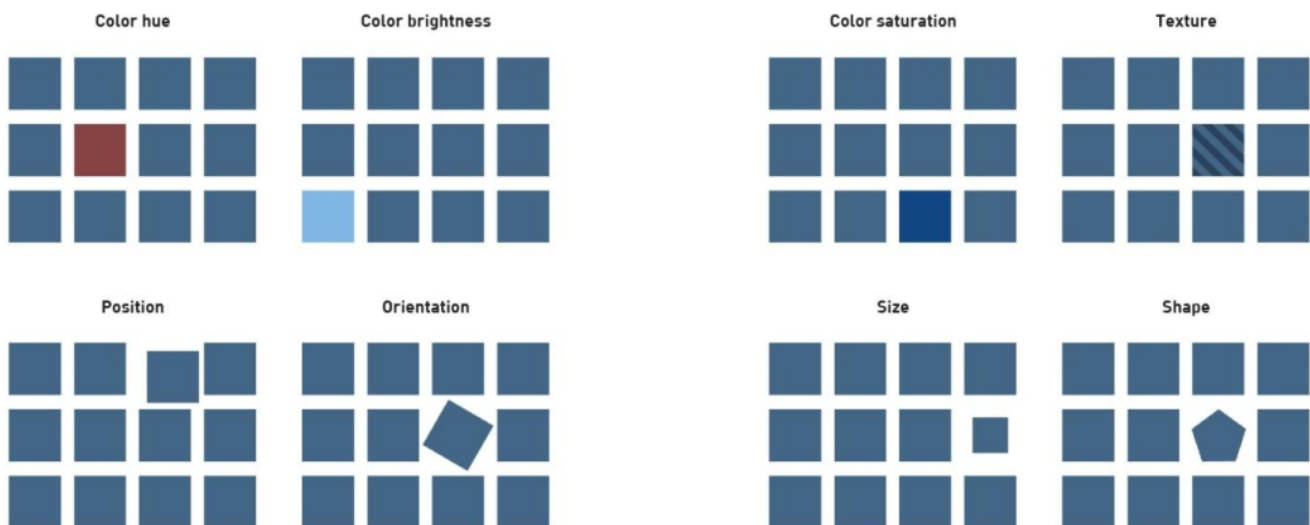


Εικόνα 2-15 Στο παράδειγμα της εικόνας, ενώ και τα τέσσερα αντικείμενα έχουν ακριβώς το ίδιο χρώμα (κατά απόλυτη τιμή), ο εγκέφαλος τα αντιλαμβάνεται ως διαφορετικά, λόγω της αλλαγής στην φωτεινότητα του φόντου. Στην συγκεκριμένη περίπτωση ο εγκέφαλος μας έχει αντιληφθεί την οπτική διαφορά του χρώματος του αντικειμένου με το χρώμα που το περιβάλλει. Μεταξύ των άλλων, ένα συμπέρασμα που μπορούμε να βγάλουμε μελετώντας το παραπάνω παράδειγμα είναι ότι πρέπει να αποφεύγουμε την χρήση χρωματικών διαβαθμίσεων (διανυσμάτων – color gradient) ως φόντο σε κάποιο γράφημα. Πηγή: Table and graph design for enlightening communication by Stephen Few. Perceptual Edge. Σελίδα 86

2.5.1 Preattentive μεταβλητές

Υπάρχουν κάποια συγκεκριμένα οπτικά χαρακτηριστικά τα οποία ο άνθρωπος αντιλαμβάνεται με την πρώτη ματιά, πριν ακόμα αυτός εστιάσει την προσοχή του και πριν ακόμα κάνει κάποια συνειδητή σκέψη. Η Jennifer Tidwell (Jennifer Tidwell, 2010) τις έχει ονομάσει "preattentive variables".

Με τον τρόπο αυτό και λαμβάνοντας υπ' όψιν ότι το ανθρώπινο μάτι διαθέτει εκατομμύρια υποδοχείς που δουλεύουν συντονισμένα και ταυτόχρονα, η preattentive αντίληψη είναι μία εξαιρετικά γρήγορη και παράλληλη διαδικασία. Όταν τα δεδομένα αναπαρίστανται με την χρήση κάποιων preattentive μεταβλητών, είμαστε σε θέση να αντιληφθούμε έναν μεγάλο όγκο πληροφορίας γρήγορα, αλλά και να αποθηκεύσουμε ένα ολόκληρο αντικείμενο που αποτελείται από πολλές ιδιότητες, ως ένα ενιαίο κομμάτι στην μνήμη μας.



Εικόνα 2-16 Οι 8 preattentive μεταβλητές του Tidwell. Πηγή: An introduction to visualizing data by Joel Laumans (σελίδα 14).

2.5.2 Περισσότερη κωδικοποίηση (Redundant Encoding)

Αν μετά την κωδικοποίηση των κύριων διαστάσεων των δεδομένων μας, μας έχουν μείνει κάποιες οπτικές ιδιότητες αχρησιμοποίητες μπορούμε να τις χρησιμοποιήσουμε για να κωδικοποιήσουμε ήδη κωδικοποιημένες διαστάσεις. Το πλεονέκτημα αυτής της προσέγγισης είναι ότι χρησιμοποιώντας περισσότερα του ενός κανάλια μπορούμε να κατανοήσουμε την ίδια πληροφορία πιο εύκολα, γρήγορα και με περισσότερη ακρίβεια.

2.5.3 Κατηγοριοποίηση οπτικών ιδιοτήτων

Φυσική διάταξη (Natural Ordering)

Το αν μια οπτική ιδιότητα έχει φυσική διάταξη εξαρτάται από το αν οι μηχανισμοί του "λογισμικού" του εγκεφάλου μας και του οπτικού μας συστήματος, αυτόματα και χωρίς σκοπό, αποδίδουν μια ταξινόμηση ή μια κατάταξη στις διάφορες πιθανές τιμές αυτής της ιδιότητας. Το "λογισμικό" που κάνει αυτήν την κρίση βρίσκεται βαθιά ενσωματωμένο στο νου μας και αξιολογεί την σχετική διάταξη ανεξάρτητα από την γλώσσα ή την κουλτούρα.

Οπτικές ιδιότητες όπως η θέση, το μήκος, το πάχος μιας γραμμής, η γωνία, η έκταση μιας περιοχής, η πυκνότητα προτύπου, η φωτεινότητα και ο κορεσμός έχουν φυσική διάταξη, ενώ το σχήμα, η υφή (texture), το στυλ της γραμμής (συνεχής, με τελείες, διακεκομμένη) και το χρώμα δεν έχουν φυσική διάταξη.

Διακριτές τιμές (Distinct Values)

Ένας άλλος παράγοντας που πρέπει να λάβουμε σοβαρά υπ' όψιν για την επιλογή της κατάλληλης οπτικής ιδιότητας είναι πόσες διακριτές τιμές αυτή έχει ώστε ο αναγνώστης να είναι ικανός να ξεχωρίσει να τις αντιληφθεί και πιθανώς να τις θυμηθεί. Για παράδειγμα, αν και υπάρχουν άπειρα χρώματα ο αναγνώστης της οπτικοποίησης δεν μπορεί να τα διακρίνει αν είναι παρόμοια. Σε αντίθεση, μπορεί πιο εύκολα να ξεχωρίσει έναν μεγάλο αριθμό σχημάτων, θέσεων, αριθμών.

Όπως σημειώσαμε πριν το χρώμα δεν έχει φυσική διάταξη και για τον λόγο αυτό δεν πρέπει να το χρησιμοποιούμε για να ταξινομήσουμε, να διατάξουμε σε σειρά ή να κατατάξουμε τα δεδομένα μας. Μπορούμε όμως να χρησιμοποιήσουμε το χρώμα για να επισημάνουμε κατηγορικά δεδομένα. Βέβαια αυτό δεν μπορεί να συμβεί αν έχουμε μεγάλο αριθμό κατηγοριών, γιατί πολύ απλά θα χρειαστεί να χρησιμοποιήσουμε παρόμοια χρώματα που δεν μπορεί κάποιος να τα ξεχωρίσει εύκολα.

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

Εικόνα 2-17 Ο παραπάνω πίνακας των συνηθισμένων οπτικών ιδιοτήτων μπορεί να μας φανεί αρκετά χρήσιμος κατά την διαδικασία κωδικοποίησης για τους τύπους των δεδομένων μας. Πηγή: Designing data visualizations by Noab Iliinski & Julie Steele (σελίδα 28).

2.6 Δυναμικές οπτικοποιήσεις δεδομένων

2.6.1 Διαδραστικότητα

Η διαδραστικότητα αναφέρεται στην δυνατότητα των χρηστών να ενεργούν επί της οπτικοποίησης, ελέγχοντας και επιλέγοντας τι και πως θα οπτικοποιηθεί. Όταν επιτρέπουμε στον χρήστη να αλληλεπιδράσει με τα δεδομένα, πλέον δεν έχουμε να κάνουμε μόνο με το πως οπτικοποιούνται τα δεδομένα αλλά και με το πως αυτά συμπεριφέρονται.

Υπάρχουν πολλές φόρμες διεπαφών που μπορούν να εφαρμοστούν πάνω σε οπτικοποιήσεις δεδομένων. Οι φόρμες αυτές μπορούν να ενταχθούν σε δύο κατηγορίες (Galitz, 2002):

- Επιλογή δεδομένων και φιλτράρισμα: Ο χρήστης ελέγχει ποια δεδομένα θα οπτικοποιηθούν. Αυτό τον βοηθά να εστιάσει στην αναζήτηση μόνο των δεδομένων που τον αφορούν, εμποδίζοντας με τον τρόπο αυτό την υπερφόρτωση της πληροφορίας.
- Διευθέτηση και πλοήγηση δεδομένων: Ο χρήστης επιλέγει τον τρόπο με τον οποίο παρουσιάζονται ή οπτικοποιούνται τα δεδομένα την επιλογή αυτή ο χρήστης μπορεί να βρει κάποιο νέο νόημα στα δεδομένα. Παρουσιάζοντας τα ίδια δεδομένα με διαφορετικό τρόπο, μπορούμε να οδηγηθούμε σε νέα συμπεράσματα αλλά και να δούμε διαφορετικές σχέσεις ανάμεσα στα δεδομένα.

Αν επιτρέψουμε τον χρήστη να ελέγχει τις δύο παραπάνω μεταβλητές, τότε έχουμε να κάνουμε με αποτελεσματικότερες οπτικοποιήσεις, αφού μπορούν να τον επικεντρώσουν στις δικές του ανάγκες.

Με βάση το πόσο διαδραστική είναι μία οπτικοποίηση μπορεί να ανήκει σε κάποια από τις παρακάτω κατηγορίες (Joel Laumans):

- Στατική οπτικοποίηση: Δεν αλλάζει καθόλου στο πέρασμα του χρόνου και δεν προσφέρει καμία διαδραστική λειτουργικότητα.
- Κινούμενη οπτικοποίηση: Μεταβάλλεται στο πέρασμα του χρόνου. Χρησιμοποιεί τον χρόνο ως μια επιπλέον μεταβλητή ή διάσταση.
- Διαδραστική οπτικοποίηση: Ο χρήστης μπορεί να αλληλεπιδράσει με την οπτικοποίηση επιλέγοντας ποια δεδομένα θα οπτικοποιηθούν και με ποιον τρόπο. Ωστόσο τα αποτελέσματα της αλληλεπίδρασης δεν είναι άμεσα ορατά.
- Απευθείας χειρισμός (Direct Manipulation): Ο χρήστης έχει τις ίδιες δυνατότητες όπως με μια διαδραστική οπτικοποίηση, με την μόνη διαφορά ότι στην περίπτωση του απευθείας χειρισμού τα αποτελέσματα της αλληλεπίδρασης γίνονται άμεσα ορατά.

2.6.2 Συστήματα πραγματικού χρόνου

Επειδή τα δεδομένα έρχονται από τον πραγματικό κόσμο, όπου δεν υπάρχουν απόλυτα, δεν είναι σταθερές τιμές προς ανάλυση, αντιθέτως αποτελούν κινούμενο στόχο. Για παράδειγμα, η θερμοκρασία αλλάζει συνεχώς, η επισκεψιμότητα ενός ιστότοπου αυξάνεται με την κυκλοφορία ενός νέου προϊόντος, κλπ.. Κρίνεται λοιπόν αναγκαίο να δημιουργήσουμε αναπαραστάσεις δεδομένων που αναπροσαρμόζονται με νέες τιμές κάθε εβδομάδα, ώρα, δευτερόλεπτο κ.ο.κ. .

2.6.3 Ροή Γραφήματος (Graph Stream)

Οι δυναμικοί γράφοι μοντελοποιούνται ως μια ροή γεγονότων που συμβαίνουν στον γράφο. Τα γεγονότα αυτά μπορεί να αφορούν είτε την δομή του γράφου (προσθήκη ή αφαίρεση στοιχείων) είτε τις ιδιότητες των στοιχείων. Ένα stream είναι η διασύνδεση μεταξύ μιας πηγής γεγονότων και ενός αποδέκτη. Μια πηγή γεγονότων μπορεί να είναι για παράδειγμα ένα αρχείο γεγονότων ή ένας αλγόριθμος παραγωγής γράφου.

Το Graph Stream παρέχει επίσης μερικά χαρακτηριστικά οπτικοποίησης γράφων. Η απόδοση των στοιχείων μπορεί να προσαρμόζεται μέσω κανόνων CSS μορφοποίησης. Με τον τρόπο αυτό ο θεατής της οπτικοποίησης δεν βλέπει απλά μια στατική οπτικοποίηση αλλά έναν γράφο που αλλάζει συνεχώς.

Ο θεατής κοιτώντας τον γράφο μπορεί να παρατηρήσει την ακολουθία των γεγονότων και να βγάλει συμπεράσματα για το πως το δίκτυο εξελίσσεται στον χρόνο.

Για την υλοποίηση του Graph Stream μπορούμε να χρησιμοποιήσουμε την ομώνυμη βιβλιοθήκη της Java. Στο λογισμικό Gephi, που θα παρουσιάσουμε αναλυτικά στην επόμενη ενότητα, μπορούμε να χρησιμοποιήσουμε το πρόσθετο λογισμικό Graph Streaming API.