

# ΟΠΤΙΚΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ

Αναπλ. Καθηγ. Στελιος Ζήμερας  
Τμηση Στατιστικής και Αναλογιστικών –  
Χρηματοοικονομικών Μαθηματικών  
Πανεπιστήμιο Αιγαίου  
Σαμος

2021

# Εισαγωγή

Η απεικόνιση των πληροφοριών είναι η διαδικασία παρουσίασης αφηρημένων επιχειρήσεων ή επιστημονικών δεδομένων με τη μορφή εικόνων που μπορούν να βοηθήσουν στην κατανόηση της έννοιας των δεδομένων

Η χρήση της απεικόνισης δεδομένων είναι μια παρουσίαση των πληροφοριών με τη μορφή διαγραμμάτων

Η οπτικοποίηση δεδομένων είναι η παρουσίαση των δεδομένων σε εικονογραφική ή γραφική μορφή

# Εισαγωγή

Κατά την ανάλυση δεδομένων χρησιμοποιώντας την απεικόνιση, χρησιμοποιείται το λεγόμενο γρήγορο πρωτότυπο - δηλαδή, δημιουργώντας ένα μεγάλο αριθμό διαφορετικών οπτικών αναπαραστάσεων των ίδιων δεδομένων. Έγινε για τη δυνατότητα να βρεθούν κρυμμένες, με την πρώτη ματιά, τις αλληλεξάρτηση και τις εξαρτήσεις, καθώς και την πρωτογενή αξιολόγηση ενός συνόλου δεδομένων σε ενδεχομένως εφαρμογές σε περαιτέρω πιο σύνθετα εργαλεία ανάλυσης

# Εισαγωγή

η οπτικοποίηση δεδομένων είναι ένας γενικός όρος που περιγράφει τη χρήση οπτικών αναπαραστάσεων για να βοηθήσουν τους ανθρώπους να αναλύσουν και να ερμηνεύσουν τα δεδομένα

κύριες χρήσεις των οπτικών αναπαραστάσεων είναι:

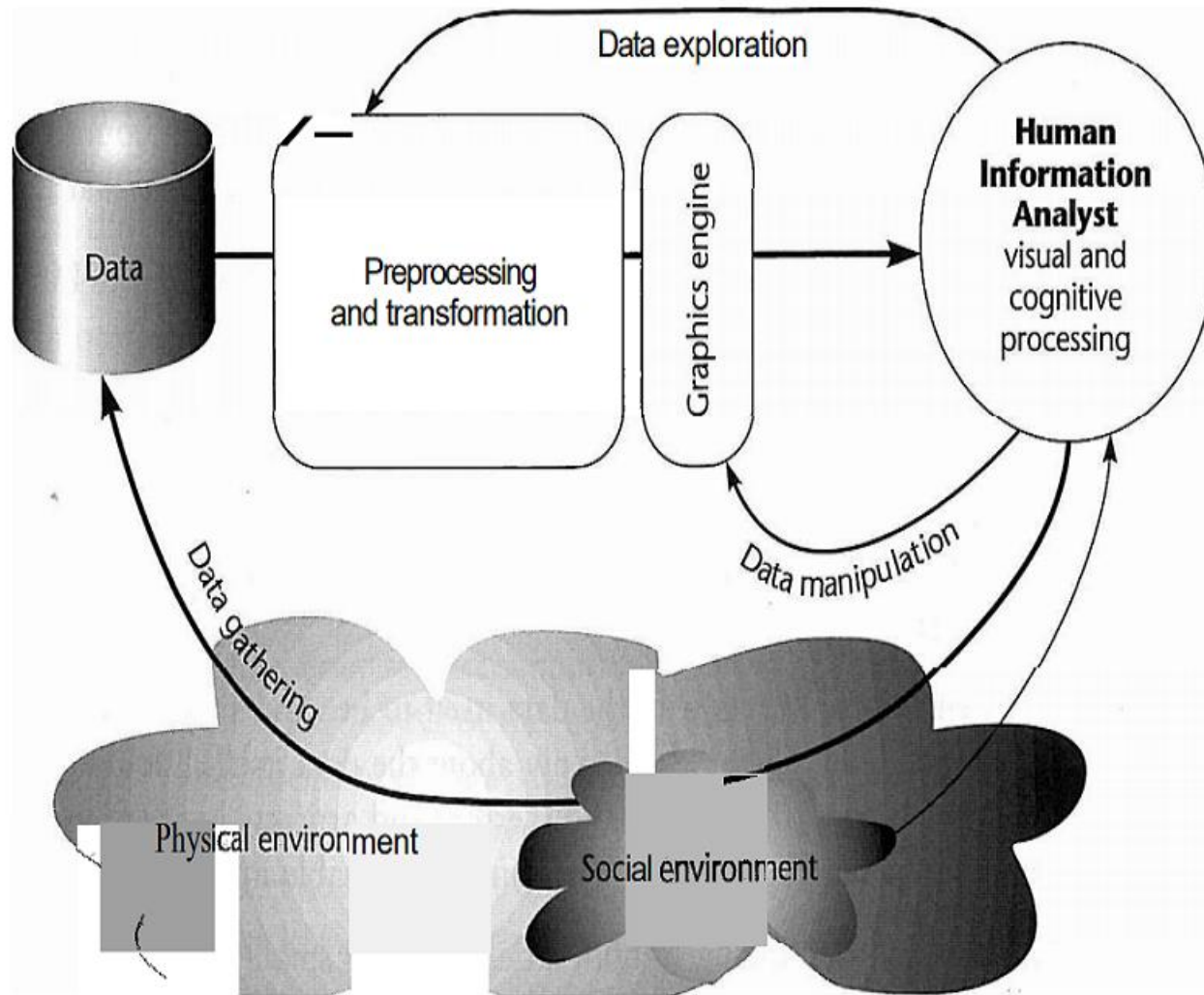
- (1) να βρουν και να εξετάσουν το νόημα των δεδομένων και να το κατανοήσουν, και
- (2) να μεταδώσουν αυτή την πληροφορία στο κοινό.

Η επίτευξη της αριστείας στην οπτικοποίηση των δεδομένων απαιτεί προγραμματισμό, γνώση των καθιερωμένων αρχών και κατάλληλη χρήση των διαθέσιμων εργαλείων.

# Εισαγωγή

1700-1799	Νέες μορφές γραφικής απεικόνισης
1800-1850	Αρχές σύγχρονων γραφικών αναπαραστάσεων
1850-1900	Η χρυσή εποχή των στατιστικών γραφικών αναπαραστάσεων
1900-1950	Μοντέρνα σκοτεινή εποχή
1950-1975	Αναγέννηση της απεικόνισης δεδομένων
1975-σήμερα	Διαδραστική και δυναμική απεικόνιση δεδομένων

# Στάδια Οπτικοποίησης



Colin Ware μοντέλο (2004)

# Στάδια Οπτικοποίησης

Σύμφωνα με την εικόνα τα τέσσερα στάδια αποτελούνται από:

- Τη συλλογή και την αποθήκευση των δεδομένων.
- Την προεργασία για την μετατροπή των δεδομένων σε κάτι που μπορούμε να καταλάβουμε.
- Τους αλγορίθμους γραφικών που παράγουν μία εικόνα σε μία οθόνη.
- Την ανθρώπινη αντίληψη.

# Διαγραμματική απεικόνιση

Edward Rolf Tufte

- Να δείχνει τα δεδομένα.
- Να μην διαταράσσει το νόημα των δεδομένων.
- Να παρουσιάζει πολλά δεδομένα σε περιορισμένο χώρο.
- Να κάνει συνεκτικά μεγάλα σύνολα δεδομένων.
- Να ενθαρρύνει την επαγωγή συμπερασμάτων, πχ με σύγκριση τιμών.
- Να δίνει διαφορετικές οπτικές γωνίες των δεδομένων, από συνοπτικές έως αναλυτικές.



# Παράγοντες κατάλληλης επιλογής

- Το πρόβλημα. Αφορά το τι πρέπει να βρεθεί η να παρουσιαστεί.
- Η φύση των δεδομένων. Τα δεδομένα μπορεί να είναι αριθμητικά, ονομαστικά, κείμενο κλπ. Ανάλογα με τον τύπο των δεδομένων υπάρχει και η κατάλληλη τεχνική οπτικοποίησης.
- Το πλήθος των διαστάσεων. Ο αριθμός των διαστάσεων των δεδομένων που πρέπει να παρουσιαστεί καθορίζει τον τύπο της αναπαράστασης που χρησιμοποιούμε.
- Η δομή των δεδομένων. Τα δεδομένα μπορεί να είναι γραμμικά(πίνακες) , χωρικά ή γεωγραφικά (χάρτης), χρονικά (δεδομένα που αλλάζουν στο χρόνο), ιεραρχικά ή να έχουν δικτυακή δομή.
- Ο τύπος της αλληλεπίδρασης, Αυτό καθορίζει εάν η οπτική αναπαράσταση είναι στατική (π.χ. μια εικόνα τυπωμένη σε χαρτί ή μια εικόνα που απεικονίζεται σε οθόνη υπολογιστή αλλά όχι τροποποιημένη από τον χρήστη), μετασχηματιστή (όταν ο χρήστης μπορεί να ελέγξει τη διαδικασία τροποποίησης και μετασχηματισμού δεδομένων )όπως μεγέθυνση δεδομένων, επιλογή δεδομένων κλπ.

# Ταξινόμηση μεθόδων οπτικοποίησης

- τον τύπο των δεδομένων,
- την τεχνική οπτικοποίησης,
- την τεχνική αλληλεπίδρασης και στρέβλωσης

Keim(2002),

# Ταξινόμηση μεθόδων οπτικοποίησης

Ως προς τον τύπο τους τα δεδομένα μπορεί να είναι :

- Μονοδιάστατα , δηλαδή τα δεδομένα έχουν μια διάσταση όπως τα χρονικά δεδομένα.
- Δυσδιάστατα. Τα δεδομένα έχουν δύο διαστάσεις, 'όπως οι γεωγραφικοί χάρτες.
- Πολυδιάστατα. Τα δεδομένα έχουν πολλές διαστάσεις, όπως σχεσιακοί πίνακες.
  - Κείμενο και Υπερκείμενο. Τα δεδομένα είναι αδόμητα και δεν μπορούν να εκφραστούν σε σχέση με διαστάσεις. Όπως άρθρα ειδήσεων και διαδικτυακά έγγραφα.
- Ιεραρχίες και Γράφοι .
- Αλγόριθμοι και λογισμικό. Αφορά δεδομένα ροής πληροφοριών σε ένα πρόγραμμα.

# Ταξινόμηση μεθόδων οπτικοποίησης

Οι τεχνικές απεικόνισης που χρησιμοποιούνται μπορεί να ταξινομηθούν σε:

- Τυπικές δύο ή τριών διαστάσεων (Standard 2D/3D displays). Σχετικά απλές τεχνικές, που συνήθως εφαρμόζονται στα πρώτα στάδια της ανάλυσης. Δεν είναι κατάλληλες για την οπτικοποίηση σύνθετων δομών.
- Γεωμετρικού Μετασχηματισμού (Geometrically Transformed). Πολυδιάστατα δεδομένα μετασχηματίζονται και προβάλλονται με γεωμετρικό τρόπο, ώστε να αποκαλυφθούν πιθανές σχέσεις τους. Περίπτωση τεχνικής γεωμετρικού μετασχηματισμού είναι τα διαγράμματα παράλληλων συντεταγμένων.
- Εικονογραφικές (Iconic Displays). Κάθε παρατήρηση (αντικείμενο) αντιστοιχίζεται σε μια εικόνα και κάθε τιμή της παρατήρησης αντιστοιχίζεται με ένα χαρακτηριστικό της εικόνας, πχ σχήμα, μέγεθος, χρώμα κλπ. Στα πρόσωπα του Chernoff, κάθε παρατήρηση αντιστοιχεί σε ένα πρόσωπο, και τα χαρακτηριστικά του προσώπου (μύτη, αυτιά κλπ.) εκφράζουν τις τιμές των μεταβλητών. Ο ερευνητής συγκρίνει τις εικόνες για να βρει ομοιότητες και διαφορές. Τεχνικές κατάλληλες για μέτριο πλήθος δεδομένων και σχετικά μικρό αριθμό μεταβλητών.
- Εικονοστοιχείων (Dense Pixel Displays). Κάθε τιμή των δεδομένων αντιστοιχίζεται σε ένα pixel, το οποίο χρωματίζεται ανάλογα με την τιμή. Τα εικονοστοιχεία μιας διάστασης τοποθετούνται σε γειτονικές περιοχές. Τεχνικές κατάλληλες για

# Ταξινόμηση μεθόδων οπτικοποίησης

απεικόνιση περίπου 1.000.000 τιμών, με αδυναμίες όμως στον εντοπισμό σύνθετων δομών δεδομένων.

- Στοιβάς (ή ιεραρχικές) (Stacked Displays). Η παρουσίαση των δεδομένων γίνεται στη βάση μιας ιεράρχησης. Οι τύποι των ιεραρχήσεων ποικίλουν. Παράδειγμα τέτοιας τεχνικής είναι η Dimensional Stacking, όπου ο χώρος απεικόνισης χωρίζεται σε τμήματα ανάλογα με δύο διαστάσεις και εντός αυτών των τμημάτων γίνεται απεικόνιση των δεδομένων ανάλογα με δύο άλλες διαστάσεις. Άλλο παράδειγμα, με τελείως διαφορετικό τρόπο ιεράρχησης, είναι τα Δενδρογράμματα, που αποτυπώνουν τη διαδικασία διαδοχικής συγχώνευσης συστάδων και χρησιμοποιούνται στην Ανάλυση Συστάδων

# Ταξινόμηση μεθόδων οπτικοποίησης

Αναφορικά με τον τρόπο αλληλεπίδρασης και στρέβλωσης οι τεχνικές κατηγοριοποιούνται ως:

- Δυναμικής προβολής (Dynamic Projections). Συνίσταται στη μεταβολή του τρόπου προβολής των δεδομένων.
- Διαδραστικής επιλογής (Interactive Filtering). Επιτρέπει την τμηματοποίηση των δεδομένων και την επικέντρωση σε ένα υποσύνολο. Το υποσύνολο των δεδομένων μπορεί να προκύψει είτε με την εκτέλεση κάποιου ερωτήματος είτε με την άμεση επιλογή από τον χρήστη.
- Διαδραστικής Διαβάθμισης Λεπτομέρειας (Interactive Zooming). Είναι η δυνατότητα προβολής σε διαφορετικό βαθμό λεπτομέρειας. Τα αντικείμενα μπορεί να μεγεθυνθούν ή μπορεί να προβληθεί διαφορετικού τύπου πληροφορία, όπως πχ κείμενο.
- Διαδραστικής στρέβλωσης (Interactive Distortion). Συνίσταται στην προβολή του συνόλου των δεδομένων με χαμηλό βαθμό λεπτομέρειας, με ταυτόχρονη προβολή τμήματος των δεδομένων με υψηλό βαθμό λεπτομέρειας.
- Διαδραστικής Σύνδεσης και Χρωματισμού (Interactive Linking and Brushing). Είναι ο συνδυασμός διαφορετικών τεχνικών οπτικοποίησης. Για παράδειγμα, σε ένα σύνολο διαγραμμάτων διασποράς μπορεί να χρωματιστούν και να συνδεθούν ορισμένα σημεία σε όλα τα διαγράμματα.

# Ταξινόμηση μεθόδων οπτικοποίησης

Οι έξι (6) «αλγόριθμοι» του Manovich για σχεδιασμό καλαίσθητων οπτικοποιήσεων είναι:

1. *Μοντέρνος σχεδιασμός / infonis.* Κάθε πληροφορία οπτικού σχεδιασμού έχει έναν αριθμό οπτικών γνωρισμάτων όπως χρώμα, μήκος γραμμής, γραμματοσειρά, μέγεθος γραμματοσειράς κ.ά. Η πιο σημαντική αρχή σχεδιασμού είναι η χρήση μερικών μόνο επιλογών για κάθε χαρακτηριστικό (π.χ. θα χρησιμοποιηθούν μόνο δύο μεγέθη γραμματοσειράς, μία χρωματική παλέτα τριών ή τεσσάρων χρωμάτων) κ.ο.κ.
2. *Σύνδεση οπτικών χαρακτηριστικών με τη σημασιολογία τους.* Δηλαδή αυτά τα σχέδια τι αναπαριστούν (π.χ. ένα χρώμα ή μία διαφορετικού μεγέθους γραμμή θα υποδηλώνει διαφορετικό τύπο περιεχομένου).
3. *Χρώμα.* Χρήση πολυάριθμων ιστοτόπων, οι οποίοι περιλαμβάνουν επαγγελματικές χρωματικές παλέτες και color palette generators.
4. *Γραμματοσειρές.* Χρήση όχι παραπάνω από μίας οικογένειας γραμματοσειράς με δύο στυλ βάρους (font weight).

# Ταξινόμηση μεθόδων οπτικοποίησης

5. *Πλέγμα (Grid)*. Είναι πολύ χρήσιμο όταν συνδυάζονται πολλές οπτικοποιήσεις μαζί σε ένα σχέδιο. Επίσης χρησιμεύει όταν προστίθενται blocks σε περιγραφικό κείμενο μέσα στην οπτικοποίηση.
6. *Αφαίρεση περιττών στοιχείων*. Ο καλύτερος τρόπος απόδοσης μιας οπτικοποίησης είναι η απλούστευση και όχι περιττές ετικέτες, γραμμές πλέγματος, γραμμές άξονα μέσα στο γράφημα.



# ΓΡΑΦΗΜΑ

Γράφημα : Αποκαλείται μία γραφική παράσταση μίας ή πολλών μεταβλητών. Τα γραφήματα χρησιμοποιούνται για την εμφάνιση σειρών αριθμητικών δεδομένων σε γραφική μορφή, ώστε να διευκολύνεται η κατανόηση μεγάλων ποσοτήτων δεδομένων και η σχέση μεταξύ διαφορετικών σειρών δεδομένων. Οι γραφικές παραστάσεις είναι το καλύτερο μέσο στατιστικής παρουσίασης γιατί δίνουν στους αφηρημένους αριθμούς μία συγκεκριμένη μορφή που μας διευκολύνει να έχουμε, με τη βοήθεια ενός γεωμετρικού σχήματος, μία άμεση αντίληψη της μορφής του φαινομένου που θέλουμε να μελετήσουμε. Μία καλά σχεδιασμένη γραφική παράσταση μπορεί να γίνει κατανοητή και να διατηρηθεί στη μνήμη του αναγνώστη ευκολότερα από ένα αριθμητικό πίνακα.

# ΓΡΑΦΗΜΑ

1. Να ξεκαθαρίσουμε τους στόχους και τις προτεραιότητές μας σε ότι αφορά το μήνυμα που θέλουμε να δώσουμε στον αναγνώστη.
2. Να επιλέξουμε κατάλληλο είδος γραφικής παράστασης (αυτό εξαρτάται κατά κύριο λόγο από τη φύση των δεδομένων που έχουμε στη διάθεσή μας).
3. Να ενημερώσουμε τον αναγνώστη σχετικά με τη φύση των απεικονιζόμενων πληροφοριών με σαφή τίτλο.
4. Να κατασκευάσουμε σε τελική ανάλυση, ένα σχεδιάγραμμα το οποίο να είναι:
  - Παραστατικό, ώστε να διευκολύνει στην κατανόηση και απομνημόνευση των απεικονιζόμενων πληροφοριών.
  - Σαφές, χωρίς να δημιουργεί οπτική σύγχυση και να παρασύρει τον αναγνώστη ή ερευνητή σε εσφαλμένα συμπεράσματα.
  - Ακριβές, δηλαδή να αποτυπώνει την πραγματικότητα και να μην δημιουργεί πλασματικές εντυπώσεις.

# ΓΡΑΦΗΜΑ

Για να επιτευχθούν τα παραπάνω θα πρέπει μεταξύ άλλων να δοθεί προσοχή :

- a) Στις διαστάσεις του διαγράμματος. Στις περισσότερες περιπτώσεις η γραφική απεικόνιση των στοιχείων γίνεται με την βοήθεια ορθογώνιων συντεταγμένων. Συνήθως στον άξονα  $XX'$  (άξονα των τετμημένων) μετρώνται οι τιμές της μεταβλητής και στον άξονα  $YY'$  (άξονα των τεταγμένων) μετρώνται οι συχνότητες της κατανομής. Κατά μήκος των αξόνων του διαγράμματος πρέπει να αναγράφονται ευκρινώς οι κλίμακες μέτρησης των συχνοτήτων και οι μονάδες μέτρησης των μεταβλητών. (π.χ. βάρος σε κιλά, ποσοστό %). Η επιλογή της σωστής κλίμακας μέτρησης των συχνοτήτων και των σωστών διαστάσεων των αξόνων είναι κρίσιμης σημασίας, διότι αλλιώς είναι δυνατόν να δημιουργηθούν λανθασμένες εντυπώσεις.

# ΓΡΑΦΗΜΑ

- b) Στις χρησιμοποιούμενες γραμμοσκιάσεις και χρωματισμούς. Με τη χρήση διαγραμμάτων γίνεται προσπάθεια συμπύκνωσης ταυτόχρονα πολλών πληροφοριών. Θα πρέπει γενικά να αποφεύγεται η υπερβολή σε ότι αφορά την ποσότητα της πληροφόρησης που περιλαμβάνεται σε ένα διάγραμμα. Μία «φορτωμένη» γραφική παράσταση με μεγάλο πλήθος πληροφοριών είναι τις περισσότερες φορές δύσκολα ερμηνεύσιμη. Η «ανάγνωση» ενός διαγράμματος διευκολύνεται πολύ από την σωστή επιλογή γραμμών, γραμμοσκιάσεων και χρωμάτων καθώς και τη χρήση ευκρινούς υπομνήματος με το οποίο να διευκρινίζεται το τι υποδηλώνει η κάθε γραμμοσκίαση ή χρώμα. Η χρήση έντονων χρωμάτων (κόκκινο, πορτοκαλί) δημιουργούν μία επιβλητική εντύπωση, ενώ απαλά χρώματα, όπως τόνοι του γκρι, ή απουσία γενικά χρωμάτων, προσδίδουν μία απλούστερη και σοβαρότερη γραφική αποτύπωση των δεδομένων χωρίς ωστόσο να υστερούν σε λεπτομέρεια ή πληροφόρηση.
- c) Στην ορθολογική χρήση των εργαλείων λογισμικού. Τα σύγχρονα εργαλεία λογισμικού παρέχουν πλέον τεράστιες σχεδιαστικές δυνατότητες. Ο ερευνητής ωστόσο, δεν θα πρέπει να παρασύρετε από την ευελιξία και την παρεχόμενη ποικιλία των προγραμμάτων αυτών, ούτε και να υποτιμά την αμεσότητα και αποτελεσματικότητα που παρέχει μία «απλή» γραφική απεικόνιση.
- d) Σε τυχόν διευκρινήσεις που είναι απαραίτητο να δοθούν. Πρέπει να γίνεται αναφορά στην πηγή των στατιστικών στοιχείων, ενώ συμπληρωματικές επεξηγήσεις παρέχονται υπό μορφή υποσημειώσεων.

# ΓΡΑΦΗΜΑ

Τέλος πριν επιλέξουμε το είδος του διαγράμματος, καλό θα είναι να σκεφτούμε:

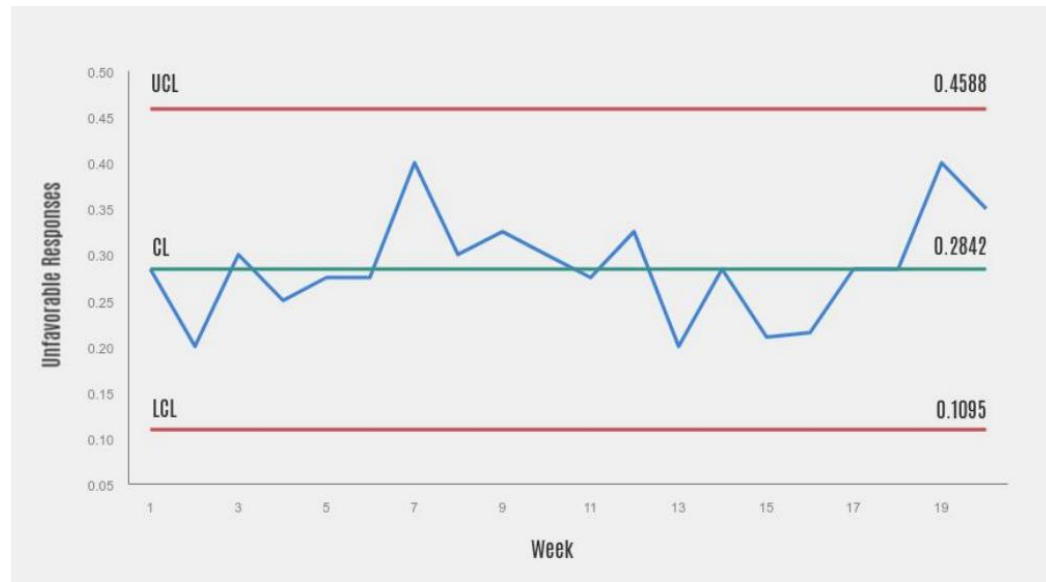
1. Τι ακριβώς θέλουμε να παρουσιάσουμε με το διάγραμμα;
2. Τι είδους είναι η μεταβλητή που εξετάζουμε;
3. Τι είδους είναι τα δεδομένα μας (π.χ. χρονολογικά, μεικτά);

# ΓΡΑΦΗΜΑ

## 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.

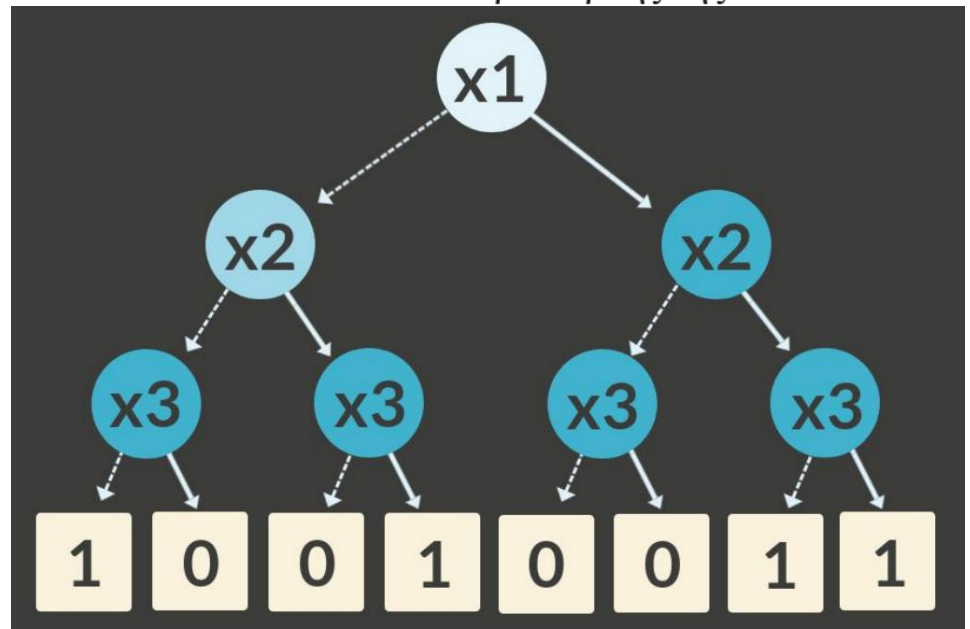


# ΓΡΑΦΗΜΑ

## Δυαδικά διαγράμματα αποφάσεων

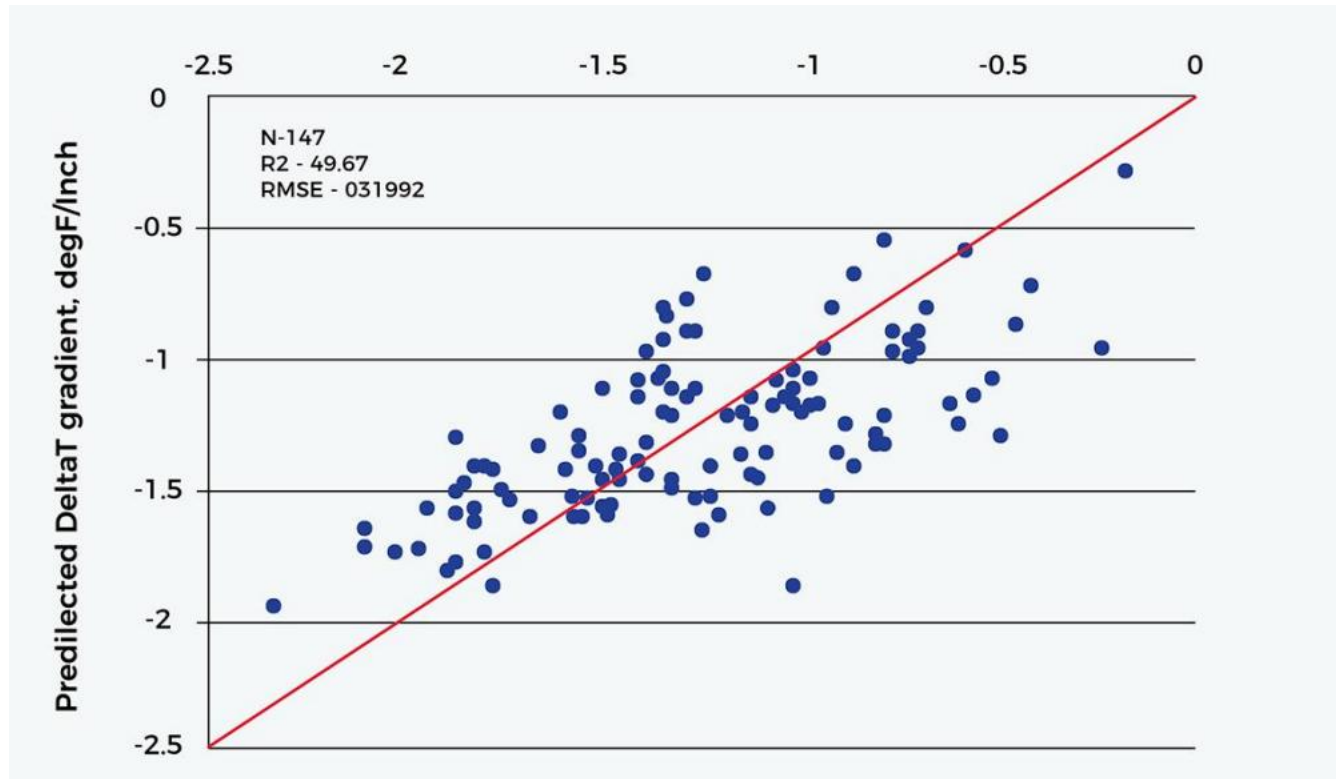
Μια δυαδική απόφαση είναι μια επιλογή μεταξύ δύο εναλλακτικών λύσεων, έτσι ένα διάγραμμα δυαδικής απόφασης απεικονίζει το μονοπάτι από τη μία απόφαση στην άλλη. Στην επιστήμη των υπολογιστών, οι δυαδικές αποφάσεις αποτελούν τον τύπο δεδομένων Boolean, στον οποίο δύο τιμές συνδέονται με διαφορετικές ενέργειες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



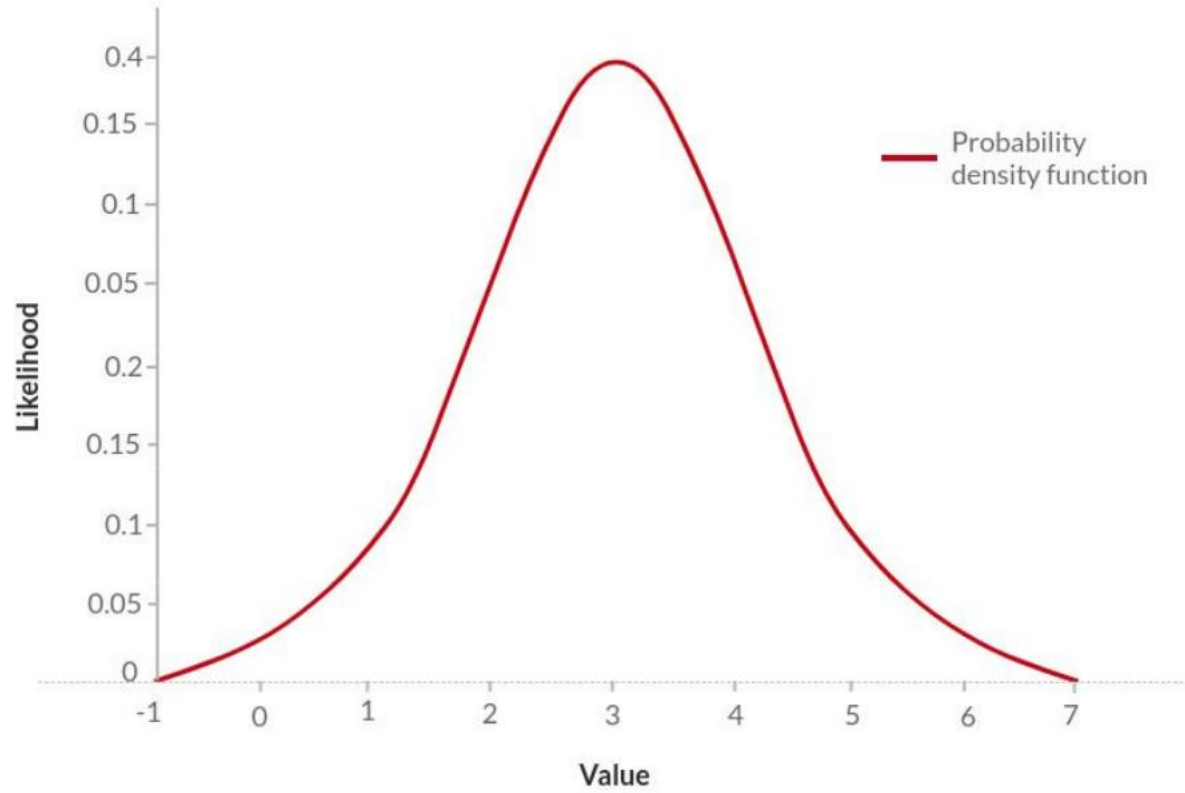
# ΓΡΑΦΗΜΑ

## Γραφήματα διασποράς



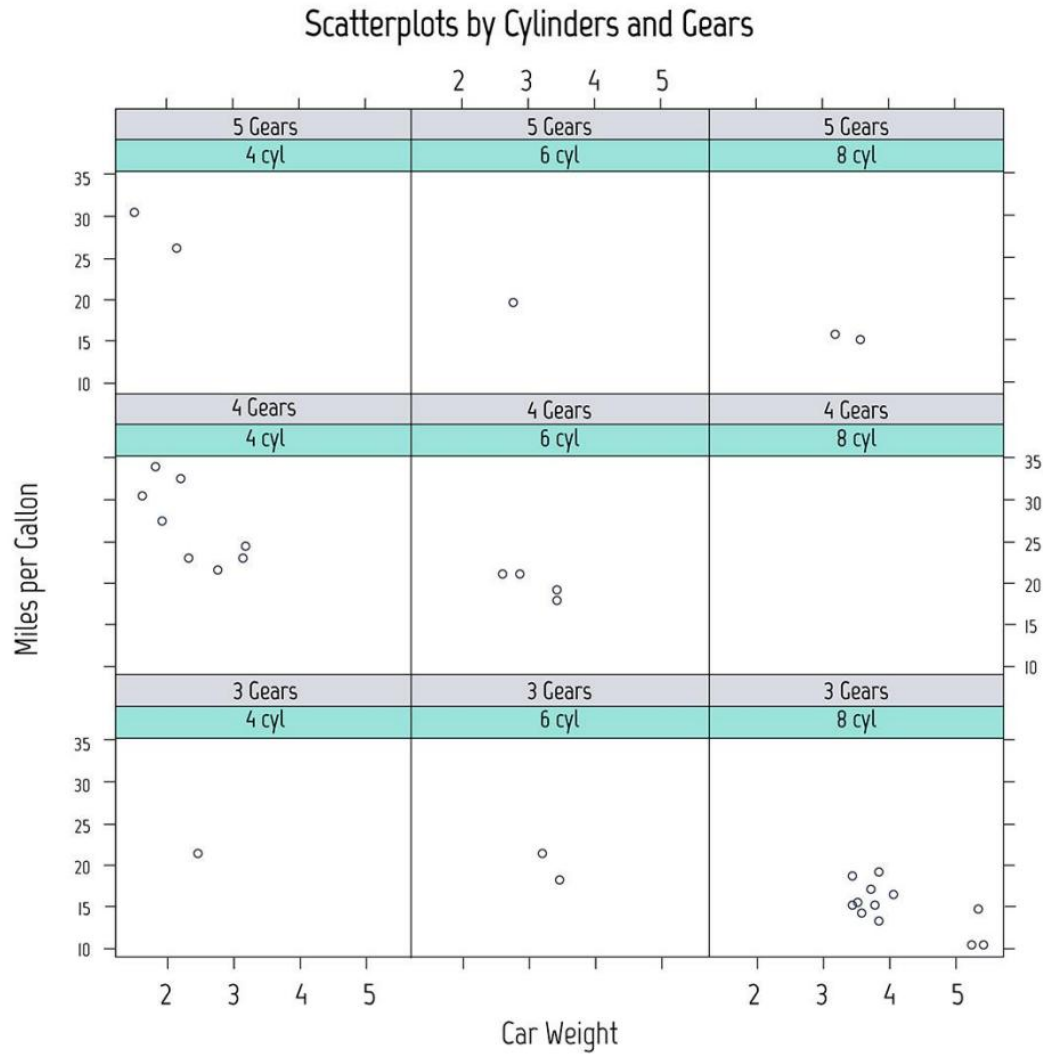


# ΓΡΑΦΗΜΑ



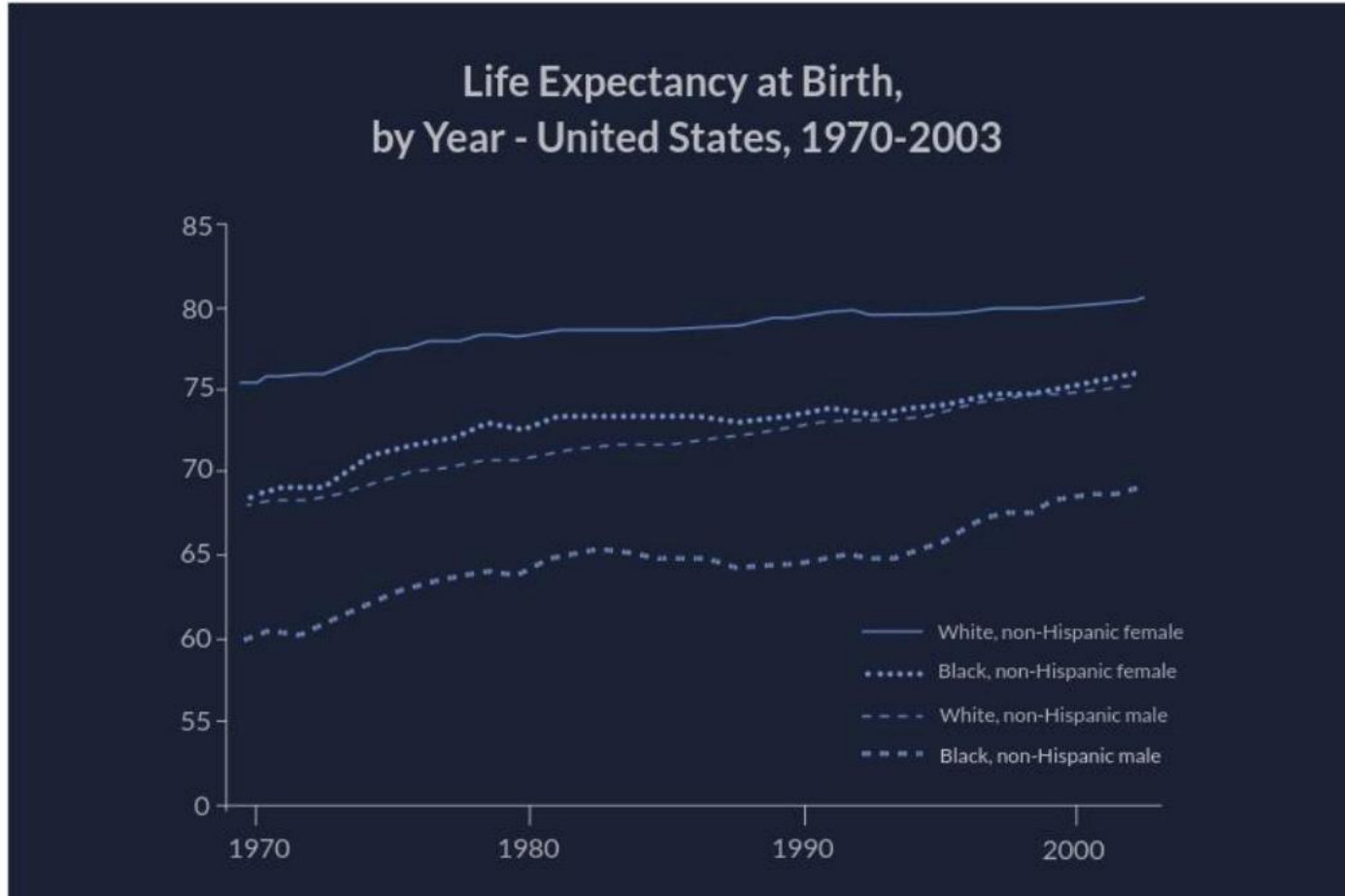
# ΓΡΑΦΗΜΑ

## Trellis plots



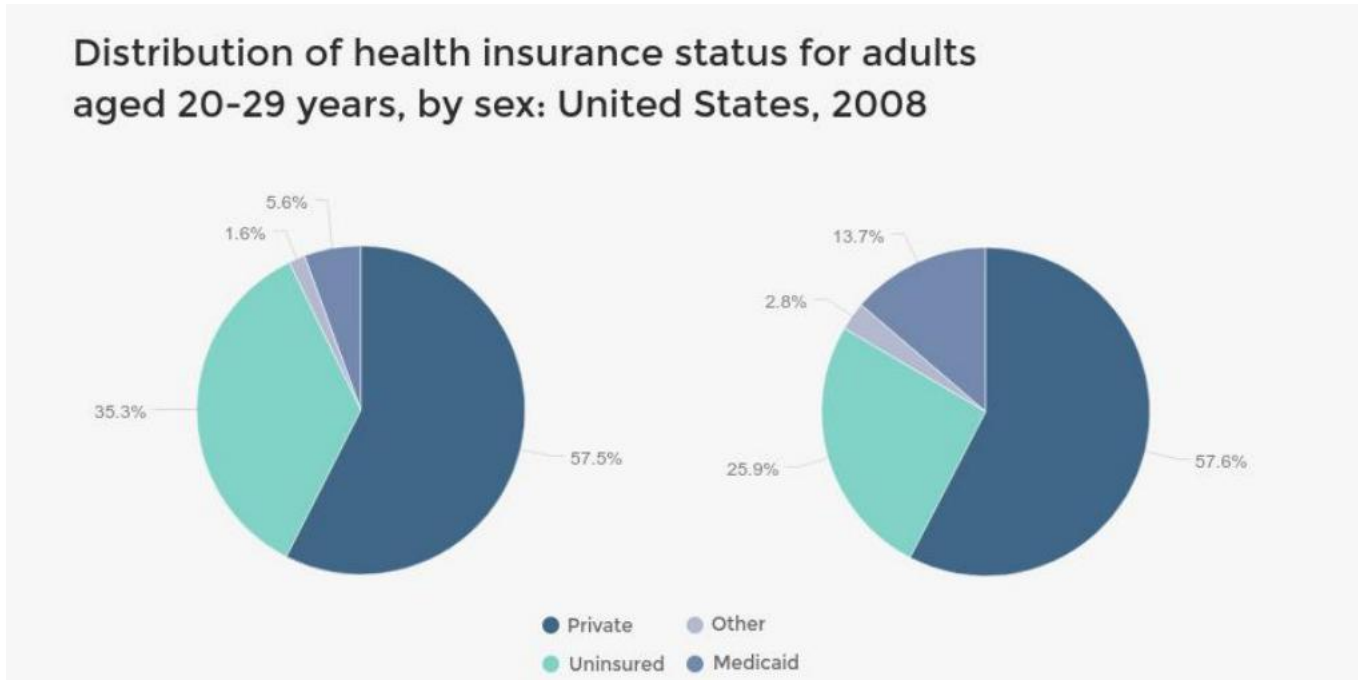
# ΓΡΑΦΗΜΑ

Γραφήματα πολλαπλών γραμμών



# ΓΡΑΦΗΜΑ

## Γραφήματα πολλαπλών πιτών



# ΓΡΑΦΗΜΑ

## Pictographs

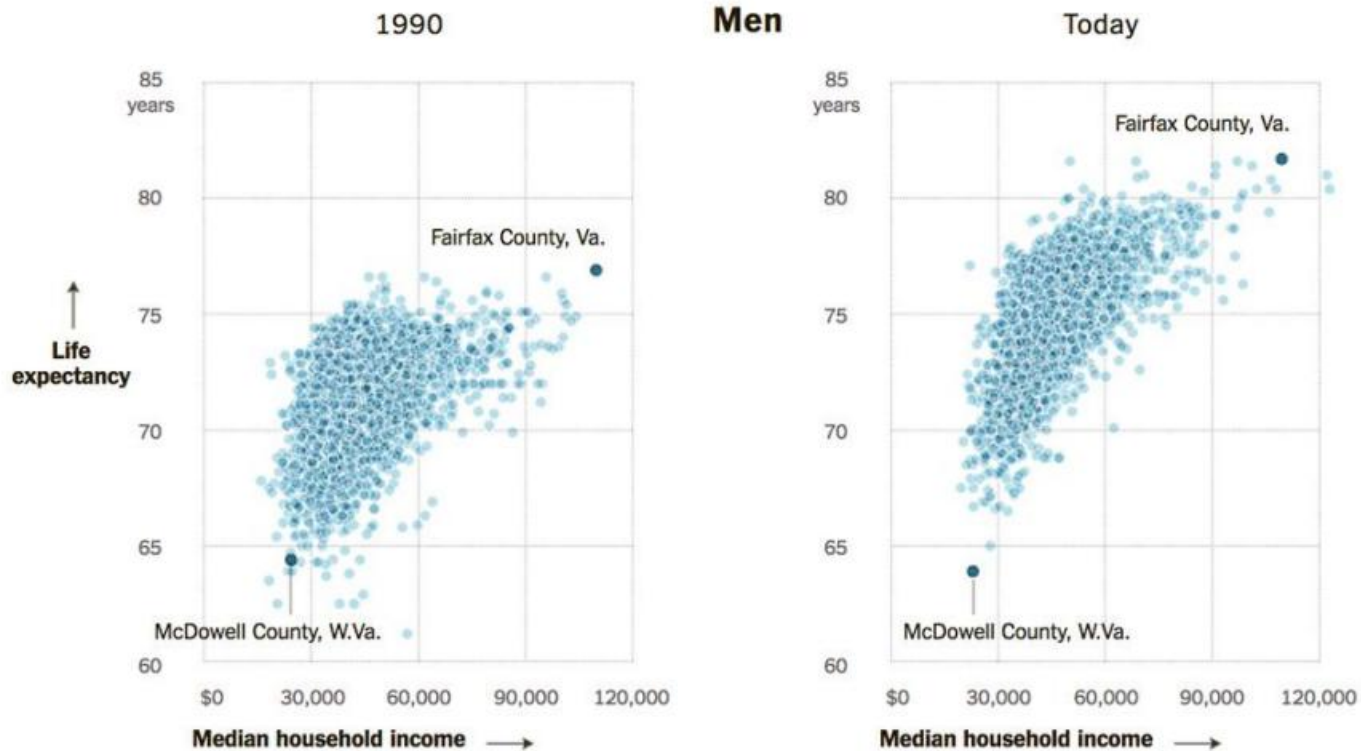


# ΓΡΑΦΗΜΑ

## Γραφήματα διασποράς

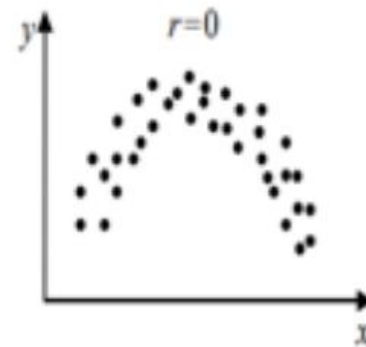
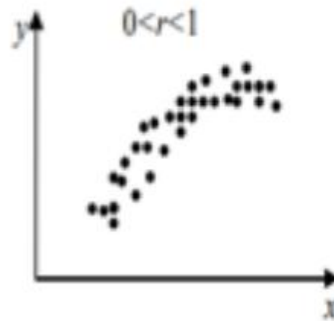
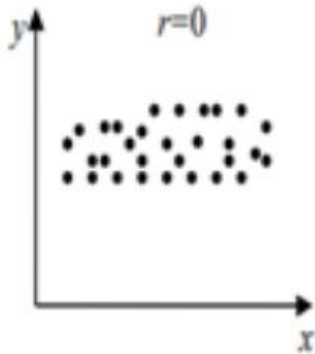
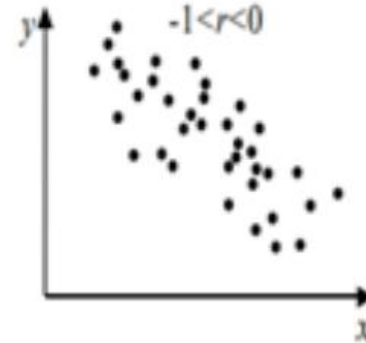
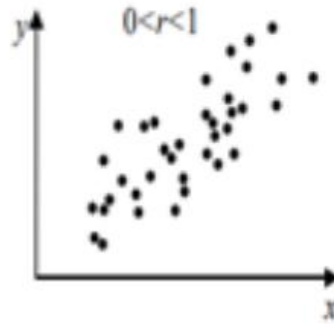
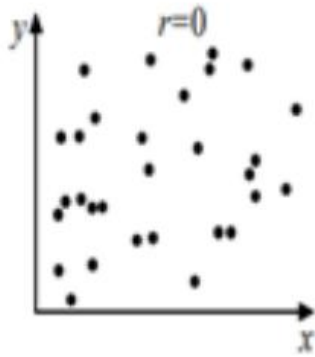
As incomes have diverged between the country's richest counties, like Fairfax County, Va., and its poorest ones, like McDowell County, W.Va., so have the life expectancies of their residents.

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



# ΓΡΑΦΗΜΑ

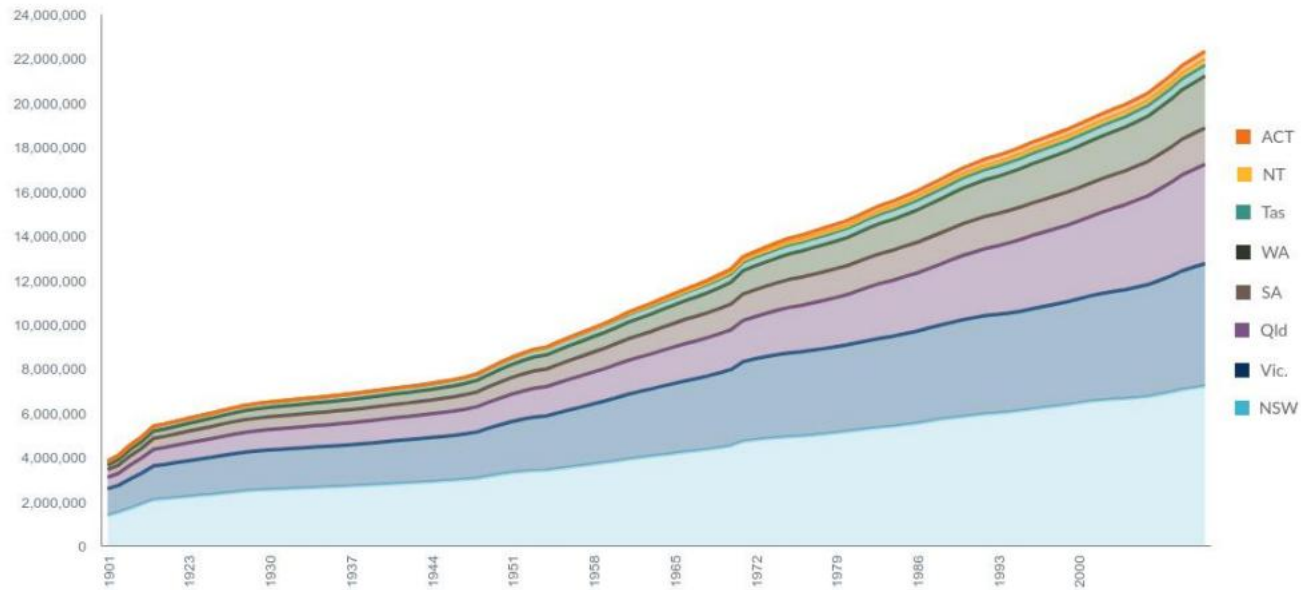
## Γραφήματα διασποράς



# ΓΡΑΦΗΜΑ

## Stacked Area charts

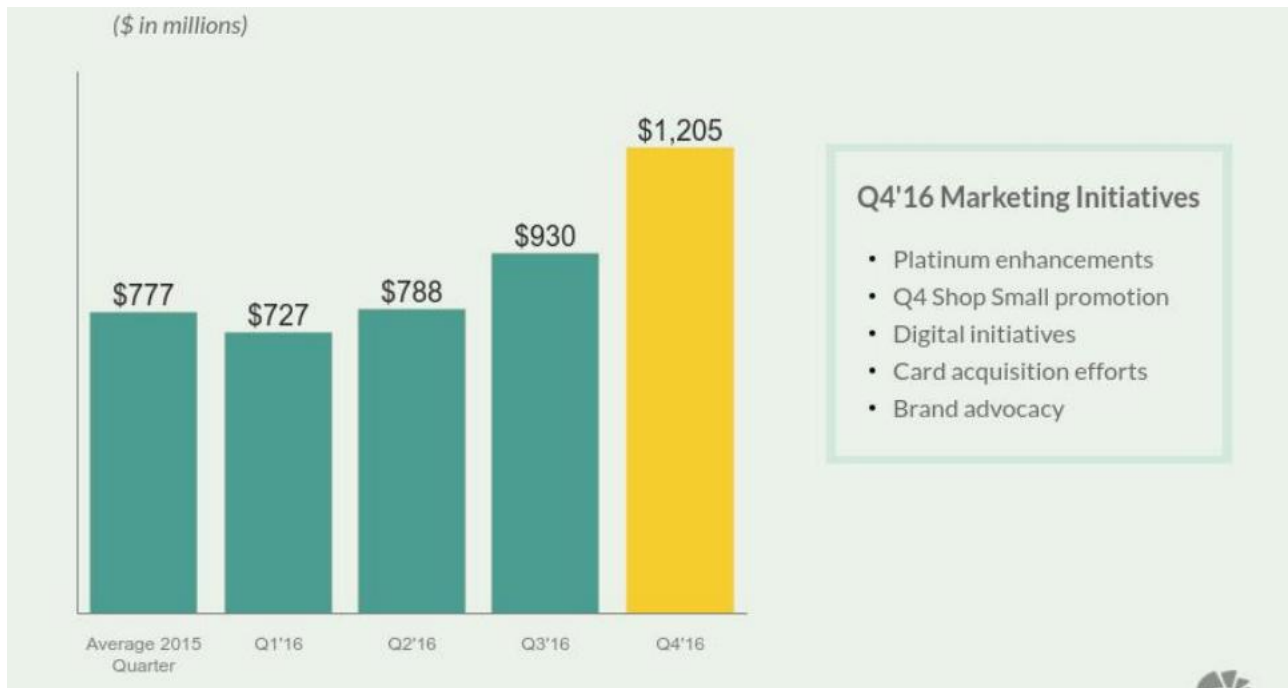
Year/Population (millions)





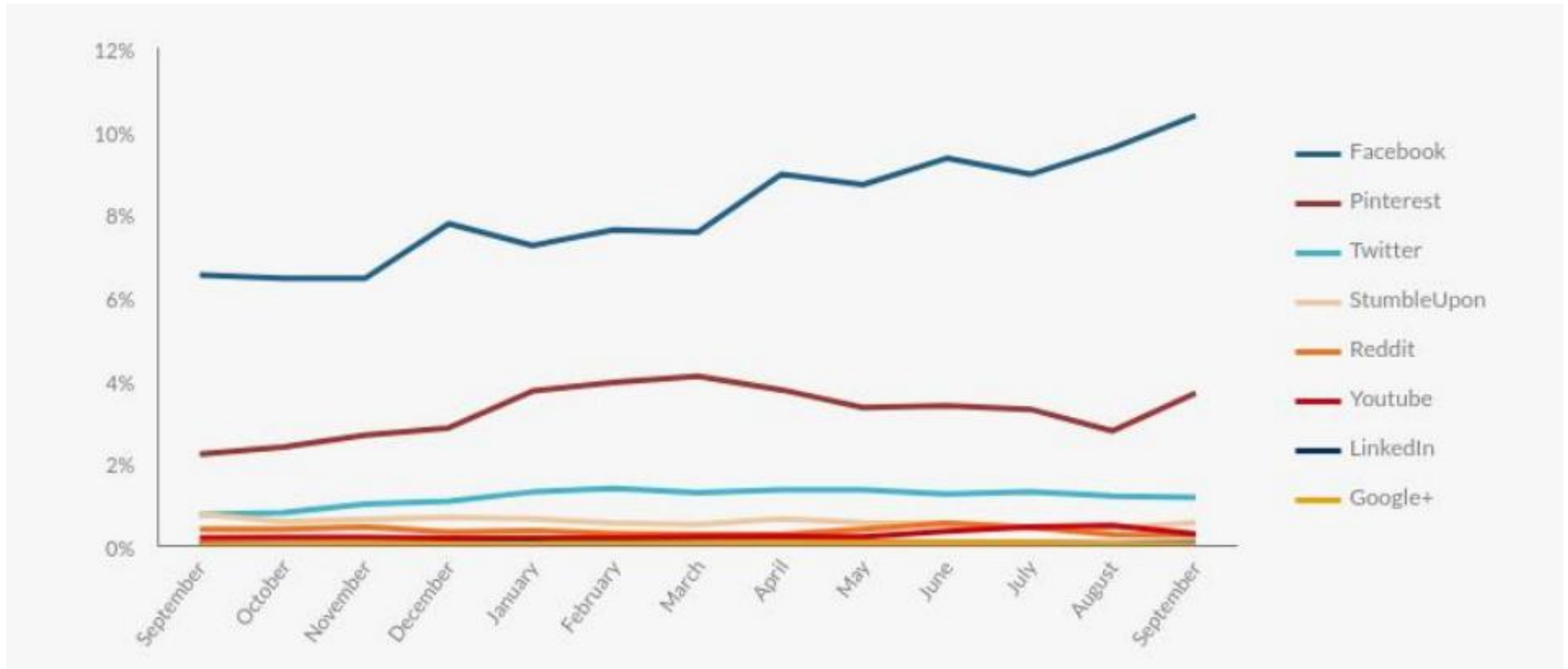
# ΓΡΑΦΗΜΑ

## Column bar graphs



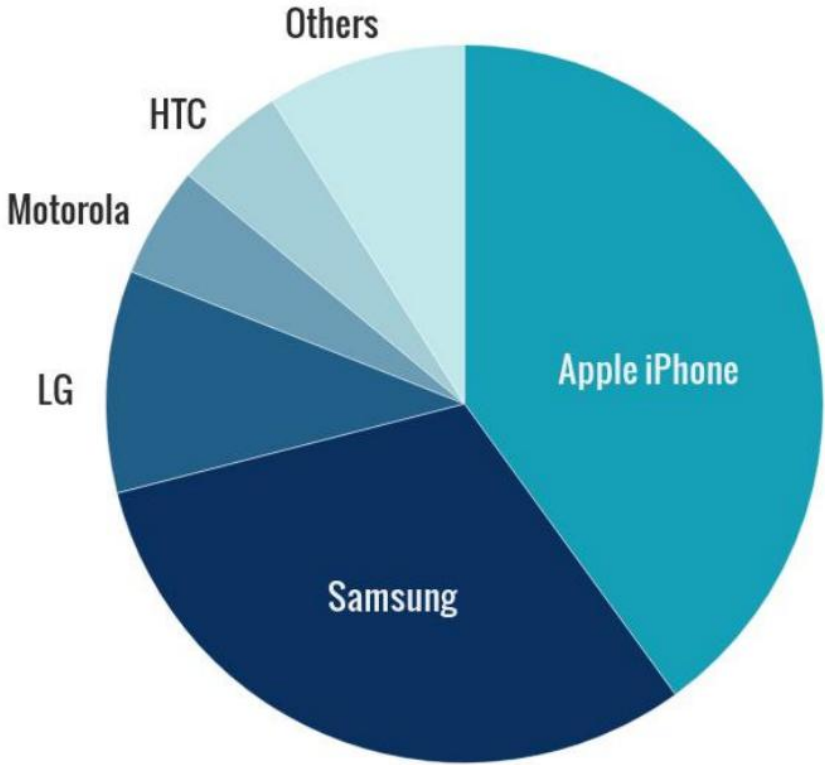
# ΓΡΑΦΗΜΑ

## Line graphs



# ΓΡΑΦΗΜΑ

## Pie charts



# ΓΡΑΦΗΜΑ

## Spider charts

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



# ΓΡΑΦΗΜΑ

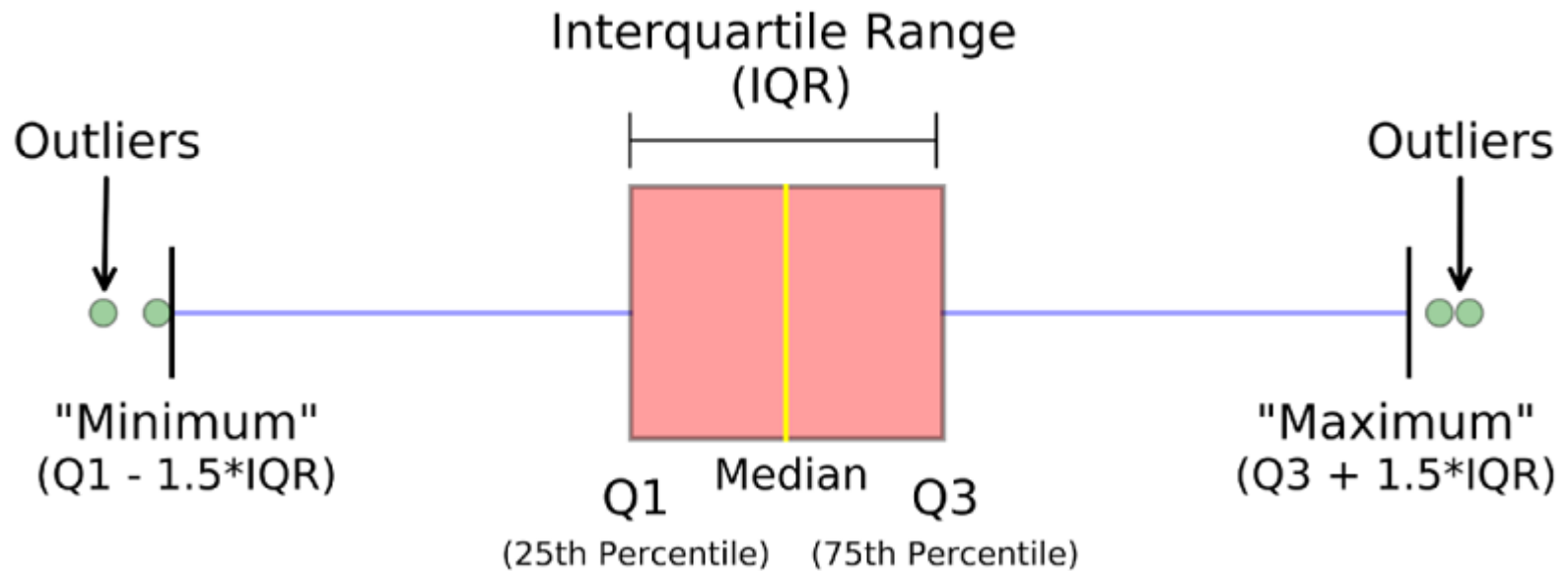
## Διάγραμμα Πλαισίου Απολήξεων/Θηκόγραμμα (Box Plot)

Το Θηκόγραμμα είναι ένα διάγραμμα το οποίο δίνει μία οπτική παρουσίαση της κατανομής των δεδομένων. Εκτός αυτού είναι ένα εξαιρετικό εργαλείο για τη εξαγωγή πληροφοριών σχετικά με τα μέτρα θέσης και μεταβλητότητας ενός συνόλου δεδομένων ή τυχόν διαφορών μεταξύ των μέτρων αυτών όταν πρόκειται για διαφορετικές ομάδες δεδομένων. Οι παράμετροι είναι οι εξής :

- Η ελάχιστη τιμή (min).
- Το πρώτο τεταρτημόριο ( $Q_{25}$  ή  $Q_1$ ).
- Τη διάμεσο ( $Q_{50}$  ή  $Q_2$ ).
- Το τρίτο τεταρτημόριο ( $Q_{75}$  ή  $Q_3$ ).
- Η μέγιστη τιμή (max).

# ΓΡΑΦΗΜΑ

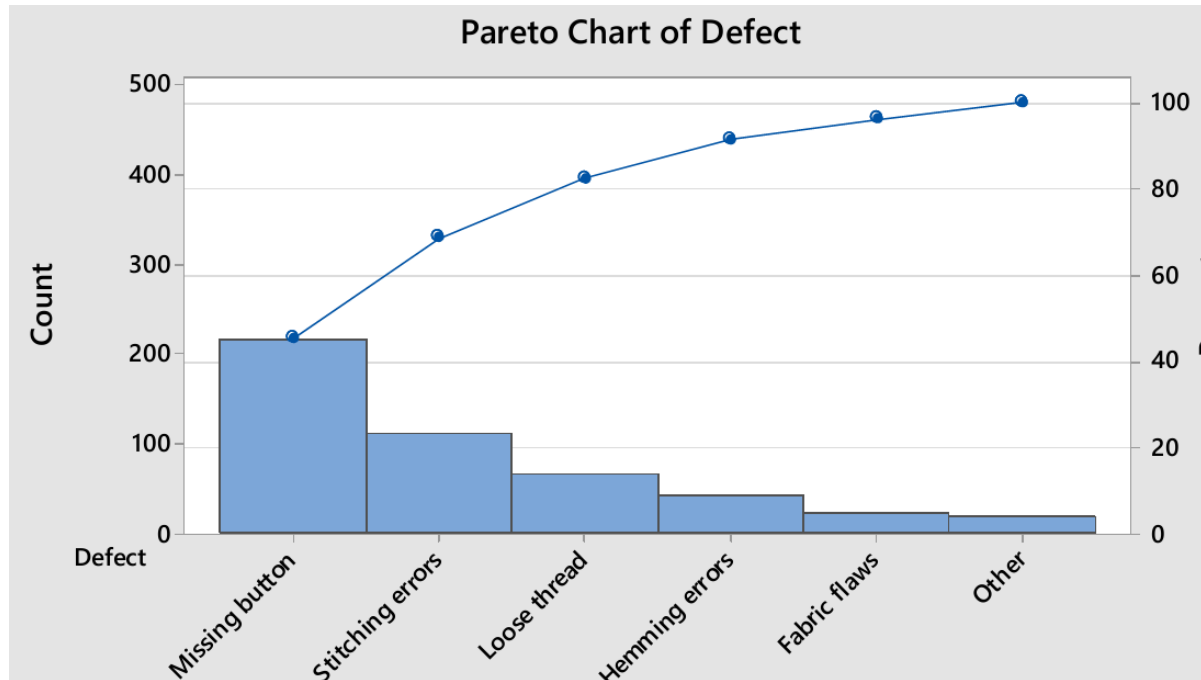
## Διάγραμμα Πλαισίου Απολήξεων/Θηκόγραμμα (Box Plot)



# ΓΡΑΦΗΜΑ

## Διάγραμμα Pareto

Συνήθως, αυτό το γράφημα αποτελείται από μια κάθετη μπάρα και μερικές φορές μια γραφική παράσταση γραμμών και γραμμών. Η κατακόρυφη γραμμή αντιπροσωπεύει τη συχνότητα των ελαττωμάτων και η γραμμή αντιπροσωπεύει ένα σωρευτικό ποσοστό των ελαττωμάτων. Η φθίνουσα σειρά των ράβδων είναι από αριστερά προς τα δεξιά. Η μπάρα στα αριστερά θα έχει το υψηλότερο ύψος και η μπάρα στη δεξιά πλευρά θα έχει το χαμηλότερο ύψος.



# ΓΡΑΦΗΜΑ

## Διαγράμματα αστέρων (Star Plots)

Για κάθε μία παρατήρηση ο ερευνητής κατασκευάζει ένα αστέρι με τόσες ακτίνες όσες είναι και οι μεταβλητές, δηλαδή το μέγεθος κάθε ακτίνας αναπαριστά την τιμή της παρατήρησης για κάποια μεταβλητή.

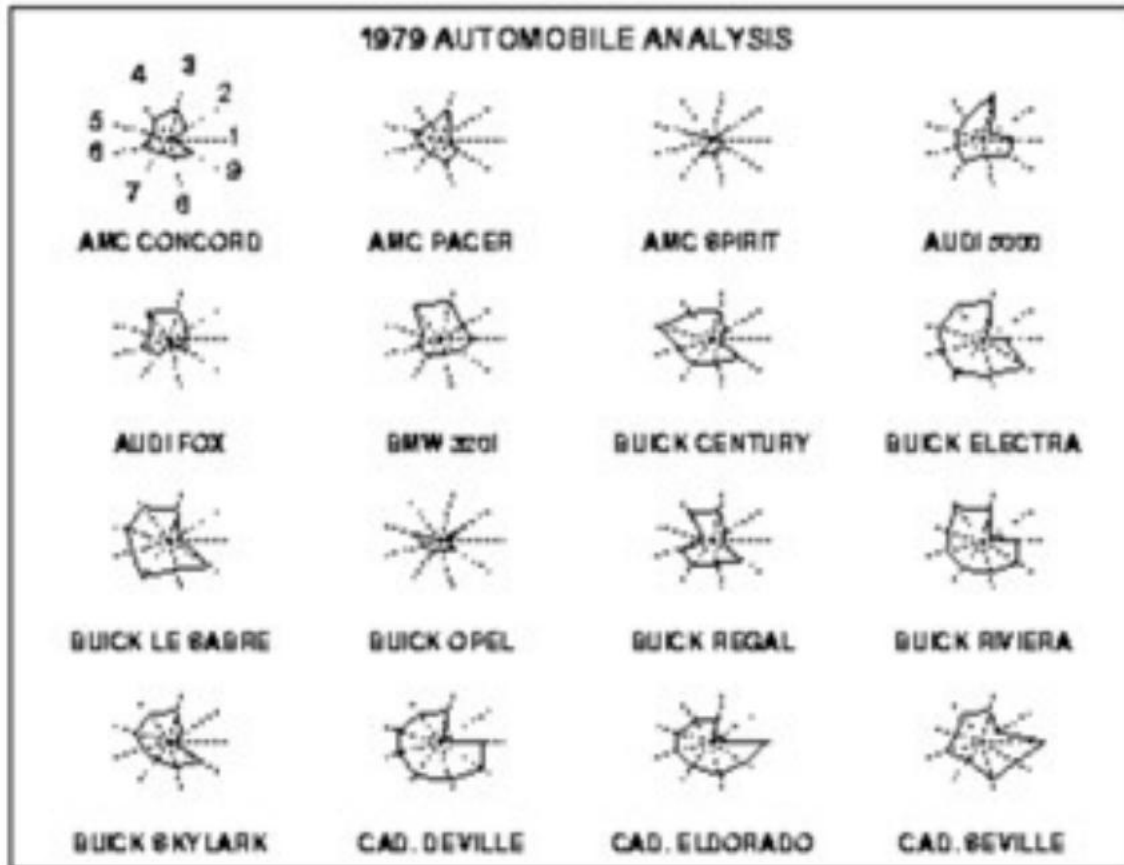
Τα διαγράμματα αστέρων δίνουν απάντηση στις εξής ερωτήσεις :

- Ποιες μεταβλητές κυριαρχούν για μια δεδομένη παρατήρηση;
- Ποιες παρατηρήσεις έχουν κοινά χαρακτηριστικά, δηλαδή, υπάρχουν ομάδες παρατηρήσεων;
- Υπάρχουν ακραίες παρατηρήσεις;



# ΓΡΑΦΗΜΑ

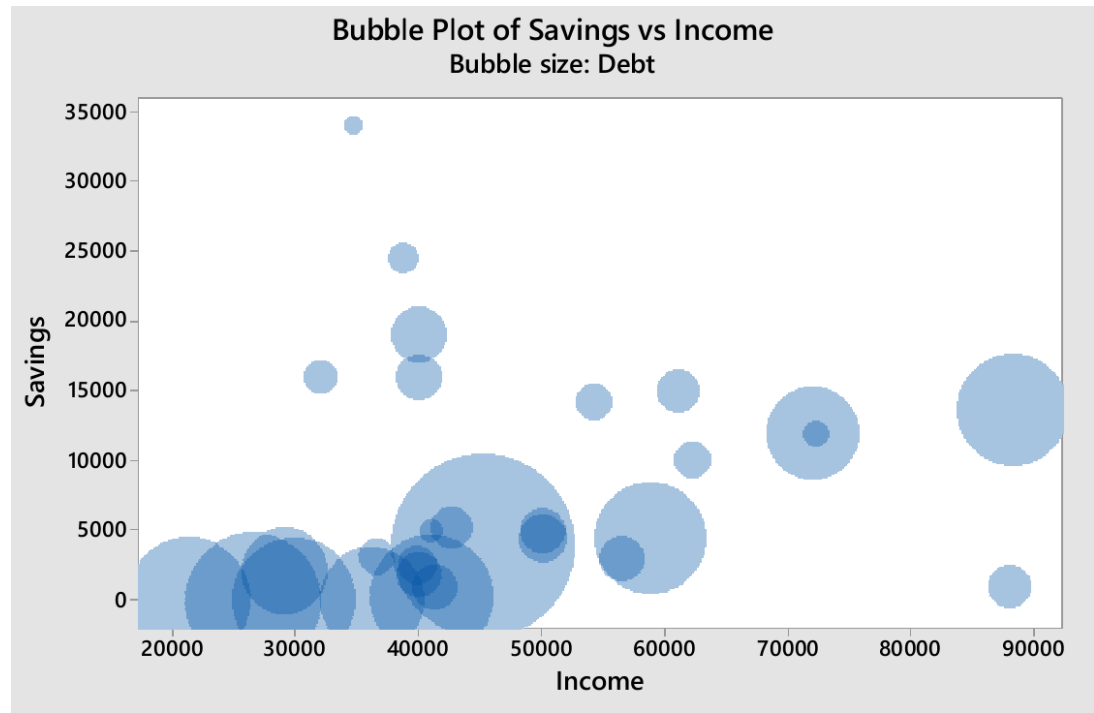
## Διαγράμματα αστέρων (Star Plots)



# ΓΡΑΦΗΜΑ

## Διάγραμμα Φυσαλίδας (Bubble Chart)

Τα διαγράμματα φυσαλίδας δεν είναι παρά απλά διαγράμματα σημείων για τα οποία μία τρίτη μεταβλητή απεικονίζεται ανάλογα με το μέγεθος των κύκλων που αναπαριστούν με κάθε παρατήρηση. Δηλαδή, ενώ στα απλά διαγράμματα σημείων κάθε παρατήρηση συμβολίζεται με ένα σύμβολο ίδιου μεγέθους, για κάθε παρατήρηση, στα bubble charts, χρησιμοποιούμε διαφορετικό μέγεθος ανάλογα με μία τρίτη μεταβλητή. Συνήθως, το σύμβολο είναι κύκλος και για αυτό το γράφημα μοιάζει με σαπουνόφουσκες, από όπου πηγάζει και όνομα.



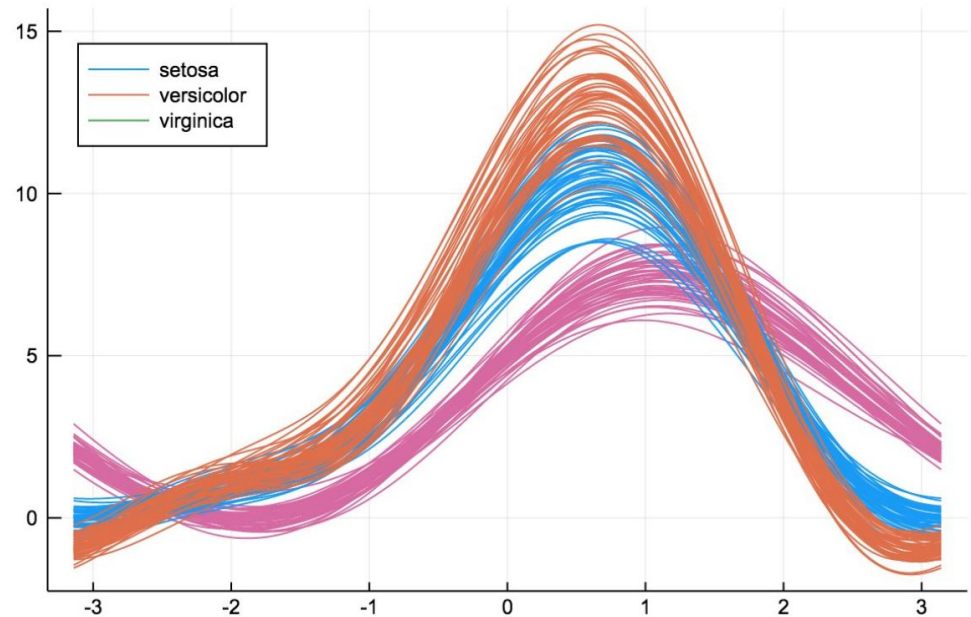
# ΓΡΑΦΗΜΑ

## Οι καμπύλες του Andrews

Για κάθε παρατήρηση σχηματίζουμε την καμπύλη της συνάρτησης

$f(x)(t) = x_1/\sqrt{2} + X_2 \cdot \sin(t) + X_3 \cdot \cos(t) + X_4 \cdot \sin(2t) + X_5 \cdot \cos(2t) + \dots$  όπου  $x_1, x_2, \dots, x_p$  είναι το σύνολο των μεταβλητών και  $t \in [-\pi, \pi]$ .

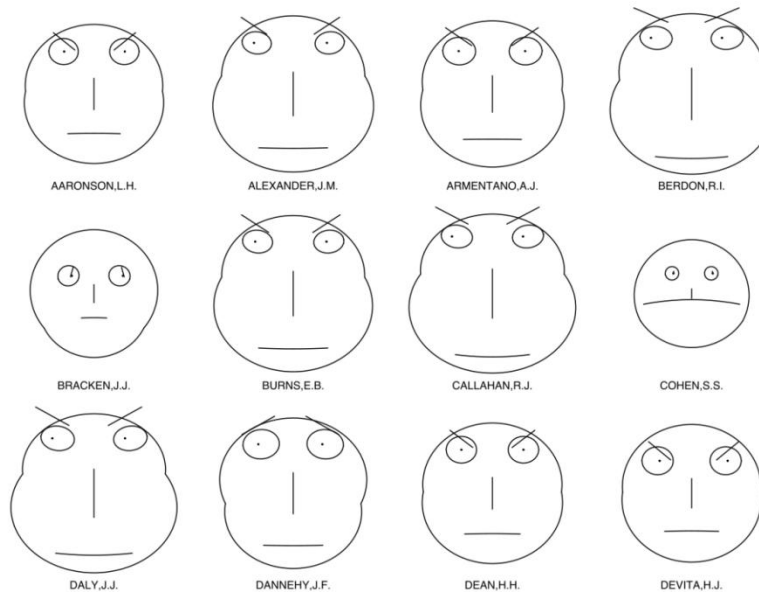
Για διαφορετικές τιμές του  $t$  και στη συνέχεια φτιάχνουμε το γράφημα  $(t, f(t))$  για το διάστημα  $(-\pi, \pi)$ .



# ΓΡΑΦΗΜΑ

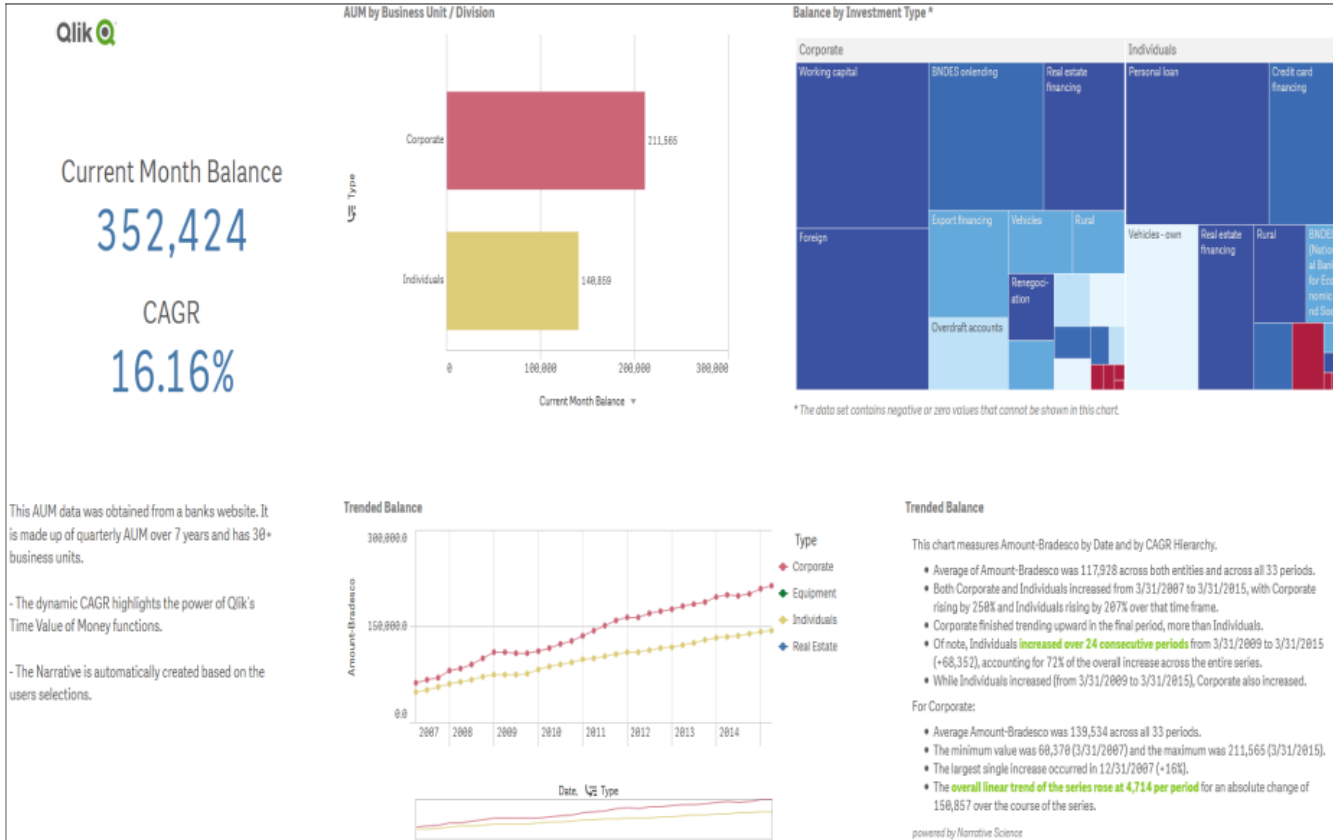
## Τα πρόσωπα του Chernoff

Για κάθε παρατήρηση δημιουργούμε ένα πρόσωπο του οποίου τα χαρακτηριστικά (π.χ μέγεθος, θέση και μέγεθος μύτης και αυτιών κλπ) καθορίζονται από την τιμή της παρατήρησης για κάποια μεταβλητή. Έτσι, ανάλογα με τις τιμές των παρατηρήσεων προκύπτουν διαφορετικά πρόσωπα. Φυσικά ανάλογα με το πόσες μεταβλητές χρησιμοποιούμε, χρειαζόμαστε ανάλογο αριθμό χαρακτηριστικών του προσώπου για να το κατασκευάσουμε.



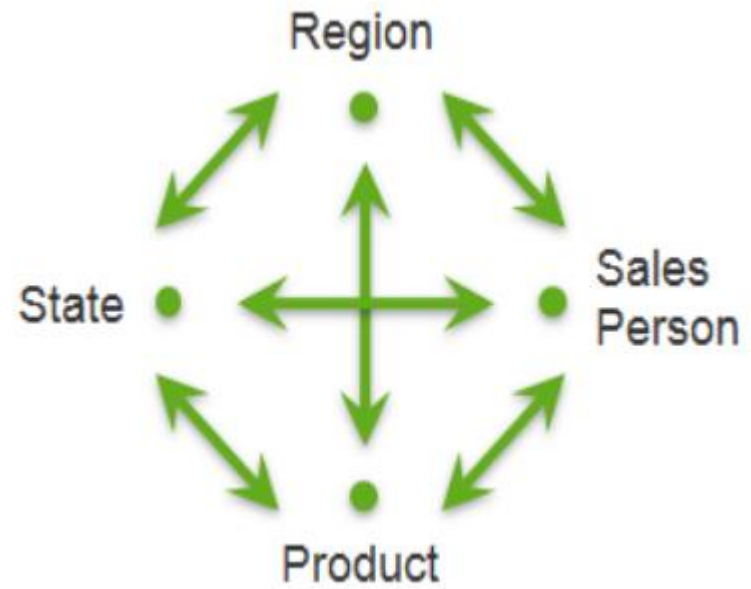
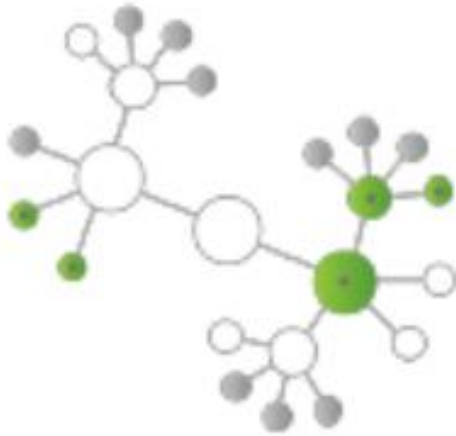
# ΓΡΑΦΗΜΑ

## Ταμπλό (Dashboard)



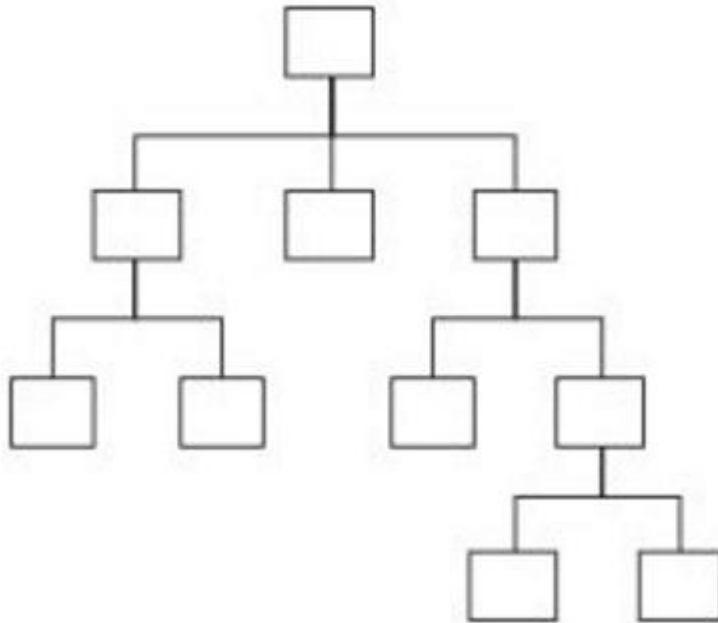
# ΓΡΑΦΗΜΑ

Συσχετίσεις



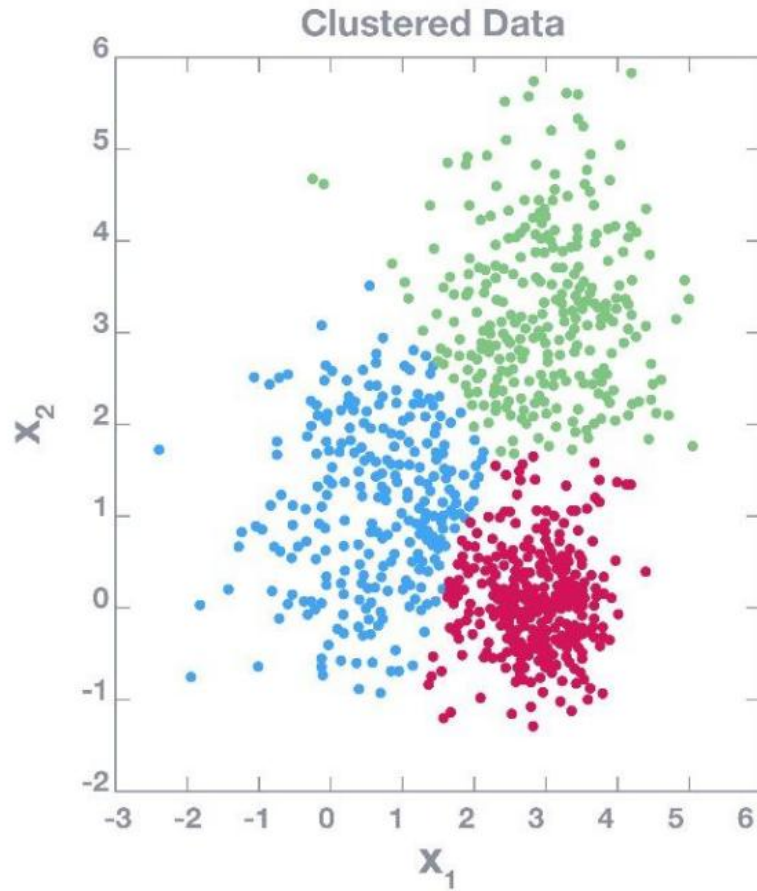
# ΓΡΑΦΗΜΑ

Δενδροχάρτες



# ΓΡΑΦΗΜΑ

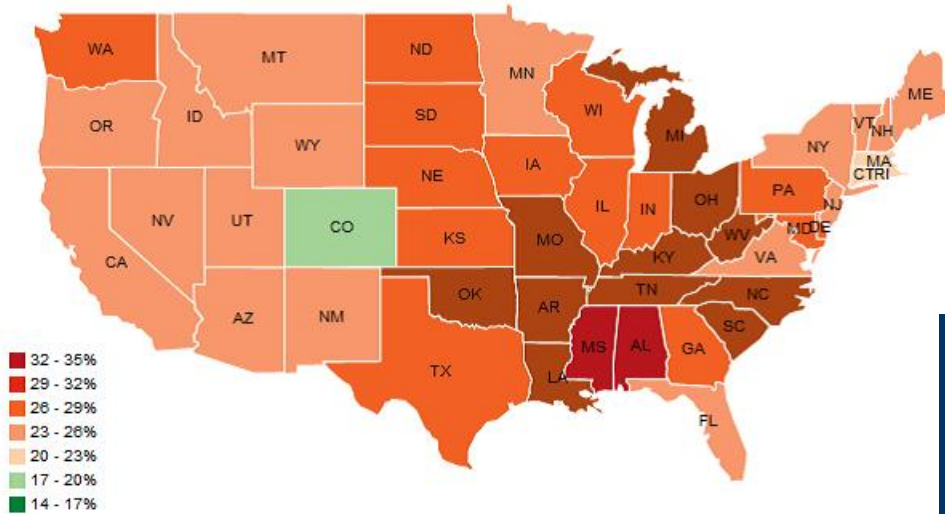
*Παράδειγμα ανάλυσης κατά συστάδες*





# ΓΡΑΦΗΜΑ

## Χάρτες



# ΓΡΑΦΗΜΑ

Χωρικά δεδομένα

