

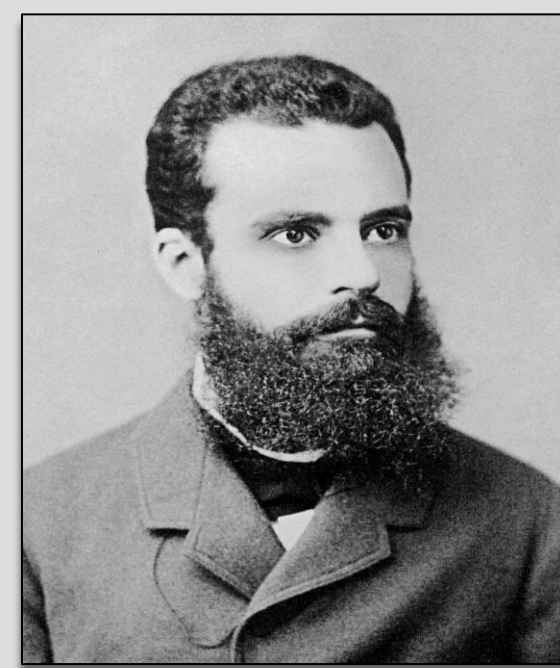
e-ΠΕΡΙΣΚΟΠΙΟ

No.1/2020

Το e-Περίσκοπιο του Εργαστηρίου Στατιστικής και Ανάλυσης Δεδομένων του Πανεπιστημίου Αιγαίου σε συνεργασία με το Ελληνικό Στατιστικό Ινστιτούτο αποτελεί μια πρωτοβουλία των φοιτητών - ερευνητών που το πλαισιώνουν και δεν απευθύνεται αποκλειστικά σε άτομα με στατιστικό υπόβαθρο.

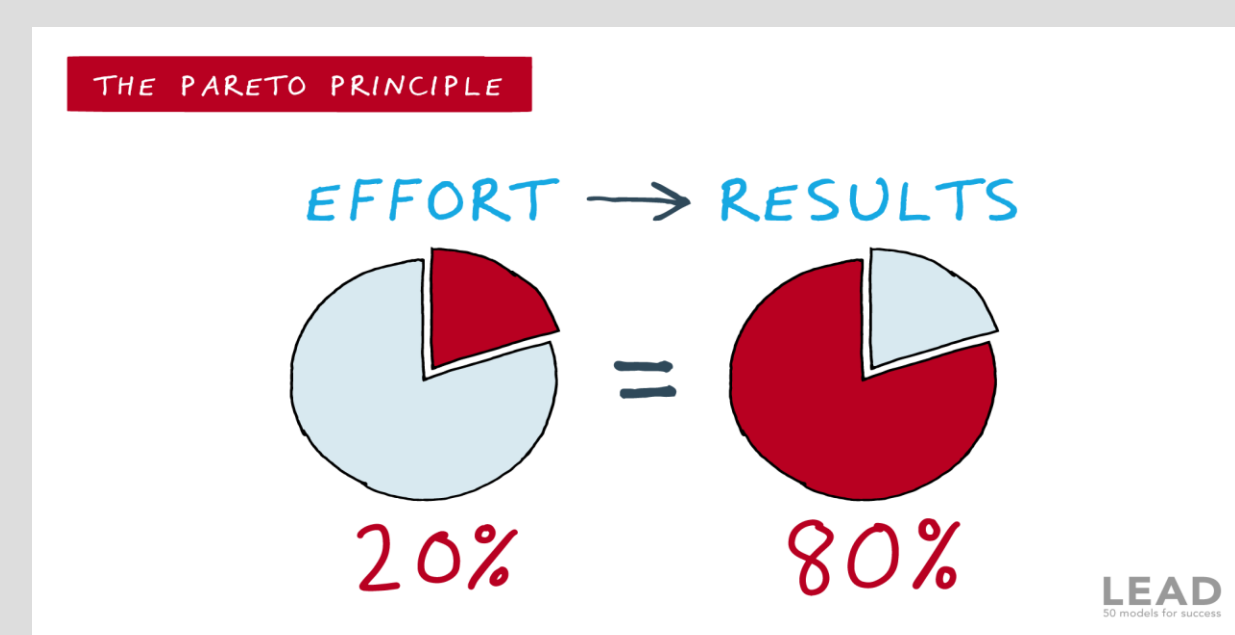
Το κάθε τεύχος είναι ανεξάρτητο των υπολοίπων και ακολουθεί συγκεκριμένη θεματολογία. Συγκεκριμένα απαρτίζεται από: (1) Βιογραφικό σημείωμα ενός ατόμου που έχει συνδεθεί με το υπό ανάλυση θέμα και η συμβολή του στην στατιστική ήταν καθοριστική και θεμελιώδης; (2) Εισαγωγική συζήτηση του θέματος συνοδευόμενη από κάποια funny corners καθώς και ένα quiz / paradox. Στόχος του e-Περίσκοπίου είναι η ενημέρωση, η ψυχαγωγία και ο προβληματισμός των αναγνωστών σε θέματα που έχουν ως κεντρικό άξονα την στατιστική. Αν επιθυμείτε να συμβάλετε στο περιεχόμενο θέτοντας κάποιο θέμα προς ανάλυση, επισκεφτείτε την σελίδα του εργαστηρίου ή/και ελάτε σε επικοινωνία μαζί μας μέσω των πληροφοριών που βρίσκονται στην καρτέλα "ΣΤΟΙΧΕΙΑ ΕΠΙΚΟΙΝΩΝΙΑΣ".

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ



Ο Vilfredo Pareto (1848-1923) ήταν ένας Ιταλικής καταγωγής Μηχανικός, Κοινωνιολόγος, Οικονομολόγος και Φιλόσοφος. Οι καινοτόμες αντιλήψεις του περί οικονομίας, οικονομετρίας καθώς και κοινωνιολογίας συνέβαλαν στη ανάδειξη του ως ενός εκ των κορυφαίων στο πεδίο έρευνας του. Ανάμεσα στα ερευνητικά του επιτεύγματα συγκαταλέγονται τα ακόλουθα:

1. **Αρχή (ή κανόνας) του Pareto** σύμφωνα με την οποία το 80% των αποτελεσμάτων προέρχονται από το 20% των μέσων ή αιτιών.
2. **Διάγραμμα Pareto**, το οποίο αποτελεί ένα στατιστικό εργαλείο που χρησιμοποιείται για την προβολή αιτιών ενός προβλήματος με τη σειρά βαρύτητας από το μεγαλύτερο στο μικρότερο και συνδέεται γραφικά με την αρχή του Pareto.
3. **Κατανομή νόμου δύναμης**, γνωστή και ως Pareto κατανομή, η οποία είναι μια συναρτησιακή σχέση μεταξύ δύο ποσοτήτων, όπου μια ποσότητα μεταβάλλεται ως δύναμη της άλλης.
4. **Δείκτης Pareto**, είναι ο δείκτης μέτρησης της ανισότητας στην κατανομή του εισοδήματος.
5. **Ανάλυση κατά Pareto**, είναι μια τεχνική στατιστικής ανάλυσης που βοηθάει στην σωστή λήψη αποφάσεων για την βελτίωση μιας κατάστασης. Με την ανάλυση Pareto δεν επικεντρωνόμαστε σε όλες τις αιτίες ενός προβλήματος, αλλά μόνο στις πιο σημαντικές.



ΠΑΡΑΔΟΞΟ ΤΩΝ ΓΕΝΕΘΛΙΩΝ

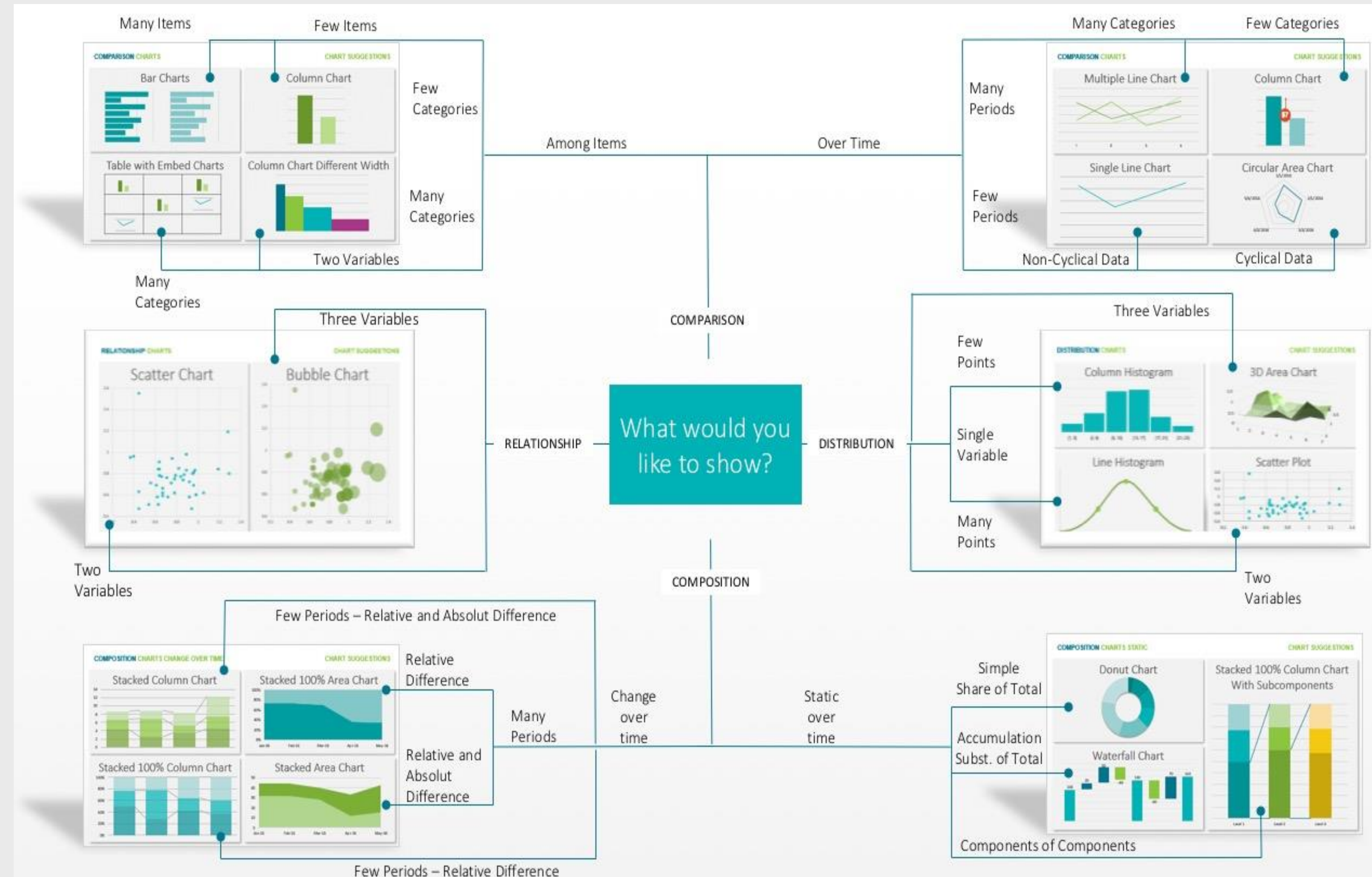
Σε μία ομάδα 23 ατόμων ποια είναι η πιθανότητα δύο από αυτά τα άτομα να έχουν την ίδια ημέρα γενέθλια;

ΑΠΑΝΤΗΣΗ



ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΟΠΤΙΚΟΠΟΙΗΣΗΣ ΔΕΔΟΜΕΝΩΝ

1. **Κατανόηση από το κοινό.** Αναλογισμός για το αν είναι πιο αποδοτικό ένα σύντομο καλογραμμμένο report ή μια εις βάθος ανάλυση του αντικειμένου της παρουσίασης.
2. **Διαμερισμός πολύπλοκων γραφημάτων.** Η διάσπαση ενός περίπλοκου γραφήματος σε μικρότερα κομμάτια μπορεί να είναι το "κλειδί" για αποτελεσματικότερη απεικόνιση των δεδομένων ή της πληροφορίας προς μετάδοση.
3. **Προσθήκη εξεζητηματικών σχολίων.** Μια τέτοια προσθήκη θα βοηθήσει τον αναγνώστη στην καλύτερη κατανόηση και χρήση ενός γραφήματος.
4. **Χρήση colorblind-friendly παλέτα χρωμάτων όπου αυτό είναι εφικτό.** Σύμφωνα με μελέτη του WHO το 2018, τουλάχιστον 2.2 δις ανθρώπων παγκοσμίως έχουν μειωμένη όραση ή τύφλωση. Χρήση τέτοιων παλετών καθιστούν αποδοτικότερη την αφομοίωση των γραφημάτων.
5. **Αποφυγή 3D γραφημάτων** όταν δεν υπάρχει τρίτη μεταβλητή προς απεικόνιση. Στην απόπειρα δημιουργίας ενός "εντυπωσιακού" γραφήματος ενδέχεται πιθανή στρέβλωση της αντίληψης των δεδομένων.
6. **Τοποθέτηση ετικετών** με ευανάγνωστο τρόπο. Ετικέτες π.χ. με συντομογραφίες ή πλάγια γραμματοσειρά καθιστούν δυσκολότερη την κατανόηση του γραφήματος.



"A PICTURE IS WORTH A THOUSAND WORDS"

Ως γράφημα (ή διάγραμμα) ορίζεται η διαγραμματική απεικόνιση ενός συνόλου δεδομένων. Πολύπλοκες και μερικές φορές πολλαπλές ιδέες μπορούν να μεταφερθούν μέσω μιας εικόνας αποτελεσματικότερα από οποιαδήποτε λεκτική περιγραφή. Για τον λόγο αυτό τα διαγράμματα αποτελούν, αποτελούν και θα συνεχίσουν να αποτελούν έναν από τους πιο εύστοχους τρόπους παρουσίασης και μετάδοσης μιας πληροφορίας. Τα γραφήματα, αν και στατιστικού περιεχομένου, χρήζουν ευρείας χρήσης σε όλες τις επιστήμες, και όχι μόνο, όταν πρόκειται για ανάλυση δεδομένων. Τα τελευταία χρόνια πολλές μελέτες έχουν υποστηρίξει ότι η (στατιστική) συμπερασματολογία μέσω γραφημάτων, μπορεί να αποδειχθεί περισσότερο αξιόπιστη απ' ό,τι η χρήση ενός στατιστικού test (Kozak and Piepho, <https://doi.org/10.1111/jac.12220>). Επιπλέον, η χρήση τους στις επιχειρήσεις θεωρείται ζωτικής σημασίας για την μετάδοση μιας πληροφορίας ή ενός μηνύματος στους πελάτες τους.

Όμως δεν είναι όλα για όλους! Αν υπάρχουν δεδομένα προς απεικόνιση, πρέπει ο χρήστης να βεβαιωθεί ότι χρησιμοποιεί τα σωστά διαγράμματα. Παρόλο που τα δεδομένα ενδέχεται να λειτουργούν με πολλούς τύπους γραφημάτων, εξαρτάται από τον αναλυτή η επιλογή που θα εξασφαλίσει ότι το μήνυμά είναι σαφές και ακριβές. Always remember ότι τα δεδομένα είναι πολύτιμα μόνο αν είναι γνωστός ο τρόπος απεικόνισης και παρουσίασης του περιεχομένου τους. Πριν την δημιουργία ενός γραφήματος είναι σημαντική η κατανόηση του λόγου κατασκευής του. Τα διαγράμματα, οι χάρτες και τα infographics βοηθούν τους ανθρώπους να κατανοούν περίπλοκα δεδομένα, να βρίσκουν μοτίβα, να εντοπίζουν τάσεις και να αφηγούνται μια ιστορία. Σκεφτείτε το μήνυμα που θέλετε να μοιραστείτε με το ακροατήριό σας. Με βάση αυτό το μήνυμα θα πρέπει να επιλέξετε με ποια από τις 4 εναρκτηρίες κατηγορίες θα πρέπει να κινηθείτε. Οι κατηγορίες είναι οι εξής:

1. **Κατανομή (distribution):** Τα διαγράμματα κατανομής χρησιμοποιούνται για να δείξουν πώς κατανέμονται οι μεταβλητές με την πάροδο του χρόνου, συμβάλλοντας στον εντοπισμό των αποδόσεων και των τάσεων.
2. **Σύνδεση (Relationship):** Τα διαγράμματα σύνδεσης χρησιμοποιούνται για να δείξουν μια σύνδεση ή συσχέτιση μεταξύ δύο ή περισσότερων μεταβλητών.

3. **Σύγκριση (Comparison):** Οι πίνακες σύγκρισης χρησιμοποιούνται για να συγκρίνουν ένα ή περισσότερα σύνολα δεδομένων. Μπορούν να συγκρίνουν στοιχεία ή να εμφανίσουν διαφορές με την πάροδο του χρόνου.

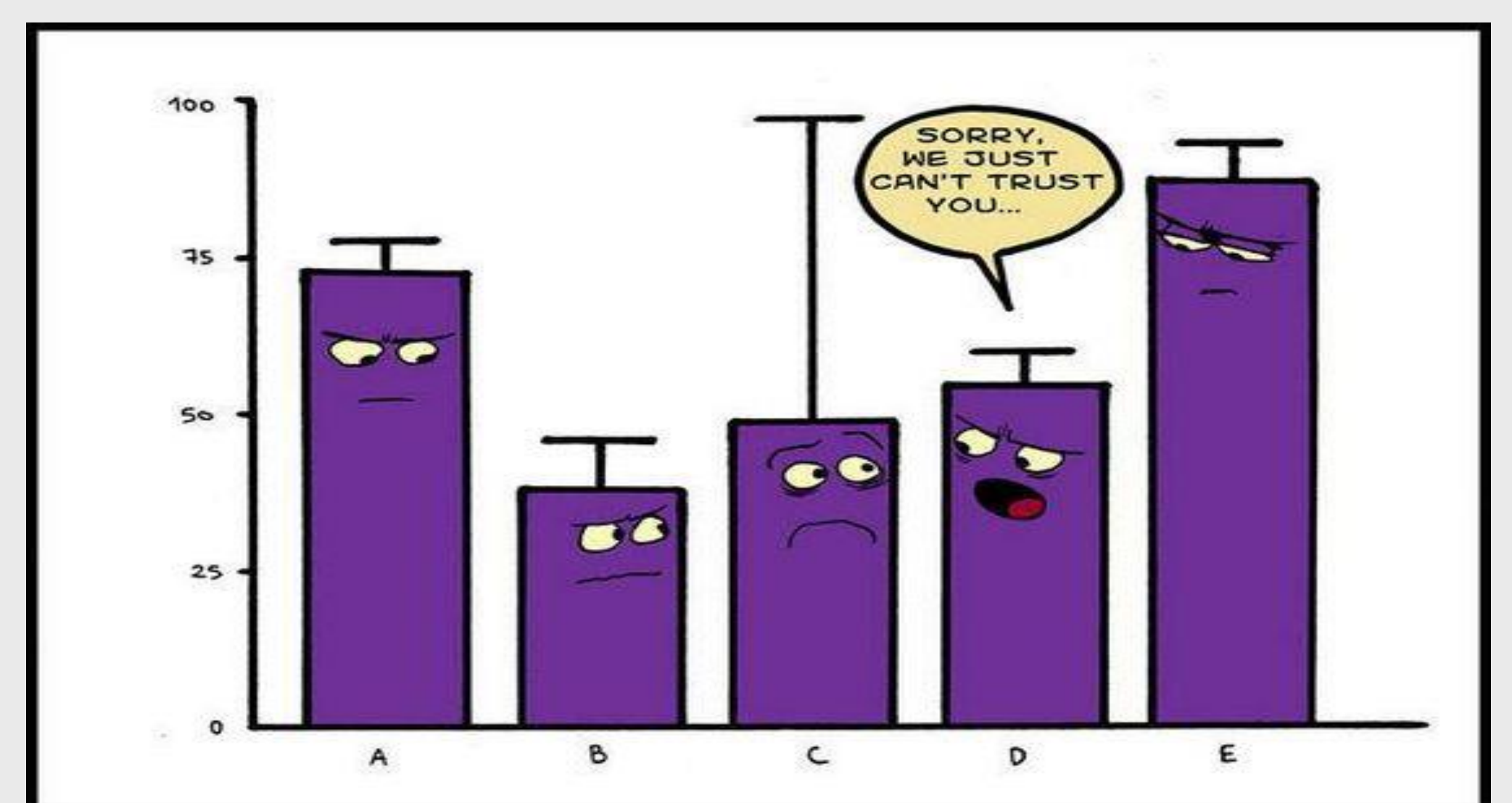
4. **Σύνθεση (Composition):** Τα γραφήματα σύνθεσης χρησιμοποιούνται για την εμφάνιση τμημάτων ενός συνόλου και την εξέλιξή τους με την πάροδο του χρόνου.

Αφού εντοπιστεί η κατάλληλη κατηγορία που ανήκει η πληροφορία προς μετάδοση, οι επιλογές γραφημάτων περιορίζονται στις πιο αποδοτικές για τον προς επίτευξη στόχο. Περαιτέρω ερωτήματα, μεταξύ άλλων, που μπορούν να βοηθήσουν στην λήψη της τελικής απόφασης, είναι "πόσες μεταβλητές θα απεικονιστούν;", "πόσα σημεία δεδομένων υπάρχουν;" και "πώς θα οριστούν οι άξονες;". Έχοντας τις απαντήσεις των παραπάνω ερωτημάτων, θα είναι πλέον ξεκάθαρο στον αναλυτή ποιο είναι το "ιδανικό" γράφημα απεικόνισης που πρέπει να επιλέξει.

The greatest value of a picture is when it forces us to notice what we never expected to see.

-John Tukey

And that's all for the day fellow statistician folks. Until next time take care and remember... The probability to be killed by a cow is low but never zero!



ΣΤΟΙΧΕΙΑ ΕΠΙΚΟΙΝΩΝΙΑΣ

Laboratory of Statistics and Data Analysis

