promise PROMISE Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation

Winter School 2012
Zinal, Valais - Switzerland
23 - 27 January 2012

Number Visualization

Giuseppe Santucci

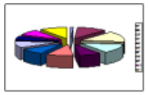
University of Rome "La Sapienza"
santucci@dis.uniroma1.it

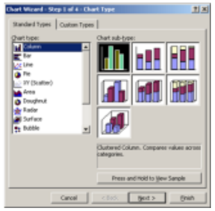
Thanks to:
Ross Ihaka (very inspiring lectures)

Number visualization ?

- Obviously information visualization is, in general, about numbers

- In some cases, however, the numerical part is relevant, and the use of tables, graphs and other visual means to communicate **quantitative** information is commonplace in business today (pie chart, diagrams, boxplots, scatterplots, etc.)





Clustered Column. Compares values across categories.

- Actual software applications allows for easy (?) development of different typologies of charts

- I will discuss the basic relationships and the logical steps that allows for moving from quantitative data to suitable visualizations.

Types of Data

- Quantitative (allows arithmetic operations)
  - *123, 29.56, …*
- Categorical (group, identify & organize; no arithmetic)

  Nominal (name only, no ordering)
  - *Direction: North, East, South, West*

  Ordinal (ordered, not measurable)
  - *First, second, third …*
  - *Hot, warm, cold*

  Interval (starts out as quantitative, but it is made categorical by subdividing into ordered ranges)
  - *0-999, 1000-4999, 5000-9999, 10000-19999, …*

  Hierarchical (successive inclusion)
  - *Region: Continent > Country > State > City*
  - *Animal > Mammal > Horse*
- Relationships
  - Correlation
  - ....

Table and graphs,

- **Table** and **graphs** are widely used to communicate quantitative information
- The goals of presenting quantitative data are
  - Analyzing
  - Monitoring
  - Planning
  - Communicating
- Remember that we are dealing with data that is
  - Quantitative
  - Categorical
- Not all numbers carry quantitative information
  - Categorical intervals
  - IDs (e.g., order number)
- The problem is to map such data to the right visualization, and clear indications about that exists

uhmmm...

- Boring ?

- I do agree !

- I have changed my mind !
- It is plenty of books that teach about quantitative data and how to represent it (see references).
- Read all of them! I'll go for another way...

Outline
(basically what you have NOT to do)

- An introductive example
- Good and bad graphs
  - Basic rules
  - Some additional considerations
- Visual issues
  - Quantitative perception (basic rules)
  - The role of interaction
- Two examples for IR

A lotto game

- Forms of lotto are played world-wide and many people have theories about how to make money at the game

- User task ? ---> Money !!!

- We will examine a particular lotto game, to see whether it might be possible to play it profitably

- The game we’ll look at is the daily pick-it lottery run by the state of New Jersey in the USA

Lotto rules

- Each player selects a number between 000 and 999

- A winning number is selected by independently picking three digits between 0 and 9 at random

- All players that hold the winning number split the prize money for the game

Available data

- The results of the games (winning number and winning amount) are publicly available

- Does this data contain information which will enable us to choose a profitable strategy for this game?

- We will use the results of 254 consecutive games to look for a profitable strategy

The data (254 values)

(winning number, winning amount)

- (810, $190.0), (156, $120.5), (140, $285.5), (542, $184.0), (507, $384.5),
- (972, $324.5), (431, $114.0), (981, $506.5), (865, $290.0), (499, $869.5),
- (020, $668.5), (123, $83.0), (356, $188.0), (015, $449.0), (011, $289.5),
- (160, $212.0), (507, $466.0), (779, $548.5), (286, $260.0), (268, $300.5),
- (698, $556.5), (640, $371.5), (136, $112.5), (854, $254.5), (069, $368.0),
- (199, $510.0), (413, $102.0), (192, $206.5), (602, $261.5), (987, $361.0),
- (112, $167.5), (245, $187.0), (174, $146.5), (913, $205.0), (828, $348.5),
- (539, $283.5), (434, $447.0), (357, $102.5), (178, $219.0), (198, $292.5),
- (406, $343.0), (079, $332.5), (034, $532.5), (089, $445.5), (257, $127.0),
- (662, $557.5), (524, $203.5), (809, $373.5), (527, $142.0), (257, $230.5),
- (008, $482.5), (446, $512.5), (440, $330.0), (781, $273.0), (615, $171.0),
- (231, $178.0), (580, $463.5), (987, $476.0), (391, $290.0), (267, $176.0),
- (808, $195.0), (258, $159.5), (479, $296.0), (516, $177.5), (964, $406.0),
- (742, $182.0), (537, $164.5), (275, $137.0), (112, $191.0), (230, $298.0),
- (310, $110.0), (335, $353.0), (238, $192.5), (294, $308.5), (854, $287.0),
- (309, $203.5), (026, $377.5), (960, $211.5), (200, $342.0), (604, $259.0),
- (841, $231.0), (659, $348.0), (735, $159.0), (105, $130.5), (254, $176.0),
- (117, $128.5), (751, $159.0), (781, $290.0), (937, $335.0), (020, $514.0),
- (348, $191.0), (653, $304.5), (410, $167.0), (468, $257.0), (077, $640.0),
- (921, $142.0), (314, $146.0), (683, $356.0), (000, $96.0), (963, $295.0),

Visualizing the data

- Humans can really only make sense of three or four numbers at a time

- By representing the values in a graphical form we make it easier to handle large numbers of values

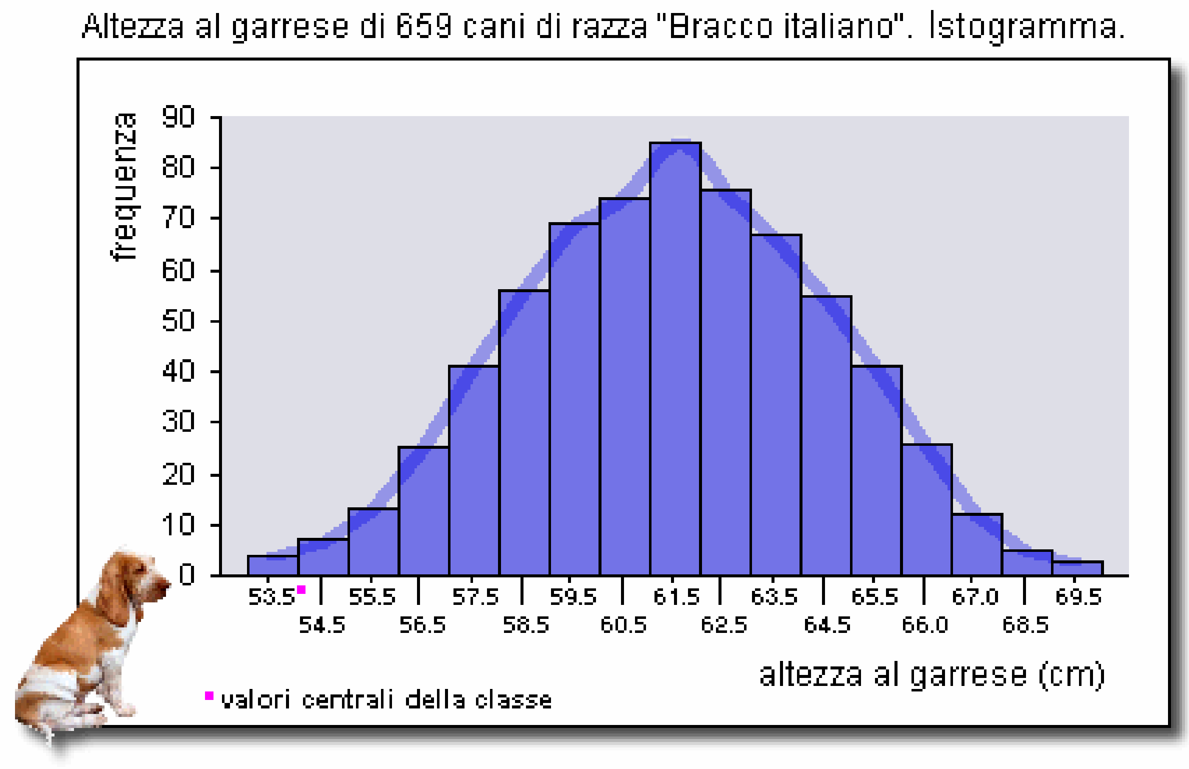
- Using visualizations should make it possible to learn more about this data

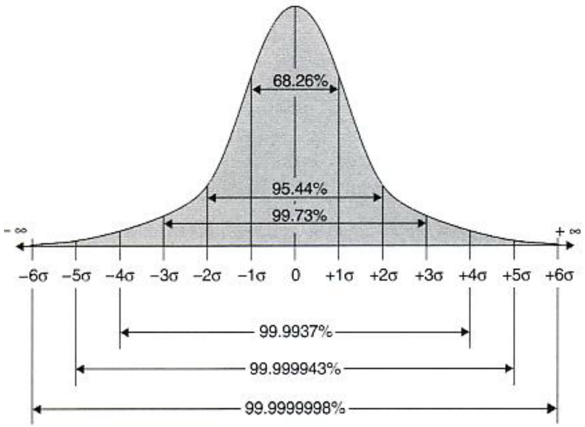
- We have NOT to **lie** or make **noise** !!!

User task and visualization

- One approach to making money at “Pick It” is to try to select numbers which are more likely to win
- We can look at the distribution of the winning numbers to see whether some ranges of values are more like to produce a winner than others
- One way to do this is to produce a histogram of the winning numbers

Histogram example

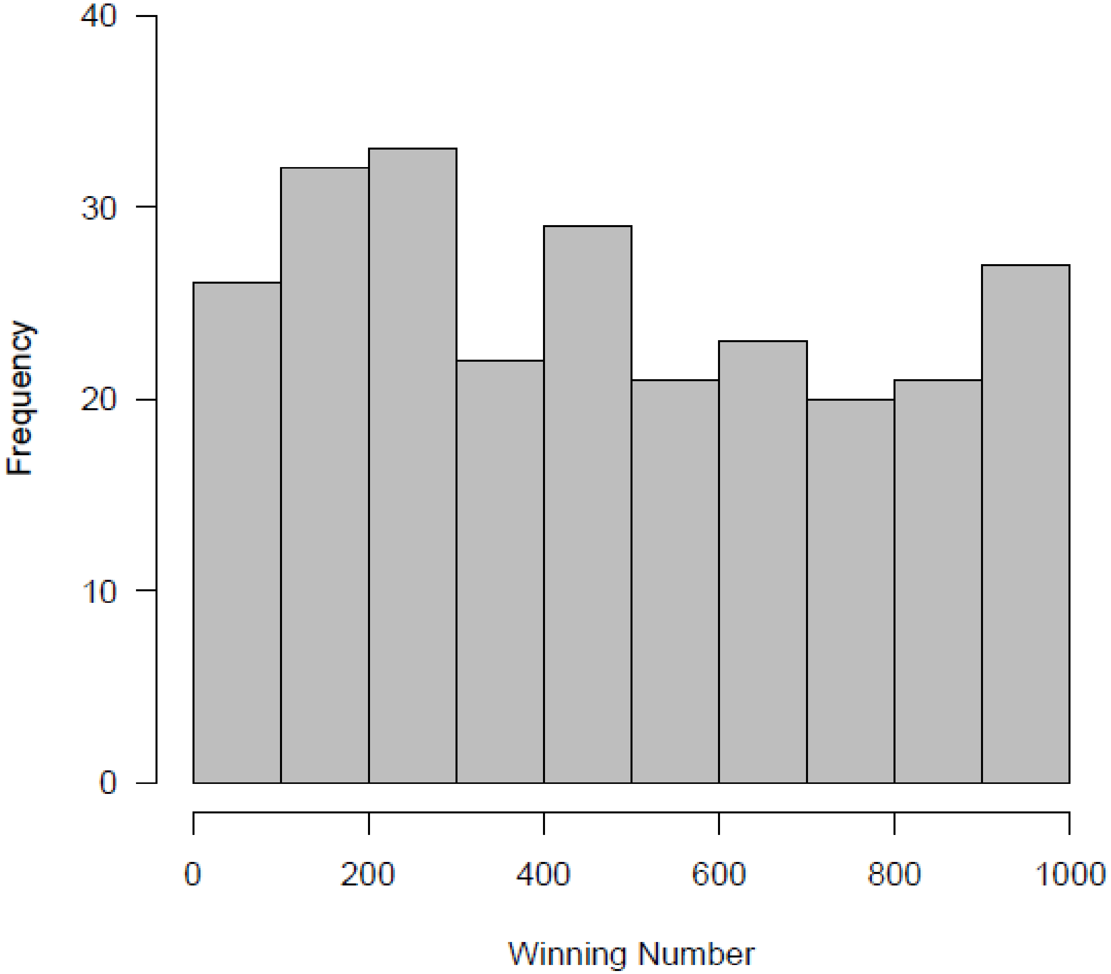

Altezza al garrese di 659 cani di razza "Bracco italiano". Istogramma.



bin

Winter School 2012
Zinal, Valais - Switzerland
23 - 27 January 2012

Data distribution

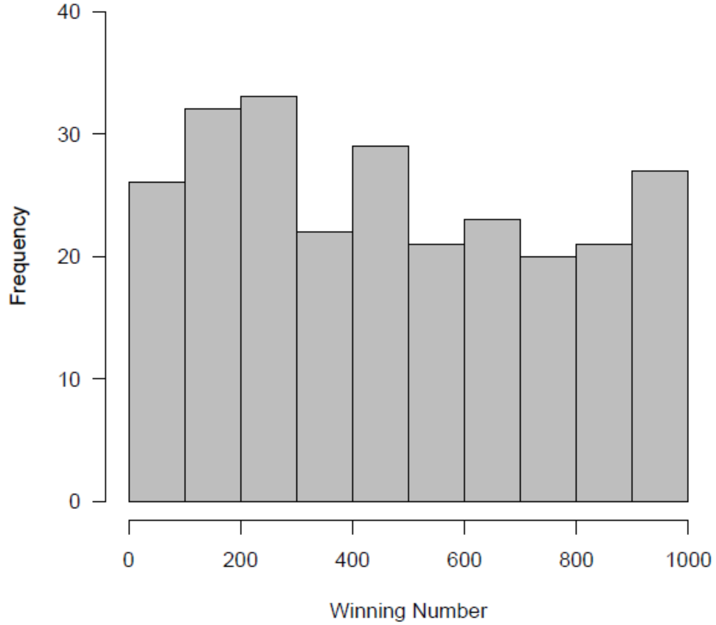


What can we infer from this histogram?

(Is the bin size ok?)

Analysis

- It looks there tend to be more winners in the region from 100 to 300 than in other regions
- This suggests that we might be best to choose numbers in this range



Do you agree ?

Winter School 2012
Zinal, Valais - Switzerland
23 - 27 January 2012

We are telling lies...

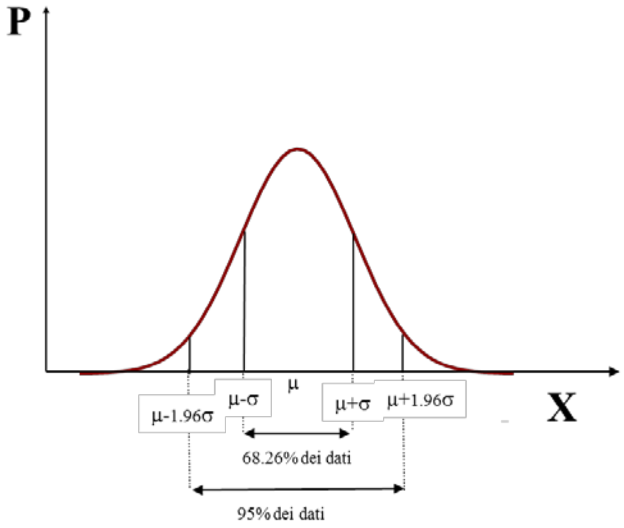
(wrong number understanding)

- Even if the winning numbers are chosen randomly, we can expect some “random variability” in a sample

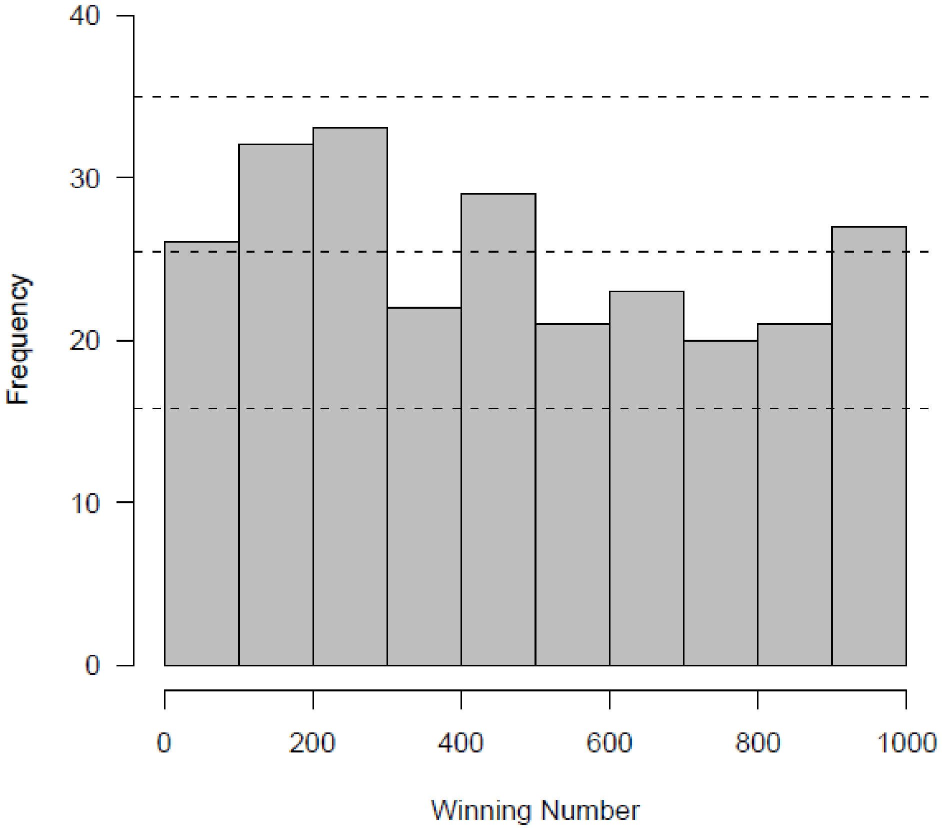
- To judge the significance of what we see in the histogram we have to recall some formal statistical theory

The mean is not enough !

- There are 254 values. We would expect the number of values in each cell to be approximately: 25.4 = 254/10
- Such a number is a random variable as well, with normal distribution
- 95% of the observations fall within +/- 2σ



Better number visualization



+ 2σ

mean

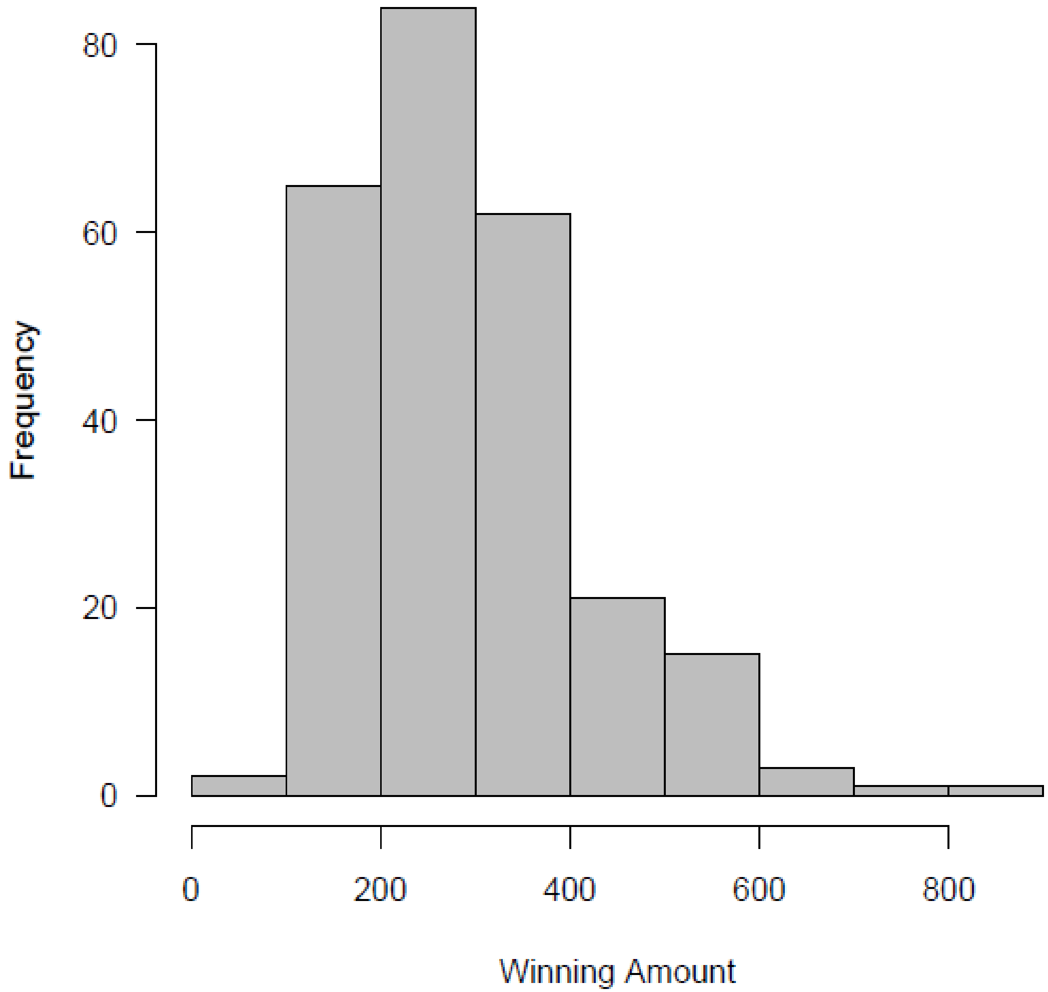
- 2σ

- Variance analysis AND visualization

Conclusions and new task

- Winning numbers are totally random
- It makes no sense to look for a " lucky " number
- However, we can change our task:
  - to increase the amount won !
- So we study the distribution of winning amount

New visualization



Looking for new insights

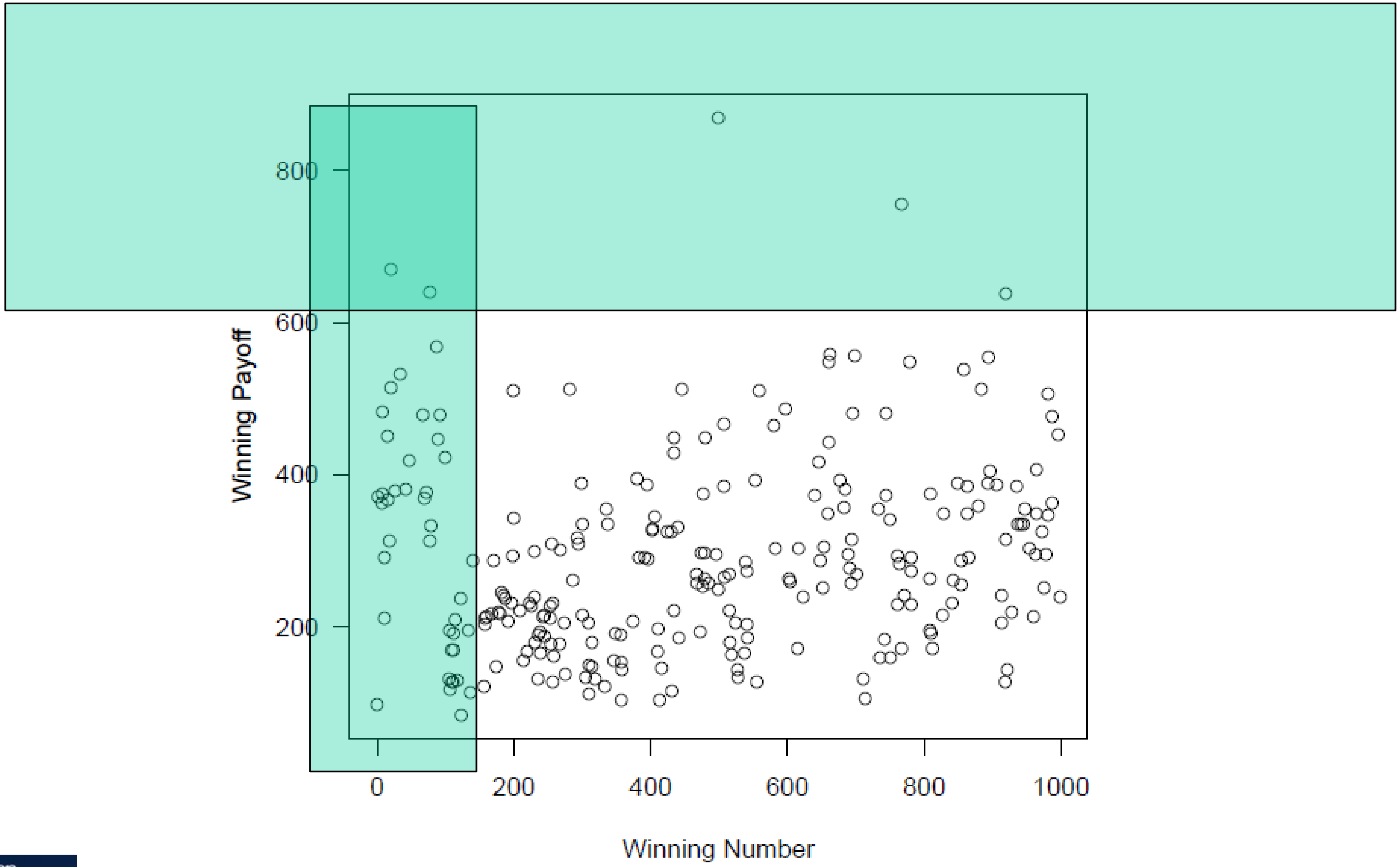
- The histogram shows that there is a wide (more than 2σ) range amounts won in the game

- It *might* be possible to choose the numbers which win larger amounts

- We search for relationship between ticket number and winning amount

- A scatter plot is the natural way to look for such a relationship.

New visualization

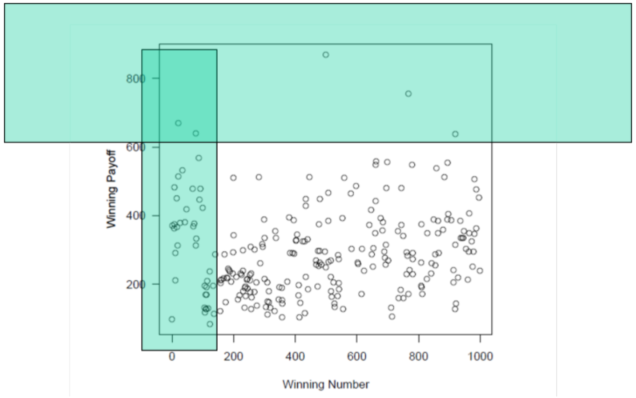


Winter School 2012
Zinal, Valais - Switzerland
23 - 27 January 2012

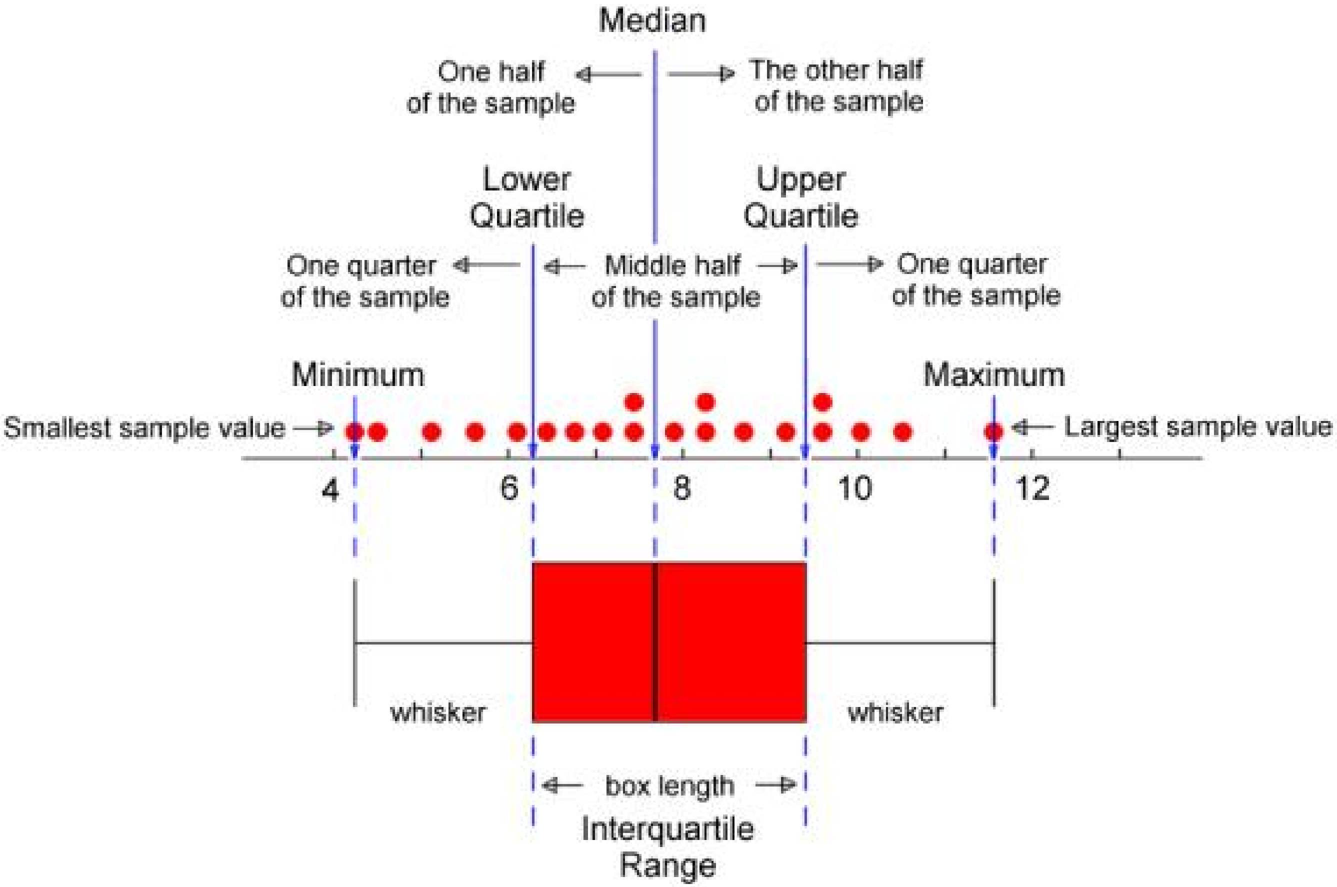
Insights from the scatterplot

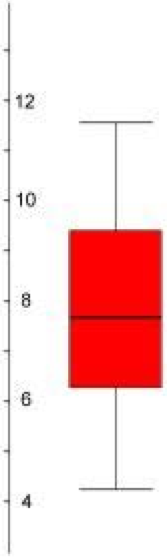
- The winning amounts in a band to the left of the plot appear to generally be higher than those in the rest of the plot

- We can investigate this further by separating the numbers into groups according to the first digit of the ticket number and drawing box plots for each group

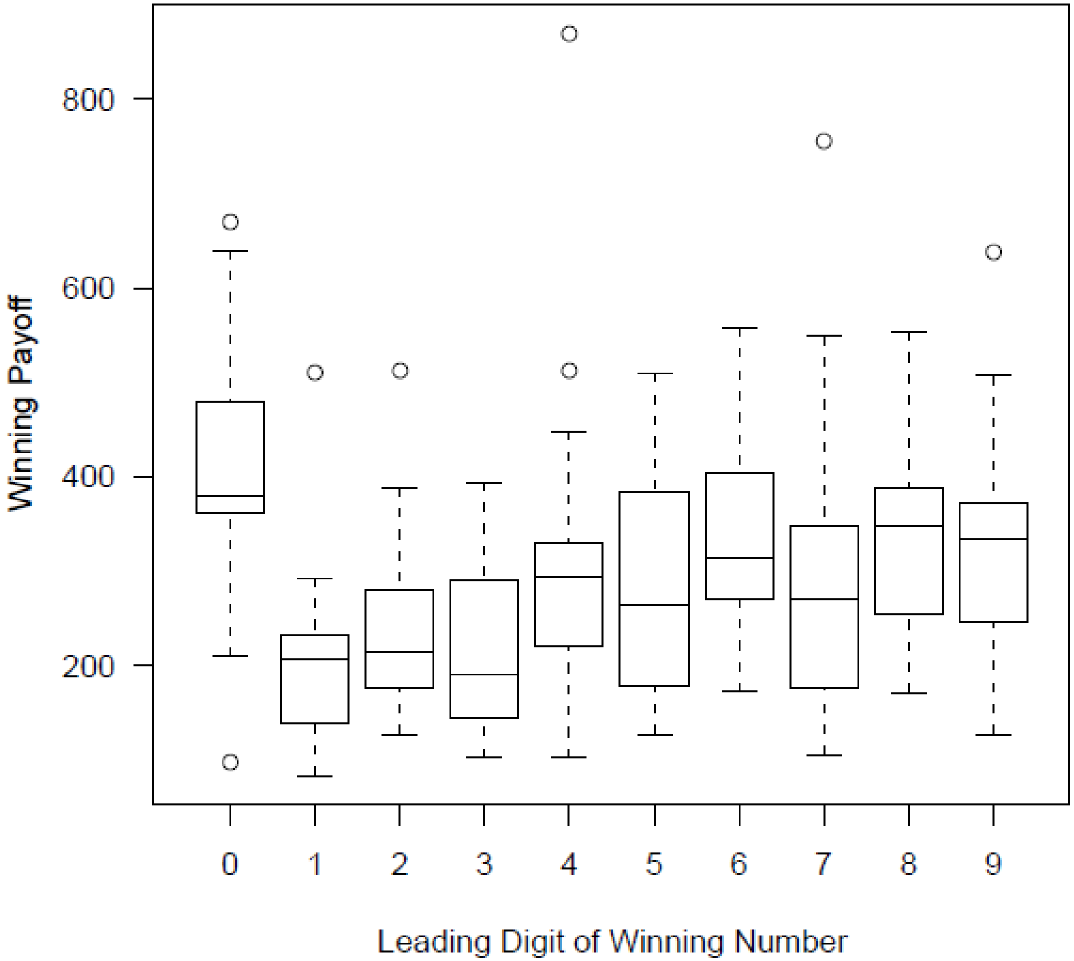


Boxplot


Median
One half of the sample
The other half of the sample
Lower Quartile
Upper Quartile
One quarter of the sample
Middle half of the sample
One quarter of the sample
Minimum
Maximum
Smallest sample value
Largest sample value
4
6
8
10
12
whisker
whisker
box length
Interquartile Range

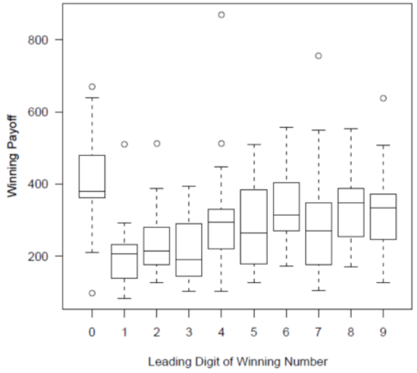

12
10
8
6
4

Lottery's boxplots

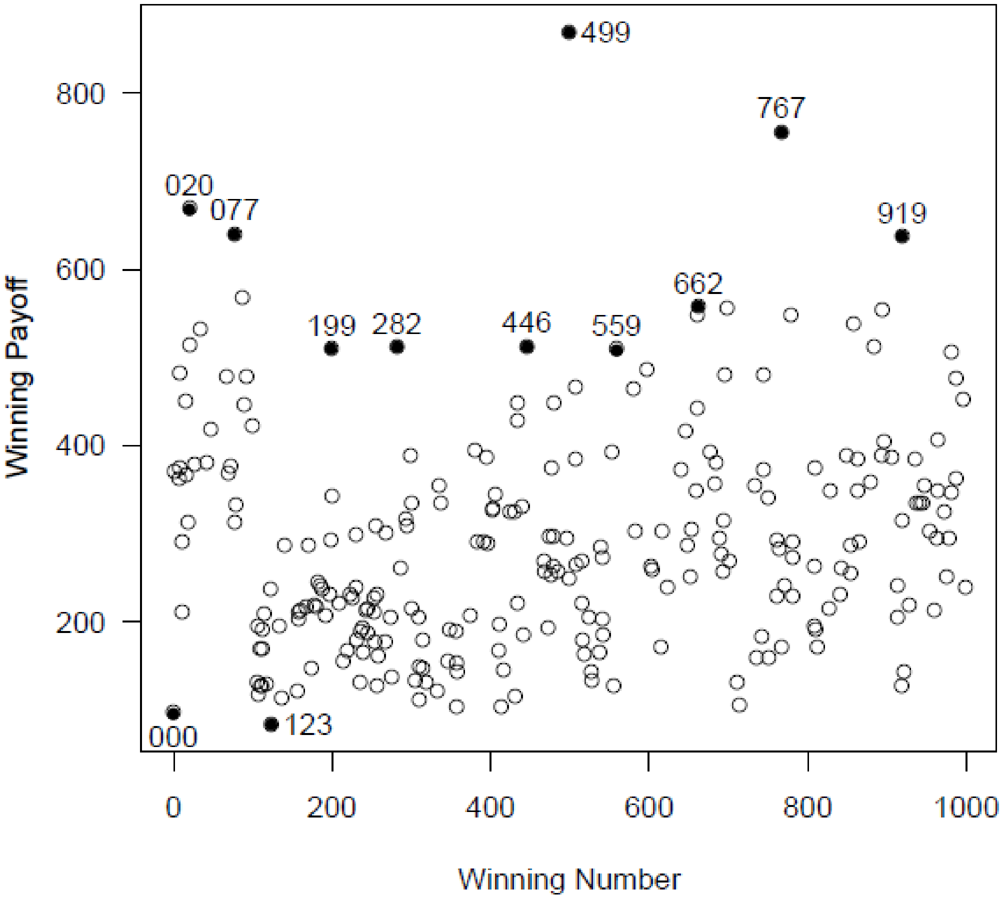


New insights

- Tickets with a leading zero digit clearly tend to produce larger winnings
- It is also apparent that there are some very large and some very small winning amounts
- It is probably of interest to identify the ticket numbers corresponding to these extremes



High and low winning numbers



Winter School 2012
Zinal, Valais - Switzerland
23 - 27 January 2012

Lotto strategy

- While winning numbers are non predictable, players' choices are!

- Choose numbers which are less likely to be chosen by other players
- Then, when you win (if), you will tend to win more
- Possible ways to choose:
  - Choose a number with a leading zero
  - Choose a number with repeated digits
  - Avoid “obvious” numbers like, e.g. 000, 123, 246, . . .

Lessons learned

- Define clearly the task
- Use basic visualizations
  - bar charts
  - scatterplots
  - boxplots
- Be ready to switch among them
- Look for precise values when needed
- Do not lie !

Outline
(basically what you have NOT to do)

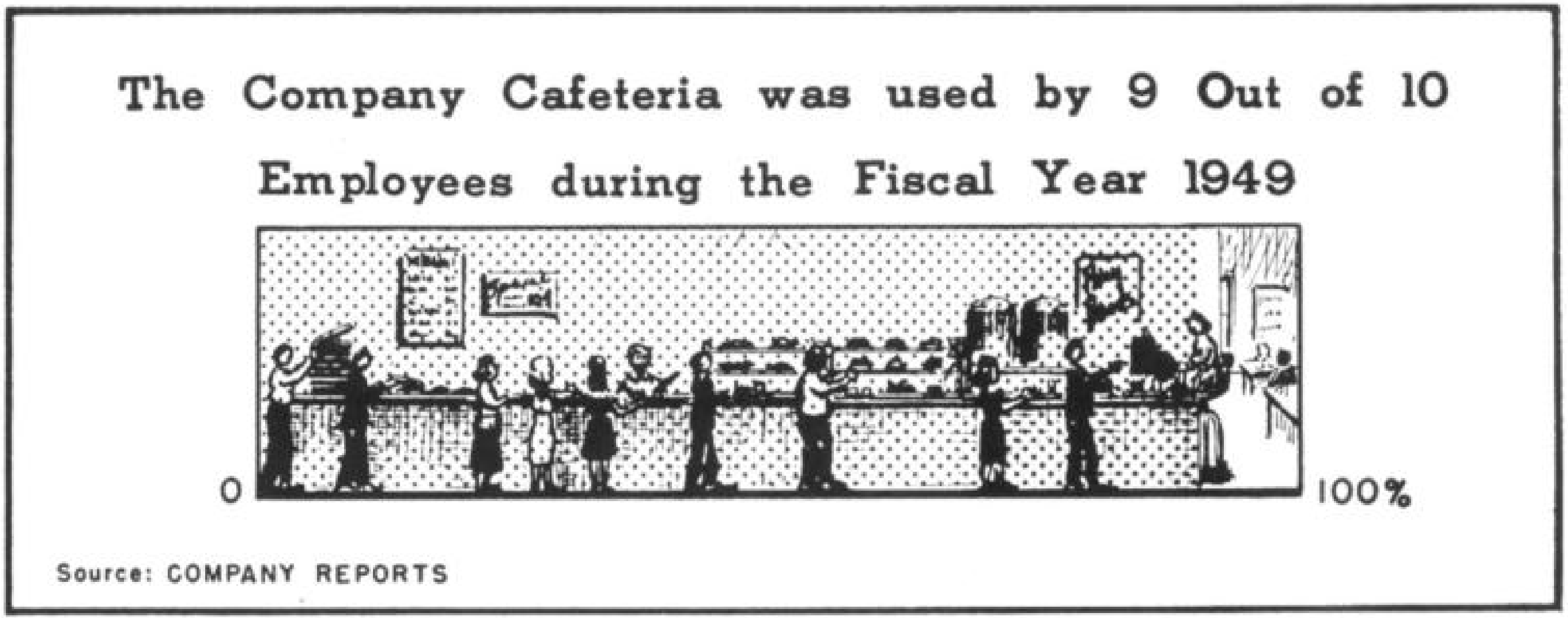
- An introductive example
- Good and bad graphs
  - Basic rules
  - Some additional considerations
- Visual issues
  - Quantitative perception (basic rules)
  - The role of interaction
- Two examples for IR

Rule 0:
Do not use diagrams when handling few numbers

- It does not make sense to use graphs to display very small amounts of data

- The human brain is quite capable of grasping one two, or even three values

Rule 0 violation (and also rule 2)

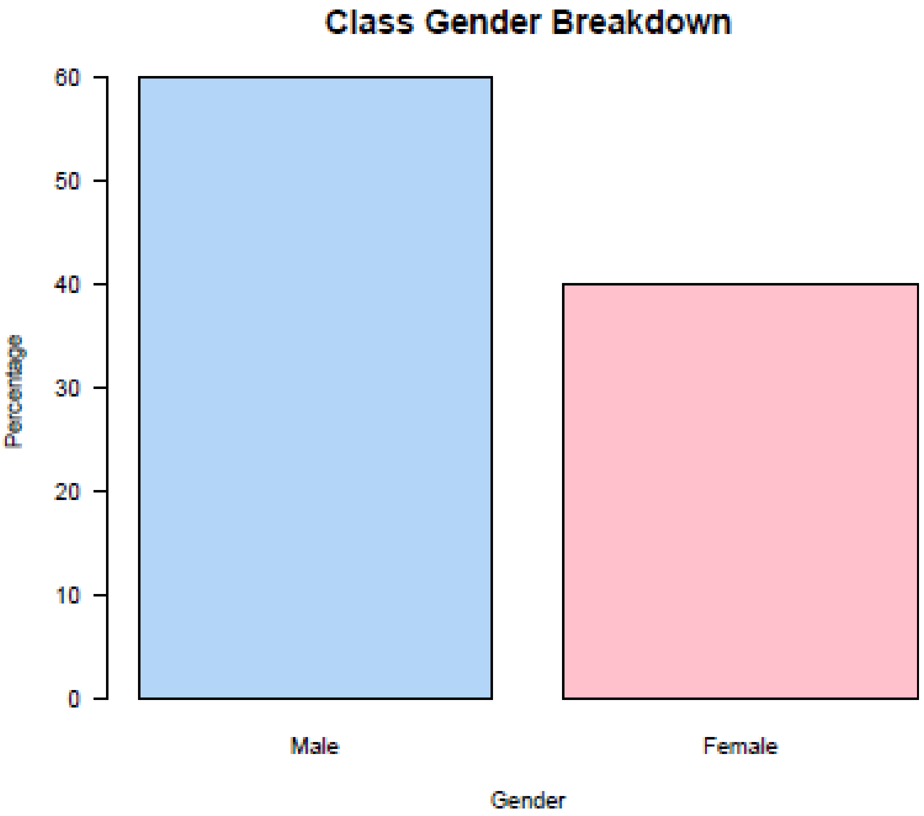


The Company Cafeteria was used by 9 Out of 10 Employees during the Fiscal Year 1949

Source: COMPANY REPORTS

90%

Rule 0 violation



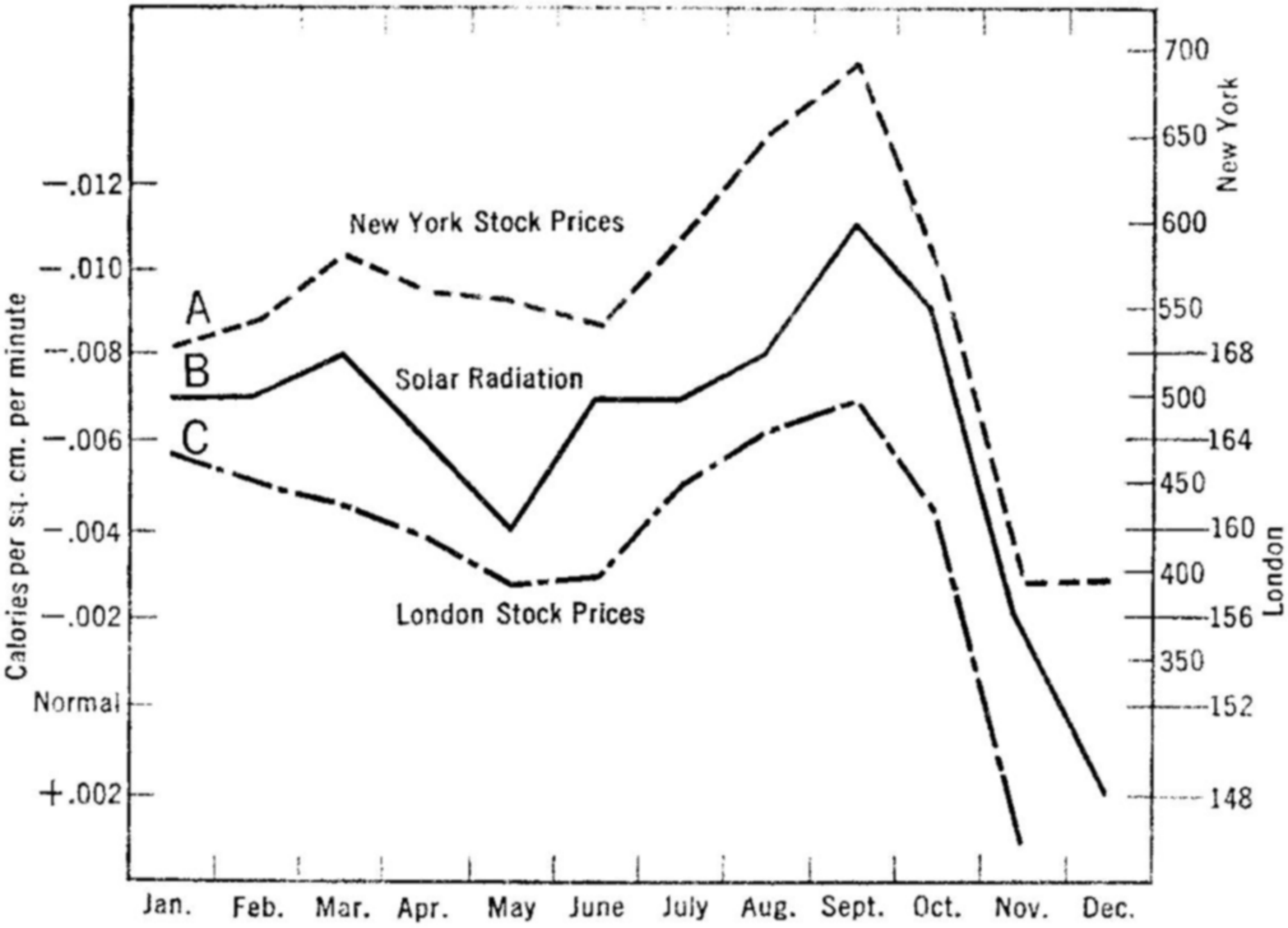
Male	**60%**
Female	**40%**

Rule 1:
Insure data quality / significance

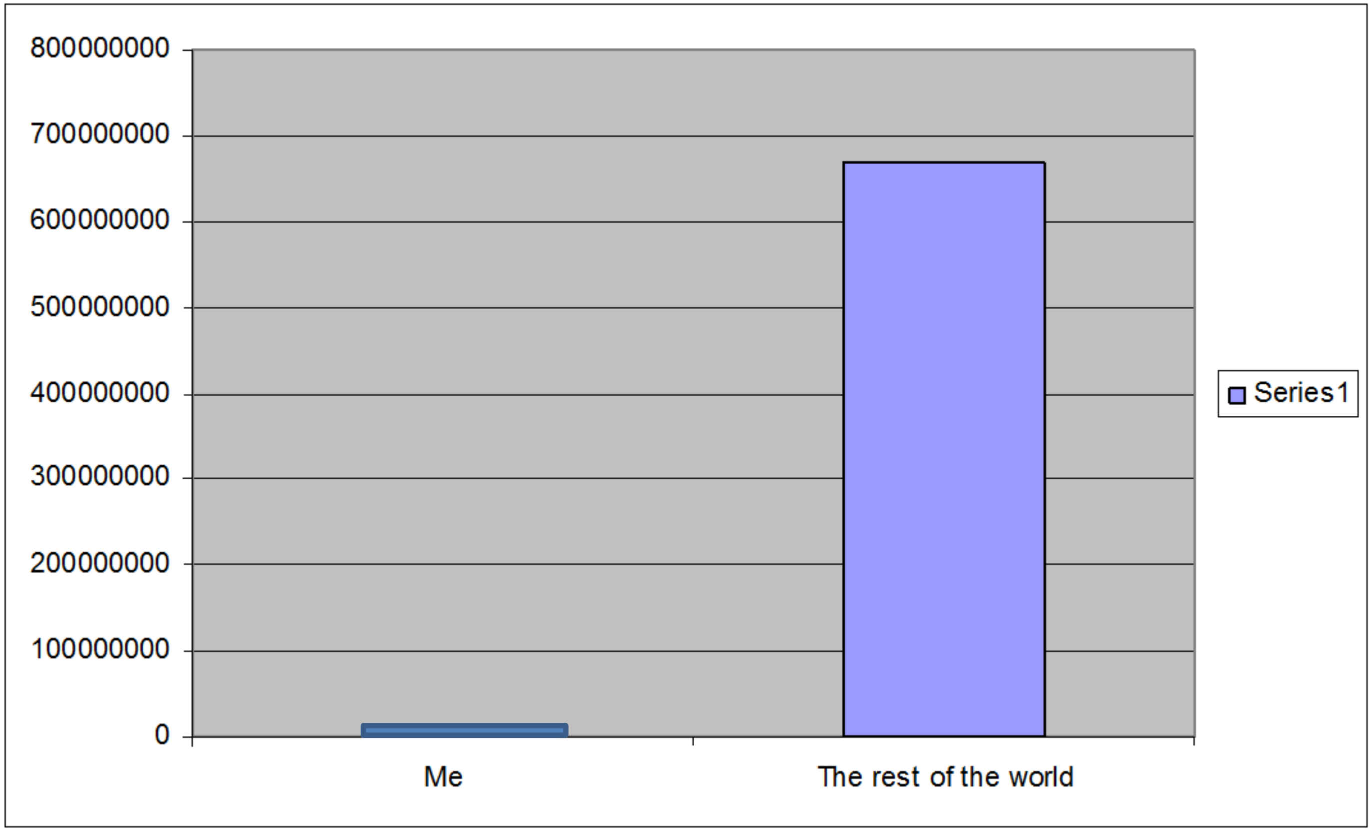
- Graphs are only as good as the data they display

- No amount of creativity can produce a good graph from dubious or non relevant data

Rule 1 violation



Rule 1 violation (and also rule 0)



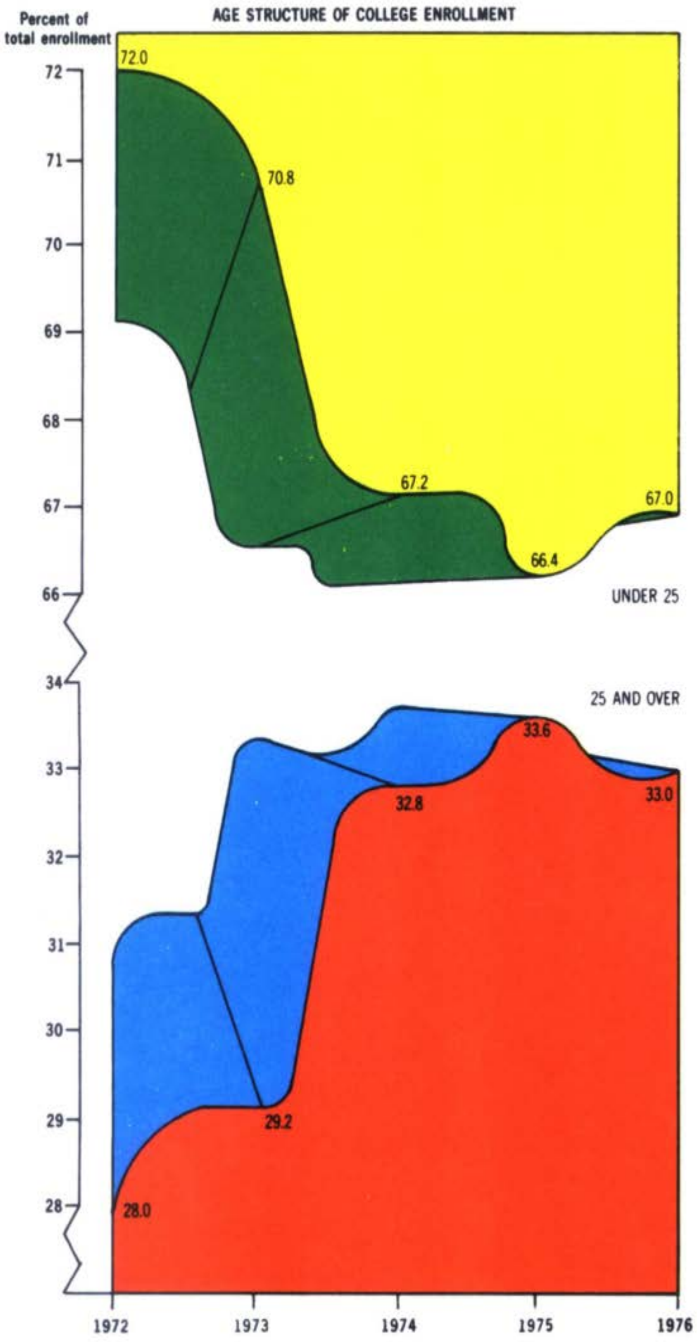
Not very significant data but good example of distortion

Rule 2:
Insure chart simplicity

- Graphs should be no more complex than the data which they portray

- Unnecessary complexity can be introduced by
  - irrelevant decorations
  - colors
  - 3d effects
  - ...
- These are collectively known as “chartjunk”

- For a very comprehensive set of chartjunk effects look at Microsoft Excel
  - the later the version the larger the set !



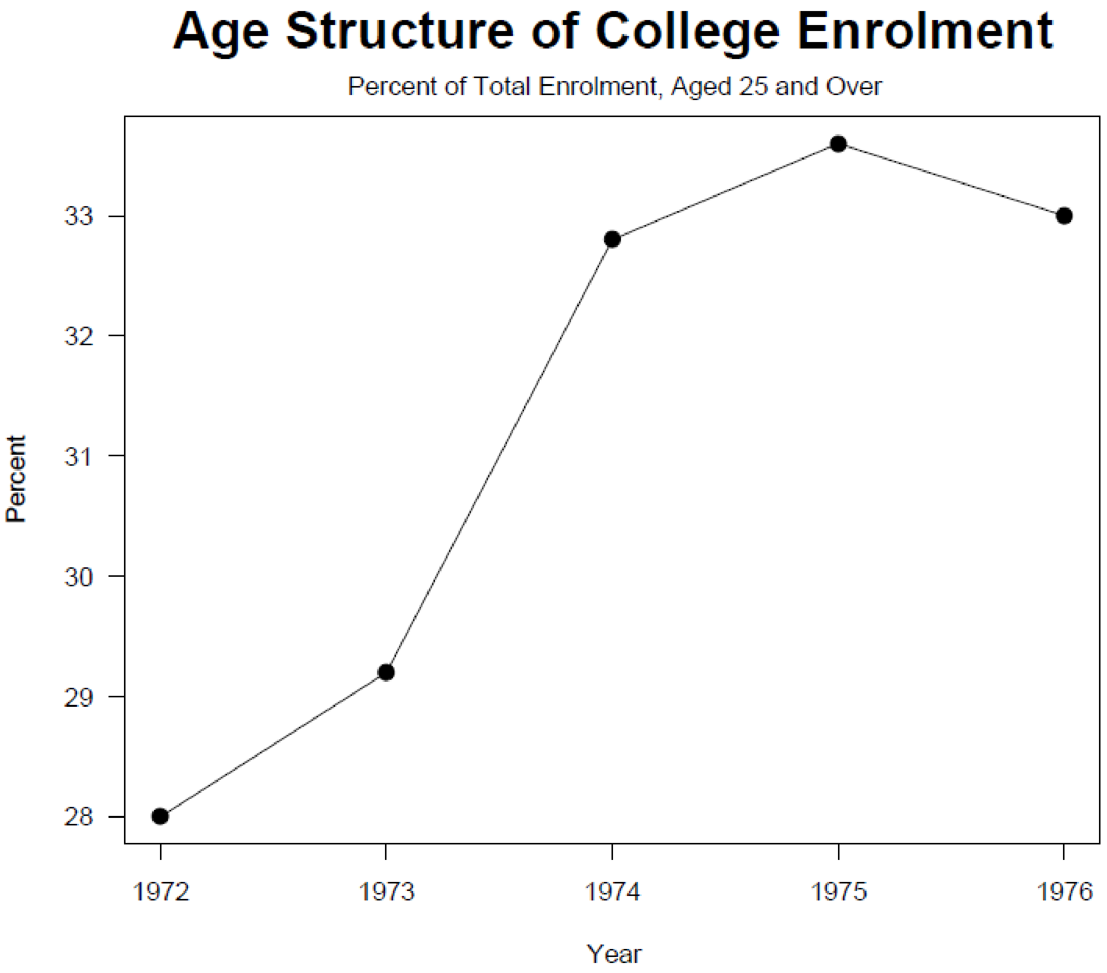
Age structure of College enrollment
(percentage of enrolled people above 25 years)

Rule 2 violation (and also rule 3)

- A very good bad example!
- Only 5 numbers on it but
  - 4 meaningless colors
  - useless 3D
  - useless axes split
  - confusing and wrong visual attributes (size)
  - nonsense interpolation
- Designers of this graph are now working in the Microsoft Excel's team, inspiring the new Excel's versions ...

American Education Magazine

The same data...

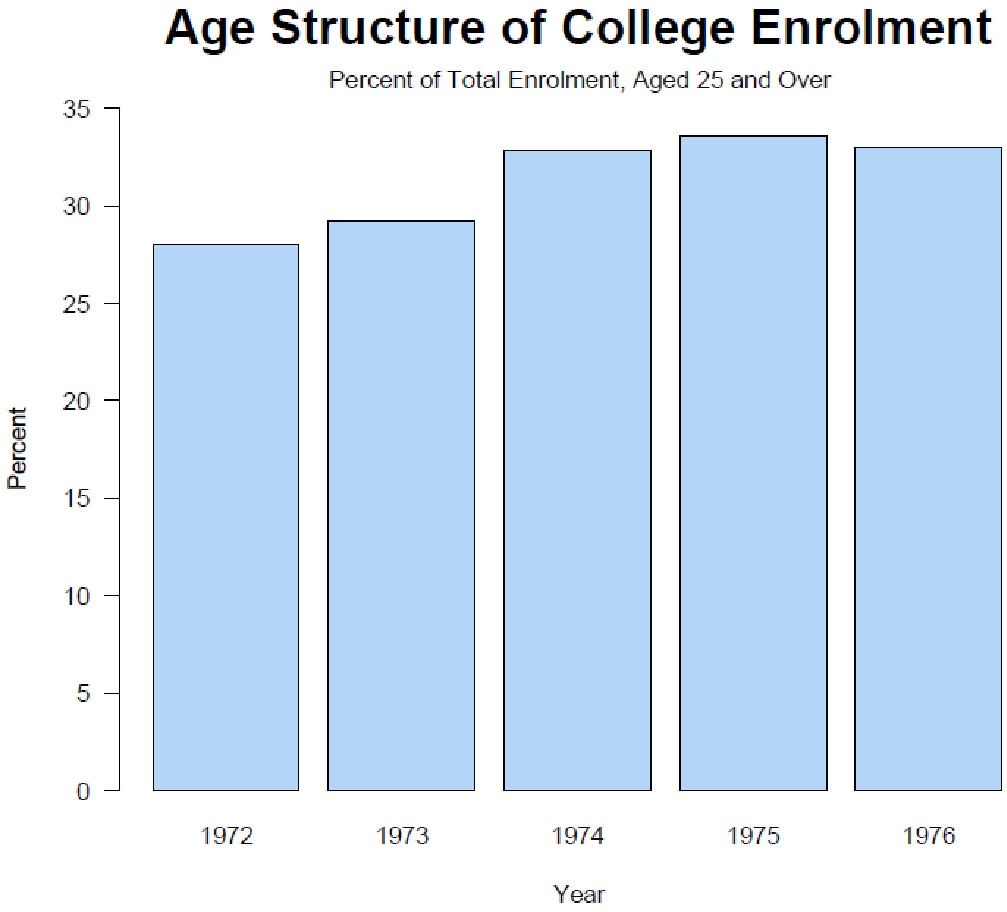


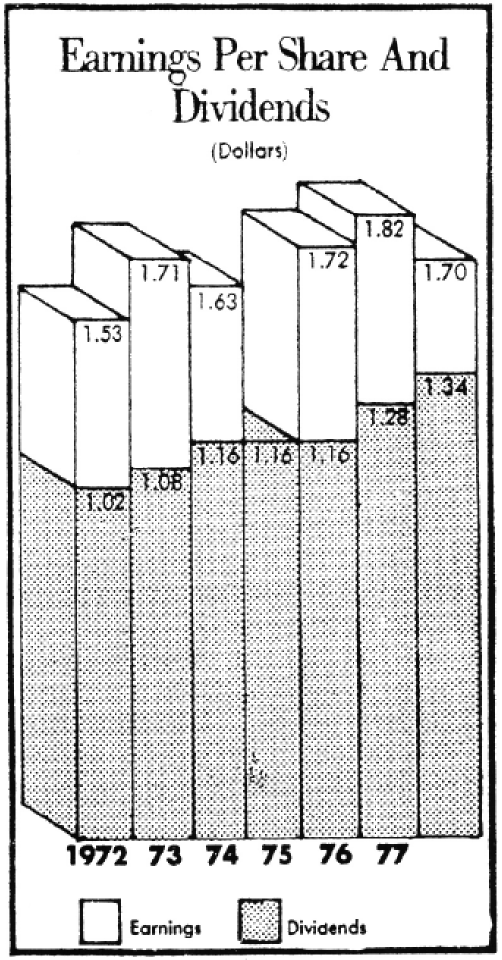
Winter School 2012
Zinal, Valais - Switzerland
23 - 27 January 2012

The same data...

Year	Percentage above 25
1972	28.0
1973	29.2
1974	32.8
1975	33.6
1976	33.0

The same data...



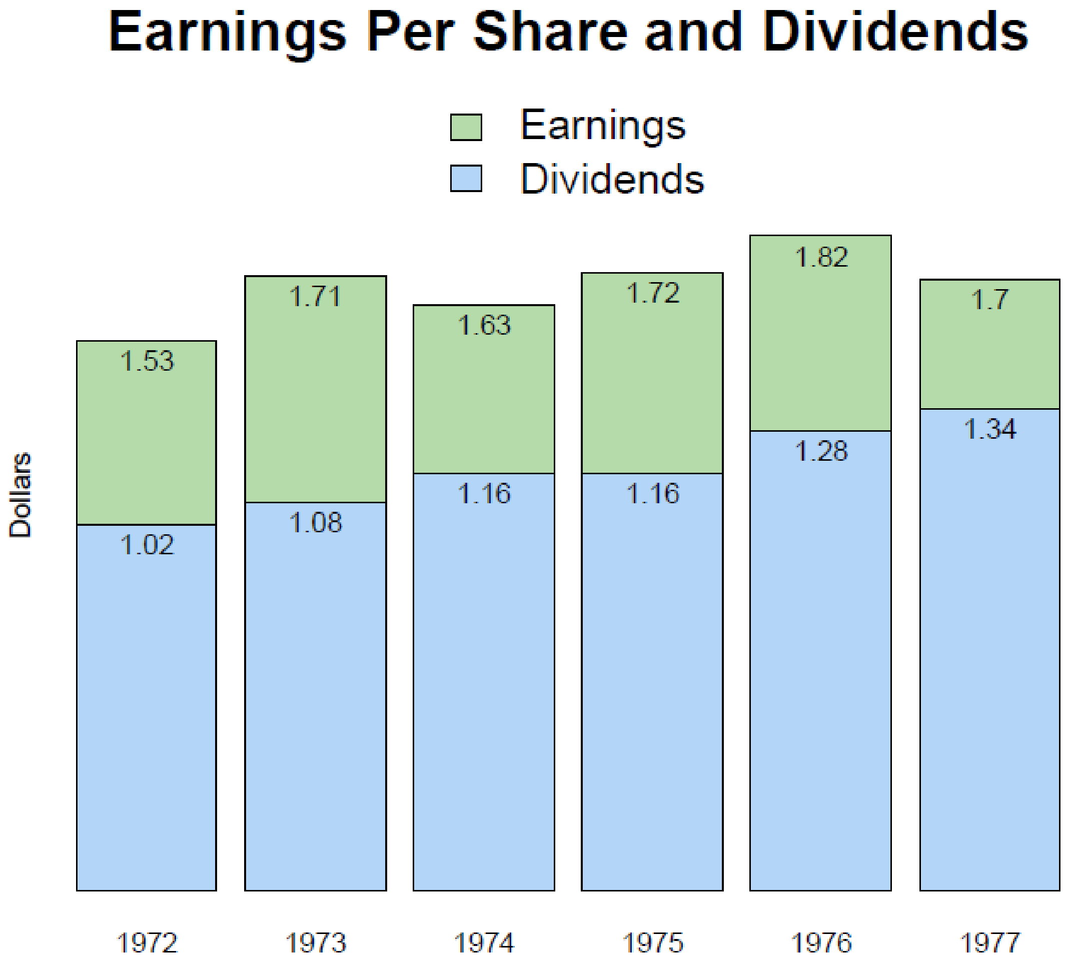


Rule 2 violation

- Why 3D?
- The extra dimension used in this graph has confused even the person who created it..

The Washington Post, 1979

The same data...



Winter School 2012
Zinal, Valais - Switzerland
23 - 27 January 2012

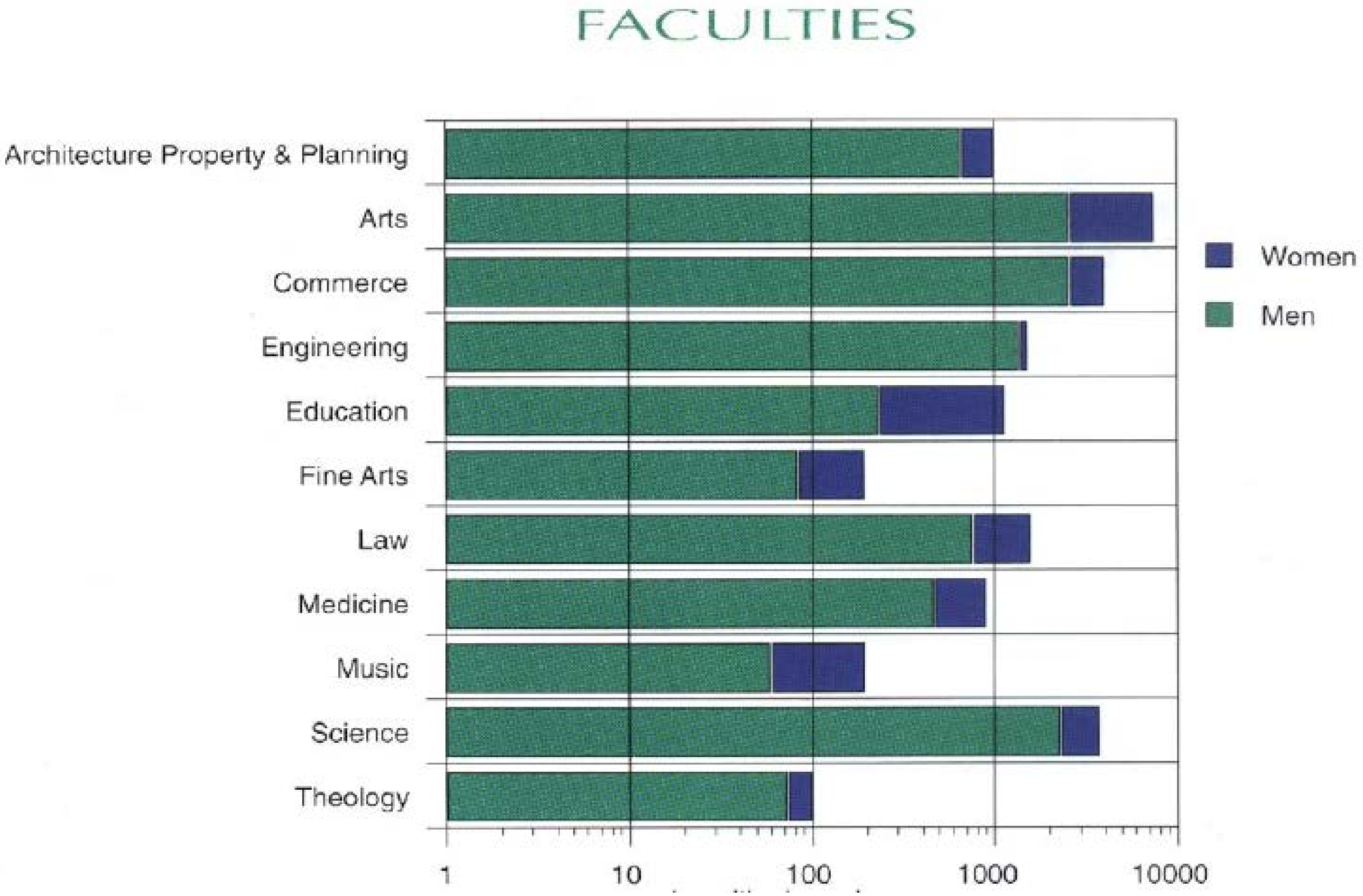
Rule 3:
Do not distort data

- Graphs should not provide a distorted picture of the values they portray

- Distortion can be:
  - deliberate
  - accidental

- Of course, it could be useful to know how to produce a graph which bends the truth...

Rule 3 violation

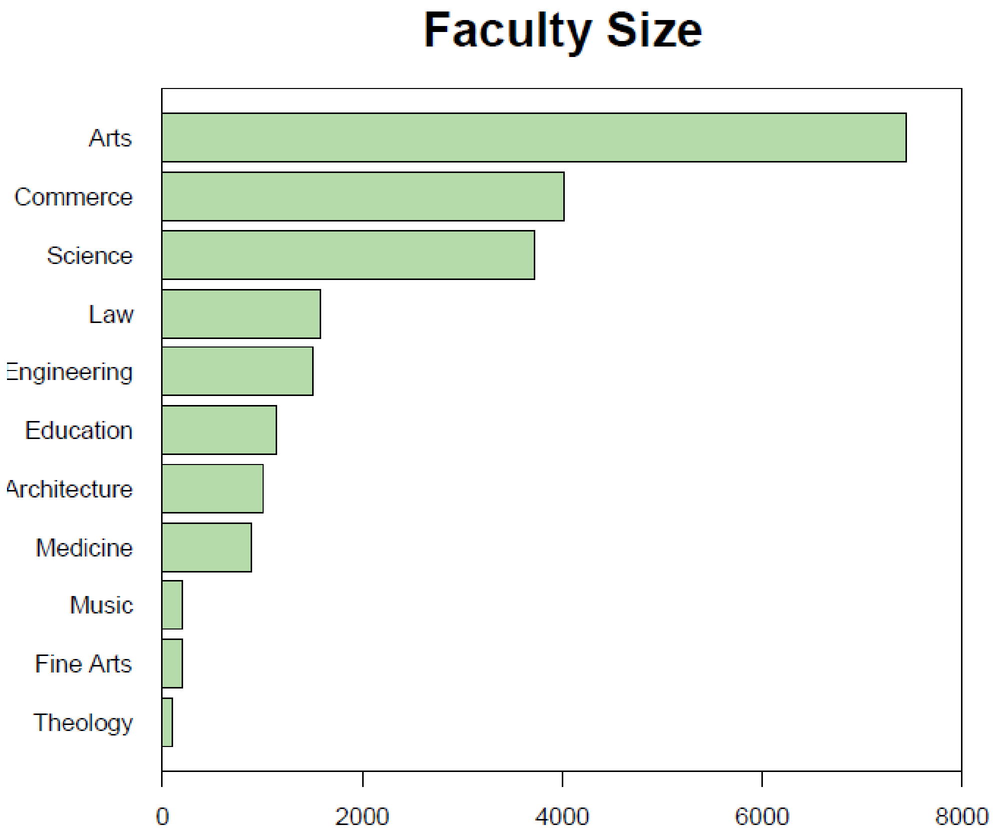


At a very quick glance:

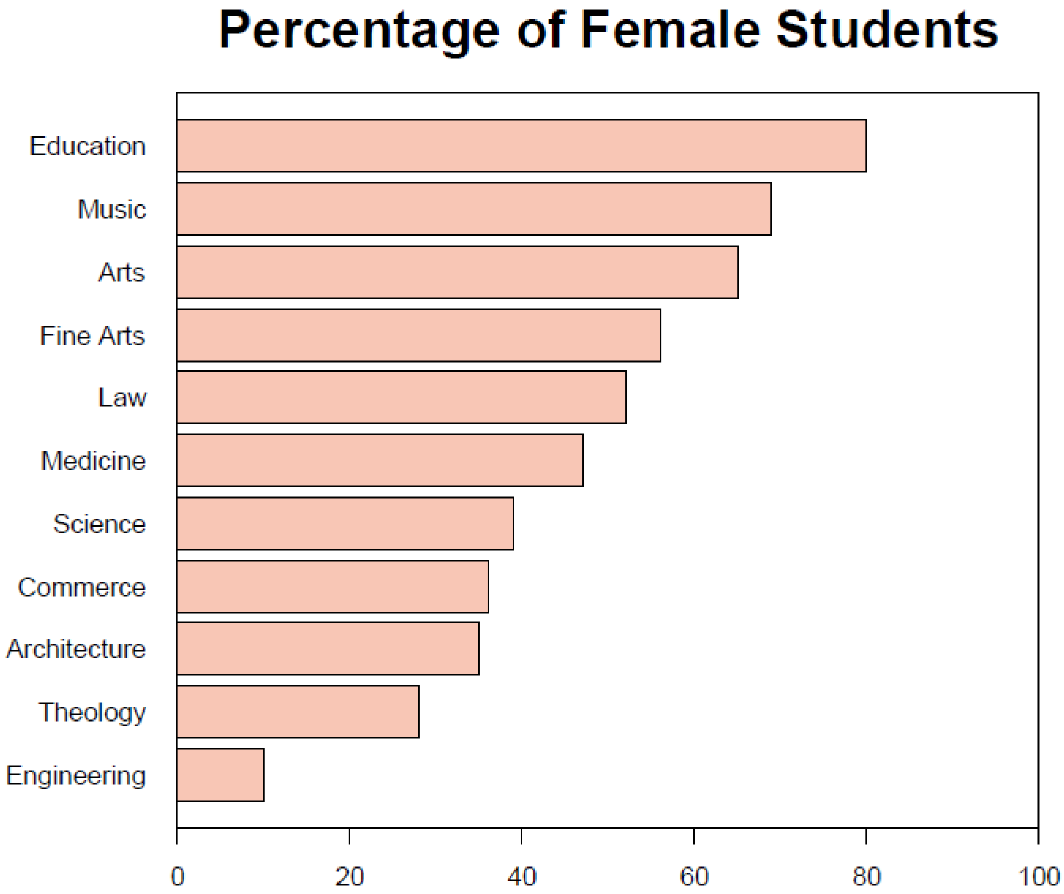
- balanced faculty population
- most male students

What's wrong with this graph?

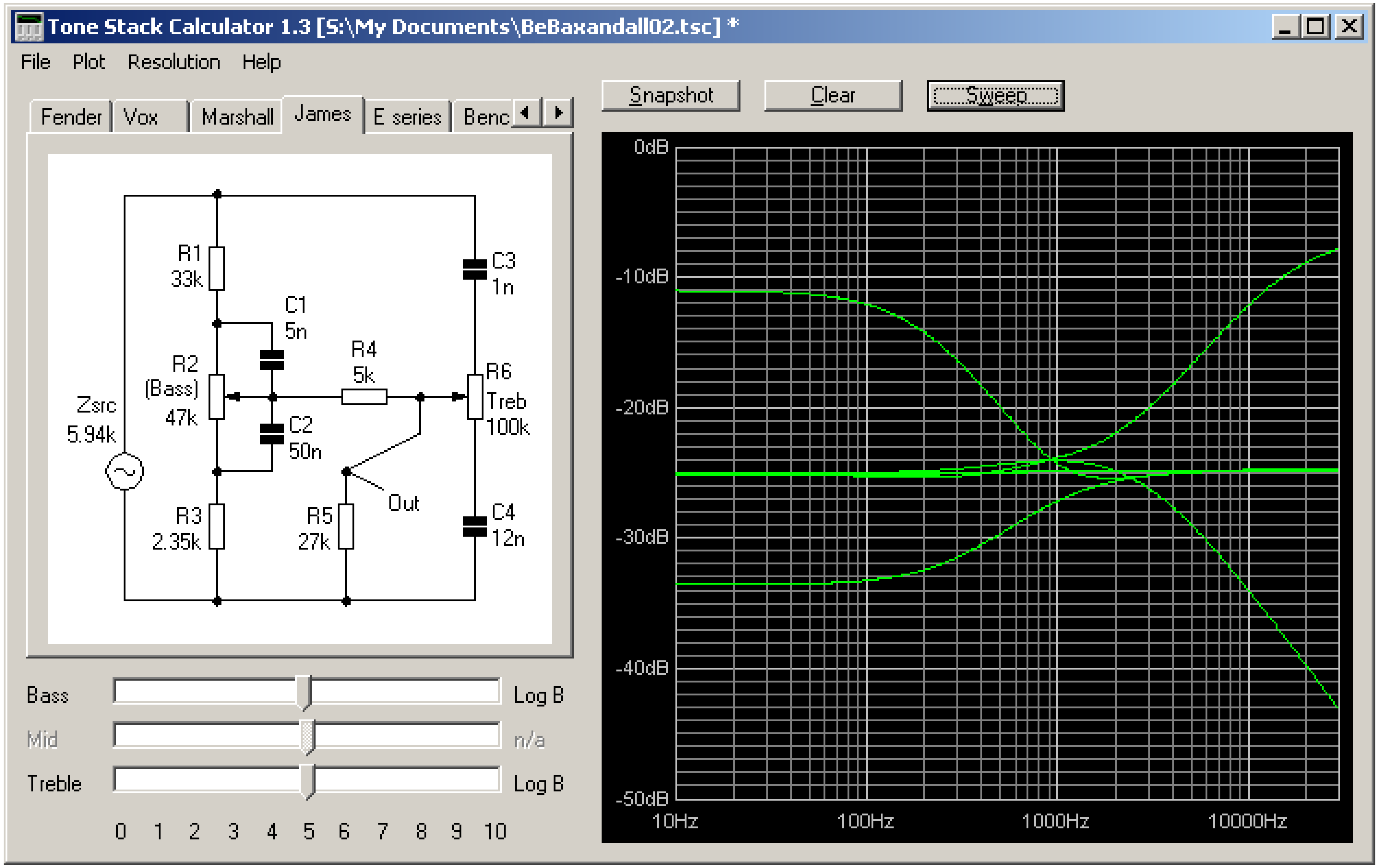
The truth : population size



The truth : male /female ratio



In other cases distortion is ok...



The lie factor

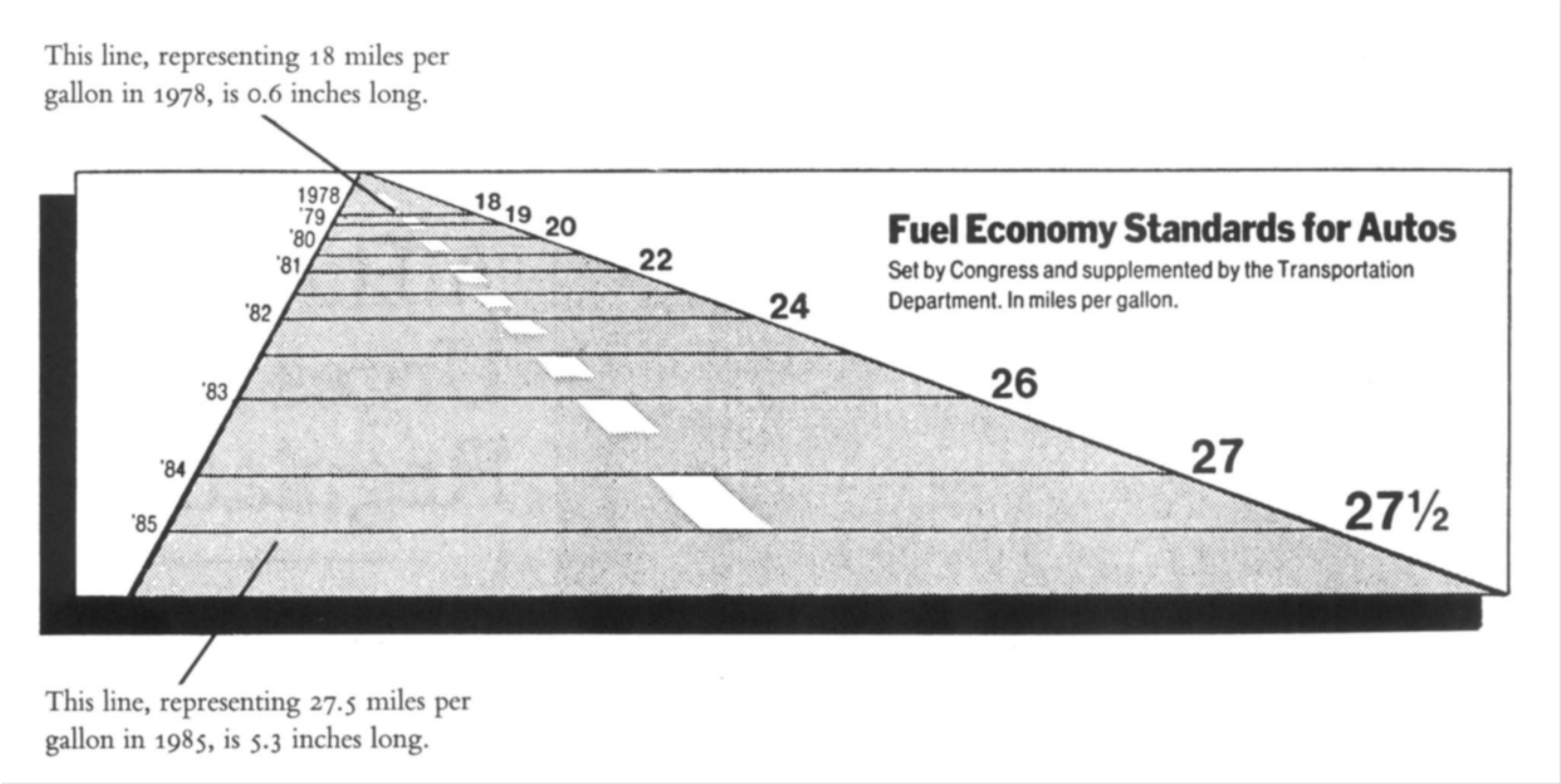
- Edward Tufte of Yale University has defined the "lie factor" as a measure of the amount of distortion

Lie Factor =

size of effect in graphic / size of effect in data

- If the lie factor is greater than 1, the graph is exaggerating the size of the effect

Measuring distortion through the lie factor (miles per gallon across years)



This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

Fuel Economy Standards for Autos

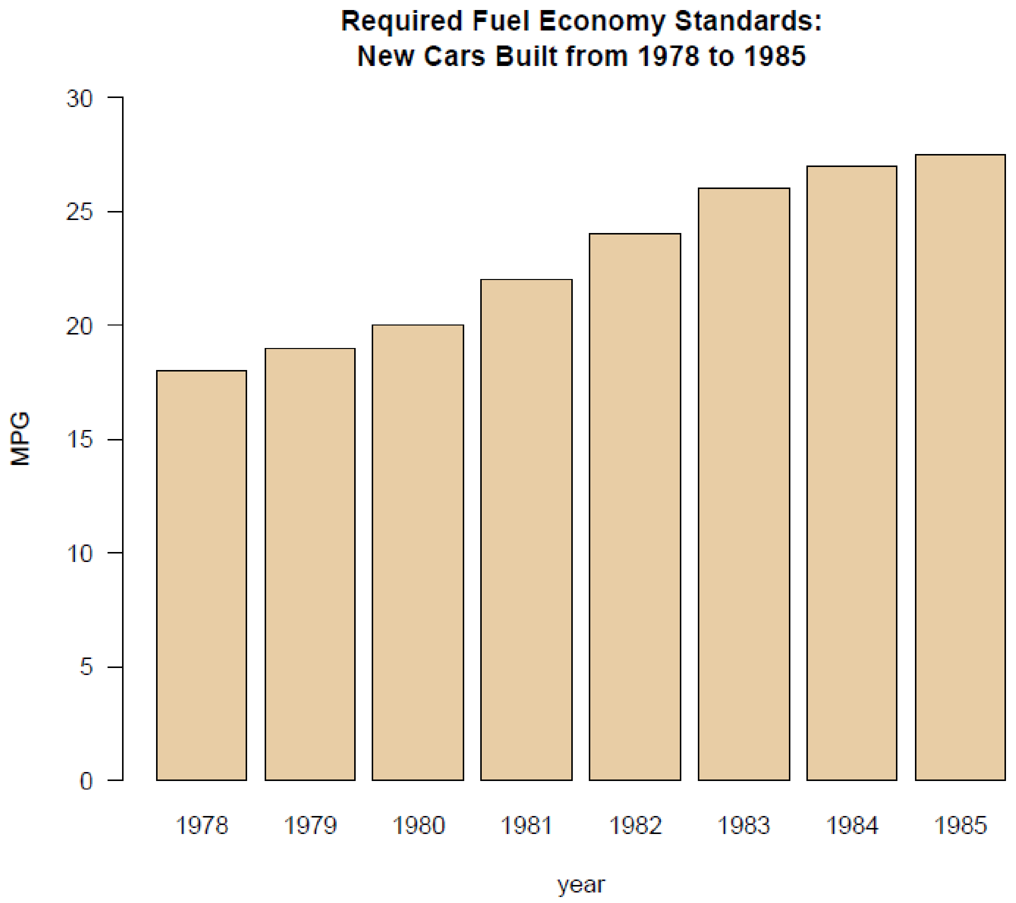
Set by Congress and supplemented by the Transportation Department. In miles per gallon.

This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

$$\text{Data Effect} = \frac{27.5 - 18}{18} = 0.53, \qquad \text{Graph Effect} = \frac{5.3 - .6}{.6} = 7.83,$$

$$\text{Lie Factor} = 14.8$$

The same data with lie factor=1 (and following the previous roles)



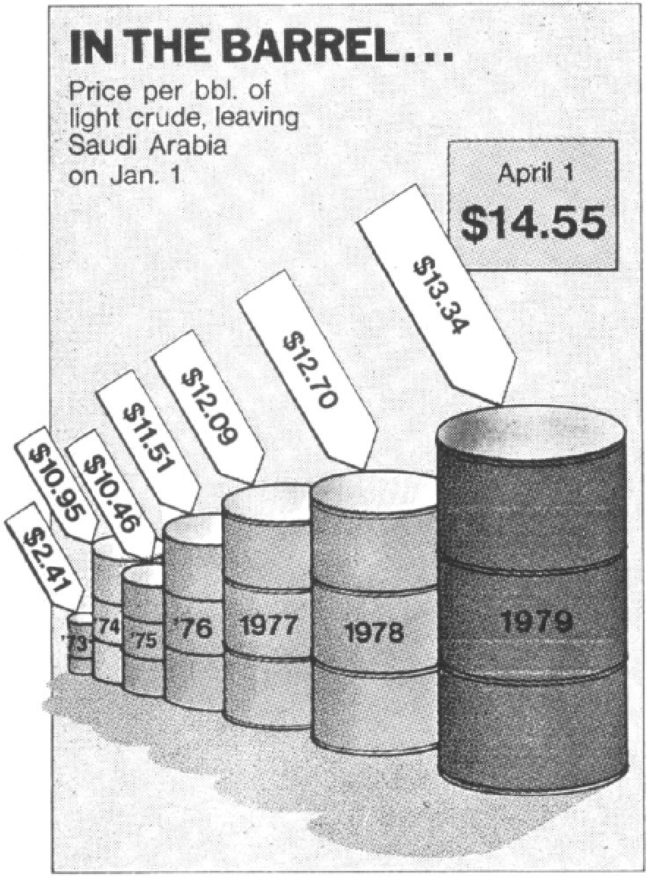
Required Fuel Economy Standards:
New Cars Built from 1978 to 1985

Common sources of distortion

- The use of 3 dimensional “effects” is a common source of distortions in graphs (and of occlusion)

- Another common source is the inappropriate (or deliberate?) use of linear scaling when using area or volume to represent values

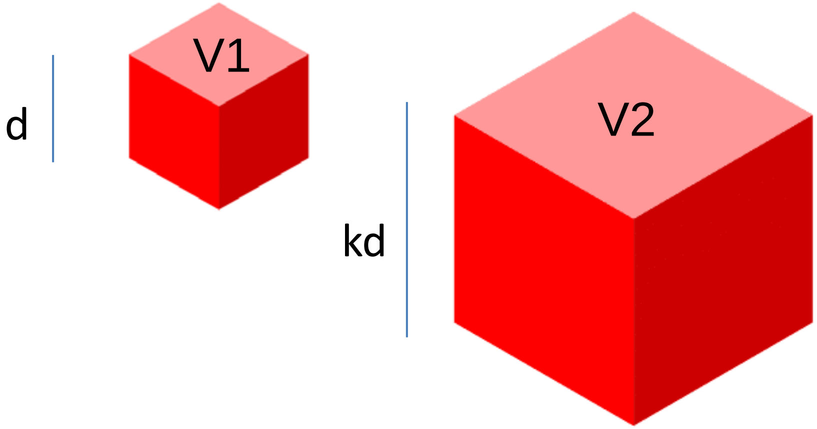
Distortion through volumes



Lie factor= ~9

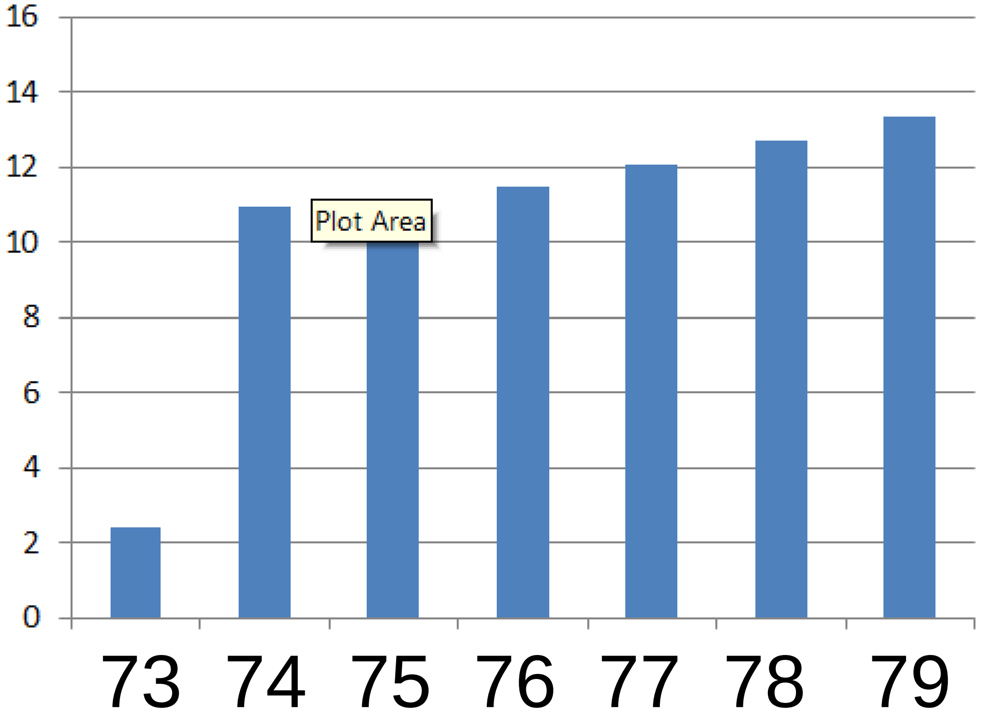
$V1=d^3$
$V2=k^3d^3$

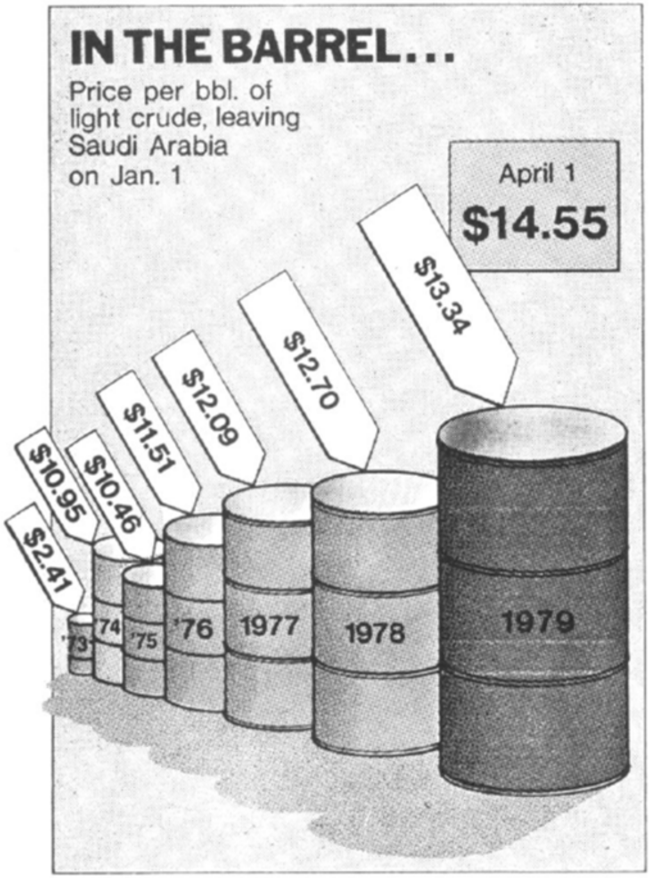
$V1/V2 = k^3$
$kd/d \quad = k$



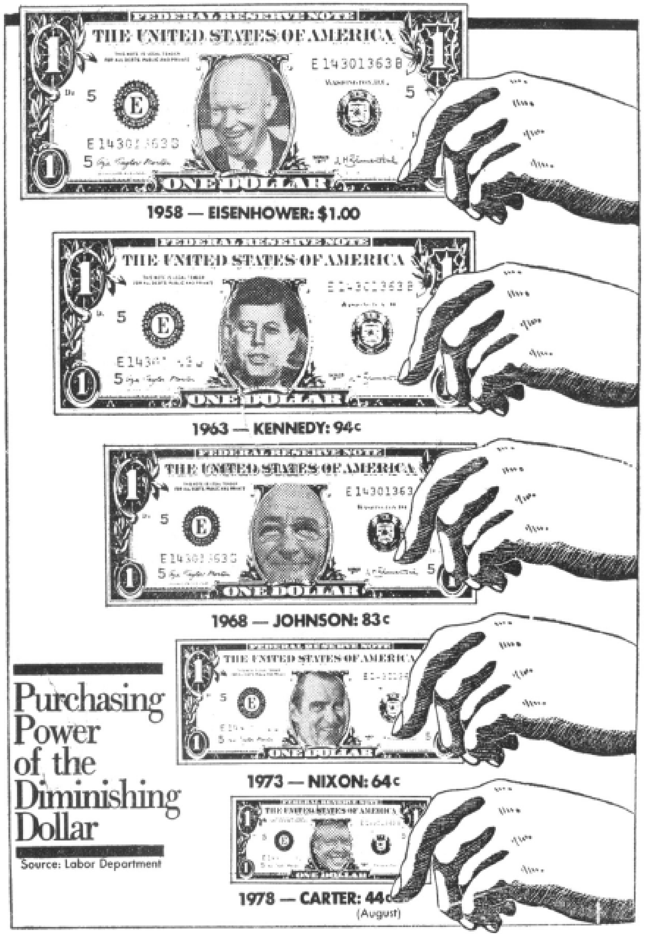
Lie factor ~= $k^3/k = k^2 =$
size_of_effect_in_data^2

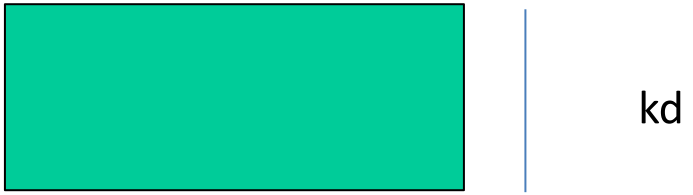
The same data



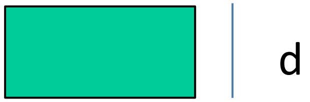


Distortion through areas





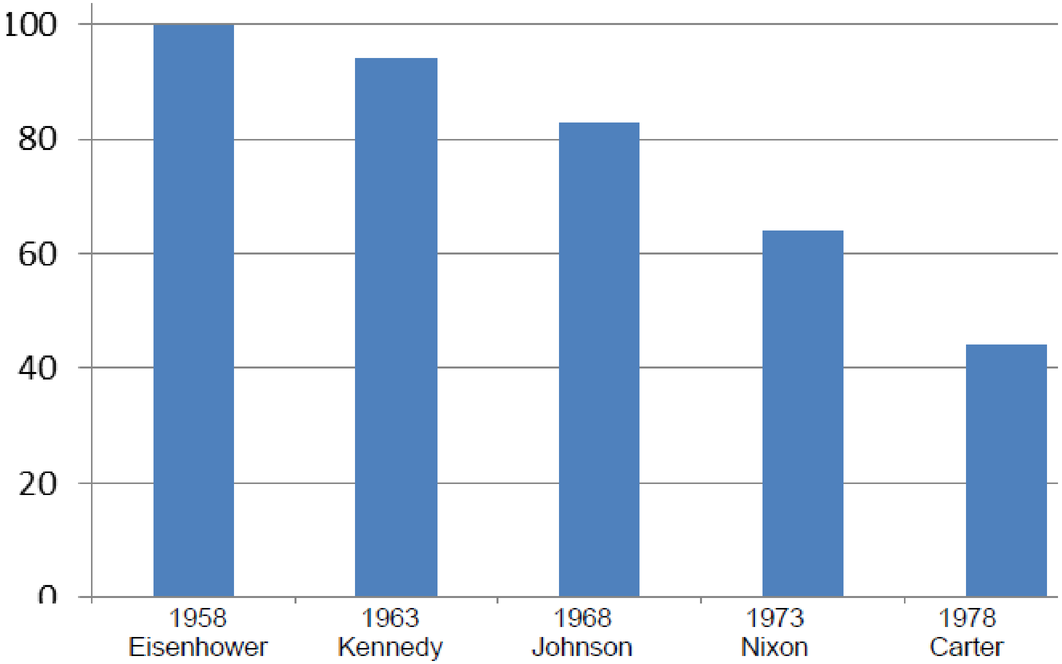
Lie factor ~= $k^2/k = k =$ **size_of_effect_in_data**



Is the bottom dollar roughly half the size of the top one?

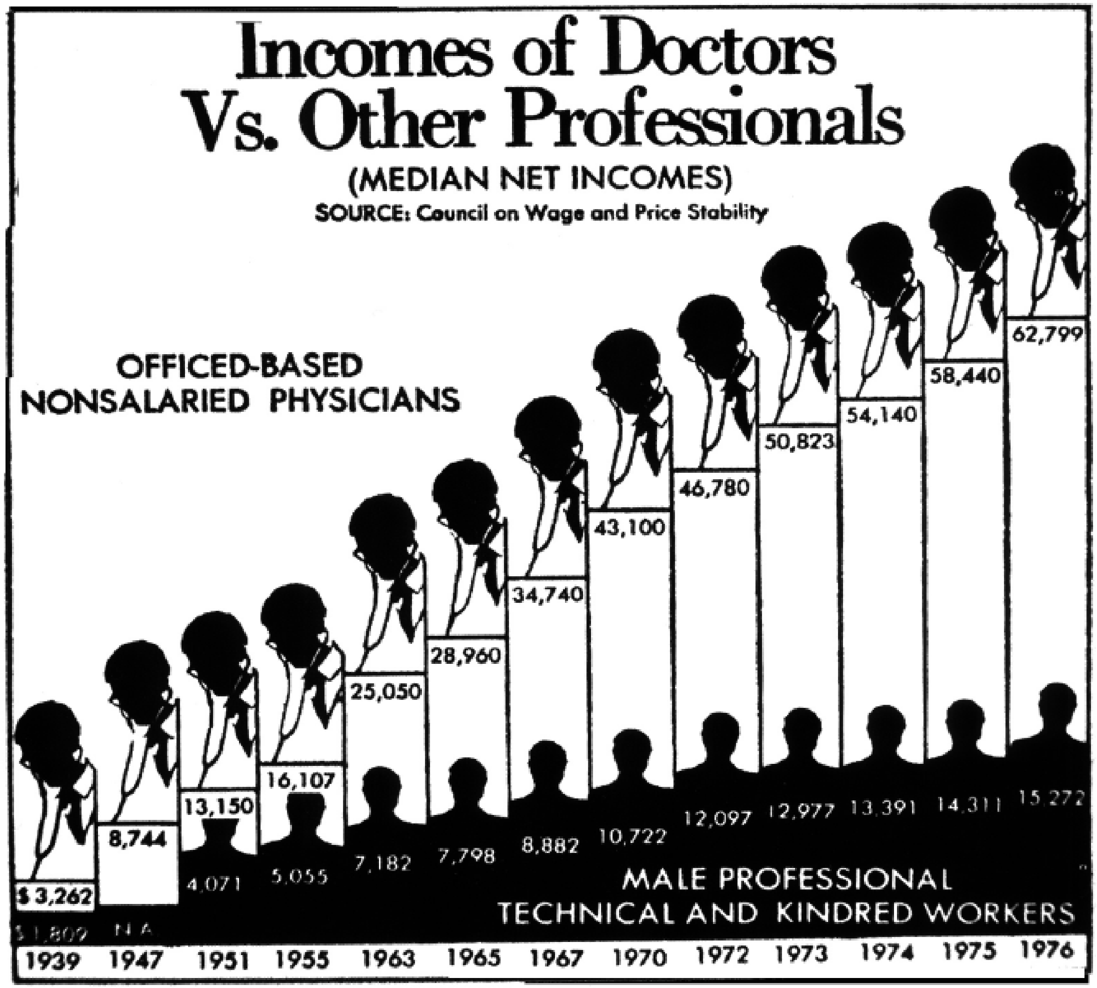
The same data with lie factor = 1

Note that in a histogram you are comparing **lengths**, not **areas**



This is why it is better to use thin bars...

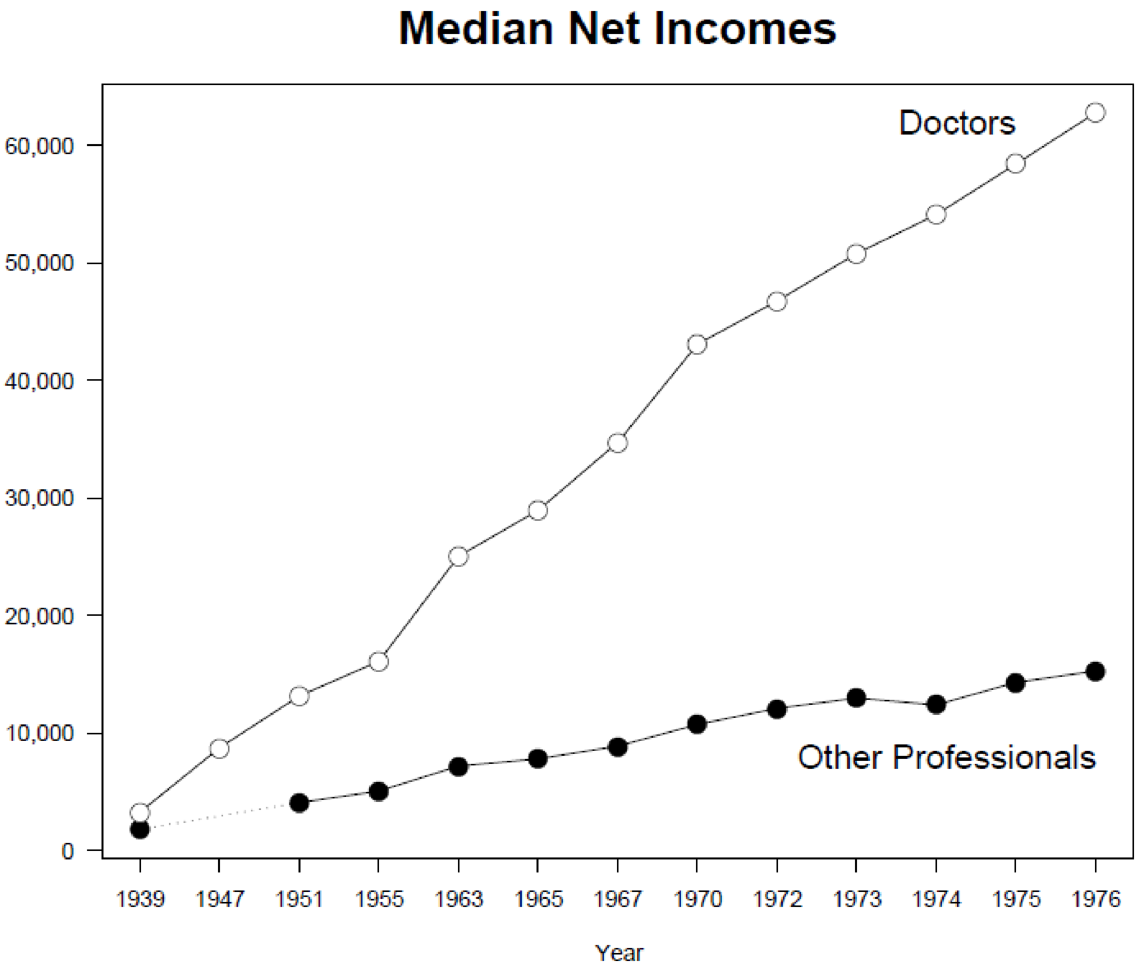
Distortion (deliberate?)



What's wrong with this graph?

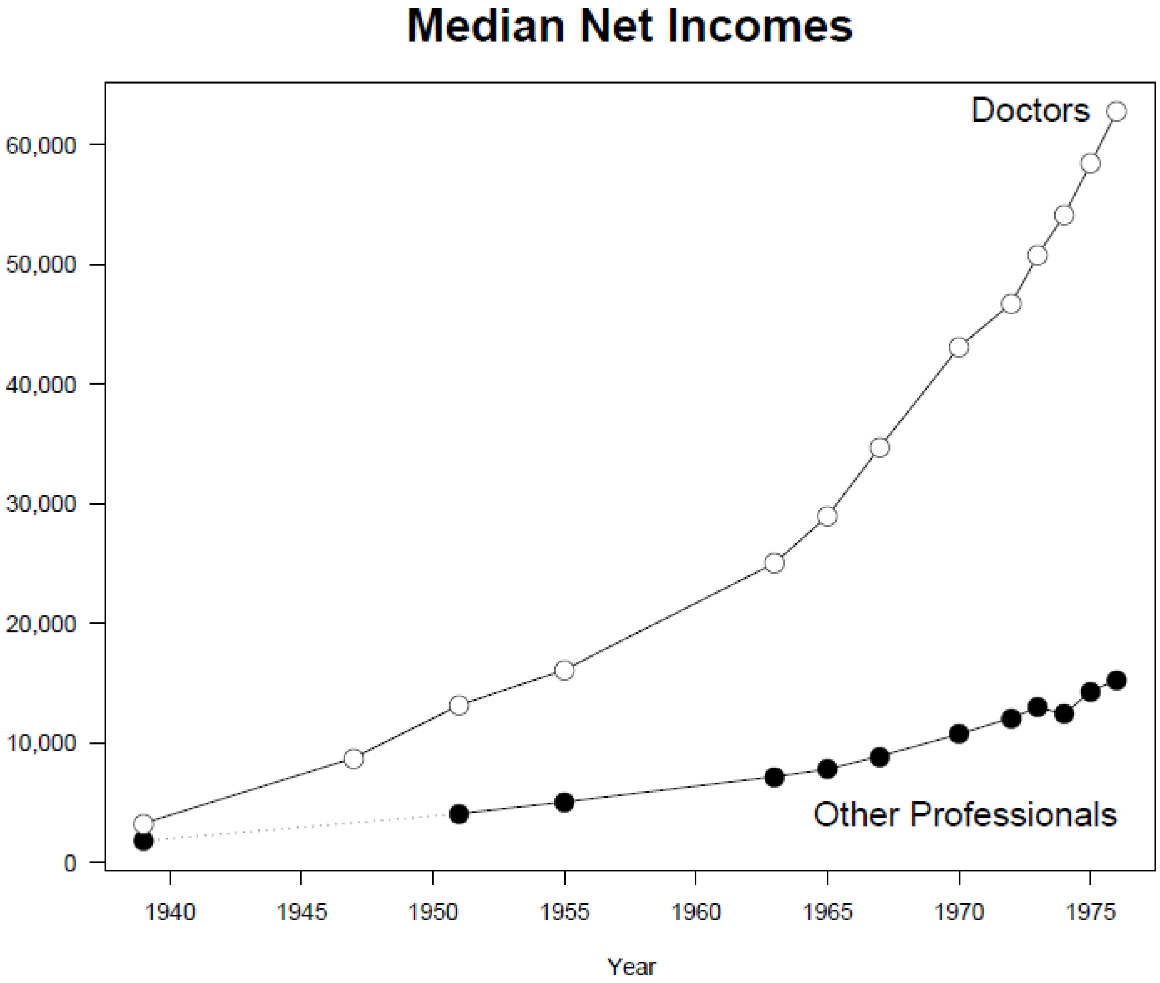
Neglecting chartjunk...

Removing chartjunk



It suggests
a linear trend

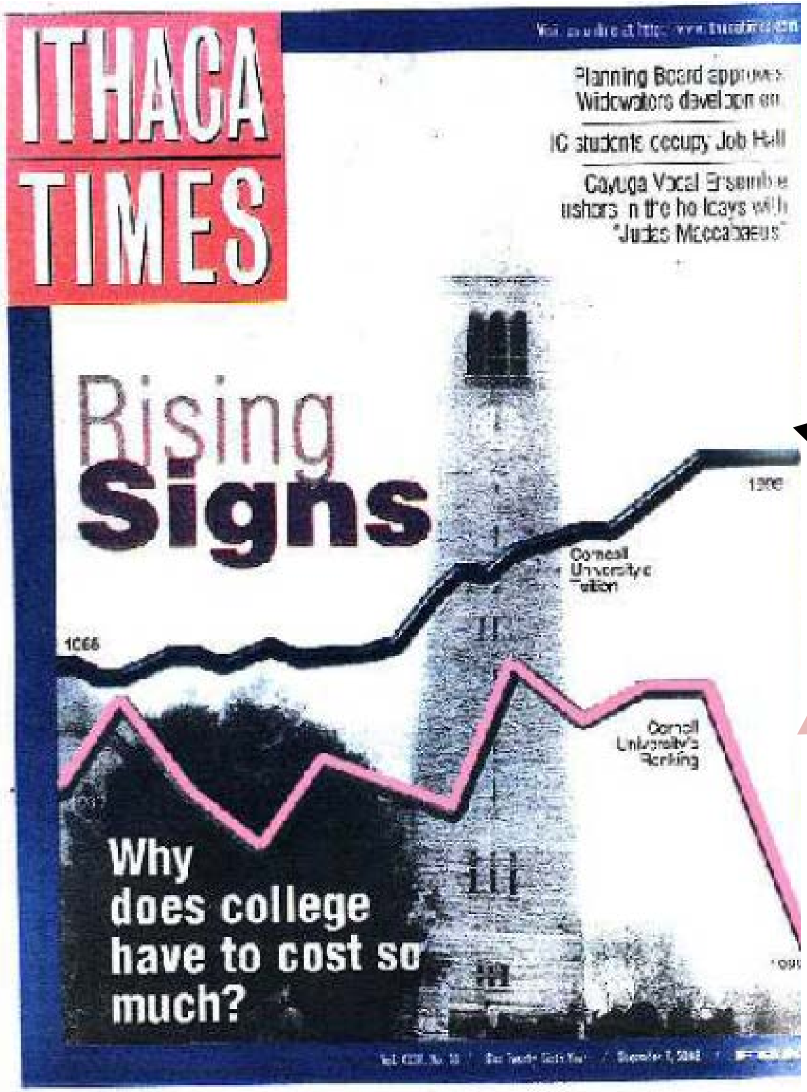
Real data...



The time scale was not consant!

Exponential trend !

One of the best graph lie...

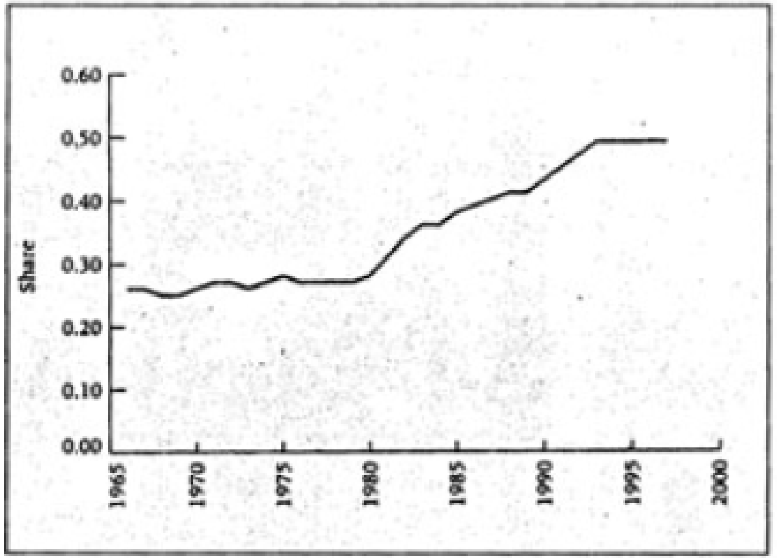


- The cover story, "Why does college have to cost so much?" shows a large graph superimposed on a scene from the Cornell campus. There are two jagged lines running across the graph
  - "Cornell's Tuition" = MONEY
  - "Cornell's Ranking"= QUALITY
- The clear impression is that students are paying more for far less
- What is wrong with it?

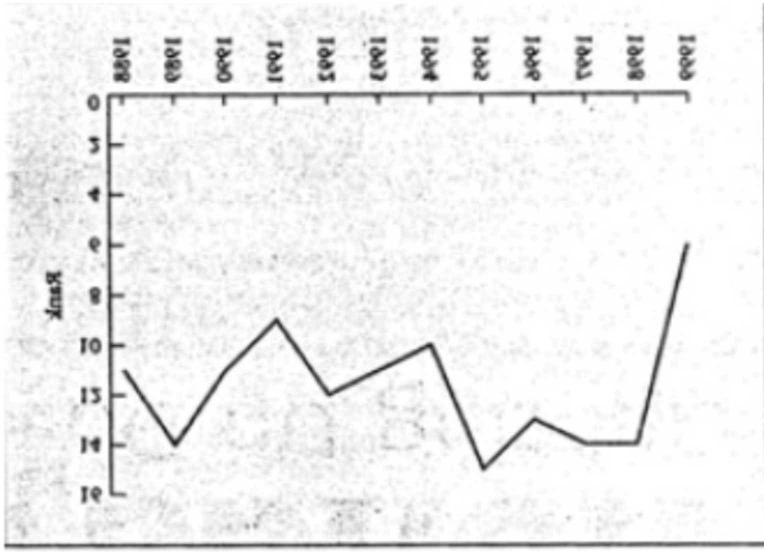
The lie

- The ranking graph covers an 11 year period, the tuition graph 35 years, yet they are shown simultaneously (the same apparent width) on the same horizontal "scale".
- The vertical scale for tuition and ranking could not possibly have common units, but the ranking graph is placed under the tuition graph creating the impression that cost exceeds quality.
- And here is the masterstroke: the sharp "drop" in the ranking graph over the past few years actually represents the fact that Cornell's rank has IMPROVED from 15th TO 6th ...

The real data



Money



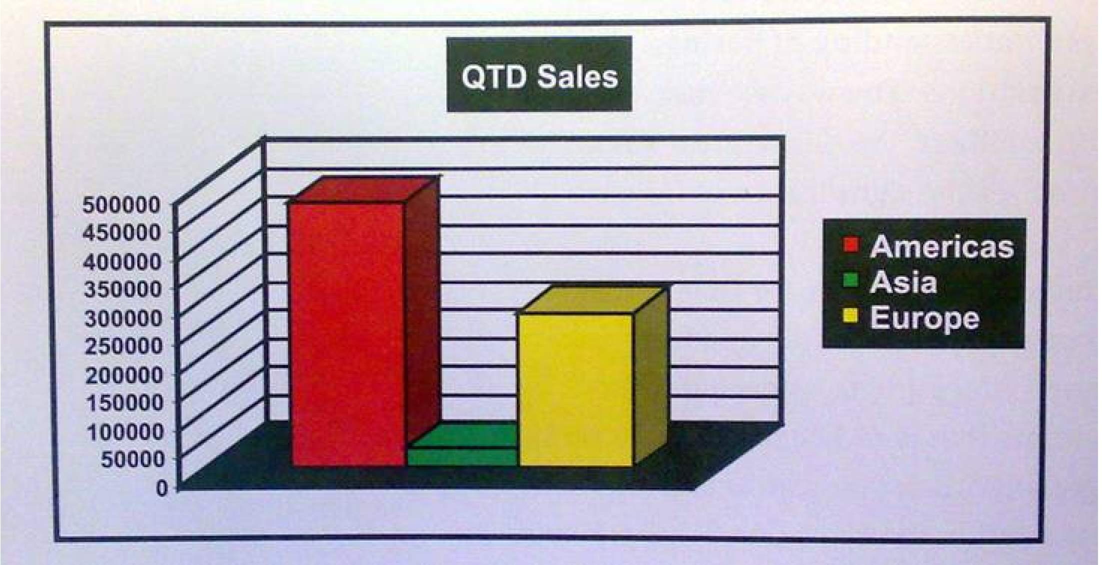
Rank

Outline
(basically what you have NOT to do)

- An introductive example
- Good and bad graphs
  - Basic rules
  - Some additional considerations
- Visual issues

Another bad example

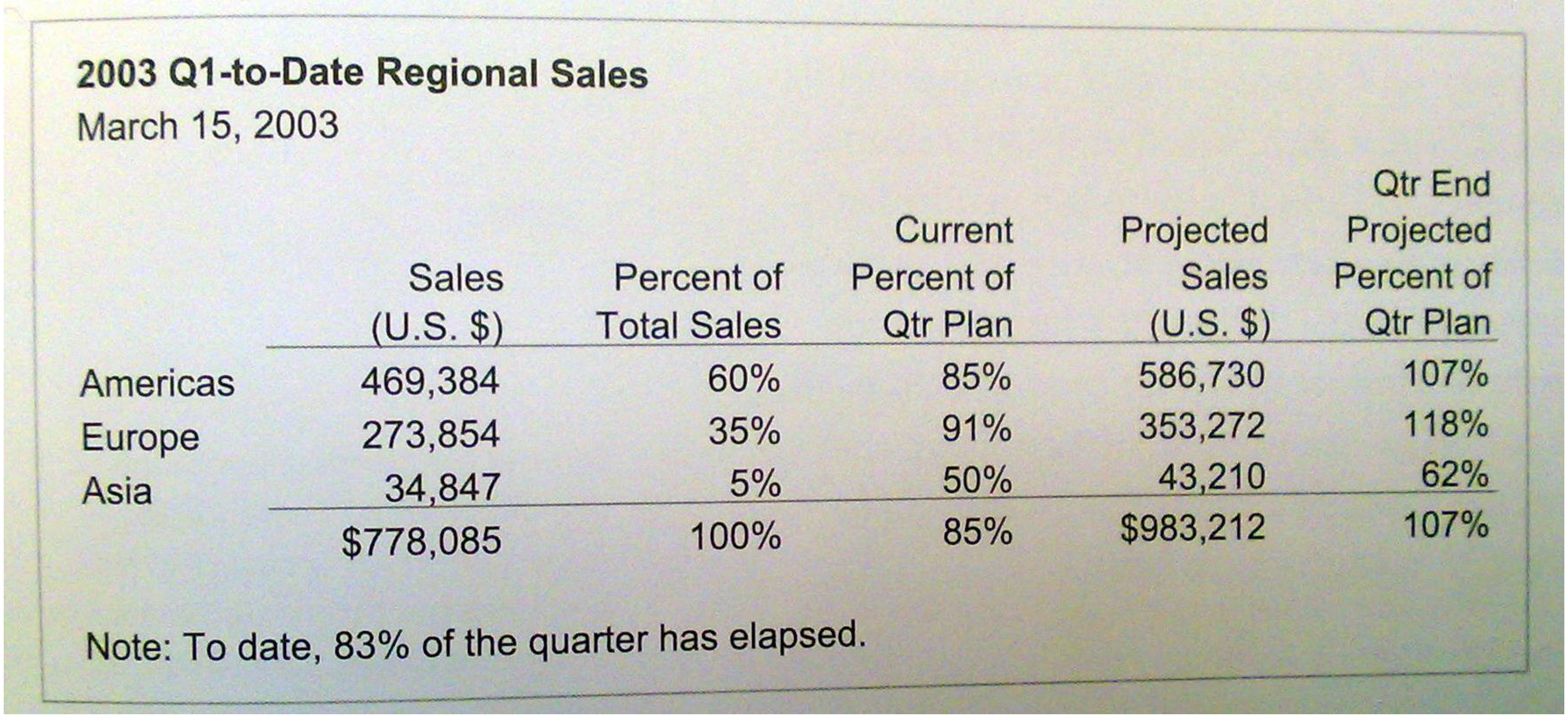
- You are a manager of a big company
- You need to control and to report, every Monday, the current state of quarterly sales in the Americas, Asia, and Europe, with the goal of verifying your forecast
- Someone presents you with this graph
- Are you happy with it? (disregarding chartjunk)



•YOU MISS :

•Units !
•The actual date !
•Some additional summarizing information (e.g., percentages)
•Planned sales v.s. actual sales

All the needed information



2003 Q1-to-Date Regional Sales

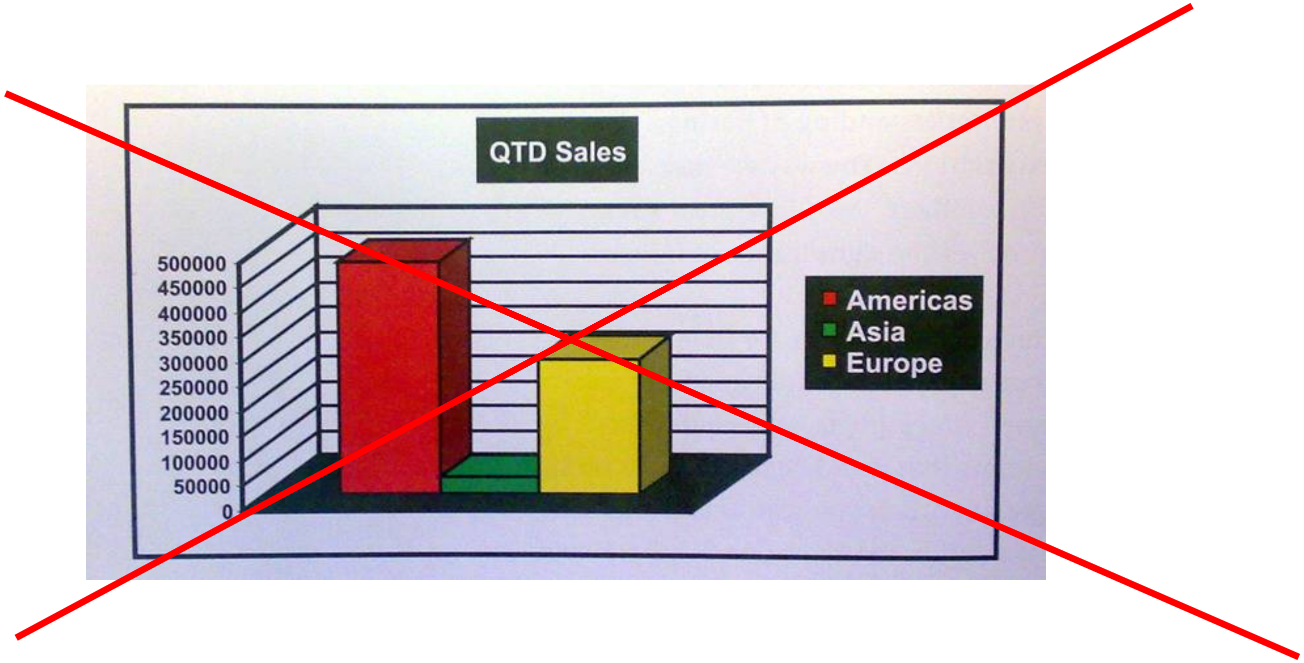
March 15, 2003

					Qtr End
			Current	Projected	Projected
	Sales	Percent of	Percent of	Sales	Percent of
	(U.S. $)	Total Sales	Qtr Plan	(U.S. $)	Qtr Plan
Americas	469,384	60%	85%	586,730	107%
Europe	273,854	35%	91%	353,272	118%
Asia	34,847	5%	50%	43,210	62%
	$778,085	100%	85%	$983,212	107%

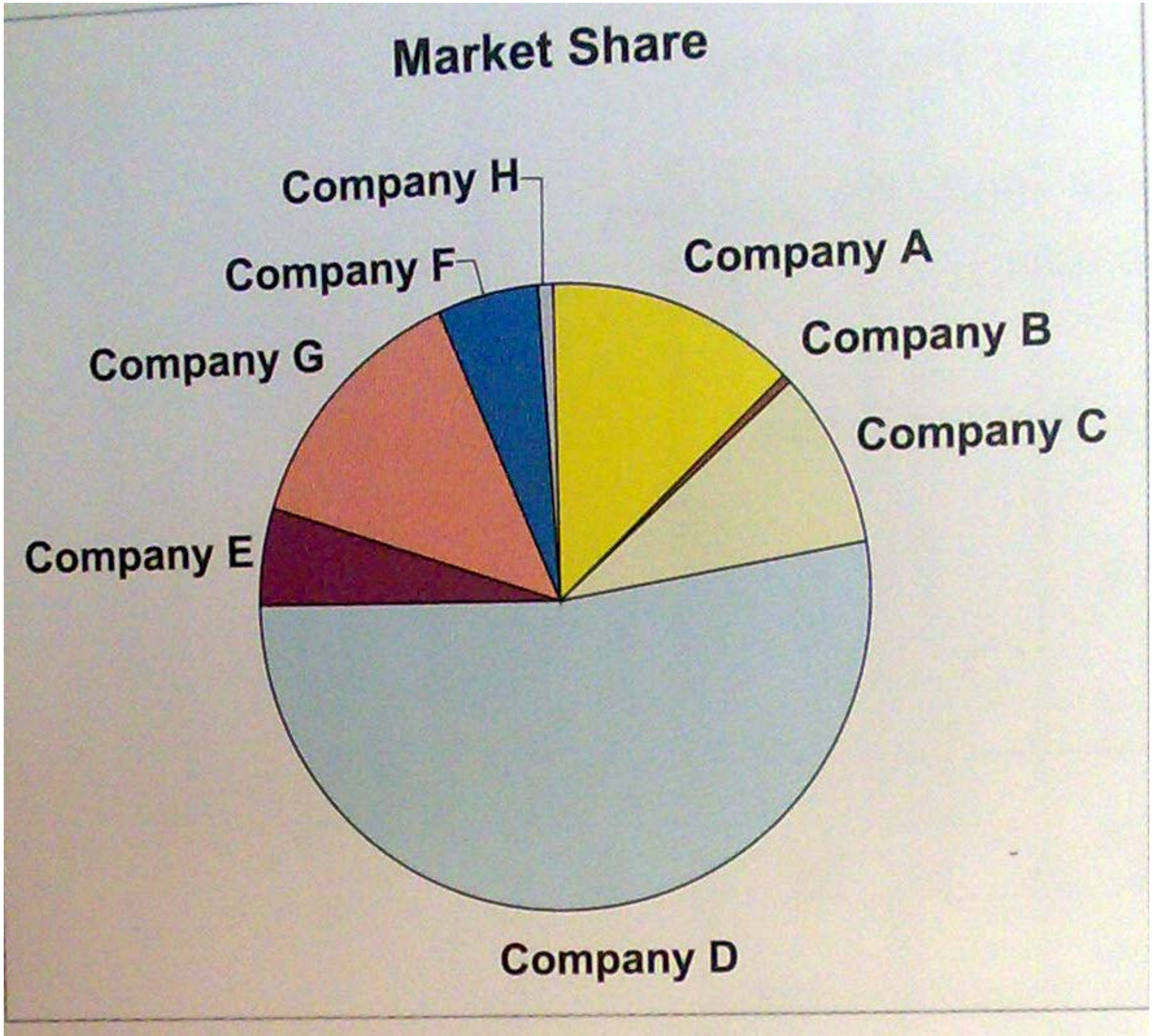
Note: To date, 83% of the quarter has elapsed.

Always remember to (in addition to rules 0..3):

- Label your axes
- Make your units clear
- Use appropriate and readable label values
- Add useful ancillary pieces of information

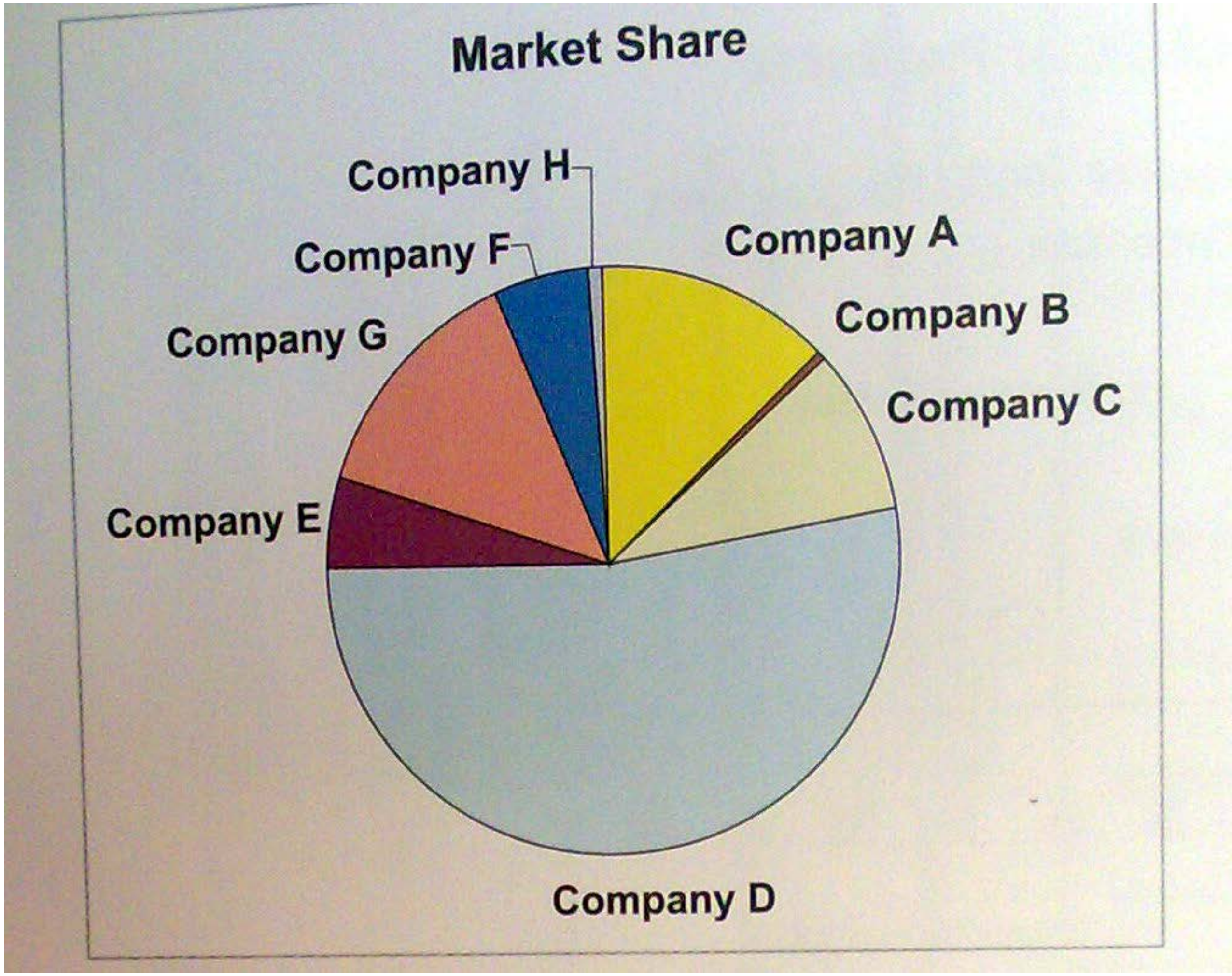


The last example: our company against the world!



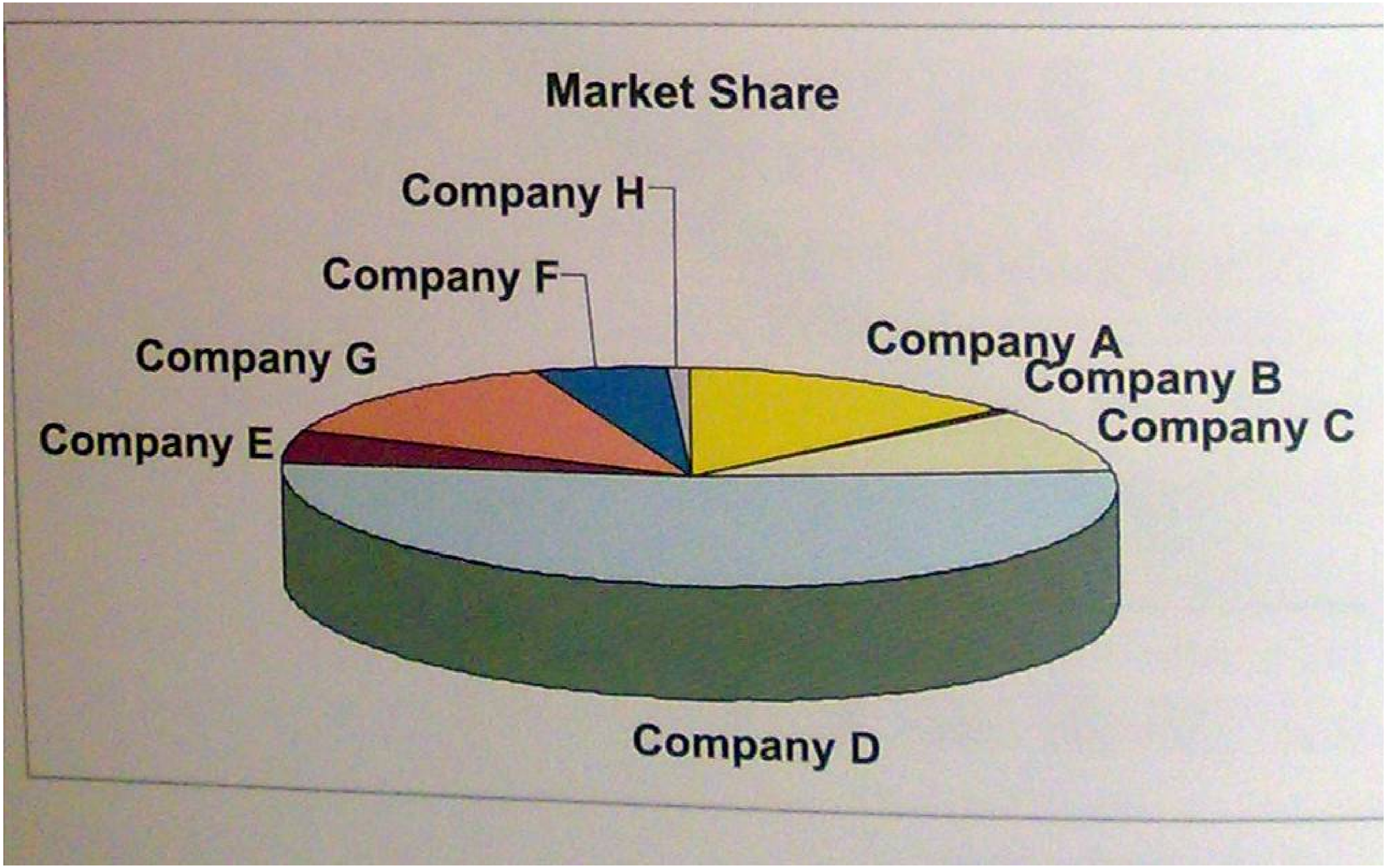
- •What is the purpose of this chart?
- •Comparison !
- •What is wrong whit it?

The last example



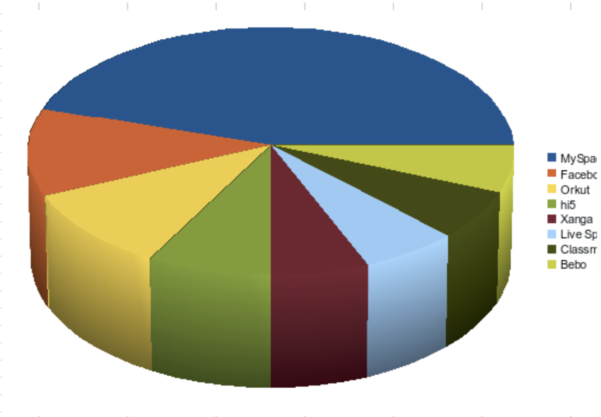
- Is the order clear?
- Which is my company?
- Who is bigger G or A?

Even worst : 3D!!!

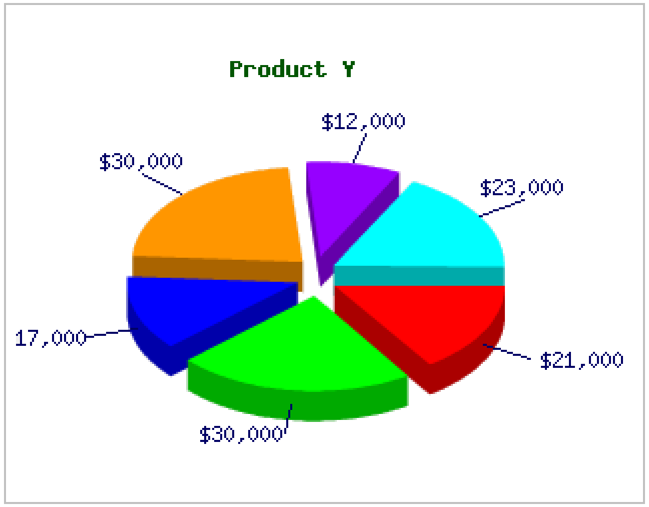


I pie charts!

HATE

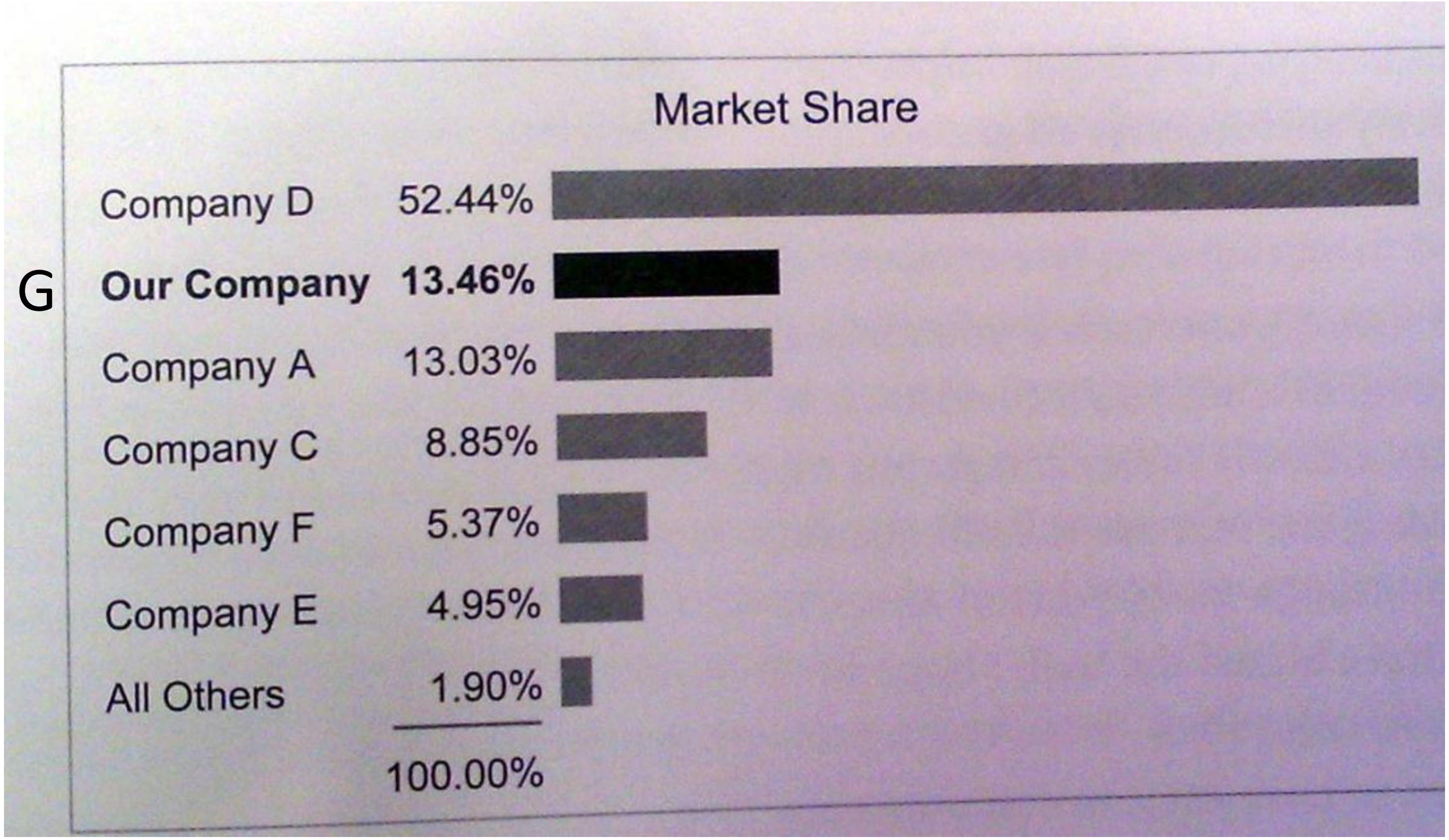


🟦 MySpa
🟧 Facebo
🟨 Orkut
🟩 hi5
🟥 Xanga
🟦 Live Sp
⬛ Classm
🟨 Bebo



Product Y
$12,000
$30,000
$23,000
17,000
$21,000
$30,000

A better solution



If you have ordering (ranking) alternatives think about that!

Chartjunk is not the unique enemy...

- Before PCs, building graphs was a matter of paper and pencil
  - requiring time and effort
  - pushing you to better understand :
    - the meaning of numbers
    - the graph purpose
    - the graph organization
    - …
- now, with Excel you can produce graphs so fast that you might loose control...
  - you select predefined solutions
  - you might not understand how the graph is built (row, columns, headings, ...)
  - you can make mistakes (e.g., missing a row...)

So...

1. Look at the numbers (plus statistics) and at the task

2. Plan a graph (even on the paper!)
   - kind of graph(s) / or even plain numbers
   - label your axes
   - units
   - scale

3. Look for an Excel implementation of your design

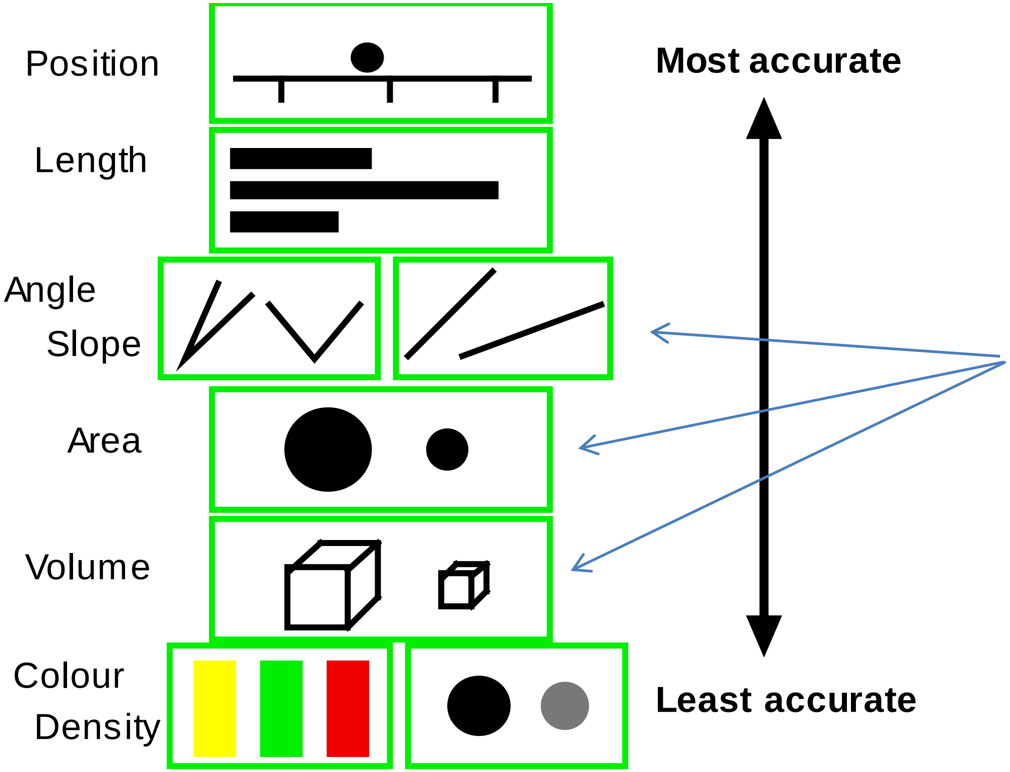
4. If step 3 fails, proceed without Excel ! You can also consider more serious visualization tools, e.g., R (http://cran.r-project.org/bin/windows/base/).

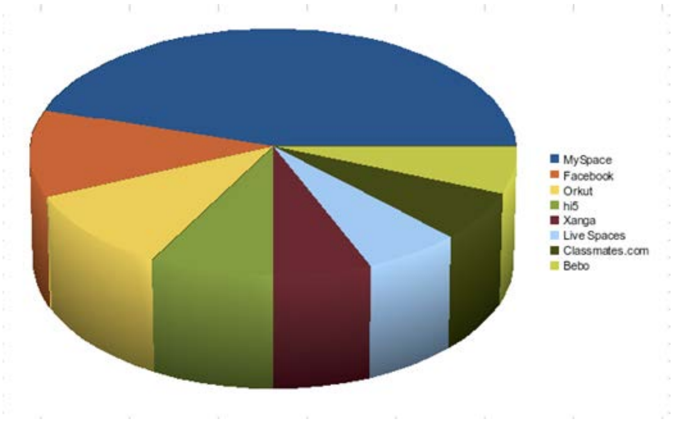
Outline
(basically what you have NOT to do)

- An introductive example
- Good and bad graphs
  - Basic rules
  - Some additional considerations
- Visual issues
  - Quantitative perception (basic rules)
  - The role of interaction
- Two examples for IR

Why do I pie-charts?

The relative difficulty of assessing **quantitative** value as a function of visual encoding mechanism, as established by Cleveland and McGill

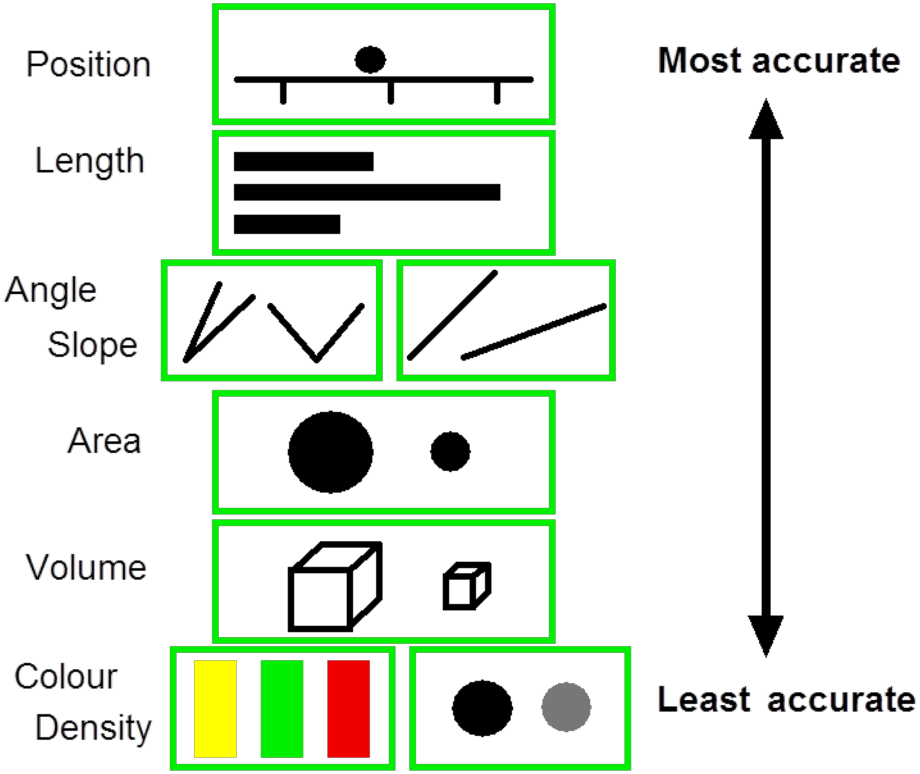




Pie-charts discards the two first choices

I do NOT see ANY reason to use them

What about quantitative comparison?

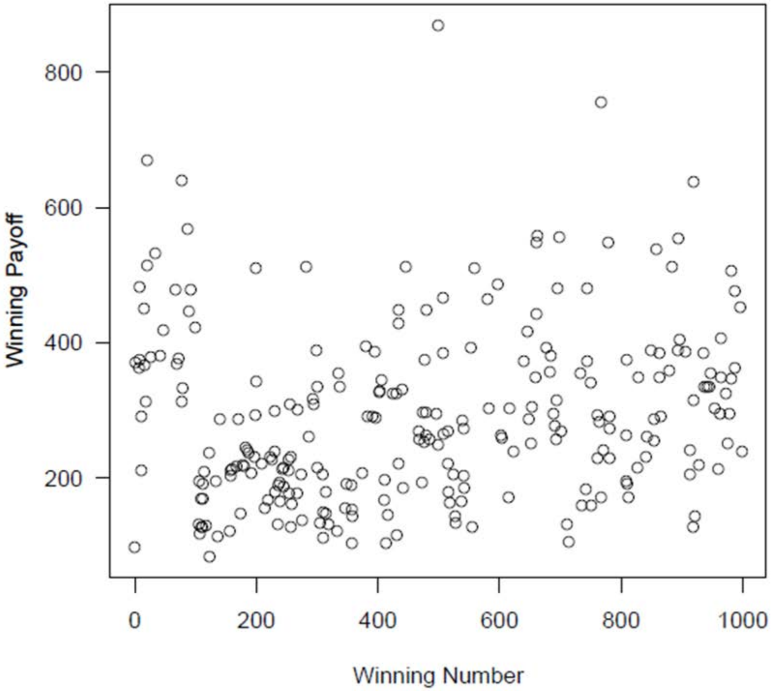


Use position and length
Avoid angles
Avoid areas
Avoid volumes

Use colors carefully

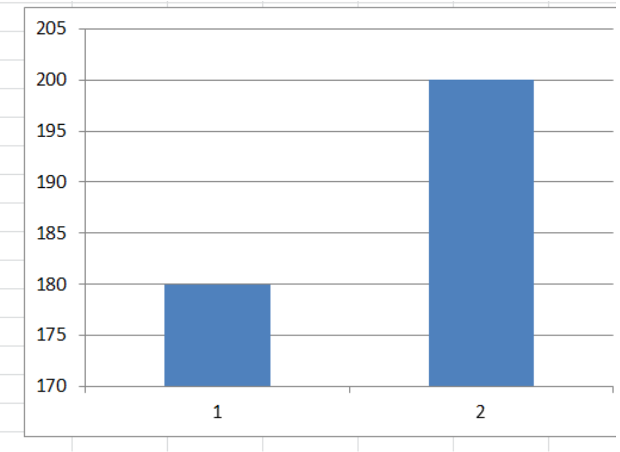
Position

- It works fine

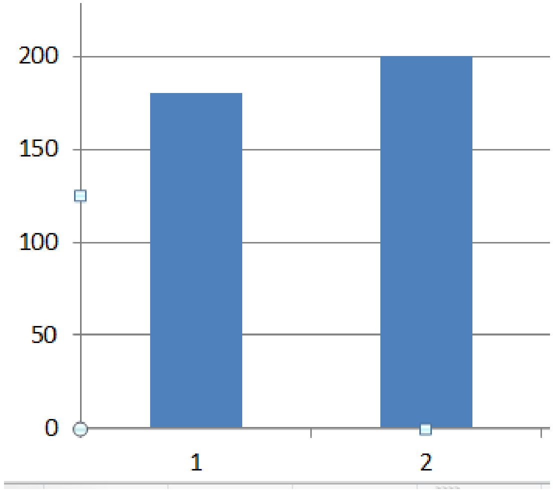


Length?

- Length is fine as well , but use the right scale!



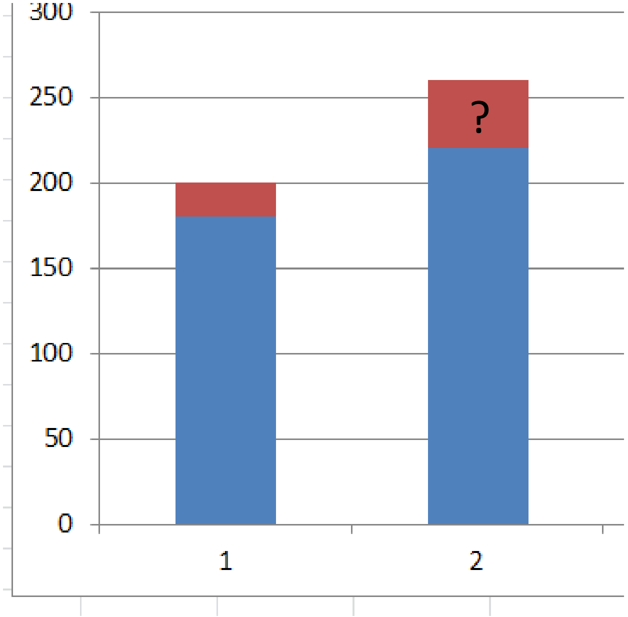
Automatically produced
by Excel

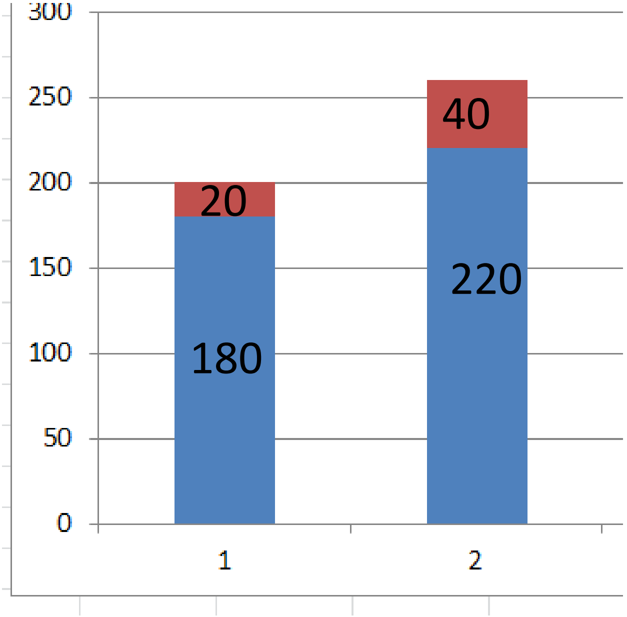


The reality

Length?

- The lookup of precise number might be difficult if the position is not evident (e.g., stacked bar chart)

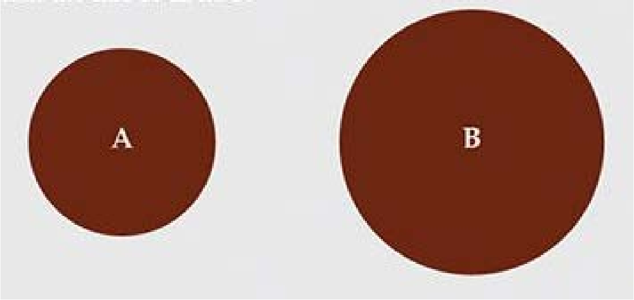




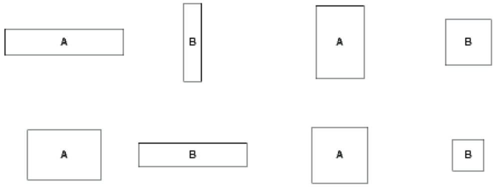
It makes sense to explicitly add figures

Areas: some new surprising issues

- Human being are very bad in estimating area ratios



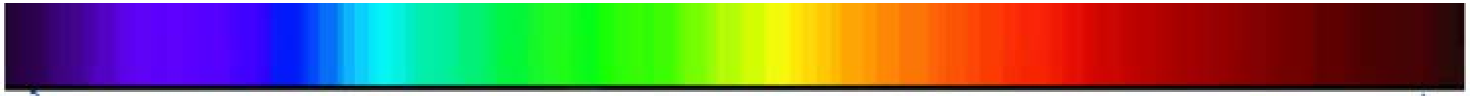
- What is the ratio between this two circles?
  35% 40% 45% 50% 55% 60%   ?
- What is the shape that produces the biggest error?



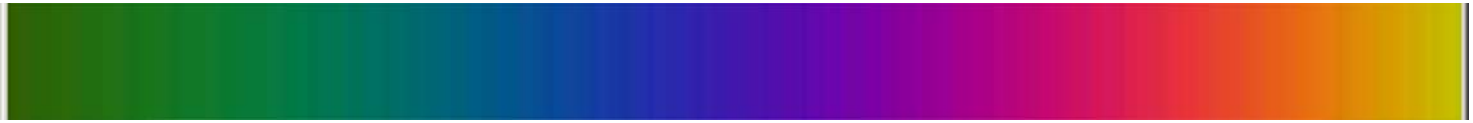
- The square!
- Perceptual Guidelines for Creating Rectangular Treemaps (Nicholas Kong et al., Infovis 2010)

Colors / Numerical data

- Someone already thought how to associate quantitative values to colors and different choices are available
- Do not reinvent the wheel
- (The rainbow scale does not work)



rainbow scale



HSI color model
(Keim and Kriegel) - Issues in visualizing
large databases. Proc. of the IFIP working conference
on Visual database Systems, 1995

Other choices (Colin Ware)

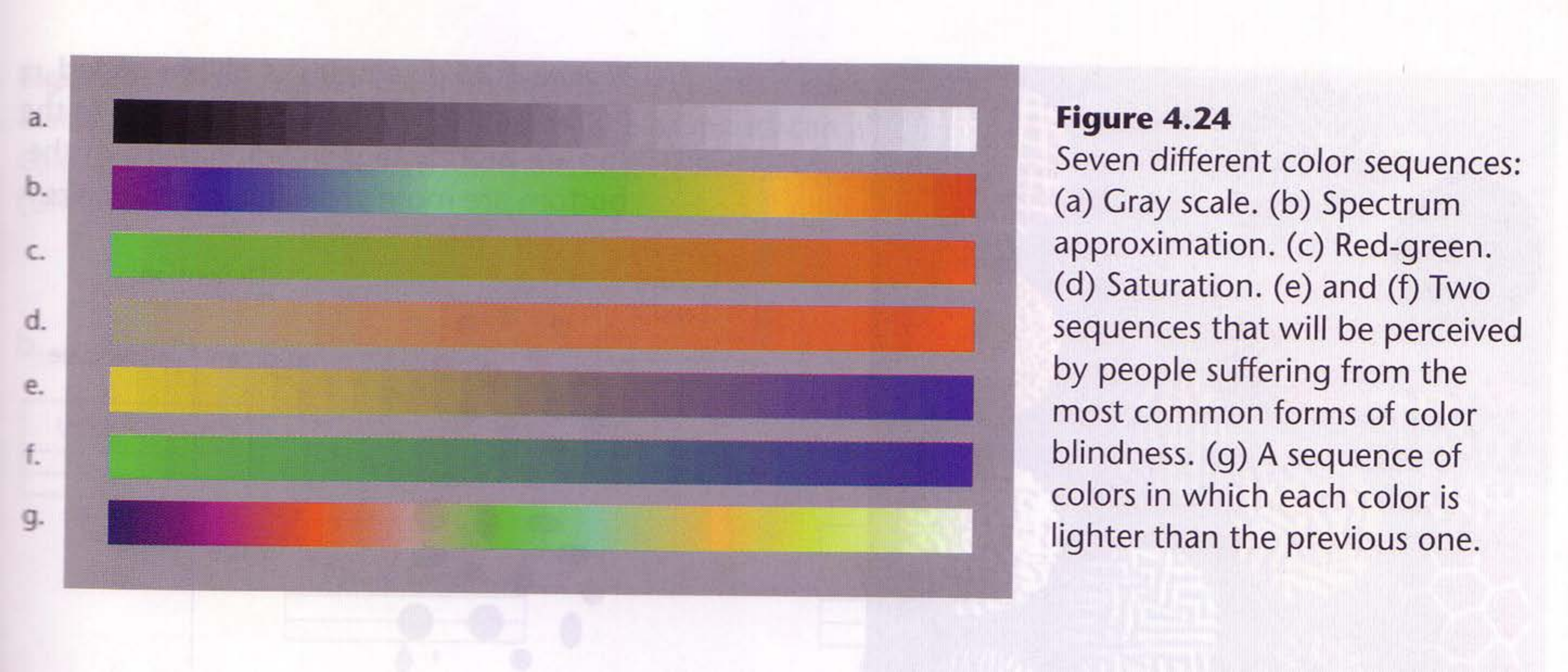
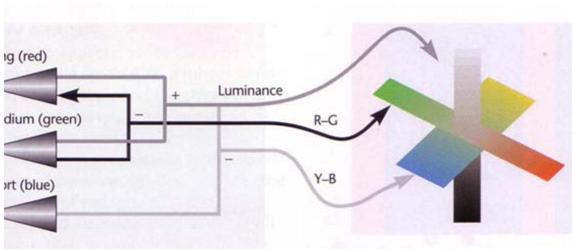
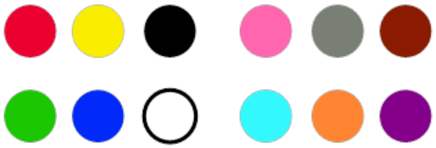


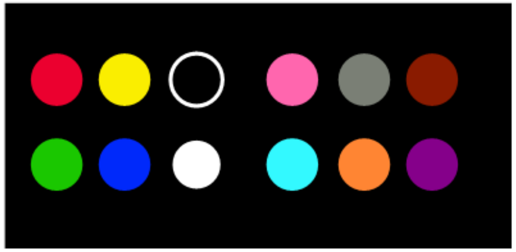
Figure 4.24
Seven different color sequences: (a) Gray scale. (b) Spectrum approximation. (c) Red-green. (d) Saturation. (e) and (f) Two sequences that will be perceived by people suffering from the most common forms of color blindness. (g) A sequence of colors in which each color is lighter than the previous one.

Colors /Categorical data



- Colors are fine with categorical data
- Do not reinvent the wheel (again)
- The Ewald Hering idea is that there are only 6 elementary colors arranged in three pairs
- That gives us up to 12 (6+6) colors easily distinguishable (11!)



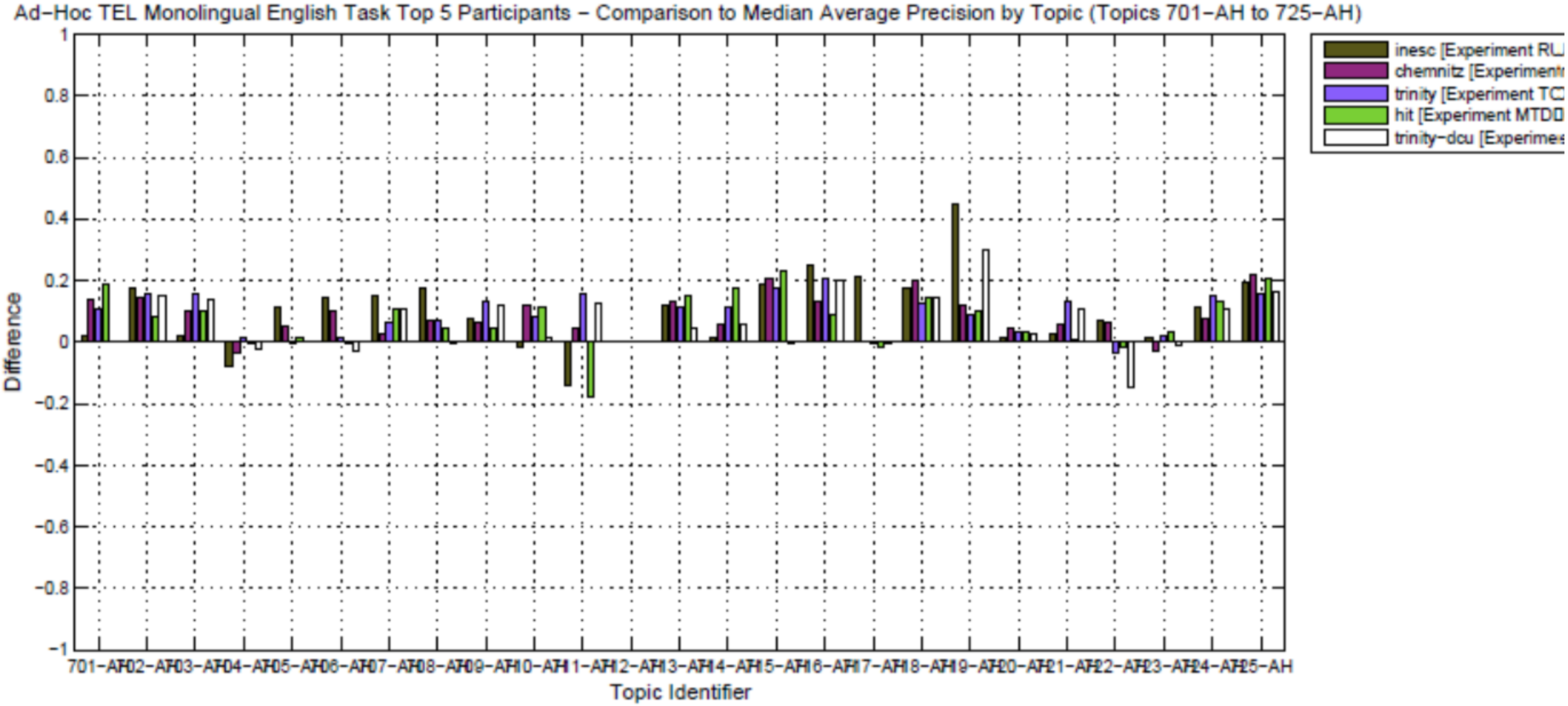


12 Colors
for labeling

Outline
(basically what you have NOT to do)

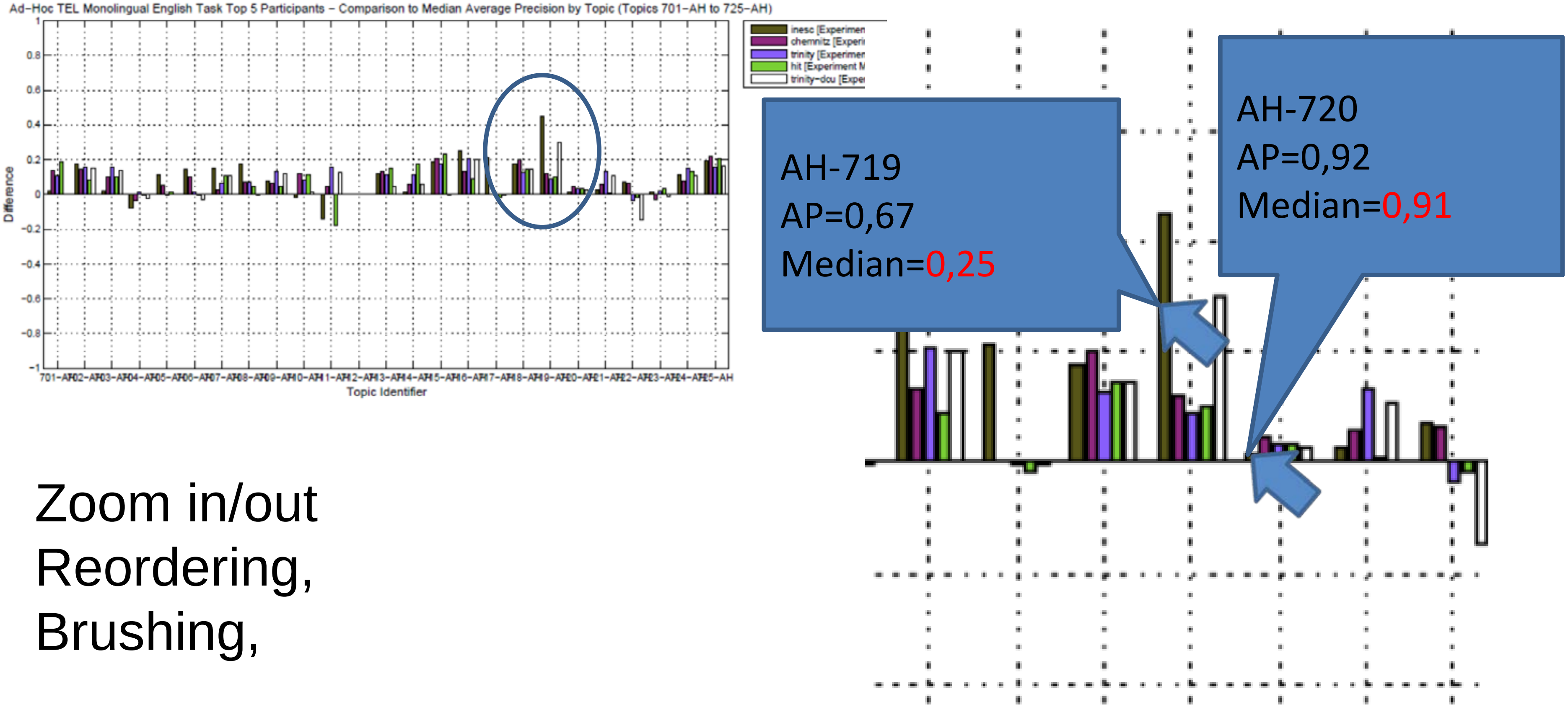
- An introductive example
- Good and bad graphs
  - Basic rules
  - Some additional considerations
- Visual issues
  - Quantitative perception (basic rules)
  - The role of interaction
- Two examples for IR

Interaction?



- Average precision (y axis) compared to the topic median (5 experiments)

Interaction ?



AH-719
AP=0,67
Median=0,25

AH-720
AP=0,92
Median=0,91

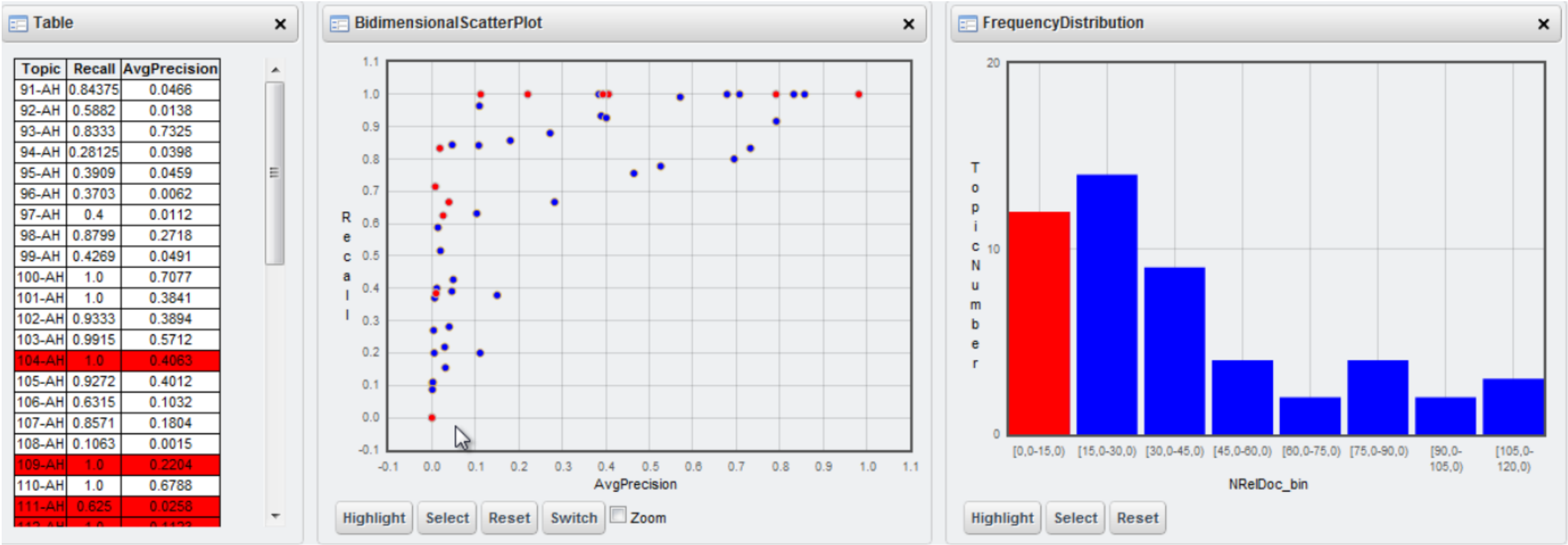
Zoom in/out
Reordering,
Brushing,
...

Winter School 2012
Zinal, Valais - Switzerland
23 - 27 January 2012

Outline
(basically what you have NOT to do)

- An introductive example
- Good and bad graphs
  - Basic rules
  - Some additional considerations
- Visual issues
  - Quantitative perception (basic rules)
  - The role of interaction
- Two examples for IR

Parallel coordinated views



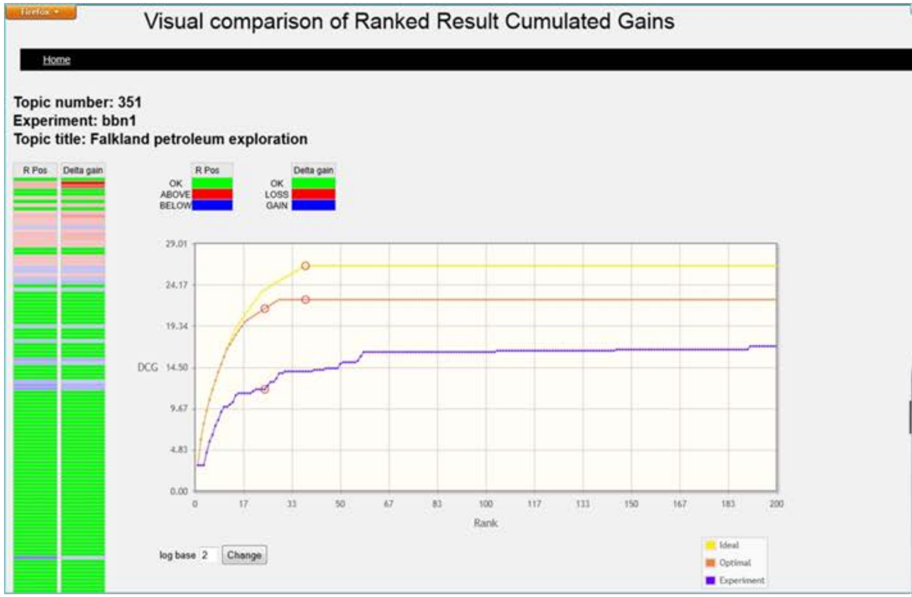
Topic	Recall	AvgPrecision
91-AH	0.84375	0.0466
92-AH	0.5882	0.0138
93-AH	0.8333	0.7325
94-AH	0.28125	0.0398
95-AH	0.3909	0.0459
96-AH	0.3703	0.0062
97-AH	0.4	0.0112
98-AH	0.8799	0.2718
99-AH	0.4269	0.0491
100-AH	1.0	0.7077
101-AH	1.0	0.3841
102-AH	0.9333	0.3894
103-AH	0.9915	0.5712
104-AH	1.0	0.4063
105-AH	0.9272	0.4012
106-AH	0.6315	0.1032
107-AH	0.8571	0.1804
108-AH	0.1063	0.0015
109-AH	1.0	0.2204
110-AH	1.0	0.6788
111-AH	0.625	0.0258
112-AH	1.0	0.1123

Table
- Topic
- Recall
- AvgPrecision

Scatterplot
- Y Recall
- X AvgPrecision

Histogram
- X Number of relevant docs
  - bin = 15
- Y Number of topics in the interval

Rank analysis

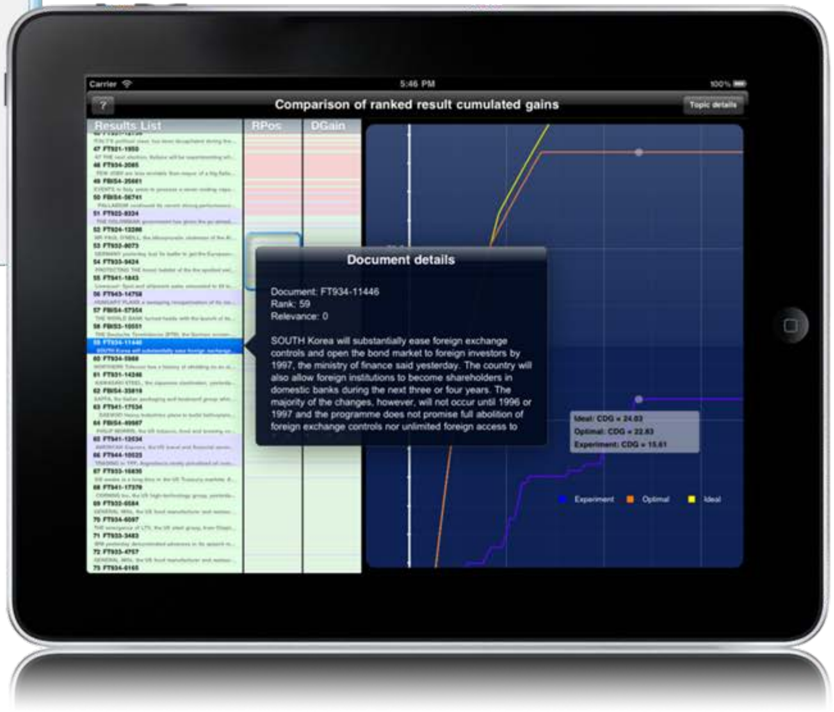


Visual comparison of Ranked Result Cumulated Gains

Home

Topic number: 351
Experiment: bbn1
Topic title: Falkland petroleum exploration

log base 2 Change



Comparison of ranked result cumulated gains

47 FT921-1950
48 FT934-2085
49 FBIS4-25661
50 FBIS4-56741
51 FT922-8334
52 FT924-13286
53 FT932-8073
54 FT933-9424
55 FT941-1843
56 FT943-14758
57 FBIS4-57354
58 FBIS3-10551
59 FT934-11446
SOUTH Korea will substantially ease foreign exchange...
60 FT934-5968
61 FT931-14246
62 FBIS4-33816
63 FT941-17534
64 FBIS4-49987
65 FT941-12634
66 FT944-10525
67 FT933-16836
68 FT941-17379
69 FT932-6584
70 FT934-6097
71 FT933-3483
72 FT935-4757
73 FT934-6165

Document details

Document: FT934-11446
Rank: 59
Relevance: 0

SOUTH Korea will substantially ease foreign exchange controls and open the bond market to foreign investors by 1997, the ministry of finance said yesterday. The country will also allow foreign institutions to become shareholders in domestic banks during the next three or four years. The majority of the changes, however, will not occur until 1996 or 1997 and the programme does not promise full abolition of foreign exchange controls nor unlimited foreign access to

Ideal: CDG = 24.03
Optimal: CDG = 22.83
Experiment: CDG = 15.61

Rank analysis (relevance 0-3)

The actual result

GT(V)	DF	DCG
🟩 3	3,00	3,00
🟥 1	1,00	4,00
🟥 2	1,26	5,26
🟩 3	1,50	6,76
🟩 2	0,86	7,62
🟩 2	0,77	8,40
🟦 3	1,07	9,47
🟩 2	0,67	10,13
🟥 0	0,00	10,13
🟩 1	0,30	10,43
🟩 0	0,00	10,43
🟦 3	0,84	11,27

🟩 OK

🟥 ABOVE

🟦 BELOW

The optimal result

GT(O)	DF	DCG
3	3,00	3,00
3	3,00	6,00
3	1,89	7,89
3	1,50	9,39
2	0,86	10,25
2	0,77	11,03
2	0,71	11,74
2	0,67	12,41
1	0,32	12,72
1	0,30	13,02
0	0,00	13,02
0	0,00	13,02

Books worth to read

- Stephen Few - Show me the number - Analytic press
- Stephen Few - Now You See It: Simple Visualization Techniques for Quantitative Analysis - Analytic press
- Robert Spence - Information Visualization: Design for Interaction (2nd Edition) - Addison-Wesley (ACM Press)
- Edward Tufte - The Visualization of quantitative information - Graphics Pr
- Colin Ware - Information Visualization, Third Edition: Perception for Design (Interactive Technologies) - Morgan Kaufmann