

# Visual Analytics

Giuseppe Santucci

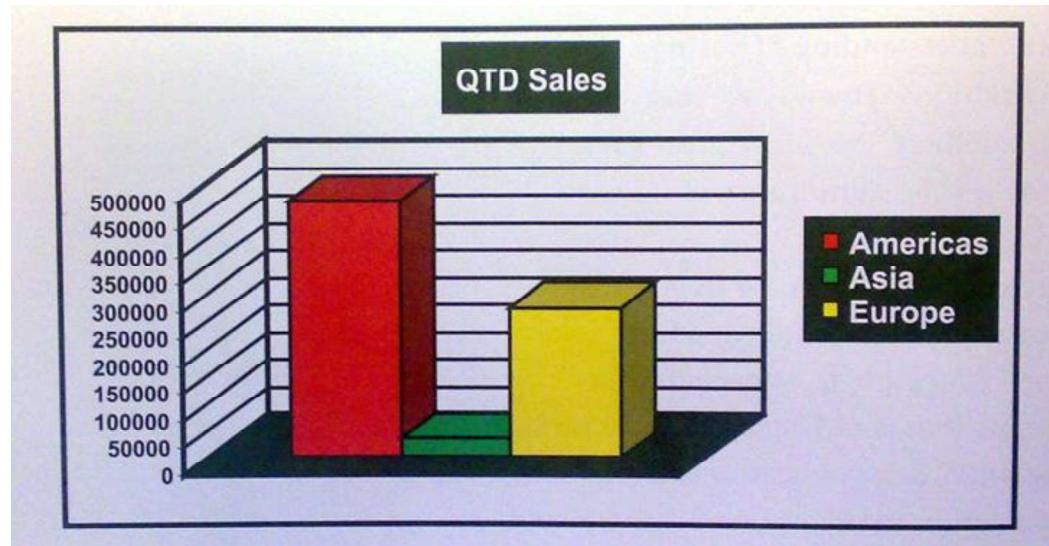
## 3 – Visualizing quantitative Information

# Outline

- New ideas about good and bad graphs
- Meaning of numbers
- Tables and graphs

# An 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



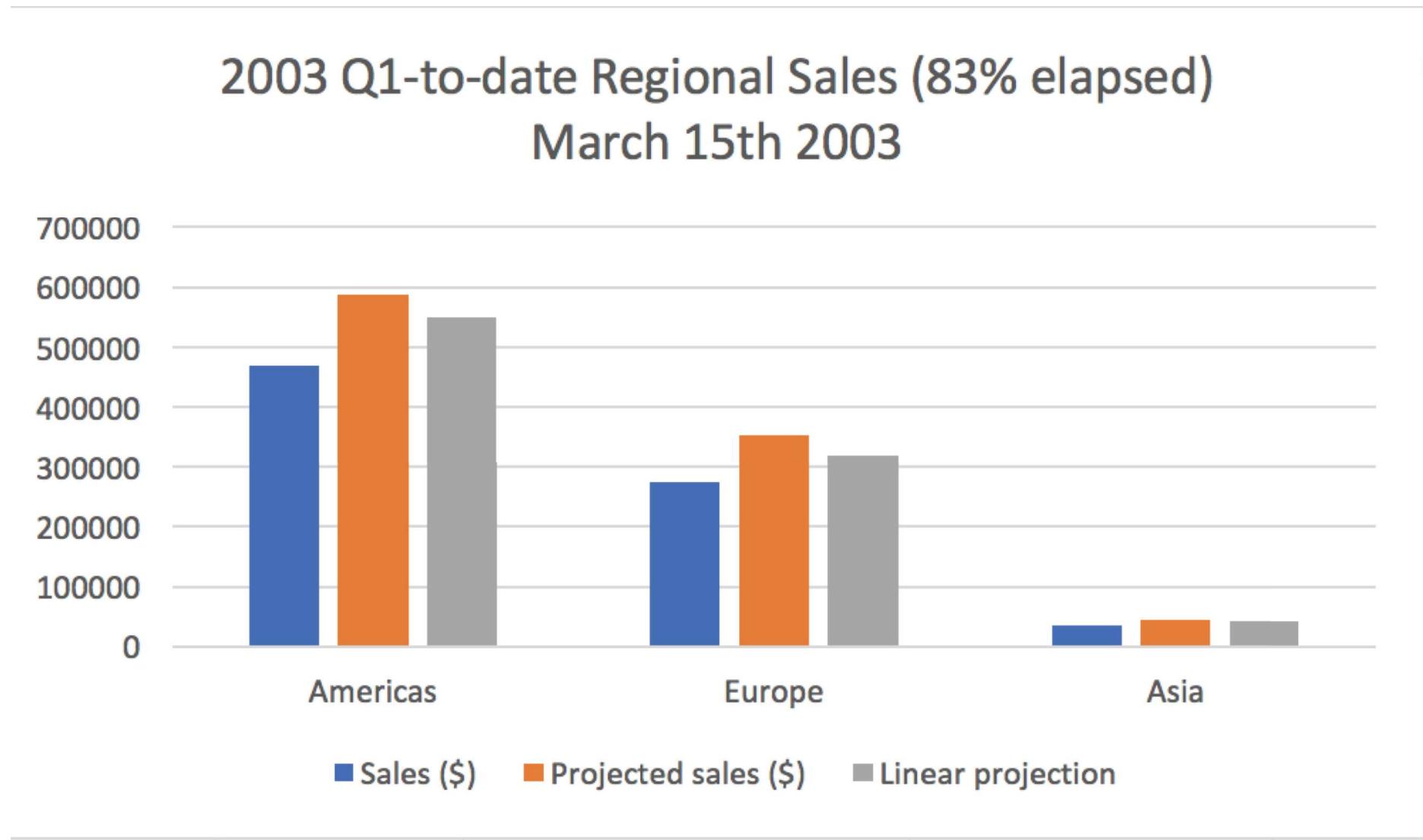
- Are you happy with it?
- Think how to design something that is more informative for your job

# All the needed information (plus some support to the task)

2003 Q1-to-date Regional Sales - Elapsed quarter: 83%				
March 15th 2003				
	Sales (\$)	Projected sales (\$)	Linear projection	Actual progress
Americas	469384	586730	549179	0.80
Europe	273854	353272	320409	0.78
Asia	34847	43210	40771	0.81

- Units !
- The actual **date** and quarter progress (**task!**)!
- Some additional summarizing information (sales progress)
- Planned sales vs actual sales (according to the **task!**)
- Linear forecast

# A better quantitative visualization (even with Excel...)



Comparison is a frequent task!

# Comparing with other products

Is it ok?

Perceptual (color)

and

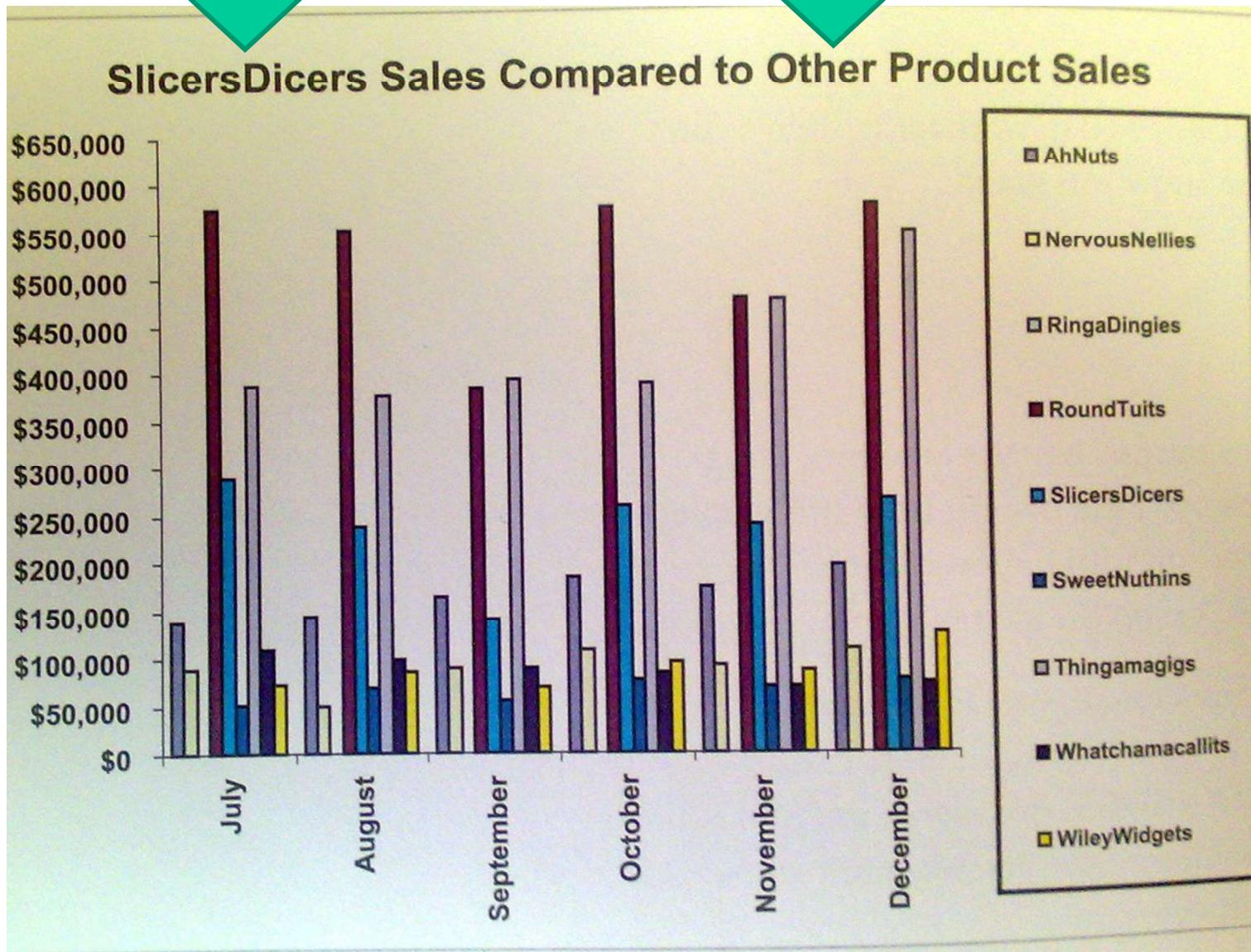
HCI (recognition rather than recall)

issues

The focus is the **comparison** of your product:

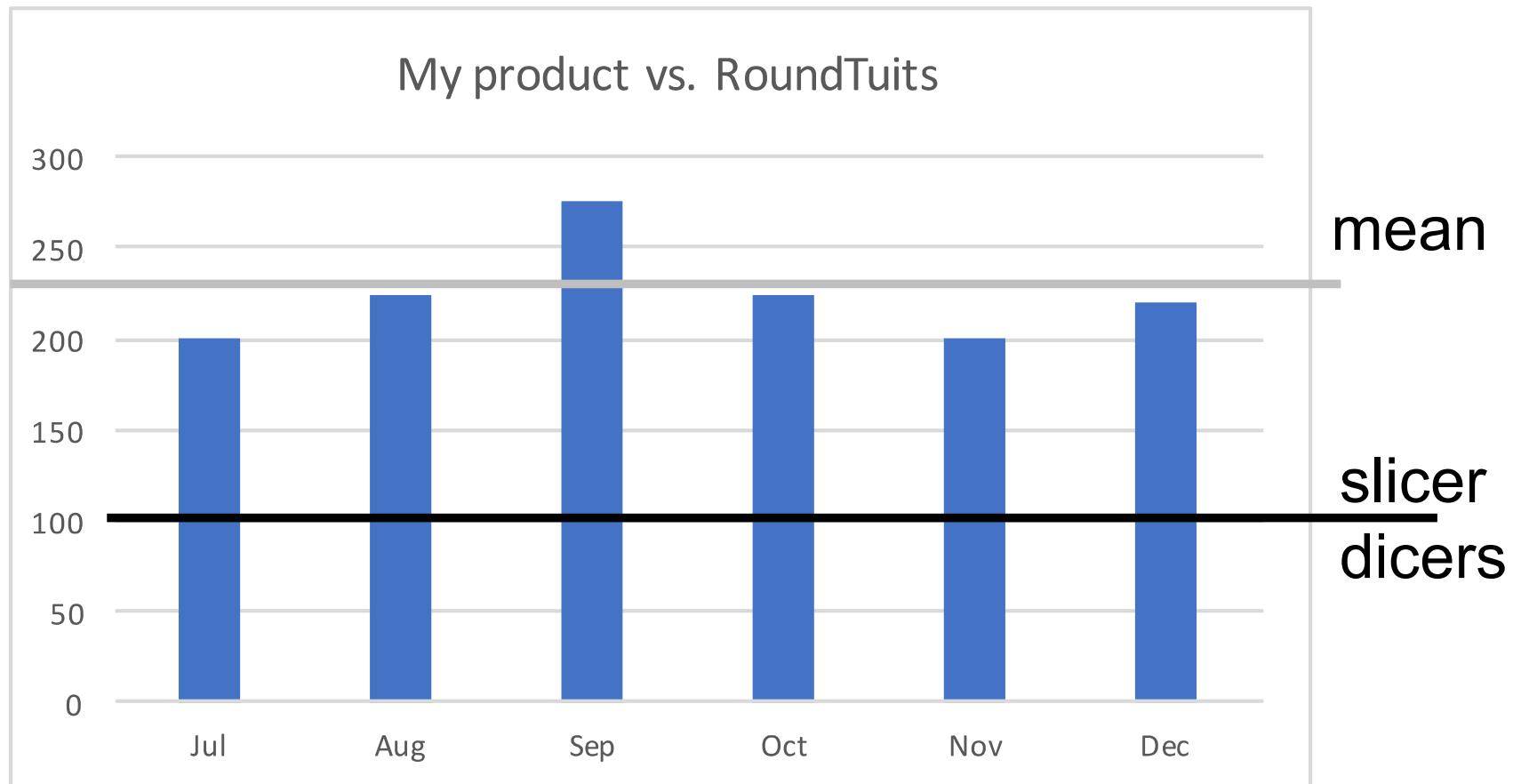
- vs others
- on 6 months

Let's start with just one competitor...



# Comparison percentage!

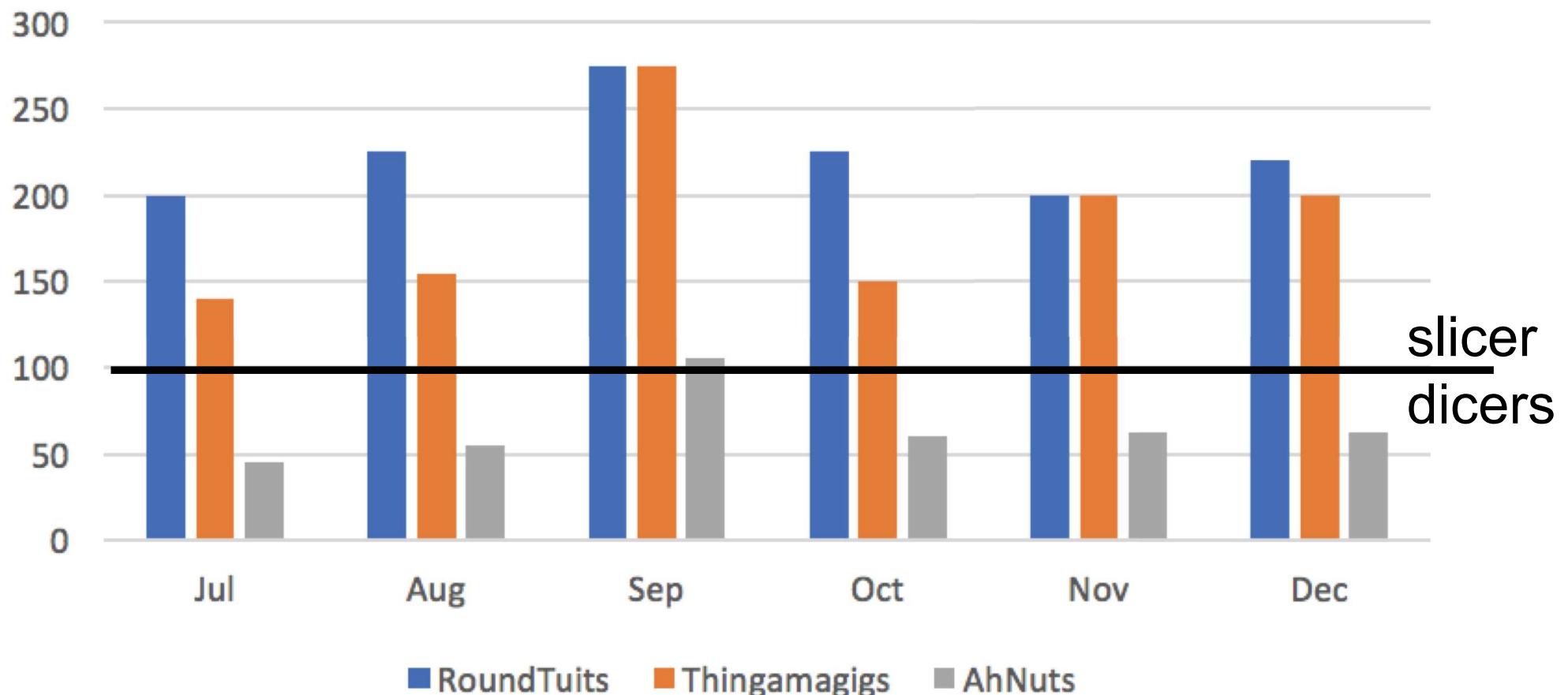
Not supported by Excel



# Comparison!

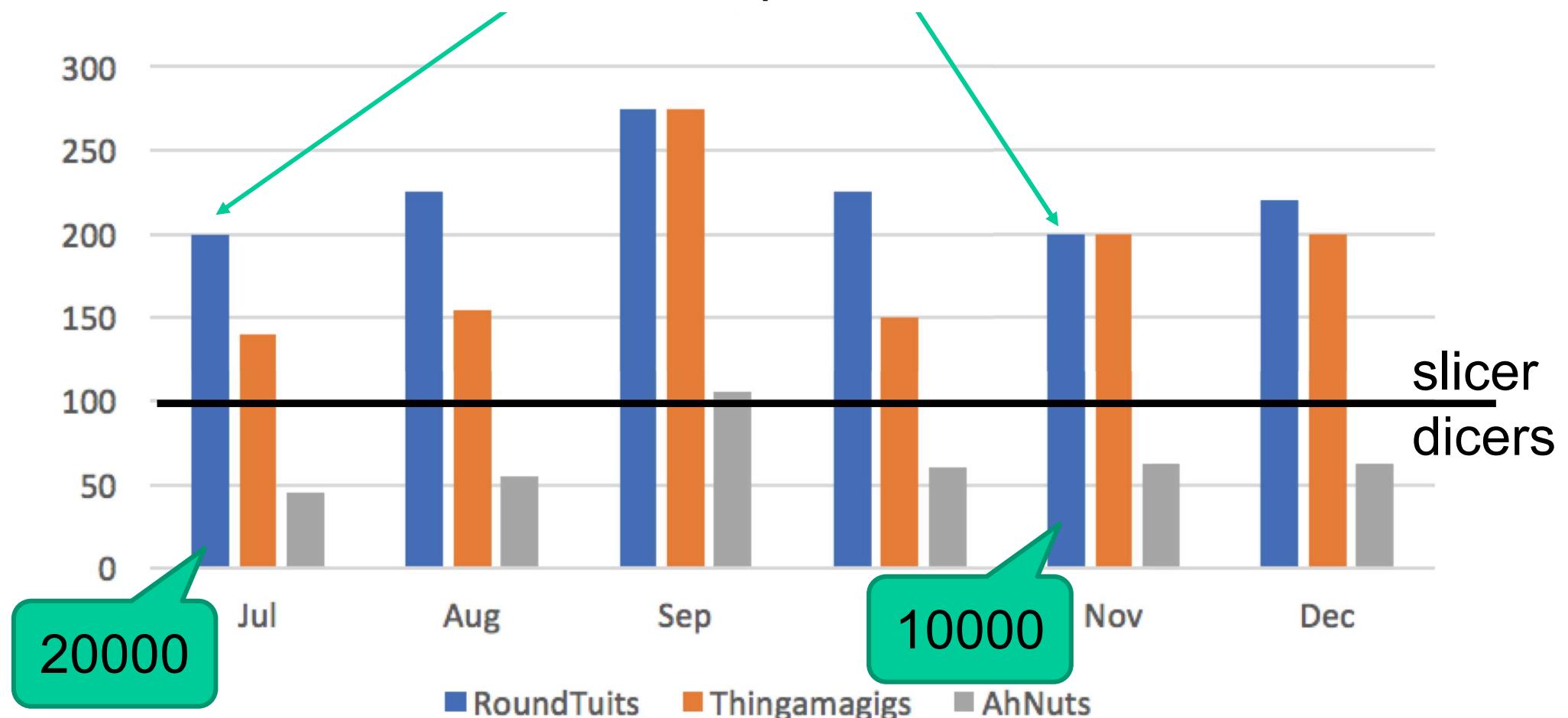
Not supported by Excel

## Slicer Dicers vs 3 competitors

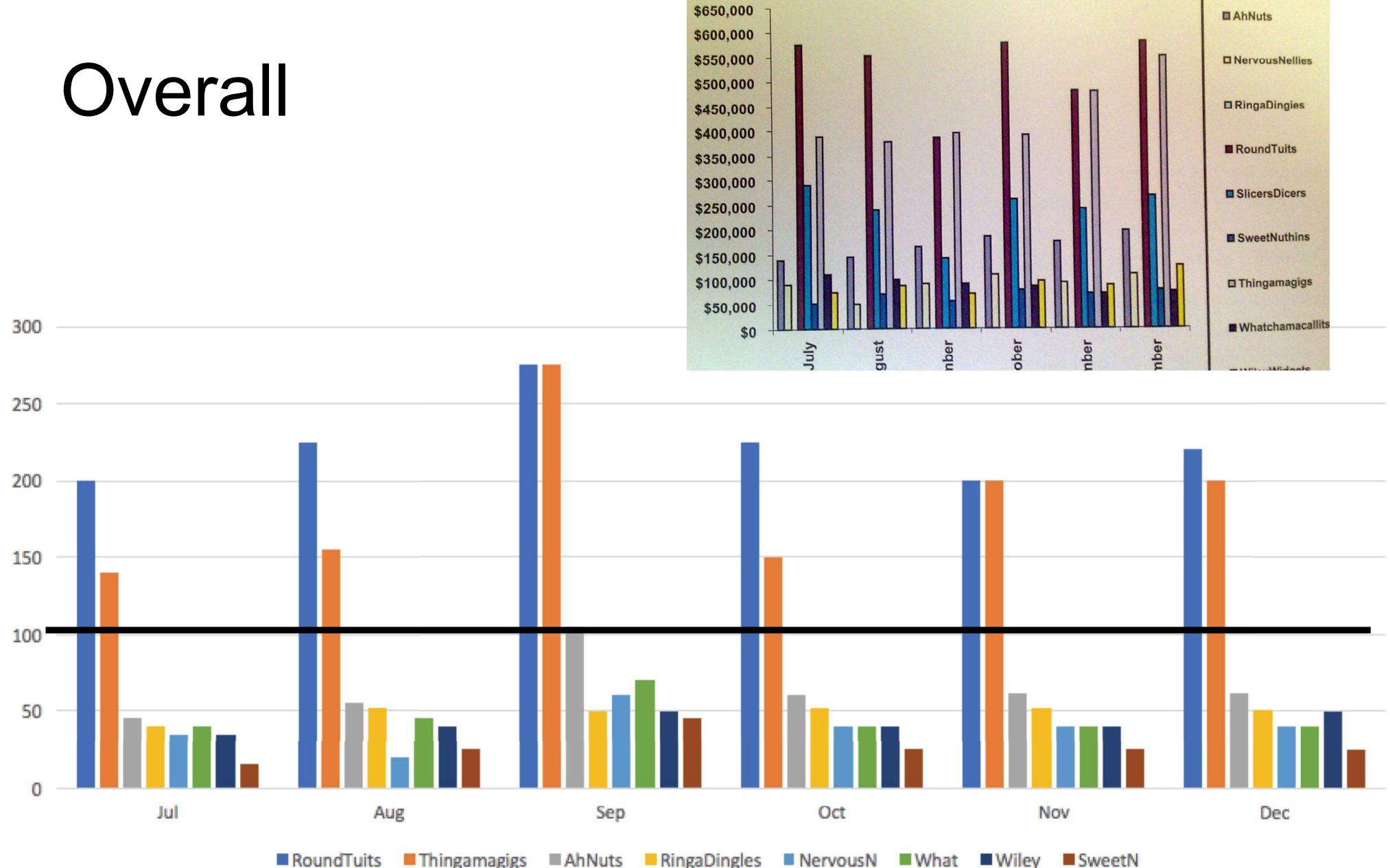


# Be careful! Are these the same?

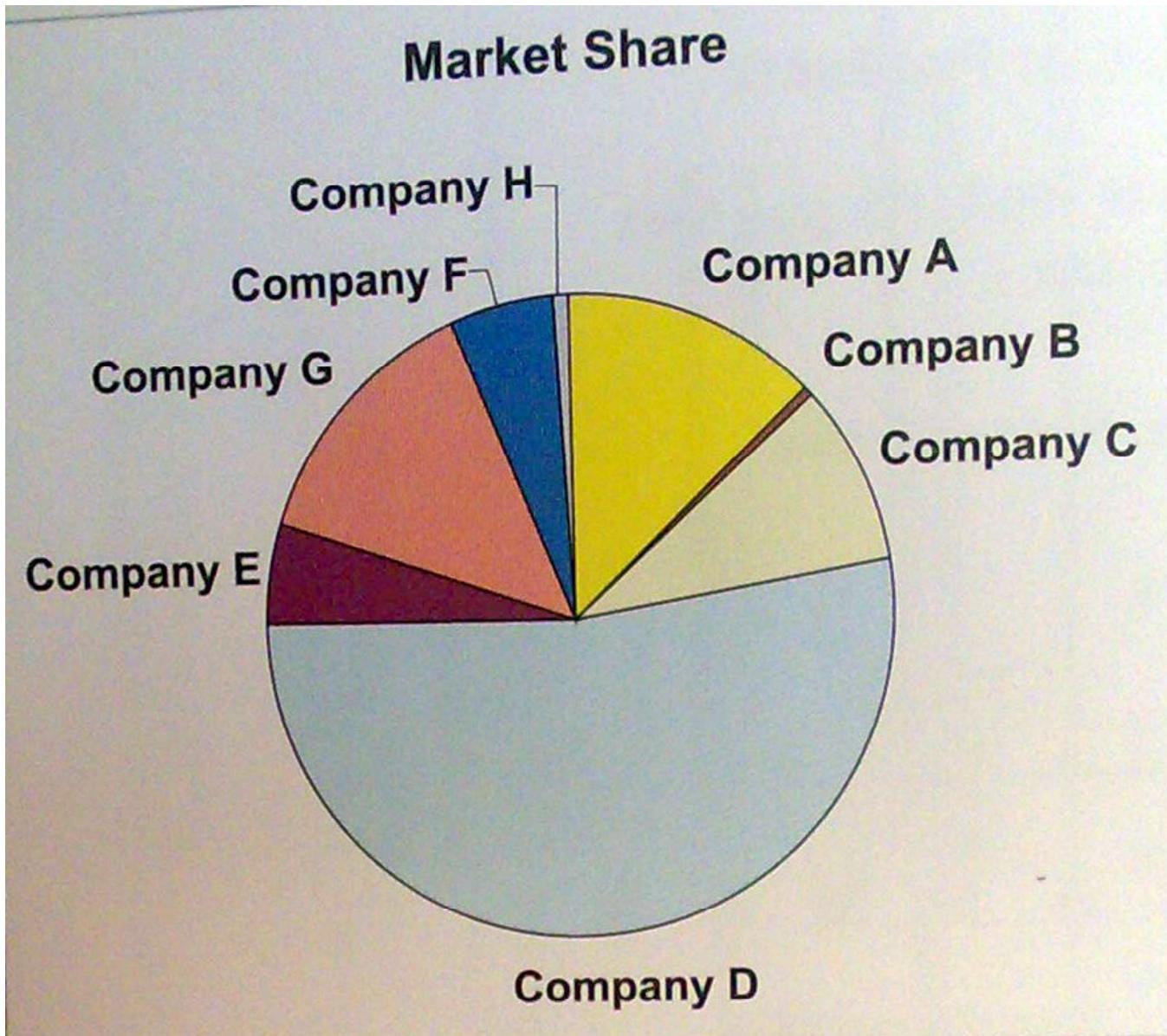
Slicer Dicers vs 3 competitors



# Overall

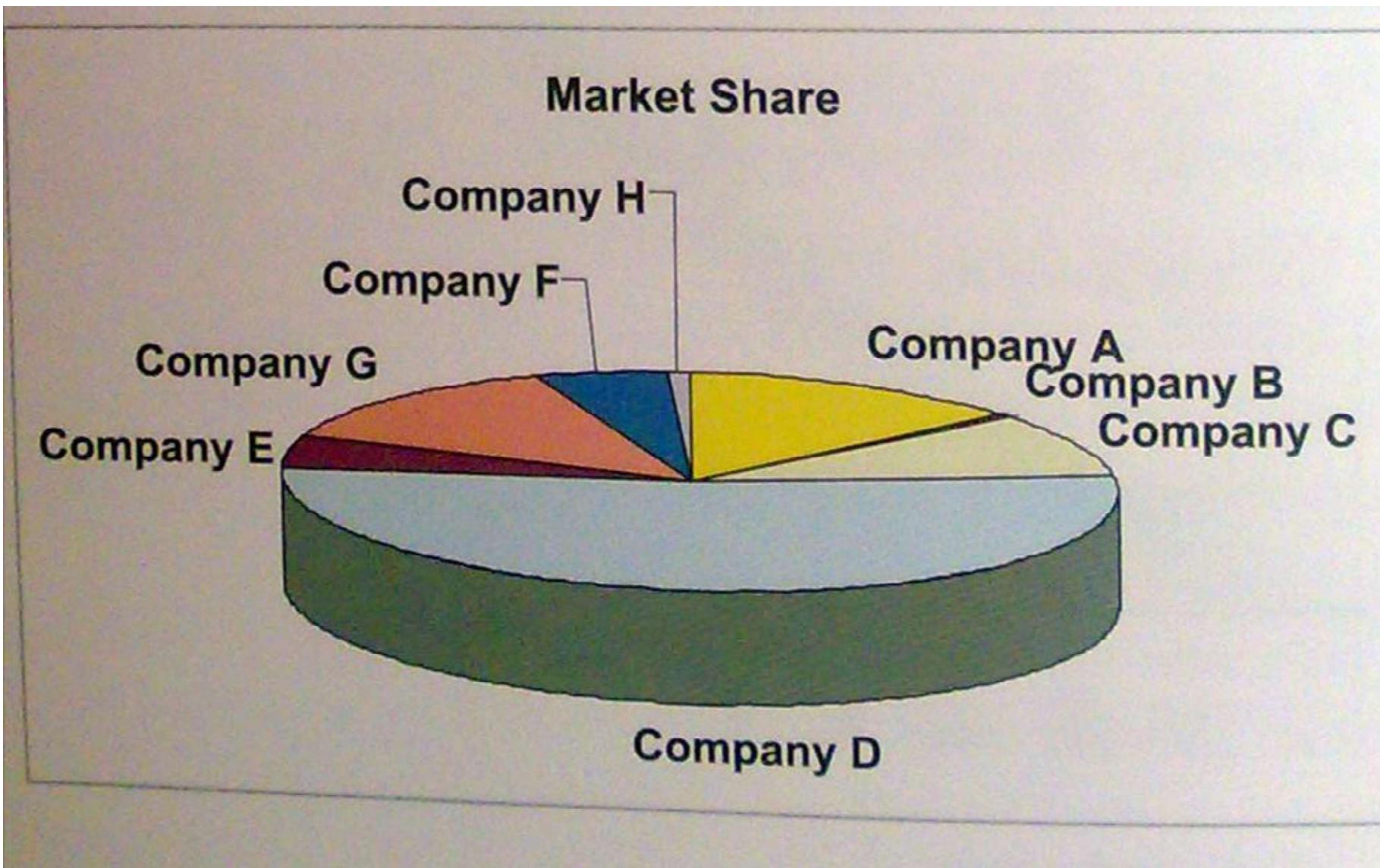


# Another example of comparison: our company against the world!

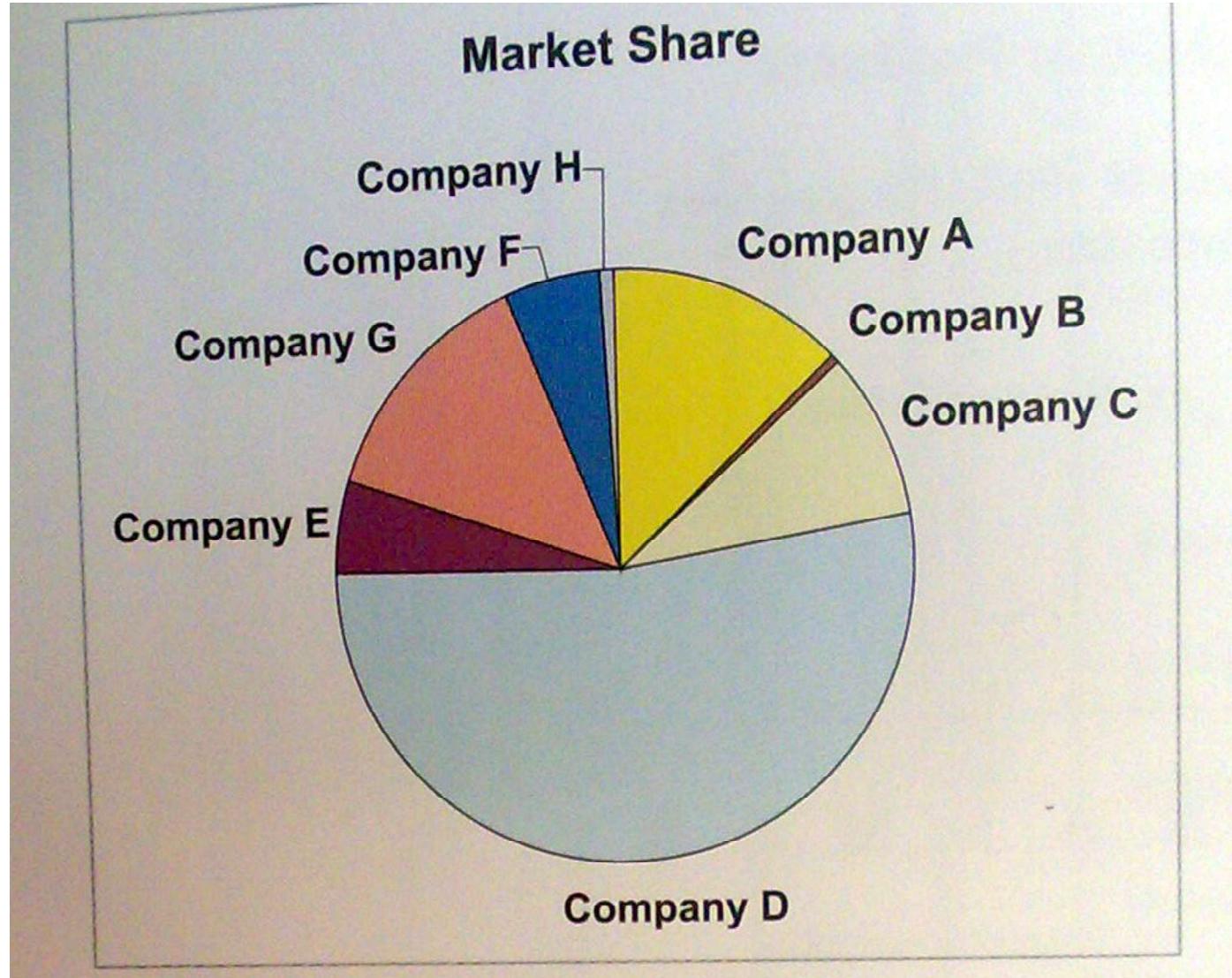


- What is the purpose of this chart?
- What is wrong with it?

# Even worst : 3D!!!



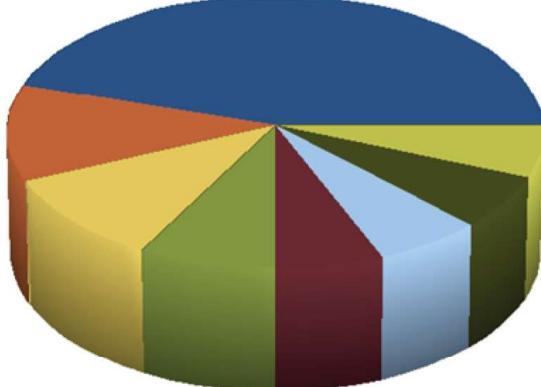
# Issues



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



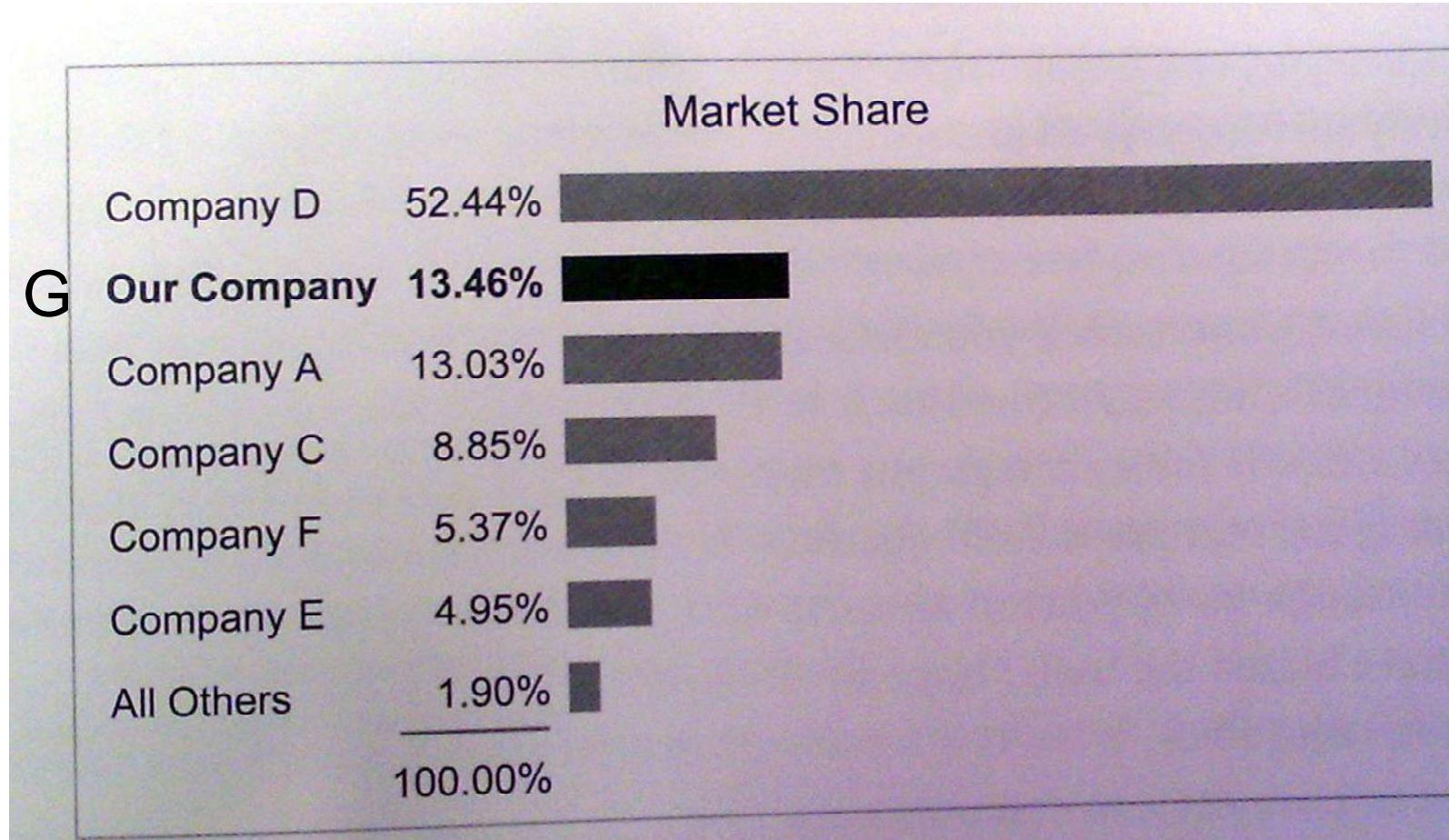
pie charts!



# At least most of them...



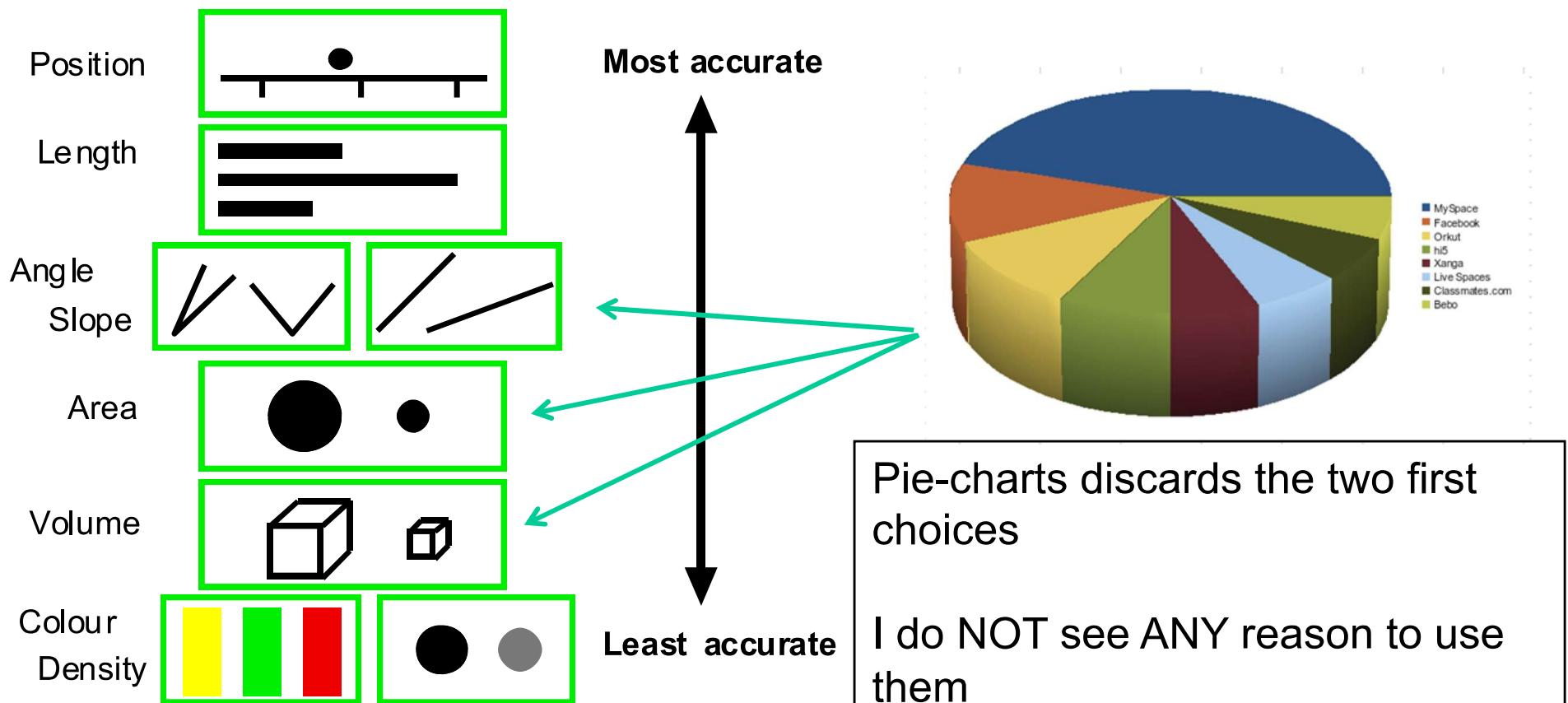
# A better solution



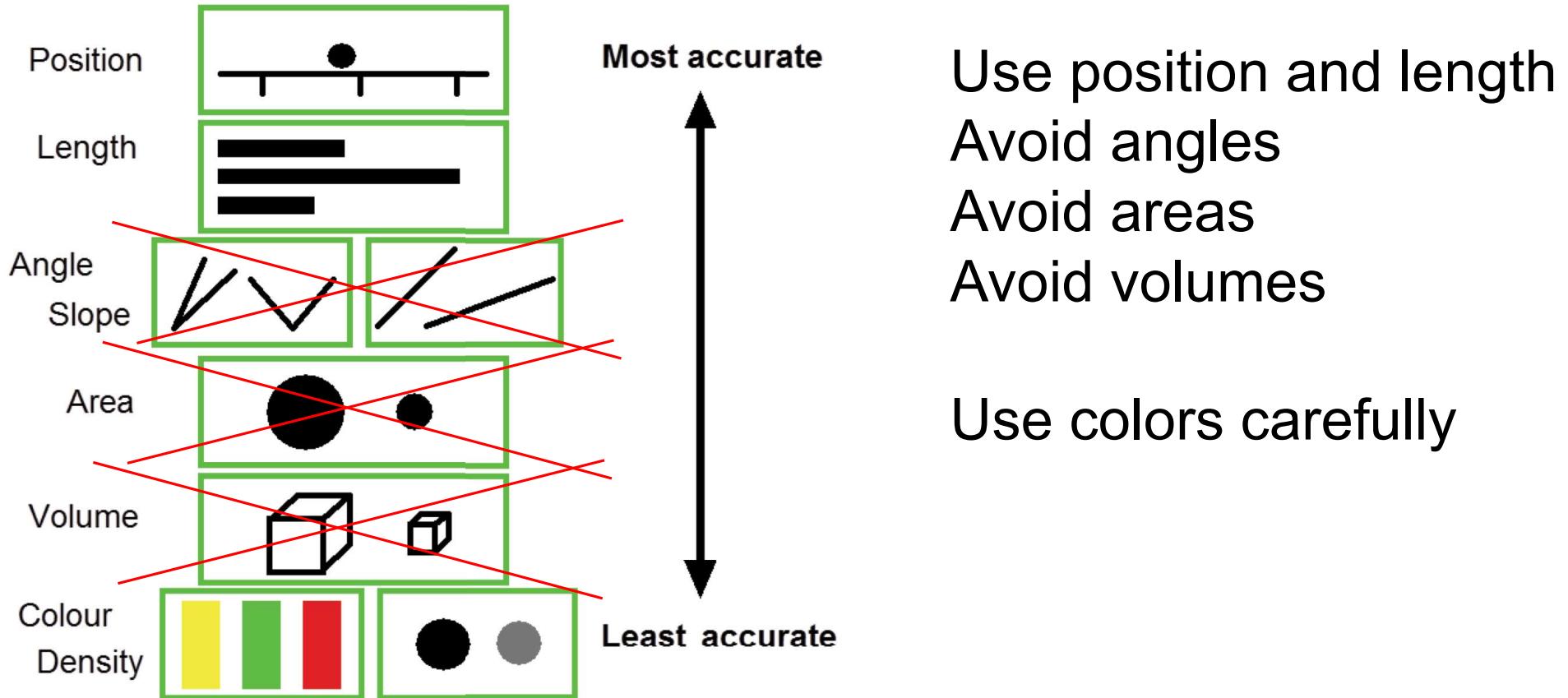
If you have ordering (ranking) alternatives think about that!

# 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 (1984) and restated on 2010



# What about quantitative comparison?

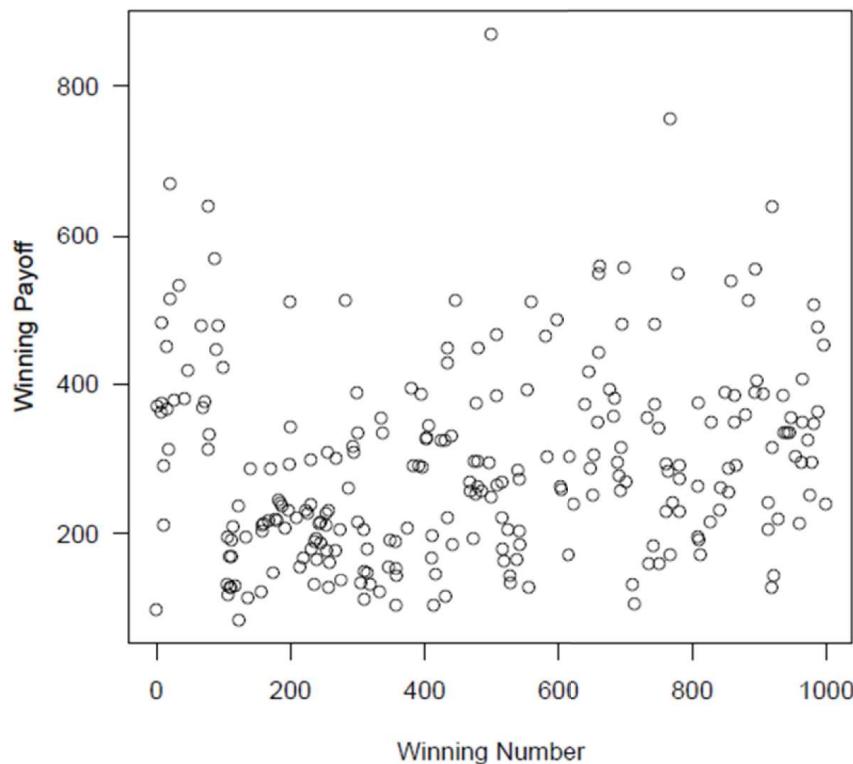


Use position and length  
Avoid angles  
Avoid areas  
Avoid volumes

Use colors carefully

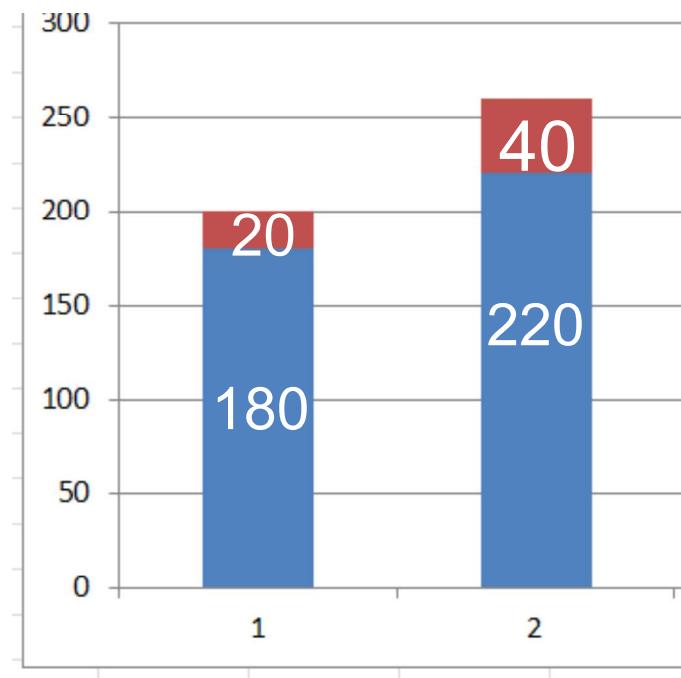
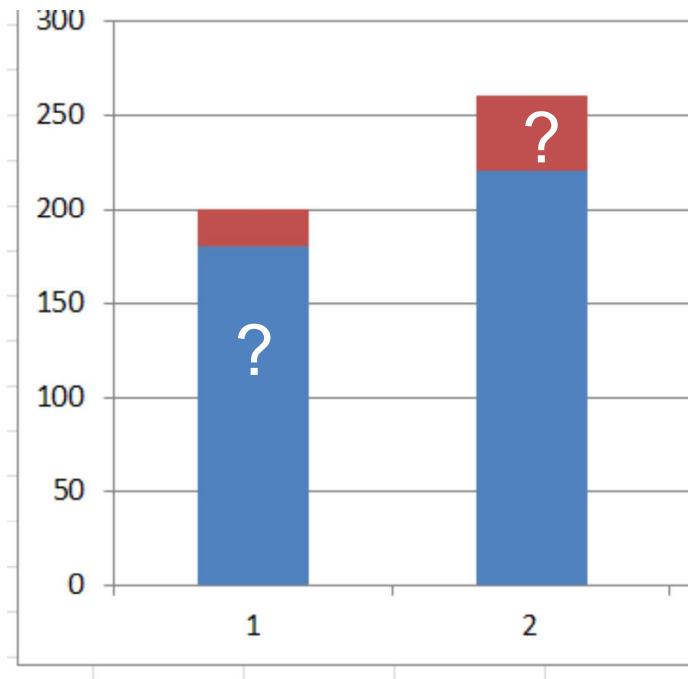
# Position

- It works fine



# Length?

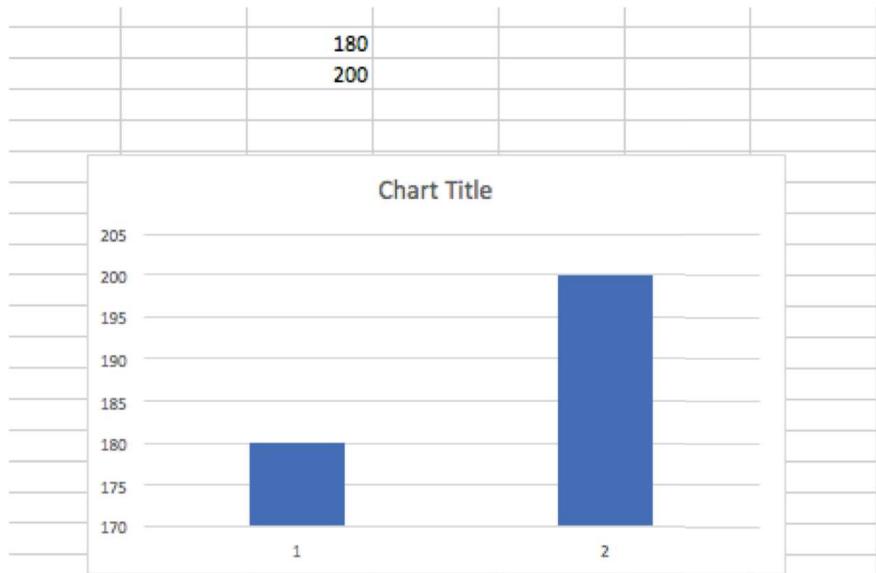
- The lookup of precise number might be difficult if the position is not evident (e.g., stacked bar chart)



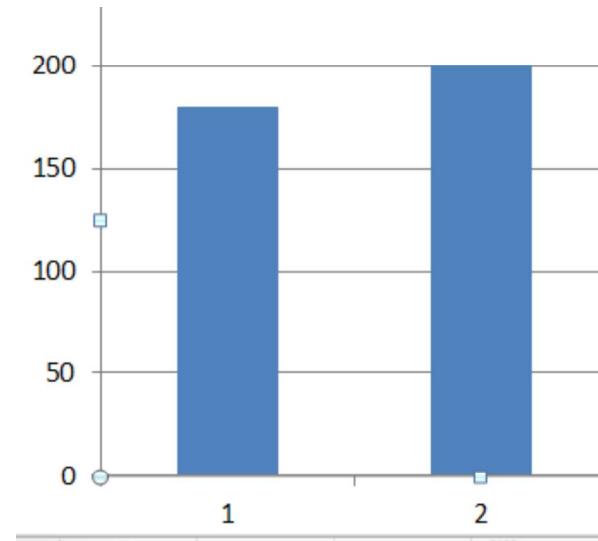
It makes sense to explicitly add figures

# Length?

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



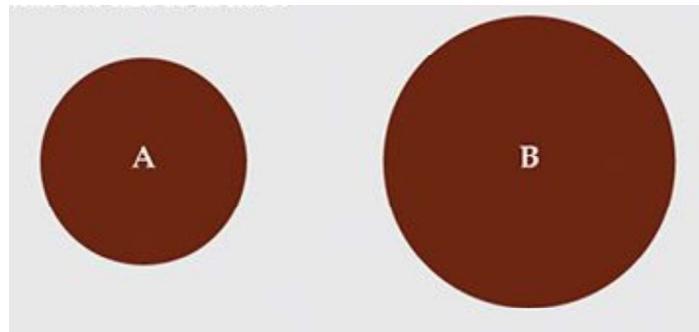
Automatically produced  
by Excel



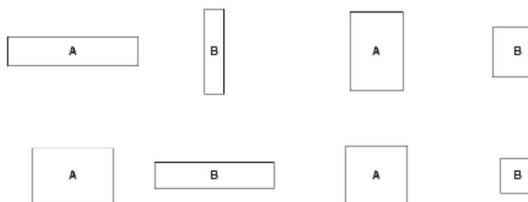
The reality

# Areas: some new surprising issues

- Human being are very bad in estimating area ratios



- What is the ratio between this two circles A/B ?  
25% 35% 40% 45% 50% 55% 60% 70% ?
- 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!!!
- It is a very common error (I did it as well, 20 years ago...)



rainbow scale



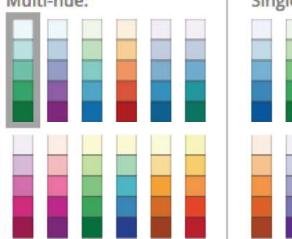
HSI color model

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

# colorbrewer2.org

Number of data classes: 3

Nature of your data:  
 sequential  diverging  qualitative

Pick a color scheme:  
Multi-hue:  Single hue: 

Only show:  
 colorblind safe  
 print friendly  
 photocopy safe

Context:  
 roads  
 cities  
 borders

Background:  
 solid color  terrain  
 color transparency

how to use | updates | downloads | credits

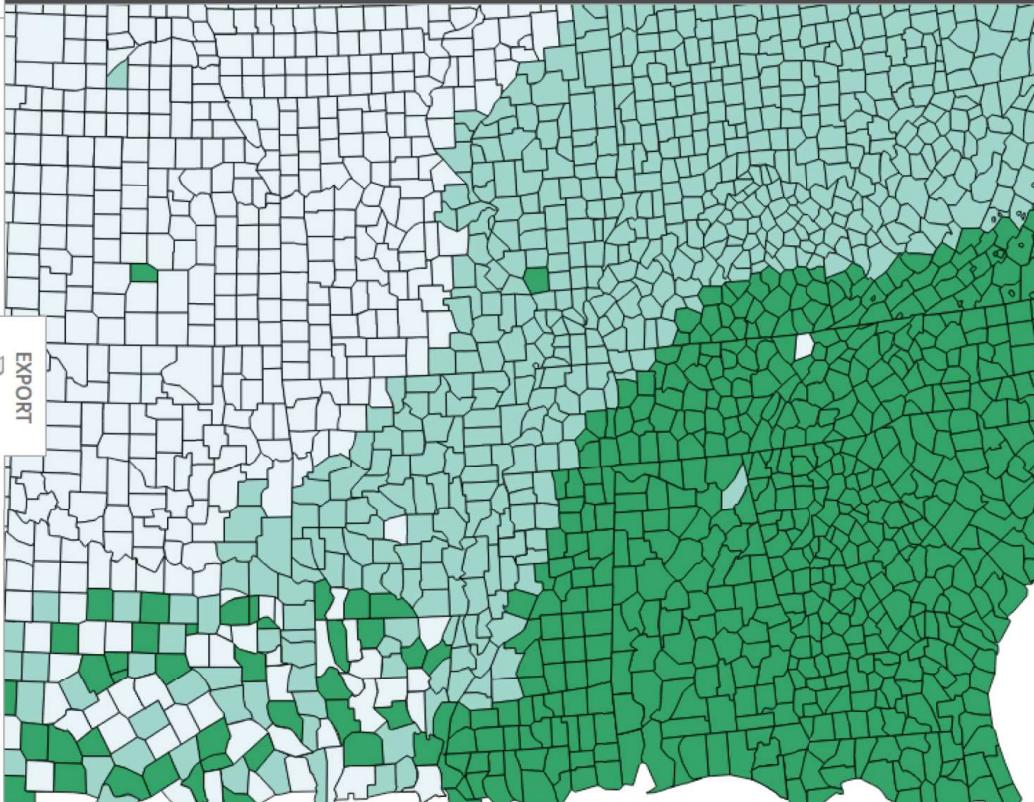
**COLORBREWER 2.0**  
color advice for cartography

3-class BuGn

EXPORT

HEX

#e5f5f9  
#99d8c9  
#2ca25f

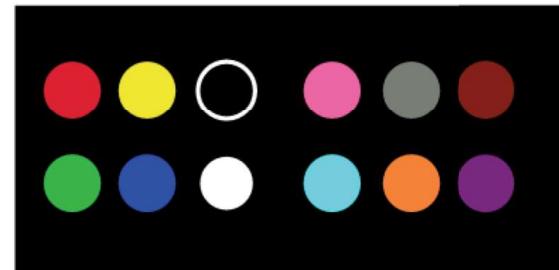
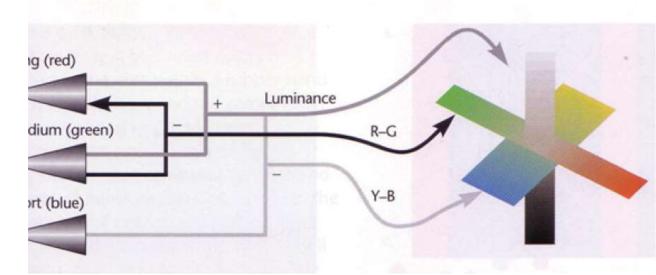


© Cynthia Brewer, Mark Harrower and The Pennsylvania State University  
 Source code and feedback  
[Back to Flash version](#)  
[Back to ColorBrewer 1.0](#)

 axismaps

# Colors /Categorical data

- Colors are fine with categorical data
- Do not reinvent the wheel (again)
- There are only 6 elementary colors arranged in three pairs
  - black-white
  - red-green
  - yellow-blue
- That gives us up to 12 (6+6) colors easily distinguishable (11!)



12 Colors  
for labeling

# Some new considerations

- 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 and at the task
2. Plan a graph (even on the paper!), considering perceptual issues
3. Look for an Excel implementation of your design
4. If 3 fails, proceed without Excel !

# Outline

- New ideas about good and bad graphs
- Meaning of numbers
- Tables and graphs

# Types of Data (rough classification)

- **Quantitative** (allows arithmetic operations)
  - 123, 29.56, ...
- **Categorical** (group, identify & organize; no arithmetic **BUT** they give a **structure** to the graphs!)
  - 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*
- Numbers are arranged on a chart using relationships among data

# Number <-> Category Relationships

Quantitative information	Relationship
<b>Unit of products</b> sold per geographical region	Sales count related to <b>geography</b>
<b>Expenses</b> by department and month	Expenses related to organizational <b>structure</b> and <b>time</b>
The <b>number of students</b> that got one of the possible exam score	Students counts related to <b>exam performance</b>

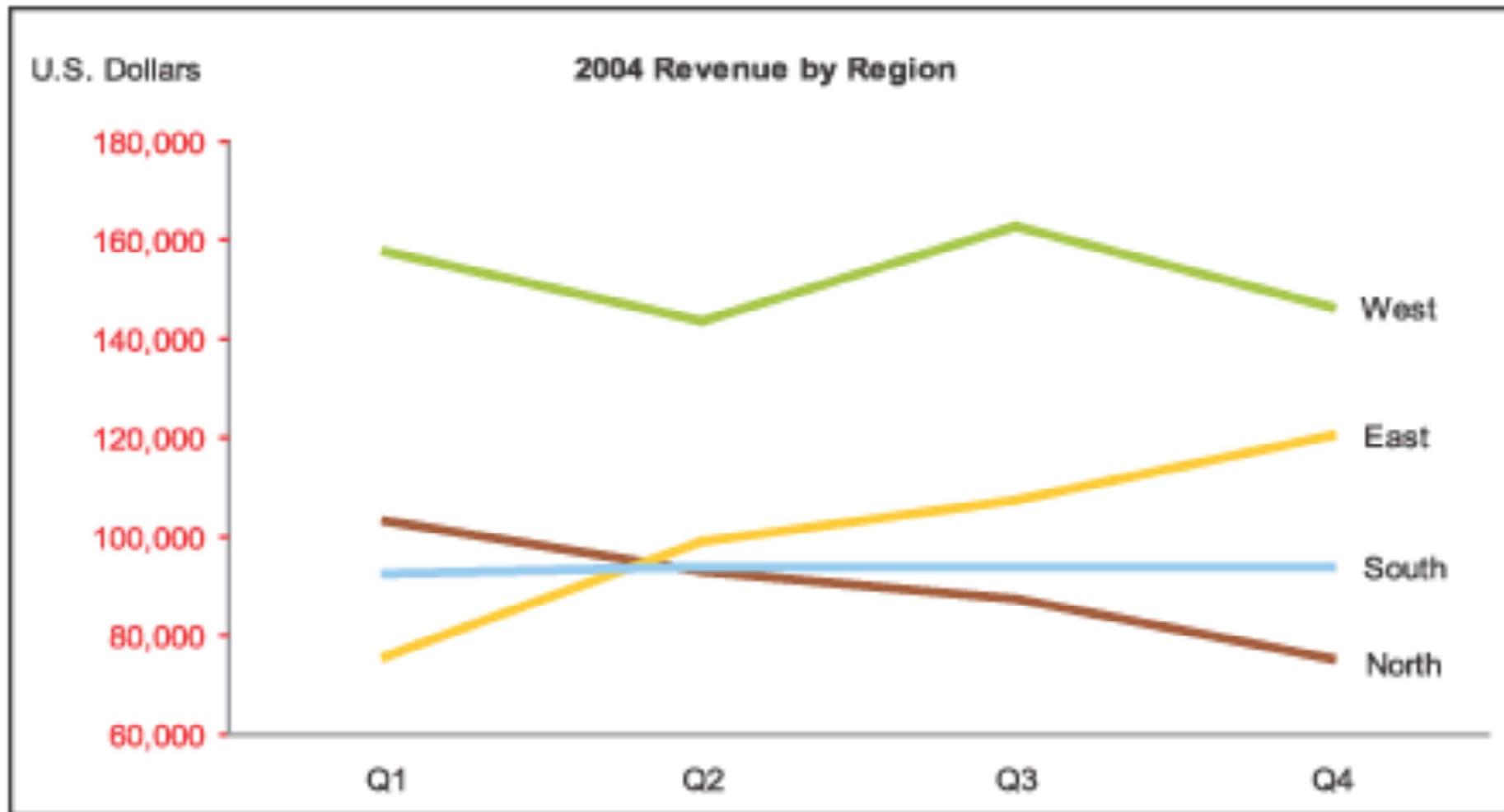
Numbers



Categories

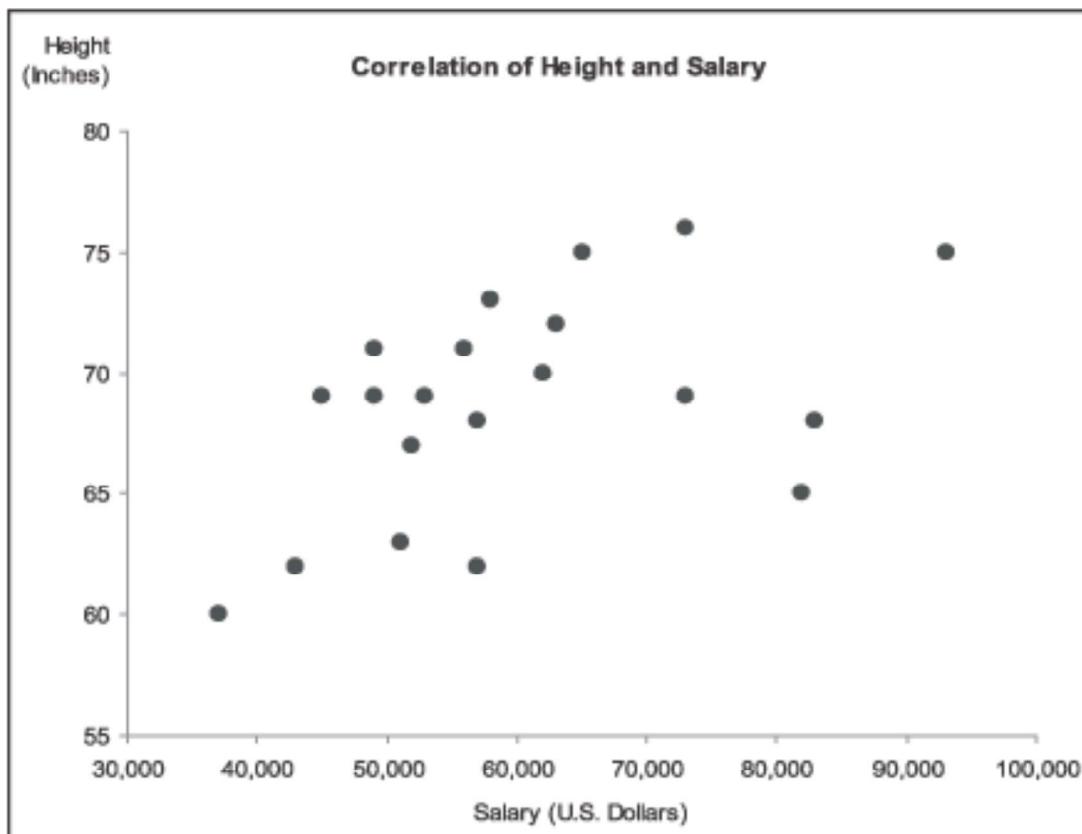


## Relationship between categories and numbers



Quantitative (y axis) vs categorical data (x axis and colors)  
An order exists among Q1..Q4...

# Relationship between numbers and numbers



- Quantitative vs quantitative data

# Nominal relationship

Region	Sales
North	50,000
South	20,000
East	40,000
West	20,000
Total	130,000

- Order is not relevant
  - Be aware of some artificial orders (conventional/ alphabetical order)
  - Maintain **consistence** across different graphs
- Just divide up the quantitative value

# Ordinal relationship (the order is in the category)

<b>Production office</b>	<b>Sales</b>
First office (1977)	50,000
Second office (2000)	20,000
Third office (2005)	40,000
Total	110,000

- Order is relevant
- Altering it is not a good idea

# Interval relationship

Order size	Count	Sum
[0, 1000)	25	2000785
[1000, 2000)	19	20086356256
[2000, 3000)	13	134555
[4000, 5000)	14	700005254

- Several equal intervals (bins) covering the whole range
  - Frequency distribution (count)
  - Other math's (sum of values, mean, etc.)
  - You can order any of the columns

# Time series relationship

Dept	Jan	Feb	Mar	Qtotal
Marketing	83,883	98,883	95,939	273,655
Sales	38,838	39,848	39,488	118,174

- Which kind of relationship best describes the categorical subdivision of time?
  - Obviously is ordinal
  - But months represent **intervals** as well

# Hierarchical relationship

Division	Dept	Group	Expenses
G&A	Human Resources	Recruiting	42,292
		Compensation	118,174
	Info Systems	Operations	512,885
		Applications	442,909
Finance	Accounting	AP	73,302
		AR	83,392
	Corp Finance	Fin Planning	93,027
		Fin Reporting	74,383

- Multiple categories, closely related to each other as separate levels in a ranked arrangement
- Commonly used in tables to arrange quantitative information (e.g., OLAP, On-Line Analytical Processing)
- <http://www.tableausoftware.com/products/desktop>



# Relationships among quantities

- Ranking
- Ratio / Proportion
- Correlation

# Ranking

Rank	Order Number	Order Amount
1	100303	1,939,393
2	100374	875,203
3	100482	99,303
4	100310	87,393
5	100398	67,939
		\$3,069,231

- It is an ordinal relationship in which the order is based on the associated quantitative values

# Proportion vs Ratio

- It is a relationship involving two quantitative values, compared by dividing one by the other
- If one is a part of the whole ( e.g.,  $a/a+b$  ) it is a **proportion** and it is typically represented as a percentage (ranging between 0 and 100)
- If the two values come from different sets it is a **ratio**, and it can assume any value, also above 100 and if the two values come from the same domain it makes sense consider the difference as well, that could be negative

# Proportion example

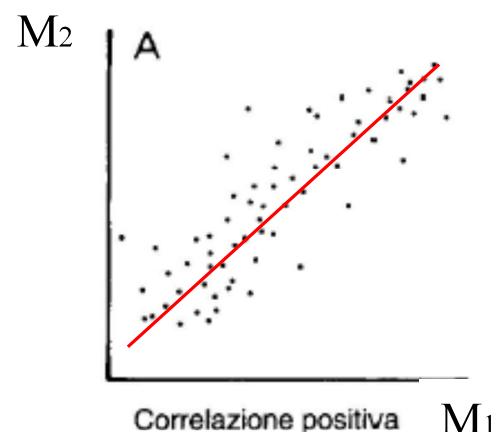
Company	Sales	Sales %
Company A	239,949,993	15%
Company B	873,777,473	54%
Company C	37,736,336	2%
Company D	63,874,773	4%
Company E	399,399,948	24%
Total	\$1,614,738,523	100%

# Ratio example

<b>Department</b>	<b>Jan</b>	<b>Feb</b>	<b>Feb/Jan</b>	<b>Variation</b>	<b>Change %</b>
Sales	9,933	9,293	0.93	-640	-6%
Marketing	5,385	5,832	1.08	+447	+8 %

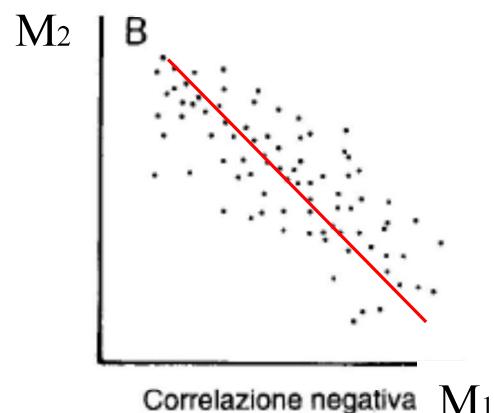
# Correlation relationship

Correlation is a relationship in which the values of two paired set of quantities are compared, looking for a (usually linear) function between them



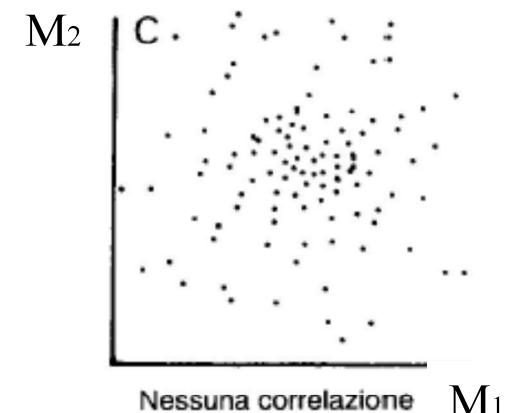
Directly proportional

+1



Inversely proportional

-1

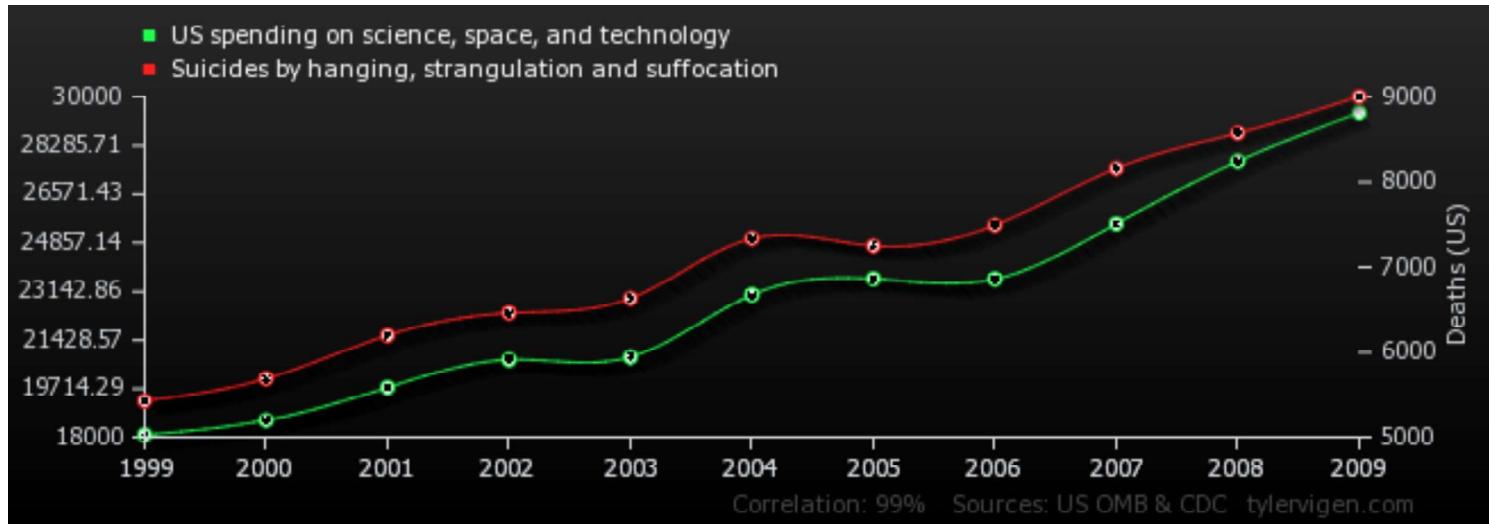


No correlation

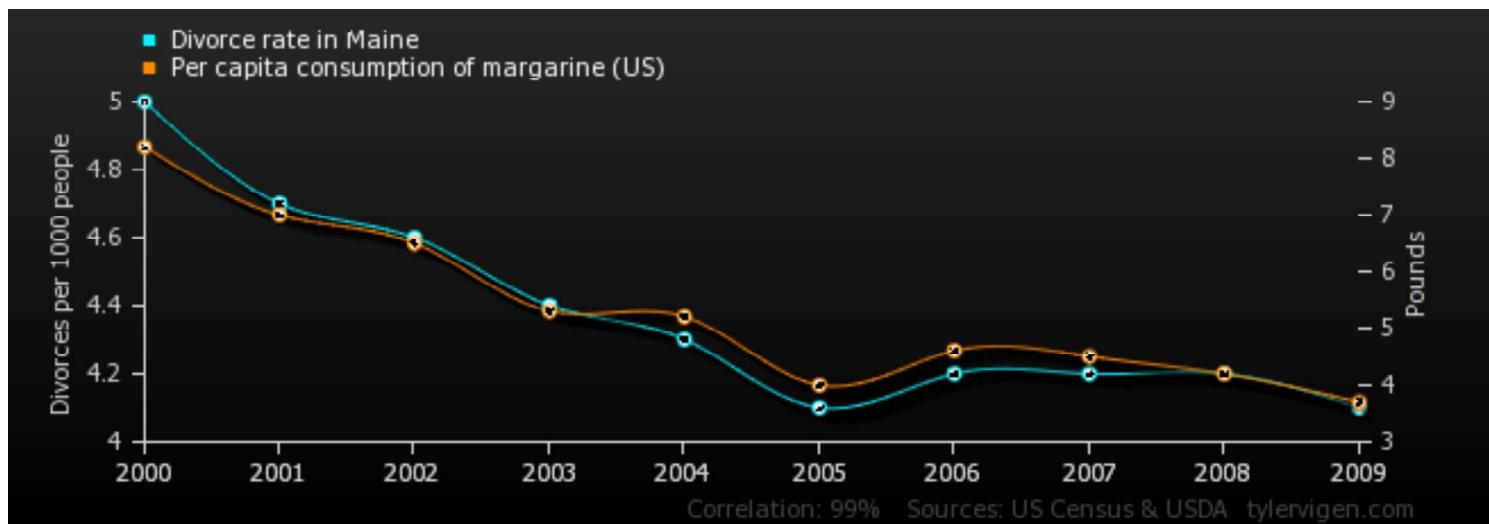
0

# Do not always trust statistics... spurious correlation site

<http://www.tylervigen.com/>



Correlation:  
0.992082



Correlation:  
0.992558

# Data & relationships summary

- Information consist (roughly) of two main types of data
  - Quantitative
  - Categorical
- Relationship among data could be
  - Simple associations between **quantitative** and **categorical** subdivision
  - More complex association among multiple set of values/intervals
- Four types of relationships within **categories**
  - Nominal
  - Ordinal
  - Interval
  - Hierarchical
- Three types of relationships between **quantitative** values
  - Ranking
  - Ratio / Proportion
  - Correlation

# Numbers that summarize (do not lie)

- Measures of average
  - Mean
  - Median
  - Mode
  - Midrange
- Measures of distribution
  - Range
  - Variance
  - Standard deviation

# Mean

- Sometimes it is not informative

Quarter	Units Sold
Q1	339
Q2	373
Q3	437
Q4	563
Sum	1,712
Count	4
Mean (per Qtr)	428

Employee	Position	Annual Salary
Employee A	Vice President	475,000
Employee B	Manager	165,000
Employee C	Manager	165,000
Employee D	Admin Assistant	43,000
Employee E	Admin Assistant	39,000
Employee F	Analyst	65,000
Employee G	Analyst	63,000
Employee H	Writer	54,000
Employee I	Writer	52,000
Employee J	Graphic Artist	64,000
Employee K	Graphic Artist	62,000
Employee L	Intern	28,000
Employee M	Intern	25,000
Mean Salary		\$100,000

# Median

- It splits the sorted distribution in two

Rank	Position	Annual Salary
1	Vice President	475,000
2	Manager	165,000
3	Manager	165,000
4	Analyst	65,000
5	Graphic Artist	64,000
6	Analyst	63,000
7	Graphic Artist	62,000
8	Writer	54,000
9	Writer	52,000
10	Admin Assistant	43,000
11	Admin Assistant	39,000
12	Intern	28,000
13	Intern	25,000
Median Salary		\$62,000

# Distribution

Warehouse	Sum of shipping days	Delivery mean	Delivery median
A	51	4.25	4.5
B	51	4.25	4.5

- Performances of delivery time of 12 orders of two warehouses
- Do they perform the same?
- What is missing?

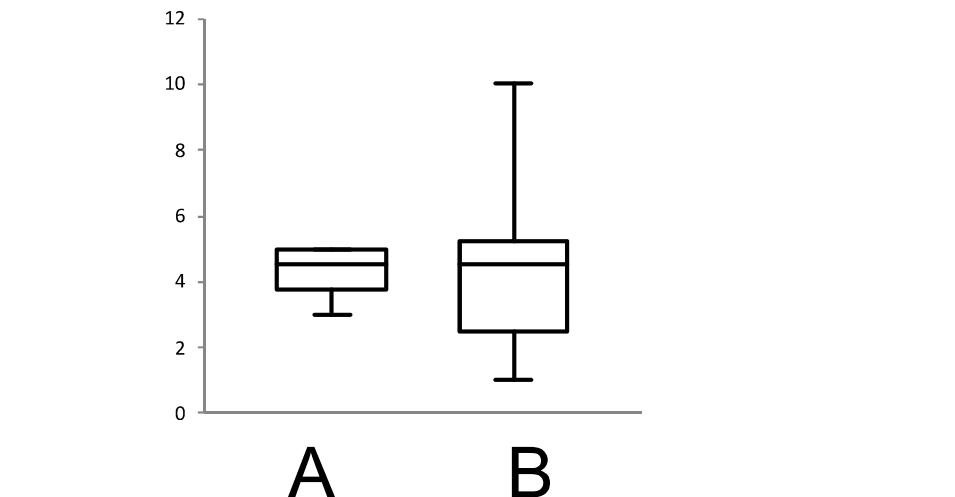
# Distribution

<b>Order #</b>	<b>Warehouse A</b>	<b>Warehouse B</b>
1	3	1
2	3	1
3	3	1
4	4	3
5	4	3
6	4	4
7	5	5
8	5	5
9	5	5
10	5	6
11	5	7
12	5	10

# Range and midrange

Order #	Warehouse A	Warehouse B
1	3	1
2	3	1
3	3	1
4	4	3
5	4	3
6	4	4
7	5	5
8	5	5
9	5	5
10	5	6
11	5	7
12	5	10

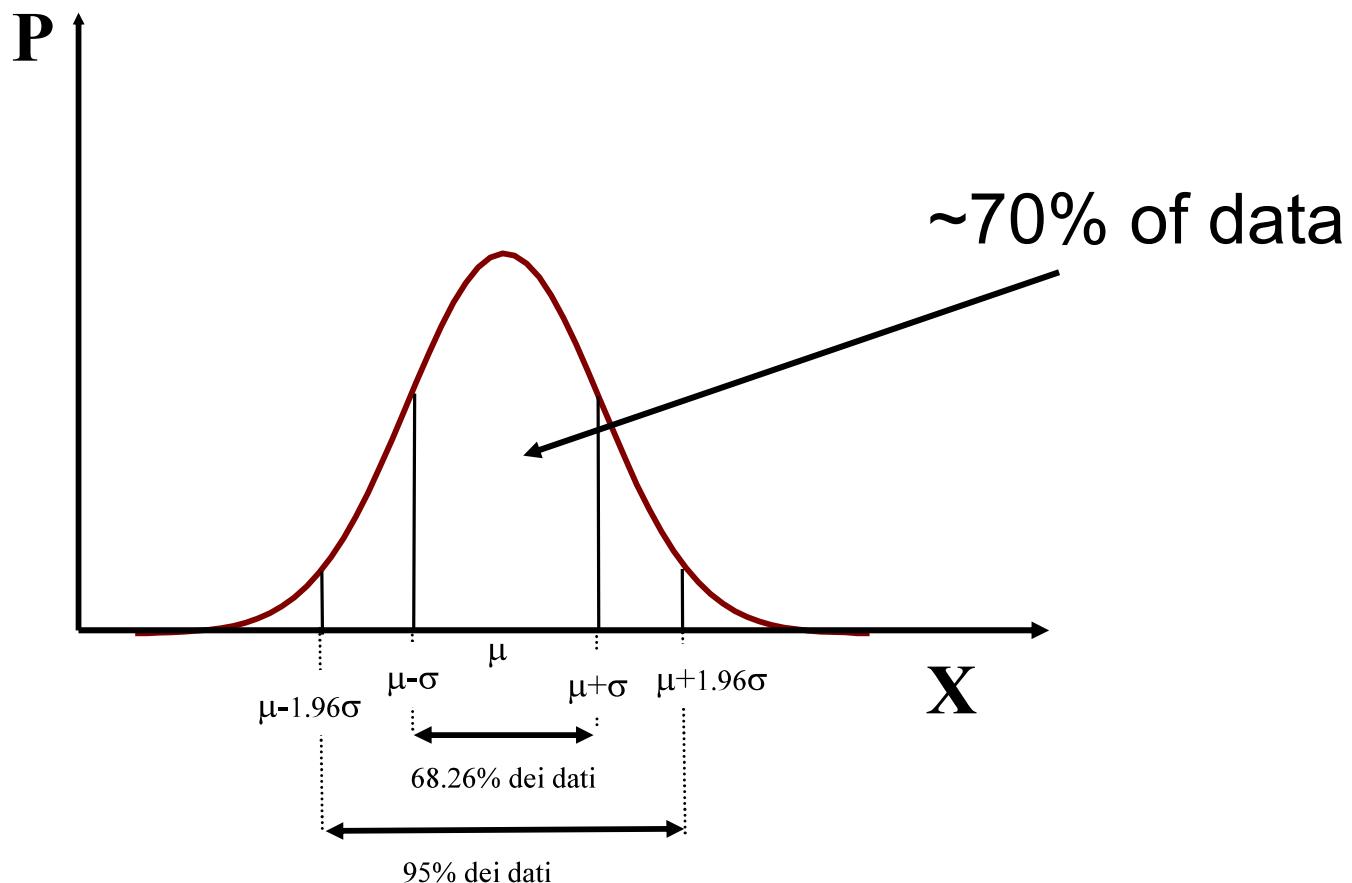
- Range is just max-min
  - Range A =  $5-3=2$
  - Range B =  $10-1=9$
- 
- Midrange is just  $(\text{max}+\text{min})/2$
  - Midrange A= $(5+3)/2=4$
  - Midrange B= $(10+1)/2=5.5$



Boxplots are more informative

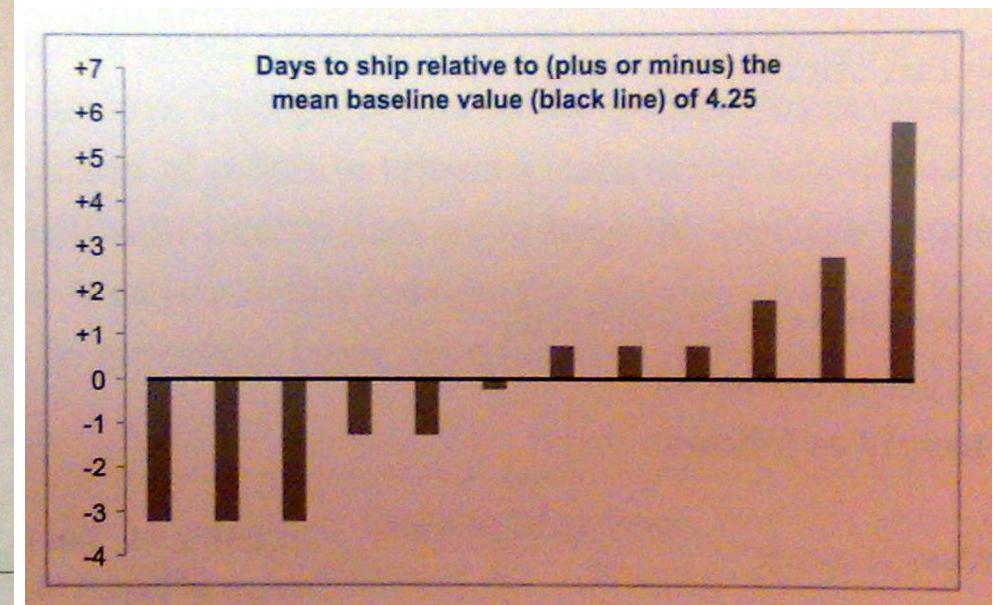
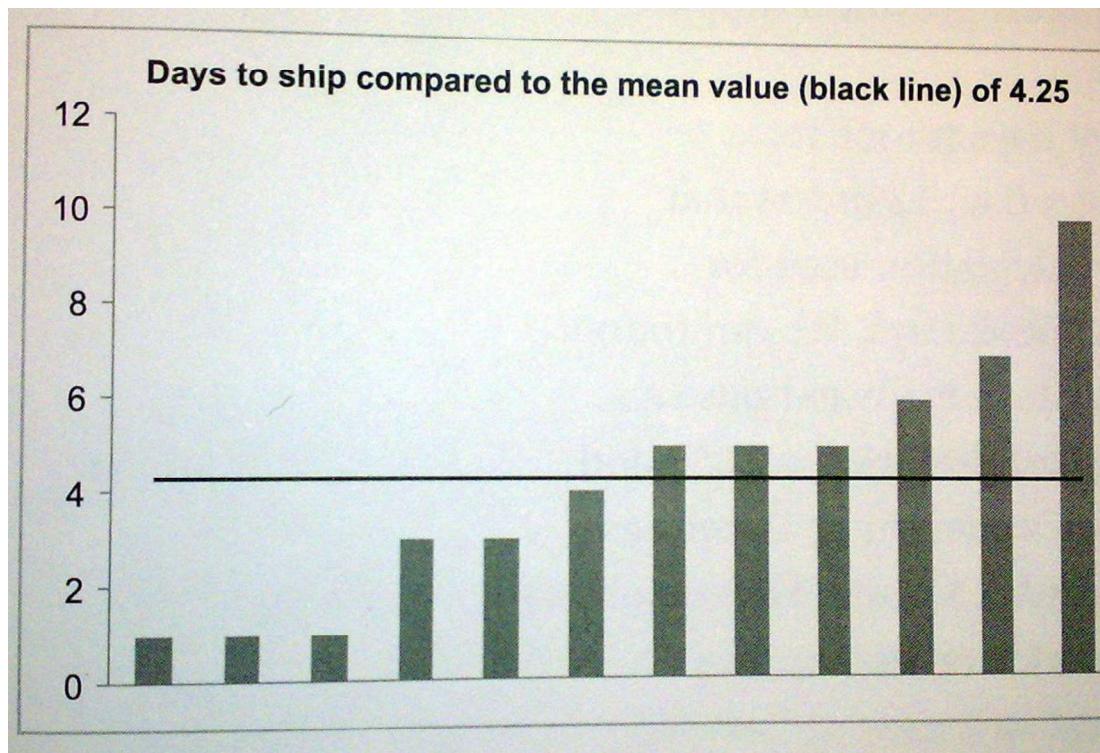
# Standard deviation

- This variability is well described by variance and standard deviation
- mean:  $\mu = (x_1 + x_2 + \dots + x_n)/n$
- variance  $\text{var} = [(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2]/n$
- standard deviation  $\sigma = \text{var}^{1/2}$
- However such concepts are hard to communicate



# Standard deviation

- These bar charts compare values with mean, providing a simpler way of communicating standard deviation



# Measures of ratio

- Simple numerical relationship between two values
- It can be used to summarize data as well



# Number that summarize

Type of Summary	Method	Note
Average	Mean	Measures the center of a set of values in a manner that is equally sensitive to all values, including extremes
	Median	Measures the center of a set of values in a manner that is insensitive to extreme values
Distribution	Range	Simple to calculate, relying entirely on the highest and lowest values, but only roughly defines a range of values
	Standard Deviation $\sigma$	Provides a rich expression of the distribution of a set of values across its entire range
Correlation	Linear Correlation Coefficient	Indicates whether a correlation exists between two paired sets of values, and if so, its direction (positive or negative) and its strength (strong or weak)
Ratio	Rate or Percentage	Measures the direct relationship between two quantitative values

# Outline

- New ideas about good and bad graphs
- Meaning of numbers
- Tables and graphs

# Table and graphs

- Table and graphs are widely used to communicate quantitative information
- Sometimes it is better to just show the (few) numbers
- 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)

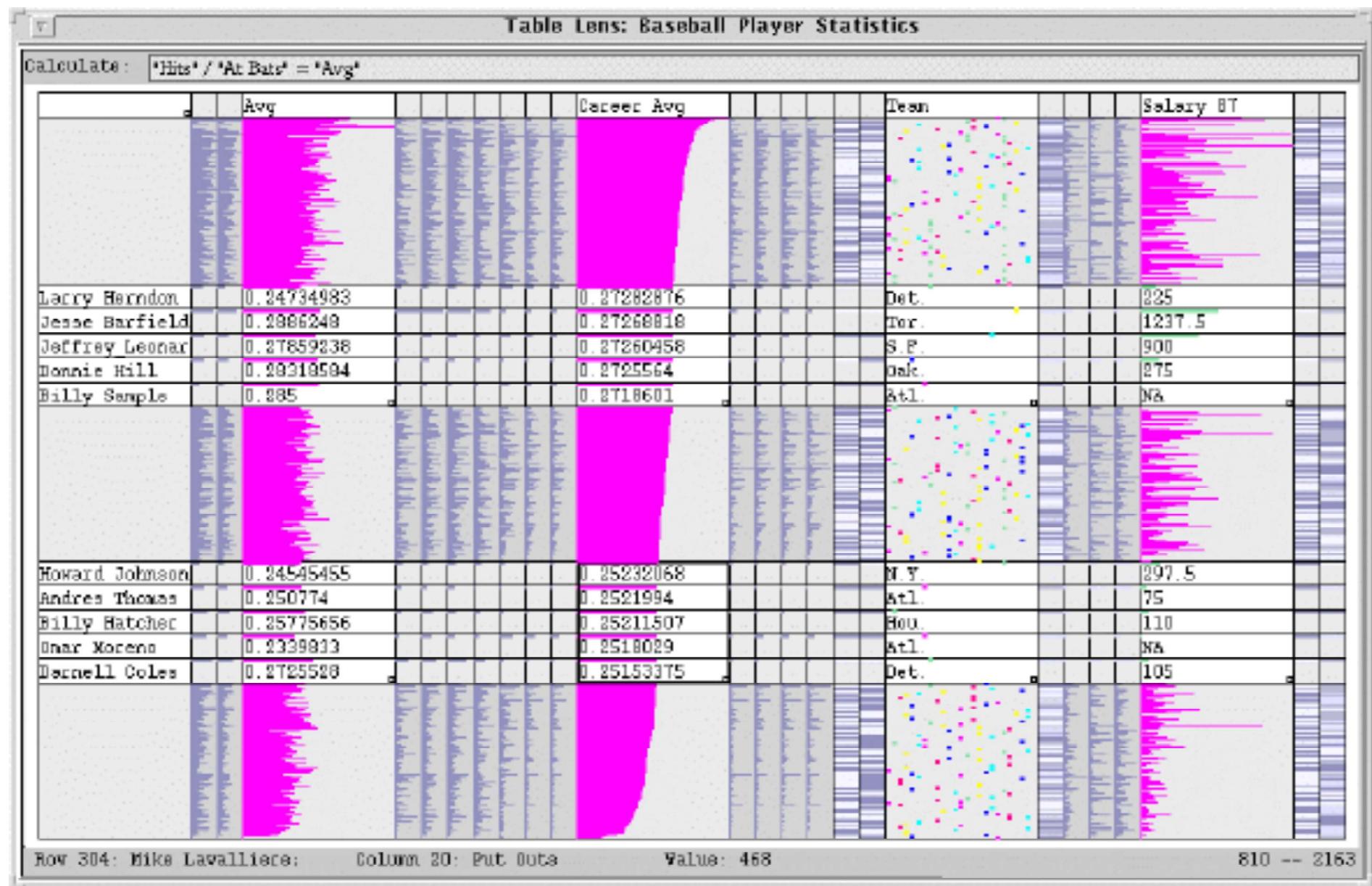
# Table

- Data is arranged in columns and row
  - Data is encoded as text (usually)
  - They are used also for non quantitative information (just spatial arrangement)
1. Tables make easy **look up** values
  2. Tables allow for displaying simple relationships between **quantitative** and **categorical** subdivision
  3. Table allow for **local** comparisons
  4. Tables provide for **high precision**
  5. Table allow for easy management of **different units** of measure

# Choose a table when...

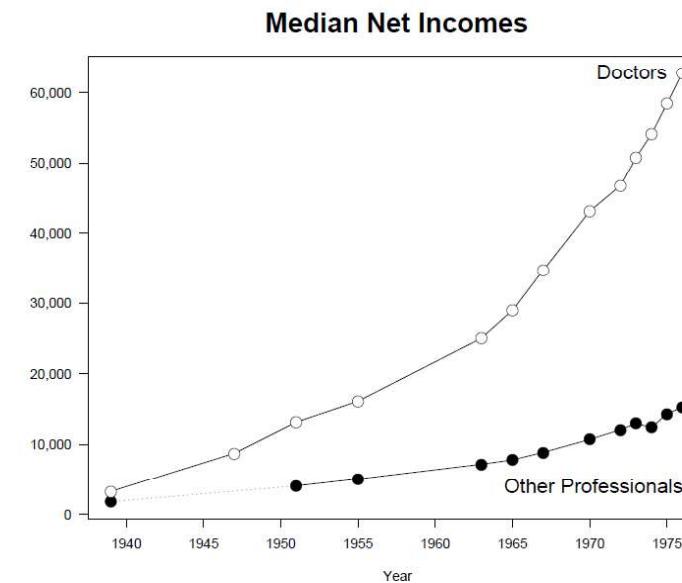
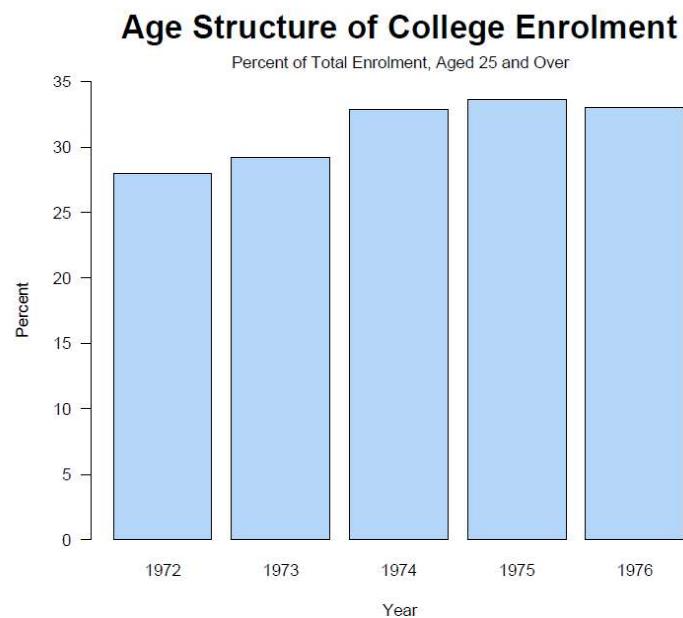
- If one of the following is true, a table could be a good choice
1. The report you produce will be used to look up single values
  2. It will be used to compare individual values
  3. Precise values are required
  4. Different units of measure are involved

# Tablelens: a table with non-numerical values



# Graphs

- A graph is a visual display of quantitative information
- **Quantitative information is encoded visually**
- More precisely, values are represented and presented on one or more axes
- Axes provide scales (quantitative or categorical)



# Graphs

- A graph provides the overall shape of the data
  - Trends
  - Outliers
  - Similarities and differences
- 
- Low precision
  - Not easy look up
  - Not easy local comparison
  - Not easy handling of different units

# Fundamental issues in table design

- Relationships in table
  - Quantitative to categorical
- Variation in table design
  - Unidirectional
  - Bidirectional
  - Table design solutions

# 1:1 - One categorical subdivision and one quantitative values (sales)

nominal

Salesperson	QTD Sales
Robert Jones	13,803
Mandy Rodriguez	20,374
Terri Moore	28,520
John Donelly	34,786
Jennifer Taylor	36,975
Total USD	\$134,458

, . ??? : different rules

# 1:n Among 1 categorical subdivision (a salesperson) and n quantitative values (sales, returns, net)

Salesperson	Sales	Returns	Net sales
Robert Jones	13,803	593	13,210
Mandy Rodriguez	20,374	1,203	19,171
Terri Moore	28,520	10,393	18,127
John Donelly	34,786	483	34,303
Jennifer Taylor	36,973	-	36,973
Total USD	\$134,456	\$12,672	\$121,784

- Here the focus is the comparison among **NOT** homogeneous numerical values

# n:1 n categories (n=2 salespersons & months) to 1 quantitative value (sales)

Salesperson	January	February	March
Robert Jones	2,834	4,838	6,131
Mandy Rodriguez	5,890	6,482	8,002
Terri Moore	7,398	9,374	11,748
John Donelly	9,375	12,387	13,024
Jennifer Taylor	10,395	12,383	14,197
Total USD	\$35,892	\$45,464	\$53,102

nominal + interval (time)

# hn:1 - hn hierarchical categories

(Product Line -> Family -> Product) and 1 quantitative value (sales)

Product line	Product family	Product	Sales
Hardware	Printer	PPS	6,131
		PXT	8,002
		PQT	11,748
		RRZ	13,024
		RTS	14,197
		RQZ	23,293
Software	Business	ACT	12,393
		SPR	9,393
		DBM	5,392
	Game	ZAP	10,363
		ZAM	15,709
		ZOW	13,881
Total			\$143,526

# Unidirectional - Bidirectional

Dept	Expense type	Expenses
Finance	Compensation	160,383
	Supplies	5,038
	Travel	10,385
Sales	Compensation	683,879
	Supplies	193,378
	Travel	125,705

Unidirectional

	Departments	
Expense Types	Finance	Sales
Compensation	160,383	683,879
Supplies	5,038	193,378
Travel	10,385	125,705

Bidirectional (pivot table)

They save space

# Graphs

- Relationships in graphs
  - Quantitative to categorical
  - Quantitative to quantitative !

# Graphs' variation

- The primary source of variation is the choice (or combination) of the different **visual encoding** used to represent **quantitative values** and **categorical subdivision**. The most common choices are:

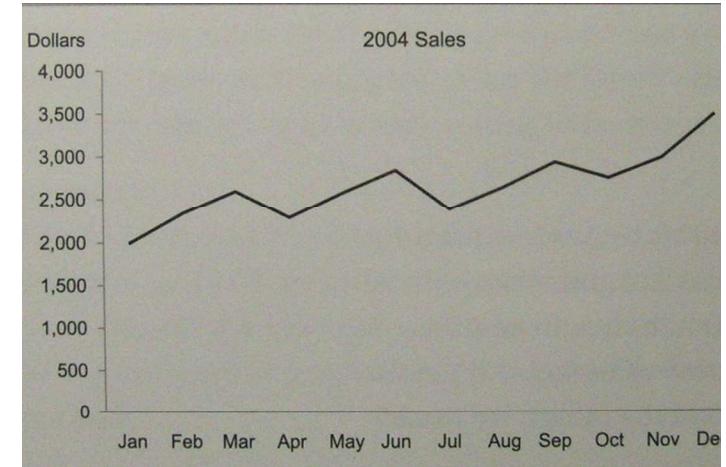
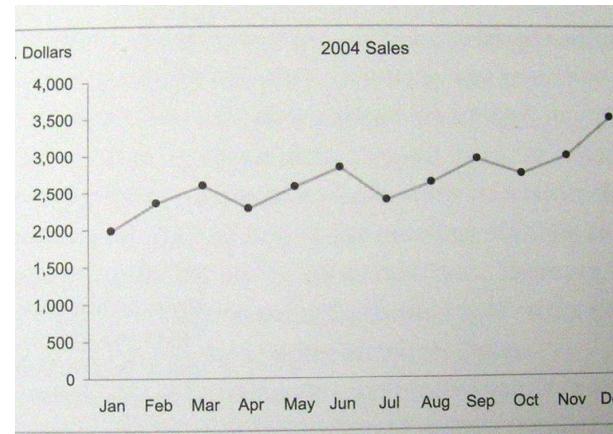
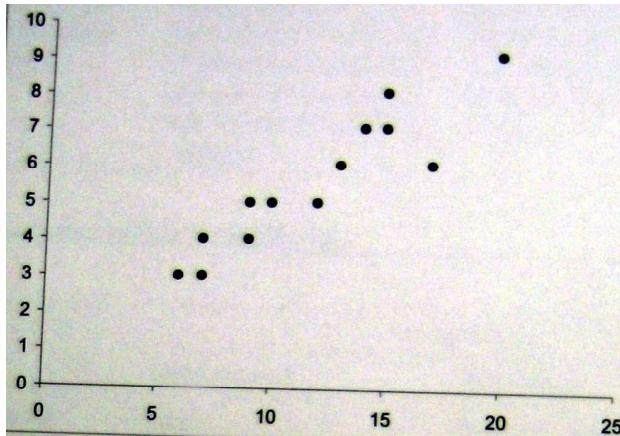
- Shape
- Position
- Length
- Area
- Color
- Pattern

Do not use them randomly!

# Encoding quantitative values

- Position
  - Length
  - Areas

# Position using points and/or lines



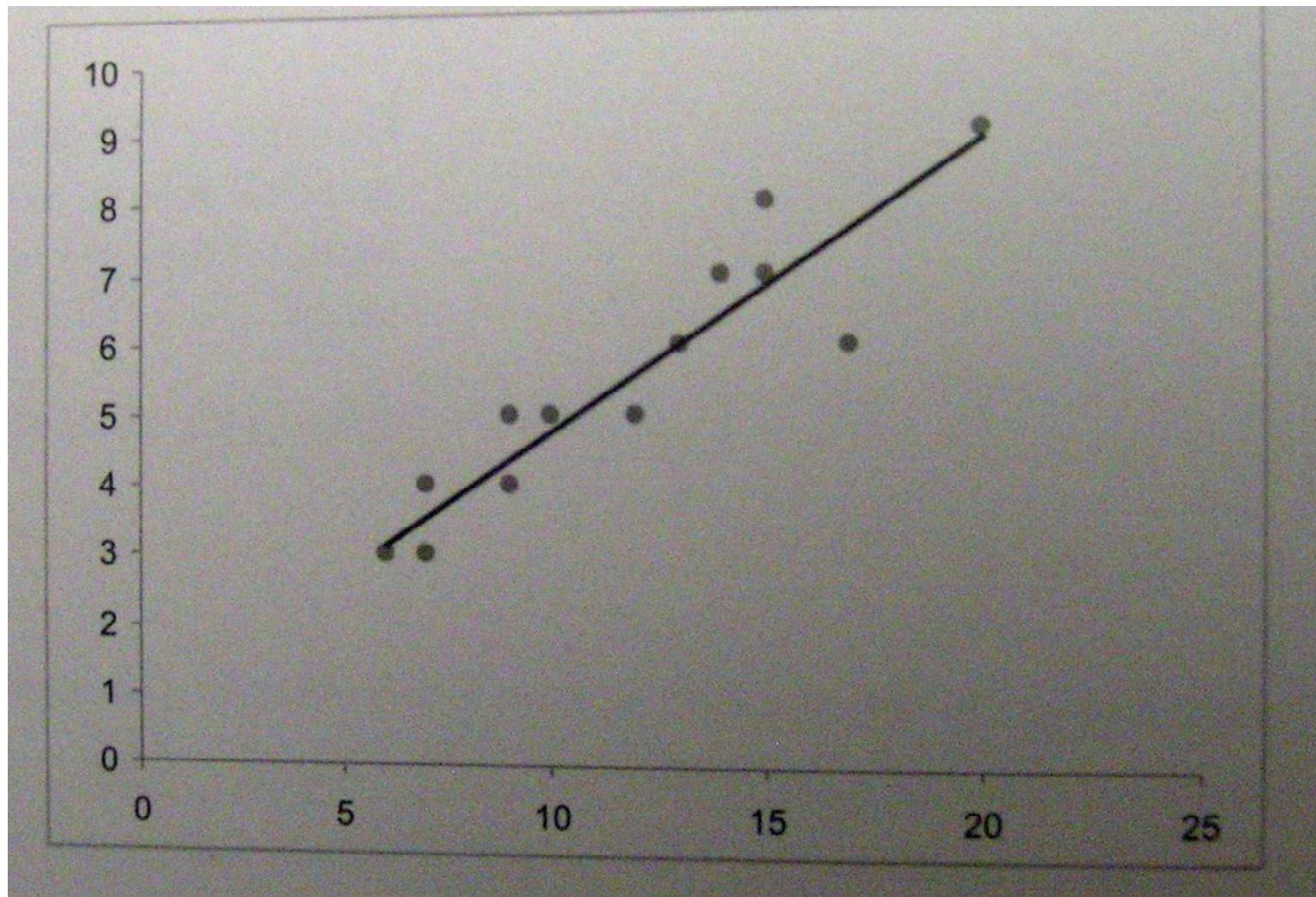
quantitative-quantitative:  
use points

Categorical – Numerical

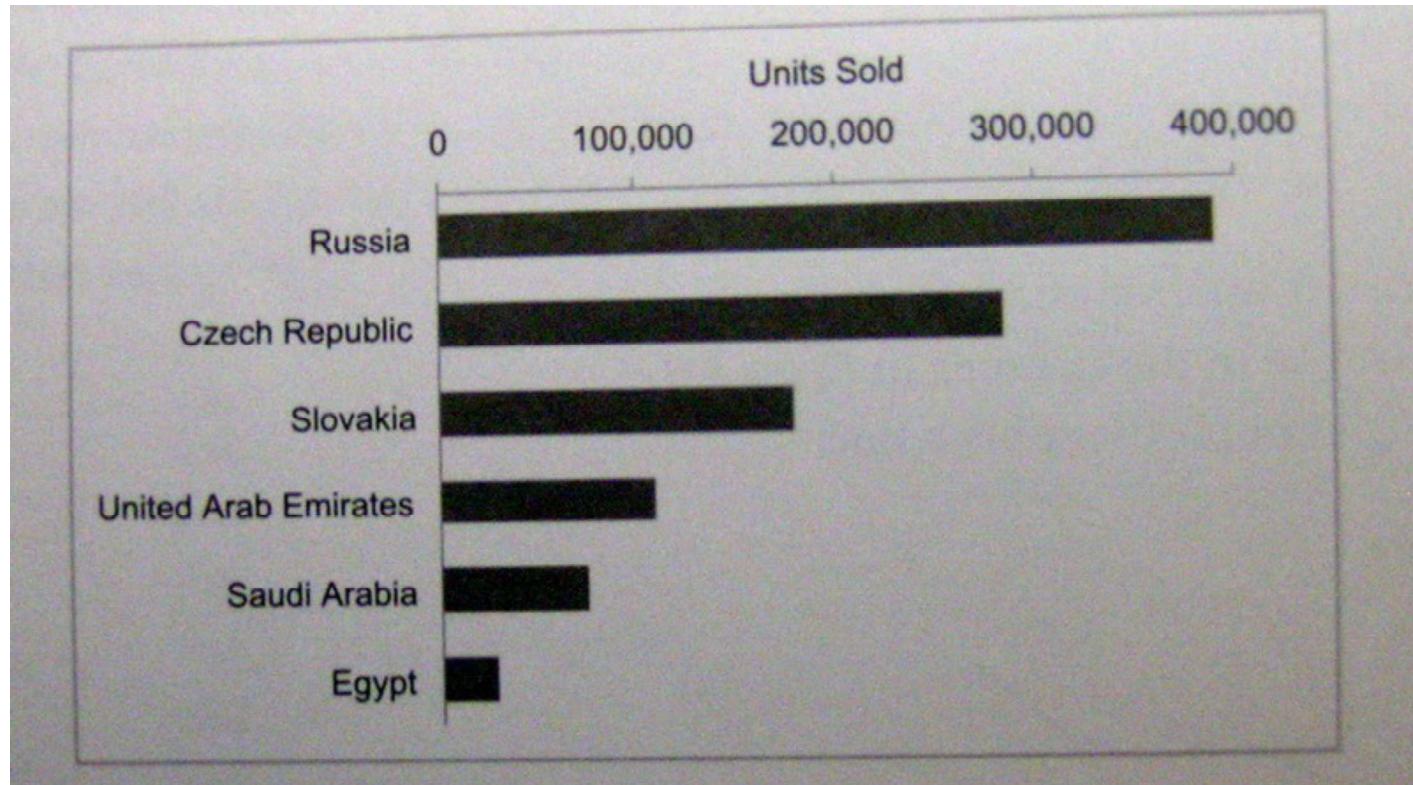
- Use lines **only** when
  - both axes are numerical, or
  - there exists an order on the categorical axis (e.g., intervals)

Do not leave the decision to Excel ...

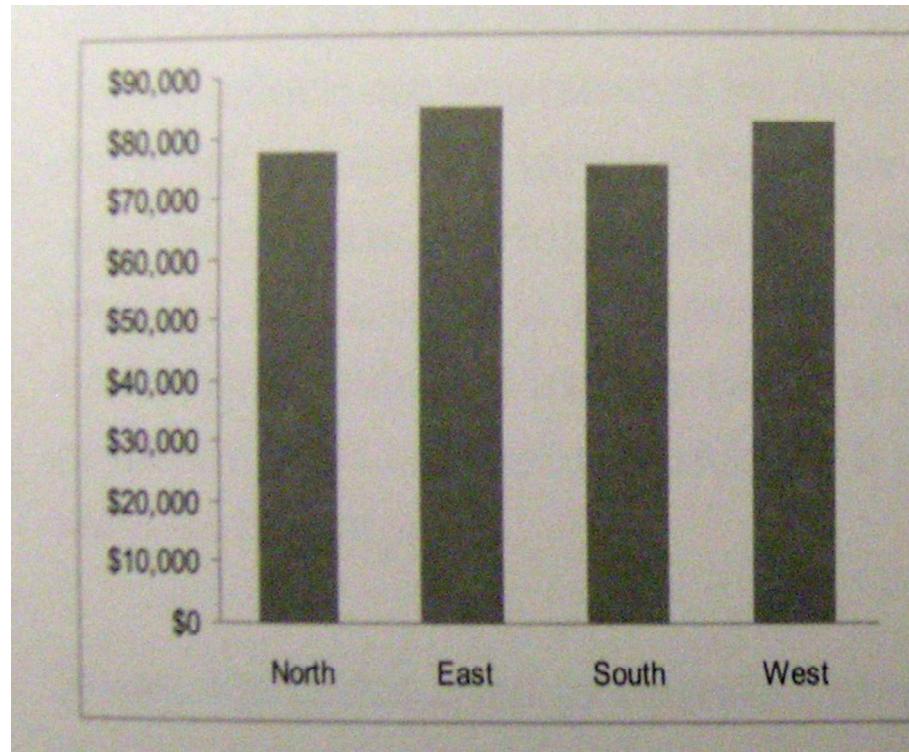
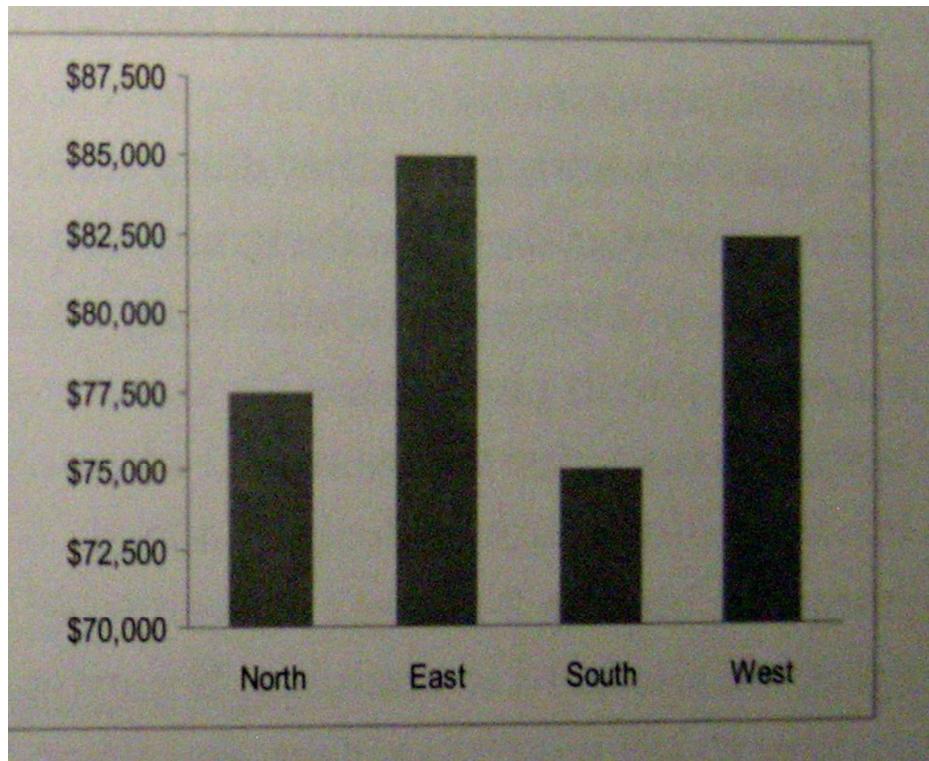
# Position using line (correlation)



# Length using bars



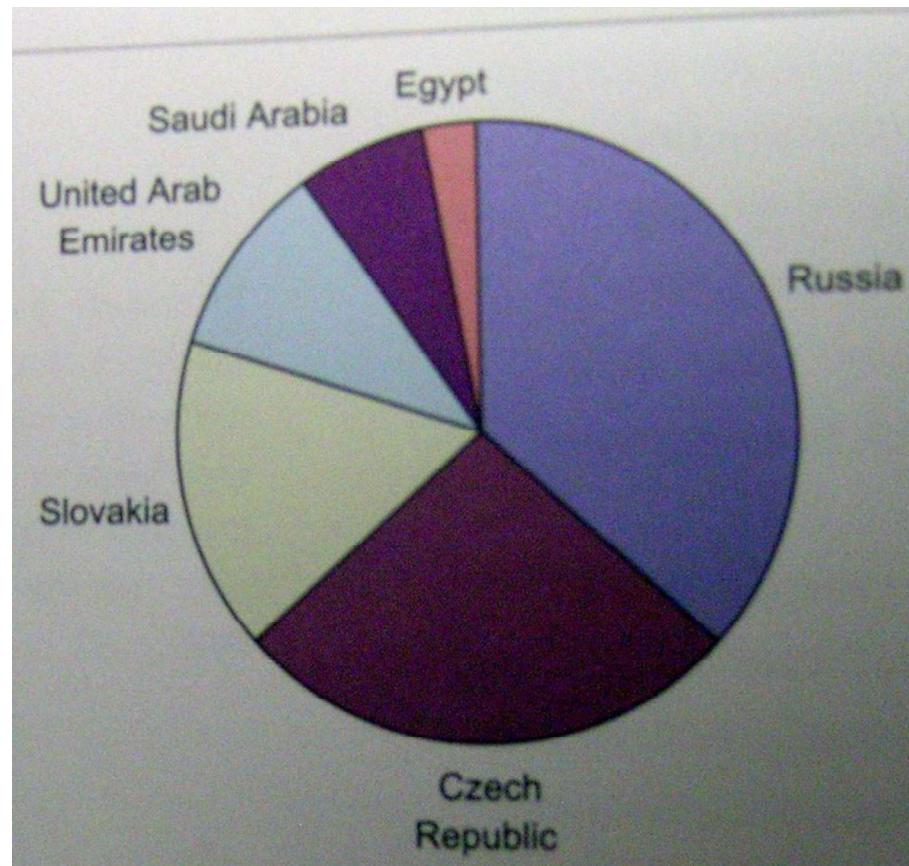
- Thickness is not relevant
- Thickness must be constant (and little)



- Do not lie! (Or allow Excel to lie for you!!!)
- Start scale by zero, allowing full lengths to be compared!

# Areas

- Classical pie chart
- Part of a larger family of area graphs
- Remember its limitations
- Where is the scale ?
- Our visual perception is not good to accurately assess and compare quantitative values using areas (or worst, slices)

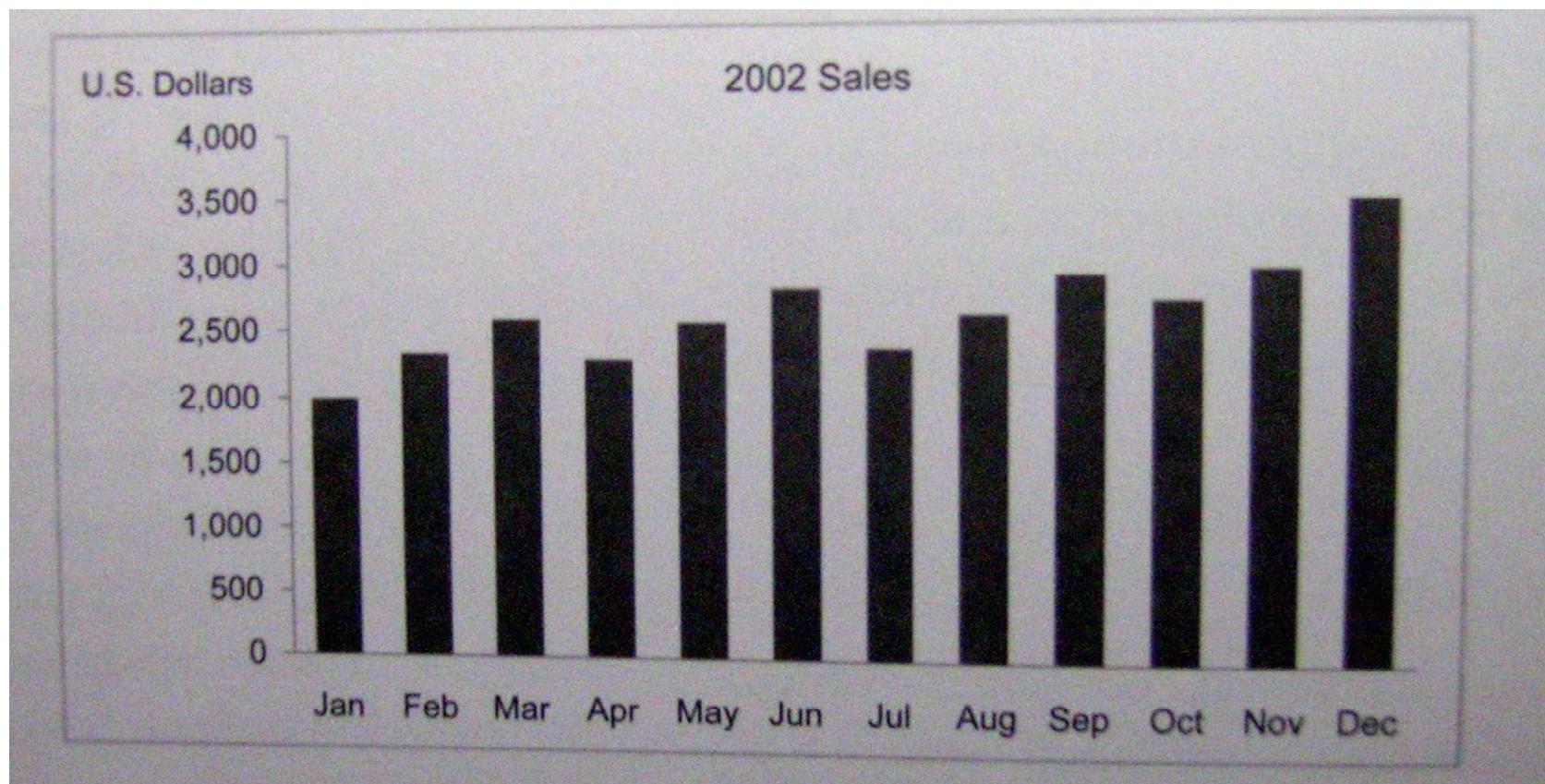


So, simply, do not use them at all !!  
Unless you do that for psychological reasons  
Add figures on it!

# Encoding categorical values

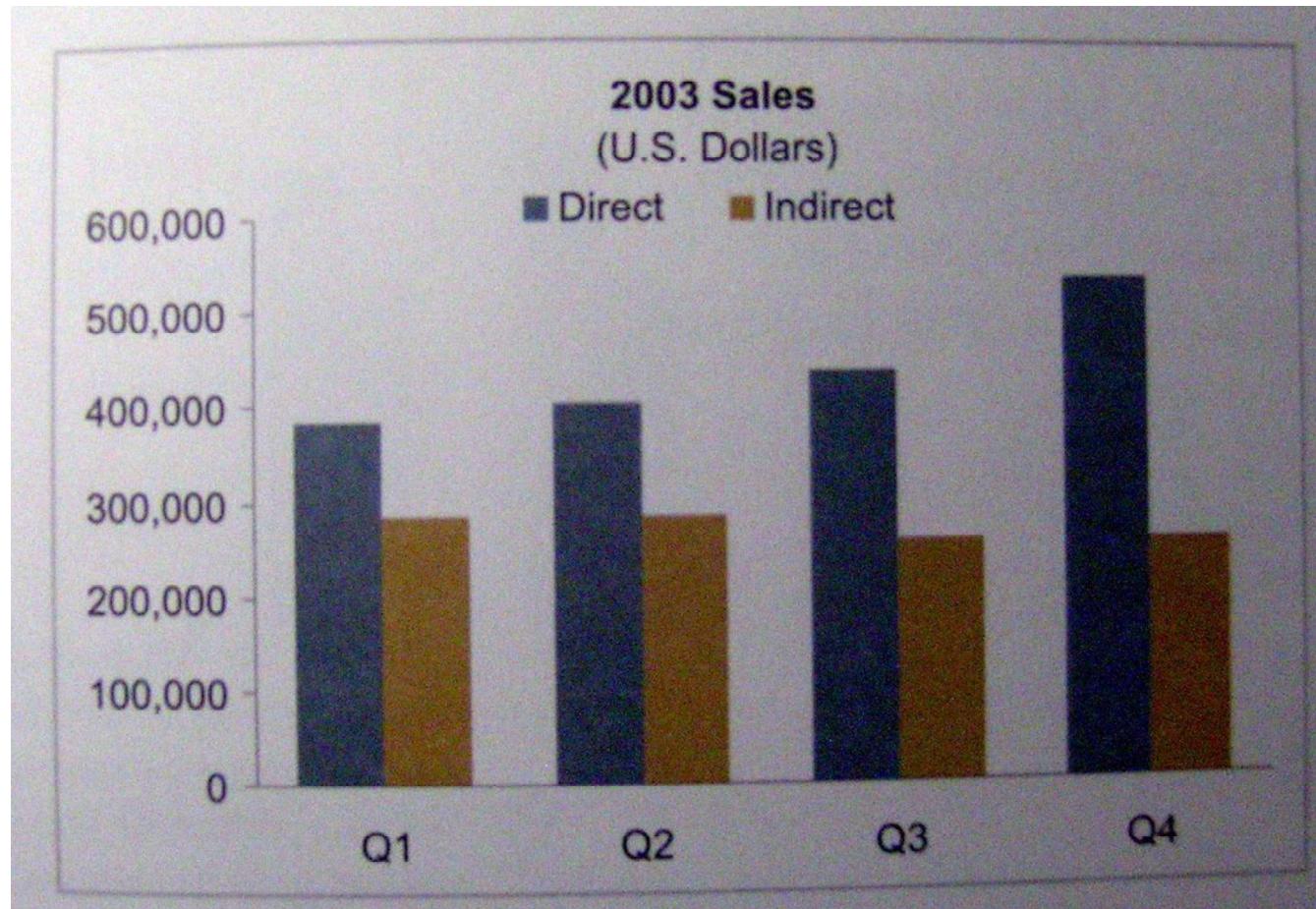
- Position
- Color
- Point shape
- Fill pattern
- Line style

# Position

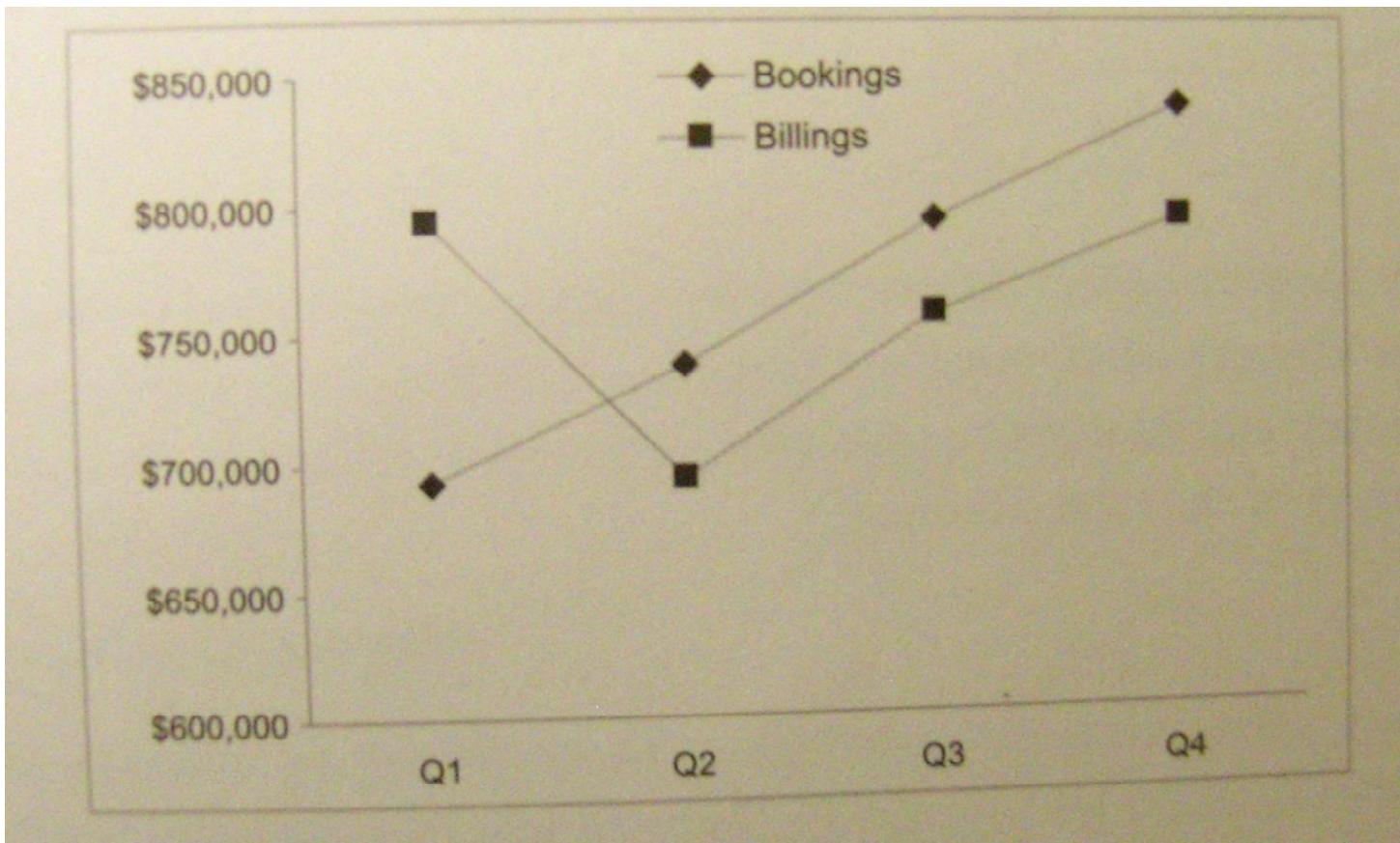


- X axis

# Color

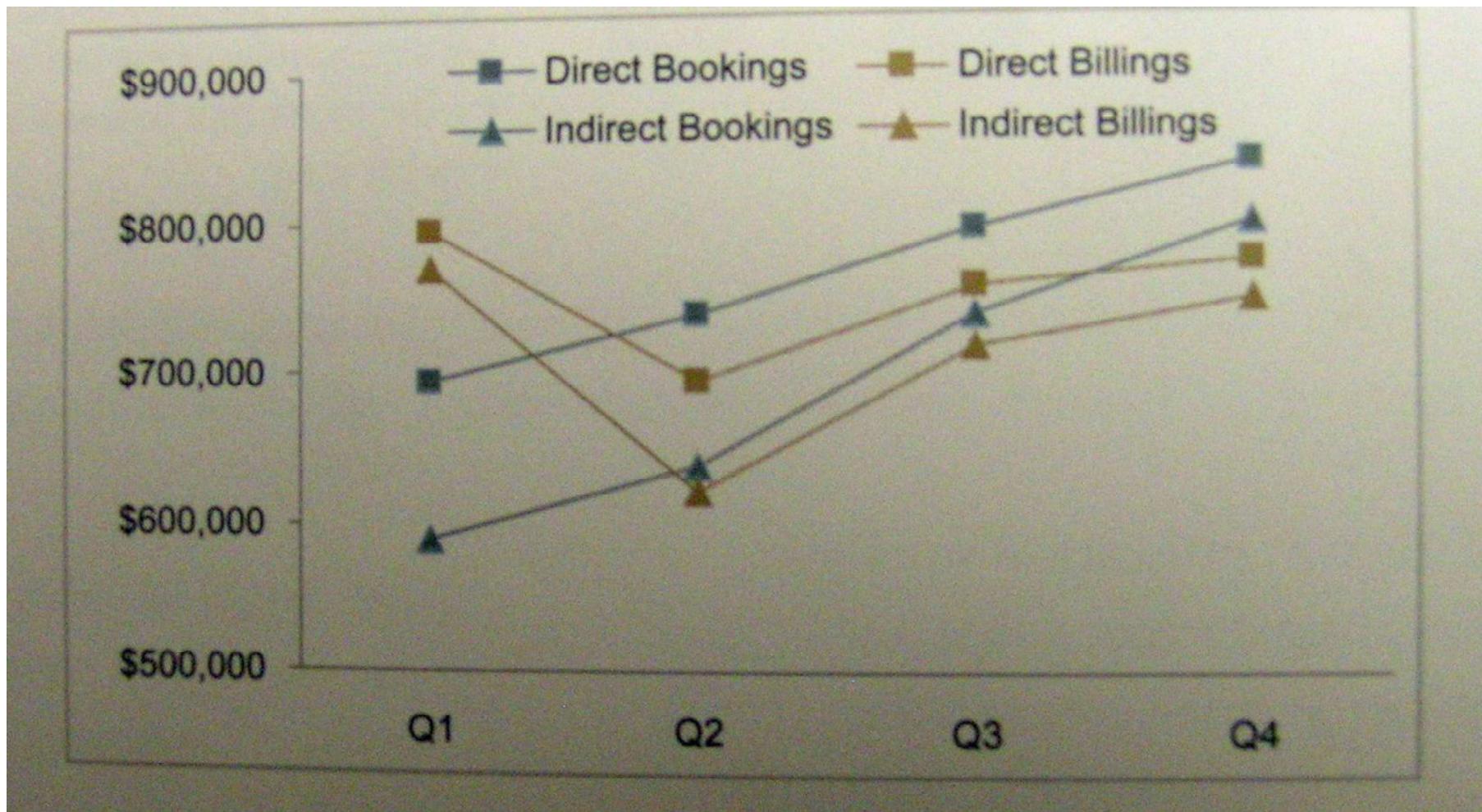


# Point shape

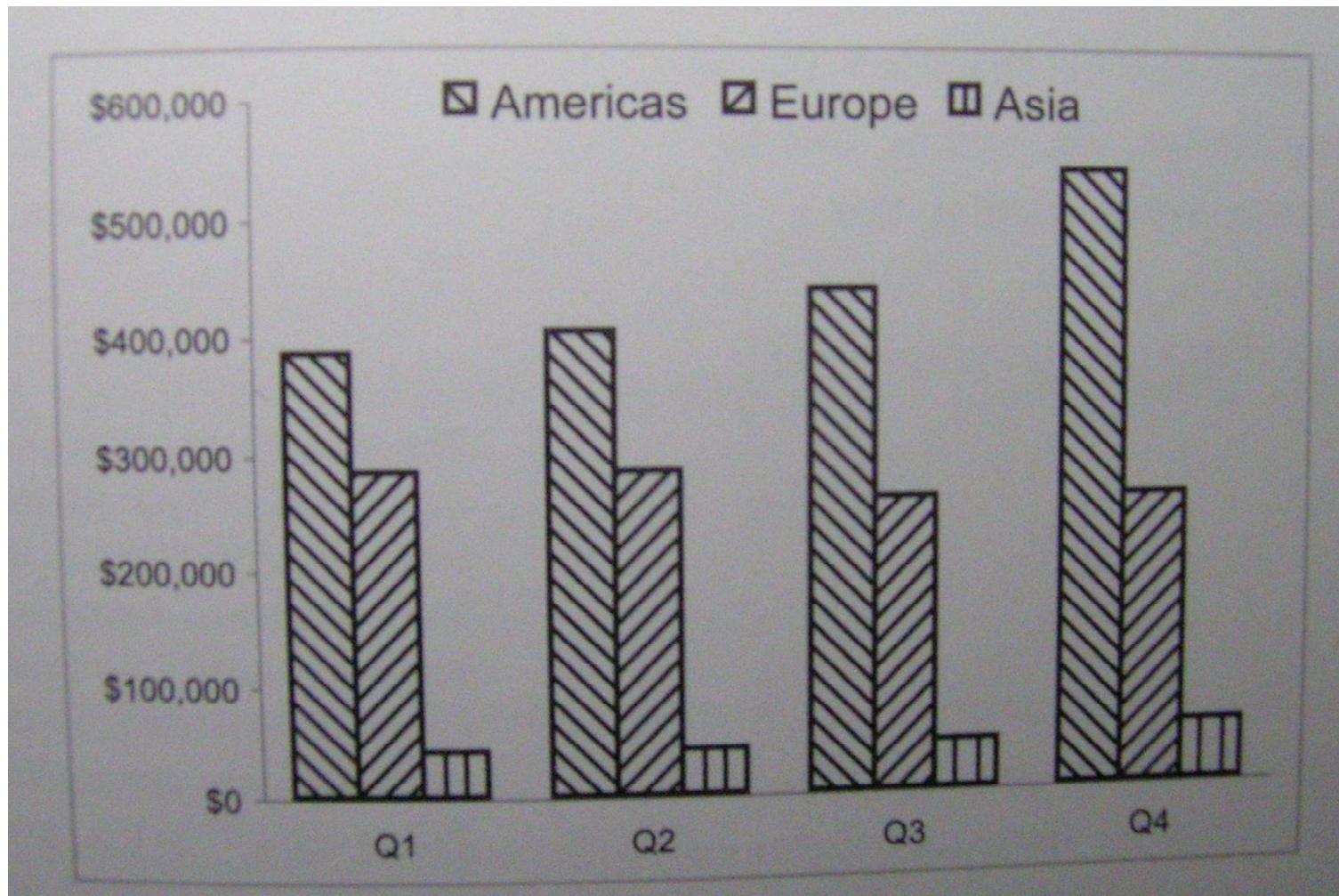


- Only applicable when points represents also quantitative values

# Position, Color, Point shape

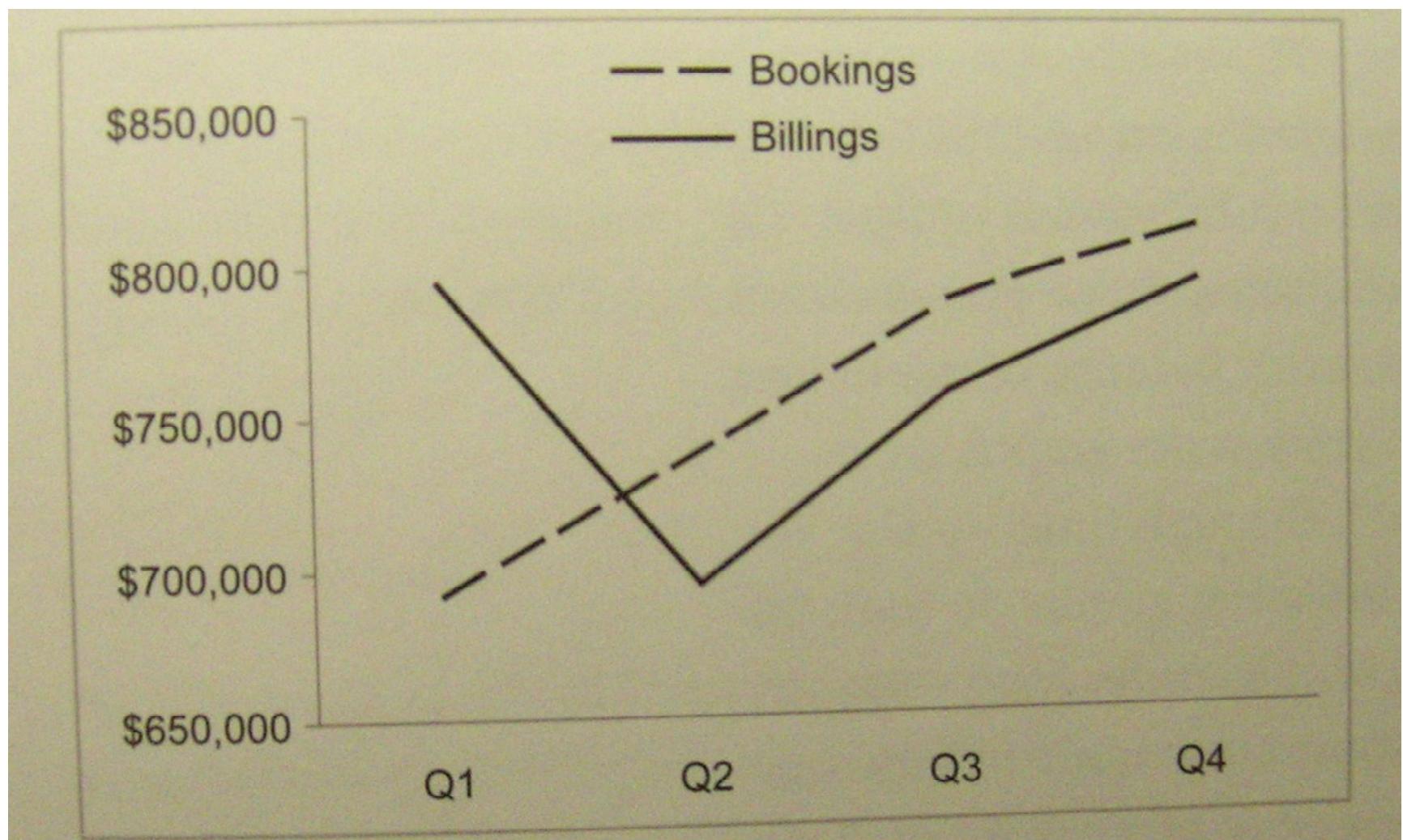


# Fill pattern



mmm, hard to see and causing **moirè vibration**  
use as the last resource

# Line style

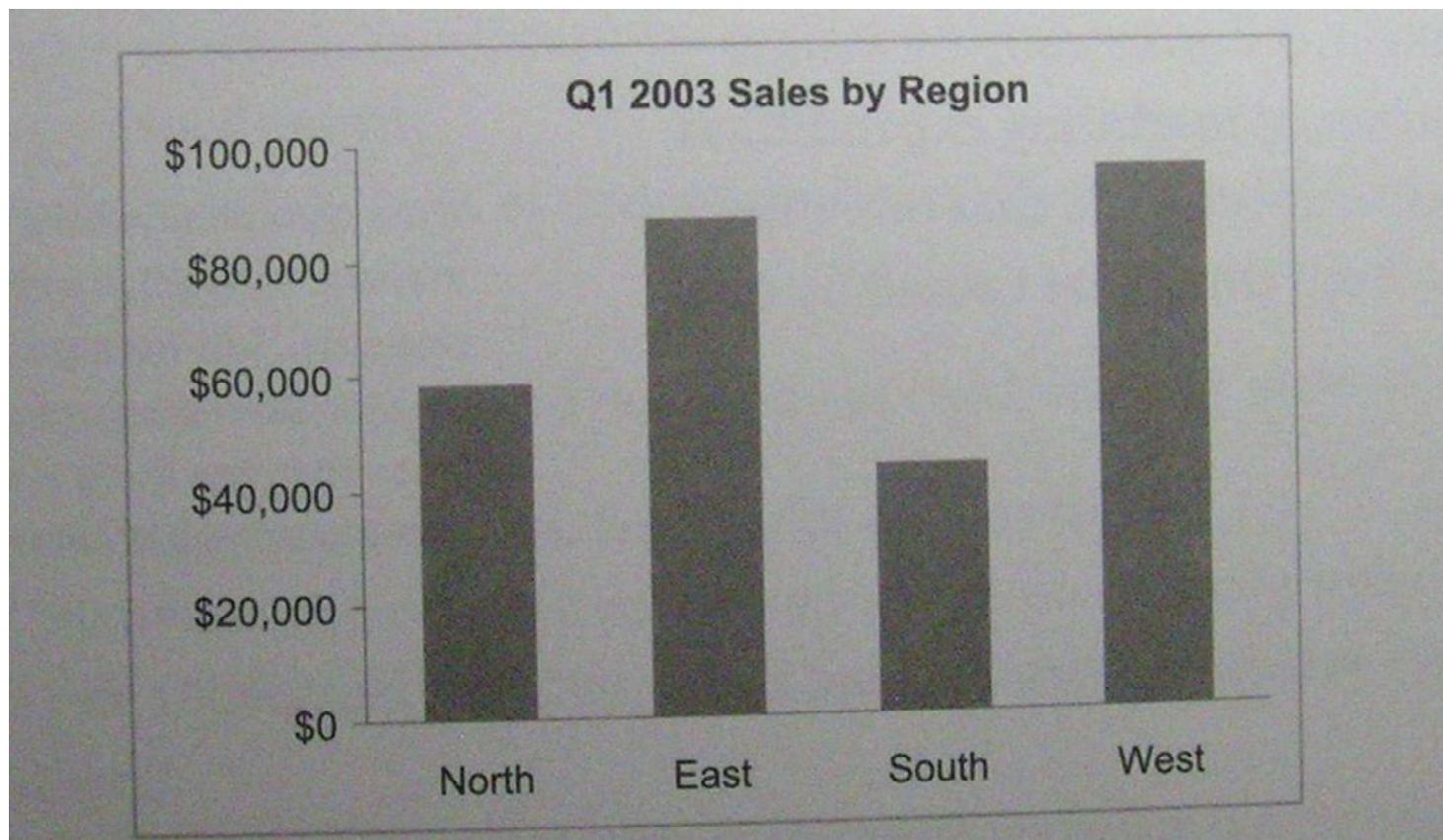


# Relationships in Graphs

- Nominal comparison
- Time series
- Ranking
- Part-to-whole
- Deviation
- Distribution
- Correlation

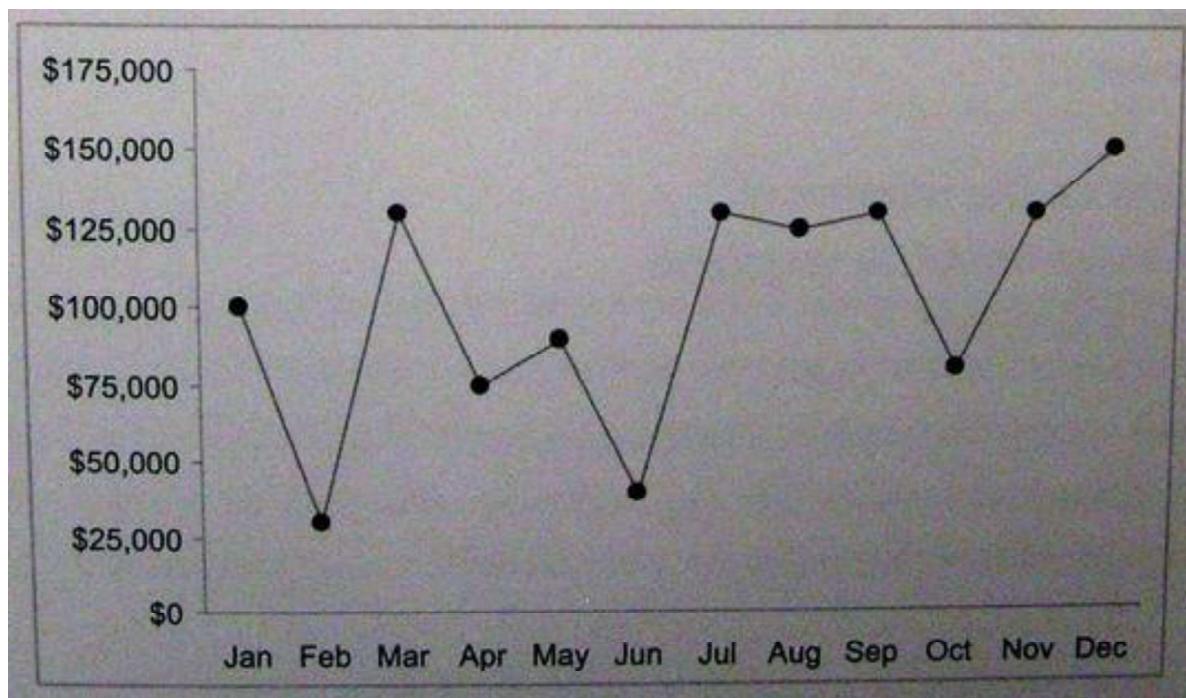
# Nominal comparison

- Nominal categorical attribute
- Quantitative values that are compared each other



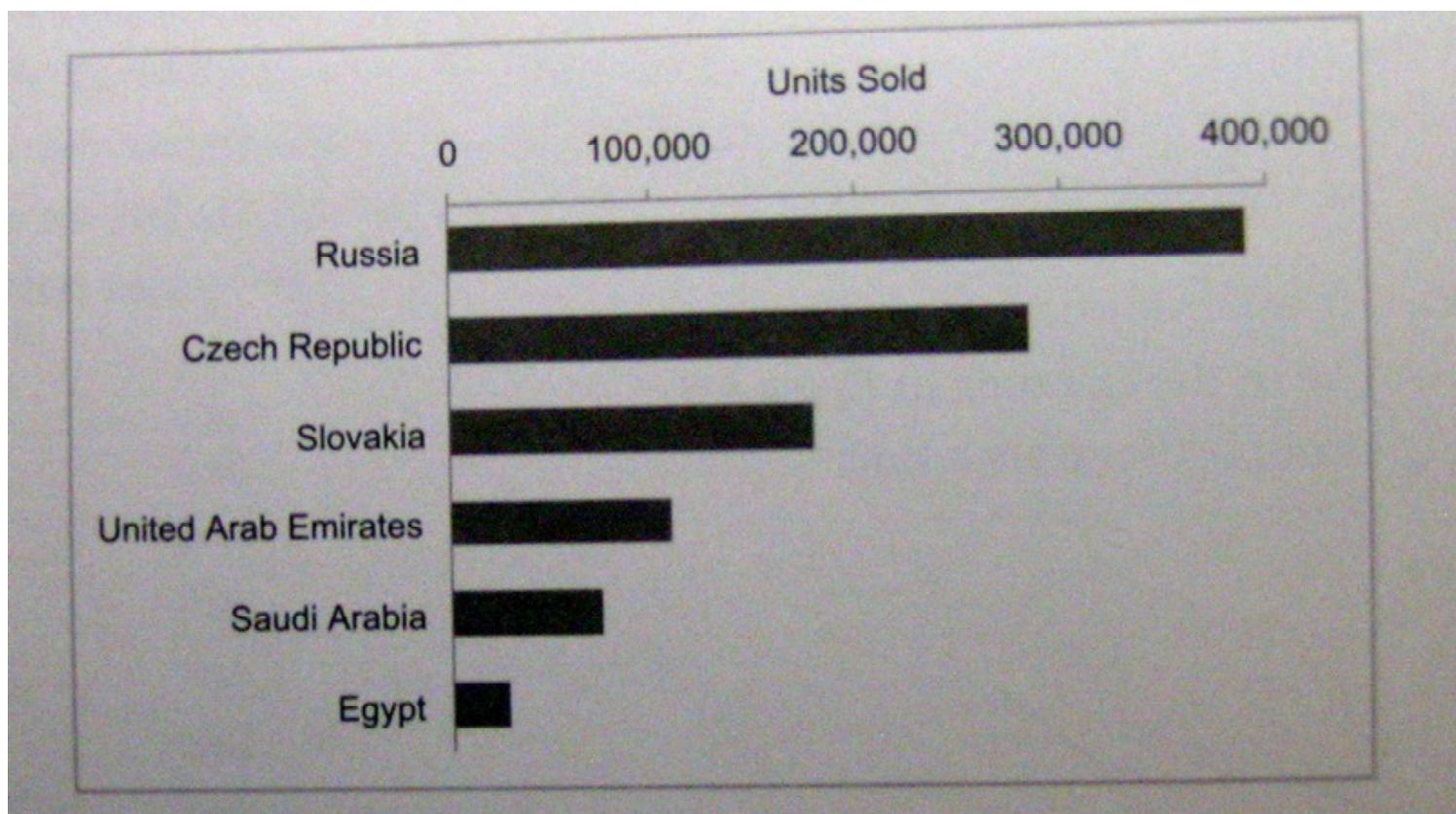
# Time series

- Time categorical subdivision
- Quantitative values that are compared each other for
  - Change
  - Rise
  - Fluctuate
  - Decline
  - Trend
  - ..



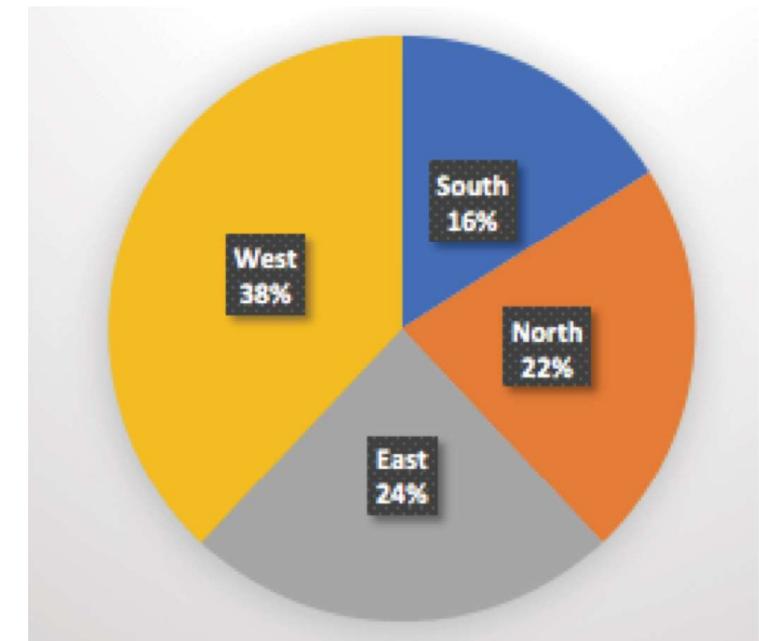
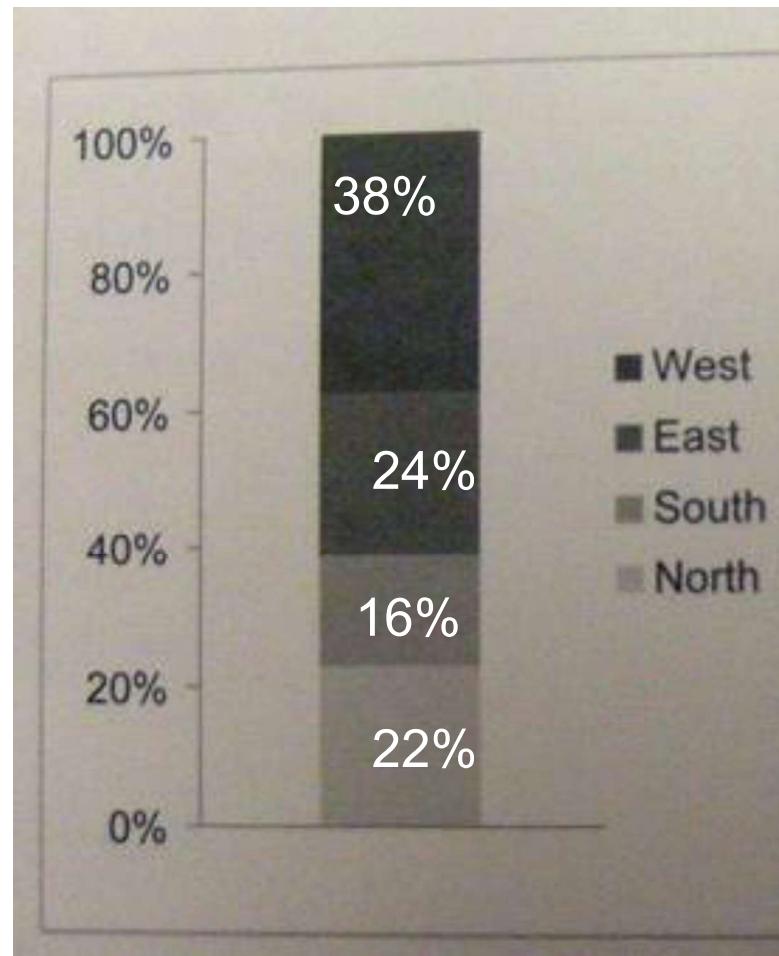
# Ranking

- Categorical subdivision sorted by size
- Quantitative values that are compared each other for
  - Larger than
  - Smaller than
  - Equal to
  - $n^{\text{th}}$  position
  - ...



# Part-to-whole (Proportions)

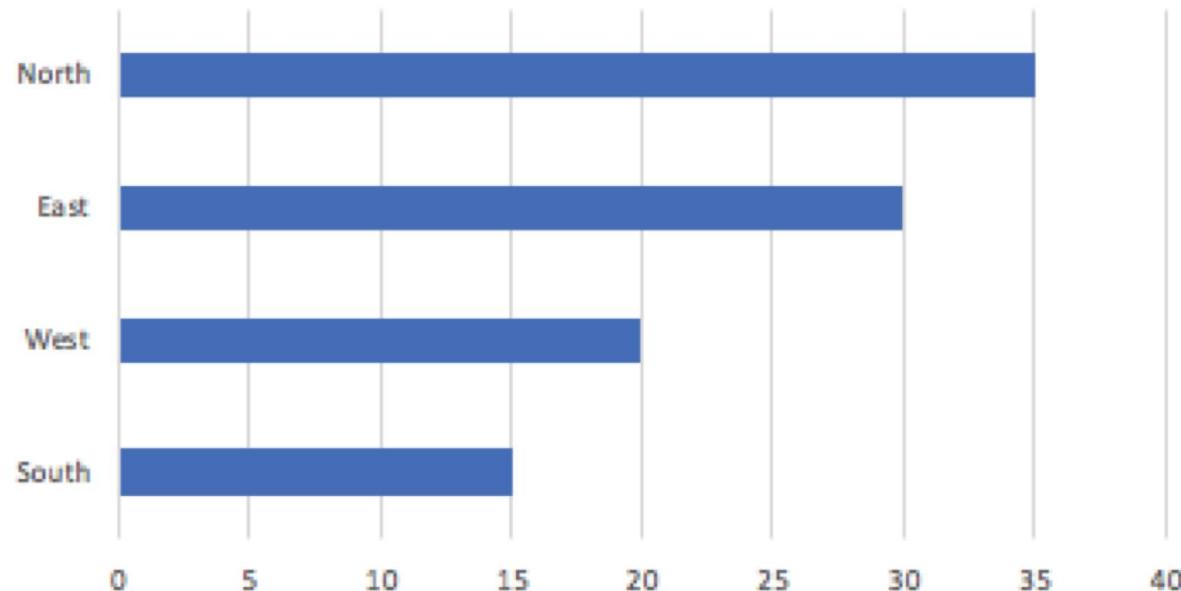
- How individual quantitative values, associated to categorical relate to the complete set of values
- Usually expressed as percentage
- Quantitative values that are compared each other for
  - Percent
  - Share
  - ...



Problems with areas  
use numbers!

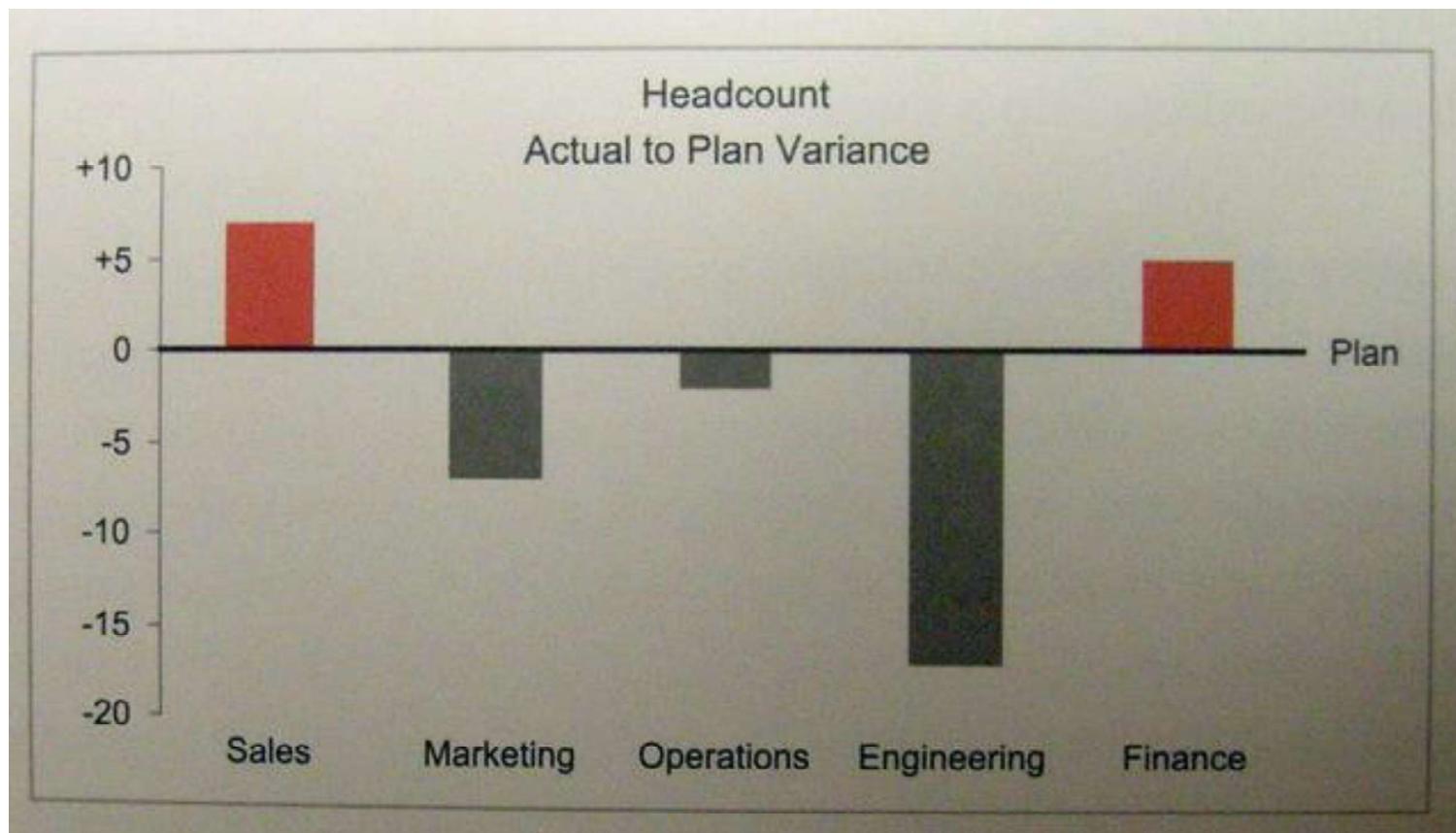
# Part-to-whole comparison

- Better (but uses more space...)



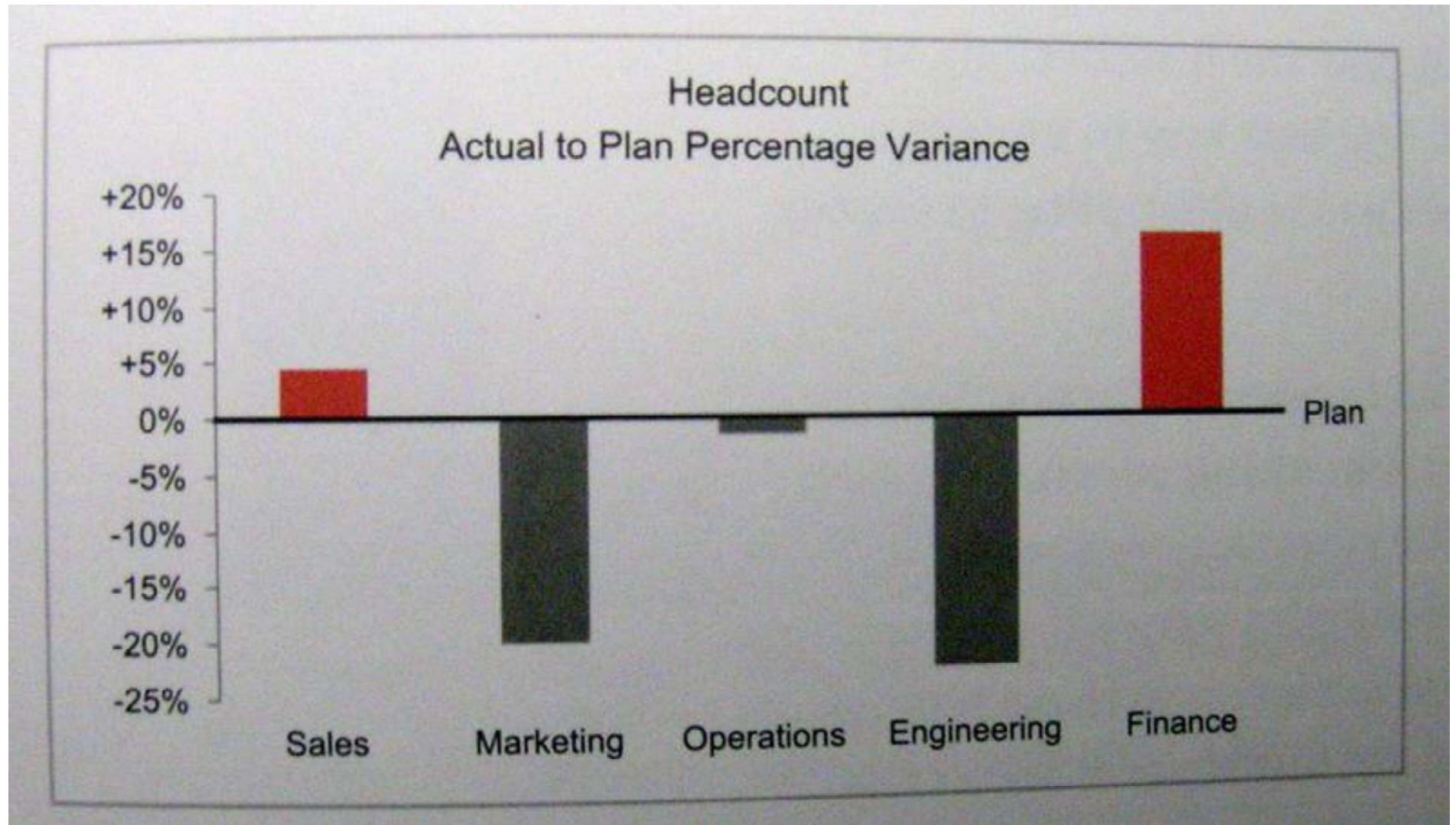
# Deviation

- The degree to which one or more quantitative values differ in relation to a primary set of values
- Color is categorical: bad vs good data



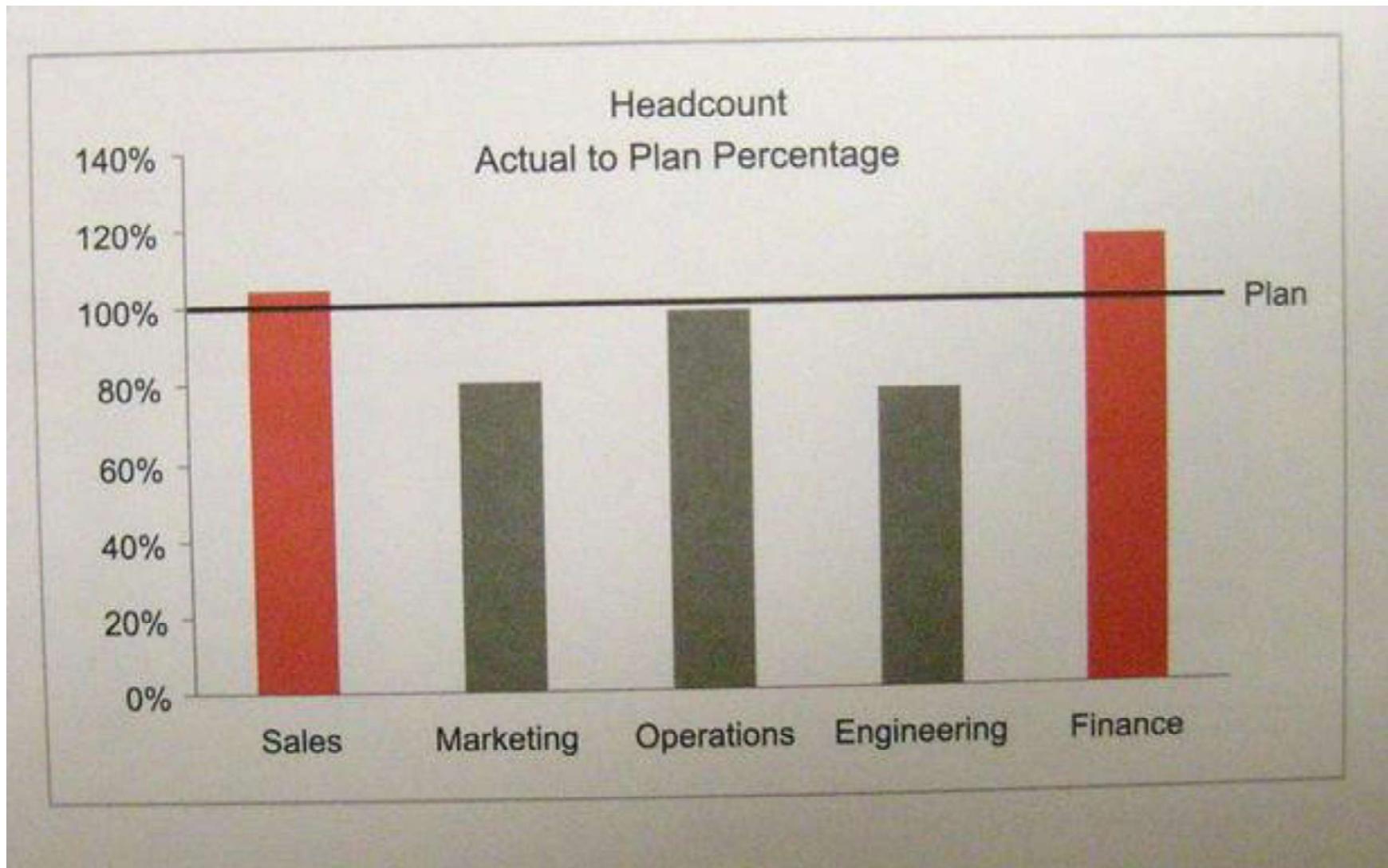
# Deviation design

- Same data as percentage

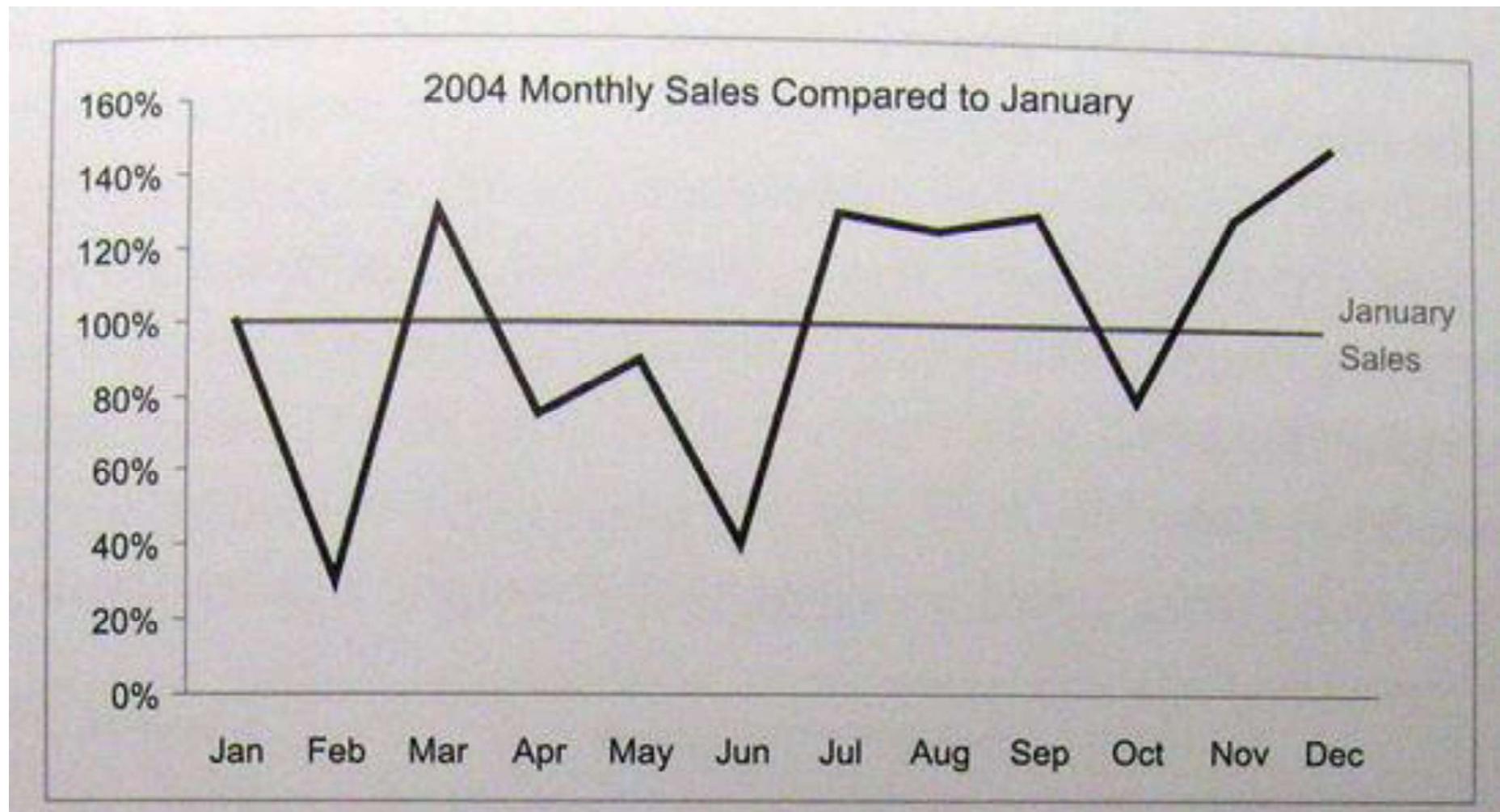


# Deviation design

- Same data as percentage

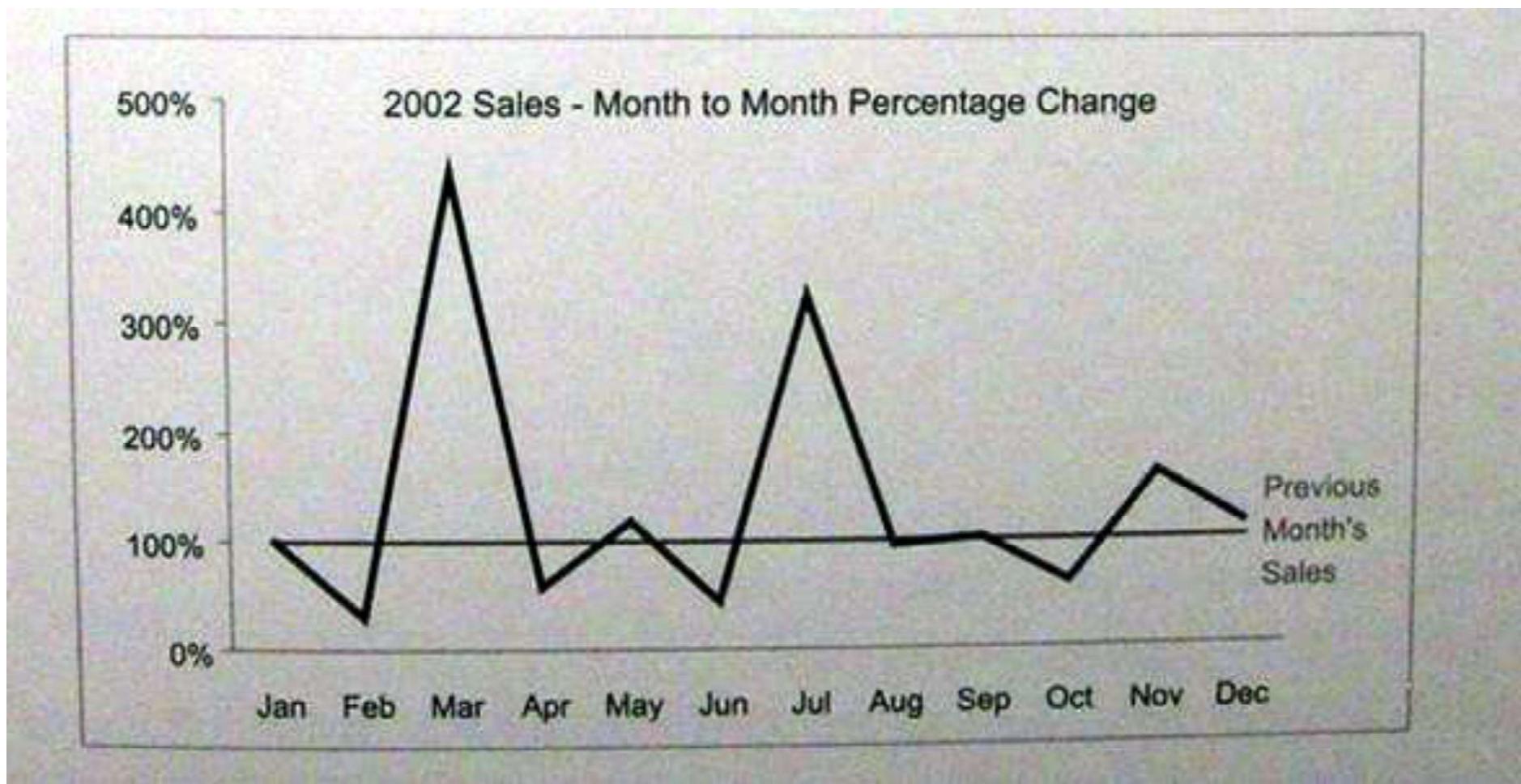


# Deviation design + time-series



# Deviation design + time-series

- Note that the horizontal line represents very different values



# Control chart

$\mu$  and  $\sigma$  can give more information to expert people

The graph represents software components (x) and errors per kloc (y)

In the bottom graph outliers have been removed

What is wrong with this graph?

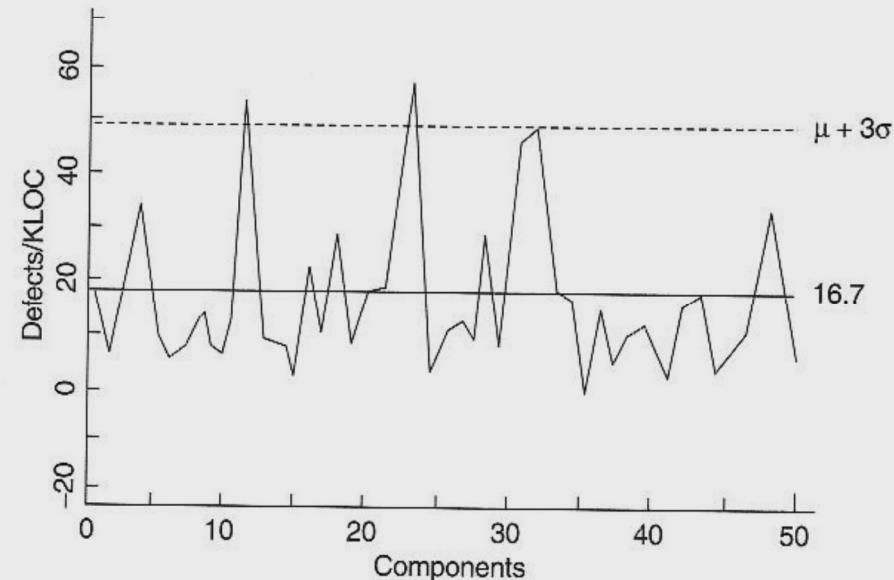
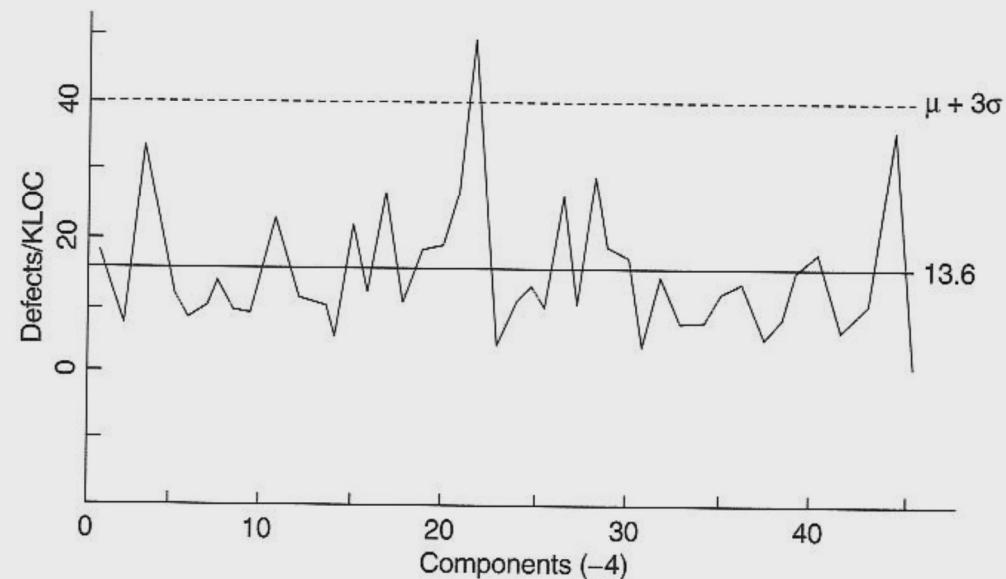
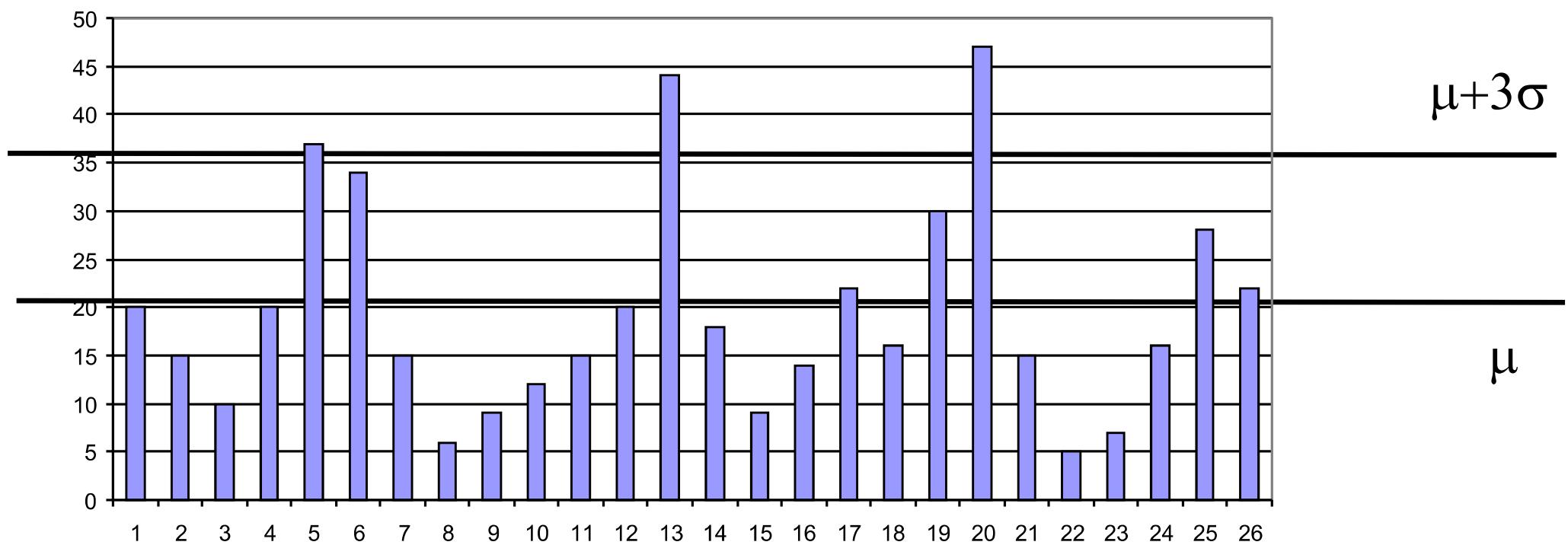


FIGURE 5.12  
Pseudo-Control Chart of Test Defect Rate—First Iteration



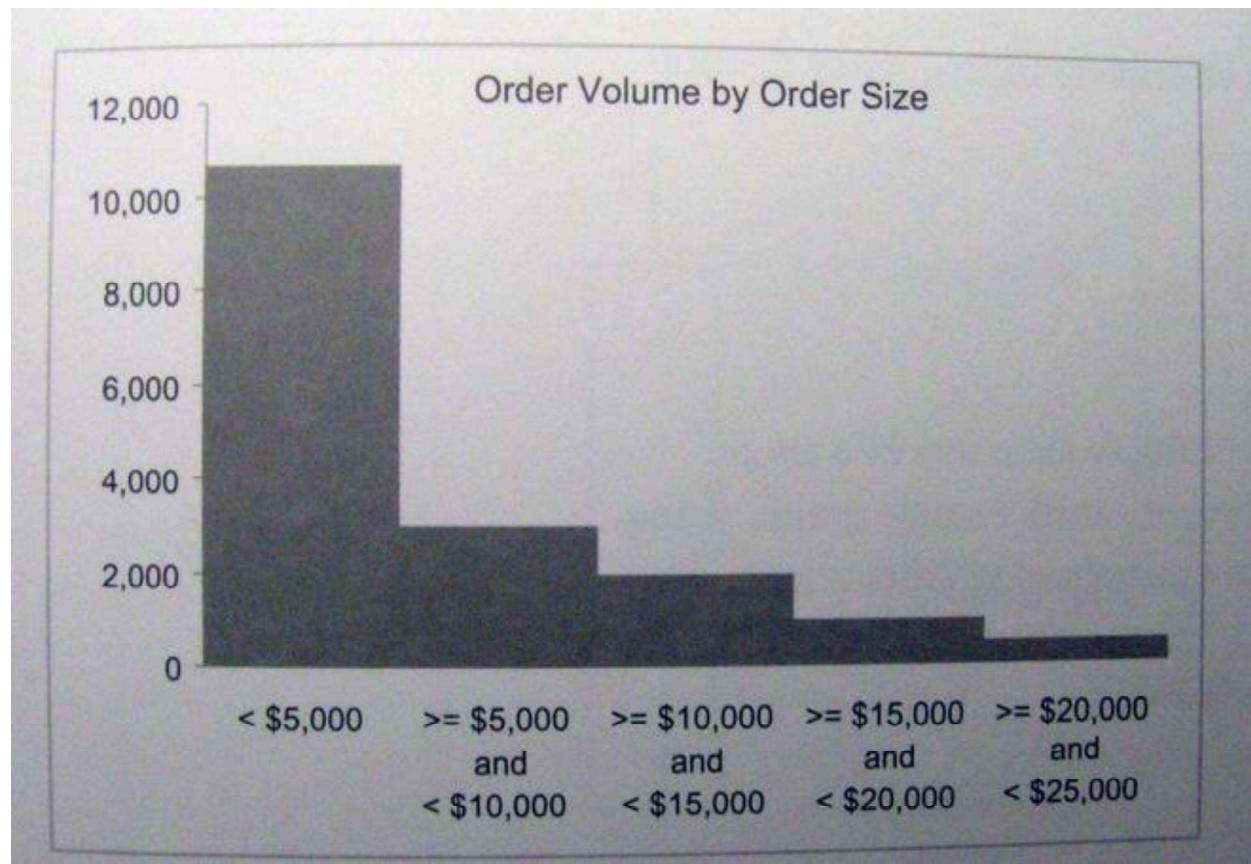
# Lines are wrong here!



$\mu + 3\sigma$

$\mu$

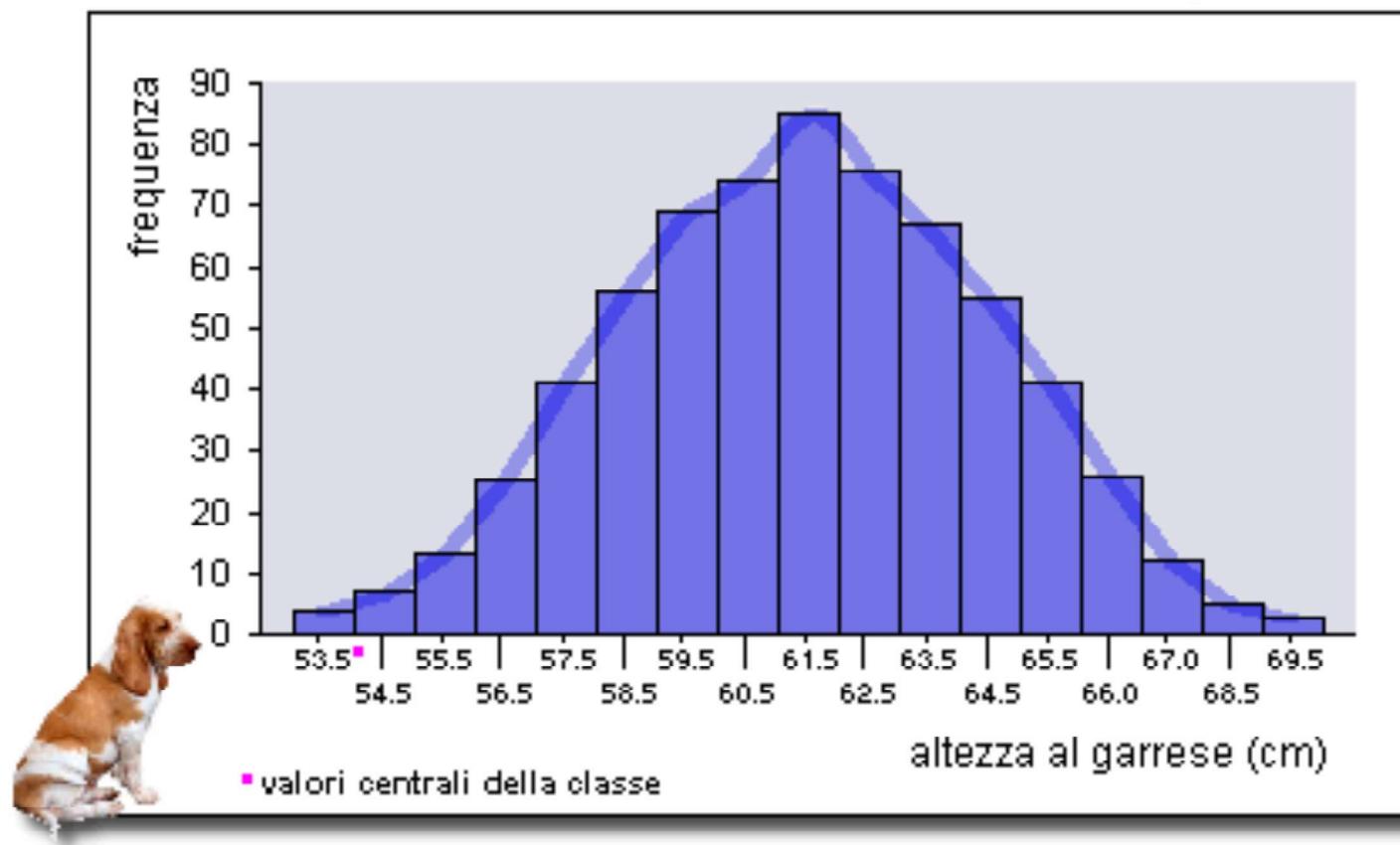
# Distribution (values)



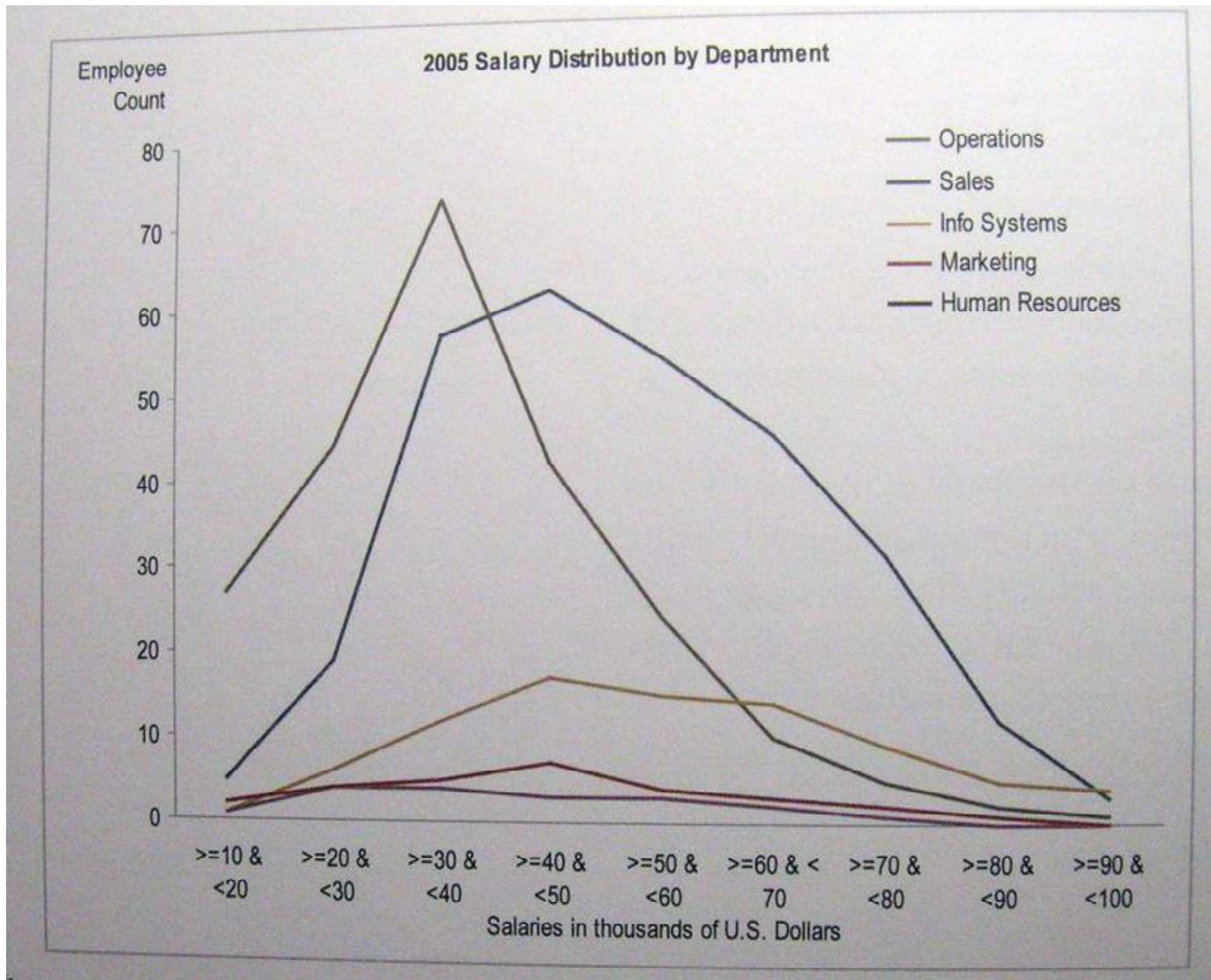
- Histogram

# Distribution (values + shape)

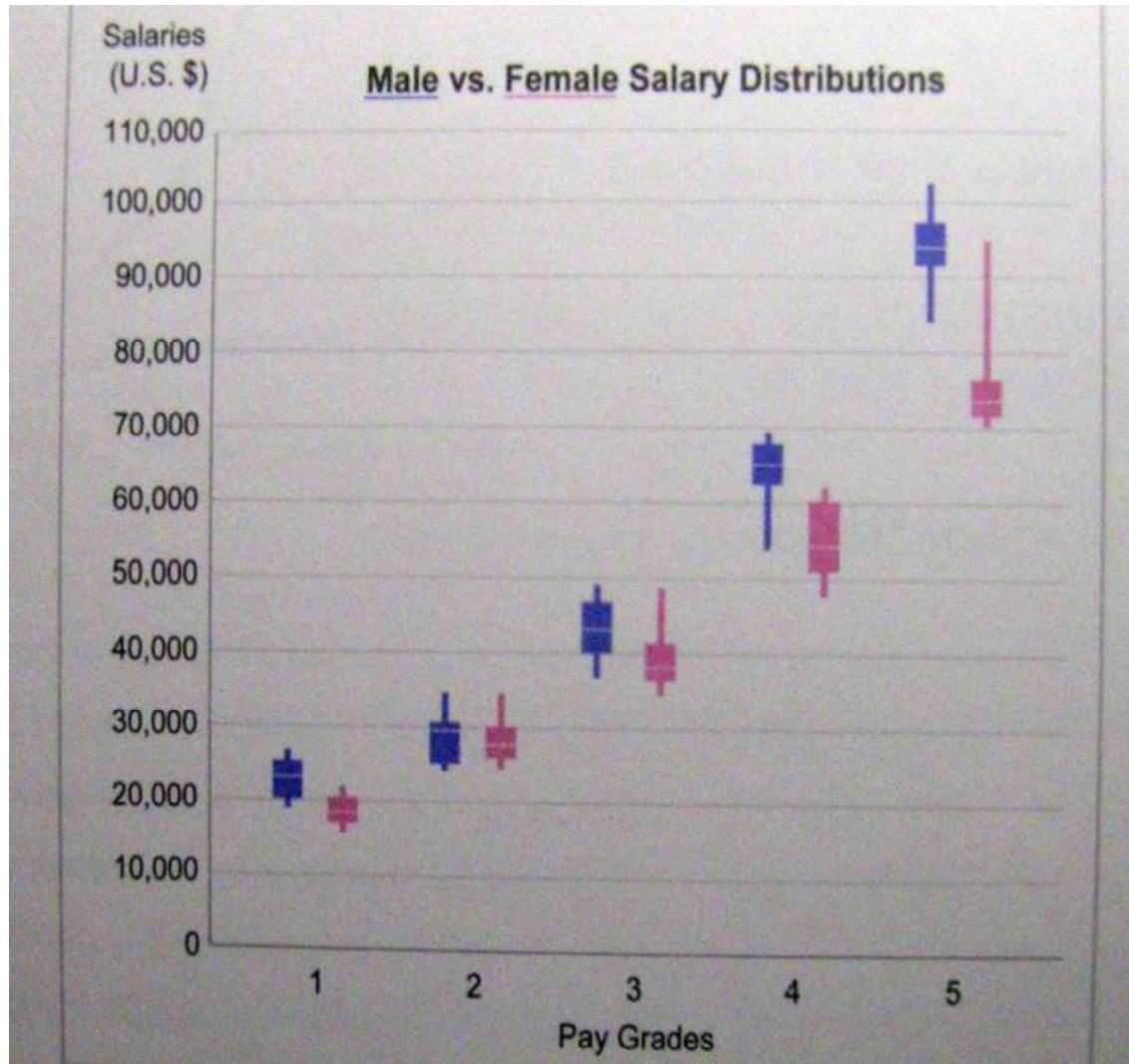
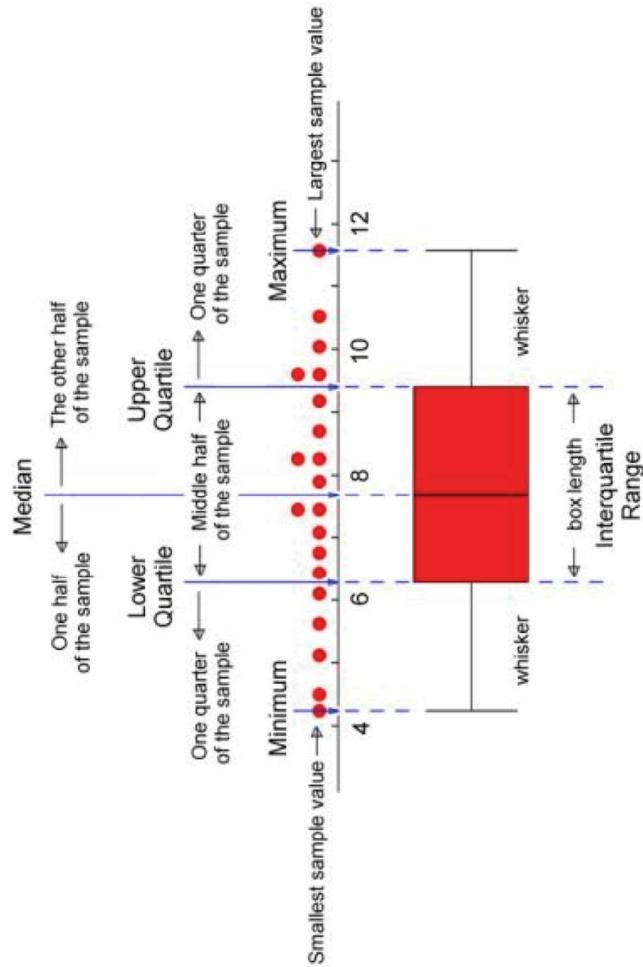
Altezza al garrese di 659 cani di razza "Bracco italiano". Istogramma.



# Multiple distributions (shapes)

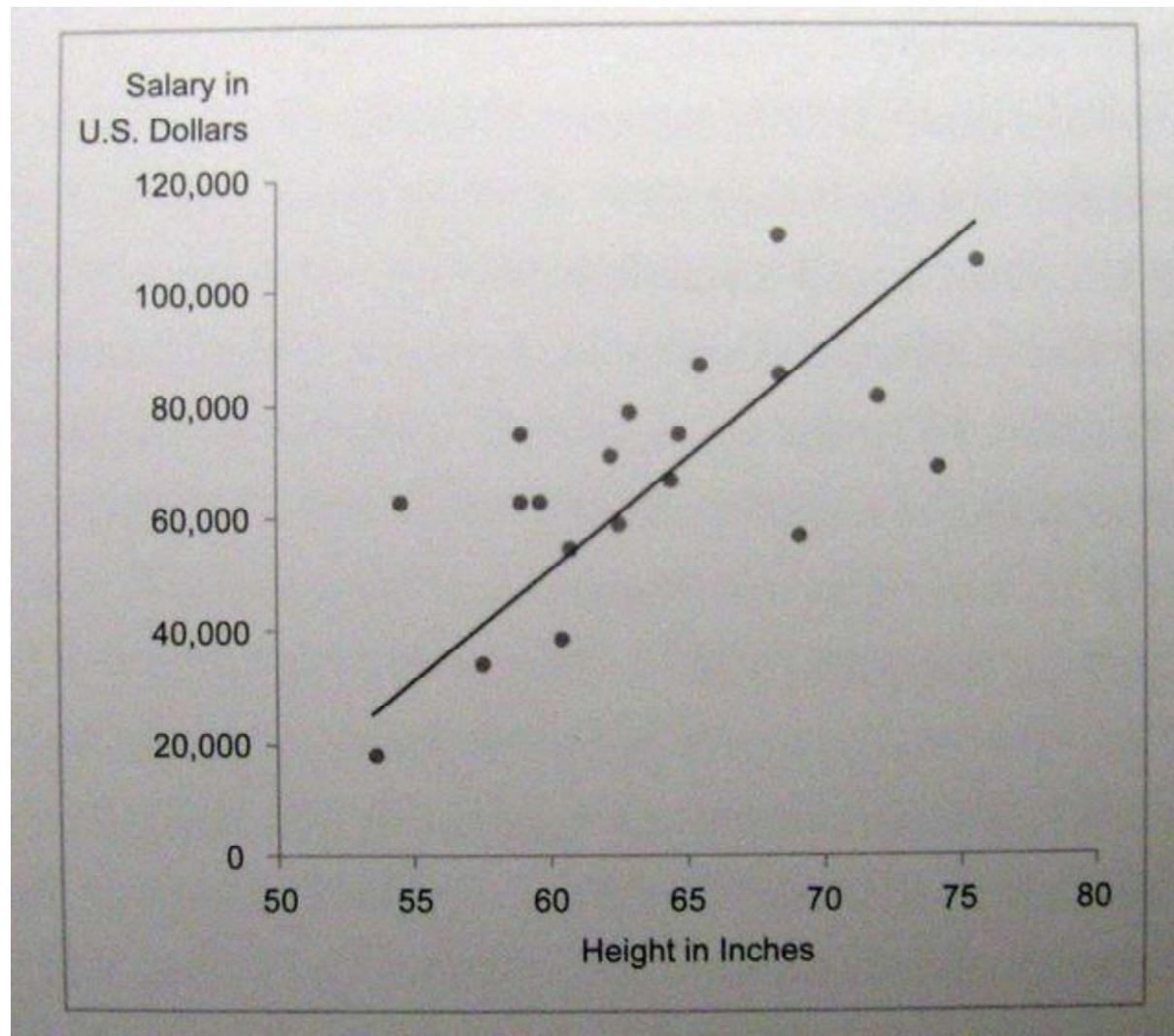


# Multiple distributions (boxplots)



1. On average women are paid less
2. The disparity becomes increasingly greater as grade increases
3. Salaries vary the most for women in the highest salary grades (long upper tail)

# Correlation



# Relationships Summary

Relationship	Points	Lines	Points & Lines	Bars
Nominal comparison	When narrowing the scale and removing the zero	Avoid	Avoid	horizontal or vertical
Time series	Avoid	x=time y= quantitative emphasis on trends	x=time y= quantitative emphasis on trends and individual values	x=time y= quantitative emphasis on individual values
Ranking	When narrowing the scale and removing the zero	Avoid	Avoid	horizontal or vertical
Part-to-whole	Avoid	Avoid	Avoid	horizontal or vertical
Deviation	Avoid	Useful combined with time series	Useful combined with time series and emphasis on individual values	horizontal or vertical vertical with time series
Single Distribution	Avoid	emph. on pattern	Avoid	Histogram
Multiple Distribut.	Use to mark median in boxplots	up to 5 distributions	Avoid	As boxplots
Correlation	Scatter plot	Avoid	Only as a trend (not connecting points)	horizontal or vertical

# Last Remark!

If time is involved take care of scales  
e.g., money (but also college grades)

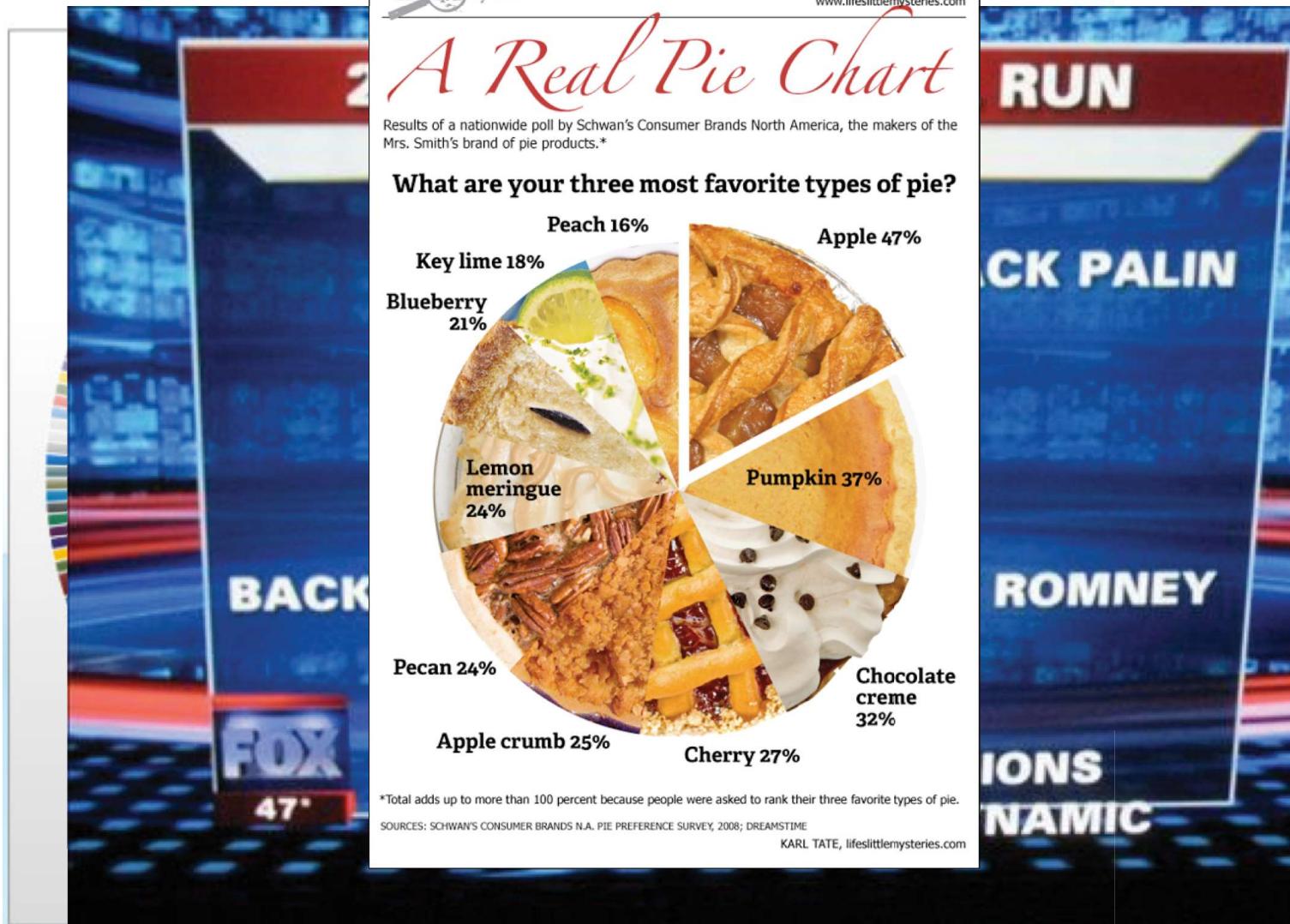
- It is one of the measure whose scale changes across time
  - inflation / deflation
  - change rate
- In comparisons you have to take that into account

# Summarizing

1. Have a good understanding of:
  - involved data (categorical, numerical)
  - involved relationships
    - categorical (nominal, ordinal, interval ...)
    - quantitative (ranking, ratio, proportion ...)
2. If numerical data is involved :
  - select the right summarizing numbers (task, people) to add to the original data (mean, sigma, etc.) Do not lie!
  - Consider time dependent values (e.g., money)
3. Analyze the task(s) (comparison, trend, lookup, etc.)
  - Data
  - Relationships
4. To table or to graph?
5. Table: select table variation
6. Graph: select visual variables



# And avoid abusing of pies...



# Home exercises

- Collect all the scores of your Bachelor exams
- Visualize them showing
  - Temporal distribution
  - Local and global performance (score, average, etc.)
- Assume that you worked within a group of 3-4 people, studying and passing the same exam together
- Visualize the story of the group (use random data)
  - Showing group and individual performance w.r.t. an exam
- In both cases classify data, relationships, visual variable/variations
- Use what you want: Excel, PowerPoint, pencil and paper...
- Send to [santucci@diag.uniroma1.it](mailto:santucci@diag.uniroma1.it) the results