Data visualization

Basic design principles and types

David Hoksza
http://siret.ms.mff.cuni.cz/hoksza

Challenge of data visualization

- **Determining** the **medium** (visualization) which tells the story best
  - Table
  - Graph
  - Schema
  - …
- **Design** the **components** of the medium in such a way that the story is relayed clearly
  - Which data to emphasize and which to play down
  - Colors
  - …

Tables vs graphs

Tables

- Looking up **individual values**
- Required reading of **precise values**
- **Comparing** individual **items** rather than whole series
- More than one unit of measure
- **Multiple levels of aggregation** are needed (summary, average)

Graphs

- Message is contained in **patterns**, **trends** and **exceptions**
- **Set of values** needs to be seen **as a whole** or compared

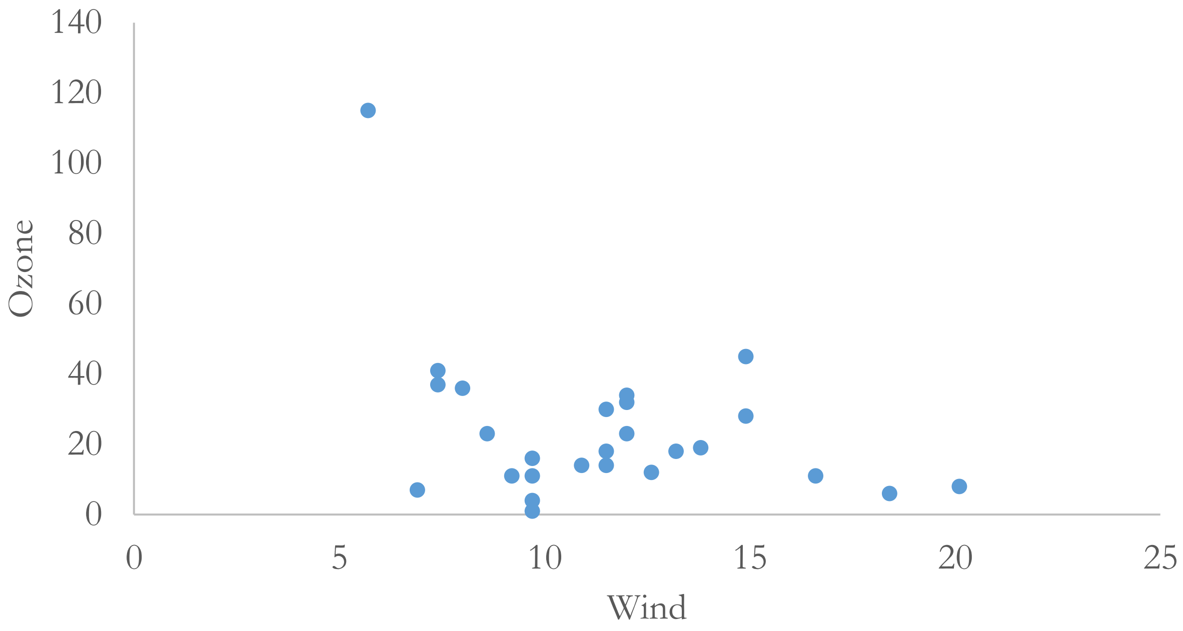
Encoding quantitative values in graphs

- **Means to encode** quantitative values (sales, temperature, …)
  - **Points**
  - **Lines**
  - **Bars**
  - **Boxes**
  - Shapes with varying **2D areas**
  - Shapes with varying **color intensity**

- Each encoding has its strengths and limitations

Points

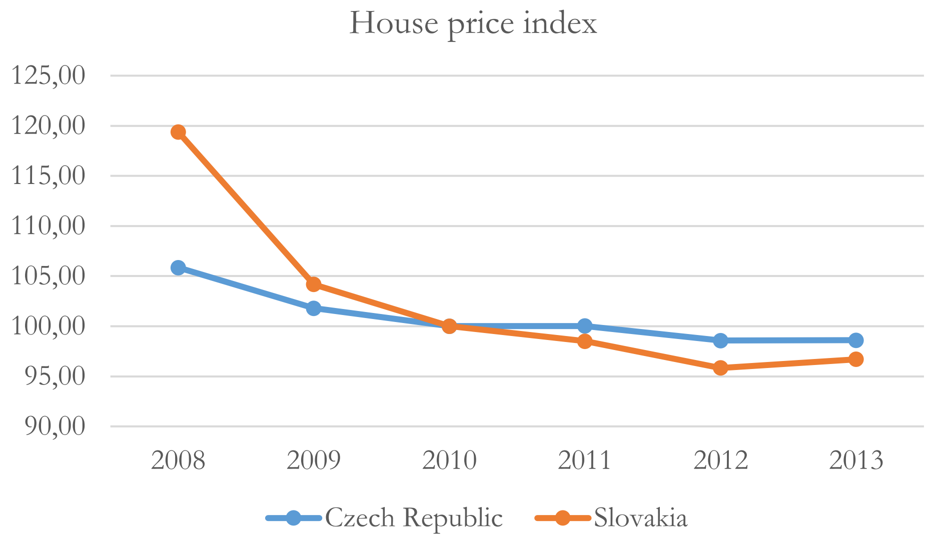
- Small, simple geometrical object used to mark a location on a graph
- **Scatter plot**



Lines

Patterns

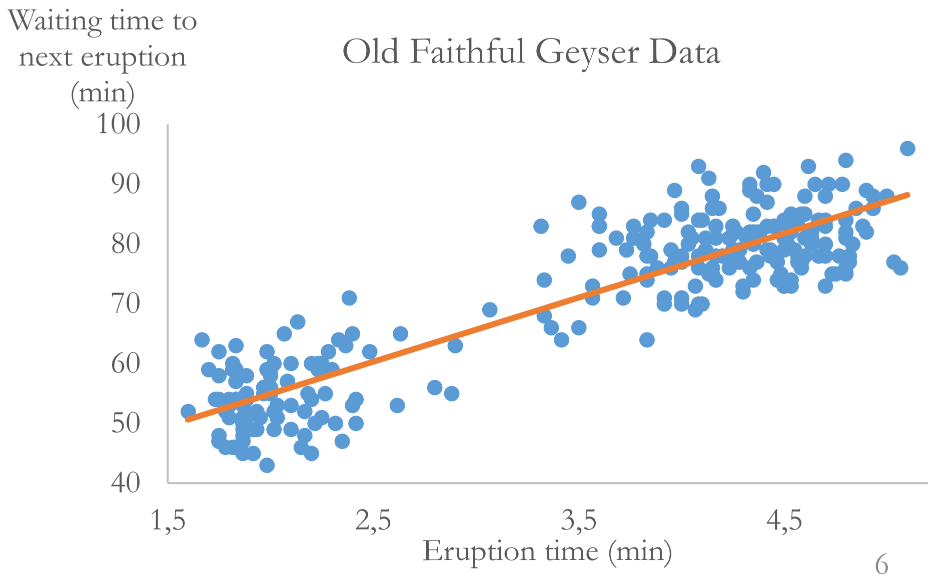
- **Connecting points** by a line enables one to see an entire series of values as a single pattern



data source: Eurostat

Trends

- **Trend lines** (lines of best fits)



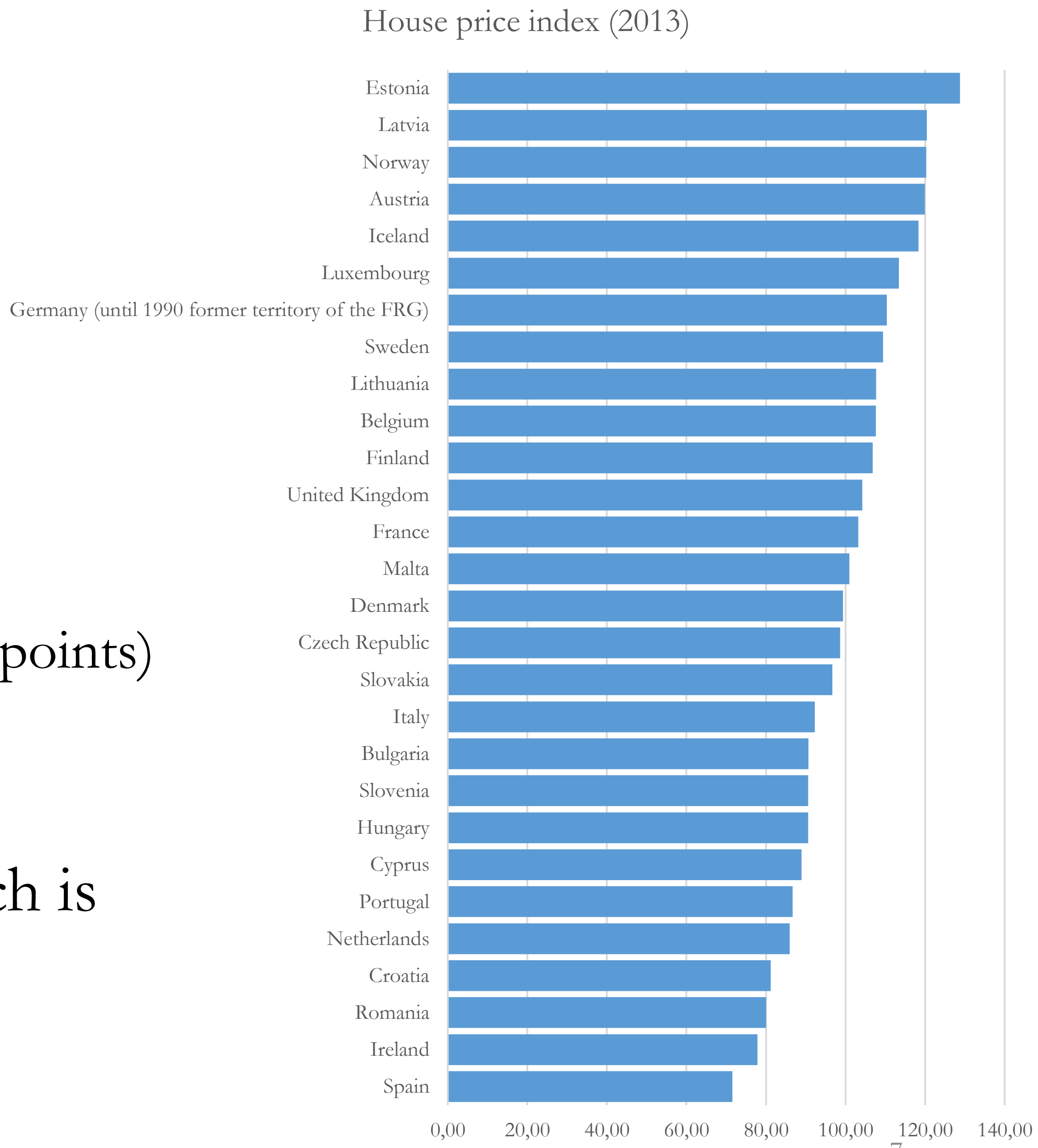
data source: R datasets (faithful)

Bars (1)

- **Bar chart**

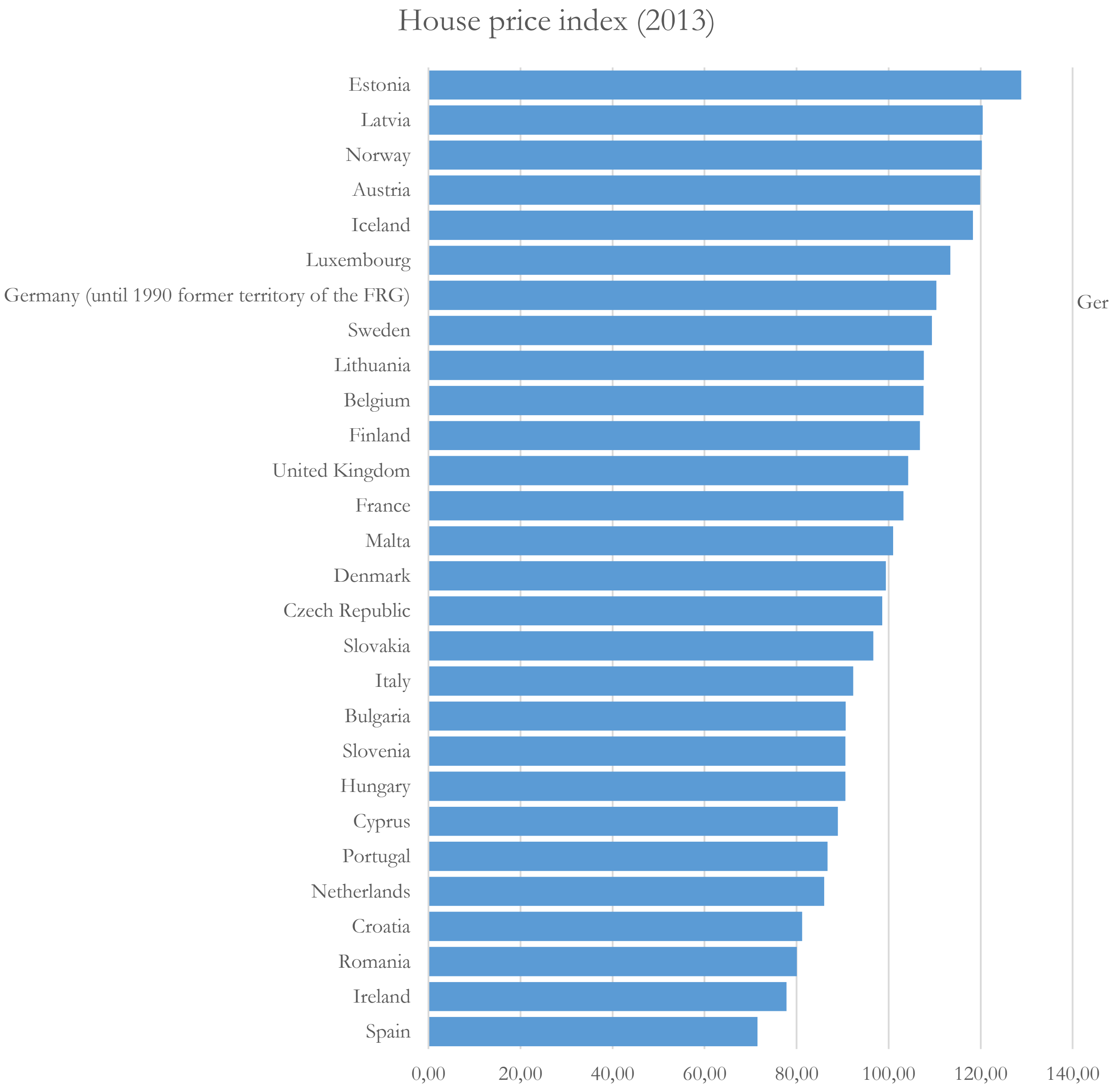
- **Connects** well **labels** with the **values**
  - Well-suited for **comparison** (better than points)
  - Can run both horizontally and vertically

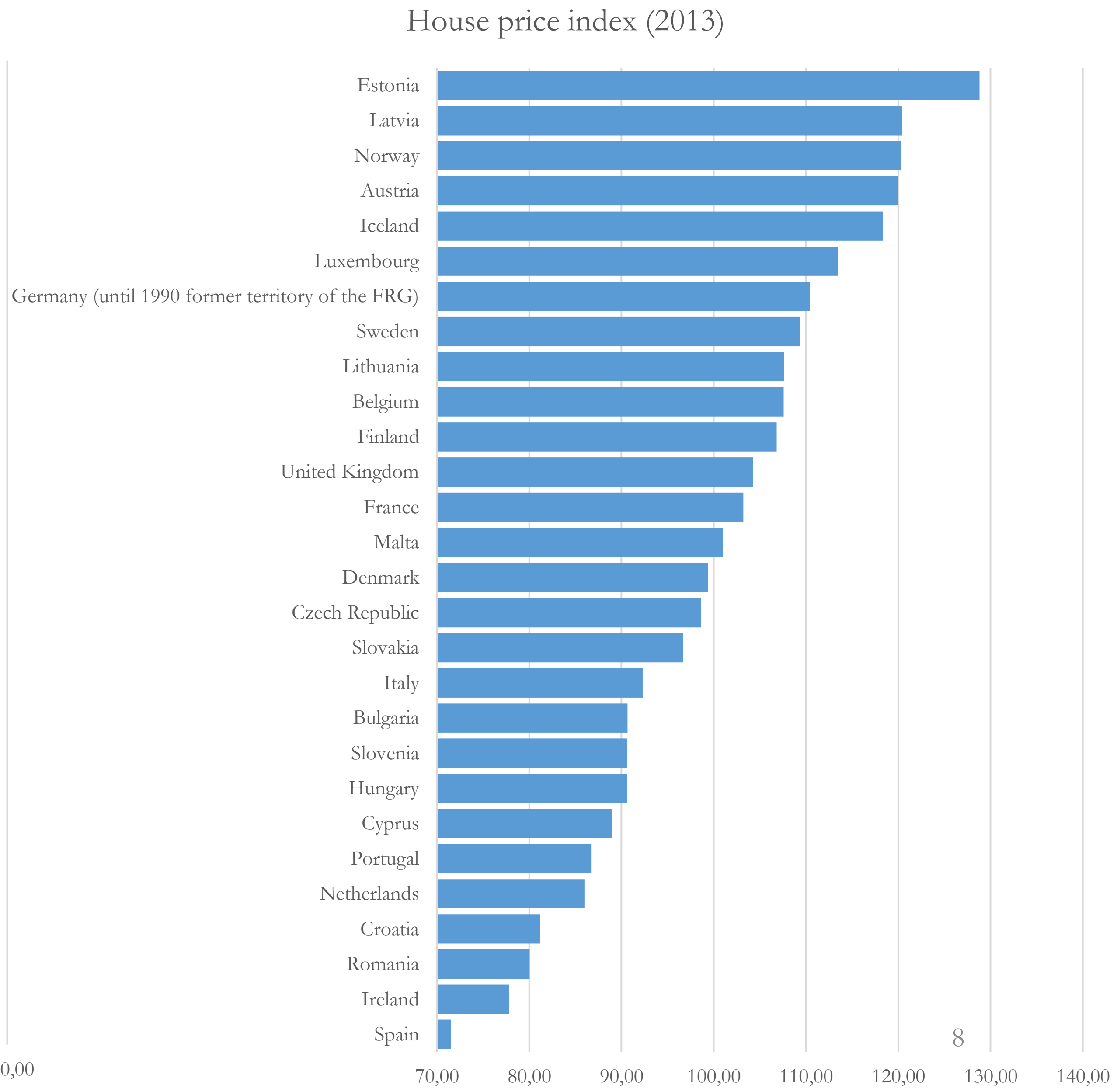
- Adds **second dimension** (width) which is usually not used (and should not)



data source: Eurostat

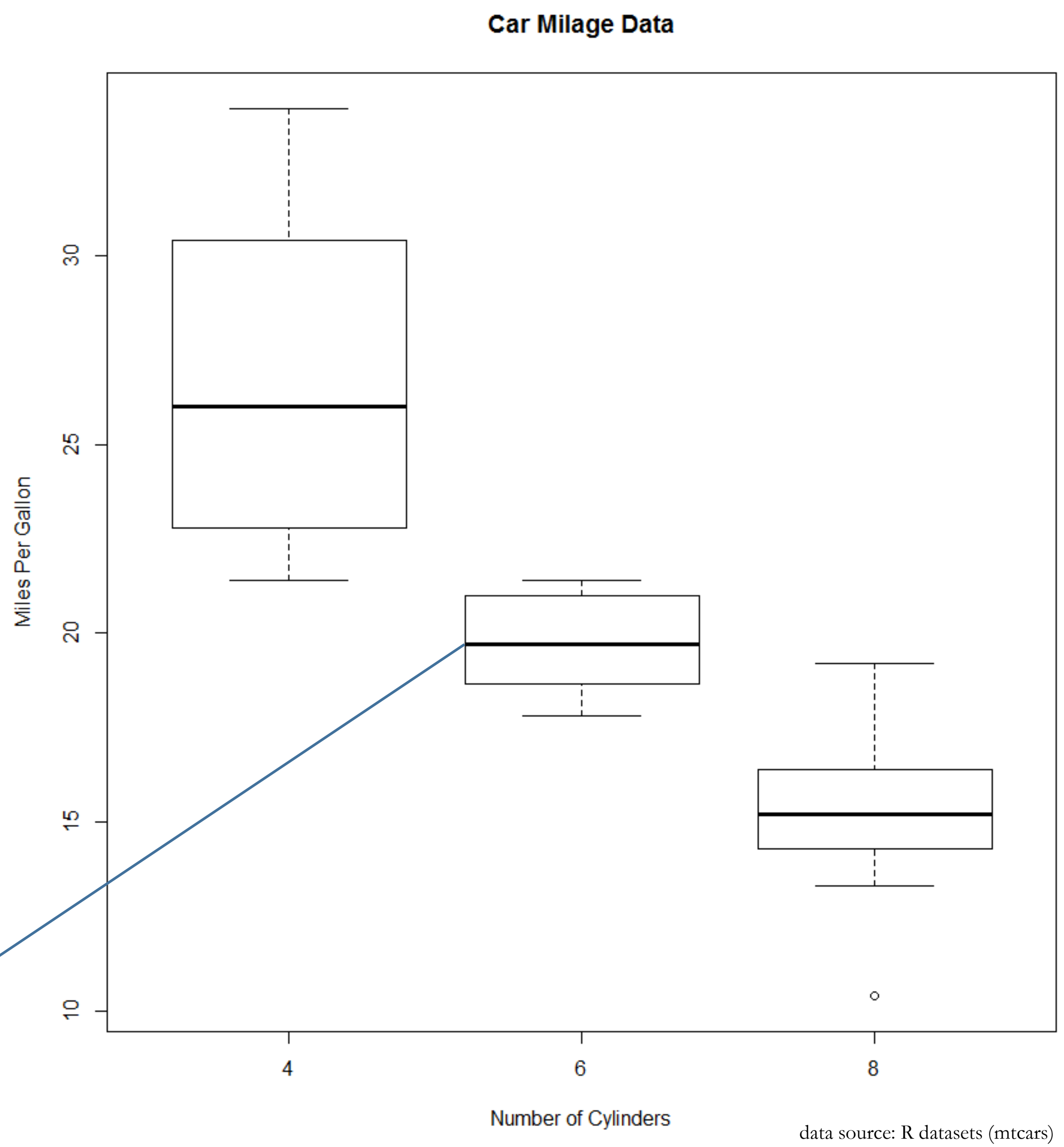
Since bars are good for comparison they also good for "cheating"





Boxes

- Comparison of **distributions** of sets of values → every box represents a set of values → **box plot**



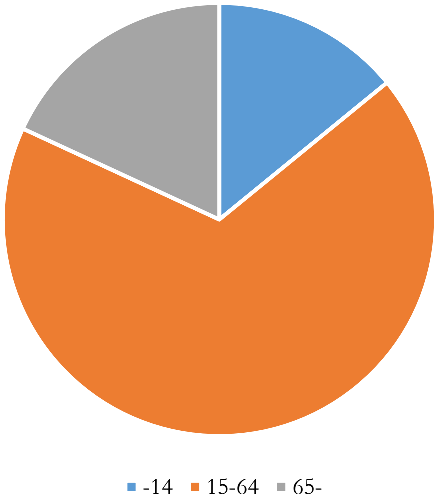
data source: R datasets (mtcars)

Center of distribution (usually median)

Shapes with areas

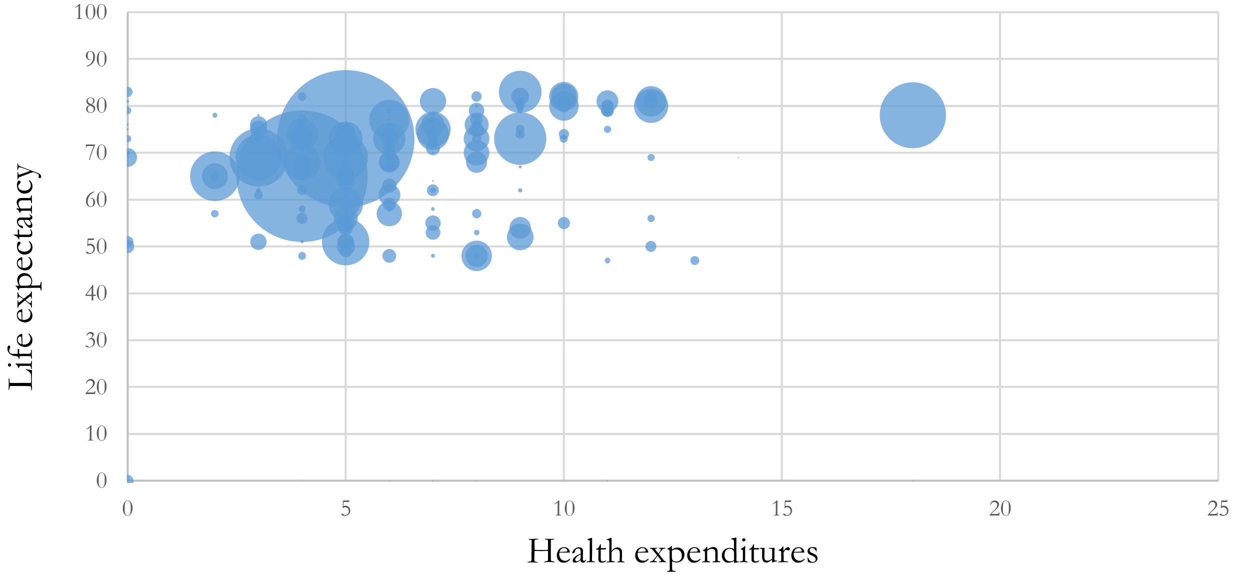
- Representing **values in proportion to their area** (rather than location)
  - Area graphs → **pie chart**
  - Bubbles → **bubble chart**

Age structure in Prague (2013)



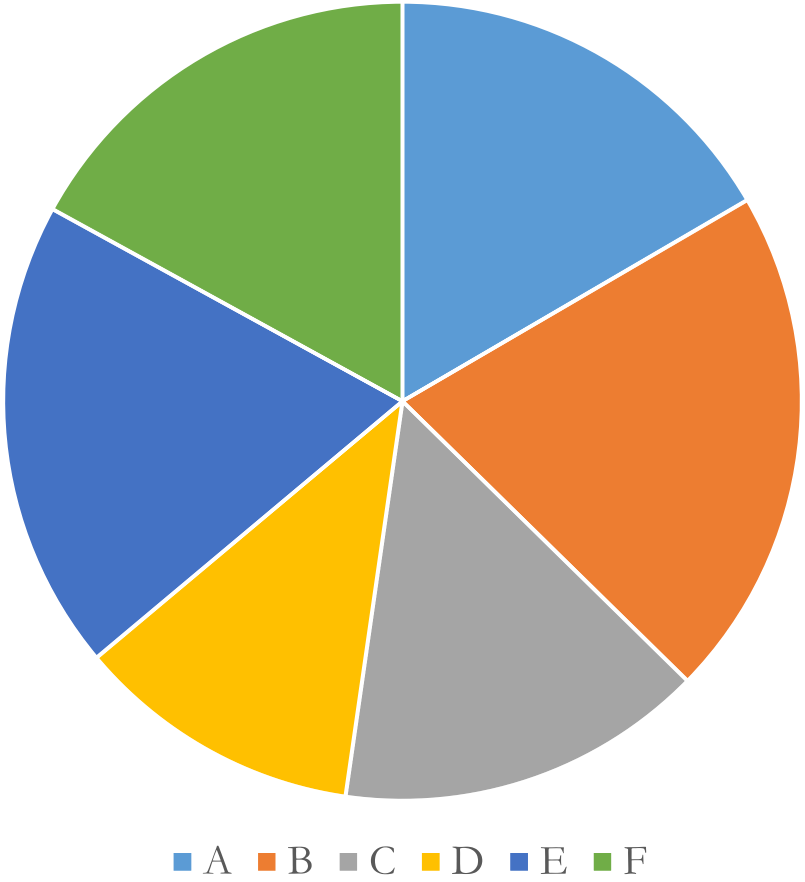
data source: Český statistický úřad

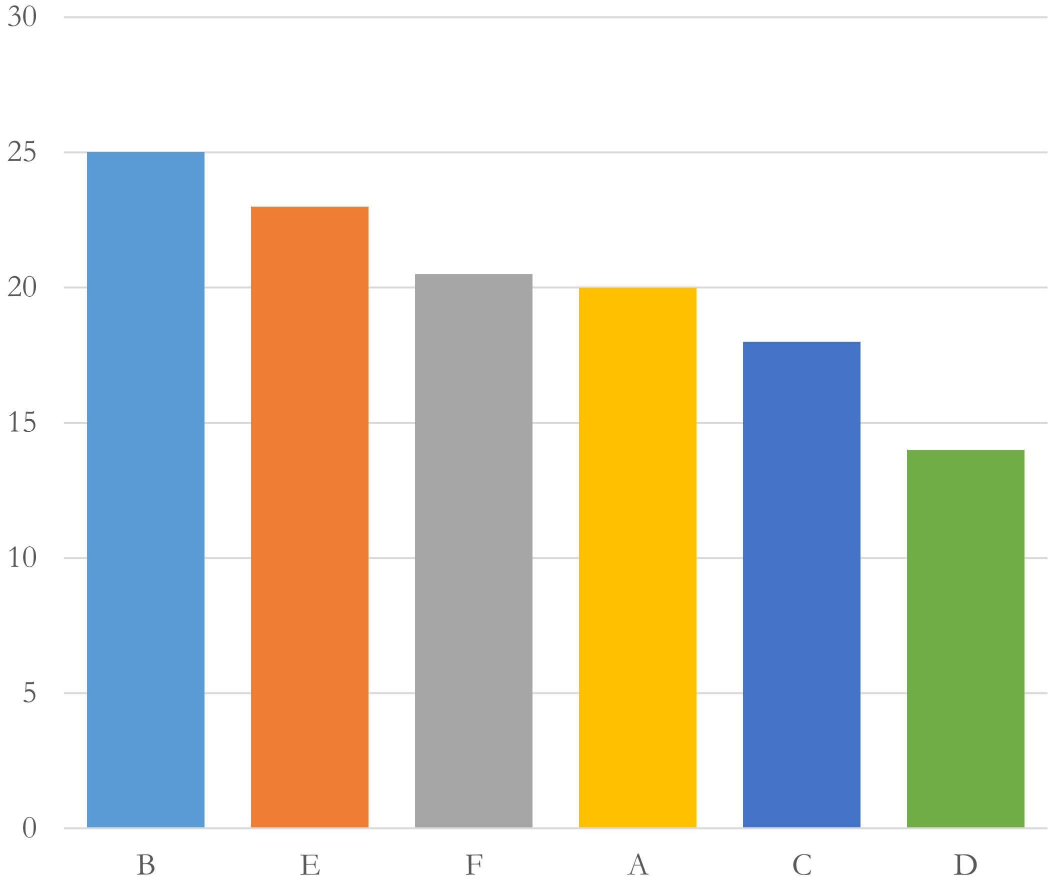
Life expectancy by country
(bubble sizes correspond to population size)



data source: http://www.tableausoftware.com/public/community/sample-data-sets

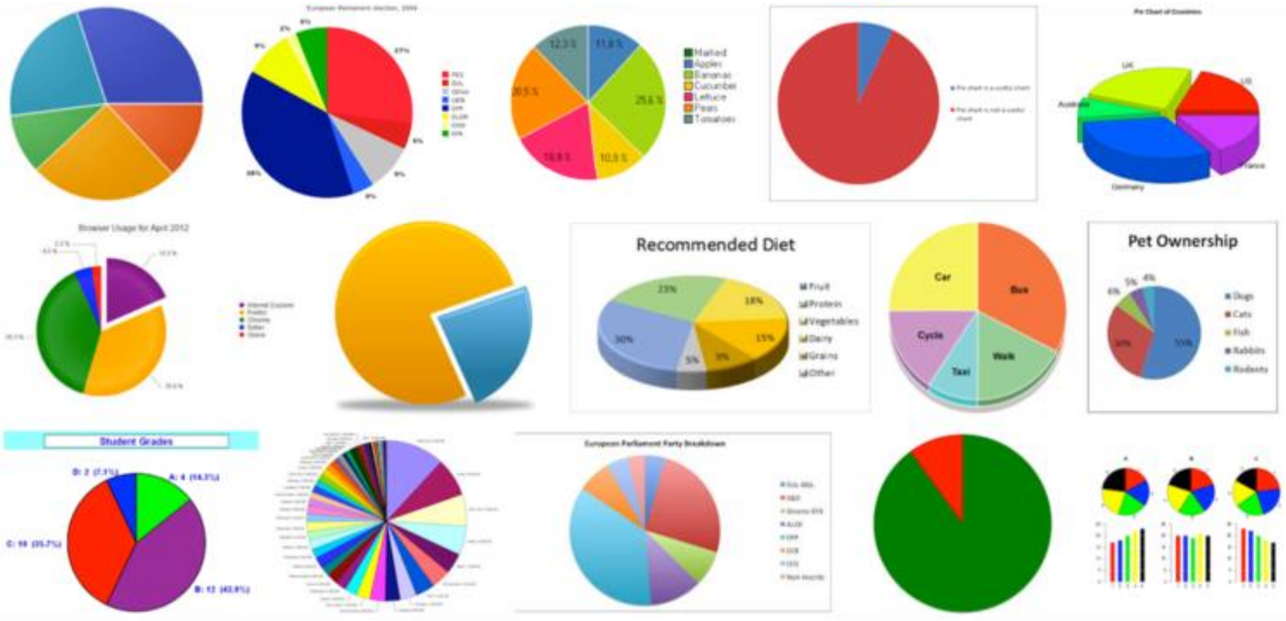
Areas are not suitable for comparison





What not to do with pie charts

- Don't use **3D effects** or explode your pie
- If the pie is depicting percents, it must **sum to 100%**
- Don't have a **ton of slices**
- Don't use a pie if the primary goal is to **compare the size of the slices**
- Don't use **multiple pies** and ask your audience to **compare across** them

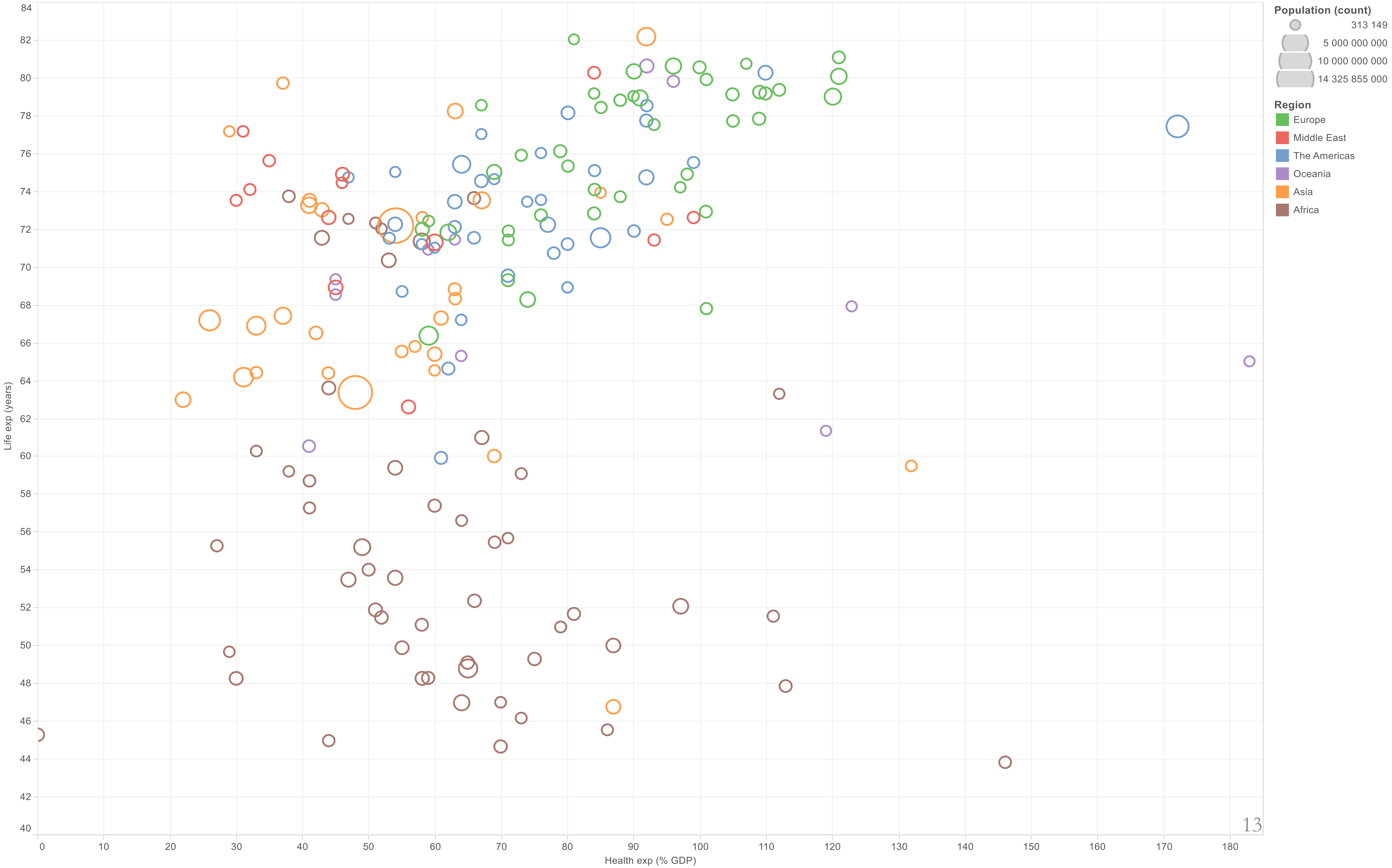


source: http://www.storytellingwithdata.com/blog/2017/1/10/an-updated-post-on-pies

http://www.storytellingwithdata.com/blog/2017/1/10/an-updated-post-on-pies

Shapes with color

- Bubble plot with **varying hue** or **intensity**

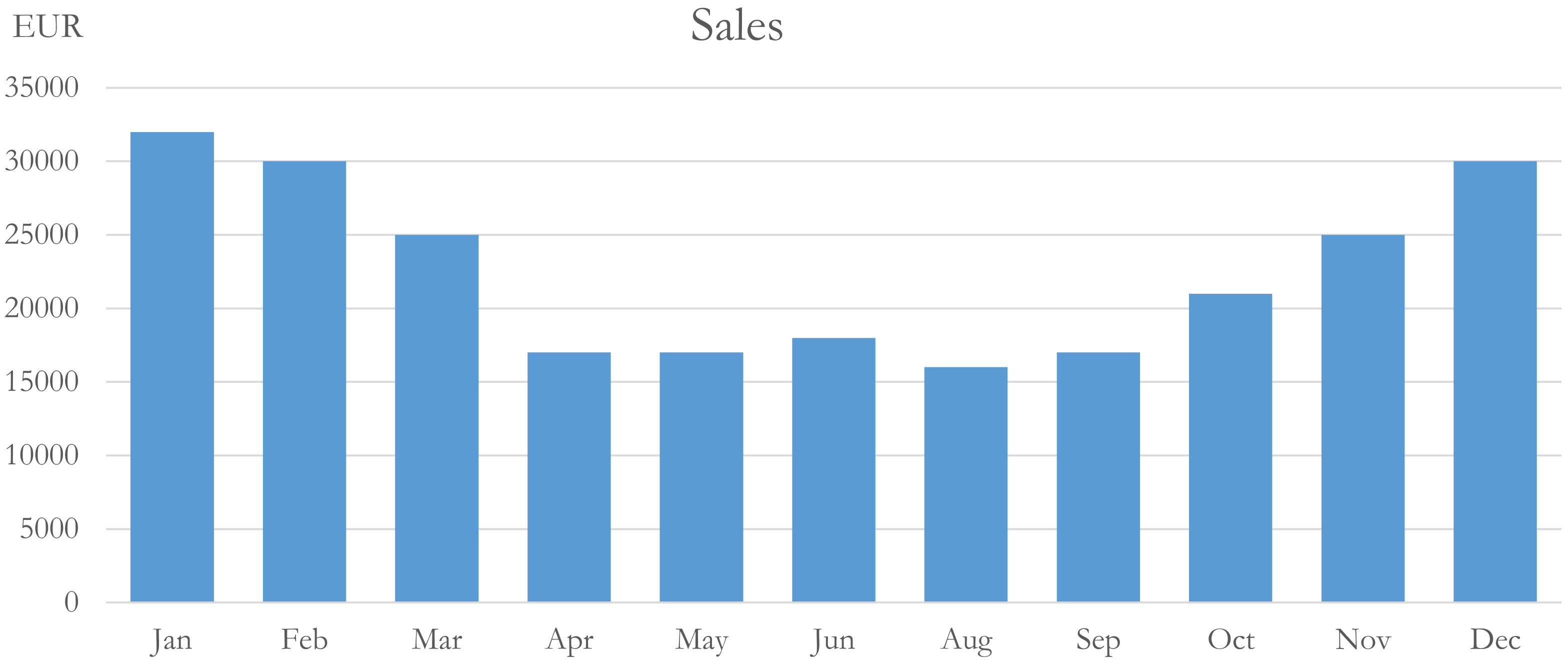


Encoding categorical values in charts

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

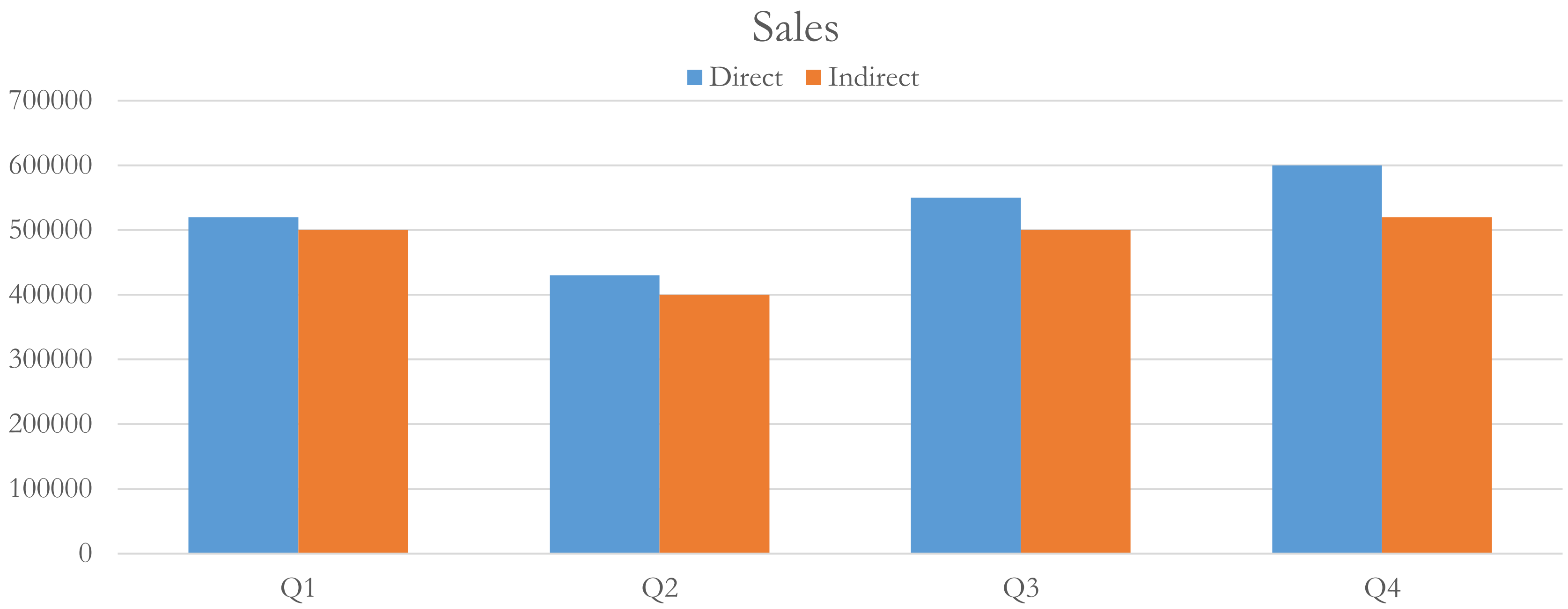
Position

- Most common to identify categorical items
- Works with bars, points, lines or boxes



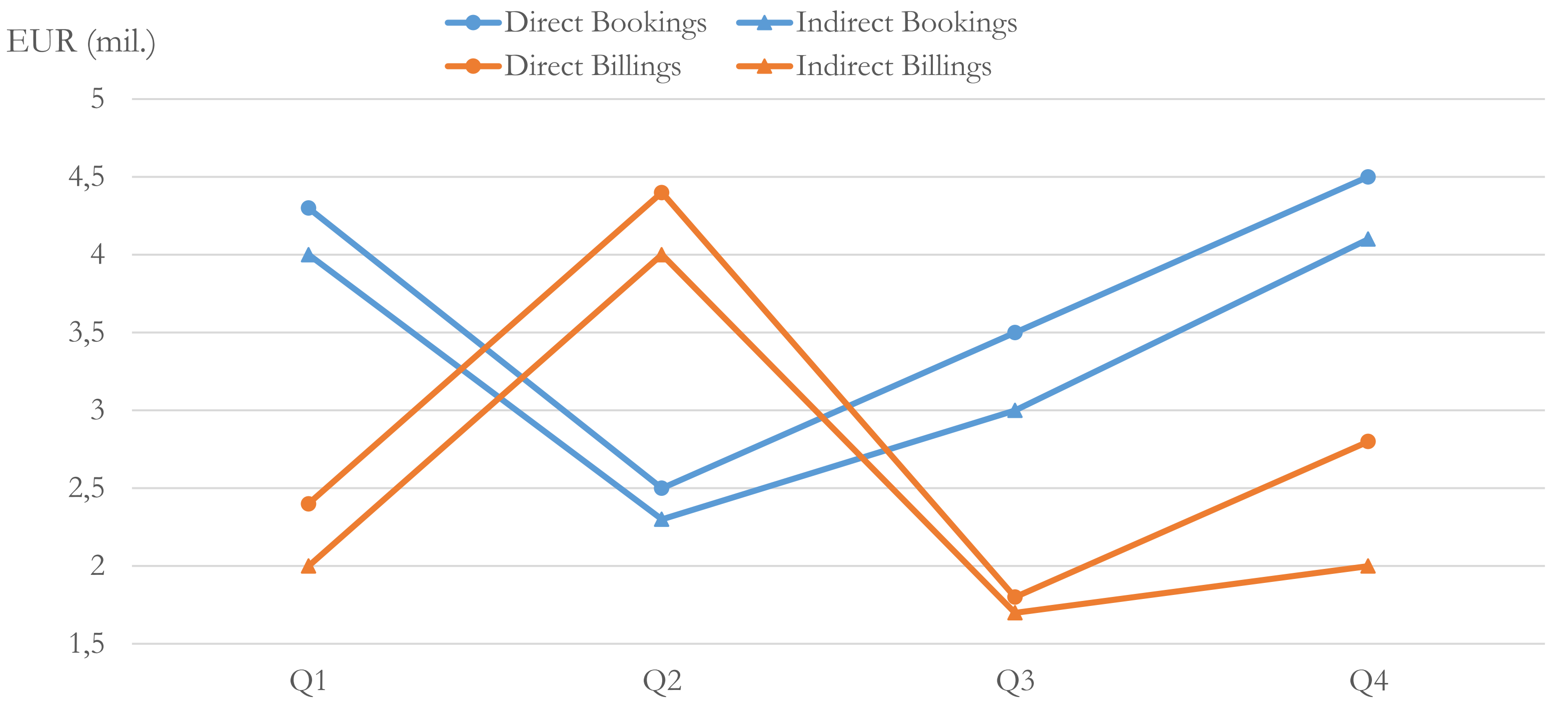
Hue

- **When position is taken**, hue can be used to differentiate categorical items



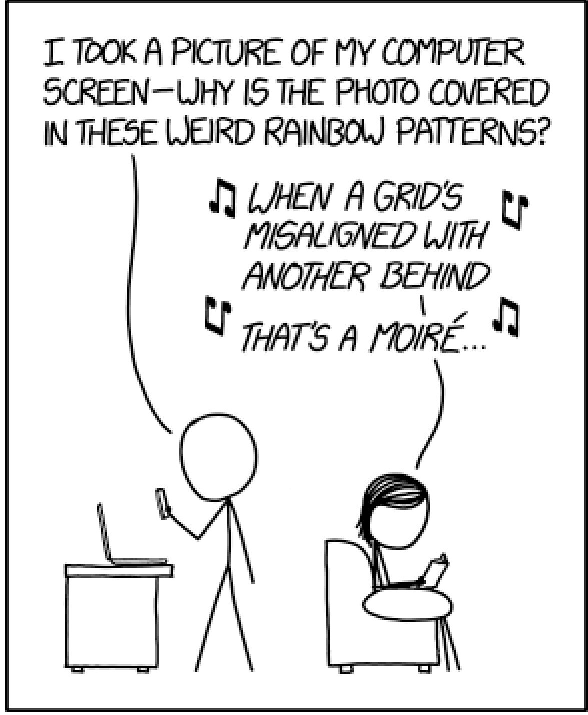
Point shape

- A bit **more difficult to discern** than position and color
  - **When color** is not available or **already taken**



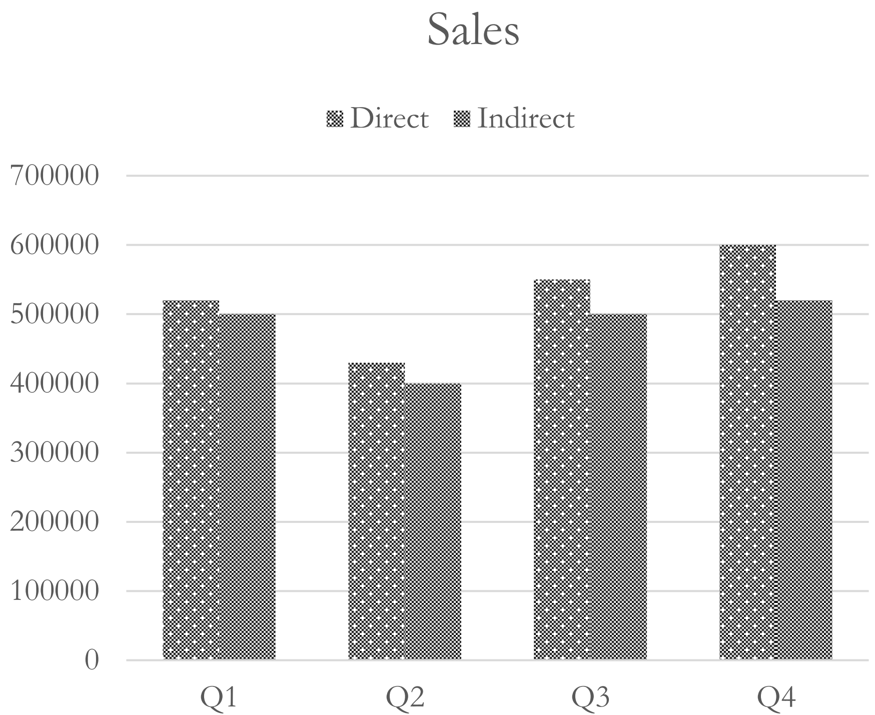
Fill pattern

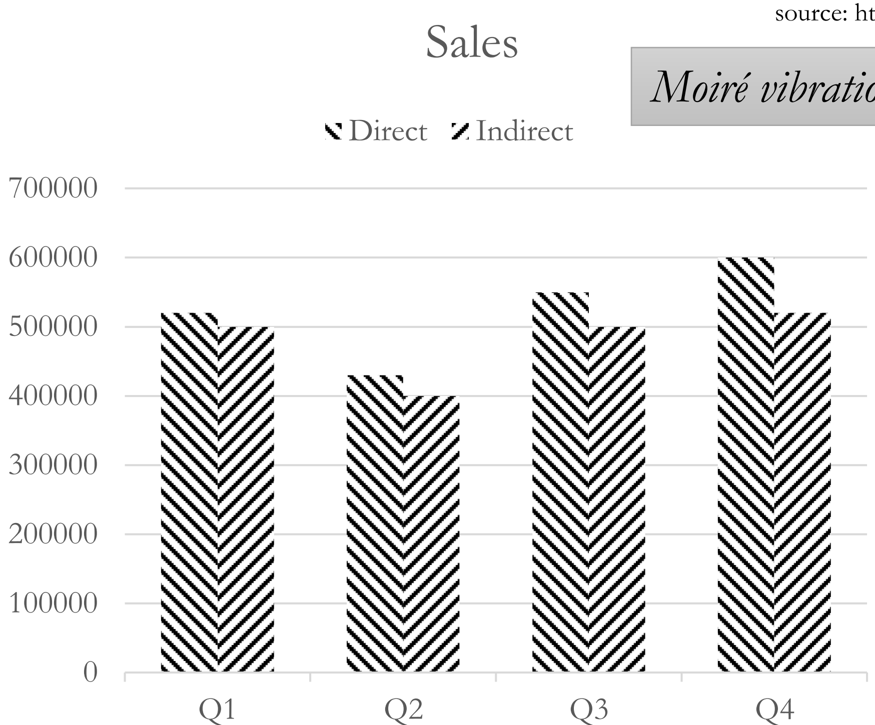
- Used to encode categorical items **when the quantitative values are encoded as bars** (or boxes)
- **Harder** to distinguish **than color**


I TOOK A PICTURE OF MY COMPUTER SCREEN—WHY IS THE PHOTO COVERED IN THESE WEIRD RAINBOW PATTERNS?
♫ WHEN A GRID'S MISALIGNED WITH ANOTHER BEHIND ♫ THAT'S A MOIRÉ... ♫

source: https://xkcd.com/1814/

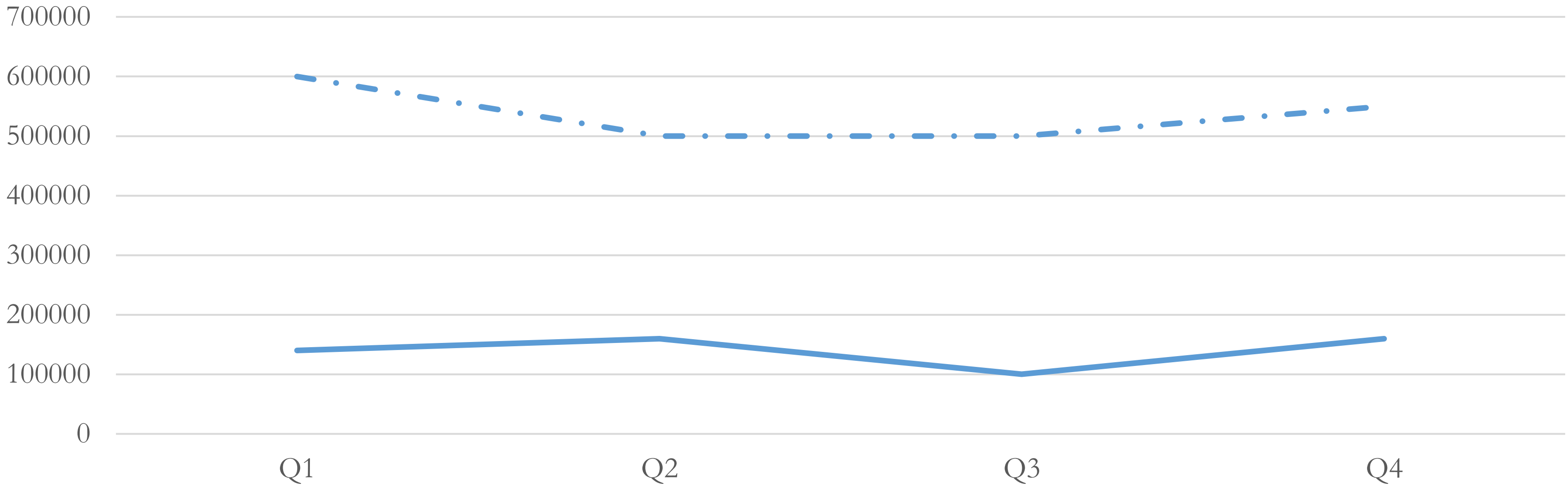
Moiré vibration/effect/pattern





Line style

- Lines bare a feeling of continuity which might be disrupted by breaks in the lines



Relationships in graphs

- Shaping **relationships of quantitative information**
- Different types of graphs are suitable for communicating different types of quantitative relationships

- Time series
- Ranking
- Part-to-whole
- Deviation
- Distribution
- Correlation
- Geospatial relation

Time series

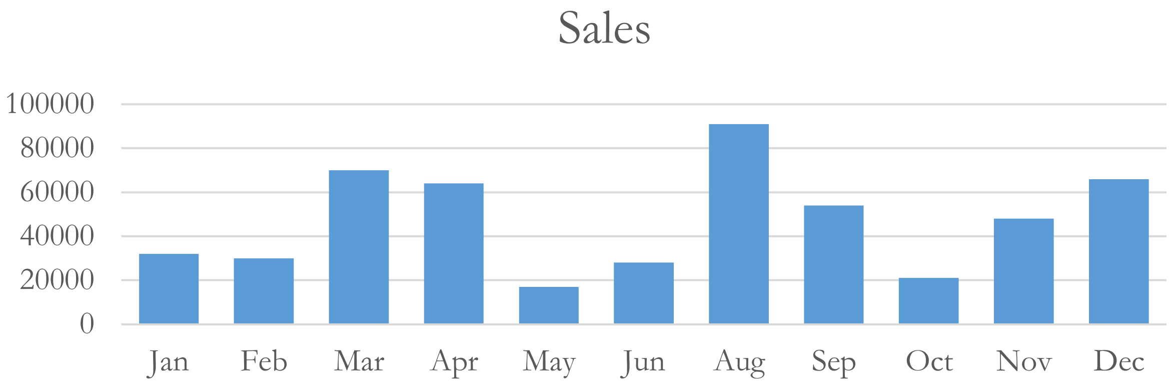
- **Series** of quantitative values featuring how an **attribute changes in time**

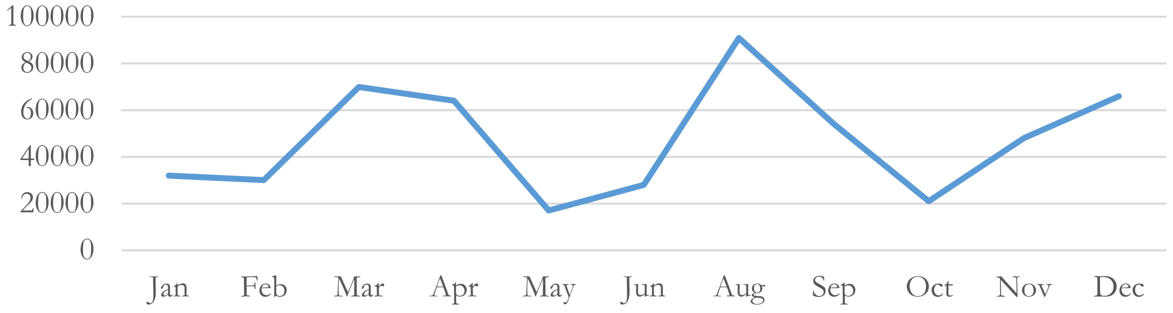
- Captures **patterns** and **trends**

- Quantitative messages involving time series usually include **words like**
  - *change, rise, increase, fluctuate, grow, decline, decrease, trend*

Time series design (1)

- Due to convention in most cultures, the layout of time should be **from left to right along the X axis →** vertical designs (bar, boxes) should be avoided in general

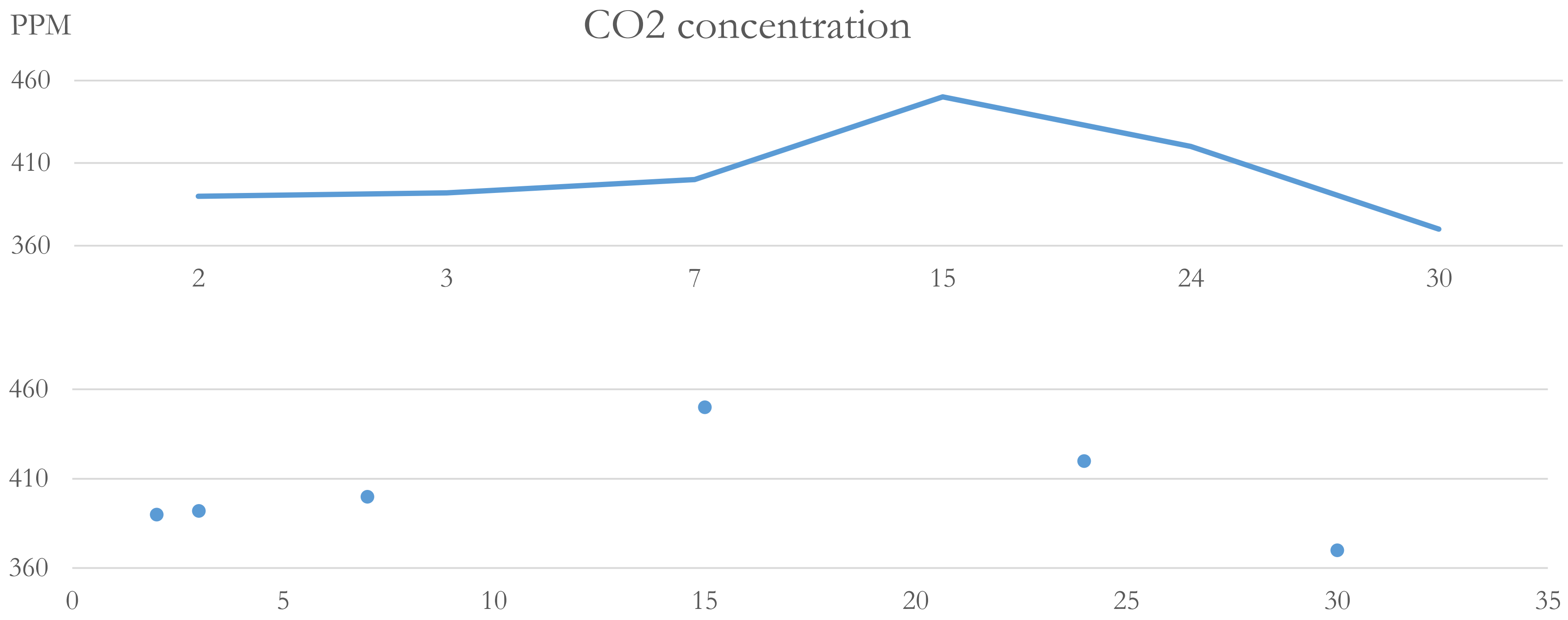




- **Bars** better when the goal is to emphasize **individual values**
- **Lines** more suitable for showing a **pattern of change** throughout the time

Time series design (2)

- **Points** suitable for display of values recorded at **irregular intervals**



Ranking

- Also called **item comparison**

- Display of how a set of quantitative values relate to each other sequentially

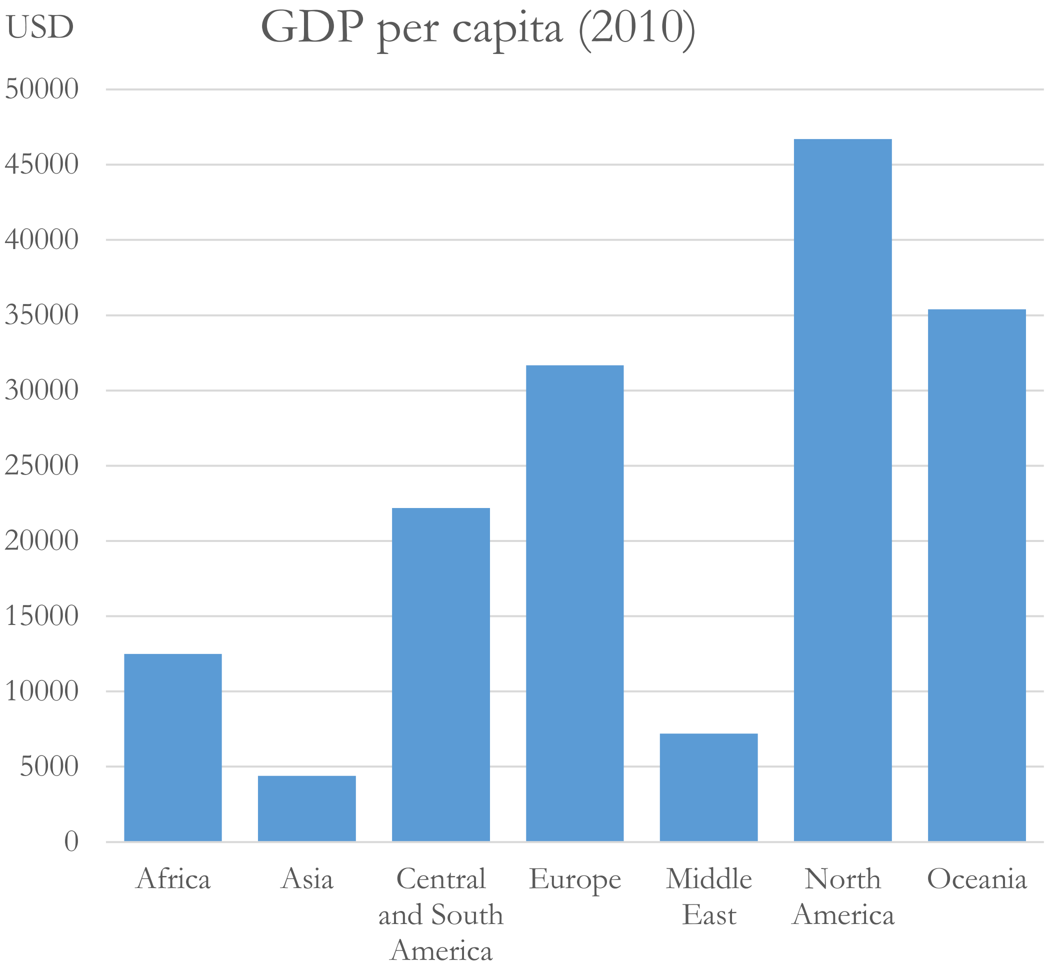
- **Sorted by size**

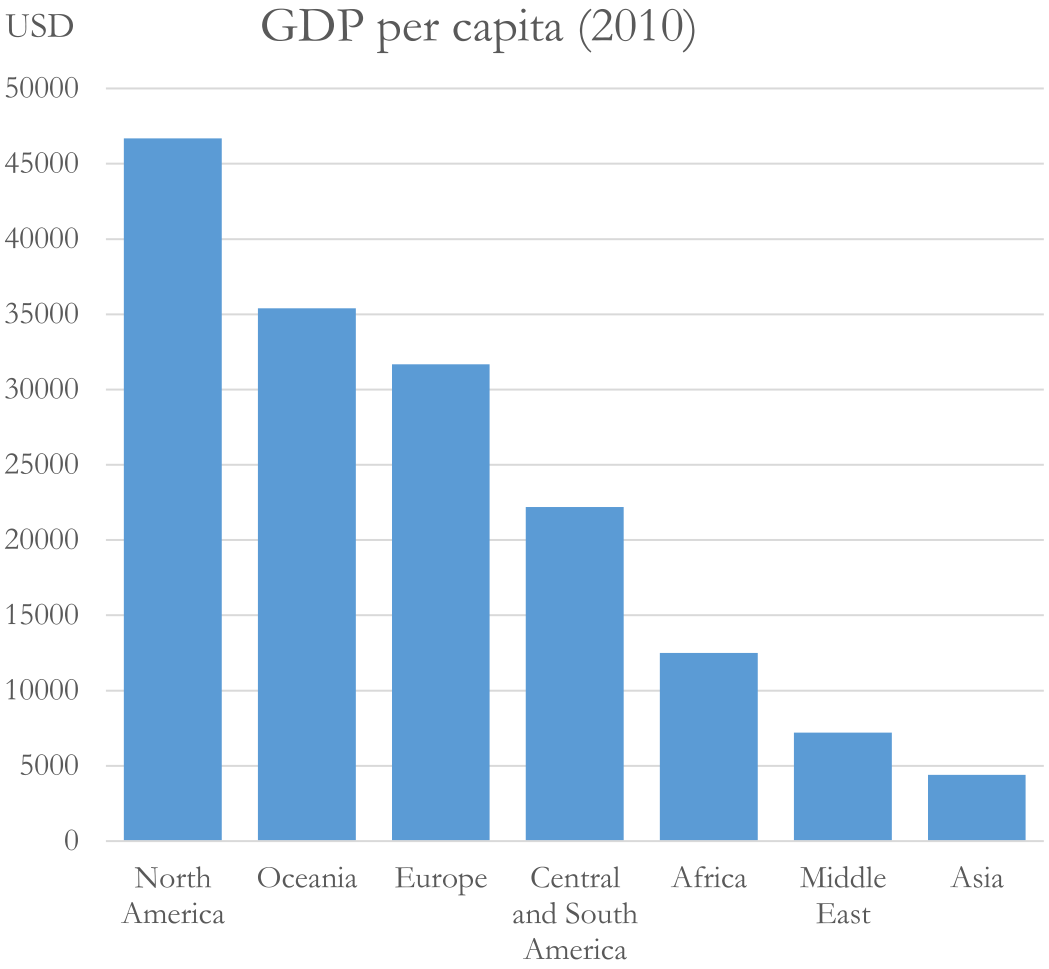
- Quantitative messages involving ranking usually include **words like**
  - *larger than, smaller than, equal to, greater than, less than*

Ranking design

- The goal is to **emphasize** each **individual item** → **bars**
- Both vertical and horizontal design is acceptable

Purpose	Sort order	Bar position
Emphasize the **highest** value	**Descending**	Vertical bars: highest bar on left Horizontal bars: highest value on top
Emphasize the **lowest** value	**Ascending**	Vertical bars: lowest bar on left Horizontal bars: lowest value on top





data source: http://www.tableausoftware.com/public/community/sample-data-sets

Part-to-whole

- Also called **component comparison**

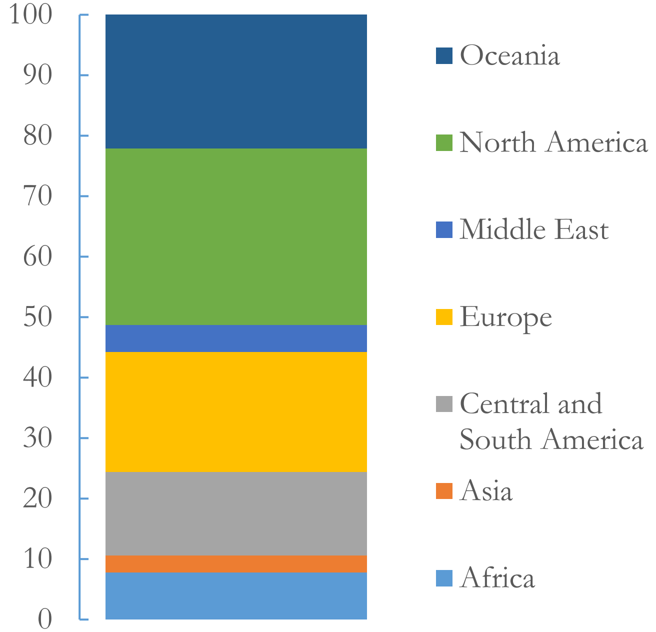
- Display of how **individual values** (parts, components) **make up a whole**

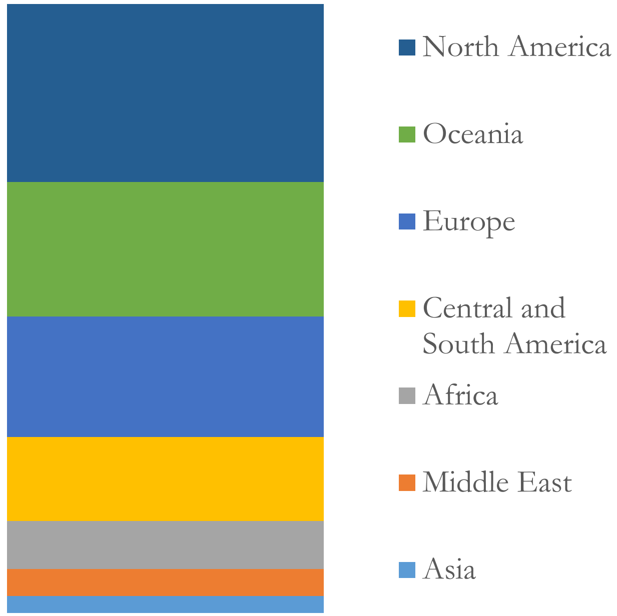
- Percentages (sum up to 100%), rates (sum up to 1)

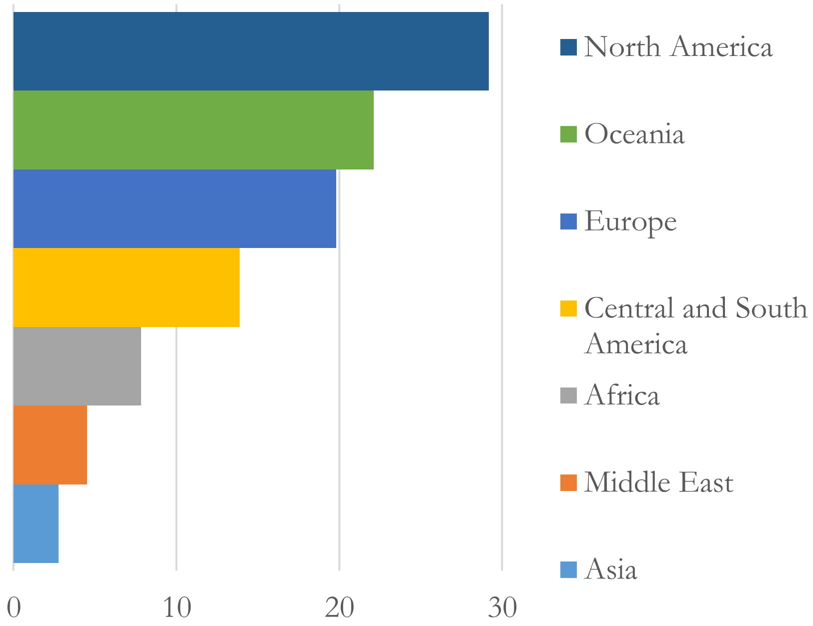
- Quantitative messages involving part-to-whole relationship usually include **words like**
    - *rate, percent, share, accounts for N percent*

Part-to-whole design

- Pie charts, although commonly used, are not very suitable (see slide 11)







Stacked bar graph

(% GDP per capita)

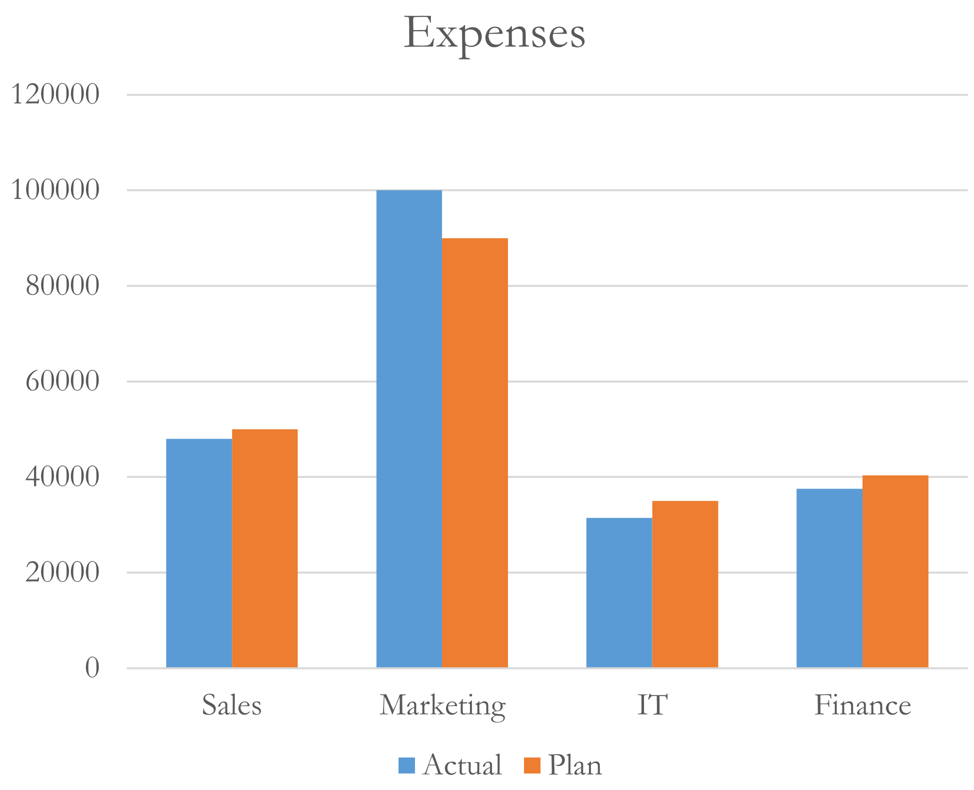
Deviation

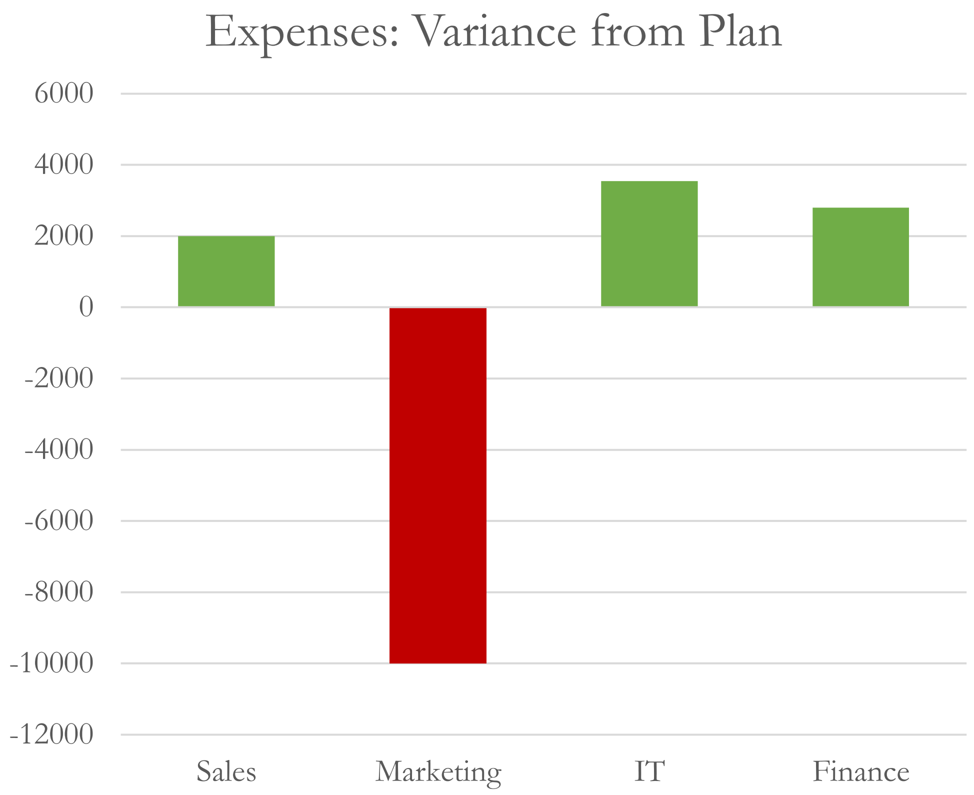
- Display of how one or more sets of quantitative values **differ from** a **reference set** (baseline)

- Usually expressed as positive or negative amount relative to the reference values or positive or negative rates or percentages relative to the reference value

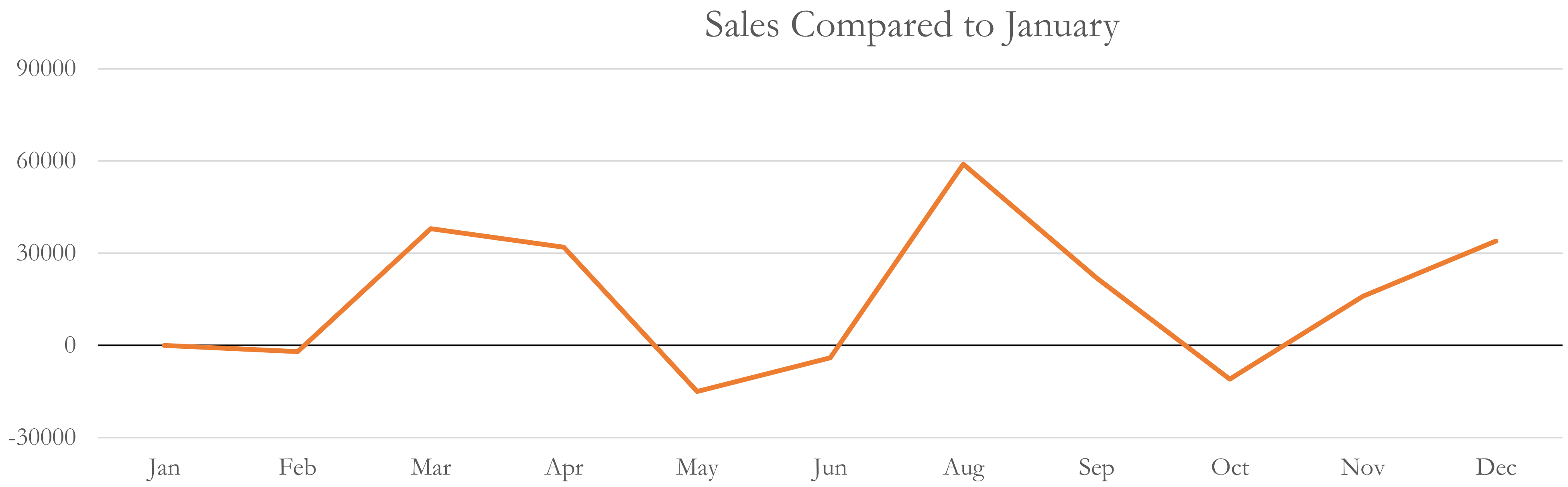
- Quantitative messages involving deviation usually include **words like**
  - *plus or minus, variance, difference, relative to*

Deviation design (1)





Deviation design (2)



Distribution

- Display of how quantitative values are **distributed across** an entire **range**

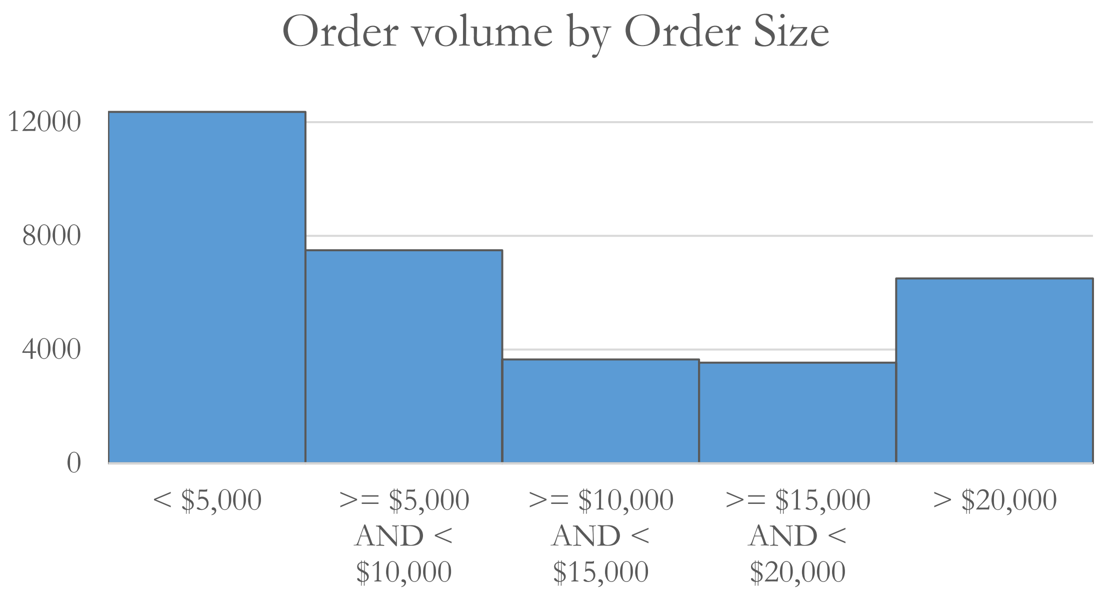
- Range commonly split into small ranges (intervals)

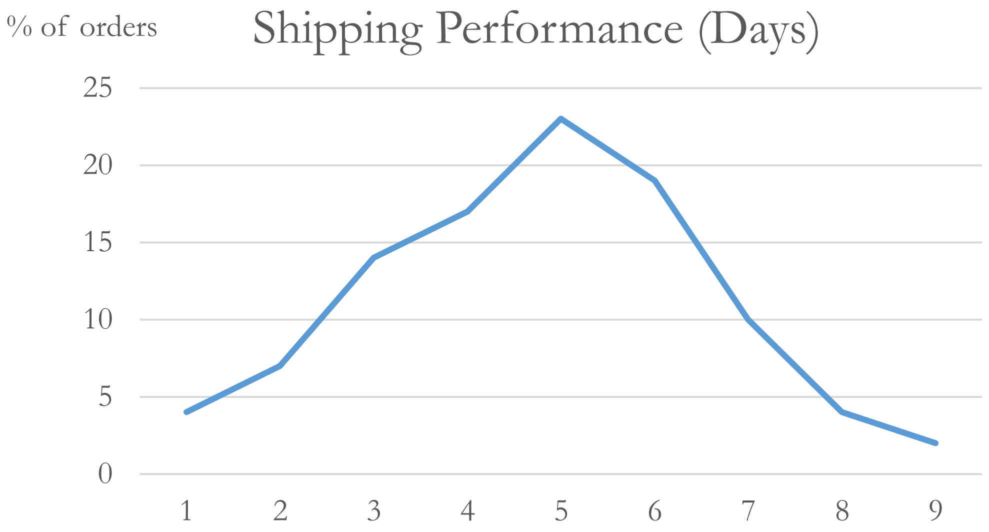
- A single visualization can cover multiple distributions

- Quantitative messages involving distribution usually include **words like**
    - *frequency, distribution, range, concentration*

Distribution design (1)

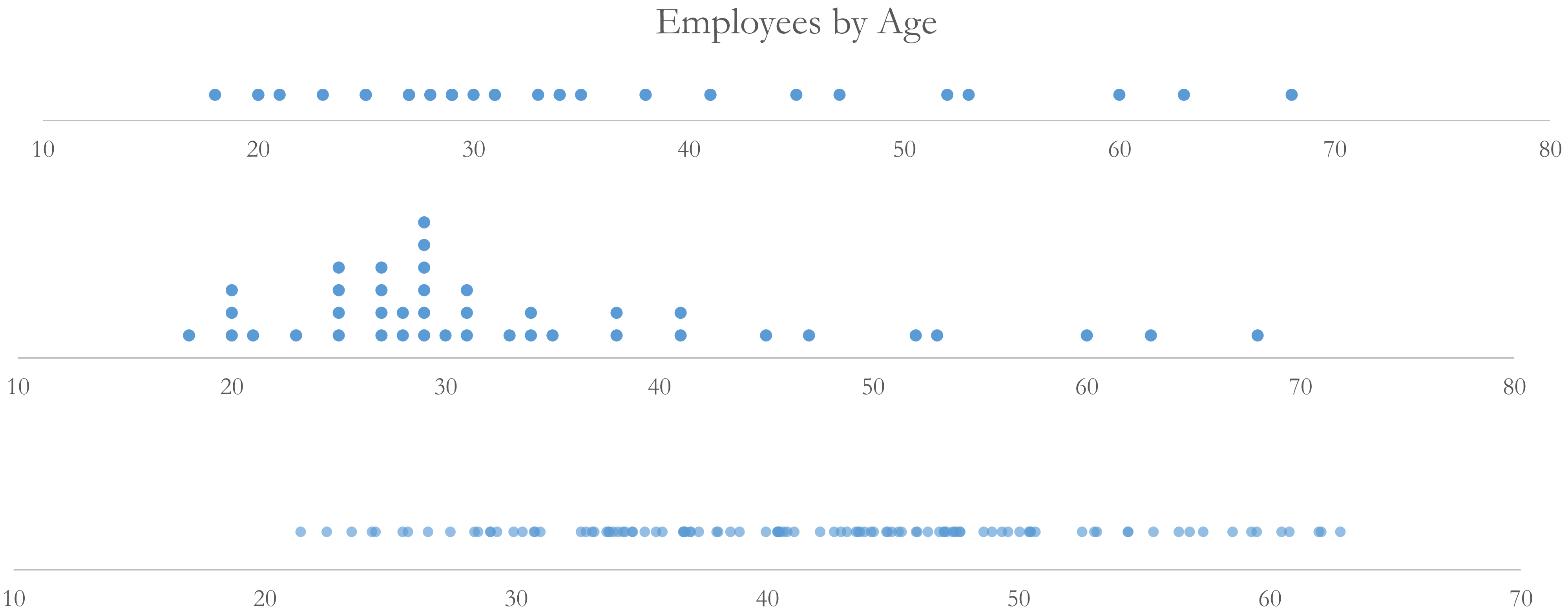
- Emphasis on
  - The **number of occurrences** in each interval → bars (**histogram**)
  - The overall **shape of the distribution** across the entire range → line (**frequency polygon**)





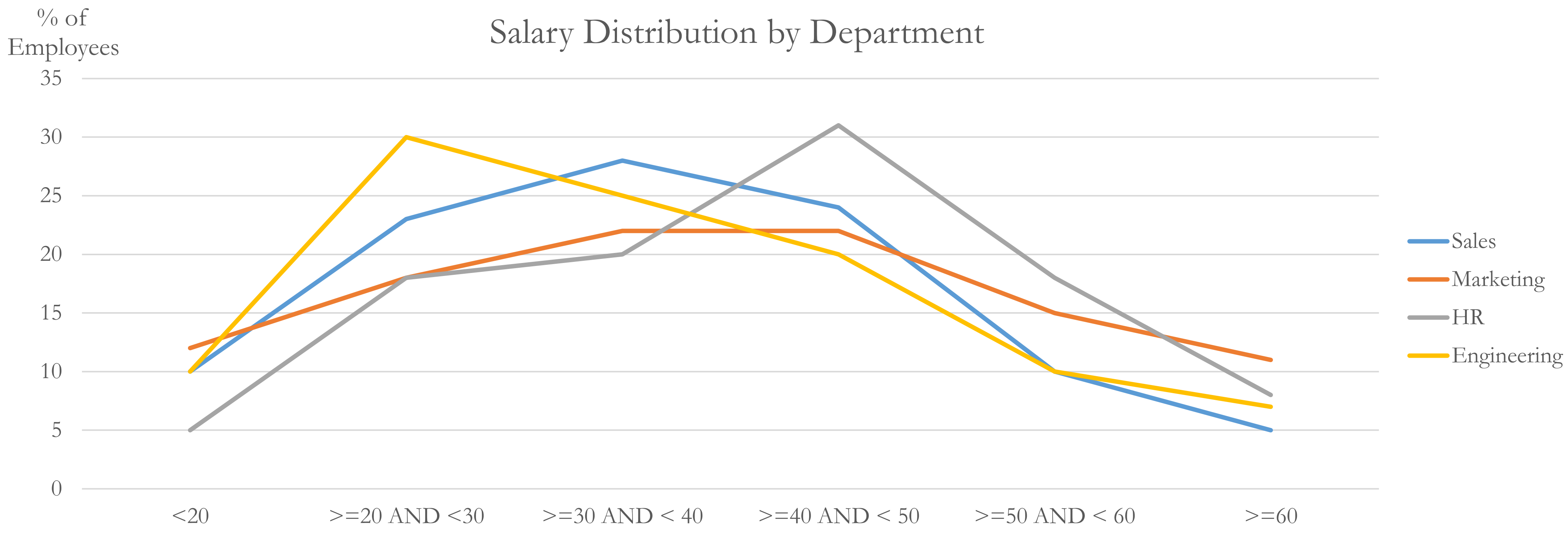
Distribution design (2)

- If we have a small number of values and want to see the individual items → **strip plot**



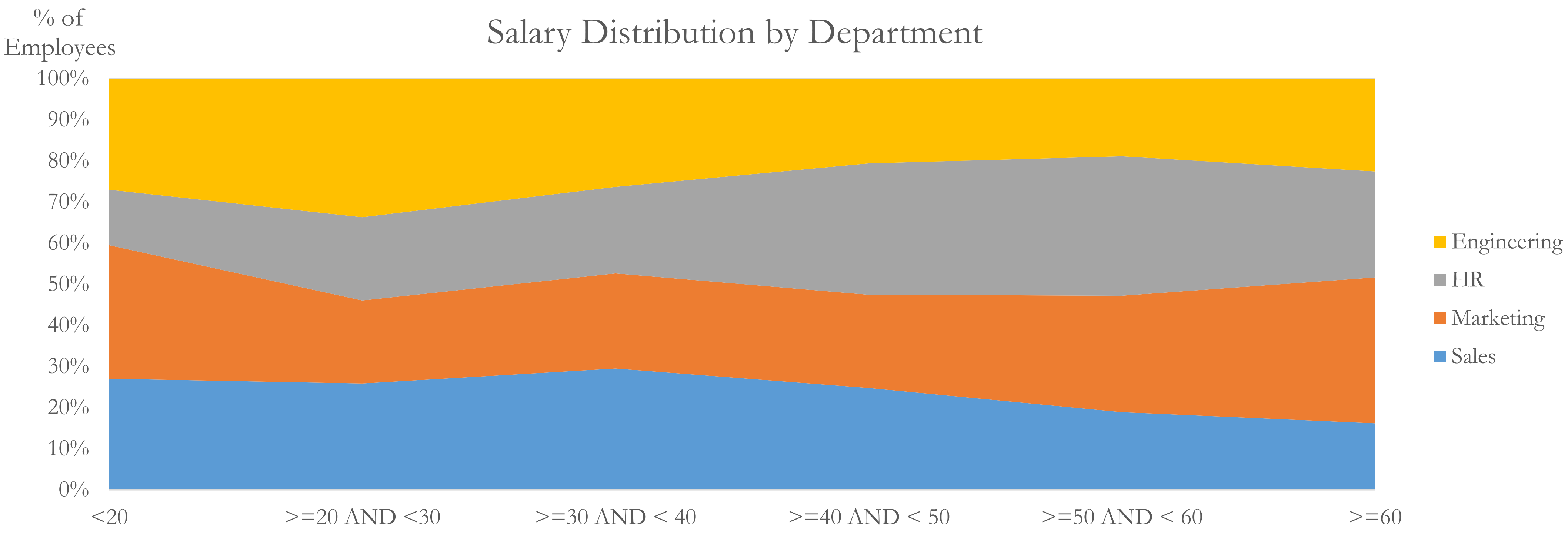
Distribution design (3)

- Frequency polygon can capture **multiple distributions**



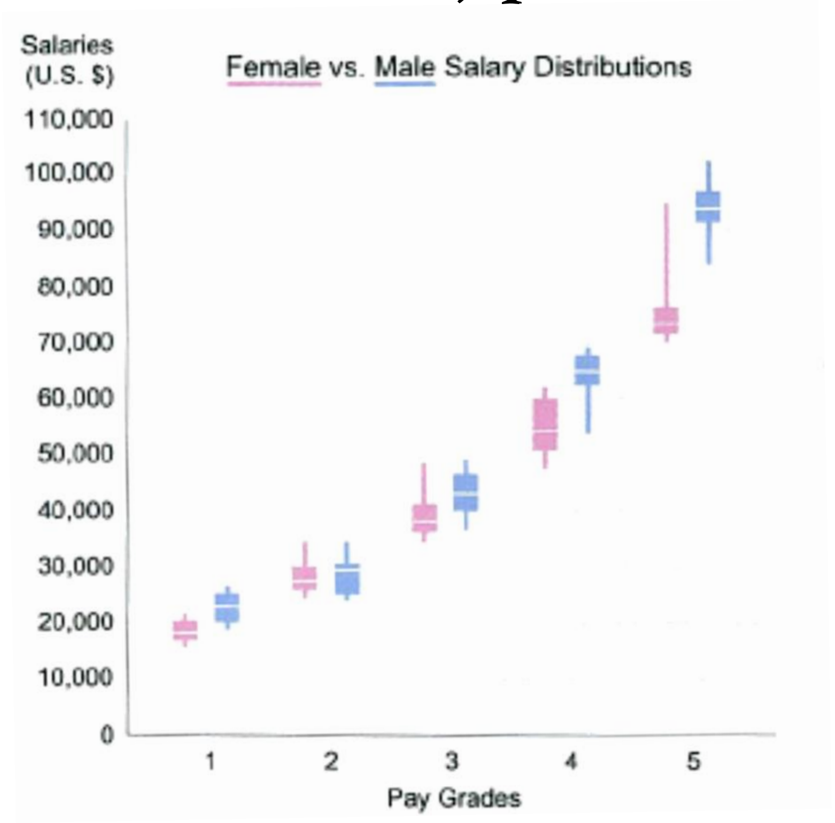
Distribution design (4)

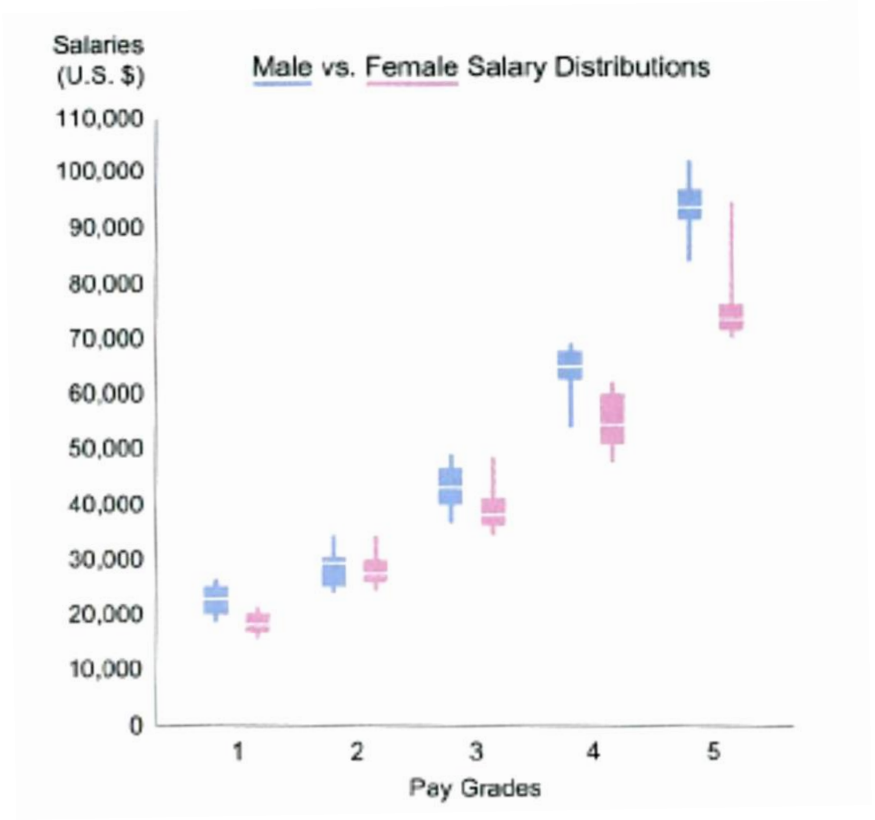
- Frequency plots do not work for **more** than a few **distributions** → **stacked density chart**



Distribution design (5)

- Frequency plots do not work for **more** than a few **distributions → box (box-and-whisker) plot → candlestick chart**

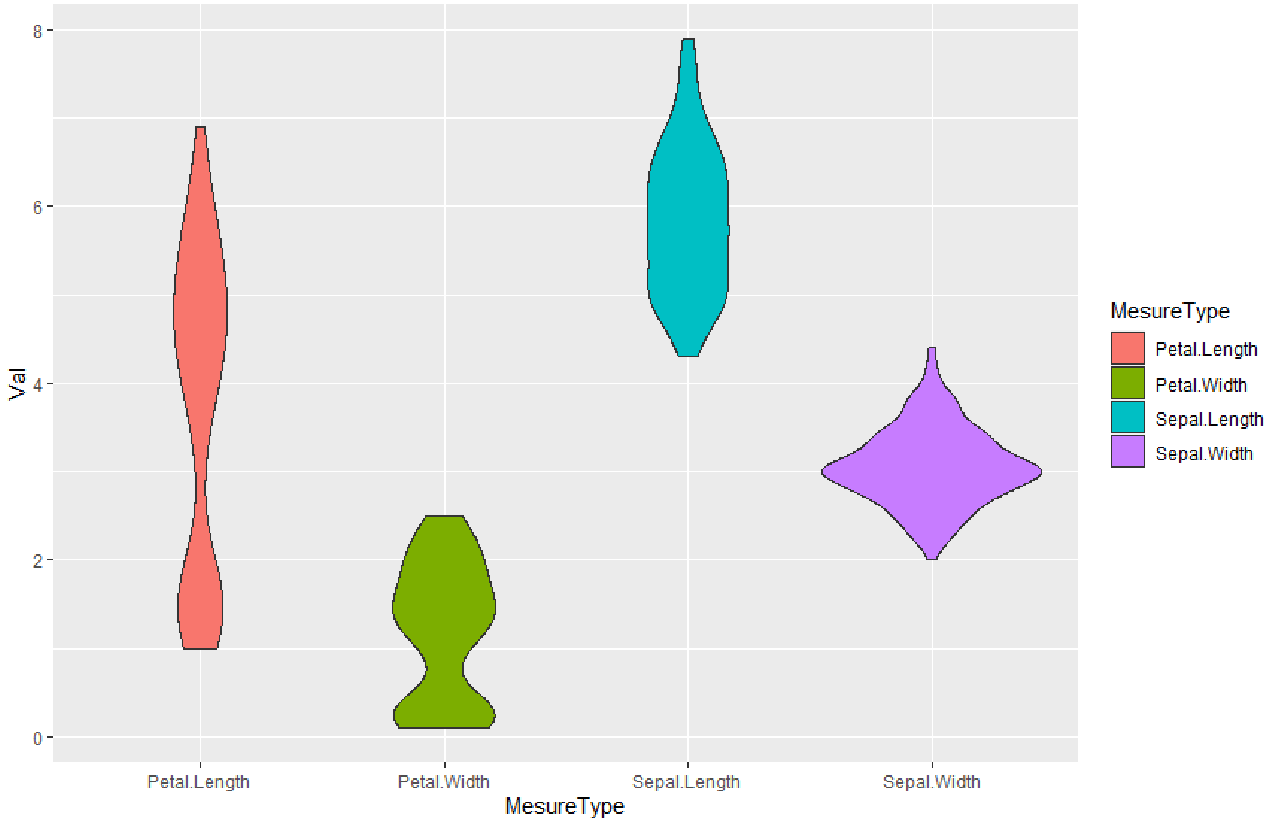




source: Stephen Few (2012) Show me the numbers – Designing Graphs and Tables to Enlighten

Distribution design (6)

- When more detail about distribution is required → **violin plot**



Correlation

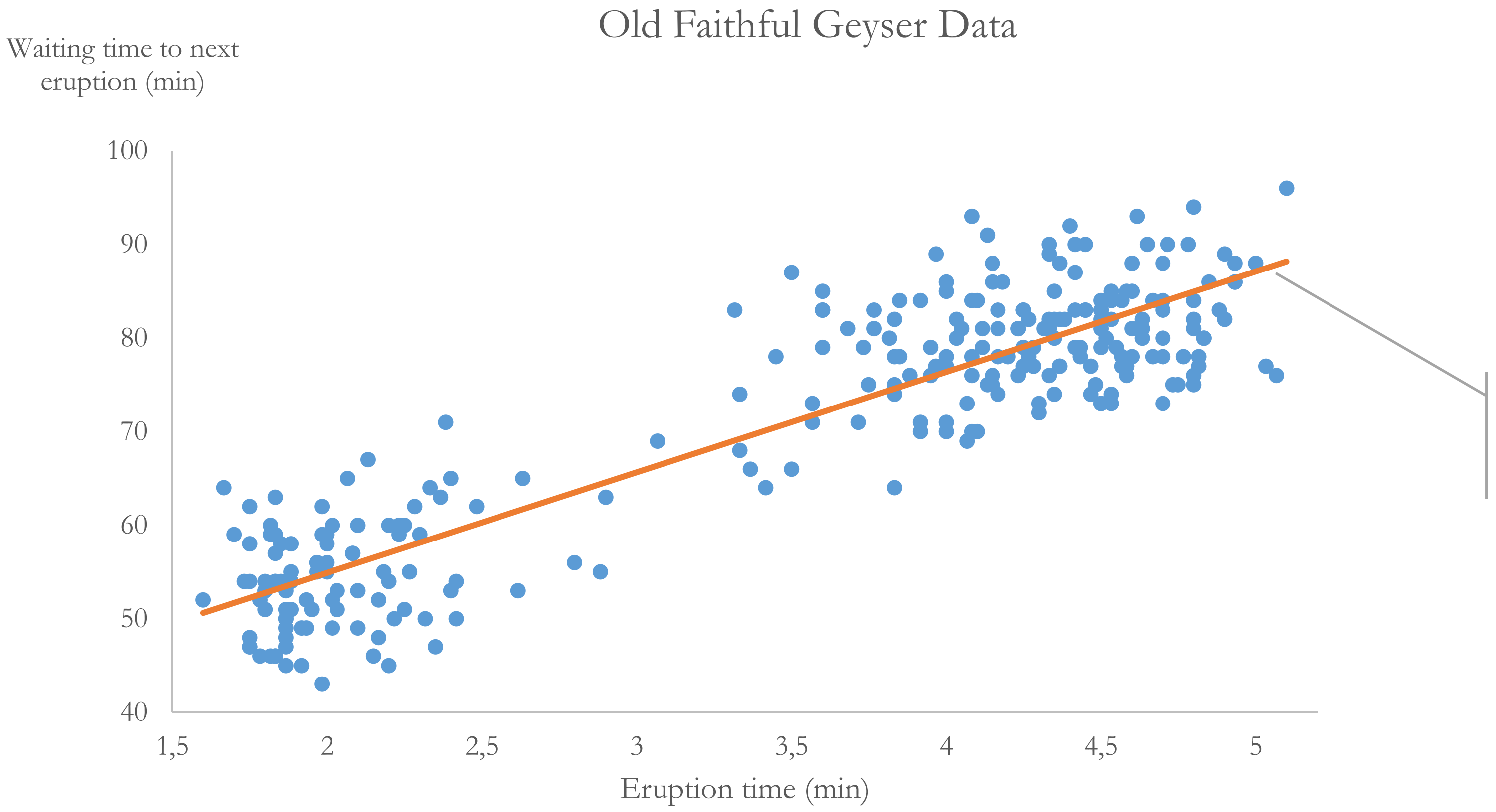
- Display of how (or whether) **two sets** of quantitative values **vary in relation** to each other (**covary**)
- Should show **direction** (positive, negative) and **degree** (low, high)
- Correlation does not imply causality ("*Correlation does not imply causation*")

http://www.tylervigen.com/spurious-correlations

- Quantitative messages involving correlation usually include **words like**
  - *increases with, decreases with, changes with, varies with, caused by, affected by, follows*

Correlation design

- Relationship between two quantitative values → **scatter plot**

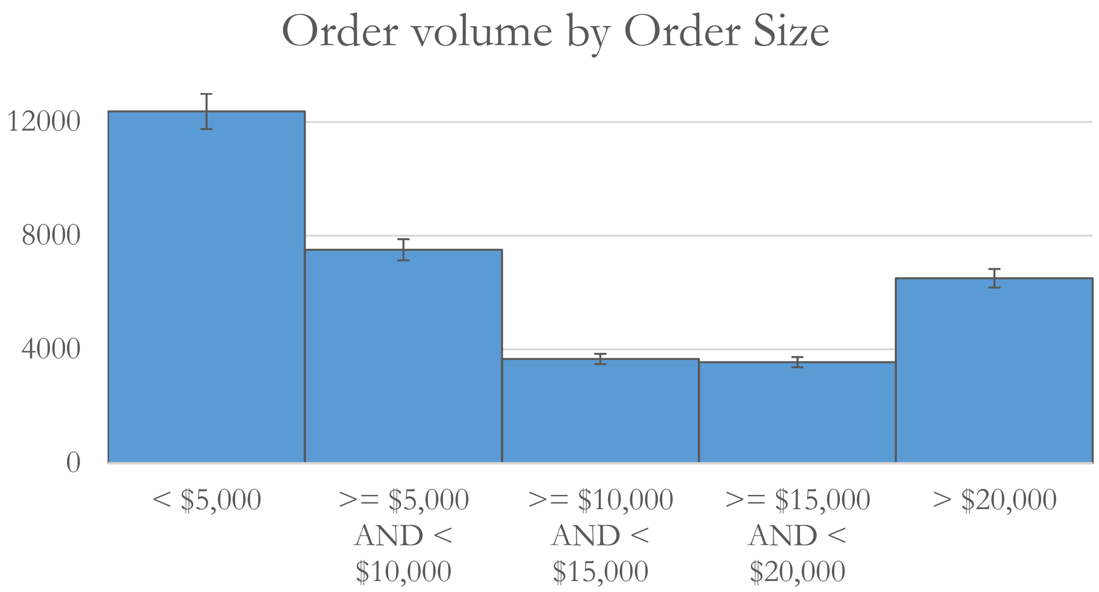


Old Faithful Geyser Data

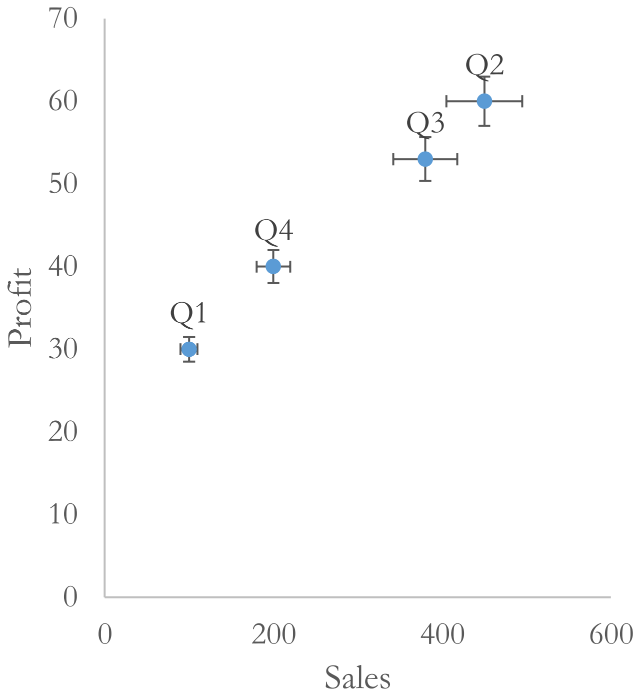
Trend line

Uncertainty (1)

- Values of estimates or measures with uncertainty can be visualized with this estimate



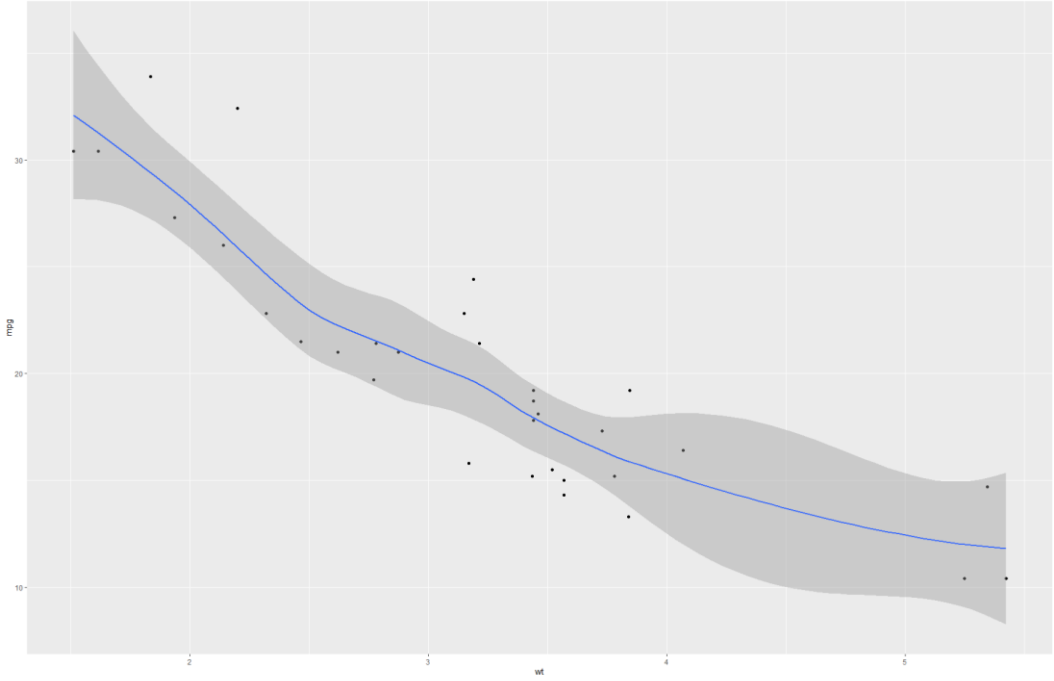
Error bars

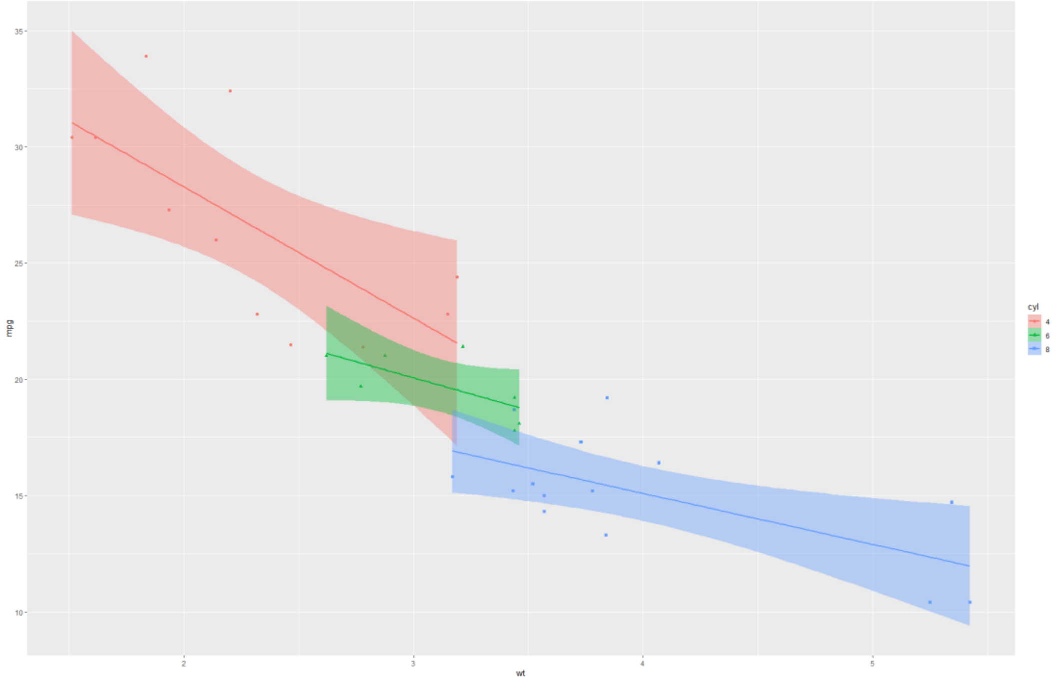


2D error bars

Uncertainty (2)

- Equivalent of an error bar **for line graphs**





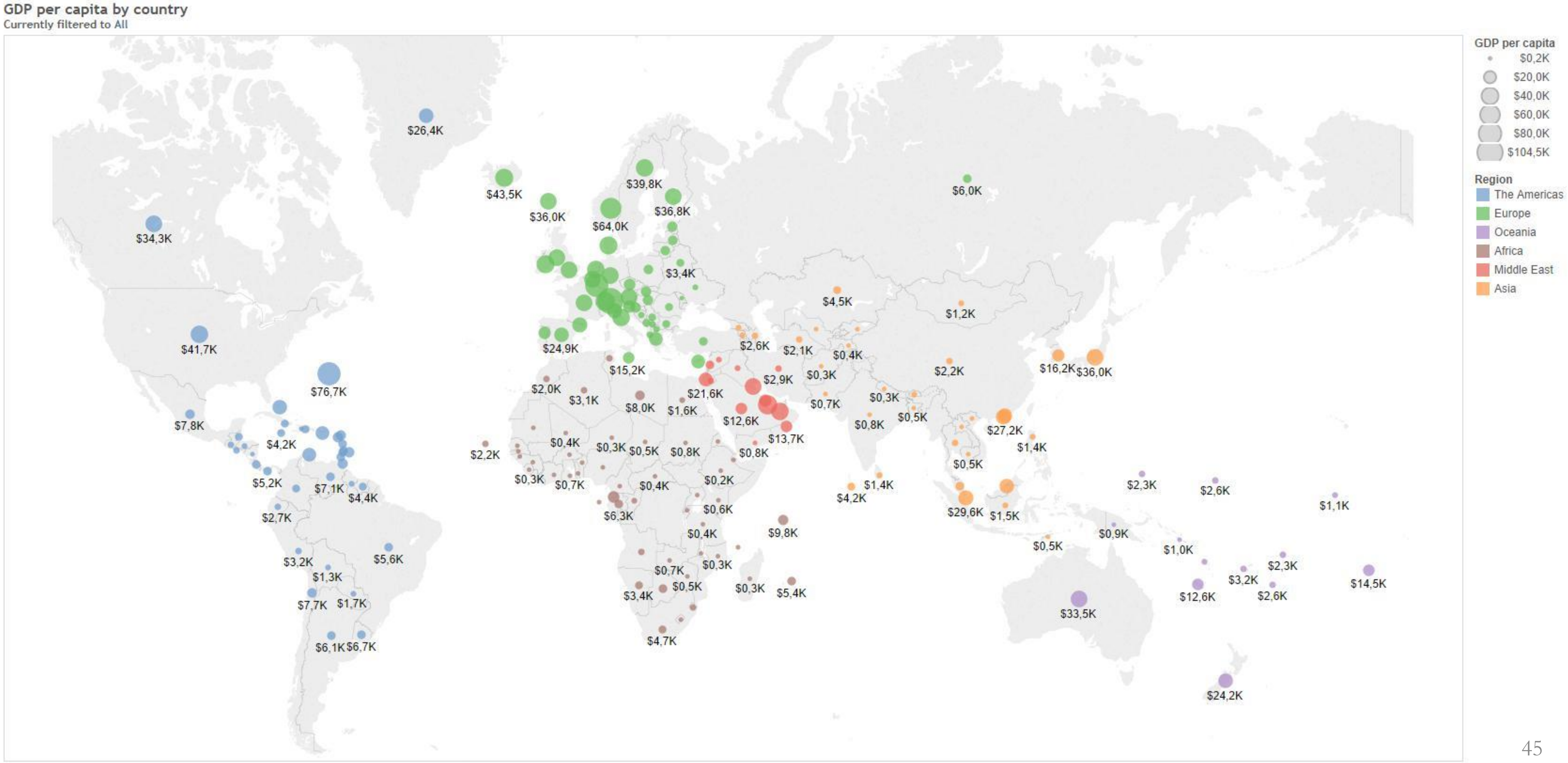
Geospatial relationship

- Display where quantitative **values** are **located** (**spatial relation**)

- The spatial location is commonly geographic, but does not have to be (e.g. buildings plans)

- Quantitative messages involving geospatial relation include **words like**
  - *geography, location, where, region, territory, country, state, city*

Geospatial design



GDP per capita by country

Currently filtered to All

Map based on Longitude (generated) and Latitude (generated). Color shows details about Region. Size shows average of GDP per capita (curr $). Details are shown for Subregion and Country / Region. The data is filtered on Action (YEAR(Date (year))), which keeps 11 members. The view is filtered on average of GDP per capita (curr $) and Region. The average of GDP per capita (curr $) filter ranges from $0,2K to $104,5K. The Region filter keeps 6 of 6 members.

Principles of graph design

- **Highlight data and suppress everything else**
  - *"Above all else show the data"* (Tufte, 1983)

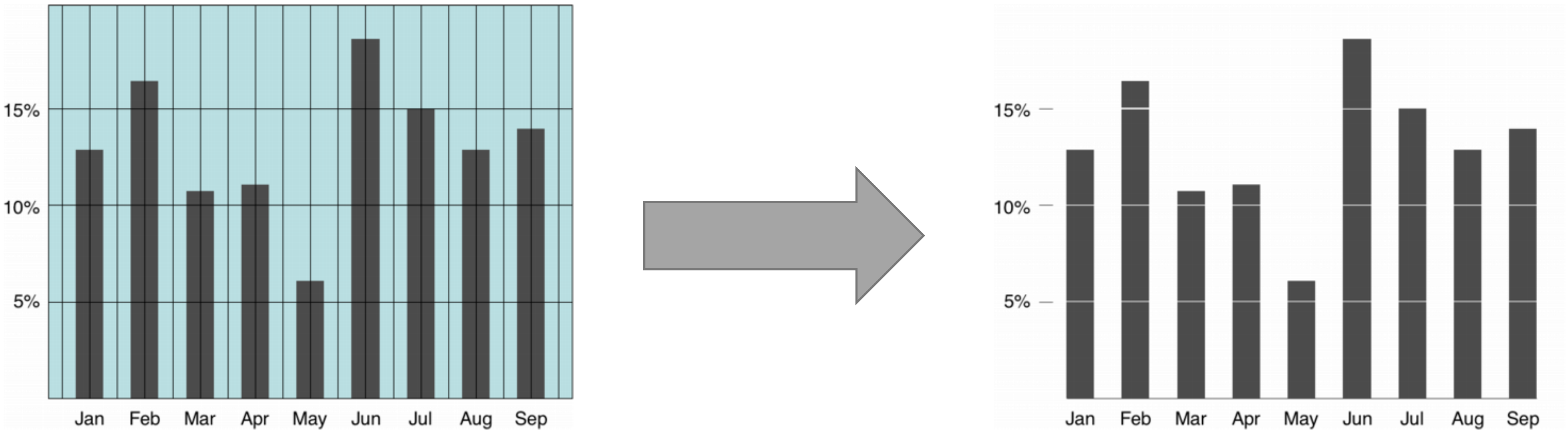
- **Maintain visual correspondence with numerical quantities**
  - Quantity is best expressed as length (bars, boxes) or 2D position (points, lines)
  - Distance in the axis scale (distance between tick marks) should always correspond with the difference of the corresponding quantitative values

- **Avoid 3D**
  - Adding third dimension without adding a third scale → makes the graph more difficult to read
  - Adding third dimension with adding a third scale → some values probably won't be visible at all and all will be difficult to compare

Data-ink ratio

- *"Above all else show the data"* (Tufte, 1983)

$$\text{DataInkRatio} = \frac{\text{data ink}}{\text{total ink used to print the graphics}}$$

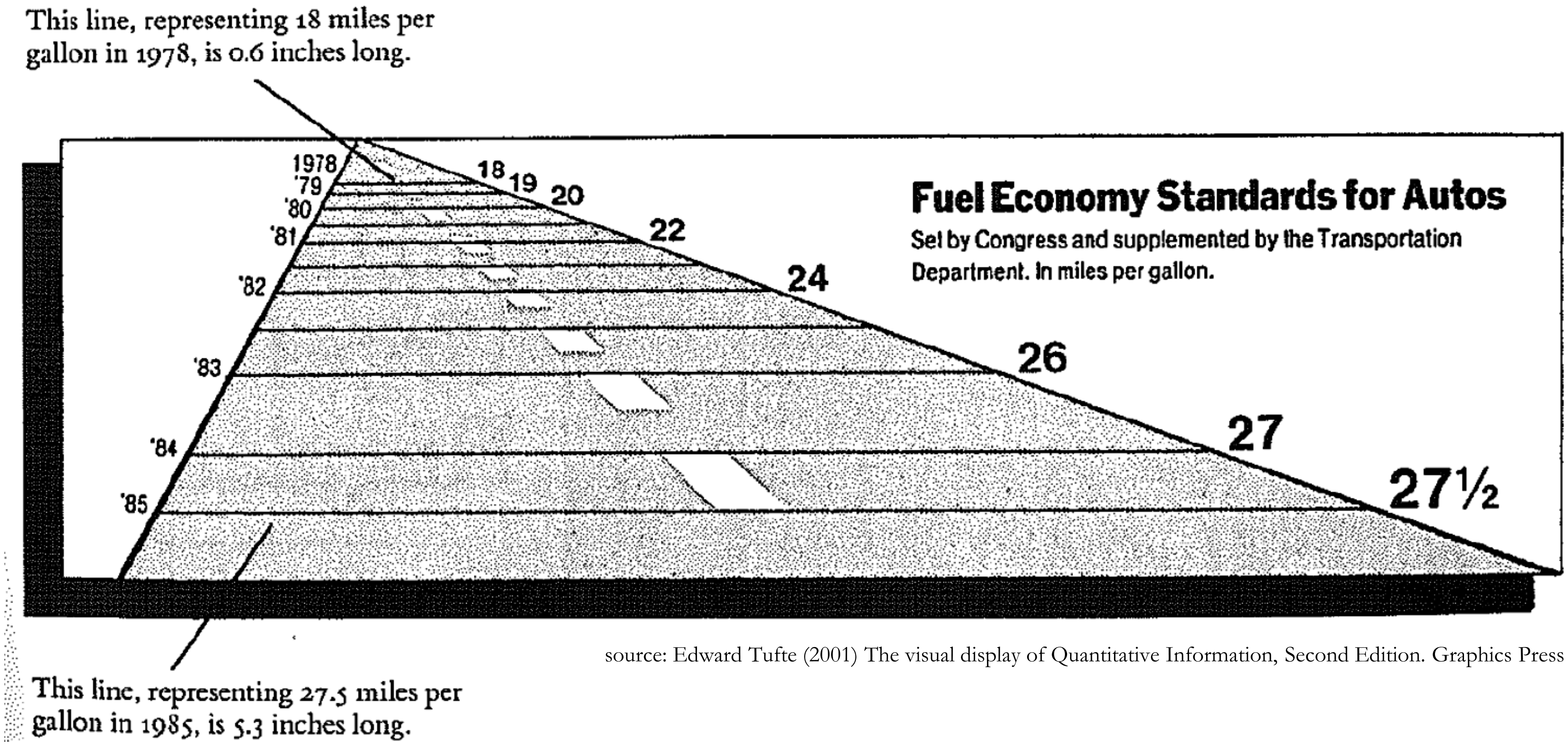


Misleading (lying) with graphs

- The visual **image** (perceived visual effect) should **represent** the underlying **numbers** → how to **measure** such thing?
  - Conduct an **experiment** on visual perception of graphics
    - E.g., approximate laws in perceiving have been discovered (perceived area of a circle = (actual area)^x, x=0.8 ± 0.3)
    - The perception is context dependent
  - Define a **measure of "misperception"** → **Lie Factor**

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

- $LF > 1.05$ or $LF < .95$ suggests substantial distortion



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

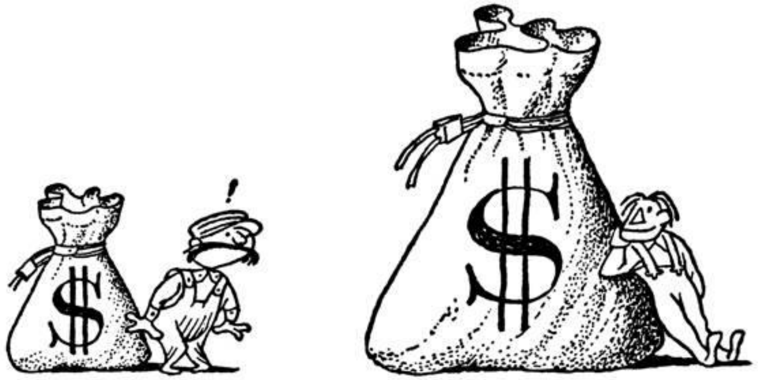
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

source: Edward Tufte (2001) The visual display of Quantitative Information, Second Edition. Graphics Press

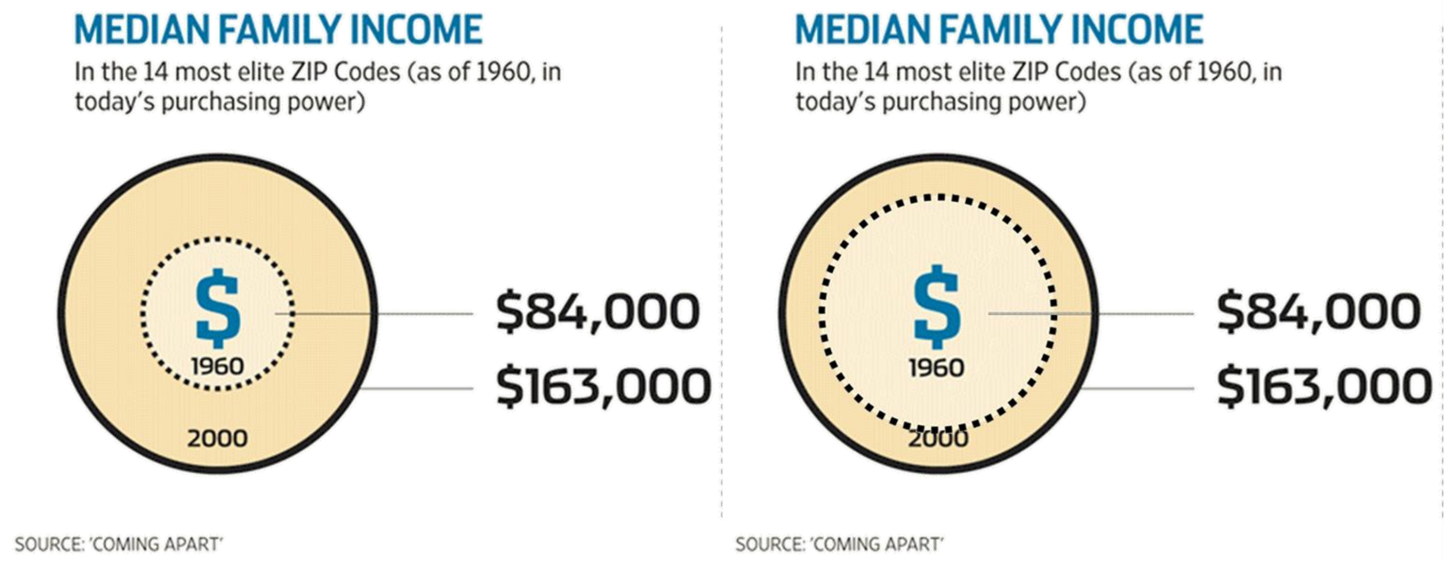
$$\left.\begin{aligned}\text{effect in data} &= \frac{27.5 - 18.0}{18.0} \times 100 = 53\% \\ \text{effect in graphics} &= \frac{5.3 - 0.6}{0.6} \times 100 = 783\%\end{aligned}\right\} \quad \text{Lie Factor} = \frac{783}{53} = 14.8$$

Beware of the effect of size

- If the visualization uses **area** (or even volume) then the **area** (not length) **should reflect the change** in the quantitative value



source: Darrel Huff (1954) How to lie with statistics, W.W. Norton & Company Inc



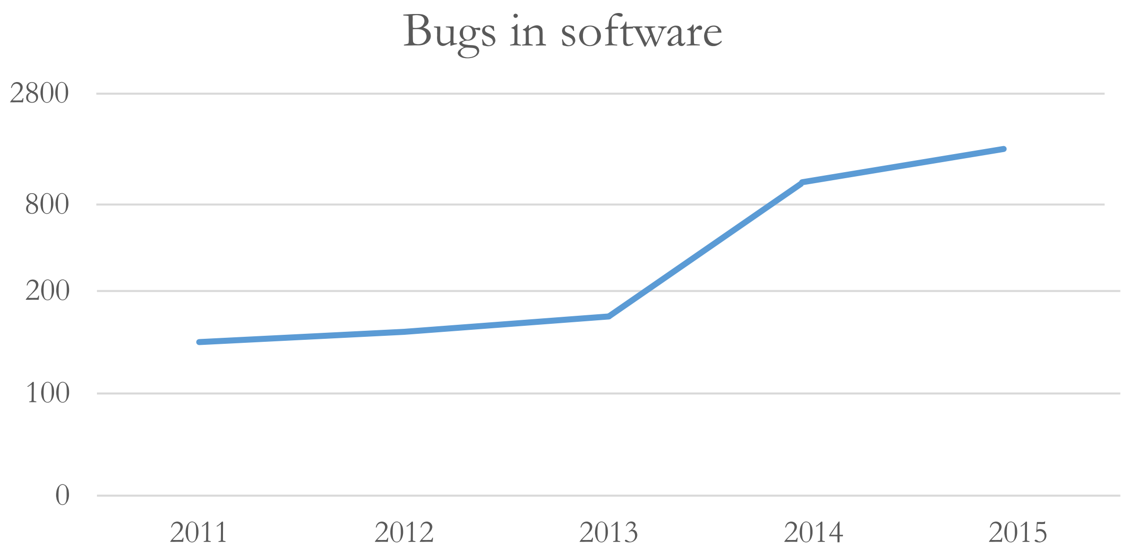
Inaccurate graph as it appeared in the *Wall Street Journal* (1/21/2012)

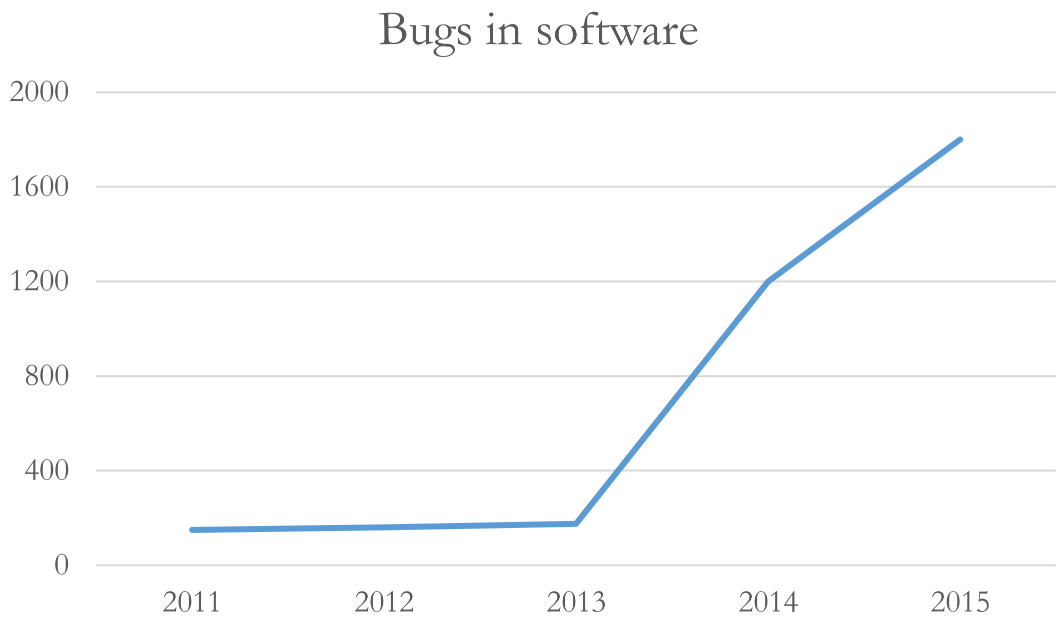
Accurate graph constructed by EvalBlog.com

source: http://evalblog.com/tag/how-to-lie-with-statistics/

Y-axis manipulation (1)

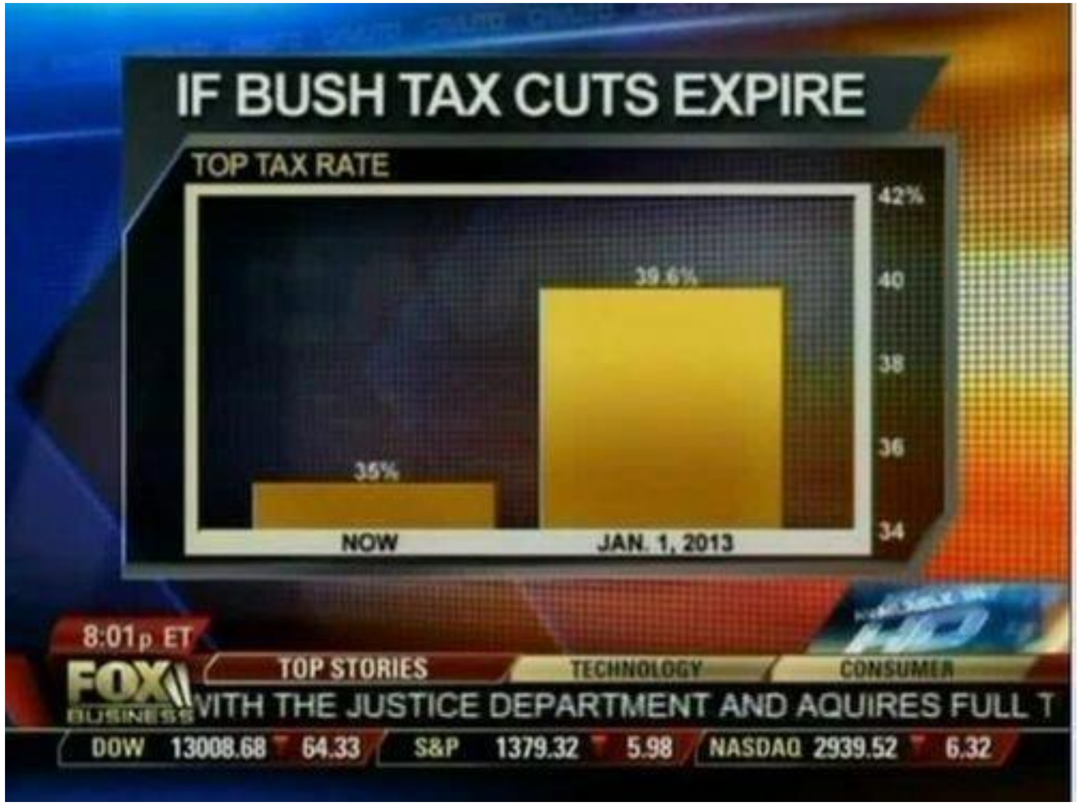
- The distance between tick marks on the scale line should be consistent with the difference in the quantitative values

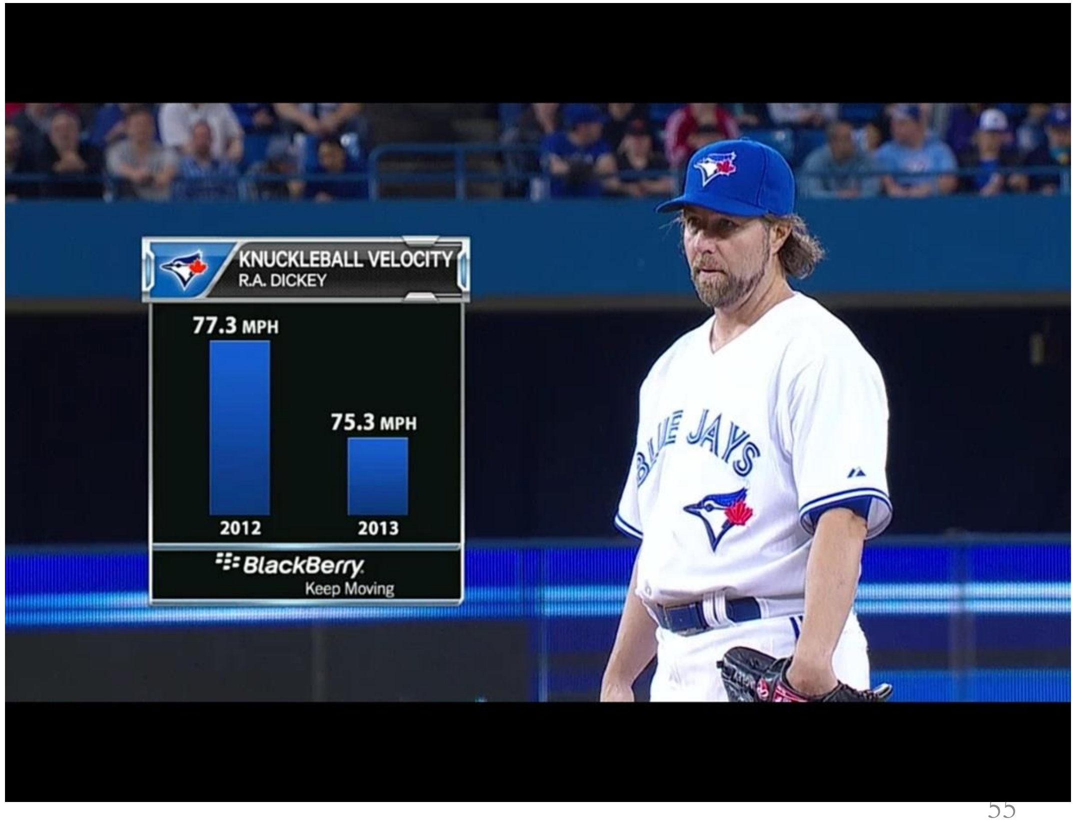




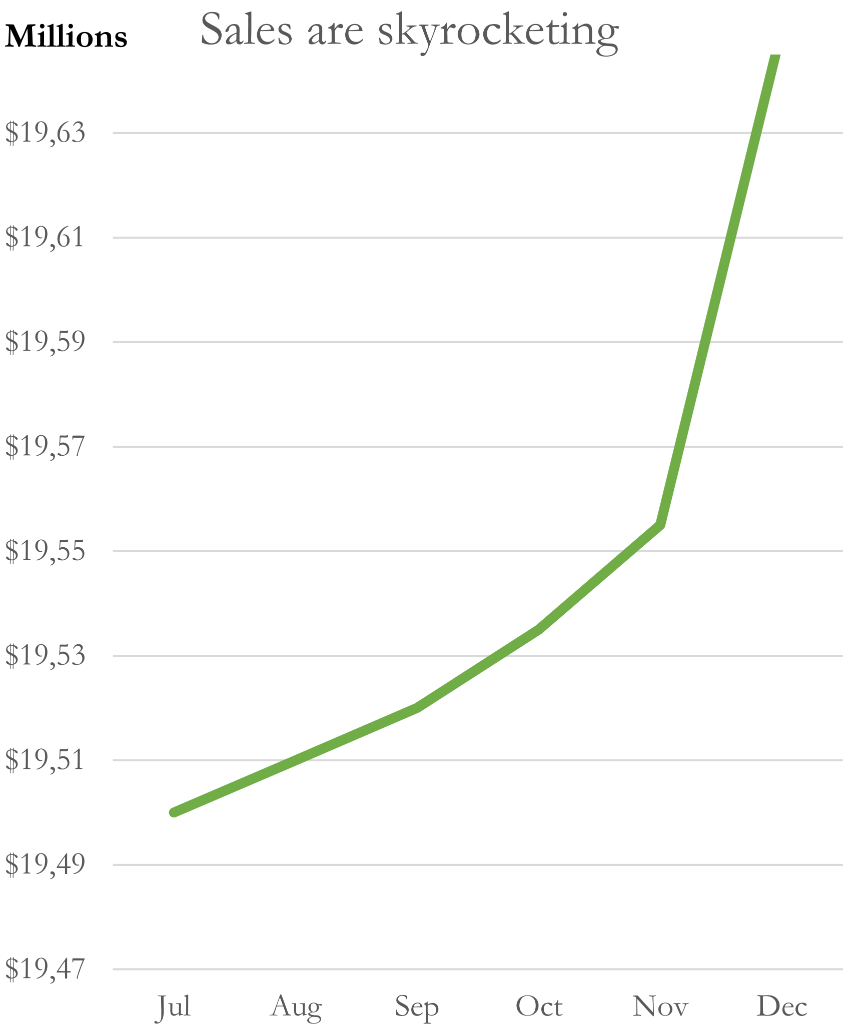
Y-axis manipulation (2)

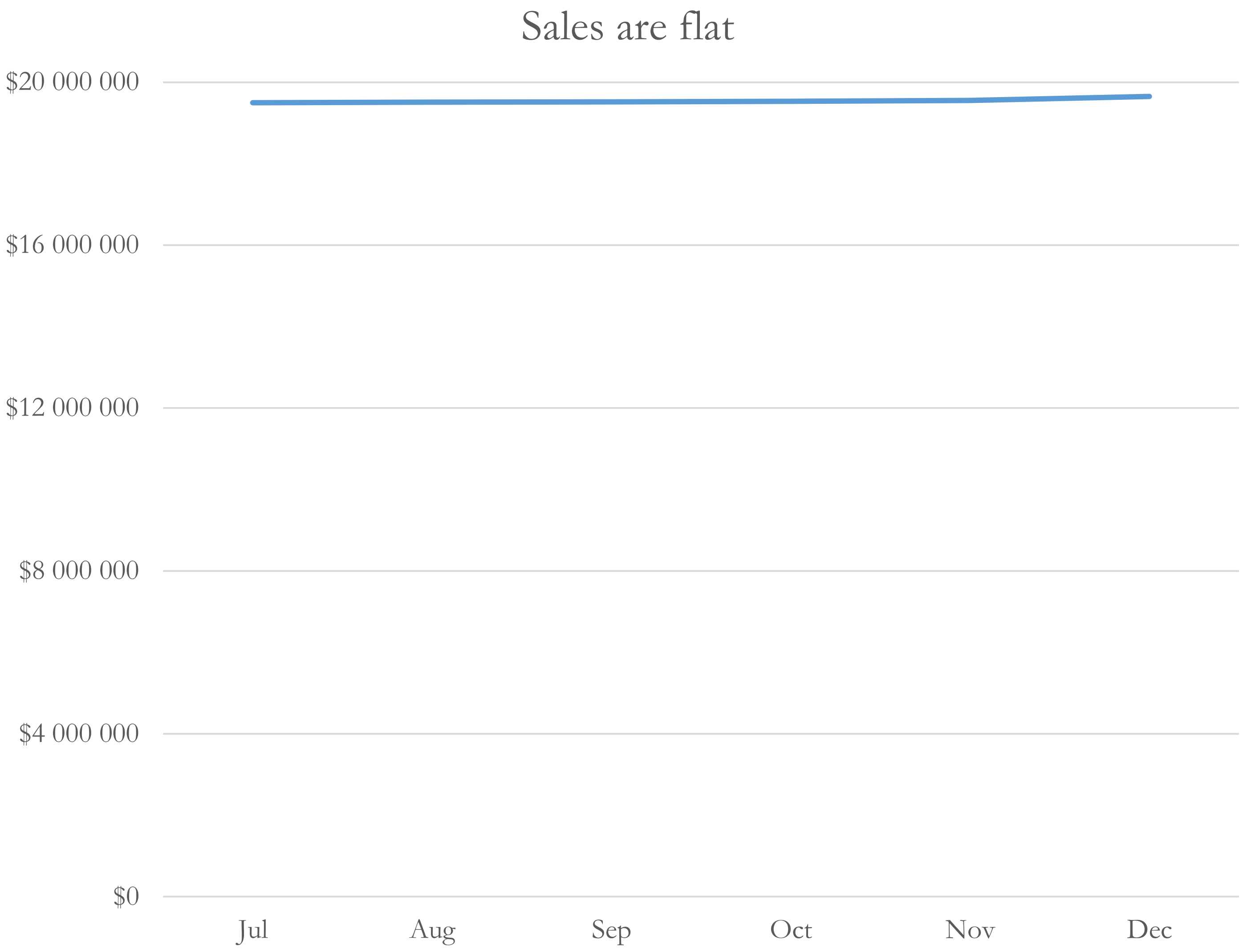
- You should **never eliminate zero** from the scale with **bars**

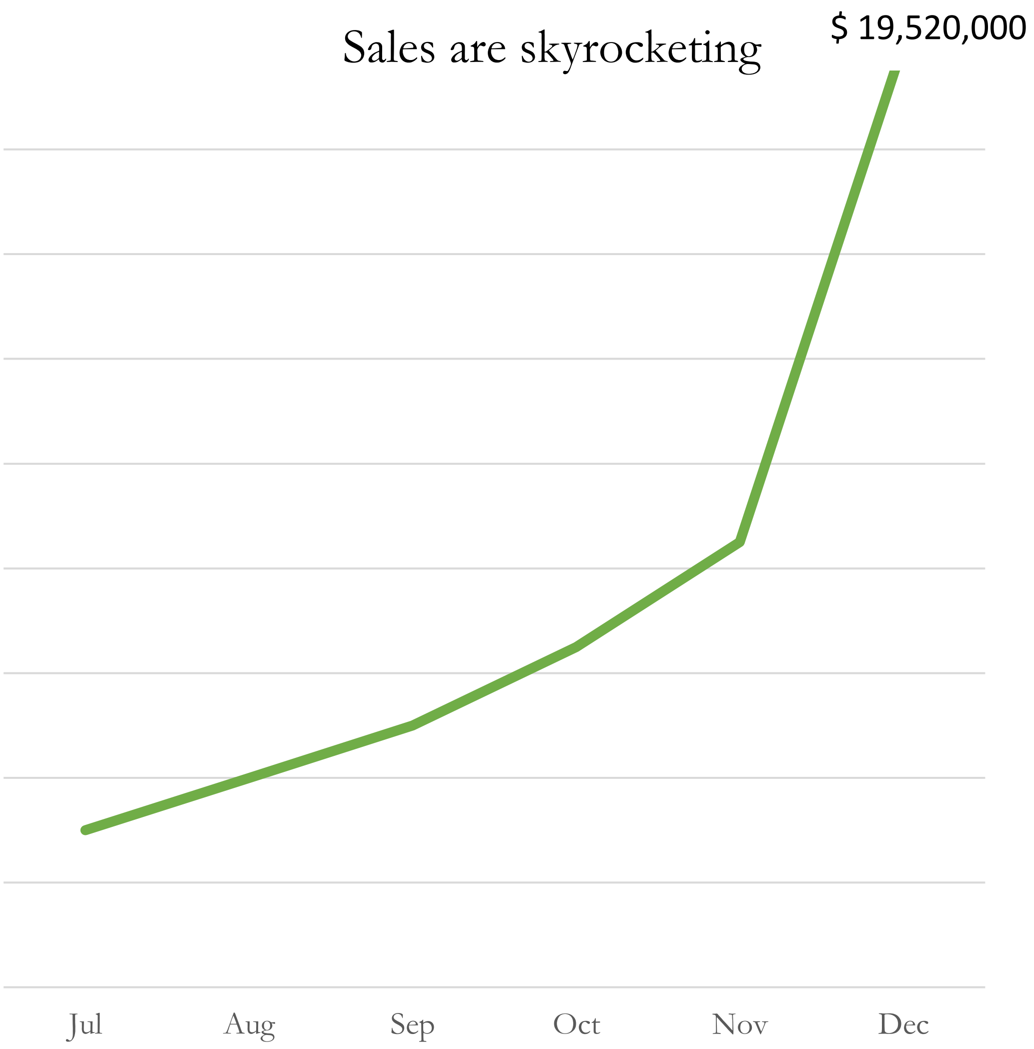




source: http://data.heapanalytics.com/how-to-lie-with-data-visualization


Millions
Sales are skyrocketing
$19,63
$19,61
$19,59
$19,57
$19,55
$19,53
$19,51
$19,49
$19,47
Jul
Aug
Sep
Oct
Nov
Dec


Sales are flat
$20 000 000
$16 000 000
$12 000 000
$8 000 000
$4 000 000
$0
Jul
Aug
Sep
Oct
Nov
Dec


Sales are skyrocketing
$ 19,520,000
Jul
Aug
Sep
Oct
Nov
Dec

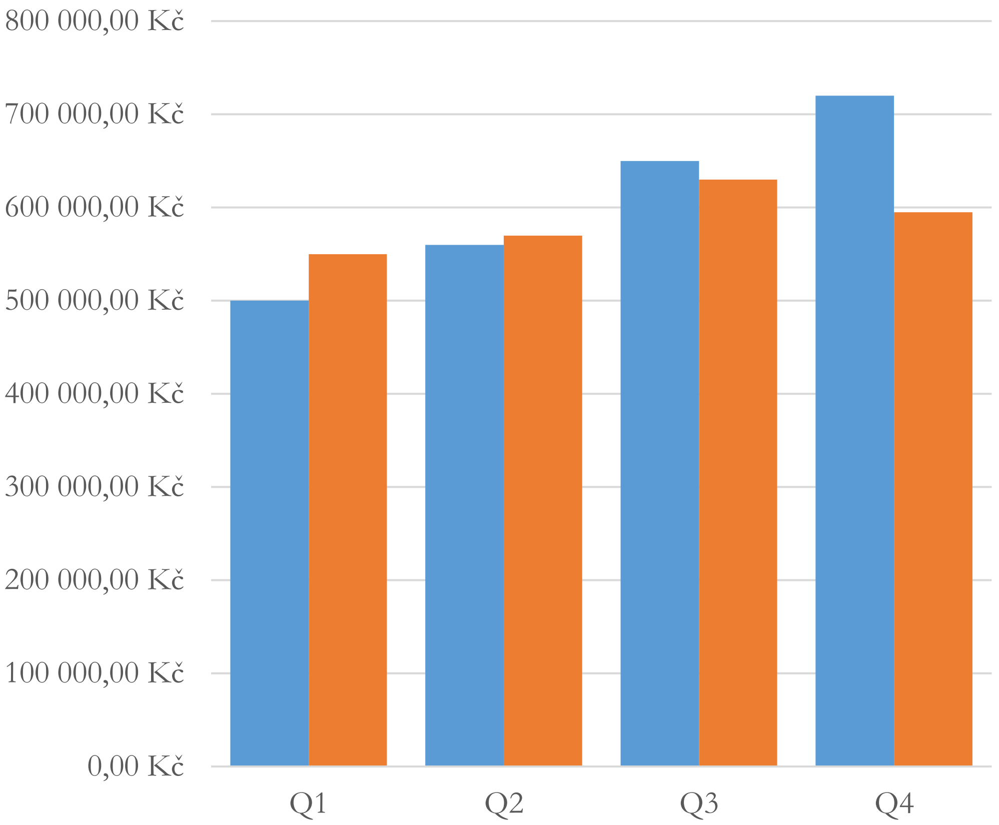
Axis scaling

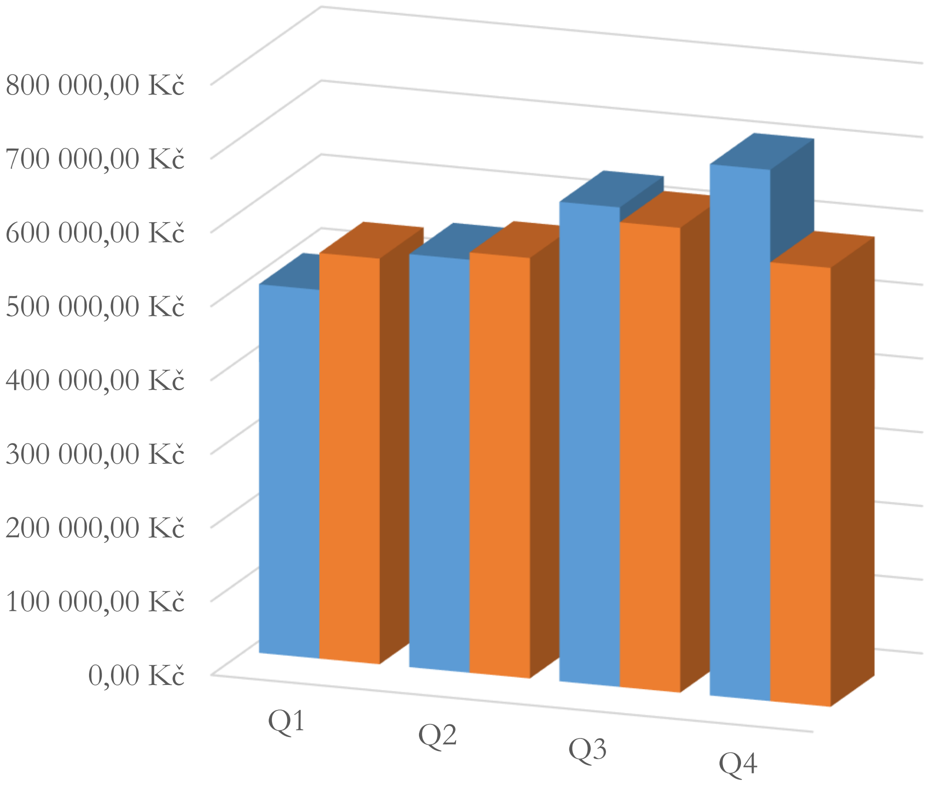
- Scale is a transformation of the data to the axis
  - Determines the min and max values on the axis, offsets, intervals between tick marks, …

- **Linear** scale
  - 1 unit on the axis correspond to n data units

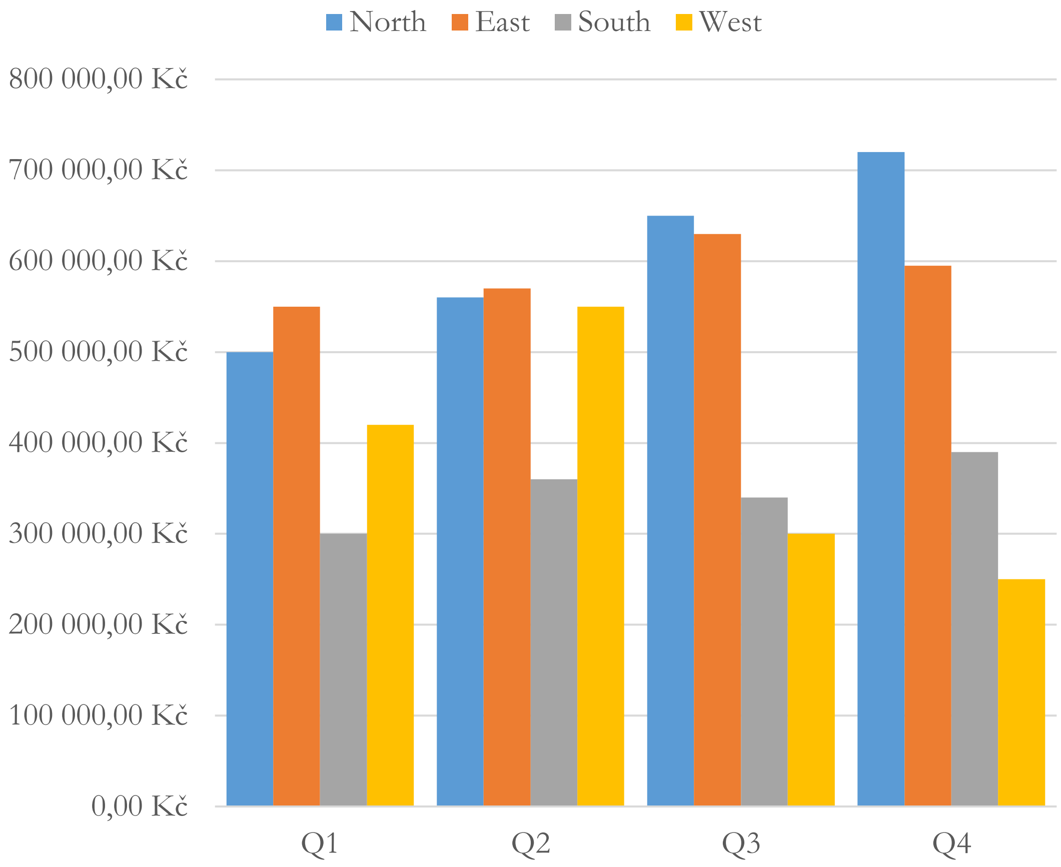
- **Logarithmic** scale
  - 1 unit on the axis correspond to $\log_m(n)$ data units

3D (1)

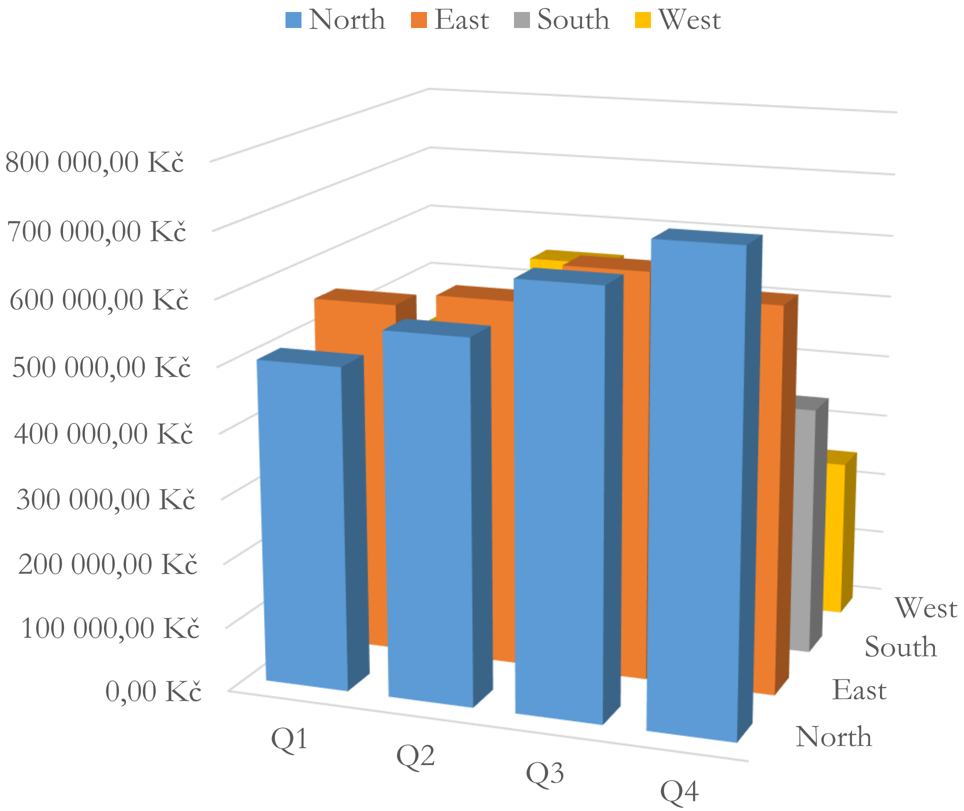

800 000,00 Kč
700 000,00 Kč
600 000,00 Kč
500 000,00 Kč
400 000,00 Kč
300 000,00 Kč
200 000,00 Kč
100 000,00 Kč
0,00 Kč
Q1
Q2
Q3
Q4


800 000,00 Kč
700 000,00 Kč
600 000,00 Kč
500 000,00 Kč
400 000,00 Kč
300 000,00 Kč
200 000,00 Kč
100 000,00 Kč
0,00 Kč
Q1
Q2
Q3
Q4

3D (2)



■ North ■ East ■ South ■ West
800 000,00 Kč
700 000,00 Kč
600 000,00 Kč
500 000,00 Kč
400 000,00 Kč
300 000,00 Kč
200 000,00 Kč
100 000,00 Kč
0,00 Kč
Q1
Q2
Q3
Q4



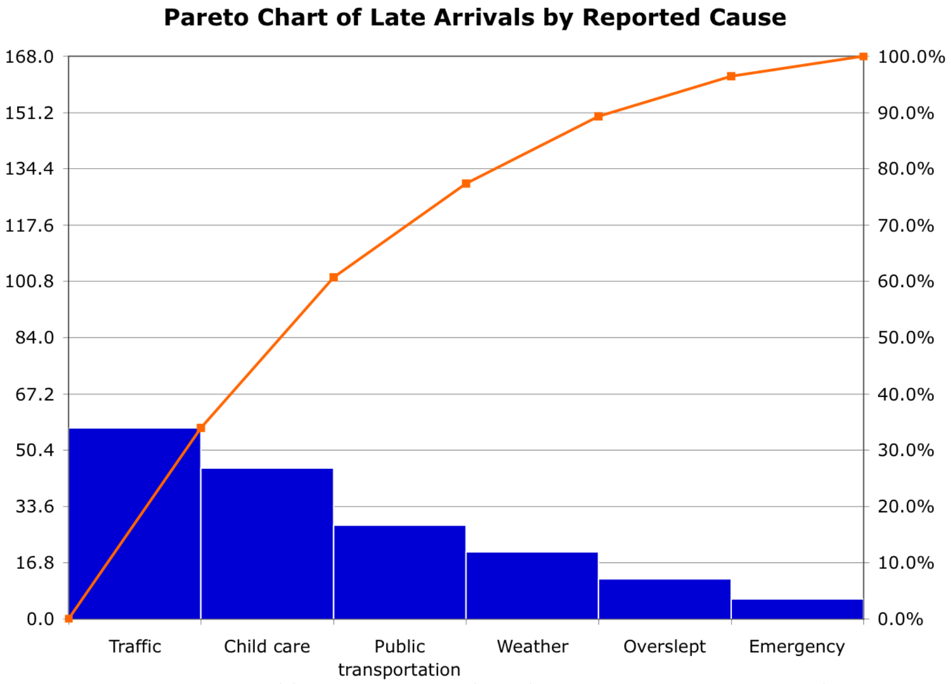
■ North ■ East ■ South ■ West
800 000,00 Kč
700 000,00 Kč
600 000,00 Kč
500 000,00 Kč
400 000,00 Kč
300 000,00 Kč
200 000,00 Kč
100 000,00 Kč
0,00 Kč
Q1
Q2
Q3
Q4
West
South
East
North

“Less traditional” visualizations

- **Combination**
  - Pareto chart
  - Small multipple
  - Scatterplot matrix
- **Part-to-whole**
  - Treemap
- **Correlation**
  - Heatmap
- **Distribution**
  - Steam-and-leaf
  - Bag plot
- **Network**
  - Arc diagram
  - Arc maps
  - Radial chart
  - Hive plots
  - BioFabric
- **Hierarchies**
  - Treemap
  - Icicle
  - Sunburst
  - Circle packing
  - Hierarchical edge bundling
- **Multivariate data**
  - Bag plot
  - Parallel coordinates
  - Parallel sets
  - Radar chart
- **Time**
  - Watterfall chart
  - Gantt chart
  - Slopegraph
  - Sparklines
- **Others**
  - Word cloud

Pareto chart

- **Combination** of one **unit of measure** and a **cumulative percentage** (or running total) of that measure
  - The individual measures are usually visualized using **bar chart**
  - The cumulative measure visualized as a **line graph**



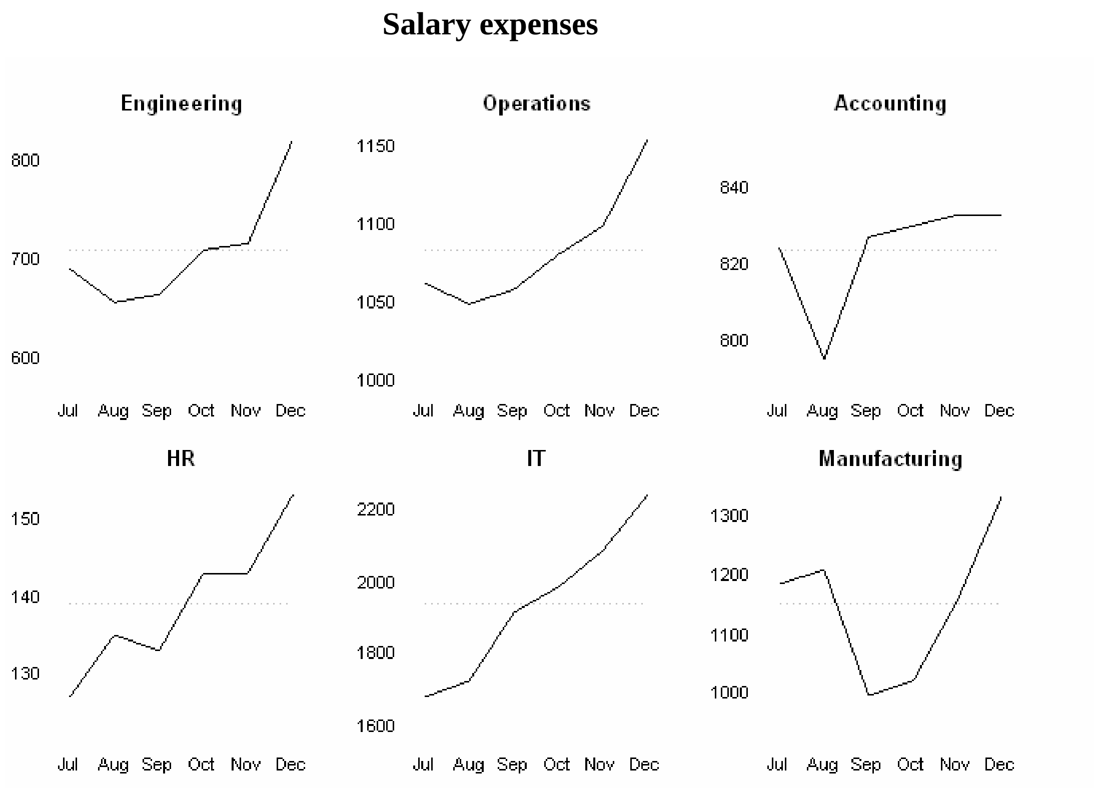
source: http://en.wikipedia.org/wiki/Pareto_chart#mediaviewer/File:Pareto.PNG

Small multiple

- Also called **trellis chart, lattice chart, grid chart**, or **panel chart**

- Series of graphs using the same scale and axes

- Allows to see different slices of the same data using the same base graphics

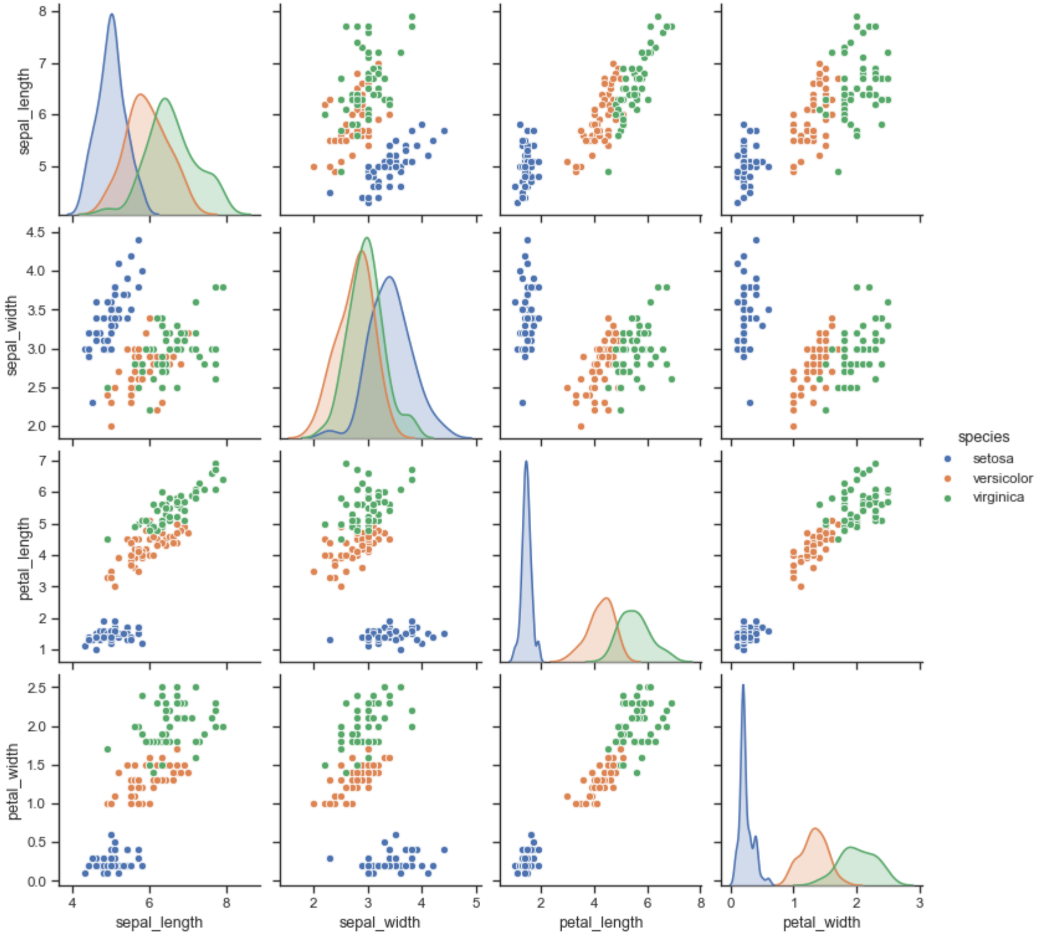


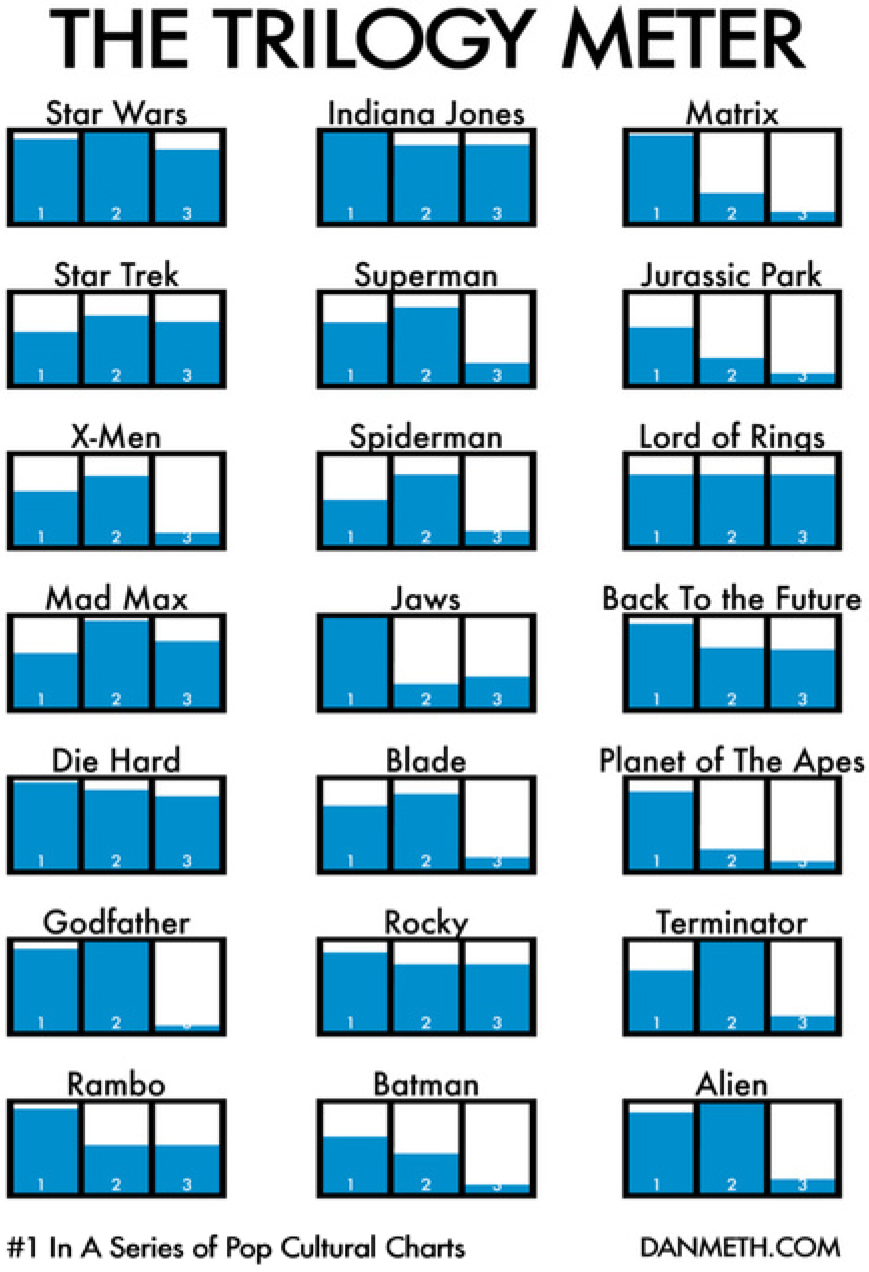
source: http://upload.wikimedia.org/wikipedia/en/a/a6/Smallmult.png

Scatterplot matrix

- **All pairwise scatter plot** of given variables

- Typically used to **get feeling** for the data to be investigated



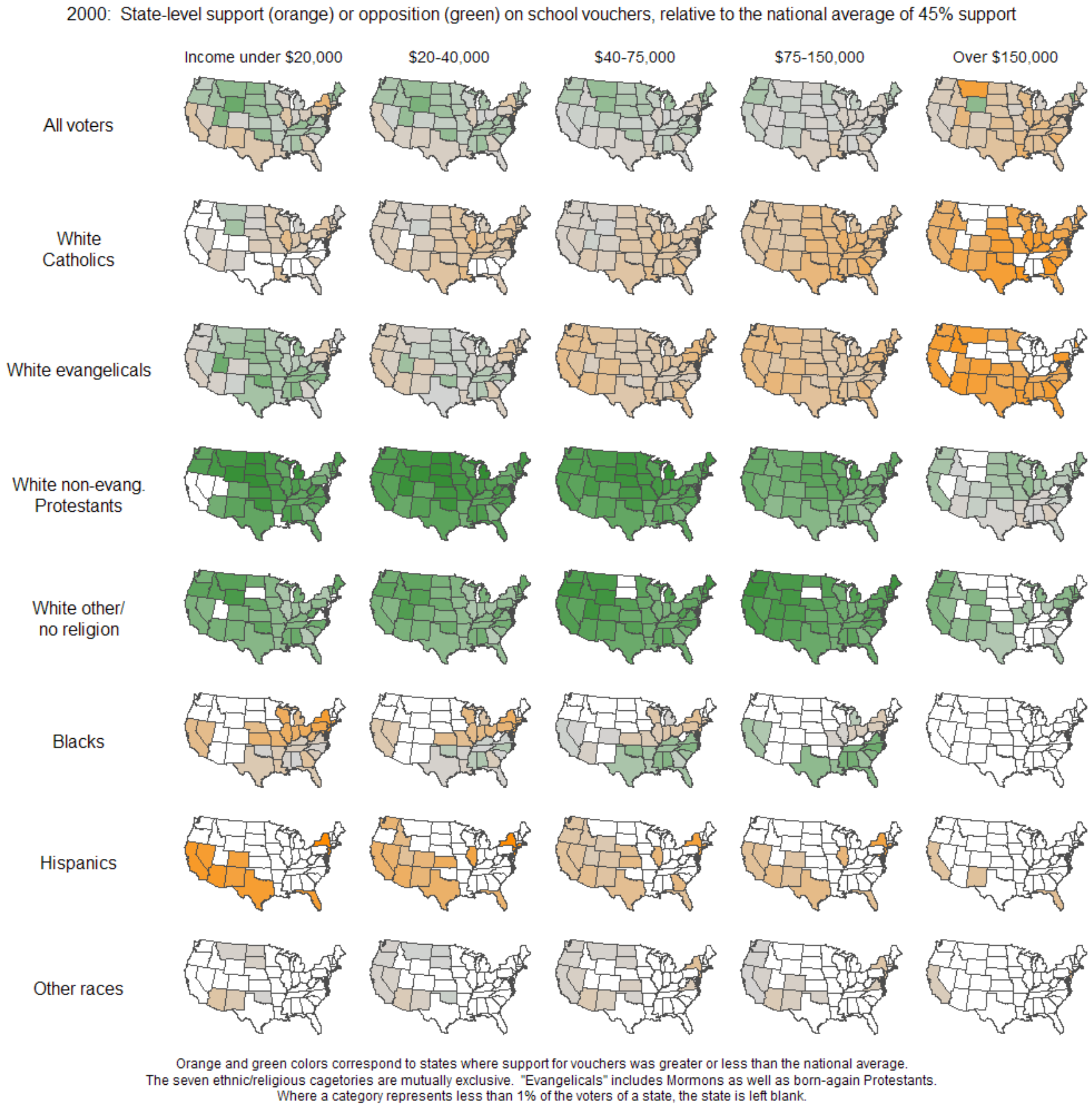


THE TRILOGY METER

#1 In A Series of Pop Cultural Charts

DANMETH.COM

source: http://danmeth.com/post/77471620/my-trilogy-meter-1-in-a-series-of-pop-cultural



2000: State-level support (orange) or opposition (green) on school vouchers, relative to the national average of 45% support

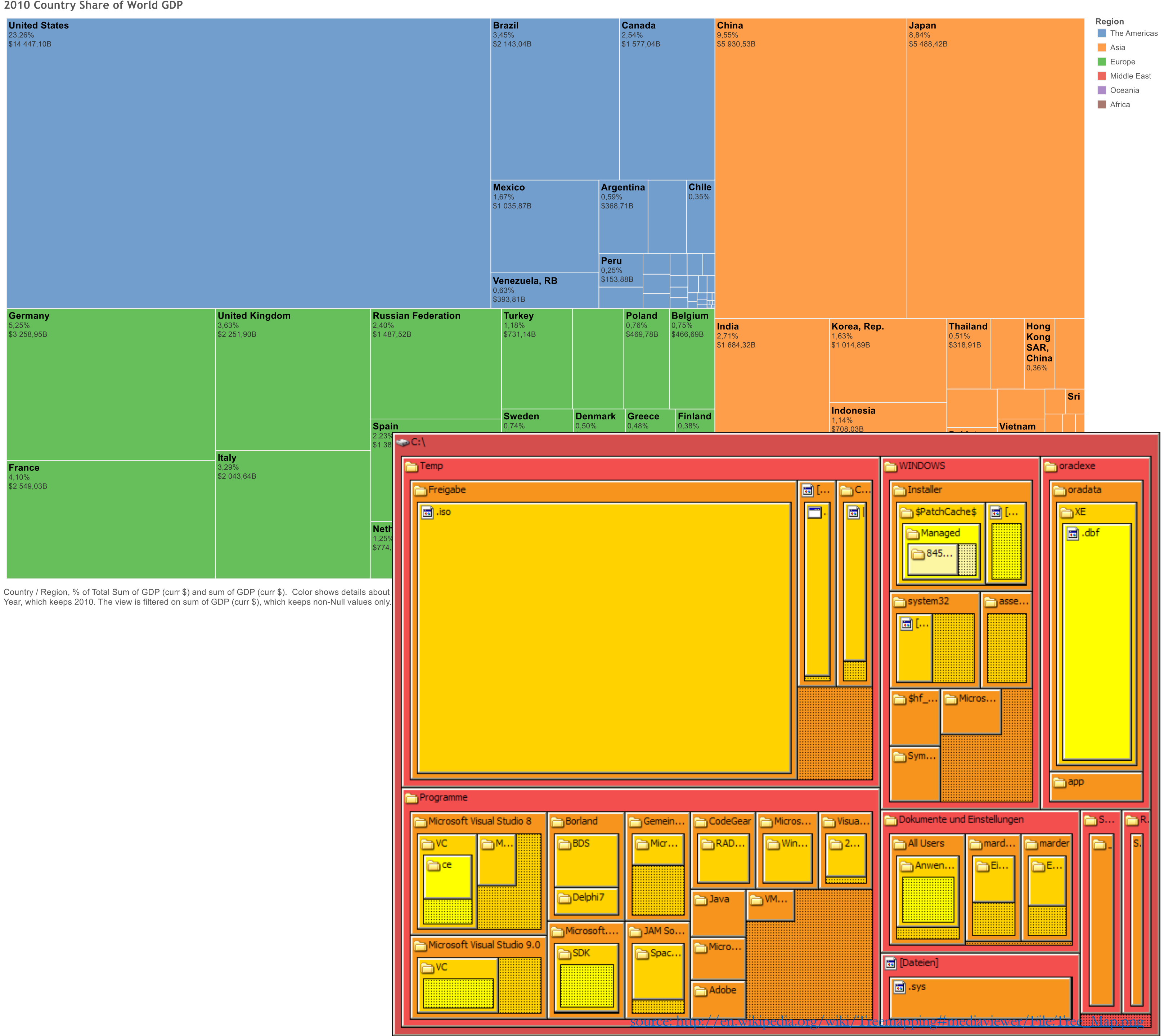
Orange and green colors correspond to states where support for vouchers was greater or less than the national average.
The seven ethnic/religious cagetories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants.
Where a category represents less than 1% of the voters of a state, the state is left blank.

source: http://andrewgelman.com/2009/07/15/hard_sell_for_b/

Treemap

- **Part-to-whole** and/or **hierarchical** design

- Nested rectangles can capture hierarchy (if any is present)

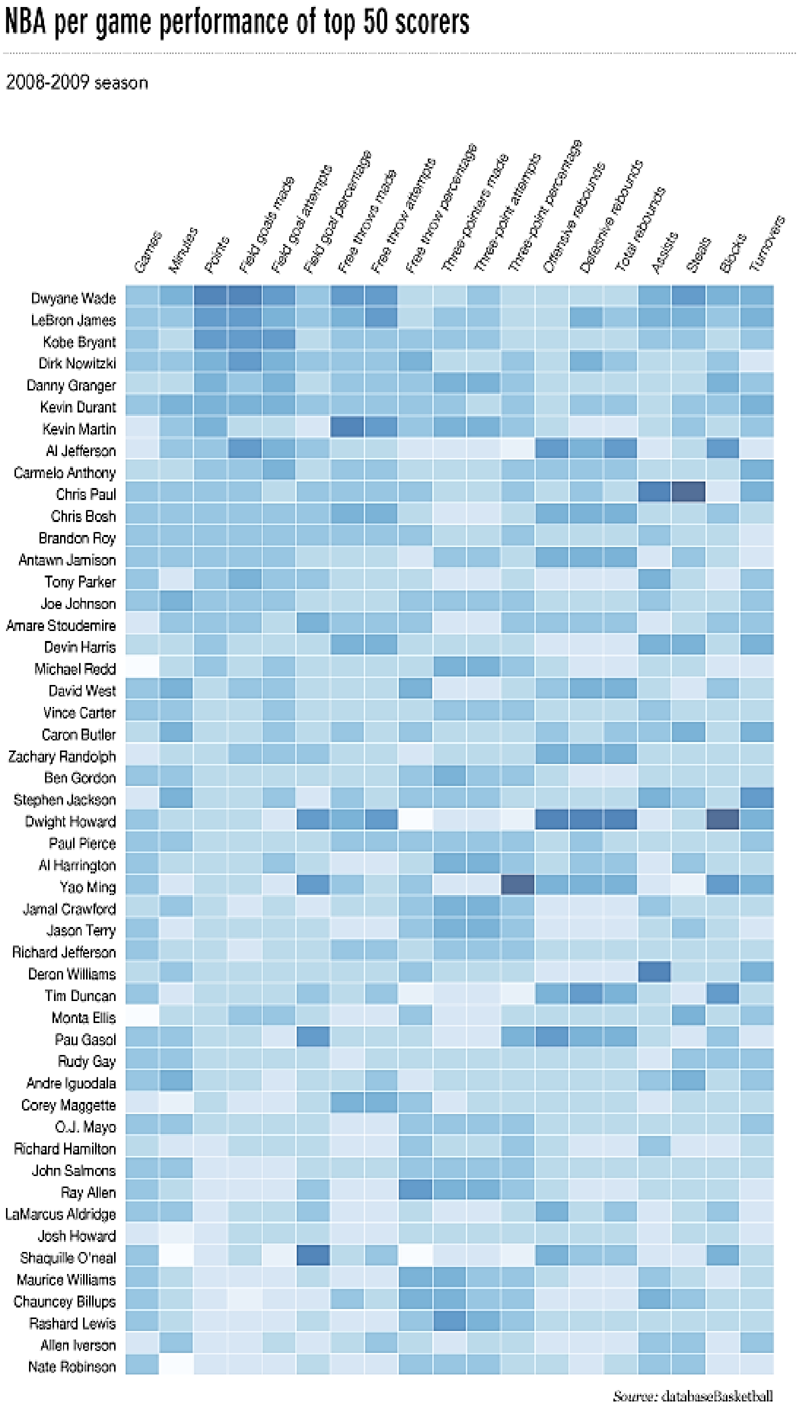


Country / Region, % of Total Sum of GDP (curr $) and sum of GDP (curr $). Color shows details about Year, which keeps 2010. The view is filtered on sum of GDP (curr $), which keeps non-Null values only.

source: http://en.wikipedia.org/wiki/Treemapping#mediaviewer/File:Tree_Map.png

Correlation matrix (1)

- Also known as **heatmap** or matrix diagram
- Display of how (or whether) **two sets** of **categorical** values **relate** to each other (correlate)
- Can be used for visualization of graphs



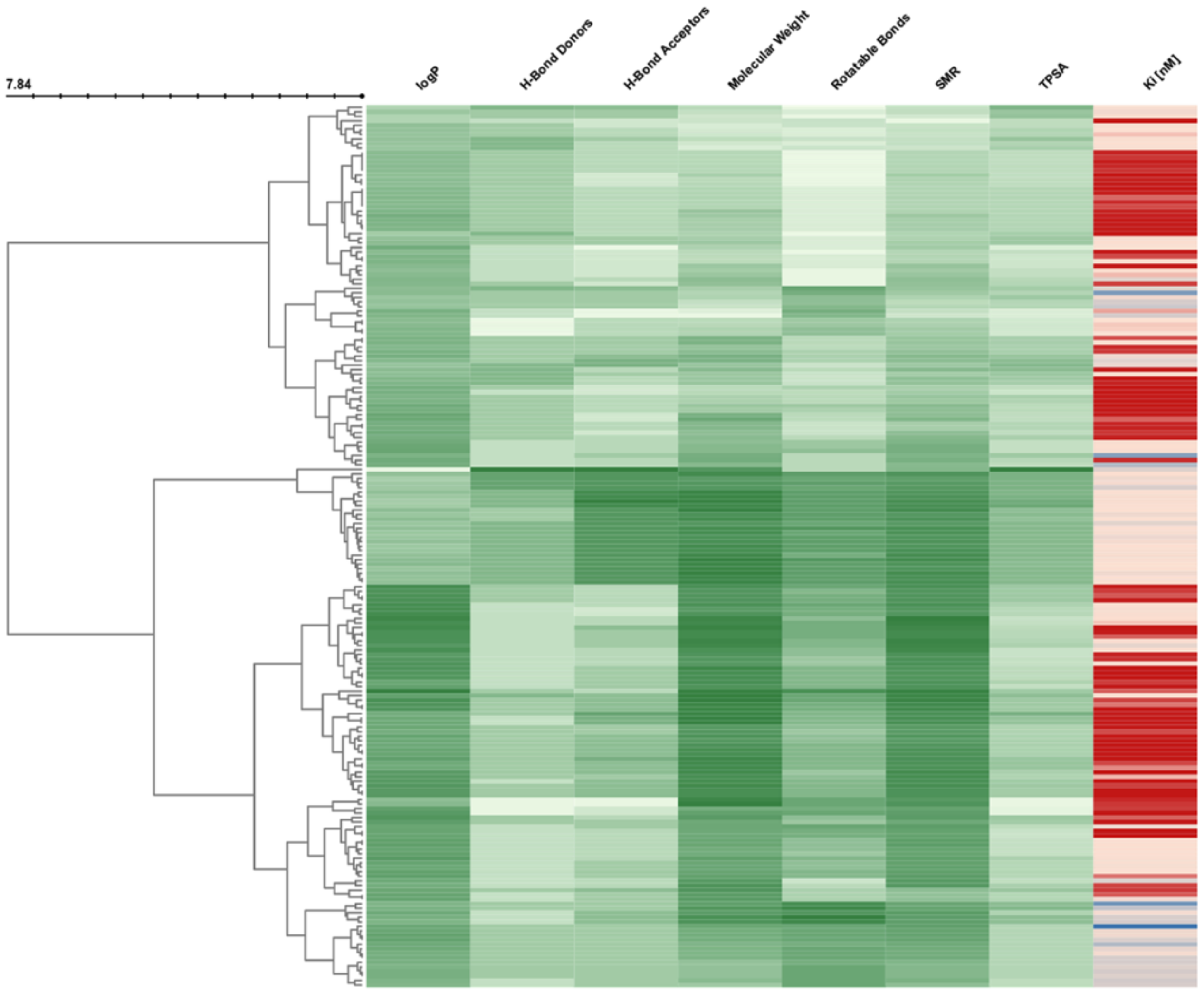
NBA per game performance of top 50 scorers

2008-2009 season

Source: databaseBasketball

Correlation matrix (2)

- The correlation information can be incorporated with the help of **dendrograms**
  - Helps to reveal clusters in data



source: InCHlib - interactive cluster heatmap for web applications

Stem-and-leaf plot

- Similar to histogram displays **frequency of each class**

- Unlike histogram, it allows to see the **original data points**

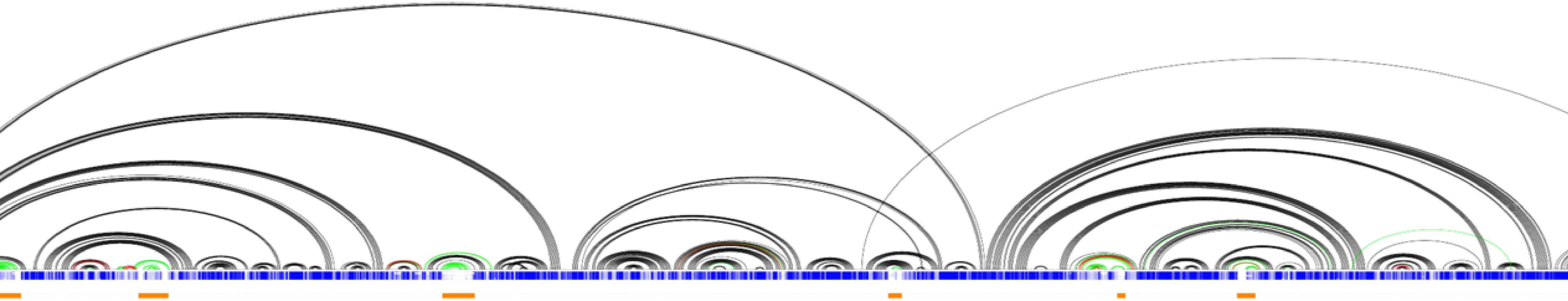
- Suitable only for **small datasets**

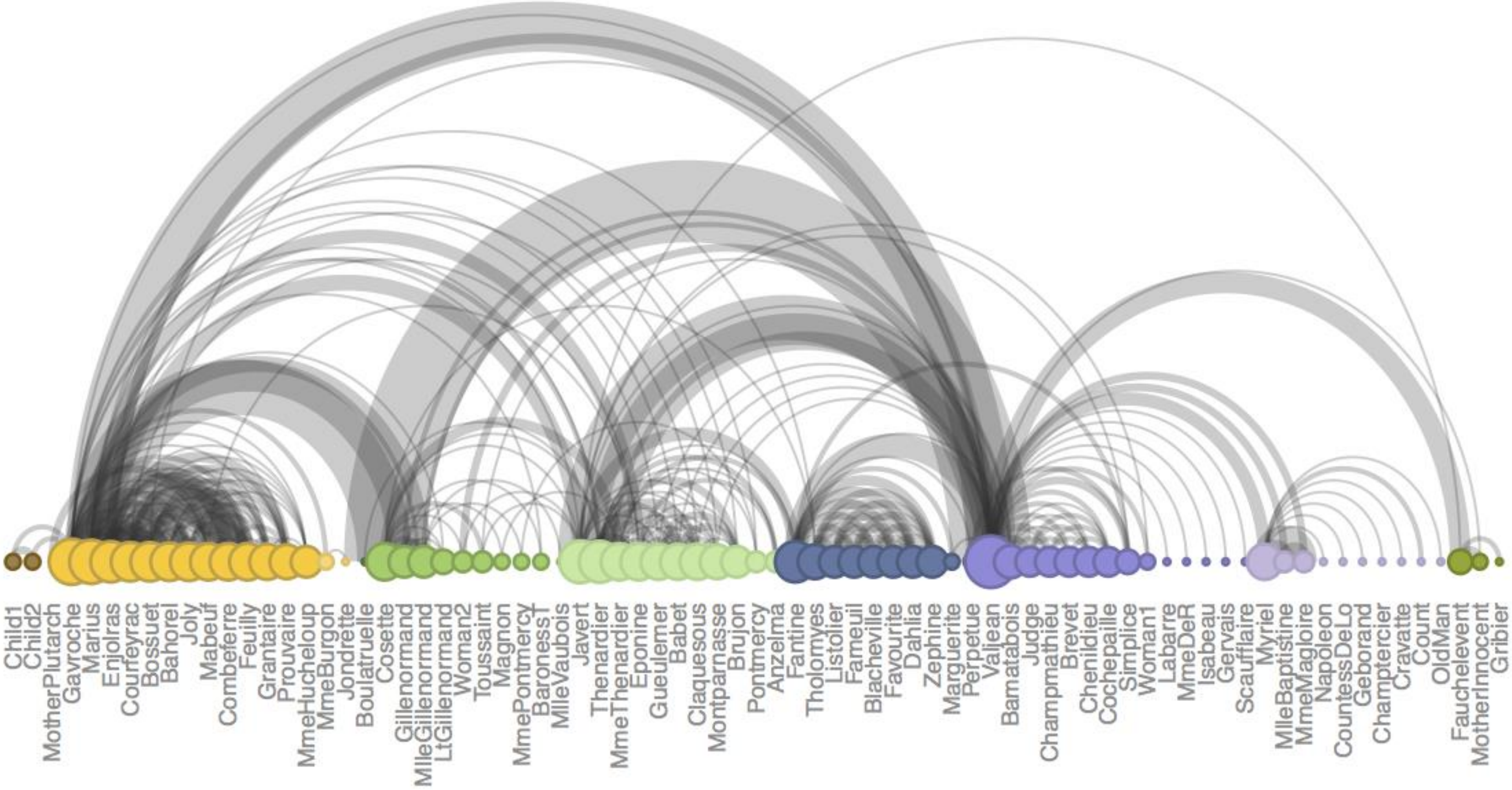
Grades

steam	**leaf**
4	2 3
5	0 5 7
6	0 0 7 9
7	2 8
8	1 1 3 8 7
9	5

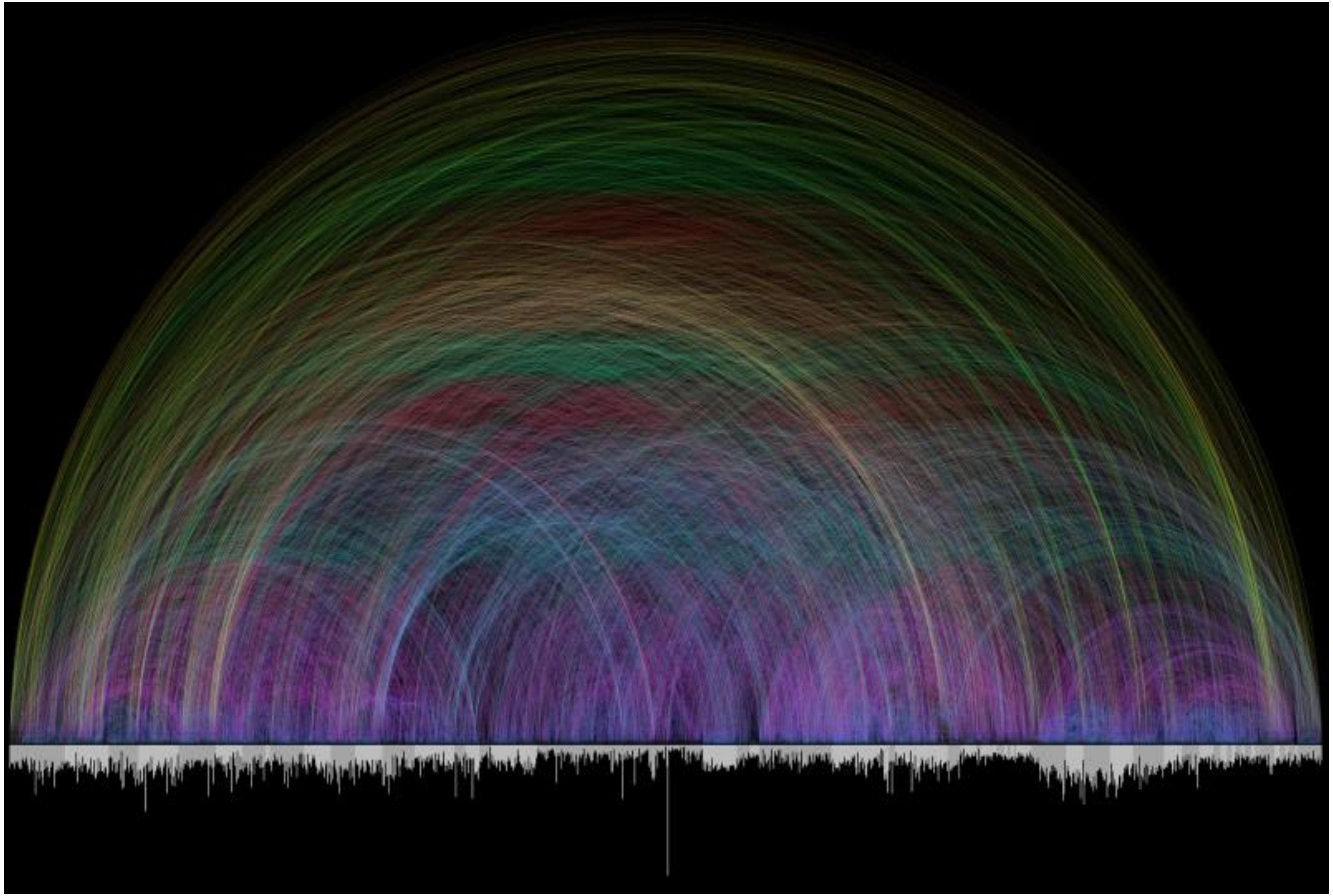
Arc diagram

- **Vertices** are placed **along a line** and edges are drawn as semicircles
  - 1D layout of a graph → suitable when the vertices have a linear ordering
  - Arcs represent relationships
  - Further visual attributes such as color can encode additional information, e.g.,





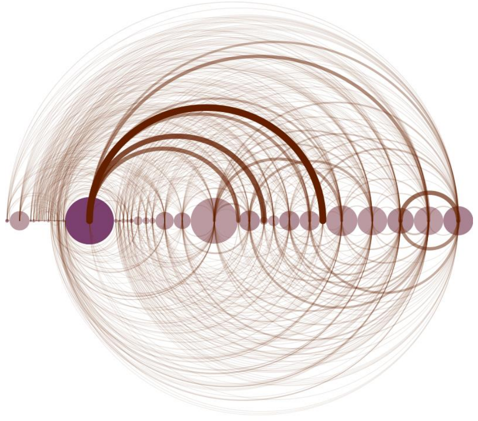
source: http://gastonsanchez.com/got-plot/how-to/2013/02/02/Arc-Diagrams-in-R-Les-Miserables/



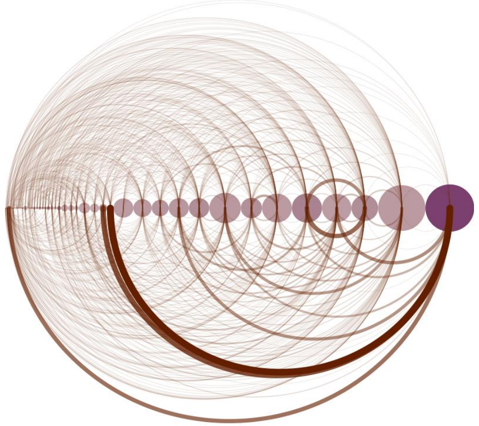
source: http://www.chrisharrison.net/index.php/Visualizations/BibleViz

A map of 63,799 **cross-references** found in the **Bible**. The bottom bars represent number of verses in the given chapter. Color of arcs represents the distance between the two chapters.

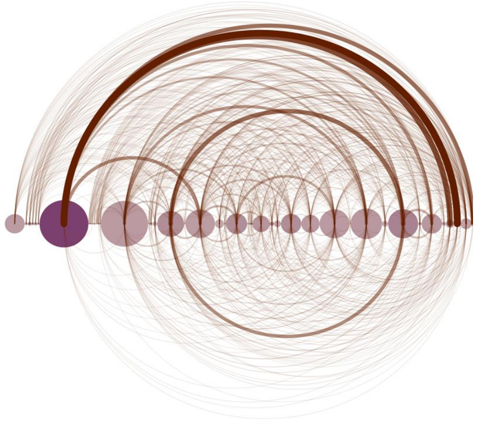
Circle size = Number of messages
Circle color = Average message length



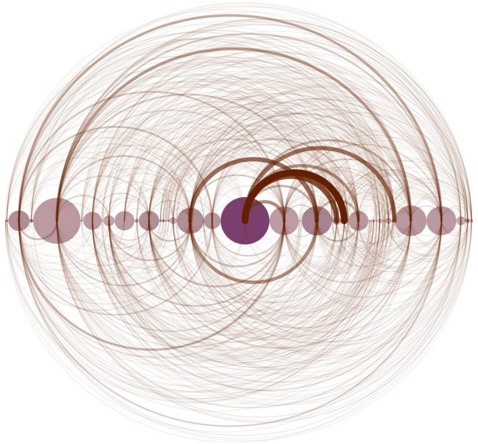
Sorted by the amount of incoming references



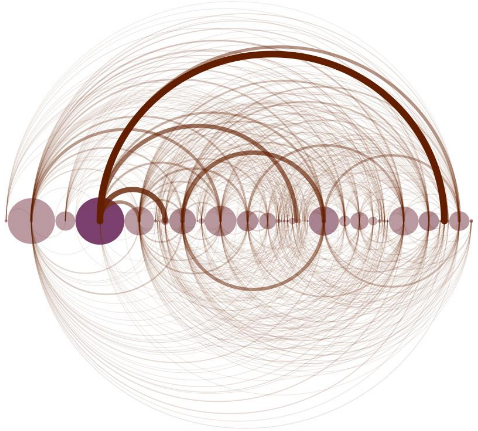
Sorted by the amount of outgoing references



Sorted by rate of incoming/outgoing references



Sorted by user name

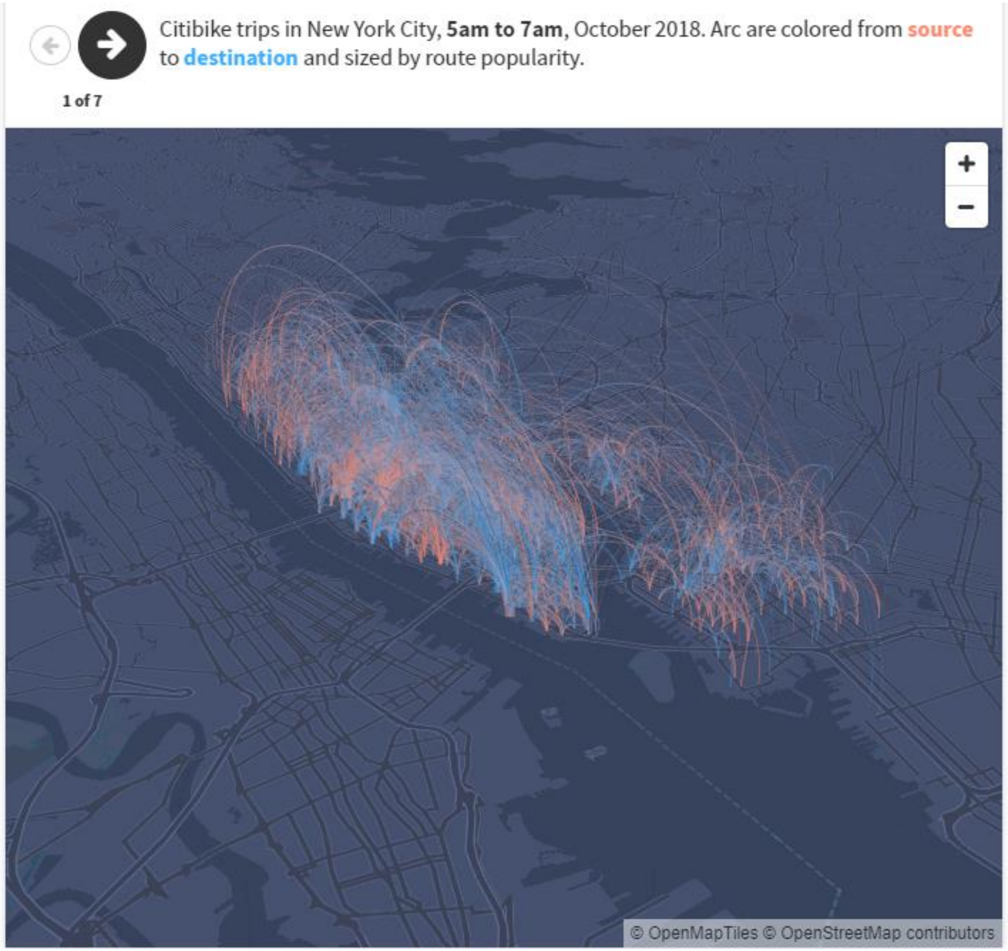


Unsorted

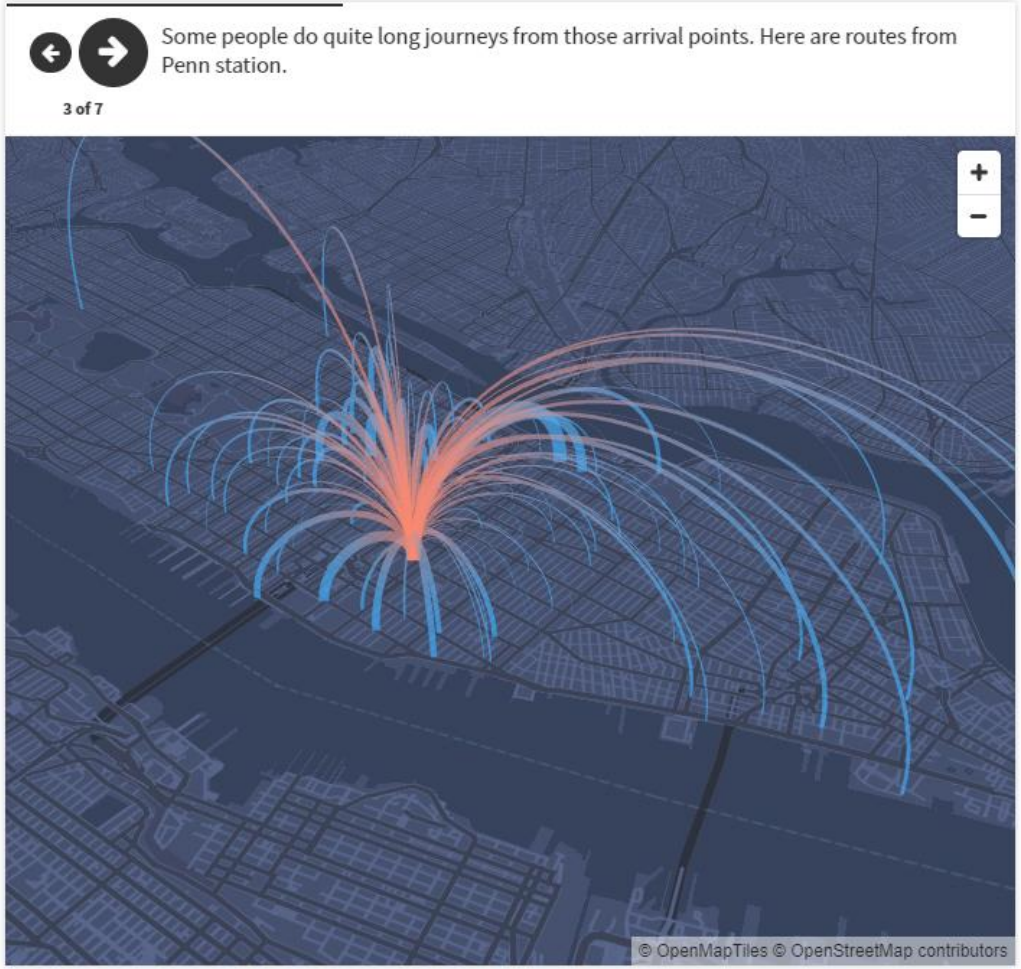
- Visualization of **IRC communication** behavior: Who is talking to whom?
- **Arcs** are **directional** and drawn clockwise:
  - In the upper half of a graph they point from left to right, in the bottom half from right to left
  - Arc strength corresponds to the number of references from the source to the target
- This visualization favors strong social connections over sociability: Frequent references between the same two users feature more prominently than combined references from several sources to a single target.

source: http://datavis.dekstop.de/irc_arcs/

Arc maps



Citibike trips in New York City, **5am to 7am**, October 2018. Arc are colored from **source** to **destination** and sized by route popularity.

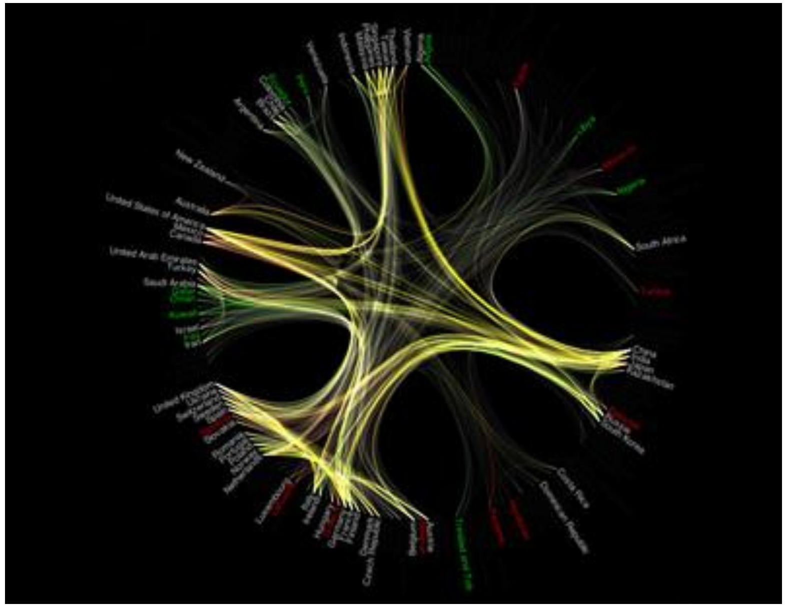


Some people do quite long journeys from those arrival points. Here are routes from Penn station.

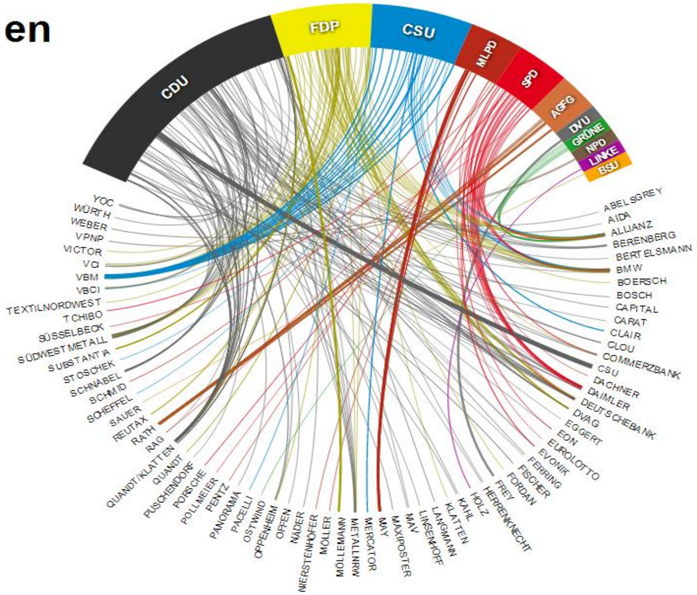
source: https://flourish.studio/2018/11/16/arc-map-webgl/

Radial chart

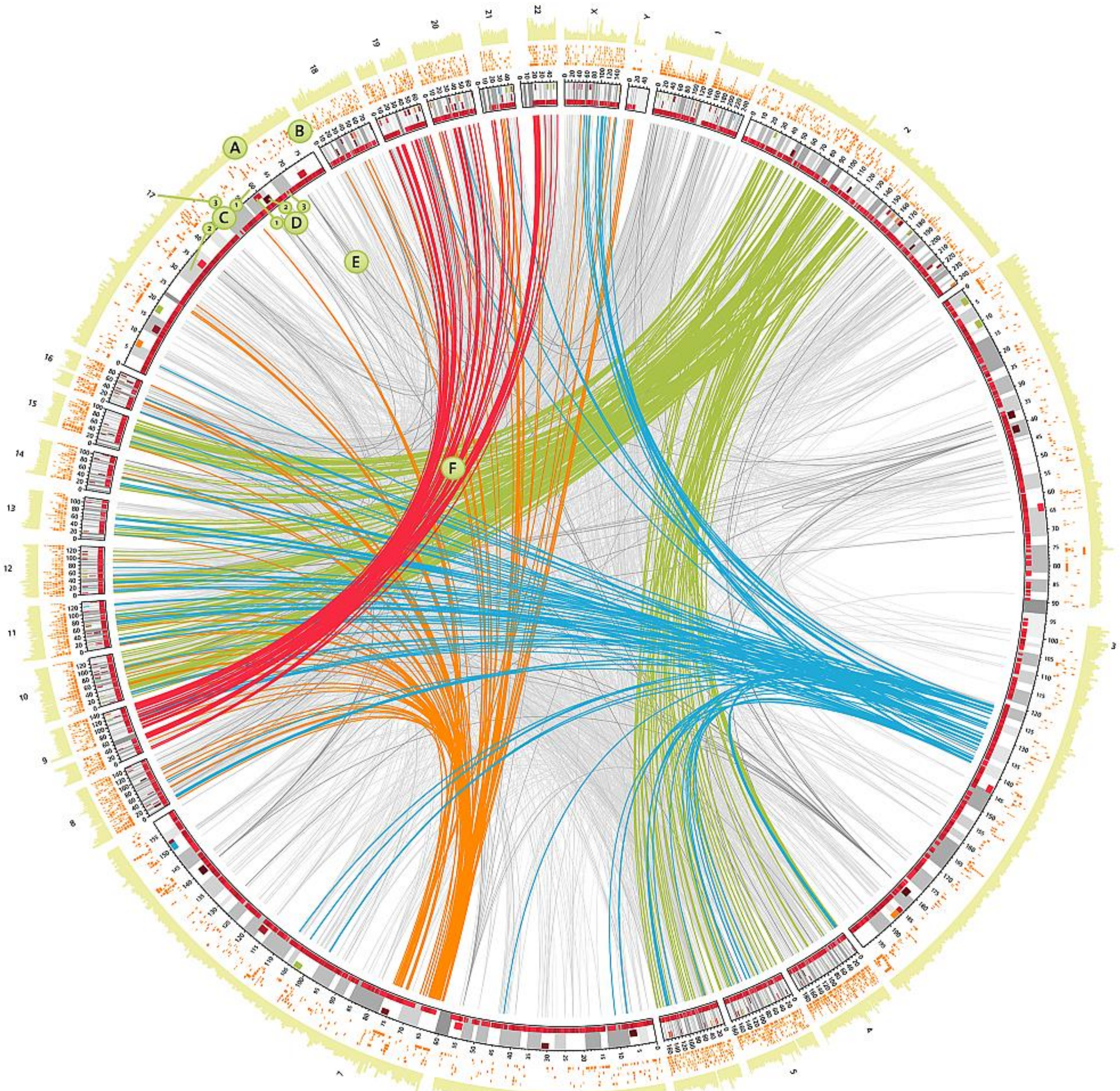
- Modification of the arc diagram where the **x-axis** forms a **ring**
- Also called **circular layout** or **chord diagram**

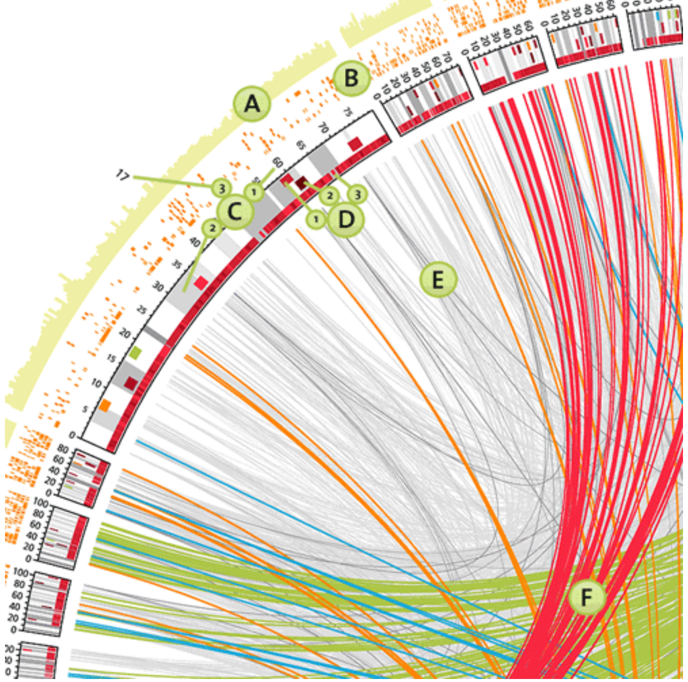


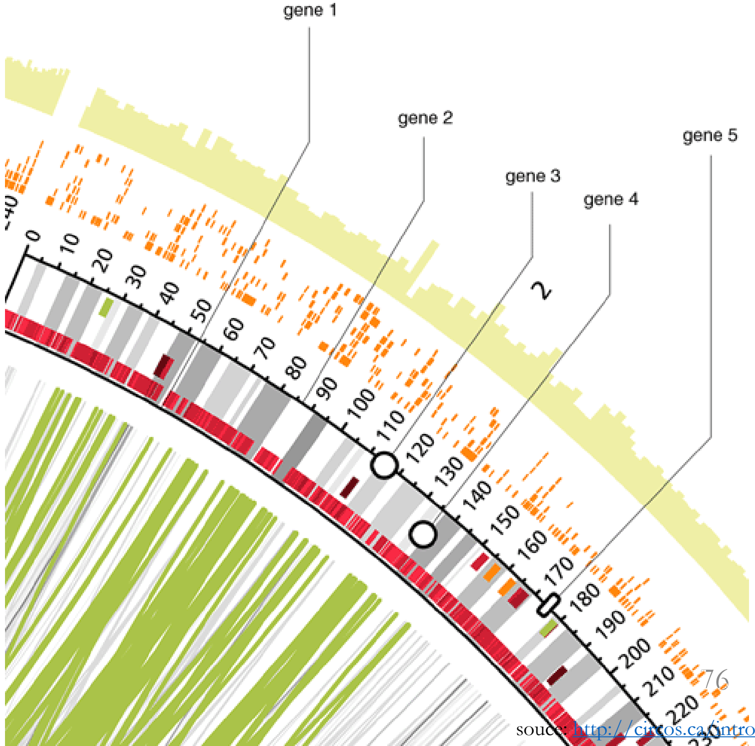
Tracking the commercial ties between most countries across the globe.
http://cephea.de/gde/



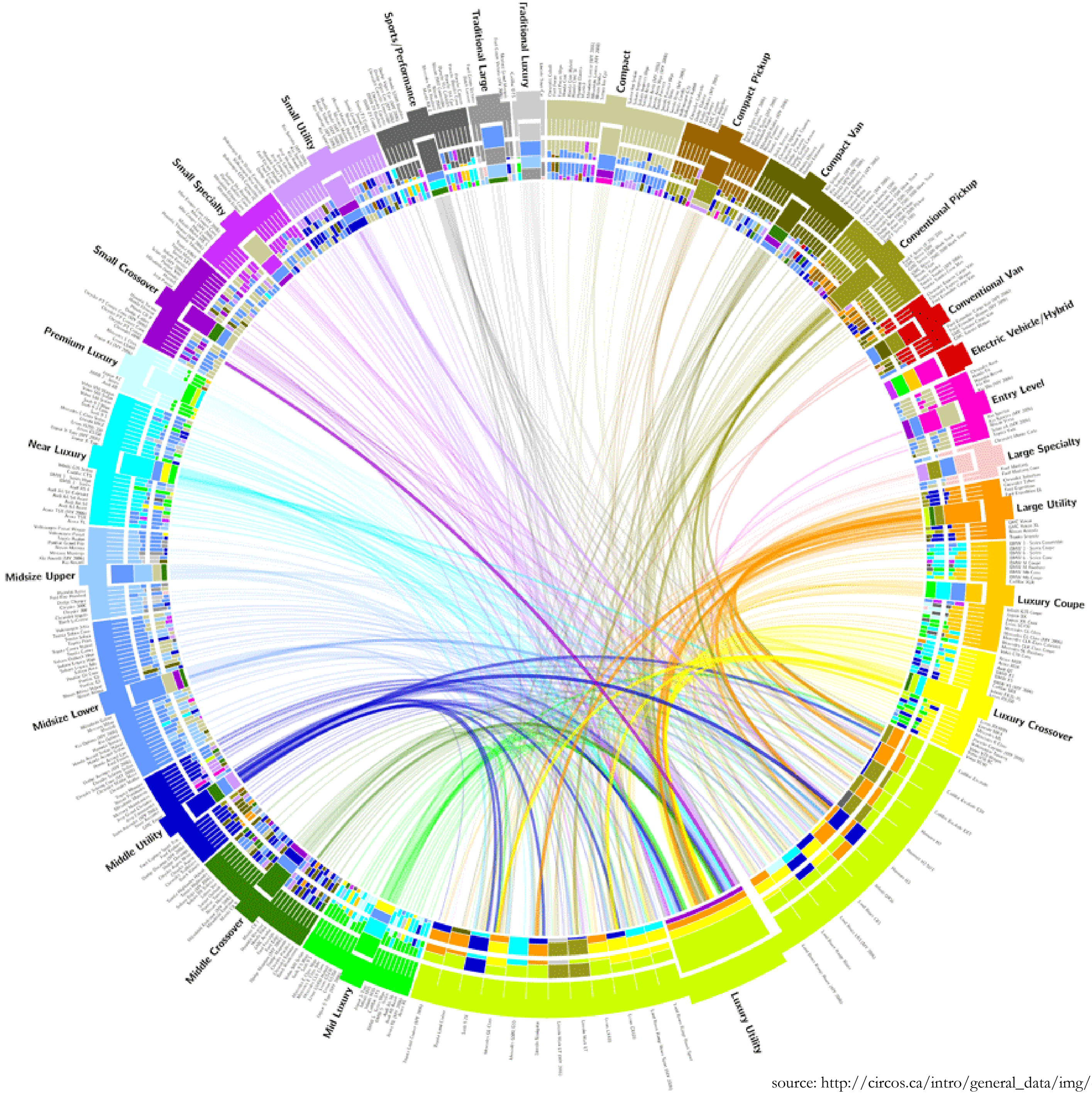
Money flow from private donators to parties in the German Bundestag (house of the parliament).
http://labs.vis4.net/parteispenden/





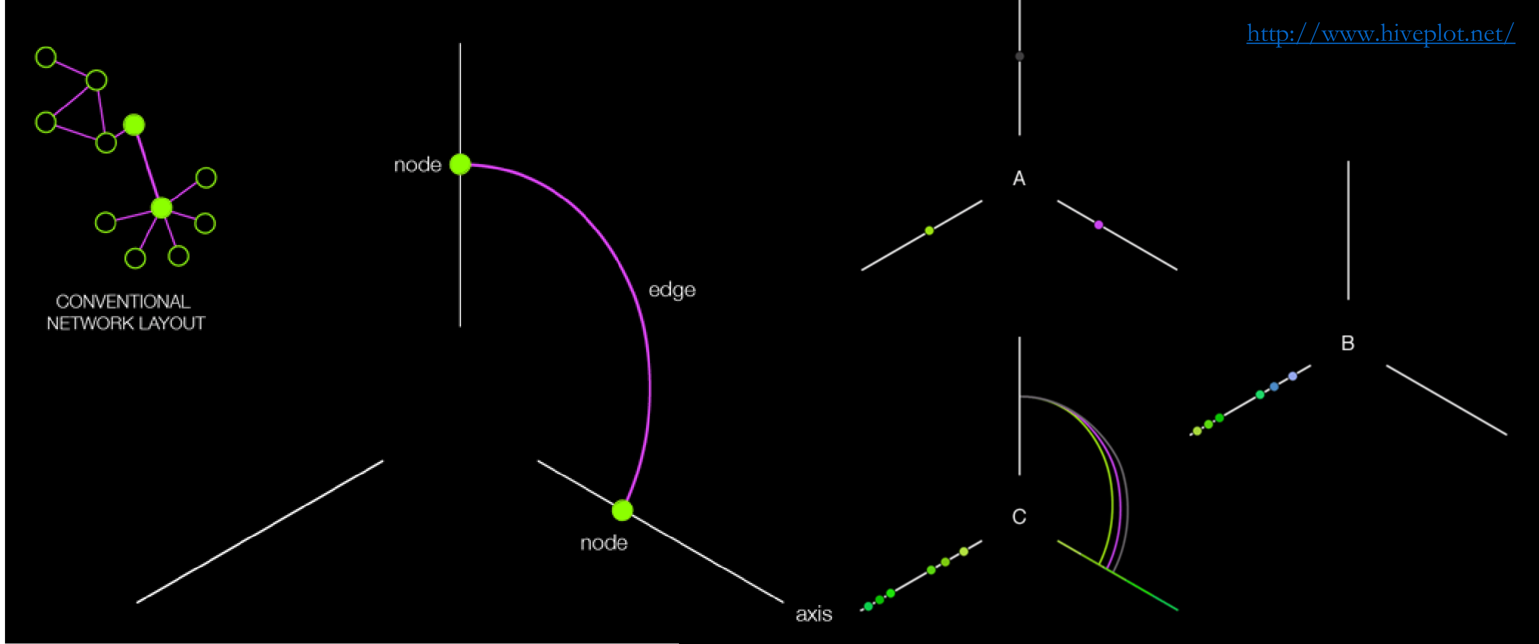


souce: http://circos.ca/intro/genomic_data/

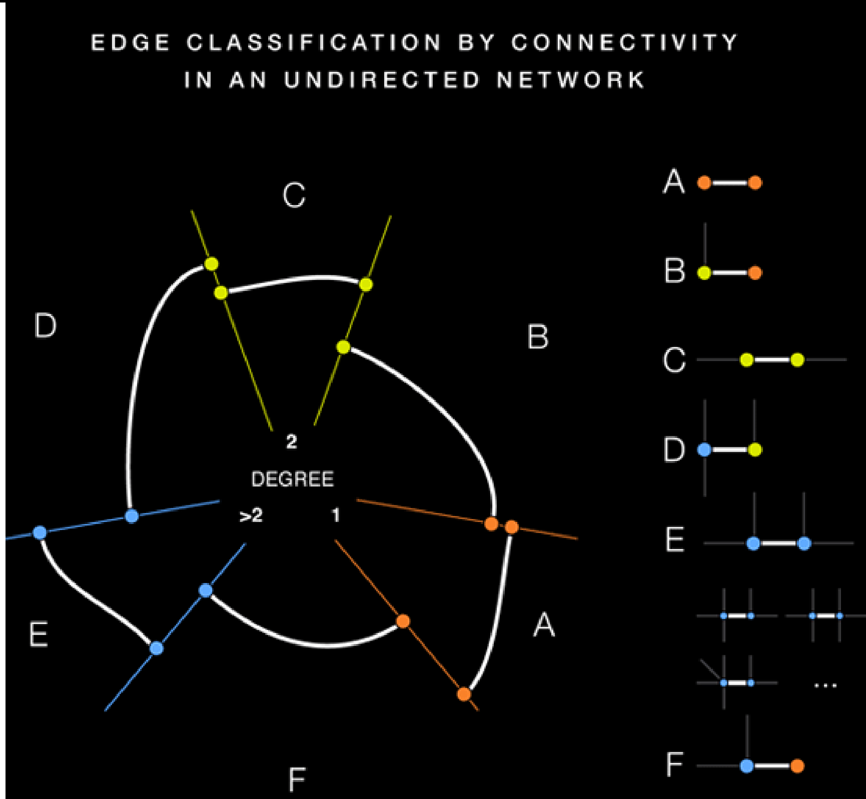


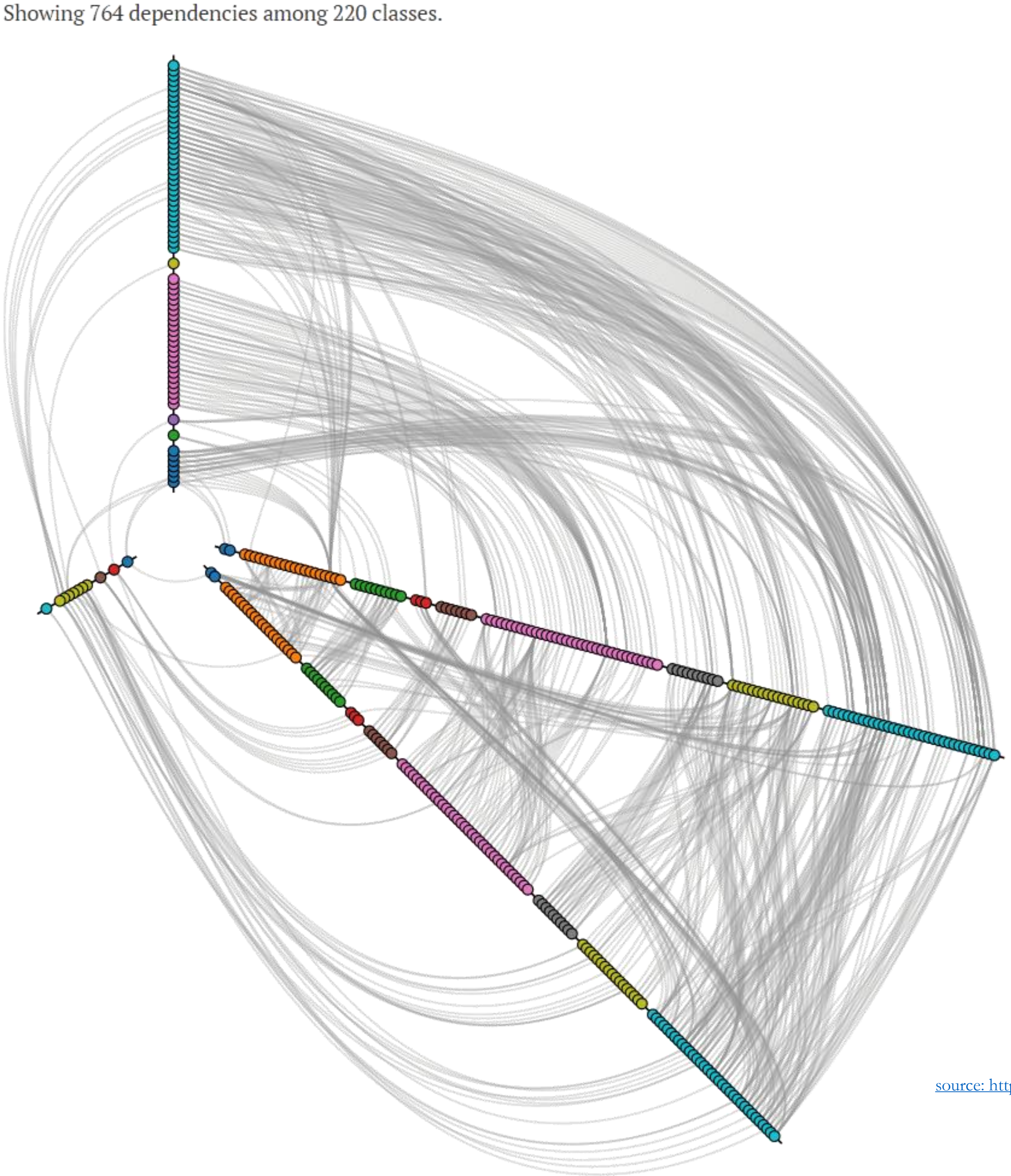
source: http://circos.ca/intro/general_data/img/circos-car-purchase.png

Hive plots


http://www.hiveplot.net/

- Visualization method for drawing **networks**
  - **Nodes** mapped to and positioned **on radially distributed linear axes** → linear layout of nodes
    - Can be divided into **segments**
  - **Edges** drawn as **curved links**
  - Graph structure can be mapped to
    - Axis
    - Position
    - Color


EDGE CLASSIFICATION BY CONNECTIVITY IN AN UNDIRECTED NETWORK



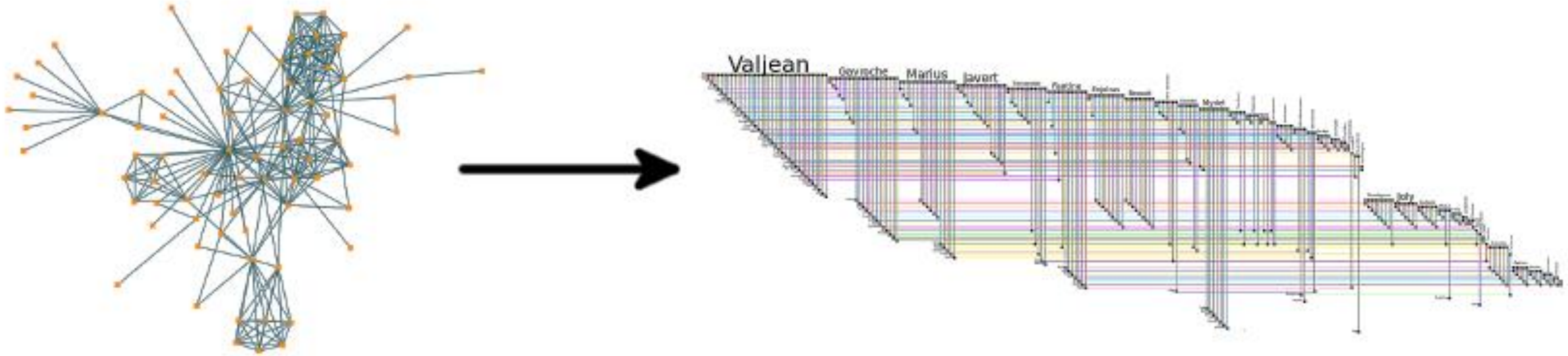
Showing 764 dependencies among 220 classes.

Each **node** represents a **class** in a software library. Nodes are divided into three categories. The **12 o'clock axis** (the top) shows ***source*** **nodes—classes** with only outgoing dependencies. The **bottom-left axis** shows ***target*** **nodes** with only incoming dependencies. The remaining nodes in the **bottom-right** have both **incoming** and **outgoing** dependencies; these are **duplicated** to reveal dependencies within this category.

source: http://bost.ocks.org/mike/hive/

BioFabric

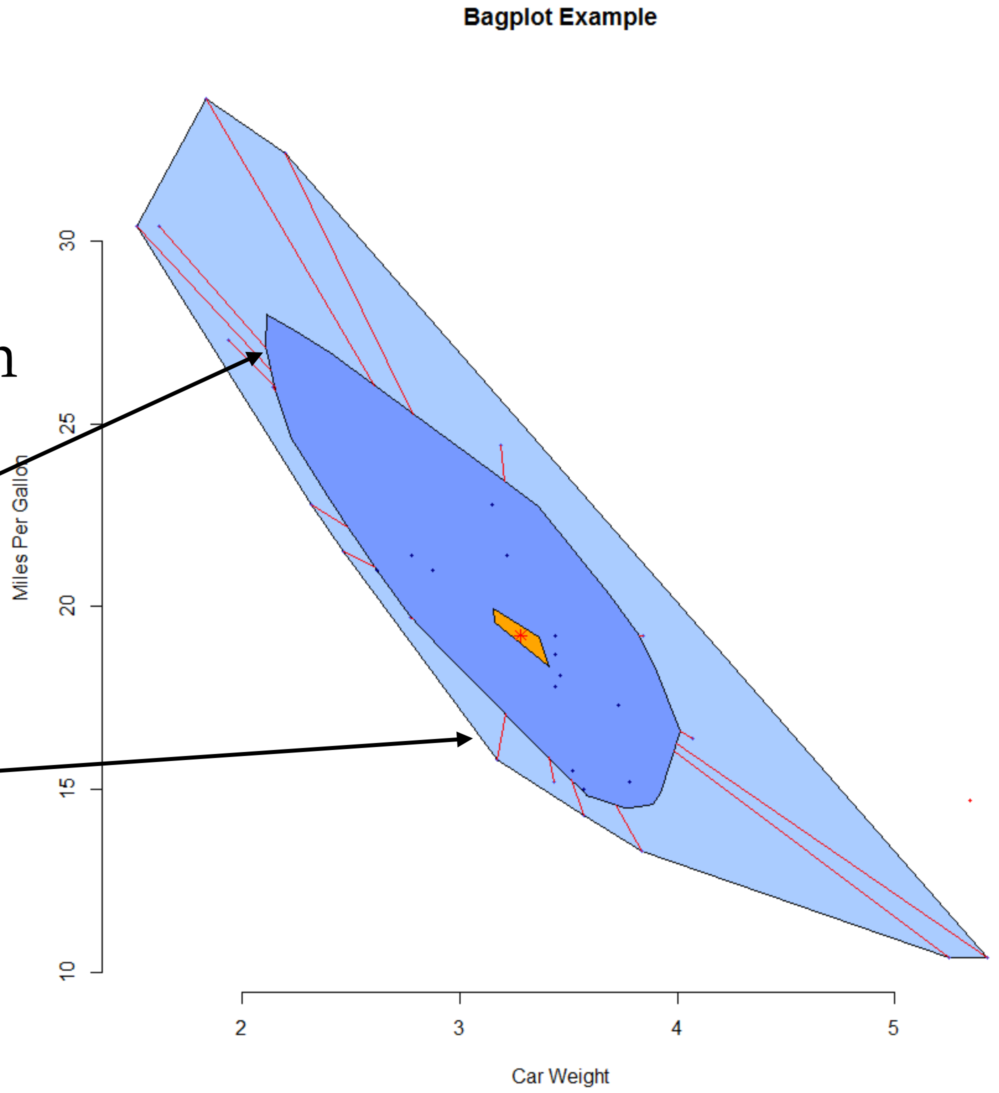
- Dealing with **large networks**
- **Nodes** as **horizontal line** segments
- **Edges** as darker **vertical line** segments, do not overlap and can originate anywhere on the line segment



http://www.biofabric.org/gallery/pages/SuperQuickBioFabric.html

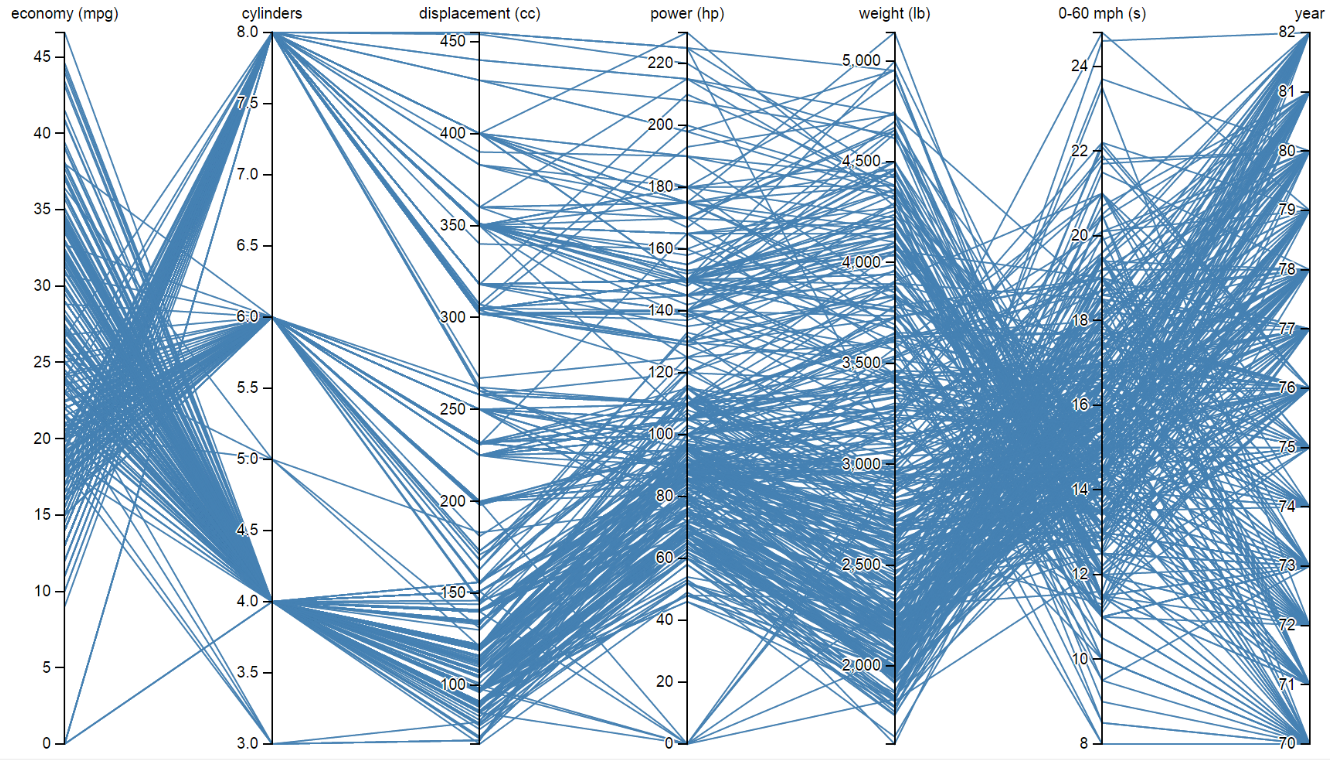
Bag plot

- Also called starburst plot
- **Bivariate generalization** of the well known **boxplot**
  - Consists of three nested polygons
    - **Bag**
      - Bag contains 50 percent of all points (IQR)
    - **Loop**
      - Convex hull of points within the fence
    - **Fence**
      - Inflation of the bag by a factor
      - Points outside of the fence are considered outliers



Parallel coordinates

- A way to visualoze high-dimensional data in 2D
- Unlike line charts, a line represent a **single object along multiple dimensions**
- Each **dimension** is **scaled** so that each data point ends up somewhere between **min** (**bottom** of scale) and **max** (**top** of the scale)



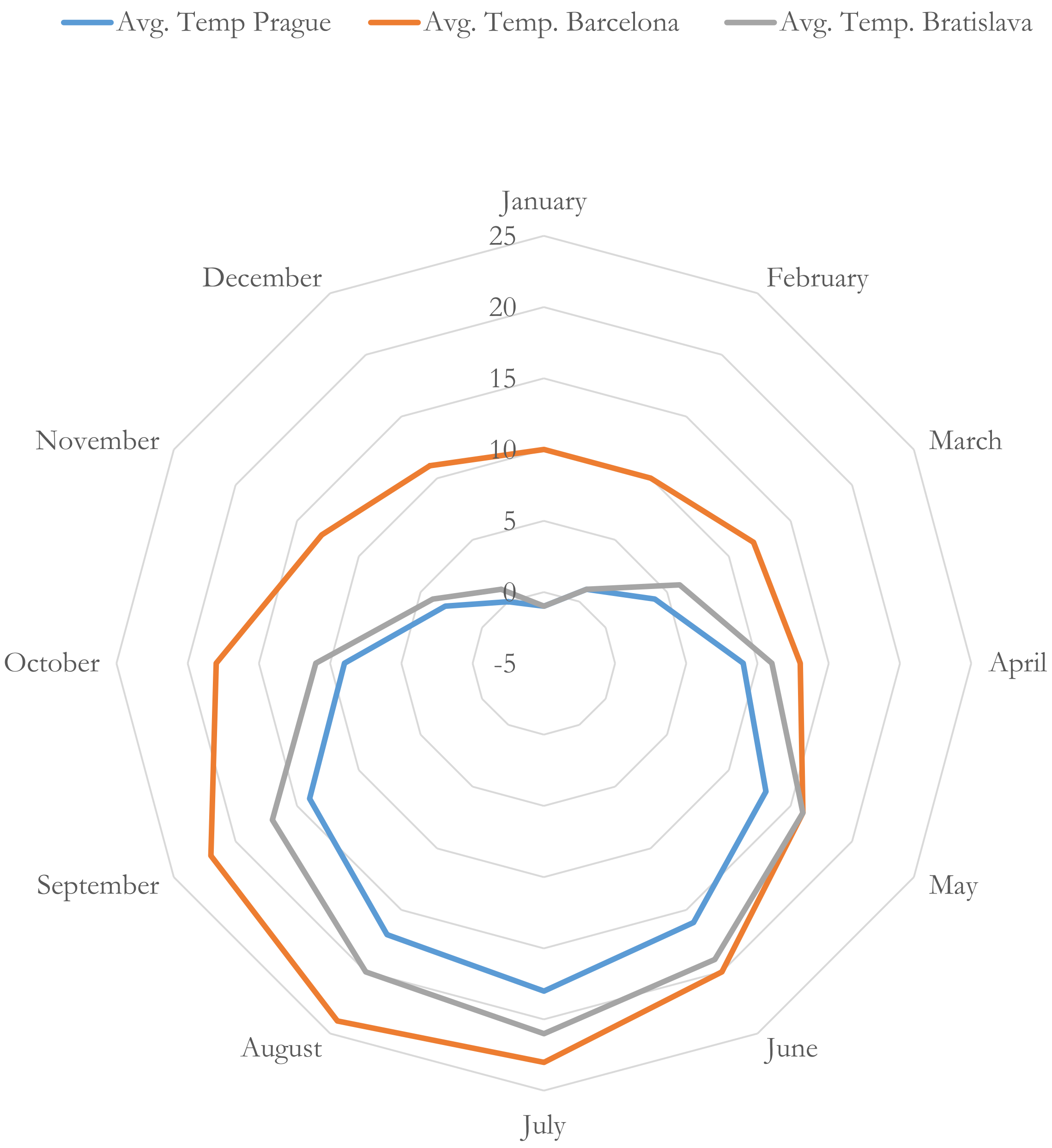
source: http://bl.ocks.org/jasondavies/1341281

Radar chart

- Also known as **spider/star chart**

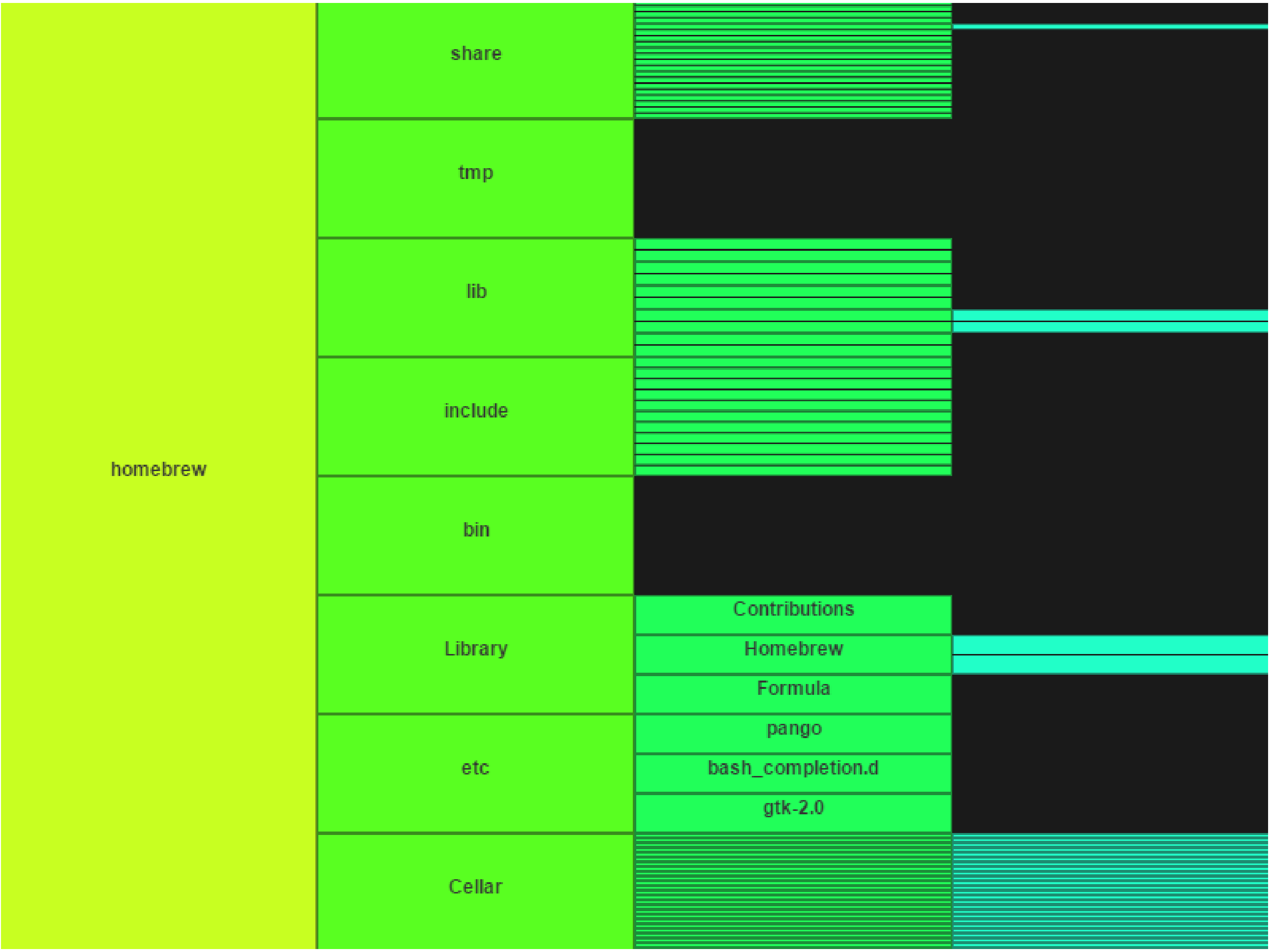
- Enables display of **three or more** quantitative variables in **2D**

- Each **axis** represents one **attribute**



Icicle tree

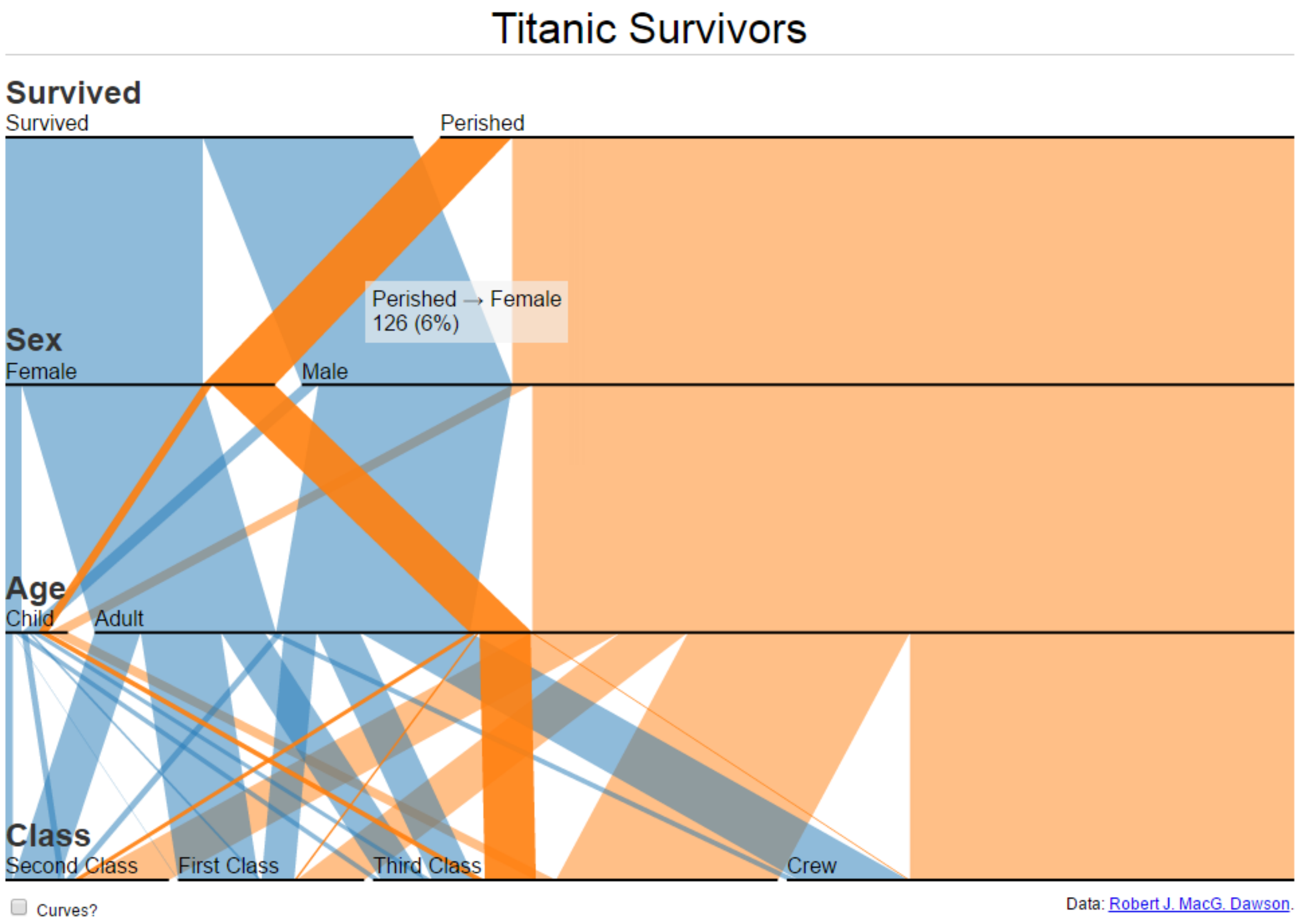
- Visualization of **clusters** during **successive steps** of a **cluster analysis**



source: http://philogb.github.io/jit/static/v20/Jit/Examples/Icicle/example2.html#

Parallel sets

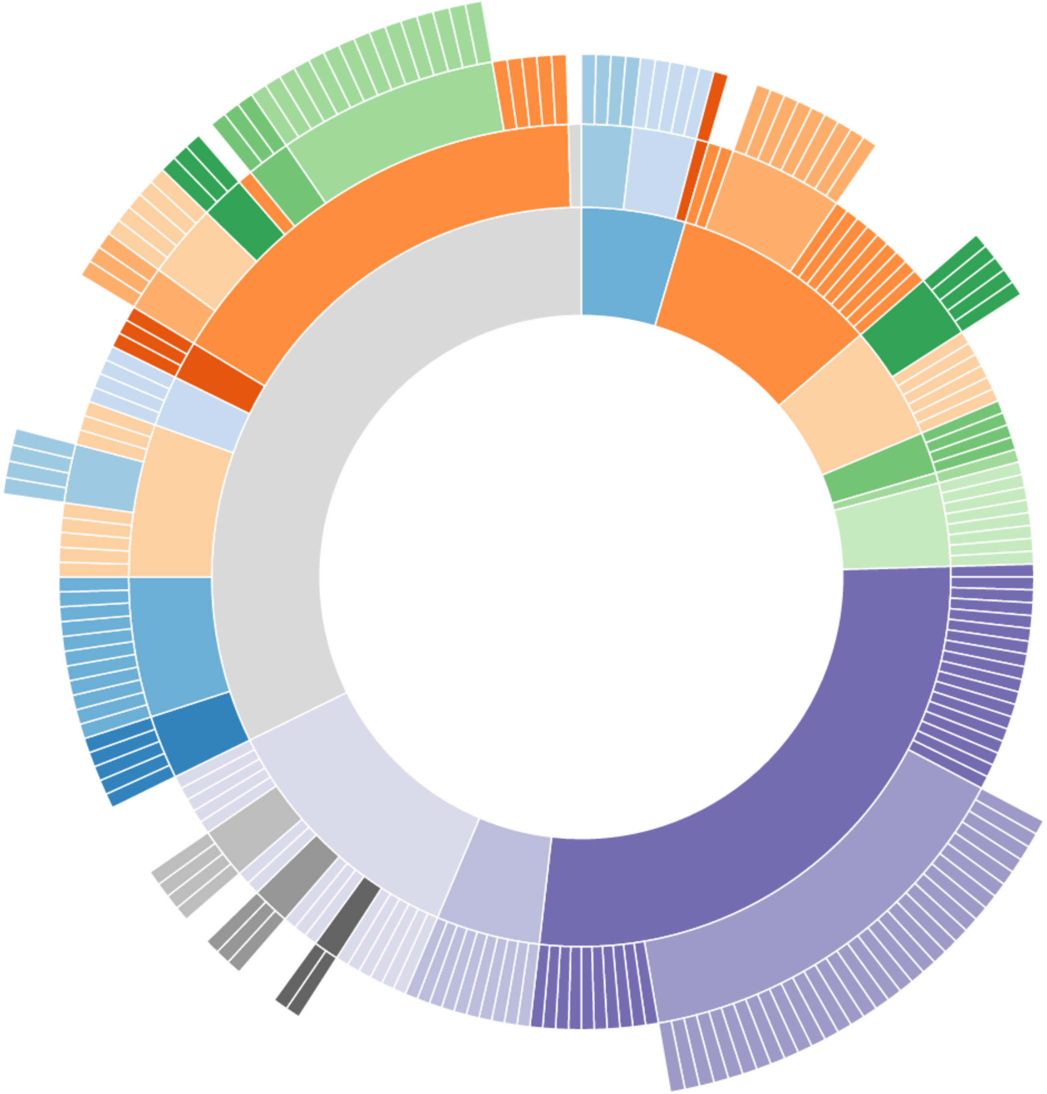
- Repetitive subdivision of categories
- One horizontal line per dimension and category
- Number of matches represented by width of bar
- Interactivity (both vertical and horizontal)



source: https://www.jasondavies.com/parallel-sets/

Sunburst

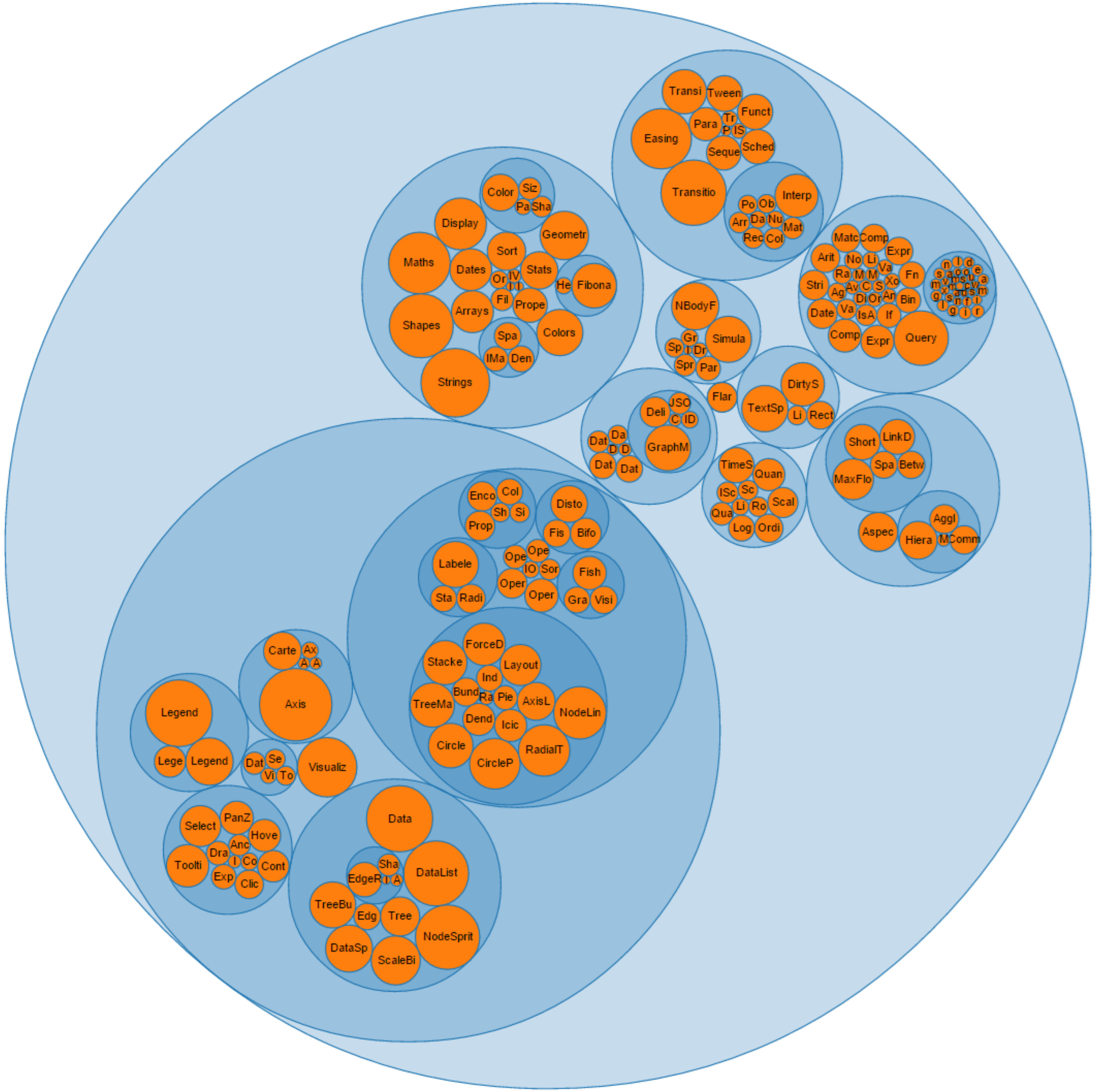
- Inspired by treemap → layout for **tree structures**
- **Root** represents **center** of the plot
- A **shell** corresponds to a **level** in the tree → **leaves** on the **circumference**
- **Area** of arcs correspond to a **value associated with given node**



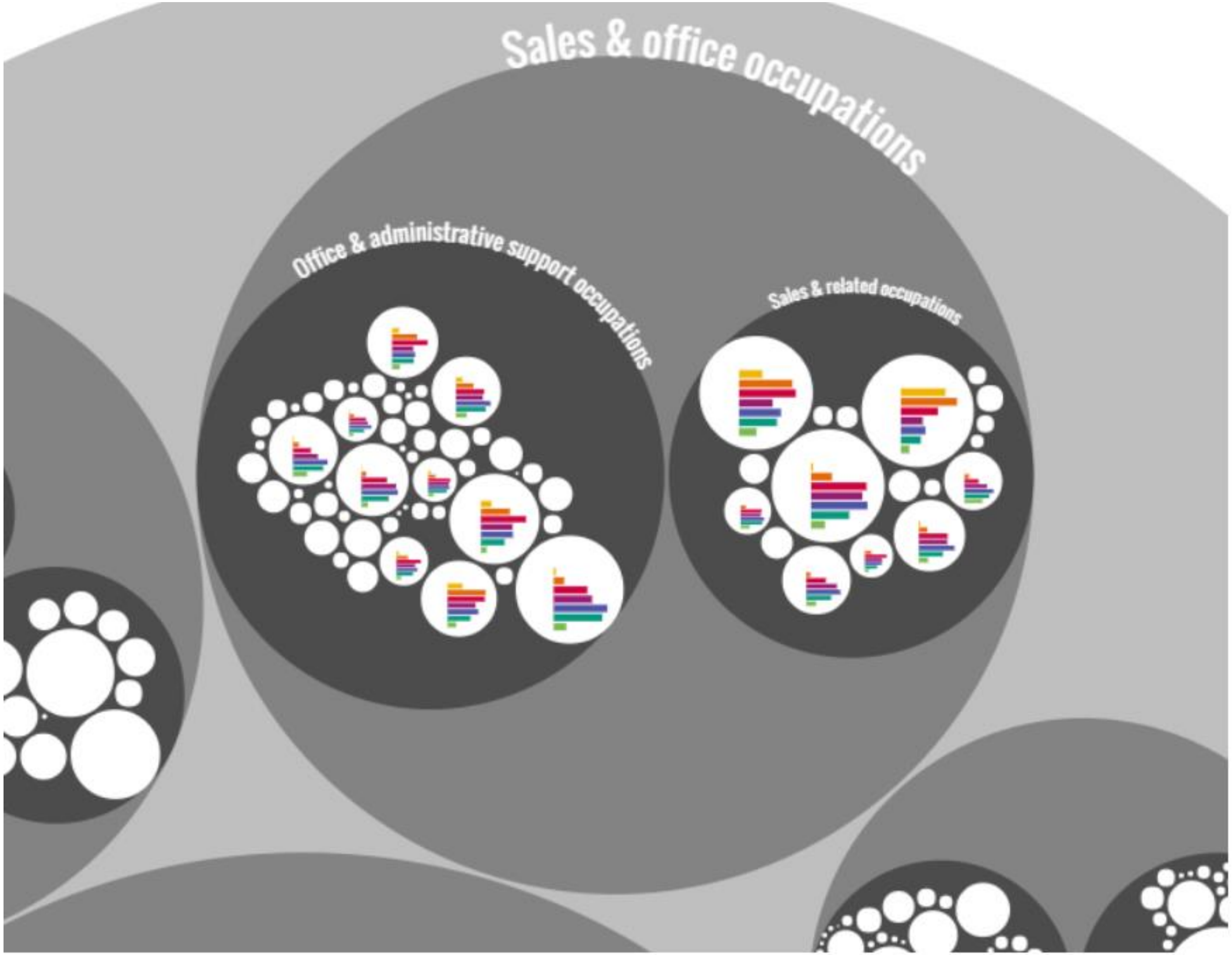
source: http://bl.ocks.org/mbostock/4063423

Circle packing

- Inspired by treemap → layout for **tree structures**
- In general, circle packing is a space filling technique dealing with **arrangement of circles** so that all circles **touch each other but do not overlap**
- **Size** of the circle can represent an **arbitrary property**



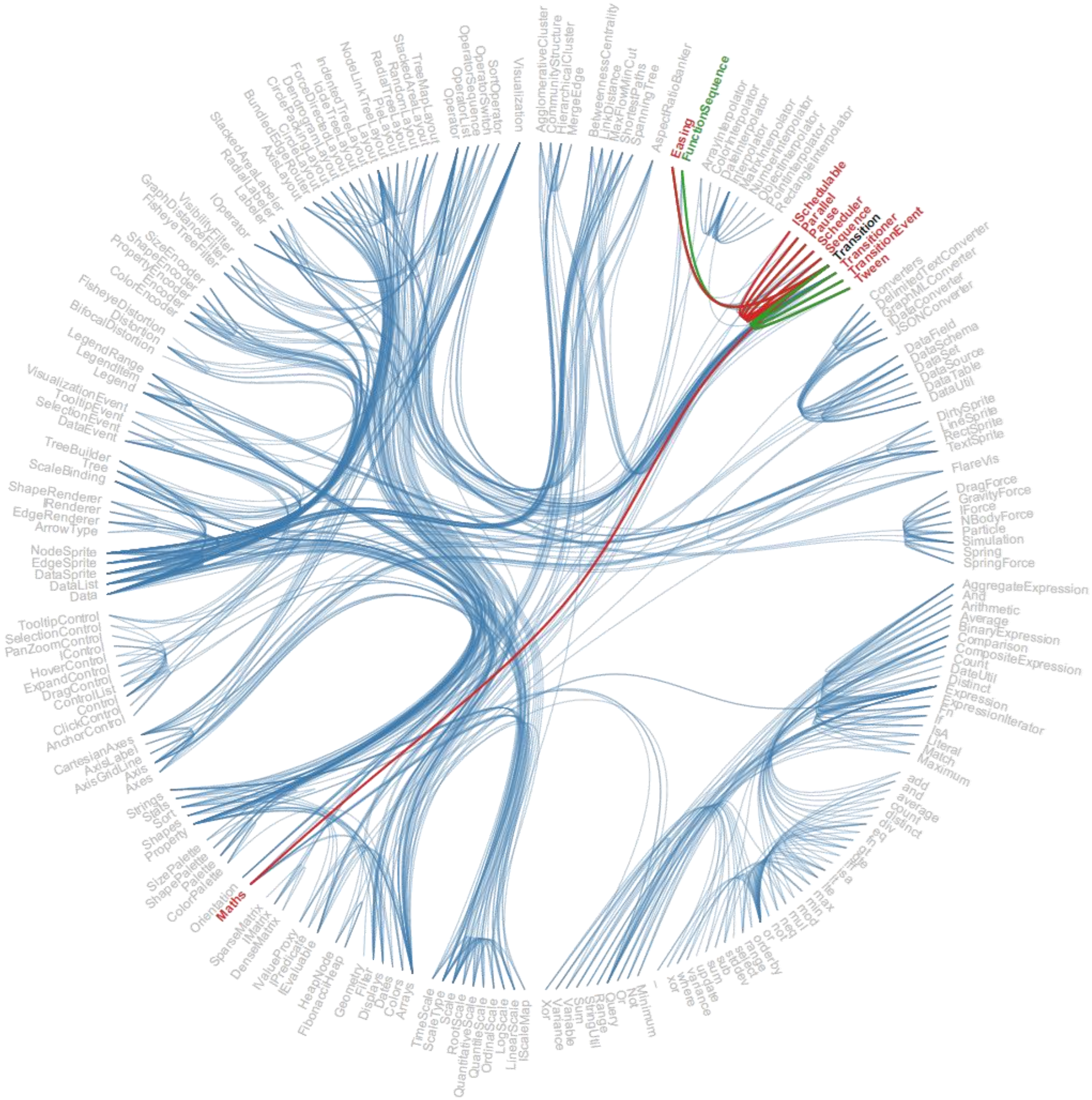
source: http://bl.ocks.org/mbostock/4063530



source: http://www.visualcinnamon.com/occupations

Hierarchical edge bundling

- Basically a **radial chart** including **hierarchical clustering**



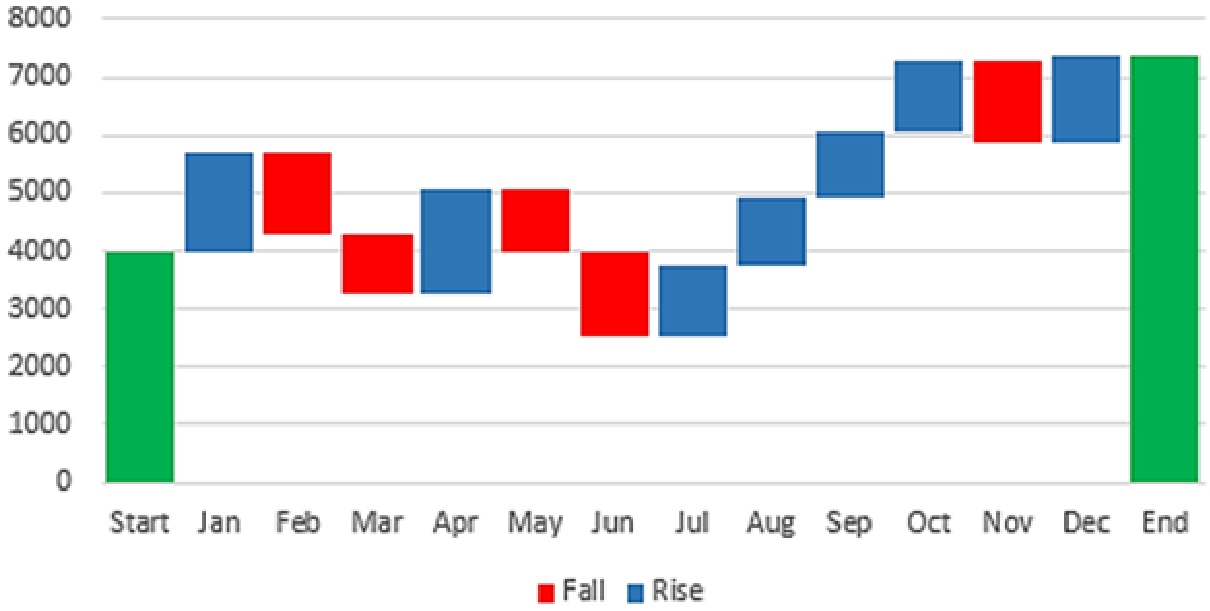
source: http://bl.ocks.org/mbostock/7607999

Waterfall chart

- Also known as **flying bricks chart**

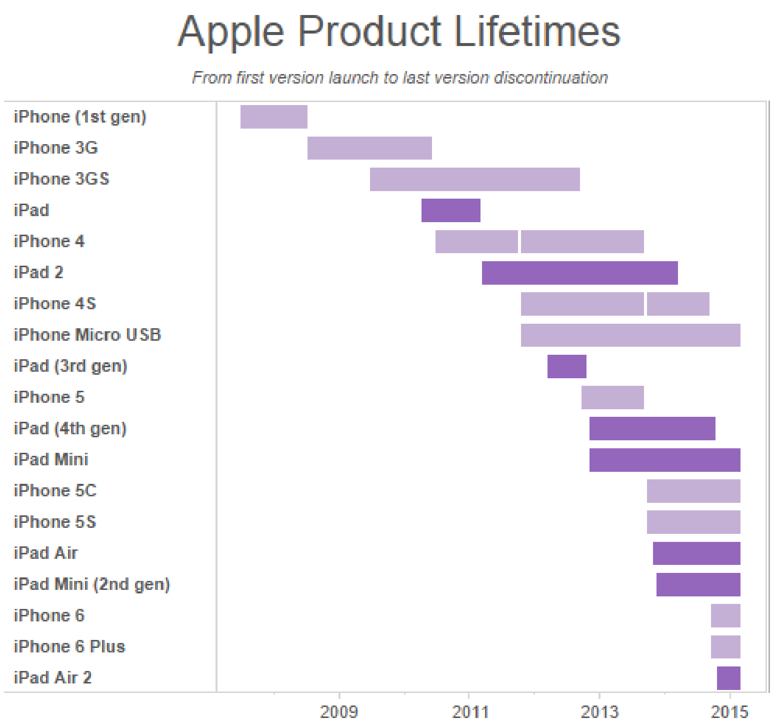
- Display of **gradual** negative or positive **effects** on an **initial** value

- Basically a bar chart

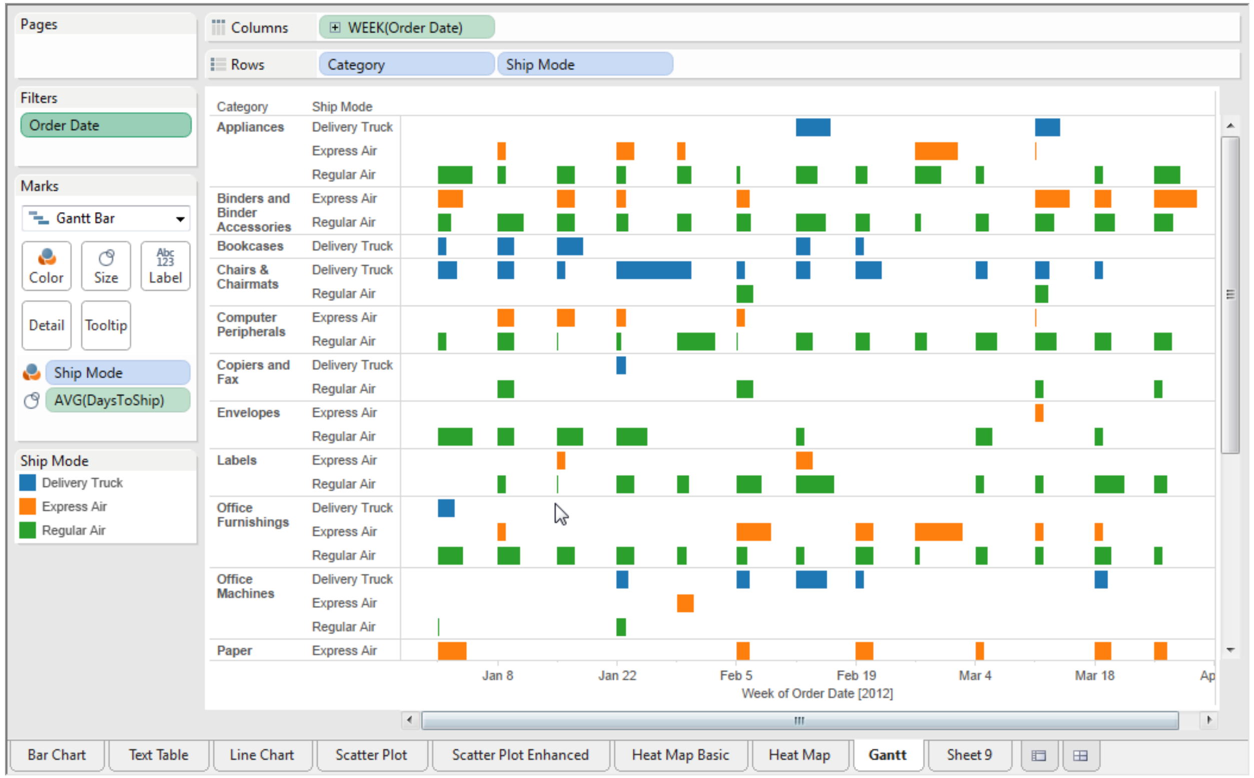


Gantt chart

- Display of the **duration** of events or activities **over time**
- Each separate **mark** (bar) shows a **duration**



source: http://www.tableau.com/about/blog/2014/12/5-chart-types-youve-never-tried-tableau-35281?inline=true#TableauGanttChart



Columns: WEEK(Order Date)

Rows: Category, Ship Mode

Filters: Order Date

Marks: Gantt Bar

Color: Ship Mode

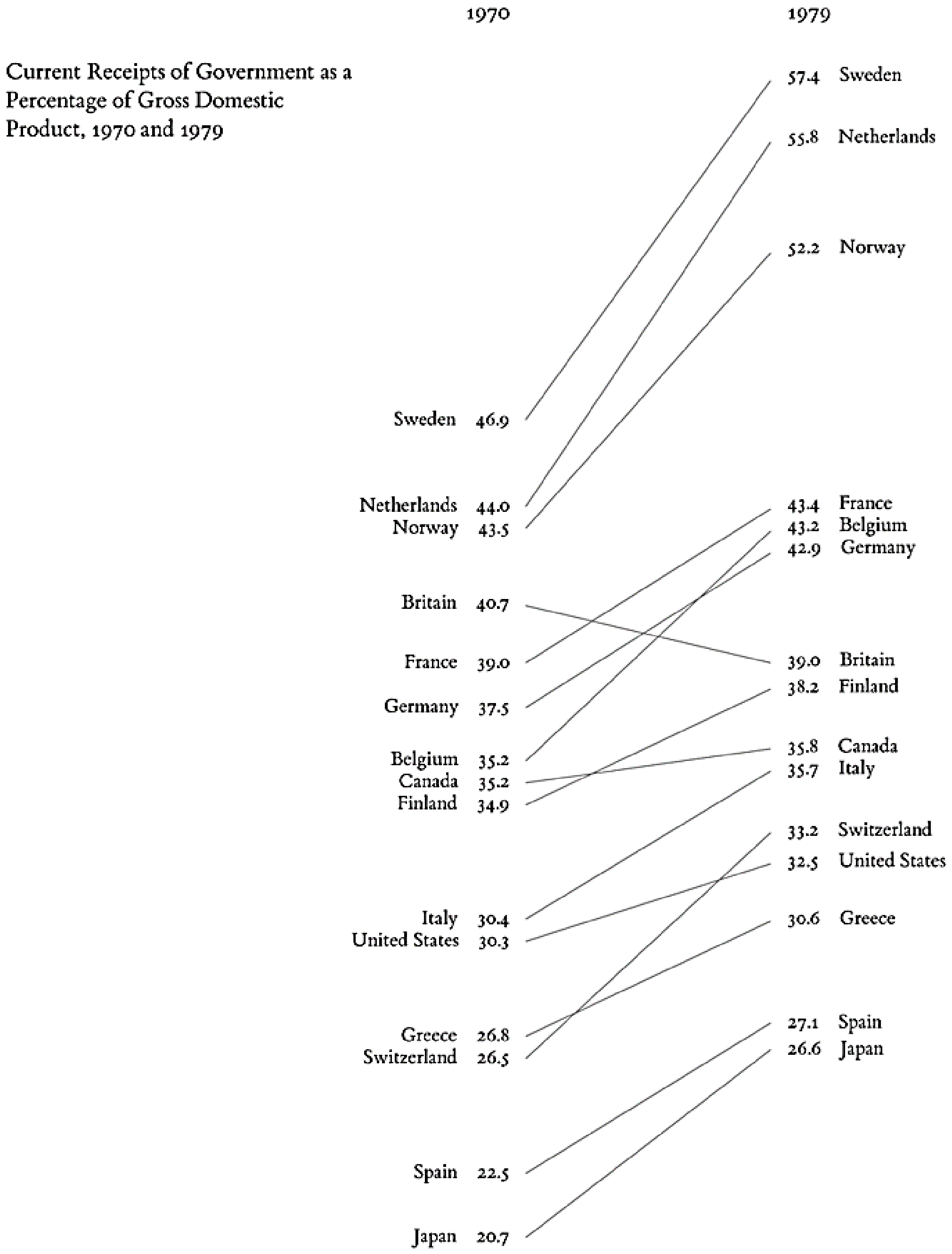
Size: AVG(DaysToShip)

source: http://onlinehelp.tableau.com/current/pro/online/mac/en-us/help.htm#buildexamples_gantt.html

Slopegraph

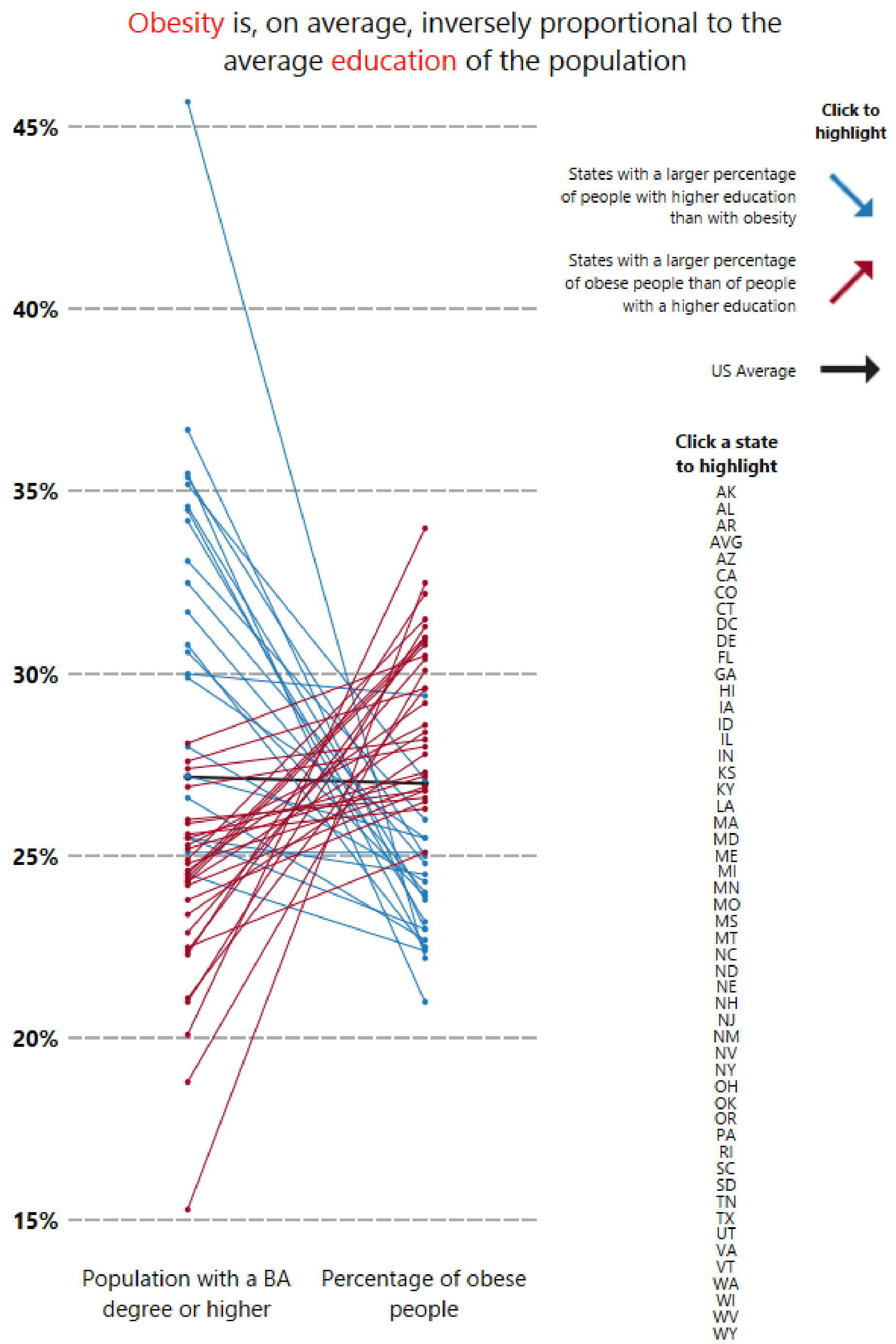
- **Comparison** of **two sets of items** having some relation to each other

- In the original version, slopegraph is basically a line graph where each item has two observations



Current Receipts of Government as a Percentage of Gross Domestic Product, 1970 and 1979

source: Edward Tufte (1983) The visual display of Quantitative Information, Second Edition. Graphics Press



Obesity is, on average, inversely proportional to the average education of the population

Click to highlight

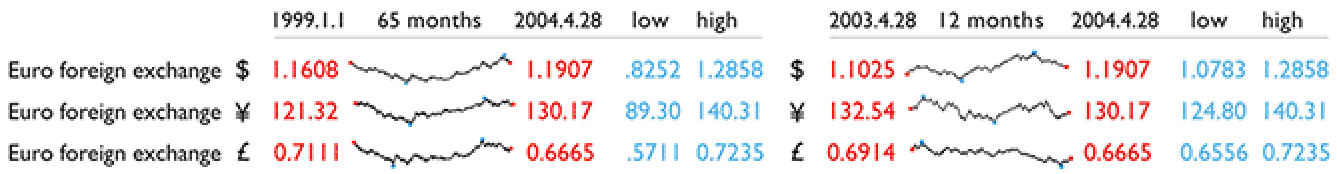
States with a larger percentage of people with higher education than with obesity

States with a larger percentage of obese people than of people with a higher education

US Average

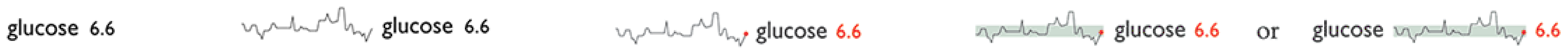
source: http://vizwiz.blogspot.cz/2013/01/alberto-cairo-three-steps-to-become.html

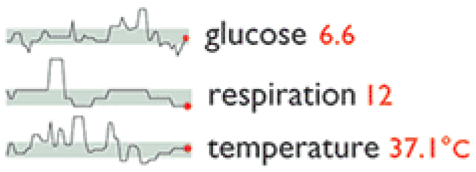
Sparklines



		1999.1.1	65 months	2004.4.28	low	high		2003.4.28	12 months	2004.4.28	low	high
Euro foreign exchange	$	1.1608		1.1907	.8252	1.2858	$	1.1025		1.1907	1.0783	1.2858
Euro foreign exchange	¥	121.32		130.17	89.30	140.31	¥	132.54		130.17	124.80	140.31
Euro foreign exchange	£	0.7111		0.6665	.5711	0.7235	£	0.6914		0.6665	0.6556	0.7235

- Small line chart goal of which is to capture general shape (over time) of a measurement (reading of an instrument)
- Small, high-resolution graphics, usually embedded in a full context of words, numbers, images → ***datawords*** (**data-intense, design-simple, word-sized graphics**)





source: http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001OR&topic_id=1

Tag cloud

- Also knows as **word cloud** or **weighted list**
- Text analysis visualization of word frequencies
  - How frequently words appear in a given text reflects in its size
  - Inner structure can be revealed with other visual attributes such as color (e.g., to differentiate groups of words)

Literature

- Stephen Few (2012) Show me the numbers – Designing Graphs and Tables to Enlighten

- Edward Tufte (2001) The visual display of Quantitative Information, Second Edition. Graphics Press

- Gene Zelazny (2001) Say It with charts