

S. Rinzivillo – rinzivillo@isti.cnr.it

DATA VISUALIZATION AND VISUAL ANALYTICS

TAXONOMY OF VISUAL VARIABLES

Cleveland McGill [1984]

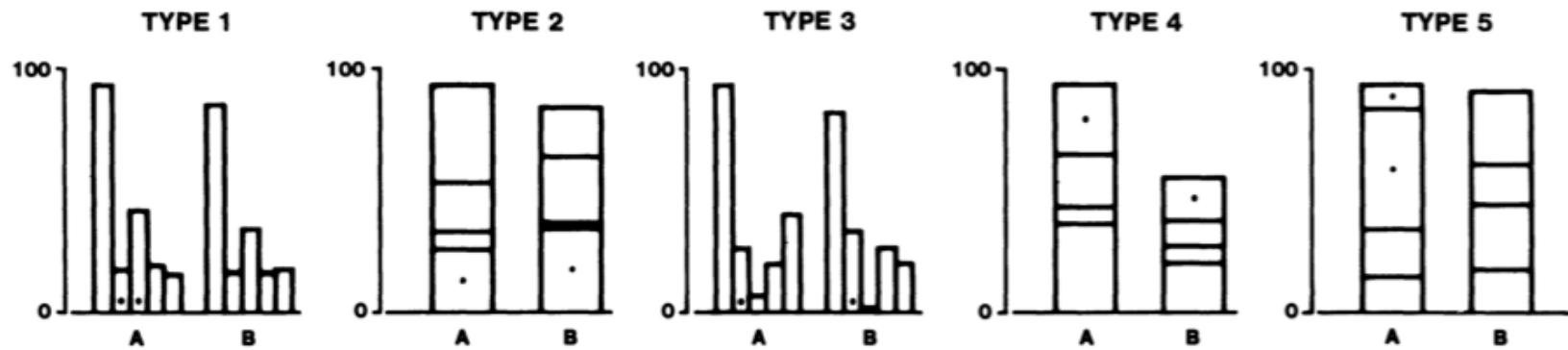


Figure 4. Graphs from position-length experiment.

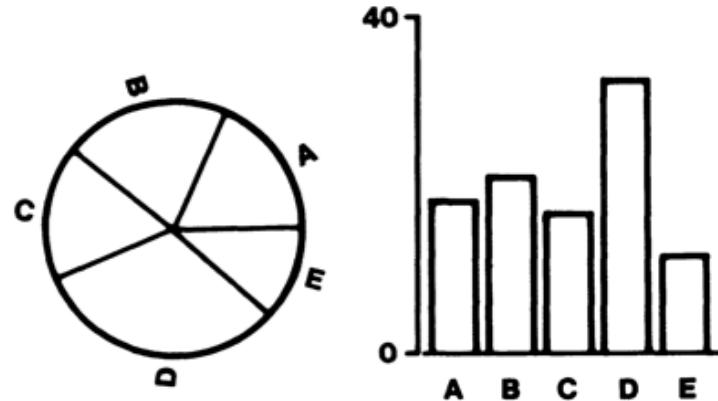
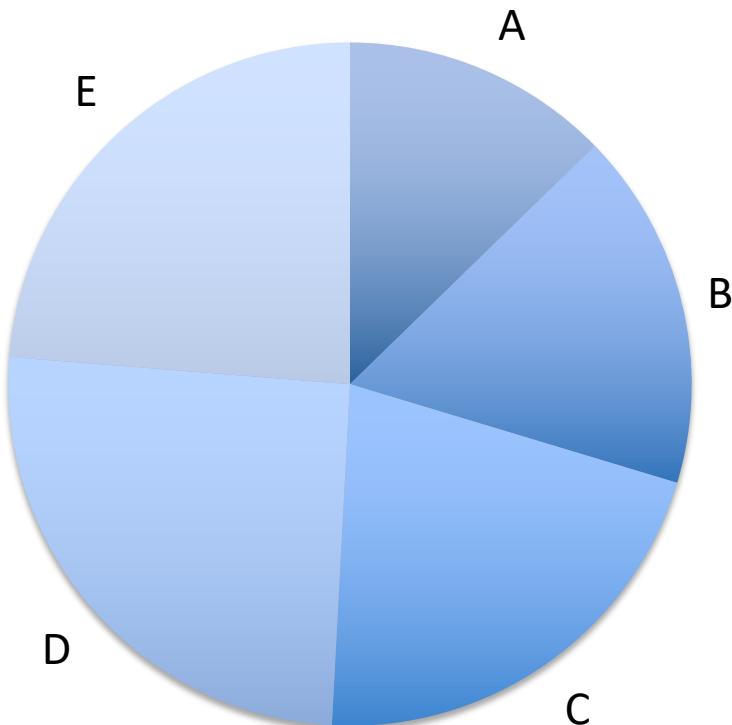


Figure 3. Graphs from position-angle experiment.

Cleveland & McGill: graphical encodings

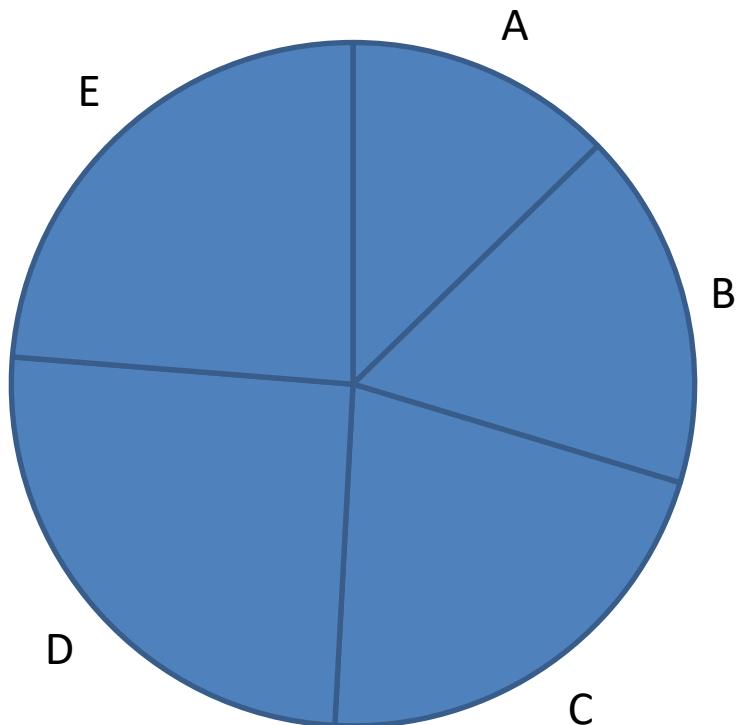
- Angle
- Area
- Color Hue
- Color Saturation
- Density
- Length
- Position on a common scale
- Position on non aligned scale
- Slope
- Volume

Angle decoding



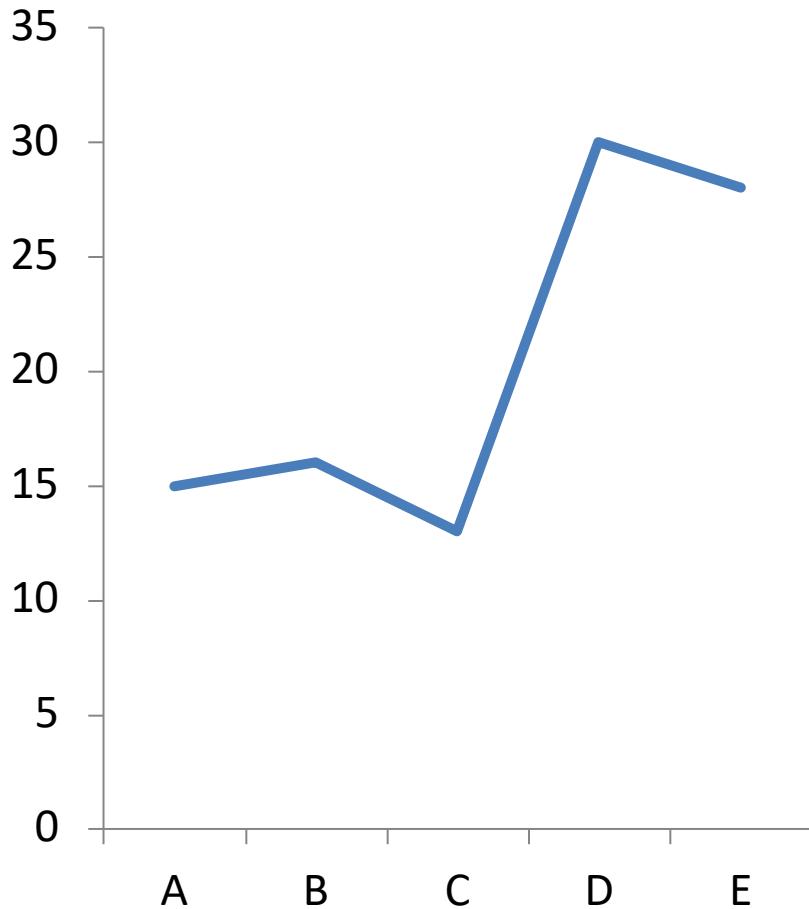
- It is difficult to compare angles
 - Underestimation of acute angles
 - Overestimation of obtuse angles
 - Easier if bisectors are aligned
- Area estimation helps

Angle decoding



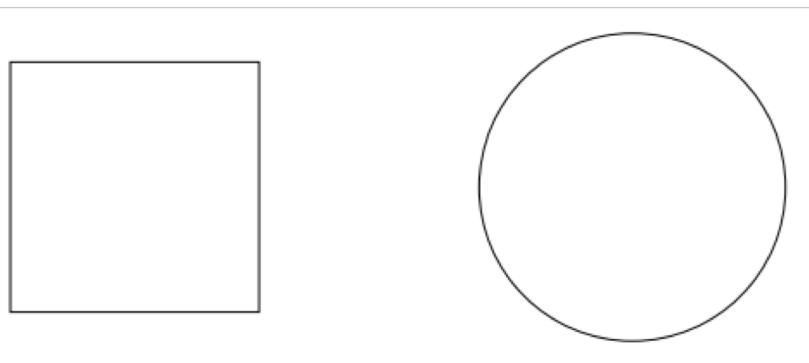
- It is difficult to compare angles
 - Underestimation of acute angles
 - Overestimation of obtuse angles
 - Easier if bisectors are aligned

Slopes Decoding



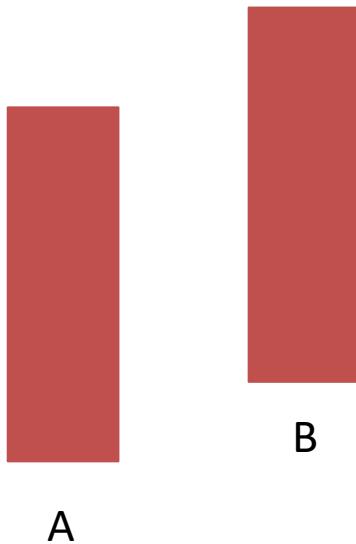
- Same difficulties as angles
- Easier task since one branch is aligned with x-axis

Area Decoding



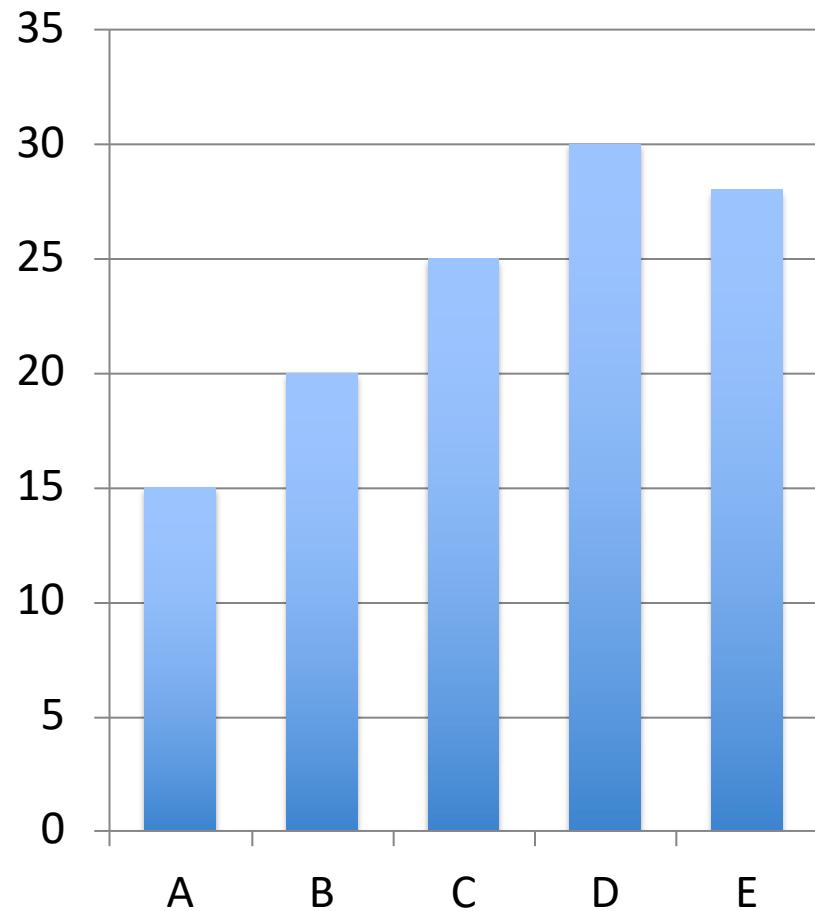
- Area is not well decoded
 - Different regular shapes
 - Irregular shapes
 - Context influences (thin area within compact thick area)

Length Decoding



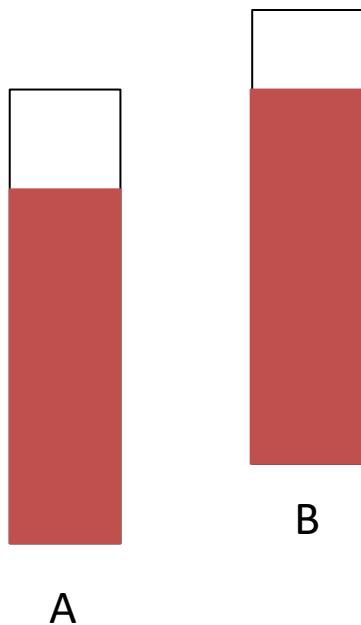
- Straight forward to encode numerical values
- Difficulties with relative lengths

Position on a common scale



- Widely used in statistical charts

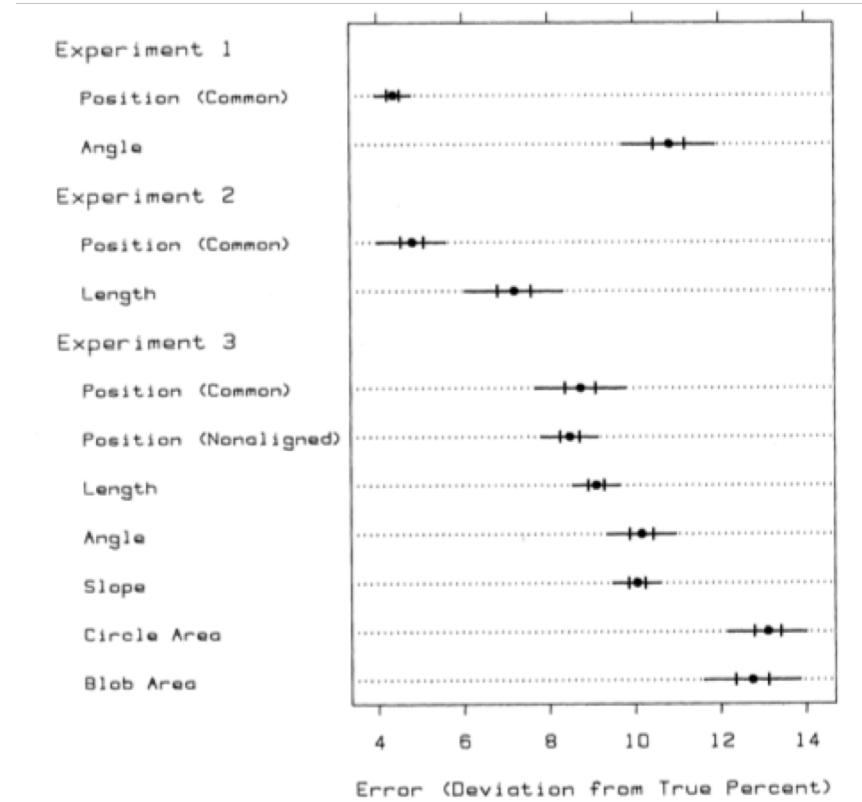
Position on non-aligned scale

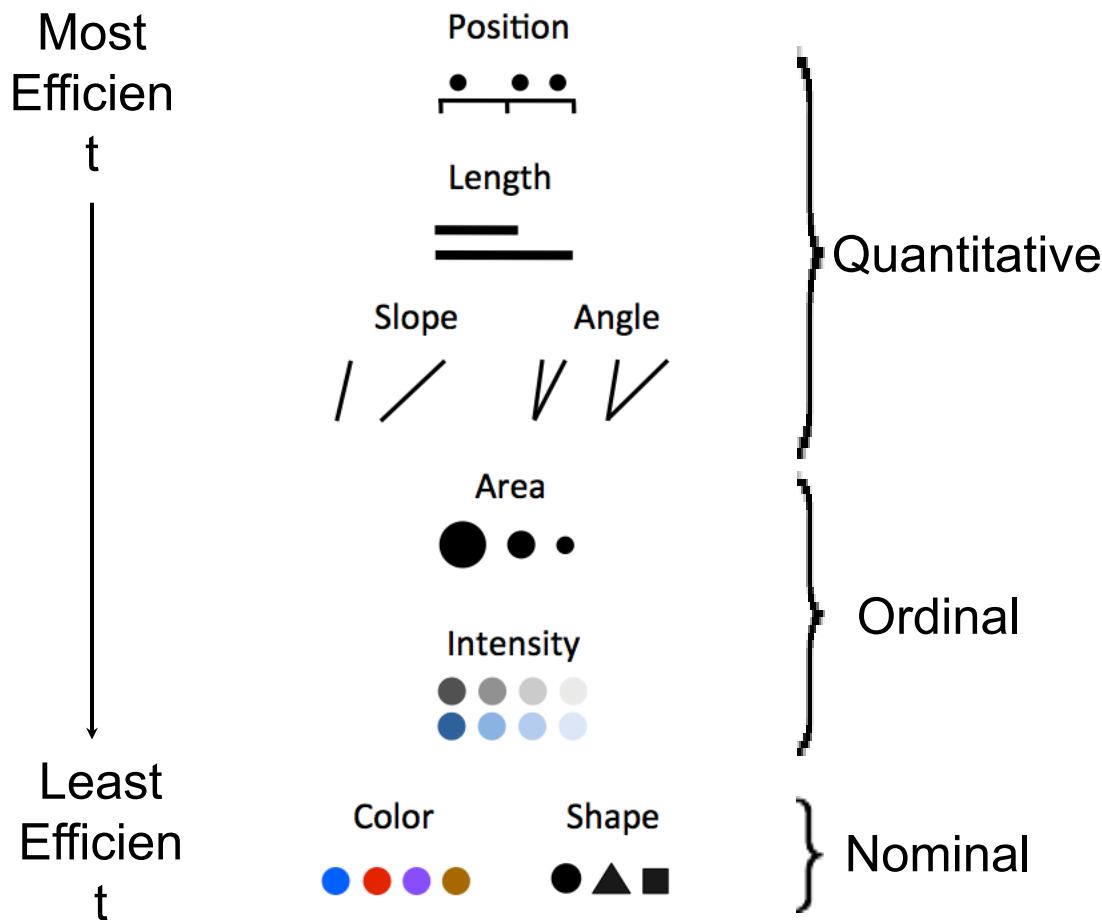


- Not as bad as common scale
- Still acceptable

Designing Effective Visualizations

- If possible, use graphical encoding that are easily decoded
- Graphical Attributes ordered(Cleveland & McGill):
 - Position along a common scale
 - Position on non aligned scales
 - Length
 - Angle and Slope
 - Area
 - Volume, density, color saturation
 - Color Hue





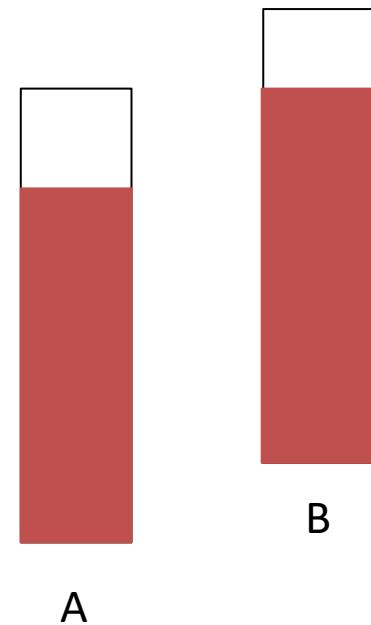
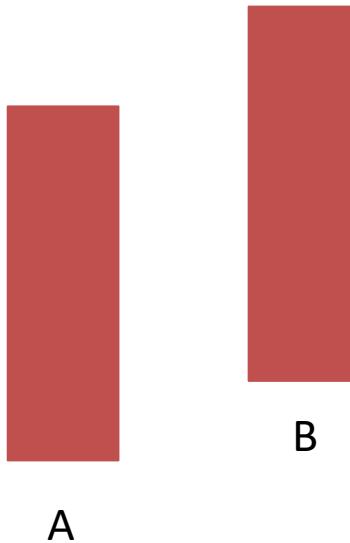
PERCEPTION LAWS

Weber's Law

- Just-noticeable difference between two stimuli is proportional to their magnitudes
- Case study on length
 - Given two lines with lengths x and $x+w$
 - If w is small, it is difficult to notice difference between the two lines
 - If w is larger, it is easier to catch the difference
- How large should w be?
 - The probability of detecting the change is proportional to the relative value w/x

Weber's Law

- Given values (90, 92)
- Detect with probability of 2/90
- Given values(90,92)
- Detect with probability of 2/10

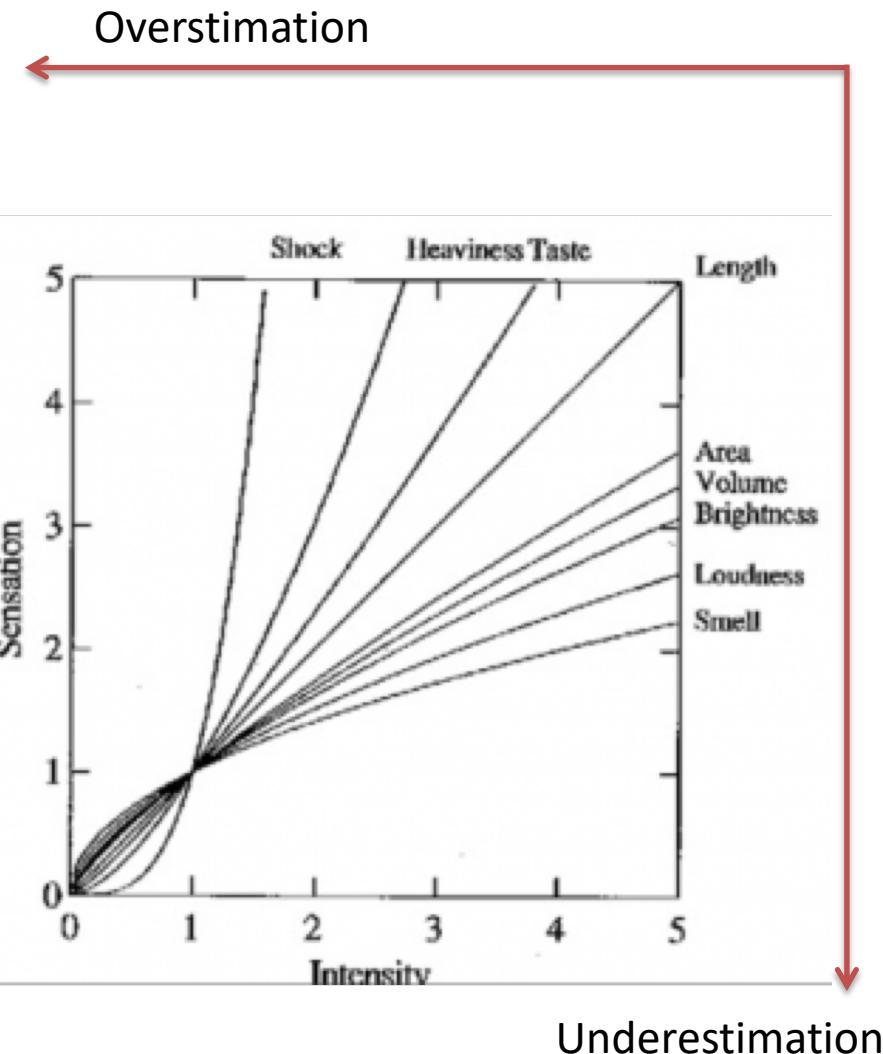


Stevens' Law

- Model the relation between a stimulus and its perceived intensity
- Given a stimulus x encoded with a visual attribute
- An observer decode a perceived value $p(x)$
- Stevens' law states that
 - $p(x) = kx^\beta$
 - where k is constant and
 - β is a constant that depends on the nature of stimulus

Stevens' law

- Better effectiveness when $p(x) = kx^\beta$ is linear
- Linearity depends only on β
- Different visual encodings yields typical ranges for β
 - Lengths: 0.9 – 1.1
 - Area: 0.6 – 0.9
 - Volume: 0.5 – 0.8



Weber and Stevens' Laws

- Given two values x_1 and x_2
- Let the perceived values be $p(x_1)$ and $p(x_2)$

$$\frac{p(x_1)}{p(x_2)} = \left(\frac{x_1}{x_2} \right)^\beta$$

Weber and Stevens' Laws: areas

- For areas $\beta=0.7$
- Let $x_1=2$ and $x_2=1$
- The perceived difference will be

$$\frac{p(2)}{p(1)} = \left(\frac{2}{1}\right)^{0.7} = 1,6245$$

- For areas $\beta=0.7$
- Let $x_1=0,5$ and $x_2=1$
- The perceived difference will be

$$\frac{p(\frac{1}{2})}{p(1)} = \left(\frac{\frac{1}{2}}{1}\right)^{0.7} = 0,6155$$

Weber and Stevens' Laws: areas vs lengths

- For areas $\beta=0.7$
- Let $x_2=x_1+w$
- The perceived difference will be

$$\left(\frac{x+w}{x}\right)^{0.7} \approx 1 + \frac{0.7w}{x}$$

- For lengths $\beta=1$
- Let $x_2=x_1+w$
- The perceived difference will be

$$\left(\frac{x+w}{x}\right)^1 = 1 + \frac{w}{x}$$

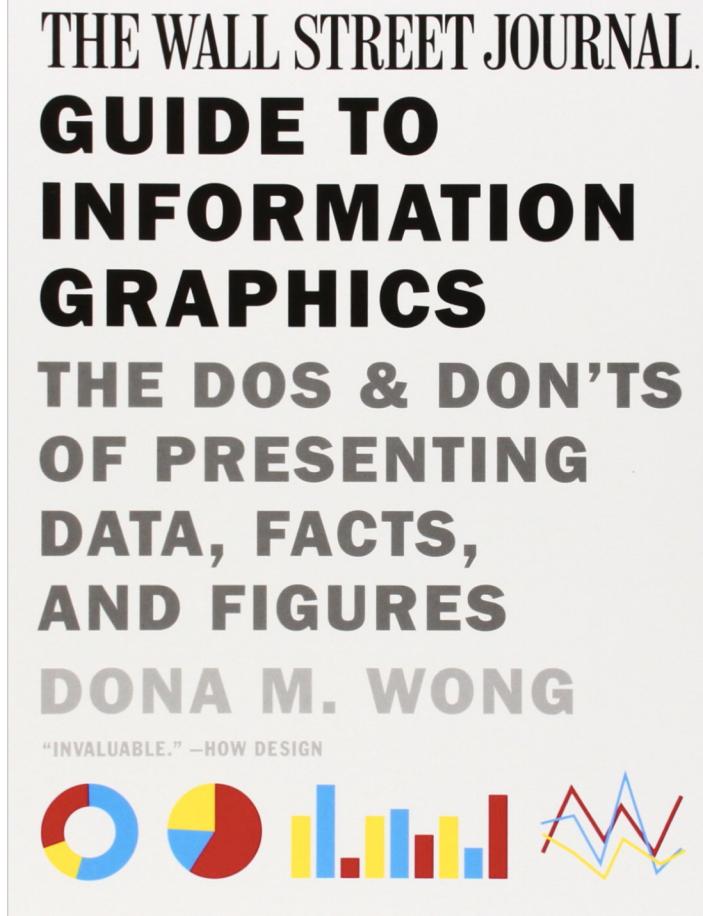
Takeaway messages

- Data type for entities and relationships
- Visual variables for representation
- Mapping of types to VVs
- Some VVs are more appropriate for specific data types

Visual Analytics

Dos and Don'ts for visual charts

Crash course on effective Charting



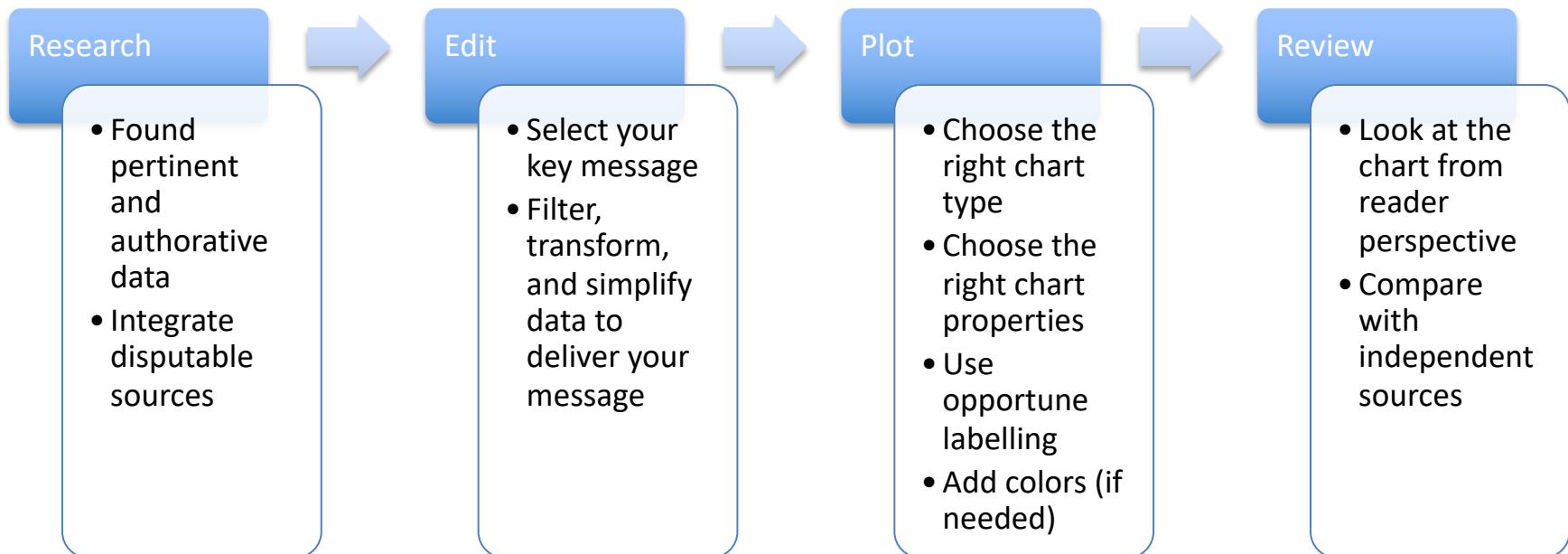
Dona M. Wong

Guide to Information Graphics

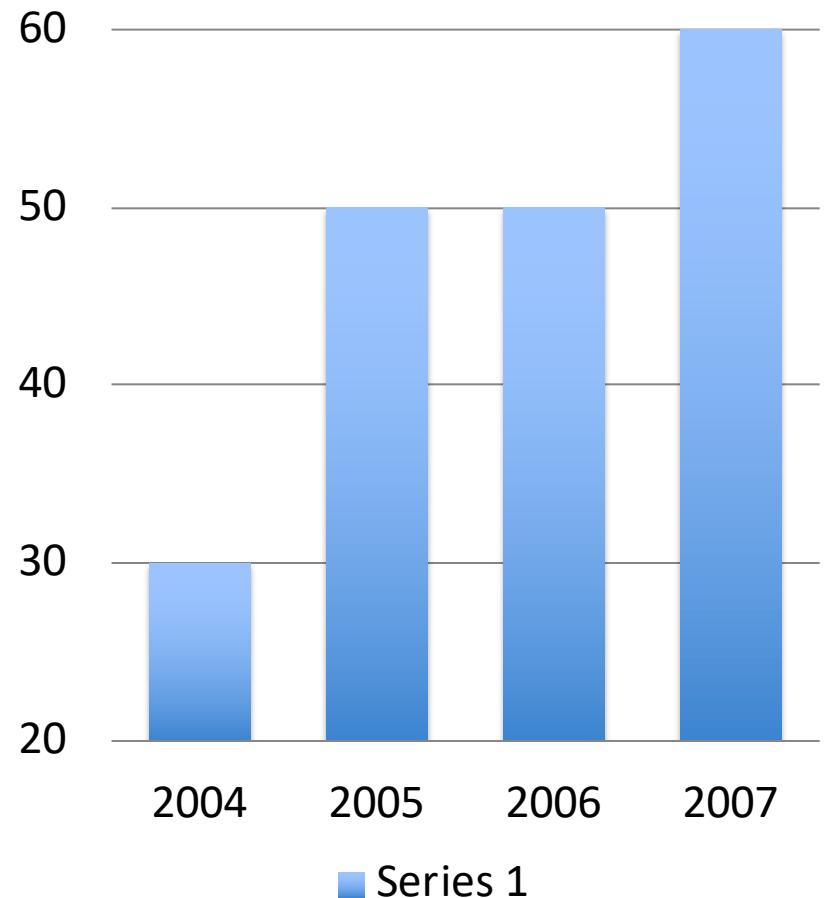
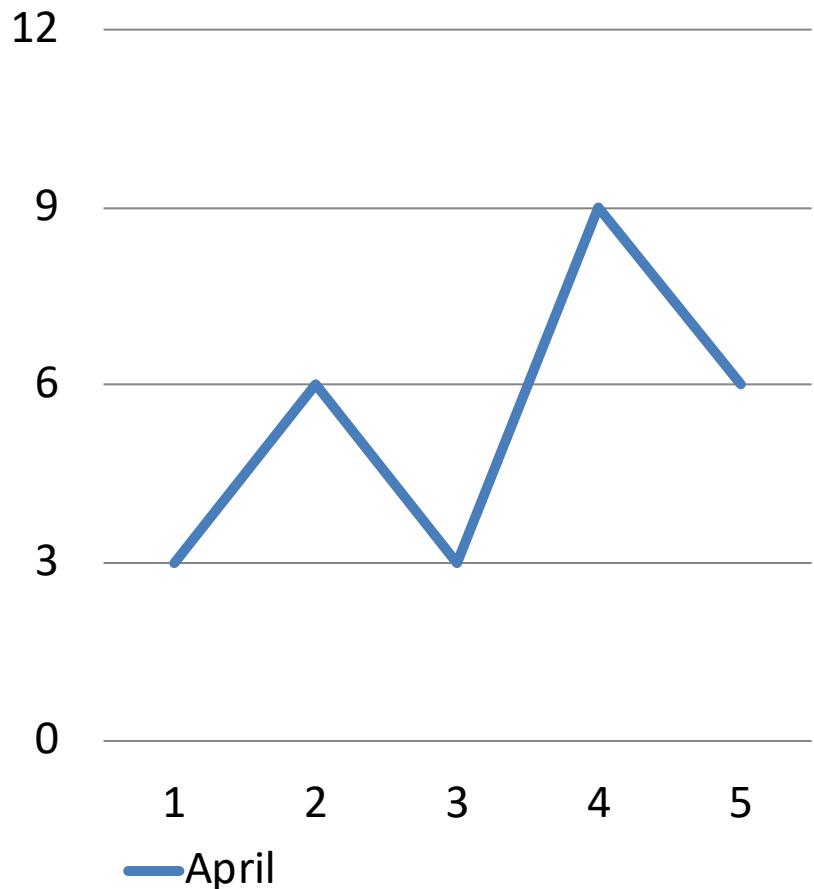
The Dos and Don'ts of Presenting Data,
Facts, and Figures

W. W. Norton & Company

Charting Pipeline

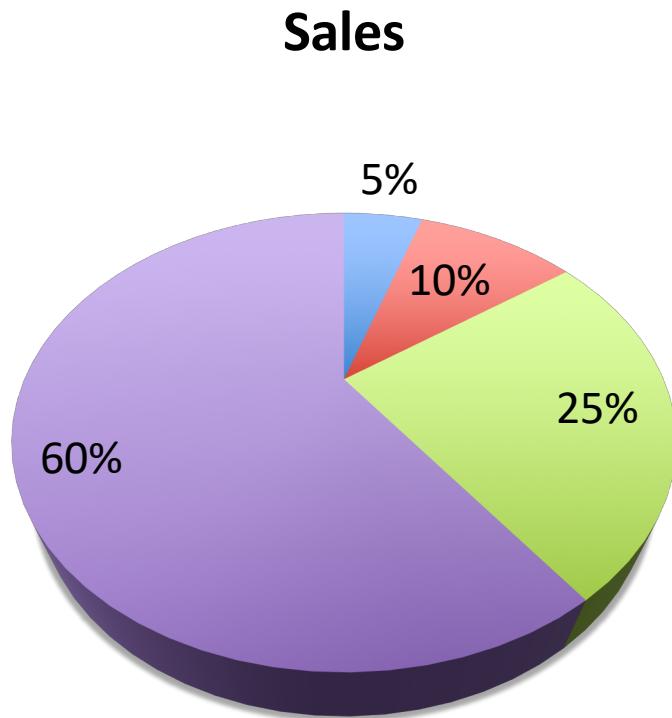
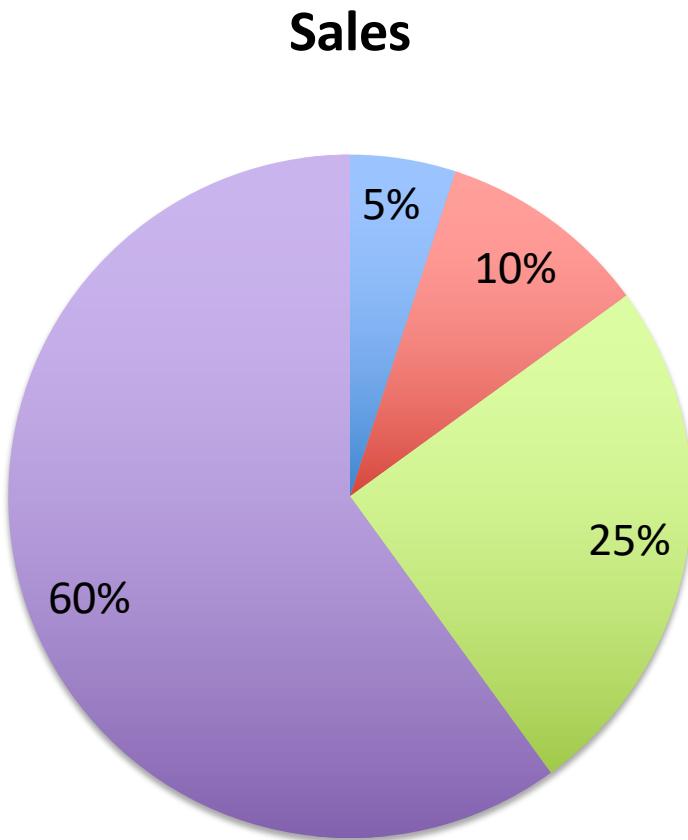


Charting Examples



May these charts be improved? Why? How?

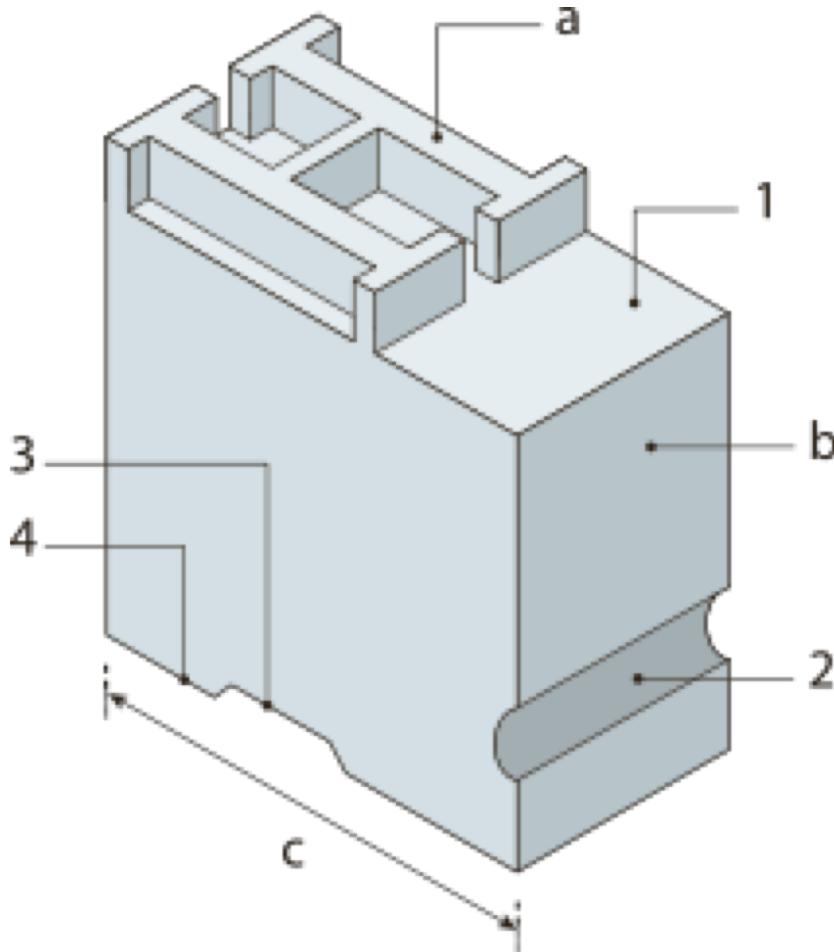
Charting Examples



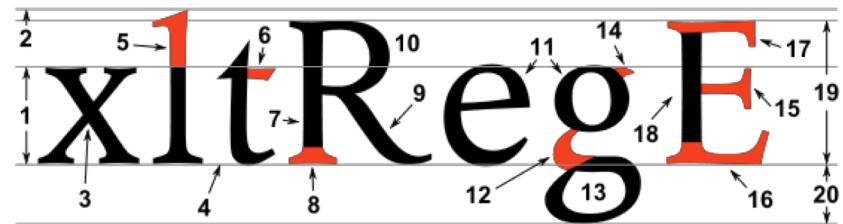
May these charts be improved? Why? How?

FONTS

Fonts



"Metal type". Licensed under Public Domain via Wikimedia Commons -
http://commons.wikimedia.org/wiki/File:Metal_type.svg#mediaviewer/File:Metal_type.svg



Typographic parts of a glyph:
1) x-height; 2) ascender line; 3) apex; 4) baseline; 5) ascender; 6) crossbar; 7) stem; 8) serif; 9) leg; 10) bowl; 11) counter; 12) collar; 13) loop; 14) ear; 15) tie; 16) horizontal bar; 17) arm; 18) vertical bar; 19) cap height; 20) descender line.

$$\begin{aligned}\text{Font size} &= (1) + (2) + (20) \\ &= (19) + (20)\end{aligned}$$

"Typoghaphia" by F l a n k e r (typographic font designed by myself, named Imperator). Licensed under Public Domain via Wikimedia Commons -
<http://commons.wikimedia.org/wiki/File:Typoghaphia.svg#mediaviewer/File:Typoghaphia.svg>

Fonts: general rules

- Leading should be 2 points larger then type size
- Avoid too small or condensed type faces
- Keep style simple: use **bold** or *italic* to emphasize a word (better not **both**)
- Avoid ALL CAPS
- Avoid *styled fonts*
- Avoid C***C Sans Serif
- Reduce type at an angle
- Avoid tracking

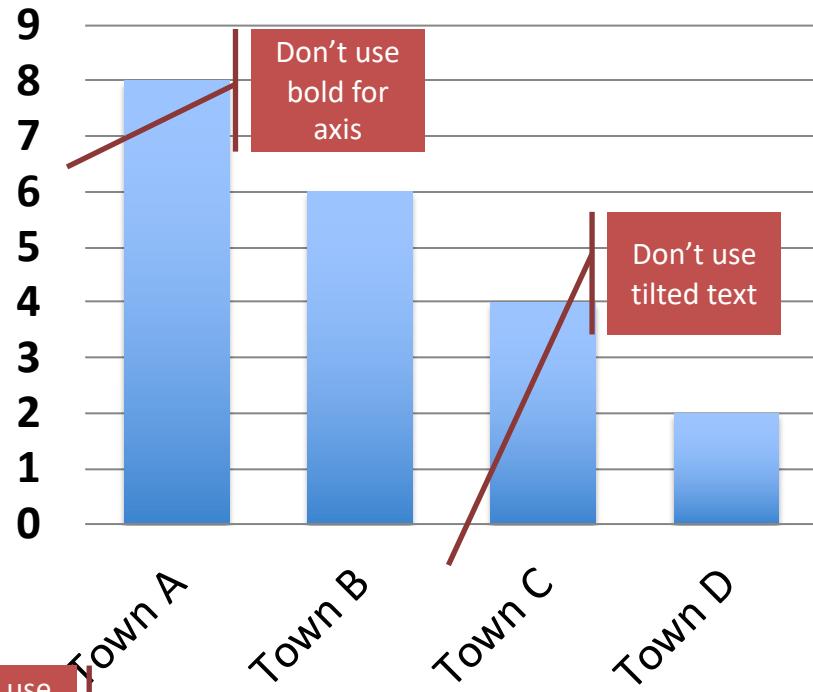
Fonts are meant to describe, not to adorn

Typography in Charts

Don't

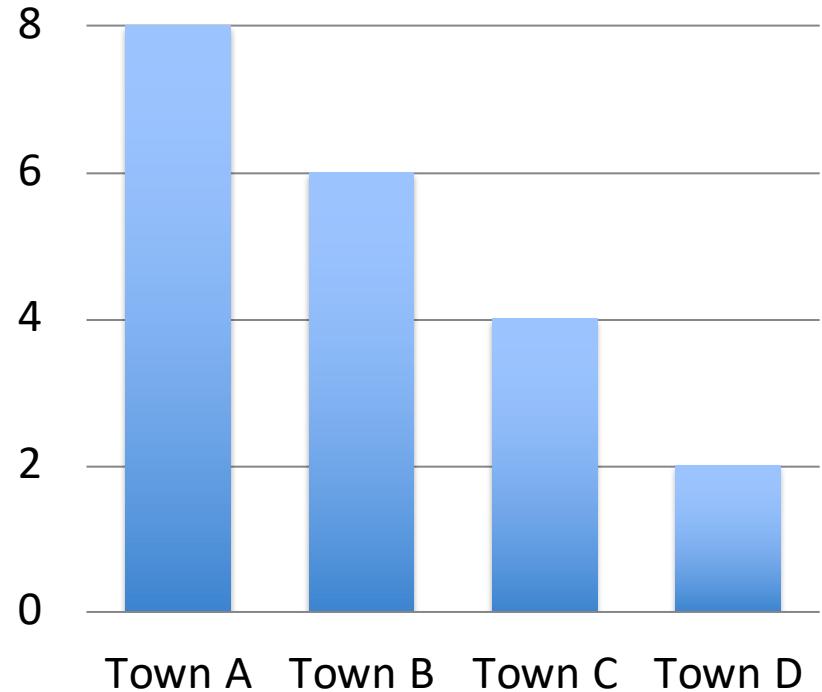
Don't use all caps or
high contrast white
type out of black

HEADLINE OF THE CHART



Do

Headline of the chart



Don't use
bold and
italic

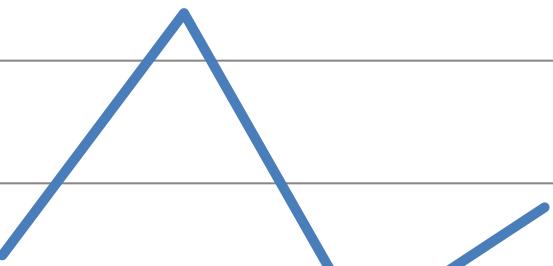
***A brief description that outlines
what the data shows***

Typography in Charts

Don't

Headline of the chart

Title of y-axis

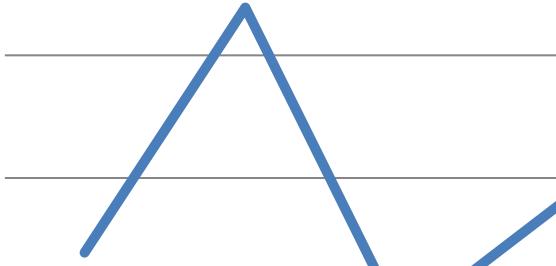


Title of x-axis

Do

Headline of the chart

Title of
y-axis



Title of x-axis

Typography in Charts

Name	Data	Data	Data
Company A	0.0	0.0	0.0
Company B	0.0	0.0	0.0
Company C	0.0	0.0	0.0
Company D	0.0	0.0	0.0

Name	Data	Data	Data
Company A	0.0	0.0	0.0
Company B	0.0	0.0	0.0
Company C	0.0	0.0	0.0
Company D	0.0	0.0	0.0

Many elements in bold. Which part is highlighted?



Give emphasis to relevant results



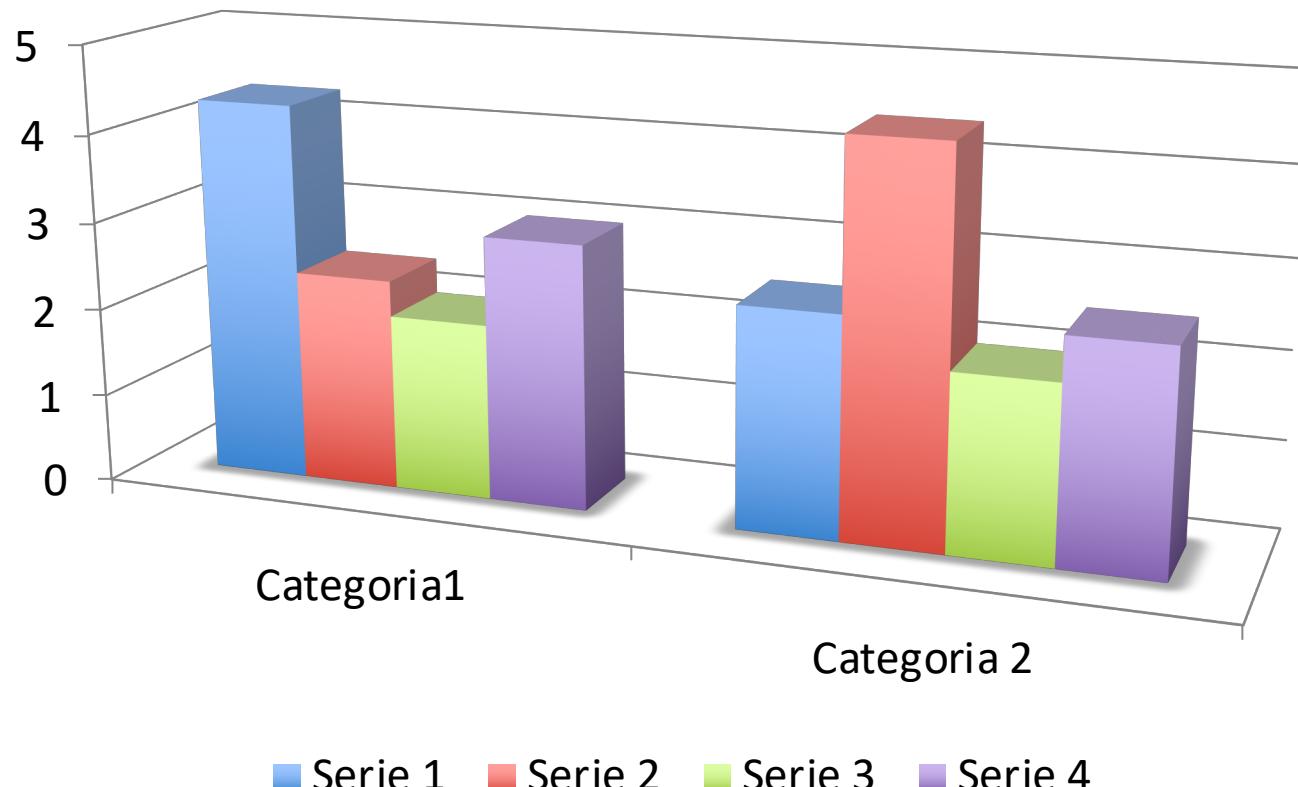


Visual Display of Quantitative Data
Edward Tufte, 1983

DATA-INK RATIO

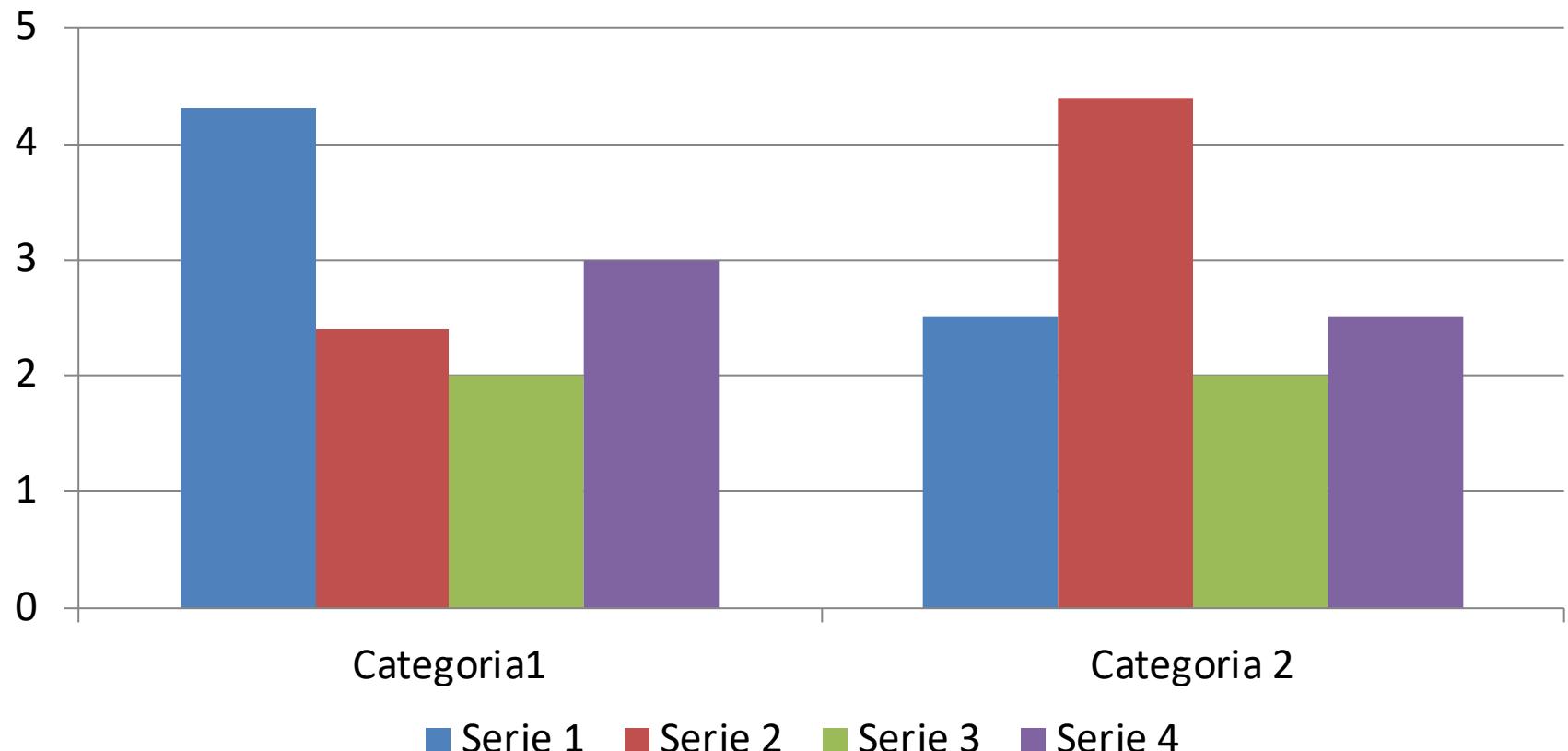
Data-ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



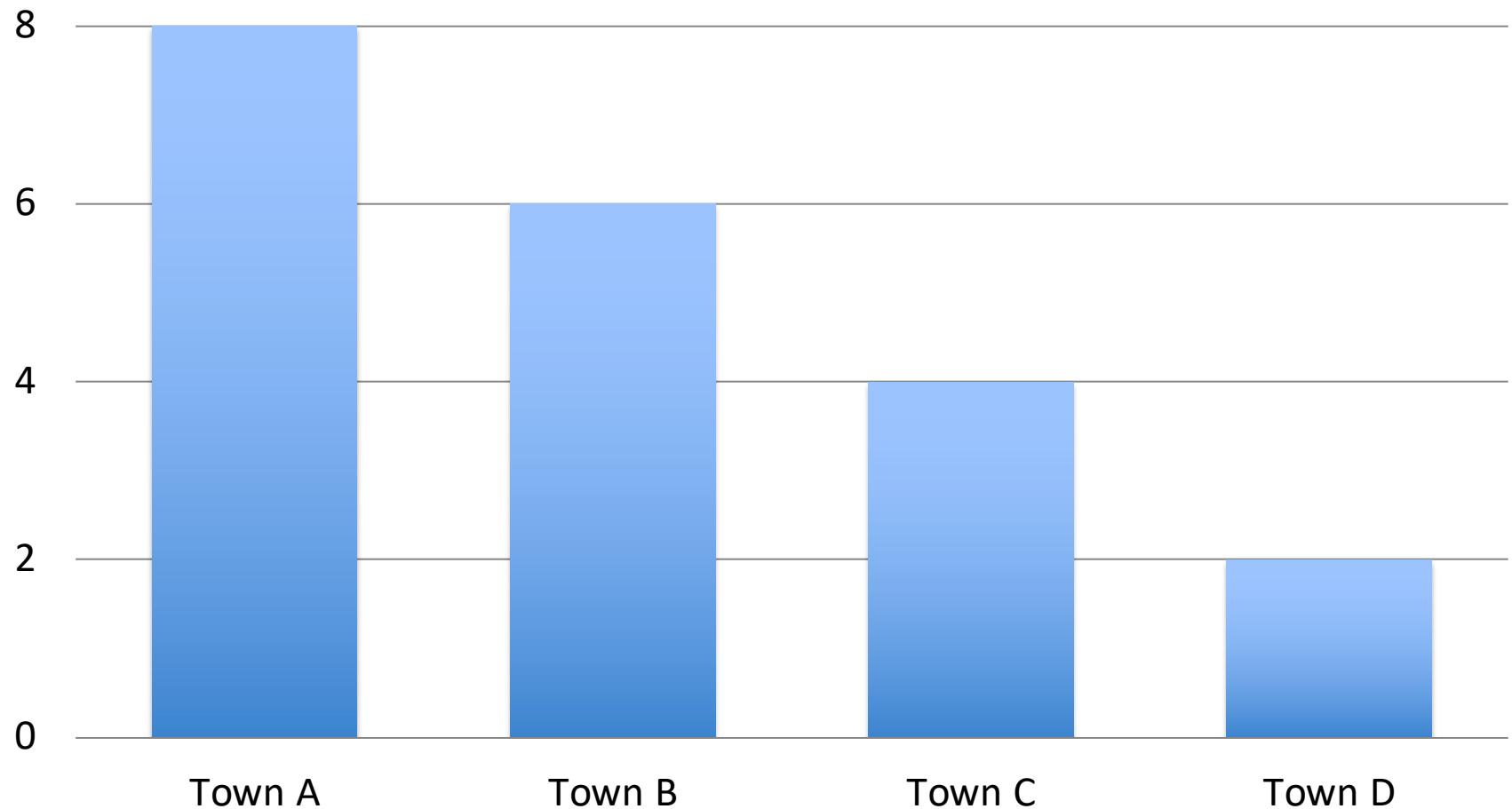
Data-ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



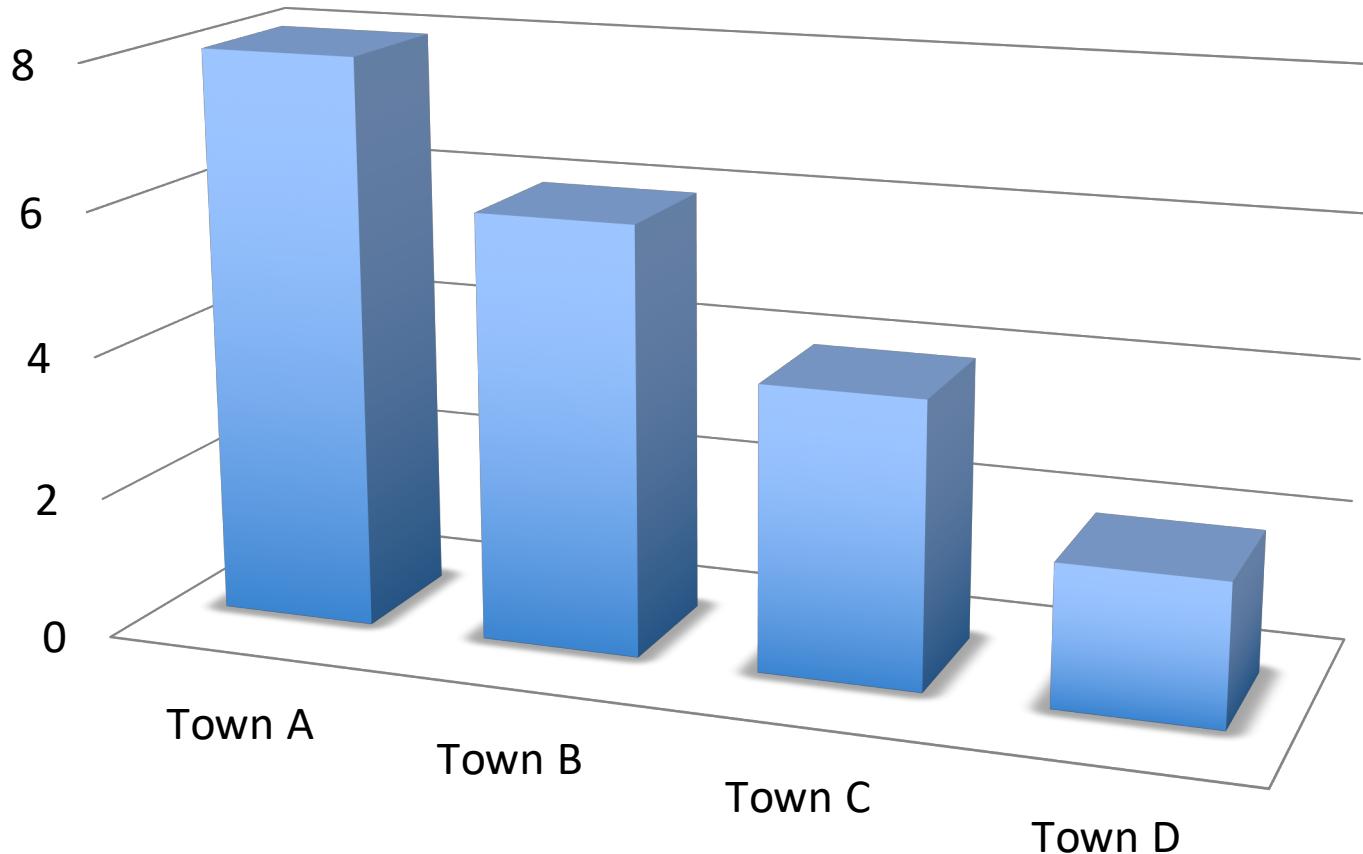
Bar Charts

Represent discrete quantities

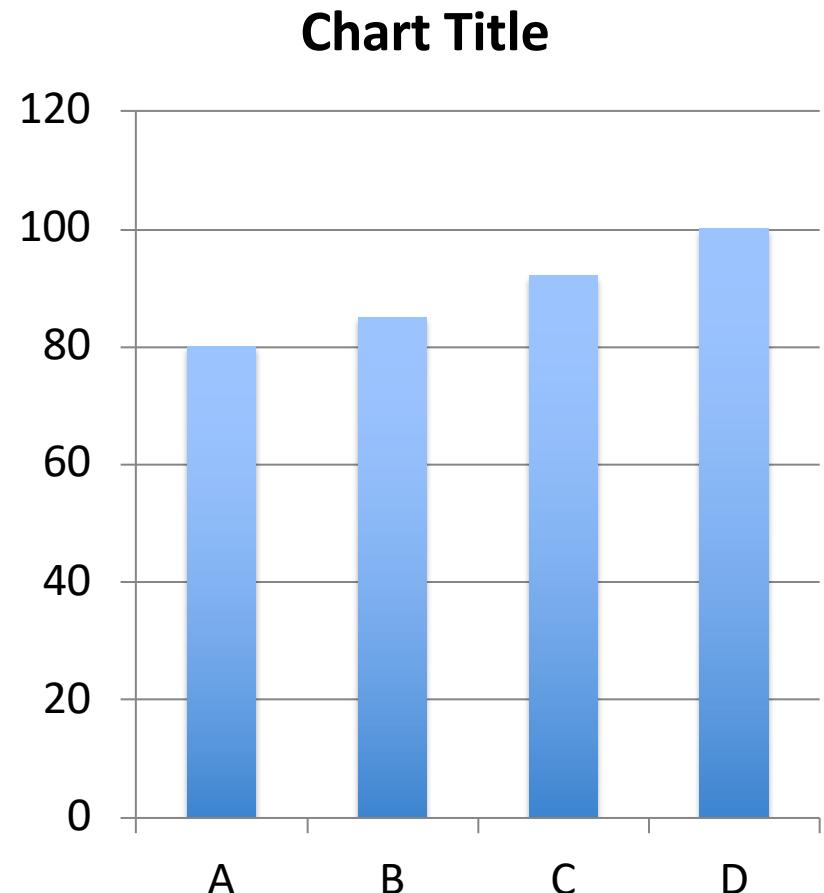
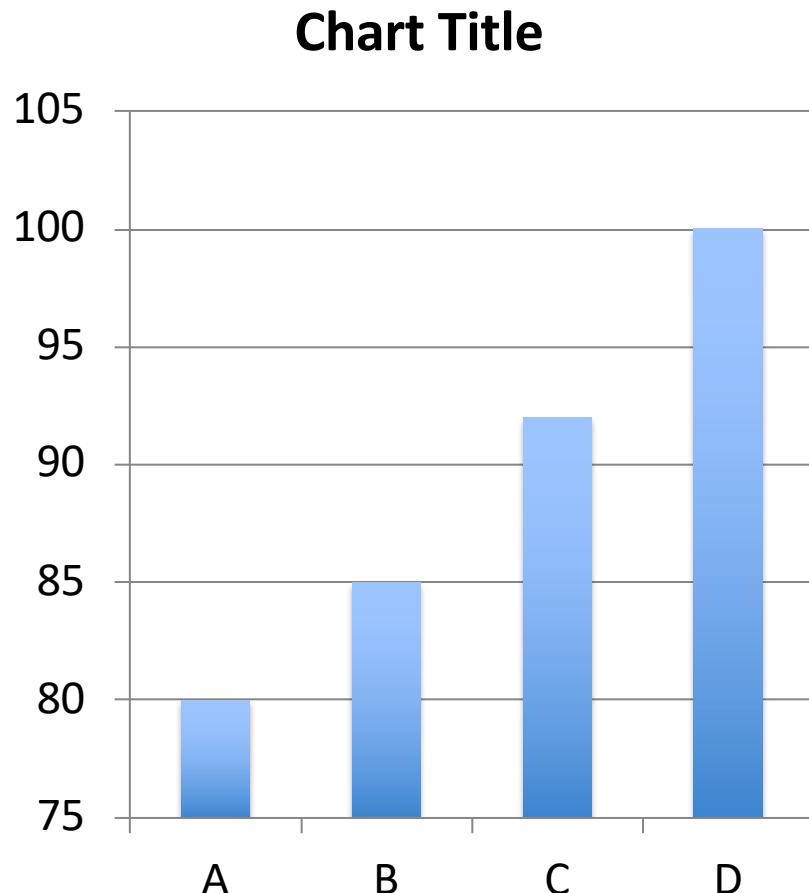


Bar Charts

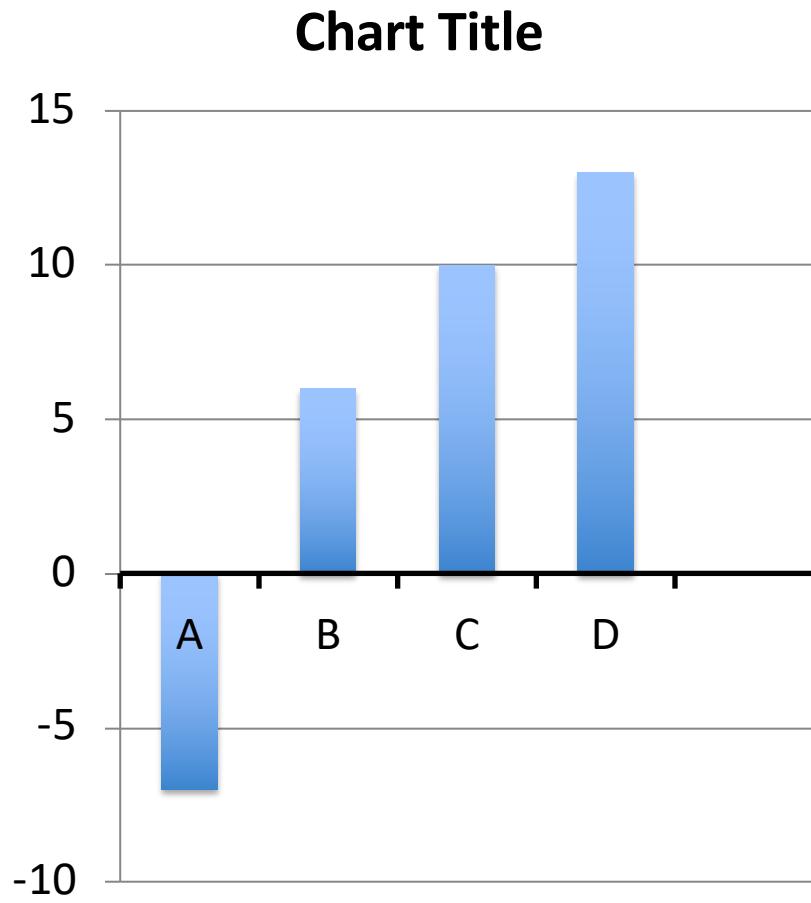
Avoid non-functional adornation



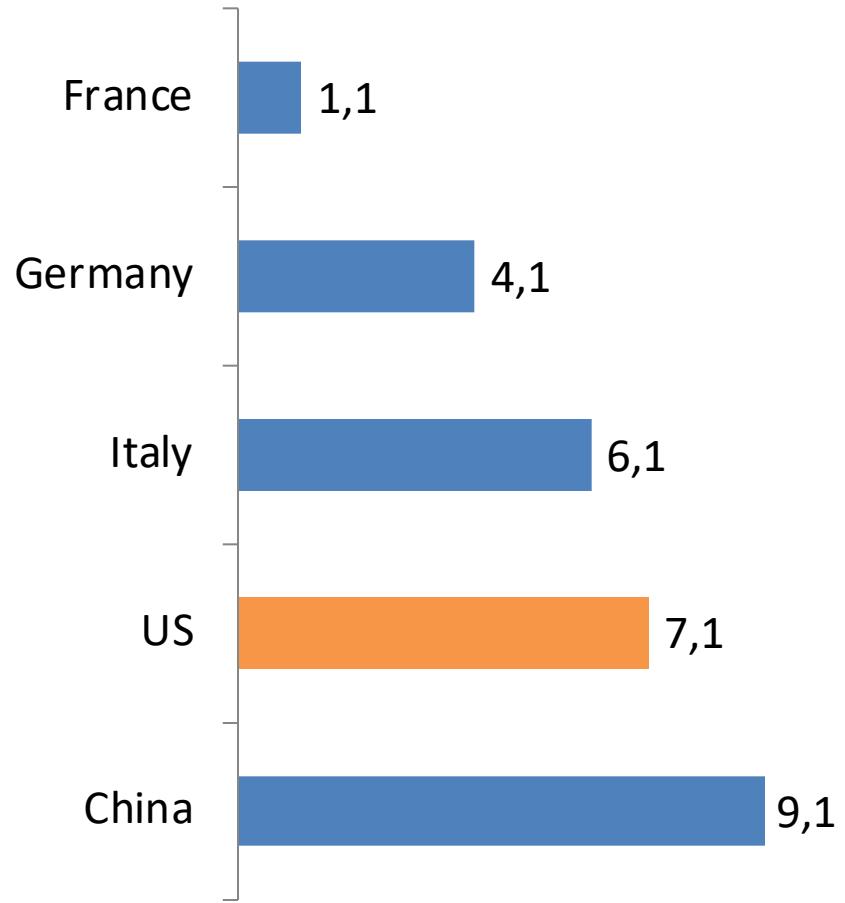
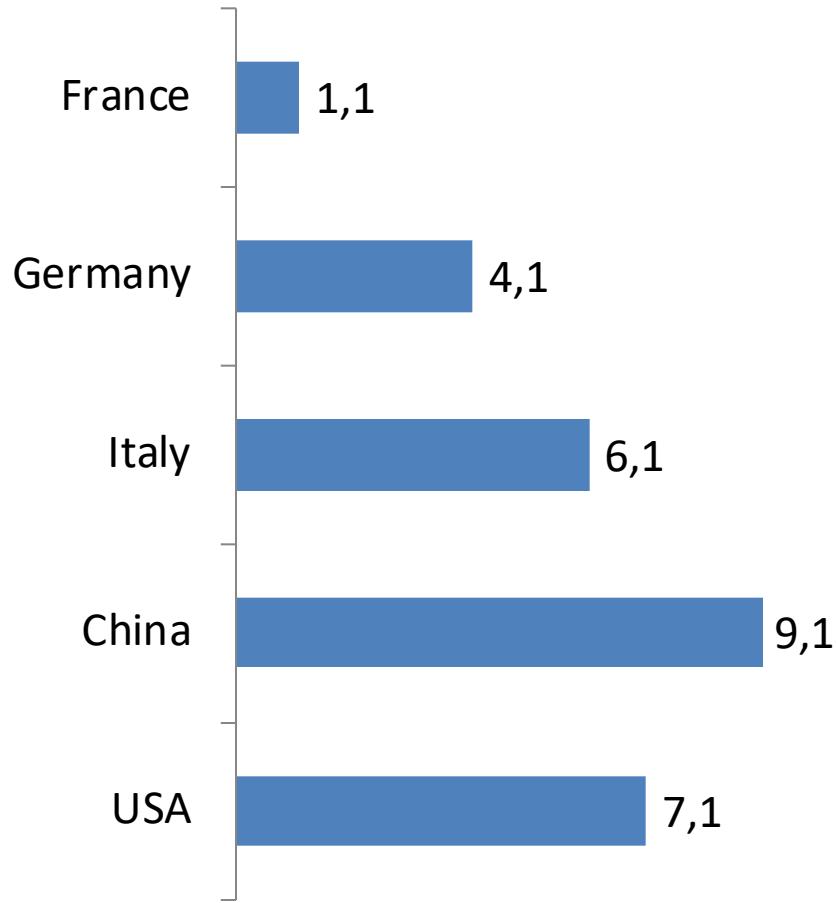
Bar Charts: baseline



Bar Charts: baseline

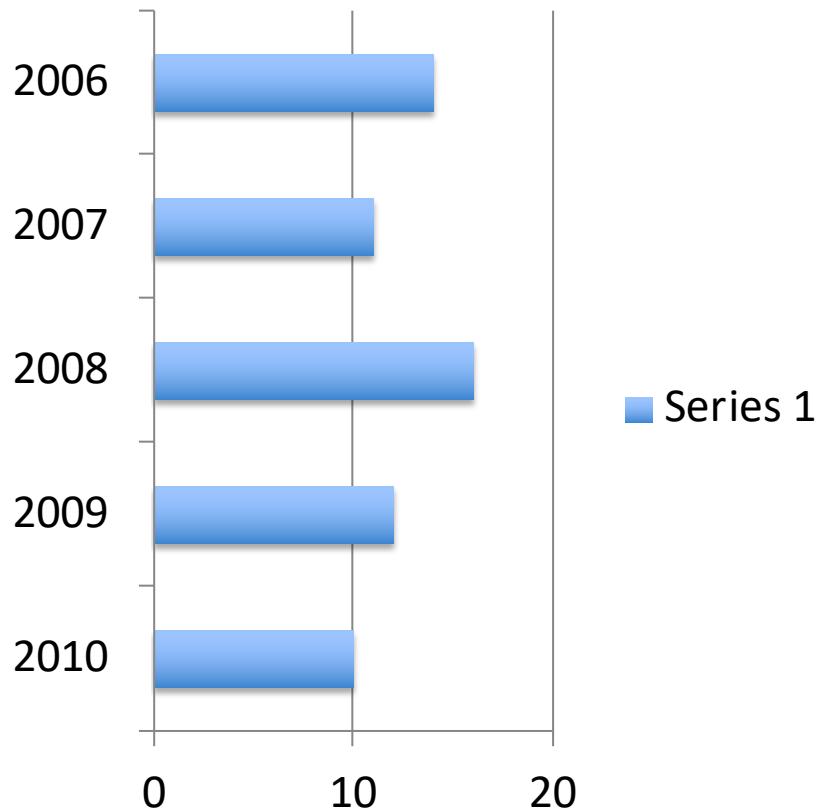


Bar Charts: ordering

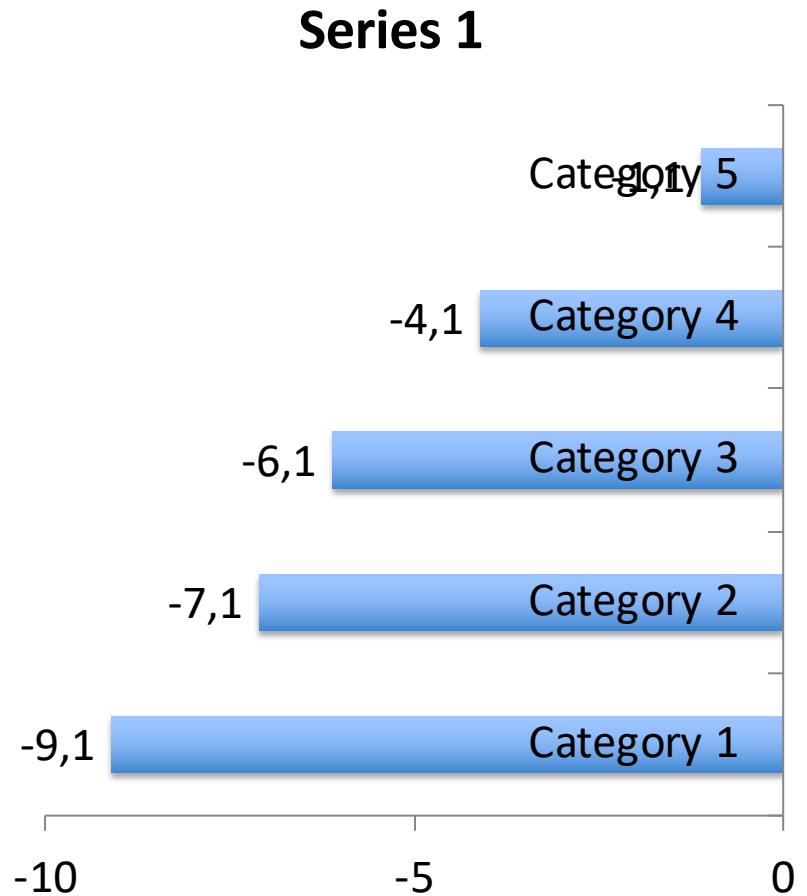




Series 1

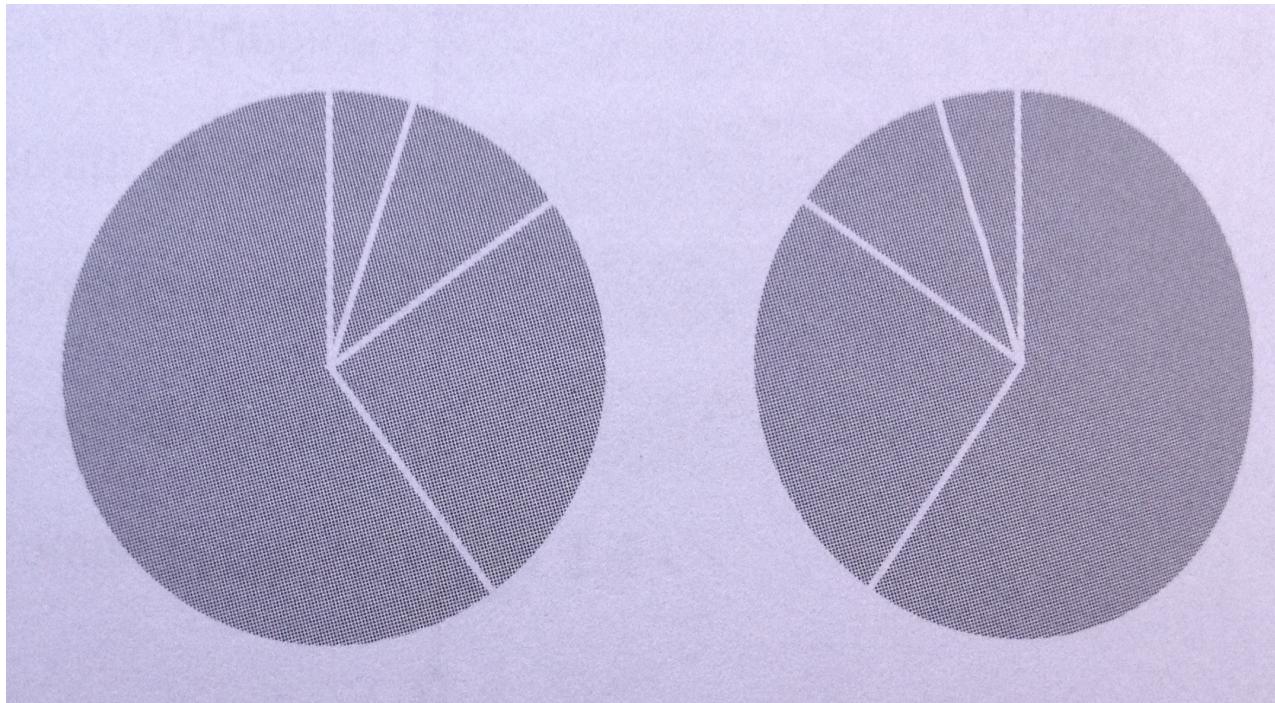


Series 1

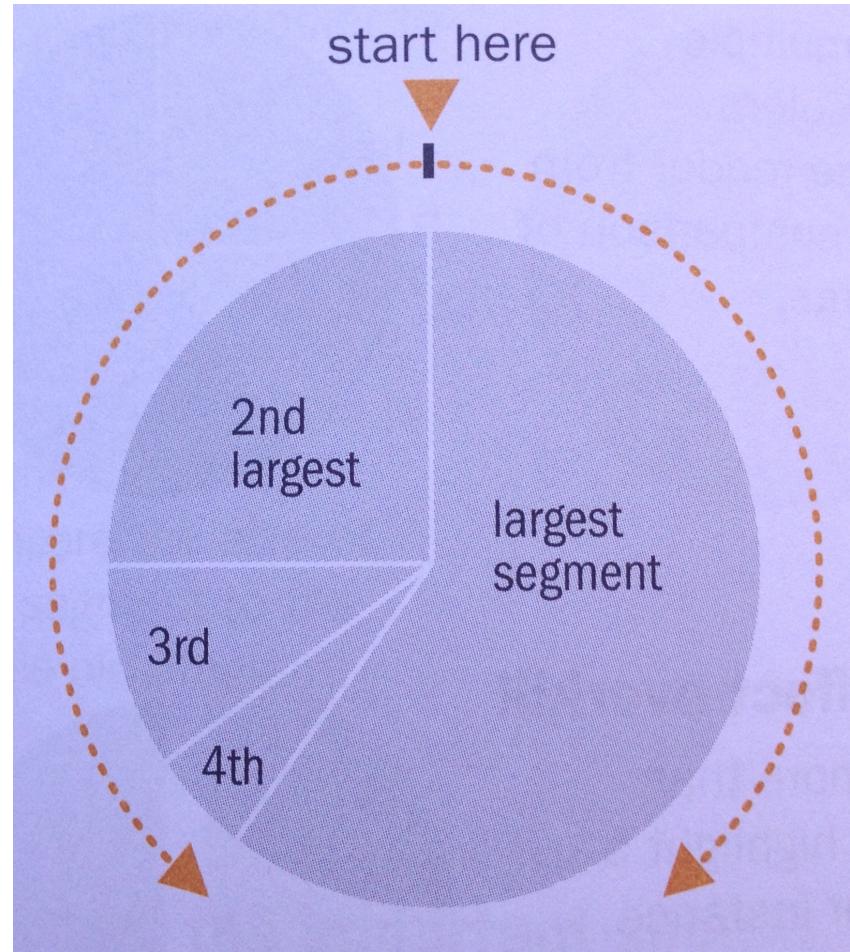


Pie Charts

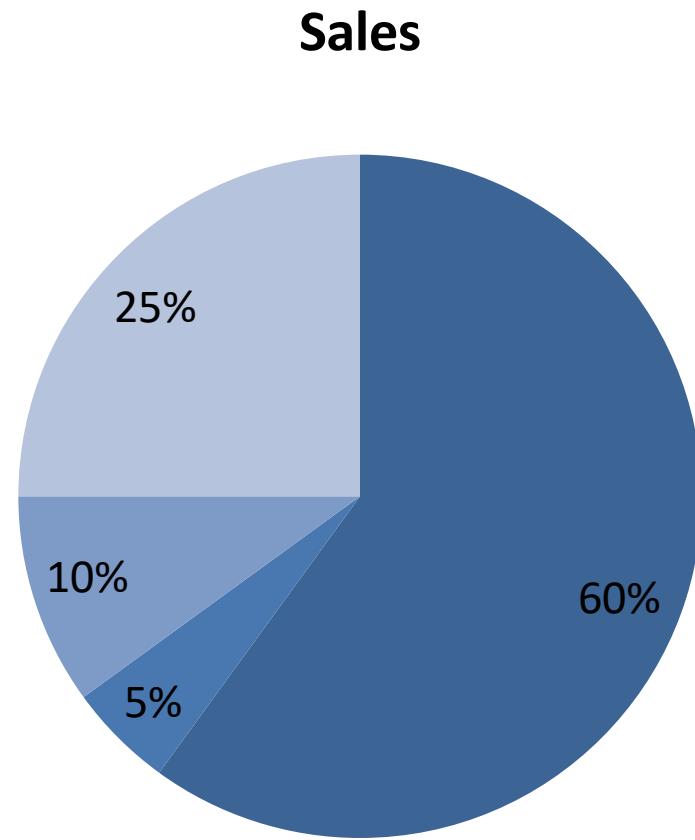
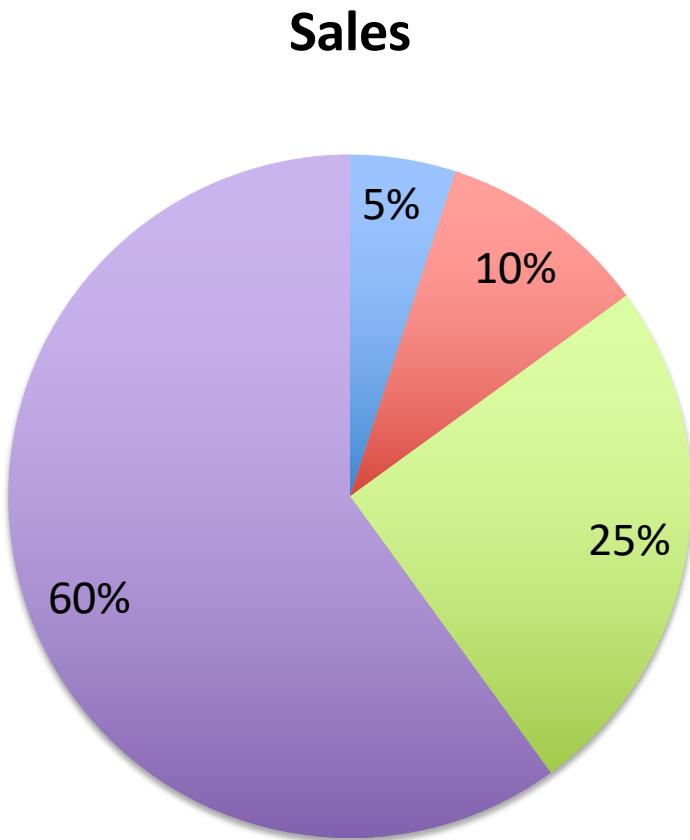
- Pie Charts compares relative sizes and contributions



Pie Charts: ordering slices



Charting Examples



May these charts be improved? Why? How?

Takeaway Messages

- Charts exploit position on scale
- Best practice to reduce biases and misinterpretation of charts

Visualization Taxonomy

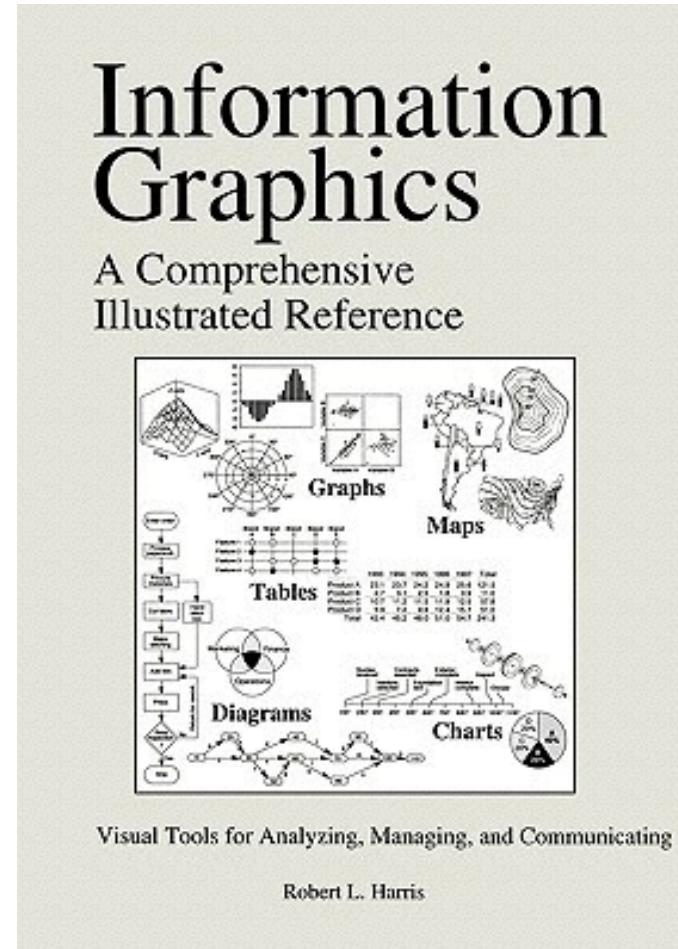
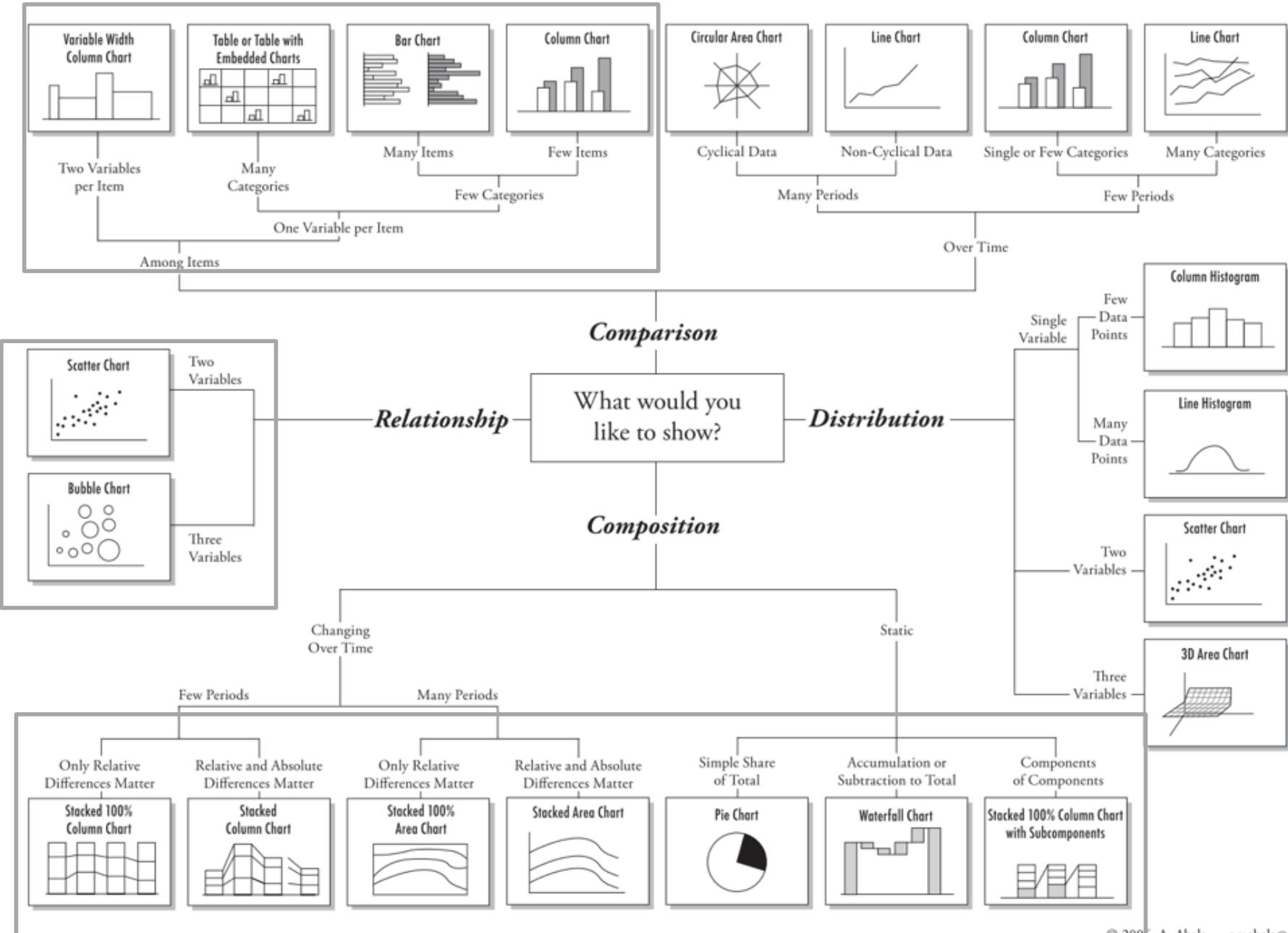
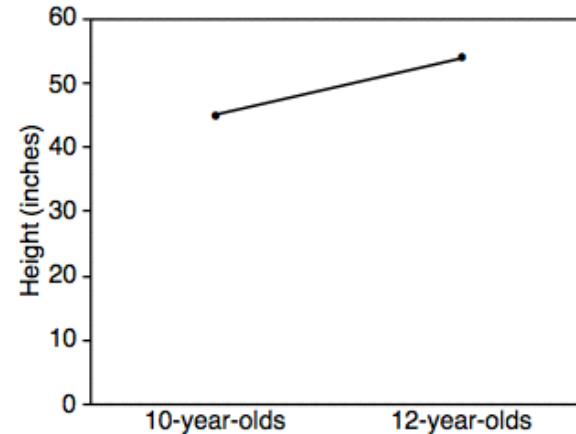
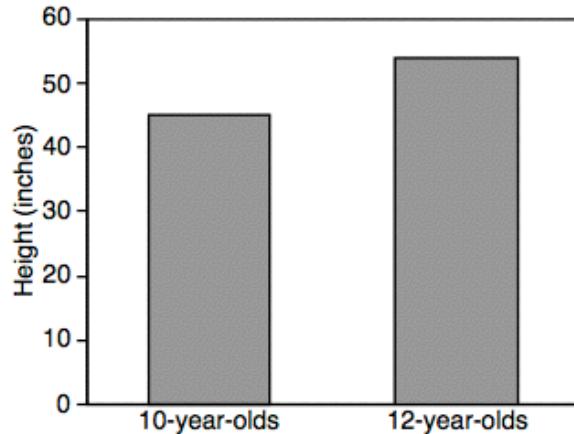
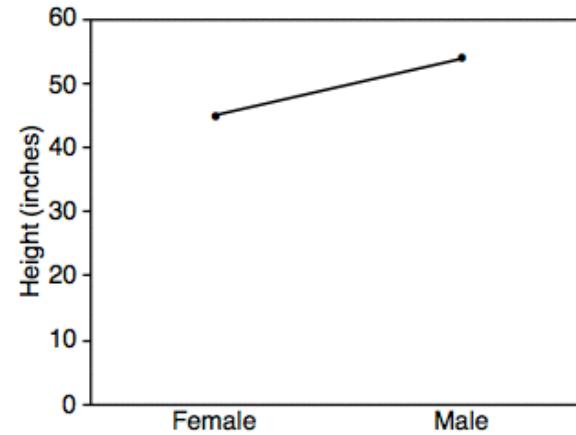
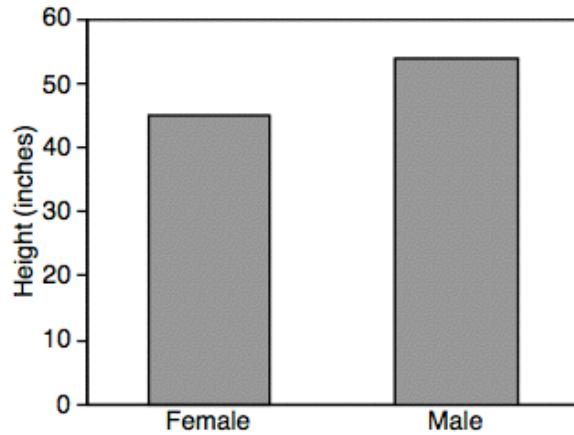


Chart Suggestions—A Thought-Starter



Bars vs. Lines

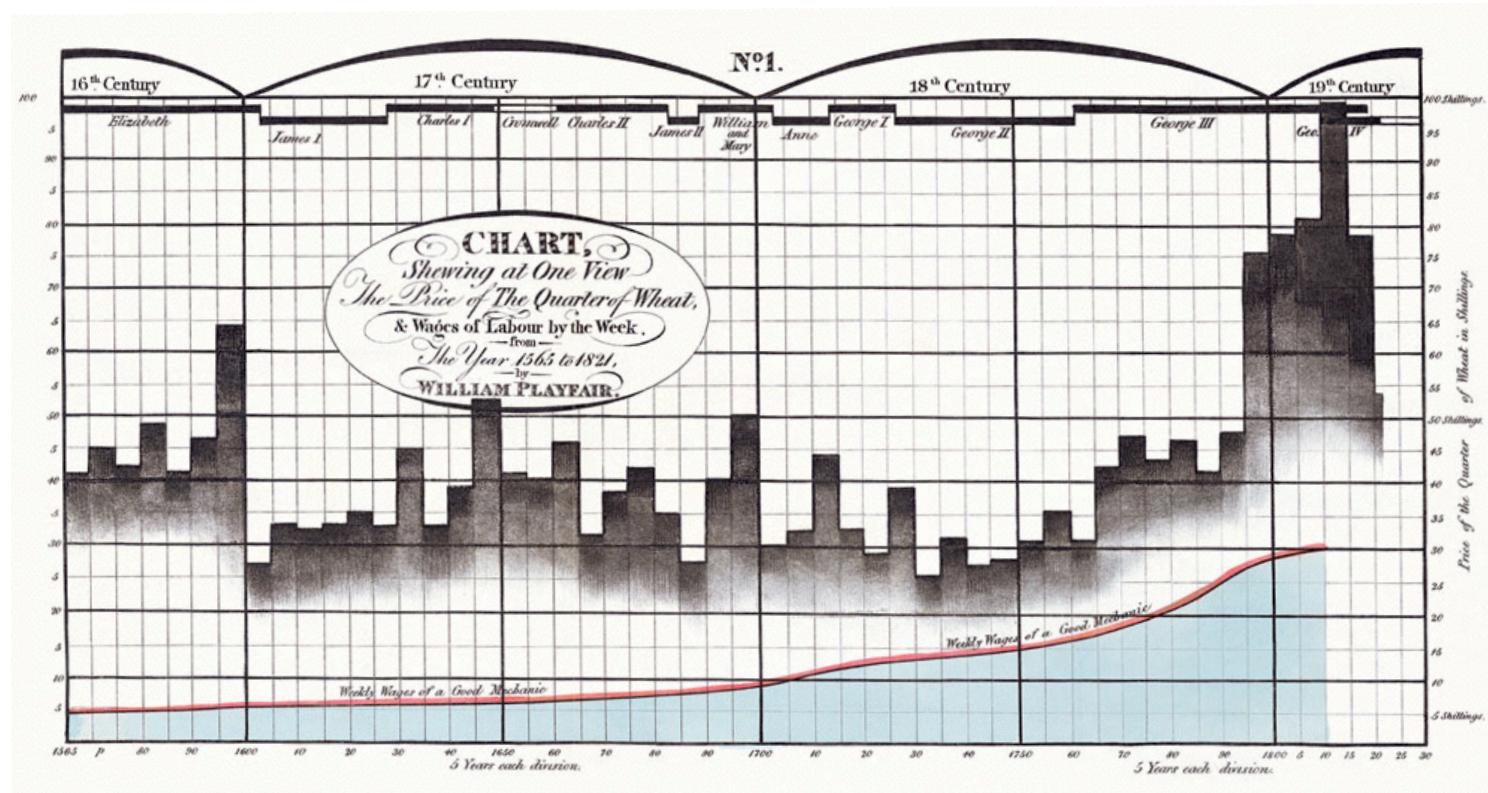


Line implies trends. Do not use for categorical data

Trend over time

WILLIAM PLAYFAIR

1759-1823



Trend over time



Trend over time

Published: February 2, 2010

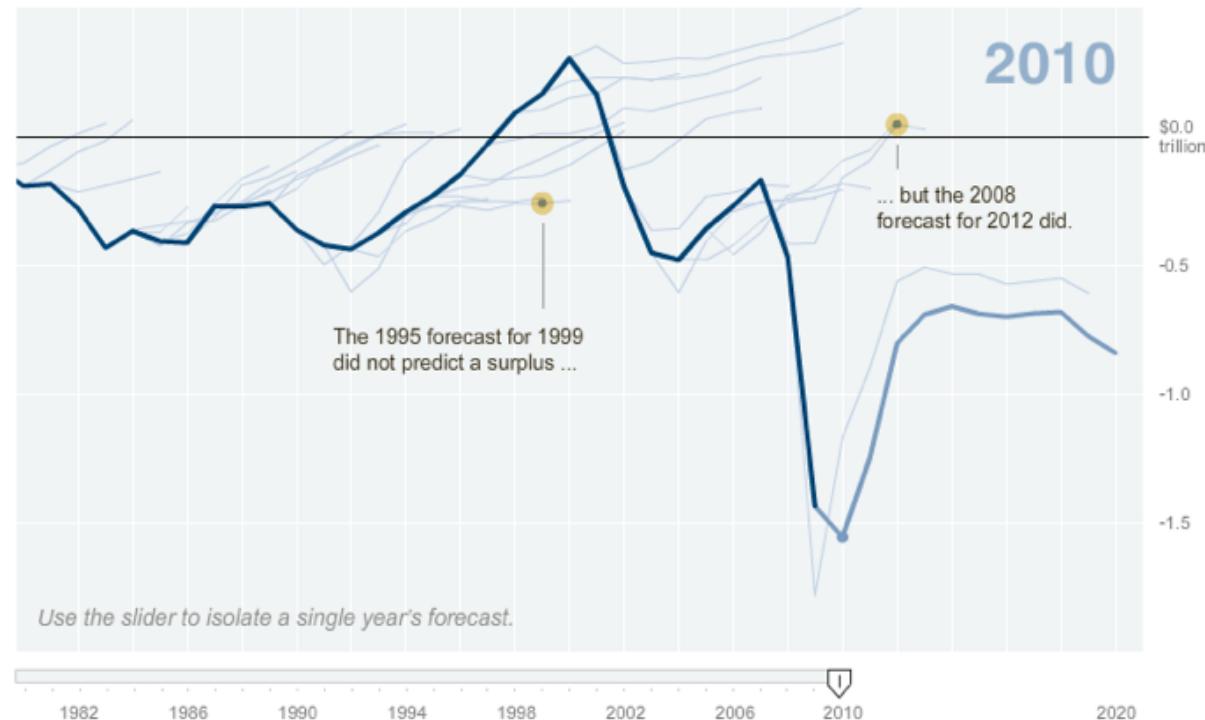
Budget Forecasts, Compared With Reality

Just two years ago, surpluses were predicted by 2012. How accurate have past White House budget forecasts been?

1 2 3 4 5 6 NEXT ►

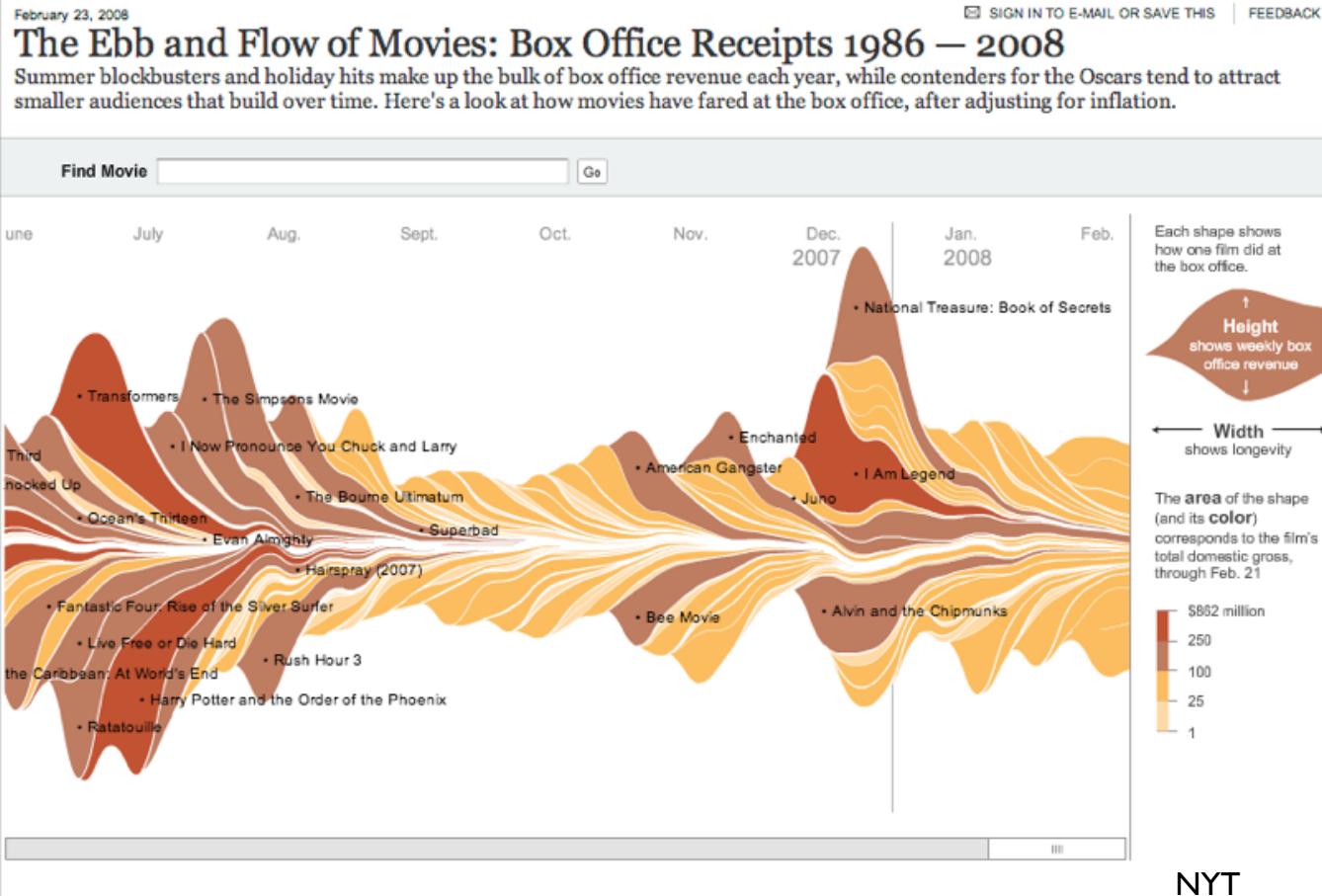
Latest forecast

Today, with a better understanding of the severity of the economic downturn, the deficit situation is much more dire.



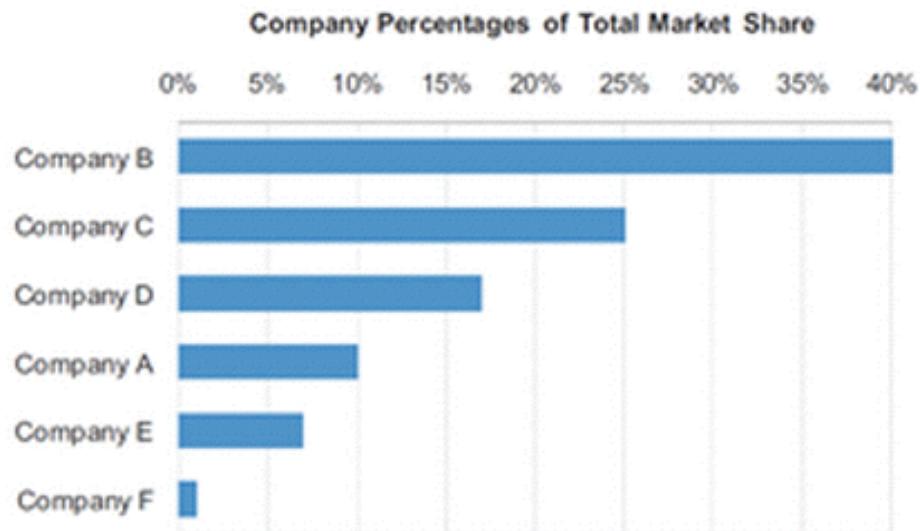
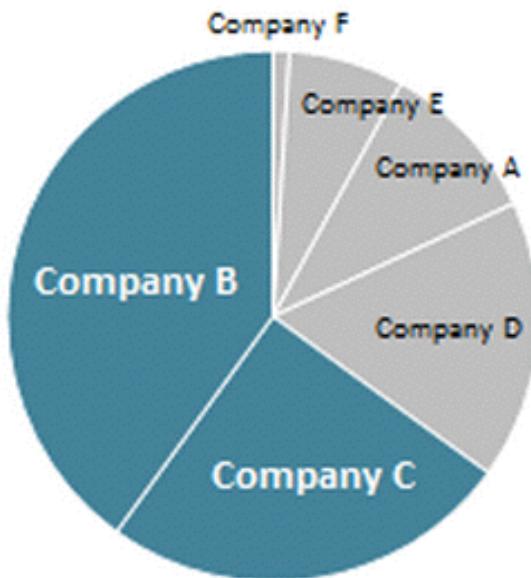
Make clear distinction between data and prediction

Streamgraphs

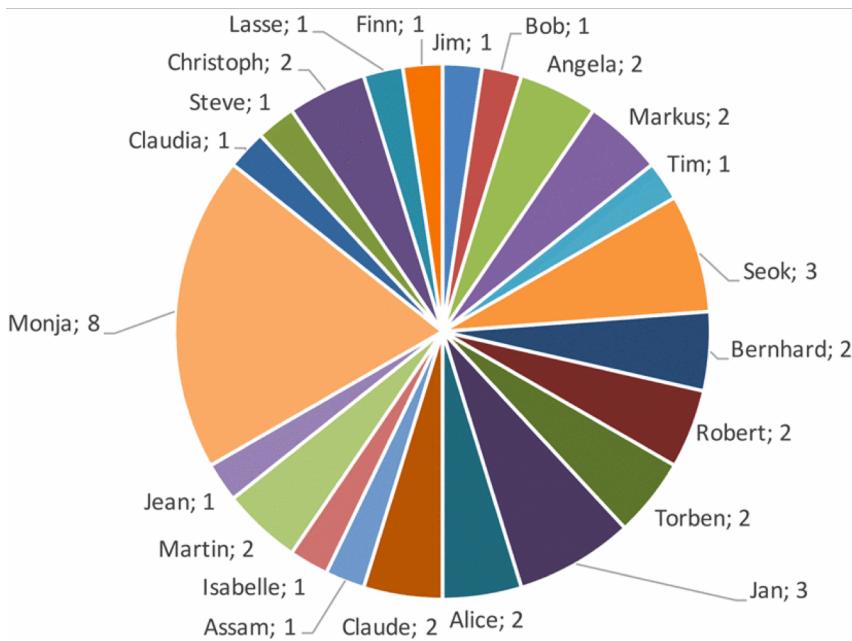


Pie vs Bar charts

65% of the market is controlled by companies B and C

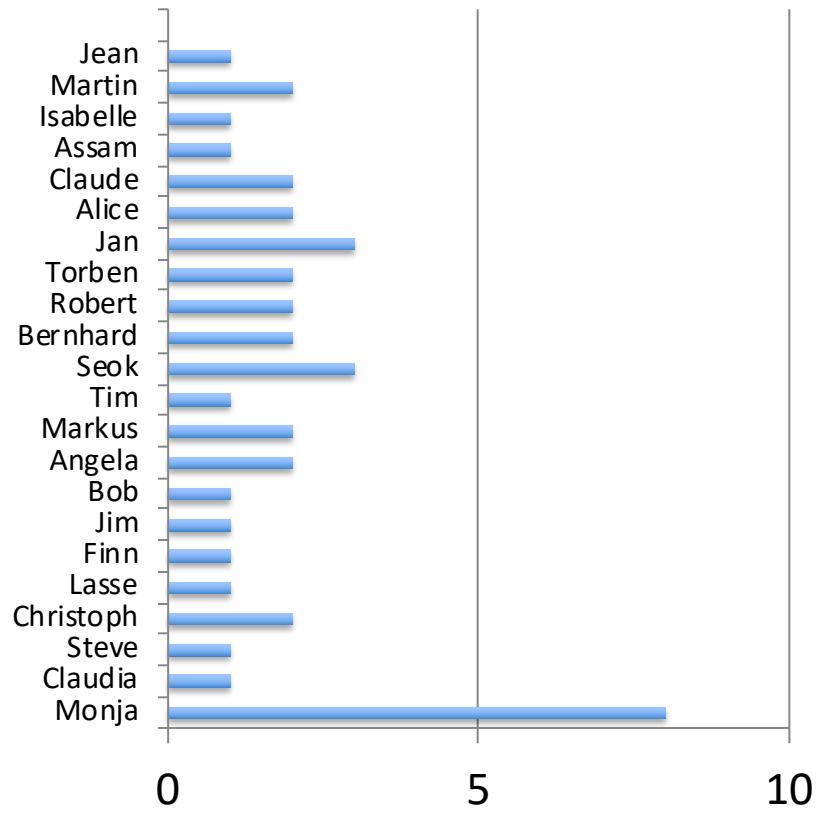


Pies vs Bar charts

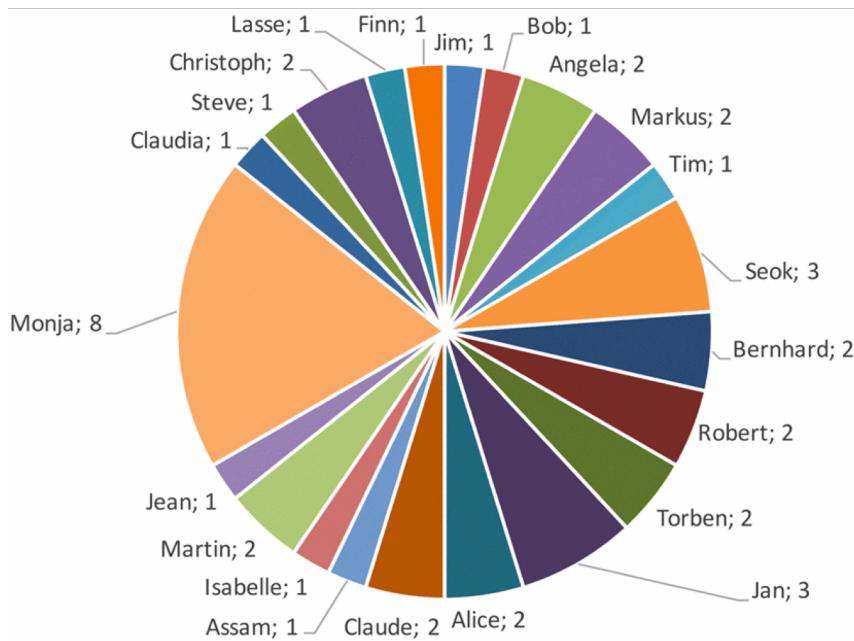


Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the names are pseudonyms for real people. The person that suffered the most attacks is Monja a secretary with overall 8 attacks. In contrast, all other victims suffered between 1 and 3 attacks.

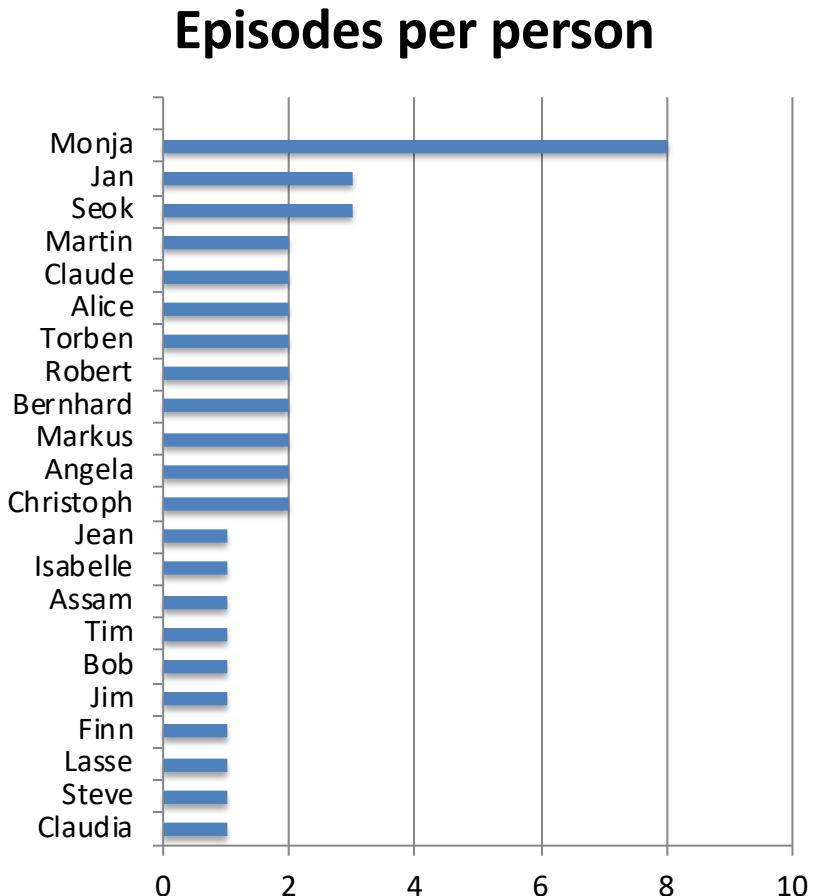
Episodes



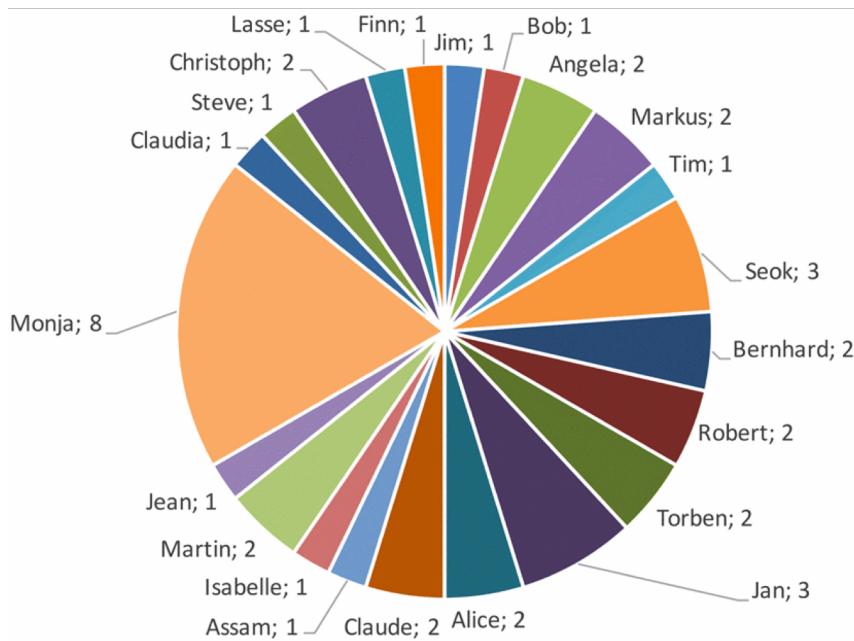
Pies vs Bar charts (improved)



Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the names are pseudonyms for real people. The person that suffered the most attacks is Monja a secretary with overall 8 attacks. In contrast, all other victims suffered between 1 and 3 attacks.

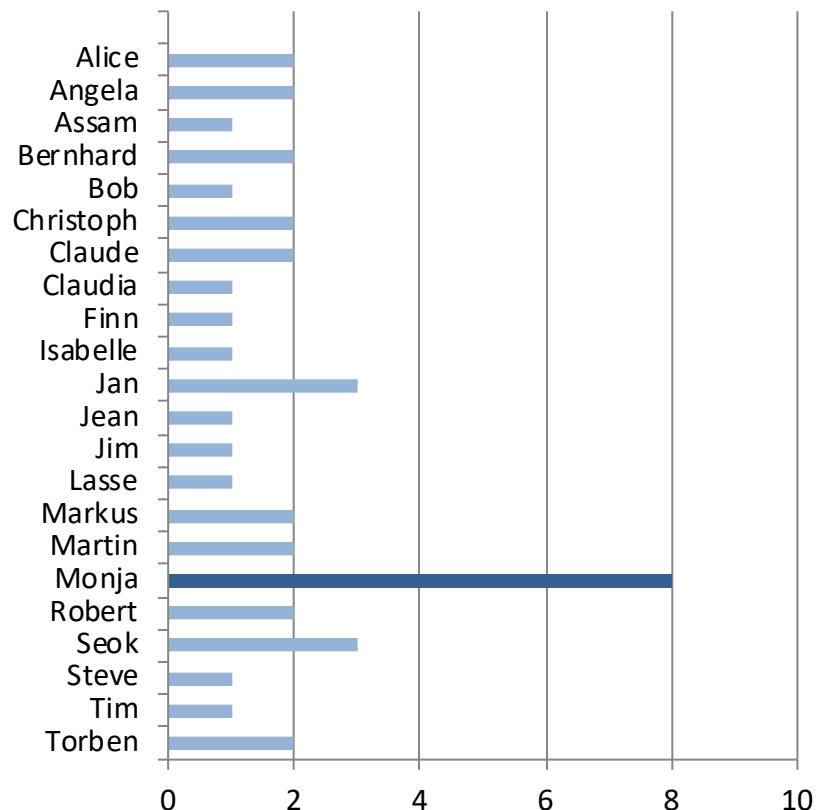


Pies vs Bar charts (improved)

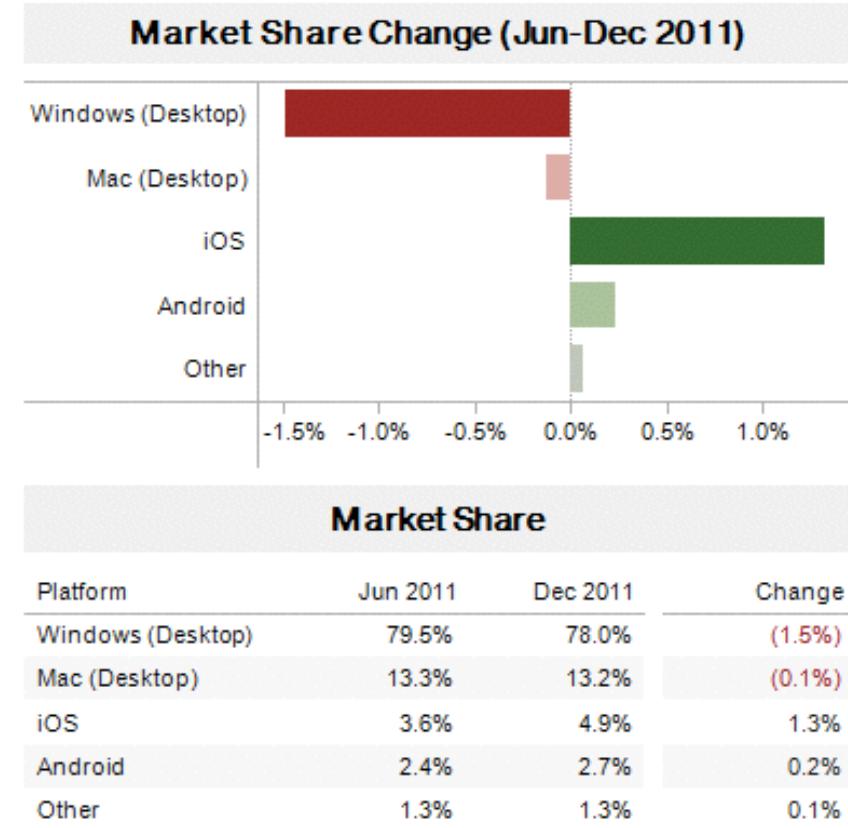
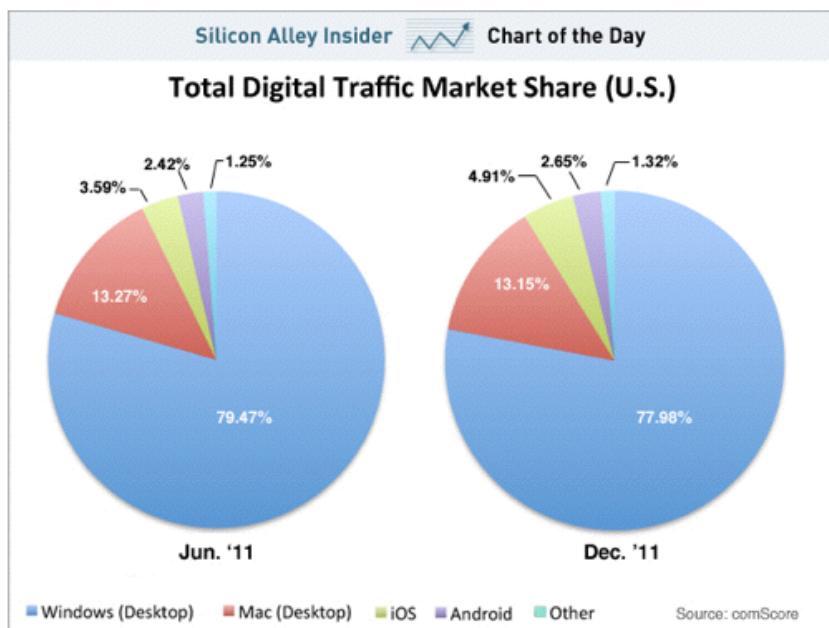


Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the names are pseudonyms for real people. The person that suffered the most attacks is Monja a secretary with overall 8 attacks. In contrast, all other victims suffered between 1 and 3 attacks.

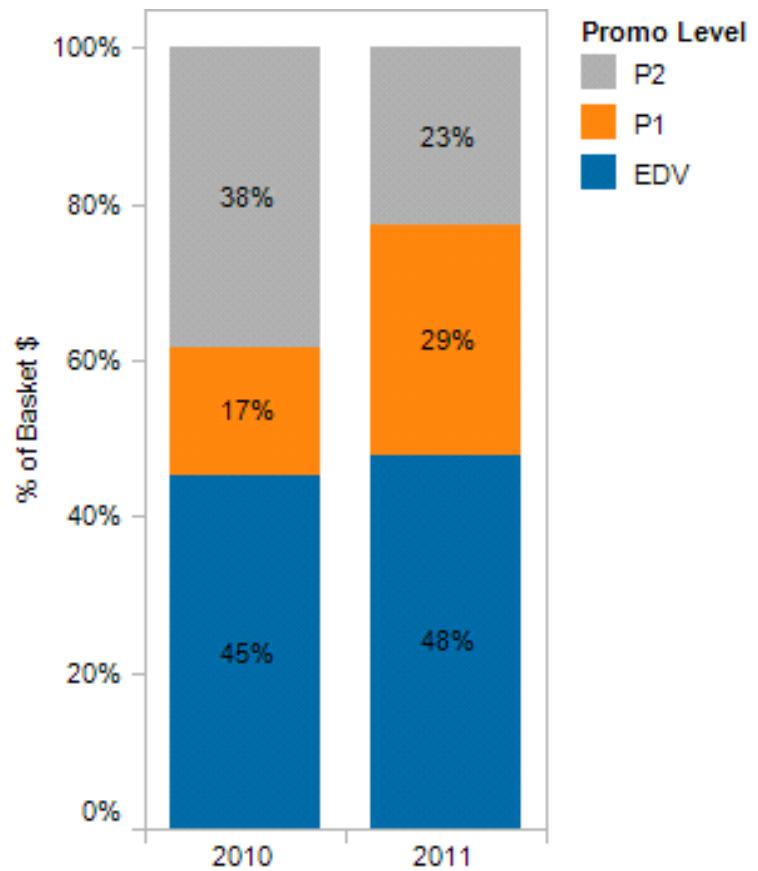
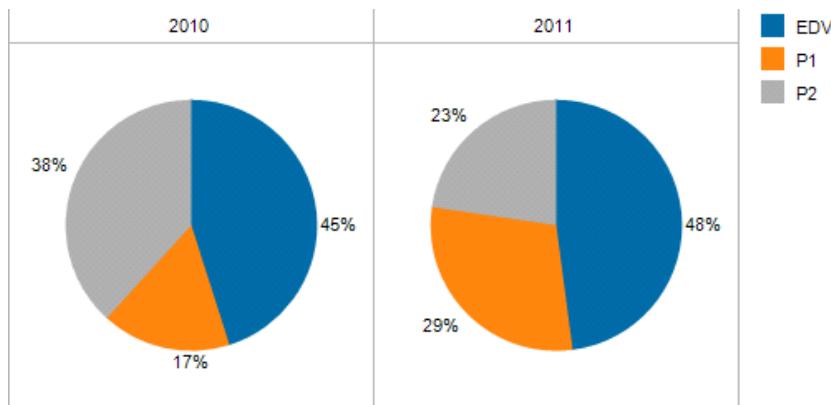
Episodes per person



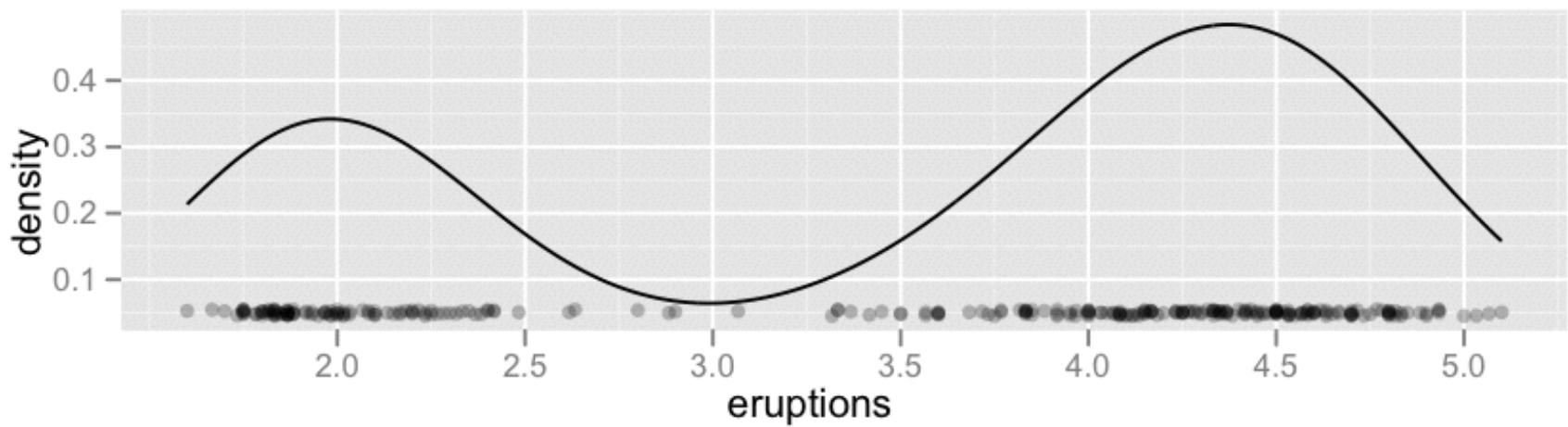
Showing changes



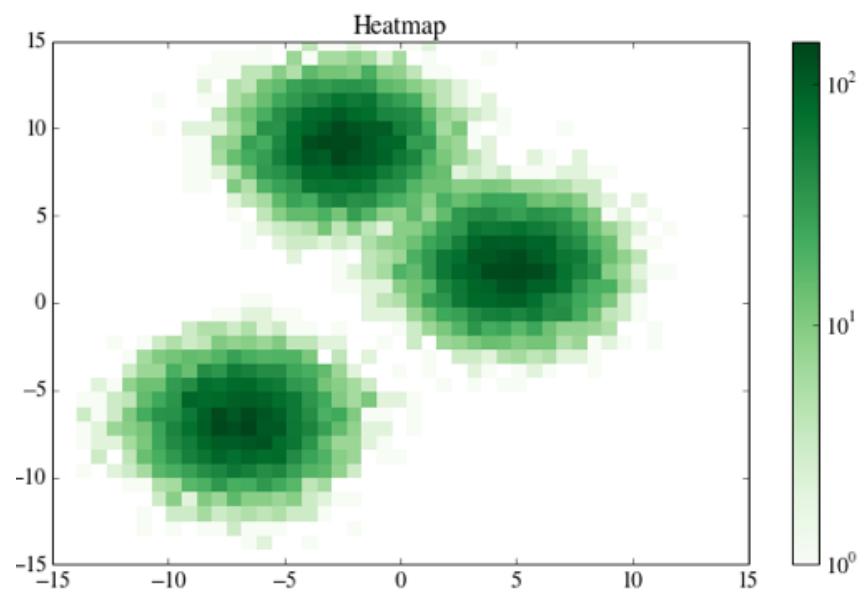
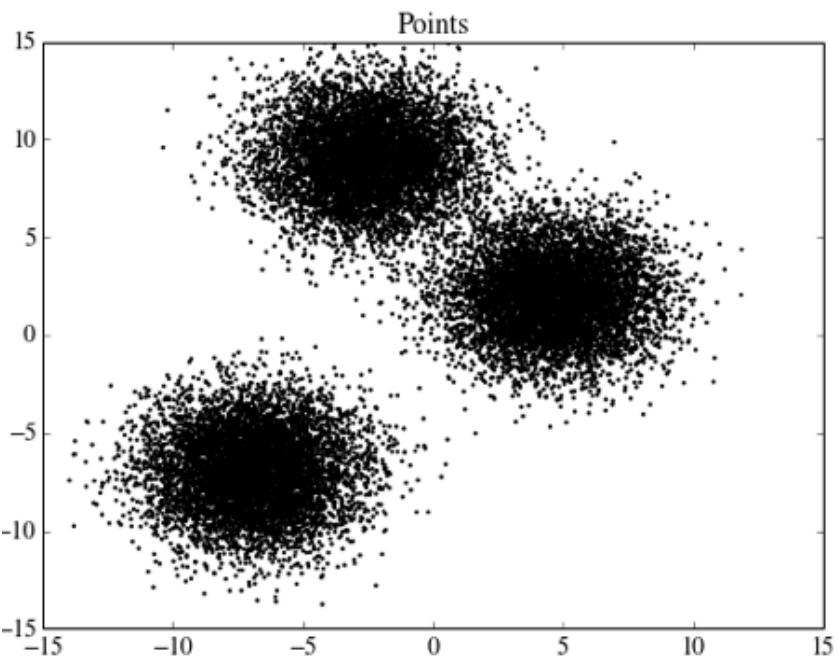
Showing Changes



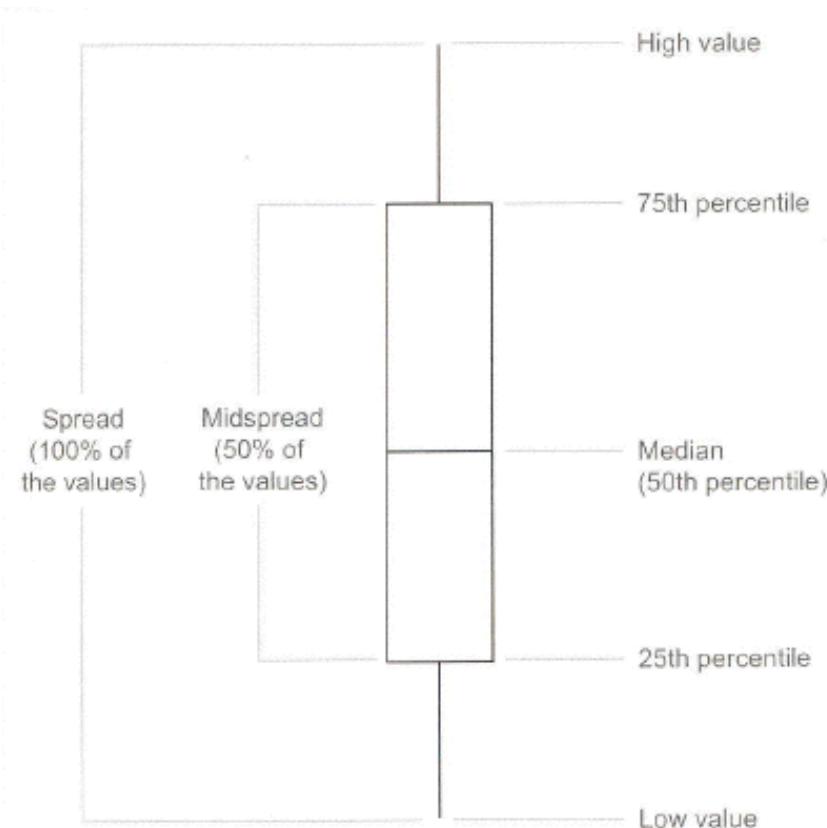
Density Plot



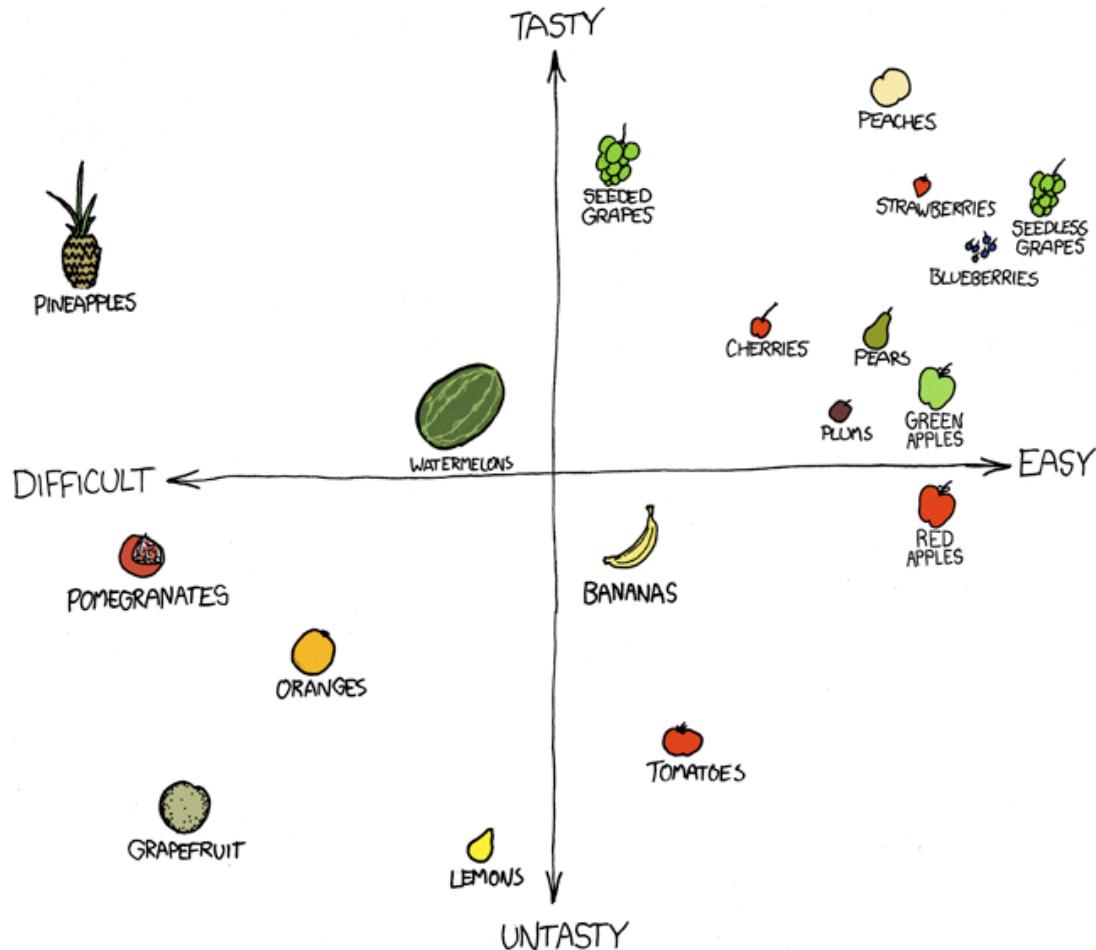
2D Density Plots



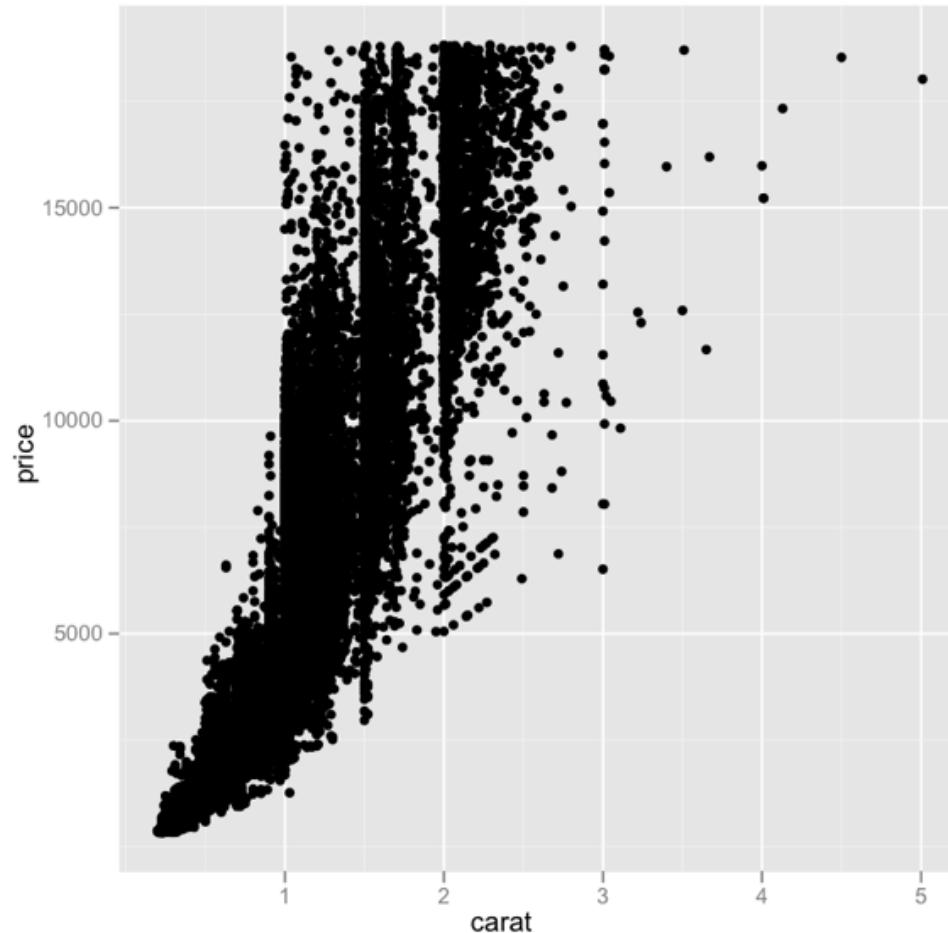
Box Plots



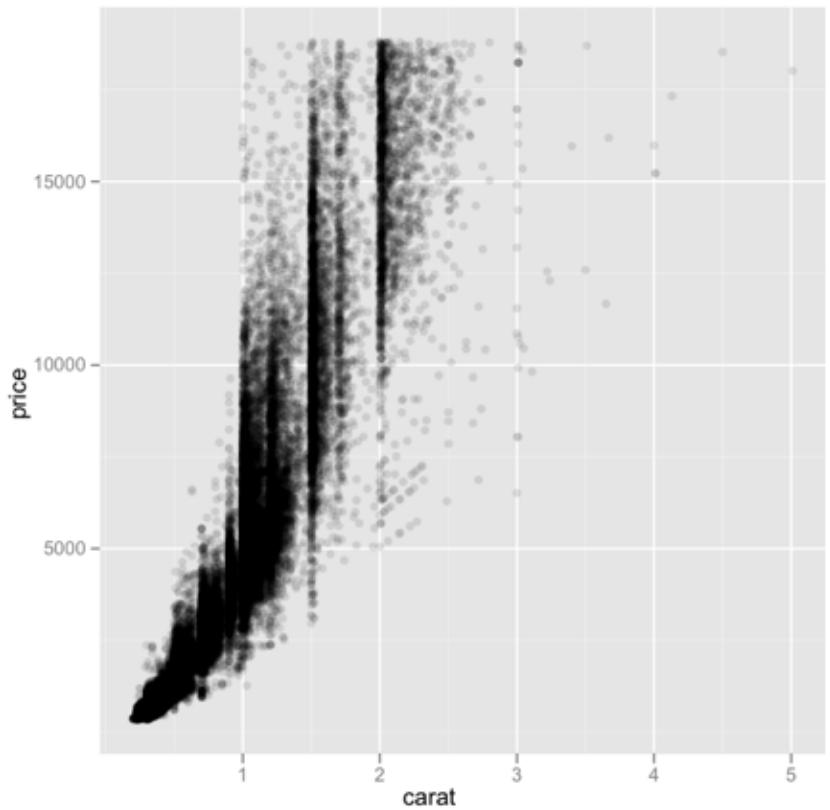
Scatterplot



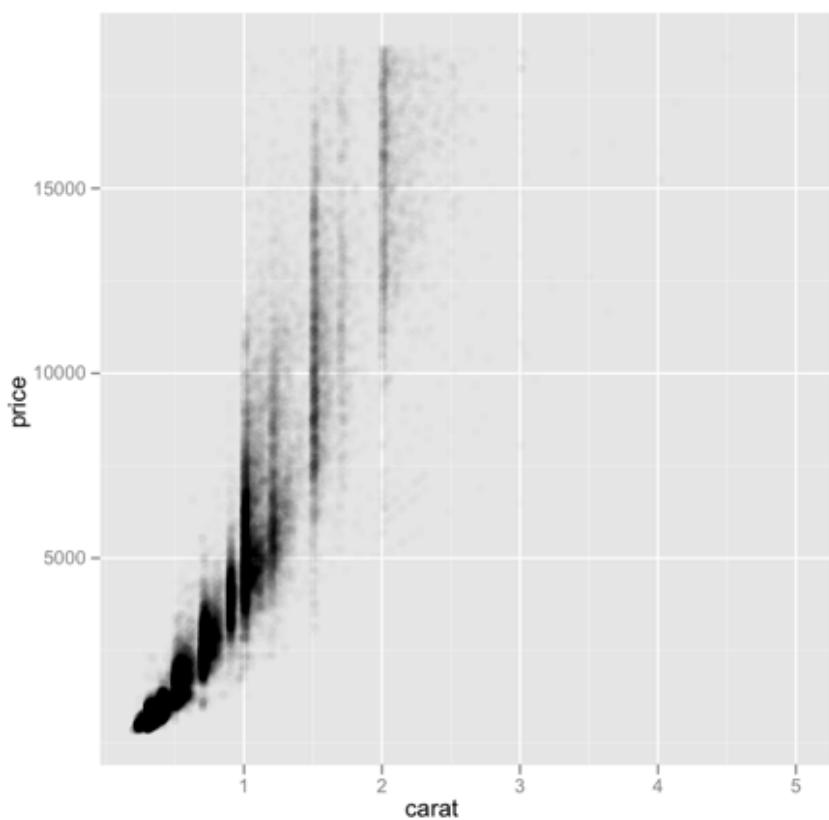
Cluttering, Overplotting



alpha=1/10



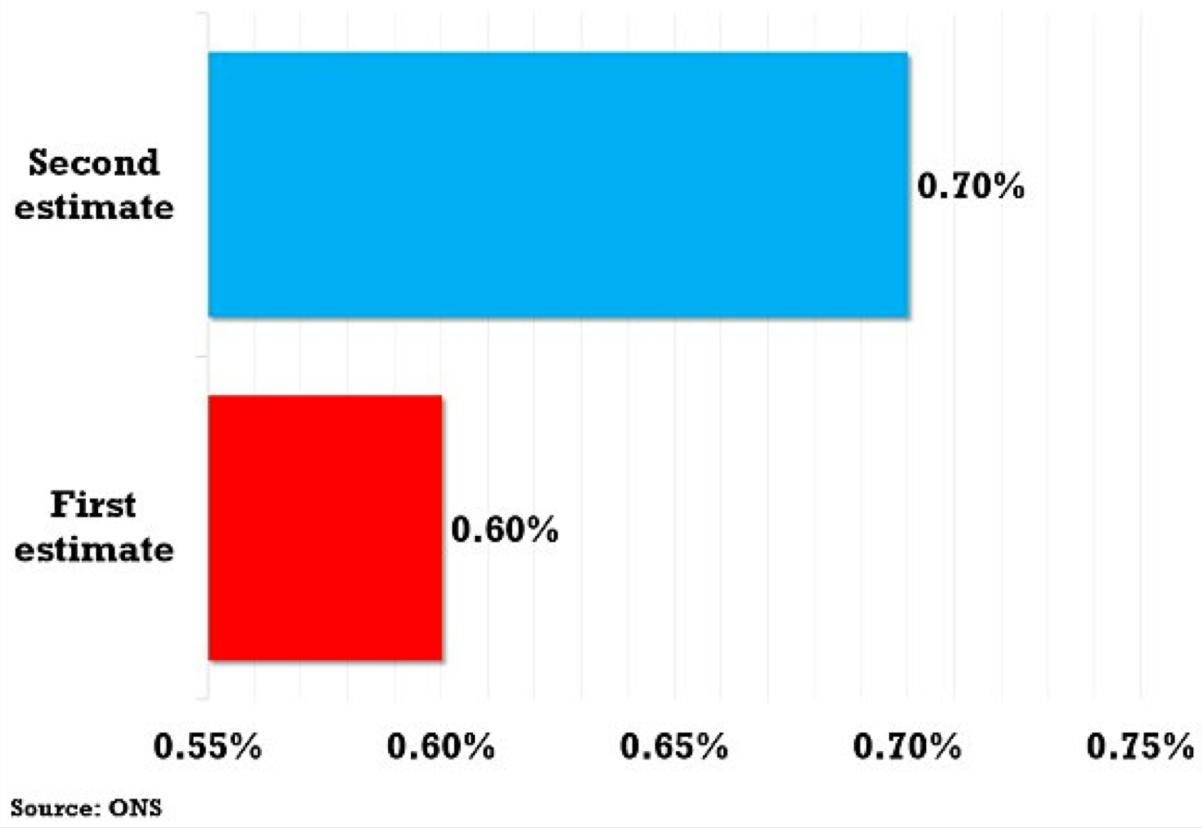
alpha=1/100





A FEW EXAMPLES AND CASE STUDIES

2016 Q4 GROWTH UPGRADED



The Office for National Statistics (ONS) said gross domestic product (GDP) expanded by 0.7 per cent in the fourth quarter - an increase from the 0.6 per cent calculated on the watchdog's first look at the economy

Source: <http://www.dailymail.co.uk/news/article-4248690/Economy-grew-0-7-final-three-months-2016.html>

Awareness Engagement Tickets

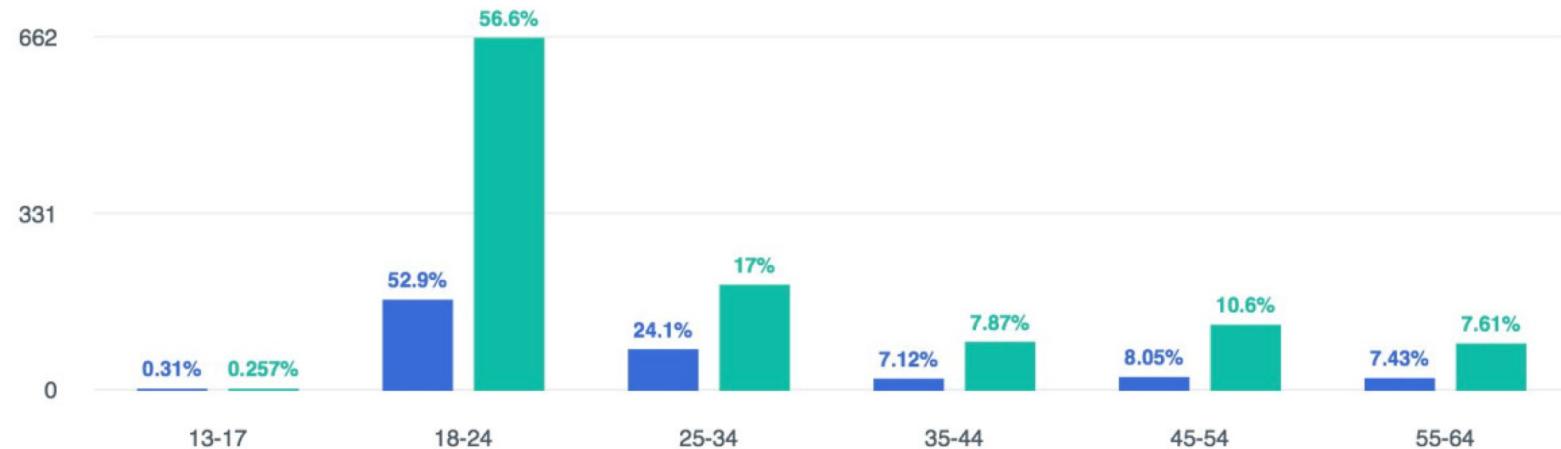
Audience

Include events that your Page is co-hosting

Last 7 days ▾

Demographics

Men Women

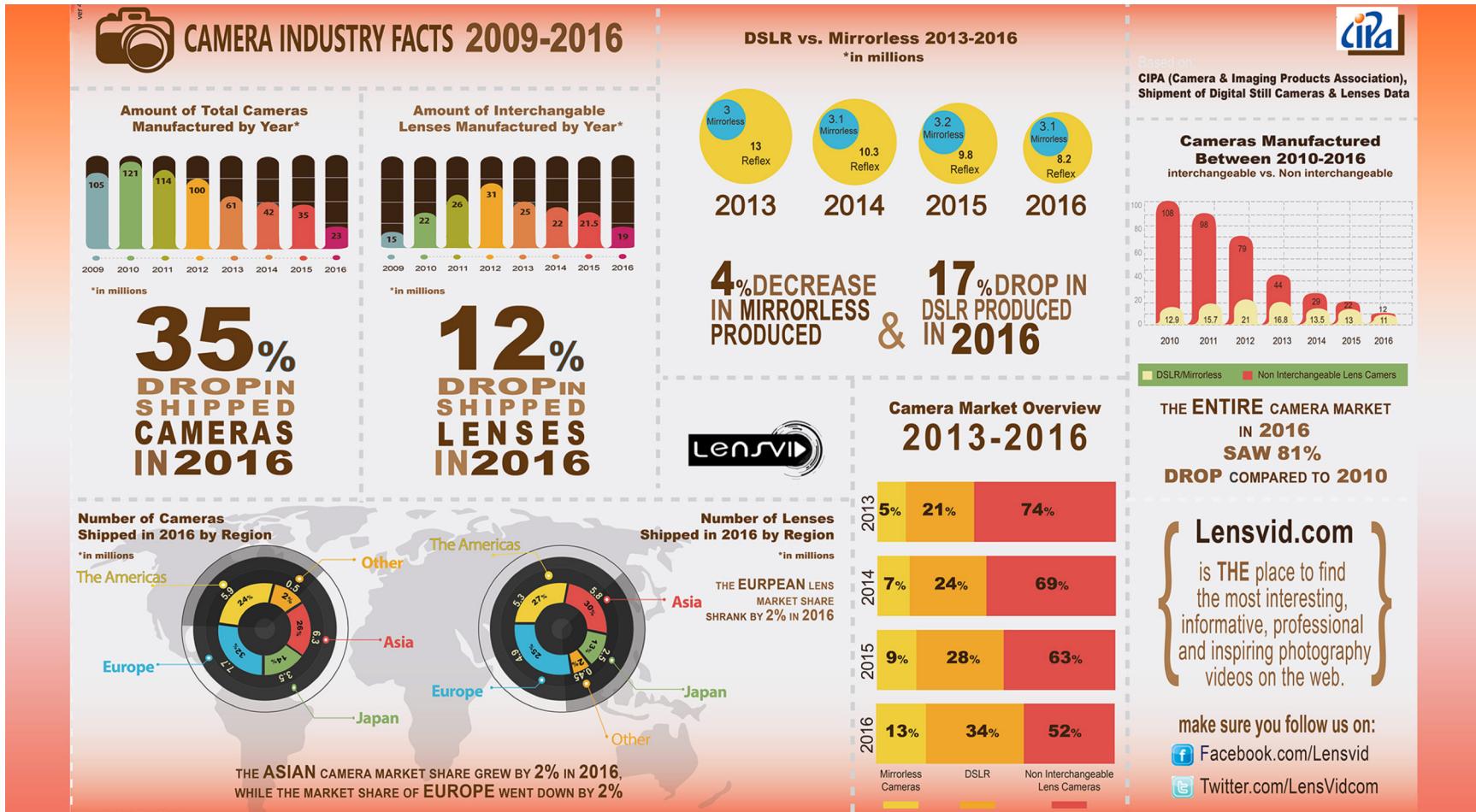


Source: Facebook Analytics

Procent użytków rolnych w gospodarstwach > niż 50 ha:

1989

25%

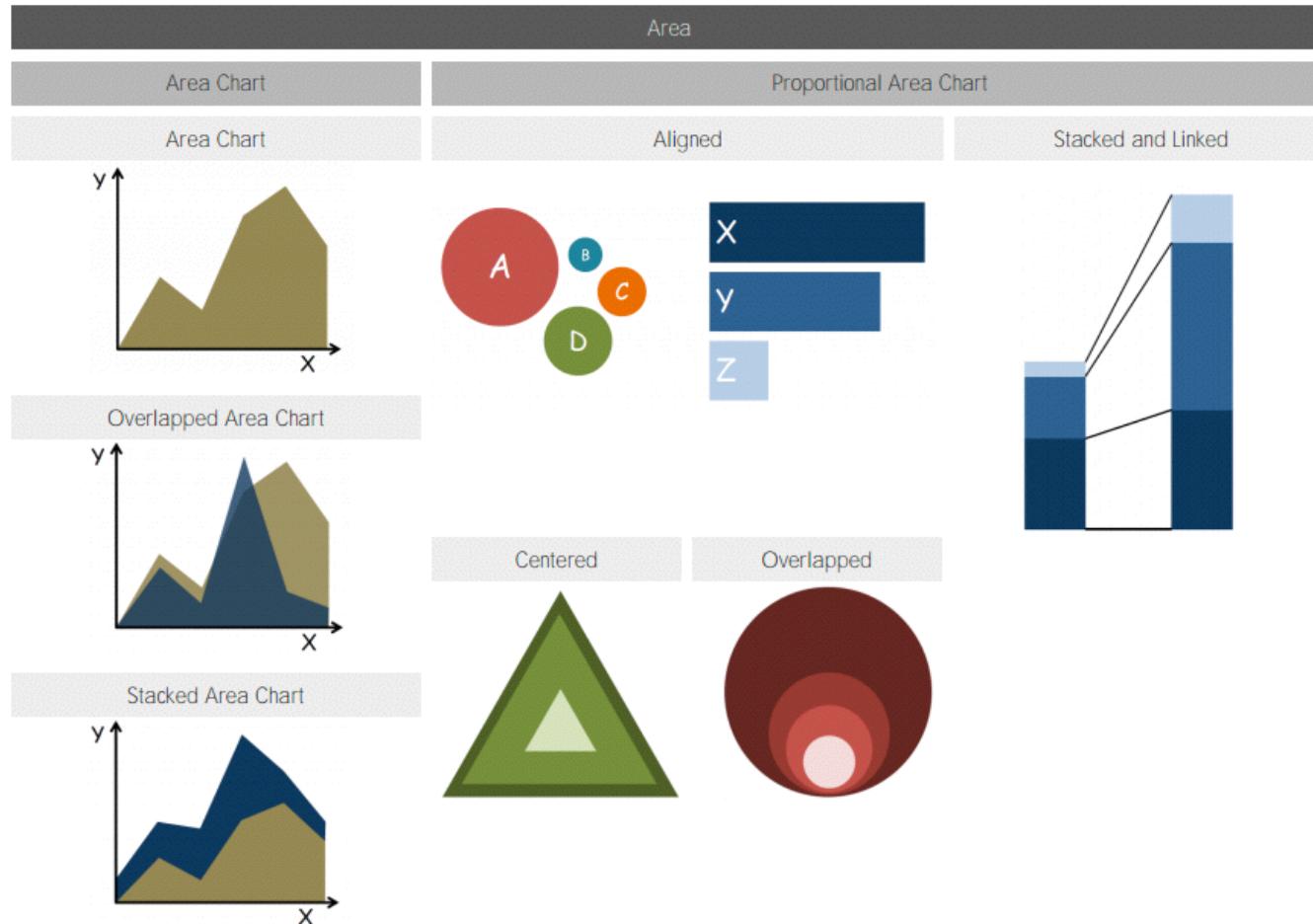


Borkin MA, Vo AA, Bylinskii Z, Isola P, Sunkavalli S, Oliva A, Pfister H.
What Makes a Visualization Memorable?
IEEE Transactions on Visualization and Computer Graphics (InfoVis 2013).

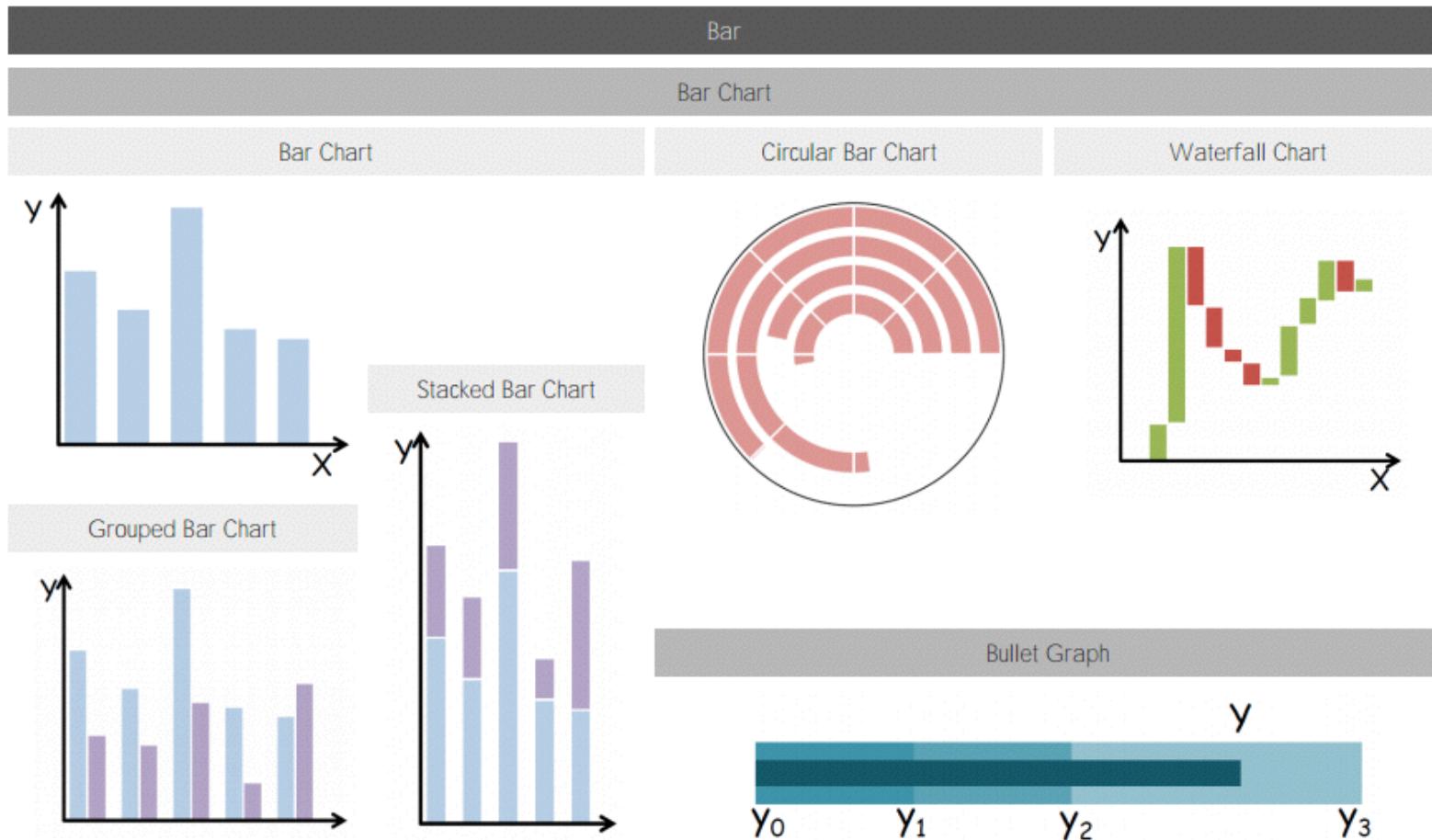
[http://vcg.seas.harvard.edu/publications/ what-makes-visualization-memorable](http://vcg.seas.harvard.edu/publications/what-makes-visualization-memorable)

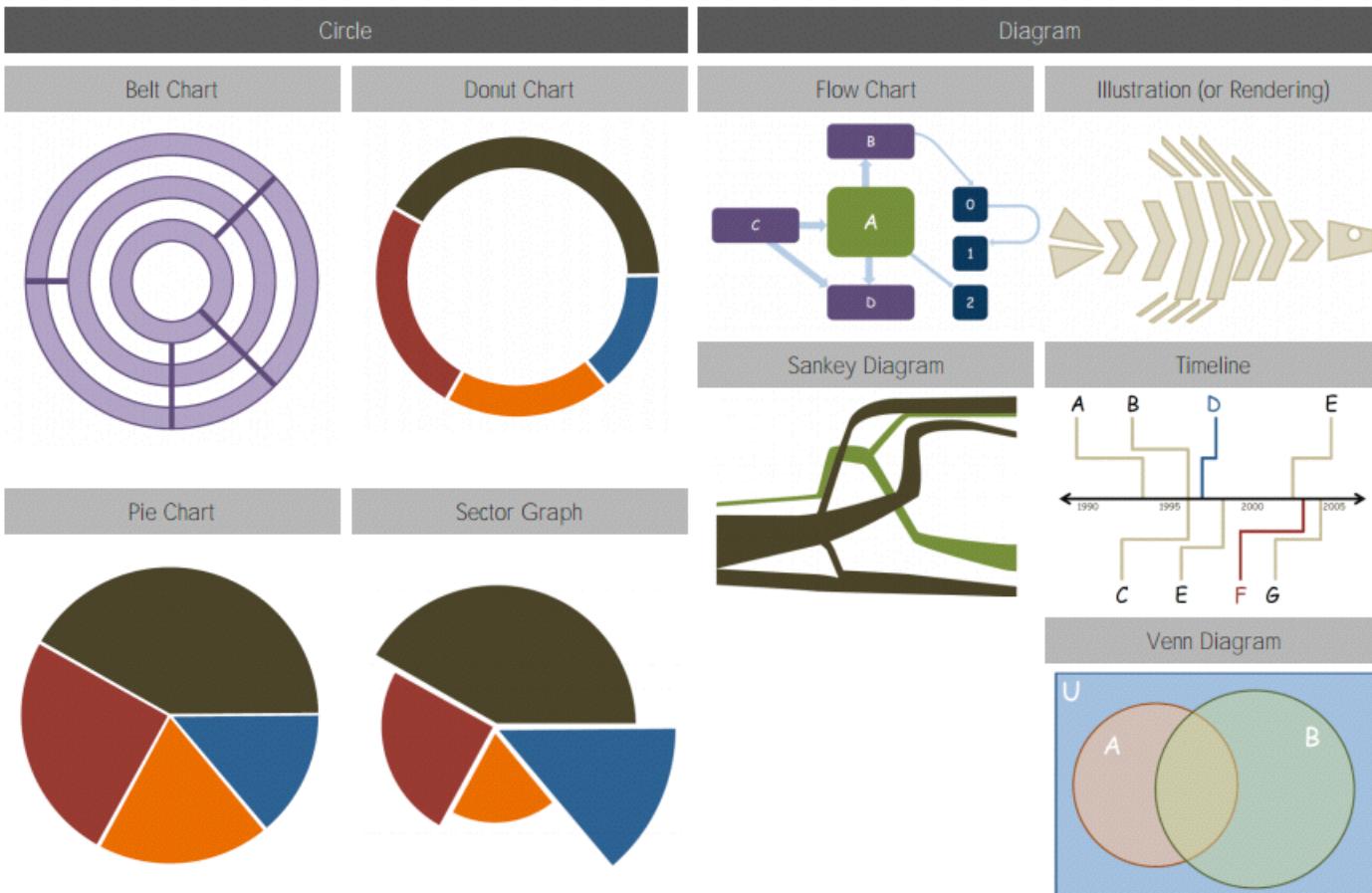
VISUALIZATION TAXONOMY

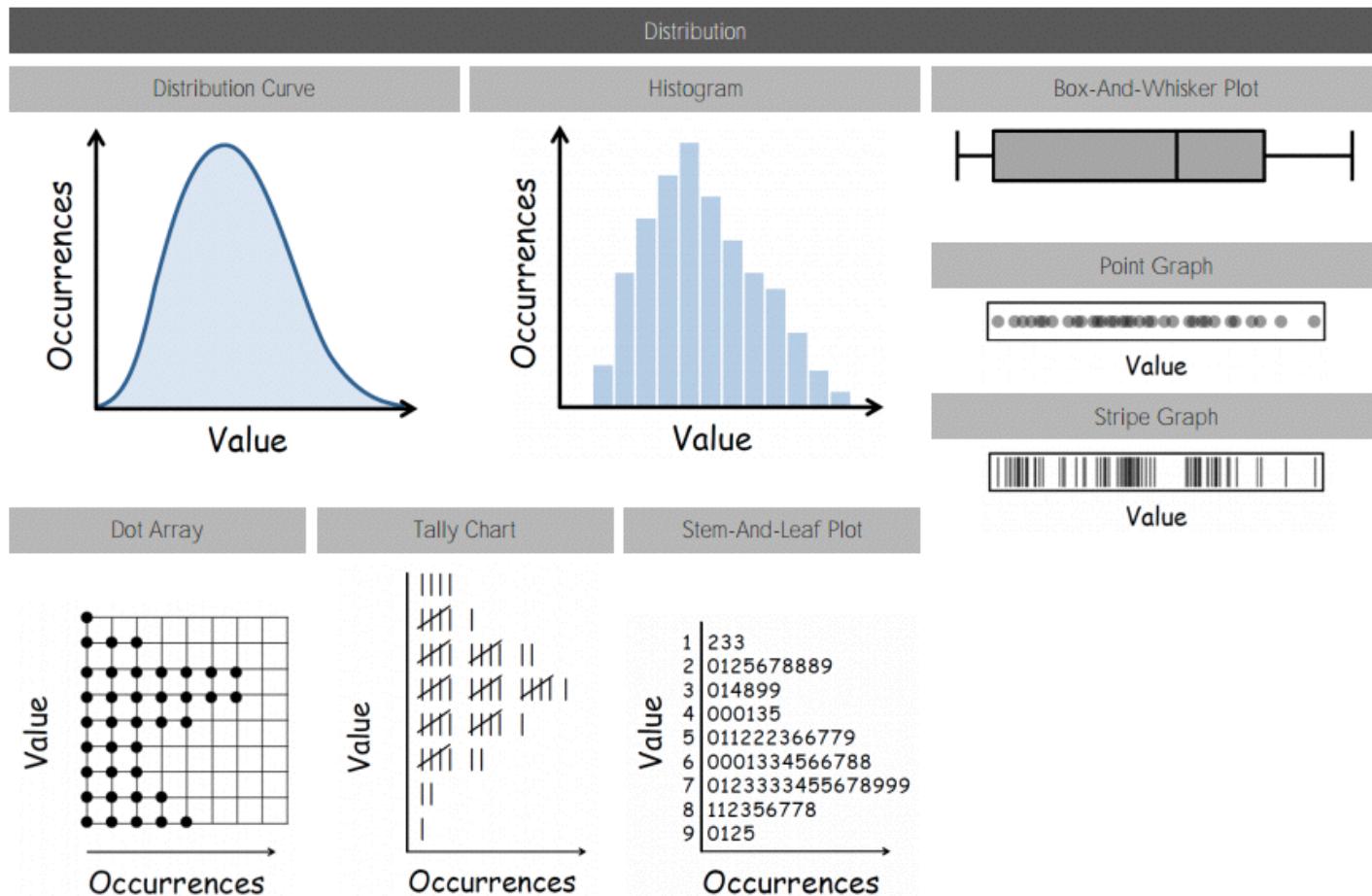
Area

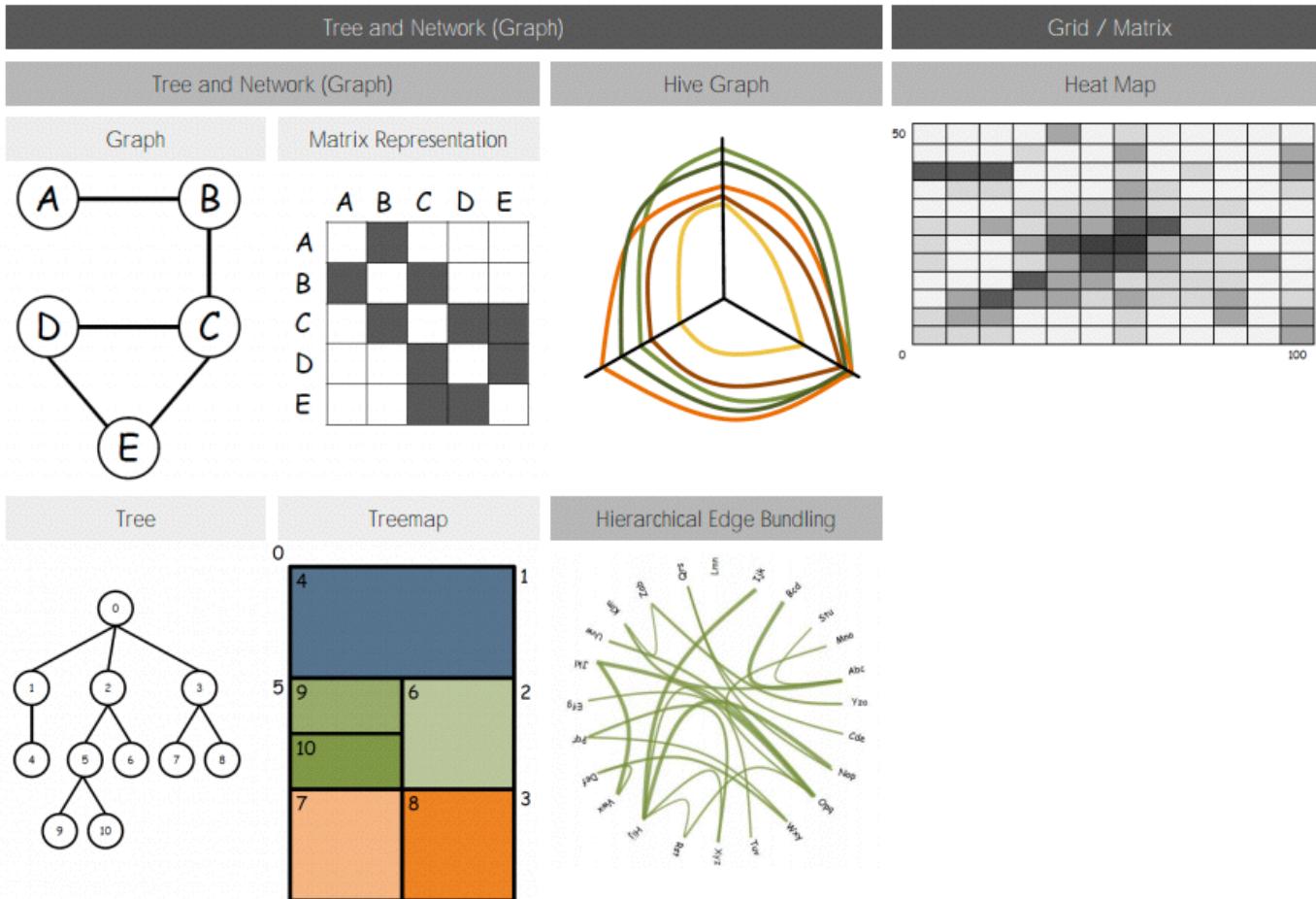


Bar



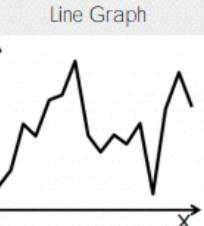




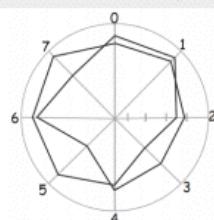


Line

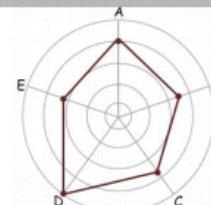
Line Graph



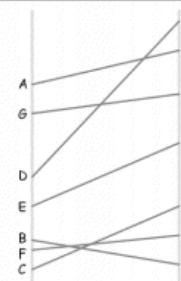
Circular Line Graph



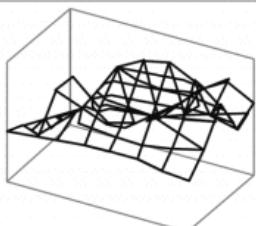
Star Plot



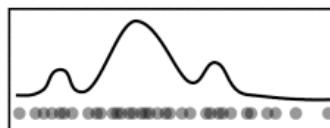
Slopegraph



Surface Graph



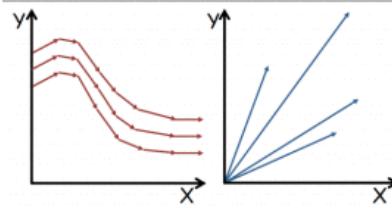
Density Plot



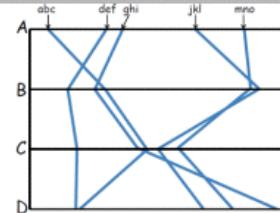
Surface Graph



Vector Graph

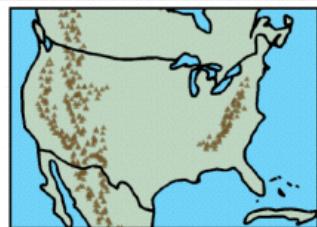


Parallel Coordinates

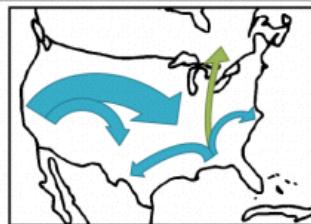


Map

Geographic Map



Flow Map

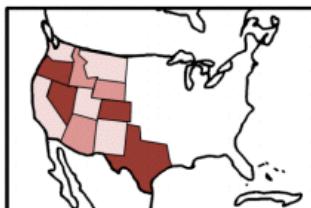


Statistical Map

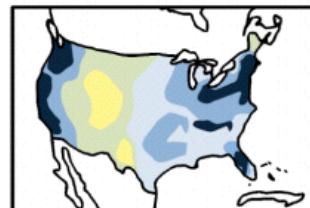
Street Map



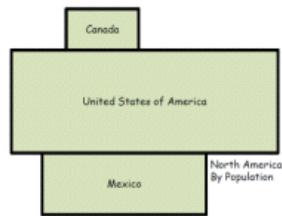
Choropleth Map



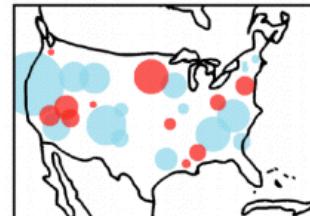
Contour Map (Isopleth)



Distorted Map (Cartogram)



Statistical Plot Map



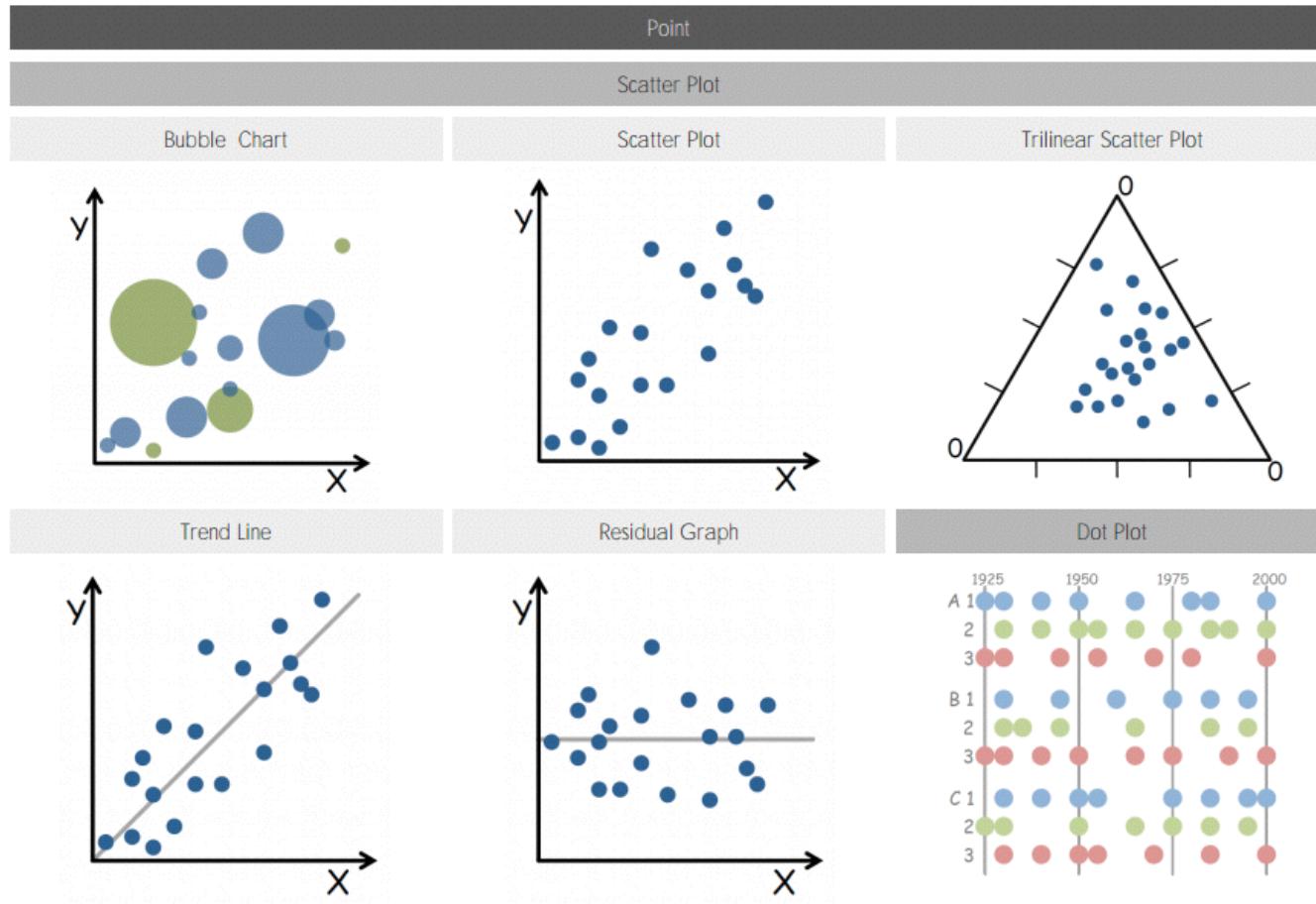


Table			Text Based		
Table			Phrase Net		
ABC	1234	X45	omnis espiciatis unde fugit iste sit voluptatem natus laudantium accusantium doloremque vitae dicta eaque error sunt rem aperiam	quae laudantium doloremque eaque aperiam	
Category	543.2109	7%			
Group	45.67	45%			
Unit	9876	98%			
Class	123.78	12%			

Text Chart

Title	Heading 1
<ul style="list-style-type: none"> Sed dignissim vehicula Nisl quis congue Sed vitae rhoncus odio Integer at odio 	<p>"Nunc aliquam turpis at tellus varius hendrerit. Ut nec magna tortor. Proin adipiscing dolor eget odio semper ut commodo lacus imperdiet."</p> <p>- Lorem</p>
Heading 2	Heading 3
<p>Aenean tincidunt sem vel massa cursus non tempus quam auctor. In nisi mi, commodo sit.</p> <p>Amet rutrum vitae, fringilla non urna. Quisque sagittis ultrices sapien, quis pesuere massa interdum quiz.</p>	<ul style="list-style-type: none"> ✓ Chart 1 ✓ Chart 2 ✓ Chart 3 ✓ Chart 4

Word Cloud

word word word
 word **text** word
 word word
 word

Word Tree

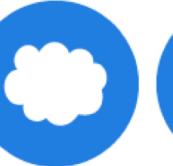
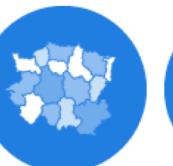
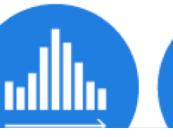
the
 process of graphical use of study
 visualization techniques are of involves uses the computer

Visual Taxonomy

The Data Visualisation Catalogue

About • Suggest • Shop • Resources

Search by Function View by List

 Arc Diagram	 Area Graph	 Bar Chart	 Box & Whisker Plot	 Brainstorm	 Bubble Chart
 Bubble Map	 Calendar	 Chord Diagram	 Choropleth Map	 Circle Packing	 Connection Map
 Donut Chart	 Dot Map	 Dot Grid	 Flow Map	 Histogram	 Line Chart

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

Takeaway Messages

- Appropriate chart type for specific data type and visualization task