S. Rinzivillo – rinzivillo@isti.cnr.it

DATA VISUALIZATION AND VISUAL ANALYTICS

TAXONOMY OF VISUAL VARIABLES

Cleveland McGill [1984]

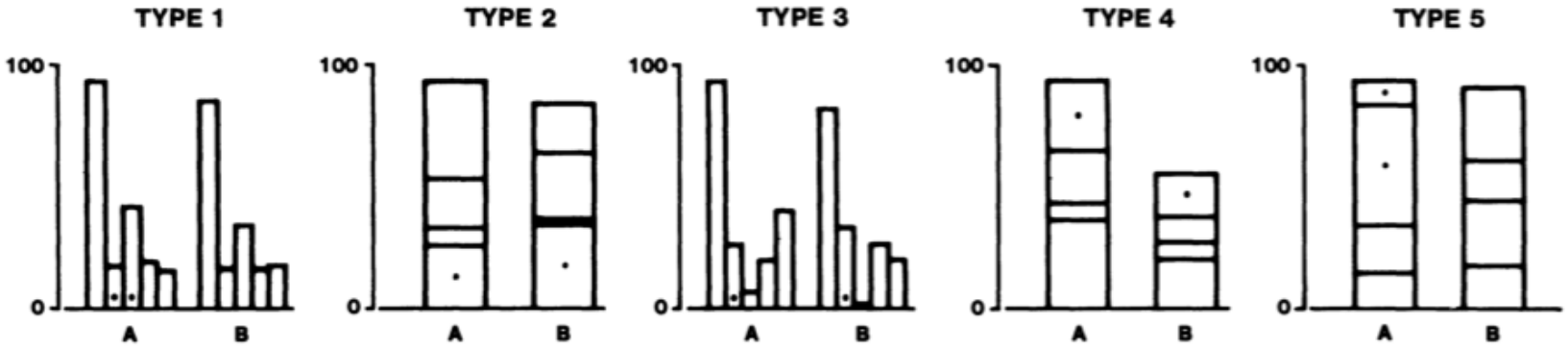


Figure 4. Graphs from position–length experiment.

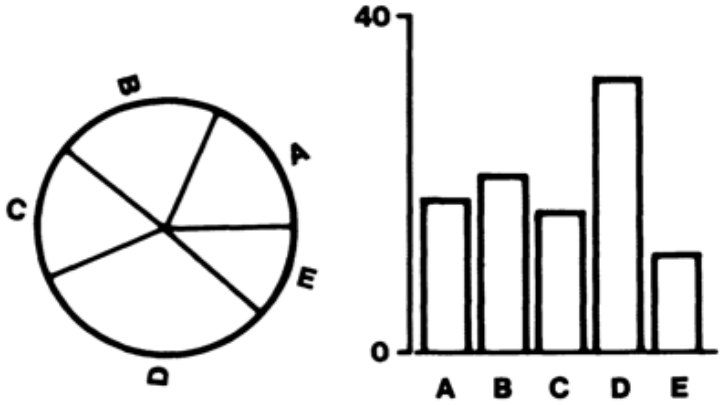
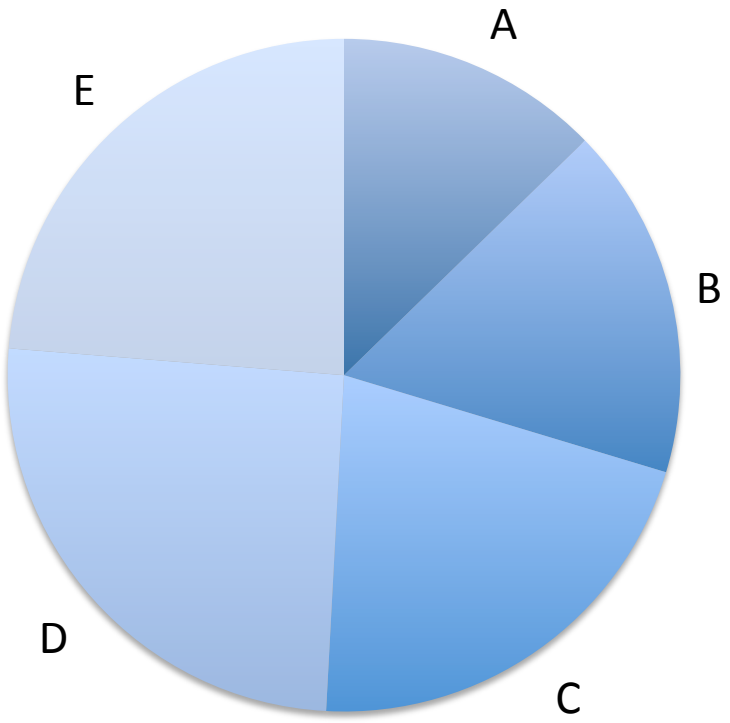


Figure 3. Graphs from position–angle experiment.

Cleveland & McGill: graphical encodings

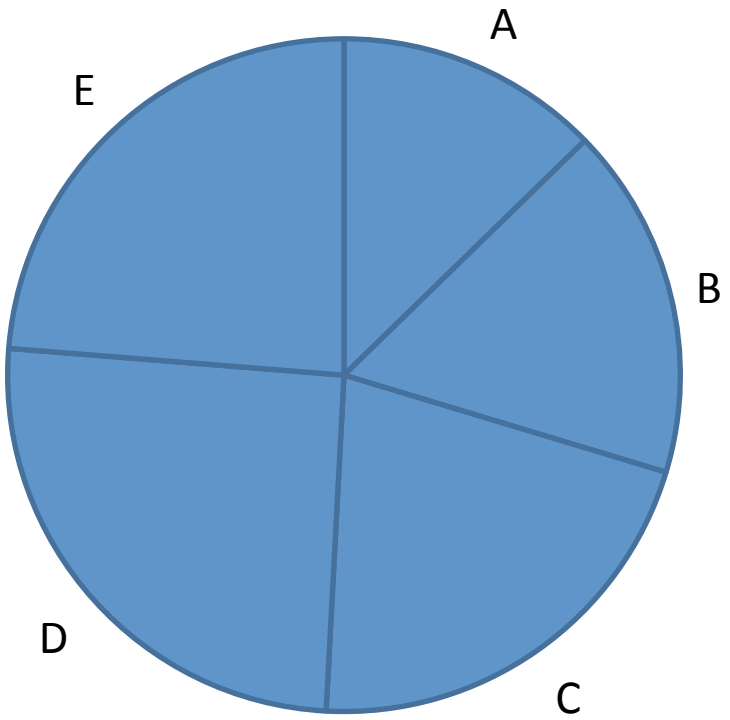
- Angle
- Area
- Color Hue
- Color Saturation
- Density
- Length
- Position on a common scale
- Position on non aligned scale
- Slope
- Volume

Angle decoding



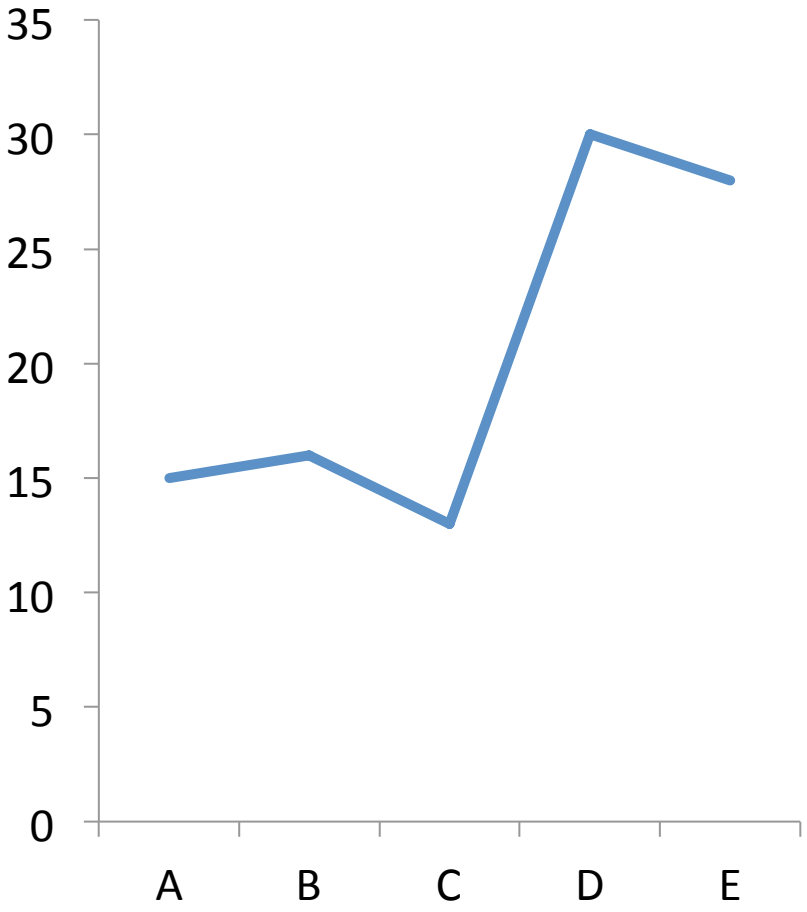
- It is difficult to compare angles
  - Underestimation of acute angles
  - Overestimation of obtuse angles
  - Easier if bisectors are aligned
- Area estimation helps

Angle decoding



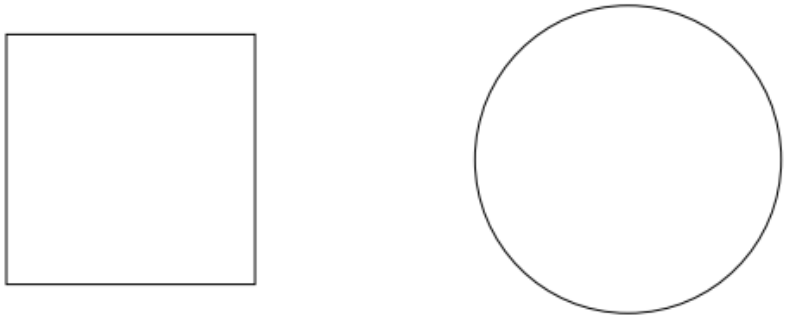
- It is difficult to compare angles
  - Underestimation of acute angles
  - Overestimation of obtuse angles
  - Easier if bisectors are aligned

Slopes Decoding



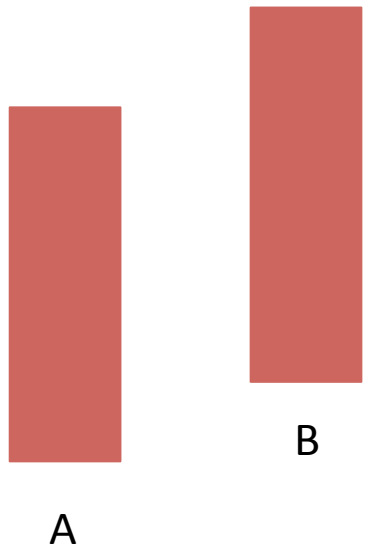
- Same difficulties as angles
- Easier task since one branch is aligned with x-axis

Area Decoding



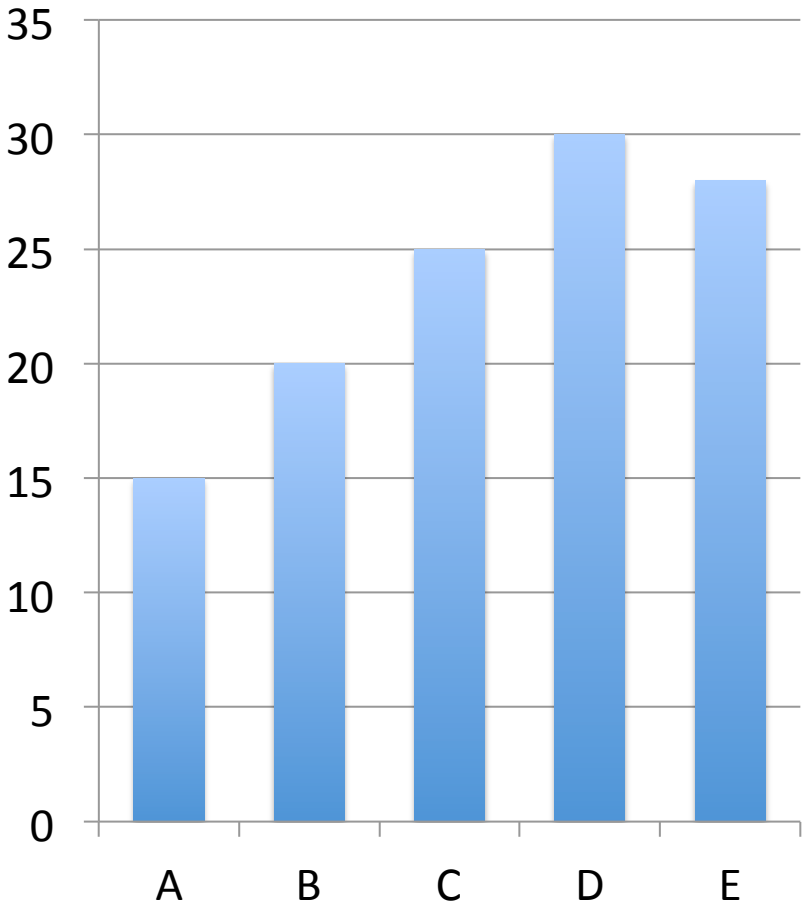
- Area is not well decoded
  - Different regular shapes
  - Irregular shapes
  - Context influences (thin area within compact thick area)

Length Decoding



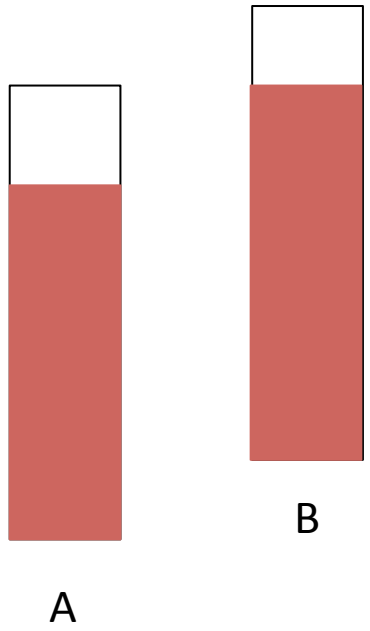
- Straight forward to endoce numerical values
- Difficulties with relative lengths

Position on a common scale



- Widely used in statistical charts

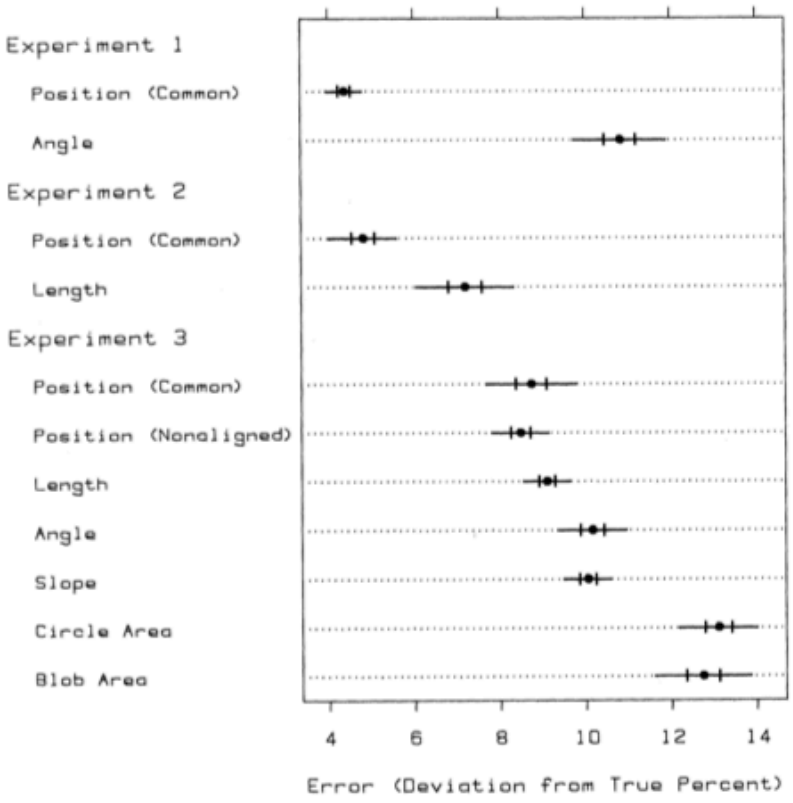
Position on non-aligned scale

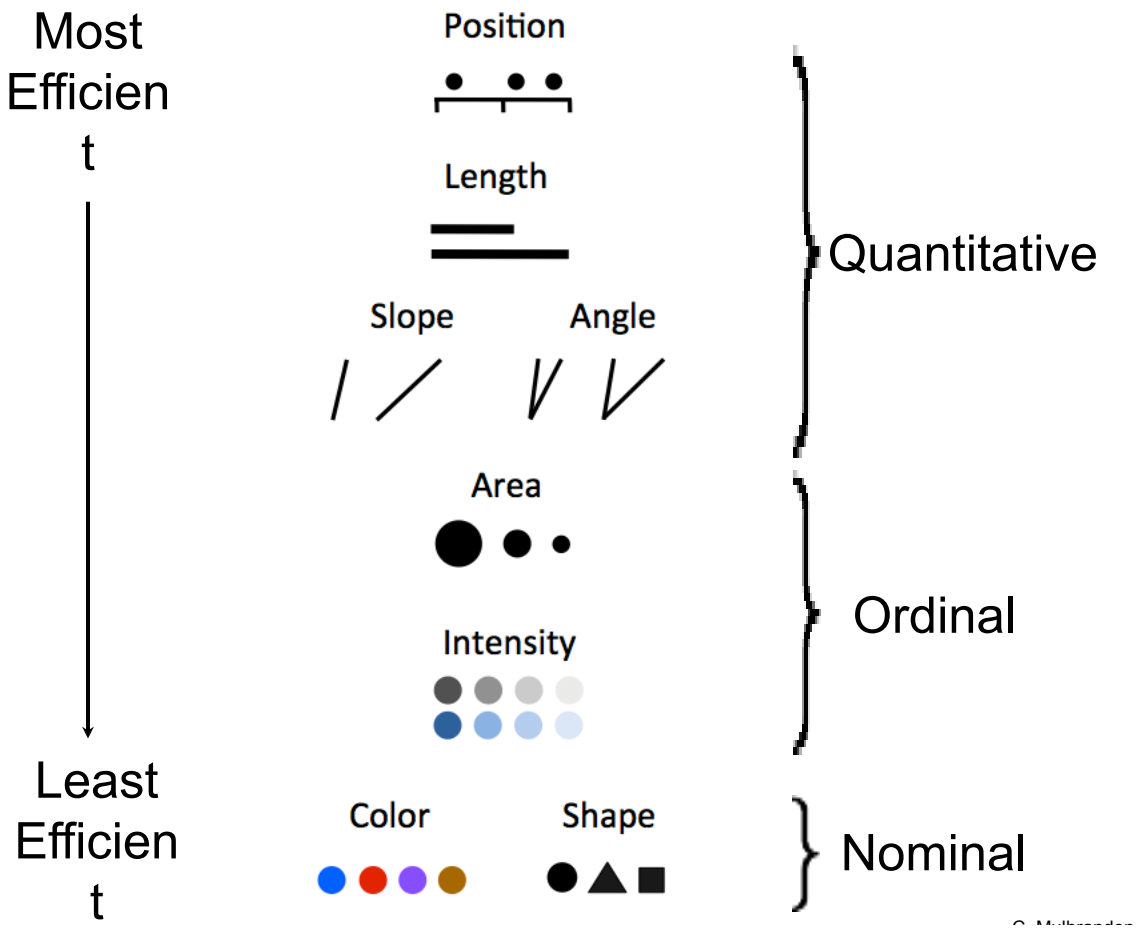


- Not as bas as common scale
- Still acceptable

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




Most Efficient
Least Efficient
Position
Length
Slope
Angle
Area
Intensity
Color
Shape
Quantitative
Ordinal
Nominal

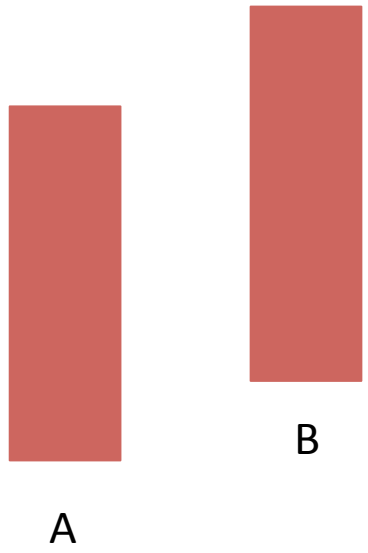
PERCEPTION LAWS

Weber's Law

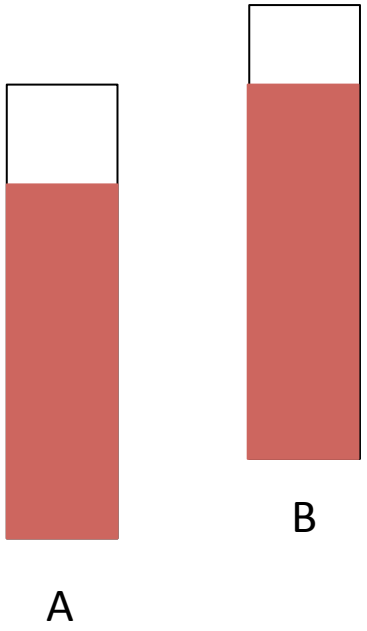
- Just-noticeable difference between two stimuli is proportional to their magnitudes
- Case study on length
  - Given two lines with lengths x and $x+w$
  - If w is small, it is difficult to notice difference between the two lines
  - If w is larger, it is easier to catch the difference
- How large should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

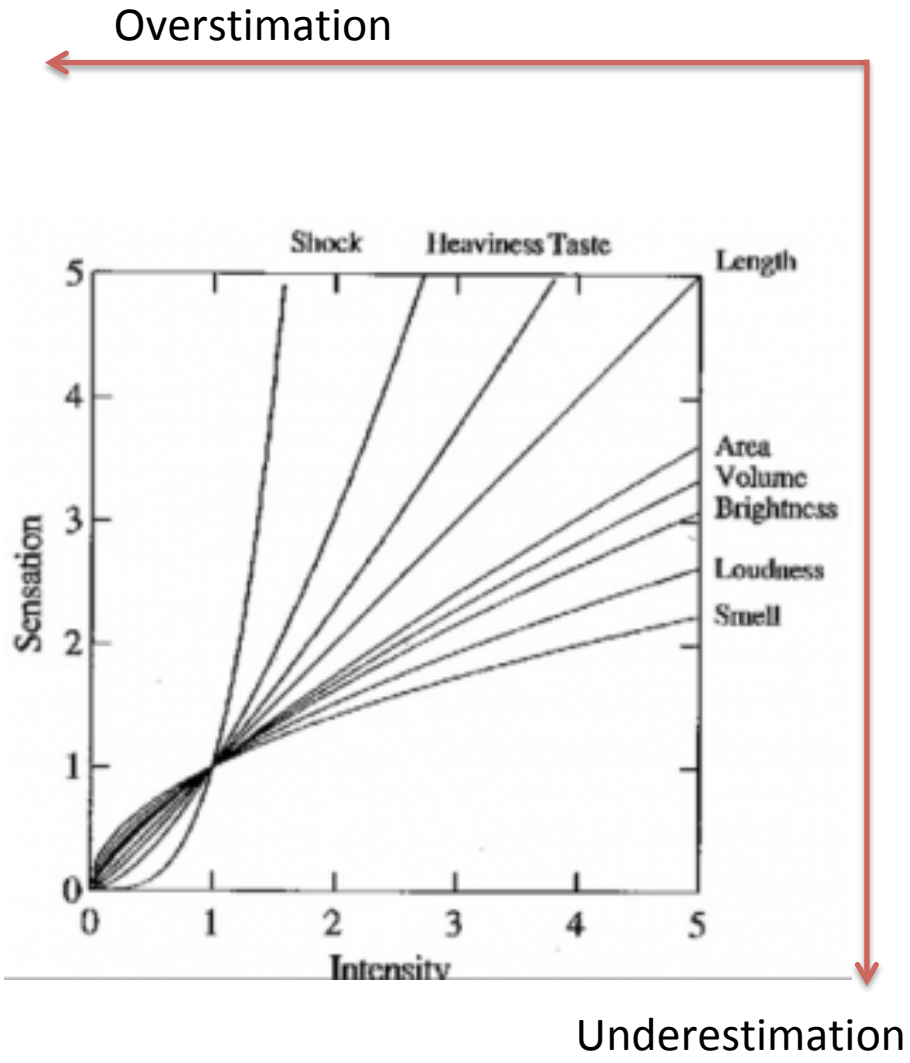


Stevens' Law

- 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 β=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 β=0.7
- Let x_1=0,5 and x_2=1
- The perceived difference will be

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

Weber and Stevens' Laws: areas vs lengths

- For areas β=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 β=1
- Let $x_2=x_1+w$
- The perceived difference will be

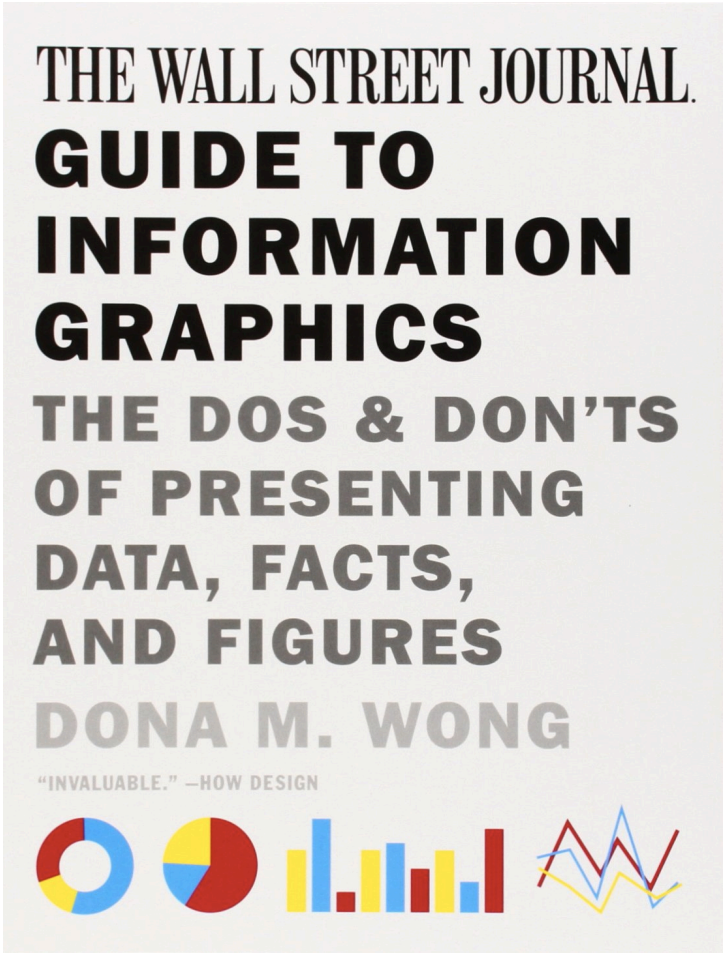
$$\left(\frac{x+w}{x}\right)^{1} = 1+\frac{w}{x}$$

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



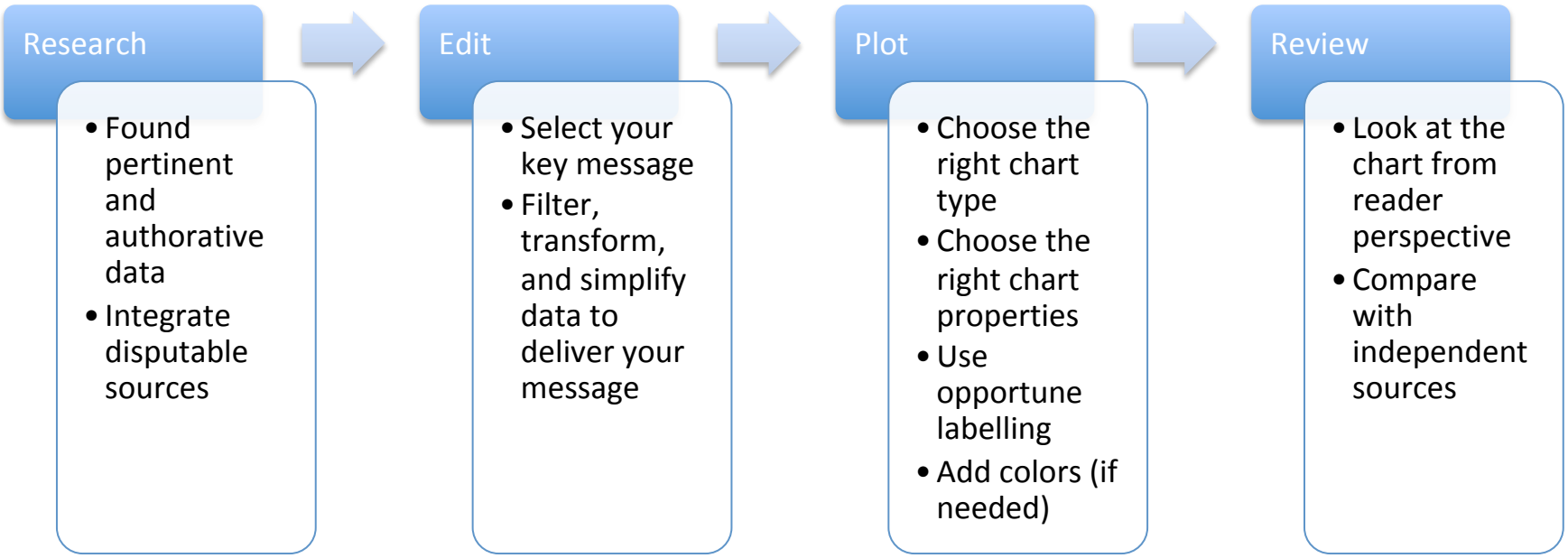
Dona M. Wong

Guide to Information Graphics

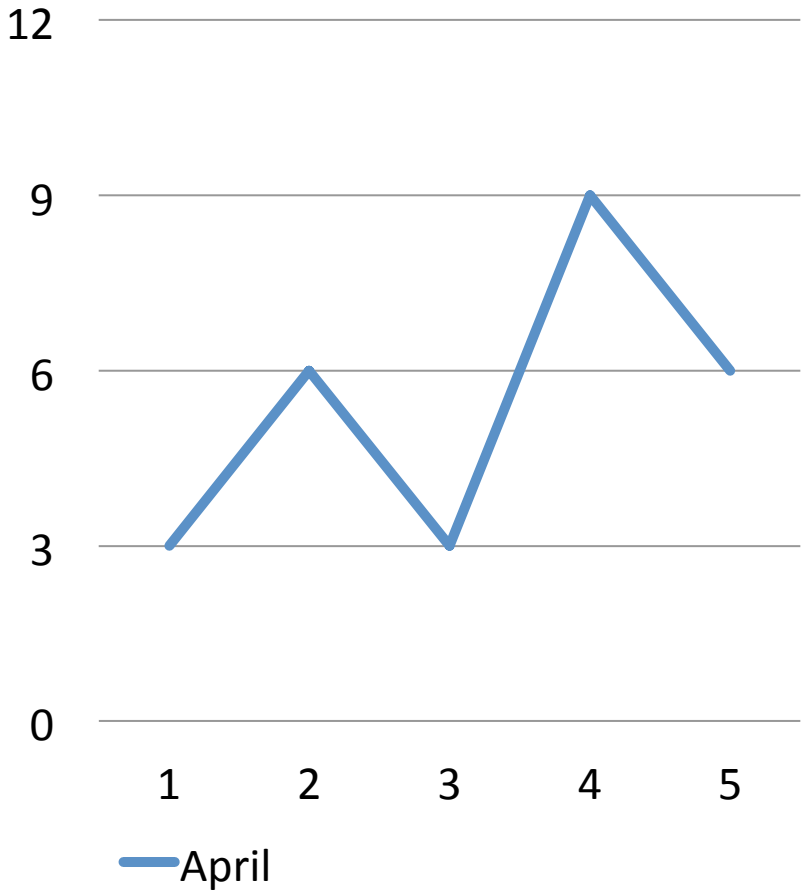
The Dos and Don'ts of Presenting Data, Facts, and Figures

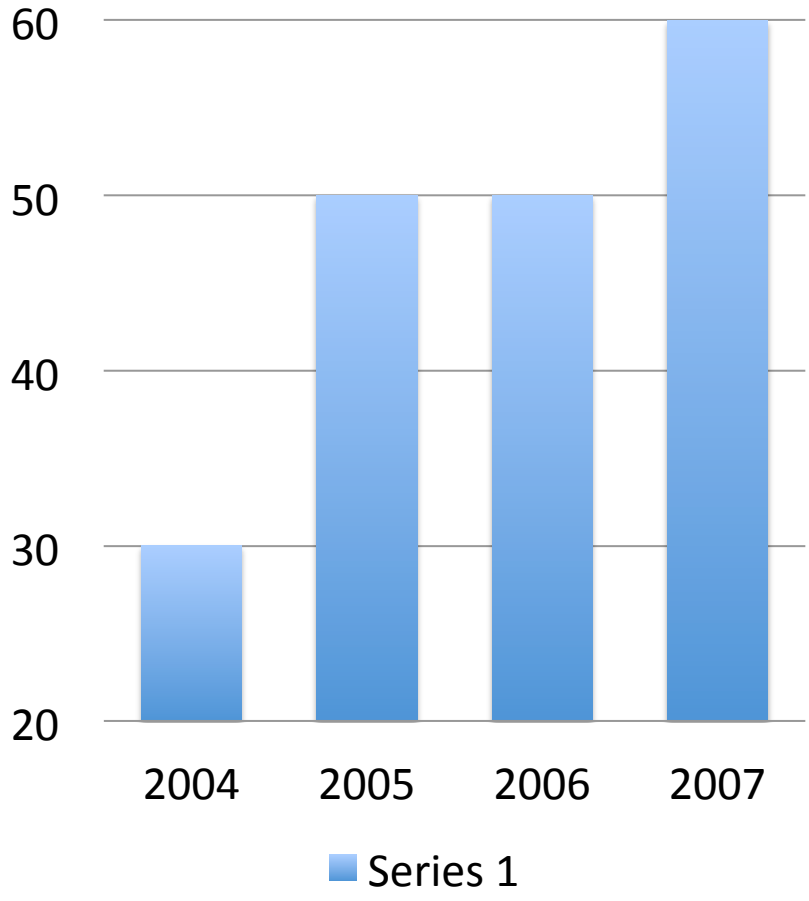
W. W. Norton & Company

Charting Pipeline



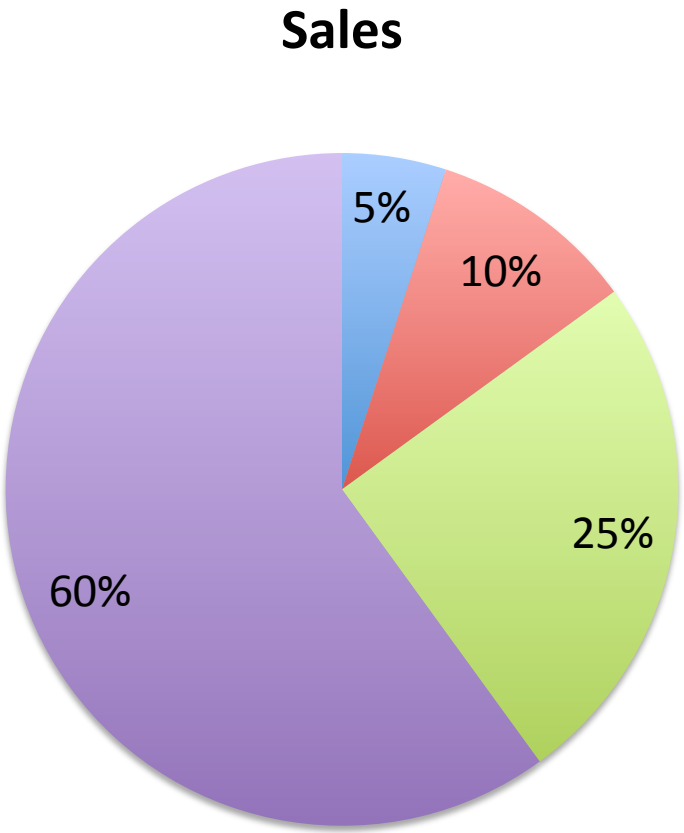
Charting Examples

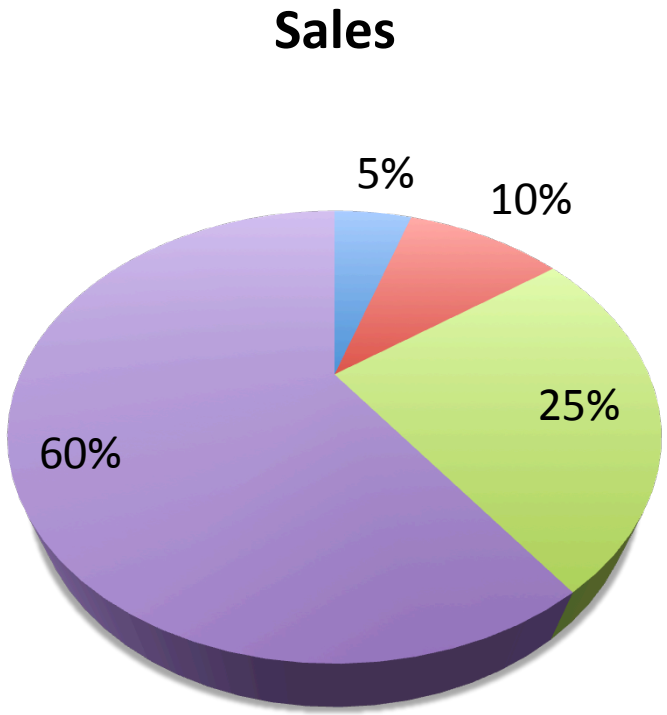




May these charts be improved? Why? How?

Charting Examples

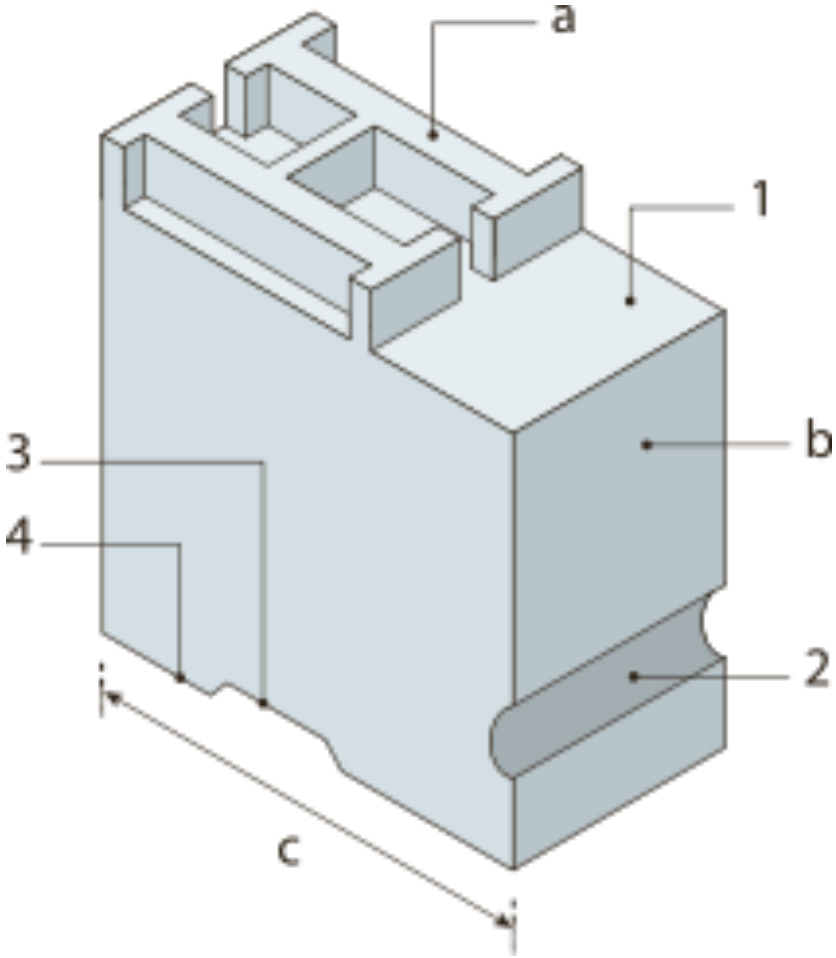




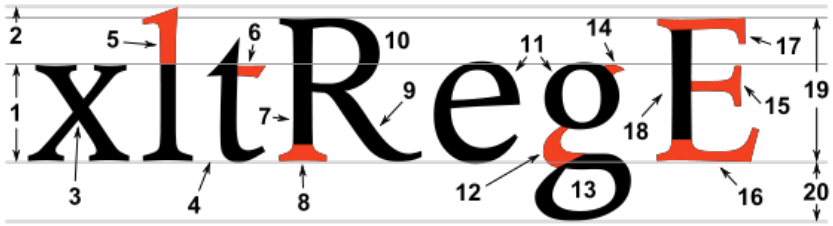
May these charts be improved? Why? How?

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.**

Font size = (1) + (2) + (20)
= (19) + (20)

"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

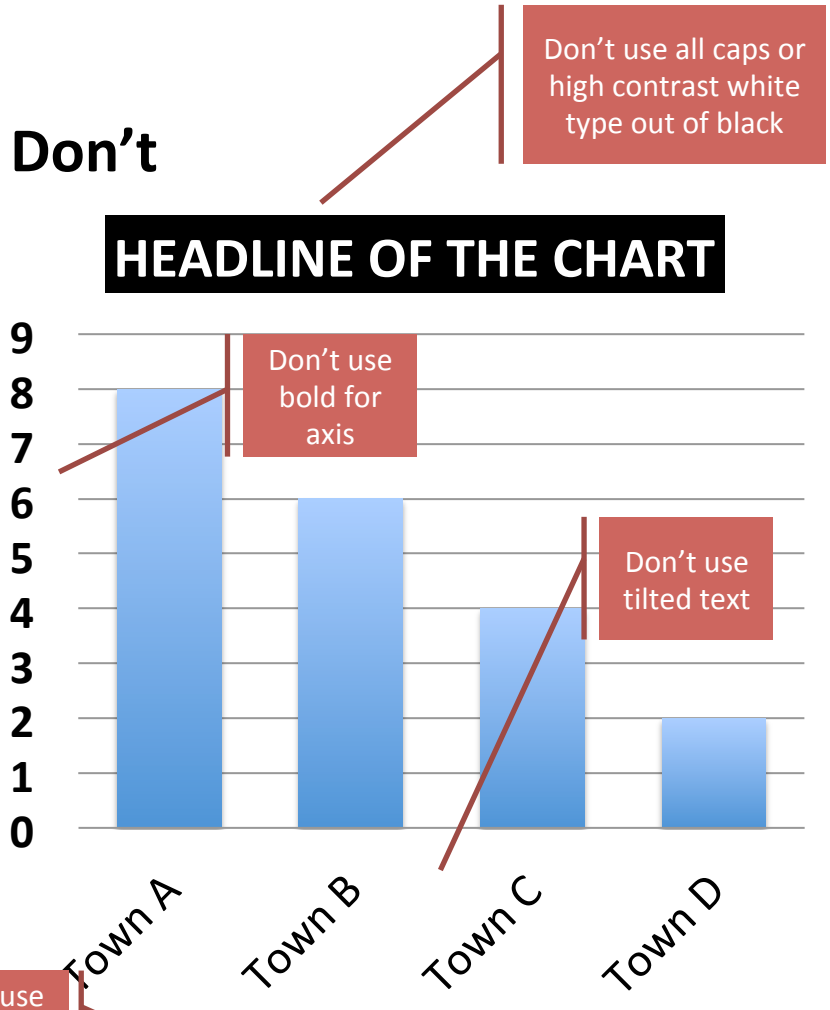
- Leading should be 2 points larger then type size
- Avoid too small or condensed type faces
- Keep style simple: use **bold** or *italic* to emphasize a word (better not ***both***)

- Avoid ALL CAPS
- Avoid *styled fonts*
- Avoid C***C Sans Serif
- Reduce type at an angle
- Avoid t r a c k i n g

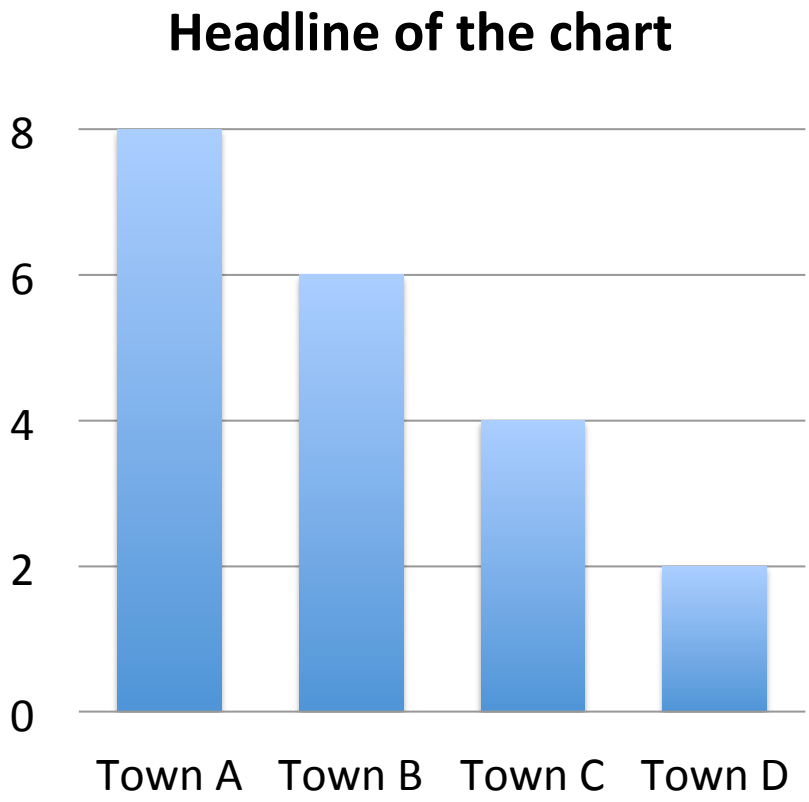
Fonts are meant to describe, not to adorn

Typography in Charts

Don't

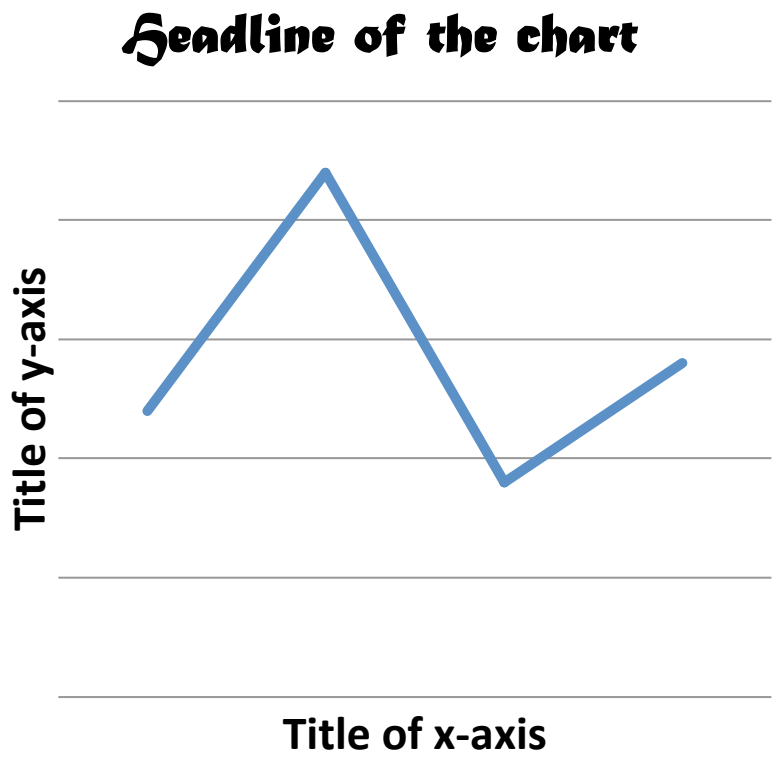


Do

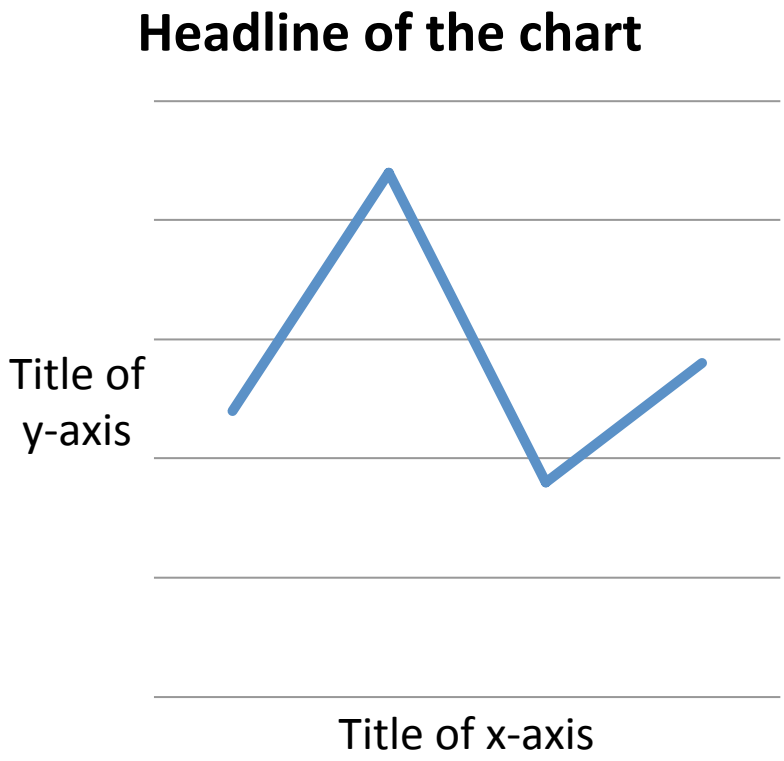


Typography in Charts

Don’t



Do



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**

Many elements in bold. Which part is highlighted?

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**

Give emphasis to relevant results

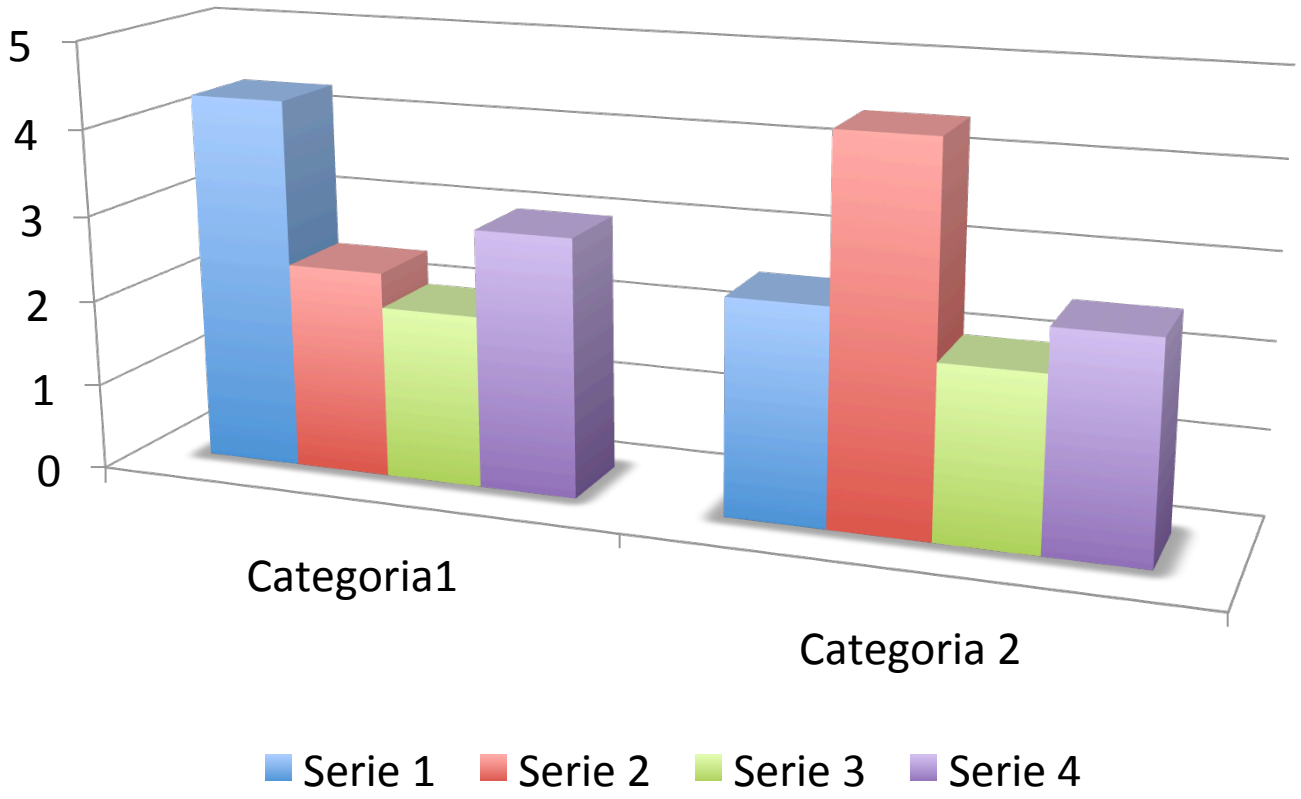
Visual Display of Quantitative Data

Edward Tufte, 1983

DATA-INK RATIO

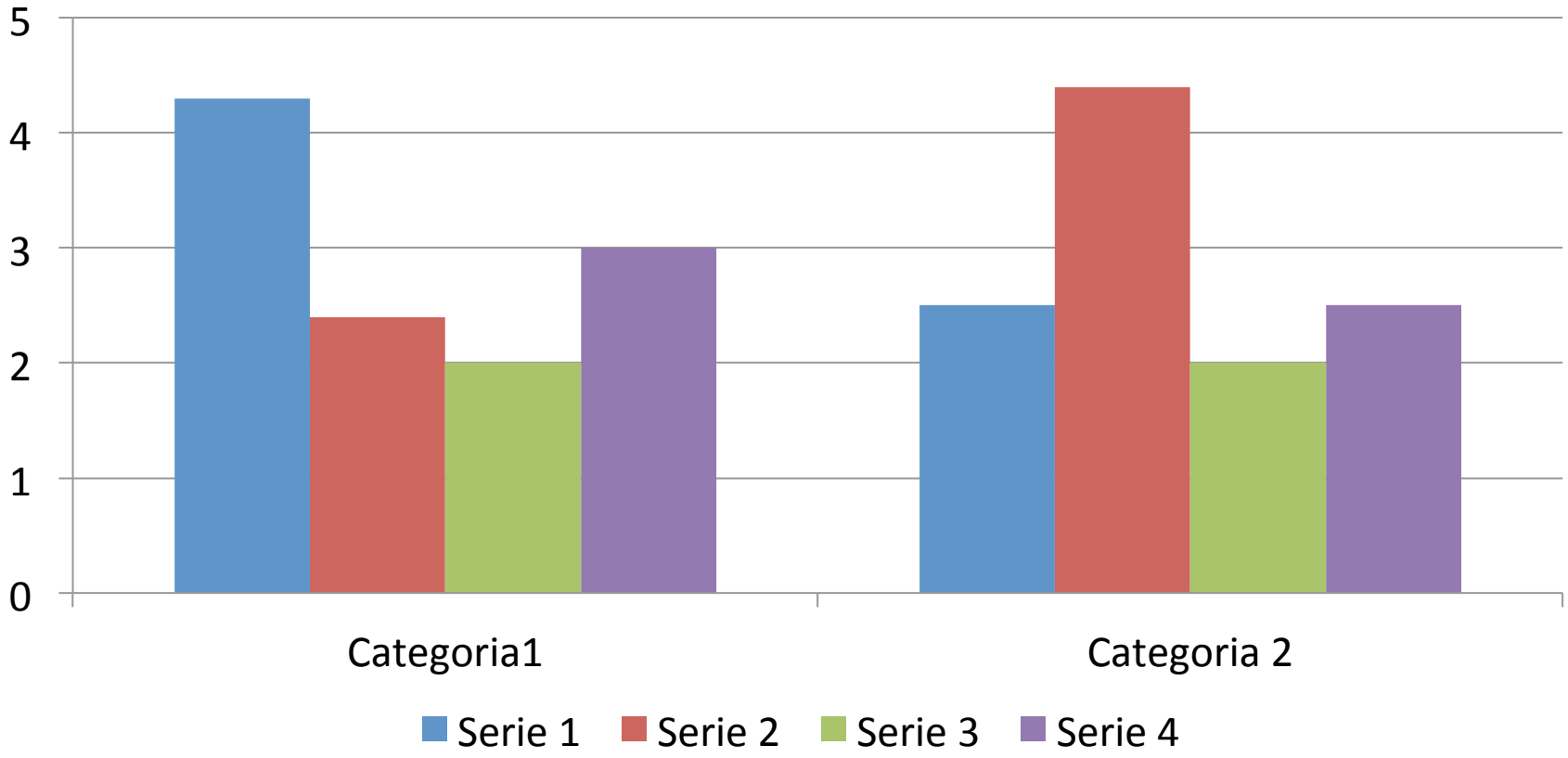
Data-ink Ratio

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



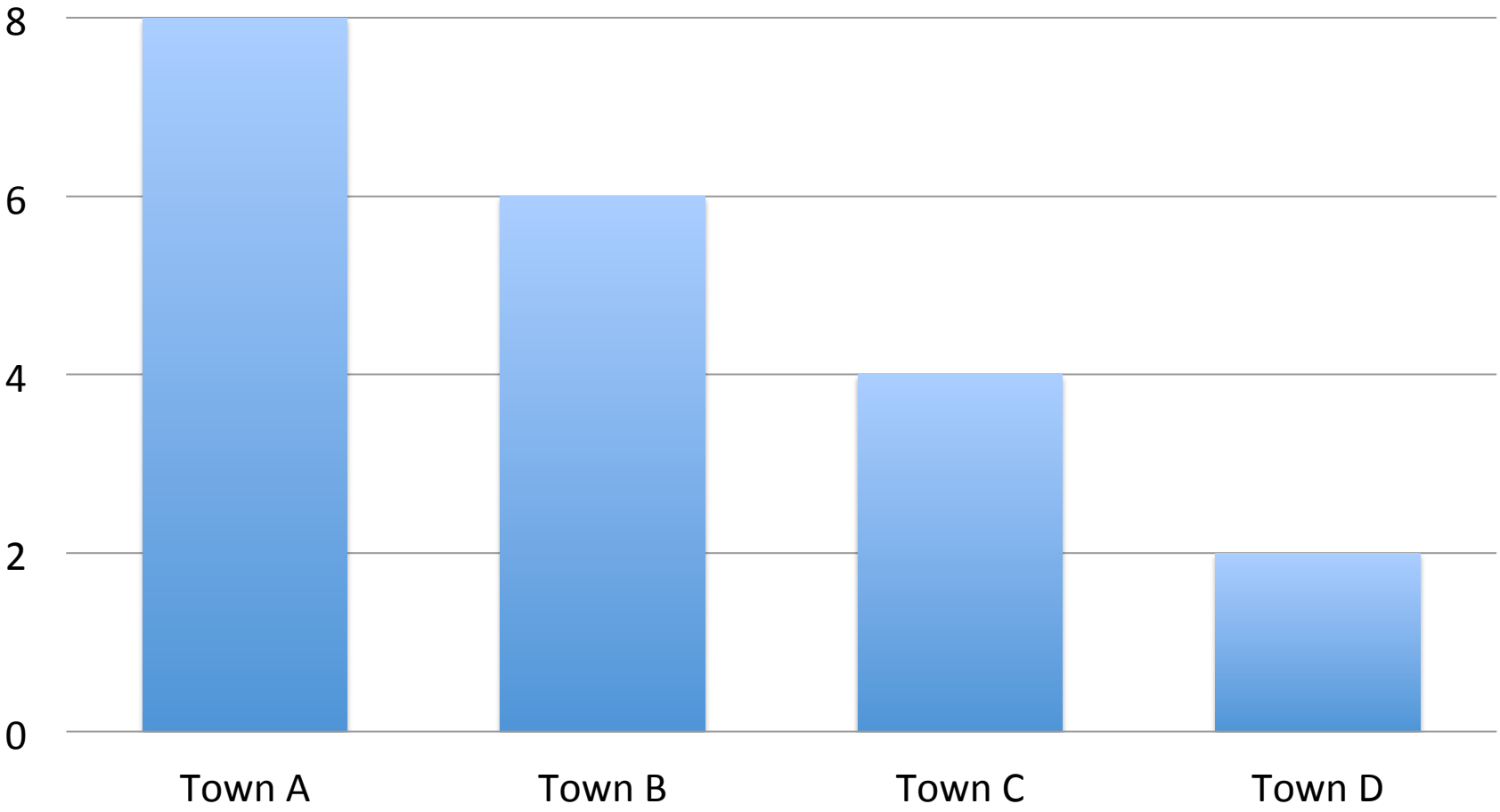
Data-ink Ratio

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



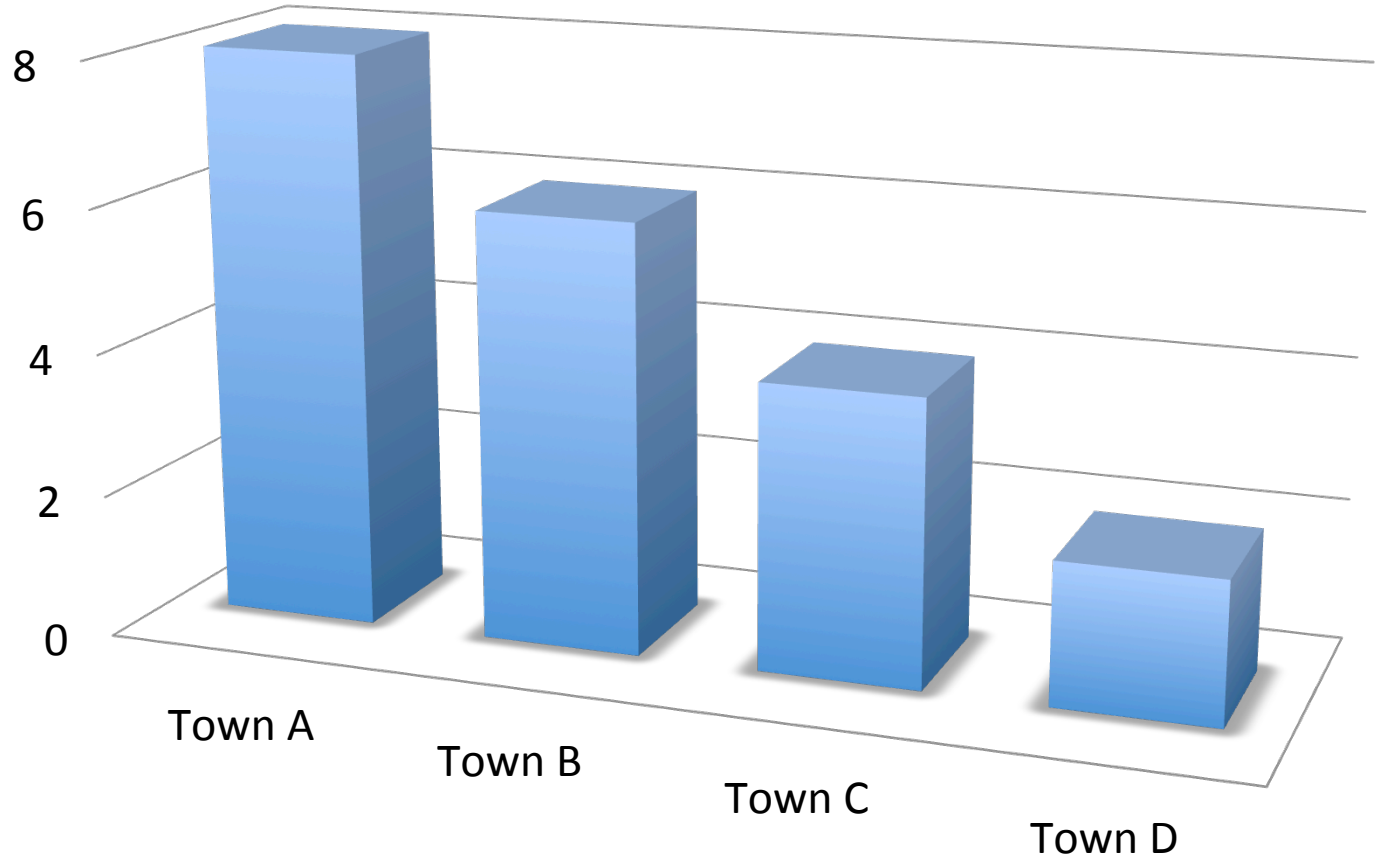
Bar Charts

Represent discrete quantities

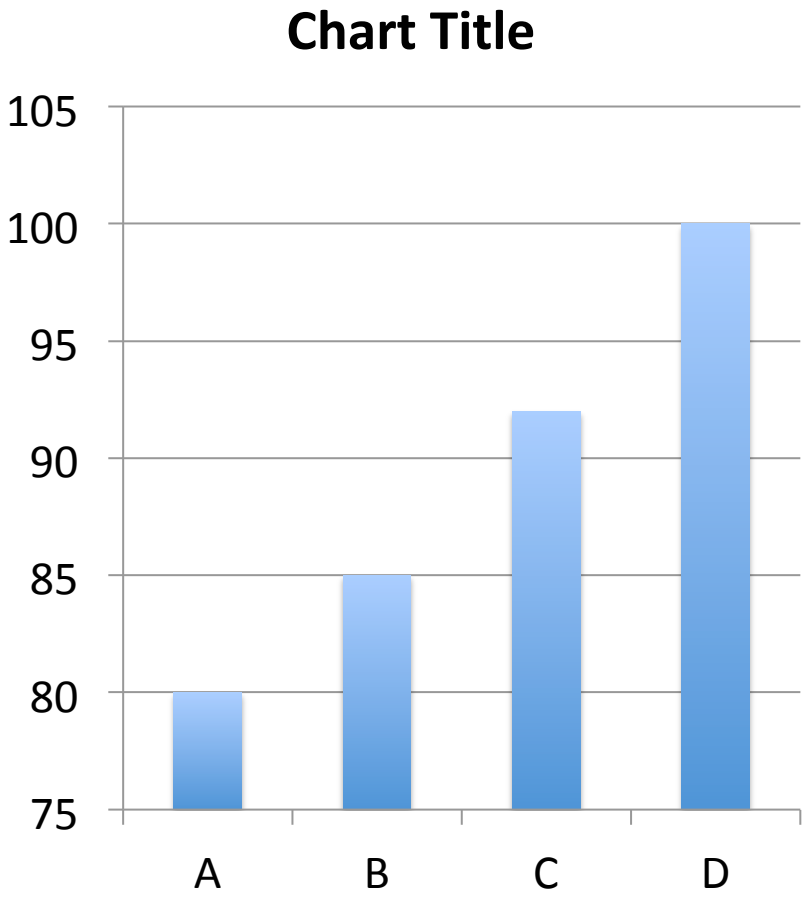


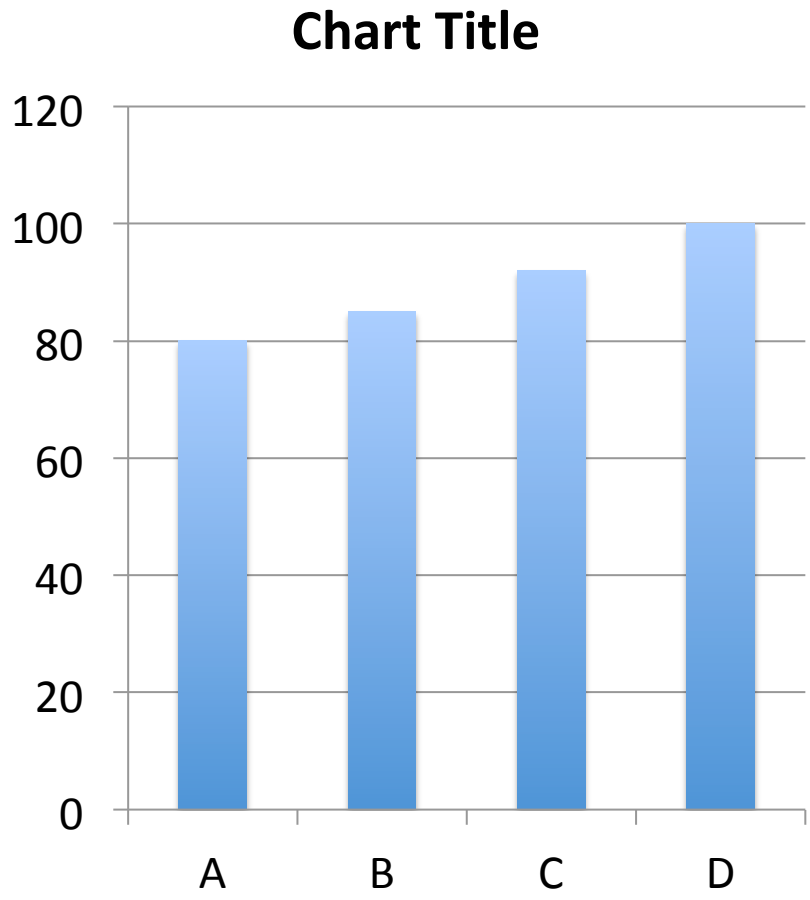
Bar Charts

Avoid non-functional adornation

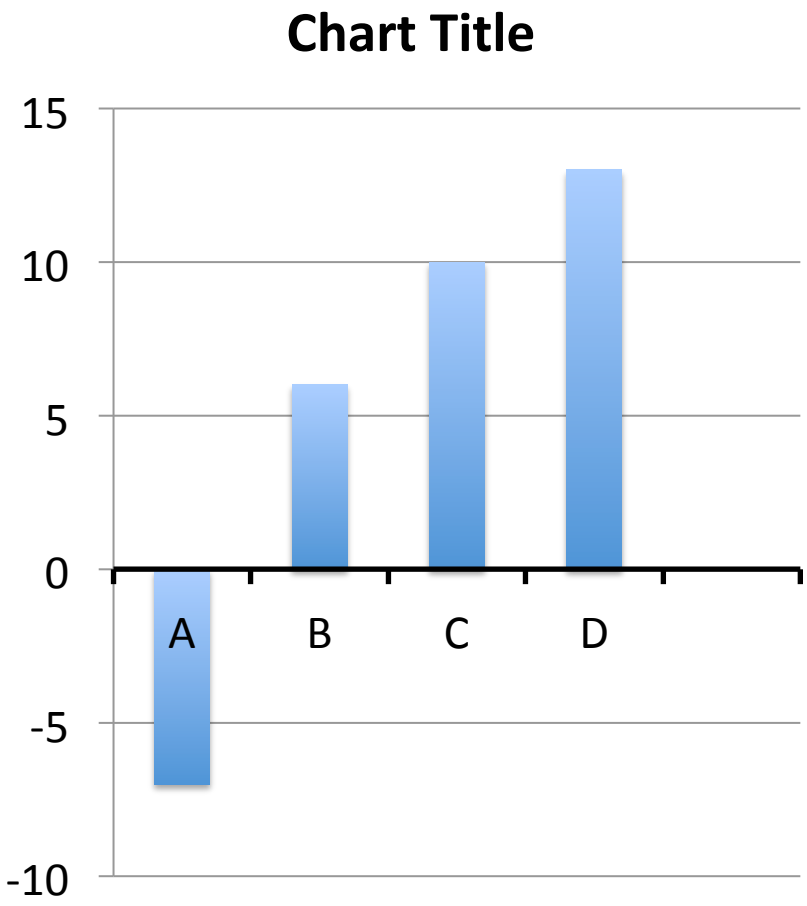


Bar Charts: baseline

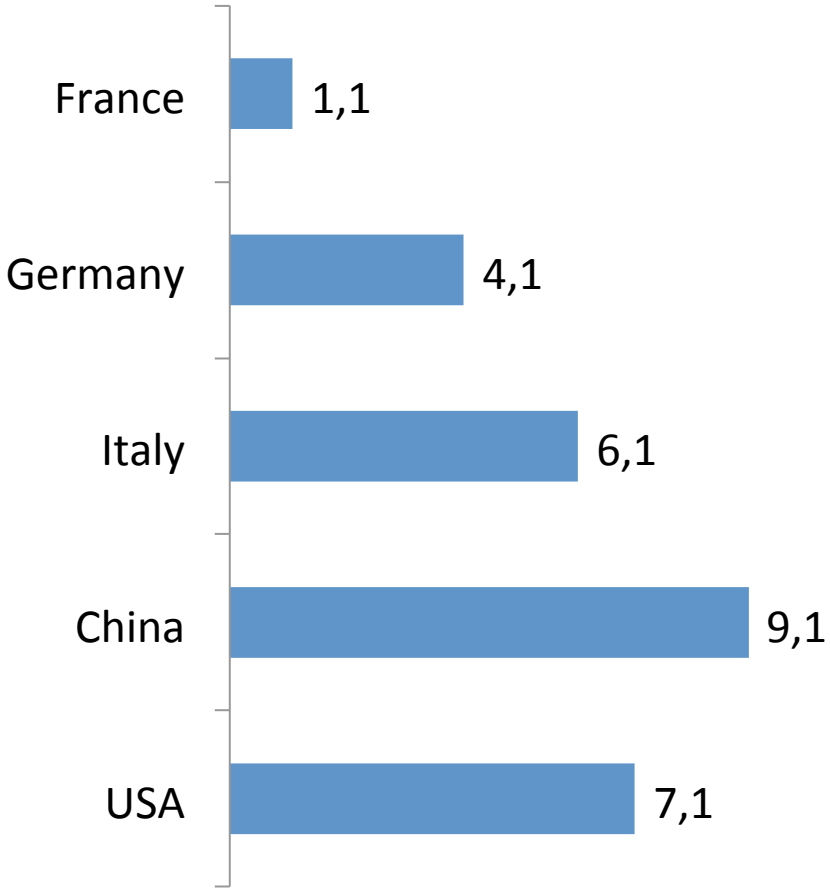


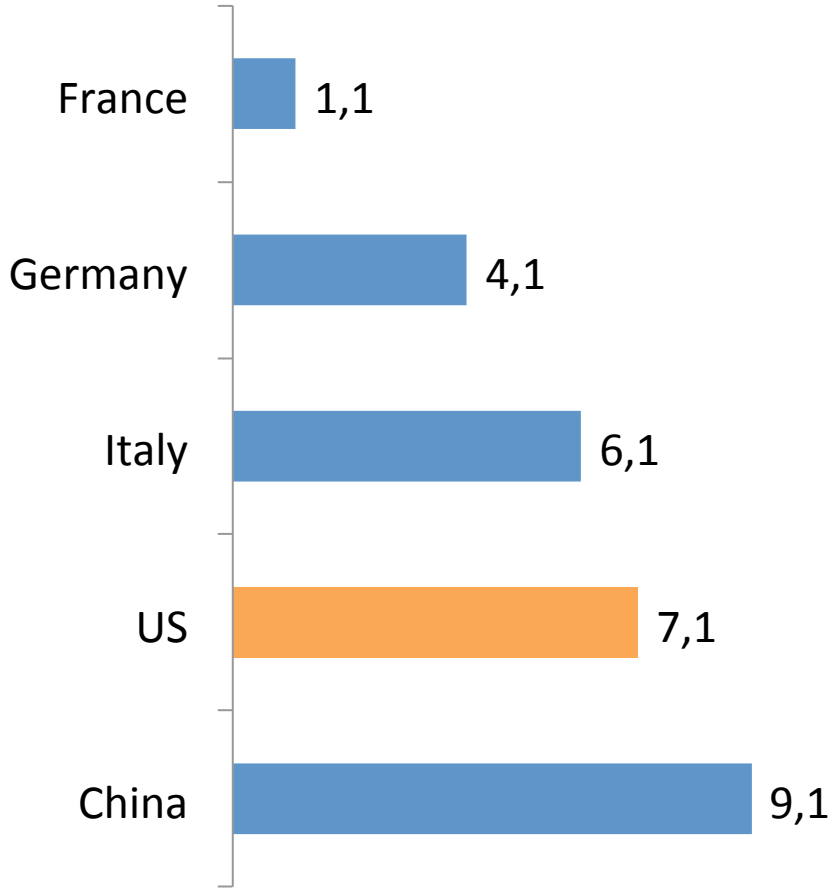


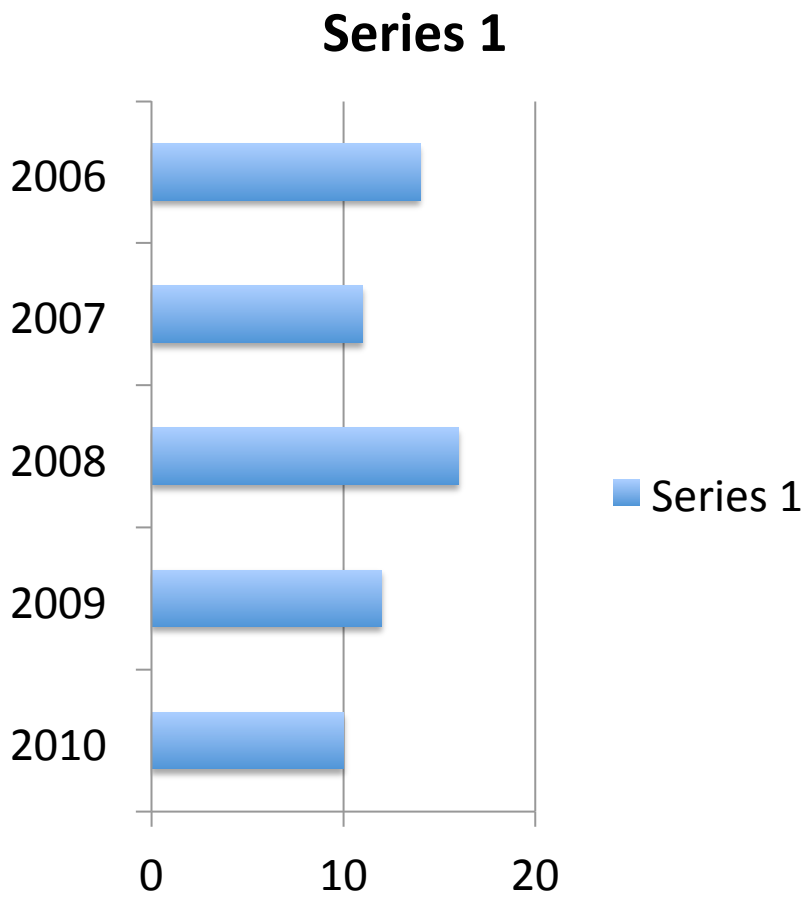
Bar Charts: baseline

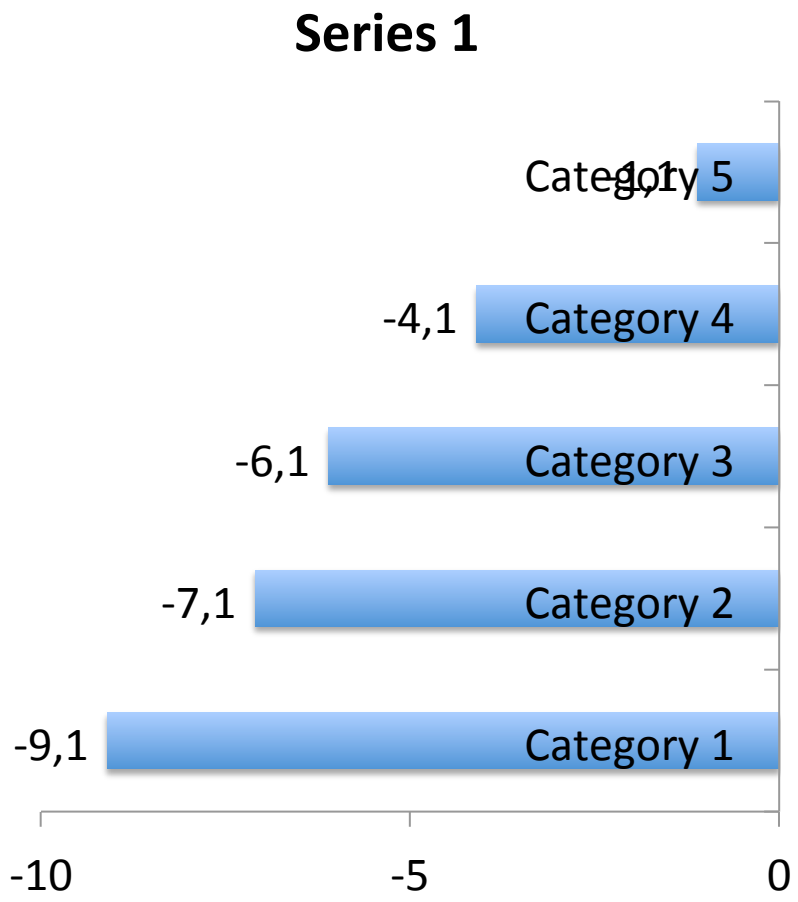


Bar Charts: ordering



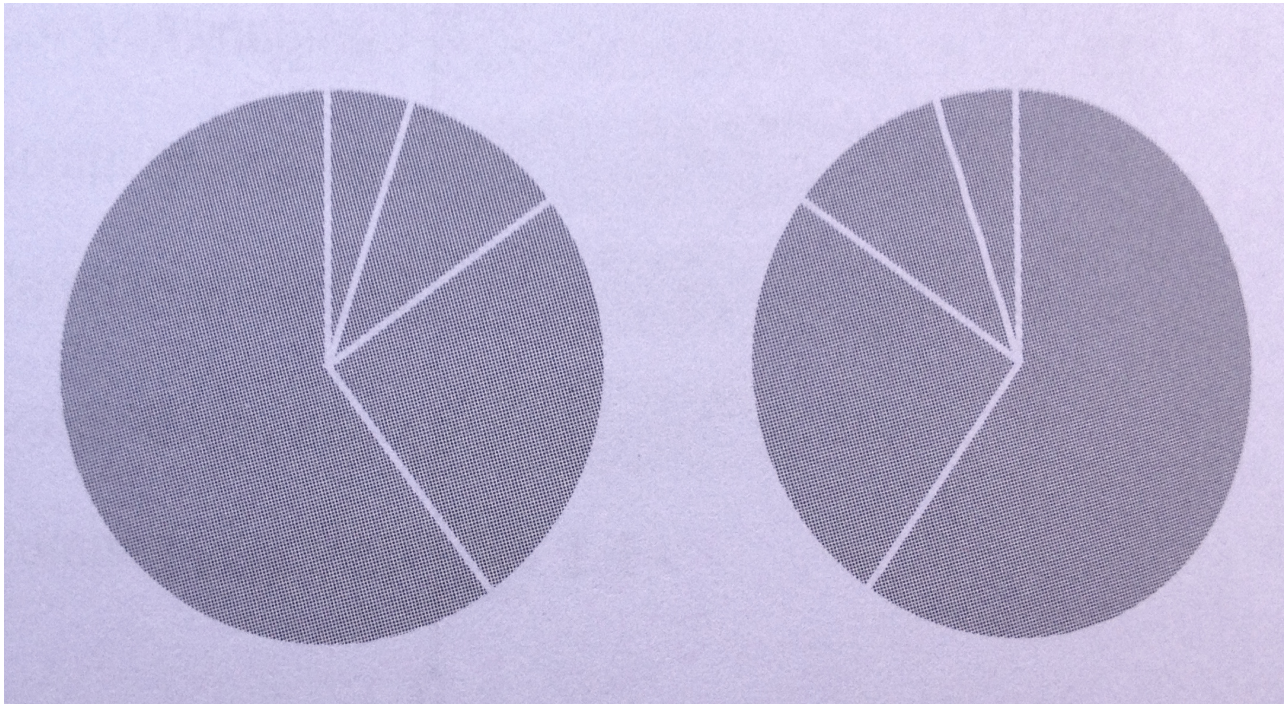



Series 1
2006
2007
2008
2009
2010
0
10
20
Series 1

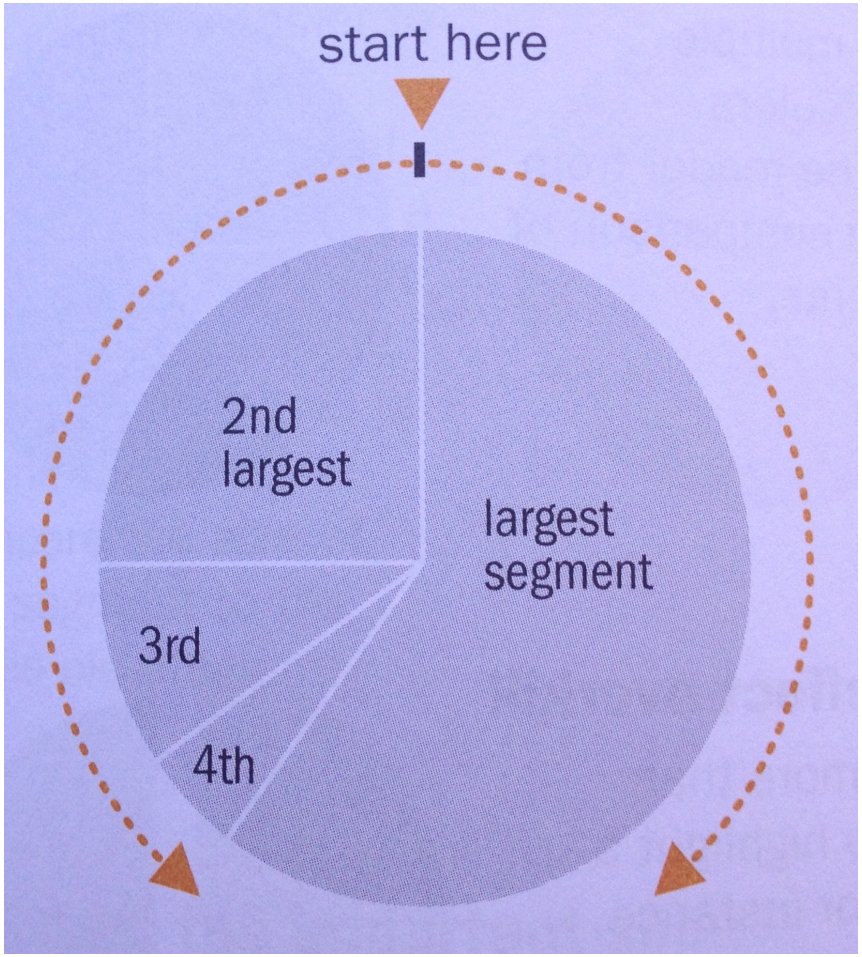

Series 1
-1,1
Category 5
-4,1
Category 4
-6,1
Category 3
-7,1
Category 2
-9,1
Category 1
-10
-5
0

Pie Charts

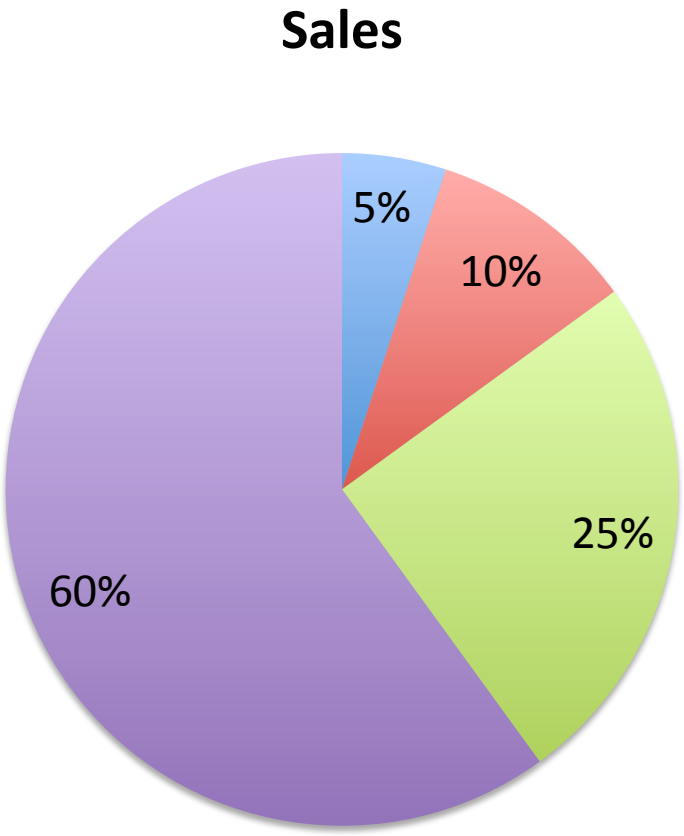
- Pie Charts compares relative sizes and contributions

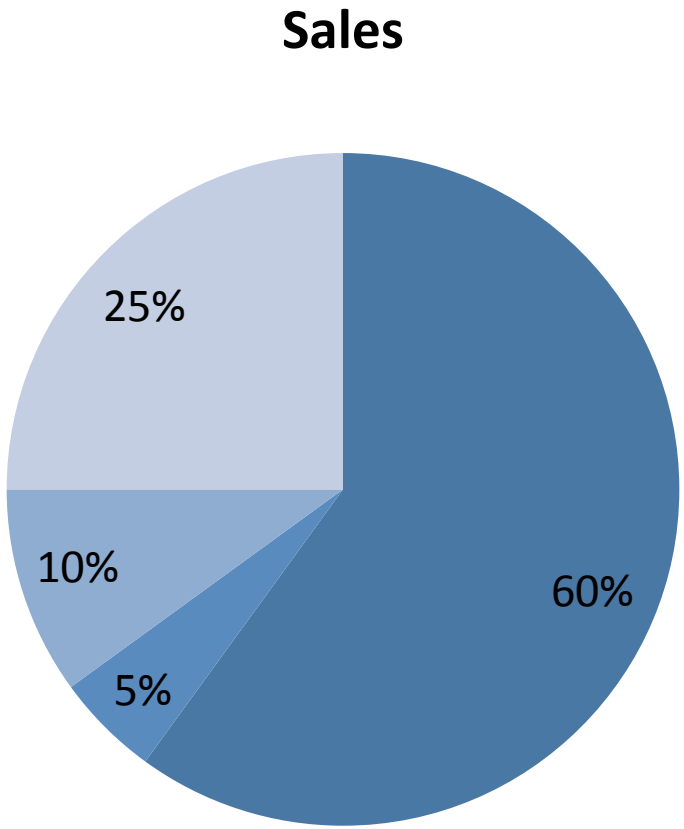


Pie Charts: ordering slices



Charting Examples





May these charts be improved? Why? How?

Takeaway Messages

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

Visualization Taxonomy

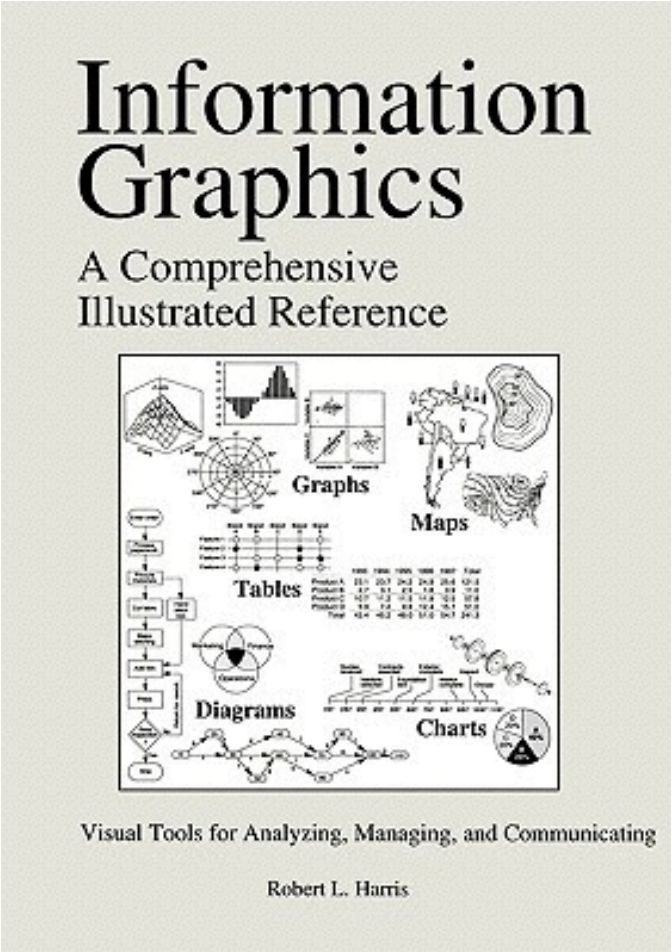
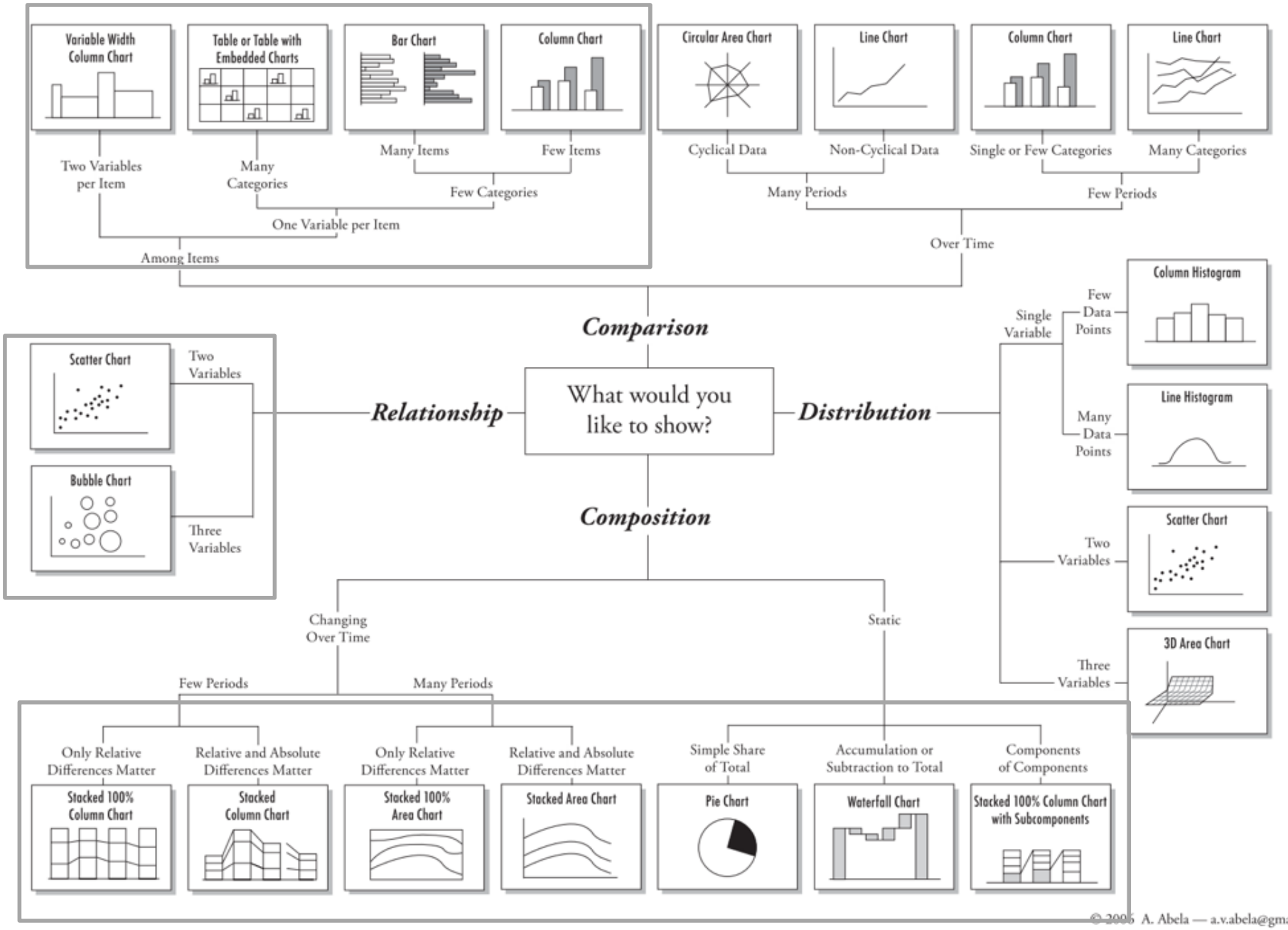
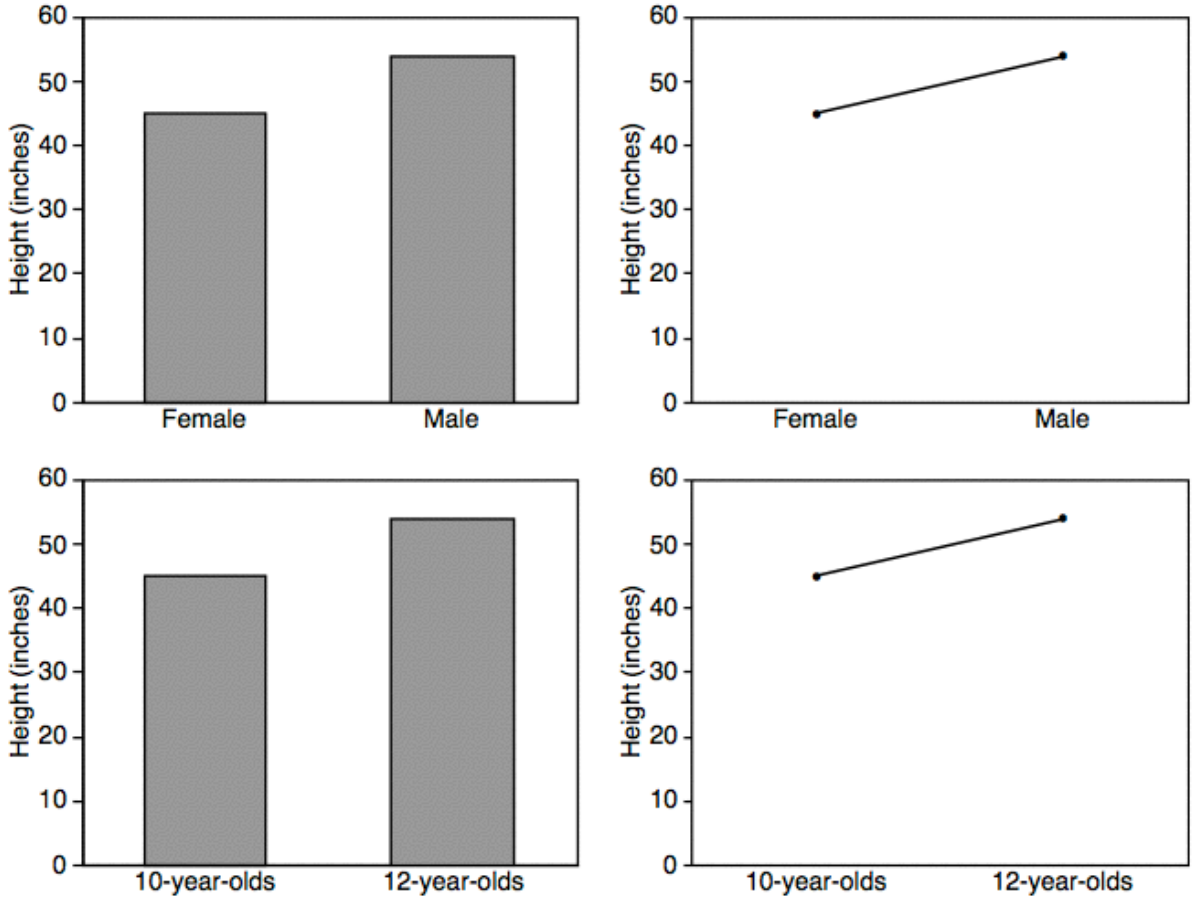


Chart Suggestions—A Thought-Starter



© 2006 A. Abela — a.v.abela@gmail.com

Bars vs. Lines

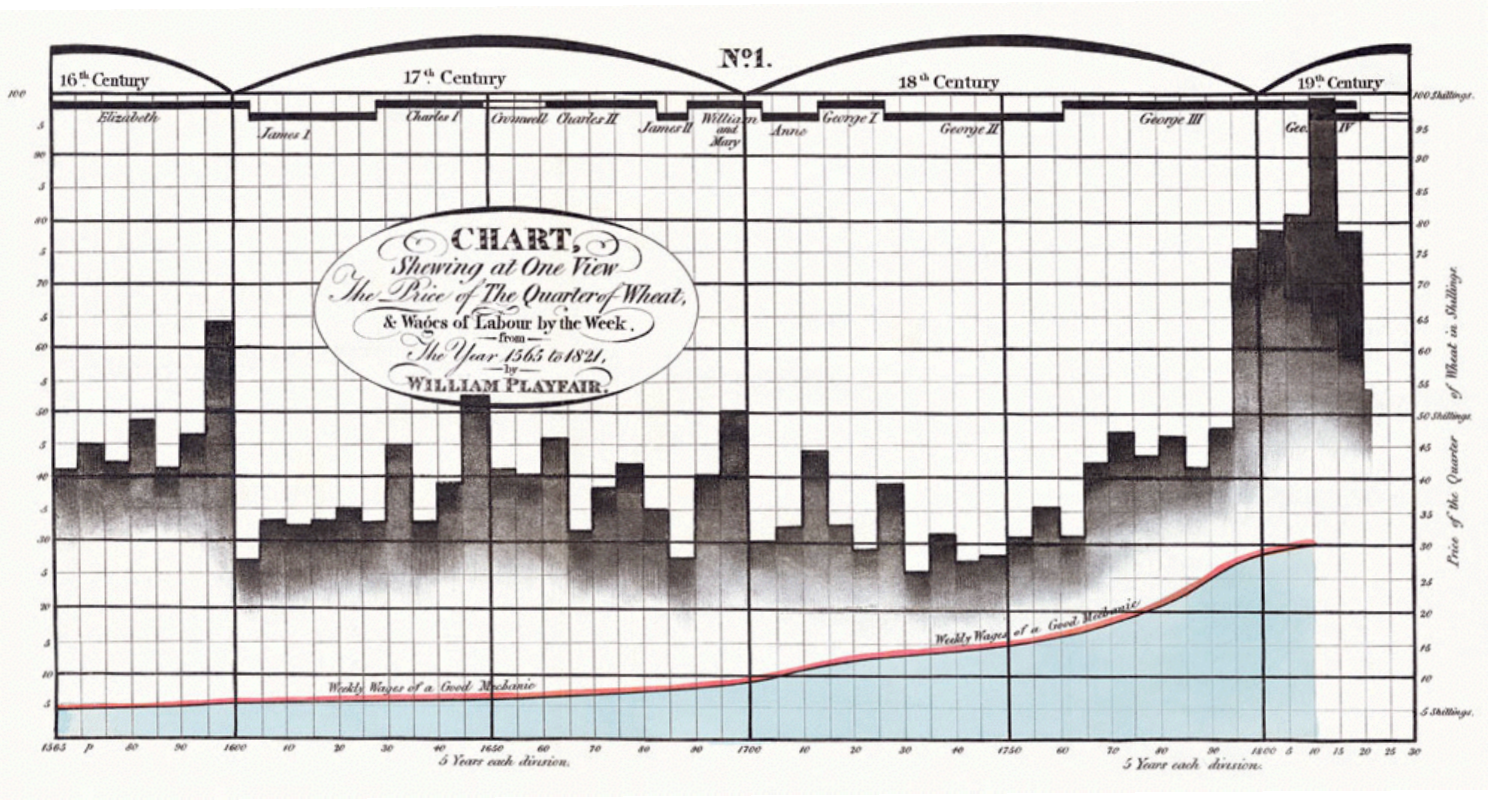


Line implies trends. Do not use for categorical data

Trend over time

WILLIAM PLAYFAIR

1759-1823



Trend over time

Trend over time

Published: February 2, 2010

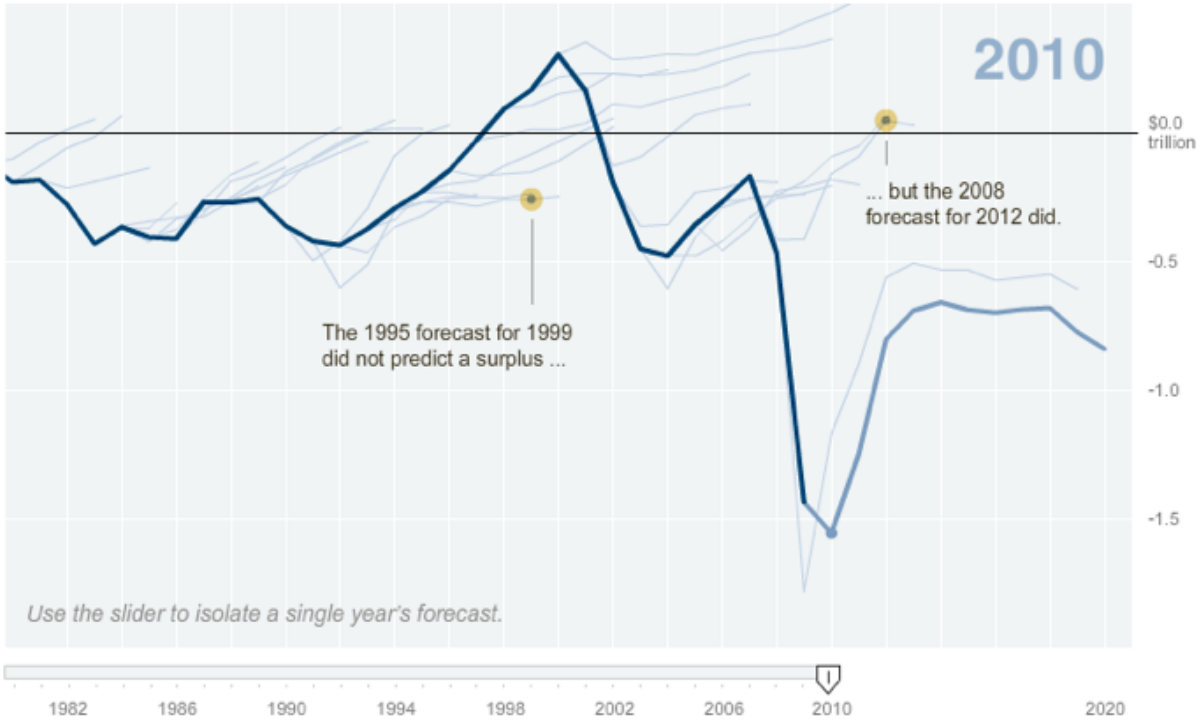
Budget Forecasts, Compared With Reality

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 ▸

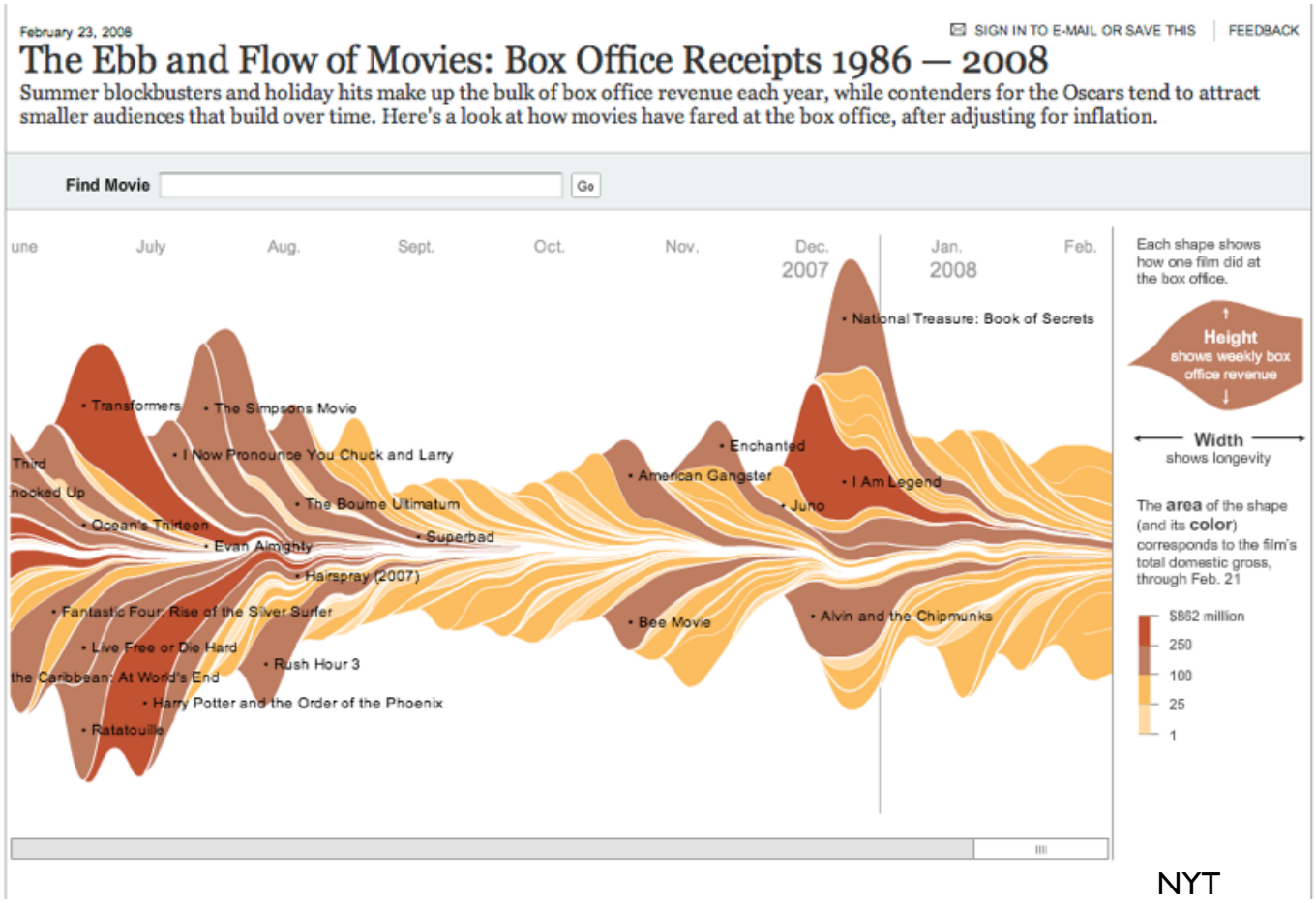
Latest forecast

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



NYT

Vision Statement

Scott Berinato is an editor at HBR. **Jeff Clark** is a developer and information visualizer based in Toronto.

Six Ways to Find Value in Twitter's Noise

Text by Scott Berinato, data and visualization by Jeff Clark

It's easy to dismiss Twitter as jabber, but smart marketers will recognize it as a stream of free consumer data to be mined in near-real time. Online visualization tools can help pinpoint what consumers are reading and sharing, elucidate memes in the chatter, and unearth trends. To show marketers how they can gain insight from Twitter, we captured more than a half million tweets containing the word "iPad" that were broadcast during the product's launch weekend in April. We then mapped key words that appeared in those tweets on the graph below.

HBR Graphic #10002

The iPad Launch by the Numbers

547,898

TOTAL TWEETS USING "IPAD" DURING PRODUCT'S LAUNCH WEEKEND

15% TWEETS FLAGGED AS SPAM (CONTAINED WORD "GIVEAWAY")

88% PERCENTAGE OF POSITIVE TWEETS

8% PERCENTAGE OF NEGATIVE TWEETS

1

Learn about the competitive landscape. Tweets about your product that include the names of rival brands can reveal a lot about market positioning. Most of the **KINDLE** tweets didn't focus on the iPad's being a Kindle killer (OK, for that matter, a laptop killer). Instead, they hailed the arrival of the Kindle app for the iPad—sending Apple stock to own books app will face serious competition.

2

Look for unexpected themes. Keyboard words point to persistent ideas. The word **CASE** became far more common right after people started buying iPads. But sources? Offers and ask for best cases, and customers trading tips about protecting their expensive new gadgets.

3

Dig deeper into the stream. While stream graphs give an overall impression of what people are tweeting about, it's important to know what other words are being used in relation to those in the stream. A few ribbons of tweets containing the word **EVIL** came not from Apple haters but from gamers tweeting about the arrival of Angry Birds HD 4 for the iPad. In our stream, it might look as though a lot of tweeters were calling the iPad **AMAZING**, but in fact, most of this traffic came from users retweeting a joke from comedian Conan O'Brien.

4

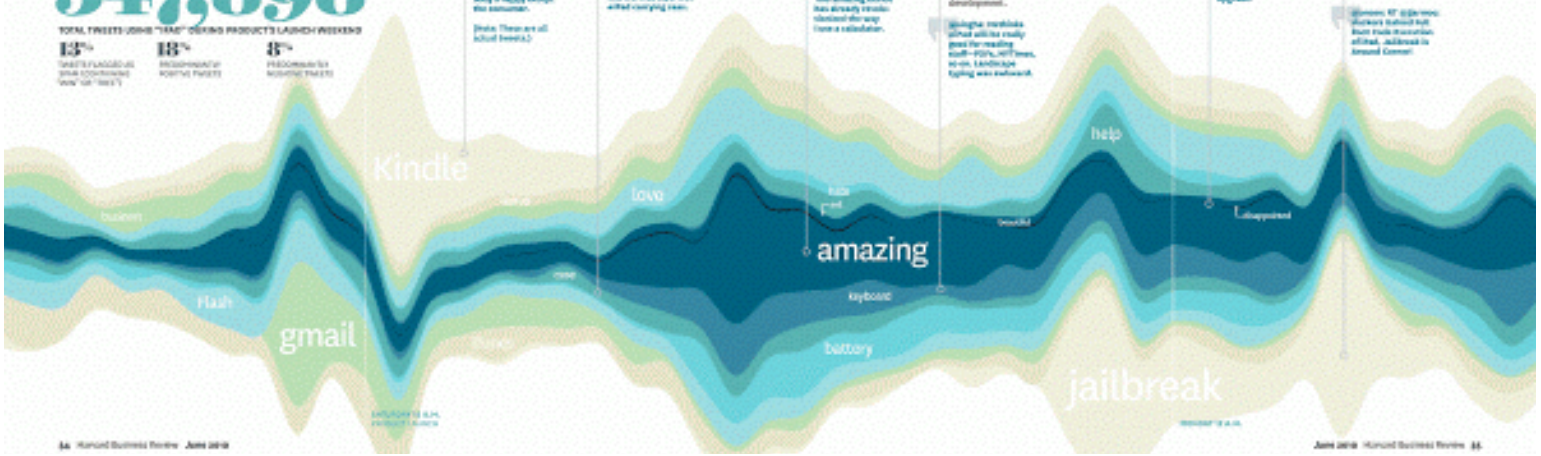
Look for user experiences. Product ratings and reviews can't replace user reactions. Looking at iPad tweets with the words **TYPING** and **KEYBOARD**, and then singling out the most common words in those tweets (see above), gives you feedback from thousands of users. The negative words could help prioritize marketing messages and product development.

5

Learn why negative words are coming up. Finding negative words is a good way to locate consumers' pain points. Though not frequently, **DISAPPOINTED** and **FRUSTRATION** appeared in tweets that focused on product functionality. Customer service could be adjusted to address the most common complaints.

6

Learn about conversation dominators. Words that suddenly dominate the tweet stream mean something has happened that's worth tracking down. The term **JAILBREAK** refers to hackers' use of iPad source code to run software not authorized by Apple. Twitter's most strong links to stories, code, and video tutorials on how to hack the device.

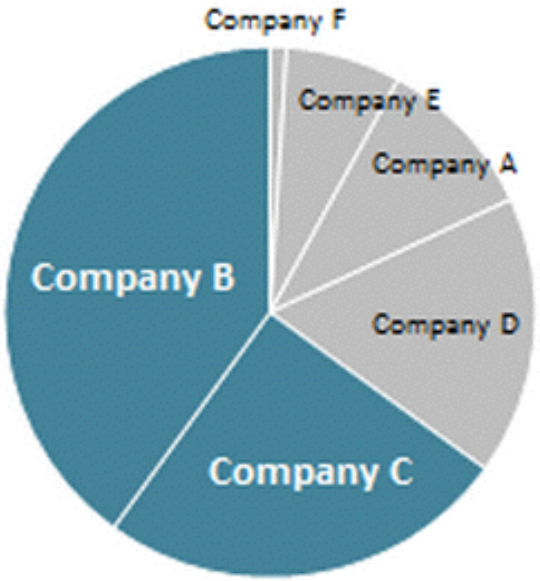


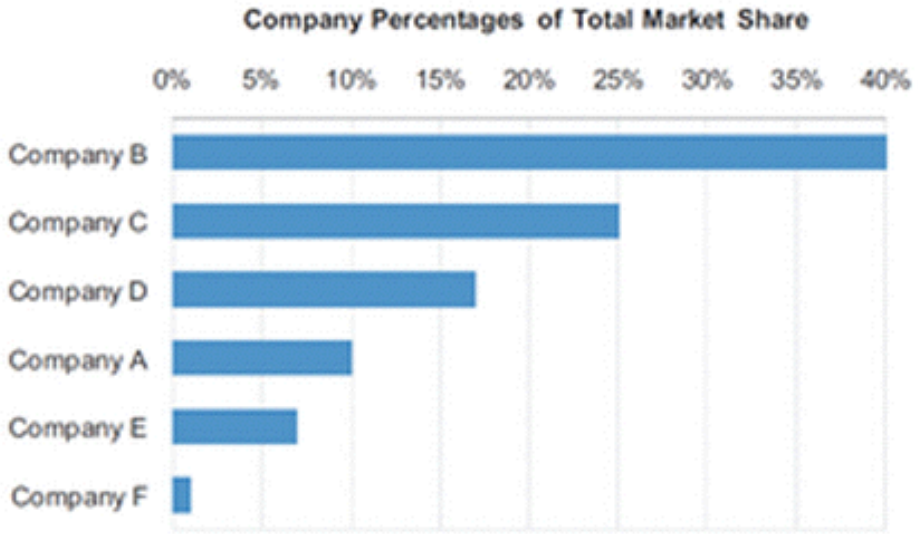
How to Read This Graph

This stream graph shows tweet volume over time. Each ribbon illustrates the proportion of iPad tweets containing a given word, such as **HELP**. Color is used only to distinguish ribbons.

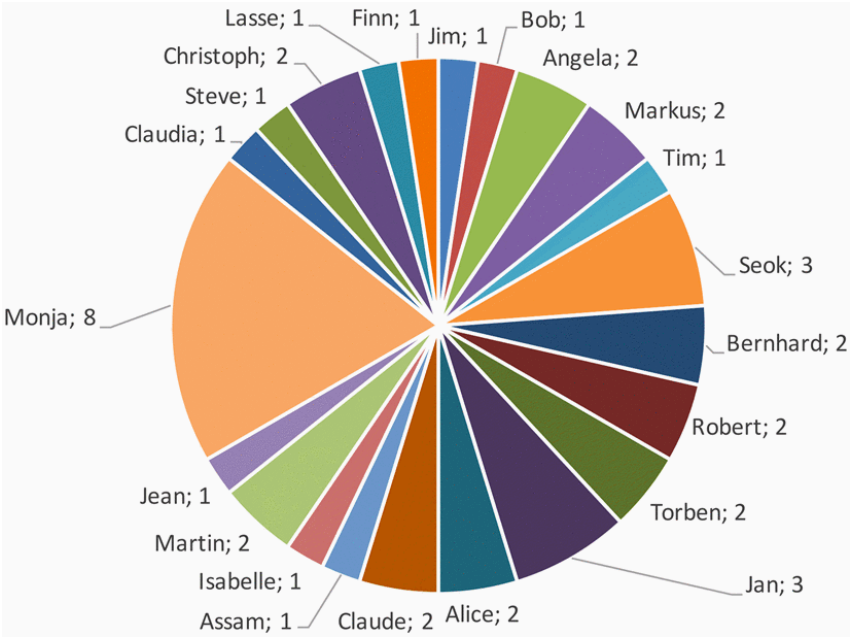
Pie vs Bar charts

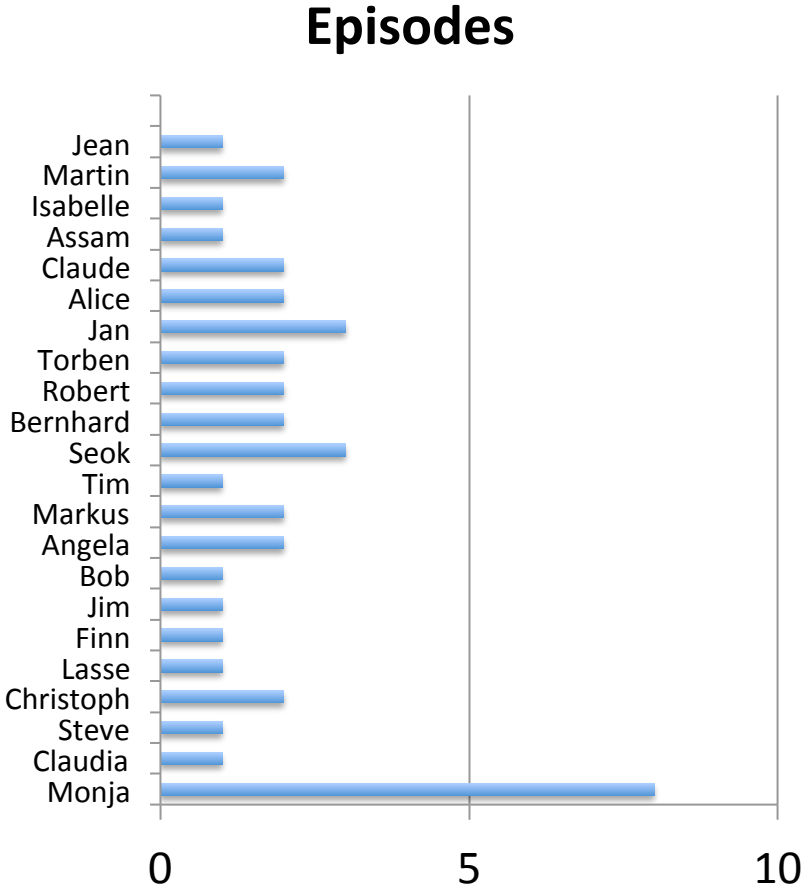
65% of the market is controlled by companies B and C





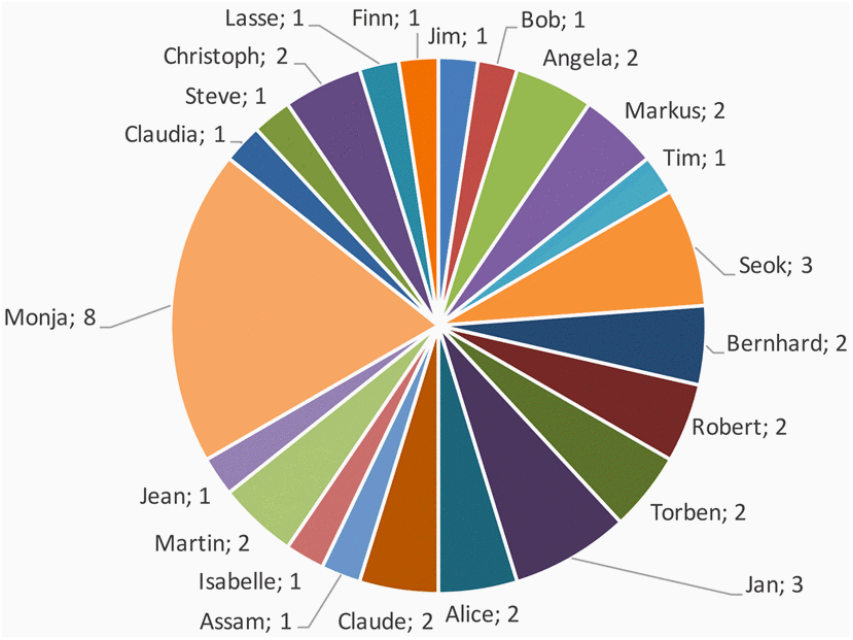
Pies vs Bar charts

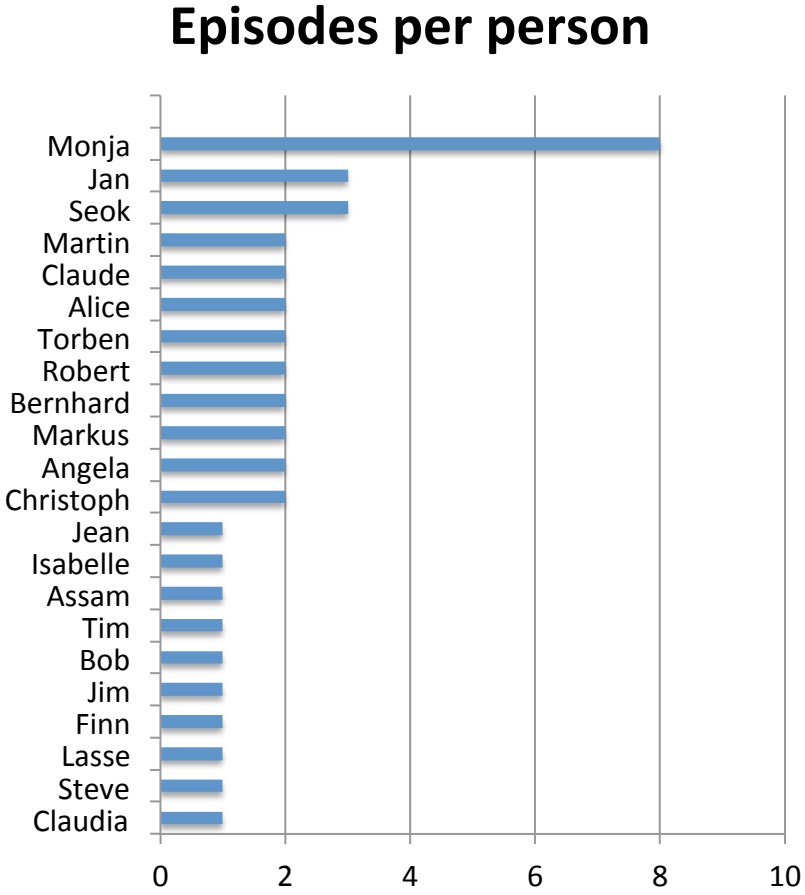




Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the 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.

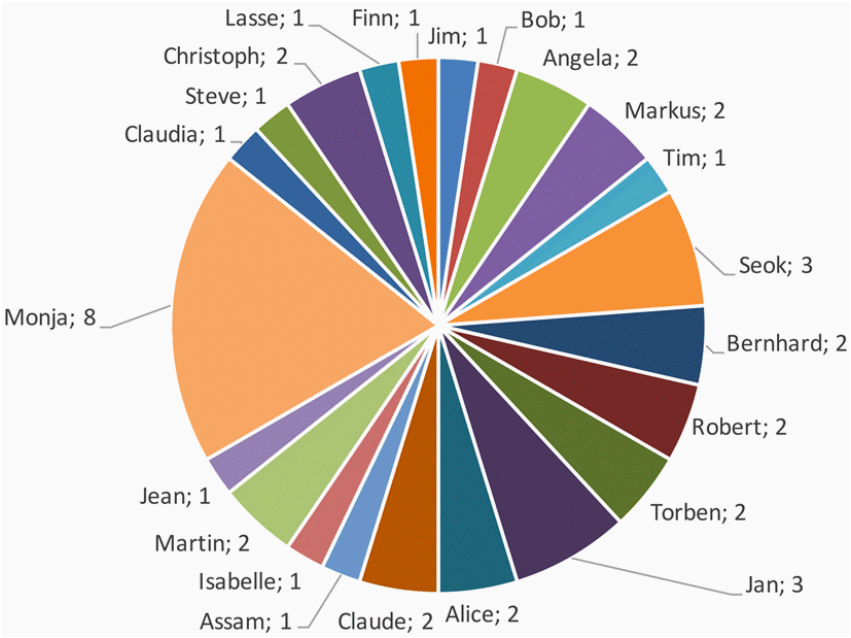
Pies vs Bar charts (improved)

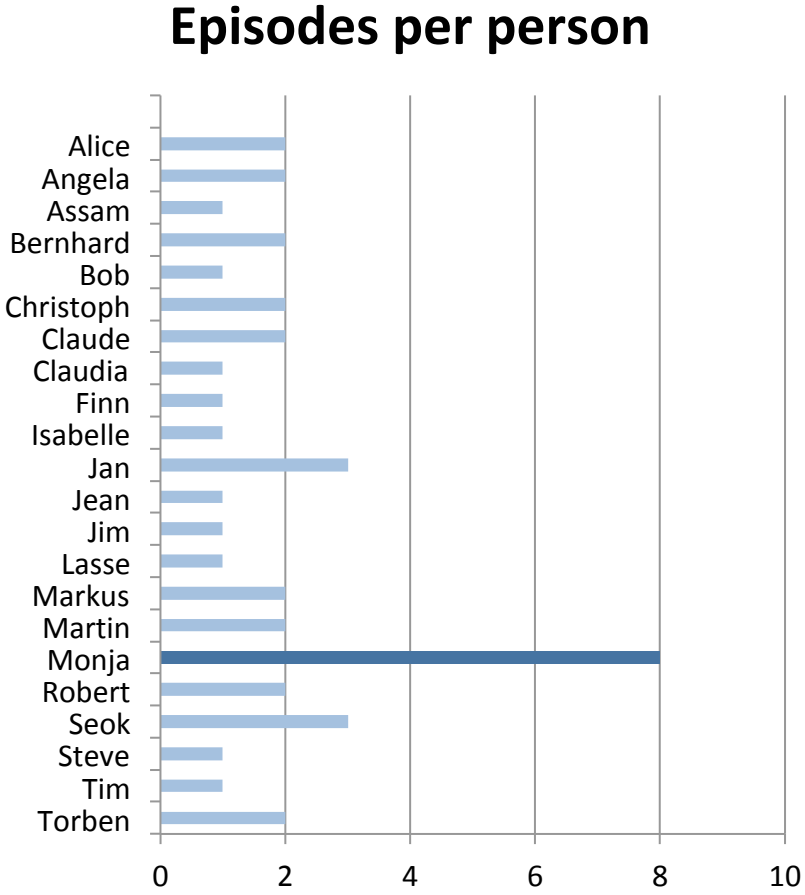




Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the 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.

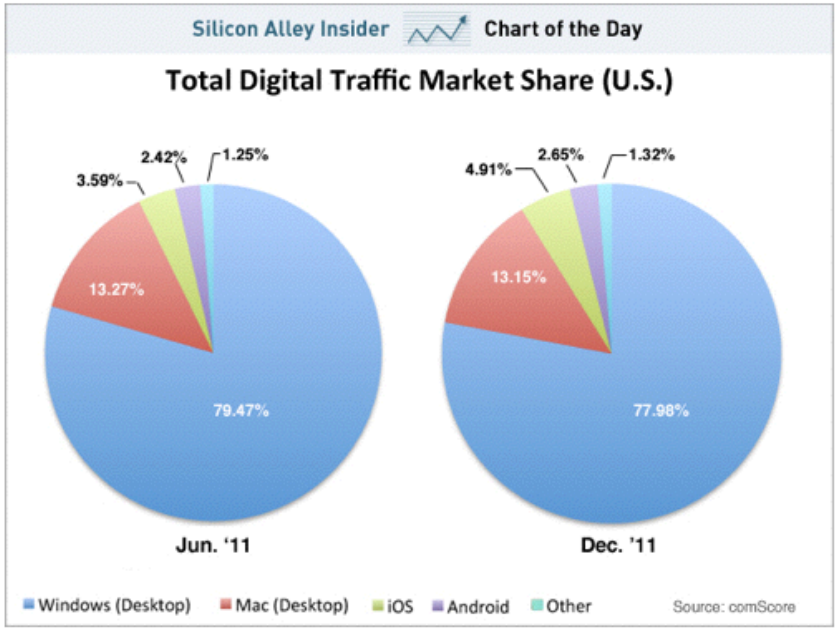
Pies vs Bar charts (improved)

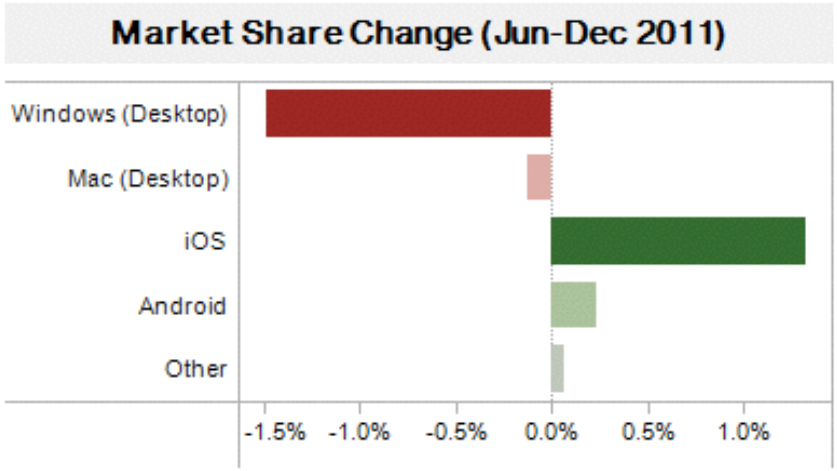




Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the 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.

Showing changes

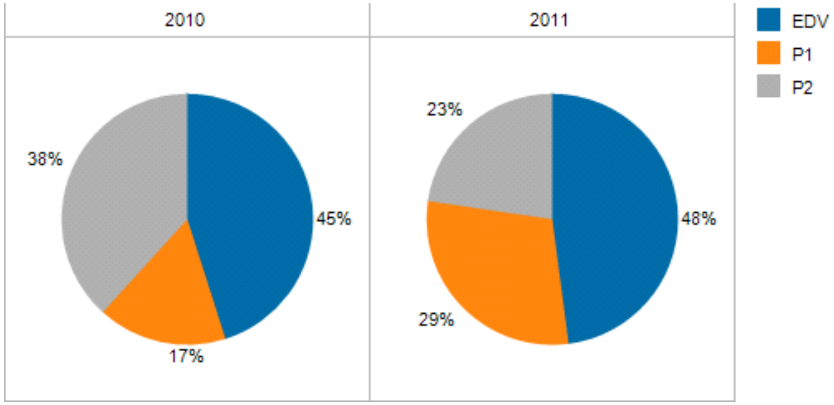


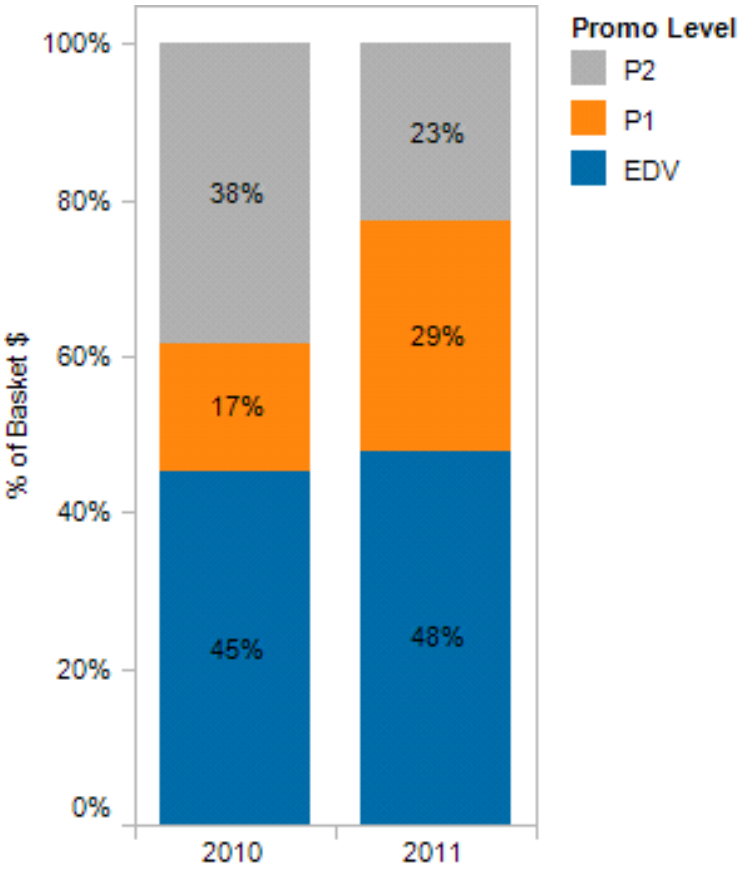


Market Share

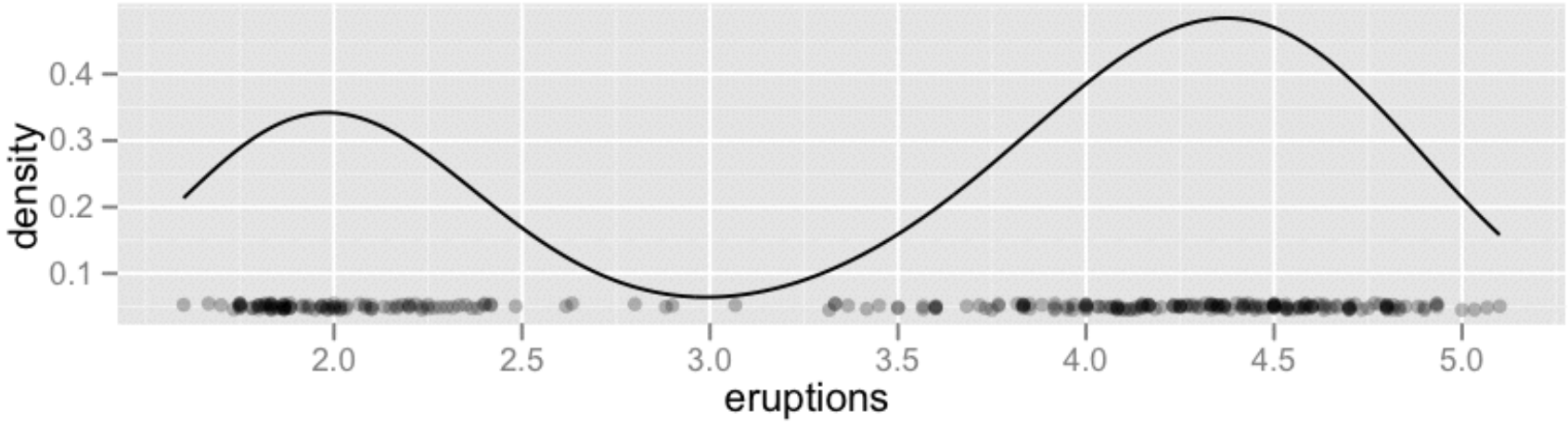
Platform	Jun 2011	Dec 2011	Change
Windows (Desktop)	79.5%	78.0%	(1.5%)
Mac (Desktop)	13.3%	13.2%	(0.1%)
iOS	3.6%	4.9%	1.3%
Android	2.4%	2.7%	0.2%
Other	1.3%	1.3%	0.1%

Showing Changes

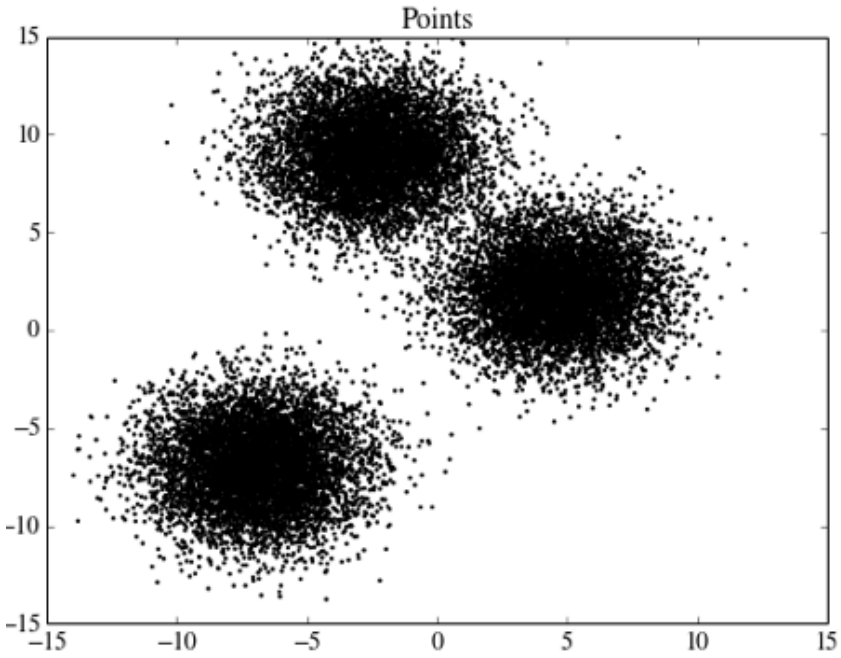


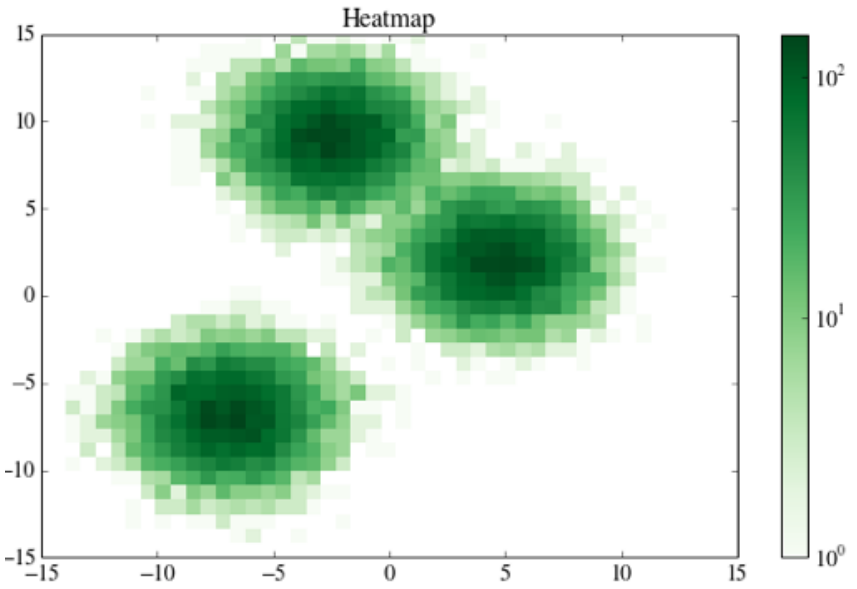


Density Plot

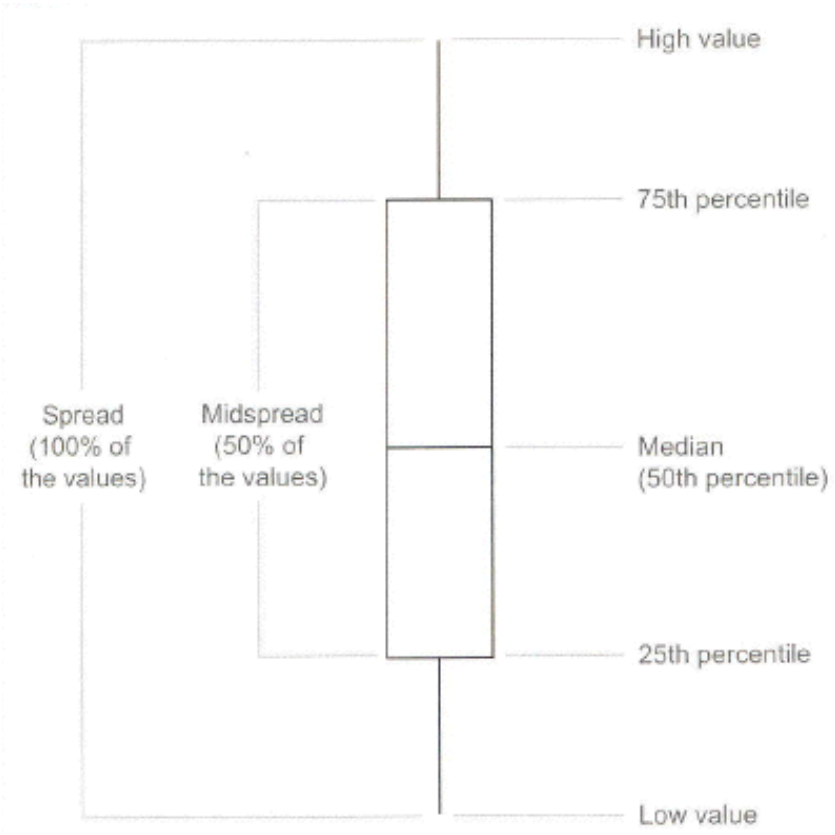


2D Density Plots

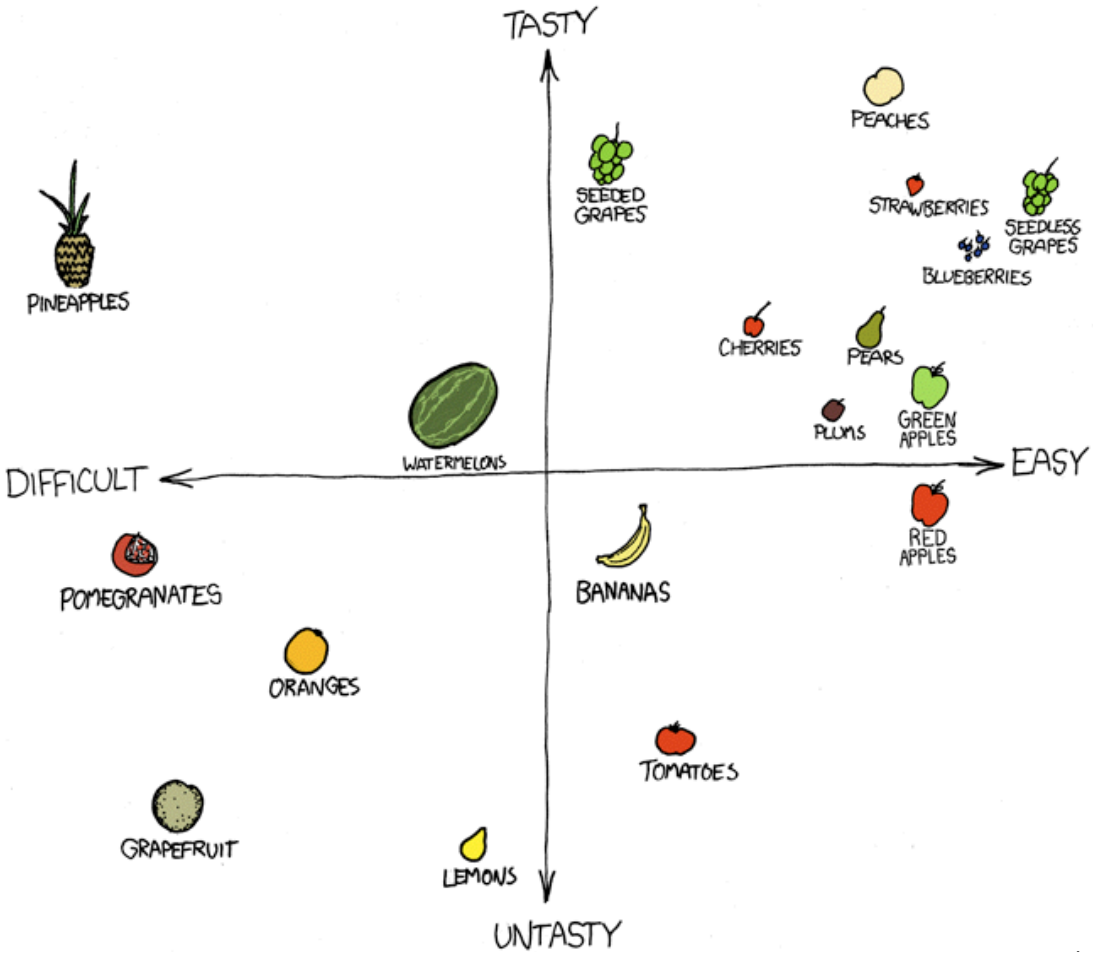




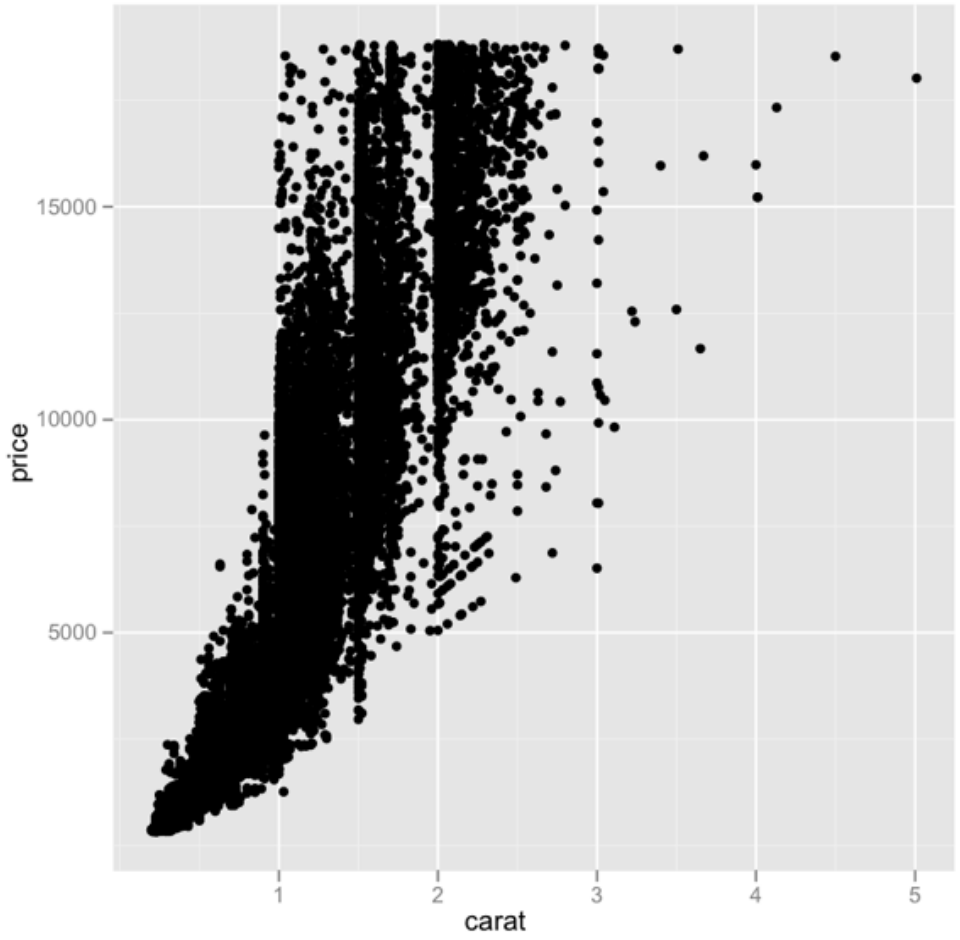
Box Plots


High value
75th percentile
Spread
(100% of
the values)
Midspread
(50% of
the values)
Median
(50th percentile)
25th percentile
Low value

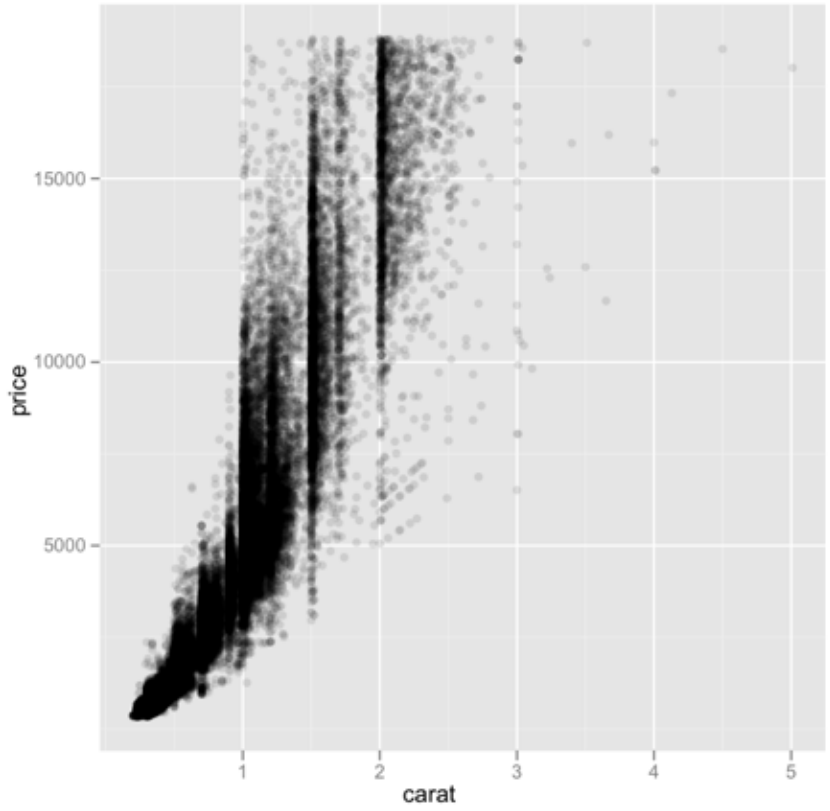
Scatterplot



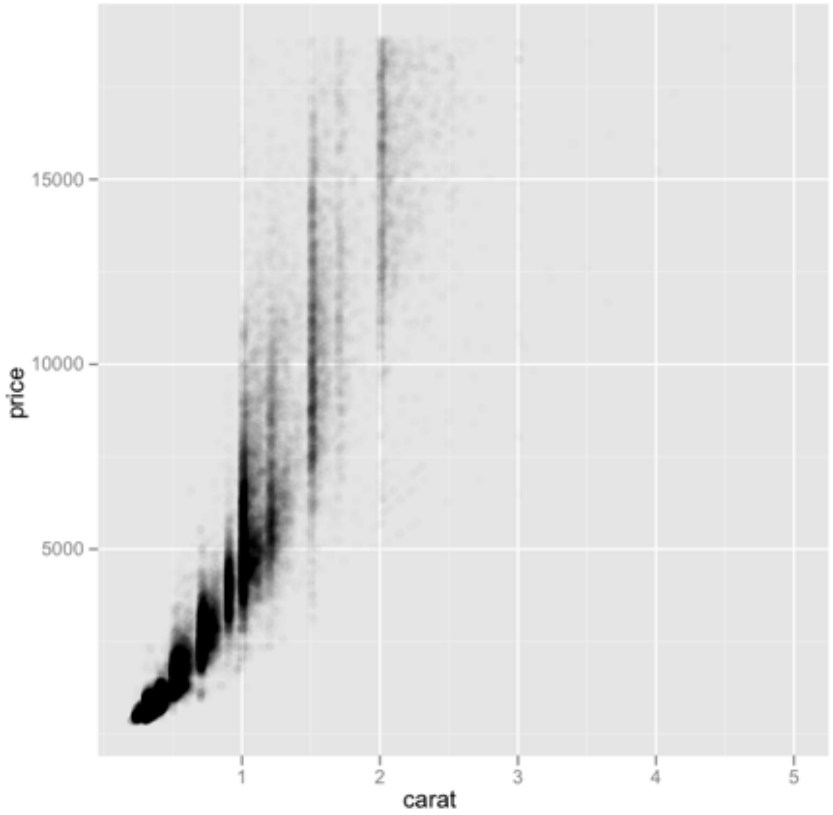
Cluttering, Overplotting



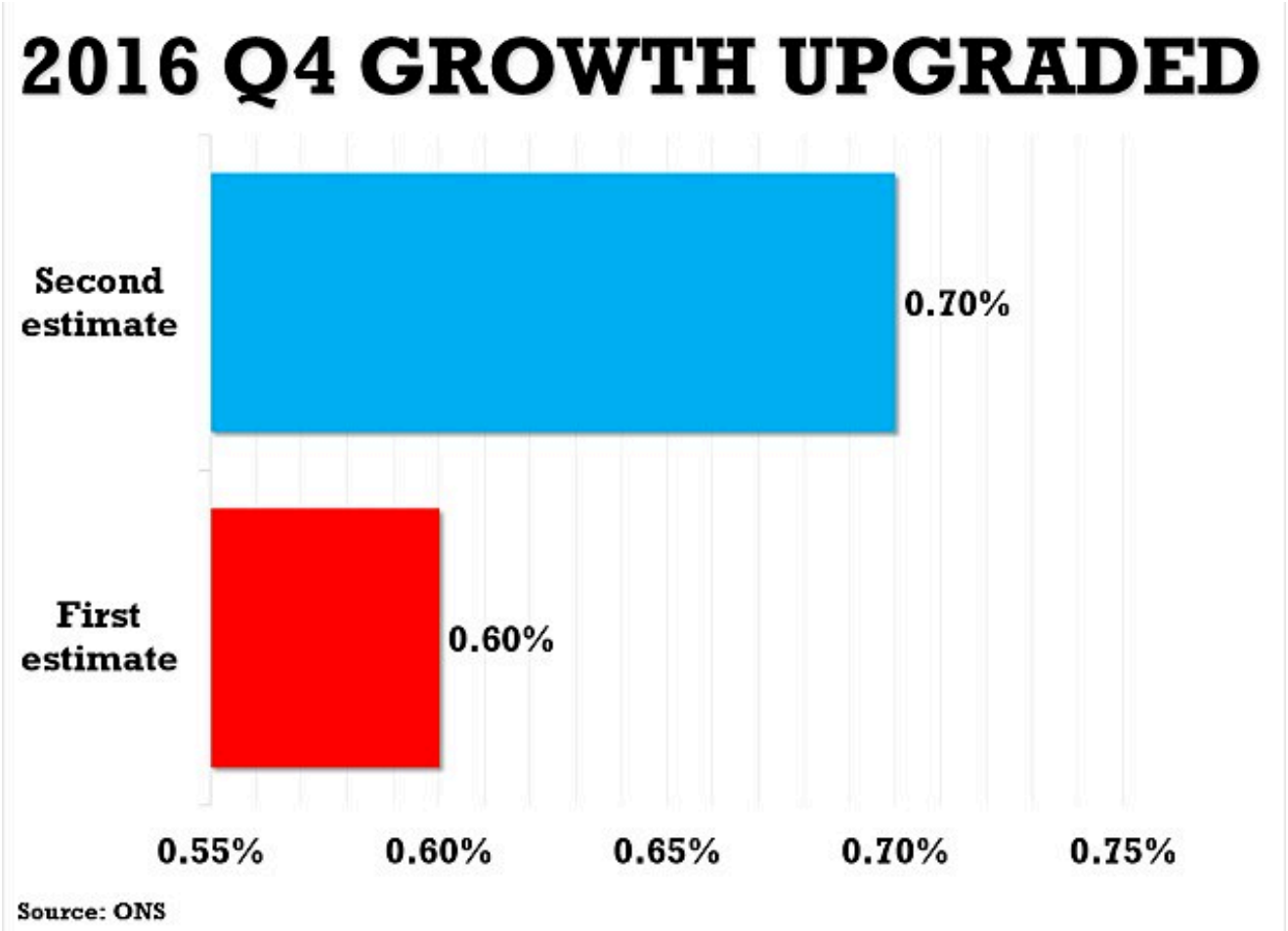
alpha=1/10



alpha=1/100

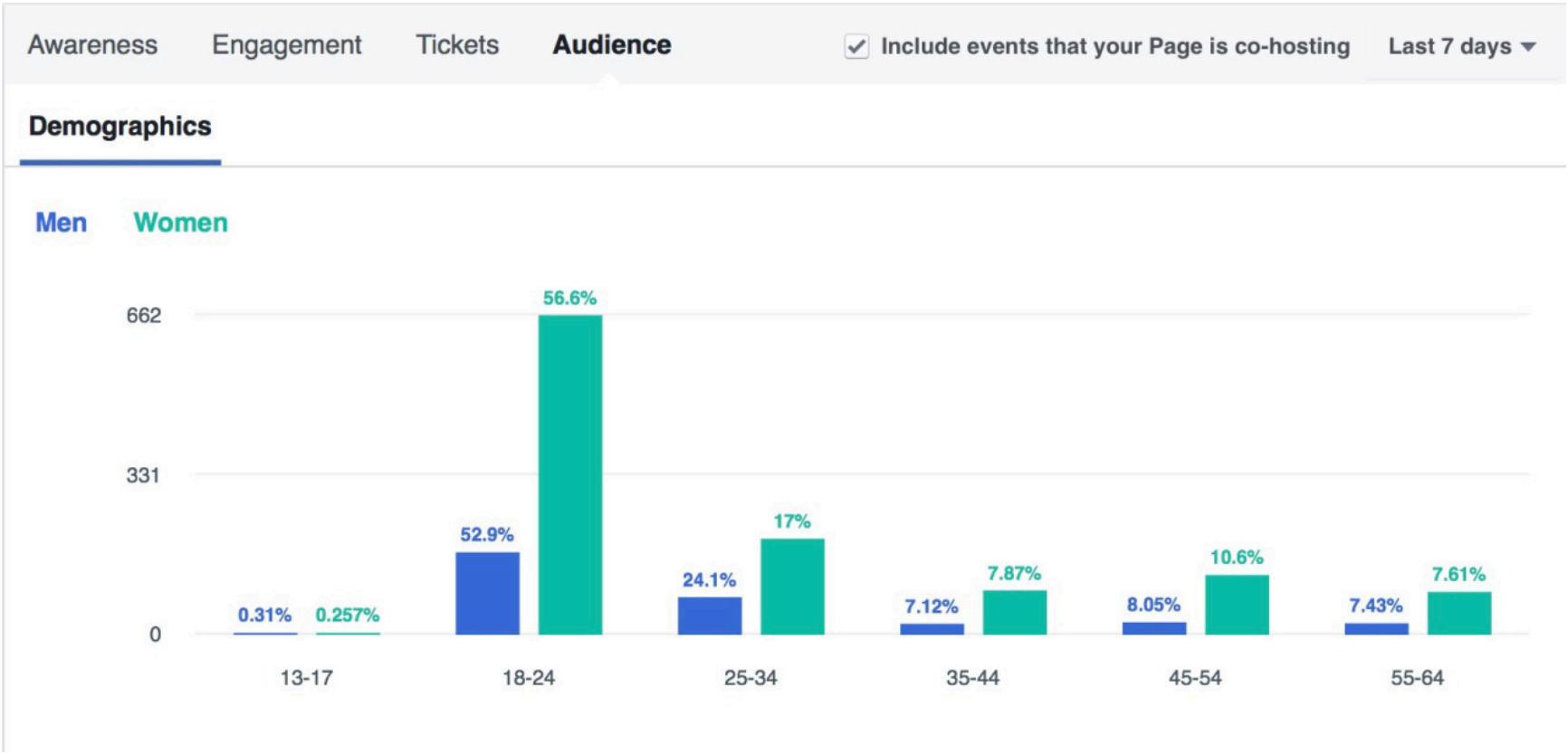


A FEW EXAMPLES AND CASE STUDIES

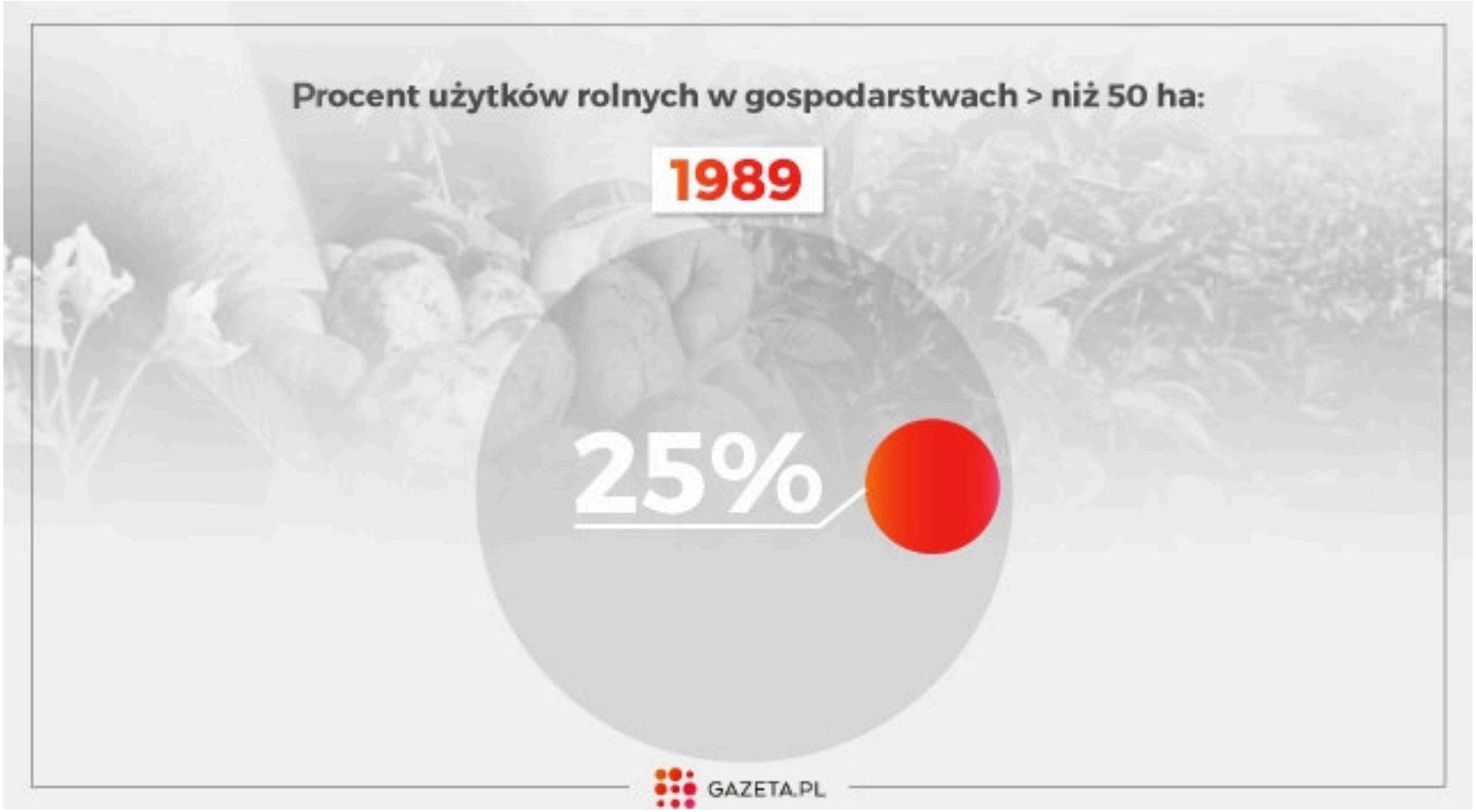


The Office for National Statistics (ONS) said gross domestic product (GDP) expanded by 0.7 per cent in the fourth quarter - an increase from the 0.6 per cent calculated on the watchdog's first look at the economy

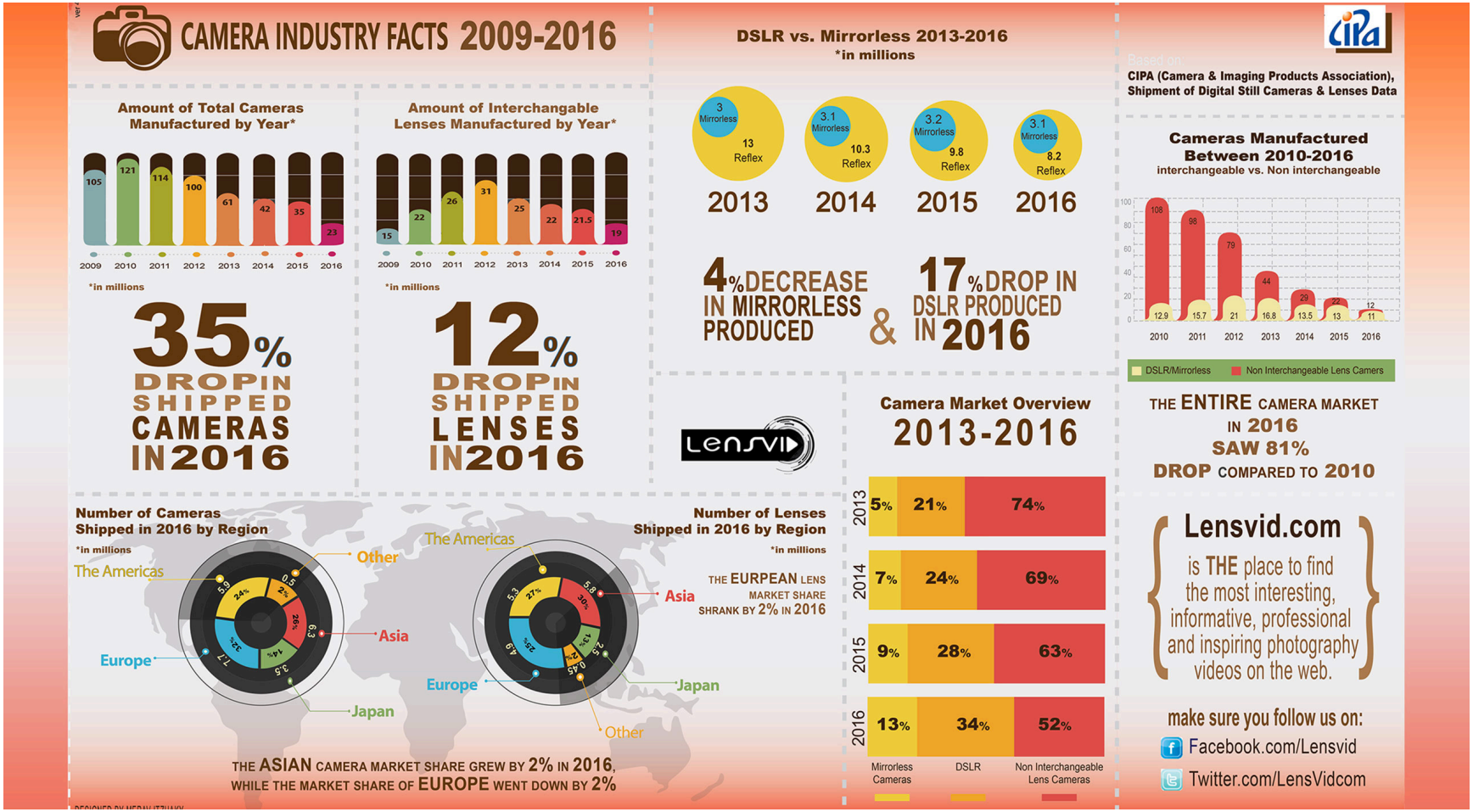
Source: http://www.dailymail.co.uk/news/article-4248690/Economy-grew-0-7-final-three-months-2016.html



Source: Facebook Analytics



Source: http://weekend.gazeta.pl/weekend/1,152121,20528386,wierza-mocniej-niz-miastowi-za-to-zarabiaja-mniej-i-calym.html



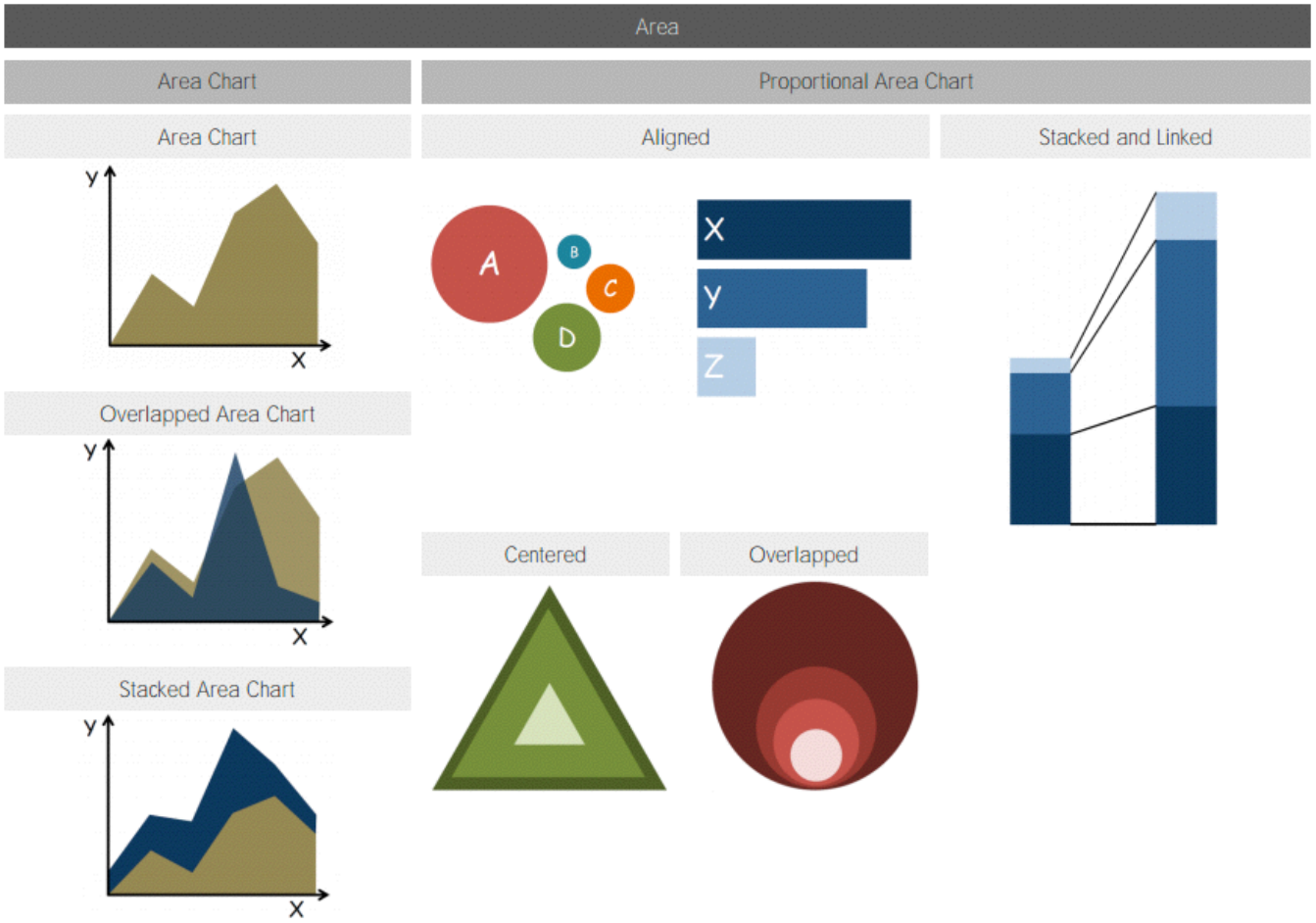
Source: https://lensvid.com/gear/lensvid-exclusive-happened-photography-industry-2016/

Borkin MA,VoAA,Bylinskii Z,Isola P,Sunkavalli S,OlivaA,Pfister H.
What Makes a Visualization Memorable?
IEEETransactions on Visualization and Computer Graphics (InfoVis 2013).

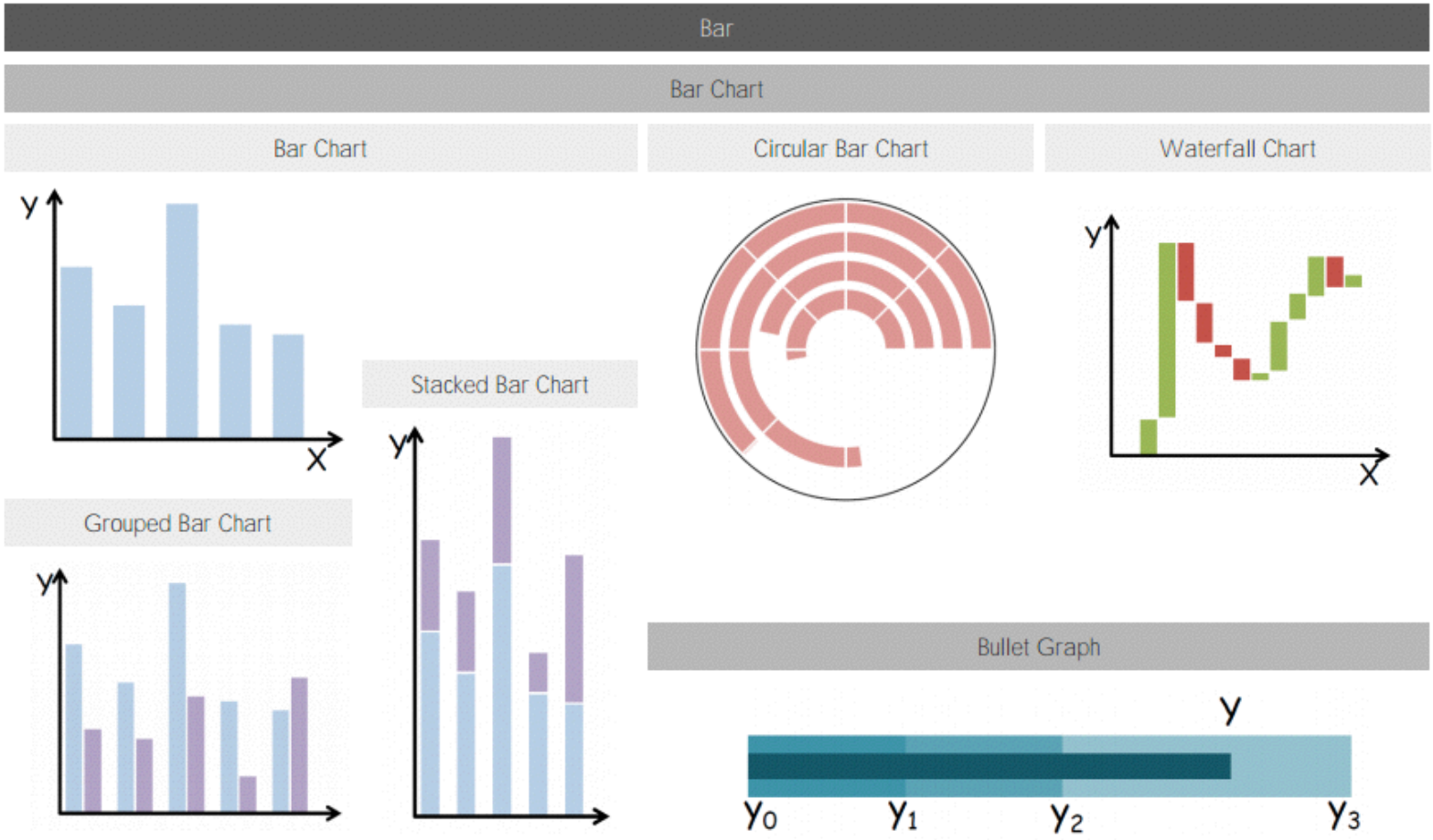
http://vcg.seas.harvard.edu/publications/ what-makes-visualization-memorable

VISUALIZATION TAXONOMY

Area


Area
Area Chart
Proportional Area Chart
Area Chart
Aligned
Stacked and Linked
y
X
A
B
C
D
X
y
Z
Overlapped Area Chart
y
X
Centered
Overlapped
Stacked Area Chart
y
X

Bar

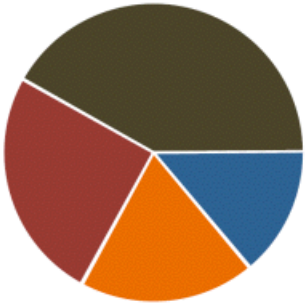

Bar
Bar Chart
Bar Chart
Circular Bar Chart
Waterfall Chart
y
x
Stacked Bar Chart
Grouped Bar Chart
Bullet Graph
y
y_0
y_1
y_2
y_3

Circle

Belt Chart

Donut Chart

Pie Chart



Sector Graph

Diagram

Flow Chart

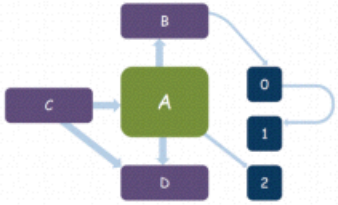
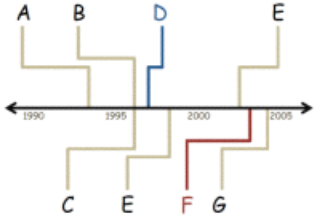


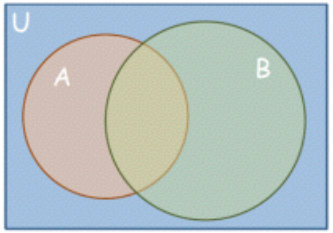
Illustration (or Rendering)

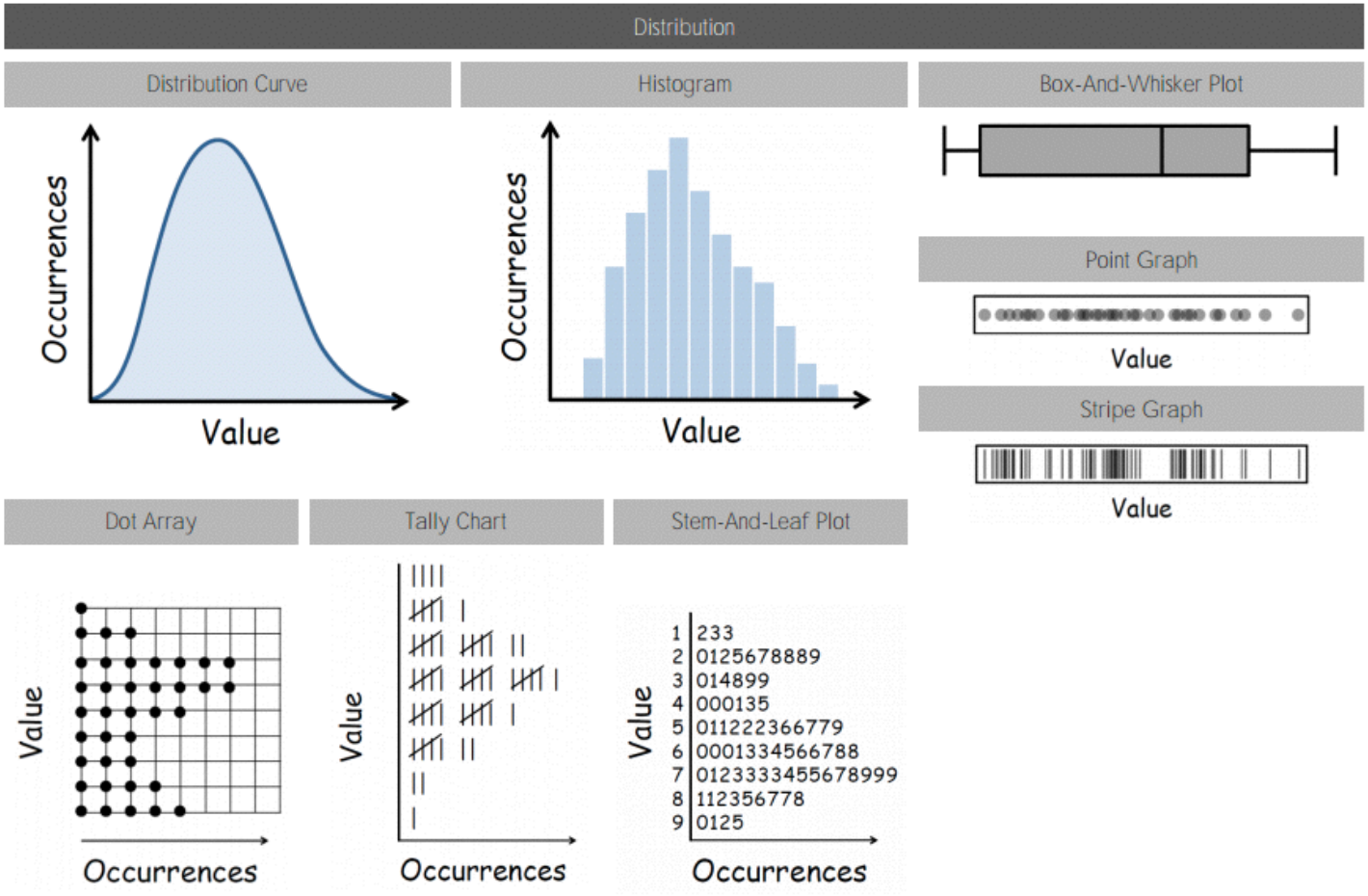
Sankey Diagram

Timeline



Venn Diagram

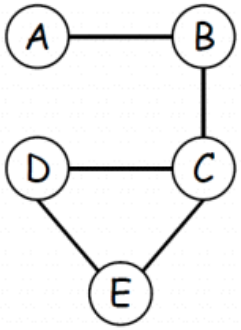



Distribution
Distribution Curve
Occurrences
Value
Histogram
Occurrences
Value
Box-And-Whisker Plot
Point Graph
Value
Stripe Graph
Value
Dot Array
Value
Occurrences
Tally Chart
Value
Occurrences
Stem-And-Leaf Plot
Value
1 233
2 0125678889
3 014899
4 000135
5 011222366779
6 0001334566788
7 0123333455678999
8 112356778
9 0125
Occurrences

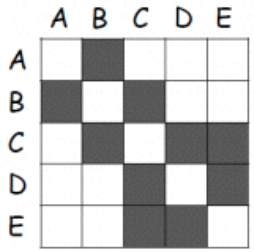
Tree and Network (Graph)

Tree and Network (Graph)

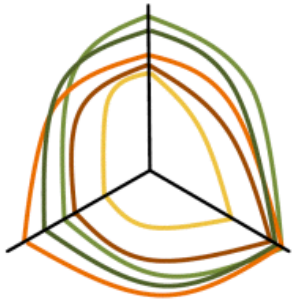
Graph



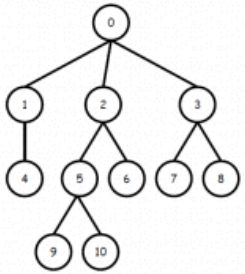
Matrix Representation



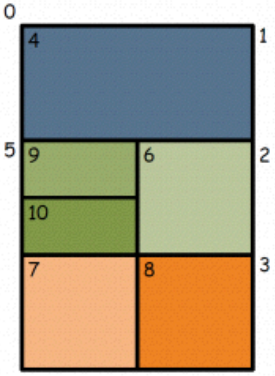
Hive Graph



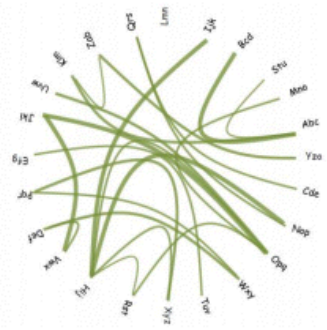
Tree



Treemap

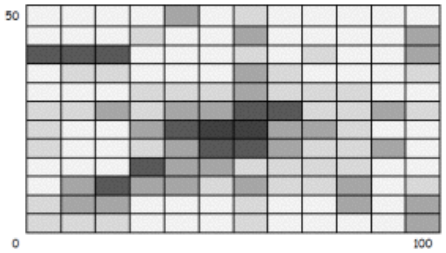


Hierarchical Edge Bundling



Grid / Matrix

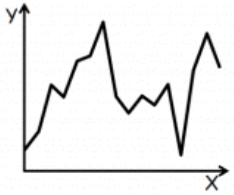
Heat Map



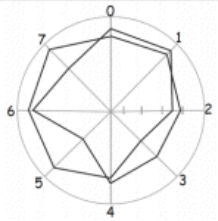
Line

Line Graph

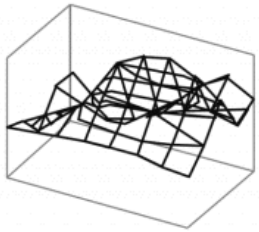
Line Graph



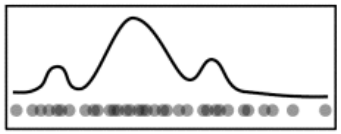
Circular Line Graph



Surface Graph

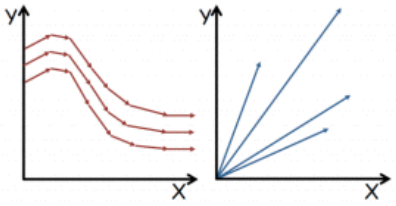


Density Plot

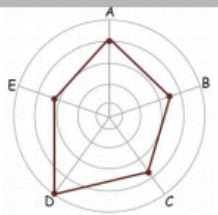


Surface Graph

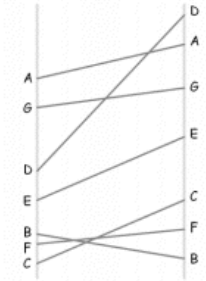
Vector Graph



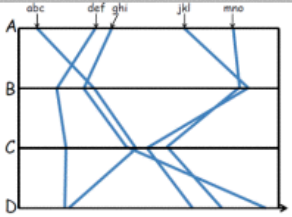
Star Plot



Slopegraph



Parallel Coordinates



Map

Geographic Map

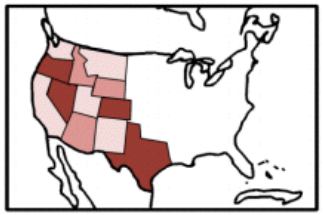
Geographic Map

Street Map

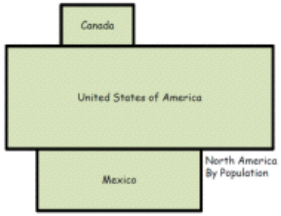
Flow Map

Statistical Map

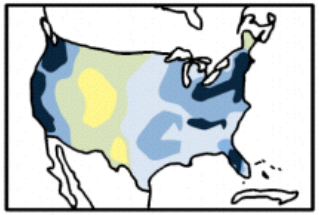
Choropleth Map



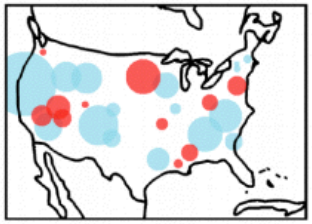
Distorted Map (Cartogram)



Contour Map (Isopleth)



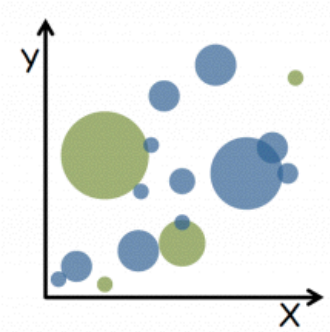
Statistical Plot Map



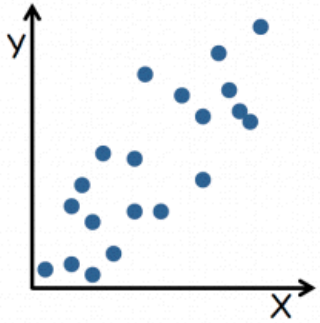
Point

Scatter Plot

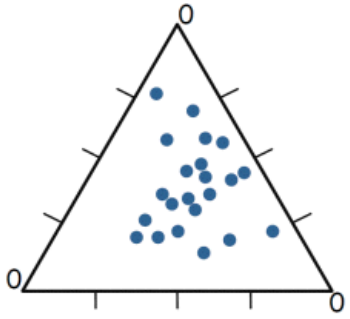
Bubble Chart



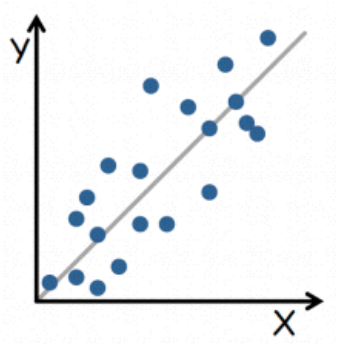
Scatter Plot



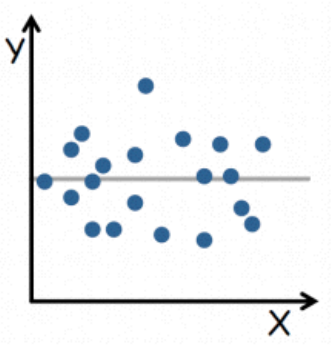
Trilinear Scatter Plot



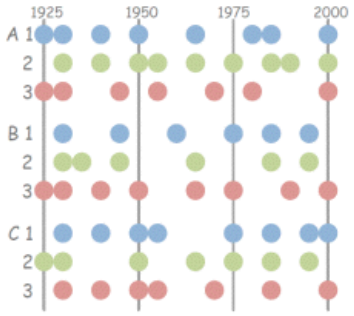
Trend Line



Residual Graph



Dot Plot



Table

Table

ABC	1234	X45
Category	543.2109	7%
Group	45.67	45%
Unit	9876	98%
Class	123.78	12%

Text Chart

Title

- Sed dignissim vehicula
- Nisl quis congue
- Sed vitae rhoncus odio
- Integer at odio

Heading 1

"Nunc aliquam turpis at tellus varius hendrerit. Ut nec magna tortor. Proin adipiscing dolor eget odio semper ut commodo lacus imperdiet."

- Lorem

Heading 2

Aenean tincidunt sem vel massa cursus non tempus quam auctor. In nisi mi, commodo sit.

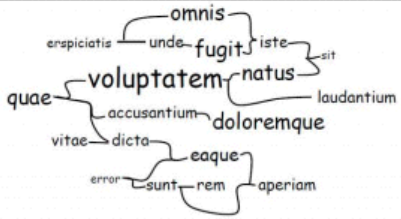
Amet rutrum vitae, fringilla non urna. Quisque sagittis ultrices sapien, quis posuere massa interdum quis.

Heading 3

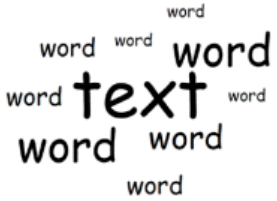
- ✓Chart 1
- ✓Chart 2
- ✓Chart 3
- ✓Chart 4

Text Based

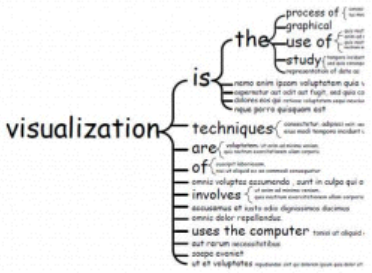
Phrase Net



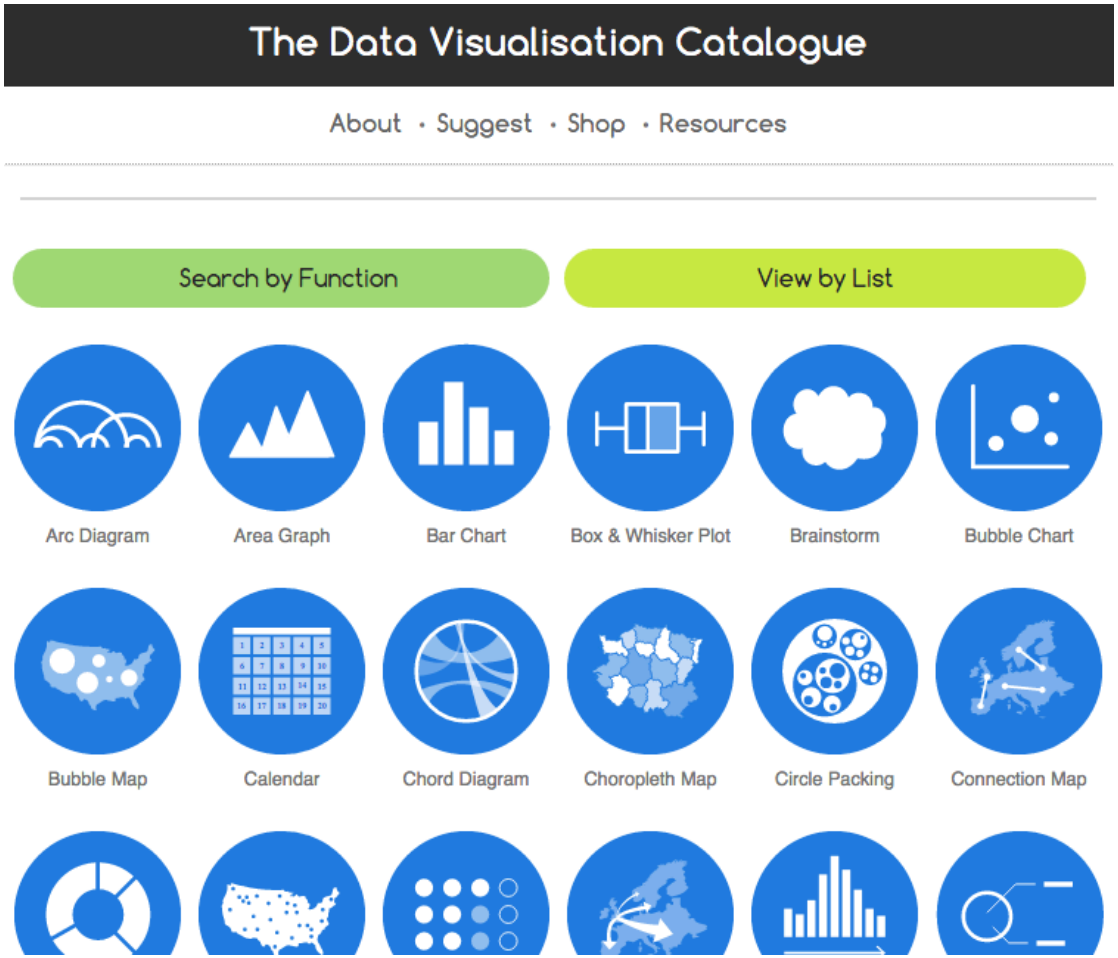
Word Cloud



Word Tree



Visual Taxonomy



http://www.datavizcatalogue.com/

Takeaway Messages

- Appropriate chart type for specific data type and visualization task