GridDynamics

Visualization for dummies

... from dummies

Iana Klementeva 2022

slack
@iklementeva
~ 2y. CustomDev/BigData

Main questions

Visualisation?

- Representation

- Data explorational/Find insights
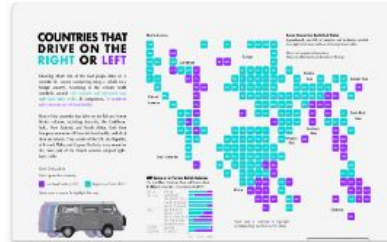
What we will talking about

- Visualisation technique

- Graphical Excellence
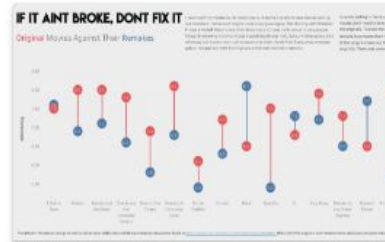
- Lies

- Charts

Viz of the day

What will you watch next? #B2VB

Agata Ketterick

☆ 43 👁 4,140

CountriesThat Drive on the Right or Left

Adedamola (Damola) Ladipo

☆ 70 👁 5,073

If It Ain't Broke, Don't Fix It - Movie Remakes
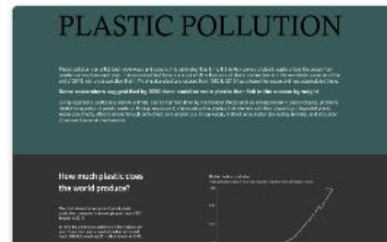
David Potter

☆ 70 👁 5,384

Winners of The Open

Jessica Moon

☆ 15 👁 2,798

Insurance Dashboard

Autumn Battani

☆ 171 👁 15,848

Plastic Pollution

Saverio Rocchetti

☆ 31 👁 3,455

Cost of Living in the World

Hypntic Data

☆ 85 👁 611,213

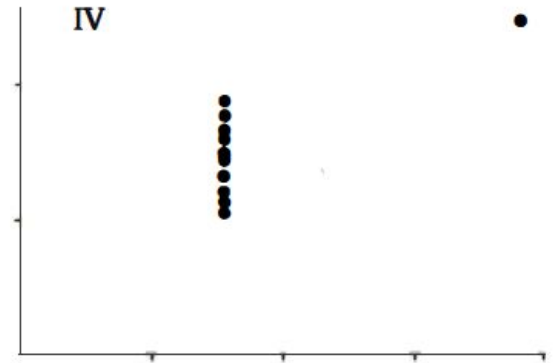**Sales Funnel Dashboard**
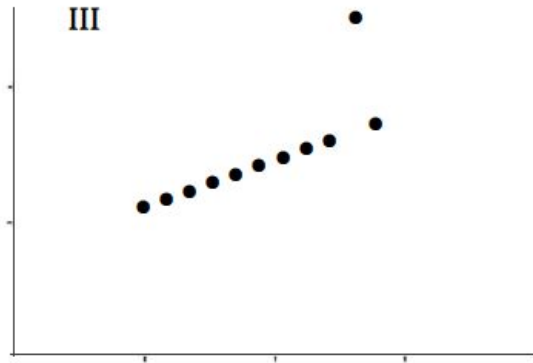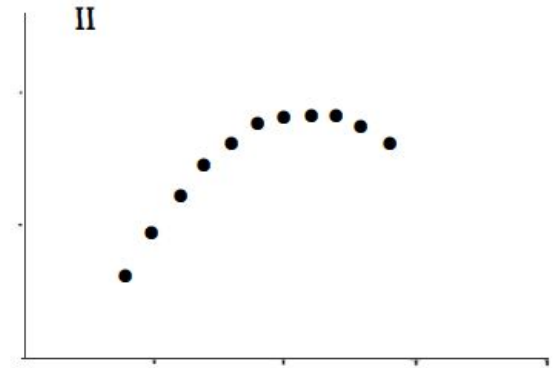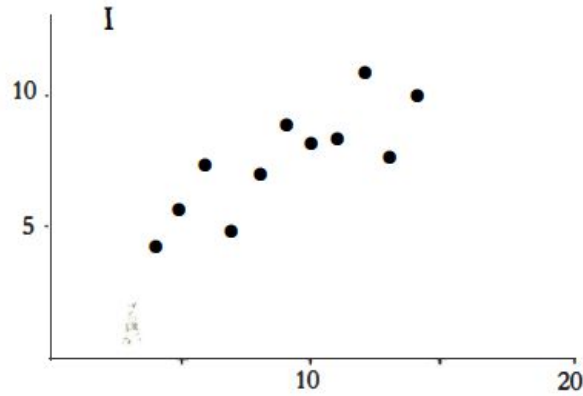
Seoyeon Jun

☆ 232 👁 13,409

https://public.tableau.com/app/discover/viz-of-the-day

Visualize it

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

$N = 11$

mean of X's $= 9.0$

mean of Y's $= 7.5$

equation of regression line: $Y = 3 + 0.5X$

standard error of estimate of slope $= 0.118$

$t = 4.24$

sum of squares $X - \overline{X} = 110.0$

regression sum of squares $= 27.50$

residual sum of squares of $Y = 13.75$

correlation coefficient $= .82$

$r^2 = .67$

Visualize it

	I		II		III		IV
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Visualize it

Graphical displays:

- Show the data (and only the data)

Graphical displays:

- Show the data (and only the data)

- Avoid distortion of what the data have to say

Graphical displays:

- Show the data (and only the data)

- Avoid distortion of what the data have to say

- Present many numbers in a small space

Graphical displays:

- Show the data (and only the data)

- Avoid distortion of what the data have to say

- Present many numbers in a small space

- Make large data sets coherent

Graphical displays:

- Show the data (and only the data)

- Avoid distortion of what the data have to say

- Present many numbers in a small space

- Make large data sets coherent

- Encourage the eye to compare different pieces of data

Graphical displays:

- Show the data (and only the data)

- Avoid distortion of what the data have to say

- Present many numbers in a small space

- Make large data sets coherent

- Encourage the eye to compare different pieces of data

- Reveal the data at several levels of detail from broad overview to the fine structure

Graphical displays:

- Show the data (and only the data)

- Avoid distortion of what the data have to say

- Present many numbers in a small space

- Make large data sets coherent

- Encourage the eye to compare different pieces of data

- Reveal the data at several levels of detail from broad overview to the fine structure

- Serve a clear purpose: description, exploration, tabulation, decoration

Graphical displays:

- Show the data (and only the data)

- Avoid distortion of what the data have to say

- Present many numbers in a small space

- Make large data sets coherent

- Encourage the eye to compare different pieces of data

- Reveal the data at several levels of detail from broad overview to the fine structure

- Serve a clear purpose: description, exploration, tabulation, decoration

- Be closely integrated with the statistical and narrative description of a data set

There are three kinds of lies: Lies, Damned Lies, and Statistics
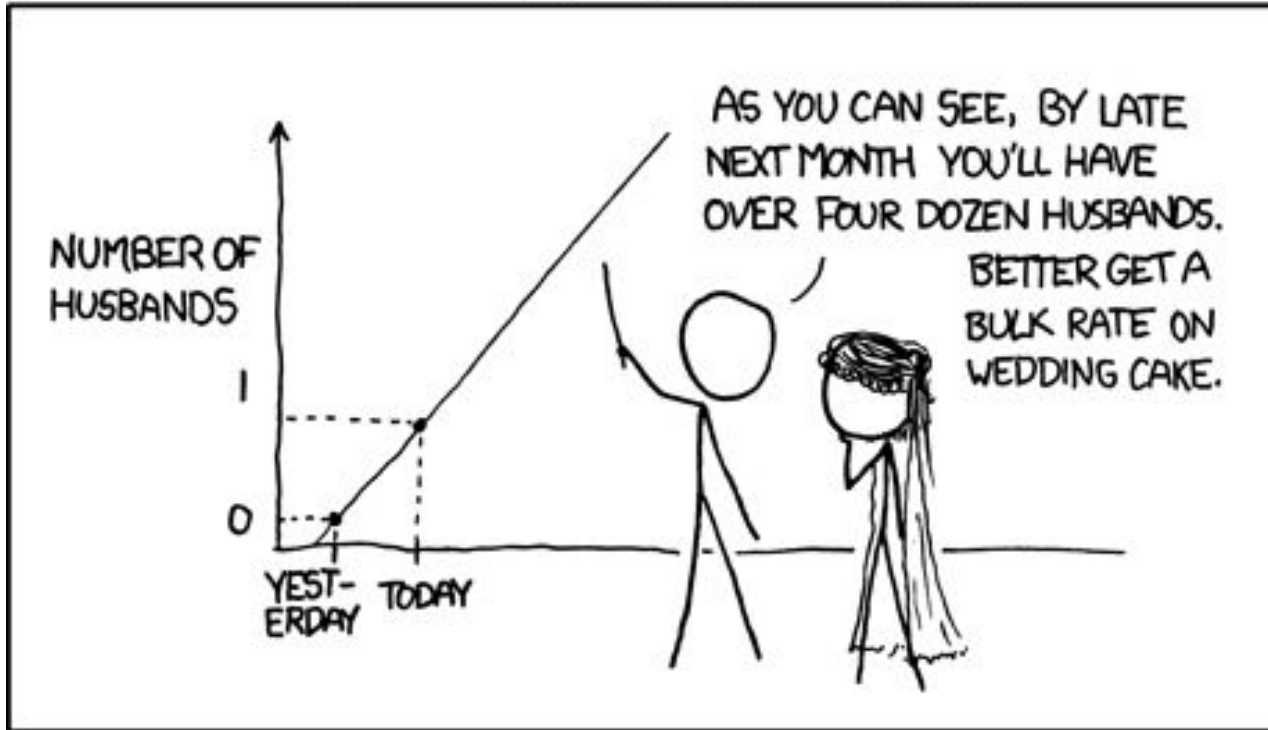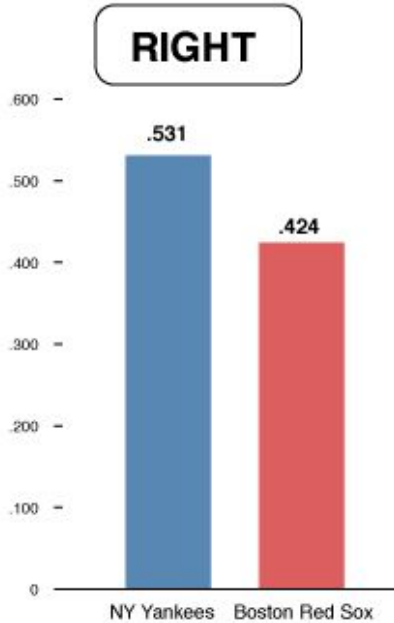
There are three kinds of lies: Lies, Damned Lies, and Statistics
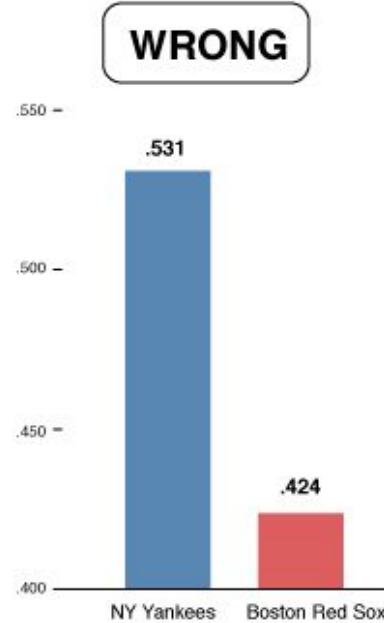
www.relativelyinteresting.com/how-to-mislead-people-with-graphics-that-lie

There are three kinds of lies: Lies, Damned Lies, and Statistics

There are three kinds of lies: Lies, Damned Lies, and Statistics
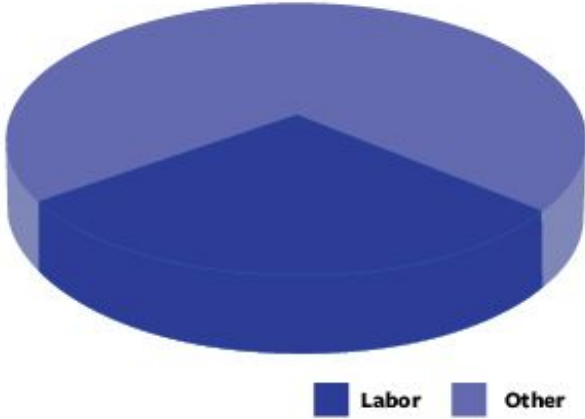
So many ways to ~~lie~~ visualize
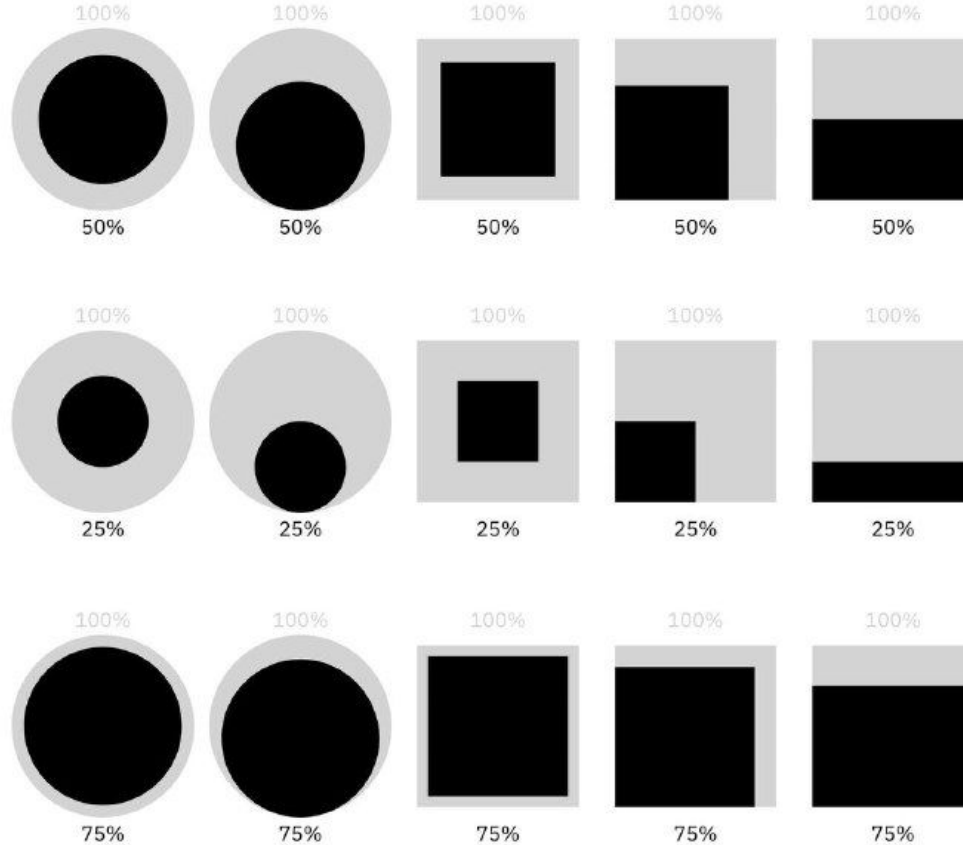
Spurious Correlations
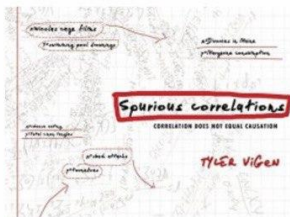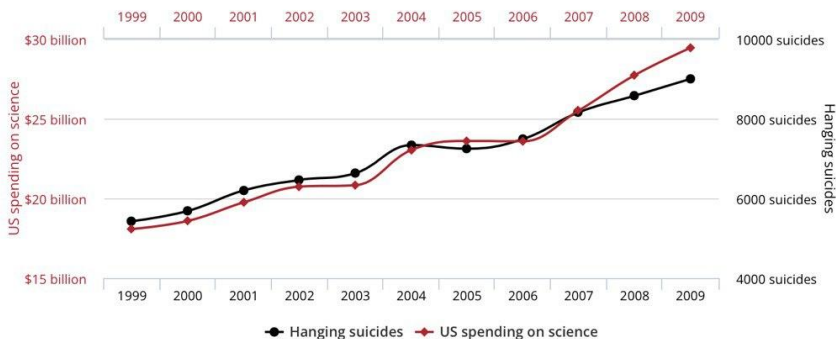
http://www.tylervigen.com/spurious-correlations

Lie Factor (E. Tufte)

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

$$size\ of\ effect = \frac{|second\ value\ -\ first\ value|}{first\ value}$$

If equal to 1, then the graphic might be doing a reasonable job of accurately representing the underlying numbers

Lie Factor

https://scienceplusstory.com/the-lie-factor-vs-the-baseline-principle

Lie Factor

$$\frac{(2011.5 - 2009)}{\frac{(108-94)}{94}}$$

Lie Factor

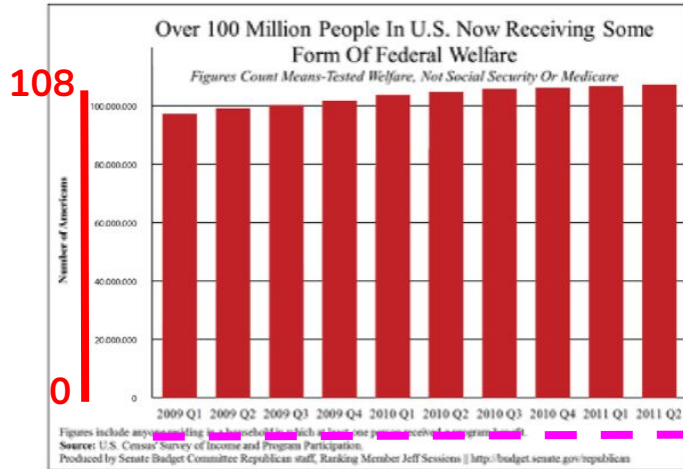lie factor: 1

lie factor: 16,08

Lie Factor

THE BLOG

Over 100 Million Now Receiving Federal Welfare
2:40 PM, AUG 8, 2012 · BY DANIEL HALPER

A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."
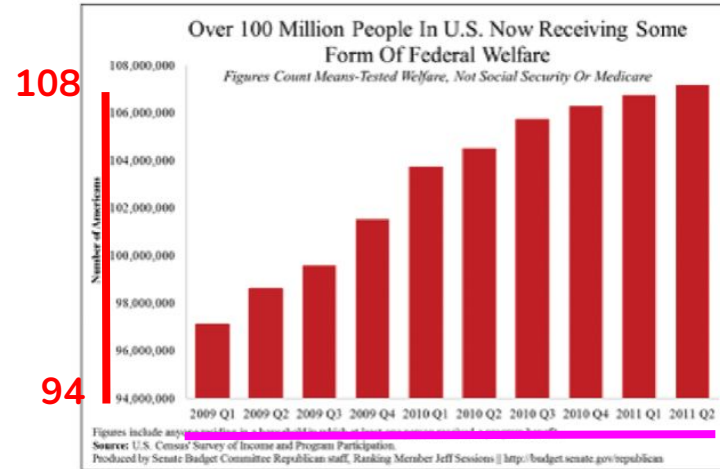
SHOWING OVERALL DIFFERENCE
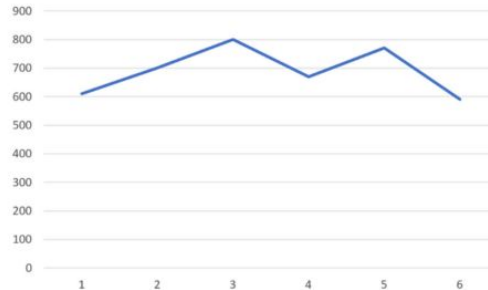
THE BLOG

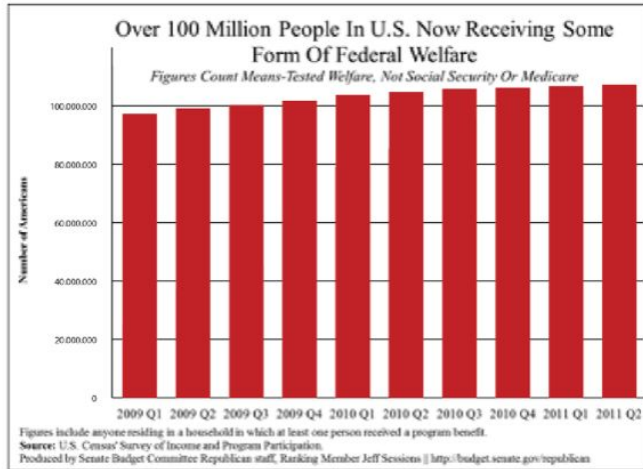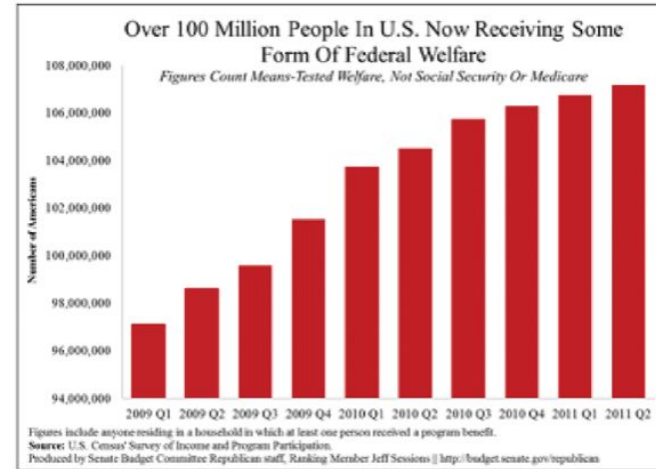### Over 100 Million Now Receiving Federal Welfare
2:40 PM, AUG 8, 2012 · BY DANIEL HALPER

A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."
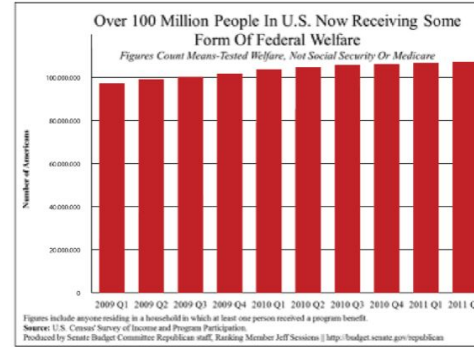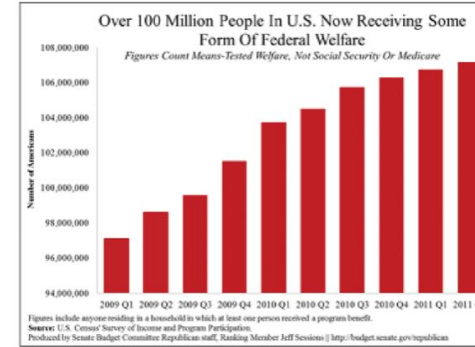
SHOW FLUCTUATION
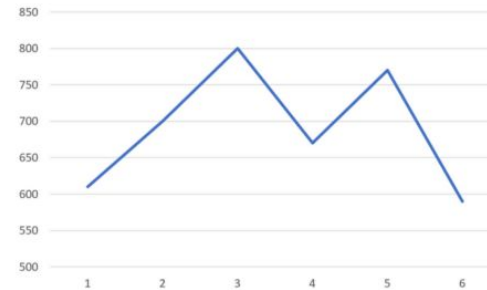
Graphical Excellence:

- …is the well-designed presentation of interesting data - a matter of substance, of statistics, and of design.
- …consists of complex ideas communicated with clarity, precision, and efficiency.
- …is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- …is nearly always multivariate.
- …requires telling the truth about the data.

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

$$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$$

$$= 1.0 - \text{proportion of a graphic that can be erased}$$

Where '**data-ink'**, is defined as the total ink on a graph that represents data

Ink - if it was printed on paper. In our case it can be amount of pixels, used to represent information

Data-Ink ratio principles
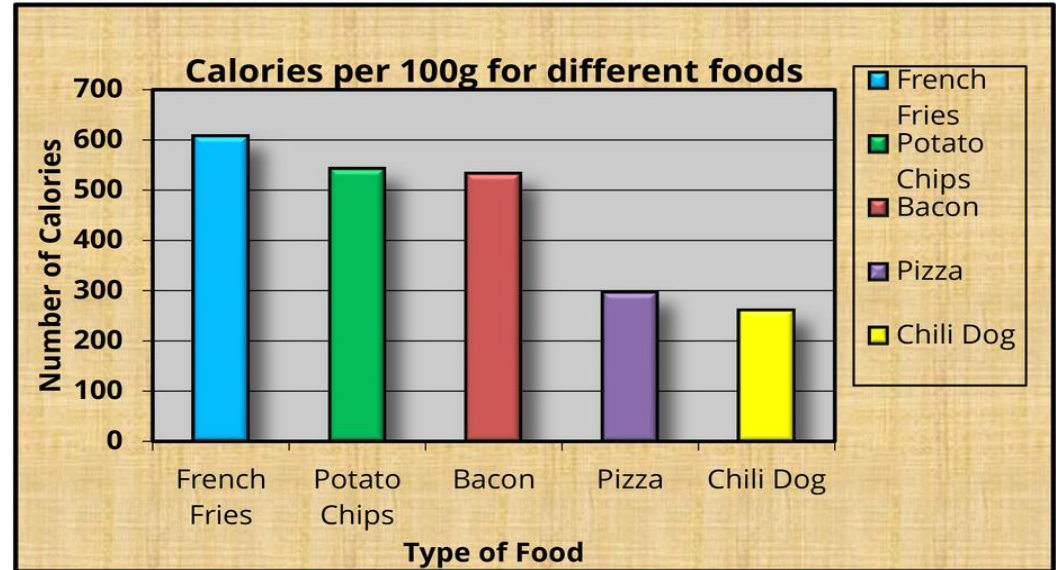
1. Above all else show the data.

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.

Remove backgrounds
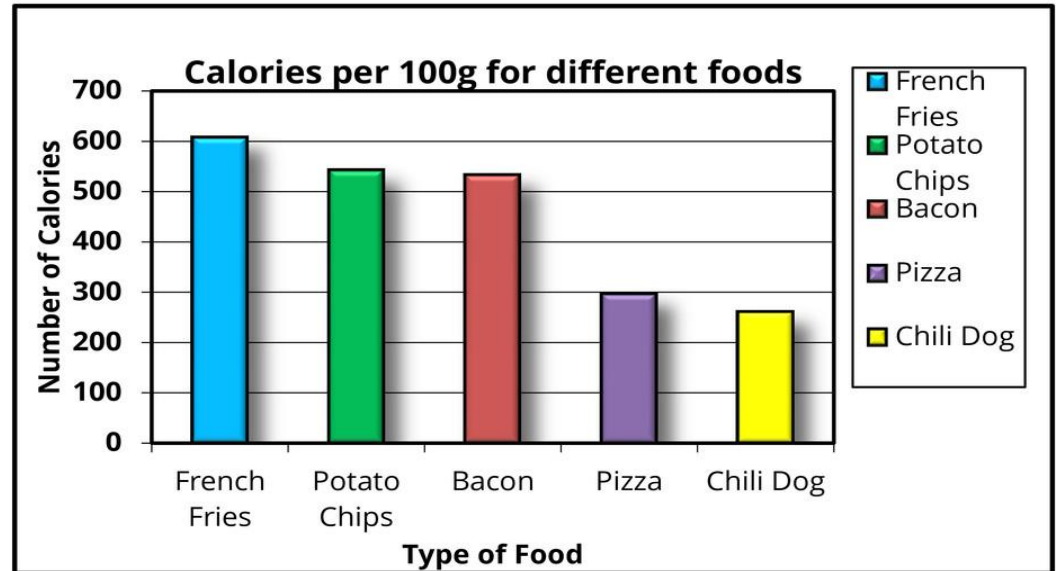
Calories per 100g for different foods

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.

3. Erase non-data ink.

Remove redundant labels

Calories per 100g

French Fries, Potato Chips, Bacon, Pizza, Chili Dog

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.

3. Erase non-data ink.

Remove borders

Calories per 100g

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.

3. Erase non-data ink.

4. Erase redundant data-ink.

Remove special effects

Calories per 100g

A bar chart with values:
- French Fries: ~605
- Potato Chips: ~540
- Bacon: ~530
- Pizza: ~295
- Chili Dog: ~255

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.

3. Erase non-data ink.

4. Erase redundant data-ink.

Remove bolding

Calories per 100g

French Fries | Potato Chips | Bacon | Pizza | Chili Dog

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.

3. Erase non-data ink.
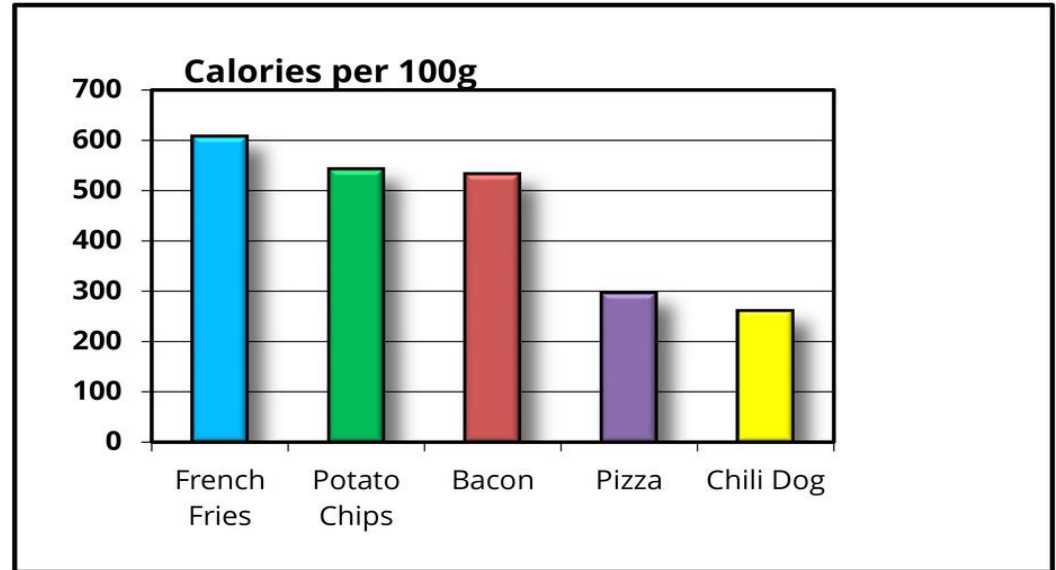
4. Erase redundant data-ink.

Lighten labels

Calories per 100g

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.

3. Erase non-data ink.

4. Erase redundant data-ink.

Lighten lines

Calories per 100g

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.

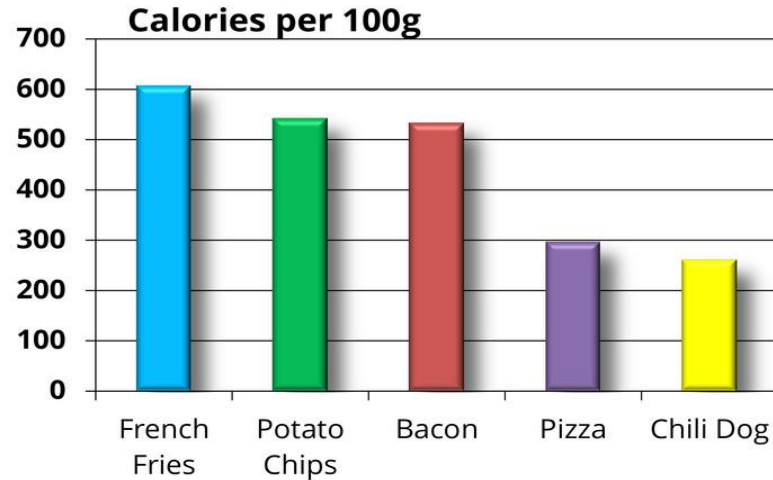3. Erase non-data ink.

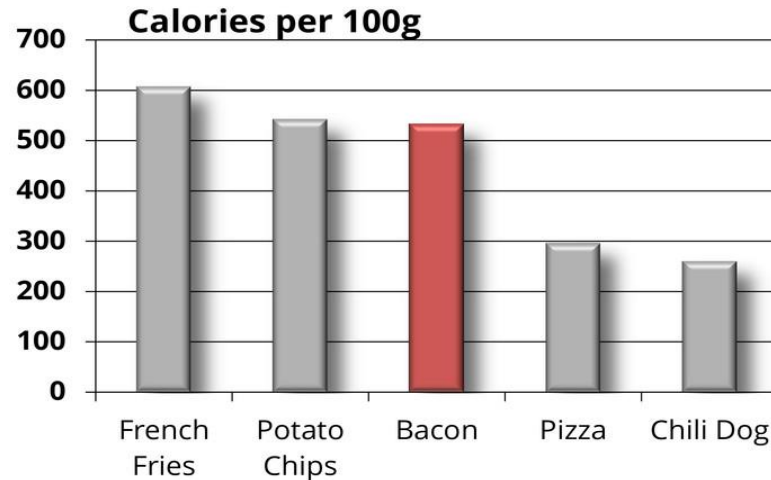4. Erase redundant data-ink.

Or remove lines

Calories per 100g

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.
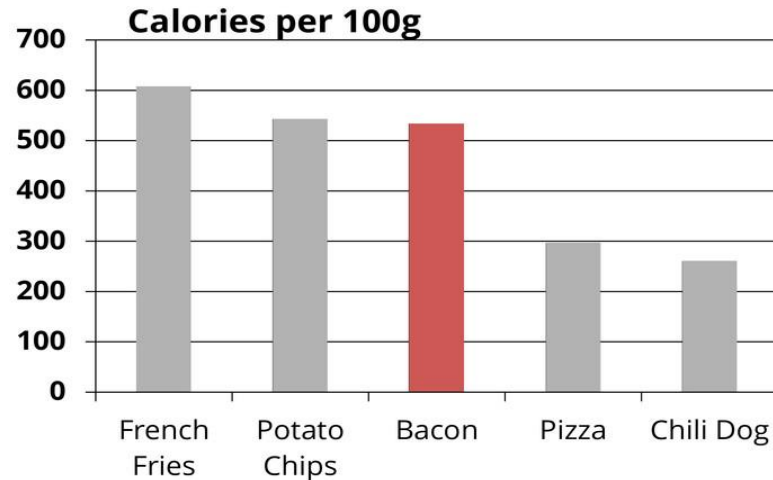
3. Erase non-data ink.

4. Erase redundant data-ink.

Direct label

Calories per 100g

| 607 | 542 | 533 | 296 | 260 |
| French Fries | Potato Chips | Bacon | Pizza | Chili Dog |

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.
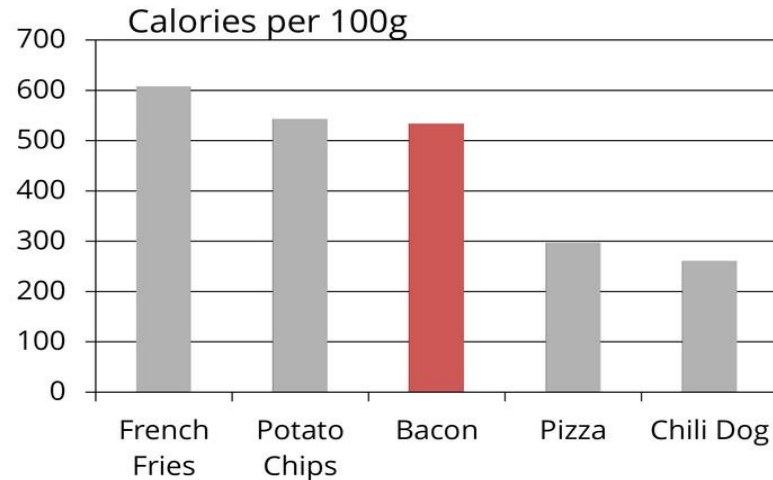
3. Erase non-data ink.

4. Erase redundant data-ink.

5. Revise and edit.

After

Calories per 100g

607	542	533	296	260
French Fries	Potato Chips	Bacon	Pizza	Chili Dog

Data-Ink ratio principles

1. Above all else show the data.

2. Maximize the data-ink ratio.
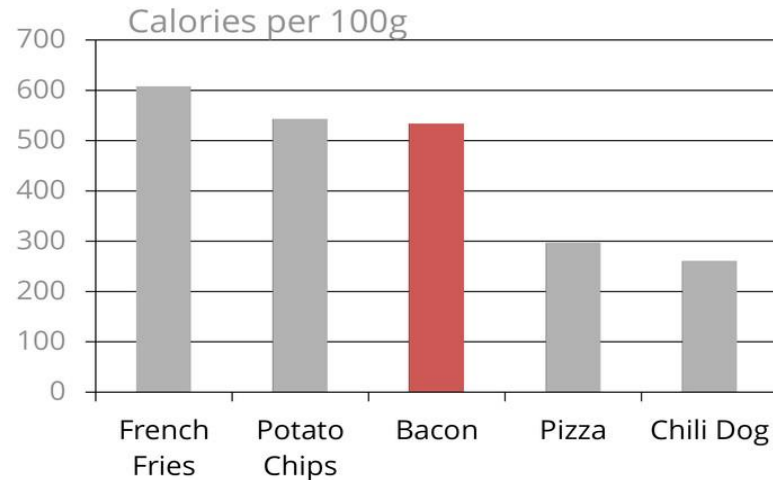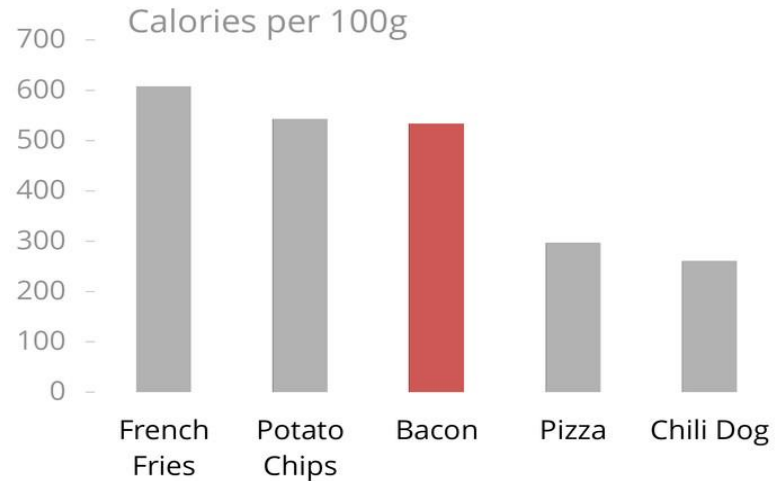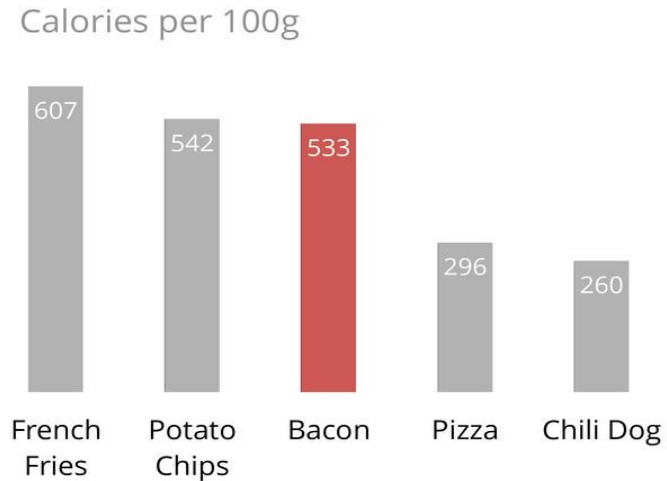
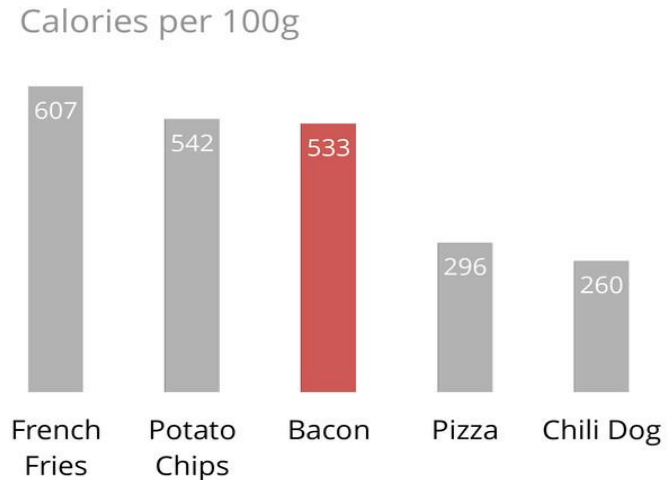3. Erase non-data ink.

4. Erase redundant data-ink.

5. Revise and edit.

Remove
to improve
(the **data-ink** ratio)

Created by **Darkhorse Analytics**   www.darkhorseanalytics.com

Texture and colour
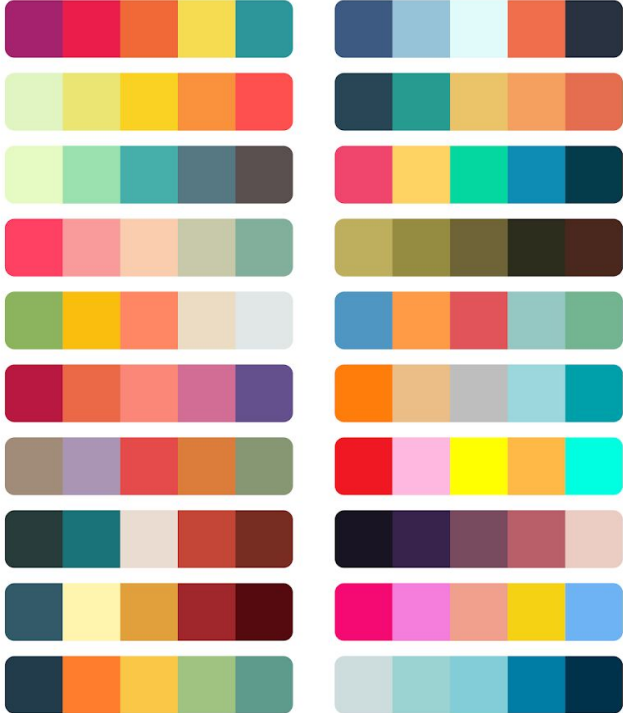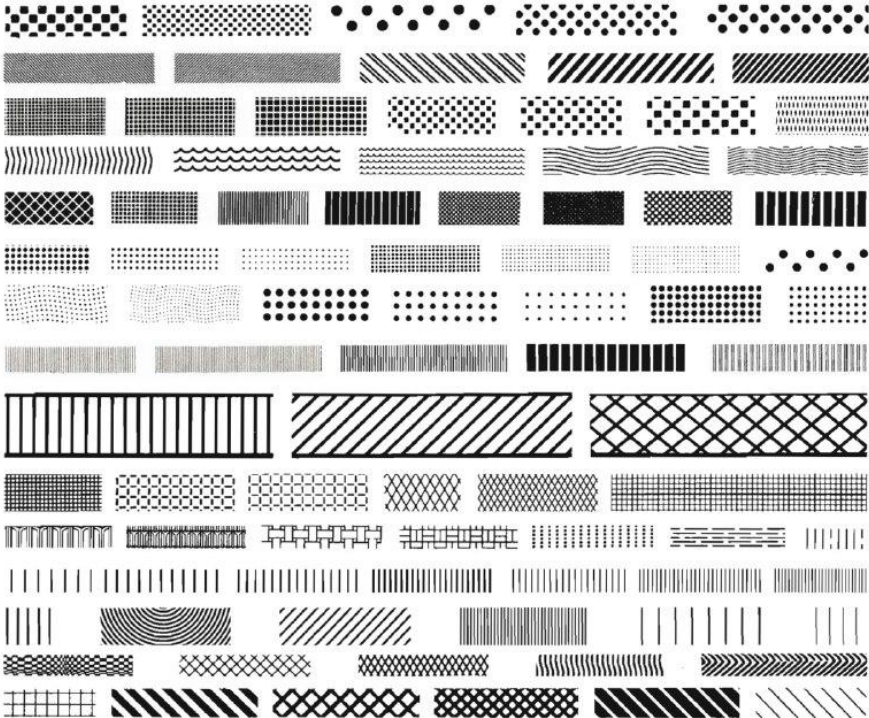
Data Ink good or bad examples

Types of Charts and Graphs to use for your Data

https://coolinfographics.com/dataviz-guides   https://datavizcatalogue.com (check resources page too)

Chart Suggestions—A Thought-Starter

© Dr. Andrew Abela https://extremepresentation.com

1. Comparison

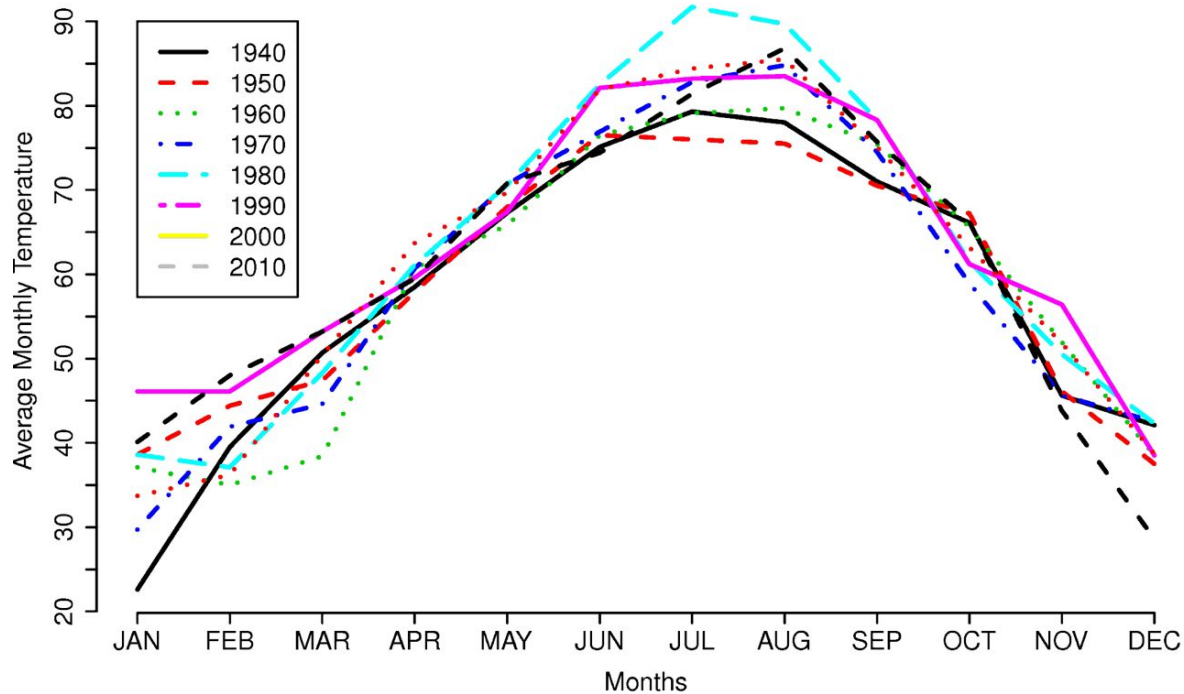| Variable Width Column Chart	Table or Table with Embedded Charts	Bar Chart	Column Chart	Circular Area Chart	Line Chart	Column Chart	Line Chart
Two Variables per Item	Many Categories	Many Items / Few Items		Cyclical Data	Non-Cyclical Data	Single or Few Categories	Many Categories
	One Variable per Item	Few Categories		Many Periods		Few Periods	
Among Items				Over Time			

Common questions are:
- What product sells best?
- How are our sales compared to last year?
- How many new users are we acquiring every day?
- Which referrers are driving the most traffic to our website?

Bar Charts, Error Bars and Dot Plots
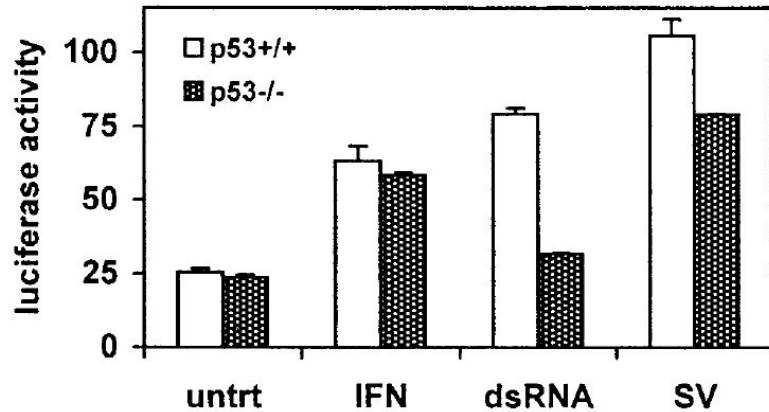
Hummer BT, Li XL, Hassel BA (2001) Role for p53 in gene induction by double-stranded RNA
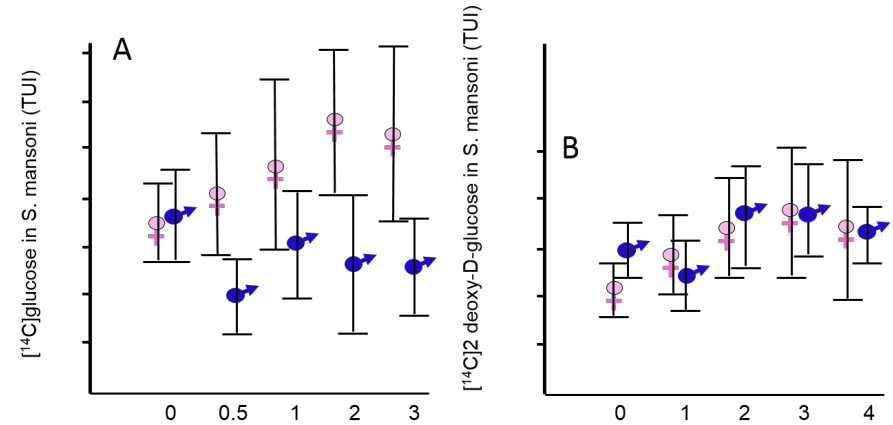
Bar Charts, Error Bars and Dot Plots

Bar Charts, Error Bars and Dot Plots

Bar Charts, Error Bars and Dot Plots
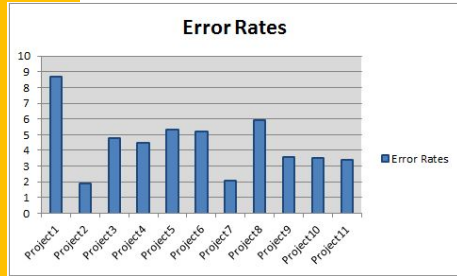
Bar Charts, Error Bars and Dot Plots
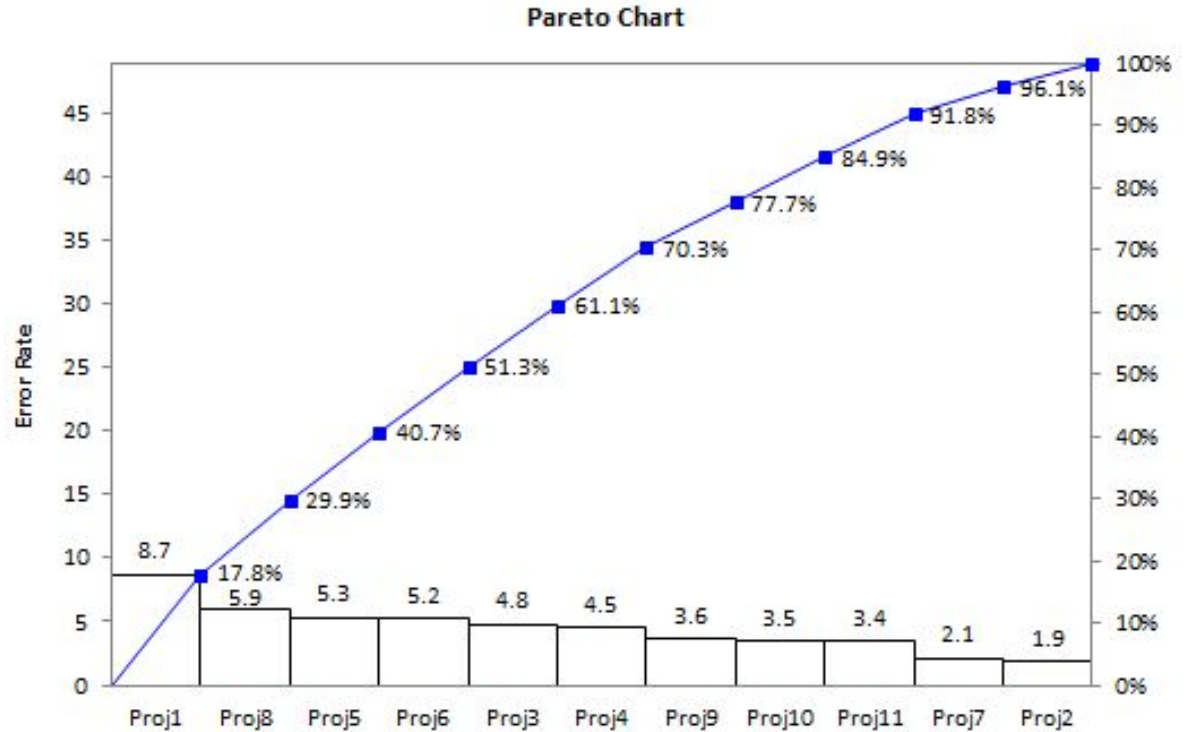
2. Composition
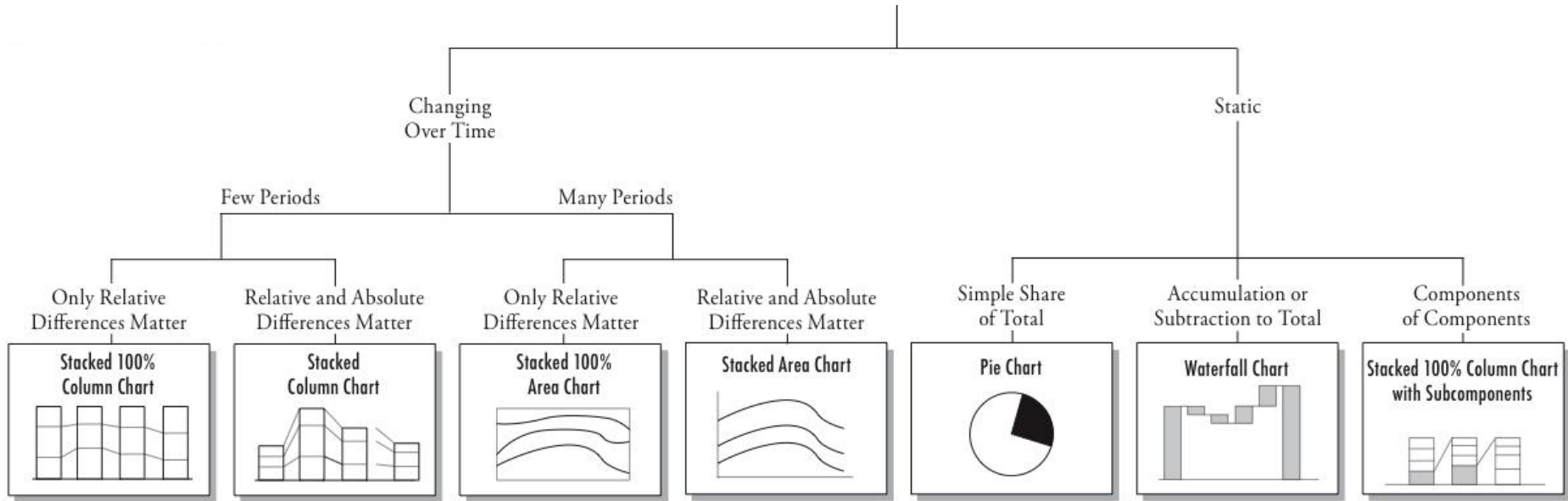
Common questions are:
- How big part of market to we have in a region?
- What percentages of our total sales come from which regions?
- What channels are these new users coming from?

Quick note about Pie Chart

The Worst Chart In The World https://www.businessinsider.com/pie-charts-are-the-worst-2013-6
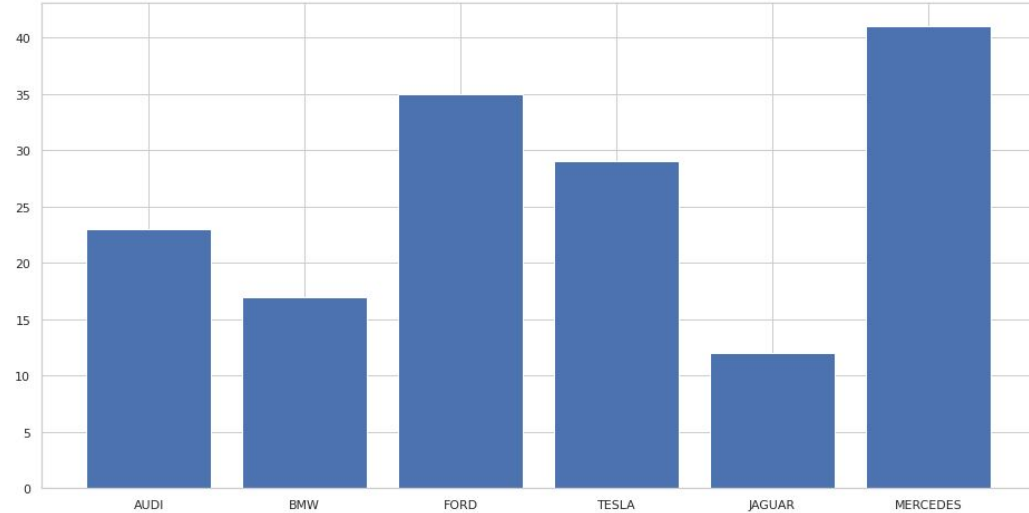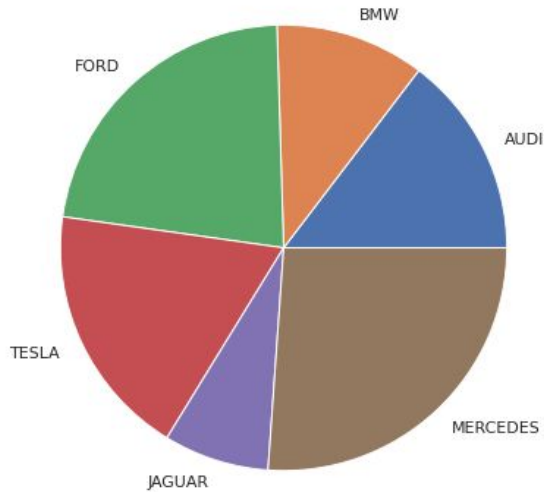
3. Distribution

Common questions are:
- Number of customers per age group?
- How many days late are our payments?

Single Variable — Few Data Points — **Column Histogram**

Single Variable — Many Data Points — **Line Histogram**

Two Variables — **Scatter Chart**

Three Variables — **3D Area Chart**

Single Continuous Numeric Variable
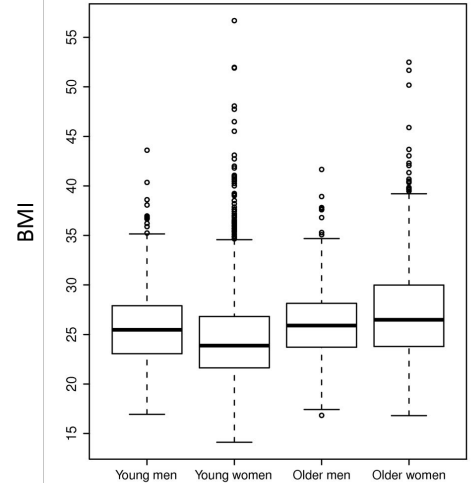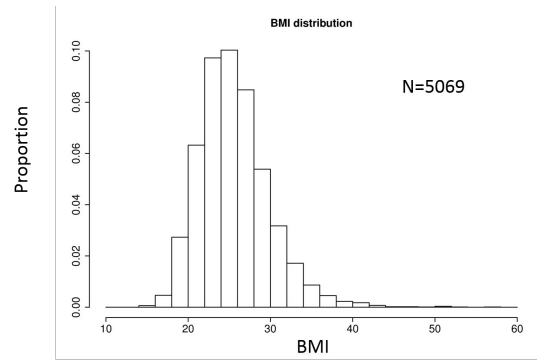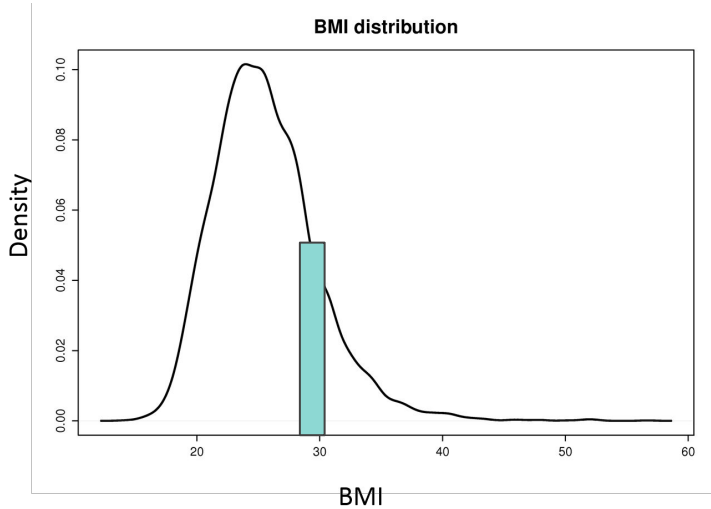
4. Relationship

Common questions are:
- Is there a correlation between advertising spend and sales for our products?
- How does expenses and income vary per night and what's the deviation?
- Which landing pages are driving the most engagement by channel?
- Where do we have opportunities to drive more traffic to high-performing web pages?

Scatter Chart — Two Variables

Bubble Chart — Three Variables
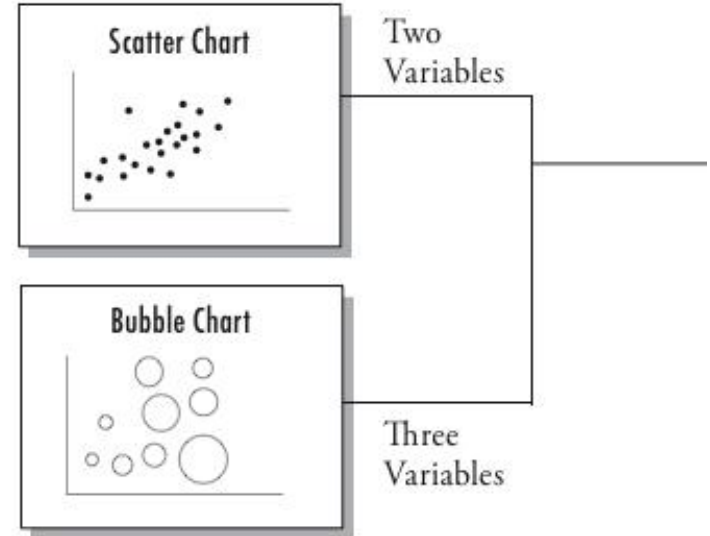
Multivariate Data
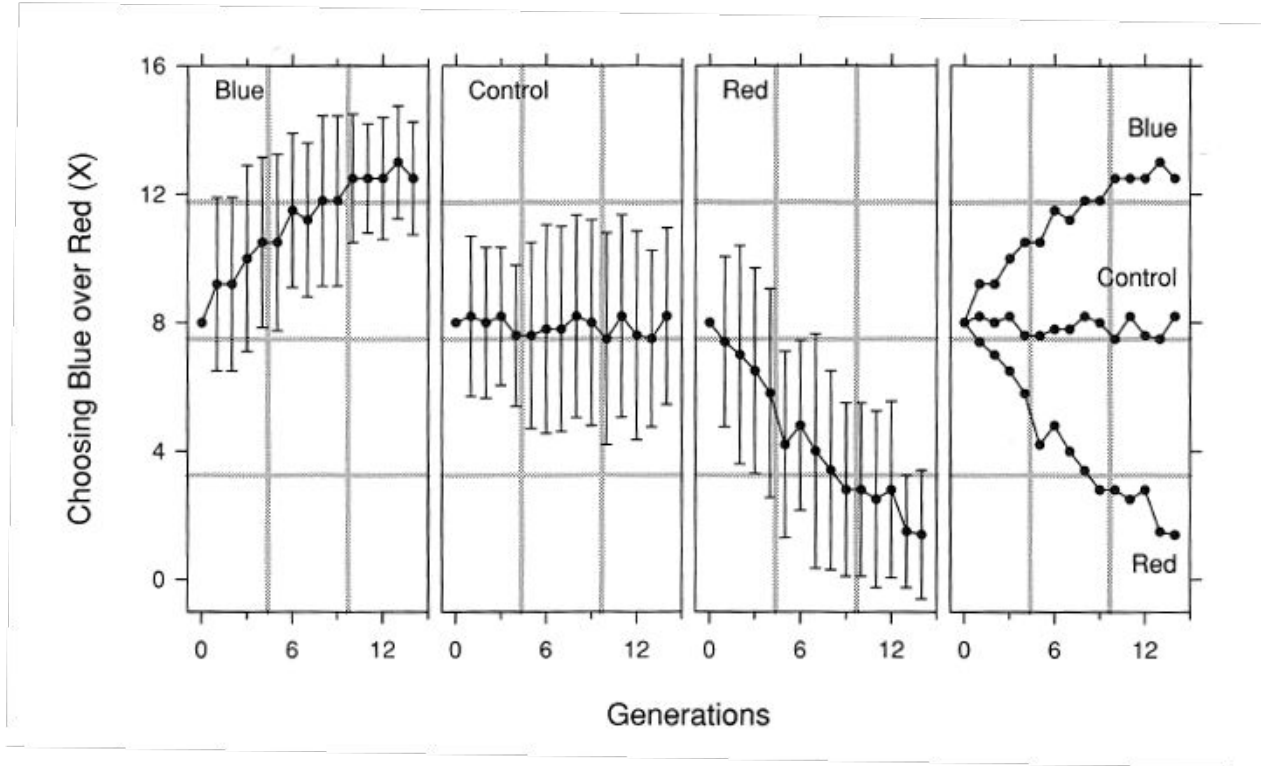
Q&A break

public.tableau.com/
app/discover/viz-of-t
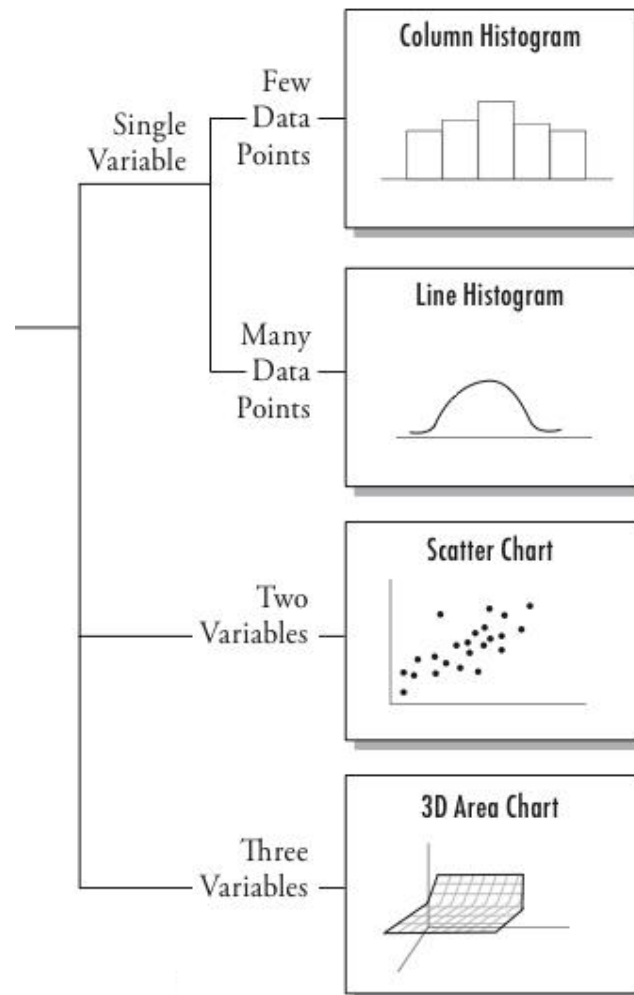he-day

www.tylervigen.com/
spurious-correlations

Datavizcatalogue.com
(check resources page too)

The Worst Chart In The World
www.businessinsider.com/pie-
charts-are-the-worst-2013-6?I
R=T#ixzz2WU7bwUlY

coolinfographics.com/
dataviz-guides

Practice

BI tools:

1. Microsoft Power BI

2. Tableau

3. Qlik Sense

4. ThoughtSpot

5. Looker

6. Google Data Studio

7. Gliffy, Datylon, Datamatic, Flourish, Vizzlo

8. R, Matplotlib, hvplot

9. etc

Summary:

- Avoid chart junk. Keep it simple and clear.

Summary:

- Avoid chart junk. Keep it simple and clear.

- Avoid pie charts, because humans have difficulty perceiving relative angles.

Summary:

- Avoid chart junk. Keep it simple and clear.

- Avoid pie charts, because humans have difficulty perceiving relative angles.

- Pay attention to scale.

Summary:

- Avoid chart junk. Keep it simple and clear.

- Avoid pie charts, because humans have difficulty perceiving relative angles.

- Pay attention to scale.

- Ask questions to the data about what you want to show

Summary:

- Avoid chart junk. Keep it simple and clear.

- Avoid pie charts, because humans have difficulty perceiving relative angles.

- Pay attention to scale.

- Ask questions to the data about what you want to show

- Know your data.

Summary:

- Avoid chart junk. Keep it simple and clear.

- Avoid pie charts, because humans have difficulty perceiving relative angles.

- Pay attention to scale.

- Ask questions to the data about what you want to show

- Know your data.

- Do not overload your chart - one chart one piece of information.

Summary:

- Avoid chart junk. Keep it simple and clear.

- Avoid pie charts, because humans have difficulty perceiving relative angles.

- Pay attention to scale.

- Ask questions to the data about what you want to show

- Know your data.

- Do not overload your chart - one chart one piece of information.

- Explore several ways to display the data

Tips on How to Display Data Badly：

- Show as few data as possible

- Hide what data you do show; minimize the data-ink ratio

- Ignore the visual metaphor altogether

- Only order matters

- Graph data out of context

- Change scales in mid-axis

- Emphasize the trivial; ignore the important

- Jiggle the baseline

- Alphabetize everything.

- Make your labels illegible, incomplete, incorrect, and ambiguous.

- More is murkier: use a lot of decimal places and make your graphs three dimensional whenever possible.

- If it has been done well in the past, think of another way to do it

Links

Books

- Show Me the Numbers
  https://www.amazon.com/Show-Me-Numbers-Designing-Enlighten/dp/0970601999

- The Visual Display of Q Information by Edward Tufte and his blog https://www.edwardtufte.com/tufte

How to choose visualisation tool

- https://coolinfographics.com/dataviz-guides and https://datavizcatalogue.com

- Stephen Few: Designing Effective Tables and Graphs.
  http://www.perceptualedge.com/images/Effective_Chart_Design.pdf

Other links

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

- Vis of day https://public.tableau.com/app/discover/viz-of-the-day

- https://infovis-wiki.net

Grid Dynamics

trusted engineering partner for digital transformation

Thank you!

Let's chat!
Slack: @iklementeva

Grid Dynamics Holdings Inc.

5000 Executive Parkway,

Suite 520 / San Ramon, CA

+1 650-523-5000

info@griddynamics.com

www.griddynamics.com