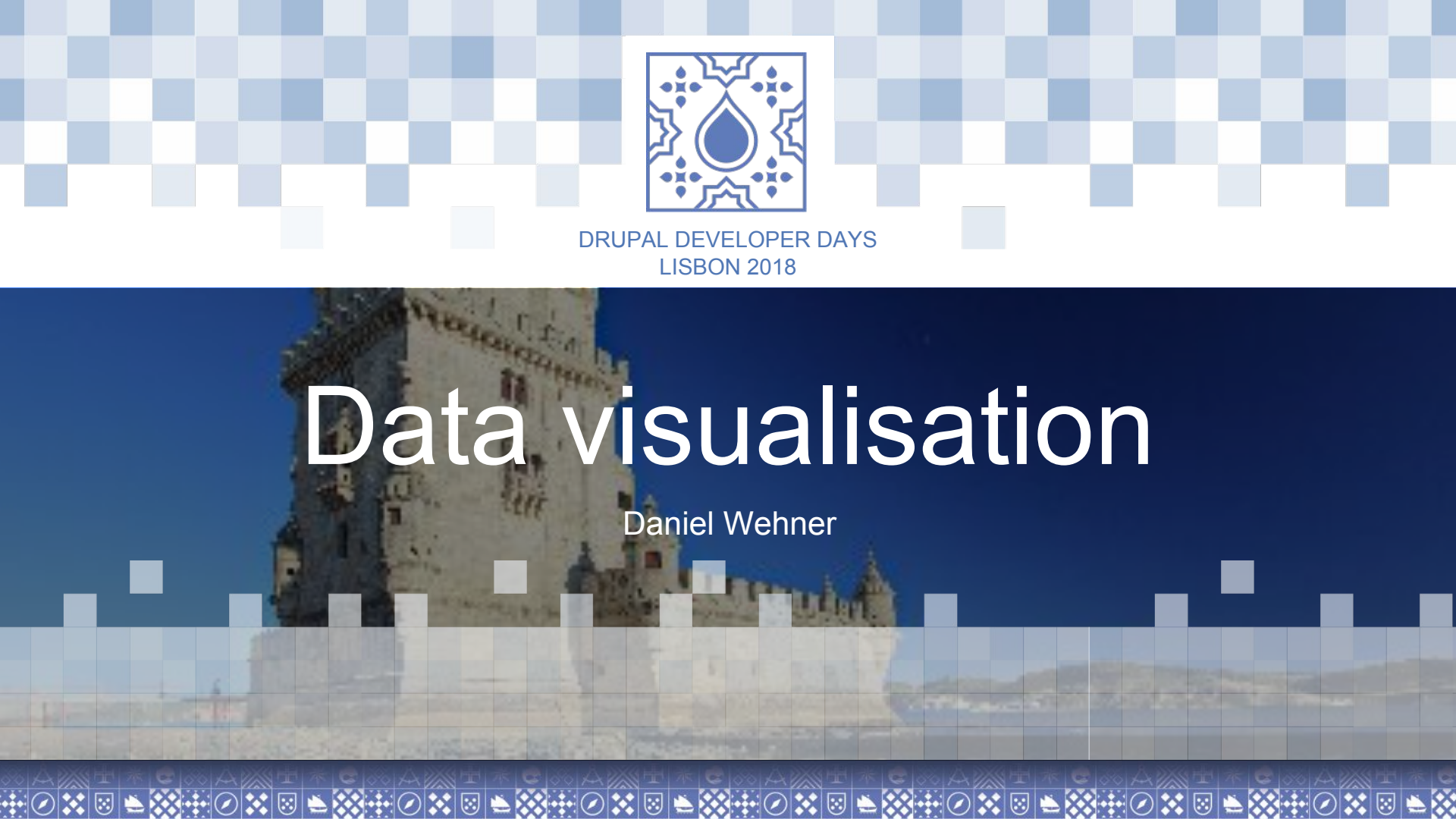

DRUPAL DEVELOPER DAYS
LISBON 2018
Data visualisation
Daniel Wehner

Diamond Sponsor

Platinum Sponsors

Gold Sponsors

Data visualisation

Daniel Wehner

Engineer as Times Educational Supplement (TES)

twitter.com/da_wehner

twitter.com/dawehnerquotes

Yards
50 0 50 100 150 200
× Pump • Deaths from cholera
OXFORD STREET
GT. MARLBOROUGH ST.
REGENT STREET
KING STREET
BROAD STREET
WARDOUR STREET
DEAN STREET
CONDUIT STREET
SAVILLE ROW
GOLDEN SQUARE
BREWER STREET
NEW BOND STREET
REGENTS QUADRANT
PICCADILLY

OXFORD
REGENT STREET
GT. MARLBOROUGH ST.
BROAD STREET
WARDOUR STREET
DEAN STREET
KING STREET
NDUIT STREET
S T R E E T
GOLDEN SQUARE

Explore data

Explain data

Exploratory data analysis

“If a thing is not worth doing, it is not worth doing well.”

— **John W. Tukey**

Content model

Flat format

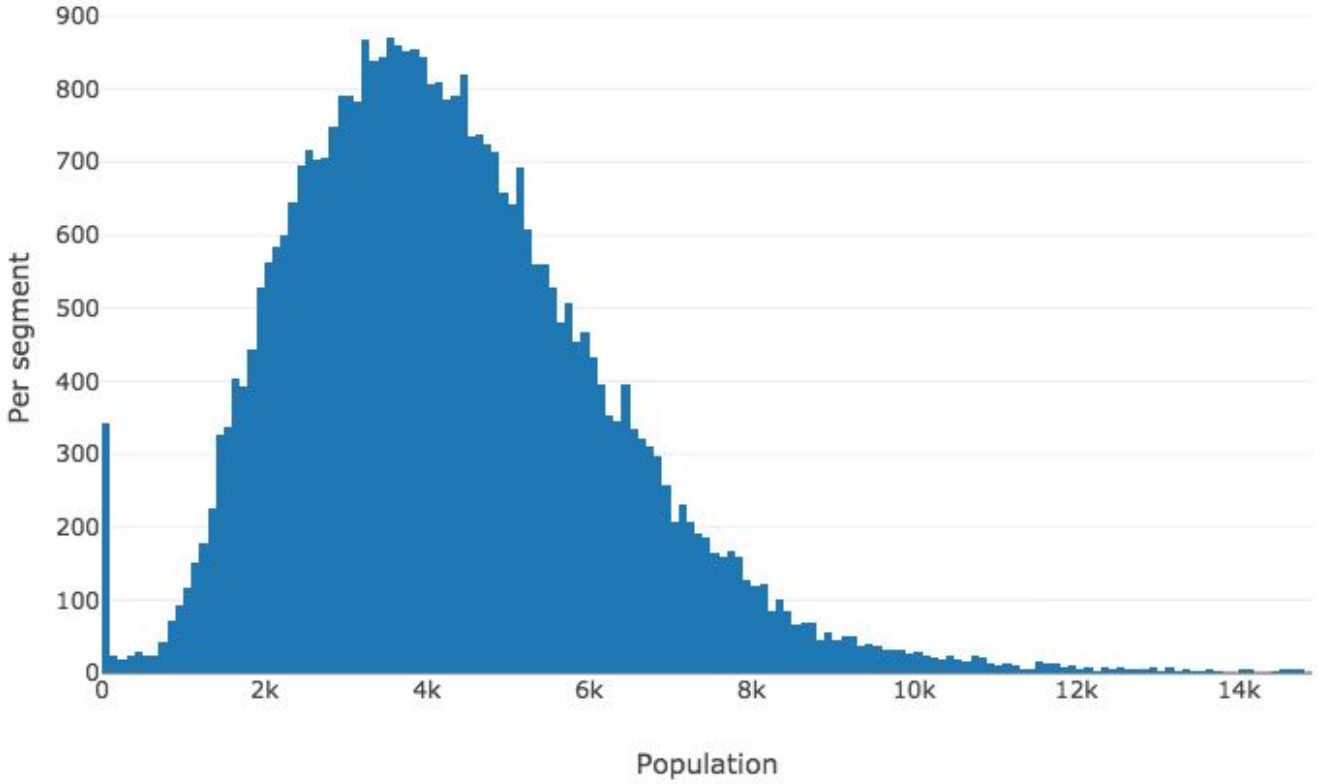
	UID	BLOCKID	SUMLEVEL	COUNTYID	STATEID	state	state_ab	city	place	type	...	female_age_mean
0	220336	NaN	140	16	2	Alaska	AK	Unalaska	Unalaska City	City	...	32.78177
1	220342	NaN	140	20	2	Alaska	AK	Eagle River	Anchorage	City	...	38.97956
2	220343	NaN	140	20	2	Alaska	AK	Jber	Anchorage	City	...	22.20427
3	220345	NaN	140	20	2	Alaska	AK	Anchorage	Point Mackenzie	City	...	37.00750
4	220347	NaN	140	20	2	Alaska	AK	Anchorage	Anchorage	City	...	34.96611
5	220348	NaN	140	20	2	Alaska	AK	Anchorage	Anchorage	City	...	33.63770
6	220349	NaN	140	20	2	Alaska	AK	Anchorage	Anchorage	City	...	34.54987
7	220350	NaN	140	20	2	Alaska	AK	Anchorage	Point Mackenzie	City	...	31.90301
8	220352	NaN	140	20	2	Alaska	AK	Anchorage	Point Mackenzie	City	...	30.64458

click to unsc

Ask a question

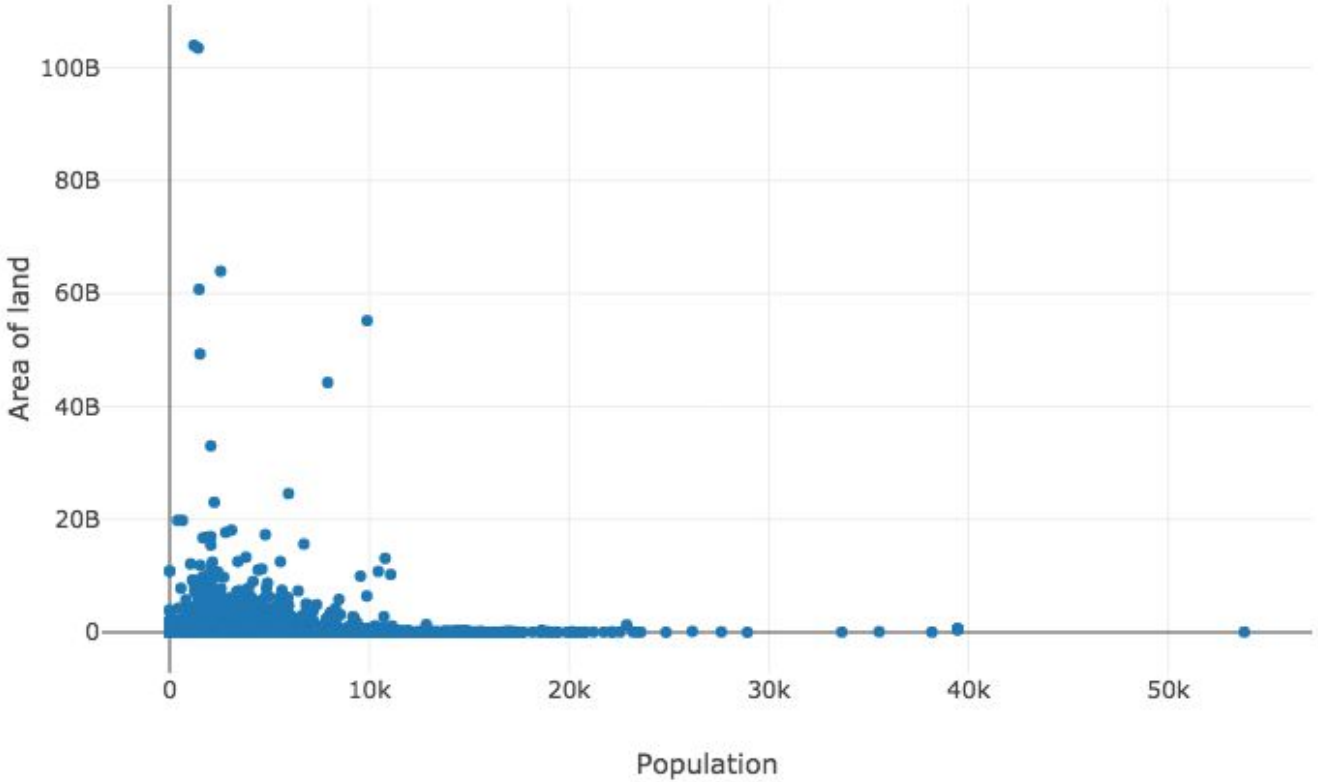
One continuous variable

	pop
0	4619
1	3727
2	8736
3	1941
4	5981
5	5476


900
800
700
600
500
400
300
200
100
0
Per segment
0
2k
4k
6k
8k
10k
12k
14k
Population

Two continuous variables

	ALand	pop
0	2823180154	4619
1	509234898	3727
2	270593047	8736
3	2371512	1941
4	1979230	5981
5	2089774	5476
6	2936374	5893
7	2494426	7481
8	4467475	5501
9	1146156	3225
10	1749931	3416


100B
80B
60B
40B
20B
0
Area of land
0
10k
20k
30k
40k
50k
Population

Data crisis

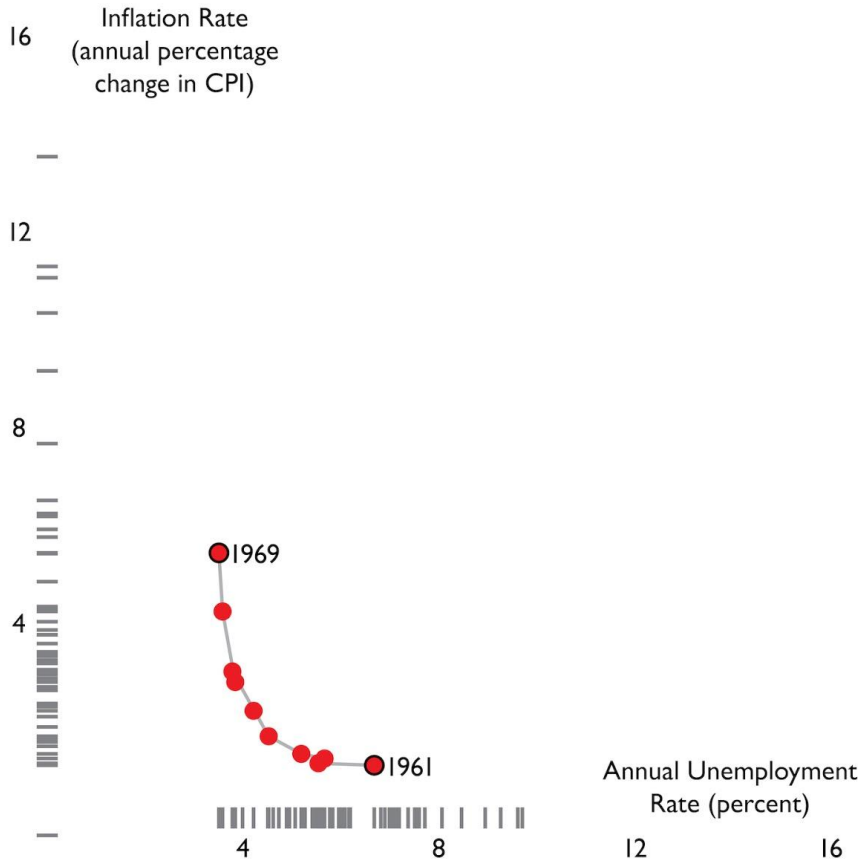
"The data may not contain the answer.

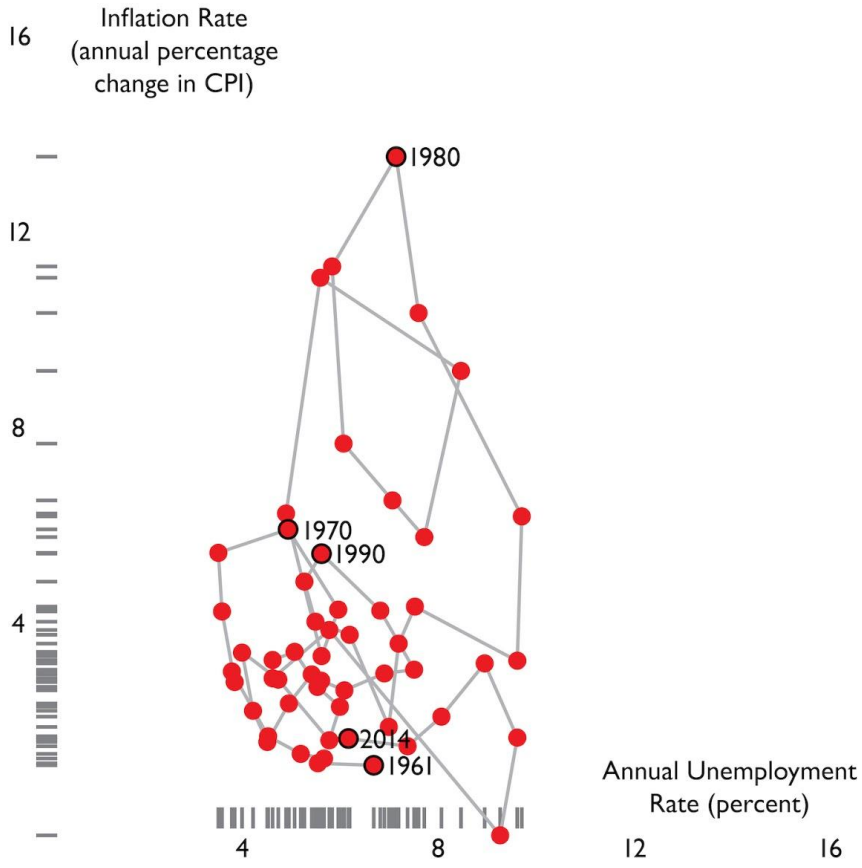
And, if you torture the data long enough,

it will tell you anything."

— **John W. Tukey**

Phillips curve, 1961 – 2014, United States





Cleanup

One categorical variable

One continuous variable

Married percentage
1
0.8
0.6
0.4
0.2
0
Alaska
Arkansas
California
Connecticut
Delaware
Georgia
Iowa
Illinois
Kansas
Louisiana
Maryland
Michigan
Missouri
Montana
North Dakot
New Hampsl
New Mexico
New York
Oklahoma
Pennsylvanil
Rhode Islan
South Dakot
Texas
Virginia
Washington
West Virgini

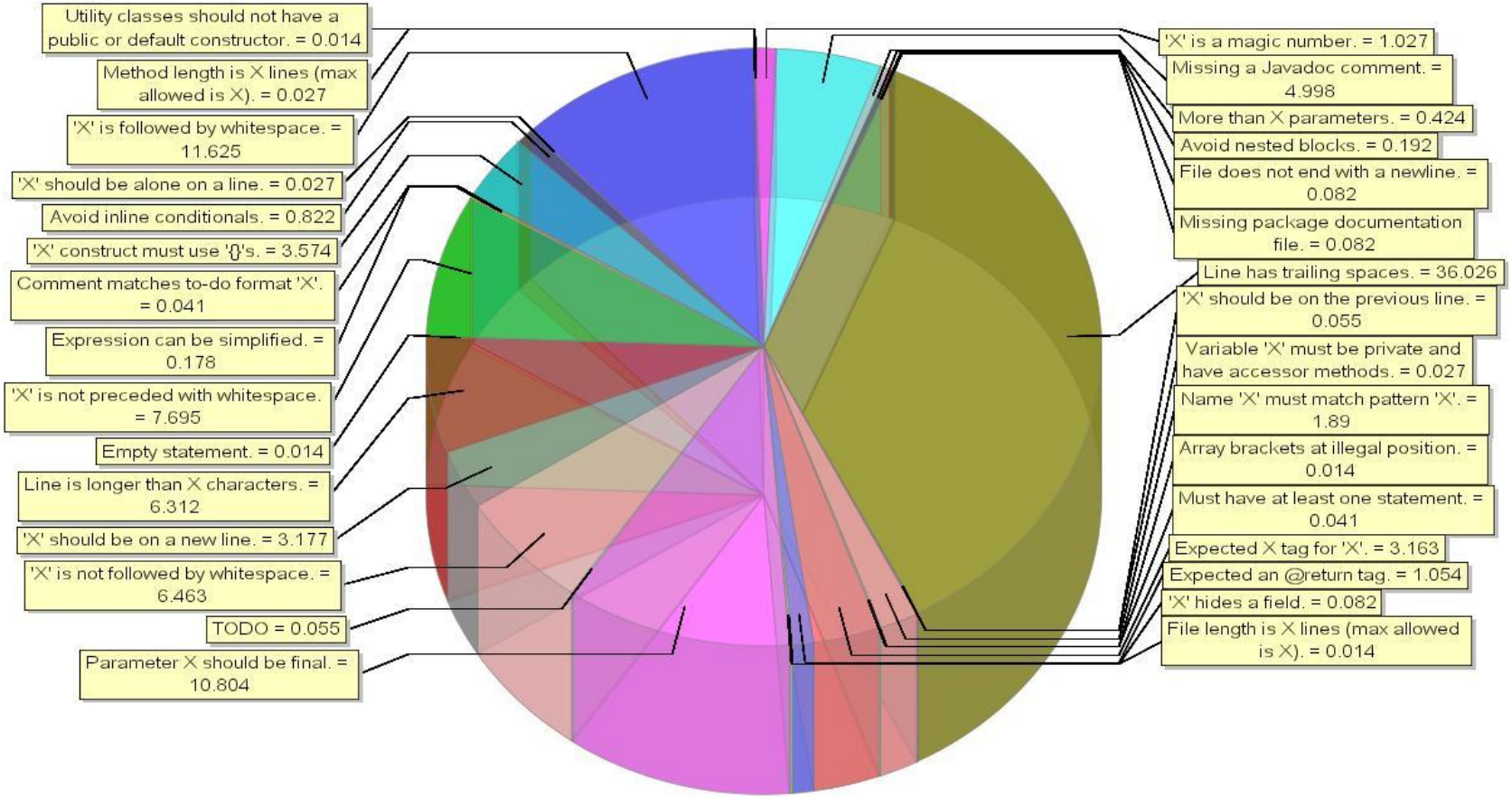
Married percentage
0.6
0.5
0.4
0.3
0.2
0.1
0
District of C
Puerto Rico
Louisiana
Mississippi
Nevada
Alaska
New York
Michigan
New Mexico
Illinois
Massachuse
Ohio
Delaware
California
Florida
Pennsylvani
Rhode Islan
Georgia
Alabama
Indiana
Maryland
Arizona
Wisconsin
Connecticut
Missouri
West Virgini
South Caroli
North Dakot
Oklahoma
Hawaii
Tennessee
Washington
New Jersey
Oregon
Kentucky
Texas
Minnesota
Colorado
Arkansas
Maine
Virginia
Vermont
Kansas
North Caroli
New Hampsi
South Dakot
Montana
Iowa
Wyoming
Nebraska
Idaho
Utah

Summary

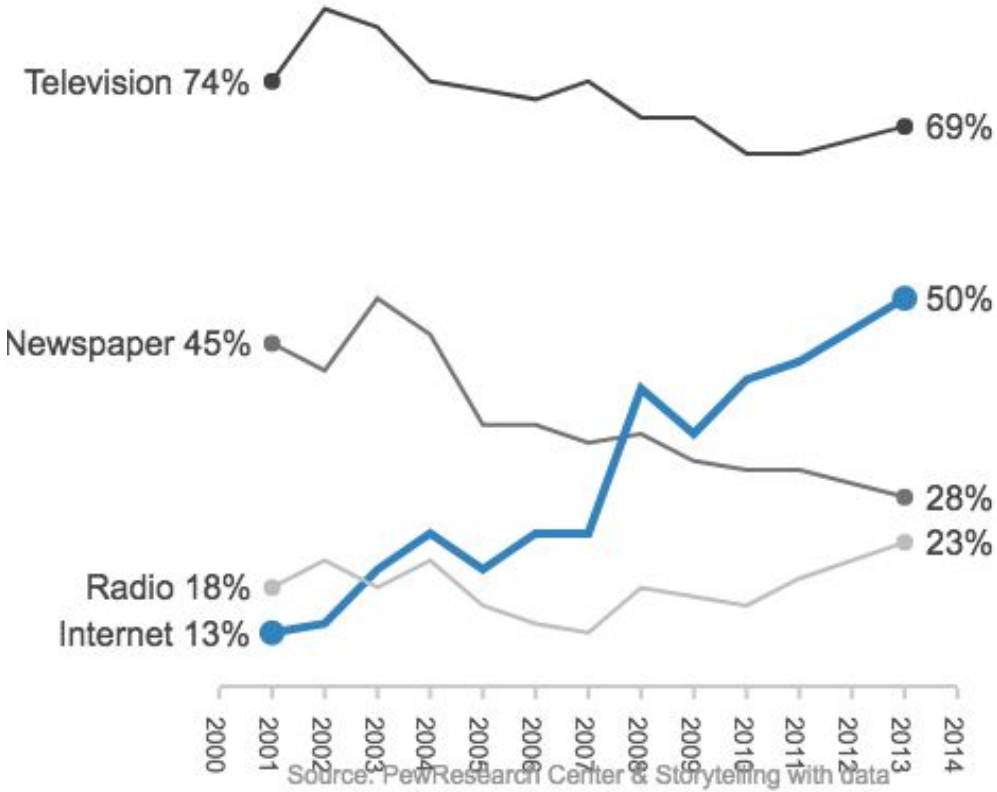
1. Flatten your data
2. Ask a question
3. Choose what data to plot
4. Pick your plot
5. Refine / Cleanup

Explain the data

Explain data

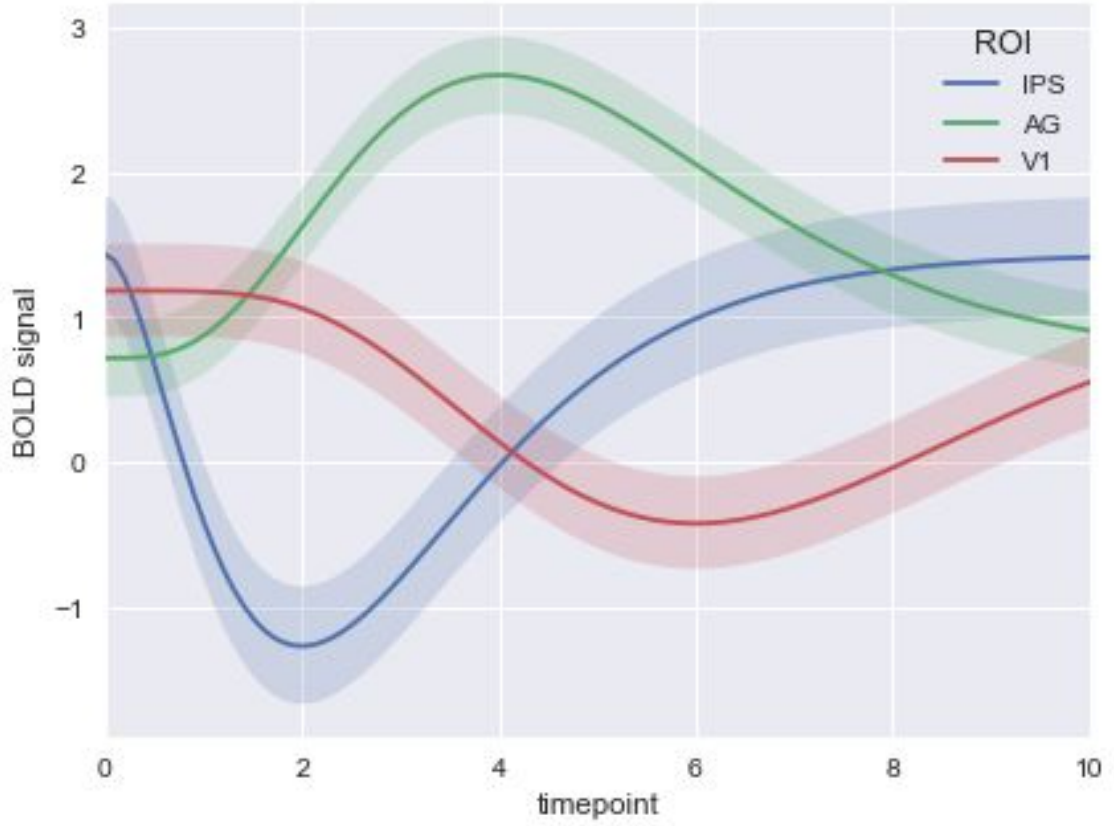

Utility classes should not have a public or default constructor. = 0.014
Method length is X lines (max allowed is X). = 0.027
'X' is followed by whitespace. = 11.625
'X' should be alone on a line. = 0.027
Avoid inline conditionals. = 0.822
'X' construct must use '{}'s. = 3.574
Comment matches to-do format 'X'. = 0.041
Expression can be simplified. = 0.178
'X' is not preceded with whitespace. = 7.695
Empty statement. = 0.014
Line is longer than X characters. = 6.312
'X' should be on a new line. = 3.177
'X' is not followed by whitespace. = 6.463
TODO = 0.055
Parameter X should be final. = 10.804
'X' is a magic number. = 1.027
Missing a Javadoc comment. = 4.998
More than X parameters. = 0.424
Avoid nested blocks. = 0.192
File does not end with a newline. = 0.082
Missing package documentation file. = 0.082
Line has trailing spaces. = 36.026
'X' should be on the previous line. = 0.055
Variable 'X' must be private and have accessor methods. = 0.027
Name 'X' must match pattern 'X'. = 1.89
Array brackets at illegal position. = 0.014
Must have at least one statement. = 0.041
Expected X tag for 'X'. = 3.163
Expected an @return tag. = 1.054
'X' hides a field. = 0.082
File length is X lines (max allowed is X). = 0.014

Main Source for News



Reduce thinking

Headings, labels, 0-start



Distracting design

- Add visual noise
  - Don’t emphasize out the important point
- Bullet points due to lack of better ideas

Distracting design

- Add visual noise
  - Don’t emphasize out the important point
- Bullet points due to lack of better ideas

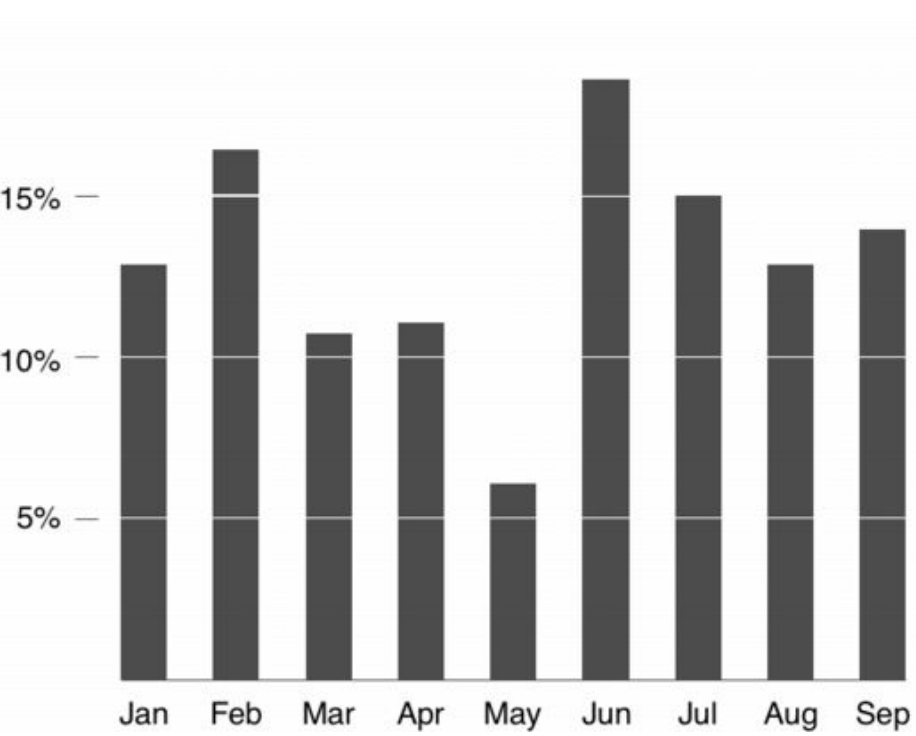
Distracting design

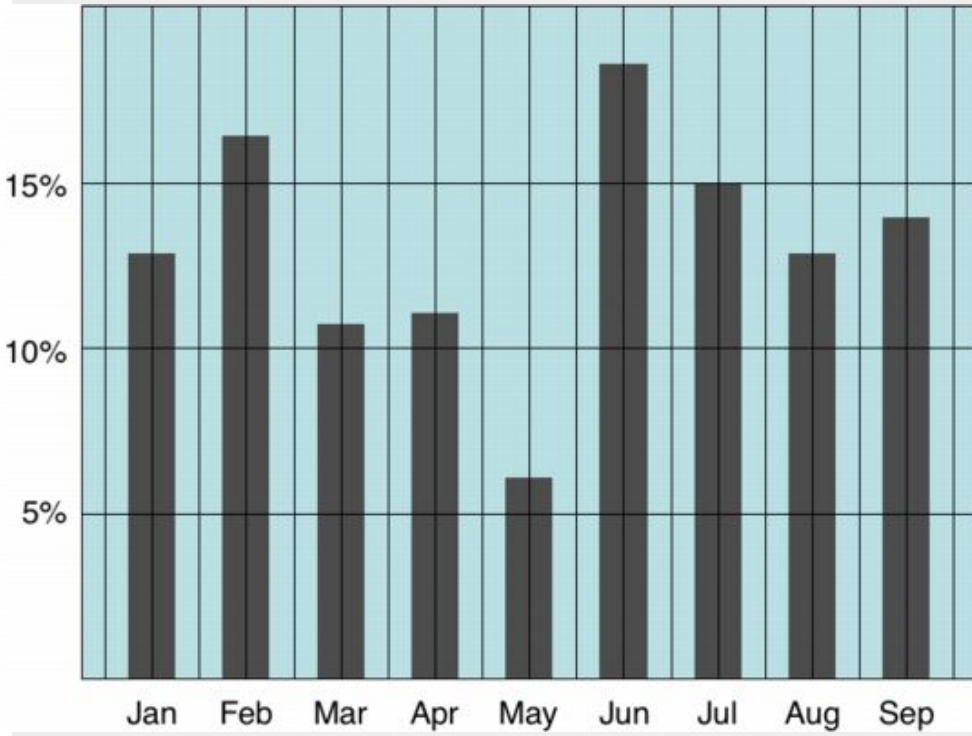
- Add visual noise
  - Don’t emphasize out the important point
- Bullet points due to lack of better ideas

Distracting design

- Add visual noise
  - Don’t emphasize out the important point
- Bullet points due to lack of better ideas

Data ink ratio


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep

Use tables

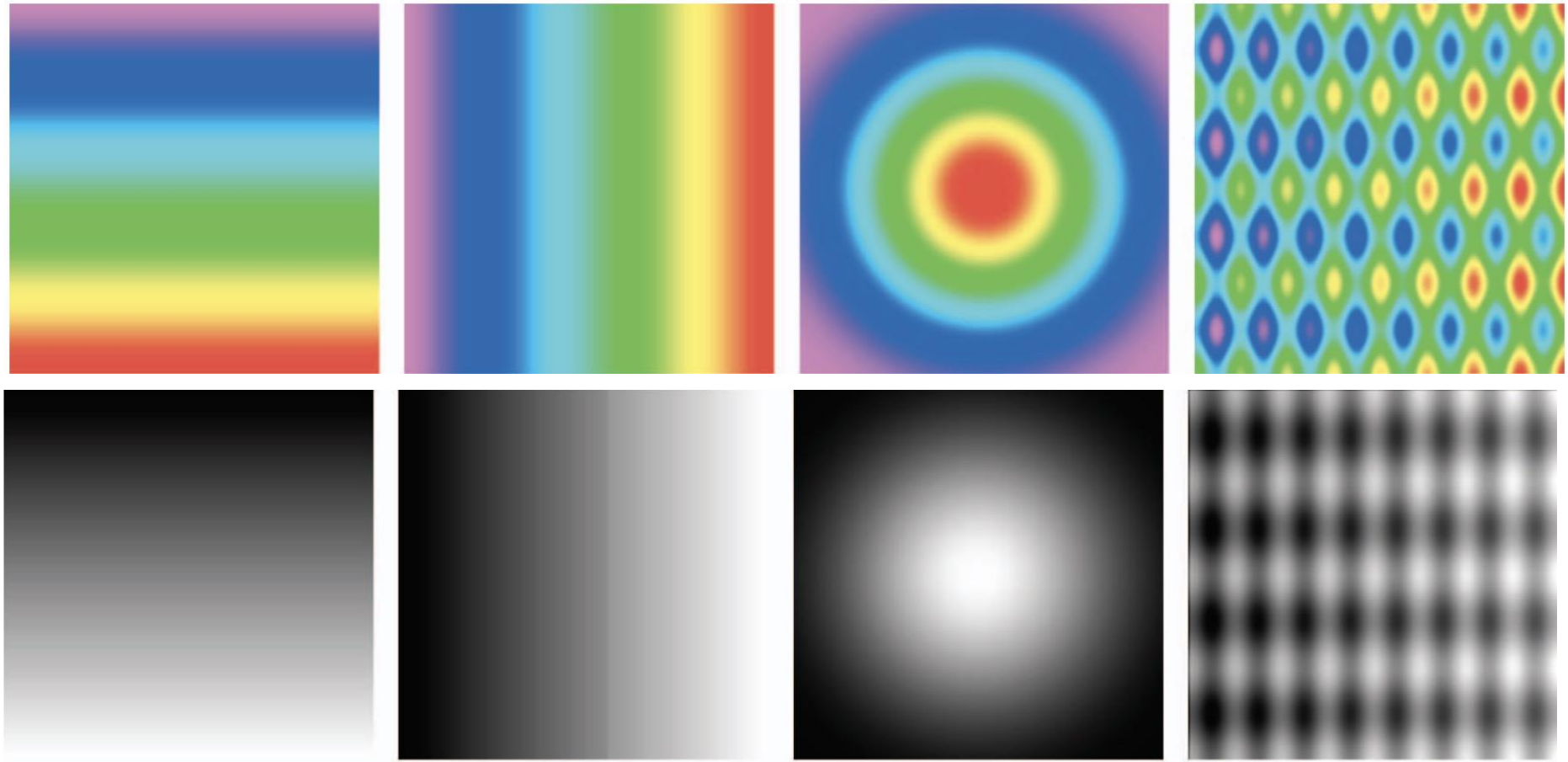
Use tables

slices	abs. error (%)		abs. error (slices)	
	avg.	max.	avg.	max
< 5000	7.4	73.5	116	625
5000–10000	3.1	27.2	209	1807
10000–15000	2.4	15.6	297	2133
> 15000	1.8	9.0	317	1609

Colors

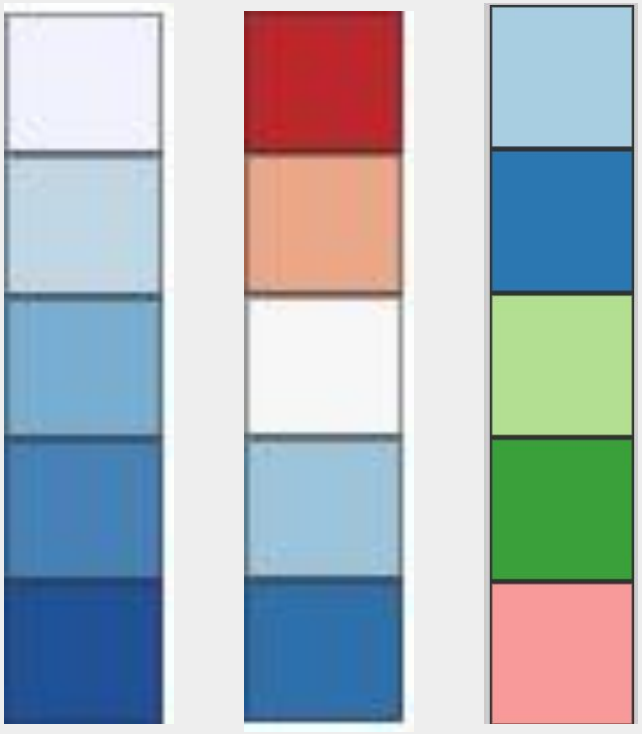
Colours

Find the differences



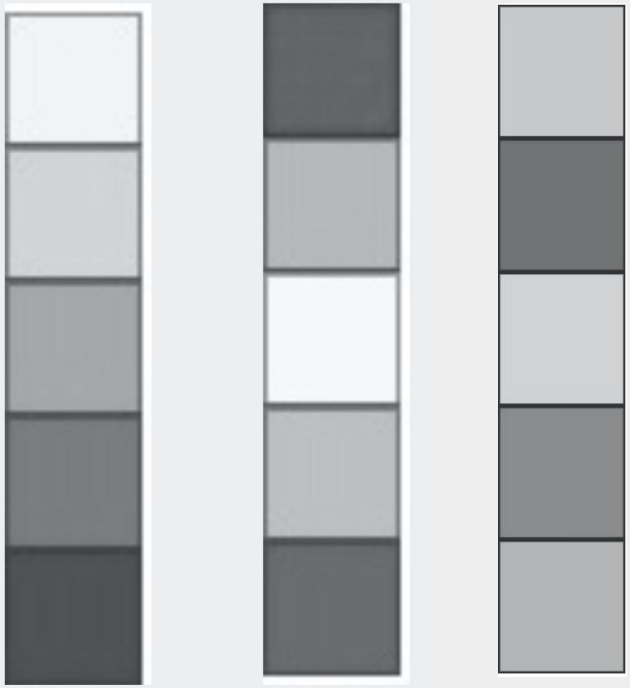
More colours

https://colorbrewer2.org

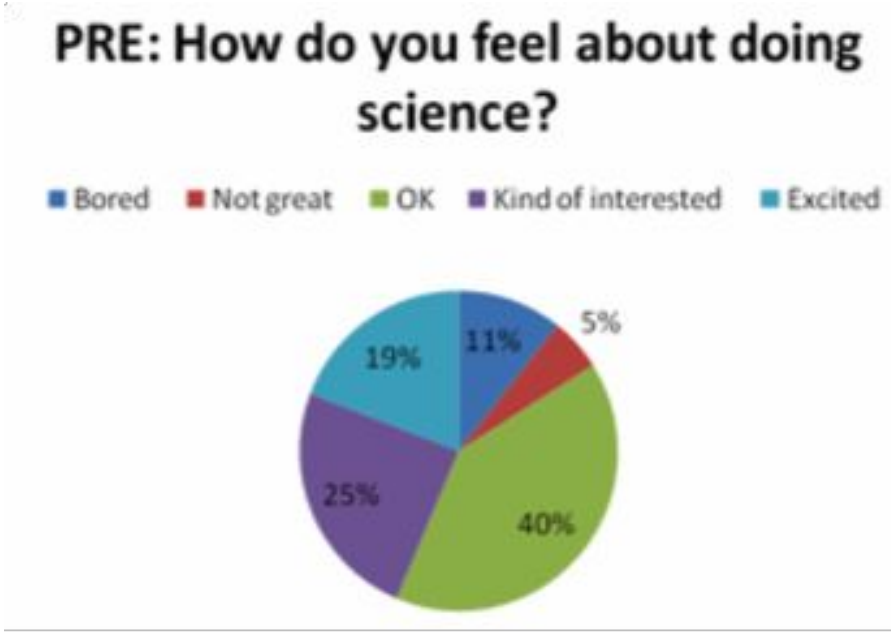


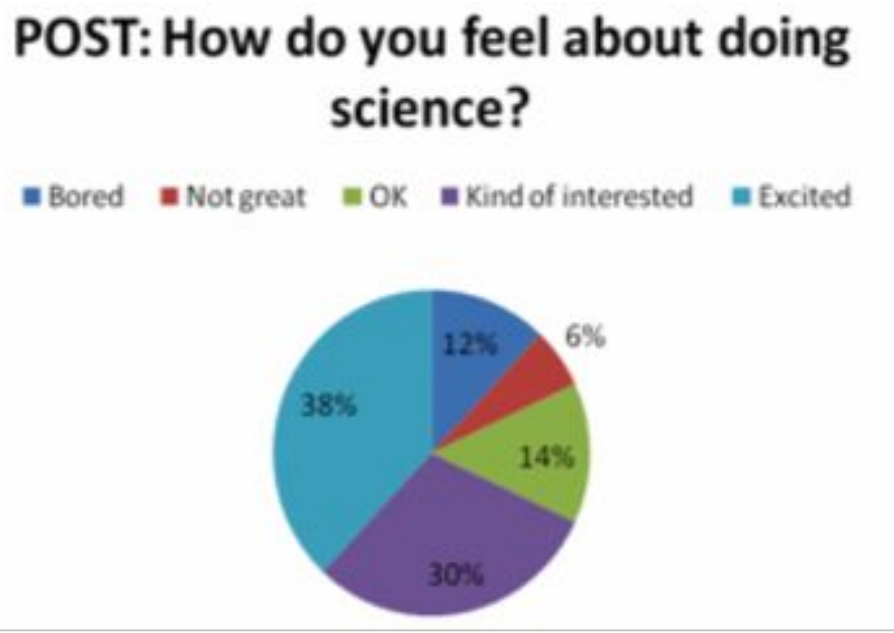
More colours

https://colorbrewer2.org

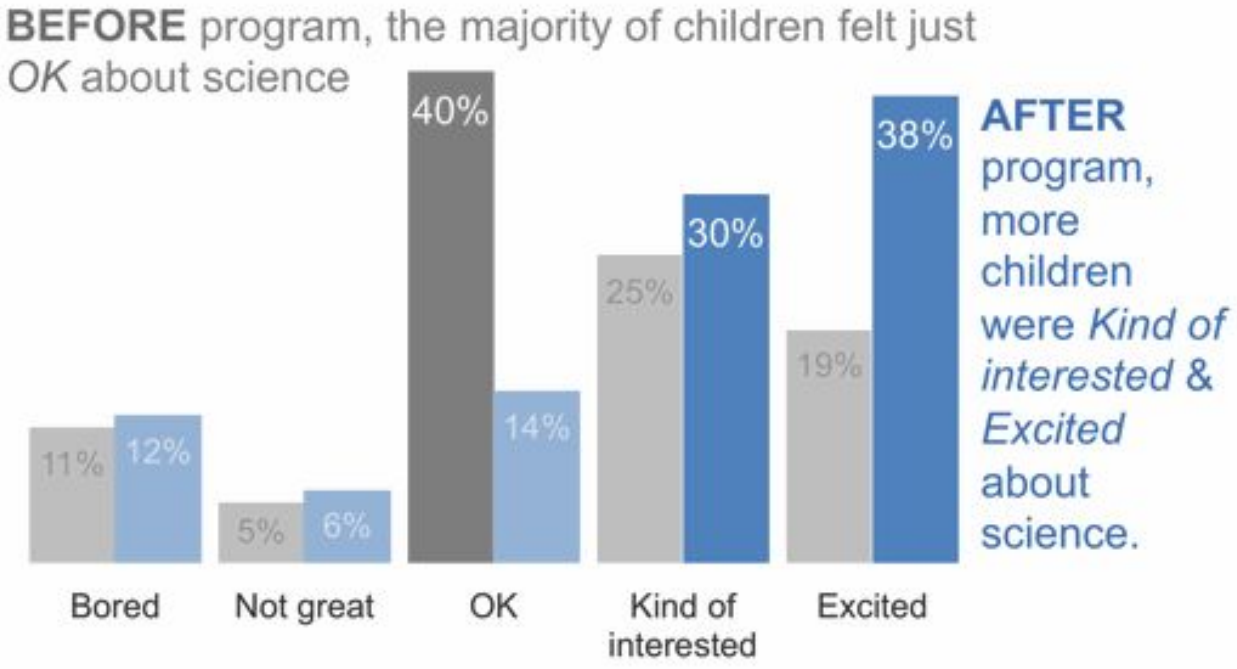


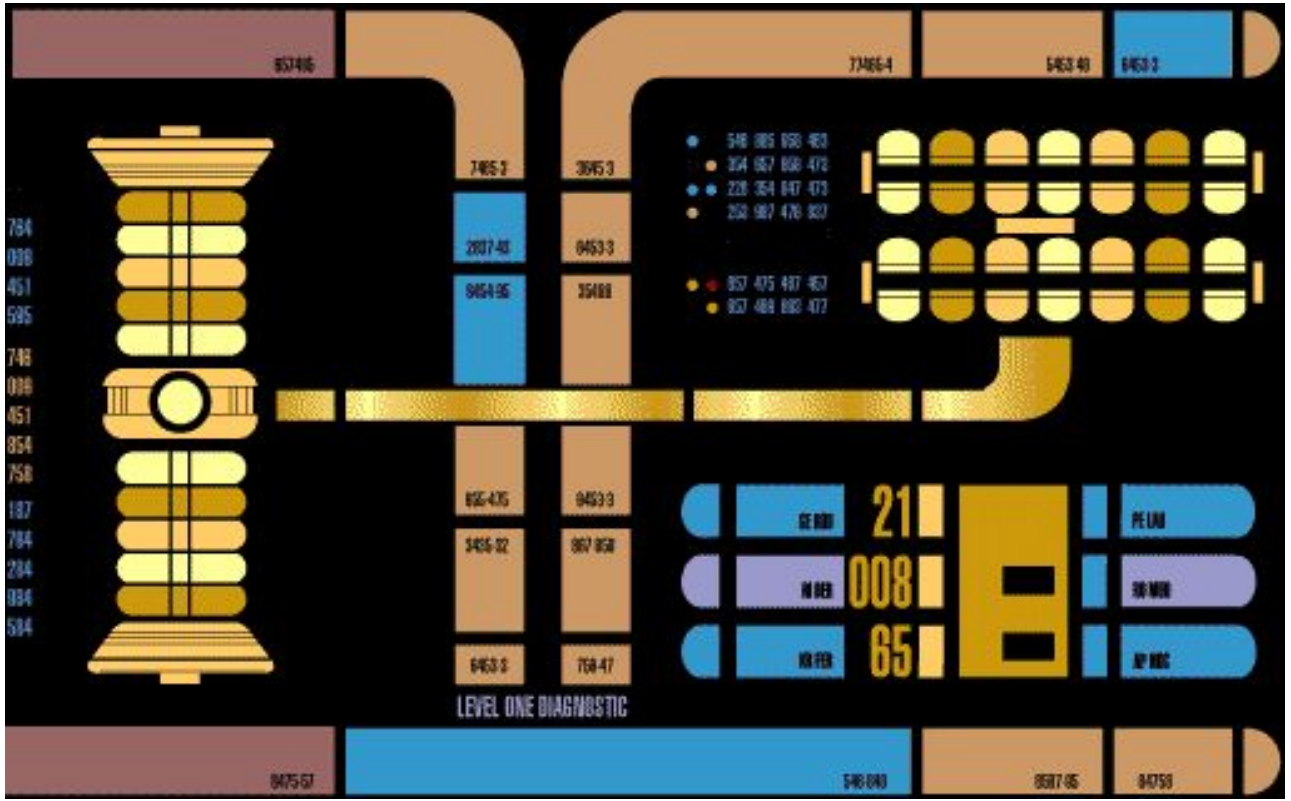
Comparing data





How do you feel about science?

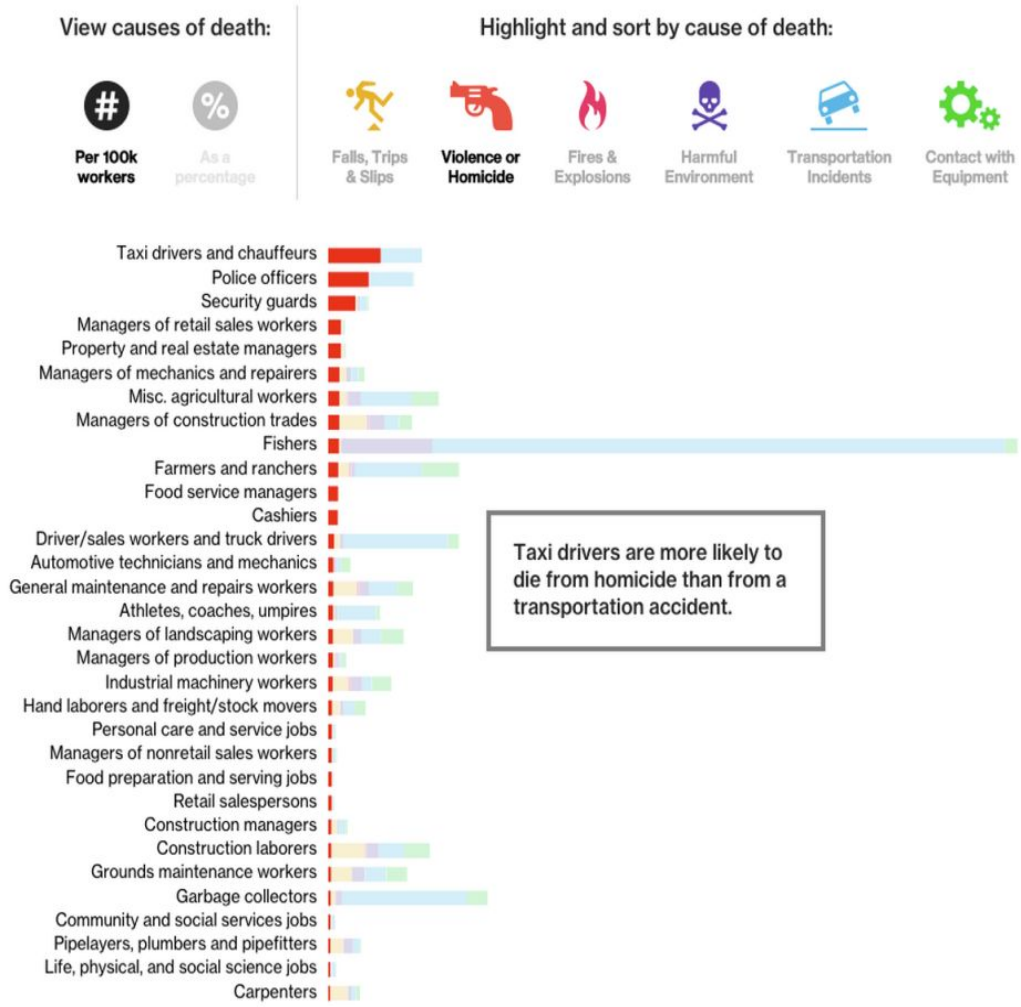




https://blog.libero.it/startrekmaniab/

Dashboards

1. Typical tasks
2. Information needed
3. Be boring
4. Iterative process



Data journalism

You are not one?

Tools

- Jupyter notebooks: Interactive computing https://nbviewer.jupyter.org/
- Python: matplotlib, seaborn, pandas
- Plotly https://plot.ly/create/?fid=chelsea_lyn:8648#/
- Javascript: plotly, charts.js, echarts …
- R: Rstudio, ggplot
- Drupal:
    - https://www.drupal.org/project/charts
    - https://www.drupal.org/project/plotly_js

Links

- https://en.wikipedia.org/wiki/John_Tukey
- https://www.kaggle.com/
- https://yahwes.github.io/
- https://www.edwardtufte.com/
- https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0000bx
- http://www.storytellingwithdata.com/
- https://www.kaggle.com/goldenoakresearch/us-acs-mortgage-equity-loans-rent-statistics/version/10
- https://www.inf.ethz.ch/personal/markusp/teaching/guides/guide-tables.pdf

Thank you / Questions?