McMaster University

Family Medicine

David Braley Primary Care Research Collaborative

@McMasterFamMed

Data Visualization

Watch a recording of the session here

Subhanya Sivajothy

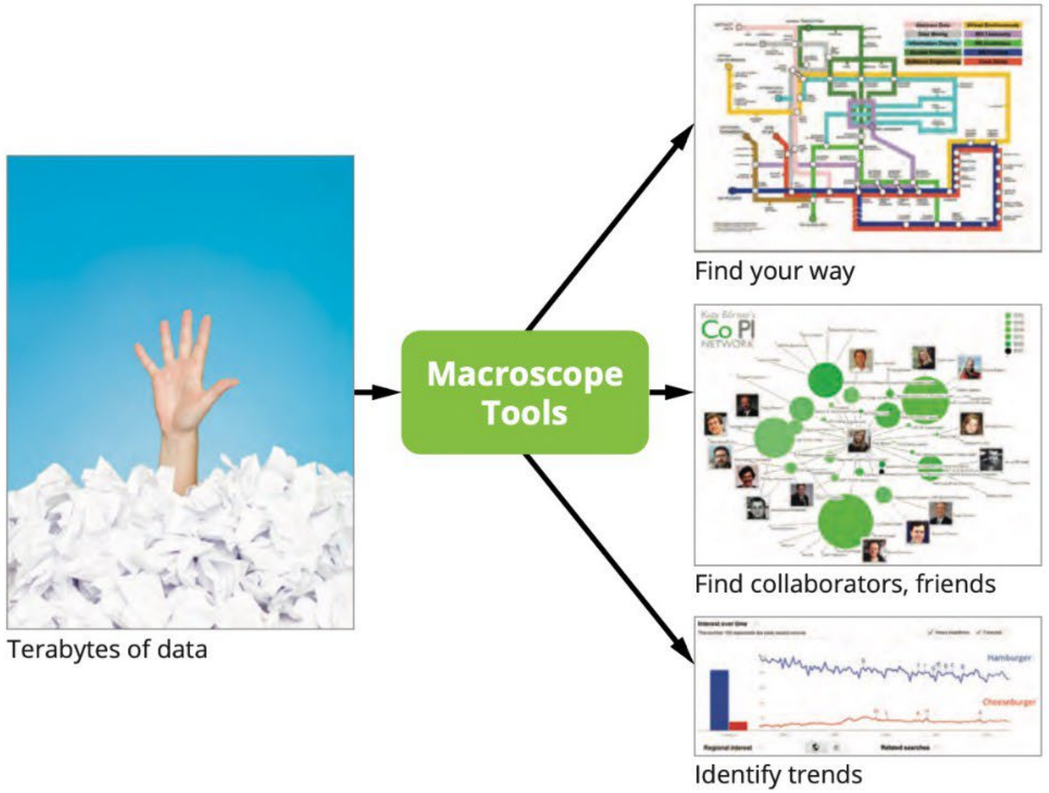
September 20th, 2022

Michael G. DeGroote SCHOOL OF MEDICINE

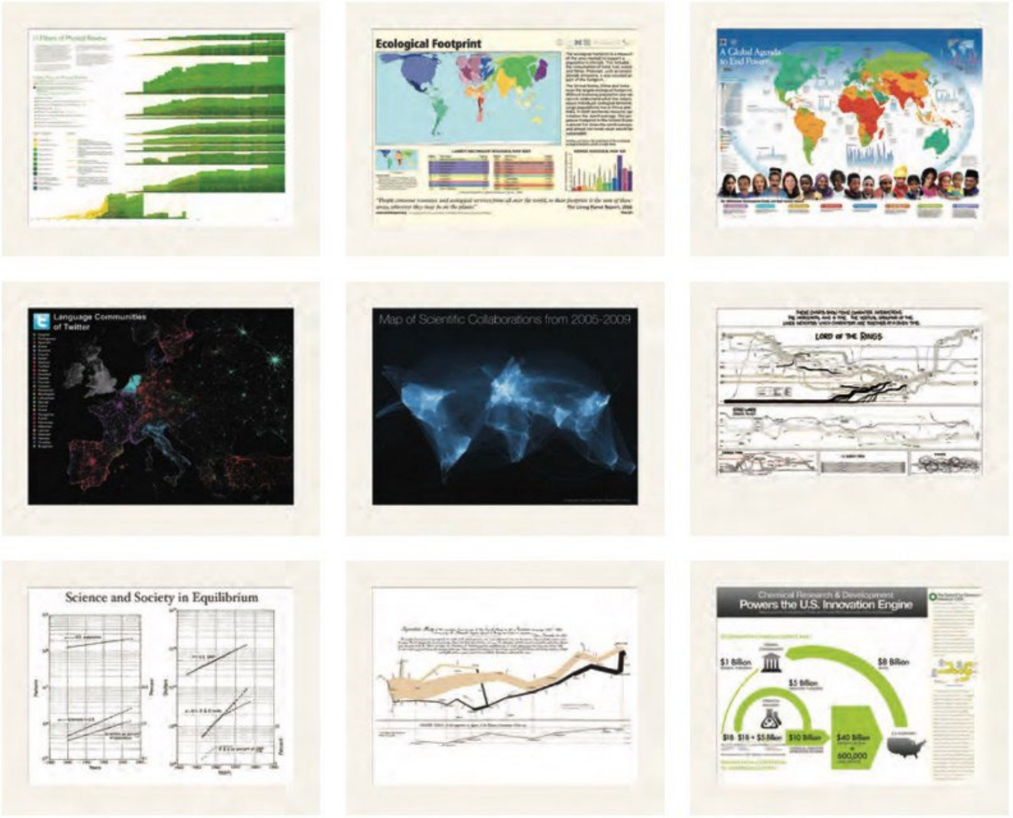
Data Visualization in Research

By Subhanya Sivajothy

What is Data Visualization?

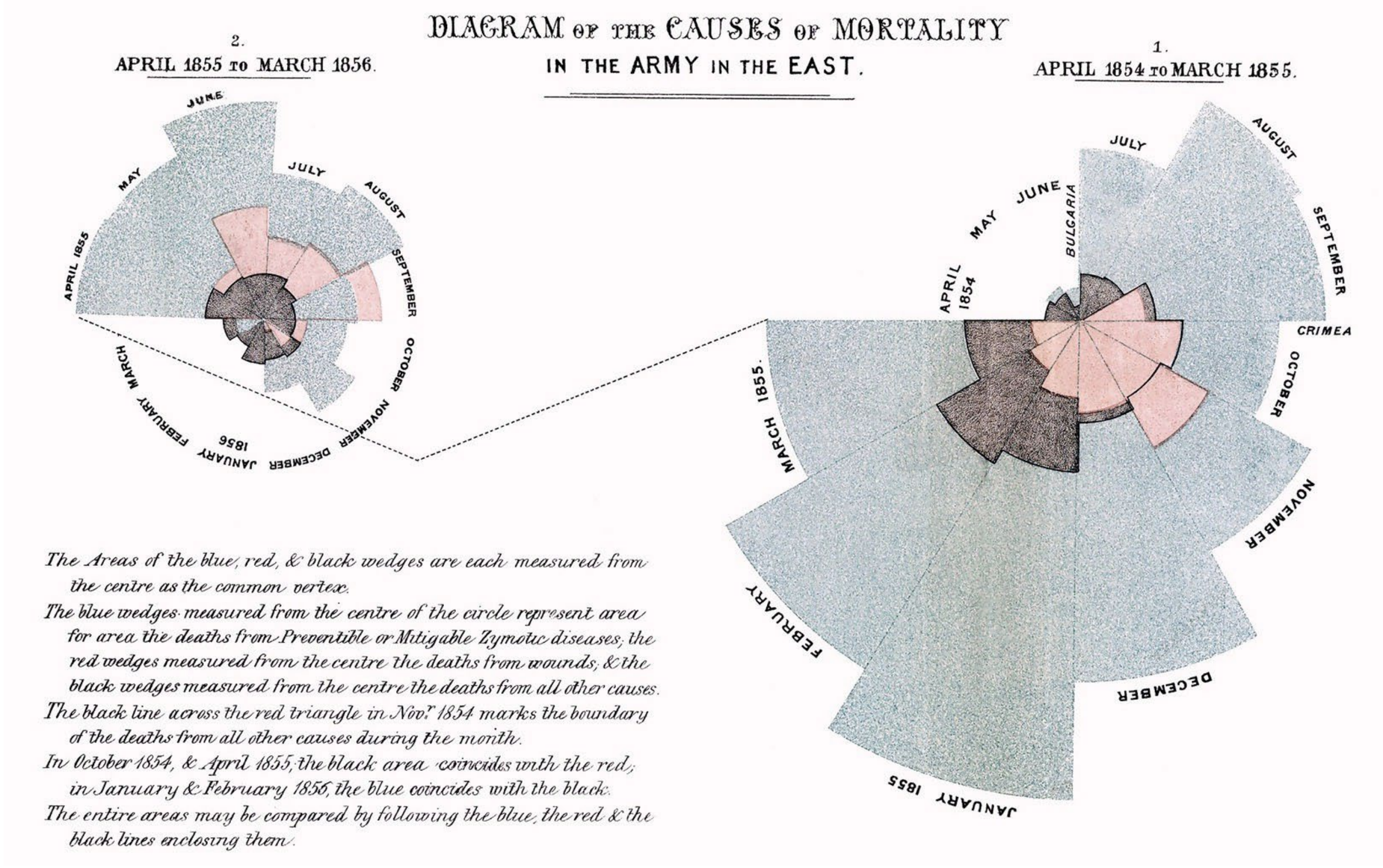


Why Use Data Visualizations?



It's all about you!

Dr. John Snow's Cholera Map in 1854

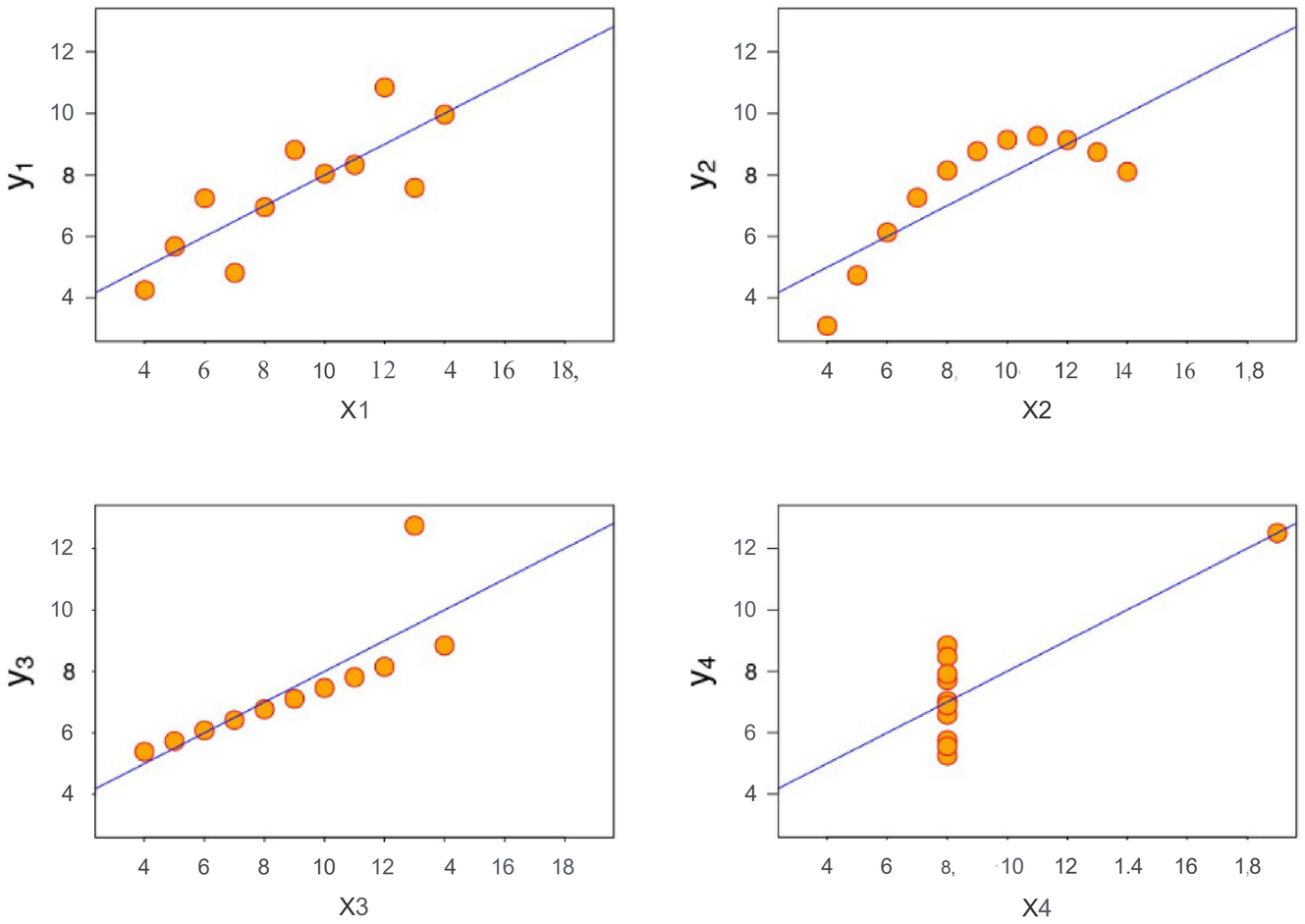


Florence Nightingale's Rose Diagram

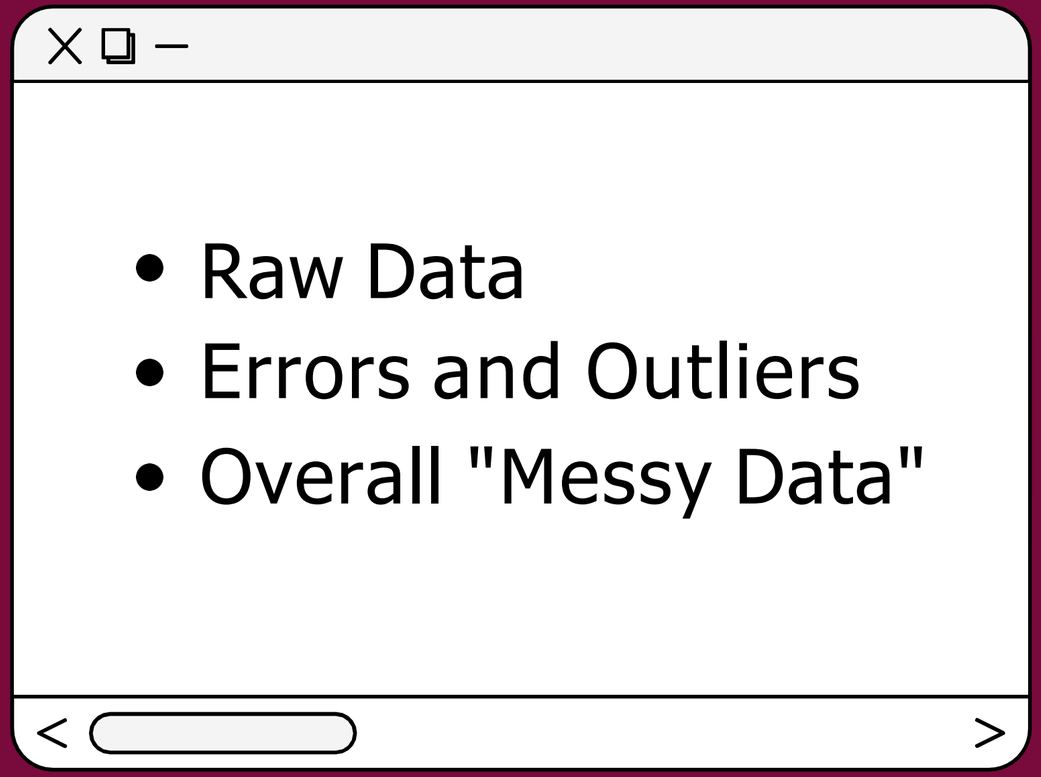
Exploratory Data Analysis

Anscombe's Quartet

Anscombe's Data										
Observation	x1	y1		x2	y2		x3	y3		x4	y4
1	10	8.04		10	9.14		10	7.46		8	6.58
2	8	6.95		8	8.14		8	6.77		8	5.76
3	13	7.58		13	8.74		13	12.74		8	7.71
4	9	8.81		9	8.77		9	7.11		8	8.84
5	11	8.33		11	9.26		11	7.81		8	8.47
6	14	9.96		14	8.1		14	8.84		8	7.04
7	6	7.24		6	6.13		6	6.08		8	5.25
8	4	4.26		4	3.1		4	5.39		19	12.5
9	12	10.84		12	9.13		12	8.15		8	5.56
10	7	4.82		7	7.26		7	6.42		8	7.91
11	5	5.68		5	4.74		5	5.73		8	6.89
				Summary Statistics							
N	11	11		11	11		11	11		11	11
mean	9.00	7.50		9.00	7.500909		9.00	7.50		9.00	7.50
SD	3.16	1.94		3.16	1.94		3.16	1.94		3.16	1.94
r	0.82			0.82			0.82			0.82	


y1
12
10
8
6
4
4 6 8 10 12 4 16 18,
X1
y2
12
10
8
6
4
4 6 8, 10 12 14 16 1,8
X2
y3
12
10
8
6
4
4 6 8 10 12 4 16 18
X3
y4
12
10
8
6
4
4 6 8, 10 12 1.4 16 1,8
X4

Exploratory Data Analysis: Stage One



Exploratory Data Analysis: Stage Two

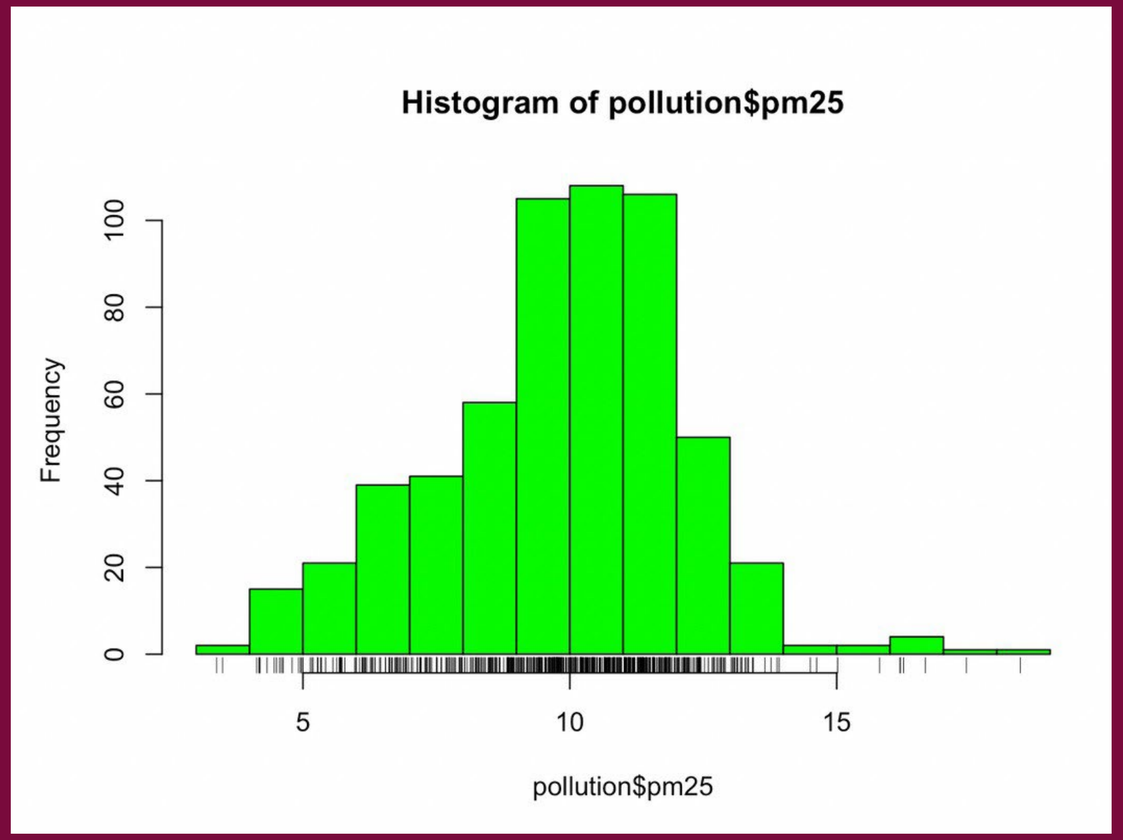
1. Generate questions about your data
2. Search for answers by visualising, transforming, and modeling your data
3. Use what you learn to refine your questions and or generate new questions
4. Repeat process until you have a graph you'd like to publish

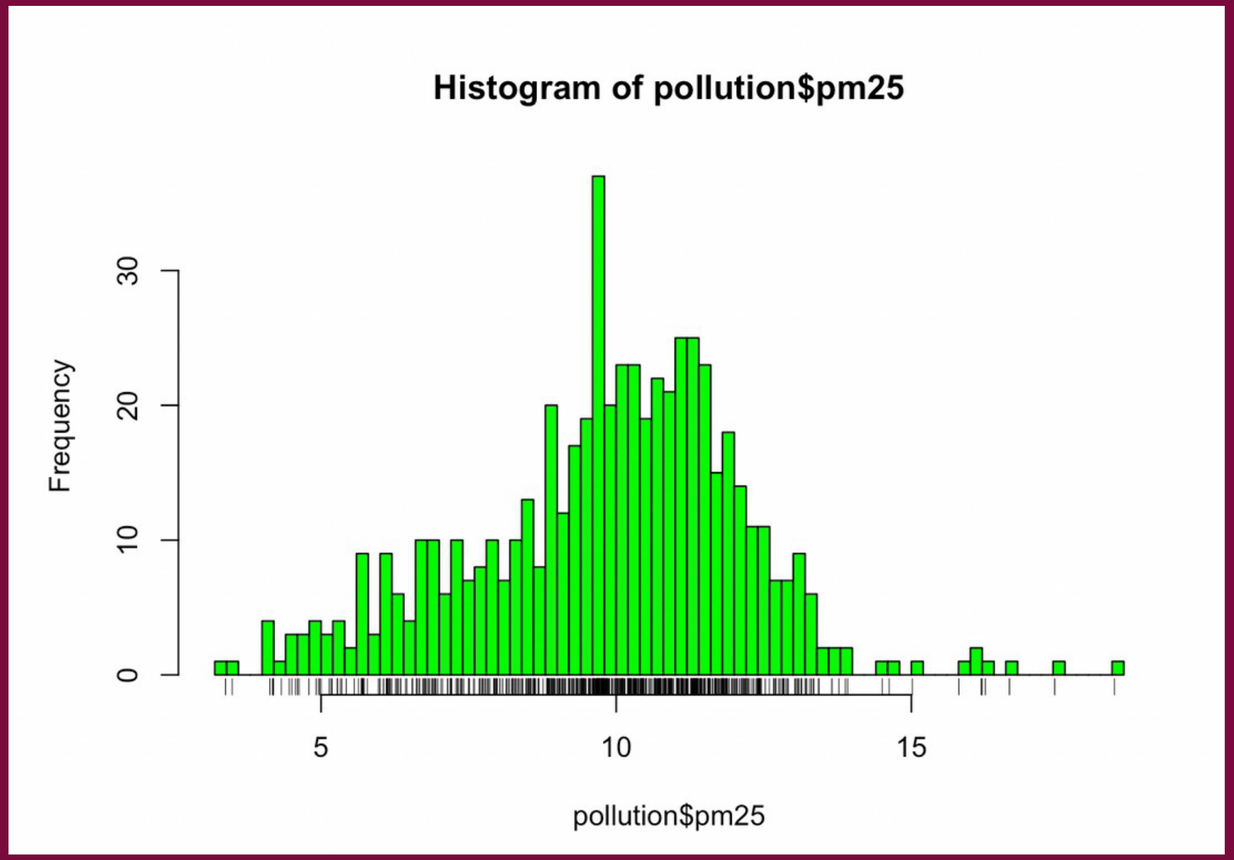
adapted from Hadley Wickham
https://cfss.uchicago.edu/notes/exploratory-data-analysis/

Exploratory Data Analysis: Stage Two

1. What type of variation occurs **within** my variables?
2. What type of covariation occurs **between** my variables?

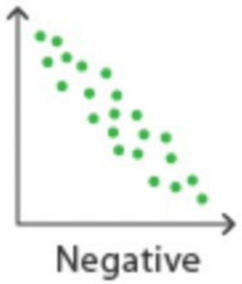
adapted from Hadley Wickham
https://cfss.uchicago.edu/notes/exploratory-data-analysis/



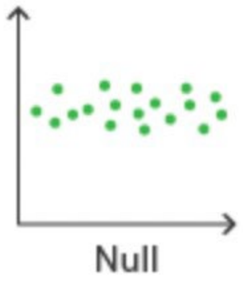


https://bookdown.org/rdpeng/exdata/exploratory-graphs.html

Positive

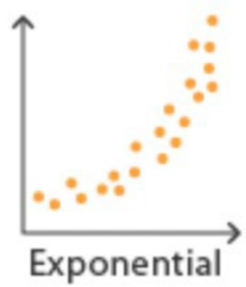


Negative



Null

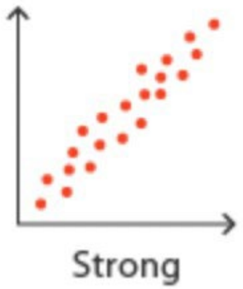
Linear



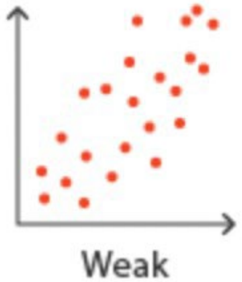
Exponential

U-Shaped

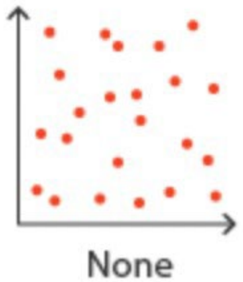
Correlation Strength:



Strong

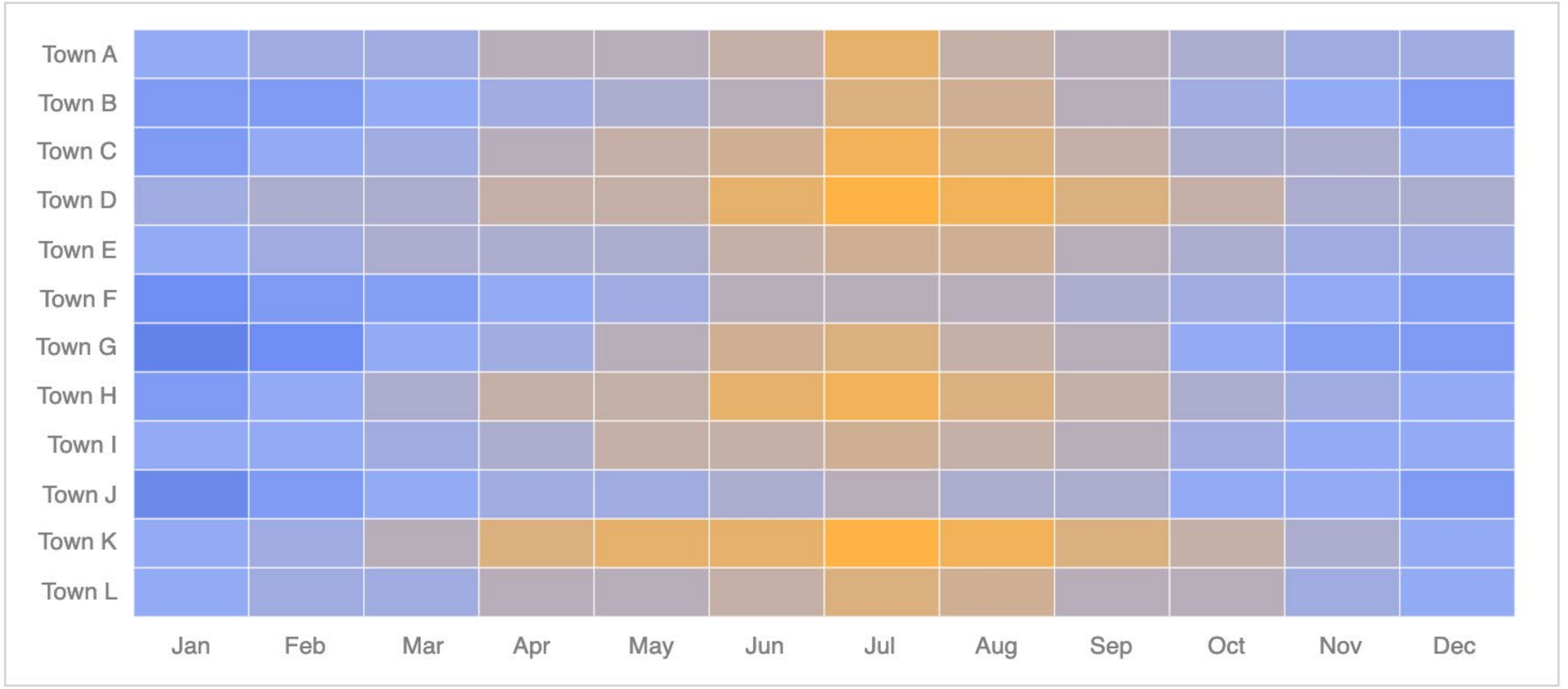


Weak



None

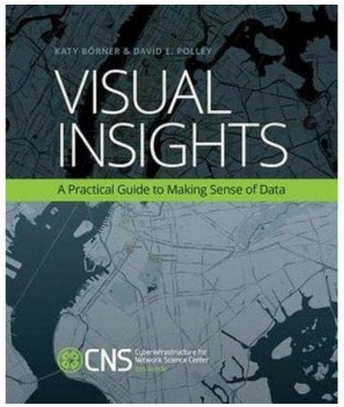
Heatmap (Matrix)



Data Visualization for Communication

Workflow

1. Identify Audience and Purpose
2. Analysis: Explore and Prepare Data
3. Visualization Idioms
4. Select Visualization Elements
5. Share for Interpretation and Critique

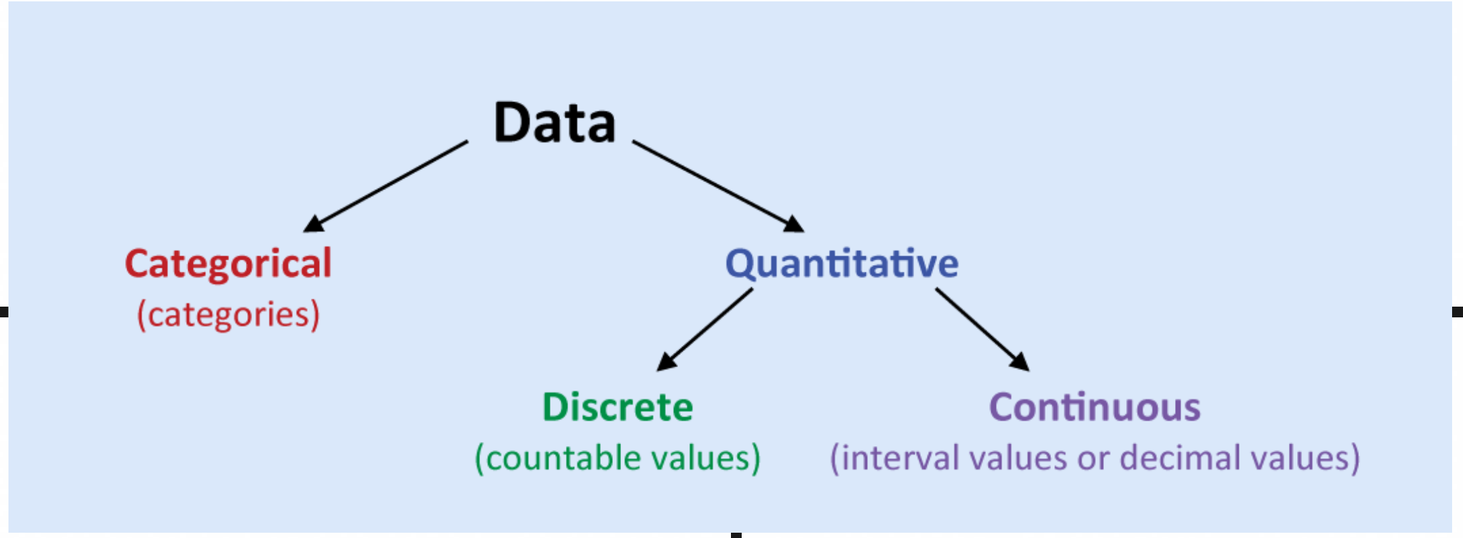


Katy Börner, David E. Polley
+ Kelly Schulz

1. Identify Audience and Purpose

01	Who is your audience for your visualization?
02	What level of familiarity do they have with your topic?
03	What is the purpose of your visualization?
04	Is it to communicate a finding, or is it exploratory for your own analysis?
05	What is the story that I'm trying to tell?

2. Analysis: Explore and Prepare Data



Categorical

Categorical variables contain a finite number of categories or distinct groups. Categorical data might not have a logical order. Qualitative data is often categorical.

Continuous

Continuous variables are numeric variables that have an infinite number of values between any two values. A continuous variable can be numeric or date/time. Continuous data is always quantitative.

Discrete

Discrete variables are numeric variables that have a countable number of values between any two values. A discrete variable is always numeric.

2. Analysis: Explore and Prepare Data

Common Tasks

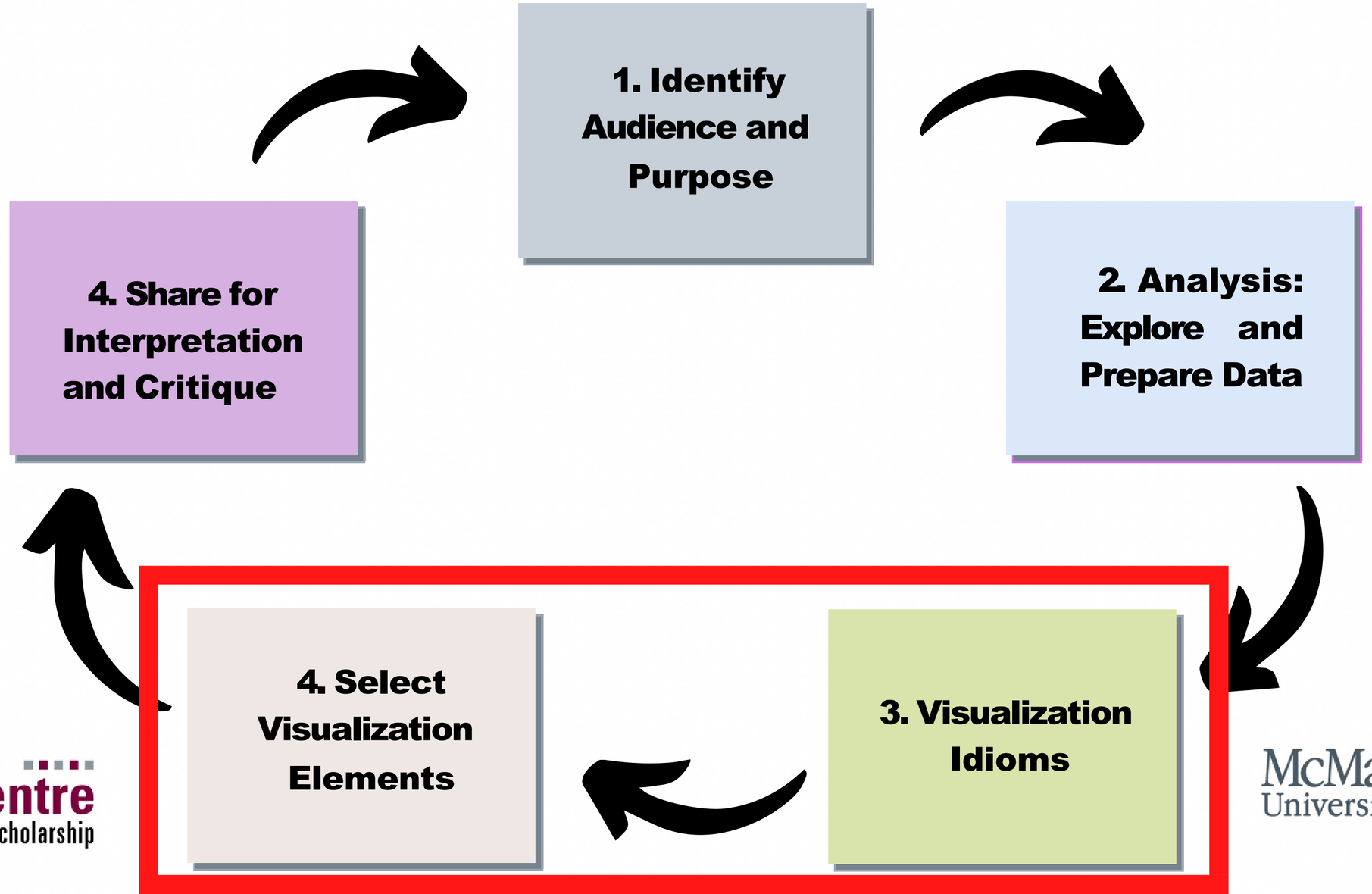
formatting values

anomalies and missing data

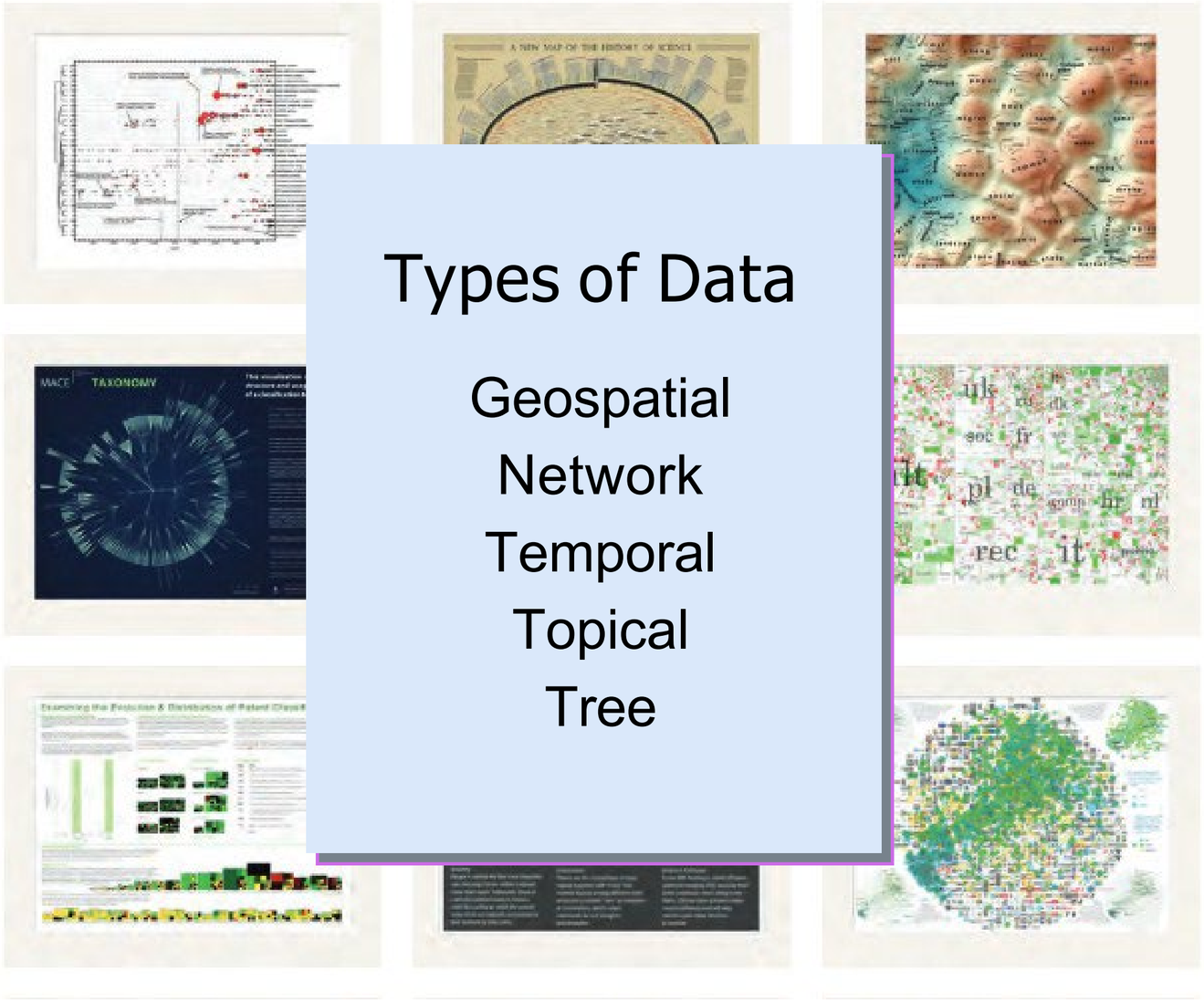
standardizing values and
remove pre-aggregated data

readable headings

Workflow


1. Identify Audience and Purpose
2. Analysis: Explore and Prepare Data
3. Visualization Idioms
4. Select Visualization Elements
4. Share for Interpretation and Critique

3. Visualization Idioms



Types of Data

Geospatial
Network
Temporal
Topical
Tree

Choosing Idioms

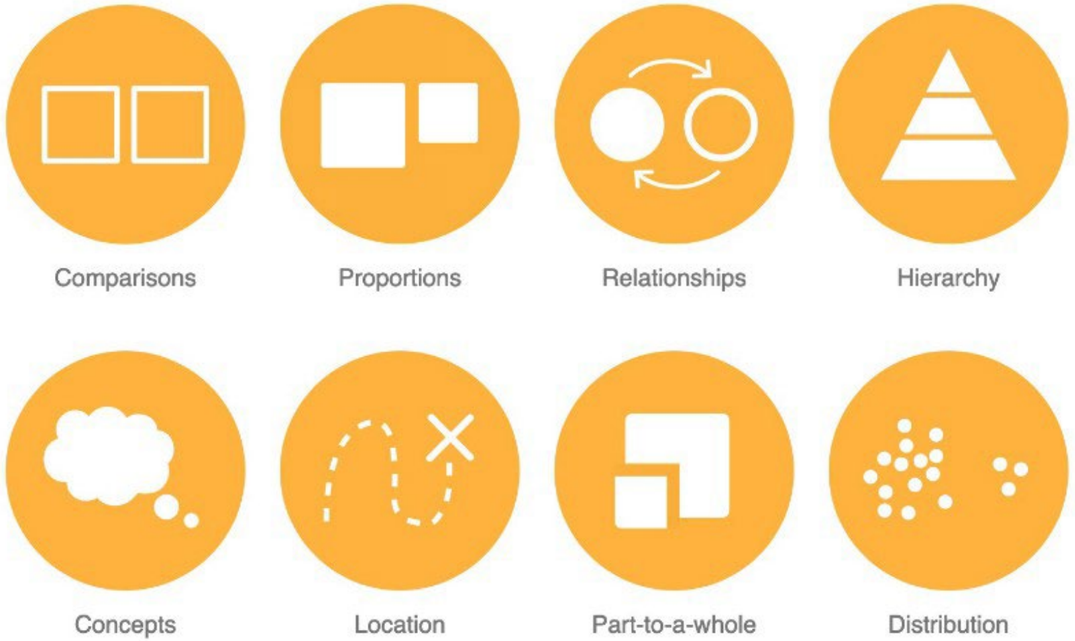
01	Geospatial	Bubble Map, Choropleth Map
02	Temporal	Timeline, Line Graph, Area Chart, Histogram, Bubble Chart
03	Network	Arc Diagram, Chord Diagram, Network Diagram
04	Topical	Wordclouds, Bar Graph, Tree Maps
05	Tree	Sunburst diagram, Tree Map, Flowchart

3. Visualization Idioms

The Data Visualisation Catalogue

What do you want to show?

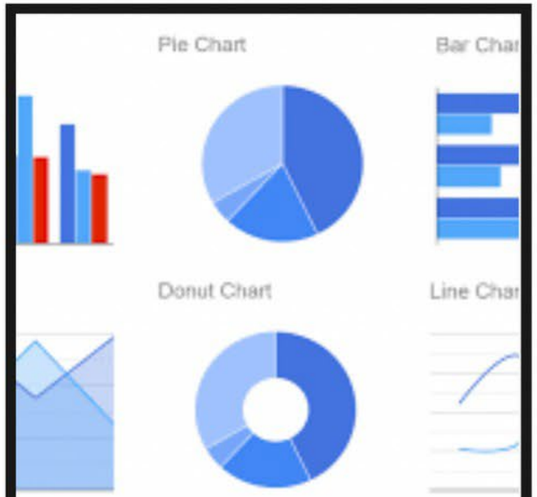
Here you can find a list of charts categorised by their data visualization functions or by what you want a chart to communicate to an audience. While the allocation of each chart into specific functions isn't a perfect system, it still works as a useful guide for selecting chart based on your analysis or communication needs.



Tools for Data Visualization

MS Excel

Tableau



Google Charts

D3.js

4. Select Visual Elements

Marks

- basic graphical element in an image

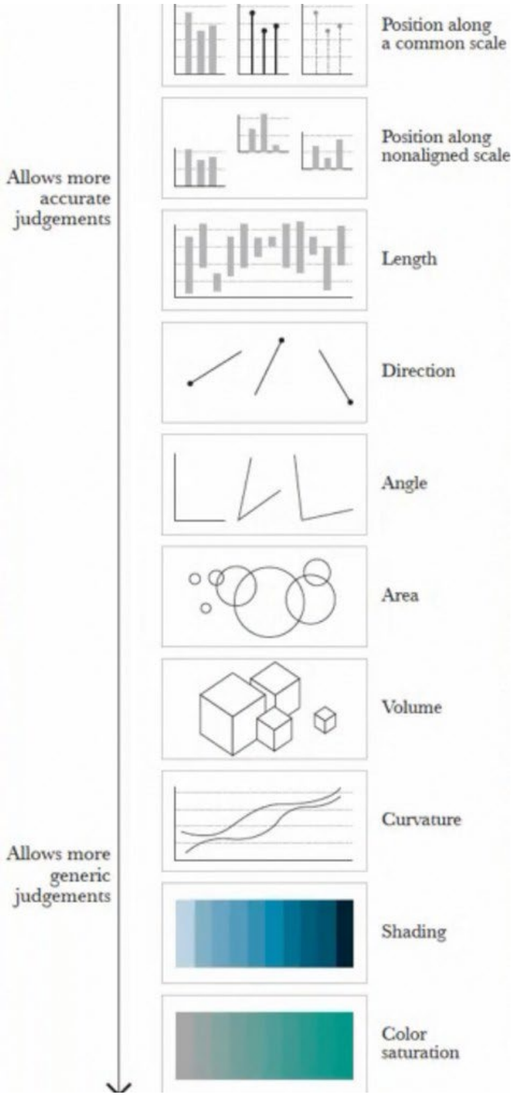
i.e the points, bars, lines, areas

Channels

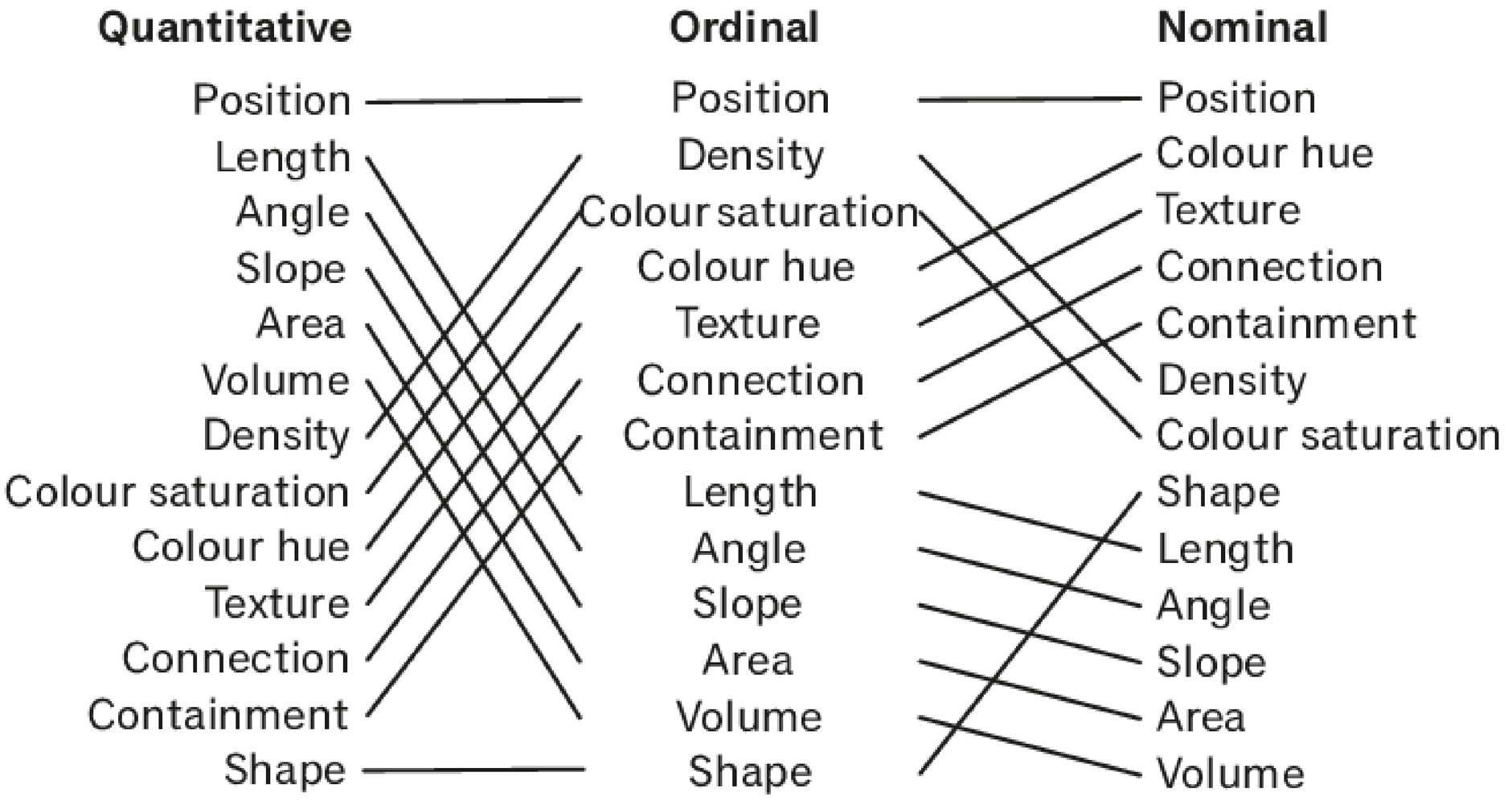
- The attributes of a mark.

i.e position, shape, size, or color.

4. Select Visual Elements



Perception of graphical elements (Cleveland & McGill, 1984, P532)



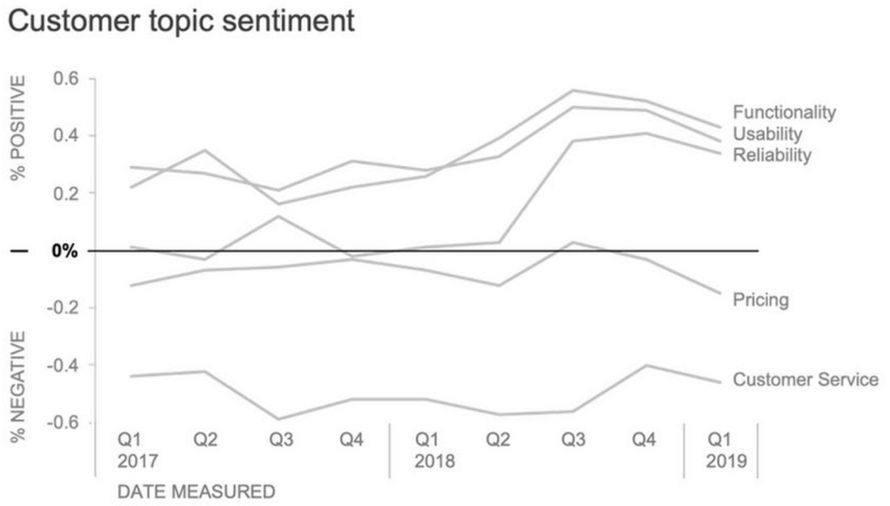
The Mackinlay ranking of perceptual task

5. Share for Interpretation and Receive Feedback

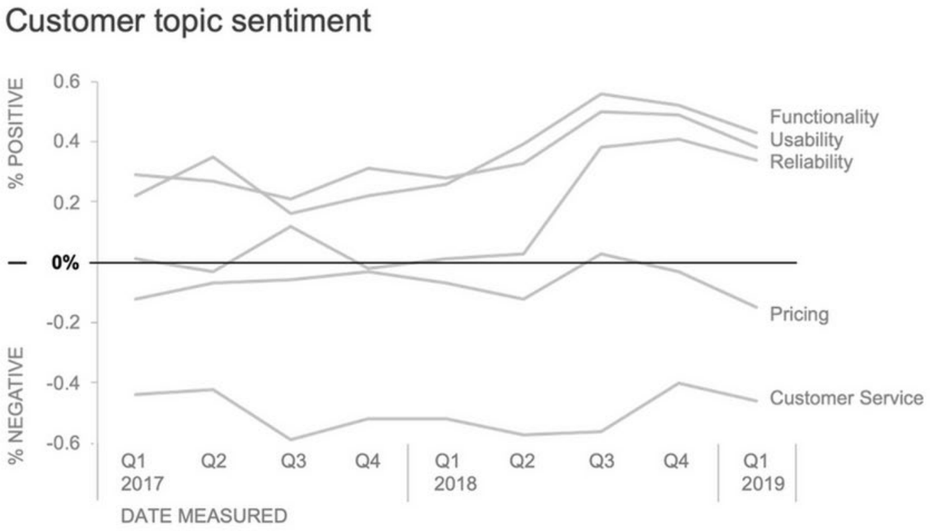
01	Would a user be able to understand the basics in 15 seconds?
02	Is this visualization honest about what isn't represented?
03	Have I properly attributed the work?

Ethics & Accessibility in Visualization, and Critical Design Practices

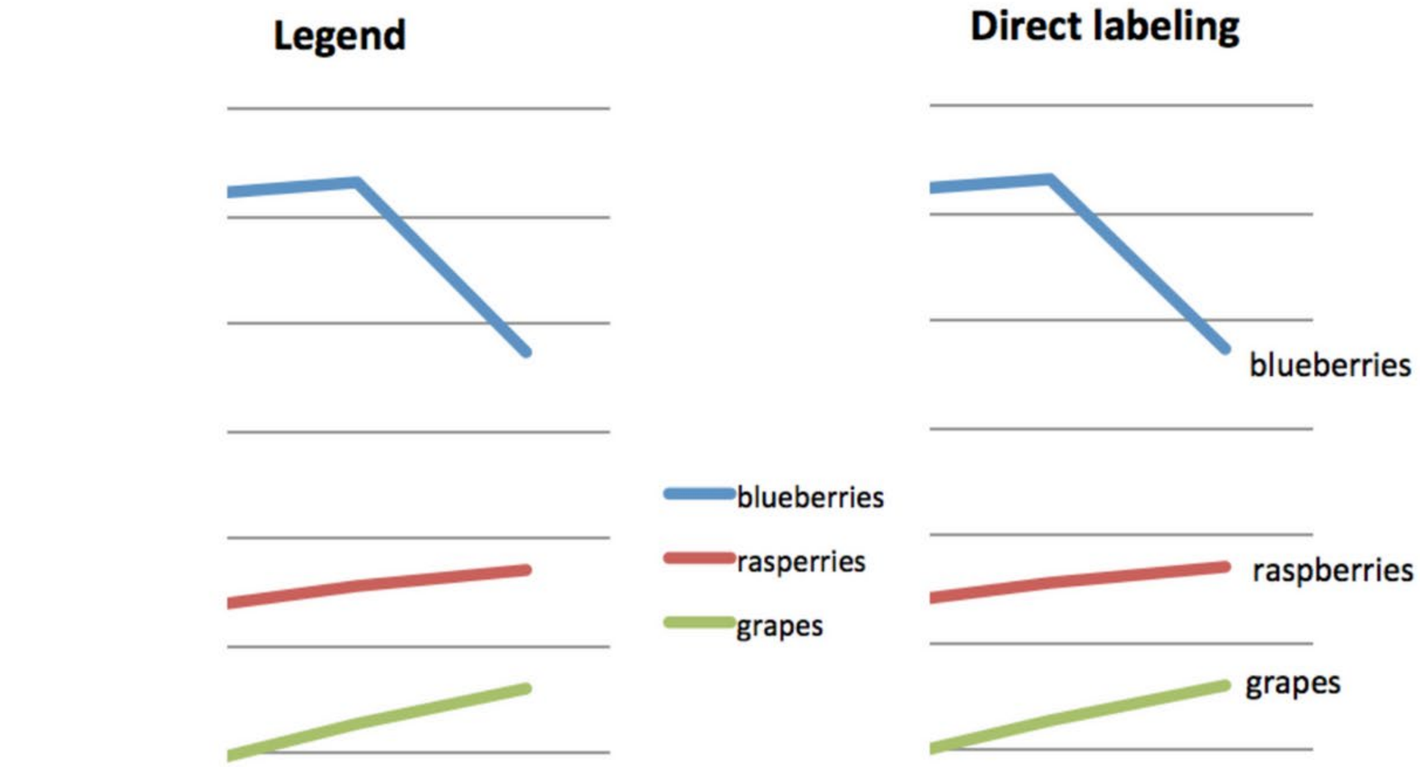
Action needed to address **recent decline**



Success: efforts to increase **reliability** worked!



https://www.storytellingwithdata.com/blog/2018/6/26/accessible-data-viz-is-better-data-viz



An example graph using legend vs. direct labeling

https://www.storytellingwithdata.com/blog/2018/6/26/accessible-data-viz-is-better-data-viz

Check type and colour contrast

Small Non-Bold Text (less than 18pt, or approximately 1.5em rendered) for FFFFFF

Color Code	Sample Text	Sample Text (inverted)	Pass or Fail	Ratio (pass>=4.5)
0072CE	Lorem ipsum	Lorem ipsum	PASS	4.89
4497DC	Lorem ipsum	Lorem ipsum	FAIL	3.13

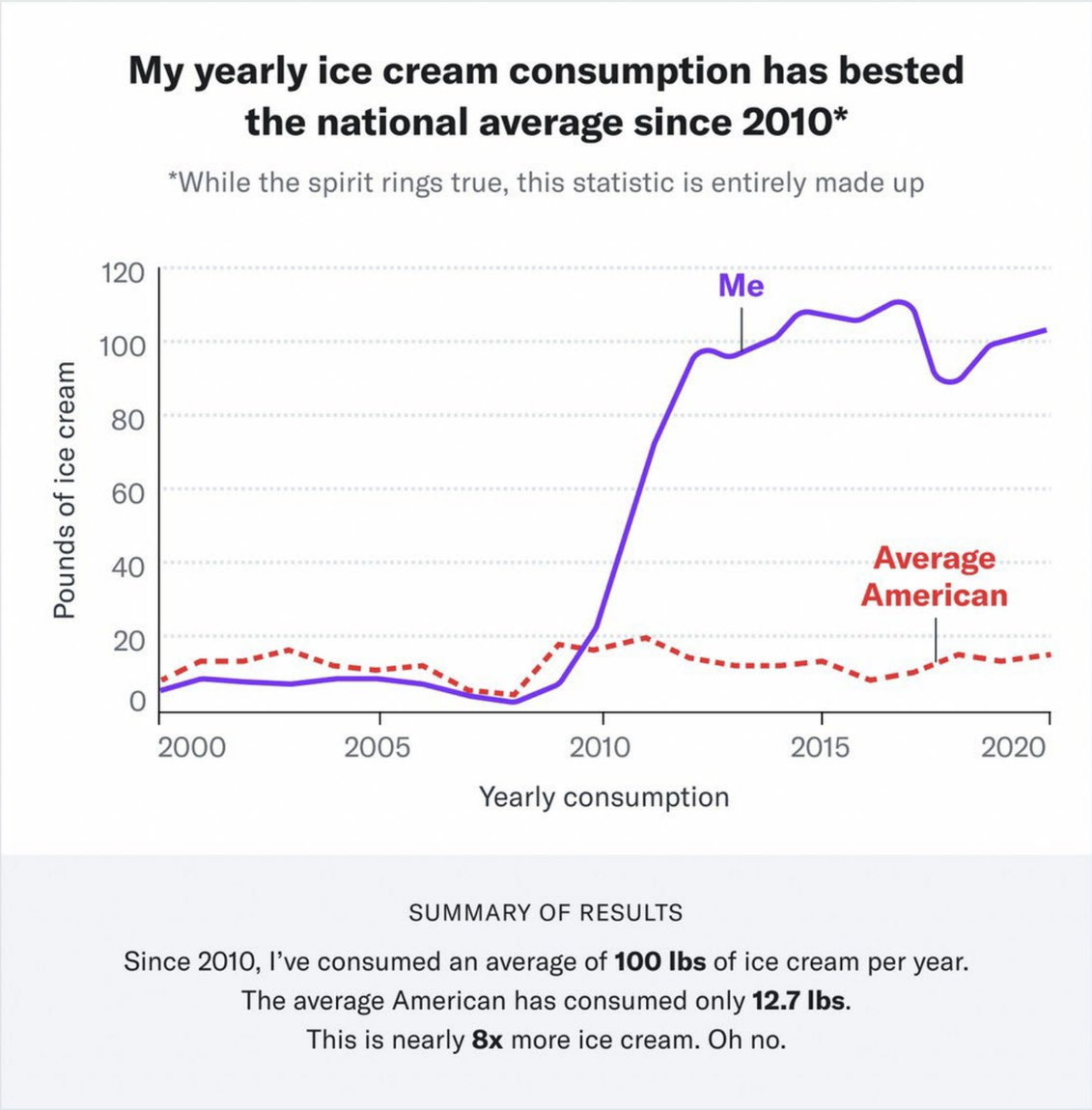
https://www.storytellingwithdata.com/blog/2018/6/26/accessible-data-viz-is-better-data-viz

Example of the for the color palette contrast evaluation tool WCAG standards

Adding Alt Text

```
<img src="pancakes.png" alt="Stack of blueberry pancakes with powdered sugar">
```

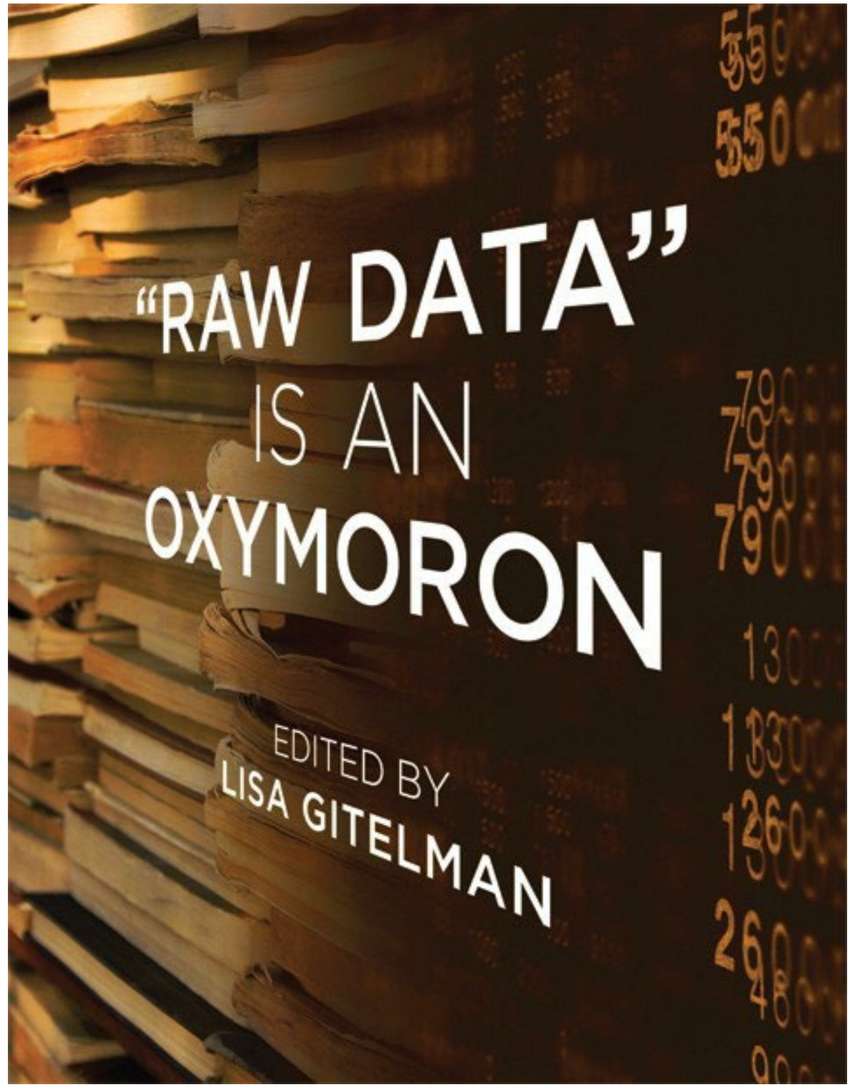
source: https://www.storytellingwithdata.com/blog/2018/6/26/accessible-data-viz-is-better-data-viz



https://www.betterment.com/design/accessible-data-visualization

'The world does not spontaneously quantify, curate, or data-mine itself. Rather, the process of observing the world and quantifying it is a political act, and deserves ethical consideration"

- Michael Correll



Michael Correll. 2018. "Ethical Dimensions of Visualization Research."
https://arxiv.org/pdf/1811.07271.pdf

Make the Invisible Visible

Visualize Hidden Labour

Visualize Hidden Uncertainty

Visualize Hidden Impacts

… Managing Complexity?

Michael Correll. 2018. "Ethical Dimensions of Visualization Research." https://arxiv.org/pdf/1811.07271.pdf

Collect Data With Empathy

Encourage 'Small Data'

Anthropomorphize Data

Anonymity By Design (and Right to Due Process)

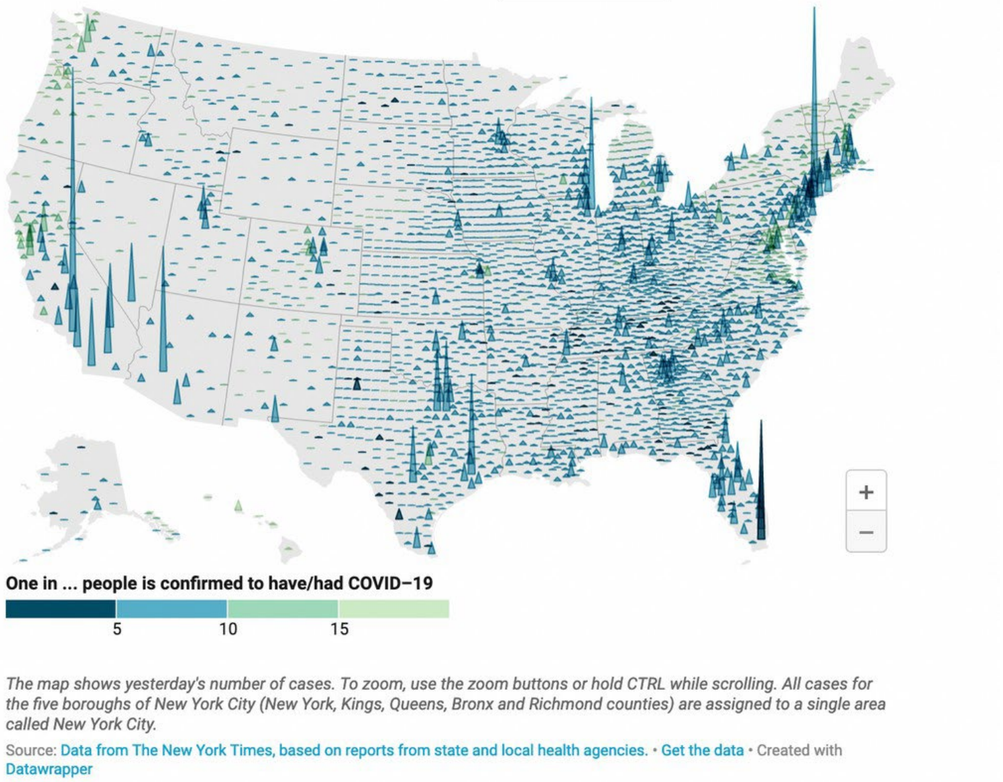
... Including Context?

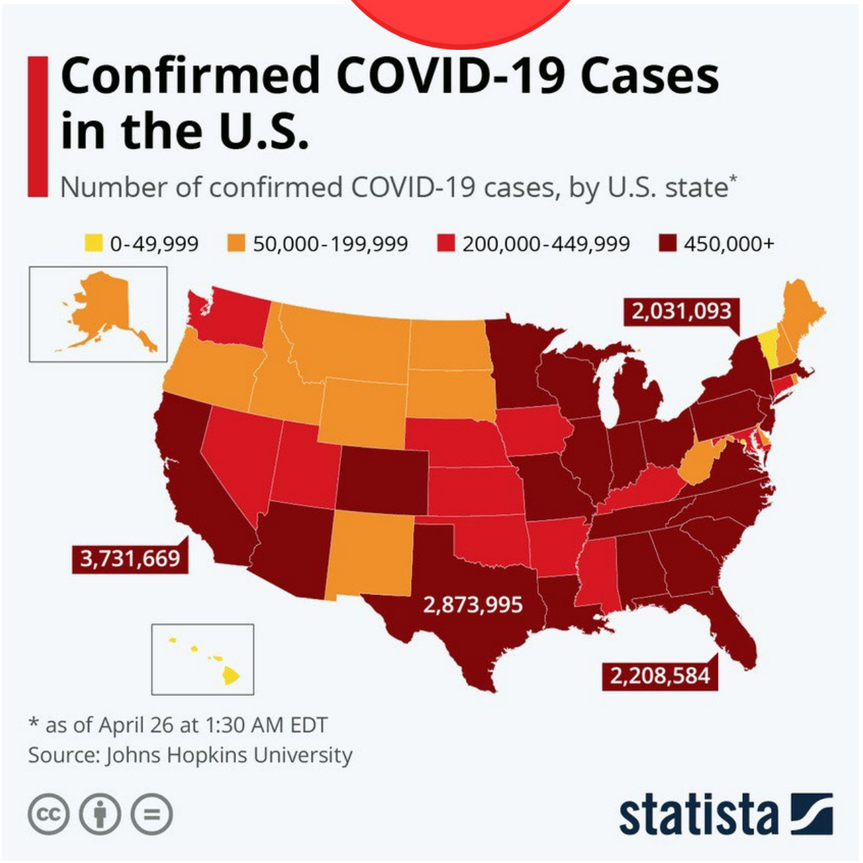
Michael Correll. 2018. "Ethical Dimensions of Visualization Research." https://arxiv.org/pdf/1811.07271.pdf

Logistcal Considerations

- Are you disclosing the limits of your data in a way that is clear? For example, what all influences how cases are reported?
- If you're visualizing information that is time-sensitive such as cases, are you willing to commit to daily updates and for how long?
- Are you disclosing all sources?
- How does storage affect updating and relationships?
- What inaccuracies exist?
- What transformations are you making in storage?
- What does NULL mean?
- What decisions have you made in storage? (ex: how locations are classified?)

https://www.tableaufit.com/the-ethics-of-visualizing-during-a-pandemic/





source: https://blog.datawrapper.de/coronaviruscharts/

Visualization Considerations

- How does the visualization limit what you can understand? (ex: A line chart shows trend, but not potential geographic patterns)
- What expertise is required to follow it? (ex: log scales)
- Who have you considered as your audience?
- Who have you (intentionally or unintentionally) ignored?
- What is your motivation for doing this?
- What values are you espousing in your visualization? Do they support or conflict with other values?
- Do you create overt alarm? Most people are already stressed.
- Does it overly abstract and calm too much? Some people aren't taking this seriously enough.

https://www.tableaufit.com/the-ethics-of-visualizing-during-a-pandemic/

Visualization Provenance

- Documenting what steps were involved in creating the visualization
- Helps you with tracing errors and reproducing steps
- Allows others to evaluating the quality of the data and the validity of the visualization output