

Intro to Data Visualizations

Subhanya Sivajothy

DASH Workshop Series

February 10th 2022





McMaster University sits on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the “Dish With One Spoon” wampum agreement.

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences.

In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Code of Conduct

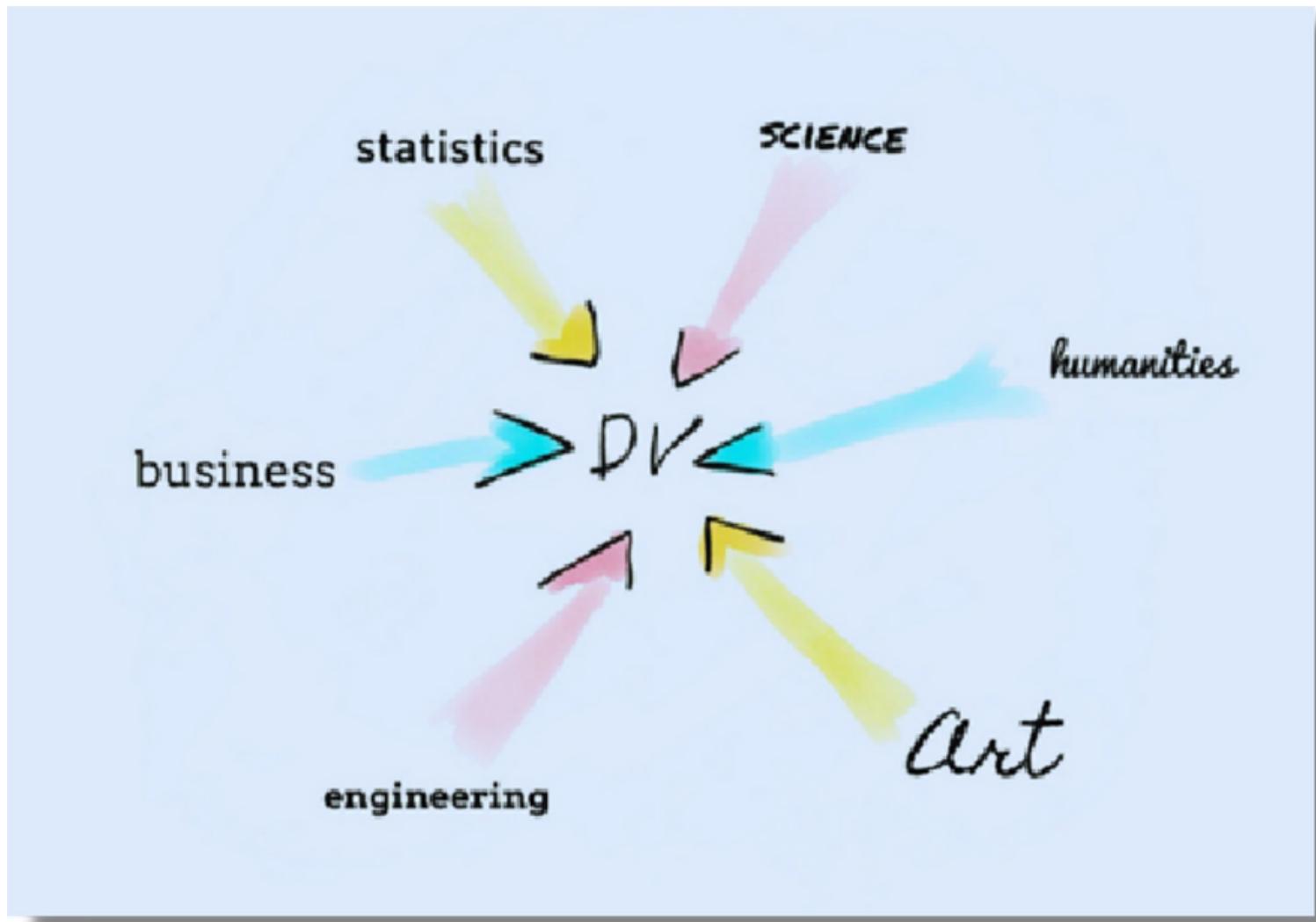
The Sherman Centre and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information:

scds.ca/events/code-of-conduct/

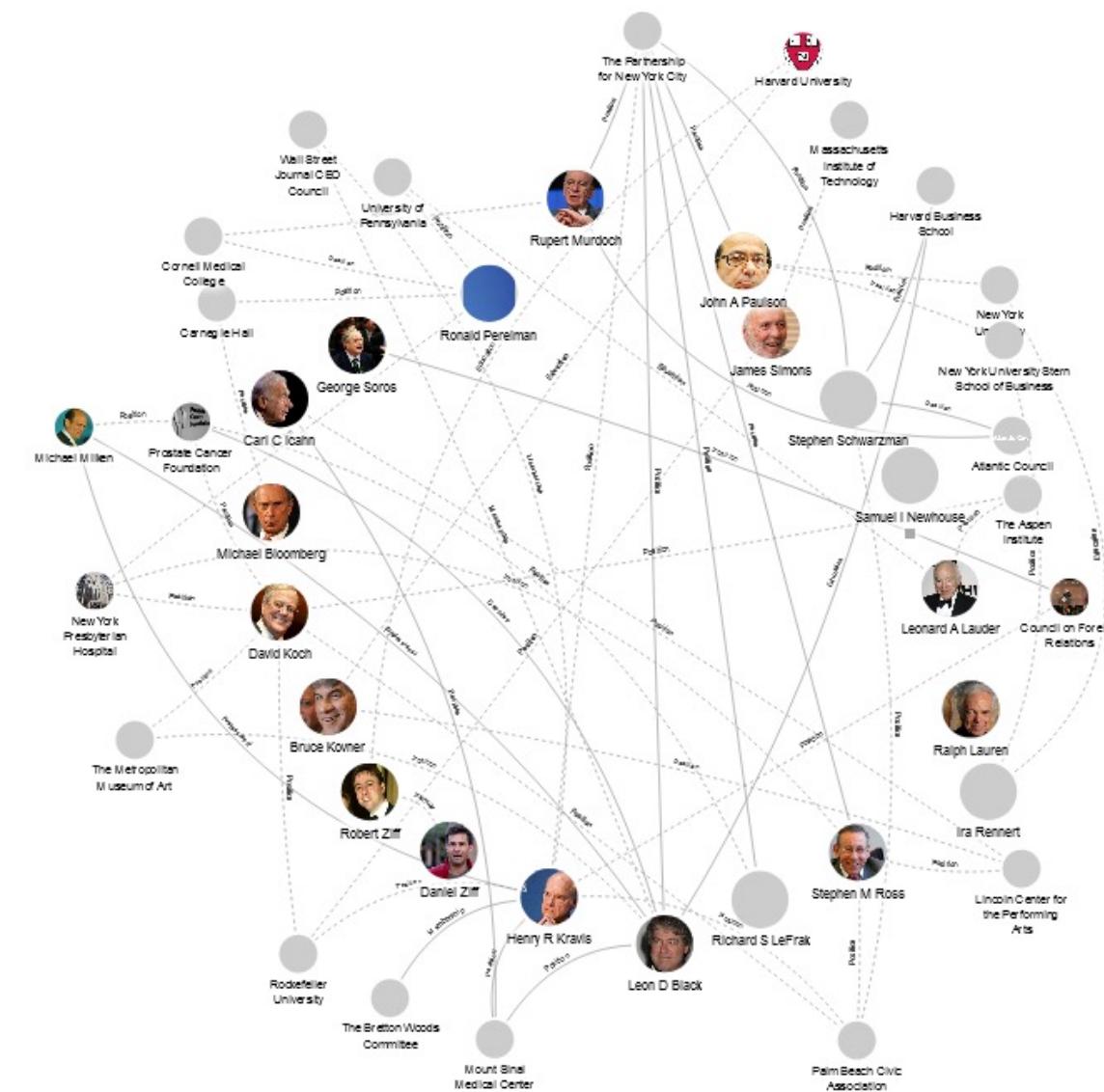
Introduction



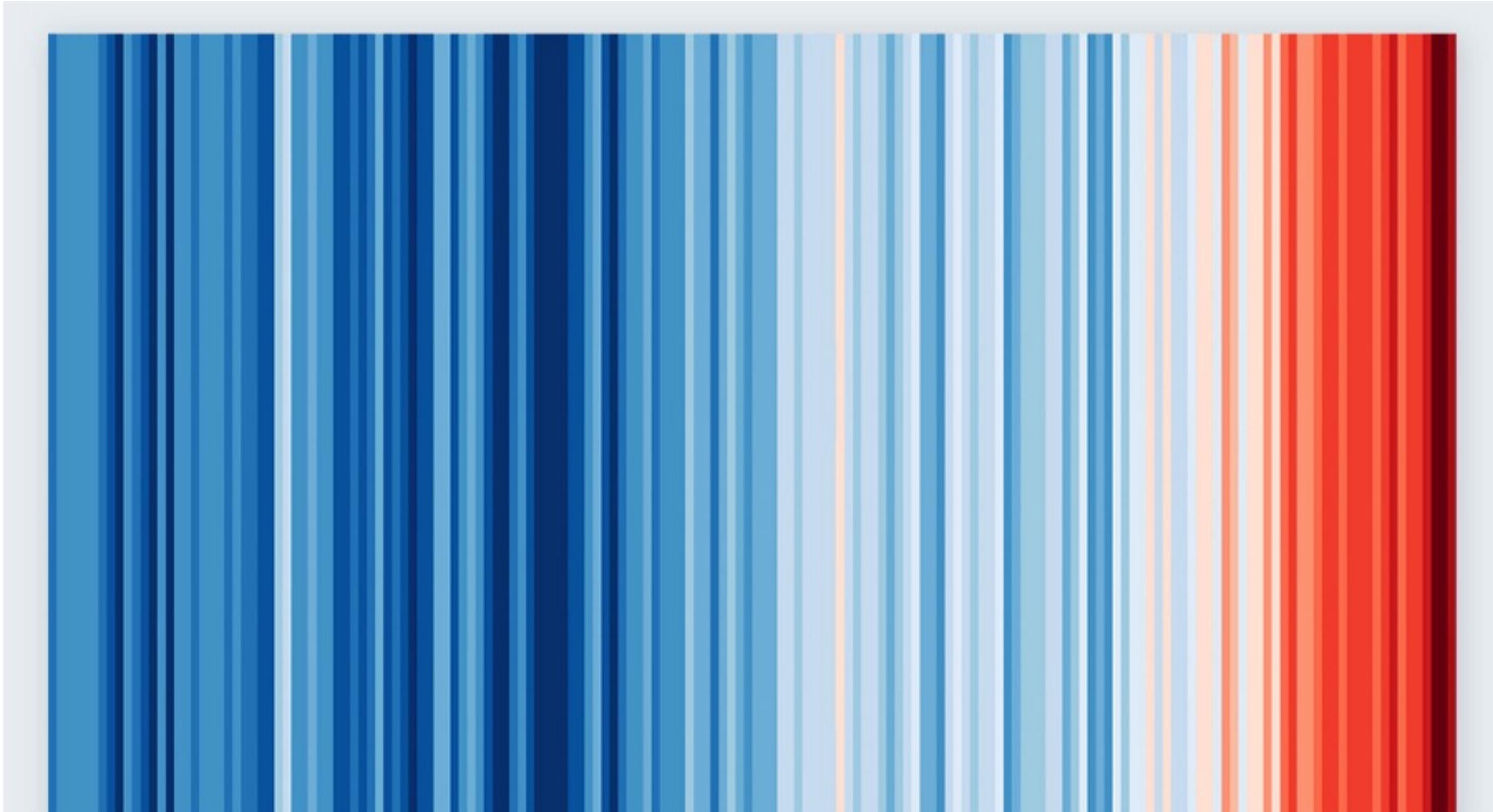
Poll #1



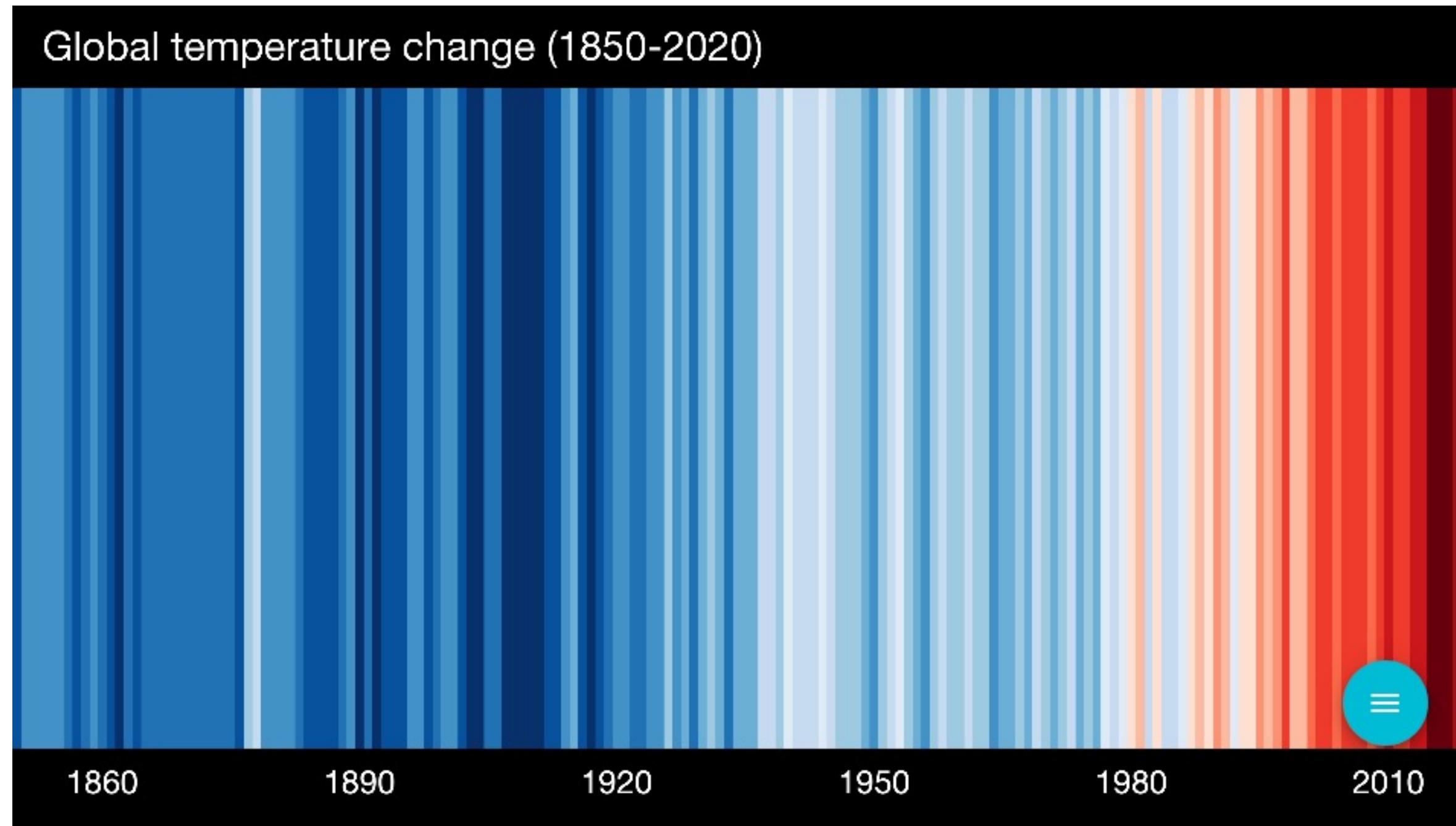
Poll #2



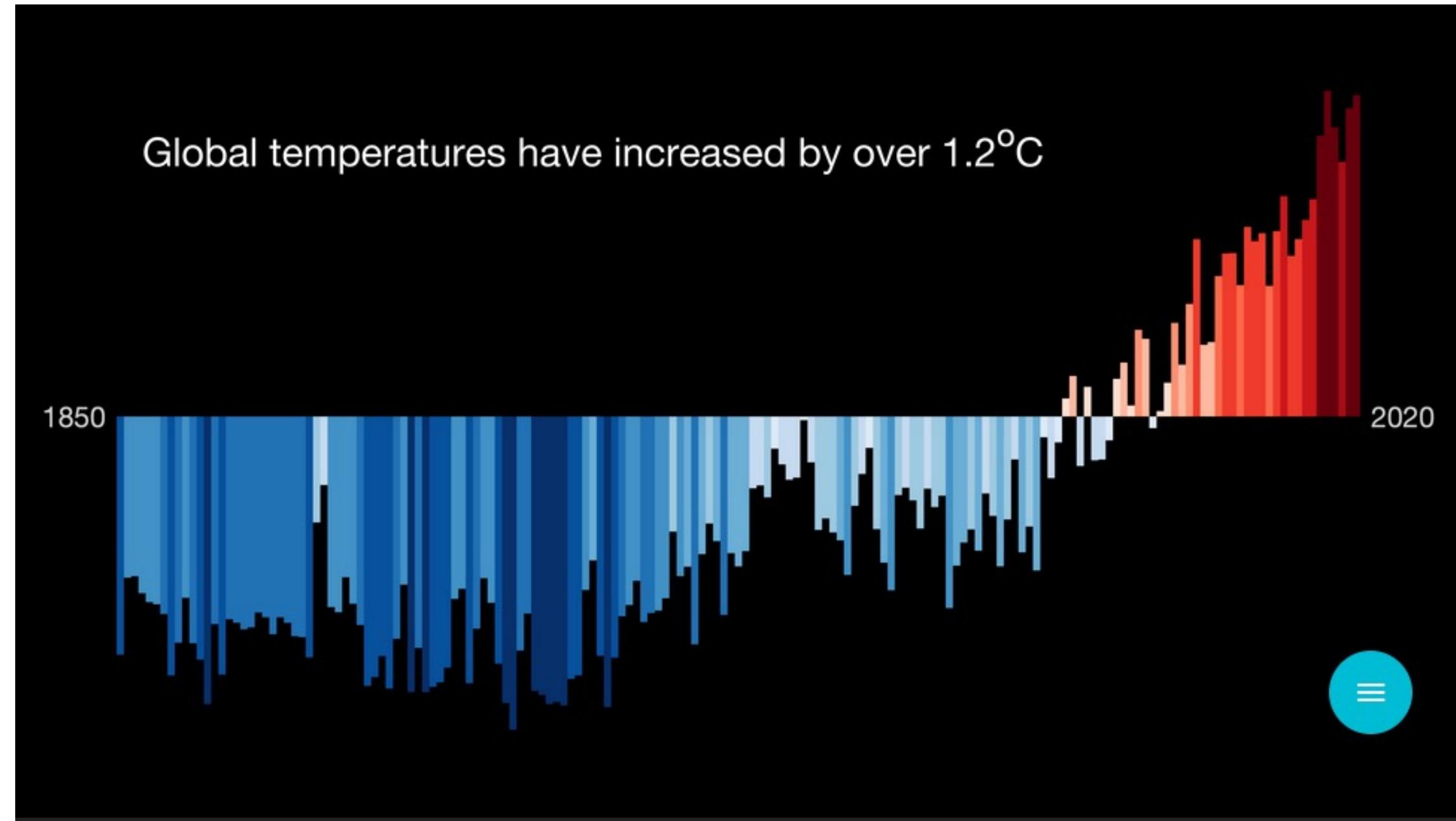
Poll #3



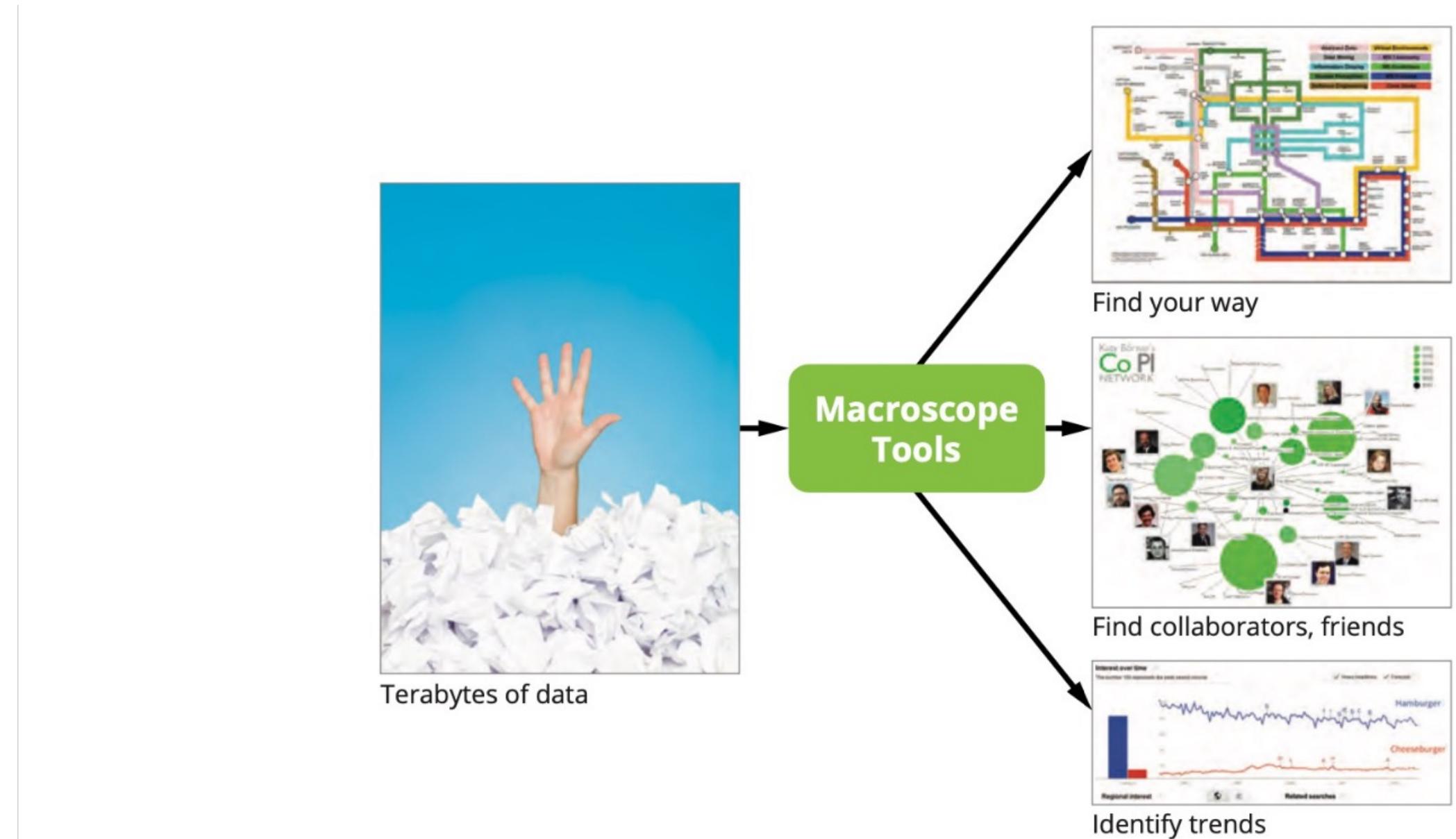
Poll #3



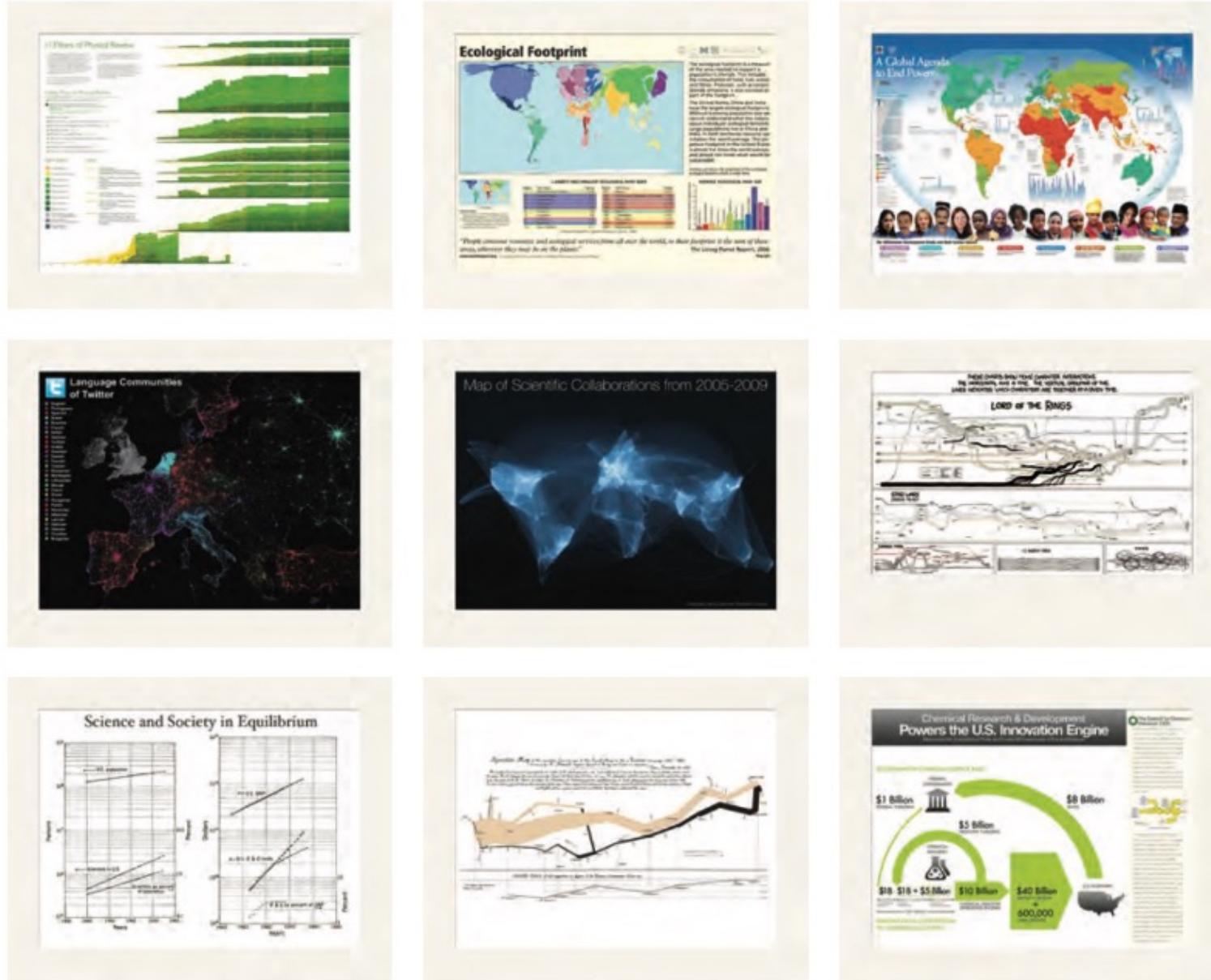
Poll #3



What is Data Visualization?



Why Use Data Visualizations?



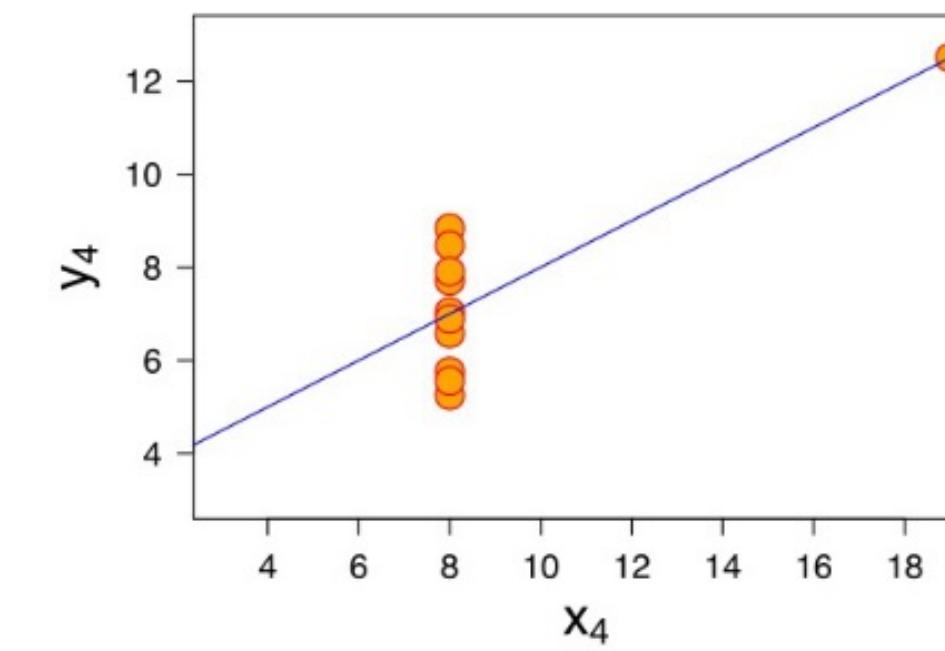
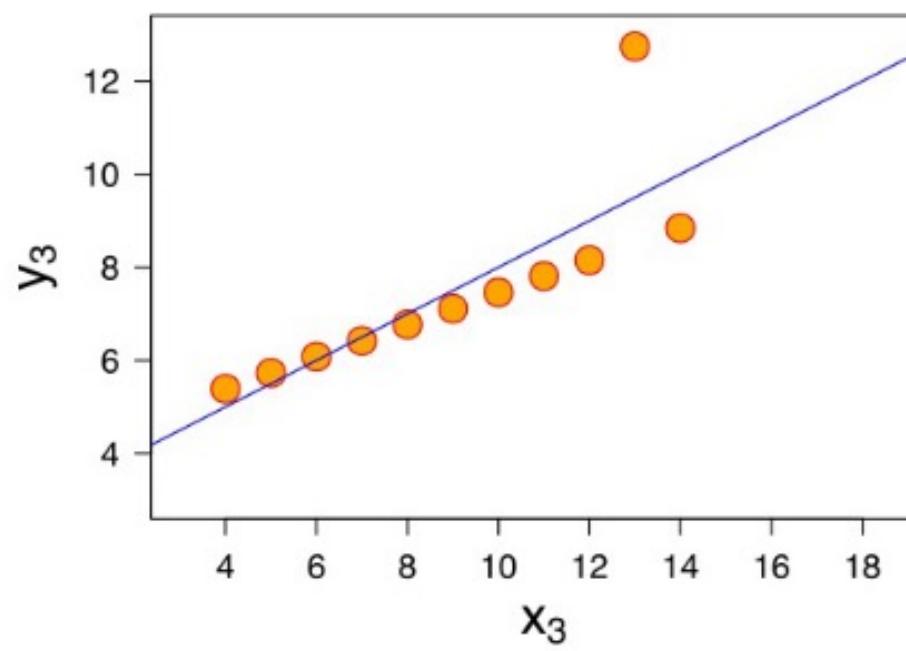
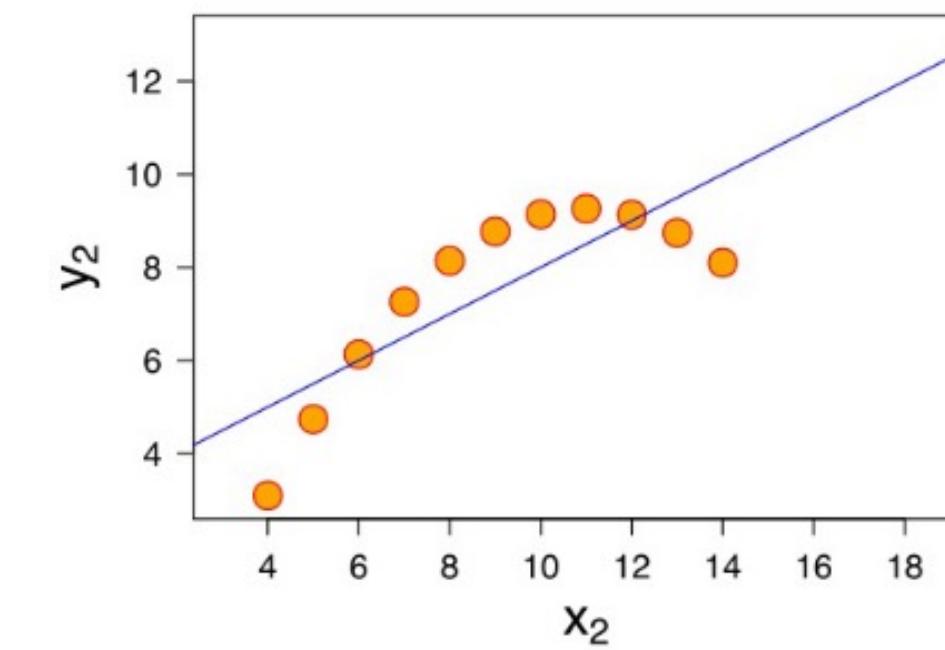
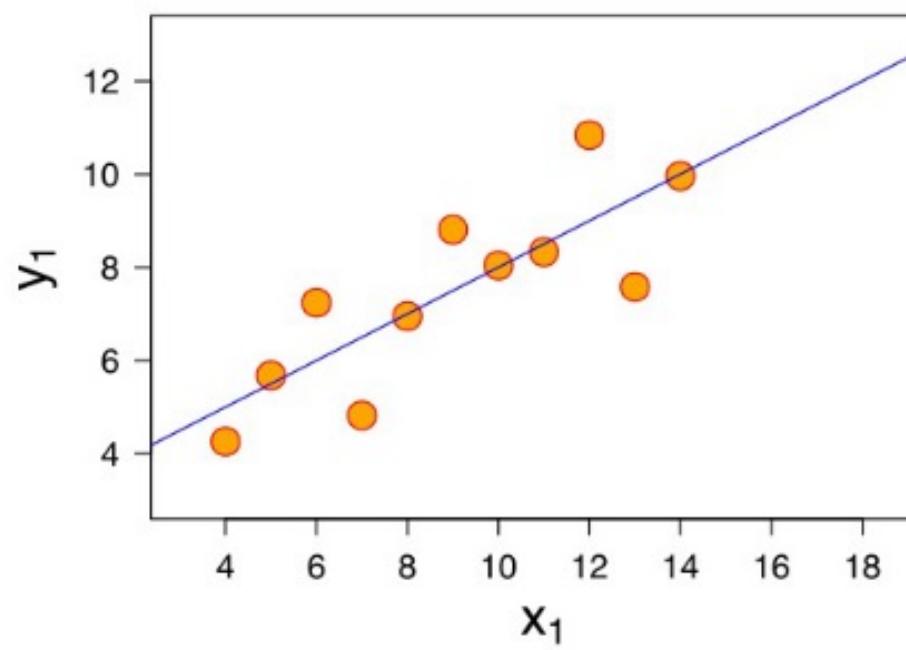
It's all about
you!



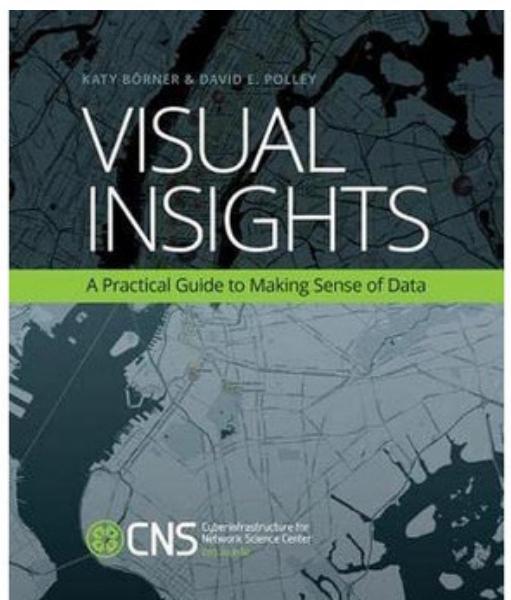
Anscombe's Quartet

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	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		

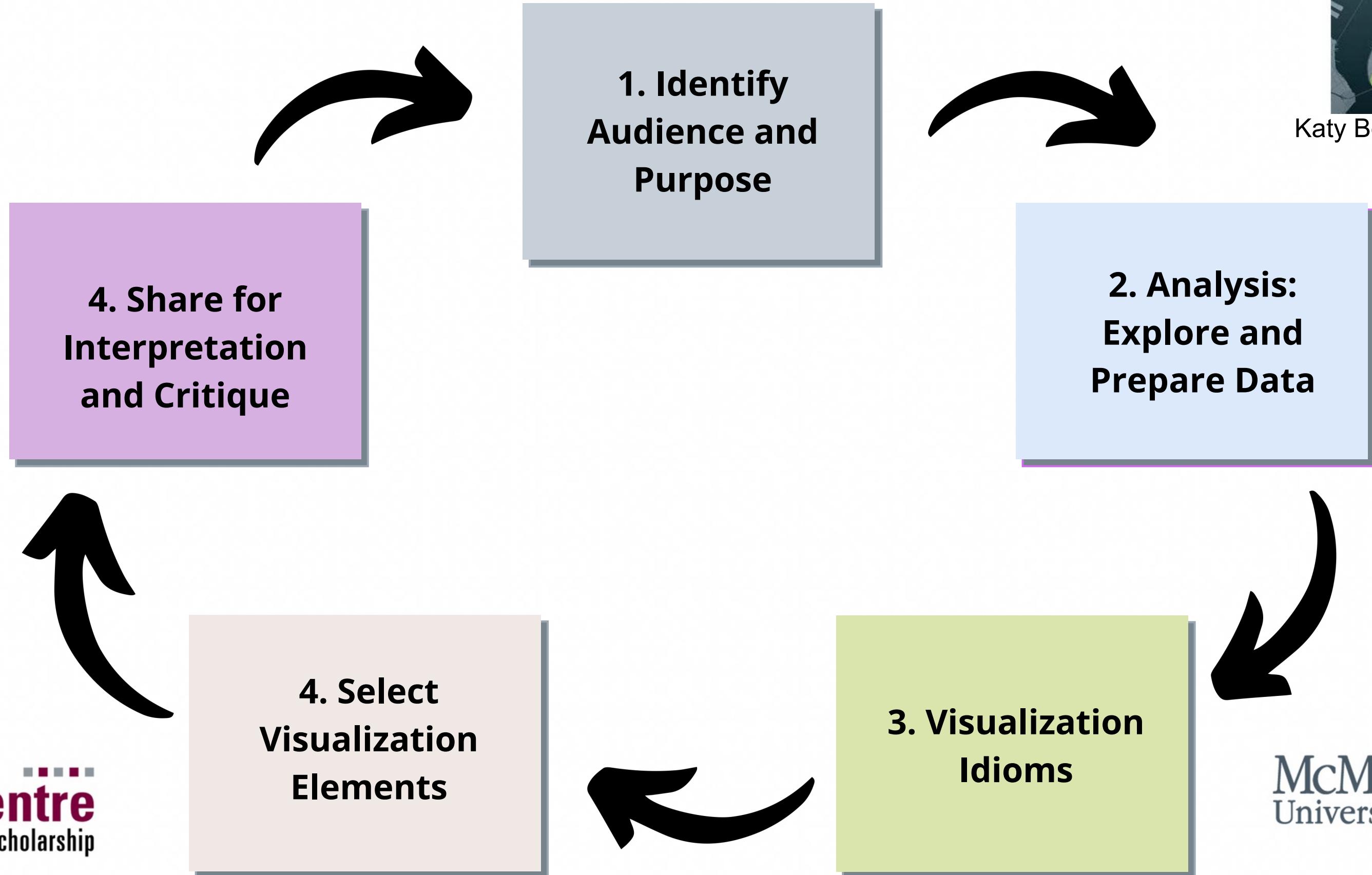




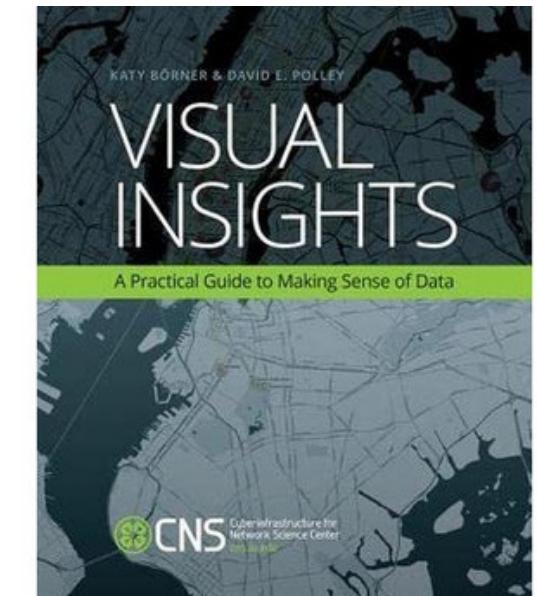
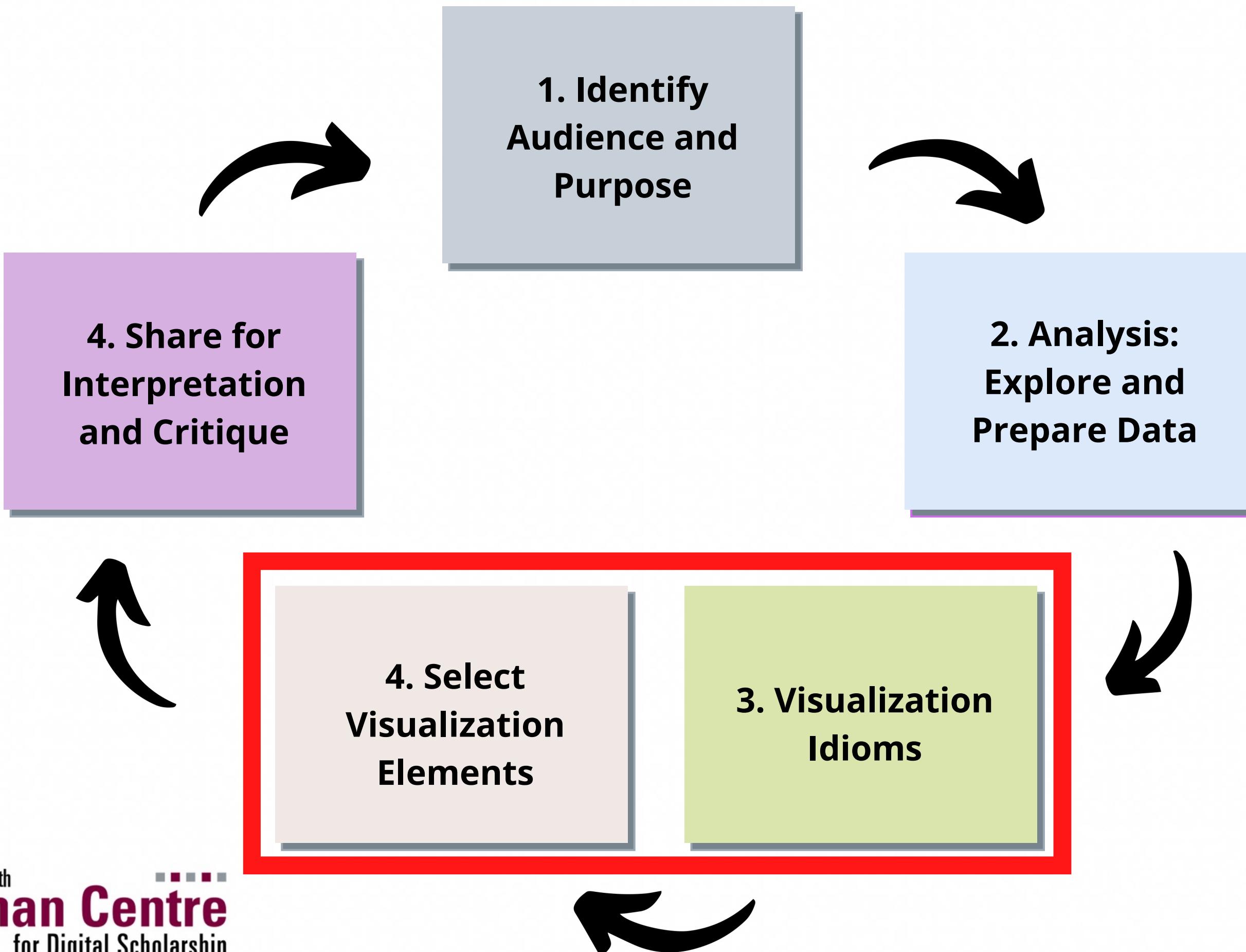
Workflow



Katy Börner, David E. Polley
+ Kelly Schulz



Workflow



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?**

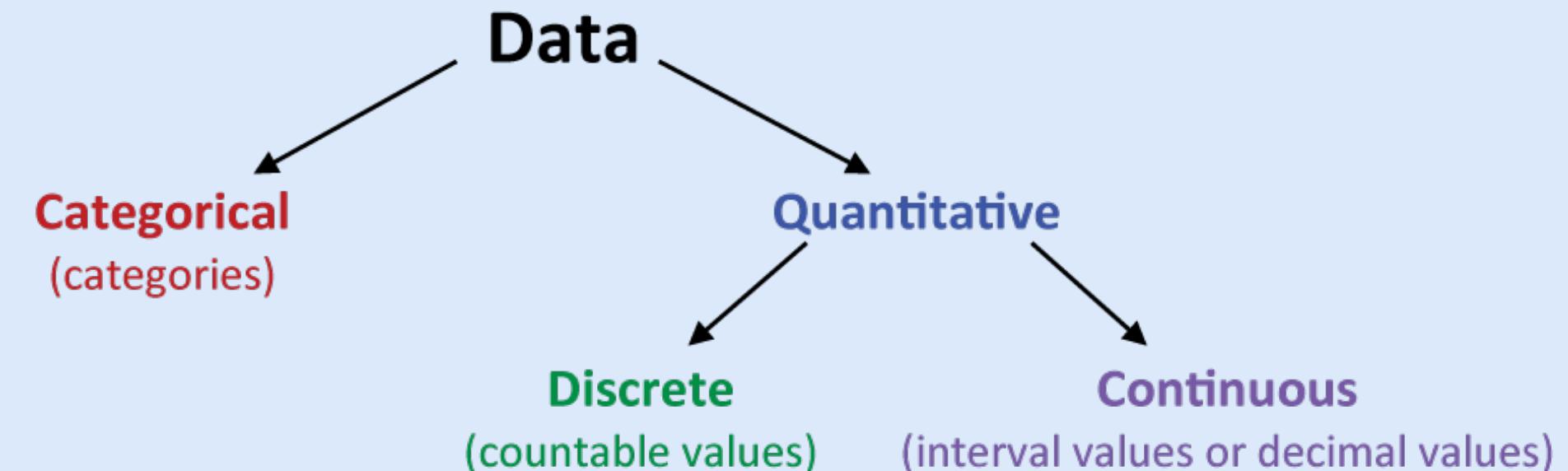
04

**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.

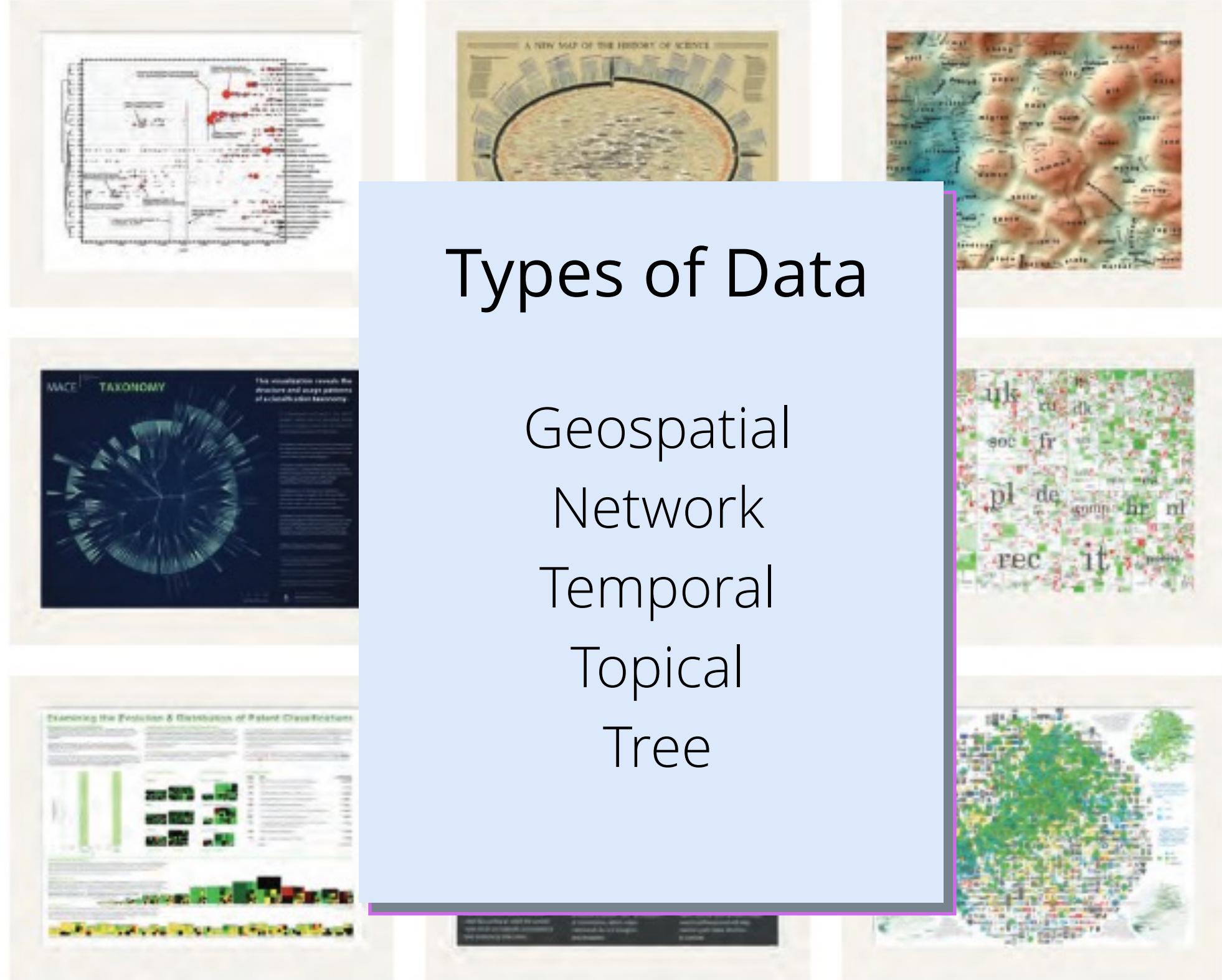
Continuous

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



Types of Data

Geospatial
Network
Temporal
Topical
Tree



Common Tasks

2. Analysis: Explore and Prepare Data

formatting values

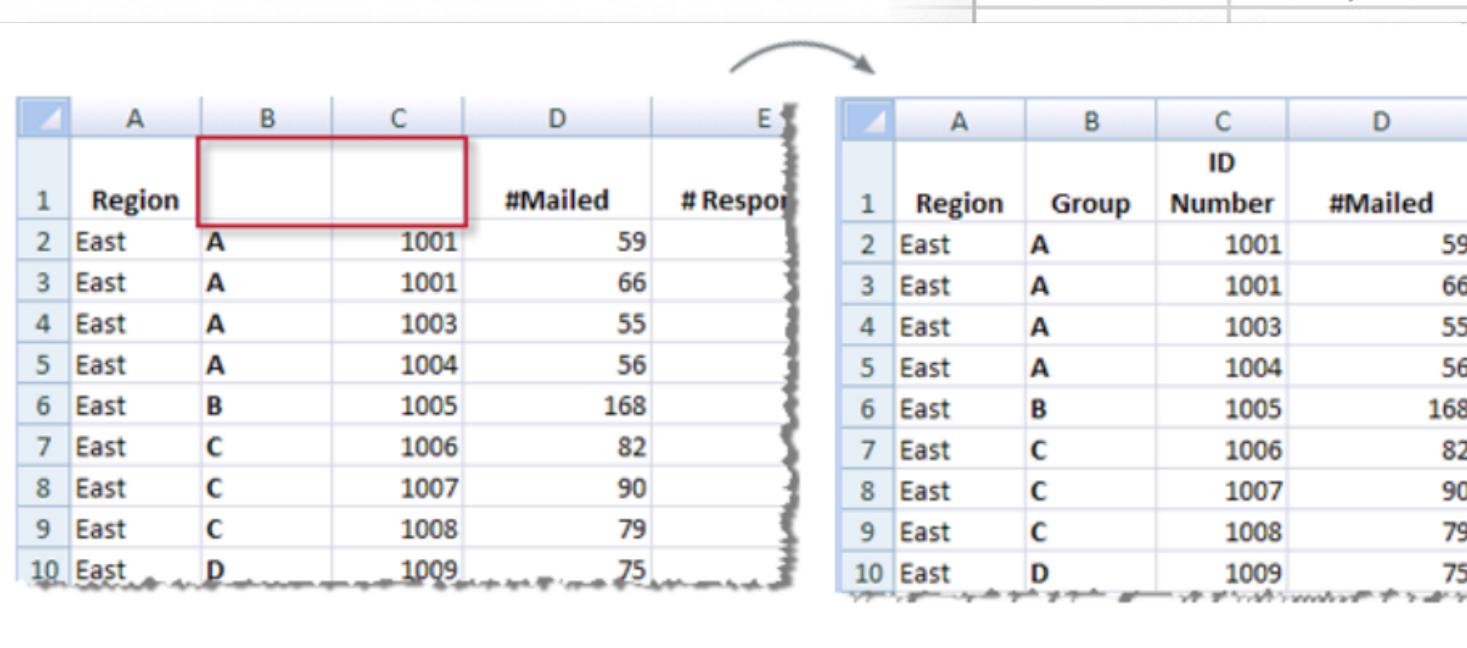
anomalies and missing data

**standardizing values and
remove pre-aggregated data**

readable headings

2. Analysis: Explore and Prepare Data

	A	B	C	D
1	Year	Country	FundingAgency	FundingAmount
2	2000	Korea, N	Dept of Agriculture	\$32 242 376
3	2000	Korea–North	Dept of Agriculture	\$86,151,301
4	2000	Korea North	department of State	166855
5	2000	SouthKorea	U.S. Agency for International Development	282,805a
6	2000	south Korea	Trade and Development Agency	735718
7	2001	North Korea	US Agency for International Development	345,399
8	2001	N Korea	Department of Argic	117715223
9	2001	So Korea	Department of agriculture	2260293
10	2001	Korea, North	State Department	183,752
11	2001	Korea, South	Trade and Development Agency	329,953
12	2002	Korea, N	Department of Agriculture	37,322,244.00
13	2002	Korea, South	U.S. Agency for International Development	67,990.00
14	2002	Korea, South	Trade and Development Agency	\$294,340
15	2003	Korea, North	U.S. Agency for International Development	\$333 823



The diagram illustrates a data transformation process. It shows two tables side-by-side. A curved arrow originates from the bottom right of the first table and points to the top left of the second table, indicating a flow or relationship between the two datasets.

	A	B	C	D	E
1	Region			#Mailed	# Respon
2	East	A	1001	59	
3	East	A	1001	66	
4	East	A	1003	55	
5	East	A	1004	56	
6	East	B	1005	168	
7	East	C	1006	82	
8	East	C	1007	90	
9	East	C	1008	79	
10	East	D	1009	75	

	A	B	C	D
1	Region	Group	ID Number	#Mailed
2	East	A	1001	59
3	East	A	1001	66
4	East	A	1003	55
5	East	A	1004	56
6	East	B	1005	168
7	East	C	1006	82
8	East	C	1007	90
9	East	C	1008	79
10	East	D	1009	75

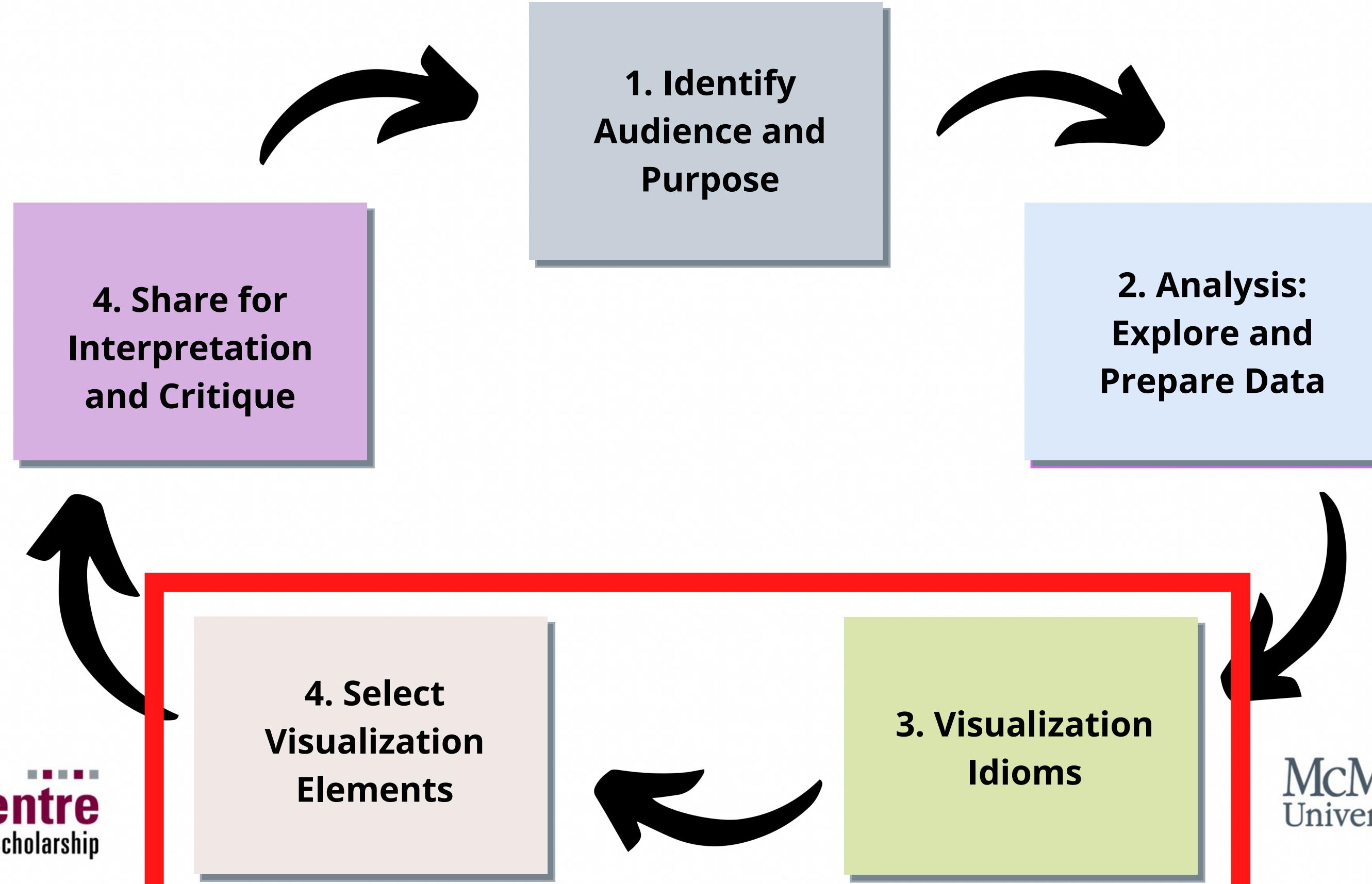
2. Analysis: Explore and Prepare Data



OpenRefine



Workflow



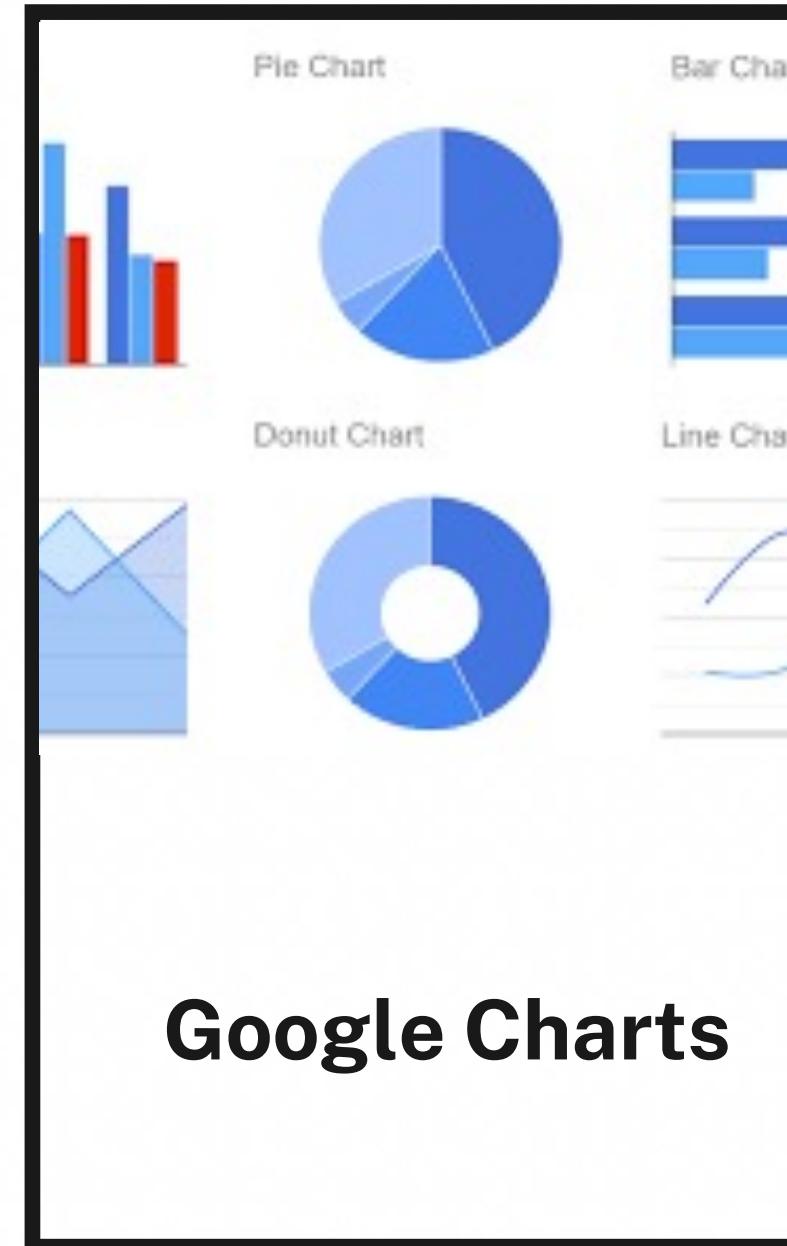
Tools for Data Visualization



MS Excel



Tableau

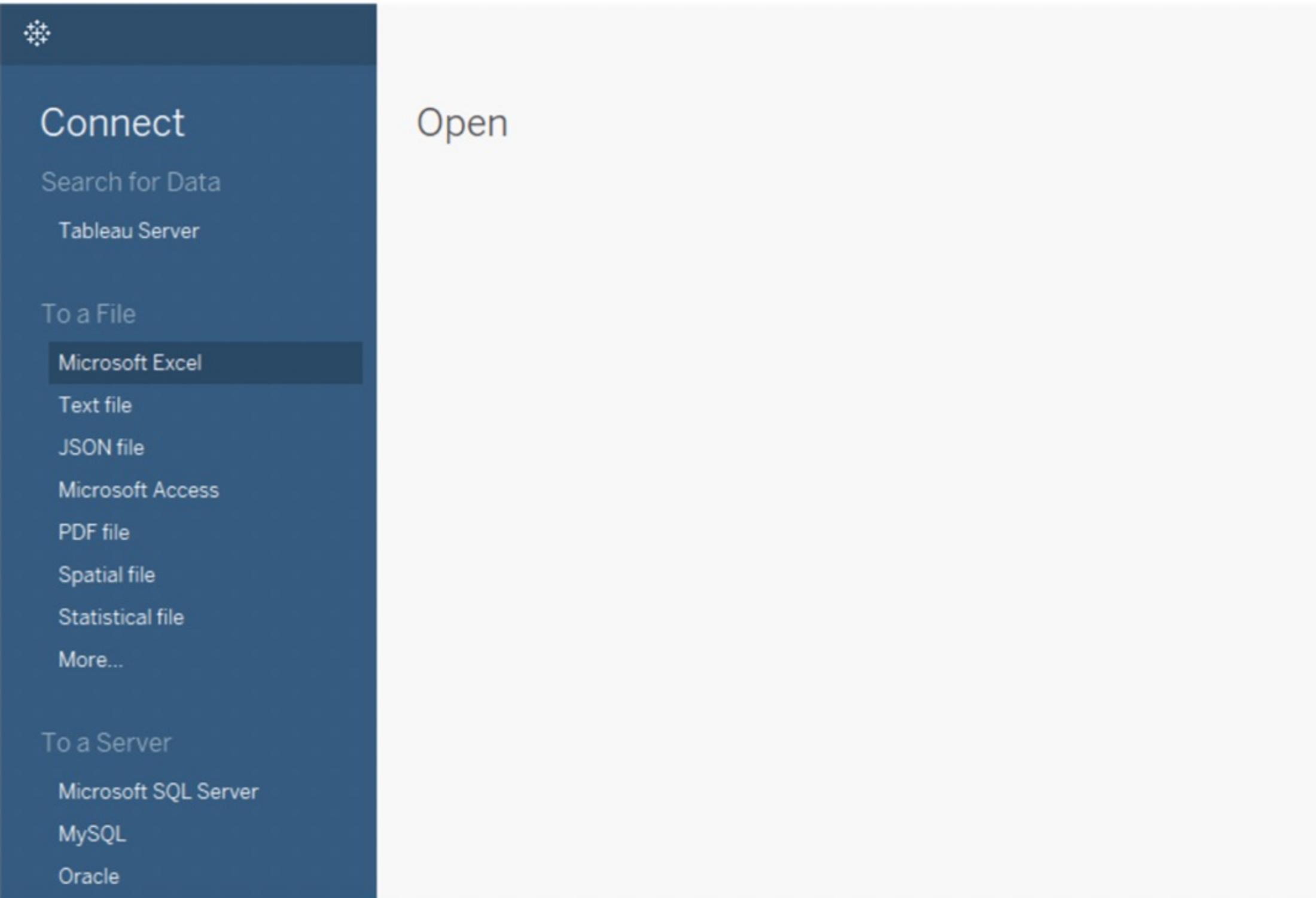


Google Charts



D3.js





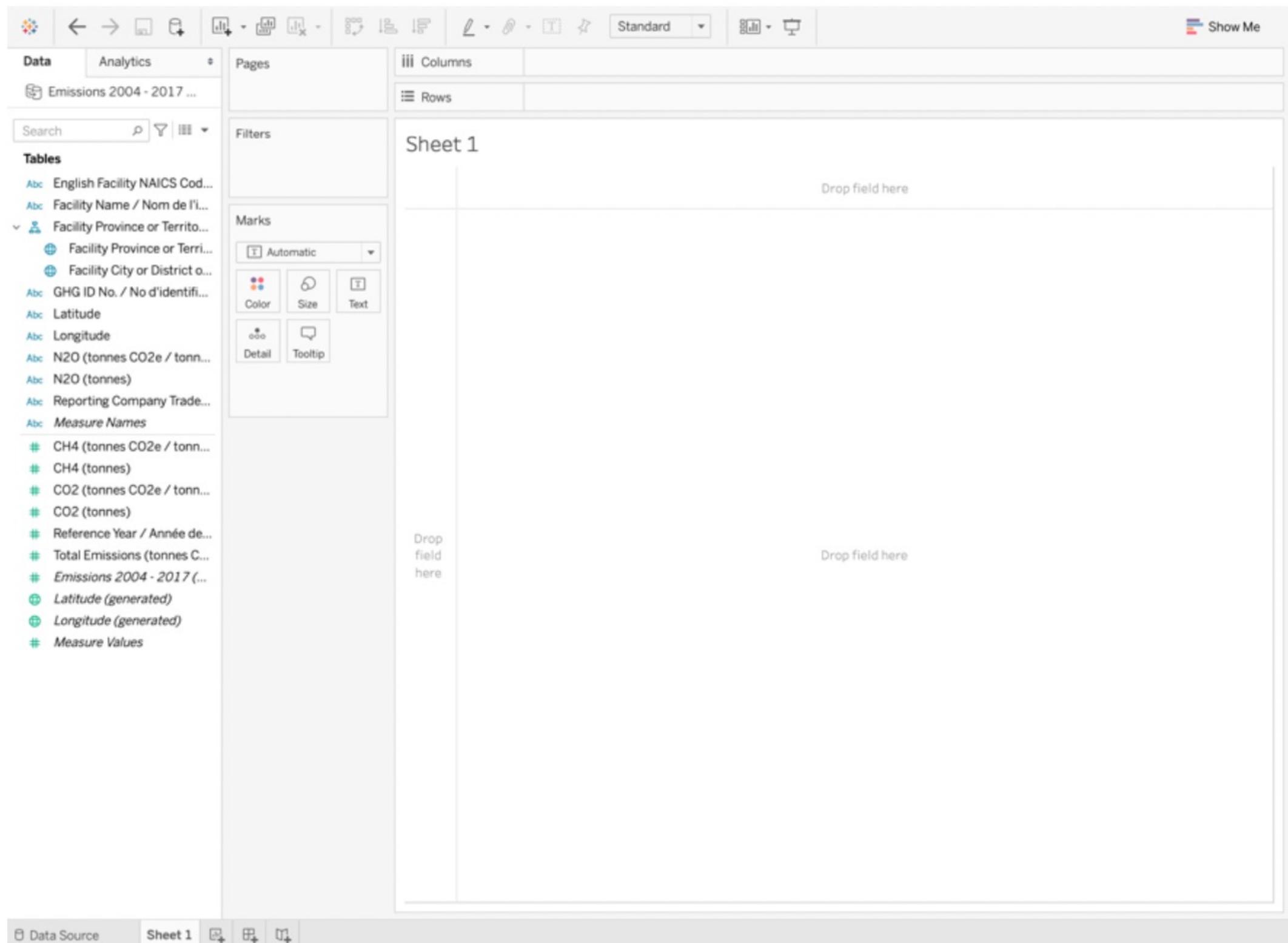
The screenshot shows the Microsoft Power BI Data Studio interface. At the top, there are navigation icons: a star, back, forward, refresh, and a circular arrow. Below them, the title "GHGEmissionsGES-2004-Present" is displayed. On the left side, there's a sidebar titled "Connections" with an "Add" button, showing one entry: "GHGEmission...004-Present" (Microsoft Excel). Under "Sheets", there are several options: "Use Data Interpreter" (unchecked), a note about the Data Interpreter, and a list of sheets: "Emissions 2004 - 2017", "Parent Company Info '04", "Read Me", "À lire", and "New Union". The "Emissions 2004 - 2017" sheet is currently selected. A tooltip "Drag sheets here" is visible near the top right. At the bottom of the sidebar, there are buttons for "Sort fields" and "Data source order".



	G10008
	G10009
	G10010
	G10011
	G10012
	G10013
	G10014
	G10015

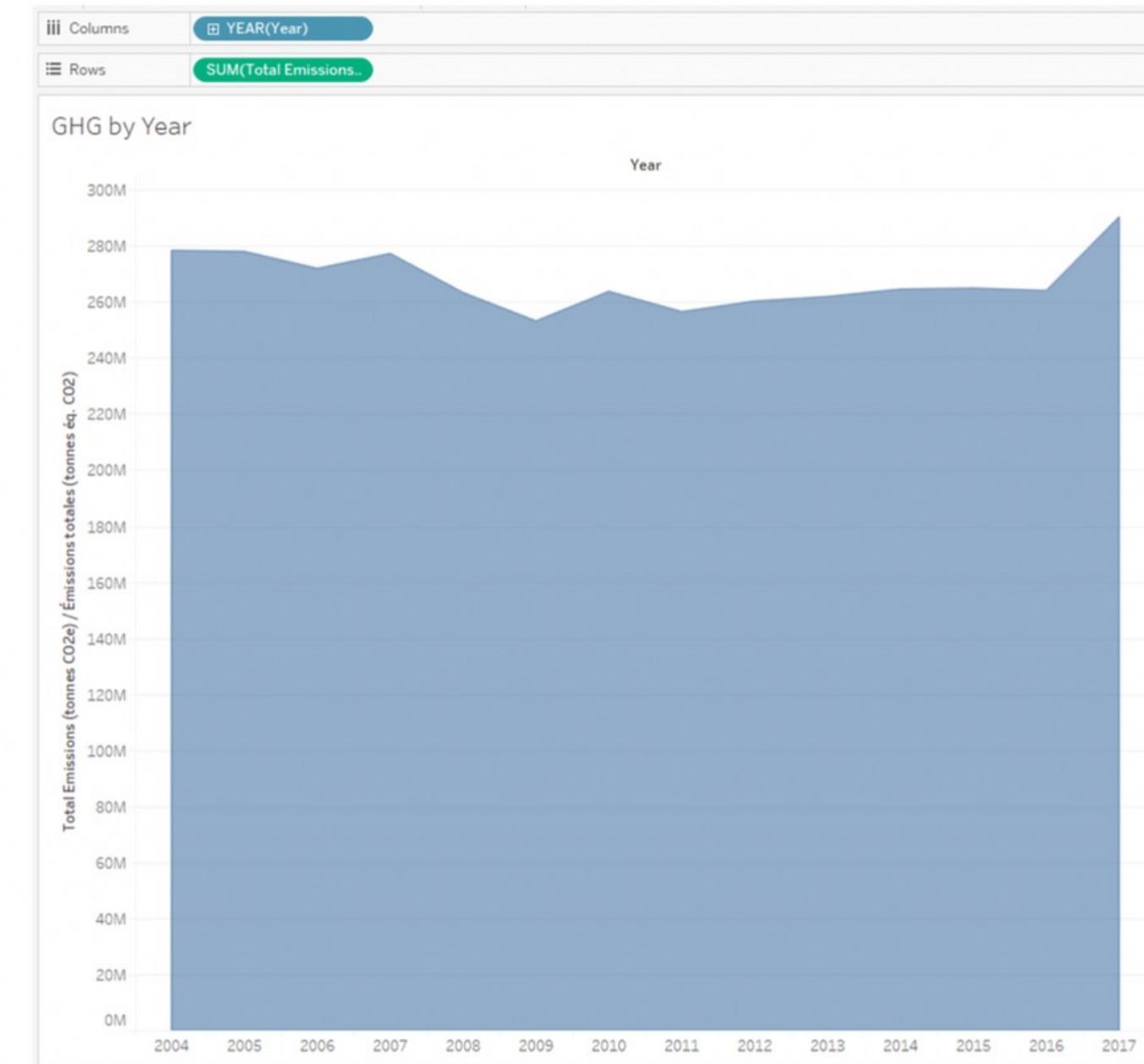
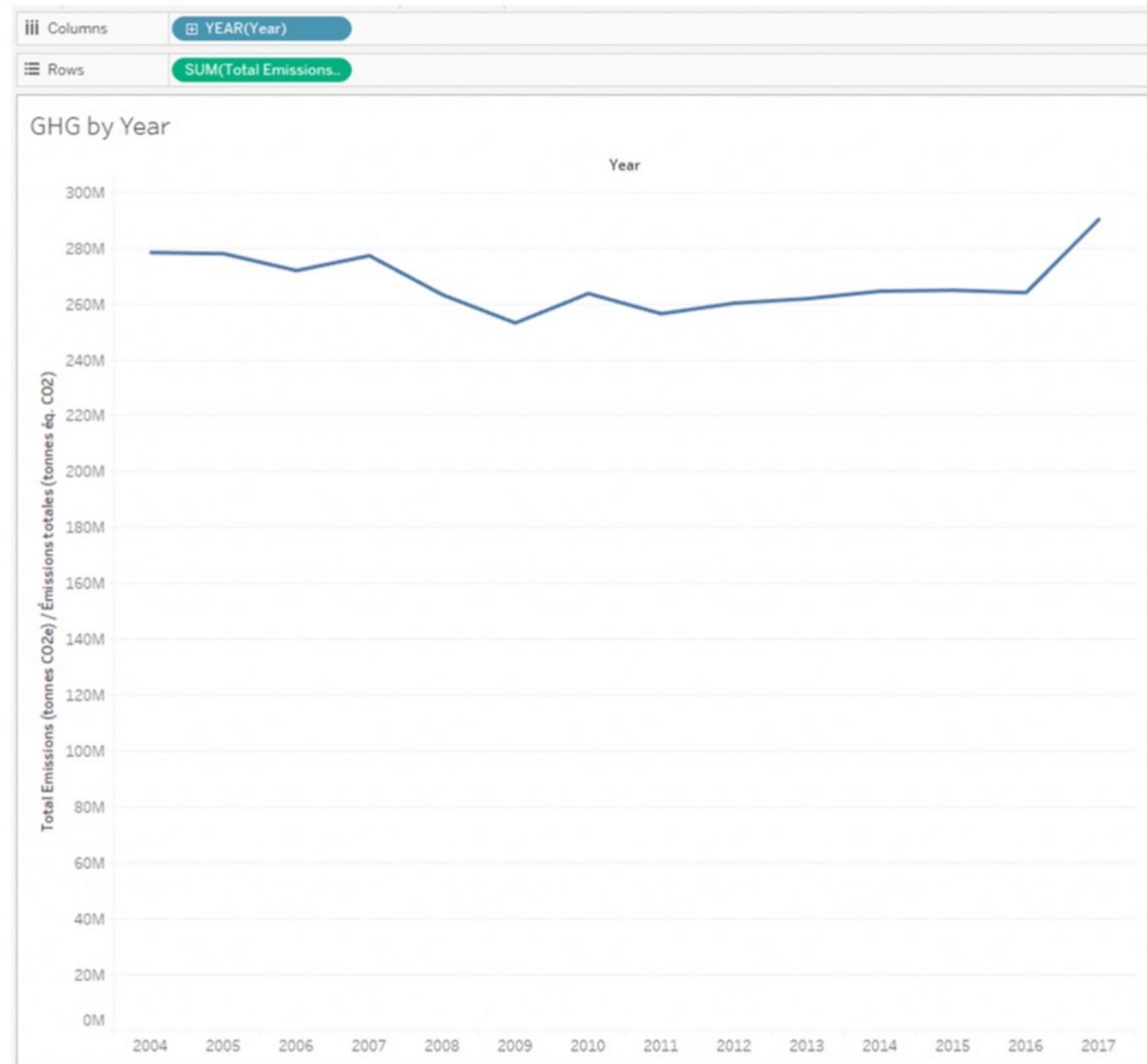
Go to Worksheet ×

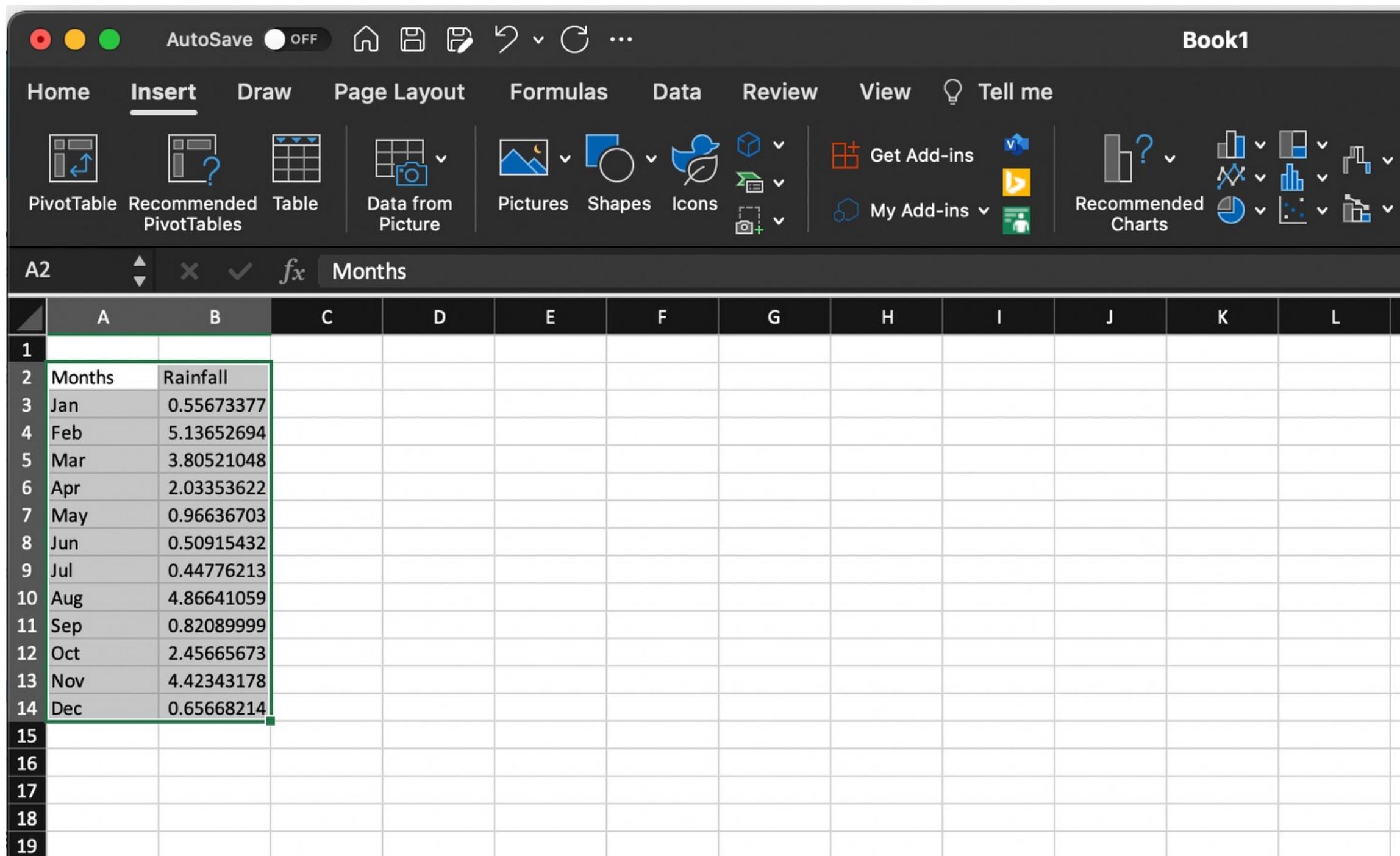
 Data Source Sheet 1   



The screenshot shows two separate context menus from the Tableau interface:

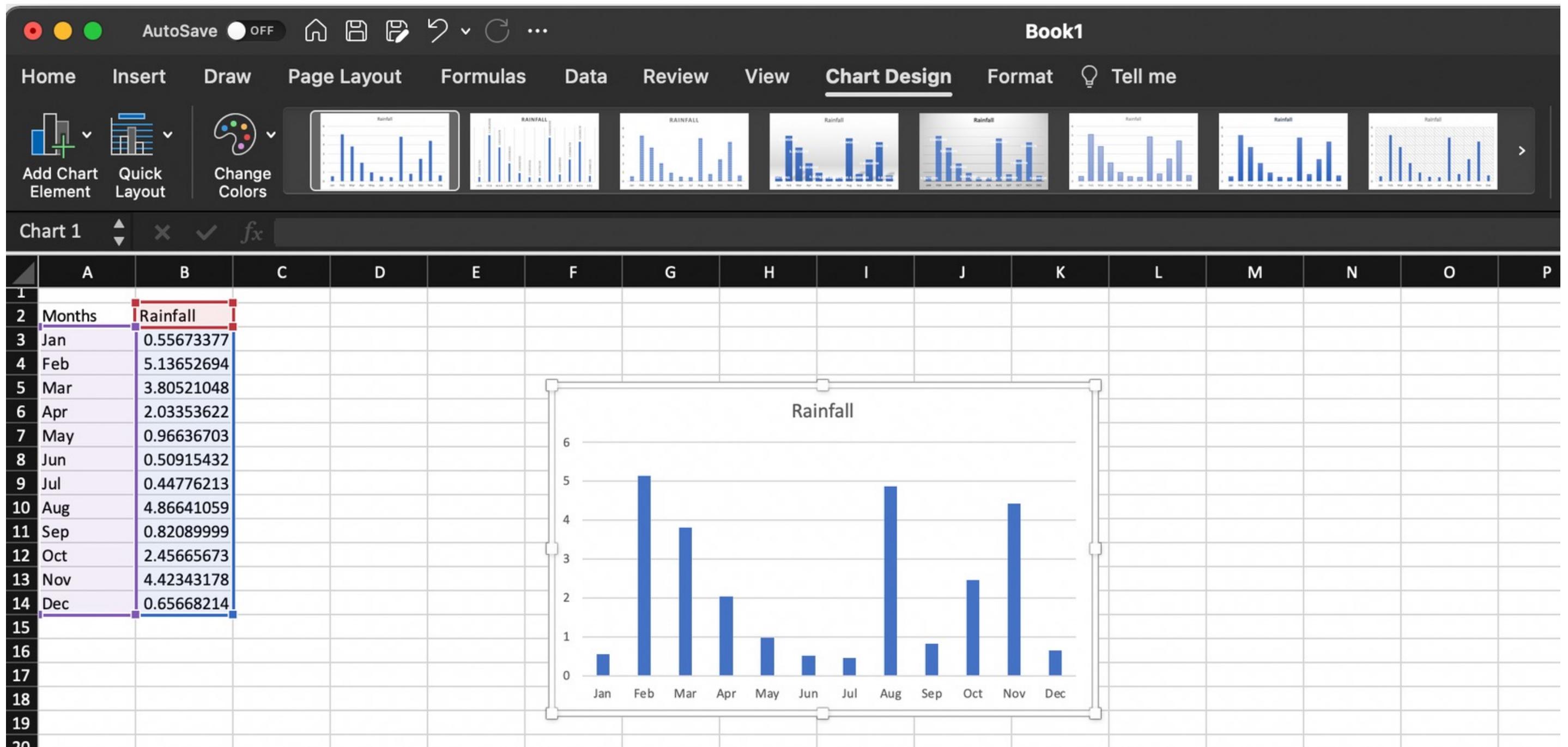
- Left Context Menu (for 'Latitude' mark):**
 - Marks**
 - Add to Sheet**
 - Duplicate
 - Rename
 - Hide
 - Aliases...
 - Create
 - Transform
 - Convert to Measure
 - Change Data Type
 - Geographic Role
 - Default Properties
 - Group by
 - Folders
 - Hierarchy
 - Replace References...
 - Describe...
- Right Context Menu (for 'Longitude' mark):**
 - Marks**
 - Add to Sheet**
 - Duplicate
 - Rename
 - Hide
 - Aliases...
 - Create
 - Transform
 - Convert to Continuous
 - Convert to Measure
 - Change Data Type
 - Geographic Role**
 - None
 - Airport
 - Area Code (U.S.)
 - CBSA/MSA (U.S.)
 - City
 - Congressional District (U.S.)
 - Country/Region
 - County
 - Latitude**
 - Longitude
 - NUTS Europe
 - State/Province
 - ZIP Code/Postcode
 - Create from





A screenshot of Microsoft Excel showing a table of monthly rainfall data. The table has two columns: 'Months' and 'Rainfall'. The data starts from row 2 and continues to row 14. The 'Months' column lists the months from Jan to Dec. The 'Rainfall' column lists the corresponding rainfall values.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Months	Rainfall										
3	Jan	0.55673377										
4	Feb	5.13652694										
5	Mar	3.80521048										
6	Apr	2.03353622										
7	May	0.96636703										
8	Jun	0.50915432										
9	Jul	0.44776213										
10	Aug	4.86641059										
11	Sep	0.82089999										
12	Oct	2.45665673										
13	Nov	4.42343178										
14	Dec	0.65668214										
15												
16												
17												
18												
19												



Google Charts

Home Guides Reference Support

Filter

Chart types

- Chart Gallery
- Annotation Charts
- Area Charts
- Bar Charts
- Bubble Charts**
- Calendar Charts
- Candlestick Charts
- Column Charts
- Combo Charts
- Diff Charts
- Donut Charts
- Gantt Charts
- Gauge Charts
- GeoCharts
- Histograms
- Intervals
- Line Charts
- Maps
- Org Charts
- Pie Charts
- Sankey Diagrams
- Scatter Charts
- Stepped Area Charts
- Table Charts
- Timelines
- Tree Map Charts
- Trendlines
- VegaChart
- Waterfall Charts
- Word Trees
- Miscellaneous Examples

Home > Products > Charts > Guides

Was this helpful?

Visualization: Bubble Chart

[Send feedback](#)

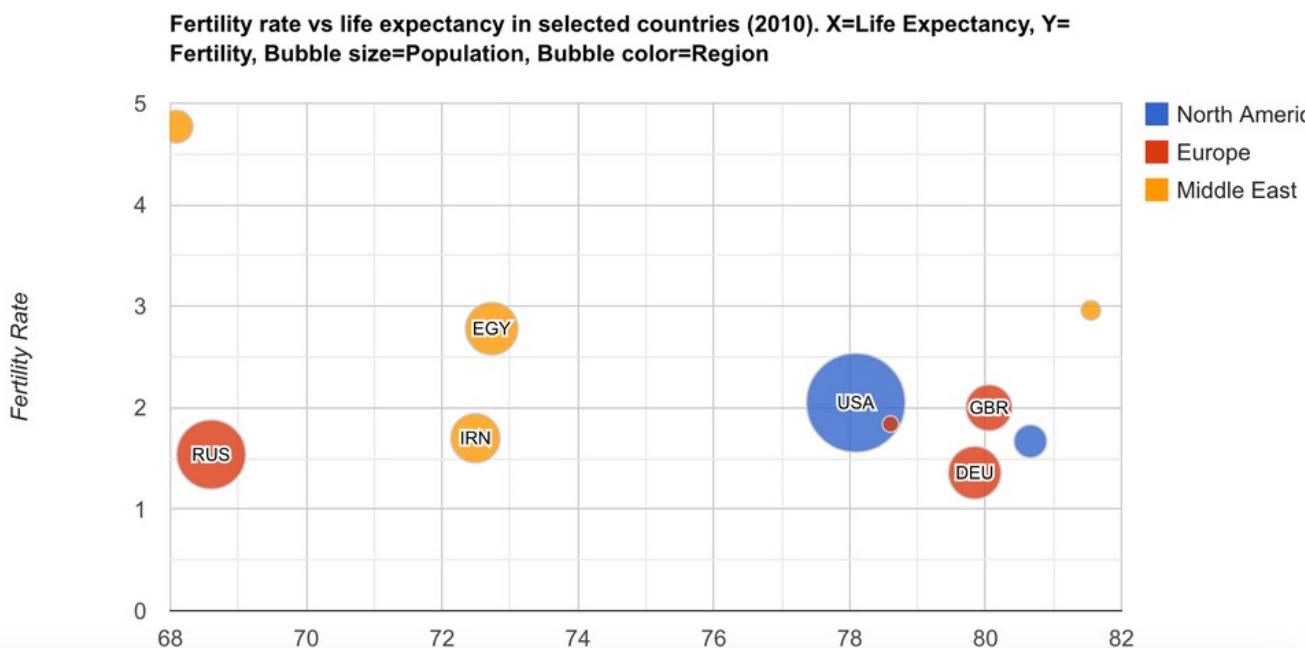
Overview

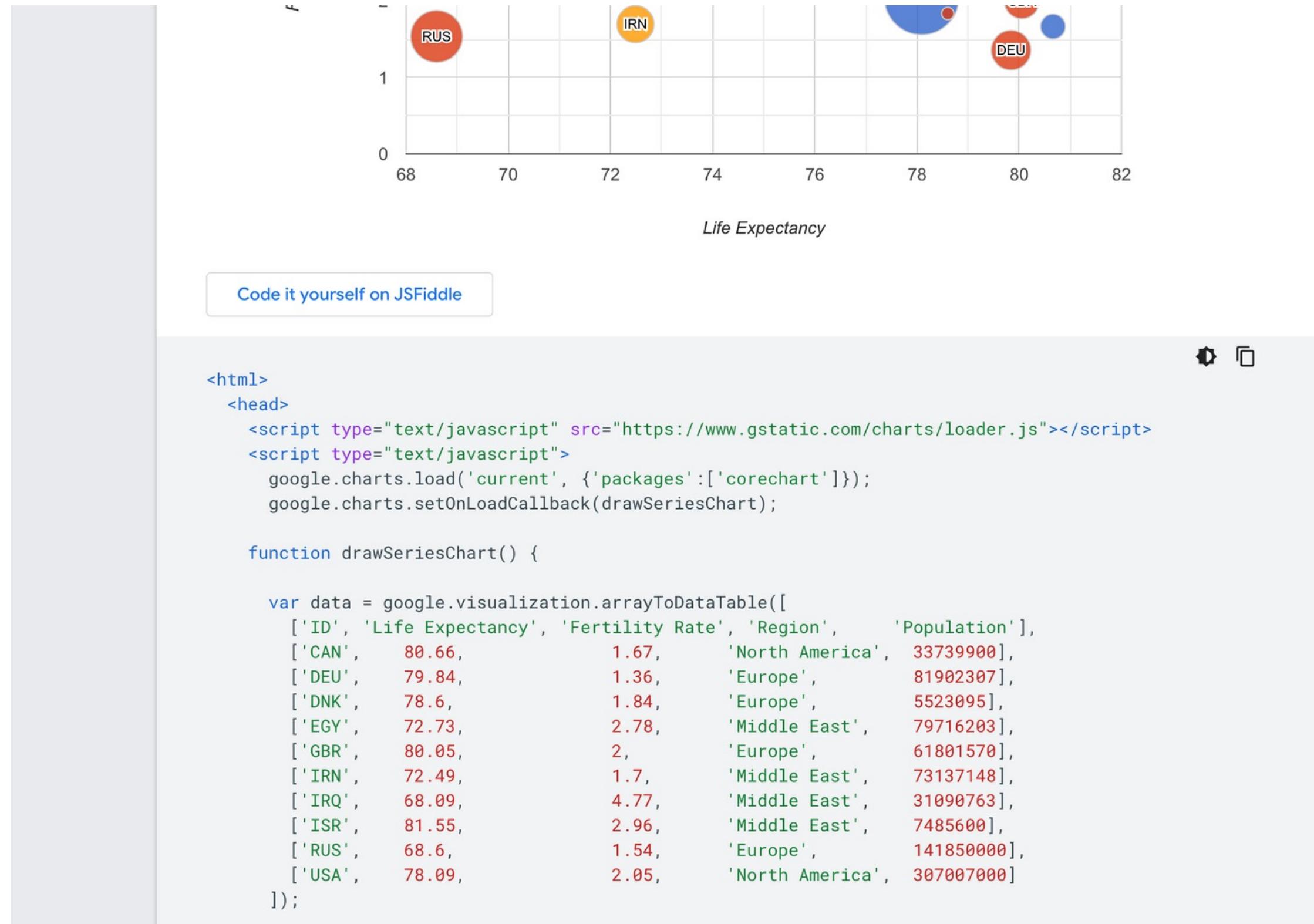
A bubble chart that is rendered within the browser using [SVG](#) or [VML](#). Displays tips when hovering over bubbles.

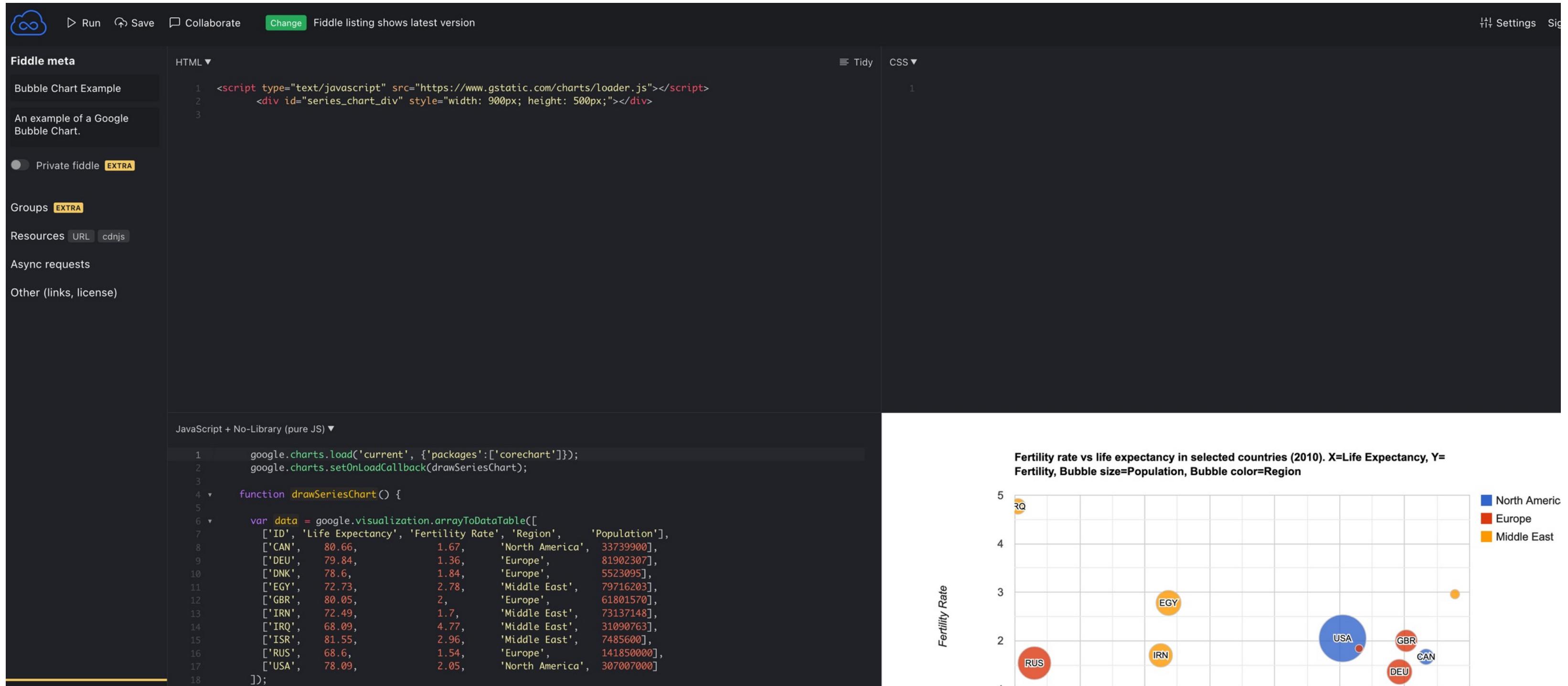
A bubble chart is used to visualize a data set with two to four dimensions. The first two dimensions are visualized as coordinates, the third as color and the fourth as size.

Example

Series Example







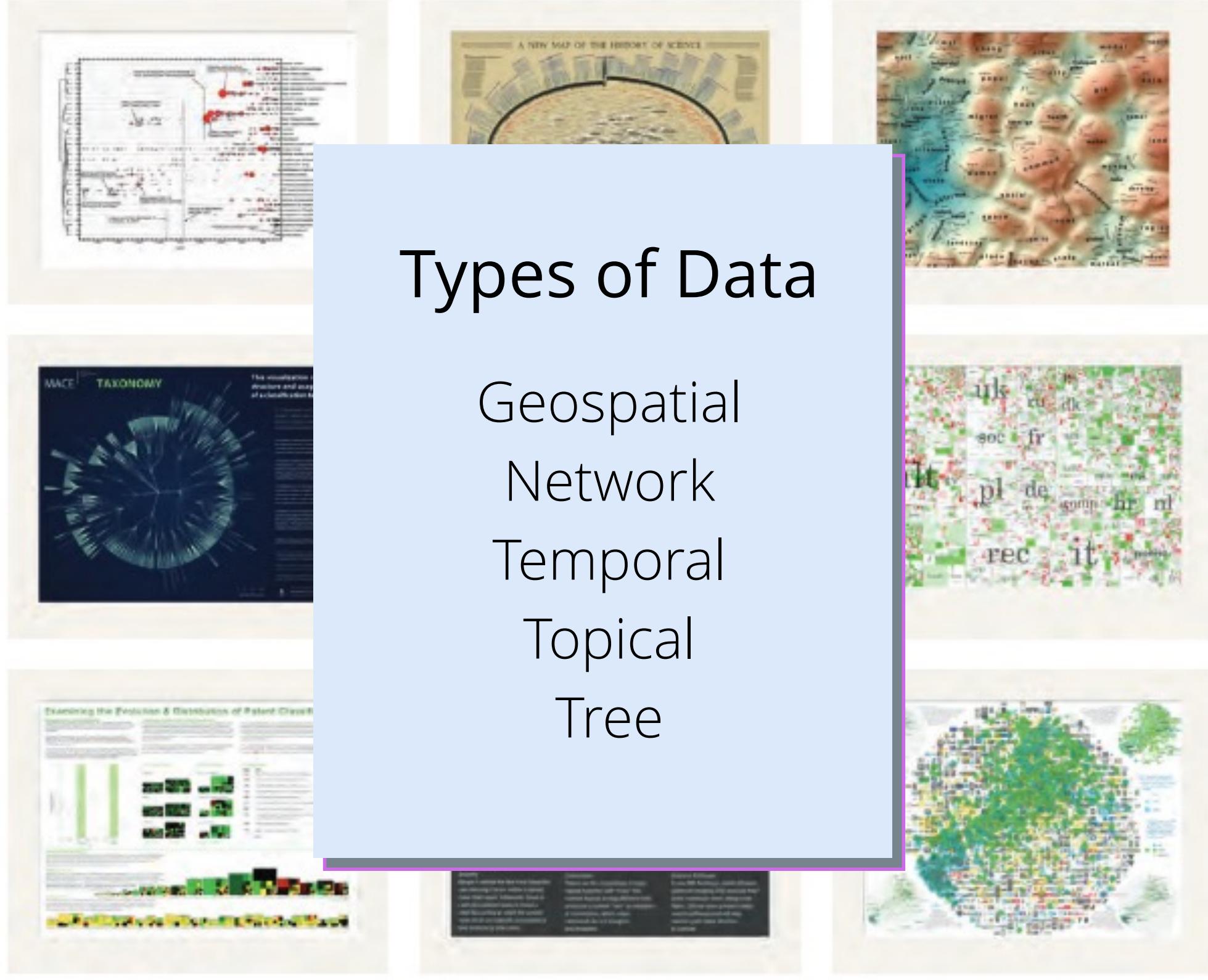
Configuration Options

Name	
animation.duration	<p>The duration of the animation, in milliseconds. For details, see the animation documentation.</p> <p>Type: number Default: 0</p>
animation.easing	<p>The easing function applied to the animation. The following options are available:</p> <ul style="list-style-type: none">• 'linear' - Constant speed.• 'in' - Ease in - Start slow and speed up.• 'out' - Ease out - Start fast and slow down.• 'inAndOut' - Ease in and out - Start slow, speed up, then slow down. <p>Type: string Default: 'linear'</p>
animation.startup	<p>Determines if the chart will animate on the initial draw. If <code>true</code>, the chart will start at the baseline and animate to its final state.</p> <p>Type: boolean Default: false</p>
axisTitlesPosition	<p>Where to place the axis titles, compared to the chart area. Supported values:</p> <ul style="list-style-type: none">• in - Draw the axis titles inside the chart area.• out - Draw the axis titles outside the chart area.• none - Omit the axis titles. <p>Type: string Default: 'out'</p>
backgroundColor	<p>The background color for the main area of the chart. Can be either a simple HTML color string, for example: '<code>red</code>' or '<code>#00cc00</code>', or an object with the following properties.</p> <p>Type: string or object</p>

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

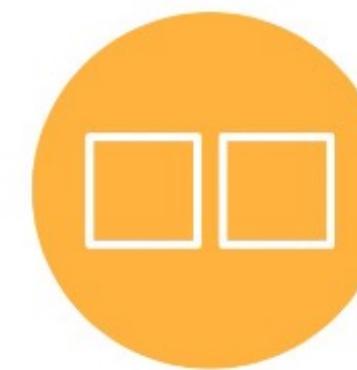
05 Tree Sunburst diagram, Tree
Map, Flowchart

The Data Visualisation Catalogue

3. Visualization Idioms

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.



Comparisons



Proportions



Relationships



Hierarchy



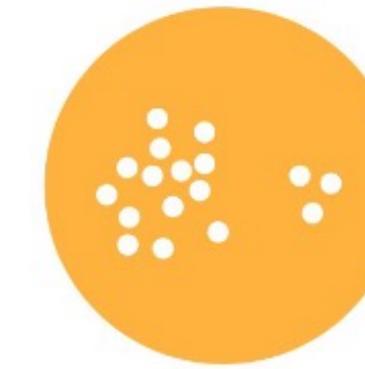
Concepts



Location



Part-to-a-whole

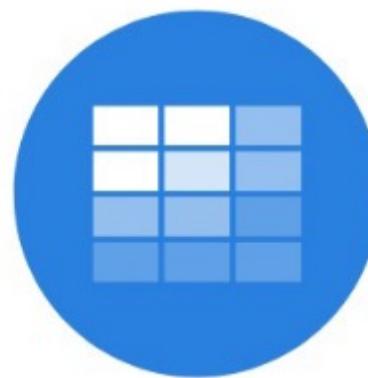


Distribution

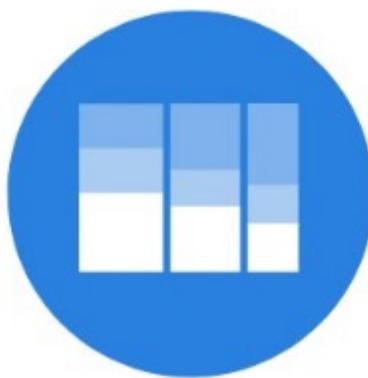
3. Visualization Idioms

Relationships

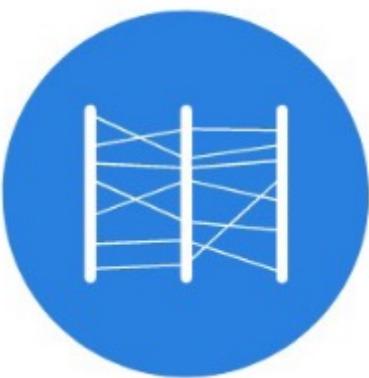
Relationships: Visualization methods that show relationships and connections between the data or show correlations between two or more variables.



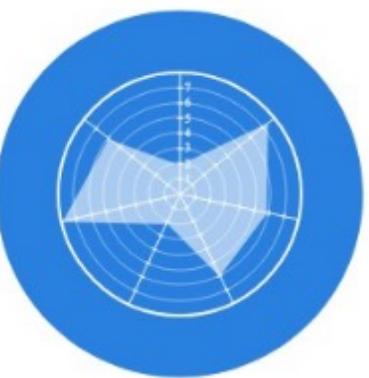
Heatmap



Marimekko Chart



Parallel
Coordinates Plot



Radar Chart



Venn Diagram

3. Visualization Idioms

Heatmap (Matrix)

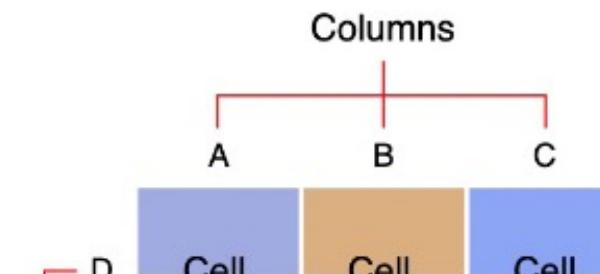


Description

Heatmaps visualise data through variations in colouring. When applied to a tabular format, Heatmaps are useful for cross-examining multivariate data, through placing variables in the rows and columns and colouring the cells within the table. Heatmaps are good for showing variance across multiple variables, revealing any patterns, displaying whether any variables are similar to each other, and for detecting if any correlations exist in-between them.

Anatomy

Heatmap using numerical data:



3. Visualization Idioms

Tools to Generate Visualisation

MS Excel and Apple Numbers

[AnyChart \(code\)](#)

[amCharts \(code\)](#)

[Britecharts \(code\)](#)

[FusionCharts \(code\)](#)

[Highcharts](#)

[jChartFX](#)

[JSCharting \(JS Library\)](#)

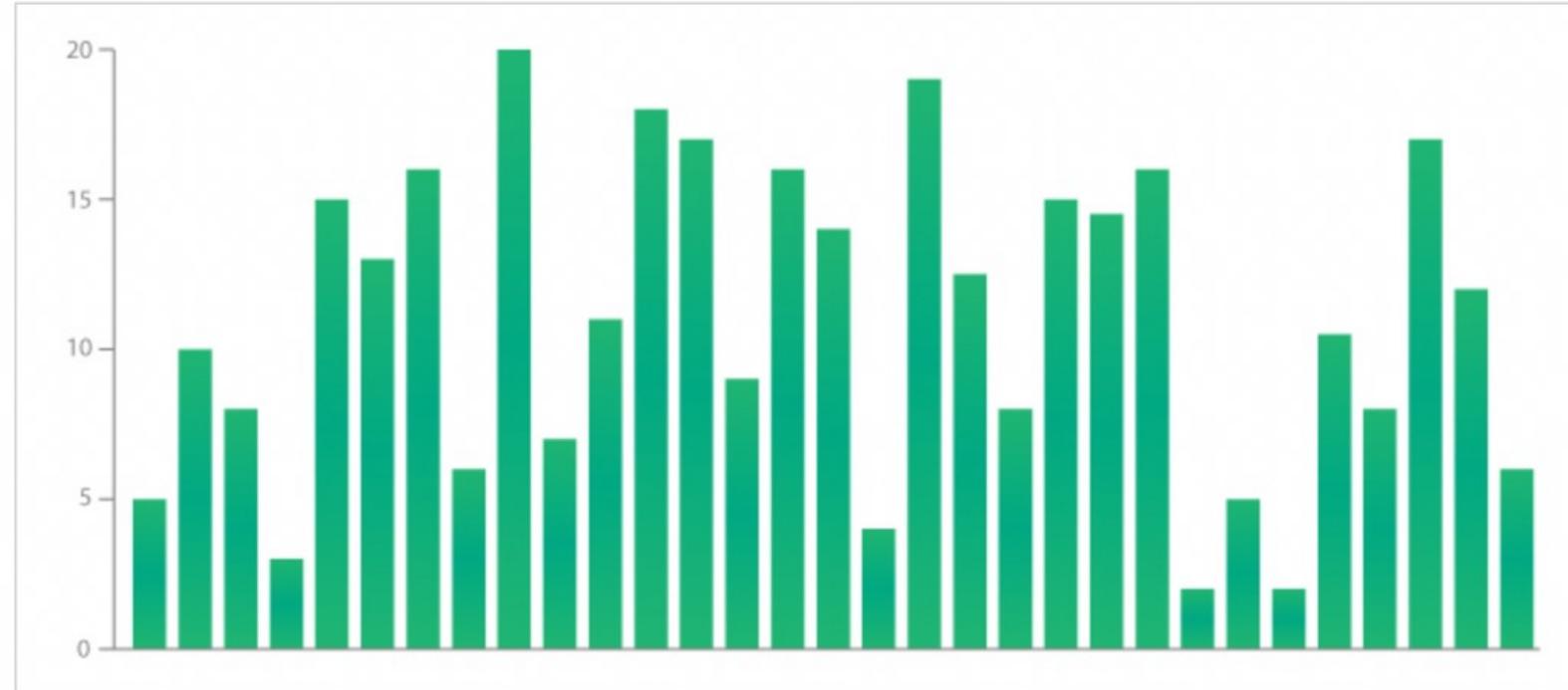
[plot.ly](#)

[R Graph Gallery & Python Graph Gallery \(code\)](#)

[Zing Chart \(code\)](#)

3. Visualization Idioms

Bar Chart



Description

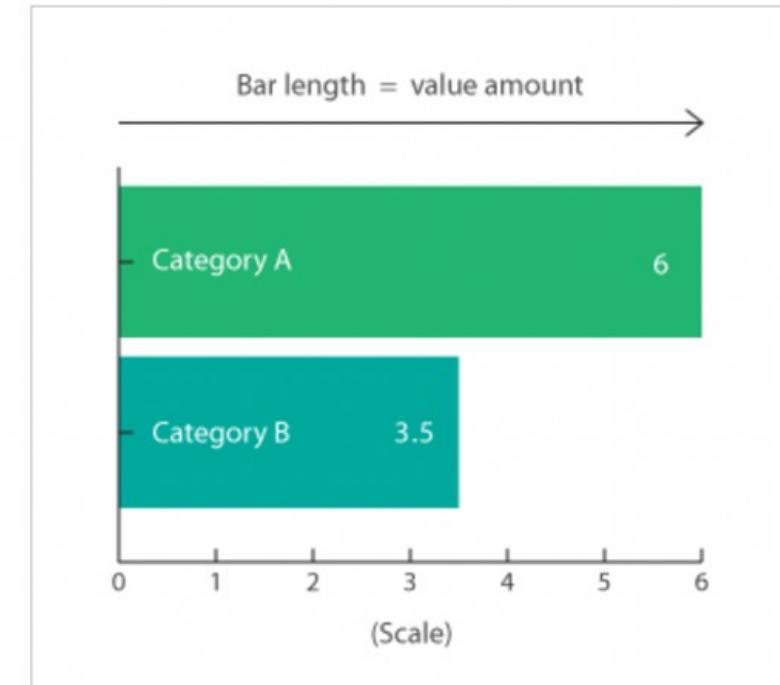
As known as *Bar Graph* or *Column Graph*.

The classic Bar Chart uses either horizontal or vertical bars (column chart) to show discrete, numerical comparisons across categories. One axis of the chart shows the specific categories being compared and the other axis represents a discrete value scale.

Bars Charts are distinguished from [Histograms](#), as they do not display continuous developments over an interval. Bar Chart's discrete data is categorical data and therefore answers the question of "how many?" in each category.

One major flaw with Bar Charts is that labelling becomes problematic when there are a large number of bars.

Anatomy

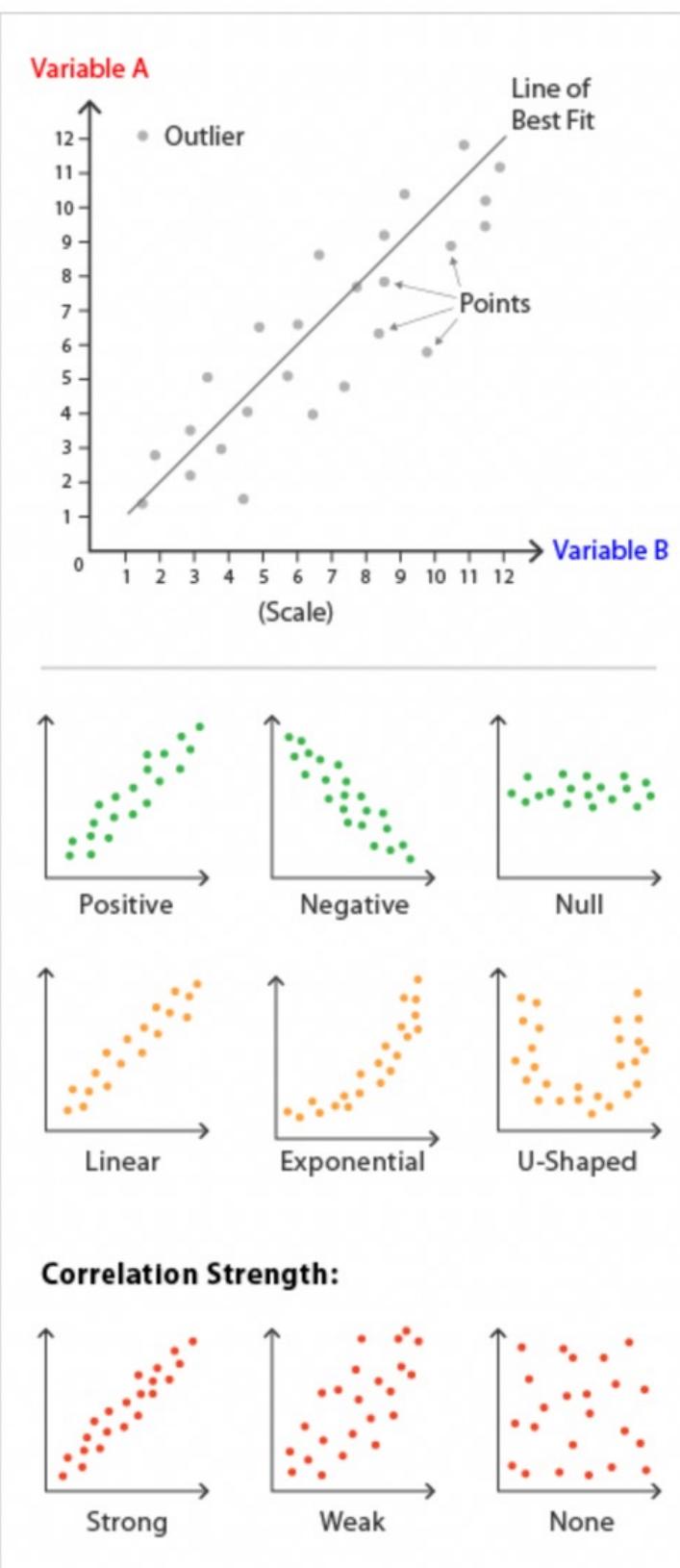


Functions

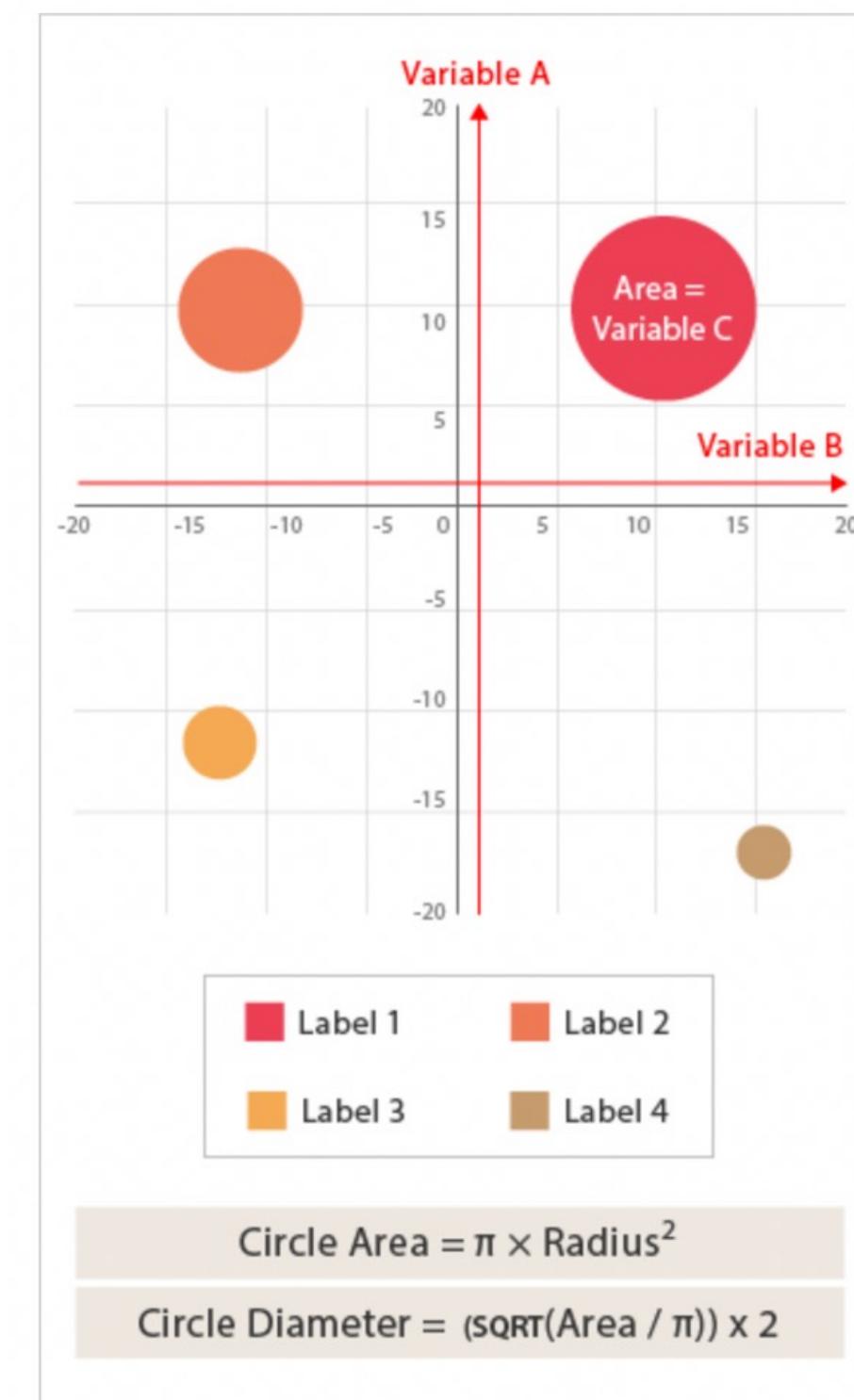


3. Visualization Idioms

Anatomy



Anatomy



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.

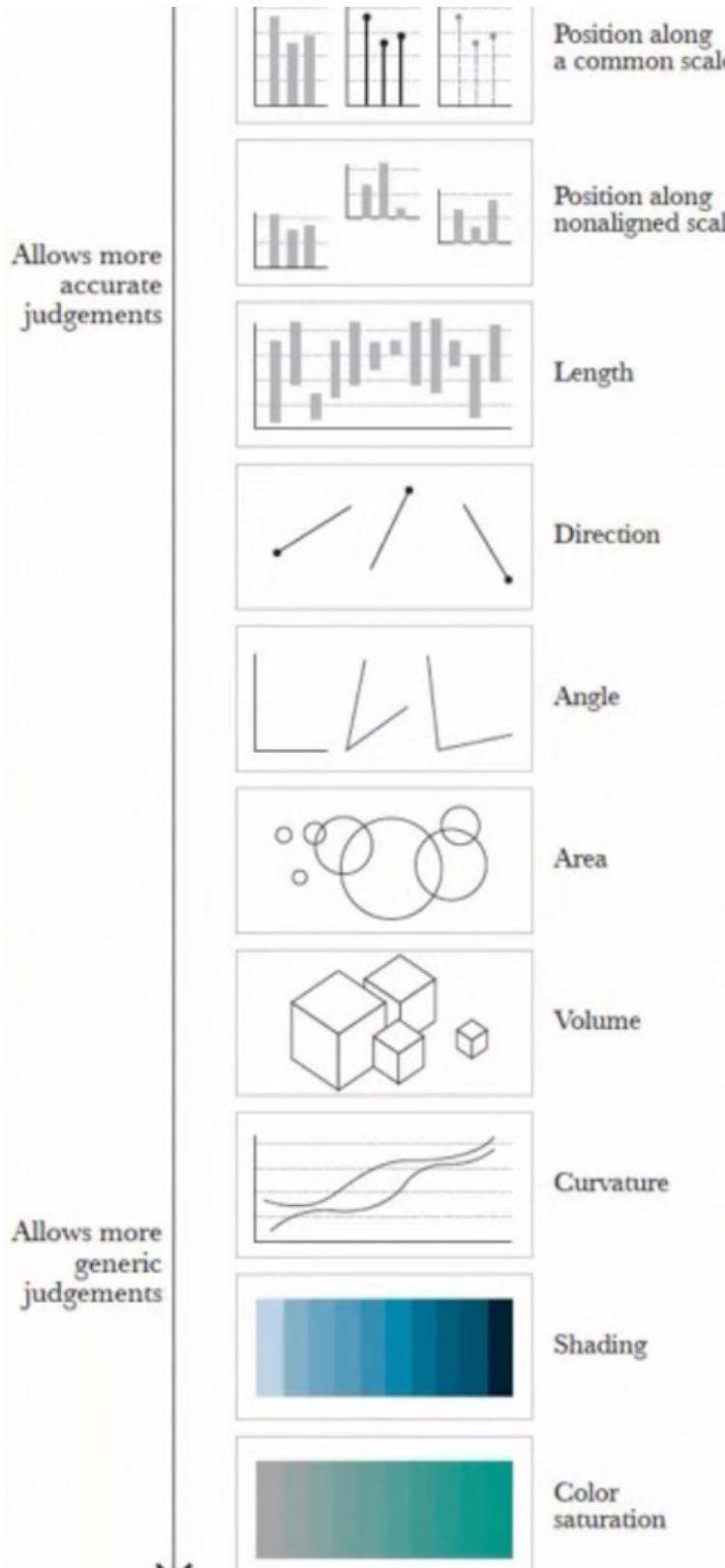


**What's wrong
with this
graph?**

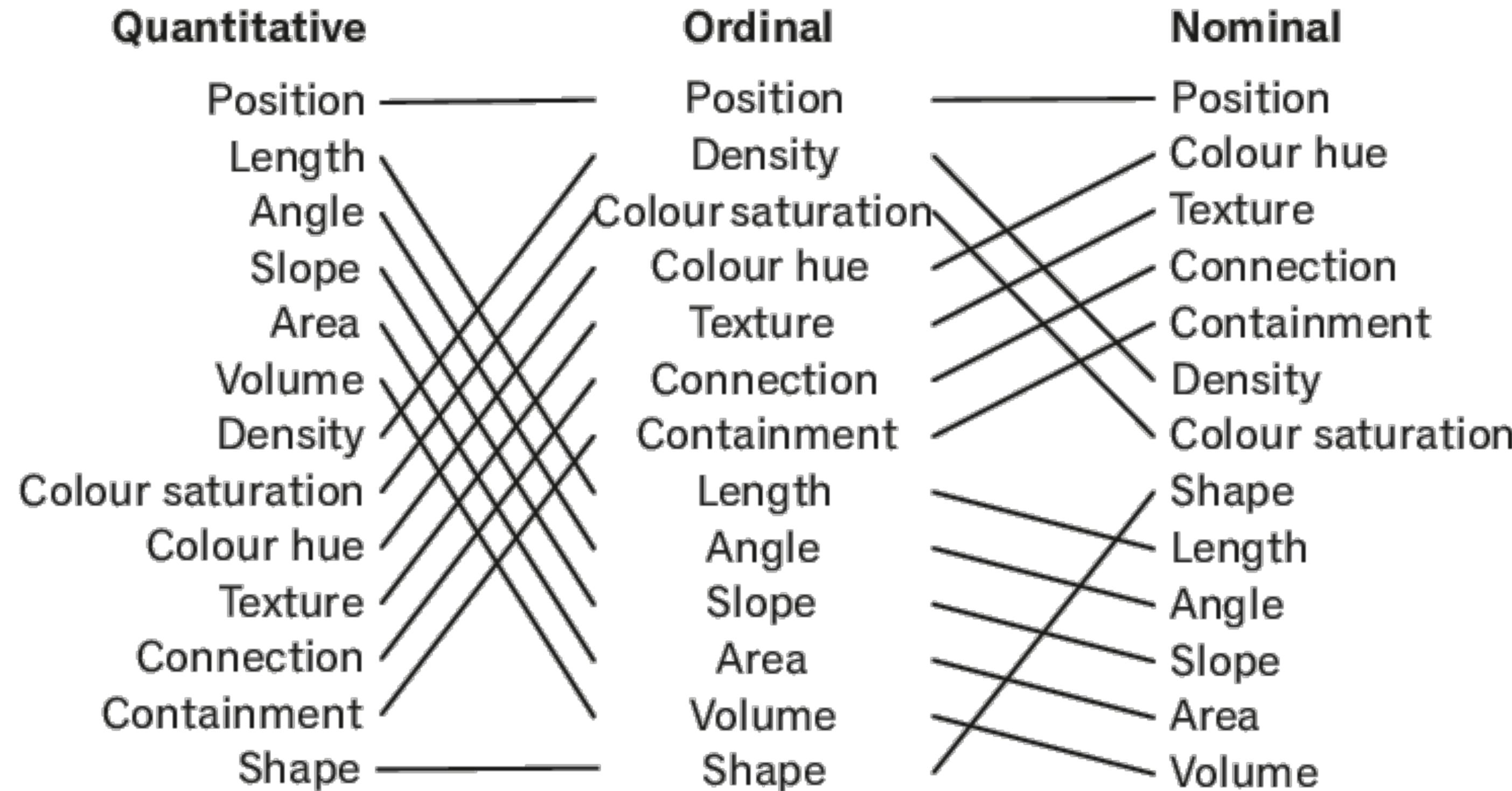


Source: Information Visualization Marks & Channels Presentation, Munzner &
<https://twitter.com/ChaseThomason/status/1118478036507164672?s=19>

4. Select Visual Elements



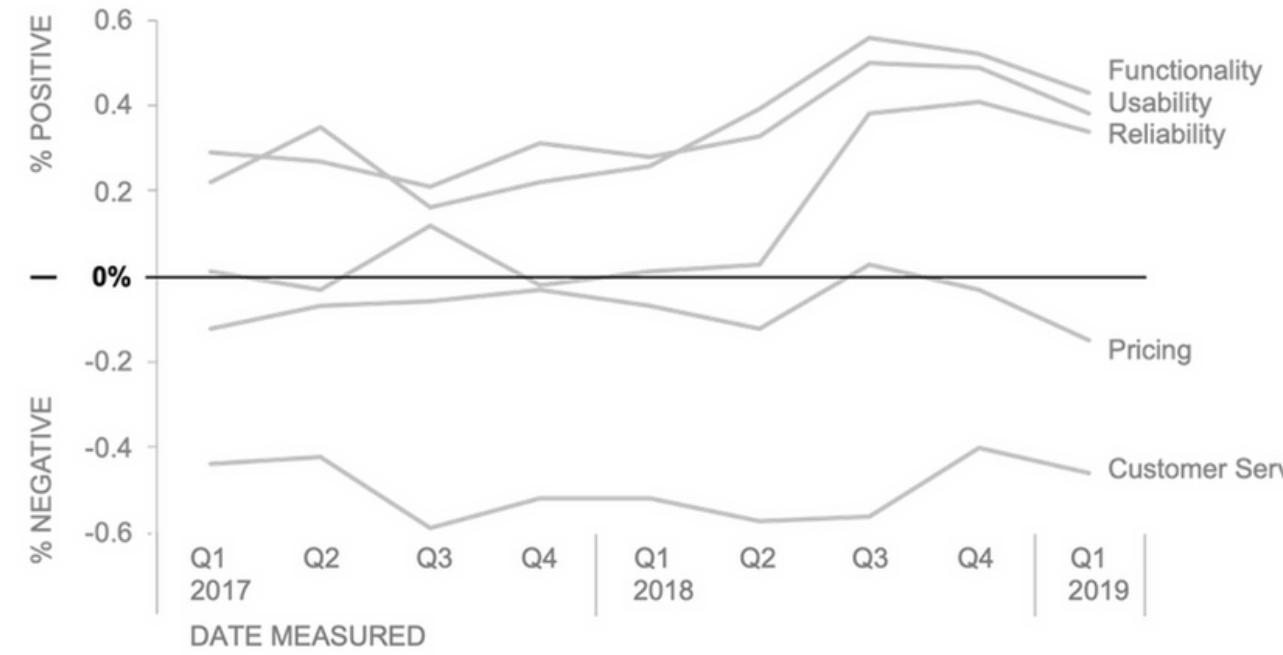
Perception of graphical elements (Cleveland & McGill, 1984, P532)



The Mackinlay ranking of perceptual task

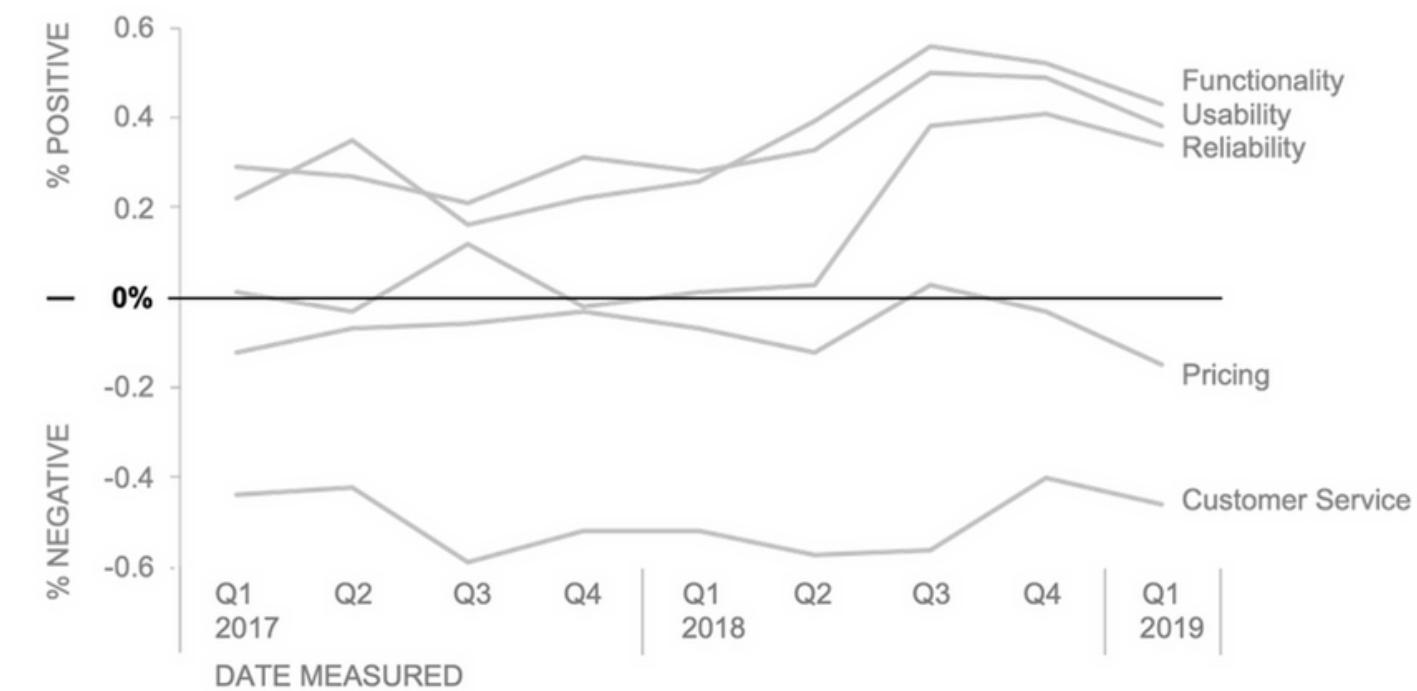
Action needed to address **recent decline**

Customer topic sentiment



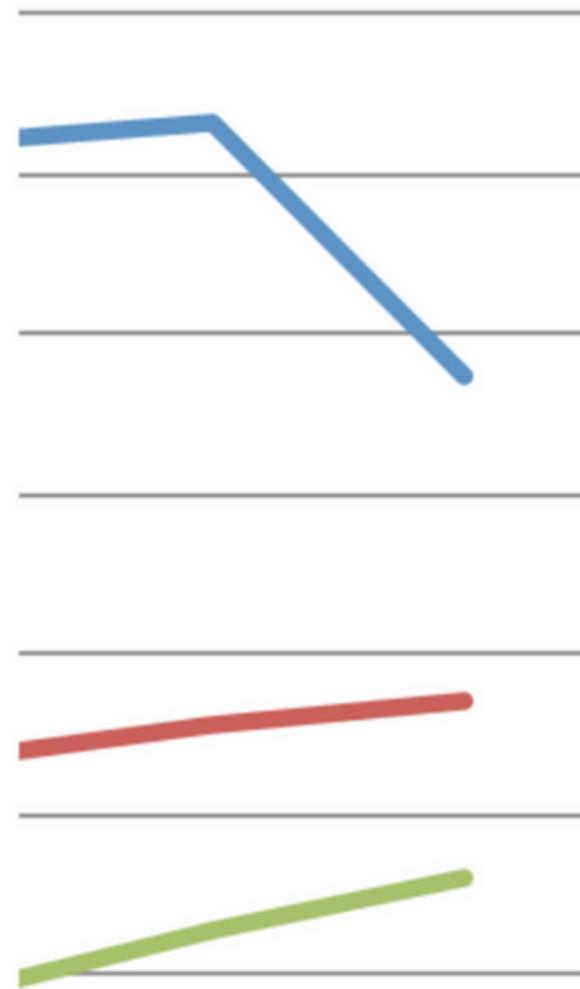
Success: efforts to increase **reliability** worked!

Customer topic sentiment

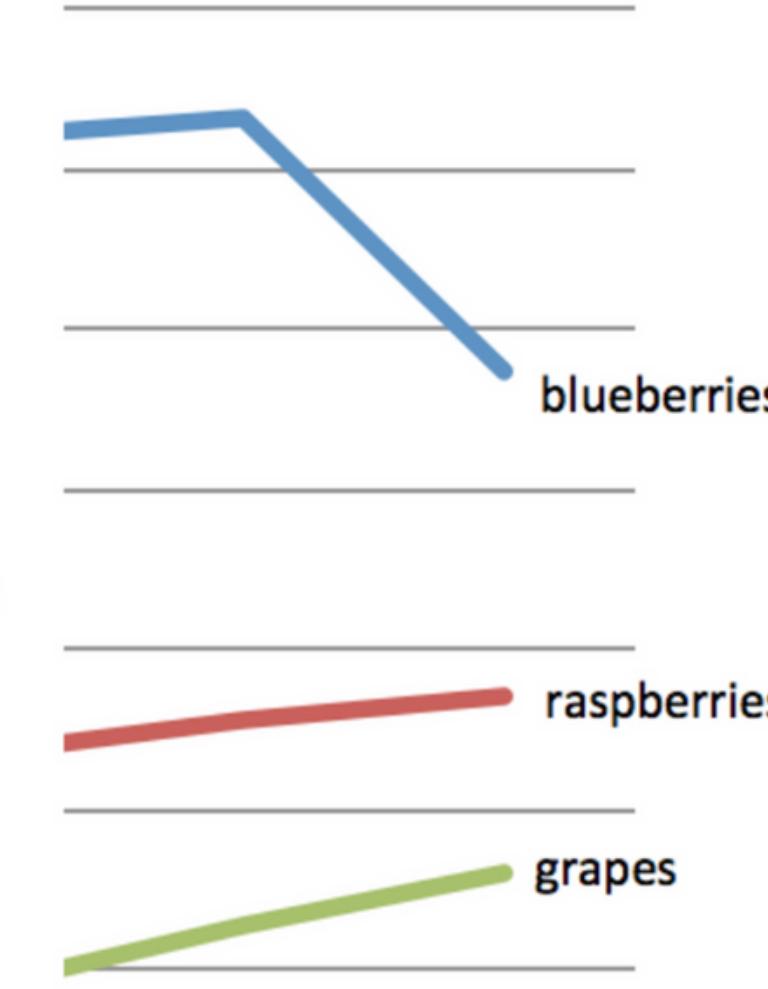


<https://www.storytellingwithdata.com/blog/2018/6/26/accessible-data-viz-is-better-data-viz>

Legend



Direct labeling



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



```

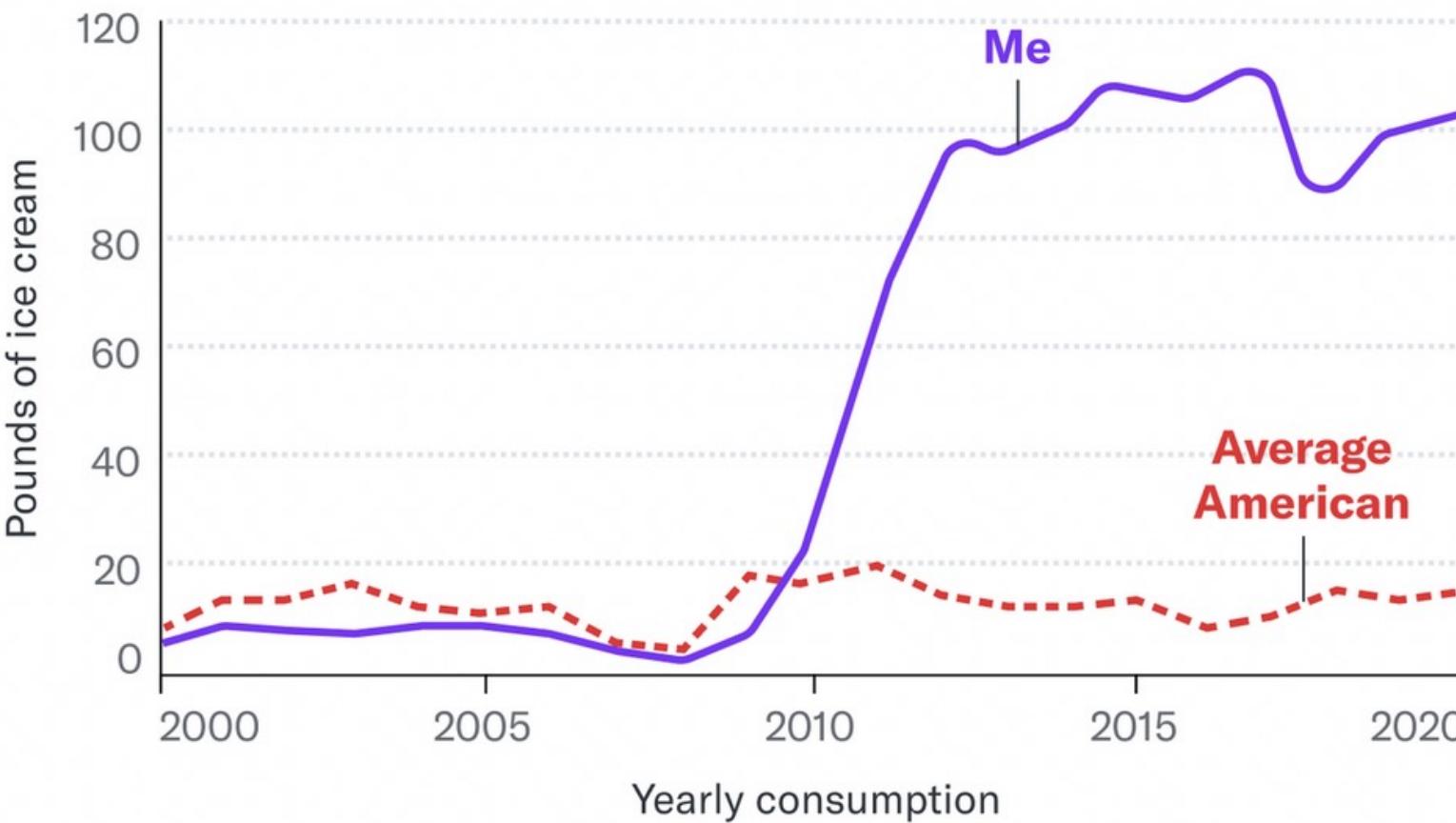
```

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



My yearly ice cream consumption has bested the national average since 2010*

*While the spirit rings true, this statistic is entirely made up

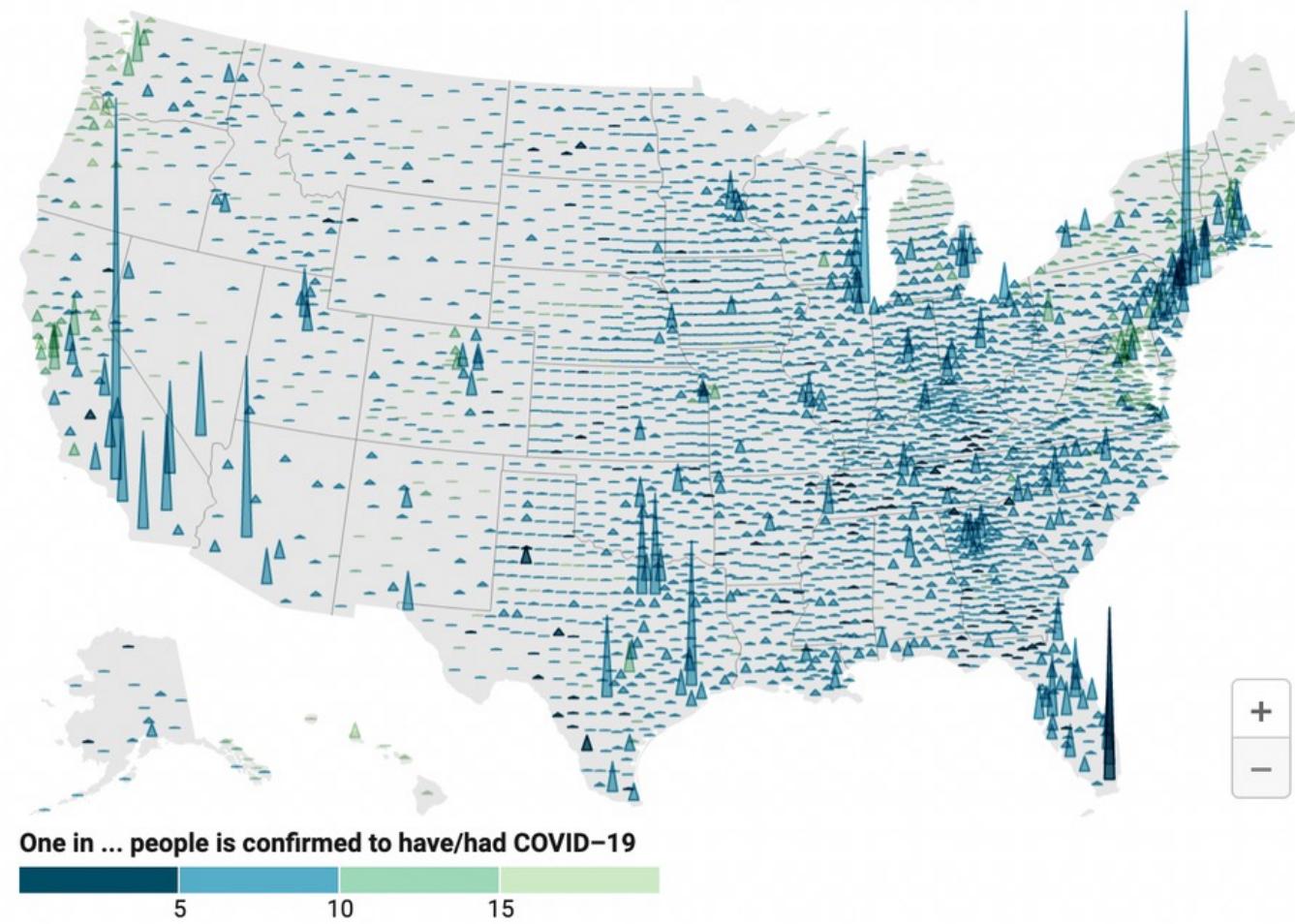


SUMMARY OF RESULTS

Since 2010, I've consumed an average of **100 lbs** of ice cream per year.
The average American has consumed only **12.7 lbs**.
This is nearly **8x** more ice cream. Oh no.



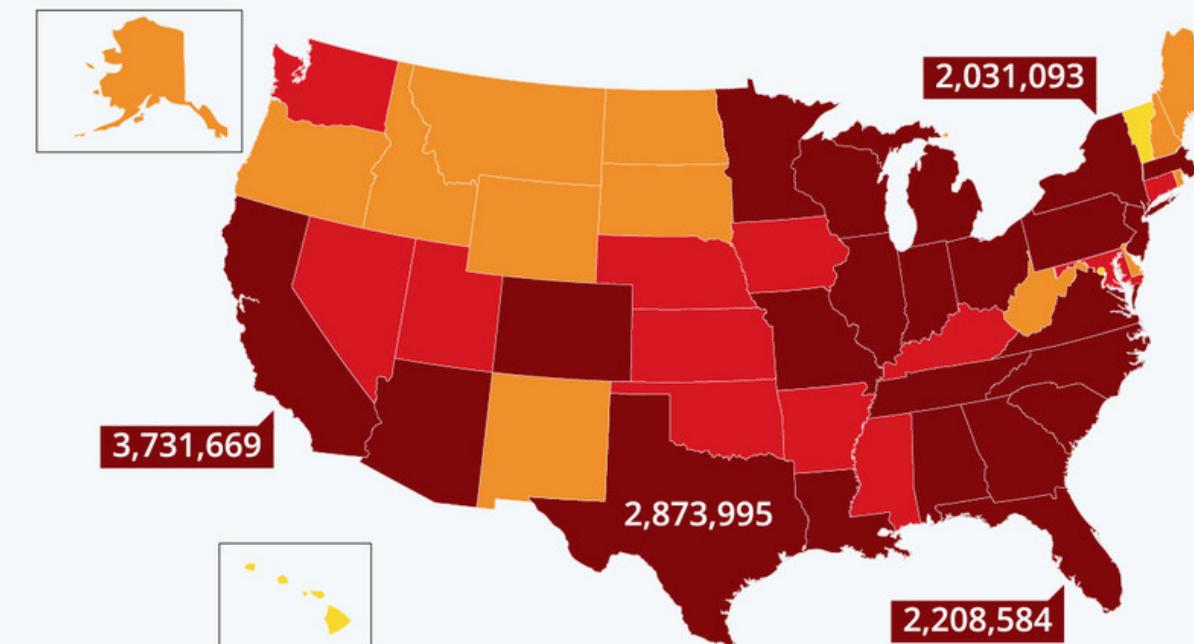
Total confirmed COVID-19 cases in US counties



Confirmed COVID-19 Cases in the U.S.

Number of confirmed COVID-19 cases, by U.S. state*

0-49,999 50,000-199,999 200,000-449,999 450,000+



* as of April 26 at 1:30 AM EDT

Source: Johns Hopkins University



statista

5. Share for Interpretation and Receive Feedback

01

Would a user be able to understand the basics in 15 seconds?

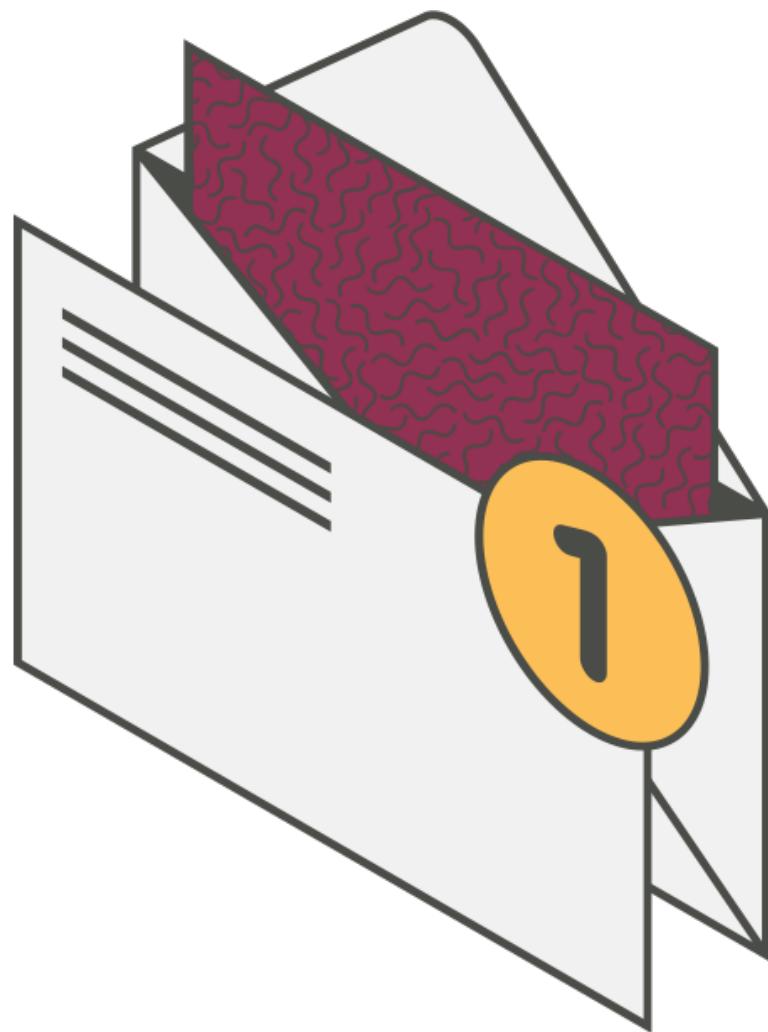
02

Is this visualization honest about what isn't represented?

03

Have I properly attributed the work?

Questions?



If you need additional help or information, please get in touch!
sivajos@mcmaster.ca

