Data Visualization: Choosing the Right Visualization

```
$ echo "Data Sciences Institute"
```

We're going to...

- Explore how to choose the right data visualization for a given situation
- Explore Chapter 3 (On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints) of D'Ignazio and Klein (2020). Data Feminism. MIT Press.
- Discuss how ideas of neutrality and objectivity apply to data visualization
- Understand how different elements and types of data visualization are generally perceived, and use this understanding to decide what kind of visualization we should use for a particular situation

- So far, we have learned how to make and modify different types of data visualizations

- How do we decide which of these types of data visualization to use, and when?

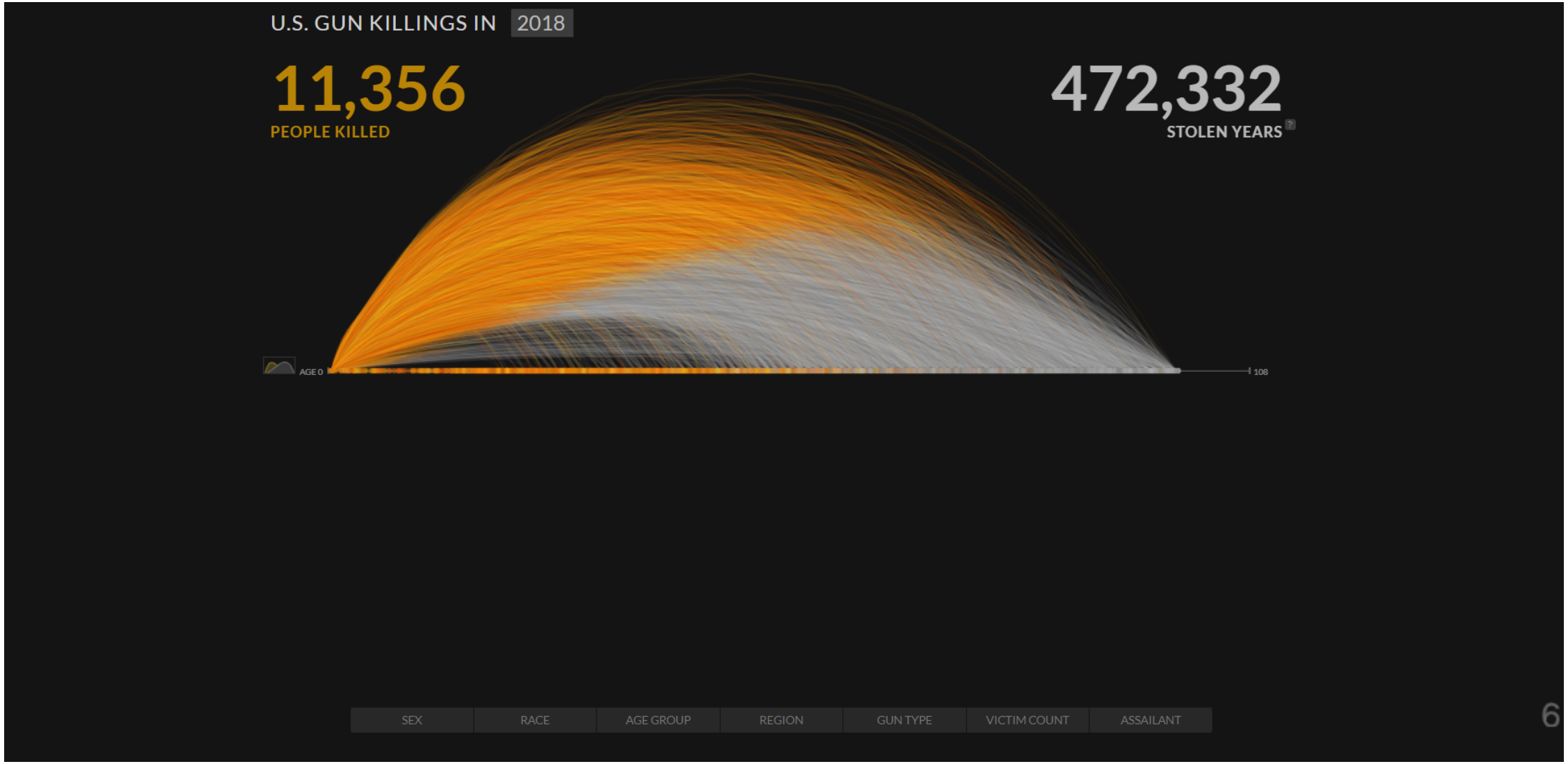
- If we are accurately and honestly displaying our data, does the type of visualization even matter?

How do we choose the 'right' visualization?

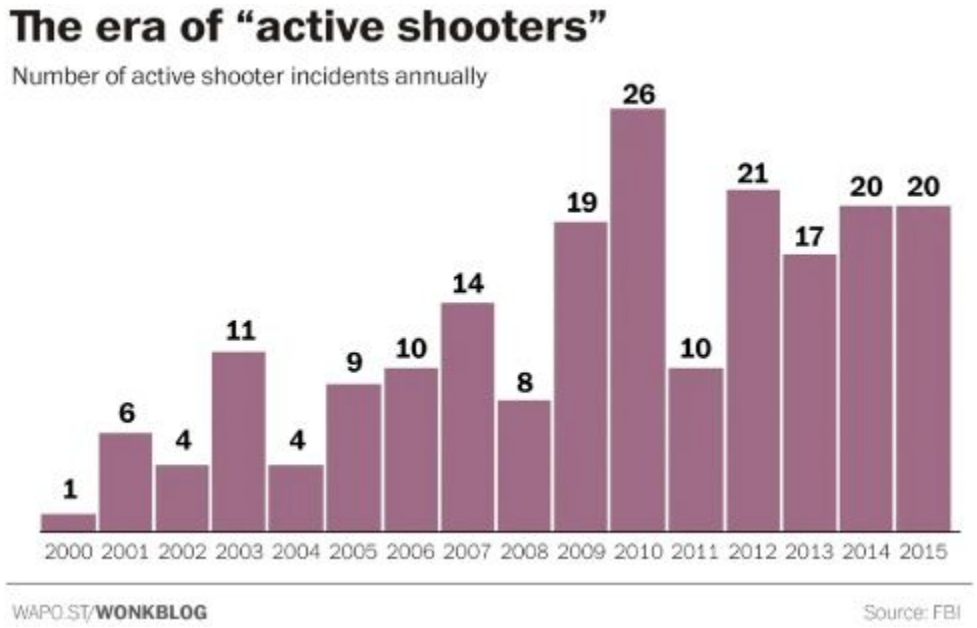
Activity

- We will explore two data visualizations, each showing similar datasets with different techniques
- For each visualization, discuss the following questions:
  - What information can we learn from this visualization?
  - Is this an example of objective, neutral data visualization? Why or why not?

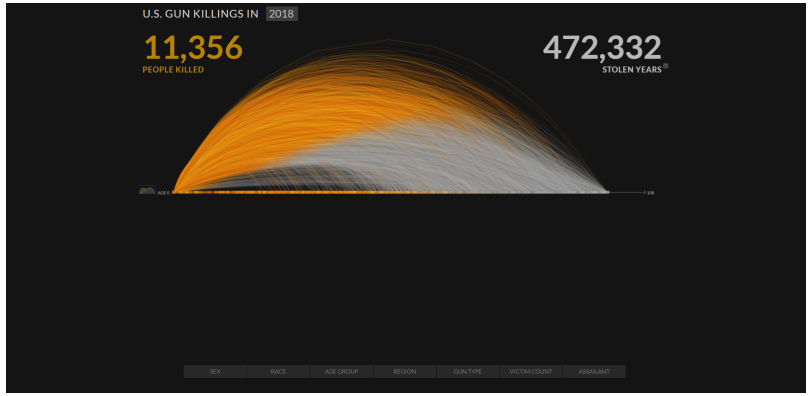
Visualization #1: US Gun Killings in 2018



Visualization #2: Washington Post Active Shooters

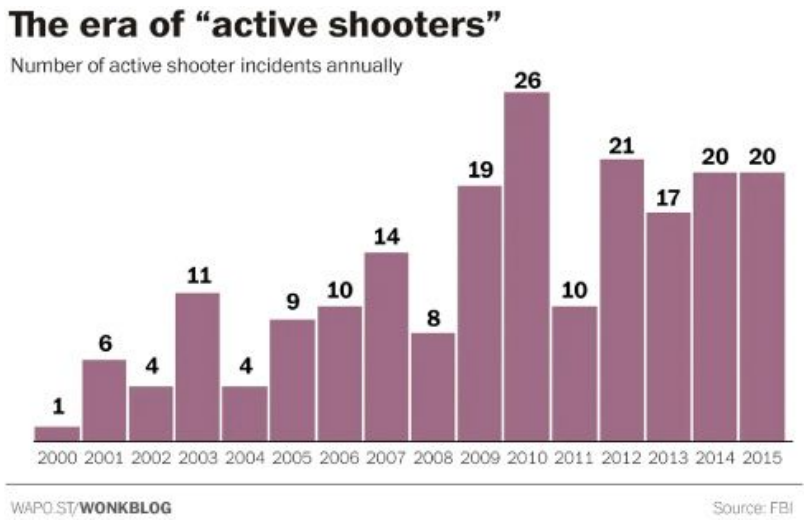


Visualization #1: US Gun Killings in 2018



- Periscopic's animated visualization shows the expected years of life lost to gun violence in the United States in 2018
- It emphasizes an emotion: a sense of loss
- This visualization has been criticized as "actively [shaping] data to support a cause" (in this case, highlighting a lack of gun control in the United States)

Visualization #2: Washington Post Active Shooters

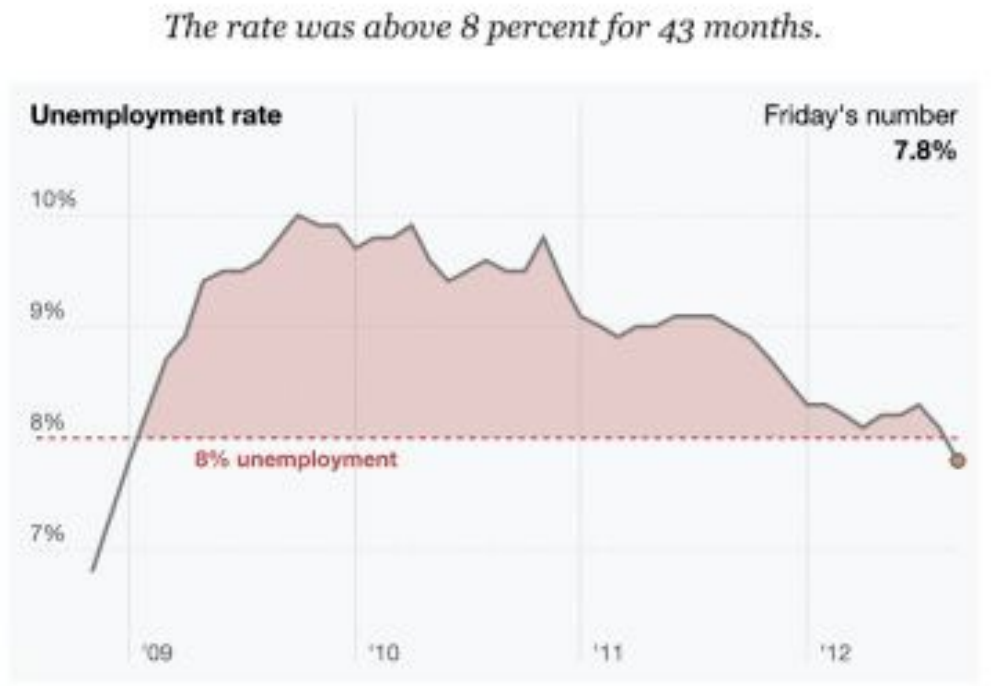


- The next visualization shows a related dataset (about gun violence in the United States)
- Viewers will likely reach a similar conclusion as in Visualization #1, but this plot is intended to present “a deliberately neutral emotional field, a blank page in effect, upon which viewers are more free to choose their own response to the information”

What qualities or visual elements of Visualization #2 help to make it a “blank page”?

A blank page

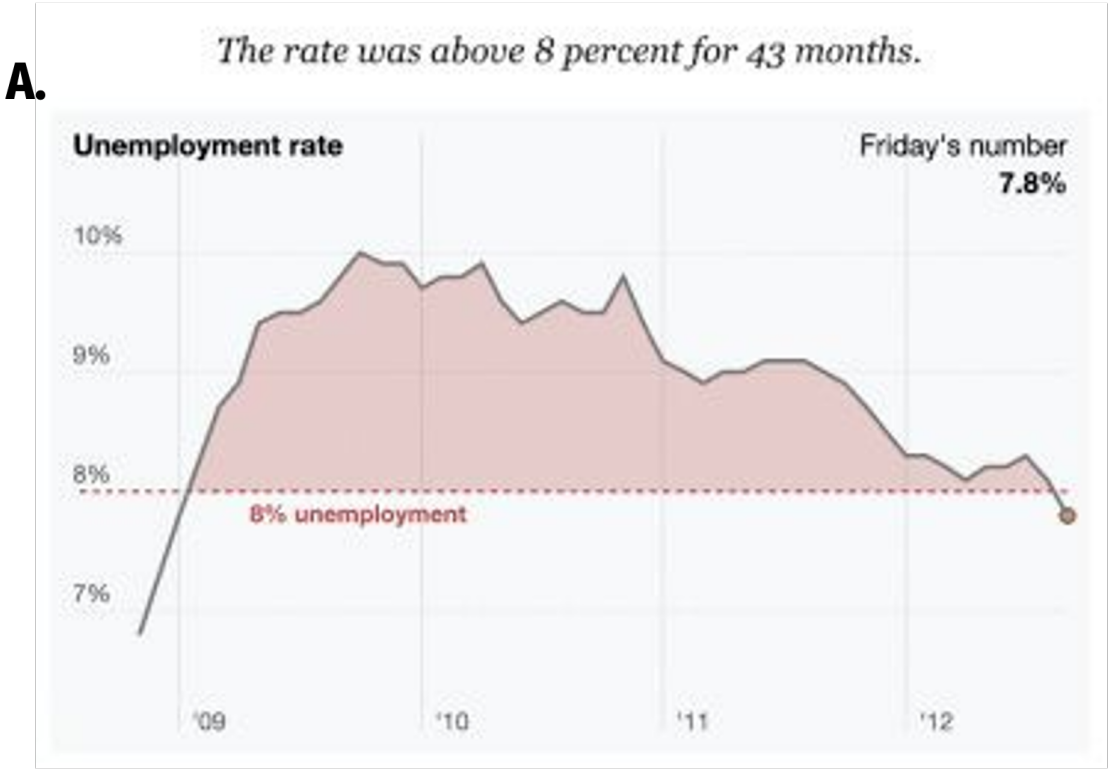
- Some of the same design elements from our ‘blank page’ Visualization #2 can be seen in this New York Times visualization of the September 2012 Jobs Report
- The clean, 2D layout is designed to avoid conveying an emotional narrative to the audience



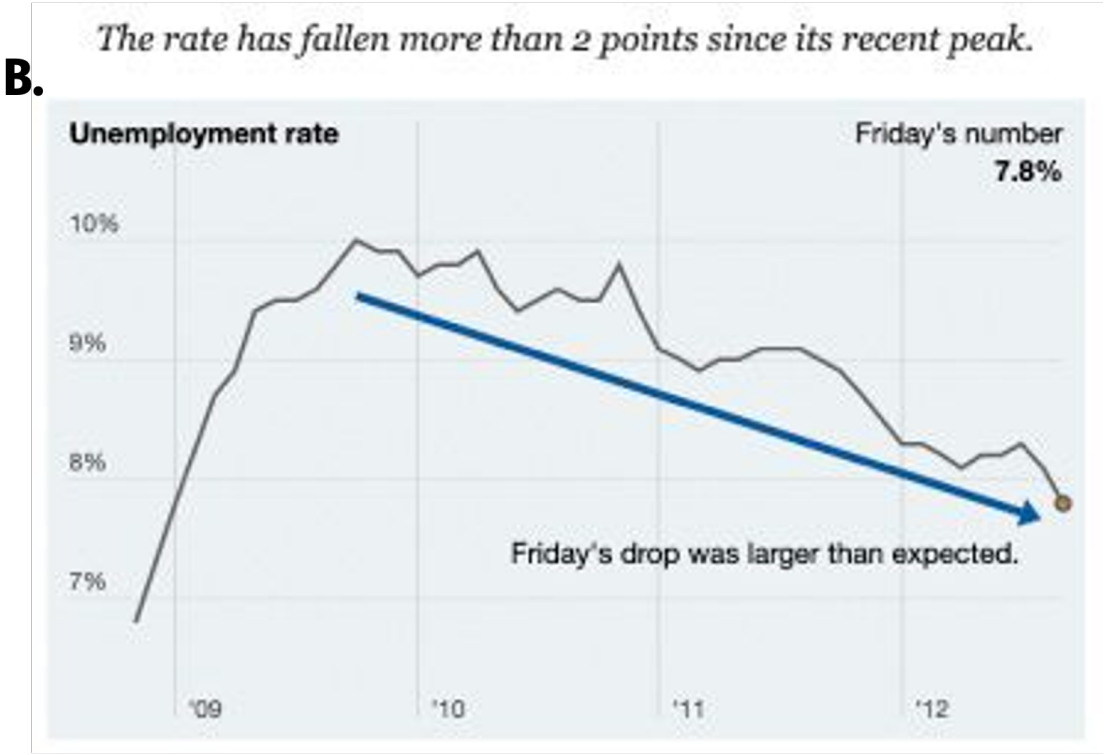
A blank page...right?

- The Jobs Report graphic was published alongside another image:

A.



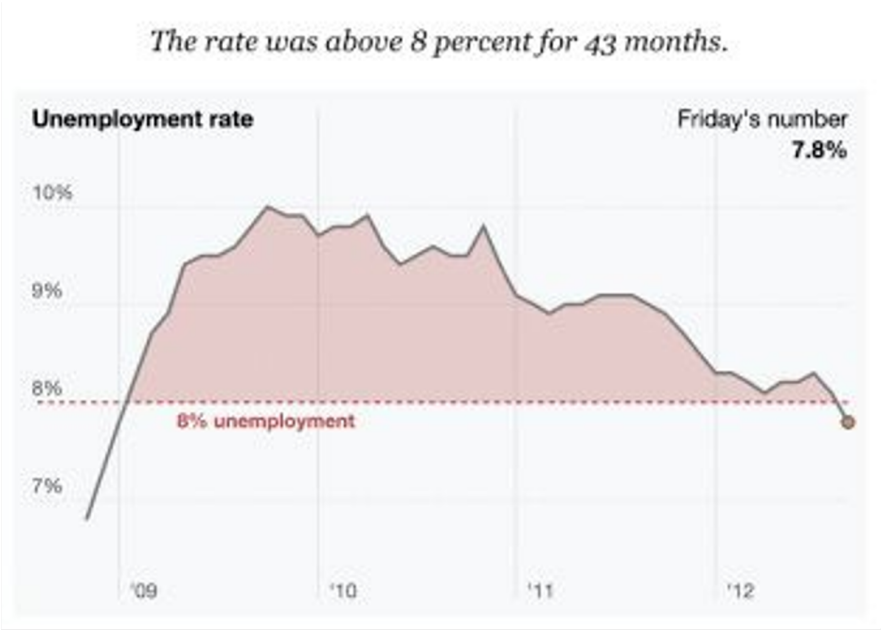
B.



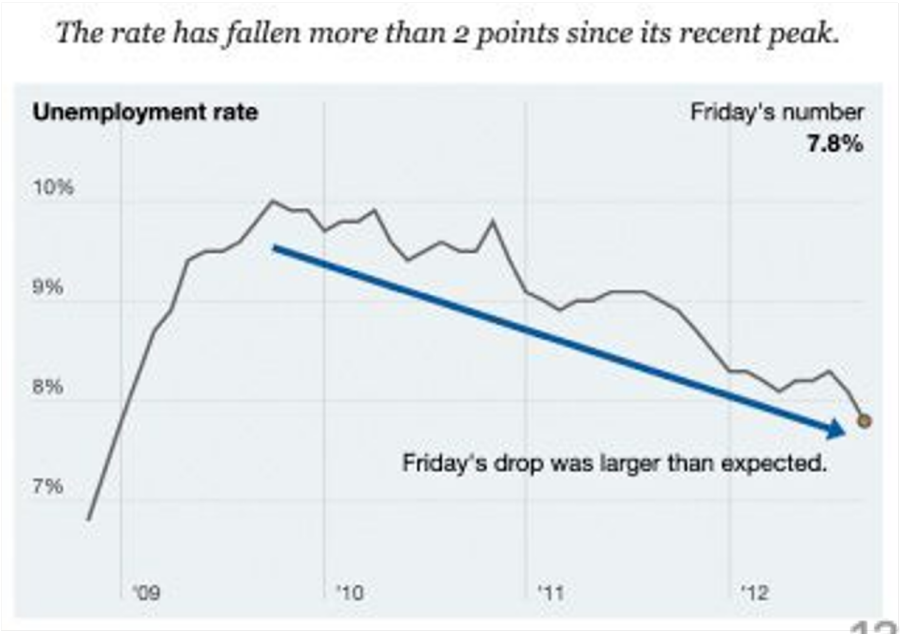
A blank page

- Images A and B were designed to show the exact same data from the perspectives of Republicans and Democrats, respectively
- Image A emphasizes the unemployment rate staying above 8%, while Image B emphasizes the rate's decline
- **Neither is technically dishonest!**

A.

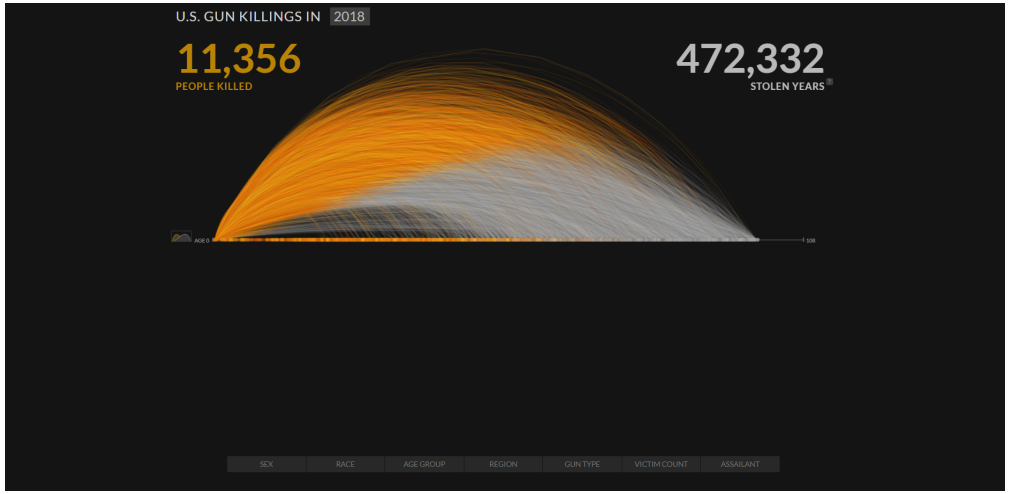


B.



Supporting a cause

- Periscopic's Gun Killings visualization won several year-end information visualization awards
- This visualization is not neutral, and conveys an emotional narrative to the audience…
- **..But this visualization is not dishonest either!**



So where does this leave us in our search for neutral, objective data visualization?

Can data visualization be neutral?

Short answer:

NO!

“The constraints of truth leave a very wide space for interpretation…”

(Stray, 2016)

Data visualizations as rhetorical objects

- **Rhetoric** is the act of communicating effectively and persuasively
- From D'Ignazio and Klein (2020),

> "Any communicating object that reflects choices about the selection and representation of reality is a rhetorical object. Whether or not it is rhetorical (it always is) has nothing to do with whether or not it is true (it may or may not be)."

- That is, **we make choices about how to visualize our data, so these visualizations are not neutral...**
- **..BUT data visualizations can be factual without being neutral**

Data visualization as an interpretative, rhetorical act is not necessarily a bad thing, but one that we should be aware of.

Recall

- Three important qualities of data visualization:
  - Is the visualization pleasing to look at? → **Aesthetic**
  - Does the visualization accurately and honestly present data? → **Substantive**
  - Can we understand what message the maker of the visualization is attempting to convey? → **Perceptual**

- Two data visualizations can share the same substantive qualities while, intentionally or not, being perceived completely differently
- When we are aware of the choices we make while creating data visualizations, we can design data visualizations that are suited to the situation at hand (perceptual qualities) without sacrificing honesty and accuracy (substantive qualities)

What do we want our data visualization to do?

Intended purpose

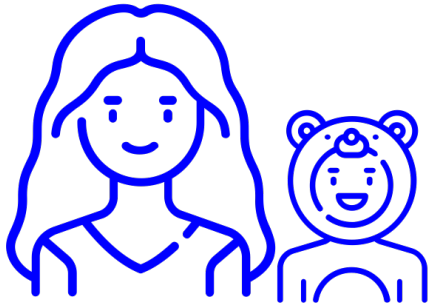
Persuading

Comparison

Evaluating

Exploring

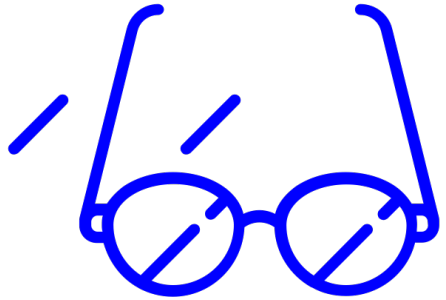
Intended audience



Age

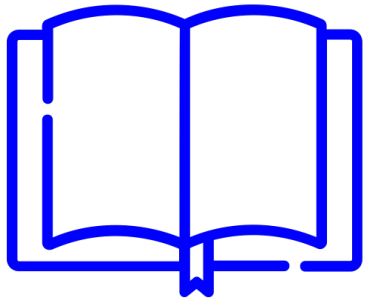
Education

Expertise



Accessibility

Intended medium



Print

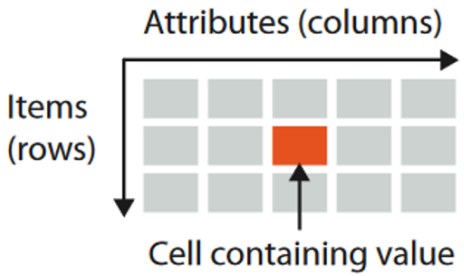
Web

Poster

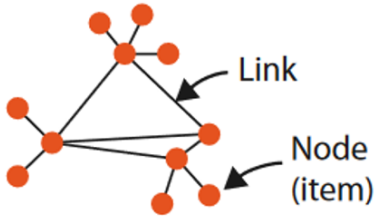
Presentation

Dataset Types

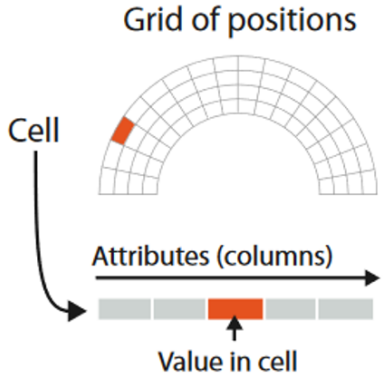
Tables



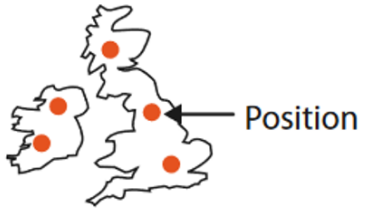
Networks



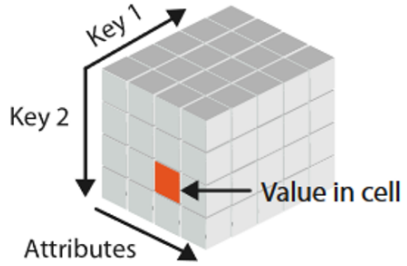
Fields (Continuous)



Geometry (Spatial)



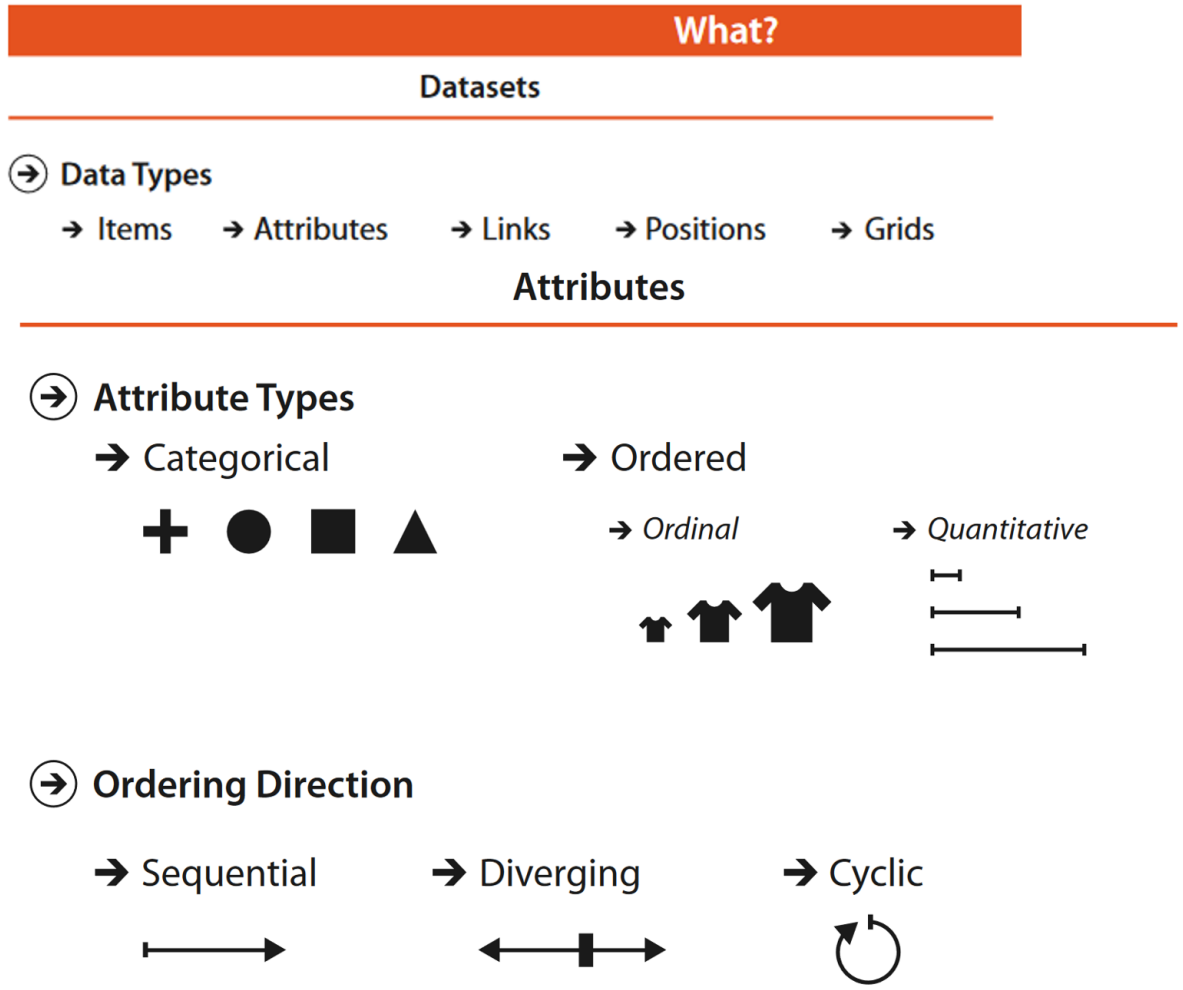
Multidimensional Table



Trees

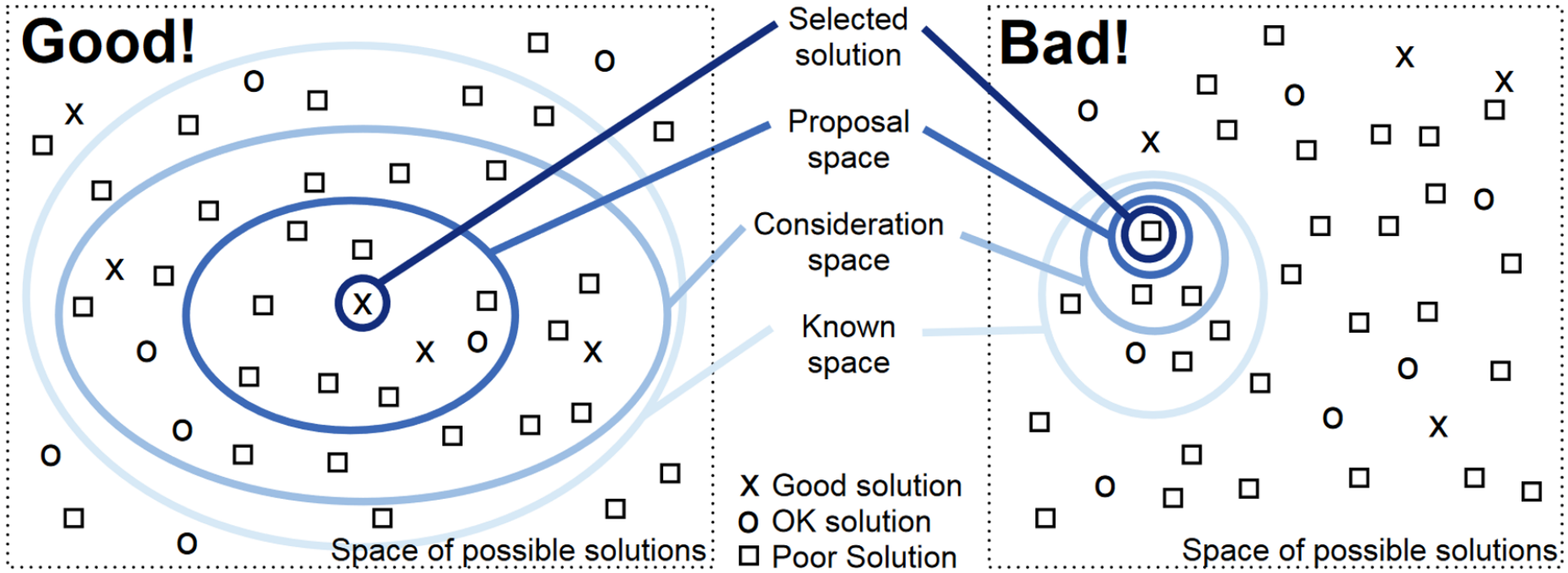
- Tables, networks, fields, and geometries are four basic dataset types.
- For any of these dataset types, the full dataset could be available immediately as a static file or it might be dynamic data processed gradually in the form of a stream.

- Identifying the attribute types for our data will help us make choices about how to visualize it.

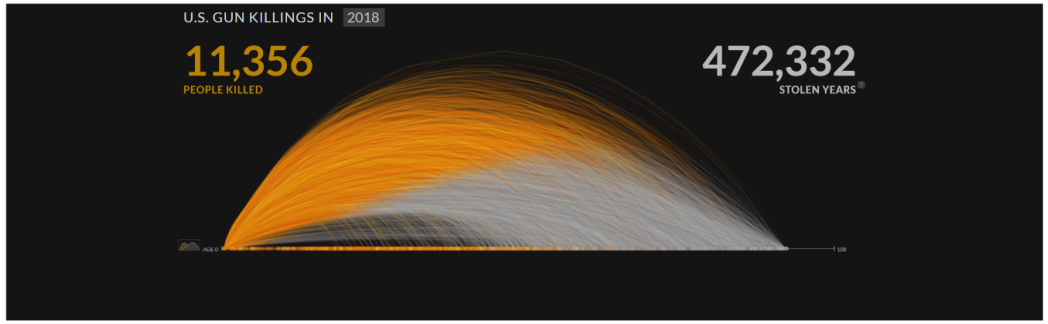


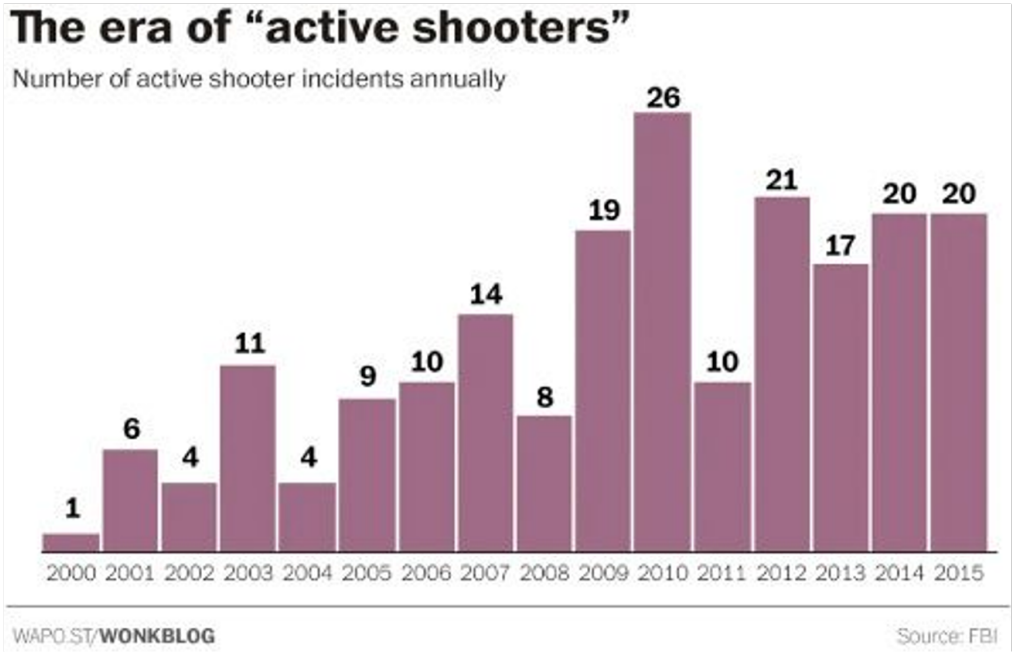
Effective Visualization

- Depends on purpose, audience, and medium!
- A good goal is to satisfy: “to find one of the many possible good solutions rather than one of the even larger number of bad ones”



Different purpose, different results



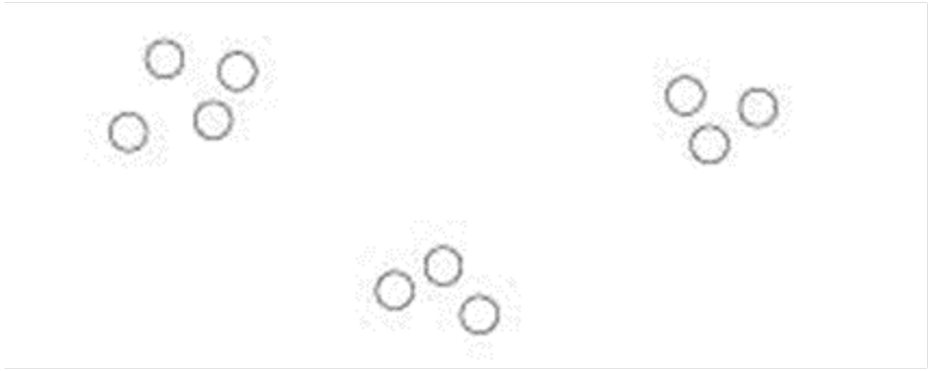


How is our data visualization perceived?

Taking advantage of cognitive psychology

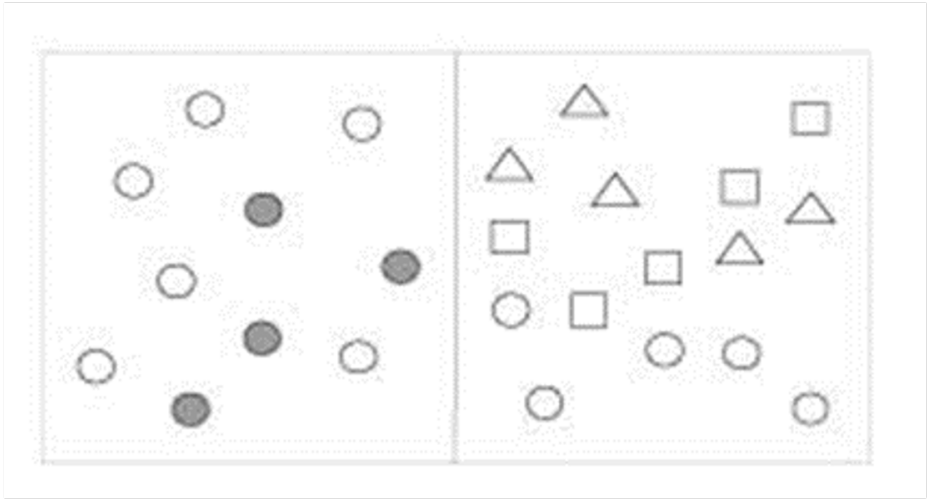
- In general, data visualization takes advantage of human cognition to help us understand data more intuitively than we can if it is presented to us as a list or a table (Li, 2020)
- By learning about how humans tend to process visual information, we can communicate more effectively with our graphs. For example...
- **Gestalt principles** (Gestalt is German for shape) are a set of cognitive theories for how people tend to organize visual information; and are commonly used in UX design and data visualization (Wong, 2010)

Gestalt principles



Proximity

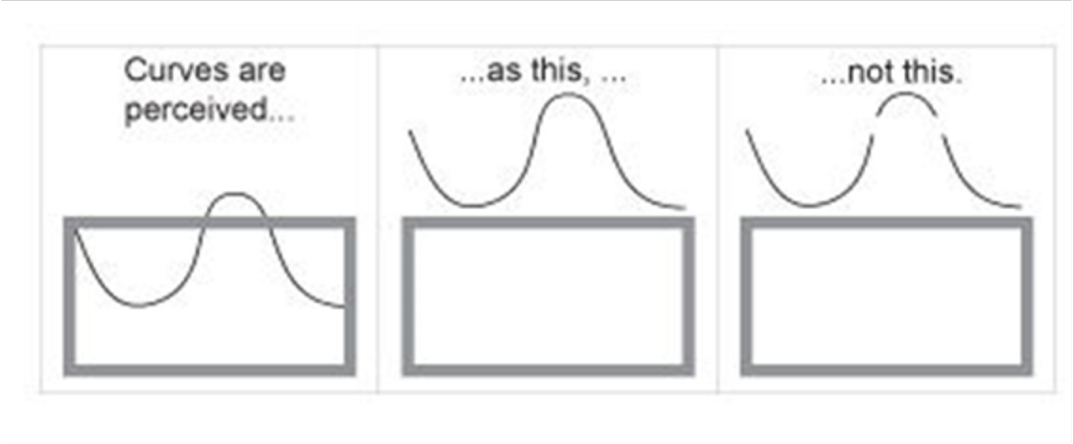
Objects that are close together are perceived as belonging to a group



Similarity

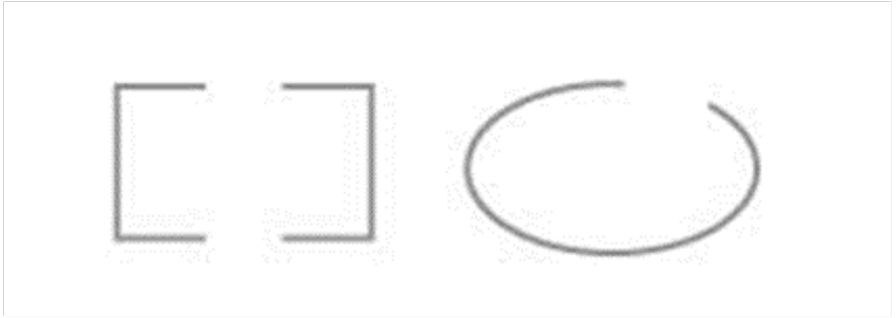
Similar objects are grouped, regardless of proximity

Gestalt principles



Continuity

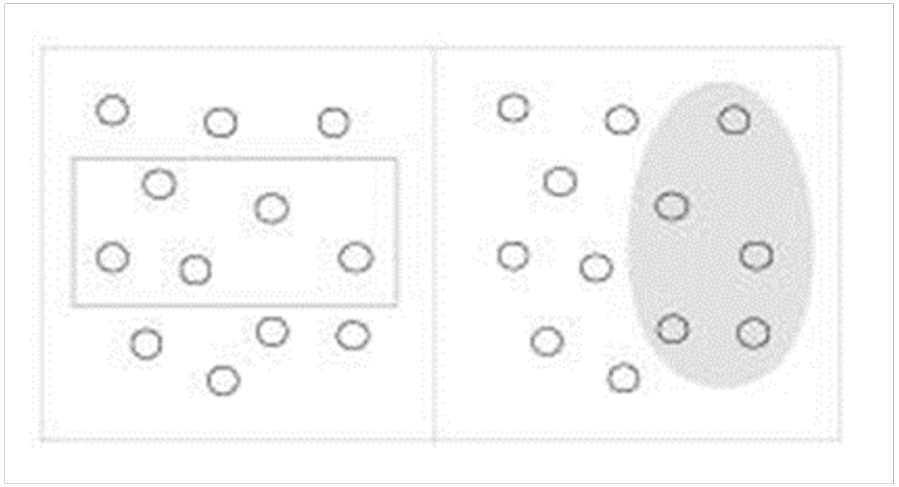
Aligned objects or objects that appear to continue are perceived as a group



Closure

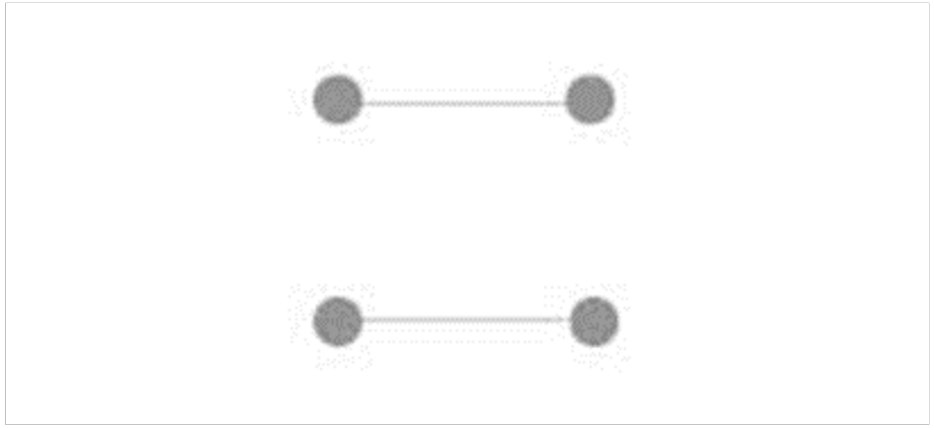
Open structures are perceived as closed/complete (our brains fill in the gaps)

Gestalt principles



Enclosure

Objects with a boundary around them are perceived as a group



Connection

Connected objects are perceived as related/as a group

Cognitive load

- It can also be helpful to consider **cognitive load** , or the amount of work required to take in new information
- Cognitive load can be divided into:
  - **Intrinsic**(the intrinsic complexity of the new information)
  - **Germane**(the audience's familiarity with the information)
  - **Extraneous**(complexity from how the information is presented)
- In a data visualization context, extraneous cognitive load is most within our control

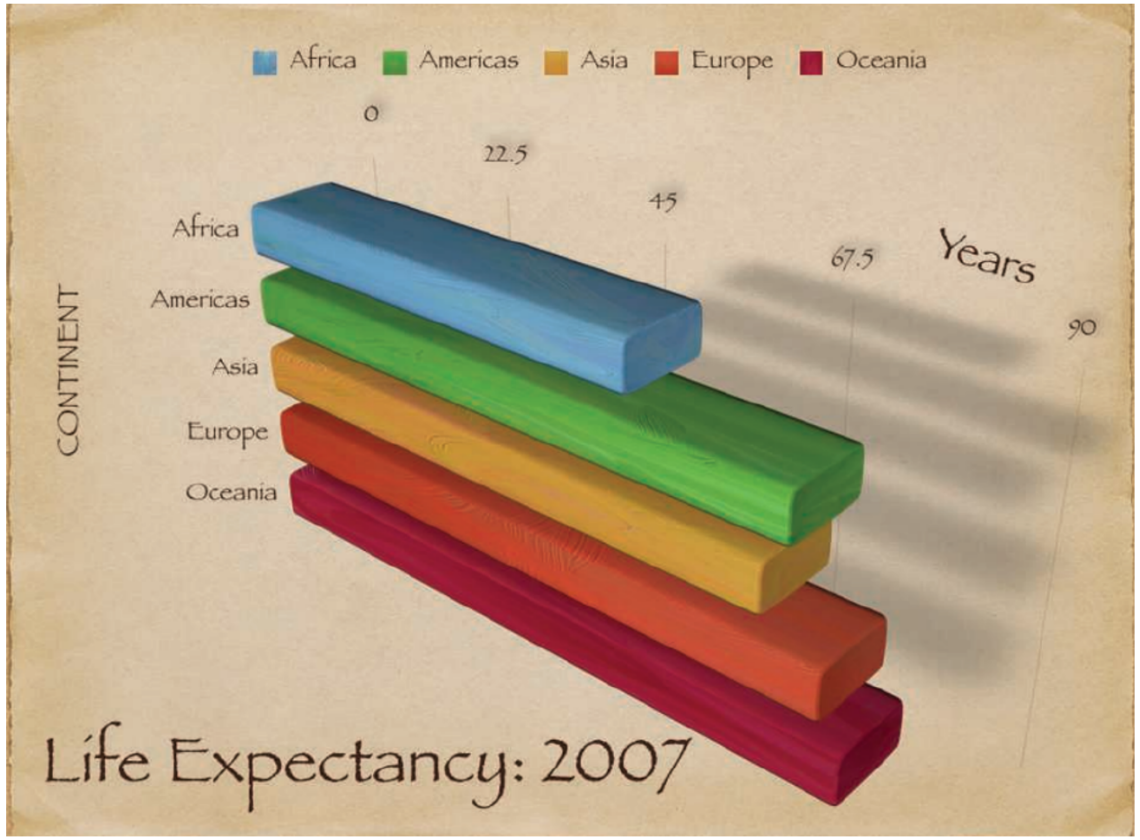
Cognitive load

- Elements of a visualization that can affect cognitive load include:
  - **Familiar vs. Rare chart types** → rare types increase cognitive load
  - **Accurate vs.Approximate interpretation** → relational values or areas (approximate) increase cognitive load compared to absolute values or position (accurate)
  - **Concise vs. Detailed composition** → more visual elements increases cognitive load
  - **Explanatory vs. Exploratory composition** → a chart that the audience navigates alone increases cognitive load compared to a chart that they are guided through step-by-step

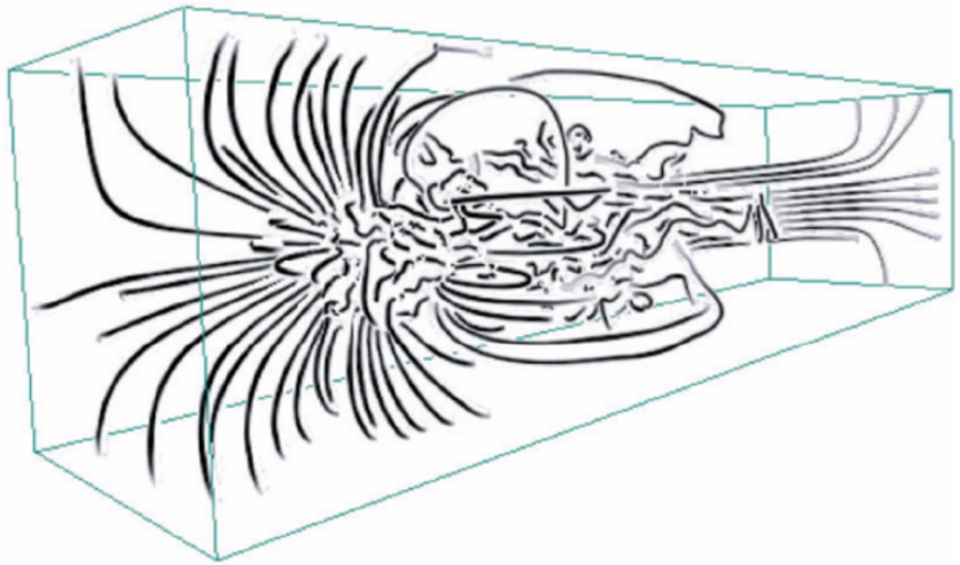
Cognition over Memory

- Working memory is limited!
- It’s better to show side-by-side views than using our memory

No 3D without Cause



VS



Perceived factual basis

- Sociologists Kennedy et al. (2016) find that adherence to **four conventions of data visualization** reinforces the perceived objectivity and factual basis of a visualization:
  - Two-dimensional image
  - Clean layouts
  - Geometric shapes and lines
  - Inclusion of data sources at the bottom of the image

Provenance rhetoric

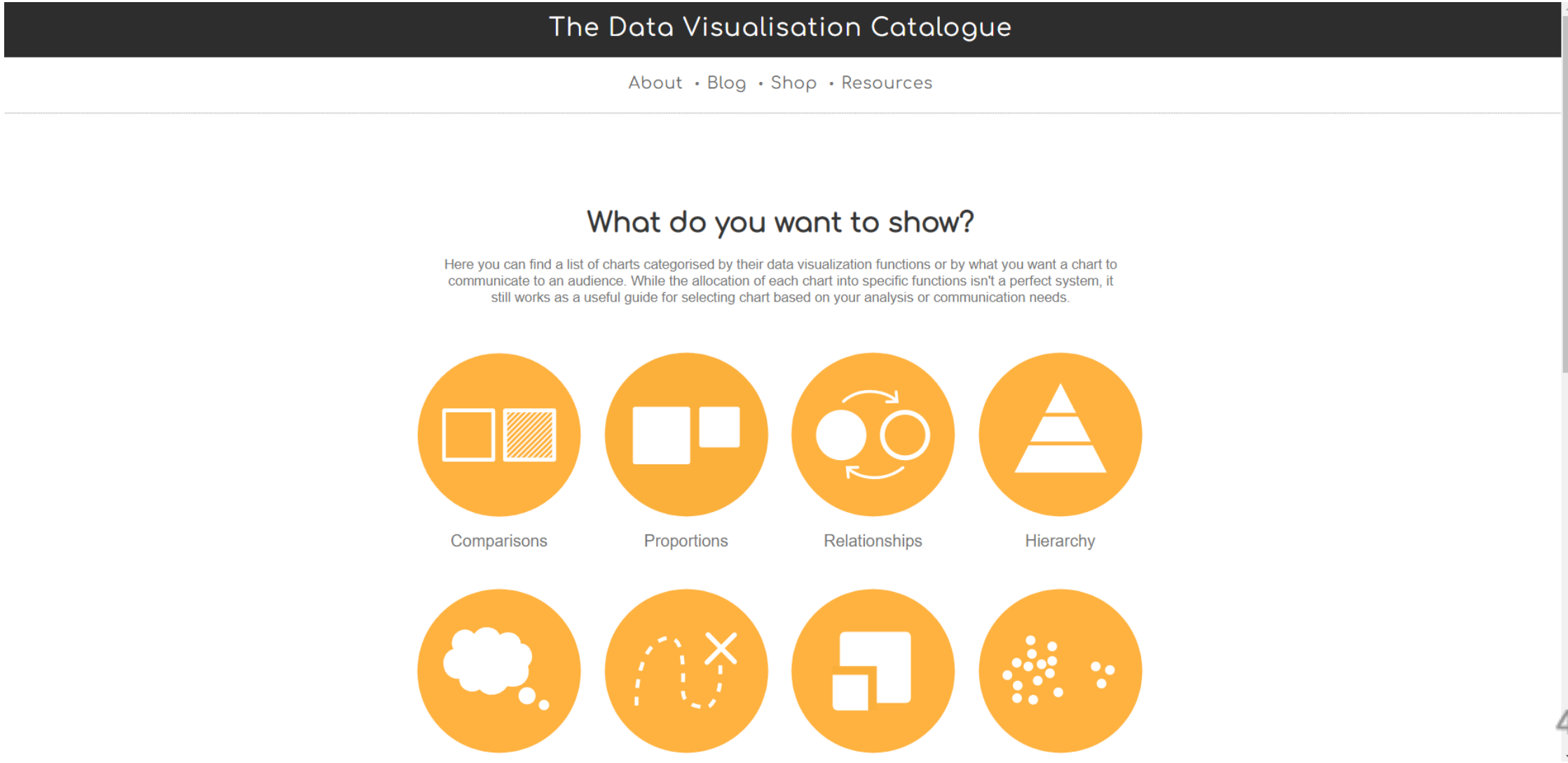
- Citing the source(s) of our data is not only best practice (reproducibility!), but also helps people to trust our data visualizations more
- **Provenance rhetoric** is the idea that the inclusion of a data source with our graphic signals "transparency and trustworthiness" to the audience
- This increases the persuasiveness of the visualization, since viewers are more likely to believe what they see

Resources for choosing data visualization types

Decision making tools

There are resources available online that incorporate visualization purpose and cognitive principles into reference guides to help us decide the most suitable data visualization in a given situation

The Data Visualization Catalogue

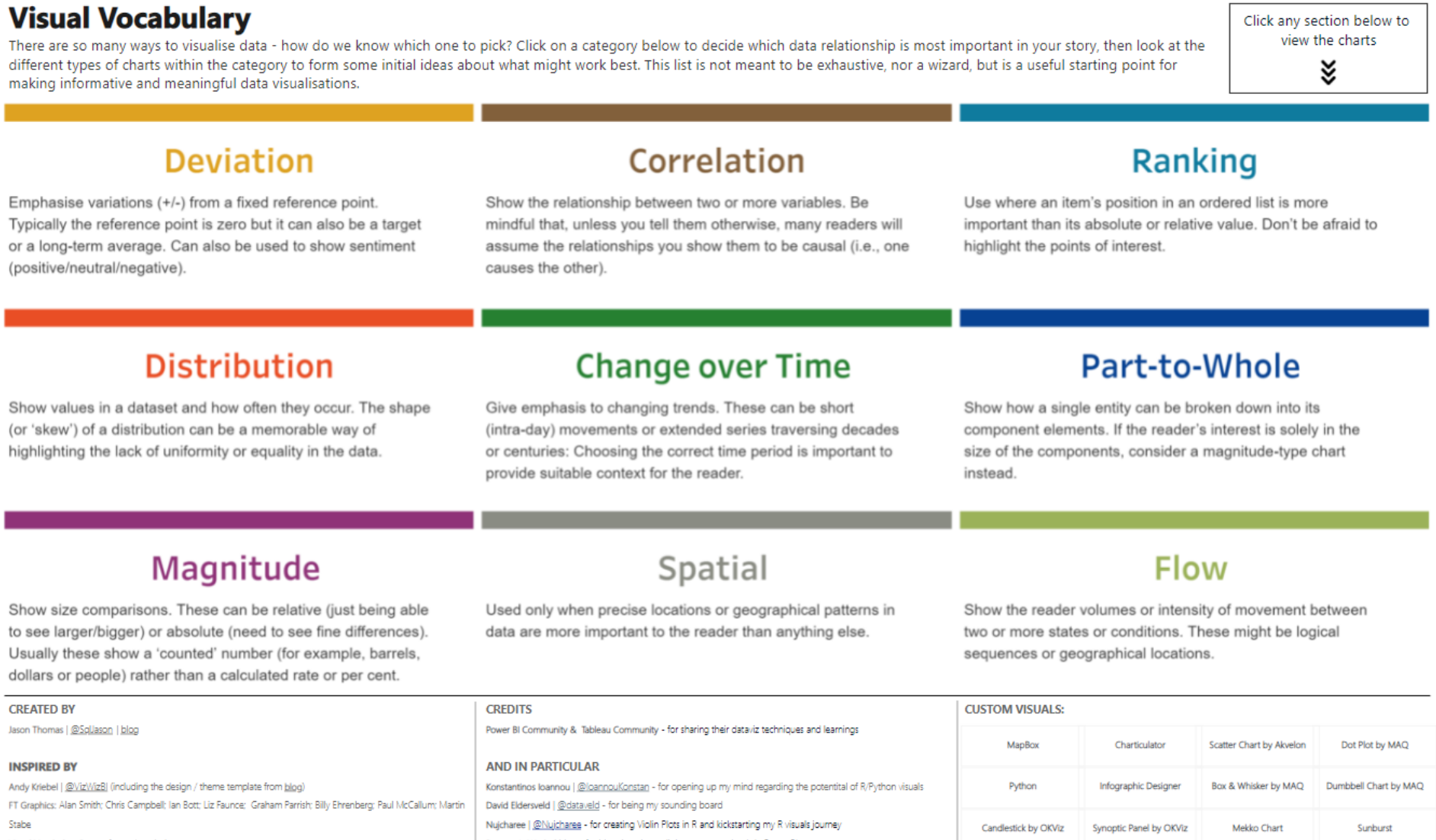


Financial Times Visual Vocabulary

The Financial Times Visual Vocabulary help sheet is available in both interactive (online PowerBI dashboard) and PDF versions

In both forms, the Visual Vocabulary offers a list of potential functions of visualizations, and several corresponding chart types and examples for each

Financial Times Visual Vocabulary



Assignment 2

Next session, we'll discuss:

- More matplotlib!