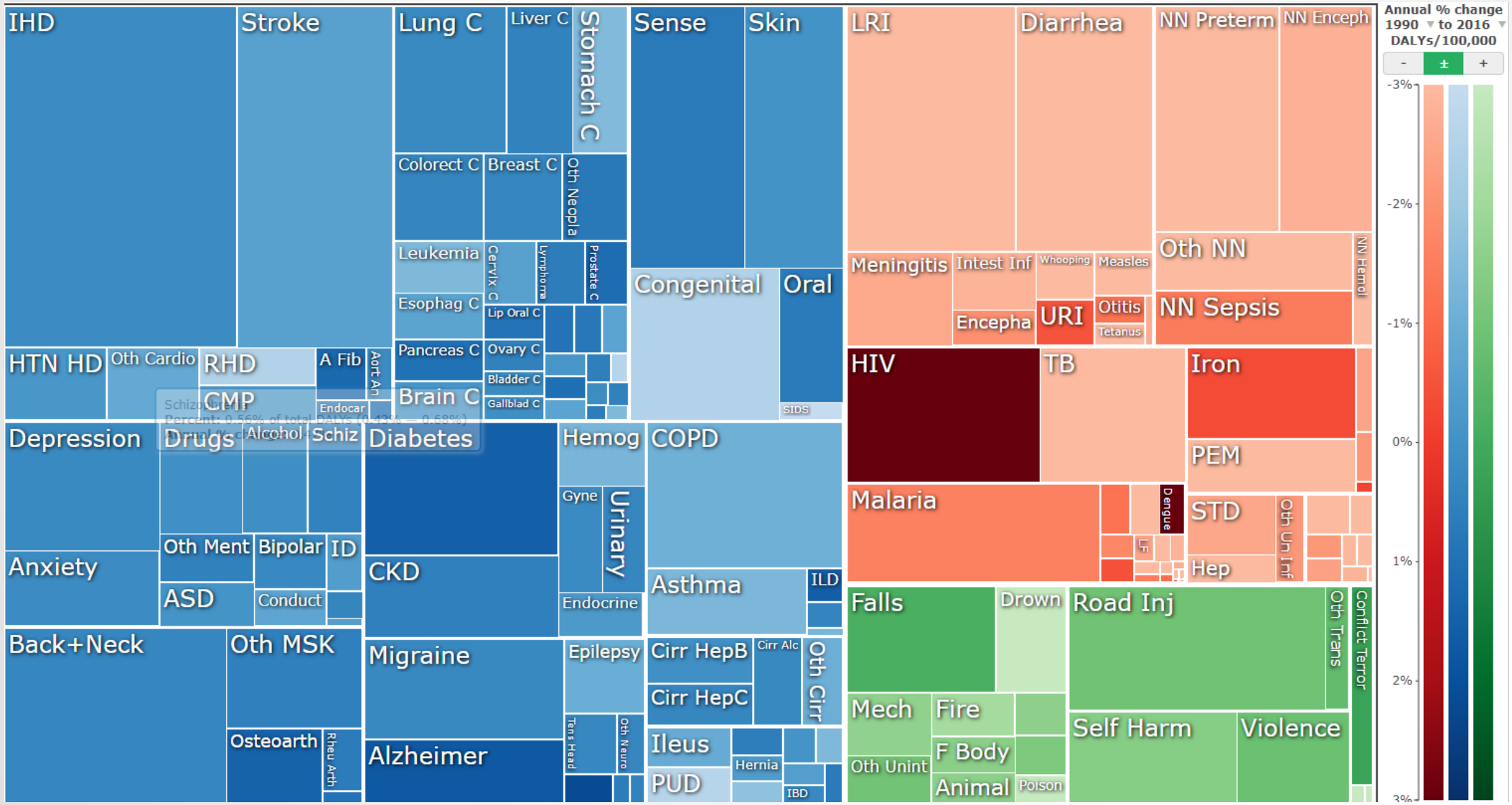
DATA VISUALIZATION: PRINCIPLES AND PRACTICE

Drs. Silas Bergen and Todd Iverson
Winona State University
ICOTS 2018
Kyoto, Japan
July 8, 2018

IN THE VISUALIZATIONS THAT FOLLOW:

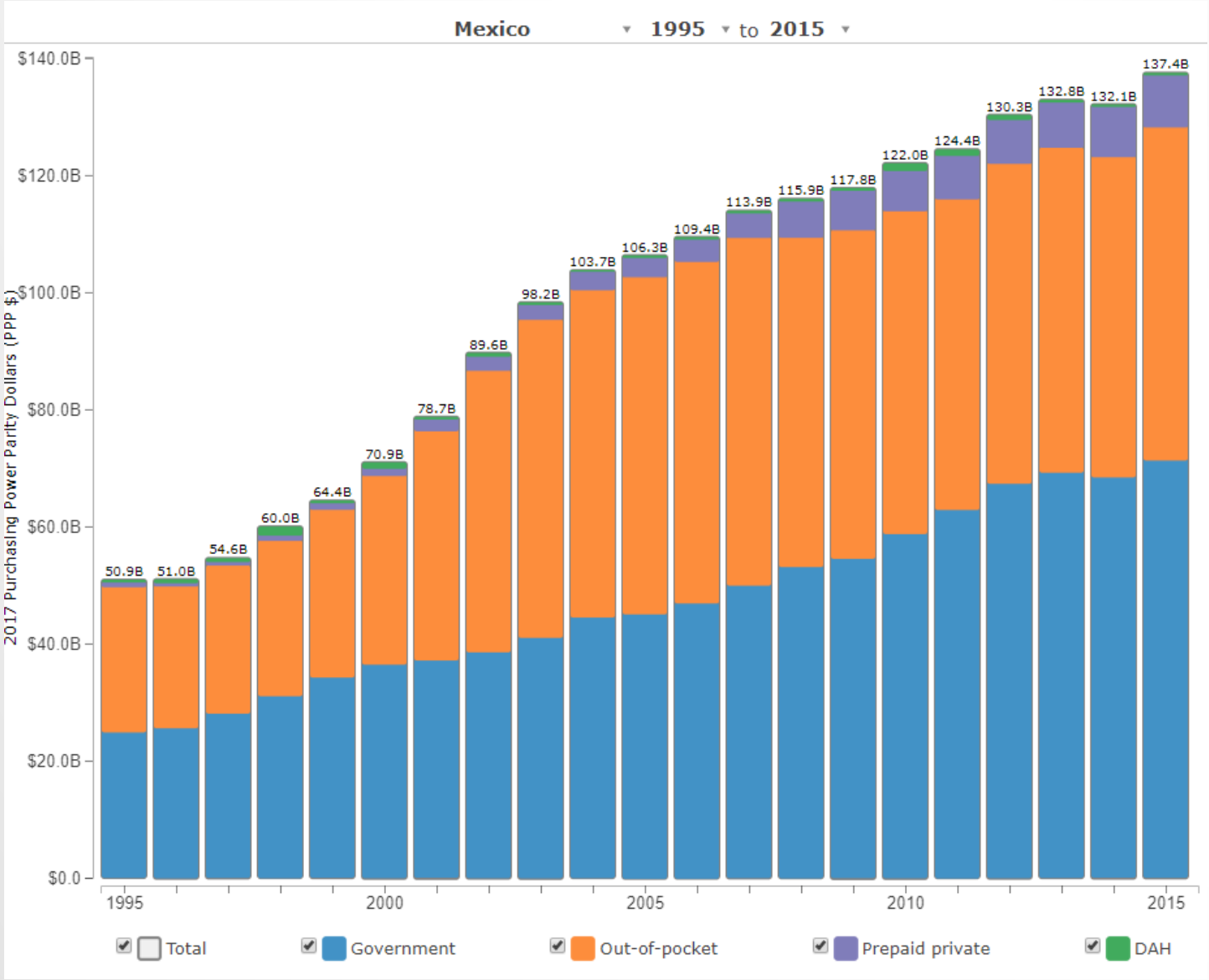
- What information can you extract?
- How are you extracting this information?

IHME GBD VISUALIZATION



https://vizhub.healthdata.org/gbd-compare/

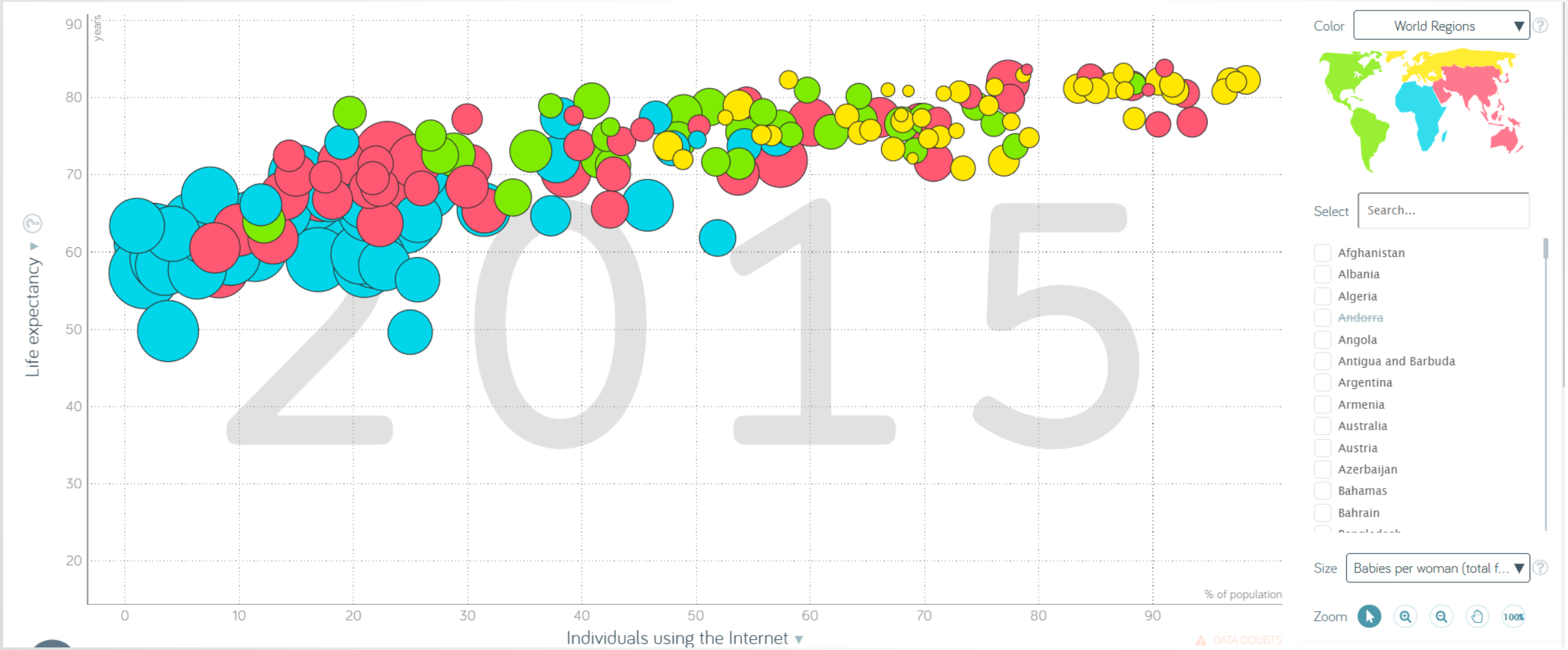
IHME HEALTH CARE SPENDING VISUALIZATION



(DAH = Development Assistance for Health)

http://ihmeuw.org/4iuu

GAPMINDER VISUALIZATION



https://www.gapminder.org/tools/#_state_marker_axis/_x_which=internet/_users&domainMin:null&domainMax:null&zoomedMin:null&zoomedMax:null&scaleType=linear;&size_which=children/_per/_woman/_total/_fertility&domainMin:null&domainMax:null;;;&chart-type=bubbles

PRIMARY SOURCES

- Wilkinson, L. **The Grammar of Graphics (2nd ed).** Springer Science. 2005

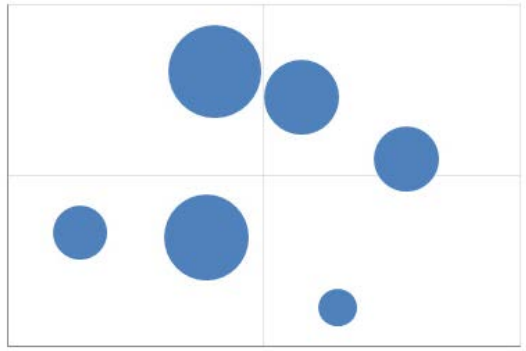
- Cleveland, WS and McGill, R. **Graphical perception: theory, experimentation, and application to the development of graphical methods.** *Journal of the American Statistical Association.* 79(387): 531-554. 1984.

WHAT IS A DATA VISUALIZATION?

A set of visual **geometries** whose **aesthetics** are mapped from **data**

Aesthetics

County	State	Resident t	Resident t	Resident t	Resident p	Resident p	Resident p	F
Autauga C	AL	55514	54571	1.7	54571	6.5	26	
Baldwin C	AL	190790	182265	4.7	182265	5.9	22.6	
Barbour C	AL	27201	27457	-0.9	27457	5.6	21.2	
Bibb Coun	AL	22597	22919	-1.4	22915	5.4	21.6	
Blount Co	AL	57826	57322	0.9	57322	6.2	24	
Bullock Cc	AL	10474	10915	-4	10914	6.8	21.2	
Butler Cou	AL	20307	20946	-3.1	20947	6.1	23.1	
Calhoun C	AL	117296	118572	-1.1	118572	6.1	22.7	
Chambers	AL	34064	34215	-0.4	34215	5.8	22	
Cherokee	AL	26021	25986	0.1	25989	4.8	20.8	



Data

Geometries

Many major visualization software (Tableau, ggplot in R, python, graph builder in JMP) are based on this grammar

GEOMETRY

- A ***geometry*** is a visual entity in space.
- Some common geometries encountered in data visualizations:

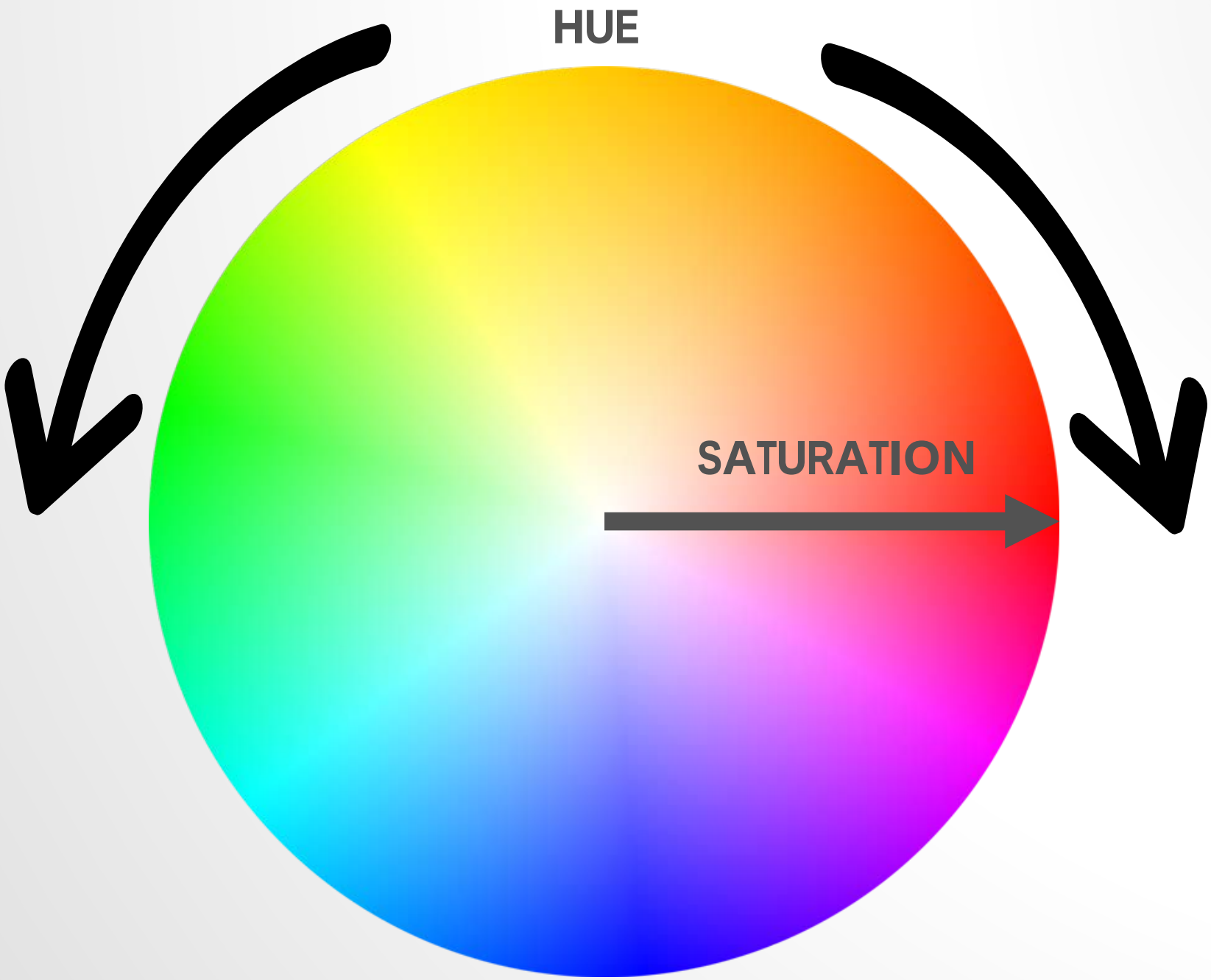
Point

Line

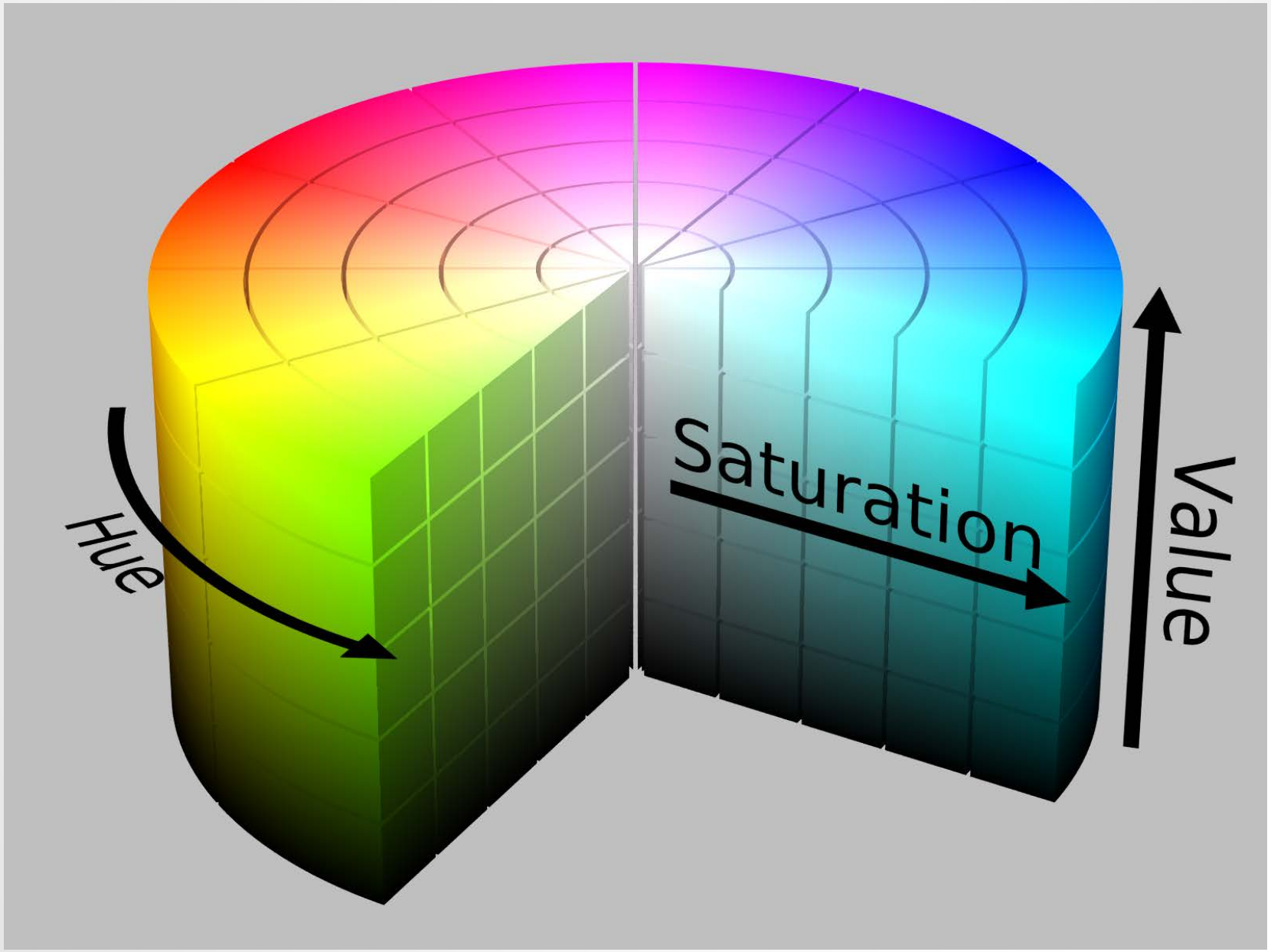
Bar

AESTHETIC

- An ***aesthetic*** is a visual attribute of a geometry
- Common aesthetics:
  - Position on horizontal (X)
  - Position on vertical (Y)
  - Shape
  - Size
  - Color
    - Hue
    - Saturation (“intensity”)
    - Value (“brightness”)
  - Text
- Not all aesthetics are available for every geometry



https://color.adobe.com/build2.0.0-buildNo/resource/img/kuler/color_wheel_730.png

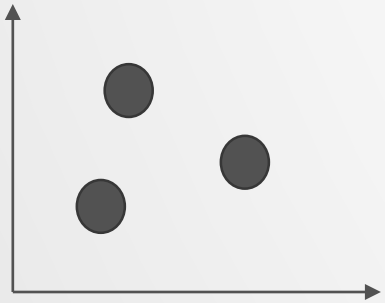


https://upload.wikimedia.org/wikipedia/commons/0/0d/HSV_color_solid_cylinder_alpha_lowgamma.png

AESTHETIC ATTRIBUTES OF A POINT GEOMETRY

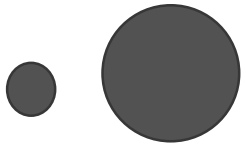
What distinguishes one point from another point?

X, Y



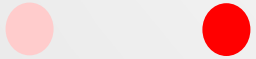
Shape

Size



Hue

Saturation



Value

Text

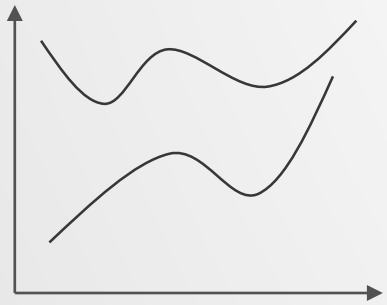
Point A

Point B

AESTHETIC ATTRIBUTES OF A LINE GEOMETRY

What distinguishes one line from another line?

X, Y



Shape

Size

Hue

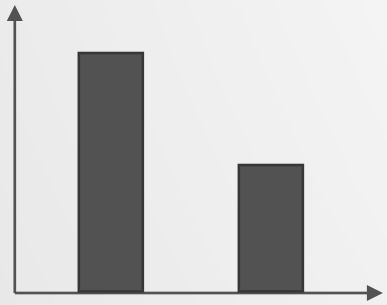
Saturation

Value

AESTHETIC ATTRIBUTES OF A BAR GEOMETRY

What distinguishes one bar from another bar?

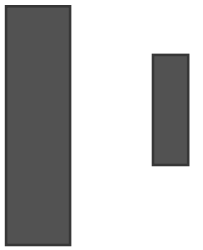
X, Y
(what happens if we use r, θ?)



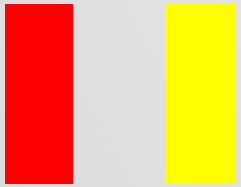
Shape

NA
(a bar is a bar!)

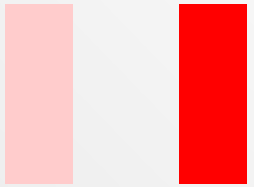
Size



Hue



Saturation



Value

DATA

- To visualize, must have data in row-by-column format where:
  - Rows represent cases: at most one geometry per case (assuming no aggregation)
  - Columns represent variables: to be mapped to aesthetic attributes

CONSTRUCTING A DATA VISUALIZATION

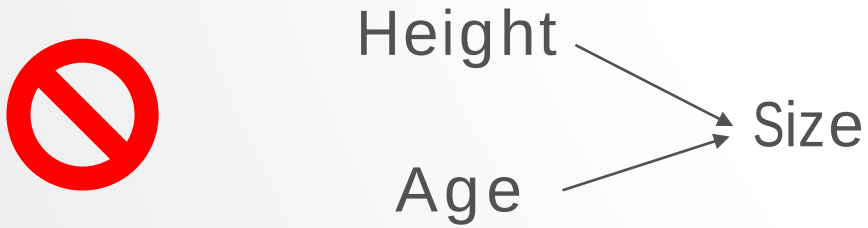
- Differences in **geometry aesthetics** map to differences in **data variables**

- Available mappings depend on whether **data variable** is *continuous* (height) or *discrete* (race)

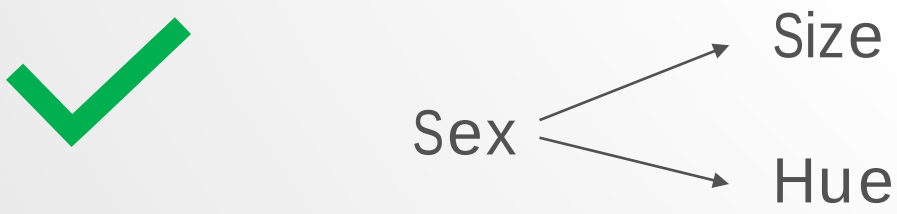
CONSTRUCTING A DATA VISUALIZATION

The following caveats apply:

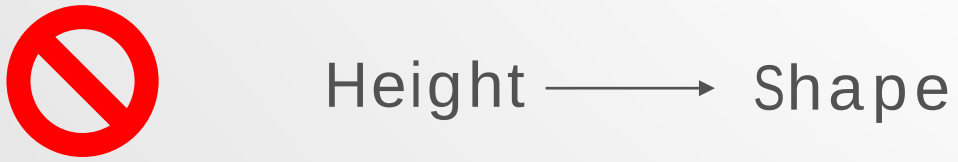
1. An aesthetic attribute can be mapped back to at most one variable



2. A variable can be mapped to more than one aesthetic

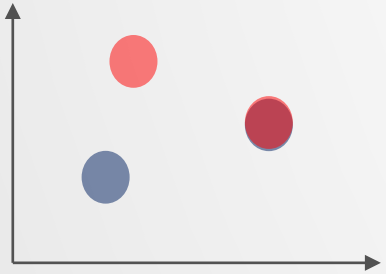


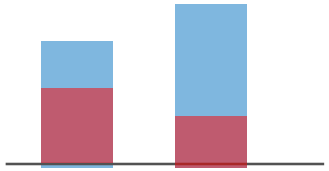
3. Not all mappings make sense



MODIFIERS

- Ties: when two cases yield overlapping geometries under a given mapping

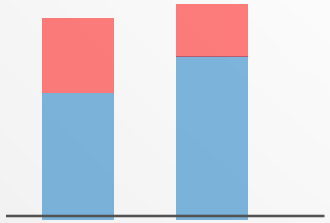




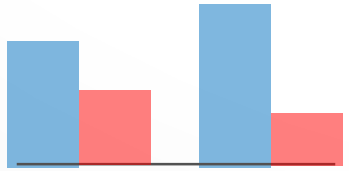
- Some common modifiers:

Jitter:

Stack:



Dodge:

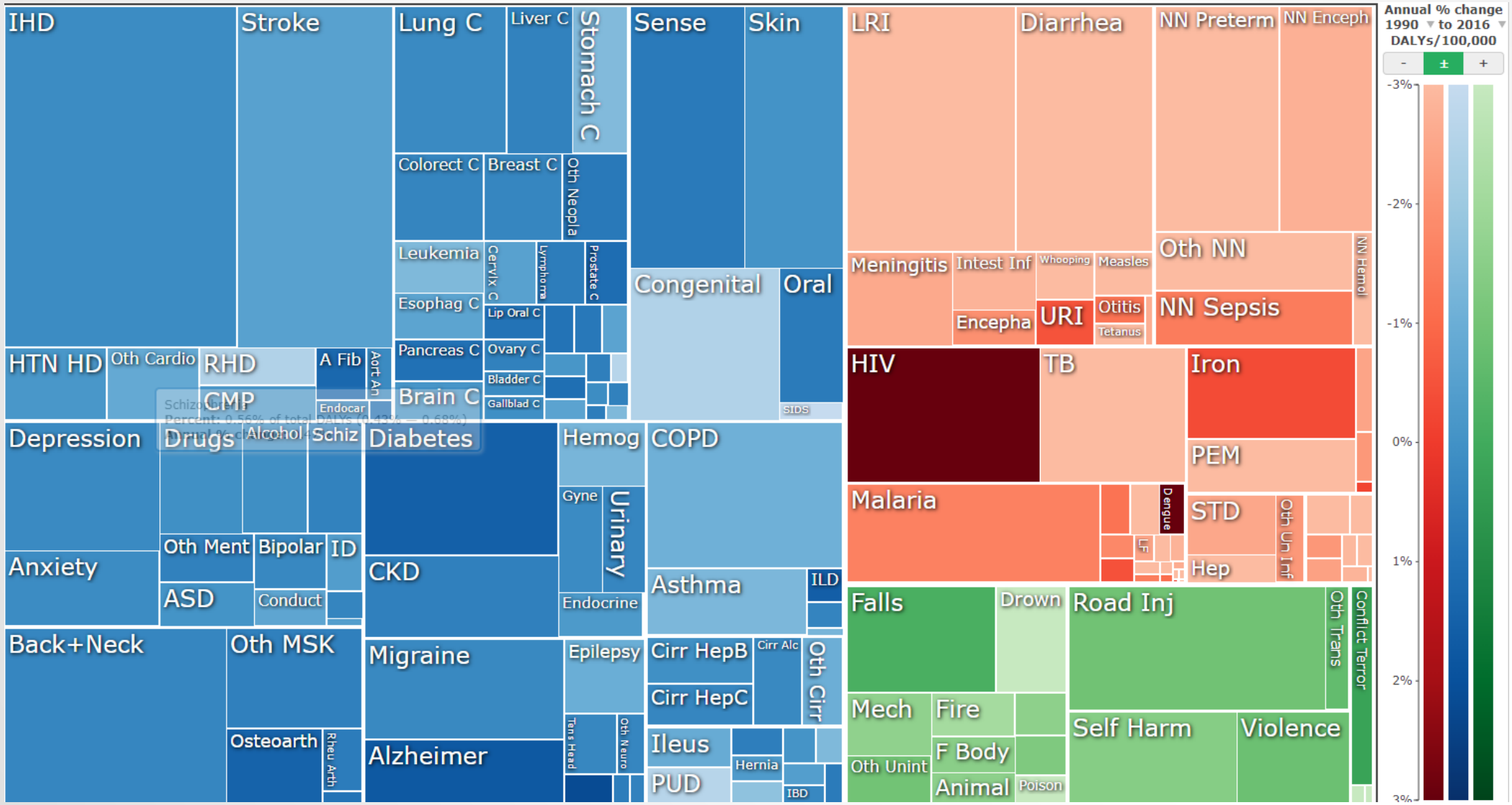


FIND THE MAPPINGS!

- Revisit the three visualizations we encountered earlier. Identify:

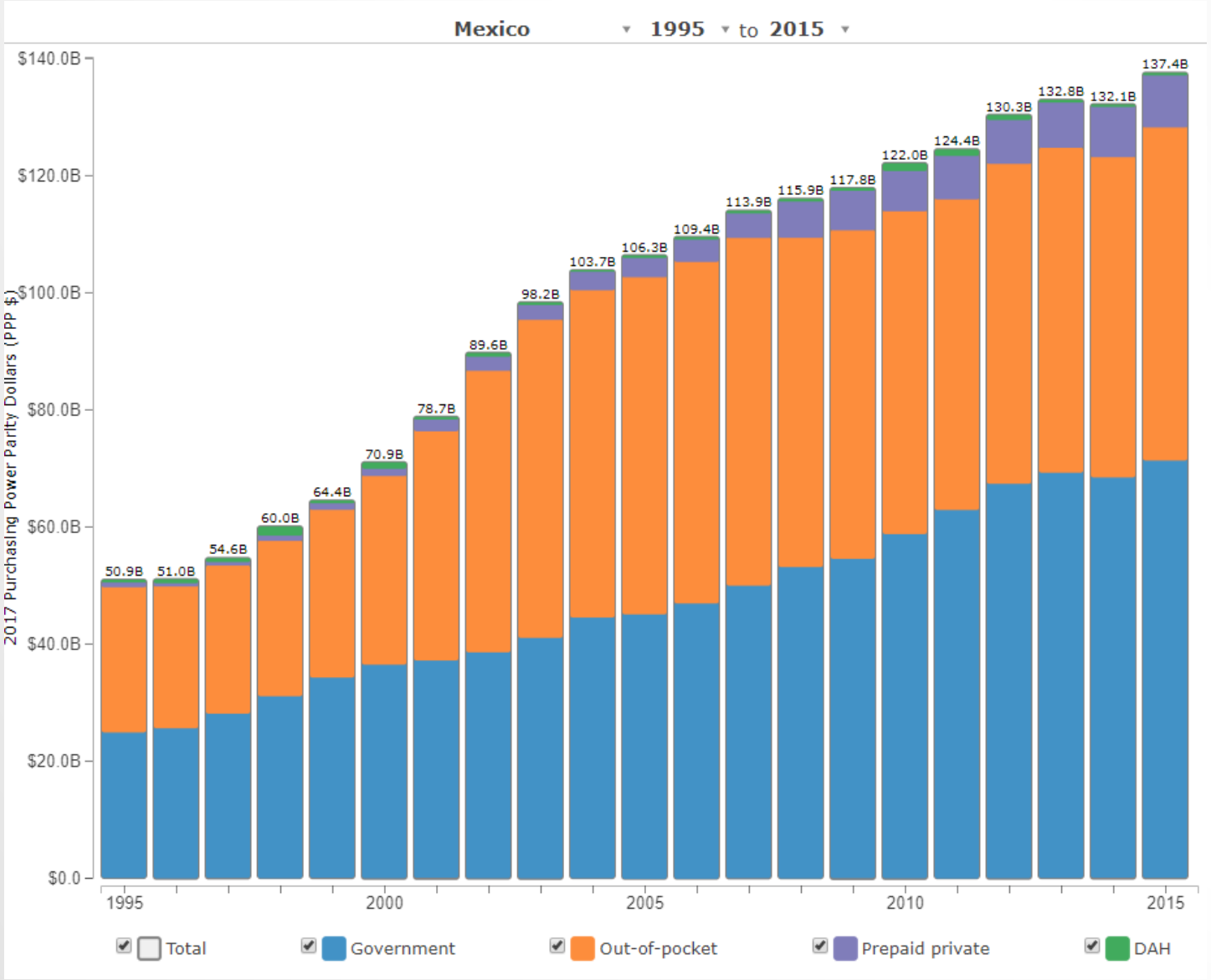
1. The data cases (assuming one geometry per case);
2. The geometries;
3. The aesthetic attributes that are varied;
4. The variables that control the differences in aesthetic attributes (bonus: are they continuous or discrete?)
5. Modifiers (if any)

IHME GBD VISUALIZATION



https://vizhub.healthdata.org/gbd-compare/

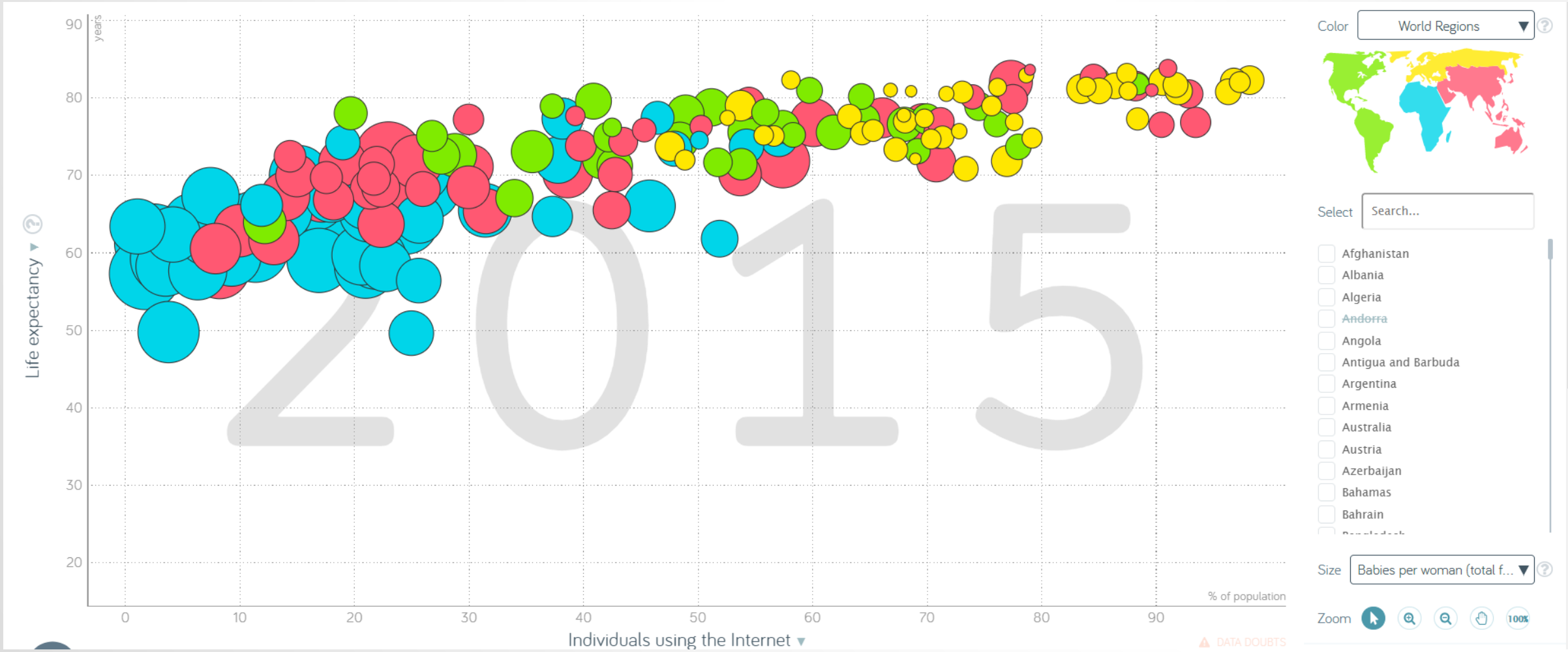
IHME HEALTH CARE SPENDING VISUALIZATION



(DAH = Development Assistance for Health)

http://ihmeuw.org/4iuu

GAPMINDER VISUALIZATION



https://www.gapminder.org/tools/#_state_marker_axis/_x_which=internet/_users&domainMin:null&domainMax:null&zoomedMin:null&zoomedMax:null&scaleType=linear;&size_which=children/_per/_woman/_total/_fertility&domainMin:null&domainMax:null;;;&chart-type=bubbles

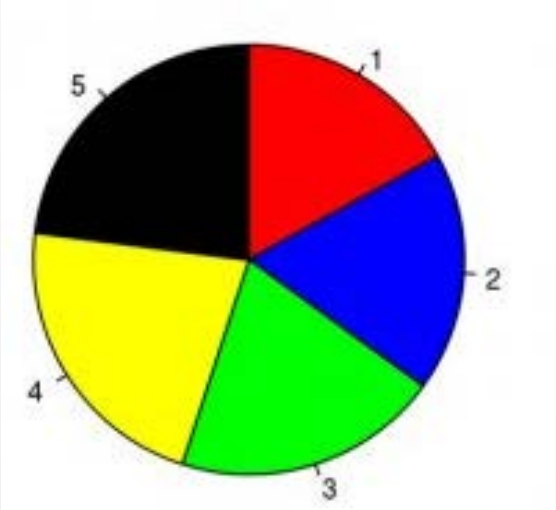
RANKING THE AESTHETIC ATTRIBUTES

The work of William S. Cleveland

- Recall the following aesthetic attributes:
  - Position on horizontal (X)
  - Position on vertical (Y)
  - Shape
  - Size
  - Color
    - Hue
    - Saturation (“intensity”)
    - Value (“brightness”)

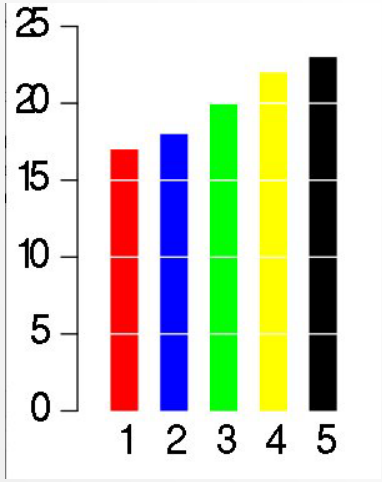
- Cleveland & McGill created a ranking of these when mapping a geometry to a quantitative variable

- Hypothetical polling data on 5 candidates.
- Geometry: “pie wedge” (a bar in polar coordinates)
- Aesthetic mappings:
  - Candidate (discrete) → hue & text
  - share of vote (continuous) → size, specifically angle
- Rank the candidates’ vote shares.



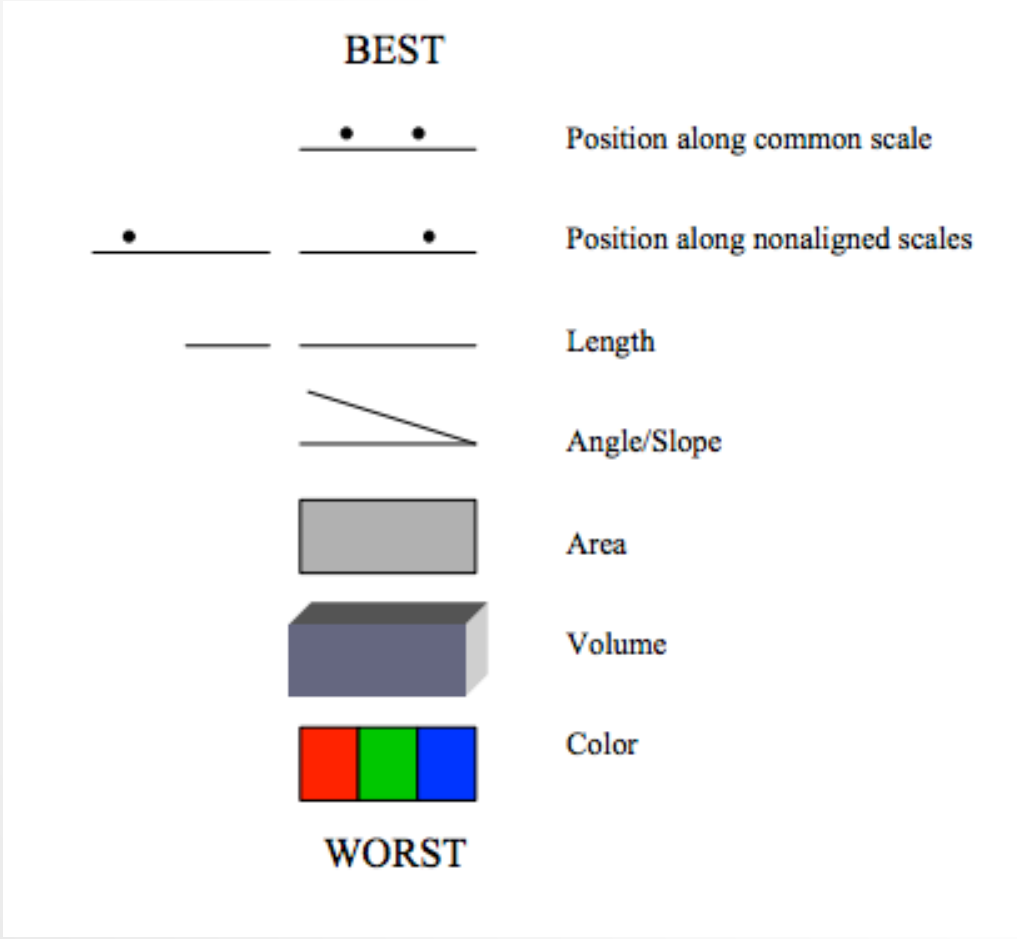
Walter Hickey at Business Insider

- Same data, different visualization
- Geometry: bar
- Aesthetic mappings:
  - candidate → hue & X
  - share of vote → Y
- Rank the candidates' vote shares.



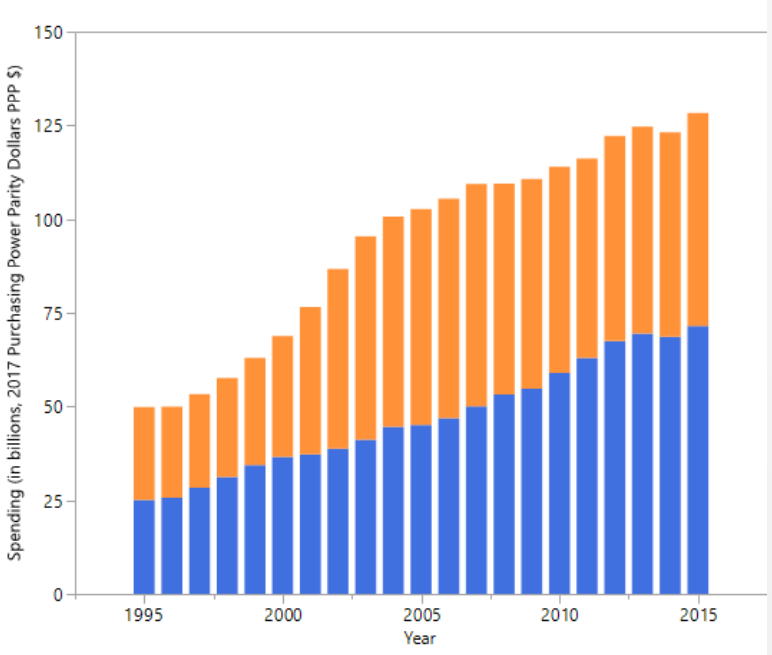
Walter Hickey at Business Insider

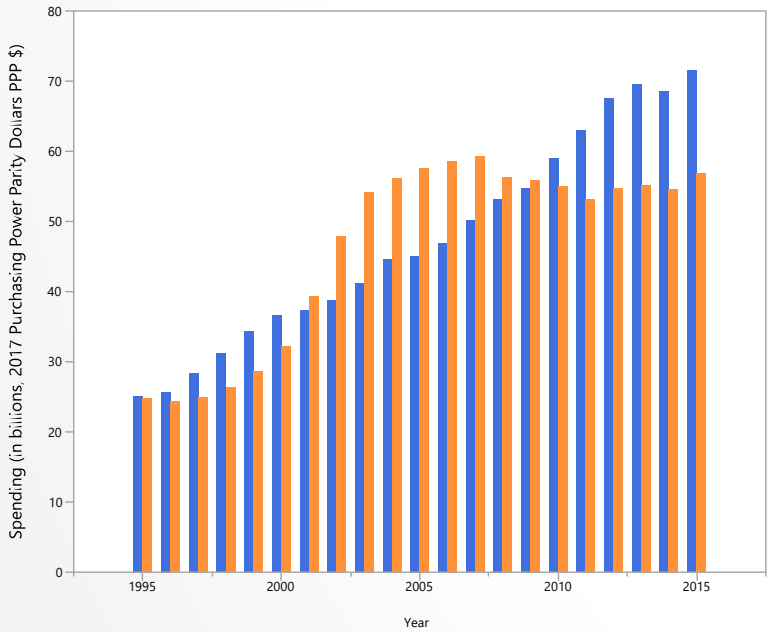
Cleveland's hierarchy for mapping quantitative data:

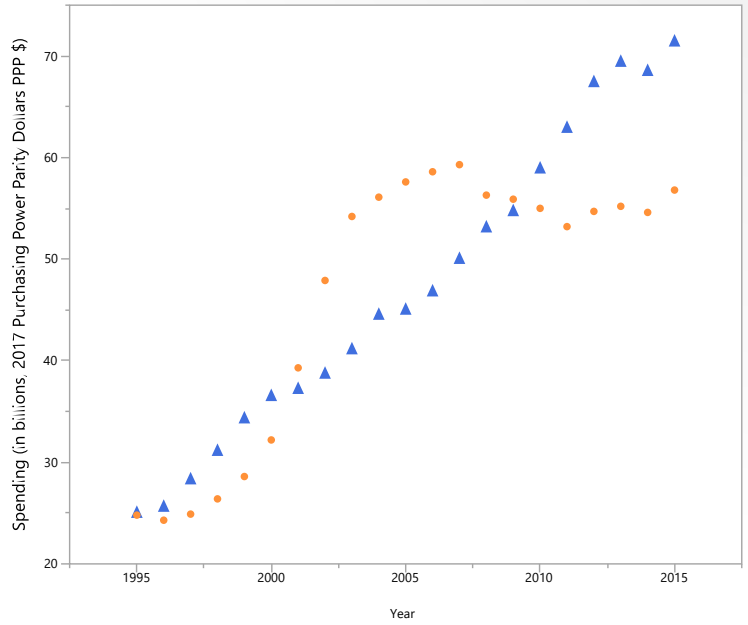


Wilkinson, *Presentation Graphics*

Mexico's health spending (**Government** and **Out-of-pocket** only)







Geometry: Bar
Year → X
Spending → Y
Source → Hue
Modifier: Stack

Geometry: Bar
Year → X
Spending → Y
Source → Hue
Modifier: Dodge

Geometry: Point
Year → X
Spending → Y
Source → Hue, Shape
Modifier: None