



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:

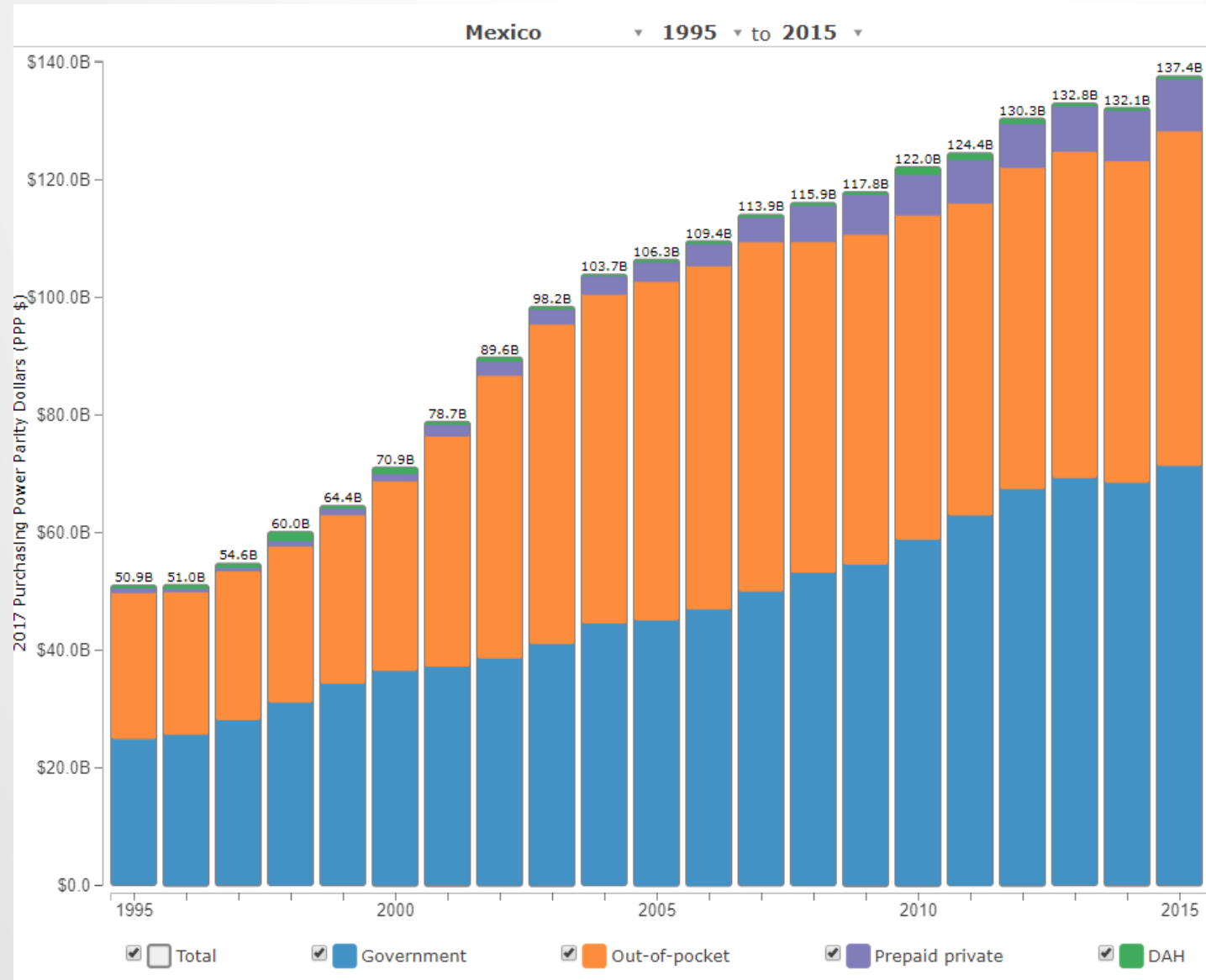
- What information can you extract?
- How are you extracting this information?

This treemap visualizes the burden of diseases and injuries (DALYs) in 2016, categorized by the annual percentage change from 1990 to 2016. The color scale ranges from -3% (dark blue) to +3% (dark red), with 0% being white. The size of each rectangle represents the number of DALYs, and the area is divided into three main sections: Non-communicable diseases (left, blue), Infectious diseases (middle, orange), and Injuries (right, green).

Non-communicable diseases (left section):

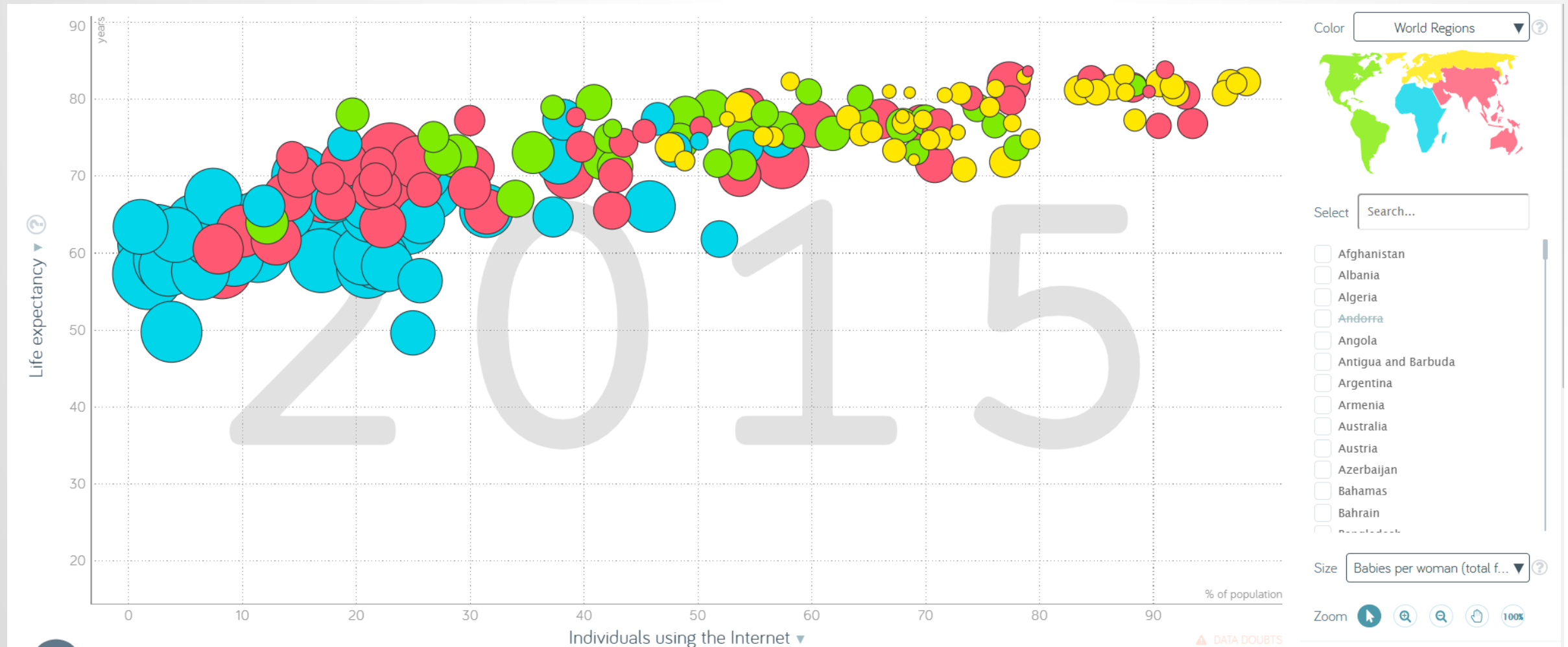
- IHD** (Ischemic Heart Disease): -3.0%
- Stroke**: -2.5%
- Lung C** (Lung Cancer): -2.0%
- Liver C** (Liver Cancer): -1.5%
- Stomach C** (Stomach Cancer): -1.0%
- Colorect C** (Colorectal Cancer): -0.5%
- Breast C** (Breast Cancer): -0.5%
- Oth Neopla** (Other Neoplasms): -0.5%
- Leukemia**: -0.5%
- Cervix C** (Cervical Cancer): -0.5%
- Lymphoma**: -0.5%
- Prostate C** (Prostate Cancer): -0.5%
- Esophag C** (Esophageal Cancer): -0.5%
- Lip Oral C** (Lip and Oral Cancers): -0.5%
- Pancreas C** (Pancreatic Cancer): -0.5%
- Ovary C** (Ovarian Cancer): -0.5%
- Bladder C** (Bladder Cancer): -0.5%
- Gallblad C** (Gallbladder Cancer): -0.5%
- Brain C** (Brain Cancer): -0.5%
- HTN HD** (Hypertensive Heart Disease): -2.0%
- Oth Cardio** (Other Cardiovascular Diseases): -1.5%
- RHD** (Rheumatic Heart Disease): -1.0%
- A Fib** (Atrial Fibrillation): -0.5%
- Aortic An** (Aortic Aneurysm): -0.5%
- CMP** (Congestive Heart Failure): -0.5%
- Endocar** (Endocarditis): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog** (Hemoglobinopathy): -0.5%
- COPD** (Chronic Obstructive Pulmonary Disease): -0.5%
- Depression**: -0.5%
- Drugs** (Drug Use): -0.5%
- Alcohol** (Alcohol Use): -0.5%
- Schiz** (Schizophrenia): -0.5%
- Diabetes**: -0.5%
- Hemog</**

IHME HEALTH CARE SPENDING VISUALIZATION



(DAH = Development Assistance for Health)

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**



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

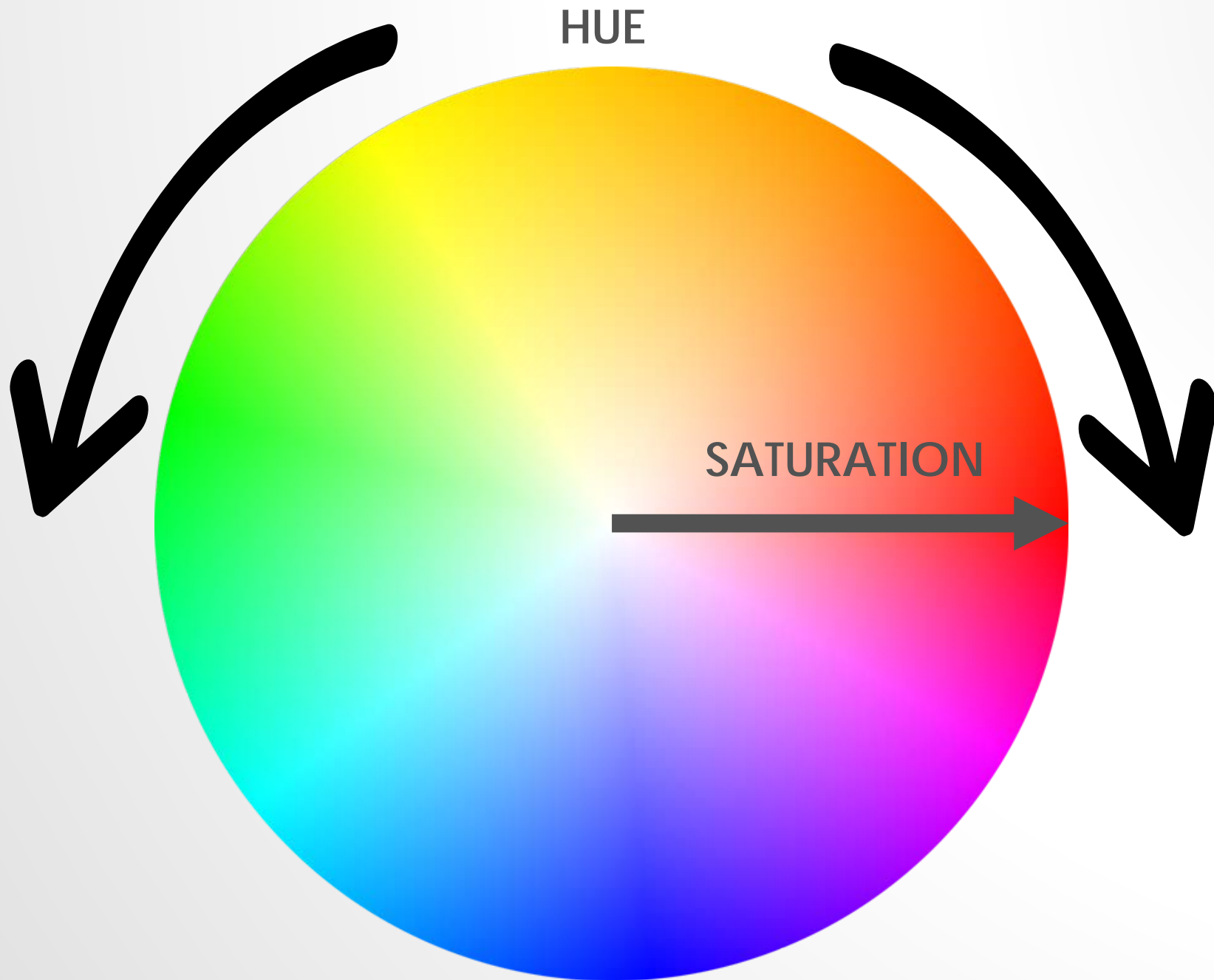


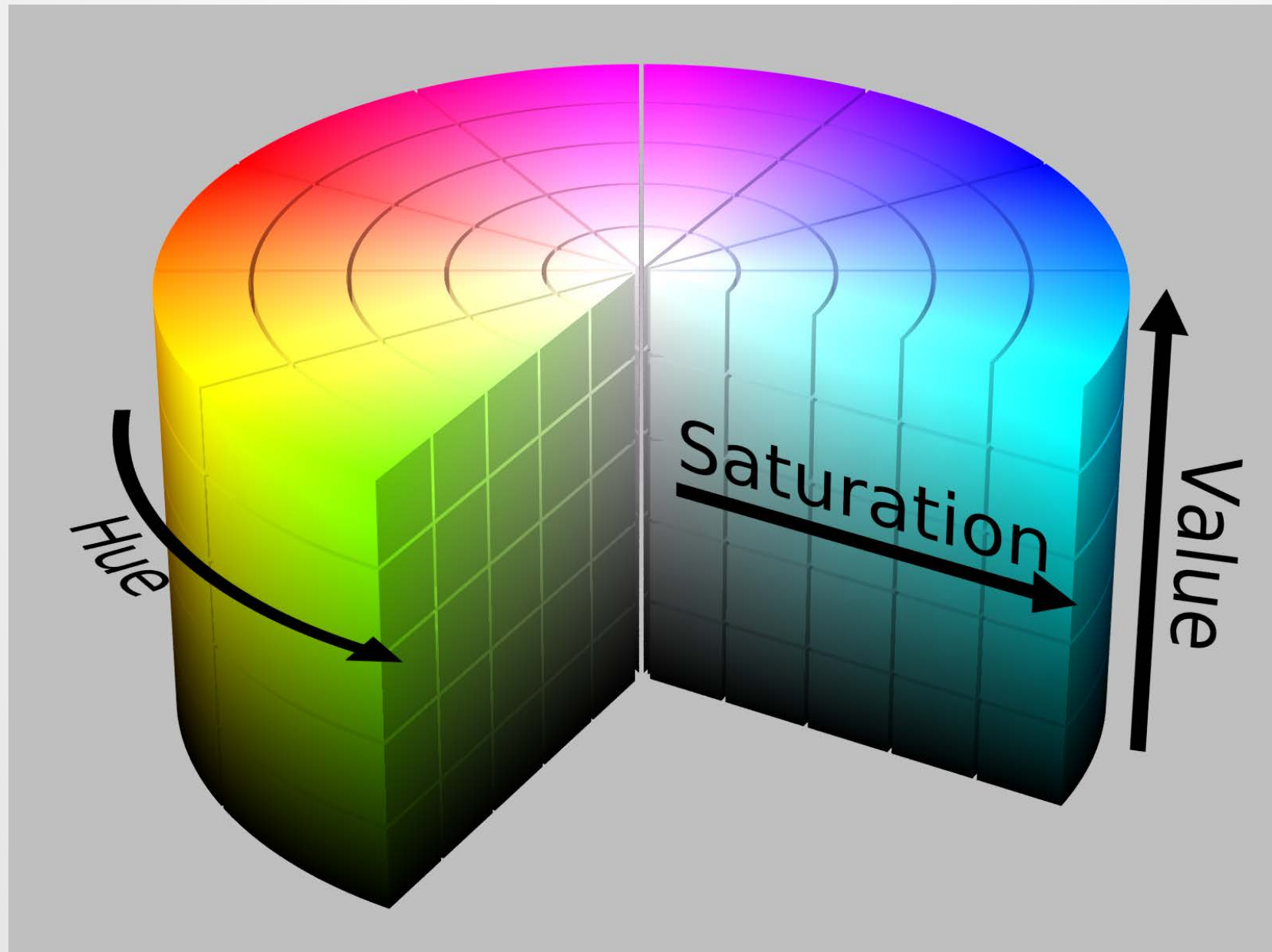
Bar



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



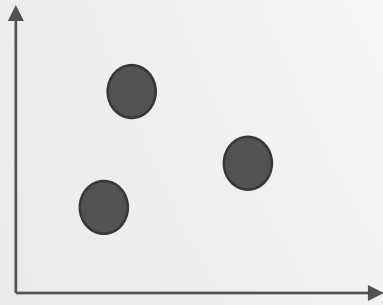


https://upload.wikimedia.org/wikipedia/commons/0/0d/HSV_color_solid_cylinder_alpha_lowgamma.png

AESTHETIC ATTRIBUTES OF A POINT GEOMETRY

What distinguishes one point from another point?

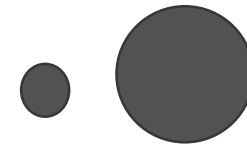
X, Y



Shape



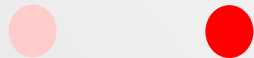
Size



Hue



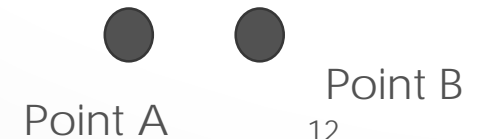
Saturation



Value



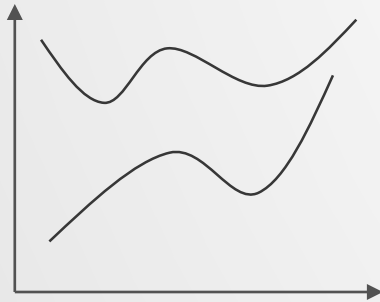
Text



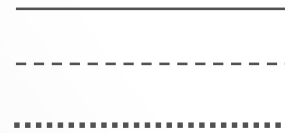
AESTHETIC ATTRIBUTES OF A LINE GEOMETRY

What distinguishes one line from another line?

X, Y



Shape



Size



Hue



Saturation



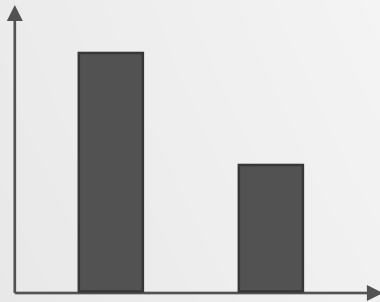
Value



AESTHETIC ATTRIBUTES OF A BAR GEOMETRY

What distinguishes one bar from another bar?

X, Y
(what happens if we use r, θ ?)



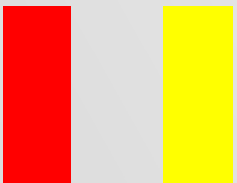
Shape

NA
(a bar is a bar!)

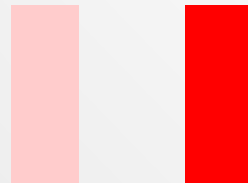
Size



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

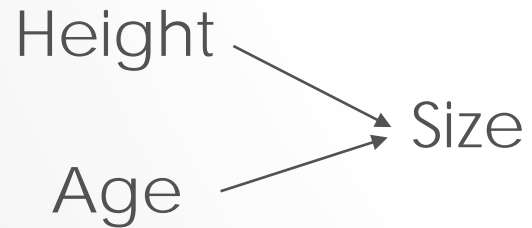
CONSTRUCTING A DATA VISUALIZATION

- Differences in **geometry aesthetics** map to differences in **data variables**
- Available mappings depend on whether **data variable** is *continuous* (height) or *discrete* (race)

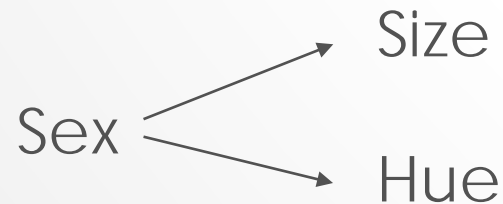
CONSTRUCTING A DATA VISUALIZATION

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

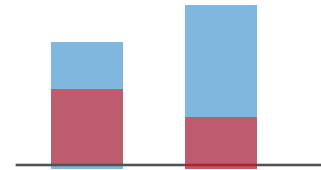
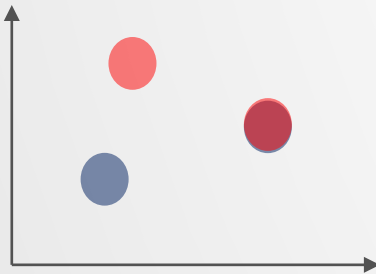


3. Not all mappings make sense



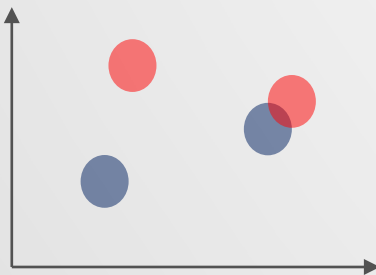
MODIFIERS

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

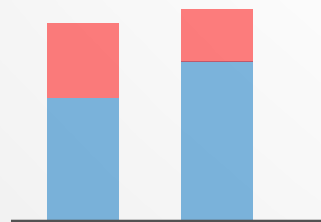


- Some common modifiers:

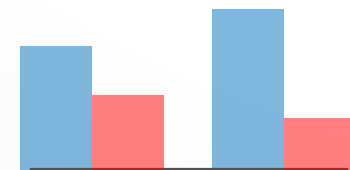
Jitter:



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)

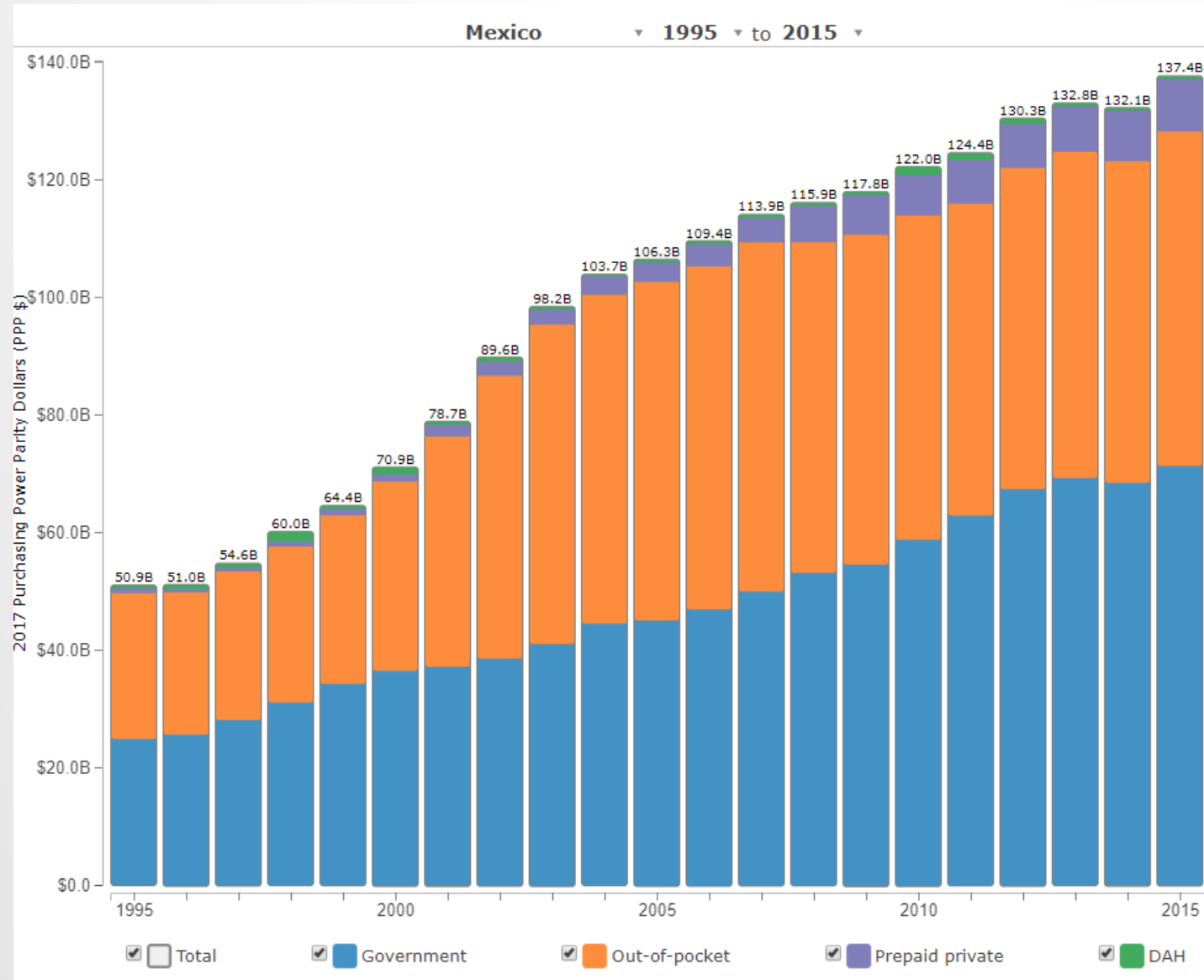
This treemap visualizes the burden of diseases and injuries (DALYs) in 2016, categorized by the annual percentage change from 1990 to 2016. The color scale ranges from -3% (dark blue) to +3% (dark red), with 0% being white. The size of each rectangle represents the number of DALYs, and the area is divided into three main sections: Non-communicable diseases (left, blue), Infectious diseases (middle, orange), and Injuries (right, green).

Legend: Annual % change 1990 to 2016 DALYs/100,000. Scale: -3%, -2%, -1%, 0%, 1%, 2%, 3%.

Key Disease Categories and Sub-categories:

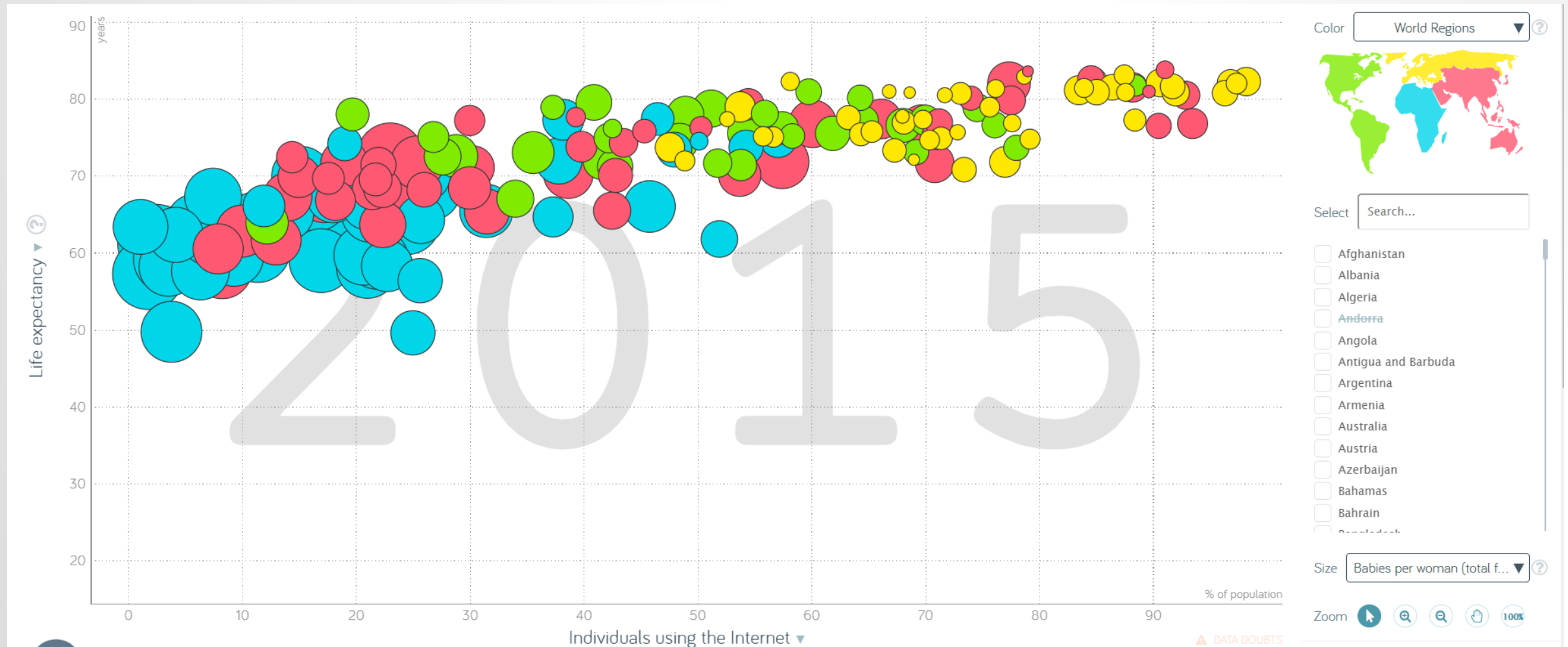
- Non-communicable diseases (Left):** IHD, Stroke, Lung C, Liver C, Stomach C, Colorect C, Breast C, Oth Neopla, Leukemia, Esophag C, Pancreas C, Brain C, Diabetes, Hemog, COPD, Asthma, CKD, Migraine, Alzheimer, Oth MSK, Osteoarth, RHD, A Fib, Aortic A, CMP, Schiz, Depression, Anxiety, Back+Neck, Oth Cardio, HTN HD, Oth Ment, Bipolar, ID, ASD, Conduct, Endocar, Gyne, Urinary, Endocrine, Epilepsy, Cirr HepB, Cirr HepC, Ileus, PUD, Hernia, IBD, Oth Cirr, Oth Neuro, Tens Head.
- Infectious diseases (Middle):** LRI, Diarrhea, NN Preterm, NN Enceph, Meningitis, Intest Inf, Whooping, Measles, Oth NN, NN Sepsis, NN Hemol, HIV, TB, Iron, PEM, STD, Oth Un Inf, Hep, Dengue, LF, Malaria.
- Injuries (Right):** Falls, Drown, Road Inj, Oth Trans, Conflict Terror, Mech, Fire, Self Harm, Violence, Oth Unint, F Body, Animal, Poison.

IHME HEALTH CARE SPENDING VISUALIZATION



(DAH = Development Assistance for Health)

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

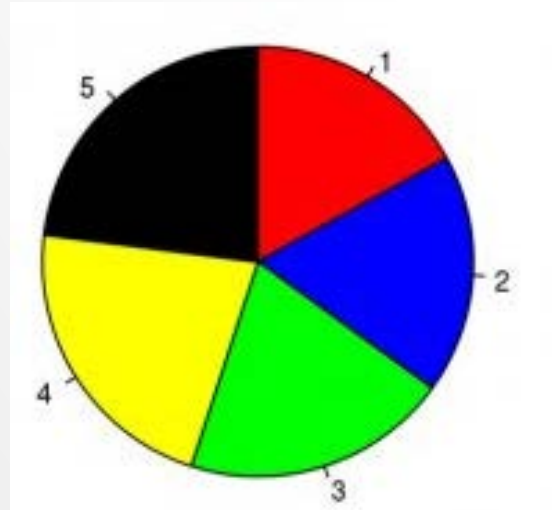


RANKING THE AESTHETIC ATTRIBUTES

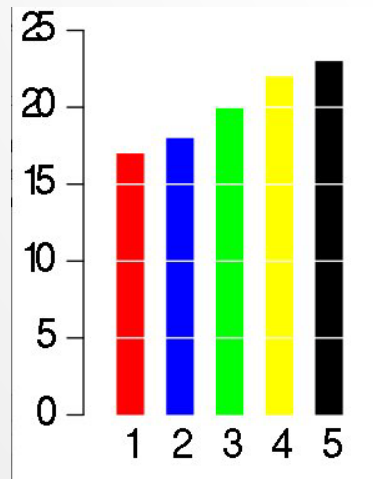
The work of William S. Cleveland

- 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.

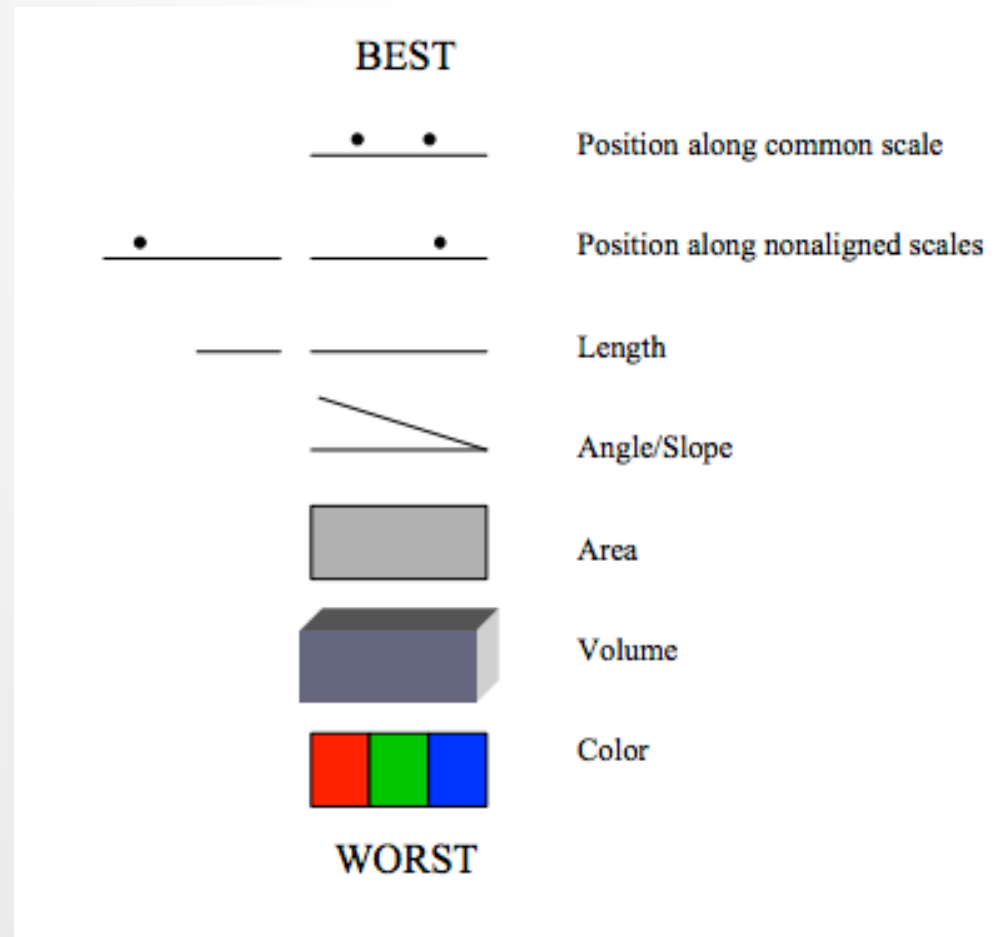


- Same data, different visualization
- Geometry: bar
- Aesthetic mappings:
 - candidate → hue & X
 - share of vote → Y
- Rank the candidates' vote shares.

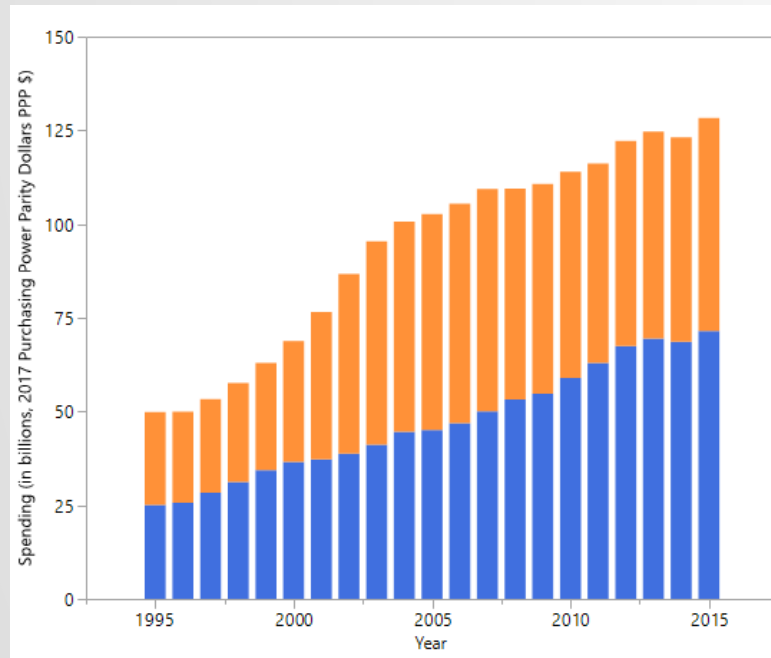


Walter Hickey at [Business Insider](#)

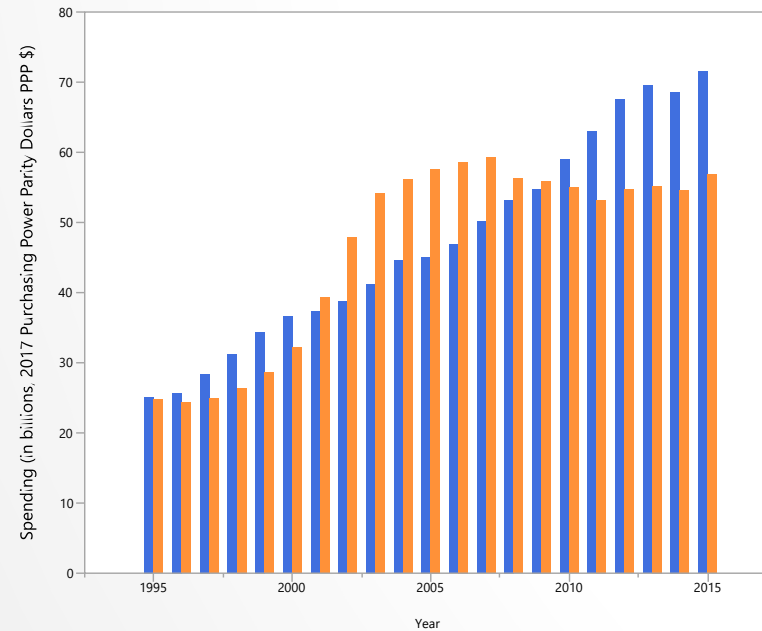
Cleveland's hierarchy for mapping quantitative data:



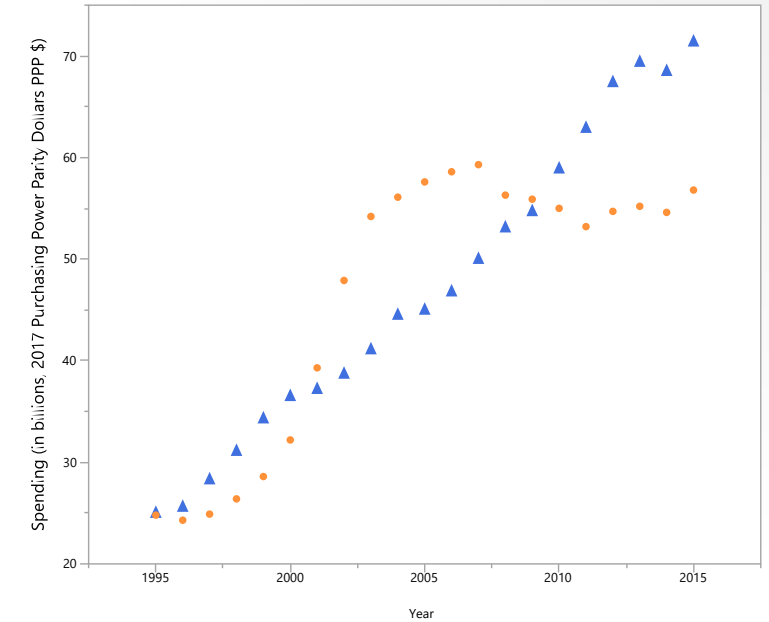
Mexico's health spending (Government and Out-of-pocket only)



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