The Data Visualization Process

DEFINE A QUESTION & GOAL

- What do you want to learn or find out?
- What story or message do you want to tell?

DETERMINE THE AUDIENCE

Who will use this information?

How will they use it?
- interactive / online
- presentation
- one pager
- report
- poster

Why will they use it?
- to learn
- to explore/understand
- to make decisions
- to communicate data
- ...

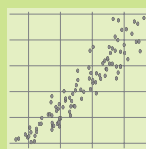## EXPLORE THE DATA

Do the data make sense?
- How are the data distributed?
- Are there any outliers?
- Are there missing data?
- Do the data fall within a reasonable range?

← Is this right?

Plot bar charts (categorical) or histograms (numerical)

What do they mean?
- Are variables within the data related?
- How are they related?

Plot scatter or line plots between two variables
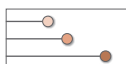
Do new variables shed more insight?
- Average values (point or running)
- Create an index or composite to reduce dimensions
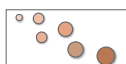- Calculate a percent, ratio, or normalize

DEFINE THE COMPARISON

I want to show the **relationship** between _____ and _____

The **primary comparison** I want to show is:

- ☐ Magnitude / size
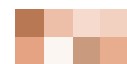- ☐ Relationship between 2+ variables
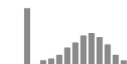- ☐ Trend over time
- ☐ Ratio / composition of a group

- ☐ Differences *between* groups (geographies, groups of people)
- ☐ Differences *within* groups (distribution)

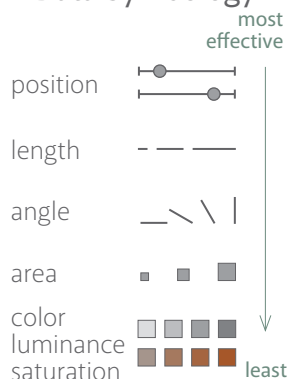 target →
- ☐ Between different data (provide context)
- ☐ Geographic

SKETCH

- Try different chart types
- Organize how to represent the information
- Mock-up in software and refine

Data Symbology

most effective

- position
- length
- angle
- area
- color
  - luminance
  - saturation

least

EVALUATE — Refine the visualization by asking if it is:
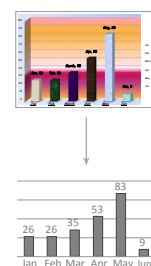
Effective
- ☐ Is the relationship conveyed through the visualization?
- ☐ Is the chart type the best way of explaining the relationship?
- ☐ Is the most important comparison encoded with position?
- ☐ Should other data or comparisons be added to the chart?
- ☐ Will the visualization be useful to the audience?

Simple
- ☐ Does the visualization focus on the data?
- ☐ Is there chart junk (heavy lines, patterns, 3D, etc.)?
- ☐ Is every dot, symbol, color, line, and variable necessary?
- ☐ Did you get it right in black and white? Is color used judiciously? Consistently?

Organized
- ☐ Should the chart be broken into small multiples?
- ☐ Should categories be sorted/grouped?
- ☐ Are annotations used effectively?
- ☐ Are values labeled directly? (no legend)

Responsible
- ☐ Does the plot faithfully represent the data?
- ☐ Has data been excluded? Is this noted?
- ☐ Are the numbers meaningful? Should they be raw numbers? Some transformation?
- ☐ Do the axes have labels?
- ☐ Does averaging values smooth out the noise or wash away the signal?
- ☐ Are data sources and manipulations clear?

GEOCENTER