Agenda:

Introduction:

Part 1: Some data is meant to be consumed visually.

Visualization lets you see the big picture.

Visual: Excel worksheet with numbers that correspond to RGB values. When zoomed out, viewers can see a picture.

Translating visuals to action:

Visual: A dataset that contains a x,y grid with values. Upon converting the data into a visualization with elevation bands, viewers understand that this a Topographic map. Furthermore, armed with this information and tasked with plotting a path to navigate the map, viewers can easily pick the best path, but only because they can see the data visually.

Beyond the numbers with Anscombe Quartet

Visual: This one is famous. I am sure you have heard of it. The lesson to portray to the audience is that measures of central tendency can obscure the underlying patterns of the data and hide otherwise obvious insights.

Part 2: The Grammar of Graphics

Encoding: Converting something from one system of communication to another. How numbers and information are converted for visual consumption.

Data observations:

Transformations: How do we need to manipulate our data for analysis?

Granular/Aggregated?

Group by/Calculations/Standardizations?

Scaffolding: Plot area. Canvas.

Axis, Scale, Orientation

Encode data as objects in space

Objects:

Categoric: Shape, Color, Label

Continuous: Shade, Size

Spatial:

Categoric: Series, Facets, order(time)

Continuous: Position, proximity, density, size

Part 3: Put it into practice

Pro Tip: Attempt to describe the top of exploration in plain English.

Example 1: Compare two means with a p val of 99%

So really I am comparing two distributions that have been found to likely be dis-similar. Ok easy, we should be able to see this visually.

Visual creation process.

Data Transformation:

1. Binning - Group by
2. KDE – Standardize for compatibility
3. Plot
4. Evaluate

Example 2: Wolves