Warm Up

- The moving geospatial visualization talks about the connectedness between people in their local communities based on region. This was used by compiling Facebook data on different users and their friend's locations.

Topics of the Day

- Data Types In Visualizations
- Distribution
  - The original slide that talks about this, the slide titled Data Types, the Profesor proposes a series of questions. In these questions, the professor centers a few ideas that he talks about: geometric, categories, quantity, etc.
  - Geometric objects refer to examples of vectors and colors as set columns.
    - The example that he uses is the example of the pot that is given depth and demission to look like a realistic teapot. These vectors allow for the bowl to look more realistic, from the handle to the lid the pot contained depth. But the significant thing to notice is that the pot has a lot
    - Quantity: "We can encode the values associated with a data point by modifying how we express it. To do so, we need to be able to identify the different components of representation, and how we can scale between them."

- In the analysis above it talks about the representation of objects. Thes objects can be represented through teh following dimensions:
- Position, color, size, shape, relationship, and motion
- When plotting these different representations, you have to make sure that the visual is not overpacked with detail because although the quantity is there, you don't want to overwhelm the user with too much information.

Distributions

- Distribution is all about the way the data is placed on the graph and the way the information is portrayed in the distribution.
  - The professor plotted a series of points on the graph that he premade and started drawing lines through them to show different variations of the same position but explained in many different ways.
  - He then goes on to pose a question, Given a set of points, how do we make them into a continuous distribution?
  - The answer to this question is simple, bins.
  - This is the syntax for bins: bin_id = floor((value - left_edge) / bin_width)
  - This basically allocates the prospective value of a given row to a particular bin so that the distribution is adequately placed.
  - This can take the form of columns, rows, etc.

- When conducting this type of distribution on large datasets, the issues become about processing speed. How can you get the processing of distribution of large data sets down to a minimum
- summation: total quantity measured
- Arithmetic Average: average or mean quantity definer.
- Weighted Average: mean but not the arithmetic mean
- After this the professor had us get into groups to do our own binning examples.