DOM-E5057 (M01) Information Graphics; Formats and Genres
Workshop by Prof. Lily Díaz
26 – 30 Jan, 2014
Day 1

Some basic information about data values and information graphics

Qualitative = Descriptive
Quantitative = Numerical

Categorical variables/data/elements

A categorical value/variable, also called a nominal variable, is one that has two or more categories, but there is no intrinsic ordering to the categories. Some examples:

- Gender (Female, Male)
- Hair color (Blondes, Brunettes, Redheads)
- The type of blood of a person
- The country where (s)he lives.

There is usually no agreed way to order these variables (e.g. in terms of highest or lowest)

> "A purely categorical variable is one that simply allows you to assign categories but you cannot clearly order the variables."

> "A set of data is said to be *categorical* if the values / observations belonging to it can be sorted according to category. Each value is chosen from a set of non-overlapping categories. For example, shoes in a cupboard can be sorted according to color; the characteristic 'color' can have non-overlapping categories 'black', 'brown', 'red' and 'other'."

Ordinal variables/data

"The difference between categorical and ordinal values is that there is a clear ordering of the variables. For example, suppose you have a variable, **economic status**, with three categories (low, medium and high). In addition to being able to classify people into these three categories, you can order the categories as low, medium and high. Now consider a variable like **educational experience** that can be ordered as elementary school, high school, university and university graduate. We can order these from lowest to highest, the spacing between the values may not be the same across the levels of the variables. Say we assign scores 1, 2, 3 and 4 to these four levels of educational experience and we compare the difference in education between categories one and two with the difference in educational experience between categories two and three, or the difference between categories three and four. The difference between categories one and two (elementary and high school) is probably much bigger than the difference between categories two and three (high school and some college). In this example, we can order the people in level of educational experience but the size of the difference between categories is inconsistent (because the spacing between categories one and two is bigger than categories two and three)."

Interval
An interval variable is similar to an ordinal variable, except that the intervals between the values of the interval variable are equally spaced. For example, suppose you have a variable such as annual income that is measured in dollars, and we have three people who make $10,000, $15,000 and $20,000. The second person makes $5,000 more than the first person and $5,000 less than the third person, and the size of these intervals is the same. If there were two other people who make $90,000 and $95,000, the size of that interval between these two people is also the same ($5,000).

Discreet data
A set of data is said to be *discrete* if the values / observations belonging to it are distinct and separate. Examples include:

- Number of students in a class (you cannot half a student)
- Number of results possible when throwing dice (11)
- Number of customers in a shop at a given time.
- Whole numbers

Continuous
A set of data is said to be *continuous* if the values / observations belonging to it may take on any value within a finite or infinite interval. You can count, order and measure continuous data. Some examples:

- A person's height, weight, age
- The length of a football field
- The time it takes to run around a building.
- Can take any value within a range

Area graphs
Display quantitative data graphically. It is used to "communicate differences in size, quantity, value, etc. (Harris 9)

Based on the line chart, an area graph evaluates data values to a total over time. Usually quantitative values are placed on the y-axis and analyzed (compared) with a category, quantitative or sequence scale on the x-axis. Each line (or shaded area) on the graph represents a cumulative sum. You can see the contribution of each data series to the sum and how it changes over time. ("With the category scale a data point should be placed directly above every label on the horizontal scale." Harris 10)

> **Example**: You could examine growth of data centers that make up the cloud, over time, comparing different providers', such as for example Amazon (AWS), Google, Microsoft, ect.

100% Stacked column bar
Displays multiple data series as stacked columns, and the cumulative proportion of each stacked element always totals 100%. It is useful for measuring multiple series as a proportion versus time. For example, use this chart type for displaying the

proportion of a monthly mortgage payment that is applied to interest and principal over time. In this example, the mortgage payment amount represents 100%, while the interest and the principal values are the two stacked elements that make up one column.

Data ink

"Good graphics should include only data-Ink. Non-Data-Ink is to be deleted everywhere where possible. The reason for this is to avoid drawing the attention of viewers of the data presentation to irrelevant elements.The goal is to design a display with the highest possible data-ink ratio (that is, as close to the total of 1.0), without eliminating something that is necessary for effective communication." InfoVis wiki, http://www.infovis-wiki.net/index.php/Data-Ink_Ratio, (Accessed 25.01.2015.)

Source

Harris, Robert, *Information Graphics, A Comprehensive and Illustrated Reference*, Oxford, UK: Oxford University Press, 1999.

What statistical analysis should I use?, Institute for digital research and education (IDRE), University of California Los Angeles, http://www.ats.ucla.edu/stat/mult_pkg/whatstat/nominal_ordinal_interval.htm Accessed 23.01.2015

Tools

Color brewer
http://colorbrewer2.org/#

I mentioned this tool in class that will allow you to study the color combinations you use in your graphics.