Preparing the Data for Visualization

1. ***Develop your research question***

It is important to have a clear understanding of the goal of your research. This will determine what sort of data is needed, the type of analysis necessary, and the types of visualizations that would be most effective to communicate your explorations or findings.

2. ***Get or create your data***

The Library provides:

Access to a large collection of numerical, statistical and geospatial data. There is also a great wealth of open data freely available for download on the web.

Advice and technical assistance with the design, creation, and dissemination of surveys using the Qualtrics web survey platform to assist you in collecting your own data.

Assistance in creating your own data through methods such as digitizing imagery and documents, collecting data from APIs, and scraping data from web sources.

3. ***Clean your data***

This is an essential step to perform before creating a visualization. Clean, consistent data will be much easier to visualize.

Clean data is data that is free of errors or anomalies which may make it hard to use or analyze the data. Starting from a clean dataset allows you to focus on creating an effective visualization rather than trying to diagnose and and fix issues while creating visualizations.

Data cleaning tasks will be very dependent on the dataset that you're working with. In most cases, data cleaning involves:

Removing unnecessary variables

Deleting duplicate rows/observations

Addressing outliers or invalid data

Dealing with missing values

Standardizing or categorizing values

Correcting typographical errors

For more information and best practices for data cleaning visit Clean and Prepare Your Data.

4. ***Choose a chart type***

It is important to pick a chart or graph which will best communicate the message you wish to tell your audience. The first step to choosing a chart is to determine what message you're trying to deliver. Are you:

Showing how variables compare to each other?

Showing relationships between variables?

Showing patterns in the data?

Showing how the whole dataset can be broken down into smaller parts?

Picking one of these can help you to select the chart that will work best for you. There are a number of resources you can use to pick the right chart:

Chart Chooser: This site provides templates to make various charts using Excel and Powerpoint

Data Visualization Catalogue: A useful guide for selecting a chart based on your analysis or communications need.

5. ***Choose your tool***

There are many tools to choose from ranging from freely available open web based tools to licensed desktop tools.

Tools and Resources.

6. ***Prepare data***

The data preparation steps you'll need to go through are determined by the type of chart or visualization you'd like to create and the tool which you have chosen. Once you've selected your chart you may need to perform transformations to create the required data.

Typical data preparation tasks include:

Formatting columns appropriately (numbers are treated as numbers, dates as dates)

Convert values into appropriate units

Filter your data to focus on the specific data that interests you.

Group data and create aggregate values for groups (Counts, Min, Max, Mean, Median, Mode)

Extract values from complex columns

Combine variables to create new columns.

7. ***Create chart***

Following the procedures specific to your tool, the general process for creating charts includes:

Import data into the software

Select the chart type you wish to create

Evaluate the effectiveness of the chart.

Refine by applying design principles. The way in which you design your chart can have a big impact on the effectiveness of the chart. Consider these design principles.

Design Considerations.