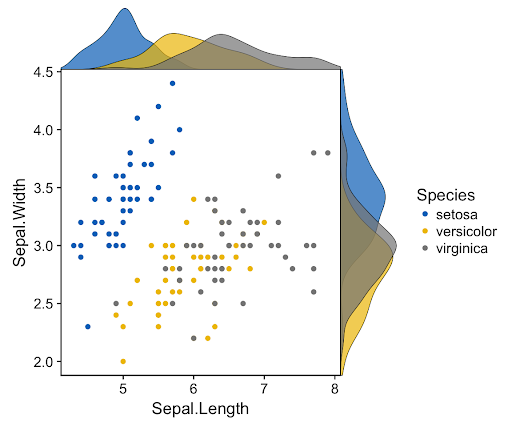
**STATS 506, Fall 2020**

**Midterm Project Proposal**

**Group 2: EunSeon Ahn, Yanyu Long, Tianshi Wang**

Given the high relevance of COVID-19 in 2020 and the impact it has on the vast majority of individuals around the globe, staying informed regarding the current status and spread of COVID around the world has become a daily task for many people. While the severity of the situation necessitates daily monitoring and continuous tracking of the virus, thereby generating enormous amounts of data on a daily basis, this also makes it easy for the general population to be overwhelmed by the overabundance of COVID-related data made available from the scientific community. This is where data visualization can serve the critical role of making large data more accessible and comprehensible to the public at large. In our project, we will explore a few of these different graphical concepts that may facilitate a better understanding of the large amounts of information contained in the dataset.

* Roles
  + EunSeon Ahn - MATLAB
  + Yanyu Long - R
  + Tianshi Wang – Python
  + All – Stata
    - We will attempt to demonstrate the use of same graphical concepts in Stata but may be constrained by the limited graphical capabilities of the software (noted which of the graphical concepts will be attempted in the graphical concepts below)
* Datasets (COVID-19)
  + The COVID Racial Data Tracker: <https://covidtracking.com/race>
  + Coronavirus (COVID-19) Hospitalizations & Preventative Measures: <https://ourworldindata.org/covid-hospitalizations>
* Graphics concepts
  + Marginal Plots (e.g. marginal histogram, marginal distribution plot, or marginal boxplot)
    - All Languages
    - Example: Comparing the # of cases and the # of deaths across different races in United States to compare how COVID-19 may be affecting different races disproportionately ([Dataset](https://covidtracking.com/race))



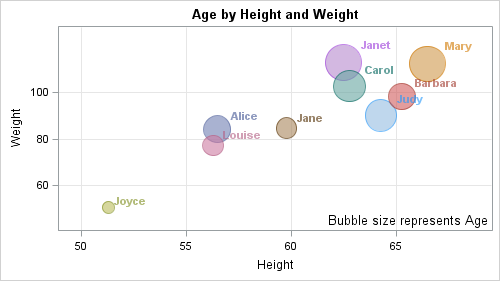
X-axis: # of cases, Y-axis: # of Deaths, with each race/ethnicity represented by a different color

* + Interactive Plots (e.g. manipulating view options with legends)
    - Only MATLAB, R, and Python
    - Example: histogram comparing the # of cases, deaths, tests performed, and hospitalized across a few different countries ([Dataset)](https://ourworldindata.org/covid-hospitalizations). Clicking the legend in the plot will either hide or show the bar associated with those values

|  |  |
| --- | --- |
|  |  |
| A plot showing all series of data | Click the legend to hide some series of data |

For the selected countries, we have all of the following variables available: # of deaths, # of cases, # of tests, # hospitalizations

* + Bubble Plots
    - Possibly in all languages
    - Example: Plotting the # of COVID cases across different countries to study the relationship between COVID preventative measures (face covering & stay-at-home order) and the # of cases ([Dataset](https://ourworldindata.org/policy-responses-covid))



* + - Each bubble will represent a different country with the size of the bubble proportionate to the # of cases or deaths (will decide which makes more sense). X-axis will show face mask enforcement rating. Y-axis will show a stay-at-home restriction rating. Both ratings will be categorical on the following scale:   
      0 - No measures, 1 - Recommended, 2- Required (except essentials), 3-Required (few exceptions)
* Software/Tools
  + R – tidyverse (ggplot2), plotly, ggExtra
  + Python – matplotlib, numpy, pandas, Plotly, Bokeh, pygal
  + MATLAB (no specific libraries required)
  + STATA – graph combine, scatter