R Workshop Day 1

Warmup

- Introductions
- What is your main goal for this workshop?
- What type of data analysis are you most interested in?

RStudio basics

- What are R and RStudio?[1]
- Why might one use RStudio over something like Excel or SPSS?[2]
- Logging in to the RStudio server.[3]
- Type `library(hanoverbase)` in the "console" window and press Enter. We will talk about what this does later.
- The different sections and panes in RStudio.[4]
- Using the console, including help and history[5]
- Basic arithmetic and use of scientific notation in RStudio.[6]
- Assigning to variables.[7]
- Working with vectors/sequences.[8]
- Quickly entering some data using `scan()`.[9]
- Basic statistics for a vector[10]
- Basic graphs for a vector[11]

Analyzing data

- Packages[12]
- Loading data from CSV files[13]
- Loading built-in datasets with `data()`[14]
- Your Rstudio workspace vs your computer[15]
- Loading data from Excel.[16]

[1] ../morsels/whatAreRAndRStudio.html
[2] ../morsels/whyRStudio.html
[3] ../morsels/loggingToRStudio.html
[4] ../morsels/RStudioSections.html
[5] ../morsels/WorkingWithConsole.html
[6] ../morsels/arithmetic.html
[7] ../morsels/variableAssignment.html
[8] ../morsels/vectors.html
[9] ../morsels/usingScan.html
[10] ../morsels/basicStatisticsVector.html
[11] ../morsels/basicGraphsVector.html
[12] ../morsels/packagesAsToolboxes.html
[13] ../morsels/loadingDataCSV.html
[14] ../morsels/builtInDataSets.html
[15] ../morsels/workspaceVsComputer.html
[16] ../morsels/loadingDataExcel.html

- Viewing a loaded dataset[17]
- Basic statistics for variables in datasets[18]
- Basic graphs for variables in datasets[19]
- Restricting a dataset using `filter`[20]
- Viewing only the top or bottom of a list: `head` and `tail`[21]

Practice

1. Which counties have the smallest/largest population?
2. Draw a histogram of the different counties populations. How would we describe the distribution? Restrict the dataset to smaller and smaller ranges to get a better view.
3. What are the mean and median populations for counties? Which one is larger, how does that make sense?
4. The following command draws histograms of the percent of foreign_born on each county for each of three states. What can we learn about the states from this graph?

```
counties %>%
    filter(state %in% c("California", "Indiana", "West Virginia")) %>%
    gf_histogram(~foreign_born, bins=20) %>% gf_facet_grid(state~.)
```

5. Change the above code to instead show the percent of residents that have graduated from high school for the counties in Indiana, Kentucky and Ohio, and compare the states.
6. Do the same, looking at the percent of residents with incomes below the poverty line.
7. Produce a dotplot or barchart with one dot/bar for each state, where the value is the number (tally) of counties in that state. What stands out?
8. What are some popular county names?
9. Create a new dataset `lincolnCounties` that contains only those counties named `Lincoln County`.
10. Draw histograms and calculate statistics for the percent of females in the different counties, and discuss any interesting patterns.
11. Identify counties with extremely small or extremely large percentages of female population. Repeat for some specific state.

[17] ../morsels/viewCommand.html
[18] ../morsels/basicStatisticsDataset.html
[19] ../morsels/basicGraphsDataset.html
[20] ../morsels/filteringDatasets.html
[21] ../morsels/headAndTail.html