R Workshop Day 2

RStudio projects and RMarkdown

- RStudio projects[1]
- Creating an RMarkdown document[2]
- Markdown syntax[3]
- Knitting[4]
- Rchunks[5]
- The GGFormula system for graphics based on a dataset[6]

Practice

- We will be working with the "driving" dataset:
  - Start a new project
  - Start a new RMarkdown document
  - Add a R-chunk to the document, that contains the code to load the packages hanoverbase and ggformula.
  - Download this file[7]
  - Upload the file into your new project space
  - Import the dataset into your RMarkdown document, but on the Import window screen, after the data was appeared, do the following:
    - Select the first column heading (day), and change it to the "Date" format, with format string \%Y/\%m/\%d.
    - For the "leaveTime" and "arrTime" columns, change their format to "Time" with format \%H:\%M. We learned these formats by looking at the actual file in a program like Excel and reading the documentation for col_date (try ?col_date).
    - Now copy the updated code in the import window and paste into a code chunk in your document (set it to show no output but run code). Remember to use the View line only in the console, not in the RMarkdown.
- In a new chunk, use gf_histogram to draw a histogram of the leaveTime variable, and describe and explain the result.
- Draw a similar histogram of the miles variable. What patterns do you notice?
- Draw a scatterplot (gf_point) comparing miles and time, restricting to the part of the dataset where miles is at most 48.

[1] ../morsels/rstudioProjects.html
[2] ../morsels/rmarkdown.html
[3] ../morsels/markdown.html
[4] ../morsels/knitting.html
[5] ../morsels/RChunks.html
[6] ../morsels/ggformula.html
[7] http://hanoverstatslabs.github.io/resources/datasets/driving.csv

More Practice

1. Create a simple RMarkdown report that includes some data and graphics from one of the datasets we have used so far.
2. Choose 3-5 states and filter the counties data to these states. Then draw a scatterplot comparing the populations in 2000 and 2010, with a different color for each state, then add smooth lines for each state as well as overall. Add title and caption to the graph, set the x and y labels, and change the label that goes with the colors.