Class Notes: Week 1

Orientation to Class Resources

- R and RStudio. R does the computations. RStudio provides an interface to R that makes it easier to document your work, to access many of the features of R, and other things.
  - Macalester RStudio Server: . Login with your Macalester email name (e.g. `brosenberg`). The initial password is the last 4 digits of your student ID number. This can be changed. Instructions here
  - Installing on your own computer. You need both of these:
    * R software
    * RStudio desktop software
- Packages you will need:
  - `yaml` & `devtools` through regular installation in RStudio
  - `devtools::install_github( "rstudio/rmarkdown" )`
  - Not today, but soon: `devtools::install_github( "dtkaplan/DCFdevel" )`
- DCF Homepage. Bookmark this page so you can get to it easily.
- The FAQ for the course: http://dtkaplan.github.io/DCF-Course-2014/Notes/FAQ.html

The conceit of this course

A very large part of the use and presentation of data draws on a small set of concepts and techniques that are not difficult individually and that can be taught individually. The complexity of data use and presentation comes from combining these concepts and techniques in various ways to achieve our specific purposes.

We're going to start today with the infrastructure for these techniques:

- the basic structure of data
- commanding the computer to handle data and present the story we want to tell, usually a combination of narration and graphic summary. (There's also modeling, but that's more complicated and won't play a big role in this course.)

In coming weeks, we'll deal with the relationship between data and graphical presentations, the proper form of data to make a graphic (or a model), and how to transfigure the data we have into the form needed for a graphic. We'll also talk about some widely used forms for graphic, to give you some ideas for how you might want to present your own data to achieve your own purpose.

There will be some difficulties — it's not so easy to instruct a computer to do the right thing. But we will try to adopt a work-flow and notation that helps to reduce the inevitable problems.

The basic problem is lack of instantaneous feedback of the sort given by a word processor or web browser. This happens because a (potentially complicated) computer command can't be evaluated until it's complete .

Some examples of contemporary data

Taxicabs and the Shared Economy

A team of mathematicians and engineers has calculated that if taxi riders were willing to share a cab, New York City could reduce the current fleet of 13,500 taxis up to 40 percent. Link to news story and an interactive site with the data.

Medicare Spending

Newspaper article here

Data available here.

DTK notes

Examples from many fields

Infrastructure

RStudio

- Windows, panes, and tabs in RStudio.
- R/Markdown
  - Opening an Rmd file for editing.
  - Saving Rmd files
  - Compiling Rmd to HTML
- Handing in files
  - Upload HTML files to Moodle.
  - Downloading from RStudio server to your desktop so that you can upload them to Moodle.

Creating an Rmd File

Create an Rmd file named `Class-1.Rmd`. Eventually, you will upload your HTML file to Moodle, under In-class, Week 1

Markdown for ...

- Headings, lists, mathematics
- Links, images
- R code chunks

TASK: Create an narrative description of your classes this term. Include links to the Moodle site, links to a relevant Wikipedia (or other) article, and an embedded figure (perhaps from Wikipedia).

Tidy Data

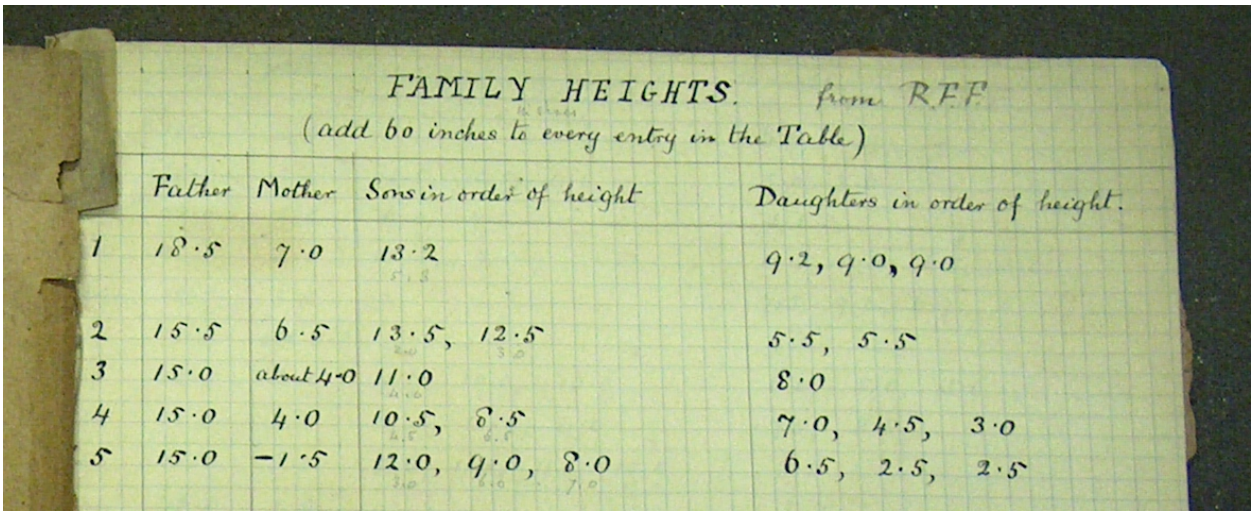
Basics: Cases, Variables, rows, columns, quantitative, categorical

Exercises: Put these into tidy form.

Divide into groups and put your answer to the following in these spreadsheets:Group-1, Group-2, Group-3, Group-4, Group-5, Group-6

Make a separate tab for each table.

Height measurements In the 1880s, Francis Galton started to make a mathematical theory of evolution. Here's part of a page from his lab notebook. Translate it into tidy form.



FAMILY HEIGHTS. from R.F.F.

(add 60 inches to every entry in the Table)

	Father	Mother	Sons in order of height	Daughters in order of height.
1	18·5	7·0	13·2	9·2, 9·0, 9·0
2	15·5	6·5	13·5, 12·5	5·5, 5·5
3	15·0	about 4·0	11·0	8·0
4	15·0	4·0	10·5, 8·5	7·0, 4·5, 3·0
5	15·0	−1·5	12·0, 9·0, 8·0	6·5, 2·5, 2·5

Marital status in the US armed forces Here's the original, untidy spreadsheet.

Back to your Week-1 Document

1. Add in answers to the above two questions, along with links to your group's spreadsheet.
2. Add in answers to the Week-1 drill problems

When done, upload it to Moodle, under In-class, Week 1.

The assignment document for next week

Start on Assignment 1.

- Create an Rmd document named `AssignmentOne-XXX.Rmd` (where `XXX` is your initials).
- Eventually, by class next week, you'll hand it in on Moodle, Week-1 Assignment hand in site.