Assignment 3

Haviland Wright

Version: 19sep16

It's time to take the next steps in R – maybe two steps.

This assignment is about data and also about functions. Like last week, the functions will save you from using loops in your R programs and scripts.

First, you need data. So far, you know one source – R's built-in data. But there's plenty of data.

Take a look at the US dept of Ag – https://quickstats.nass.usda.gov/

When you open this site, look at the upper righthand corner – 34 million records, and climbing. But now select crops, Massachusetts, 2015. Only 592 records. You see farms in Massachusetts, but . . .

Explore this database. It's about food so it's interesting for everyone. See if you can assemble a dataset that looks interesting for analysis. USDA will only let the casual visitor (like you and me) download 50K records at a time, but you may fine that your first dataset definitions result in sizes far less than 50K.

In class we are going to use a dataset that I assembled. Over the next few days we will clean and organize the dataset so that it can be explored and analyzed. The dataset for class is on Blackbaord in Class Notes, Class 7.

Our goal will be to make the dataset "tidy," so you should read Wickham's Tidy article before no latter then Monday afternoon 26Sept16.

The other readers discuss data "wrangling" using commands that manipulate groups of data – apply, dplyr, tidyr

Reading

- Lander: Chapter 11
- Wickham: Tidy Data In the assignment
- RStudio blog:
  - https://blog.rstudio.org/2014/01/17/introducing-dplyr/
  - https://blog.rstudio.org/2016/06/27/dplyr-0-5-0
- Rpubs:
  - https://rpubs.com/bradleyboehmke/data_wrangling

Make sure that you do the examples in the reading. Doing the examples is the most important thing.