R Open Lab - 2016 Oct 19th

During the first 20-30 minutes of this open lab, we will talk about how to merge datasets and filter data using R. The rest of the open lab will be free discussion about any topic/personal project in R among participants, or between participants and the instructors.

Continuing from last week’s discussion about the Wal-Mart Store Sales data set, we will cover these topics for today’s open lab:

**1. How to merge datasets?**

**2. How to filter data (extract data for each store)?**

**3. How to represent categorical data?**

**4. How to handle missing values?**

**Learning with Swirl**

To install swirl, first install R if you haven’t already and open it. In the command line, type

install.packages("swirl")

and hit Enter. You need a working internet connection. Once R has installed the package, you also need to load it. Type

library(swirl)

and hit Enter. Once you do that, swirl will take over and start giving you instructions (and peppy feedback!) to take you through the basics of R. Have fun!

**Topics covered in previous weeks**

Oct. 12 R Open Lab

<https://drive.google.com/drive/folders/0B4uO8w_uGtsCWDgzbElHZ2ZMYlU?usp=sharing>

Oct. 5 R Open Lab

<https://drive.google.com/drive/folders/0B4uO8w_uGtsCazBabFAxcmc4eXc?usp=sharing>

**Digital Social Science Center Blog**

<https://blogs.cul.columbia.edu/dssc/>

**Sign up for future workshops**

<http://library.columbia.edu/research/workshops.html>

**Useful Reference Books**

R in action : data analysis and graphics with R:<https://clio.columbia.edu/catalog/11557863>

R for Statistical Analysis: <http://guides.library.columbia.edu/Rstats>

Bad Data Handbook: <https://clio.columbia.edu/catalog/11548061>

**Data source**

Walmart Store Sales Forecasting: <https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting>

**Next Session:**

car package