Expository Data Analysis with R (ECON-181)

## About this Course

Conducting economics research requires both an understanding of theoretical concepts, and practical knowledge of empirical methods. Today most empirical work in conducted using statistical programming languages such as Stata and SAS In this course, students will learn how to use one such language, R, as a means of building their empirical toolkit. R has become one of the leading languages in data science and statistics and therefore a very valuable programming language to learn. R is a free software environment used by a variety of fields including economics.

## Learning Objectives

* Create, read, modify, and store R datasets
* Use available R packages and write custom functions
* Create figures and plots
* Perform efficient dataset manipulation
* Perform and interpret multiple linear regression
* Create, manage, and share reproducible project files using R markdown

## Course Structure

* **Lectures** -- Lectures will be held Fridays from 9:00 am to 12:00 p.m. Class time will include lectures with Federal Reserve Board staff.
  + ***Class will meet in the Federal Reserve Board’s building at 1801 K-Street, NW. Washington, DC.***
* **Office Hours** -- TAs will hold office hours once a week in the Howard University Economics department for two hours from 6-8pm.
  + Additional office hours will be held from noon-1pm immediately after lecture
* **Github Site** -- All the lectures and homework will be posted to the [github](https://github.com/wampeh1/Ecog314_Spring2017) site.
* **Piazza Site** -- We will use a [piazza](https://piazza.com/federal_reserve_boardhoward_univeristy/spring2017/ecog314econ181/home) site as our course wiki.
  + You will be asked to create a Piazza account as part of your first homework
  + All work submission and all communication with staff should take place via Piazza

## Course Prerequisites

All applicants must have completed a college level course in Econometrics with a grade of **B or higher**. No prior training in programming or data science is required.

## Computer

A Windows or Mac laptop is required with the following minimum configuration: 4 GB RAM or higher; 320 GB hard disk; configured to allow the installation of R and RStudio software. \* You should have installed R and Rstudio prior to the first day of class \* A limited number of loaner laptops will be made available if needed, **for in-class use only**.

## Software

R and selected R packages will be the primary software for this class. R is free. Substantial instruction will be provided in lecture notes and assignments, and additional instructions will also be available in the online reference materials.

Link to Download R: [Comprehensive R Archive](http://cran.us.r-project.org/)

RStudio is the recommended R integrated development environment

RStudio Download: See <https://www.rstudio.com/products/rstudio/download/>

RStudio is easy to install and the installation does not require any instruction. However, the following links provide additional setup and navigation guidance:

<http://web.cs.ucla.edu/~gulzar/rstudio/index.html>

<http://dss.princeton.edu/training/RStudio101.pdf>

<https://support.rstudio.com/hc/en-us/sections/200107586-Using-RStudio>

## Textbook

*R for Data Science* by Garrett Grolemund and Hadley Wickham -- The R4DS textbook is [freely available online](r4ds.had.co.nz) and provides an excellent practical introduction to using R for data science. The book is written by Hadley Wickham, the author of many packages used in the course and provides insight into how to most effectively utilize R for data analysis.

## Grading

Numerical class grades will be based on:

* Homework (40%)
  + Assignments will be due by midnight the Wednesday after the class they are assigned.
  + Late assignments will lose 10% of the total for each day late.
  + If all assignments are turned in on time, lowest will be dropped
* Participation (5%)
  + Good class citizenship (helping other students, contributing to the piazza site)
* Attendance -- 2% dropped from final grade per missed class (first absence is excused)
  + If you miss 2 or more of the first five classes you may be dropped from the course
* Take-home Exam (20%)
  + This will be a take home exam given around the midpoint of the semester
  + This will be an individual assignment where you will not be able to ask questions of the TAs.
  + There will be no office hours while the mid-term is live.
* Final project check in and the final project (35%).
  + Separate syllabus to follow

The instructor reserves the right to amend weighting.

## Lecture dates

* The course will meet each Friday of the semester at 9:00 am at the 1801 K street building of the Federal Reserve
* We currently plan to hold lecture on Friday March 2 (Charter Day Convocation) to make up for lost class on January 12