ECON/DCS 368 Week 1 Notes

Lecture 1: Introduction

- Research fields of Prof. Coombs is Public and Labor Econ, but is interested in econometrics and data science
- One goal of this course is to help students with their coding skills and have only one string of code that needs to be run when the time to write thesis comes; this class fills in the gaps left by econometrics and methods classes
- Expect this course to be extraordinarily challenging as we will be teaching ourselves new skills that cannot be covered in 12 weeks, but it should also be a rewarding class
- Grading
    - Make-up
        - 7 homework assignments (top 5 graded) = 50% of grade
        - 2 Short presentations (best graded) = 10% of grade
        - 1 Final Project = 40% of grade
    - Extensions: You receive 3 grace periods (details in syllabus)
- Homework
    - Assigned and submitted via GitHub
    - 2 of the assignments will be group assignments and the remaining are individual
    - Submit homework in a .tex or .Rmd file only
- Final Project
    - 5-10 page research paper
    - Will be broken into 6 different components
    - See Christine Murray in library for finding potential data sources as well as Kaggle online
    - See more details in syllabus
- Data science is the scientific discipline that deals with transforming data into useful information using a variety of states/ML techniques
- Pillars of Data Science
    - Programming
    - Visualization
    - Cousin\asl inference
    - Machine learning
    - (Calc and Stats foundations are necessary)
- Big Data can be…
    - Wild data (unstructured such as twitter, unlike census data, which is organized)

    - Wide data (K>N, customer data sets where each click is a variable) (Good for prediction. I.e. predicting income on target ads)
    - Long data (good for identifying causal effects. I.e. effects of improving schools on income)
    - Machine learning allows for computers to learn for themselves without explicitly being programmed. AI is constructing machines to think like humans and is a subset of machine learning
- We are encouraged to use ChatGPT and GitHub CoPilot
- Causal inference requires a model
- Software installation and registration
    - Download R
    - Download RStudio
- OS-Specific Extras
    - Windows: install Rtools and Chocolately
    - Mac: install homebrew
- Checklist
    - Make sure you have the necessary software installed
    - Make sure all of this software is up to date
- Next week: we will check and make sure everything is working
- Review of R
    - Why R and RStudio?
        - R is free
        - R is useful in the job market
        - R is THE data science software
        - Economists that don't use stata use R
        - It is smart to learn multiple coding languages and it gets easier the more languages you learn
- (See lecture slides for a R Code example for constructing a regression)
- Ggplot2 is a fantastic data analysis tool to use in your code
- Example Activity: Do increases in GDP cause life? (btw: the gapminder dataset contains panel data on life expectancy, population size, and GDP per capital for 142 countries since the 1950s)
- Once the packages are installed, leopard them into your R session with the library() function
- When selecting a country to examine, you must consider country specific business cycle issues, changes in government, changes in how accurate the government provided data is, etc.
- Elements of ggplot2
    - Hadley Wickham's ggplot2 is one of the most popular packages

    - Aesthetic Mappings: aes(x = gdpPercap, y = lifeExp) (see lecture slides for full code) Note: Aesthetics must be mapped to variables, not descriptions!!!
    - Geoms: once your variable relationships have been defined by the aesthetic mappings, you can invoke and combine different geoms. They only accepts a subset of mappings, such as geom_density (note that geom_density cannot take a y variable)
    - There are ways that you can change the appearance of graphs, so you can use the aesthetics of fivethirtyeight or harvard business for example
- Correlation v. Causation
- Next lecture: Deep dive into GitHub