Week 1. Course Introduction

Assignment: Very Basic Data in R (Due Week 2)

Prof. Jack Reilly

F2025

Readings

Lecture: Course Introduction

- Recommended reference reading:
  - **DV**, chapter 1
  - **RDS** chapter 2
  - **FCSP** chapters 1-2
    - *We will not cover all of this material in ch 2 this week; don't get too caught up on things not in course slides or course R scripts*
  - **DMSS** chapters 1-2
    - *Ignore, for now, things that relate to Database Management Systems (DBSMs) or SQL*
  - **The Plain Person's Guide to Plain Text Social Science**, ch 1-2 (https://plain-text.co/index.html)
    - *Note: The Plain Person's Guide is a little out of date techinically - some of the tools discussed have newer, and better, versions - but the overlying principles and philosophy are what you should pay attention to*

Data & Computational Work

Last time, you installed software. This time, we're going to begin to use it.

Submit: .R File 1

(If you did not submit this last week)

In an .R file, write code to answer the following questions. Make sure your file is appropriately titled and headered.

1. Create an object named `aardvark` that stores a 3 as a single number
2. Create a second object named `boomba` that stores a 6 as a single number
3. Create a third object named `centauri` that is the addition of `aardvark` and `boomba`
4. Create a fourth object named `diabolical` that is the multiplication `aardvark` and `boomba`
5. Create an object named `ebullient` that stores three numbers as a vector: 4,5,and 6
6. Create an object named `fastidious` that stores three numbers as a vector: 8,9, and 11
7. Add `ebullient` and `fastidious` together, and store it in an object named `george`
8. Find the mean (average) of `fastidious`, and store it in an object named `zoinks`

Submit: .R File 2

Create an R file to write all of your code in. Make sure this file is properly titled and headered, etc. Write code that answers the following:

1. Roll a 12 sided die 50 times. Store the result in an object `roll`. What is the mean (average) of your 10 die rolls?
2. Create a object named `trusttheforce` that stores the following information:
    - Information for five Star Wars characters: Luke, Han, Leia, Vader, Rey
    - One string variable storing their name: `name`
    - One boolean (true/false) variable indicating if the character can use the force (F - Han, else - T): `force`
    - One boolean (true/false) variable indicating if the character is a Sith (T - Vader, else - F): `sith`
    - One numeric variable indicating the total number of Star Sars movie the character appeared in as a non-baby, non-ghost character (Vader/6, Luke/5, Leia/6, Han/4, Rey/3): `movies`
3. Draw a histogram of `trusttheforce$movies` showing the distribution of number of movies
4. Let's take the Obi-Wan perspective, and say it only counts for "Vader" to show up if he's actually wearing the suit after "betraying and murdering" Luke's father.
    - Replace "Vader's" number of movies with 4.
    - Draw another histogram of `trusttheforce$movies` showing the distribution of number of movies according to Obi-Wan

5. Load `npsvisitation0910.csv`. It has three variables: `park`, `year2009`, and `year2010`, with the second two giving the number of visitors to the park in question in 2009 and 2010.

    - What is the average number of park visitors in 2009? 2010?
    - Draw a histogram of the number of park visitors in 2009.

6. From the list of parks, identify Yellowstone National Park and Mount Rainier National Park. (They are in rows 359 and 235).

    - How many more visitors did each park, individually, have in 2010 compared to 2009? (Subtract one cell from the other)
    - What is the difference in the number of visitors between those two parks in 2009? (Subtract one cell from the other)
    - How many visitors did the parks have, combined, in 2010? (Add one cell to the other)

7. **BONUS**:[1] Roll one twelve sided die 1000 times. Then, roll *two* six sided dice (together) 1000 times. Add the two six sided die rolls together, so you can get a "combined" roll that goes from 2 to 12 (the way you would in a normal board game). What is the mean of the twelve sided die rolls and two six sided dice rolls? How do the *distributions* (histogram) of these two rolls differ?

Submit this .R file to Blackboard.

Submit: PDF file

Answer the following questions and upload as a PDF to Blackboard.

1. Include your histogram showing the distribution of movies. Include both the "normal" version and the "Obi-Wan" version.
2. Include your histogram of number of park visitors in 2009.
3. How many visitors did Yellowstone and Rainier have in 2009 and 2010, respectively?
4. **BONUS** If you did the bonus problem, include your histograms for your two die rolls above.

[1] This bonus is for no actual bonus points - it's an optional problem to help you begin thinking your way through code.