Week 5 Overview

As mentioned in the previous week – we are sticking with data transformations for another week. This will really cement your R knowledge and help you with your final project and other future assignments. A lot of these functions are very common and will be necessary for you to complete your project. It is also a great introduction into what you can expect in your Data Prep course. If you haven't already, start thinking about your final project that was outlined the first week. Time will fly by, so it is a good idea to get a head-start on what you want your project topic to be.

Contents of the Week

Overview

Readings, Assignments and Tasks

5.1 Discussion/Participation

5.2 Exercise

Objectives

After completing this week, you should be able to:

- Understanding and application of population R packages for transforming data

- Data formats and data structures

- Know when and how to clean data

Weekly Resources

Comprehensive R Archive Network

R Studio

Sage Publications. (2021). Discovering Statistics Using R

RStudio, PBC. (2021). RStudio Documentation

RStudio, PBC. (2021). R Studio Cheatsheets

Week 5 Readings, Assignments, and Tasks

Here are your tasks for this week:

Read the following:

- *R for Everyone:* Chapters 12 – 16

Complete the following:
- 5.1 Discussion/Participation
- 5.2 Exercises

5.1 Discussion/Participation

Here are optional topics for discussion via Teams this week. Remember, these topics aren't required, but if you are struggling to know what to post about, these can be used to initiate discussion!

1. What is the dplyr package for? What uses does it have?
2. Why does order of loading packages in R matter?
3. What are pipes?
4. What is the purrr package?
5. When does data need to be re-arranged from column oriented to row oriented and vice versa? When does it need to be combined into one dataset or split into multiple datasets?
6. What are cbind and rbind functions?
7. How are joins handled in R?
8. How do you convert from long/wide formats?
9. How do we work with strings? Or how do we split strings? When do we need to do these transformations?

5.2 Exercise

1. Using either the same dataset(s) you used in the previous weeks' exercise or a brand-new dataset of your choosing, perform the following transformations (Remember, anything you learn

about the Housing dataset in these two weeks can be used for a later exercise!)

a. Using the dplyr package, use the 6 different operations to analyze/transform the data - GroupBy, Summarize, Mutate, Filter, Select, and Arrange – Remember this isn't just modifying data, you are learning about your data also – so play around and start to understand your dataset in more detail

b. Using the purrr package – perform 2 functions on your dataset. You could use zip_n, keep, discard, compact, etc.

c. Use the cbind and rbind function on your dataset

d. Split a string, then concatenate the results back together

Submission Instructions

For all assignments in this course, you must export the script or Markdown file to PDF. You are welcome to submit your URL to GitHub in addition, but all submissions must include a PDF (no zip files will be accepted either).

The assignment is due by Sunday, 11:59 p.m. CT.