

Group S: Milestone #2

PhuongTseng & Carri Beshears

2022-10-02

This is a team assignment; each team should complete and turn in a PDF created from an Rmd via Github. Please include code and output for the following components:

Description of dataset

What is the data source? (1-2 sentences on where the data is coming from, dates included, etc.) How does the dataset relate to the group problem statement and question?

ANSWER:

Import statement

NOTE: Please use datasets available in the PHW251 Project Data github repoLinks to an external site. (This is important to make sure everyone is using the same datasets)

Use appropriate import function and package based on the type of file Utilize function arguments to control relevant components (i.e. change column types, column names, missing values, etc.)

Document the import process

- Identify data types for 5+ data elements/columns/variables
- Identify 5+ data elements required for your specified scenario. If <5 elements are required to complete the analysis, please choose additional variables of interest in the data set to explore in this milestone.
- Utilize functions or resources in RStudio to determine the types of each data element (i.e. character, numeric, factor)
- Identify the desired type/format for each variable—will you need to convert any columns to numeric or another type?

Provide a basic description of the 5+ data elements

- Numeric: mean, median, range
- Character: unique values/categories
- Or any other descriptives that will be useful to the analysis