Assignment 3

Getting Data (LMS Module 4)

Lacey Hartigan

due 6/22/21

For this assignment, you'll need to open up, clean, and save the following datasets using the tools we've gone over in class. For each dataset, make sure that when you're done you have a nice, neatly labeled dataset that would be easy for you or another analyst to open and analyze. Save the result in an `RData` file using the names provided.

For this submission, you need to upload the following SIX files:

- [File 1] An .Rmd file (preferably named `04-assignment_<lastname>.Rmd`) that cleanly creates all of the datasets requested
- [Files 2-6] One copy of each saved R dataset you created.

NO KNIT file is required for this submission.

PLEASE NOTE: You can't just copy these links into your code. You need to figure out the format for the data. Take a look at the website to figure out what kind of data is there; then you can download the dataset and use the appropriate code in R to import it. Additionally, a CLEAN dataset means one that could be easily opened by another analyst and used immediately. You should tidy up as needed.

Packages you will need:

- tidyverse
- haven

1. Wine data. Save as file name `wine.Rds`

http://archive.ics.uci.edu/ml/machine-learning-databases/wine/

2. 2016 Health Education District Data

https://www.cdc.gov/healthyyouth/data/shpps/data.htm

3. County level replication file for "Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States"

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BT3LXD

4. Airline dataset

http://www.principlesofeconometrics.com/sas.htm

5. King county births

http://courses.washington.edu/b517/Datasets/datasets.html