Data Dictionary

Dataset 1: Trump's Tweets (2020-01-20 to 2020-09-01)
Source: http://www.trumptwitterarchive.com/archive

Variable Name	Variable Description
id_str	Unique identifier for the tweet
source	Was this tweet posted via an iPhone, Android, the Twitter web client, or something else?
text	Text of the tweet
created_at	Date and time that the tweet was posted
retweet_count	Retweet count of the tweet (at the time of the data collection)
favorite_count	Favorite (heart) count of the tweet (at he time of the data collection)
is_retweet	Is the tweet a retweet of another tweet
sentiment_score	Sentiment score using the AFINN dictionary
urls	URLs that were embedded in the tweet
hours	The hour that the tweet was posted (00 = midnight, 12 = noon)
timeofday	Time of the day that the tweet was posted (06-11 am = morning, 12-4 pm = afternoon, 5-8 pm = evening, 9pm-5am = night)
year	Year that the tweet was posted

Use this dataset to:

1. Practice looking for and fixing "dirty" values in vectors (e.g., a character value in what should be a numeric vector)
2. Practice making decisions about what time structure to use
3. Practice turning factors into numerics

Bonus thing to learn: Get a head start on the NLP class and learn how to work with text data in the "text" / column)! You can use the "text" variable to work through some of the first few chapters in the TidyText book: https://www.tidytextmining.com/index.html

Dataset 2: COVID-19 Public Opinion Data (Marist)
Source: https://ropercenter.cornell.edu/ipoll/study/31117235

Note: There are 53 variables in this dataset, but the ones below are the ones (at minimum) you should focus on.

Variable Name	Variable Description
TRUDP105R	Do you approve or disapprove of the job Donald Trump is doing as president? [And, would you say you strongly approve/disapprove of the job he is doing or just approve/disapprove?]
CNRNCMCV1	Are you very concerned, concerned, not very concerned, or not concerned at all about the spread of coronavirus to your community?
AGEEPWT	Age Group
AGER	Age
GENRATNX2	Generation Cohort
PARTYID	Party Affiliation
collegep	College Grad
TRSTCV1A	Do you trust the information you hear about coronavirus from President Trump a great deal, a good amount, not very much, or not at all?
TRSTCV1B	Do you trust the information you hear about coronavirus from the News Media a great deal, a good amount, not very much, or not at all?
TRSTCV1D	Do you trust the information you hear about coronavirus from public health experts a great deal, a good amount, not very much, or not at all?

Use this dataset to:

4. Learn how to clean public opinion data
    a. Removing unused factor levels
    b. Re-leveling
    c. Factor → Numeric
5. Practice regression models with different dependent variables

Bonus thing to learn: Play with more variables not listed above!

Dataset 3: Reform/Defund/Abolish the Police Facebook Data (2020-03-01 – 2020-09-08)
Dataset: CrowdTangle

Note: There are 29 variables in this dataset, but the ones below are the ones (at minimum) you should focus on.

Variable Name	Variable Description
Likes at Posting	Number of likes that the page had at the time the post was made
Type	Type of Facebook post (Video, Link, Photo, etc.)
Likes	Number of likes that the Facebook post received
Comments	Number of comments that the Facebook post received
Shares	Number of times that the Facebook post was shared
Love	Number of "love" reacts that the Facebook post received
Wow	Number of "wow" reacts that the Facebook post received
Haha	Number of "haha" reacts that the Facebook post received
Sad	Number of "sad" reacts that the Facebook post received
Angry	Number of "angry" reacts that the Facebook post received
Care	Number of "care" reacts that the Facebook post received
Video Share Status	Is there a video and, if so, is it original or from a share?
Link	What is the link in the post (if there is one)
Total Interactions	Total number of interactions with post
BLM Page	Is the page that the post was on a self-identified BlackLivesMatter group?

Use this dataset to:

1. Practice looking for and fixing "dirty" values in vectors (e.g., a character value in what should be a numeric vector)
2. Work with lots of numeric values
3. Learn how to use (and "fix") variable names with spaces

Bonus thing to learn: Turn created_at into a "date" object! Learn more about datetimes through the R4DS textbook (https://r4ds.had.co.nz/dates-and-times.html) or this tutorial from Cole Beck (http://biostat.mc.vanderbilt.edu/wiki/pub/Main/ColeBeck/datestimes.pdf).