If you are looking for a dataset for your final project, consider these websites:

- https://www.openintro.org/data/
  - This is a list of datasets used in our textbook. You can read them into R using the `read.csv(...)` command. The links in the middle column show the data documentation, where you can learn more about the dataset.
- https://vincentarelbundock.github.io/Rdatasets/datasets.html
  - This is a list of datasets that are publicly available, either via different R packages or via reading the link using `read.csv(...)`. The "DOC" links show the data documentation, where you can learn more about the dataset. I frequently use this page to find datasets for our course notes, homeworks, labs, etc.
- https://dasl.datadescription.com/
  - This is another list of datasets, this time sorted by different statistical methods. You can read the datasets in using `read.csv(..., sep="\t")`–if you are having trouble getting the data loaded into R, let me know. I also use this page fairly frequently.
- https://data.worldbank.org/
  - This website has several different datasets on global development. You can sort by country or by topic (e.g., population, poverty, etc.). I have looked up statistics on this site, but I have never downloaded a dataset for use in class. This means that I can't promise the data will be useful! Please also be careful that you are not relying on time series data, which violates the assumption of independence we need for both linear and logistic regression. If you are interested in looking at coronavirus-related datasets, you might find them here.
- https://www.kaggle.com/datasets
  - Kaggle is a website that hosts different data-related challenges. These challenges are often focused on machine learning, so the datasets might be best suited for methods beyond the scope of our course. I have used some of these websites for other projects, but never for teaching–some of them are not "clean" (e.g., missing values, duplicated columns, etc.), so it might require some data wrangling, which may make them even more complicated. If you are interested in looking at coronavirus-related datasets, you might find them here.
- https://cloud.google.com/blog/products/data-analytics/publicly-available-covid-19-data-for-analytics
  - If you want to analyze coronavirus-related data, this is your site! I haven't used any of these datasets, but they might be interesting. However, they may also require some data wrangling, which might put them beyond the scope of our class.
- https://github.com/awesomedata/awesome-public-datasets
  - This is a nice list of datasets sorted by topic (e.g., agriculture, biology, etc.). I've never used any of these datasets, and furthermore, the datasets look like they come from a variety of different sources. They may or may not be appropriate to use for the project, and they may or may not be clean.
- https://www.reddit.com/r/datasets/
  - This is probably the least organized of all of the resources, but the datasets do look really interesting! I have never used a dataset downloaded from Reddit, but I did some browsing and it looks like at least one person packaged their data for R. Similar to the Awesome Public Datasets link, I cannot vouch for how appropriate the datasets might be to use for the project, or how clean they are. If you are interested in looking at coronavirus-related datasets, you might find them here (pinned post at the top).