VAERS
Have to query the data ourselves. The system limits the output to 10,000 rows. May not have enough numeric columns and I'm not sure what we would analyze, but I'm beyond myself that it's publicly available. Technically time series... but we do not have to treat it as such if we don't consider time as a variable.
https://wonder.cdc.gov/controller/datarequest/D8

Statistical "Which Character are You?" Personality quiz
2.5 million rows, open data from the ultimate what character are you personality quiz that was huge on Twitter/social media last summer. May be pretty fun.
https://openpsychometrics.org/tests/characters/

Mental Health of Young People in England
Has 2000+ variables, had a bunch of kids screened for a variety of common disorders, parent interviews, demographics, etc.
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8467#!/documentation

Covid-19 and Sleep
A ton of variables, multiple measures, we would probably have to score the data though :/ and their documentation is complicated af
https://osf.io/ptykj/

Airplanes that Hit Birds
Simple! And keeps with the "every final project is about airplanes" theme
https://www.openintro.org/data/index.php?data=birds

Data on Whether or not the perceived "hotness" of a Professor Affects Their End of Term Evaluations (omg) ✿
I…. honestly this is my favorite so far I think this is so funny
https://www.openintro.org/data/index.php?data=evals

Machiavellianism (one factor in the Dark Triad) Predicted by Demographic factors
https://openpsychometrics.org/_rawdata/

Candy popularity
Just about predicting most popular candy, has a ton of quant/qual variables
https://www.kaggle.com/fivethirtyeight/the-ultimate-halloween-candy-power-ranking/

Cat Ownership and the GSS ✿
… so, in the 2018 survey there's a variable for 'owns a cat'. So, we pick like 10-15 other variables that we hypothesize may be related to cat ownership and try and predict cat ownership from them for the logistic regression. For the linear regression, we can try and predict… idk… maybe something completely unrelated for fun.
https://gssdataexplorer.norc.org/
wow so many variables @_@

Speed Dating Experiment
What characteristics (there's a ton) made someone want to see their data again (categorical) and led to scores for likeability (participants were given scores at the end, I think they're on a likert scale so we can take the mean of them to create the likeability variable which would then be continuous)
https://data.world/annavmontoya/speed-dating-experiment

Very Boring Data -

Heart Attack CSV
Just grabbed this from Kaggle, it won't require cleaning and it fits the data description for the project.
https://www.kaggle.com/andrewmvd/heart-failure-clinical-data

Cereal Data
Has data on 80 cereals, though lacking on categorical variables.
https://www.kaggle.com/crawford/80-cereals