IDEAS:

Audio Classification

<https://search.datacite.org/works/10.5281/zenodo.831188>

<https://zenodo.org/record/831189#.X6mGjZNKhBw>

<http://millionsongdataset.com/>

This project will analyze music data. The two sub-ideas within this topic are either to analyze the musical elements of songs and find what elements are most important to song popularity, or to classify each song by artist. This one may be hard; I have never worked with audio data.

Sports outcome prediction (regression/classification)

<http://nflsavant.com/about.php> (maybe try putting plays into graph format?)

<https://github.com/ryurko/nflscrapR-data>

The ultimate goal of this project is to predict the final score of American football games. American football was chosen because it is very popular, and data is plentiful. The outcome of games should be predictable because it depends significantly on human actions.

News bias (sentiment analysis)

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GMFCTR>

<https://webhose.io/free-datasets/english-news-articles/>

<https://www.kaggle.com/harishcscode/all-news-articles-from-home-page-media-house>

<https://www.kaggle.com/snapcrack/all-the-news/>

In this project, many news articles will be parsed and checked for sentiment. News articles should be even-handed and unemotional, and so should express little sentiment. Therefore, I would argue that news providers whose articles regularly express strong sentiment, regardless of what that sentiment is, are more biased than news providers whose articles are less sentimental.

Painting Classification

<https://www.kaggle.com/ikarus777/best-artworks-of-all-time>

<https://www.kaggle.com/ipythonx/van-gogh-paintings>

This project will create a model that can identify the artist behind a painting from a picture of that painting. Ideally, it will also be able to detect stylistic features of paintings and say, of a new painting, “This painting most resembles the work of [artist]”.