

# Resources For Data

## Introduction to Data Science

Fall 2018

The following is an extremely incomplete list of options for accessing publicly available data. This list will likely continue to grow with time.

**Warning:** The resources listed may be greatly varying in quality, and some data sets found in the provided links may be very challenging to analyze.

- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/index.php>
- Kaggle <https://www.kaggle.com/> this resource provides not only a variety of interesting data sets but also many other resources for data science.
- Publicly available government data <https://www.data.gov/>
- Scientific weather/climate data <https://www.climate.gov/maps-data/datasets>
- Many R packages provide access to interesting data sets. For example:
  - R data sets package `datasets`
  - `Lahman` a package with baseball data
- Gapminder data sets <https://www.gapminder.org/data/> provides data centered around global social, health, and economic issues. Also available as a R package `gapminder`
- FiveThirtyEight data <https://github.com/fivethirtyeight/data>, also available as an R package `fivethirtyeight`
- This site provides links to public data and is sorted topically, <https://github.com/awesomedata/awesome-public-datasets#biology>. The data found by following links on this site can vary greatly in quality.
- These articles provide links to free public data sets on a variety of topics: <https://www.springboard.com/blog/free-public-data-sets-data-science-project/>, <https://www.dataquest.io/blog/free-datasets-for-projects/>, <https://goo.gl/MUzxNB>