

# CAP 5300, Project: Data Description.

Provide a detailed description of the data you've selected for your project, including

1. The source and inspiration for selecting this particular data set. If you don't have any data preferences off the top of your head, some examples of data sources are *Kaggle.com*, UCI Machine Learning repository, *sports-reference.com*, among many others.

**Note: It shouldn't be among data sets that we've already used in class, or that are contained in any of the *R* packages. It should be obtained from external source (most likely a *.csv* file, or via web scraping).**

2. The size of the data (# of observations and # of variables).
3. Description of all variables (similarly to descriptions you encounter for *R* data sets, e.g. <https://stat.ethz.ch/R-manual/R-devel/library/MASS/html/Boston.html>)

Make sure that your data set contains:

- A total of at least 10 variables and 50 observations.
- Several numerical variables, one of which may act as a response variable in linear regression (your main variable of interest).
- At least one categorical variable.

The report should be

- **no more (!) than 2 pages long.**