**Project Title:**

**Application domain and datasets (150 words):**

The application domain will be focused on boxing, specifically the top 50 male boxers within each weight category (heavyweight to minimumweight).

The data sets I am using I have scraped from [www.boxrec.com](http://www.boxrec.com), the largest boxing related database on the internet.

The first data set has details on 793 rows and 20 columns of data comprising of data such as win/loss/draw as well as information on debut date, division, division rating, and birthplace. Unfortunately, after scraping the data it became clear there are fields with missing values, however I have replaced these values where reasonable (such as missing reach data).

The second data set has also been scraped from [www.boxrec.com](http://www.boxrec.com) and contains information on every bout of the respective fighters in the first data set, from their first to their most recent bout, including data such as outcome (by method of victory or loss/round) and locations.

**Well-motivated analytical questions (150 words):**

What are your analytical questions and what is your motivation for answering them?

**Plan (maximum 200 words):**

What is your plan? Include data processing, data derivation, model building and validation. The aim of your plan is to address your research questions.

**Specific questions to us**

1. Considering that a person’s reach is generally related to their height, is it reasonable that I am calculating the average difference between height and reach per division, and replacing missing reach values with their height + that average difference value?
   1. Similarly, I have calculated a mean height value by division and used this for missing values, however I think this could lead to inaccuracies.
2. X
3. X
4. X
5. X

**Initial investigations**

Provide your initial exploratory work on your data that informed your plan