Problem Set 1
Advanced Methods in Applied Statistics

Magnus Berg Sletfjerding

February 12, 2019

1 Exercise 1

For the first part of the assignment, I used scrapy (`https://scrapy.org/`) to parse the data on the given website to a csv.

The main challenge lay in figuring out the fact that teams without a seed had two (!) fewer columns. This took longer than I care to admit, but I extracted all the data into csv files. (attached)

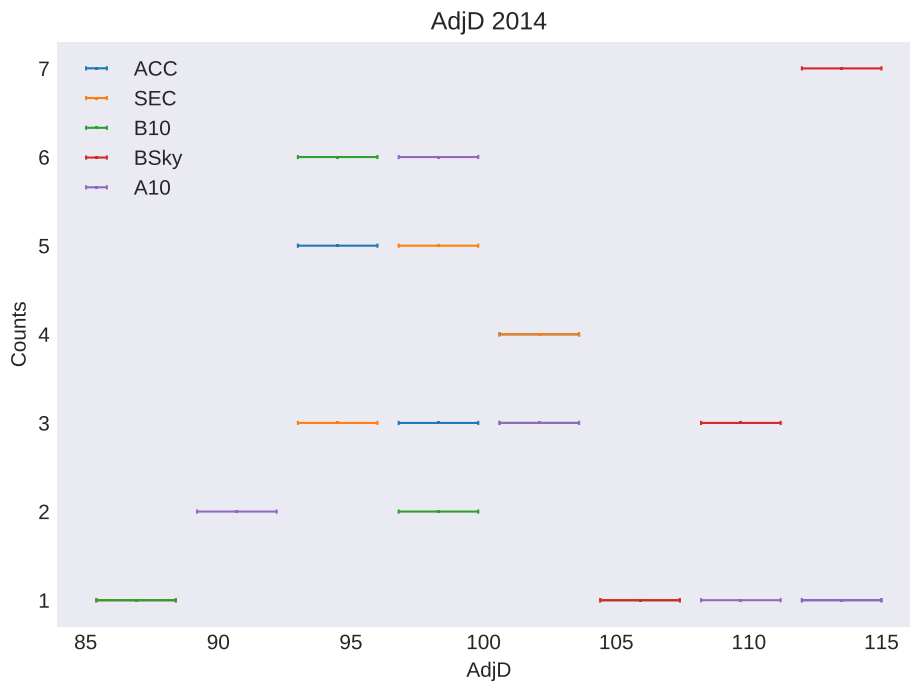I treated the data with pandas dataframes, and plotted the data in figure 1.

Figure 1: Exercise 1 - histogram of all the teams

2 Exercise 2

Importing this csv took three clicks. Thanks for the tip on code robustness, Jason.

With Pandas, subtracting the two columns was simple enough. Eliminating NaN values was also straightforward.

This resulted in figure 4.
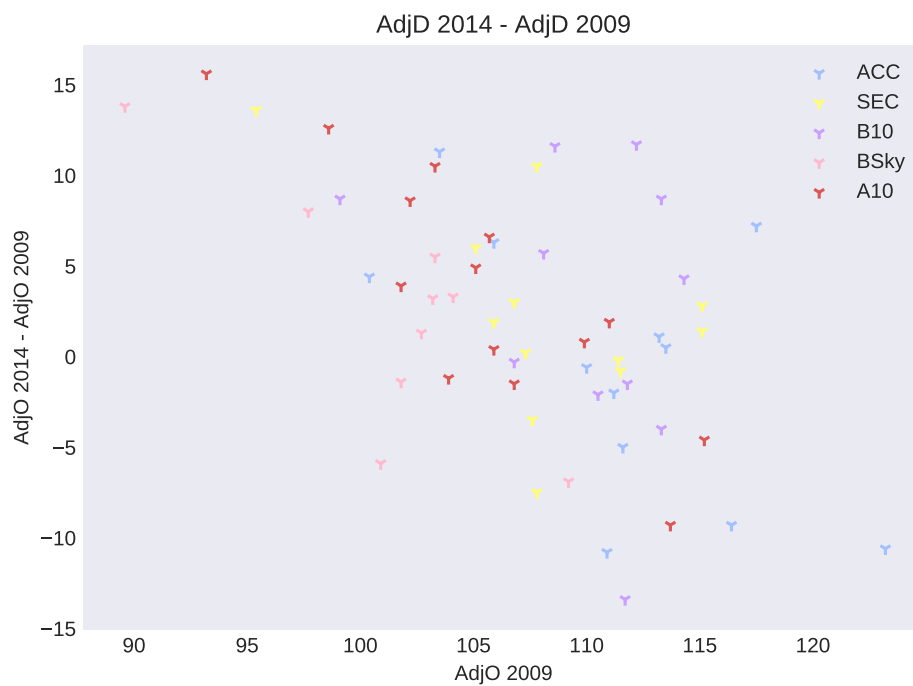
AdjD 2014 - AdjD 2009

Figure 2: Exercise 2 - plot of the AdjO of all the teams as fcn of 2009 AdjO

Pandas' mean and median functions also came in handy here, and I generated the following .txt file.

AdjO-diff-Md is Median, while AdjO-diff-Mn is Mean

```
Conference    AdjO diff-Md
ACC                  -0.05
SEC                   1.65
B10                   4.30
BSky                  3.20
A10                   2.90
Outsiders             1.30
Conference    AdjO diff-Mn
ACC                  -0.63
SEC                   2.28
```

```
B10                2.67
BSky               2.32
A10                3.51
Outsiders          2.40
```

3   Exercise 3

Again, Jason's tip about robust and scalable code came in handy.
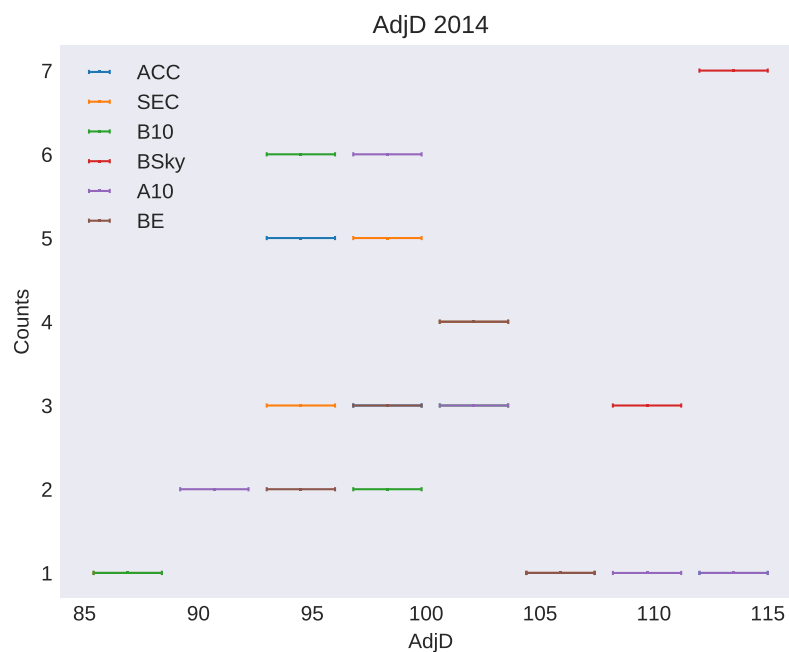
3.1   Exercise 1 redone

Figure 3: Exercise 3-1 - histogram of all the teams

3.2   Exercise 2 redone

Another text file was generated:

```
Conference    AdjO diff-Md
ACC                  -0.05
SEC                   1.65
B10                   4.30
BSky                  3.20
```
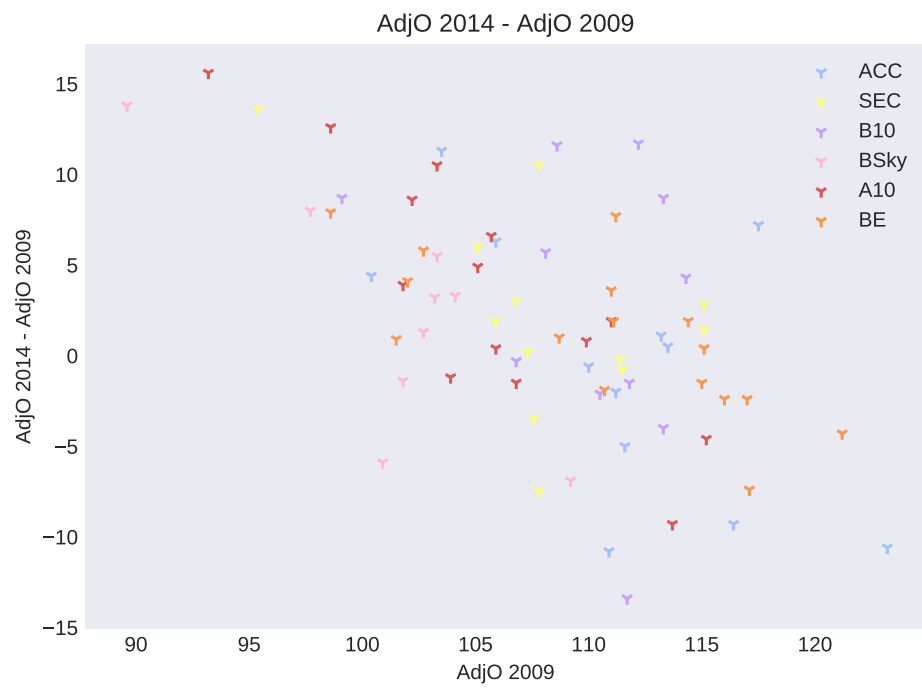
Figure 4: Exercise 3-2 - plot of the AdjO of all the teams as fcn of 2009 AdjO

```
A10                    2.90
BE                     0.95
Outsiders              1.50
Conference    AdjO diff-Mn
ACC                   -0.63
SEC                    2.28
B10                    2.67
BSky                   2.32
A10                    3.51
BE                     0.96
Outsiders              2.49
```