Data Analysis Fixing

May 1, 2016

1 Data Analysis Fixing

1.1 College Scorecard and University Ranking

The original analysis did not distinguish between public and private universities. In this analysis I will answer the following research questions in the scope of public universities only.

1. How does the ranking of the university correspond with admission rate?
2. Is there a correlation between cost of tuition and university ranking?
3. What is the correlation between average SAT score and admission rate?

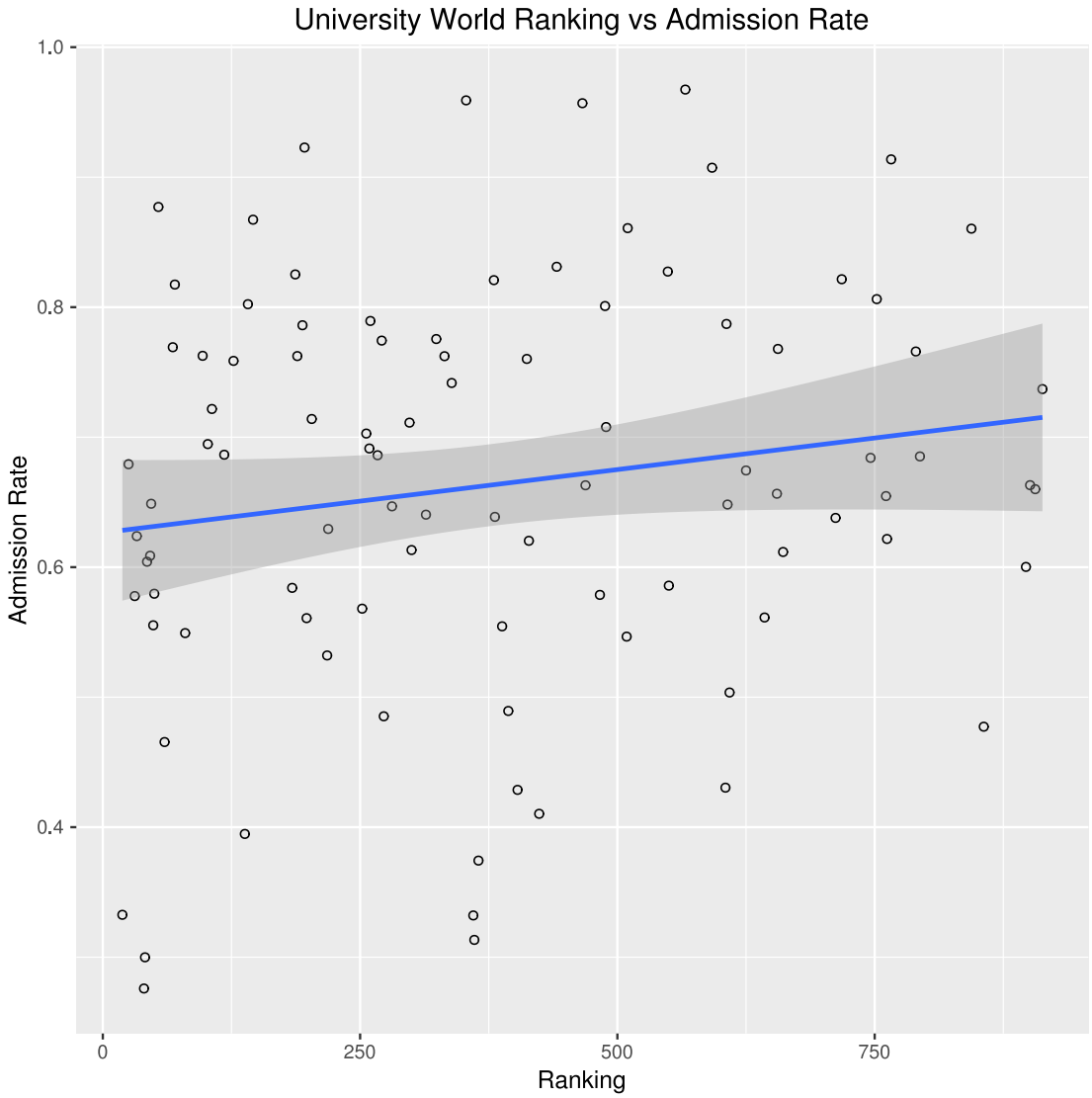
The first thing to do is to load the datasets and restrict them public universities only.

```
In [1]: TIMES_SCORE = read.csv("TIMES_SCORE.csv", encoding="UTF-8")
        CWUR_SCORE = read.csv("CWUR_SCORE.csv", encoding="UTF-8")

        CWUR_SCORE = subset(CWUR_SCORE, CONTROL == "Public")
        TIMES_SCORE = subset(TIMES_SCORE, CONTROL == "Public")
```

The graph below answers question one. It shows that the admission rate has a positive correlation with the ranking of the university. The closer the university is to being the top ranked school, number one, the lower the admission rate it. There are a large amount of data points that are not inside the confidence region but there is still a good correlation. This trend is what I expected, the better the school is ranked the harder it is to be admitted into the univeristy.

```
In [2]: require(ggplot2)

        ggplot(CWUR_SCORE, aes(x=world_rank, y = ADM_RATE_ALL)) +
          geom_point(shape = 1) + geom_smooth(method=lm) +
          ggtitle("University World Ranking vs Admission Rate") +
          labs(x = "Ranking") + ylab("Admission Rate")
```

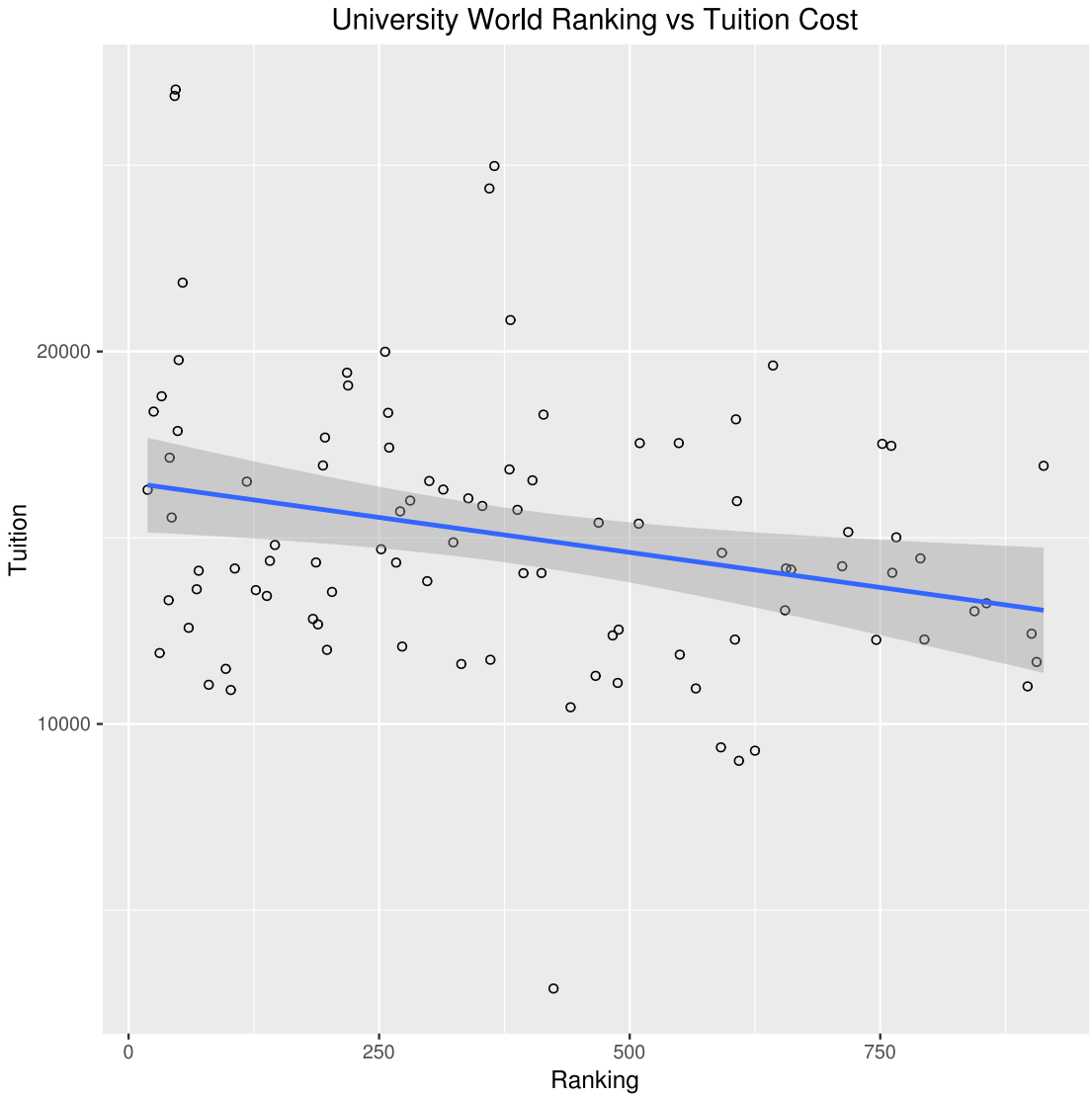
```
Loading required package: ggplot2
Warning message:
: Removed 7 rows containing non-finite values (stat_smooth).Warning message:
: Removed 7 rows containing missing values (geom_point).
```



The graph below answers question number two. The is a inverse correlation between the cost of tuition and the world rank of the university. The confidence region contains a significant amount of data points meaning the trend line is a good representation of the data. The trend shows that the lower the ranking of the school, the higher the cost of tuition. This is what I expected, the closer the school is to being ranked number one in the world the more expensive it would be to attend.

```
In [3]: ggplot(CWUR_SCORE, aes(x=world_rank, y = NPT4_PUB)) +
          geom_point(shape = 1) + geom_smooth(method=lm) +
          ggtitle("University World Ranking vs Tuition Cost") +
          labs(x = "Ranking") + ylab("Tuition")
```

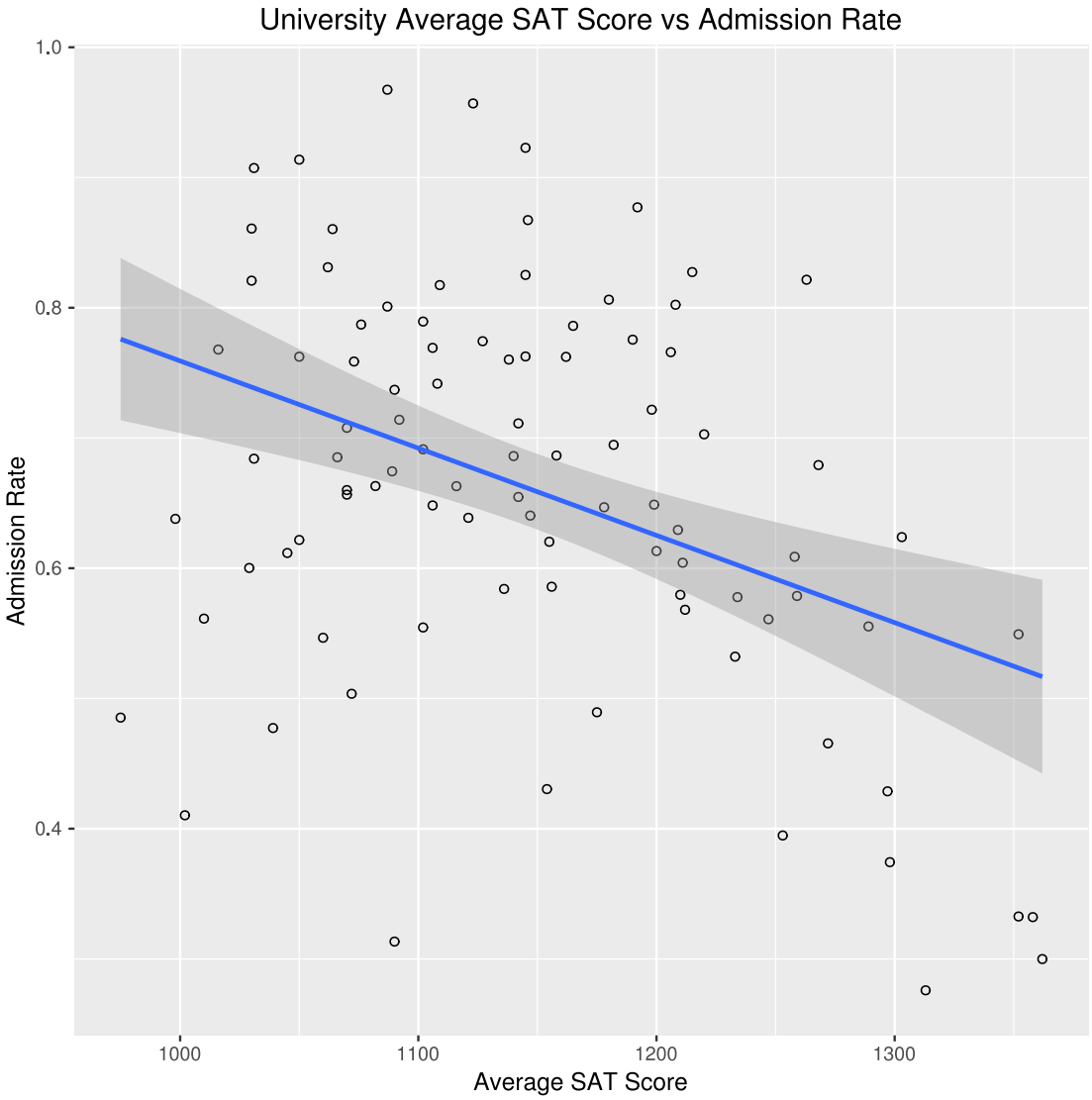
```
Warning message:
: Removed 6 rows containing non-finite values (stat_smooth).Warning message:
: Removed 6 rows containing missing values (geom_point).
```



The graph below answers question three. There is a strong inverse correlation between universitiy admission rates and average SAT scores. The lower the average SAT score is for the university the higher the admission rate. This is a logical outcome. I would expect it to be more challenging to be admitted into a school that has high performing students.

```
In [4]: ggplot(CWUR_SCORE, aes(x=SAT_AVG, y = ADM_RATE_ALL)) +
          geom_point(shape = 1) + geom_smooth(method=lm) +
          ggtitle("University Average SAT Score vs Admission Rate") +
          labs(x = "Average SAT Score") + ylab("Admission Rate")
```

```
Warning message:
: Removed 8 rows containing non-finite values (stat_smooth).Warning message:
: Removed 8 rows containing missing values (geom_point).
```


University Average SAT Score vs Admission Rate
Admission Rate
1.0
0.8
0.6
0.4
1000
1100
1200
1300
Average SAT Score