Predicting Atlantic 10 Coaches' Salaries

My intent was to look at the fourteen Atlantic 10 (A10) basketball teams and evaluate if there is a linear relationship between the coach's salary and the team's performance. Team performance can be measured by win percentage overall, in conference, and out of conference (this year, last year, over the past 'x' number of years). It can also be measured by the number of wins vs. the Ratings Percentage Index (RPI) Top 50 and/or Top 100 (this year, last year, over the past 'x' number of years). Finally, team performance could be measured by final regular season conference standings or conference tournament final standings. I hoped to be able to conclude who is being overpaid with respect to their salary and who is being underpaid.

I combed the internet for the salaries of the A10 coaches. In Figure 1 below, you will find their raw 2014-2015 salaries, sorted from highest to lowest. Immediately, you can see that VCU's Shaka Smart was the highest paid coach in the A10 by a large margin. His salary was over $600,000 more than St. Joe's Phil Martelli. But these raw salaries need to be normalized for comparison-sake.

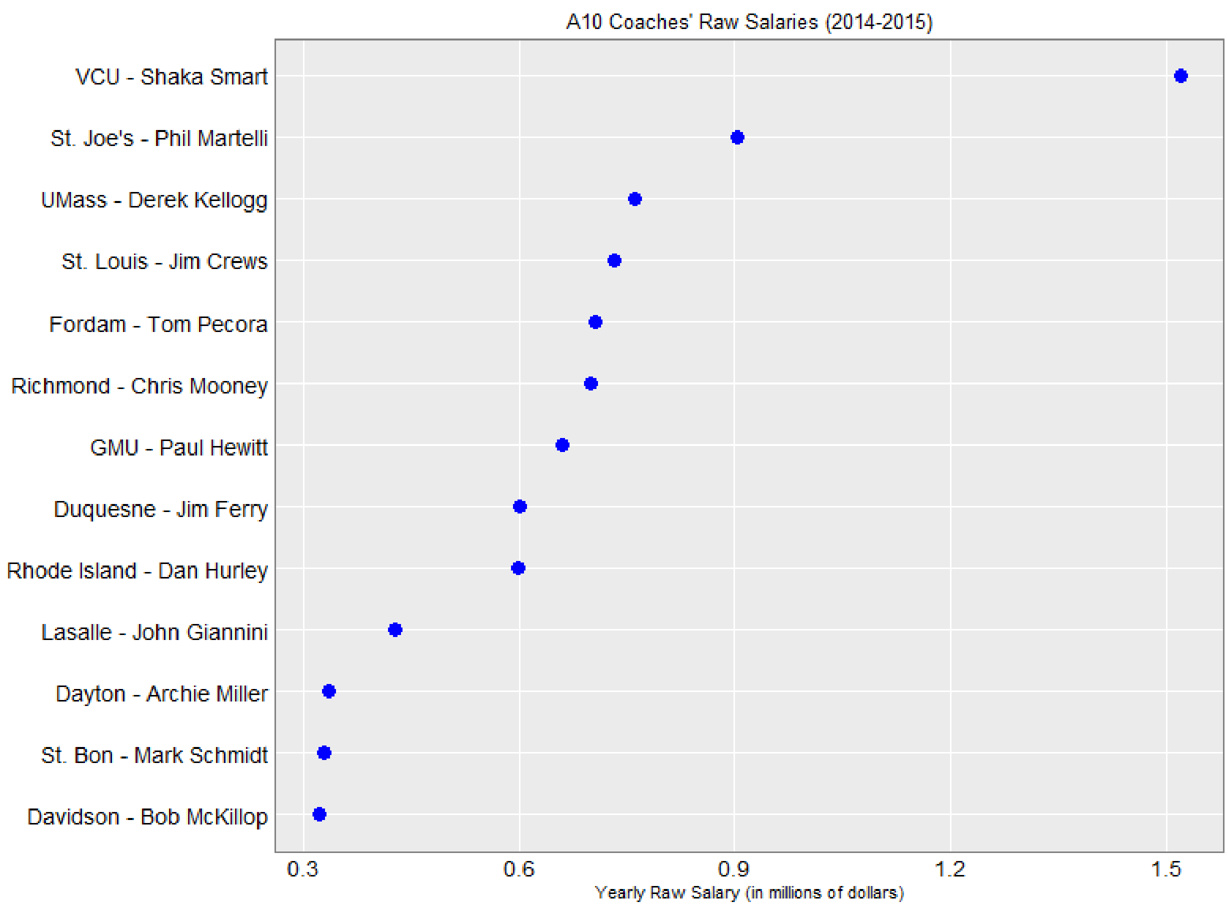

Figure 1.

I normalized all raw salaries to "DC dollars" (because a dollar in DC is not worth the same as a dollar in New York, for example). This realization to normalize the salaries to "DC dollars" came about because I was wondering why the Fordham head coach was making so much more than the Davidson head coach when it was obvious that the Fordham head coach was not performing nearly as well. I hypothesized that cost of living was higher in certain parts of the country, like NY and DC. After researching on CNN Money, I found this hypothesis to be true, so I also discounted all salaries to 2015 "DC dollars." Please see Figure 2 below. When I adjusted for cost of living, VCU's Shaka Smart's salary skyrocketed to over $2.5 million per year. All coaches' salaries, except Fordham's Tom Pecora and George Mason's Paul Hewitt, went up. The greater the distance of the adjusted salary (red dot) from the raw salary (blue triangle) means that cost of living adjustment is more significant. Quickly, you can see that St. Louis's Jim Crews and Richmond's Chris Mooney are next two highest paid coaches after Shaka Smart. The lowest paid coach is Fordham's Tom Pecora. This is easy to see because the data are plotted along the scale.

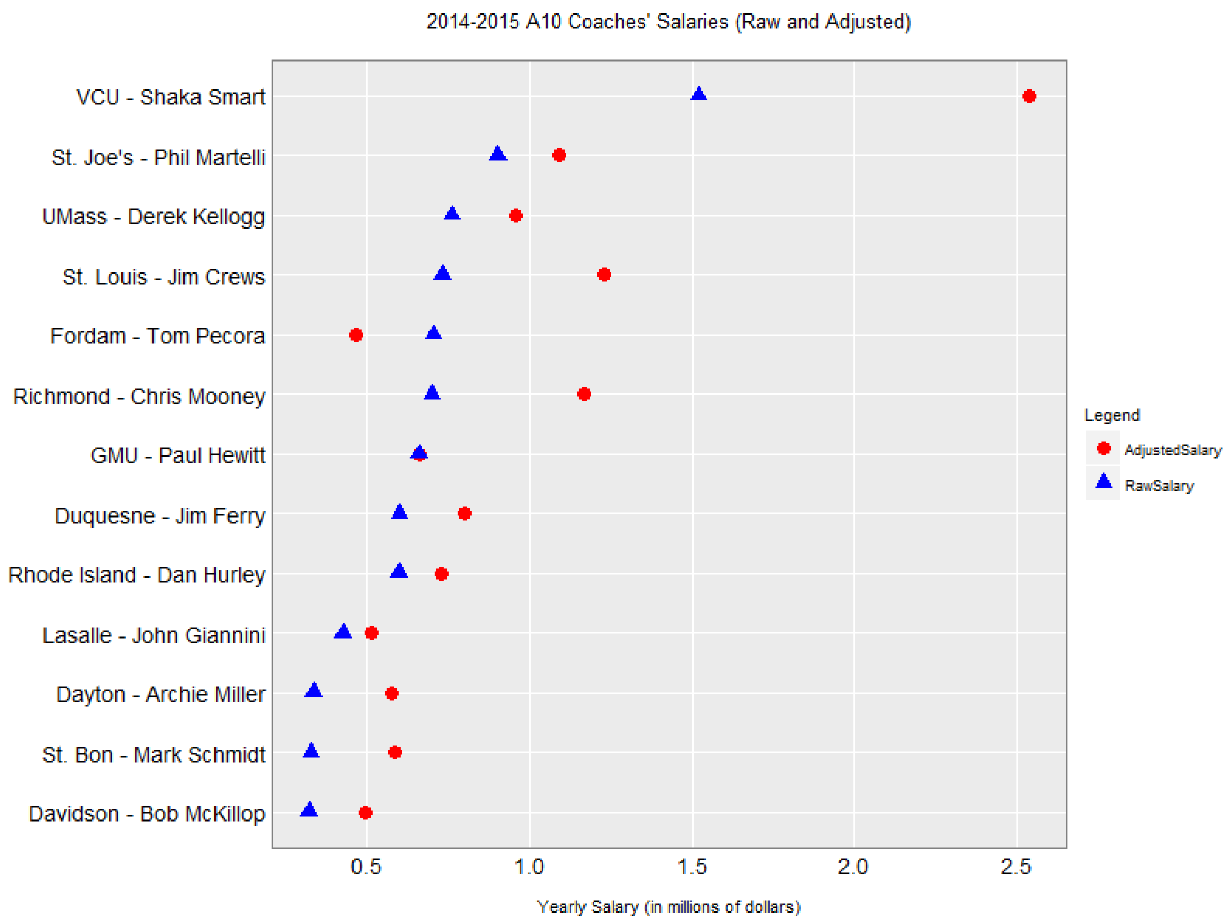

Figure 2.

To get a sense of what states make up the A10, I used micro-maps and plotted the average adjusted salaries of the head coaches from each state. Please see Figure 3 below. Virginia, Missouri, and Massachusetts all have adjusted salaries above the A10 average of $908,920 (which is represented by the vertical green dashed line). Figure 3 also illustrates how the A10 is composed of teams from states in the northeast, with the exception of St. Louis, which is found in Missouri.

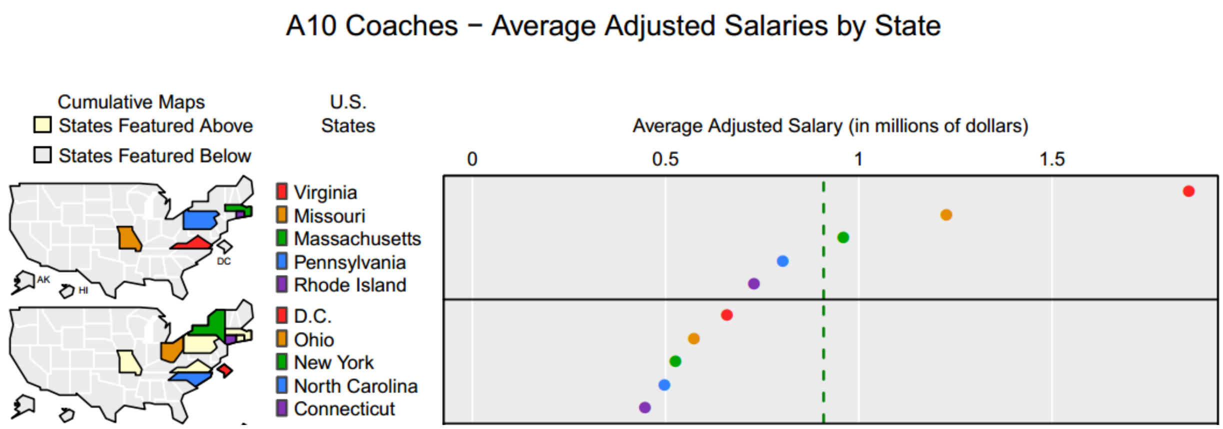

Figure 3.

Before I could conduct linear regressions, I had to collect performance data on each of the coaches. The performance data include both the team's performance under that coach over the last few seasons as well as the coach's historical performance as a head coach. I also hypothesized that there were some non-performance factors, such as age of the coach and tenure at their current school, that would influence the predicted salary of the coach. I used a scatterplot to plot all variables against adjusted salary. See Figure 4 on the next page. You can see that most of the variables are not Gaussian distributed and only of couple of the scatterplots show any kind of promise with regards to a linear relationship. I plotted Career NCAA Tournament Appearances and ESPN 3 & 4 Star Recruits against Adjusted Salaries. Please refer to Figures 5a & 5b and 6a & 6b on pages 5 and 6. In Figure 5a, you will see the linear regression for Career NCAA Tournament Appearances vs. Adjusted Salary, and in Figure 5b, you will find the R^2 value of .1063. The R^2 is skewed greatly by the Shaka Smart data point. In Figure 6a, you will see the linear regression for ESPN 3 & 4 Star Recruits vs. Adjusted Salary, and in Figure 6b, you will find the R^2 value of .3954. Both the dependent variable (ESPN 3 & 4 Star Recruits) and intercept are statistically significant at a significance level of 0.05. However, the R^2 is skewed greatly by the Shaka Smart data point.

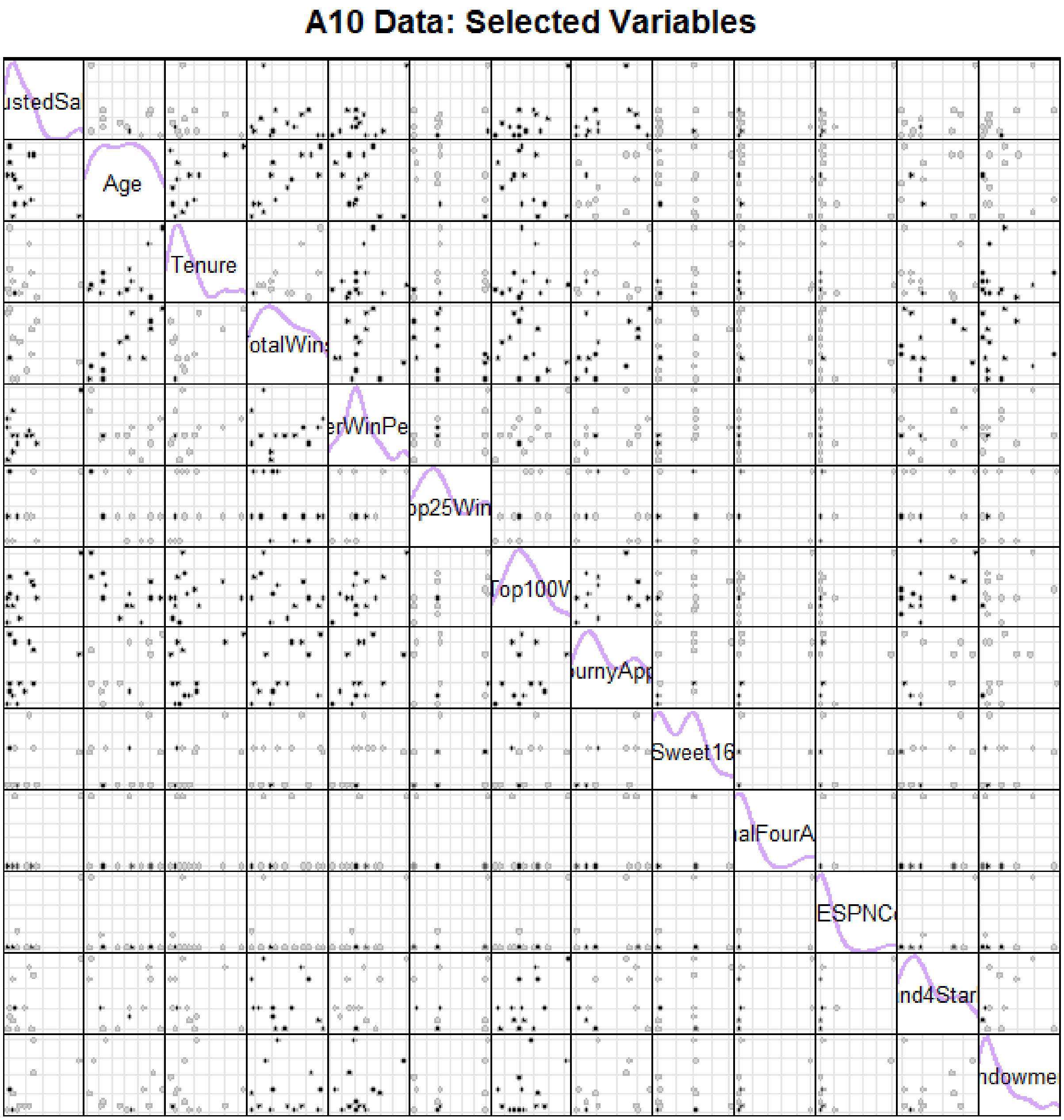
A10 Data: Selected Variables
ıstedSa
Age
Tenure
otalWin
erWinPe
op25Win
Top100V
urnyApp
Sweet16
ıalFourA
ESPNC
nd4Star
ndowme

Figure 4.

Figure 5a.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.69684    0.23991   2.905   0.0143 *
TourneyApps  0.07255    0.06343   1.144   0.2769
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.549 on 11 degrees of freedom
Multiple R-squared:  0.1063,    Adjusted R-squared:  0.02506
F-statistic: 1.308 on 1 and 11 DF,  p-value: 0.2769
```

Figure 5b.

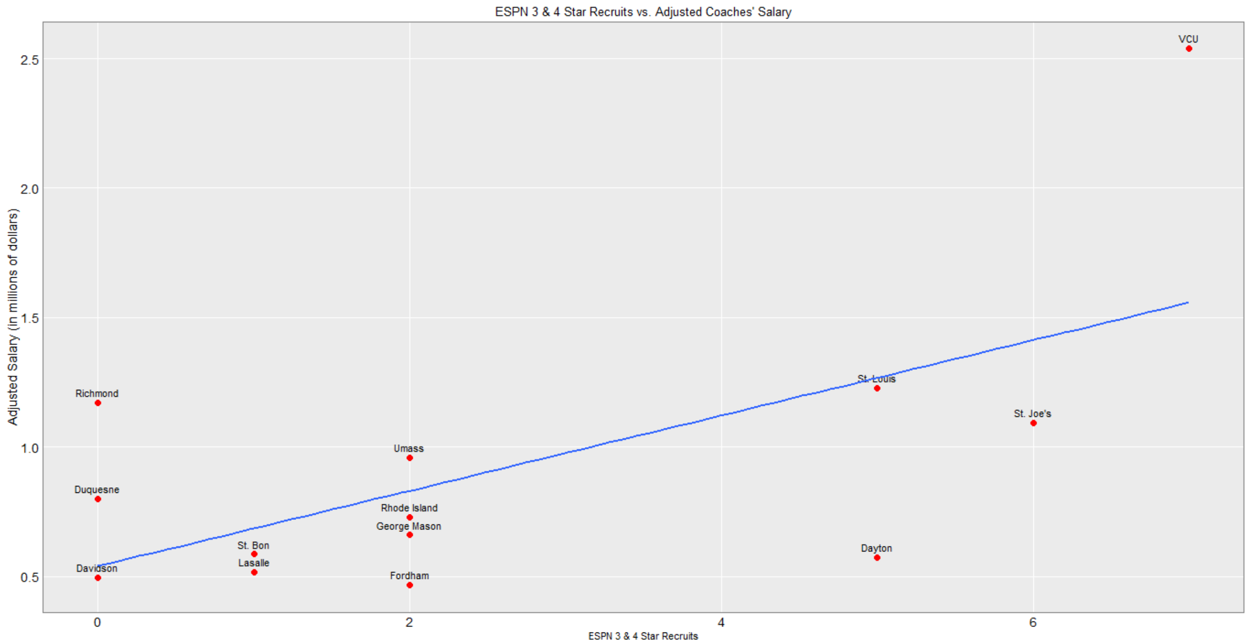

Figure 6a.

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            0.53943    0.18619   2.897   0.0145 *
ESPN3and4StarRecruits  0.14556    0.05427   2.682   0.0213 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4516 on 11 degrees of freedom
Multiple R-squared:  0.3954,    Adjusted R-squared:  0.3404
F-statistic: 7.193 on 1 and 11 DF,  p-value: 0.02133
```

Figure 6b.

The low correlation and flatness of the data also was evident in non-performance factors that were considered. For instance, tenure at the current school intuitively may lead one to think that the coach has had success (otherwise he could have been fired) and that the coach would have had chances to renegotiate their contract for a higher salary. However, the R^2 was extremely low (at 0.0253) for Tenure vs. Adjusted Salary and even more surprising is that the tenure is negatively correlated with adjusted salary among Atlantic 10 coaches. Age vs. Adjusted Salary produced a similar inversely correlated relationship, with a better (still low) R^2. Again, however, there is an outlier in Shaka Smart, a young coach who has only been with the current school for 6 years and makes a large salary. This pulls the regression line up slightly on the left side, leading to the appearance of an inversely correlated relationship.

Given that much of the data set was flat and provided no statically significant correlations, I went back to the drawing board to explore other factors that could influence a coach's salary in the A10. I looked into the financial factors, such as the school's endowment, the previous coach's salary, the salary the existing coach was making under a previous contract with the same school to see if there was a historical trend at the schools in particular. However, for most of these factors, they could not be completely attributed to the A10. Bob McKillop at Davidson was very successful in the 2014-2015 season in the A10, winning the regular season title. On the other hand, Paul Hewitt at George Mason was making one of the highest salaries in the A10 and had not had a winning record in his last two seasons. Why was it that McKillop was making one of the lowest salaries and Hewitt one of the highest? It turns out that Davidson was newly admitted to the A10, and was coming from a conference where the athletic budgets, let alone the men's basketball budget, was significantly less. Hewitt, on the other hand, was an NCAA Tournament Semi-Finalist at Georgia Tech, an ACC school renowned for top basketball programs, thus his previous salary prior to coming to George Mason might have influenced his George Mason salary. It would follow then that with less money for the program, less money would be available to pay the coaching staff.

This led us to look at the men's basketball budgets. Could the head coach's salary simply be dependent on the basketball budget? Please refer to Figure 7a and 7b on the next page. The data appears to show a relationship, however, R^2 was only 0.4219. That is because there are two outliers in Dayton and VCU. There definitely could be an argument made that Archie Miller was being underpaid. Interestingly, at the end of the 2014-2015 season, Coach Miller was awarded a new contract and a nice pay raise from $335,152 to $652,049 per year.

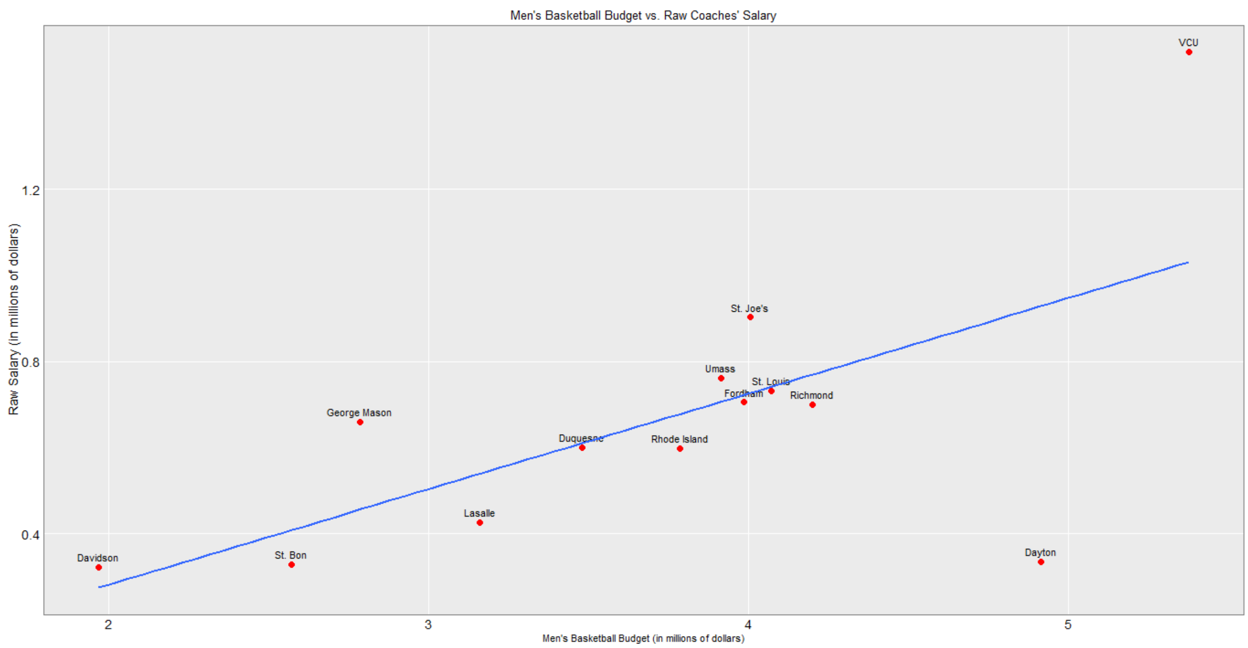

Figure 7a.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.16320    0.29906  -0.546   0.5962
MBB_Budget   0.22210    0.07838   2.834   0.0163 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2521 on 11 degrees of freedom
Multiple R-squared:  0.4219,    Adjusted R-squared:  0.3694
F-statistic: 8.029 on 1 and 11 DF,  p-value: 0.01627
```

Figure 7b.

I updated the data set with Archie Miller's new salary as well as Bob McKillop's renegotiated contract with Davidson (worth $381,285). I also removed the Shaka Smart data point from the data set because his salary is so much higher than all the rest of the coaches in the A10. As a result, I found that the men's basketball budgets were highly correlated with the updated salaries. Please see Figure 8a and 8b on the next page. R^2 improved to 0.5097 (from 0.4219) and the dependent variable (MBB Budget) was highly significant (p-value of less than 0.01).

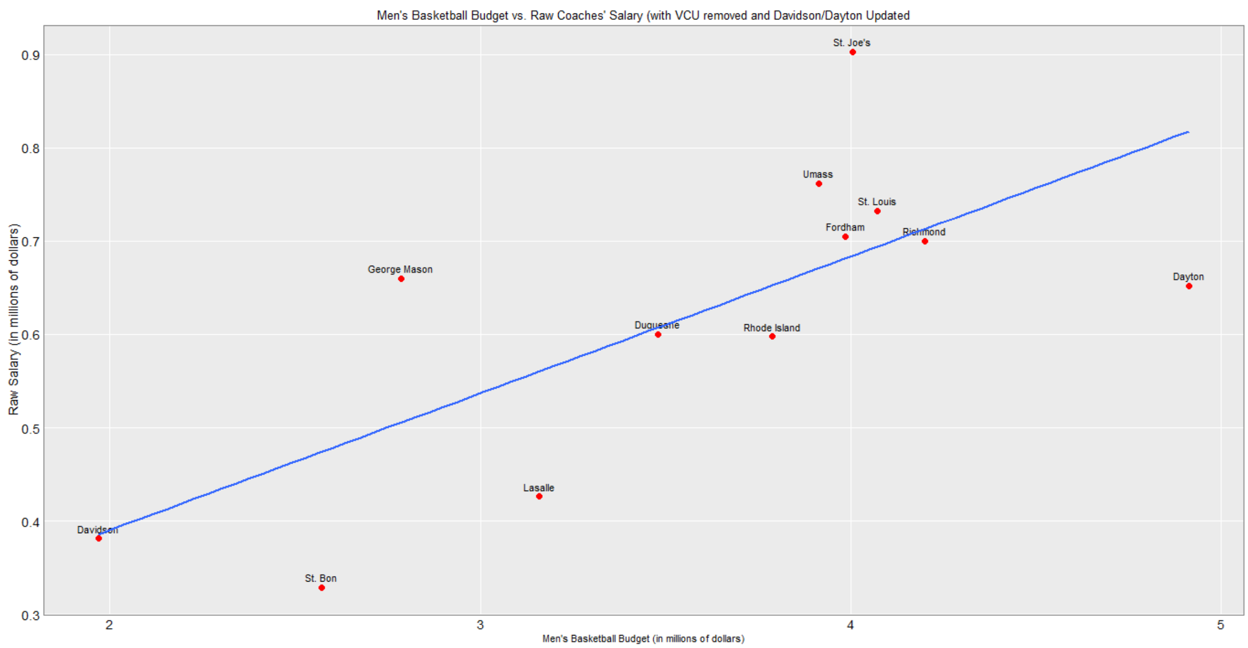

Figure 8a.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09711    0.16619   0.584  0.57192
MBB_Budget   0.14659    0.04547   3.224  0.00911 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1231 on 10 degrees of freedom
Multiple R-squared:  0.5097,	Adjusted R-squared:  0.4607
F-statistic: 10.39 on 1 and 10 DF,  p-value: 0.009111
```

Figure 8b.

After finding a link between men's basketball budget and the coach's salary in the A10, I turned my efforts toward trying to see if this link extended beyond the A10. A10 coach salaries were, on average, about 15-20% of the men's basketball budget. I wanted to see if that same percentage held for the major basketball conferences (Big 12, Big 10, Pac 12, ACC, SEC, Big East). So, I analyzed 33 teams from major basketball conferences that made the 2015 NCAA tournament. In Figure 9a, the data visually seems to suggest that the higher the budget, the higher the salary of the coach. Figure 9b gives credence to that conjecture as the R^2 is 0.7227, and the dependent variable (MBB Budget) is highly significant with a p-value of less than 0.001. Therefore, the linear regression results would suggest that I could estimate coach salary in the major basketball conferences with the following equation:

Estimated Salary = 0.399*MBB Budget – 0.543

Figure 9c shows the results of a 95% confidence interval on the dependent variable (MBB Budget) is (0.308, 0.490). This means that there is statistical evidence (at an alpha = 0.05) that the major basketball conferences do pay their coaches a higher percent of their budget. Even Shaka Smart did not make over 30% of VCU's basketball budget. (He only made 28%, and that was top in the A10.)

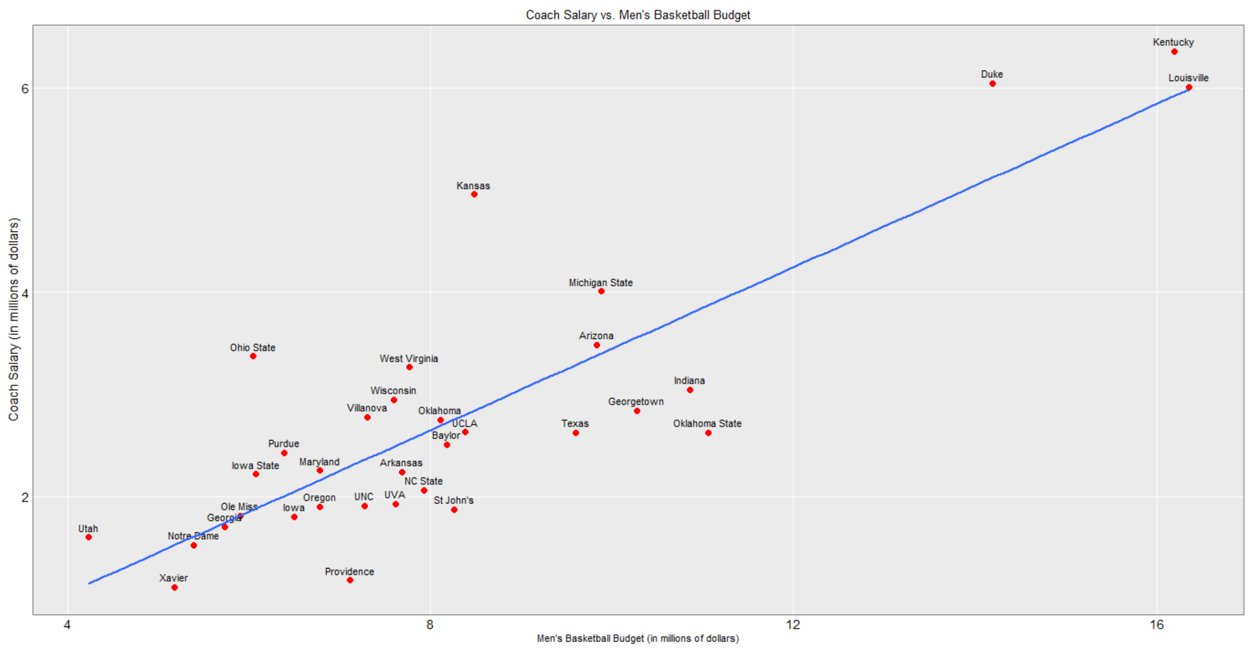

Figure 9a.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.54264    0.39024  -1.391    0.174
MBBBudget    0.39900    0.04439   8.988 3.84e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7165 on 31 degrees of freedom
Multiple R-squared:  0.7227,	Adjusted R-squared:  0.7138
F-statistic: 80.79 on 1 and 31 DF,  p-value: 3.837e-10
```

Figure 9b.

```
                 2.5 %    97.5 %
(Intercept) -1.338537 0.2532629
MBBBudget    0.308468 0.4895378
> |
```

Figure 9c.

In order to determine if I can actually use a linear regression to model the data, I need to look at a few plots. Figure 10 below shows two important plots that test the linear regression assumptions. The “Residuals vs. Fitted” plot helps me to check for equal variance and the “Normal Q-Q” plot assists me in checking that underlying data is normally distributed. It does not appear that the “Residuals vs. Fitted” plot shows equal variance. And the “Normal Q-Q” plot shows that the data may have a thick right tail and may not be normally distributed.

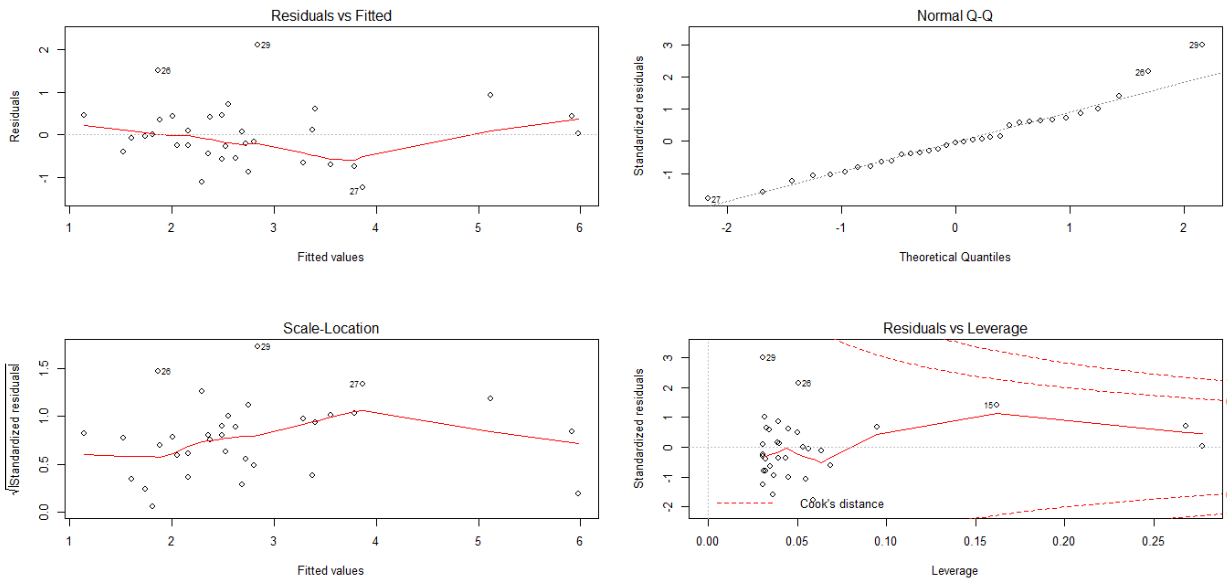

Figure 10.

Figures 11 and 12 (on the next page) show two EDA plots. Figure 11 is the EDA plot for the men’s basketball budget data set (n=33). From the density plot and box plot, it is easy to see the thick right tail. The median budget is $7.7 million and is displayed on the box plot. Figure 12 is the EDA plot for coaches’ salaries (n=33). Again, it is clear to see the thick right tail from the density and box plots. The median salary is $2.5 million.

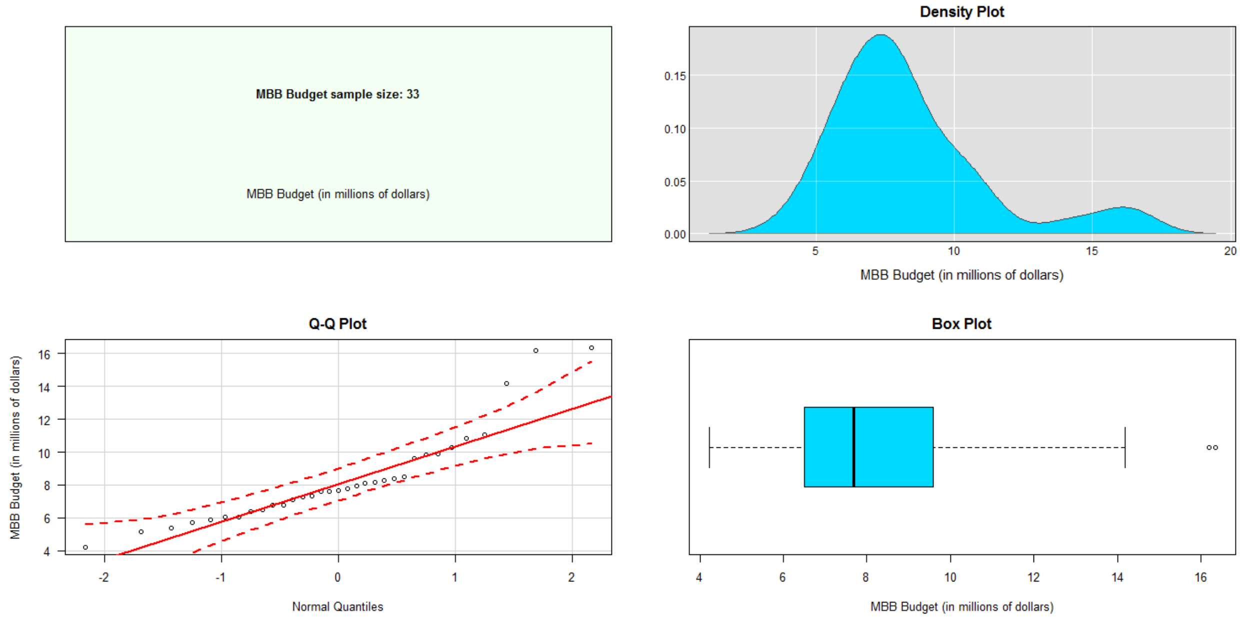

Figure 11.

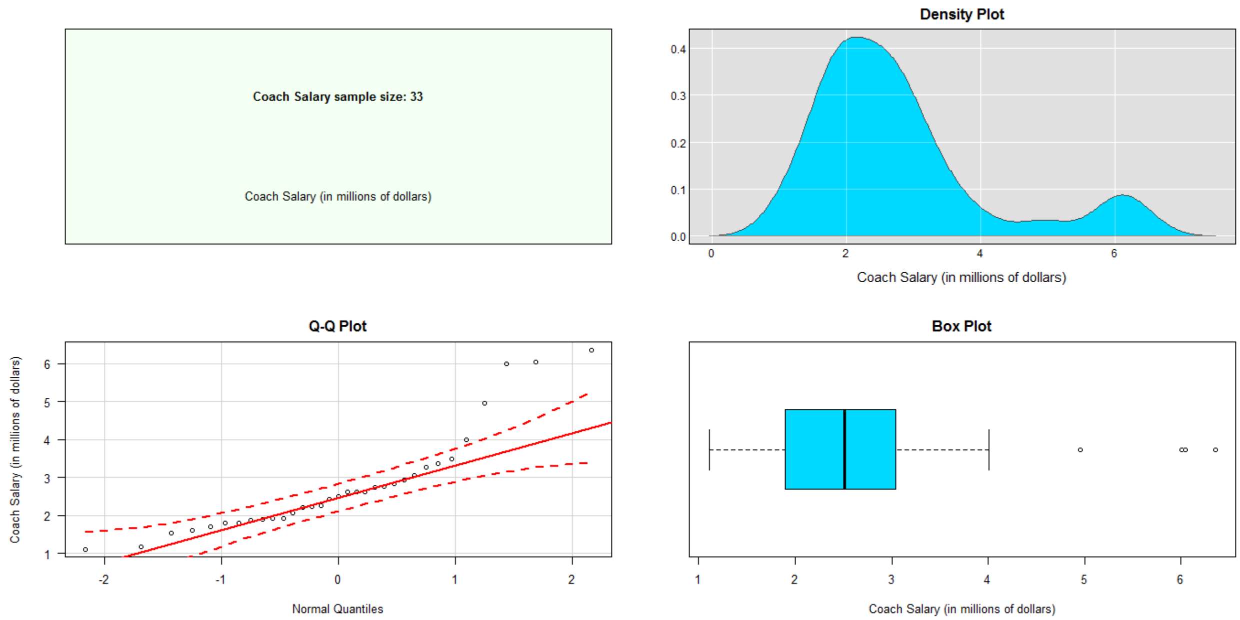

Figure 12.

My final two charts show how much of each of the schools pay their coach as a percentage of their budget. In Figure 13 below, I sorted the data from highest to lowest and plotted the points along a scale. It is easy to see that the coaches from Kansas (Big 12) and Ohio State (Big 10) receive well over 50% of the men's basketball budget while Providence's coach (Big East) is the only coach to receive less than 20%.

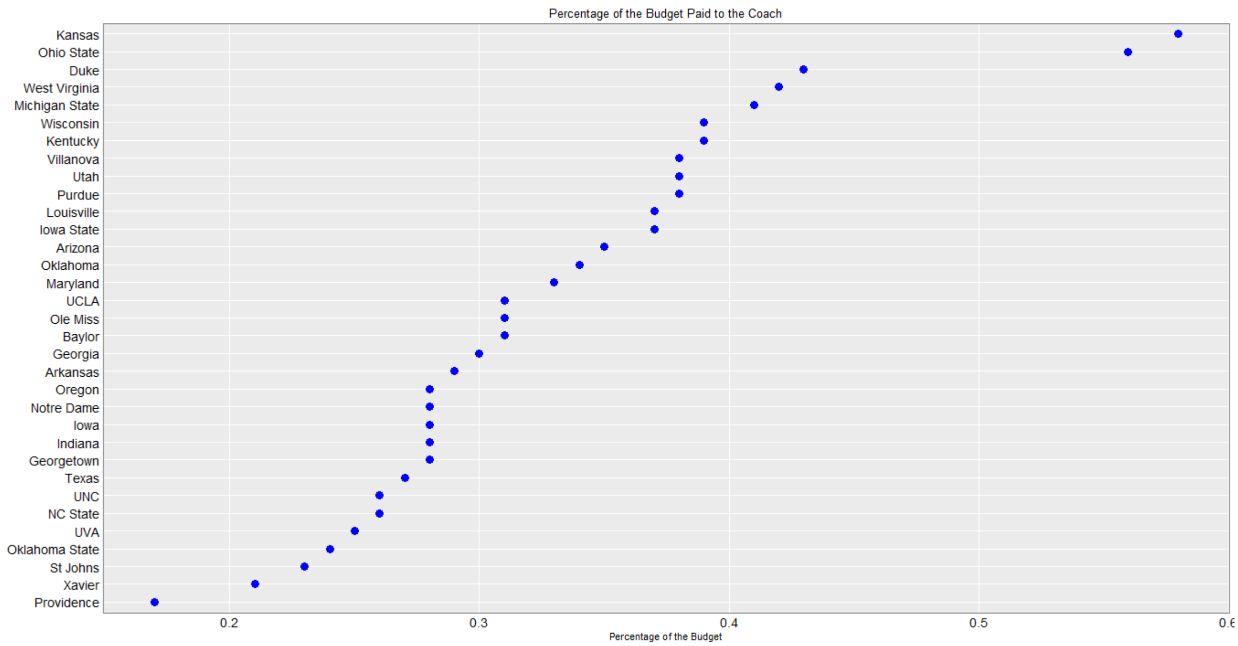

Figure 13.

Figure 14 (on the next page) sorts the data by conference from highest to lowest. By color coding the conferences and adding a line to join the conference data points, you can quickly see that the Big East has the three lowest paid coaches (as measured by the percentage of the budget paid to the coach). The Big 12 has the greatest variance while the SEC and PAC 12 have the least. The schools from each conference that pay the largest percentage of the budget are: Kansas (Big 12), Ohio State (Big 10), Duke (ACC), Kentucky (SEC), Utah (PAC 12), and Villanova (Big East).

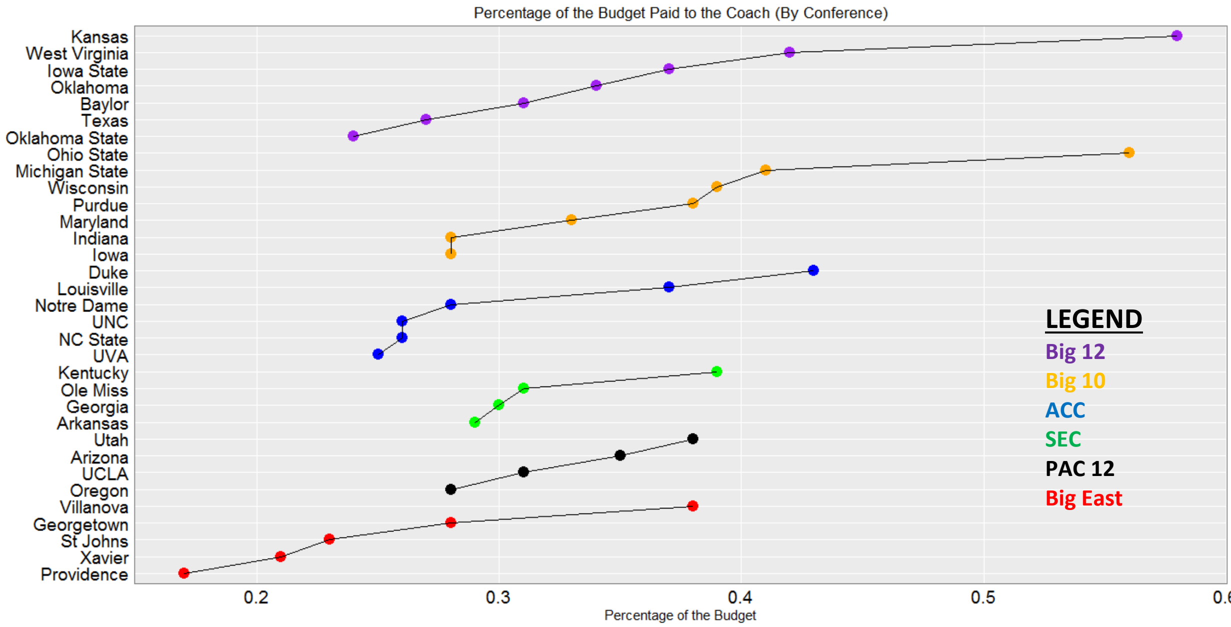

Figure 14.

You might recall that my intent was to look at the fourteen Atlantic 10 basketball teams and evaluate if there is a linear relationship between the coach's salary and the team's performance. Unfortunately, the data set was sparse, consisted of a few outliers, and did not yield expected results. However, after more research, I found a variable that appeared to predict the salary of a coach. That variable was the men's basketball budget of each school. With this apparent relationship, I was able to develop a linear regression model that showed promise. Then, I extended that model to other NCAA schools and found that the linear relationship remained. Unfortunately, when checking the linear regression assumptions, I found that the equal variance assumption and normality assumption appeared to be violated. Therefore, with more rich data, one might be able to re-analyze this relationship between men's basketball budgets and the coaches' salaries.