Pierre Casco
Lab 03 – Coaches
IST718

Introduction

Division 1 football, also known as the FBS, is a gigantic money maker for the NCAA, sponsors, gamblers, and the schools who put their product on the field every Saturday during the fall. It can be inferred that the more you win, the more money gets generated for everyone involved (except the players, but that's another conversation).A large part of the on-field product involves coaching the players, some naturally talented and some not as talented but hard working and coachable. Because of this, it's important to find a great coach to make the team as competitive as possible.

Research Question

How can we recommend the best salary (total compensation, minus bonus) for our next head football coach?

Data

Data was compiled from different sources on the internet. The primary dataset included salary information for 129 division 1 coaches spanning ten conferences and one group of independents. After combining the primary dataset with 3 other datasets (stadium information, graduation rates, and records from last year, the dataset grew to 21 different columns. The columns are explained below, (figure 1). One new column was added, point differential, to see how teams did overall with their scoring.

School	School name
Conference	Conference school belongs to
Coach	Coach name
SchoolPay	School pay
TotalPay	Total pay including bonus
Bonus	Bonus amount
BonusPaid	Bonus paid
Buyout	Contract buyout if coach gets fired or leaves
W	Wins
L	Losses
PF	Points for
PA	Points against
FED_RATE	Federal graduation rate
GSR	Graduation success rate
Stadium	Stadium Name
City	City stadium is located in
state	State stadium is located in
Capacity	Stadium capacity
Latitude	Latitude of stadium

longitude	Longitude of stadium
pointDiff	Point differential

Figure 1 – Column explanation

The head of the dataset is pictured below (figure 2).

School	Conference	Coach	SchoolPay	TotalPay	Bonus	BonusPaid	Buyout	W	L	PF	PA	FED_RATE	GSR	Stadium Name	stadium	city	state	conference	capacity	built	expanded	div	latitude	longitude
Air Force	Mt. West	Troy Calhoun	885000	885000	247000	NaN	NaN	5	7	363	309	NaN	77	Falcon	Stadium	Colorado Springs	CO	Mountain West	46692	1962	NaN	fbs	38.996907	-104.843688
Akron	MAC	Terry Bowden	411000	412500	225000	50000	688500	4	8	227	326	61	72	Summa Field at InfoCision	Stadium	Akron	OH	MAC	30000	2009	NaN	fbs	41.07257	-81.508384
Alabama	SEC	Nick Saban	8307000	8307000	1100000	500000	33600000	14	1	684	271	63	84	Bryant–Denny	Stadium	Tuscaloosa	AL	SEC	101821	1929	2010	fbs	33.20749	-87.550392
Alabama at Birmingham	C-USA	Bill Clark	900000	900000	950000	165471	3847500	11	3	418	238	58	71		NaN	NaN	NaN	NaN	71594	NaN	NaN	NaN	33.5115	-86.8425
Appalachian State	Sun Belt	Scott Satterfield	712500	712500	295000	145000	2160417	11	2	485	201	67	71	Kidd Brewer	Stadium	Boone	NC	Sun Belt	24050	1962	2009	fbs	36.211515	-81.685506

Figure 2 – Head of dataset

During compiling of data, I had to drop four schools from the dataset due to missing salary information. These four schools were Baylor, BYU, Rice, and SMU.

Observations

Initial data analysis shows the top 5 (figure 3) paid coaches in the dataset as well as the bottom 5 (figure 4) paid coaches as seen below. The top 5 paid coaches reside in either the SEC or the Big Ten. This makes sense as these are part of the Power 5 conferences which generate the most revenue and produce national champions on a yearly basis. The Power 5 include the SEC, ACC, Big 10, Big 12, and Pac 12. The bottom 5 paid coaches belong to conferences that are considered Mid-Major schools. These include the MAC, Sun Belt, and AAC amongst the list.

```
The top 5 paid coaches are:
              Coach   TotalPay Conference
2       Nick Saban  8307000.0        SEC
82     Urban Meyer  7600000.0    Big Ten
62    Jim Harbaugh  7504000.0    Big Ten
105   Jimbo Fisher  7500000.0        SEC
10     Gus Malzahn  6705656.0        SEC
```

Figure 3 – Top 5 paid coaches

```
The bottom 5 paid coaches are:
             Coach  TotalPay Conference
11       Mike Neu  435689.0        MAC
74    Doug Martin  419640.0       Ind.
1   Terry Bowden  412500.0        MAC
24     Joe Moglia  400000.0   Sun Belt
53    Matt Viator  390000.0   Sun Belt
```

Figure 4 – Bottom 5 paid coaches

The next thing I looked at was a boxplot of the school pay by conference (figure 5). The highest average pay resides in the Big Ten with the average totaling just over ~$4 million. The largest

range was in the SEC. There were some outliers among the Power 5 conferences. One in the ACC and 2 in the Big Ten.

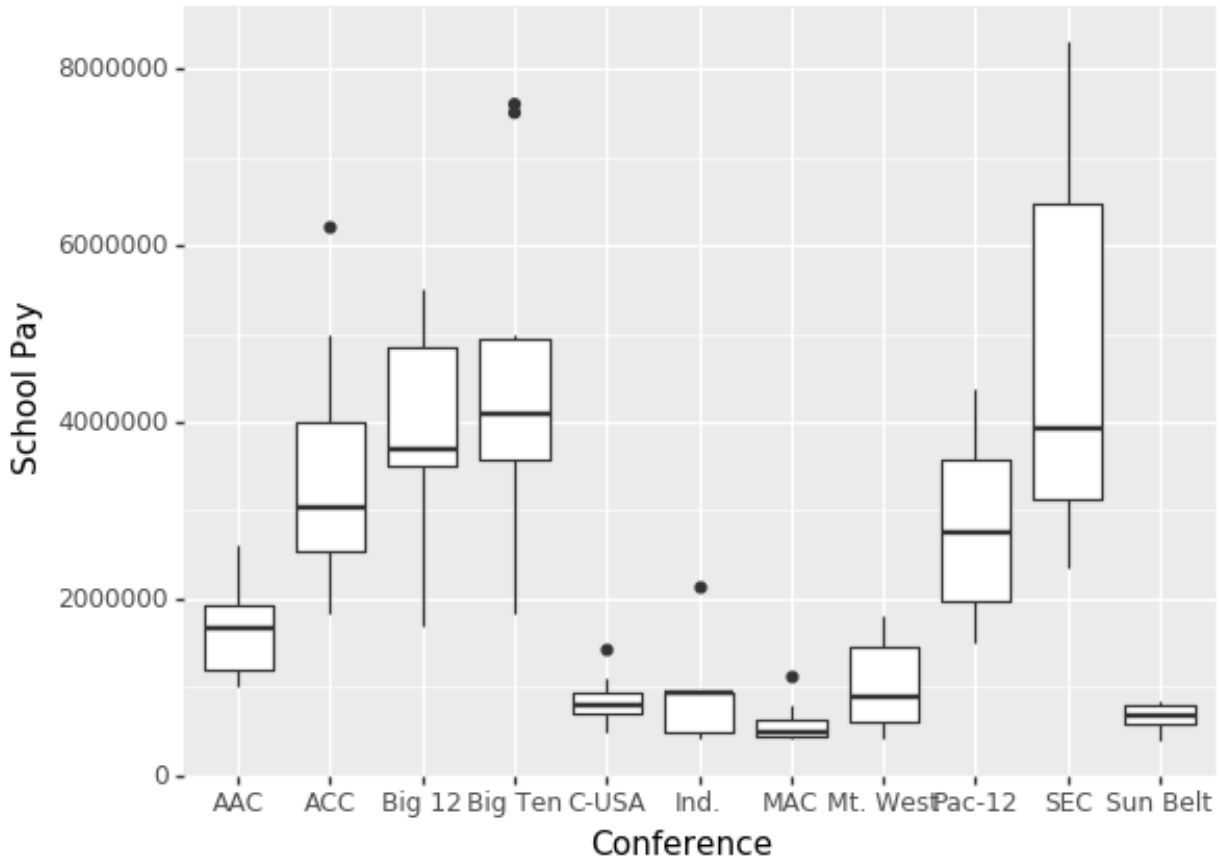


Figure 5 – Pay by conference

I wanted to explore the relationship between school pay and total points scored. The modern offenses of college football are built to score points in bunches. So, I was curious to see if higher scoring offenses amounted to higher paying coaches. Last season's win data was also added to the plot to see if there were any possible relations. One can infer that higher points equal more wins. Although this thought process can be debunked with a high scoring game (56-49) on both sides.

The plot (figure 6) showed a positive linear relationship between total points scored and school pay. It also showed the higher paying coaches also had a win total between 10 and 15 games last season. On the lower end, the coaches with the lower school pay, scored less points and won less than 5 games last season.

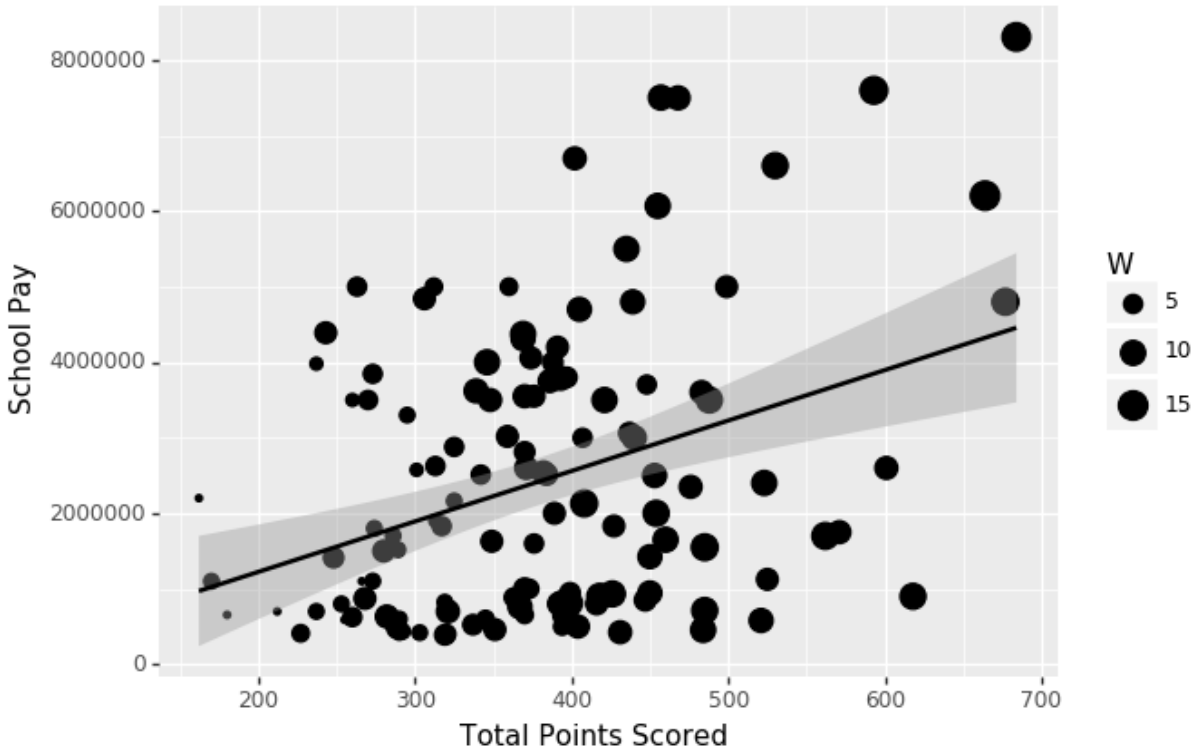


Figure 6 – Scatterplot of points scored, and school pay with dot size representing wins.

Bar graphs were utilized to study conference data further. Totals for points scored, wins, capacity, and graduation rate were graphed to study any trends in performance on the field, as well as stadium size and performance in the classroom. The charts for wins, points scored, and stadium capacity were as expected, with Power 5 conferences leading the pack. The graduation rate chart was even. This is encouraging considering that a vast majority of these kids won't make it to the next level, the NFL. So, it was nice to see they are graduating at a solid rate for their post college career. This also shows that coaches are following through on their promises of getting players an education.

The other graphs also explain why Power 5 coaches make more money than their Mid-Major counterparts. The explosive offenses lead to wins which bring in fans to fill their massive stadiums.

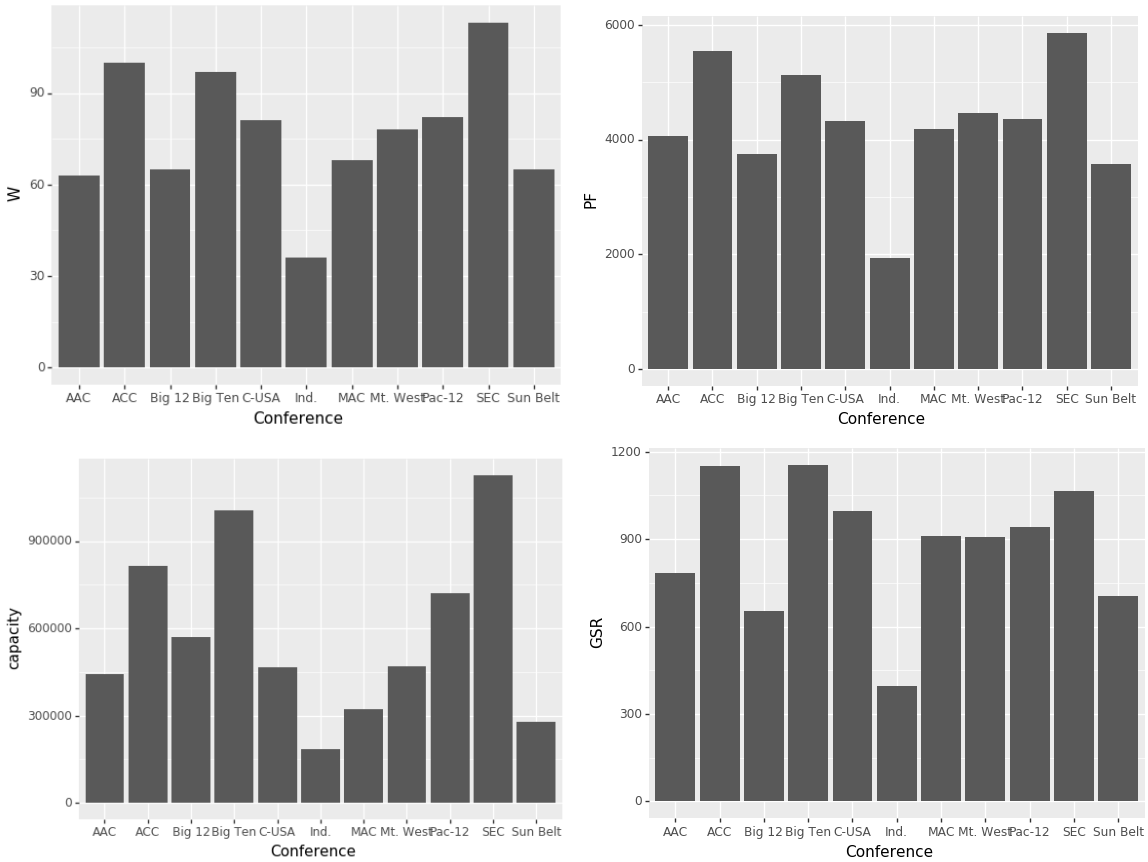
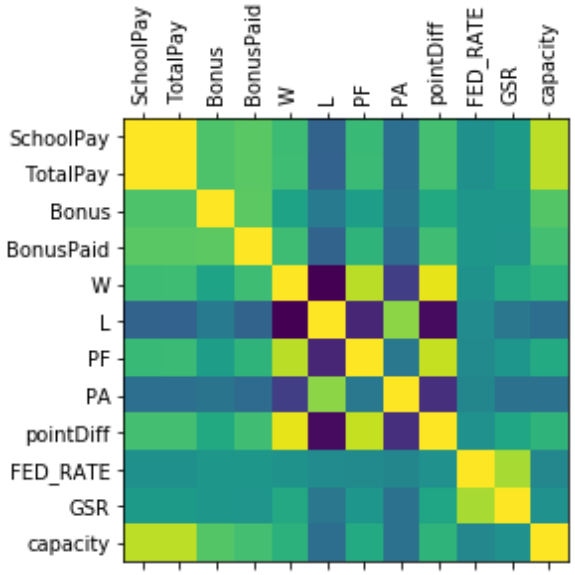


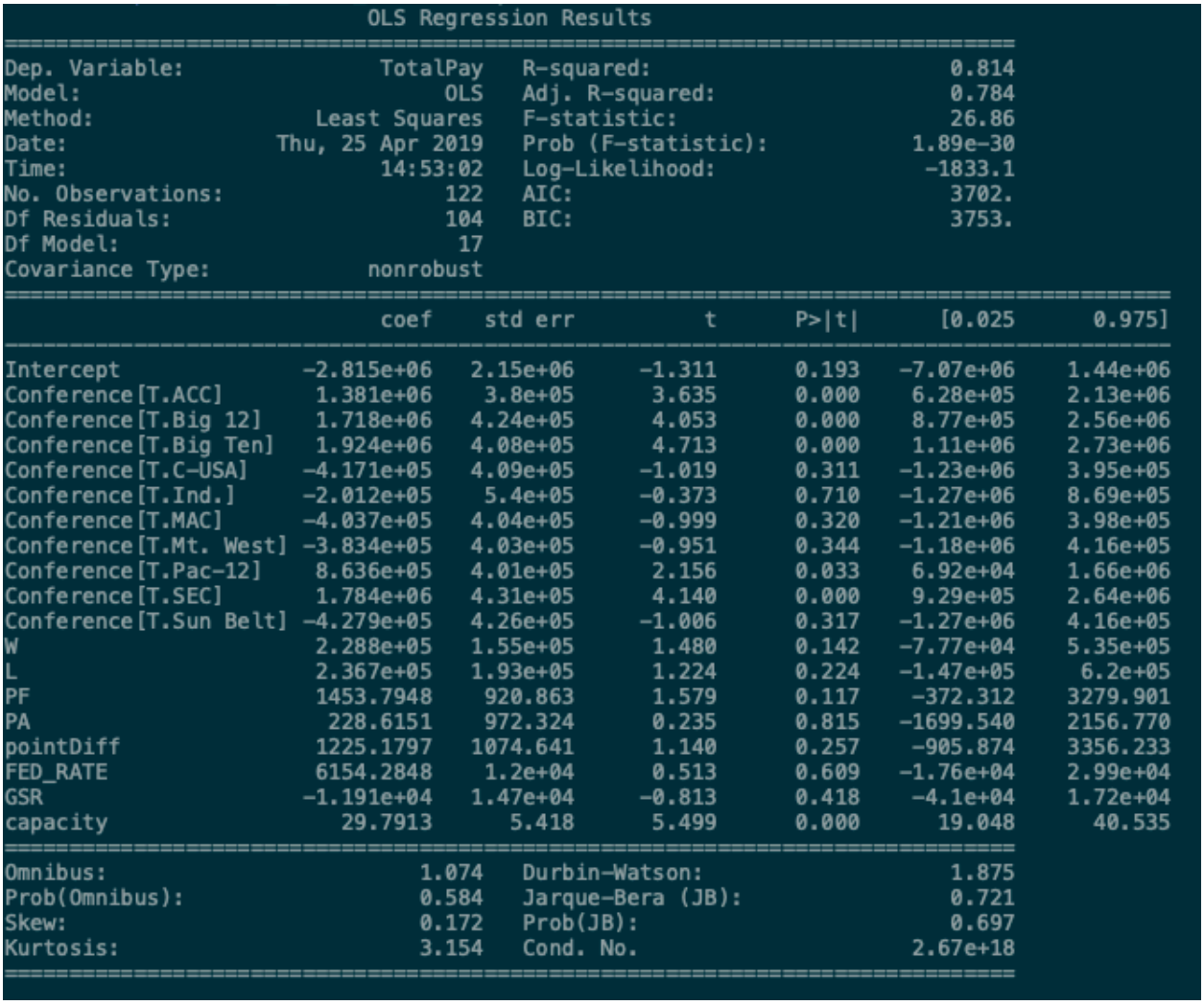
Figure 7 – Bar graphs showing wins, points scored, stadium capacity, and graduation rates by conference

A correlation matrix was created to see what correlations could be found between non categorical variables. Some noteble results were a correlation between points scored and wins, stadium capacity and school pay, points given up and losses.



Analysis

A regression model (figure 8) was created using pay as the output and conference, wins, losses, points scored, points against, federal graduation rate, graduation success rate, and stadium capacity. Using an OLS model, the adjusted R-squared value was 78%, meaning it is a decent model. The variables that had the highest significance were Power 5 conferences and stadium capacity. Of all the variable listed, the single biggest impact on salary size is the conference they're coaching in, with the Big Ten having the largest impact of the conferences.



OLS Regression Results

Dep. Variable:	TotalPay	R-squared:	0.814
Model:	OLS	Adj. R-squared:	0.784
Method:	Least Squares	F-statistic:	26.86
Date:	Thu, 25 Apr 2019	Prob (F-statistic):	1.89e-30
Time:	14:53:02	Log-Likelihood:	-1833.1
No. Observations:	122	AIC:	3702.
Df Residuals:	104	BIC:	3753.
Df Model:	17		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2.815e+06	2.15e+06	-1.311	0.193	-7.07e+06	1.44e+06
Conference[T.ACC]	1.381e+06	3.8e+05	3.635	0.000	6.28e+05	2.13e+06
Conference[T.Big 12]	1.718e+06	4.24e+05	4.053	0.000	8.77e+05	2.56e+06
Conference[T.Big Ten]	1.924e+06	4.08e+05	4.713	0.000	1.11e+06	2.73e+06
Conference[T.C-USA]	-4.171e+05	4.09e+05	-1.019	0.311	-1.23e+06	3.95e+05
Conference[T.Ind.]	-2.012e+05	5.4e+05	-0.373	0.710	-1.27e+06	8.69e+05
Conference[T.MAC]	-4.037e+05	4.04e+05	-0.999	0.320	-1.21e+06	3.98e+05
Conference[T.Mt. West]	-3.834e+05	4.03e+05	-0.951	0.344	-1.18e+06	4.16e+05
Conference[T.Pac-12]	8.636e+05	4.01e+05	2.156	0.033	6.92e+04	1.66e+06
Conference[T.SEC]	1.784e+06	4.31e+05	4.140	0.000	9.29e+05	2.64e+06
Conference[T.Sun Belt]	-4.279e+05	4.26e+05	-1.006	0.317	-1.27e+06	4.16e+05
W	2.288e+05	1.55e+05	1.480	0.142	-7.77e+04	5.35e+05
L	2.367e+05	1.93e+05	1.224	0.224	-1.47e+05	6.2e+05
PF	1453.7948	920.863	1.579	0.117	-372.312	3279.901
PA	228.6151	972.324	0.235	0.815	-1699.540	2156.770
pointDiff	1225.1797	1074.641	1.140	0.257	-905.874	3356.233
FED_RATE	6154.2848	1.2e+04	0.513	0.609	-1.76e+04	2.99e+04
GSR	-1.191e+04	1.47e+04	-0.813	0.418	-4.1e+04	1.72e+04
capacity	29.7913	5.418	5.499	0.000	19.048	40.535

Omnibus:	1.074	Durbin-Watson:	1.875
Prob(Omnibus):	0.584	Jarque-Bera (JB):	0.721
Skew:	0.172	Prob(JB):	0.697
Kurtosis:	3.154	Cond. No.	2.67e+18

Figure 8 – OLS regression model

With this current model, it is possible to recommend pay for different coaches around division 1, college football. Dino Babers, head coach at Syracuse, is currently making $2,401,210. When looking at his team's performance for last season, the model is recommending that he make $3,501,090. At this point, according to the model, we are under paying our coach by $1,099,880.

Syracuse, as well as a number of other schools, used to belong to the Big East. That conference no longer has any college football schools; it only supports college basketball. To predict what his salary would be if he were still in the Big East, I decided to use the Independent school rate. The predicting equation for Independents is:

$$pay = -2.815e^6 - 2.012e^5 + 2.88e^5(w) + 2.367e^5(l) + 2678.9745(pf) - 996.5647(pa) + 1225.1797(pointDiff) + 6154.2848(fed_{rate}) - 1.191e^4(gsr) + 29.7913(capacity)$$

When inserting variables into the equation, the recommended pay for the Syracuse coach if he were in the Big East is, $2,510,881, which is close to what he is getting paid now. This can be explained by the strong performance his team had in 2018, winning 10 games and losing only 3. He also had a point differential of +172.

If there was to be a conference reshuffle, and Syracuse ended up in the Big Ten, his pay will likely change. The predicting equation for the Big Ten is:

$$pay = -2.815e^6 + 1.924e^6 + 2.88e^5(w) + 2.367e^5(l) + 2678.9745(pf) - 996.5647(pa) + 1225.1797(pointDiff) + 6154.2848(fed_{rate}) - 1.191e^4(gsr) + 29.7913(capacity$$

When inserting last season's info into the equation, the recommended pay for the Syracuse coach if he were in the Big Ten is, $4,646,080. This makes sense as the Big Ten has some of the highest salaries of all the schools.

Graduation rate is an important factor since, as mentioned above, not all of these players will make it to the NFL. Of the 2 rates, federal rate and graduation success rate, the federal rate has a larger effect than the GSR. However, both are not as significant as conference, wins, and losses.

Summary

- The recommended salary for the Syracuse football coach is $3,501,090.
- His salary if he were still in the Big East would be $2,510,881.
- His salary if he were in the Big Ten would be $4,646,080.
- I dropped 4 schools from my data because they were missing salary information (Baylor, BYU, SMU, Rice)
- Graduation rate has a small effect on projected salary.
- Our model was decent at an adjusted R-squared of 78%
- The single biggest impact on salary size is conference, specifically the Big Ten.

Bonus:

A map showing median pay by conference (figure 9) was plotted using ggplot. The color of the dot is the conference placed where the conference headquarters are located. The size of the dot represents the median pay for the conference. The independents don't have a conference headquarters, so they were placed in Indianapolis, IN, which is where the NCAA headquarters is located. As expected, the median pay for the Power 5 conferences are the largest.

Figure 9 – Median pay by conference plotted on US map.