Final

Sienna Mori

Dec 15th, 2017

I chose to create an analysis based off of NFL Quarterback Statistics, dating back to 1996. These statistics include completions, attempts, yards, yards per attempt, touchdowns, intercepts, longest throw, game points, whether the game was home or away, and rating. Each year has multiple data points for a particular player, due to the fact that there are multiple games in a season.

In this anlaysis, my goal was to answer questions related to the strength of a player based on his stats, as well as find whether certain statistics have a significant relationship with a player's game points.

In order to begin answering these questions, first I coded a Player object, which includes the players, the year, completions (cmp), touchdown interception ratio (td/int), and yards per attempt (ypa). I had to calculate the td/int ratio. Since there are multiple data points for each player for each year, in my Player object, I averaged the data points for each statistic.

In order to answer the following questions, I used get methods. My first question aimed to find the player with the highest yards per attempt for each year. Second, I found the player with the highest number of completions for each year.

Finally, I found the td/int, ypa, and cmp for my favorite players in 2016. I wanted to look at the most recent year provided. I wanted to test if my favorite quarterbacks (Eli Manning, Peyton Manning ,and Aaron Rodgers) do in fact possess the skill I believe them to possess, indicated by their statistics.

I found that in 2016, Aaron Rodgers had the highest number of completions, ypa, and td/int ratio of the three.

Next, in order to answer my correlation questions, I coded a Game Object, where I included game points, rating, whether a game was home or away, and percentage completion. I had to calculate percentage completion. I then ran regressions of game points against both rating and percentage completion. Percentage completion is thought to have a high correlation with game points, and thus I wanted to test this theory. I also wanted to test whether a quarterback's rating has a high correlation with game points. I ran these regressions for each year and provided the regression plots for the year that had the most significant p values. I found that 2009 and 1997 had the most significant p values for percentage completion and rating, respectively. See below for the graphs. Finally, I looked at the relationship between whether a game was home or away (I converted this variable into a dummy variable) and game points for all of the years. 1 represents an Away game. I found, as predicted, there is a highly significant negative relationship between whether a game was away and the game points.

Finally, I ran a PCA analysis of my Game Points Object in an effort to see if there were any trends among players with high or low ratings/game points/percentage completion. See below for the PCA. I interpret the PCA as showing me that similar types of players (those who tend to have weaker statistics) tend to perform similarly. It makes sense that those on the low end of pc1 are also on the low end of pc2 and similarly for those on the high end.

Overall, this analysis was incredibly interesting, in terms of understanding how you can analyze NFL quarterback statistics in order to assess an individual player's strength, as well as the relationships between an individual player's stats and his success in a game (measured by game points).

```
##  [1] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
##  [3] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
##  [5] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
##  [7] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
##  [9] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
```

```
## [11] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
## [13] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
## [15] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
## [17] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
## [19] "Kevin LockettK.\xe5\xcaLockett" "Kevin LockettK.\xe5\xcaLockett"
## [21] "Kevin LockettK.\xe5\xcaLockett"
```

```
##  [1] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
##  [3] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
##  [5] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
##  [7] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
##  [9] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
## [11] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
## [13] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
## [15] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
## [17] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
## [19] "Shannon WoodsS.\xe5\xcaWoods" "Shannon WoodsS.\xe5\xcaWoods"
## [21] "Shannon WoodsS.\xe5\xcaWoods"
```

```
## [1] 1.469767
```

```
## [1] 7.030808
```

```
## [1] 20.39899
```

```
## [1] 2.120482
```

```
## [1] 7.754373
```

```
## [1] 23.08365
```

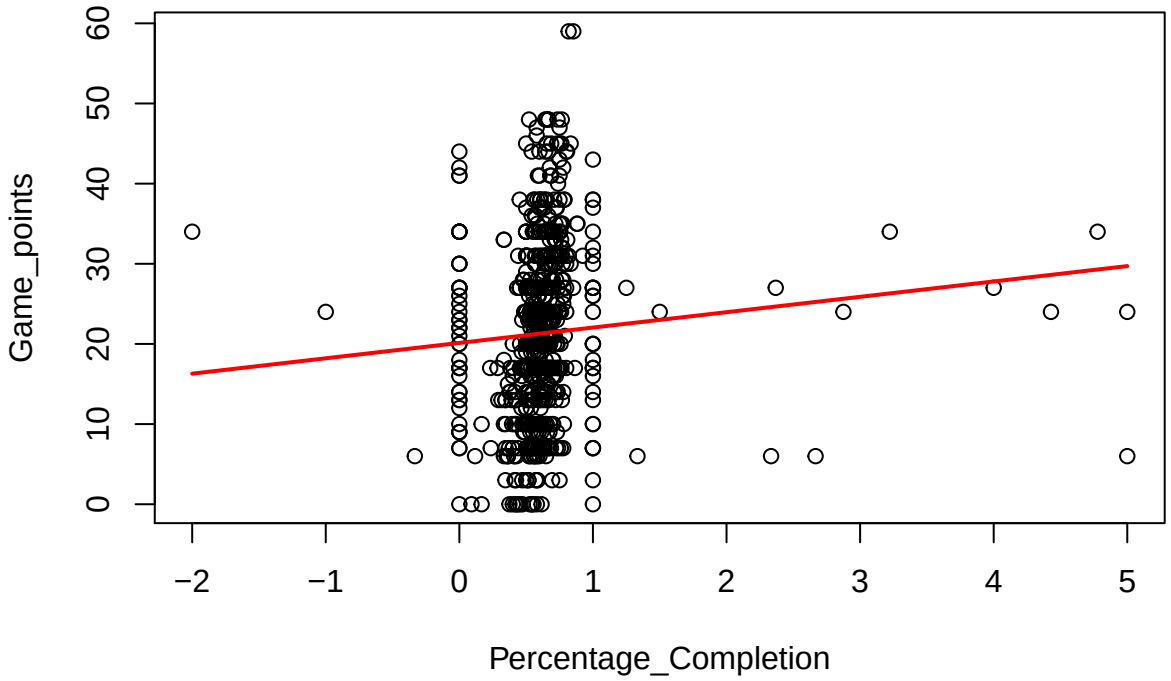
```
## [1] 4.125
```

```
## [1] 7.941844
```

```
## [1] 21.51773
```

```
##  [1] 4.840764e-06 1.019146e-06 3.317634e-11 2.381971e-09 2.591620e-08
##  [6] 2.729284e-09 5.550851e-12 1.331299e-06 3.051283e-06 7.805592e-16
## [11] 3.068178e-08 1.143451e-12 3.423662e-09 3.713291e-02 5.759117e-09
## [16] 1.047251e-07 9.432365e-05 4.649347e-10 1.072449e-07 1.586008e-08
## [21] 2.001955e-05
```

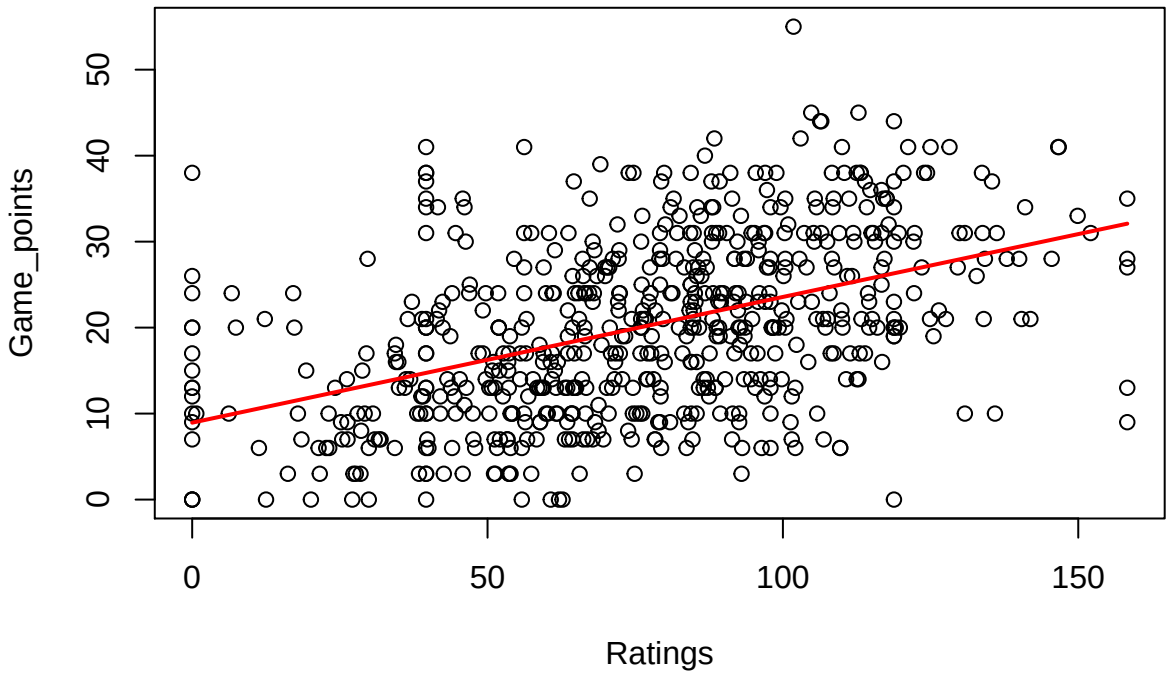
```
## [1] 2009
```




```
##
## Call:
## lm(formula = Game_points ~ Percentage_Completion, data = gp_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.714  -8.075  -0.946   6.874  37.305
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            20.1263     0.7134  28.212   <2e-16 ***
## Percentage_Completion   1.9176     0.9182   2.088   0.0371 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.06 on 668 degrees of freedom
## Multiple R-squared:  0.006487,   Adjusted R-squared:  0.005
## F-statistic: 4.362 on 1 and 668 DF,  p-value: 0.03713
```

```
##  [1] 3.838234e-34 1.011881e-33 1.348898e-40 5.478440e-43 8.228526e-42
##  [6] 2.331255e-48 6.742391e-54 6.767466e-39 2.987750e-38 2.265642e-46
## [11] 1.103306e-42 1.306244e-41 5.244455e-47 6.053162e-46 1.246540e-52
## [16] 3.394942e-37 2.266954e-38 1.944774e-46 7.041707e-44 1.719524e-46
## [21] 7.266132e-42
```

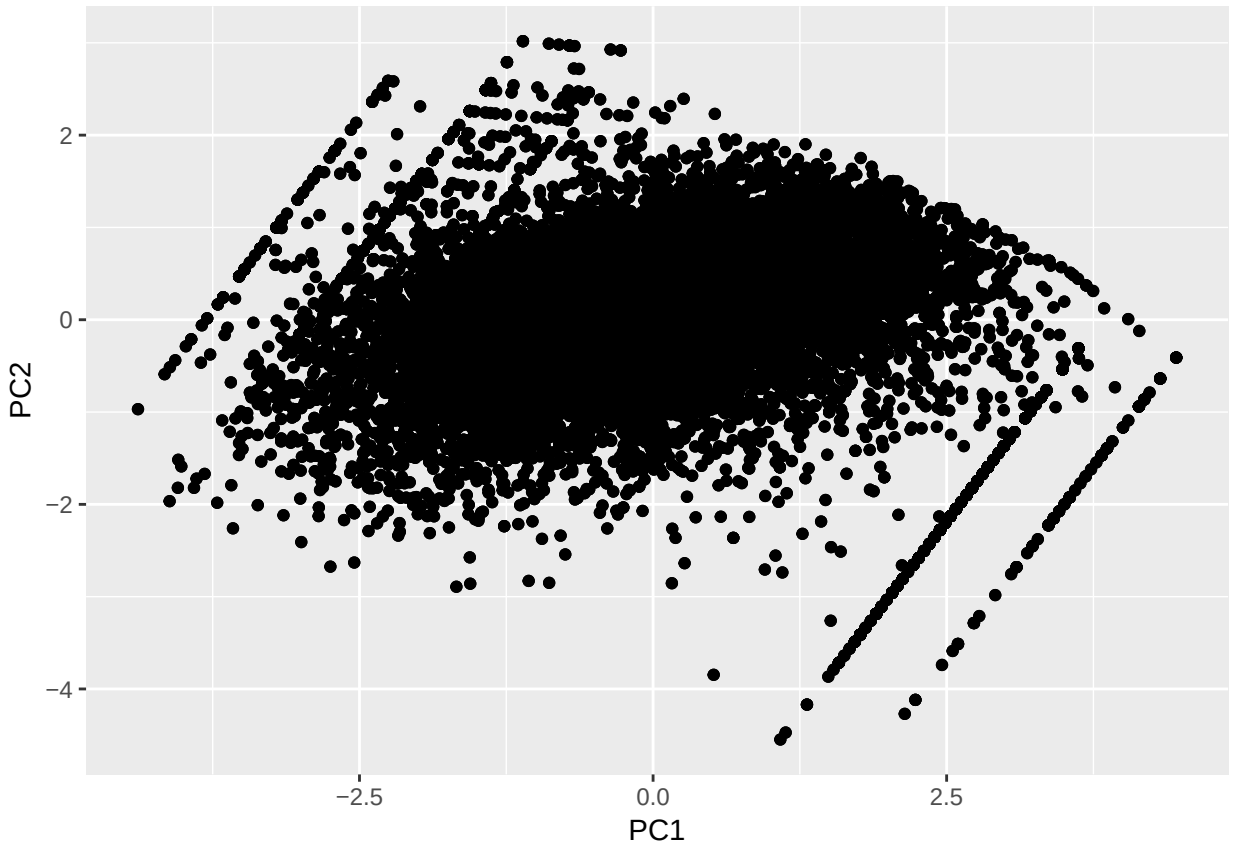
```
## [1] 1997
```




```
##
## Call:
## lm(formula = Game_points ~ Ratings, data = gp_df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.3147  -6.4613  -0.7964   6.2734  31.1710
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.94454    0.93906   9.525   <2e-16 ***
## Ratings      0.14621    0.01137  12.858   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.18 on 622 degrees of freedom
## Multiple R-squared:   0.21,  Adjusted R-squared:  0.2087
## F-statistic: 165.3 on 1 and 622 DF,  p-value: < 2.2e-16
```

```
## Called from: get_home_relationship.Game_points(gp)
## debug: reg <- lm(gp$Game_points ~ gp$Home_Away)
## debug: coef <- reg$coefficients
## debug: m <- coef[2]
## debug: b <- coef[1]
## debug: p_value <- summary(reg)$coefficients[2, 4]
## debug: return(p_value)
```

```
## [1] 6.524456e-53
```


2
0
−2
−4
PC2
−2.5
0.0
2.5
PC1