DATA 605 Discussion Week 11 (12)

CUNY Spring 2021

Philip Tanofsky

17 April 2021

```
library(tidyverse)
library(stats)
library(broom)
library(ggplot2)
```

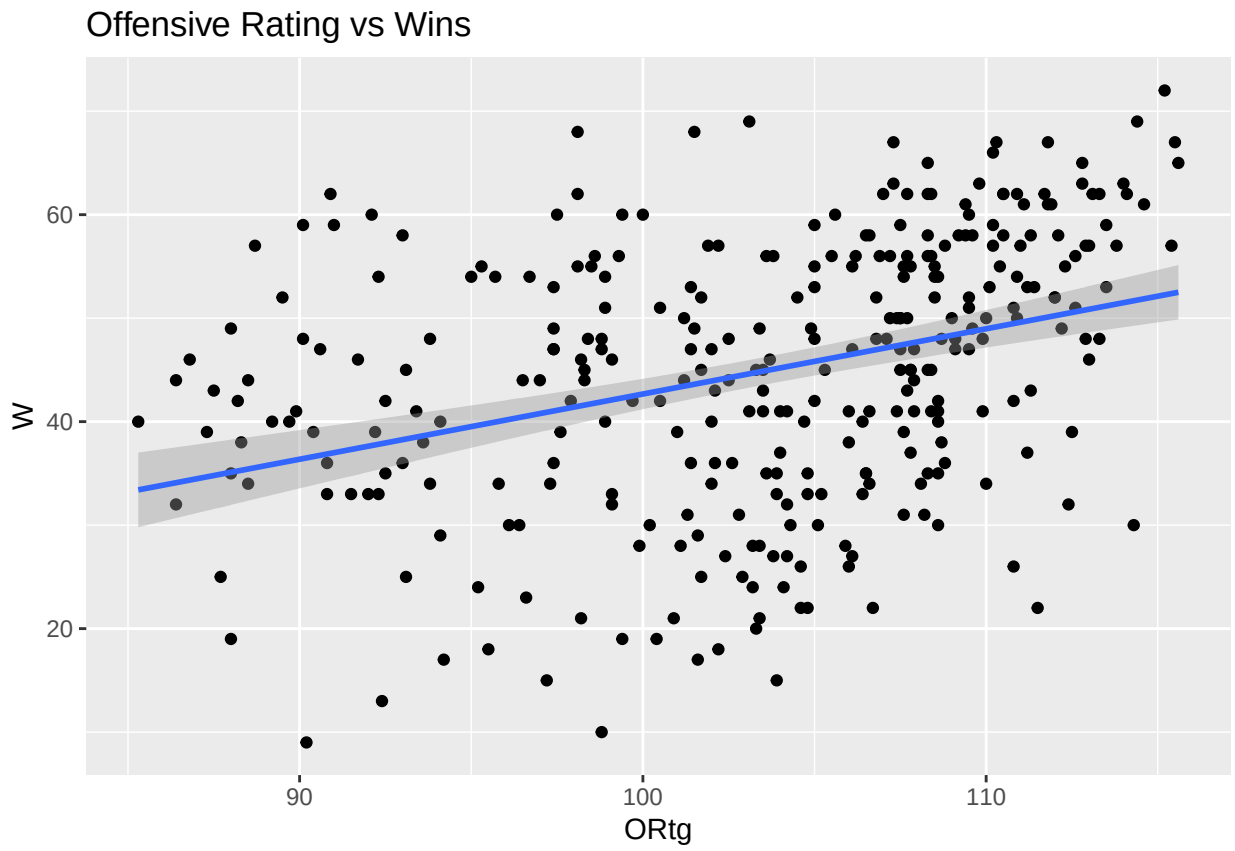
```
nba_data <- read_csv("https://raw.githubusercontent.com/fivethirtyeight/nba-player-advanced-metrics/mas

nba_data <- nba_data %>% filter(year_id >= 2016) %>% select( 'P/36',  'A/36', 'R/36', 'Raptor O')

nba_data <- rename_with(nba_data, ~ gsub(" ", "_", .x, fixed = TRUE))
nba_data <- rename_with(nba_data, ~ gsub("/", "_", .x, fixed = TRUE))

nba_data$P_36 <- as.numeric(nba_data$P_36)
nba_data$A_36 <- as.numeric(nba_data$A_36)
nba_data$R_36 <- as.numeric(nba_data$R_36)

nba_data <- na.omit(nba_data)

nba_data <- nba_data %>% filter(P_36 <= 50)

dim(nba_data)

summary(nba_data)

plot(x = nba_data$P_36, y = nba_data$Raptor_O, main = "Raptor Offensive Rating by Points per 36")

plot(x = nba_data$A_36, y = nba_data$Raptor_O, main = "Raptor Offensive Rating by Assists per 36")

plot(x = nba_data$R_36, y = nba_data$Raptor_O, main = "Raptor Offensive Rating by Rebounds per 36")

model <- lm(Raptor_O ~ ., data=nba_data)
#model

summary(model)

model.diag.metrics <- augment(model)
head(model.diag.metrics)

#autoplot(model)
```

```r
library(nbastatR)

#players_2019 <- teams_players_stats(seasons = 2019, types=player, return_message = T)

players_2018 <- teams_players_stats(seasons = 2018,
                    types = c("player", "team"),
                    modes = c("PerGame", "Totals"),
                    tables = c("general", "defense", "clutch", "hustle", "shots", "shot locations")
                   )

dim(players_2018)

summary(players_2018)
```

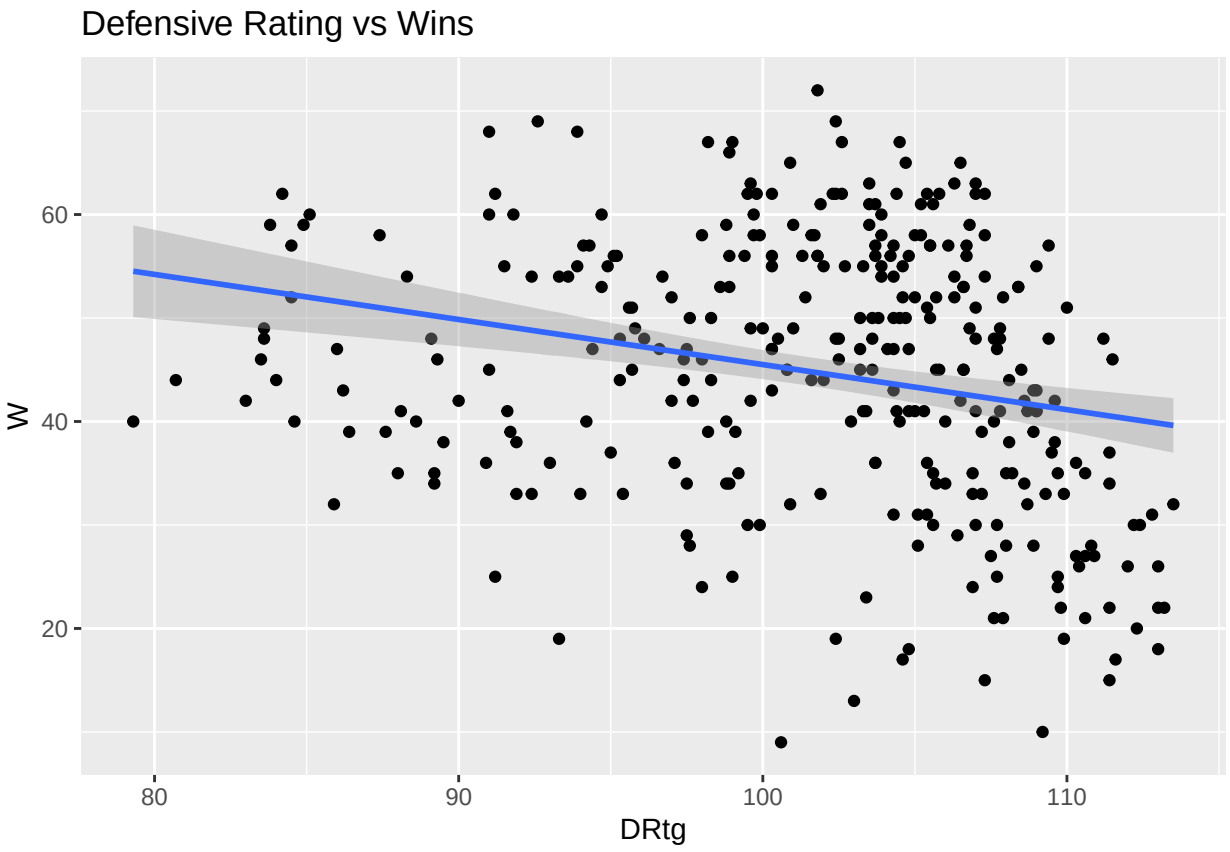
```r
library(NBAloveR)

bos <- getTeamHistory(team_code = "bos")
lal <- getTeamHistory(team_code = "lal")
chi <- getTeamHistory(team_code = "chi")
sas <- getTeamHistory(team_code = "sas")
phi <- getTeamHistory(team_code = "phi")

nba <- rbind(bos, lal)
nba <- rbind(nba, chi)
nba <- rbind(nba, sas)
nba <- rbind(nba, phi)

nba$RtgDif <- nba$ORtg - nba$DRtg

head(nba)
```

```
##    Season  Lg           Team  W  L  W/L%   Finish  SRS Pace RelativePace  ORtg
## 1 2020-21 NBA Boston Celtics 30 26 0.536 3rd of 5 2.40 97.9         -1.3 114.3
## 2 2019-20 NBA Boston Celtics 48 24 0.667 2nd of 5 5.83 99.5         -0.8 113.3
## 3 2018-19 NBA Boston Celtics 49 33 0.598 3rd of 5 3.90 99.6         -0.4 112.2
## 4 2017-18 NBA Boston Celtics 55 27 0.671 2nd of 5 3.23 96.0         -1.3 107.6
## 5 2016-17 NBA Boston Celtics 53 29 0.646 1st of 5 2.25 96.8          0.4 111.2
## 6 2015-16 NBA Boston Celtics 48 34 0.585 2nd of 5 2.84 98.5          2.7 106.8
##   RelativeORtg  DRtg RelativeDRtg               Playoffs           Coaches
## 1          2.2 112.2          0.1                        B. Stevens (30-26)
## 2          2.7 107.0         -3.6   Lost E. Conf. Finals B. Stevens (48-24)
## 3          1.8 107.8         -2.6    Lost E. Conf. Semis B. Stevens (49-33)
## 4         -1.0 103.9         -4.7   Lost E. Conf. Finals B. Stevens (55-27)
## 5          2.4 108.4         -0.4   Lost E. Conf. Finals B. Stevens (53-29)
## 6          0.4 103.6         -2.8 Lost E. Conf. 1st Rnd. B. Stevens (48-34)
##        TopWinShare RtgDif
## 1   J. Tatum (4.9)    2.1
## 2   J. Tatum (6.9)    6.3
## 3  K. Irving (9.1)    4.4
## 4  K. Irving (8.9)    3.7
## 5 I. Thomas (12.5)    2.8
## 6  I. Thomas (9.7)    3.2
```

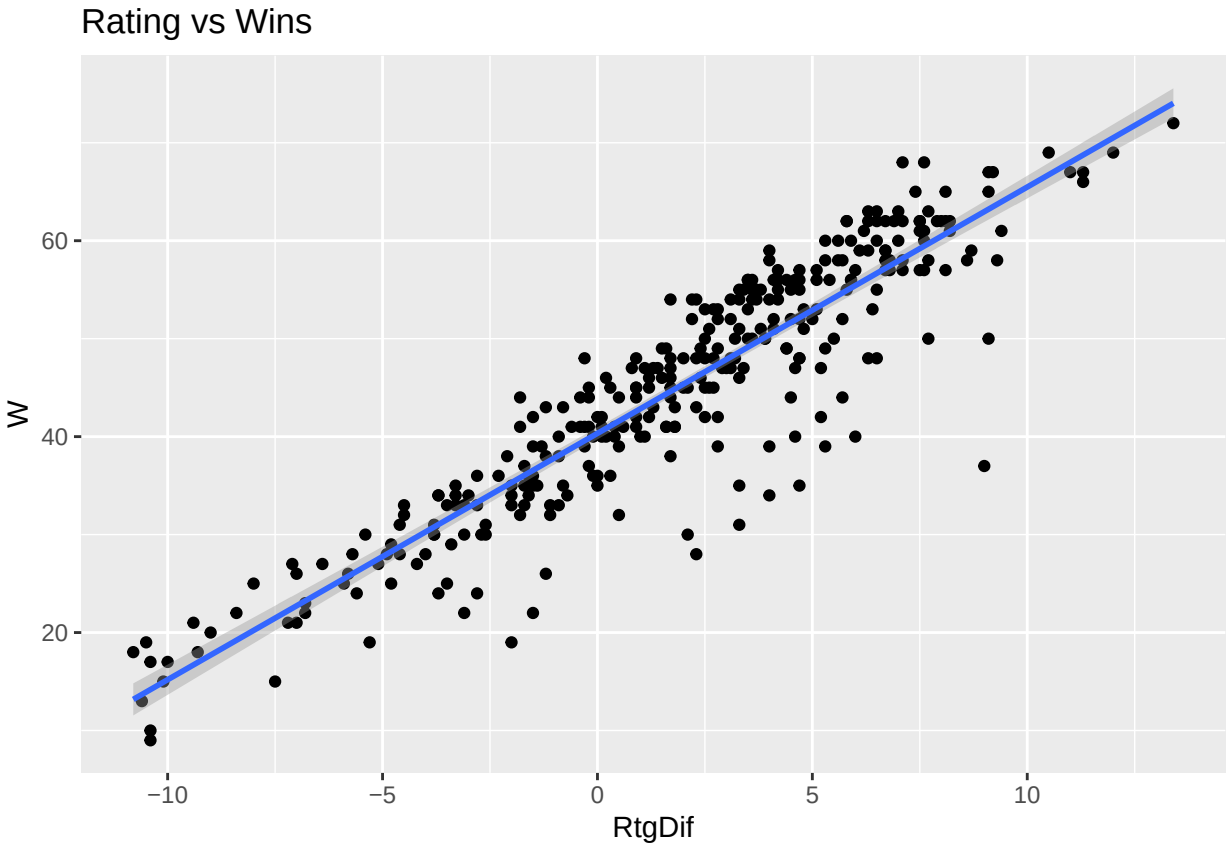
```
nba %>%
  ggplot(aes(x=ORtg, y=W)) +
  geom_point() +
  labs(title = 'Offensive Rating vs Wins') + geom_smooth(method='lm', formula= y~x)
```




```
nba %>%
  ggplot(aes(x=DRtg, y=W)) +
  geom_point() +
  labs(title = 'Defensive Rating vs Wins') + geom_smooth(method='lm', formula= y~x)
```




```
nba %>%
  ggplot(aes(x=RtgDif, y=W)) +
  geom_point() +
  labs(title = 'Rating vs Wins') + geom_smooth(method='lm', formula= y~x)
```




```
corr <- cor(nba$RtgDif, nba$W, use = "complete.obs")
corr
```

```
## [1] 0.9151873
```

```
dim(nba)
```

```
## [1] 329  18
```

```
summary(nba)
```

```
##     Season              Lg                Team                W         
##  Length:329         Length:329         Length:329         Min.   : 9.00  
##  Class :character   Class :character   Class :character   1st Qu.:35.00  
##  Mode  :character   Mode  :character   Mode  :character   Median :46.00  
##                                                           Mean   :44.56  
##                                                           3rd Qu.:55.00  
##                                                           Max.   :72.00  
##                                                                          
##        L              W/L%           Finish               SRS         
##  Min.   :10.00   Min.   :0.1100   Length:329         Min.   :-11.500  
##  1st Qu.:25.00   1st Qu.:0.4560   Class :character   1st Qu.: -1.340  
##  Median :33.00   Median :0.5850   Mode  :character   Median :  2.250  
##  Mean   :34.58   Mean   :0.5628                      Mean   :  1.693  
```

```
##  3rd Qu.:42.00   3rd Qu.:0.6830                      3rd Qu.:  5.020
##  Max.   :73.00   Max.   :0.8780                      Max.   : 11.800
##
##       Pace        RelativePace         ORtg         RelativeORtg
##  Min.   : 87.3   Min.   :-8.500   Min.   : 85.30   Min.   :-10.1000
##  1st Qu.: 93.1   1st Qu.:-1.300   1st Qu.: 98.65   1st Qu.: -1.5000
##  Median : 98.5   Median : 0.400   Median :104.75   Median :  0.7500
##  Mean   :101.3   Mean   : 1.083   Mean   :103.38   Mean   :  0.5311
##  3rd Qu.:106.5   3rd Qu.: 2.275   3rd Qu.:108.60   3rd Qu.:  2.6000
##  Max.   :136.3   Max.   :16.700   Max.   :115.60   Max.   :  7.7000
##  NA's   :7       NA's   :7        NA's   :7        NA's   :7
##       DRtg        RelativeDRtg       Playoffs           Coaches
##  Min.   : 79.3   Min.   :-10.800   Length:329         Length:329
##  1st Qu.: 97.6   1st Qu.: -3.575   Class :character   Class :character
##  Median :103.5   Median : -1.000   Mode  :character   Mode  :character
##  Mean   :101.6   Mean   : -1.243
##  3rd Qu.:107.0   3rd Qu.:  0.800
##  Max.   :113.5   Max.   :  5.600
##  NA's   :7       NA's   :7
##  TopWinShare            RtgDif
##  Length:329         Min.   :-10.800
##  Class :character   1st Qu.: -1.200
##  Mode  :character   Median :  2.400
##                     Mean   :  1.777
##                     3rd Qu.:  5.175
##                     Max.   : 13.400
##                     NA's   :7
```

```
model_ortg <- lm(W ~ ORtg, data=nba)
summary(model_ortg)
```

```
##
## Call:
## lm(formula = W ~ ORtg, data = nba)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.914  -8.165   1.713   9.013  26.528
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.34924    9.76658  -2.084    0.038 *
## ORtg          0.63019    0.09424   6.687 1.02e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.14 on 320 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.1226, Adjusted R-squared:  0.1199
## F-statistic: 44.71 on 1 and 320 DF,  p-value: 1.017e-10
```

```
model_drtg <- lm(W ~ DRtg, data=nba)
summary(model_drtg)
```

```
##
## Call:
## lm(formula = W ~ DRtg, data = nba)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.239  -9.704   0.741  10.228  27.284
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 89.06285    9.78860   9.099  < 2e-16 ***
## DRtg        -0.43562    0.09609  -4.533 8.21e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.56 on 320 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.06035,    Adjusted R-squared:  0.05741
## F-statistic: 20.55 on 1 and 320 DF,  p-value: 8.214e-06
```

```
model_rtgdif <- lm(W ~ RtgDif, data=nba)
summary(model_rtgdif)
```

```
##
## Call:
## lm(formula = W ~ RtgDif, data = nba)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.9646  -2.2960   0.7784   3.3895   9.8133
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.33265    0.31117  129.62   <2e-16 ***
## RtgDif       2.51466    0.06191   40.62   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.223 on 320 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.8376, Adjusted R-squared:  0.8371
## F-statistic:  1650 on 1 and 320 DF,  p-value: < 2.2e-16
```