Doubles for Dollars

Load Libraries

```r
library(Lahman)
library(tidyverse)
library(broom)
library(modelsummary)
library(fixest)
library(quantmod)
library(scales)
```

Load Datasets & Filter

```r
Batting <- Lahman::Batting
Appearances <- Lahman::Appearances

#DF of position players
PositionAppearances <- Appearances %>%
  #Gross filter to find position players
  #Arbitrarily selected 5 because that filtered out a lot of PO's who came in a couple tim
  filter(G_c >= 5 | G_1b >= 5 | G_2b >= 5 | G_3b >= 5 | G_ss >= 5 | G_lf >= 5 | G_cf >= 5
  arrange(desc(G_p))

#Filter batting data by position players
PositionBatting <- Batting %>%
  filter(playerID %in% PositionAppearances$playerID) %>%
  #Only grab players who played in or after 1985
  #Omit 2020 because it creates 30 outliers with only 60 games played
  filter(yearID >= 1985 & yearID <= 2016) %>%
```

```
  #Isolate singles as X1B to keep naming convention
  mutate(X1B = H - X2B - X3B - HR, .after = H) %>%
  #Introduce "catch-all" term for "free base"
  mutate(OB = BB + HBP + IBB, .after = HBP)


PositionBatting$yearID <- ISOdate(PositionBatting$yearID, 12, 1)
PositionBatting$yearID <- as.Date(PositionBatting$yearID)
PositionBatting %>%
  glimpse()
```

```
Rows: 20,566
Columns: 24
$ playerID <chr> "adamsri02", "aguaylu01", "aikenwi01", "allenga01", "almonbi0~
$ yearID   <date> 1985-12-01, 1985-12-01, 1985-12-01, 1985-12-01, 1985-12-01, ~
$ stint    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
$ teamID   <fct> SFN, PHI, TOR, TOR, PIT, LAN, BOS, HOU, CLE, NYN, HOU, LAN, C~
$ lgID     <fct> NL, NL, AL, AL, NL, NL, AL, NL, AL, NL, NL, NL, AL, AL, AL, A~
$ G        <int> 54, 91, 12, 14, 88, 77, 103, 65, 46, 145, 114, 74, 160, 15, 1~
$ AB       <int> 121, 165, 20, 34, 244, 221, 385, 189, 76, 520, 332, 118, 640,~
$ R        <int> 12, 27, 2, 2, 33, 24, 50, 20, 10, 77, 47, 8, 86, 4, 48, 74, 1~
$ H        <int> 23, 46, 4, 4, 66, 44, 102, 53, 19, 142, 88, 29, 198, 5, 92, 1~
$ X1B      <int> 17, 30, 2, 3, 43, 34, 57, 37, 10, 112, 64, 25, 144, 4, 62, 80~
$ X2B      <int> 3, 7, 1, 1, 17, 6, 17, 8, 7, 24, 14, 3, 29, 1, 15, 28, 4, 4, ~
$ X3B      <int> 1, 3, 0, 0, 0, 0, 5, 0, 0, 5, 0, 1, 3, 0, 1, 2, 1, 1, 9, 0, 0~
$ HR       <int> 2, 6, 1, 0, 6, 4, 23, 8, 2, 1, 10, 0, 22, 0, 14, 36, 0, 1, 27~
$ RBI      <int> 10, 21, 5, 3, 29, 18, 64, 25, 15, 38, 45, 7, 113, 1, 52, 88, ~
$ SB       <int> 1, 1, 0, 0, 10, 5, 0, 0, 0, 30, 0, 1, 1, 0, 2, 1, 0, 8, 22, 0~
$ CS       <int> 1, 0, 0, 0, 7, 4, 0, 0, 0, 12, 2, 0, 1, 0, 1, 1, 1, 2, 8, 1, ~
$ BB       <int> 5, 22, 3, 0, 22, 35, 18, 24, 4, 36, 67, 3, 42, 0, 50, 52, 22,~
$ SO       <int> 23, 26, 6, 10, 61, 42, 90, 27, 17, 72, 70, 5, 89, 9, 47, 166,~
$ IBB      <int> 3, 5, 0, 0, 0, 3, 4, 2, 1, 1, 13, 0, 8, 0, 0, 4, 0, 0, 5, 0, ~
$ HBP      <int> 1, 6, 0, 0, 1, 1, 2, 1, 0, 1, 1, 1, 1, 0, 0, 5, 0, 0, 4, 0, 2~
$ OB       <int> 9, 33, 3, 0, 23, 39, 24, 27, 5, 38, 81, 4, 51, 0, 50, 61, 22,~
$ SH       <int> 3, 4, 0, 0, 4, 4, 0, 1, 0, 14, 1, 8, 0, 0, 0, 0, 2, 1, 0, 0, ~
$ SF       <int> 0, 3, 1, 0, 3, 1, 5, 1, 1, 3, 1, 0, 10, 0, 3, 5, 2, 0, 3, 0, ~
$ GIDP     <int> 2, 7, 1, 1, 6, 4, 14, 9, 2, 3, 16, 3, 23, 0, 12, 14, 6, 1, 14~
```

```
Teams <- Lahman::Teams


Teams1985 <- Teams %>%
```

```
    #Omit 2020 because it creates 30 outliers with only 60 games played
    filter(yearID >= 1985 & yearID <= 2016) %>%
    mutate(X1B = H - X2B - X3B - HR, .after = H) %>%
    mutate(OB = BB + HBP, .after = HBP)

  #Put this after because ">=" no work with factor variables
  Teams1985$yearID <- as.factor(Teams1985$yearID)
```

Create Models

Fixed Effects for Linear Weights

Variable	Min	Q1	Median	Q3	Max	Mean	SD
X1B	646	925	971	1015.75	1186	965.84	77.97
X2B	159	255	276	297	376	275.29	34.16
X3B	6	24	30	37	61	30.97	9.08
HR	58	130	157	183	264	158.06	37.18
OB	350	521	571.5	623	841	574.71	77.51
SB	19	77	100	126	314	104.26	36.78
SF	23	39	45	51	75	45.36	8.89

```
  TeamLW <- lm(data = Teams1985, R ~ X1B + X2B + X3B + HR + OB + SB + SF + yearID + teamID)
  summary(TeamLW)
```

```
Call:
lm(formula = R ~ X1B + X2B + X3B + HR + OB + SB + SF + yearID +
    teamID, data = Teams1985)

Residuals:
    Min      1Q  Median      3Q     Max
-73.858 -12.798   0.544  13.313  86.015

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -402.10588   20.16094 -19.945  < 2e-16 ***
X1B            0.46682    0.01514  30.829  < 2e-16 ***
X2B            0.71788    0.03483  20.612  < 2e-16 ***
```

```
X3B           1.09385    0.10181  10.744  < 2e-16 ***
HR            1.42873    0.03021  47.291  < 2e-16 ***
OB            0.31567    0.01316  23.979  < 2e-16 ***
SB            0.16677    0.02409   6.924 8.69e-12 ***
SF            0.75466    0.10369   7.278 7.75e-13 ***
yearID1986   -3.55445    5.89572  -0.603 0.546746
yearID1987   -0.08698    5.98758  -0.015 0.988413
yearID1988   -0.69266    5.94858  -0.116 0.907330
yearID1989   -0.20663    5.92743  -0.035 0.972200
yearID1990   -6.93298    5.91537  -1.172 0.241516
yearID1991    2.13870    5.90047   0.362 0.717097
yearID1992  -13.98692    5.96269  -2.346 0.019220 *
yearID1993    0.13140    5.85821   0.022 0.982110
yearID1994  118.68557    7.42787  15.978  < 2e-16 ***
yearID1995   45.65247    5.97297   7.643 5.75e-14 ***
yearID1996    2.47230    6.03597   0.410 0.682207
yearID1997  -12.36503    5.98414  -2.066 0.039104 *
yearID1998   -2.94150    5.96651  -0.493 0.622139
yearID1999   -2.48392    6.03841  -0.411 0.680917
yearID2000   -1.03954    6.10273  -0.170 0.864784
yearID2001   -6.54870    6.03034  -1.086 0.277807
yearID2002  -10.10806    5.97533  -1.692 0.091085 .
yearID2003   -4.16172    6.00779  -0.693 0.488675
yearID2004  -10.25438    6.06503  -1.691 0.091256 .
yearID2005  -10.27609    6.03991  -1.701 0.089242 .
yearID2006  -10.19673    6.10878  -1.669 0.095450 .
yearID2007   -8.16734    6.09555  -1.340 0.180643
yearID2008  -11.21230    6.05830  -1.851 0.064557 .
yearID2009  -17.12226    5.96238  -2.872 0.004185 **
yearID2010  -10.48366    5.93208  -1.767 0.077542 .
yearID2011  -13.82462    5.91885  -2.336 0.019740 *
yearID2012  -10.07002    5.94498  -1.694 0.090659 .
yearID2013  -20.25306    5.98170  -3.386 0.000742 ***
yearID2014  -11.18874    5.95179  -1.880 0.060466 .
yearID2015  -13.00858    5.97584  -2.177 0.029767 *
yearID2016  -11.20570    6.04378  -1.854 0.064075 .
teamIDARI   -18.07340    9.20543  -1.963 0.049934 *
teamIDATL    -6.02032    8.53718  -0.705 0.480888
teamIDBAL   -11.89746    8.56039  -1.390 0.164947
teamIDBOS   -12.89464    8.72166  -1.478 0.139657
teamIDCAL     5.40744    9.91391   0.545 0.585595
teamIDCHA     1.48086    8.51471   0.174 0.861972
teamIDCHN   -12.35559    8.54335  -1.446 0.148485
```

```
teamIDCIN     -15.14258     8.55634  -1.770 0.077129 .
teamIDCLE      -4.36965     8.51004  -0.513 0.607757
teamIDCOL       1.66507     8.90601   0.187 0.851737
teamIDDET      -8.86224     8.57126  -1.034 0.301456
teamIDFLO     -21.93233     9.06902  -2.418 0.015800 *
teamIDHOU      -9.93657     8.54685  -1.163 0.245320
teamIDKCA      -0.99434     8.51908  -0.117 0.907110
teamIDLAA      13.02415     9.84307   1.323 0.186134
teamIDLAN      -8.32157     8.51866  -0.977 0.328916
teamIDMIA     -18.01891    12.39275  -1.454 0.146321
teamIDMIL     -20.48999     9.13261  -2.244 0.025116 *
teamIDMIN      -1.79462     8.50806  -0.211 0.832991
teamIDML4      16.10230     9.71212   1.658 0.097697 .
teamIDMON     -16.66640     9.13323  -1.825 0.068383 .
teamIDNYA      -2.18639     8.60548  -0.254 0.799503
teamIDNYN      -5.44366     8.54168  -0.637 0.524098
teamIDOAK       6.71799     8.62423   0.779 0.436217
teamIDPHI     -11.94623     8.59939  -1.389 0.165139
teamIDPIT     -11.39043     8.59013  -1.326 0.185201
teamIDSDN      -9.69842     8.55733  -1.133 0.257391
teamIDSEA     -10.92560     8.49511  -1.286 0.198759
teamIDSFN      -6.23563     8.57087  -0.728 0.467098
teamIDSLN      -4.45423     8.50820  -0.524 0.600748
teamIDTBA     -18.30928     9.09523  -2.013 0.044425 *
teamIDTEX       0.74155     8.53919   0.087 0.930818
teamIDTOR      -7.97915     8.59718  -0.928 0.353615
teamIDWAS     -16.99768     9.92703  -1.712 0.087215 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.16 on 845 degrees of freedom
Multiple R-squared:  0.9489,     Adjusted R-squared:  0.9446
F-statistic:   218 on 72 and 845 DF,  p-value: < 2.2e-16
```

```r
feols(R ~ X1B + X2B + X3B + HR + OB + SB + SF | yearID + teamID, data = Teams1985)
```

```
OLS estimation, Dep. Var.: R
Observations: 918
Fixed-effects: yearID: 32,  teamID: 35
Standard-errors: Clustered (yearID)
    Estimate Std. Error  t value   Pr(>|t|)
```

```
X1B 0.466818   0.015246 30.61960   < 2.2e-16 ***
X2B 0.717882   0.038288 18.74943   < 2.2e-16 ***
X3B 1.093846   0.113169  9.66563 7.1744e-11 ***
HR  1.428731   0.033677 42.42390   < 2.2e-16 ***
OB  0.315669   0.013954 22.62246   < 2.2e-16 ***
SB  0.166770   0.021111  7.89956 6.4494e-09 ***
SF  0.754662   0.105617  7.14525 4.9617e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 20.3     Adj. R2: 0.944559
             Within R2: 0.880088
```

Formula

Current formula:

$$xRC = .47X1B + .72X2B + 1.09X3B + 1.43HR + .32OB + .17SB + .75SF$$

```
exRC <- function(a, b, c, d, e, f, g){
  return(.466818*a + .717882*b + 1.093846*c + 1.428731*d + .315669*e + .166770*f + .754662
}
```

Find Player's xRC

```
PositionBatting <- PositionBatting %>%
  mutate(xRC = exRC(X1B, X2B, X3B, HR, OB, SB, SF), .after = lgID) %>%
  select(playerID, yearID, teamID, xRC, R, X1B, X2B, X3B, HR, OB, SB, SF)
```

Salary

```
Salaries <- Lahman::Salaries
Salaries %>%
  glimpse()
```

```
Rows: 26,428
Columns: 5
```

```
$ yearID   <int> 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1~
$ teamID   <fct> ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, A~
$ lgID     <fct> NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, N~
$ playerID <chr> "barkele01", "bedrost01", "benedbr01", "campri01", "ceronri01~
$ salary   <int> 870000, 550000, 545000, 633333, 625000, 800000, 150000, 48333~
```

```r
getSymbols("CPIAUCSL", src='FRED')
```

```
[1] "CPIAUCSL"
```

```r
avg.cpi <- apply.yearly(CPIAUCSL, mean)
cf <- avg.cpi/as.numeric(avg.cpi['2021'])
dat <- merge(Salaries, cf, all=FALSE)
newdat <- dat %>%
  mutate(INFadj = salary / CPIAUCSL)


Salaries$yearID <- ISOdate(Salaries$yearID, 12, 1)
Salaries$yearID <- as.Date(Salaries$yearID)

cf2 <- cf %>%
  fortify.zoo() %>%
  as_tibble() %>%
  rename(yearID = Index)

Salaries <- inner_join(Salaries, cf2)

Salaries <- Salaries %>%
  mutate(real_salary = salary/CPIAUCSL, .after = salary)

Salaries %>%
  glimpse()
```

```
Rows: 26,428
Columns: 7
$ yearID      <date> 1985-12-01, 1985-12-01, 1985-12-01, 1985-12-01, 1985-12-0~
$ teamID      <fct> ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL, ATL~
$ lgID        <fct> NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL, NL~
$ playerID    <chr> "barkele01", "bedrost01", "benedbr01", "campri01", "ceronr~
$ salary      <int> 870000, 550000, 545000, 633333, 625000, 800000, 150000, 48~
$ real_salary <dbl> 2190891.4, 1385046.3, 1372454.9, 1594900.9, 1573916.2, 201~
$ CPIAUCSL    <dbl> 0.3970986, 0.3970986, 0.3970986, 0.3970986, 0.3970986, 0.3~
```

Salary and Runs Created

```r
df1 <- PositionBatting %>%
  select(playerID, xRC,yearID)

df2 <- Salaries %>%
  select(playerID, real_salary,yearID)

df3 <- inner_join(df2,df1, by=c("playerID", "yearID"))
```

```r
ggplot(data=df3)+
  aes(x=xRC,
      y=real_salary)+
  geom_point(aes(color=yearID))+
  geom_smooth(method="lm")+
  scale_y_continuous(labels = label_comma())+
  labs(x="Expected Runs Created (xRC)",
       y="Real Salary (2021, $)",
       title="Real Salaries From 1985 to 2016")
```


Real Salaries From 1985 to 2016