PARK FACTOR

VICTOR KYLE SHIH

WHAT IS PARK FACTOR?

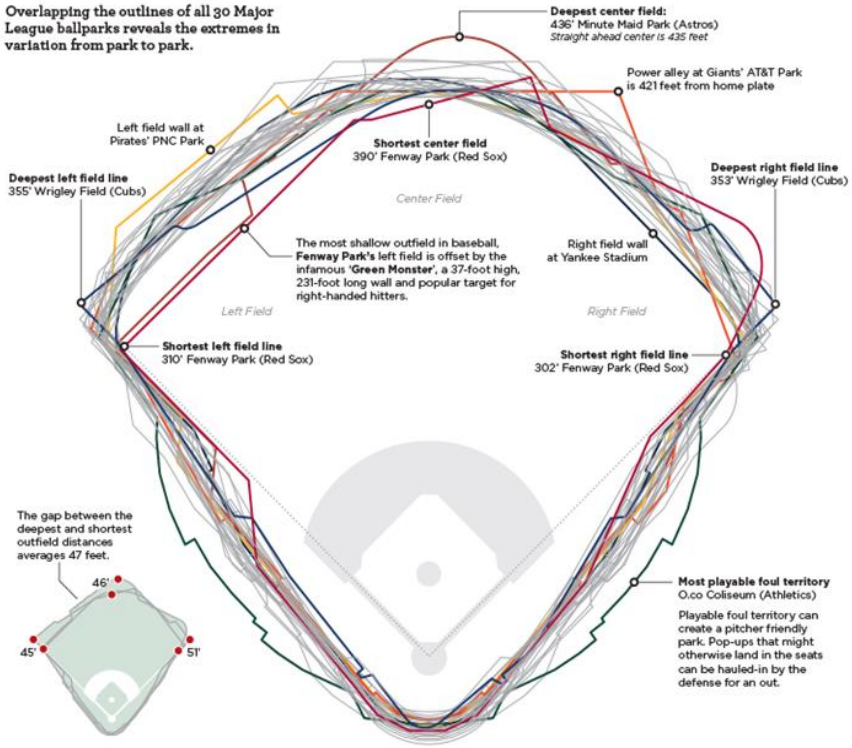
- Relatively new statistic
- The rate of offensive performance at home versus rate on the road (collective)
- A rate higher than 1.0 favors the hitter
- A rate below 1.0 favors the pitcher
- ((homeRS + homeRA)/(homeG)) / ((roadRS + roadRA)/(roadG))*

*ESPN

Baseball's Many Physical Dimensions

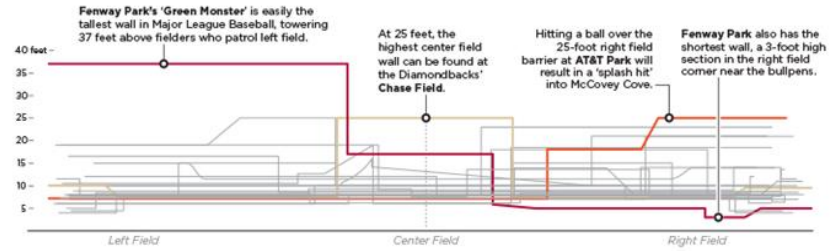
Unlike any other professional sport, baseball's contest is played upon fields that vary in size from park to park. With the exception of the infield diamond, where strict rules regulate the location and height of the pitchers mound and distance between the bases, no two baseball stadiums are alike. From the shape of the field to the distance and height of the outfield walls, the cathedrals of Major League Baseball exhibit unique physical characteristics that distinguishes each from any other.

Overlapping the outlines of all 30 Major League ballparks reveals the extremes in variation from park to park.



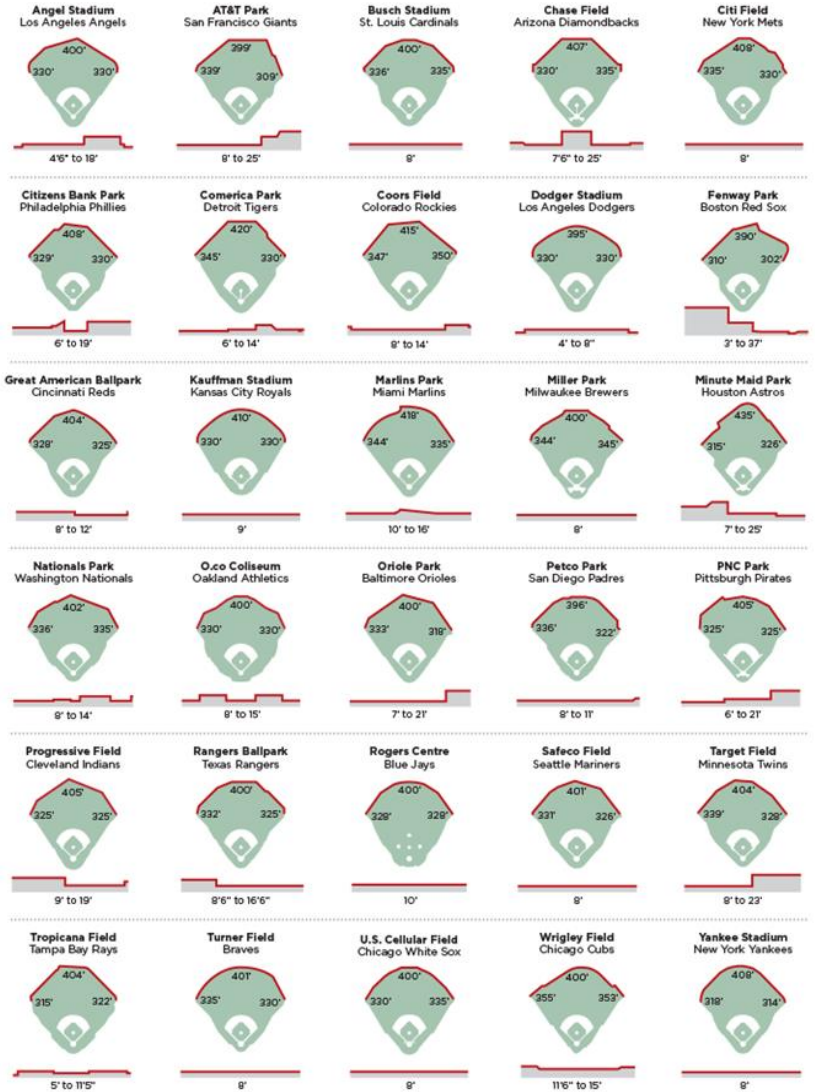
While the typical outfield wall height is about eight to ten feet, they range widely from a scant three feet to a lofty 37 feet as illustrated by overlapping all 30 outfield wall profiles.

Note: Vertical height exaggerated for clarity.



A 'Spotters Guide' for the 30 Major League Baseball venues illustrates the shape and depth of each field in addition to the minimum and maximum height of the outfield wall.

Note: Vertical height exaggerated for clarity.



Sources: Google Maps, MLB teams sites, original research. All information current as of Opening Day 2013.

© 2013 Louis J. Spirito | THIRTY81Project.com

OUR RESEARCH

Is there a significant relationship between park factor and other commonly used baseball statistics such as ERA, OBP, SLG, and OFA for teams?

Data Collected (Part 1)

- Data span (2008-2016)
- Wins
- Losses
- Total Runs Allowed
- R.S.H. (Runs Scored Home)
- R.A.H. (Runs Allowed Home)
- R.S.A. (Runs Scored Away)
- R.A.A (Runs Allowed Away)

Data Collected (Part 2)

- **E.R.A. Ratio (Earned Run Average) = Home ERA/Away ERA**

$$\text{ERA} = 9 \times \frac{\text{Earned Runs Allowed}}{\text{Innings Pitched}}$$

- **O.B.P. (On Base Percentage)**

$$OBP = \frac{H + BB + HBP}{AB + BB + HBP + SF}$$

Data Collected (Part 3)

- SLG (Slugging)

$$\text{SLG} = \frac{(1B) + (2 \times 2B) + (3 \times 3B) + (4 \times HR)}{AB}$$

- OFA (Total Outfield Area)
    - Shape/depth of outfield (varies at each stadium)
    - distance to the wall

DATA (2016)

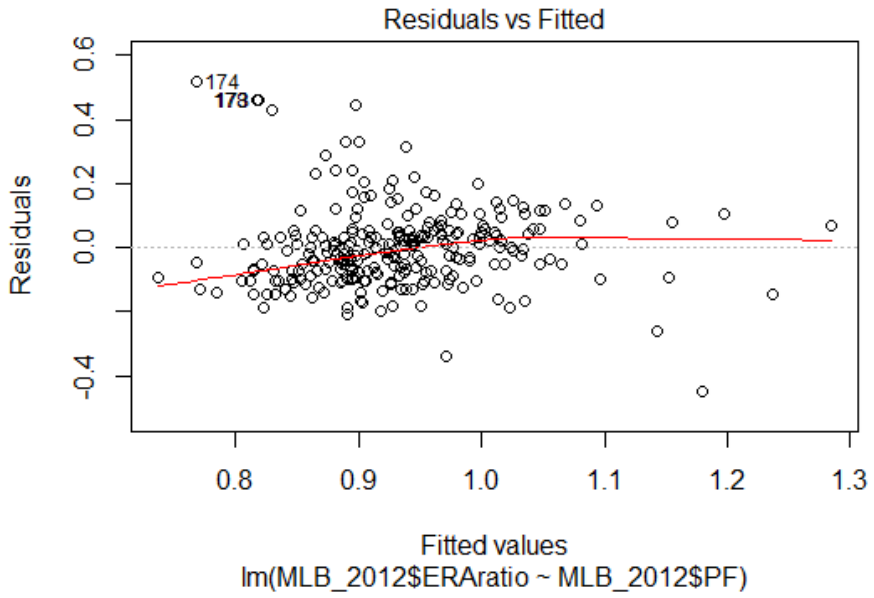
TEAM	Year	W	L	HomeERA	AwayERA	ERA	Total Run	RSH	RSA	RAH	RAA	HomeOBP	AwayOBP	OBP	HomeSLG	AwaySLG	SLG	OF Area	ERA RATIO
Baltimore	2016	89	73	3.81	4.64	4.22	715	376	368	336	379	0.324	0.31	0.317	0.455	0.433	0.443	87.8	0.821121
Boston	2016	93	69	4.3	3.69	4	694	477	401	380	314	0.365	0.332	0.348	0.492	0.43	0.461	83.5	1.165312
LA Angels	2016	74	88	3.97	4.62	4.28	727	337	380	351	376	0.317	0.326	0.322	0.399	0.41	0.405	89.2	0.859307
Chicago S	2016	78	84	3.74	4.47	4.1	715	345	341	329	386	0.323	0.311	0.317	0.418	0.402	0.41	87.8	0.836689
Kansas Cit	2016	81	81	4.27	4.15	4.21	712	377	298	371	341	0.323	0.302	0.312	0.41	0.39	0.4	97.9	1.028916
Milwauke	2016	73	89	3.76	4.43	4.08	733	341	330	351	382	0.325	0.32	0.322	0.423	0.391	0.407	91.1	0.848758
Minnesot	2016	59	103	5.12	5.03	5.08	889	348	374	475	414	0.313	0.32	0.316	0.424	0.419	0.421	90.4	1.017893
NY Yanke	2016	84	78	3.97	4.36	4.16	702	363	317	340	362	0.322	0.307	0.314	0.43	0.382	0.405	87.6	0.91055
Oakland	2016	69	93	4.12	4.92	4.51	761	288	365	353	408	0.291	0.316	0.304	0.367	0.421	0.395	88.4	0.837398
Seattle	2016	86	76	3.81	4.19	4	707	370	398	345	362	0.33	0.323	0.326	0.431	0.429	0.43	87.8	0.909308
Texas	2016	95	67	4.4	4.33	4.37	757	425	340	391	366	0.344	0.299	0.322	0.453	0.414	0.433	92.7	1.016166
Toronto	2016	89	73	4.07	3.47	3.78	666	401	358	363	303	0.341	0.318	0.33	0.444	0.408	0.426	91.8	1.172911
Chicago C	2016	103	58	2.72	3.6	3.15	556	389	419	247	309	0.348	0.339	0.343	0.419	0.437	0.429	89.7	0.755556
Cincinnati	2016	68	94	4.62	5.21	4.91	854	365	351	416	438	0.324	0.308	0.316	0.417	0.399	0.408	87.1	0.886756
Houston	2016	84	78	3.4	4.79	4.06	701	334	390	303	398	0.311	0.326	0.319	0.407	0.426	0.417	88.6	0.709812
LA Dodge	2016	91	71	2.97	4.46	3.7	638	350	375	261	377	0.323	0.315	0.319	0.42	0.399	0.409	91.1	0.665919
Washingt	2016	95	67	3.42	3.62	3.51	612	365	398	307	305	0.325	0.326	0.326	0.423	0.428	0.426	88.8	0.944751
NY Mets	2016	87	75	3.41	3.76	3.58	617	339	332	301	316	0.317	0.315	0.316	0.41	0.422	0.417	91.5	0.906915
Philadelp	2016	71	91	3.92	5.39	4.63	796	279	331	363	433	0.291	0.31	0.301	0.372	0.397	0.385	86.2	0.727273
Pittsburgh	2016	78	83	4.14	4.29	4.21	758	359	370	387	371	0.344	0.322	0.332	0.416	0.389	0.402	90.2	0.965035
St. Louis	2016	86	76	4.08	4.07	4.08	712	355	424	360	352	0.322	0.329	0.325	0.431	0.454	0.443	91.1	1.002457
San Diego	2016	68	94	4.13	4.74	4.43	770	350	336	383	387	0.308	0.29	0.299	0.393	0.387	0.39	90.8	0.871308
San Franci	2016	87	75	3.5	3.82	3.65	631	371	344	306	325	0.347	0.312	0.329	0.415	0.381	0.398	92.2	0.91623
Colorado	2016	75	87	5.4	4.37	4.91	860	508	337	477	383	0.369	0.302	0.336	0.516	0.398	0.457	97.3	1.235698
Arizona	2016	69	93	5.54	4.61	5.09	890	411	341	493	397	0.328	0.313	0.32	0.467	0.397	0.432	94.1	1.201735
Tampa Ba	2016	68	94	3.8	4.62	4.2	713	314	358	338	375	0.301	0.313	0.307	0.413	0.438	0.426	89.6	0.822511
Cleveland	2016	94	67	3.87	3.81	3.84	676	452	325	347	329	0.359	0.3	0.329	0.469	0.391	0.43	85.6	1.015748
Detroit	2016	86	75	4.17	4.3	4.24	721	381	369	357	364	0.338	0.324	0.331	0.449	0.428	0.438	95.8	0.969767
Atlanta	2016	68	93	4.39	4.63	4.51	779	335	314	404	375	0.327	0.314	0.321	0.377	0.392	0.384	94.1	0.948164
Miami	2016	79	82	3.53	4.6	4.05	682	302	353	302	380	0.31	0.333	0.322	0.379	0.408	0.394	93.4	0.767391

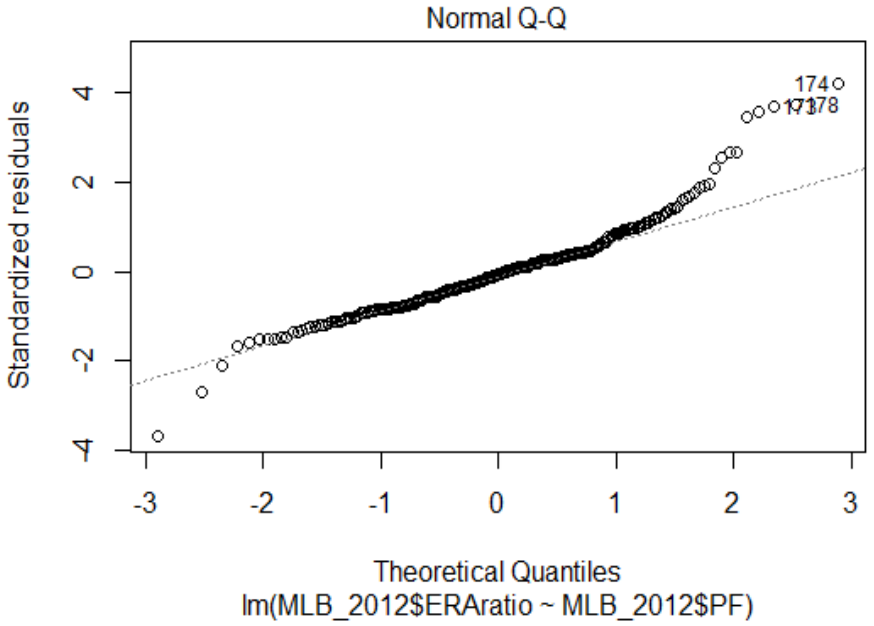
CLEANING

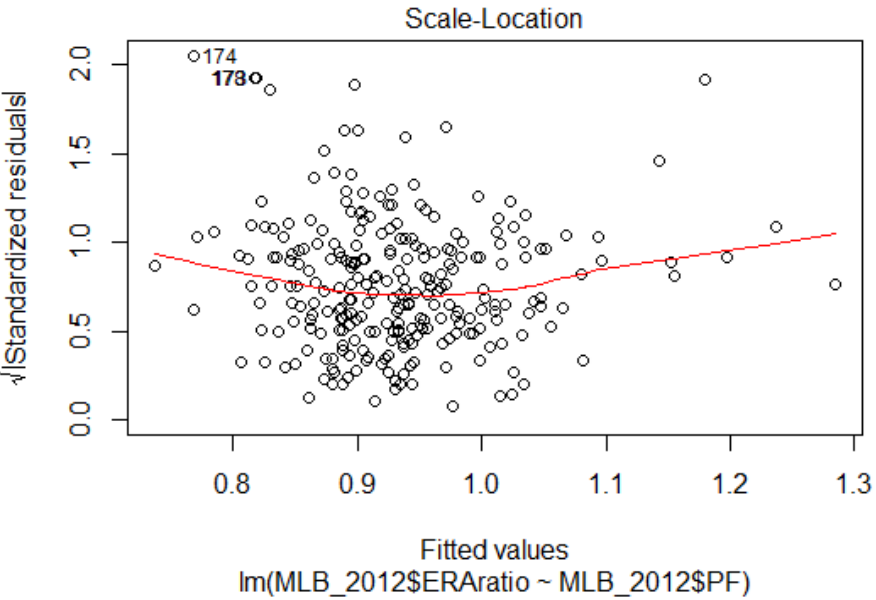
- Criteria of data:
    - 2008 – 2016
    - Old stadium's data removed if new stadium was built during time period

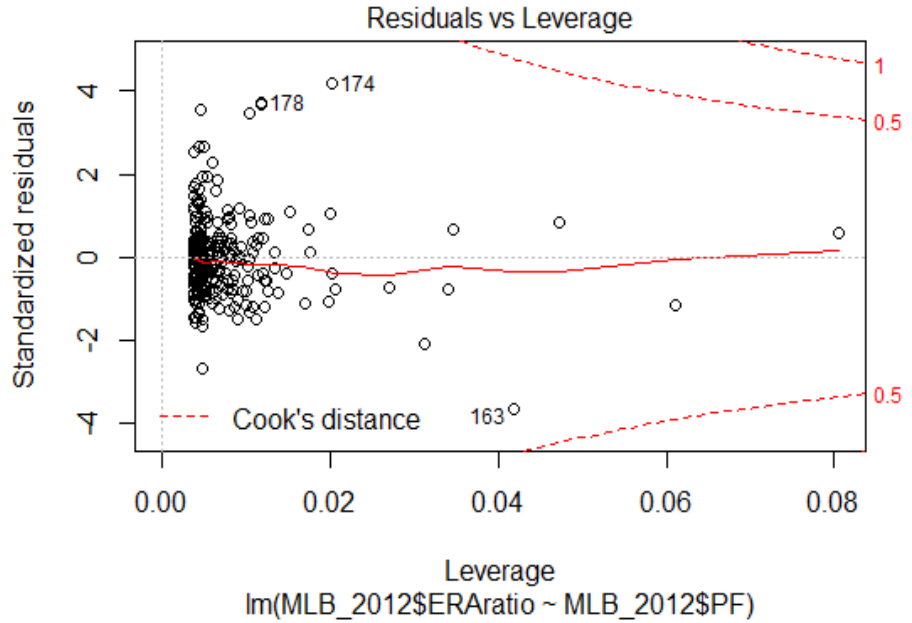
PARK FACTOR VS ERA

```
Call:
lm(formula = MLB_2012$ERAratio ~ MLB_2012$PF)

Residuals:
     Min       1Q   Median       3Q
-0.45013 -0.07904 -0.00742  0.05178
     Max
 0.52031

Coefficients:
             Estimate Std. Error
(Intercept)  0.31611    0.06087
MLB_2012$PF  0.61328    0.06017
             t value Pr(>|t|)
(Intercept)    5.193 4.18e-07 ***
MLB_2012$PF   10.192  < 2e-16 ***
---
Signif. codes:
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05
  ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1251 on 260 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.2855,    Adjusted R-squared:  0.2827
F-statistic: 103.9 on 1 and 260 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
Fitted values
lm(MLB_2012$ERAratio ~ MLB_2012$PF)
174
178


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(MLB_2012$ERAratio ~ MLB_2012$PF)
174
178

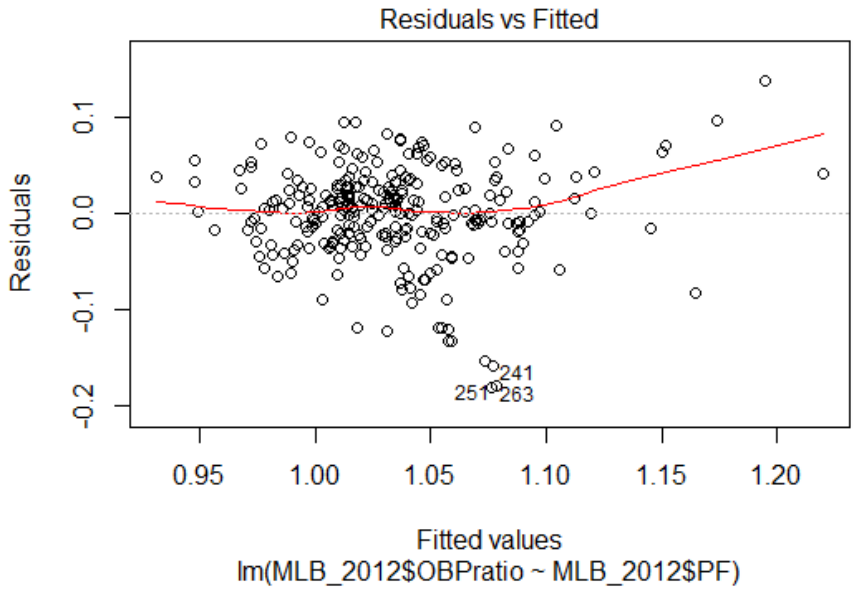

Scale-Location
√|Standardized residuals|
Fitted values
lm(MLB_2012$ERAratio ~ MLB_2012$PF)
174
178

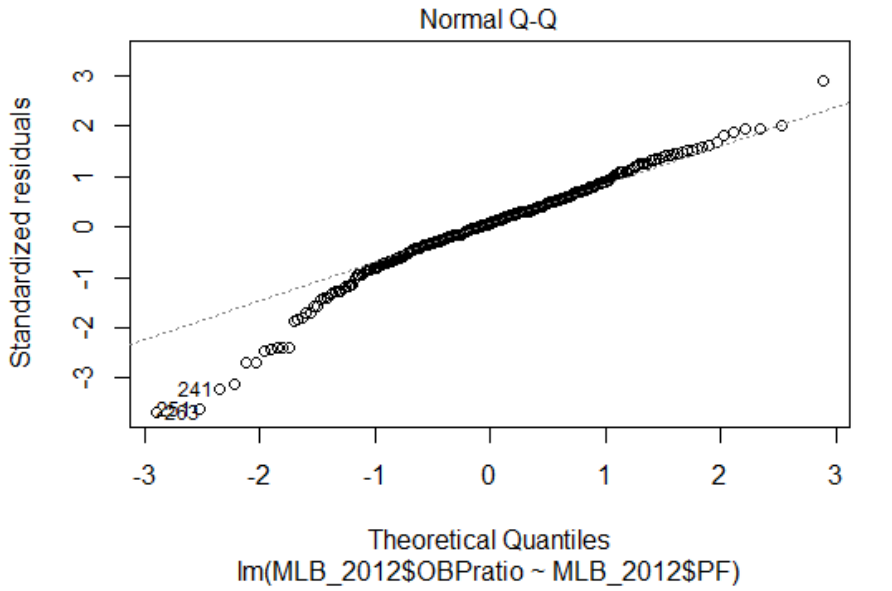

Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(MLB_2012$ERAratio ~ MLB_2012$PF)
174
178
163

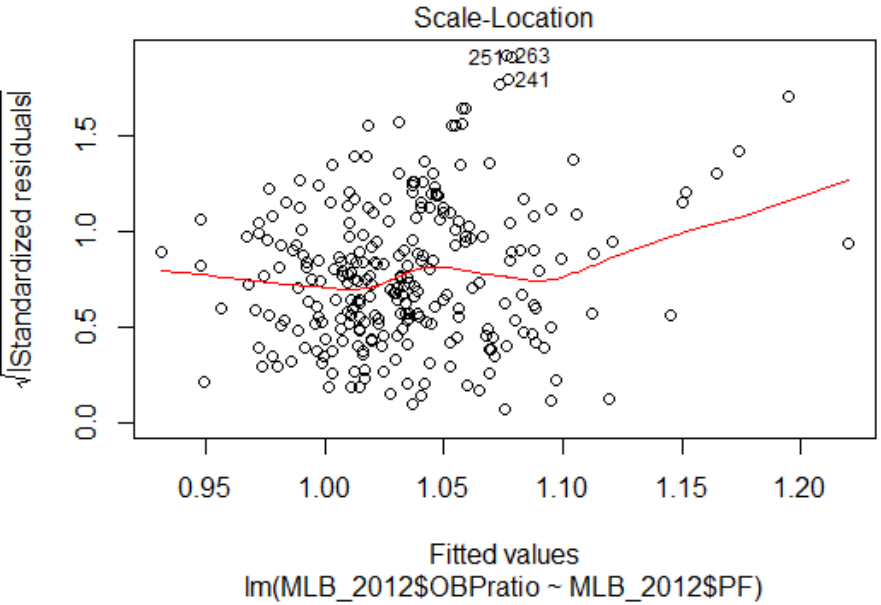
```
Call:
lm(formula = MLB_2012$ERAratio ~ MLB_2012$PF)

Coefficients:
(Intercept)  MLB_2012$PF
     0.3161       0.6133
```

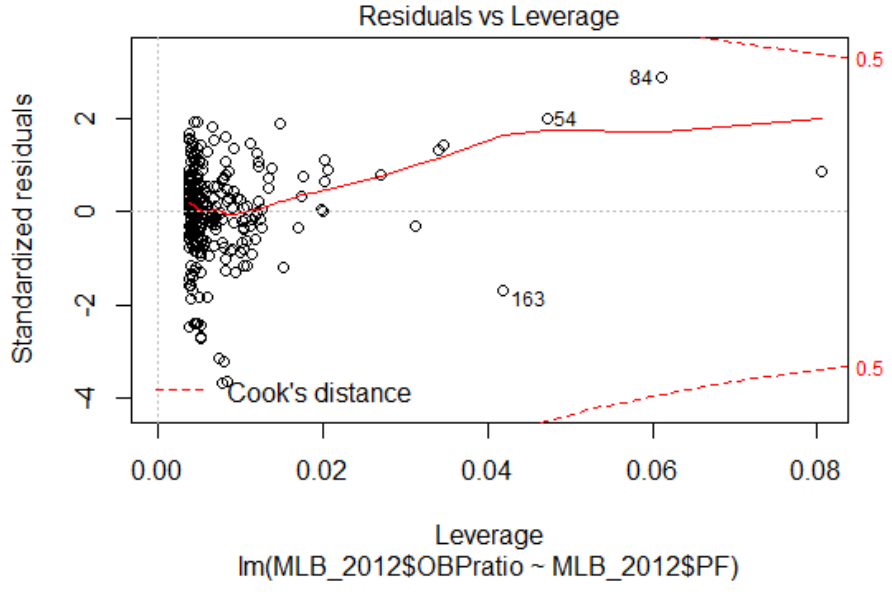
PARK FACTOR VS OBP

```
Call:
lm(formula = MLB_2012$OBPratio ~ MLB_2012$PF)

Residuals:
      Min        1Q    Median        3Q       Max
-0.180655 -0.022281  0.003427  0.028748  0.138734

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.70962    0.02399   29.59   <2e-16 ***
MLB_2012$PF  0.32315    0.02371   13.63   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04928 on 260 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.4167,    Adjusted R-squared:  0.4144
F-statistic: 185.7 on 1 and 260 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
0.1
0.0
-0.1
-0.2
241
251
263
0.95
1.00
1.05
1.10
1.15
1.20
Fitted values
lm(MLB_2012$OBPratio ~ MLB_2012$PF)


Normal Q-Q
Standardized residuals
3
2
1
0
-1
-2
-3
241
251
263
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(MLB_2012$OBPratio ~ MLB_2012$PF)

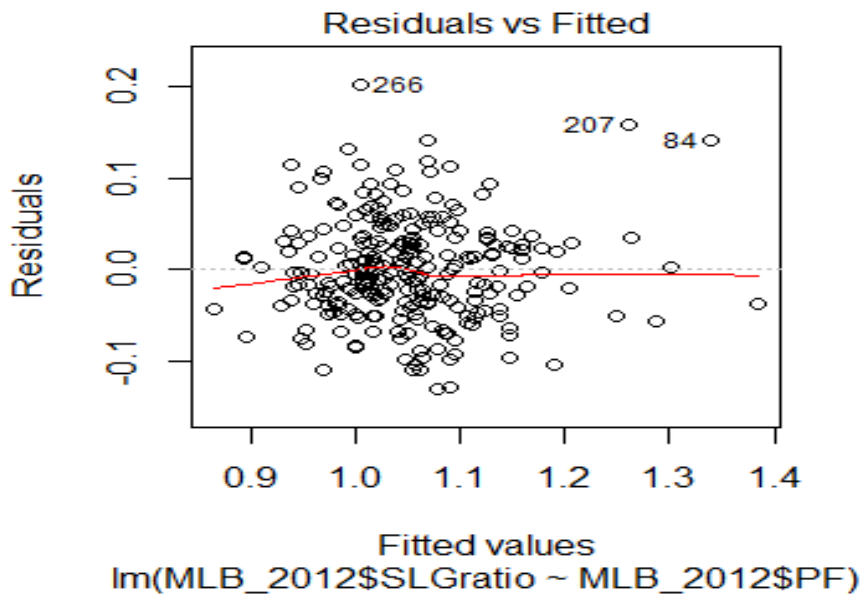

Scale-Location
√|Standardized residuals|
251
263
241
1.5
1.0
0.5
0.0
0.95
1.00
1.05
1.10
1.15
1.20
Fitted values
lm(MLB_2012$OBPratio ~ MLB_2012$PF)

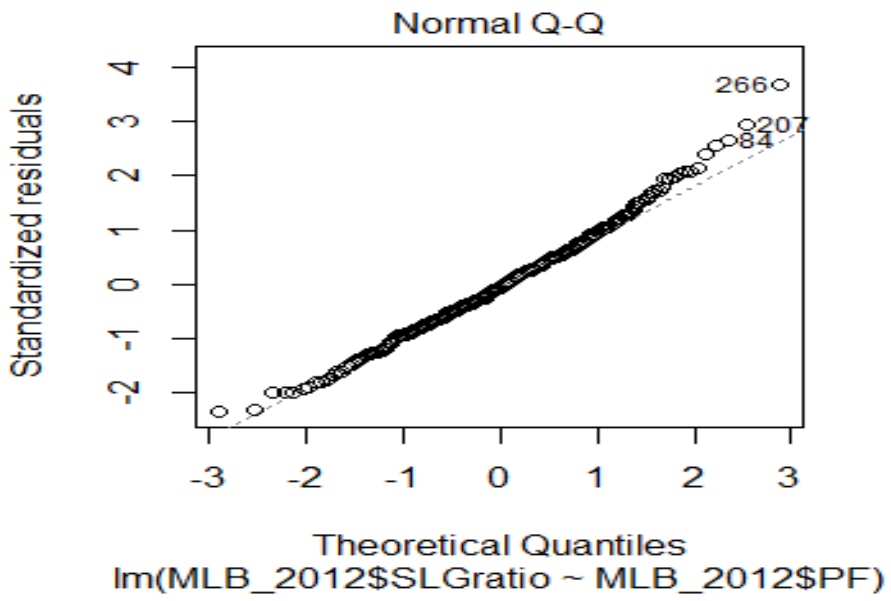

Residuals vs Leverage
Standardized residuals
2
0
-2
-4
84
54
163
Cook's distance
0.5
0.5
0.00
0.02
0.04
0.06
0.08
Leverage
lm(MLB_2012$OBPratio ~ MLB_2012$PF)

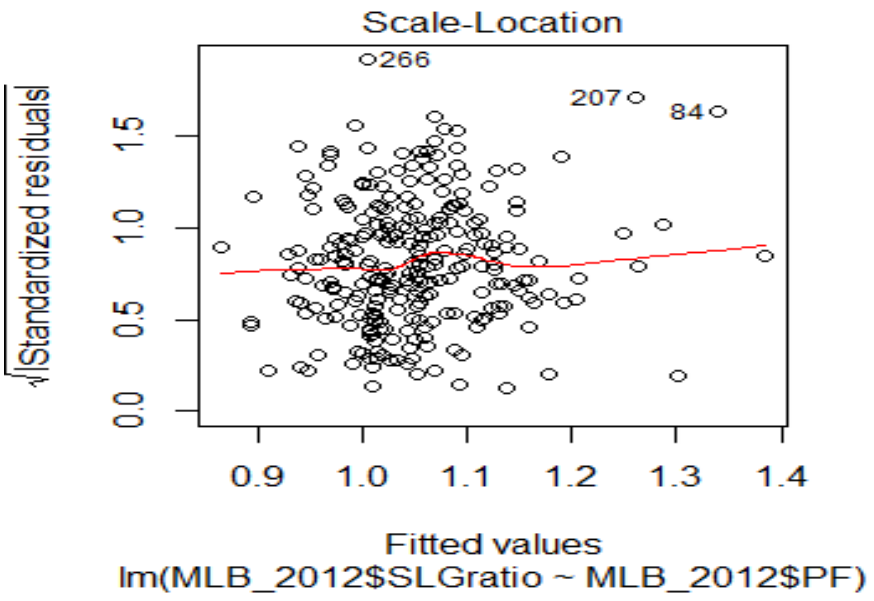
```
Call:
lm(formula = MLB_2012$OBPratio ~ MLB_2012$PF)

Coefficients:
(Intercept)  MLB_2012$PF
     0.7096       0.3231
```

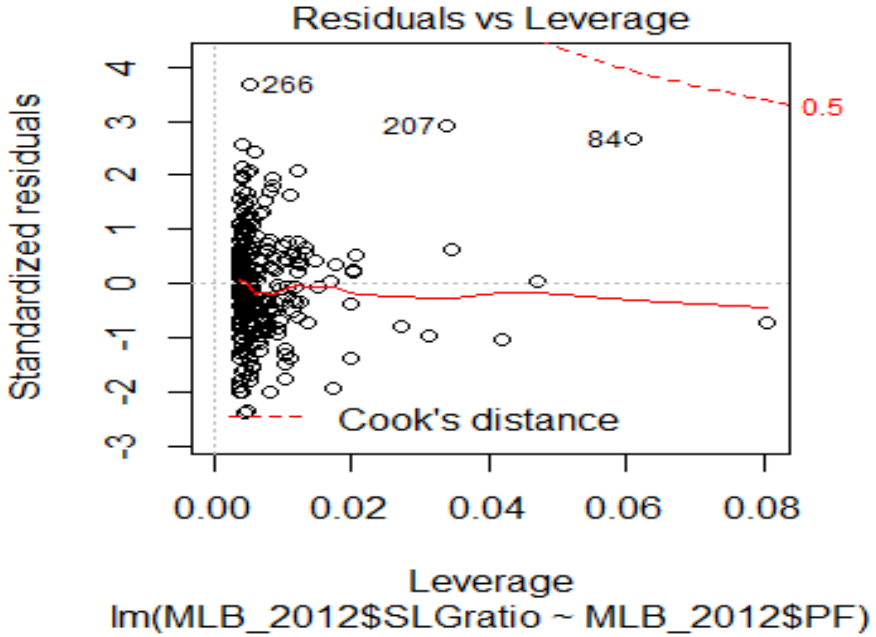
PARK FACTOR VS SLG

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.129931 -0.036910 -0.003724  0.031860  0.201797

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.46374    0.02687   17.26   <2e-16 ***
MLB_2012$PF  0.58366    0.02656   21.98   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05519 on 260 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.6501,    Adjusted R-squared:  0.6487
F-statistic:   483 on 1 and 260 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
Fitted values
lm(MLB_2012$SLGratio ~ MLB_2012$PF)
266
207
84


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(MLB_2012$SLGratio ~ MLB_2012$PF)
266
207
84

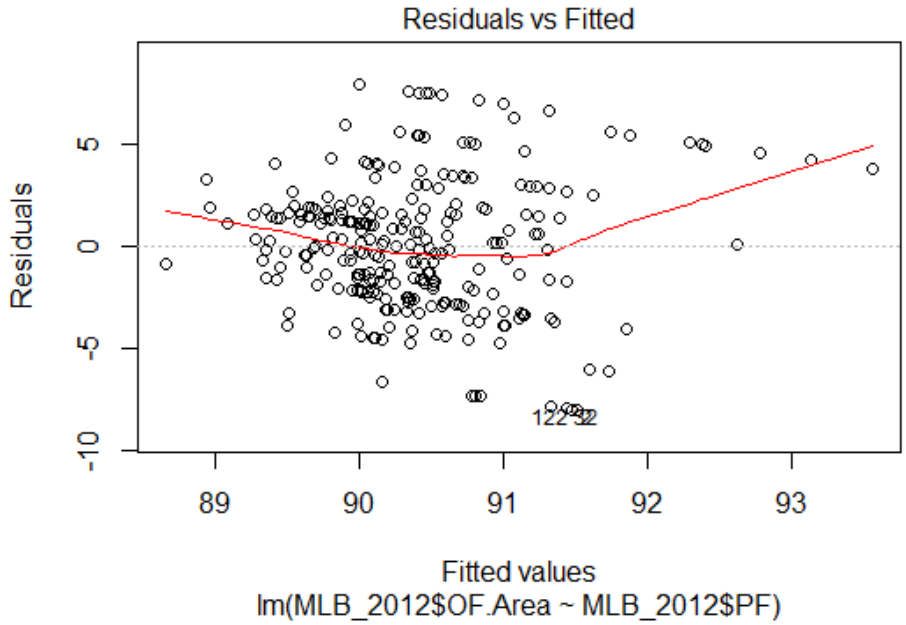

Scale-Location
√|Standardized residuals|
Fitted values
lm(MLB_2012$SLGratio ~ MLB_2012$PF)
266
207
84

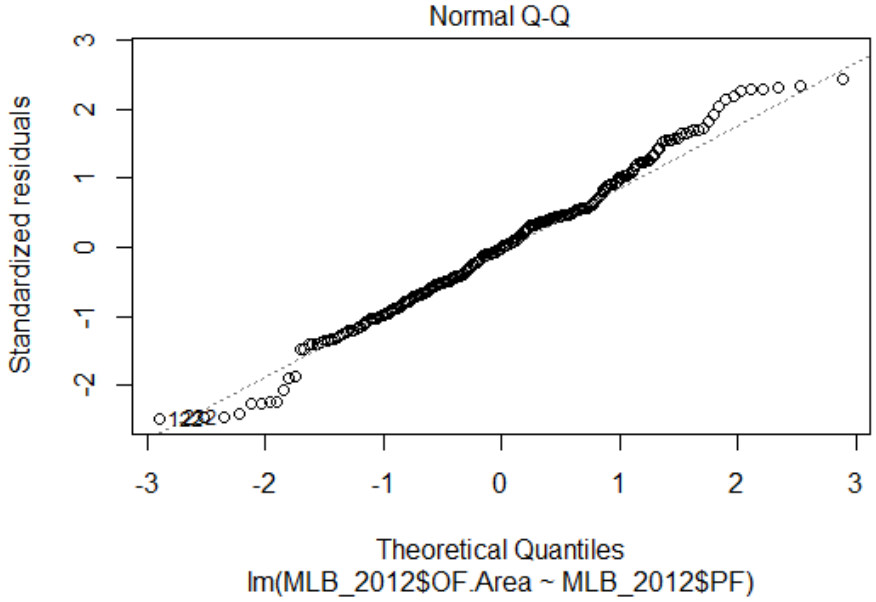

Residuals vs Leverage
Standardized residuals
Cook's distance
0.5
Leverage
lm(MLB_2012$SLGratio ~ MLB_2012$PF)
266
207
84

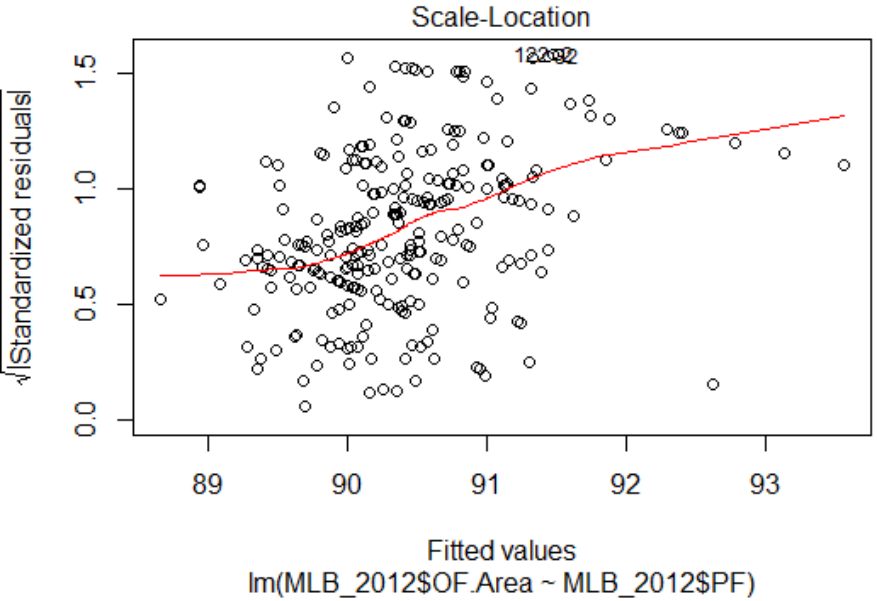
```
Call:
lm(formula = MLB_2012$SLGratio ~ MLB_2012$PF)

Coefficients:
(Intercept)  MLB_2012$PF
     0.4637       0.5837
```

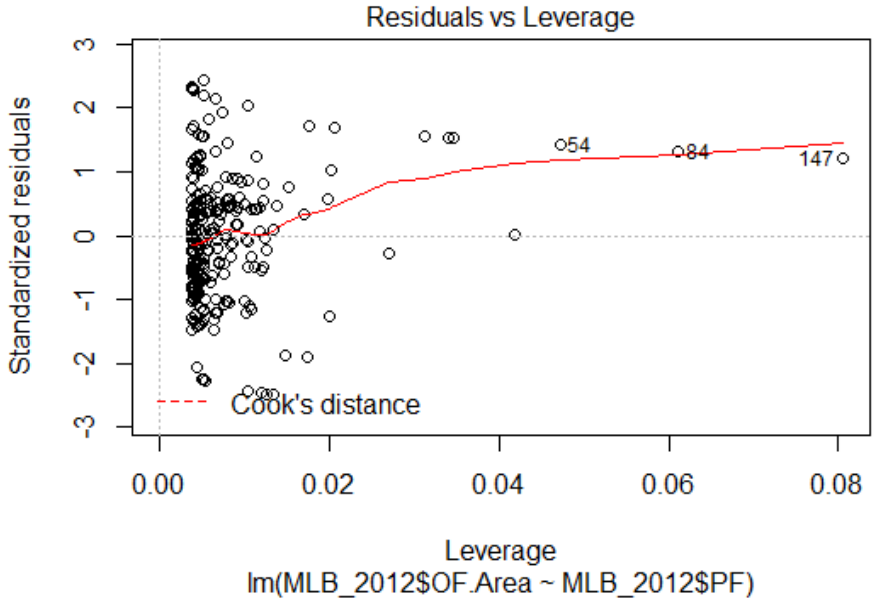
PARK FACTOR VS OUTFIELD

```
Call:
lm(formula = MLB_2012$OF.Area ~ MLB_2012$PF)

Residuals:
    Min      1Q  Median      3Q     Max
-8.0152 -2.1876 -0.0245  1.7948  7.9052

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   84.896      1.576  53.855  < 2e-16 ***
MLB_2012$PF    5.487      1.558   3.521 0.000507 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.239 on 260 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.04552,   Adjusted R-squared:  0.04185
F-statistic:  12.4 on 1 and 260 DF,  p-value: 0.000507
```


Residuals vs Fitted
Residuals
Fitted values
lm(MLB_2012$OF.Area ~ MLB_2012$PF)


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(MLB_2012$OF.Area ~ MLB_2012$PF)

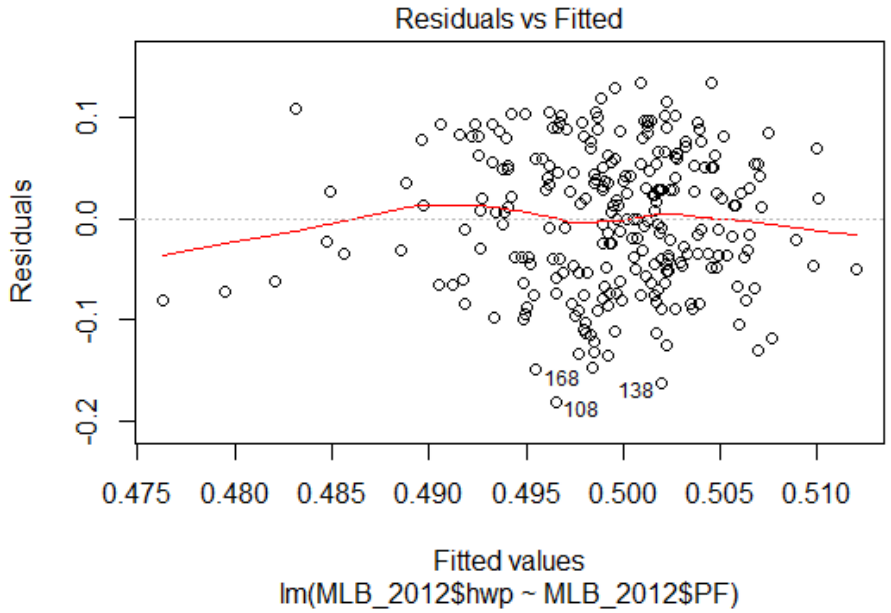

Scale-Location
√|Standardized residuals|
Fitted values
lm(MLB_2012$OF.Area ~ MLB_2012$PF)

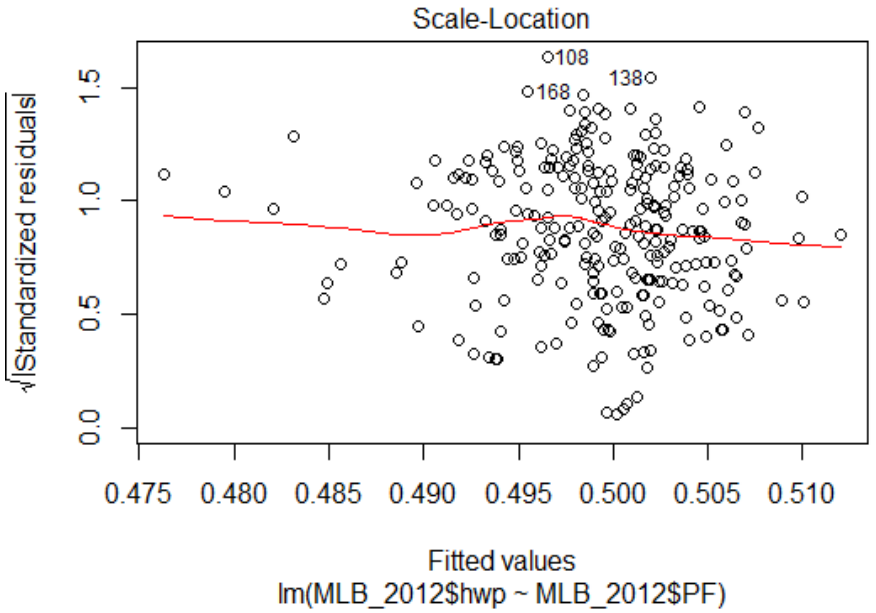

Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(MLB_2012$OF.Area ~ MLB_2012$PF)

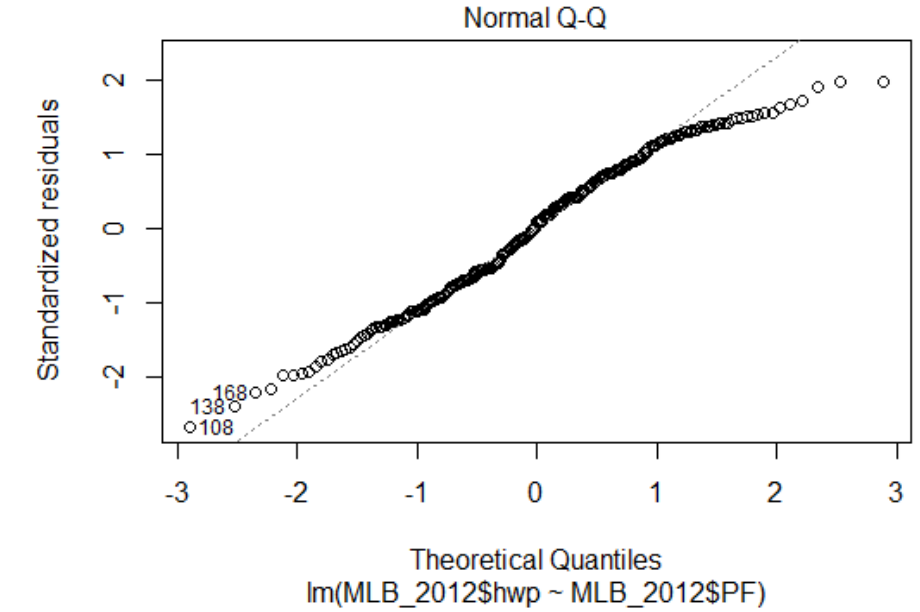
```
Call:
lm(formula = MLB_2012$OF.Area ~ MLB_2012$PF)

Coefficients:
(Intercept)  MLB_2012$PF
     84.896        5.487
```

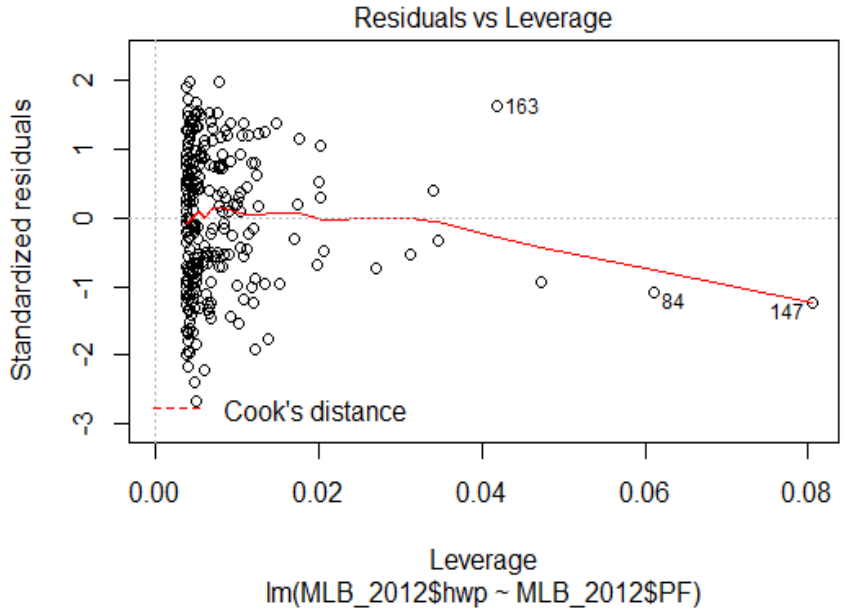
PARK FACTOR VS HWP

```
Call:
lm(formula = MLB_2012$hwp ~ MLB_2012$PF)

Residuals:
      Min        1Q    Median        3Q       Max
-0.181709 -0.051207  0.006351  0.054466  0.135215

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.53948    0.03319  16.254   <2e-16 ***
MLB_2012$PF -0.04000    0.03281  -1.219    0.224
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06818 on 260 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.005684,  Adjusted R-squared:  0.00186
F-statistic: 1.486 on 1 and 260 DF,  p-value: 0.2239
```


Residuals vs Fitted
Residuals
0.1
0.0
-0.1
-0.2
168
108
138
0.475
0.480
0.485
0.490
0.495
0.500
0.505
0.510
Fitted values
lm(MLB_2012$hwp ~ MLB_2012$PF)


Scale-Location
√|Standardized residuals|
1.5
1.0
0.5
0.0
108
138
168
0.475
0.480
0.485
0.490
0.495
0.500
0.505
0.510
Fitted values
lm(MLB_2012$hwp ~ MLB_2012$PF)

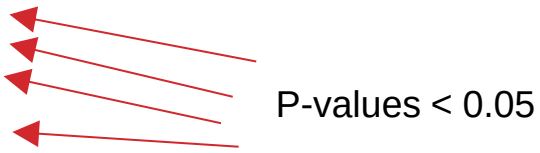

Normal Q-Q
Standardized residuals
2
1
0
-1
-2
168
138
108
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(MLB_2012$hwp ~ MLB_2012$PF)


Residuals vs Leverage
Standardized residuals
2
1
0
-1
-2
-3
163
84
147
Cook's distance
0.00
0.02
0.04
0.06
0.08
Leverage
lm(MLB_2012$hwp ~ MLB_2012$PF)

```
Call:
lm(formula = MLB_2012$hwp ~ MLB_2012$PF)

Coefficients:
(Intercept)  MLB_2012$PF
     0.5395      -0.0400
```

Multiple Linear Regression

```
> fullmultiplemlb <- lm(MLB_2012$SLGratio ~ MLB_2012$PF + MLB_2012$OBP + MLB_2012$OF.Area + MLB_2012$RSH + MLB_2012$hwp)
> summary(multiplemlb)

Call:
lm(formula = MLB_2012$SLGratio ~ MLB_2012$PF + MLB_2012$OBP +
    MLB_2012$OF.Area + MLB_2012$RSH + MLB_2012$hwp)

Residuals:
      Min        1Q    Median        3Q
-0.106981 -0.033414 -0.005505  0.030260
      Max
 0.210265

Coefficients:
                  Estimate Std. Error t value
(Intercept)       0.9323567  0.1281924   7.273
MLB_2012$PF       0.2787454  0.0420628   6.627
MLB_2012$OBP     -2.1187501  0.3934705  -5.385
MLB_2012$OF.Area  0.0013164  0.0009385   1.403
MLB_2012$RSH      0.0012220  0.0001392   8.780
MLB_2012$hwp     -0.0754536  0.0529638  -1.425
                 Pr(>|t|)
(Intercept)      4.28e-12 ***
MLB_2012$PF      2.01e-10 ***
MLB_2012$OBP     1.64e-07 ***
MLB_2012$OF.Area    0.162
MLB_2012$RSH     2.38e-16 ***
MLB_2012$hwp        0.155
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04834 on 256 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.7357,    Adjusted R-squared:  0.7306
F-statistic: 142.6 on 5 and 256 DF,  p-value: < 2.2e-16
```



P-values < 0.05

Multiple Linear Regression

```
> reducedmultiplemlb <- lm(MLB_2012$SLGratio ~ MLB_2012$PF + MLB_2012$OBP + MLB_2012$RSH)
> summary(reducedmultiplemlb)

Call:
lm(formula = MLB_2012$SLGratio ~ MLB_2012$PF + MLB_2012$OBP +
    MLB_2012$RSH)

Residuals:
      Min        1Q    Median        3Q       Max
-0.111936 -0.031494 -0.006639  0.033128  0.209997

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.0102693  0.1102399   9.164  < 2e-16 ***
MLB_2012$PF   0.3129469  0.0385451   8.119 1.94e-14 ***
MLB_2012$OBP -2.1204667  0.3874187  -5.473 1.05e-07 ***
MLB_2012$RSH  0.0011373  0.0001328   8.567 9.82e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04853 on 258 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.7316,    Adjusted R-squared:  0.7284
F-statistic: 234.4 on 3 and 258 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression

Full Model

```
> anova(reducedmultiplemlb,fullmultiplemlb)
```

```
Residual standard error: 0.04834 on 256 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.7357,    Adjusted R-squared:  0.7306
F-statistic: 142.6 on 5 and 256 DF,  p-value: < 2.2e-16
```

Reduced Model

```
2$OBP + MLB_2012$OF.Area +
    MLB_2012$RSH + MLB_2012$hwp
```

```
Residual standard error: 0.04853 on 258 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.7316,    Adjusted R-squared:  0.7284
F-statistic: 234.4 on 3 and 258 DF,  p-value: < 2.2e-16
```

Park Factor vs ERA (2015-2016)

```
Residual standard error: 0.05914 on 58 degrees of freedom
  (939 observations deleted due to missingness)
Multiple R-squared:  0.8104,    Adjusted R-squared:  0.8071
F-statistic: 247.8 on 1 and 58 DF,  p-value: < 2.2e-16
```

```
Residual standard error: 0.1251 on 260 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.2855,    Adjusted R-squared:  0.2827
F-statistic: 103.9 on 1 and 260 DF,  p-value: < 2.2e-16
```

RESULTS SUMMARY

Significant: SLG

Nonsignificant: OBP, ERA, Outfield, HWP

- **We learned that Park Factor only has a significant relationship with slugging and insignificant relationships with the other predictors.**

Limitations

- **random variation of the performances of the pitchers and hitters**
- **other uncontrolled factors influence relationship:**
  - climate/weather and altitude
  - skewed right or left hand power hitters
  - injuries lead to slugger playing more games
  - ball characteristics
    - dry storage vs humid storage
- **Differences in National and American League**