Modeling

Group 6

```r
# Get couldabeens
couldabeens <- read_csv("../data-gen/couldabeens.csv")
# Get payroll revenue data
payroll_rev <- read_csv("../data/revenue-payroll.csv")
payroll <- find_labShare(payroll_rev)
```

```
## [1] -0.4509946
```

```
## [1] 0.9913644
```

```
## [1] -0.3523251
```

Various Subsets

Subset 1

```
##   Year       prop totRev  labShare
## 1 2004 0.08163265   4269 0.4851876
## 2 2005 0.15909091   4733 0.4631073
## 3 2006 0.12000000   5111 0.4552351
## 4 2007 0.16923077   5489 0.4512095
## 5 2008 0.07547170   5819 0.4616658
## 6 2009 0.14583333   5898 0.4502196
```

```
## 
## Call:
## lm(formula = prop ~ . - Year, data = dataset1)
## 
## Residuals:
##       Min        1Q    Median        3Q       Max 
## -0.067477 -0.026412 -0.001198  0.021475  0.056960 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.929e-02  6.090e-01   0.081    0.937
## totRev      7.382e-06  1.258e-05   0.587    0.568
## labShare    4.977e-02  1.197e+00   0.042    0.968
## 
## Residual standard error: 0.03959 on 12 degrees of freedom
## Multiple R-squared:  0.09891,    Adjusted R-squared:  -0.05127 
## F-statistic: 0.6586 on 2 and 12 DF,  p-value: 0.5353
```

Subset 2

```
##   Year      prop totRev  labShare postMoneyball
## 1 1990 0.1372549 1346.1 0.3375848             0
```

```
## 2 1991 0.3023256 1459.1 0.4317788            0
## 3 1992 0.2000000 1583.6 0.4950075            0
## 4 1993 0.1590909 1774.5 0.4827547            0
## 5 1994 0.2105263 1687.0 0.5245621            0
## 6 1995 0.2765957 1410.5 0.6253683            0
```

```
##
## Call:
## lm(formula = prop ~ . - Year, data = dataset2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.064602 -0.029048  0.001389  0.026921  0.109327
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.498e-02  8.694e-02   0.747  0.46178
## totRev         6.001e-06  5.818e-06   1.031  0.31223
## labShare       2.762e-01  1.690e-01   1.634  0.11480
## postMoneyball -1.070e-01  3.158e-02  -3.387  0.00234 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04269 on 25 degrees of freedom
## Multiple R-squared:  0.6152, Adjusted R-squared:  0.569
## F-statistic: 13.32 on 3 and 25 DF,  p-value: 2.156e-05
```

Subset 3

```
##   Year       prop totRev totPayroll
## 1 2004 0.08163265   4269   2071.266
## 2 2005 0.15909091   4733   2191.887
## 3 2006 0.12000000   5111   2326.707
## 4 2007 0.16923077   5489   2476.689
## 5 2008 0.07547170   5819   2686.433
## 6 2009 0.14583333   5898   2655.395
```

```
##
## Call:
## lm(formula = prop ~ . - Year, data = dataset3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07104 -0.02383 -0.00009  0.02391  0.05541
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.479e-02  8.553e-02   1.108    0.289
## totRev       2.485e-05  6.614e-05   0.376    0.714
## totPayroll  -4.740e-05  1.741e-04  -0.272    0.790
##
## Residual standard error: 0.03947 on 12 degrees of freedom
## Multiple R-squared:  0.1043, Adjusted R-squared:  -0.04497
## F-statistic: 0.6988 on 2 and 12 DF,  p-value: 0.5163
```

Subset 4

```
##   Year      prop totRev totPayroll postMoneyball
## 1 1990 0.1372549 1346.1   454.4229             0
## 2 1991 0.3023256 1459.1   630.0085             0
## 3 1992 0.2000000 1583.6   783.8939             0
## 4 1993 0.1590909 1774.5   856.6482             0
## 5 1994 0.2105263 1687.0   884.9363             0
## 6 1995 0.2765957 1410.5   882.0820             0
```

```
##
## Call:
## lm(formula = prop ~ . - Year, data = dataset4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06605 -0.03038  0.00185  0.02670  0.09676
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.967e-01  2.252e-02   8.734 4.57e-09 ***
## totRev        -8.722e-06  2.568e-05  -0.340  0.73696
## totPayroll     3.423e-05  5.929e-05   0.577  0.56889
## postMoneyball -1.184e-01  3.208e-02  -3.689  0.00109 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04462 on 25 degrees of freedom
## Multiple R-squared:  0.5797, Adjusted R-squared:  0.5292
## F-statistic: 11.49 on 3 and 25 DF,  p-value: 6.327e-05
```

Subset 5

```
##   Year       prop totRev totPayroll
## 1 2004 0.08163265   4269   2071.266
## 2 2005 0.15909091   4733   2191.887
## 3 2006 0.12000000   5111   2326.707
## 4 2007 0.16923077   5489   2476.689
## 5 2008 0.07547170   5819   2686.433
## 6 2009 0.14583333   5898   2655.395
```

```
##
## Call:
## lm(formula = prop ~ . - Year, data = dataset5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07104 -0.02383 -0.00009  0.02391  0.05541
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.479e-02  8.553e-02   1.108    0.289
## totRev       2.485e-05  6.614e-05   0.376    0.714
## totPayroll  -4.740e-05  1.741e-04  -0.272    0.790
##
## Residual standard error: 0.03947 on 12 degrees of freedom
```

```
## Multiple R-squared:  0.1043, Adjusted R-squared:  -0.04497
## F-statistic: 0.6988 on 2 and 12 DF,  p-value: 0.5163
```

Subset 7

```
##   Year       prop totPayroll
## 1 2004 0.08163265   2071.266
## 2 2005 0.15909091   2191.887
## 3 2006 0.12000000   2326.707
## 4 2007 0.16923077   2476.689
## 5 2008 0.07547170   2686.433
## 6 2009 0.14583333   2655.395
```

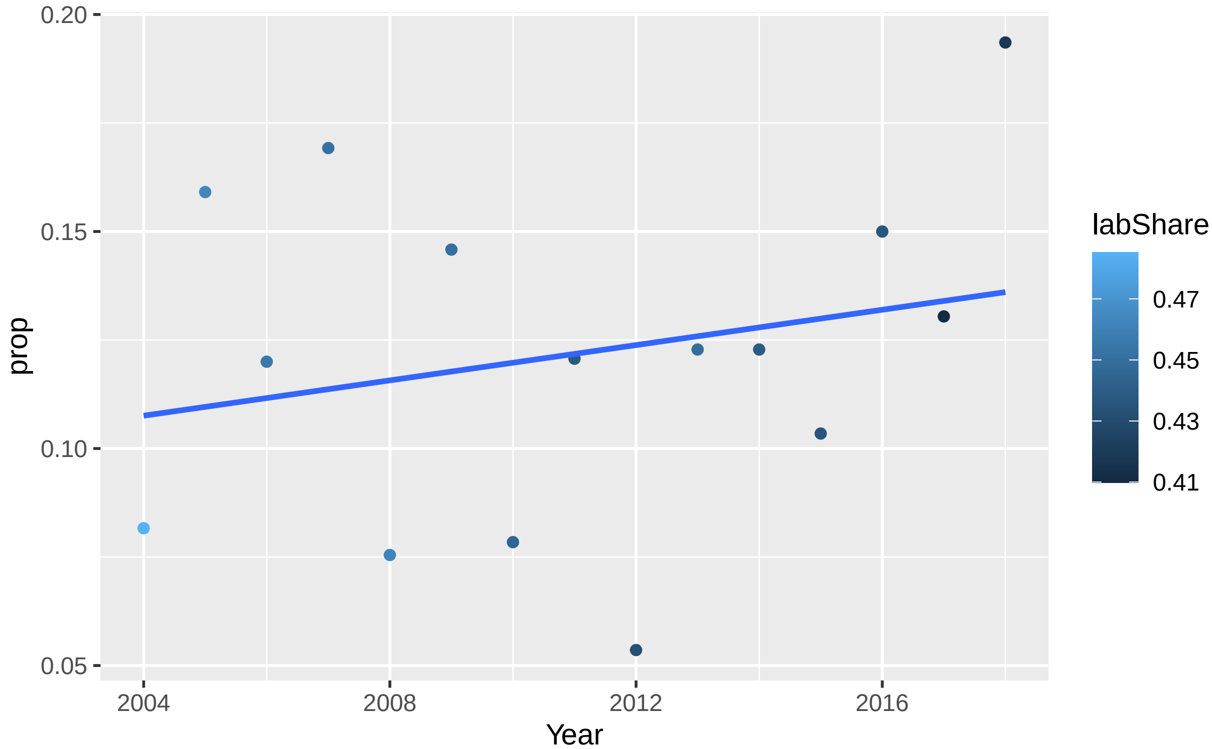
```
##
## Call:
## lm(formula = prop ~ . - Year, data = dataset7)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.067029 -0.026713 -0.002178  0.020995  0.056869
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.838e-02  4.710e-02   1.452    0.170
## totPayroll  1.776e-05  1.531e-05   1.160    0.267
##
## Residual standard error: 0.03815 on 13 degrees of freedom
## Multiple R-squared:  0.09378,    Adjusted R-squared:  0.02407
## F-statistic: 1.345 on 1 and 13 DF,  p-value: 0.267
```

Subset 8

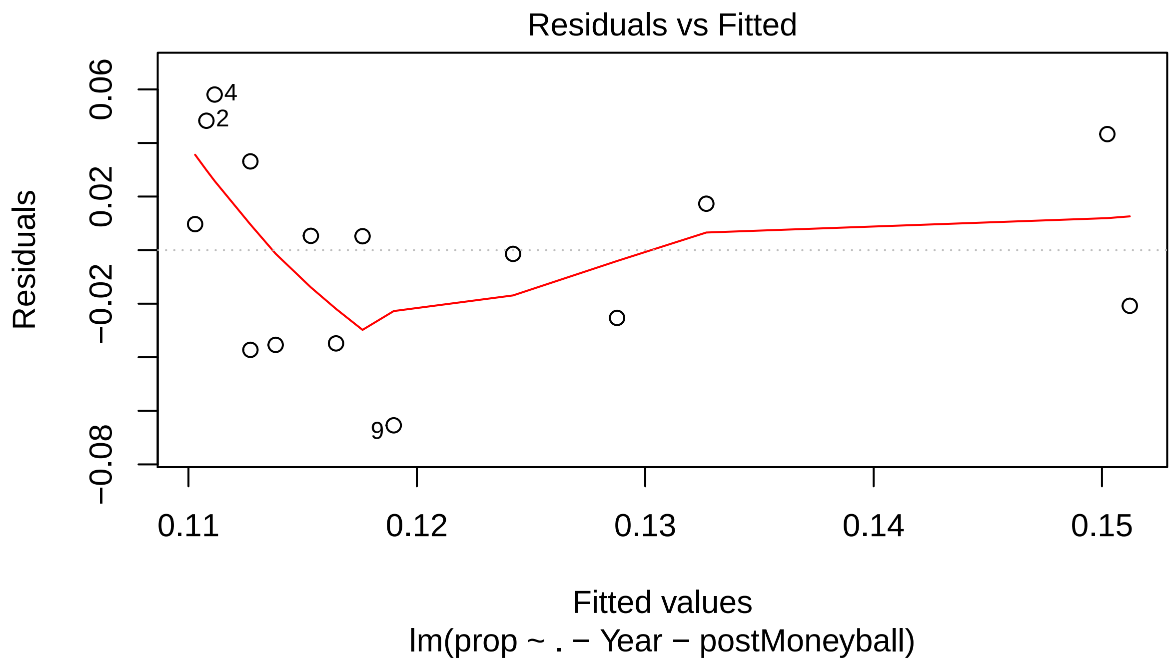
```
##   Year      prop totPayroll postMoneyball
## 1 1990 0.1372549   454.4229             0
## 2 1991 0.3023256   630.0085             0
## 3 1992 0.2000000   783.8939             0
## 4 1993 0.1590909   856.6482             0
## 5 1994 0.2105263   884.9363             0
## 6 1995 0.2765957   882.0820             0
```

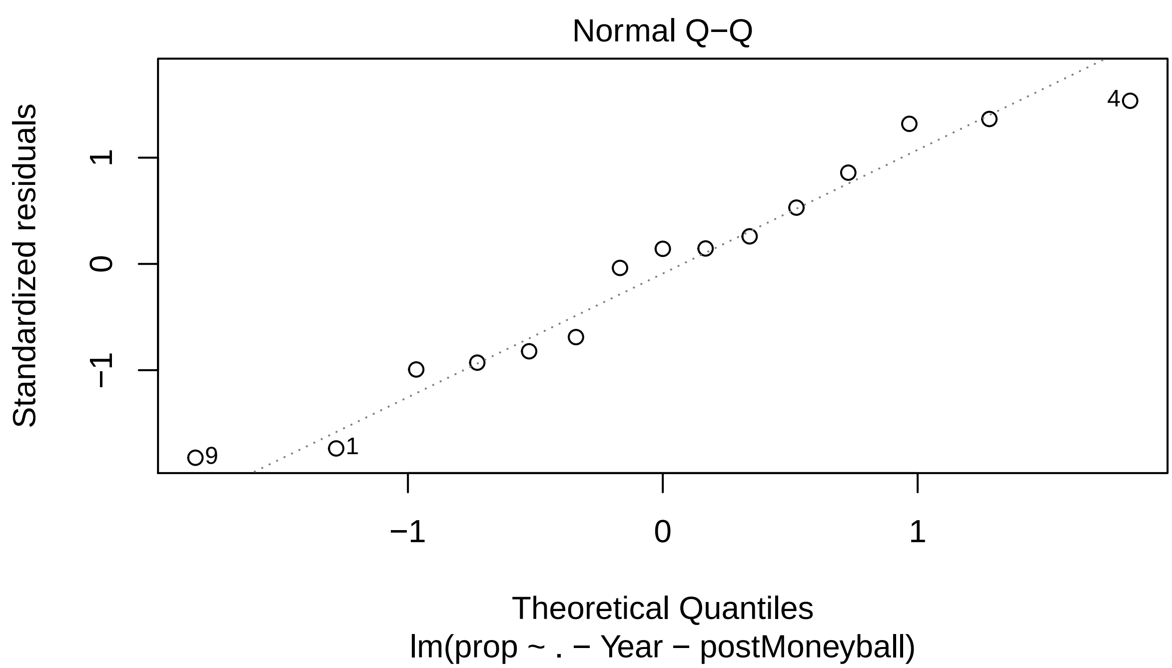
```
##
## Call:
## lm(formula = prop ~ . - Year, data = dataset8)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.069135 -0.027884 -0.001623  0.029206  0.093361
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.997e-01  2.036e-02   9.808 3.17e-10 ***
## totPayroll     1.466e-05  1.379e-05   1.063 0.297428
## postMoneyball -1.220e-01  2.970e-02  -4.109 0.000352 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

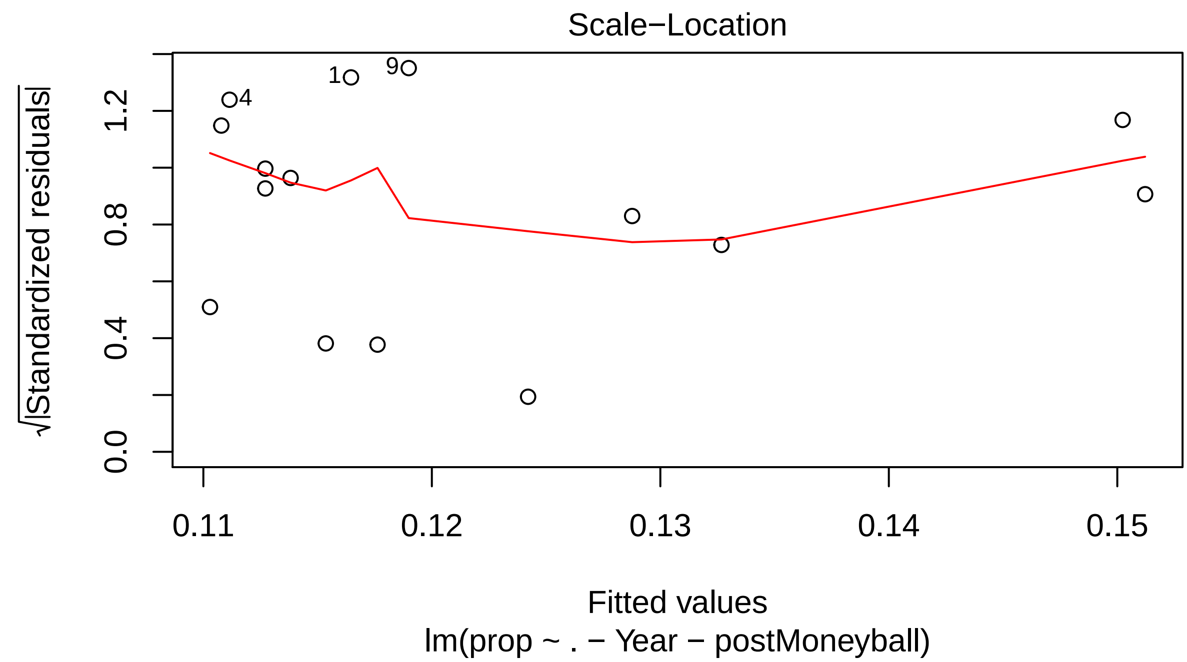
```
##
## Residual standard error: 0.04385 on 26 degrees of freedom
## Multiple R-squared:  0.5777, Adjusted R-squared:  0.5453
## F-statistic: 17.79 on 2 and 26 DF,  p-value: 1.357e-05
```

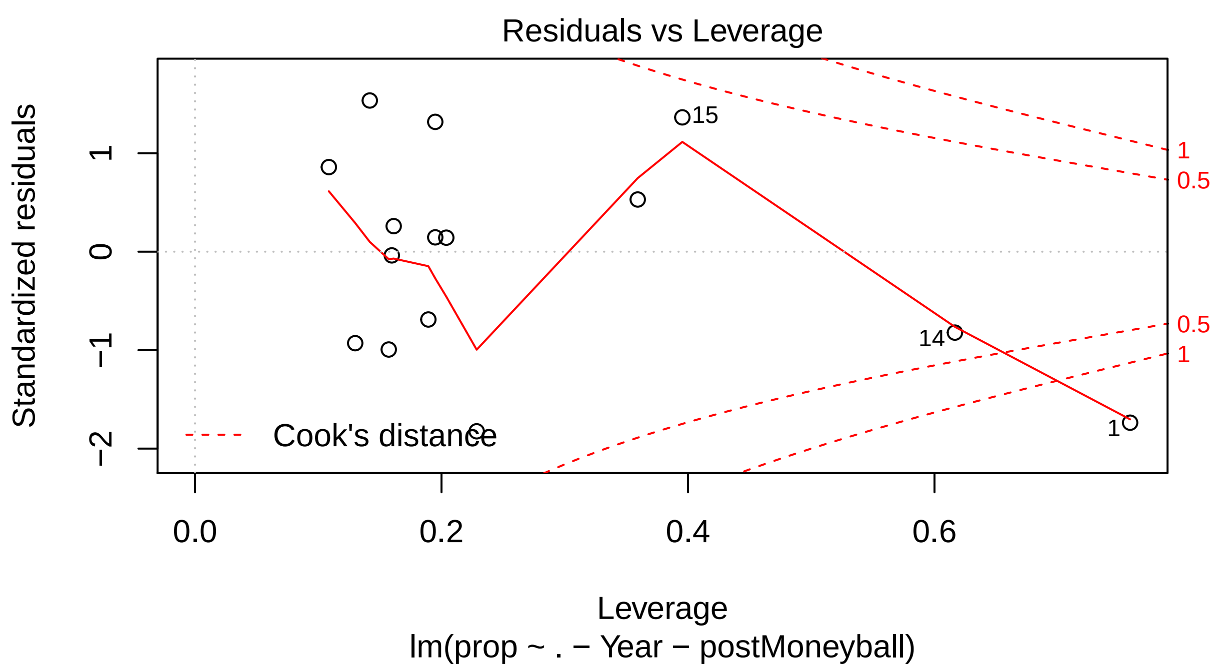



```
##
## Call:
## lm(formula = prop ~ . - Year - postMoneyball, data = couldabeens_post)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.065416 -0.030073  0.005184  0.025223  0.058085
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.752e-01  1.014e+00  -0.370    0.718
## totRev       8.165e-05  1.400e-04   0.583    0.571
## totPayroll  -1.724e-04  3.235e-04  -0.533    0.605
## labShare     1.031e+00  2.217e+00   0.465    0.651
##
## Residual standard error: 0.04083 on 11 degrees of freedom
## Multiple R-squared:  0.1216, Adjusted R-squared:  -0.118
## F-statistic: 0.5075 on 3 and 11 DF,  p-value: 0.6851
```


Residuals vs Fitted
Residuals
0.06
0.02
−0.02
−0.08
4
2
9
0.11
0.12
0.13
0.14
0.15
Fitted values
lm(prop ~ . − Year − postMoneyball)


Normal Q−Q
Standardized residuals
1
0
−1
4
1
9
−1
0
1
Theoretical Quantiles
lm(prop ~ . − Year − postMoneyball)


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
0.11
0.12
0.13
0.14
0.15
Fitted values
lm(prop ~ . − Year − postMoneyball)
1
9
4


Residuals vs Leverage
Standardized residuals
−2
−1
0
1
0.0
0.2
0.4
0.6
Leverage
lm(prop ~ . − Year − postMoneyball)
Cook's distance
15
14
1
1
0.5
0.5
1

Resampling