Modeling

Group 6

The Data

```
## # A tibble: 6 x 6
##    Year  prop totRev totPayroll labShare postMoneyball
##   <dbl> <dbl>  <dbl>      <dbl>    <dbl>         <dbl>
## 1  1989 0.130  1346.       454.    0.338             0
## 2  1990 0.137  1459.       630.    0.432             0
## 3  1991 0.302  1584.       784.    0.495             0
## 4  1992 0.2    1774.       857.    0.483             0
## 5  1993 0.159  1687        885.    0.525             0
## 6  1994 0.211  1410.       882.    0.625             0
```

Yearly Payroll Data

```
## # A tibble: 6 x 4
##    Year totRev totPayroll labShare
##   <dbl>  <dbl>      <dbl>    <dbl>
## 1  1990  1346.       454.    0.338
## 2  1991  1459.       630.    0.432
## 3  1992  1584.       784.    0.495
## 4  1993  1774.       857.    0.483
## 5  1994  1687        885.    0.525
## 6  1995  1410.       882.    0.625
```

Checking Correlations

```
# Revenue and Labor Share
cor(couldabeens$totRev, couldabeens$labShare)
```

```
## [1] -0.5044864
```

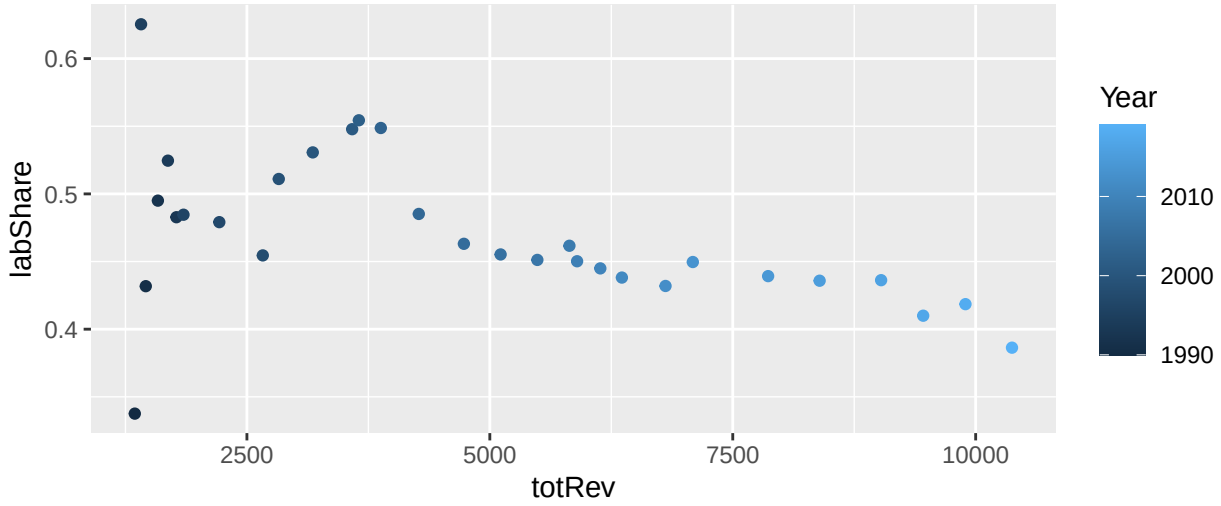
```
# Revenue and Payroll
cor(couldabeens$totRev, couldabeens$totPayroll)
```

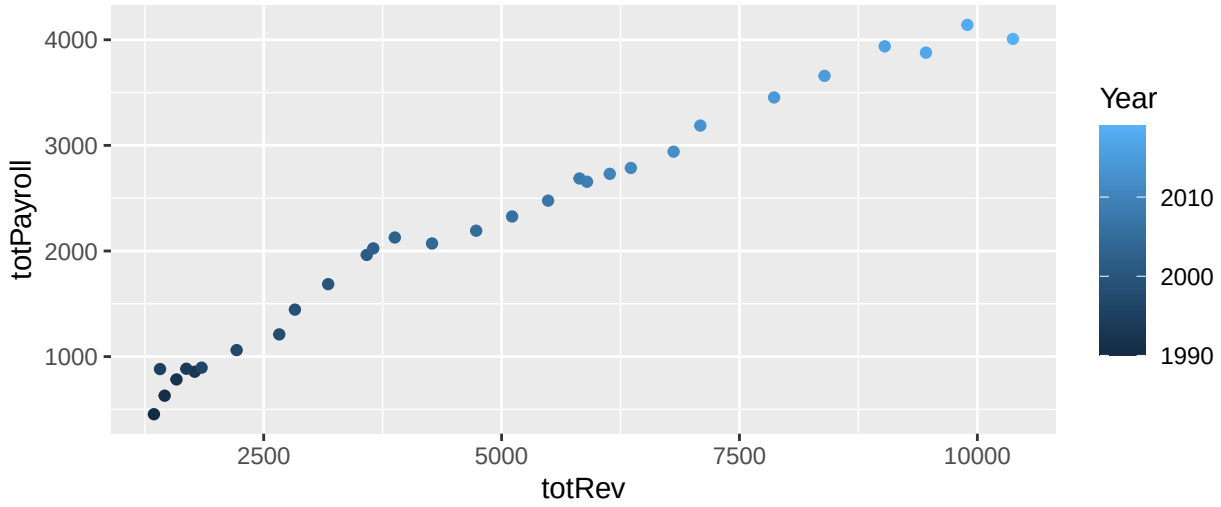
```
## [1] 0.9897413
```

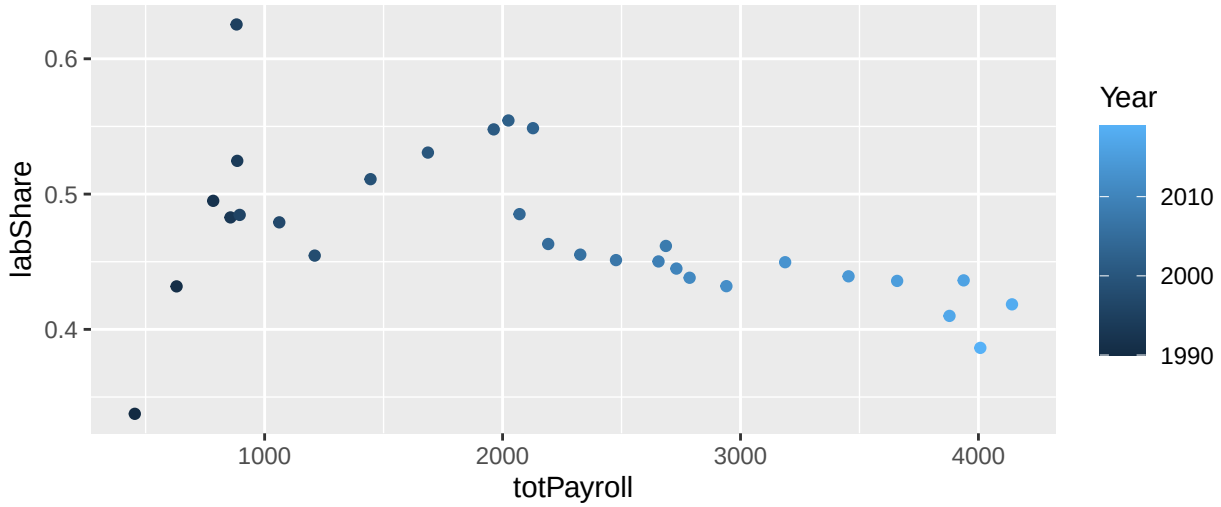
```
# Payroll and Labor Share
cor(couldabeens$totPayroll, couldabeens$labShare)
```

```
## [1] -0.4059079
```

Predictor Visualization

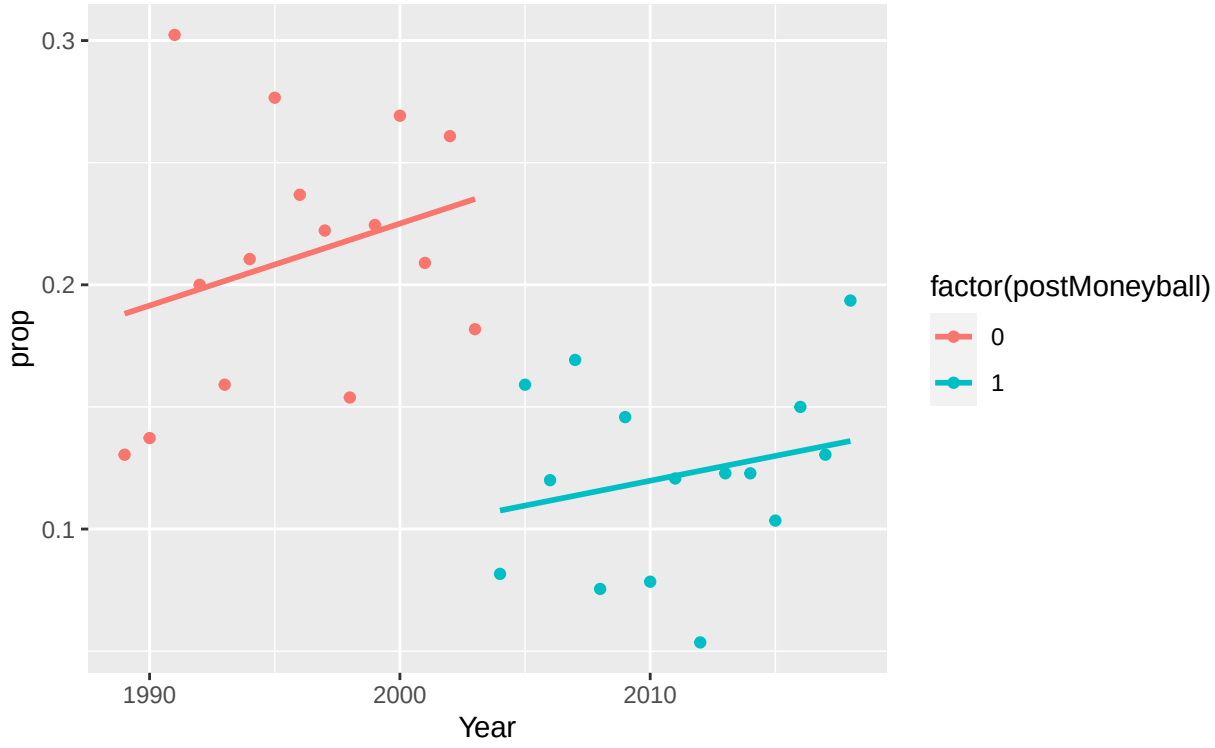




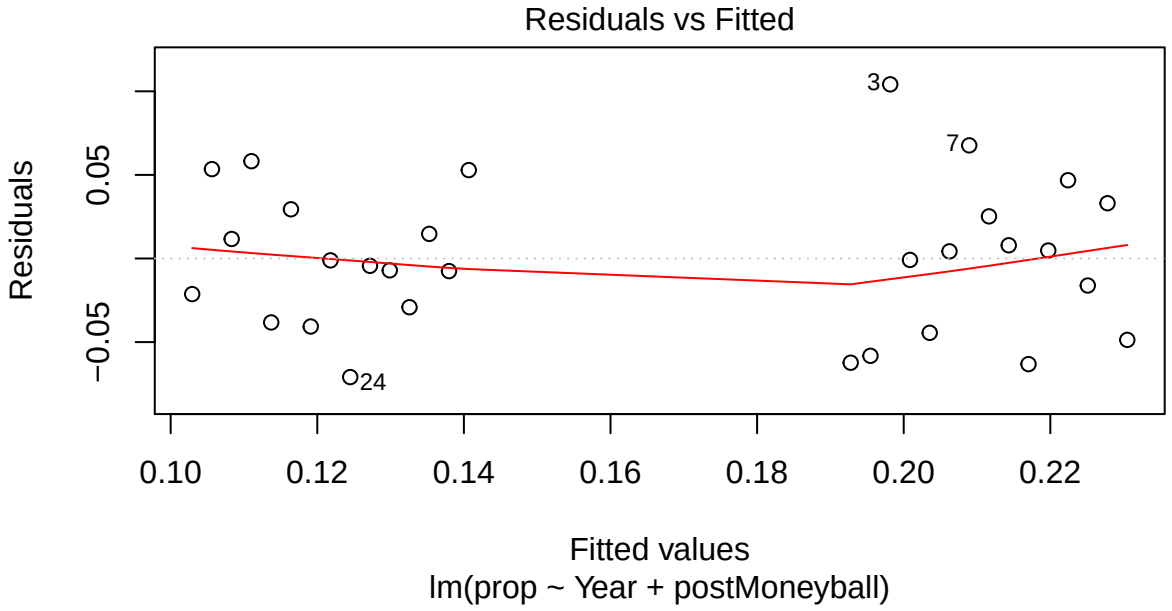


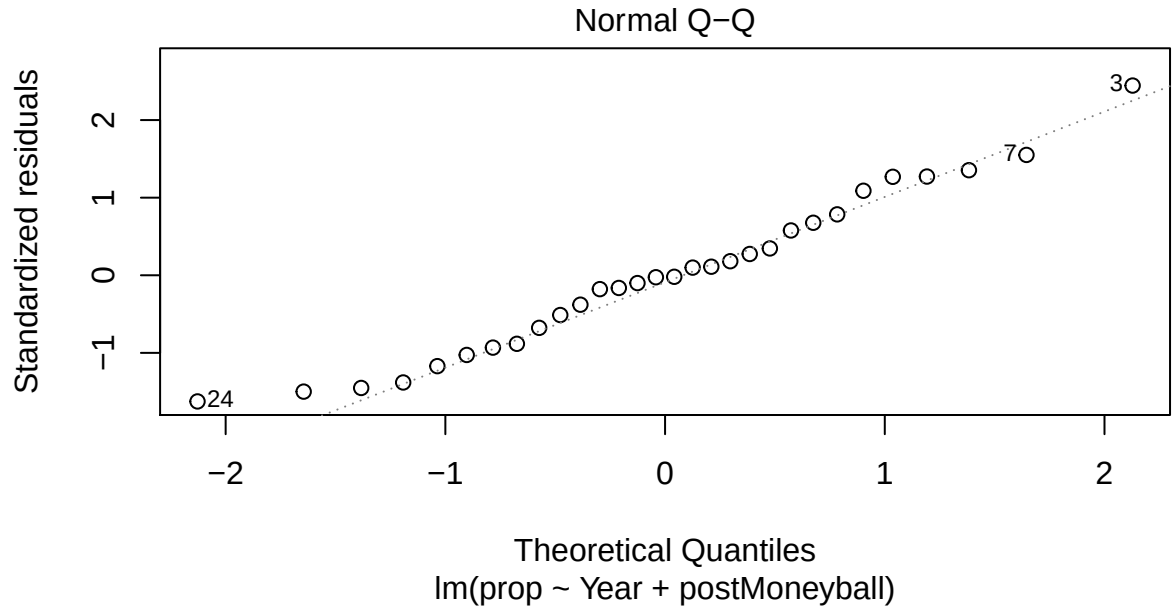
Linear Model: Year (Same Slopes)

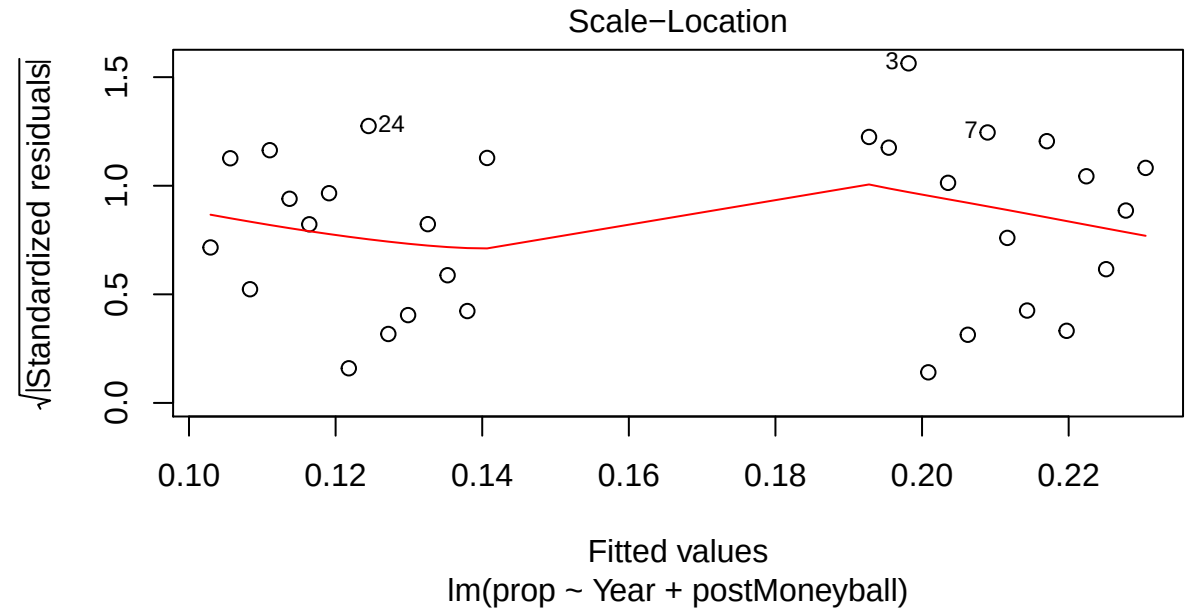
```
##
## Call:
## lm(formula = prop ~ Year + postMoneyball, data = couldabeens)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.070924 -0.035965 -0.000981  0.028370  0.104168
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -5.168095   3.812552  -1.356 0.186475
## Year           0.002695   0.001910   1.411 0.169645
## postMoneyball -0.130262   0.033065  -3.940 0.000519 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0452 on 27 degrees of freedom
## Multiple R-squared:  0.5394, Adjusted R-squared:  0.5052
## F-statistic: 15.81 on 2 and 27 DF,  p-value: 2.853e-05
```

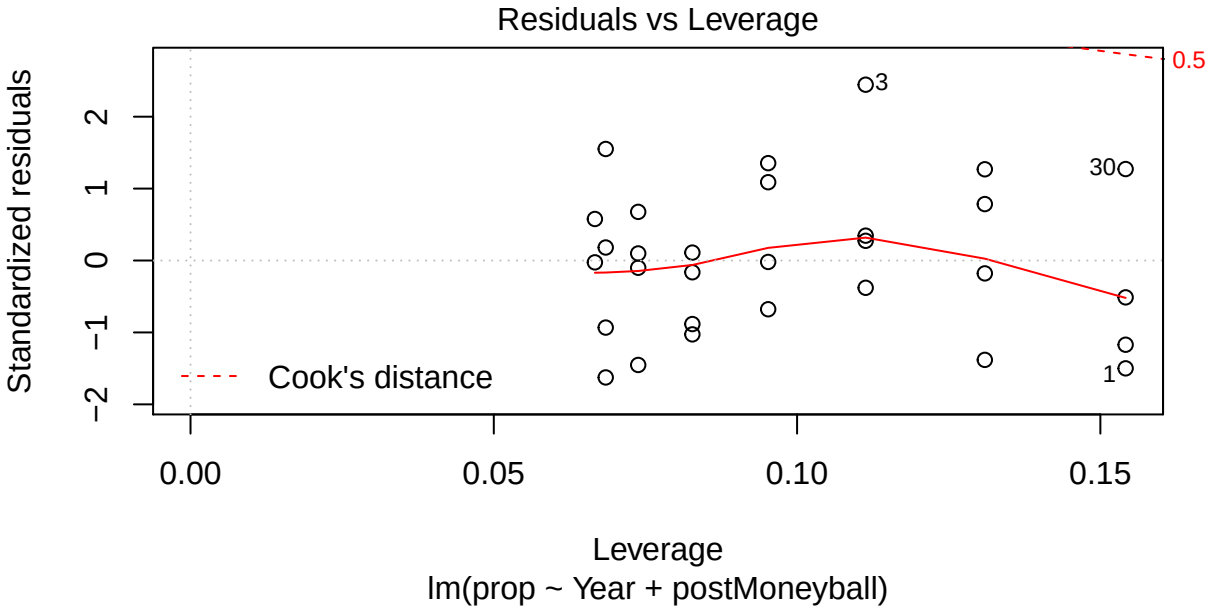


Diagnostic Plots: Year Linear Model






Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
24
3
7
0.10
0.12
0.14
0.16
0.18
0.20
0.22
Fitted values
lm(prop ~ Year + postMoneyball)

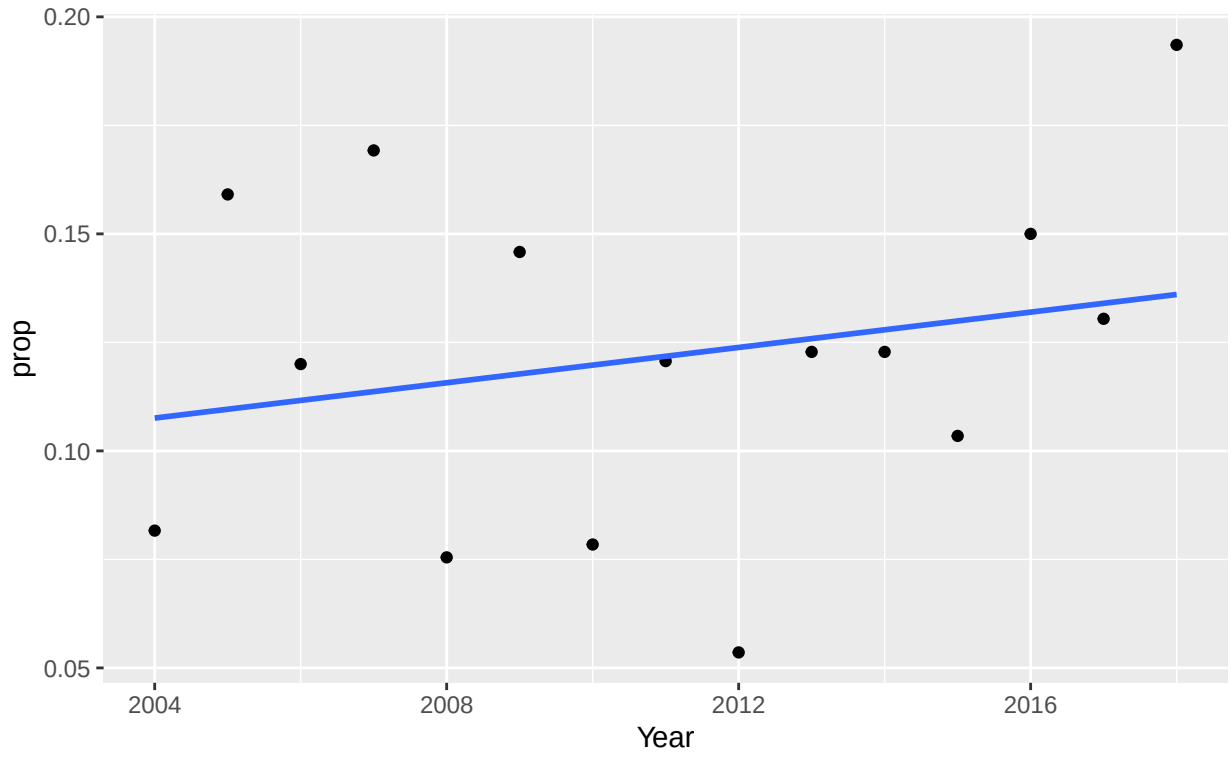

Residuals vs Leverage
Standardized residuals
−2
−1
0
1
2
0.5
3
30
1
Cook's distance
0.00
0.05
0.10
0.15
Leverage
lm(prop ~ Year + postMoneyball)

Linear Model: Year (Different Slopes)

Since we realize `postMoneyball` is a statistically significant variable, we decide to attempt a different slopes model and attempt to measure the effect sizes of the partitioned data.

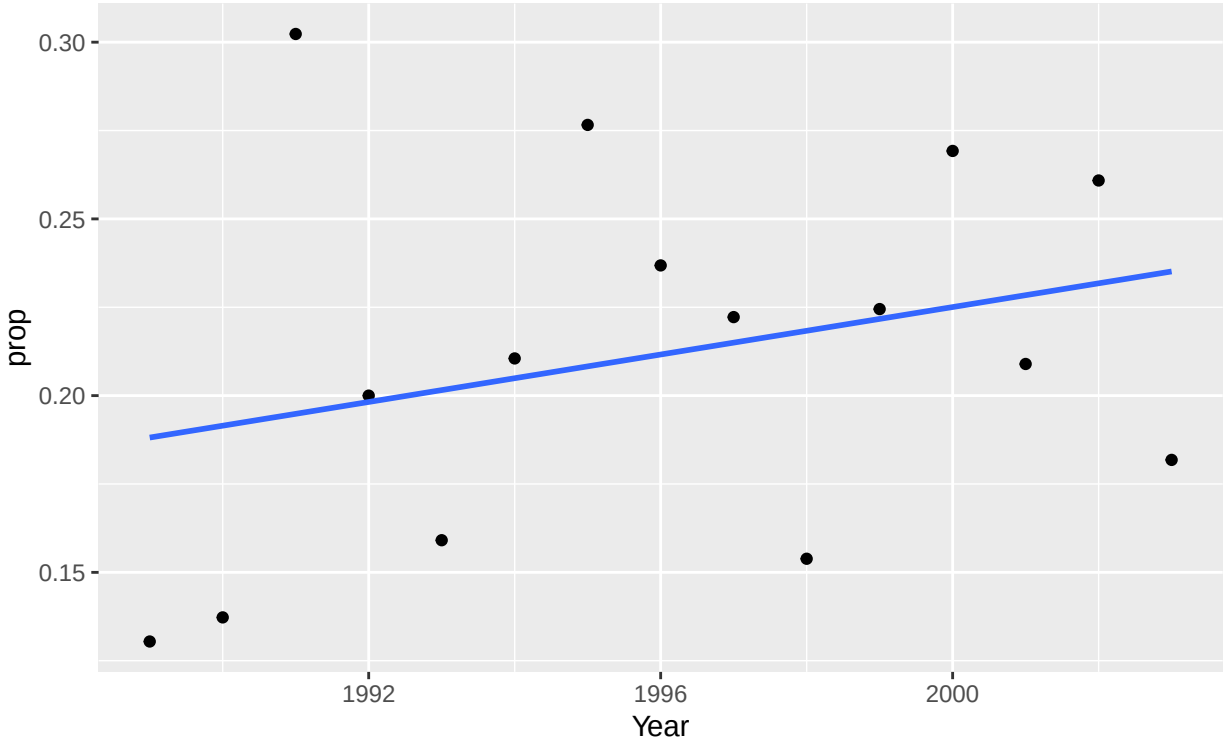
Post-Moneyball

```
##
## Call:
## lm(formula = prop ~ Year, data = couldabeens_post)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07026 -0.02621 -0.00306  0.02307  0.05751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.967977   4.680332  -0.848    0.412
## Year         0.002034   0.002327   0.874    0.398
##
## Residual standard error: 0.03894 on 13 degrees of freedom
## Multiple R-squared:  0.05548,    Adjusted R-squared:  -0.01718
## F-statistic: 0.7636 on 1 and 13 DF,  p-value: 0.3981
```



Pre-Moneyball

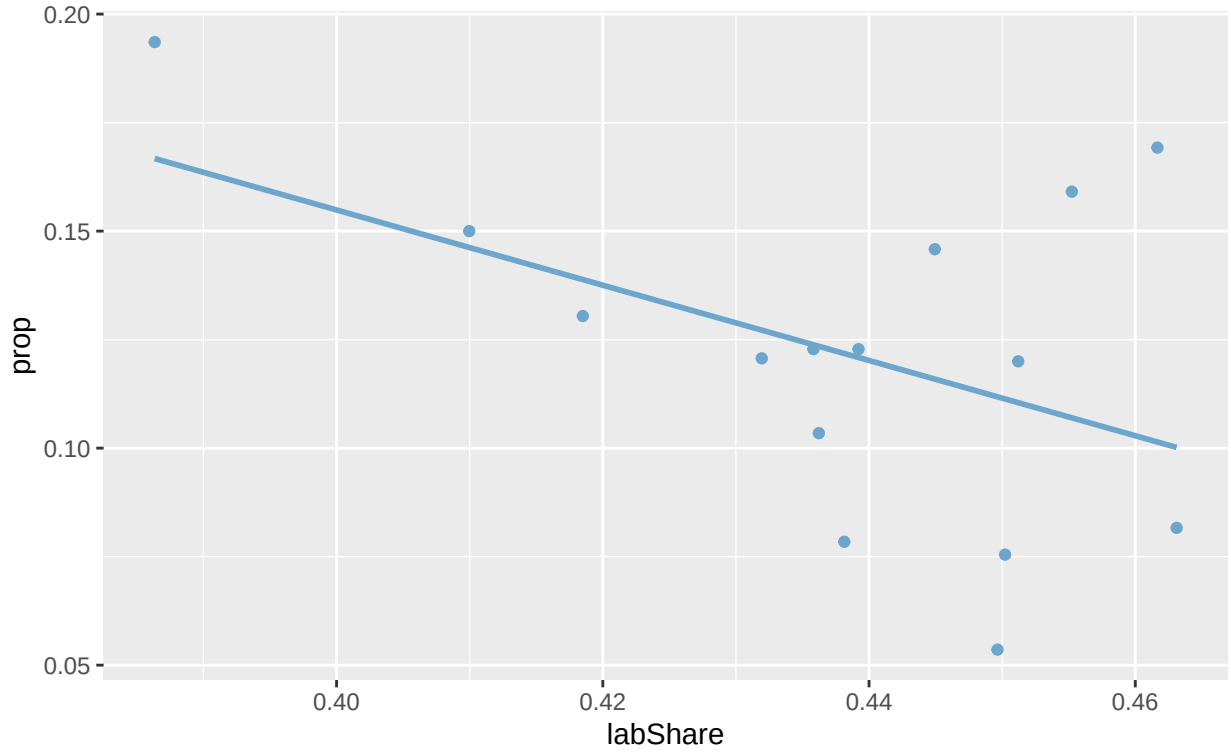
```
##
## Call:
## lm(formula = prop ~ Year, data = couldabeens_pre)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.064501 -0.047893  0.002786  0.027152  0.107476
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.488553   6.207251  -1.045    0.315
## Year         0.003357   0.003110   1.079    0.300
##
## Residual standard error: 0.05204 on 13 degrees of freedom
## Multiple R-squared:  0.08225,    Adjusted R-squared:  0.01166
## F-statistic: 1.165 on 1 and 13 DF,  p-value: 0.3
```



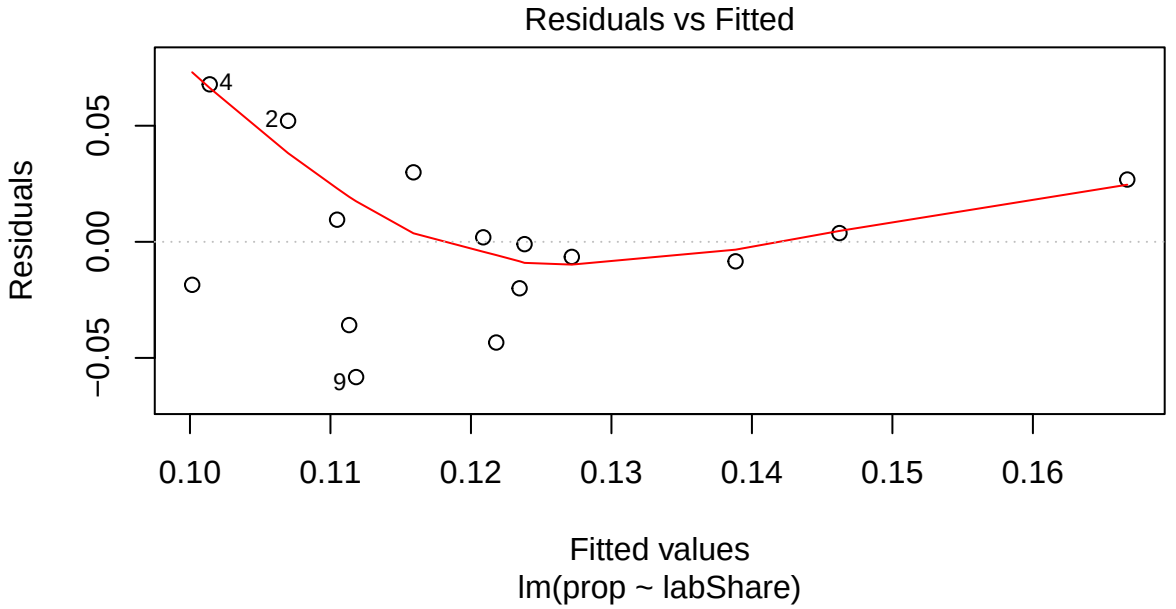
Linear Model: Labor Share

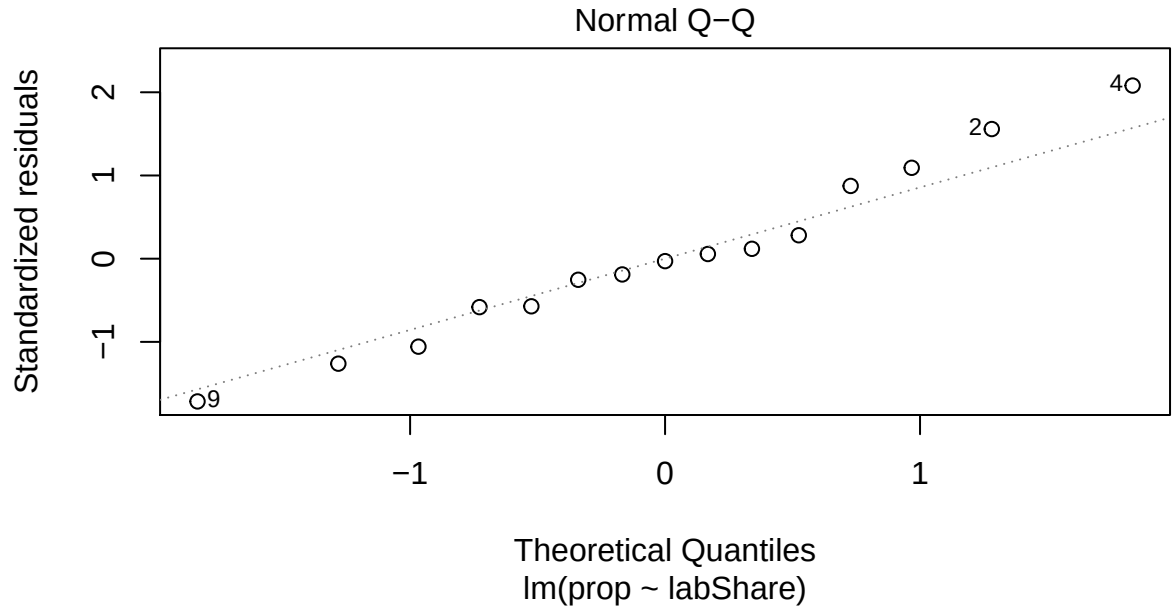
```
##
## Call:
## lm(formula = prop ~ labShare, data = couldabeens_post)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.058253 -0.019268 -0.001008  0.018178  0.067824
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5017     0.2036   2.464   0.0284 *
## labShare     -0.8670     0.4642  -1.868   0.0845 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03558 on 13 degrees of freedom
## Multiple R-squared:  0.2116, Adjusted R-squared:  0.151
## F-statistic:  3.49 on 1 and 13 DF,  p-value: 0.08446
```

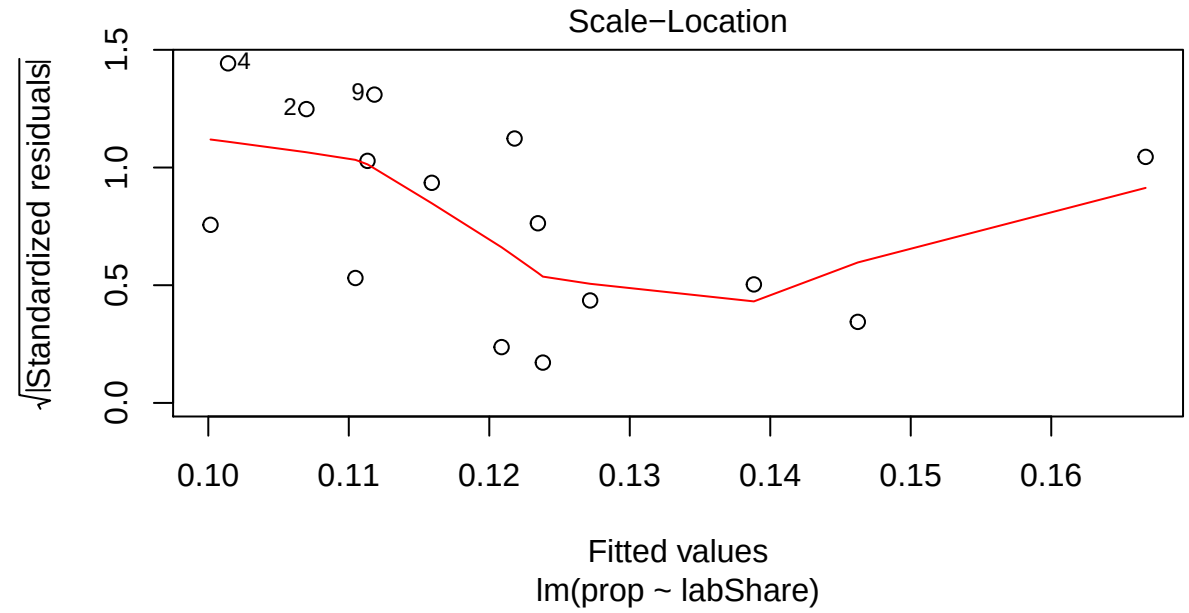
Relationship between Labor Share and Couldabeen Rates

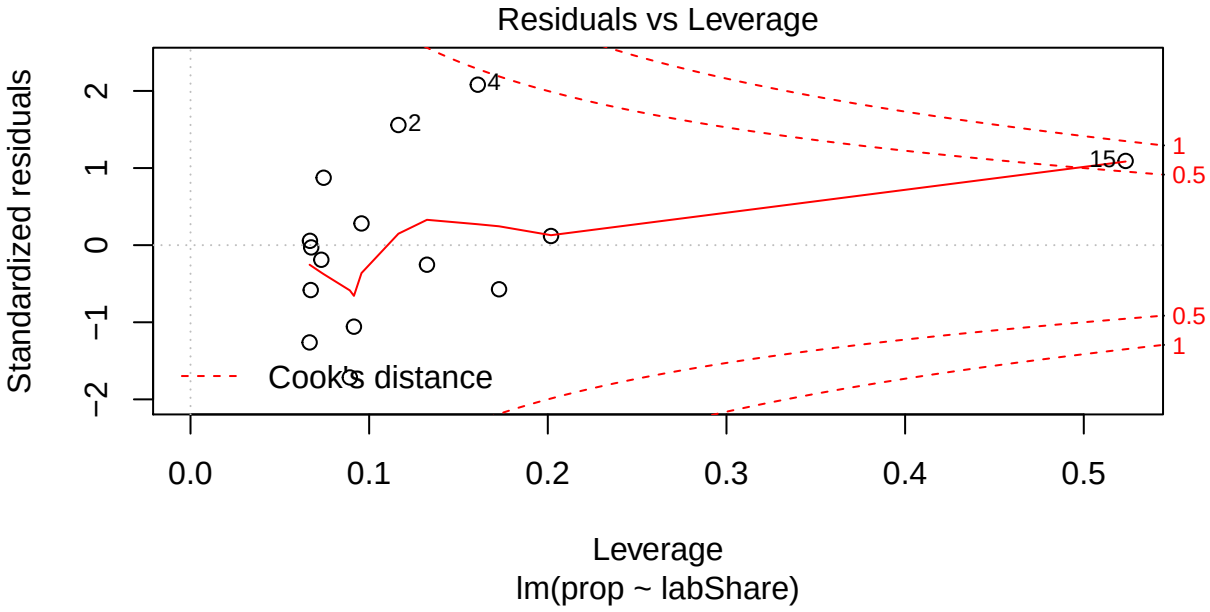


Diagnostic Plots: Labor Share Linear Model






Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
4
2
9
0.10
0.11
0.12
0.13
0.14
0.15
0.16
Fitted values
lm(prop ~ labShare)


Residuals vs Leverage
Standardized residuals
−2
−1
0
1
2
4
2
15
1
0.5
0.5
1
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
Leverage
lm(prop ~ labShare)

Couldabeen Rates across the Years