Modeling

Group 6

The Data

```
## # A tibble: 6 x 6
##    Year   prop totRev totPayroll labShare postMoneyball
##   <dbl>  <dbl>  <dbl>      <dbl>    <dbl>          <dbl>
## 1  1969 0.125      NA         NA       NA              0
## 2  1970 0.0857     NA         NA       NA              0
## 3  1971 0.0909     NA         NA       NA              0
## 4  1972 0.122      NA         NA       NA              0
## 5  1973 0.171      NA         NA       NA              0
## 6  1974 0.111      NA         NA       NA              0
```

```
## # A tibble: 6 x 6
##    Year  prop totRev totPayroll labShare postMoneyball
##   <dbl> <dbl>  <dbl>      <dbl>    <dbl>          <dbl>
## 1  2013 0.123   7864      3454.    0.439              1
## 2  2014 0.123   8394      3658.    0.436              1
## 3  2015 0.103   9027      3938.    0.436              1
## 4  2016 0.15    9460      3878.    0.410              1
## 5  2017 0.130   9895      4141.    0.418              1
## 6  2018 0.194  10374      4008.    0.386              1
```

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
```

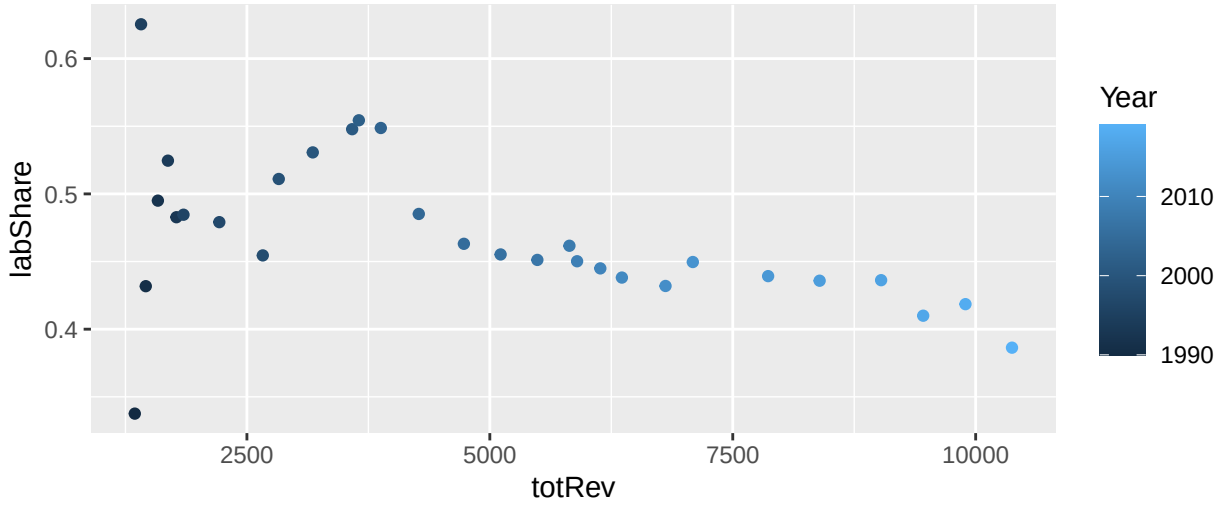
```
# Revenue and Payroll
cor(couldabeens$totRev, couldabeens$totPayroll)
```

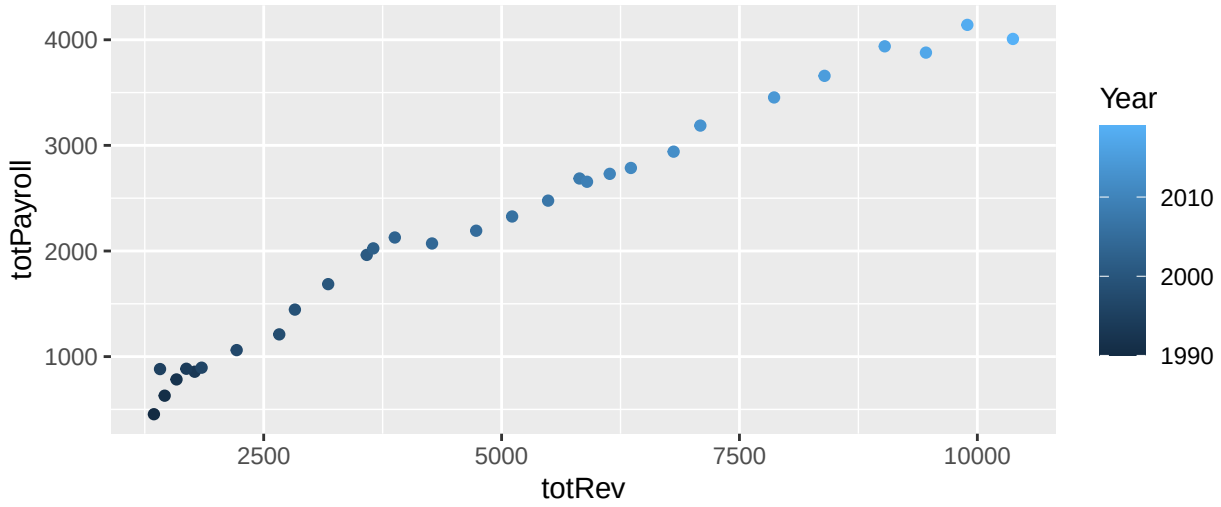
```
## [1] 0.9897413
```

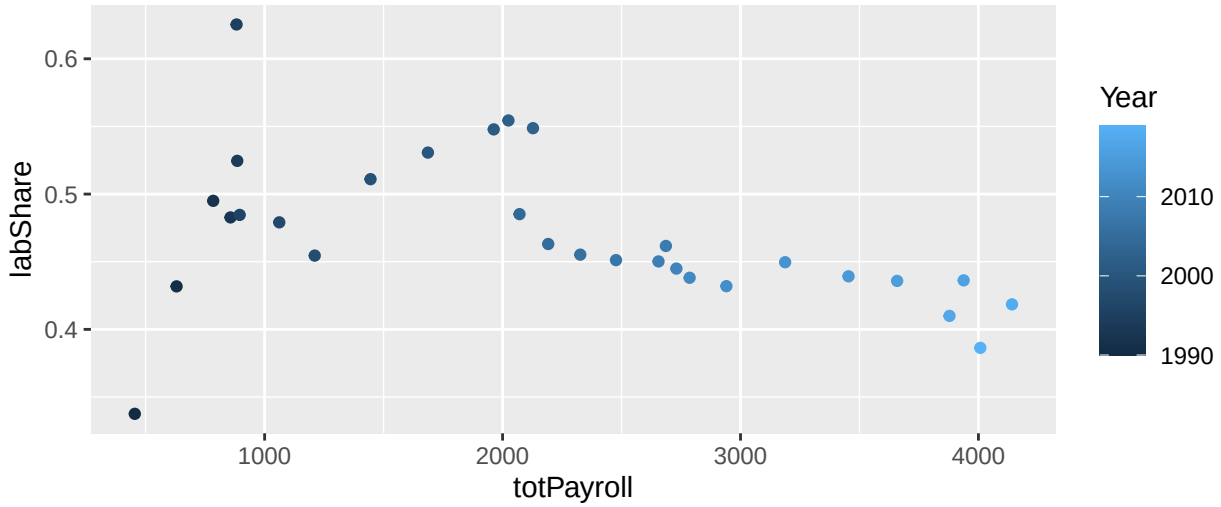
```r
# Payroll and Labor Share
cor(couldabeens$totPayroll, couldabeens$labShare)
```

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

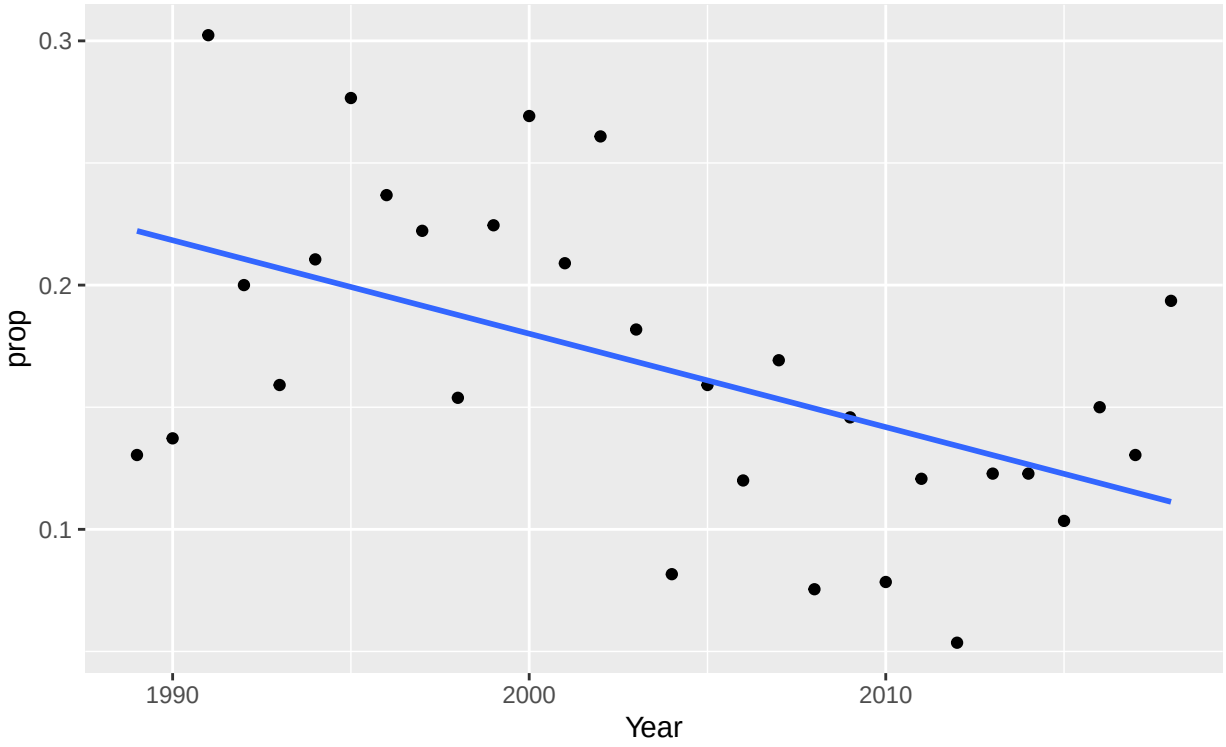
Predictor Visualization







Couldabeens across the Years



Hypothesis Test: Splitting the Data

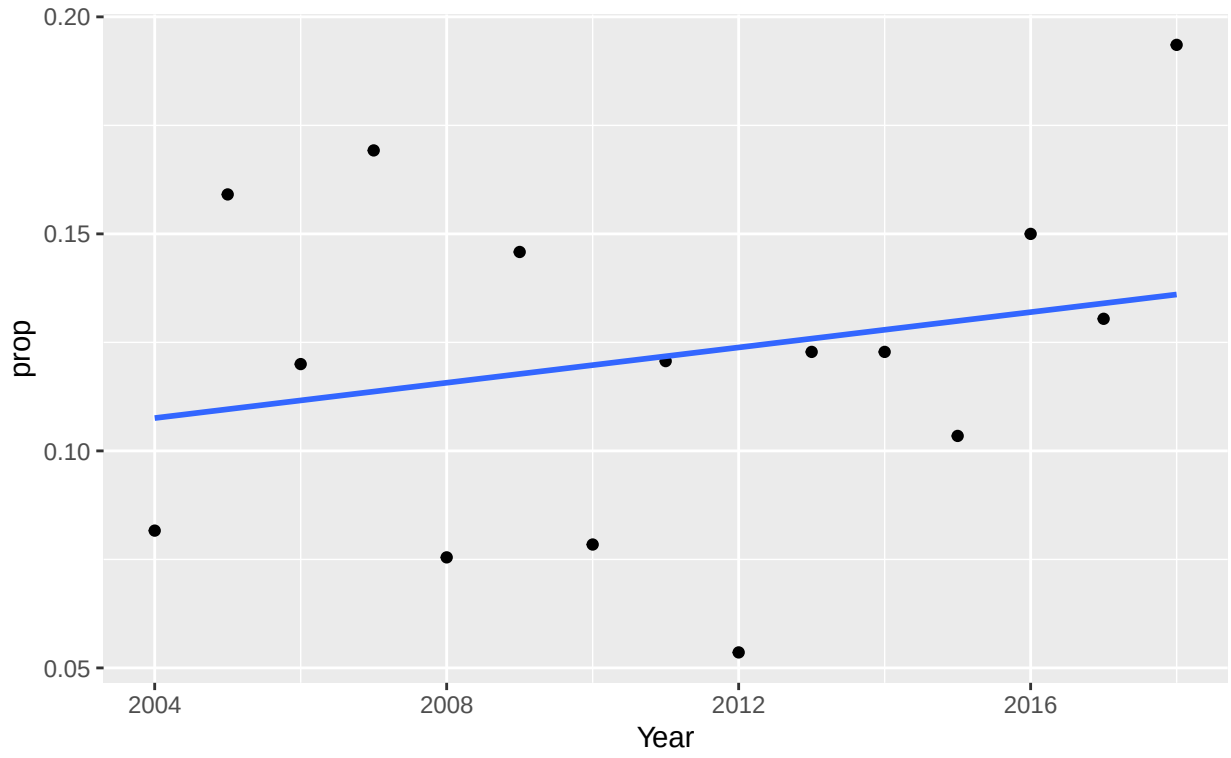
```
##
## Call:
## lm(formula = prop ~ postMoneyball, data = couldabeens)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.08120 -0.03758 -0.00005  0.02745  0.09069
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.21163    0.01188  17.821  < 2e-16 ***
## postMoneyball -0.08983    0.01679  -5.349 1.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04599 on 28 degrees of freedom
## Multiple R-squared:  0.5054, Adjusted R-squared:  0.4877
## F-statistic: 28.61 on 1 and 28 DF,  p-value: 1.069e-05
```

Linear Model: Year (Different Eras)

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.

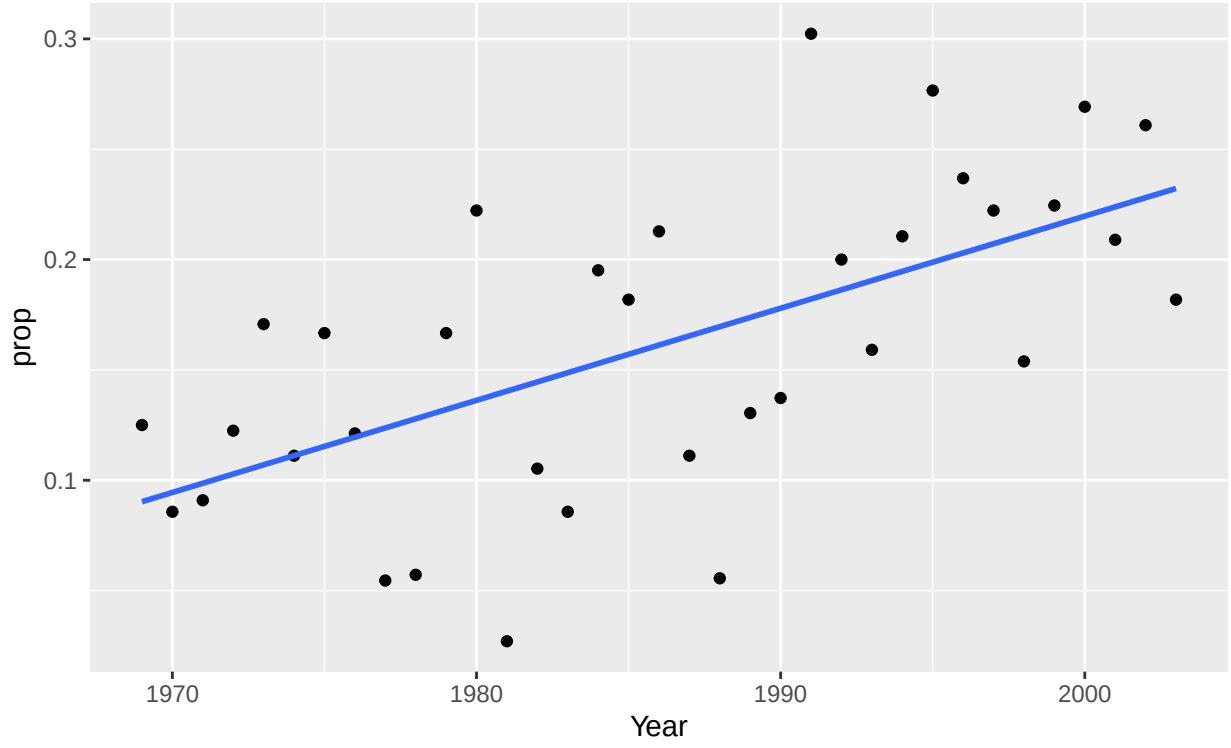
Post-Moneyball

```
##
## 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

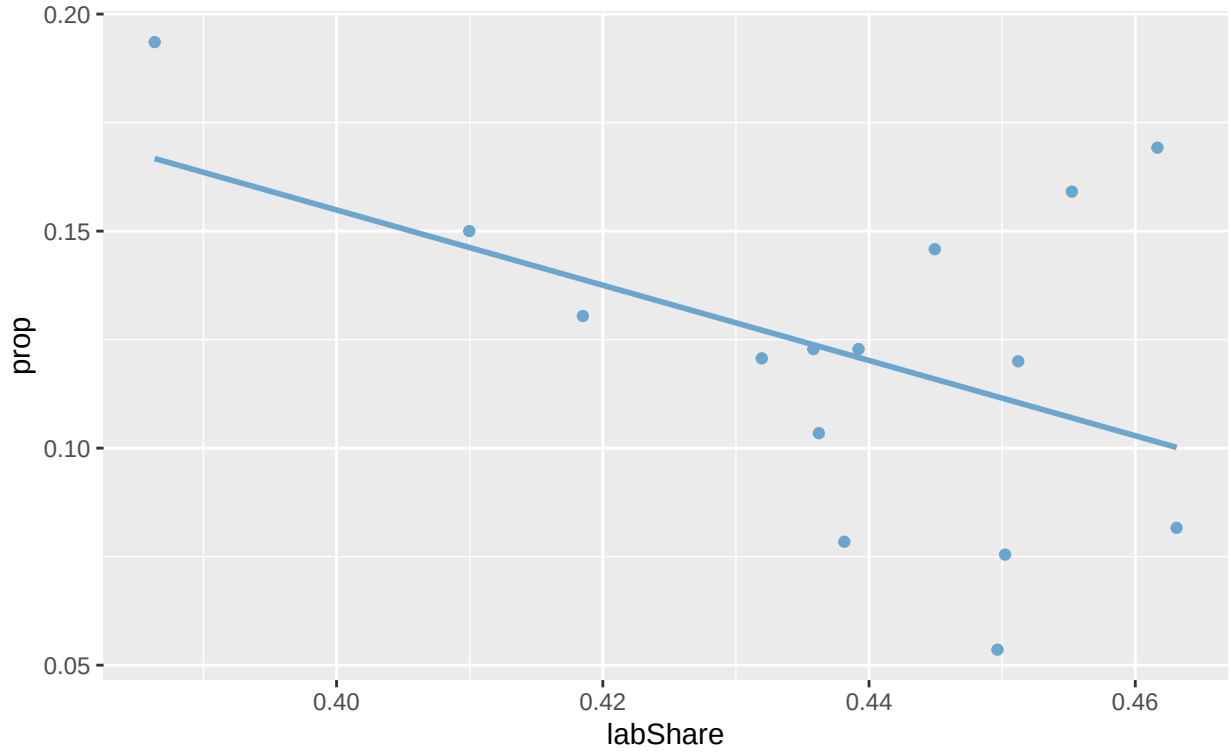
```
##
## Call:
## lm(formula = prop ~ Year, data = couldabeens_pre)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.114028 -0.042000  0.008993  0.034685  0.120220
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.1282589  1.8541569  -4.384 0.000112 ***
## Year         0.0041740  0.0009336   4.471 8.69e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05578 on 33 degrees of freedom
## Multiple R-squared:  0.3772, Adjusted R-squared:  0.3583
## F-statistic: 19.99 on 1 and 33 DF,  p-value: 8.69e-05
```



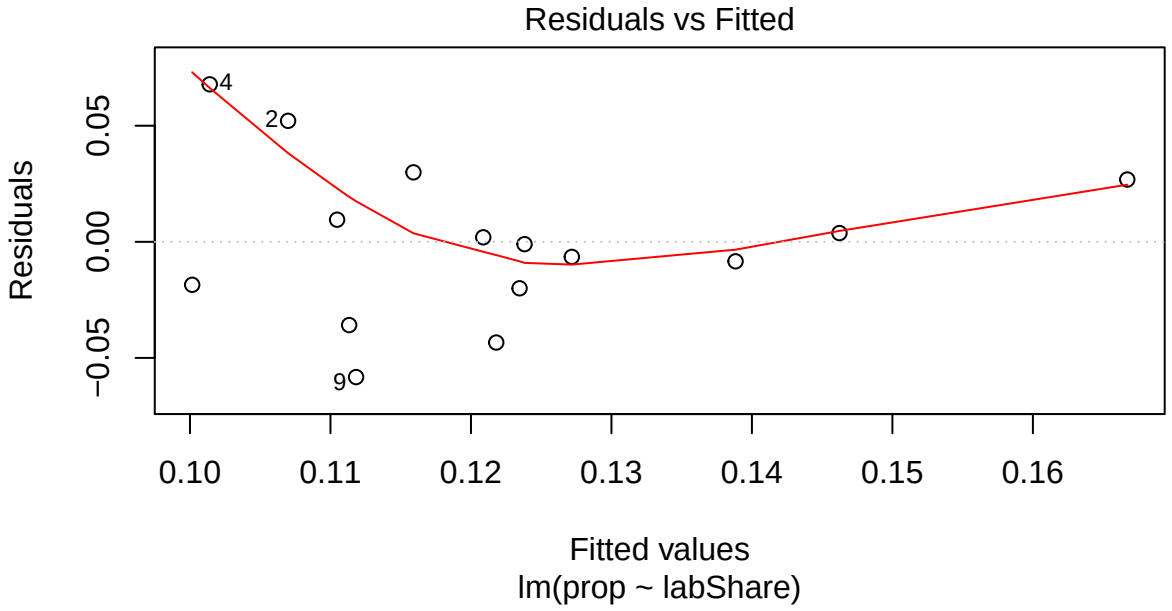
Linear Model: Labor Share

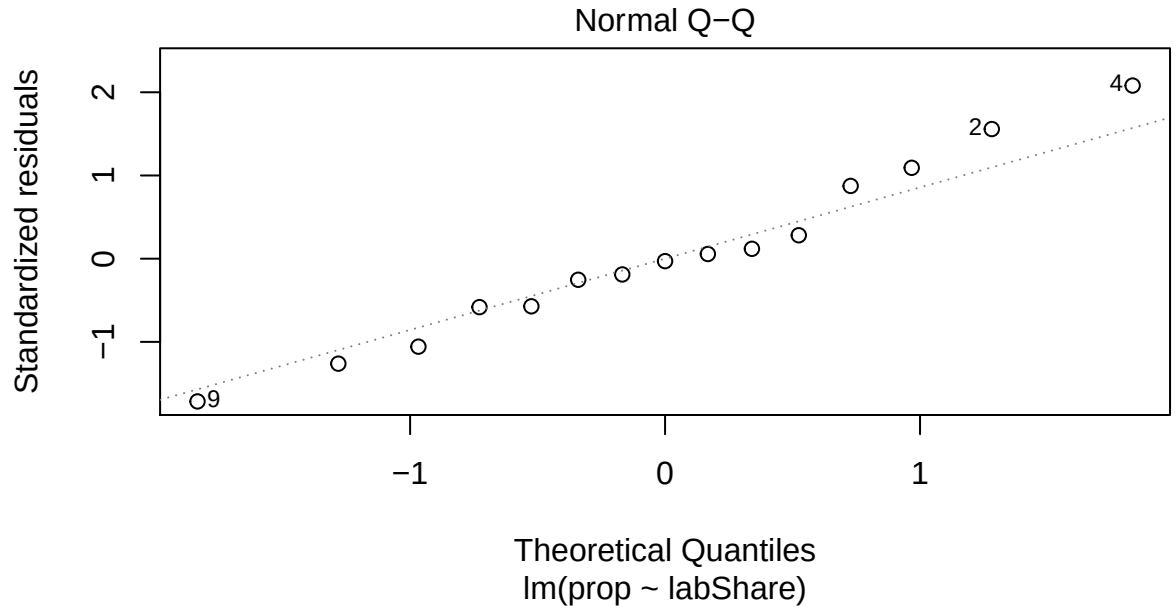
```
##
## Call:
## lm(formula = prop ~ labShare, data = couldabeens_post)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.058253 -0.019268 -0.001008  0.018178  0.067824
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5017     0.2036   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
```

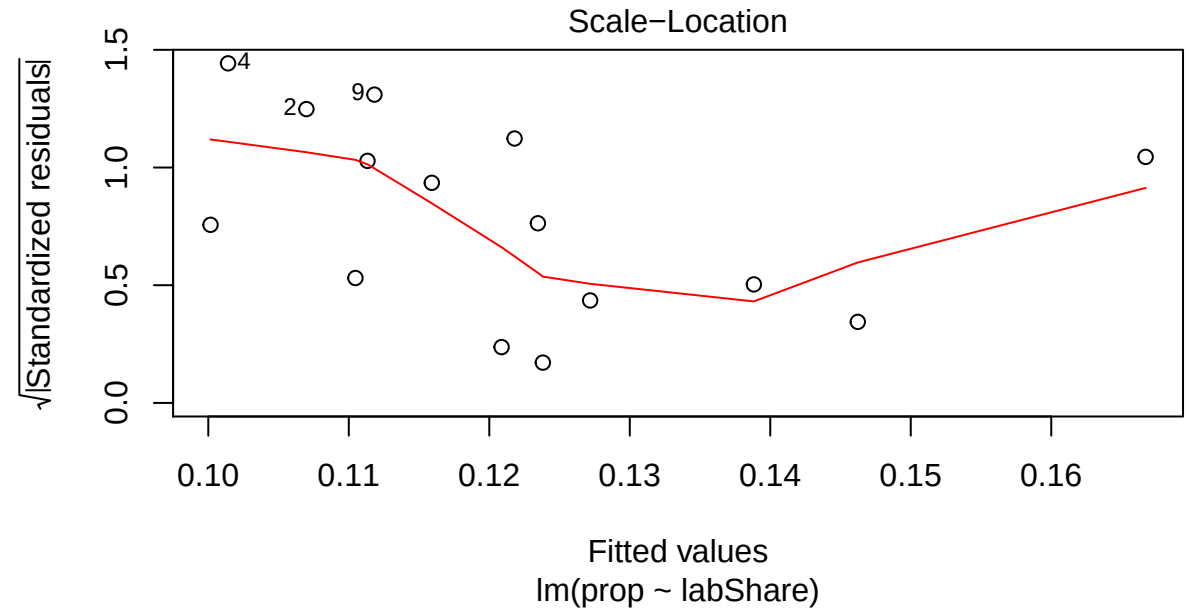
Relationship between Labor Share and Couldabeen Rates

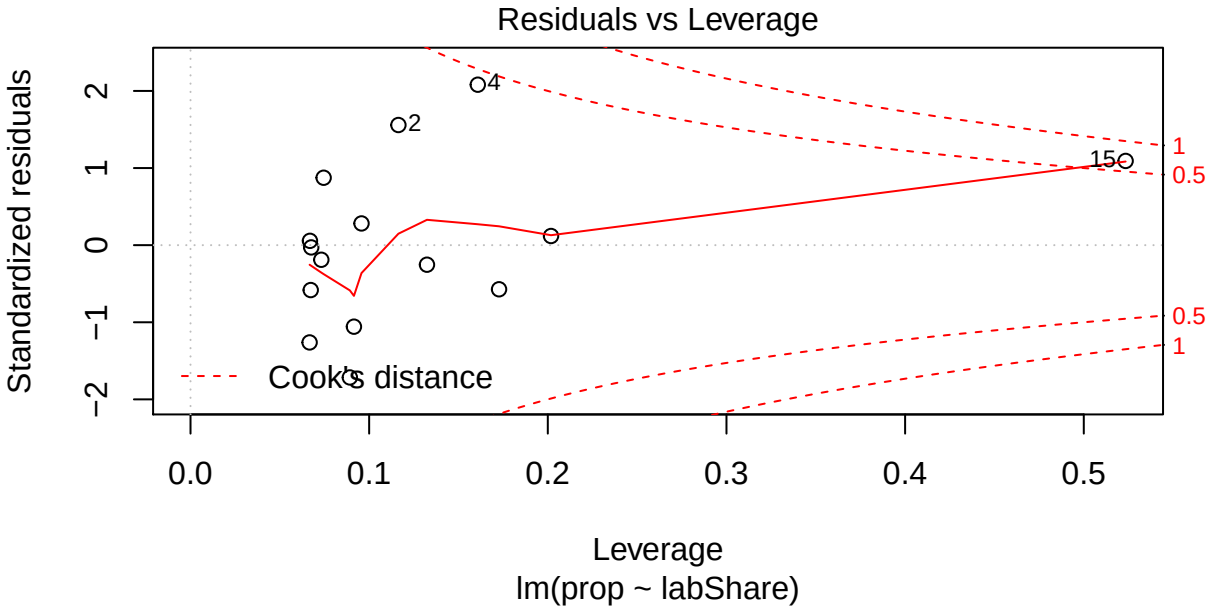


Diagnostic Plots: Labor Share Linear Model






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