

Report

Group 6

Abstract

Introduction

Methods

We mainly compare the proportions of couldabeen retirees across years in different time periods.

Exploratory Data Analysis

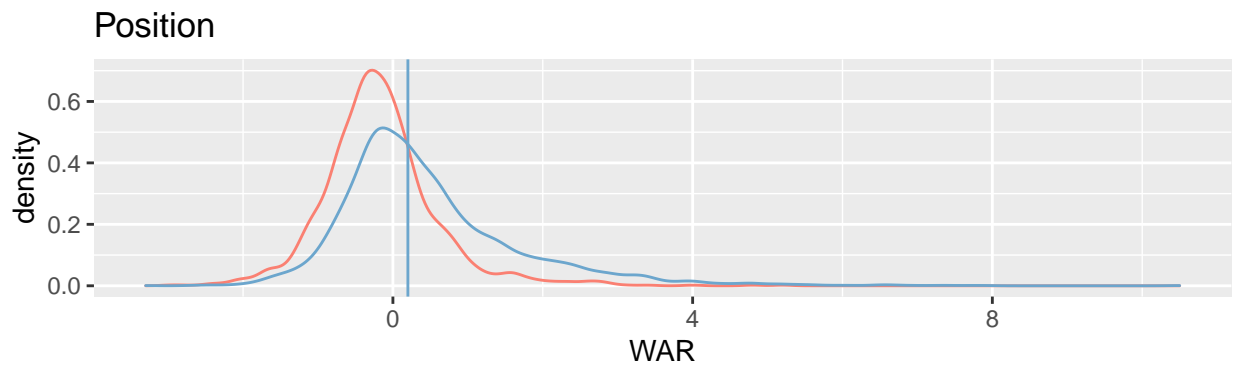
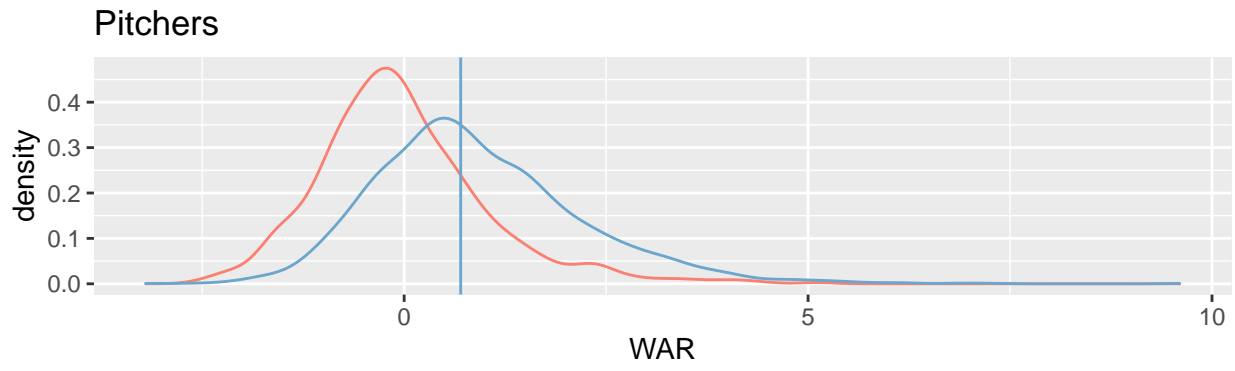
The Data

Our data comes from baseball-reference.com and we mainly have four sets of data.

1. Rookie pitchers
2. Rookie positions
3. Retired pitchers
4. Retired positions

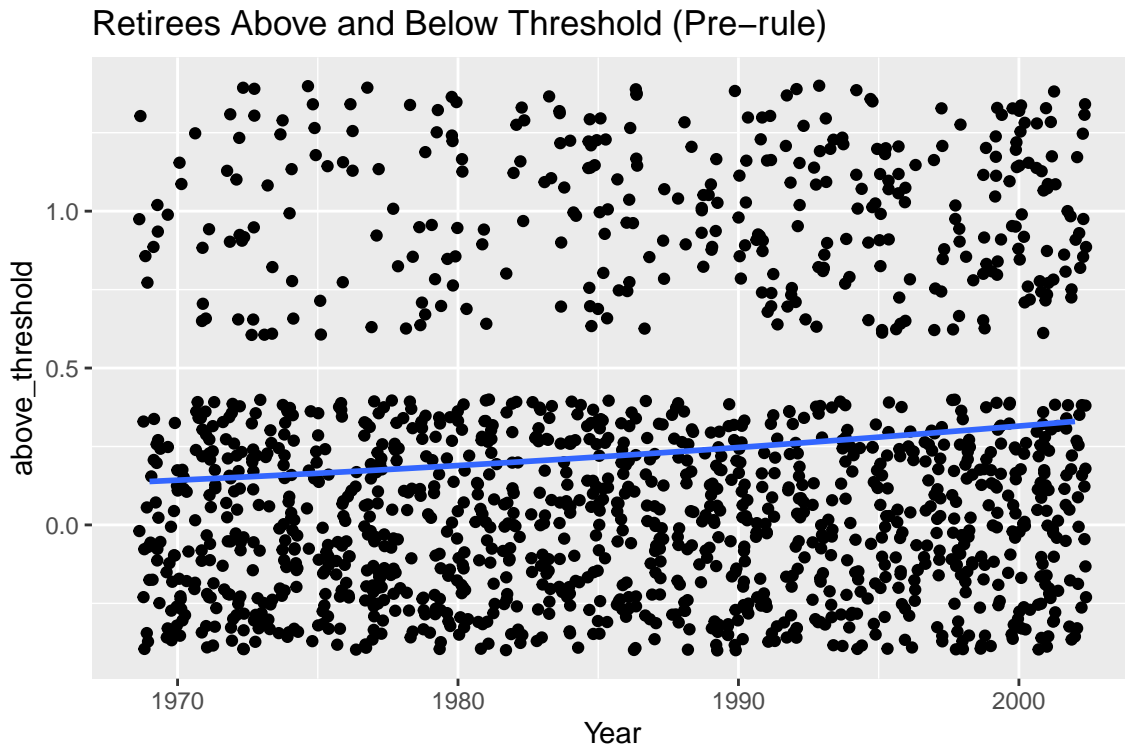
Then we compare the ratio of couldabeens from the retirees pool over two different time periods, 1969-2002 and 2003-2018.

Here are our models



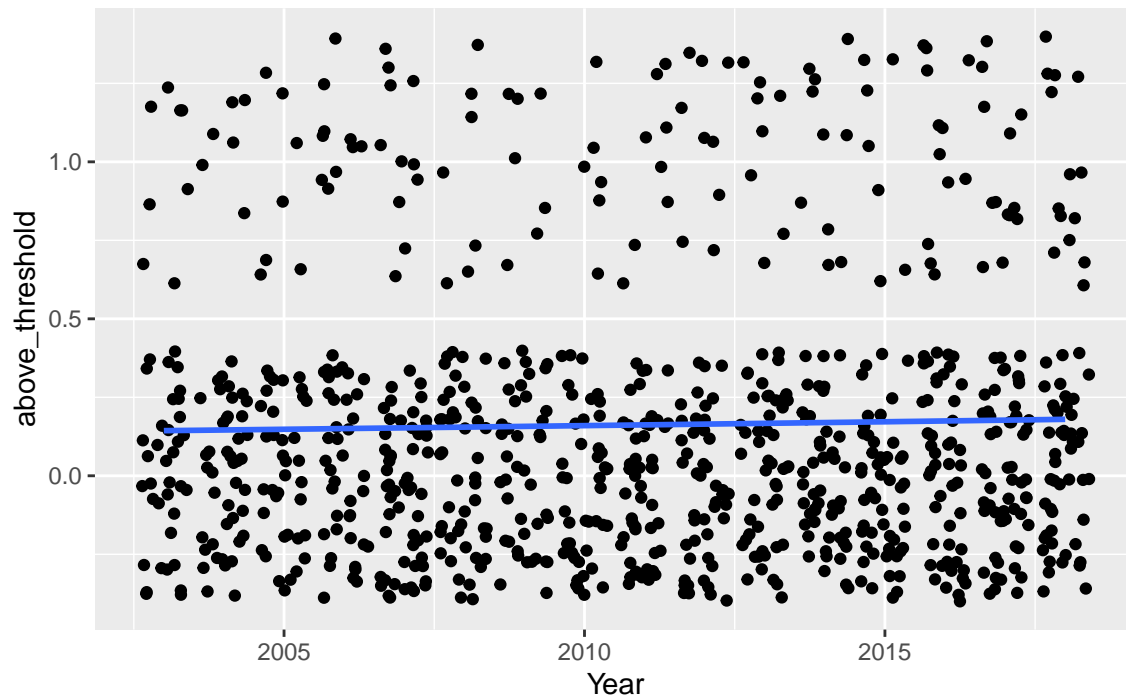
Results

Logistic Model



```
##
## Call:
## glm(formula = above_threshold ~ Year, family = "binomial", data = dataset)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8949  -0.7642  -0.6479  -0.5547   1.9883
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -68.617412  12.853177  -5.339 9.37e-08 ***
## Year          0.033921   0.006465   5.247 1.55e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1583.3  on 1470  degrees of freedom
## Residual deviance: 1554.9  on 1469  degrees of freedom
## AIC: 1558.9
##
## Number of Fisher Scoring iterations: 4
```

Retirees Above and Below Threshold (Post-rule)



```
##
## Call:
## glm(formula = above_threshold ~ Year, family = "binomial", data = dataset)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6297  -0.6096  -0.5900  -0.5617   1.9697
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -37.46198   40.21840  -0.931   0.352
## Year          0.01781    0.02000   0.891   0.373
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 781.43  on 880  degrees of freedom
## Residual deviance: 780.63  on 879  degrees of freedom
## AIC: 784.63
##
## Number of Fisher Scoring iterations: 4
```

Linear Model

```
# Partition dataset into years before and after rule
couldabeens_pre <- prerule(couldabeens)
couldabeens_post <- postrule(couldabeens)
```

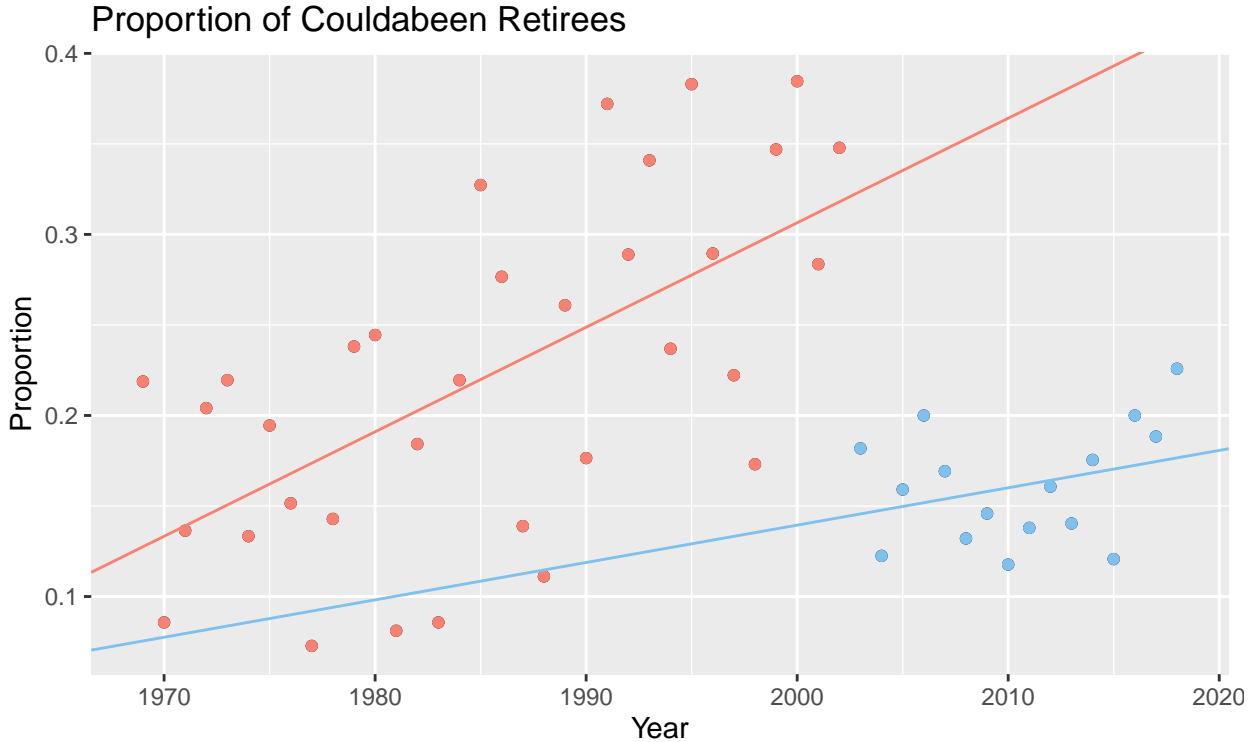
```
# Obtain linear model for pre-rule years
model_pre <- lm(formula = prop ~ I(Year), data = couldabeens_pre)
coefs_pre <- model_pre$coefficients
coefs_pre
```

```
##      (Intercept)          I(Year)
## -11.238735292    0.005772587
```

```
# Obtain linear model for post-rule years
model_post <- lm(formula = prop ~ I(Year), data = couldabeens_post)
coefs_post <- model_post$coefficients
coefs_post
```

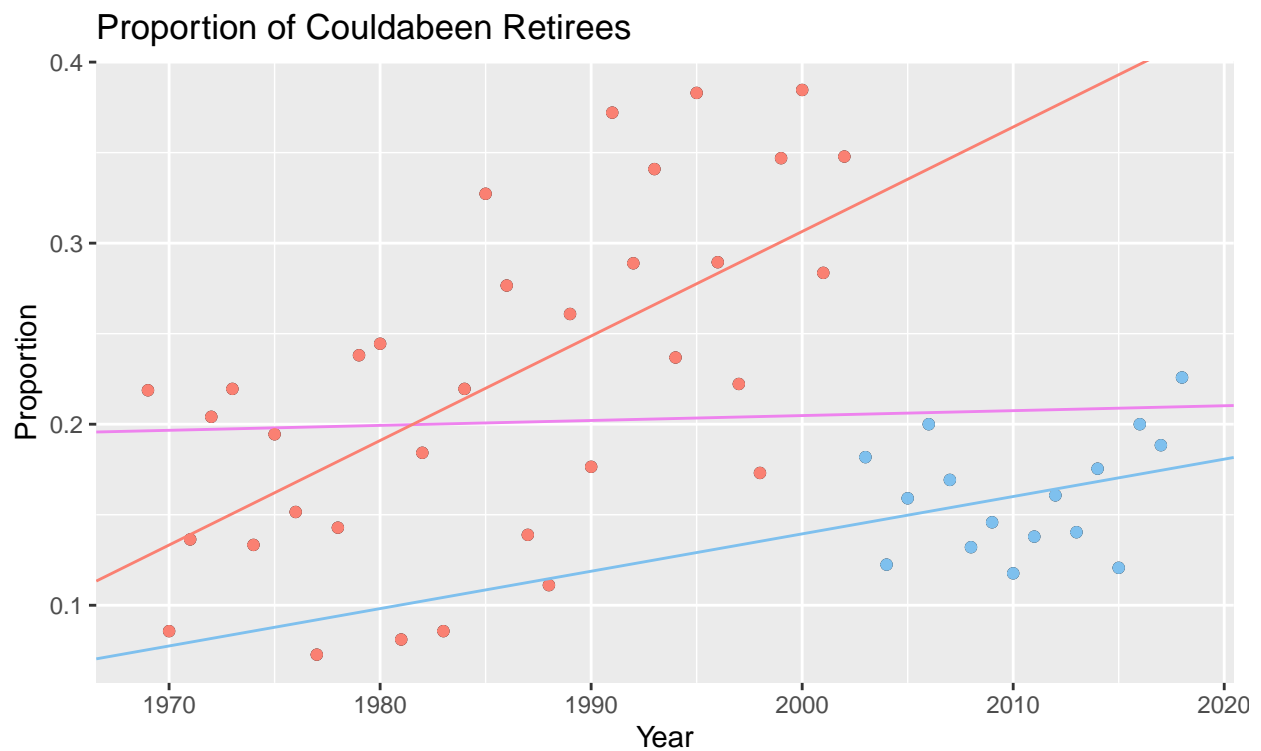
```
##      (Intercept)          I(Year)
## -3.987716737    0.002063571
```

Couldabeens Proportions



Simpson's Paradox

```
##  
## Call:  
## lm(formula = prop ~ I(Year), data = couldabeens)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.12581 -0.06211 -0.01833  0.04358  0.17984   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -0.3379371  1.6395066  -0.206   0.838      
## I(Year)      0.0002714  0.0008224   0.330   0.743      
##  
## Residual standard error: 0.08392 on 48 degrees of freedom  
## Multiple R-squared:  0.002263,    Adjusted R-squared:  -0.01852   
## F-statistic: 0.1089 on 1 and 48 DF,  p-value: 0.7429
```



Couldabeens: Pre-rule Era (1969-2002)

```
##  
## Call:  
## lm(formula = prop ~ I(Year), data = couldabeens_pre)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.126056 -0.044806  0.005781  0.053314  0.117608   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -11.238735    2.549750  -4.408  0.00011 ***  
## I(Year)       0.005773    0.001284   4.495 8.56e-05 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.07346 on 32 degrees of freedom  
## Multiple R-squared:  0.3871, Adjusted R-squared:  0.3679   
## F-statistic: 20.21 on 1 and 32 DF,  p-value: 8.56e-05
```

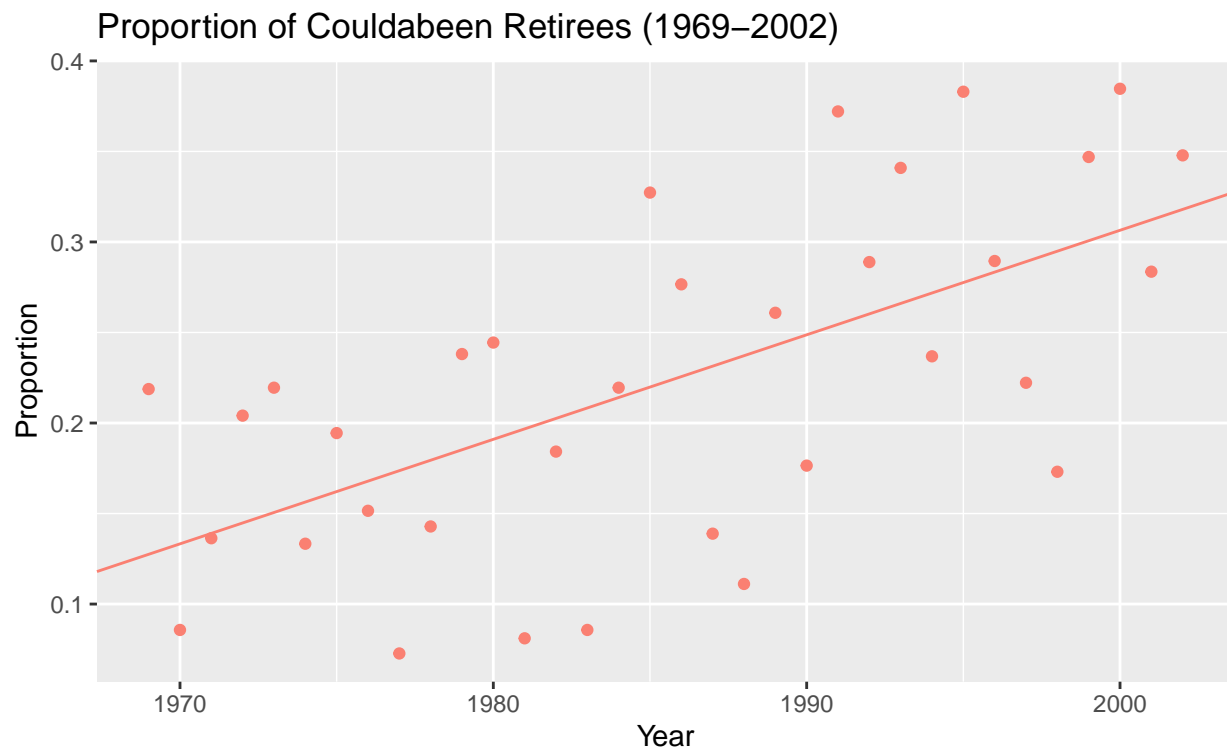


Figure 1: Proportion of Retirees who were Couldabeens prior to the implementation of the Luxury Tax

Couldabeens: Post-rule Era (2003-2018)

```
##
## Call:
## lm(formula = prop ~ I(Year), data = couldabeens_post)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.049689 -0.024453  0.001825  0.018410  0.049237
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.987717   3.481582  -1.145   0.271
## I(Year)      0.002064   0.001732   1.192   0.253
##
## Residual standard error: 0.03193 on 14 degrees of freedom
## Multiple R-squared:  0.09209,    Adjusted R-squared:  0.02724
## F-statistic:  1.42 on 1 and 14 DF,  p-value: 0.2532
```

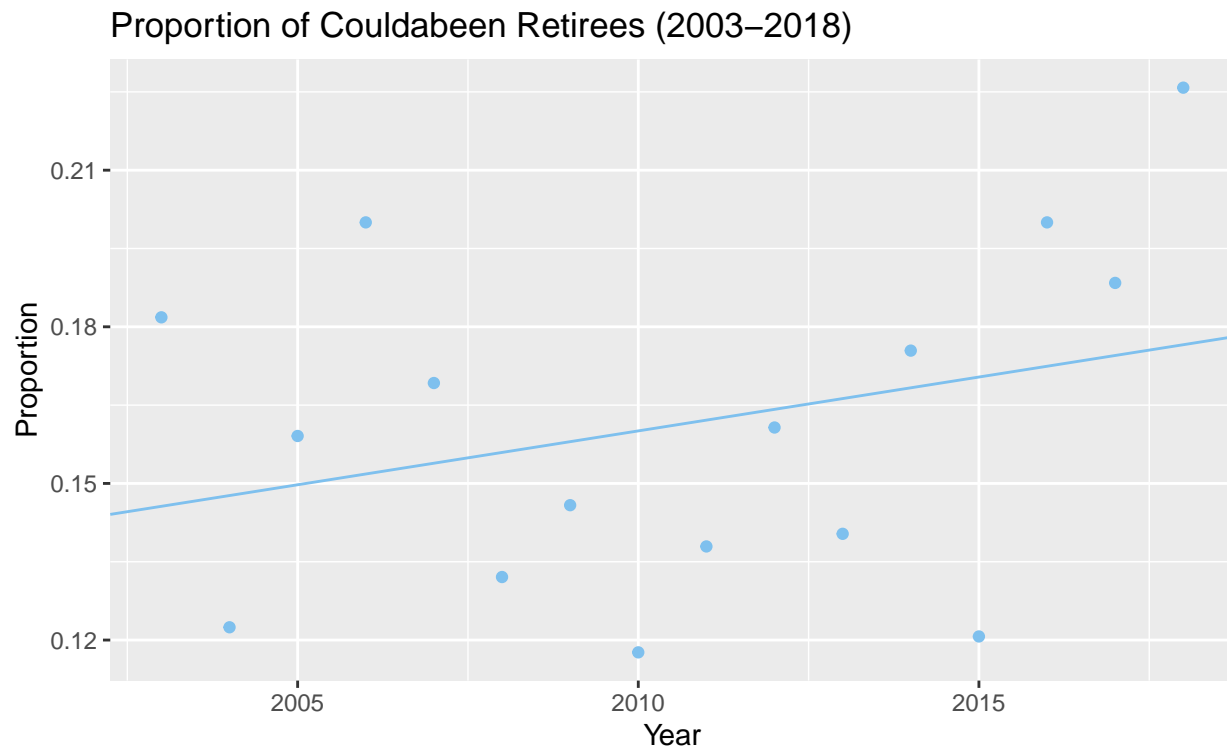


Figure 2: Proportion of Retirees who were Couldabeens after the implementation of the Luxury Tax

Discussion

Code Appendix

References

add source: baseball-reference.com