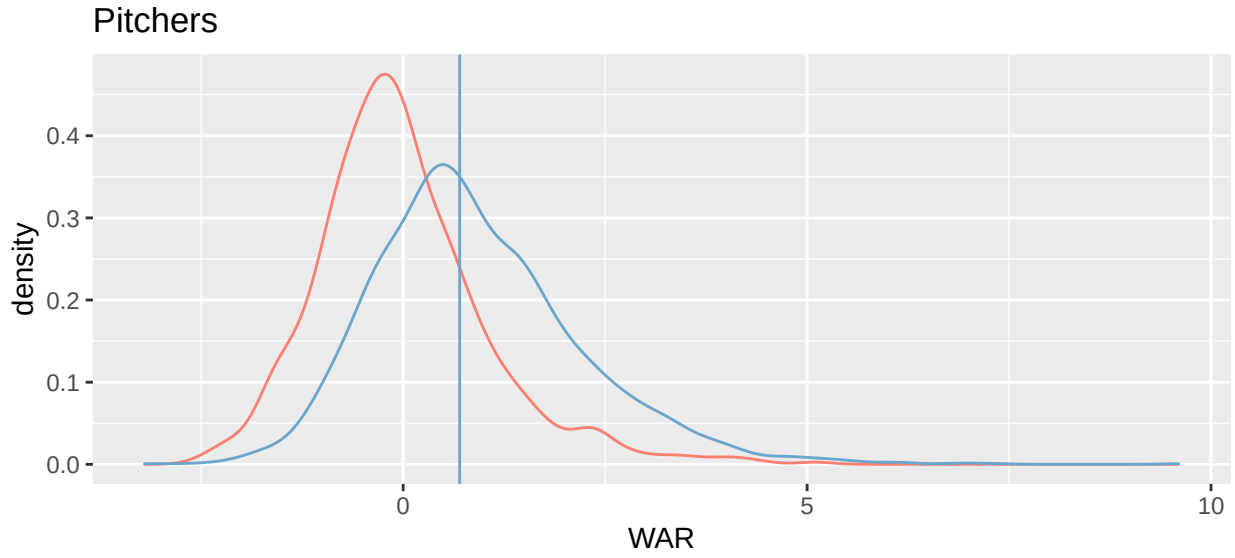
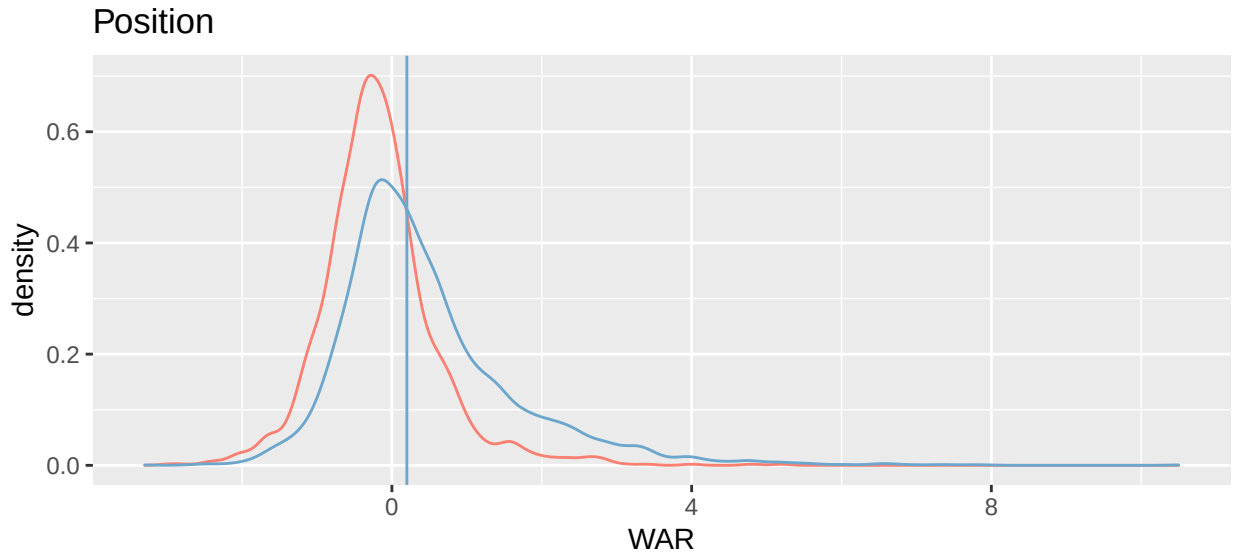
Modeling

Group 6

The Couldabeen Classification Problem





Counting Couldabeens

```
#==================================#
#        Counting: Couldabeens     #
#==================================#
# Combine the threshold-classified retiree datasets
retirees <- rbind(pit_ret,pos_ret)
# Count couldabeens
couldabeens <- count_cbns(retirees)
```

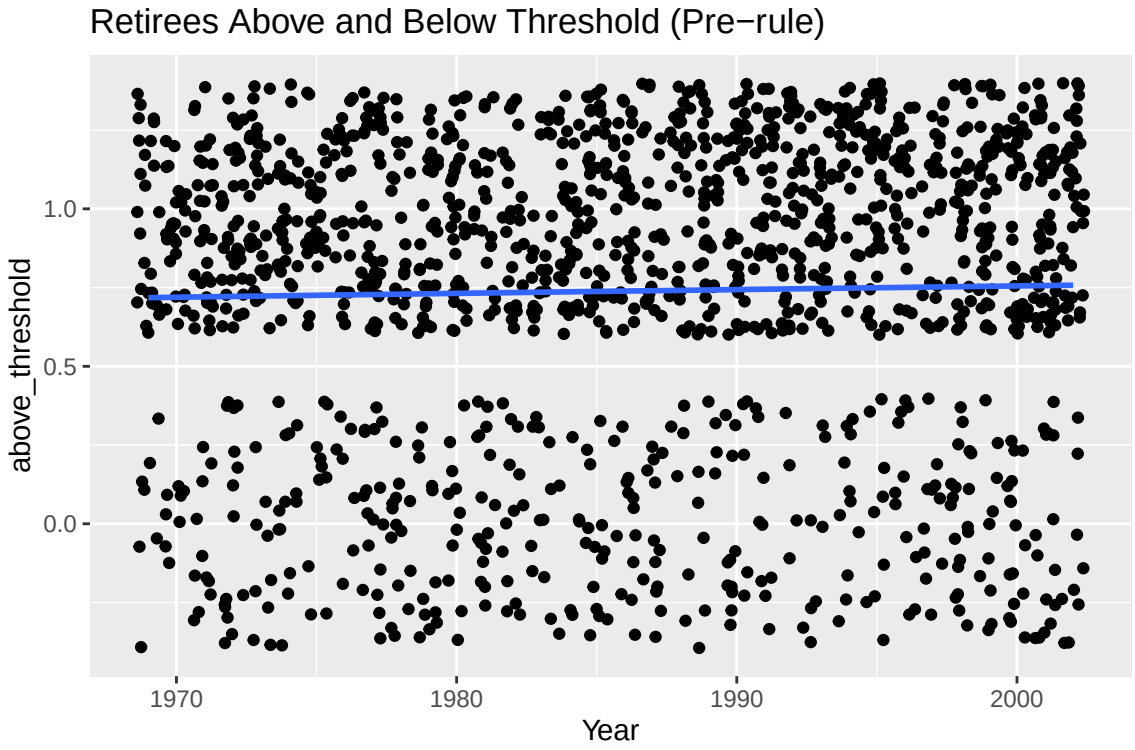
Our retirees dataframe looks like this:

```
##    WAR Year above_threshold
## 1  1.8 1972            TRUE
## 2  0.1 1974            TRUE
## 3  0.3 1976            TRUE
## 4 -0.5 1977           FALSE
## 5  0.4 1977            TRUE
## 6 -1.8 1974           FALSE
```

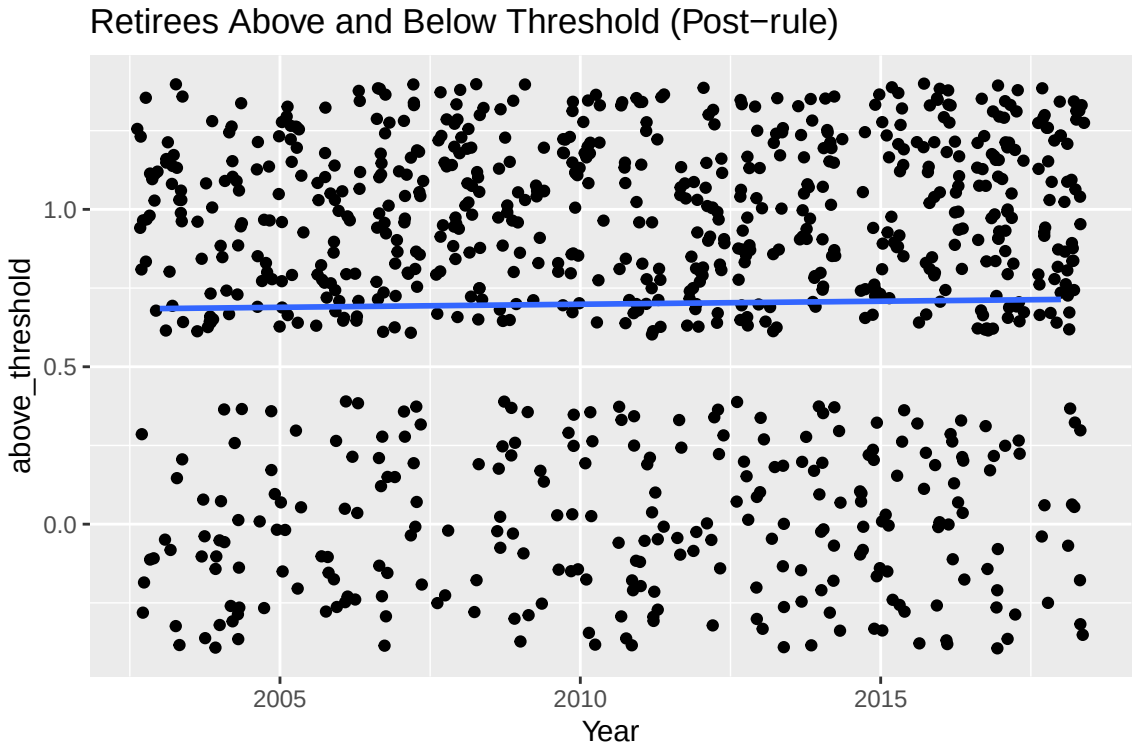
Our couldabeens dataframe looks like this:

```
## # A tibble: 6 x 2
##    Year  cbns
##   <dbl> <int>
## 1  1969    25
## 2  1970    27
## 3  1971    33
## 4  1972    34
## 5  1973    32
## 6  1974    32
```

First Look: A Logistic Model




```
##
## Call:
## glm(formula = above_threshold ~ Year, family = "binomial", data = dataset)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6833  -1.5969   0.7656   0.7884   0.8138
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -11.248491  11.890601  -0.946    0.344
## Year          0.006188   0.005987   1.033    0.301
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1689.1  on 1470  degrees of freedom
## Residual deviance: 1688.1  on 1469  degrees of freedom
## AIC: 1692.1
##
## Number of Fisher Scoring iterations: 4
```




```
##
## Call:
## glm(formula = above_threshold ~ Year, family = "binomial", data = dataset)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5819  -1.5280   0.8307   0.8535   0.8701
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -17.638285  32.211759  -0.548    0.584
## Year          0.009193   0.016019   0.574    0.566
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1075.8  on 880  degrees of freedom
## Residual deviance: 1075.5  on 879  degrees of freedom
## AIC: 1079.5
##
## Number of Fisher Scoring iterations: 4
```

Computing Retiree Proportions

```
#==================================#
#      Proportions: Couldabeens    #
#==================================#
# Find number of retirees by year
num_retirees <- total_retirees_by_yr(df_pit_ret, df_pos_ret)
num_retirees <- data.frame(retirees = num_retirees$retirees)
# Append number of retirees that year
couldabeens <- cbind(couldabeens, num_retirees)
# Find proportion of couldabeens : retirees
couldabeens <- couldabeens %>% mutate(prop = cbns/retirees)
```

Here is what the proportion-appended couldabeen dataframe looks like:

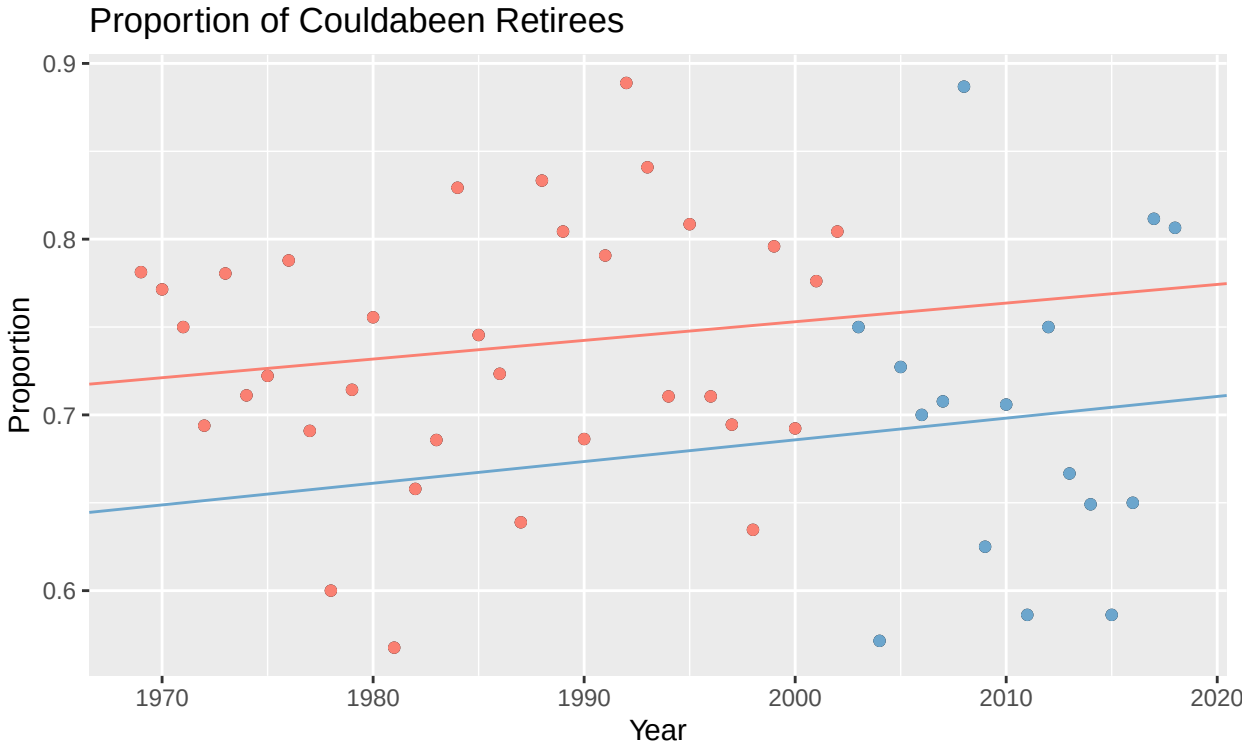
```
##   Year cbns retirees      prop
## 1 1969   25       32 0.7812500
## 2 1970   27       35 0.7714286
## 3 1971   33       44 0.7500000
## 4 1972   34       49 0.6938776
## 5 1973   32       41 0.7804878
## 6 1974   32       45 0.7111111
```

Year as Predictor: Linear Modeling

```
#==================#
#     Modeling     #
#==================#
# 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 <- linear_model(couldabeens_pre)
coefs_pre <- model_pre$coefficients
# Obtain linear model for post-rule years
model_post <- linear_model(couldabeens_post)
coefs_post <- model_post$coefficients
# Obtain linear model for all years
model_comp <- linear_model(couldabeens)
coefs_comp <- model_comp$coefficients
```

Couldabeens: A Comprehensive Look

```
##
## Call:
## lm(formula = prop ~ I(Year), data = dataset)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.165628 -0.045789 -0.008786  0.054675  0.170899
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.0026801  1.5512916   1.291    0.203
## I(Year)     -0.0006408  0.0007782  -0.824    0.414
##
## Residual standard error: 0.0794 on 48 degrees of freedom
## Multiple R-squared:  0.01393,    Adjusted R-squared:  -0.006611
## F-statistic: 0.6782 on 1 and 48 DF,  p-value: 0.4143
```



Couldabeens: Pre-rule Era (1969-2002)

```
##
## Call:
## lm(formula = prop ~ I(Year), data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16527 -0.04650  0.00206  0.05467  0.14437
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.370105   2.515373  -0.545    0.590
## I(Year)      0.001062   0.001267   0.838    0.408
##
## Residual standard error: 0.07247 on 32 degrees of freedom
## Multiple R-squared:  0.02147,    Adjusted R-squared:  -0.009108
## F-statistic: 0.7022 on 1 and 32 DF,  p-value: 0.4083
```

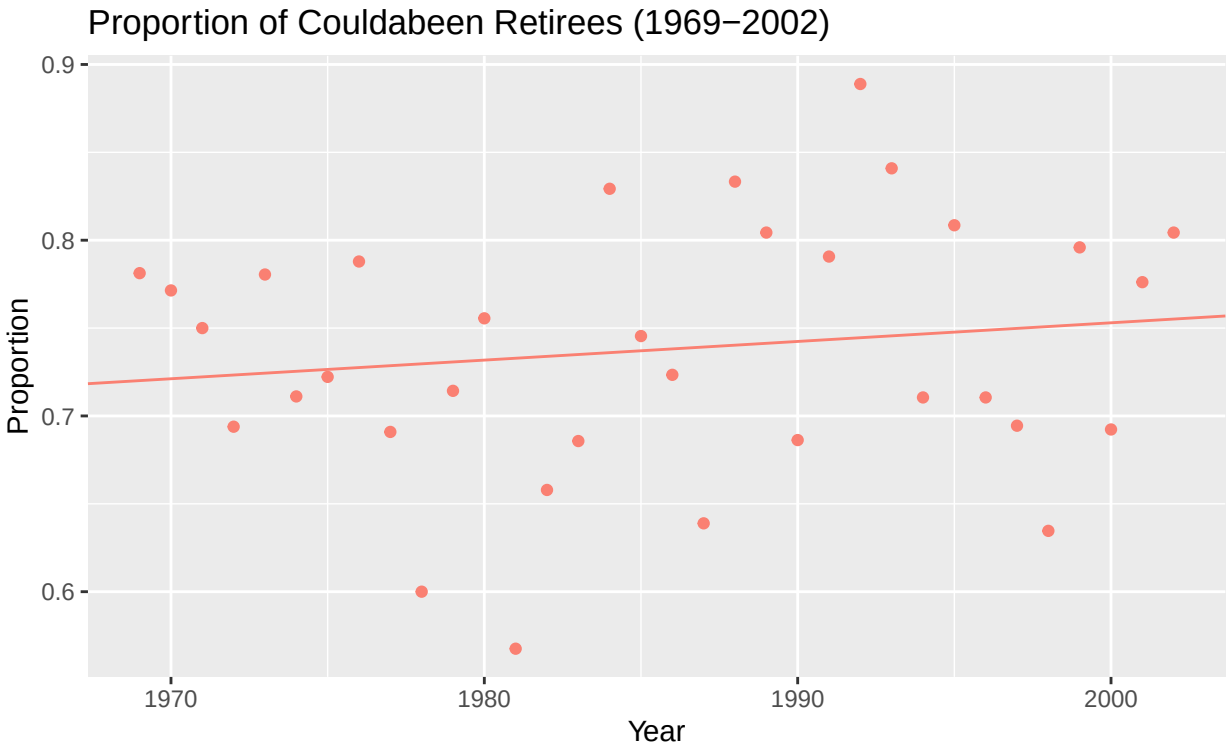


Figure 1: Proportion of Retirees who were Coulabeens prior to the implementation of the Luxury Tax

Couldabeens: Post-rule Era (2003-2018)

```
##
## Call:
## lm(formula = prop ~ I(Year), data = dataset)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.119317 -0.059649  0.007258  0.052156  0.191109
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.783306  10.019728  -0.178    0.861
## I(Year)      0.001235   0.004984   0.248    0.808
##
## Residual standard error: 0.09189 on 14 degrees of freedom
## Multiple R-squared:  0.004364,   Adjusted R-squared:  -0.06675
## F-statistic: 0.06136 on 1 and 14 DF,  p-value: 0.8079
```

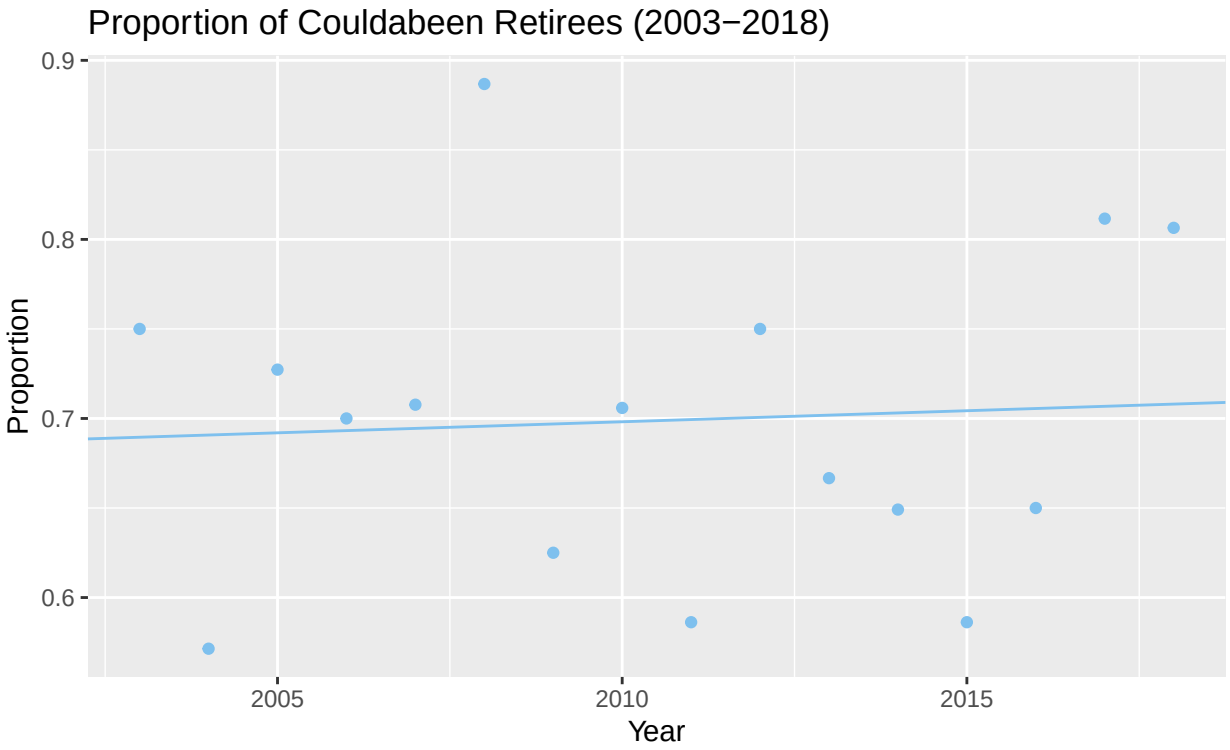


Figure 2: Proportion of Retirees who were Coulabeens after the implementation of the Luxury Tax