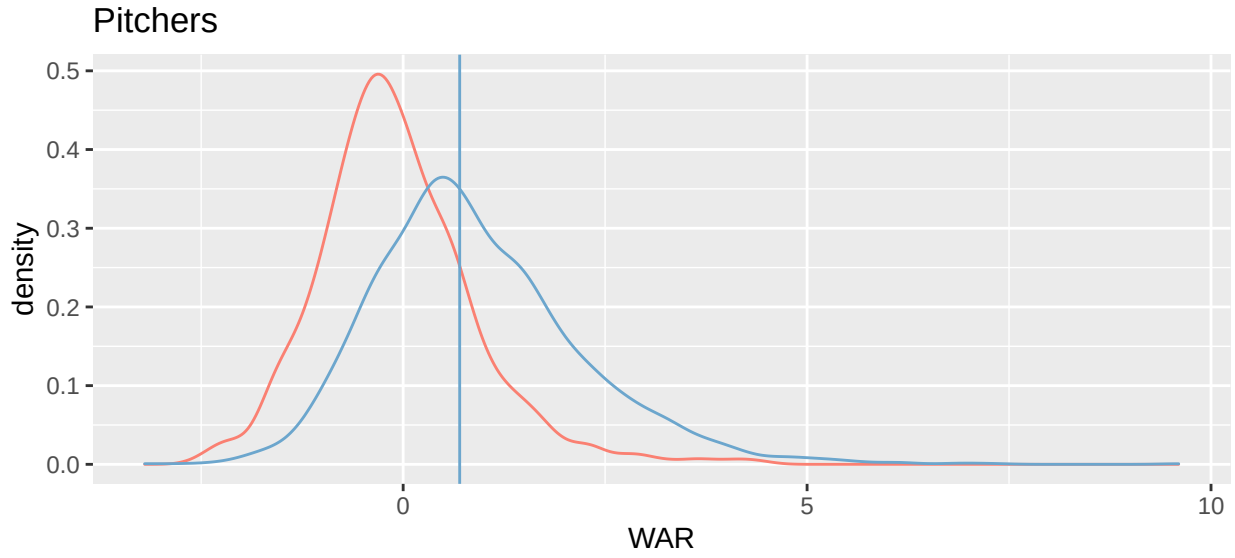
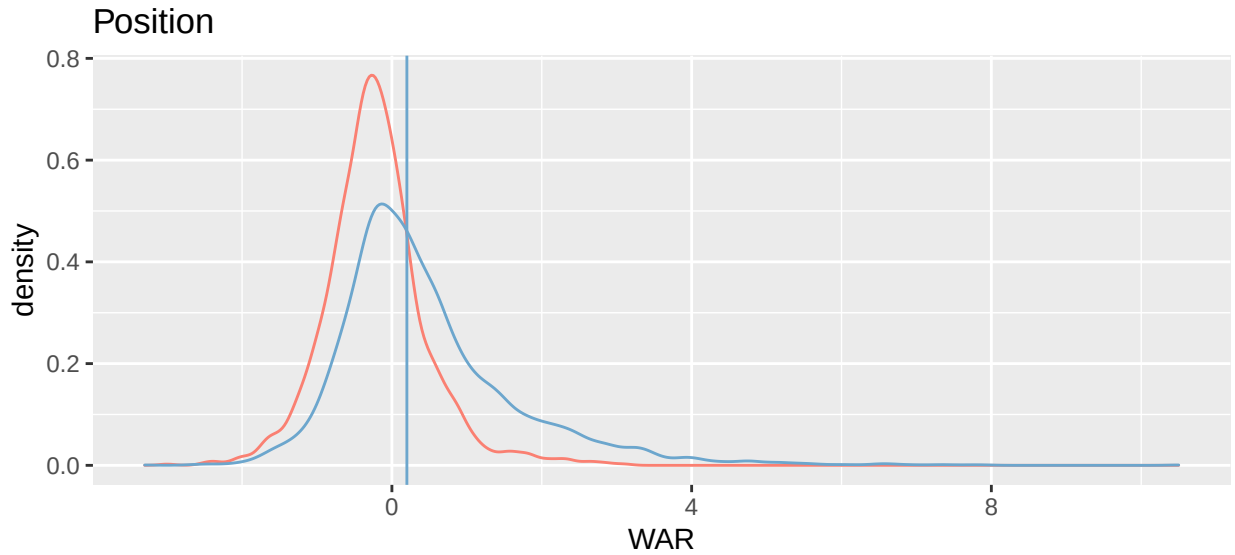
Modeling

Group 6

The Couldabeen Classification Problem





```r
#=====================#
#      Couldabeens     #
#=====================#
# Get the threshold-classified retiree dataset of pitchers
retirees <- rbind(pit_ret,pos_ret)
couldabeens <- count_cbns(retirees)
# Find number of retirees that year
num_retirees <- df_tot_ret[order(df_tot_ret$Year),]
num_retirees <- data.frame(retirees = num_retirees$retirees)
# Append number of retirees that year
couldabeens <- cbind(couldabeens, num_retirees)
# Find proportion of couldabeens : retirees
couldabeens <- couldabeens %>% mutate(prop = cbns/retirees)
```

```r
#==================#
#     Analysis     #
#==================#
# Partition dataset into years before and after rule
couldabeens_pre <- couldabeens %>% filter(Year <= 2002)
couldabeens_post <- couldabeens %>% filter(Year > 2002)
# Obtain linear model for pre-rule years
model_pre <- lm(formula = prop ~ I(Year), data = couldabeens_pre)
coefs_pre <- model_pre$coefficients
# Obtain linear model for post-rule years
model_post <- lm(formula = prop ~ I(Year), data = couldabeens_post)
coefs_post <- model_post$coefficients
```

Couldabeens: Pre-rule Era (1969-2002)

```
##
## Call:
## lm(formula = prop ~ I(Year), data = couldabeens_pre)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.151821 -0.076389  0.001651  0.062593  0.166725
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.310781   3.013385   1.762   0.0876 .
## I(Year)     -0.002436   0.001518  -1.605   0.1184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08682 on 32 degrees of freedom
## Multiple R-squared:  0.07449,    Adjusted R-squared:  0.04556
## F-statistic: 2.575 on 1 and 32 DF,  p-value: 0.1184
```

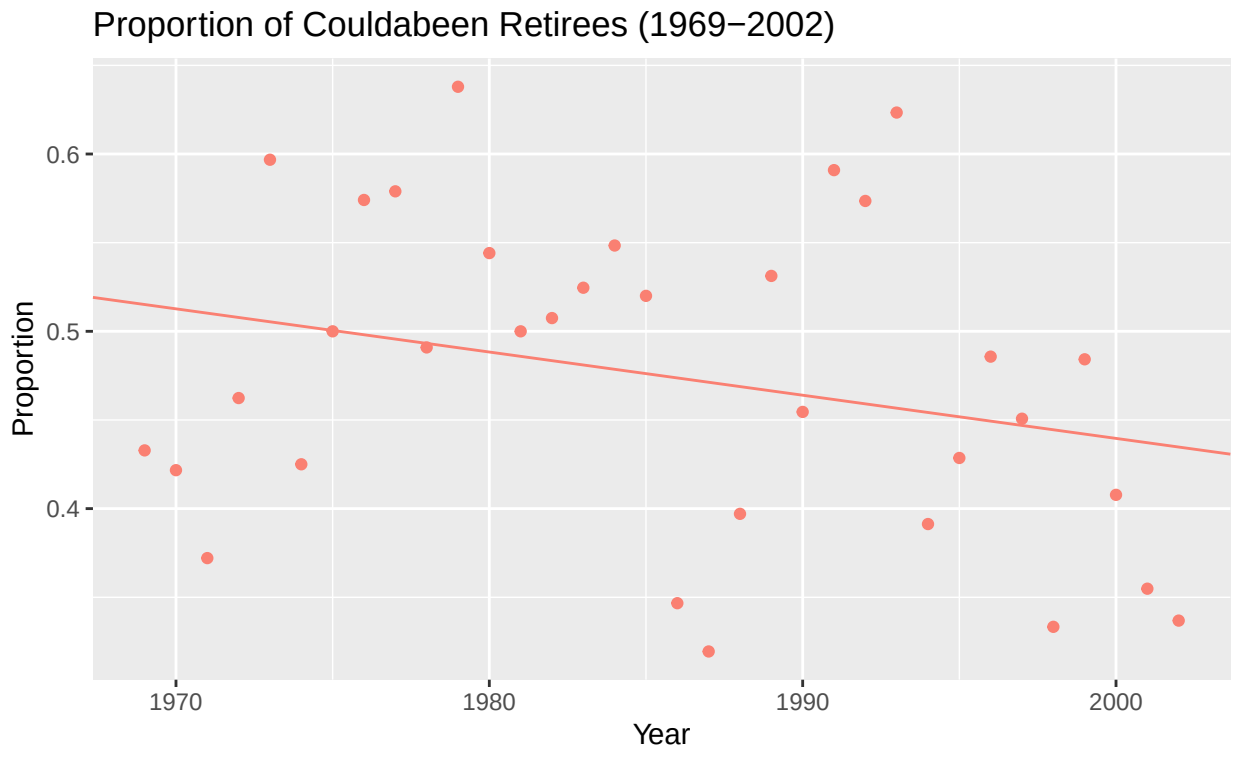


Figure 1: Proportion of Retirees who were Coulabeens prior to the implementation of the Luxury Tax

Couldabeens: Post-rule Era (2003-2018)

```
##
## Call:
## lm(formula = prop ~ I(Year), data = couldabeens_post)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.074697 -0.026968 -0.000475  0.022815  0.092860
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -9.689947   5.138559  -1.886   0.0803 .
## I(Year)      0.004965   0.002556   1.943   0.0725 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04713 on 14 degrees of freedom
## Multiple R-squared:  0.2123, Adjusted R-squared:  0.1561
## F-statistic: 3.774 on 1 and 14 DF,  p-value: 0.07246
```

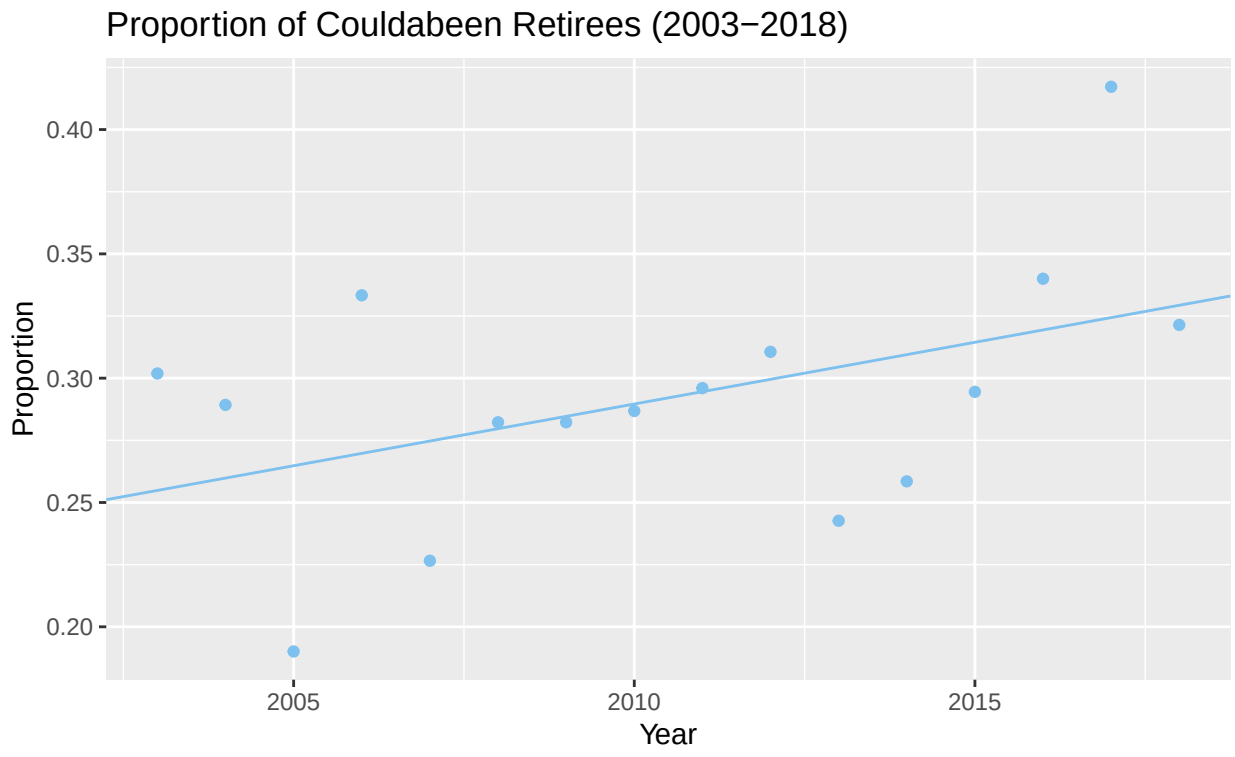
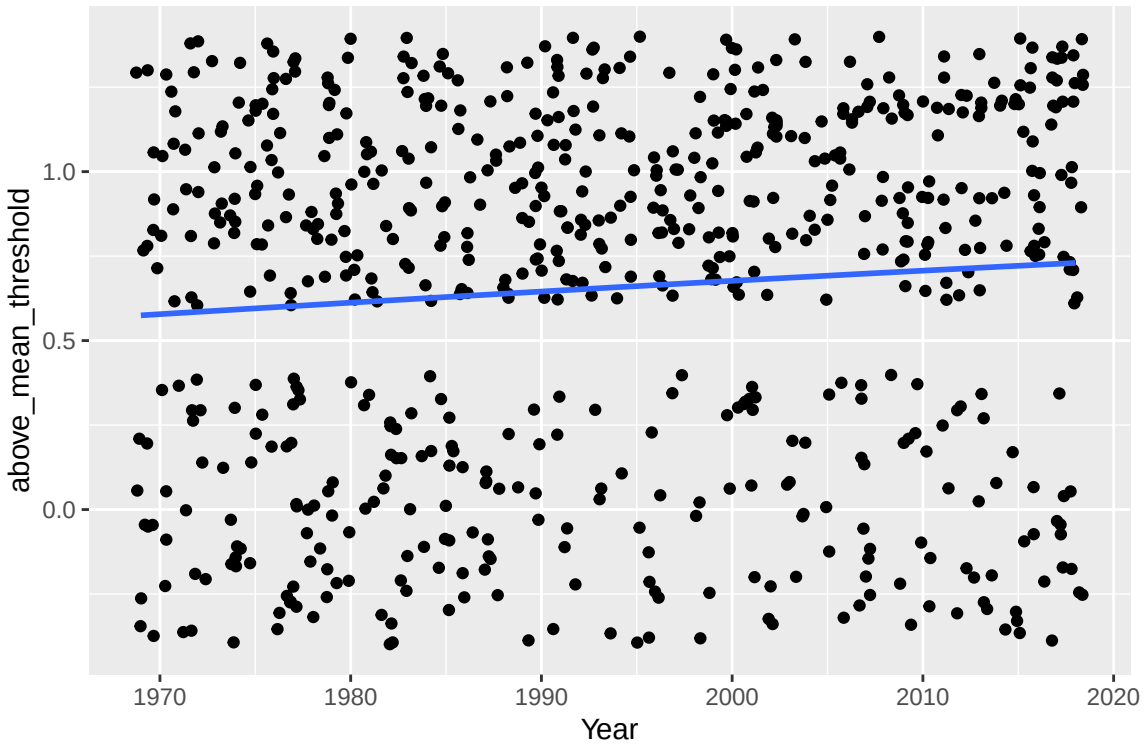


Figure 2: Proportion of Retirees who were Coulabeens after the implementation of the Luxury Tax




```
##
## Call:
## glm(formula = above_mean_threshold ~ Year, family = "binomial",
##     data = pit_ret)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6177  -1.3702   0.8428   0.9362   1.0525
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -27.515639  11.239649  -2.448   0.0144 *
## Year          0.014127   0.005641   2.504   0.0123 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 865.29  on 670  degrees of freedom
## Residual deviance: 858.95  on 669  degrees of freedom
## AIC: 862.95
##
## Number of Fisher Scoring iterations: 4
```