# 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)
# 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
## -0.183410 -0.028906 -0.008431  0.047255  0.162481
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.876935   2.489627   2.361   0.0245 *
## I(Year)     -0.002615   0.001254  -2.086   0.0451 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07173 on 32 degrees of freedom
## Multiple R-squared:  0.1197, Adjusted R-squared:  0.09215
## F-statistic:  4.35 on 1 and 32 DF,  p-value: 0.04508
```

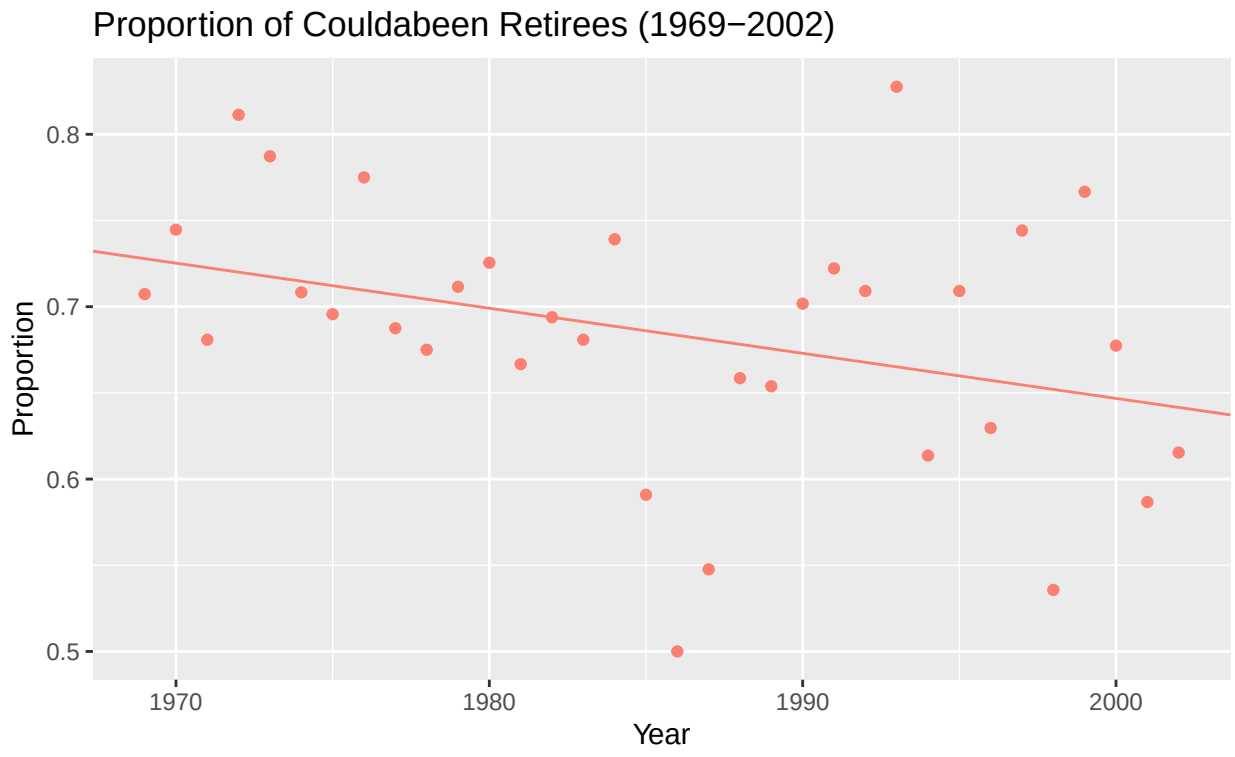



Figure 1: Proportion of Retirees who were Coulabeens prior to the implementation of the Luxury Tax

### Couldabeens: Post-rule Era (2003-2018)

```
##
## Call:
## lm(formula = prop ~ I(Year), data = couldabeens_post)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.151527 -0.034281  0.009201  0.036678  0.129404
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.124022   7.778200  -2.587   0.0215 *
## I(Year)       0.010312   0.003869   2.665   0.0185 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07134 on 14 degrees of freedom
## Multiple R-squared:  0.3366, Adjusted R-squared:  0.2892
## F-statistic: 7.104 on 1 and 14 DF,  p-value: 0.01847
```

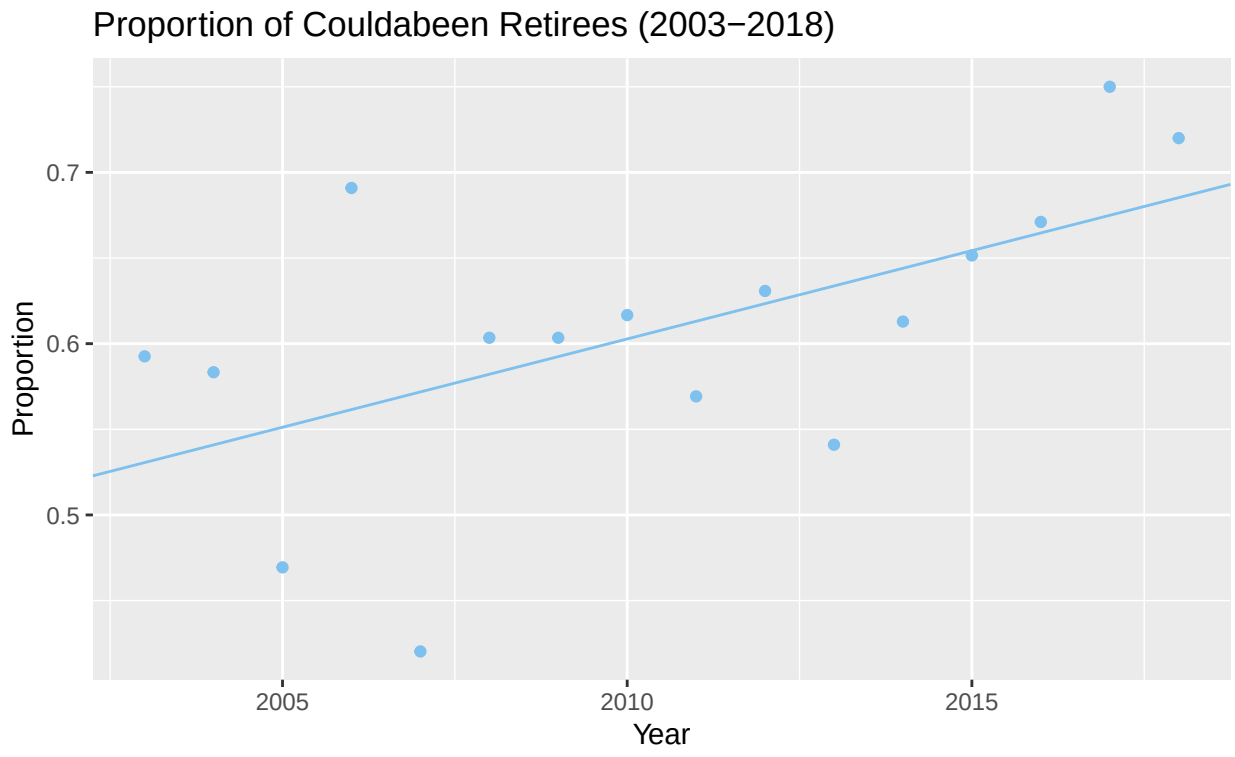
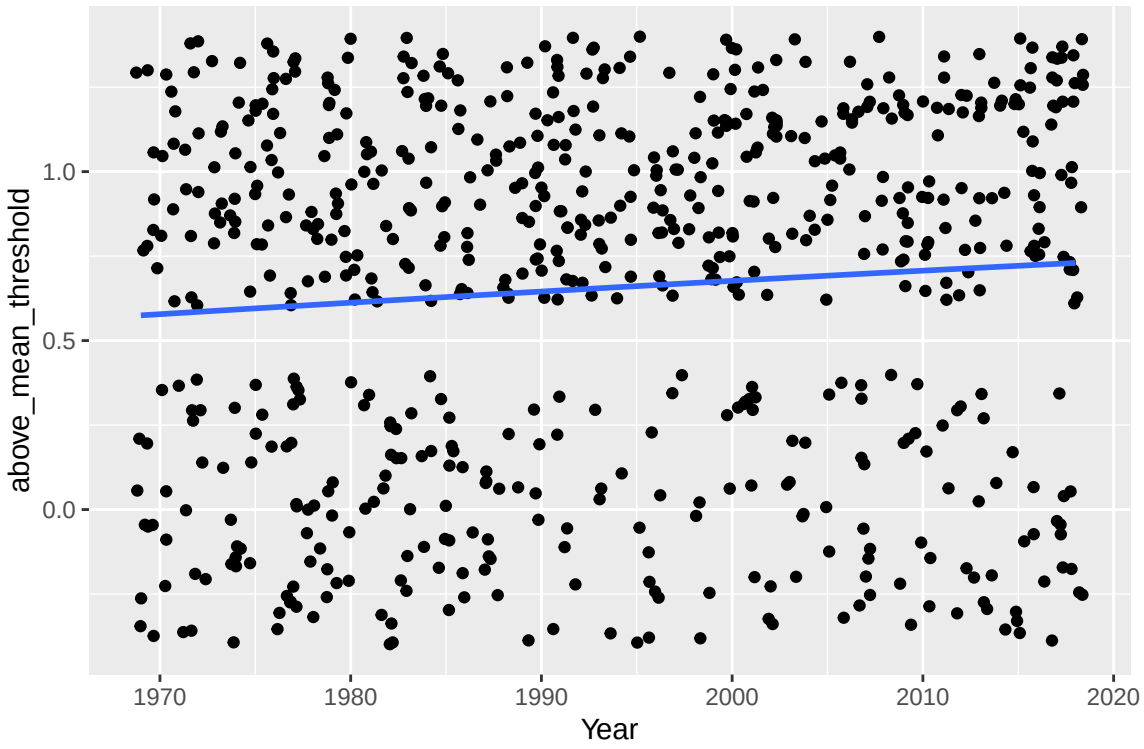



Figure 2: Proportion of Retirees who were Coulabeens after the implementation of the Luxury Tax




```
##
## 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
```