Modeling

Group 6

Thresholds

Smooth thresholds

Couldabeen Proportions

```
#==================================#
#        Counting: Couldabeens     #
#==================================#
# Combine the threshold-classified retiree datasets
retirees <- rbind(pit_ret,pos_ret)
# Count couldabeens
couldabeens <- count_cbns(retirees)
#==================================#
#      Proportions: Couldabeens    #
#==================================#
# Find number of retirees by year
num_retirees <- total_retirees_by_yr(pit_ret, 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)
```

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
```

Simpson's Paradox

Couldabeens: Pre-rule Era (1969-2002)

```
##
## Call:
## lm(formula = prop ~ I(Year), data = dataset)
##
## 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
```

Couldabeens: Post-rule Era (2003-2018)

```
##
## Call:
## lm(formula = prop ~ I(Year), data = dataset)
##
## 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
```