

Load necessary packages

Load data and extract classes with both CR and OR

Old data

```
files = list.files("C:/Users/Cole/Documents/DATA/PLIC_DATA/Raw_Surveys/BC (Before Cole)/",
                  recursive = TRUE, full.names = TRUE)

files <- files[!grepl('*UBC119_RAW_POST*', files)]

read.old.file <- function(file){
  df <- read.csv(file)
  df <- df[-1, ]
  df$Survey <- ifelse((df$Q1b != '') | (df$Q4b != ''), 'F', 'C')
  df[df$Q4a == '', 'Q4a'] <- df[df$Q4a == '', 'Q4a.1']
  df <- df %>%
    select(Survey, Q4a)

  Name <- strsplit(file, '/')[[1]]
  df$Class_ID <- strsplit(Name[length(Name)], '\\.')[[1]][1]
  return(df)
}

df.old = ldply(files, read.old.file)
```

New data

```
df.new.post <- read.csv('C:/Users/Cole/Documents/DATA/PLIC_DATA/Collective_Surveys/POST_Valid/POST_Valid_Collec.csv')
select(Survey, Q4a, Class_ID)
Class_IDs <- unique(df.new.post[df.new.post$Survey == 'F', 'Class_ID'])
df.new.post <- subset(df.new.post, df.new.post$Class_ID %in% Class_IDs) %>%
  mutate(Class_ID = paste(Class_ID, 'post', sep = '.'))

df.new.pre <- read.csv('C:/Users/Cole/Documents/DATA/PLIC_DATA/Collective_Surveys/PRE_Valid/PRE_Valid_Collec.csv')
select(Survey, Q4a, Class_ID)
Class_IDs <- unique(df.new.pre[df.new.pre$Survey == 'F', 'Class_ID'])
df.new.pre <- subset(df.new.pre, df.new.pre$Class_ID %in% Class_IDs) %>%
  mutate(Class_ID = paste(Class_ID, 'pre', sep = '.'))

df <- rbind(df.old, df.new.post, df.new.pre) %>%
  mutate(Group = as.factor(case_when(
    Q4a == '1' ~ '1',
    Q4a == '2' ~ '2',
    Q4a == '3' ~ 'B',
    TRUE ~ ''
  ))) %>%
  filter(Group != '' & Survey != '') %>%
  select(Survey, Group, Class_ID)
```

# Analysis

## Overall differences

```
df %>%
  group_by(Survey, Group) %>%
  summarize(N = n()) %>%
  mutate(freq = N / sum(N))

## `summarise()` regrouping output by 'Survey' (override with `.groups` argument)

## # A tibble: 6 x 4
## # Groups:   Survey [2]
##   Survey Group      N freq
##   <chr>   <fct> <int> <dbl>
## 1 C       1       922 0.468
## 2 C       2       804 0.409
## 3 C       B       242 0.123
## 4 F       1       462 0.439
## 5 F       2       439 0.417
## 6 F       B       152 0.144

chisq.test(df$Survey, df$Group)

##
## Pearson's Chi-squared test
##
## data:  df$Survey and df$Group
## X-squared = 3.8466, df = 2, p-value = 0.1461
```

## Multilevel models

We fit separate Multilevel logit models because packages for multinomial models can't handle random effects and packages for mixed effects can't handle multinomial responses. I also ran a chi-squared test to check with accounting for nesting.

```
summary(glmer(Group ~ Survey + (1 | Class_ID), df %>%
  filter(Group != 'B'), family = binomial(link = 'logit'))

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Group ~ Survey + (1 | Class_ID)
## Data: df %>% filter(Group != "B")
##
##      AIC      BIC   logLik deviance df.resid
## 3614.1   3631.7 -1804.1   3608.1     2624
##
## Scaled residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -1.4570 -0.9287 -0.7667  1.0694  1.3043
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   Class_ID (Intercept) 0.09225  0.3037
## Number of obs: 2627, groups:  Class_ID, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.08399    0.07856  -1.069   0.285
## SurveyF      0.05537    0.08847   0.626   0.531
##
## Correlation of Fixed Effects:
##          (Intr)
## SurveyF -0.369
```

```
summary(glmmer(Group ~ Survey + (1 | Class_ID), df %>%
  filter(Group != '2'), family = binomial(link = 'logit')))
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Group ~ Survey + (1 | Class_ID)
## Data: df %>% filter(Group != "2")
##
##      AIC      BIC   logLik deviance df.resid
## 1862.7   1879.1   -928.3   1856.7     1775
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -0.7738 -0.5523 -0.4903 -0.3969  2.5193
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   Class_ID (Intercept) 0.1508  0.3884
## Number of obs: 1778, groups:  Class_ID, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.3363    0.1085 -12.318 <2e-16 ***
## SurveyF      0.1474    0.1286   1.146   0.252
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## SurveyF -0.397
```

```
summary(glmmer(Group ~ Survey + (1 | Class_ID), df %>%
  filter(Group != '1'), family = binomial(link = 'logit')))
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
```

```

## Approximation) [glmerMod]
## Family: binomial (logit)
## Formula: Group ~ Survey + (1 | Class_ID)
## Data: df %>% filter(Group != "1")
##
##      AIC      BIC    logLik deviance df.resid
## 1795.6    1811.8   -894.8    1789.6     1634
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.6976 -0.6037 -0.5147 -0.4071  2.4562
##
## Random effects:
## Groups Name Variance Std.Dev.
## Class_ID (Intercept) 0.1362  0.369
## Number of obs: 1637, groups: Class_ID, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.2429     0.1063 -11.690  <2e-16 ***
## SurveyF       0.0864     0.1273  0.679    0.497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr)
## SurveyF -0.400

```