Stats 209 Final - EDA

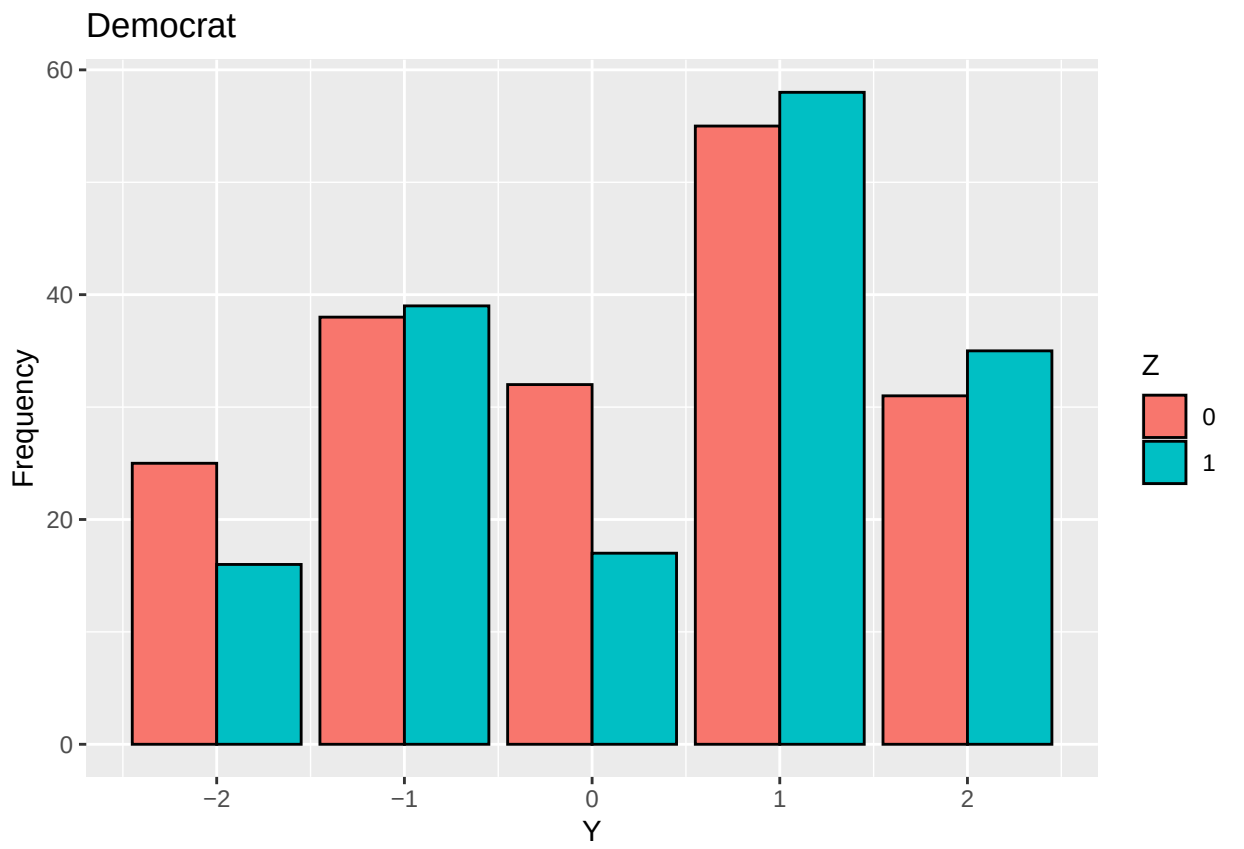
Andri, Eden, Veronica

11/8/2023

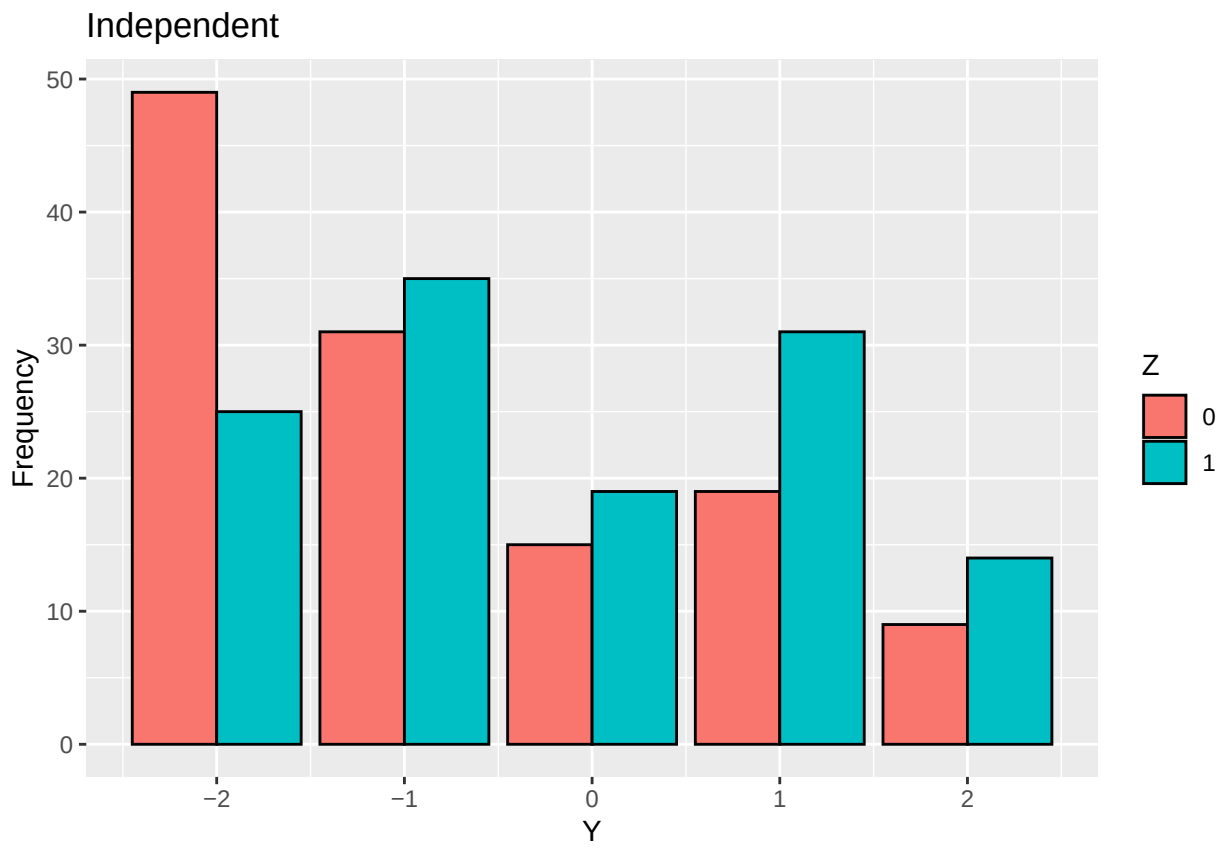
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(corrplot)
source('dataProcessing.R')
```

```
dlexp = loadData()
```

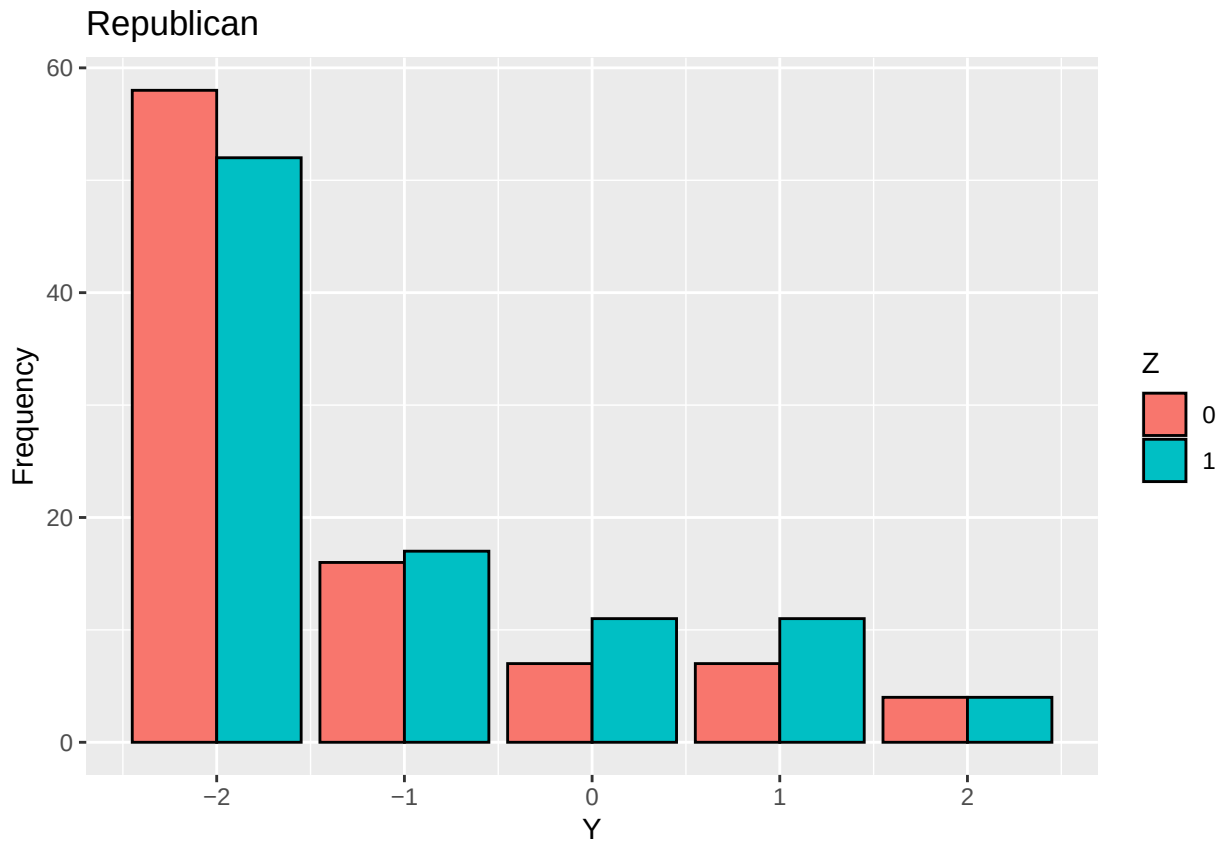
```
party_names = c('1' = "Democrat", '2' = "Independent", '3' = "Republican")
dat = dlexp %>%
   filter(party.cues == 1) %>%
   group_by(across(all_of(c("pid3", "Z", "Y")))) %>%
   summarize(Frequency = n())

dat$Z = as.factor(dat$Z)

# barplot showing responses for treatment/control for selected political party
party_bar = function(pid){
  dat_ = filter(dat, pid3 == pid)
  ggplot(data=dat_, aes(x=Y, y=Frequency, fill=Z)) +
    geom_bar(stat="identity",  position="dodge", colour = "black") + ggtitle(party_names[pid])

}

party_bar(1)
```



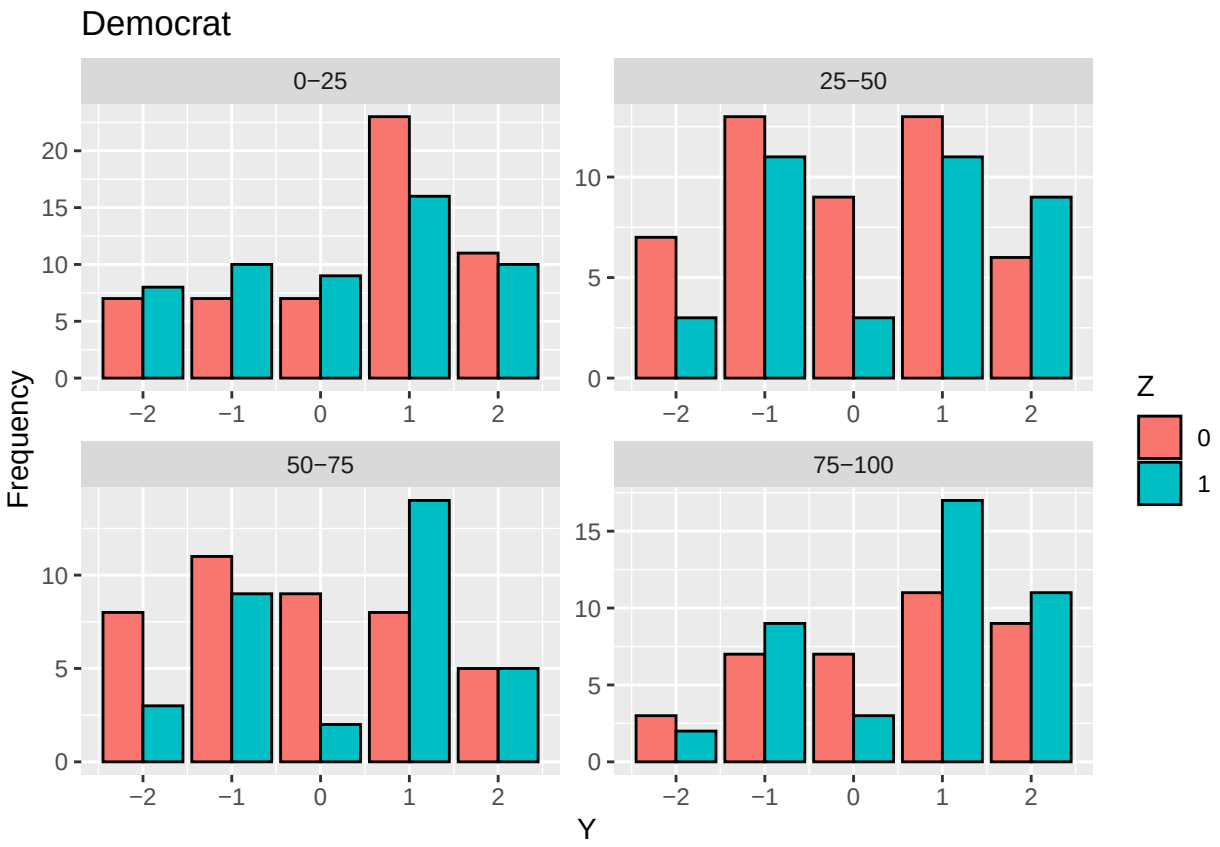

```
party_bar(2)
```



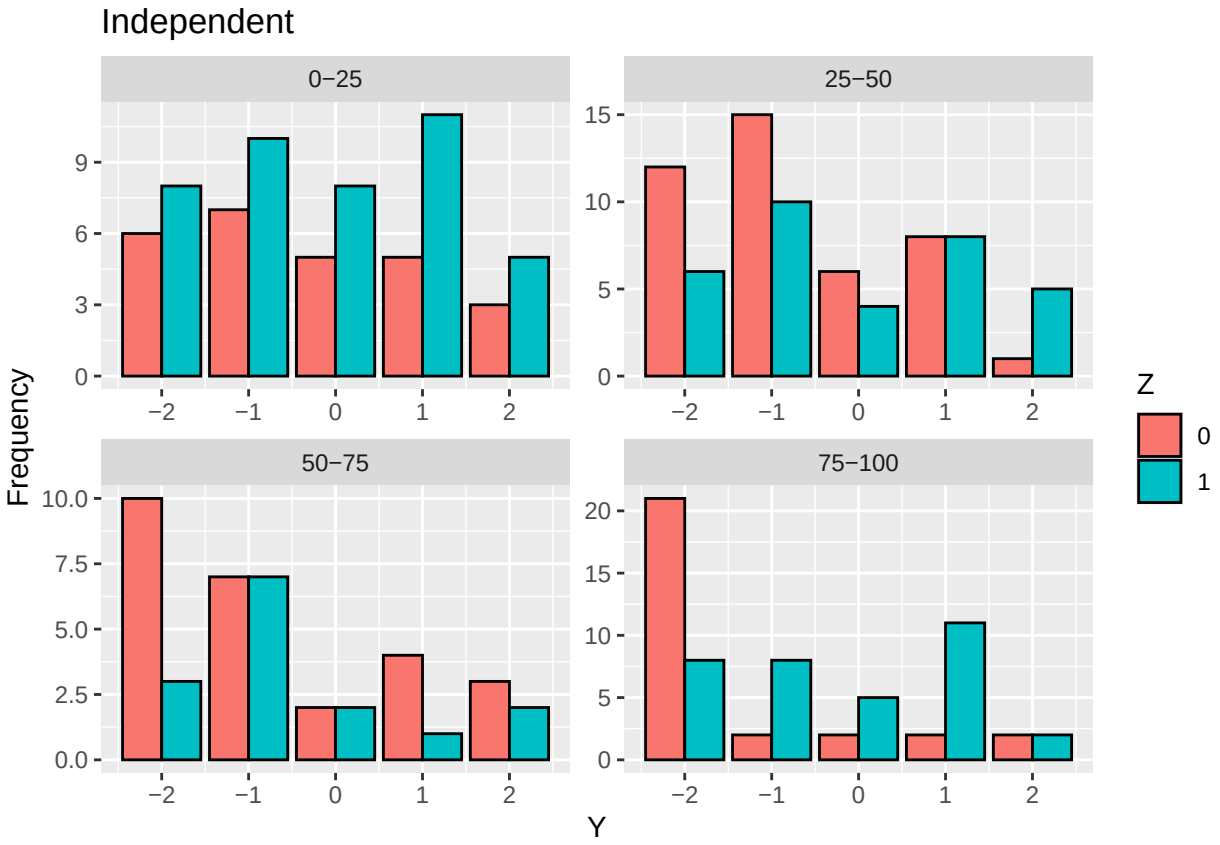

```
party_bar(3)
```



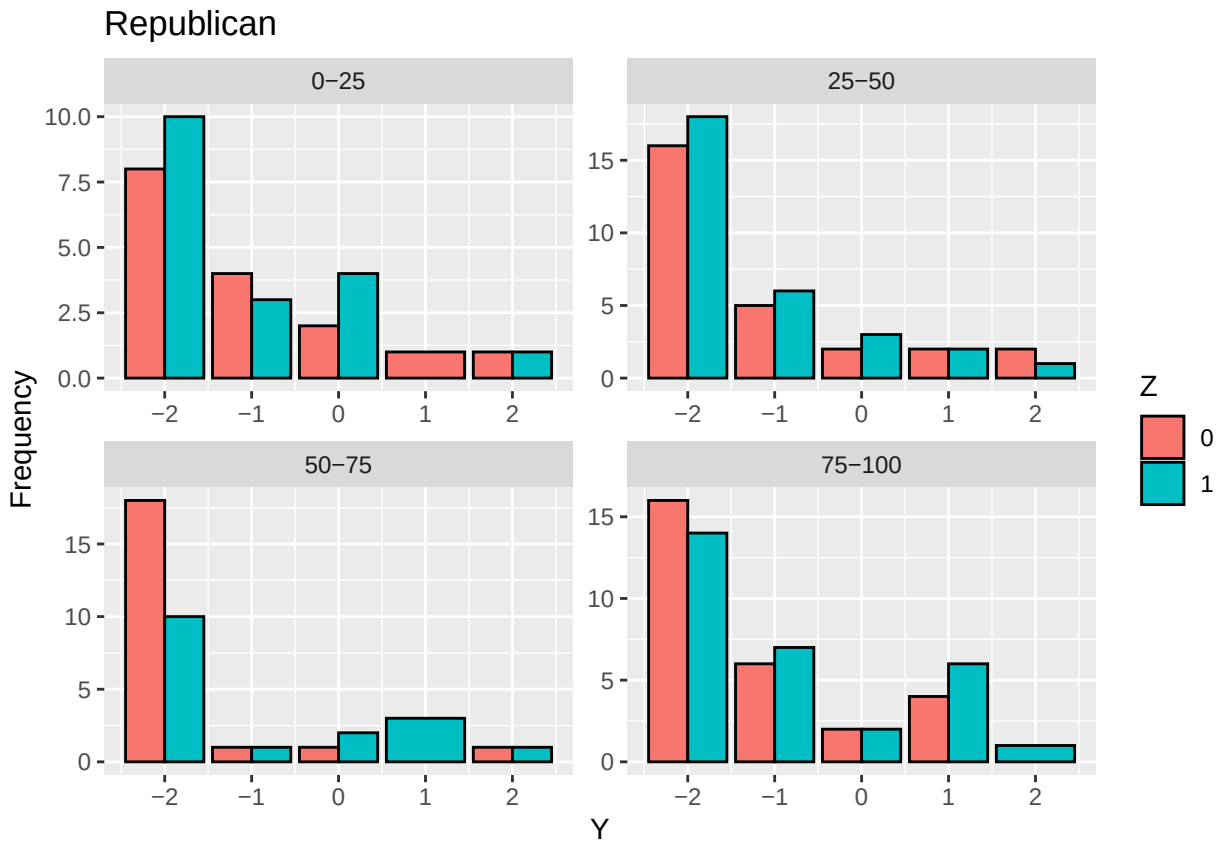

```
datInc = dlexp %>%
   filter(party.cues == 1) %>%
   group_by(across(all_of(c("pid3", "incomeQuant", "Z", "Y")))) %>%
   summarize(Frequency = n())

datInc$Z = as.factor(datInc$Z)

# show breakdown for responses by income quantile, per group
income_plot = function(pid){
  dat = filter(datInc, pid3 == pid)
  ggplot(data=dat, aes(x=Y, y=Frequency, fill=Z)) +
    geom_bar(stat="identity",  position="dodge", colour = "black") +
    facet_wrap(~incomeQuant, scales = "free") +
    ggtitle(party_names[pid])

}

income_plot(1)
```



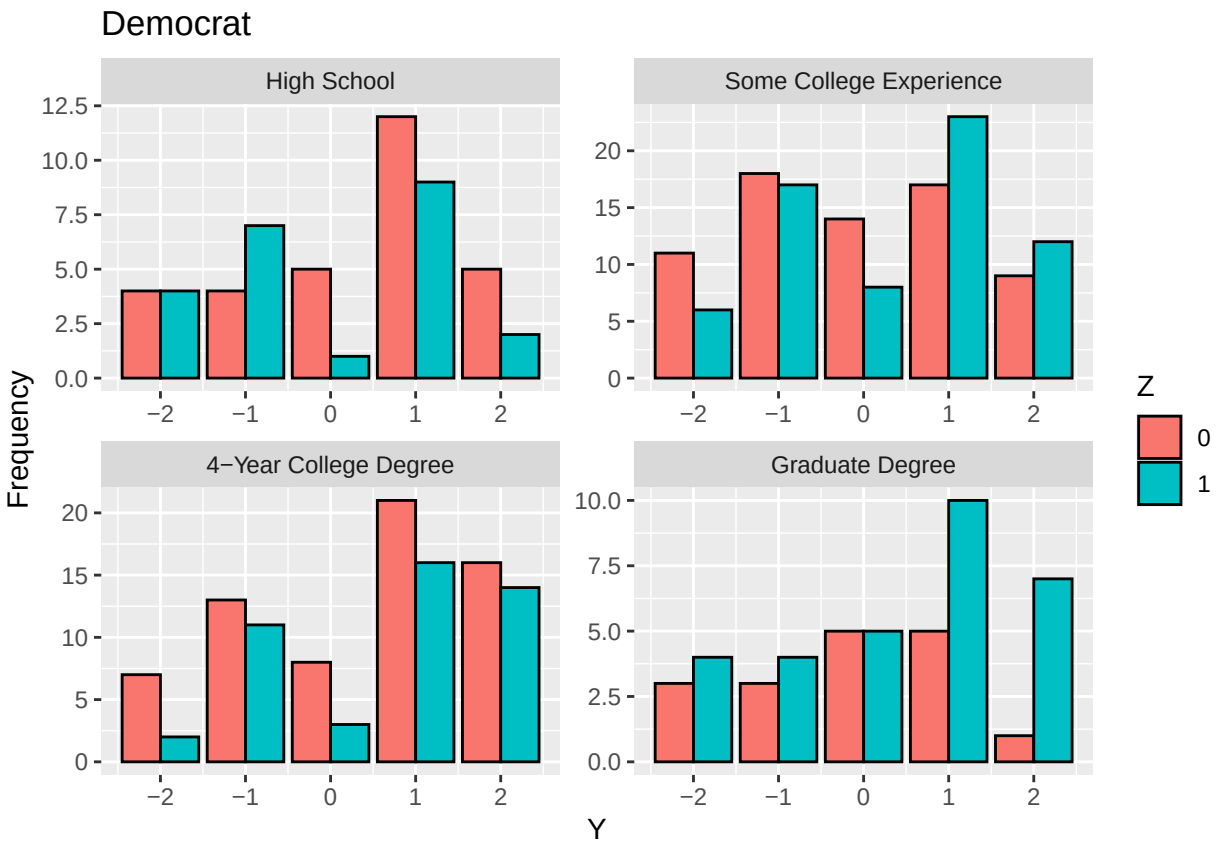

```
income_plot(2)
```



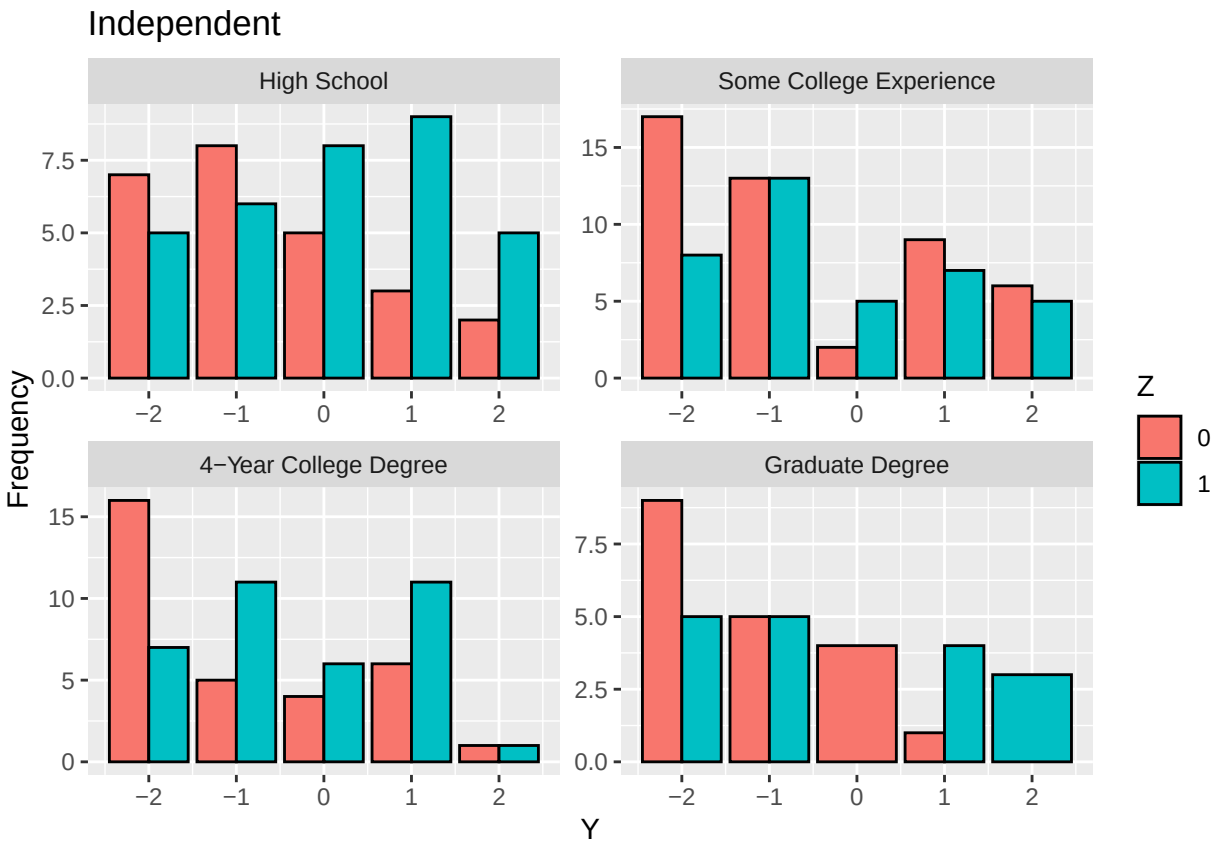

```
income_plot(3)
```



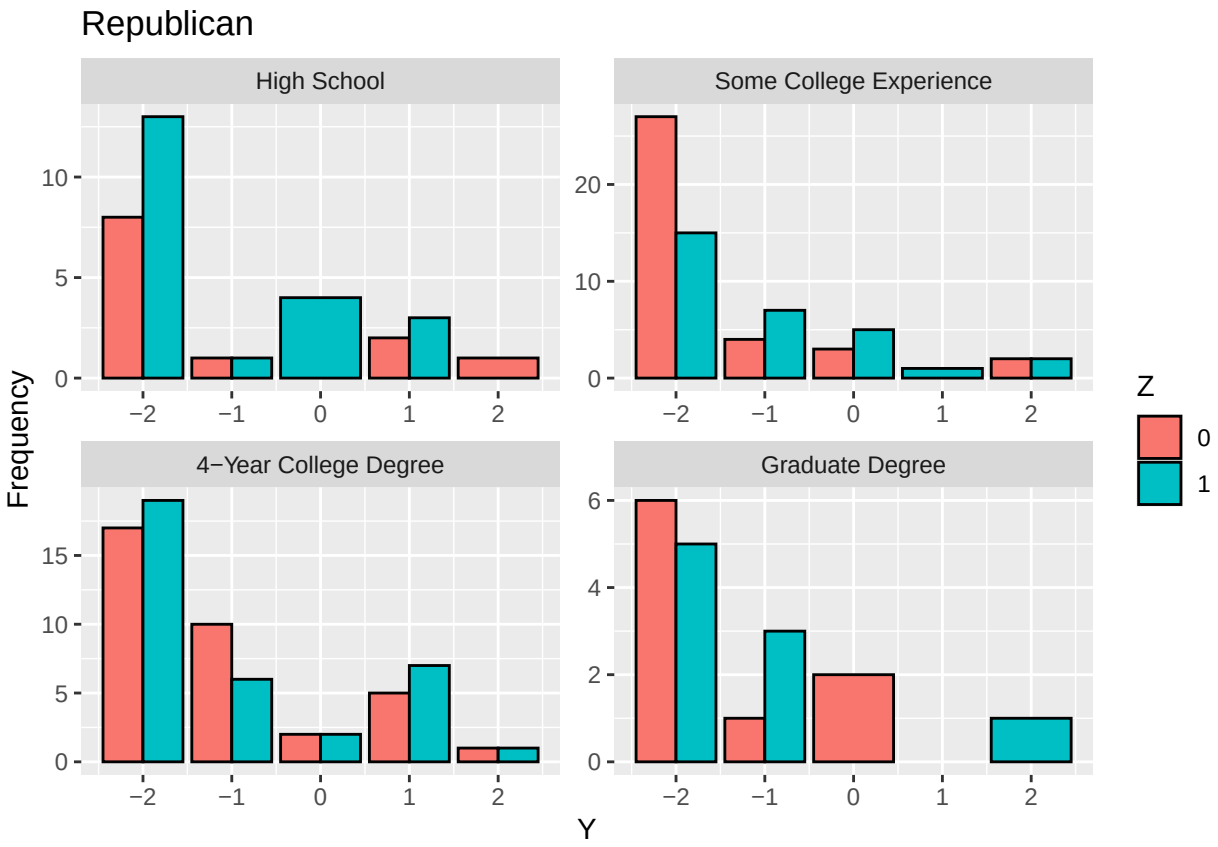

```
# USE THIS
datEduc = dlexp %>%
   filter(party.cues == 1) %>%
   group_by(across(all_of(c("pid3", "educationLevel", "Z", "Y")))) %>%
   summarize(Frequency = n())

datEduc$Z = as.factor(datEduc$Z)

# show breakdown for responses by education level, per group
educ_plot = function(pid){
  dat = filter(datEduc, pid3 == pid)

  # maintain order of education - high school, college, ....
  dat = transform(dat,
      educationLevel=factor(educationLevel,
                            levels=c("High School","Some College Experience","4-Year College Degree", "

  ggplot(data = dat,
      aes(x=Y, y=Frequency, fill=Z)) +
    geom_bar(stat="identity",  position="dodge", colour = "black") +
    facet_wrap(~educationLevel, scales = "free") +
    ggtitle(party_names[pid])

}

educ_plot(1)
```


Democrat
High School
Some College Experience
4−Year College Degree
Graduate Degree
Frequency
Y
Z
0
1

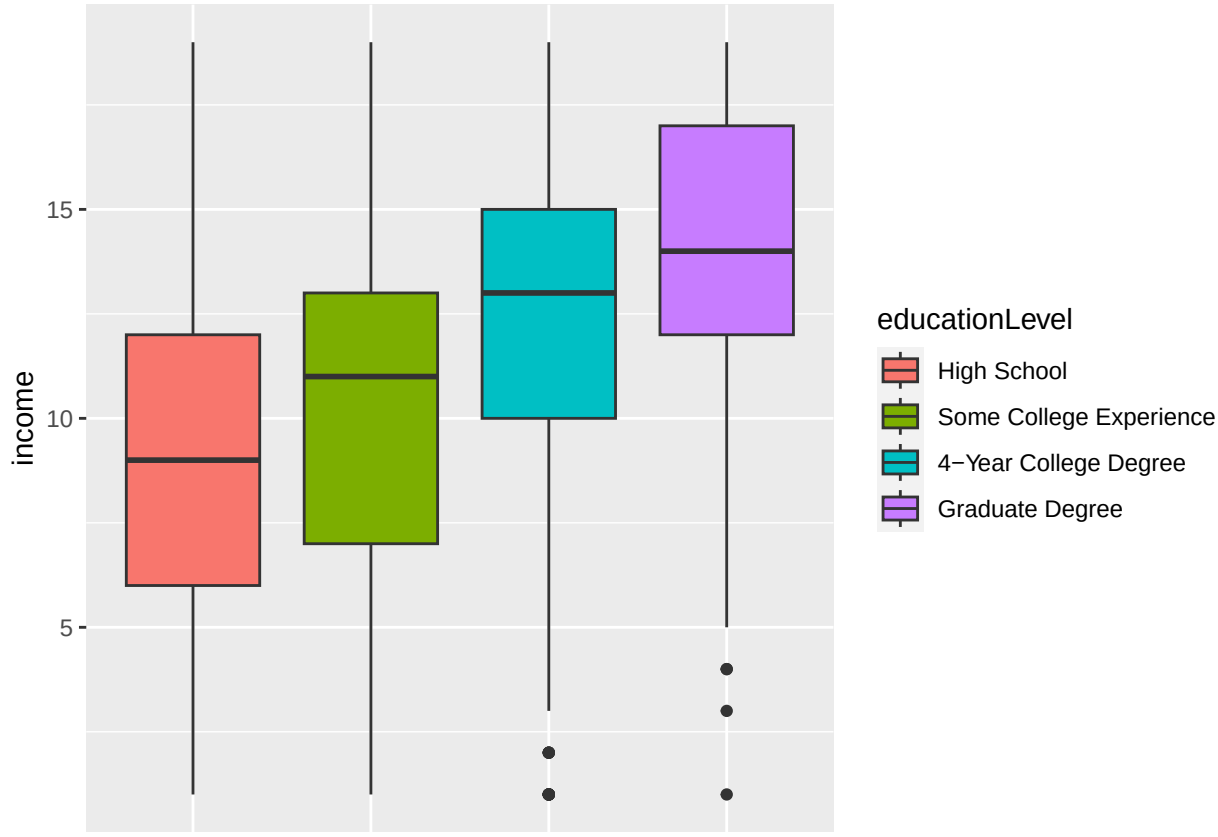
```
educ_plot(2)
```


Independent
High School
Some College Experience
4−Year College Degree
Graduate Degree
Frequency
Y
Z
0
1

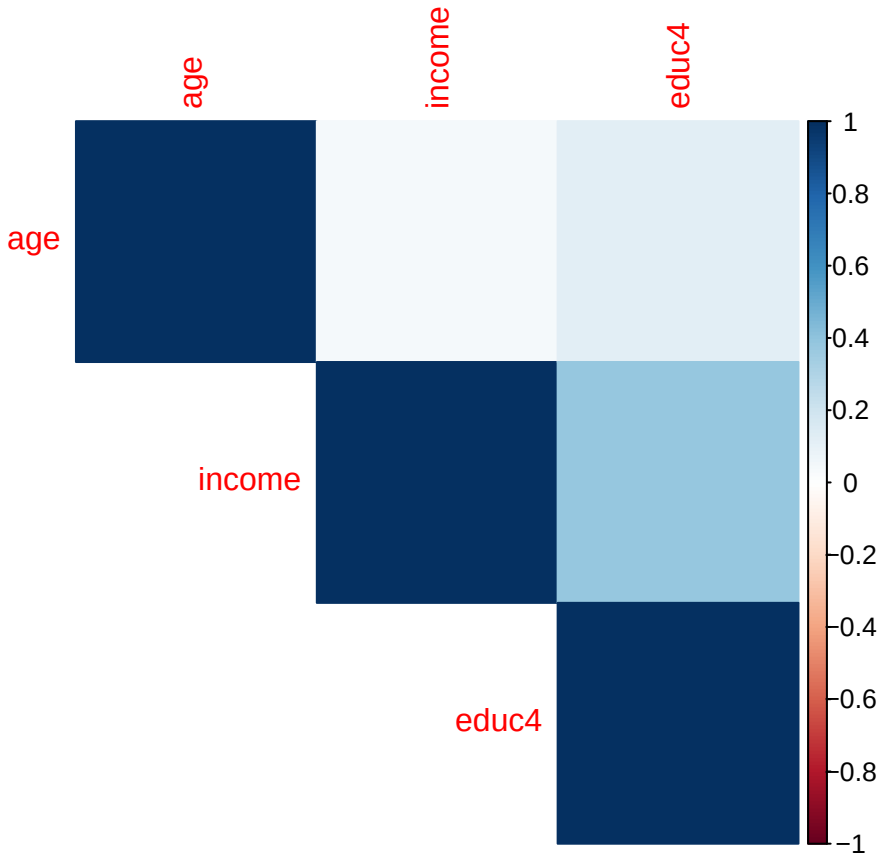
```
educ_plot(3)
```




```
# Education-Income Correlation

dat = transform(dlexp,
      educationLevel=factor(educationLevel,levels=c("High School","Some College Experience","4-Year Col
ggplot(aes(y = income, x = educationLevel), data = dat) +
geom_boxplot(aes(fill=educationLevel)) + theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```




```
dat_sub = dlexp[, c("income", "educ4", "age")]
cor_matrix <- cor(dat_sub)

corrplot(cor_matrix, method = "color", type = "upper", order = "hclust")
```


age
income
educ4
age
income
educ4
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1