Data Analysis for Problem Set 3

Olivia Bogiages

2025-11-17

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

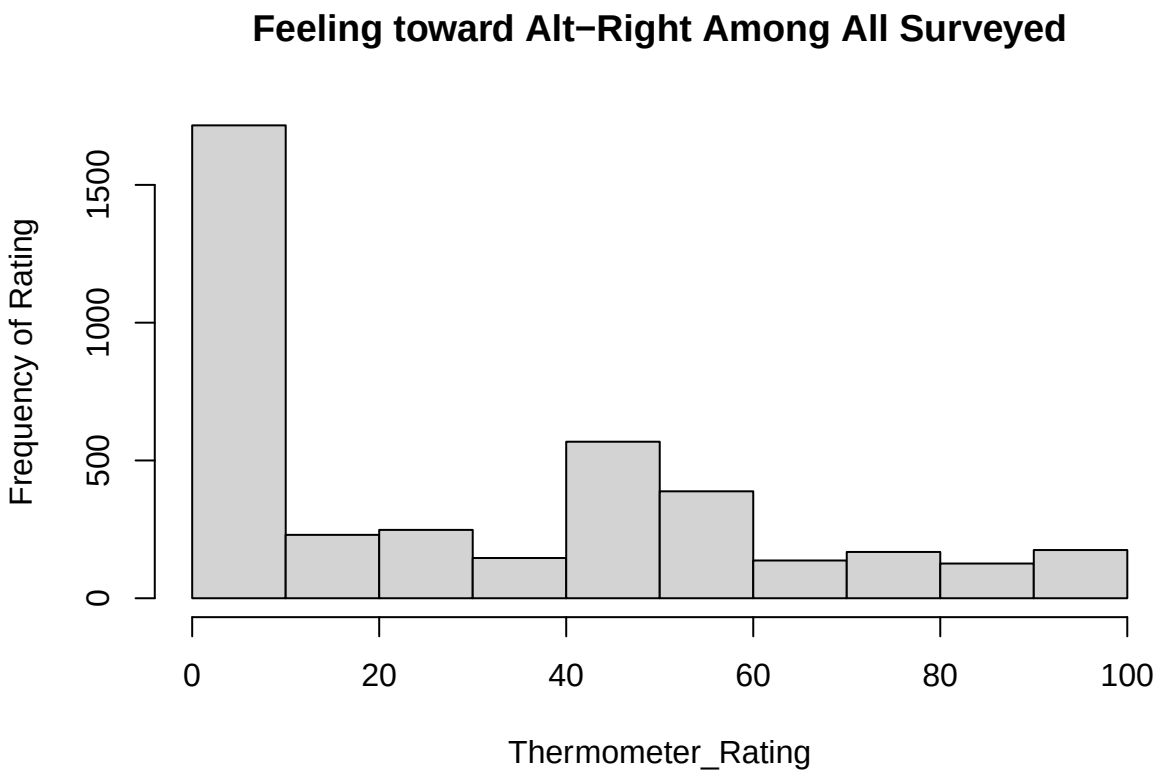
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#load the data
thermometer<-read.csv("https://raw.githubusercontent.com/MLBurnham/pols_602/refs/heads/main/data/thermo

#make an age variable from the birth year
thermometer$age<-2017-(thermometer$birth_year)

#pick one feeling thermometer and one categorical demographic variable.
#Pick: thermometer on the alt-right and variable race


#Spread and central tendency for all observations relating to the alt-right
hist(thermometer$ft_altright, xlab="Thermometer_Rating", ylab="Frequency of Rating", main="Feeling towa
```




```
median(thermometer$ft_altright, na.rm=TRUE)
```

```
## [1] 21
```

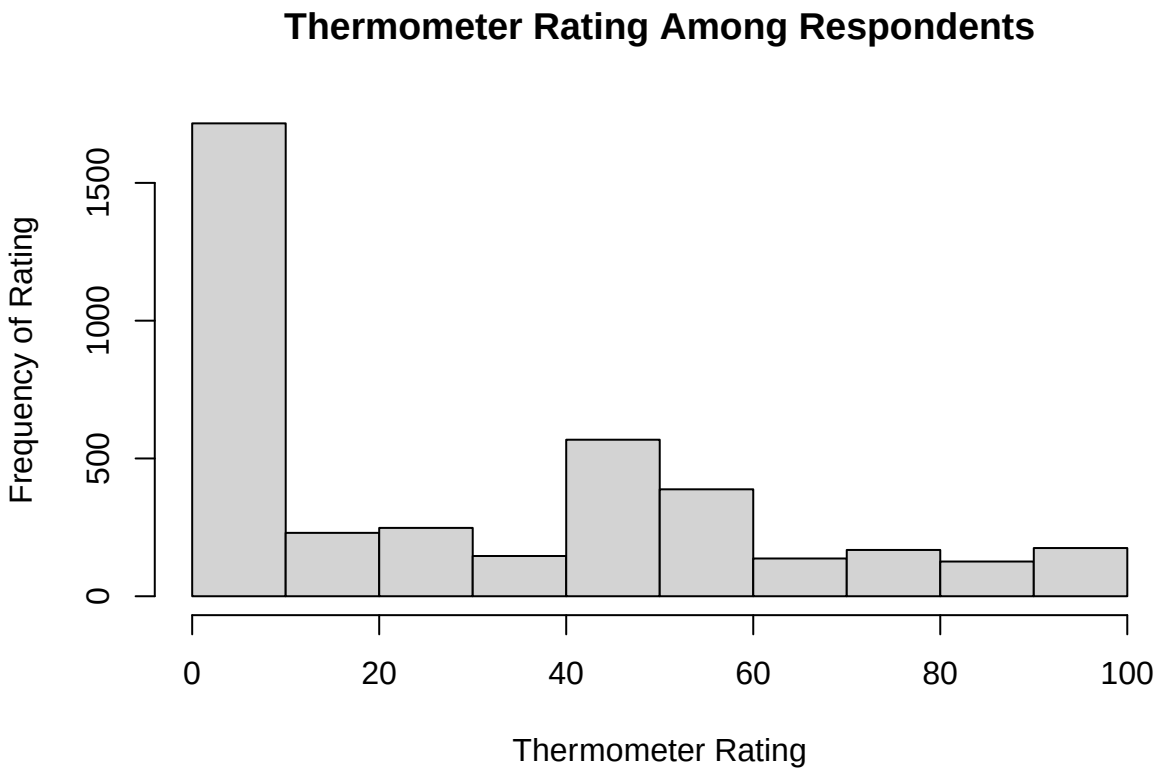
```
mean(thermometer$ft_altright, na.rm=TRUE)
```

```
## [1] 30.05484
```

```
sd(thermometer$ft_altright, na.rm=TRUE)
```

```
## [1] 29.97812
```

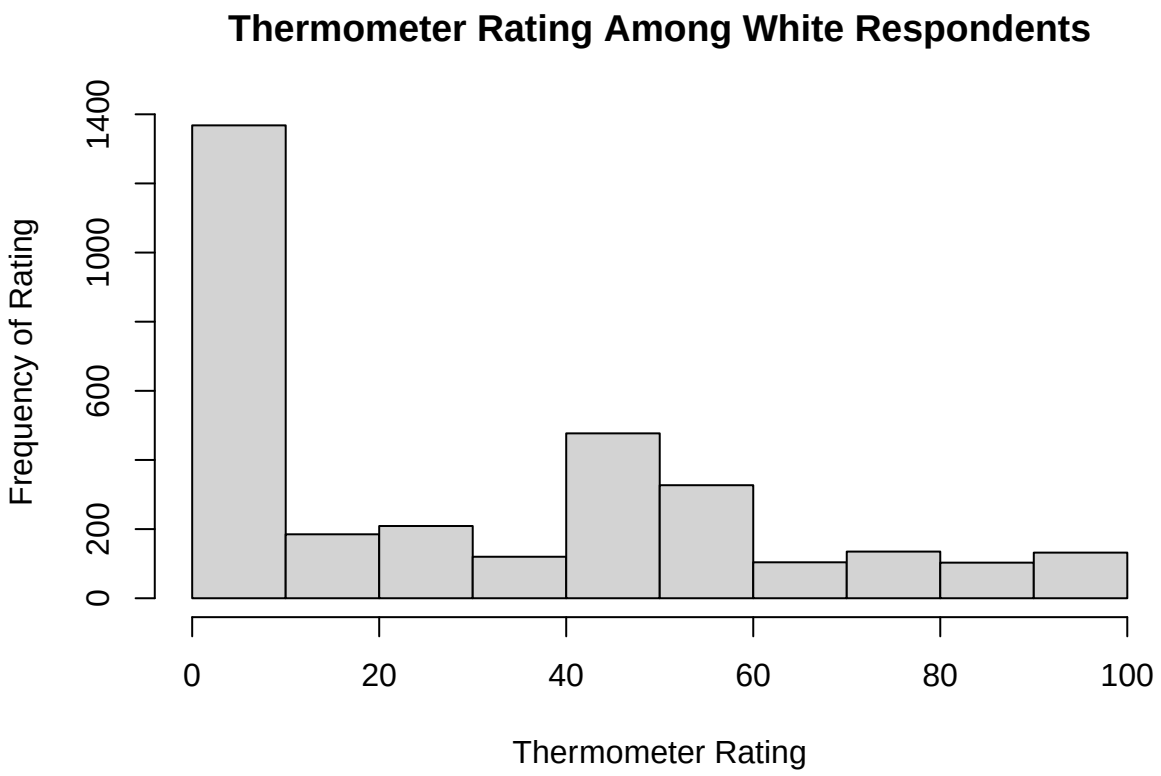
```
 #histogram of thermometer rating for alt right among all
hist (thermometer$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Thermometer
```




```
#spread and central tendency for each category in the demographic variable race
#Central tendency and spread of thermometer rating of alt right among White people
white_ftaltright<-thermometer %>% filter(race=="White") %>%
    select(ft_altright)
white_ftaltright %>% summarise(
mean=mean(ft_altright, na.rm=TRUE), median=median(ft_altright, na.rm=TRUE), sd=sd(ft_altright, na.rm=TRU
```

```
##      mean median      sd    n
## 1 30.1807     21 29.6938 4023
```

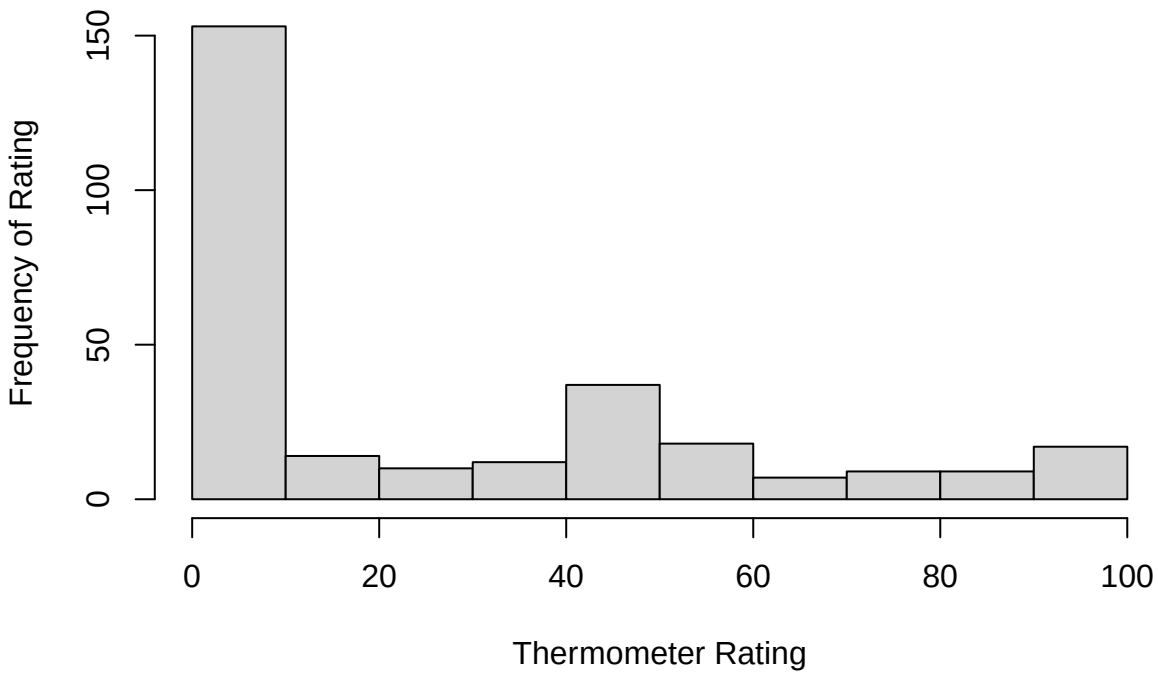
```
 #histogram of thermometer rating for alt right among White people
hist (white_ftaltright$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Thermo
```




```
#Spread and central tendency of thermometer rating about for alt right among Black people
black_ftaltright<-thermometer %>% filter(race=="Black") %>%
    select(ft_altright)
black_ftaltright %>% summarise(
mean=mean(ft_altright, na.rm=TRUE), median=median(ft_altright, na.rm=TRUE), sd=sd(ft_altright, na.rm=TRU
```

```
##       mean median       sd   n
## 1 26.27622      8 30.86321 405
```

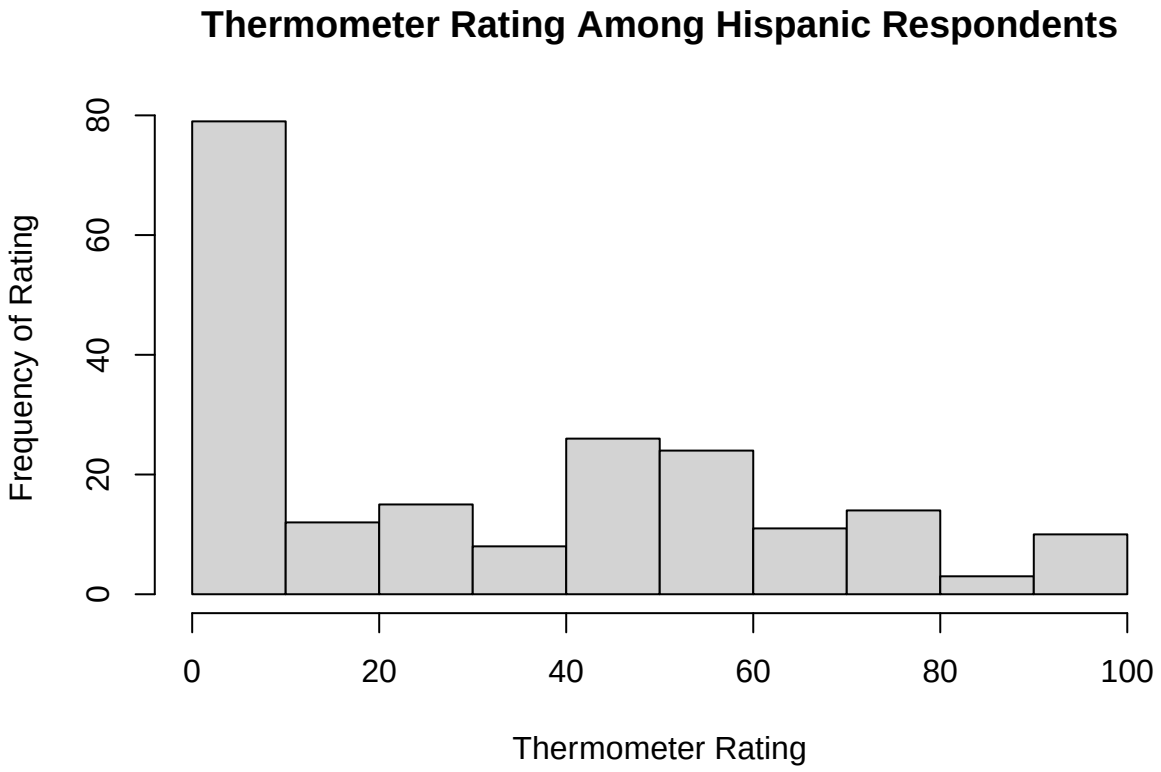
```
#histogram of thermometer rating for alt right among Black people
hist (black_ftaltright$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Thermo
```




```
#spread and central tendency of thermometer rating about for alt right among Hispanic people
hispanic_ftaltright<-thermometer %>% filter(race=="Hispanic") %>%
    select(ft_altright)
hispanic_ftaltright %>% summarise(
mean=mean(ft_altright, na.rm=TRUE), median=median(ft_altright, na.rm=TRUE), sd=sd(ft_altright, na.rm=TRU
```

```
##       mean median       sd   n
## 1 32.56436   29.5 30.09369 246
```

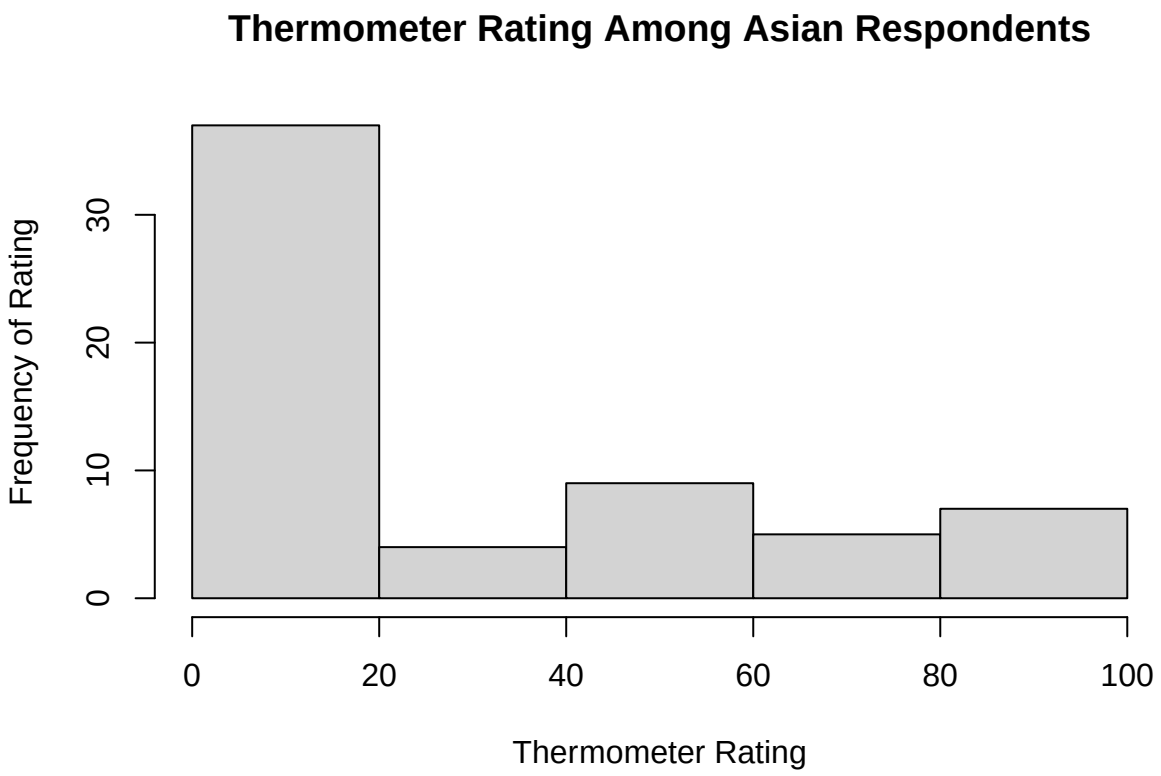
```
#histogram of thermometer rating for alt right among Hispanic people
hist (hispanic_ftaltright$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Ther
```




```
#Spread and central tendency of thermometer rating about for alt right among Asian people
asian_ftaltright<-thermometer %>% filter(race=="Asian") %>%
    select(ft_altright)
asian_ftaltright %>% summarise(
mean=mean(ft_altright, na.rm=TRUE), median=median(ft_altright, na.rm=TRUE), sd=sd(ft_altright, na.rm=TRU
```

```
##       mean median       sd  n
## 1 28.54839     11 33.04427 80
```

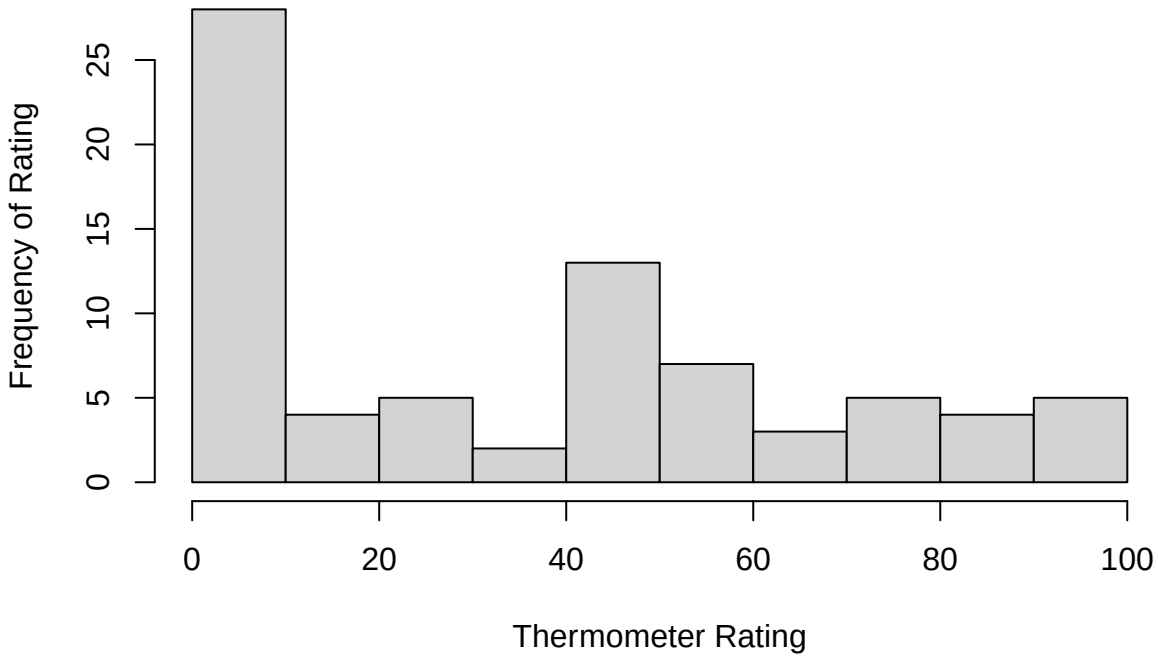
```
#histogram of thermometer rating for alt right among Asian people
hist (asian_ftaltright$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Thermo
```




```
#Spread and central tendency of thermometer rating about for alt right among Other races
other_ftaltright<-thermometer %>% filter(race=="Other") %>%
    select(ft_altright)
other_ftaltright %>% summarise(
mean=mean(ft_altright, na.rm=TRUE), median=median(ft_altright, na.rm=TRUE), sd=sd(ft_altright, na.rm=TRU
```

```
##       mean median       sd  n
## 1 35.97368   35.5 32.58874 90
```

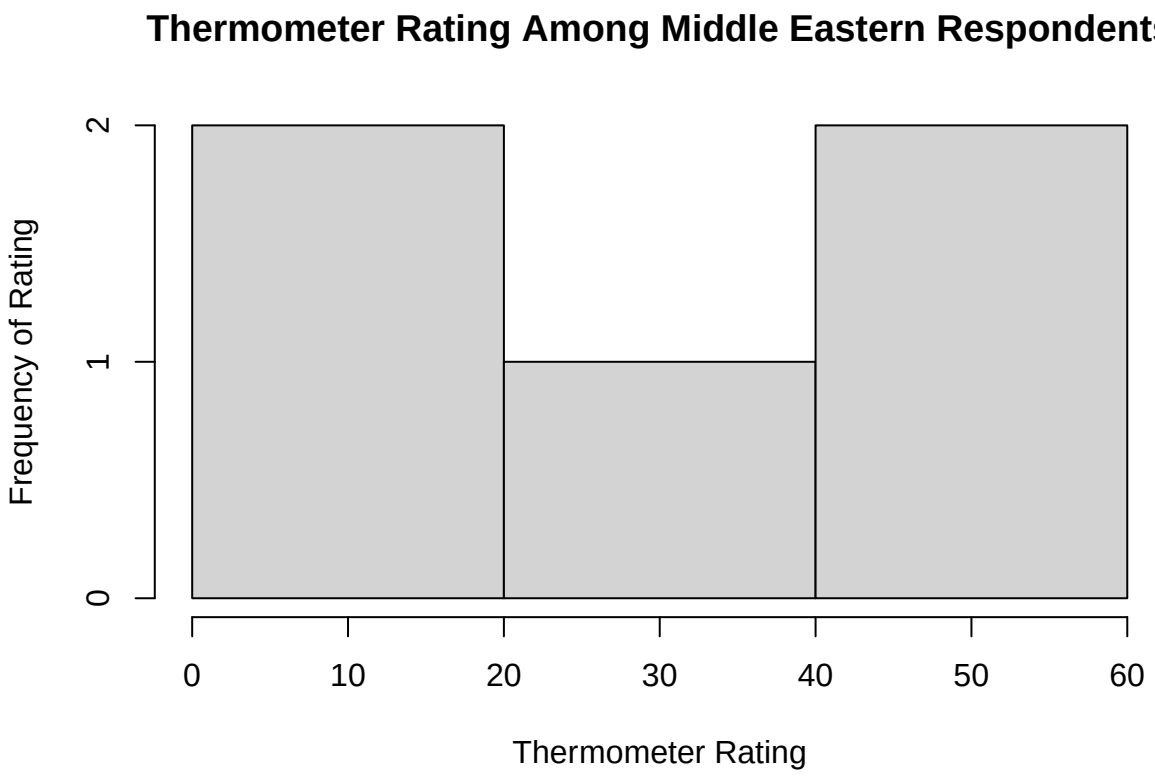
```
#histogram of thermometer rating for alt right among Other races
hist (other_ftaltright$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Thermo
```




```
#Spread and central tendency of thermometer rating about for alt right among Middle Easterner
ME_ftaltright<-thermometer %>% filter(race=="Middle Eastern") %>%
    select(ft_altright)
ME_ftaltright %>% summarise(
mean=mean(ft_altright, na.rm=TRUE), median=median(ft_altright, na.rm=TRUE), sd=sd(ft_altright, na.rm=TRU
```

```
##   mean median       sd n
## 1   31     36 22.53886 5
```

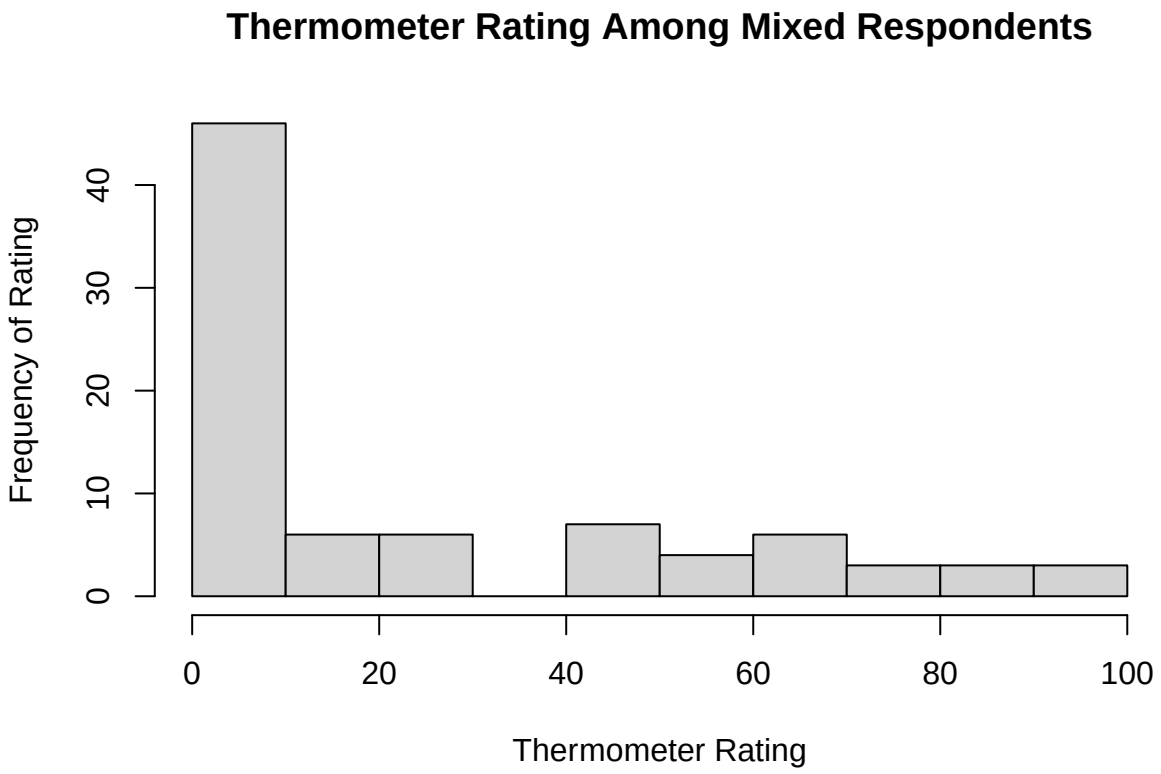
```
#histogram of thermometer rating for alt right among Middle Easterners
hist (ME_ftaltright$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Thermomet
```




```
#Spread and central tendency of thermometer rating about for alt right among Mixed people
mixed_ftaltright<-thermometer %>% filter(race=="Mixed") %>%
    select(ft_altright)
mixed_ftaltright %>% summarise(
mean=mean(ft_altright, na.rm=TRUE), median=median(ft_altright, na.rm=TRUE), sd=sd(ft_altright, na.rm=TRU
```

```
##       mean median       sd   n
## 1 24.89286     10 29.67577 106
```

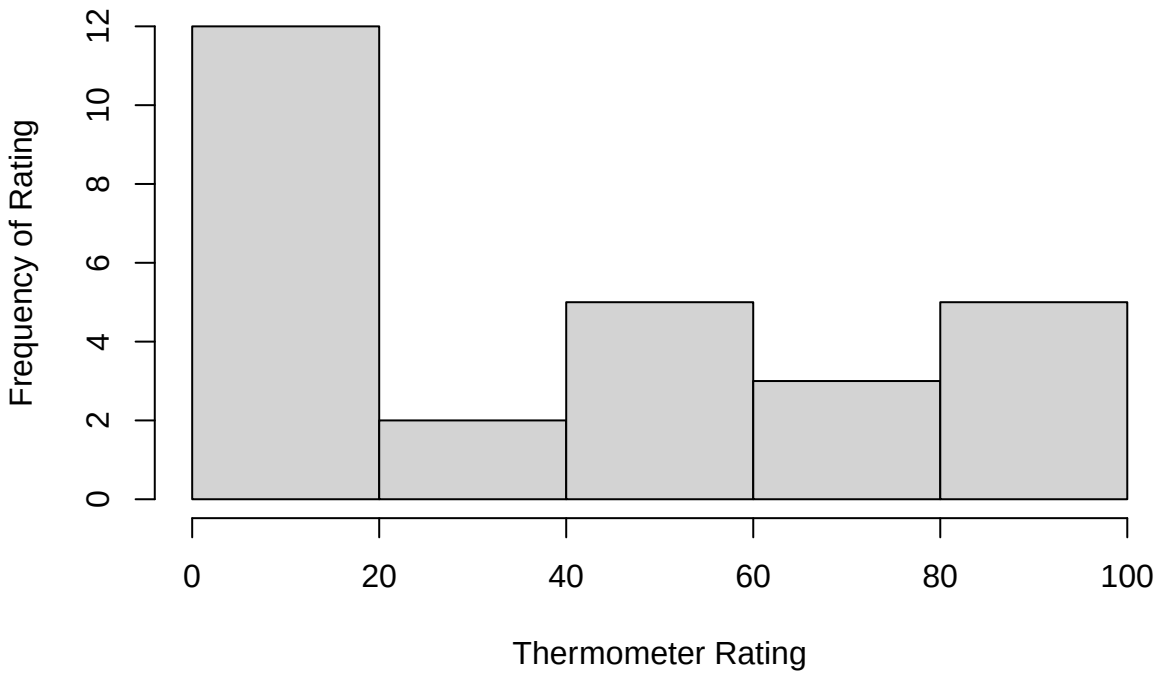
```
#histogram of thermometer rating for alt right among Mixed people
hist (mixed_ftaltright$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Thermo
```




```
#Spread and central tendency of thermometer rating about for alt right among Native Americans
NA_ftaltright<-thermometer %>% filter(race=="Native American") %>%
    select(ft_altright)
NA_ftaltright %>% summarise(
mean=mean(ft_altright, na.rm=TRUE), median=median(ft_altright, na.rm=TRUE), sd=sd(ft_altright, na.rm=TRU
```

```
##       mean median       sd  n
## 1 39.25926     31 35.51337 34
```

```
#histogram of thermometer rating for alt right among NA races
hist (NA_ftaltright$ft_altright, xlab="Thermometer Rating", ylab="Frequency of Rating", main="Thermomet
```




```r
#fit a regression model to estimate the conditional mean of the feeling thermometer for each category o


#conditional mean, regression model for race and rating the alt-right
thermometer$race<-as.factor(thermometer$race)
model<-lm(thermometer$ft_altright~thermometer$race)
summary(model)
```

```
##
## Call:
## lm(formula = thermometer$ft_altright ~ thermometer$race)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.259 -27.181  -9.181  20.819  75.107
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       28.548      3.804   7.505 7.55e-14 ***
## thermometer$raceBlack             -2.272      4.196  -0.542    0.588
## thermometer$raceHispanic           4.016      4.348   0.924    0.356
## thermometer$raceMiddle Eastern     2.452     13.924   0.176    0.860
## thermometer$raceMixed             -3.656      5.015  -0.729    0.466
## thermometer$raceNative American   10.711      6.906   1.551    0.121
## thermometer$raceOther              7.425      5.126   1.449    0.148
## thermometer$raceWhite              1.632      3.841   0.425    0.671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.95 on 3894 degrees of freedom
```

```
##   (1087 observations deleted due to missingness)
## Multiple R-squared:  0.003634,   Adjusted R-squared:  0.001843
## F-statistic: 2.029 on 7 and 3894 DF,  p-value: 0.04799
```

```
#create a new data frame that only contains rows for democrats and republicans , new binary variable fo
party_df<-subset.data.frame(thermometer, party_id %in% c("Democrat", "Republican"))
party_df$dem_rep_bin<-ifelse(party_df$party_id=="Democrat",1,0)
table(party_df$party_id)
```

```
##
##   Democrat Republican
##       1734       1412
```

```
#add age into df
party_df$age <- 2017 - party_df$birth_year
#use multiples linear regression to build a model that predicts your binary party_id variable, include
regression_model<-lm(
  dem_rep_bin~ft_altright*educ,
  data=party_df
)
summary(regression_model)
```

```
##
## Call:
## lm(formula = dem_rep_bin ~ ft_altright * educ, data = party_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8355 -0.3898  0.1751  0.2842  1.0701
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          0.7785013  0.0337420  23.072  < 2e-16 ***
## ft_altright                         -0.0068626  0.0007425  -9.243  < 2e-16 ***
## educ4-year                           0.0402974  0.0408346   0.987 0.323819
## educHigh school graduate            -0.1114899  0.0434656  -2.565 0.010377 *
## educNo HS                           -0.2356091  0.1119237  -2.105 0.035386 *
## educPost-grad                        0.0570107  0.0430829   1.323 0.185867
## educSome college                     0.0458861  0.0484549   0.947 0.343741
## ft_altright:educ4-year              -0.0018043  0.0009268  -1.947 0.051666 .
## ft_altright:educHigh school graduate  0.0014264  0.0009580   1.489 0.136641
## ft_altright:educNo HS                0.0035338  0.0023675   1.493 0.135655
## ft_altright:educPost-grad           -0.0037917  0.0010460  -3.625 0.000295 ***
## ft_altright:educSome college        -0.0009974  0.0010711  -0.931 0.351862
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4359 on 2423 degrees of freedom
##   (711 observations deleted due to missingness)
## Multiple R-squared:  0.2368, Adjusted R-squared:  0.2334
## F-statistic: 68.35 on 11 and 2423 DF,  p-value: < 2.2e-16
```

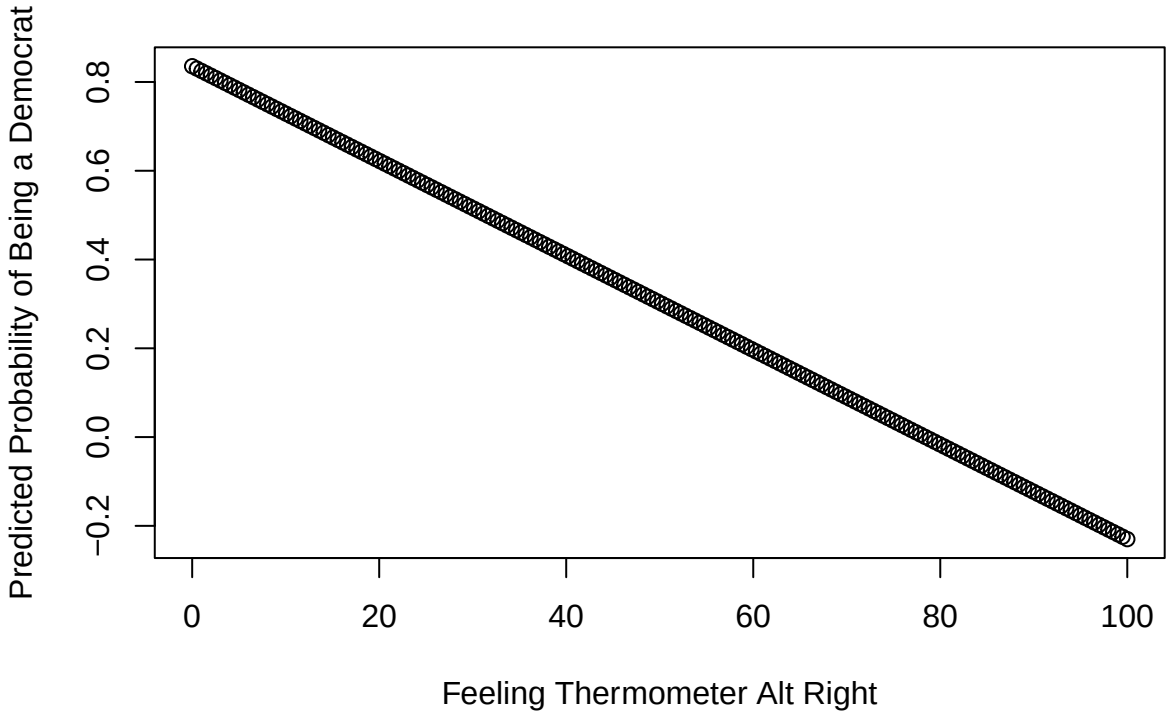
The coefficients in your model represent what change?

The negative coefficients represents a negative change in probability of being a democrat based on the education of the respondent and the positive coefficients represent a positive change in probability of being a democrat based on the education of the respondent.

```r
#Select one of the feeling thermometers in your model and plot how your
#predicted values change as the feeling thermometer changes. Interpret your results. Can this reasonabl
ft_seq<-seq(0,100, length.out=200)

#hold constant education level
pred_data<-(data.frame(educ="Post-grad", ft_altright=ft_seq
                       ))
#Predicted probabilities
pred_values<-predict(regression_model, newdata = pred_data)

plot(ft_seq, pred_values,
     xlab = "Feeling Thermometer Alt Right",
     ylab = "Predicted Probability of Being a Democrat")
```



Interpret your results. Can this reasonably be interpreted as a causal effect?

These results cannot be interpreted as causal given that unobserved confounders were not controlled for. This model shows that as a person rates their feeling toward the alt-right higher, their predicted probability of being a democrat decreases (when education is constant). This does not indicate that ranking the alt-right very warmly causes the individual to be a Republican.