Survey on Climate Change

Isabel Heard

4/30/2023

Introduction

The National Surveys on Energy and Environment (NSEE) is a biannual opinion survey conducted on energy and climate policy. The NSEE covers issues such as public opinion on attitudes toward climate change, public policy approaches to address climate issues, carbon taxes, and much more. The data was collected in the fall of 2017. Data was gathered by calling and surveying 929 adults within the United States. During this survey they were asked questions regarding their opinions to climate change along with their demographic information. The goal of this analysis is to see if a survey taker says they believe in climate change for one question, how likely are they to answer other questions with that same stance? This research question will be answered using the gee model in R.

Methods

Data Compilation and Cleaning

For the data cleaning, it began with re-coding the survey questions, so they became binary variables. For instance, with the variable '*gw_adapt*', there were originally 5 categories (1 = Strongly Agree, 2 = Somewhat Agree, 3 = Somewhat Disagree, 4 = Strongly Agree, 98 = Not Sure). This was then changed to '0', being people believing in focusing on adapting (combining options 1 and 2), and '1', being people who do not think we should focus on adapting (combining options 3 and 4). 'Not Sure' ended up being dropped as I was not interested in people who did not have a concrete answer.

I continued this for all of the other survey questions where '0' always meant someone was against believing in climate change and against pro climate change ideas, and a '1' always meant someone was pro believing in climate change and had pro climate change ideas. For the demographic variables the only change that occurred was removing people who were unsure about their political views, political party, and highest education level. Table 1 provides a list of all the variables and their descriptions.

I then turned the original data set into a long data set where there was now a '*Question'* column that had all six questions, and an '*Answer'* column that shows whether the person answered a '0' or '1'. I chose to pivot longer so I could see every answer that a person gave, so now every user takes up six rows, one for each question they answered. With this new data set, there were 528 participants left.

Graphical Summaries

Figure 1 shows the counts for each survey question. From those graphs, for every question asked, more people answered '1' (pro believing in climate change) than '0'. Figure 2 effectively summarizes the demographic variables. There are more males than females who took this survey, and most people has some college education or more. The majority of people who took they survey were between the ages of 45 and 64. Also, most people said that they were moderate.

Model Selection

Initially before learning about the 'gee' package in R, I decided to create a model using the 'glm' function. With this, I was able to see what explanatory variables were significant in my model for predicting *Answer*. When I ran my initial model 'fit_all', with all the predictors, *Region* and *demog_polparty* came back as not significant in the model. So, I then created a new mode, 'fit_new' without those two variables. When I ran an ANOVA test to see which model performed better, they both performed the same, with a p-value output of 0.062, and almost identical AIC values. So, I decided that it would be best to work with the simpler model, just containing *Question, AgeRecode, demog_polviews, demog_gender*, and *demog_education*. Also, including *demog_polparty* did not seem necessary, as the variable demog_polviews gave us just as much insight, if not more for where people stand politically.

Next, I began to test for marginal homogeneity between questions. I choose to use a gee model because I am dealing with binary data. Also, since I am looking at the same individual over multiple questions, the model could consider the correlation within a person. The gee model also lets you use the 'id' argument, so you can separate each person within the model. I also decided to use the 'exchangeable' correlation structure since I need to account for repeated measures, and not assume independence. With these structures in place, a final model was made, called 'fit_gee_reduced'.

Accuracy and Interpretation

Based on the coefficients from the gee model seen in Output 1 (Appendix A), it seems that for people who answered '1' to the baseline question, would answer '1' to all other questions as well. For the questions *govt_fed, gw_belief, and gw_adapt*, were very likely to answer '1'. For the questions *state_impactsfelt* and *state_vehichles_v1*, people were still likely to answer '1', but just on a lower scale. For instance, you could say that holding all other variables constant, the odds of answering '1' are 2.07 times higher for every one unit increase for the question *gw_belief* compared to other questions. Looking at another question, holding all other variables constant, the odds of answering '1' are 1.2 times higher for every one unit increase for the question *state_impactsfelt* compared to other questions.

Within this model, *AgeRecode* and *demog_gender* are not significant predictors, suggesting that age and gender do not have a big role is predicting if someone will answer a '0' or '1'. *Demog_polviews* had the greatest significance when it comes to the demographic variables. For every one unit increase in *demog_polviews*, the odds of answering a '1' are 2.09. For age, which was less significant, for every one unit increase in age, the odds of answering a '1' decrease by a factor of 0.9, meaning that older participants are less likely to answer '1' on a question compared to younger ones. Next, an ANOVA test was run to see the accuracy of the model. With 5 degrees of freedom, a X2 value of 103.6, and a p-value of 2e-16, this gives strong evidence against the null hypothesis of no homogeneity from question to question, adjusted for age, political views, gender, and highest education level.

Discussion and Conclusion

With this data and results, we can conclude that there is an association between the responses to the different questions. The distribution of answers for one question is not independent of the answers to the other questions. This makes sense as, if you believe in climate change in one instance, one would think that you would continue your beliefs in other questions as well. Overall, if you were a younger participant, you were more likely to answer a '1'. A more liberal participant is more likely to answer a '1' as well as males, and people who are more highly educated. In all, if you answered a '1' for one question, you answered a '1' for all other questions. Although it was interesting to see that gender and age were not significant predictors.

In further research it would be interesting to see if this marginal homogeneity would still hold true if there were more questions asked that related to what personal steps people would be willing to take in order to help climate change. Furthermore, the NSEE has surveys from 2016 and 2019, so it would be interesting to see how people's opinions on climate change have changed over the years through a longitudinal study.

References

Link to dataset

Appendix A

This appendix contains all tables and figures referenced within this report.

Table 1: Description of Variables in Modelling Process

Variable	Description
gw_belief	Is there solid evidence that the average temperature on earth has been warming in past four decades?
gw_adapt	Instead of trying to stop global warming from occurring we should focus on adapting?
state_impactsfelt	My state has already felt negative effects from global warming.
govt_fed	How much responsibility federal goverment has for taking actions to reduce global warming.
state_vehicles_v1	State governments should require auto makers to increase the fuel efficiency of their vehicles even if it increases the cost of the vehicle.
carbon_support_v3	Support for reducing GHG by taxing carbon-based fuels.
demog_polviews	Participants political views
AgeRecode	Participants age
demog_edu	Highest Level of education
Region	Region located in the US
demog_polparty	Political Party
demog_gender	Participants gender

Figure 1: Bar Chart of Question Variables

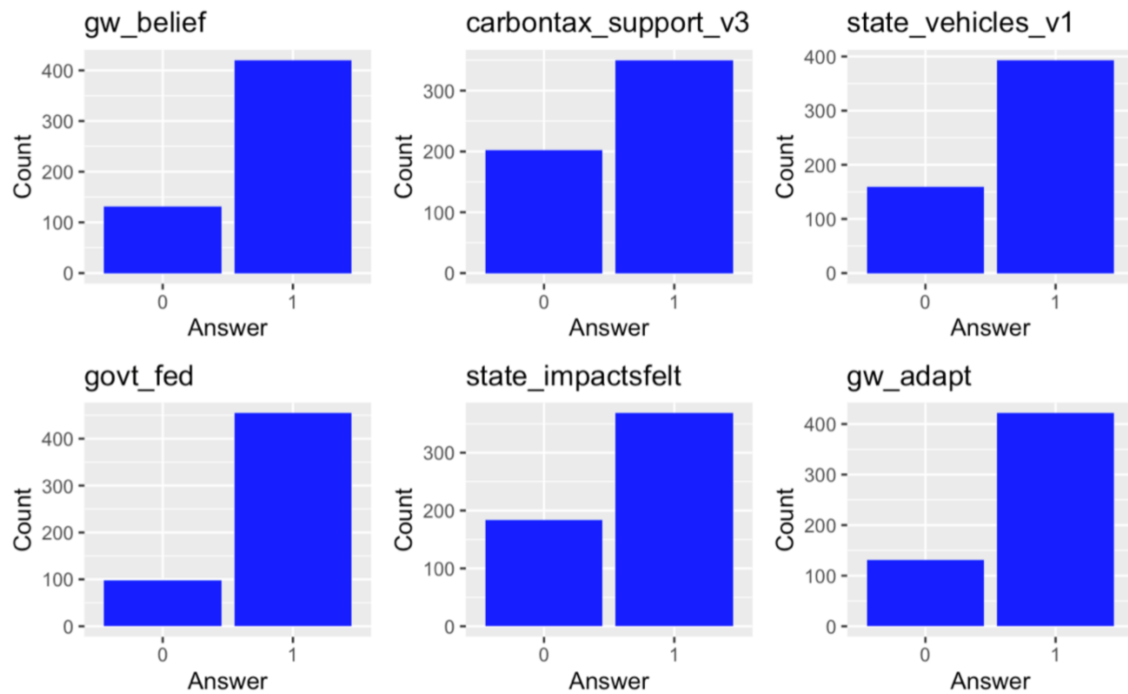
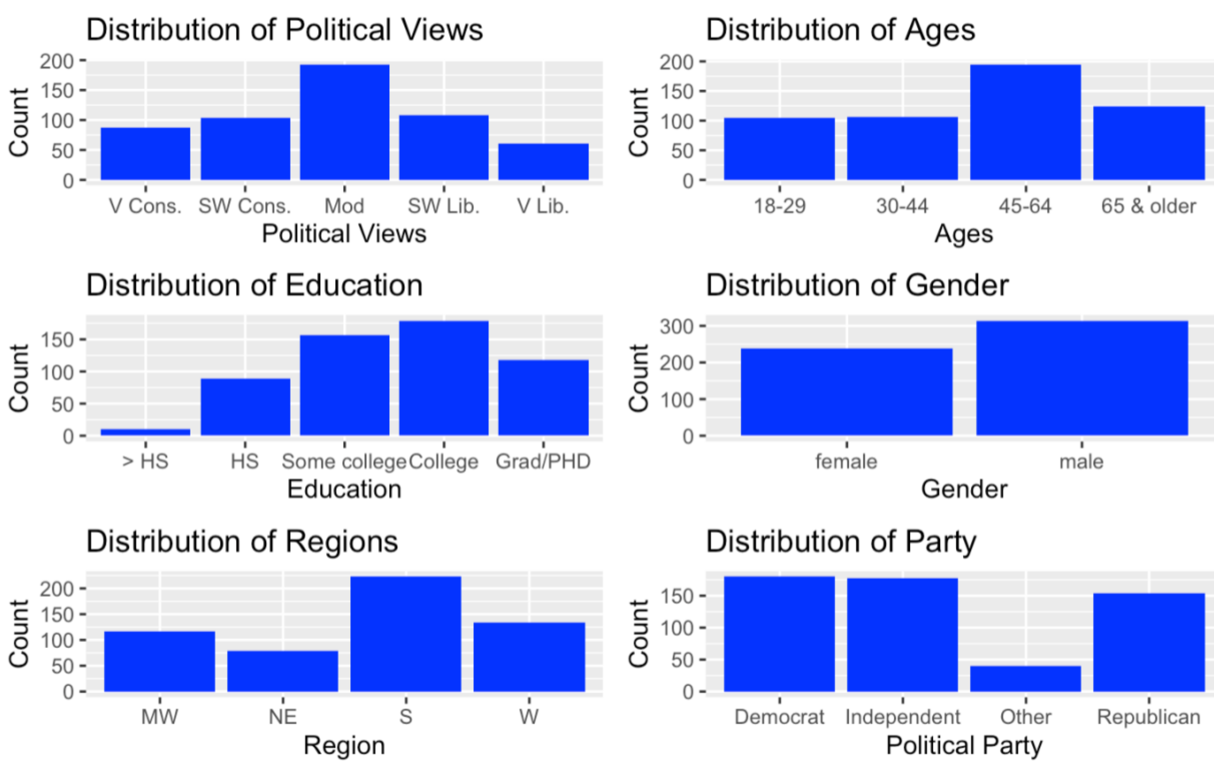


Figure 2: Bar Chart of Demographic Variables



Output 1: fit_gee_reduced model output

```
Call:
geeglm(formula = Answer ~ Question + AgeRecode + demog_polviews +
   demog_gender + demog_edu, family = binomial(link = "logit"),
   data = df_long_clean, id = UserID, corstr = "exchangeable")

 Coefficients:
                     Estimate Std.err   Wald Pr(>|W|)
(Intercept)           -2.2405  0.3774  35.24  2.9e-09 ***
Questiongovt_fed        1.2062  0.1355  79.19  < 2e-16 ***
Questiongw_adapt         0.7158  0.1439  24.73  6.6e-07 ***
Questiongw_belief        0.7281  0.1277  32.49  1.2e-08 ***
Questionstate_impactsfelt  0.1812  0.1156   2.46   0.1169
Questionstate_vehicles_v1  0.3929  0.1244   9.98   0.0016 **
AgeRecode             -0.0882  0.0637   1.92   0.1663
demog_polviews          0.7389  0.0643 132.25  < 2e-16 ***
demog_gender            0.2458  0.1361   3.26   0.0709 .
demog_edu              0.1801  0.0627   8.25   0.0041 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Correlation structure = exchangeable
Estimated Scale Parameters:
 Link = identity
Estimated Correlation Parameters:
Number of clusters:   528  Maximum cluster size: 6
```

Appendix B

This appendix contains all code used for this report.

Imports

```
library(gee)

## Warning: package 'gee' was built under R version 4.0.5

library(geepack)
library(multgee)
```

```
## Loading required package: gnm

library(corrplot)

## corrplot 0.92 loaded

library(ggplot2)
library(tidyr)

## Warning: package 'tidyr' was built under R version 4.0.5

library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:multgee':
##
##     housing

library(gridExtra)
data <- read.csv("data.csv")
```

Creating a data frame

```
library("data.table")
df <- fread("data.csv", select = c("RSPDNUM", "gw_belief", "gw_adapt", "state
_impactsfelt",  "govt_fed", "state_vehicles_v1", "carbontax_support_v3", "Age
Recode", "Region", "demog_edu", "demog_polparty", "demog_polviews", "demog_ge
nder"))
colnames(df)[1] <- 'UserID' #Renaming RSPDNUM
```

Looking at NA values and dimensions

```
sort(sapply(df, function(y) sum(length(which(is.na(y))))))
dim(df)
```

Re-coding variables

```
# 1 - pro believing in climate change
# 0 - against believing in climate change


#gw_belief - Is there solid evidence avg temp on earth has been warming in pa
st 4 decades?
#table(df$gw_belief)
df<-df[!(df$gw_belief==98),]
df$gw_belief <- replace(df$gw_belief, df$gw_belief == 1, 1) # yes you believe
in CC
df$gw_belief <- replace(df$gw_belief, df$gw_belief == 2, 0) # no you don't be
lieve in CC
#table(df$gw_belief)

#gw_adapt:Instead of trying to stop GW from occurring we should focus on adap
ting
#table(df$gw_adapt)
df<-df[!(df$gw_adapt==98),]
df$gw_adapt <- replace(df$gw_adapt, df$gw_adapt == 1, 0) # Focus on adapting
df$gw_adapt <- replace(df$gw_adapt, df$gw_adapt == 2, 0) # Focus on adapting
df$gw_adapt <- replace(df$gw_adapt, df$gw_adapt == 3, 1) # Not focus on adapt
ing
df$gw_adapt <- replace(df$gw_adapt, df$gw_adapt == 4, 1) # Not focus on adapt
ing
#table(df$gw_adapt)

#state_impactsfelt: My state has already felt negative effects from global wa
rming
#table(df$state_impactsfelt)
df<-df[!(df$state_impactsfelt==98),]
df$state_impactsfelt <- replace(df$state_impactsfelt, df$state_impactsfelt ==
1, 1) # My state has felt the effect
df$state_impactsfelt <- replace(df$state_impactsfelt, df$state_impactsfelt ==
2, 1) # My state has felt the effect
```

```
df$state_impactsfelt <- replace(df$state_impactsfelt, df$state_impactsfelt ==
3, 0) # My state has not felt the effect
df$state_impactsfelt <- replace(df$state_impactsfelt, df$state_impactsfelt ==
4, 0) # My state has not felt the effect
#table(df$state_impactsfelt)

#govt_fed: How much responsibility federal govt has for taking actions to red
uce GW
#table(df$govt_fed)
df<-df[!(df$govt_fed==98),]
df$govt_fed <- replace(df$govt_fed, df$govt_fed == 1, 1) # A great deal of re
sponsibility
df$govt_fed <- replace(df$govt_fed, df$govt_fed == 2, 1) # A great deal of re
sponsibility
df$govt_fed <- replace(df$govt_fed, df$govt_fed == 3, 0) # No responsibility
#table(df$govt_fed)

#state_vehicles_v1 - State governments should require auto makers to increase
the fuel efficiency of their vehicles even if it increases the cost of the ve
hicle
#table(df$state_vehicles_v1)
df<-df[!(df$state_vehicles_v1==98),]
df$state_vehicles_v1 <- replace(df$state_vehicles_v1, df$state_vehicles_v1 ==
1, 1) # States should take action
df$state_vehicles_v1 <- replace(df$state_vehicles_v1, df$state_vehicles_v1 ==
2, 1) # States should take action
df$state_vehicles_v1 <- replace(df$state_vehicles_v1, df$state_vehicles_v1 ==
3, 0) # States should not take action
df$state_vehicles_v1 <- replace(df$state_vehicles_v1, df$state_vehicles_v1 ==
4, 0) # States should nnot take action
#table(df$state_vehicles_v1)

#carbontax_support_v3 - Support for reducing GHG by taxing carbon based fuels
```

```r
#table(df$carbontax_support_v3)
df<-df[!(df$carbontax_support_v3==98),]
df$carbontax_support_v3 <- replace(df$carbontax_support_v3, df$carbontax_supp
ort_v3 == 1, 1) # Yes to tax
df$carbontax_support_v3 <- replace(df$carbontax_support_v3, df$carbontax_supp
ort_v3 == 2, 1) # Yes to tax
df$carbontax_support_v3 <- replace(df$carbontax_support_v3, df$carbontax_supp
ort_v3 == 3, 0) # No to tax
df$carbontax_support_v3 <- replace(df$carbontax_support_v3, df$carbontax_supp
ort_v3 == 4, 1) # No to tax
df$carbontax_support_v3 <- replace(df$carbontax_support_v3, df$carbontax_supp
ort_v3 == 5, 0) # No to tax
#table(df$carbontax_support_v3)

#demog_polviews: (1 = Very conservative, 2 = Somewhat conservative, 3 = Moder
ate, 4 = Somewhat liberal, 5 = Very liberal, 98 = Not sure, 99 = Refused)
df<-df[!(df$demog_polviews==98),]
#table(df$demog_polviews)

#AgeRecode: (1 = 18-29, 2 = 30-44, 3 = 45-64, 4 = 65 & older, 98 = Not Sure,
99 = Refused)
#table(df$AgeRecode)

#demog_edu: (1 = Less than HS, 2 = HS, 3 = Some college, 4 = College, 5 = Gra
d/PHD)
df<-df[!(df$demog_edu==98),]
#table(df$demog_edu)

#Region: (1 = NE, 2 = S, 3 = MW, 4 = W)
#table(df$Region)

#demog_polparty: (1 = dem, 2 = rep, 3 = other, 4 = ind)
df<-df[!(df$demog_polparty==98),]
```

```
#table(df$demog_polparty)


#demog_gender: (1 = male, 2 = female)
table(df$demog_gender)
```

Bar Charts for Questions

```
#gw_belief Bar Chart
#table(df$gw_belief)
gw_belief_counts <- data.frame(
  belief = c("0", "1"),
  count = c(132, 420))
q1 <- ggplot(gw_belief_counts, aes(x = belief, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Answer") +
  ylab("Count") +
  ggtitle("gw_belief")

#carbontax_support_v3 Bar Chart
#table(df$carbontax_support_v3)
carbontax_support_v3_counts <- data.frame(
  belief = c("0", "1"),
  count = c(202, 350))
q2 <- ggplot(carbontax_support_v3_counts, aes(x = belief, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Answer") +
  ylab("Count") +
  ggtitle("carbontax_support_v3")

#df$state_vehicles_v1  Bar Chart
#table(df$state_vehicles_v1 )
state_vehicles_v1_counts <- data.frame(
  belief = c("0", "1"),
  count = c(159, 393))
```

```r
q3 <- ggplot(state_vehicles_v1_counts, aes(x = belief, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Answer") +
  ylab("Count") +
  ggtitle("state_vehicles_v1")

#df$govt_fed  Bar Chart
#table(df$govt_fed )
govt_fed_counts <- data.frame(
  belief = c("0", "1"),
  count = c(98, 454))
q4 <- ggplot(govt_fed_counts, aes(x = belief, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Answer") +
  ylab("Count") +
  ggtitle("govt_fed")

#df$state_impactsfelt  Bar Chart
#table(df$state_impactsfelt )
state_impactsfelt_counts <- data.frame(
  belief = c("0", "1"),
  count = c(184, 368))
q5 <- ggplot(state_impactsfelt_counts, aes(x = belief, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Answer") +
  ylab("Count") +
  ggtitle("state_impactsfelt")

#df$gw_adapt  Bar Chart
#table(df$gw_adapt )
gw_adapt_counts <- data.frame(
  belief = c("0", "1"),
  count = c(131, 421))
```

```
q6 <- ggplot(gw_adapt_counts, aes(x = belief, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Answer") +
  ylab("Count") +
  ggtitle("gw_adapt")

#grid.arrange(q1, q2, q3, q4, q5, q6, ncol = 3)
```

Bar charts for demographic variables

```
#demog_polviews Bar Chart
#table(df$demog_polviews)
demog_polviews_counts <- data.frame(
  polviews = c("V Cons.", "SW Cons.", "Mod",
               "SW Lib.", "V Lib."),
  count = c( 87,104, 192, 108, 61 ))

demog_polviews_counts$polviews <- factor(demog_polviews_counts$polviews, leve
ls = c("V Cons.", "SW Cons.", "Mod",
               "SW Lib.", "V Lib."))
graph2 <- ggplot(demog_polviews_counts, aes(x = polviews, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Political Views") +
  ylab("Count") +
  ggtitle("Distribution of Political Views")



#AgeRecode Bar Chart
#table(df$AgeRecode)
AgeRecode_counts <- data.frame(
  ages = c("18-29", "30-44", "45-64", "65 & older"),
  count = c(104, 106, 194, 124 ))

graph3 <- ggplot(AgeRecode_counts, aes(x = ages, y = count)) +
```

```
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Ages") +
  ylab("Count") +
  ggtitle("Distribution of Ages")

#demog_edu Bar Chart
#table(df$demog_edu)
demog_edu_counts <- data.frame(
  edu = c("> HS", "HS", "Some college", "College", "Grad/PHD"),
  count = c( 10, 89, 156, 179, 118))

demog_edu_counts$edu <- factor(demog_edu_counts$edu, levels = c("> HS", "HS",
"Some college", "College", "Grad/PHD"))
graph4 <- ggplot(demog_edu_counts, aes(x = edu, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Education") +
  ylab("Count") +
  ggtitle("Distribution of Education")


#demog_gender: (1 = male, 2 = female)
#table(df$demog_gender)
demog_edu_counts <- data.frame(
  gender = c("male", "female"),
  count = c(314, 238 ))

graph5 <- ggplot(demog_edu_counts, aes(x = gender, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Gender") +
  ylab("Count") +
  ggtitle("Distribution of Gender")
```

```r
#Region: (1 = NE, 2 = S, 3 = MW, 4 = W)
#table(df$Region)
Region_counts <- data.frame(
  region = c("NE", "S", "MW", "W"),
  count = c( 79,223,117,133))

graph6 <- ggplot(Region_counts, aes(x = region, y = count)) +
  geom_bar(stat = "identity", fill = "blue") +
  xlab("Region") +
  ylab("Count") +
  ggtitle("Distribution of Regions")

#grid.arrange(graph2,graph3, graph4, graph5, graph6, ncol = 2)
```

Contingency Tables

```r
#gw_belief vs Age
gw_belief_age <- xtabs(~gw_belief + AgeRecode, data = df)
gw_belief_age
chisq.test(gw_belief_age)
stdres <- chisq.test(gw_belief_age)$stdres
stdres
#The distribution of "gw_belief" values does significantly differ across the
age groups. Believers who are between the ages of 18-64 had large positive re
siduals meaning that more believers between those ages occurred than the hypo
thesis of independence predicted.

#gw_belief vs demog_polviews
gw_belief_pol <- xtabs(~gw_belief + demog_polviews, data = df)
gw_belief_pol
chisq.test(gw_belief_pol)
stdres <- chisq.test(gw_belief_pol)$stdres
stdres
#The distribution of "gw_belief" values does significantly different across t
```

```
he political views groups.

#gw_belief vs demog_gender
gw_belief_gender <- xtabs(~gw_belief + demog_gender, data = df)
gw_belief_gender
chisq.test(gw_belief_gender)
#The distribution of "gw_belief" values does significantly different across t
he genders.

#gw_belief vs demog_edu
gw_belief_edu <- xtabs(~gw_belief + demog_edu, data = df)
gw_belief_edu
chisq.test(gw_belief_edu)
```

```
## Warning in chisq.test(gw_belief_edu): Chi-squared approximation may
be incorrect
```

```
#The distribution of "gw_belief" values does not  significantly differe
nt across the education groups.

#gw_belief vs demog_polparty
gw_belief_party <- xtabs(~gw_belief + demog_polparty, data = df)
gw_belief_party
chisq.test(gw_belief_party)
#The distribution of "gw_belief" values does significantly different across t
he political party groups.

#gw_belief vs Region
gw_belief_region <- xtabs(~gw_belief + Region, data = df)
gw_belief_region
chisq.test(gw_belief_region)
stdres <- chisq.test(gw_belief_region)$stdres
stdres
#The distribution of "gw_belief" values does not  significantly different acr
```

```
oss Regions.

#3-way contingency table - gw_belief + demog_polviews + AgeRecode
mytable <- xtabs(~gw_belief+demog_polviews+AgeRecode, data=df)
ftable(mytable)

#3 way contingency table with percentages
ftable(addmargins(prop.table(mytable, c(1, 2)), 3))*100
```

GLM with just gw_belief variable

```
#Fitting glm for gw_belief with demographic explanatory variables
fit <- glm(gw_belief ~ AgeRecode + Region + demog_polviews + demog_polparty +
demog_gender + demog_edu, data = df, family = binomial)
summary(fit) #AIC: 475.2
anova(fit)
stepAIC(fit) # gw_belief ~ AgeRecode + demog_polviews + demog_gender + demog_
edu -> AIC=AIC: 471.7

fit2 <- glm(gw_belief ~ AgeRecode + demog_polviews + demog_gender + demog_edu
, data = df, family = binomial)
summary(fit2)

anova(fit2, fit, test = "Chisq")
#The results of the LRT indicate that the difference in deviance between the
two models is 0.43261, and the p-value associated with the test is 0.8055. Si
nce the p-value is greater than 0.05, we fail to reject the null hypothesis a
nd conclude that Model 1 is not significantly worse than Model 2. Therefore,
we might choose to use Model 1 since it is simpler and contains fewer predict
ors.
```

Pivoting to a long data set

```
    df_long <- pivot_longer(df, cols = c(gw_belief, gw_adapt, state_impacts
felt, govt_fed, carbontax_support_v3, state_vehicles_v1), names_to = "Questio
```

```
n",  values_to = "Answer")
df_long_clean <- na.omit(df_long)
```

Contingency table for Questions and Answers

```
    questionTable <- xtabs(~Answer + Question, data = df_long_clean)
questionTable
chisq.test(questionTable)
stdres <- chisq.test(questionTable)$stdres
stdres
```

GLM with Question & AIC

```
fit_all <- glm(Answer ~ Question + AgeRecode + Region + demog_polviews + demo
g_polparty + demog_gender + demog_edu, data = df_long_clean, family = binomia
l(link = "logit"))
summary(fit_all) #AIC: 3155, Region and demog_polparty are not significant
```

```
      ##
## Call:
## glm(formula = Answer ~ Question + AgeRecode + Region + demog_polviews +
##     demog_polparty + demog_gender + demog_edu, family = binomial(link = "l
ogit"),
##     data = df_long_clean)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6035  -0.8603   0.5036   0.7638   1.8852
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -2.27650    0.30599  -7.440 1.01e-13 ***
## Questiongovt_fed             1.21334    0.16028   7.570 3.73e-14 ***
## Questiongw_adapt             0.71999    0.14955   4.814 1.48e-06 ***
## Questiongw_belief            0.73234    0.14976   4.890 1.01e-06 ***
## Questionstate_impactsfelt  0.18227    0.14241   1.280  0.20058
```

```
## Questionstate_vehicles_v1  0.39522    0.14473   2.731  0.00632 **
## AgeRecode                 -0.10251    0.04306  -2.381  0.01727 *
## Region                     0.08251    0.04517   1.827  0.06775 .
## demog_polviews             0.75257    0.04235  17.770  < 2e-16 ***
## demog_polparty            -0.05948    0.03745  -1.588  0.11225
## demog_gender               0.23753    0.09197   2.583  0.00980 **
## demog_edu                  0.17759    0.04121   4.309 1.64e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3682.8  on 3167  degrees of freedom
## Residual deviance: 3130.8  on 3156  degrees of freedom
## AIC: 3154.8
##
## Number of Fisher Scoring iterations: 4
```

```
anova(fit_all)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Answer
##
## Terms added sequentially (first to last)
##
##
##                Df Deviance Resid. Df Resid. Dev
## NULL                            3167     3682.8
## Question        5    67.96      3162     3614.9
## AgeRecode       1    34.53      3161     3580.3
## Region          1     2.50      3160     3577.8
```

```
## demog_polviews  1   417.88       3159      3159.9
## demog_polparty  1     3.73       3158      3156.2
## demog_gender    1     6.73       3157      3149.5
## demog_edu       1    18.66       3156      3130.8
```

```
stepAIC(fit_all, direction="both")
```

```
## Start:  AIC=3154.82
## Answer ~ Question + AgeRecode + Region + demog_polviews + demog_polparty +
##     demog_gender + demog_edu
##
##                  Df Deviance    AIC
## <none>                3130.8 3154.8
## - demog_polparty  1   3133.3 3155.3
## - Region          1   3134.2 3156.2
## - AgeRecode       1   3136.5 3158.5
## - demog_gender    1   3137.5 3159.5
## - demog_edu       1   3149.5 3171.5
## - Question        5   3210.6 3224.6
## - demog_polviews  1   3503.0 3525.0
```

```
##
## Call:  glm(formula = Answer ~ Question + AgeRecode + Region + demog_polvie
ws +
##     demog_polparty + demog_gender + demog_edu, family = binomial(link = "l
ogit"),
##     data = df_long_clean)
##
## Coefficients:
##               (Intercept)           Questiongovt_fed
##                  -2.27650                    1.21334
##           Questiongw_adapt           Questiongw_belief
##                   0.71999                    0.73234
## Questionstate_impactsfelt  Questionstate_vehicles_v1
##                   0.18227                    0.39522
```

```
##                  AgeRecode                        Region
##                   -0.10251                       0.08251
##              demog_polviews                demog_polparty
##                    0.75257                      -0.05948
##                demog_gender                     demog_edu
##                    0.23753                       0.17759
##
## Degrees of Freedom: 3167 Total (i.e. Null);  3156 Residual
## Null Deviance:       3683
## Residual Deviance: 3131  AIC: 3155
```

```
fit_new <- glm(Answer ~ Question + AgeRecode + demog_polviews + demog_g
ender + demog_edu, data = df_long_clean, family = binomial(link = "logit"))
summary(fit_new) #AIC: 3156
```

```
##
## Call:
## glm(formula = Answer ~ Question + AgeRecode + demog_polviews +
##     demog_gender + demog_edu, family = binomial(link = "logit"),
##     data = df_long_clean)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6068  -0.8637   0.5056   0.7666   1.8278
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -2.24646    0.26651  -8.429  < 2e-16 ***
## Questiongovt_fed            1.21107    0.16012   7.563 3.93e-14 ***
## Questiongw_adapt            0.71858    0.14940   4.810 1.51e-06 ***
## Questiongw_belief           0.73091    0.14961   4.885 1.03e-06 ***
## Questionstate_impactsfelt  0.18190    0.14227   1.279  0.20104
## Questionstate_vehicles_v1  0.39443    0.14459   2.728  0.00637 **
## AgeRecode                  -0.10148    0.04299  -2.360  0.01825 *
```

```
## demog_polviews            0.75118    0.04222  17.792  < 2e-16 ***
## demog_gender              0.25990    0.09137   2.845  0.00445 **
## demog_edu                 0.17933    0.04122   4.351 1.36e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3682.8  on 3167  degrees of freedom
## Residual deviance: 3136.4  on 3158  degrees of freedom
## AIC: 3156.4
##
## Number of Fisher Scoring iterations: 4
```

```
anova(fit_new)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Answer
##
## Terms added sequentially (first to last)
##
##
##                Df Deviance Resid. Df Resid. Dev
## NULL                            3167     3682.8
## Question        5    67.96      3162     3614.9
## AgeRecode       1    34.53      3161     3580.3
## demog_polviews  1   416.68      3160     3163.6
## demog_gender    1     8.24      3159     3155.4
## demog_edu       1    19.02      3158     3136.4
```

```
#Test to see if one model fits better than the other
anova(fit_all, fit_new, test = "Chisq") #no significant difference
```

```
## Analysis of Deviance Table
##
## Model 1: Answer ~ Question + AgeRecode + Region + demog_polviews + demog_p
olparty +
##     demog_gender + demog_edu
## Model 2: Answer ~ Question + AgeRecode + demog_polviews + demog_gender +
##     demog_edu
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      3156     3130.8
## 2      3158     3136.4 -2   -5.562  0.06198 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test for marginal homogeneity

```
fit_gee_ex <- gee(Answer ~ Question + AgeRecode + demog_polviews + demog_gend
er + demog_edu, id = UserID, family = binomial(link = "logit"), corstr ="exch
angeable", data = df_long_clean)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

## running glm to get initial regression estimate

##                 (Intercept)           Questiongovt_fed           Questio
ngw_adapt
##                  -2.2464611                  1.2110663                0.7185
837
##          Questiongw_belief Questionstate_impactsfelt Questionstate_vehicles
_v1
##                   0.7309118                  0.1819018                0.3944
275
##                   AgeRecode             demog_polviews             demog_gen
der
##                  -0.1014826                  0.7511798                0.2599
019
```

```
##                   demog_edu
##                   0.1793331
```

```r
#summary(fit_gee_ex)

#Geeglm - test the situation effect, adjusting for gender, polviews, age, and edu
fit_gee_reduced <- geeglm(Answer ~ Question + AgeRecode + demog_polviews + demog_gender + demog_edu, id = UserID, family = binomial(link = "logit"), corstr = "exchangeable",data = df_long_clean)
summary(fit_gee_reduced)
```

```
##
## Call:
## geeglm(formula = Answer ~ Question + AgeRecode + demog_polviews +
##     demog_gender + demog_edu, family = binomial(link = "logit"),
##     data = df_long_clean, id = UserID, corstr = "exchangeable")
##
##  Coefficients:
##                             Estimate  Std.err    Wald Pr(>|W|)
## (Intercept)                 -2.24054  0.37743  35.239 2.92e-09 ***
## Questiongovt_fed             1.20623  0.13555  79.193  < 2e-16 ***
## Questiongw_adapt             0.71581  0.14393  24.733 6.59e-07 ***
## Questiongw_belief            0.72809  0.12773  32.493 1.20e-08 ***
## Questionstate_impactsfelt  0.18122  0.11558   2.458  0.11692
## Questionstate_vehicles_v1  0.39293  0.12438   9.980  0.00158 **
## AgeRecode                   -0.08818  0.06371   1.916  0.16634
## demog_polviews               0.73891  0.06425 132.249  < 2e-16 ***
## demog_gender                 0.24583  0.13611   3.262  0.07091 .
## demog_edu                    0.18014  0.06271   8.252  0.00407 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = exchangeable
```

```
## Estimated Scale Parameters:
##
##             Estimate Std.err
## (Intercept)   0.9709  0.1068
##   Link = identity
##
## Estimated Correlation Parameters:
##       Estimate Std.err
## alpha   0.2561  0.0421
## Number of clusters:   528  Maximum cluster size: 6
```

```
anova(fit_gee_reduced)
```

```
## Analysis of 'Wald statistic' Table
## Model: binomial, link: logit
## Response: Answer
## Terms added sequentially (first to last)
##
##               Df    X2 P(>|Chi|)
## Question       5 103.6   < 2e-16 ***
## AgeRecode      1  12.2   0.00048 ***
## demog_polviews  1 148.2   < 2e-16 ***
## demog_gender   1   3.2   0.07339 .
## demog_edu      1   8.3   0.00407 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```