Results.Rmd

Leo Costa

12/14/2021

# Introduction

Latinx people make up a significant portion of the population of the United States and that percentage has been consistently grown for decades. These days, Latinx people are the US’s second largest racial/ethnic group. Latinx people also come from a variety of different cultural backgrounds despite belonging to an overarching umbrella term, whether that is informed by their (or their family’s) country of origin or how long their family has lived in the United States, among other factors. Naturally, this means that there is immense variation in this population: age, socioeconomic status, political views, and more. I am interested in better understanding the diversity of this demographic population, in particular because I am part of it.

The 2018 National Survey of Latinos was conducted between July 26, 2018 and September 9, 2018 by the Pew Research Center to understand the views of the US’s Latinx population. The survey was conducted via telephone in both English and Spanish and asked questions of Latinx adults living in the United States. This data was collected during the Trump administration, and President Trump was well known, both during his campaign and his presidency, for being outspoken about and representing a conservative stance on immigration, particularly towards Latinx immigrants. I am curious by how different variables within the Latinx community, specifically their citizenship status, English language proficiency (which may indicate their ability to assimilate within US culture), and experiences of discrimination based on being Latinx may predict participants’ political stances.

I investigated two outcome variables: the degree to which participants had a negative opinion of Trump and the degree to which participants had a conservative immigration policy stance. Models to predict each of these outcome variables were created using three predictor variables: the citizenship status of participants, the participants’ English language proficiency, and the participants’ experiences of discrimination. How these variables were derived from the data will be elaborated upon under the univariate analysis section.

I had different hypotheses for the two outcome variables: • [Negative] Opinion of Trump: I expect those who were naturalized citizens and non-citizens to have more negative opinions of Trump relative to those who are natural born citizens. I expect those with without English proficiency to have more negative opinions of Trump because Trump is known for having xenophobic stances. I expect those with greater experiences of discrimination to have a greater negative opinion of Trump because Trump has been known for making comments that may encourage discrimination against Latinx immigrants and descendants. • [Conservative] Immigration Policy Stance: I anticipate those who are naturalized citizens and non-citizens to have less conservative immigration policy stances and than natural born citizens because they are immigrants. I anticipate those who are not proficient in English to have less conservative immigration policy stances because they are more likely to be immigrants or to be closely tied to them. Further, I anticipate those with higher experiences of discrimination to have a less conservative immigration policy stance.

# Univariate Analyses

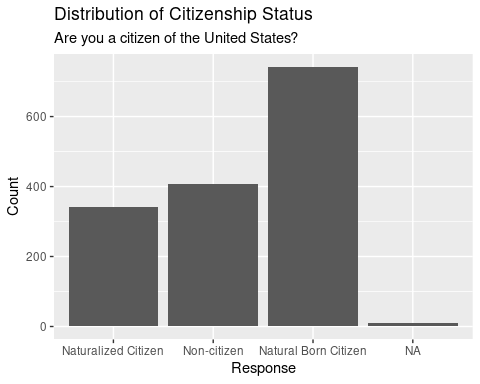
## Univariate Analyses: Predictor Variables: Citizenship Status

Citizenship status was gleaned from qn9 (“Are you a citizen of the United States?”). Those who responded yes were labeled as naturalized citizens, those who responded no were labeled as non-citizens, and missing values were treated as natural born citizens, as those who were born in the United States (a previous question) were not asked this question. This produced a factor of three levels. The sample is 49.77% natural born citizens, 27.36% non-citizens, and 22.87% naturalized citizens. There were 10 participants who refused to answer the question and were thus treated as missing values. This could potentially be because questions about citizenship feel more vulnerable for those who are undocumented.

NSL$qn9[is.na(NSL$qn9)] <- 3  
NSL$qn9[NSL$qn9 == 9] <- NA  
NSL$qn9 <- as.factor(NSL$qn9)  
levels(NSL$qn9)[levels(NSL$qn9)=="1"] <- "Naturalized Citizen"  
levels(NSL$qn9)[levels(NSL$qn9)=="2"] <- "Non-citizen"  
levels(NSL$qn9)[levels(NSL$qn9)=="3"] <- "Natural Born Citizen"  
  
prop.table(table(NSL$qn9))

##   
## Naturalized Citizen Non-citizen Natural Born Citizen   
## 0.2287056 0.2736419 0.4976526

ggplot(NSL, aes(qn9)) +   
 geom\_bar() +   
 labs(title = "Distribution of Citizenship Status", subtitle = "Are you a citizen of the United States?", x = "Response", y = "Count")



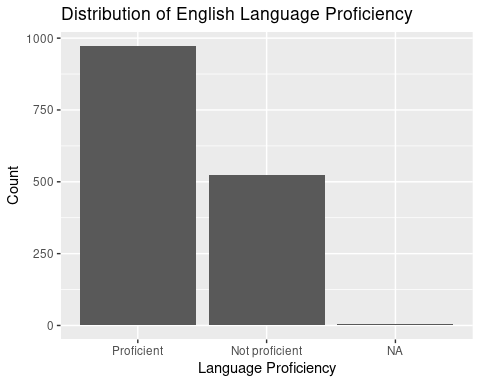
## Univariate Analyses: Predictor Variables: English Language Proficiency

English language proficiency was gathered from qnlan3 (“Would you say you can carry on a conversation in English, both understanding and speaking – very well, pretty well, just a little, or not at all?”). The responses were coded into a categorical variable with two factors: proficient (those who responded “very well” or “pretty well”) or not proficient (“just a little” or “not at all”). 65.11% of the participants were proficient and 34.89% of the participants were not proficient. There were 5 missing numbers: 2 people responded “Don’t know” and 3 people refused to respond and were thus excluded from analyses.

NSL$qnlan3 <- as.factor(NSL$qnlan3)  
levels(NSL$qnlan3)[levels(NSL$qnlan3)=="1"] <- "Proficient" #Very well  
levels(NSL$qnlan3)[levels(NSL$qnlan3)=="2"] <- "Proficient" #Pretty well  
levels(NSL$qnlan3)[levels(NSL$qnlan3)=="3"] <- "Not proficient" #Just a little  
levels(NSL$qnlan3)[levels(NSL$qnlan3)=="4"] <- "Not proficient" #Not at all  
levels(NSL$qnlan3)[levels(NSL$qnlan3)=="8"] <- NA  
levels(NSL$qnlan3)[levels(NSL$qnlan3)=="9"] <- NA # 8: 2, 9: 3  
  
prop.table(table(NSL$qnlan3))

##   
## Proficient Not proficient   
## 0.6510695 0.3489305

ggplot(NSL, aes(qnlan3)) +  
 geom\_bar() +  
 labs(title = "Distribution of English Language Proficiency", x = "Language Proficiency", y = "Count")



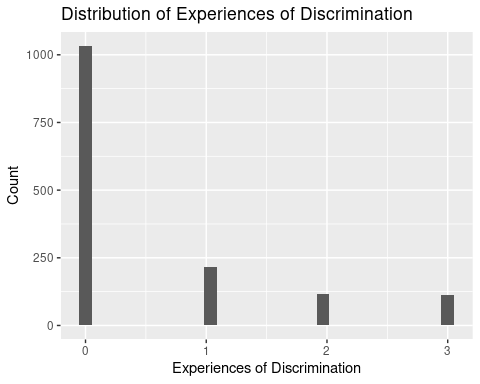
## Univariate Analyses: Predictor Variables: Experiences of Discrimination

Participants’ degree of experiences of discrimination were quantified as the response to 3 survey questions: qn23c (“In the past twelve months, have you been called offensive names because you are (Hispanic/Latino), or not?”), qn23e (“In the past twelve months, has someone made a remark that you should go back to your home country, or not?”), qn23f (“In the past twelve months, have you personally experienced discrimination or been treated unfairly because of your (Hispanic/Latino) background, or not?”). The responses to these questions were coded such that those who hadn’t experienced the discrimination corresponded to 0 and those who had corresponded to 1. The answers were then added together for each participant so that higher scores corresponded to greater experiences of discrimination. qn23d (“In the past twelve months, have you been criticized for speaking Spanish in public, or not?”) was excluded because it was only asked of Spanish language speakers, which would have thus made the combined variable exclude a large portion of the dataset. “Don’t know” and “Refused” were treated as missing data points and were excluded from subsequent analyses. Two people responded “Don’t know” to qn23c, two people responded “Don’t know” and two people “Refused” for qn23e, and nine people responded “Don’t know” and four people “Refused” for qn23f. This produced 19 missing values. The resulting variable has a right skew, meaning that participants tended to have experienced little to none of these types of discrimination. The variable had a minimum of 0, a median of 0, a maximum of 3, and an IQR of 1.

NSL$qn23c[NSL$qn23c == 2] <- 0  
NSL$qn23d[NSL$qn23d == 2] <- 0  
NSL$qn23e[NSL$qn23e == 2] <- 0  
NSL$qn23f[NSL$qn23f == 2] <- 0  
NSL$qn23c[NSL$qn23c == 8 | NSL$qn23c == 9] <- NA  
NSL$qn23d[NSL$qn23d == 8 | NSL$qn23d == 9] <- NA  
NSL$qn23e[NSL$qn23e == 8 | NSL$qn23e == 9] <- NA  
NSL$qn23f[NSL$qn23f == 8 | NSL$qn23f == 9] <- NA  
  
NSL <- NSL %>%  
 mutate(discrimination = (qn23c + qn23e + qn23f))  
  
ggplot(NSL, aes(discrimination)) +   
 geom\_histogram() +  
 labs(title = "Distribution of Experiences of Discrimination", x = "Experiences of Discrimination", y = "Count")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 19 rows containing non-finite values (stat\_bin).

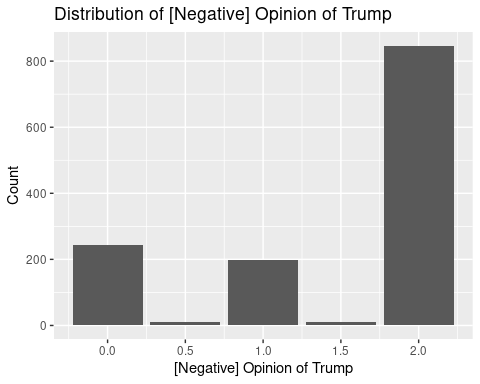


## Univariate Analyses: Outcome Variables: [Negative] Opinion of Trump

The degree that participants had a negative opinion of Donald Trump was quantified as the response to 2 survey questions: qn14a (“Do you approve or disapprove of the way Donald Trump is handling his job as President?”) and qn17 (“Overall, do you think that the Trump administration’s policies have been helpful to (Hispanics/Latinos), harmful to (Hispanics/Latinos), or have they had no particular effect on (Hispanics/Latinos)?”). Positive or neutral responses regarding Trump were coded as 0, negative ones were coded as 1, and mixed answers were coded as 0.5. The answers to the two questions were added together to generate a score. Thus, higher numbers correspond to having a more negative opinion of the president. To qn14a, 107 responded “Don’t know” and 32 “Refused,” while for qn17, 73 responded “Don’t know” and 14 “Refused.” These were excluded from analyses because we wanted to only observe those with formed opinions. This produced 191 missing values. The resulting variable had a left skew, indicating that participants tend to have a more negative opinion of Trump. The variable had a minimum of 0, a median of 2, a maximum of 2, and an IQR of 1.

NSL$qn14a[NSL$qn14a == 8 | NSL$qn14a == 9] <- NA  
NSL$qn17[NSL$qn17 == 8 | NSL$qn17 == 9] <- NA  
NSL$qn14a[NSL$qn14a == 1] <- 0  
NSL$qn14a[NSL$qn14a == 2] <- 1  
  
NSL$qn17[NSL$qn17 == 1] <- 0  
NSL$qn17[NSL$qn17 == 2] <- 1  
NSL$qn17[NSL$qn17 == 3] <- 0  
NSL$qn17[NSL$qn17 == 4] <- 0.5  
  
NSL <- NSL %>%   
 mutate(trump = (qn14a + qn17))  
ggplot(NSL, aes(trump)) +   
 geom\_bar() +  
 labs(title = "Distribution of [Negative] Opinion of Trump", x = "[Negative] Opinion of Trump", y = "Count")

## Warning: Removed 191 rows containing non-finite values (stat\_count).

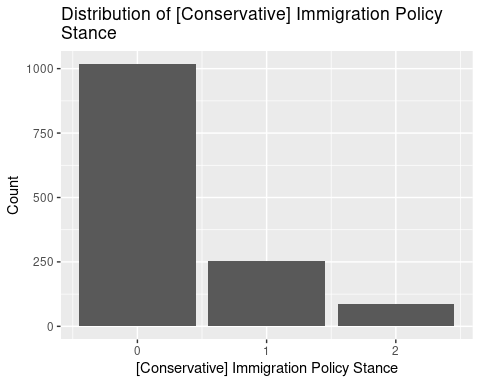


## Univariate Analyses: Outcome Variables: [Conservative] Immigration Policy Stance

The degree that participants had a conservative stance on immigration was quantified as the response to two survey questions: qn28 (“As you may know, many immigrants who came illegally to the U.S. when they were children now have temporary legal status that may be ending. Would you favor or oppose Congress passing a law granting them permanent legal status?”) and qn29 (“As you may know, there is a proposal to substantially expand the wall along the U.S. border with Mexico. In general, do you favor or oppose this proposal?”). Answers were coded so that higher numbers would indicate a more conservative stance and for each participant, their answer to each question was added together. The resulting variable had a right skew, indicating that participants tended not to have a particularly conservative stance on immigration. The variable had a minimum of 0, a median of 0, a maximum of 2, and an IQR of 1. There were 141 missing values because for certain sections of the survey, not all participants were asked all of the questions. This means that despite the high number of missing data, this should not skew our analyses because the questions a person was asked were randomized.

NSL$qn28[NSL$qn28 == 8 | NSL$qn28 == 9] <- NA  
NSL$qn29[NSL$qn29 == 8 | NSL$qn29 == 9] <- NA  
NSL$qn28[NSL$qn28 == 1] <- 0  
NSL$qn28[NSL$qn28 == 2] <- 1  
NSL$qn29[NSL$qn29 == 2] <- 0  
  
NSL <- NSL %>%  
 mutate(immigration = (qn28 + qn29))  
  
ggplot(NSL, aes(immigration)) +   
 geom\_bar() +  
 labs(title = "Distribution of [Conservative] Immigration Policy\nStance", x = "[Conservative] Immigration Policy Stance", y = "Count")

## Warning: Removed 141 rows containing non-finite values (stat\_count).



sample\_NSL <- NSL %>%   
 select(citizen = qn9, language = qnlan3, discrimination, trump, party, immigration)  
  
#making a dataset with no missing data  
NSL\_clean <- sample\_NSL %>%   
 drop\_na()  
  
#write.csv(NSL\_clean, file = "NSL\_clean")

# Model: [Negative] Opinion of Trump

## Bivariate Analyses: [Negative] Opinion of Trump

All three levels of citizenship status had the same median opinion of Trump and IQR (median = 2, IQR = 1), suggesting that there is not much difference between the groups and that all groups tend to have a more negative opinion of Trump. This does not support my hypothesis that there would be a difference between these groups, with naturalized citizens and non-citizens having more negative opinions of Trump than natural born citizens.

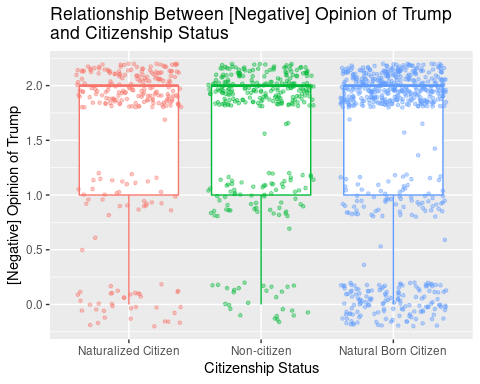
Both levels of English language proficiency had the same median opinion of Trump and IQR (median = 2, IQR = 1), suggesting that there is not much difference between the groups and that all groups tend to have a more negative opinion of Trump. This did not support my hypothesis that these two groups are different and that not proficient participants would have a more negative view.

There was a small-medium positive correlation between experiences of discrimination and opinion of Trump (r = 0.1986), meaning that greater experiences of discrimination were correlated with a more negative opinion of Trump, which is in alignment with the hypothesis regarding this relationship.

#Citizenship Status (categorical)  
NSL\_clean %>% group\_by(citizen) %>%  
 summarize(median = median(trump, na.rm = TRUE),  
 min = min(trump, na.rm = TRUE),  
 max = max(trump, na.rm = TRUE),  
 IQR = IQR(trump, na.rm = TRUE))

## # A tibble: 3 × 5  
## citizen median min max IQR  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 Naturalized Citizen 2 0 2 1  
## 2 Non-citizen 2 0 2 1  
## 3 Natural Born Citizen 2 0 2 1

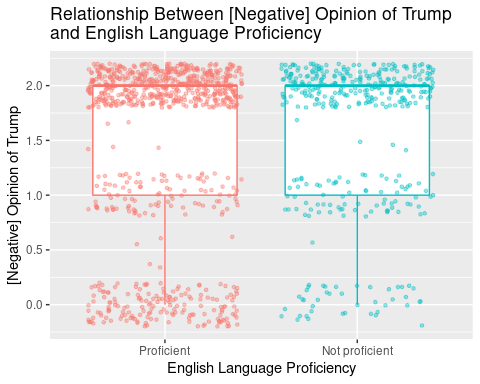
ggplot(NSL\_clean, aes(citizen, trump, color = citizen)) +  
 geom\_boxplot() +   
 geom\_jitter(size = 1, alpha = .4) +  
 theme(legend.position = "none") +  
 labs(title = "Relationship Between [Negative] Opinion of Trump\nand Citizenship Status", x = "Citizenship Status", y = "[Negative] Opinion of Trump")



#English Language Proficiency (categorical)  
NSL\_clean %>% group\_by(language) %>%  
 summarize(median = median(trump, na.rm = TRUE),  
 min = min(trump, na.rm = TRUE),  
 max = max(trump, na.rm = TRUE),  
 IQR = IQR(trump, na.rm = TRUE))

## # A tibble: 2 × 5  
## language median min max IQR  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 Proficient 2 0 2 1  
## 2 Not proficient 2 0 2 1

ggplot(NSL\_clean, aes(language, trump, color = language)) +  
 geom\_boxplot() +   
 geom\_jitter(size = 1, alpha = .4) +  
 theme(legend.position = "none") +  
 labs(title = "Relationship Between [Negative] Opinion of Trump\nand English Language Proficiency", x = "English Language Proficiency", y = "[Negative] Opinion of Trump")

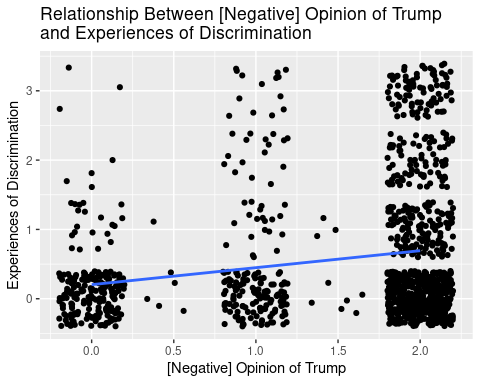


#Experiences of Discrimination (numerical)  
cor(NSL\_clean$trump, NSL\_clean$discrimination, use = "complete.obs")

## [1] 0.1988358

ggplot(NSL\_clean, aes(trump, discrimination)) +   
 geom\_jitter() +  
 geom\_smooth(method = "lm", se = FALSE) +   
 labs(title = "Relationship Between [Negative] Opinion of Trump\nand Experiences of Discrimination", x = "[Negative] Opinion of Trump", y = "Experiences of Discrimination")

## `geom\_smooth()` using formula 'y ~ x'



## Regression Model: [Negative] Opinion of Trump

For the outcome variable [negative] opinion of Trump, I compared three models. The first model using just the three predictor variables had an r-squared value of 0.05131, meaning that it accounted for about 5.13% of variation in scores regarding participants’ opinion of Trump. According to Cohen’s guide, this is a small-medium effect size (between 1% and 9%). The model overall had a p-value of 1.329e-13 and an AIC score of 2713. Since opinion of Trump is left skewed and experiences of discrimination is right skewed, another model was made, replacing opinion of Trump with the natural log of reflected opinion of Trump and by using the natural log of experiences of discrimination. These transformations improved the AIC score from 2713.842 to 1385.622. A third model removed language score as a predictor variable because it was not a significant difference between being and not being proficient in English in either the first (p = 0.65525) or second (0.81067) models. The reduced model only slightly improved the model’s AIC score (from 1385.622 to 1383.627), which suggests that the models are not reliably different. Since the third model is the most parsimonious (due to the exclusion of language) and had the best AIC score, this is the model I will elaborate on in this analysis.

The model had an r-squared value of 0.05079, meaning that it accounted for about 5.079% of variation in scores regarding participants’ opinion of Trump (p = 5.028e-14). According to Cohen’s guide, this is a small-medium effect size (between 1% and 9%). The intercept was significant, indicating that when a participant is a natural born citizen and the natural log of experiences of discrimination score is 0 then the natural log of the participant’s reflected opinion of Trump score is predicted to be 0.40760 (p < 2e-16). Thus, this model predicts that by default – natural born citizens without experiences of discrimination, participants tend not to have a very favorable opinion of Trump. When all else is held constant, an increase of 1 in the natural log of experiences of discrimination score is expected to produce a 0.180587 decrease in the natural log of the reflected opinion of Trump (p = 4.23e-12), indicating that the model predicts increased experiences of discrimination will increase negative opinions of Trump. When all else is held constant, a change from a participant being a natural born citizen to being a naturalized citizen is anticipated to produce a 0.10089 decrease in the natural log of the reflected opinion of Trump (p = 0.00169) and a change from a natural born citizen to being a non-citizen is anticipated to produce a 0.09566 decrease in the natural log of the reflected opinion of Trump (p = 0.00143). Thus, the model predicts that not being a natural born citizen will increase one’s negative opinion of Trump.

The Residuals vs. Fitted showed that the residuals appeared to be more positive than negative, suggesting a slight deviation from the assumptions of Mean of Residuals is ~0 and Constant Variability.

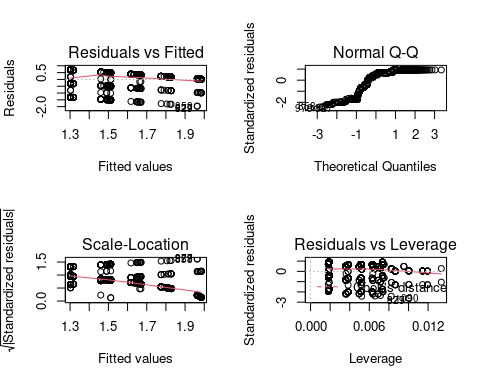
The Normal Q-Q suggested some deviation from the assumption that Residuals are Nearly Normal – there were light right and left tails, and deviations both above and below the 45° line in the center.

In the Scale-Location plot, the line was not very flat and the points were not equally spread, also suggesting some deviation from the assumption of Constant Variability.

The Residuals vs. Leverage plot showed no data points beyond Cook’s distance, indicating that that assumption of No Bad (high leverage) Outliers was met. As a result, no trimming and Winsorizing of the data was attempted to improve the model.

Overall, however, this model had difficulty meeting the assumptions of linear regression.

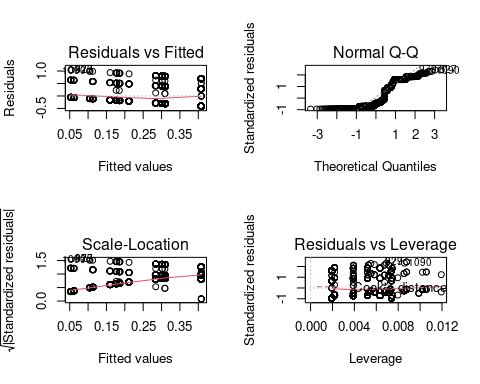
NSL\_clean$citizen <- relevel(NSL\_clean$citizen, "Natural Born Citizen")  
  
modelfit\_trump\_1 <- lm(formula = trump ~ citizen + language + discrimination, data = NSL\_clean)  
par(mfrow = c(2, 2))  
plot(modelfit\_trump\_1)



summary(modelfit\_trump\_1)

##   
## Call:  
## lm(formula = trump ~ citizen + language + discrimination, data = NSL\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.9633 -0.5059 0.3823 0.5398 0.6973   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.30269 0.03297 39.507 < 2e-16 \*\*\*  
## citizenNaturalized Citizen 0.18808 0.05926 3.174 0.00154 \*\*   
## citizenNon-citizen 0.19431 0.06736 2.885 0.00399 \*\*   
## languageNot proficient 0.01514 0.06047 0.250 0.80237   
## discrimination 0.15751 0.02284 6.896 8.64e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7566 on 1199 degrees of freedom  
## Multiple R-squared: 0.05607, Adjusted R-squared: 0.05292   
## F-statistic: 17.8 on 4 and 1199 DF, p-value: 3.284e-14

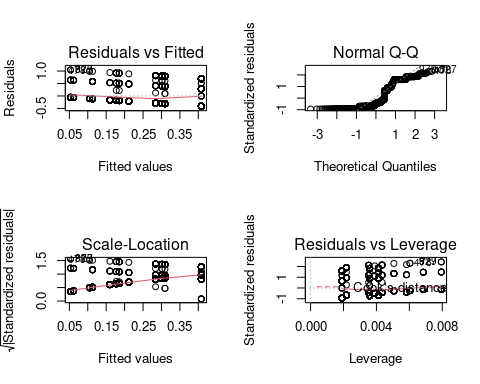
modelfit\_trump\_2 <- lm(formula = log(3 - trump) ~ citizen + language + log(discrimination + 1), data = NSL\_clean)  
par(mfrow = c(2, 2))  
plot(modelfit\_trump\_2)



summary(modelfit\_trump\_2)

##   
## Call:  
## lm(formula = log(3 - trump) ~ citizen + language + log(discrimination +   
## 1), data = NSL\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.4082 -0.3092 -0.2115 0.3840 1.0448   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.408242 0.019050 21.430 < 2e-16 \*\*\*  
## citizenNaturalized Citizen -0.106187 0.033871 -3.135 0.00176 \*\*   
## citizenNon-citizen -0.097544 0.038508 -2.533 0.01143 \*   
## languageNot proficient -0.001522 0.034555 -0.044 0.96488   
## log(discrimination + 1) -0.179042 0.025684 -6.971 5.19e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4324 on 1199 degrees of freedom  
## Multiple R-squared: 0.05372, Adjusted R-squared: 0.05056   
## F-statistic: 17.02 on 4 and 1199 DF, p-value: 1.395e-13

modelfit\_trump\_3 <- lm(formula = log(3 - trump) ~ citizen + log(discrimination + 1), data = NSL\_clean)  
par(mfrow = c(2, 2))  
plot(modelfit\_trump\_3)



summary(modelfit\_trump\_3)

##   
## Call:  
## lm(formula = log(3 - trump) ~ citizen + log(discrimination +   
## 1), data = NSL\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.4081 -0.3095 -0.2115 0.3836 1.0453   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.40813 0.01887 21.626 < 2e-16 \*\*\*  
## citizenNaturalized Citizen -0.10670 0.03179 -3.356 0.000816 \*\*\*  
## citizenNon-citizen -0.09862 0.02982 -3.307 0.000970 \*\*\*  
## log(discrimination + 1) -0.17899 0.02564 -6.980 4.86e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4323 on 1200 degrees of freedom  
## Multiple R-squared: 0.05372, Adjusted R-squared: 0.05135   
## F-statistic: 22.71 on 3 and 1200 DF, p-value: 2.661e-14

AIC(modelfit\_trump\_1) #I was helping my friend with her own R project (for another class) and saw this is how she calculated AIC

## [1] 2752.14

AIC(modelfit\_trump\_2)

## [1] 1405.083

AIC(modelfit\_trump\_3)

## [1] 1403.085

## Bivariate Analyses: [Conservative] Immigration Policy Stance

All three levels of citizenship status had the same median immigration policy stance (median = 0), though there was a different IQR for natural born citizens (IQR = 1) than naturalized and non-citizens (IQR = 0), suggesting that there is not much difference between the groups, with all groups tending to not have conservative immigration policy stances, though there may be more variability for natural born citizens. While this does not support our hypothesis that naturalized and non-citizens will have less conservative stances than natural born citizens, it leaves the possibility that this could potentially be a finding with a larger sample size.

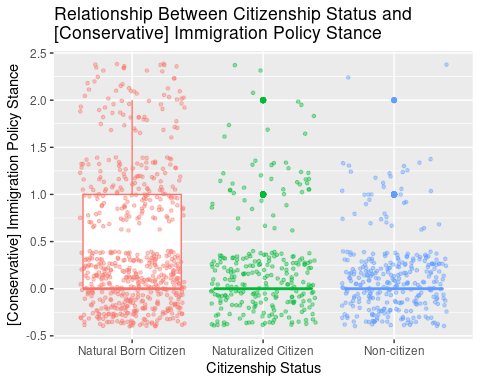
Both levels of English language proficiency had the same median immigration policy stance (median = 0), though there was more variability in those who were proficient in English (IQR = 1) compared to those who were not proficient (IQR = 0). This suggests that there is not much difference between the groups, with both groups tending not to have conservative immigration policy stances, though there may be more variability for those proficient in English. Like with citizenship status, while this does not support our hypothesis that those who are not proficient in English will have less conservative stances than those who are, it leaves the possibility that this could be a finding with a larger sample size.

There was a weak negative correlation between experiences of discrimination and immigration policy stance (r = -0.1425), meaning that greater experiences of discrimination may be weakly related to not having as conservative of an immigration policy stance, which would support the hypothesis about this relationship.

#Citizenship Status (categorical)  
NSL\_clean %>% group\_by(citizen) %>%  
 summarize(median = median(immigration, na.rm = TRUE),  
 min = min(immigration, na.rm = TRUE),  
 max = max(immigration, na.rm = TRUE),  
 IQR = IQR(immigration, na.rm = TRUE))

## # A tibble: 3 × 5  
## citizen median min max IQR  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 Natural Born Citizen 0 0 2 1  
## 2 Naturalized Citizen 0 0 2 0  
## 3 Non-citizen 0 0 2 0

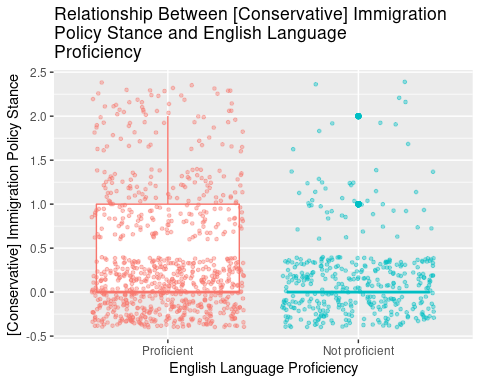
ggplot(NSL\_clean, aes(x = reorder(citizen, immigration, FUN = median), y = immigration, color = citizen)) +  
 geom\_boxplot() +   
 geom\_jitter(size = 1, alpha = .4) +  
 theme(legend.position = "none") +  
 labs(title = "Relationship Between Citizenship Status and\n[Conservative] Immigration Policy Stance", x = "Citizenship Status", y = "[Conservative] Immigration Policy Stance")



#English Language Proficiency (numerical)  
NSL\_clean %>% group\_by(language) %>%  
 summarize(median = median(immigration, na.rm = TRUE),  
 min = min(immigration, na.rm = TRUE),  
 max = max(immigration, na.rm = TRUE),  
 IQR = IQR(immigration, na.rm = TRUE))

## # A tibble: 2 × 5  
## language median min max IQR  
## <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 Proficient 0 0 2 1  
## 2 Not proficient 0 0 2 0

ggplot(NSL\_clean, aes(language, immigration, color = language)) +   
 geom\_boxplot() +   
 geom\_jitter(size = 1, alpha = .4) +  
 theme(legend.position = "none") +  
 labs(title = "Relationship Between [Conservative] Immigration\nPolicy Stance and English Language\nProficiency", x = "English Language Proficiency", y = "[Conservative] Immigration Policy Stance")

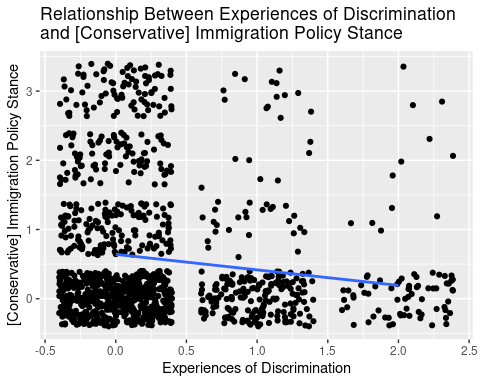


#Experiences of Discrimination (numerical)  
cor(NSL\_clean$immigration, NSL\_clean$discrimination, use = "complete.obs")

## [1] -0.1389922

ggplot(NSL\_clean, aes(immigration, discrimination)) +   
 geom\_jitter() +  
 geom\_smooth(method = "lm", se = FALSE) +   
 labs(title = "Relationship Between Experiences of Discrimination\nand [Conservative] Immigration Policy Stance", x = "Experiences of Discrimination", y = "[Conservative] Immigration Policy Stance")

## `geom\_smooth()` using formula 'y ~ x'



## Regression Model: [Conservative] Immigration Policy Stance

For the outcome variable [conservative] immigration policy stance, I compared three models. The first model using just the three predictor variables had an r-squared value of 0.08266, meaning that it accounted for about 8.266% of variation in scores regarding participants’ immigration policy stance. According to Cohen’s guide, this is approaching a medium effect size (~9%). The model overall had a p-value of < 2.2e-16 and an AIC score of 2041.644. Since both immigration policy stance and experiences of discrimination are right skewed, another model was made, replacing immigration policy stance with the natural log of immigration policy stance and by using the natural log of experiences of discrimination. These transformations improved the AIC score from 2041.644 to 845.8828. A third model removed language score as a predictor variable because it was not a significant difference between being and not being proficient in English in either the first (p = 0.0991) or second (0.0587) models. The reduced model had a slightly worse AIC score (from 845.8828 to 847.4731), but the scores are so similar that the models are likely not reliably different. Thus, since the third model contains only significant variables, this is the model I will elaborate on in this analysis.

The model had an r-squared value of 0.08518, meaning that it accounted for about 8.518% of variation in scores regarding participants’ immigration policy stance (p < 2.2e-16). According to Cohen’s guide, this is an approximately medium effect size (~9%). The intercept was significant, indicating that when a participant is a natural born citizen and the natural log of experiences of discrimination score is 0 then the participant’s immigration policy stance score is predicted to be 0.32320 (p < 2e-16), indicating that by default – natural born citizens without experiences of discrimination – participants did not have very conservative stances on immigration. When all else is held constant, an increase of 1 in the natural log of experiences of discrimination score is expected to produce a 0.10703 decrease in the natural log of immigration policy stance (p = 2.10e-07), meaning that in this model greater experiences of discrimination predicted less conservative immigration stances. When all else is held constant, a change from a participant being a natural born citizen to being a naturalized citizen is anticipated to produce a 0.14617 decrease in the natural log of immigration policy stance (p = 1.29e-08) and a change from a natural born citizen to being a non-citizen is anticipated to produce a 0.20107 decrease in the natural log of immigration policy stance score (< 2e-16). This means that in this model, being a naturalized citizen and being a non-citizen is predicted to produce a less conservative immigration policy stance compared to natural born citizens.

The Residuals vs. Fitted showed that the residuals appeared to be more negative than positive and that there was an uneven distribution of the residuals above and below the red regression line, suggesting a deviation from the assumptions of Mean of Residuals is ~0 and Constant Variability.

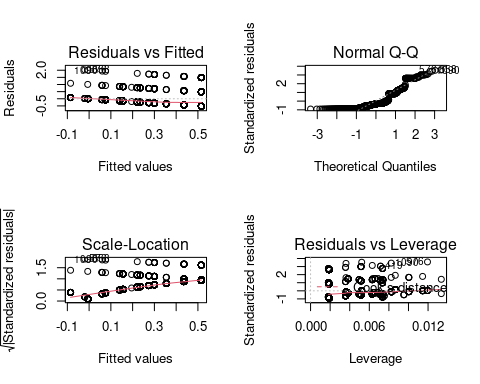
The Normal Q-Q suggested some deviation from the assumption that Residuals are Nearly Normal – there were light right and left tails, and a gap towards the center of the 45° line.

In the Scale-Location plot, the line was not very flat and the points were not equally spread, also suggesting some deviation from the assumption of Constant Variability.

The Residuals vs. Leverage plot showed no data points beyond Cook’s distance, indicating that that assumption of No Bad (high leverage) Outliers was met. As a result, no trimming and Winsorizing of the data was attempted to improve the model.

Overall, however, this model had difficulty meeting the assumptions of linear regression.

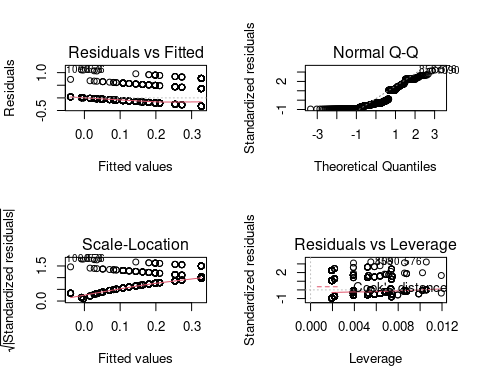
modelfit\_immigration\_1 <- lm(formula = immigration ~ citizen + language + discrimination, data = NSL\_clean)  
par(mfrow = c(2, 2))  
plot(modelfit\_immigration\_1)



summary(modelfit\_immigration\_1)

##   
## Call:  
## lm(formula = immigration ~ citizen + language + discrimination,   
## data = NSL\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.5156 -0.4351 -0.1562 0.4844 2.0201   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.51561 0.02476 20.824 < 2e-16 \*\*\*  
## citizenNaturalized Citizen -0.21543 0.04450 -4.841 1.46e-06 \*\*\*  
## citizenNon-citizen -0.28057 0.05058 -5.547 3.58e-08 \*\*\*  
## languageNot proficient -0.07886 0.04541 -1.737 0.0827 .   
## discrimination -0.08046 0.01715 -4.691 3.03e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5681 on 1199 degrees of freedom  
## Multiple R-squared: 0.08572, Adjusted R-squared: 0.08267   
## F-statistic: 28.11 on 4 and 1199 DF, p-value: < 2.2e-16

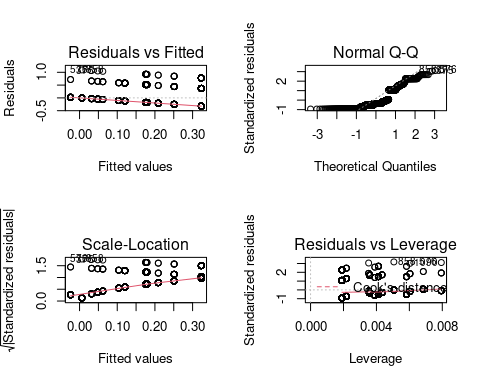
modelfit\_immigration\_2 <- lm(formula = log(immigration + 1) ~ citizen + language + log(discrimination + 1), data = NSL\_clean)  
par(mfrow = c(2, 2))  
plot(modelfit\_immigration\_2)



summary(modelfit\_immigration\_2)

##   
## Call:  
## lm(formula = log(immigration + 1) ~ citizen + language + log(discrimination +   
## 1), data = NSL\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.3271 -0.2537 -0.1082 0.3660 1.1069   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.32711 0.01516 21.575 < 2e-16 \*\*\*  
## citizenNaturalized Citizen -0.12882 0.02696 -4.779 1.98e-06 \*\*\*  
## citizenNon-citizen -0.16446 0.03065 -5.366 9.64e-08 \*\*\*  
## languageNot proficient -0.05447 0.02750 -1.981 0.0479 \*   
## log(discrimination + 1) -0.10597 0.02044 -5.184 2.54e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3442 on 1199 degrees of freedom  
## Multiple R-squared: 0.09001, Adjusted R-squared: 0.08697   
## F-statistic: 29.65 on 4 and 1199 DF, p-value: < 2.2e-16

modelfit\_immigration\_3 <- lm(formula = log(immigration + 1) ~ citizen + log(discrimination + 1), data = NSL\_clean)  
par(mfrow = c(2, 2))  
plot(modelfit\_immigration\_3)



summary(modelfit\_immigration\_3)

##   
## Call:  
## lm(formula = log(immigration + 1) ~ citizen + log(discrimination +   
## 1), data = NSL\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.3231 -0.2511 -0.1203 0.3700 1.0925   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.32310 0.01504 21.477 < 2e-16 \*\*\*  
## citizenNaturalized Citizen -0.14717 0.02535 -5.807 8.16e-09 \*\*\*  
## citizenNon-citizen -0.20285 0.02377 -8.534 < 2e-16 \*\*\*  
## log(discrimination + 1) -0.10394 0.02044 -5.085 4.26e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3446 on 1200 degrees of freedom  
## Multiple R-squared: 0.08703, Adjusted R-squared: 0.08475   
## F-statistic: 38.13 on 3 and 1200 DF, p-value: < 2.2e-16

AIC(modelfit\_immigration\_1)

## [1] 2062.38

AIC(modelfit\_immigration\_2)

## [1] 855.2996

AIC(modelfit\_immigration\_3)

## [1] 857.2327