Report 2

Naomi Zubeldia

2022-12-12

Abstract

#Say what was found from the study

Contents

1 Introduction

Driving is a normal part of people's every day lives. However, one must keep the laws of the land in order to keep the privilege of driving. There is a police organization that patrols and makes sure everyone is keeping the driving laws. If not, then you may be pulled over and fined, which is all up to the digression of the police officer that pulled you over. Unfortunately, sometimes police officers may give certain people fines or citations and others not all based on profiling. We live in a democratic society and want to avoid any profiling when trying to deal with people fairly. This analysis focused on testing data to see if police officers in the Montgomery, Maryland area were fairly giving out citations to people that they pulled over. The factors being looked at specifically are if the person was wearing a seat belt, their gender, their race, and the car color. The focused scientific question for this analysis is to see which race and gender of a person is more likely to get a citation based on their car color and if they were wearing a seat belt or not.

The first step to find the answer will be to clean up the data from the Montgomery police department's record of pull overs to get a more summarized set. The reason for this is the dataset is very large, which makes it harder to focus on the scientific question. Also, run time on a local machine is often very long because it cannot handle the size of the set. Next, the summarized data will be fit to an appropriate linear model to be run through the anova function.

From the results of the anova, the most significant factors will be identified. Also, interactions between factors can be found and evaluated to see if they are significant as well. From that point, the significant values can be used to pin point specific gender/race combinations that are more likely to be given a citation based on their car color and if they were wearing a seat belt. The function called "GLM Tree" will be used to make that determination quick and efficiently compared to just splitting up the data by hand. Lastly, from the results of the glm tree, the probabilities of each gender/ race combination can be found to help make conclusions.

The overall hypotheses of this analysis are that all the gender/race combinations are the same for any car color and belt wearing status according to the null hypothesis. The alternative hypothesis claims that at least one of the combinations is different from the rest.Therefore, the null and alternative hypotheses are as follows:

Null: $H_0 : \mu_{Race/Gender/Belts/CarColor} = \mu$

Alternative: $H_a : \mu_{Race/Gender/Belts/CarColor} \neq \mu$

2 Data

The Montgomery police department data is a large dataset with around 800,000 values. It has over 10 different factors as well to describe each pull over. The time of the data starts around 2014. It seems to be updated many times throughout the year. The set was summarized to specifically look into the scientific question.

The summary set contains the factors of citations, race, gender, belts, car color, and counts. The citations column originated from the 'Traffic Violations' column with a filter that indicated 0s and 1s for if a person received a citation or not. The gender column was cleaned up by removing any unknown and NA values.The car color column was filtered down to just colors: blue, black, brown, green, red, silver/white, and other. These colors are all pretty common, which is why so many are kept to answer the scientific question as specifically as possible. The last column of counts was created to condense the set by counting the number of each citation/race/gender/belts/carcolor combinations there are.

The following table shows some basic summary statistics of the summarized data to be familiar with the set that is being used throughout this analysis:

Table 1: Figure 1. Summary Statistics

	Citation	Race	Gender	Belts	CarColor	count
	Min. :0.0000	Length:371	Length:371	Length:371	Length:371	Min. : 1.0
	1st Qu.:0.0000	Class :character	Class :character	Class :character	Class :character	1st Qu.: 77.5
	Median :0.0000	Mode :character	Mode :character	Mode :character	Mode :character	Median : 339.0
	Mean :0.4987					Mean : 2238.0
	3rd Qu.:1.0000					3rd Qu.: 2465.5
	Max. :1.0000					Max. :20139.0

3 Testing

3.1 Anova

The first method used for testing is logistic regression on a linear model of the data through the anova function in R. The citation column was set up to be the response with the other factors labeled as explanatory variables except for the counts column, which is the weights of the model. To first try and fit the data, the glm model was set to have three way interactions. From the anova output, the degrees of freedom were not correct for some of the interactions. Due to them being smaller than what they should have been, the model was reduced down to just having two way interactions, which fit well since the degrees of freedom were right.

According to the anova output of the model, all the factors and interactions are significant. The Gender, Belts, and Race factors are the most significant out of all of them, which was found by dividing the degrees of freedom into the deviance. Most of the interactions are not very significant, but those that include Gender, Belts, or Race are the most out of all of them.

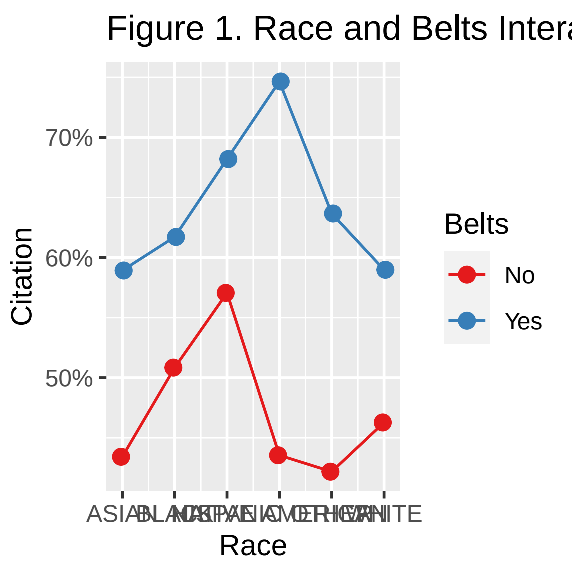
The following output comes from the anova of the best fit glm to observe the significant factors:

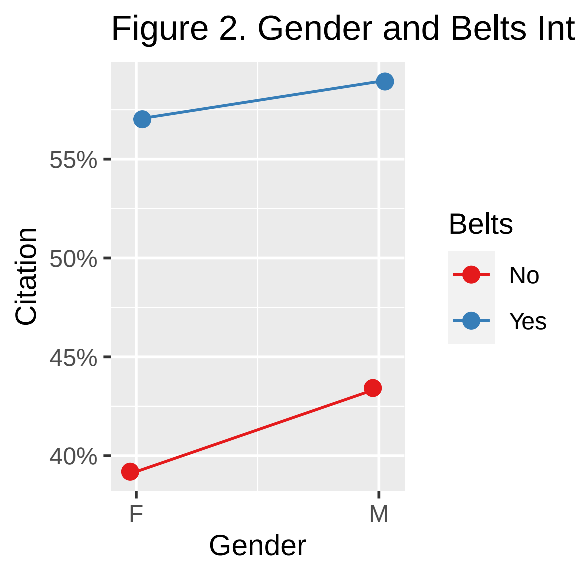
Table 2: Anova Output

	Df	Deviance	Resid. Df	Resid. Dev
NULL			370	1149278
Race	5	6418.10227	365	1142860
Gender	1	3064.57956	364	1139795
Belts	1	2045.79357	363	1137749
CarColor	7	260.67911	356	1137489
Race:Gender	5	249.75219	351	1137239
Race:Belts	5	91.04359	346	1137148
Race:CarColor	35	200.42301	311	1136947
Gender:Belts	1	15.67753	310	1136932
Gender:CarColor	7	93.91764	303	1136838
Belts:CarColor	7	87.51699	296	1136750

3.2 Observing Interactions

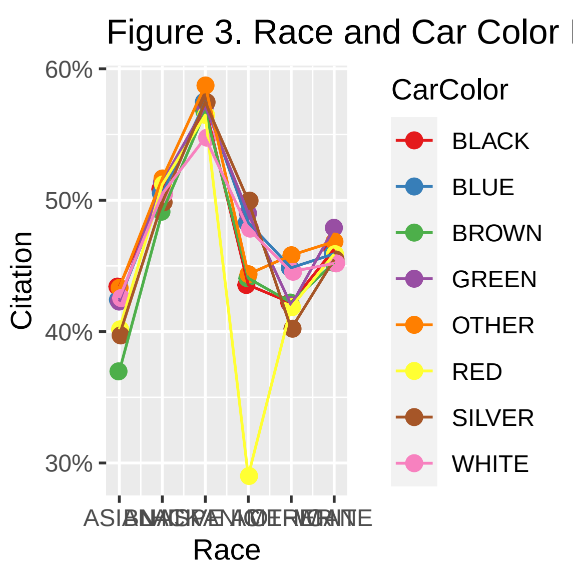
The interactions between the factors were slightly significant indicating that there may be some influence between factors. To visualize the interactions, the following outputs show the most interesting ones. (To see all interaction plots, please refer to the appendix)





In figure one, the lines look mostly parallel except for a couple races. It is interesting to see the drop for both genders as Native Americans, which may be due to the lack of observations that include them. Figure two examines the interaction between gender and belt wearing, which looks like there is a bit of interaction but it does not seem very compelling.

The interactions that include car color were not very significant but their plots are very intersting.



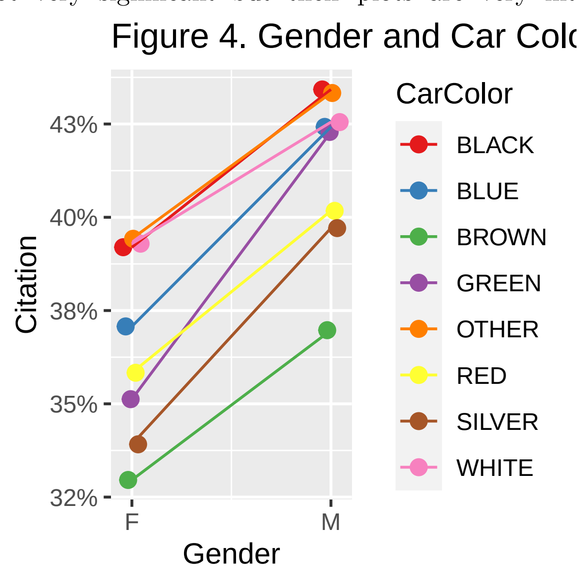


Figure three shows the interaction of car color and race, which looks to be very significant for specifically red, silver, and maybe white. The next graph, figure four, shows multiple interactions of car color across the genders. These interactions seem to appear more significant than the others, which may be because of the amount of observations per color. To examine the data better without the effect of car color, the next step is to observe the data through a glm decision tree.

3.3 GLM Decision Trees

A glm decision tree is a machine learning algorithm that can perform tasks such as regression and classification (cite). With the tree, the interactions can be visualized better. To compute this method, the following R code was used:

tree <- glmtree(Citation~ Race+Gender|Belts+CarColor, data = test1, weights=count,family = binomial)

Below is the tree that resulted from the R code above:

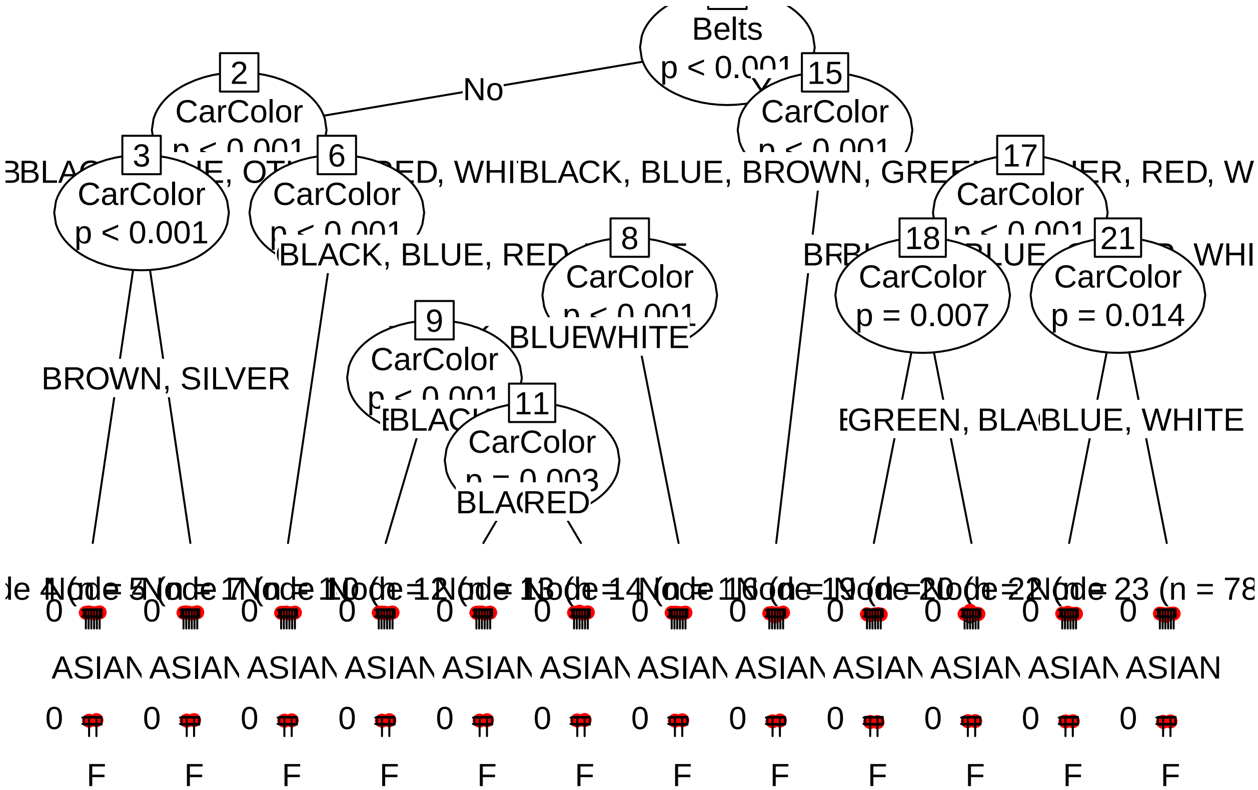


Figure 1: Whole Tree of Summarized Data

```
## NULL
```

From this particular model, the resulting tree in table 3, seen above, was very large with many splits. Most of the splits were on car color, which we saw from the interaction graphs previously. Since the tree is very large, it was best to split up the data into different car colors and create mini trees for each color. The data was then easier to observe from this split. The color trees follow the same format as the R code listed above. From the car color glm tree models, predicted values were computed through the predict function for each combination of the factors. The predicted values represent the probability of getting a new observation(cite). In other words, they are the probability that a person of a certain race and color in a certain colored car while wearing or not wearing a seat belt will get a traffic citation.

The following figures are the observed values and probabilities for each car color for individuals who were given a citation. These figures only contain the max and min values for each race. Each color table was included to allow for better comprehension to the reader of whihc observation is the most likely. To see each whole table, please refer to the appendix.

Table 3: Black Car Probabilities

Citation	Race	Gender	Belts	CarColor	count	probability
1	ASIAN	M	Yes	BLACK	121	0.6245274
1	ASIAN	F	No	BLACK	1074	0.3789883
1	BLACK	M	Yes	BLACK	793	0.6205686
1	BLACK	F	No	BLACK	7978	0.4466044
1	HISPANIC	M	Yes	BLACK	635	0.6927856
1	HISPANIC	F	No	BLACK	3285	0.4986237
1	NATIVE AMERICAN	M	Yes	BLACK	2	0.4395836
1	NATIVE AMERICAN	F	No	BLACK	46	0.3707449
1	OTHER	M	Yes	BLACK	197	0.6870684
1	OTHER	F	No	BLACK	919	0.3643949
1	WHITE	M	Yes	BLACK	866	0.5723086
1	WHITE	F	No	BLACK	8392	0.4068233

Table 4: Red Car Probabilities

Citation	Race	Gender	Belts	CarColor	count	probability
1	ASIAN	M	Yes	RED	33	0.6560052
1	ASIAN	F	No	RED	444	0.3453505
1	BLACK	M	Yes	RED	254	0.6177286
1	BLACK	F	No	RED	2898	0.4473295
1	HISPANIC	M	Yes	RED	352	0.6561792
1	HISPANIC	F	No	RED	1649	0.4911940
1	NATIVE AMERICAN	M	No	RED	28	0.2975511
1	NATIVE AMERICAN	M	Yes	RED	1	0.2063052
1	OTHER	M	Yes	RED	72	0.6263557
1	OTHER	F	No	RED	380	0.3620838
1	WHITE	M	Yes	RED	351	0.6484654
1	WHITE	F	No	RED	3433	0.3973017

Table 5: Blue Car Probabilities

Citation	Race	Gender	Belts	CarColor	count	probability
1	ASIAN	M	Yes	BLUE	72	0.5933679
1	ASIAN	F	No	BLUE	669	0.3600859
1	BLACK	M	Yes	BLUE	324	0.6130450
1	BLACK	F	No	BLUE	3621	0.4316421
1	HISPANIC	M	Yes	BLUE	331	0.6603083
1	HISPANIC	F	No	BLUE	1882	0.4918565
1	NATIVE AMERICAN	M	Yes	BLUE	1	0.9999975
1	NATIVE AMERICAN	F	No	BLUE	53	0.3937404
1	OTHER	M	Yes	BLUE	107	0.6582547
1	OTHER	F	No	BLUE	594	0.3813396
1	WHITE	M	Yes	BLUE	493	0.5895704
1	WHITE	F	No	BLUE	5684	0.3913720

Table 6: Brown Car Probabilities

Citation	Race	Gender	Belts	CarColor	count	probability
1	ASIAN	F	Yes	BROWN	27	0.7122544
1	ASIAN	F	No	BROWN	283	0.3138642
1	BLACK	F	Yes	BROWN	86	0.6390146
1	BLACK	F	No	BROWN	1834	0.4284262
1	HISPANIC	F	Yes	BROWN	43	0.6886288
1	HISPANIC	F	No	BROWN	771	0.4948705
1	NATIVE AMERICAN	M	No	BROWN	47	0.4345822
1	NATIVE AMERICAN	F	No	BROWN	6	0.3718195
1	OTHER	F	Yes	BROWN	33	0.8240154
1	OTHER	F	No	BROWN	240	0.3612563
1	WHITE	F	Yes	BROWN	115	0.6808524
1	WHITE	F	No	BROWN	1856	0.3936999

Table 7: Green Car Probabilities

Citation	Race	Gender	Belts	CarColor	count	probability
1	ASIAN	M	Yes	GREEN	37	0.6303863
1	ASIAN	F	No	GREEN	267	0.3347791
1	BLACK	M	Yes	GREEN	188	0.5705228
1	BLACK	F	No	GREEN	1918	0.4182191
1	HISPANIC	M	Yes	GREEN	280	0.6466845
1	HISPANIC	F	No	GREEN	975	0.4666245
1	NATIVE AMERICAN	M	No	GREEN	50	0.4982580
1	NATIVE AMERICAN	F	No	GREEN	13	0.4013301
1	OTHER	M	Yes	GREEN	29	0.6414319
1	OTHER	F	No	GREEN	233	0.3324411
1	WHITE	M	Yes	GREEN	179	0.5842967
1	WHITE	F	No	GREEN	2122	0.3877416

Table 8: Silver/White Car Probabilities

Citation	Race	Gender	Belts	CarColor	count	probability
1	ASIAN	M	Yes	SILVER	93	0.4489735
1	ASIAN	F	No	SILVER	1440	0.3312793
1	ASIAN	M	Yes	WHITE	98	0.5780375
1	ASIAN	F	No	WHITE	1144	0.3827702
1	BLACK	M	Yes	SILVER	490	0.5738186
1	BLACK	F	No	SILVER	7399	0.4185749
1	BLACK	M	Yes	WHITE	435	0.6091684
1	BLACK	F	No	WHITE	4471	0.4532772
1	HISPANIC	M	Yes	SILVER	485	0.6405374
1	HISPANIC	F	No	SILVER	3276	0.4834289
1	HISPANIC	M	Yes	WHITE	608	0.6738778
1	HISPANIC	F	No	WHITE	2394	0.4884773
1	NATIVE AMERICAN	M	Yes	SILVER	10	0.9999993
1	NATIVE AMERICAN	F	No	SILVER	53	0.4032375
1	NATIVE AMERICAN	M	Yes	WHITE	30	0.9378495
1	NATIVE AMERICAN	F	No	WHITE	27	0.4060548
1	OTHER	M	Yes	SILVER	76	0.5074732
1	OTHER	F	No	SILVER	978	0.3340349
1	OTHER	M	Yes	WHITE	102	0.6258660
1	OTHER	F	No	WHITE	600	0.4006353
1	WHITE	M	Yes	SILVER	517	0.5486000
1	WHITE	F	No	SILVER	8066	0.3810615
1	WHITE	M	Yes	WHITE	580	0.5862353
1	WHITE	F	No	WHITE	5837	0.4047554

4 Results

After creating decision trees for each car color, we found the probability for each combination of the tested factors.

The highest probabilities for people driving black cars were mostly men who were wearing their seat belts. The top probability was a Hispanic male where a seat belt at .6927. Other high probabilities were Asian, Black, and Other males wearing seat belts.

For people driving red cars, the highest probability was again Hispanic males wearing seat belts at .6561. Asian males were a very close second. Other and White males wearing seat belts were also quite high.

For blue cars, a Native American male wearing a seat belt has the highest probability at .999, which seems questionable because the count is only one. The next highest is Hispanic males wearing seat belts with similar probabilities to those in black and red cars. Other and Black males wearing seat belts have the next largest probabilities.

From the findings of brown cars, females of other races has the highest probability of .824, which is the first time females have had higher probabilities than males. Asian, Hispanic, and Black females wearing seats also had high probabilities.

For green cars, Hispanic males wearing seat belts are again the highest probability at .646.The next highest probabilities are from Asian, Black, and White males wearing seat belts.

Lastly, for the silver and white car probabilities, the two highest probabilities came from a Native American male driving silver and white cars while wearing seat belts, which seems a bit questionable again due to the low number of counts each has. Black, Hispanic, and Other males all wearing seat belts and driving white cars were the largest probabilities.

5 Summary of Conclusions

From the color trees, a couple trends arose of what gender/race combinations seem to have the highest probability of getting a traffic citation.

The first interesting trend observed was that all the highest probabilities were wearing seat belts. This was a surprising result to discover because wearing seat belts is a law, so breaking that law should add on to the likeliness of getting a ticket.This indicates there may have been other factors that had stronger influence over the citations not given to non-seat belt wearers than what can be shown by this data.

Another trend of interest was the for all the car colors except brown was that males had the highest probabilities to get a citation. This may be attributed to the stereotype that males are typically more reckless than driving. This merits further investigation in to the exact reason why males have this higher likelihood especially to look at the influence of interactions since Figure 1 did show a little bit of one.

The results of the brown car tree were interesting because all the high probabilities were from female drivers. There may be an influential factor here to answer why. Perhaps there is not enough data on brown cars or something in the environment that caused females to drive brown cars more. This merits for further research to find what might be influencing this.

The couple high results of Native American males having the highest probability seemed significant. However, the low amount of observations that are recorded for those drivers makes those results seem less reliable. There needs to be more data before it can be determined that Native American males have almost 100% probability for being pulled over in a blue, white, or silver car. #talk about other

It was interesting to see that people who are white did not have the highest probabilities, which may indicate that the police are selective on which races they give citation to. However, White males wearing seat belts were usually in the top five for each car color. This observation suggests that the police in that area may not have any racial prejudice, but it does require further research to come to a concise conclusion. Overall it seems that Hispanic males wearing seat belts have the highest likelihood of being given a citation when pulled over in Montgomery county, Maryland. Under each car color except brown, they have at least a 60% probability. There seems like there maybe some influence due to interaction, which needs to be inspected further.

After considering all the higher probabilities, the conclusion of this analysis is the males wearing seat belts who are of the races: Hispanic, Asian, Black, Other, and White (in descending order), are the most likely to be given a traffic citation when pulled over by the Montgomery Police. We conclude that there may be other influencing factors but that these gender/race combinations seem to be the most reckless in their driving to get a citation.

6 Future Recommendations

To get more concise results, it is recommended that additional variables be added to analyze the scientific question. The current factors only tell about the driver rather than what the driver was doing. It would be better to go through and filter the comments about the citation to see if there is a common factor like speeding or driving under the influence. Then, it could be better found if police officers were profiling people based on their race if other races were not given citations for the same law breaking.

An additional recommendation is to look further into the interactions. The predicted values in this report are include all the interactions. To see how the influence of each factor goes into the probability, it will be better to break down the dataset to avoid the influence of the interactions and to see them deeper.

7 References

Wickham, Hadley. 2022. *Tidyverse: Easily Install and Load the Tidyverse*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

———. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. https://yihui.org/knitr/.

———. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

8 Appendix

```
knitr::opts_chunk$set(echo = TRUE,warning = FALSE, message = FALSE)

library(tidyverse)
library(partykit)
library(mlbench)
library(vcd)
library(sjPlot)
library(knitr)
library(dplyr)
library(kableExtra)
#imported data here
load("C:/Users/naomi/OneDrive/Desktop/STAT 435/PermutationTest/TV.Rdata")


#First cleaning of the 800,000 observations
#Selected certain columns and renamed them
#Cleaned up the Violation Type column to just be 1s and 0s
test<-Traffic_Violations%>%
  select('Violation Type','Race','Gender','Belts','Color','DL State')%>%
  rename('CarColor'='Color', 'DLstate'='DL State')%>%
  mutate(Citation= sapply(`Violation Type`, function (x) ifelse(x == "Citation", 1, 0)))%>%
  select('Citation','Race','Gender','Belts','CarColor','DLstate')


#Created lists for renaming different car colors
oth<-c('CAMOUFLAGE'='OTHER','GOLD'='OTHER','GRAY'='OTHER','MAROON'='OTHER','N/A'='OTHER','ORANGE'='OTHE
gre<-c('GREEN, DK'='GREEN','GREEN, LGT'='GREEN')
blu<-c('BLUE, DARK'='BLUE','BLUE, LIGHT'='BLUE')
brn<-c('BEIGE'='BROWN','BRONZE'='BROWN','TAN'='BROWN','COPPER'='BROWN')
wht<-c('CREAM'='WHITE')


#Second cleaning/summarizing of the data
#renamed different car colors and filtered the genders
#added a counts column
test1<-test%>%
  group_by(Citation,Race,Gender,Belts, CarColor )%>%
  mutate(CarColor=str_replace_all(CarColor,oth),
         CarColor=str_replace_all(CarColor,gre),
         CarColor=str_replace_all(CarColor,blu),
         CarColor=str_replace_all(CarColor,brn),
         CarColor=str_replace_all(CarColor,wht))%>%
  filter(Gender=="F" | Gender=='M')%>%
  summarize(count=n())


save(test1,file='test1')
rm(list = ls())


load('test1')
```

```r
#Created Summary statistics for test1
opts <- options(knitr.kable.NA = "")
kable(summary(test1),align = "lccrr", caption='Figure 1. Summary Statistics')


#Created a linear model of test1 that runs through the anova funtion
a<-glm(Citation~(Race+Gender+Belts+CarColor)^2, data=test1, family='binomial',weights=count)
kable(anova(a), caption='Anova Output')%>%
  kable_styling(latex_options = "HOLD_position")


#interaction plots between factors of the linear model "a"
plot_model(a,type='pred',terms=c('Race','Gender','Belts'),ci.lvl=NA)+geom_line()+ labs(title = "Race, Ge

plot_model(a,type='pred',terms=c('Race','Gender'),ci.lvl=NA)+geom_line()+ labs(title = "Race and Gender

plot_model(a,type='pred',terms=c('Race','Gender', 'CarColor'),ci.lvl=NA)+geom_line()+ labs(title = "Race

plot_model(a,type='pred',terms=c('Belts', 'CarColor'),ci.lvl=NA)+geom_line()+ labs(title = "Belts and Ca

plot_model(a,type='pred',terms=c('Race','Belts'),ci.lvl=NA)+geom_line()+ labs(title = "Figure 1. Race an

plot_model(a,type='pred',terms=c('Gender','Belts'),ci.lvl=NA)+geom_line()+ labs(title = "Figure 2. Gende
plot_model(a,type='pred',terms=c('Race', 'CarColor'),ci.lvl=NA)+geom_line()+ labs(title = "Figure 3. Rac

plot_model(a,type='pred',terms=c('Gender', 'CarColor'),ci.lvl=NA)+geom_line()+ labs(title = "Figure 4. 


#Set variables in test1 to be factors
test1$Race<-factor(test1$Race)
test1$Gender<-factor(test1$Gender)
test1$Belts<- factor(test1$Belts)
test1$CarColor<- factor(test1$CarColor)

#Ran test1 through the glmtree function
tree <- glmtree(Citation~ Race+Gender|Belts+CarColor,
                     data = test1, weights=count,family = binomial)
print(plot(tree))

#Split up the data by car color to perform individual trees

#BLACK

#Filtered the data to just have black cars
tb<-test1%>%
  filter(CarColor=='BLACK')

#Performed a tree analysis
treeb <- glmtree(Citation~ Race+Gender|Belts,
                data = tb, weights=count,family = binomial)

  #Print statements to see the tree
    #print(treeb)
```

```r
    #print(treeb, node=1)
    #plot(treeb)
    #plot(treeb, terminal_panel = NULL)

#Getting the probabilities for each observation in the tree
#Also made a new dataset with the probabilities included
probability<-predict(treeb, newdata = tb, type = "response")

testtb<-tb
testtb$probability<-probability

testtb<- testtb%>%
  filter(Citation==1)

testtb

#Filtered the probability data set for just the min and max values of each race
blk1<-testtb %>% group_by(Race) %>% top_n(1, probability)
blk2<-testtb %>% group_by(Race) %>% top_n(-1, probability)
com<-rbind(blk1,blk2)
blk<-com %>% arrange(Race)


#RED

#Filtered the data to only get observations with red cars
tr<-test1%>%
  filter(CarColor=='RED')

#Performed a Tree analysis
treer <- glmtree(Citation~ Race+Gender|Belts,
                 data = tr, weights=count,family = binomial)

  #Print statements to see the tree
    #print(treer)
    #print(treer, node=1)
    #plot(treer)
    #plot(treer, terminal_panel = NULL)

#Getting the probabilities for each observation in the tree
#Also made a new dataset with the probabilities included
probr<-predict(treer, newdata = tr, type = "response")

testtr<-tr
testtr$probability<-probr

testtr<- testtr%>%
  filter(Citation==1)

testtr

#Filtered the probability data set for just the min and max values of each race
rd1<-testtr %>% group_by(Race) %>% top_n(1, probability)
```

```
rd2<-testtr %>% group_by(Race) %>% top_n(-1, probability)
comr<-rbind(rd1,rd2)
rd<-comr %>% arrange(Race)

#BLUE

#Filtered the data to just have blue cars
tbl<-test1%>%
  filter(CarColor=='BLUE')

#Performed the Tree analysis
treebl <- glmtree(Citation~ Race+Gender|Belts,
                 data = tbl, weights=count,family = binomial)

  #Print statements to see the tree
    #print(treebl)
    #print(treebl, node=1)
    #plot(treebl)
    #plot(treebl, terminal_panel = NULL)


#Getting the probabilities for each observation in the tree
#Also made a new dataset with the probabilities included
probl<-predict(treebl, newdata = tbl, type = "response")

testbl<-tbl
testbl$probability<-probl

testbl<- testbl%>%
  filter(Citation==1)

testbl

#Filtered the probability data set for just the min and max values of each race
bl1<-testbl %>% group_by(Race) %>% top_n(1, probability)
bl2<-testbl %>% group_by(Race) %>% top_n(-1, probability)
coml<-rbind(bl1,bl2)
blu<-coml %>% arrange(Race)

#GREEN

#Filtred the data to only get green cars
tg<-test1%>%
  filter(CarColor=='GREEN')

#Performed the Tree analysis
treeg <- glmtree(Citation~ Race+Gender|Belts,
                  data = tg, weights=count,family = binomial)

  #Print statments to see the tree
    #print(treeg)
    #print(treeg, node=1)
    #plot(treeg)
```

```r
    #plot(treeg, terminal_panel = NULL)

#Getting the probabilities for each observation in the tree
#Also made a new dataset with the probabilities included
probg<-predict(treeg, newdata = tg, type = "response")

testg<-tg
testg$probability<-probg

testg<- testg%>%
  filter(Citation==1)
testg

#Filtered the probability data set for just the min and max values of each race
g1<-testg %>% group_by(Race) %>% top_n(1, probability)
g2<-testg %>% group_by(Race) %>% top_n(-1, probability)
comg<-rbind(g1,g2)
grn<-comg %>% arrange(Race)

#BROWN

#Filtered the data to only get brown cars
tbr<-test1%>%
  filter(CarColor=='BROWN')

#Performed the Tree analysis
treebr <- glmtree(Citation~ Race+Gender|Belts,
                  data = tbr, weights=count,family = binomial)

  #Print statements to see the tree
    #print(treebr)
    #print(treebr, node=1)
    #plot(treebr)
    #plot(treebr, terminal_panel = NULL)


#Getting the probabilities for each observation in the tree
#Also made a new dataset with the probabilities included
probbr<-predict(treebr, newdata = tbr, type = "response")

testbr<-tbr
testbr$probability<-probbr

testbr<- testbr%>%
  filter(Citation==1)

testbr

#Filtered the probability data set for just the min and max values of each race
br1<-testbr %>% group_by(Race) %>% top_n(1, probability)
br2<-testbr %>% group_by(Race) %>% top_n(-1, probability)
combr<-rbind(br1,br2)
brn<-combr %>% arrange(Race)
```

```r
#SILVER+WHITE

#Filtered the data to only have white and silver cars
tws<-test1%>%
  filter(CarColor=='WHITE' | CarColor=='SILVER')

#Performed the Tree analysis
treews <- glmtree(Citation~ Race+Gender|Belts+CarColor,
                  data = tws, weights=count,family = binomial)

  #Print statements to see the tree
    #print(treews)
    #print(treews, node=1)
    #plot(treews)
    #plot(treews, terminal_panel = NULL)


#Getting the probabilities for each observation in the tree
#Also made a new dataset with the probabilities included
probsw<-predict(treews, newdata = tws, type = "response")

testsw<-tws
testsw$probability<-probsw

testsw<- testsw%>%
  filter(Citation==1)

testsw

#Filtered the probability data set for just the min and max values of each race
sw1<-testsw %>% group_by(CarColor,Race) %>% top_n(1, probability)
sw2<-testsw %>% group_by(CarColor,Race) %>% top_n(-1, probability)
comsw<-rbind(sw1,sw2)
sw<-comsw %>% arrange(Race,CarColor)


#print all the tables with min and max probabilities for each race of each car color

#black car
kable(blk, caption='Black Car Probabilities') %>%
  kable_styling(latex_options = "HOLD_position")
#red car
kable(rd, caption='Red Car Probabilities')%>%
  kable_styling(latex_options = "HOLD_position")
#blue car
kable(blu,caption='Blue Car Probabilities')%>%
  kable_styling(latex_options = "HOLD_position")
#brown car
kable(brn, caption='Brown Car Probabilities')%>%
  kable_styling(latex_options = "HOLD_position")
#green car
kable(grn,caption='Green Car Probabilities')%>%
  kable_styling(latex_options = "HOLD_position")
```

```r
#silver/white cars
kable(sw,caption='Silver/White Car Probabilities')%>%
  kable_styling(latex_options = "HOLD_position")
```