ANA515 Assignment 2

Fang Tham

2022-09-17

Section 1: Description of the data

The data set from the flying etiquette survey aims to understand what people think of flying etiquettes. The data collected is collected by Walt Hickey who was the chief culture writer of FiveThirtyEight. Since reading many articles about air passengers’ confrontations amongst each other, he decided to run a SuveyMonkey Audience poll with air passengers to inquire about flying etiquettes. Data from 1040 respondents were collected, 874 who had experience flying. He included questions about reclining seats, using the bath room, switching seats, bringing babies on flights etc. We are interested to know the percentage of people who thinks that certain etiquettes should be done a certain way.

Section 2: Reading the data.

#Section 2: Reading into the data, the read\_csv from readr package that is loaded through tidyverse is used.  
url <- 'https://raw.githubusercontent.com/fivethirtyeight/data/master/flying-etiquette-survey/flying-etiquette.csv'  
etiquette <- read.csv(url(url))

Section3: Cleaning the data. I cleaned the data by renaming some columns of interest.

#clean data by renaming the columns  
colnames(etiquette)[colnames (etiquette) == 'How.tall.are.you.'] <- 'Height'  
colnames(etiquette)[colnames (etiquette) == 'Is.itrude.to.recline.your.seat.on.a.plane.'] <- 'Recline\_Rude'  
colnames(etiquette)[colnames (etiquette) == 'Do.you.ever.recline.your.seat.when.you.fly.'] <- 'Recline\_Freq'

Section 4: This is to inspect the characteristics of my data

This data set has 1040 and 27. The names of the columns and a brief description of each are in the table below:

#The table below shows a summary of what each column means.  
table\_description <- data.frame(Names = c("Recline\_Freq", "Height", "Recline\_Rude", "Age", "Gender"),  
 Description = c(  
 "Frequency of reclining their seats when they fly",   
 "Height of respondent",  
 "Whether it's rude for one to recline their seat",  
 "Age of respondent", "Gender of respondent"))  
  
#Print the table  
print(table\_description)

## Names Description  
## 1 Recline\_Freq Frequency of reclining their seats when they fly  
## 2 Height Height of respondent  
## 3 Recline\_Rude Whether it's rude for one to recline their seat  
## 4 Age Age of respondent  
## 5 Gender Gender of respondent

This is the summary statistics for 4 columns. I have selected to look at the recline frequency, height, age, gender, and seeing if those factors relate to whether people think reclining chairs on an airplane is rude

#I am selecting 4 columns that I am interested in by creating a new dataframe from my subset. I kept the respondentID in as as an ID tag:  
newetq <- subset(etiquette, select=c(1,3,4,13,23,24))  
  
#I am changing the class of all columns from characters to factors  
newetq2 <- as.data.frame(unclass(newetq),  
 stringsAsFactors = TRUE)  
str(newetq2)

## 'data.frame': 1040 obs. of 6 variables:  
## $ RespondentID: num 3.44e+09 3.43e+09 3.43e+09 3.43e+09 3.43e+09 ...  
## $ Recline\_Freq: Factor w/ 6 levels "","About half the time",..: 1 2 6 3 2 6 5 5 5 4 ...  
## $ Height : Factor w/ 21 levels "","5'0\"","5'1\"",..: 1 17 12 5 11 13 16 14 14 10 ...  
## $ Recline\_Rude: Factor w/ 4 levels "","No, not rude at all",..: 1 3 2 2 2 2 3 2 2 4 ...  
## $ Gender : Factor w/ 3 levels "","Female","Male": 1 3 3 3 3 3 3 3 1 3 ...  
## $ Age : Factor w/ 5 levels "","> 60","18-29",..: 1 4 4 4 4 4 4 4 1 4 ...

Now, I am able to analyze the frequency of each survey result. I first run a general summary of all my columns of interest. In this case, the missing cells do not affect our data analysis as we are calculating frequencies and running cross tab analysis. It will not affect the descriptive analysis, so I have decided to leave the missing rows in the data.

etqsum <- summary(newetq2)  
print(etqsum)

## RespondentID Recline\_Freq Height   
## Min. :3.432e+09 :182 :182   
## 1st Qu.:3.432e+09 About half the time:118 5'4" : 79   
## Median :3.433e+09 Always :137 5'7" : 76   
## Mean :3.433e+09 Never :171 5'8" : 76   
## 3rd Qu.:3.433e+09 Once in a while :257 5'6" : 75   
## Max. :3.436e+09 Usually :175 5'9" : 72   
## (Other):480   
## Recline\_Rude Gender Age   
## :186 : 33 : 33   
## No, not rude at all:502 Female:528 > 60 :258   
## Yes, somewhat rude :281 Male :479 18-29:220   
## Yes, very rude : 71 30-44:254   
## 45-60:275   
##   
##

Then, I perform a cross tab analysis to observe if different genders had different opinions on reclining seats. I also looked into whether heights affected their opinions on reclining the seats. I then looked at whether people of different ages had opinions on whether reclining seats is rude

crosstab\_genderrecline <- xtabs(~ Gender + Recline\_Freq, data=newetq2)  
crosstab\_heightrecline <- xtabs(~ Height + Recline\_Freq, data=newetq2)  
crosstab\_agerude <- xtabs(~ Age + Recline\_Rude, data=newetq2)  
  
#I'll now convert the cross tabulation counts to percentages by column or row and round it to 0 decimal points  
Propcrosstab\_genderrecline <- round(100 \* prop.table(crosstab\_genderrecline, 1), 0)  
Propcrosstab\_heightrecline<- round(100 \* prop.table(crosstab\_heightrecline, 1), 0)  
Propcrosstab\_agerude <- round(100 \* prop.table(crosstab\_agerude, 1), 0)  
print(Propcrosstab\_genderrecline)

## Recline\_Freq  
## Gender About half the time Always Never Once in a while Usually  
## 55 6 15 15 9 0  
## Female 16 13 14 16 24 17  
## Male 16 10 13 16 27 18

print(Propcrosstab\_heightrecline)

## Recline\_Freq  
## Height About half the time Always Never Once in a while Usually  
## 100 0 0 0 0 0  
## 5'0" 0 29 18 18 12 24  
## 5'1" 0 5 21 26 32 16  
## 5'10" 0 6 13 25 33 22  
## 5'11" 0 15 22 24 30 9  
## 5'2" 0 9 13 27 36 16  
## 5'3" 0 29 19 17 21 15  
## 5'4" 0 10 14 23 29 24  
## 5'5" 0 14 15 17 32 21  
## 5'6" 0 19 13 21 23 24  
## 5'7" 0 16 17 17 33 17  
## 5'8" 0 13 16 22 22 26  
## 5'9" 0 11 11 12 42 24  
## 6'0" 0 11 21 18 35 16  
## 6'1" 0 12 23 12 23 31  
## 6'2" 0 15 7 19 41 19  
## 6'3" 0 22 11 28 22 17  
## 6'4" 0 0 18 18 36 27  
## 6'5" 0 17 33 17 17 17  
## 6'6" and above 0 0 50 0 0 50  
## Under 5 ft. 0 17 17 17 33 17

print(Propcrosstab\_agerude)

## Recline\_Rude  
## Age No, not rude at all Yes, somewhat rude Yes, very rude  
## 67 24 6 3  
## > 60 17 52 24 8  
## 18-29 22 35 34 9  
## 30-44 13 56 25 6  
## 45-60 15 51 29 5

In summary, there was no differences between the likelihood of a gender reclining the seat or not. Both females (24%) and males (27%) were most likely to recline their seats “once in a while”. When it comes to height, it was apparent that people who were 6’6” and above either “Usually” (50%) or “Always” (50%) recline their seats. Majority (33%) of the people who were under 5 Ft. said that they only recline their seats once in a while. When we looked at whether there were certain age groups that thought reclining seats were rude, there wasn’t a significant differences between the opinions of each age group. Out of the 3 selections of “No, not rude at all, Yes somewhat rude and Yes very rude), most people(48%) selected”No not rude at all”.