DSO545\_HW02

Xu Zhang

February 7, 2018

getwd()

## [1] "C:/Users/Xu Zhang/Desktop"

FlyData=read.csv("flying.csv")

1. (2 point) Use dplyr functions and the pipe operator %>% to divide the dataset into two dataframes:males\_data and female\_data. Each dataframe should include only the following variables:respondent\_id, gender, baby.

library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

library(ggplot2)  
males\_data=  
 FlyData %>%   
 select(respondent\_id, gender, baby) %>%  
 filter(gender=="Male")  
females\_data=  
 FlyData %>%   
 select(respondent\_id, gender, baby) %>%  
 filter(gender=="Female")

1. (2 points) What is the percentage of males who thinks that it is totally fine to bring a baby on the plane? What is the percentage of females who thinks that it is totally fine to bring a baby on the plane? Who is more tolerant?

FlyData\_m=males\_data[!is.na(males\_data$baby),]  
paste(round(nrow(FlyData\_m %>% filter(baby=="No"))/nrow(FlyData\_m)\*100,3),"%")

## [1] "62.095 %"

FlyData\_f=females\_data[!is.na(females\_data$baby),]  
paste(round(nrow(FlyData\_f %>% filter(baby=="No"))/nrow(FlyData\_f)\*100,3),"%")

## [1] "76.471 %"

Female is more tolerent.

1. (2 points) Compute the percentages of people who think that the following are very annoying. Which is the most annoying to fliers? . Bring a baby on a plane

FlyData1=FlyData[!is.na(FlyData$baby),]  
paste(round(nrow(FlyData1 %>% filter(baby=="Very"))/nrow(FlyData1)\*100,3),"%")

## [1] "8.834 %"

. Be chatty and talking to strangers

FlyData2=FlyData[!is.na(FlyData$talk\_stranger),]  
paste(round(nrow(FlyData2 %>% filter(talk\_stranger=="Very"))/nrow(FlyData2)\*100,3),"%")

## [1] "3.158 %"

. Wake someone up to go for a walk

FlyData3=FlyData[!is.na(FlyData$wake\_up\_walk),]  
paste(round(nrow(FlyData3 %>% filter(wake\_up\_walk=="Very"))/nrow(FlyData3)\*100,3),"%")

## [1] "20.941 %"

. Wake someone up to use the bathroom

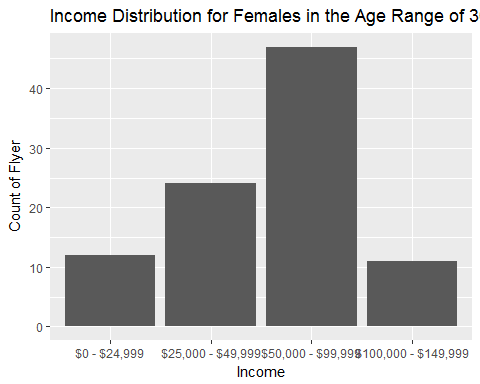
FlyData4=FlyData[!is.na(FlyData$wake\_up\_bathroom),]  
paste(round(nrow(FlyData4 %>% filter(wake\_up\_bathroom=="Very"))/nrow(FlyData4)\*100,3),"%")

## [1] "4 %"

Waking someone up to go for a walk is the most annoying to filter

(4 points) Create the following two plots.

NewFlyData=FlyData[!is.na(FlyData$household\_income),]  
lev = levels(NewFlyData$household\_income)  
lev = lev[c(1,3,4,2)]  
NewFlyData$household\_income = factor(NewFlyData$household\_income, levels = lev)  
  
NewFlyData %>%  
 filter(gender=="Female" & age=="30-44") %>%  
 group\_by(household\_income) %>%   
 ggplot(aes(x = household\_income)) +  
 geom\_bar() +  
 ggtitle("Income Distribution for Females in the Age Range of 30???44") +   
 xlab("Income") +   
 ylab("Count of Flyer")



NewFlyData=FlyData[!is.na(FlyData$household\_income),]  
lev = levels(NewFlyData$household\_income)  
lev = lev[c(1,3,4,2)]  
NewFlyData$household\_income = factor(NewFlyData$household\_income, levels = lev)  
  
NewFlyData %>%  
 filter(gender=="Male" & age=="30-44") %>%  
 group\_by(household\_income) %>%   
 ggplot(aes(x = household\_income)) +  
 geom\_bar() +  
 ggtitle("Income Distribution for Males in the Age Range of 30???44") +   
 xlab("Income") +   
 ylab("Count of Flyer")

