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")
```

(2) (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.

(3) (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

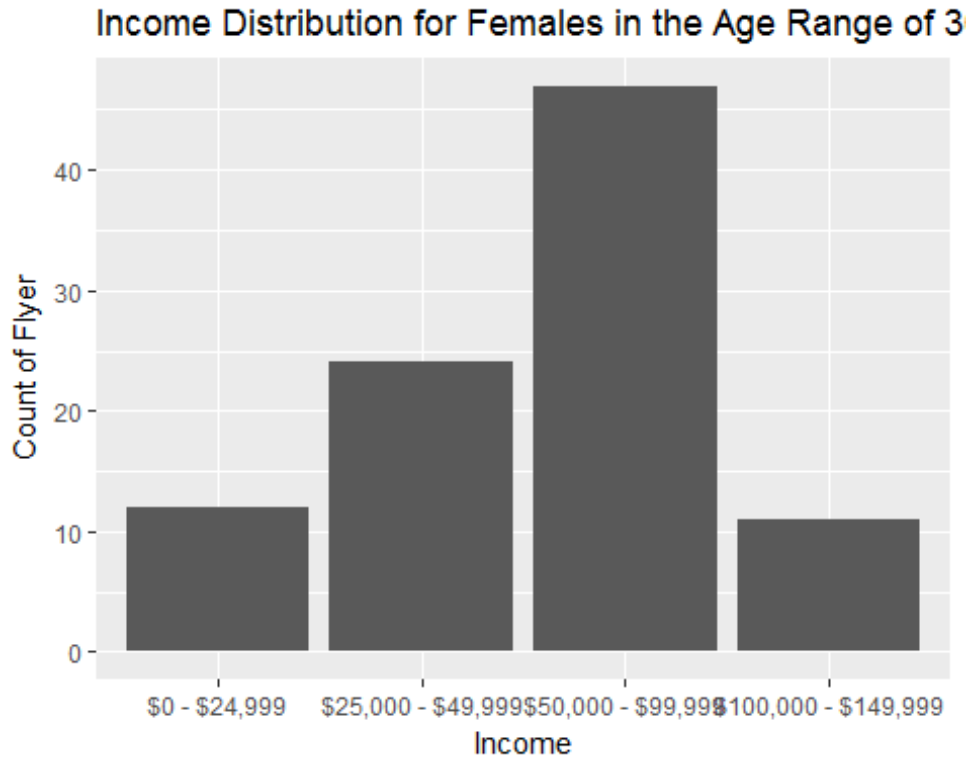
```
FlyData4=FlyData[!is.na(FlyData$wake_up_bathroom),]
paste(round(nrow(FlyData4 %>%
filter(wake_up_bathroom=="Very"))/nrow(FlyData4)*100,3),"%")

## [1] "4 %"
```

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

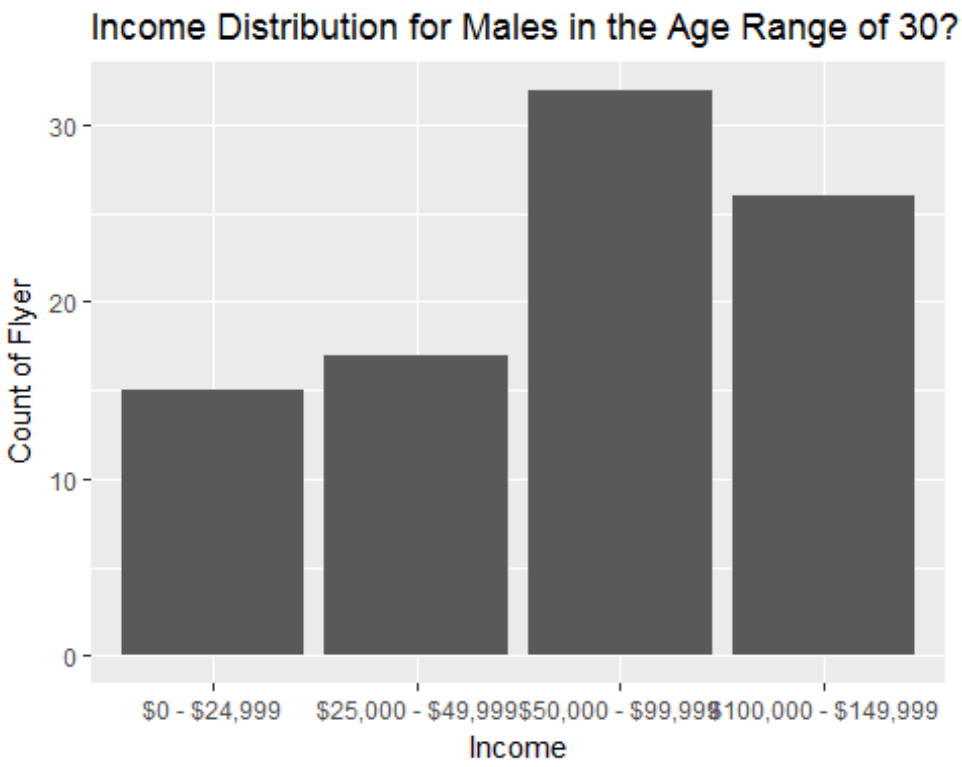
(4 points) Create the following two plots.

```
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")
```


Income Distribution for Males in the Age Range of 30?
Count of Flyer
30
20
10
0
$0 - $24,999
$25,000 - $49,999
$50,000 - $99,999
$100,000 - $149,999
Income