Stat 480 - Homework #5

Vahid Azizi

2/20/2020

Flying etiquette

1. Download the RMarkdown file with these homework instructions to use as a template for your work. Make sure to replace “Your Name” in the YAML with your name.
2. FiveThirtyEight is a website founded by Statistician and writer Nate Silver to publish results from opinion poll analysis, politics, economics, and sports blogging. One of the featured articles considers flying etiquette (https://fivethirtyeight.com/features/airplane-etiquette-recline-seat/). This article is based on data collected by FiveThirtyEight and publicly available on github. Use the code below to read in the data from the survey:

```
fly <- read.csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/flying-etiquette-survey/flying-etiquette.csv")
```

The next couple of lines of code provide a bit of cleanup of the demographic information by reordering the levels of the corresponding factor variables. Run this code in your session.

```
fly$Age <- factor(fly$Age, levels=c("18-29", "30-44", "45-60", "> 60", ""))
fly$Household.Income <- factor(fly$Household.Income, levels = c("$0 - $24,999","$25,000 - $49,999", "$50,000 - $99,999", "$100,000 - $149,999", "150000", ""))
fly$Education <- factor(fly$Education, levels = c("Less than high school degree", "High school degree", "Some college or Associate degree", "Bachelor degree",  "Graduate degree", ""))
```

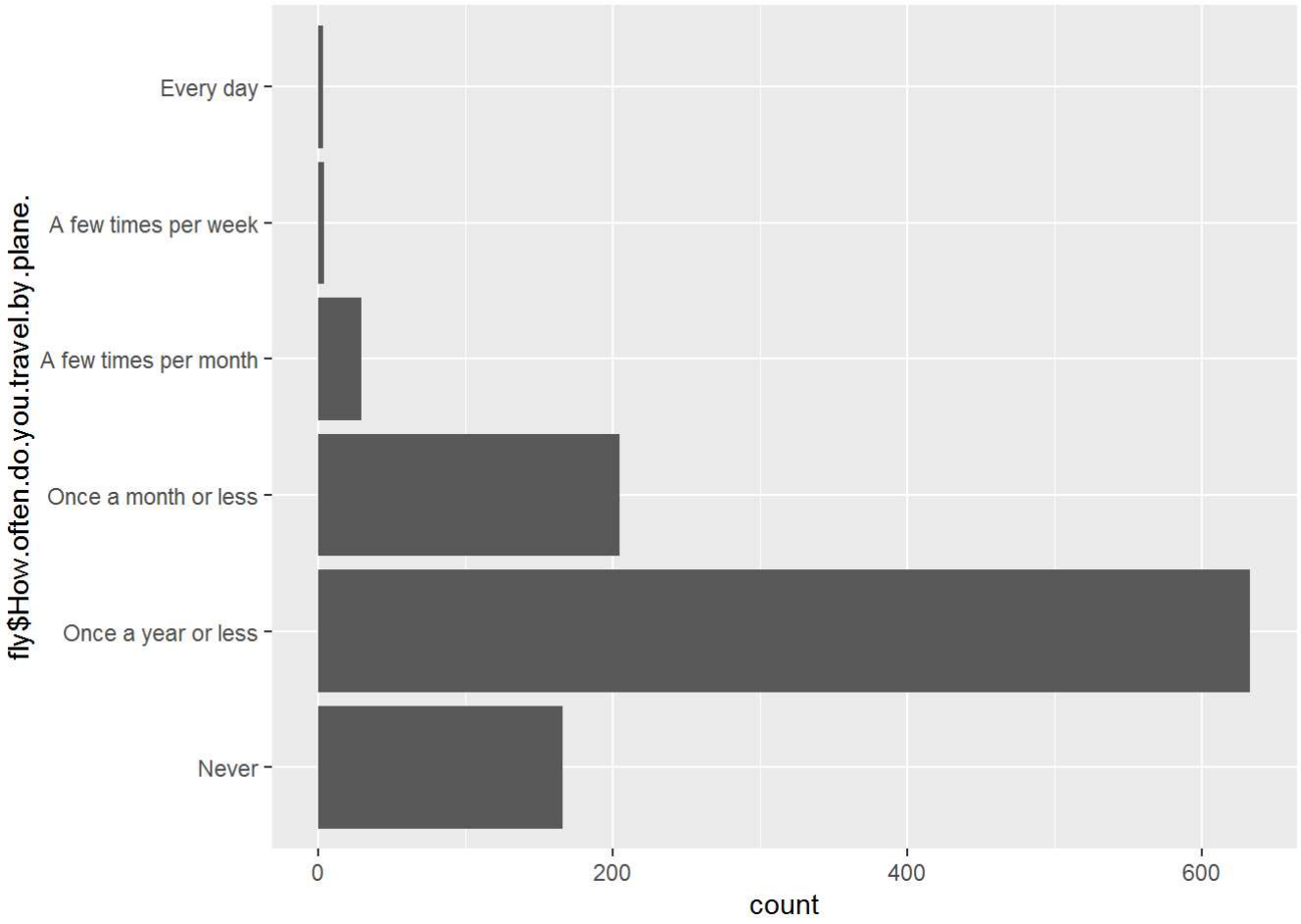
3. Some people do not travel often by plane. Provide a breakdown of travel frequency (use variable `How.often.do.you.travel.by.plane.`). Reorder the levels in the variable by travel frequency from least frequent travel to most frequent. Draw a barchart of travel frequency and comment on it.

```
#travel frequency levels
unique(fly$How.often.do.you.travel.by.plane.)
```

```
## [1] Once a year or less   Once a month or less  Never
## [4] A few times per week  Every day             A few times per month
## 6 Levels: A few times per month A few times per week Every day ... Once a year or less
```

```
#reordering the levels
fly$How.often.do.you.travel.by.plane.<-factor(fly$How.often.do.you.travel.by.plane., levels=c("Never", "Once a year or less", "Once a month or less", "A few times per month", "A few times per week", "Every day"))
```

```
#plotting bar chart for travel frequency levels
ggplot(fly,aes(x=fly$How.often.do.you.travel.by.plane.))+
  geom_bar()+
  coord_flip()
```



According to the bar chart the most of people attended in this survey travel by plane once a year or less. There are some people (almost 190) that never travel by plane. Also, small number of people travel everyday by plane.

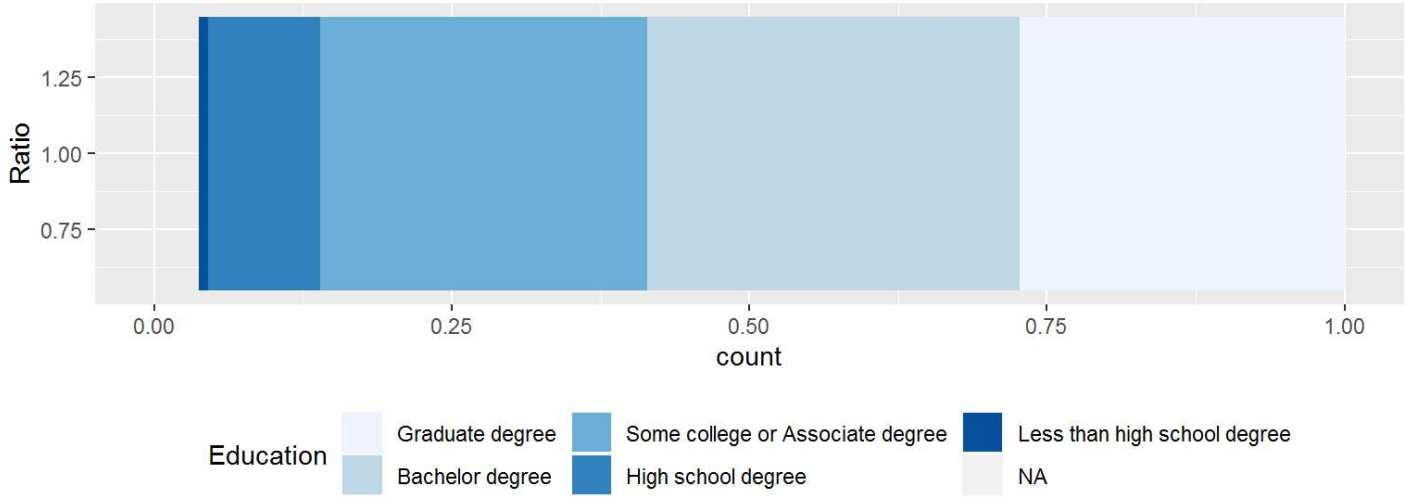
4. In the demographic variables (`Education` , `Age` , and `Houshold.Income`), replace all occurrences of the empty string “” by a missing value `NA` . How many responses do not have any missing values? (Hint: the function `is.na` might come in handy)

```
#replacing string "" values with NA
fly$Education<-replace(fly$Education,which(fly$Education==""),NA)
fly$Age<-replace(fly$Age,which(fly$Age==""),NA)
fly$Household.Income<-replace(fly$Household.Income,which(fly$Household.Income==""),NA)

#counting responses without NA values
length(which(rowSums(is.na(fly))==0))
```

```
## [1] 824
```

5. Run the command below and interpret the output. What potential purpose can you see for the chart?

```
library(ggplot2)
fly$Education = with(fly, factor(Education, levels = rev(levels(Education))))

ggplot(data = fly, aes(x = 1)) +
  geom_bar(aes(fill=Education), position="fill") +
  coord_flip() +
  theme(legend.position="bottom") +
    scale_fill_brewer() +
  xlab("Ratio")
```



This plot tries to show the ordered categories for 'Education' from 'Graduate degree' as the highest level to the 'Less than high school degree' that is the lowest level of education. Also it shows the percentage of each category out of all responses.

6. Rename the variable `In.general..is.itrude.to.bring.a.baby.on.a.plane.` to `baby.on.plane.` . How many levels does the variable `baby.on.plane` have, and what are these levels? Rename the level labeled "" to "Not answered". Reorder the levels of `baby.on.plane` from least rude to most rude. Put the level "Not answered" last. Draw a barchart of variable `baby.on.plane` . Interpret the result.

```
#number of levels
names(fly)[names(fly)=='In.general..is.itrude.to.bring.a.baby.on.a.plane.']<-'baby.on.plane'
length(levels(fly$baby.on.plane))
```

```
## [1] 4
```

```
levels(fly$baby.on.plane)
```

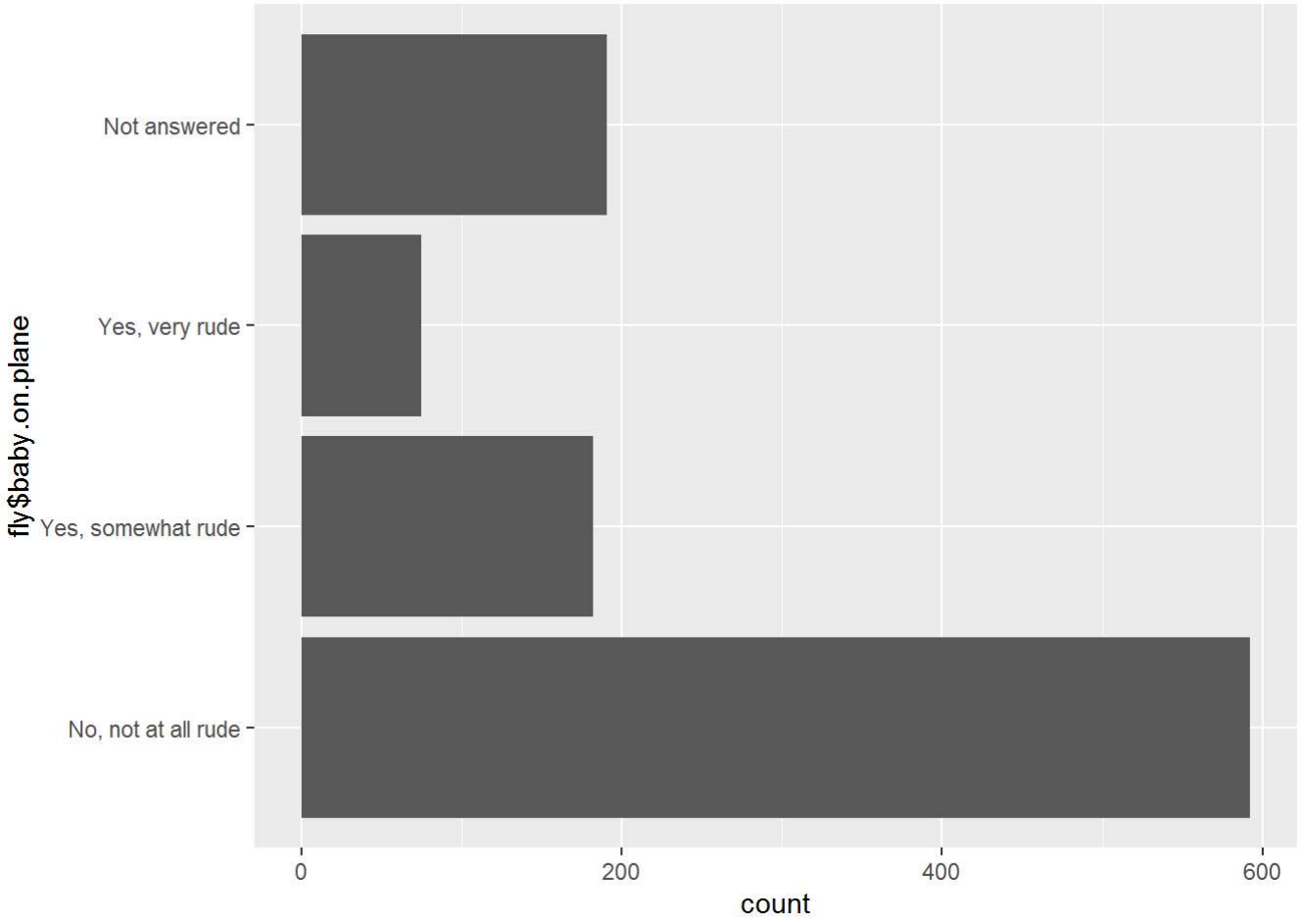
```
## [1] ""                   "No, not at all rude" "Yes, somewhat rude"
## [4] "Yes, very rude"
```

```
#replacing empty cells
fly$baby.on.plane<-as.character(fly$baby.on.plane)
fly$baby.on.plane[fly$baby.on.plane==""]<-"Not answered"
fly$baby.on.plane<-as.factor(fly$baby.on.plane)
levels(fly$baby.on.plane)
```

```
## [1] "No, not at all rude" "Not answered"        "Yes, somewhat rude"
## [4] "Yes, very rude"
```

```
#reodering levels
fly$baby.on.plane<-factor(fly$baby.on.plane, levels=c("No, not at all rude", "Yes, somewhat rud
e", "Yes, very rude", "Not answered"))
```

```
#plotting bar chart for baby.on.plane levels
ggplot(fly,aes(x=fly$baby.on.plane))+
  geom_bar()+
  coord_flip()
```



According to this bar chart, most of the responses consider the baby on the plane as bahavior that is not rude at all. Almost 200 people did not answer this question. Almost 60 people consider having baby on the plane as very rude behavior.

7. Investigate the relationship between gender and the variables `Do.you.have.any.children.under.18.` and `baby.on.plane` . How is the attitude towards babies on planes shaped by gender and having children under

18? Find a plot that summarises your findings (use `ggplot2`).

```
#replacing empty cells
fly$Gender<-as.character(fly$Gender)
fly$Gender[fly$Gender==""]<-"Not answered"
fly$Gender<-as.factor(fly$Gender)

fly$Do.you.have.any.children.under.18.<-as.character(fly$Do.you.have.any.children.under.18.)
fly$Do.you.have.any.children.under.18.[fly$Do.you.have.any.children.under.18.==""]<-"Not answere
d"
fly$Do.you.have.any.children.under.18.<-as.factor(fly$Do.you.have.any.children.under.18.)


paste("variable: Do.you.have.any.children.under.18.")
```

```
## [1] "variable: Do.you.have.any.children.under.18."
```

```
male_yes<-fly %>% group_by(Gender) %>% filter(Gender=="Male",Do.you.have.any.children.under.18.=
="Yes")
male_yes_rate<-nrow(male_yes)/nrow(fly)
paste("rate for <Gender:male, children.under.18:Yes>", male_yes_rate)
```

```
## [1] "rate for <Gender:male, children.under.18:Yes> 0.0894230769230769"
```

```
male_no<-fly %>% group_by(Gender) %>% filter(Gender=="Male",Do.you.have.any.children.under.18.==
"No")
male_no_rate<-nrow(male_no)/nrow(fly)
paste("rate for <Gender:male, children.under.18:No>", male_no_rate)
```

```
## [1] "rate for <Gender:male, children.under.18:No> 0.294230769230769"
```

```
female_yes<-fly %>% group_by(Gender) %>% filter(Gender=="Female",Do.you.have.any.children.under.
18.=="Yes")
female_yes_rate<-nrow(female_yes)/nrow(fly)
paste("rate for <Gender:female, children.under.18:Yes>", female_yes_rate)
```

```
## [1] "rate for <Gender:female, children.under.18:Yes> 0.0903846153846154"
```

```
female_no<-fly %>% group_by(Gender) %>% filter(Gender=="Female",Do.you.have.any.children.under.1
8.=="No")
female_no_rate<-nrow(female_no)/nrow(fly)
paste("rate for <Gender:female, children.under.18:No>", female_no_rate)
```

```
## [1] "rate for <Gender:female, children.under.18:No> 0.332692307692308"
```

```
paste("variable: baby.on.plane")
```

```
## [1] "variable: baby.on.plane"
```

```
male_1<-fly %>% group_by(Gender) %>% filter(Gender=="Male",baby.on.plane=="No, not at all rude")
male_1_rate<-nrow(male_1)/nrow(fly)
paste("rate for <Gender:male, baby.on.plane:not at all rude>", male_1_rate)
```

```
## [1] "rate for <Gender:male, baby.on.plane:not at all rude> 0.239423076923077"
```

```
female_1<-fly %>% group_by(Gender) %>% filter(Gender=="Female",baby.on.plane=="No, not at all ru
de")
female_1_rate<-nrow(female_1)/nrow(fly)
paste("rate for <Gender:female, baby.on.plane:not at all rude>", female_1_rate)
```

```
## [1] "rate for <Gender:female, baby.on.plane:not at all rude> 0.325"
```

```
male_2<-fly %>% group_by(Gender) %>% filter(Gender=="Male",baby.on.plane=="Yes, somewhat rude")
male_2_rate<-nrow(male_2)/nrow(fly)
paste("rate for <Gender:male, baby.on.plane:somewhat rude>", male_2_rate)
```

```
## [1] "rate for <Gender:male, baby.on.plane:somewhat rude> 0.100961538461538"
```

```
female_2<-fly %>% group_by(Gender) %>% filter(Gender=="Female",baby.on.plane=="Yes, somewhat rud
e")
female_2_rate<-nrow(female_2)/nrow(fly)
paste("rate for <Gender:female, baby.on.plane:somewhat rude>", female_2_rate)
```

```
## [1] "rate for <Gender:female, baby.on.plane:somewhat rude> 0.0730769230769231"
```

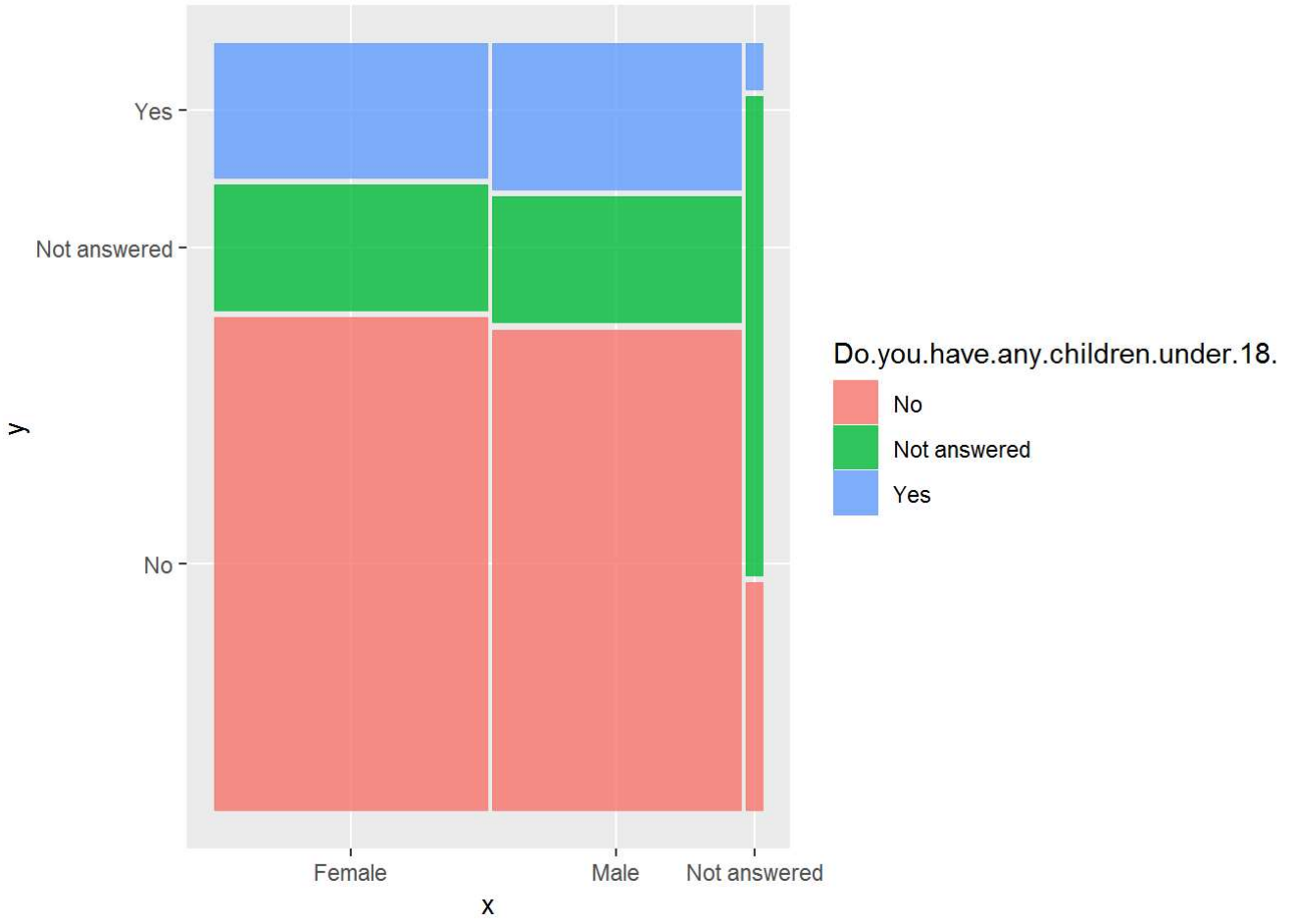
```
male_3<-fly %>% group_by(Gender) %>% filter(Gender=="Male",baby.on.plane=="Yes, very rude")
male_3_rate<-nrow(male_3)/nrow(fly)
paste("rate for <Gender:male, baby.on.plane:very rude>", male_3_rate)
```

```
## [1] "rate for <Gender:male, baby.on.plane:very rude> 0.0451923076923077"
```

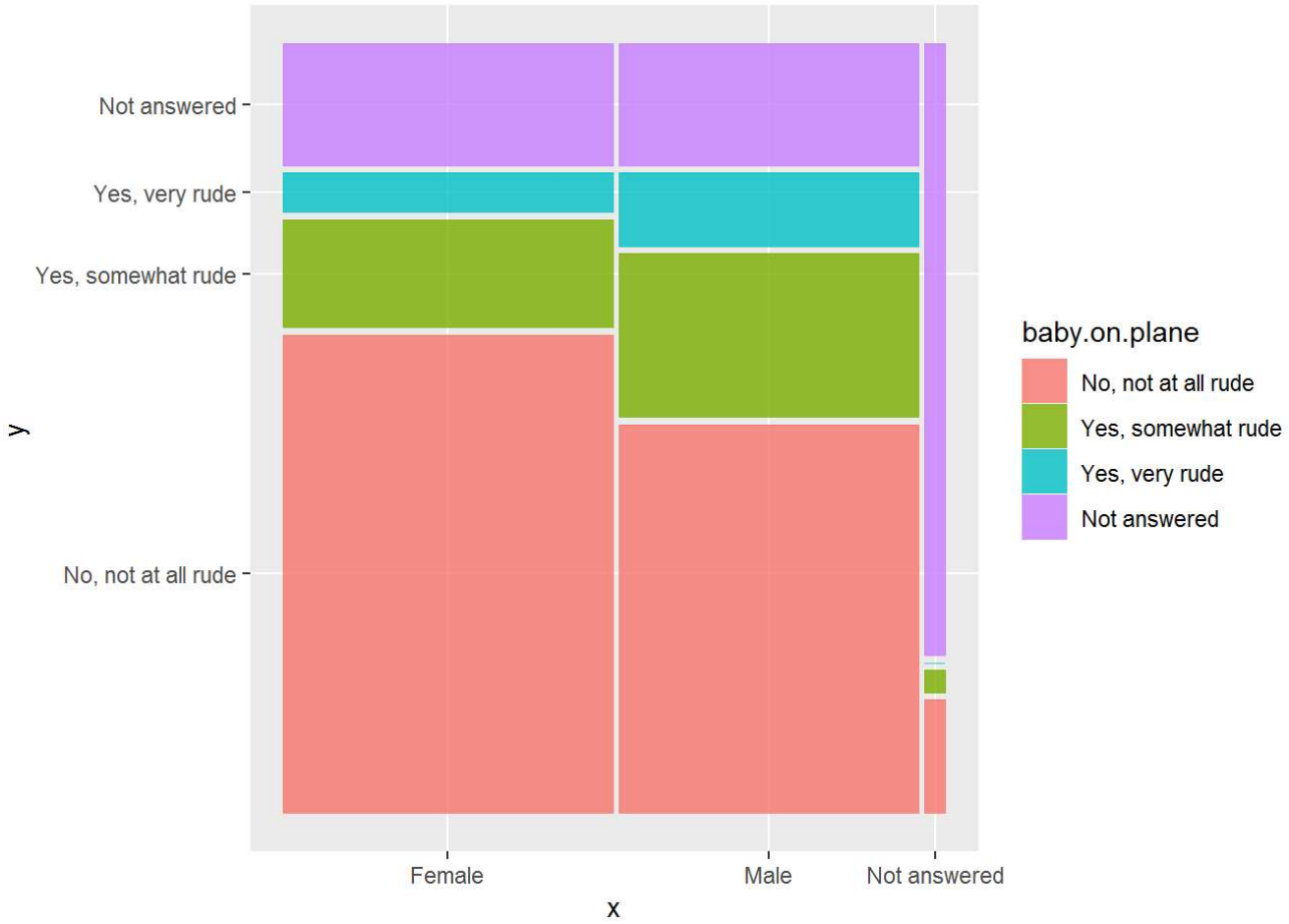
```
female_3<-fly %>% group_by(Gender) %>% filter(Gender=="Female",baby.on.plane=="Yes, very rude")
female_3_rate<-nrow(female_3)/nrow(fly)
paste("rate for <Gender:female, baby.on.plane:very rude>", female_3_rate)
```

```
## [1] "rate for <Gender:female, baby.on.plane:very rude> 0.0269230769230769"
```

```
library(ggmosaic)
ggplot(data = fly)  +
  geom_mosaic(aes(x = product(Gender), fill=Do.you.have.any.children.under.18., weight=1))
```

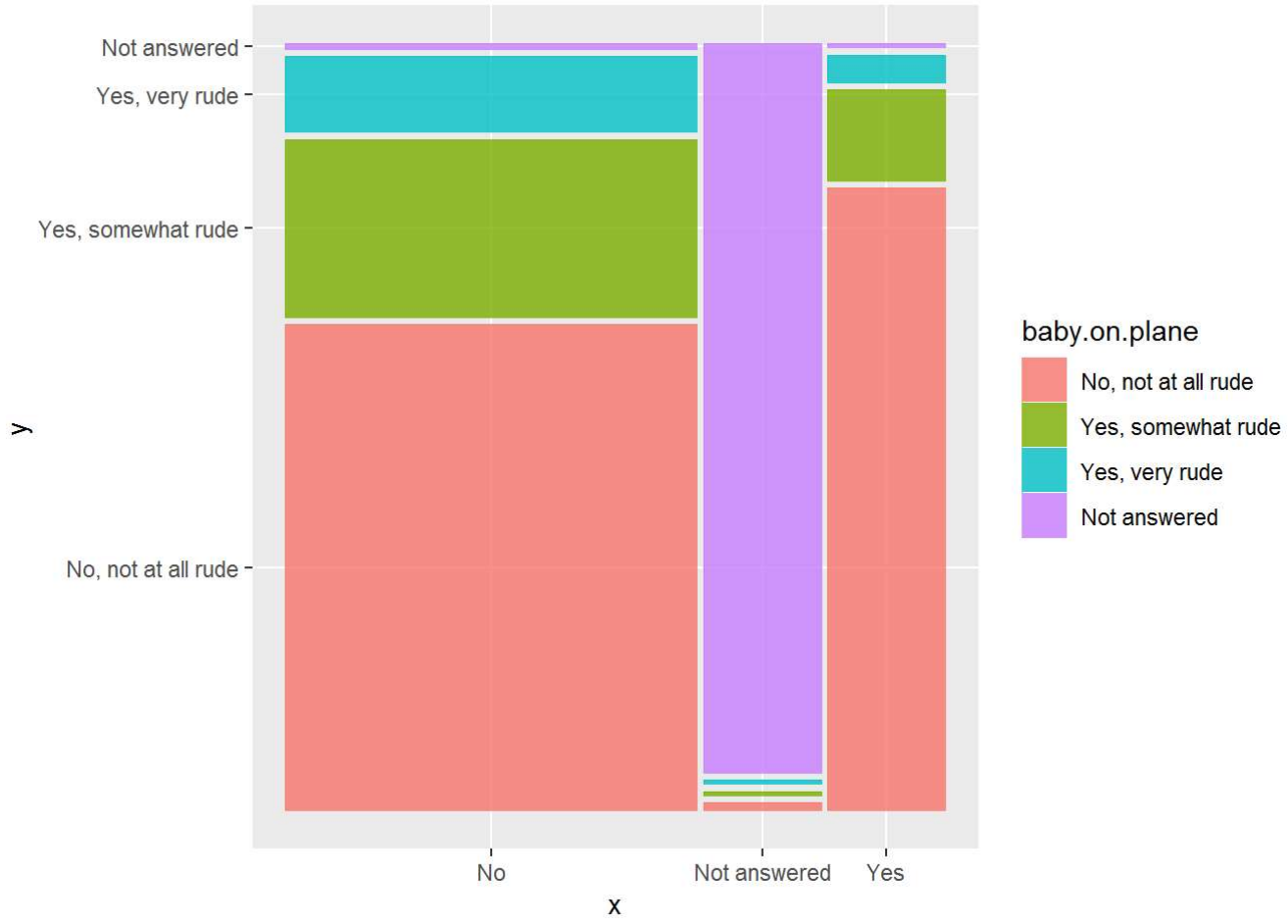



```
library(ggmosaic)
ggplot(data = fly)  +
  geom_mosaic(aes(x = product(Gender), fill=baby.on.plane, weight=1))
```



It seems females tolarate baby on plane better than males. The percentage of females who answered baby on plane is not rude at all, is greater than corresponding percentage for males.

```
library(ggmosaic)
ggplot(data = fly)  +
  geom_mosaic(aes(x = product(Do.you.have.any.children.under.18.), fill=baby.on.plane, weight=1
))
```



Apparently, people who have children under 18, are more patiant to tolarate babies on the plane compared to the people who do not have children under 18.