

HW01

Shuyi Tao
8/28/2017

Question 1

```
data = read.csv("flying.csv")

males = data[data$gender == "Male", ]
nrow(males) # number of males
## [1] 512
female = data[data$gender == "Female", ]
nrow(female) # number of females
## [1] 561
```

Question 2

```
recline_very_rude = data[data$recline_rude == "Very", ]
nrow(recline_very_rude)
## [1] 257
```

Question 3

```
female_18_29 = female[female$age == "18-29", ]
nrow(female_18_29)
## [1] 145
```

Question 4

```
males_data = males[c("respondent_id", "gender", "baby")]

female_data = female[c("respondent_id", "gender", "baby")]
```

Question 5

```
male_baby_tolerance = nrow(males_data[males_data$baby == "No", ]) / nrow(males_data)

female_baby_tolerance = nrow(female_data[female_data$baby == "No", ]) / nrow(female_data)
```

females are more tolerant.

Question 6

```
baby_annoy = nrow(data[data$baby == "Very", ]) / nrow(data)
strangers_annoy = nrow(data[data$talk_stranger == "Very", ]) / nrow(data)
walk_annoy = nrow(data[data$wake_up_walk == "Very", ]) / nrow(data)
bathroom_annoy = nrow(data[data$wake_up_bathroom == "Very", ]) / nrow(data)

AnnoyNames = c("baby_annoy", "strangers_annoy", "walk_annoy", "bathroom_annoy")
AnnoyValues = c(baby_annoy, strangers_annoy, walk_annoy, bathroom_annoy)
max = AnnoyNames[which.max(AnnoyValues)]
```

Wake someone up to go for a walk is the most annoying to fliers.

Question 7

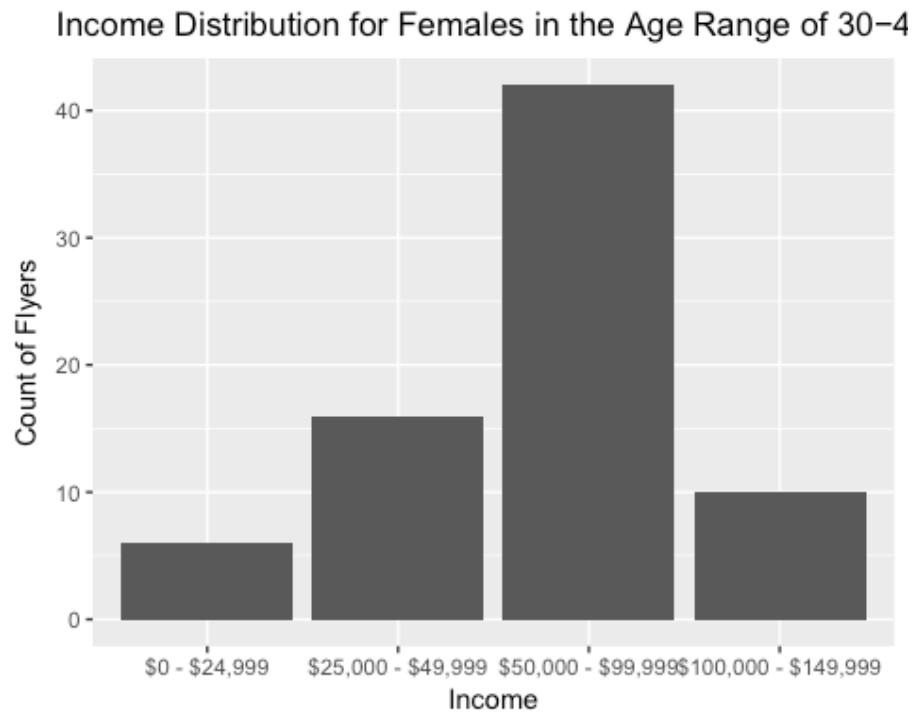
```
female_30_44 = female[female$age == "30-44", ]
male_30_44 = males[males$age == "30-44", ]

library(ggplot2)
```

Income Distribution for Females in the Age Range of 30–44

```
levs_female = levels(female_30_44$household_income)[c(1,3,4,2)]
female_30_44$household_income = factor(female_30_44$household_income, levs_female)
```

```
ggplot(na.omit(female_30_44), aes(x = household_income)) +
  geom_bar() +
  xlab("Income") +
  ylab("Count of Flyers") +
  ggtitle("Income Distribution for Females in the Age Range of 30–44") +
  theme(plot.title = element_text(hjust = 0.5))
```



Income Distribution for Males in the Age Range of 30–44

```
levs_male = levels(male_30_44$household_income)[c(1,3,4,2)]
male_30_44$household_income = factor(male_30_44$household_income, levs_male)
```

```
ggplot(na.omit(male_30_44), aes(x = household_income)) +
  geom_bar() +
  xlab("Income") +
  ylab("Count of Flyers") +
  ggtitle("Income Distribution for Males in the Age Range of 30–44") +
  theme(plot.title = element_text(hjust = 0.5))
```

