Exploring the BRFSS data

Setup

Load packages

```
library(ggplot2)
library(dplyr)
```

Load data

Make sure your data and R Markdown files are in the same directory. When loaded your data file will be called `brfss2013`. Delete this note when before you submit your work.

```
load("brfss2013.RData")
```

Part 1: Data

Data were collected through monthly phone interviews. In conducting the BRFSS landline telephone survey, interviewers collect data from a randomly selected adult in a household. In conducting the cellular telephone version of the BRFSS questionnaire, interviewers collect data from an adult who participates by using a cellular telephone and resides in a private residence or college housing.

Because of the observational nature of the collection method, the scope of the inference is generalizable but cannot infer causation. Since only a percentage of people called completed the survey, there is also non-response bias in the results.

Part 2: Research questions

Research quesion 1: Do adults consuming more fruits and vegetables have more healthy days? Do female or male consume more fruits and vegetables?

Research quesion 2: Do adults participated in 150+ minutes have longer hours of sleep than ones with 0 minutes?

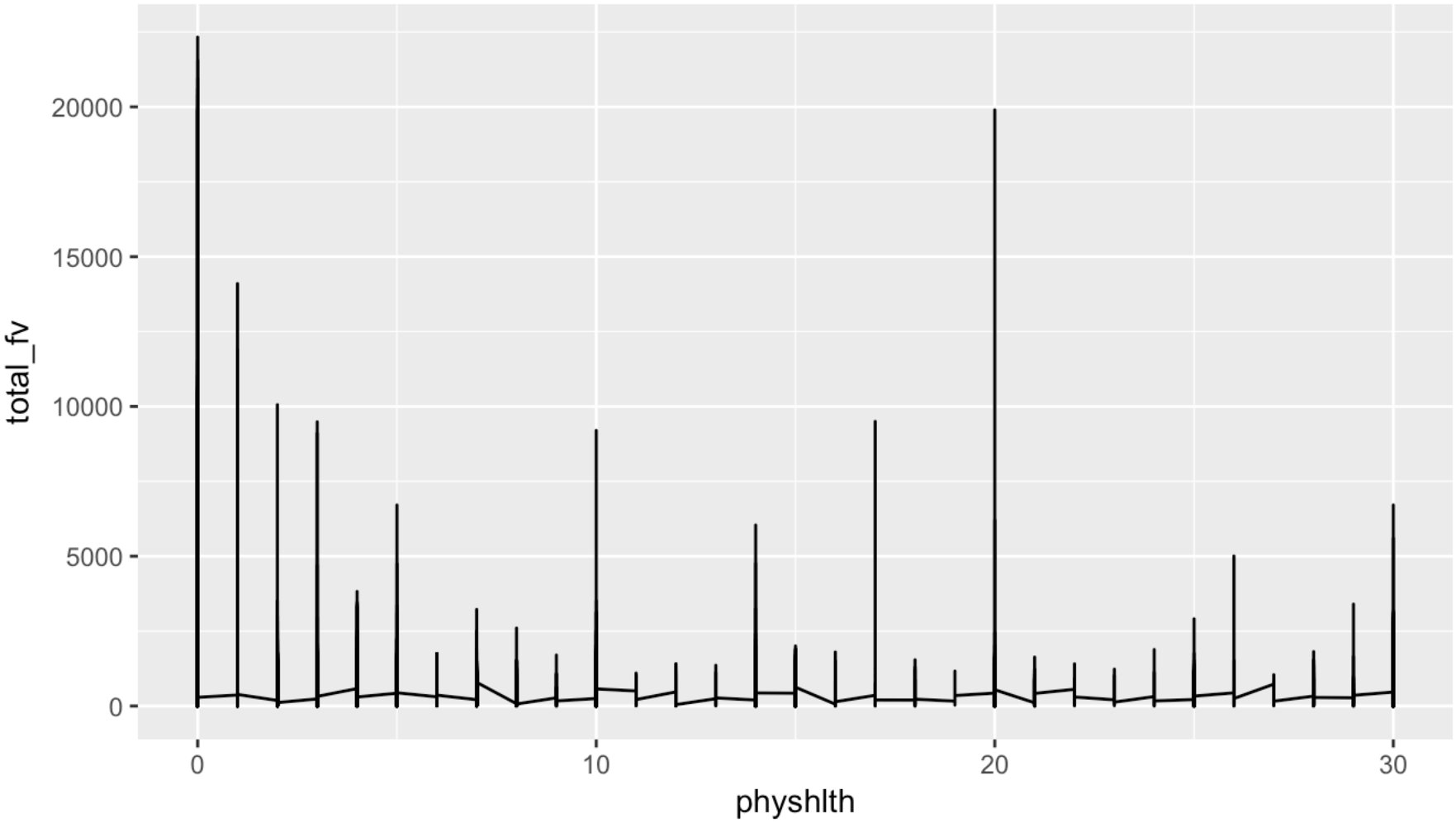
Research quesion 3: Do adults participated in 150+ minutes exercise consume more fruits and vegetabels and less sugar than ones with 0 minutes?

Part 3: Exploratory data analysis

NOTE: Insert code chunks as needed by clicking on the “Insert a new code chunk” button (green button with orange arrow) above. Make sure that your code is visible in the project you submit. Delete this note when before you submit your work.

Research quesion 1:

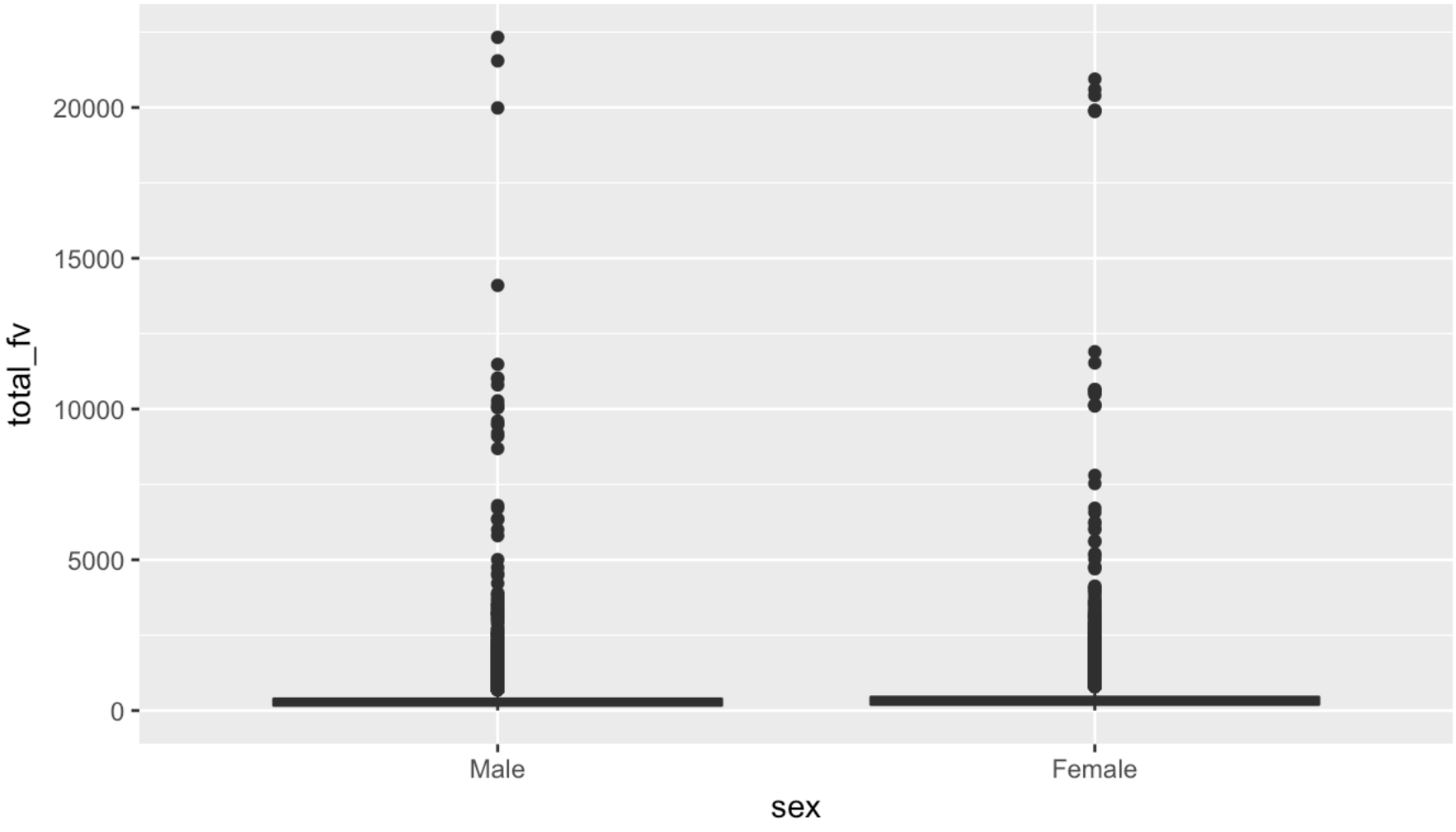
```
q1 <- brfss2013 %>% select(X_frutsum, X_vegesum, physhlth, sex) %>% filter(!is.na(X_f
rutsum), !is.na(X_vegesum), !is.na(physhlth), (sex == "Male" | sex == "Female"))
q1 <- q1 %>% mutate(total_fv = X_frutsum + X_vegesum)
ggplot(data = q1, aes(x = physhlth, y = total_fv)) + geom_line()
```




```
q1 %>% group_by(sex) %>% summarise(mean_total_fv = mean(total_fv), sd_total_fv = sd(t
otal_fv))
```

```
## # A tibble: 2 × 3
##      sex mean_total_fv sd_total_fv
##   <fctr>         <dbl>       <dbl>
## 1   Male      301.8888    238.6544
## 2 Female      350.3758    250.4559
```

```
ggplot(data = q1, aes(x = sex, y = total_fv)) + geom_boxplot()
```




```
q1 %>% group_by(sex) %>% summarise(mean_total_fv = mean(total_fv), sd_total_fv = sd(t
otal_fv), mean_hthdays = mean(physhlth), sd_hthdays = sd(physhlth))
```

```
## # A tibble: 2 × 5
##      sex mean_total_fv sd_total_fv mean_hthdays sd_hthdays
##   <fctr>         <dbl>       <dbl>        <dbl>      <dbl>
## 1   Male      301.8888    238.6544     3.909134   8.565830
## 2 Female      350.3758    250.4559     4.654360   9.012717
```

Research quesion 2:

```
q2 <- brfss2013 %>% select(X_pa150r2,sleptim1) %>% filter(!is.na(X_pa150r2), !is.na(s
leptim1))
q2 %>% group_by(X_pa150r2) %>% summarise(mean_sleep_time = mean(sleptim1), sd = sd(sl
eptim1))
```

```
## # A tibble: 3 × 3
##       X_pa150r2 mean_sleep_time       sd
##          <fctr>           <dbl>    <dbl>
## 1  150+ minutes        7.085595 1.314776
## 2 1-149 minutes        7.018778 1.354564
## 3     0 minutes        7.002169 1.710144
```

Research quesion 3:

```
q3 <- brfss2013 %>% select(X_pa150r2, X_frutsum, X_vegesum, ssbsugar, ssbfrut2) %>%
filter(!is.na(X_pa150r2), !is.na(X_frutsum), !is.na(X_vegesum), !is.na(ssbsugar), !is
.na(ssbfrut2))
q3 <- q3 %>% mutate(total_fv = X_frutsum + X_vegesum, total_sugar = ssbsugar + ssbfru
t2)
q3 %>% group_by(X_pa150r2) %>% summarise(mean_fv = mean(total_fv), sd_fv = sd(total_f
v), mean_sugar = mean(total_sugar), sd_sugar = sd(total_sugar))
```

```
## # A tibble: 3 × 5
##       X_pa150r2  mean_fv    sd_fv mean_sugar sd_sugar
##          <fctr>    <dbl>    <dbl>      <dbl>    <dbl>
## 1  150+ minutes 371.8916 238.7625   224.9966 207.1892
## 2 1-149 minutes 309.7573 193.2516   257.6872 207.7046
## 3     0 minutes 266.7621 197.3934   227.7259 197.8119
```