IDS 400 Assignment 1 Part 2

Timothy Troutman (650212956)

September 28th, 2025

Questions

1. Which day of the week has the least number of diners on average?
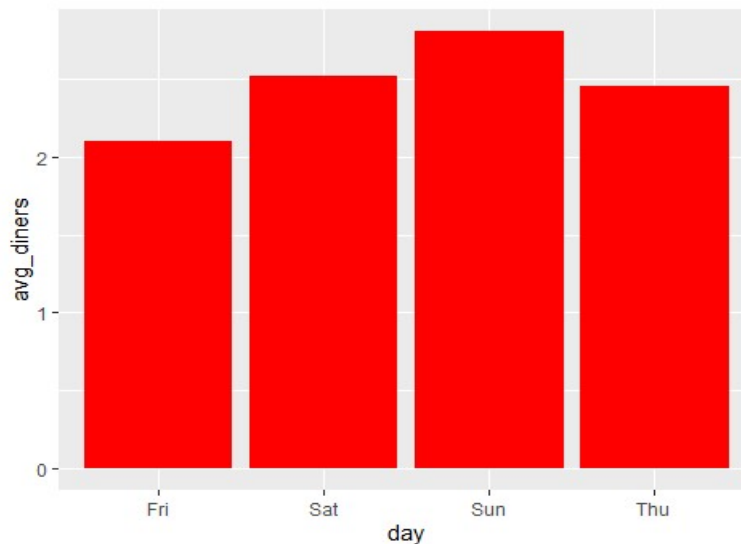
```r
answer1 <- tips_data %>%
  group_by(day) %>%
  summarize(avg_diners = mean(size, na.rm = TRUE)) %>%
  arrange(desc(avg_diners))
pander(answer1)
```

day	avg_diners
Sun	2.811
Sat	2.517
Thu	2.452
Fri	2.105

Friday has the least number of diners on average.

**2. Create a bar chart (ggplot2) showing the average number of diners on each day?**

```r
ggplot(data = answer1, mapping = aes(y = avg_diners, x = day)) +
  geom_col(fill = "red")
```

3. The default bar chart displays the bars by alphabetical order (Friday, Saturday, Sun, and Thursday).Rearrange to make the bars follow the normal week day order (Thursday, Friday, Saturday, and Sunday).

```
answer1$day <- factor(answer1$day, levels = c("Thu", "Fri", "Sat", "Sun"))

ggplot(data = answer1, mapping = aes(y = avg_diners, x = day)) +
  geom_col(fill = "blue") +
  xlab("Weekday") +
  ylab("Average Diners") +
  ggtitle("Average Diners by Weekday")
```
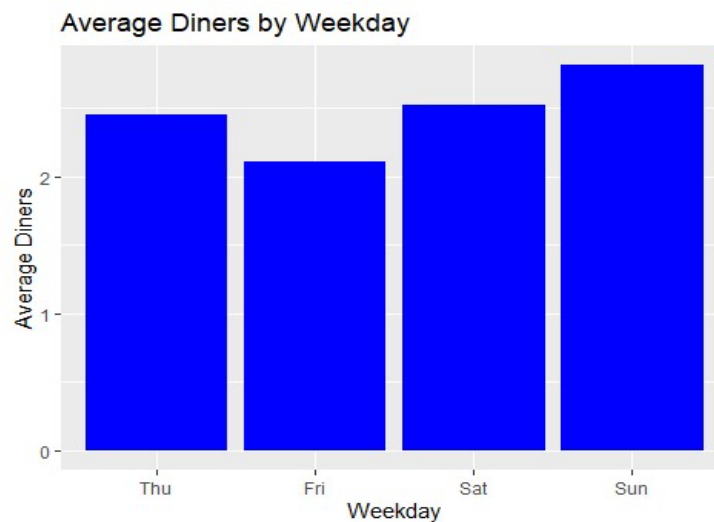

4. How many female smokers are in the tips dataset who have dined on a Friday?

```
answer4 <-
  count(filter(tips_data, smoker == "Yes", sex == "F", day == "Fri"))
pander(answer4)
```

n
7

==*7 female smokers have dined on a Friday.*==

5. On average, who tips more? Males or Females?

```
answer5 <- tips_data %>%
  group_by(sex) %>%
  summarize(avg_tip = mean(tip))
pander(answer5)
```

| sex	avg_tip
F	2.833
M	3.09

==*On average, Males tip more.*==

6. Create a new variable called tipRate (tip/totbill). Add this variable to your dataset. Create a boxplot for this variable.

```
tips_data <-
  mutate(tips_data, tipRate = tip/totbill)

ggplot(data = tips_data, mapping = aes(y = tipRate, x = "")) +
  geom_boxplot() +
  ylab("Tip Rate") +
  ggtitle("Distribution of Tip Rates")
```
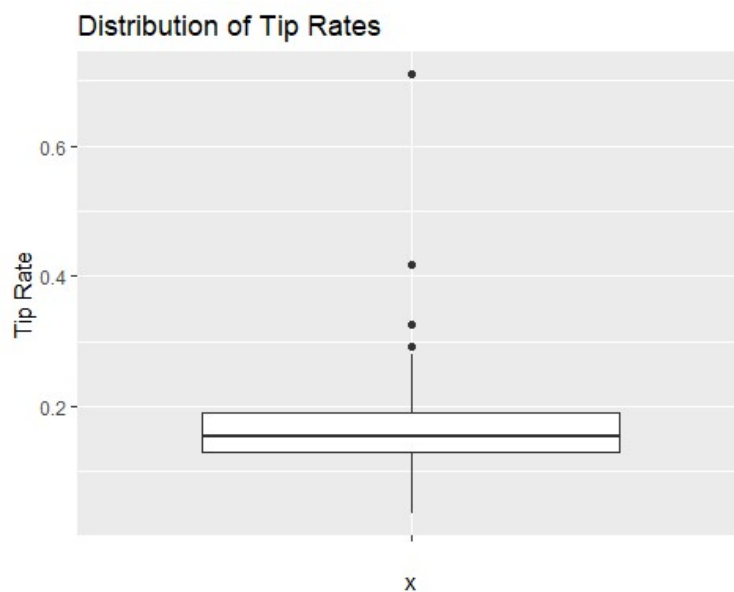


```
ggplot(data = tips_data, mapping = aes(y = tipRate, x = sex)) +
  geom_boxplot() +
  ylab("Tip Rate") +
  ggtitle("Distribution of Tip Rates by Sex")
```
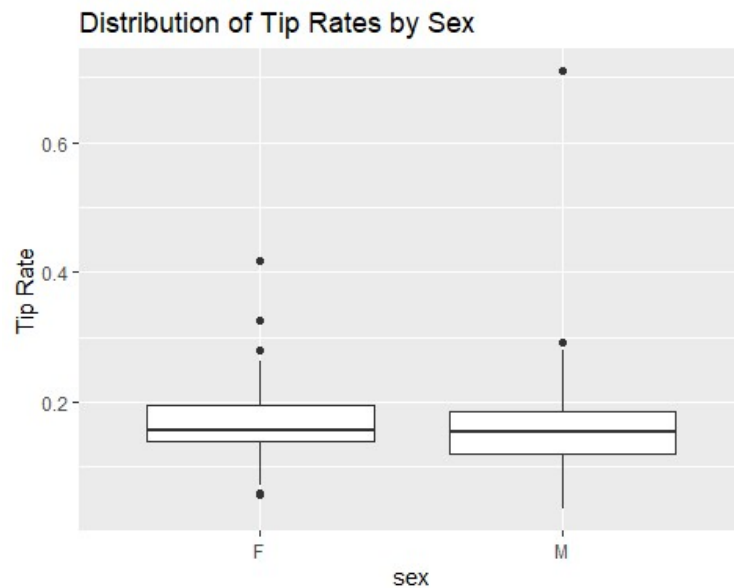
7. Using the results from questions 4 and 5 above, can we conclude that females are more generous in tipping? Explain.

==*When comparing tip rate by sex in a box plot, it is observed that the median tip rate is slightly higher for females than it is for males. The reason that the mean tip rate is higher for males is due to an extreme outlier which is pulling the average up. When looking at the Interquartile range, the 25th and 75th percentile are higher for females compared to males. This means that when looking at the middle range, females tend to tip more. Because the median and interquartile range are higher for females than males, we can conclude that females are more generous in tipping.*==

8. What is the number of females in this dataset who paid more than 15% tips?

```
answer8 <-
  count(filter(tips_data, sex == "F", tipRate > .15))
```

==*52 females paid more than 15% tips.*==

9. Create a scatter plot with totbill on the x-axis and tip on the y-axis. The chart should show data points corresponding to female customers in a different colour from points corresponding to male customers.

```
ggplot(data = tips_data, mapping = aes(x = totbill, y = tip, color = sex)) +
  geom_point() +
  xlab("Total Bill") +
  ylab("Tip") +
  ggtitle("Distribution of Total Bill vs. Tip")
```
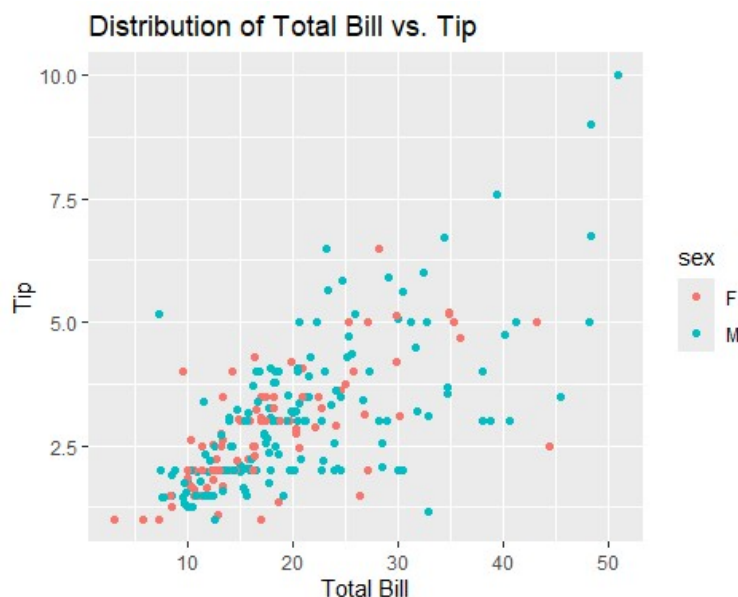

10. How many of the parties had more than 2 people? (Hint: check for NA values and make sure they are not being counted).

```
answer10 <-
  count(filter(tips_data, size > 2 & !is.na(size)))
```

==*82 parties had more than 2 people.*==