

Chapter 2

2.1

Compute the mean \bar{x} and median, m , of the six numbers: 3, 5, 8, 15, 20, 21, 24, then apply the natural log to the data.

```
x <- c(3, 5, 8, 15, 20, 21, 24)
xt <- log(x)
```

Does $\bar{x} = \tilde{x}$?

```
log(mean(x)) == mean(xt)
```

```
[1] FALSE
```

Does $m = \tilde{m}$?

```
log(median(x)) == median(xt)
```

```
[1] TRUE
```

2.2

Compute the mean \bar{x} and median of the eight numbers: 1, 2, 4, 5, 6, 8, 11, 15.

Let $f(x) = \sqrt{x}$

Apply the transformation, then compute the mean, *tildex* and median, m, of the transformed data.

```
x <- c(1, 2, 4, 5, 6, 8, 11, 15)
xt <- sqrt(x)
```

- Is $f(\bar{x}) = \tilde{x}$?

```
sqrt(mean(x)) == mean(xt)
```

```
[1] FALSE
```

- Is $f(m) = \tilde{m}$?

```
sqrt(median(x)) == median(xt)
```

```
[1] FALSE
```

2.4

Import the flights data.

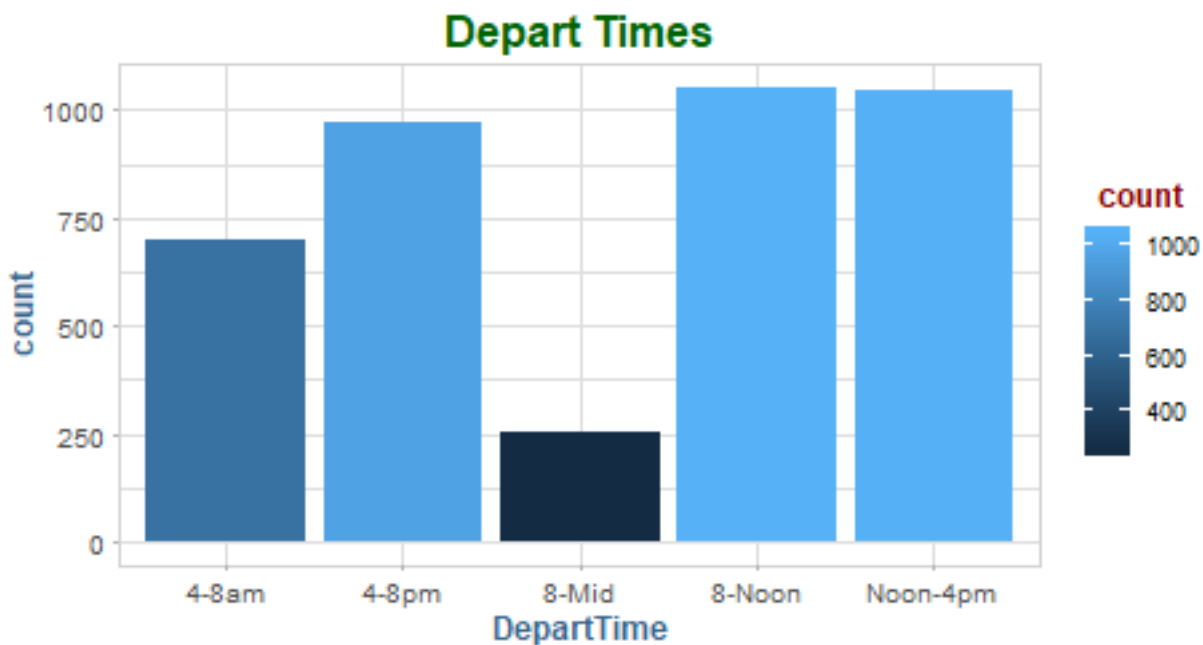
```
flights <- data.table(read.csv(paste0(data.dir, "FlightDelays.csv"),  
                               header = T))
```

a.) Create a table and bar chart of the departure times (*DepartTimes*)

```
table(flights$DepartTime)
```

4-8am	4-8pm	8-Mid	8-Noon	Noon-4pm
699	972	257	1053	1048

```
ggplot(flights, aes(DepartTime)) +  
  geom_bar(aes(fill = ..count..)) +  
  labs(title = "Depart Times")
```



b.) Create a contingency table of the variables *Day* and *Delay30*.

```
delay <- table(flights$Day, flights$Delayed30)

pretty_kable(delay, "Flight Delays")
```

Table 1: Flight Delays

	No	Yes
Fri	493	144
Mon	569	61
Sat	406	47
Sun	507	44
Thu	434	132
Tue	535	93
Wed	488	76

Show the proportions of delayed flights, by day:

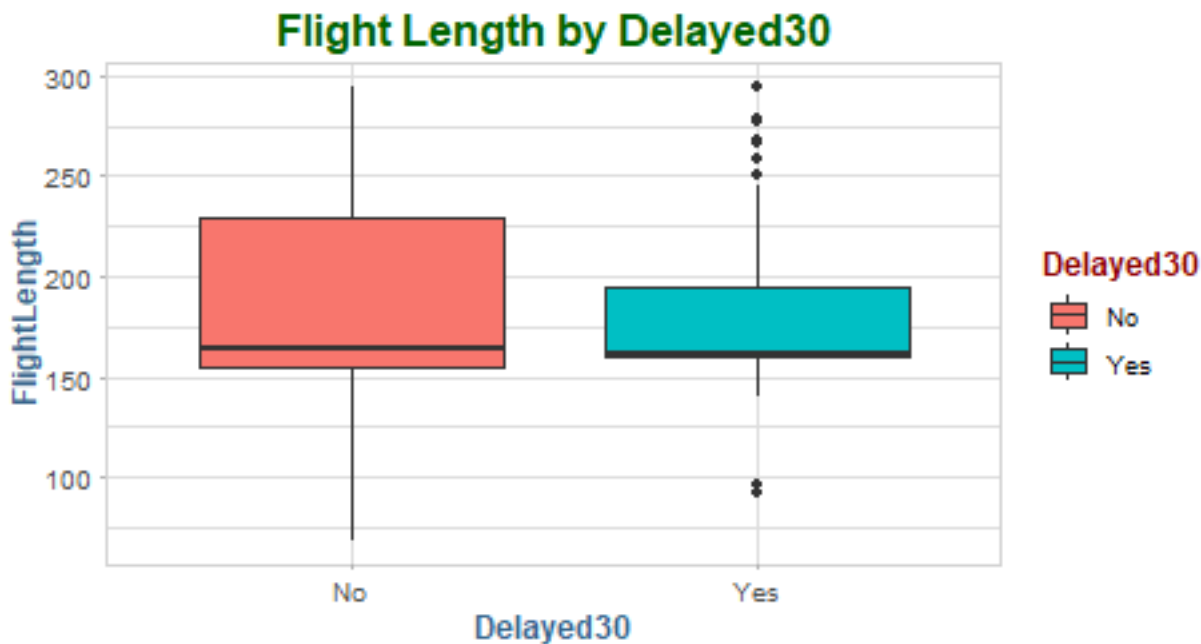
```
pretty_kable(round(prop.table(delay), 4) * 100, "Flight Delays Proportions")
```

Table 2: Flight Delays Proportions

	No	Yes
Fri	12.24	3.57
Mon	14.12	1.51
Sat	10.08	1.17
Sun	12.58	1.09
Thu	10.77	3.28
Tue	13.28	2.31
Wed	12.11	1.89

c.) Create side-by-side boxplots of the lengths of flight times, grouped by whether or not the flight was delayed at least 30 minutes:

```
ggplot(flights) +  
  geom_boxplot(aes(Delayed30, FlightLength, fill = Delayed30)) +  
  labs(title = "Flight Length by Delayed30")
```



d.) Do you think there is a relationship between the length of the flight and whether or not the departure time is delayed by at least 30 minutes?

The average flight time is the same, however, the flights that are delayed 30 minutes or more seem to be shorter overall.