

# P8130 Recitation 1: Sept 18th/20th

*Yutao Liu*

*September 1, 2017*

```
rm( list = ls() ) # clear workspace

if ( !require(pacman) ) install.packages('pacman')
pacman::p_load(readr, dplyr, ggplot2)
```

## 1) Appendectomies

### a) Data Input

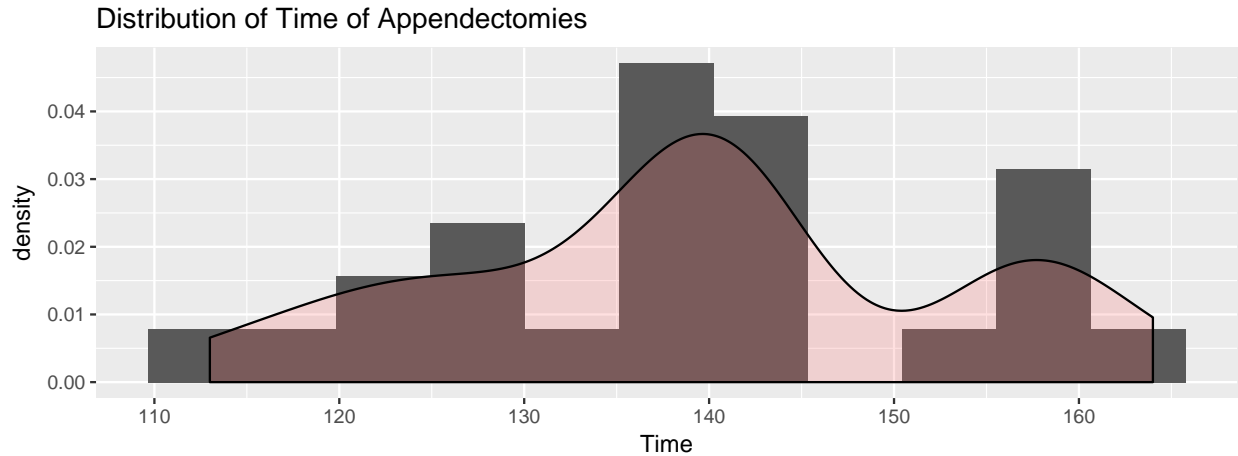
```
appy <- c(113, 118, 121, 123, 126, 128, 130, 135, 136, 137,
          138, 139, 140, 140, 142, 142, 142, 142, 143, 155,
          157, 157, 158, 159, 164) %>% data.frame(Time = .)

appy %>% head(., n = 10)
```

```
##      Time
## 1     113
## 2     118
## 3     121
## 4     123
## 5     126
## 6     128
## 7     130
## 8     135
## 9     136
## 10    137
```

### b) Histogram and Density Plot

```
ggplot(data = appy, aes(Time) ) +
  geom_histogram(aes(y=..density..), bins = 11) +
  geom_density(alpha = .2, fill="#FF6666") +
  labs(title = "Distribution of Time of Appendectomies", x = "Time")
```



How many modes does the distribution have?

## 2) Heart rates for asthmatic patients

```
asth <- c(165, 145, 115, 110, 150, 145, 38, 140, 122, 155) %>%
  data.frame(HR = .)
```

```
asth$HR %>% mean(.)
```

```
## [1] 128.5
```

```
asth$HR %>% median(.)
```

```
## [1] 142.5
```

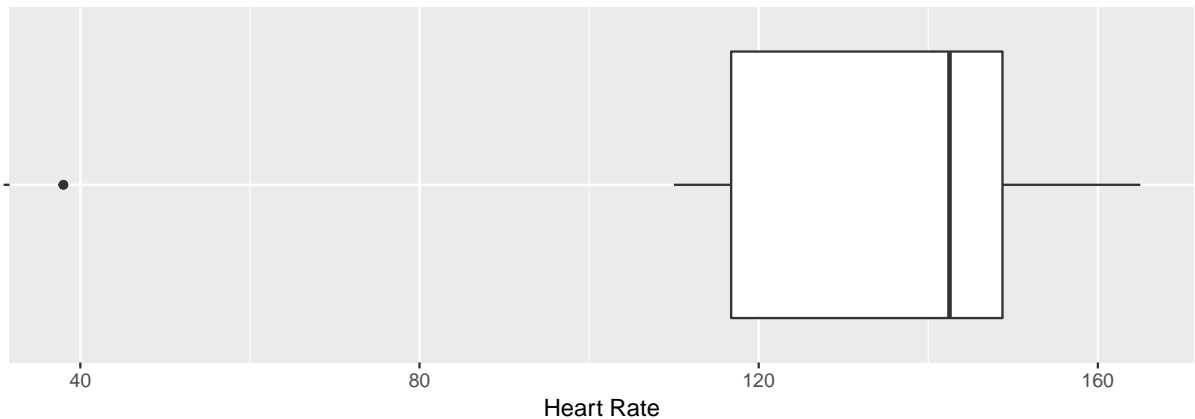
```
asth$HR %>% summary(.) # note that the algorithm for quantiles is different from the book
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##   38.0  116.8  142.5  128.5  148.8  165.0
```

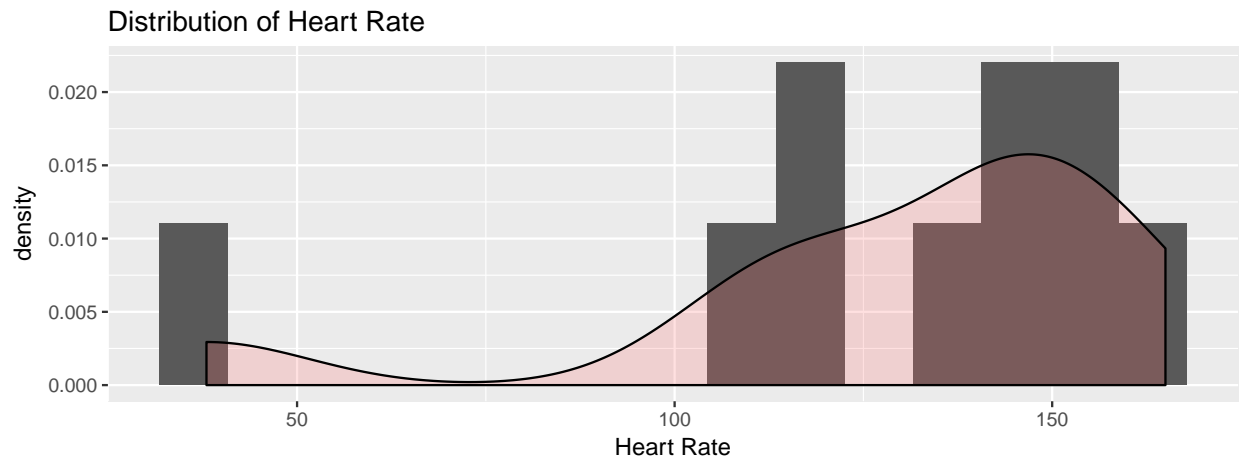
```
ggplot(data = asth, aes(x = '', y = HR)) + geom_boxplot() + coord_flip() +
  labs(title = "Box Plot of Heart Rates for Asthmatic Patients", x = "", y = "Heart Rate")
```

Box Plot of Heart Rates for Asthmatic Patients



```
ggplot(data = asth, aes(HR)) +
  geom_histogram(aes(y=..density..), bins = 15) +
```

```
geom_density(alpha = .2, fill="#FF6666") +
labs(title = "Distribution of Heart Rate", x = "Heart Rate")
```



Is the distribution symmetric?

Are there any outliers?

Which summary statistics is larger, mean or median? Why does this happen?

Which of the two summary statistics better describe the location of the distribution, mean or median? Why?

```
asth$HR %>% range(.)
```

```
## [1] 38 165
```

```
asth$HR %>% var(.)
```

```
## [1] 1325.611
```

```
asth$HR %>% sd(.)
```

```
## [1] 36.40894
```

```
sd(asth$HR) == sqrt( var(asth$HR) )
```

```
## [1] TRUE
```

```
asth$HR %>% IQR(.)
```

```
## [1] 32
```