

Week 3: Summary Statistics

Example 1

On page 100 of the OpenStax textbook, a data set shows the heights in inches for a class of 40 students. To enter the data manually, we use the function `c` to create a vector. Below we use the variable name `bh` for the heights of boys, and `gh` for girls.

```
bh <- c(66, 66, 67, 67, 68, 68, 68, 68, 68, 69, 69, 69, 70, 71, 72, 72, 72, 73, 73, 74)
gh <- c(61, 61, 62, 62, 63, 63, 63, 65, 65, 65, 66, 66, 66, 67, 68, 68, 68, 69, 69, 69)
```

We use the `fivenum` function to find the five number summary.

```
fivenum(bh)
```

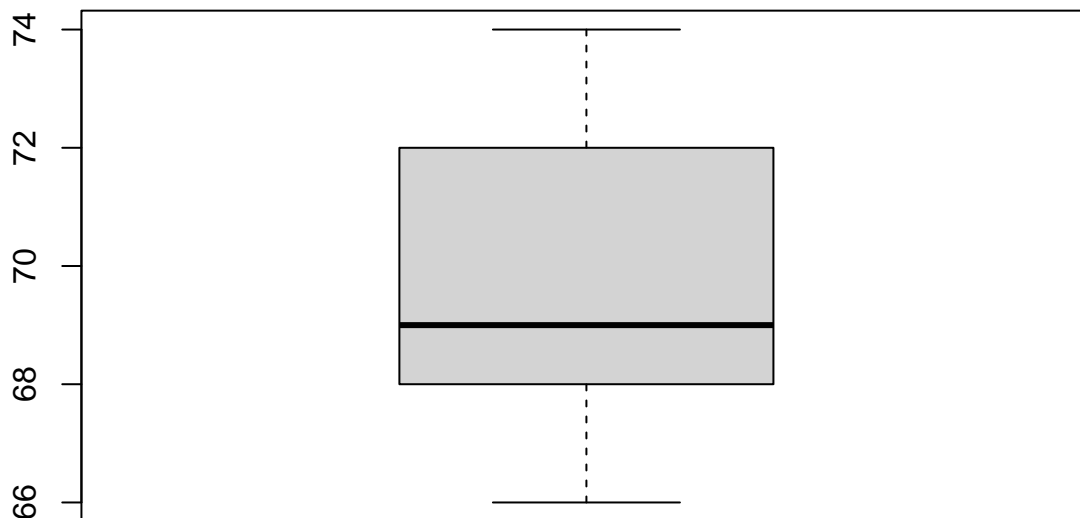
```
## [1] 66 68 69 72 74
```

```
fivenum(gh)
```

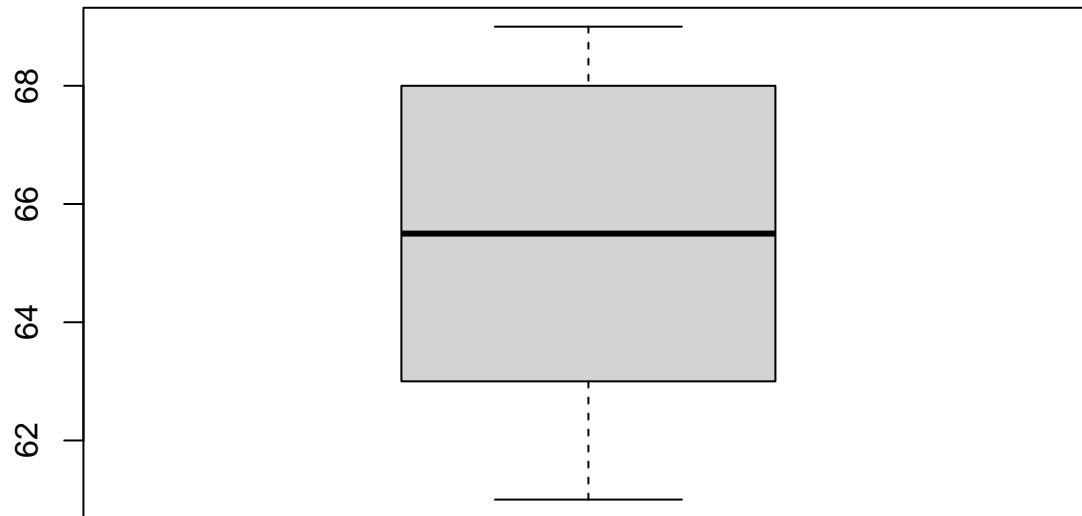
```
## [1] 61.0 63.0 65.5 68.0 69.0
```

To produce a boxplot, we do the following.

```
boxplot(bh)
```

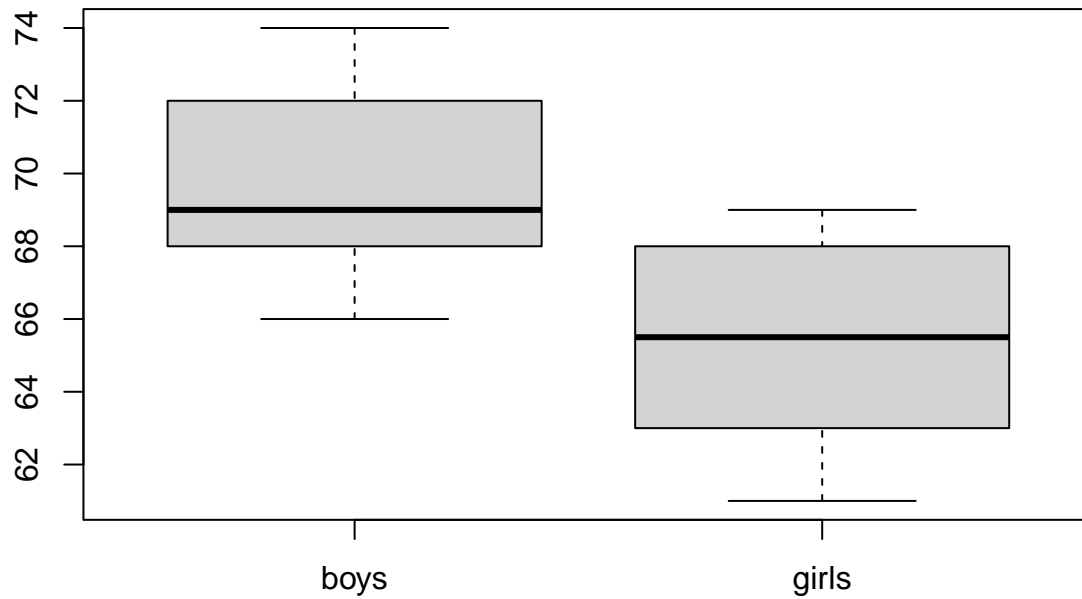


```
boxplot(gh)
```



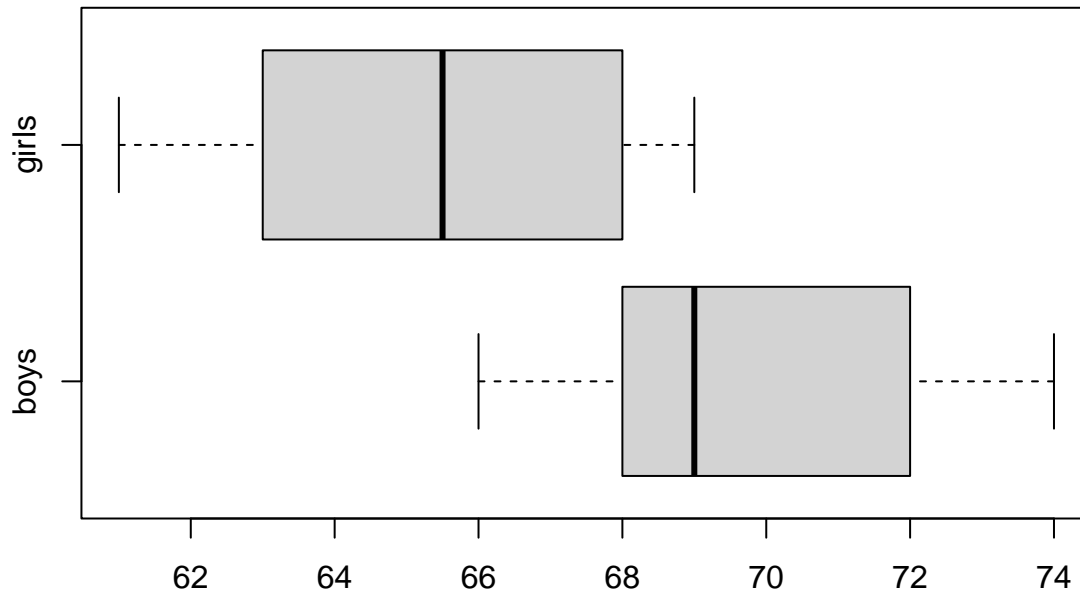
Boxplots are useful for making a comparison. We plot both girls' and boys' heights side by side.

```
boxplot(bh, gh, names = c("boys", "girls"))
```



We can change the orientation with the option shown below.

```
boxplot(bh, gh, names = c("boys", "girls"), horizontal = TRUE)
```



The sample mean and sample standard deviation can be evaluated using the mean and sd functions.

```
mean(bh)
```

```
## [1] 69.5
```

```
sd(bh)
```

```
## [1] 2.46021
```

```
mean(gh)
```

```
## [1] 65.3
```

```
sd(gh)
```

```
## [1] 2.735729
```

Example 2

A study of men's health measured 14 body characteristics of 250 men. We import the data from Professor Nick Horton's website.

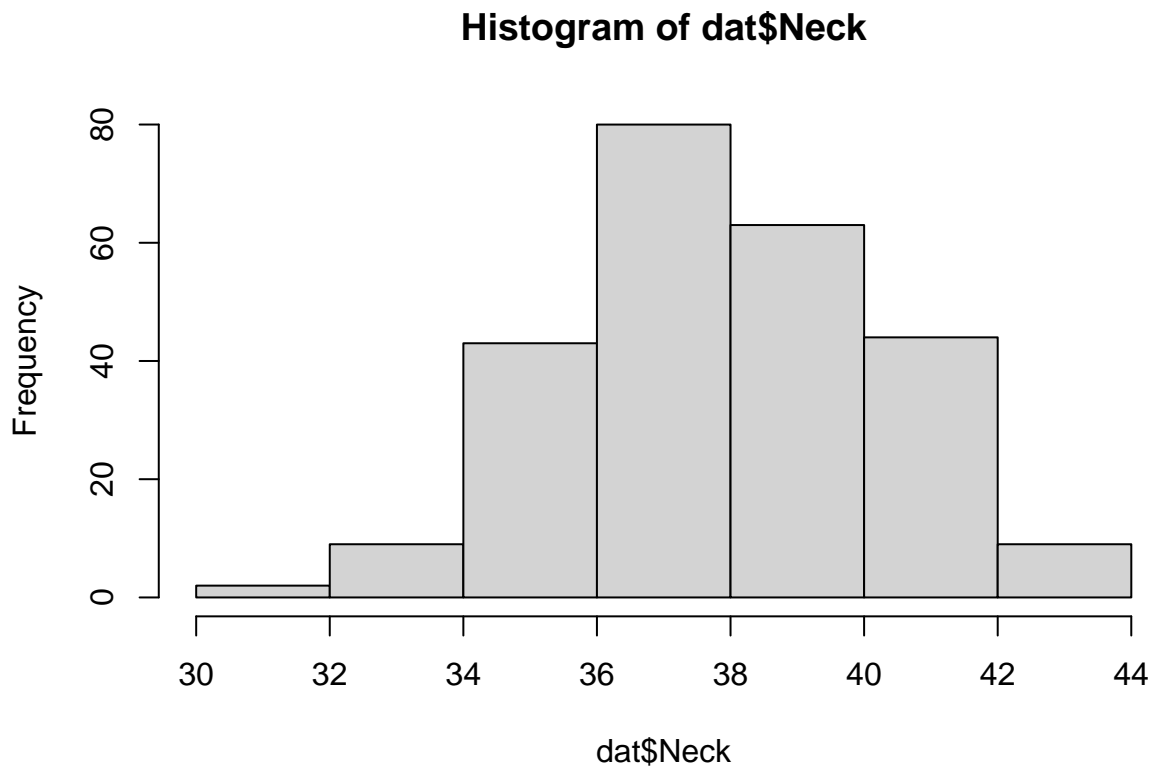
```
dat <- read.csv(url("https://nhorton.people.amherst.edu/is5/data/Bodyfat.csv"))
head(dat)
```

```
##   Density Pct.BF Age Weight Height Neck Chest Abdomen   Waist   Hip Thigh Knee
## 1  1.0708  12.3  23 154.25  67.75 36.2  93.1   85.2 33.54331  94.5  59.0 37.3
## 2  1.0853   6.1  22 173.25  72.25 38.5  93.6   83.0 32.67717  98.7  58.7 37.3
## 3  1.0414  25.3  22 154.00  66.25 34.0  95.8   87.9 34.60630  99.2  59.6 38.9
## 4  1.0751  10.4  26 184.75  72.25 37.4 101.8   86.4 34.01575 101.2  60.1 37.3
## 5  1.0340  28.7  24 184.25  71.25 34.4  97.3  100.0 39.37008 101.9  63.2 42.2
## 6  1.0502  20.9  24 210.25  74.75 39.0 104.5   94.4 37.16535 107.8  66.0 42.0
##   Ankle Bicep Forearm Wrist
## 1  21.9  32.0   27.4  17.1
## 2  23.4  30.5   28.9  18.2
## 3  24.0  28.8   25.2  16.6
## 4  22.8  32.4   29.4  18.2
## 5  24.0  32.2   27.7  17.7
```

```
## 6 25.6 35.7 30.6 18.8
```

We use the following method to product a histogram showing the neck sizes of the 250 men.

```
hist(dat$Neck)
```

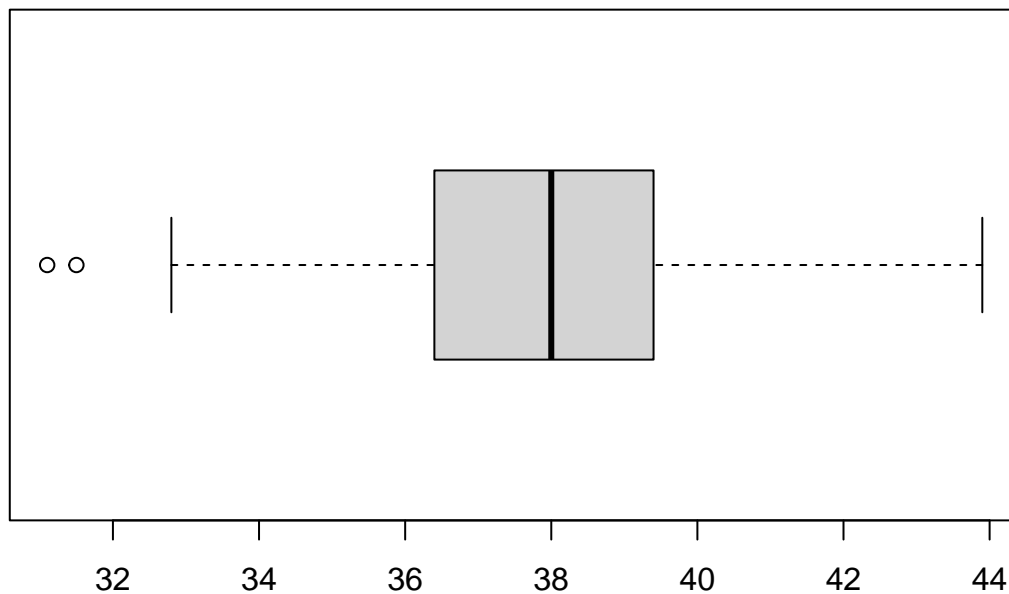


The five number summary and box plot for the neck sizes are obtained below.

```
fivenum(dat$Neck)
```

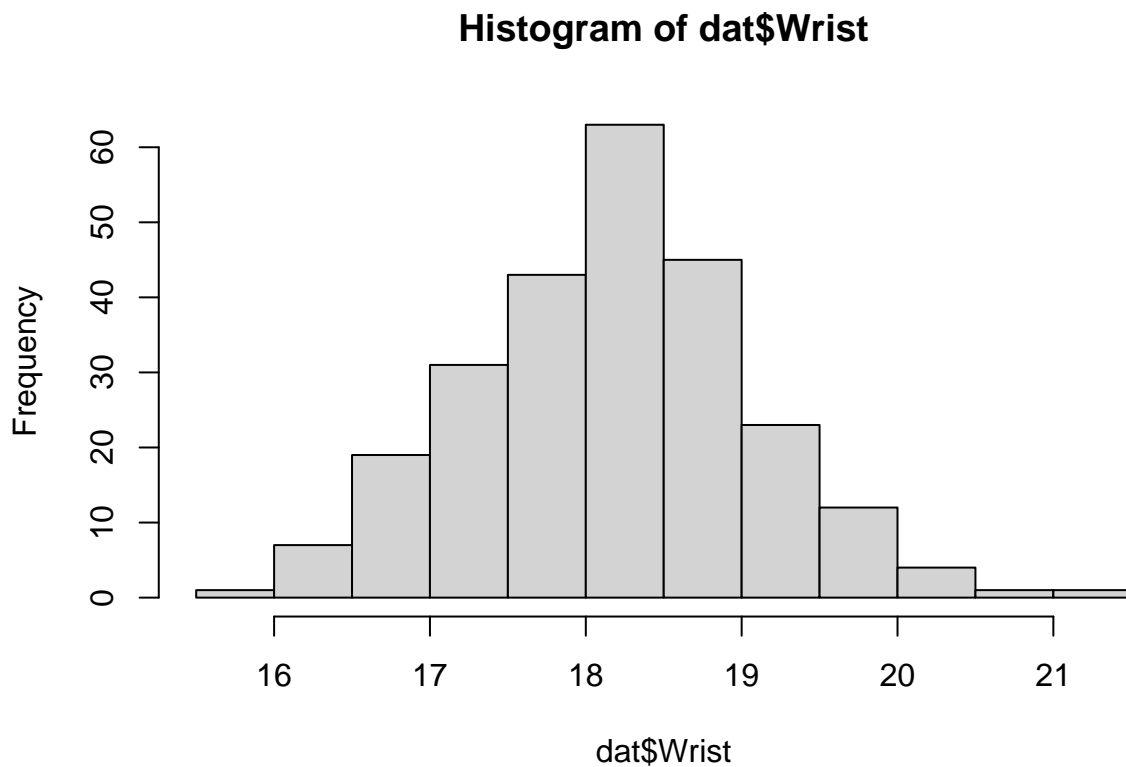
```
## [1] 31.1 36.4 38.0 39.4 43.9
```

```
boxplot(dat$Neck, horizontal = TRUE)
```



For the wrist circumference, see below.

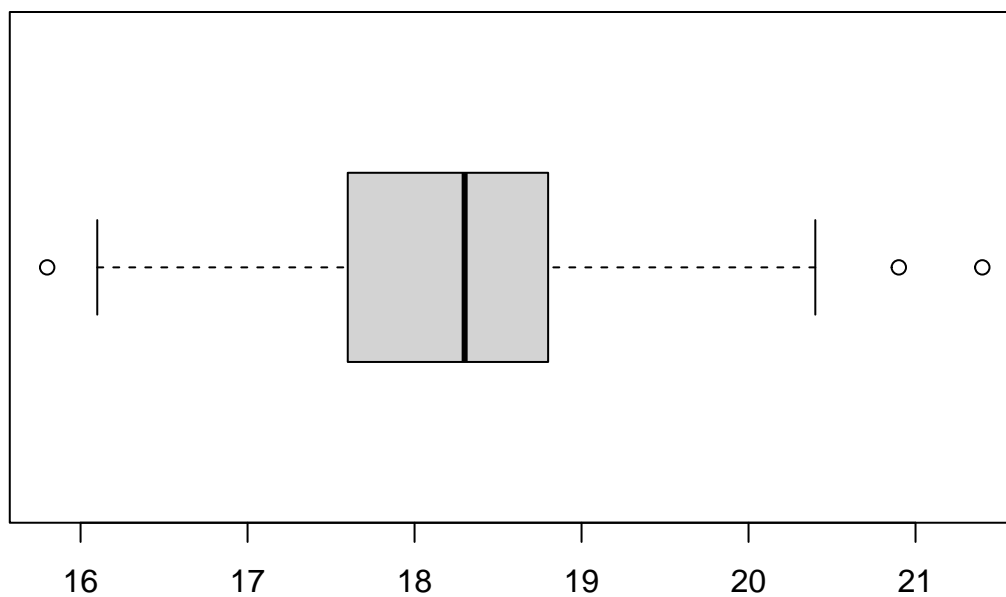
```
hist(dat$Wrist)
```



```
fivenum(dat$Wrist)
```

```
## [1] 15.8 17.6 18.3 18.8 21.4
```

```
boxplot(dat$Wrist, horizontal = TRUE)
```



Follow Up

Consider the bicep circumferences in the data set Bodyfat.

1. Construct a histogram.
2. Find the five number summary.
3. Construct a box plot.
4. Find the sample mean and standard deviation.

© 2022 Frank Wang