Do females live longer than males?

Brandon Rozek and Matt Martinez

May 6th 2017

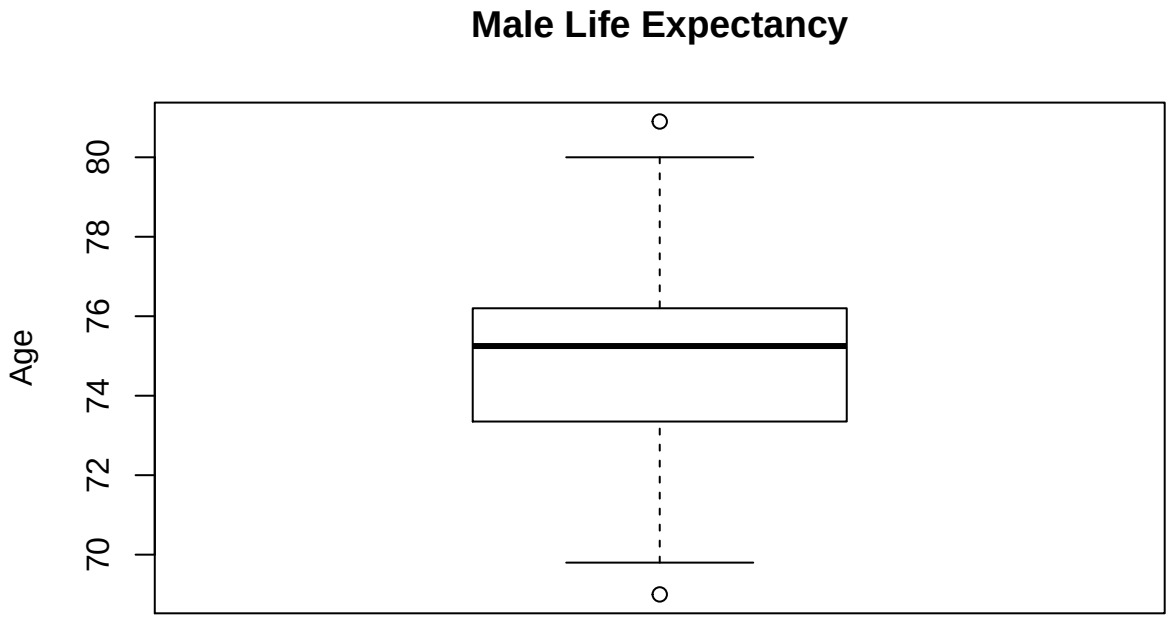
```
rm(list=ls())
# Read in file

LifeExpectancy = read.csv("~/LifeExpectancy.csv")
maleExpectancy = LifeExpectancy$Life.Expectancy.Male
femaleExpectancy = LifeExpectancy$Life.Expectancy.Female

# Summary statistics
male_row = c(min(maleExpectancy), mean(maleExpectancy), max(maleExpectancy), IQR(maleExpectancy))
female_row = c(min(femaleExpectancy), mean(femaleExpectancy), max(femaleExpectancy), IQR(femaleExpectan
summary = rbind(male_row, female_row)
colnames(summary) = c("Min", "Mean", "Max", "IQR")
rownames(summary) = c("Male", "Female")
summary
```

```
##         Min   Mean  Max   IQR
## Male   69.0 74.952 80.9 2.775
## Female 76.1 80.416 84.1 2.350
```

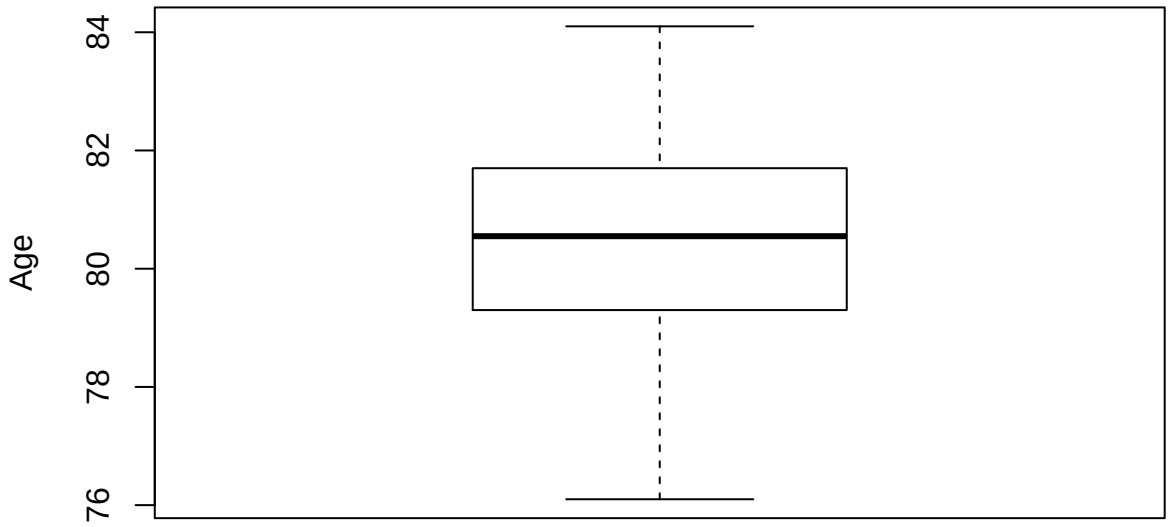
```
# Check for normality
hist(maleExpectancy, main = "Male Life Expectancy", xlab = "Age")
```


```
hist(femaleExpectancy, main = "Female Life Expectancy", xlab = "Age")
```

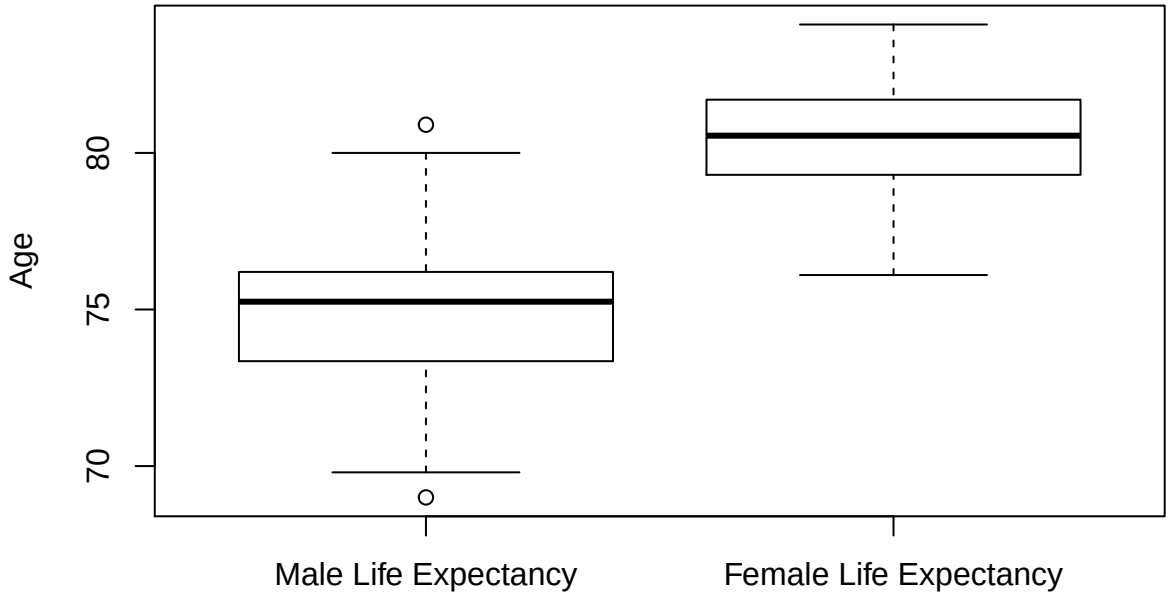

```
boxplot(maleExpectancy , ylab = "Age", main = "Male Life Expectancy")
```




```
boxplot(femaleExpectancy, ylab = "Age", main = "Female Life Expectancy")
```




```
boxplot(maleExpectancy, femaleExpectancy, names = c("Male Life Expectancy", "Female Life Expectancy"), y
```




```
# Find confidence interval
t.test(femaleExpectancy, maleExpectancy)
```

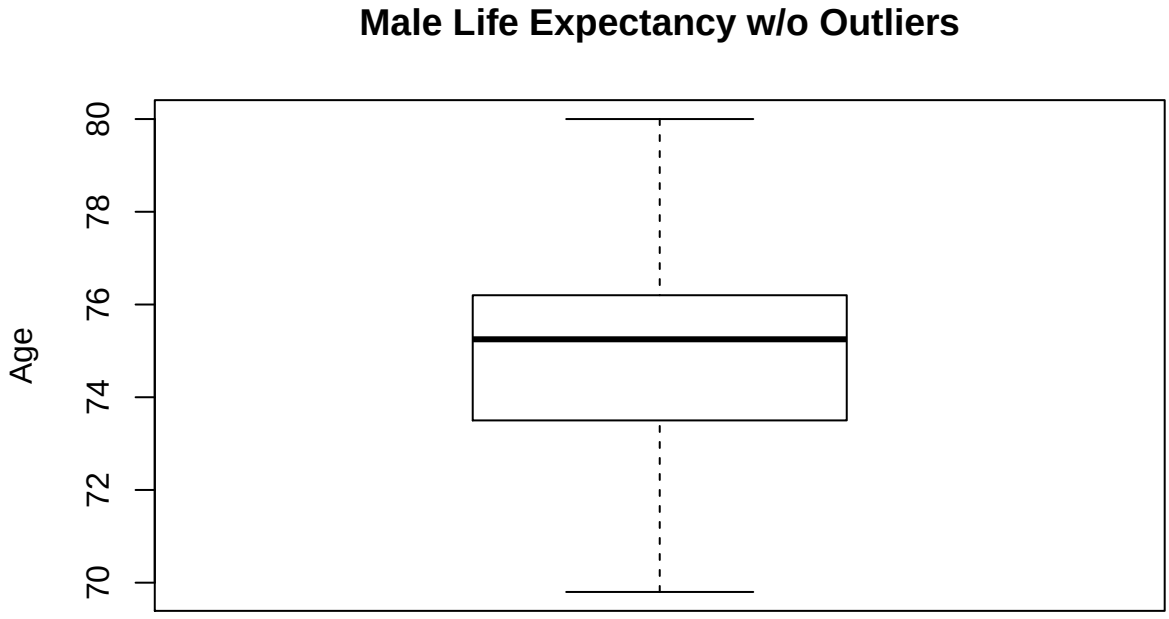
```
##
##  Welch Two Sample t-test
##
## data:  femaleExpectancy and maleExpectancy
## t = 18.858, df = 182.48, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  4.892333 6.035667
## sample estimates:
## mean of x mean of y
```

```
##    80.416    74.952
```

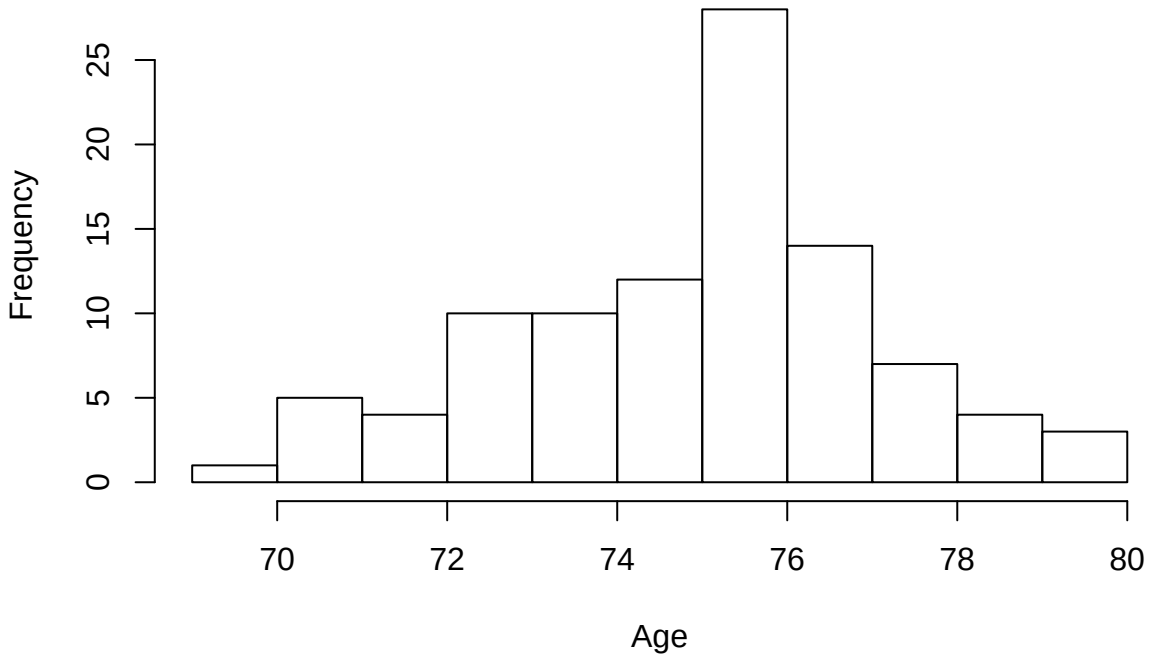
```
# Test alternative hypothesis
t.test(femaleExpectancy, maleExpectancy, alternative='g')
```

```
##
##  Welch Two Sample t-test
##
## data:  femaleExpectancy and maleExpectancy
## t = 18.858, df = 182.48, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.984992      Inf
## sample estimates:
## mean of x mean of y
##    80.416    74.952
```

```
# Remove outliers
maleExpectancy2 = maleExpectancy[!maleExpectancy %in% boxplot.stats(maleExpectancy)$out]
# Check graphs again
boxplot(maleExpectancy2, ylab = "Age", main = "Male Life Expectancy w/o Outliers")
```




```
hist(maleExpectancy2, xlab = "Age", main = "Male Life Expectancy w/o Outliers")
```




```
# Find new confidence interval
t.test(femaleExpectancy, maleExpectancy2)
```

```
##
##  Welch Two Sample t-test
##
## data:  femaleExpectancy and maleExpectancy2
## t = 19.471, df = 184.03, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  4.910317 6.017601
## sample estimates:
## mean of x mean of y
##  80.41600  74.95204
```

```
# Test new alternative
t.test(femaleExpectancy, maleExpectancy2, alternative='g')
```

```
##
##  Welch Two Sample t-test
##
## data:  femaleExpectancy and maleExpectancy2
## t = 19.471, df = 184.03, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  5.000048      Inf
## sample estimates:
## mean of x mean of y
##  80.41600  74.95204
```