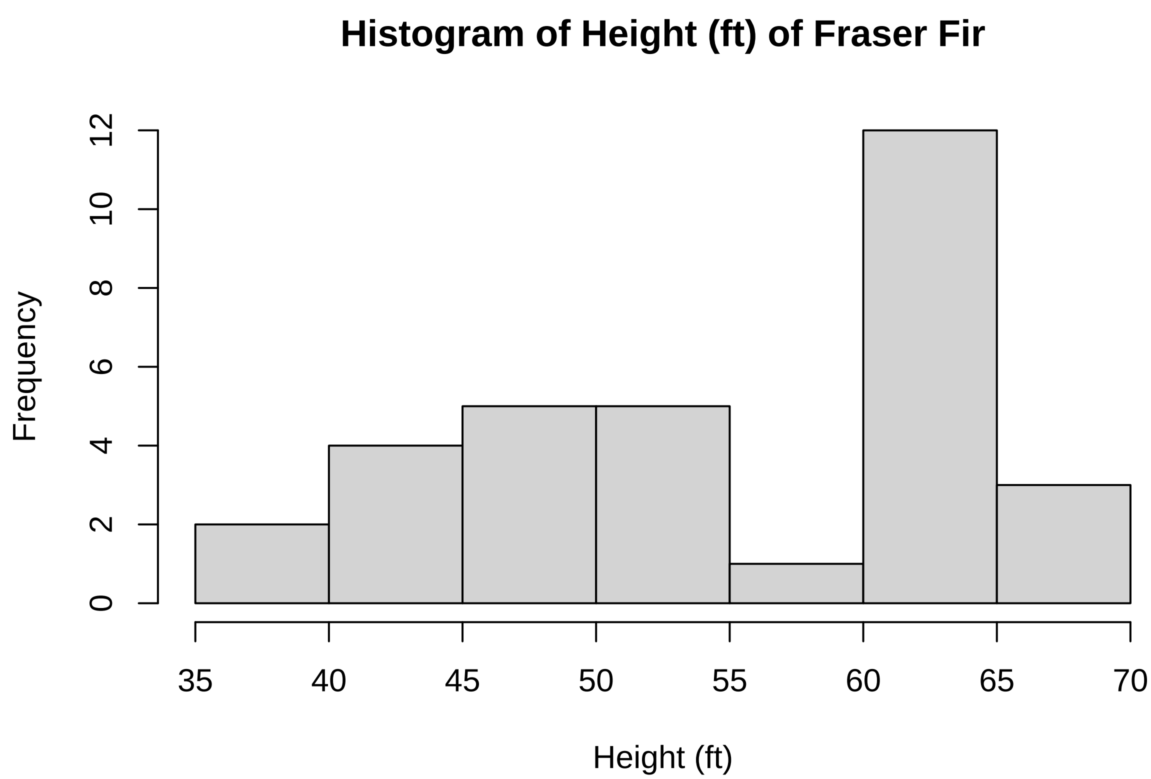
Stat 3202 Lab 9

Jane Weissberg.11

2024-07-19

1a.

```
heights <- c(43,44,54,55,44,39,48,53,
65,39,42,48,49,50,61,54,
61,61,67,53,61,64,50,63,
59,63,64,67,67,62,61,62)

hist(heights, main="Histogram of Height (ft) of Fraser Fir", xlab="Height (ft)")
```



The heights do not follow a bell-shaped curve, therefore we probably cannot assume the data is normally distributed.

1b.

```r
set.seed(515)
heights <- c(43,44,54,55,44,39,48,53,
65,39,42,48,49,50,61,54,
61,61,67,53,61,64,50,63,
59,63,64,67,67,62,61,62)

B <- 50000
muhat_storage <- c()

for (i in 1:B) {
  boot_sample <- sample(heights, replace=TRUE, size=length(heights))
  muhat_storage[i] <- mean(boot_sample)
}

quantile(muhat_storage, c(.005, .995))
```

```
##     0.5%    99.5%
## 51.40625 59.18750
```

1c.

```r
heights <- c(43,44,54,55,44,39,48,53,
65,39,42,48,49,50,61,54,
61,61,67,53,61,64,50,63,
59,63,64,67,67,62,61,62)

sample_mean <- mean(heights)
sample_sd <- sd(heights)
n <- length(heights)
alpha <- 0.01

LL <- sample_mean - qt(1 - alpha/2, df=n-1) * sample_sd / sqrt(n)
UL <- sample_mean + qt(1 - alpha/2, df=n-1) * sample_sd / sqrt(n)

round(c(LL, UL), 2)
```

```
## [1] 51.16 59.65
```

```r
t.test(heights, conf.level=0.99)
```

```
##
##  One Sample t-test
##
## data:  heights
## t = 35.833, df = 31, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  51.16331 59.64919
## sample estimates:
## mean of x
##  55.40625
```

```
differences <- c(abs(51.40625-51.16), abs(59.18750-59.65))
differences
```

```
## [1] 0.24625 0.46250
```

The confidence intervals are not too far from each other. The lower limits are 0.24625 from one another and the upper limits are 0.46250 from one another.

2a.

```
set.seed(515)
heights <- c(43,44,54,55,44,39,48,53,
65,39,42,48,49,50,61,54,
61,61,67,53,61,64,50,63,
59,63,64,67,67,62,61,62)

B <- 50000
var_storage <- c()
bs75th_storage <- c()

for (i in 1:B) {
  boot_sample <- sample(heights, replace=TRUE, size=length(heights))
  var_storage[i] <- var(boot_sample)
  bs75th_storage[i] <- quantile(boot_sample, 0.75)
}

quantile(var_storage, c(.005, .995))
```

```
##      0.5%     99.5%
##  41.57149 107.20976
```

```
quantile(bs75th_storage, c(.005, .995))
```

```
##  0.5% 99.5%
##  59.5  65.5
```

2b.

```
options(repos = c(CRAN = "https://cloud.r-project.org"))

install.packages("DescTools")
```

```
##
## The downloaded binary packages are in
##  /var/folders/66/_4vz9_y57nx1wgym4pn0zxcc0000gn/T//RtmpzMZnO6/downloaded_packages
```

```
library(DescTools)

heights <- c(43,44,54,55,44,39,48,53,
65,39,42,48,49,50,61,54,
61,61,67,53,61,64,50,63,
```

```r
59,63,64,67,67,62,61,62)

VarTest(heights, conf.level=0.99)
```

```
##
##  One Sample Chi-Square test on variance
##
## data:  heights
## X-squared = 2371.7, df = 31, p-value < 2.2e-16
## alternative hypothesis: true variance is not equal to 1
## 99 percent confidence interval:
##   43.12004 164.04461
## sample estimates:
## variance of x
##      76.50706
```

```r
differences <- c(abs(41.57149-43.12004), abs(107.20976-164.04461))
differences
```

```
## [1]  1.54855 56.83485
```

The confidence intervals are not very close. The lower limits are 1.55 away from one another, while the upper limits are 56.84 away from one another. These are fairly large differences.

2c. When we use this method, we assume that the data follows a roughly normal distribution and that the sampling distribution of the sample variance roughly follows a chi-squared distribution. The 75th percentile does not follow a distribution like that, so there is no quantile we can use to create a confidence interval in this fashion. We can use the bootstrap method like we did previously.

3a.

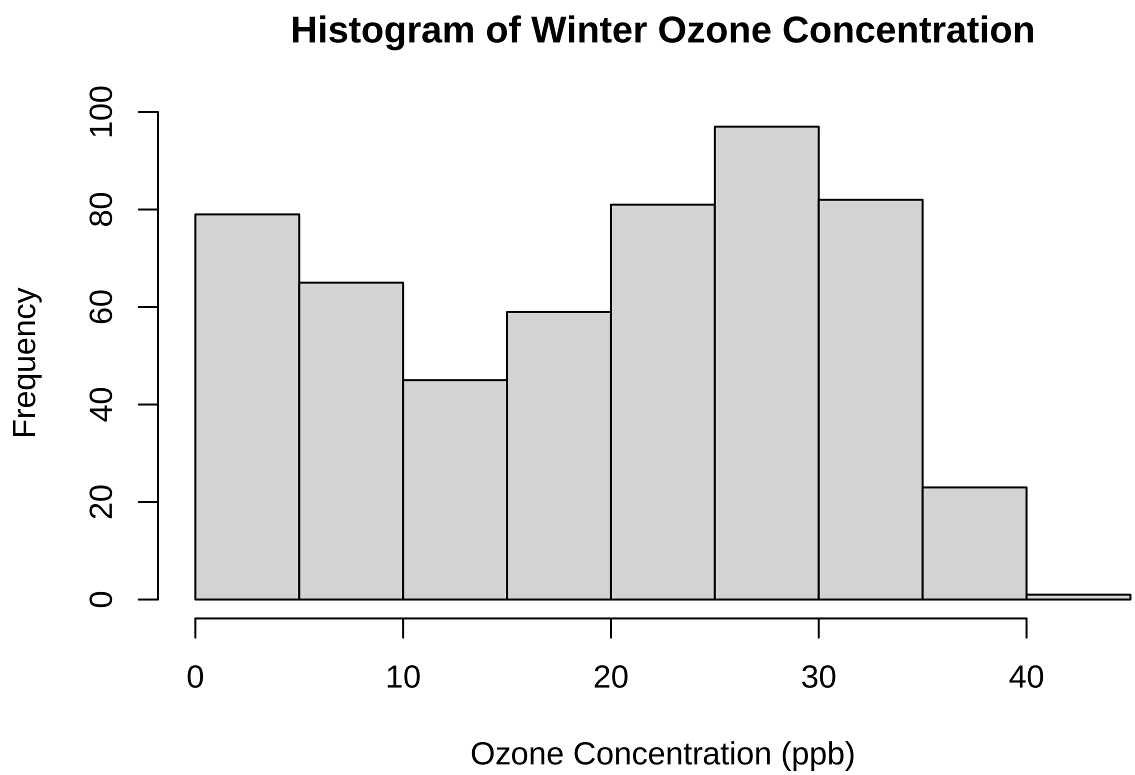
```r
install.packages("texmex")
```

```
##
## The downloaded binary packages are in
##  /var/folders/66/_4vz9_y57nx1wgym4pn0zxcc0000gn/T//RtmpzMZnO6/downloaded_packages
```

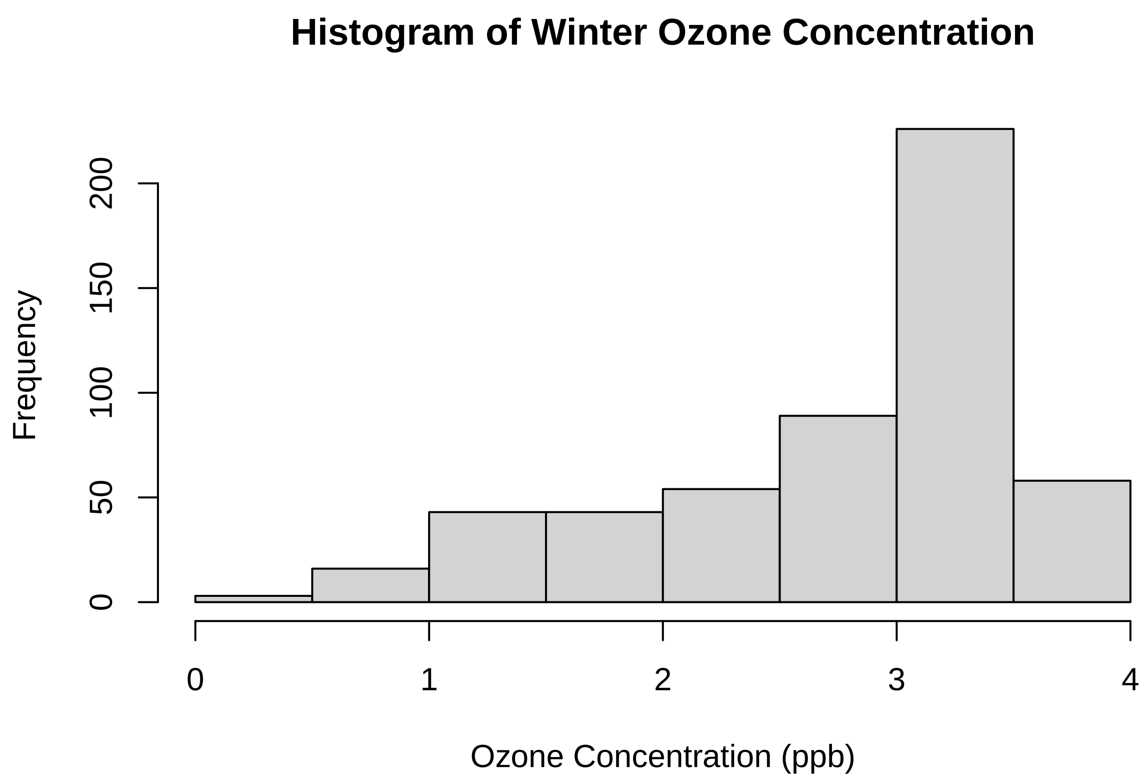
```r
library(texmex)
```

```
## Loading required package: mvtnorm

## Loading required package: ggplot2

##
## Attaching package: 'texmex'

## The following object is masked from 'package:DescTools':
##
##     rFrechet
```

```r
data("summer")
data("winter")

ozone_winter <- winter$O3
hist(ozone_winter, main = "Histogram of Winter Ozone Concentration", xlab = "Ozone Concentration (ppb)")
```




```r
hist(log(ozone_winter), main = "Histogram of Winter Ozone Concentration", xlab = "Ozone Concentration (
```




```r
set.seed(515)

B <- 50000
muhat_storage <- c()
sd_storage <- c()

for (i in 1:B) {
  boot_sample <- sample(ozone_winter, replace=TRUE, size=length(ozone_winter))
  muhat_storage[i] <- mean(boot_sample)
  sd_storage[i] <- sd(boot_sample)
}

quantile(muhat_storage, c(.005, .995))
```

```
##     0.5%    99.5%
## 18.83835 21.27257
```

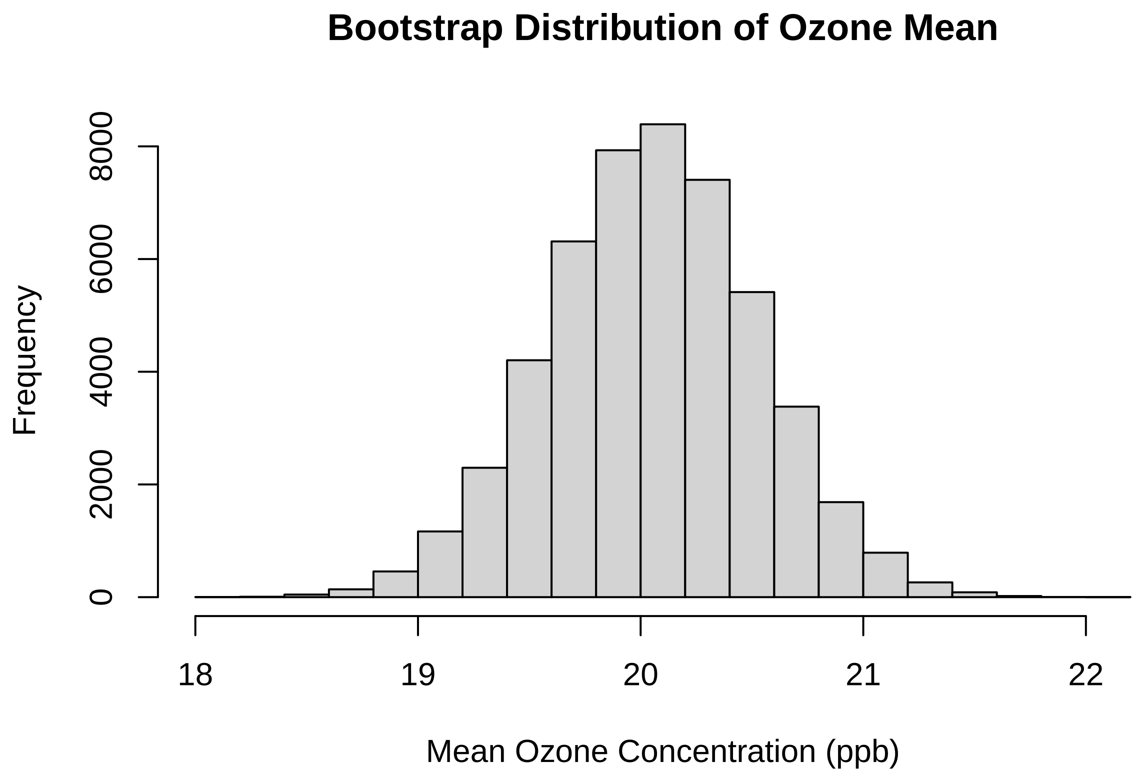
```r
quantile(sd_storage, c(.005, .995))
```

```
##     0.5%    99.5%
## 10.33765 11.42479
```

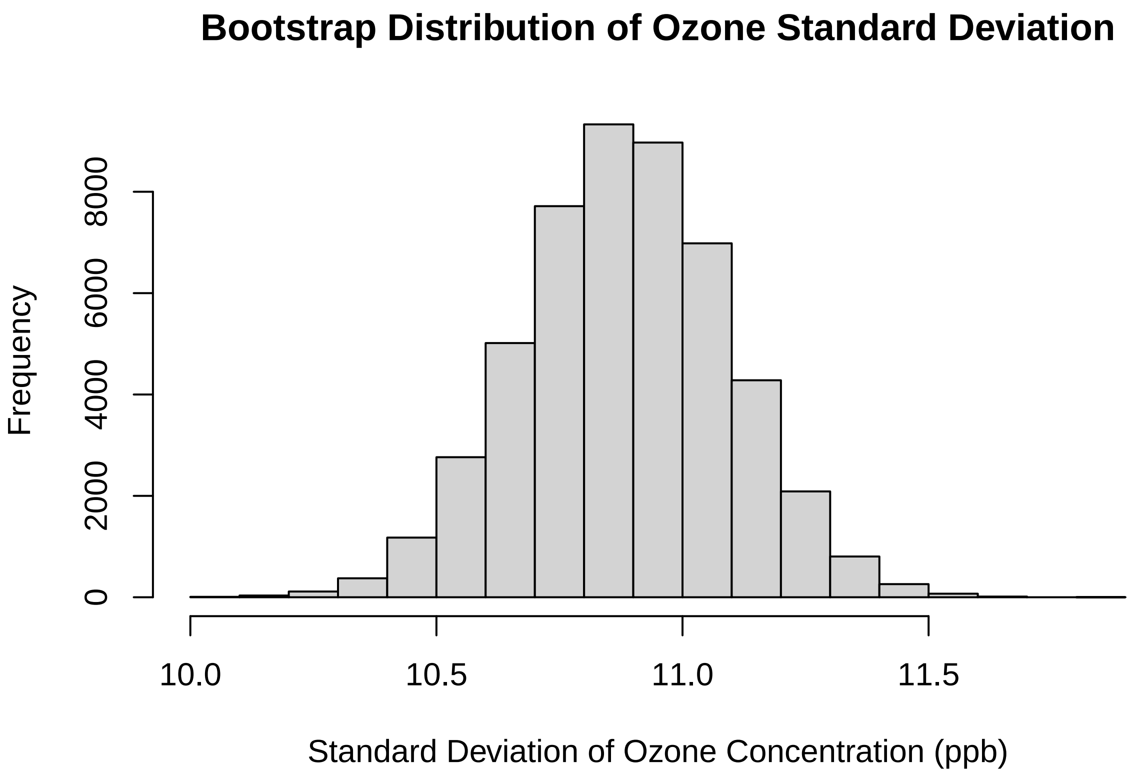
We are 99% confident that the mean of the ozone concentration in the winter is between 18.84 ppb and 21.27 ppb. We are also 99% confident that the standard deviation of the ozone concentration in the winter is between 10.34 ppb and 11.43 ppb.

3b.

```
hist(muhat_storage, main = "Bootstrap Distribution of Ozone Mean", xlab = "Mean Ozone Concentration (ppb
```




```
hist(sd_storage, main = "Bootstrap Distribution of Ozone Standard Deviation", xlab = "Standard Deviatio
```



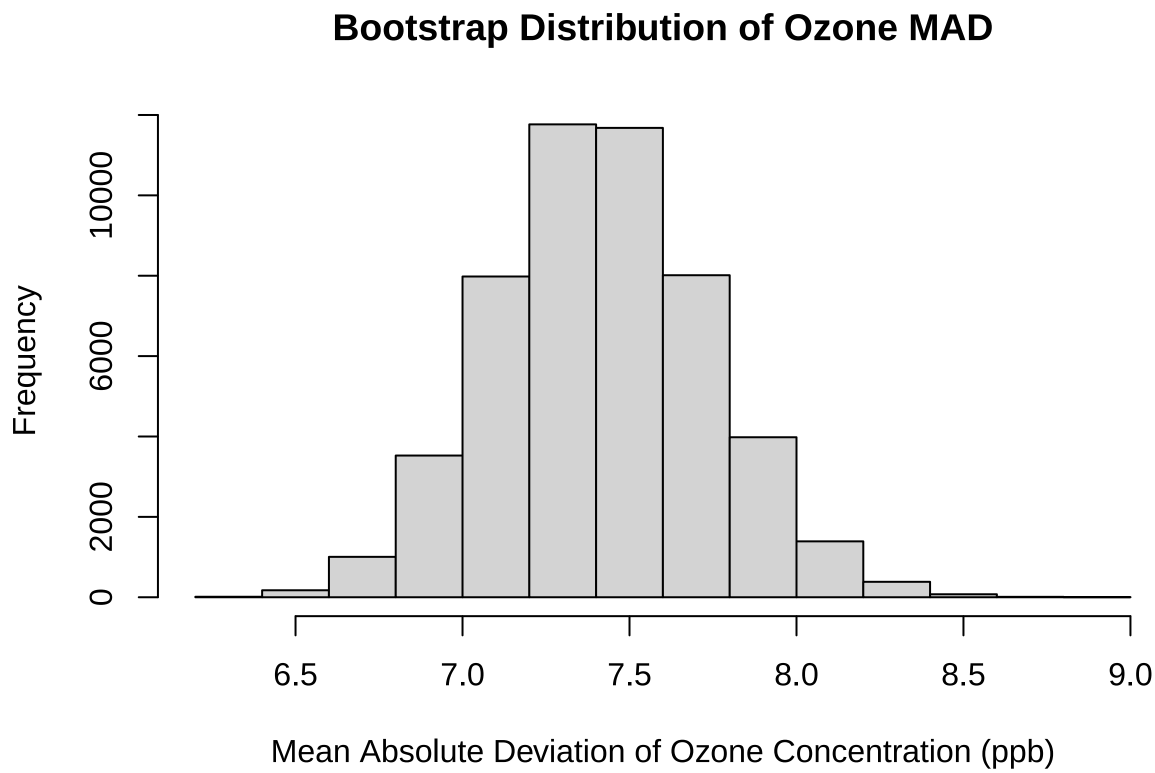
Both histograms appear approximately normal.

3c.

```r
ozone_summer <- summer$O3

B <- 50000

mad_function <- function(x) {
  mean(abs(x - mean(x)))
}

mad_storage <- c()

for (i in 1:B) {
  boot_sample <- sample(ozone_summer, length(ozone_summer), replace = TRUE)
  mad_storage[i] <- mad_function(boot_sample)
}

quantile(mad_storage, c(.005, .995))
```

```
##     0.5%    99.5%
## 6.629219 8.280889
```

```r
hist(mad_storage, main = "Bootstrap Distribution of Ozone MAD", xlab = "Mean Absolute Deviation of Ozone
```



We are 99% confident that the mean absolute deviation of Ozone Concentration is between 6.63 ppb and 8.28 ppb.