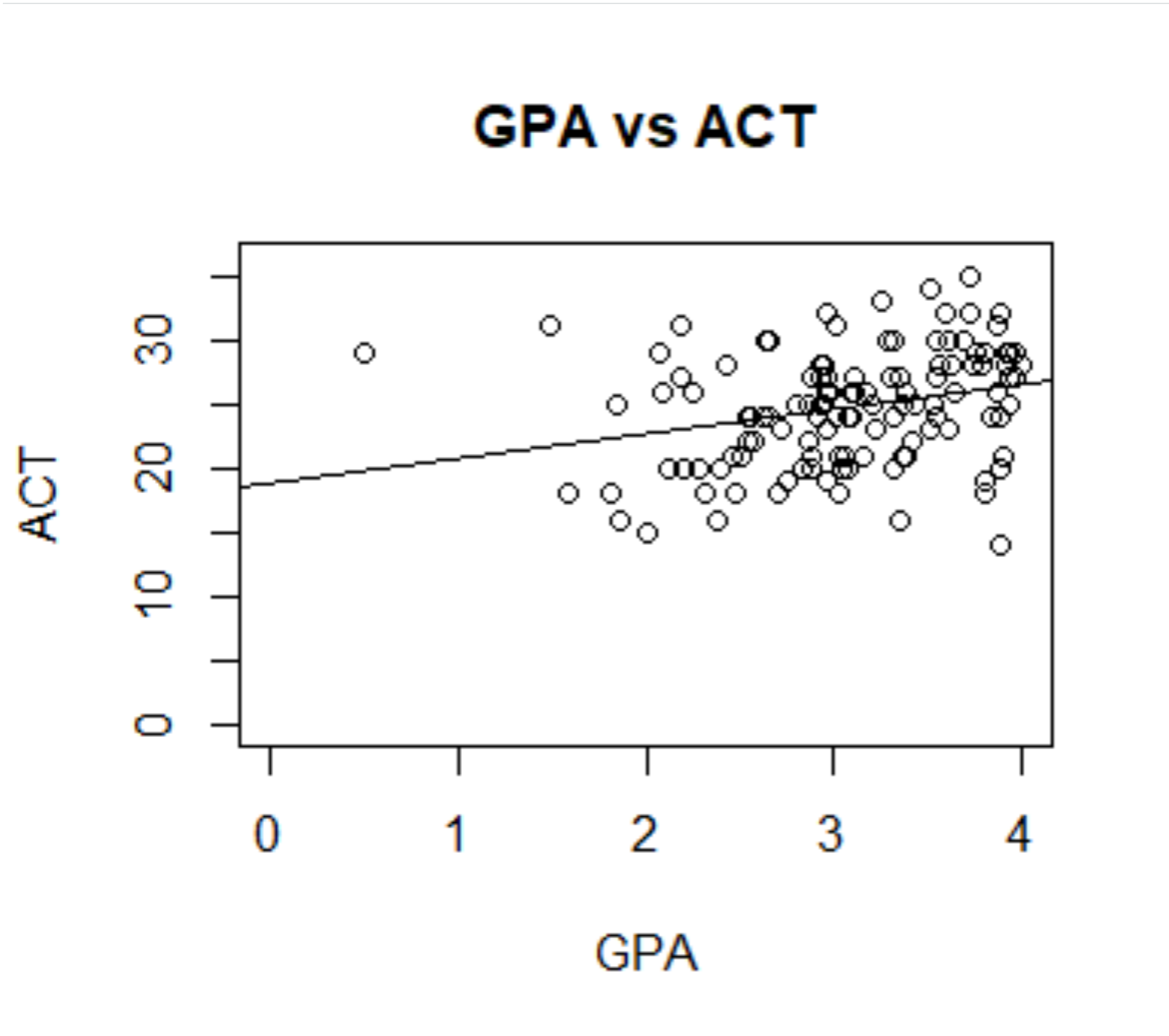
Mini Project #4
David McCormick (DTM190000)
I am in a solo group.

Section 1

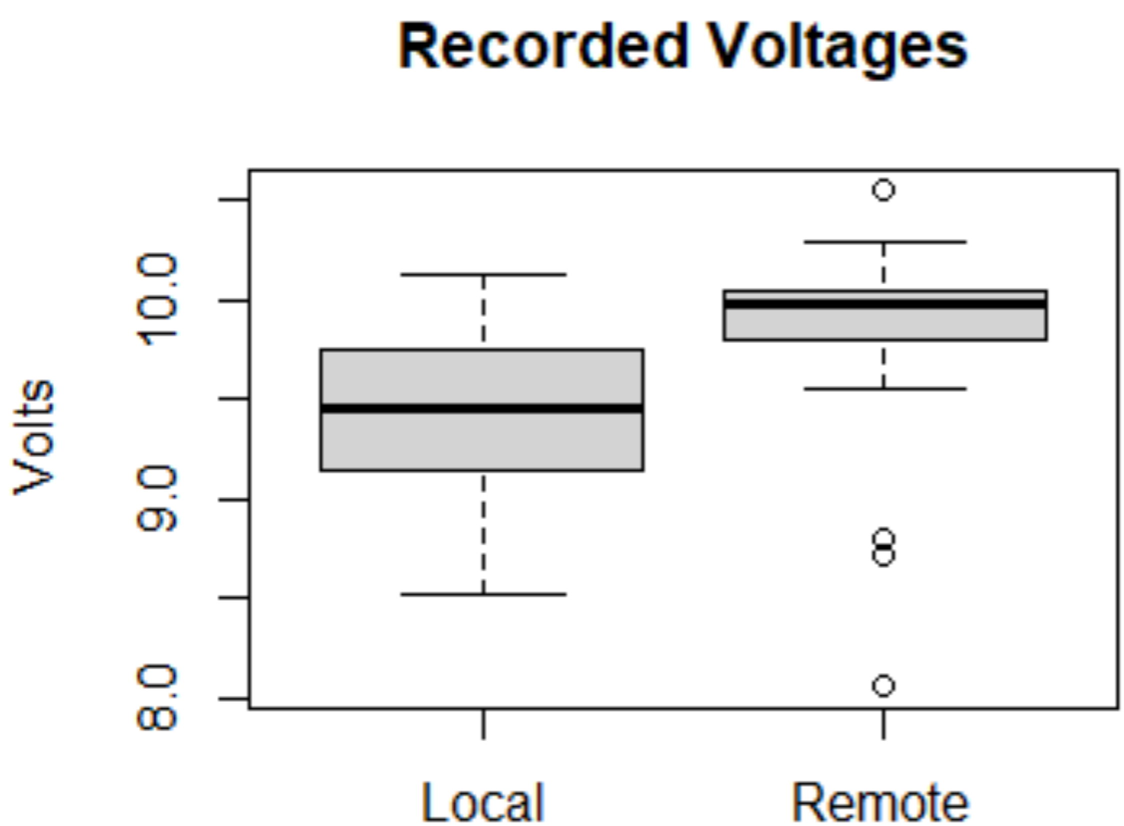
1. Part 1
   a. The graph of GPA vs ACT scores is below:



The slope of the regression line is positive, suggesting a weak positive relationship between GPA and ACT score. The calculated correlation coefficient between the two variables is .2694818.

Using the bootstrap library for R, we can calculate the point estimate, bias, and standard error. The calculated point estimate from the bootstrap samples is .2705715, bias is .001089648, and standard error is .1027537. The bootstrap library calculates a 95% confidence interval of [.0731, .4724], which we can verify by sorting the data and finding the 25th and 975th entries-confirming the confidence interval above. We can conclude that the true value for correlation between GPA and ACT would fall between .0731 and .4724 95% of the time if we took samples of 120 students.

2. Part 2
    a. The distributions do not seem similar; they have notably different variances. The local voltages seem normally distributed with a mean of about 9.4V. The remote voltages seem skewed left, with a mean of 9.8V. The remote voltages also have several outliers.



    b. We can calculate the confidence interval by finding a point estimate using the differences between the means of local and remote voltages. Afterwards, we would add or subtract the standard error times the critical value. We calculated a 95% confidence interval of [-.63985V, -.12282V]. Since the calculated confidence interval does not include 0, there is a difference between the local and remote voltages. We can conclude that the manufacturing process should not be established locally.

We can also use a t-test with a null hypothesis that there is no difference between the recorded local and remote means, and an alternate hypothesis that there is a difference between the recorded local and remote means. We will also use an alpha of .05.

$H_0 = \mu_l - \mu_r = 0$
$H_1 = \mu_l - \mu_r \neq 0$

After we run the t-test, we receive a p-value of .005419 which is less than our alpha of .05, we should reject the null hypothesis. Thus, there is a difference between the recorded local and remote means.

c. I expected that there would be a difference in the recorded local and remote voltages because one look at the boxplots shows that the distributions are quite different. Thus, the manufacturing process should not be established locally.

3. Part 3
    a. We can calculate a confidence interval between the means of the calculated and actual temperatures. Manually calculating the confidence interval entails:

$CI = \mu_c - \mu_a \pm qt(1 - \alpha/2,\ n - 1) * sd/\sqrt{n}$
$CI = \mu_d \pm qt(.975,\ 15) * sd/\sqrt{16}$
$CI = [-.00688,\ .008263]$

Since the 95% confidence interval contains 0, we can conclude that there is no difference between the calculated and actual temperatures.

We can conduct a t-test with a null hypothesis that there is no difference between the calculated and actual means, and an alternate hypothesis that there is a difference between the calculated and actual means. We will also use an alpha of .05

$H_0 = \mu_c - \mu_a = 0$
$H_1 = \mu_c - \mu_a \neq 0$

After we run the t-test, we receive a p-value of ..8492 which is greater than our alpha of .05, we should fail to reject the null hypothesis. Thus, there is no difference between the recorded local and remote means.

Section 2

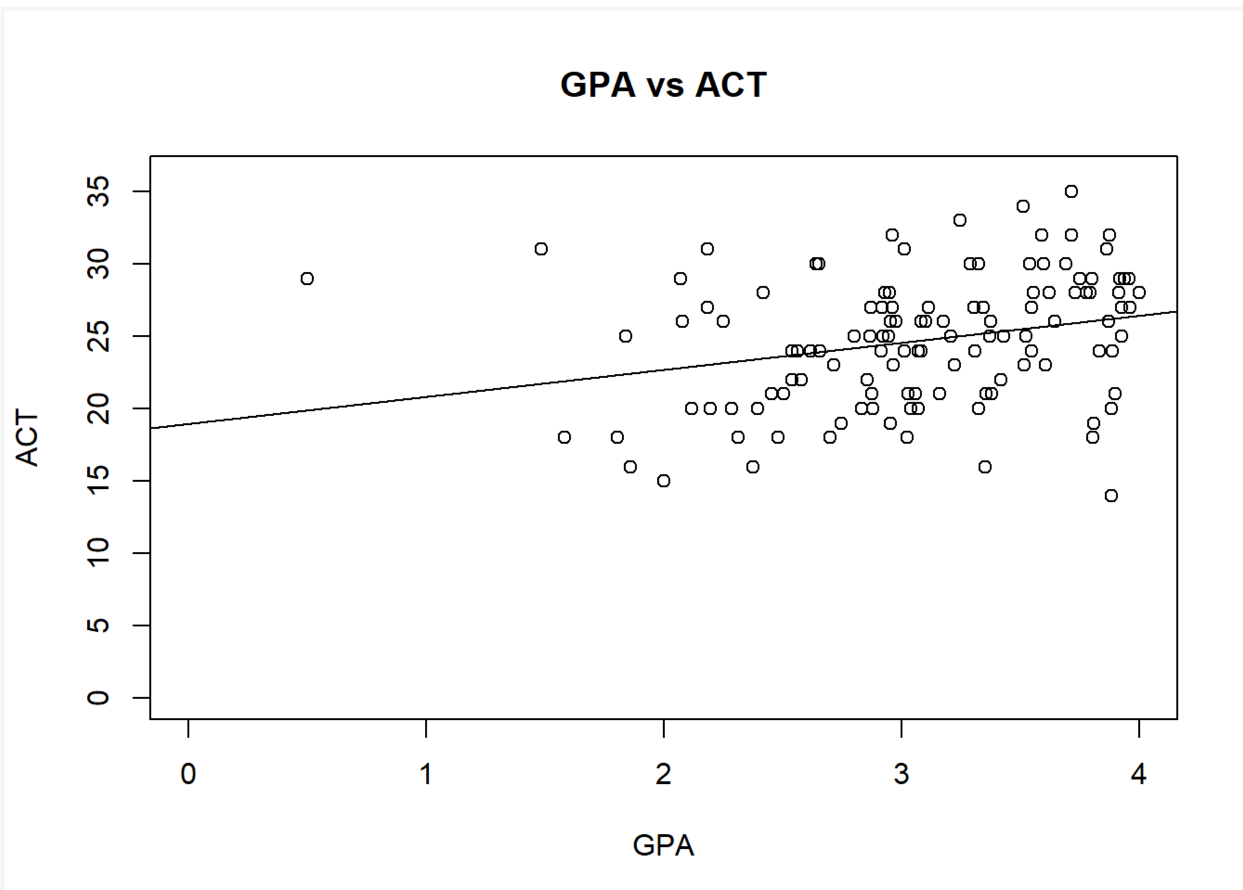
```
#### Question 1 ####

# Import the bootstrapping library
library(boot)

# Read the GPA data
data = read.csv("C:/Users/David/Desktop/gpa.csv")

GPA = data[,c(1)]
ACT = data[,c(2)]
plot(GPA, ACT, xlab="GPA", xlim=c(0,4), ylim=c(0,36), ylab="ACT", main="GPA vs ACT")

# Line of regression
abline(lm(ACT~GPA))
```




```
# Correlation
cor(GPA, ACT)
```

```
## [1] 0.2694818
```

```r
# Gets the correlation between bootstrapped indices
correlation <- function(data, indices) {
  GPA = data[,c(1)]
  ACT = data[,c(2)]
  tGPA = data$gpa[indices]
  tACT = data$act[indices]
  return(cor(tGPA, tACT))
}
```

```r
(stat.boot = boot(data, correlation, R=999, sim="ordinary", stype="i"))
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = data, statistic = correlation, R = 999, sim = "ordinary",
##     stype = "i")
##
##
## Bootstrap Statistics :
##      original      bias    std. error
## t1* 0.2694818 0.001089648   0.1027537
```

```r
mean(stat.boot$t)
```

```
## [1] 0.2705715
```

```r
boot.ci(boot.out = stat.boot, type = "perc")
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 999 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = stat.boot, type = "perc")
##
## Intervals :
## Level     Percentile
## 95%   ( 0.0731,  0.4724 )
## Calculations and Intervals on Original Scale
```

```r
# Manual 95% percentile bootstrap estimate
# alpha/2th and 1-alpha/2th quartiles
sort(stat.boot$t)[c(25,975)]
```

```
## [1] 0.07313051 0.47244272
```

```r
#### Question 2 ####
```

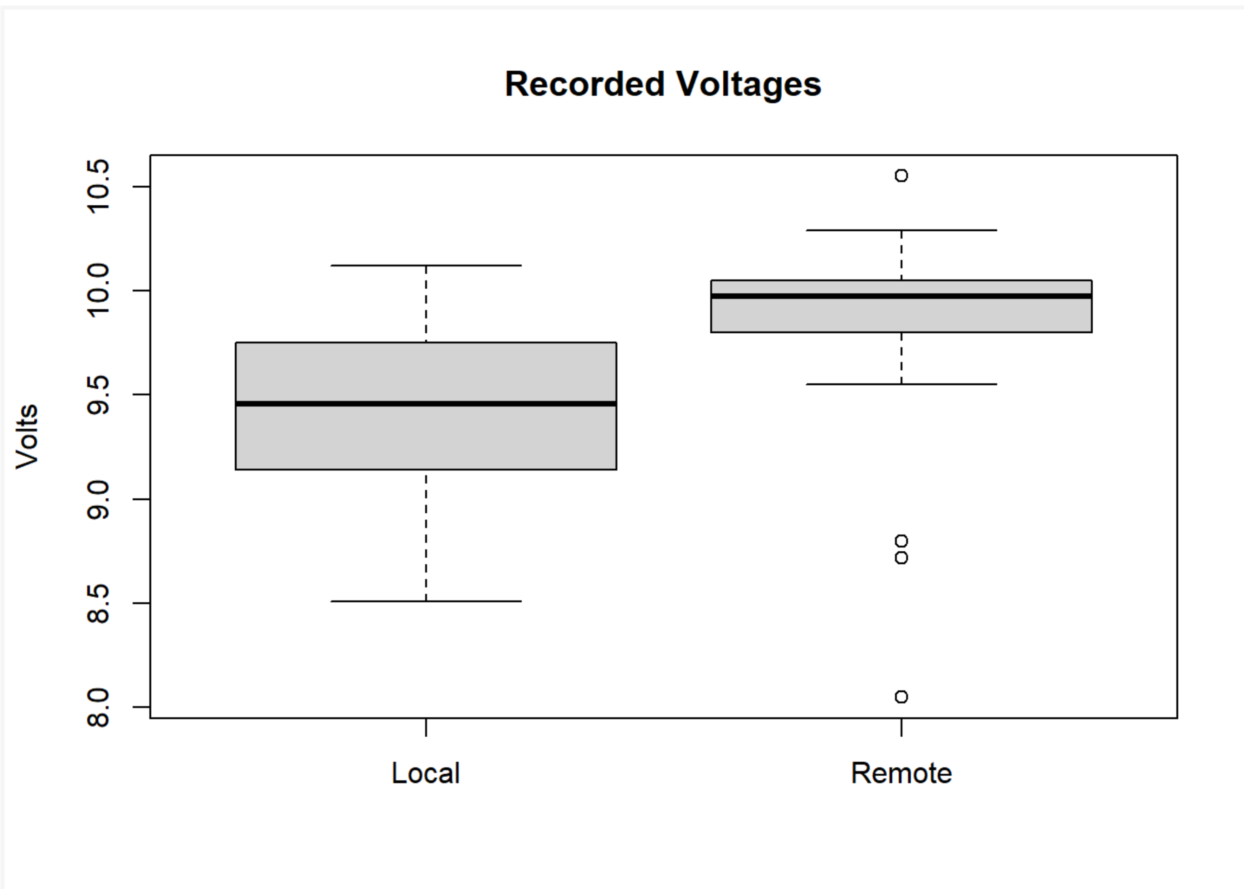
```r
# Import the bootstrapping library
library(boot)

# Read the GPA data
data = read.csv("C:/Users/David/Desktop/VOLTAGE.csv")

remote = data$voltage[which(data$location == 0)]
local = data$voltage[which(data$location == 1)]

# Descriptive stats
summary <- function(col) {
  print(mean(col))
  print(median(col))
  print(quantile(col, prob=c(.25,.75)))
  print(max(col))
  print(min(col))
  print(var(col))
}

boxplot(local, remote, names=c("Local", "Remote"), ylab="Volts", main="Recorded Voltages")
```




```r
# For descriptive stats
```

```
summary(local)

## [1] 9.422333
## [1] 9.455
##    25%    75%
## 9.1525 9.7375
## [1] 10.12
## [1] 8.51
## [1] 0.229322

summary(remote)

## [1] 9.803667
## [1] 9.975
##   25%   75%
##  9.80 10.05
## [1] 10.55
## [1] 8.05
## [1] 0.2925895

# Calculating the Confidence Interval
se = sqrt(var(local)/30 + var(remote)/30)
difference = mean(local) - mean(remote)
upper = difference + qnorm(1-.025) * se
lower = difference - qnorm(1-.025) * se
paste(lower, upper)

## [1] "-0.639848420785972 -0.122818245880696"

t.test(local, remote, alternative="two.sided", paired=FALSE, var.equal=FALSE, conf.level=.95)

##
##  Welch Two Sample t-test
##
## data:  local and remote
## t = -2.8911, df = 57.16, p-value = 0.005419
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6454382 -0.1172284
## sample estimates:
## mean of x mean of y
##  9.422333  9.803667

#### Question 3 ####

# Import the bootstrapping library
library(boot)

# Read the GPA data
data = read.csv("C:/Users/David/Desktop/VAPOR.csv")
```

```r
calculated = data[,c(2)]
actual = data[,c(3)]

#print(var(calculated))
#print(var(actual))

difference = calculated-actual
se = sd(difference)/4
difference = mean(calculated) - mean(actual)
upper = difference + qt(1-.025, 15) * se
lower = difference - qt(1-.025, 15) * se
paste(lower, upper)
```

```
## [1] "-0.00688769375004795 0.00826269375004808"
```

```r
t.test(calculated, actual, alternative="two.sided", paired=TRUE, var.equal=FALSE, conf.level=.95)
```

```
##
##  Paired t-test
##
## data:  calculated and actual
## t = 0.19344, df = 15, p-value = 0.8492
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.006887694  0.008262694
## sample estimates:
## mean of the differences
##               0.0006875
```