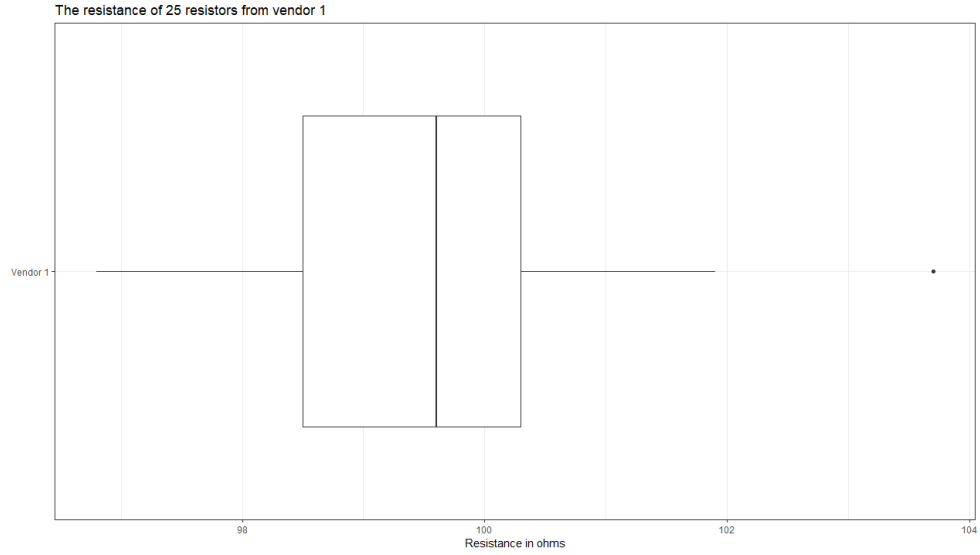
MAB205/MPB311 Statistics Coursework

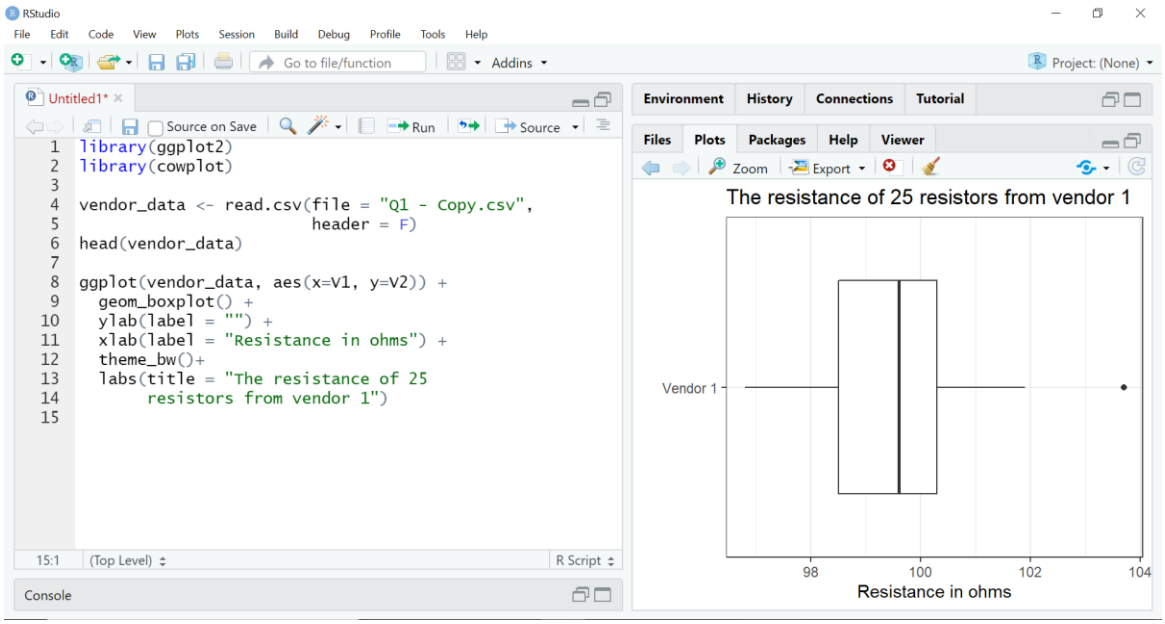
Mohammad Talib Hossain – F028130

Question 1-1

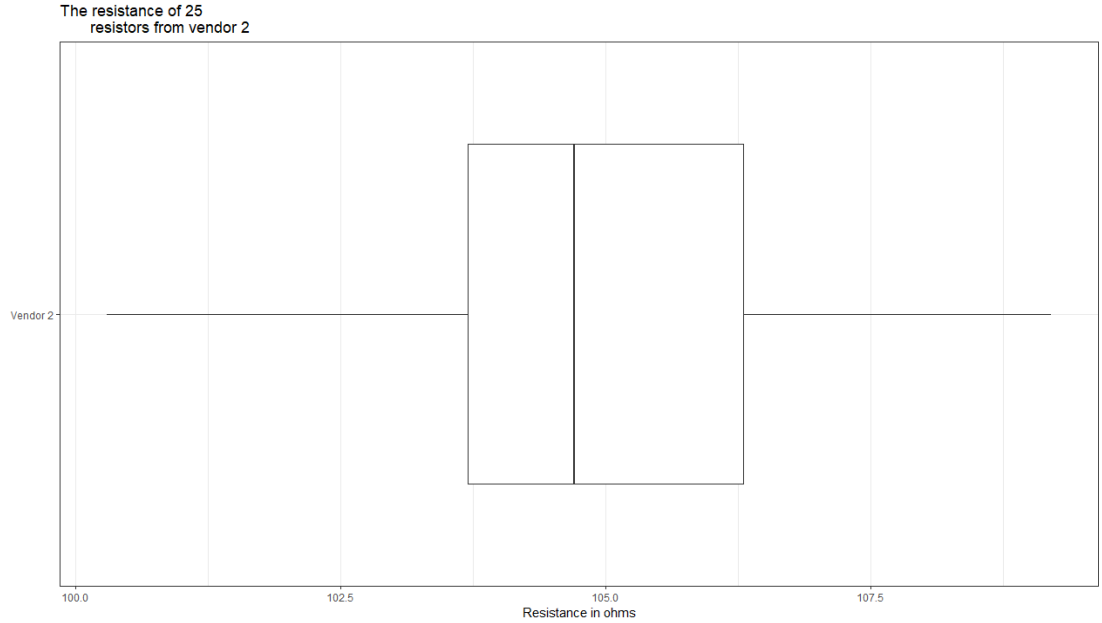
(a) Boxplot diagram

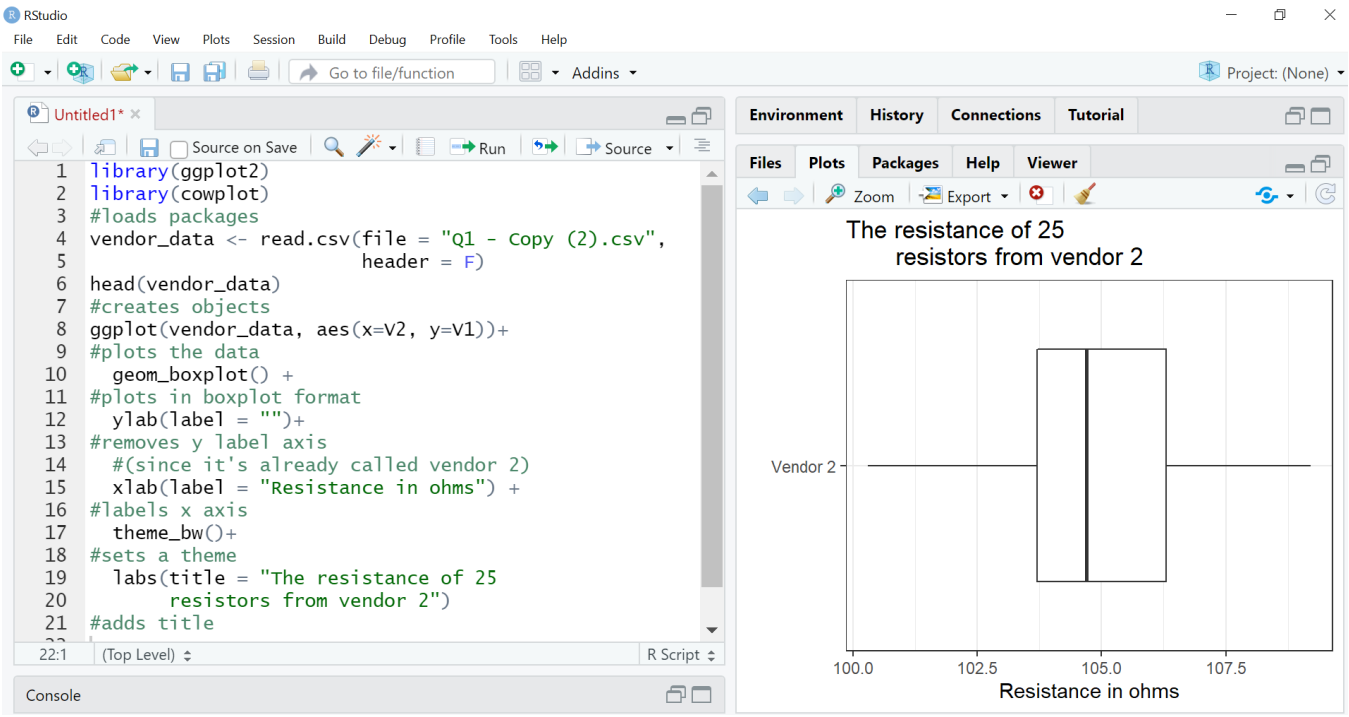
(b) Vendor 1:





Vendor 2:

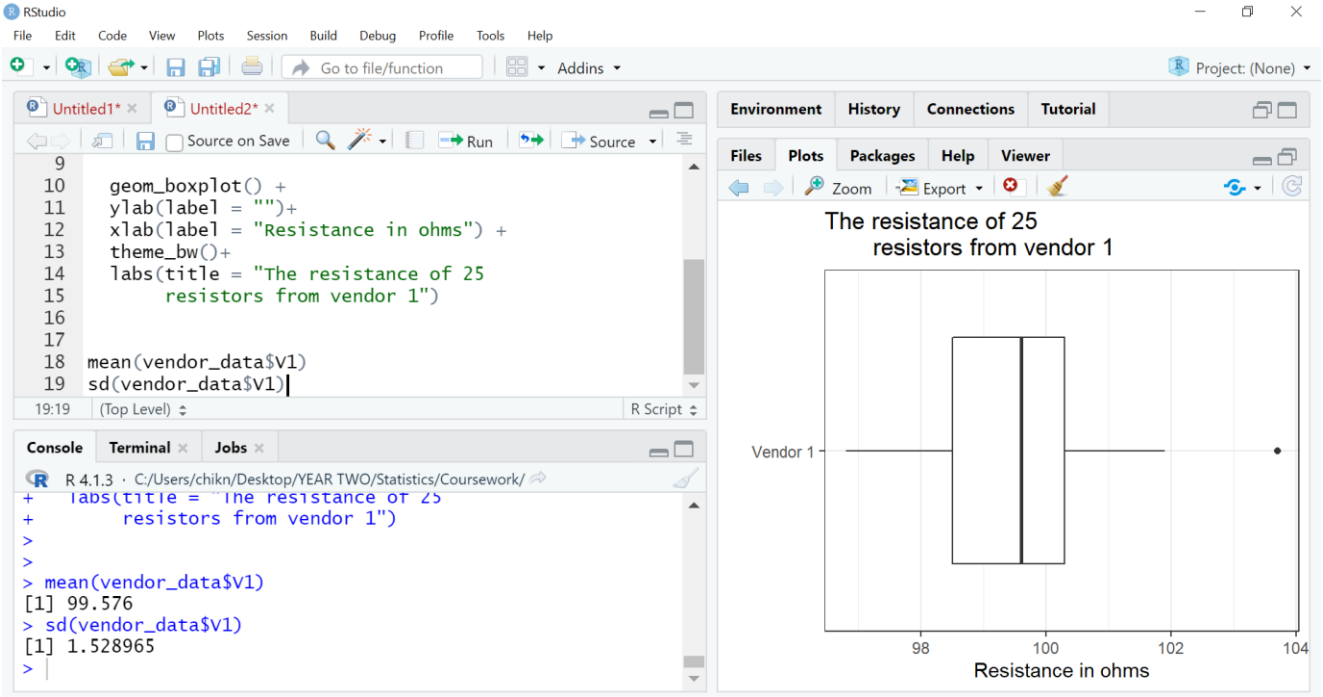

The resistance of 25
resistors from vendor 2
Vendor 2
100.0
102.5
105.0
107.5
Resistance in ohms


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Untitled1*
Source on Save
Run
Source
1 library(ggplot2)
2 library(cowplot)
3 #loads packages
4 vendor_data <- read.csv(file = "Q1 - Copy (2).csv",
5 header = F)
6 head(vendor_data)
7 #creates objects
8 ggplot(vendor_data, aes(x=V2, y=V1))+
9 #plots the data
10 geom_boxplot() +
11 #plots in boxplot format
12 ylab(label = "")+
13 #removes y label axis
14 #(since it's already called vendor 2)
15 xlab(label = "Resistance in ohms") +
16 #labels x axis
17 theme_bw()+
18 #sets a theme
19 labs(title = "The resistance of 25
20 resistors from vendor 2")
21 #adds title
22:1
(Top Level)
R Script
Console
Environment
History
Connections
Tutorial
Files
Plots
Packages
Help
Viewer
Zoom
Export
The resistance of 25
resistors from vendor 2
Vendor 2
100.0
102.5
105.0
107.5
Resistance in ohms

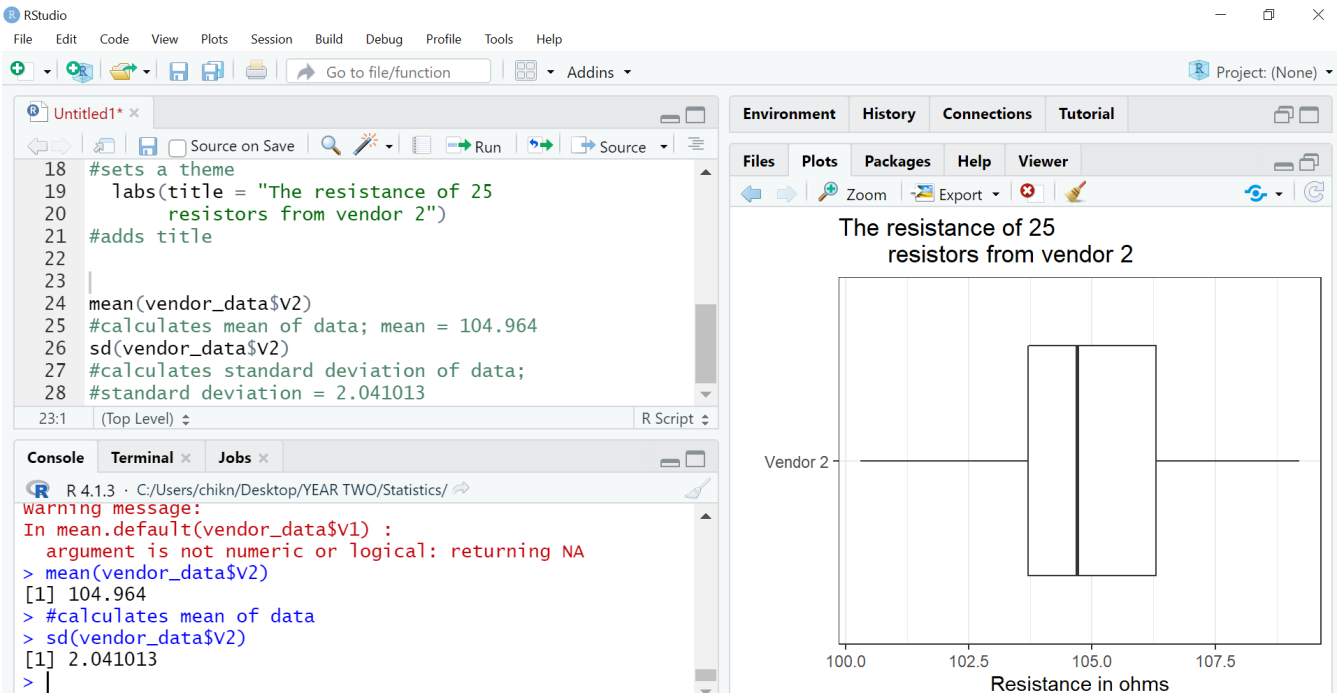
Question 1-2

(a) A measure of central tendency is mean. A measure of spread is standard deviation.

(b) Vendor 1: Mean = 99.576; Standard deviation = 1.528965



Vendor 2: Mean = 104.964; Standard deviation = 2.041013



Question 1-3

(a) Null hypothesis: The resistance of resistors from both vendors will be the same.

Alternative hypothesis: The resistance of resistors from both vendors will be different.

(b) Independent samples t-test. The resistances of resistors from both vendors are independent of each other, hence an independent samples t-test is ideal.

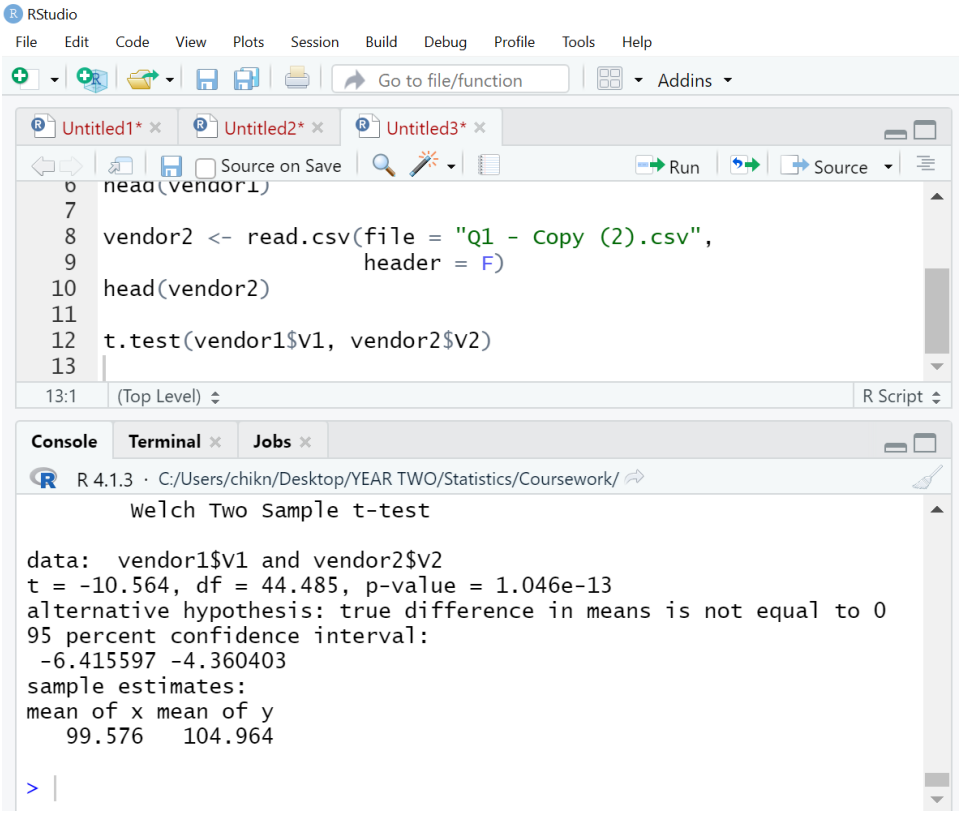
Question 1-4

(a) Data for each group is normally distributed.

(b) The assumption can be checked by doing the Shapiro-Wilk test for each group. If the p-value is greater than 0.05, the data is normally distributed.

Question 1-5

(a)



t = -10.564, df = 44.485, p-value = 1.046e-13

Since p-value is very small, the null hypothesis can be rejected.

(b) There is statistically significant evidence showing resistors of both vendors are not equal.

Question 2-1

(a) ANOVA

(b) Homogeneity of variance; variance among the groups should approximately equal. It can be checked by using Levene's test.

Question 2-2

The means of observations grouped by current are not the same.

The means of observations grouped by exposure time are not the same.

There are interactions observed between current and exposure time.

Question 2-3

(a)

RStudio

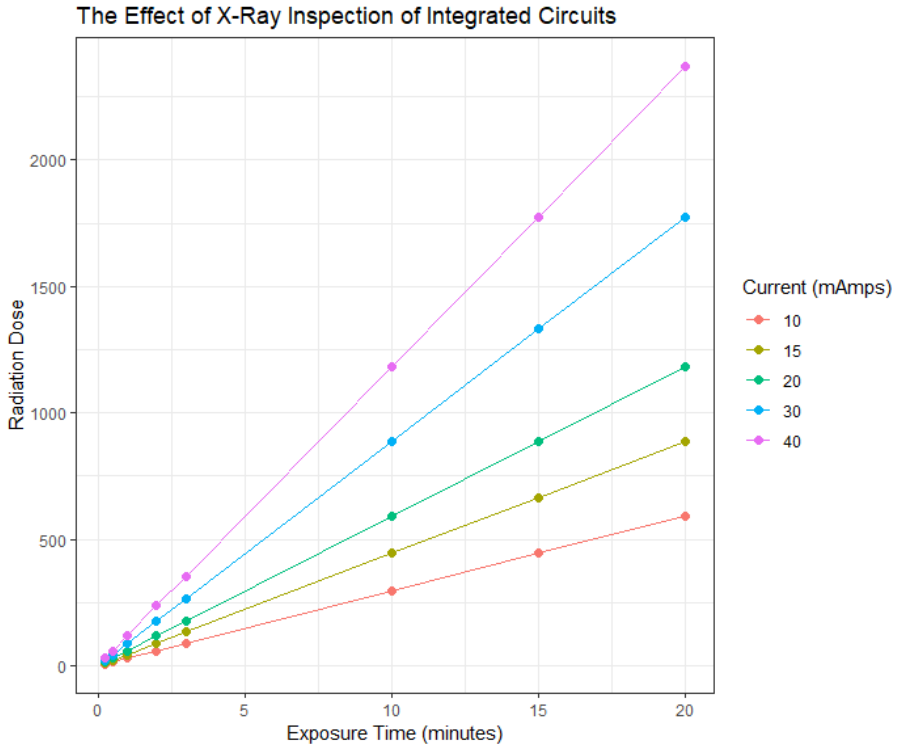
File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

q2.R | q1-1v2.R | Untitled1* | Untitled2*

Source on Save | Run | Source

```
1  library(ggplot2)
2  library(cowplot)
3
4  q2_data <- read.csv(file = "Q2.csv")
5  head(q2_data)
6
7  ggplot(data=q2_data,
8         aes(x=Exposure, y=Radiation, group = Current, colour = as.factor(Current))) +
9    geom_line() + geom_point(size = 2) +
10   xlab("Exposure Time (minutes)") + ylab("Radiation Dose") +
11   labs(colour = "Current (mAmps)") +
12   theme_bw() +
13   labs(title = "The Effect of X-Ray Inspection of Integrated Circuits")
14
15 summary(aov(Radiation ~ Current * Exposure, data = q2_data))
16
```




```
> summary(aov(Radiation ~ Current * Exposure, data = q2_data))
                 Df  Sum Sq Mean Sq   F value Pr(>F)
Current           1 1701148 1701148 3.384e+31 <2e-16 ***
Exposure          1 9375325 9375325 1.865e+32 <2e-16 ***
Current:Exposure  1 2055837 2055837 4.089e+31 <2e-16 ***
Residuals        36       0       0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

(b) The current had a main effect on the radiation, $F(1,36) = 3.384e+31$, $p<2e-16$

The exposure time also had a main effect on the radiation, $F(1,36) = 1.865e+32$, $p<2e-16$

There was a statistically significant interaction between the effects of current and exposure time on radiation, $F(1,36)=4.089e+31$, $p<2e-16$

As the p-values are very low, the alternative hypotheses are accepted.

Question 2-4

The residuals sum of squares equals zero, indicating that the model is a perfect fit.