Test 2

2023-05-21

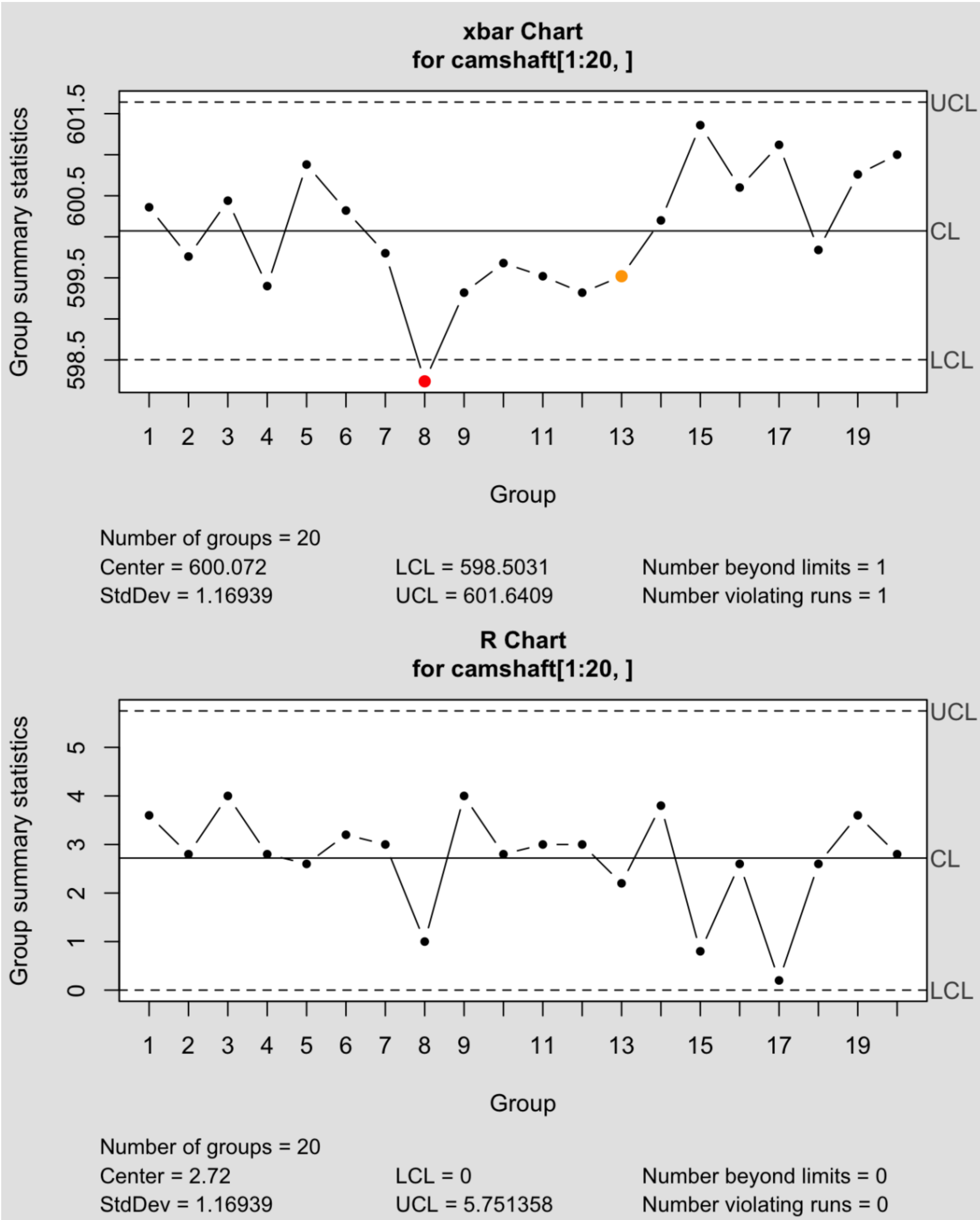
```
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✔ dplyr     1.1.0     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ──────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force
all conflicts to become errors

library(readxl)
library(qcc)

## Package 'qcc' version 3.0
## Type 'citation("qcc")' for citing this R package in publications.

library(matrixStats)

##
## Attaching package: 'matrixStats'
##
## The following object is masked from 'package:dplyr':
##
##     count

qccGroups <- function(data, x, sample)
{
  # collect x and sample from data if provided
  if(!missing(data))
  {
    x      <- eval(substitute(x), data, parent.frame())
    sample <- eval(substitute(sample), data, parent.frame())
  }
  stopifnot(length(x) == length(sample))
  #
  x <- lapply(split(x, sample), as.vector)
  lx <- sapply(x, length)
  for(i in which(lx != max(lx)))
```

```
      x[[i]] <- c(x[[i]], rep(NA, max(lx)-lx[i]))
  x <- t(sapply(x, as.vector))
  return(x)
}
```

1. For the camshaft data from lectures, produce an X-bar and a Range chart. Comment on whether the process is in statistical control.

```
# load your data, replace "camshaft.xlsx" with your actual file path
cam <- read_excel("camshaft.xlsx")

camshaft <- qccGroups(data = cam, Length, sample)
head(camshaft)

##    [,1]  [,2]  [,3]  [,4]  [,5]
## 1 601.4 601.6 598.0 601.4 599.4
## 2 600.0 600.2 601.2 598.4 599.0
## 3 601.2 601.0 600.8 597.6 601.6
## 4 599.4 601.2 598.4 599.2 598.8
## 5 601.4 599.0 601.0 601.6 601.4
## 6 601.4 598.8 601.4 598.4 601.6

(obj <- qcc(camshaft[1:20,], type="xbar"))

## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = xbar
## Data (phase I)             = camshaft[1:20, ]
## Number of groups           = 20
## Group sample size          = 5
## Center of group statistics = 600.072
## Standard deviation         = 1.16939
##
## Control limits at nsigmas  = 3
##        LCL      UCL
##   598.5031 601.6409

obj <- qcc(camshaft[1:20,], type="R")
(obj)

## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = R
## Data (phase I)             = camshaft[1:20, ]
## Number of groups           = 20
## Group sample size          = 5
## Center of group statistics = 2.72
## Standard deviation         = 1.16939
##
## Control limits at nsigmas  = 3
```

```
##   LCL       UCL
##     0 5.751358
```



Within the xbar chart there appears to be a value outside the lower limit for the 8th group and an orange value within the limits at the 13th value. The value outside the control limit suggests that the process may be out of control and is called special cause variation. The orange point within the control limits is a warning, existing when for example 2 out of 3 consecutive points falling beyond 2 standard deviations from the center line or 4 out of 5 points falling beyond 1 standard deviation from the center line. It does not directly imply the process is out of control, but it often indicates a potential issue.

The range chart appears to be within control limits. This shows a consistent proves variability for each group.

Since the x-bar chart is out of control it suggests the process is not in statistical control.

A high voltage power supply unit has a nominal output voltage of 350 V. A random sample of four units is taken each day for 20 days and tested for quality control. The data is in the file Voltage.xlsx. Recorded are ten times the difference between the observed reading on each unit and the nominal voltage, that is, xi = 10(Observed voltage on unit i − 350)

2a) Obtain X-bar and R charts for this process and decide if the process is in statistical control.

```
voltage <- read_excel("Voltage.xlsx", skip = 1)
# reshape the data to a suitable format
# Reshape the data
voltage_long <- voltage %>%
  gather(key = "measurement", value = "x", x1, x2, x3, x4)

# Check the structure of the reshaped data
str(voltage_long)

## tibble [80 × 3] (S3: tbl_df/tbl/data.frame)
##  $ Number     : num [1:80] 1 2 3 4 5 6 7 8 9 10 ...
##  $ measurement: chr [1:80] "x1" "x1" "x1" "x1" ...
##  $ x          : num [1:80] 6 10 7 8 9 12 16 7 9 15 ...

# Now, you should be able to create the control chart
voltage_data <- qccGroups(data = voltage_long, x = x, sample = Number)



# X-bar chart
xbar_obj <- qcc(voltage_data, type="xbar")

# R chart
range_obj <- qcc(voltage_data, type="R")
(xbar_obj)

## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = xbar
## Data (phase I)             = voltage_data
```

```
## Number of groups           = 20
## Group sample size          = 4
## Center of group statistics = 10.325
## Standard deviation         = 3.035454
##
## Control limits at nsigmas  = 3
##       LCL      UCL
##  5.771819 14.87818
```

```
(range_obj)
```

```
## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = R
## Data (phase I)             = voltage_data
## Number of groups           = 20
## Group sample size          = 4
## Center of group statistics = 6.25
## Standard deviation         = 3.035454
##
## Control limits at nsigmas  = 3
##  LCL      UCL
##    0 14.26188
```

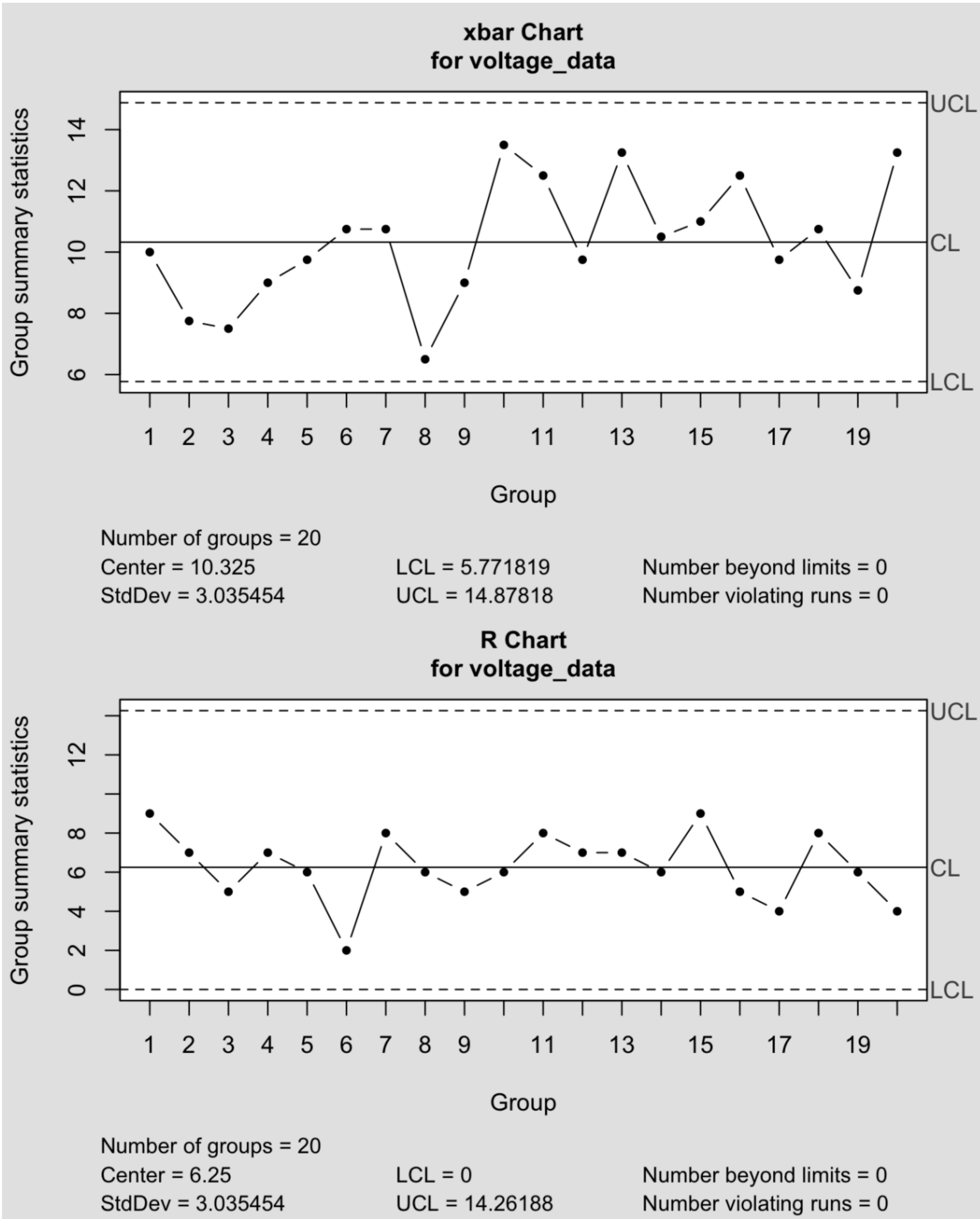
```
summary(xbar_obj)
```

```
## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = xbar
## Data (phase I)             = voltage_data
## Number of groups           = 20
## Group sample size          = 4
## Center of group statistics = 10.325
## Standard deviation         = 3.035454
##
## Control limits at nsigmas  = 3
##       LCL      UCL
##  5.771819 14.87818
```

```
summary(range_obj)
```

```
## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = R
## Data (phase I)             = voltage_data
## Number of groups           = 20
## Group sample size          = 4
## Center of group statistics = 6.25
## Standard deviation         = 3.035454
##
## Control limits at nsigmas  = 3
```

```
##  LCL      UCL
##    0 14.26188
```

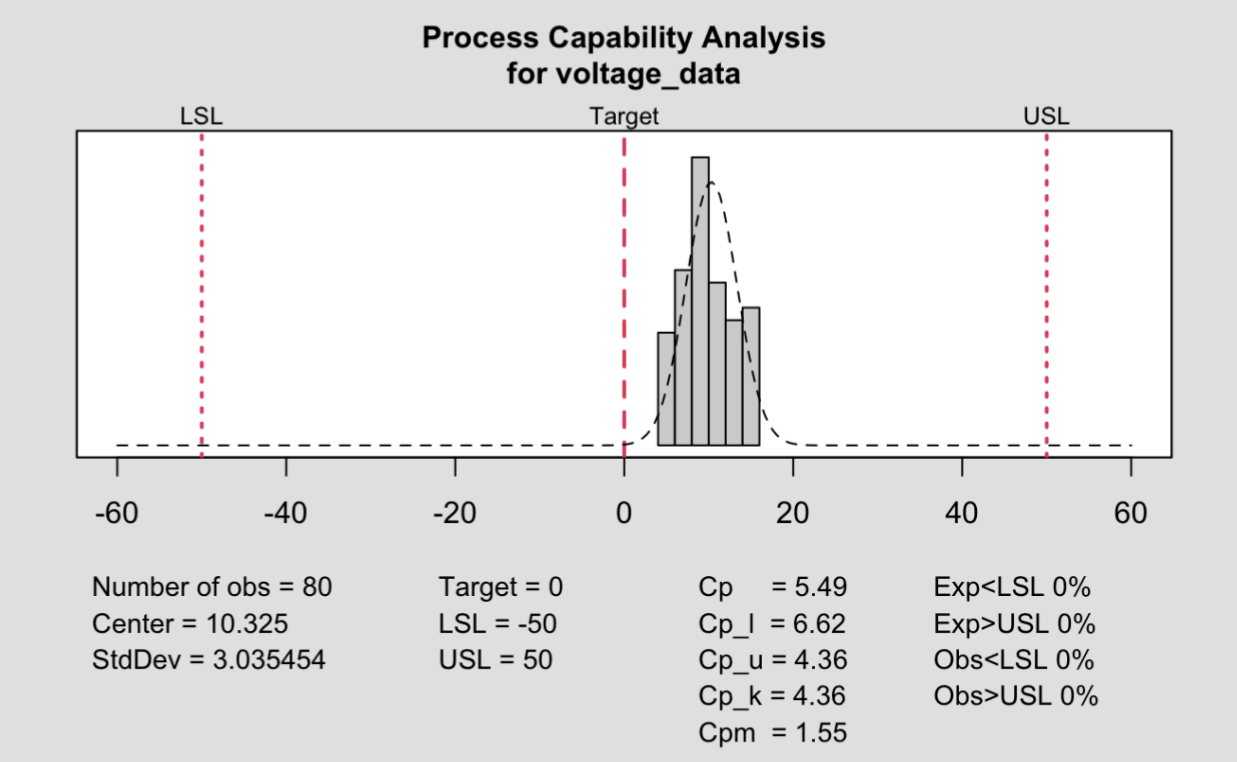


Both the R chart and xbar chart have no out of control groups and therefore imply the process is in control.

2b) If the specifications are 350 ± 5 V, what can you conclude regarding the process capability?

2b

```
# Define specification limits
USL <- 50  # upper spec limit
LSL <- -50  # lower spec limit

# Compute process capability
pc <- processCapability(xbar_obj, spec.limits = c(LSL, USL))

# Print the process capability
(pc)

## ── Process Capability Analysis ───────────────────
##
## Number of obs = 80           Target = 0
## Center        = 10.325       LSL    = -50
## StdDev        = 3.035454     USL    = 50
##
## Capability indices  Value  2.5%  97.5%
##                Cp     5.49  4.64   6.34
##                Cp_l   6.62  5.76   7.49
##                Cp_u   4.36  3.78   4.93
##                Cp_k   4.36  3.67   5.04
##                Cpm    1.55  1.22   1.88
##
## Exp<LSL 0%    Obs<LSL 0%
## Exp>USL 0%    Obs>USL 0%
```



Cp: This measures the potential capability in the process. In this case the value is 5.49, typically a Cp geater than 1 is considered good implying the model has good process capability.

Cp_I and Cp_u: These are the lower and upper process capability indices, respectively. In this case Cp_I is 6.62 and Cp_u is 4.36are both far above 1 which implies good process capability.

Cpk: The minimum process capability, is the minimum of Cp_I and Cp_u. It takes into account the centering of the process distribution within the specification limits. In this scase, Cpk is 4.36 which is quite high indiciating the process is well within the specification limits.

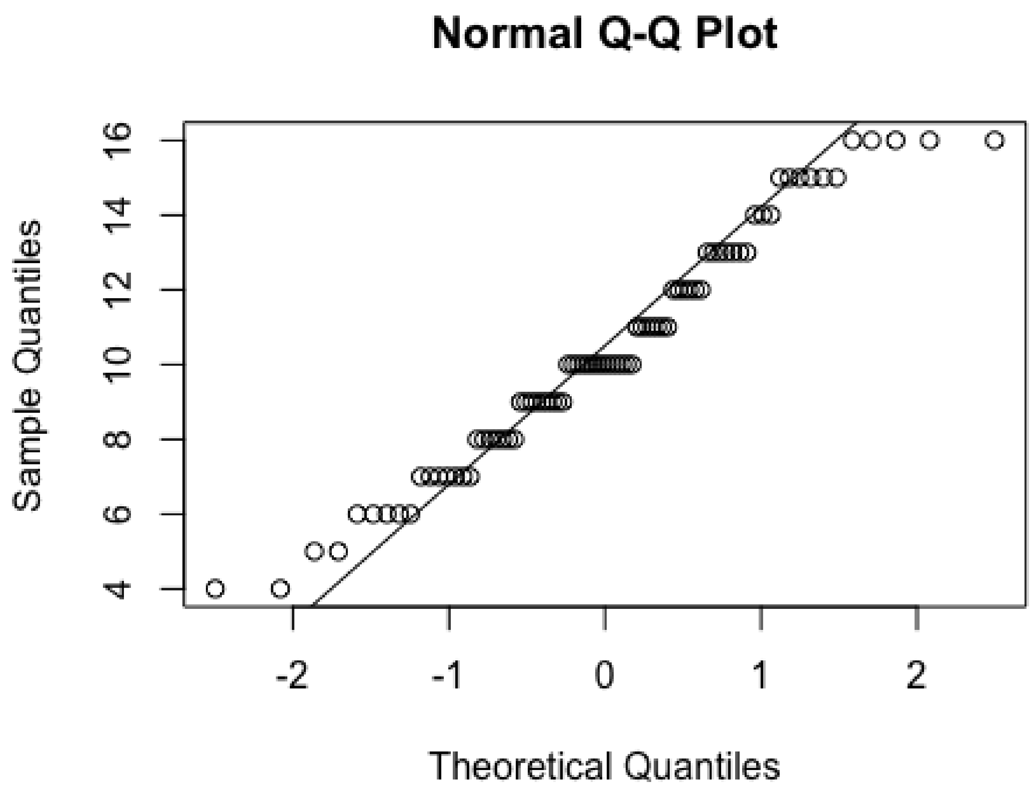
Cpm: This index is similar to Cp but also takes into account the process target. The Cpm is 1.55 which is good but not as high as Cp and Cpk and Cpk indicating the process mean is not perfectly centered.

2c Is there evidence to support the claim that the voltage is normally distributed?

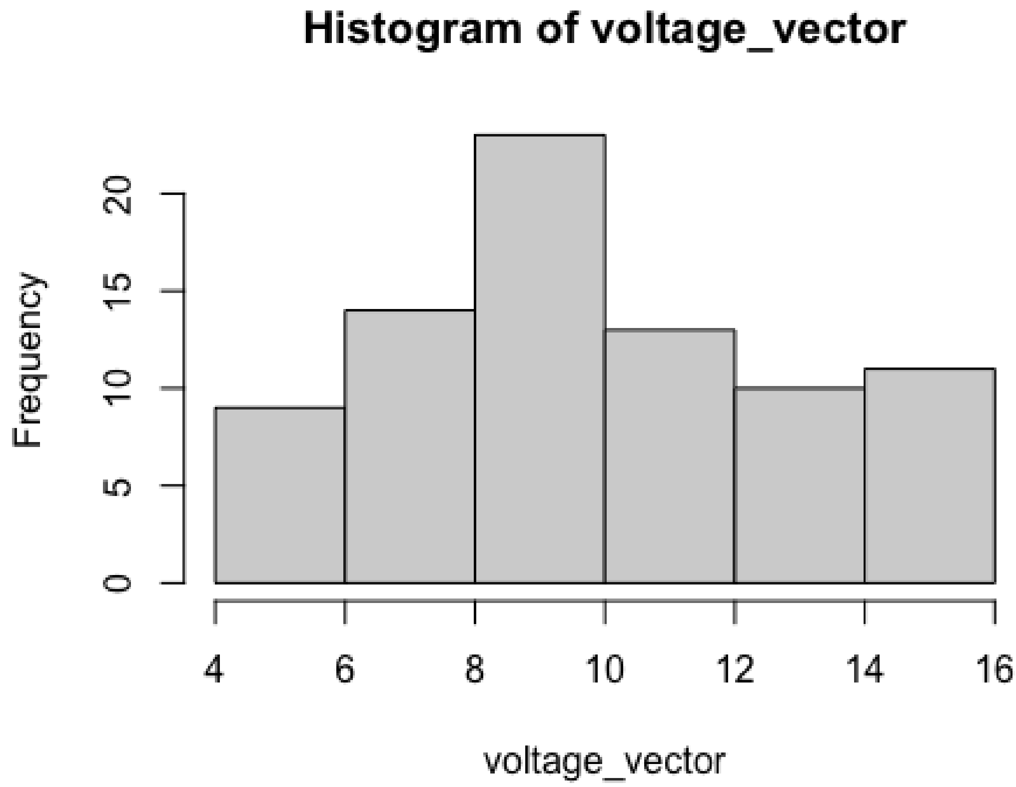
```
# flatten the data into a single vector for the normality test
voltage_vector <- as.vector(t(voltage_data))
# Shapiro-Wilk test
shapiro_test <- shapiro.test(voltage_vector)
print(shapiro_test)

##
##  Shapiro-Wilk normality test
```

```
##
## data:  voltage_vector
## W = 0.96889, p-value = 0.04826
```

```r
# Q-Q plot
qqnorm(voltage_vector)
qqline(voltage_vector)
```




```r
hist(voltage_vector)
```



Based on the results of the Shapiro-Wilk test the value, 0.0483 is less than 0.05 indicating the alternative hypothesis is true and the data is not normally distributed. It is important to note the p-value is very close to the threshold meaning the evidence against the null hypothesis, that it is normally distributed, is not very strong.

The lines on the Q-Q plot follow mostly a straight line but there are deviations at every sample quantile value once again implying the data is close to normal but not perfectly fitting a normal distribution.

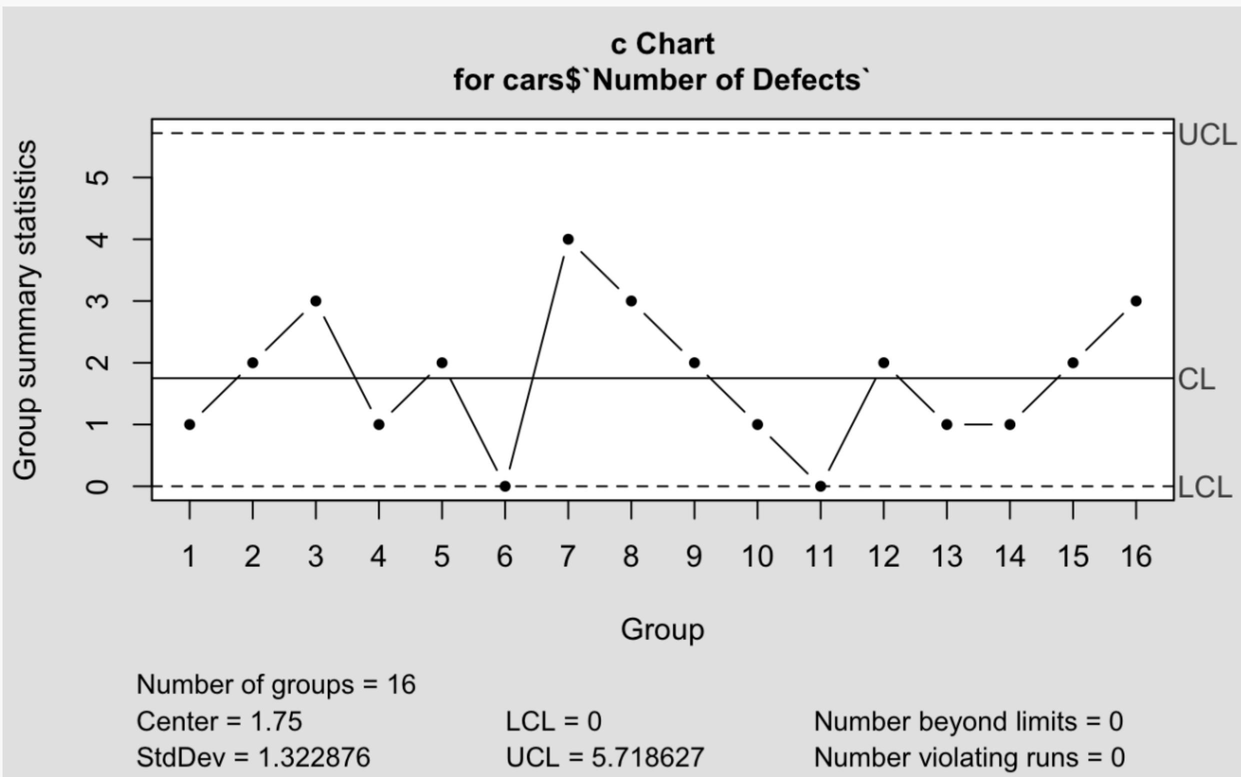
Finally the histogram loosely follows the bell curve shape of a normal distribution however the tails hold too much frequency causing them to slightly deviate from a normal distribution.

Overall, there isn't strong enough evidence to support the claim the voltage is normally distributed.

3 The number of defects in a sub-assembly producing manual transmissions for cars is recorded in the file Cars. Produces appropriate charts for the data and determine if the process is in statistical control.

```
cars <- read_excel("Cars.xlsx")

cchart_obj <- qcc(cars$`Number of Defects`, type = "c")
(cchart_obj)
```

```
## ── Quality Control Chart ─────────────────────────
##
## Chart type                  = c
## Data (phase I)              = cars$`Number of Defects`
## Number of groups            = 16
## Group sample size           = 1
## Center of group statistics = 1.75
## Standard deviation          = 1.322876
##
## Control limits at nsigmas   = 3
##  LCL      UCL
##    0 5.718627
```



The C-chart is appropriate for monitoring the number of defects in a process where sample size remains constant over time. Based on the results two values fall on the lower limit but do not cross it. Since these points are not outside the limits they do not violate any of the traditional control chart rules. Therefore the process appears to be in control.

4A

```
# Load the data
parts <- read_excel("Parts.xlsx", skip = 2)
```

```r
parts_long <- parts %>%
  gather(key = "measurement", value = "x", "1", "2", "3", "4","5")

# Check the structure of the reshaped data
str(parts_long)
```

```
## tibble [100 × 3] (S3: tbl_df/tbl/data.frame)
##  $ Sample     : num [1:100] 1 2 3 4 5 6 7 8 9 10 ...
##  $ measurement: chr [1:100] "1" "1" "1" "1" ...
##  $ x          : num [1:100] 83 88.6 85.7 80.8 83.4 75.3 74.5 79.2 80.5
75.7 ...
```

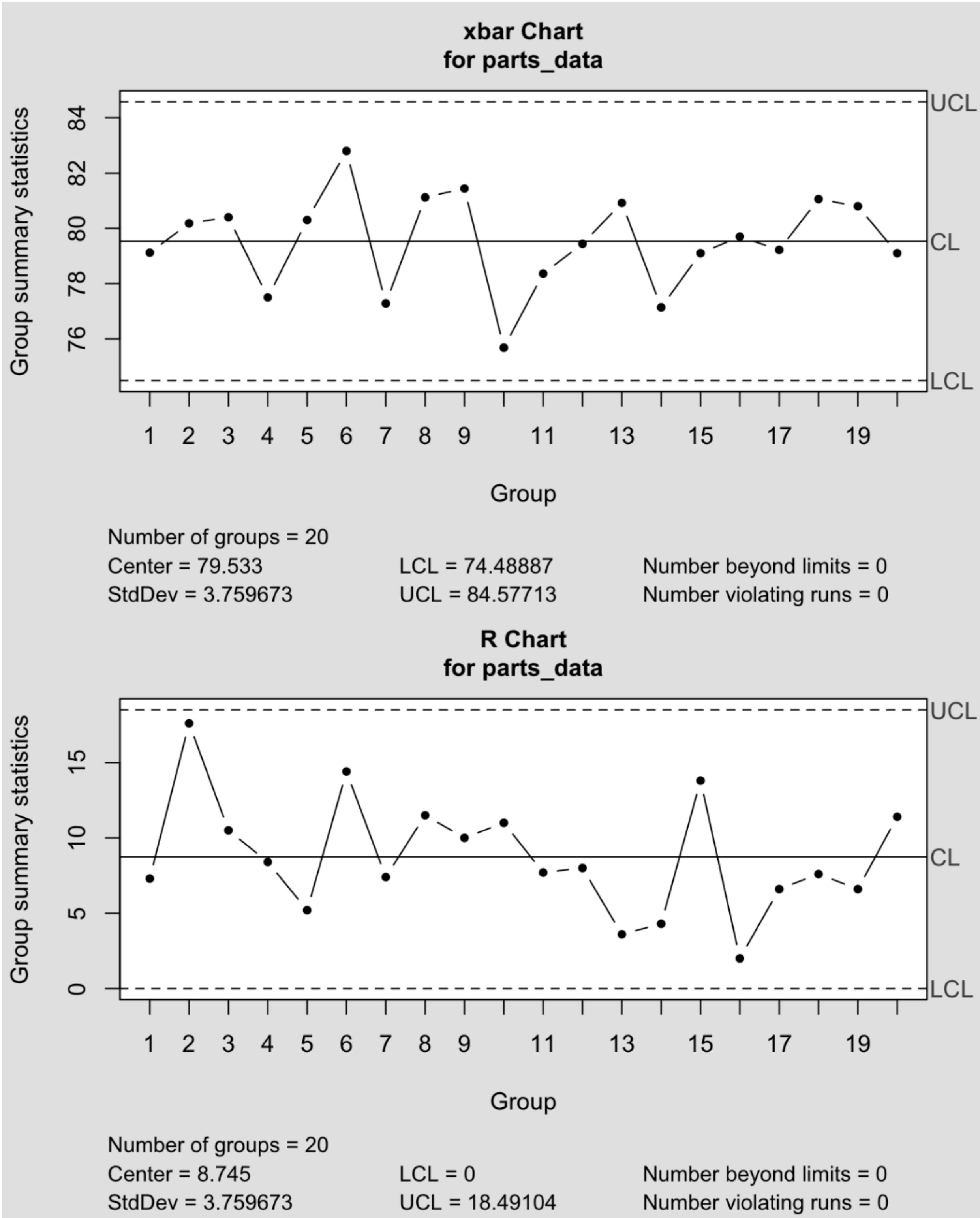
```r
parts_data <- qccGroups(data = parts_long, x = x, sample = Sample)

# Calculate and plot the X-bar and R control charts
xbar_obj <- qcc(parts_data, type = "xbar")
range_obj <- qcc(parts_data, type = "R")

# Check if the process is in statistical control
summary(xbar_obj)
```

```
## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = xbar
## Data (phase I)             = parts_data
## Number of groups           = 20
## Group sample size          = 5
## Center of group statistics = 79.533
## Standard deviation         = 3.759673
##
## Control limits at nsigmas  = 3
##       LCL      UCL
##  74.48887 84.57713
```

```r
summary(range_obj)
```

```
## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = R
## Data (phase I)             = parts_data
## Number of groups           = 20
## Group sample size          = 5
## Center of group statistics = 8.745
## Standard deviation         = 3.759673
##
## Control limits at nsigmas  = 3
##  LCL      UCL
##    0 18.49104
```

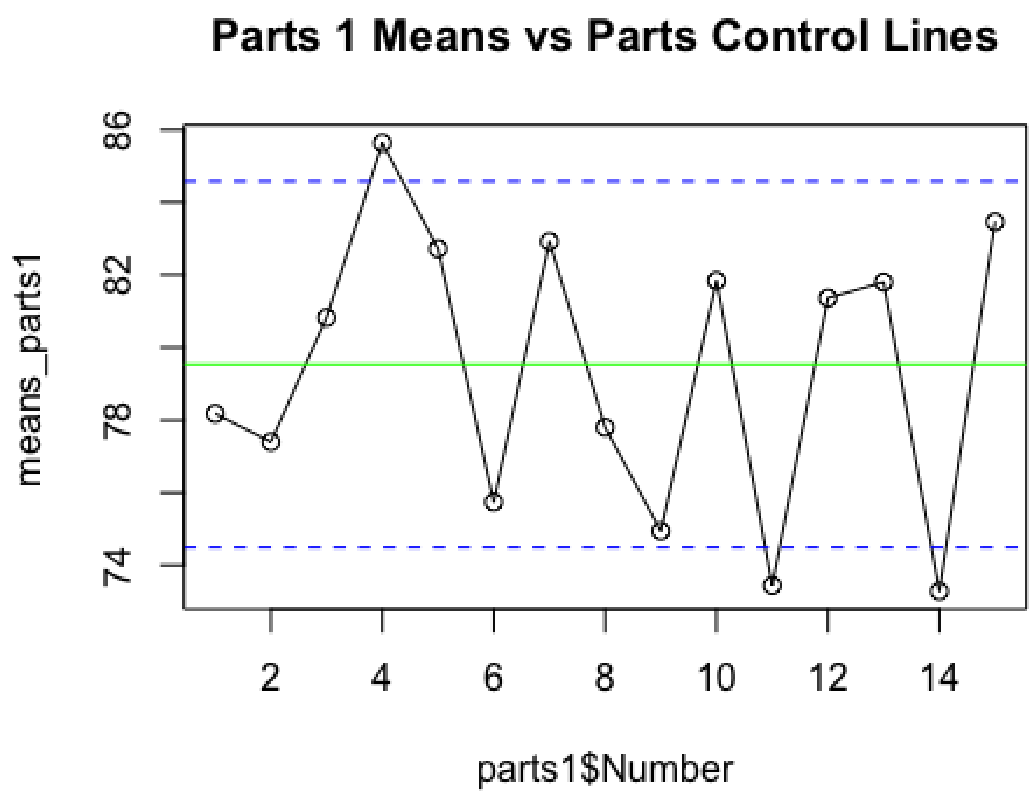
```r
# Save control limits
xbar_center <- xbar_obj$center
xbar_std_dev <- xbar_obj$std.dev

R_center <- range_obj$center
R_std_dev <- range_obj$std.dev

(xbar_obj)

## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = xbar
## Data (phase I)             = parts_data
## Number of groups           = 20
## Group sample size          = 5
## Center of group statistics = 79.533
## Standard deviation         = 3.759673
##
## Control limits at nsigmas  = 3
##       LCL      UCL
##  74.48887 84.57713

(range_obj)

## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = R
## Data (phase I)             = parts_data
## Number of groups           = 20
## Group sample size          = 5
## Center of group statistics = 8.745
## Standard deviation         = 3.759673
##
## Control limits at nsigmas  = 3
##  LCL      UCL
##    0 18.49104
```



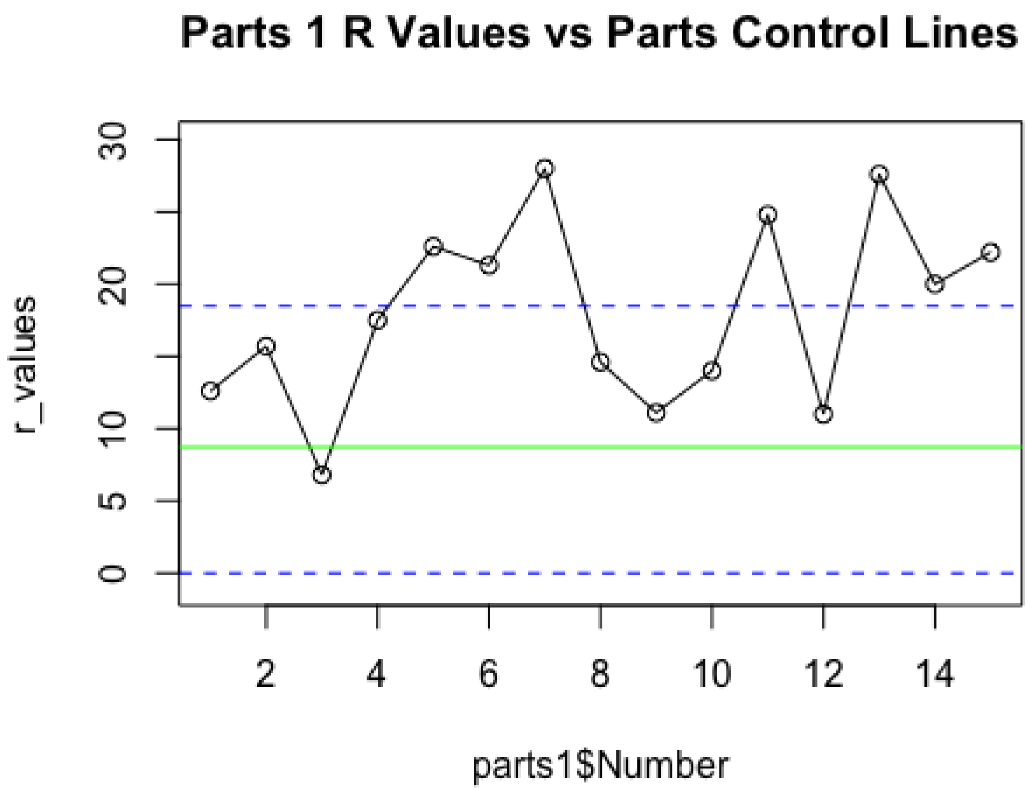
Both x-bar chart and r chart appear in control with all values falling within the range of the lower and upper limits signaling the process is in control for the parts data.

4B

```
# Load the new data
parts1 <- read_excel("Parts1.xlsx", skip = 1)

# Calculate means for parts1_data
means_parts1 <- rowMeans(parts1[, -1])
r_values <- apply(parts1[, -1], 1, function(x) diff(range(x)))

plot(means_parts1 ~ parts1$Number, type = "o", main = "Parts 1 Means vs Parts
Control Lines")
abline(h = xbar_obj$center, col = "green")
abline(h = xbar_obj$limits[1], col = "blue", lty = 2)
abline(h = xbar_obj$limits[2], col = "blue", lty = 2)
```




```
plot(r_values ~ parts1$Number, type = "o", ylim = c(-1, 30), main = "Parts 1
R Values vs Parts Control Lines")
abline(h = range_obj$center, col = "green")
abline(h = range_obj$limits[1], col = "blue", lty = 2)
abline(h = range_obj$limits[2], col = "blue", lty = 2)
```

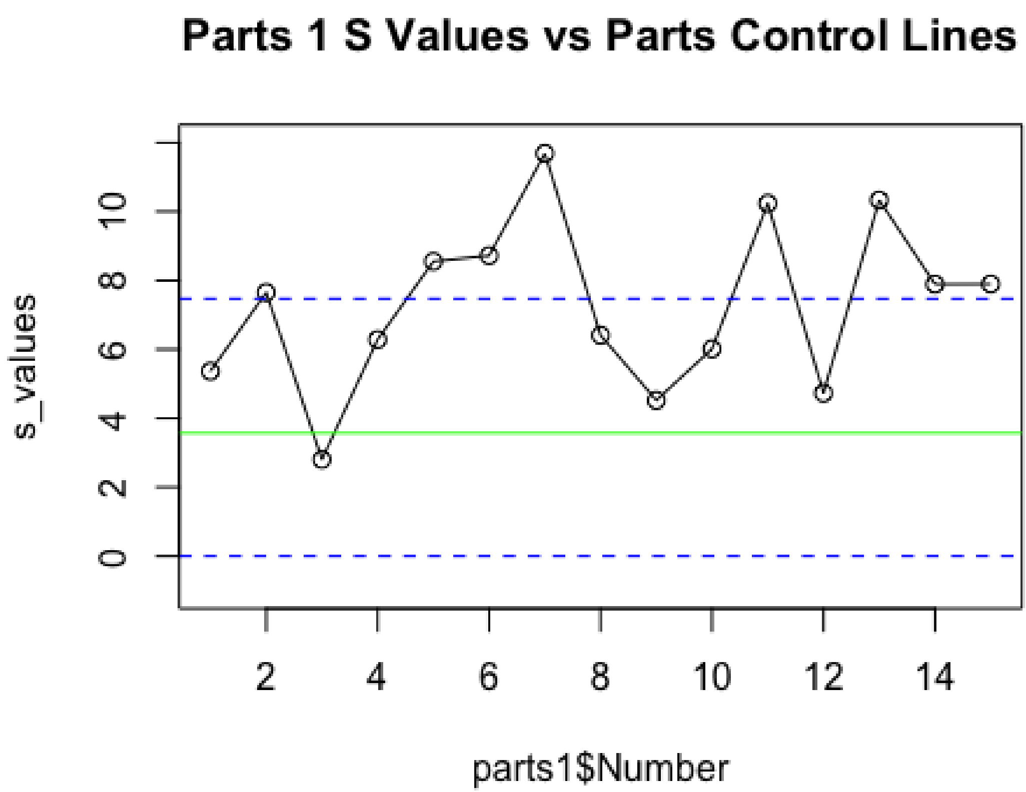


Based on this output the parts1 sample means and range values were plotted against the established control lines. When looking at both charts there are several points that appear out of the limits meaning the process is not in statistical control.

4C Repeat the above two parts using x-bar and s charts. Comment on your findings.

```
# Calculate and plot the X-bar and s values of the new data against the
control limits of the initial data
(s_obj <- qcc(parts[,-1], type="S"))

## ── Quality Control Chart ─────────────────────────
##
## Chart type                 = S
## Data (phase I)             = parts[, -1]
## Number of groups           = 20
## Group sample size          = 5
## Center of group statistics = 3.574917
## Standard deviation         = 3.803162
##
## Control limits at nsigmas  = 3
##  LCL      UCL
##    0 7.467994
```

```
s_center <- s_obj$center
s_std_dev <- s_obj$std.dev

s_values <- apply(parts1[, -1], 1, sd)
plot(s_values ~ parts1$Number, type = "o", ylim = c(-1, 12), main = "Parts 1
S Values vs Parts Control Lines")
abline(h = s_center, col = "green")
abline(h = s_obj$limits[1], col = "blue", lty = 2)
abline(h = s_obj$limits[2], col = "blue", lty = 2)
```

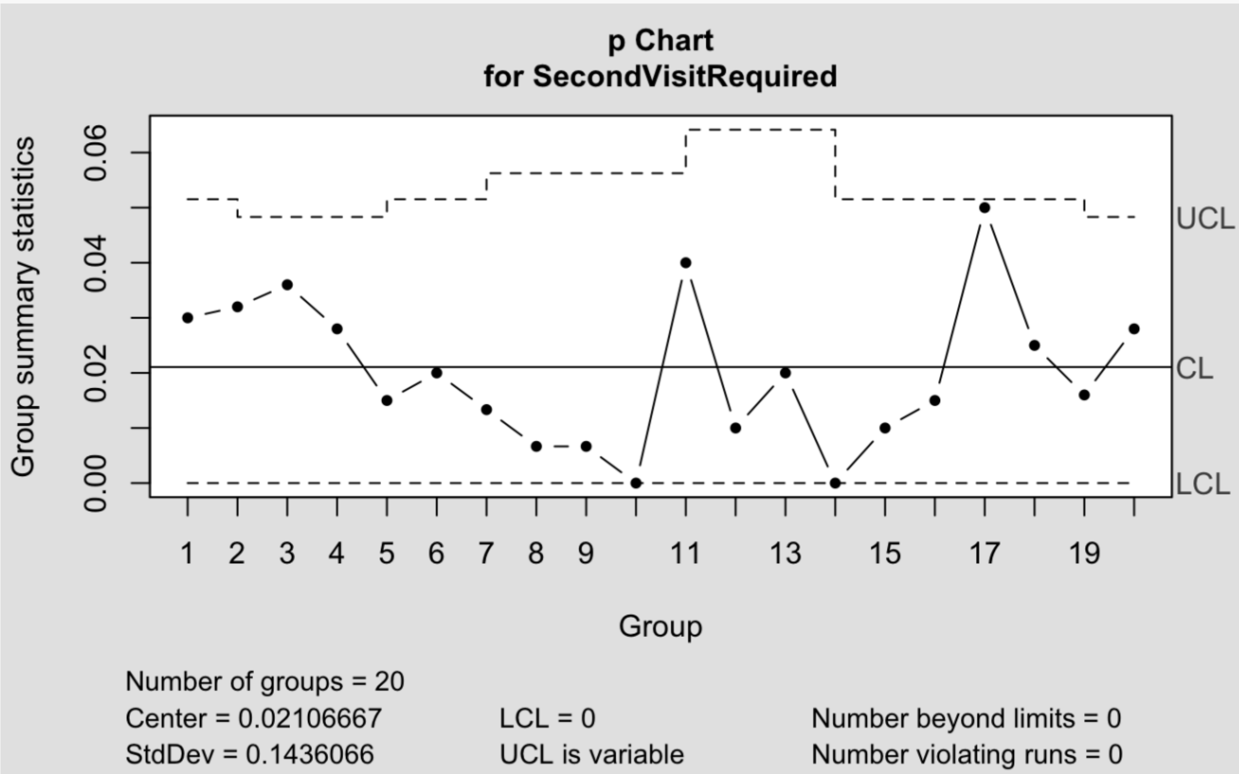


Once again the s_values which are the standard deviations are out of control on the chart when being fitted to the control values from the original parts dataset.

5. The file Work.xlsx records the number of auto repair jobs at a workshop that require a second call to complete. Plot appropriate charts to determine if the process is in statistical control. What would you recommend to the workshop?

```
work = read_excel("Work.xlsx")
names(work) <- c("Week", "TotalRequests", "SecondVisitRequired")
q = with(work, qcc(SecondVisitRequired, sizes = TotalRequests, type = "p"))
(q)

## ── Quality Control Chart ─────────────────────────
##
```

```
## Chart type                 = p
## Data (phase I)             = SecondVisitRequired
## Number of groups           = 20
## Group sample sizes         =
##   sizes  100 150 200 250
##   counts   3   4   8   5
## Center of group statistics = 0.02106667
## Standard deviation         = 0.01015452 0.00908248 0.00908248 ...
##
## Control limits at nsigmas  = 3
##   LCL        UCL
##     0 0.05153023
##     0 0.04831411
## :
##     0 0.04831411
```



Based on the output of the selected chart, p chart since it can be used to track the proportion of repair jobs that require a second visit out of the total number of requests. Based on the results of the p chart the process is in control since no points are outside the limits.

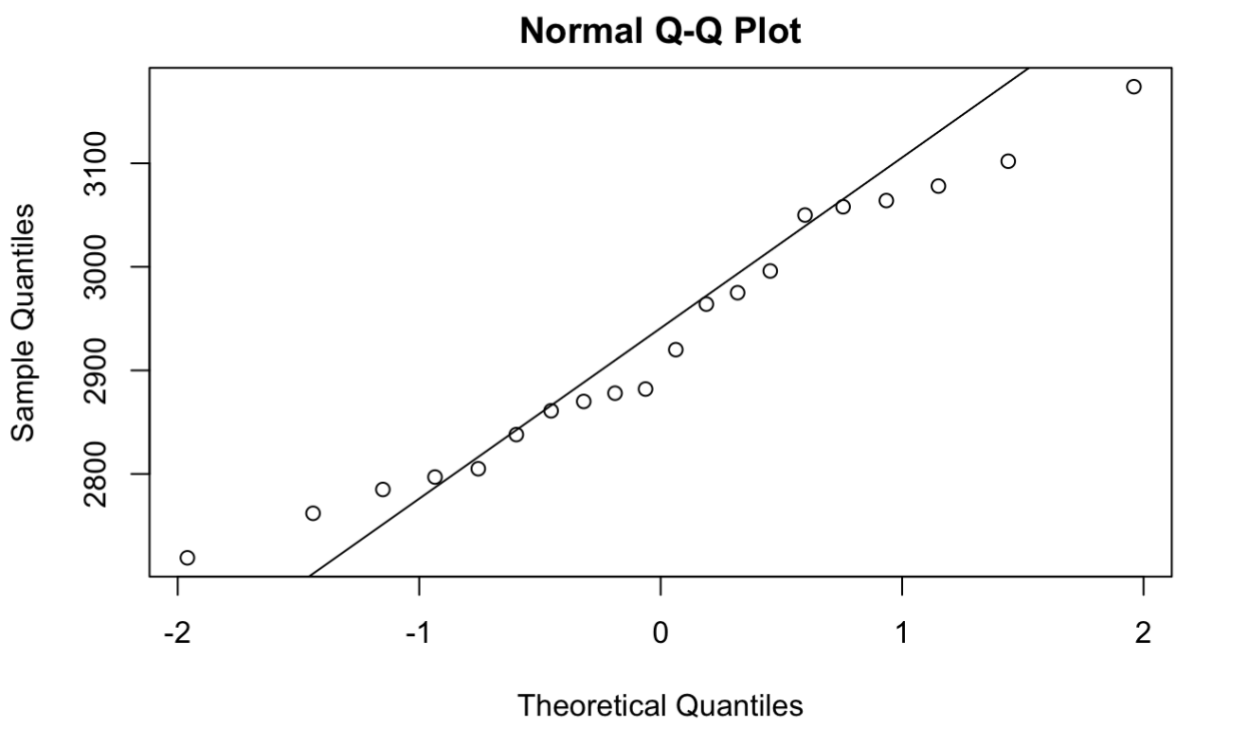
6 (a) Does viscosity follow a normal distribution?

```
viscosity <- read_excel("Viscosity.xlsx")
shapiro.test(viscosity$Viscosity)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  viscosity$Viscosity
## W = 0.96211, p-value = 0.5869
```

```
qqnorm(viscosity$Viscosity)
```

```
qqline(viscosity$Viscosity)
```



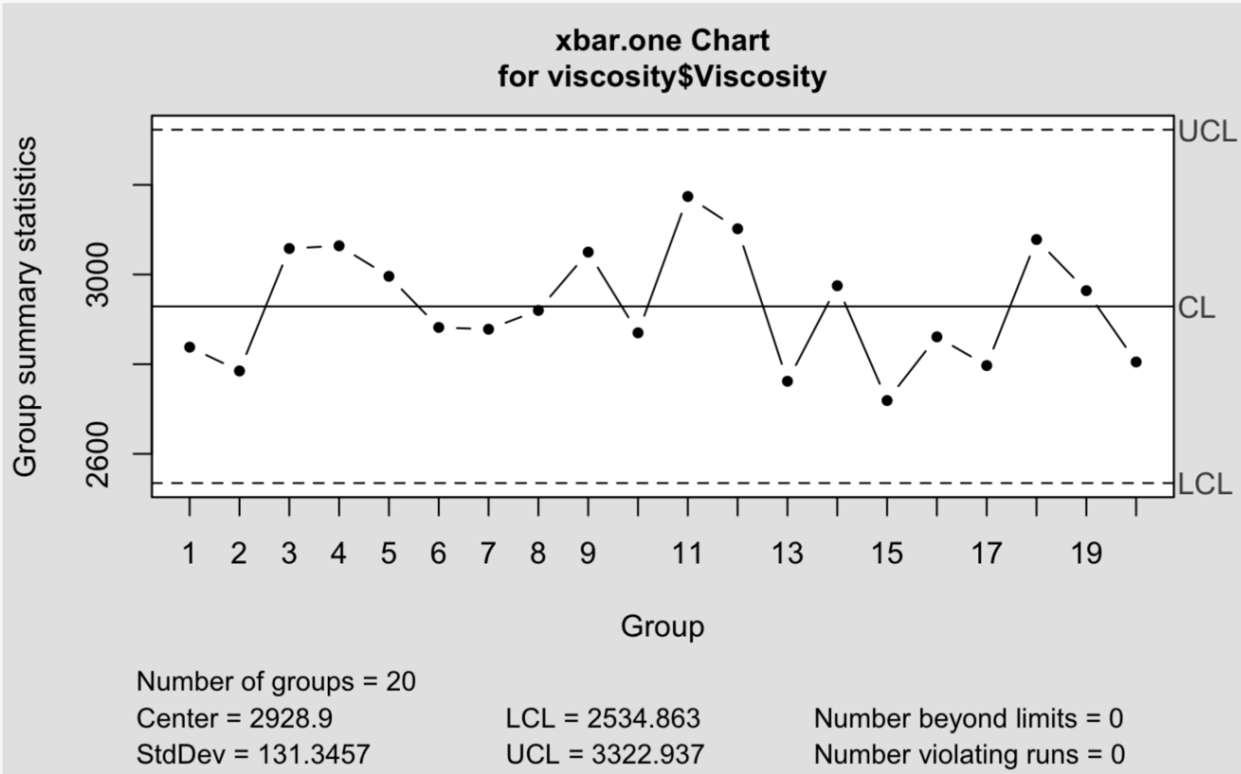
Based on the very large p-value 0.587 > 0.05, it implies the null hypothesis is true and that the data is normally distributed. Also the Normal Q-Q Plot supports this since most points follow the line indicating a normal fit, the points at the end that deviate imply the tails do deviate a little bit but overall it seems to follow a normal distribution.

Plot an appropriate control chart for the data. Does the process exhibit statistical control?

```
control_chart <- qcc(viscosity$Viscosity, type = "xbar.one") # Replace
MeasurementColumn with the correct column name.
(control_chart)
```

```
## ── Quality Control Chart ─────────────────────────
##
## Chart type                  = xbar.one
## Data (phase I)              = viscosity$Viscosity
## Number of groups            = 20
## Group sample size           = 1
```

```
## Center of group statistics = 2928.9
## Standard deviation          = 131.3457
##
## Control limits at nsigmas   = 3
##       LCL      UCL
##  2534.863 3322.937
```




```
The process appears to be in control since all the values fall within the
limits.
```

Estimate the process mean and standard deviation.

```r
mean_viscosity = mean(viscosity$Viscosity)
sd_viscosity = sd(viscosity$Viscosity)

print(paste("Process mean: ", mean_viscosity))

## [1] "Process mean:  2928.9"

print(paste("Process standard deviation: ", sd_viscosity))

## [1] "Process standard deviation:  129.023416887258"
```

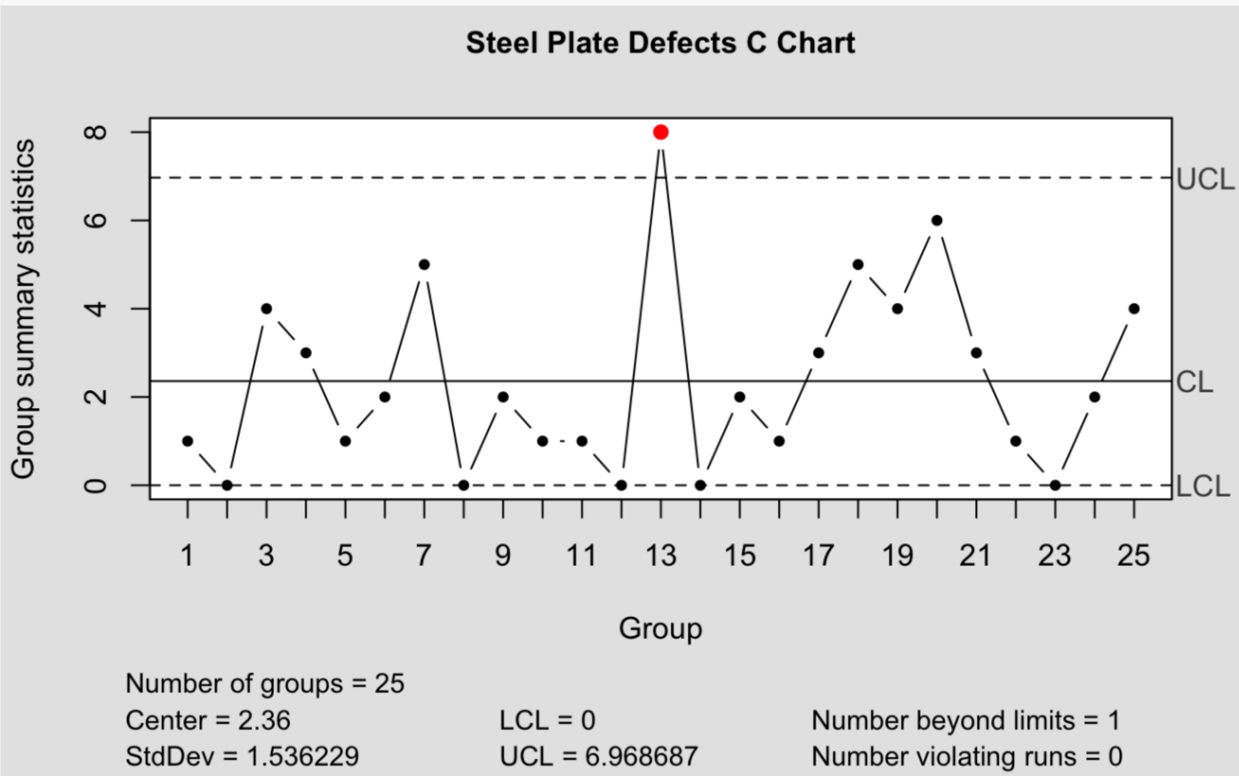
MEAN = 2928.9

SD = 129.0

Surface defects have been counted on twenty-five rectangular steel plates, and the data are given in the file steel.xlsx. Set up a control chart for non-conformities using these data. Does the process producing the plates appear to be in statistical control?

```
steel = read_excel("Steel.xlsx")
steel_chart <- qcc(steel$`No of Defects`, type = "c", title = "Steel Plate
Defects C Chart")
(steel_chart)
```

```
## ── Quality Control Chart ────────────────────────
##
## Chart type                = c
## Data (phase I)            = steel$`No of Defects`
## Number of groups          = 25
## Group sample size         = 1
## Center of group statistics = 2.36
## Standard deviation        = 1.536229
##
## Control limits at nsigmas  = 3
##  LCL      UCL
##    0 6.968687
```



Based on this plot the system appears to not be in statistical control since the point at 13 violates the upper boundary.