

# R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
library(data.table)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##   between, first, last

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(qcc)
```

```
## Package 'qcc' version 2.7
## Type 'citation("qcc")' for citing this R package in publications.
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##   select
```

```
library(graphics)
Pre=fread(input = "C:\\Users\\ethan\\Downloads\\PreJune3_Pitchdata.csv" )
Post=fread(input = "C:\\Users\\ethan\\Downloads\\PostJune3rdData_correctfile.csv")
low1=Pre %>% arrange(xwoba)
low2=Post %>% arrange(xwoba)
```

```
top1=low1[!is.na(low1[,xwoba])]
top2=low2[!is.na(low2[,xwoba])]
```

```
topbefore=tail(top1,20)
topafter=tail(top2,20)
```

```
cat("Best Pitchers' xwoba before June 3rd","\n")
```

```

## Best Pitchers' xwoba before June 3rd
summary(topbefore[,xwoba])

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.4380 0.4557 0.4935 0.5259 0.5700 0.7760
cat("Best Pitchers' xwoba after June 3rd","\n")

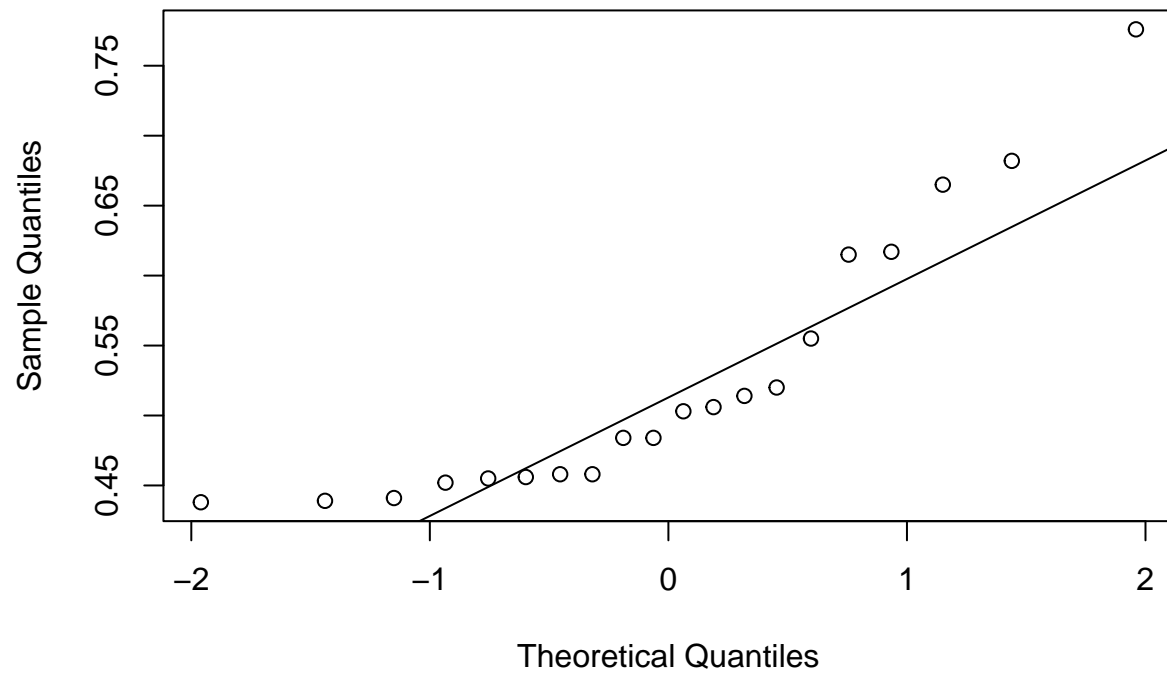
## Best Pitchers' xwoba after June 3rd
summary(topafter[,xwoba])

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.4550 0.4828 0.4995 0.5331 0.5707 0.7220
t.test(x = topbefore[,xwoba],y = topafter[,xwoba],paired = FALSE,var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data: topbefore[, xwoba] and topafter[, xwoba]
## t = -0.25789, df = 36.559, p-value = 0.7979
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.0633483 0.0490483
## sample estimates:
## mean of x mean of y
## 0.52590 0.53305
qqnorm(y = unlist(topbefore[,xwoba]),plot.it = TRUE)
qqline(y = unlist(topbefore[,xwoba]))

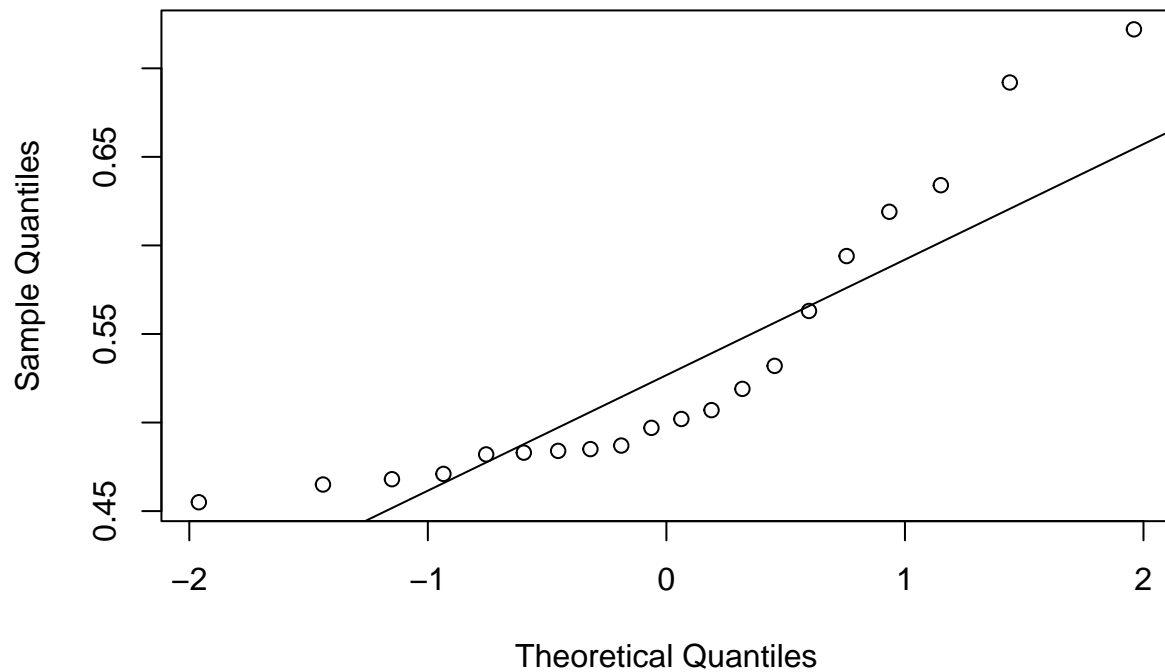
```

### Normal Q-Q Plot



```
qqnorm(y = unlist(topafter[,xwoba]),plot.it = TRUE)  
qqline(y = unlist(topafter[,xwoba]))
```

## Normal Q-Q Plot



```
shapiro.test(unlist(topbefore[,xwoba]))
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  unlist(topbefore[, xwoba])  
## W = 0.83152, p-value = 0.002653
```

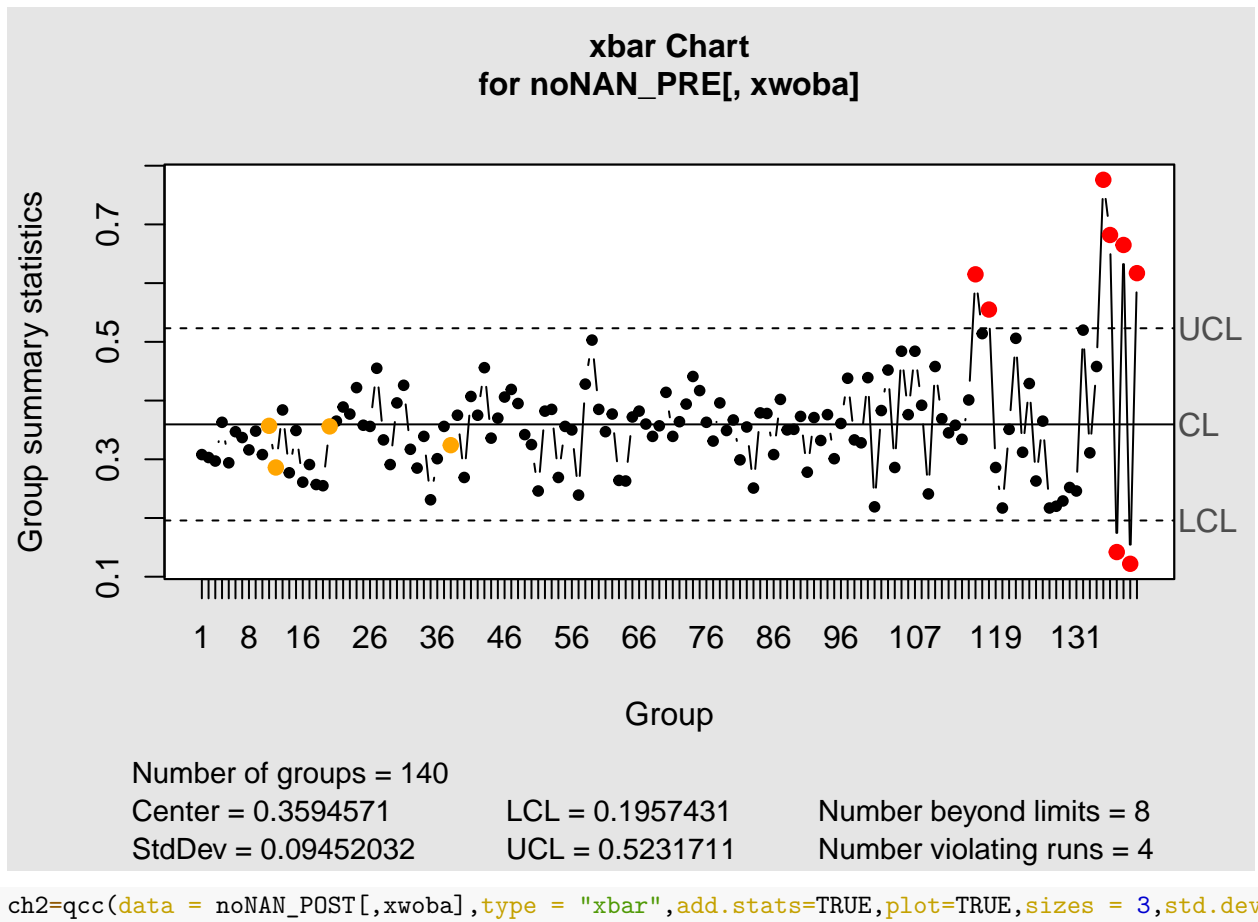
```
shapiro.test(unlist(topafter[,xwoba]))
```

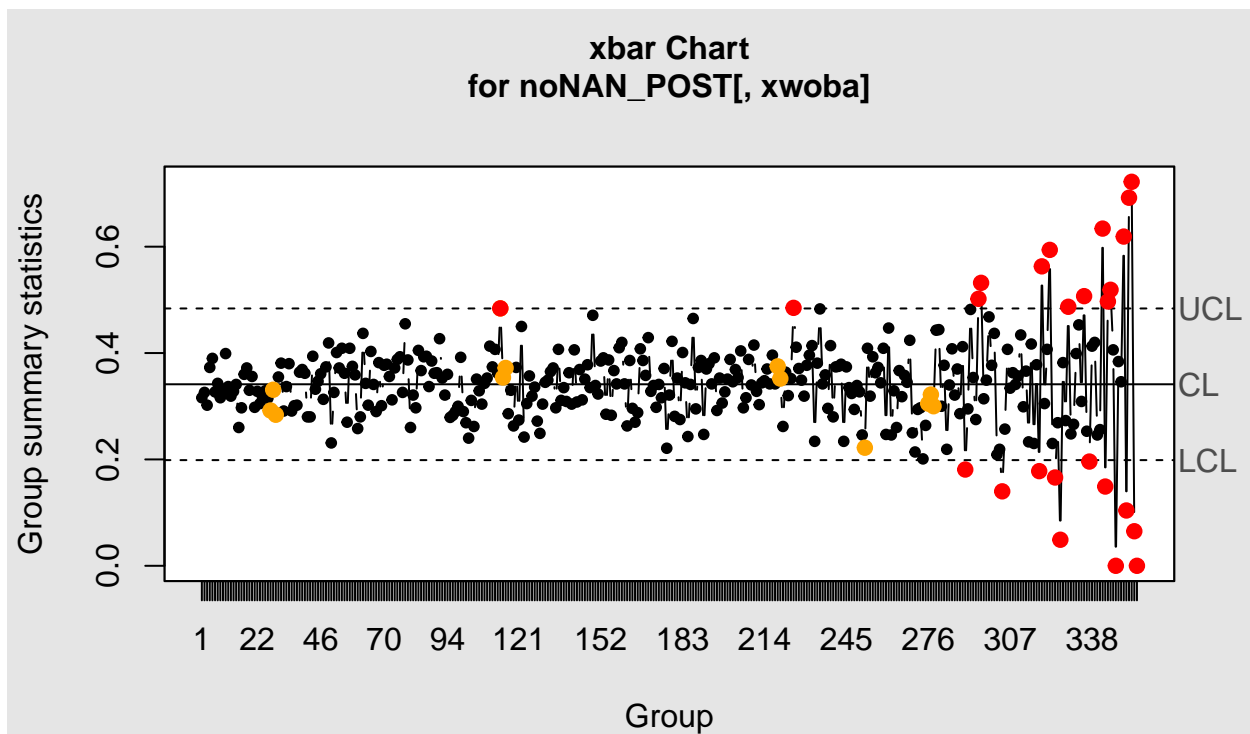
```
##  
##  Shapiro-Wilk normality test  
##  
## data:  unlist(topafter[, xwoba])  
## W = 0.82356, p-value = 0.001983
```

```
sdxwoba1=sd(unlist(top1[,xwoba]))  
sdxwoba2=sd(unlist(top2[,xwoba]))
```

```
noNAN_PRE=Pre[!is.na(Pre[,xwoba])]  
noNAN_POST=Post[!is.na(Post[,xwoba])]
```

```
ch1=qcc(data = noNAN_PRE[,xwoba],type = "xbar",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sdxwoba1)
```





Number of groups = 355

Center = 0.3412817

StdDev = 0.08230939

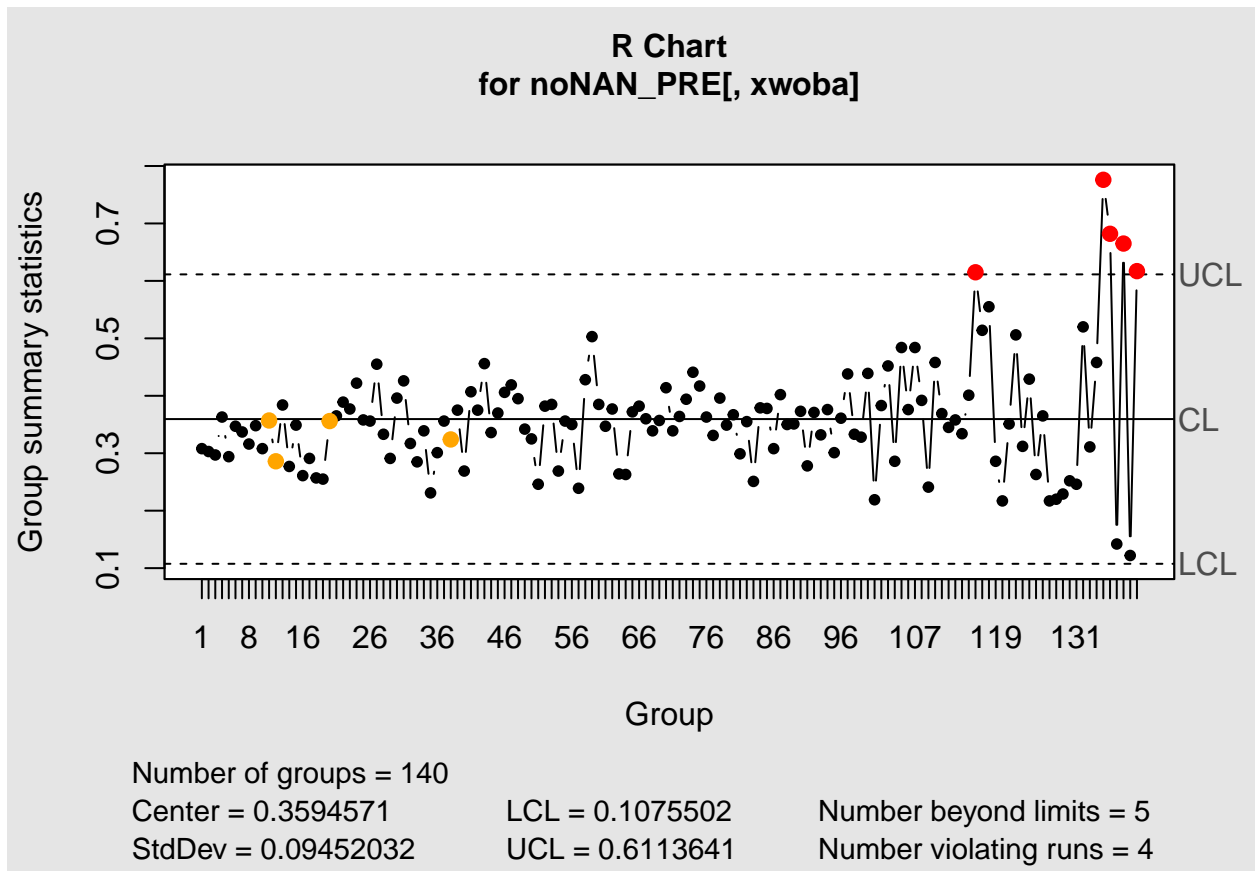
LCL = 0.1987176

UCL = 0.4838457

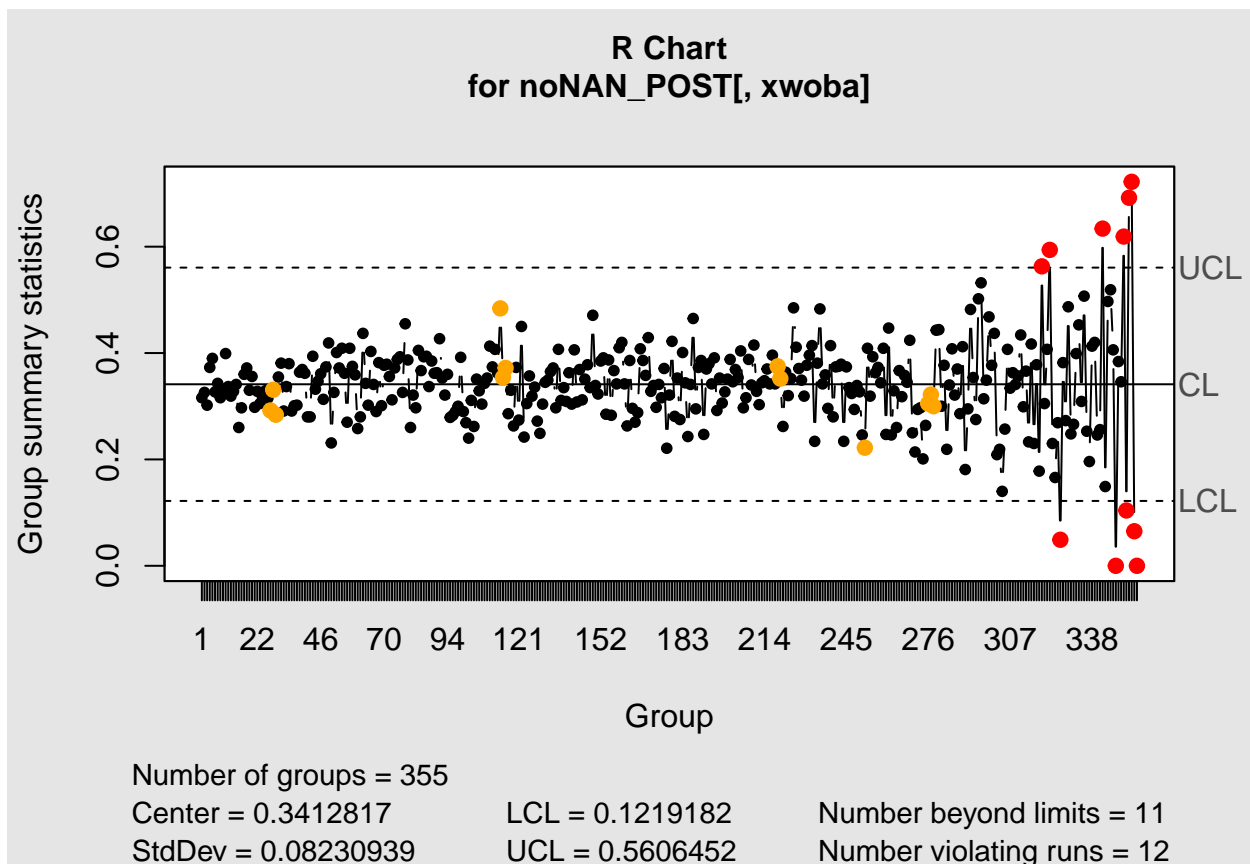
Number beyond limits = 25

Number violating runs = 12

```
ch3=qcc(data = noNAN_PRE[,xwoba],type = "R",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sdwoba1)
```



```
ch4=qcc(data = noNAN_POST[,xwoba],type = "R",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sdwxwoba2)
```



```
cat("Best Pitchers' spin rate before June 3rd","\n")
```

```
## Best Pitchers' spin rate before June 3rd
```

```
summary(topbefore[,spin_rate])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1858   2139   2186   2190   2262   2443
```

```
cat("Best Pitchers' spin rate after June 3rd","\n")
```

```
## Best Pitchers' spin rate after June 3rd
```

```
summary(topafter[,spin_rate])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1779   2070   2204   2190   2291   2576
```

```
t.test(x = top1[,spin_rate],y =top2[,spin_rate],paired = FALSE,var.equal = FALSE)
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: top1[, spin_rate] and top2[, spin_rate]
```

```
## t = 3.1873, df = 229.92, p-value = 0.001636
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

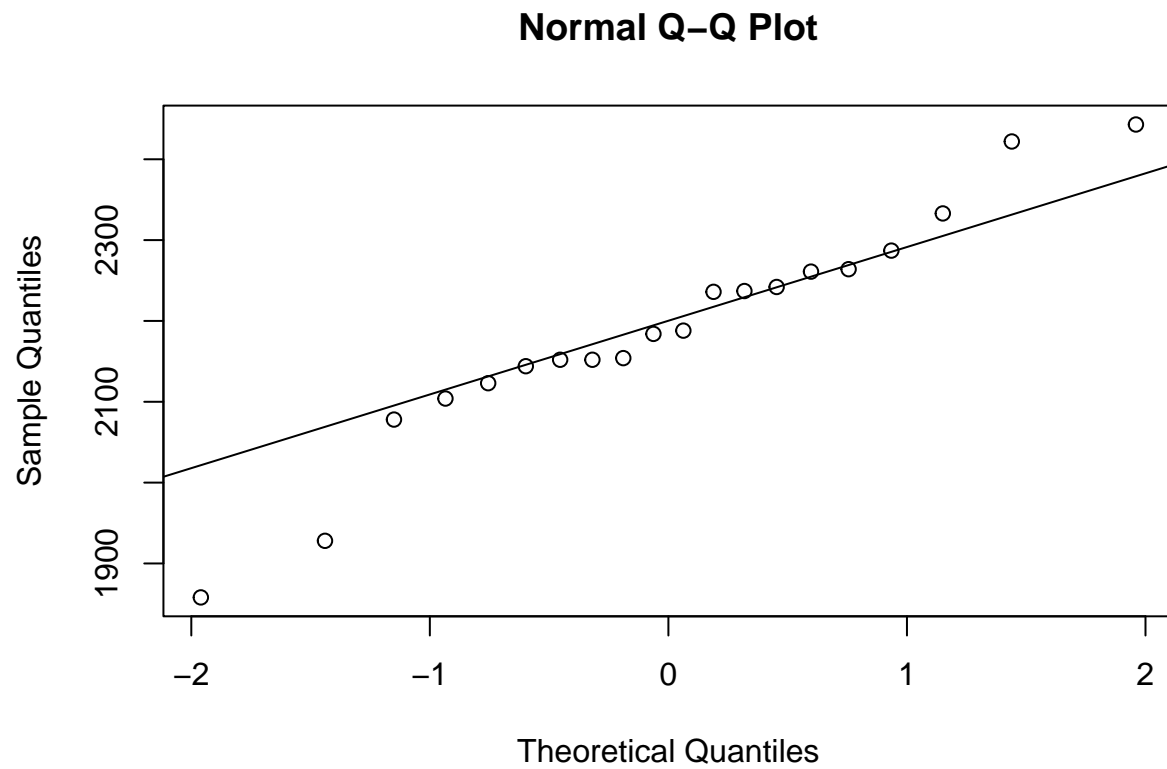
```
##  20.25538 85.84502
```

```
## sample estimates:
```

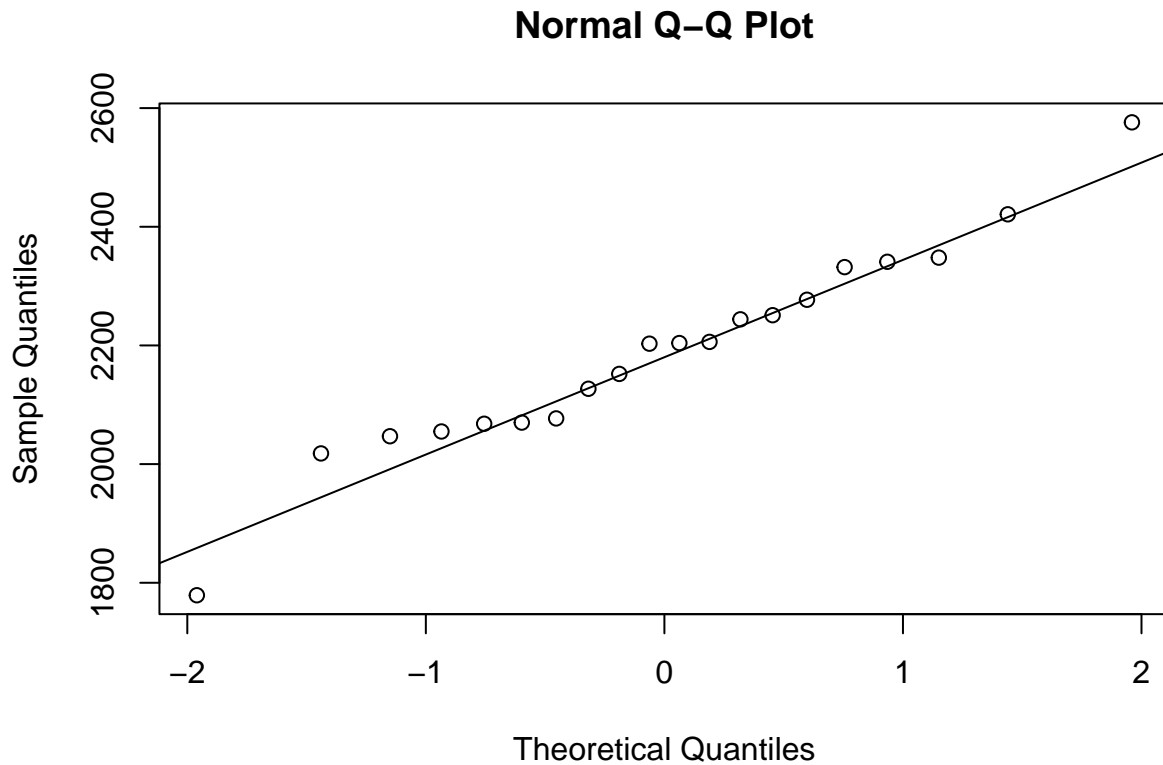


```
## mean of x mean of y
## 2290.064 2237.014

qqnorm(y = unlist(topbefore[,spin_rate]),plot.it = TRUE)
qqline(y = unlist(topbefore[,spin_rate]))
```



```
qqnorm(y = unlist(topafter[,spin_rate]),plot.it = TRUE)
qqline(y = unlist(topafter[,spin_rate]))
```



```
shapiro.test(unlist(topbefore[,spin_rate]))
```

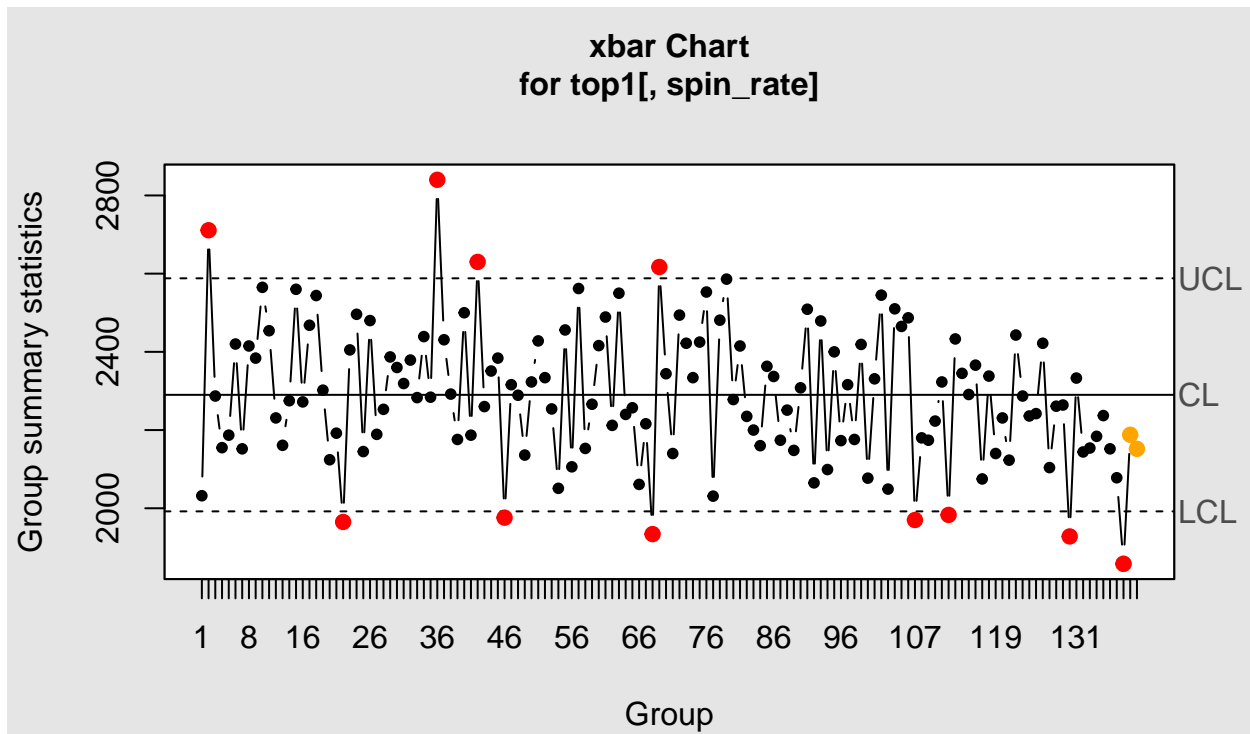
```
##
##  Shapiro-Wilk normality test
##
## data:  unlist(topbefore[, spin_rate])
## W = 0.95171, p-value = 0.3939
```

```
shapiro.test(unlist(topafter[,spin_rate]))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  unlist(topafter[, spin_rate])
## W = 0.97136, p-value = 0.7833
```

```
sdspin1=sd(unlist(top1[,spin_rate]))
sdspin2=sd(unlist(top2[,spin_rate]))
```

```
ch5=qcc(data = top1[,spin_rate],type = "xbar",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sdspin1)
```



Number of groups = 140

Center = 2290.064

StdDev = 172.0536

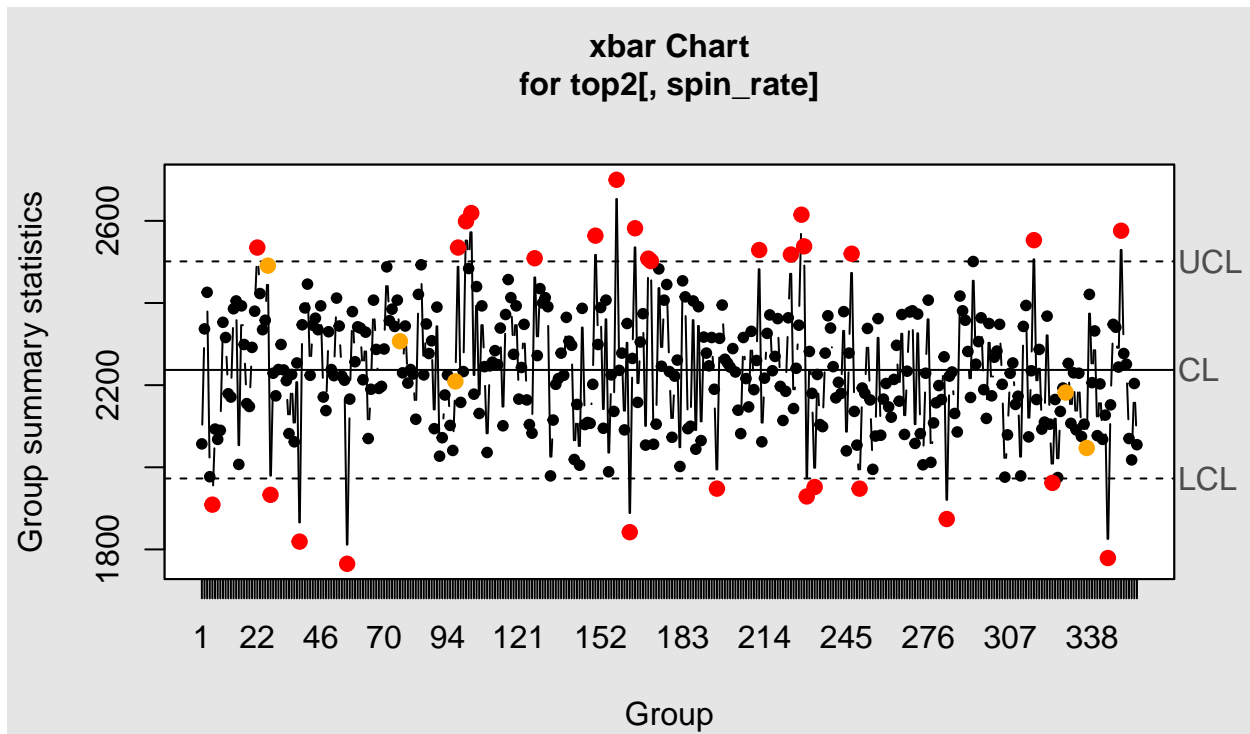
LCL = 1992.059

UCL = 2588.07

Number beyond limits = 11

Number violating runs = 3

```
ch6=qcc(data = top2[,spin_rate],type = "xbar",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sdspin2)
```



Number of groups = 355

Center = 2237.014

StdDev = 152.5879

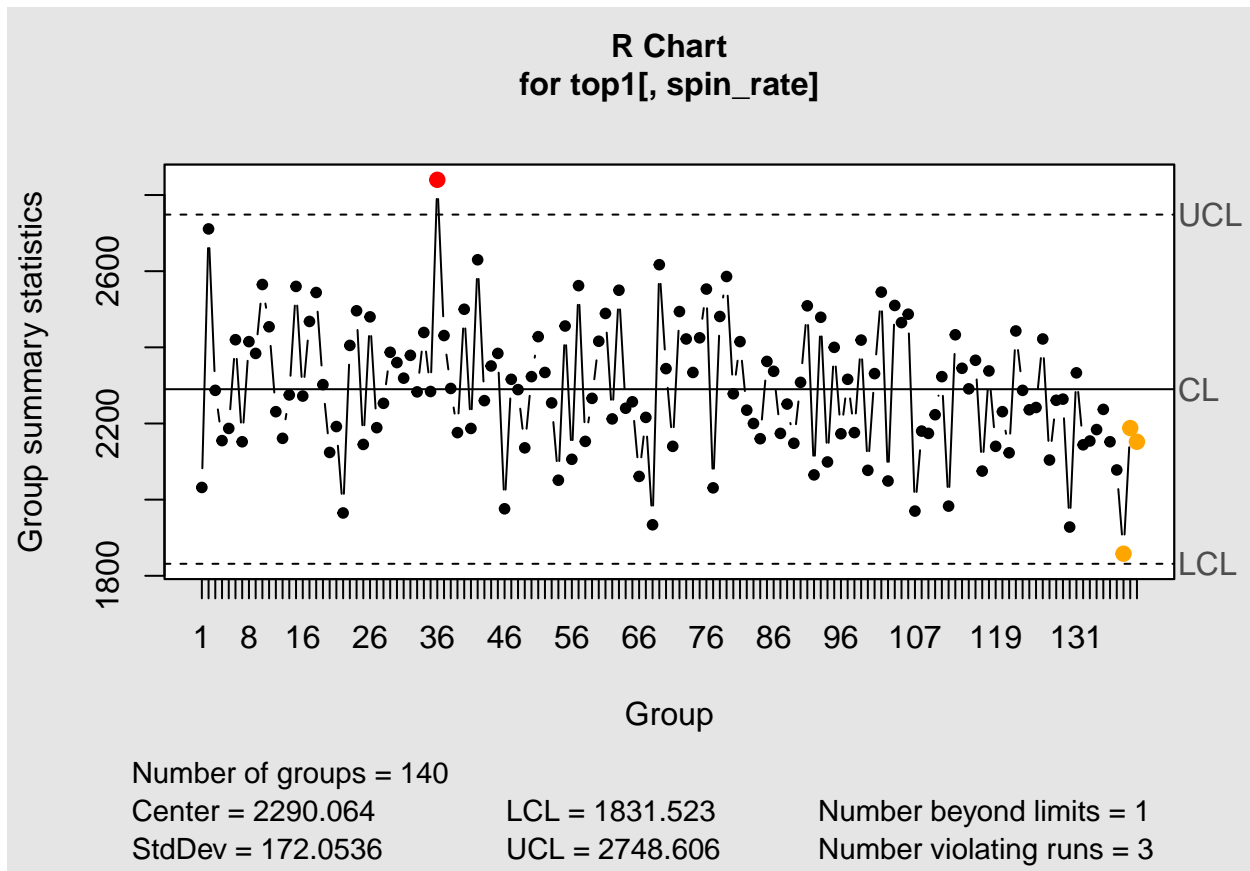
LCL = 1972.724

UCL = 2501.304

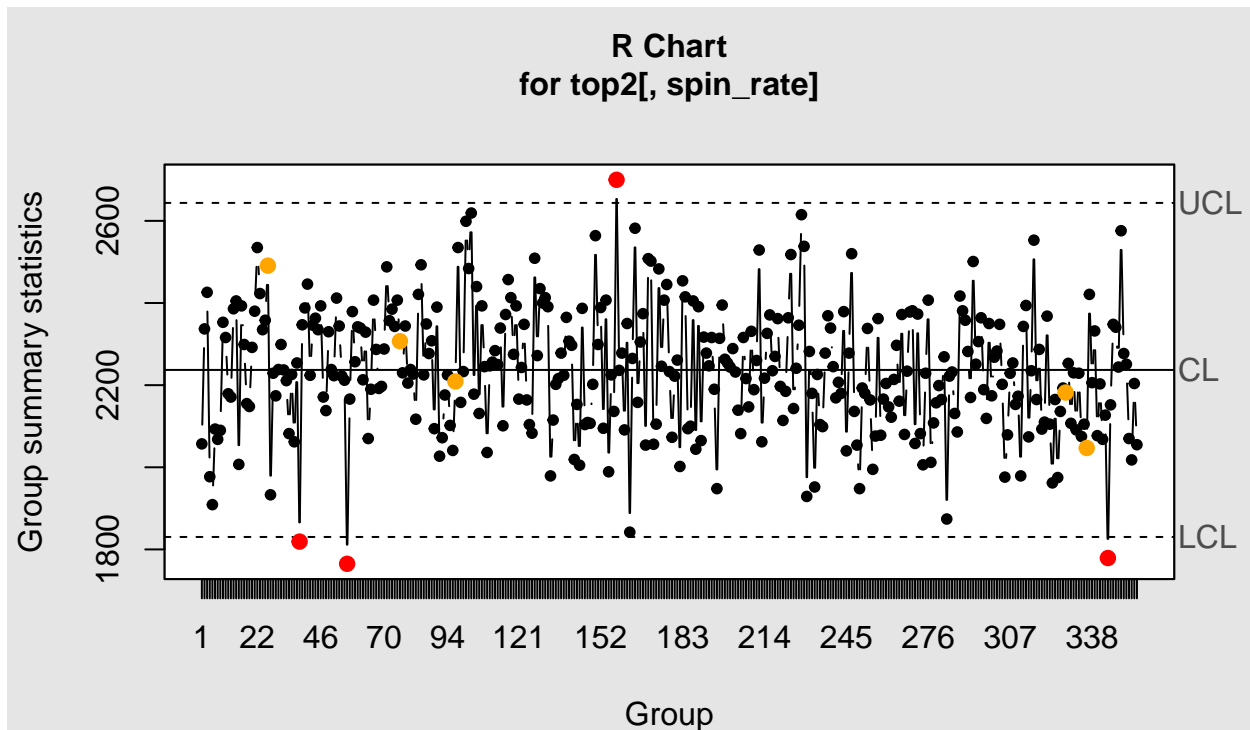
Number beyond limits = 29

Number violating runs = 5

```
ch7=qcc(data = top1[,spin_rate],type = "R",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sdspin1)
```



```
ch8=qcc(data = top2[,spin_rate],type = "R",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sdspin2)
```



Number of groups = 355

Center = 2237.014

LCL = 1830.351

Number beyond limits = 4

StdDev = 152.5879

UCL = 2643.677

Number violating runs = 5

```
Precurve=fread("C:\\Users\\ethan\\Downloads\\Curveball_data_PreJune3.csv")
Postcurve=fread("C:\\Users\\ethan\\Downloads\\curveballData_postJune3rd.csv")
```

```
low1=Precurve %>% arrange(spin_rate)
low2=Postcurve %>% arrange(spin_rate)
```

```
top1curve=low1[!is.na(low1[,spin_rate])]
top2curve=low2[!is.na(low2[,spin_rate])]
```

```
topbeforecurve=tail(top1curve,20)
topaftercurve=tail(top2curve,20)
```

```
cat("Best Pitchers' spin rate before June 3rd","\n")
```

```
## Best Pitchers' spin rate before June 3rd
```

```
summary(topbeforecurve[,spin_rate])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2694   2731   2810   2838   2884   3271
```

```
cat("Best Pitchers' spin rate after June 3rd","\n")
```

```
## Best Pitchers' spin rate after June 3rd
```

```
summary(topaftercurve[,spin_rate])
```

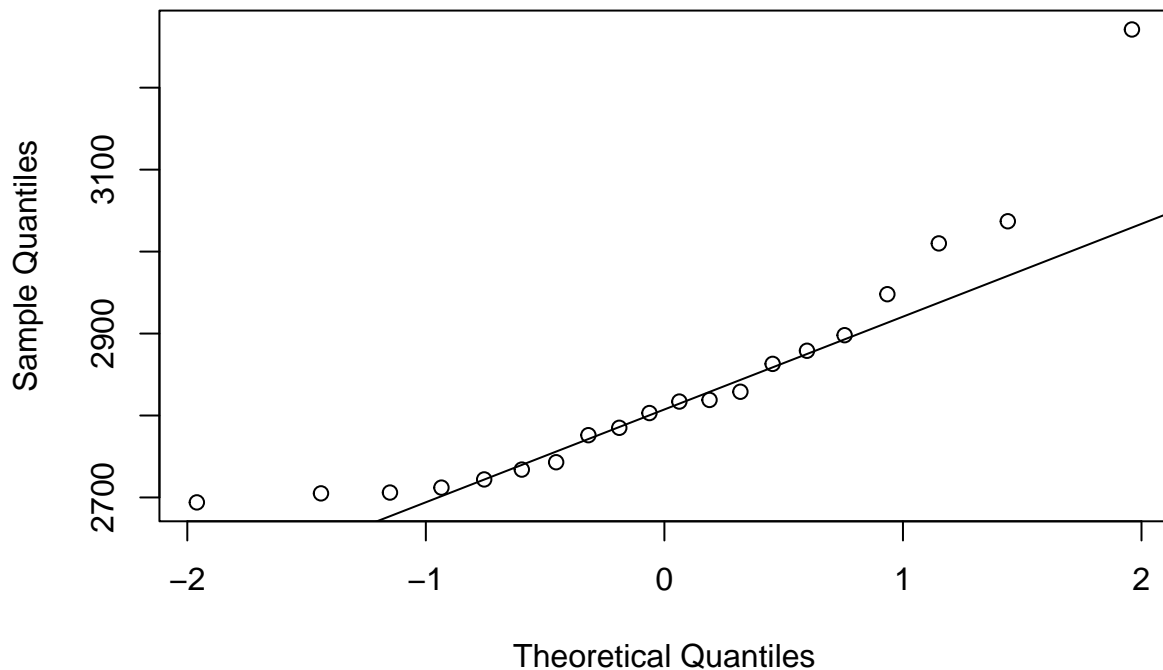
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##      2793      2848      2873      2949      3027      3261
t.test(x = topbeforecurve[,spin_rate],y = topaftercurve[,spin_rate],paired = FALSE,var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data: topbeforecurve[, spin_rate] and topaftercurve[, spin_rate]
## t = -2.5301, df = 37.916, p-value = 0.01568
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -201.25897  -22.34103
## sample estimates:
## mean of x mean of y
##  2837.55  2949.35

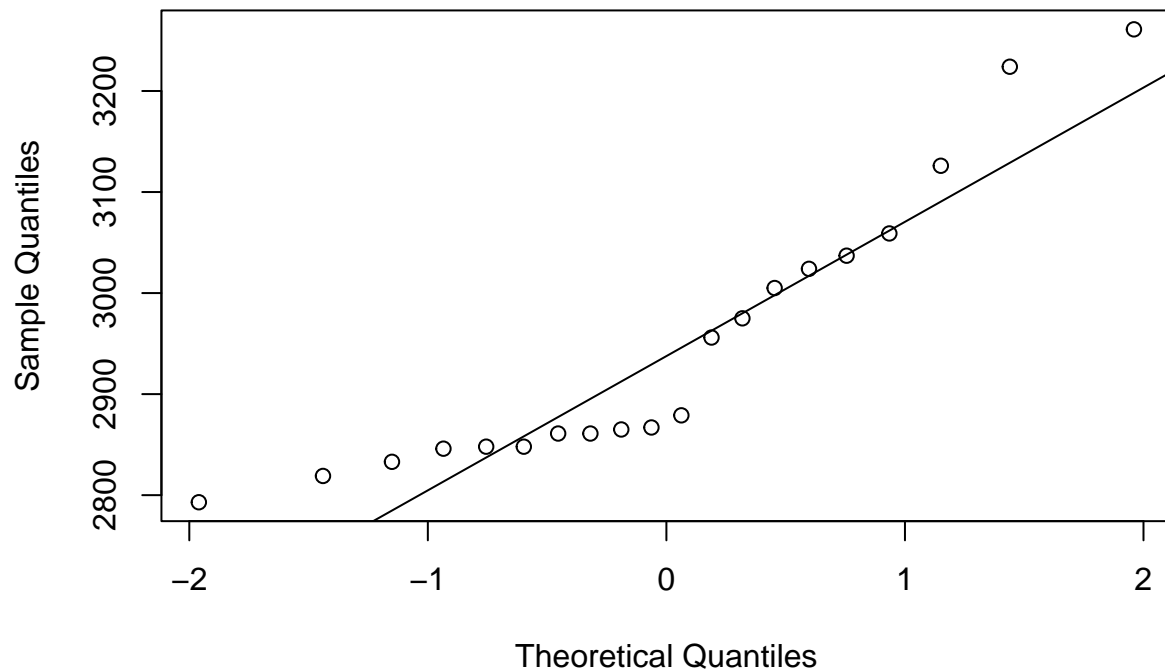
qqnorm(y = unlist(topbeforecurve[,spin_rate]),plot.it = TRUE)
qqline(y = unlist(topbeforecurve[,spin_rate]))
```

### Normal Q-Q Plot



```
qqnorm(y = unlist(topaftercurve[,spin_rate]),plot.it = TRUE)
qqline(y = unlist(topaftercurve[,spin_rate]))
```

## Normal Q-Q Plot



```
shapiro.test(unlist(topbeforecurve[,spin_rate]))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  unlist(topbeforecurve[, spin_rate])
## W = 0.84661, p-value = 0.004679
```

```
shapiro.test(unlist(topaftercurve[,spin_rate]))
```

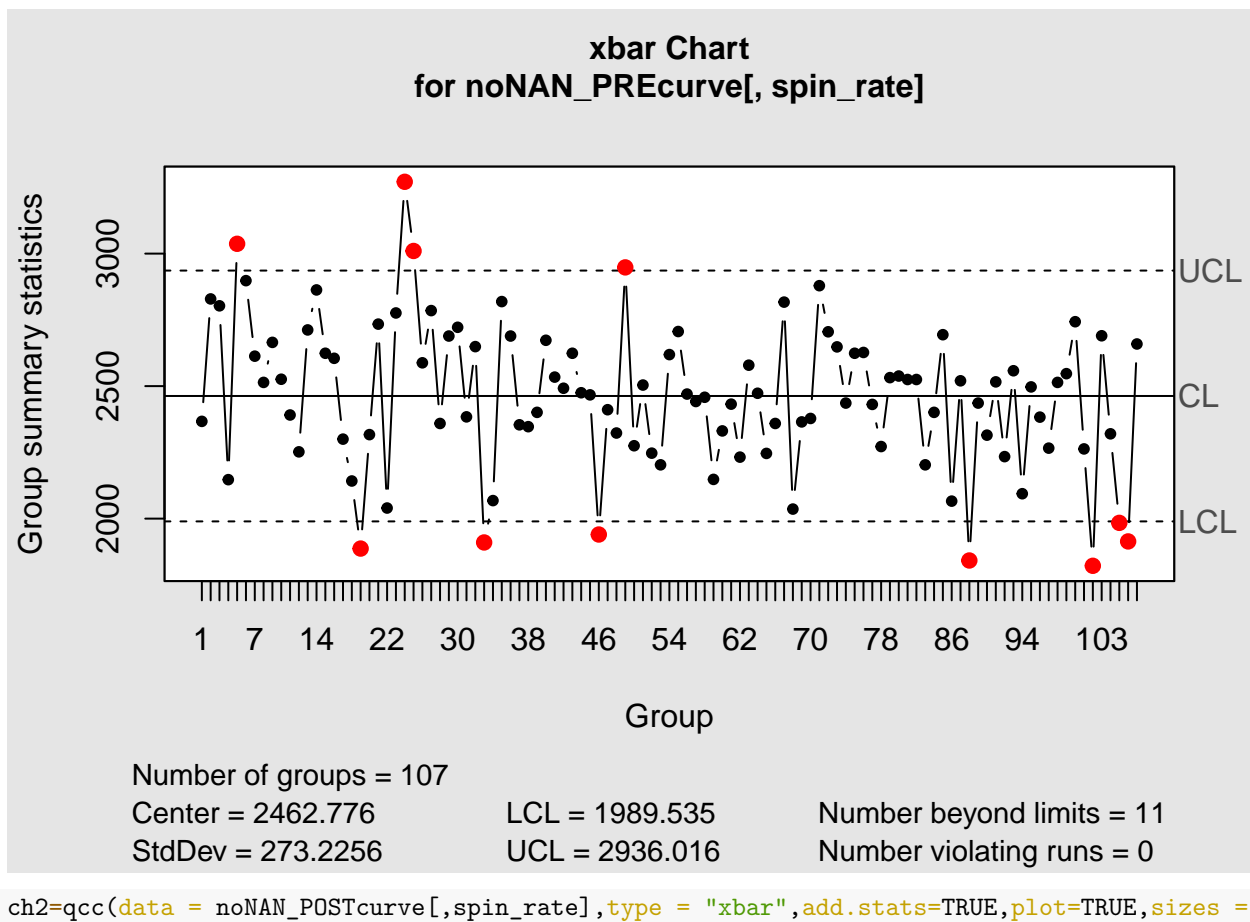
```
##
##  Shapiro-Wilk normality test
##
## data:  unlist(topaftercurve[, spin_rate])
## W = 0.86592, p-value = 0.009968
```

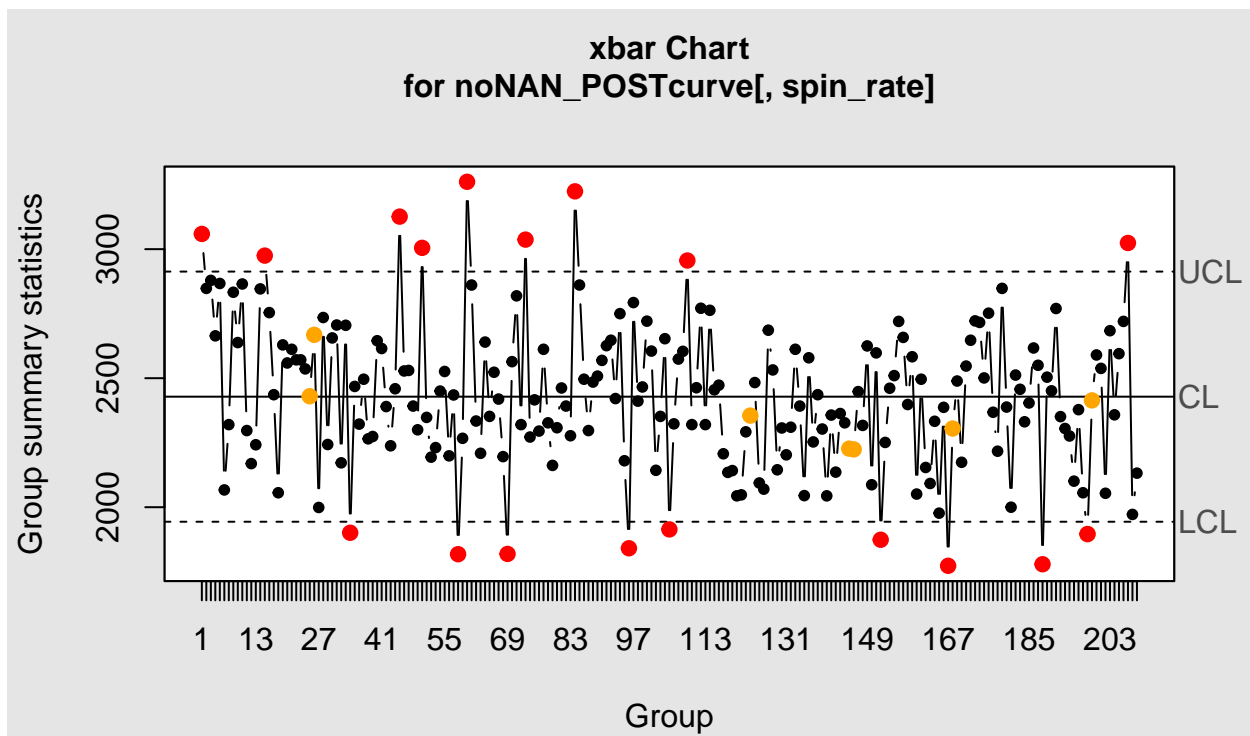
```
sdxwoba1=sd(unlist(top1curve[,spin_rate]))
sdxwoba2=sd(unlist(top2curve[,spin_rate]))
```

```
noNAN_PREcurve=Precurve[!is.na(Precurve[,spin_rate])]
noNAN_POSTcurve=Postcurve[!is.na(Postcurve[,spin_rate])]
```

```
ch1=qcc(data = noNAN_PREcurve[,spin_rate],type = "xbar",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sd
```







Number of groups = 209

Center = 2428.364

StdDev = 279.8259

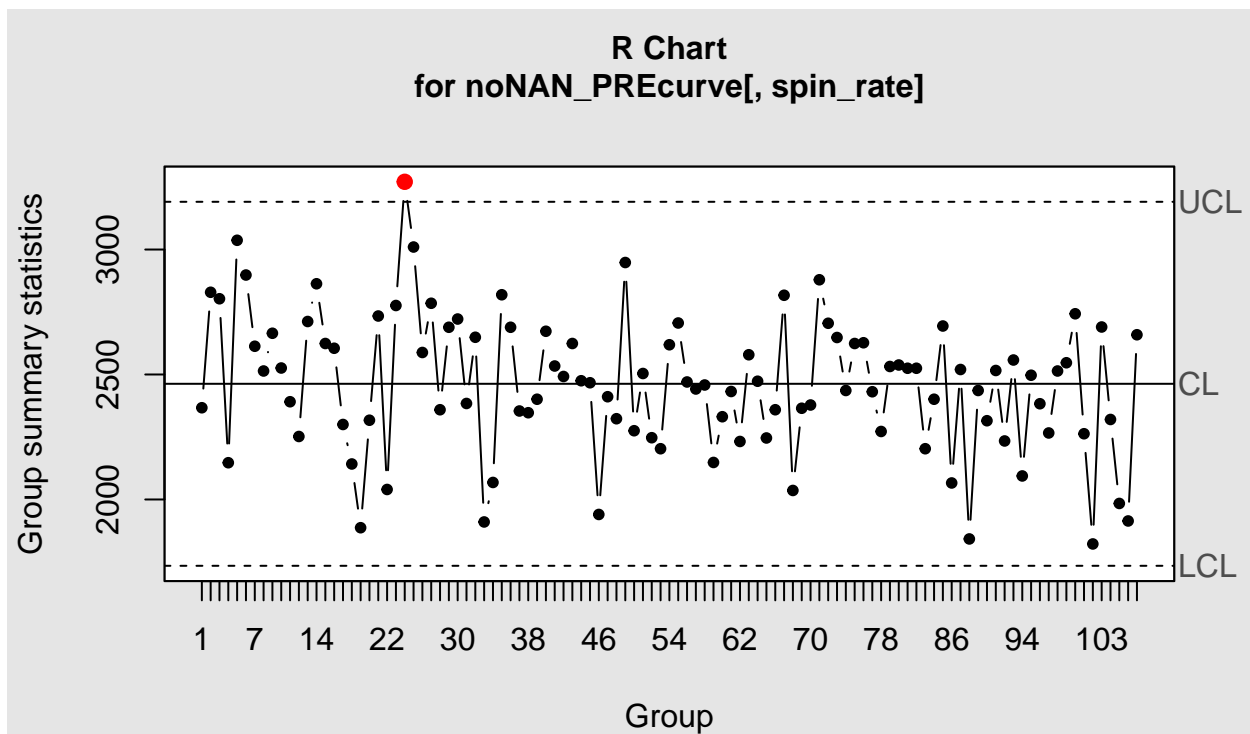
LCL = 1943.691

UCL = 2913.036

Number beyond limits = 18

Number violating runs = 8

```
ch3=qcc(data = noNAN_PREcurve[,spin_rate],type = "R",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sd(xwo
```



Number of groups = 107

Center = 2462.776

StdDev = 273.2256

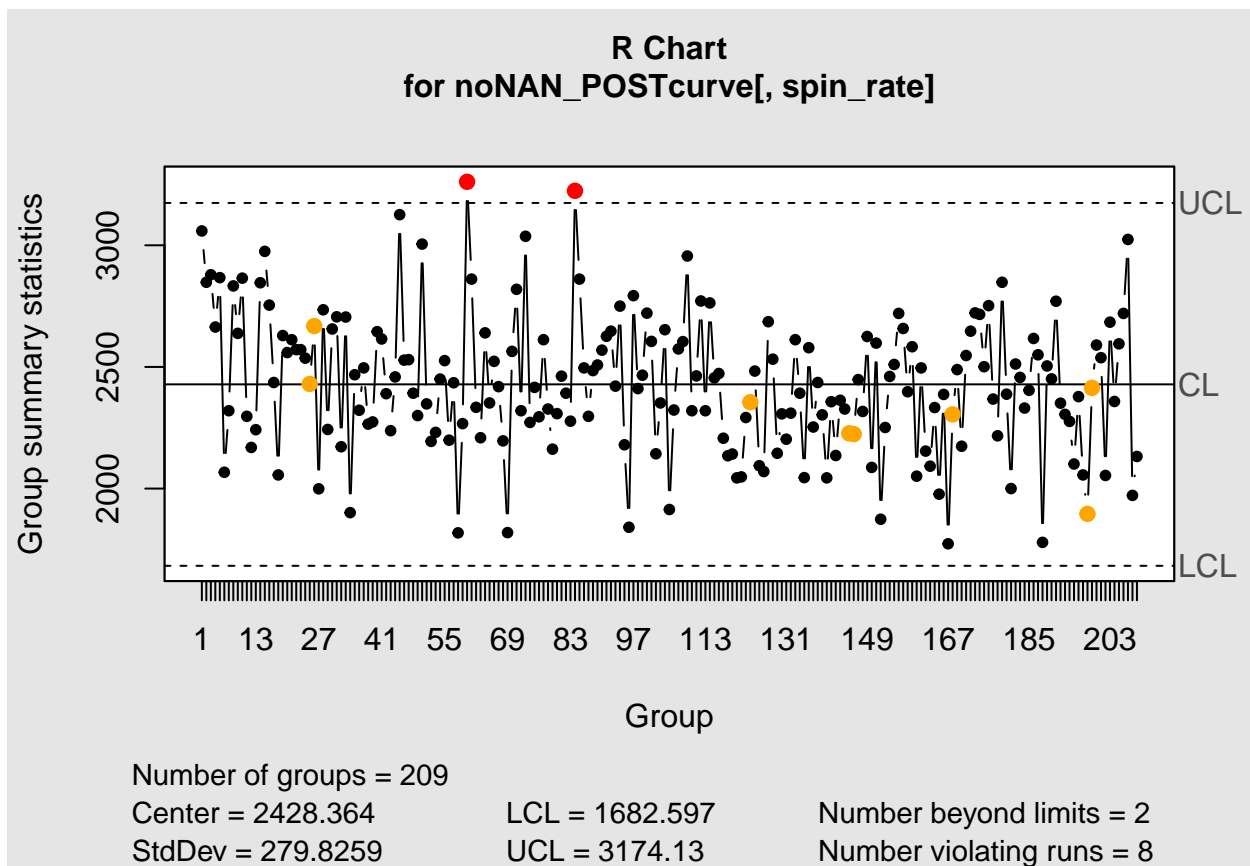
LCL = 1734.6

UCL = 3190.952

Number beyond limits = 1

Number violating runs = 0

```
ch4=qcc(data = noNAN_POSTcurve[,spin_rate],type = "R",add.stats=TRUE,plot=TRUE,sizes = 3,std.dev = sdw
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.