

# ProjectML

```
setwd("C:/Users/Administrator/Documents/archive (3)")
ktd <- read.csv("kdd.csv",header = TRUE )
```

```
ktd<-na.omit(ktd)
#ktd
nrow(ktd)
```

```
## [1] 203
```

```
#sum(is.na(ktd$pe))
```

```
#sum(is.na(ktd$cad))
```

```
#sum(is.na(ktd$bp))
```

```
#sum(is.na(ktd$rbc))
```

```
#view(ktd)
```

```
#the number of records after removing missing values
```

```
kdn<-ktd
```

```
summary(ktd)
```

```
##      age      bp      sg      al
##  Min.   : 6.00  Min.   : 50.00  Min.   :1.005  Min.   :0.0000
##  1st Qu.:42.00  1st Qu.: 70.00  1st Qu.:1.015  1st Qu.:0.0000
##  Median :55.00  Median : 80.00  Median :1.020  Median :0.0000
##  Mean   :51.86  Mean   : 74.93  Mean   :1.019  Mean   :0.8227
##  3rd Qu.:63.00  3rd Qu.: 80.00  3rd Qu.:1.025  3rd Qu.:1.5000
##  Max.   :90.00  Max.   :110.00  Max.   :1.025  Max.   :4.0000
##      su      rbc      pc      pcc
##  Min.   :0.0000  Length:203  Length:203  Length:203
##  1st Qu.:0.0000  Class :character  Class :character  Class :character
##  Median :0.0000  Mode  :character  Mode  :character  Mode  :character
##  Mean   :0.3596
##  3rd Qu.:0.0000
##  Max.   :5.0000
##      ba      bgr      bu      sc
##  Length:203  Min.   : 70.0  Min.   : 10.00  Min.   : 0.400
##  Class :character  1st Qu.: 95.5  1st Qu.: 26.00  1st Qu.: 0.800
##  Mode  :character  Median :117.0  Median : 40.00  Median : 1.100
```

```
##           Mean   :138.0   Mean   : 53.26   Mean   : 2.251
##           3rd Qu.:133.5   3rd Qu.: 51.50   3rd Qu.: 2.200
##           Max.    :490.0   Max.    :309.00   Max.    :15.200
##           sod           pot           hemo           pcv
## Min.    :111.0   Min.    : 2.500   Min.    : 3.10   Min.    : 9.00
## 1st Qu.:135.0   1st Qu.: 3.750   1st Qu.:11.30   1st Qu.:34.00
## Median :139.0   Median : 4.500   Median :13.90   Median :42.00
## Mean    :138.7   Mean    : 4.569   Mean    :13.34   Mean    :40.67
## 3rd Qu.:142.5   3rd Qu.: 4.900   3rd Qu.:15.50   3rd Qu.:48.00
## Max.    :150.0   Max.    :47.000   Max.    :17.80   Max.    :54.00
##           wc           rc           htn           dm
## Min.    : 3800   Min.    :2.100   Length:203   Length:203
## 1st Qu.: 6700   1st Qu.:4.050   Class :character   Class :character
## Median : 8100   Median :4.800   Mode  :character   Mode  :character
## Mean    : 8593   Mean    :4.789
## 3rd Qu.: 9800   3rd Qu.:5.500
## Max.    :26400   Max.    :8.000
##           cad           appet           pe           ane
## Length:203   Length:203   Length:203   Length:203
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
## classification
## Length:203
## Class :character
## Mode  :character
##
##
##
```

## R Markdown

```
#ktd[25] <- ifelse(ktd$classification == "ckd", 1, 0)

b1<-ggplot(ktd, aes(x=classification)) + geom_bar(fill = "#FF6666")+
theme(text = element_text(size = 12))

b2<-ggplot(ktd, aes(x=al)) + geom_bar(fill = "#FF6666")+
theme(text = element_text(size = 12))

b3<-ggplot(ktd, aes(x=su)) + geom_bar(fill = "#FF6666")+
theme(text = element_text(size = 12))

b4<- ggplot(ktd, aes(x=rbc)) + geom_bar(fill = "#FF6666")+
theme(text = element_text(size = 12))

b5<- ggplot(ktd, aes(x=pc)) + geom_bar(fill = "#FF6666")+
theme(text = element_text(size = 12))
```

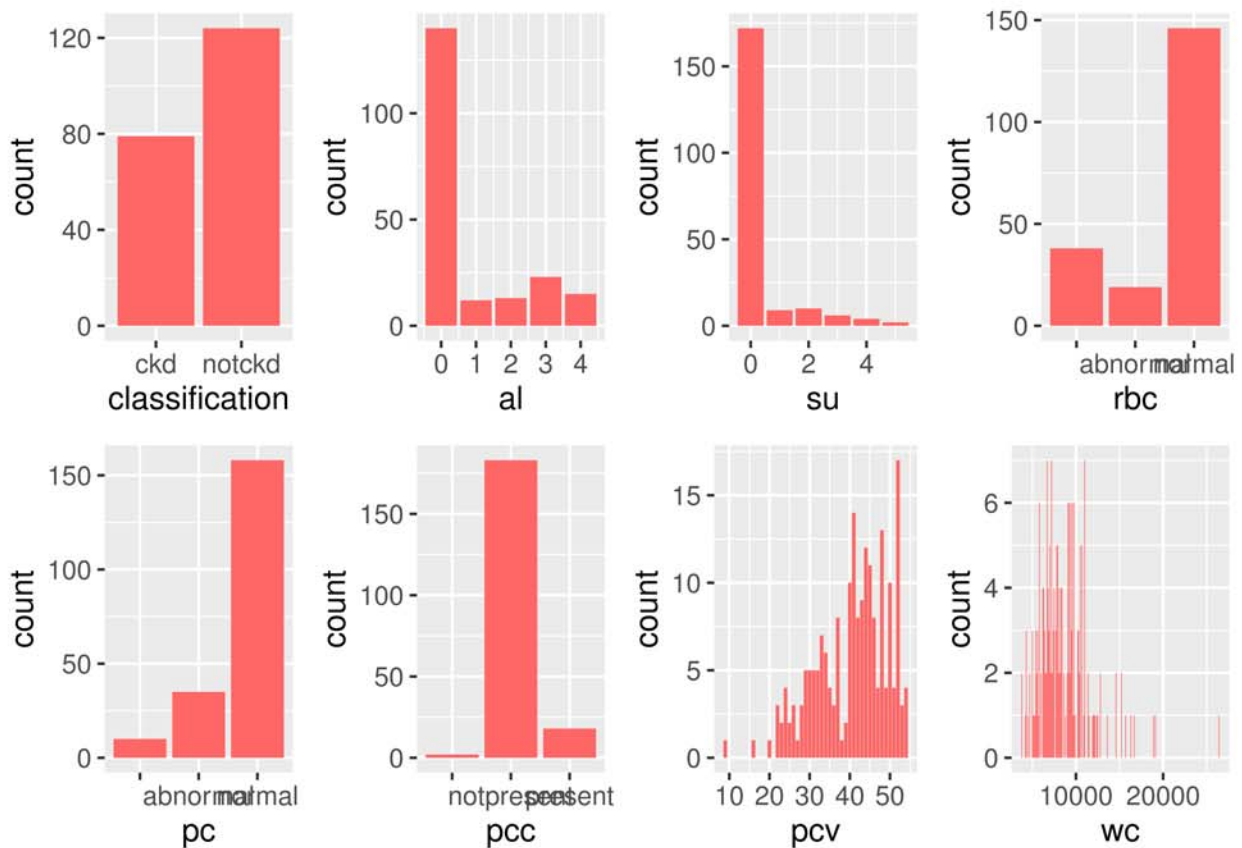
```

b6<-ggplot(ktd, aes(x=pcc)) + geom_bar(fill = "#FF6666")+
  theme(text = element_text(size = 12))

b7<-ggplot(ktd, aes(x=pcv)) + geom_bar(fill = "#FF6666")+
  theme(text = element_text(size = 12))
b8<-ggplot(ktd, aes(x=wc)) + geom_bar(fill = "#FF6666")+
  theme(text = element_text(size = 12))

b <- plot_grid(b1,b2,b3,b4,b5,b6,b7,b8, ncol = 4)
b

```



*# The relationship between age and Chronic Kidney disease*

```

r1<-ggplot(ktd,aes(x=classification, y=age, fill=classification )) +
  geom_boxplot() +
  ggtitle("Kidney disease & age")+
  theme(text = element_text(size = 9))+
  scale_fill_brewer(palette="Set1")

```

```

r2<-ggplot(ktd,aes(x=classification, y=bp, fill=classification )) +
  geom_boxplot() +
  ggtitle("Kidney disease& bp")+
  theme(text = element_text(size = 9))+
  scale_fill_brewer(palette="Set1")

```

```
r3<-ggplot(ktd,aes(x=classification, y=sg, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease & age")+
theme(text = element_text(size = 9))+
scale_fill_brewer(palette="Set1")
```

```
r4<-ggplot(ktd,aes(x=classification, y=al, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease&al")+
theme(text = element_text(size = 9))+
scale_fill_brewer(palette="Set1")
```

```
r5<-ggplot(ktd,aes(x=classification, y=bgr, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease&bgr")+
theme(text = element_text(size = 9))+
scale_fill_brewer(palette="Set1")
```

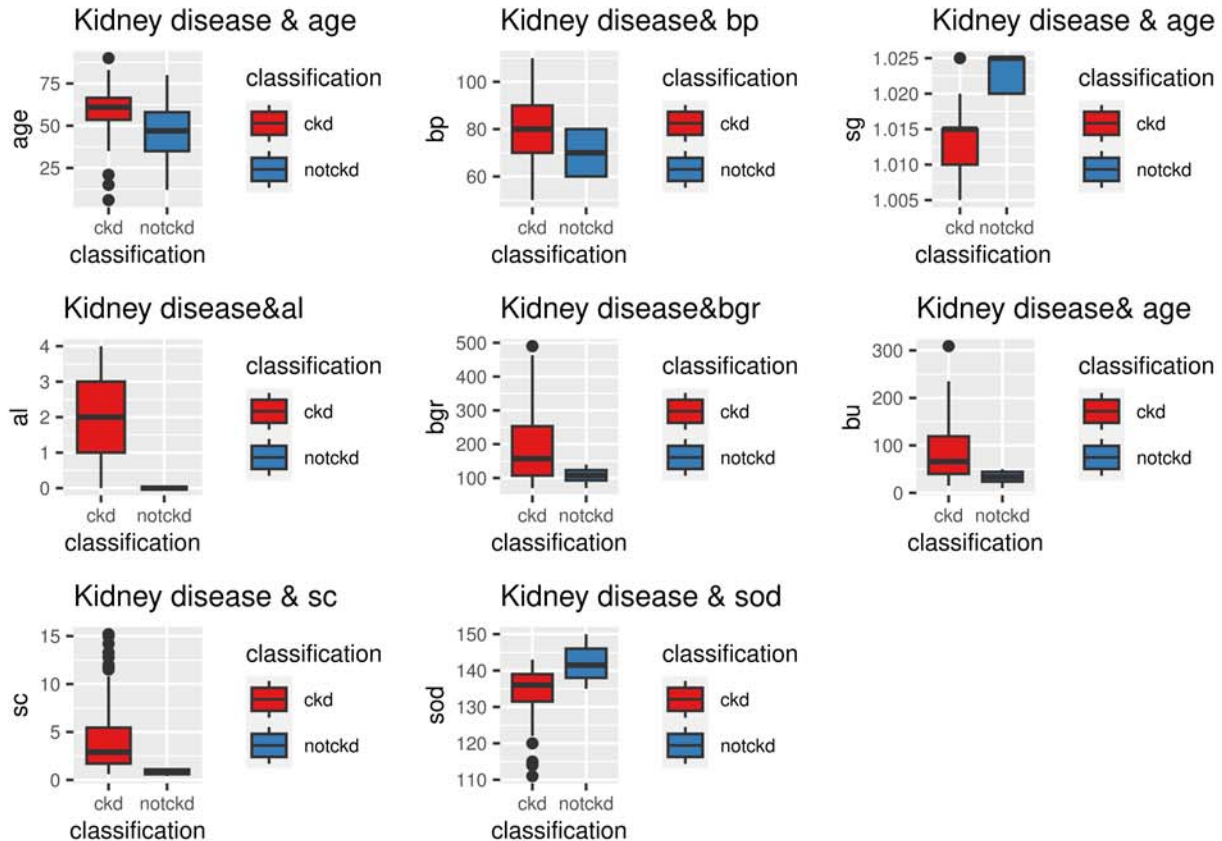
```
r6<-ggplot(ktd,aes(x=classification, y=bu, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease& age")+
theme(text = element_text(size = 9))+
scale_fill_brewer(palette="Set1")
```

```
r7<-ggplot(ktd,aes(x=classification, y=sc, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease & sc")+
theme(text = element_text(size = 9))+
scale_fill_brewer(palette="Set1")
```

```
r8<-ggplot(ktd,aes(x=classification, y=sod, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease & sod")+
theme(text = element_text(size = 9))+
scale_fill_brewer(palette="Set1")
```

```
r<- plot_grid(r1,r2,r3,r4,r5,r6,r7,r8, ncol = 3)
r
```





```
# The relationship between age and Chronic Kidney disease
u1<-ggplot(ktd,aes(x=classification, y=pot, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease &pot")+
theme(text = element_text(size = 8))+
scale_fill_brewer(palette="Set1")

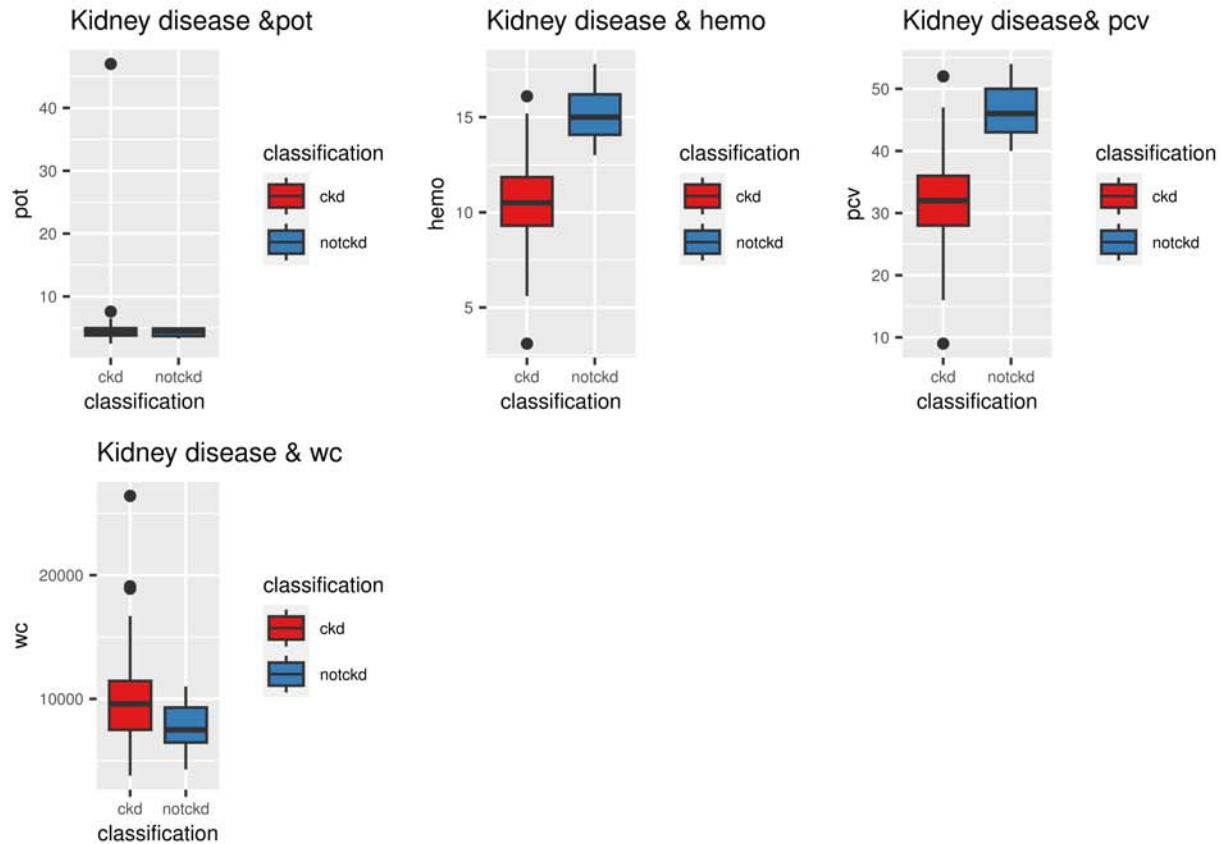
u2<-ggplot(ktd,aes(x=classification, y=hemo, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease & hemo")+
theme(text = element_text(size = 8))+
scale_fill_brewer(palette="Set1")

u3<-ggplot(ktd,aes(x=classification, y=pcv, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease& pcv")+
theme(text = element_text(size = 8))+
scale_fill_brewer(palette="Set1")

u4<-ggplot(ktd,aes(x=classification, y=wc, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease & wc")+
theme(text = element_text(size = 8))+
```

```
scale_fill_brewer(palette="Set1")

b <- plot_grid(u1,u2,u3,u4, ncol = 3)
b
```



```
#it shows outliers
a1<-ggplot(data = ktd, aes(y= wc))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")

a2<-ggplot(data = ktd, aes(y= pcv))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")

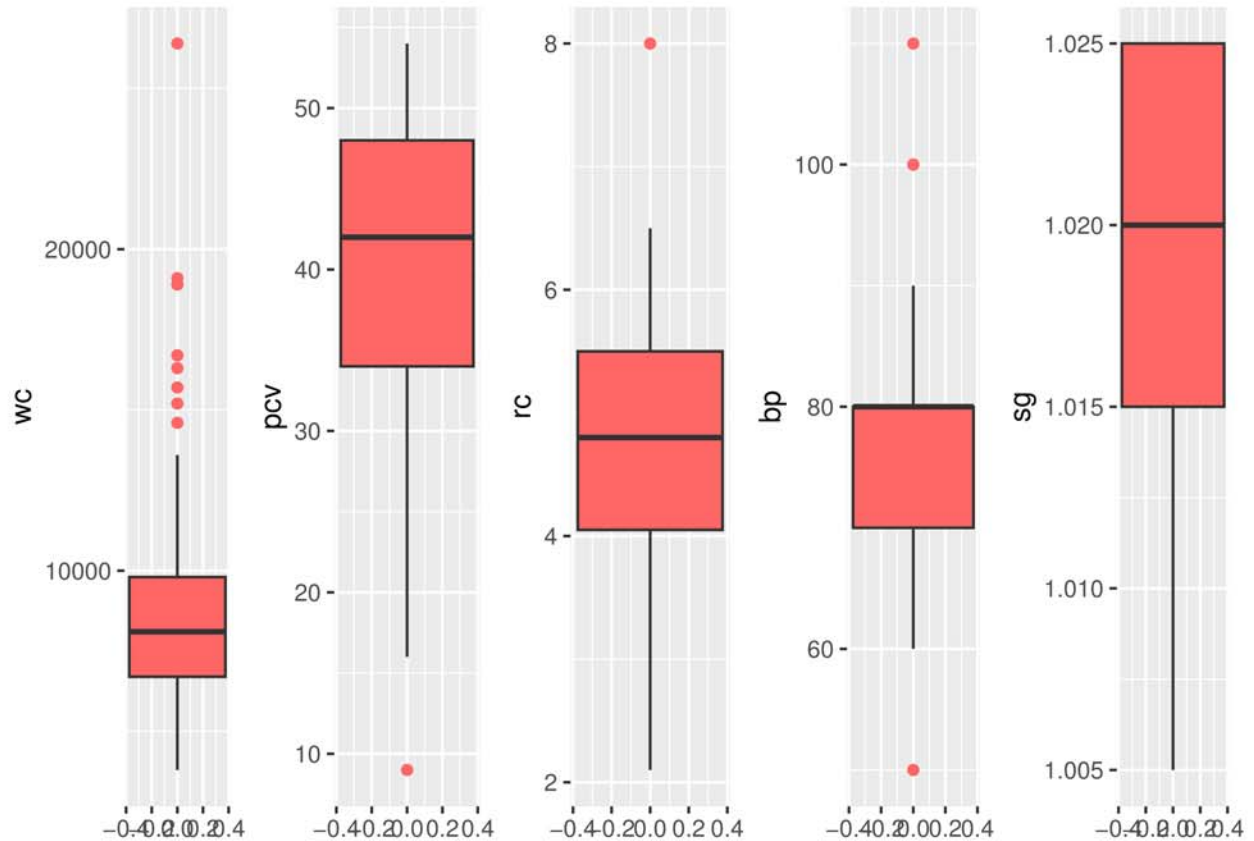
#ktd[18]<-scale(ktd$rc)

a3<-ggplot(data = ktd, aes(y= rc))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")

a4<-ggplot(data = ktd, aes(y= bp))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")

a5<-ggplot(data = ktd, aes(y= sg))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")
```

```
b <- plot_grid(a1,a2,a3,a4,a5, ncol =5 )
b
```



```
#view(ktd)
```

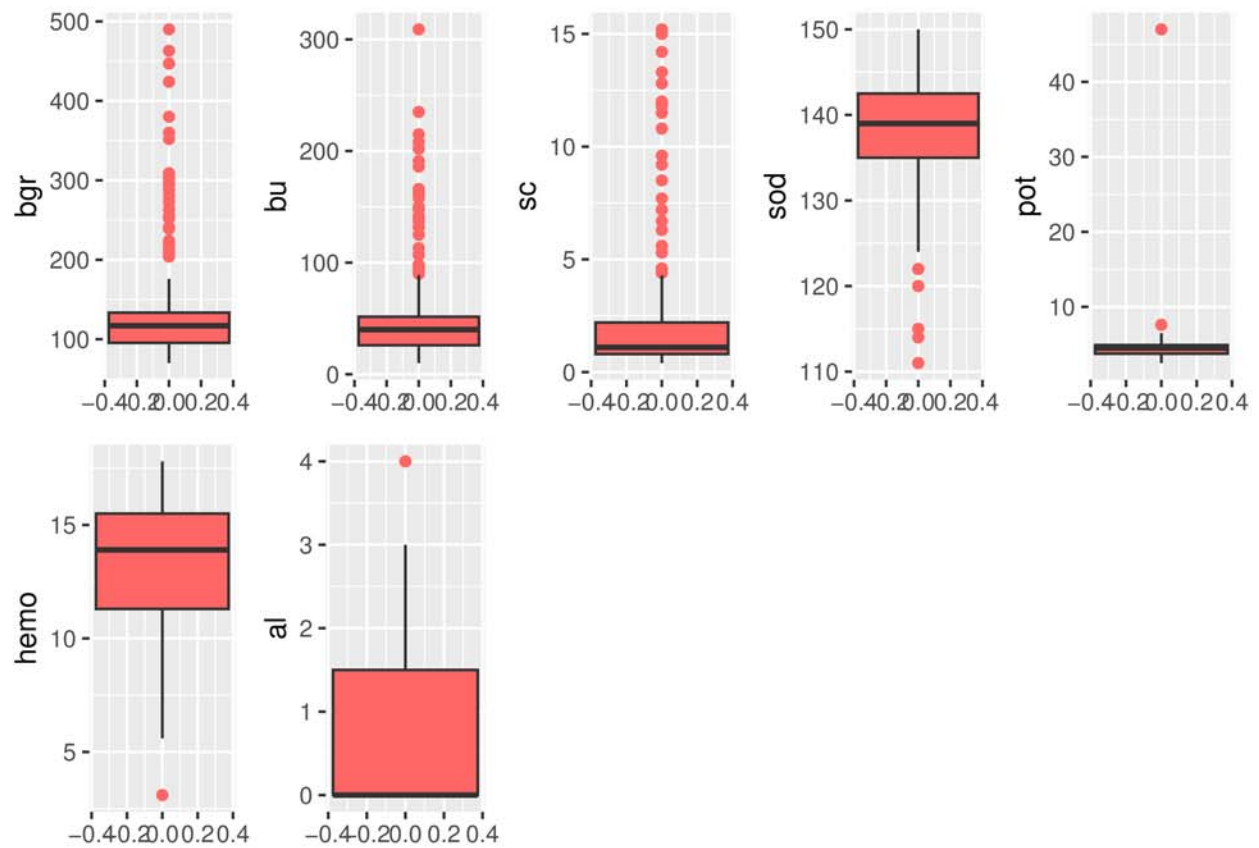
```
#it shows outlier
```

```
a6<-ggplot(data = ktd, aes(y= bgr))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")

a7<-ggplot(data = ktd, aes(y= bu))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")
a8<-ggplot(data = ktd, aes(y= sc))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")
a9<-ggplot(data = ktd, aes(y= sod))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")
a10<-ggplot(data = ktd, aes(y= pot))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")

a11<-ggplot(data = ktd, aes(y= hemo))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")
a12<-ggplot(data = ktd, aes(y= al))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#FF6666")
```

```
b <- plot_grid(a6,a7,a8,a9,a10,a11,a12, ncol =5 )
b
```

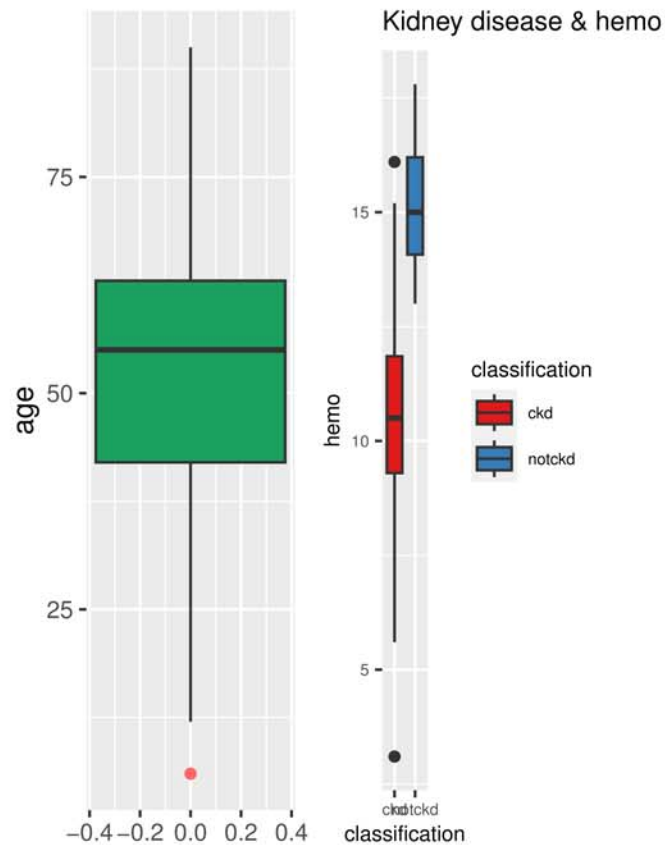


```
u0 <- ggplot(data = ktd, aes(y=pcc ))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#1CA160")

u1 <- ggplot(data = ktd, aes(y=age ))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#1CA160")

b <- plot_grid(u1,u2, ncol =4 )
b
```





*#handling outliers, Dataset is changed manually to replace outliers with upper or lower interquantile*

```
Q1 <- quantile(ktd$wc, .25)
Q3 <- quantile(ktd$wc, .75)
IQR <- IQR(ktd$wc)
```

```
# Lower bound Quantile Range:
l = Q1-1.5*IQR
# Upper bound Quantile Range:
u = Q3+1.5*IQR
#lower
l
```

```
## 25%
## 2050
```

```
#upper
u
```

```
## 75%
## 14450
```

```
#find Q1, Q3, and interquartile range for values in column A
Q1 <- quantile(ktd$bgr, .25)
Q3 <- quantile(ktd$bgr, .75)
IQR <- IQR(ktd$bgr)
```

```
# Lower bound Quantile Range:
l = Q1-1.5*IQR
# Upper bound Quantile Range:
u = Q3+1.5*IQR
#lower
l
```

```
## 25%
## 38.5
```

```
#upper
u
```

```
## 75%
## 190.5
```

```
#only keep rows in dataframe that have values within 1.5*IQR of Q1 and Q3
no_outliers <- subset(ktd, ktd$bgr > (Q1 - 1.5*IQR) & ktd$bgr < (Q3 + 1.5*IQR))
```

```
#view row and column count of new data frame
#dim(no_outliers)
Q1 <- quantile(ktd$pcv, .25)
Q3 <- quantile(ktd$pcv, .75)
IQR <- IQR(ktd$pcv)
```

```
# Lower bound Quantile Range:
l = Q1-1.5*IQR
# Upper bound Quantile Range:
u = Q3+1.5*IQR
#lower
l
```

```
## 25%
## 13
```

```
#upper
u
```

```
## 75%
## 69
```

```
Q1 <- quantile(ktd$bu, .25)
Q3 <- quantile(ktd$bu, .75)
IQR <- IQR(ktd$bu)
```

```

# Lower bound Quantile Range:
l = Q1-1.5*IQR
# Upper bound Quantile Range:
u = Q3+1.5*IQR
#lower
l

```

```

## 25%
## -12.25

```

```

#upper
u

```

```

## 75%
## 89.75

```

```

Q1 <- quantile(ktd$bp, .25)
Q3 <- quantile(ktd$bp, .75)
IQR <- IQR(ktd$bp)

```

```

# Lower bound Quantile Range:
l = Q1-1.5*IQR
# Upper bound Quantile Range:
u = Q3+1.5*IQR
#lower
l

```

```

## 25%
## 55

```

```

#upper
u

```

```

## 75%
## 95

```

```

Q1 <- quantile(ktd$sc, .25)
Q3 <- quantile(ktd$sc, .75)
IQR <- IQR(ktd$sc)

```

```

# Lower bound Quantile Range:
l = Q1-1.5*IQR
# Upper bound Quantile Range:
u = Q3+1.5*IQR
#lower
l

```

```

## 25%
## -1.3

```

```
#upper  
u
```

```
## 75%  
## 4.3
```

```
Q1 <- quantile(ktd$pot, .25)  
Q3 <- quantile(ktd$pot, .75)  
IQR <- IQR(ktd$pot)
```

```
# Lower bound Quantile Range:  
l = Q1-1.5*IQR  
# Upper bound Quantile Range:  
u = Q3+1.5*IQR  
#lower  
l
```

```
## 25%  
## 2.025
```

```
#upper  
u
```

```
## 75%  
## 6.625
```

```
Q1 <- quantile(ktd$hemo, .25)  
Q3 <- quantile(ktd$hemo, .75)  
IQR <- IQR(ktd$hemo)
```

```
# Lower bound Quantile Range:  
l = Q1-1.5*IQR  
# Upper bound Quantile Range:  
u = Q3+1.5*IQR  
#lower  
l
```

```
## 25%  
## 5
```

```
#upper  
u
```

```
## 75%  
## 21.8
```



```
Q1 <- quantile(ktd$al, .25)
Q3 <- quantile(ktd$al, .75)
IQR <- IQR(ktd$al)
```

```
# Lower bound Quantile Range:
l = Q1-1.5*IQR
# Upper bound Quantile Range:
u = Q3+1.5*IQR
#lower
l
```

```
## 25%
## -2.25
```

```
#upper
u
```

```
## 75%
## 3.75
```

```
kdn[9]<- ifelse(kdn$ba == "present", 1, 0)
kdn[20]<- ifelse(kdn$dm == "yes", 1, 0)
kdn[21] <- ifelse(kdn$cad == "yes", 1, 0)

kdn[6] <- ifelse(kdn$rbc == "normal", 1, 0)
kdn[7] <- ifelse(kdn$pc == "normal", 0, 1) #####changes

kdn[8] <- ifelse(kdn$pcc == "present", 1, 0)

kdn[22] <- ifelse(kdn$appet == "good", 1, 0)
kdn[23] <- ifelse(kdn$pe== "yes", 1, 0)
kdn[24] <- ifelse(kdn$sane == "yes", 1, 0)
kdn[19] <- ifelse(kdn$htn == "yes", 0, 1) #####change

kdn[25] <- ifelse(kdn$classification == "ckd", 1, 0)
```

```
#correlation
```

```
set.seed(7)
# load the library
library(mlbench)
```

```
## Warning: package 'mlbench' was built under R version 4.2.3
```

```
library(caret)
# load the data
#data(PimaIndiansDiabetes)
# calculate correlation matrix
correlationMatrix <- cor(kdn[,1:25])
```

```
#correlationMatrix2 <- cor(ktd[,10:18])
```

```
# summarize the correlation matrix
print(correlationMatrix)
```

```
##          age          bp          sg          al          su
## age      1.000000000  0.14935044 -0.29953914  0.2073196  0.22975609
## bp       0.149350444  1.00000000 -0.26148915  0.2768790  0.26076293
## sg      -0.299539136 -0.26148915  1.00000000 -0.6090625 -0.39243756
## al       0.207319587  0.27687899 -0.60906254  1.0000000  0.42308406
## su       0.229756093  0.26076293 -0.39243756  0.4230841  1.00000000
## rbc     -0.267773017 -0.30412201  0.49685398 -0.3240833 -0.24315965
## pc       0.164502565  0.15288422 -0.47757378  0.6204952  0.24245790
## pcc      0.136131934  0.19779929 -0.36221968  0.5006244  0.11668496
## ba       0.046240241  0.15032585 -0.40200805  0.4643083  0.23507779
## bgr      0.282864108  0.21307714 -0.46496179  0.4364680  0.79233295
## bu       0.162911904  0.26090151 -0.48308076  0.6049761  0.21250453
## sc       0.155684446  0.34078646 -0.50617996  0.6298400  0.23178787
## sod     -0.126011461 -0.19877538  0.51199202 -0.5603697 -0.27343497
## pot      0.003477226  0.09473102 -0.04889704  0.1906210  0.18717097
## hemo    -0.238897609 -0.30227715  0.68176469 -0.7137224 -0.29621949
## pcv     -0.254352041 -0.33091901  0.67422251 -0.6940612 -0.32261505
## wc       0.172883519  0.07542072 -0.26854740  0.2584450  0.21865487
## rc      -0.243763314 -0.23517235  0.61108364 -0.5932926 -0.28578300
## htn     -0.377857545 -0.33884786  0.58436514 -0.6674557 -0.54598491
## dm       0.353267592  0.22935900 -0.59127233  0.5508653  0.55067158
## cad      0.228813984  0.19254889 -0.29097682  0.3481802  0.40070040
## appet   -0.194102310 -0.08948144  0.43912771 -0.4734967 -0.11688235
## pe       0.212774284  0.05925116 -0.51278789  0.5430750  0.21386052
## ane      0.041467469  0.26226278 -0.32982537  0.4736358  0.09091403
## classification 0.360866364  0.38711217 -0.77886443  0.7597193  0.46477340
##          rbc          pc          pcc          ba          bgr
## age     -0.26777302  0.1645026  0.13613193  0.04624024  0.28286411
## bp      -0.30412201  0.1528842  0.19779929  0.15032585  0.21307714
## sg       0.49685398 -0.4775738 -0.36221968 -0.40200805 -0.46496179
## al      -0.32408328  0.6204952  0.50062443  0.46430832  0.43646796
## su      -0.24315965  0.2424579  0.11668496  0.23507779  0.79233295
## rbc      1.00000000 -0.3790870 -0.11359657 -0.10520976 -0.36723735
## pc      -0.37908698  1.0000000  0.50104506  0.39325471  0.30474278
## pcc     -0.11359657  0.5010451  1.00000000  0.41388542  0.22313756
## ba      -0.10520976  0.3932547  0.41388542  1.00000000  0.19552213
## bgr     -0.36723735  0.3047428  0.22313756  0.19552213  1.00000000
## bu      -0.25002077  0.4918097  0.30529303  0.18714397  0.24278216
## sc      -0.28246216  0.4542458  0.29960301  0.21123962  0.22376479
## sod      0.23611898 -0.3853658 -0.39925925 -0.19627893 -0.30966035
## pot      0.04676506  0.1232835 -0.01036711  0.01244311  0.06251438
## hemo     0.45934608 -0.5505358 -0.45250716 -0.33658775 -0.34713324
## pcv      0.46137393 -0.5496542 -0.45112346 -0.32438105 -0.36225595
## wc      -0.09505660  0.1349710  0.14856039  0.15777000  0.20573462
## rc       0.39018026 -0.4937920 -0.42670900 -0.28710274 -0.32027348
## htn      0.46916532 -0.3897241 -0.29648473 -0.18707902 -0.54414611
## dm      -0.51227272  0.3932121  0.21394644  0.21526316  0.56680158
```

## cad	-0.22405114	0.2400115	0.29463208	0.22218138	0.39834920
## appet	0.29049056	-0.3368089	-0.37777601	-0.12877498	-0.23626141
## pe	-0.20720124	0.4196032	0.28607633	0.31231805	0.21313636
## ane	-0.14770181	0.3639043	0.40616353	0.10935672	0.13865356
## classification	-0.67039757	0.5469928	0.39079382	0.32771187	0.51553296
##	bu	sc	sod	pot	hemo
## age	0.16291190	0.1556844	-0.12601146	0.003477226	-0.2388976
## bp	0.26090151	0.3407865	-0.19877538	0.094731021	-0.3022771
## sg	-0.48308076	-0.5061800	0.51199202	-0.048897041	0.6817647
## al	0.60497614	0.6298400	-0.56036967	0.190621034	-0.7137224
## su	0.21250453	0.2317879	-0.27343497	0.187170974	-0.2962195
## rbc	-0.25002077	-0.2824622	0.23611898	0.046765061	0.4593461
## pc	0.49180969	0.4542458	-0.38536583	0.123283454	-0.5505358
## pcc	0.30529303	0.2996030	-0.39925925	-0.010367106	-0.4525072
## ba	0.18714397	0.2112396	-0.19627893	0.012443112	-0.3365878
## bgr	0.24278216	0.2237648	-0.30966035	0.062514381	-0.3471332
## bu	1.00000000	0.8594977	-0.47538661	0.222661490	-0.6668152
## sc	0.85949770	1.00000000	-0.49462098	0.127120957	-0.6636695
## sod	-0.47538661	-0.4946210	1.00000000	-0.049011674	0.5410705
## pot	0.22266149	0.1271210	-0.04901167	1.000000000	-0.1614173
## hemo	-0.66681516	-0.6636695	0.54107053	-0.161417286	1.00000000
## pcv	-0.65443539	-0.6583689	0.53637210	-0.188458580	0.8655518
## wc	0.09133445	0.0785811	-0.18028681	-0.104277873	-0.3019262
## rc	-0.58867887	-0.5899689	0.44065284	-0.175752401	0.7687802
## htn	-0.49158910	-0.5019193	0.47522573	-0.137574907	0.6875897
## dm	0.45989034	0.4270302	-0.43215564	0.134242798	-0.5878093
## cad	0.26582488	0.2507298	-0.22112424	0.024331790	-0.3506681
## appet	-0.41263187	-0.3831177	0.42262114	-0.004033118	0.5339970
## pe	0.49311671	0.4979533	-0.40660753	0.013846493	-0.5168700
## ane	0.55499785	0.5618905	-0.45982680	0.219414583	-0.5753838
## classification	0.55426204	0.5880926	-0.55720766	0.098974697	-0.8151133
##	pcv	wc	rc	htn	dm
## age	-0.2543520	0.17288352	-0.2437633	-0.3778575	0.3532676
## bp	-0.3309190	0.07542072	-0.2351724	-0.3388479	0.2293590
## sg	0.6742225	-0.26854740	0.6110836	0.5843651	-0.5912723
## al	-0.6940612	0.25844500	-0.5932926	-0.6674557	0.5508653
## su	-0.3226150	0.21865487	-0.2857830	-0.5459849	0.5506716
## rbc	0.4613739	-0.09505660	0.3901803	0.4691653	-0.5122727
## pc	-0.5496542	0.13497102	-0.4937920	-0.3897241	0.3932121
## pcc	-0.4511235	0.14856039	-0.4267090	-0.2964847	0.2139464
## ba	-0.3243811	0.15777000	-0.2871027	-0.1870790	0.2152632
## bgr	-0.3622559	0.20573462	-0.3202735	-0.5441461	0.5668016
## bu	-0.6544354	0.09133445	-0.5886789	-0.4915891	0.4598903
## sc	-0.6583689	0.07858110	-0.5899689	-0.5019193	0.4270302
## sod	0.5363721	-0.18028681	0.4406528	0.4752257	-0.4321556
## pot	-0.1884586	-0.10427787	-0.1757524	-0.1375749	0.1342428
## hemo	0.8655518	-0.30192617	0.7687802	0.6875897	-0.5878093
## pcv	1.0000000	-0.30223072	0.7665879	0.6858457	-0.5935414
## wc	-0.3022307	1.00000000	-0.2339079	-0.1989072	0.2526941
## rc	0.7665879	-0.23390789	1.0000000	0.6335552	-0.5661479
## htn	0.6858457	-0.19890720	0.6335552	1.0000000	-0.7428080
## dm	-0.5935414	0.25269409	-0.5661479	-0.7428080	1.0000000
## cad	-0.3562755	0.04707028	-0.3517483	-0.4569771	0.3728097
## appet	0.5476970	-0.25688011	0.4817237	0.4675952	-0.3870829



```
## pe -0.5257780 0.22206325 -0.5067597 -0.5160550 0.5345088
## ane -0.5698999 0.08395135 -0.5127225 -0.4594276 0.1712251
## classification -0.8057846 0.35494930 -0.7121427 -0.8019403 0.7352085
## cad appet pe ane classification
## age 0.22881398 -0.194102310 0.21277428 0.04146747 0.3608664
## bp 0.19254889 -0.089481445 0.05925116 0.26226278 0.3871122
## sg -0.29097682 0.439127713 -0.51278789 -0.32982537 -0.7788644
## al 0.34818021 -0.473496722 0.54307503 0.47363585 0.7597193
## su 0.40070040 -0.116882353 0.21386052 0.09091403 0.4647734
## rbc -0.22405114 0.290490556 -0.20720124 -0.14770181 -0.6703976
## pc 0.24001149 -0.336808885 0.41960319 0.36390430 0.5469928
## pcc 0.29463208 -0.377776011 0.28607633 0.40616353 0.3907938
## ba 0.22218138 -0.128774982 0.31231805 0.10935672 0.3277119
## bgr 0.39834920 -0.236261407 0.21313636 0.13865356 0.5155330
## bu 0.26582488 -0.412631873 0.49311671 0.55499785 0.5542620
## sc 0.25072982 -0.383117705 0.49795329 0.56189055 0.5880926
## sod -0.22112424 0.422621141 -0.40660753 -0.45982680 -0.5572077
## pot 0.02433179 -0.004033118 0.01384649 0.21941458 0.0989747
## hemo -0.35066812 0.533997010 -0.51686998 -0.57538378 -0.8151133
## pcv -0.35627550 0.547696968 -0.52577800 -0.56989990 -0.8057846
## wc 0.04707028 -0.256880108 0.22206325 0.08395135 0.3549493
## rc -0.35174833 0.481723690 -0.50675966 -0.51272246 -0.7121427
## htn -0.45697709 0.467595172 -0.51605495 -0.45942757 -0.8019403
## dm 0.37280971 -0.387082890 0.53450876 0.17122507 0.7352085
## cad 1.00000000 -0.148082666 0.26228731 0.14076051 0.3664684
## appet -0.14808267 1.000000000 -0.55849915 -0.38012158 -0.4718375
## pe 0.26228731 -0.558499146 1.00000000 0.24802270 0.4907077
## ane 0.14076051 -0.380121576 0.24802270 1.00000000 0.4255702
## classification 0.36646836 -0.471837542 0.49070767 0.42557024 1.0000000
```

```
#print(correlationMatrix2)
# find attributes that are highly correlated (ideally >0.75)
highlyCorrelated <- findCorrelation(correlationMatrix, cutoff=0.7)
# print indexes of highly correlated attributes
print(highlyCorrelated)
```

```
## [1] 25 16 15 19 12 10
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.