

# ProjectML

```
setwd("C:/Users/Administrator/Documents/archive (3)")
ktd <- read.csv("kdd3.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)
```

```
kdn<-ktd
```

```
summary(ktd)
```

```
##      age      bp      sg      al
##  Min.   : 6.00  Min.   :55.00  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   :74.61 Mean   :1.019 Mean   :0.8042
## 3rd Qu.:63.00 3rd Qu.:80.00 3rd Qu.:1.025 3rd Qu.:1.5000
## Max.   :90.00 Max.   :95.00 Max.   :1.025 Max.   :3.7500
##      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   :122.4   Mean   :44.87   Mean   :1.667
##           3rd Qu.:133.5   3rd Qu.:51.50   3rd Qu.:2.200
##           Max.   :190.5   Max.   :89.75   Max.   :4.300
##           sod           pot           hemo           pcv
## Min.   :111.0   Min.   :2.500   Min.   : 5.00   Min.   :13.00
## 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   :4.365   Mean   :13.35   Mean   :40.69
## 3rd Qu.:142.5   3rd Qu.:4.900   3rd Qu.:15.50   3rd Qu.:48.00
## Max.   :150.0   Max.   :6.625   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   : 8454   Mean   :4.789
## 3rd Qu.: 9800   3rd Qu.:5.500
## Max.   :14450   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))
```

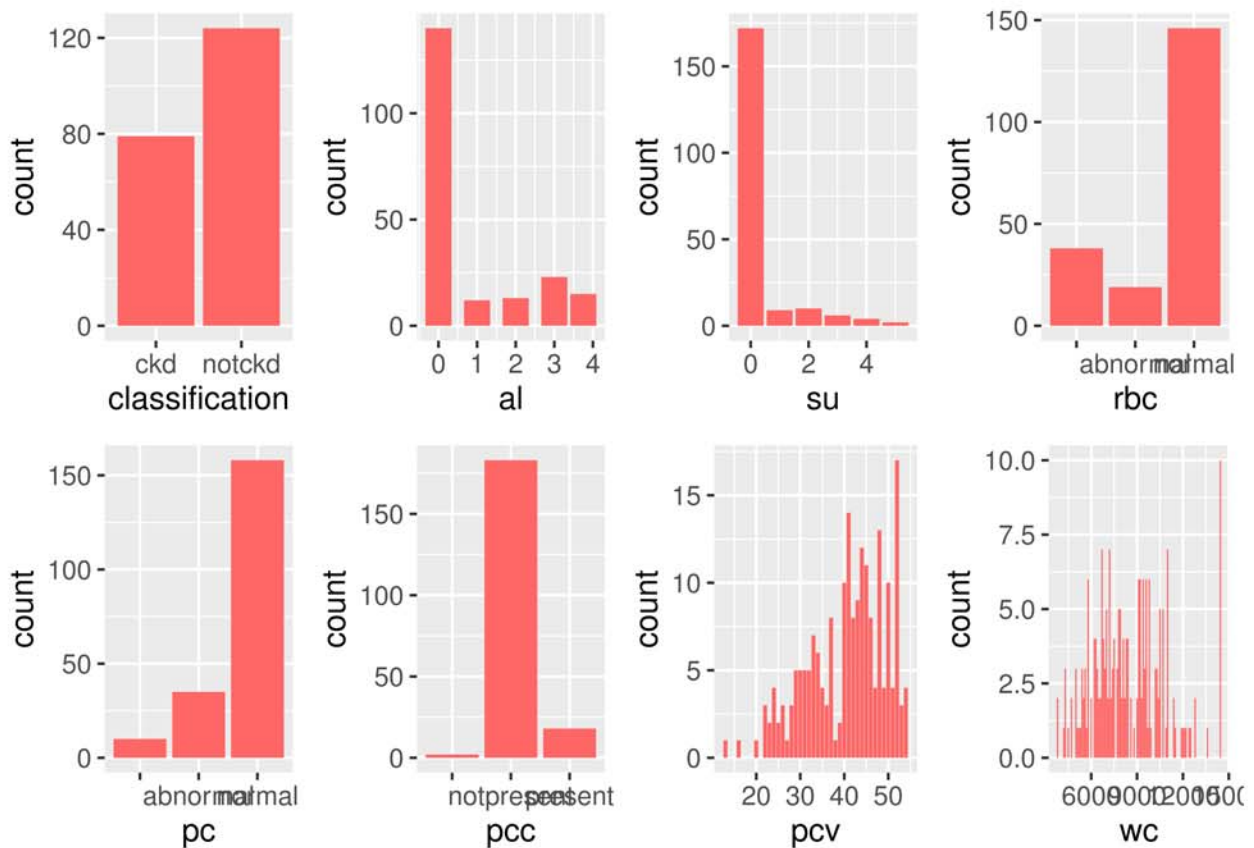
```

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")
```

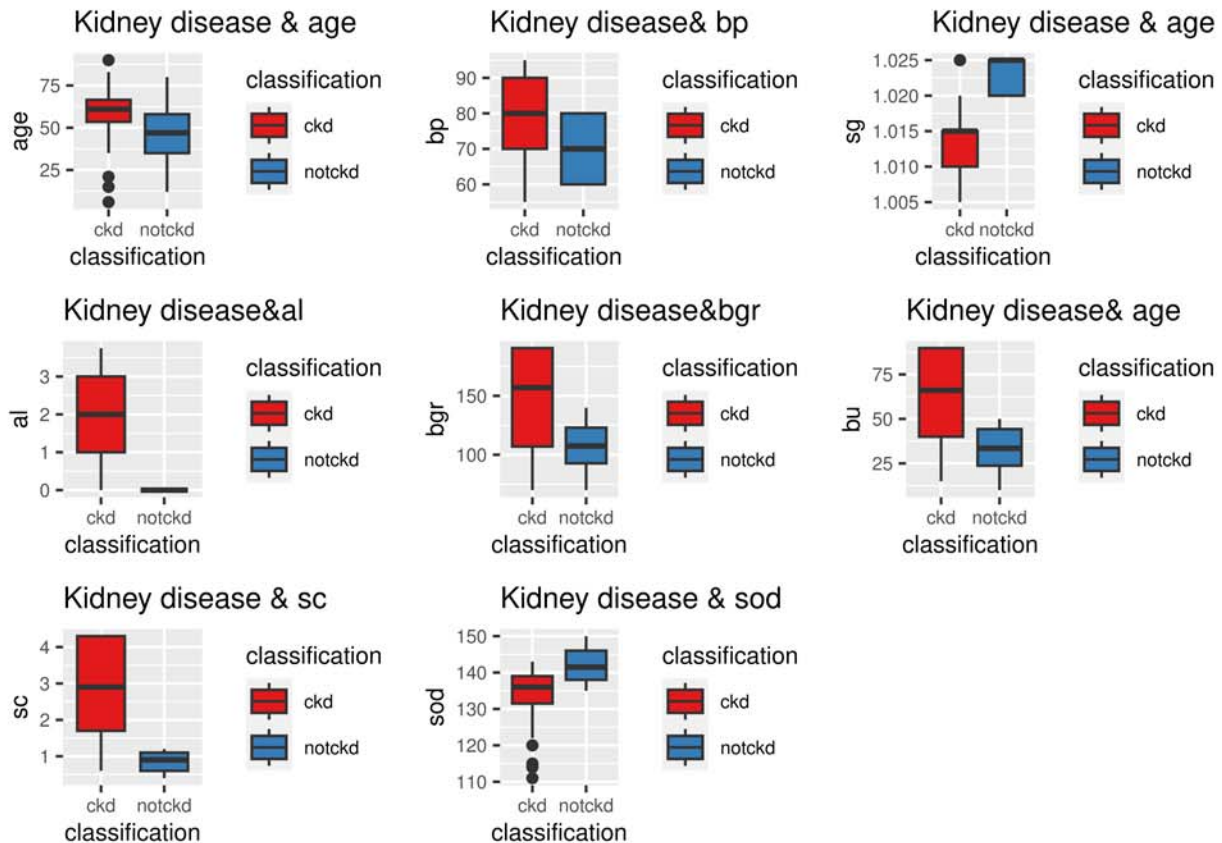
```
r6<-ggplot(ktd,aes(x=classification, y=bu, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease& age")+
theme(text = element_text(size = 9))+
scale_fill_brewer(palette="Set1")
```

```
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")

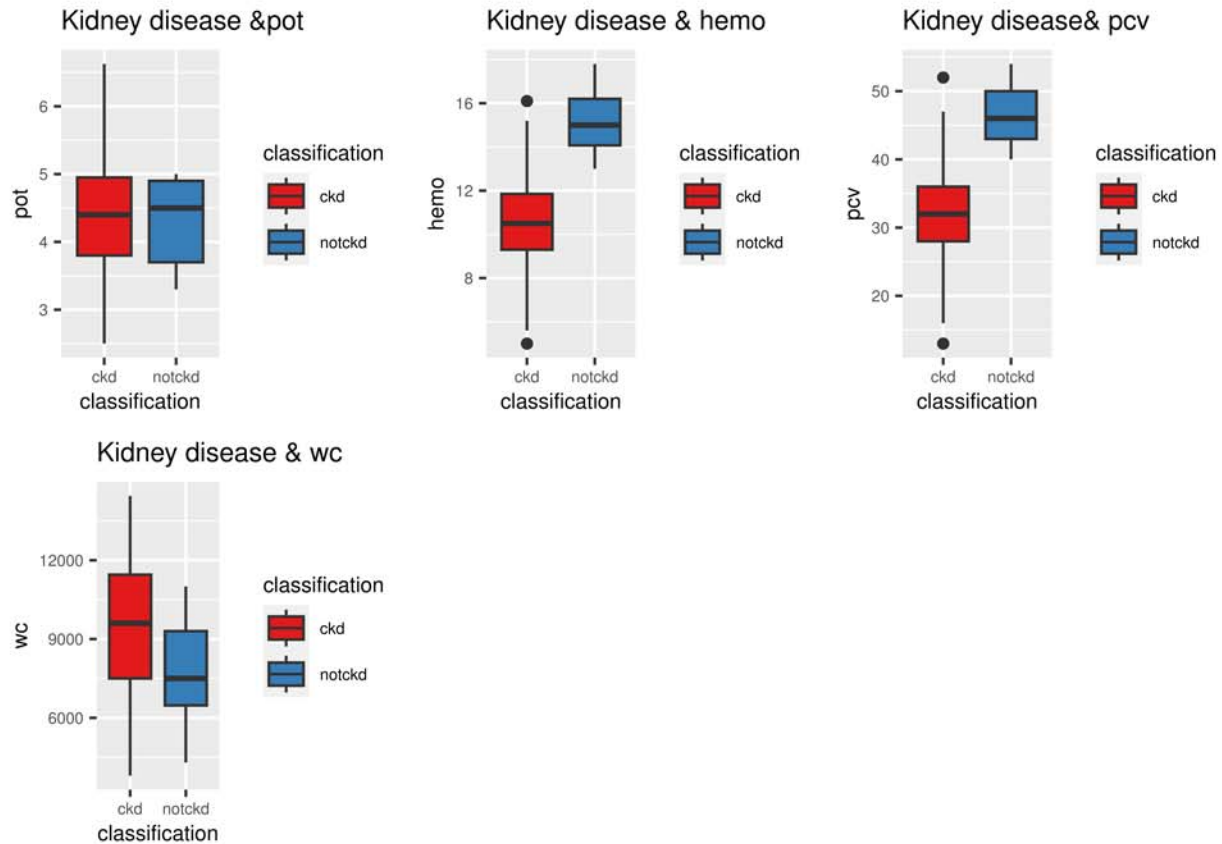
u2<-ggplot(ktd,aes(x=classification, y=hemo, fill=classification )) +
geom_boxplot() +
ggtitle("Kidney disease & hemo")+
theme(text = element_text(size = 8))+
scale_fill_brewer(palette="Set1")

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")

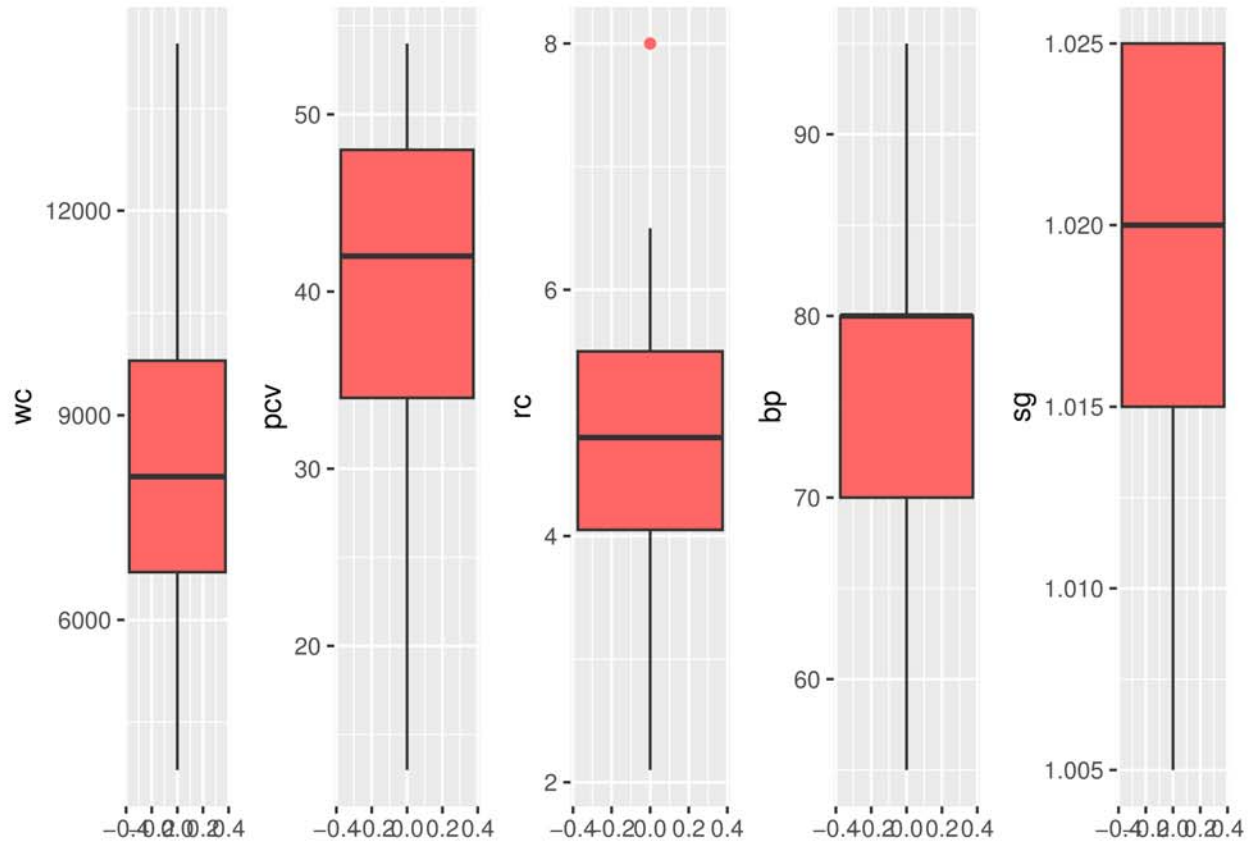
#ktd[18]<-scale(ktd$rc)

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)
```

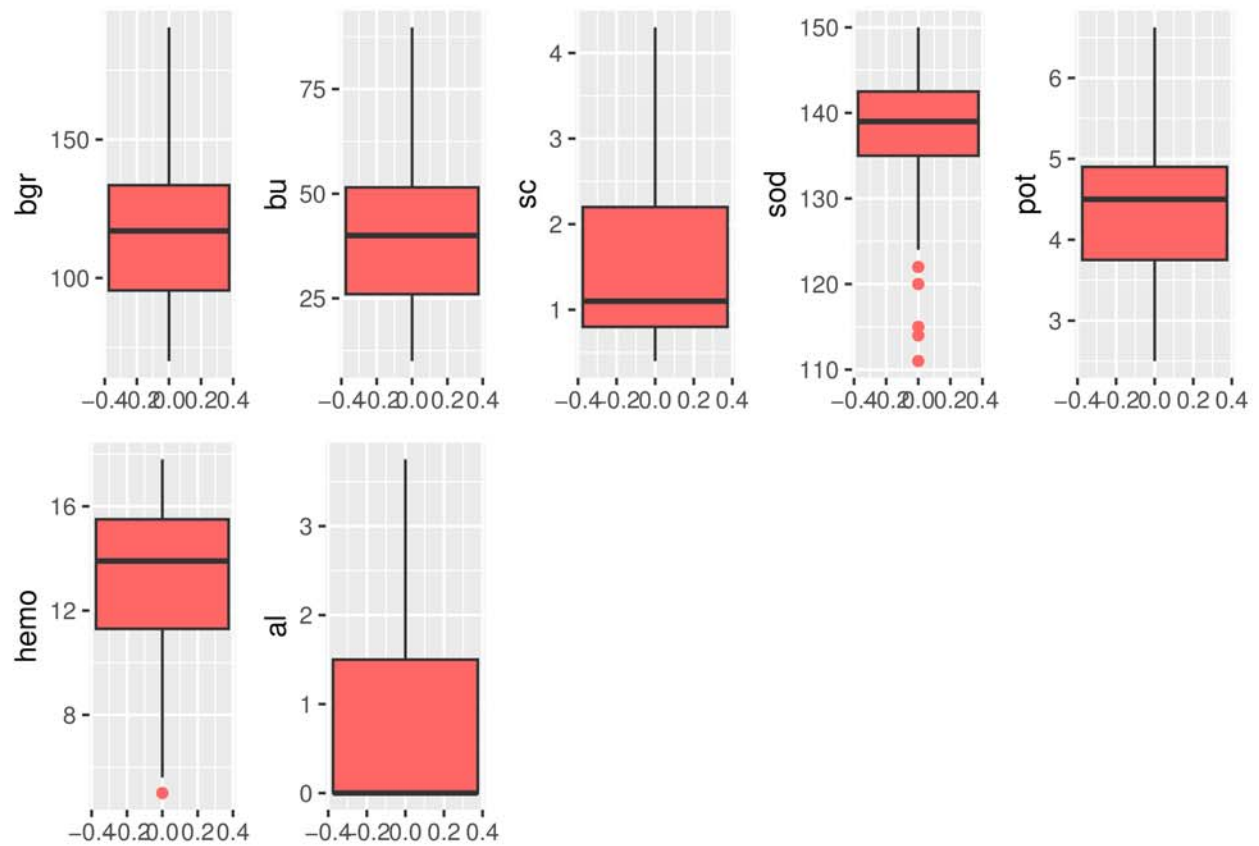
```
#it shows outlier
```

```
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
```

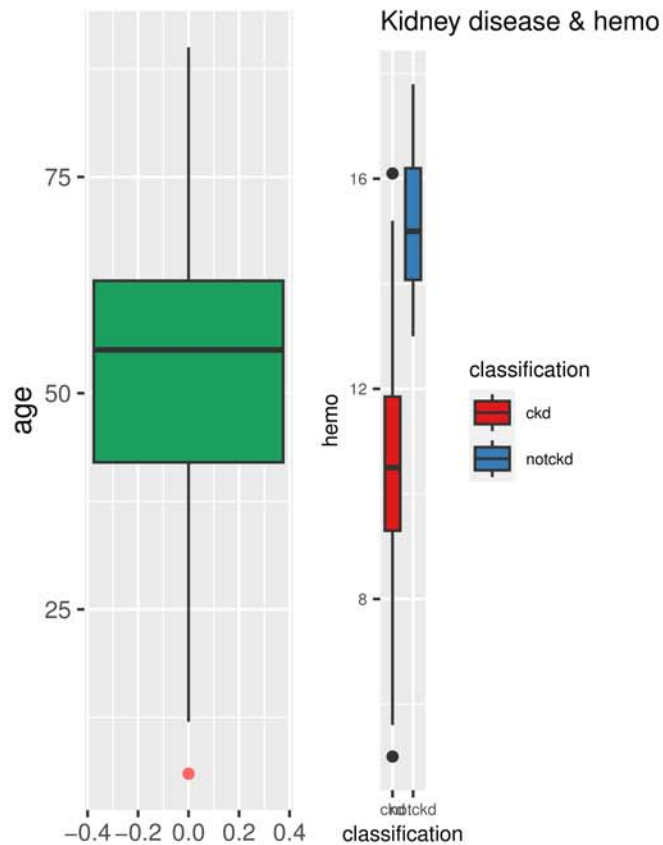


```
u0 <- ggplot(data = ktd, aes(y=pcc ))+
  geom_boxplot(outlier.color = "#FF6666", fill = "#1CA160")

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  0.14960165 -0.2995391364  0.2132106  0.22975609
## bp       0.149601647  1.000000000 -0.2572469667  0.2624886  0.25071749
## sg      -0.299539136 -0.25724697  1.00000000000 -0.6136142 -0.39243756
## al       0.213210611  0.26248860 -0.6136141709  1.0000000  0.42743124
## su       0.229756093  0.25071749 -0.3924375566  0.4274312  1.00000000
## rbc     -0.267773017 -0.29912666  0.4968539830 -0.3328825 -0.24315965
## pc       0.164502565  0.16347208 -0.4775737812  0.6254313  0.24245790
## pcc      0.136131934  0.18932545 -0.3622196837  0.5079547  0.11668496
## ba       0.046240241  0.15042067 -0.4020080520  0.4599817  0.23507779
## bgr      0.322211480  0.23114231 -0.4758074767  0.4730501  0.66136266
## bu       0.282469738  0.22080254 -0.5373770081  0.6527854  0.30720824
## sc       0.321719794  0.36934702 -0.6378669160  0.7803694  0.36525085
## sod     -0.126011461 -0.19664334  0.5119920192 -0.5637894 -0.27343497
## pot      0.008931901  0.03407559 -0.0001720557  0.1489699  0.04829191
## hemo    -0.240655099 -0.28812057  0.6839425631 -0.7190514 -0.29728895
## pcv     -0.255653646 -0.31971041  0.6756654838 -0.6989695 -0.32354476
## wc       0.198456159  0.05844127 -0.2921034415  0.2275768  0.23185209
## rc      -0.243763314 -0.22726627  0.6110836441 -0.6020627 -0.28578300
## htn     -0.377857545 -0.31709070  0.5843651361 -0.6752590 -0.54598491
## dm       0.353267592  0.22690298 -0.5912723260  0.5575439  0.55067158
## cad      0.228813984  0.15574486 -0.2909768206  0.3565557  0.40070040
## appet   -0.194102310 -0.08801784  0.4391277131 -0.4780718 -0.11688235
## pe       0.212774284  0.06177449 -0.5127878936  0.5440043  0.21386052
## ane      0.041467469  0.24641924 -0.3298253721  0.4721341  0.09091403
## classification 0.360866364  0.37814815 -0.7788644304  0.7665899  0.46477340
##          rbc          pc          pcc          ba          bgr
## age     -0.26777302  0.164502565  0.13613193  0.04624024  0.32221148
## bp      -0.29912666  0.163472082  0.18932545  0.15042067  0.23114231
## sg       0.49685398 -0.477573781 -0.36221968 -0.40200805 -0.47580748
## al      -0.33288253  0.625431330  0.50795475  0.45998172  0.47305015
## su      -0.24315965  0.242457901  0.11668496  0.23507779  0.66136266
## rbc      1.00000000 -0.379086976 -0.11359657 -0.10520976 -0.29295004
## pc      -0.37908698  1.000000000  0.50104506  0.39325471  0.30546598
## pcc     -0.11359657  0.501045059  1.00000000  0.41388542  0.23773954
## ba      -0.10520976  0.393254708  0.41388542  1.00000000  0.23584062
## bgr     -0.29295004  0.305465978  0.23773954  0.23584062  1.00000000
## bu      -0.31946153  0.500867080  0.34671686  0.30450548  0.35901063
## sc      -0.40791064  0.574256963  0.41006837  0.38364840  0.46991832
## sod      0.23611898 -0.385365835 -0.39925925 -0.19627893 -0.30541285
## pot      0.02227545 -0.002967037  0.04227078  0.12239794  0.04125286
## hemo     0.46652704 -0.550360391 -0.44678532 -0.34119242 -0.42428302
## pcv     0.46619891 -0.549571918 -0.44734147 -0.32737760 -0.43649622
## wc      -0.13897485  0.139356854  0.14449897  0.17041362  0.20712396
## rc       0.39018026 -0.493792005 -0.42670900 -0.28710274 -0.38874510
## htn     0.46916532 -0.389724063 -0.29648473 -0.18707902 -0.55285059
## dm     -0.51227272  0.393212075  0.21394644  0.21526316  0.55421450
```

## cad	-0.22405114	0.240011486	0.29463208	0.22218138	0.43485047
## appet	0.29049056	-0.336808885	-0.37777601	-0.12877498	-0.29559441
## pe	-0.20720124	0.419603193	0.28607633	0.31231805	0.29569727
## ane	-0.14770181	0.363904302	0.40616353	0.10935672	0.17339889
## classification	-0.67039757	0.546992779	0.39079382	0.32771187	0.53630216
##	bu	sc	sod	pot	hemo
## age	0.2824697	0.3217198	-0.12601146	0.0089319010	-0.2406551
## bp	0.2208025	0.3693470	-0.19664334	0.0340755885	-0.2881206
## sg	-0.5373770	-0.6378669	0.51199202	-0.0001720557	0.6839426
## al	0.6527854	0.7803694	-0.56378945	0.1489698817	-0.7190514
## su	0.3072082	0.3652508	-0.27343497	0.0482919129	-0.2972890
## rbc	-0.3194615	-0.4079106	0.23611898	0.0222754531	0.4665270
## pc	0.5008671	0.5742570	-0.38536583	-0.0029670367	-0.5503604
## pcc	0.3467169	0.4100684	-0.39925925	0.0422707755	-0.4467853
## ba	0.3045055	0.3836484	-0.19627893	0.1223979414	-0.3411924
## bgr	0.3590106	0.4699183	-0.30541285	0.0412528617	-0.4242830
## bu	1.0000000	0.8062202	-0.47620160	0.2135299791	-0.6483460
## sc	0.8062202	1.0000000	-0.54822025	0.1858411648	-0.7488169
## sod	-0.4762016	-0.5482202	1.00000000	0.0220707684	0.5399172
## pot	0.2135300	0.1858412	0.02207077	1.0000000000	-0.1487577
## hemo	-0.6483460	-0.7488169	0.53991718	-0.1487576873	1.0000000
## pcv	-0.6289479	-0.7429468	0.53561033	-0.2302989038	0.8630132
## wc	0.1746400	0.1737633	-0.17198434	-0.1072650252	-0.3091930
## rc	-0.5715641	-0.6451290	0.44065284	-0.2031060910	0.7682838
## htn	-0.5858537	-0.7001620	0.47522573	-0.1406340641	0.6899758
## dm	0.5563189	0.6086356	-0.43215564	0.1030288045	-0.5886152
## cad	0.3638113	0.3800856	-0.22112424	0.1797261648	-0.3555324
## appet	-0.3979437	-0.4514644	0.42262114	-0.1068709160	0.5315416
## pe	0.4571142	0.5592488	-0.40660753	0.1446861754	-0.5140469
## ane	0.4312891	0.4816326	-0.45982680	0.1011594982	-0.5719004
## classification	0.6316191	0.7800113	-0.55720766	0.0660805863	-0.8199482
##	pcv	wc	rc	htn	dm
## age	-0.2556536	0.19845616	-0.2437633	-0.3778575	0.3532676
## bp	-0.3197104	0.05844127	-0.2272663	-0.3170907	0.2269030
## sg	0.6756655	-0.29210344	0.6110836	0.5843651	-0.5912723
## al	-0.6989695	0.22757676	-0.6020627	-0.6752590	0.5575439
## su	-0.3235448	0.23185209	-0.2857830	-0.5459849	0.5506716
## rbc	0.4661989	-0.13897485	0.3901803	0.4691653	-0.5122727
## pc	-0.5495719	0.13935685	-0.4937920	-0.3897241	0.3932121
## pcc	-0.4473415	0.14449897	-0.4267090	-0.2964847	0.2139464
## ba	-0.3273776	0.17041362	-0.2871027	-0.1870790	0.2152632
## bgr	-0.4364962	0.20712396	-0.3887451	-0.5528506	0.5542145
## bu	-0.6289479	0.17463997	-0.5715641	-0.5858537	0.5563189
## sc	-0.7429468	0.17376331	-0.6451290	-0.7001620	0.6086356
## sod	0.5356103	-0.17198434	0.4406528	0.4752257	-0.4321556
## pot	-0.2302989	-0.10726503	-0.2031061	-0.1406341	0.1030288
## hemo	0.8630132	-0.30919304	0.7682838	0.6899758	-0.5886152
## pcv	1.0000000	-0.29497029	0.7662987	0.6874708	-0.5941638
## wc	-0.2949703	1.00000000	-0.2225593	-0.1941537	0.2500957
## rc	0.7662987	-0.22255931	1.0000000	0.6335552	-0.5661479
## htn	0.6874708	-0.19415366	0.6335552	1.0000000	-0.7428080
## dm	-0.5941638	0.25009565	-0.5661479	-0.7428080	1.0000000
## cad	-0.3595783	0.07320854	-0.3517483	-0.4569771	0.3728097
## appet	0.5462057	-0.22725751	0.4817237	0.4675952	-0.3870829

```
## pe -0.5240052 0.16727941 -0.5067597 -0.5160550 0.5345088
## ane -0.5675852 0.10188866 -0.5127225 -0.4594276 0.1712251
## classification -0.8089913 0.35453941 -0.7121427 -0.8019403 0.7352085
## cad appet pe ane classification
## age 0.22881398 -0.19410231 0.21277428 0.04146747 0.36086636
## bp 0.15574486 -0.08801784 0.06177449 0.24641924 0.37814815
## sg -0.29097682 0.43912771 -0.51278789 -0.32982537 -0.77886443
## al 0.35655570 -0.47807175 0.54400432 0.47213407 0.76658992
## su 0.40070040 -0.11688235 0.21386052 0.09091403 0.46477340
## rbc -0.22405114 0.29049056 -0.20720124 -0.14770181 -0.67039757
## pc 0.24001149 -0.33680889 0.41960319 0.36390430 0.54699278
## pcc 0.29463208 -0.37777601 0.28607633 0.40616353 0.39079382
## ba 0.22218138 -0.12877498 0.31231805 0.10935672 0.32771187
## bgr 0.43485047 -0.29559441 0.29569727 0.17339889 0.53630216
## bu 0.36381133 -0.39794367 0.45711421 0.43128907 0.63161914
## sc 0.38008564 -0.45146440 0.55924881 0.48163262 0.78001128
## sod -0.22112424 0.42262114 -0.40660753 -0.45982680 -0.55720766
## pot 0.17972616 -0.10687092 0.14468618 0.10115950 0.06608059
## hemo -0.35553236 0.53154161 -0.51404691 -0.57190042 -0.81994817
## pcv -0.35957829 0.54620565 -0.52400523 -0.56758525 -0.80899130
## wc 0.07320854 -0.22725751 0.16727941 0.10188866 0.35453941
## rc -0.35174833 0.48172369 -0.50675966 -0.51272246 -0.71214274
## htn -0.45697709 0.46759517 -0.51605495 -0.45942757 -0.80194033
## dm 0.37280971 -0.38708289 0.53450876 0.17122507 0.73520846
## cad 1.00000000 -0.14808267 0.26228731 0.14076051 0.36646836
## appet -0.14808267 1.00000000 -0.55849915 -0.38012158 -0.47183754
## pe 0.26228731 -0.55849915 1.00000000 0.24802270 0.49070767
## ane 0.14076051 -0.38012158 0.24802270 1.00000000 0.42557024
## classification 0.36646836 -0.47183754 0.49070767 0.42557024 1.00000000
```

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

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

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