Data Visualization - CSE 3020
Lab 3 - Correlation Plots in R

Name: Ayush Sharma
Reg. No: 15BCE1335
Faculty: Dr. Priyadarshini J.

1. **Assume your own dataset (may be downloaded) and apply the correct correlation techniques to identify the relationship among the variables for the your dataset also find the positive and negative correlation.**

Dataset: Vertebral Column Dataset is taken from the following website..
http://archive.ics.uci.edu/ml/datasets/vertebral+column

Code: (For Correlogram)
```
data <- read.delim("C:/Users/TEMP/Downloads/column_3C.dat", header = TRUE, sep="")
colnames(data) <- c("pelvic incidence", "pelvic tilt", "lumbar lordosis angle", "sacral slope",
"pelvic radius", "grade of spondylolisthesis", "class")
df<-data[,c("pelvic incidence", "pelvic tilt", "lumbar lordosis angle",
        "sacral slope", "pelvic radius", "grade of spondylolisthesis")]
corr <- round(cor(df), 1)
# Plot
ggcorrplot(corr, hc.order = TRUE,
        type = "lower",
        lab = TRUE,
        lab_size = 3,
        method="circle",
        colors = c("tomato2", "white", "springgreen3"),
        title="Correlogram of Vertebral Column Dataset",
        ggtheme=theme_bw)
```

Screenshot:

Correlogram of Vertebral Column Dataset
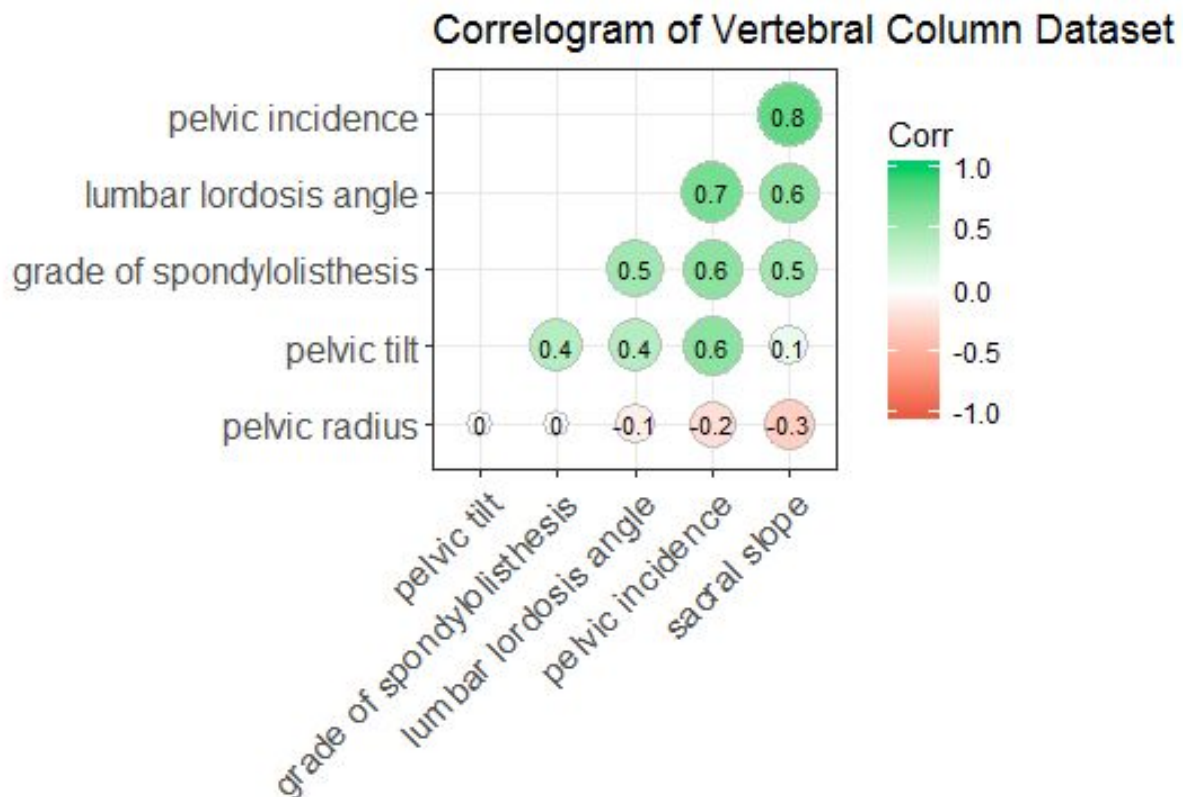
Code: (Scatter Plot for min and max)

```
# install.packages("ggplot2")
# load package and data
options(scipen=999) # turn-off scientific notation like 1e+48
library(ggplot2)
theme_set(theme_bw()) # pre-set the bw theme.
data <- read.delim("C:/Users/TEMP/Downloads/column_3C.dat", header = TRUE, sep="")
colnames(data) <- c("pelvic_incidence", "pelvic_tilt", "lumbar_lordosis_angle",
            "sacral_slope", "pelvic_radius", "grade_of_spondylolisthesis",
            "class")
# Scatterplot
gg <- ggplot(data, aes(x=pelvic_incidence, y=sacral_slope)) +
  geom_point(aes(col=class, size=grade_of_spondylolisthesis)) +
  geom_smooth(method="loess", se=F) +
  labs(y="Sacral Slope",
     x="Pelvic Incidence",
     title="Scatterplot")
plot(gg)
```

```
gg1 <- ggplot(data, aes(x=pelvic_radius, y=sacral_slope)) +
  geom_point(aes(col=class, size=grade_of_spondylolisthesis)) +
  geom_smooth(method="loess", se=F) +
  labs(y="Sacral Slope",
       x="Pelvic Radius",
       title="Scatterplot")
plot(gg1)
```

Screenshots:
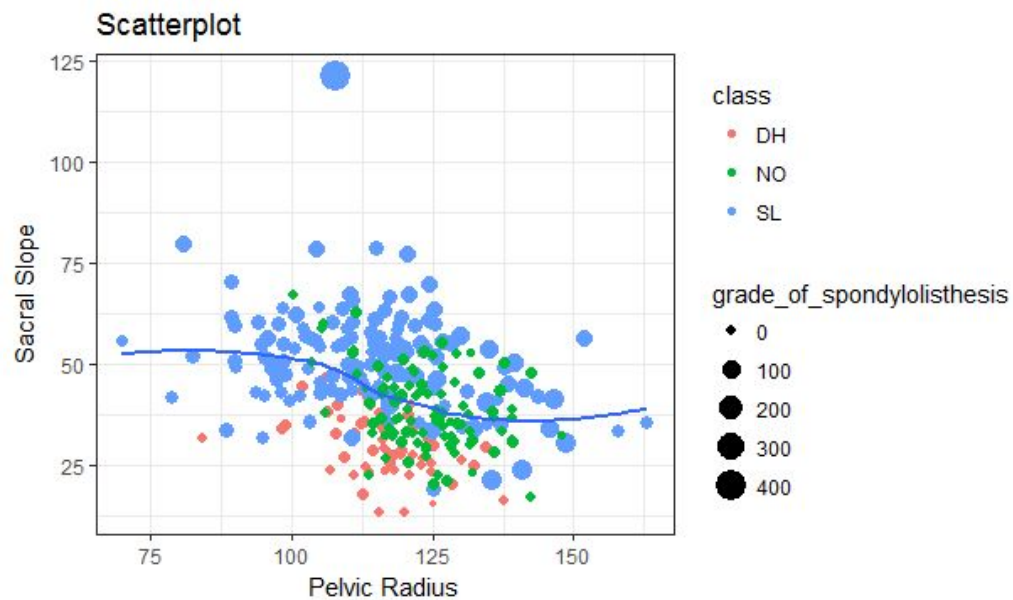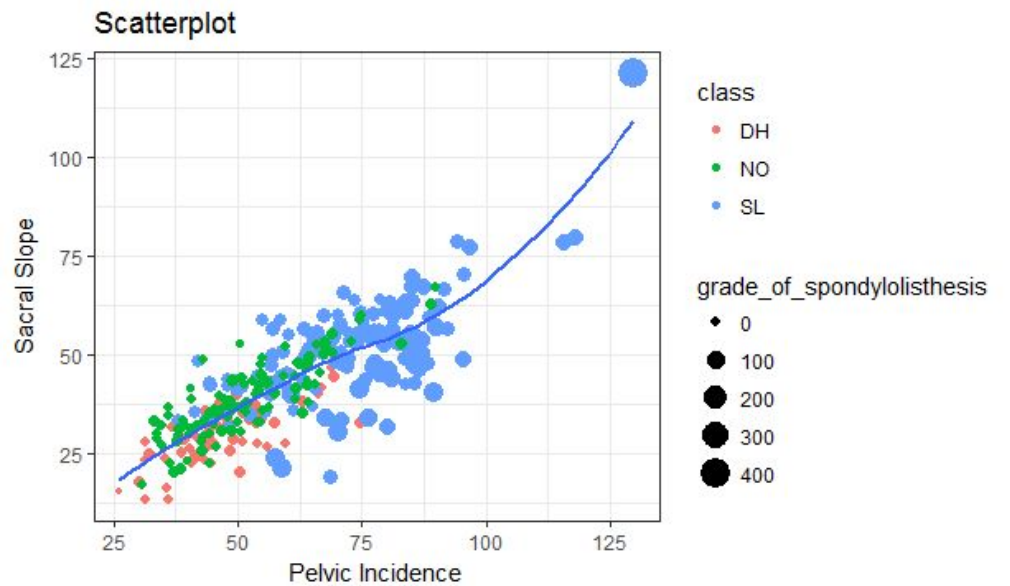
Explanation:

From the Correlogram, we can see that the correlation is majorly positive and is negative for few couples of attributes. Correlation is a measure of the strength and direction of the linear relationship between two variables.

We have plotted the scatter plots for these two couple of attributes having the maximum and minimum correlation.

We can clearly see that the line representation of the scatter plots shows a gradual increase and decrease respectively.

In the scatter plot, we can also see that the size of every point varies representing another attribute for the medical report (i.e. grade_of_spondylolisthesis) which varies from 0-400.