

Data Visualization-01 Base R Plots

Created by H. M. Samadhi Chathuranga Rathnayake

```
setwd("D:\\Workshops\\R Programming for Data Science Workshop\\Part 03 - Data
Visualization\\Datasets")
data=read.csv("iris.CSV")
head(data)

#Univariate Analysis
#Categorical Data

#Bar Charts
tab=table(data$Species)
barplot(tab)

barplot(tab,xlab ="Species",ylab = "Frequency",main ="Iris Species",col
="red")

barplot(tab,xlab ="Species",ylab = "Frequency",main ="Iris Species",col
=c("red","green","blue"))

bp=barplot(tab,xlab ="Species",ylab = "Frequency",main ="Iris Species",col
=c("red","green","blue"))
text(x = bp, y = data.frame(tab)$Freq, label = data.frame(tab)$Freq, pos = 1,
cex = 0.9)

#Pie Charts
tab=table(data$Species)
pc=pie(tab,xlab ="Species",ylab = "Frequency",main ="Iris Species",col
=c("red","green","blue"))

lb=paste(round(prop.table(tab)*100,2), "%", sep = "")
pc=pie(tab,xlab ="Species",ylab = "Frequency",main ="Iris Species",col
=c("red","green","blue"),labels = lb)

#Numerical Data

#Line Charts
plot(data$Id,data$Sepal.Length,type="l",xlab ="ID",ylab = "Values",main
="Sepal Length")

#Histograms
hist(data$Sepal.Length,xlab ="Sepal Length",ylab = "Frequency",main ="Sepal
Length Distribution",col="blue")

nmdata=rnorm(1000,50,5)
hist(nmdata)
```

```

nmdata=rnorm(1000)
hist(nmdata)

#Density Charts
den=density(data$Sepal.Length)
plot(den, frame = FALSE, xlab = "Sepal Length",ylab = "Density",main = "Sepal
Length Distribution",col="blue")
polygon(den, col = "purple")

#Boxplots
boxplot(data$Sepal.Length,xlab = "Sepal Length",ylab = "Spread",main = "Sepal
Length Distribution",col="green")

#Bivariate Analysis

#Numrical VS Numerical

#Multiple Line Charts
dat=cbind(data$Sepal.Length,data$Petal.Length)
matplot(dat, type = c("l","l"),col = c("red","blue"),xlab = "Index",ylab =
"Length",main="Sepal Length VS Petal Length")
legend("topleft", legend = c("SL","PL"), col=c("red","blue"), pch=1)

#Scatter Plots
plot(data$Sepal.Length~data$Sepal.Width,xlab="Sepal Width",ylab="Sepal
Length",main="Sepal Length VS Sepal Width",col="red",pch=10)

scatter.smooth(data$Sepal.Length~data$Sepal.Width,xlab="Sepal
Width",ylab="Sepal Length",main="Sepal Length VS Sepal
Width",col="red",pch=10)

#Pair Plot
num_data=data[c("Sepal.Length","Sepal.Width","Petal.Length","Petal.Width")]
head(num_data)

pairs(num_data)

#Categorical VS Categorical

#Stacked Bar Charts
data=read.csv("Living_Area.CSV")
head(data)

tab=table(data$Gender,data$Living)
tab

barplot(tab,col=c("blue","green"),legend=c("Female","Male"),xlab="Living
Area",ylab="Frequency",main="Gender VS Living Area")

#Multiple Bar Charts
tab=table(data$Gender,data$Living)
tab

```

```

barplot(tab,col=c("blue","green"),legend=c("Female","Male"),xlab="Living
Area",ylab="Frequency",main="Gender VS Living Area",beside = TRUE)

#100% Stacked Bar Charts
fun_perc=function(x){
  return(100*x/sum(x,na.rm = TRUE))
}
per_tab=apply(tab,2,fun_perc)
per_tab

barplot(per_tab,col=c("blue","green"),legend=c("Female","Male"),xlab="Living
Area",ylab="Frequency",main="Gender VS Living Area")

#Numerical VS Categorical
data=read.csv("iris.CSV")
head(data)

boxplot(data$Sepal.Length~data$Species,xlab ="Species",ylab = "Sepal
Length",main ="Sepal Length VS Species",col=c("green","red","blue"))

#Multiple Plots in a same figure
par(mfrow=c(2,2))
hist(data$Sepal.Length)
boxplot(data$Sepal.Width)
scatter.smooth(data$Sepal.Length~data$Petal.Length)
boxplot(data$Sepal.Length~data$Species,xlab ="Species",ylab = "Sepal
Length",main ="Sepal Length VS Species",col=c("green","red","blue"))

par(mfcol=c(2,2))
hist(data$Sepal.Length)
boxplot(data$Sepal.Width)
scatter.smooth(data$Sepal.Length~data$Petal.Length)
boxplot(data$Sepal.Length~data$Species,xlab ="Species",ylab = "Sepal
Length",main ="Sepal Length VS Species",col=c("green","red","blue"))

par(mfrow=c(1,1))
hist(data$Sepal.Length)

```