

Data Visualization in RStudio

```
#-----  
# install.packages("DataExplorer","tidyverse")  
# load library  
library(tidyverse)  
library(datasets)  
data()          # Data sets in package 'datasets'
```

```
#-----  
# Motor Trend Car Road Tests  
#-----  
mtcars          # load built-in data  
str(mtcars)     # view structure  
?mtcars        # view documentation  
|  
#-----  
head(mtcars)  
tail(mtcars)  
head(mtcars, 25)  
nrow(mtcars)  
ncol(mtcars)
```

```
#-----  
# Running individual summary functions  
min(mtcars$mpg)      # [, 1] mpg Miles/(US) gallon  
mean(mtcars$mpg)  
median(mtcars$mpg)  
max(mtcars$mpg)  
quantile(mtcars$mpg)  
quantile(mtcars$mpg, probs = c(0.25,0.75))    # use option
```

```
#-----  
#SHORTCUT: summary() function  
summary(mtcars)  
summary(mtcars$mpg)      # Input a vector  
summary(mtcars[, 1:3])   # Input a data frame
```

```
#-----  
# table()  
mtcars$cyl      # [, 2] cyl Number of cylinders  
mtcars$am       # [, 9] am  Transmission (0 = automatic, 1 = manual)  
table(mtcars$cyl)      # input 1 vector  
table(mtcars$cyl,mtcars$am) # input 2 vector  
table(Cylinders = mtcars$cyl, Manual = mtcars$am) # set labels  
table(mtcars[, c('cyl','am')]) # input data frame
```

```
#-----  
# Plotting Functions  
# Base R functions  
# boxplot(), hist(), barplot(), plot()  
#-----  
# Boxplot  
boxplot(mtcars$mpg, main='Miles per Gallon')  
  
# A formula as an argument "\mpg conditional on cyl"  
boxplot(mpg ~ cyl, data = mtcars,  
        main = "Miles per Gallon by Number of Cylinders",  
        xlab = "Cylinders",  
        ylab = "Miles per Gallon")  
  
boxplot(mpg ~ cyl + am, data = mtcars,  
        main = "MPG by Number of Cylinders & Transmissions")
```

```
#-----  
# Histogram  
# (Takes a vector, and plots the distribution of values)  
  
hist(mtcars$mpg)  
mtcars %>% plot_histogram()
```

```
#-----  
# Bar Chart  
# (Use the table function to create a two-way frequency table, and  
plotting options to group bars)  
  
counts <- table(mtcars$cyl, mtcars$am)  
colnames(counts) <- c("Auto", "Manual")  
barplot(counts,  
        main = "Number of Cars by Transmission and Cylinders",  
        xlab = "Transmission",  
        beside = TRUE,  
        legend = rownames(counts))
```

```
#-----  
# Scatterplot  
plot(mtcars$mpg, mtcars$hp,  
      xlab = "Miles per Gallon",  
      ylab = "Horsepower")  
  
# create a vector for conditional color coding  
colorcode <- ifelse(mtcars$am == 0, "red", "blue")  
plot(mtcars$mpg, mtcars$hp,  
      xlab = "Miles per Gallon",  
      ylab = "Horsepower",  
      col = colorcode)
```

```
#-----  
#ggplot package  
#-----  
# Biochemical Oxygen Demand  
#-----  
BOD  
?BOD          # view documentation  
ggplot(data = BOD,  
        mapping = aes(x = Time,  
                       y = demand))+  
  geom_point(size = 5)+  
  geom_line(colour = 'red')  
  
ggplot(BOD,  
        mapping = aes(Time, demand))+  
  geom_point(size = 3)+  
  geom_line(colour = 'red')
```

```

#-----
# Carbon Dioxide Uptake in Grass Plants
#-----
C02
View(C02)
?C02
names(C02)
C02 %>%
  ggplot(aes(conc, uptake, colour =Treatment))+
  geom_point(size = 3, alpha = 0.5)+
  geom_smooth(method = lm, se = F) +
  facet_wrap(~Type)+
  labs(title = 'Concentration of co2')+
  theme_bw()

```

```

C02 %>%
  ggplot(aes(Treatment, uptake))+
  geom_boxplot()+
  geom_point(alpha = 0.5,
             aes(size = conc, colour = Plant))+
  facet_wrap(~Type)+
  coord_flip()+
  theme_bw()+
  labs(title = 'Chilled Vs Non-chilled')

```

```
#-----
mpg %>%
  ggplot(aes(displ, cty))+
  geom_point(aes(colour = drv, size = trans),
             alpha = 0.5)+
  geom_smooth(method = lm)+
  facet_wrap(~year, nrow = 1)+
  labs(x = 'Engine size',
       y = 'MPG in the city',
       title = 'Fuel efficiency')+
  theme_bw()
```

Activity by using RStudio

1. Create a chart that highlights the relationship between Sales and Sales FTE (full-time employees)

Region	Sales (USD M)	Share of Region	YoY Growth (%)	Sales FTE
North	300	13%	2.3	12
Northwest	233	10%	-0.4	11
East	335	12%	-1.0	14
East Central	411	15%	4.0	17
Southwest	194	8%	0.2	9
Southeast	168	10%	-2.1	5
South Central	590	18%	3.5	24
West	372	12%	1.8	14

2. Create an EDA for the dataset 'starwars' in package 'datasets'