Visualization for Univariate

View of data

- head(df,n) – displays n records/values from x object. N by default is 6.
- tail(df) – displays last 6 records/values from x object
- View(df) – helps to view the dataframe. Opens in the script window.
- table(x) – displays the count/frequencies of x with each category (categorical variable)
- summary(numeric_vector)- displays the fivenumber summary (min,max,median,first and third quartile)

Visualization

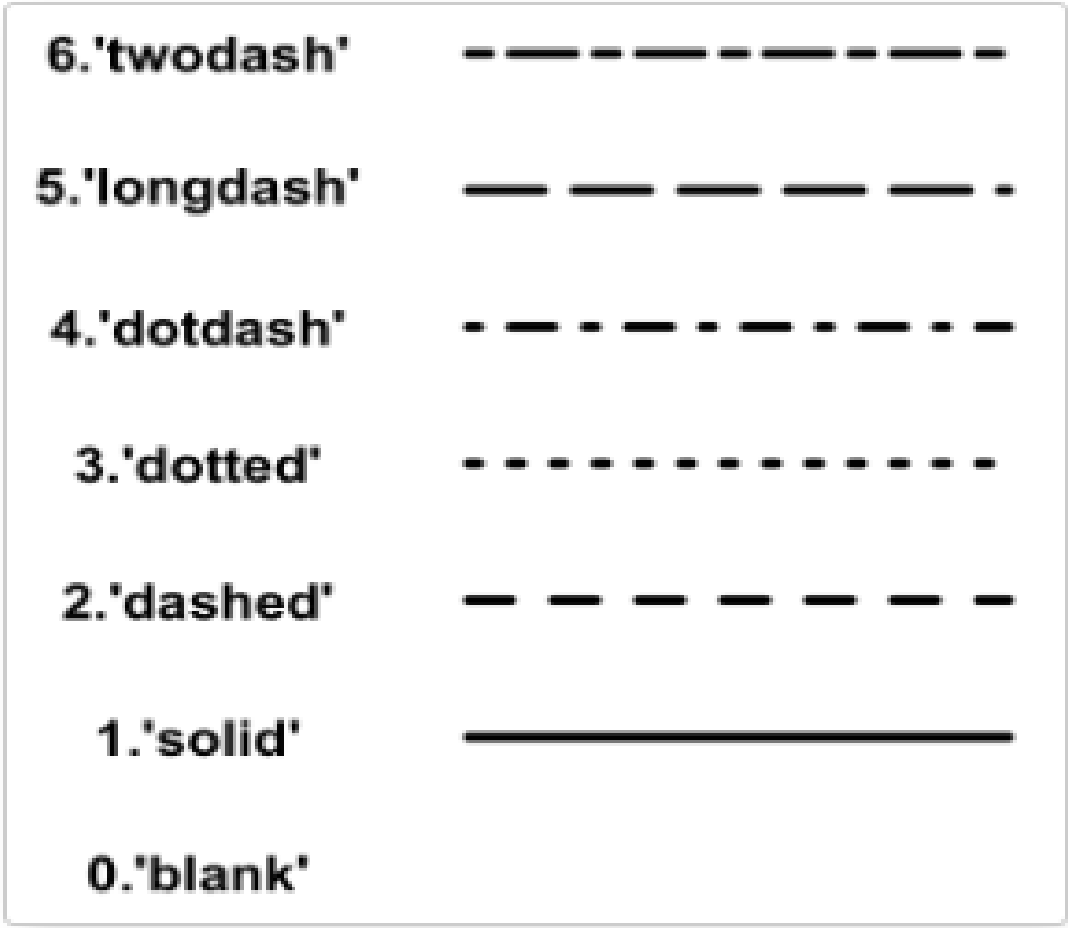
- Basic Visualization
  - Bar/Line chart
  - Pie chart
  - Histogram
  - Box plot

General parameters in any plot

- bg – plot background color
- lty – line type (e.g. dot, dash, solid)
- lwd – line width
- col – color
- cex – text size /symbol size inside plot
- xlab, ylab – axes labels
- main – title
- pch – plotting symbol
- xlim,ylim – axes scale (vector of 2 elements, start and end)

lty

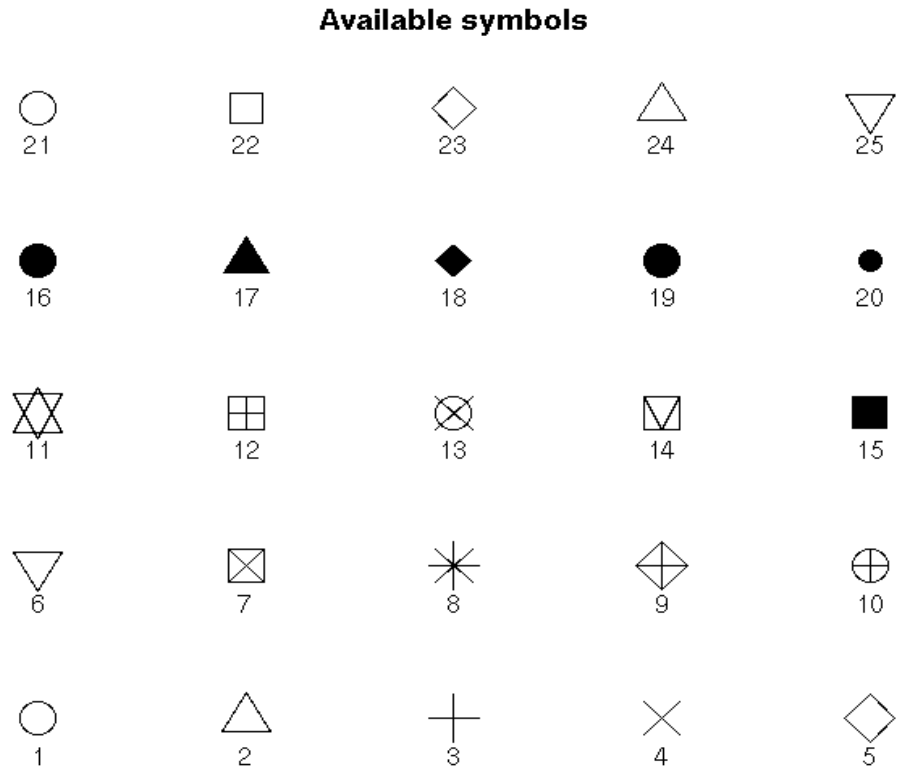
- Represented with text or number
- Text: **“blank”, “solid”, “dashed”, “dotted”, “dotdash”, “longdash”, “twodash”**
- Number - **0, 1, 2, 3, 4, 5, 6**



Line width and cex

- cex – default=1, 1.5 is 50% larger, 0.5 is 50% smaller
- lwd - default=1, 2 is twice as wide

pch


Available symbols
21
22
23
24
25
16
17
18
19
20
11
12
13
14
15
6
7
8
9
10
1
2
3
4
5

col - colors

- black, blue, blueviolet, brown, darkgreen, darkred, darkorange,darkviolet, deeppink, green, magenta, maroon, orangered, tomato, skyblue, yellow, wheat, violet, steelblue …
- Some function for colors

  rainbow(*n*), **heat.colors(*n*)**, **terrain.colors(*n*)**, **topo.colors(*n*)**, and **cm.colors(*n*)**.
- For further color range refer

  http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf

Barplot

- Used for ploting categorical variable expressed as numerical values using table function /numerical values.
- barplot(vector)

Bar Plot

- barplot(H, xlab, ylab, main, names.arg, col)
  - **H** is a vector or matrix containing numeric values used in bar chart. It can be used for categorical variable by giving table function output to the barplot
  - **xlab** is the label for x axis.
  - **ylab** is the label for y axis.
  - **main** is the title of the bar chart.
  - **names.arg** is a vector of names appearing under each bar.
  - **col** is used to give colors to the bars in the graph.

Barplot – Example for numeric

```
H <- c(7,12,28,3,41)
M <- c("Mar","Apr","May","Jun","Jul")
# Plot the bar chart.
barplot(H,names.arg = M,xlab = "Month",ylab = "Revenue",col = "blue", main = "Revenue chart",border = "red")
```

Barplot Example

```
y=table(bike$cyc_freq)
barplot(y,
        main="Freq Distribution",
         xlab="Department",
         ylab="Frequency",
         col = c("red", "yellow", "green","violet"))
```

Pie chart

- pie(x, labels, radius, main, col, clockwise)
  - **x** is a vector containing the numeric values used in the pie chart.
  - **labels** is used to give description to the slices.
  - **radius** indicates the radius of the circle of the pie chart.(value between −1 and +1).
  - **main** indicates the title of the chart.
  - **col** indicates the color palette.
  - **clockwise** is a logical value indicating if the slices are drawn clockwise or anti clockwise.
- Eg: pie (table(bike$cyc_freq), col = c("red", "yellow", "green","violet"))

```
# Create data for the graph.
x <- c(21, 62, 10, 53)
labels <- c("London", "New York", "Singapore", "Mumbai")

 # Plot the chart with title and rainbow color pallet.
pie(x, labels, main = "City pie chart", col = rainbow(length(x)))
piepercent<- round(100*x/sum(x), 1)
# plot with more parameters
pie(x, labels = piepercent, main = "City pie chart",col = rainbow(length(x)))
legend("topright", c("London","New York","Singapore","Mumbai"), cex = 0.8, fill =
rainbow(length(x)))
```

Histogram

- Histogram specifies the count/frequencies on numerical data.
- It usually makes bins and then counts the values according to the bins. So x-axis represents the bins and y axis represents the count.
- hist(numerical_vector)

Histogram

- hist(v,main,xlab,xlim,ylim,breaks,col,border)
  - **v** is a vector containing numeric values used in histogram.
  - **main** indicates title of the chart.
  - **col** is used to set color of the bars.
  - **border** is used to set border color of each bar.
  - **xlab** is used to give description of x-axis.
  - **xlim** is used to specify the range of values on the x-axis.
  - **ylim** is used to specify the range of values on the y-axis.
  - **breaks** is used to mention the width of each bar.

Histogram Example

```
# Create data for the graph.
v <- c(9,13,21,8,36,22,12,41,31,33,19)
# Create the histogram.
hist(v,
     main="histogram",
     xlab = "Weight",
     xlim = c(0,40), ylim = c(0,5), breaks = 5,
     col = "yellow",border = "blue")
```

Line Chart

- plot(v,type,col,xlab,ylab)
  - **v** is a vector containing the numeric values.
  - **type** takes the value "p" to draw only the points, "l" to draw only the lines and "o" to draw both points and lines.
  - **xlab** is the label for x axis.
  - **ylab** is the label for y axis.
  - **main** is the Title of the chart.
  - **col** is used to give colors to both the points and lines.

Line chart-Example

```
# Create the data for the chart.
v <- c(7,12,28,3,41)
t <- c(14,7,6,19,3)

# Plot the bar chart.
plot(v, type = "o",col = "red", xlab = "Month", ylab = "Rain fall", main = "Rain fall chart")
lines(t, type = "o", col = "blue")
```

Stacked Bar Plot

```
# Create the input vectors.
   colors <- c("green","orange","brown")
   months <- c("Mar","Apr","May","Jun","Jul")
   regions <- c("East","West","North")
 # Create the matrix of the values.
    Values <- matrix(c(2,9,3,11,9,4,8,7,3,12,5,2,8,10,11),nrow = 3,ncol =
        5,byrow = TRUE)
# Create the bar chart.
   barplot(Values,main = "total revenue",names.arg = months,xlab =
        "month",ylab = "revenue", col = colors)
# Add the legend to the chart.
        legend("topleft", regions, cex = 1.3, fill = colors)
```

Box Plot

- boxplot(x, data, notch, varwidth, names, main)
  - **x** is a vector [univariate] or a formula[bivariate].
  - **data** is the data frame.
  - **notch** is a logical value. Set as TRUE to draw a notch.
  - **varwidth** is a logical value. Set as true to draw width of the box proportionate to the sample size.
  - **names** are the group labels which will be printed under each boxplot.
  - **main** is used to give a title to the graph.
- boxplot(bike$distance, data = bike, xlab = “distance", main = “Sales Data")

Multiple plots

- The number of plots on a page, and their placement on the page, can be controlled using *par()* or *layout().*

- The number of figure regions can be controlled using *mfrow* and *mfcol.*

  e.g. par(mfrow=c(3,2)) # Creates 6 figures arranged in
  3 rows and 2 columns

- *Layout()* allows the creation of multiple figure regions of unequal sizes.

  e.g. layout(matrix(c(1,2)), heights=c(2,1))

- *To reset to the original*

  par(mfrow=c(1,1))

  layout(1)

Save the output

- Could be files
  - **Not Scalable:** JPG, BMP,PNG
  - **Scalable**: Postscript, Pdf
- Specify destination of graphics output or simply right click and copy

Or

```
# Give the chart file a name.
png(file = "mychart.jpg")
# Plot the chart.
.....
# Save the file.
dev.off()
```