Semester : VI Subject : DAV Academic Year: 2023 - 2024

VISUALIZATION BEFORE ANALYSIS:

Data visualization is the technique used to deliver insights in data using visual cues such as graph, charts, maps and many others.

Data visualization in R programming Language:

The R is a language that is designed for statistical computing, graphical data analysis, and scientific research. It is usually preferred for data visualization as it offers flexibility and minimum required coding through its packages.

Consider the following air quality data set for visualization in R:

Ozone	Solar R.	Wind	Temp	Month	Day
41	119	7.4	67	5	1
36	118	8.0	72	5	2
12	149	12.6	74	5	3
18	313	11.5	62	5	4
NA	NA	14.3	56	5	5
28	NA	14.9	66	5	6

Types of Data Visualizations:

Some of the various types of visualizations offered are:

BarPlot:-

There are two types of barplots - horizontal and vertical which represents data points as horizontal or

Semester : VI Subject : DAV Academic Year: 2023-2024

Barplots are used for the following scenarios :

* To perform comparitive study between the various data categories in the dataset.

* To analyze a change of a variable over time in months or years.

Histogram:

→ A histogram is like a barchart as it uses bars of varying height to represent data distribution.

→ In histogram values are grouped into consecutive intervals called bins.

→ Continuous values are grouped and displayed in these bins whose size can be varied.

Example:

```
#Histogram for Maximum Daily Temperature.
data(airquality)
hist(airquality$Temp, main = "Temperature Analysis",
     xlab = "Temperature (Fahrenheit)",
     xlim = c(50,125), col = "yellow", freq = TRUE)
```

For a histogram, the temperature xlim can be used to specify the interval within which all values are to be displayed.

Another parameter freq when set to true denotes the frequency of various values in the histogram and when set to

FALSE, the probability densities are represented on the y-axis such that they are of the histogram adds up to one.

Histograms are used in the following scenarios:

* To verify an equal and symmetric distribution of the data.
* To identify deviations from expected value.

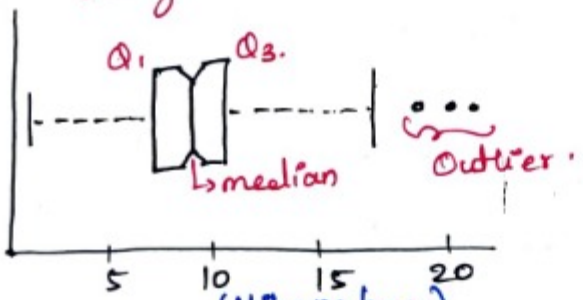
Box Plot:

The statistical summary of the given data is presented graphically using a boxplot.

A boxplot depicts information like the minimum and maximum data point, the median value, first and third quartile and IQR.

Example:

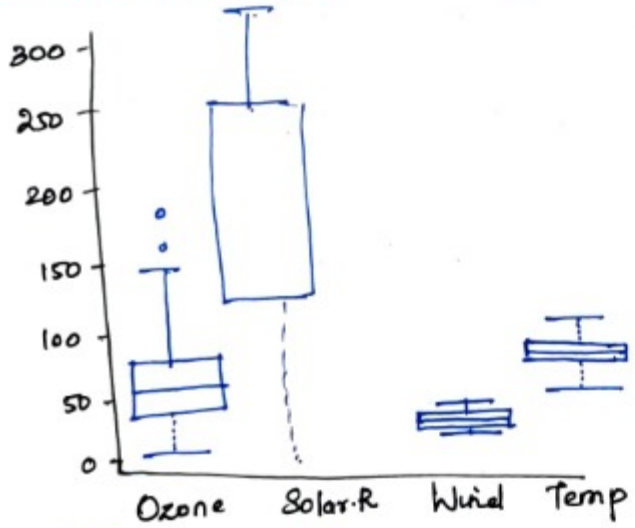
```
#Boxplot for average wind speed
data(airquality)
boxplot(airquality$Wind, main = "Average wind speed",
        xlab = "Miles per hour", ylab = "Wind",
        col = "orange", border = "brown",
        horizontal = TRUE, notch = TRUE)
```



Semester: VI Subject: DAV Academic Year: 2023 - 2024

Multiple box plots can also be generated through the following code:

```
#Multiple BoxPlots each representing an airquality
boxplot (airquality [,0:4],
main = 'Box Plots for Air Quality Parameters').
```



Box plots are used for:

* To give a comprehensive statistical description of the data through a visual cue.
* To identify the outlier points that do not lie in the inter-quartile range of data.

Scatter Plot:

A scatter plot is composed of many points on a cartesian plane. Each point denotes the value taken by two parameters and helps us easily identify the relationship between them.

```
#Scatterplot for Ozone Concentration per month
data(airquality)
```

```
plot(airquality$Ozone, airquality$Month, main = "Scatterplot",
     xlab = "Ozone Concentration",
     ylab = "Month of observation")
```

Scatterplots are used in the following scenarios:

To show whether an association exists between bivariate data.

To measure the strength and direction of such a relationship.

Heatmap:

Heat map is defined as a graphical representation of data using colors to visualize the value of the matrix. heatmap() function is used to plot heatmap.

```
#set seed for reproducibility
set.seed(110)

#Create example data
data <- matrix(rnorm(50,0,5), nrow=5, ncol=5)

#Column names.
colnames(data) <- paste0("col",1:5)
rownames(data) <- paste0("row",1:5)

#Draw a heatmap
heatmap(data)
```