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

Exploratory Data Analysis (EDA):

Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data. These patterns include outliers and the features of the data that might be unexpected. EDA is an important first step in any data analysis.

The following step-by-step examples shows how to use functions to perform exploratory data analysis on starwars dataset that comes built-in with the tidyverse packages.

Example:-

```
library(tidyverse)
data(starwars)
```

The data() is used to load the starwars dataset.

```
?starwars
```

? is used to know about the starwars dataset.

```
dim(starwars)
```

The dim() shows the dimensions of starwars dataset. It displays the number of rows and columns in the dataset.

```
glimpse(starwars)
```

The glimpse() gives the entire glimpse of the dataset. It includes No. of rows and no. of columns, column names, datatype of each column and the values of each column.

```
head(starwars)
```

It displays the first few rows of starwars dataset

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

tail(starwars):
It displays the last few rows of the starwars.
starwars$name.
It displays all the values of the name coloumn.
names(starwars)
The names() displays only the coloumn names of the dataset.
length(starwars)
The length() methods shows the total count of the Coloumn names.
class(hair_color)
The class() method displays the datatype of the coloumn hair_color.
unique(hair_color)
If anyone wants to know the unique values of hair_color coloumn, then unique() method displays it.
sort(table(hair_color), decreasing = TRUE)
The table() method will count the unique occurrence of the values in the dataset. The count is displayed.
The sort() method will sort those values in descending order.
This can be also written as:

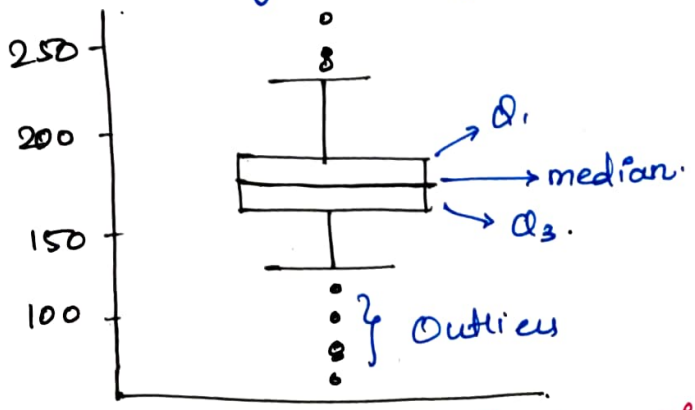
```
starwars %>%
    select(hair_color) %>%
    count(hair_color) %>%
    arrange(desc(n))
```

Semester : VI Subject : DAV Academic Year: 2023 2024

The datas are also visualized using Graph in EDA:

boxplot(starwars $height).

This will display the height coloumn as boxplot.



barplot(sort(table(hair_color), decreasing = TRUE)).

This will plot the barplot with the count of unique hair colors in decreasing order.