



Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Artificial Intelligence and Data Science

Experiment No: 7

Aim: Visualization in R using Libraries.

Theory:

Data visualization is a crucial aspect of data analysis, as it allows for the representation of complex data in a visually appealing and easily understandable manner. R, being a powerful statistical computing language, offers numerous packages and functions for creating high-quality visualizations. Here are some key theories and concepts related to data visualization in R:

- **Aesthetics Mapping:** Aesthetic mapping involves mapping variables in your dataset to visual properties such as color, size, shape, and position within the plot. In ggplot2, aesthetic mappings are defined within the `aes()` function.
- **Geometric Objects (Geoms):** Geometric objects, or geoms, represent the actual visual elements in a plot, such as points, lines, bars, or polygons. Each geom in ggplot2 corresponds to a different type of plot (e.g., `geom_point()` for scatter plots, `geom_bar()` for bar plots).
- **Statistical Transformations (Stats):** Statistical transformations, or stats, allow you to compute summary statistics or transformations of your data before plotting. For example, you can use `stat_smooth()` to add a smoothed line to a scatter plot.
- **Faceting:** Faceting involves splitting your data into subsets based on one or more categorical variables and creating separate plots for each subset. In ggplot2, you can use the `facet_wrap()` or `facet_grid()` functions to create faceted plots.
- **Customization and Themes:** R provides extensive options for customizing the appearance of plots, including modifying axis labels, titles, colors, fonts, and more. ggplot2 allows for further customization using themes, which provide pre-defined sets of appearance settings.



Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Artificial Intelligence and Data Science

Code:

The figure shows two RStudio sessions side-by-side. Both sessions have the following interface elements:

- File**, **Edit**, **Code**, **View**, **Plots**, **Session**, **Build**, **Debug**, **Profile**, **Tools**, **Help**, **Addins**.
- A toolbar with icons for file operations like Open, Save, Run, Source, and a search bar.
- An **Environment** tab showing variables: CPS85 (534 obs. of 11 variables) and plotdata (533 obs. of 11 variables).
- A **Data** pane showing the same two variables.
- A **Console** tab showing R code and its output.
- A **Terminal** tab showing R code and its output.
- A **Background Jobs** tab.

Top Session (R 4.3.2):

```
datavisual1.R* x
Source on Save Run Source Environment History Connections Tutorial
R Data Global Environment
CPS85 534 obs. of 11 variables
plotdata 533 obs. of 11 variables

1 # load data
2 data(CPS85, package = "mosaicData")
3 
4 # The first function in building a graph is the ggplot function. It specifies the
5 # data frame containing the data to be plotted the mapping of the variables to visual properties
6 # The mappings are placed within the aes function (where aes stands for aesthetics).
7 
8 # specify dataset and mapping
9 library(ggplot2)
10 ggplot(data = CPS85,
11         mapping = aes(x = exper, y = wage))
12 
13 # Geoms are the geometric objects (points, lines, bars, etc.) that can be placed on a graph.
14 # They are added using functions that start with geom_. In this example, we'll add points using the
15 # creating a scatterplot.
16 
17 # In ggplot2 graphs, functions are chained together using the + sign to build a final plot.
18 
19 # add points
20 ggplot(data = CPS85,
21         mapping = aes(x = exper, y = wage)) +
22         geom_point()
23 
24 # delete outlier
25 library(dplyr)
26 plotdata <- filter(CPS85, wage < 40)
27 
28 # redraw scatterplot
29 ggplot(data = plotdata,
30         mapping = aes(x = exper, y = wage)) +
31         geom_point()
32 
33 # make points blue, larger, and semi-transparent
34 ggplot(data = plotdata,
35         mapping = aes(x = exper, y = wage)) +
36         geom_point(mapping = aes(alpha = runif(exper)),
37                    alpha = .7,
38                    size = 3)
39 
40 # add a line of best fit.
41 ggplot(data = plotdata,
42         mapping = aes(x = exper, y = wage)) +
43         geom_point(mapping = aes(color = "cornflowerblue",
44                             alpha = .7,
45                             size = 3)) +
46         geom_smooth(method = "lm")
47 
48 # indicate sex using color
49 ggplot(data = plotdata,
50         mapping = aes(x = exper,
51                     y = wage,
52                     color = sex)) +
53 
54 <--
```

R 4.3.2 - /Users/.../Desktop/

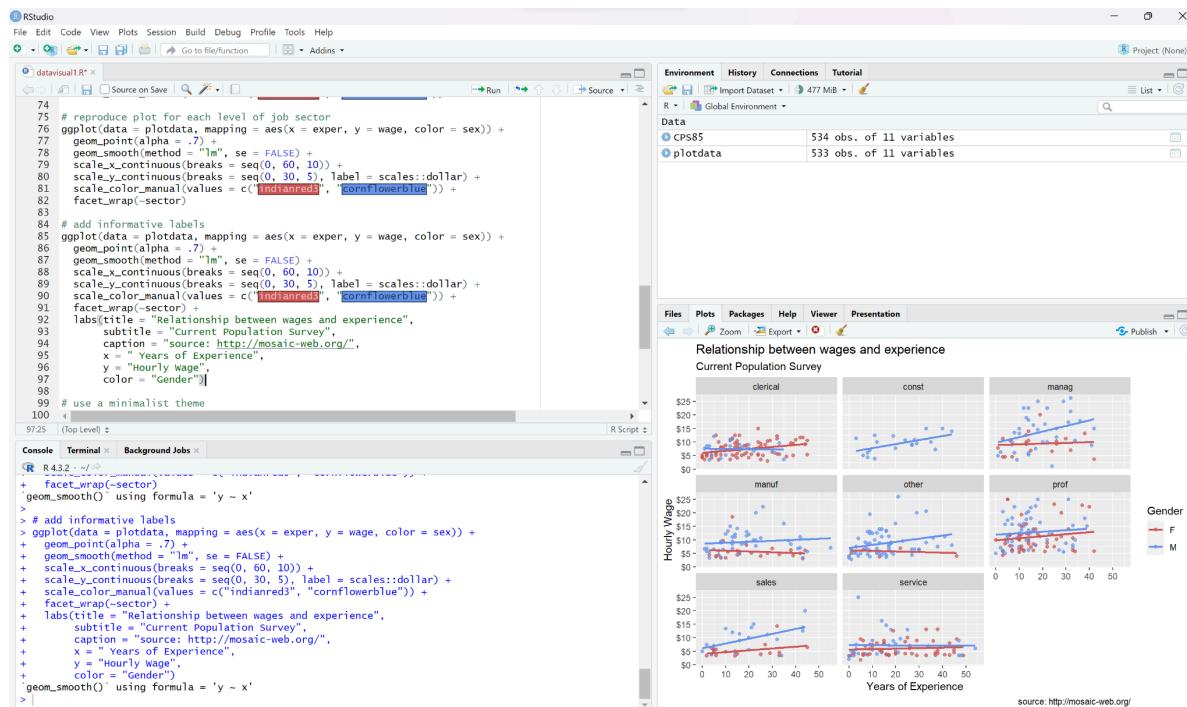
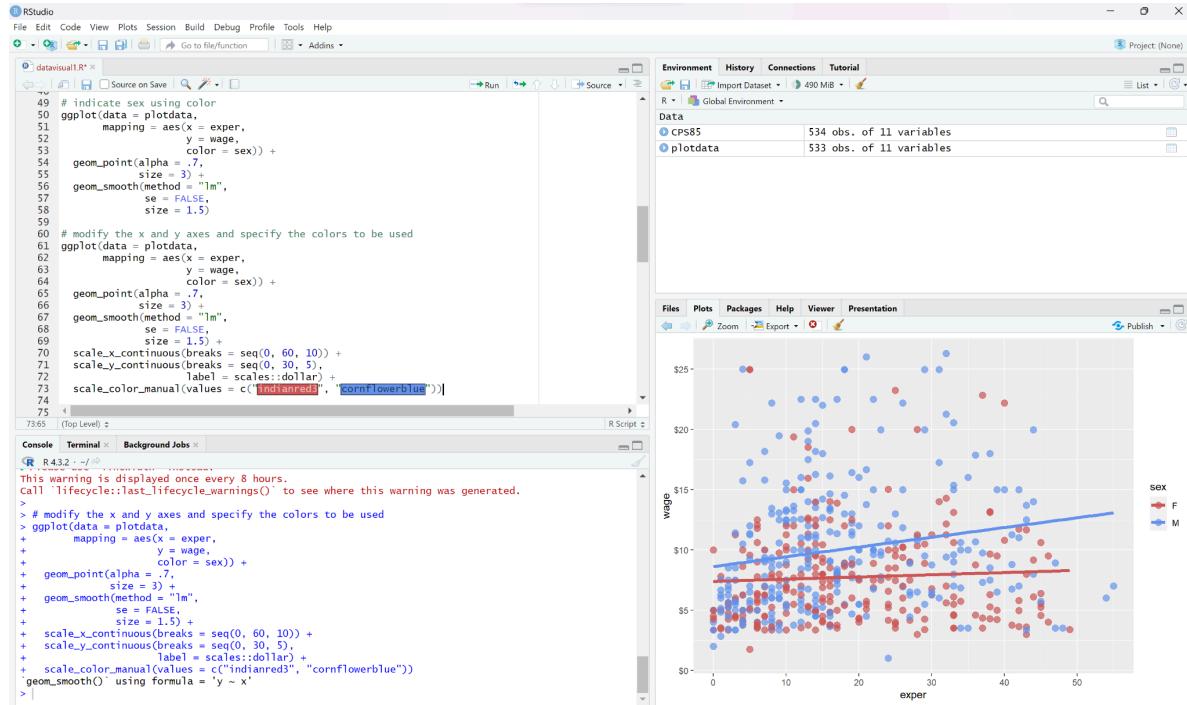
```
+ geom_point()
> 
> # make points blue, larger, and semi-transparent
ggplot(data = plotdata,
+       mapping = aes(x = exper, y = wage)) +
+       geom_point(mapping = aes(color = "cornflowerblue",
+                               alpha = .7,
+                               size = 3)) +
+       geom_smooth(method = "lm")
geom_smooth() using formula = 'y ~ x'
> 
```

Bottom Session (R 4.3.2):

This session shows the same data as the top one, but with the points colored by sex (male and female). The plot shows a positive correlation between experience and wage, with a linear regression line fitted to the data.

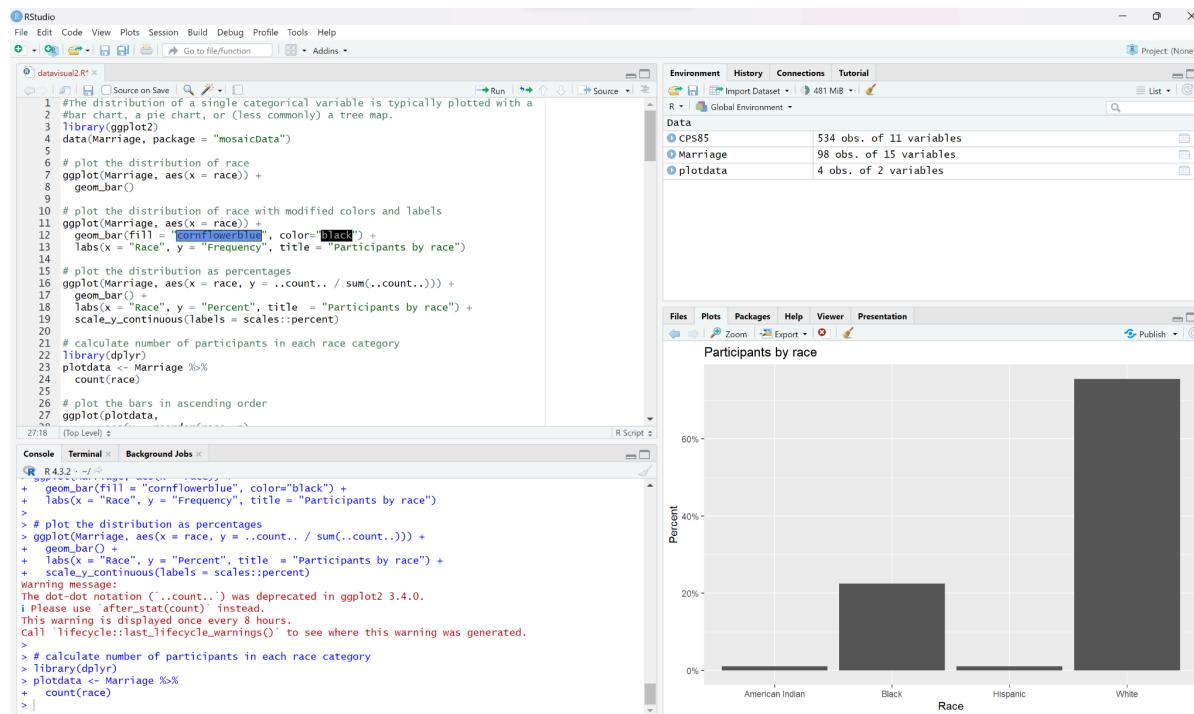
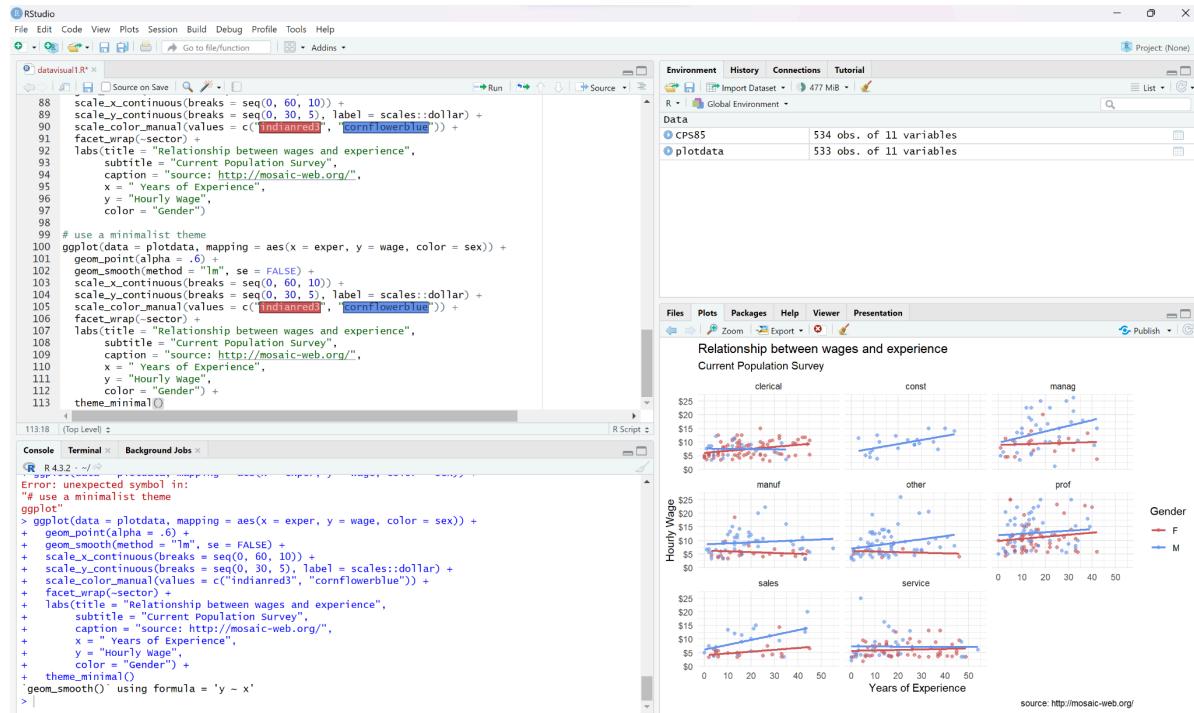


Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Artificial Intelligence and Data Science





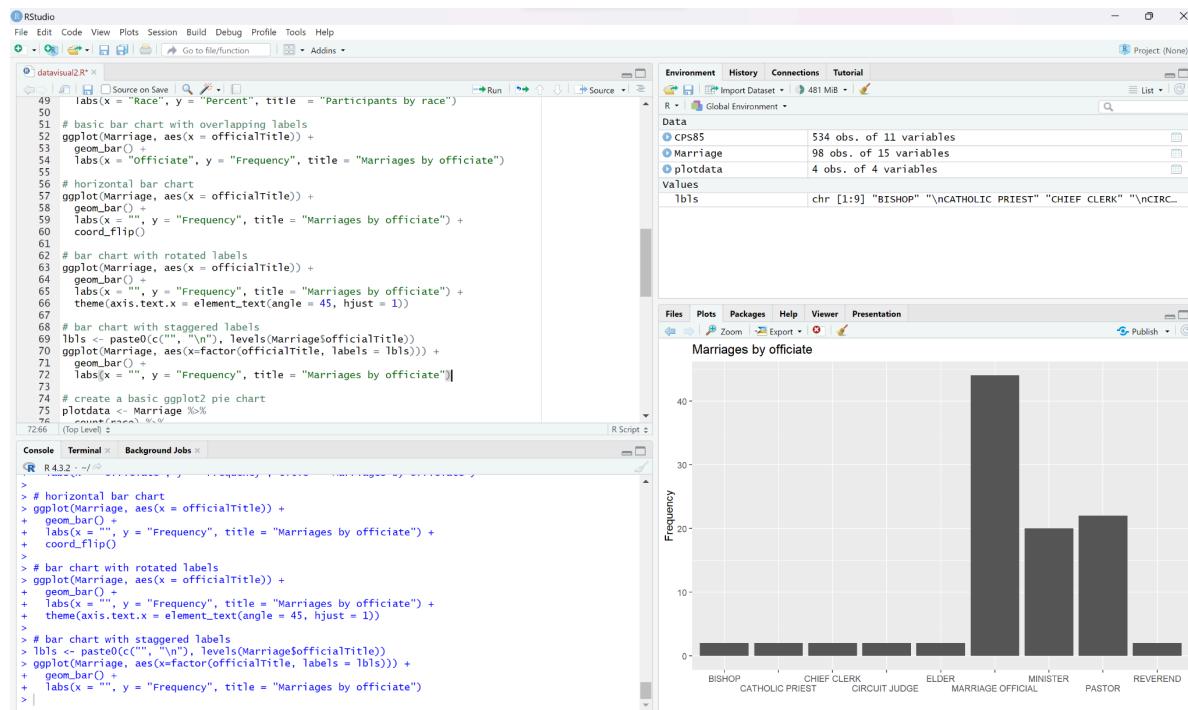
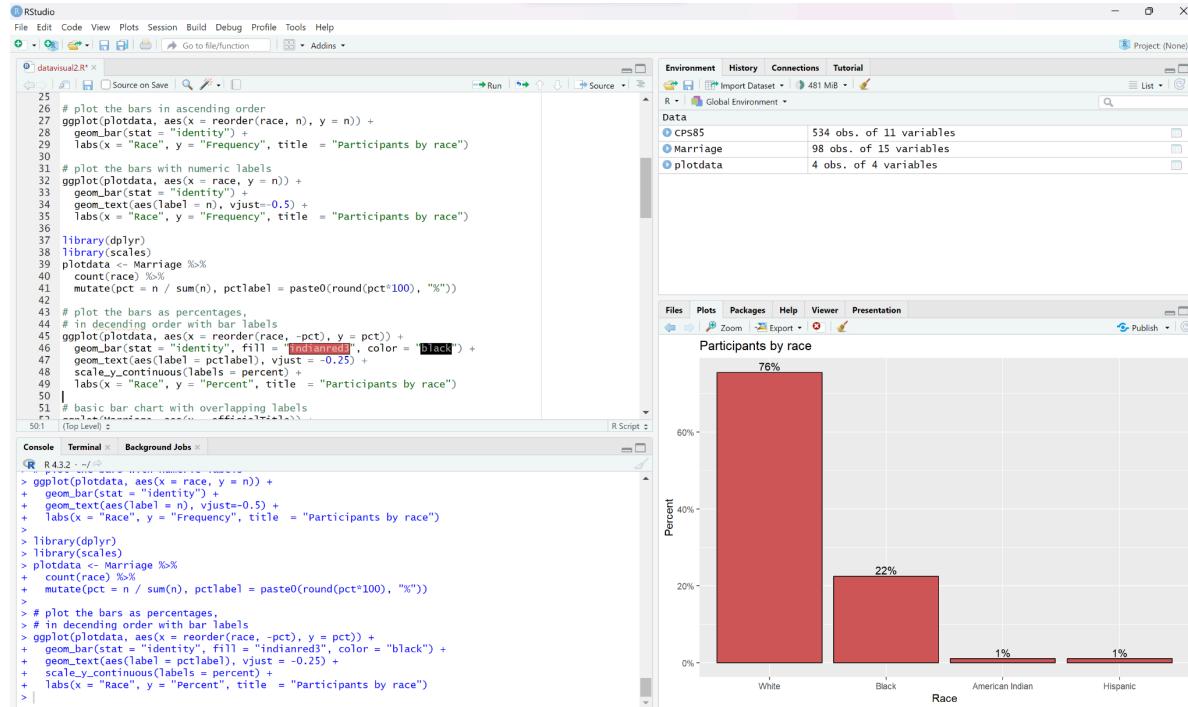
Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Artificial Intelligence and Data Science





**Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088**

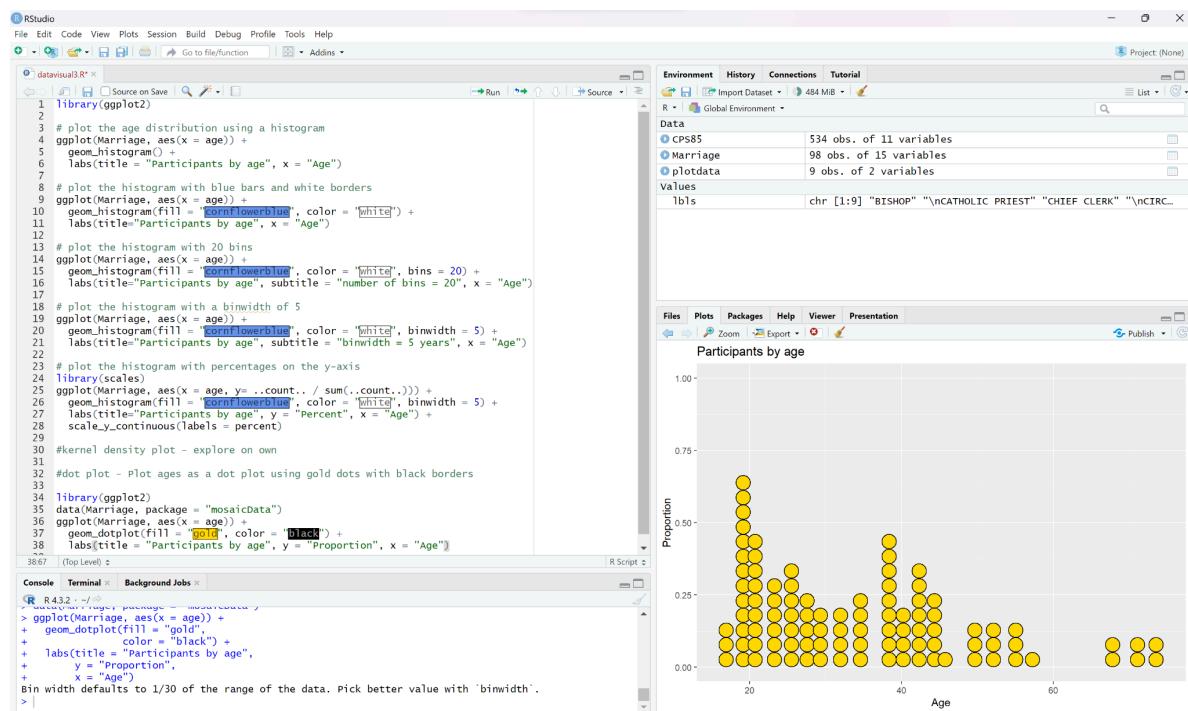
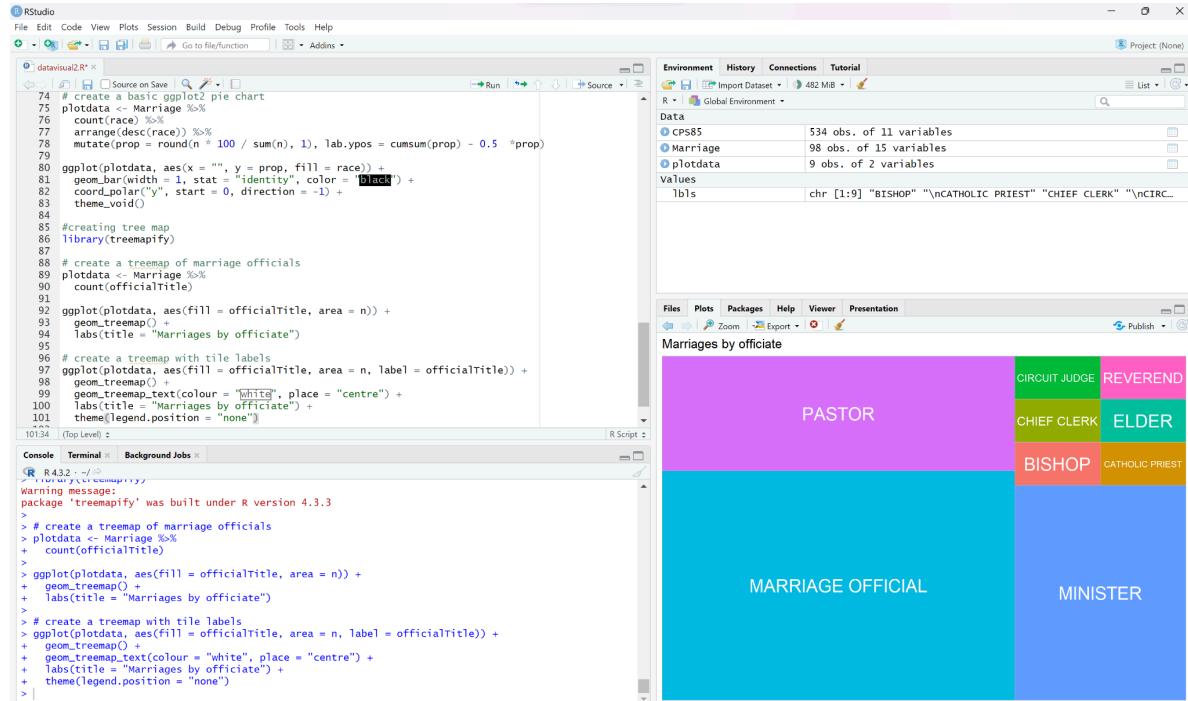
UG Program in Artificial Intelligence and Data Science





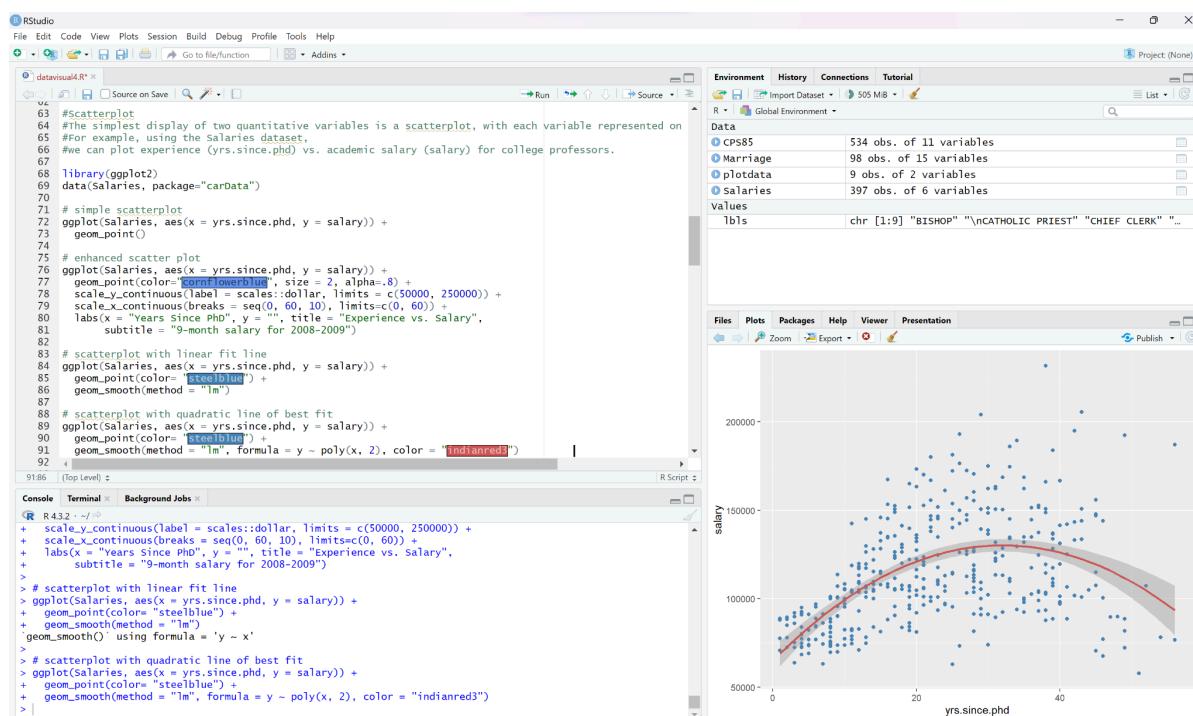
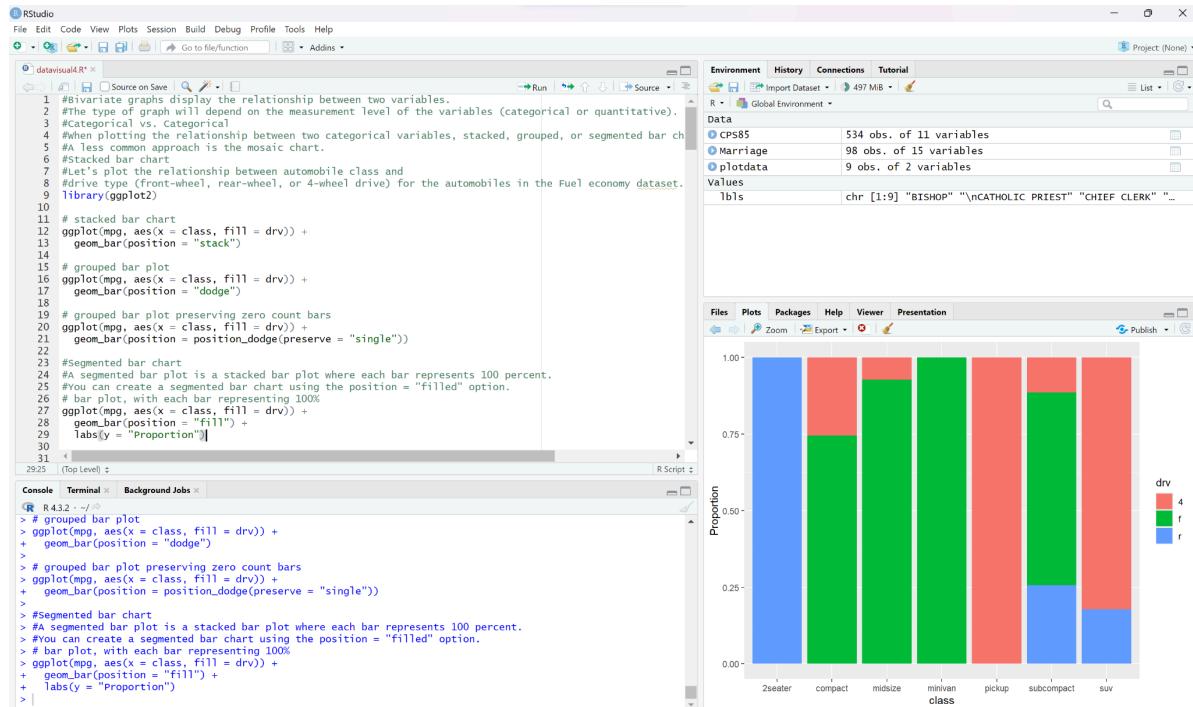
**Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088**

UG Program in Artificial Intelligence and Data Science



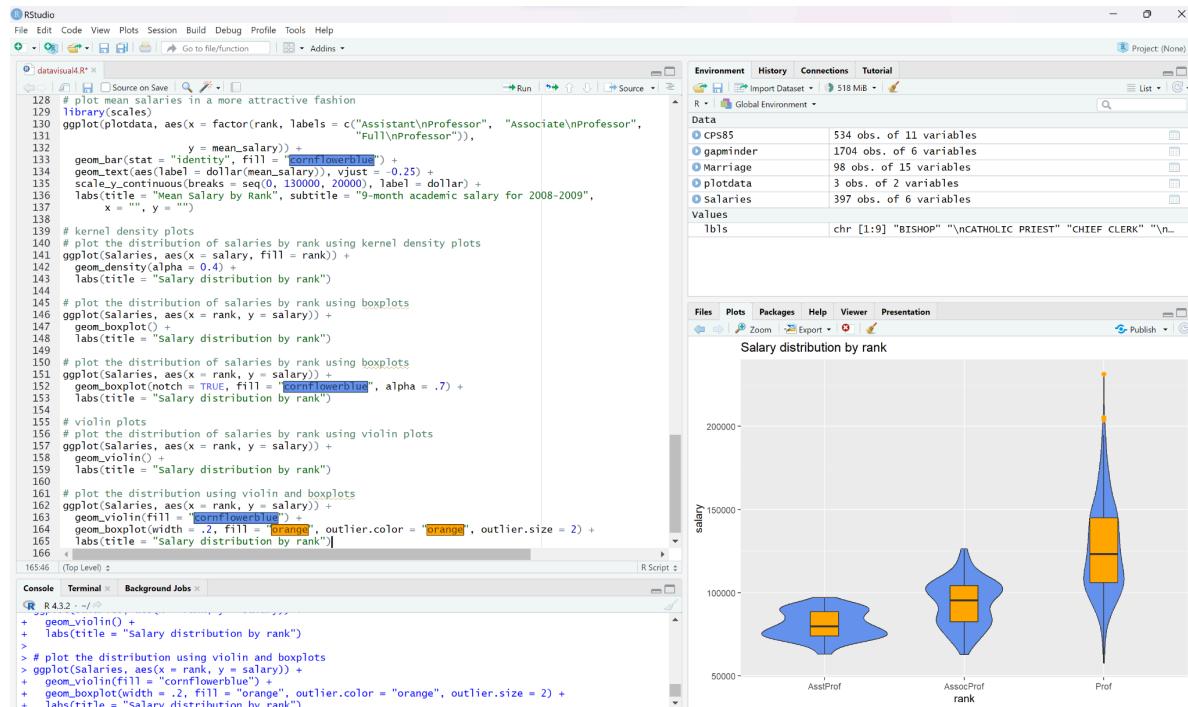
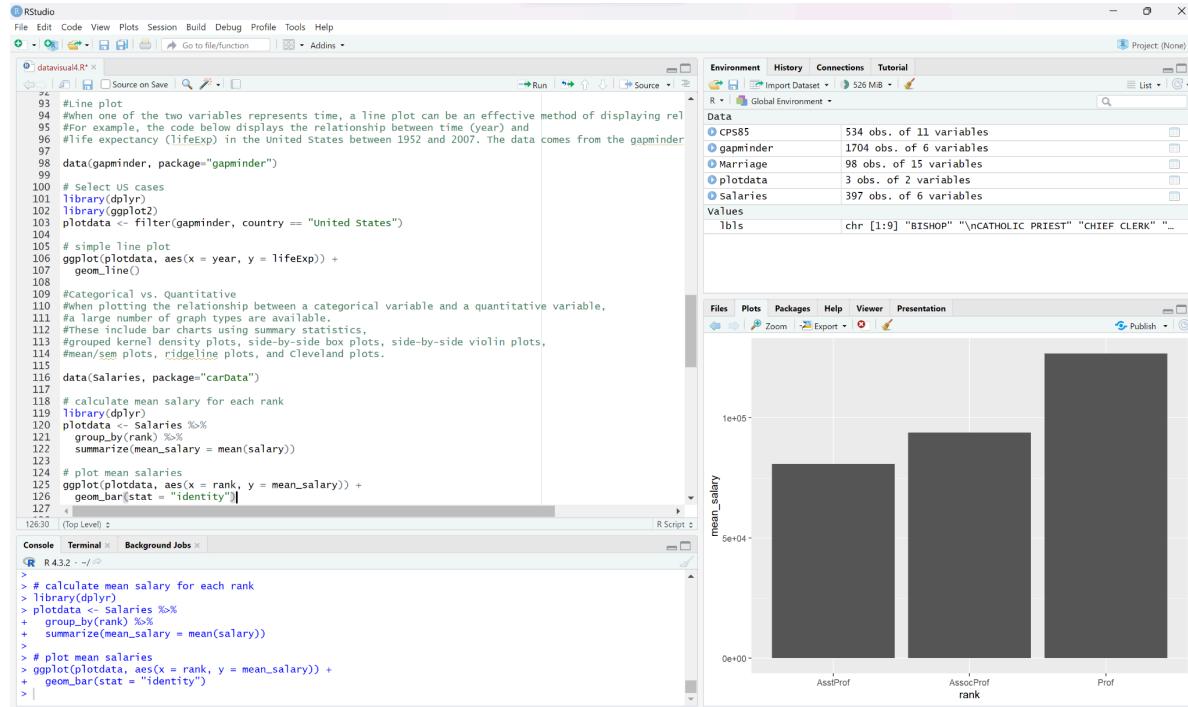


Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Artificial Intelligence and Data Science



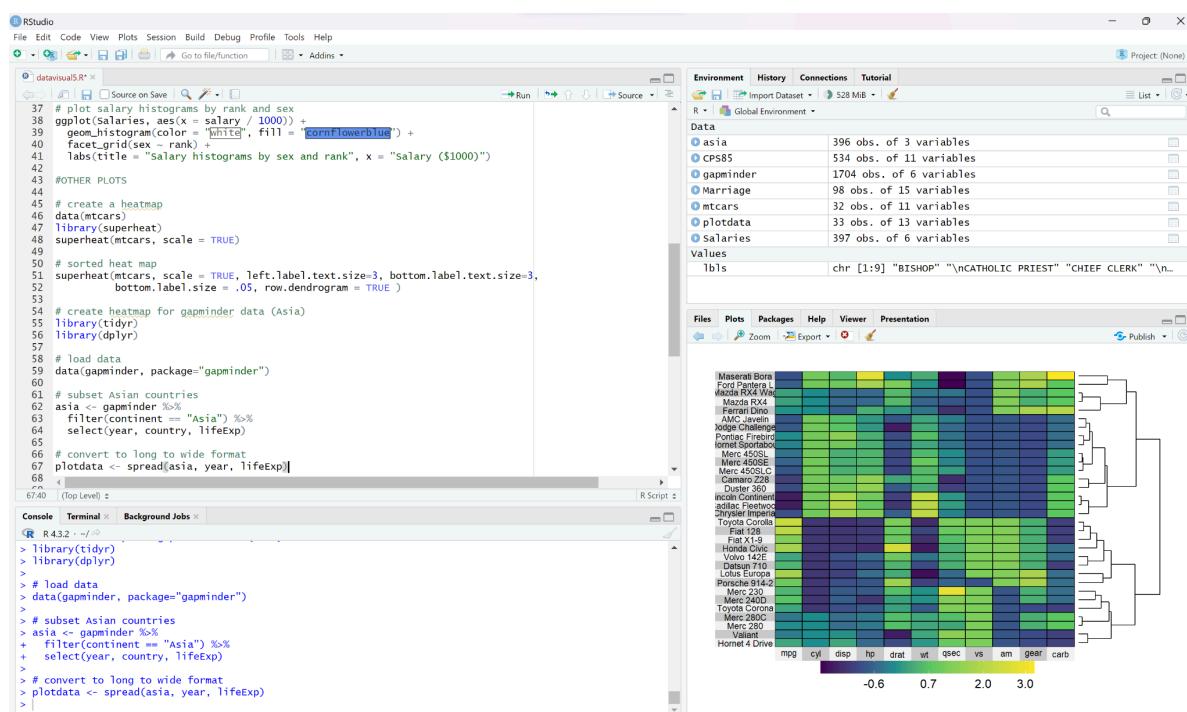
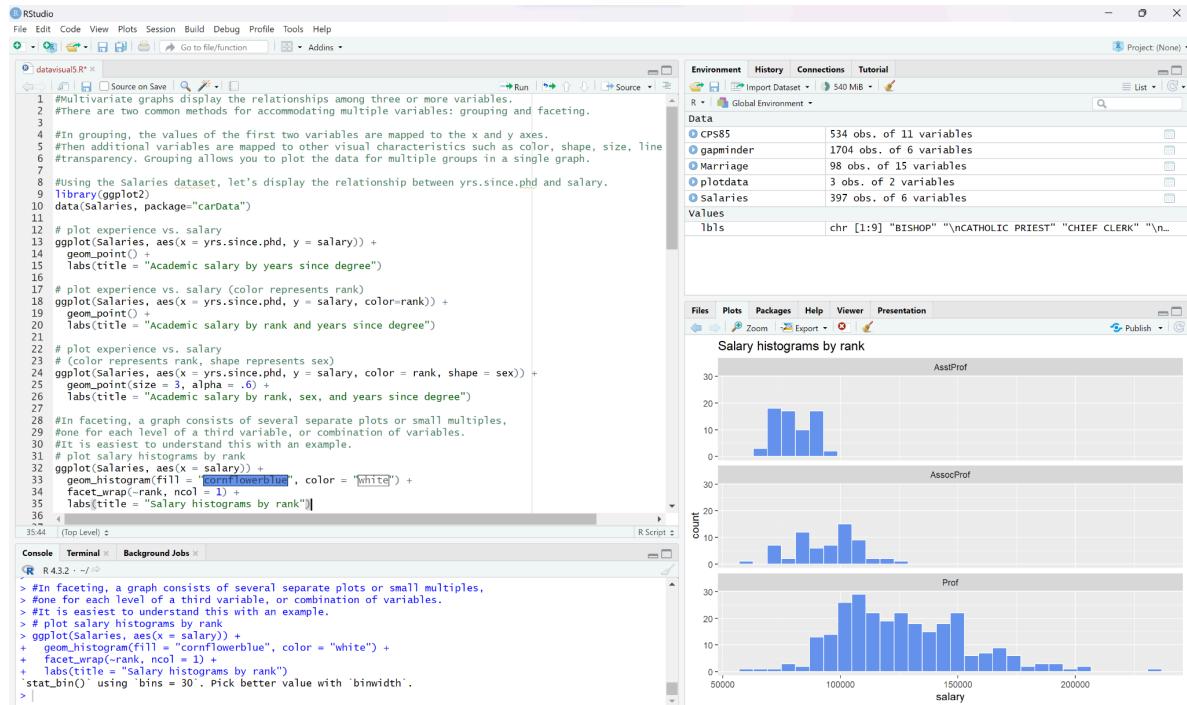


Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Artificial Intelligence and Data Science





Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Artificial Intelligence and Data Science





**Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088**

UG Program in Artificial Intelligence and Data Science

Conclusion: Hence we successfully implemented text analytics and Sentiment analysis in python.