EXPT NO : 12	DATA VISUALIZATION
DATE : 16.04.2025	

AIM:

To Implement data visualization in R programming.

Write an R program for Data visualization using Base R and ggplot2. (Download Titanic dataset from kaggle)

```
library(ggplot2)

titanic_data <- read.csv("titanic.csv", stringsAsFactors = TRUE)

titanic_clean <- titanic_data[!is.na(titanic_data$Age) & !is.na(titanic_data$Fare), ]
```

Data Visualization using Base R

Program 1: Bar Chart – Survival Count

1. Create a bar chart showing the count of Survived vs. Not Survived passengers.

2. Use different colors for visualization.

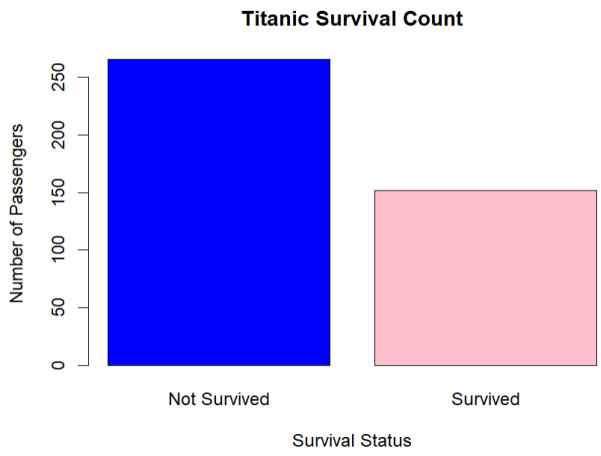
3. Add proper labels and a title.

CODE:

```
survival_count <- table(titanic_data$Survived)

barplot(survival_count,
    col = c("blue", "pink"),
    names.arg = c("Not Survived", "Survived"),
    main = "Titanic Survival Count",
    xlab = "Survival Status",
    ylab = "Number of Passengers")
```

OUTPUT:



Program 2: Line Graph – Age Distribution

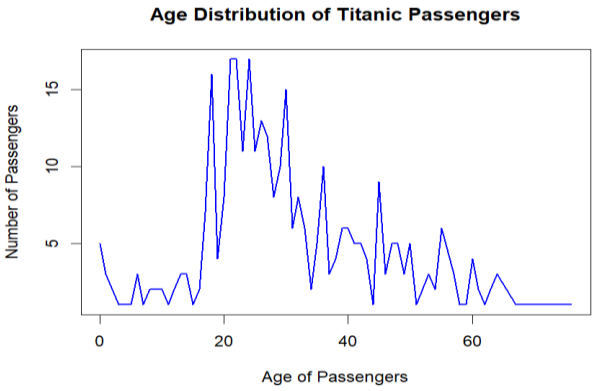
1. Create a line graph showing the distribution of passengers' ages.

2. Use the Age column and adjust the graph's appearance.

CODE:

```
age_data <- na.omit(titanic_data$Age)
age_freq <- table(floor(age_data))
plot(as.numeric(names(age_freq)),
    as.numeric(age_freq),
    type = "l",
    col = "blue",
    lwd = 2,
    xlab = "Age of Passengers",
    ylab = "Number of Passengers",
    main = "Age Distribution of Titanic Passengers")
```

OUTPUT:



Program 3: Scatter Plot – Age vs Fare

1. Create a scatter plot between Age and Fare.

2. Differentiate the points by Survived status (use different colors).

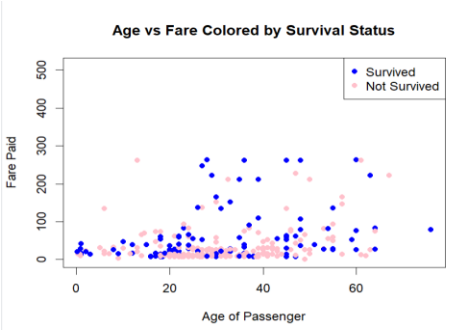
3. Add axis labels, a title, and a legend.

CODE:

```
titanic_clean <- titanic_data[!is.na(titanic_data$Age) & !is.na(titanic_data$Fare), ]
```

```
colors <- ifelse(titanic_clean$Survived == 1, "blue", "pink")
plot(titanic_clean$Age, titanic_clean$Fare,
     col = colors,
     pch = 19,                          # Solid circle points
     xlab = "Age of Passenger",
     ylab = "Fare Paid",
     main = "Age vs Fare Colored by Survival Status")
legend("topright",
       legend = c("Survived", "Not Survived"),
       col = c("blue", "pink"),
       pch = 19)
```

OUTPUT:



Data Visualization using ggplot2

Program 4: Histogram – Age Distribution

1. Create a histogram using ggplot2 to visualize the age distribution of Titanic passengers.

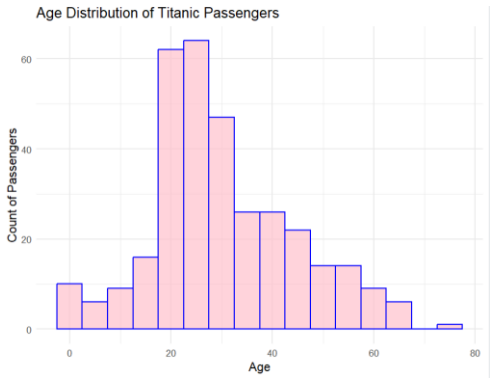
2. Use proper bin width and color customization.

CODE:

```
titanic_clean <- titanic_data[!is.na(titanic_data$Age), ]
ggplot(titanic_clean, aes(x = Age)) +
  geom_histogram(binwidth = 5,
                 fill = "pink",
                 color = "blue",
                 alpha = 0.7) +
  labs(title = "Age Distribution of Titanic Passengers",
```

```
    x = "Age",
    y = "Count of Passengers") +
  theme_minimal()
```

OUTPUT:



Program 5: Box Plot – Fare vs Passenger Class

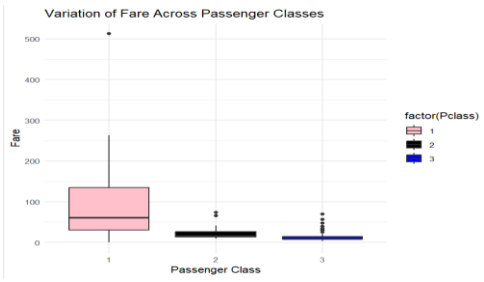
1. Create a box plot showing the variation of Fare across different Pclass (Passenger Class).

2. Customize colors and add appropriate labels.

CODE:

```
ggplot(titanic_data, aes(x = factor(Pclass), y = Fare, fill = factor(Pclass))) +
  geom_boxplot() +
  scale_fill_manual(values = c("pink", "black", "blue")) +
  labs(title = "Variation of Fare Across Passenger Classes",  x = "Passenger Class", y = "Fare") +
  theme_minimal()
```

OUTPUT:



RESULT:

Thus , the R program is implemented and output is verified successfully.