**# 1) Load Titanic dataset**

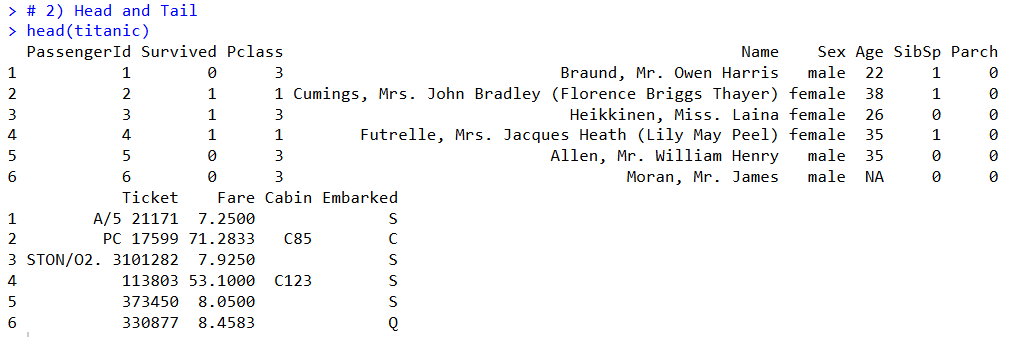
titanic <- read.csv(file.choose(), stringsAsFactors = FALSE)

**# 2) Head and Tail**

head(titanic)

tail(titanic)

**Output :**

****

**# 3) Dimensions**

dim(titanic)

**Output :**

**A close up of a number

AI-generated content may be incorrect.**

**# 4) Structure & Summary**

str(titanic)

summary(titanic)

**Output :**

A computer code on a white background

AI-generated content may be incorrect.

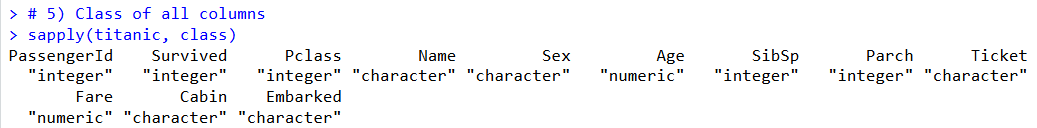
A screenshot of a computer screen

AI-generated content may be incorrect.

**# 5) Class of all columns**

sapply(titanic, class)

**Output :**

****

**# 6) Convert Survived & Sex to factor**

titanic$Survived <- factor(titanic$Survived, labels = c("Not Survived", "Survived"))

titanic$Sex <- factor(titanic$Sex)

**Output :**

**A screen shot of a computer code

AI-generated content may be incorrect.**

**# 7) Convert Pclass, Embarked, SibSp to factor**

titanic$Pclass <- factor(titanic$Pclass)

titanic$Embarked <- factor(titanic$Embarked)

titanic$SibSp <- factor(titanic$SibSp)

**Output :**

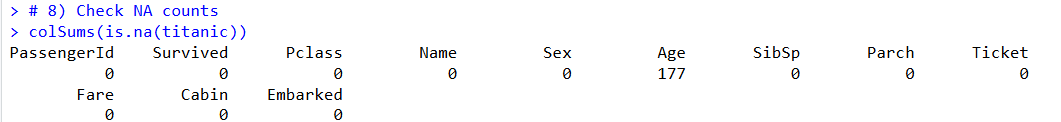
**A close-up of a number

AI-generated content may be incorrect.**

**# 8) Check NA counts**

colSums(is.na(titanic))

**Output :**

****

**# 9) Boxplot of Age**

boxplot(titanic$Age, main="Boxplot of Age", col="lightblue")

**Output :**

**A diagram of a box plot

AI-generated content may be incorrect.**

**# 10) Impute missing Age with median**

titanic$Age[is.na(titanic$Age)] <- median(titanic$Age, na.rm = TRUE)

sum(is.na(titanic$Age))

**Output :**

****

**# 11) Barplot - Pclass**

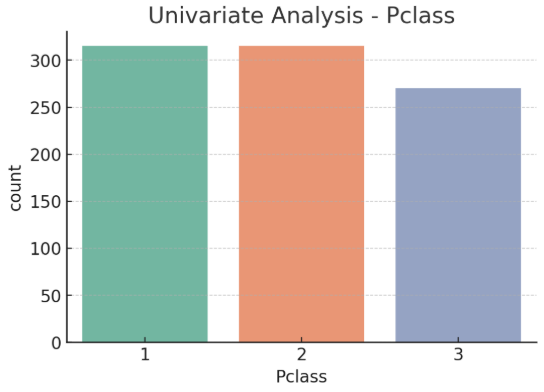
library(ggplot2)

ggplot(titanic, aes(x=Pclass, fill=Pclass)) +

geom\_bar() +

labs(title="Univariate Analysis - Pclass")

**Output :**

****

**# 12) Barplot - Survived**

ggplot(titanic, aes(x=Survived, fill=Survived)) +

geom\_bar() +

labs(title="Univariate Analysis - Survived", y="No. of Passengers")

**Output** :

A graph of a number of people

AI-generated content may be incorrect.

**# 13) Histogram of Age**

ggplot(titanic, aes(x=Age)) +

geom\_histogram(binwidth=5, fill="steelblue", color="black") +

labs(title="Univariate Analysis - Age", x="Age")

**Output :**

**A graph of age and age

AI-generated content may be incorrect.**

**# 14) Barplot - Survived & Sex**

ggplot(titanic, aes(x=Survived, fill=Sex)) +

geom\_bar(position="dodge") +

geom\_text(stat="count", aes(label=..count..),

position=position\_dodge(0.9), vjust=-0.3) +

labs(x="Survived", y="Count", title="Survival by Sex")

**Output** :

A graph with blue and orange bars

AI-generated content may be incorrect.

**# 15) Barplot - Survived & Pclass**

ggplot(titanic, aes(x=Survived, fill=Pclass)) +

geom\_bar(position="dodge") +

geom\_text(stat="count", aes(label=..count..),

position=position\_dodge(0.9), vjust=-0.3) +

labs(x="Survived", y="Count", title="Survival by Pclass")

**Output :**

**A graph of survival by pc class

AI-generated content may be incorrect.**

**# 16) Mean & Median of Age**

mean(titanic$Age, na.rm=TRUE)

median(titanic$Age, na.rm=TRUE)

**Output :**

**A close-up of a white background

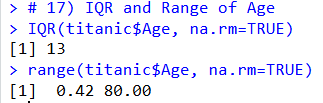
AI-generated content may be incorrect.**

**# 17) IQR and Range of Age**

IQR(titanic$Age, na.rm=TRUE)

range(titanic$Age, na.rm=TRUE)

**Output :**

****

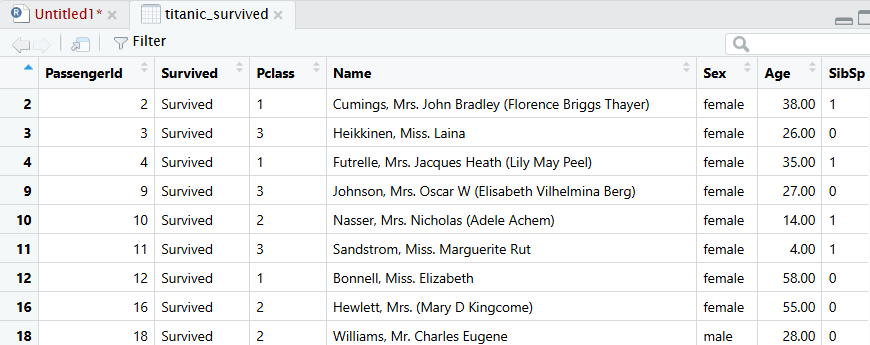
**# 18) Split dataset by Survived**

titanic\_survived <- subset(titanic, Survived=="Survived")

titanic\_notsurvived <- subset(titanic, Survived=="Not Survived")

**Output :**

**“Survived”**



**“Not Survived”**

A screenshot of a computer

AI-generated content may be incorrect.

**# 19) Select Age & Survived (first 10 rows)**

library(dplyr)

titanic %>% select(Age, Survived) %>% head(10)

**Output :**

**A screenshot of a computer

AI-generated content may be incorrect.**

**# 20) Drop unnecessary columns**

library(dplyr)

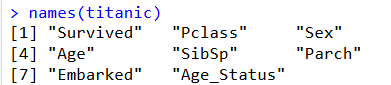
cols\_to\_remove <- c("Name", "Fare", "Ticket", "PassengerId", "Cabin")

existing\_cols <- intersect(names(titanic), cols\_to\_remove)

titanic <- titanic %>% select(-all\_of(existing\_cols))

names(titanic)

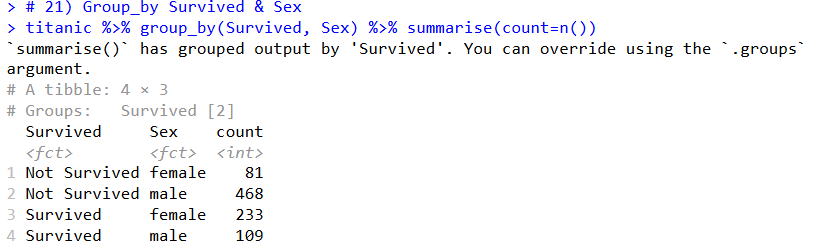
**Output :**

****

**# 21) Group\_by Survived & Sex**

titanic %>% group\_by(Survived, Sex) %>% summarise(count=n())

**Output :**

****

**# 22) Group\_by Sex & Survived**

titanic %>% group\_by(Sex, Survived) %>% summarise(count=n())

**Output :**

**A screenshot of a computer

AI-generated content may be incorrect.**

**# 23) Group\_by Survived, summarise mean age**

titanic %>% group\_by(Survived) %>% summarise(mean\_age=mean(Age, na.rm=TRUE))

**Output :**

**A white background with blue text

AI-generated content may be incorrect.**

**# 24) Mutate new column Age\_Status**

titanic <- titanic %>% mutate(Age\_Status = ifelse(Age < 18, "Minor", "Major"))

head(titanic %>% select(Age, Age\_Status))

**Output :**

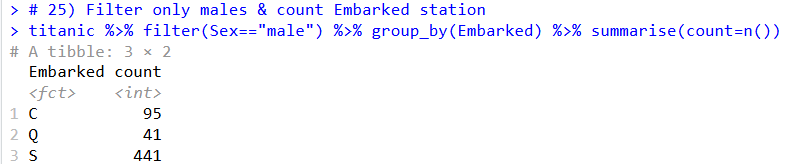
**A computer screen with blue text

AI-generated content may be incorrect.**

**# 25) Filter only males & count Embarked station**

titanic %>% filter(Sex=="male") %>% group\_by(Embarked) %>% summarise(count=n())

**Output :**

****