Customer_Behavior

Name:Koduri gokul

19bcd7006

DataSet :

Mall_Customers.

DATA PREPROCESSING:

Importing Data:

Data Exploration:

Head(Mall_customers)

```
> head(Mall_Customers)
  CustomerID  Genre Age Annual_Income_.k.. Spending_Score
1          1   Male  19                 15             39
2          2   Male  21                 15             81
3          3 Female  20                 16              6
4          4 Female  23                 16             77
5          5 Female  31                 17             40
6          6 Female  22                 17             76
> |
```

Tail(Mall_customers)

```
> tail(Mall_Customers)
    CustomerID  Genre Age Annual_Income_.k.. Spending_Score
195        195 Female  47                120             16
196        196 Female  35                120             79
197        197 Female  45                126             28
198        198   Male  32                126             74
199        199   Male  32                137             18
200        200   Male  30                137             83
> |
```

dim(Mall_Customers)

```
> dim(Mall_Customers)
[1] 200   5
```

str(Mall_Customers)

```
> str(Mall_Customers)
'data.frame':   200 obs. of  5 variables:
 $ CustomerID         : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Genre              : chr  "Male" "Male" "Female" "Female" ...
 $ Age                : int  19 21 20 23 31 22 35 23 64 30 ...
 $ Annual_Income_.k.. : int  15 15 16 16 17 17 18 18 19 19 ...
 $ Spending_Score     : int  39 81 6 77 40 76 6 94 3 72 ...
>
```

Data Cleaning: apply(Mall_Customers, 2, function(row)

sum(is.na(row)))

```
> apply(Mall_Customers, 2, function(row) sum(is.na(row)))
        CustomerID            Genre            Age Annual_Income_.k..     Spending_Score
                 0                0              0                  0                  0
> |
```

Data Transformation

Mall_Customers <- Mall_Customers %>%

+   mutate(Age.Group = case_when(Age <= 35 ~ "Young Adult", Age <= 55 ~ "Middle Age Adult", Age > 55 ~ "Older Adult"))

```
> head(Mall_Customers)
  CustomerID  Genre Age Annual_Income_.k.. Spending_Score
1          1   Male  19                 15             39
2          2   Male  21                 15             81
3          3 Female  20                 16              6
4          4 Female  23                 16             77
5          5 Female  31                 17             40
6          6 Female  22                 17             76
> |
```

Data Reduction

sum(duplicated(Mall_Customers$CustomerID))

```
> sum(duplicated(Mall_Customers$CustomerID))
[1] 0
> |
```

	Genre	Age	Annual_Income_.k..	Spending_Score
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6
8	Female	23	18	94
9	Male	64	19	3
10	Female	30	19	72
11	Male	67	19	14
12	Female	35	19	99
13	Female	58	20	15
14	Female	24	20	77
15	Male	37	20	13
16	Male	22	20	79
17	Female	35	21	35
18	Male	20	21	66

DATA VISUALIZATION
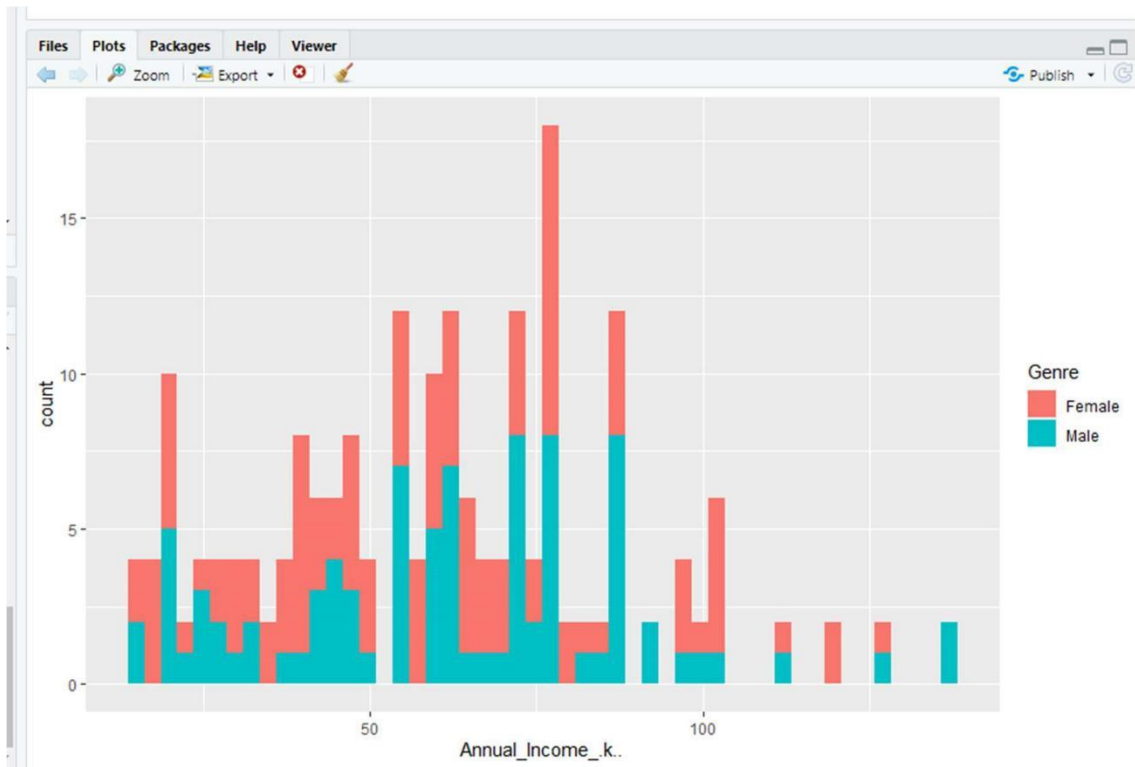
Histograms library(ggplot2)

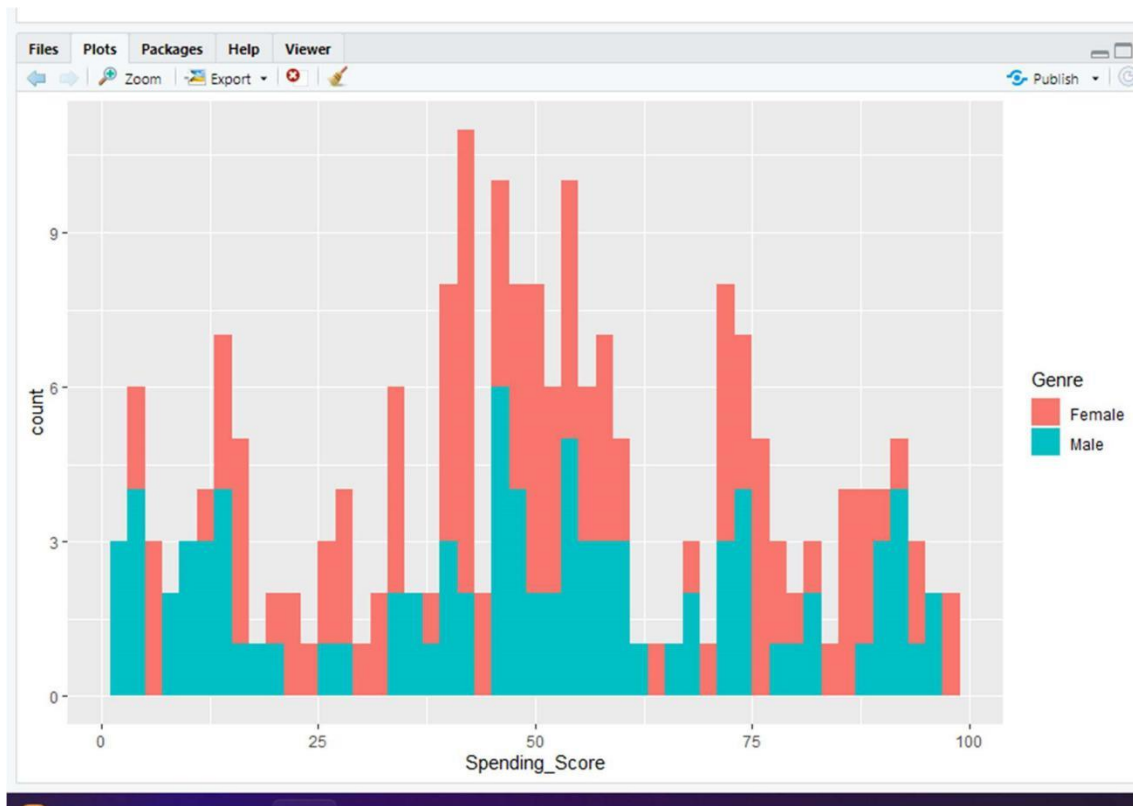> ggplot(Mall_Customers,aes(x= Age, fill=Genre))+geom_histogram(bins = 50)

ggplot(Mall_Customers,aes(x= `Annual_Income_.k..`,fill=Genre))
+geom_histogram(bins = 50)

ggplot(Mall_Customers,aes(x= `Spending_Score`,fill=Genre))
+geom_histogram(bins=50)

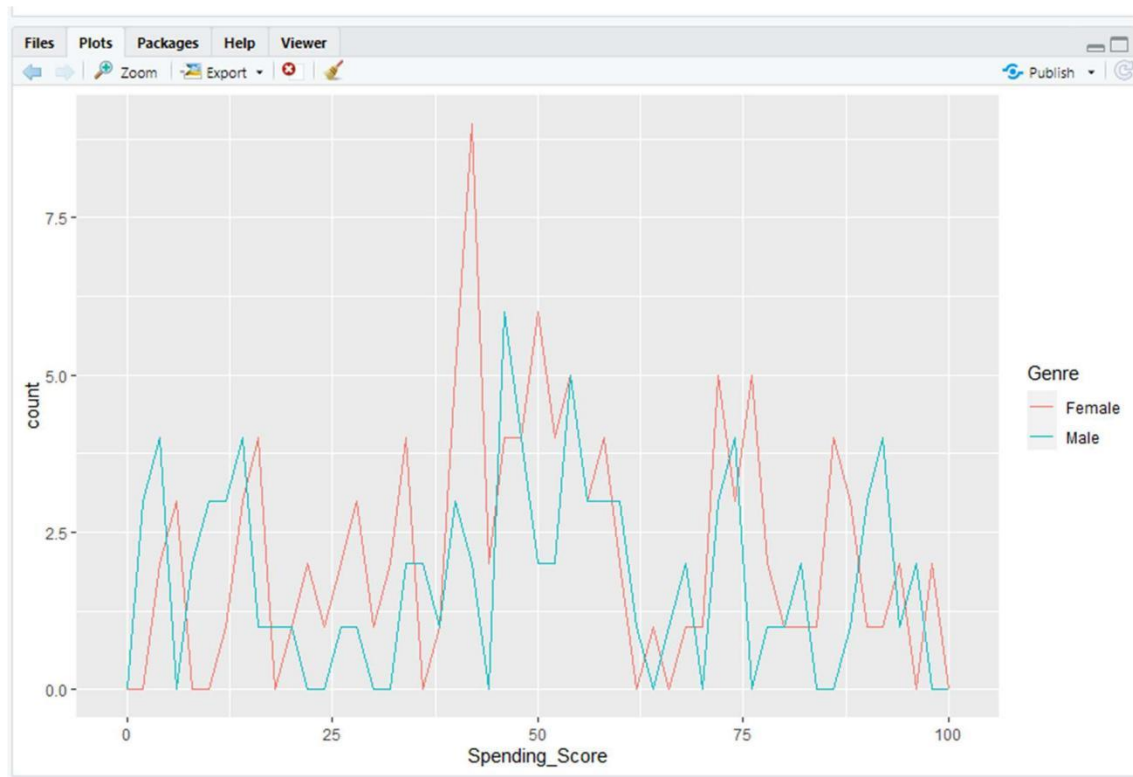Bar plot: ggplot(Mall_Customers,aes(x=

Genre))+geom_bar ()

Frequency Polygon

library(ggplot2)

> ggplot(Mall_Customers,aes(x= `Spending_Score`, col=Genre)) +
geom_freqpoly(bins=50)

Box Plot:
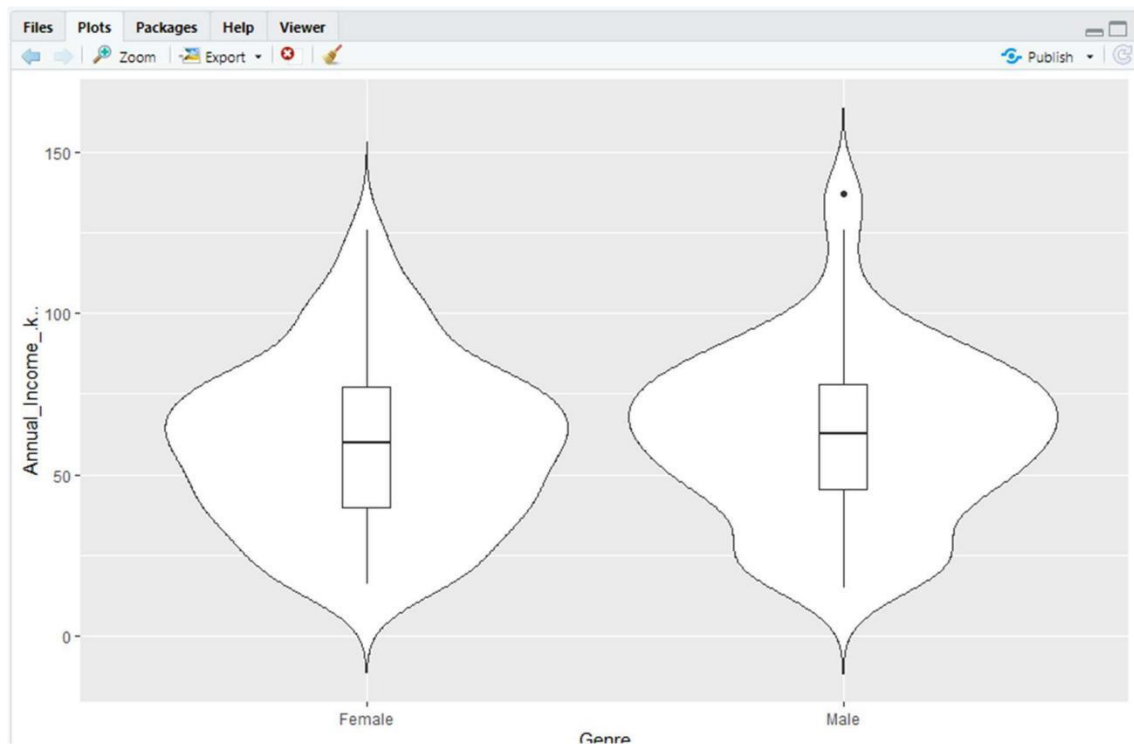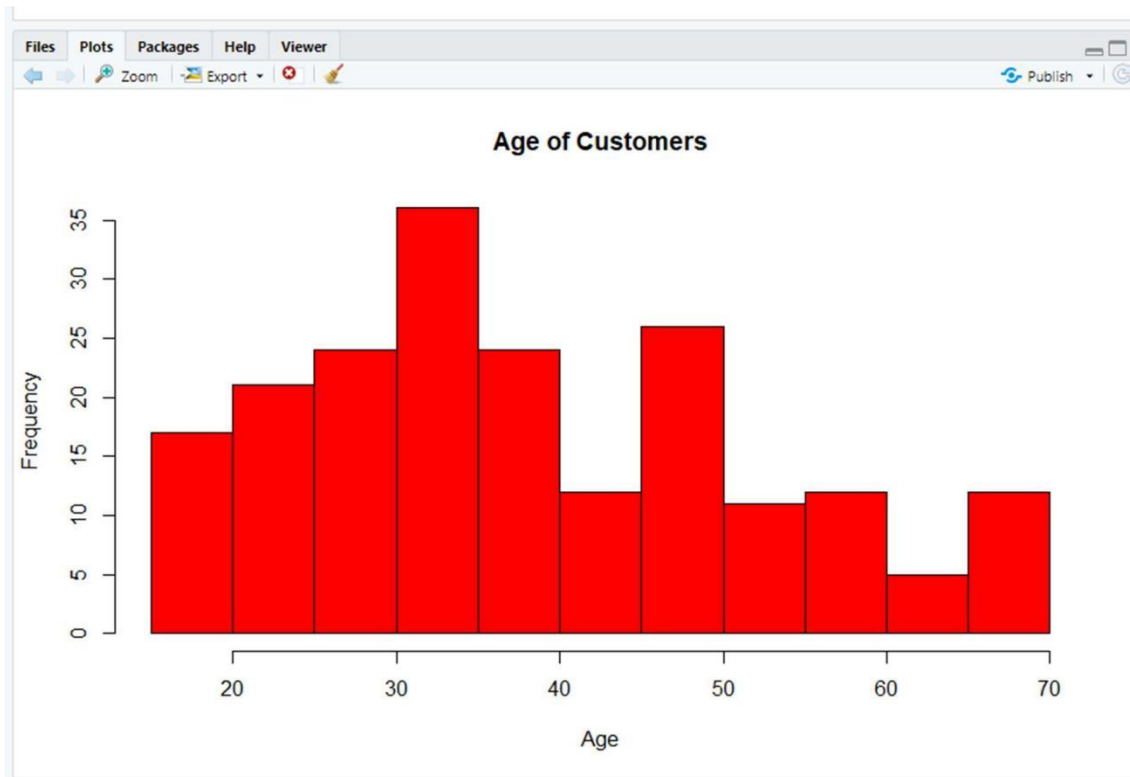
library(ggplot2)

> p<-ggplot(Mall_Customers,aes(y= `Annual_Income_.k..`, x= Genre))+geom_violin(trim=FALSE)

> p + geom_boxplot(width=0.1)

Data Analysis

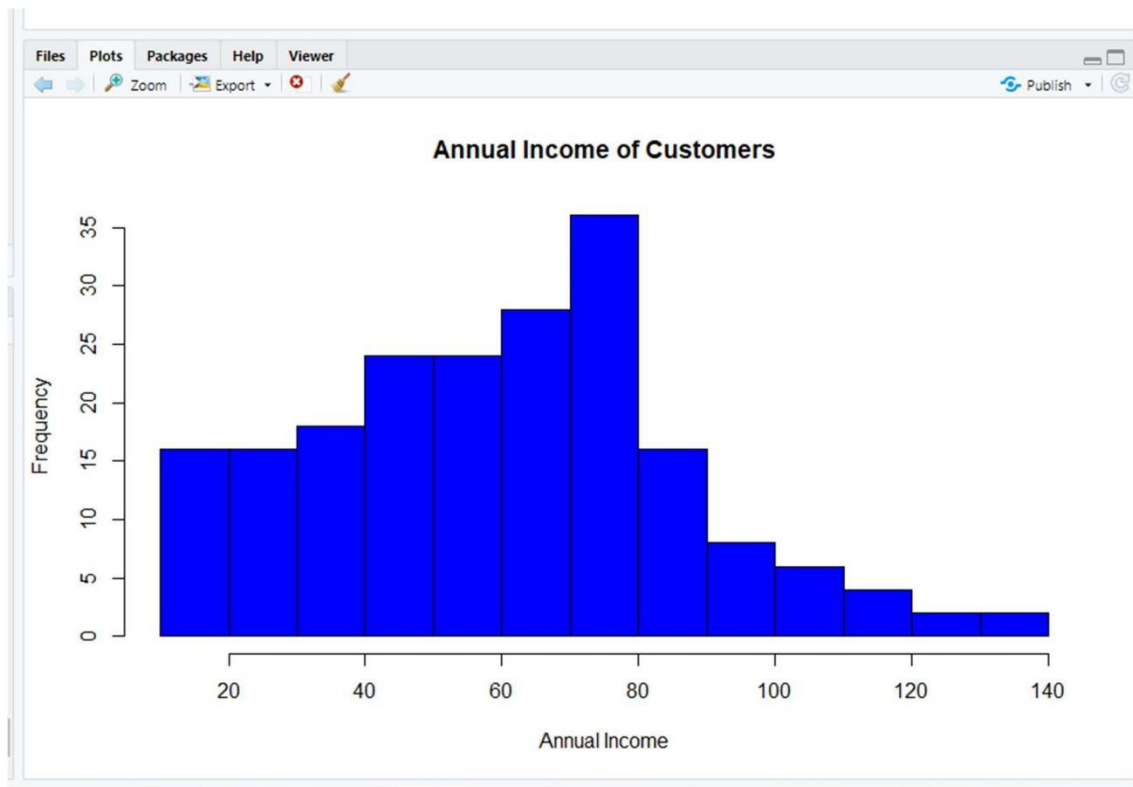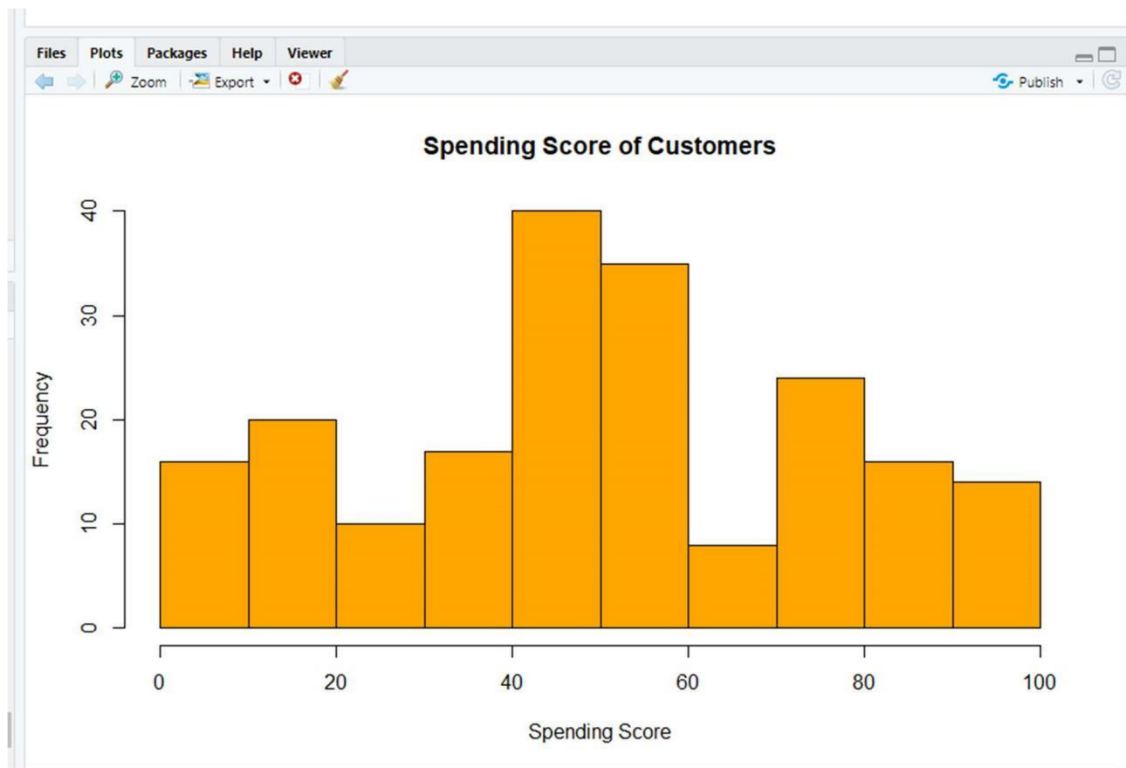hist(Mall_Customers$Age,

    col = 'red',

    main = 'Age of Customers',

xlab = 'Age')

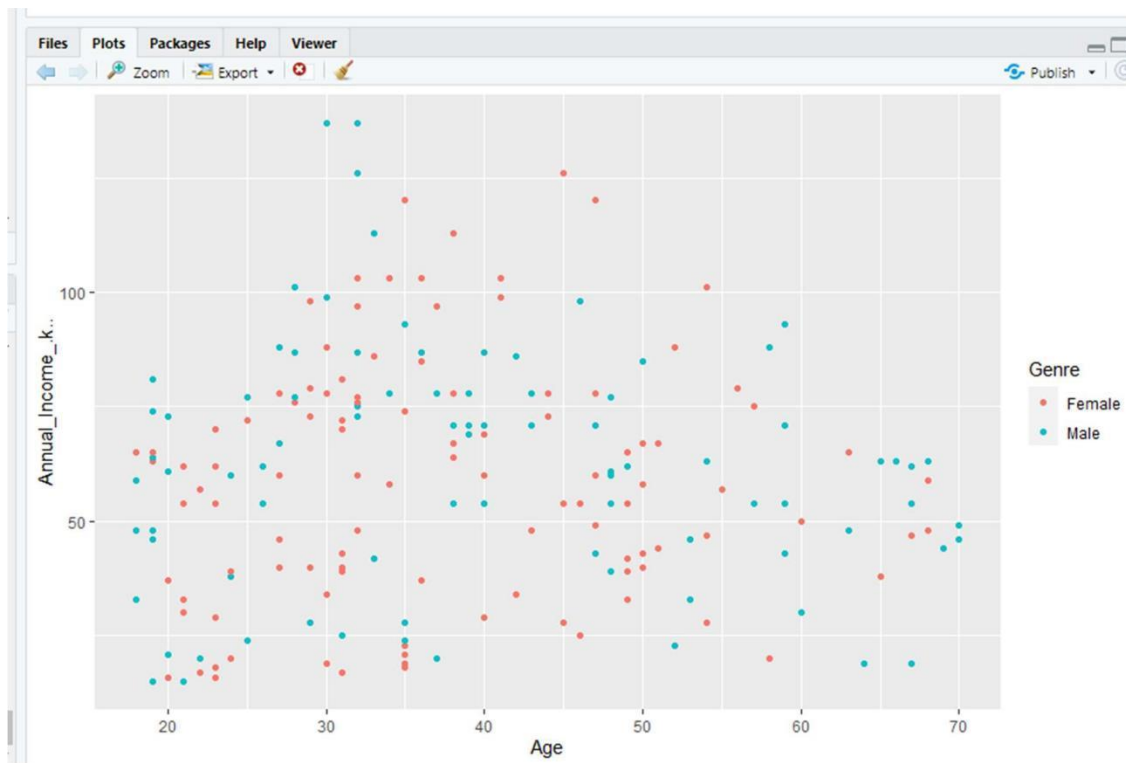**Age of Customers**

hist(Mall_Customers$Annual_Income_.k..,

+   col = 'blue',

+   main = 'Annual Income of Customers',

+   xlab = 'Annual Income')

Annual Income of Customers

```
hist(Mall_Customers$Spending_Score,
+     col = 'orange',
+     main = 'Spending Score of Customers',
+     xlab = 'Spending Score')
```

Spending Score of Customers

> ggplot(Mall_Customers) +

+   geom_point(aes(x = Age, y = Annual_Income_.k.., col = Genre))

ggplot(Mall_Customers) +

+ geom_point(aes(x = Annual_Income_.k.., y = Spending_Score, col = Genre))