


Nama: Francisco NIM: 064002300044	 Praktikum Probabilitas & Statistika	MODUL 6 Nama Dosen: Drs. Joko Riyono, M.Si.
Hari/Tanggal: sabtu, 04-05-2024		Nama Asisten Labratorium: 1. Adzriel Yusak Noah (064.20.11) 2. Mohammad Akhmal Firdaus (064.20.34)

Explorasi Data

1. Teori Singkat

Boxplot

Box plot atau boxplot (juga dikenal sebagai diagram box-and-whisker) merupakan suatu teknik grafikal dalam statistik deskriptif untuk menggambarkan secara grafik dari data numerik melalui lima ukuran sebagai berikut:

- Nilai observasi terkecil (minimum)
- Kuartil terendah atau kuartil pertama (Q1), yang memotong 25 % dari data terendah
- Median (Q2) atau nilai pertengahan
- Kuartil tertinggi atau kuartil ketiga (Q3), yang memotong 25 % dari data tertinggi
- Nilai observasi terbesar (maksimum)

Histogram

Histogram adalah salah satu grafik statistik untuk mengevaluasi bentuk dan sebaran dari data numerik. Histogram membagi nilai sampel menjadi beberapa interval yang disebut bins. Batangnya menggambarkan jumlah pengamatan (frekuensi) yang jatuh di dalam bin.

|

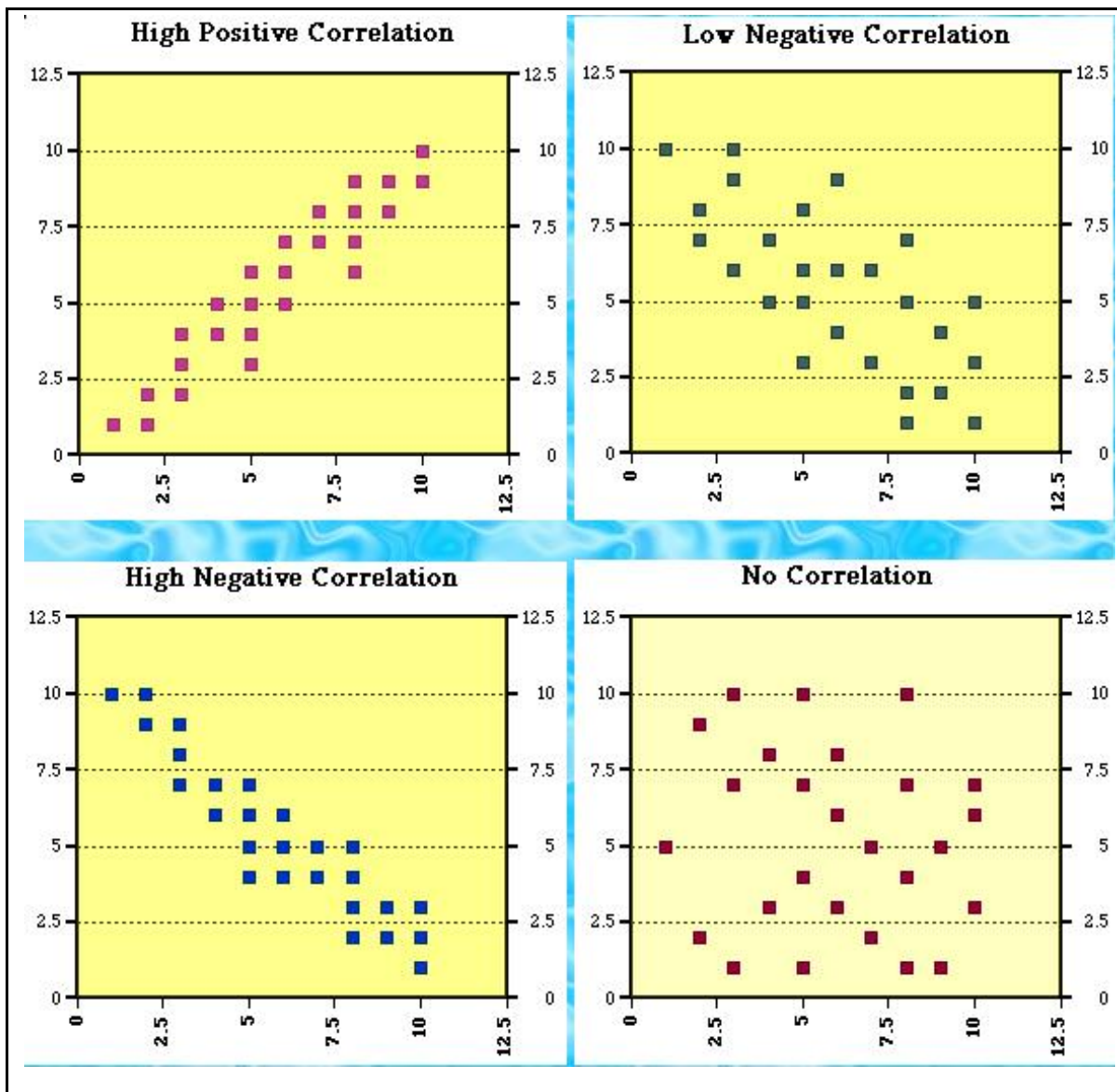
Scatter Plot

Francsico
064002300044

|

|

Scatter plot dapat digunakan untuk melihat hubungan antara dua variabel numeric berdasarkan dari pola tebaran titiknya.



2. Alat dan Bahan

Hardware : Laptop/PC

Software : R Studio

3. Elemen Kompetensi

Data yang digunakan: Iris.csv

a. Latihan pertama – Boxplot

1. Script

```
summary(iris)
```

Output:

```
> summary(iris)
   Sepal.Length      Sepal.width
Min.   :4.300      Min.   :2.000
1st Qu.:5.100      1st Qu.:2.800
Median :5.800      Median :3.000
Mean   :5.843      Mean   :3.057
3rd Qu.:6.400      3rd Qu.:3.300
Max.   :7.900      Max.   :4.400
   Petal.Length      Petal.width
Min.   :1.000      Min.   :0.100
1st Qu.:1.600      1st Qu.:0.300
Median :4.350      Median :1.300
Mean   :3.758      Mean   :1.199
```

2. Script

```
head(iris)
```

Output:

```
> head(iris)
   Sepal.Length Sepal.width
1           5.1           3.5
2           4.9           3.0
3           4.7           3.2
4           4.6           3.1
5           5.0           3.6
6           5.4           3.9
   Petal.Length Petal.width species
1           1.4           0.2  setosa
2           1.4           0.2  setosa
3           1.3           0.2  setosa
4           1.5           0.2  setosa
5           1.4           0.2  setosa
6           1.5           0.2  setosa
```

3. Script

```
by(iris$Petal.Length, iris$Species, summary)
```

|

Francsico
064002300044

|

|

Output:

```
> by(iris$Petal.Length, iris$Species, summary)
iris$Species: setosa
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  1.400   1.500   1.462  1.575   1.900
-----
iris$Species: versicolor
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 3.00   4.00   4.35   4.26   4.60   5.10
-----
iris$Species: virginica
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 4.500  5.100   5.550   5.552  5.875   6.900
>
```

4. Script

```
mean(iris$Sepal.Length)
```

Output:

```
> mean(iris$Sepal.Length)
[1] 5.843333
```

5. Script

```
var(iris$Sepal.Length)
```

Output:

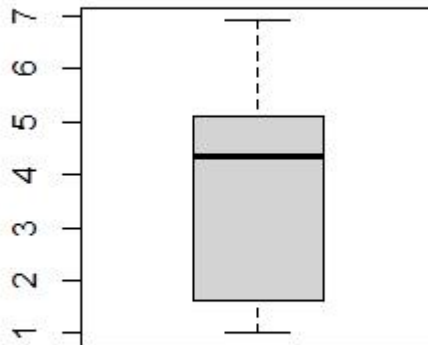
```
> var(iris$Sepal.Length)
[1] 0.6856935
```

6. Script

```
boxplot(iris$Petal.Length)
```

Output:

```
> boxplot(iris$Petal.Length)
```

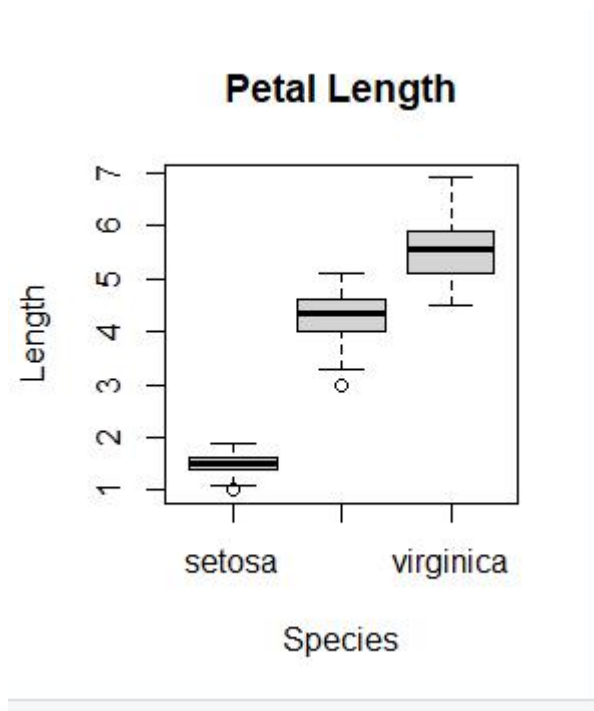


7. Script

```
boxplot(Petal.Length ~ Species, data=iris, main="Petal Length", xlab="Species",  
ylab="Length")
```

Output:

```
> boxplot(Petal.Length ~ species, data=  
iris, main="Petal Length", xlab="Specie  
s", ylab="Length")
```



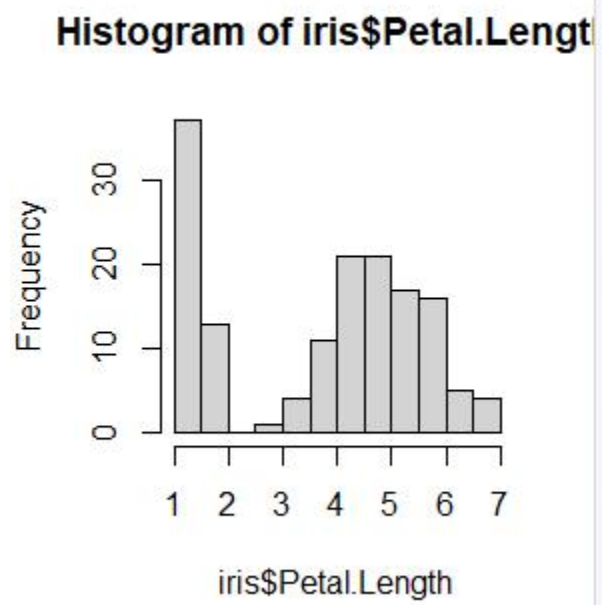
b. Latihan Kedua – Histogram & Destiny

1. Script

```
hist(iris$Petal.Length)
```

Output:

```
> hist(iris$Petal.Length)
```

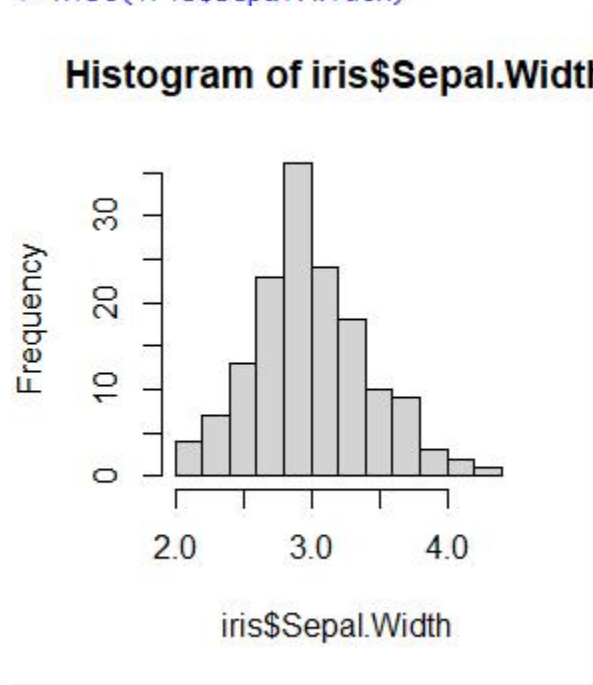



2. Script

```
hist(iris$Sepal.Width)
```

Output:

```
> hist(iris$Sepal.Width)
```

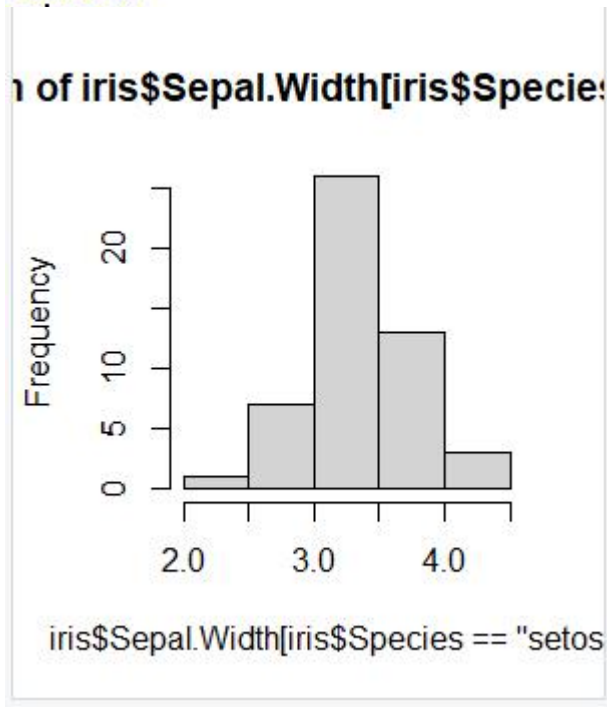


3. Script

```
hist(iris$Sepal.Width[iris$Species=="setosa"])
```

Output:

```
> hist(iris$Sepal.Width[iris$Species=="setosa"])
```



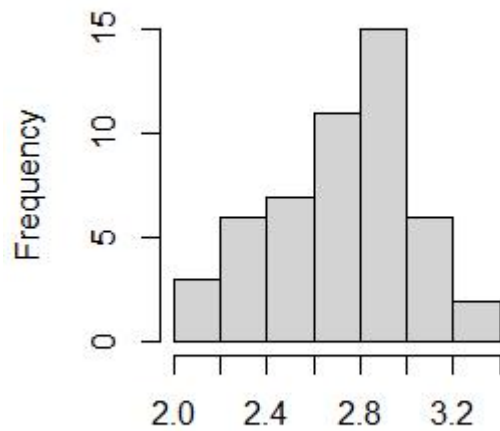
4. Script

```
hist(iris$Sepal.Width[iris$Species=="versicolor"])
```

Output:

```
> hist(iris$Sepal.Width[iris$Species=="versicolor"])
```

of iris\$Sepal.Width[iris\$Species == "versicol"]



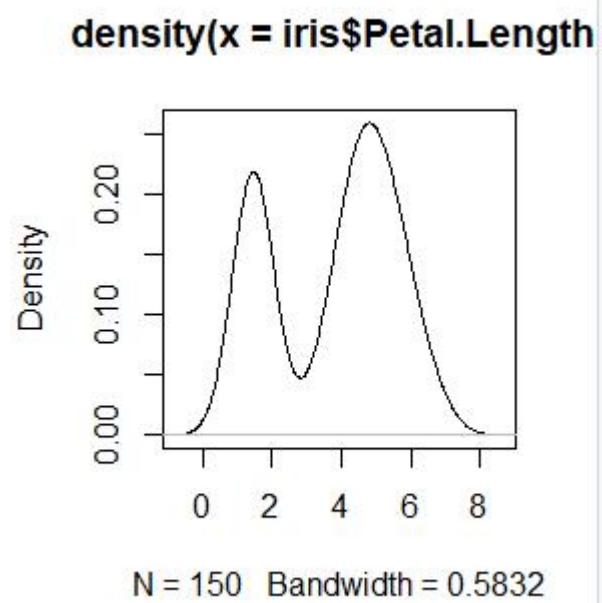
iris\$Sepal.Width[iris\$Species == "versicol"]

5. Script

```
d=density(iris$Petal.Length)
plot(d)
```

Output:

```
> d=density(iris$Petal.Length)
> plot(d)
```



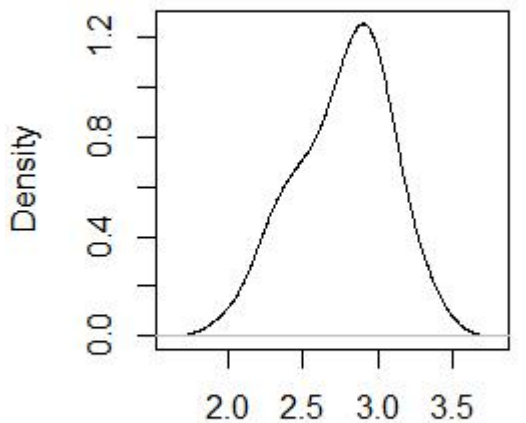
6. Script

```
d=density(iris$Sepal.Width[iris$Species=="versicolor"])  
plot (d)
```

Output:

```
> d=density(iris$sepal.width[iris$species=="versicolor"])  
> plot(d)
```

iris\$Sepal.Width[iris\$Species ==



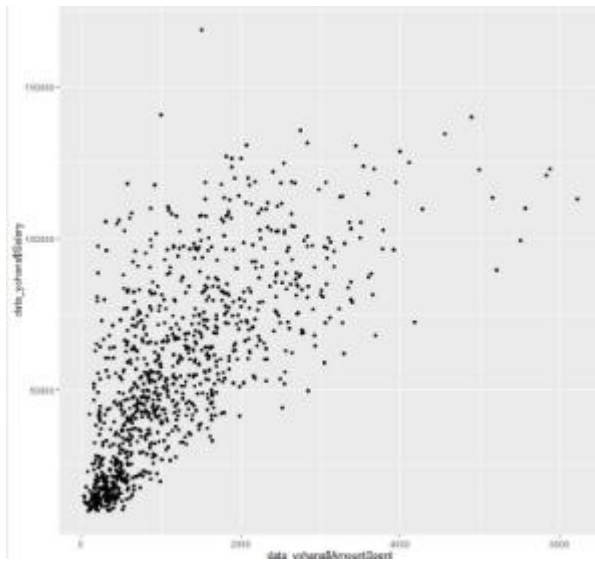
c. Latihan Ketiga – Scatter Plot

1. Script

```
qplot(data_namapraktikan$AmountSpent, data_namapraktikan$Salary)
```

Output:

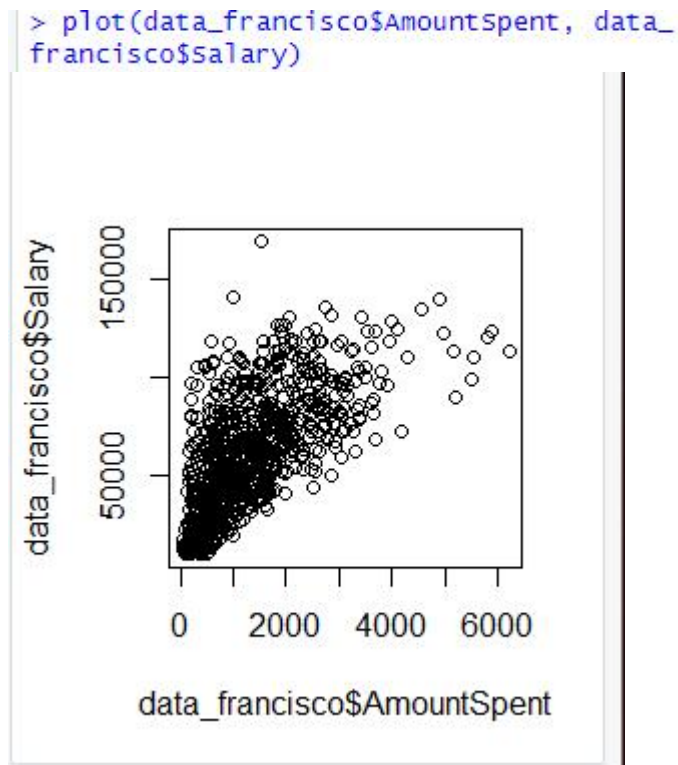
```
> data_francisco=read.csv("C:/Users/Francisco/Downloads/DirectMarketing - Direct Marketing.csv")  
> qplot(data_francisco$AmountSpent, data_francisco$Salary)
```



2. Script

```
plot(data_namapraktikan$AmountSpent, data_namapraktikan$Salary)
```

Output:



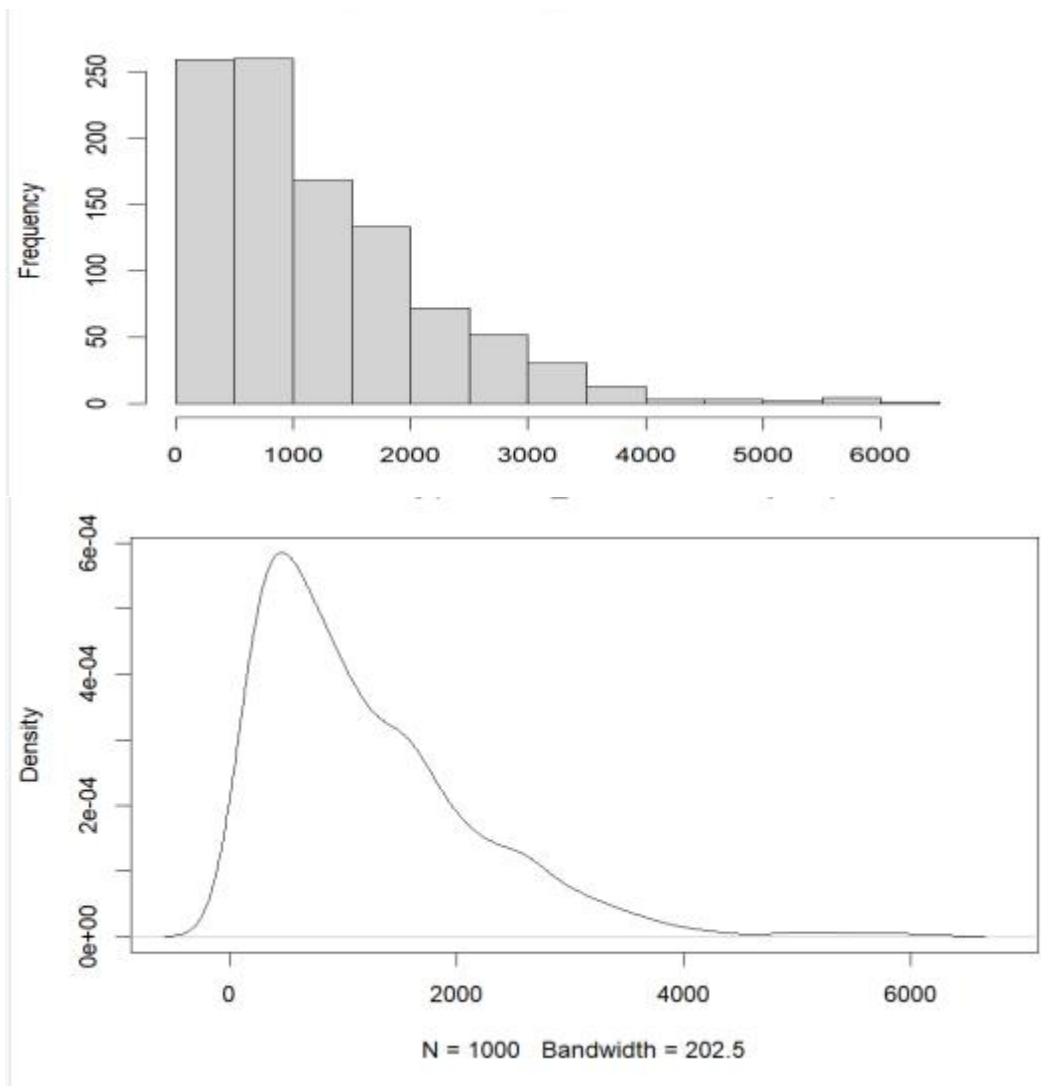
d. Latihan Keempat – Tugas

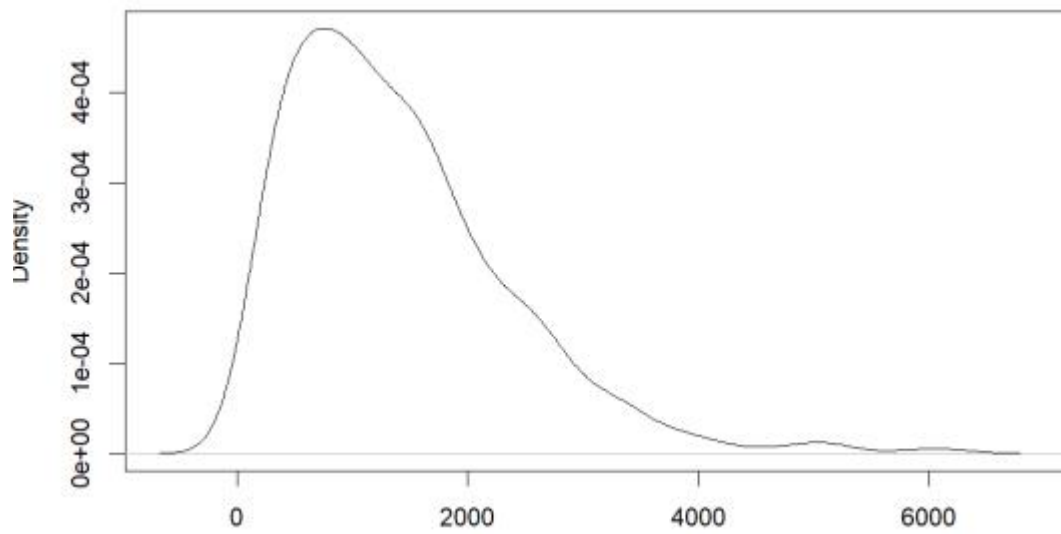
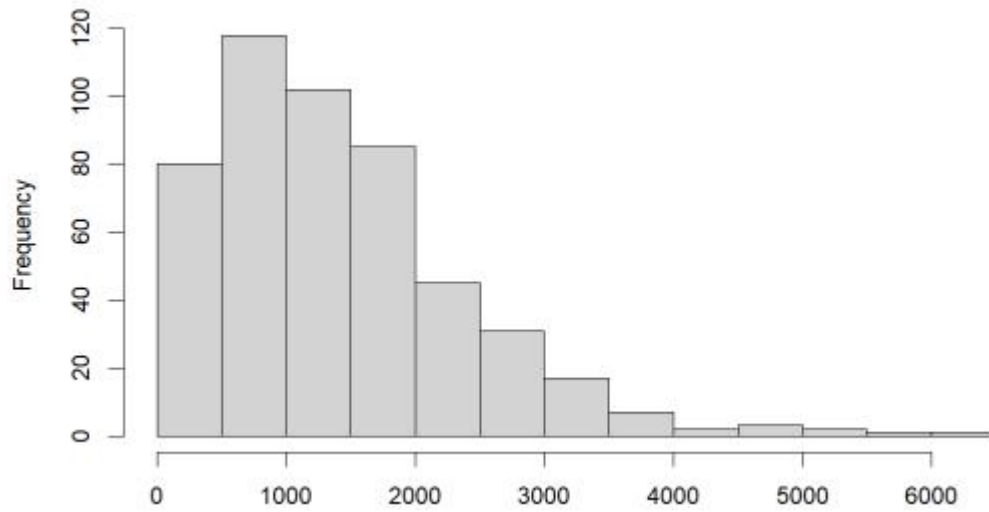
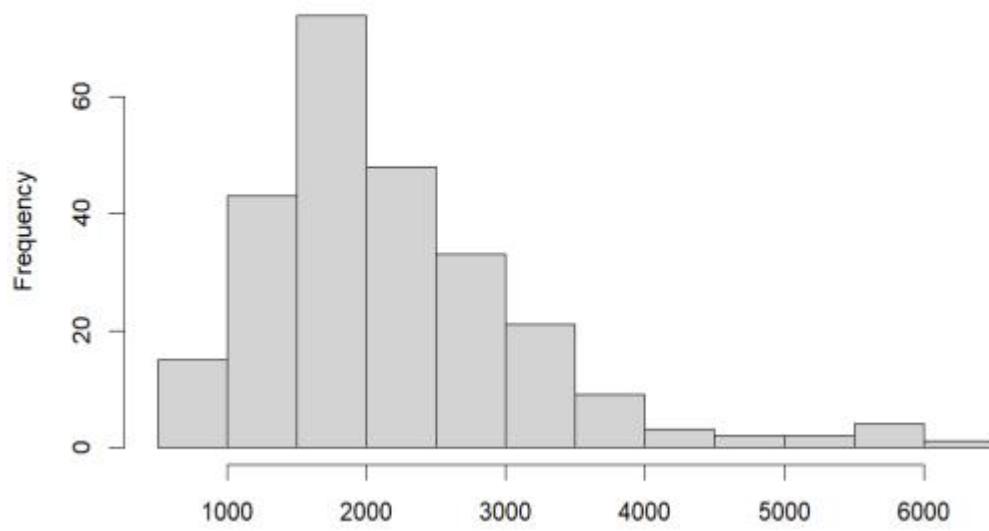
1. Gunakan data DirectMarketing.csv. Lakukan analisis dengan menggunakan histogram dan density plot untuk mengetahui faktor-faktor yang mempengaruhi pengeluaran (AmountSpent)

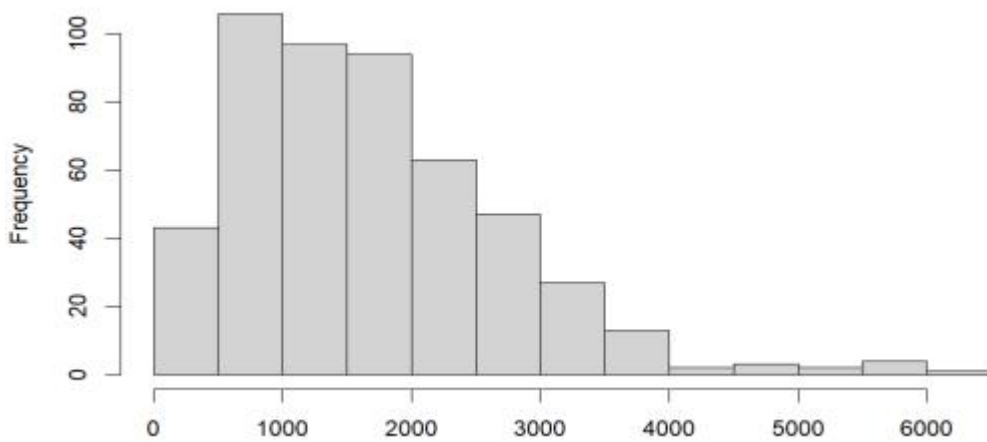
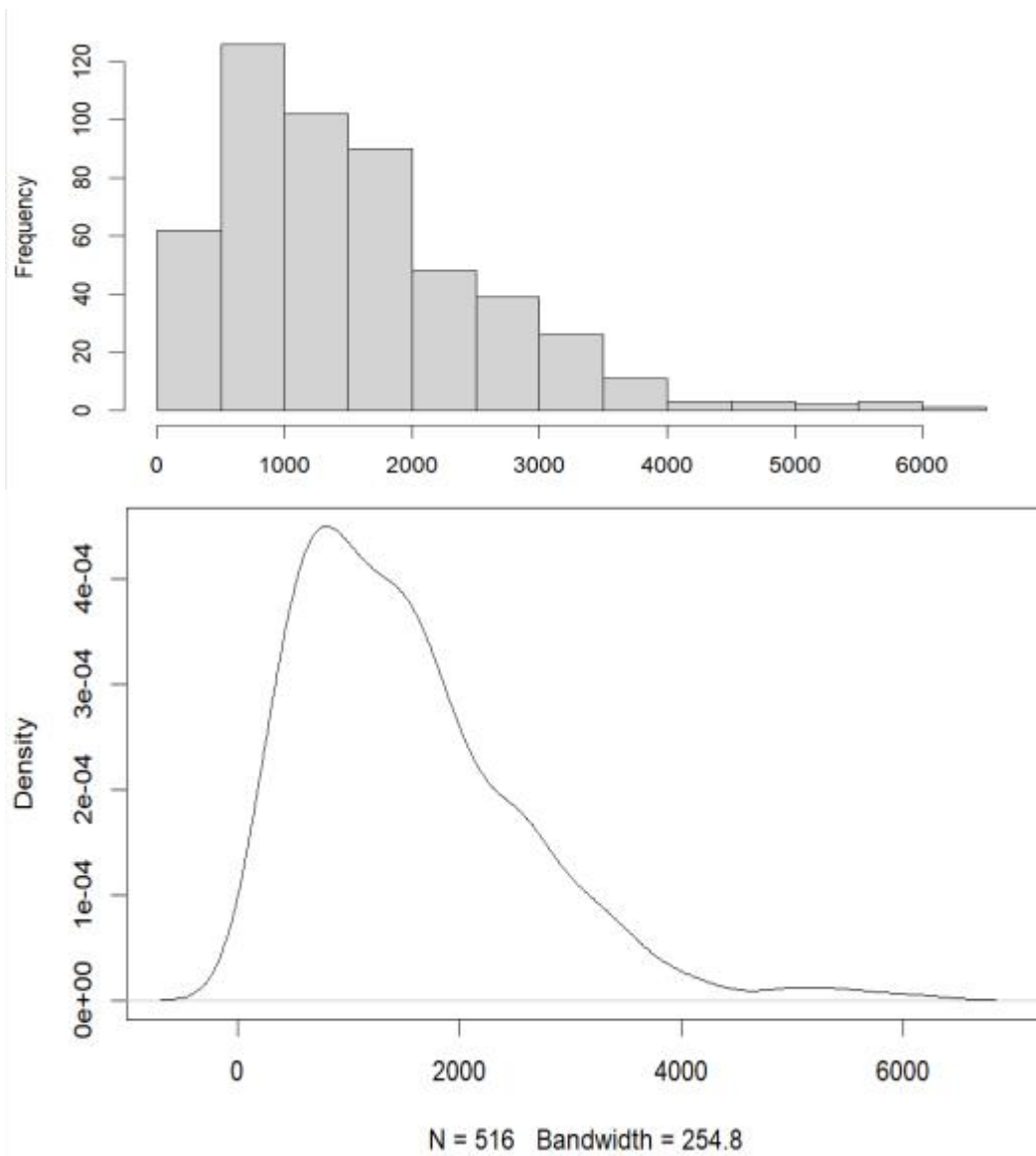
Script

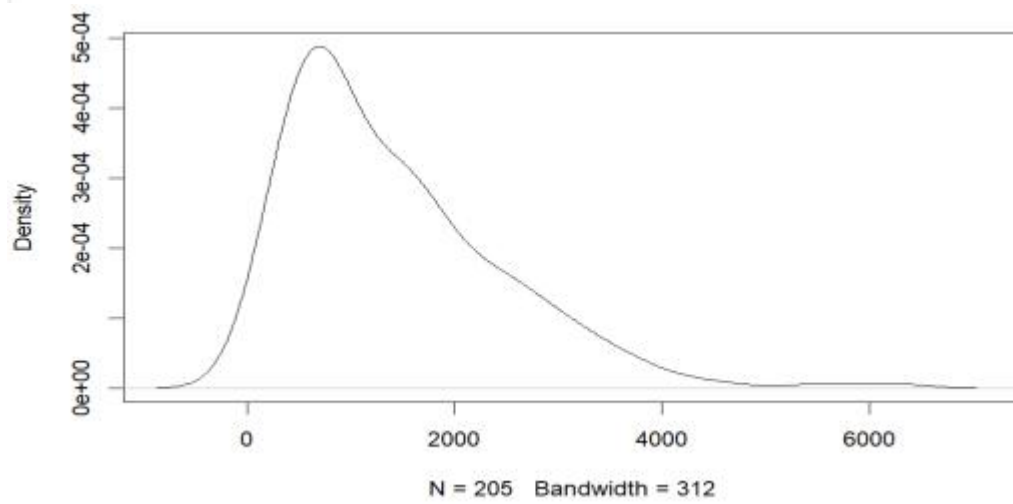
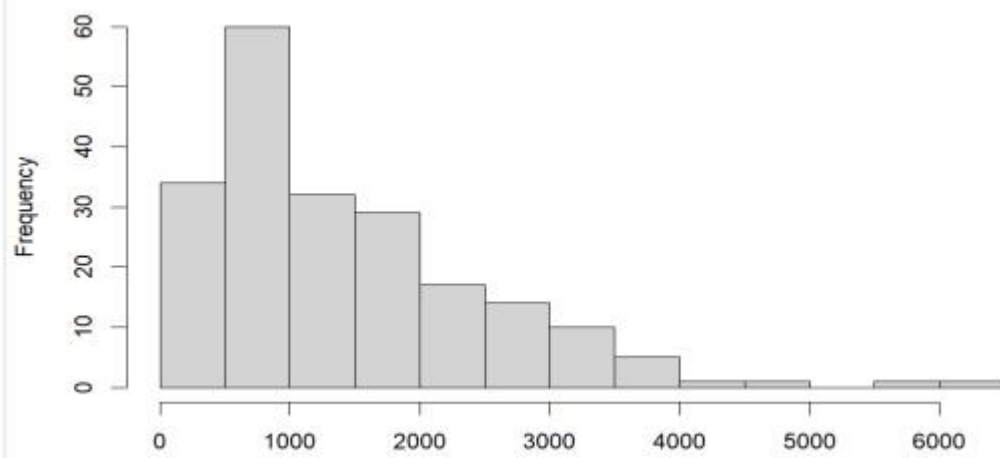
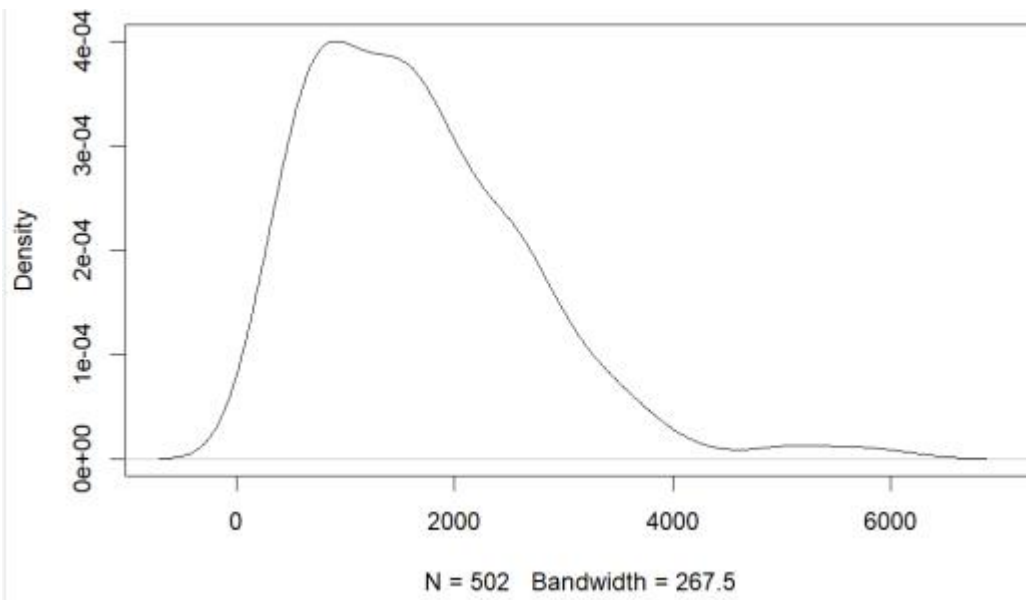
```
> df_francisco = PlantGrowth
> hist(data_francisco$AmountSpent)
> d=density(data_francisco$AmountSpent)
> plot(d)
> hist(data_francisco$AmountSpent[data_francisco$History=="High"])
> hist(data_francisco$AmountSpent[data_francisco$Gender=="Male"])
> d=density(data_francisco$AmountSpent[data_francisco$Gender=="Male"])
> plot(d)
> hist(data_francisco$AmountSpent[data_francisco$OwnHome=="Own"])
> d=density(data_francisco$AmountSpent[data_francisco$OwnHome=="Own"])
> plot(d)
> hist(data_francisco$AmountSpent[data_francisco$Married=="Married"])
> d=density(data_francisco$AmountSpent[data_francisco$Married=="Married"])
```

Output







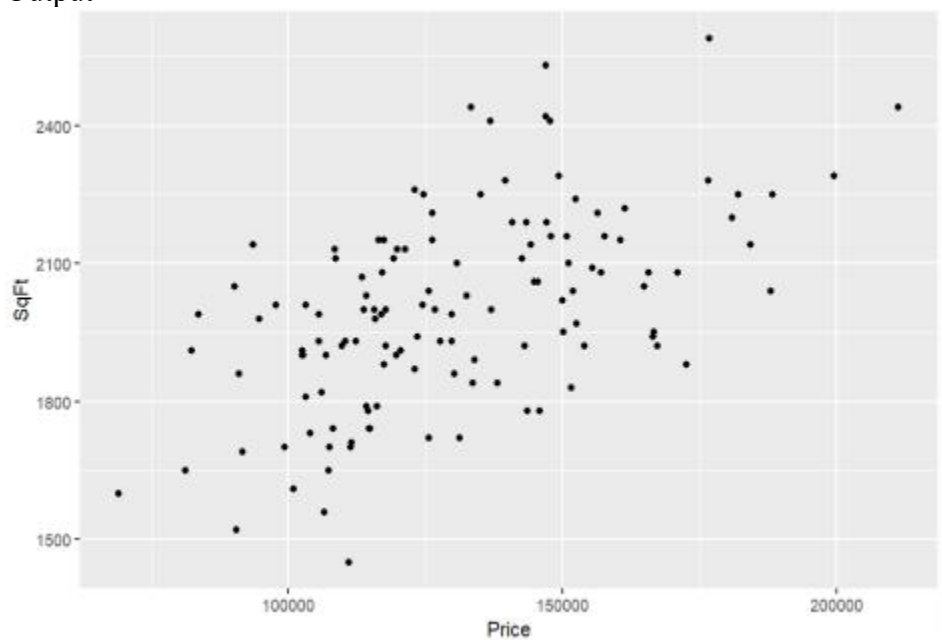


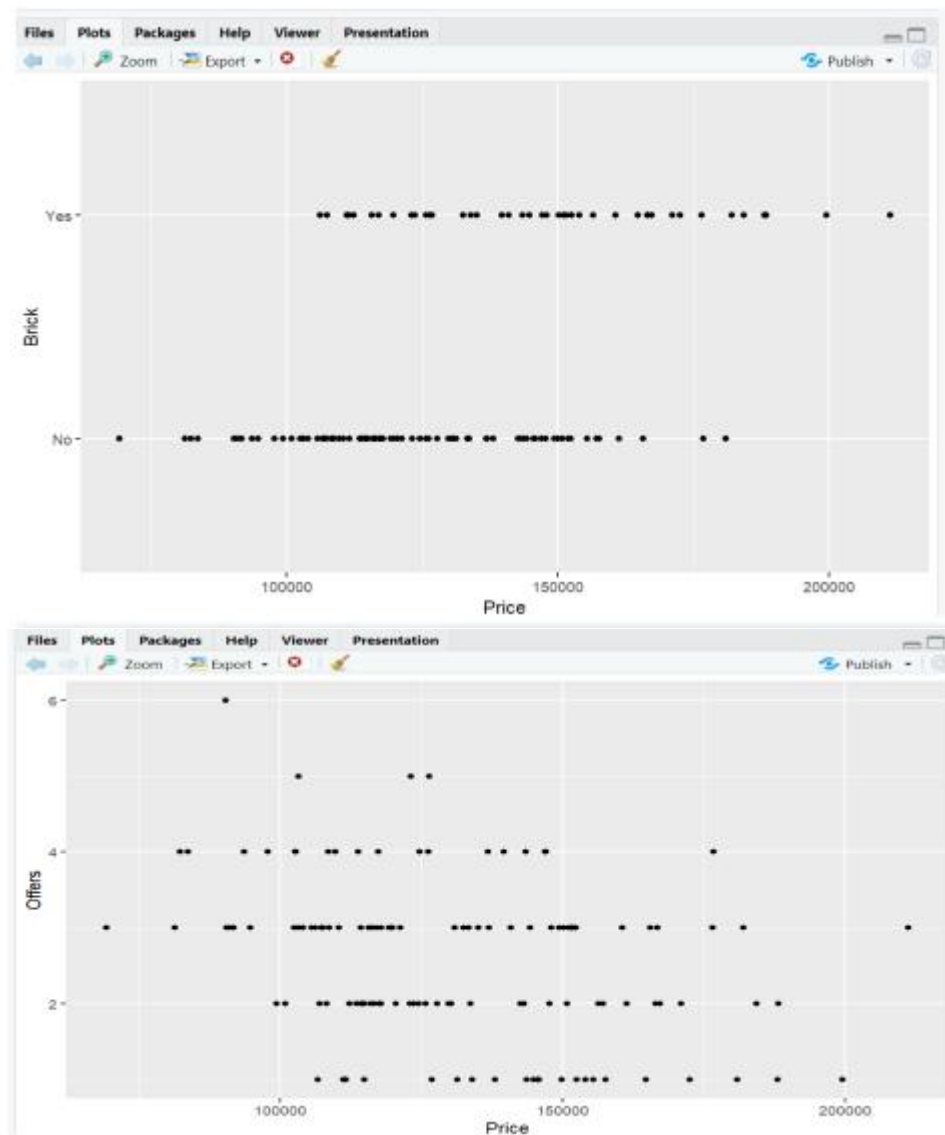
2. Gunakan data Houseprices.csv. Lakukan analisis dengan menggunakan scatter plot untuk mengetahui berbagai hubungan variabel numeric dengan harga rumah.

Script

```
> library(ggplot2)
> data_francisco=read.csv("E:/houseprice.csv")
> ggplot(data_francisco, aes(x = price, y = SqFt))
+   geom_point()
> ggplot(data_francisco, aes(x = price, y = Brick))
+   geom_point()
> ggplot(data_francisco, aes(x = price, y = offers))
+   geom_point()
```

Output





4. File Praktikum

Github Repository:

https://github.com/franafuk/Praktikum_stat

5. Soal Latihan

Soal:

1. Apa perbedaan utama pada grafik boxplot, histogram, dan density?
2. Jelaskan kegunaan dan fungsi dari Scatter Plot?

Jawaban:

1. Perbedaan diantara ketiganya yaitu Boxplot memberikan ringkasan visual tentang distribusi data, termasuk median, kuartil, rentang interkuartil (IQR), dan keberadaan pencilan (outlier), Histogram menampilkan distribusi frekuensi data dalam bentuk batang-batang tegak, membantu mengidentifikasi pola distribusi dan struktur data, sedangkan Density Plot menunjukkan perkiraan distribusi probabilitas dari suatu variabel dengan menggunakan kurva berkelanjutan, membantu memperkirakan kepadatan relatif data di berbagai nilai.
2. Scatter plot memiliki kegunaan yaitu membantu dalam menjelajahi hubungan antara variabel, mendeteksi korelasi, mengidentifikasi pola, mendeteksi pencilan, memodelkan prediksi, dan berkomunikasi hasil analisis secara visual.

6. Kesimpulan

- a. Dalam pengerjaan praktikum Statistika, terbilang cukup mudah walau sedikit rumit.
- b. Kita juga dapat mengetahui bahwasannya diantara ketiga jenis grafik di modul ini, Boxplot memberikan gambaran singkat tentang distribusi data dan perbedaan di antara kelompok, histogram menyajikan distribusi frekuensi data dengan jelas, sementara scatter plot membantu memahami hubungan antara dua variabel numerik. Dengan ini kita paham dengan apa yang materi ini sampaikan.

7. Cek List (✓)

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Latihan Pertama	v	

2.	Latihan Kedua	v	
3.	Latihan Ketiga	v	
4.	Latihan Keempat	v	

8. Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Latihan Pertama	10 Menit	Menarik
2.	Latihan Kedua	10 Menit	Menarik
3.	Latihan Ketiga	10 Menit	Menarik
4.	Latihan Keempat	10 Menit	Menarik

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang