


Nama : Ibnu Fajar Setiawan NIM : 065002000006	 UNIVERSITAS TRISAKTI Praktikum Data Analitik	Modul 2 Nama Dosen: Syandra Sari, S.Kom, M.Kom
Hari/Tanggal : Kamis, 29 Oktober 2022		Nama Aslab : 1. Ida Jubaidah (06500190037) 2. Azzahra Nuranisa (065001900044)

Praktikum 2 – Data Analitik
HUBUNGAN ANTAR DUA VARIABEL KATEGORIK DAN
ANTAR DUA VARIABEL NUMERIK

DESKRIPSI MODUL : Melihat hubungan antara variabel Katagorik dengan Variabel Numerik.

No	Elemen Kompetensi	Indikator Kinerja	Jml Jam	hlm
1	Mampu memahami cara menganalisis hubungan antar dua variable kategorik dan dua variable numerik	Dapat menganalisis hubungan antar dua variable kategorik dan dua variable numerik	2	

TEORI SINGKAT

Dalam praktikum ini akan dipelajari dan dipraktekkan bagaimana mengetahui adanya hubungan antar variabel dalam sebuah data multivariat atau data yang berisi lebih dari dua variabel. Variabel sendiri dapat diartikan sebagai sebuah konsep yang memiliki variasi nilai. Hubungan (relationship) antar variable dapat terjadi antara sebuah variabel kategorik dengan sebuah variabel kontinu, antar dua variabel kategorik serta antar dua variabel kontinu.

LAB SETUP

Untuk dapat menjalankan praktikum ini maka yang harus disiapkan adalah :

1. Aplikasi RStudio
2. Xampp

ELEMEN KOMPETENSI I

Deskripsi : Dapat menganalisis hubungan antar dua variabel

Kompetensi Dasar : Mampu memahami cara menganalisis hubungan antar dua variable

PRAKTIKUM

Dalam praktikum ini akan dipelajari dan dipraktekkan bagaimana mengetahui adanya hubungan antar dua variabel kategorik serta antar dua variabel numerik. Hubungan antar dua variabel kategori dapat diketahui menggunakan teknik tabulasi silang atau cross tabulation serta uji statistiknya menggunakan uji khi-kuadrat atau uji kebebasan. Sedangkan hubungan antar dua variabel numerik dapat dilihat menggunakan Teknik scatter plot serta koefisien korelasi. Nilai koefisien korelasi berkisar dari -1 sd 1.

Hubungan antar dua variabel kategorik

```
> data()
> View(Titanic)
> cross <- xtabs(Freq ~ Class+Sex, data=Titanic)
> cross
> prop.table(cross, 1)
> chisq.test(cross)
```

Output:

	Class	Sex	Age	Survived	Freq
1	1st	Male	Child	No	0
2	2nd	Male	Child	No	0
3	3rd	Male	Child	No	35
4	Crew	Male	Child	No	0
5	1st	Female	Child	No	0
6	2nd	Female	Child	No	0
7	3rd	Female	Child	No	17
8	Crew	Female	Child	No	0
9	1st	Male	Adult	No	118
10	2nd	Male	Adult	No	154
11	3rd	Male	Adult	No	387
12	Crew	Male	Adult	No	670

```
> view Titanic
> cross <- xtabs(Freq ~ Class+Sex)
> cross
```

```
      Sex
Class Male Female
1st    180    145
2nd    179    106
3rd    510    196
Crew   862     23
```

```
> prop.table(cross, 1)
```

```
      Sex
Class   Male   Female
1st  0.5538462 0.4461538
2nd  0.6280702 0.3719298
3rd  0.7223796 0.2776204
Crew 0.9740113 0.0259887
```

```
> chisq.test(cross)
```

Pearson's Chi-squared test

data: cross

X-squared = 349.91, df = 3, p-value < 2.2e-16

Hubungan antar dua variabel numerik

```
> head(iris)
```

```
> cor(iris$sepal.length, iris$sepal.width)
```

```
> cor(iris$petal.length, iris$petal.width)
```

```
> library(ggplot2)
```

```
> ggplot(iris, aes(x=petal.length, y=petal.width)) + geom_point()
```

Output:

```
> head(iris)
```

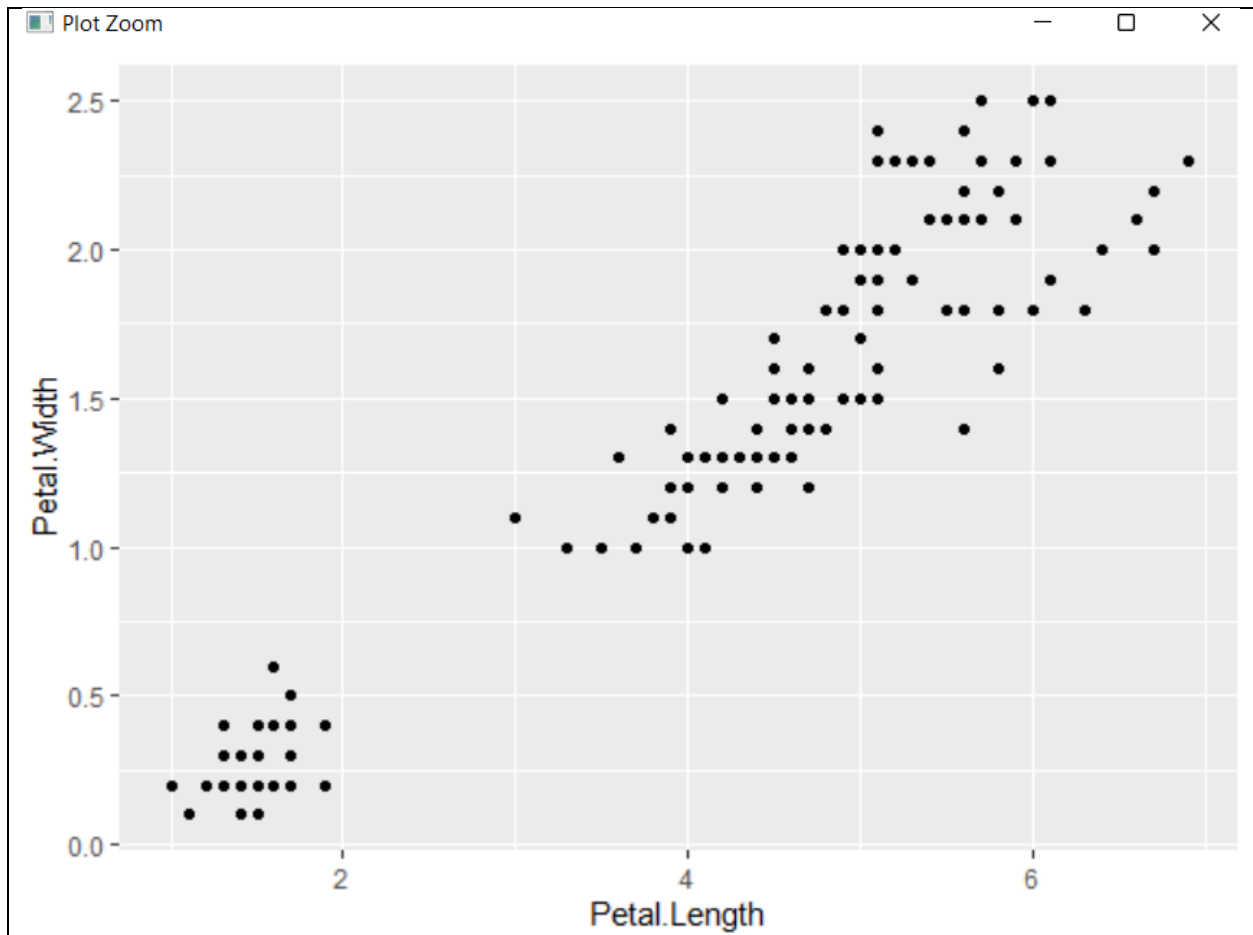
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

```
> cor(iris$Sepal.Length, iris$Sepal.Width)
```

```
[1] -0.1175698
```

```
> cor(iris$Petal.Length, iris$Petal.Width)
```

```
[1] 0.9628654
```



SCATTER PLOT

ELEMEN KOMPETENSI II

Tugas 1:

Deskripsi : Dapat menganalisis hubungan antar dua variable kategorik dan dua variable numerik

Kompetensi Dasar : Mampu memahami cara menganalisis hubungan antar dua variable kategorik dan dua variable numerik dari data Houseprices dengan menggunakan crostab-uji chi square, koefisien korelasi dan scatter plot

Data: House Price Data

Tujuan utama analisis dalam tugas ini adalah melihat adanya hubungan antar dua variabel kategorik (misalnya bahan dinding dengan lokasi) serta hubungan antar dua variabel numerik

(misalnya harga rumah dengan luas bangunan). Jawaban atas pertanyaan-pertanyaan ini bisa membantu

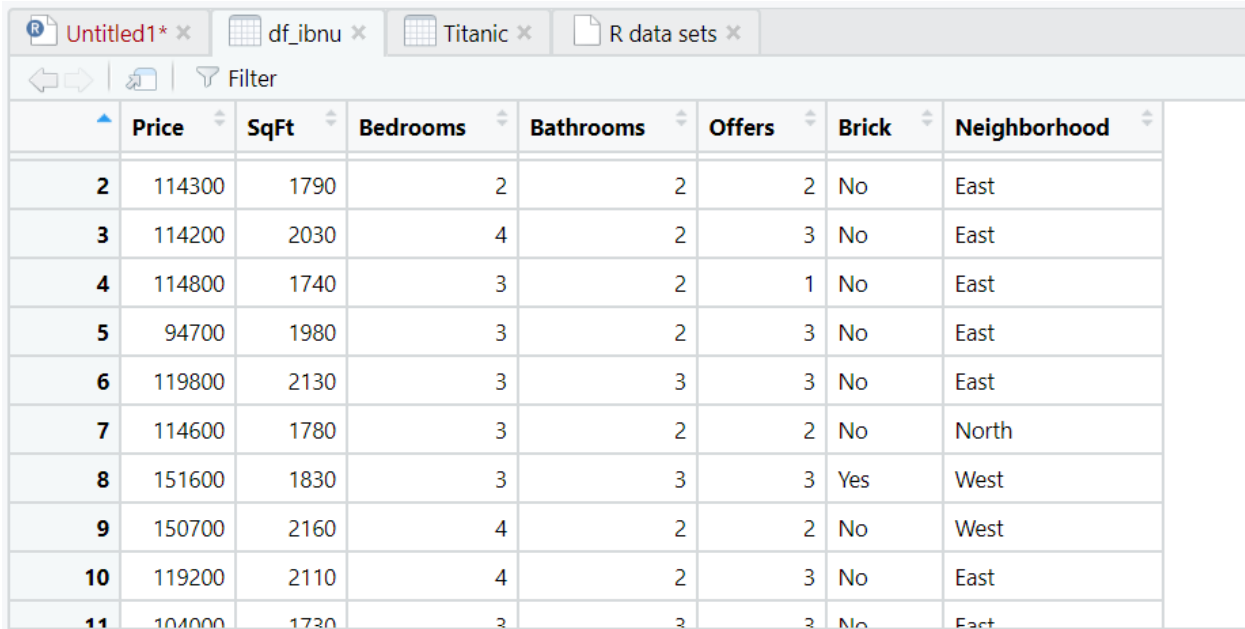
calon pembeli memutuskan berapa besar tawaran untuk rumah.

Berikan penjelasan untuk setiap output yang dihasilkan serta tulis R code yang digunakan. Tampilan grafik yang berwarna tentunya lebih baik.

Catatan : Kalimat yang dipergunakan untuk membuat penjelasan setiap output antar praktikan tidak boleh sama

```
> library(RMySQL)
> con = dbConnect(MySQL(), user = 'root', password = '', dbname = 'houseprices', host = 'localhost')
> myQuery <- "select * from tabel1;"
> df_nama <- dbGetQuery(con, myQuery)
> View(df_nama)
```

Output:



	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
2	114300	1790	2	2	2	No	East
3	114200	2030	4	2	3	No	East
4	114800	1740	3	2	1	No	East
5	94700	1980	3	2	3	No	East
6	119800	2130	3	3	3	No	East
7	114600	1780	3	2	2	No	North
8	151600	1830	3	3	3	Yes	West
9	150700	2160	4	2	2	No	West
10	119200	2110	4	2	3	No	East
11	104000	1730	3	3	3	No	East

Hubungan antar dua variabel kategorik

```
> cross <- xtabs(Price ~ Brick+Neighborhood, data=df_nama)
> cross
> prop.table(cross, 1)
> chisq.test(cross)
```

Output:

```

> View(df_ibnu)
> cross <- xtabs(Price ~ Brick+Neighborhood, data=df_ibnu)
> cross
      Neighborhood
Brick   East Neighborhood   North   West
Brick      0      0      0      0
No    3061500      0 4017600 3409300
Yes   2573900      0  829200 2803200
> prop.table(cross, 1)
      Neighborhood
Brick   East Neighborhood   North   West
Brick
No    0.2918939    0.0000000 0.3830518 0.3250543
Yes   0.4147237    0.0000000 0.1336062 0.4516701
> chisq.test(cross)

Pearson's Chi-squared test

data:  cross
X-squared = 1177926, df = 2, p-value < 2.2e-16

```

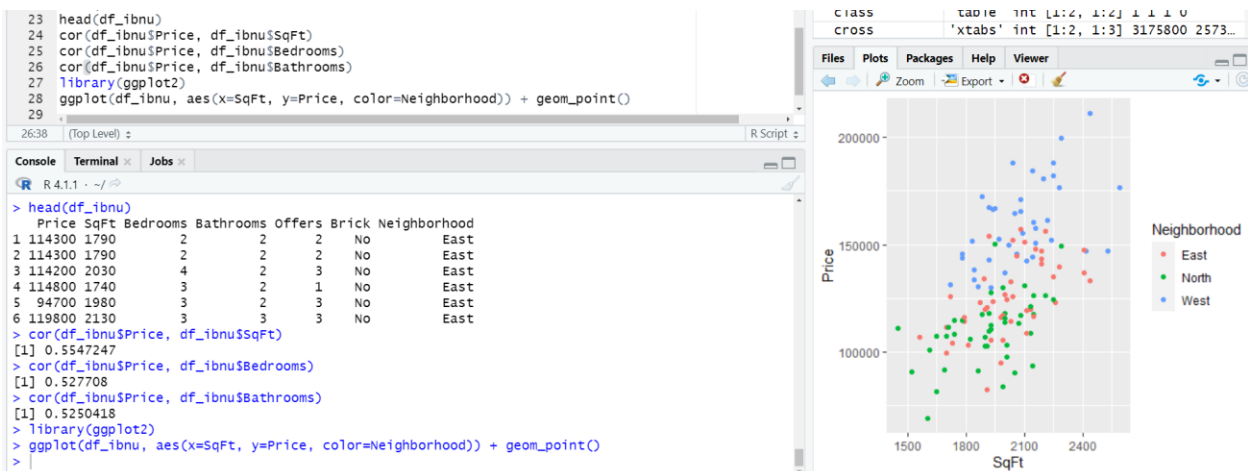
Hubungan antar dua variabel numerik

```

> head(df_nama)
> cor(df_nama$Price, df_nama$SqFt)
> cor(df_nama$Price, df_nama$Bedrooms)
> cor(df_nama$Price, df_nama$Bathrooms)
library(ggplot2)
ggplot(df_nama, aes(x=SqFt, y=Price, color=Neighborhood)) + geom_point()

```

Output:



Dari hasil grafik diatas dapat disimpulkan bahwa ...

Tugas 2:

Deskripsi : Dapat menganalisis hubungan antar dua variable kategorik dan dua variable numerik

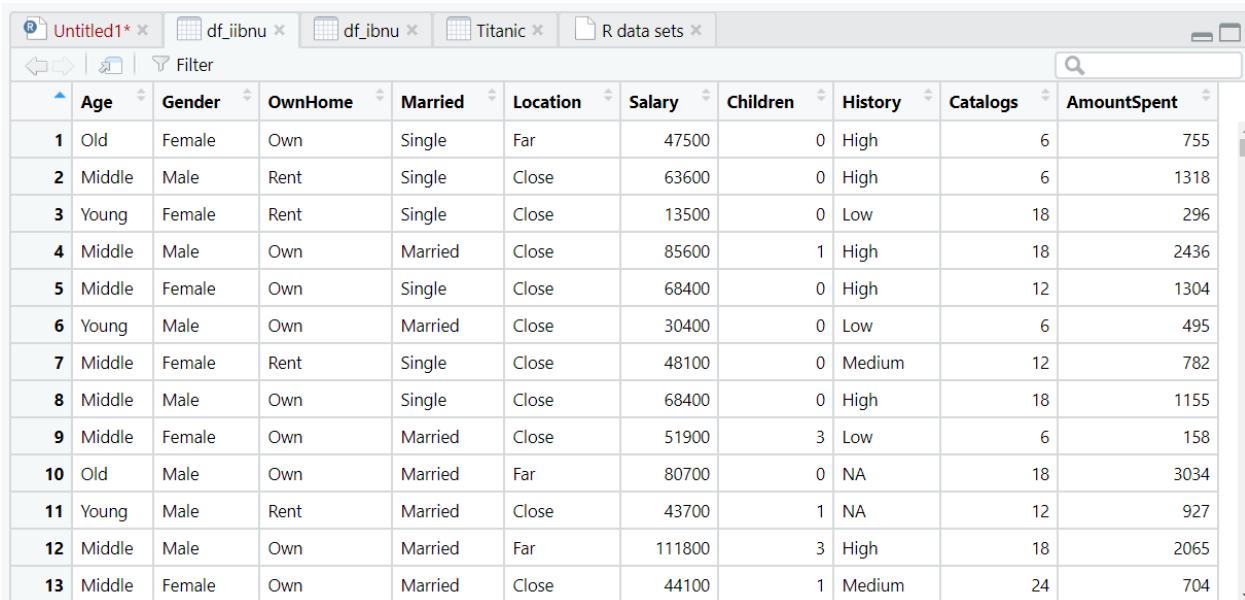
Kompetensi Dasar : Mampu memahami cara menganalisis hubungan antar dua variable kategorik dan dua variable numerik dari data Direct Marketing dengan menggunakan crosstab-uji chi square, koefisien korelasi dan scatter plot

Data : DirectMarketing.

Lihat apakah di dalam datanya terdapat hubungan antar variable kategorik dan antar variable numerik. Buat tampilan grafiknya berwarna.

```
> library(RMySQL)
> con = dbConnect(MySQL(), user = 'root', password = '', dbname = 'directmarketing', host = 'localhost')
> myQuery <- "select * from tabel1;"
> df_nama <- dbGetQuery(con, myQuery)
> View(df_nama)
```

Output:



	Age	Gender	OwnHome	Married	Location	Salary	Children	History	Catalogs	AmountSpent
1	Old	Female	Own	Single	Far	47500	0	High	6	755
2	Middle	Male	Rent	Single	Close	63600	0	High	6	1318
3	Young	Female	Rent	Single	Close	13500	0	Low	18	296
4	Middle	Male	Own	Married	Close	85600	1	High	18	2436
5	Middle	Female	Own	Single	Close	68400	0	High	12	1304
6	Young	Male	Own	Married	Close	30400	0	Low	6	495
7	Middle	Female	Rent	Single	Close	48100	0	Medium	12	782
8	Middle	Male	Own	Single	Close	68400	0	High	18	1155
9	Middle	Female	Own	Married	Close	51900	3	Low	6	158
10	Old	Male	Own	Married	Far	80700	0	NA	18	3034
11	Young	Male	Rent	Married	Close	43700	1	NA	12	927
12	Middle	Male	Own	Married	Far	111800	3	High	18	2065
13	Middle	Female	Own	Married	Close	44100	1	Medium	24	704

Hubungan antar dua variabel kategorik

```
> cross <- xtabs(Salary ~ OwnHome+Married, data=df_nama)
> cross
> prop.table(cross, 1)
> chisq.test(cross)
```

Output:


```

> cross <- xtabs(Salary ~ OwnHome+Married, data=df_iibnu)
> cross
      Married
OwnHome Married  Single
Own   27080000  8915500
Rent  11421600  8686800
> prop.table(cross, 1)
      Married
OwnHome Married  Single
Own   0.7523163 0.2476837
Rent  0.5680014 0.4319986
> chisq.test(cross)

Pearson's Chi-squared test with Yates' continuity correction

data:  cross
X-squared = 2035593, df = 1, p-value < 2.2e-16

```

Hubungan antar dua variabel numerik

```

> head(df)
> cor(df_nama$Salary, df_nama$AmountSpent)
> cor(df_nama$Salary, df_nama$Catalogs)
> cor(df_nama$Salary, df_nama$Children)

> library(ggplot2)
> ggplot(df_nama, aes(x=AmountSpent, y=Salary, color=Age)) + geom_point()

> head(df_iibnu)
  Age Gender OwnHome Married Location Salary Children History Catalogs AmountSpent
1 Old Female Own Single Far 47500 0 High 6 755
2 Middle Male Rent Single Close 63600 0 High 6 1318
3 Young Female Rent Single Close 13500 0 Low 18 296
4 Middle Male Own Married Close 85600 1 High 18 2436
5 Middle Female Own Single Close 68400 0 High 12 1304
6 Young Male Own Married Close 30400 0 Low 6 495
> cor(df_iibnu$Salary, df_iibnu$AmountSpent)
[1] 0.6995957
> cor(df_iibnu$Salary, df_iibnu$Catalogs)
[1] 0.1835509
> cor(df_iibnu$Salary, df_iibnu$Children)
[1] 0.04966316
> ggplot(df_iibnu, aes(x=AmountSpent, y=Salary, color=Age)) + geom_point()
>

```



Dari hasil grafik diatas dapat disimpulkan bahwa age young banyak yg belum berpenghasilan/salary dan amount spend nya dominan 0-2000.

Output:

1. Cek List

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak
1	Elemen Kompetensi I	✓	

	Memahami hubungan antar variabel katagorik dengan Numerik		
2	Elemen Kompetensi II TUGAS 1 : Mampu memahami cara menganalisis hubungan antar dua variable kategorik dan dua variable numerik dari data Houseprices dengan menggunakan Scatter Plot	✓	
3	Elemen Kompetensi II TUGAS 2 : Mampu memahami cara menganalisis hubungan antar dua variable kategorik dan dua variable numerik dari data Direct Marketing dengan menggunakan Scatter Plot	✓	

2. Form Umpan Balik

Elemen Kompetensi	Waktu Pengerjaan	Kriteria
Elemen Kompetensi I Memahami hubungan antar variabel katagorik dengan Numerik	30	1
Elemen Kompetensi II TUGAS 1 : Mampu memahami cara menganalisis hubungan antar dua variable kategorik dan dua variable numerik dari data Houseprices dengan menggunakan Scatter Plot	30	1
Elemen Kompetensi II TUGAS 2 : Mampu memahami cara menganalisis hubungan antar dua variable kategorik dan dua variable numerik dari data Direct Marketing dengan menggunakan Scatter Plot	30	1

Kriteria

1.Sangat Menarik

- 2.Cukup Menarik
- 3.Kurang Menarik
- 4.Sangat Kurang Menarik