

Tugas 1

[NIM_NAMA]

30/3/2021

Contents

Soal 1	1
Soal 2	2
Soal 3	3
Soal 4	4
Soal 5	5
Soal 6	6
Grouping	7

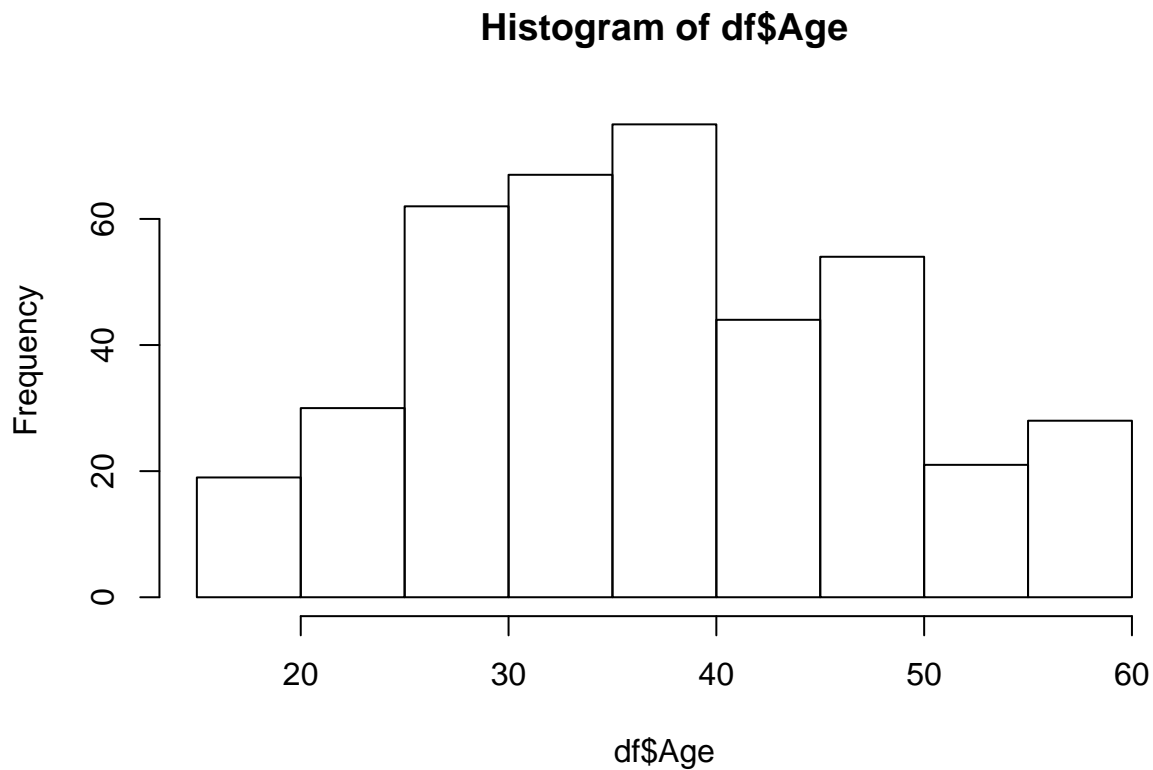
Kembali pada kasus PT. METSTAT. Setelah kamu memberikan hasil analisis kemarin, manajer memuji hasil analisis kamu! Manajer pun mengusulkan kamu untuk mempresentasikan hasil analisis kamu pada rapat nanti. Namun sebelum itu, manajer meminta kamu untuk memvisualisasikan hasil analisis kamu terlebih dahulu. Di samping untuk mempercantik slide presentasi, visualisasi data berguna untuk mempermudah dalam menjelaskan hasil analisis pada divisi-divisi yang masih awam akan statistika.

Note: Data tetap menggunakan `Social_Network_Ads.csv` (<https://www.kaggle.com/rakeshrau/social-network-ads>) . Khusus nomor 6 menggunakan data `sales_data_sample_grouped.csv` (<https://www.kaggle.com/kyanyoga/sample-sales-data>)

Soal 1

Visualisasikan distribusi data usia pengguna. Plot apa yang sebaiknya digunakan? Berikan alasan dan informasi yang diperoleh dari visualisasi tersebut!

```
df = read.csv("data/Social_Network_Ads.csv")
hist(df$Age)
```



Interpretasi

ketik di sini

Kesimpulan

ketik di sini

Soal 2

Visualisasikan banyak pengguna berdasarkan kelompok usia:

- 15-24 tahun: Kelompok usia muda
- 25-34 tahun: Kelompok usia pekerja awal
- 35-44 tahun: Kelompok usia paruh baya
- 45-54 tahun: Kelompok usia pra-pensiun
- 55-64 tahun: Kelompok usia pensiun

Source: https://sepakat.bappenas.go.id/wiki/Kelompok_Usia

Plot apa yang sebaiknya digunakan? Berikan alasan dan informasi yang diperoleh dari visualisasi tersebut!

```
usia_muda = length(subset(df$Age, df$Age >= 15 & df$Age <= 24))
pekerja_awal = length(subset(df$Age, df$Age >= 25 & df$Age <= 34))
paruh_baya = length(subset(df$Age, df$Age >= 35 & df$Age <= 44))
```

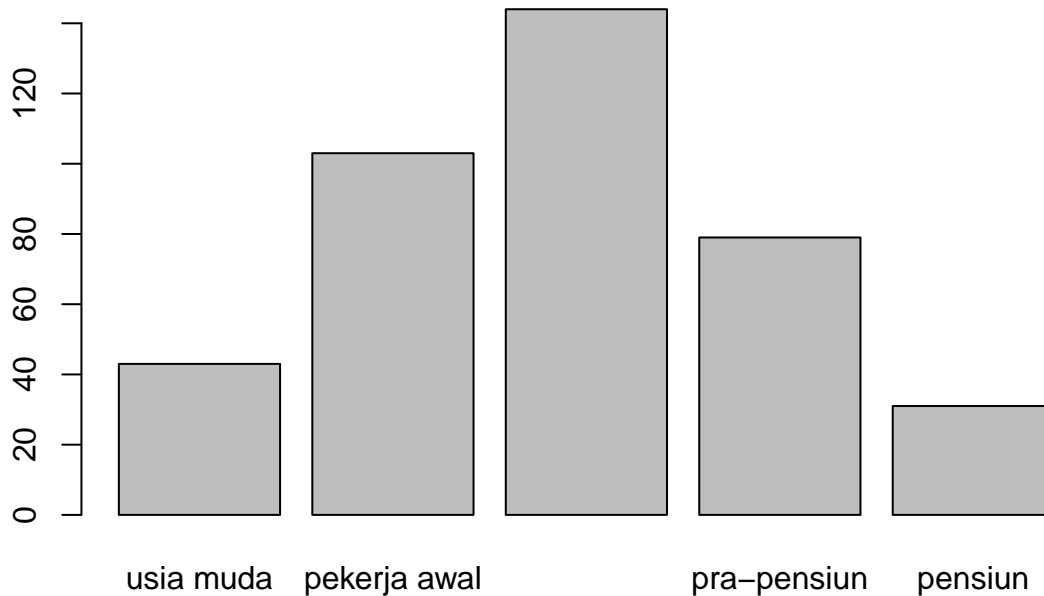
```

pra_pensiun = length(subset(df$Age, df$Age >= 45 & df$Age <= 54))
pensiun = length(subset(df$Age, df$Age >= 55 & df$Age <= 64))

label = c("usia muda", "pekerja awal", "paruh baya", "pra-pensiun", "pensiun")
frekuensi = c(usia_muda, pekerja_awal, paruh_baya, pra_pensiun, pensiun)

barplot(frekuensi, names.arg = label)

```



Soal 3

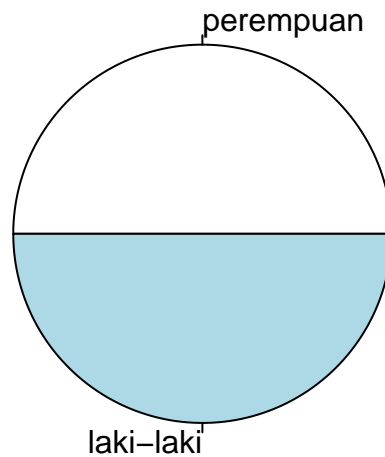
Visualisasikan perbandingan jumlah pengguna laki-laki dan perempuan yang berada pada kelompok usia pekerja awal. Plot apa yang sebaiknya digunakan? Berikan alasan dan informasi yang diperoleh dari visualisasi tersebut!

```

pekerja_awal.lk = length(subset(df, df$Age >= 25 & df$Age <= 34 & df$Gender == "Male"))
pekerja_awal.pr = length(subset(df, df$Age >= 25 & df$Age <= 34 & df$Gender == "Female"))
frekuensi = c(pekerja_awal.pr, pekerja_awal.lk)
label = c("perempuan", "laki-laki")

pie(frekuensi, labels = label)

```



Interpretasi

ketik di sini

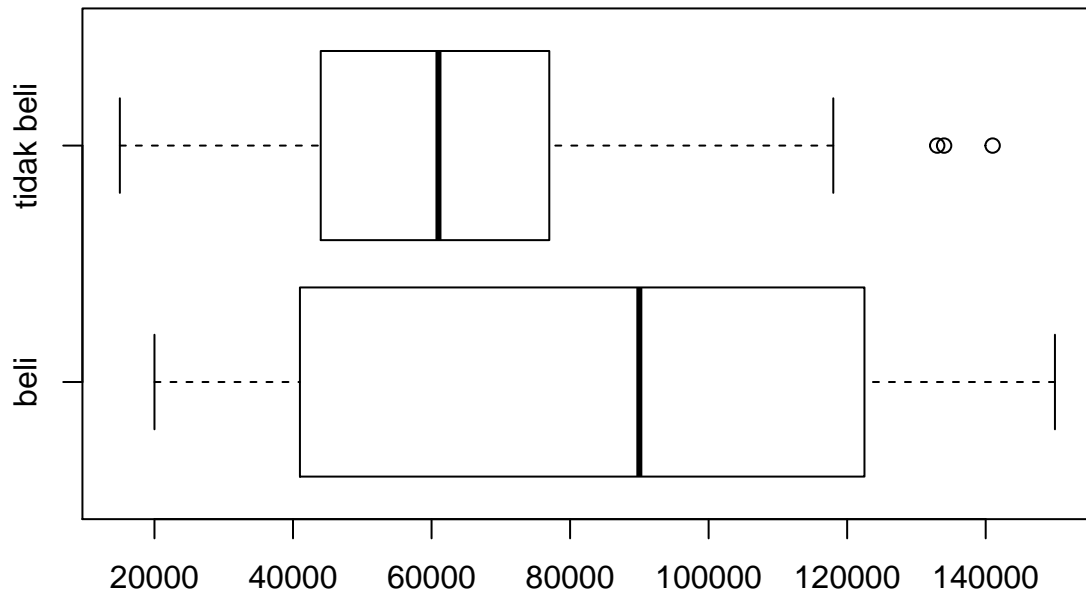
Kesimpulan

ketik di sini

Soal 4

Visualisasikan perbandingan distribusi data gaji pengguna yang membeli produk dengan yang tidak. Plot apa yang sebaiknya digunakan? Berikan alasan dan informasi yang diperoleh dari visualisasi tersebut!

```
df.beli = subset(df, df$Purchased == 1)
df.g_beli = subset(df, df$Purchased == 0)
boxplot(df.beli$EstimatedSalary, df.g_beli$EstimatedSalary,
        horizontal = TRUE,
        names = c("beli", "tidak beli"))
```



Interpretasi

ketik di sini

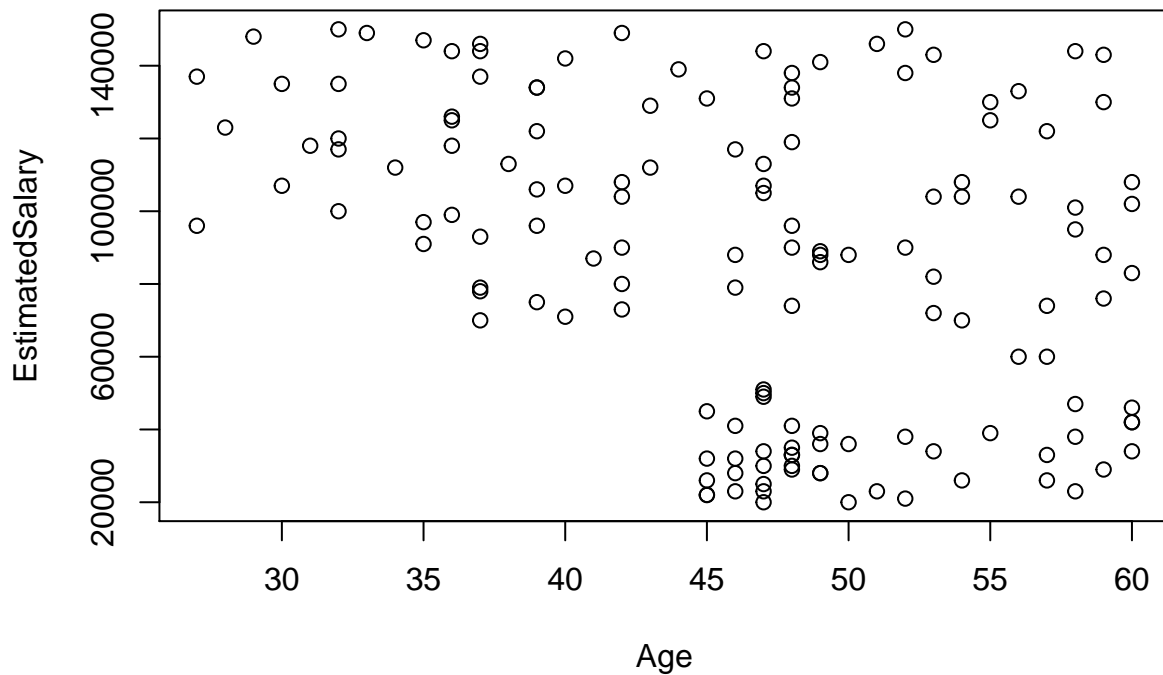
Kesimpulan

ketik di sini

Soal 5

Visualisasikan hubungan/korelasi antara usia dengan gaji pengguna yang membeli produk. Plot apa yang sebaiknya digunakan? Berikan alasan dan informasi yang diperoleh dari visualisasi tersebut!

```
plot(subset(df, df$Purchased == 1)[, c(3,4)])
```



Interpretasi

ketik di sini

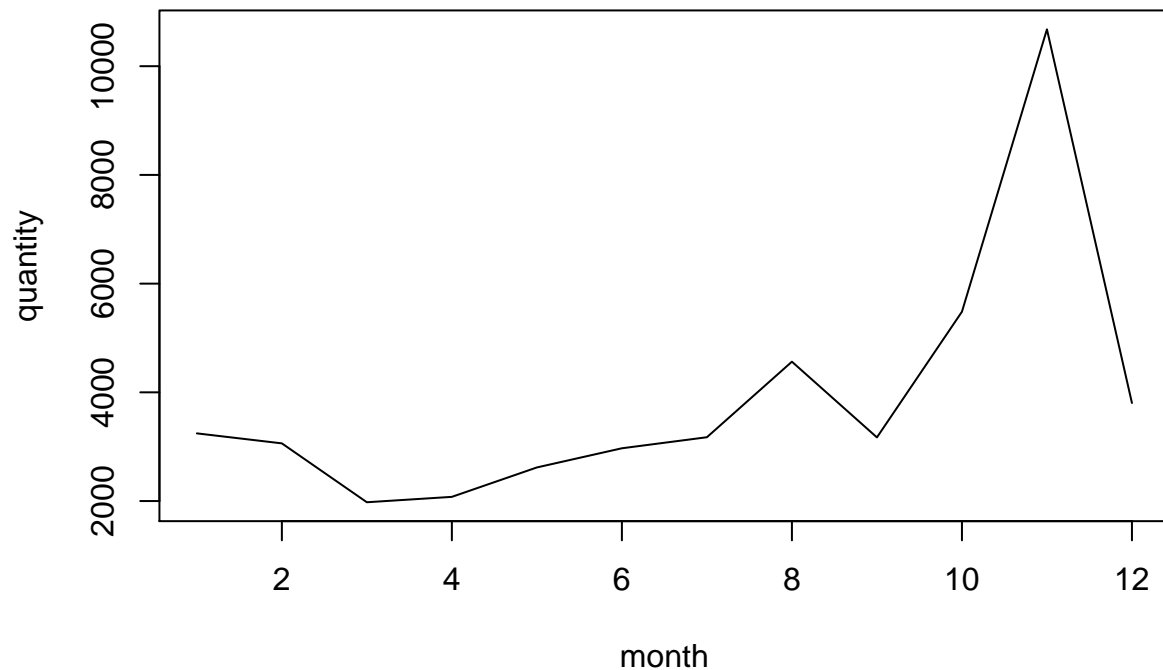
Kesimpulan

ketik di sini

Soal 6

Visualisasikan data jumlah penjualan tahun lalu tiap bulannya. Plot apa yang sebaiknya digunakan? Berikan alasan dan informasi yang diperoleh dari visualisasi tersebut!

```
df.penjualan = read.csv("data/sales_data_sample_grouped.csv")  
plot(df.penjualan, type="l")
```



Interpretasi

ketik di sini

Kesimpulan

ketik di sini

Grouping

Ngga usah dijalanin

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union  
  
df = read.csv("data/sales_data_sample.csv")
```

```
# select and grouping data
```

```

df_grouped =
  df %>%
  filter(YEAR_ID == 2004) %>%
  select(MONTH_ID, QUANTITYORDERED) %>%
  group_by(MONTH_ID) %>%
  summarise(QUANTITYORDERED = sum(QUANTITYORDERED)) %>%
  rename(month = MONTH_ID, quantity = QUANTITYORDERED) %>%
  arrange(month)

## `summarise()` ungrouping output (override with `.groups` argument)

# save to csv
write.csv(df_grouped, "data/sales_data_sample_grouped.csv", row.names = FALSE)

```