

# Praktikum 3

Fariza Shielda Akzatria\_162112133026

9/3/2022

Unduhlah (download) 1 dataset (bebas) dari Kaggle (<https://www.kaggle.com/>) atau UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets.php>) kemudian identifikasi jenis atributnya dan visualisasikan menggunakan teknik visualisasi yang sudah dijelaskan pada modul ini.

## Visualisasi Jenis Atribut

### Dataset

Dataset yang saya gunakan yaitu *Brain Stroke Prediction Dataset* yang didapat dari *Kaggle*, dapat diakses pada link berikut : <<https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-brain-stroke-dataset>>

##	gender	age	hypertension	heart_disease	ever_married	work_type
## 1	Female	61	0	0	Yes	Self-employed
## 2	Female	59	0	0	Yes	Private
## 3	Male	78	0	1	Yes	Private
## 4	Male	57	0	1	No	Govt_job
## 5	Male	58	0	0	Yes	Private
## 6	Male	59	0	0	Yes	Private

##	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
## 1	Rural	202.21	31.55560	never smoked	1
## 2	Rural	76.15	30.24294	Unknown	1
## 3	Urban	219.84	30.69895	Unknown	1
## 4	Urban	217.08	33.80841	Unknown	1
## 5	Rural	189.84	31.37853	Unknown	1
## 6	Rural	211.78	33.48457	formerly smoked	1

### 1. Atribut Gender

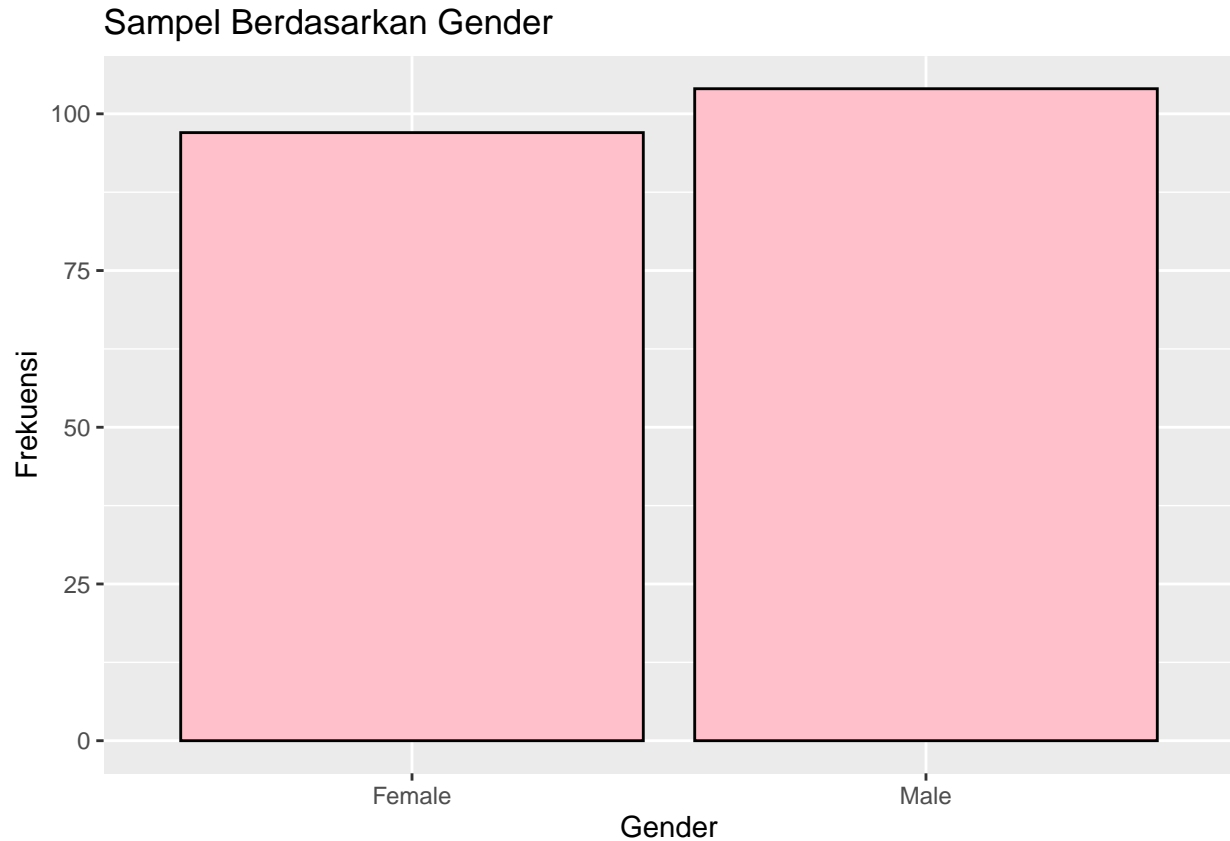
- Jenis Atribut : Termasuk tipe data kualitatif, subkategori nominal karena data berupa kategori dan hanya sebagai penggolongan saja.
- Untuk visualisasi dapat menggunakan *bar chart* karena grafik tersebut cocok untuk menggambarkan banyaknya jumlah *male* dan *female* dalam dataset.

```
library(readr)
data<-read.csv("C:/Users/FARIZA SHIELDA/Documents/File/Semester 3/Eksplorasi dan Visualisasi Data/Praktikum 3/Brain Stroke Prediction Dataset/brain_stroke_prediction_dataset.csv")
library(ggplot2)
ggplot(data, aes(x = gender)) +
  geom_bar(fill = "pink",
```

```

    color = "black") +
labs(x = "Gender",
     y = "Frekuensi",
     title = "Sampel Berdasarkan Gender")

```



## 2. Atribut Age

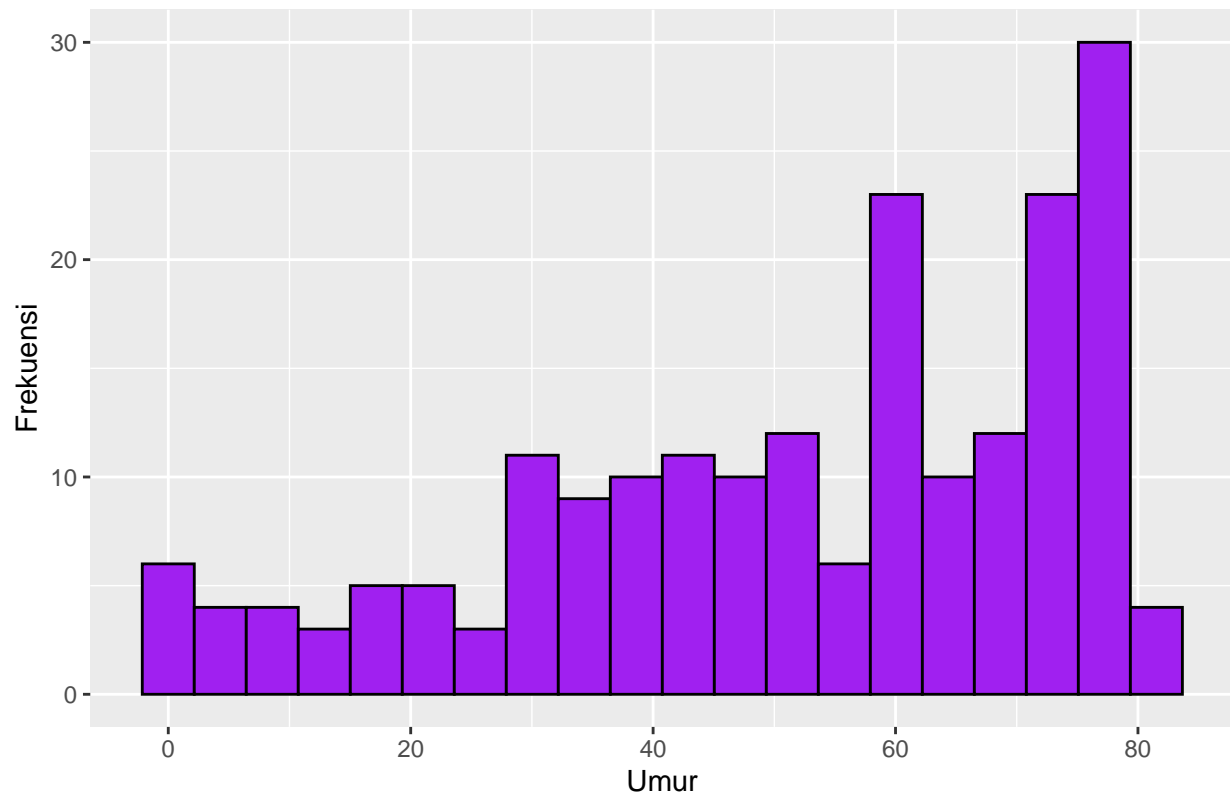
- Jenis Atribut : Termasuk tipe data kuantitatif, subkategori rasio karena data berupa angka dan memiliki nilai nol mutlak serta memiliki sifat perbandingan.
- Untuk visualisasi dapat menggunakan *histogram* karena atributnya berupa data kuantitatif dan termasuk model visualisasi yang paling sederhana.

```

ggplot(data, aes(x = age)) +
  geom_histogram(fill = "purple",
                 color = "black",
                 bins = 20) +
labs(title = "Partisipan Berdasarkan Umur",
     x = "Umur",
     y = "Frekuensi")

```

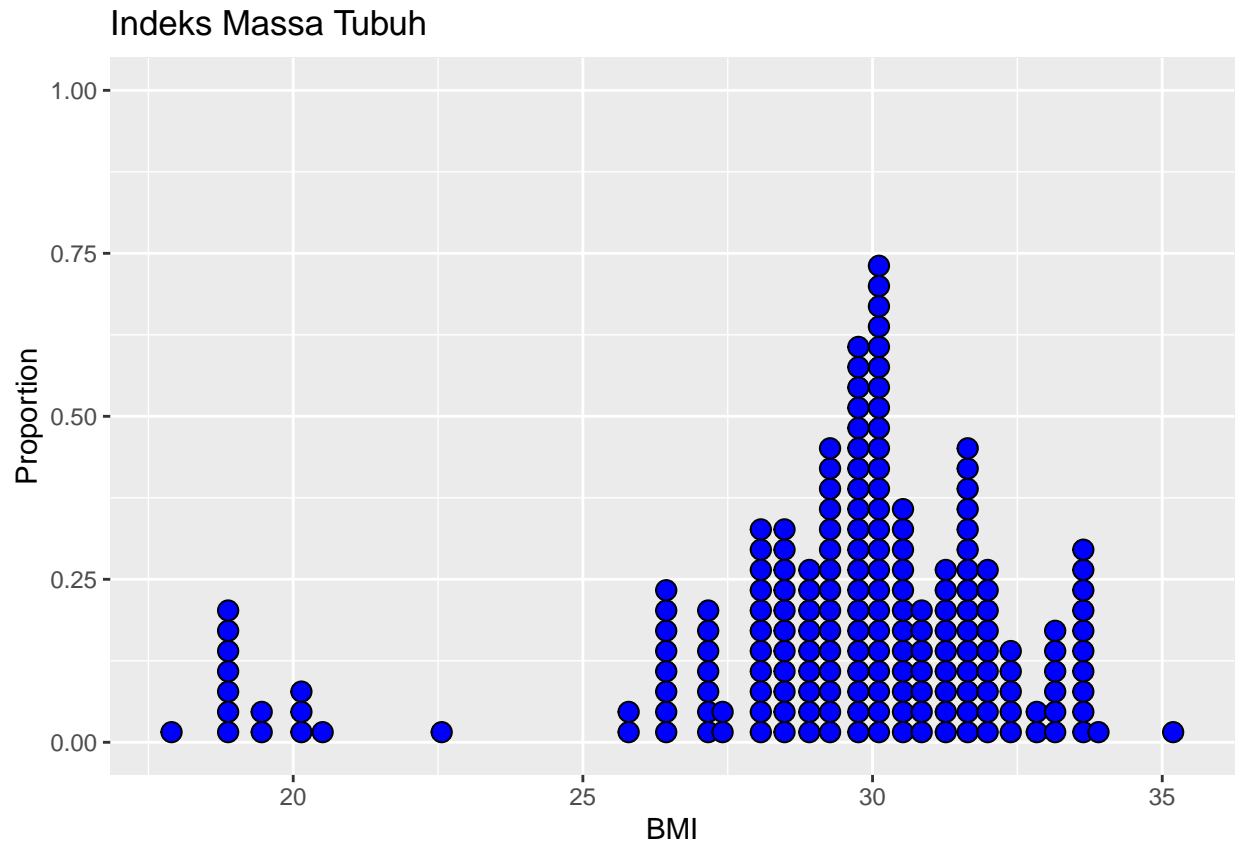
### Partisipan Berdasarkan Umur



### 3. Atribut Body Mass Index

- Jenis Atribut : Termasuk tipe data kuantitatif, subkategori interval karena data berupa angka dan tidak memiliki nilai nol mutlak serta dapat dibagi ke dalam suatu rentang.
- Untuk visualisasi dapat menggunakan *dot chart* karena atributnya berupa data kuantitatif dan rentang-rentangnya dapat dibagi ke dalam beberapa bin.

```
ggplot(data, aes(x = bmi)) +  
  geom_dotplot(fill = "blue",  
               color = "black",  
               binwidth = 0.35) +  
  labs(title = "Indeks Massa Tubuh",  
        x = "BMI",  
        y = "Proportion")
```



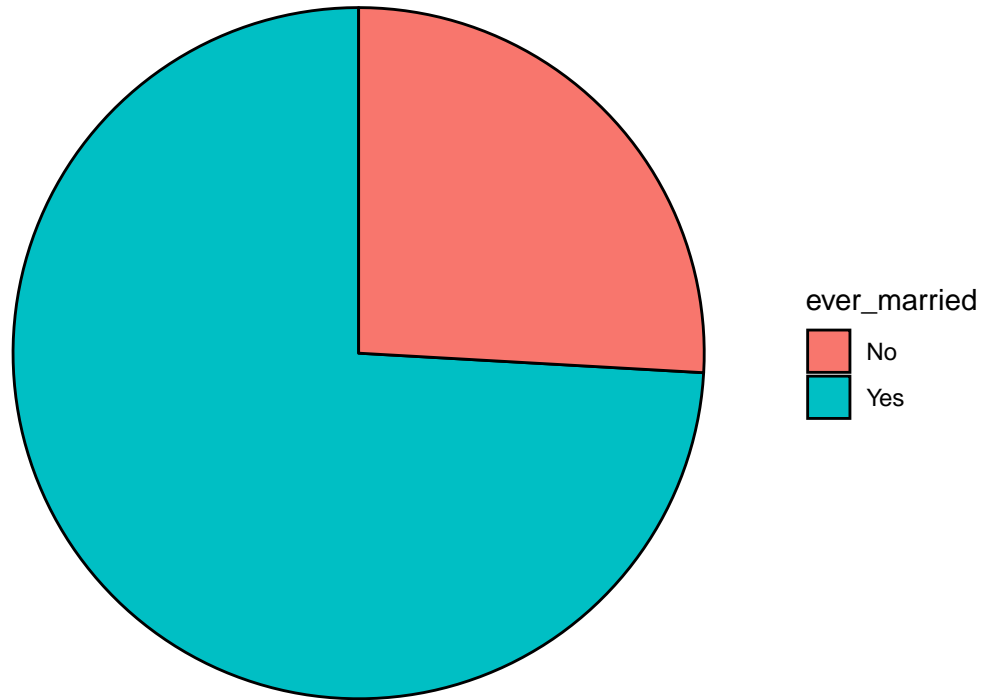
#### 4. Atribut Ever Married

- Jenis Atribut : Termasuk tipe data kualitatif, subkategori biner karena data berupa kategori dan hanya memiliki dua kategori ya/tidak saja.
- Untuk visualisasi dapat menggunakan *pie chart* karena dapat memperlihatkan perbandingan antar kategori dalam variabel.

```
library(dplyr)
dataplot <- data %>%
  count(ever_married) %>%
  arrange(desc(ever_married)) %>%
  mutate(prop = round(n * 100 / sum(n), 1),
         lab.ypos = cumsum(prop) - 0.5 * prop)

ggplot(dataplot,
       aes(x = "",
           y = prop,
           fill = ever_married)) +
  geom_bar(width = 1,
           stat = "identity",
           color = "black") +
  coord_polar("y",
             start = 0,
```

```
direction = -1) +  
theme_void()
```



## 5. Atribut Work Type

- Jenis Atribut : Termasuk tipe data kualitatif, subkategori nominal karena data berupa kategori dan hanya sebagai penggolongan saja.
- Untuk visualisasi dapat menggunakan *bar chart* karena grafik tersebut cocok untuk menggambarkan banyaknya jumlah tiap kategori dalam variabel *Work Type*

```
ggplot(data, aes(x = work_type)) +  
  geom_bar(fill = "green",  
           color = "black") +  
  labs(x = "Jenis Pekerjaan",  
       y = "Frekuensi",  
       title = "Sampel Berdasarkan Jenis Pekerjaan")
```

Sampel Berdasarkan Jenis Pekerjaan

