TUGAS MEMBUAT DATA PROFILE DIAGRAM

MATA KULIAH MANAJEMEN DAN ANALISIS DATA DENGAN R

ASTI OKTOVIANTI SUNMAYA ANANDA PUTRI

NPM. 131520220005

PROGRAM STUDI EPIDEMIOLOGI

FAKULTAS KEDOKTERAN

UNIVERSITAS PADJADJARAN

2023

LANGKAH-LANGKAH

Skrip:

```
#Mengaktifkan package
library(readr)
library(dplyr)
library(janitor)
#1. Reading data pef
pef = read_csv("https://raw.githubusercontent.com/dwi-agustian/biostat/main/pef.csv")
#1a. deduplikasi
pef = pef[!duplicated(pef$pidlink),]

#2. Reading data w5
w5 = read_csv("https://raw.githubusercontent.com/dwi-agustian/biostat/main/w5.csv")
length(unique(w5$pidlink))
get_dupes(w5,pidlink)
#2a. deduplikasi
w5 <- w5[!duplicated(w5$pidlink),]

#3. Konversi pidlink di w5 dari character jadi angka(numeric)
w5$pidlink = as.numeric(w5$pidlink)
summary(w5$pidlink)

#4. Observasi pidlink tidak valid
str(w5$pidlink)

#5. Memilih berdasarkan kritera pidlink missing
w5 = filter(w5,!is.na(pidlink))

#6. Combining dataset w5 (58.297) vs pef (58.297)
w5_pef_lj = left_join(w5, pef, by = "pidlink")
w5_pef_rj = right_join(w5, pef, by = "pidlink")
w5_pef_ij = inner_join(w5, pef, by = "pidlink")
w5_pef_fj = full_join(w5, pef, by = "pidlink")
names (w5_pef_lj)
```

#7. Mengaktifkan packages

- library(readr)
- library(dplyr)

#8. Reading data

- pefbaru <- dplyr::select(w5_pef_lj, sex, age, height, pef)

#9. Melihat missing data (N/A)

- summary(pefbaru$age)
- summary(pefbaru$height)
- summary(pefbaru$sex)
- summary(pefbaru$pef)

#10. Memilih yang tidak missing dari seluruh variabel

- pef_final = filter(pefbaru,!is.na(height), !is.na(pef), !is.na(sex), !is.na(age))

DATA PROFILE DIAGRAM