

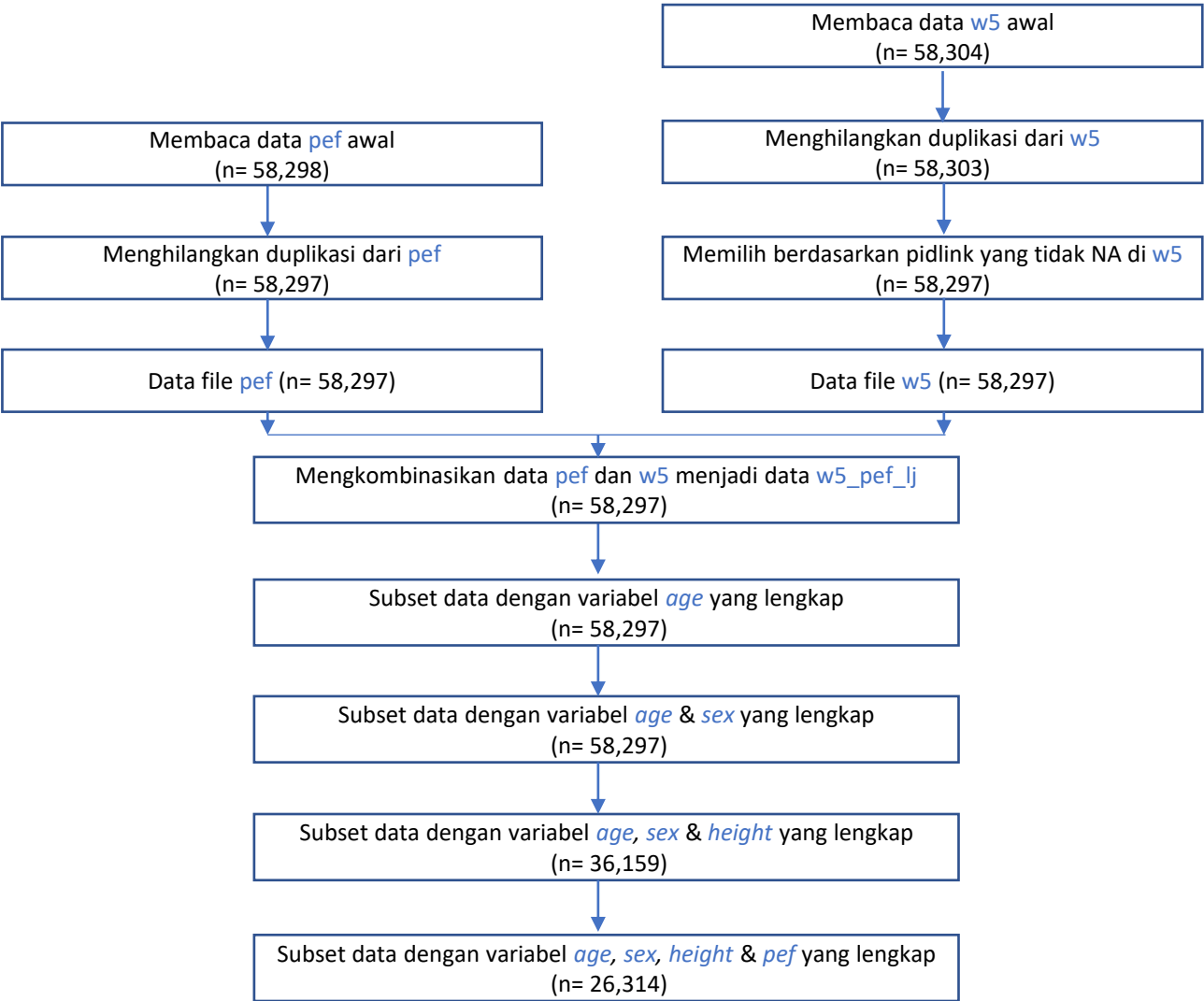
# Subsetting Data (Data Profile Diagram)

Sri Hartati

131520220006

Mata Kuliah Manajemen Data dengan R  
Prodi Magister Epidemiologi  
Universitas Padjadjaran  
2022/2023

Data Profile Diagram



R script untuk tugas subsetting data

#1. mengaktifkan paket

```
library(readr)
```

```
library(dplyr)
```

```
library(janitor)
```

```
library(dplyr)
```

#2. membaca data pef (n=58,298)

```
pef = read_csv("https://raw.githubusercontent.com/dwi-agustian/biostat/main/pef.csv")
```

#3. menghilangkan duplikasi dari pef (n=58,297)

```
pef_nodup <- pef[!duplicated(pef$pidlink),]
```

#4. membaca data w5 (n=58,304)

```
w5 = read_csv("https://raw.githubusercontent.com/dwi-agustian/biostat/main/w5.csv")
```

```
names(w5)
```

```
get_dupes(w5,pidlink)
```

#5. menghilangkan duplikasi dari w5 (n=58,303)

```
w5 <- w5[!duplicated(w5$pidlink),]
```

#6. merubah jenis variabel pidlink di w5 dari karakter menjadi numerik

```
w5$pidlink = as.numeric(w5$pidlink)
```

#7. memeriksa variabel pidlink

```
summary(w5$pidlink)
```

#8. memilih observasi di w5 yang tidak NA berdasarkan kriteria pidlink (n=58,297)

```
w5 = filter(w5,!is.na(pidlink))
```

#10. menggabungkan data w5\_na (n=58,297) dan data pef\_nodup (n=58,297) dengan key variable pidlink

```
w5_pef_lj = left_join(w5, pef_nodup, by = "pidlink")  
w5_pef_rj = right_join(w5, pef_nodup, by = "pidlink")  
w5_pef_ij = inner_join(w5, pef_nodup, by = "pidlink")  
w5_pef_fj = full_join(w5, pef_nodup, by = "pidlink")
```

#11. memilih observasi berdasarkan kriteria age yang tidak NA (n=58,297)

```
w5_pef_lj_age = filter(w5_pef_lj, !is.na(age))
```

#12. memilih observasi berdasarkan kriteria sex yang tidak NA (n=58,297)

```
w5_pef_lj_age_sex = filter(w5_pef_lj_age, !is.na(sex))
```

#13. memilih observasi berdasarkan kriteria height yang tidak NA (n=36,159)

```
w5_pef_lj_age_sex_height = filter(w5_pef_lj_age_sex, !is.na(height))
```

#14. memilih observasi berdasarkan kriteria pef yang tidak NA (n=26,314)

```
w5_pef_lj_age_sex_height_pef = filter(w5_pef_lj_age_sex_height, !is.na(pef))
```