

R-seminar 1 Oppgaver

STV1020 Vår 2021

Uke 9

1. Gå sammen i gruppe på 4-6 i seminaret. Sett navn, alder, hjemby og favoritt-film inn i ett datasett.

```
# Her vil jeg lage 4 vektorer som inneholder de dataene jeg vil ha.
Navn <-c("Audun", "Jonas", "Trygve", "Kjell", "Guri", "Erna")
Alder <-c(20, 48, 39, 12, 32, 40)
By <-c("A", "B", "C", "D", "E", "F")
Film <-c("A1", "B2", "C3", "D4", "E5", "F6")

# Her er det viktig at alle er like lange. Skulle det mange noen
# opplysninger, må dere skrive NA for å fylle ut lengden.
# Med vektorene på plass, kan datasettet lages.
New_Friends <-data.frame(Navn, Alder, By, Film)
```

2. Finn gjennomsnitt og median av alder i datasettet.

```
# Det er to måter vi kan gjøre dette på.
# Enten med de spesifikke funksjonene:
mean(New_Friends$Alder)

## [1] 31.83333

median(New_Friends$Alder)

## [1] 35.5

# Eller ved å bruke summary():
summary(New_Friends$Alder)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    12.00   23.00   35.50   31.83   39.75   48.00
```

3. Finn avstanden fra hjembyene til Oslo, og legg inn dette som en ny kolonne i datasettet. Er du usikker på hvordan, se om du finner ut av det på google.

```
# For å finne avstanden har jeg bare brukt Google Maps, ca avstand i km.
# Så lager jeg en ny vektor med avstandene i samme rekkefølge
# som byene kommer i datasettet.
Avstand <-c(201, 170, 211, 236, 179, 188)

# Så kan jeg legge til denne i datasettet ved å definere en variabel som vanlig
New_Friends$Avstand <- Avstand
```

4. `table()`-funksjonen gjør at du kan lage krysstabeller. Mellom parentesene setter du inn navnet på variablene du vil ha med. Lag en tabell med navn og avstand fra hjembyen. Hvem er lengst hjemmefra?

```
# Med table()-funksjonen setter vi inn de to variablene
table(New_Friends$Navn, New_Friends$Avstand)
```

```
##
##          170 179 188 201 211 236
##   Audun    0  0  0  1  0  0
##   Erna     0  0  1  0  0  0
##   Guri     0  1  0  0  0  0
##   Jonas    1  0  0  0  0  0
##   Kjell    0  0  0  0  0  1
##   Trygve   0  0  0  0  1  0
```

```
# Her kan vi se at den som har 1 på det høyeste tallet, helt til høyre,
# er Kjell.
```

5. Lag et diagram som viser fordelingen av alder. Om du skriver `geom_` og så trykker på tab-knappen, vil du kunne bla gjennom alle de forskjellige plottypene. Velg det du synes passer best, og prøv deg frem.

```
# For alder tenker er et histogram greit
# Først må funksjonen aktiveres med library
library(tidyverse)

ggplot(New_Friends,aes(Alder))+
  geom_bar()
```

