

# Statistik og dataanalyse

## Tællende aktivitet 1

Peter Heilbo Ratgen - perat17@student.sdu.dk

11. oktober 2020

# 1 Databehandling

Datafilen indeholder data om landes CO2-udledning per indbygger i metriske ton per indbygger. Der er to kolonner i datafilen, "Country" og "CO2". Først gemmes .xlsx-filen som en .csv fil. Denne data er dog ikke komplet, det indeholder også de lande eller territorier der ikke har opgivet data. Det giver ikke mening at betragte disse lande i statistikken, derfor fjernes disse fra datasættet. Dette gøres ved at fjerne alle datapunkter der har "." som værdi. Også datapunkter der ikke er lande skal også fjernes fra datasættet, derfor fjernes alle datapunkter hvor "income" indgår i navnet med kommandolinjeværktøjet `sed`. Her er der fjernet 10 datapunkter. Derudover fjernes datapunktet "world" også, siden det noget der kan regnes frem til. Til sidst i databehandlingen skal , ændres til ., for at R importerer data som den korrekte type.

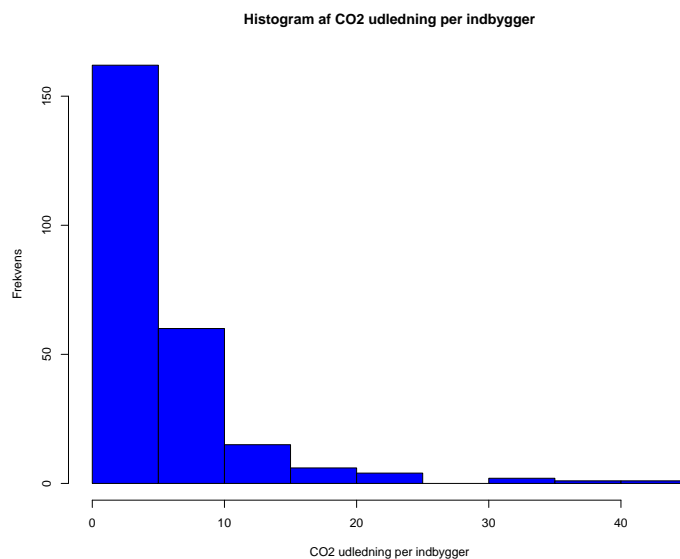
Nu kan data indporteres til R med:

```
co2 <- read.csv("CO2.csv", sep = ";", row.names = NULL)
```

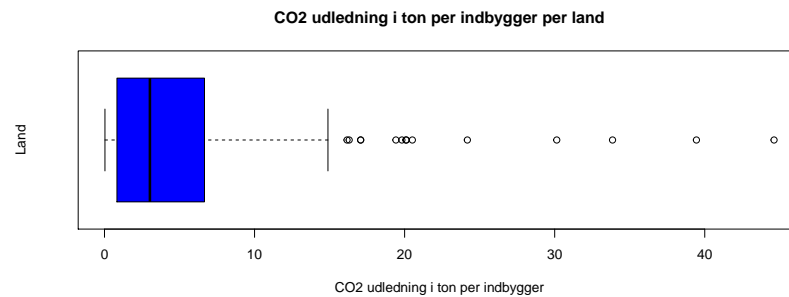
Her bruges `sep = ";"` til at indikere at værdierne i .csv formatet ikke er separeret med et komma, men med et semikolon.

## 2 Grafiske illustrationer

### 2.1 Histogram



## 2.2 Boxplot



Boxplottet fortæller at langt de fleste lande (75%, eller dem inden for det 3. kvartil) ikke udleder over 10 ton CO<sub>2</sub> per indbygger. Det fortæller også at der er nogle lande der udleder langt mere CO<sub>2</sub> per indbygger, disse er de lande der ligger uden for 1.5 gange den interkvartile afstand.

Danmark ligger indenfor det 3. kvartil med sine 6.51 ton CO<sub>2</sub> udledt per indbygger. Det vil også sige at Danmark ligger over medianen.

## 3 Deskriptiv analyse