

# Untitled

Mohamad Ali Al-Mousawy

2/8/2022

## Oppgave 1

Jeg valgte å bruke pakken jsonlite for å få inn json urlen som vi skulle bruke. Lastet så inn Tidyverse fordi den inneholder pakker som kan være nødvendige. Og ggrepel fordi den har funksjoner som jeg kan bruke på labele statene i plott. Countrycode pakken gir oss statene, der vi kan forkorte de.

```
library(jsonlite)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.1      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x purrr::flatten() masks jsonlite::flatten()
## x dplyr::lag()     masks stats::lag()
```

```
library(ggrepel)
library(rvest)
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:readr':
##
##      guess_encoding
```

```
library(countrycode)
```

```
## Warning: package 'countrycode' was built under R version 4.0.5
```

Henter inn datasett fra Json og mutater en ny variabel med forkortelse av stater.

```

Jsontdata <- fromJSON("https://static01.nyt.com/newsgraphics/2021/12/20/us-coronavirus-deaths-2021/ff0ad
Jsontdata <- Jsontdata %>%
  mutate(states = state.abb[match(name,state.name)])

Jsontdata[is.na(Jsontdata)] = "D.C."

```

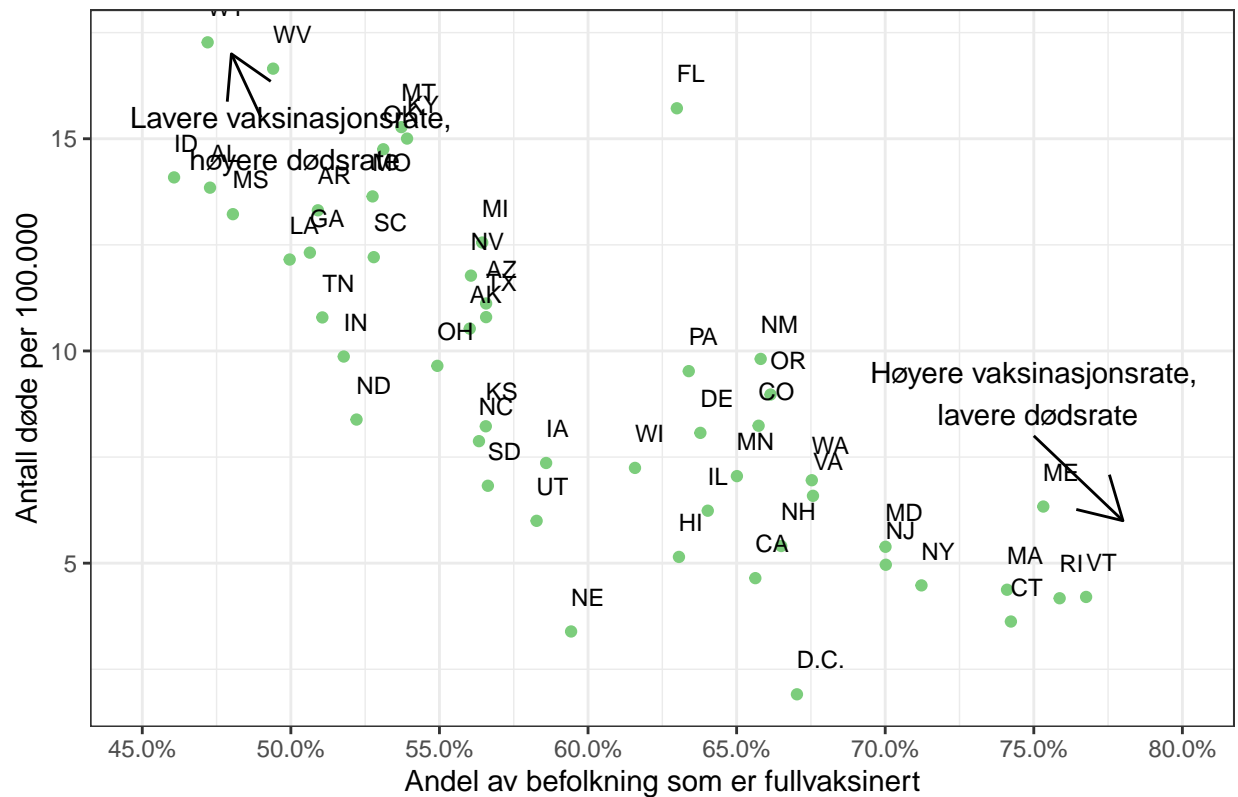
Plotter deretter datasettet, bruker diverse text og repel funksjoner som gir meg navn på stater, titler og piler. Ved å bruke annotate kan jeg ha tekst på plotten vedsiden av pilene.

```

ggplot(Jsontdata, aes(x = fully_vaccinated_pct_of_pop, y = deaths_per_100k)) +
  geom_point(colour = "palegreen3") +
  geom_text(aes(label = states),hjust=0, vjust=-1.5, size = 3) +
  labs(title = "Andel av befolkning som er fullvaksinert","Antall døde per 100.000") +
  ggtitle("20 månedelige døde i gjennomsnitt per 100k") +
  scale_x_continuous(labels = scales::percent, limits=c(0.45, 0.80), breaks=seq(0.45, 0.80, by = 0.05))
  theme_bw() +
  xlab("Andel av befolkning som er fullvaksinert") +
  ylab("Antall døde per 100.000") +
  annotate("text", x=0.50, y=15,
    label= "Lavere vaksinasjonsrate,\n høyere dødsrate") +
  annotate("segment", x = 0.49,
    xend = 0.48, y = 15.5,
    yend = 17, colour = "black", arrow = arrow()+
  annotate("text", x=0.75, y=9,
    label= "Høyere vaksinasjonsrate,\n lavere dødsrate") +
  annotate("segment", x = 0.75,
    xend = 0.78, y = 8,
    yend = 6, colour = "black", arrow = arrow())

```

## 20 månedlige døde i gjennomsnitt per 100k



## Oppgave 2

En linjær funksjon har likningen  $Y = a + bX$ , hvor verdien av  $Y$  er den avhengige variabelen. Der verdien av  $y$  når  $x = 0$ .

Modellen sier oss at temperaturen er den avhengige variabelen ( $y$ ), som blir forklart av en fullvaksinert prosent av populasjonen. Der 1 enhets endring i variabelen  $x$  gir  $B$  endring i variabelen  $y$ . Ved å sette inn en  $(lm)$  funksjon som er linær funksjon, får vi korrelasjonen mellom Dødsfall per 100k og Fullvaksinert prosentandel. der den har en stigningstall på -36,66.

```
lm(deaths_per_100k ~ fully_vaccinated_pct_of_pop, data = Jsondata)
```

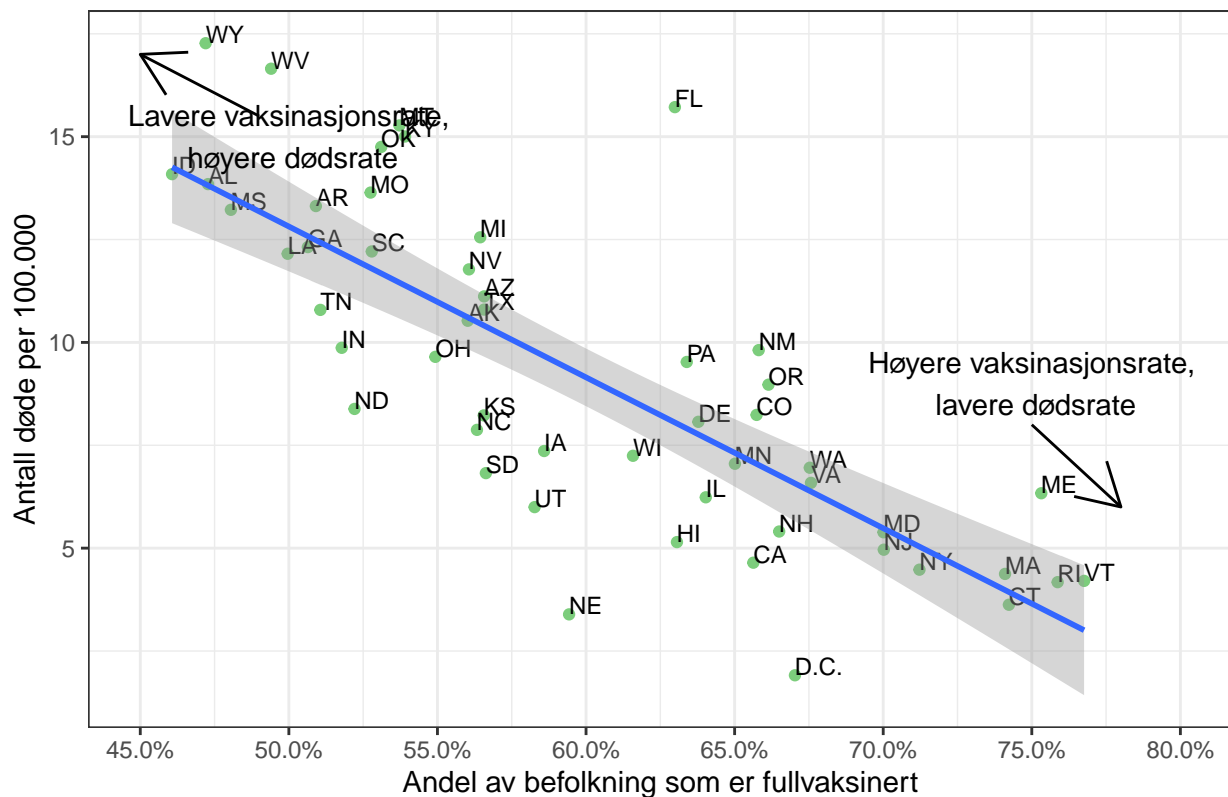
```
##
## Call:
## lm(formula = deaths_per_100k ~ fully_vaccinated_pct_of_pop, data = Jsondata)
##
## Coefficients:
##              (Intercept)  fully_vaccinated_pct_of_pop
##                   31.15                   -36.66
```

Ved å se tallene ovenfor ser vi at det er en negativ sammenheng mellom  $y \sim x$ . I plotten ser vi dette enda klarere ved å se at linjen "faller".

```
ggplot(Jsondata, aes(x = fully_vaccinated_pct_of_pop, y = deaths_per_100k)) +
  geom_point(colour = "palegreen3") +
  geom_text(aes(label = states), hjust=0, vjust=0, size = 3) +
  geom_smooth(method = lm) +
  xlab("Andel av befolkning som er fullvaksinert") +
  ylab("Antall døde per 100.000") +
  ggtitle("20 månedelige døde i gjennomsnitt pr 100.000") +
  theme(plot.title = element_text(hjust = 0.5, size = 15)) +
  theme(axis.title.x = element_text(hjust = 0.5, size = 10)) +
  theme(axis.title.y = element_text(hjust = 0.5, size = 10)) +
  scale_x_continuous(labels = scales::percent, limits=c(0.45, 0.80), breaks=seq(0.45, 0.80, by = 0.05)) +
  theme_bw() +
  annotate("text", x=0.50, y=15,
    label= "Lavere vaksinasjonsrate,\n høyere dødsrate") +
  annotate("segment", x = 0.49,
    xend = 0.45, y = 15.5,
    yend = 17, colour = "black", arrow = arrow()) +
  annotate("text", x=0.75, y=9,
    label= "Høyere vaksinasjonsrate,\n lavere dødsrate") +
  annotate("segment", x = 0.75,
    xend = 0.78, y = 8,
    yend = 6, colour = "black", arrow = arrow())
```

## 'geom\_smooth()' using formula 'y ~ x'

## 20 månedelige døde i gjennomsnitt pr 100.000



Samme plot som oppgave 2 uten 95% signifikant som skygger i plotten ovenfor. Ved å ha det samme de samme funksjonene som plotten over, men la til `se = FALSE` funksjon for å få bort skyggen som er 95% signifikant.

```
ggplot(Jsondata, aes(x = fully_vaccinated_pct_of_pop, y = deaths_per_100k)) +
  geom_point(colour = "palegreen3") +
  geom_text(aes(label = states), hjust=0, vjust=0, size = 3) +
  geom_smooth(method = "lm", se = FALSE) +
  xlab("Andel av befolkning som er fullvaksinert") +
  ylab("Antall døde per 100.000") +
  ggtitle("20 månedelige døde i gjennomsnitt pr 100.000") +
  theme(plot.title = element_text(hjust = 0.5, size = 15)) +
  theme(axis.title.x = element_text(hjust = 0.5, size = 10)) +
  theme(axis.title.y = element_text(hjust = 0.5, size = 10)) +
  scale_x_continuous(labels = scales::percent, limits=c(0.45, 0.80), breaks=seq(0.45, 0.80, by = 0.05))
  theme_bw() +
  annotate("text", x=0.50, y=15,
    label= "Lavere vaksinasjonsrate,\n høyere dødsrate") +
  annotate("segment", x = 0.49,
    xend = 0.45, y = 15.5,
    yend = 17, colour = "black", arrow = arrow()) +
  annotate("text", x=0.75, y=9,
    label= "Høyere vaksinasjonsrate,\n lavere dødsrate") +
  annotate("segment", x = 0.75,
    xend = 0.78, y = 8,
    yend = 6, colour = "black", arrow = arrow())
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

## 20 månedelige døde i gjennomsnitt pr 100.000

