

# Arbeidskrav 3

“[8]”

## Arbeidskrav 3

```
rm (list = ls())  
library(tidyverse)
```

```
-- Attaching packages ----- tidyverse 1.3.2 --  
v ggplot2 3.4.0      v purrr   0.3.4  
v tibble  3.1.8      v dplyr  1.0.9  
v tidyr   1.2.0      v stringr 1.4.0  
v readr   2.1.2      v forcats 0.5.1  
-- Conflicts ----- tidyverse_conflicts() --  
x dplyr::filter() masks stats::filter()  
x dplyr::lag()    masks stats::lag()
```

```
library(janitor)
```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

```
library(lubridate)
```

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

```
date, intersect, setdiff, union
```

```
library(jsonlite)
```

Attaching package: 'jsonlite'

The following object is masked from 'package:purrr':

```
flatten
```

```
df <- fromJSON("https://static01.nyt.com/newsgraphics/2021/12/20/us-coronavirus-deaths-2021")
```

```
#dataset from the webpage
```

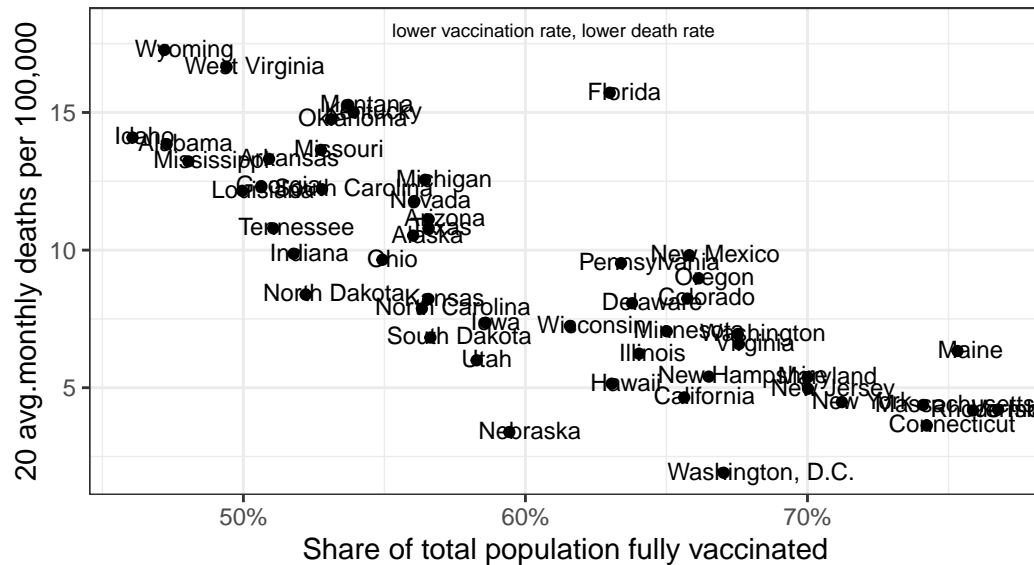
```
df <- df %>%  
  rename(states = name)
```

```
#changing name on the variabel
```

```
#plotting in dataset
```

```
df %>%  
  ggplot(aes(x=fully_vaccinated_pct_of_pop, y=deaths_per_100k, label=states))+  
  geom_point() +  
  geom_text(hjust=.3, vjust=.4, size=3.2)+  
  annotate("text", x=0.61, y=18, label = "lower vaccination rate, lower death rate", size=3.2)+  
  ggtitle("COVID-19 deaths since universal adult vaccine eligibility  
    compared with vaccination rates")+  
  ylab("20 avg.monthly deaths per 100,000") +  
  xlab("Share of total population fully vaccinated") +  
  scale_x_continuous(labels = scales::percent)+  
  theme_bw()
```

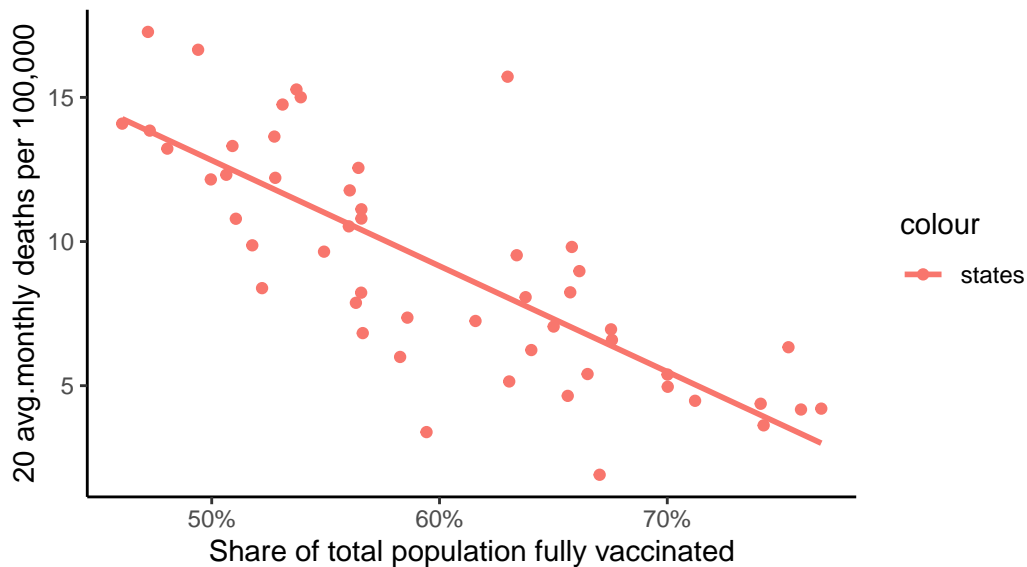
## COVID-19 deaths since universal adult vaccine eligibility compared with vaccination rates



```
#plotting in with the lm function
df %>%
  ggplot(aes(x=fully_vaccinated_pct_of_pop, y=deaths_per_100k, col="states")) +
  geom_point() +
  ggtitle("COVID-19 deaths since universal adult vaccine eligibility
    compared with vaccination rates")+
  ylab("20 avg.monthly deaths per 100,000") +
  xlab("Share of total population fully vaccinated") +
  geom_smooth(method = "lm", se = FALSE) +
  scale_x_continuous(labels = scales::percent)+
  theme_classic()
```

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

## COVID-19 deaths since universal adult vaccine eligibility compared with vaccination rates



By looking at the graph we can see that there are a negative correlation between the numbers of death and the proportion of the population vaccinated.

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

Call:

```
lm(formula = deaths_per_100k ~ fully_vaccinated_pct_of_pop, data = df)
```

Coefficients:

(Intercept)	fully_vaccinated_pct_of_pop
31.15	-36.66

There is a connection between the number of death and the proportion of the population vaccinated. We can see that the states where the proportion of the population is vaccinated, the death rate is much lower, compared too the states where there is a low proportion of the population who is vaccinated. The number -36.66 gives us the information about how much the deaths rate would go down, if the proportion of the population got vaccinated. The numbers give us the information about the effect the vaccine has in a state.

```
lm(`deaths_per_100k` ~ `fully_vaccinated_pct_of_pop`, data = df) + geom_smooth(method = lm)
```

NULL

```
summary(lm(`deaths_per_100k` ~ `fully_vaccinated_pct_of_pop`, data = df))
```

Call:

```
lm(formula = deaths_per_100k ~ fully_vaccinated_pct_of_pop, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.9688	-1.4000	-0.0834	1.1955	7.6635

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	31.149	2.531	12.305	< 2e-16 ***
fully_vaccinated_pct_of_pop	-36.663	4.183	-8.764	1.33e-11 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.49 on 49 degrees of freedom

Multiple R-squared: 0.6105, Adjusted R-squared: 0.6026

F-statistic: 76.81 on 1 and 49 DF, p-value: 1.328e-11

By using this function we get the a summery of min, median and the max.