# Mappeoppgave 3

## Mappeoppgave 3

```
suppressPackageStartupMessages(library(tidyverse, warn.conflicts = FALSE))
```

```
## Warning: package 'tibble' was built under R version 4.1.2
```

```
## Warning: package 'tidyr' was built under R version 4.1.2
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

```
suppressPackageStartupMessages(library(rvest, warn.conflicts = FALSE))
```

```
## Warning: package 'rvest' was built under R version 4.1.2
```

```
suppressPackageStartupMessages(library(proto, warn.conflicts = FALSE))
```

```
## Warning: package 'proto' was built under R version 4.1.2
```

```
df <- read_html("https://www.motor.no/aktuelt/motors-store-vintertest-av-rekkevidde-pa-elbiler/217132")

df <- df %>% html_table()
df <- df[[1]]
df <- df %>%
  rename(Modell = X1,
         WLTP = X2,
         STOPP = X3,
         Avvik = X4)

df = select(df, -c(1, 4))
df =df[-1,]
df <-df[!grepl("x",df$STOPP),]

df$STOPP<-gsub("km","",as.character(df$STOPP))
df$STOPP <- as.numeric(as.character(df$STOPP))
df$WLTP <- sub("^(\\d{3}).*$", "\\1",df$WLTP)
```

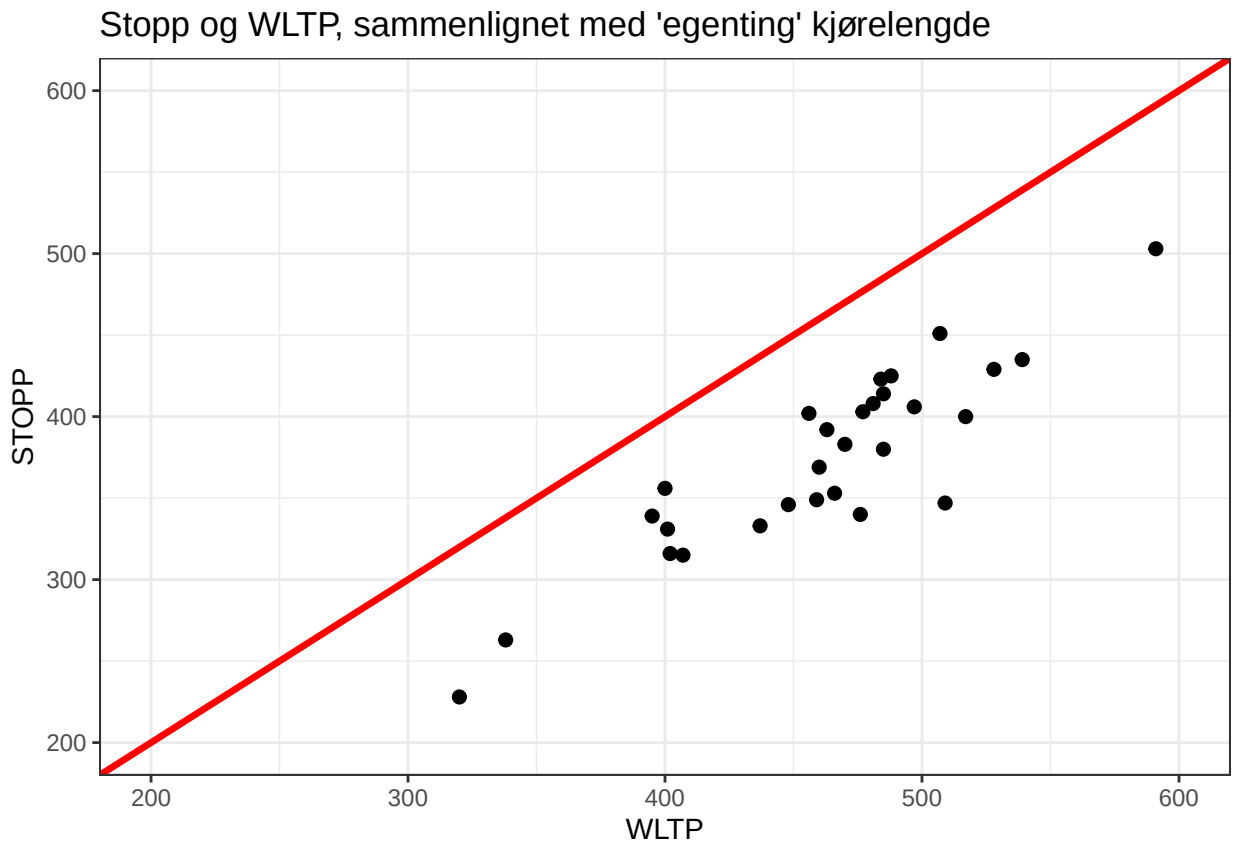
Brukte stackoverflow som kilde: https://stackoverflow.com/questions/21675379/r-only-keep-the-3-x-first-characters-in-a-all-rows-in-a-column/21675473

## Oppgave 1

```
df$WLTP <- as.numeric(as.character(df$WLTP))

df %>%
  ggplot(aes(x=WLTP, y=STOPP)) +
  geom_point(size = 2, col="black") +
  geom_abline(size = 1.2, col = "red") +
  scale_x_continuous(limits= c(200, 600), breaks = seq(200, 600, by = 100)) +
  scale_y_continuous(limits= c(200, 600), breaks = seq(200, 600, by = 100)) +
  labs(title="Stopp og WLTP, sammenlignet med 'egenting' kjørelengde",
       x ="WLTP",
       y = "STOPP") +
  theme_bw()
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```




## Oppgave 2

```
lm(STOPP ~ WLTP, data =df)
```

```
##
## Call:
## lm(formula = STOPP ~ WLTP, data = df)
```

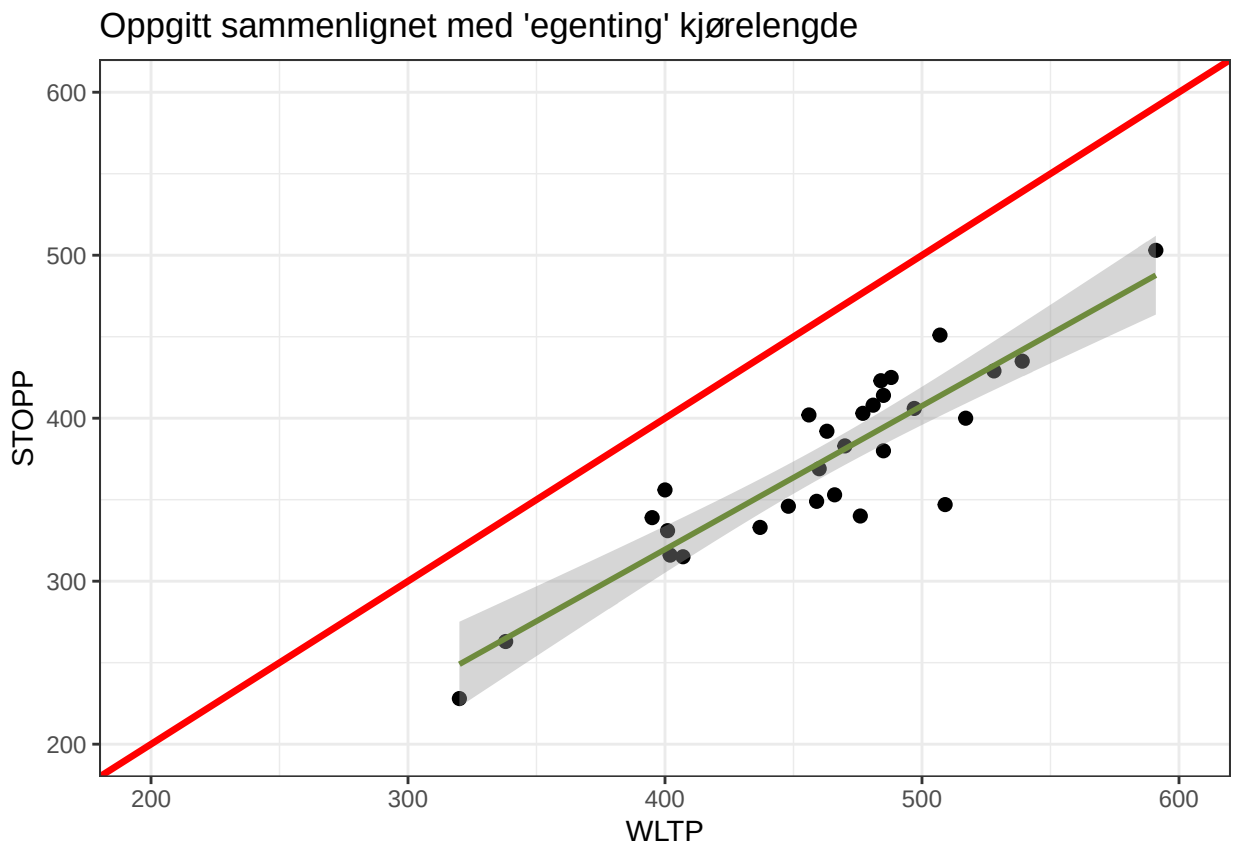
```
##
## Coefficients:
## (Intercept)         WLTP
##    -26.6450       0.8671
```

```
df %>%
  ggplot(aes(x=WLTP, y=STOPP)) +
  geom_point(size = 2, col="black") +
  geom_abline(size = 1.2, col = "red") +
  geom_smooth(method = lm, col ="darkolivegreen4") +
  scale_x_continuous(limits= c(200, 600), breaks = seq(200, 600, by = 100)) +
  scale_y_continuous(limits= c(200, 600), breaks = seq(200, 600, by = 100)) +
  labs(title="Oppgitt sammenlignet med 'egenting' kjørelengde",
       x ="WLTP",
       y = "STOPP") +
  theme_bw()
```

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

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```




Den grønne linjen lager en lineær modell ut av de punktene som vises på figuren. Dette gjør det lettere å sammenligne reell og oppgitt kjørelengde, noe vi lett kan se at ikke er det samme. Det grå område rundt den grønne linjen er et konfidensintervall, Det viser hvordan linja kan endre seg ved gjenstatte forsøk.