SOK-1005-assignment-2

"Marte J. Skadssem Haetta Eikelmann"

Prepare to answer the assignment:

```
#removing all previous data to start on a clean sheet.

rm(list=ls())


#loading necessary packages

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(rvest)
```

```
Attaching package: 'rvest'

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

    guess_encoding
```

```
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(readr)
```

Task 1

```
#scraping the table from the URL

url <- "https://www.motor.no/aktuelt/motors-store-vintertest-av-rekkevidde-pa-elbiler/2171

table <- url %>%
  read_html() %>%
  html_nodes("table") %>%
  html_table()
head(table[[1]])
```

```
# A tibble: 6 x 4
  X1                                   X2            X3      X4
  <chr>                                <chr>         <chr>   <chr>
1 Modell (temp. varierte fra 0° til -10°) WLTP-tall     STOPP   Avvik
2 Tesla Model 3 LR Dual motor          614 km/14,7 kWh 521 km -15,15 %
```

```
3 Mercedes-Benz EQS 580 4matic                645 km/18,3 kWh 513 km -20,47 %
4 BMW iX xDrive50                             591 km/21,4 kWh 503 km -14,89 %
5 Tesla Model Y LR Dual motor                 507 km/16,9 kWh 451 km -11,05 %
6 Volkswagen ID.3 PRO S                       539 km/16,3 kWh 435 km -19,29 %
```

```r
#naming the first table car_test

car_test <- table[[1]]


car_test <- car_test %>%
  row_to_names(1, remove_rows_above = FALSE) %>% #set the first row to columnnames
  rename("modell" = "Modell (temp. varierte fra 0° til -10°)","wltp_tall" = "WLTP-tall","a


#extracting numbers from each string

car_test <- car_test %>%
  mutate(stopp = parse_number(stopp), wltp_tall = parse_number(wltp_tall), avvik = parse_n
  mutate(avvik = avvik/100)
```

```
Warning: 2 parsing failures.
row col expected actual
 19  -- a number      x
 26  -- a number      x

Warning: 2 parsing failures.
row col expected actual
 19  -- a number      x
 26  -- a number      x
```

```r
#plotting the data

car_test %>%
  ggplot(aes(x=wltp_tall,y=stopp, label=modell)) +
  geom_point(color="blue") +
  geom_abline(intercept = 0, slope = 1, size = 0.3, color="red") + #making a 45 degree lin
  geom_text(hjust=0.5, vjust=0.5, check_overlap = TRUE) +
  xlim(300,650) +
  ylim(250,650) +
  labs(title="Sammenheng mellom WLTP-tall og antall faktisk kjørte kilometer", x="WLTP-tal
  theme_classic()
```
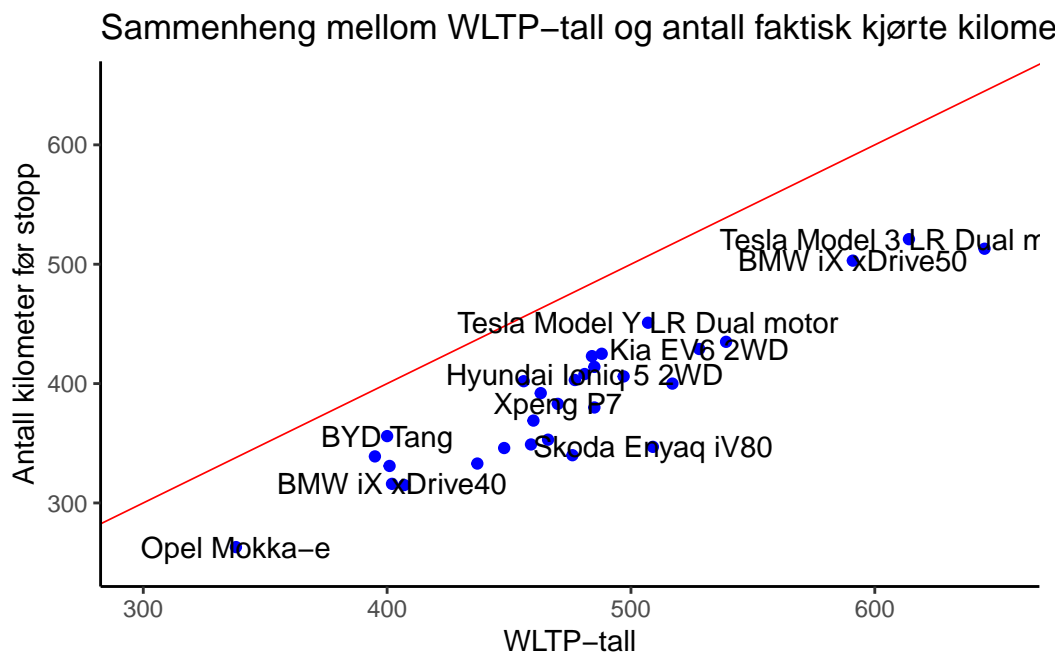
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.

Warning: Removed 3 rows containing missing values (`geom_point()`).

Warning: Removed 3 rows containing missing values (`geom_text()`).
```


Sammenheng mellom WLTP−tall og antall faktisk kjørte kilome

Task 2

```
lm(stopp ~ wltp_tall, data = car_test)
```

```
Call:
lm(formula = stopp ~ wltp_tall, data = car_test)

Coefficients:
(Intercept)     wltp_tall
   -26.6450        0.8671
```

The lm() function is used to fit a linear model based on the dataframe we put into it. The two values that appeare when I run the code represent the linear model's intercept with the y-axis and the rate of increase. Both values are constants, which means that the values can be multiplied with other values in an expression, but do not change itself.

```
#plotting the data

car_test %>%
  ggplot(aes(x=wltp_tall,y=stopp, label=modell)) +
  geom_point(color="blue") +
  geom_abline(intercept = 0, slope = 1, size = 0.3, color="red") + #making a 45 degree lin
  geom_smooth(method = lm) + #plotting the fitted linear model
  geom_text(hjust=0.5, vjust=0.5, check_overlap = TRUE) +
  xlim(300,650) +
  ylim(250,650) +
  labs(title="Sammenheng mellom WLTP-tall og antall faktisk kjørte kilometer", x="WLTP-tal
  theme_classic()
```

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

Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).

Warning: The following aesthetics were dropped during statistical transformation: label
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?

Warning: Removed 3 rows containing missing values (`geom_point()`).

Warning: Removed 3 rows containing missing values (`geom_text()`).

Sammenheng mellom WLTP−tall og antall faktisk kjørte kilome

(figure: scatter plot of "Antall kilometer før stopp" (y-axis) versus "WLTP−tall" (x-axis) with a blue regression line, grey confidence band, and a red reference line. Labeled points include: Tesla Model 3 LR Dual m, BMW iX xDrive50, Tesla Model Y LR Dual motor, Kia EV6 2WD, Hyundai Ioniq 5 2WD, Xpeng P7, Skoda Enyaq iV80, BYD Tang, BMW iX xDrive40, Opel Mokka−e)