D2 DoE project AS 2025

...

D2 DoE Project AS 2025: Tasks & Answers

Preliminaries

For this project you will perform **your own small scale experiment**. You will have to choose an interesting response variable that you can easily measure, define multiple explanatory variables and plan an experiment. After doing the experiment, you will proceed to analyze your results and provide an accurate presentation of the results and your conclusions.

This file contains your tasks, as well as fields where you should put your answers. Please edit this .qmd file and render it as pdf afterwards. Upload the pdf on the moodle, as well as your collected data (.xlsx or .csv).
Deadline for the project is December 12th, 9 am.

You do the project as **small groups of 2-4 persons**, as defined by the breakout rooms of the Teams session today. If you have problems finding a group or working together, please contact Stefanie Feiler as soon as possible.

Before you start with the project, also make sure that you have added your names in the "author" part above.

Note that we do *not* judge whether your experiments as such were successful, only whether you did your data analysis well and whether your answers are consistent. **If your model is not satisfying in the end, this does not matter** - as long as you comment on it correctly.

If you have questions, **feel free to ask** the lecturers and/or your local coaches.

Step 1: Project definition

You should choose a project which does not cost you too much time. The requirements are that you find some activity that

- All of you can do, with at least 3 complete repetitions per person
- Allows you to measure some quantitative outcome
- You can also repeat under different conditions.

Examples are

- Doing sports (running up stairs, do push-ups, …)
- Simple cooking (microwave popcorn, tea, drinks, …)
- Online tests (reaction, calculus, puzzles, …)
- …

We are sure that you will find something and are looking forward to your ideas!

This is our research question:

Does the WiFi download speed depend on the device type, the location (home vs. uni), and the distance to the router?

Step 2: Planning

1. Pretend that this is a real large project:
   Think about 10 different explanatory variables (= "influence factors") which may influence your outcome. We don't mind you getting creative ("wearing a hat: yes/no", …). Please note: when the word "factor" is used this report, it does *not* define the data type of R. In DoE, it is another word for "input variable".

    1. Narrow down the possibilities until you have decided on

        - one quantitative (numerical) outcome
        - (at least) two influence factors (quantitative or discrete)
        - and the third influence factor is the person who performs the experiment, i.e. the persons of your small group.

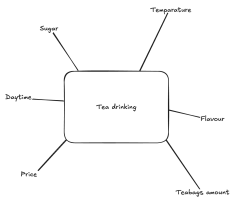
    This is only the compromise for feasibility in the course. In a real project, starting with ten factors is fine — you can test them all, as long as you do the thinking before the doing.

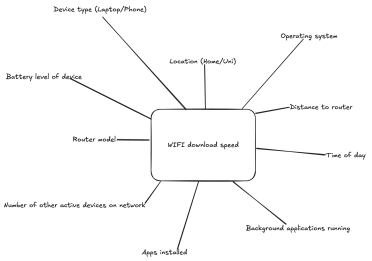
For instance, you could decide to measure the time needed for doing a certain number of calculations with an online-tool (e.g., http://sikore.schiffner-tischer.de/), or an online reaction test, typing with your dominant / non-dominant hand, in the morning and evening.

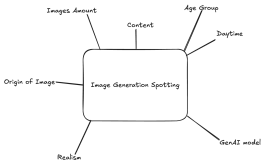
However, exactly that last option is forbidden, as well as using the website *humanbenchmark.com* - otherwise we'll get too many too similar projects and that will be boring ;)

Document your process:

This picture shows our brainstorming results (mindmap, Ishikawa diagram, …):







These is what we have decided on:

Outcome:

Download speed (Mbps)

… measured in this way:

Using the speedtest.net online tool

Our remaining influence factors are:

	Name	Range / levels	factor type: random or fixed?
Factor 1	Decive	Phone, Laptop	Fixed
Factor 2	Location	Home, Uni	Fixed
Factor 3	Approx. distance to Router	1m, 5m, 10m	Fixed
Factor 4	Persons (of your group)	Elia, Yannik, Viviane	Random
Factor 5	Daytime	Any, HH:MM:SS	Random

Step 3: Experimental Setup

We are going to proceed as follows:

Each group member measures the download speed at two locations (Home, Uni). At each location, measurements are taken with two devices (Phone, Laptop) at varying distances (1m, 5m, 10m) from the router.

Do you need to randomize the order of your runs?

Yes, to minimize the effect of temporary network fluctuations or time-dependent variables.

Is there anything else to consider?

Ensure no heavy background tasks (like App Updates, Backups) are running during the test, to avoid blocking factors.

Is there anything which is not ideal in terms of experimentation, but where you did not find a better solution?

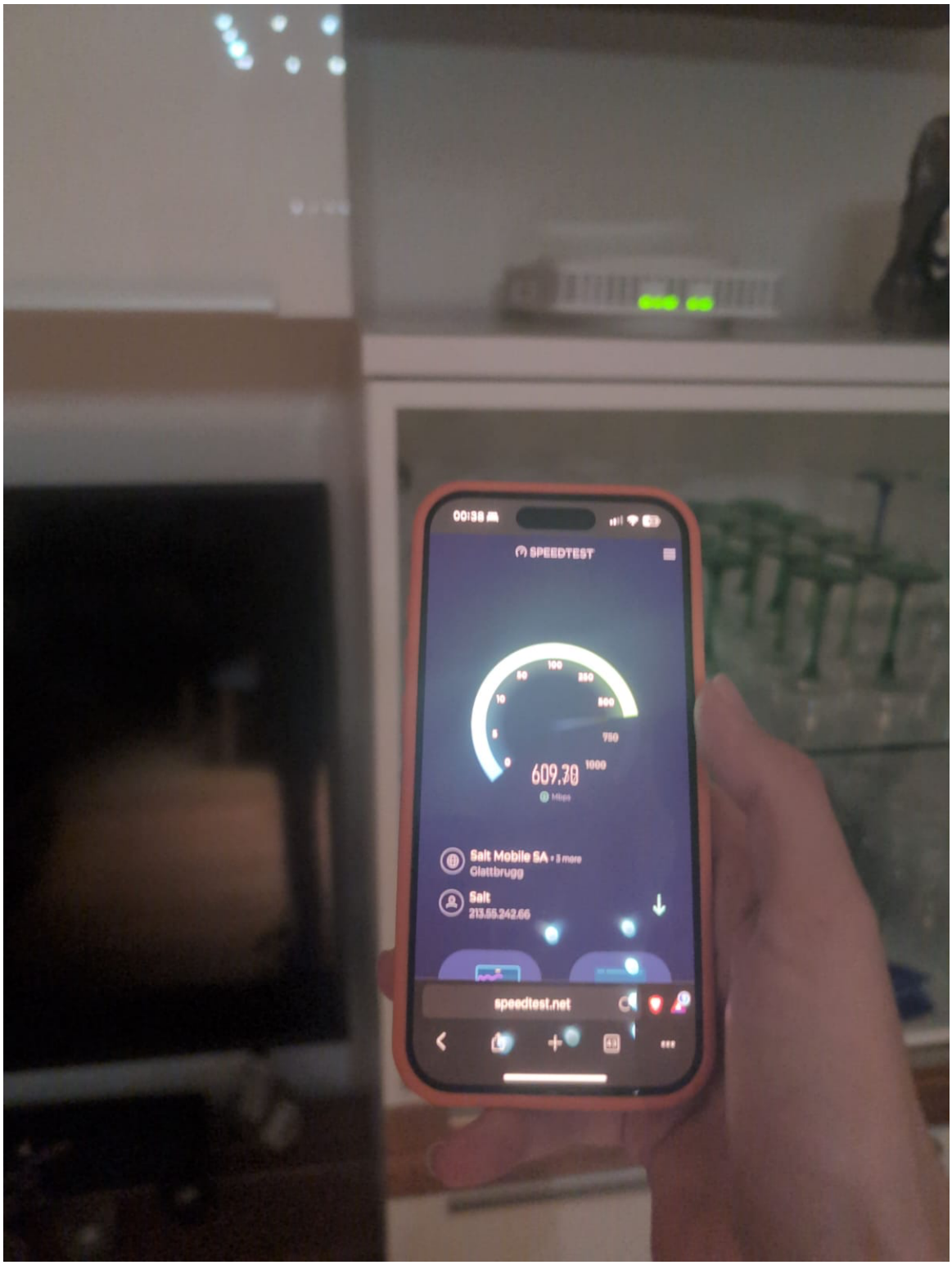
In some cases we were unsure about the exact distance to the router and had to put our best guess, which is not ideal.

Step 4: Experiments!

Do your experiments, and collect your data. Remember that it is obligatory to do **at least 3 replicates** per person and factor combination.

Do not forget to upload the data as .xslx or .csv on the moodle.

This is a fun picture showing how we did the experiments:



Are there any remarks you'd like to add with respect to the experimentation step?

Some measurements at the university varied highly due to network load.

Step 5: Data analysis

We strongly recommend that **each of your group does this part** separately - then meet up, exchange your approaches, compare the results, decide on the best plots, and on what to enter into the report.
Only by actually *doing* these tasks you'll see whether you have mastered this part of the course content.

Here is the R code used for the analysis:

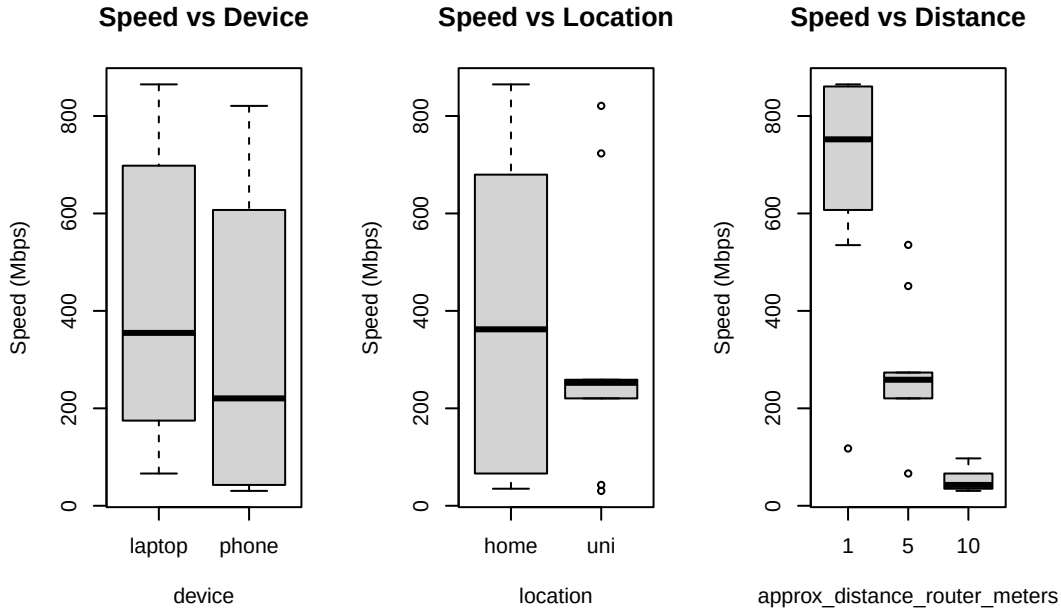
Entering the data into R:

```
df <- read.csv("d2_doe_data.csv")

df <- na.omit(df) # remove of missing data, we can remove as soon as we have all

df$person <- as.factor(df$person)
df$device <- as.factor(df$device)
df$location <- as.factor(df$location)
df$approx_distance_router_meters <- as.factor(df$approx_distance_router_meters)

head(df)
```

```
  person device approx_distance_router_meters download_speed_mbit_result
1   elia  phone                             10                      35.02
2   elia laptop                              5                     258.70
3   elia  phone                              1                     752.10
4   elia laptop                             10                      66.20
5   elia laptop                              1                     864.88
6   elia  phone                              5                     220.40
  location     time
1     home 09:56:00
2      uni 16:34:00
3     home 11:54:00
4     home 10:02:00
5     home 09:05:00
6      uni 15:54:00
```

This plot shows our data:

```
par(mfrow=c(1,3))
boxplot(download_speed_mbit_result ~ device, data = df, main = "Speed vs Device", ylab = "
boxplot(download_speed_mbit_result ~ location, data = df, main = "Speed vs Location", ylab
boxplot(download_speed_mbit_result ~ approx_distance_router_meters, data = df, main = "Spe
```



Now we calculate the model:

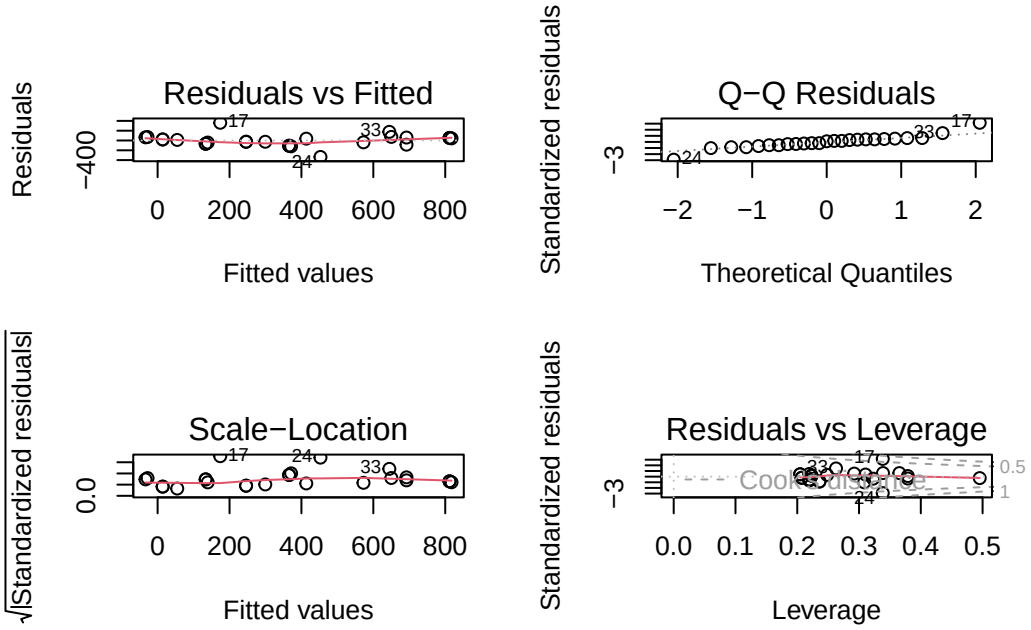
```
model <- lm(download_speed_mbit_result ~ person + device + location + approx_distance_rout
summary(model)
```

```
Call:
lm(formula = download_speed_mbit_result ~ person + device + location +
    approx_distance_router_meters, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-335.37  -67.50   11.74   52.95  361.25

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                       811.93      64.02  12.682 2.06e-10 ***
personviviane                    -239.31      89.01  -2.689   0.0150 *
personyannik                        5.61      68.82   0.082   0.9359
devicephone                      -119.64      59.73  -2.003   0.0605 .
locationuni                       -47.77      70.76  -0.675   0.5082
approx_distance_router_meters5   -398.43      71.73  -5.555 2.84e-05 ***
approx_distance_router_meters10  -678.23      73.62  -9.213 3.11e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 141.5 on 18 degrees of freedom
Multiple R-squared:  0.8415,    Adjusted R-squared:  0.7886
F-statistic: 15.92 on 6 and 18 DF,  p-value: 2.556e-06
```

These are the model diagnostics, i.e. how we have checked for the assumptions:

```
par(mfrow=c(2,2))
plot(model)
```



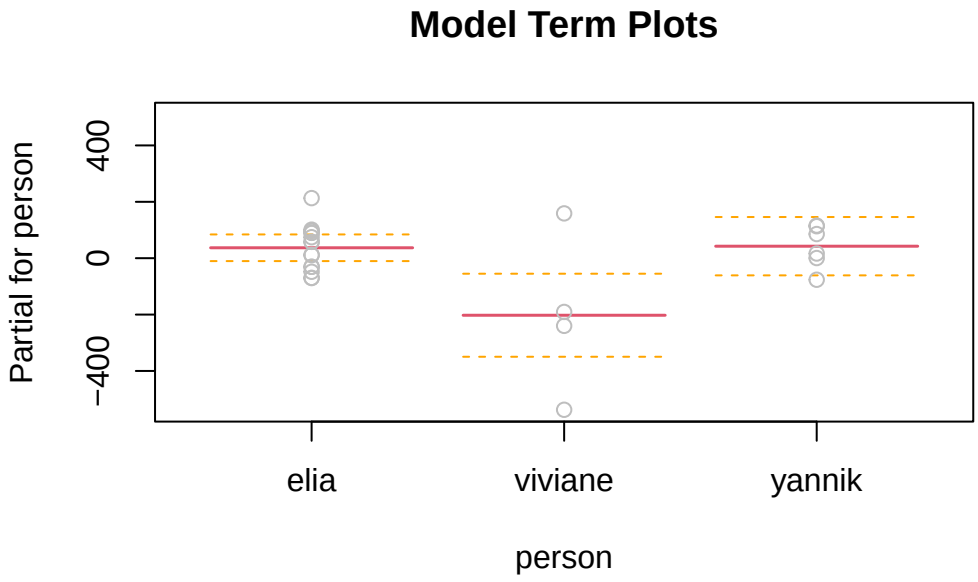

```
summary(model)
```

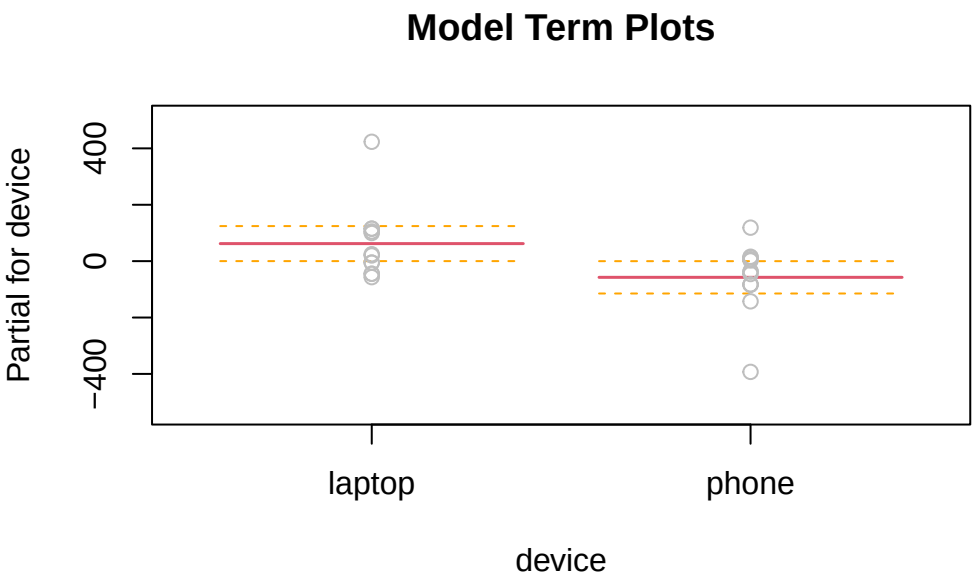
```
Call:
lm(formula = download_speed_mbit_result ~ person + device + location +
    approx_distance_router_meters, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-335.37  -67.50   11.74   52.95  361.25

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       811.93      64.02  12.682 2.06e-10 ***
personviviane                    -239.31      89.01  -2.689   0.0150 *
personyannik                        5.61      68.82   0.082   0.9359
```

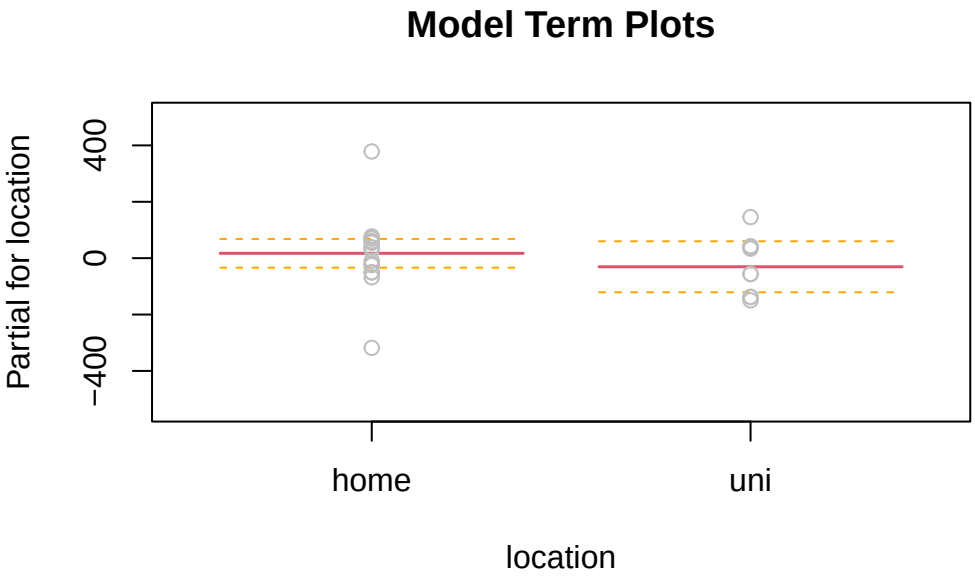
```
devicephone                      -119.64      59.73  -2.003   0.0605 .
locationuni                       -47.77      70.76  -0.675   0.5082
approx_distance_router_meters5   -398.43      71.73  -5.555 2.84e-05 ***
approx_distance_router_meters10  -678.23      73.62  -9.213 3.11e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 141.5 on 18 degrees of freedom
Multiple R-squared:  0.8415,    Adjusted R-squared:  0.7886
F-statistic: 15.92 on 6 and 18 DF,  p-value: 2.556e-06
```

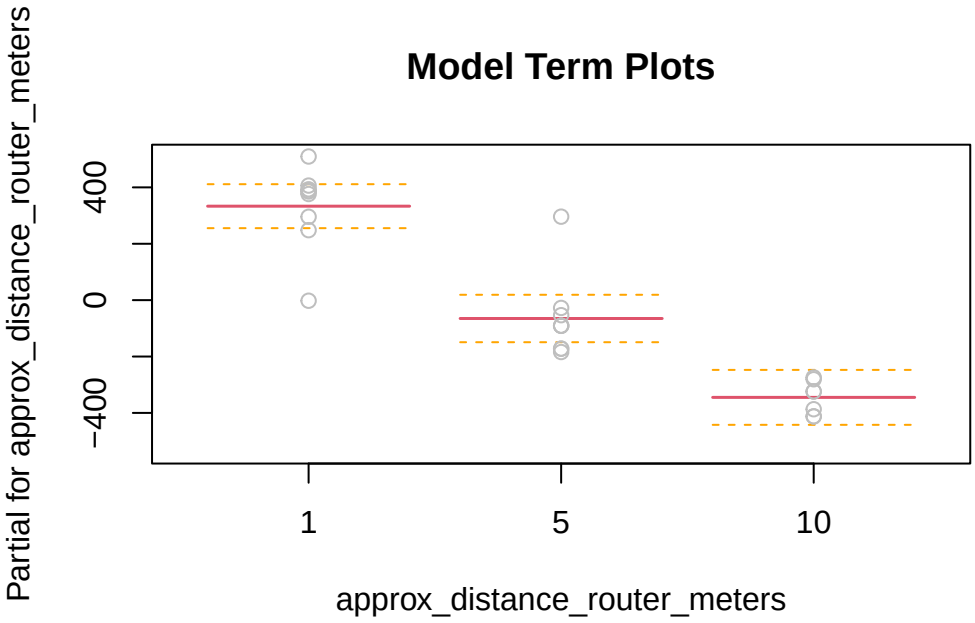
This is a plot which visualizes the model we have found:

```r
termplot(model, partial.resid = TRUE, se = TRUE, main = "Model Term Plots")
```




Model Term Plots
Partial for device
400
0
−400
laptop
phone
device


Model Term Plots
Partial for location
400
0
−400
home
uni
location



Our conclusions on the model reliability are:

[short answer]

Altogether we have found out that:

[short answer]

Step 6: Final Comments

This is what we can say about interactions in the data:

[short answer]

Is there anything else you want to add, e.g.,

- What you would do differently next time?
- What you have learnt in this project which surprised you?
- …

[add, if applicable]

Step 7: Report

Render the .qmd file (the “render” button with the large arrow at the top of the qmd file editor in RStudio) in order to obtain a **pdf** and upload it, together with your data (as .xlsx or .csv file), on the moodle.

For rendering the file, you need the packages *rmarkdown* and for the pdf generation some latex installation like e.g. *tinytex*. These can be installed by

- running these commands below in RStudio:
```
install.packages("rmarkdown")
install.packages("tinytex")
tinytex::install_tinytex()
```
- or via the terminal in RStudio as shown in this short video.

Of course, neither your RStudio nor your R version should be too old versions.

The installation should be easy. However, if your group does not manage to generate the pdf, please stop before you get too frustrated. In this case, edit the header of this document such that it produces an html file, which you then can print to pdf - if the layout is not optimal, it does not matter that much. Nevertheless: producing nice markdown documents is very satisfying, so give it a try!

Please also submit a post with your research question, a nice picture, and comments on your experiences to this Padlet - we would like that the other groups can see and admire your work. It’s anonymous, so no need to be embarrassed.

Grading

For this report, there are 0.5 points per item below, i.e. 4 points altogether:

Item	OK	something's missing	not answered / wrong
research question, outcome & measuring it			
picture of brainstorming results			
factors			
experimental setup & data collection, picture			
analysis in R: plot, model (consistent with your planning)			
analysis in R: plot of results & stating the findings			
model diagnostics (R and comments)			
interactions			

This means that the final grade is combined of these 4 points, 4 points for attendance, and 32 points in the exam.