

Thesis Progress Overview

Ties van der Veen

13-4-2020

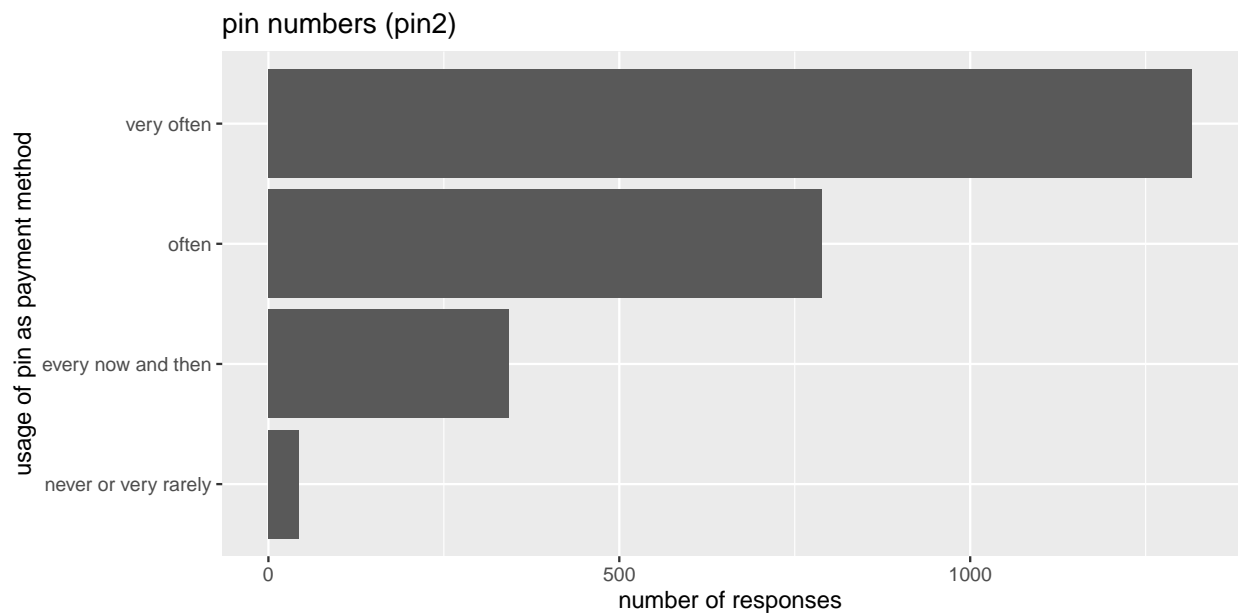
Progress overview

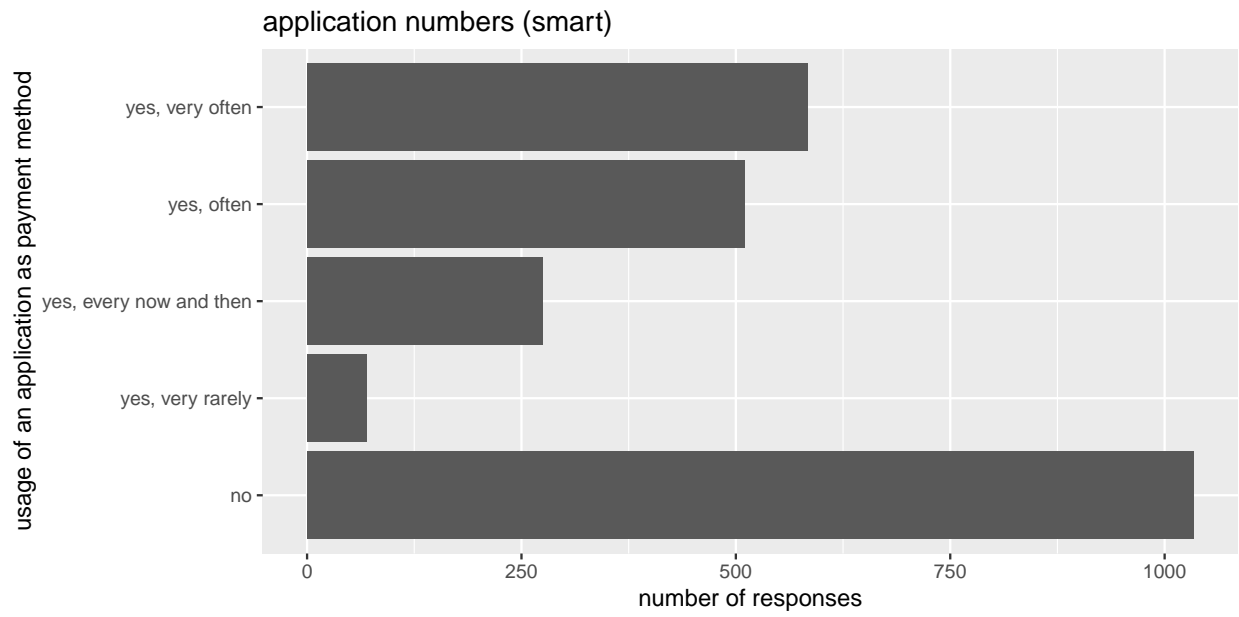
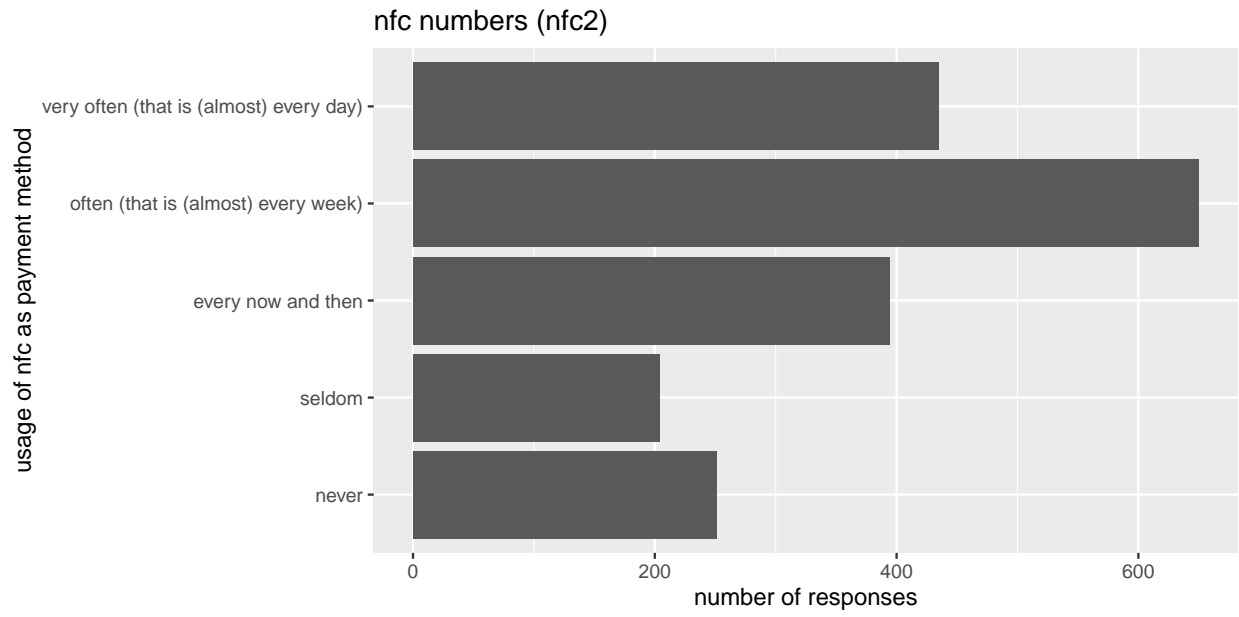
After I got the feedback on my Thesis Proposal, I prioritized factors to work on. These were:

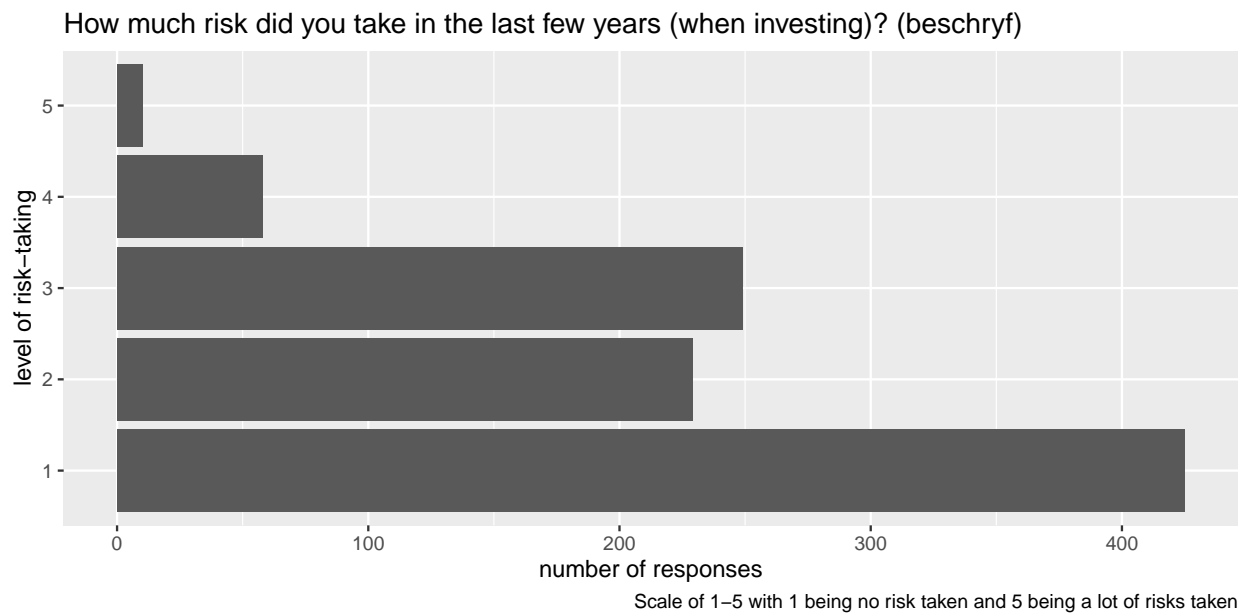
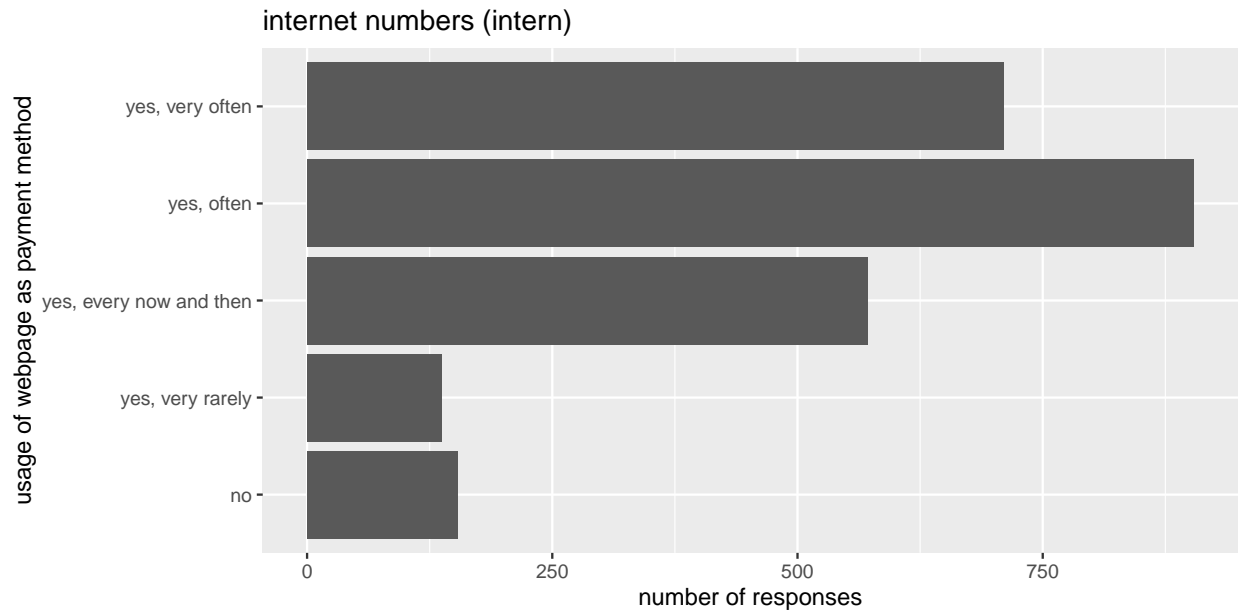
1. Flesh out the Method-section (find a better way to analyze the data, convey my findings, see what is applicable to certain hypotheses and what isn't, etc.)
2. Work down the list of hypotheses to see what kind of analysis they would need (can I use the same analysis method everywhere? Does the hypothesis still make sense looking at what I have already done?)
3. Read through papers related to my subject, and look at how they approached their Method-section (or equivalent). What can I learn from this?

Data exploration

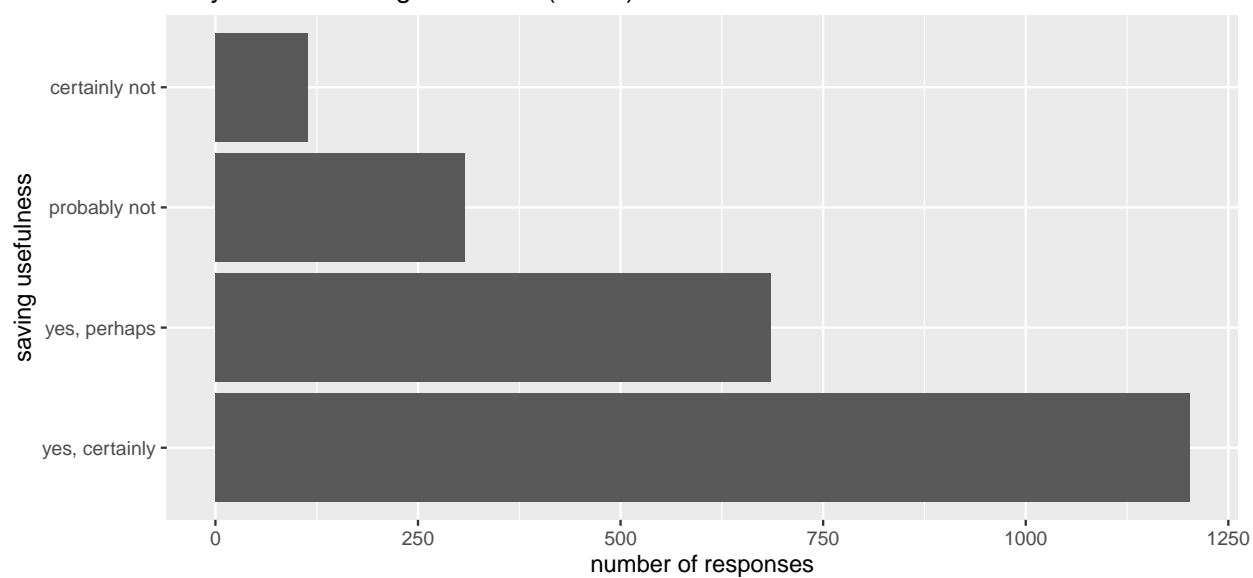
As per your email, I started with data exploration and visualization. I downloaded the data from the CentERdata website and used the survey that was most interesting for my first hypothesis: the PSY-module (containing questions on economic and psychological concepts). To get a clear view of the spread of the data, I made some histograms:



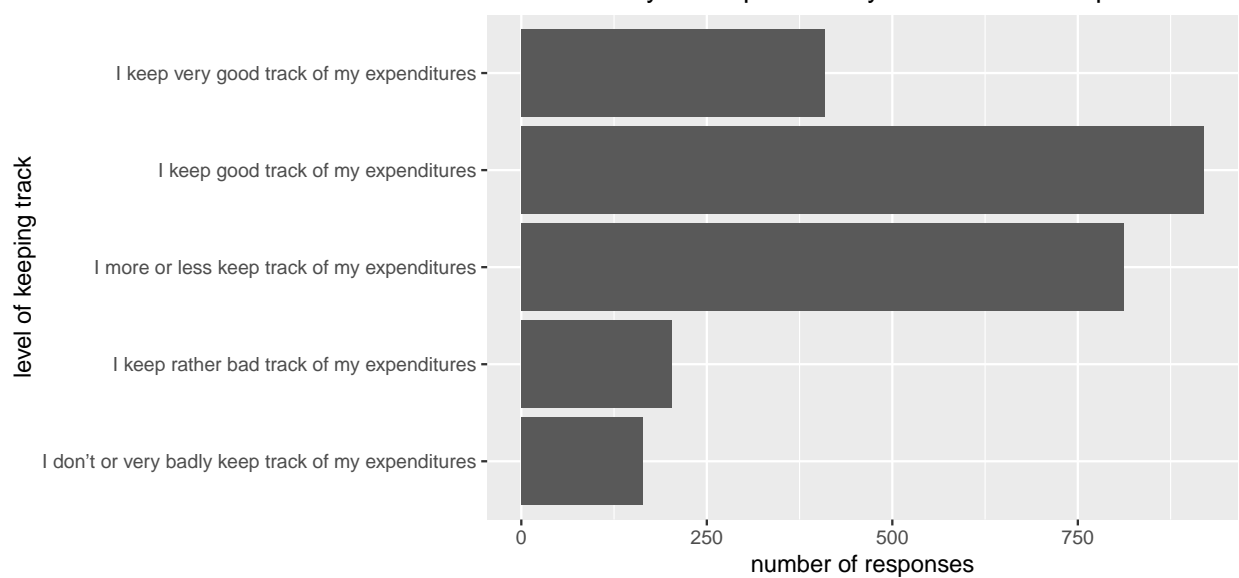


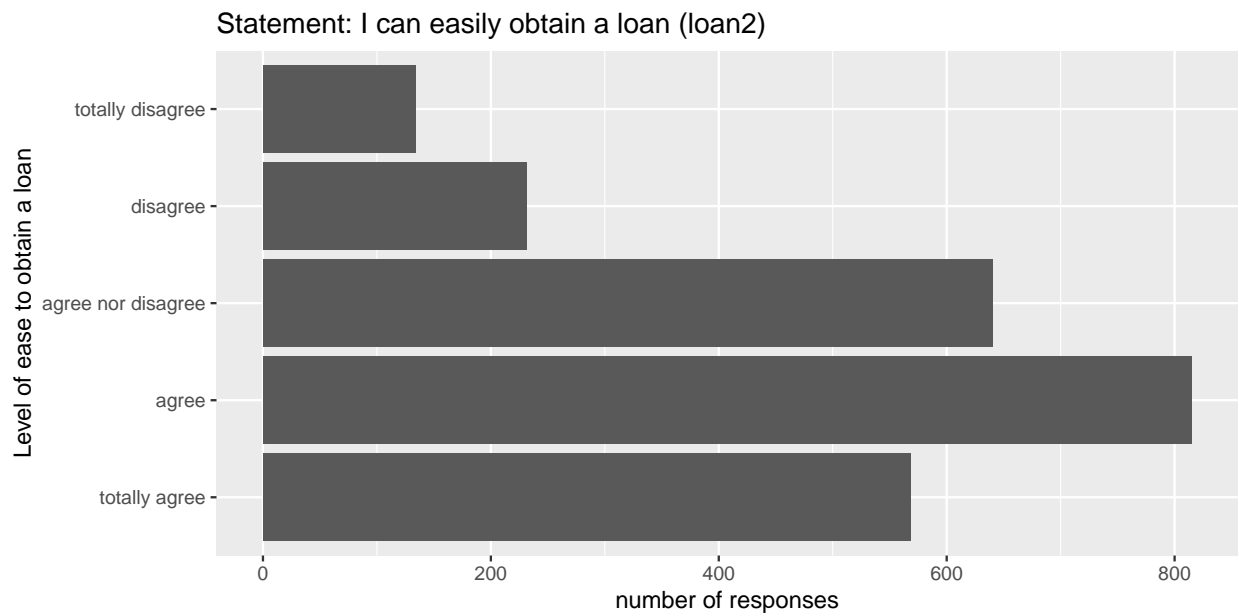
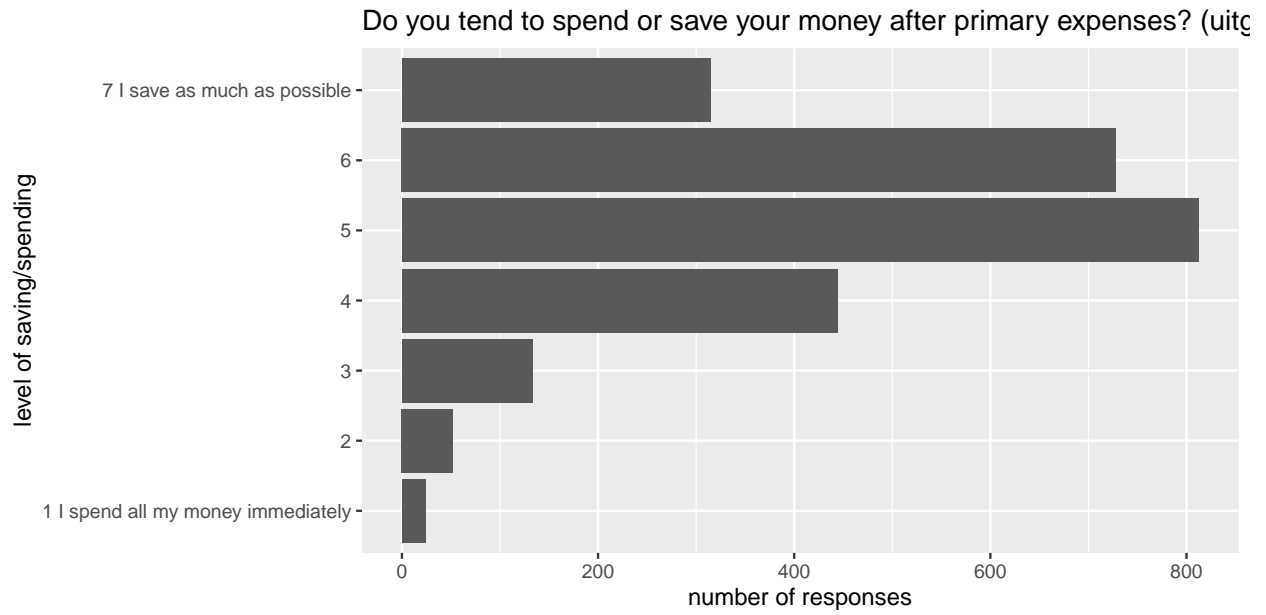


Do you think saving is useful? (zinvöl)

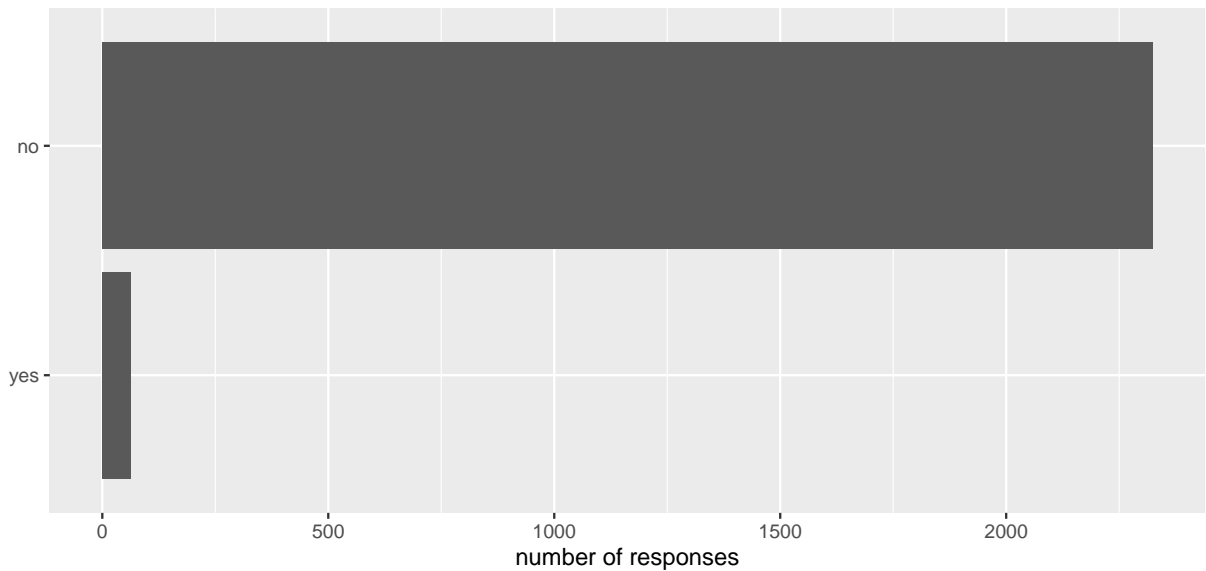


How well do you keep track of your household expenditures?

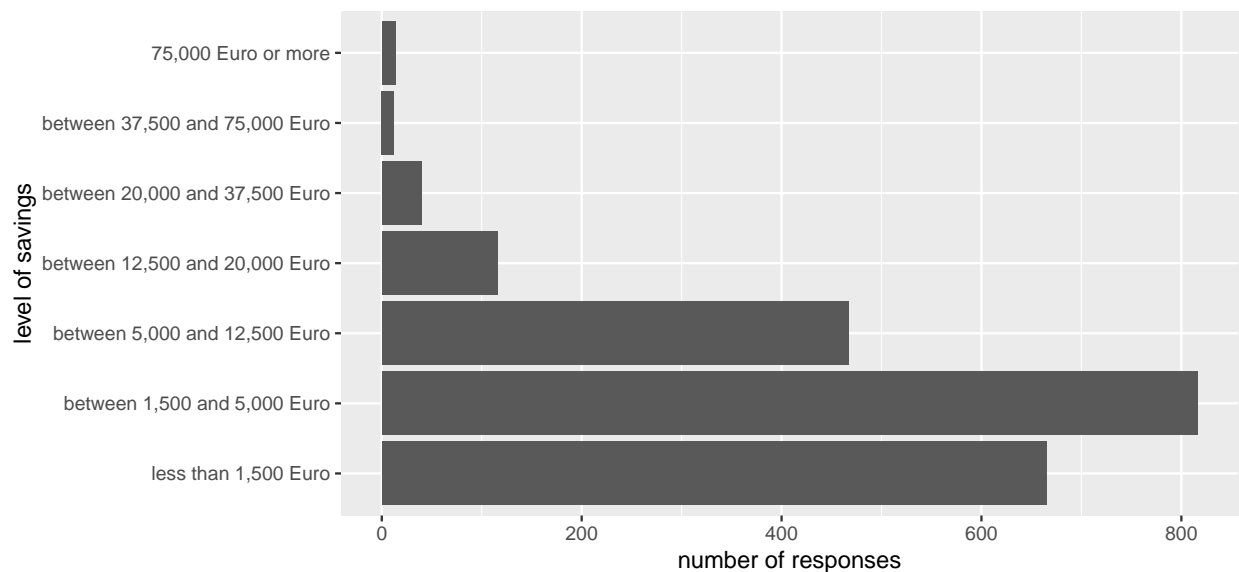




Have you ever had debt assistance? (loan3)



How much do you have saved? (hoevspa)



Looking through these histograms, I noticed some things that were relevant for further analysis:

1. Most respondents use PIN, NFC and internet payment methods often; applications fall behind on usage. This might be an issue for further analysis of applications.
2. The trend of internet and NFC users is comparable, allowing for good research on those.
3. The questions on saving and risk show that people tend to favor saving over spending and are risk averse, yet the reported amount of money they save seems low. This could suggest relatively low income, or having different types of savings (e.g. saving for leisure/holiday, but not for unexpected events).
4. Only a very low number of respondents has had debt assistance. This likely means that debt assistance cannot be used as a proxy of debt.

As I had already done some data exploration in my Proposal, I felt that I had progressed far enough that I could start looking for promising analysis methods.

Data analysis

With the previously mentioned points in mind I started looking for ways to get a more in-depth analysis than (only) linear regression. First I tried to implement a ridge/lasso design to try and find relevant variables in the full dataset (so all 6 questionnaires in the DHS). I spent some time trying to get this to work, however I ran into some errors I could not seem to fix. So I moved on to my next method: causal trees. I wanted to test this method to see if I could determine causality in my model. While I tried to fix errors I got with this model I came across another model that seemed like a better fit, and thus switched my effort to that model. However I could try and come back to this in the future if I want to try and prove causality.

The model I switched to was ordinal logistic regression. This seemed promising to me because most of the data I work with in the DHS is ordinal. Numerical output of this regression method is difficult to interpret, so I use graphs to make it easier. The graphs show the likelihood of being included in each level of the dependent variable, for each level of the independent variable. As an example look at the `zinvol`-plot for using PIN. The probability of being included in the classification of “very often” (thus using the PIN payment method very often) rises substantially as respondents’ answers to the question corresponding to `zinvol` (“Do you think saving is useful?”) goes from “probably not” to “certainly not”. Interpreting this in a quicker way: people who do not value savings are very likely to use PIN as a payment method.

I ran these regressions on each variable shown in the histograms earlier, together with a Confusion-matrix. I hid the regression output as it would just clutter the document, but it is still visible in the code if you would want to take a look. The graphs for the “`beschryf`” variable did not provide interesting results as by far most people reported taking very few risks. I did still include them for completeness, however.

As a note for the graphs:

The “`beschryf`”-scale goes from 1(no risk taken) to 5(a lot of risks taken).

The “`bijhoud`”-scale levels:

1. I don’t or very badly keep track of my expenditures
2. I keep rather bad track of my expenditures
3. I more or less keep track of my expenditures
4. I keep good track of my expenditures
5. I keep very good track of my expenditures

The “`hoevspa`”-scale levels:

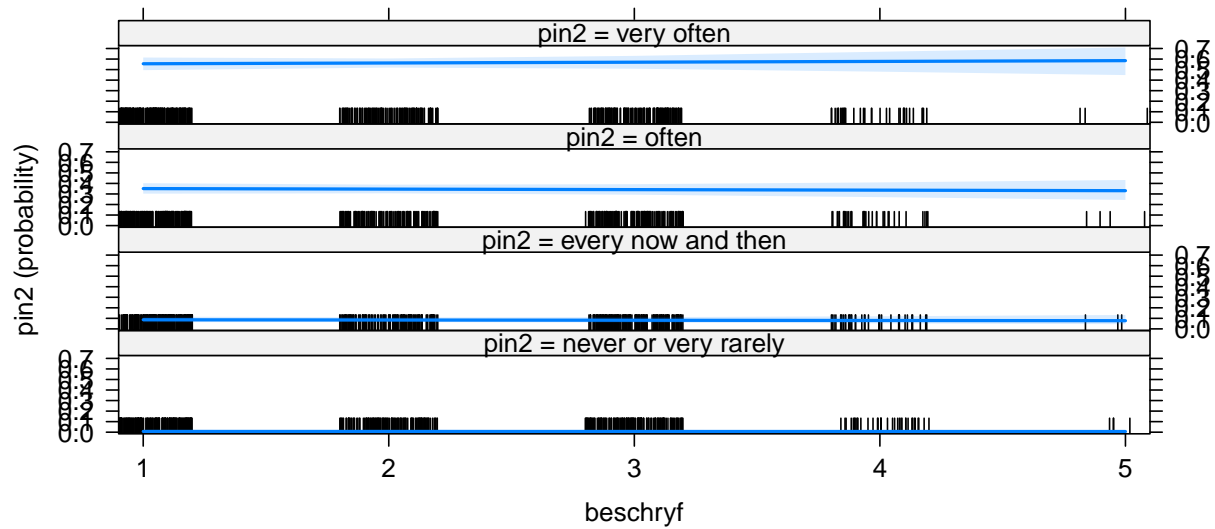
1. less than 1,500 Euro
2. between 1,500 and 5,000 Euro
3. between 5,000 and 12,500 Euro
4. between 12,500 and 20,000 Euro
5. between 20,000 and 37,500 Euro
6. between 37,500 and 75,000 Euro
7. 75,000 Euro or more

I will fix these layouts in the future.

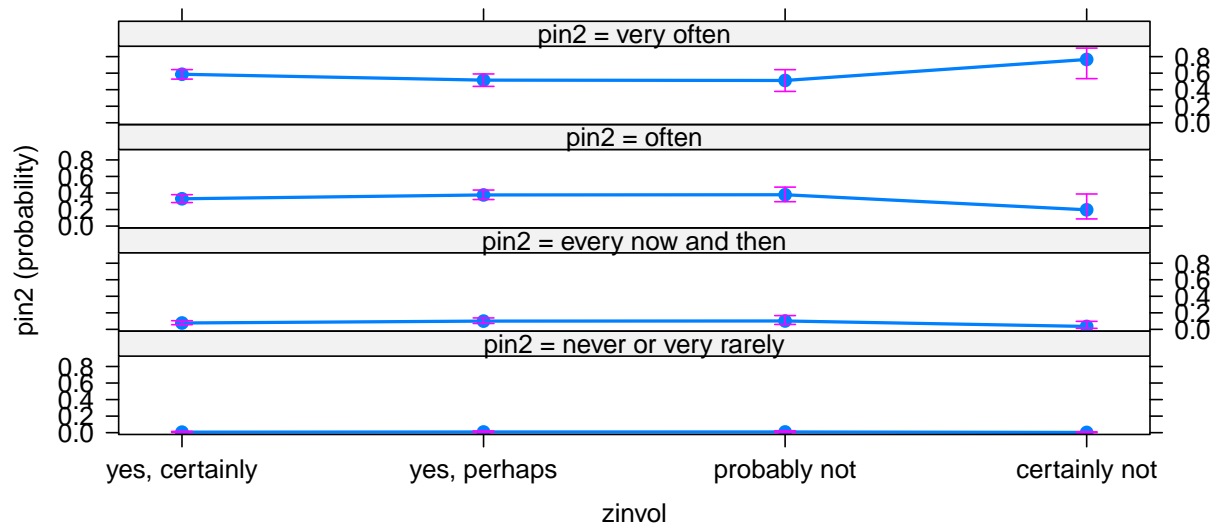
```
##                                predictpin
##                                never or very rarely every now and then often very often
##  never or very rarely          0                    0      0          6
##  every now and then            0                    0     12         32
##  often                        0                    0     29         77
##  very often                   0                    0     48        158

## [1] NA
```

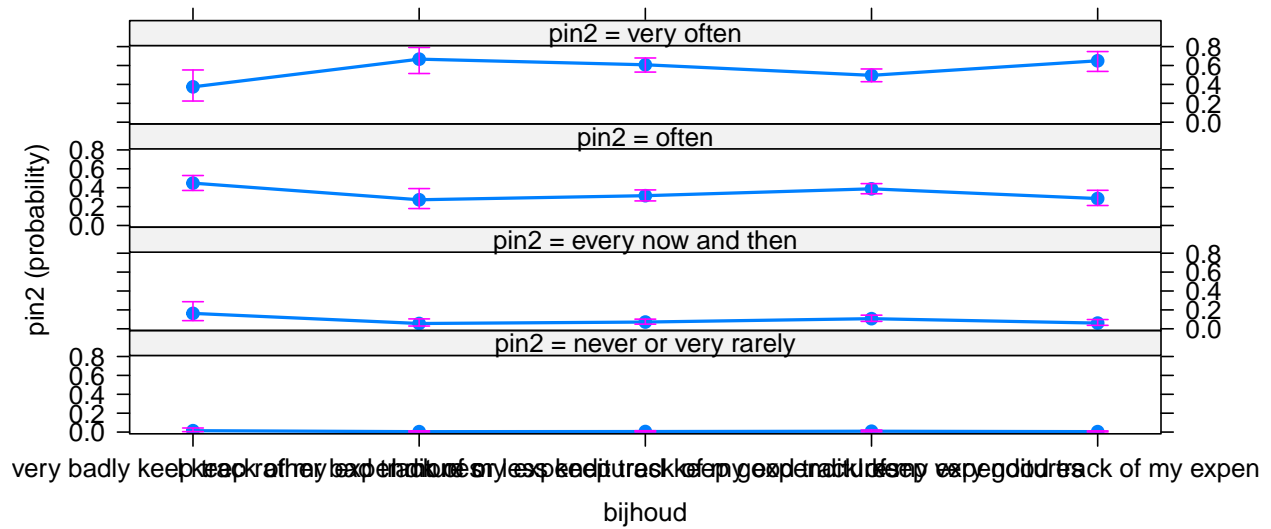
beschryf effect plot



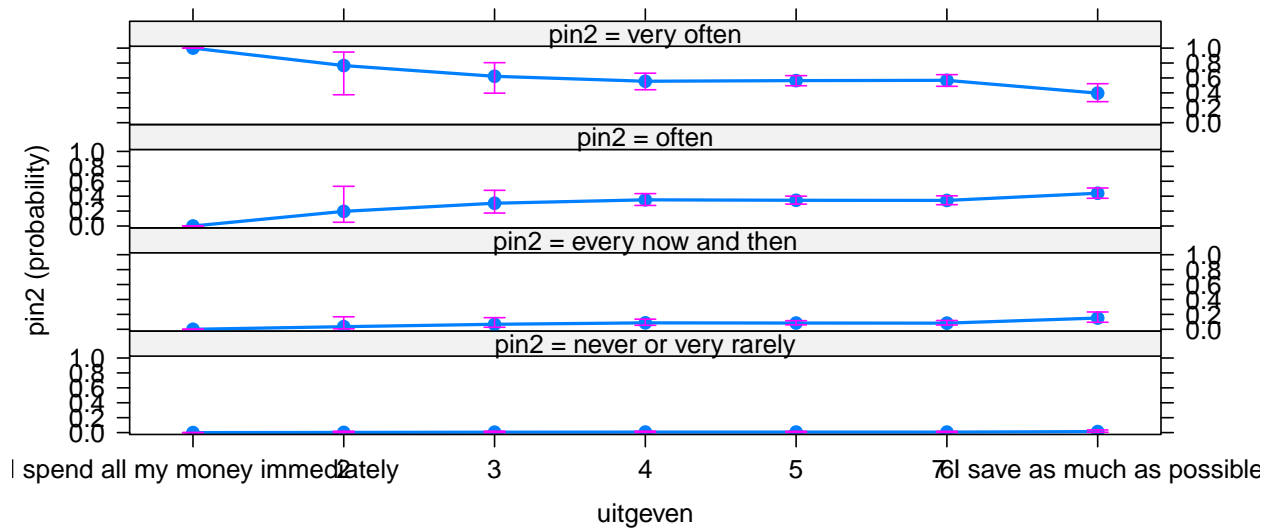
zinvol effect plot

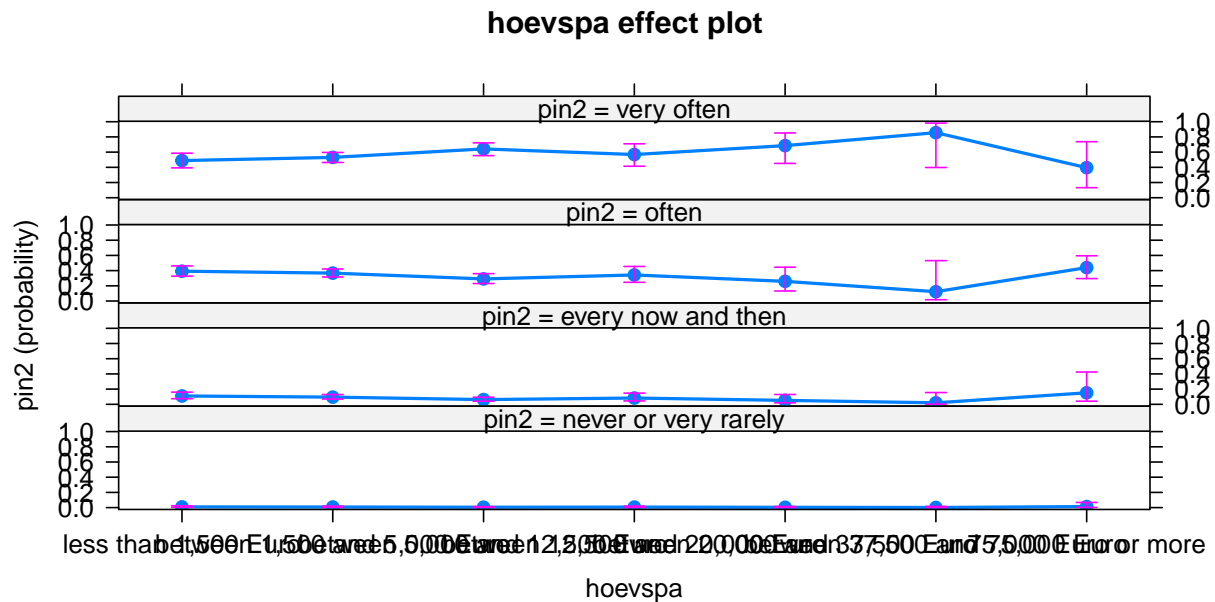
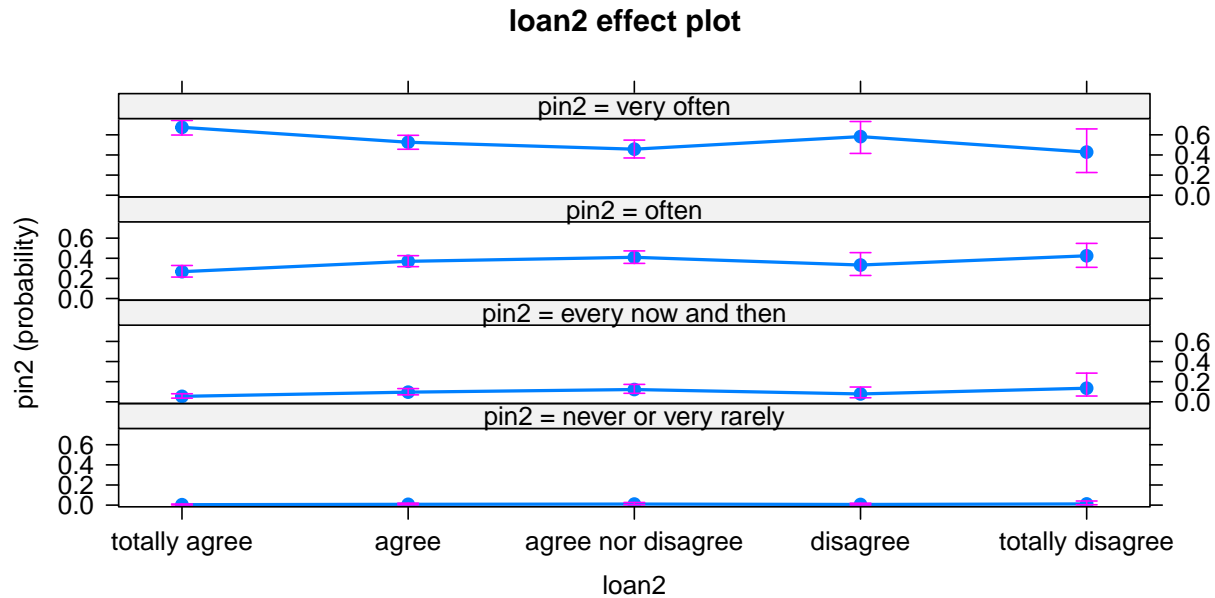


bijhoud effect plot



uitgeven effect plot





Looking at the graphs, there are some interesting things to note per dependent variable:

Pin

1. Zinvol graph

- “Often” and “very often” show opposite trends as the value for “zinvol” moves from “probably not” to “certainly not”. This could suggest that people who classify themselves as using Pin “often” are more moderate in their views (saving behaviour closer to the mean), whereas “very often” shows more deviance from this.

2. Bijhoud graph

- “Often” and “very often” show opposite trends here as well, for the values of “don’t or very badly/rather bad/more or less keep track of household expenses”.

3. Uitgeven graph

- The first three levels of this variable shows that a very high share (if not all) of the respondents who indicated they immediately spend their money also used pin as their payment method with the highest frequency. The other three frequency levels see the opposite effect, suggesting that the moderation effect we saw earlier seems to be present here as well.
- A small decline in the “very often” category for people who indicate they save as much as possible.

4. Loan2 graph

- Opposite overall trends for “very often” and “often” categories.
- High probability for people who indicate either “totally agree” or “disagree” to have the highest frequency of paying with pin. Seems to be a contradictory result, yet could also be dependent on financial knowledge of the respondent. Another explanation could be that the “very often” group is rather divided.

5. Hoevspa graph

- Total opposite effect for “very often” and “often” categories at the “37,500-75.000 saved” level. This might indicate that some of the “very often”-respondents are wealthier, thus having more savings. The “often”-respondents could be more average households which cannot reach this level of savings (yet has some outliers at the 75.000+ level).

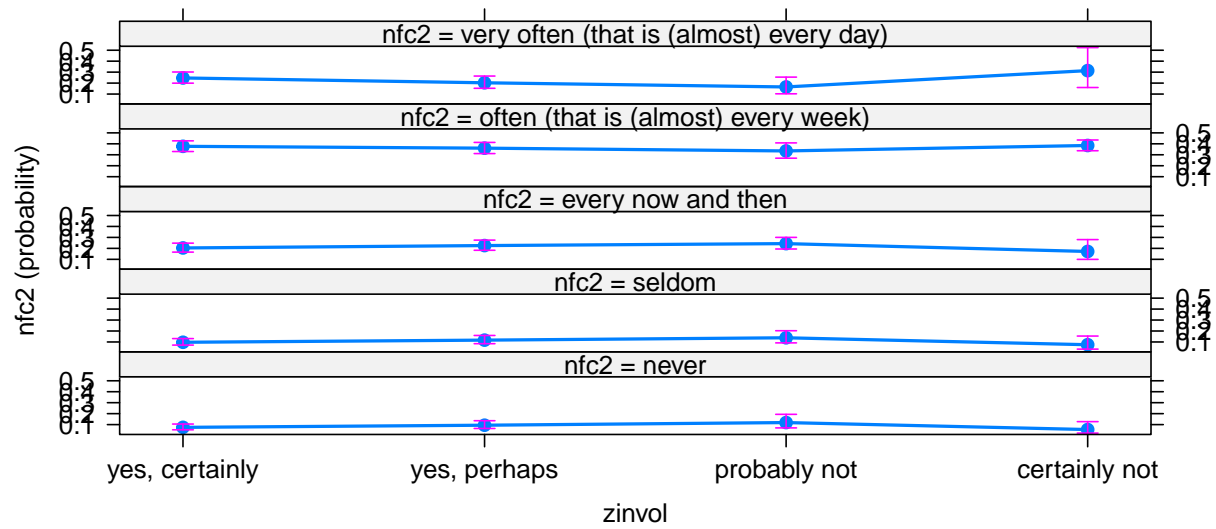
```
##                                predictnfc
##                                never seldom every now and then
##  never                        1      0                        6
##  seldom                       1      0                        4
##  every now and then           2      0                        1
##  often (that is (almost) every week)  4      0                        4
##  very often (that is (almost) every day) 1      0                        2
##                                predictnfc
##                                often (that is (almost) every week)
##  never                        27
##  seldom                       18
##  every now and then           39
##  often (that is (almost) every week) 84
##  very often (that is (almost) every day) 53
##                                predictnfc
##                                very often (that is (almost) every day)
##  never                        2
##  seldom                       3
##  every now and then           6
##  often (that is (almost) every week) 14
##  very often (that is (almost) every day) 7

## [1] NA
```

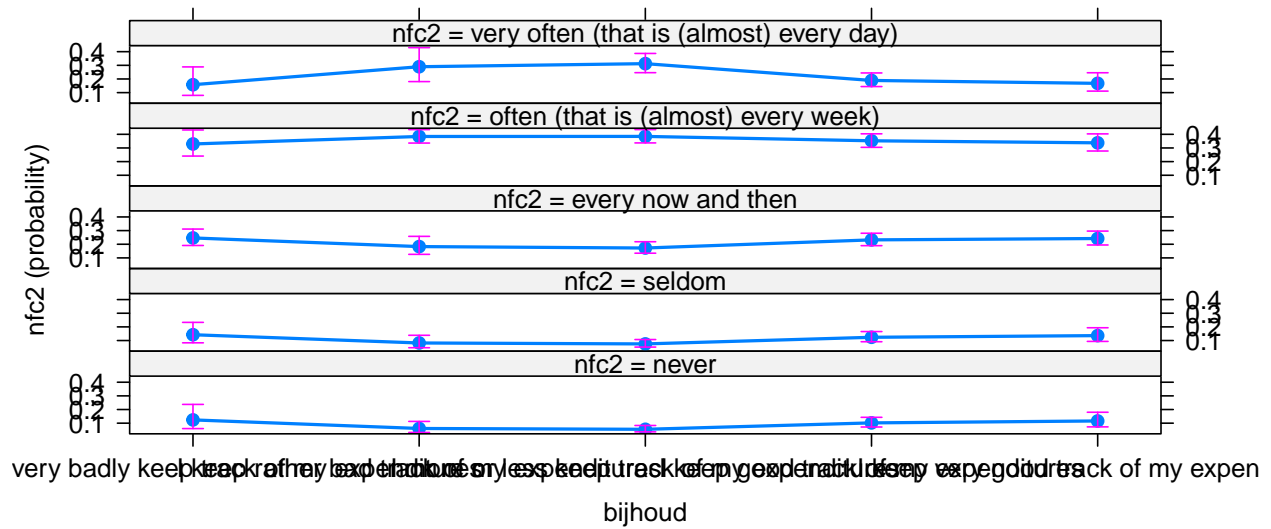
beschryf effect plot



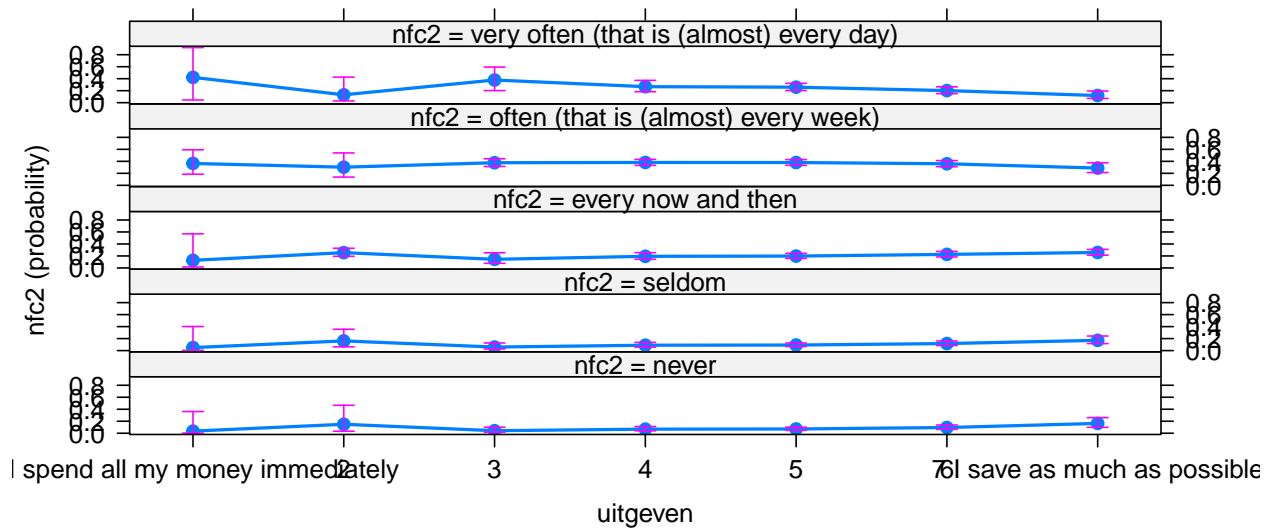
zinvol effect plot

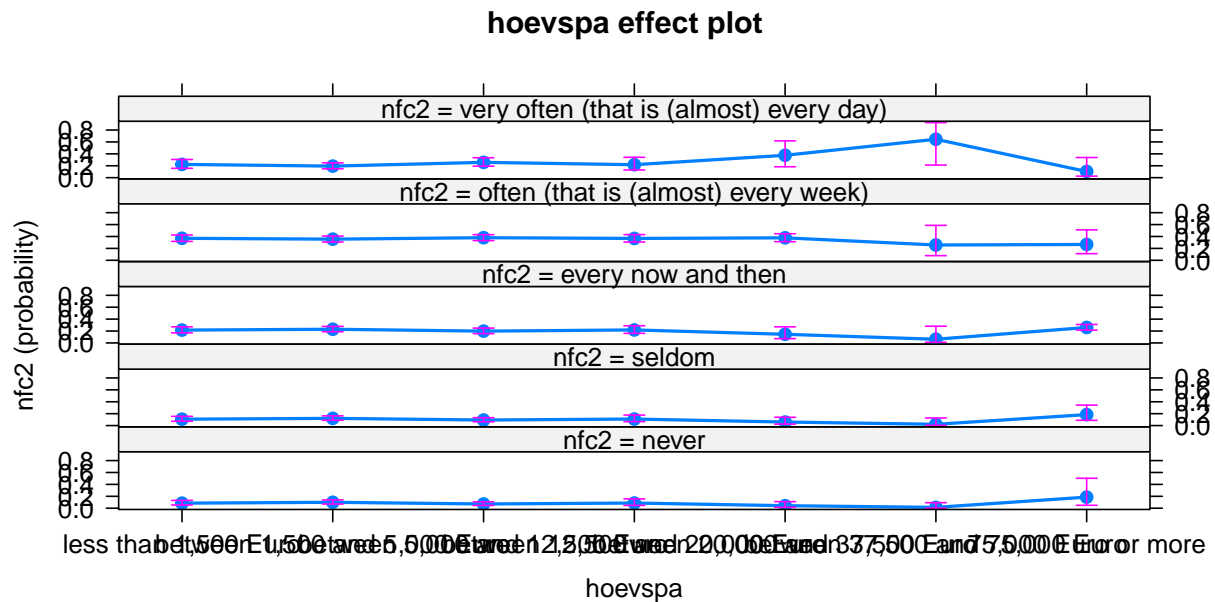
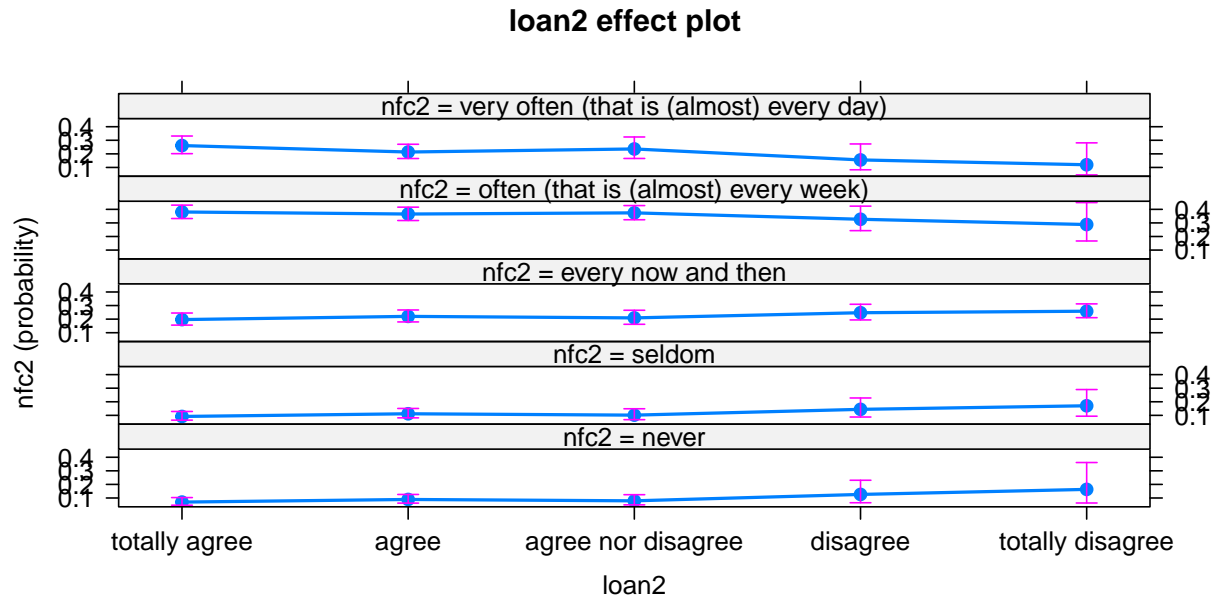


bijhoud effect plot



uitgeven effect plot





NFC

1. Zinvol graph

- Shows very little variation at each level, suggesting that the decision to use NFC is not impacted by people's views on saving.

2. Bijhoud graph

- The trend of the highest frequency of NFC usage suggests that NFC is used less by people who tend to keep better track of their expenditures.

3. Uitgeven graph

- People who say they never use NFC have a surprisingly high representation in the group of people who spend their money immediately.
- Stays consistent in the middle, most of the trendchanges happen at the tails.

4. Loan2 graph

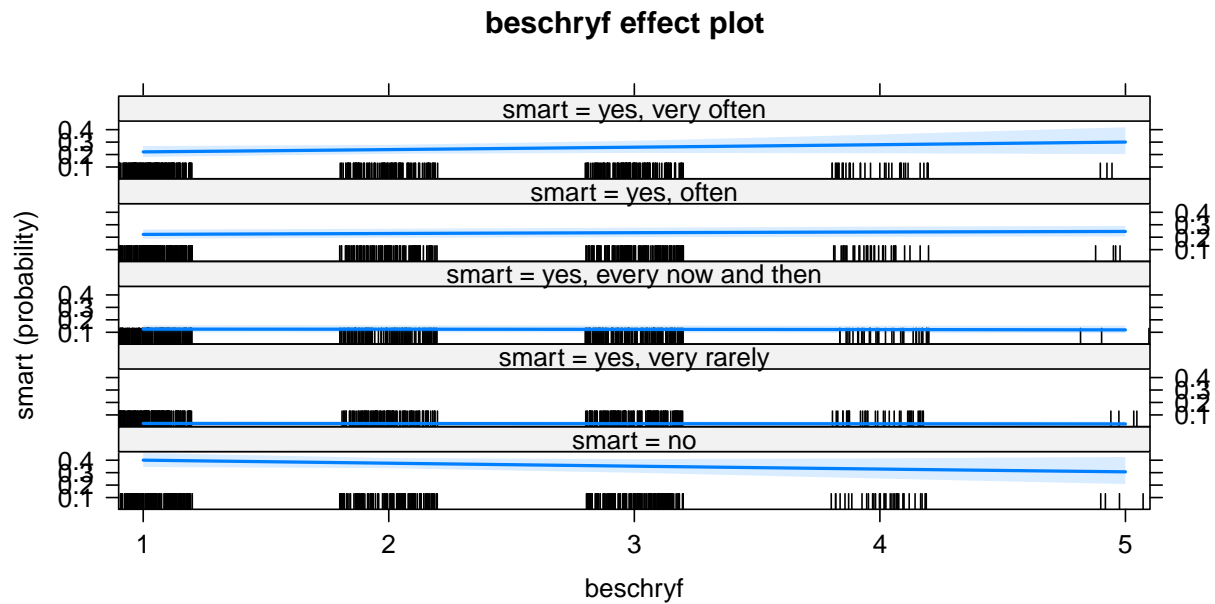
- Similar to the “Uitgeven” graph, but mirrored. People who use NFC at a high frequency think they are easily able to get a loan.

5. Hoevspa graph

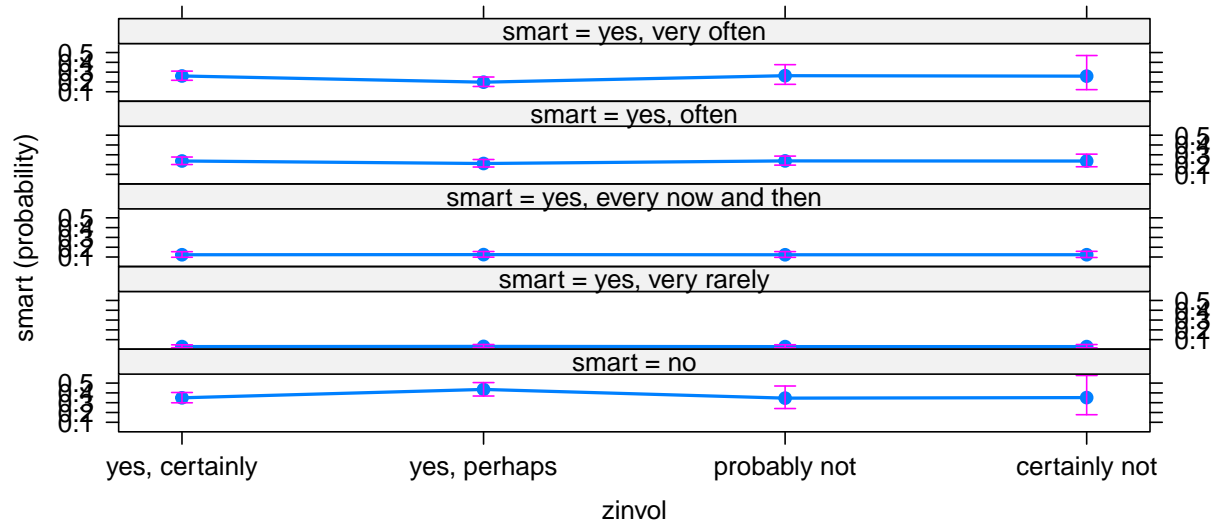
- Same effect here as seen with Pin. A trend/commonality or a mistake in the data?

```
##                                predictsmart
##                                no yes, very rarely yes, every now and then
## no                                105                                0                                0
## yes, very rarely                    4                                0                                0
## yes, every now and then            35                                0                                0
## yes, often                          48                                0                                0
## yes, very often                     60                                0                                0
##                                predictsmart
##                                yes, often yes, very often
## no                                0                                48
## yes, very rarely                  0                                5
## yes, every now and then           0                                9
## yes, often                        0                                21
## yes, very often                   0                                28

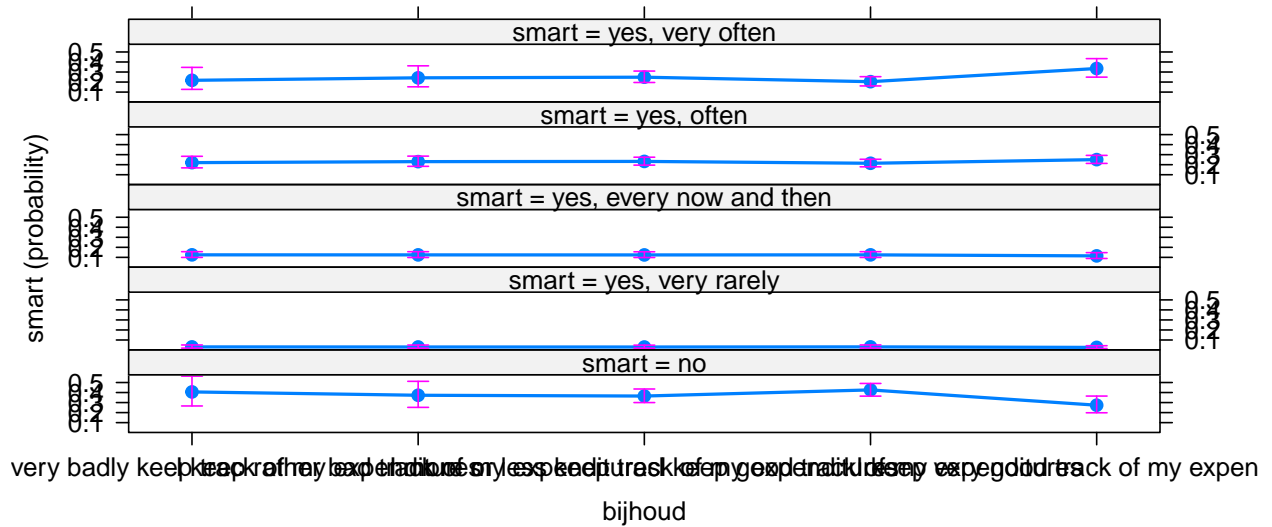
## [1] NA
```



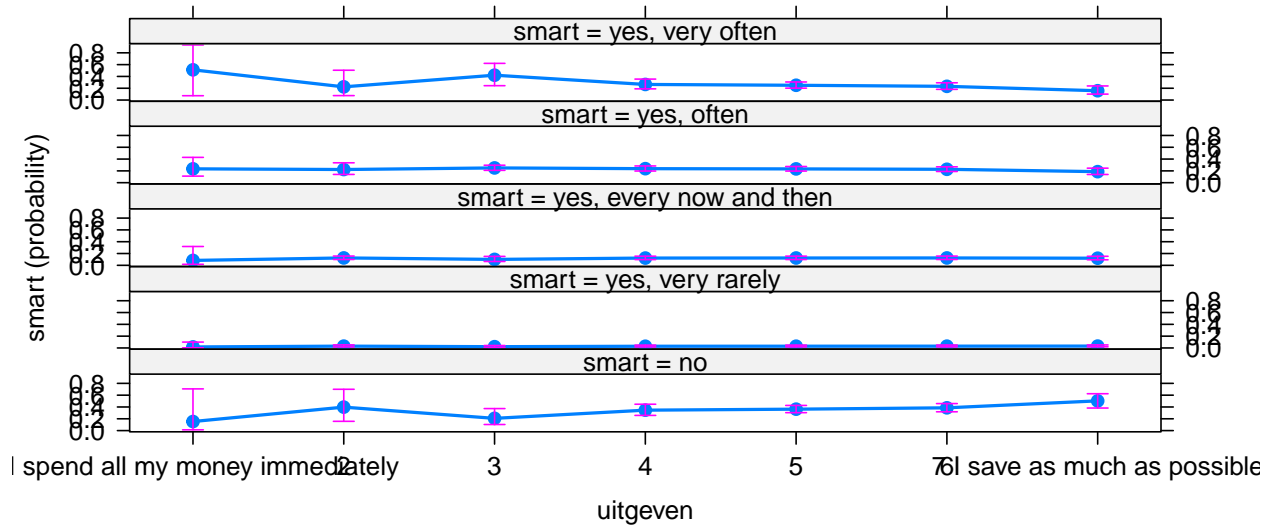
zinvol effect plot



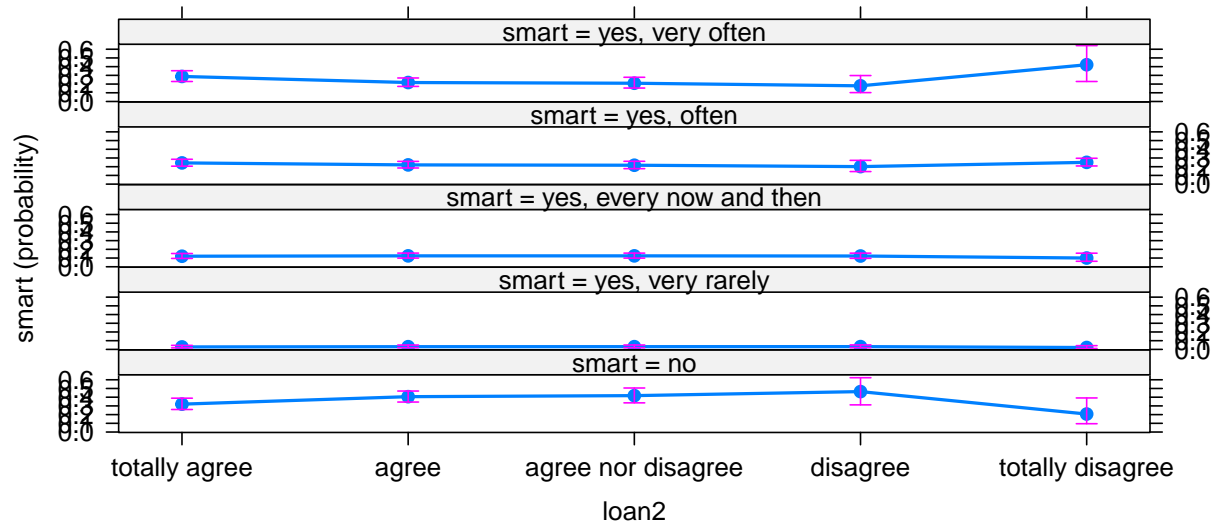
bijhoud effect plot

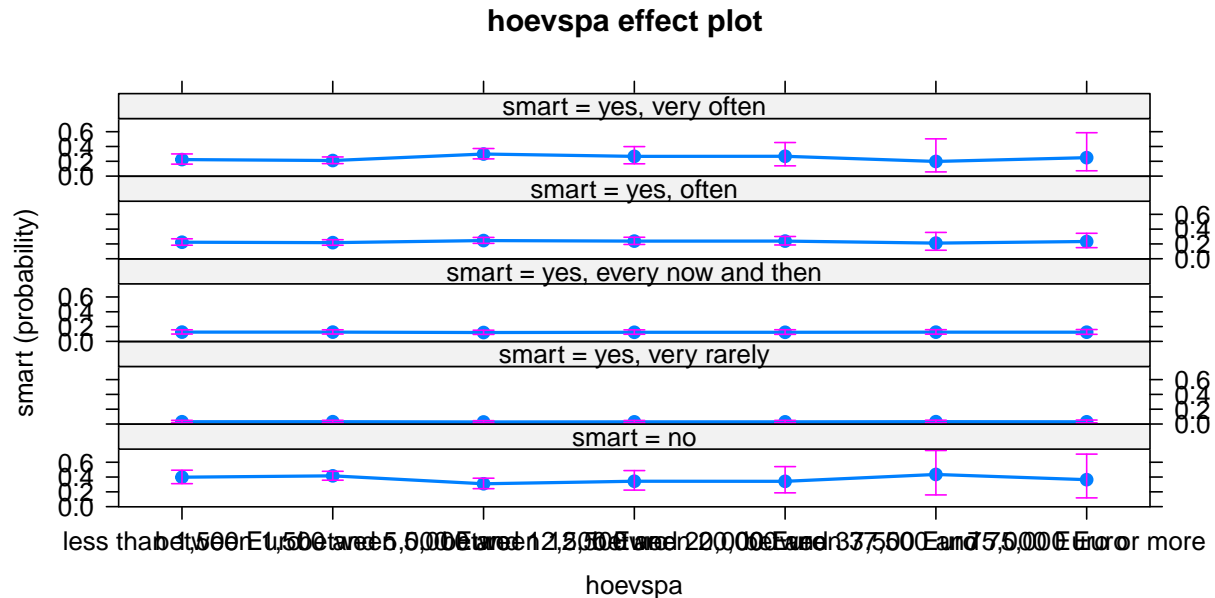


uitgeven effect plot



loan2 effect plot





Smart (application)

1. Zinvol graph

- Much less variation than with the previous two payment methods. Though the “no” group is bigger here, which allows us to look at the differences between users and non-users better.

2. Bijhoud graph

- The difference between users and non-users shows up here already. People reporting they keep moderate/good track of their expenses are more likely to *not* use applications as payment method. Though interestingly the group that is the best at keeping track of their expenditures has a roughly equal chance to use or not use applications. This could aim at potential nullification of the transparency effect when people are very aware of their expenses.

3. Uitgeven graph

- A small fluctuation around the quick spenders. This effect size seems too small to attribute anything to it.

4. Loan2 graph

- Little information can be taken from this graph. The increase in “totally disagree” is likely not correlated with people’s use of payment applications, since it does not inform us of the payment method(s) these people *do* use.

5. Hoevspa graph

- An interesting change in this graph from the previous two payment methods. As people report they have more savings, they are less likely to use applications as payment method. This could point to an age effect where younger people (generally low savings) are more technologically advanced, and thus use this method more. Older people (generally high(er) savings) are less familiar with this payment methods, reducing their usage frequency.

##		predict	intern
##		no	yes, very rarely yes, every now and then
##	no	0	0 2
##	yes, very rarely	0	0 1

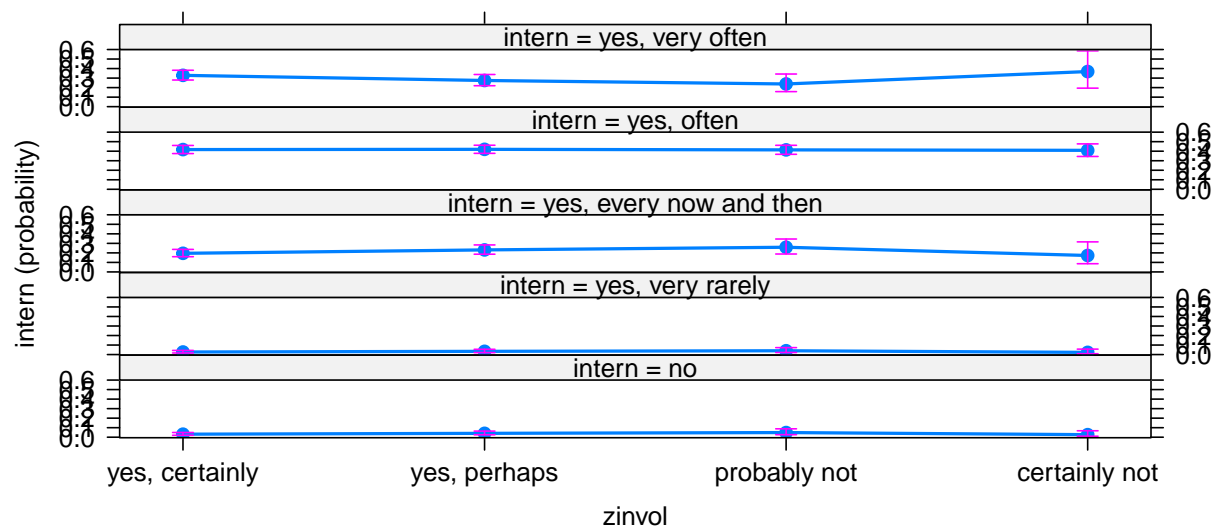
```
## yes, every now and then 0 0 3
## yes, often 0 0 11
## yes, very often 0 0 2
##
## predictintern
## yes, often yes, very often
## no 3 2
## yes, very rarely 14 4
## yes, every now and then 62 10
## yes, often 103 19
## yes, very often 94 33

## [1] NA
```

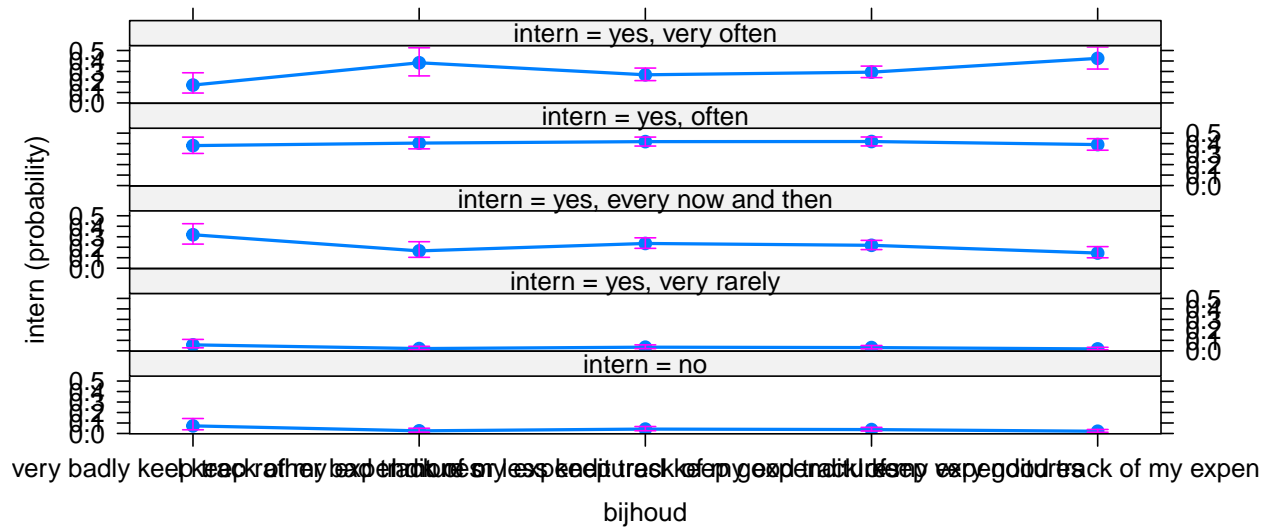
beschryf effect plot



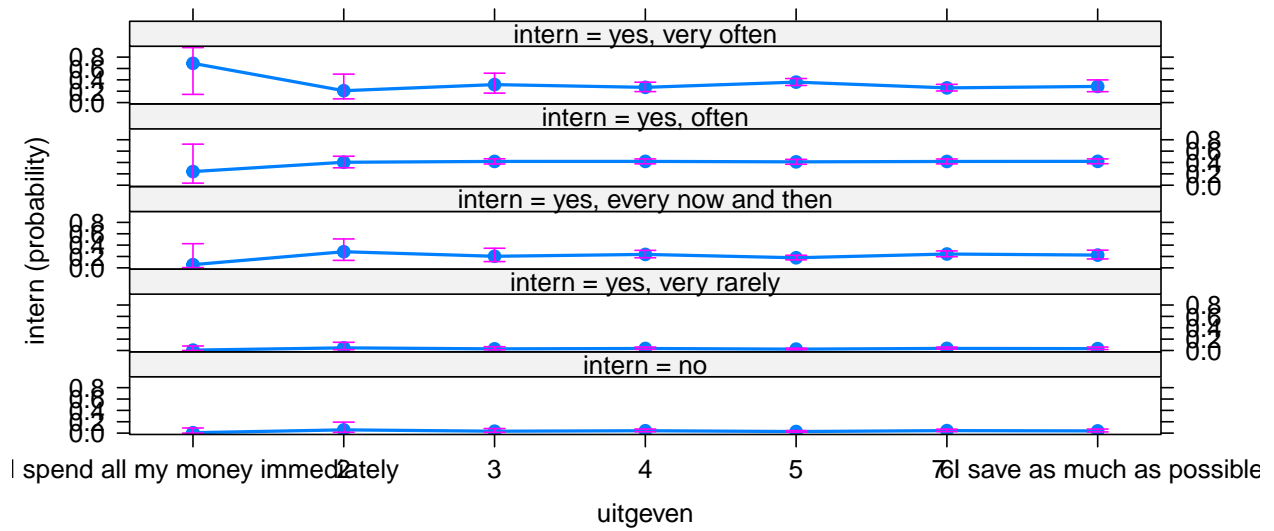
zinvol effect plot

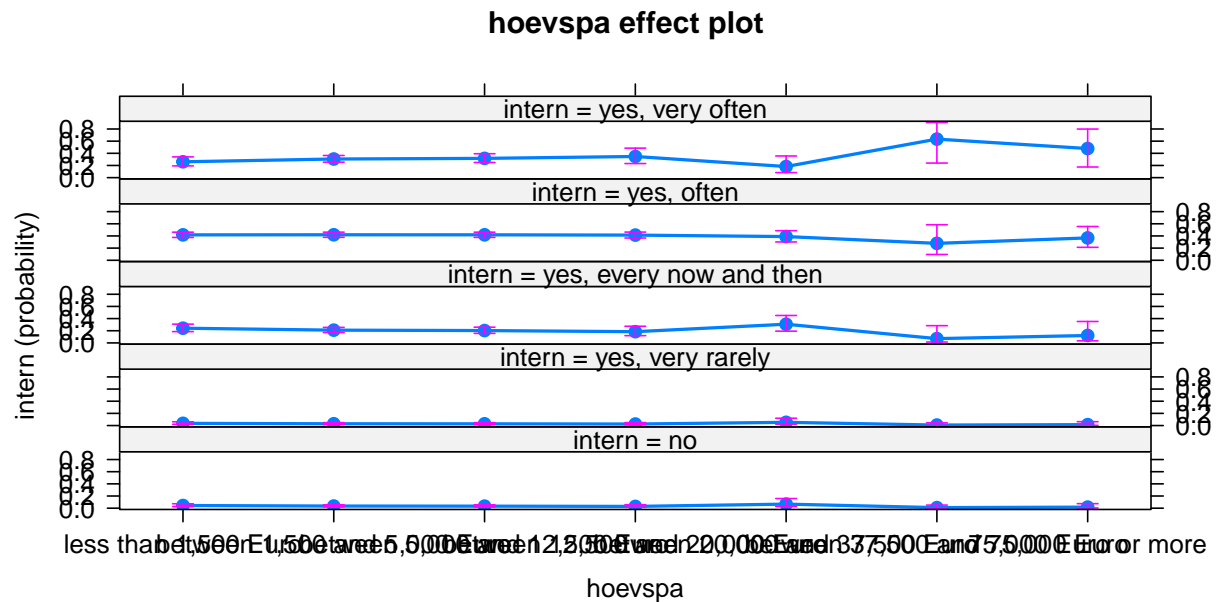
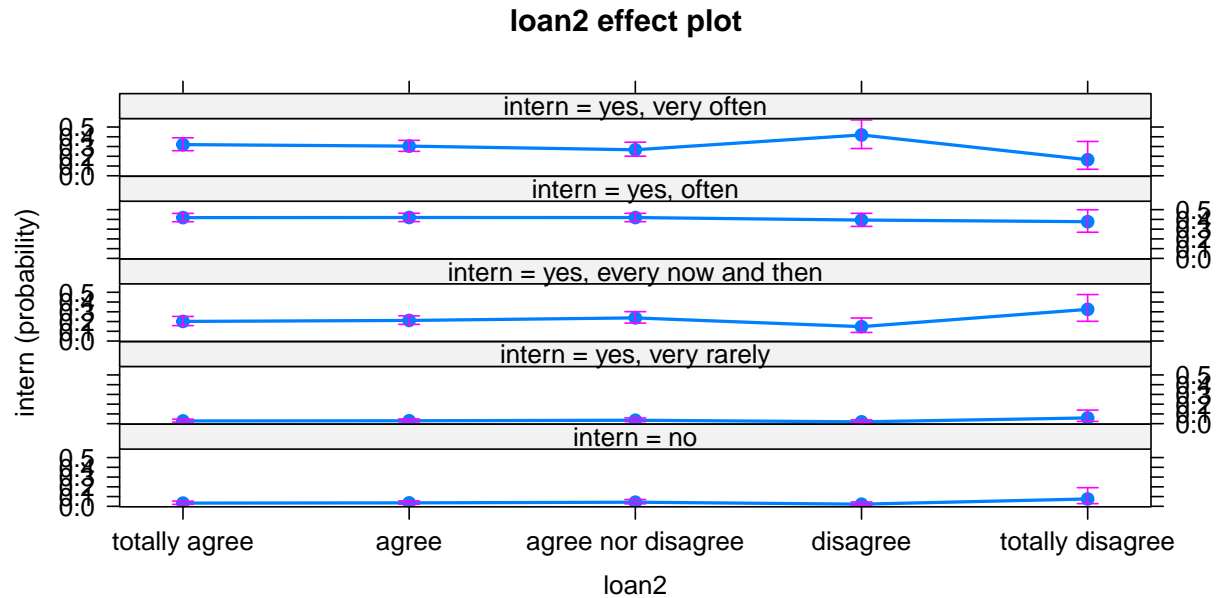


bijhoud effect plot



uitgeven effect plot





Internet

1. Zinvol graph

- Seems very similar to the same plot under Pin.

2. Bijhoud graph

- Interestingly, people who keep track of their expenditures very well are very likely to use the internet to pay their bills. This could be due to the fact that using your bank's online environment allows you to both pay your bills and have a detailed overview of your incoming and outgoing cash.
- People who say they use this method "every now and then" seem to be less likely to keep track of their expenditures, suggesting that checking your online account regularly is one of the key ways of keeping your expenses in check.

3. Uitgeven graph

- This graph seems odd, given the Bijhoud graph results. People's behaviours are evenly spread at every level, except for the group that says they spend their money immediately. It could be that this is a group with sizable savings, thus allowing them to spend more. More research is needed on this part.

4. Loan2 graph

- The difference between "very often" and "every now and then" is distinct, but seems too small to have any weight. Especially since the "totally disagree" and "agree nor disagree" levels are comparable, this seems to be an oddity.

5. Hoevspa graph

- People who use the internet quite often to pay their bills seem to have higher savings. This seems to correspond with the conclusions from the Bijhoud graph.

As can be seen in the confusion-matrices, the models perform poorly in areas where there are few observations. Often it does not predict any of these cases correctly. However, the cases with a higher number of observations seems to be predicted a lot better. Therefore I think these regressions can be used reliably to draw conclusions from the groups with relatively high observation counts.

Overall I feel like these graphs show promise in explaining some of the effects at work. I want to work this out further by looking at more specific parts of the data (e.g. for the question I raised in the Uitgeven graph when paying via internet is the dependent variable).

Current research plans

The things I will be working on in the near future:

1. Exploring whether ordinal logistic regression is an appropriate approach for the analysis needed in the other hypotheses.
2. Finding papers that research (parts of) topics related to my hypotheses and start documenting theoretical similarities & differences (to get an overview of my literature review).
3. Collecting papers that have research methods comparable with mine and see what I can learn from them. This was one of my goals to start doing directly after the Proposal feedback, however I got so caught up with finding more advanced analysis methods that I have not made much progress.

Due to the extra research needed to find appropriate analysis tools I have fallen a bit behind the schedule I had set up for my Proposal. However I believe that I can get back on track within the next week, as the ordinal logistic regression results seem promising.