Assignment 2 - Ximena Moure, Omar Lopez

Statistical Inference and Modelling

1 Loading and preprocessing of the data

1.1 Load libraries and clear workspace

```
# Clear plots
if(!is.null(dev.list())) dev.off()
```

```
## null device
##           1
```

```
# Clean workspace
rm(list=ls())
options(contrasts=c("contr.treatment","contr.treatment"))

library(tidymodels)
```

```
## -- Attaching packages -------------------------------------- tidymodels 1.0.0 --

## v broom        1.0.2      v recipes      1.0.3
## v dials        1.1.0      v rsample      1.1.1
## v dplyr        1.0.10     v tibble       3.1.8
## v ggplot2      3.3.6      v tidyr        1.2.1
## v infer        1.0.4      v tune         1.0.1
## v modeldata    1.0.1      v workflows    1.1.2
## v parsnip      1.0.3      v workflowsets 1.0.0
## v purrr        0.3.4      v yardstick    1.1.0

## -- Conflicts ----------------------------------------- tidymodels_conflicts() --
## x purrr::discard() masks scales::discard()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x recipes::step()  masks stats::step()
## * Use tidymodels_prefer() to resolve common conflicts.
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v readr   2.1.1     v forcats 0.5.1
## v stringr 1.4.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x readr::col_factor() masks scales::col_factor()
## x purrr::discard()    masks scales::discard()
## x dplyr::filter()     masks stats::filter()
## x stringr::fixed()    masks recipes::fixed()
## x dplyr::lag()        masks stats::lag()
## x readr::spec()       masks yardstick::spec()
```

```
library(ggpubr)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(ROCR)
library(effects)
```

```
## Loading required package: carData
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(nnet)
```

1.2 Load data

We load the data and we take a look at the raw data to see if there are any unusual values.

The data set consists of five numerical variables, which are `egoposition_immigration`, `ostwest`, `political_interest`, `income` and `gender`,and one categorical variable, which is `vote`. Considering the statement of the project, we will transform all the numerical variables, as they are in fact qualitative. Variables `egoposition_immigration`, `income` , and `political_interest` are ordered factors, while `ostwest` and `gender` are binary.

```
# Load data
df <- MNLpred::gles

# We look at the data to see if there are any unusual values
summary(df)
```

```
##      vote           egoposition_immigration    ostwest      political_interest
##  Length:1000        Min.   : 0.000          Min.   :0.000   Min.   :0.000
##  Class :character   1st Qu.: 3.000          1st Qu.:1.000   1st Qu.:2.000
##  Mode  :character   Median : 4.000          Median :1.000   Median :3.000
##                     Mean   : 4.361          Mean   :0.759   Mean   :2.874
##                     3rd Qu.: 6.000          3rd Qu.:1.000   3rd Qu.:4.000
##                     Max.   :10.000          Max.   :1.000   Max.   :4.000
##      income          gender
##  Min.   :0.000   Min.   :0.000
##  1st Qu.:3.000   1st Qu.:0.000
##  Median :3.000   Median :0.000
##  Mean   :2.906   Mean   :0.462
##  3rd Qu.:3.000   3rd Qu.:1.000
##  Max.   :4.000   Max.   :1.000
```

```
df %>% head(10)
```

```
## # A tibble: 10 x 6
##    vote    egoposition_immigration ostwest political_interest income gender
##    <chr>                     <dbl>   <dbl>              <dbl>  <dbl>  <dbl>
##  1 FDP                           4       1                  3      3      0
##  2 SPD                           8       0                  2      2      1
##  3 CDU/CSU                       3       1                  1      3      1
##  4 CDU/CSU                       7       1                  2      3      0
##  5 SPD                           2       1                  3      3      1
##  6 CDU/CSU                       4       0                  2      3      0
##  7 Gruene                        4       1                  3      3      0
##  8 Gruene                        1       1                  2      4      1
##  9 Gruene                        2       1                  4      4      1
## 10 FDP                           5       1                  3      3      0
```

```
# Look at data types
typeof(df$vote)
```

```
## [1] "character"
```

```
typeof(df$egoposition_immigration)
```

```
## [1] "double"
```

```
typeof(df$ostwest)
```

```
## [1] "double"
```

```
typeof(df$political_interest)
```

```
## [1] "double"
```

```r
typeof(df$income)
```

```
## [1] "double"
```

```r
typeof(df$gender)
```

```
## [1] "double"
```

```r
sapply(df, class)
```

```
##                    vote egoposition_immigration                 ostwest
##             "character"               "numeric"               "numeric"
##       political_interest                  income                  gender
##               "numeric"               "numeric"               "numeric"
```

1.3 Check missing data, duplicates and misspellings

In this section, we check for missing data and duplicates. There are no NA present in the data set. When looking at the duplicates, we can see that there are 359 duplicates. Considering that people with the same profile/characteristics can vote for the same party, we are not going to remove the duplicates. There are no misspellings.

```r
# Check missing data
cbind(lapply(lapply(df, is.na), sum))
```

```
##                         [,1]
## vote                    0
## egoposition_immigration 0
## ostwest                 0
## political_interest      0
## income                  0
## gender                  0
```

```r
# Check for blanks
which(df=="") # no blanks
```

```
## integer(0)
```

```r
# check for duplicates
sum(duplicated(df))
```

```
## [1] 359
```

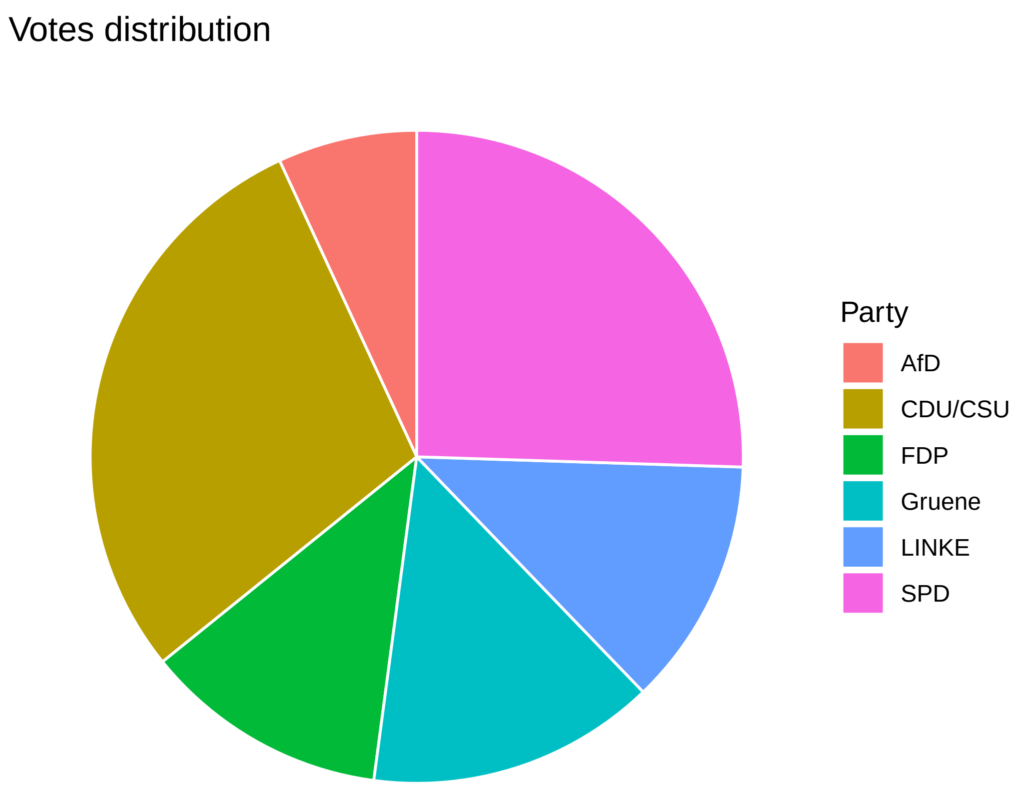
```r
# Check misspelling
df %>%
  mutate(vote = as_factor(vote)) %>%
  count(vote)
```

```
## # A tibble: 6 x 2
##   vote        n
##   <fct>   <int>
## 1 FDP       121
## 2 SPD       255
## 3 CDU/CSU   289
## 4 Gruene    143
## 5 AfD        69
## 6 LINKE     123
```

1.4 Data balancing

From the plot below we can see that the dataset is unbalanced on its target variable. This will be taken into account and dealt with when doing the modeling part.

```r
ggplot(data.frame(table(df$vote)), aes(x="", y=Freq, fill=Var1)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar("y", start=0) +
  ggtitle("Votes distribution") +
   guides(fill = guide_legend(title = "Party"))+
  theme_void()
```



1.5 Creating factors

Lets apply some preprocessing: - Convert the `vote` variable into a categorical variable - Other variables, like egoposition_immigration, political_interest and income are ordinal - Transform dummy variables like gender or ostwest.

For our analysis, let's see which political parties are considered left, right and center, and create a new variable to store it. Lets assume the following:

- Linke: left
- AfD: right
- Gruene: Left
- CDU/CSU: center
- SPD: center
- FDP: center

```
df_transformed <- df %>%
  mutate(vote = as_factor(vote),
         egoposition_immigration = factor(egoposition_immigration, ordered = T),
         political_interest = factor(political_interest, ordered = T),
         income = factor(income, ordered = T),
         ostwest = factor(ostwest, labels = c("West Germany", "East Germany")),
         gender = factor(gender, labels = c("Male", "Female")),
         )

df_transformed <- df_transformed %>%
  mutate(compass = ifelse(vote == "FDP", "center",
                   ifelse(vote == "SPD", "center",
                   ifelse(vote == "CDU/CSU", "center",
                   ifelse(vote == "Gruene", "left",
                   ifelse(vote == "LINKE", "left",
                  ifelse(vote == "AfD", "right", "")))))))
df_transformed$compass <- as_factor(df_transformed$compass)
df_transformed$income_factored <- ifelse(df$income <= 1, "Not satisfied",
                                       ifelse(df$income <= 2, "Neutral",
                                              "Satisfied"))

df_transformed$income_factored <- as_factor(df_transformed$income_factored)

df_transformed <- df_transformed %>%
    mutate(income_factored = factor(income_factored,
                                    levels = c("Not satisfied", "Neutral",
                                               "Satisfied"), ordered = T))
```

Now, we are going to create a discrete scale of 3 levels for some variables.

For the variable `egoposition_inmigration` we create the 3 following levels: Con, Neutral and Pro.

For the variable `political_interest` we create the following 3 levels: Not interested, Neutral and Interested.

For the variable `income` we create the following 3 levels: Not satisfied, Neutral and Satisfied.

```r
df_preproc <- df %>%
  mutate(egoposition_factored = ifelse(egoposition_immigration <= 4, "Pro",
                              ifelse(egoposition_immigration <= 6, "Neutral",
                                     "Con"))) %>%
  mutate(political_interest_factored = ifelse(political_interest <= 1,
                                              "Not interested",
                                       ifelse(political_interest <= 2,
                                              "Neutral", "Interested"))) %>%
    mutate(income_factored = ifelse(income <= 1, "Not satisfied",
                                    ifelse(income <= 2, "Neutral",
                                           "Satisfied"))) %>%
  mutate(ostwest = factor(ostwest, labels = c("West Germany", "East Germany")),
         gender = factor(gender, labels = c("Male", "Female")))
```

We proceed to order the new created factors.

```r
df_preproc <- df_preproc %>%
  mutate(egoposition_factored = factor(egoposition_factored,
                                       levels = c("Pro", "Neutral", "Con"),
                                       ordered = T)) %>%
    mutate(political_interest_factored = factor(political_interest_factored,
                                                levels = c("Not interested",
                                                           "Neutral",
                                                           "Interested"),
                                                ordered = T)) %>%
    mutate(income_factored = factor(income_factored, levels = c("Not satisfied",
                                                                "Neutral",
                                                                "Satisfied"),
                                    ordered = T))


df_preproc <- df_preproc %>%
    mutate(compass = ifelse(vote == "FDP", "center",
                     ifelse(vote == "SPD", "center",
                     ifelse(vote == "CDU/CSU", "center",
                     ifelse(vote == "Gruene", "left",
                     ifelse(vote == "LINKE", "left",
                    ifelse(vote == "AfD", "right", ""))))))) %>%
  mutate(clear_party = compass != "center") %>%
  mutate(right_wing = compass == "right")

df_preproc$compass <- as_factor(df_preproc$compass)
```

2 Exploratory Data Analysis

2.1 Univariate Descriptive Analysis

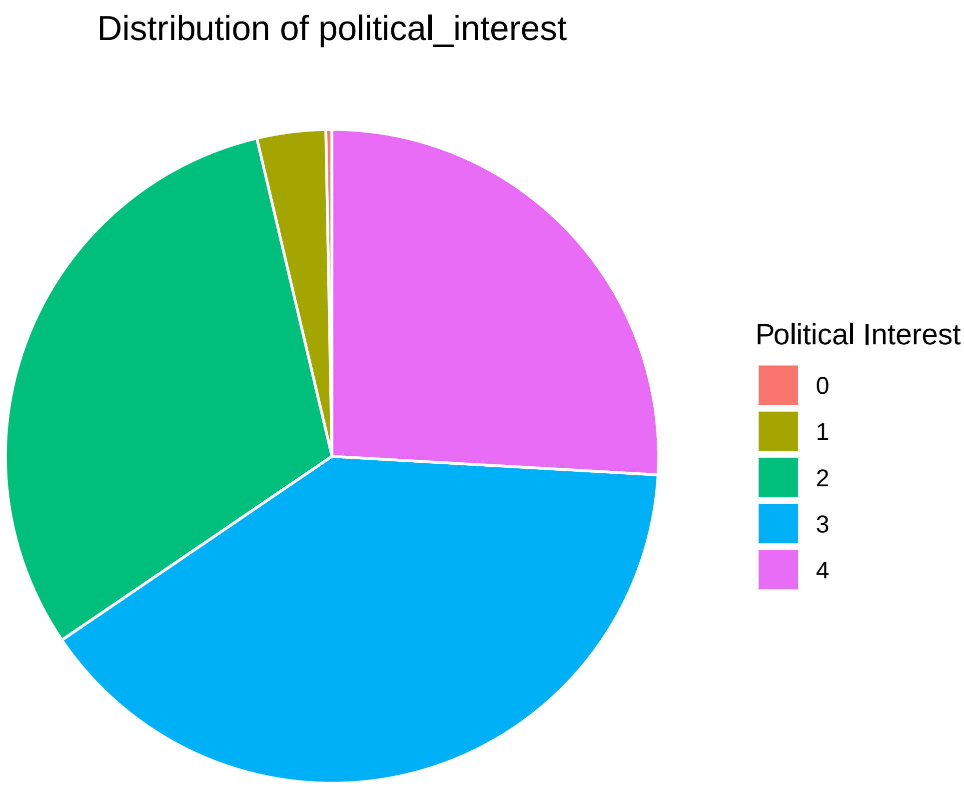
From the graphs we can observe several things. Regarding the position on immigration, the distribution we have seems to be centered on the value 4-5. That is, they are neither very open nor very restrictive to it. It also seems that there are more people who are extremely open than extremely closed, if we look at the two extremes.

In the political parties we have that CDU/CSU and SPD have by far the most voters. So, the data set contains individuals that mostly voted for parties that belong to the center wing, followed by the left wing and lastly the right wing. On the location of the observations, we have that almost all of them come from East Germany, and about 230 from West Germany.

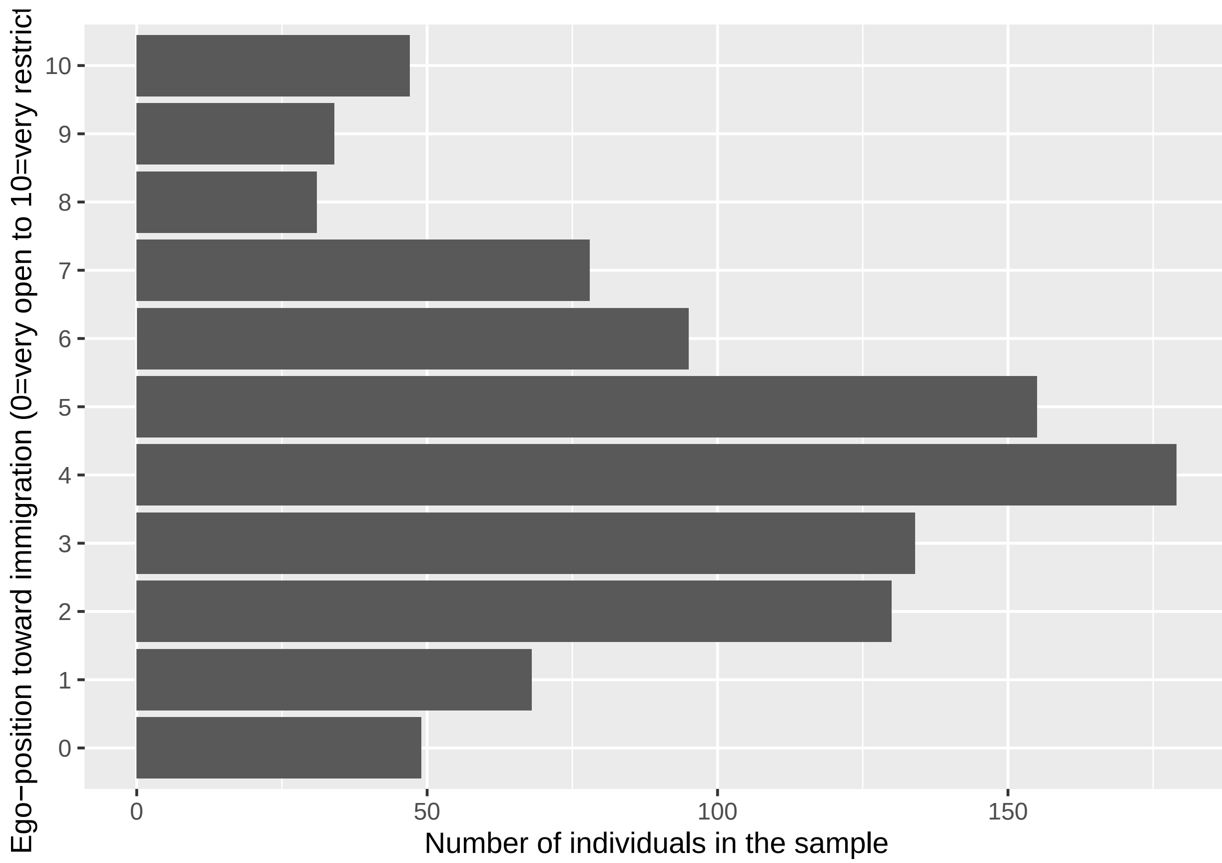
In the graph of political interest we can see that the majority of people have a medium-high interest, tending upwards. In the graph on salary satisfaction we have that most people have a medium-high satisfaction with what they earn (level 3) and there are very few observations with a satisfaction of 1 or lower.

Finally, the last thing we can see is that among the observations in the dataset we have more men (about 80) than women.

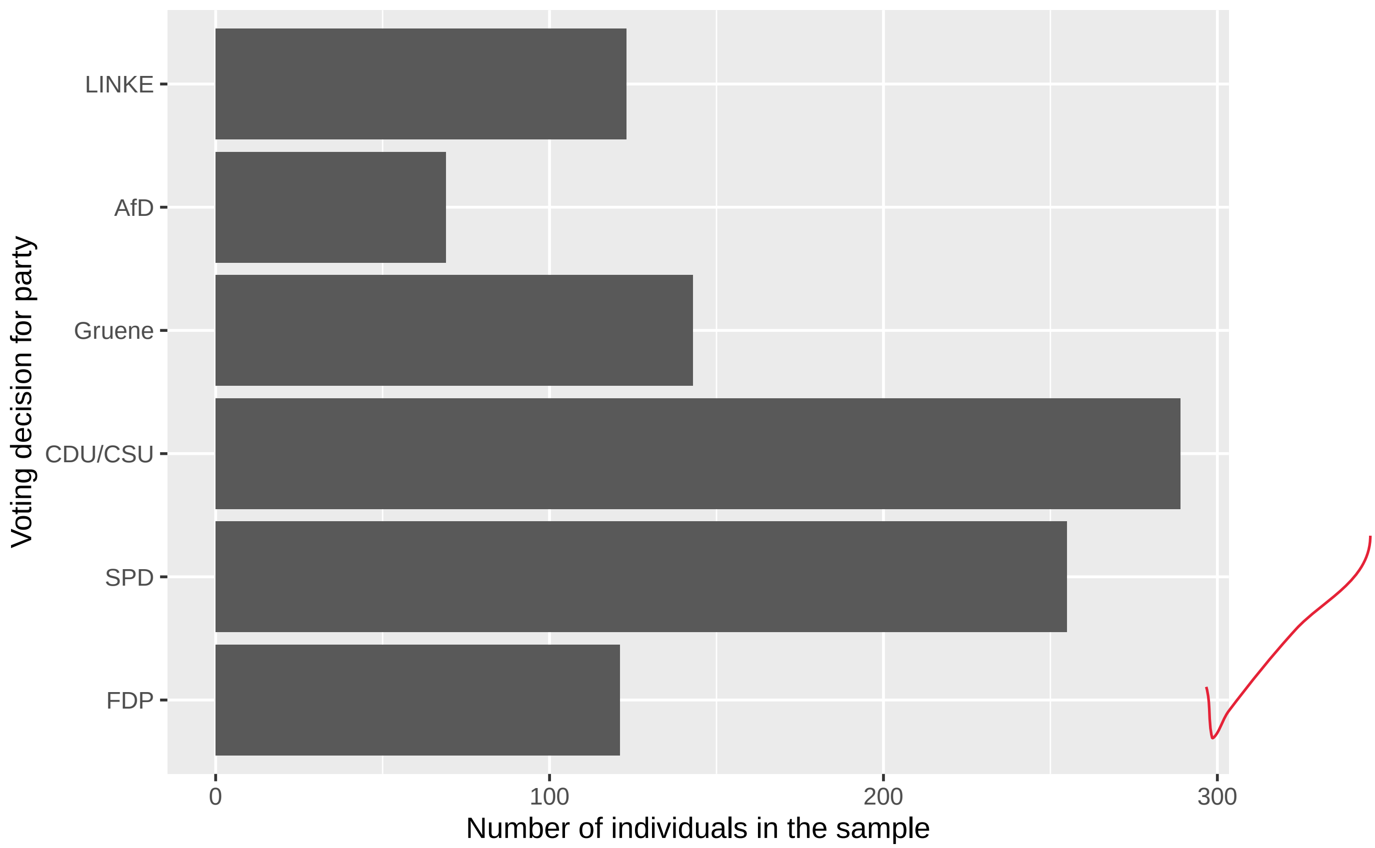
```
imm_countplot <- df_transformed %>%
  ggplot(aes(x = egoposition_immigration)) + geom_bar() + coord_flip() +
 xlab("Ego-position toward immigration (0=very open to 10=very restrictive)") +
  ylab("Number of individuals in the sample")

vote_countplot <- df_transformed %>%
  ggplot(aes(x = vote)) + geom_bar() + coord_flip() +
  xlab("Voting decision for party") +
  ylab("Number of individuals in the sample")

loc_countplot <- df_transformed %>%
  ggplot(aes(x = ostwest)) + geom_bar() + coord_flip() +
  xlab("Respondent location") +
  ylab("Number of individuals in the sample")


polint_countplot <- df_transformed %>%
  ggplot(aes(x = political_interest)) + geom_bar() + coord_flip() +
  xlab("Measurement for political interst (0 = low, 4 = high)") +
  ylab("Number of individuals in the sample")


ggplot(data.frame(table(df$political_interest)), aes(x="", y=Freq, fill=Var1)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar("y", start=0) +
  ggtitle("Distribution of political_interest") +
  theme_void() +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_fill_discrete(name = "Political Interest")
```



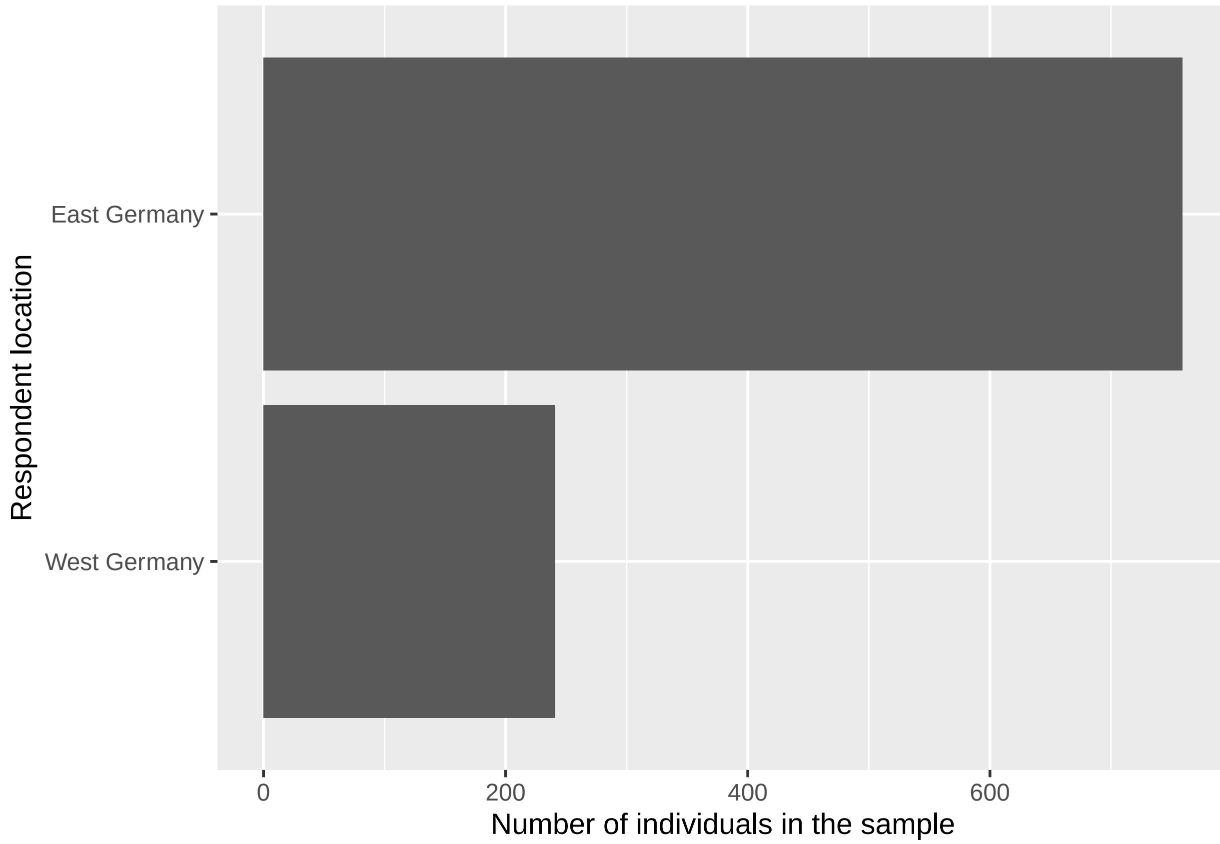

```
inc_countplot <- df_transformed %>%
  ggplot(aes(x = income)) + geom_bar() + coord_flip() +
  xlab("Self-reported income satisfaction (0 = low, 4 = high") +
  ylab("Number of individuals in the sample")

gend_countplot <- df_transformed %>%
  ggplot(aes(x = gender)) + geom_bar() + coord_flip() +
  xlab("Self-reported gender") +
  ylab("Number of individuals in the sample")

vote_withLevels_countplot <- df_transformed %>%
  ggplot(aes(x = compass)) + geom_bar() + coord_flip() +
  xlab("Political compass") +
  ylab("Number of individuals in the sample")

imm_countplot
```



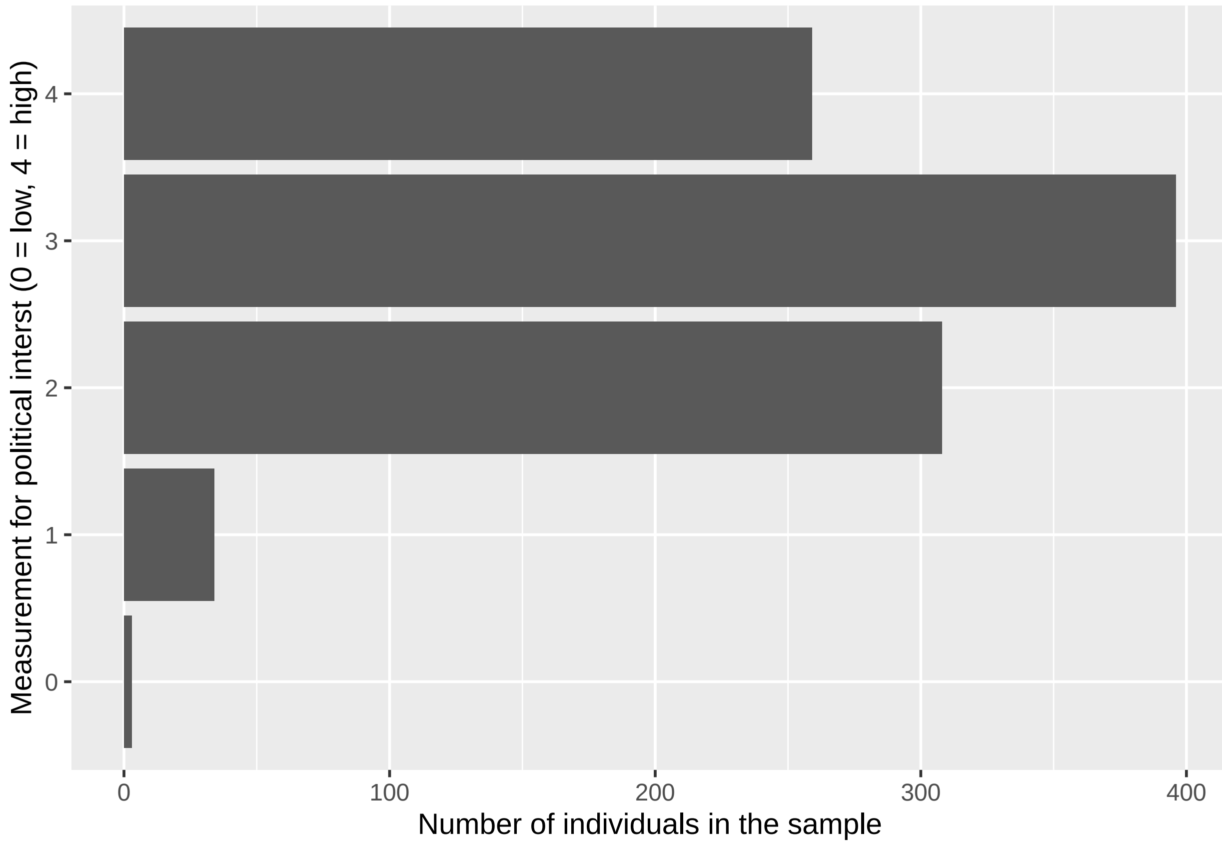

```
vote_countplot
```



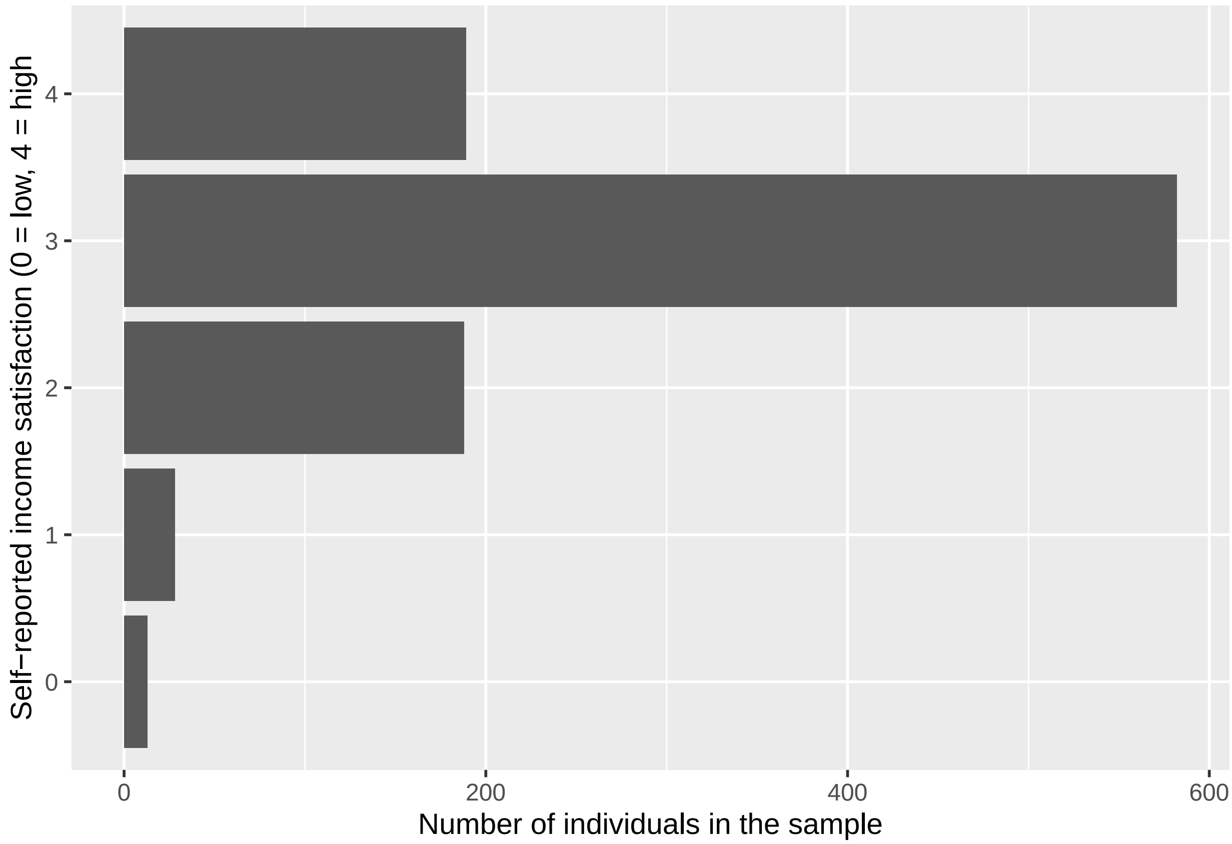

```
loc_countplot
```



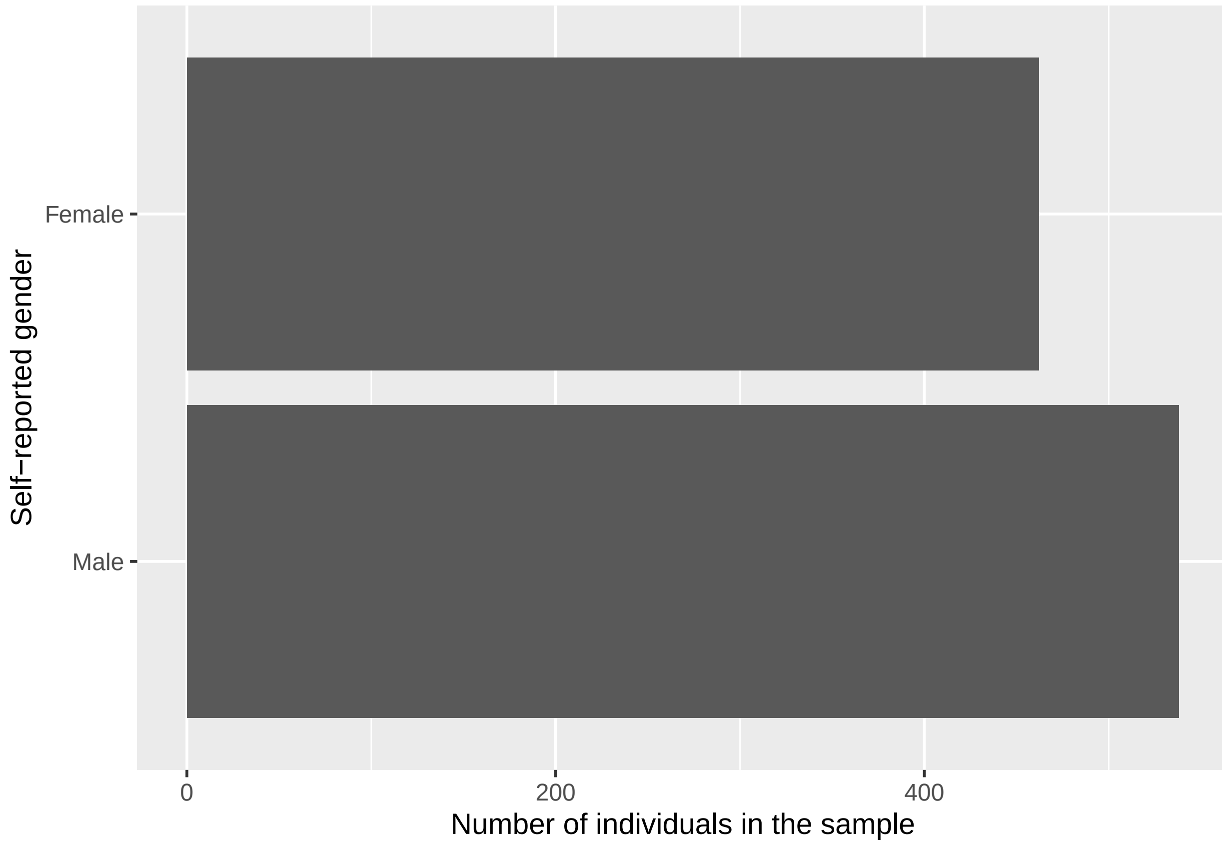

```
polint_countplot
```



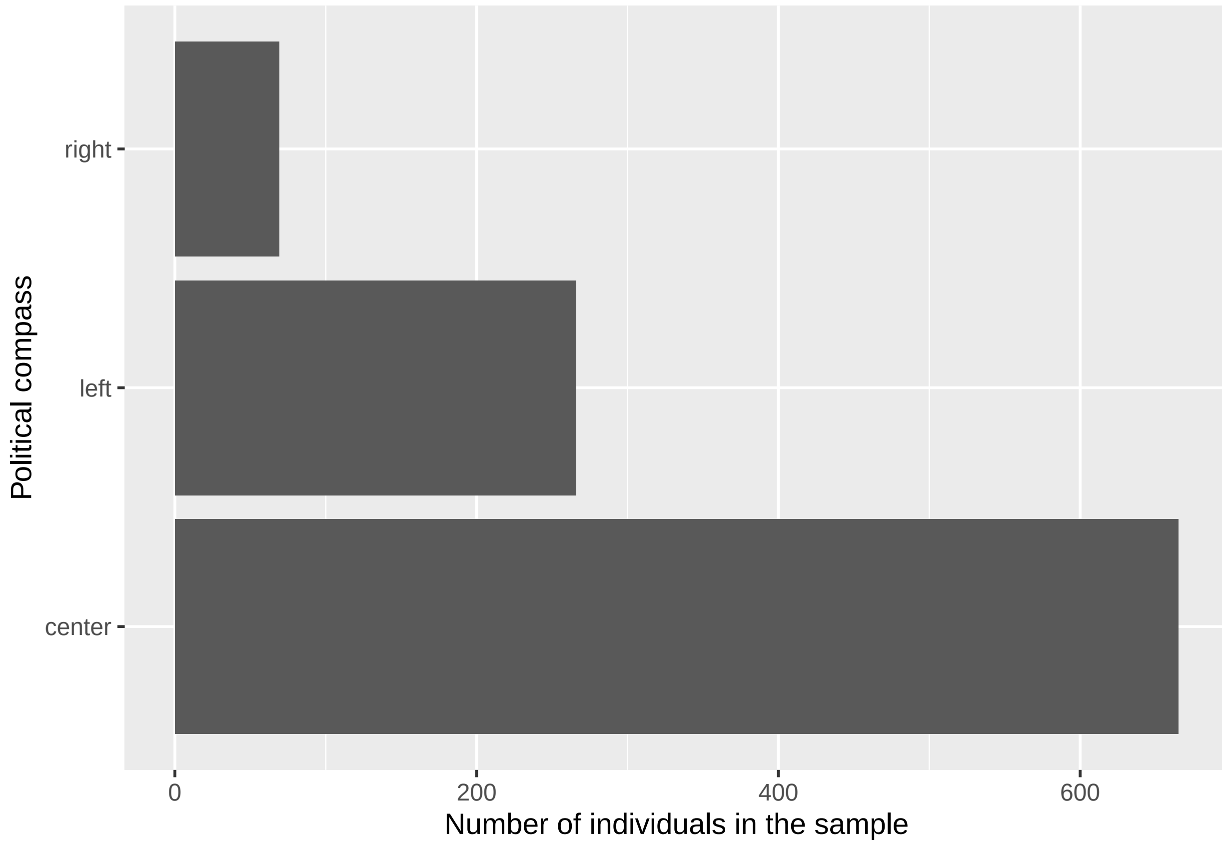

```
inc_countplot
```




```
gend_countplot
```

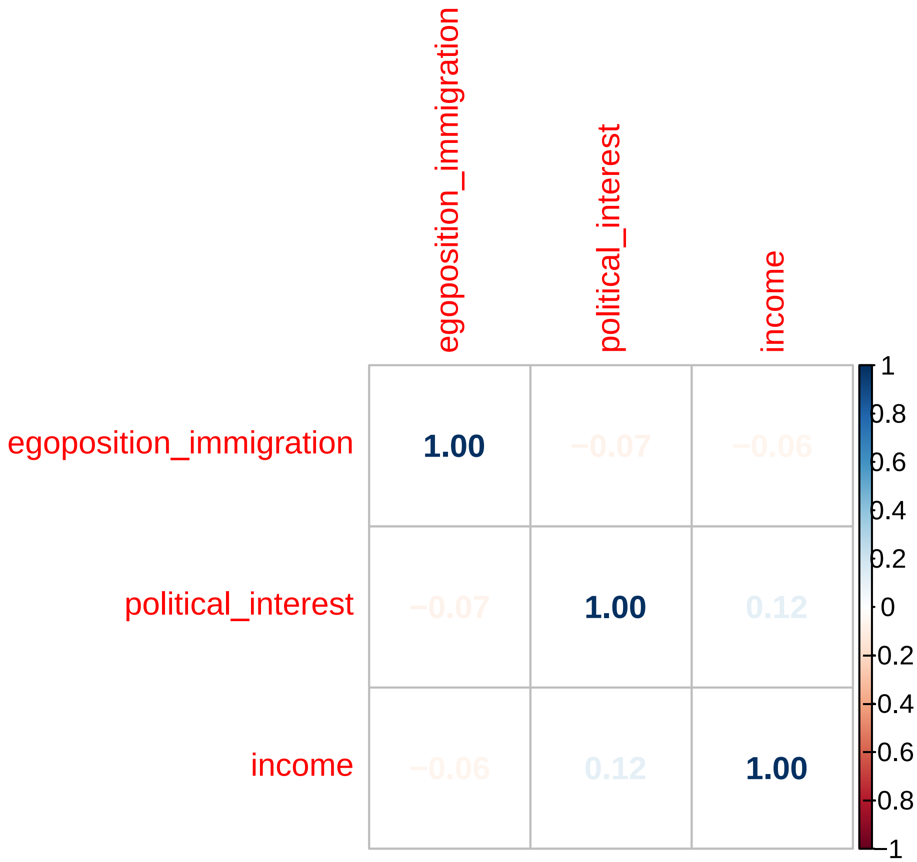



```
vote_withLevels_countplot
```



We compute the correlation matrix using the spearman coefficient. From it we can establish that there is no correlation between the numerical representation of the variables.

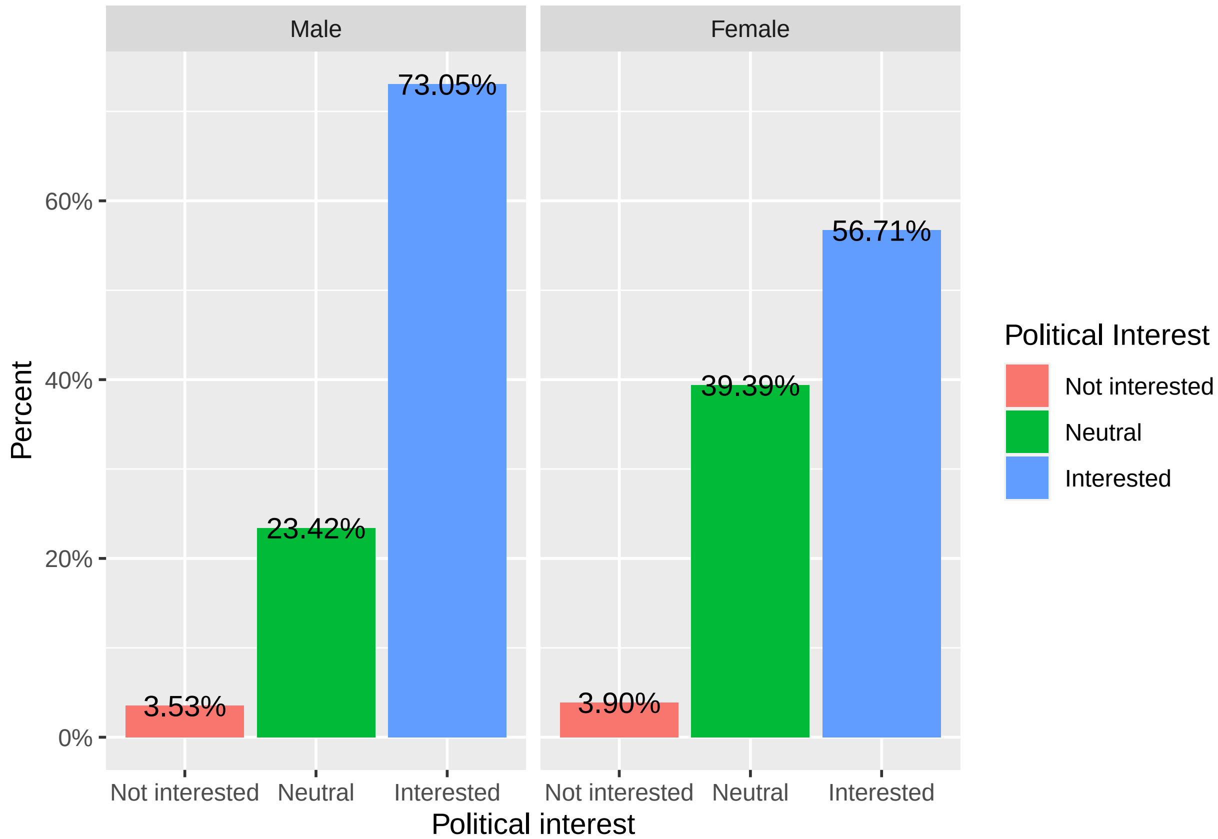
```
df_preproc %>%
  dplyr::select(where(is.numeric)) %>%
  cor() %>%
  corrplot(method="number")
```



2.2 Political interest vs Gender

We can see that no matter the gender the dataset is conformed by individuals that are interested in politics.

```
ggplot(df_preproc, aes(x= political_interest_factored,  group=gender)) +
    geom_bar(aes(y = ..prop..,
                 fill = factor(..x..,
                    labels=levels(df_preproc$political_interest_factored)))) +
    geom_text(aes( label = scales::percent(..prop..),
                   y= ..prop.. ), stat= "count") +
    labs(y = "Percent", fill="Political Interest") +
    xlab("Political interest")+
    facet_grid(~gender) +
    scale_y_continuous(labels = scales::percent)
```



2.3 Vote vs Gender

We can see that for all the parties except AFD, the percentage of male and female participants is quite equitative. For the party AfD there is quite a large difference between male and female participants, 81.2% are male and 18.8% are female.

```
ggplot(df_transformed, aes(x= gender,  group=vote)) +
    geom_bar(aes(y = ..prop..,
                 fill = factor(..x..,labels=levels(df_transformed$gender))),
             stat="count") +
    geom_text(aes( label = scales::percent(..prop..),
                   y= ..prop.. ), stat= "count", vjust = -.5) +
    labs(y = "Percentage", fill="Gender") +
    xlab("Gender")+
    facet_grid(~vote) +
    scale_y_continuous(labels = scales::percent) +
    theme(axis.text.x=element_blank(),
          axis.ticks.x=element_blank()
    )
```

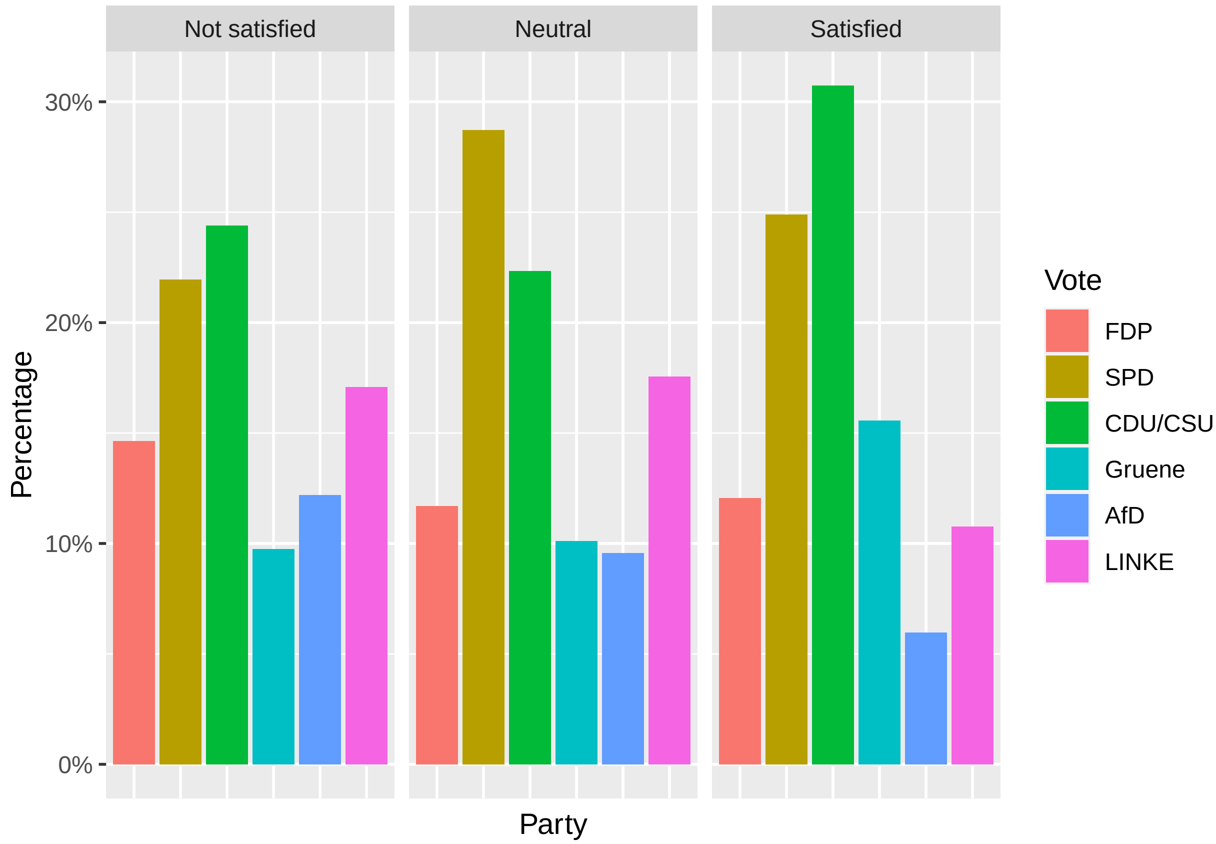
Vote vs income

It can be observed that the majority of people with a high self-reported income satisfaction vote for the CDU/CSU party.This can also be observed for people with a low income satisfaction.

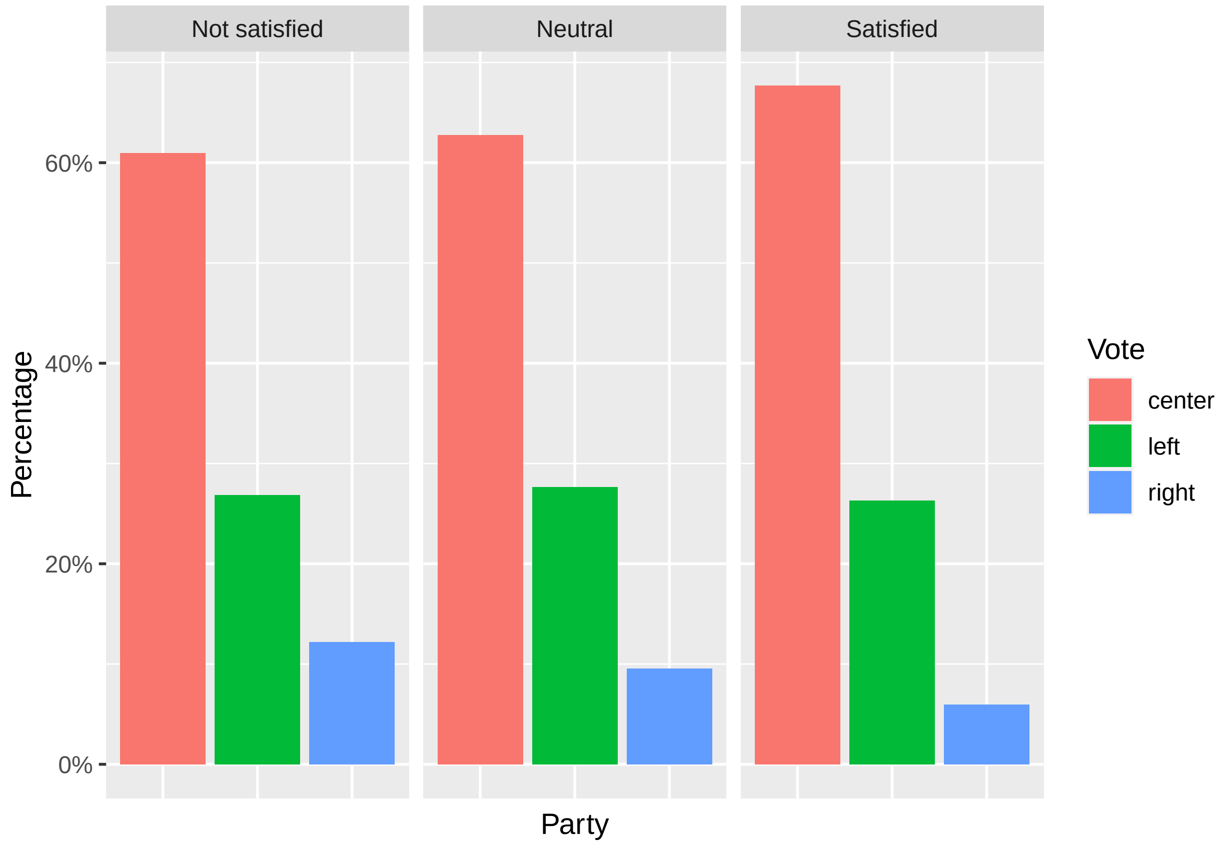
The majority of people with a medium income satisfaction (level 3) voted for the party SDP.

It can also be observed that no matter the level of income satisfaction the most voted party is the center one, followed by the left one and finally the right one.

```
ggplot(df_transformed, aes(x= vote,  group=income_factored)) +
    geom_bar(aes(y = ..prop..,
                 fill = factor(..x..,labels=levels(df_transformed$vote))),
             stat="count") +
    labs(y = "Percentage", fill="Vote") +
    xlab("Party")+
    facet_grid(~income_factored) +
    scale_y_continuous(labels = scales::percent) +
    theme(axis.text.x=element_blank(),
          axis.ticks.x=element_blank()
    )
```



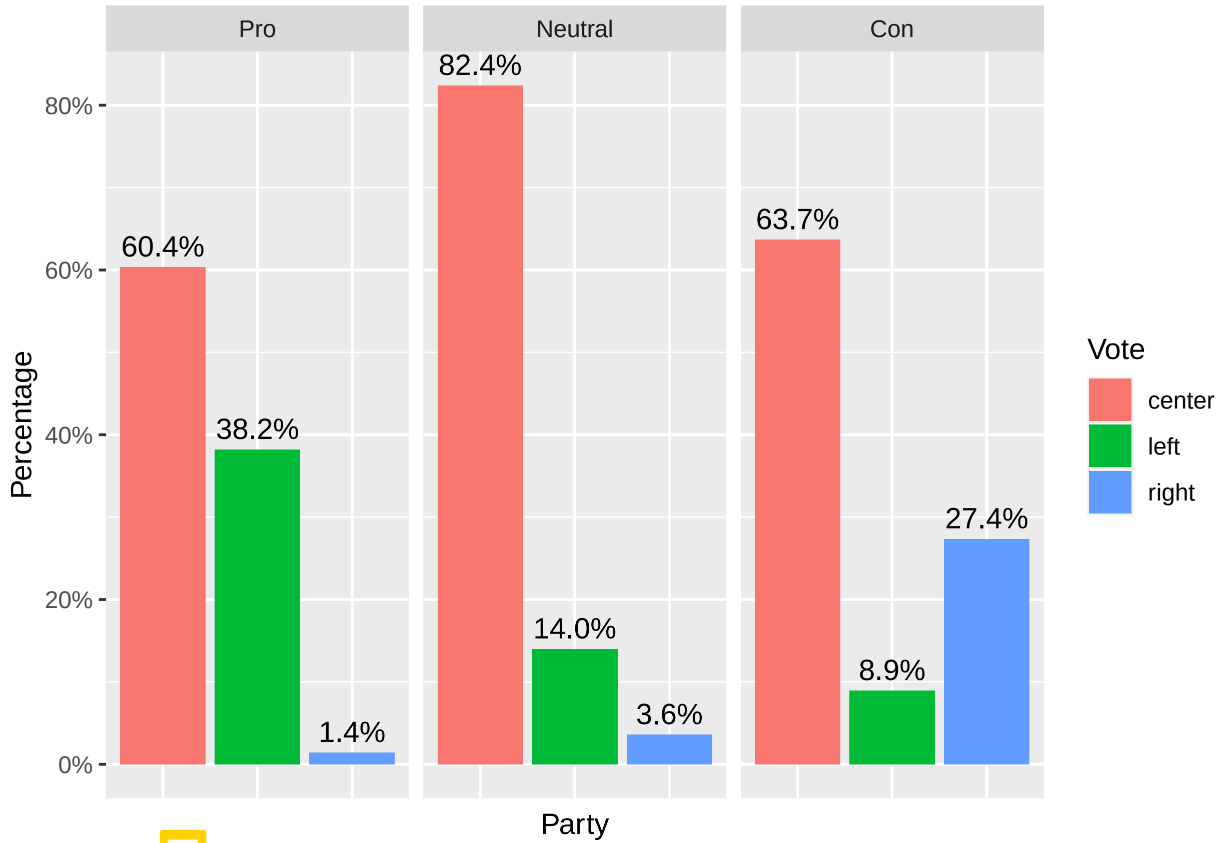

```
ggplot(df_transformed, aes(x= compass,  group=income_factored)) +
    geom_bar(aes(y = ..prop..,
                 fill = factor(..x..,labels=levels(df_transformed$compass))),
             stat="count") +
    labs(y = "Percentage", fill="Vote") +
    xlab("Party")+
    facet_grid(~income_factored) +
    scale_y_continuous(labels = scales::percent) +
    theme(axis.text.x=element_blank(),
          axis.ticks.x=element_blank()
    )
```



2.4 Vote vs egposition_inmigration

We can see that individuals belonging to the center are mainly neutral with respect to ego-position towards immigration. Individuals belonging to the right party are in the majority more restrictive with respect to ego-position towards immigration.

```
ggplot(df_preproc, aes(x= compass,  group=egoposition_factored)) +
    geom_bar(aes(y = ..prop..,
                 fill = factor(..x..,labels=levels(df_preproc$compass))),
             stat="count") +
    labs(y = "Percentage", fill="Vote") +
      xlab("Party")+
      geom_text(aes( label = scales::percent(..prop..),
                   y= ..prop.. ), stat= "count", vjust = -.5) +
    facet_grid(~egoposition_factored) +
    scale_y_continuous(labels = scales::percent) +
    theme(axis.text.x=element_blank(),
          axis.ticks.x=element_blank()
    )
```



3 Modeling

For the modeling part we decided to first model only using the numerical variables. And then do the same procedure but using the new factors created.

For the hierarchical the first level of the hierarchical approach will deal with the separation of the observations to right vs others (center and left), and the second level with the separation of observations between left and center.

3.1 Modeling according to a party

For the first problem, it should be noted that the dataset is unbalanced. Therefore, we are going to use stratification to create the train and test set.

```
set.seed(06041996)
party_split <- initial_split(df_preproc %>% mutate(vote = as_factor(vote)),
                             prop = 0.8, strata = vote)
train_data <- training(party_split)
test_data <- testing(party_split)
```

3.1.1 Multinomial

We start with the null model and we start adding factors as interactions and as additives. We then run step() to get the best model with the interactions. The best model we get is mm4m with an AIC of 2530.228.

```r
#Null model
mm0 <- multinom( vote ~ 1, data=train_data)
```

```
## # weights:  12 (5 variable)
## initial  value 1429.824056
## final  value 1341.119385
## converged
```

```r
summary(mm0)
```

```
## Call:
## multinom(formula = vote ~ 1, data = train_data)
##
## Coefficients:
##         (Intercept)
## SPD       0.7691290
## CDU/CSU   0.8842028
## Gruene    0.1910530
## AfD      -0.5835871
## LINKE     0.0512871
##
## Std. Errors:
##         (Intercept)
## SPD       0.1241143
## CDU/CSU   0.1219595
## Gruene    0.1386432
## AfD       0.1714474
## LINKE     0.1432701
##
## Residual Deviance: 2682.239
## AIC: 2692.239
```

```r
mm1 <- multinom( vote ~ political_interest, data=train_data)
```

```
## # weights:  18 (10 variable)
## initial  value 1429.824056
## iter  10 value 1343.670492
## final  value 1338.744187
## converged
```

```r
mm2 <- multinom( vote ~ income + egoposition_immigration,
                 data=train_data, Hess=T)
```

```
## # weights:  24 (15 variable)
## initial  value 1429.824056
## iter  10 value 1290.007555
## iter  20 value 1252.932545
## final  value 1252.932459
## converged
```

```r
mm3 <- multinom( vote ~ poly(income,2), data=train_data)
```

```
## # weights:  24 (15 variable)
## initial  value 1429.824056
## iter  10 value 1332.193956
## iter  20 value 1331.821614
## iter  30 value 1331.810542
## final  value 1331.810502
## converged
```

```r
summary(mm3)
```

```
## Call:
## multinom(formula = vote ~ poly(income, 2), data = train_data)
##
## Coefficients:
##         (Intercept) poly(income, 2)1 poly(income, 2)2
## SPD      0.76976104        0.8342696       -1.2936637
## CDU/CSU  0.87912589        3.7792247        0.4742663
## Gruene   0.17866782        5.6663059       -1.0748300
## AfD     -0.58878305       -2.0494448       -0.1366940
## LINKE   -0.01504744       -8.8550555       -5.6328532
##
## Std. Errors:
##         (Intercept) poly(income, 2)1 poly(income, 2)2
## SPD       0.1241957         3.452678         3.396269
## CDU/CSU   0.1222420         3.387357         3.315943
## Gruene    0.1396198         4.096220         4.091000
## AfD       0.1721304         4.649487         4.451249
## LINKE     0.1486304         4.151730         4.068457
##
## Residual Deviance: 2663.621
## AIC: 2693.621
```

```r
mm4 <- multinom( vote ~ poly(political_interest,3), data=train_data, Hess=T)
```

```
## # weights:  30 (20 variable)
## initial  value 1429.824056
## iter  10 value 1331.750280
## iter  20 value 1329.828835
## iter  30 value 1329.730415
## iter  40 value 1329.729293
## final  value 1329.729253
## converged
```

```r
mm4m <- multinom(vote ~ income*political_interest*egoposition_immigration +
                   I(income^2) + I(political_interest^2) +
                   I(egoposition_immigration^2), data=train_data, Hess=T)
```

```
## # weights:  72 (55 variable)
## initial  value 1429.824056
```

```
## iter  10 value 1263.717307
## iter  20 value 1251.161051
## iter  30 value 1239.385354
## iter  40 value 1228.887424
## iter  50 value 1226.711368
## iter  60 value 1226.461272
## final  value 1226.460784
## converged
```

```
mm4m <- stats::step(mm4m)
```

```
## Start:  AIC=2562.92
## vote ~ income * political_interest * egoposition_immigration +
##     I(income^2) + I(political_interest^2) + I(egoposition_immigration^2)
##
## trying - I(income^2)
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1263.445785
## iter  20 value 1250.514884
## iter  30 value 1240.362694
## iter  40 value 1229.347276
## iter  50 value 1228.493868
## iter  60 value 1228.413878
## final  value 1228.409633
## converged
## trying - I(political_interest^2)
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1266.508096
## iter  20 value 1252.827629
## iter  30 value 1243.092636
## iter  40 value 1231.258399
## iter  50 value 1230.067412
## iter  60 value 1230.000789
## final  value 1229.994776
## converged
## trying - I(egoposition_immigration^2)
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1267.798781
## iter  20 value 1257.794514
## iter  30 value 1244.015764
## iter  40 value 1234.629982
## iter  50 value 1233.461741
## iter  60 value 1233.417746
## final  value 1233.417067
## converged
## trying - income:political_interest:egoposition_immigration
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1301.800722
## iter  20 value 1280.874381
## iter  30 value 1253.111269
```

```
## iter  40 value 1231.271899
## iter  50 value 1228.379568
## iter  60 value 1228.360064
## final  value 1228.359911
## converged
##                                                   Df     AIC
## - income:political_interest:egoposition_immigration 50 2556.720
## - I(income^2)                                     50 2556.819
## - I(political_interest^2)                         50 2559.990
## <none>                                            55 2562.922
## - I(egoposition_immigration^2)                    50 2566.834
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1301.800722
## iter  20 value 1280.874381
## iter  30 value 1253.111269
## iter  40 value 1231.271899
## iter  50 value 1228.379568
## iter  60 value 1228.360064
## final  value 1228.359911
## converged
##
## Step:  AIC=2556.72
## vote ~ income + political_interest + egoposition_immigration +
##     I(income^2) + I(political_interest^2) + I(egoposition_immigration^2) +
##     income:political_interest + income:egoposition_immigration +
##     political_interest:egoposition_immigration
##
## trying - I(income^2)
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1299.849369
## iter  20 value 1272.999936
## iter  30 value 1240.176774
## iter  40 value 1231.102115
## iter  50 value 1230.466930
## final  value 1230.456227
## converged
## trying - I(political_interest^2)
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1302.819780
## iter  20 value 1278.559647
## iter  30 value 1240.852232
## iter  40 value 1232.577965
## iter  50 value 1231.900702
## final  value 1231.885800
## converged
## trying - I(egoposition_immigration^2)
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1268.430959
## iter  20 value 1255.711717
## iter  30 value 1238.630662
```

```
## iter  40 value 1235.379653
## iter  50 value 1235.209143
## final  value 1235.208834
## converged
## trying - income:political_interest
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1299.441751
## iter  20 value 1280.397210
## iter  30 value 1248.529759
## iter  40 value 1232.910034
## iter  50 value 1232.456280
## final  value 1232.453906
## converged
## trying - income:egoposition_immigration
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1309.641388
## iter  20 value 1280.180852
## iter  30 value 1237.988459
## iter  40 value 1229.626778
## iter  50 value 1229.161411
## final  value 1229.160485
## converged
## trying - political_interest:egoposition_immigration
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1313.348203
## iter  20 value 1279.69[illegible]6
## iter  30 value 1247.080153
## iter  40 value 1232.947843
## iter  50 value 1231.533899
## final  value 1231.518597
## converged
##                                               Df      AIC
## - income:egoposition_immigration             45 2548.321
## - I(income^2)                                45 2550.912
## - political_interest:egoposition_immigration 45 2553.037
## - I(political_interest^2)                    45 2553.772
## - income:political_interest                  45 2554.908
## <none>                                       50 2556.720
## - I(egoposition_immigration^2)               45 2560.418
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1309.641388
## iter  20 value 1280.180852
## iter  30 value 1237.988459
## iter  40 value 1229.626778
## iter  50 value 1229.161411
## final  value 1229.160485
## converged
##
## Step:  AIC=2548.32
## vote ~ income + political_interest + egoposition_immigration +
```

```
##     I(income^2) + I(political_interest^2) + I(egoposition_immigration^2) +
##     income:political_interest + political_interest:egoposition_immigration
##
## trying - I(income^2)
## # weights:  54 (40 variable)
## initial  value 1429.824056
## iter  10 value 1314.502801
## iter  20 value 1284.274944
## iter  30 value 1236.977047
## iter  40 value 1231.297581
## iter  50 value 1231.243800
## final  value 1231.243692
## converged
## trying - I(political_interest^2)
## # weights:  54 (40 variable)
## initial  value 1429.824056
## iter  10 value 1317.940516
## iter  20 value 1283.042259
## iter  30 value 1238.781043
## iter  40 value 1232.775807
## iter  50 value 1232.696318
## final  value 1232.696005
## converged
## trying - I(egoposition_immigration^2)
## # weights:  54 (40 variable)
## initial  value 1429.824056
## iter  10 value 1290.338975
## iter  20 value 1267.198237
## iter  30 value 1237.987173
## iter  40 value 1236.179770
## final  value 1236.177824
## converged
## trying - income:political_interest
## # weights:  54 (40 variable)
## initial  value 1429.824056
## iter  10 value 1308.052028
## iter  20 value 1283.796999
## iter  30 value 1240.123674
## iter  40 value 1233.427760
## iter  50 value 1233.345976
## iter  50 value 1233.345975
## iter  50 value 1233.345975
## final  value 1233.345975
## converged
## trying - political_interest:egoposition_immigration
## # weights:  54 (40 variable)
## initial  value 1429.824056
## iter  10 value 1323.359262
## iter  20 value 1289.335094
## iter  30 value 1240.204727
## iter  40 value 1232.499810
## iter  50 value 1232.468199
## final  value 1232.468172
## converged
```

```
##                                              Df      AIC
## - I(income^2)                                40 2542.487
## - political_interest:egoposition_immigration 40 2544.936
## - I(political_interest^2)                    40 2545.392
## - income:political_interest                  40 2546.692
## <none>                                       45 2548.321
## - I(egoposition_immigration^2)               40 2552.356
## # weights:  54 (40 variable)
## initial  value 1429.824056
## iter  10 value 1314.502801
## iter  20 value 1284.274944
## iter  30 value 1236.977047
## iter  40 value 1231.297581
## iter  50 value 1231.243800
## final  value 1231.243692
## converged
##
## Step:  AIC=2542.49
## vote ~ income + political_interest + egoposition_immigration +
##     I(political_interest^2) + I(egoposition_immigration^2) +
##     income:political_interest + political_interest:egoposition_immigration
##
## trying - I(political_interest^2)
## # weights:  48 (35 variable)
## initial  value 1429.824056
## iter  10 value 1310.275595
## iter  20 value 1271.826332
## iter  30 value 1237.103048
## iter  40 value 1234.781921
## final  value 1234.767516
## converged
## trying - I(egoposition_immigration^2)
## # weights:  48 (35 variable)
## initial  value 1429.824056
## iter  10 value 1292.249962
## iter  20 value 1270.563881
## iter  30 value 1238.925962
## iter  40 value 1238.355357
## final  value 1238.354666
## converged
## trying - income:political_interest
## # weights:  48 (35 variable)
## initial  value 1429.824056
## iter  10 value 1300.579140
## iter  20 value 1275.534825
## iter  30 value 1240.119480
## iter  40 value 1236.212568
## final  value 1236.204323
## converged
## trying - political_interest:egoposition_immigration
## # weights:  48 (35 variable)
## initial  value 1429.824056
## iter  10 value 1331.487378
## iter  20 value 1289.021836
```

```
## iter  30 value 1237.672158
## iter  40 value 1234.459457
## final  value 1234.455823
## converged
##                                              Df      AIC
## - political_interest:egoposition_immigration 35 2538.912
## - I(political_interest^2)                    35 2539.535
## - income:political_interest                  35 2542.409
## <none>                                       40 2542.487
## - I(egoposition_immigration^2)               35 2546.709
## # weights:  48 (35 variable)
## initial  value 1429.824056
## iter  10 value 1331.487378
## iter  20 value 1289.021836
## iter  30 value 1237.672158
## iter  40 value 1234.459457
## final  value 1234.455823
## converged
##
## Step:  AIC=2538.91
## vote ~ income + political_interest + egoposition_immigration +
##     I(political_interest^2) + I(egoposition_immigration^2) +
##     income:political_interest
##
## trying - egoposition_immigration
## # weights:  42 (30 variable)
## initial  value 1429.824056
## iter  10 value 1331.647824
## iter  20 value 1288.372277
## iter  30 value 1252.049528
## final  value 1251.887663
## converged
## trying - I(political_interest^2)
## # weights:  42 (30 variable)
## initial  value 1429.824056
## iter  10 value 1330.558915
## iter  20 value 1261.410477
## iter  30 value 1239.320931
## iter  40 value 1238.174148
## final  value 1238.173372
## converged
## trying - I(egoposition_immigration^2)
## # weights:  42 (30 variable)
## initial  value 1429.824056
## iter  10 value 1321.593659
## iter  20 value 1265.578112
## iter  30 value 1242.538089
## final  value 1242.361537
## converged
## trying - income:political_interest
## # weights:  42 (30 variable)
## initial  value 1429.824056
## iter  10 value 1323.386777
## iter  20 value 1267.263887
```

```
## iter  30 value 1239.519487
## iter  40 value 1239.365040
## iter  40 value 1239.365037
## iter  40 value 1239.365037
## final  value 1239.365037
## converged
##                               Df      AIC
## - I(political_interest^2)      30 2536.347
## - income:political_interest    30 2538.730
## <none>                         35 2538.912
## - I(egoposition_immigration^2) 30 2544.723
## - egoposition_immigration      30 2563.775
## # weights:  42 (30 variable)
## initial  value 1429.824056
## iter  10 value 1330.558915
## iter  20 value 1261.410477
## iter  30 value 1239.320931
## iter  40 value 1238.174148
## final  value 1238.173372
## converged
##
## Step:  AIC=2536.35
## vote ~ income + political_interest + egoposition_immigration +
##     I(egoposition_immigration^2) + income:political_interest
##
## trying - egoposition_immigration
## # weights:  36 (25 variable)
## initial  value 1429.824056
## iter  10 value 1347.499412
## iter  20 value 1265.248216
## iter  30 value 1255.848785
## final  value 1255.840704
## converged
## trying - I(egoposition_immigration^2)
## # weights:  36 (25 variable)
## initial  value 1429.824056
## iter  10 value 1284.187463
## iter  20 value 1252.926284
## iter  30 value 1246.204391
## final  value 1246.194220
## converged
## trying - income:political_interest
## # weights:  36 (25 variable)
## initial  value 1429.824056
## iter  10 value 1310.756634
## iter  20 value 1257.767162
## iter  30 value 1242.462082
## final  value 1242.308683
## converged
##                               Df      AIC
## - income:political_interest    25 2534.617
## <none>                         30 2536.347
## - I(egoposition_immigration^2) 25 2542.388
## - egoposition_immigration      25 2561.681
```

```
## # weights:  36 (25 variable)
## initial  value 1429.824056
## iter  10 value 1310.756634
## iter  20 value 1257.767162
## iter  30 value 1242.462082
## final  value 1242.308683
## converged
##
## Step:  AIC=2534.62
## vote ~ income + political_interest + egoposition_immigration +
##     I(egoposition_immigration^2)
##
## trying - income
## # weights:  30 (20 variable)
## initial  value 1429.824056
## iter  10 value 1296.238027
## iter  20 value 1251.002626
## iter  30 value 1250.013407
## final  value 1250.013388
## converged
## trying - political_interest
## # weights:  30 (20 variable)
## initial  value 1429.824056
## iter  10 value 1301.343696
## iter  20 value 1246.577320
## iter  30 value 1245.114066
## iter  30 value 1245.114057
## iter  30 value 1245.114057
## final  value 1245.114057
## converged
## trying - egoposition_immigration
## # weights:  30 (20 variable)
## initial  value 1429.824056
## iter  10 value 1329.510503
## iter  20 value 1263.109717
## final  value 1259.694044
## converged
## trying - I(egoposition_immigration^2)
## # weights:  30 (20 variable)
## initial  value 1429.824056
## iter  10 value 1319.914753
## iter  20 value 1250.964495
## final  value 1250.136764
## converged
##                                Df      AIC
## - political_interest           20 2530.228
## <none>                         25 2534.617
## - income                       20 2540.027
## - I(egoposition_immigration^2) 20 2540.274
## - egoposition_immigration      20 2559.388
## # weights:  30 (20 variable)
## initial  value 1429.824056
## iter  10 value 1301.343696
## iter  20 value 1246.577320
```

```
## iter  30 value 1245.114066
## iter  30 value 1245.114057
## iter  30 value 1245.114057
## final  value 1245.114057
## converged
##
## Step:  AIC=2530.23
## vote ~ income + egoposition_immigration + I(egoposition_immigration^2)
##
## trying - income
## # weights:  24 (15 variable)
## initial  value 1429.824056
## iter  10 value 1316.407928
## iter  20 value 1252.482595
## final  value 1252.408790
## converged
## trying - egoposition_immigration
## # weights:  24 (15 variable)
## initial  value 1429.824056
## iter  10 value 1305.942249
## iter  20 value 1262.634521
## final  value 1262.630418
## converged
## trying - I(egoposition_immigration^2)
## # weights:  24 (15 variable)
## initial  value 1429.824056
## iter  10 value 1290.007555
## iter  20 value 1252.932545
## final  value 1252.932459
## converged
##                                Df      AIC
## <none>                         20 2530.228
## - income                       15 2534.818
## - I(egoposition_immigration^2) 15 2535.865
## - egoposition_immigration      15 2555.261
```

```
summary(mm4m)
```

```
## Call:
## multinom(formula = vote ~ income + egoposition_immigration +
##     I(egoposition_immigration^2), data = train_data, Hess = T)
##
## Coefficients:
##         (Intercept)       income egoposition_immigration
## SPD       2.6650833 -0.001170858              -0.6808289
## CDU/CSU   0.8214006  0.175525335              -0.1305398
## Gruene    1.7952506  0.169916824              -0.5314969
## AfD      -4.0695551  0.050439509               0.6266280
## LINKE     3.4360629 -0.373838330              -0.8244800
##         I(egoposition_immigration^2)
## SPD                     0.0463329174
## CDU/CSU                 0.0066846298
## Gruene                  0.0009253267
## AfD                    -0.0128331806
```

```
## LINKE                     0.0501352530
##
## Std. Errors:
##         (Intercept)    income egoposition_immigration
## SPD       0.7141512 0.1655960              0.2081526
## CDU/CSU   0.7340268 0.1635743              0.2142396
## Gruene    0.8003723 0.1940883              0.2494100
## AfD       1.6619029 0.2336241              0.4718865
## LINKE     0.7570061 0.1829856              0.2282694
##         I(egoposition_immigration^2)
## SPD                       0.01921165
## CDU/CSU                   0.01943278
## Gruene                    0.02811157
## AfD                       0.03507256
## LINKE                     0.02225984
##
## Residual Deviance: 2490.228
## AIC: 2530.228
```

```
anova( mm0, mm1, test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: vote
##                Model Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
## 1                  1      3985   2682.239
## 2 political_interest      3980   2677.488 1 vs 2     5 4.750396 0.4470952
```

```
anova( mm2, mm3, test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: vote
##                              Model Resid. df Resid. Dev   Test    Df   LR stat.
## 1 income + egoposition_immigration      3975   2505.865
## 2                   poly(income, 2)      3975   2663.621 1 vs 2     0 -157.7561
##   Pr(Chi)
## 1
## 2       1
```

```
anova( mm3, mm4, test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: vote
##                           Model Resid. df Resid. Dev   Test    Df LR stat.
## 1               poly(income, 2)      3975   2663.621
## 2 poly(political_interest, 3)      3970   2659.459 1 vs 2     5 4.162498
##     Pr(Chi)
## 1
## 2 0.5262654
```

```r
anova( mm3, mm4, test="Chisq")
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: vote
##                           Model Resid. df Resid. Dev   Test    Df LR stat.
## 1              poly(income, 2)      3975   2663.621
## 2 poly(political_interest, 3)      3970   2659.459 1 vs 2     5 4.162498
##     Pr(Chi)
## 1
## 2 0.5262654
```

```r
AIC(mm1,mm4,mm2,mm3,mm4m)
```

```
##      df      AIC
## mm1  10 2697.488
## mm4  20 2699.459
## mm2  15 2535.865
## mm3  15 2693.621
## mm4m 20 2530.228
```

3.1.2 Metrics

From the ROC curves we can see that the model is not performing very well. We get a recall of 24 and a precision of 33.

```r
summary(mm4m)
```

```
## Call:
## multinom(formula = vote ~ income + egoposition_immigration +
##     I(egoposition_immigration^2), data = train_data, Hess = T)
##
## Coefficients:
##         (Intercept)       income egoposition_immigration
## SPD       2.6650833 -0.001170858              -0.6808289
## CDU/CSU   0.8214006  0.175525335              -0.1305398
## Gruene    1.7952506  0.169916824              -0.5314969
## AfD      -4.0695551  0.050439509               0.6266280
## LINKE     3.4360629 -0.373838330              -0.8244800
##         I(egoposition_immigration^2)
## SPD                     0.0463329174
## CDU/CSU                 0.0066846298
## Gruene                  0.0009253267
## AfD                    -0.0128331806
## LINKE                   0.0501352530
##
## Std. Errors:
##         (Intercept)    income egoposition_immigration
## SPD       0.7141512 0.1655960               0.2081526
## CDU/CSU   0.7340268 0.1635743               0.2142396
## Gruene    0.8003723 0.1940883               0.2494100
```

```
## AfD       1.6619029 0.2336241               0.4718865
## LINKE     0.7570061 0.1829856               0.2282694
##         I(egoposition_immigration^2)
## SPD                        0.01921165
## CDU/CSU                    0.01943278
## Gruene                     0.02811157
## AfD                        0.03507256
## LINKE                      0.02225984
##
## Residual Deviance: 2490.228
## AIC: 2530.228
```

```
sum(predict(mm4m, type="class") == train_data$vote) / nrow(train_data)
```

```
## [1] 0.3408521
```

```
sum(predict(mm4m, test_data, type="class") == test_data$vote) / nrow(test_data)
```

```
## [1] 0.3415842
```

```
preds_train <- tibble(
  pred = predict(mm4m, type="class"), true = train_data$vote
  )

preds_train <- preds_train  %>%
  mutate(
         true = fct_relevel(true, levels(pred)))

preds_test <- tibble(pred = predict(mm4m, test_data, type="class"),
                     true = test_data$vote)

preds_test <- preds_test %>%
    mutate(
    true = fct_relevel(true, levels(pred)))

preds_train %>%
  recall(true, pred)
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 recall  macro          0.242
```

```
preds_train %>%
  precision(true, pred)
```

```
## Warning: While computing multiclass ‘precision()‘, some levels had no predicted events (i.e. ‘true_po
## Precision is undefined in this case, and those levels will be removed from the averaged result.
## Note that the following number of true events actually occured for each problematic event level:
## 'FDP': 95
## 'Gruene': 115
```

```
## # A tibble: 1 x 3
##   .metric   .estimator .estimate
##   <chr>     <chr>          <dbl>
## 1 precision macro          0.325
```

```
preds_train %>%
  f_meas(true, pred)
```

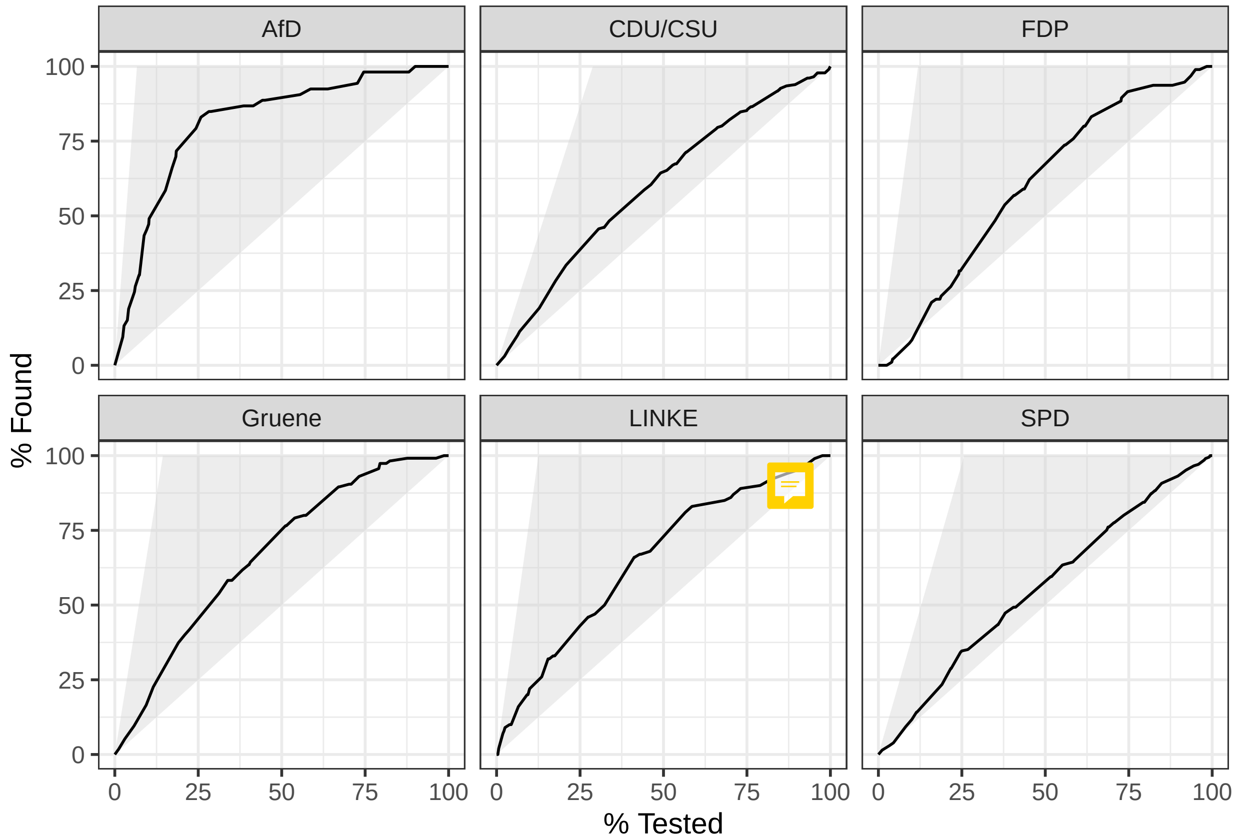
```
## Warning: While computing multiclass `precision()`, some levels had no predicted events (i.e. `true_po
## Precision is undefined in this case, and those levels will be removed from the averaged result.
## Note that the following number of true events actually occured for each problematic event level:
## 'FDP': 95
## 'Gruene': 115
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 f_meas  macro          0.308
```

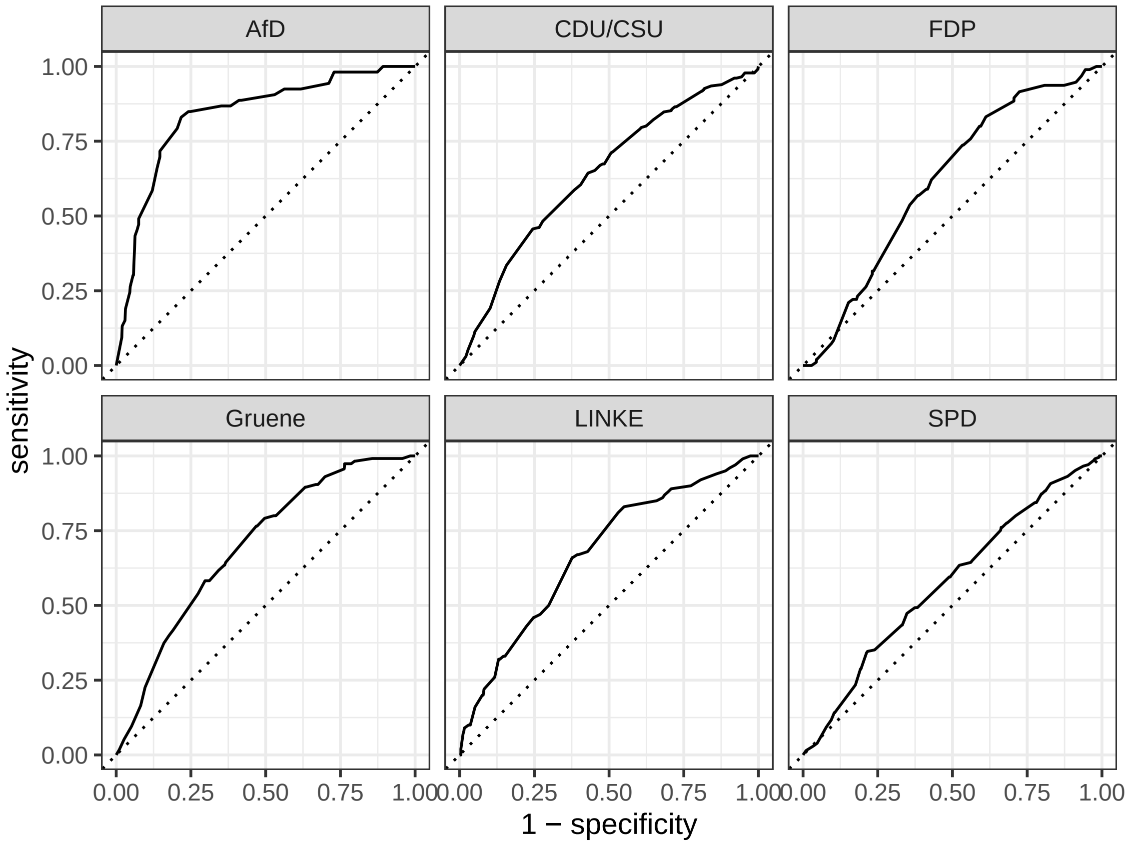
```
preds_train %>%
  conf_mat(true, pred)
```

```
##           Truth
## Prediction FDP SPD CDU/CSU Gruene AfD LINKE
##    FDP       0   0       0      0   0     0
##    SPD      17  86      49     66   4    52
##    CDU/CSU  69 104     165     45  37    35
##    Gruene    0   0       0      0   0     0
##    AfD       7  10       9      1  12     4
##    LINKE     2   5       7      3   0     9
```

```
predict(mm4m, type="prob") %>%
  bind_cols(train_data) %>%
  gain_curve(vote, FDP:LINKE) %>%
  autoplot()
```




```
predict(mm4m, type="prob") %>%
  bind_cols(train_data) %>%
  roc_curve(vote, FDP:LINKE) %>%
  autoplot()
```




```r
predict(mm4m, type="prob") %>%
  bind_cols(train_data) %>%
  roc_auc(vote, FDP:LINKE)
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 roc_auc hand_till      0.681
```

```r
preds_test %>%
  recall(true, pred)
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 recall  macro          0.267
```

```r
preds_test %>%
  precision(true, pred)
```

```
## Warning: While computing multiclass ‘precision()‘, some levels had no predicted events (i.e. ‘true_po
## Precision is undefined in this case, and those levels will be removed from the averaged result.
## Note that the following number of true events actually occured for each problematic event level:
## ’FDP’: 26
## ’Gruene’: 28
```

```
## # A tibble: 1 x 3
##   .metric   .estimator .estimate
##   <chr>     <chr>          <dbl>
## 1 precision macro          0.364
```

```r
preds_test %>%
  f_meas(true, pred)
```

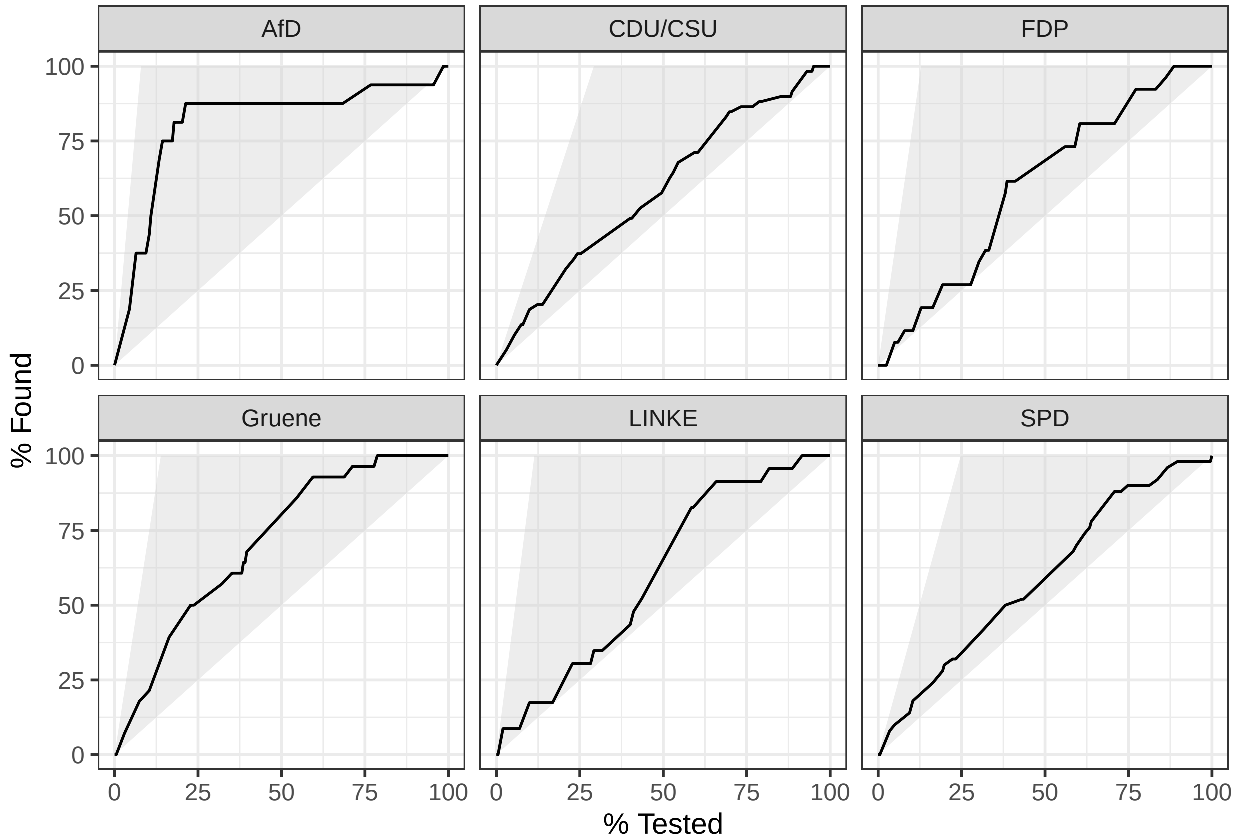
```
## Warning: While computing multiclass ‘precision()‘, some levels had no predicted events (i.e. ‘true_po
## Precision is undefined in this case, and those levels will be removed from the averaged result.
## Note that the following number of true events actually occured for each problematic event level:
## 'FDP': 26
## 'Gruene': 28
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 f_meas  macro          0.348
```

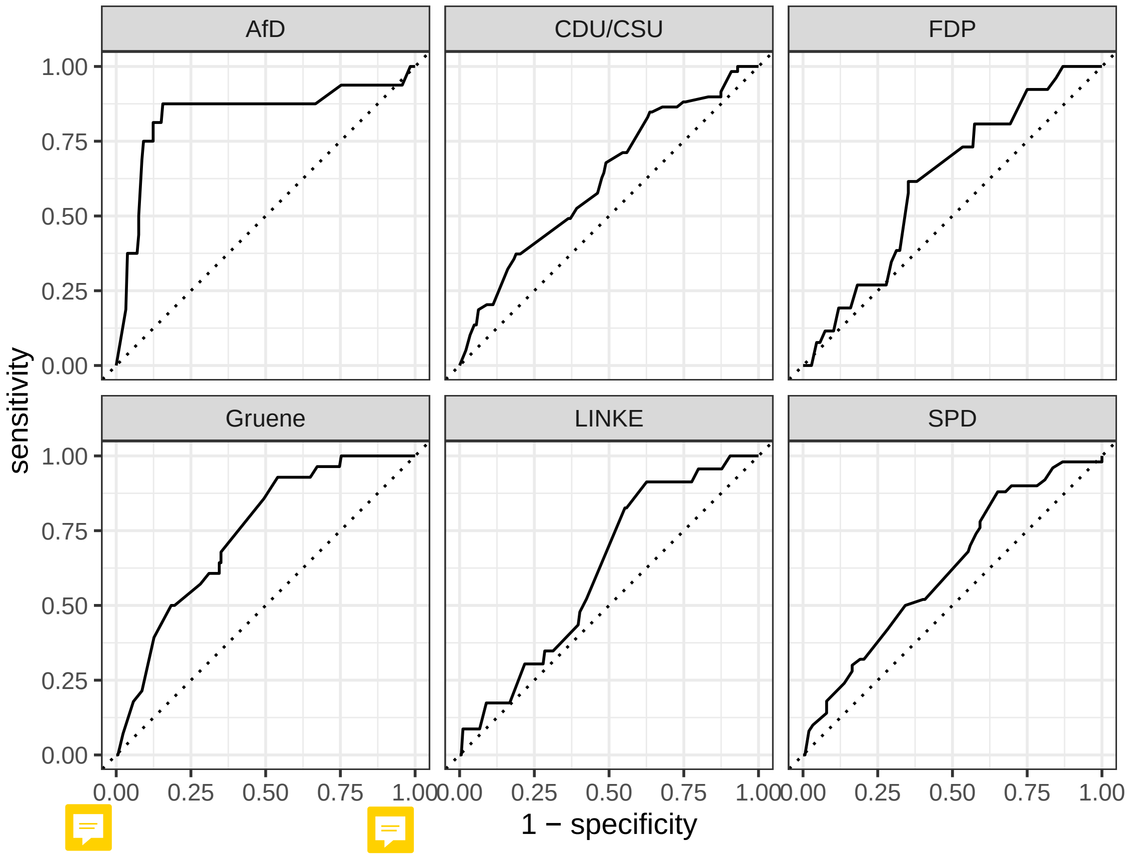
```r
preds_test %>%
  conf_mat(true, pred)
```

```
##           Truth
## Prediction FDP SPD CDU/CSU Gruene AfD LINKE
##    FDP       0   0       0      0   0     0
##    SPD       3  19      13     15   2     8
##    CDU/CSU  23  26      41     12   7    12
##    Gruene    0   0       0      0   0     0
##    AfD       0   2       5      0   7     1
##    LINKE     0   3       0      1   0     2
```

```r
predict(mm4m, test_data, type="prob") %>%
  bind_cols(test_data) %>%
  gain_curve(vote, FDP:LINKE) %>%
  autoplot()
```




```
predict(mm4m, test_data, type="prob") %>%
  bind_cols(test_data) %>%
  roc_curve(vote, FDP:LINKE) %>%
  autoplot()
```



3.1.3 Modeling according to compass

Here we will model using a hierarchical approach as described before.

From the plots we can see that the first model (hm1m1) seems to be working a bit better than the second one (hm2m1).

```
#level 1: left_wing or center, 1 right_wing
hm1m1 <- glm(right_wing ~ income*political_interest*egoposition_immigration +
               I(income^2) + I(political_interest^2) +
               I(egoposition_immigration^2), data=train_data, family=binomial)

hm1m1 <- stats::step(hm1m1)
```

```
## Start:  AIC=325.11
## right_wing ~ income * political_interest * egoposition_immigration +
##     I(income^2) + I(political_interest^2) + I(egoposition_immigration^2)
##
##                                                   Df Deviance    AIC
## - I(political_interest^2)                          1   303.25 323.25
## - I(income^2)                                      1   303.37 323.37
## - income:political_interest:egoposition_immigration  1   304.50 324.50
## - I(egoposition_immigration^2)                     1   305.06 325.06
## <none>                                                 303.11 325.11
##
```

```
## Step:  AIC=323.25
## right_wing ~ income + political_interest + egoposition_immigration +
##     I(income^2) + I(egoposition_immigration^2) + income:political_interest +
##     income:egoposition_immigration + political_interest:egoposition_immigration +
##     income:political_interest:egoposition_immigration
##
##                                                     Df Deviance    AIC
## - I(income^2)                                        1   303.48 321.48
## - income:political_interest:egoposition_immigration  1   304.62 322.62
## <none>                                                   303.25 323.25
## - I(egoposition_immigration^2)                       1   305.25 323.25
##
## Step:  AIC=321.48
## right_wing ~ income + political_interest + egoposition_immigration +
##     I(egoposition_immigration^2) + income:political_interest +
##     income:egoposition_immigration + political_interest:egoposition_immigration +
##     income:political_interest:egoposition_immigration
##
##                                                     Df Deviance    AIC
## - income:political_interest:egoposition_immigration  1   304.86 320.86
## <none>                                                   303.48 321.48
## - I(egoposition_immigration^2)                       1   305.55 321.55
##
## Step:  AIC=320.86
## right_wing ~ income + political_interest + egoposition_immigration +
##     I(egoposition_immigration^2) + income:political_interest +
##     income:egoposition_immigration + political_interest:egoposition_immigration
##
##                                              Df Deviance    AIC
## - income:egoposition_immigration              1   304.88 318.88
## - political_interest:egoposition_immigration  1   305.25 319.25
## - income:political_interest                   1   305.29 319.29
## - I(egoposition_immigration^2)                1   306.60 320.60
## <none>                                            304.86 320.86
##
## Step:  AIC=318.88
## right_wing ~ income + political_interest + egoposition_immigration +
##     I(egoposition_immigration^2) + income:political_interest +
##     political_interest:egoposition_immigration
##
##                                              Df Deviance    AIC
## - political_interest:egoposition_immigration  1   305.25 317.25
## - income:political_interest                   1   305.33 317.33
## - I(egoposition_immigration^2)                1   306.61 318.60
## <none>                                            304.88 318.88
##
## Step:  AIC=317.25
## right_wing ~ income + political_interest + egoposition_immigration +
##     I(egoposition_immigration^2) + income:political_interest
##
##                                Df Deviance    AIC
## - income:political_interest     1   305.62 315.62
## - I(egoposition_immigration^2)  1   307.09 317.09
## <none>                              305.25 317.25
```

```
## - egoposition_immigration       1   314.05 324.05
##
## Step:  AIC=315.62
## right_wing ~ income + political_interest + egoposition_immigration +
##     I(egoposition_immigration^2)
##
##                                Df Deviance    AIC
## - income                        1   305.63 313.63
## - political_interest            1   306.82 314.82
## - I(egoposition_immigration^2)  1   307.39 315.39
## <none>                              305.62 315.62
## - egoposition_immigration       1   314.26 322.26
##
## Step:  AIC=313.63
## right_wing ~ political_interest + egoposition_immigration + I(egoposition_immigration^2)
##
##                                Df Deviance    AIC
## - political_interest            1   306.82 312.83
## - I(egoposition_immigration^2)  1   307.39 313.39
## <none>                              305.63 313.63
## - egoposition_immigration       1   314.26 320.27
##
## Step:  AIC=312.83
## right_wing ~ egoposition_immigration + I(egoposition_immigration^2)
##
##                                Df Deviance    AIC
## - I(egoposition_immigration^2)  1   308.76 312.77
## <none>                              306.82 312.83
## - egoposition_immigration       1   315.71 319.71
##
## Step:  AIC=312.77
## right_wing ~ egoposition_immigration
##
##                           Df Deviance    AIC
## <none>                         308.77 312.77
## - egoposition_immigration  1   389.85 391.85
```

```
summary(hm1m1)
```

```
##
## Call:
## glm(formula = right_wing ~ egoposition_immigration, family = binomial,
##     data = train_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0361  -0.3072  -0.2359  -0.1385   3.2206
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -5.71807    0.48382 -11.819  < 2e-16 ***
## egoposition_immigration  0.53763    0.06641   8.095 5.72e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.85  on 797  degrees of freedom
## Residual deviance: 308.77  on 796  degrees of freedom
## AIC: 312.77
##
## Number of Fisher Scoring iterations: 6
```

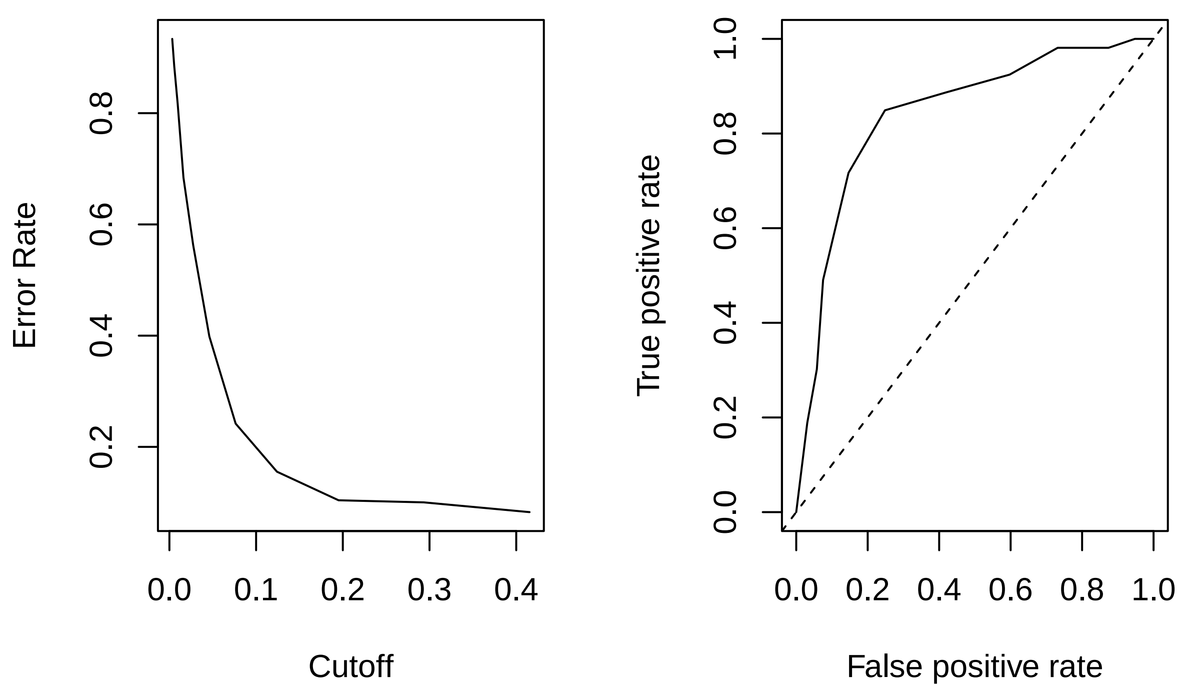
```
sum(ifelse(predict(hm1m1,type="response")>0.5,T,F) ==
      train_data$right_wing) / nrow(train_data)
```

```
## [1] 0.933584
```

```
dadesroc<-prediction(predict(hm1m1,type="response"),train_data$right_wing)
par(mfrow=c(1,2))
performance(dadesroc,"auc")
```

```
## A performance instance
##   'Area under the ROC curve'
```

```
plot(performance(dadesroc,"err"))
plot(performance(dadesroc,"tpr","fpr"))
abline(0,1,lty=2)
```



```r
coef(hm1m1)
```

```
##             (Intercept) egoposition_immigration
##              -5.7180669               0.5376297
```

```r
exp(coef(hm1m1)[2])
```

```
## egoposition_immigration
##                1.711944
```

```r
# Level 2
train_data_clear_party <- subset(train_data, right_wing == F)
hm2m1 <- glm(clear_party ~ income*political_interest*egoposition_immigration +
               I(income^2) + I(political_interest^2) +
               I(egoposition_immigration^2),
             data=train_data_clear_party,family=binomial)
hm2m1 <- stats::step(hm2m1)
```

```
## Start:  AIC=850.91
## clear_party ~ income * political_interest * egoposition_immigration +
##     I(income^2) + I(political_interest^2) + I(egoposition_immigration^2)
##
##                                                  Df Deviance    AIC
## - I(egoposition_immigration^2)                    1   828.91 848.91
## - I(political_interest^2)                         1   828.97 848.97
## - I(income^2)                                     1   829.92 849.92
## <none>                                                828.91 850.91
## - income:political_interest:egoposition_immigration  1   831.66 851.66
##
## Step:  AIC=848.91
## clear_party ~ income + political_interest + egoposition_immigration +
##     I(income^2) + I(political_interest^2) + income:political_interest +
##     income:egoposition_immigration + political_interest:egoposition_immigration +
##     income:political_interest:egoposition_immigration
##
##                                                  Df Deviance    AIC
## - I(political_interest^2)                         1   828.97 846.97
## - I(income^2)                                     1   829.92 847.92
## <none>                                                828.91 848.91
## - income:political_interest:egoposition_immigration  1   831.66 849.66
##
## Step:  AIC=846.97
## clear_party ~ income + political_interest + egoposition_immigration +
##     I(income^2) + income:political_interest + income:egoposition_immigration +
##     political_interest:egoposition_immigration + income:political_interest:egoposition_immigration
##
##                                                  Df Deviance    AIC
## - I(income^2)                                     1   829.98 845.98
## <none>                                                828.97 846.97
## - income:political_interest:egoposition_immigration  1   831.74 847.74
##
## Step:  AIC=845.98
```

```
## clear_party ~ income + political_interest + egoposition_immigration +
##     income:political_interest + income:egoposition_immigration +
##     political_interest:egoposition_immigration + income:political_interest:egoposition_immigration
##
##                                                     Df Deviance    AIC
## <none>                                                   829.98 845.98
## - income:political_interest:egoposition_immigration  1   832.95 846.95
```

```
summary(hm2m1)
```

```
##
## Call:
## glm(formula = clear_party ~ income + political_interest + egoposition_immigration +
##     income:political_interest + income:egoposition_immigration +
##     political_interest:egoposition_immigration + income:political_interest:egoposition_immigration,
##     family = binomial, data = train_data_clear_party)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1149  -0.8108  -0.6556   1.1986   2.2858
##
## Coefficients:
##                                                    Estimate Std. Error z value
## (Intercept)                                        -3.76259    2.13534  -1.762
## income                                              1.01779    0.73476   1.385
## political_interest                                  1.66385    0.74950   2.220
## egoposition_immigration                             0.68273    0.45332   1.506
## income:political_interest                          -0.44682    0.24925  -1.793
## income:egoposition_immigration                     -0.25184    0.16051  -1.569
## political_interest:egoposition_immigration         -0.36863    0.16929  -2.177
## income:political_interest:egoposition_immigration  0.09643    0.05736   1.681
##                                                    Pr(>|z|)
## (Intercept)                                          0.0781 .
## income                                               0.1660
## political_interest                                   0.0264 *
## egoposition_immigration                              0.1320
## income:political_interest                            0.0730 .
## income:egoposition_immigration                       0.1167
## political_interest:egoposition_immigration           0.0294 *
## income:political_interest:egoposition_immigration   0.0927 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 895.32  on 744  degrees of freedom
## Residual deviance: 829.98  on 737  degrees of freedom
## AIC: 845.98
##
## Number of Fisher Scoring iterations: 4
```

```
sum(ifelse(
  predict(
```

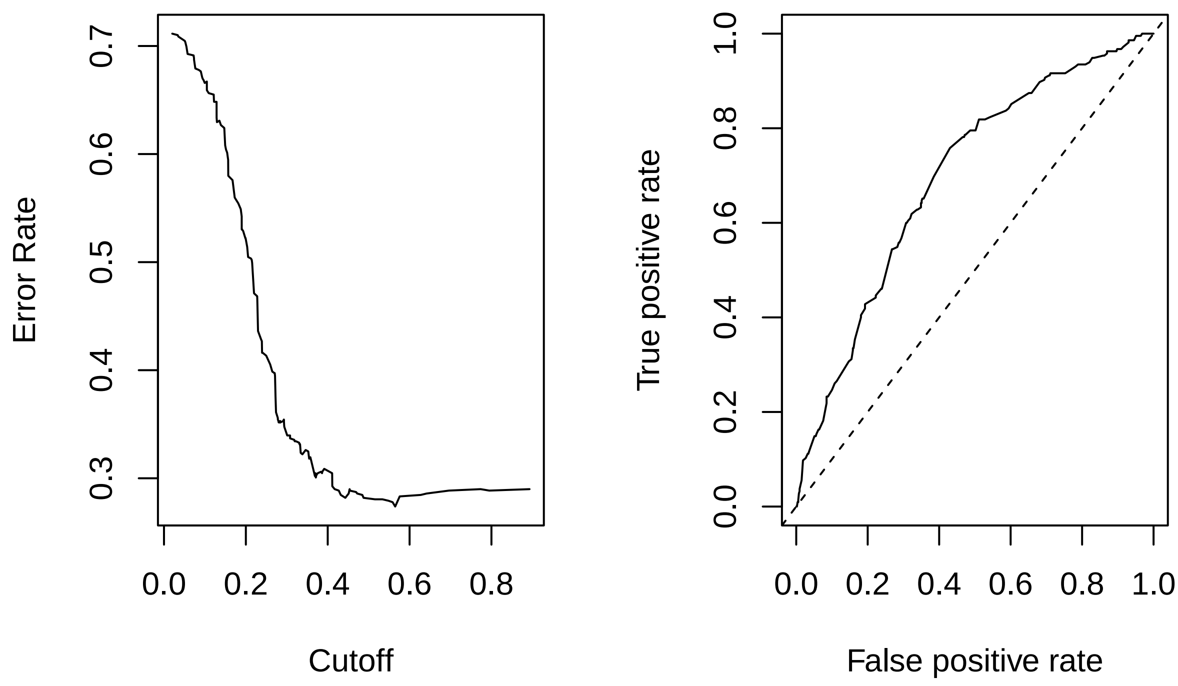
```r
    hm2m1,type="response")>0.5,T,F) == train_data_clear_party$clear_party) /
  nrow(train_data_clear_party)
```

```
## [1] 0.7194631
```

```r
dadesroc<-prediction(predict(hm2m1,type="response"),
                      train_data_clear_party$clear_party)
par(mfrow=c(1,2))
performance(dadesroc,"auc")
```

```
## A performance instance
##   'Area under the ROC curve'
```

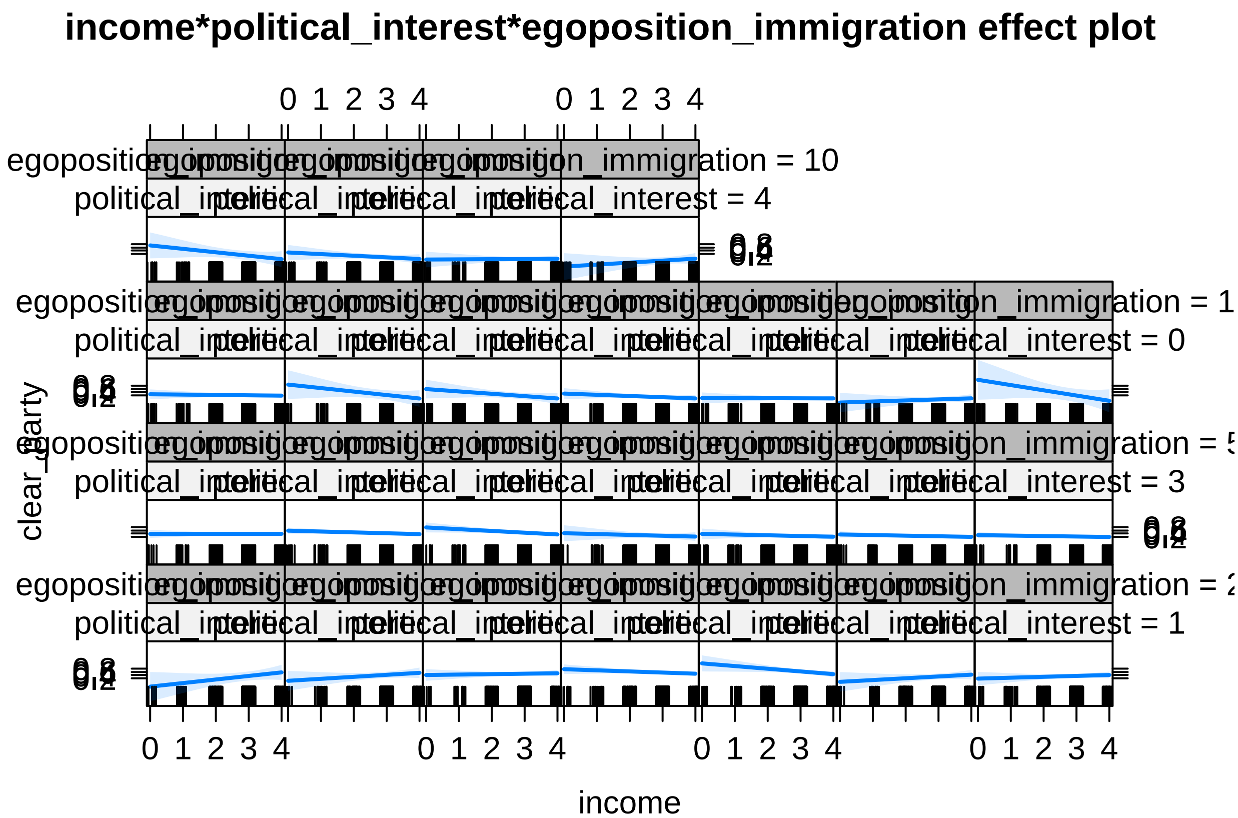
```r
plot(performance(dadesroc,"err"))
plot(performance(dadesroc,"tpr","fpr"))
abline(0,1,lty=2)
```




```r
AIC(hm2m1) + AIC(hm1m1)
```

```
## [1] 1158.745
```

```
plot(allEffects(hm2m1))
```



We get an accuracy of 64%.

```
right <- test_data[ifelse(predict(hm1m1, test_data, type="response") >
                              0.5, T, F), ]
left_center <- test_data[!ifelse(predict(hm1m1, test_data, type="response") >
                                     0.5, T, F), ]

left <- left_center[ifelse(predict(hm2m1, left_center, type="response") >
                               0.5, T, F), ]
center <- left_center[!ifelse(predict(hm2m1, left_center, type="response") >
                                  0.5, T, F), ]

# accuracy
(sum((right$right_wing == T)) + sum(left$clear_party == T &
                                      left$right_wing == F) +
  sum((center$clear_party == F))) / nrow(test_data)
```

```
## [1] 0.6435644
```

3.2 Modeling with new factors created

We select as best model the mmf1 model which is the following: vote ~ egoposition_factored + ostwest + gender.

Even though this model doesn't have the lowest Akaike, it is just above the one with the lowest one (mmf4), we choose it because it is simpler so it will be easier to explain.

```r
mmf0 <- multinom( vote ~ 1, data=train_data)
```

```
## # weights:  12 (5 variable)
## initial  value 1429.824056
## final  value 1341.119385
## converged
```

```r
summary(mmf0)
```

```
## Call:
## multinom(formula = vote ~ 1, data = train_data)
##
## Coefficients:
##         (Intercept)
## SPD       0.7691290
## CDU/CSU   0.8842028
## Gruene    0.1910530
## AfD      -0.5835871
## LINKE     0.0512871
##
## Std. Errors:
##         (Intercept)
## SPD       0.1241143
## CDU/CSU   0.1219595
## Gruene    0.1386432
## AfD       0.1714474
## LINKE     0.1432701
##
## Residual Deviance: 2682.239
## AIC: 2692.239
```

```r
mmf1 <- multinom(vote ~ egoposition_factored + ostwest + gender,
                 data=train_data, Hess=T)
```

```
## # weights:  36 (25 variable)
## initial  value 1429.824056
## iter  10 value 1259.245545
## iter  20 value 1238.632598
## iter  30 value 1237.306387
## final  value 1237.305978
## converged
```

```r
summary(mmf1) # Residual Deviance: 2474.612 , AIC: 2524.612
```

```
## Call:
## multinom(formula = vote ~ egoposition_factored + ostwest + gender,
##     data = train_data, Hess = T)
##
```

```
## Coefficients:
##         (Intercept) egoposition_factoredNeutral egoposition_factoredCon
## SPD       0.9480883                  -0.9082969             -0.74125277
## CDU/CSU   0.8785733                  -0.2886108             -0.02202811
## Gruene    0.5912086                  -1.4935088             -2.70586853
## AfD      -0.9454406                   0.3072299              2.52305202
## LINKE     1.1957158                  -2.2009876             -1.33566166
##         ostwestEast Germany genderFemale
## SPD             0.033957940    0.4896953
## CDU/CSU         0.068000044    0.1640554
## Gruene          0.004955595    0.6161539
## AfD            -0.817301196   -0.9529005
## LINKE          -0.734429488    0.2223909
##
## Std. Errors:
##         (Intercept) egoposition_factoredNeutral egoposition_factoredCon
## SPD       0.3288064                   0.2834010               0.3555903
## CDU/CSU   0.3246919                   0.2737108               0.3375728
## Gruene    0.3659746                   0.3346397               0.6549153
## AfD       0.5331645                   0.5775136               0.5291469
## LINKE     0.3444112                   0.4210263               0.4347073
##         ostwestEast Germany genderFemale
## SPD               0.3096768    0.2556062
## CDU/CSU           0.3023680    0.2501326
## Gruene            0.3516144    0.2893580
## AfD               0.3980410    0.4109790
## LINKE             0.3365035    0.2996852
##
## Residual Deviance: 2474.612
## AIC: 2524.612
```

```r
anova(mmf0, mmf1)
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: vote
##                                    Model Resid. df Resid. Dev   Test    Df
## 1                                      1      3985   2682.239
## 2 egoposition_factored + ostwest + gender      3965   2474.612 1 vs 2    20
##   LR stat. Pr(Chi)
## 1
## 2 207.6268       0
```

```r
mmf2 <- multinom( vote ~ political_interest_factored + income_factored +
                    egoposition_factored + gender + ostwest, data=train_data,
                  Hess=T)
```

```
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1256.516241
## iter  20 value 1222.153873
## iter  30 value 1219.299219
## iter  40 value 1219.118890
```

```
## final  value 1219.116058
## converged
```

```
summary(mmf2)
```

```
## Call:
## multinom(formula = vote ~ political_interest_factored + income_factored +
##     egoposition_factored + gender + ostwest, data = train_data,
##     Hess = T)
##
## Coefficients:
##         (Intercept) political_interest_factoredNeutral
## SPD      -2.0084363                          2.5236876
## CDU/CSU   0.2038051                          0.5047972
## Gruene   -0.4654136                          0.3323492
## AfD      -1.4040264                          0.4668861
## LINKE     1.0986145                          0.3720429
##         political_interest_factoredInterested income_factoredNeutral
## SPD                                 2.7120746              0.6834096
## CDU/CSU                             0.3044733              0.2352602
## Gruene                              0.6418106              0.2098160
## AfD                                 1.2131139             -0.3763003
## LINKE                               0.4432033              0.2058237
##         income_factoredSatisfied egoposition_factoredNeutral
## SPD                    0.3048870                  -0.9120213
## CDU/CSU                0.3920010                  -0.2918045
## Gruene                 0.5888457                  -1.4622199
## AfD                   -0.6321032                   0.3469935
## LINKE                 -0.4773899                  -2.2132634
##         egoposition_factoredCon genderFemale ostwestEast Germany
## SPD                -0.738706842    0.5006471          0.02135455
## CDU/CSU            -0.008520425    0.1368363          0.04608381
## Gruene             -2.631714867    0.6728621         -0.03267459
## AfD                 2.609426335   -0.9007533         -0.87685502
## LINKE              -1.385449489    0.2028389         -0.70889029
##
## Std. Errors:
##         (Intercept) political_interest_factoredNeutral
## SPD       1.2591835                          1.1137965
## CDU/CSU   0.7793709                          0.5675632
## Gruene    1.0309739                          0.7350568
## AfD       1.0884406                          0.8985394
## LINKE     0.8984423                          0.7001274
##         political_interest_factoredInterested income_factoredNeutral
## SPD                                 1.1017431              0.6920386
## CDU/CSU                             0.5473606              0.6504939
## Gruene                              0.7070641              0.8468870
## AfD                                 0.8497072              0.8518956
## LINKE                               0.6732203              0.7206908
##         income_factoredSatisfied egoposition_factoredNeutral
## SPD                    0.6428580                   0.2858107
## CDU/CSU                0.5935922                   0.2749445
## Gruene                 0.7730132                   0.3360813
## AfD                    0.7776515                   0.5806194
```

```
## LINKE                  0.6680486                   0.4231901
##         egoposition_factoredCon genderFemale ostwestEast Germany
## SPD                   0.3608973    0.2623484           0.3130797
## CDU/CSU               0.3400949    0.2555422           0.3038686
## Gruene                0.6567135    0.2962376           0.3539710
## AfD                   0.5358699    0.4234001           0.4057272
## LINKE                 0.4394372    0.3069011           0.3394001
##
## Residual Deviance: 2438.232
## AIC: 2528.232
```

```
mmf3 <- multinom( vote ~ egoposition_factored + gender + ostwest +
                    egoposition_factored * gender +
                    egoposition_factored * ostwest + gender * ostwest
                  + income_factored * egoposition_factored +
                    income_factored * gender +
                    political_interest_factored * income_factored  ,
                  data=train_data, Hess=T)
```

```
## # weights:  150 (120 variable)
## initial  value 1429.824056
## iter  10 value 1247.205441
## iter  20 value 1195.821345
## iter  30 value 1185.028906
## iter  40 value 1177.869962
## iter  50 value 1174.487149
## iter  60 value 1172.220430
## iter  70 value 1171.423538
## iter  80 value 1171.197135
## iter  90 value 1171.151657
## iter 100 value 1171.146112
## final  value 1171.146112
## stopped after 100 iterations
```

```
summary(mmf3)
```

```
## Warning in sqrt(diag(vc)): NaNs produced
```

```
## Call:
## multinom(formula = vote ~ egoposition_factored + gender + ostwest +
##     egoposition_factored * gender + egoposition_factored * ostwest +
##     gender * ostwest + income_factored * egoposition_factored +
##     income_factored * gender + political_interest_factored *
##     income_factored, data = train_data, Hess = T)
##
## Coefficients:
##         (Intercept) egoposition_factoredNeutral egoposition_factoredCon
## SPD       -8.168833                  -0.2378813               -23.29420
## CDU/CSU   11.069187                  -1.5114781               -11.58358
## Gruene     9.748463                 -22.4892427               -39.11650
## AfD      -19.489502                 -11.9675287                35.55084
## LINKE     -5.682832                 -24.7365085               -34.28480
```

```
##          genderFemale ostwestEast Germany income_factoredNeutral
## SPD       0.145906184          -0.1729950             -11.259829
## CDU/CSU  -1.583422460          -0.4941850              -8.876544
## Gruene   -0.002164518          -0.8694465             -19.881227
## AfD     -34.785273238          -1.7733036              21.059587
## LINKE    -1.680566071          -1.1013314               8.494625
##         income_factoredSatisfied political_interest_factoredNeutral
## SPD                     7.055230                         18.5053931
## CDU/CSU               -10.347252                          0.4580720
## Gruene                 -9.033071                          0.3148829
## AfD                    -7.790071                        -33.1495204
## LINKE                   6.292696                         19.6865375
##         political_interest_factoredInterested
## SPD                                  7.650866
## CDU/CSU                             -9.949011
## Gruene                             -10.883892
## AfD                                  2.488357
## LINKE                                7.417973
##         egoposition_factoredNeutral:genderFemale
## SPD                                    0.7846405
## CDU/CSU                                0.6875927
## Gruene                                 0.4241494
## AfD                                   31.1588108
## LINKE                                  0.3367951
##         egoposition_factoredCon:genderFemale
## SPD                                1.0471782
## CDU/CSU                            0.7344202
## Gruene                             1.3274252
## AfD                               32.6193436
## LINKE                              1.0141652
##         egoposition_factoredNeutral:ostwestEast Germany
## SPD                                         -0.07464234
## CDU/CSU                                      0.17142330
## Gruene                                      22.53902713
## AfD                                          1.06215221
## LINKE                                       -0.83864576
##         egoposition_factoredCon:ostwestEast Germany
## SPD                                      -0.5968525
## CDU/CSU                                   0.7229911
## Gruene                                   18.0791822
## AfD                                       0.3516883
## LINKE                                    -1.1897853
##         genderFemale:ostwestEast Germany
## SPD                            0.6252113
## CDU/CSU                        0.7199122
## Gruene                         1.0194961
## AfD                            0.3478177
## LINKE                          1.0700811
##         egoposition_factoredNeutral:income_factoredNeutral
## SPD                                              -0.6725781
## CDU/CSU                                           1.5416770
## Gruene                                           -2.1871735
## AfD                                              10.9515977
## LINKE                                            23.9967528
```

```
##         egoposition_factoredCon:income_factoredNeutral
## SPD                                           22.71926
## CDU/CSU                                       10.31962
## Gruene                                       -23.90790
## AfD                                          -35.97523
## LINKE                                         32.89494
##         egoposition_factoredNeutral:income_factoredSatisfied
## SPD                                                -1.0717747
## CDU/CSU                                             0.7002444
## Gruene                                             -1.4485153
## AfD                                                11.3388020
## LINKE                                              22.5926207
##         egoposition_factoredCon:income_factoredSatisfied
## SPD                                             22.40892
## CDU/CSU                                         10.88806
## Gruene                                          18.50038
## AfD                                            -33.44729
## LINKE                                           33.06129
##         genderFemale:income_factoredNeutral
## SPD                              -1.7771091
## CDU/CSU                           0.1718434
## Gruene                           -1.4449946
## AfD                               2.4428833
## LINKE                            -0.5596065
##         genderFemale:income_factoredSatisfied
## SPD                                -0.2717601
## CDU/CSU                             0.9784226
## Gruene                             -0.2476790
## AfD                                 1.3959200
## LINKE                               1.1097729
##         income_factoredNeutral:political_interest_factoredNeutral
## SPD                                                     3.0747121
## CDU/CSU                                                -0.8815798
## Gruene                                                 11.5729522
## AfD                                                    32.4538725
## LINKE                                                 -20.0972581
##         income_factoredSatisfied:political_interest_factoredNeutral
## SPD                                                     -16.37653726
## CDU/CSU                                                   0.41152096
## Gruene                                                    0.05031779
## AfD                                                      59.69325383
## LINKE                                                   -18.87461827
##         income_factoredNeutral:political_interest_factoredInterested
## SPD                                                         14.026303
## CDU/CSU                                                      9.180998
## Gruene                                                      22.831835
## AfD                                                         -2.496416
## LINKE                                                       -7.349500
##         income_factoredSatisfied:political_interest_factoredInterested
## SPD                                                           -5.163191
## CDU/CSU                                                       10.743516
## Gruene                                                        11.762786
## AfD                                                           24.869673
## LINKE                                                         -6.578461
```

```
##
## Std. Errors:
##         (Intercept) egoposition_factoredNeutral egoposition_factoredCon
## SPD       9.1459152                   1.6671085               1.5722429
## CDU/CSU   9.9722727                   1.5761833              27.2859061
## Gruene   10.0329940                   0.9455473              65.4019860
## AfD       0.9803917                   0.8259155               0.7180220
## LINKE     9.1396189                   0.6251537               0.6119319
##         genderFemale ostwestEast Germany income_factoredNeutral
## SPD        1.5751943           0.6231251              9.1289389
## CDU/CSU    1.4469934           0.6107357             10.0165536
## Gruene     1.9103816           0.6650416              9.0644680
## AfD        0.4955816           0.9998404              0.9983473
## LINKE      1.7193929           0.6356867              9.1687002
##         income_factoredSatisfied political_interest_factoredNeutral
## SPD                    9.1389633                         18.1859945
## CDU/CSU                9.9758925                         17.3668634
## Gruene                10.0437847                         17.3939448
## AfD                    0.3805037                          0.8937809
## LINKE                  9.1489695                         18.2008748
##         political_interest_factoredInterested
## SPD                                 9.1284253
## CDU/CSU                             9.9638550
## Gruene                             10.0304576
## AfD                                 0.8538461
## LINKE                               9.1227501
##         egoposition_factoredNeutral:genderFemale
## SPD                                    0.6115624
## CDU/CSU                                0.5933455
## Gruene                                 0.7391947
## AfD                                    0.8937502
## LINKE                                  0.9285324
##         egoposition_factoredCon:genderFemale
## SPD                                0.7917342
## CDU/CSU                            0.7418310
## Gruene                             1.4276709
## AfD                                0.6819370
## LINKE                              0.9837642
##         egoposition_factoredNeutral:ostwestEast Germany
## SPD                                           0.7651213
## CDU/CSU                                       0.7222981
## Gruene                                        0.9455474
## AfD                                           1.3382749
## LINKE                                         0.9747521
##         egoposition_factoredCon:ostwestEast Germany
## SPD                                       0.8395071
## CDU/CSU                                   0.8274612
## Gruene                                   65.4531749
## AfD                                       1.2210525
## LINKE                                     1.0175454
##         genderFemale:ostwestEast Germany
## SPD                            0.6655757
## CDU/CSU                        0.6464290
## Gruene                         0.7775060
```

```
## AfD                          1.0125272
## LINKE                        0.7294851
##         egoposition_factoredNeutral:income_factoredNeutral
## SPD                                                1.6943889
## CDU/CSU                                            1.6608751
## Gruene                                             2.2204033
## AfD                                                0.9724652
## LINKE                                              0.6949664
##         egoposition_factoredCon:income_factoredNeutral
## SPD                                            1.5720887
## CDU/CSU                                       27.2890211
## Gruene                                               NaN
## AfD                                            0.8818472
## LINKE                                          0.6516166
##         egoposition_factoredNeutral:income_factoredSatisfied
## SPD                                                  1.5476046
## CDU/CSU                                              1.4968895
## Gruene                                               1.8808575
## AfD                                                  0.7041909
## LINKE                                                0.5151013
##         egoposition_factoredCon:income_factoredSatisfied
## SPD                                              1.5269631
## CDU/CSU                                         27.2824004
## Gruene                                         130.8514722
## AfD                                              0.6839380
## LINKE                                            0.5392675
##         genderFemale:income_factoredNeutral
## SPD                               1.5978911
## CDU/CSU                           1.4898398
## Gruene                            1.9921755
## AfD                               0.6721371
## LINKE                             1.7923201
##         genderFemale:income_factoredSatisfied
## SPD                                 1.4910228
## CDU/CSU                             1.3743498
## Gruene                              1.8313704
## AfD                                 0.6135231
## LINKE                               1.6795478
##         income_factoredNeutral:political_interest_factoredNeutral
## SPD                                                      18.1802428
## CDU/CSU                                                  17.3956263
## Gruene                                                   18.2740016
## AfD                                                       0.9856723
## LINKE                                                    18.2192782
##         income_factoredSatisfied:political_interest_factoredNeutral
## SPD                                                        18.1874328
## CDU/CSU                                                    17.3723685
## Gruene                                                     17.4051469
## AfD                                                         0.3438449
## LINKE                                                      18.2100492
##         income_factoredNeutral:political_interest_factoredInterested
## SPD                                                          9.1177844
## CDU/CSU                                                     10.0172194
## Gruene                                                       9.0549154
```

```
## AfD                                                          0.8864779
## LINKE                                                        9.1589720
##         income_factoredSatisfied:political_interest_factoredInterested
## SPD                                                            9.1293400
## CDU/CSU                                                        9.9720136
## Gruene                                                        10.0463953
## AfD                                                            0.2652807
## LINKE                                                          9.1383763
##
## Residual Deviance: 2342.292
## AIC: 2582.292
```

```
mmf4 <- stats::step(mmf3)
```

```
## Start:  AIC=2582.29
## vote ~ egoposition_factored + gender + ostwest + egoposition_factored *
##     gender + egoposition_factored * ostwest + gender * ostwest +
##     income_factored * egoposition_factored + income_factored *
##     gender + political_interest_factored * income_factored
##
## trying - egoposition_factored:gender
## # weights:  138 (110 variable)
## initial  value 1429.824056
## iter  10 value 1248.751628
## iter  20 value 1197.222004
## iter  30 value 1186.540197
## iter  40 value 1181.611943
## iter  50 value 1178.754024
## iter  60 value 1176.458631
## iter  70 value 1175.786940
## iter  80 value 1175.682226
## iter  90 value 1175.656322
## iter 100 value 1175.654636
## final  value 1175.654636
## stopped after 100 iterations
## trying - egoposition_factored:ostwest
## # weights:  138 (110 variable)
## initial  value 1429.824056
## iter  10 value 1266.284608
## iter  20 value 1204.116420
## iter  30 value 1194.890395
## iter  40 value 1187.652394
## iter  50 value 1184.331284
## iter  60 value 1182.532279
## iter  70 value 1181.773833
## iter  80 value 1181.607623
## iter  90 value 1181.576472
## iter 100 value 1181.572802
## final  value 1181.572802
## stopped after 100 iterations
## trying - gender:ostwest
## # weights:  144 (115 variable)
## initial  value 1429.824056
## iter  10 value 1248.930363
```

```
## iter  20 value 1196.942978
## iter  30 value 1185.492437
## iter  40 value 1178.386726
## iter  50 value 1175.292064
## iter  60 value 1173.316423
## iter  70 value 1172.627470
## iter  80 value 1172.493427
## iter  90 value 1172.461280
## iter 100 value 1172.459611
## final  value 1172.459611
## stopped after 100 iterations
## trying - egoposition_factored:income_factored
## # weights:  126 (100 variable)
## initial  value 1429.824056
## iter  10 value 1263.537561
## iter  20 value 1203.532857
## iter  30 value 1195.335362
## iter  40 value 1191.335615
## iter  50 value 1189.812988
## iter  60 value 1189.015116
## iter  70 value 1188.390489
## iter  80 value 1188.288980
## iter  90 value 1188.284916
## final  value 1188.284698
## converged
## trying - gender:income_factored
## # weights:  138 (110 variable)
## initial  value 1429.824056
## iter  10 value 1247.594966
## iter  20 value 1198.428209
## iter  30 value 1188.813286
## iter  40 value 1182.902614
## iter  50 value 1180.056738
## iter  60 value 1177.946742
## iter  70 value 1177.458683
## iter  80 value 1177.338658
## iter  90 value 1177.328074
## final  value 1177.327733
## converged
## trying - income_factored:political_interest_factored
## # weights:  126 (100 variable)
## initial  value 1429.824056
## iter  10 value 1252.159604
## iter  20 value 1198.441022
## iter  30 value 1190.079145
## iter  40 value 1185.921947
## iter  50 value 1184.297032
## iter  60 value 1183.823928
## iter  70 value 1183.555131
## iter  80 value 1183.503292
## iter  90 value 1183.498220
## final  value 1183.497580
## converged
##                                                Df      AIC
```

```
## - income_factored:political_interest_factored 100 2566.995
## - egoposition_factored:gender                  110 2571.309
## - gender:income_factored                       110 2574.655
## - gender:ostwest                               115 2574.919
## - egoposition_factored:income_factored         100 2576.569
## <none>                                         120 2582.292
## - egoposition_factored:ostwest                 110 2583.146
## # weights:  126 (100 variable)
## initial  value 1429.824056
## iter  10 value 1252.159604
## iter  20 value 1198.441022
## iter  30 value 1190.079145
## iter  40 value 1185.921947
## iter  50 value 1184.297032
## iter  60 value 1183.823928
## iter  70 value 1183.555131
## iter  80 value 1183.503292
## iter  90 value 1183.498220
## final  value 1183.497580
## converged
##
## Step:  AIC=2567
## vote ~ egoposition_factored + gender + ostwest + income_factored +
##     political_interest_factored + egoposition_factored:gender +
##     egoposition_factored:ostwest + gender:ostwest + egoposition_factored:income_factored +
##     gender:income_factored
##
## trying - political_interest_factored
## # weights:  114 (90 variable)
## initial  value 1429.824056
## iter  10 value 1258.776217
## iter  20 value 1208.328248
## iter  30 value 1200.863251
## iter  40 value 1196.178873
## iter  50 value 1194.904413
## iter  60 value 1194.433222
## iter  70 value 1194.254738
## iter  80 value 1194.249082
## iter  90 value 1194.246147
## final  value 1194.246101
## converged
## trying - egoposition_factored:gender
## # weights:  114 (90 variable)
## initial  value 1429.824056
## iter  10 value 1253.484390
## iter  20 value 1199.342684
## iter  30 value 1191.889133
## iter  40 value 1189.658169
## iter  50 value 1188.463675
## iter  60 value 1188.106519
## iter  70 value 1187.920507
## iter  80 value 1187.898642
## iter  90 value 1187.896167
## final  value 1187.896048
```

```
## converged
## trying - egoposition_factored:ostwest
## # weights:  114 (90 variable)
## initial  value 1429.824056
## iter  10 value 1254.560917
## iter  20 value 1208.041702
## iter  30 value 1199.849452
## iter  40 value 1195.900526
## iter  50 value 1194.404872
## iter  60 value 1194.019888
## iter  70 value 1193.812928
## iter  80 value 1193.785908
## iter  90 value 1193.783038
## final  value 1193.782973
## converged
## trying - gender:ostwest
## # weights:  120 (95 variable)
## initial  value 1429.824056
## iter  10 value 1252.424970
## iter  20 value 1199.712864
## iter  30 value 1191.443624
## iter  40 value 1187.260438
## iter  50 value 1185.659071
## iter  60 value 1185.308793
## iter  70 value 1185.079836
## iter  80 value 1185.049854
## iter  90 value 1185.046271
## final  value 1185.046196
## converged
## trying - egoposition_factored:income_factored
## # weights:  102 (80 variable)
## initial  value 1429.824056
## iter  10 value 1250.023121
## iter  20 value 1207.257389
## iter  30 value 1200.997993
## iter  40 value 1198.473877
## iter  50 value 1197.669283
## iter  60 value 1197.426716
## iter  70 value 1197.390130
## iter  80 value 1197.385140
## final  value 1197.385007
## converged
## trying - gender:income_factored
## # weights:  114 (90 variable)
## initial  value 1429.824056
## iter  10 value 1253.928584
## iter  20 value 1203.015961
## iter  30 value 1195.229494
## iter  40 value 1192.255557
## iter  50 value 1191.122131
## iter  60 value 1190.668859
## iter  70 value 1190.574159
## iter  80 value 1190.558727
## final  value 1190.557536
```

```
## converged
##                                          Df      AIC
## - egoposition_factored:income_factored  80 2554.770
## - egoposition_factored:gender           90 2555.792
## - gender:ostwest                        95 2560.092
## - gender:income_factored                90 2561.115
## <none>                                 100 2566.995
## - egoposition_factored:ostwest          90 2567.566
## - political_interest_factored           90 2568.492
## # weights:  102 (80 variable)
## initial  value 1429.824056
## iter  10 value 1250.023121
## iter  20 value 1207.257389
## iter  30 value 1200.997993
## iter  40 value 1198.473877
## iter  50 value 1197.669283
## iter  60 value 1197.426716
## iter  70 value 1197.390130
## iter  80 value 1197.385140
## final  value 1197.385007
## converged
##
## Step:  AIC=2554.77
## vote ~ egoposition_factored + gender + ostwest + income_factored +
##     political_interest_factored + egoposition_factored:gender +
##     egoposition_factored:ostwest + gender:ostwest + gender:income_factored
##
## trying - political_interest_factored
## # weights:  90 (70 variable)
## initial  value 1429.824056
## iter  10 value 1259.933780
## iter  20 value 1216.816429
## iter  30 value 1210.641434
## iter  40 value 1208.070819
## iter  50 value 1207.619459
## iter  60 value 1207.484453
## iter  70 value 1207.465461
## final  value 1207.464932
## converged
## trying - egoposition_factored:gender
## # weights:  90 (70 variable)
## initial  value 1429.824056
## iter  10 value 1250.178997
## iter  20 value 1207.983172
## iter  30 value 1202.464469
## iter  40 value 1201.264040
## iter  50 value 1200.894139
## iter  60 value 1200.742610
## iter  70 value 1200.725264
## final  value 1200.724788
## converged
## trying - egoposition_factored:ostwest
## # weights:  90 (70 variable)
## initial  value 1429.824056
```

```
## iter  10 value 1249.728686
## iter  20 value 1215.238960
## iter  30 value 1210.054817
## iter  40 value 1208.657568
## iter  50 value 1208.202759
## iter  60 value 1208.089040
## iter  70 value 1208.073886
## final  value 1208.072400
## converged
## trying - gender:ostwest
## # weights:  96 (75 variable)
## initial  value 1429.824056
## iter  10 value 1246.860597
## iter  20 value 1208.027010
## iter  30 value 1202.063599
## iter  40 value 1199.852995
## iter  50 value 1199.086729
## iter  60 value 1198.898589
## iter  70 value 1198.877950
## final  value 1198.876103
## converged
## trying - gender:income_factored
## # weights:  90 (70 variable)
## initial  value 1429.824056
## iter  10 value 1246.996396
## iter  20 value 1211.558833
## iter  30 value 1205.740825
## iter  40 value 1204.021495
## iter  50 value 1203.300600
## iter  60 value 1203.188879
## iter  70 value 1203.183089
## final  value 1203.183040
## converged
##                               Df      AIC
## - egoposition_factored:gender  70 2541.450
## - gender:income_factored       70 2546.366
## - gender:ostwest               75 2547.752
## <none>                         80 2554.770
## - political_interest_factored  70 2554.930
## - egoposition_factored:ostwest 70 2556.145
## # weights:  90 (70 variable)
## initial  value 1429.824056
## iter  10 value 1250.178997
## iter  20 value 1207.983172
## iter  30 value 1202.464469
## iter  40 value 1201.264040
## iter  50 value 1200.894139
## iter  60 value 1200.742610
## iter  70 value 1200.725264
## final  value 1200.724788
## converged
##
## Step:  AIC=2541.45
## vote ~ egoposition_factored + gender + ostwest + income_factored +
```

```
##     political_interest_factored + egoposition_factored:ostwest +
##     gender:ostwest + gender:income_factored
##
## trying - political_interest_factored
## # weights:  78 (60 variable)
## initial  value 1429.824056
## iter  10 value 1261.058976
## iter  20 value 1218.037757
## iter  30 value 1212.723887
## iter  40 value 1211.729724
## iter  50 value 1211.404321
## iter  60 value 1211.339034
## final  value 1211.338089
## converged
## trying - egoposition_factored:ostwest
## # weights:  78 (60 variable)
## initial  value 1429.824056
## iter  10 value 1250.782670
## iter  20 value 1216.060846
## iter  30 value 1212.000280
## iter  40 value 1211.501762
## iter  50 value 1211.370861
## iter  60 value 1211.319941
## iter  70 value 1211.317713
## final  value 1211.317661
## converged
## trying - gender:ostwest
## # weights:  84 (65 variable)
## initial  value 1429.824056
## iter  10 value 1247.826936
## iter  20 value 1208.380564
## iter  30 value 1203.439262
## iter  40 value 1202.619551
## iter  50 value 1202.248060
## iter  60 value 1202.132607
## iter  70 value 1202.125634
## final  value 1202.125443
## converged
## trying - gender:income_factored
## # weights:  78 (60 variable)
## initial  value 1429.824056
## iter  10 value 1248.007401
## iter  20 value 1211.955072
## iter  30 value 1207.682509
## iter  40 value 1206.840225
## iter  50 value 1206.539579
## iter  60 value 1206.488069
## final  value 1206.487321
## converged
##                                Df      AIC
## - gender:income_factored       60 2532.975
## - gender:ostwest               65 2534.251
## <none>                         70 2541.450
## - egoposition_factored:ostwest 60 2542.635
```

```
## - political_interest_factored  60 2542.676
## # weights:  78 (60 variable)
## initial  value 1429.824056
## iter  10 value 1248.007401
## iter  20 value 1211.955072
## iter  30 value 1207.682509
## iter  40 value 1206.840225
## iter  50 value 1206.539579
## iter  60 value 1206.488069
## final  value 1206.487321
## converged
##
## Step:  AIC=2532.97
## vote ~ egoposition_factored + gender + ostwest + income_factored +
##     political_interest_factored + egoposition_factored:ostwest +
##     gender:ostwest
##
## trying - income_factored
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1260.407042
## iter  20 value 1222.916303
## iter  30 value 1216.079904
## iter  40 value 1215.409393
## iter  50 value 1215.202429
## iter  60 value 1215.196069
## final  value 1215.196039
## converged
## trying - political_interest_factored
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1259.807006
## iter  20 value 1222.647727
## iter  30 value 1217.415275
## iter  40 value 1216.771212
## iter  50 value 1216.583603
## final  value 1216.579545
## converged
## trying - egoposition_factored:ostwest
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1251.541180
## iter  20 value 1221.205867
## iter  30 value 1217.794809
## iter  40 value 1217.459720
## final  value 1217.454229
## converged
## trying - gender:ostwest
## # weights:  72 (55 variable)
## initial  value 1429.824056
## iter  10 value 1251.582088
## iter  20 value 1213.490171
## iter  30 value 1209.384846
## iter  40 value 1208.587821
```

```
## iter  50 value 1208.247615
## iter  60 value 1208.220400
## final  value 1208.219986
## converged
##                                Df      AIC
## - gender:ostwest               55 2526.440
## - income_factored              50 2530.392
## <none>                         60 2532.975
## - political_interest_factored  50 2533.159
## - egoposition_factored:ostwest 50 2534.908
## # weights:  72 (55 variable)
## initial  value 1429.824056
## iter  10 value 1251.582088
## iter  20 value 1213.490171
## iter  30 value 1209.384846
## iter  40 value 1208.587821
## iter  50 value 1208.247615
## iter  60 value 1208.220400
## final  value 1208.219986
## converged
##
## Step:  AIC=2526.44
## vote ~ egoposition_factored + gender + ostwest + income_factored +
##     political_interest_factored + egoposition_factored:ostwest
##
## trying - gender
## # weights:  66 (50 variable)
## initial  value 1429.824056
## iter  10 value 1260.071895
## iter  20 value 1222.861660
## iter  30 value 1218.713604
## iter  40 value 1218.180370
## iter  50 value 1218.004979
## iter  60 value 1218.001842
## iter  60 value 1218.001831
## iter  60 value 1218.001831
## final  value 1218.001831
## converged
## trying - income_factored
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1265.328754
## iter  20 value 1223.699591
## iter  30 value 1217.538630
## iter  40 value 1216.968914
## iter  50 value 1216.849741
## final  value 1216.847844
## converged
## trying - political_interest_factored
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1259.564352
## iter  20 value 1223.545963
## iter  30 value 1219.189975
```

```
## iter  40 value 1218.594477
## iter  50 value 1218.451852
## final  value 1218.450134
## converged
## trying - egoposition_factored:ostwest
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1256.516241
## iter  20 value 1222.153873
## iter  30 value 1219.299219
## iter  40 value 1219.118890
## final  value 1219.116058
## converged
##                               Df      AIC
## - income_factored             45 2523.696
## <none>                        55 2526.440
## - political_interest_factored  45 2526.900
## - egoposition_factored:ostwest 45 2528.232
## - gender                      50 2536.004
## # weights:  60 (45 variable)
## initial  value 1429.824056
## iter  10 value 1265.328754
## iter  20 value 1223.699591
## iter  30 value 1217.538630
## iter  40 value 1216.968914
## iter  50 value 1216.849741
## final  value 1216.847844
## converged
##
## Step:  AIC=2523.7
## vote ~ egoposition_factored + gender + ostwest + political_interest_factored +
##     egoposition_factored:ostwest
##
## trying - gender
## # weights:  54 (40 variable)
## initial  value 1429.824056
## iter  10 value 1272.193671
## iter  20 value 1232.001348
## iter  30 value 1227.602083
## iter  40 value 1227.089318
## iter  50 value 1227.053445
## final  value 1227.053310
## converged
## trying - political_interest_factored
## # weights:  48 (35 variable)
## initial  value 1429.824056
## iter  10 value 1266.238718
## iter  20 value 1231.609758
## iter  30 value 1227.159040
## iter  40 value 1226.741160
## iter  50 value 1226.728653
## iter  50 value 1226.728642
## iter  50 value 1226.728642
## final  value 1226.728642
```

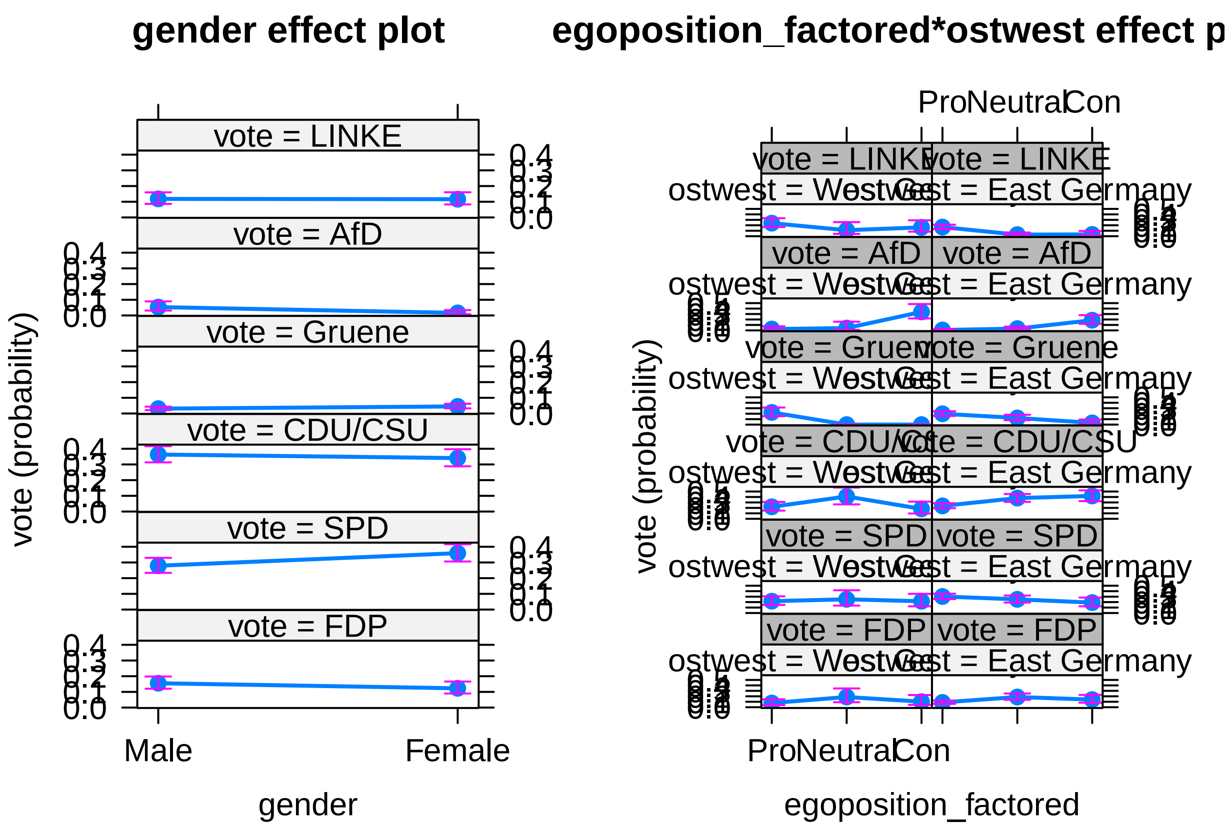
```
## converged
## trying - egoposition_factored:ostwest
## # weights:  48 (35 variable)
## initial  value 1429.824056
## iter  10 value 1276.269611
## iter  20 value 1232.515965
## iter  30 value 1227.497127
## iter  40 value 1227.384477
## iter  40 value 1227.384470
## iter  40 value 1227.384470
## final  value 1227.384470
## converged
##                              Df      AIC
## - political_interest_factored  35 2523.457
## <none>                       45 2523.696
## - egoposition_factored:ostwest 35 2524.769
## - gender                     40 2534.107
## # weights:  48 (35 variable)
## initial  value 1429.824056
## iter  10 value 1266.238718
## iter  20 value 1231.609758
## iter  30 value 1227.159040
## iter  40 value 1226.741160
## iter  50 value 1226.728653
## iter  50 value 1226.728642
## iter  50 value 1226.728642
## final  value 1226.728642
## converged
##
## Step:  AIC=2523.46
## vote ~ egoposition_factored + gender + ostwest + egoposition_factored:ostwest
##
## trying - gender
## # weights:  42 (30 variable)
## initial  value 1429.824056
## iter  10 value 1263.342951
## iter  20 value 1238.787760
## iter  30 value 1237.062579
## iter  40 value 1236.852177
## final  value 1236.849893
## converged
## trying - egoposition_factored:ostwest
## # weights:  36 (25 variable)
## initial  value 1429.824056
## iter  10 value 1259.245545
## iter  20 value 1238.632598
## iter  30 value 1237.306387
## final  value 1237.305978
## converged
##                              Df      AIC
## <none>                       35 2523.457
## - egoposition_factored:ostwest 25 2524.612
## - gender                     30 2533.700
```

```r
summary(mmf4) # vote ~ egoposition_factored + gender + ostwest +
```

```
## Call:
## multinom(formula = vote ~ egoposition_factored + gender + ostwest +
##     egoposition_factored:ostwest, data = train_data, Hess = T)
##
## Coefficients:
##         (Intercept) egoposition_factoredNeutral egoposition_factoredCon
## SPD       0.8207585                  -0.7480036              -0.2979200
## CDU/CSU   0.9807185                  -0.2745095              -0.4785381
## Gruene    0.7933178                 -15.2332171             -14.6737731
## AfD      -0.6792630                  -0.3597128               2.3163566
## LINKE     1.0420190                  -1.6742616              -0.6827649
##         genderFemale ostwestEast Germany
## SPD        0.4858914          0.19493769
## CDU/CSU    0.1690080         -0.06536663
## Gruene     0.6109501         -0.25732096
## AfD       -0.9555017         -1.29213500
## LINKE      0.2159550         -0.49787889
##         egoposition_factoredNeutral:ostwestEast Germany
## SPD                                         -0.19824431
## CDU/CSU                                     -0.00922736
## Gruene                                      13.99039752
## AfD                                          1.01880230
## LINKE                                       -0.83667603
##         egoposition_factoredCon:ostwestEast Germany
## SPD                                     -0.6273732
## CDU/CSU                                  0.5866957
## Gruene                                  12.3664220
## AfD                                      0.3876703
## LINKE                                   -1.4249231
##
## Std. Errors:
##         (Intercept) egoposition_factoredNeutral egoposition_factoredCon
## SPD       0.4276897                   0.6521504               0.6794603
## CDU/CSU   0.4247385                   0.6156315               0.6928924
## Gruene    0.4328527                   0.1792237               0.3395382
## AfD       0.6884204                   1.0523349               0.8493879
## LINKE     0.4277163                   0.7478429               0.7005731
##         genderFemale ostwestEast Germany
## SPD        0.2558474           0.4621539
## CDU/CSU    0.2503449           0.4645955
## Gruene     0.2897087           0.4660726
## AfD        0.4115668           0.9100920
## LINKE      0.2998446           0.4666753
##         egoposition_factoredNeutral:ostwestEast Germany
## SPD                                           0.7232480
## CDU/CSU                                       0.6863881
## Gruene                                        0.1792220
## AfD                                           1.2838715
## LINKE                                         0.9172870
##         egoposition_factoredCon:ostwestEast Germany
## SPD                                      0.7994419
```

```
## CDU/CSU                                              0.7937314
## Gruene                                               0.3395356
## AfD                                                  1.1034881
## LINKE                                                0.9765942
##
## Residual Deviance: 2453.457
## AIC: 2523.457
```

```
#political_interest_factored + egoposition_factored:ostwest
#Residual Deviance: 2453.457
#AIC: 2523.457

plot(allEffects(mmf4))
```



3.2.1 Metrics

We report the ROC AUC measure as it takes into account both the type-I and type-II errors. Furthermore, it is insensitive to imbalanced datasets.

From the ROC curves below we can see that the model is not performing well. We get a recall of 24% and a precision of 33% for the train data and a recall of 24% and and 33% for the test data.

```
summary(mmf1)
```

```
## Call:
```

```
## multinom(formula = vote ~ egoposition_factored + ostwest + gender,
##     data = train_data, Hess = T)
##
## Coefficients:
##         (Intercept) egoposition_factoredNeutral egoposition_factoredCon
## SPD       0.9480883                  -0.9082969             -0.74125277
## CDU/CSU   0.8785733                  -0.2886108             -0.02202811
## Gruene    0.5912086                  -1.4935088             -2.70586853
## AfD      -0.9454406                   0.3072299              2.52305202
## LINKE     1.1957158                  -2.2009876             -1.33566166
##         ostwestEast Germany genderFemale
## SPD             0.033957940    0.4896953
## CDU/CSU         0.068000044    0.1640554
## Gruene          0.004955595    0.6161539
## AfD            -0.817301196   -0.9529005
## LINKE          -0.734429488    0.2223909
##
## Std. Errors:
##         (Intercept) egoposition_factoredNeutral egoposition_factoredCon
## SPD       0.3288064                   0.2834010               0.3555903
## CDU/CSU   0.3246919                   0.2737108               0.3375728
## Gruene    0.3659746                   0.3346397               0.6549153
## AfD       0.5331645                   0.5775136               0.5291469
## LINKE     0.3444112                   0.4210263               0.4347073
##         ostwestEast Germany genderFemale
## SPD               0.3096768    0.2556062
## CDU/CSU           0.3023680    0.2501326
## Gruene            0.3516144    0.2893580
## AfD               0.3980410    0.4109790
## LINKE             0.3365035    0.2996852
##
## Residual Deviance: 2474.612
## AIC: 2524.612
```

```r
sum(predict(mmf1, type="class") == train_data$vote) / nrow(train_data)
```

```
## [1] 0.3295739
```

```r
sum(predict(mmf1, test_data, type="class") == test_data$vote) / nrow(test_data)
```

```
## [1] 0.3217822
```

```r
preds_train <- tibble(
  pred = predict(mmf1, type="class"), true = train_data$vote
  )

preds_train <- preds_train  %>%
  mutate(
         true = fct_relevel(true, levels(pred)))

preds_test <- tibble(pred = predict(mmf1, test_data, type="class"),
                     true = test_data$vote)
```

```r
preds_test <- preds_test %>%
    mutate(
    true = fct_relevel(true, levels(pred)))

preds_train %>%
  recall(true, pred)
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 recall  macro          0.240
```

```r
preds_train %>%
  precision(true, pred)
```

```
## Warning: While computing multiclass `precision()`, some levels had no predicted events (i.e. `true_p
## Precision is undefined in this case, and those levels will be removed from the averaged result.
## Note that the following number of true events actually occured for each problematic event level:
## 'FDP': 95
## 'Gruene': 115

## # A tibble: 1 x 3
##   .metric   .estimator .estimate
##   <chr>     <chr>          <dbl>
## 1 precision macro          0.328
```

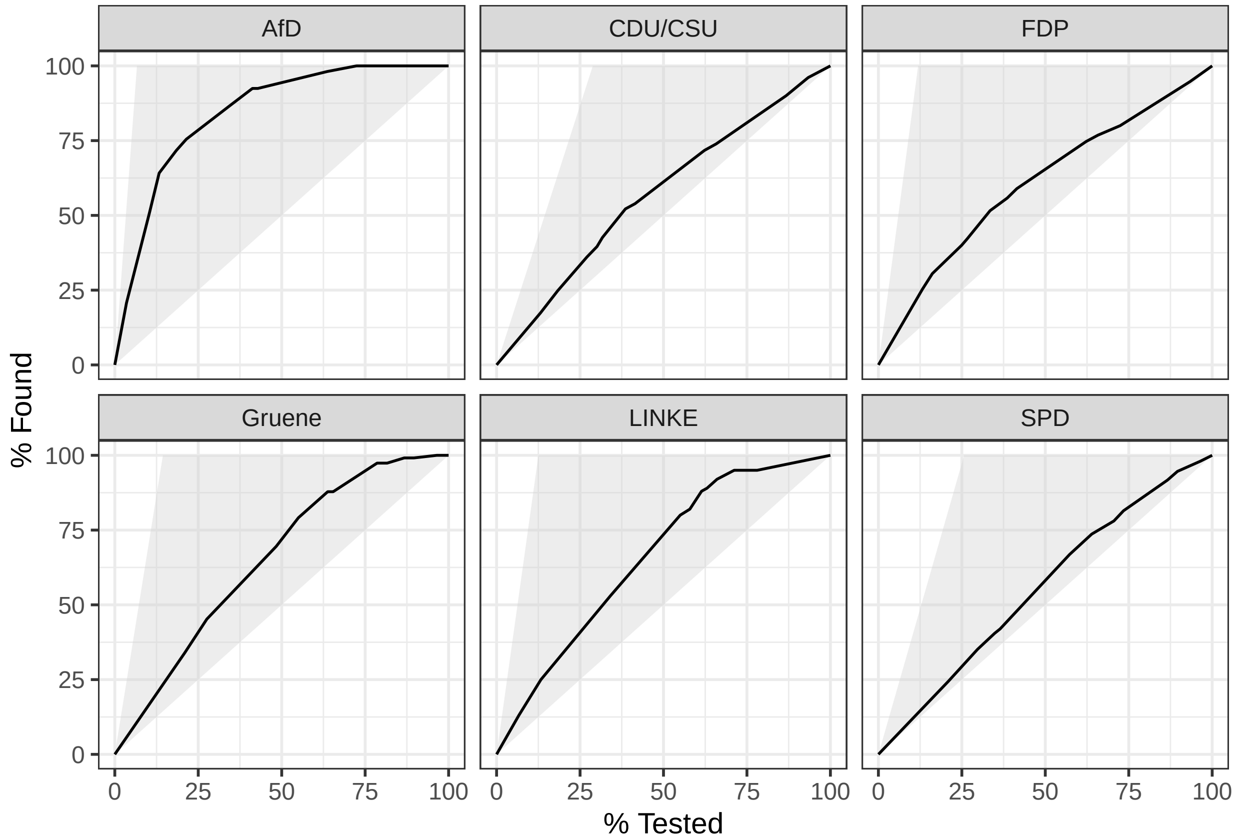
```r
preds_train %>%
  f_meas(true, pred)
```

```
## Warning: While computing multiclass `precision()`, some levels had no predicted events (i.e. `true_p
## Precision is undefined in this case, and those levels will be removed from the averaged result.
## Note that the following number of true events actually occured for each problematic event level:
## 'FDP': 95
## 'Gruene': 115

## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 f_meas  macro          0.318
```

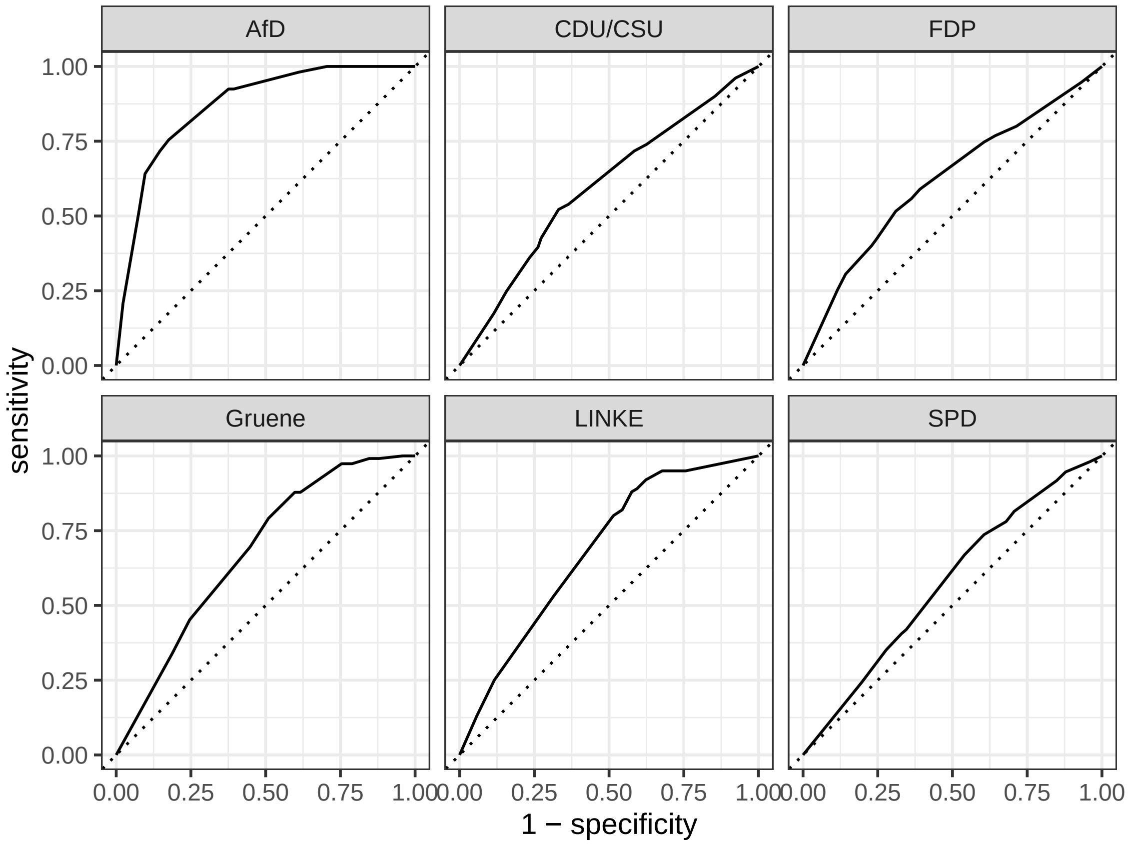
```r
preds_train %>%
  conf_mat(true, pred)
```

```
##           Truth
## Prediction FDP SPD CDU/CSU Gruene AfD LINKE
##    FDP       0   0       0      0   0     0
##    SPD      34 115      87     80   3    67
##    CDU/CSU  56  77     124     24  36    14
##    Gruene    0   0       0      0   0     0
##    AfD       2   4       5      0  11     6
##    LINKE     3   9      14     11   3    13
```

```
predict(mmf1, type="prob") %>%
  bind_cols(train_data) %>%
  gain_curve(vote, FDP:LINKE) %>%
  autoplot()
```




```
predict(mmf1, type="prob") %>%
  bind_cols(train_data) %>%
  roc_curve(vote, FDP:LINKE) %>%
  autoplot()
```




```
predict(mmf1, type="prob") %>%
  bind_cols(train_data) %>%
  roc_auc(vote, FDP:LINKE)
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 roc_auc hand_till      0.683
```

```
preds_test %>%
  recall(true, pred)
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 recall  macro          0.241
```

```
preds_test %>%
  precision(true, pred)
```

```
## Warning: While computing multiclass ‘precision()‘, some levels had no predicted events (i.e. ‘true_po
## Precision is undefined in this case, and those levels will be removed from the averaged result.
## Note that the following number of true events actually occured for each problematic event level:
## ’FDP’: 26
## ’Gruene’: 28
```

```
## # A tibble: 1 x 3
##   .metric   .estimator .estimate
##   <chr>     <chr>          <dbl>
## 1 precision macro          0.330
```

```
preds_test %>%
  f_meas(true, pred)
```

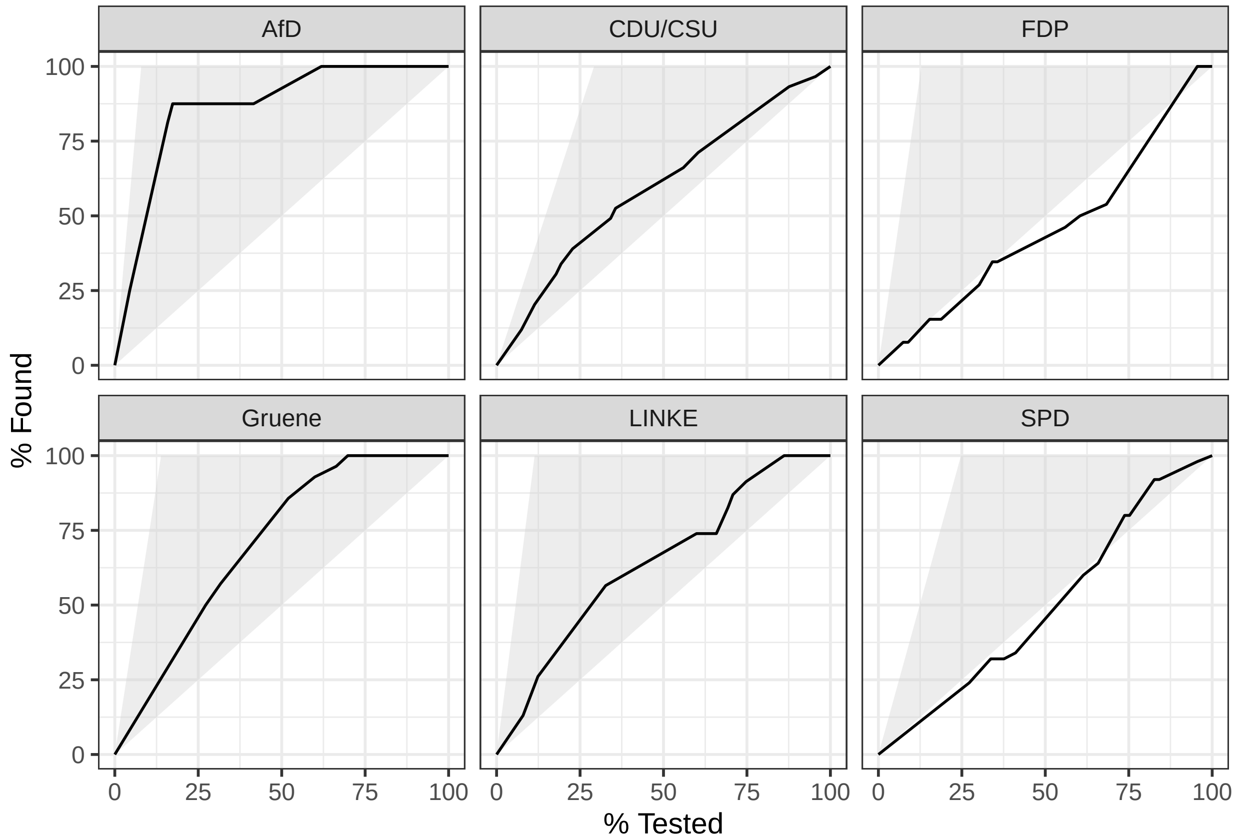
```
## Warning: While computing multiclass `precision()`, some levels had no predicted events (i.e. `true_po
## Precision is undefined in this case, and those levels will be removed from the averaged result.
## Note that the following number of true events actually occured for each problematic event level:
## 'FDP': 26
## 'Gruene': 28
```

```
## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 f_meas  macro          0.324
```

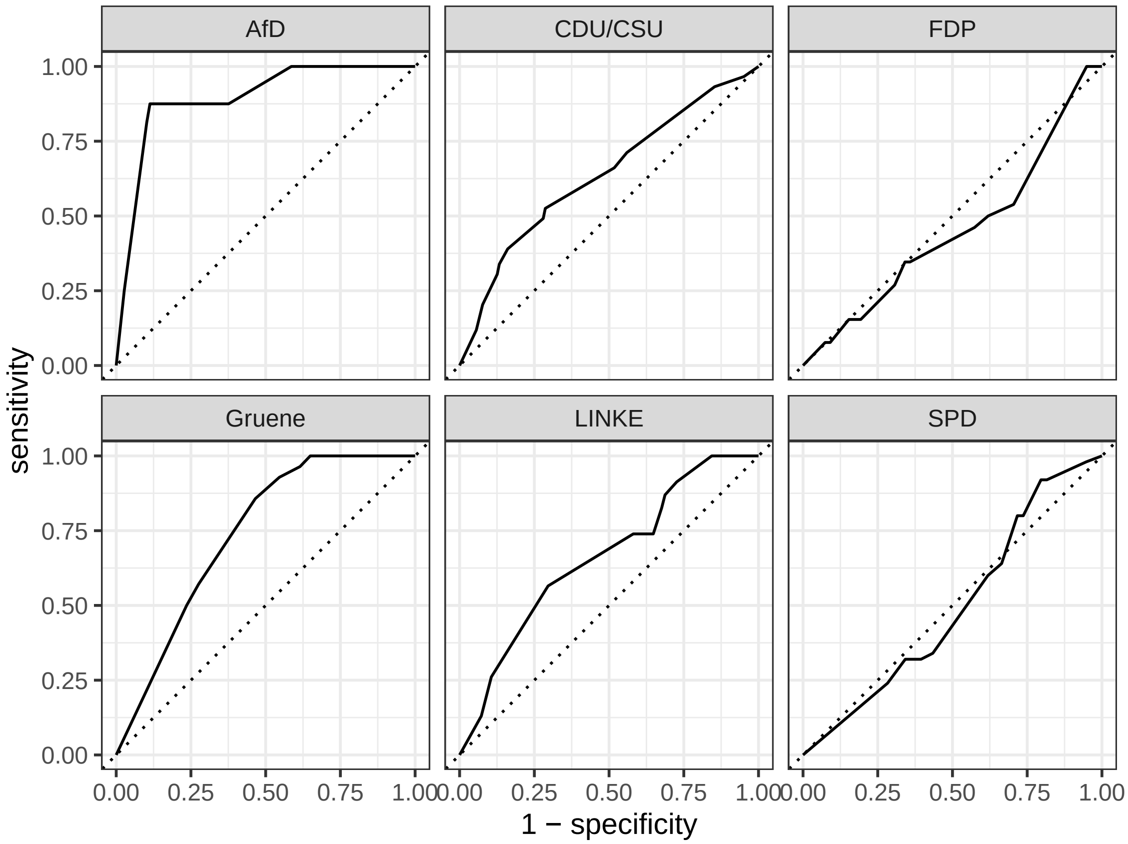
```
preds_test %>%
  conf_mat(true, pred)
```

```
##           Truth
## Prediction FDP SPD CDU/CSU Gruene AfD LINKE
##    FDP       0   0       0      0   0     0
##    SPD      15  27      23     24   2    14
##    CDU/CSU   9  14      31      2  10     6
##    Gruene    0   0       0      0   0     0
##    AfD       1   1       3      0   4     0
##    LINKE     1   8       2      2   0     3
```

```
predict(mmf1, test_data, type="prob") %>%
  bind_cols(test_data) %>%
  gain_curve(vote, FDP:LINKE) %>%
  autoplot()
```




```
predict(mmf1, test_data, type="prob") %>%
  bind_cols(test_data) %>%
  roc_curve(vote, FDP:LINKE) %>%
  autoplot()
```



3.2.2 Modeling according to compass

In this section we do a hierarchical model with the factors with the levels explained before.

```
hm0m2 <- glm(right_wing ~ 1, data=train_data, family=binomial)
summary(hm0m2)
```

```
##
## Call:
## glm(formula = right_wing ~ 1, family = binomial, data = train_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.3707  -0.3707  -0.3707  -0.3707   2.3289
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.6431     0.1422  -18.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.85  on 797  degrees of freedom
## Residual deviance: 389.85  on 797  degrees of freedom
```

```
## AIC: 391.85
##
## Number of Fisher Scoring iterations: 5
```

```
hm1m2 <- glm(right_wing ~ political_interest_factored+income_factored,
             family = "binomial", data = train_data)
summary(hm1m2)
```

```
##
## Call:
## glm(formula = right_wing ~ political_interest_factored + income_factored,
##     family = "binomial", data = train_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6069  -0.3755  -0.3755  -0.2743   2.5686
##
## Coefficients:
##                                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             -1.5984     0.7574  -2.110   0.0348 *
## political_interest_factoredNeutral     -0.8544     0.6970  -1.226   0.2203
## political_interest_factoredInterested  -0.2101     0.6422  -0.327   0.7436
## income_factoredNeutral                 -0.4398     0.6188  -0.711   0.4772
## income_factoredSatisfied               -0.8084     0.5694  -1.420   0.1557
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.85  on 797  degrees of freedom
## Residual deviance: 383.84  on 793  degrees of freedom
## AIC: 393.84
##
## Number of Fisher Scoring iterations: 5
```

```
hm2m2 <- stats::step(hm1m2)
```

```
## Start:  AIC=393.84
## right_wing ~ political_interest_factored + income_factored
##
##                               Df Deviance    AIC
## - income_factored              2   386.31 392.31
## - political_interest_factored  2   387.63 393.63
## <none>                             383.84 393.84
##
## Step:  AIC=392.31
## right_wing ~ political_interest_factored
##
##                               Df Deviance    AIC
## - political_interest_factored  2   389.85 391.85
## <none>                             386.31 392.31
##
## Step:  AIC=391.85
## right_wing ~ 1
```

```r
summary(hm2m2)
```

```
##
## Call:
## glm(formula = right_wing ~ 1, family = "binomial", data = train_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.3707  -0.3707  -0.3707  -0.3707   2.3289
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.6431     0.1422  -18.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.85  on 797  degrees of freedom
## Residual deviance: 389.85  on 797  degrees of freedom
## AIC: 391.85
##
## Number of Fisher Scoring iterations: 5
```

```r
hm3m2 <- glm(right_wing ~ egoposition_factored*ostwest + egoposition_factored +
               ostwest + egoposition_factored*gender + ostwest*gender,
             data=train_data, family=binomial)
summary(hm3m2)
```

```
##
## Call:
## glm(formula = right_wing ~ egoposition_factored * ostwest + egoposition_factored +
##     ostwest + egoposition_factored * gender + ostwest * gender,
##     family = binomial, data = train_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.01718  -0.35204  -0.19098  -0.00005   3.03906
##
## Coefficients:
##                                                Estimate Std. Error z value
## (Intercept)                                     -2.8134     0.5944  -4.733
## egoposition_factoredNeutral                      0.2338     0.9618   0.243
## egoposition_factoredCon                          2.4241     0.7058   3.435
## ostwestEast Germany                             -1.1817     0.8323  -1.420
## genderFemale                                   -15.6216   692.2058  -0.023
## egoposition_factoredNeutral:ostwestEast Germany   1.0113     1.1809   0.856
## egoposition_factoredCon:ostwestEast Germany       0.6527     0.9599   0.680
## egoposition_factoredNeutral:genderFemale         14.4855   692.2066   0.021
## egoposition_factoredCon:genderFemale             15.1200   692.2058   0.022
## ostwestEast Germany:genderFemale                 -0.7219     0.7925  -0.911
##                                                Pr(>|z|)
```

```
## (Intercept)                                      2.21e-06 ***
## egoposition_factoredNeutral                      0.807933
## egoposition_factoredCon                          0.000594 ***
## ostwestEast Germany                              0.155680
## genderFemale                                     0.981995
## egoposition_factoredNeutral:ostwestEast Germany 0.391811
## egoposition_factoredCon:ostwestEast Germany     0.496536
## egoposition_factoredNeutral:genderFemale        0.983304
## egoposition_factoredCon:genderFemale            0.982573
## ostwestEast Germany:genderFemale                0.362346
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.85  on 797  degrees of freedom
## Residual deviance: 282.55  on 788  degrees of freedom
## AIC: 302.55
##
## Number of Fisher Scoring iterations: 18
```

```
hm4m2 <- stats::step(hm3m2)
```

```
## Start:  AIC=302.55
## right_wing ~ egoposition_factored * ostwest + egoposition_factored +
##     ostwest + egoposition_factored * gender + ostwest * gender
##
##                                Df Deviance    AIC
## - egoposition_factored:ostwest  2   283.32 299.32
## - ostwest:gender                1   283.38 301.38
## <none>                              282.55 302.55
## - egoposition_factored:gender   2   286.64 302.64
##
## Step:  AIC=299.32
## right_wing ~ egoposition_factored + ostwest + gender + egoposition_factored:gender +
##     ostwest:gender
##
##                               Df Deviance    AIC
## - ostwest:gender               1   284.03 298.03
## - egoposition_factored:gender  2   287.18 299.18
## <none>                             283.32 299.32
##
## Step:  AIC=298.03
## right_wing ~ egoposition_factored + ostwest + gender + egoposition_factored:gender
##
##                               Df Deviance    AIC
## <none>                             284.03 298.03
## - egoposition_factored:gender  2   288.06 298.06
## - ostwest                      1   289.25 301.25
```

```
summary(hm4m2)
```

```
##
```

```
## Call:
## glm(formula = right_wing ~ egoposition_factored + ostwest + gender +
##     egoposition_factored:gender, family = binomial, data = train_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.07184  -0.32777  -0.20988  -0.00007   3.02542
##
## Coefficients:
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                               -3.0555     0.4589  -6.658 2.78e-11
## egoposition_factoredNeutral                0.9075     0.5541   1.638   0.1015
## egoposition_factoredCon                    2.8021     0.4780   5.862 4.59e-09
## ostwestEast Germany                       -0.7490     0.3230  -2.319   0.0204
## genderFemale                             -15.9835   719.3046  -0.022   0.9823
## egoposition_factoredNeutral:genderFemale  14.3143   719.3054   0.020   0.9841
## egoposition_factoredCon:genderFemale      15.1016   719.3047   0.021   0.9832
##
## (Intercept)                              ***
## egoposition_factoredNeutral
## egoposition_factoredCon                  ***
## ostwestEast Germany                      *
## genderFemale
## egoposition_factoredNeutral:genderFemale
## egoposition_factoredCon:genderFemale
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.85  on 797  degrees of freedom
## Residual deviance: 284.03  on 791  degrees of freedom
## AIC: 298.03
##
## Number of Fisher Scoring iterations: 18
```

```
 AIC(hm4m2, hm3m2)
```

```
##       df      AIC
## hm4m2  7 298.0261
## hm3m2 10 302.5481
```

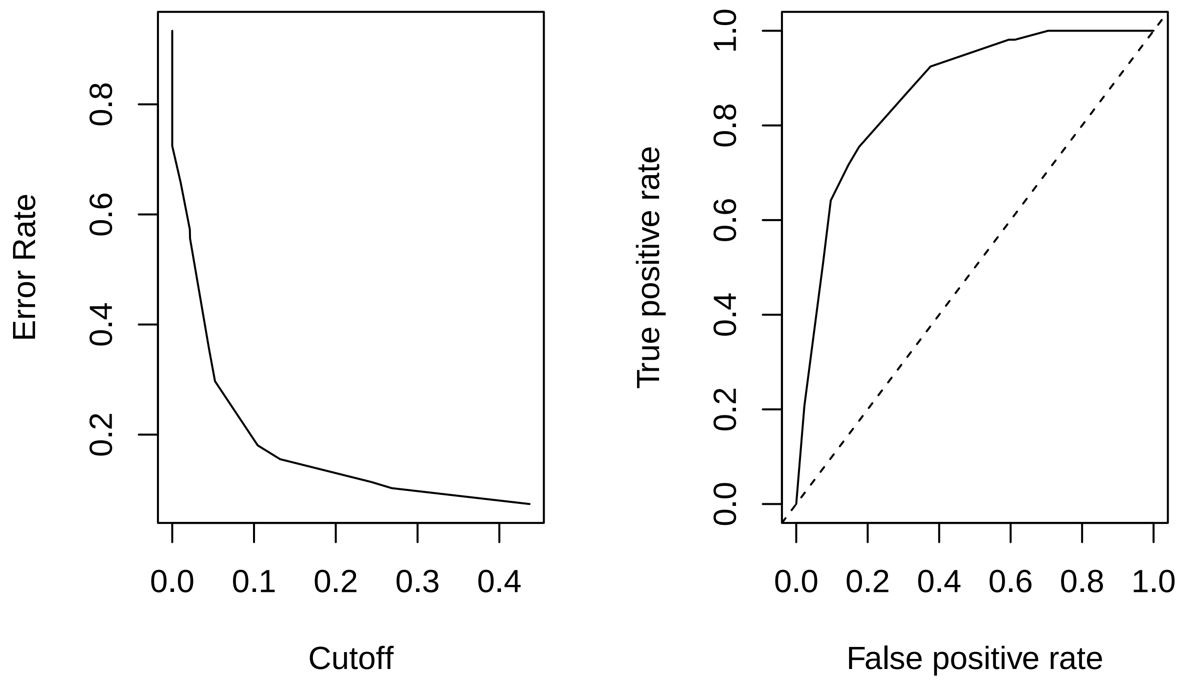
```
sum(ifelse(predict(hm4m2,type="response")>0.5,T,F) ==
      train_data$right_wing) / nrow(train_data)
```

```
## [1] 0.933584
```

```
dadesroc<-prediction(predict(hm4m2,type="response"),train_data$right_wing)
par(mfrow=c(1,2))
performance(dadesroc,"auc")
```

```
## A performance instance
##   'Area under the ROC curve'
```

```r
plot(performance(dadesroc,"err"))
plot(performance(dadesroc,"tpr","fpr"))
abline(0,1,lty=2)
```




```r
coef(hm4m2)
```

```
##                             (Intercept)
##                              -3.0555396
##           egoposition_factoredNeutral
##                               0.9075346
##               egoposition_factoredCon
##                               2.8020703
##                   ostwestEast Germany
##                              -0.7490227
##                          genderFemale
##                             -15.9835464
## egoposition_factoredNeutral:genderFemale
##                              14.3143291
##     egoposition_factoredCon:genderFemale
##                              15.1016195
```

```r
exp(coef(hm4m2)[2])
```

```
## egoposition_factoredNeutral
##                    2.478205
```

```r
train_data_clear_party <- subset(train_data, right_wing == F)
hm5m2 <- glm(clear_party ~ egoposition_factored+political_interest+income,
             family = "binomial", data = train_data_clear_party)
summary(hm5m2)
```

```
##
## Call:
## glm(formula = clear_party ~ egoposition_factored + political_interest +
##     income, family = "binomial", data = train_data_clear_party)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2552  -0.9841  -0.5592   1.2979   2.1408
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   0.03888    0.42268   0.092   0.9267
## egoposition_factoredNeutral -1.33847    0.22173  -6.036 1.58e-09 ***
## egoposition_factoredCon     -1.51703    0.30403  -4.990 6.05e-07 ***
## political_interest            0.03557    0.10330   0.344   0.7306
## income                       -0.19448    0.11031  -1.763   0.0779 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 895.32  on 744  degrees of freedom
## Residual deviance: 831.53  on 740  degrees of freedom
## AIC: 841.53
##
## Number of Fisher Scoring iterations: 4
```

```r
hm6m2 <- stats::step(hm5m2)
```

```
## Start:  AIC=841.53
## clear_party ~ egoposition_factored + political_interest + income
##
##                        Df Deviance    AIC
## - political_interest    1   831.65 839.65
## <none>                      831.53 841.53
## - income                1   834.62 842.62
## - egoposition_factored  2   893.26 899.26
##
## Step:  AIC=839.65
## clear_party ~ egoposition_factored + income
##
##                        Df Deviance    AIC
## <none>                      831.65 839.65
## - income                1   834.63 840.63
## - egoposition_factored  2   893.78 897.78
```

```r
hm7m2 <- glm(clear_party ~ egoposition_factored*ostwest +
               egoposition_factored*gender + ostwest*gender +
               political_interest_factored +
               political_interest_factored * ostwest + ostwest + gender +
               egoposition_factored +
               egoposition_factored*political_interest_factored +
               political_interest_factored * gender,
             data=train_data_clear_party,family=binomial)
hm8m2 <- stats::step(hm7m2)
```

```
## Start:  AIC=846.17
## clear_party ~ egoposition_factored * ostwest + egoposition_factored *
##     gender + ostwest * gender + political_interest_factored +
##     political_interest_factored * ostwest + ostwest + gender +
##     egoposition_factored + egoposition_factored * political_interest_factored +
##     political_interest_factored * gender
##
##                                                    Df Deviance    AIC
## - egoposition_factored:gender                       2   806.18 842.18
## - gender:political_interest_factored                2   807.57 843.57
## - ostwest:political_interest_factored               2   808.74 844.74
## - ostwest:gender                                    1   807.98 845.98
## <none>                                                  806.17 846.17
## - egoposition_factored:ostwest                      2   811.37 847.37
## - egoposition_factored:political_interest_factored  4   819.29 851.29
##
## Step:  AIC=842.18
## clear_party ~ egoposition_factored + ostwest + gender + political_interest_factored +
##     egoposition_factored:ostwest + ostwest:gender + ostwest:political_interest_factored +
##     egoposition_factored:political_interest_factored + gender:political_interest_factored
##
##                                                    Df Deviance    AIC
## - gender:political_interest_factored                2   807.61 839.61
## - ostwest:political_interest_factored               2   808.80 840.80
## - ostwest:gender                                    1   808.04 842.04
## <none>                                                  806.18 842.18
## - egoposition_factored:ostwest                      2   811.39 843.39
## - egoposition_factored:political_interest_factored  4   819.31 847.31
##
## Step:  AIC=839.61
## clear_party ~ egoposition_factored + ostwest + gender + political_interest_factored +
##     egoposition_factored:ostwest + ostwest:gender + ostwest:political_interest_factored +
##     egoposition_factored:political_interest_factored
##
##                                                    Df Deviance    AIC
## - ostwest:gender                                    1   809.60 839.60
## <none>                                                  807.61 839.61
## - egoposition_factored:ostwest                      2   812.56 840.56
## - ostwest:political_interest_factored               2   813.03 841.03
## - egoposition_factored:political_interest_factored  4   819.55 843.55
##
## Step:  AIC=839.6
## clear_party ~ egoposition_factored + ostwest + gender + political_interest_factored +
```

```
##     egoposition_factored:ostwest + ostwest:political_interest_factored +
##     egoposition_factored:political_interest_factored
##
##                                                   Df Deviance    AIC
## - gender                                           1   810.74 838.74
## <none>                                                 809.60 839.60
## - egoposition_factored:ostwest                     2   814.27 840.27
## - ostwest:political_interest_factored              2   814.63 840.63
## - egoposition_factored:political_interest_factored  4   822.04 844.04
##
## Step:  AIC=838.74
## clear_party ~ egoposition_factored + ostwest + political_interest_factored +
##     egoposition_factored:ostwest + ostwest:political_interest_factored +
##     egoposition_factored:political_interest_factored
##
##                                                   Df Deviance    AIC
## <none>                                                 810.74 838.74
## - egoposition_factored:ostwest                     2   815.43 839.43
## - ostwest:political_interest_factored              2   816.18 840.18
## - egoposition_factored:political_interest_factored  4   822.86 842.86
```

```
summary(hm8m2)
```

```
##
## Call:
## glm(formula = clear_party ~ egoposition_factored + ostwest +
##     political_interest_factored + egoposition_factored:ostwest +
##     ostwest:political_interest_factored + egoposition_factored:political_interest_factored,
##     family = binomial, data = train_data_clear_party)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6246  -0.9745  -0.6042   1.1728   2.5437
##
## Coefficients:
##                                                                     Estimate
## (Intercept)                                                          -2.8197
## egoposition_factoredNeutral                                          -1.3823
## egoposition_factoredCon                                               2.6208
## ostwestEast Germany                                                   2.3434
## political_interest_factoredNeutral                                    2.6084
## political_interest_factoredInterested                                 2.8305
## egoposition_factoredNeutral:ostwestEast Germany                       0.7203
## egoposition_factoredCon:ostwestEast Germany                          -1.1357
## ostwestEast Germany:political_interest_factoredNeutral               -2.7597
## ostwestEast Germany:political_interest_factoredInterested            -2.8523
## egoposition_factoredNeutral:political_interest_factoredNeutral       -0.6683
## egoposition_factoredCon:political_interest_factoredNeutral           -3.3685
## egoposition_factoredNeutral:political_interest_factoredInterested  -0.4482
## egoposition_factoredCon:political_interest_factoredInterested        -4.1820
##                                                                     Std. Error
## (Intercept)                                                             1.4886
## egoposition_factoredNeutral                                             1.4358
## egoposition_factoredCon                                                 1.3825
```

```
## ostwestEast Germany                                               1.4399
## political_interest_factoredNeutral                                1.5050
## political_interest_factoredInterested                             1.4923
## egoposition_factoredNeutral:ostwestEast Germany                   0.6220
## egoposition_factoredCon:ostwestEast Germany                       0.7203
## ostwestEast Germany:political_interest_factoredNeutral            1.4563
## ostwestEast Germany:political_interest_factoredInterested         1.4433
## egoposition_factoredNeutral:political_interest_factoredNeutral    1.3757
## egoposition_factoredCon:political_interest_factoredNeutral        1.3958
## egoposition_factoredNeutral:political_interest_factoredInterested 1.3350
## egoposition_factoredCon:political_interest_factoredInterested     1.3849
##                                                                 z value
## (Intercept)                                                      -1.894
## egoposition_factoredNeutral                                      -0.963
## egoposition_factoredCon                                           1.896
## ostwestEast Germany                                               1.627
## political_interest_factoredNeutral                                1.733
## political_interest_factoredInterested                             1.897
## egoposition_factoredNeutral:ostwestEast Germany                   1.158
## egoposition_factoredCon:ostwestEast Germany                      -1.577
## ostwestEast Germany:political_interest_factoredNeutral           -1.895
## ostwestEast Germany:political_interest_factoredInterested        -1.976
## egoposition_factoredNeutral:political_interest_factoredNeutral   -0.486
## egoposition_factoredCon:political_interest_factoredNeutral       -2.413
## egoposition_factoredNeutral:political_interest_factoredInterested  -0.336
## egoposition_factoredCon:political_interest_factoredInterested    -3.020
##                                                                 Pr(>|z|)
## (Intercept)                                                      0.05821 .
## egoposition_factoredNeutral                                      0.33569
## egoposition_factoredCon                                          0.05800 .
## ostwestEast Germany                                              0.10364
## political_interest_factoredNeutral                               0.08306 .
## political_interest_factoredInterested                            0.05786 .
## egoposition_factoredNeutral:ostwestEast Germany                  0.24684
## egoposition_factoredCon:ostwestEast Germany                      0.11485
## ostwestEast Germany:political_interest_factoredNeutral           0.05809 .
## ostwestEast Germany:political_interest_factoredInterested        0.04812 *
## egoposition_factoredNeutral:political_interest_factoredNeutral   0.62712
## egoposition_factoredCon:political_interest_factoredNeutral       0.01581 *
## egoposition_factoredNeutral:political_interest_factoredInterested  0.73709
## egoposition_factoredCon:political_interest_factoredInterested    0.00253 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 895.32  on 744  degrees of freedom
## Residual deviance: 810.74  on 731  degrees of freedom
## AIC: 838.74
##
## Number of Fisher Scoring iterations: 5
```

```
AIC(hm8m2, hm7m2)
```

```
##       df      AIC
## hm8m2 14 838.7361
## hm7m2 20 846.1706
```

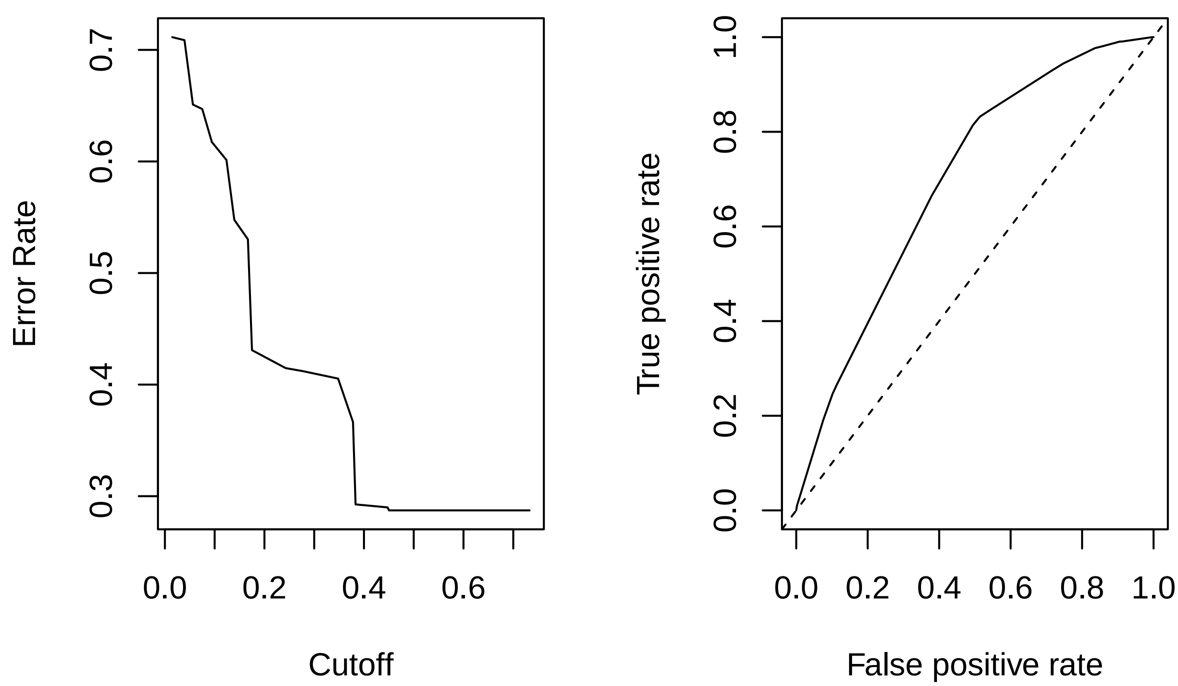
```
sum(ifelse(
  predict(
    hm8m2,type="response")>0.5,T,F) == train_data_clear_party$clear_party) /
  nrow(train_data_clear_party)
```

```
## [1] 0.7127517
```

```
dadesroc<-prediction(predict(hm8m2,type="response"),
                     train_data_clear_party$clear_party)
par(mfrow=c(1,2))
performance(dadesroc,"auc")
```

```
## A performance instance
##   'Area under the ROC curve'
```

```
plot(performance(dadesroc,"err"))
plot(performance(dadesroc,"tpr","fpr"))
abline(0,1,lty=2)
```



```
AIC(hm4m2)+AIC(hm8m2)
```

```
## [1] 1136.762
```

From the results below we see that the model has an accuracy of 65%.

```
right <- test_data[ifelse(predict(hm4m2, test_data, type="response") > 0.5, T, F), ]
left_center <- test_data[!ifelse(predict(hm4m2, test_data, type="response") > 0.5, T, F), ]

left <- left_center[ifelse(predict(hm8m2, left_center, type="response") > 0.5, T, F), ]
center <- left_center[!ifelse(predict(hm8m2, left_center, type="response") > 0.5, T, F), ]

# accuracy
(sum((right$right_wing == T)) + sum(left$clear_party == T & left$right_wing == F) +
  sum((center$clear_party == F))) / nrow(test_data)
```

```
## [1] 0.6485149
```

3.2.3 Model interpretation

Here we interpret the values of the coefficients obtained.

We get a coefficient of 0.9075 for the position neutral and 2.8021 for the Con level of egposition_factored, eastern people have a -0.749 coefficient and females a coefficient of -0.296.

From the plots we can see that people from the east and that they are open about immigration are more likely to have a clear party to which they are going to vote. People that are from the west and have a high interest in politics are also likely to have a clear party to which they are going to vote.

```
summary(hm4m2)
```

```
##
## Call:
## glm(formula = right_wing ~ egoposition_factored + ostwest + gender +
##     egoposition_factored:gender, family = binomial, data = train_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.07184  -0.32777  -0.20988  -0.00007   3.02542
##
## Coefficients:
##                                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)                              -3.0555     0.4589  -6.658 2.78e-11
## egoposition_factoredNeutral               0.9075     0.5541   1.638   0.1015
## egoposition_factoredCon                   2.8021     0.4780   5.862 4.59e-09
## ostwestEast Germany                      -0.7490     0.3230  -2.319   0.0204
## genderFemale                            -15.9835   719.3046  -0.022   0.9823
## egoposition_factoredNeutral:genderFemale  14.3143   719.3054   0.020   0.9841
## egoposition_factoredCon:genderFemale      15.1016   719.3047   0.021   0.9832
##
## (Intercept)                             ***
## egoposition_factoredNeutral
```

```
## egoposition_factoredCon                  ***
## ostwestEast Germany                      *
## genderFemale
## egoposition_factoredNeutral:genderFemale
## egoposition_factoredCon:genderFemale
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.85  on 797  degrees of freedom
## Residual deviance: 284.03  on 791  degrees of freedom
## AIC: 298.03
##
## Number of Fisher Scoring iterations: 18
```

```
coef(hm4m2)
```

```
##                              (Intercept)
##                               -3.0555396
##              egoposition_factoredNeutral
##                                0.9075346
##                  egoposition_factoredCon
##                                2.8020703
##                      ostwestEast Germany
##                               -0.7490227
##                             genderFemale
##                              -15.9835464
## egoposition_factoredNeutral:genderFemale
##                               14.3143291
##     egoposition_factoredCon:genderFemale
##                               15.1016195
```

```
summary(hm8m2)
```

```
##
## Call:
## glm(formula = clear_party ~ egoposition_factored + ostwest +
##     political_interest_factored + egoposition_factored:ostwest +
##     ostwest:political_interest_factored + egoposition_factored:political_interest_factored,
##     family = binomial, data = train_data_clear_party)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6246  -0.9745  -0.6042   1.1728   2.5437
##
## Coefficients:
##                                                                   Estimate
## (Intercept)                                                        -2.8197
## egoposition_factoredNeutral                                        -1.3823
## egoposition_factoredCon                                             2.6208
## ostwestEast Germany                                                 2.3434
## political_interest_factoredNeutral                                  2.6084
```

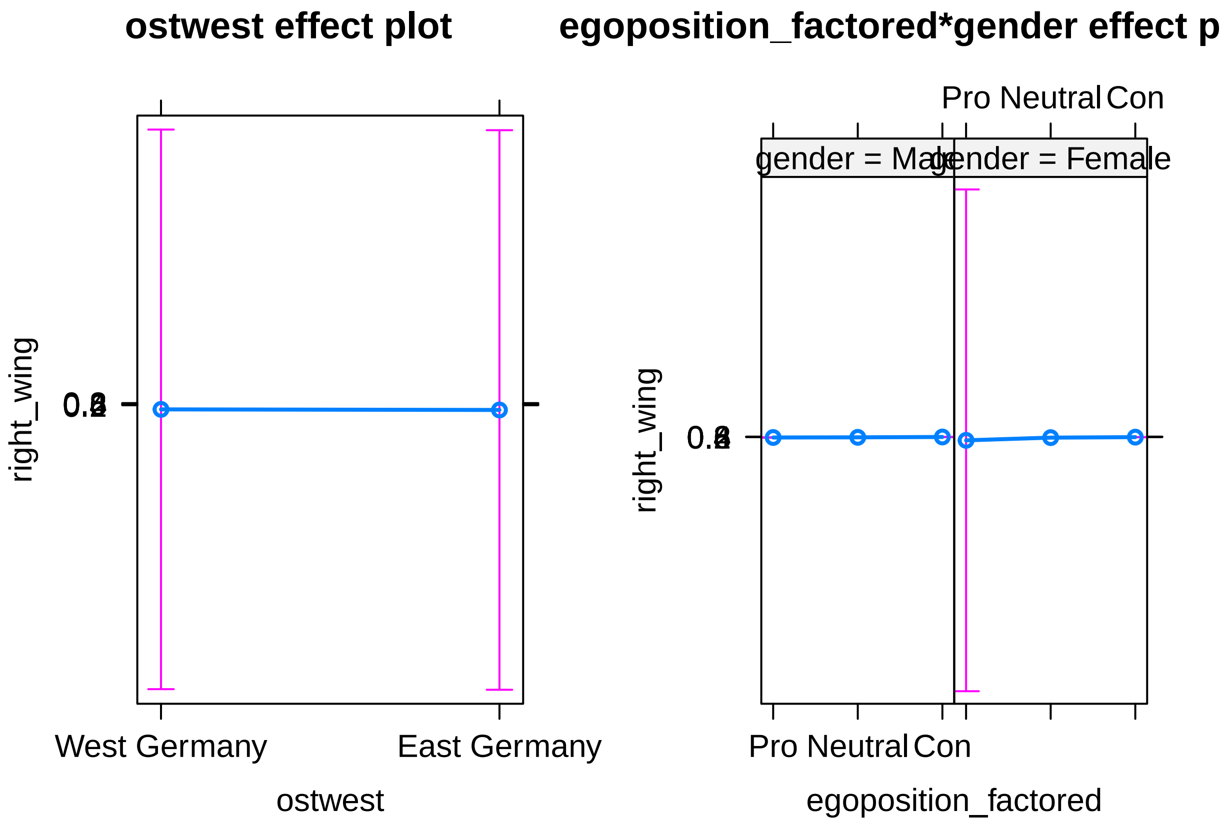
```
## political_interest_factoredInterested                            2.8305
## egoposition_factoredNeutral:ostwestEast Germany                   0.7203
## egoposition_factoredCon:ostwestEast Germany                      -1.1357
## ostwestEast Germany:political_interest_factoredNeutral           -2.7597
## ostwestEast Germany:political_interest_factoredInterested        -2.8523
## egoposition_factoredNeutral:political_interest_factoredNeutral   -0.6683
## egoposition_factoredCon:political_interest_factoredNeutral       -3.3685
## egoposition_factoredNeutral:political_interest_factoredInterested  -0.4482
## egoposition_factoredCon:political_interest_factoredInterested    -4.1820
##                                                                  Std. Error
## (Intercept)                                                          1.4886
## egoposition_factoredNeutral                                          1.4358
## egoposition_factoredCon                                              1.3825
## ostwestEast Germany                                                  1.4399
## political_interest_factoredNeutral                                   1.5050
## political_interest_factoredInterested                                1.4923
## egoposition_factoredNeutral:ostwestEast Germany                      0.6220
## egoposition_factoredCon:ostwestEast Germany                          0.7203
## ostwestEast Germany:political_interest_factoredNeutral               1.4563
## ostwestEast Germany:political_interest_factoredInterested            1.4433
## egoposition_factoredNeutral:political_interest_factoredNeutral       1.3757
## egoposition_factoredCon:political_interest_factoredNeutral           1.3958
## egoposition_factoredNeutral:political_interest_factoredInterested    1.3350
## egoposition_factoredCon:political_interest_factoredInterested        1.3849
##                                                                  z value
## (Intercept)                                                       -1.894
## egoposition_factoredNeutral                                       -0.963
## egoposition_factoredCon                                            1.896
## ostwestEast Germany                                                1.627
## political_interest_factoredNeutral                                 1.733
## political_interest_factoredInterested                              1.897
## egoposition_factoredNeutral:ostwestEast Germany                    1.158
## egoposition_factoredCon:ostwestEast Germany                       -1.577
## ostwestEast Germany:political_interest_factoredNeutral            -1.895
## ostwestEast Germany:political_interest_factoredInterested         -1.976
## egoposition_factoredNeutral:political_interest_factoredNeutral    -0.486
## egoposition_factoredCon:political_interest_factoredNeutral        -2.413
## egoposition_factoredNeutral:political_interest_factoredInterested  -0.336
## egoposition_factoredCon:political_interest_factoredInterested     -3.020
##                                                                  Pr(>|z|)
## (Intercept)                                                       0.05821 .
## egoposition_factoredNeutral                                       0.33569
## egoposition_factoredCon                                           0.05800 .
## ostwestEast Germany                                               0.10364
## political_interest_factoredNeutral                                0.08306 .
## political_interest_factoredInterested                             0.05786 .
## egoposition_factoredNeutral:ostwestEast Germany                   0.24684
## egoposition_factoredCon:ostwestEast Germany                       0.11485
## ostwestEast Germany:political_interest_factoredNeutral            0.05809 .
## ostwestEast Germany:political_interest_factoredInterested         0.04812 *
## egoposition_factoredNeutral:political_interest_factoredNeutral    0.62712
## egoposition_factoredCon:political_interest_factoredNeutral        0.01581 *
## egoposition_factoredNeutral:political_interest_factoredInterested  0.73709
## egoposition_factoredCon:political_interest_factoredInterested     0.00253 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 895.32  on 744  degrees of freedom
## Residual deviance: 810.74  on 731  degrees of freedom
## AIC: 838.74
##
## Number of Fisher Scoring iterations: 5
```

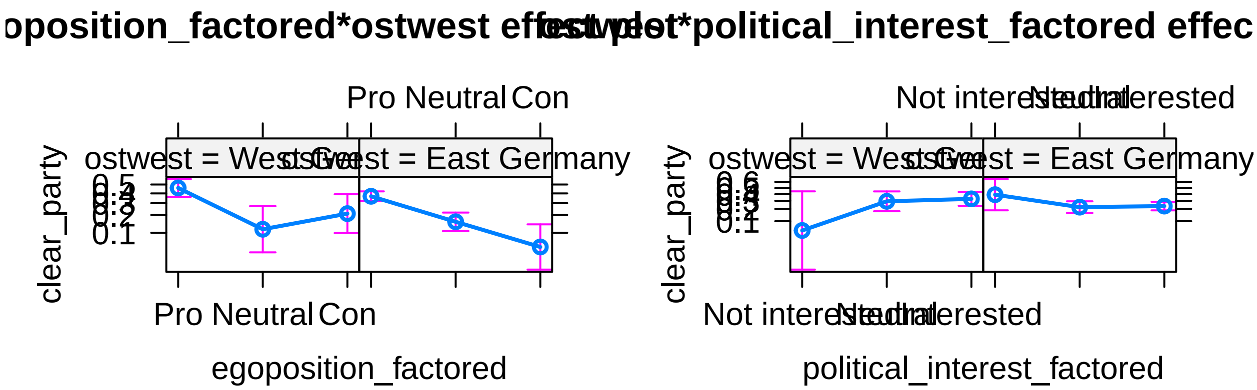
```
coef(hm8m2)
```

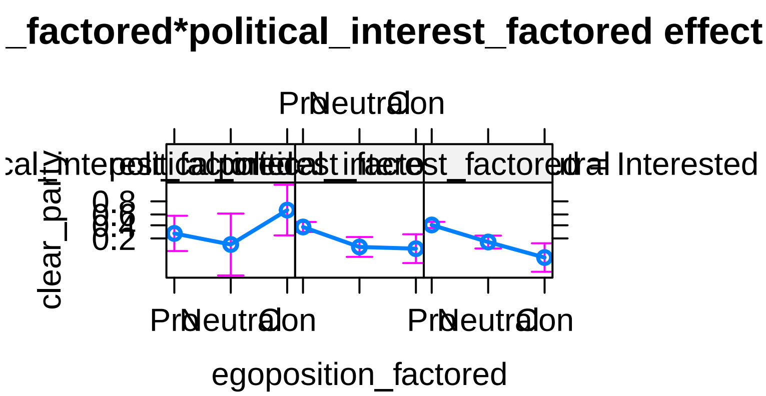
```
##                                                         (Intercept)
##                                                          -2.8197073
##                                          egoposition_factoredNeutral
##                                                          -1.3823034
##                                              egoposition_factoredCon
##                                                           2.6207587
##                                                  ostwestEast Germany
##                                                           2.3433788
##                                  political_interest_factoredNeutral
##                                                           2.6084014
##                               political_interest_factoredInterested
##                                                           2.8304964
##                     egoposition_factoredNeutral:ostwestEast Germany
##                                                           0.7203047
##                         egoposition_factoredCon:ostwestEast Germany
##                                                          -1.1357335
##              ostwestEast Germany:political_interest_factoredNeutral
##                                                          -2.7597063
##           ostwestEast Germany:political_interest_factoredInterested
##                                                          -2.8523186
##      egoposition_factoredNeutral:political_interest_factoredNeutral
##                                                          -0.6682849
##          egoposition_factoredCon:political_interest_factoredNeutral
##                                                          -3.3684799
## egoposition_factoredNeutral:political_interest_factoredInterested
##                                                          -0.4481922
##       egoposition_factoredCon:political_interest_factoredInterested
##                                                          -4.1819559
```

```
plot(allEffects(hm4m2))
```




```
plot(allEffects(hm8m2))
```






```
summary(hm8m2)
```

```
##
## Call:
## glm(formula = clear_party ~ egoposition_factored + ostwest +
##     political_interest_factored + egoposition_factored:ostwest +
##     ostwest:political_interest_factored + egoposition_factored:political_interest_factored,
##     family = binomial, data = train_data_clear_party)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6246  -0.9745  -0.6042   1.1728   2.5437
##
## Coefficients:
##                                                              Estimate
## (Intercept)                                                   -2.8197
## egoposition_factoredNeutral                                   -1.3823
## egoposition_factoredCon                                        2.6208
## ostwestEast Germany                                            2.3434
## political_interest_factoredNeutral                             2.6084
## political_interest_factoredInterested                          2.8305
## egoposition_factoredNeutral:ostwestEast Germany                0.7203
## egoposition_factoredCon:ostwestEast Germany                   -1.1357
## ostwestEast Germany:political_interest_factoredNeutral        -2.7597
## ostwestEast Germany:political_interest_factoredInterested     -2.8523
```

```
## egoposition_factoredNeutral:political_interest_factoredNeutral     -0.6683
## egoposition_factoredCon:political_interest_factoredNeutral         -3.3685
## egoposition_factoredNeutral:political_interest_factoredInterested  -0.4482
## egoposition_factoredCon:political_interest_factoredInterested      -4.1820
##                                                                   Std. Error
## (Intercept)                                                           1.4886
## egoposition_factoredNeutral                                           1.4358
## egoposition_factoredCon                                               1.3825
## ostwestEast Germany                                                   1.4399
## political_interest_factoredNeutral                                    1.5050
## political_interest_factoredInterested                                 1.4923
## egoposition_factoredNeutral:ostwestEast Germany                       0.6220
## egoposition_factoredCon:ostwestEast Germany                           0.7203
## ostwestEast Germany:political_interest_factoredNeutral                1.4563
## ostwestEast Germany:political_interest_factoredInterested             1.4433
## egoposition_factoredNeutral:political_interest_factoredNeutral        1.3757
## egoposition_factoredCon:political_interest_factoredNeutral            1.3958
## egoposition_factoredNeutral:political_interest_factoredInterested     1.3350
## egoposition_factoredCon:political_interest_factoredInterested         1.3849
##                                                                   z value
## (Intercept)                                                        -1.894
## egoposition_factoredNeutral                                        -0.963
## egoposition_factoredCon                                             1.896
## ostwestEast Germany                                                 1.627
## political_interest_factoredNeutral                                  1.733
## political_interest_factoredInterested                               1.897
## egoposition_factoredNeutral:ostwestEast Germany                     1.158
## egoposition_factoredCon:ostwestEast Germany                        -1.577
## ostwestEast Germany:political_interest_factoredNeutral             -1.895
## ostwestEast Germany:political_interest_factoredInterested          -1.976
## egoposition_factoredNeutral:political_interest_factoredNeutral     -0.486
## egoposition_factoredCon:political_interest_factoredNeutral         -2.413
## egoposition_factoredNeutral:political_interest_factoredInterested  -0.336
## egoposition_factoredCon:political_interest_factoredInterested      -3.020
##                                                                   Pr(>|z|)
## (Intercept)                                                        0.05821 .
## egoposition_factoredNeutral                                        0.33569
## egoposition_factoredCon                                            0.05800 .
## ostwestEast Germany                                                0.10364
## political_interest_factoredNeutral                                 0.08306 .
## political_interest_factoredInterested                              0.05786 .
## egoposition_factoredNeutral:ostwestEast Germany                    0.24684
## egoposition_factoredCon:ostwestEast Germany                        0.11485
## ostwestEast Germany:political_interest_factoredNeutral             0.05809 .
## ostwestEast Germany:political_interest_factoredInterested          0.04812 *
## egoposition_factoredNeutral:political_interest_factoredNeutral     0.62712
## egoposition_factoredCon:political_interest_factoredNeutral         0.01581 *
## egoposition_factoredNeutral:political_interest_factoredInterested  0.73709
## egoposition_factoredCon:political_interest_factoredInterested      0.00253 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 895.32  on 744  degrees of freedom
## Residual deviance: 810.74  on 731  degrees of freedom
## AIC: 838.74
##
## Number of Fisher Scoring iterations: 5
```

4 Conclusions and model comparision

The best models using numeric variables are:

- Best nominal model is mm4m that is described by the formula: formula = vote ~ income + egoposition_immigration + I(egoposition_immigration^2)

The best models using new factors created are:

- Best nominal model is mmf1 that is described by the formula: formula = vote ~ egoposition_factored + ostwest + gender

Taking into account the results displayed below we can say that the best model is the hierarchical that have the new created factors in it, as the AIC is the lowest one (1136.762)

General conclusions:

- We think that the dataset does not contain all the information needed to accurately predict the political orientation of the individuals.
- Most important variable to predict the political orientation is the position of the individuals towards immigration.
- Even though we weren't able to do an accurate prediction using this dataset, the exercise was a good practice that allowed us to perform statistical analysis and find insights for future work.

```
AIC(mm4m, mmf1 )
```

```
##      df      AIC
## mm4m 20 2530.228
## mmf1 25 2524.612
```

```
BIC(mm4m, mmf1)
```

```
##      df      BIC
## mm4m 20 2623.870
## mmf1 25 2641.665
```

```
AIC(hm2m1) + AIC(hm1m1) # hierarchical without the factors
```

```
## [1] 1158.745
```

```
AIC(hm4m2)+AIC(hm8m2) # hierarchical with the factors
```

```
## [1] 1136.762
```

```r
BIC(hm2m1)+BIC(hm1m1)
```

```
## [1] 1205.016
```

```r
BIC(hm4m2)+BIC(hm8m2)
```

```
## [1] 1234.124
```