Appendix

Louisa Boulaziz

May 3, 2021

Contents

1 Loading in data and packages

In this section I am loading in the data and all the packages I use to run the analysis.

2 Structure of the data

I have coded all the parties in all preliminary reference cases and their lawyers from the first judgment issued in CJEU dating to 2016. I have merged this data together with the data used in the Carrubba et al (2008) article. Thus the data consists of all cases dating from 1964 to 1995 for which Carrubba et al has coded the outcome of the case. In the sample I have 2770 observations. Each new row in the dataset is an applicant or a defendant in a preliminary reference case. The applicant and the defendant are nested in a case, and they either have a lawyer or they do not (binary indicator). If they have a lawyer or lawyer team which has experience this is also flagged (binary indicator). The parties also win or loses a case which is also a binary indicator. In the data I have information about the number of member state governments who submits observations in favor of either the applicant or the defendant. I do also have whether or not the commission submitted observations in favor of either the applicant or the defendant. I have a case salience measure – the total number of governments who submits observations.

```
# Total number of cases in the data
length(unique(data$celex))

## [1] 1301

table(data$member_state)

##
##   10   12   15    6    9
##  586 1202  110  216  656
```

3 Dependent variable

I have two variables that measures whether the applicant or defendant won the case. One is at the case level – `ecjplaintiffagree` – and the other one is at the level of the role in the dispute – `favorable_ruling`. These variables measures the same thing – the cases the applicants win and the cases the defendants win.

```
winning <- data %>%
  select(ecjplaintiffagree, celex)
winning <- distinct(winning)
```

```r
win_rate <- table(winning$ecjplaintiffagree)
# Who wins in percentage
prop.table(win_rate)*100
# Visualizing it

winning %>%
  count(ecjplaintiffagree) %>%
  mutate(perc = n / nrow(winning)) -> winning1
winning1$ecjplaintiffagree <- ifelse(winning1$ecjplaintiffagree == 0, "Defendant", "Applica
ggplot(winning1, aes(as.factor(ecjplaintiffagree), (perc*100))) +
  geom_col(color = "purple", fill = "purple") +
  scale_y_continuous() +
  ylab("Percentage of cases") +
  xlab("")+
  ggtitle("Who wins in the CJEU?")+
  theme_minimal()
```

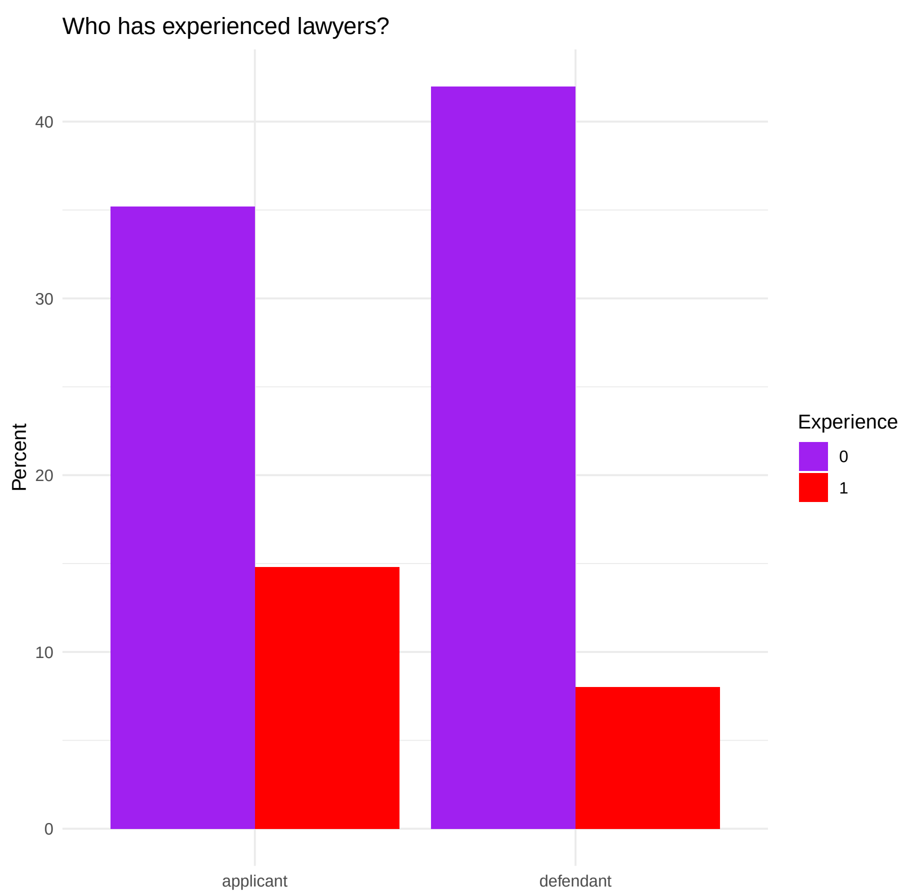
4 Main independent variables

I focus on three main explanatory variables in the analysis: `experience`, `lawyer` and `salience`. I describe them more in detail below.

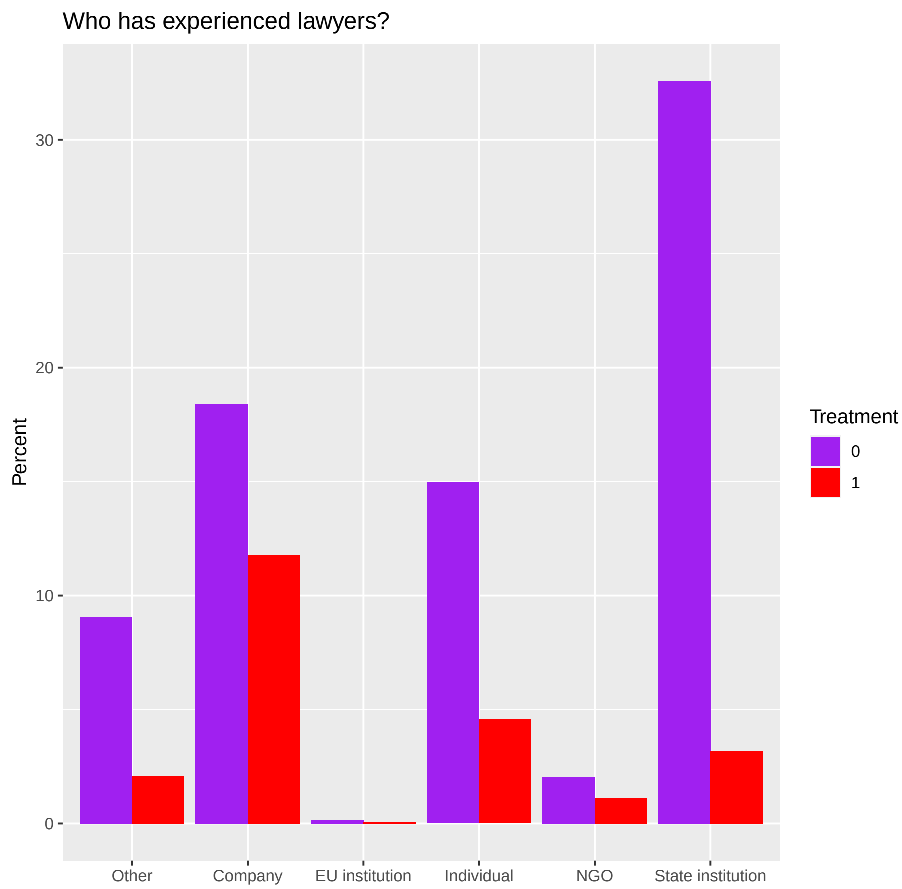
4.1 Treatment

The treatment-variable is grouped at the level of the role in the legal dispute. This means that it measures whether or not the applicants and the defendants have received the treatment. The treatment is whether or not the applicant or the defendant has a lawyer or lawyer team which have prior litigation experience. Note that all 1s represent lawyers and lawyer teams that have at least argued one prior case at the CJEU. All 0s represent non-lawyers – that means when applicants and defendants come to court without representation the treatment shows 0. The treatment also shows 0 for applicant and defendants who come to court with lawyers without prior litigation experience.

```r
data %>%
  group_by(celex, role, experience) %>%
  ggplot(aes(as.factor(role), fill = as.factor(experience))) +
  geom_bar(position = "dodge", aes(y = (..count..)/sum(..count..)*100)) +
  ylab("Percent") +
  labs(fill = "Experience") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who has experienced lawyers?") +
  xlab("") +
  theme_minimal()
```




```r
data %>%
  group_by(celex, type, experience) %>%
  ggplot(aes(as.factor(type), fill = as.factor(experience))) +
  geom_bar(position = "dodge", aes(y = (..count..)/sum(..count..)*100)) +
  ylab("Percent") +
  labs(fill = "Treatment") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who has experienced lawyers?") +
  xlab("")
```

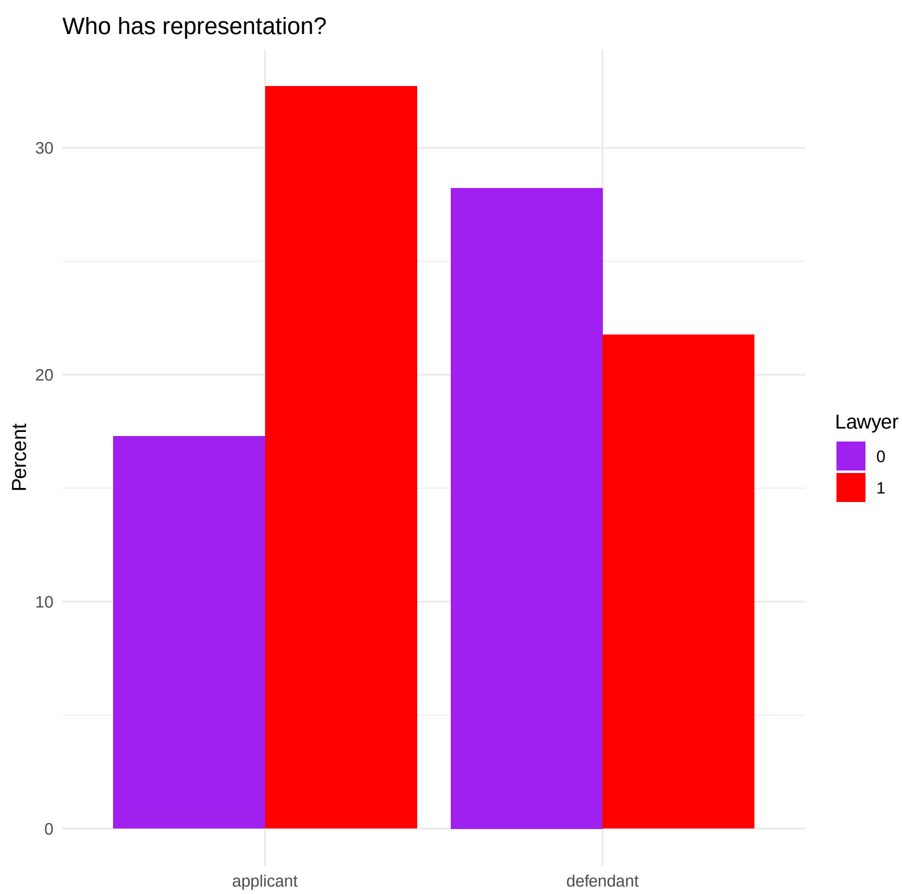



```
role_experience <- table(data$role, data$experience)
prop.table(role_experience)*100

##
##                       0        1
##   applicant 35.19856 14.80144
##   defendant 41.98556  8.01444

rm(role_experience)
```

4.2 Lawyer

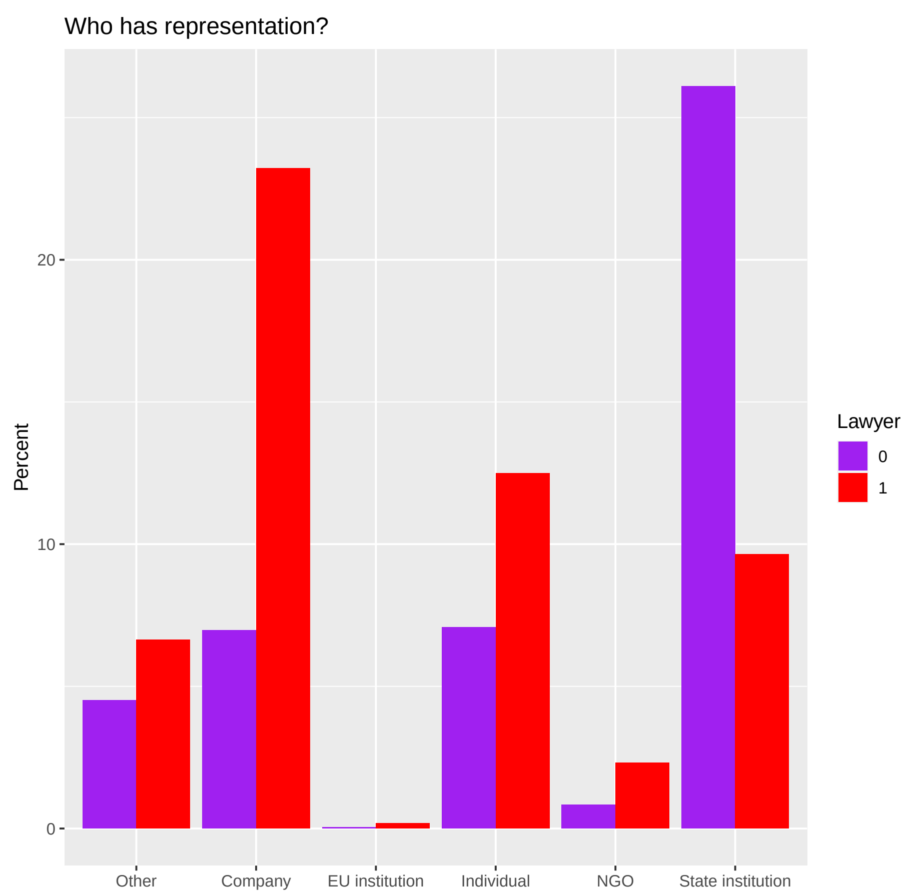
This variable is at the level of the party in the dispute. Here I recode it so that if you came to court as the applicant with another applicant who had a lawyer, you also have a lawyer.

```
data %>%
  group_by(celex, role, lawyer) %>%
  ggplot(aes(as.factor(role), fill = as.factor(lawyer))) +
  geom_bar(position = "dodge", aes(y = (..count..)/sum(..count..)*100)) +
  ylab("Percent") +
  labs(fill = "Lawyer") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who has representation?")+
  xlab("") +
  theme_minimal()
```




```
data %>%
  group_by(celex, type, lawyer) %>%
  ggplot(aes(as.factor(type), fill = as.factor(lawyer))) +
  geom_bar(position = "dodge", aes(y = (..count..)/sum(..count..)*100)) +
  ylab("Percent") +
```

```
  labs(fill = "Lawyer") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who has representation?")+
  xlab("")
```




```
role_lawyer <- table(data$role, data$lawyer)
prop.table(role_lawyer)*100
```

```
##
##                    0        1
##   applicant 17.29242 32.70758
##   defendant 28.23105 21.76895
```

```
17.29242+28.23105
```

```
## [1] 45.52347
```

```
rm(role_lawyer)
```

4.3 Salience

Salience is measured as number of EU governments that submit observations in cases referred to the CJEU. I create this variable by adding up the number of governments that submits observations in favor of the applicant and in favor of the defendant. I create a salience measure to seperate between salient and non salient cases. To make this variable I flag all cases that has above the average submitted observations

```
summary(data$salience)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  1.0000  0.8245  1.0000  7.0000

length(unique(data$celex)) # Total number of cases

## [1] 1301

# Total number of cases
data %>%
  select(celex, salience) %>%
  group_by(celex) %>%
  count(salience) %>%
  ggplot(aes(as.factor(salience))) +
  geom_bar(color = "violet", fill = "violet") +
  xlab("No. of governments") +
  ylab("No. of cases") +
  theme_minimal() +
  ggtitle("Submitted observations by EU member state governments")
# Percentage of cases
data %>%
  select(celex, salience) %>%
  group_by(celex) %>%
  count(salience) %>%
  ggplot(aes(as.factor(salience))) +
  geom_bar(color = "violet", fill = "violet", aes(y = (..count..)/sum(..count..)*100)) +
  xlab("No. of governments") +
  ylab("Percentage of cases") +
  theme_minimal() +
  ggtitle("Submitted observations by member state governments")
## Binary measure of salience

table(data$binary_salience)

##
##    0    1
## 1320 1450
```

```r
data %>%
  select(celex, binary_salience) %>%
  group_by(celex) %>%
  count(binary_salience) %>%
  ggplot(aes(as.factor(binary_salience))) +
  geom_bar(color = "violet", fill = "violet", aes(y = (..count..)/sum(..count..)*100)) +
  xlab("Salient cases flagged as 1") +
  ylab("Percent of cases") +
  theme_minimal() +
  ggtitle("Cases considered salient")
prop.table(table(data$binary_salience))*100

##
##        0        1
## 47.65343 52.34657
```

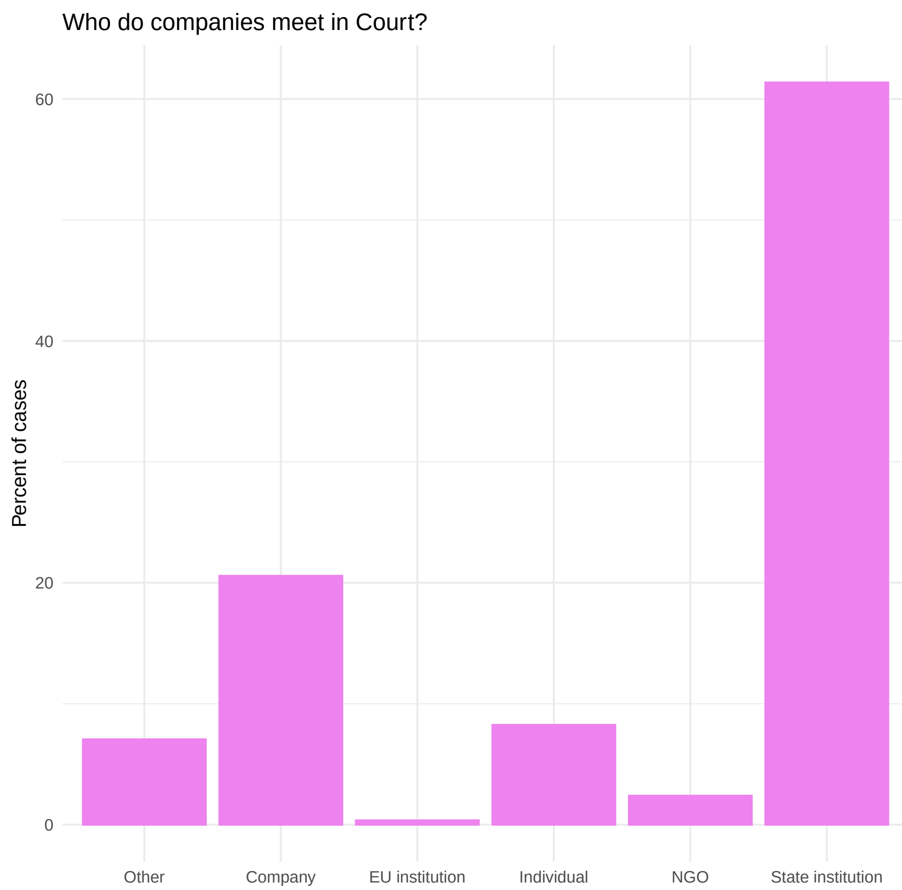
5 Control variables

5.1 Type of actor and opponent in court

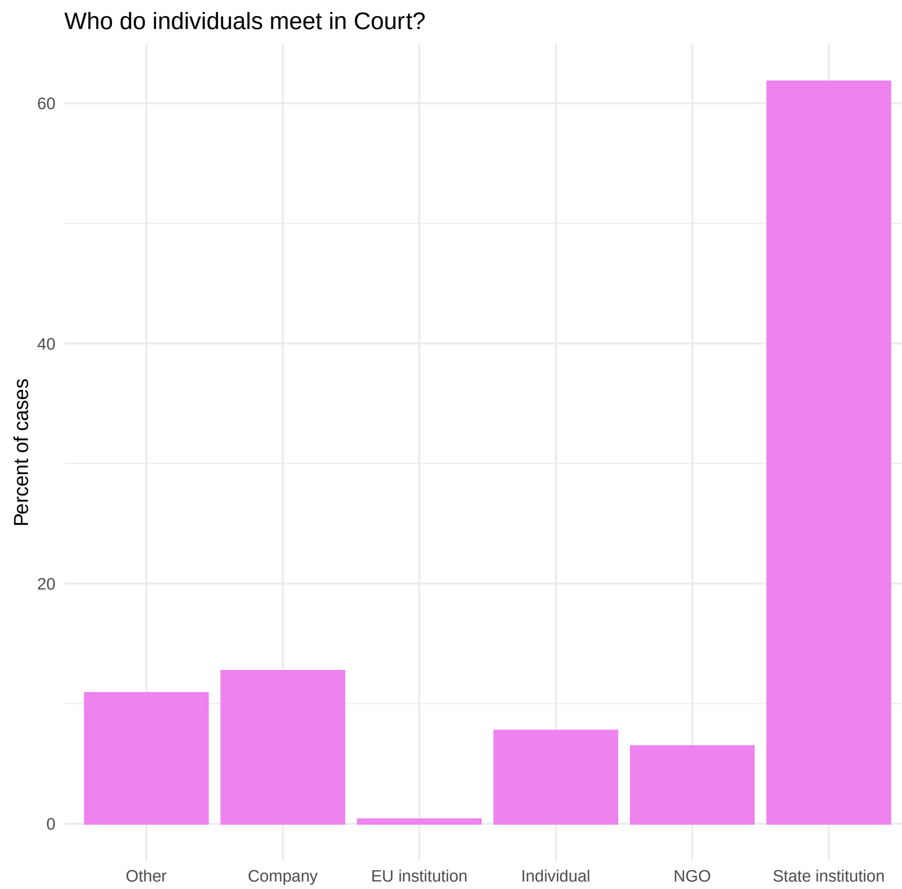
The typical case in the CJEU consists of either a company or an individual against a state institution. State institutions are involved in 40 percent of the cases. Companies meet state institutions in over 60 percent of the time (in over 60 percent of cases they are involved in). In approximately 20 percent of cases a company meets another company in the CJEU. Individuals also meet state institutions in over 60 percent of the cases. Likewise, when looking at who state institutions meet in court, in 50 percent of the cases, they meet companies and in close to 35 percent of cases they meet individuals.

```r
data %>%
  select(celex, type, role) %>%
  group_by(celex) %>%
  count(type) %>%
  ggplot(aes(as.factor(type))) +
  geom_bar(color = "violet", fill = "violet", aes(y = (..count..)/sum(..count..)*100)) +
  xlab("") +
  ylab("Percent of cases") +
  theme_minimal() +
  ggtitle("Who litigates before the CJEU?")
data %>%
  filter(type == "Company") %>%
  group_by(celex) %>%
   ggplot(aes(type_opponent)) +
   geom_bar(color = "violet", fill = "violet", aes(y = (..count..)/sum(..count..)*100)) +
```

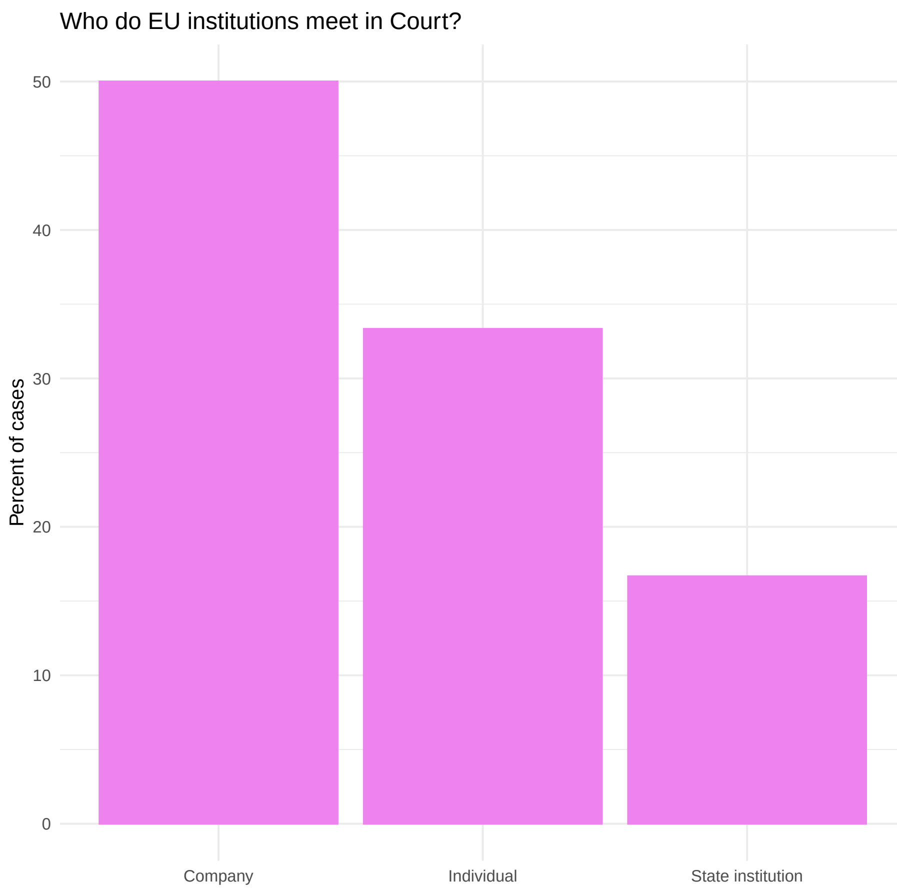
```
  xlab("") +
  ylab("Percent of cases") +
  theme_minimal() +
  ggtitle("Who do companies meet in Court?")
```



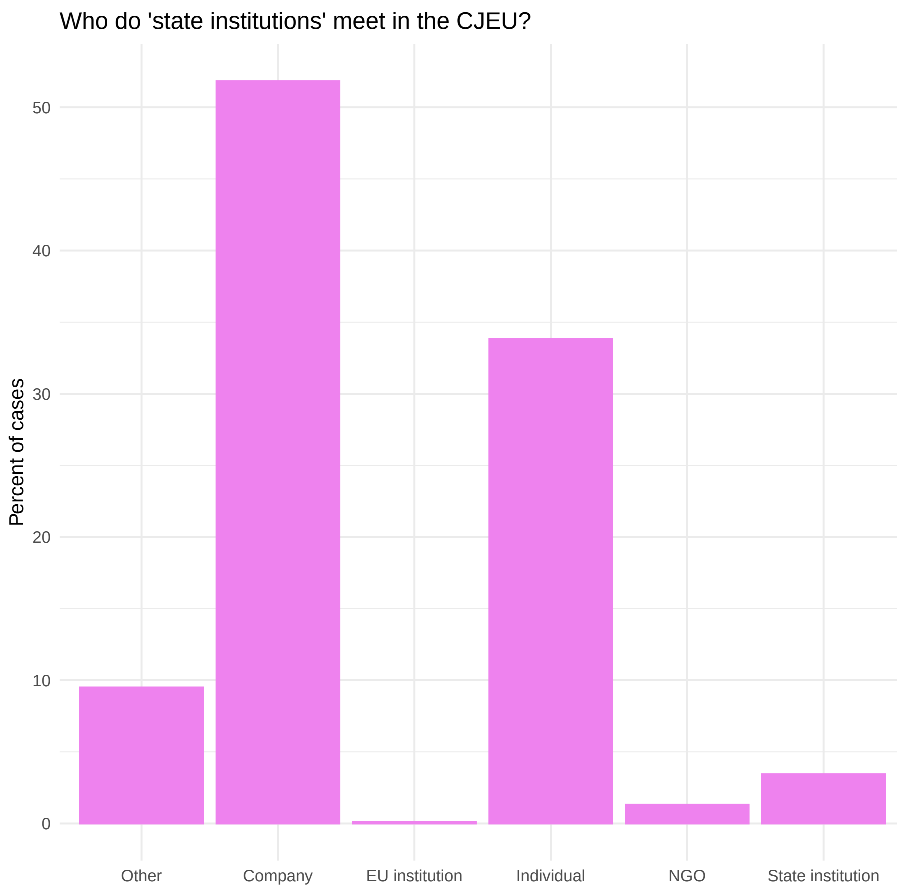

```
data %>%
  filter(type == "Individual") %>%
  group_by(celex) %>%
   ggplot(aes(type_opponent)) +
   geom_bar(color = "violet", fill = "violet", aes(y = (..count..)/sum(..count..)*100)) +
  xlab("") +
  ylab("Percent of cases") +
  theme_minimal() +
  ggtitle("Who do individuals meet in Court?")
```




```r
data %>%
  filter(type == "EU institution") %>%
  group_by(celex) %>%
   ggplot(aes(type_opponent)) +
   geom_bar(color = "violet", fill = "violet", aes(y = (..count..)/sum(..count..)*100)) +
  xlab("") +
  ylab("Percent of cases") +
  theme_minimal() +
  ggtitle("Who do EU institutions meet in Court?")
```




```
data %>%
  filter(type == "State institution") %>%
  group_by(celex) %>%
   ggplot(aes(type_opponent)) +
   geom_bar(color = "violet", fill = "violet", aes(y = (..count..)/sum(..count..)*100)) +
  xlab("") +
  ylab("Percent of cases") +
  theme_minimal() +
  ggtitle("Who do 'state institutions' meet in the CJEU?")
```



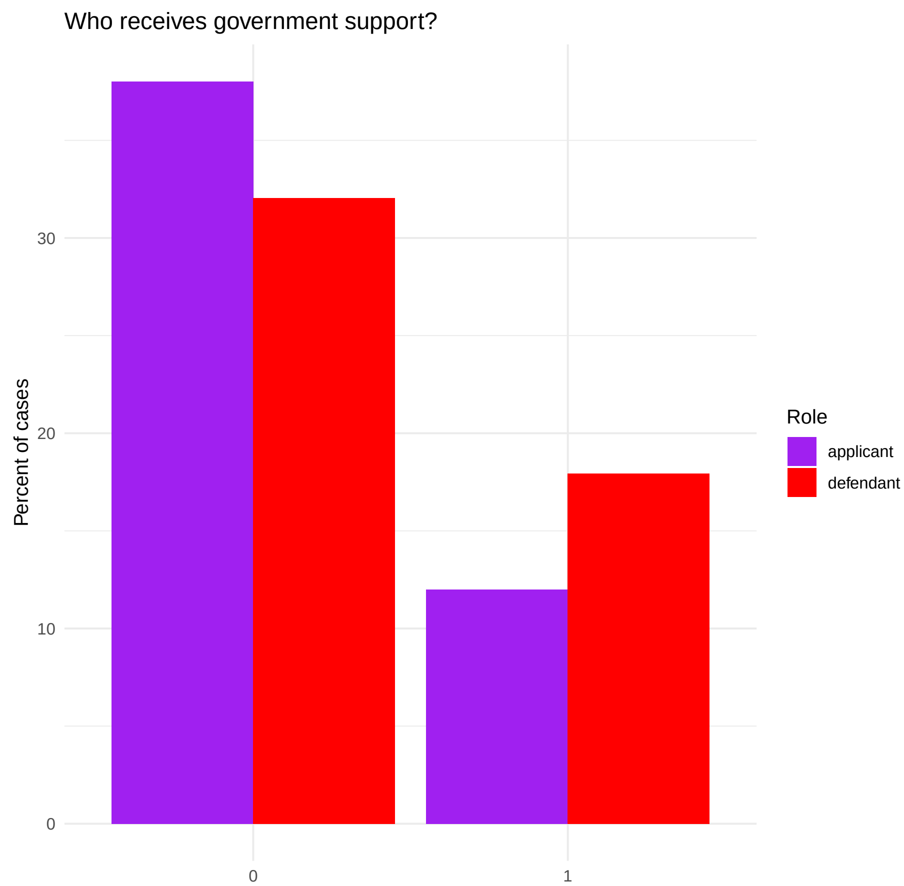
5.2 Role in legal dispute

```
data %>%
  select(celex, role) %>%
  group_by(celex) %>%
  ggplot(aes(as.factor(role))) +
  geom_bar(color = "violet", fill = "violet", aes(y = (..count..)/sum(..count..)*100)) +
  xlab("") +
  ylab("Percent of cases") +
  theme_minimal() +
  ggtitle("Who litigates before the CJEU?")
```

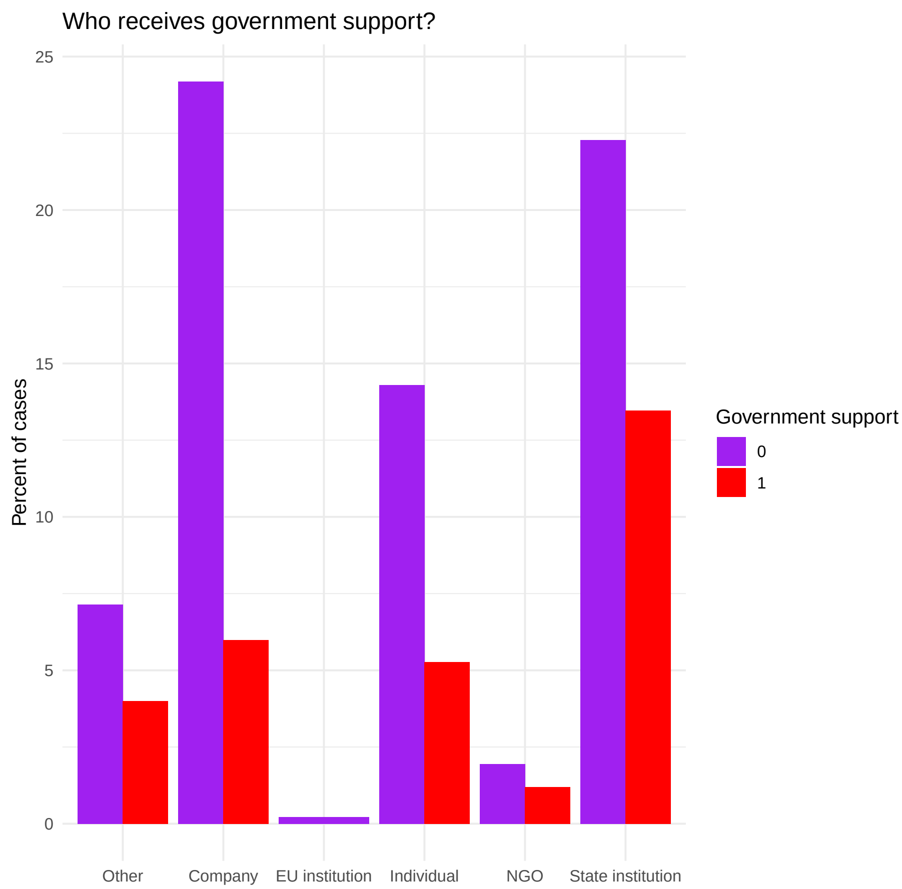
5.3 Government support

This variable measures whether or not the applicant or defendant had government support. This is measured has a binary indicator.

```
table(data$government_support)

##
##    0    1
## 1941  829

# Who has government support in Court?


data %>%
  group_by(celex, government_support, role) %>%
  ggplot(aes(as.factor(government_support), fill = as.factor(role))) +
  geom_bar(position = "dodge", aes(y = (..count..)/sum(..count..)*100)) +
  ylab("Percent of cases") +
  labs(fill = "Role") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who receives government support?")+
  xlab("") +
  theme_minimal()
```




```
data %>%
  group_by(celex, government_support, type) %>%
  ggplot(aes(as.factor(type), fill = as.factor(government_support))) +
  geom_bar(position = "dodge", aes(y = (..count..)/sum(..count..)*100)) +
  ylab("Percent of cases") +
  labs(fill = "Government support") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who receives government support?")+
  xlab("")+
  theme_minimal()
```

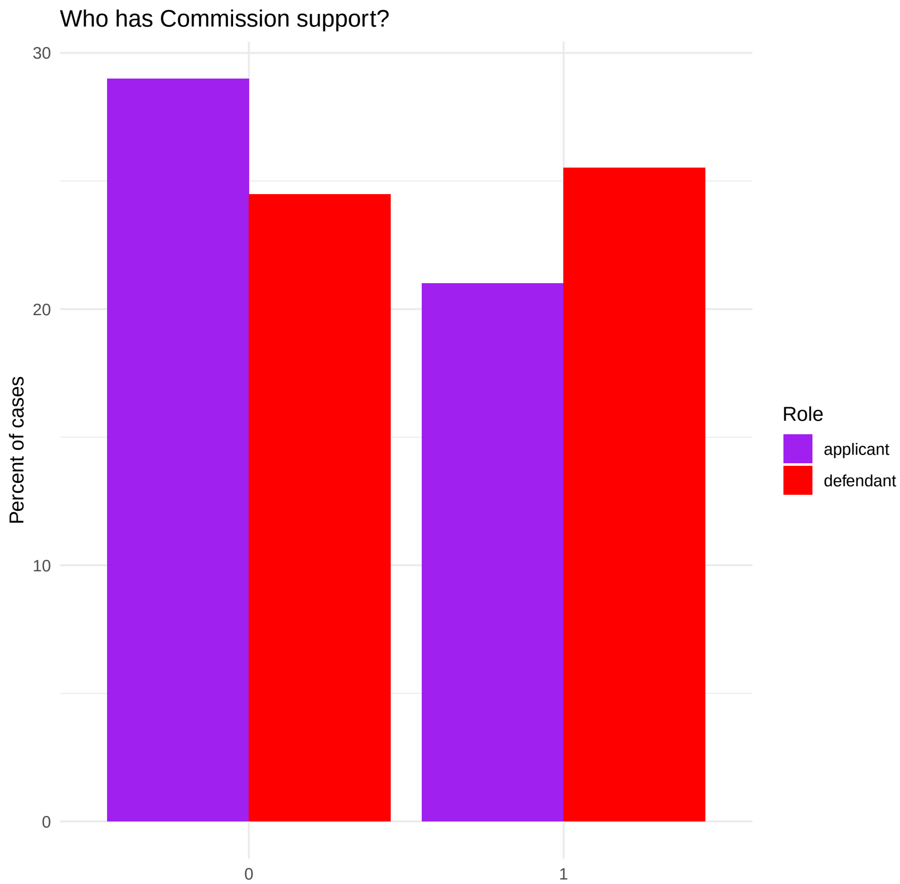


5.4 Commission support

This variable indicates whether or not the applicant or defendant in the legal dispute has support from the Commission. The variable is binary. Support from the Commission is flagged as 1.

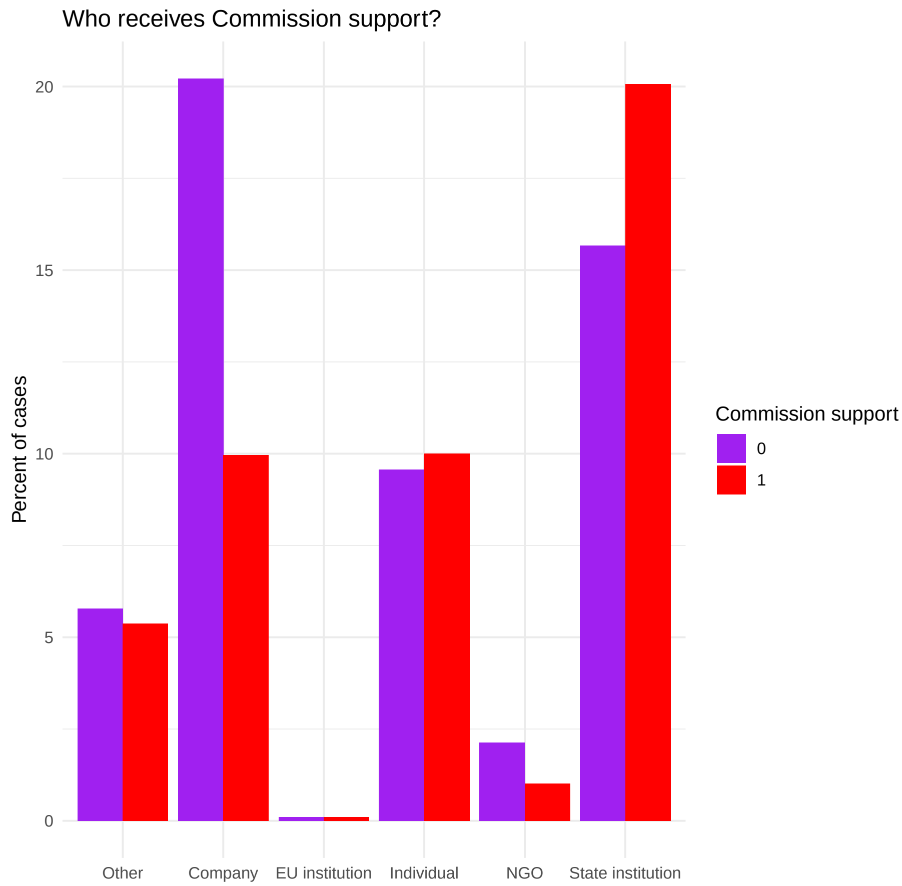
```
table(data$commission_support)

##
##    0    1
## 1481 1289

# Who has commission support in Court?


data %>%
```

```
  group_by(celex, commission_support, role) %>%
  ggplot(aes(as.factor(commission_support), fill = as.factor(role))) +
  geom_bar(position = "dodge", aes(y = (..count..)/sum(..count..)*100)) +
  ylab("Percent of cases") +
  labs(fill = "Role") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who has Commission support?")+
  xlab("")+
  theme_minimal()
```




```
# Who has commission support in Court?

data %>%
  group_by(celex, commission_support, type) %>%
  ggplot(aes(as.factor(type), fill = as.factor(commission_support))) +
  geom_bar(position = "dodge", aes(y = (..count..)/sum(..count..)*100)) +
```

```
  ylab("Percent of cases") +
  labs(fill = "Commission support") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who receives Commission support?")+
  xlab("")+
  theme_minimal()
```




```
EU <- data %>%
  filter(commission_support == 1)

prop.table(table(EU$type))*100

##
##            Other           Company    EU institution        Individual
##       11.5593483        21.4119472         0.2327386        21.4895268
##              NGO State institution
```

```
##         2.1722265        43.1342126

rm(EU)
```

5.5 Number of member states

```
table(data$member_state)

##
##   10   12   15    6    9
##  586 1202  110  216  656
```

6 Descriptive statistics

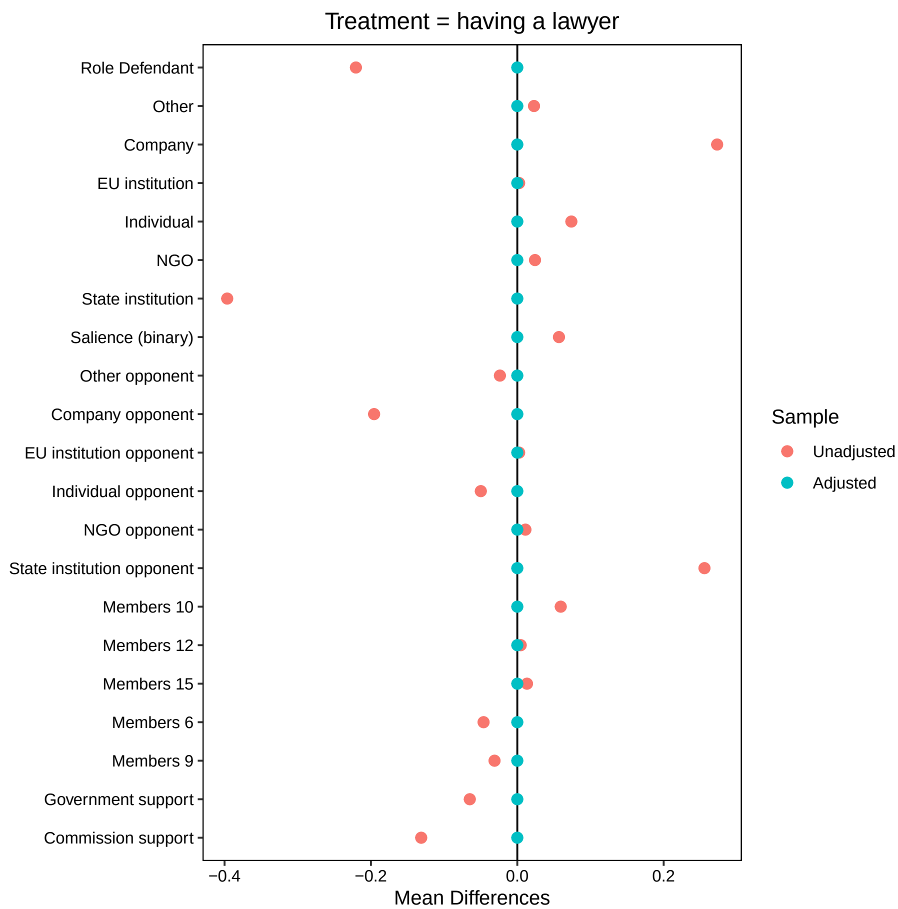
This table gives an overview of the numeric variables in the data.

```
stats_data <- data.frame(data)
names(stats_data)

##  [1] "celex"              "role"               "type"
##  [4] "experience"         "win"                "type_opponent"
##  [7] "salience"           "lawyer"             "binary_salience"
## [10] "government_support" "commission_support" "member_state"
## [13] "ecjplaintiffagree"

stats_data <- stats_data %>%
  select(4:5, 7:11)
stargazer(stats_data, type = "text")

##
## =================================================================
## Statistic            N   Mean  St. Dev. Min Pctl(25) Pctl(75) Max
## -----------------------------------------------------------------
## experience         2,770 0.228  0.420    0     0        0      1
## win                2,770 0.500  0.500    0     0        1      1
## salience           2,770 0.825  1.035    0     0        1      7
## lawyer             2,770 0.545  0.498    0     0        1      1
## binary_salience    2,770 0.523  0.500    0     0        1      1
## government_support 2,770 0.299  0.458    0     0        1      1
## commission_support 2,770 0.465  0.499    0     0        1      1
## ecjplaintiffagree  2,770 0.471  0.499    0     0        1      1
## -----------------------------------------------------------------

rm(stats_data)
```

7 Treatment-is-lawyer sample

7.1 Matching

```
# Treatment = having a lawyer
sample_1 <- matchit(lawyer ~ role +
                      type + binary_salience +
                         type_opponent
                    + member_state
                    + government_support
                    + commission_support,
                       method = "cem",
                       estimand = "ATT",
                       data = data)


sample1 <- match.data(sample_1, data=data)

love.plot(sample_1,
          title = "Treatment = having a lawyer", var.names = var_names)
```




```
### Add more plots

### Review the sample

length(unique(sample1$celex))

## [1] 1142

prop.table(table(sample1$ecjplaintiffagree))

##
##         0         1
## 0.5374942 0.4625058
```

7.2 Analysis

```
## Estimating a model with controls

law <- glm(win~ lawyer + role + type +
             type_opponent + member_state +
             binary_salience
            + lawyer*binary_salience
            + government_support
            + commission_support,
          family = binomial(link = "logit"),
          data = sample1)



beta_law <- law$coefficients

tab_law <- (exp(beta_law)-1)*100
tab_law

##                    (Intercept)                          lawyer
##                     -76.298308                       12.330685
##                  roledefendant                     typeCompany
##                       1.904952                        9.142960
##                 typeIndividual                         typeNGO
##                      22.964018                      127.691590
##           typeState institution           type_opponentCompany
##                     -14.264075                       11.046082
##        type_opponentIndividual                type_opponentNGO
##                     -14.448801                      -34.203812
## type_opponentState institution                 member_state12
##                      15.079536                       11.349618
##                 member_state15                   member_state6
##                      98.665730                       -2.812215
##                  member_state9                 binary_salience
##                      -4.169242                      -48.179588
##             government_support              commission_support
##                     189.368327                     1595.049236
##         lawyer:binary_salience
##                      -7.333599

vcov_law <- vcovHC(law, "HC1")

stargazer(law, type = "text", se = list(vcov_law))

##
## ==========================================================
##                                     Dependent variable:
```

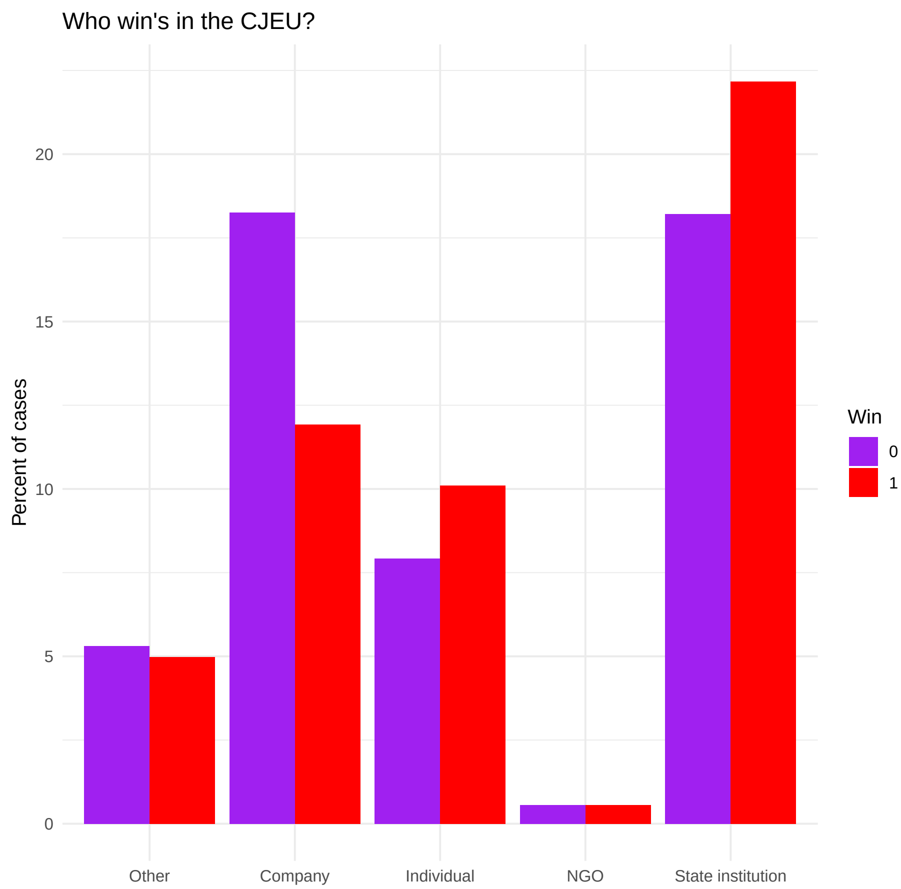
```
##                                 ---------------------------
##                                             win
## ---------------------------------------------------------
## lawyer                                    0.116***
##                                           (-0.015)
##
## roledefendant                              0.019*
##                                           (-0.011)
##
## typeCompany                               0.087***
##                                           (-0.028)
##
## typeIndividual                            0.207***
##                                           (-0.026)
##
## typeNGO                                   0.823***
##                                           (-0.043)
##
## typeState institution                     -0.154***
##                                           (-0.031)
##
## type_opponentCompany                      0.105***
##                                           (-0.029)
##
## type_opponentIndividual                   -0.156***
##                                           (-0.026)
##
## type_opponentNGO                          -0.419***
##                                           (-0.024)
##
## type_opponentState institution            0.140***
##                                           (-0.036)
##
## member_state12                            0.108***
##                                           (-0.025)
##
## member_state15                            0.686***
##                                           (-0.021)
##
## member_state6                              -0.029*
##                                           (-0.015)
##
## member_state9                             -0.043**
##                                           (-0.017)
##
```

```
## binary_salience                          -0.657***
##                                          (-0.016)
##
## government_support                       1.063***
##                                           (0.001)
##
## commission_support                       2.830***
##                                          (-0.007)
##
## lawyer:binary_salience                   -0.076***
##                                           (0.010)
##
## Constant                                 -1.440***
##                                           (0.096)
##
## ----------------------------------------------------------
## Observations                               2,147
## Log Likelihood                          -1,041.243
## Akaike Inf. Crit.                        2,120.487
## ==========================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

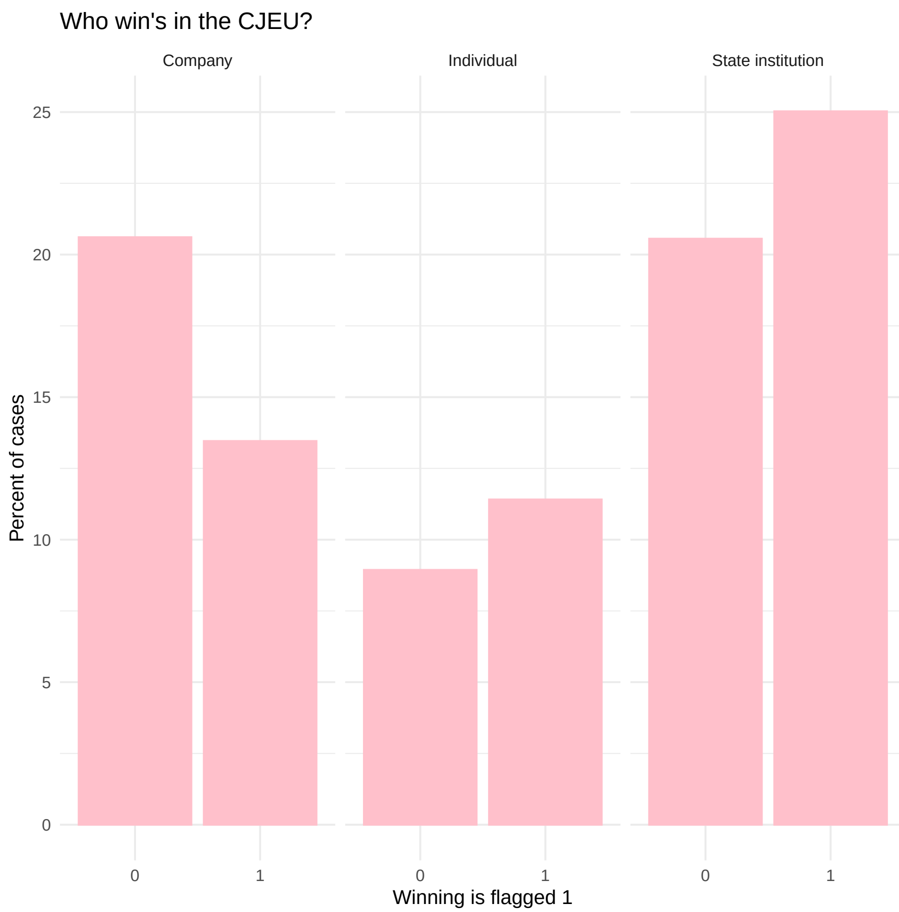
```
stargazer(law, type = "text", se = list(vcov_law), style = "all2",
          single.row = TRUE, no.space = TRUE, font.size = "small", align = TRUE,
          dep.var.caption = "Binomial logistic regression",
          dep.var.labels = "Favorable ruling",
         keep = c("lawyer", "role",
                  "type", "type_opponent",
                   "binary_salience",
                  "government_support",
                  "commission_support"),
         covariate.labels = c("Lawyer", "Defendant", "Company",
                              "Individual", "NGO",
                              "State institution",
                              "Opponent Company",
                              "Opponent Individual",
                              "Opponent NGO",
                              "Opponent State",
                              "Salience",
                              "Government support",
                              "Commission support",
                              "Interaction lawyer + salience"))
```

```
##
## ==========================================================
##                                Binomial logistic regression
```

```
##                               ----------------------------
##                                     Favorable ruling
## -----------------------------------------------------------
## Lawyer                               0.116*** (-0.015)
## Defendant                             0.019* (-0.011)
## Company                              0.087*** (-0.028)
## Individual                           0.207*** (-0.026)
## NGO                                  0.823*** (-0.043)
## State institution                    -0.154*** (-0.031)
## Opponent Company                     0.105*** (-0.029)
## Opponent Individual                  -0.156*** (-0.026)
## Opponent NGO                         -0.419*** (-0.024)
## Opponent State                       0.140*** (-0.036)
## Salience                             -0.657*** (-0.016)
## Government support                    1.063*** (0.001)
## Commission support                   2.830*** (-0.007)
## Interaction lawyer + salience        -0.076*** (0.010)
## -----------------------------------------------------------
## Observations                                2,147
## Log Likelihood                            -1,041.243
## Akaike Inf. Crit.                          2,120.487
## Residual Deviance                  2,082.487 (df = 2128)
## Null Deviance                      2,976.318 (df = 2146)
## ===========================================================
## Note:                             *p<0.1; **p<0.05; ***p<0.01

sample1 %>%
  group_by(celex) %>%
  ggplot(aes(type, fill = as.factor(win))) +
  geom_bar(aes(y = (..count..)/sum(..count..)*100), position = "dodge") +
  ylab("Percent of cases") +
  labs(fill = "Win") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who win's in the CJEU?")+
  xlab("")+
  theme_minimal()
```




```
sample1 %>%
  group_by(celex) %>%
  filter(type != "NGO" & type != "Other") %>%
  ggplot(aes(as.factor(win))) +
  geom_bar(aes(y = (..count..)/sum(..count..)*100), position = "dodge", color = "pink", fil
  ylab("Percent of cases") +
  ggtitle("Who win's in the CJEU?")+
  xlab("")+
  theme_minimal() +
  facet_wrap(~ type) +
  xlab("Winning is flagged 1")
```



7.3 Plotting effects

```
# Estimating a new model without interaction

law1 <- glm(win~ lawyer + role + type +
            type_opponent + member_state +
            binary_salience
           + government_support
           + commission_support,
          family = binomial(link = "logit"),
          data = sample1)



beta_law1 <- law1$coefficients
```

```
tab_law1 <- (exp(beta_law1)-1)*100
tab_law1
```

```
##                   (Intercept)                       lawyer
##                    -75.882042                     8.447117
##                 roledefendant                  typeCompany
##                      1.941331                     9.539142
##                typeIndividual                      typeNGO
##                     23.053552                   127.561008
##          typeState institution      type_opponentCompany
##                    -14.055699                    10.493623
##        type_opponentIndividual          type_opponentNGO
##                    -14.497113                   -34.342526
## type_opponentState institution             member_state12
##                     15.277928                    11.280267
##                member_state15               member_state6
##                     99.132130                    -2.910656
##                 member_state9              binary_salience
##                     -4.150013                   -50.543997
##            government_support           commission_support
##                    192.608628                  1592.168994
```

```
vcov_law1 <- vcovHC(law1, "HC1")

# Saving my se's in the model object
cluster_law <- cluster.vcov(law, cluster = sample1$member_state)
print(cluster_law)
sqrt(diag(cluster_law))
law$cluster_law <- cluster_law
law1$vcov_law <- vcov_law1
# Setting seed
set.seed(24)

simBetas <- mvrnorm(n = 1000,
                    mu = coefficients(law1),
                    Sigma = law1$vcov_law)


names(coefficients(law1))
```

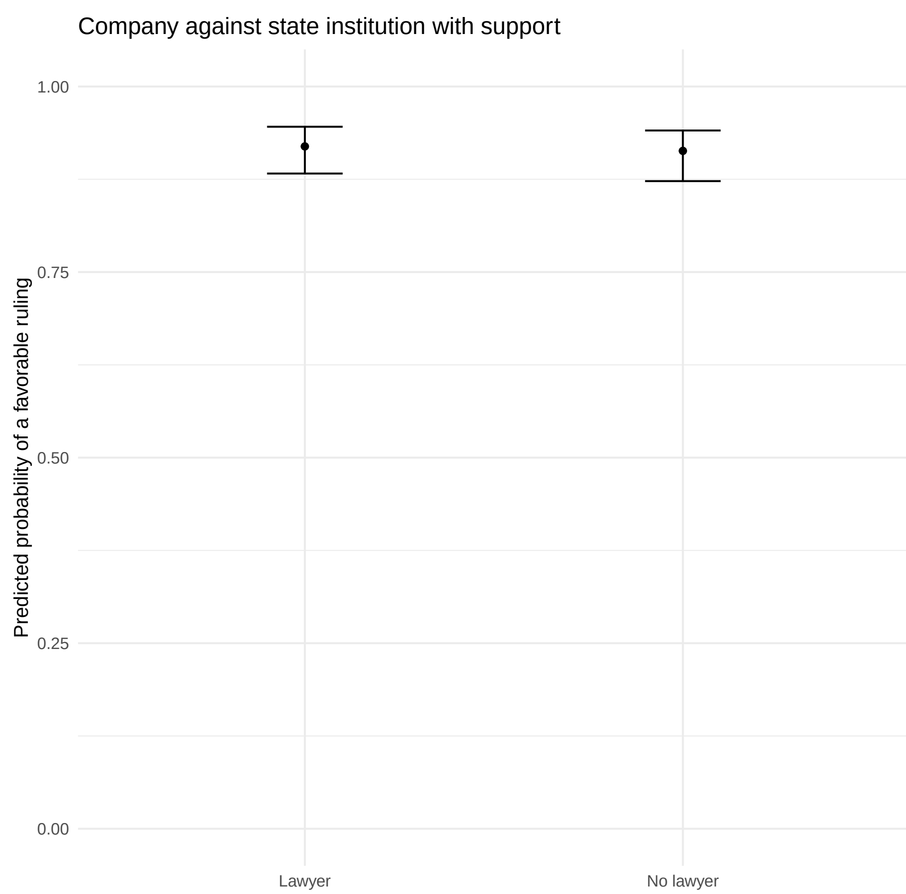
```
##  [1] "(Intercept)"                    "lawyer"
##  [3] "roledefendant"                  "typeCompany"
##  [5] "typeIndividual"                 "typeNGO"
##  [7] "typeState institution"          "type_opponentCompany"
##  [9] "type_opponentIndividual"        "type_opponentNGO"
## [11] "type_opponentState institution" "member_state12"
```

```
## [13] "member_state15"                 "member_state6"
## [15] "member_state9"                  "binary_salience"
## [17] "government_support"             "commission_support"

xMatrix <- cbind(1, #the intercept
                 c(0, 1), # Let lawyer vary
                 1, # defendant
                 1, # Company
                 0, # Individual
                 0, # NGO
                 0, # State institution
                 0, # opponent company
                 0, # opponent individual
                 0, # opponent NGO
                 1, # opponent state institution
                 0, # M 12
                 0, # M 15
                 0, # M 6,
                 1, # M 9,
                 mean(sample1$binary_salience),
                 1, # Government support,
                 1 # Commission support
                 )




ncol(simBetas) == ncol(xMatrix) #yay!!

## [1] TRUE

### Calculating predicted probabilities: Her multipliserer du simuleringen med xmatrisen
xBetaMatrix <- xMatrix %*% t(simBetas ) ## this just means x times the betas
predProbs <- 1/(1+exp(-xBetaMatrix)) #This is the predicted probability, for another type o


### Getting point estimates and confidence intervals:
quantileValues <- apply(X = predProbs,  ## read up on the apply() family of functions!
                        MARGIN = 1, ## this means we are applying a function to all the rows
                        FUN =  ## The fun argument defines what I want to do with all the ro
                          ## What we want to do here is to use quantile to get the quantiles
                          ### bounds of the confidence intervals and our point estimates:
                          quantile, probs = c(.05,.5,.95))
quantileValues <- as.data.frame(t(quantileValues))

plotPoints <- cbind(c("No lawyer", "Lawyer"),quantileValues)
```

```r
plotPoints
```

```
##   c("No lawyer", "Lawyer")        5%       50%       95%
## 1                No lawyer 0.8725370 0.9132698 0.9407584
## 2                   Lawyer 0.8827545 0.9193092 0.9457931
```

```r
colnames(plotPoints) <- c("Lawyer", "lower", "estimate", "upper")



ggplot(plotPoints,
       aes(x = Lawyer,
           y = estimate,
           ymin = lower,
           ymax = upper)) +
  geom_errorbar(width =.2)+
  geom_point()+
  ylim(0,1)+
  ylab("Predicted probability of a favorable ruling")+
  xlab("")+
  theme_minimal() +
  ggtitle("Company against state institution with support")
```

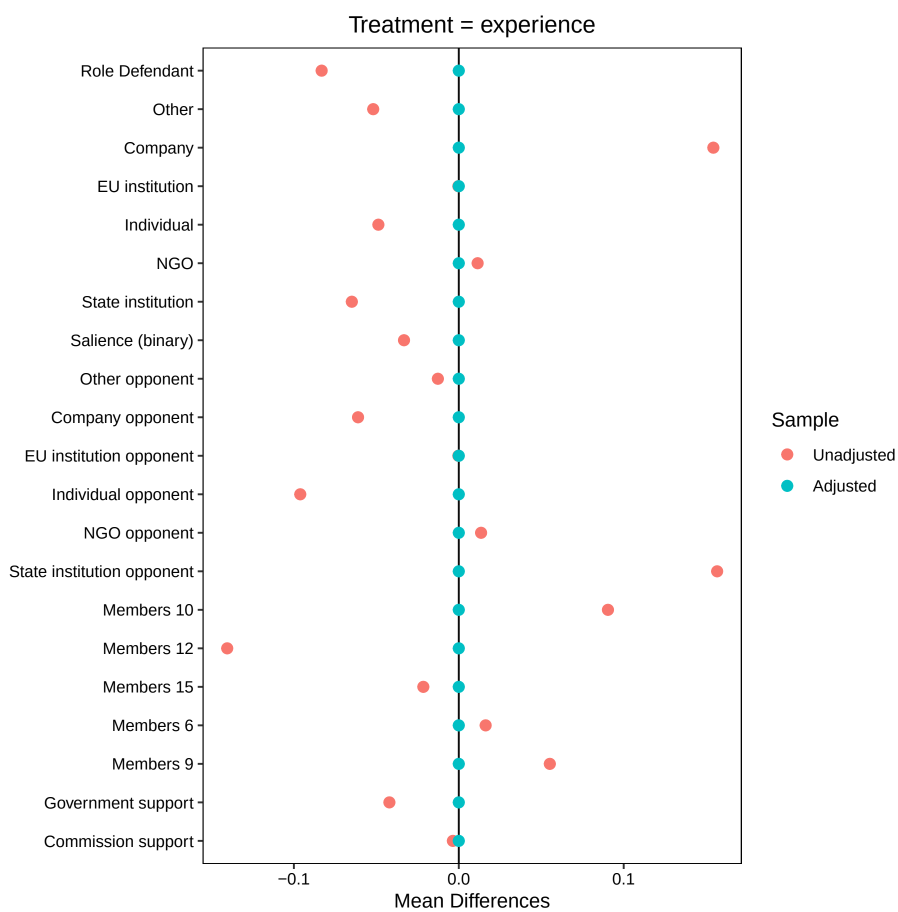


8 Treatment-is-experience sample

8.1 Matching

```
# Filtering out a sample consisting of parties with only lawyers
df <- data %>%
  filter(lawyer == 1) %>%
  ungroup()


sample_2 <- matchit(experience ~ role + type
                    + binary_salience +
                       type_opponent
                    + member_state +
```

```
                             government_support
                             + commission_support,
                                 method = "cem",
                                 estimand = "ATT",
                                 data = df)


sample2 <- match.data(sample_2, data=df)


# Checking balance

love.plot(sample_2, var.names = var_names, title = "Treatment = experience")
```



```r
# Add more balance plots

length(unique(sample2$celex))

## [1] 806

table(sample2$role)

##
## applicant defendant
##       670       356

prop.table(table(sample2$ecjplaintiffagree))

##
##         0         1
## 0.5107212 0.4892788

prop.table(table(sample2$experience, sample2$type))*100

##
##          Other    Company EU institution Individual        NGO
##   0  6.6276803 22.8070175      0.0000000 12.3781676  0.7797271
##   1  4.3859649 26.6081871      0.0000000  9.2592593  0.6822612
##
##     State institution
##   0        10.2339181
##   1         6.2378168
```

8.2 Analysis

```r
## Estimating a model with controls

m2 <- glm(win~ experience + role + type +
            type_opponent + member_state +
            binary_salience
          + government_support
          + commission_support,
          family = binomial(link = "logit"),
          data = sample2)

beta2 <- m2$coefficients

tab2 <- (exp(beta2)-1)*100
tab2
```

```
##                  (Intercept)                     experience
##                   -57.117924                     -12.876583
##                 roledefendant                    typeCompany
##                    -7.879649                     -22.513542
##                typeIndividual                        typeNGO
##                   -21.069126                       9.958094
##        typeState institution         type_opponentCompany
##                   -49.254928                     -35.041613
##      type_opponentIndividual              type_opponentNGO
##                    -7.612492                     -12.701192
## type_opponentState institution                member_state12
##                    -3.678927                       3.338763
##                member_state15                  member_state6
##                    66.724356                       1.746452
##                 member_state9                binary_salience
##                     7.227989                     -49.102830
##           government_support             commission_support
##                   196.553588                    2267.956205
```

```r
vcov_m2 <- vcovHC(m2, "HC1")

stargazer(m2, type = "text", se = list(vcov_m2))
```

```
##
## ==========================================================
##                                  Dependent variable:
##                               ---------------------------
##                                          win
## ----------------------------------------------------------
## experience                             -0.138***
##                                        (-0.008)
##
## roledefendant                          -0.082***
##                                        (-0.013)
##
## typeCompany                            -0.255***
##                                        (-0.061)
##
## typeIndividual                         -0.237***
##                                        (-0.056)
##
## typeNGO                                  0.095
##                                        (-0.104)
##
## typeState institution                  -0.678***
##                                        (-0.063)
```

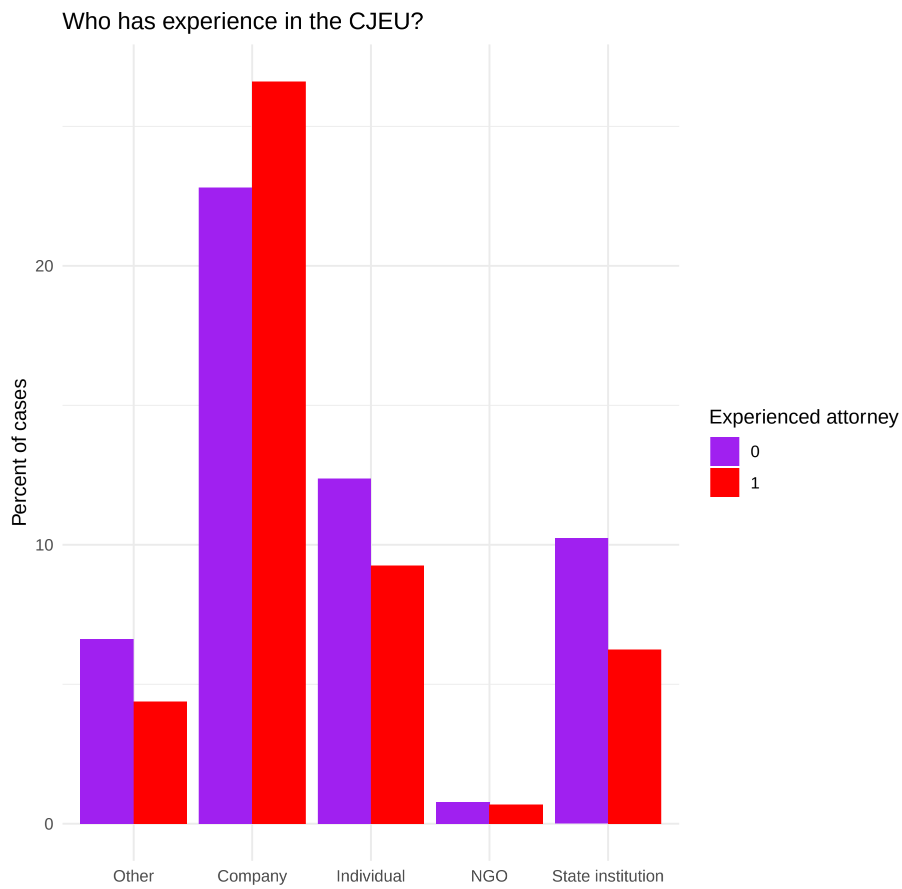
```
##
## type_opponentCompany                      -0.431***
##                                          (-0.062)
##
## type_opponentIndividual                    -0.079
##                                          (-0.072)
##
## type_opponentNGO                           -0.136
##                                          (-0.097)
##
## type_opponentState institution             -0.037
##                                          (-0.069)
##
## member_state12                              0.033
##                                          (-0.052)
##
## member_state15                            0.511***
##                                          (-0.060)
##
## member_state6                               0.017
##                                          (-0.026)
##
## member_state9                              0.070**
##                                          (-0.030)
##
## binary_salience                          -0.675***
##                                          (-0.029)
##
## government_support                        1.087***
##                                           (0.008)
##
## commission_support                        3.165***
##                                          (-0.012)
##
## Constant                                 -0.847***
##                                           (0.178)
##
## ----------------------------------------------------------
## Observations                                1,026
## Log Likelihood                           -464.936
## Akaike Inf. Crit.                          965.872
## ==========================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
stargazer(m2, type = "text",
          se = list(vcov_m2),
```

```r
          style = "all2",
          single.row = TRUE,
          no.space = TRUE,
          font.size = "small",
          align = TRUE,
          dep.var.caption =
            "Binomial logistic regression",
          dep.var.labels = "Favorable ruling",
         keep = c("experience", "role",
                  "type",
                  "type_opponent",
                   "binary_salience",
                  "government_support",
                  "commission_support"),
         covariate.labels = c("Experience",
                              "Defendant",
                              "Company",
                              "Individual", "NGO",
                              "State institution",
                              "Opponent Company",
                              "Opponent Individual",
                              "Opponent NGO",
                              "Opponent State",
                              "Salience",
                              "Government support",
                              "Commission support"
                  ))
```

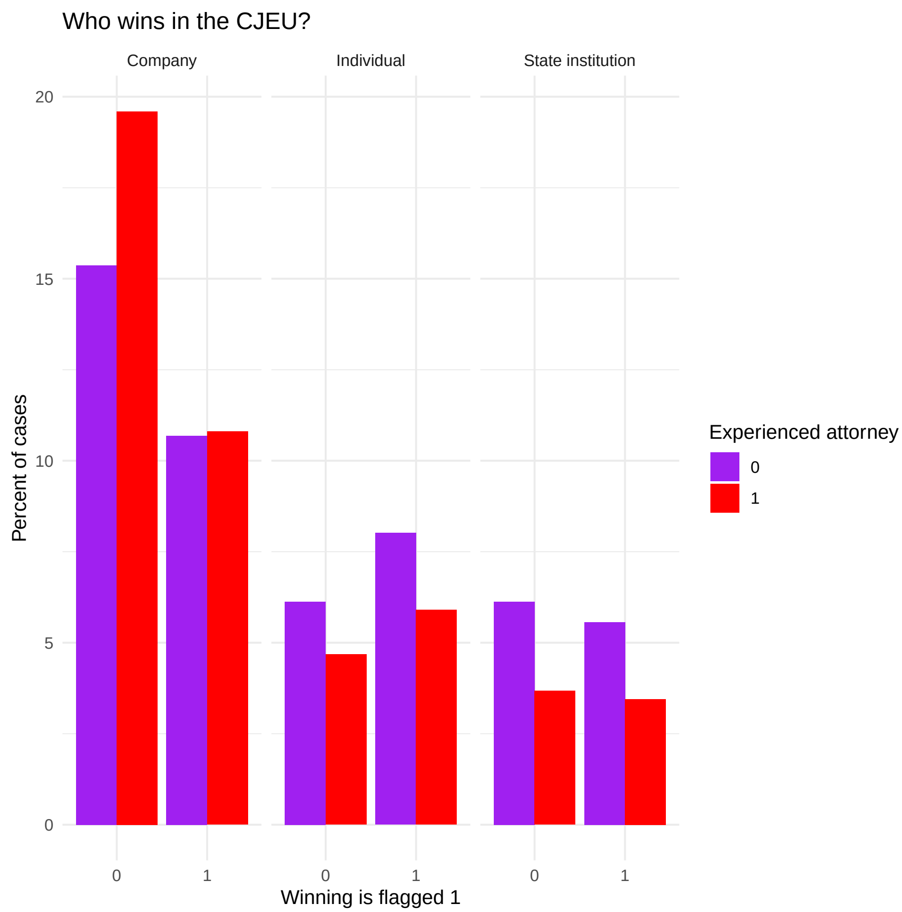
```
##
## ================================================
##                     Binomial logistic regression
##                     ----------------------------
##                           Favorable ruling
## ------------------------------------------------
## Experience                -0.138*** (-0.008)
## Defendant                 -0.082*** (-0.013)
## Company                   -0.255*** (-0.061)
## Individual                -0.237*** (-0.056)
## NGO                         0.095 (-0.104)
## State institution         -0.678*** (-0.063)
## Opponent Company          -0.431*** (-0.062)
## Opponent Individual        -0.079 (-0.072)
## Opponent NGO               -0.136 (-0.097)
## Opponent State             -0.037 (-0.069)
## Salience                  -0.675*** (-0.029)
## Government support         1.087*** (0.008)
```

```
## Commission support       3.165*** (-0.012)
## ------------------------------------------------
## Observations                   1,026
## Log Likelihood                -464.936
## Akaike Inf. Crit.              965.872
## Residual Deviance       929.872 (df = 1008)
## Null Deviance          1,415.121 (df = 1025)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
sample2 %>%
  group_by(celex) %>%
  ggplot(aes(type, fill = as.factor(experience))) +
  geom_bar(aes(y = (..count..)/sum(..count..)*100), position = "dodge") +
  ylab("Percent of cases") +
  labs(fill = "Experienced attorney") +
  scale_fill_manual(values = c( "purple", "red")) +
  ggtitle("Who has experience in the CJEU?")+
  xlab("")+
  theme_minimal()
```




```
sample2 %>%
  group_by(celex) %>%
  filter(type != "NGO" & type != "Other") %>%
  ggplot(aes(as.factor(win), fill = as.factor(experience))) +
  labs(fill = "Experienced attorney")+
  geom_bar(aes(y = (..count..)/sum(..count..)*100), position = "dodge") +
  ylab("Percent of cases") +
  ggtitle("Who wins in the CJEU?")+
  xlab("")+
  theme_minimal() +
  facet_wrap(~ type) +
  xlab("Winning is flagged 1")+
  scale_fill_manual(values = c( "purple", "red"))
```



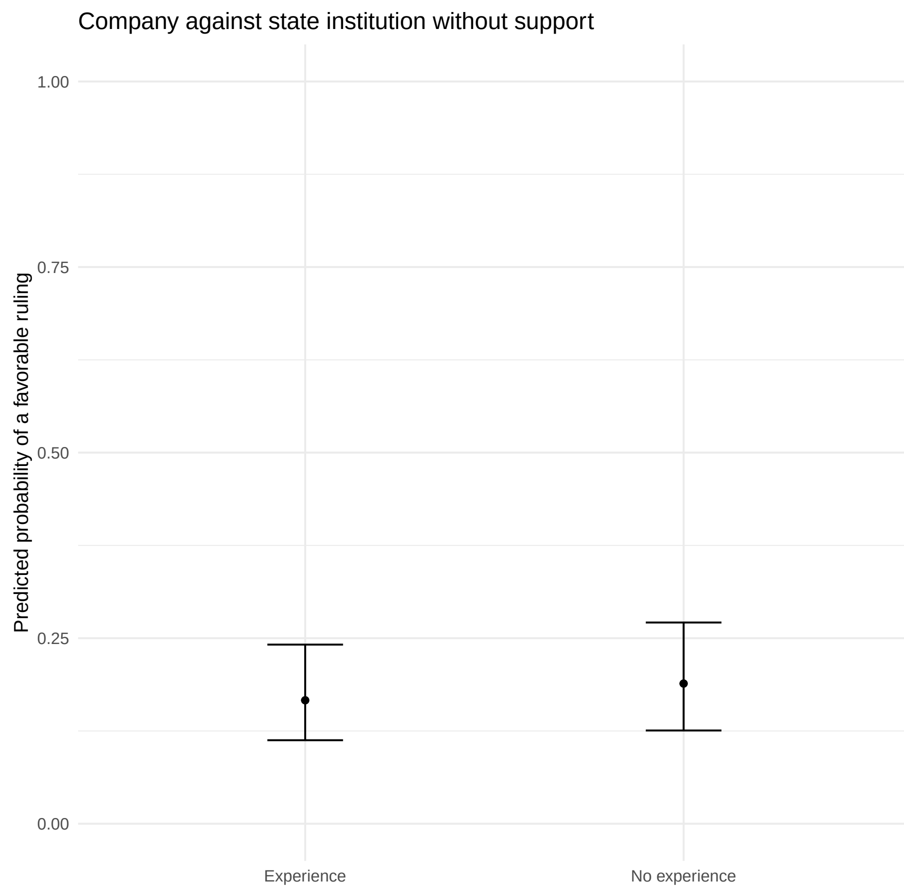

```
table(sample2$type)

##
##            Other          Company   EU institution       Individual
##              113              507                0              222
##              NGO State institution
##               15              169
```

8.3 Plotting effects – creating scenarios

```
# Saving my se's in the model object


m2$cluster_m2 <- vcov_m2
```

```r
# Setting seed
set.seed(24)

simBetas <- mvrnorm(n = 1000,
                    mu = coefficients(m2),
                    Sigma = m2$cluster_m2)


names(coefficients(m2))

##  [1] "(Intercept)"                    "experience"
##  [3] "roledefendant"                  "typeCompany"
##  [5] "typeIndividual"                 "typeNGO"
##  [7] "typeState institution"          "type_opponentCompany"
##  [9] "type_opponentIndividual"        "type_opponentNGO"
## [11] "type_opponentState institution" "member_state12"
## [13] "member_state15"                 "member_state6"
## [15] "member_state9"                  "binary_salience"
## [17] "government_support"             "commission_support"

xMatrix <- cbind(1, #the intercept
                 c(0, 1), # Let experience vary
                 1, # defendant
                 1, # Company
                 0, # Individual
                 0, # NGO
                 0, # State institution
                 0, # opponent company
                 0, # opponent individual
                 0, # opponent NGO
                 1, # opponent state institution
                 0, # M 12
                 0, # M 15
                 0, # M 6,
                 1, # M 9,
                 mean(sample1$binary_salience),
                 0, # Government support,
                 0 # Commission support
                 )


ncol(simBetas) == ncol(xMatrix) #yay!!

## [1] TRUE

### Calculating predicted probabilities: Her multipliserer du simuleringen med xmatrisen
```

```
xBetaMatrix <- xMatrix %*% t(simBetas ) ## this just means x times the betas
predProbs <- 1/(1+exp(-xBetaMatrix)) #This is the predicted probability, for another type of


### Getting point estimates and confidence intervals:
quantileValues <- apply(X = predProbs,  ## read up on the apply() family of functions!
                        MARGIN = 1, ## this means we are applying a function to all the rows
                        FUN =  ## The fun argument defines what I want to do with all the ro
                          ## What we want to do here is to use quantile to get the quantiles
                          ### bounds of the confidence intervals and our point estimates:
                          quantile, probs = c(.05,.5,.95))
quantileValues <- as.data.frame(t(quantileValues))

plotPoints <- cbind(c("No experience", "Experience"),quantileValues)
plotPoints

##   c("No experience", "Experience")        5%       50%       95%
## 1                    No experience 0.1256346 0.1888677 0.2710927
## 2                       Experience 0.1125140 0.1663308 0.2413328

colnames(plotPoints) <- c("Experience", "lower", "estimate", "upper")



ggplot(plotPoints,
       aes(x = Experience,
           y = estimate,
           ymin = lower,
           ymax = upper)) +
  geom_errorbar(width =.2)+
  geom_point()+
  ylim(0,1)+
  ylab("Predicted probability of a favorable ruling")+
  xlab("")+
  theme_minimal() +
  ggtitle("Company against state institution without support")
```



9 Digasnostics

The underlying assumption of the logistical regression model are that (1) the dependent variable is binary; (2) the probability curve is S-shaped and the logit curve is linear; (3) there are no influential observations; (4) there is no multicollinearity among the predictors; (5) there are no empty cells; (6) there is no "complete separation"; (7) no omitted variable bias and (8) the observations are independent and identically distributed.

In this section I go through each of the assumptions and evaluate the model where I control salience, role and type of actor.

The first assumption holds as the dependent variable is binary – 0 for lost case and 1 for won case. The assumption that the observations are independent and identically distributed is not relevant because I am dealing with observations data – and the data at hand is the sample of cases are more or less equal to universe of cases. No omitted variable bias is a theoretical assumption. There are many unobserved factors that may affect not just the assignment to treatment, but also the decision of the Court and the cases that end up in the CJEU in the first place. This is not controlled for because many of the mechanisms that may affect assignment to treatment, cases being referred and the decision of the CJEU are unobservable.

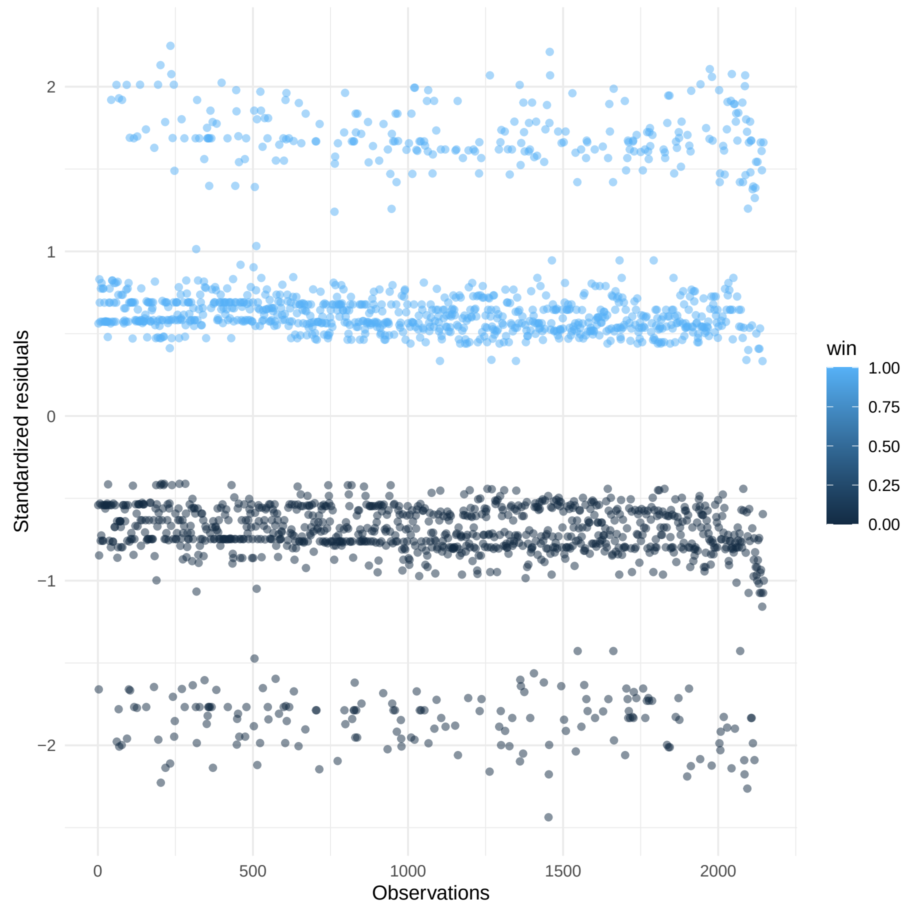
9.0.1 The regression has the shape of an S

In order to investigate if the second assumption holds I make sure that the relationship between the independent variables and the logit-outcome is linear. To show this graphically is difficult when the variables are characters.

9.0.2 Influential values

Checking for influential values. Plotting the standardized residuals. The standard normal distribution lies between -4 and 4. Values above three indicates outliers and should be further investigated as they might affect the results.

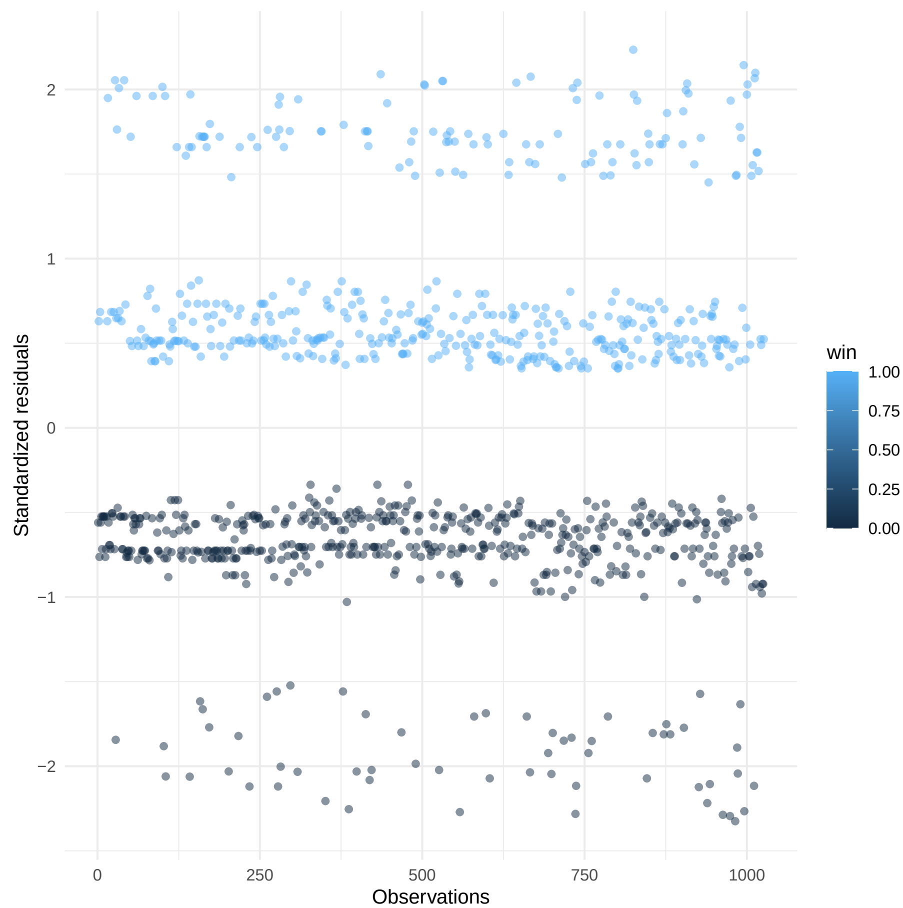
```
# Sample1 model
model.data <- augment(law) %>%
  mutate(index = 1:n())

model.data %>%
  top_n(3, .cooksd)

## # A tibble: 4 x 16
##     win lawyer role  type  type_opponent member_state binary_salience
##   <dbl>  <dbl> <chr> <fct> <fct>         <chr>                  <dbl>
## 1     1      0 appl~ Stat~ NGO           9                          0
## 2     0      1 appl~ Indi~ NGO           12                         1
## 3     1      1 defe~ Comp~ NGO           12                         1
## 4     0      1 appl~ Indi~ NGO           12                         1
## # ... with 9 more variables: government_support <dbl>,
## #   commission_support <dbl>, .fitted <dbl>, .resid <dbl>, .std.resid <dbl>,
## #   .hat <dbl>, .sigma <dbl>, .cooksd <dbl>, index <int>

ggplot(model.data, aes(index, .std.resid))+
  geom_point(aes(color = win), alpha = .5) +
  theme_minimal() +
  ylab("Standardized residuals") +
  xlab("Observations")
```




```
# Manual count
model.data %>%
  filter(abs(.std.resid) > 3)

## # A tibble: 0 x 16
## # ... with 16 variables: win <dbl>, lawyer <dbl>, role <chr>, type <fct>,
## #   type_opponent <fct>, member_state <chr>, binary_salience <dbl>,
## #   government_support <dbl>, commission_support <dbl>, .fitted <dbl>,
## #   .resid <dbl>, .std.resid <dbl>, .hat <dbl>, .sigma <dbl>, .cooksd <dbl>,
## #   index <int>

# Sample2 model
model.data <- augment(m2) %>%
  mutate(index = 1:n())

model.data %>%
  top_n(3, .cooksd)
```

```
## # A tibble: 3 x 16
##     win experience role  type  type_opponent member_state binary_salience
##   <dbl>      <dbl> <chr> <fct> <fct>         <chr>                  <dbl>
## 1     0          0 appl~ NGO   Company       12                         1
## 2     1          1 defe~ NGO   Individual    12                         1
## 3     0          1 defe~ Indi~ NGO           12                         0
## # ... with 9 more variables: government_support <dbl>,
## #   commission_support <dbl>, .fitted <dbl>, .resid <dbl>, .std.resid <dbl>,
## #   .hat <dbl>, .sigma <dbl>, .cooksd <dbl>, index <int>
```

```
ggplot(model.data, aes(index, .std.resid))+
  geom_point(aes(color = win), alpha = .5) +
  theme_minimal() +
  ylab("Standardized residuals") +
  xlab("Observations")
```



```
# Manual count
model.data %>%
  filter(abs(.std.resid) > 3)

## # A tibble: 0 x 16
## # ... with 16 variables: win <dbl>, experience <dbl>, role <chr>, type <fct>,
## #   type_opponent <fct>, member_state <chr>, binary_salience <dbl>,
## #   government_support <dbl>, commission_support <dbl>, .fitted <dbl>,
## #   .resid <dbl>, .std.resid <dbl>, .hat <dbl>, .sigma <dbl>, .cooksd <dbl>,
## #   index <int>
```

9.0.3 Multicollinearity

VIF-test measures how much of the variance in each independent variable can be explained by the other variables in the analysis. As a general rule of thumb a VIF-value under 5 indicates no multicolinearity, whilst values between 5 and 10 is considered no ideal, but yet not very problematic (Hermansen, 2019, p. 195). Values above 10 indicates strong multicolinearity(ibid).

```
vif(law)

##                             GVIF Df GVIF^(1/(2*Df))
## lawyer                  2.145390  1        1.464715
## role                    1.591941  1        1.261722
## type                    3.076044  4        1.150798
## type_opponent           2.958365  4        1.145200
## member_state            1.226589  4        1.025858
## binary_salience         3.689295  1        1.920754
## government_support      2.185847  1        1.478461
## commission_support      1.155687  1        1.075029
## lawyer:binary_salience  3.174935  1        1.781835

vif(m2)

##                         GVIF Df GVIF^(1/(2*Df))
## experience          1.034774  1        1.017238
## role                1.522512  1        1.233901
## type                3.267072  4        1.159498
## type_opponent       3.046637  4        1.149417
## member_state        1.356962  4        1.038893
## binary_salience     1.699707  1        1.303728
## government_support  1.573692  1        1.254469
## commission_support  1.169705  1        1.081529
```

Running the VIF-test I find that all my independent variables have VIF-values between 1 and 3 which indicates no multicolinearity.

9.0.4 Complete seperation

Checking for complete separation is easily done by plotting the data. The plot below indicates that also this assumption is met. Checking for empty cells is unecessary are observations with missing values are not included in the model.

9.1 Goodness of fit

9.1.1 McFadden's pseudo R2

McFadden's pseudo R2 is a measure that compares the log-likelihood value for my model and compares it to the log-likelihood value of a model without any variables – an intercept-only model (Christiphersen, 2013, p. 139). The value ranges from zero to one. Values closer to 1 indicates good predictive power. Values closer to zero indicates no predictive power. The results show that the model is better than an intercept-only model, however, the model does not explain much of the variation. The models with more variables have slightly higher McFadden scores.

```
PseudoR2(law, which = c("McFadden", "AIC", "BIC", "logLik"))

##      McFadden           AIC           BIC        logLik
##     0.3003144  2120.4866403  2228.2513495 -1041.2433202

PseudoR2(law)

##  McFadden
## 0.3003144

PseudoR2(m2, which = c("McFadden", "AIC", "BIC", "logLik"))

##     McFadden          AIC          BIC       logLik
##    0.3429027  965.8721715 1054.6737860 -464.9360858

PseudoR2(m2)

##  McFadden
## 0.3429027
```

9.1.2 Hosmler-Lemeshow-test

Tests ho good the model fits the data by comparing observed and predicted values – meaning that it compares the observed, real values of 1 and + to the models fitted values (ibid). The test does this by comparing subgroups of the population estimated. The Hosmer-Lemeshow-test is not supposed to give significant results, because this means that the model is not a good fit for the data. The results from running the test are not significant suggesting that the model is good at describing the data.

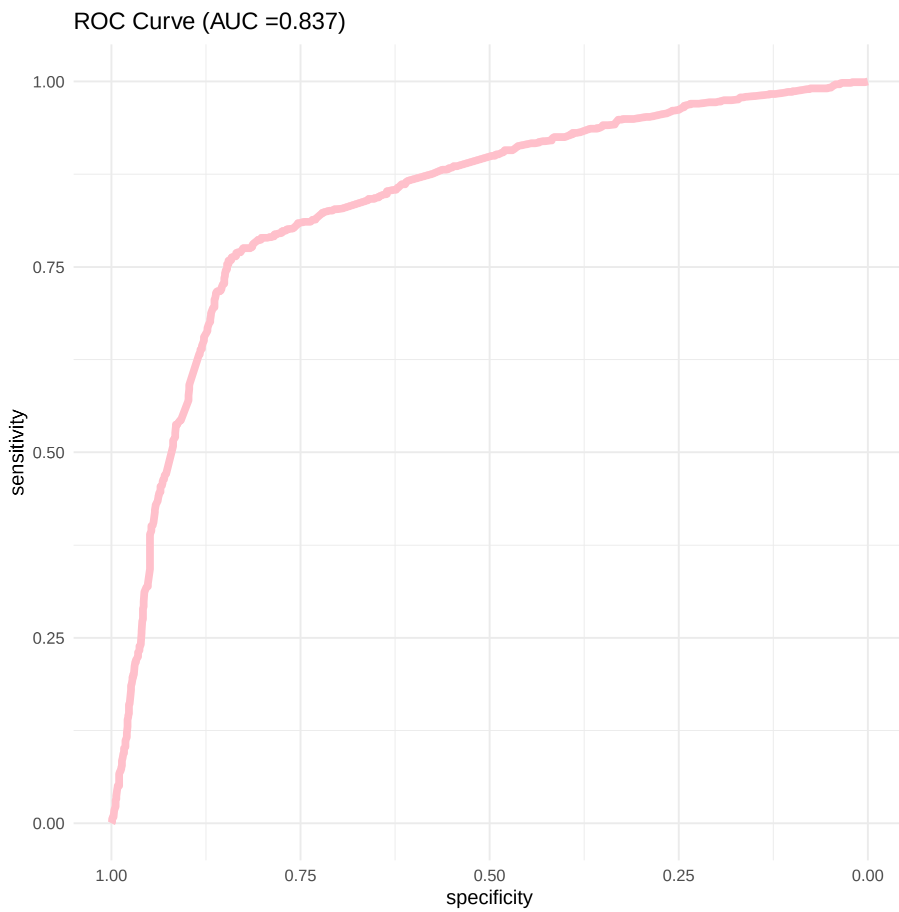
```r
############# Sample1 model
hl <- hoslem.test(law$y,
                  fitted(law),
                  g = 10)
# G= 10 ten subgroups
hl

##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  law$y, fitted(law)
## X-squared = 3.5965, df = 8, p-value = 0.8916

# Shows difference in
# observed and expected Y-values
# for ten subgroups
cbind(hl$expected, hl$observed)

##                      yhat0     yhat1  y0  y1
## [0.0809,0.145] 194.08290  27.91710 195  27
## (0.145,0.187]  176.66206  35.33794 176  36
## (0.187,0.243]  209.53455  60.46545 212  58
## (0.243,0.264]  117.42665  39.57335 113  44
## (0.264,0.314]  153.53432  60.46568 155  59
## (0.314,0.766]   90.39424 123.60576  89 125
## (0.766,0.81]    46.16726 172.83274  51 168
## (0.81,0.846]    36.25776 172.74224  33 176
## (0.846,0.867]   33.35010 198.64990  29 203
## (0.867,0.946]   21.59016 176.40984  26 172

############# Sample2 model
hl <- hoslem.test(law$y,
                  fitted(law),
                  g = 10)
# G= 10 ten subgroups
hl

##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  law$y, fitted(law)
## X-squared = 3.5965, df = 8, p-value = 0.8916

# Shows difference in
# observed and expected Y-values
# for ten subgroups
cbind(hl$expected, hl$observed)
```

```
##                      yhat0     yhat1  y0  y1
## [0.0809,0.145] 194.08290  27.91710 195  27
## (0.145,0.187]  176.66206  35.33794 176  36
## (0.187,0.243]  209.53455  60.46545 212  58
## (0.243,0.264]  117.42665  39.57335 113  44
## (0.264,0.314]  153.53432  60.46568 155  59
## (0.314,0.766]   90.39424 123.60576  89 125
## (0.766,0.81]    46.16726 172.83274  51 168
## (0.81,0.846]    36.25776 172.74224  33 176
## (0.846,0.867]   33.35010 198.64990  29 203
## (0.867,0.946]   21.59016 176.40984  26 172
```

9.1.3 How well does my model predict?

I am modelling predicted probabilitites. To estimate how well my models predict, a ROC-cruve can be helpful (Receiving Operating Characteristics). When using logistical regression, the goal is a model that predicts the outcome of the indepdent variable corectly at all times. The ROC-curve shows how well my model predics by determining the relationship between true positive values (the predictions my model predics as 1 that is observed to be 1) and false positive values (the prediction my model predicts as 1 but is actually 0) using various cut-off values. I create a ROC-curve to evaluate the overall performance of my model. The ROC-curve defines the optimal cut-off value for me – indicating at which point from 0 to 100 my model predicts correctly. Is my model correct in 70 percent of all the instances? Is my model correct in 60 percent of all instances?

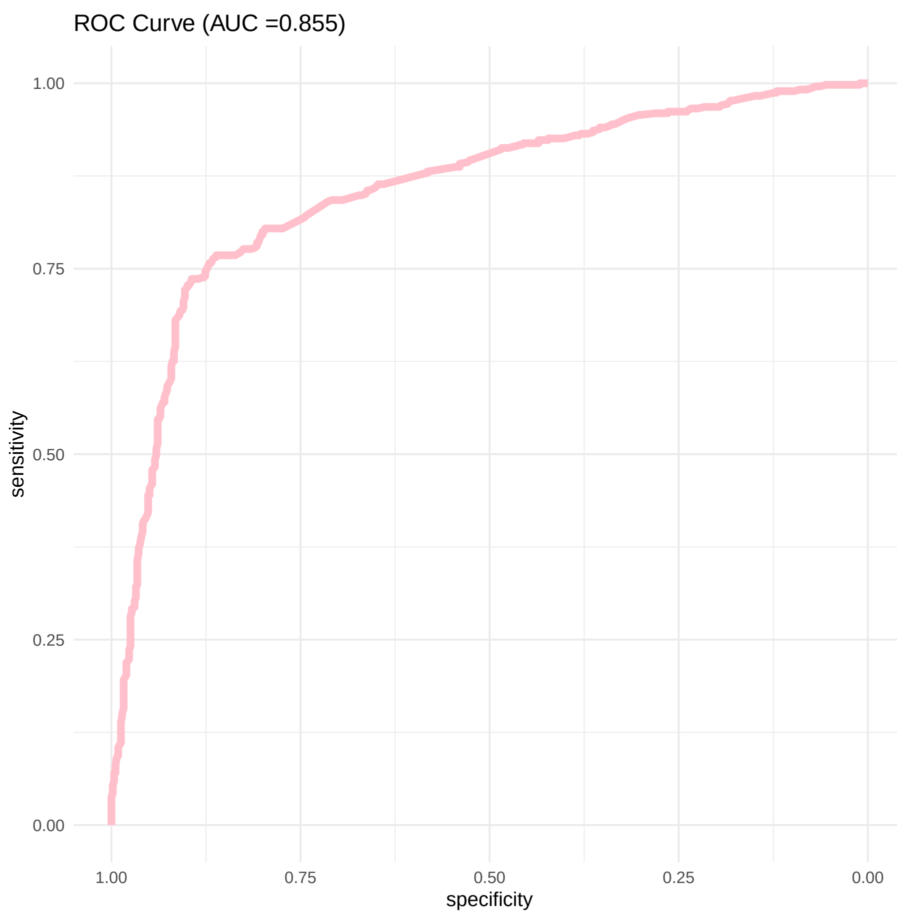
```
# Sample1 model
preds <- predict(law,
                 sample1,
                 type = "response")

roc_obj <- roc(sample1$win, preds)

## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

auc_m2 <- auc(roc_obj)

auc_m2 <- round(auc_m2, digits = 3)

law_roc <- ggroc(roc_obj, color="pink", size = 2) +
  ggtitle(paste0("ROC Curve ", "(AUC =", auc_m2, ")")) +
  theme_minimal()

law_roc
```




```r
test <- table(predicted = ifelse(preds > auc_m2, 1, 0),
              observed = sample1$win)

test

##           observed
## predicted    0    1
##          0 1022  641
##          1   57  427

# Sample3 model
preds <- predict(m2,
                 sample2,
                 type = "response")

roc_obj <- roc(sample2$win, preds)
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
auc_m2 <- auc(roc_obj)

auc_m2 <- round(auc_m2, digits = 3)

m2_roc <- ggroc(roc_obj, color="pink", size = 2) +
  ggtitle(paste0("ROC Curve ", "(AUC =", auc_m2, ")")) +
  theme_minimal()

m2_roc
```




```
test <- table(predicted = ifelse(preds > auc_m2, 1, 0),
              observed = sample2$win)
```

```
test

##          observed
## predicted   0   1
##         0 527 255
##         1  29 215
```

Area under the curve (AUC) equal to 1 means that the model makes perfect predictions, meaning that the model predicts Y = 1 when Y=1 is observed in all incidents. The model predicts correctly in approximately 84 percent of all incidents. The model I have made is able to classify Y = 1 and Y = 0 correctly in approximately 84 percent of the time.