PLSC 30600 - Lab 3 - Analyzing Experiments

01/20/2023

Analyzing an experiment: Madsen et. al. (2021)

Facts

- This lab will have you analyze and present results from an experiment conducted by Madsen, Mayoral, Strezhnev and Voeten (2021) examining attitudes towards international courts among voters in European countries.
- The experiment presented respondents with a vignette about a hypothetical court case in a European country that was then heard by a European court (the vignette was ambiguous, but one could imagine either the European Court of Justice or the European Court of Human Rights).
- The study manipulated two treatments: the outcome of the case and whether or not the European court overruled the domestic court.
- Three outcomes were observed: whether respondents agreed with the decision, whether they thought the decision should be implemented, and whether they supported the continued authority of European courts over domestic courts.
- The goal was to evaluate two competing hypotheses of what drives populist opposition to international courts: do they oppose courts because of a concern over national sovereignty (and so would support a domestic court making the same decision) or because of the content of the decisions (e.g. right-wing populists opposing pro-immigration policies).
- In other words, do controversial judgments get more support if the decision comes from a domestic court rather than from a European one. The experiment allowed the researchers to vary these attributes independently of one another in a hypothetical scenario.

More details and example

- The paper fielded multiple vignettes but here we'll focus on the immigration vignette that was fielded in all five countries in the experiment: Denmark, UK, France, Spain and Poland.
- In the vignette, respondents were asked to consider a case where the government of their home country was considering deporting a foreigner who was convicted of a crime.
- The foreigner challenged the deportation in a UK court on the grounds that deportation violated his human rights.
- Two factors were randomly varied: whether the domestic court ruled in favor of the foreigner or in favor of the government and then whether the European court overruled or agreed with the domestic court.
- This was therefore a 2x2 "factorial" design with four unique treatment combinations. The UK vignette, for example, stated:

Suppose that United Kingdom (UK) authorities decided to deport a foreigner who has been convicted of a crime. The foreigner appealed at a UK court that the decision to deport him violated his human rights. The UK court found that the authorities [CAN/CANNOT] deport the foreigner. The question was then brought before a European court, which [AGREED WITH/DISAGREED WITH] the UK court. The final decision is that the foreigner should [REMAIN IN THE UK/BE DEPORTED].

- We'll estimate the effects of whether the foreigner was deported vs. not deported ("outcome treatment") and the effect of whether the European court deferred to or overruled the domestic court ("deferrence treatment").
- Respondents were then asked 3 questions:

1. Do you agree or disagree with the final decision?
2. Do you agree or disagree that the final decision should be implemented?
3. Do you agree or disagree that the country should continue to accept the authority of European courts?

- All responses were on a 6-level scale (strongly disagree to strongly agree).
- In the pre-registration the analysis was to be conducted on the dichotomized "agree"/"disagree" outcome, so we'll focus on estimating effects on that indicator, though we'll also look at how to visualize treatment effects with multi-level or continuous outcomes.
- In practice, pre-registered "scales" combining multiple outcomes are increasingly common in recent experiments as they can be much less noisy than individual questions.
- Pre-registration avoids the risk of p-hacking and multiple testing in choosing how to create that scale.

Below is the code to read in the data

```r
### Read in dataset
data <- read_csv("final_data_clean.csv")
View(data)
```

- The analysis pre-registered four covariates that would be used to create strata for covariate adjustment: Age, Gender, Education and Country.
- For some experiments (the smaller ones run in only two countries), strata omitting age were required because they were otherwise too small and did not contain enough treated or control units.
- We also wrote a function for the stratified estimator – literally just taking the difference in means in each sub-group and averaging up to get the point and variance estimates for the ATE. Don't necessarily need to use this function, but can be useful to see the process broken down into parts in the code.

```r
# Stratified regression estimator
strat_reg <- function(formula, data, stratum){

  # Raw counts (use this do diagnostics if things go wrong)
  counts <- data %>% group_by(stratum) %>% summarize(N=n())

  # within each level of the data, fit the model in "formula", lm_robust is linear regression with robu
  stratum_regs <- data %>% group_by(stratum) %>% group_map(~lm_robust(formula=formula, data=.))

  # get sizes of each stratum
```

```
  sample_sizes <- unlist(lapply(stratum_regs, function(x) x$nobs)) #nobs for most recent version of est

  sample_shares <- sample_sizes/sum(sample_sizes)

  # get point estimates
  point_est <- sapply(stratum_regs, function(x) x$coefficients)

  if (!is.matrix(point_est)&is.vector(point_est)){
    point_est <- t(as.matrix(point_est))
  }
  # check for NAs
  if(sum(apply(point_est, 1, function(x) sum(is.na(x)))) > 0){
    print(counts[apply(point_est, 2, function(x) sum(is.na(x))) != 0,])
    stop("Error: NAs in stratified point estimates, coarsen strata to obtain enough units in each treat
  }

  var_est <- sapply(stratum_regs, function(x) abs(diag(vcov(x)))) # diagonals must be positive, some nu

  if (!is.matrix(var_est)&is.vector(var_est)){
    var_est <- t(as.matrix(var_est))
  }
    # check for NAs
  if(sum(apply(var_est, 1, function(x) sum(is.na(x)))) > 0){
    print(counts[apply(var_est, 2, function(x) sum(is.na(x))) != 0,])
    stop("Error: NAs in stratified variance estimates, coarsen strata to obtain enough units in each tre
  }


  point_combined <- apply(point_est, 1, function(x) sum(x*sample_shares))
  se_combined <- apply(var_est, 1, function(x) sqrt(sum(x*sample_shares^2)))

  # Fix names for 1-length vectors
  if(length(point_combined) < 2){
    char_names <- c("(Mean)")
  }else{
    char_names <- names(point_combined)
  }

  # Save the results
  out_results <- data.frame(term = as.character(char_names), estimate=point_combined, std.error = se_co
  out_results$statistic <- out_results$estimate/out_results$std.error
  out_results$conf.low <- out_results$estimate - qnorm(.975)*out_results$std.error
  out_results$conf.high <- out_results$estimate + qnorm(.975)*out_results$std.error

  return(out_results)
}
```

Estimating the ATEs

Our outcome variables are

- `D1Agree` (Binary agree/disagree with decision)

- `D2Agree` (Binary agree/disagree with implementing the decision)
- `D3Agree` (Binary agree/disagree on European courts should have authority)

Our two treatments are

- `caseOutcome` (0 if foreigner deported, 1 if not deported)
- `caseDefer` (0 if European court overruled domestic, 1 if European court deferred)

We'll subset the data down to the immigration vignette. One last small thing is that the immigration vignette also had some number of observations assigned to a "no court" treatment that omitted any European court involvement – since the sample size was already large, there was room to include this "placebo" treatment to see if even mentioning a European court affected attitudes. For today we'll ignore that treatment

```
# Only the immigration vignette
data_immigration <- data %>% filter(vignette == "immigration"&judgment!="nocourt")

# Any missing responses: a tiny proportion of observations was missing
sum(is.na(data_immigration$D1Agree))
```

```
## [1] 4
```

```
sum(is.na(data_immigration$D2Agree))
```

```
## [1] 6
```

```
sum(is.na(data_immigration$D3Agree))
```

```
## [1] 4
```

```
# Negligible missingness (only about 6 obs!) -- not really a problem to drop them - very likely missing
data_immigration <- data_immigration %>% filter(!is.na(D1Agree)&!is.na(D2Agree)&!is.na(D3Agree))

# Summary stats - sample size
data_immigration %>% group_by(country, caseOutcome, caseDefer) %>% summarize(N=n())
```

```
## # A tibble: 20 x 4
## # Groups:   country, caseOutcome [10]
##    country caseOutcome caseDefer     N
##    <chr>         <dbl>     <dbl> <int>
##  1 denmark           0         0   259
##  2 denmark           0         1   249
##  3 denmark           1         0   256
##  4 denmark           1         1   248
##  5 france            0         0   255
##  6 france            0         1   251
##  7 france            1         0   251
##  8 france            1         1   253
##  9 poland            0         0   252
## 10 poland            0         1   249
## 11 poland            1         0   249
```

```
## 12 poland           1         1   255
## 13 spain            0         0   254
## 14 spain            0         1   255
## 15 spain            1         0   257
## 16 spain            1         1   253
## 17 uk               0         0   254
## 18 uk               0         1   255
## 19 uk               1         0   248
## 20 uk               1         1   249
```

Let's estimate the ATE of the case outcome and the deference treatment on support for implementation (D2Agree) (without covariate adjustment)

```
# Remember: `D2Agree` (Binary agree/disagree with implementing the decision)

# Outcome treatment (effect of immigrant "remaining" in country)
ate_outcome <- lm_robust(D2Agree ~ caseOutcome, data=data_immigration)
summary(ate_outcome)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseOutcome, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  0.84248    0.00724 116.370 0.000e+00   0.8283  0.85667 5050
## caseOutcome -0.08702    0.01122  -7.759 1.027e-14  -0.1090 -0.06503 5050
##
## Multiple R-squared:  0.01179 ,   Adjusted R-squared:  0.0116
## F-statistic: 60.21 on 1 and 5050 DF,  p-value: 1.027e-14
```

```
# Court deference treatment (effect of european court agreeing with domestic)
ate_defer <- lm_robust(D2Agree ~ caseDefer, data=data_immigration)
summary(ate_defer)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseDefer, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)  CI Lower CI Upper   DF
## (Intercept)  0.78777   0.008123  96.984  0.00000 0.7718472  0.80370 5050
## caseDefer    0.02272   0.011271   2.016  0.04389 0.0006223  0.04481 5050
##
## Multiple R-squared:  0.0008036 , Adjusted R-squared:  0.0006058
## F-statistic: 4.063 on 1 and 5050 DF,  p-value: 0.04389
```

We might be interested in the effect of one treatment holding the other constant. For that, we include both indicators + an interaction (again, still a fully-saturated model, so it's still just differences in means, but we have 4 means instead of two).

```r
# Outcome treatment (effect of immigrant "remaining" in country)
joint_model <- lm_robust(D2Agree ~ caseOutcome + caseDefer + caseOutcome*caseDefer, data=data_immigrati
summary(joint_model)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseOutcome + caseDefer + caseOutcome *
##     caseDefer, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##                       Estimate Std. Error t value  Pr(>|t|)   CI Lower CI Upper
## (Intercept)            0.82810    0.01057 78.3102 0.000e+00  0.8073696  0.84883
## caseOutcome           -0.08107    0.01618 -5.0107 5.610e-07 -0.1127949 -0.04935
## caseDefer              0.02893    0.01446  2.0000 4.556e-02  0.0005721  0.05729
## caseOutcome:caseDefer -0.01204    0.02242 -0.5372 5.912e-01 -0.0559982  0.03191
##                         DF
## (Intercept)           5048
## caseOutcome           5048
## caseDefer             5048
## caseOutcome:caseDefer 5048
##
## Multiple R-squared:  0.01267 ,   Adjusted R-squared:  0.01208
## F-statistic:  22.2 on 3 and 5048 DF,  p-value: 2.816e-14
```

How do we interpret the baseline and interaction terms – what particular effect does the lower-order coefficient on correspond to (and why is it not the ATE)? How do we interpret the interaction term? What do we conclude about whether court deference moderates the effect of the case outcome treatment?

```r
# Great trick if you want a different interpretation for the lower order terms - just re-level the thin
joint_model_2 <- lm_robust(D2Agree ~ caseOutcome + I(1-caseDefer) + caseOutcome*I(1-caseDefer), data=da
summary(joint_model_2)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseOutcome + I(1 - caseDefer) +
##     caseOutcome * I(1 - caseDefer), data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##                              Estimate Std. Error t value  Pr(>|t|) CI Lower
## (Intercept)                   0.85703   0.009869 86.8390 0.000e+00  0.83768
## caseOutcome                  -0.09312   0.015520 -5.9998 2.113e-09 -0.12354
## I(1 - caseDefer)             -0.02893   0.014465 -2.0000 4.556e-02 -0.05729
## caseOutcome:I(1 - caseDefer)  0.01204   0.022421  0.5372 5.912e-01 -0.03191
##                               CI Upper   DF
## (Intercept)                  0.8763773 5048
## caseOutcome                 -0.0626920 5048
## I(1 - caseDefer)            -0.0005721 5048
## caseOutcome:I(1 - caseDefer)  0.0559982 5048
```

```
##
## Multiple R-squared:  0.01267 ,   Adjusted R-squared:  0.01208
## F-statistic:  22.2 on 3 and 5048 DF,  p-value: 2.816e-14
```

What if we were to look for heterogeneity in the effect of remain vs. deport by country?

```r
# Outcome treatment (effect of immigrant "remaining" in country)
ate_outcome_country <- lm_robust(D2Agree ~ caseOutcome + country + caseOutcome*country, data=data_immig
summary(ate_outcome_country)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseOutcome + country + caseOutcome *
##     country, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##                            Estimate Std. Error t value  Pr(>|t|) CI Lower
## (Intercept)                 0.89961    0.01335  67.403 0.000e+00  0.87344
## caseOutcome                -0.15159    0.02351  -6.447 1.247e-10 -0.19769
## countryfrance              -0.06166    0.02114  -2.916 3.557e-03 -0.10311
## countrypoland              -0.10320    0.02241  -4.604 4.245e-06 -0.14714
## countryspain               -0.09607    0.02211  -4.345 1.421e-05 -0.13942
## countryuk                  -0.02534    0.01986  -1.276 2.020e-01 -0.06428
## caseOutcome:countryfrance -0.03754    0.03570  -1.052 2.931e-01 -0.10754
## caseOutcome:countrypoland  0.12106    0.03512   3.447 5.723e-04  0.05220
## caseOutcome:countryspain   0.15198    0.03425   4.437 9.326e-06  0.08482
## caseOutcome:countryuk      0.08819    0.03284   2.686 7.265e-03  0.02381
##                            CI Upper   DF
## (Intercept)                 0.92577 5042
## caseOutcome                -0.10549 5042
## countryfrance              -0.02021 5042
## countrypoland              -0.05926 5042
## countryspain               -0.05272 5042
## countryuk                   0.01360 5042
## caseOutcome:countryfrance   0.03245 5042
## caseOutcome:countrypoland   0.18991 5042
## caseOutcome:countryspain    0.21913 5042
## caseOutcome:countryuk       0.15257 5042
##
## Multiple R-squared:  0.02727 ,   Adjusted R-squared:  0.02554
## F-statistic: 15.56 on 9 and 5042 DF,  p-value: < 2.2e-16
```

What's the interpretation of the lower order coefficient on caseOutcome? Which of the five countries is "left out"? How would we get the effect for Spain?

```r
# Easiest trick is to "re-level" the country factor in R
data_immigration$country2 <- as.factor(data_immigration$country)
data_immigration$country2 <- relevel(data_immigration$country2, "spain")

ate_outcome_country2 <- lm_robust(D2Agree ~ caseOutcome + country2 + caseOutcome*country2, data=data_imm
summary(ate_outcome_country2)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseOutcome + country2 + caseOutcome *
##     country2, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##                              Estimate Std. Error  t value  Pr(>|t|) CI Lower
## (Intercept)                 0.8035363    0.01763 45.58201 0.000e+00  0.76898
## caseOutcome                 0.0003852    0.02491  0.01547 9.877e-01 -0.04845
## country2denmark             0.0960700    0.02211  4.34490 1.421e-05  0.05272
## country2france              0.0344083    0.02408  1.42915 1.530e-01 -0.01279
## country2poland             -0.0071292    0.02520 -0.28290 7.773e-01 -0.05653
## country2uk                  0.0707269    0.02296  3.08047 2.078e-03  0.02572
## caseOutcome:country2denmark -0.1519756    0.03425 -4.43677 9.326e-06 -0.21913
## caseOutcome:country2france  -0.1895204    0.03664 -5.17279 2.396e-07 -0.26135
## caseOutcome:country2poland  -0.0309194    0.03607 -0.85715 3.914e-01 -0.10164
## caseOutcome:country2uk      -0.0637833    0.03385 -1.88411 5.961e-02 -0.13015
##                             CI Upper   DF
## (Intercept)                 0.838096 5042
## caseOutcome                 0.049217 5042
## country2denmark             0.139417 5042
## country2france              0.081608 5042
## country2poland              0.042274 5042
## country2uk                  0.115738 5042
## caseOutcome:country2denmark -0.084824 5042
## caseOutcome:country2france  -0.117694 5042
## caseOutcome:country2poland   0.039798 5042
## caseOutcome:country2uk       0.002584 5042
##
## Multiple R-squared:  0.02727 ,   Adjusted R-squared:  0.02554
## F-statistic: 15.56 on 9 and 5042 DF,  p-value: < 2.2e-16
```

Adjusting for covariates

Let's compare the covariate-adjusted to the non-covariate adjusted estimates. First, using our strata

```
## Unadjusted
# Outcome treatment (effect of immigrant "remaining" in country)
ate_outcome <- lm_robust(D2Agree ~ caseOutcome, data=data_immigration)
summary(ate_outcome)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseOutcome, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  0.84248    0.00724 116.370 0.000e+00   0.8283  0.85667 5050
```

```
## caseOutcome -0.08702    0.01122  -7.759 1.027e-14  -0.1090 -0.06503 5050
##
## Multiple R-squared:  0.01179 ,   Adjusted R-squared:  0.0116
## F-statistic: 60.21 on 1 and 5050 DF,  p-value: 1.027e-14
```

```r
## Adjusted
ate_outcome_adj <- strat_reg(D2Agree ~ caseOutcome, data=data_immigration, stratum=data_immigration$stra
ate_outcome_adj
```

```
##                    term    estimate   std.error  statistic   conf.low
## (Intercept) (Intercept)  0.84050783 0.007293655 115.238228  0.8262125
## caseOutcome caseOutcome -0.08463942 0.011223195  -7.541472 -0.1066365
##               conf.high
## (Intercept)  0.85480313
## caseOutcome -0.06264236
```

```r
## Lin estimator is the same!
ate_outcome_lin <- lm_lin(D2Agree ~ caseOutcome,
        covariates = ~stratum, data=data_immigration)
tidy(ate_outcome_lin) %>% filter(term == "caseOutcome")
```

```
##          term    estimate std.error statistic      p.value   conf.low
## 1 caseOutcome -0.08463942 0.0112232 -7.541472 5.499244e-14 -0.1066419
##     conf.high   df outcome
## 1 -0.06263696 4932 D2Agree
```

```r
# Court deference treatment (effect of european court agreeing with domestic)
ate_defer <- lm_robust(D2Agree ~ caseDefer, data=data_immigration)
summary(ate_defer)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseDefer, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)  CI Lower CI Upper   DF
## (Intercept)  0.78777   0.008123  96.984  0.00000 0.7718472  0.80370 5050
## caseDefer    0.02272   0.011271   2.016  0.04389 0.0006223  0.04481 5050
##
## Multiple R-squared:  0.0008036 , Adjusted R-squared:  0.0006058
## F-statistic: 4.063 on 1 and 5050 DF,  p-value: 0.04389
```

```r
ate_defer_adj <- strat_reg(D2Agree ~ caseDefer, data=data_immigration, stratum=data_immigration$stratum
ate_defer_adj
```

```
##                    term   estimate   std.error statistic  conf.low  conf.high
## (Intercept) (Intercept) 0.78603022 0.008209586 95.745415 0.7699397 0.80212072
## caseDefer     caseDefer 0.02453926 0.011347536  2.162519 0.0022985 0.04678002
```

Our strata didn't really do that much. This is probably because the covariates didn't really explain the outcome particularly well and at the sample sizes in this experiment (approximately 5k for the immigration study), additional marginal improvements in variance are hard to get without **very** predictive covariates.

```r
# Outcome treatment (effect of immigrant "remaining" in country)
ate_outcome <- lm_robust(D2Agree ~ caseDefer, data=data_immigration)
summary(ate_outcome)
```

```
##
## Call:
## lm_robust(formula = D2Agree ~ caseDefer, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)  CI Lower CI Upper   DF
## (Intercept)  0.78777   0.008123  96.984  0.00000 0.7718472  0.80370 5050
## caseDefer    0.02272   0.011271   2.016  0.04389 0.0006223  0.04481 5050
##
## Multiple R-squared:  0.0008036 , Adjusted R-squared:  0.0006058
## F-statistic: 4.063 on 1 and 5050 DF,  p-value: 0.04389
```

```r
# Basically same results with an additive model instead of a fully interacted one (strata)
ate_outcome_adjusted <- lm_lin(D2Agree ~ caseDefer, covariates = ~ age + college + country + gender, da
summary(ate_outcome_adjusted)
```

```
##
## Call:
## lm_lin(formula = D2Agree ~ caseDefer, covariates = ~age + college +
##     country + gender, data = data_immigration)
##
## Standard error type:  HC2
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   0.7877706  0.0081036 97.2119 0.000000
## caseDefer                     0.0228877  0.0112188  2.0401 0.041390
## age_c                        -0.0008083  0.0005120 -1.5787 0.114459
## collegeNon-College_c          0.0295978  0.0171665  1.7242 0.084740
## countryfrance_c              -0.0514773  0.0261727 -1.9668 0.049257
## countrypoland_c              -0.0474013  0.0260619 -1.8188 0.069002
## countryspain_c               -0.0034408  0.0249446 -0.1379 0.890295
## countryuk_c                   0.0258353  0.0243481  1.0611 0.288704
## genderMale_c                  0.0192416  0.0162850  1.1816 0.237438
## caseDefer:age_c               0.0022657  0.0006988  3.2426 0.001192
## caseDefer:collegeNon-College_c -0.0244931  0.0238151 -1.0285 0.303779
## caseDefer:countryfrance_c    -0.0651041  0.0366841 -1.7747 0.076004
## caseDefer:countrypoland_c     0.0047868  0.0356160  0.1344 0.893092
## caseDefer:countryspain_c     -0.0330870  0.0344701 -0.9599 0.337164
## caseDefer:countryuk_c        -0.0159254  0.0331591 -0.4803 0.631054
## caseDefer:genderMale_c       -0.0178742  0.0225634 -0.7922 0.428295
##                                CI Lower   CI Upper   DF
## (Intercept)                   0.7718839  0.8036573 5036
```

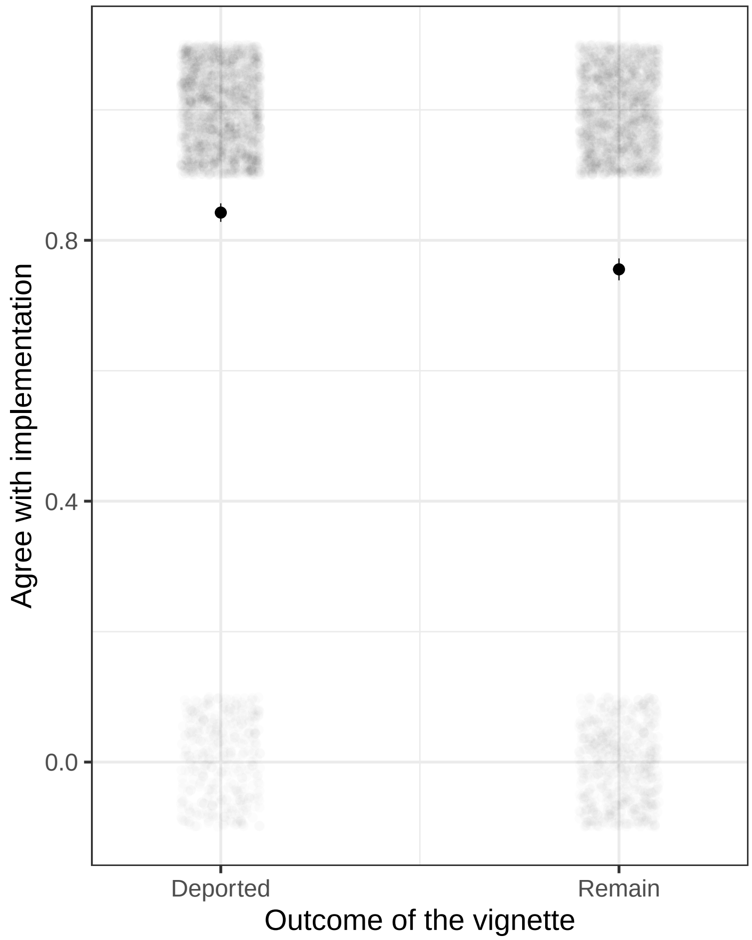
```
## caseDefer                        0.0008940  0.0448815 5036
## age_c                           -0.0018120  0.0001954 5036
## collegeNon-College_c            -0.0040560  0.0632516 5036
## countryfrance_c                 -0.1027871 -0.0001674 5036
## countrypoland_c                 -0.0984940  0.0036914 5036
## countryspain_c                  -0.0523431  0.0454615 5036
## countryuk_c                     -0.0218976  0.0735682 5036
## genderMale_c                    -0.0126841  0.0511673 5036
## caseDefer:age_c                  0.0008959  0.0036356 5036
## caseDefer:collegeNon-College_c -0.0711810  0.0221949 5036
## caseDefer:countryfrance_c       -0.1370210  0.0068127 5036
## caseDefer:countrypoland_c       -0.0650361  0.0746096 5036
## caseDefer:countryspain_c        -0.1006635  0.0344894 5036
## caseDefer:countryuk_c           -0.0809316  0.0490808 5036
## caseDefer:genderMale_c          -0.0621082  0.0263599 5036
##
## Multiple R-squared:  0.01222 ,   Adjusted R-squared:  0.009281
## F-statistic:  4.11 on 15 and 5036 DF,  p-value: 1.471e-07
```

Visualizing the results

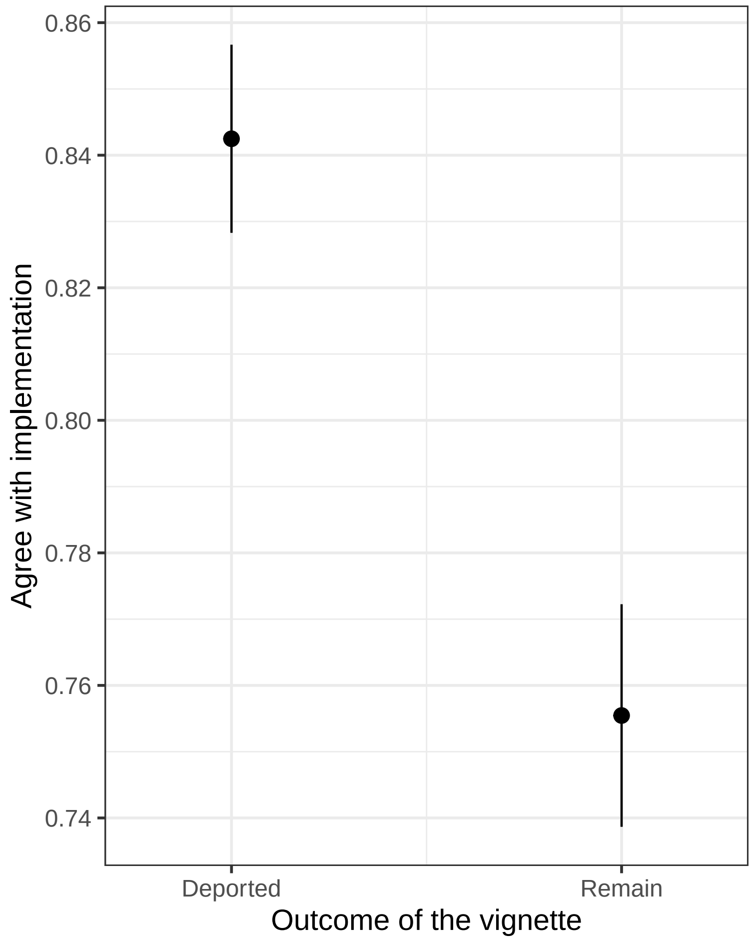
Coppock (2020) has an excellent guide, "Visualize as you Randomize" on constructing visualizations from experiments, using the underlying design to guide graphics choices. One principle is that a good visualization should give a sense of both the underlying data and the experimental design. For example, if we wanted to visualize the ATE, we could present estimates for the means of the treated and the control groups – but we might also want to use a plot of the raw observations to give a sense of the sizes of both the treated and the control groups

We'll first want to get the marginal means from the `lm_robust` fit (average under control, average under treated). You can do this manually from the components of `lm_robust` itself – the mean for control is just β_0, the mean for treated is $\beta_1 + \beta_0$ and you can calculate the SE for that sum using the variance-covariance matrix. However, this gets tedious with lots of levels and luckily there are packages that will calculate *marginal means* for you – `ggeffects` is a good one (and was imported above)

```
# Plot the marginal means and the data
ate_outcome <- lm_robust(D2Agree ~ caseOutcome, data=data_immigration)
ate_margins <- ggeffects::ggeffect(ate_outcome, terms = "caseOutcome")
ate_margins <- as_tibble(ate_margins)

# First let's plot the raw data -- add a jitter since it's binary, then add the margins
# And add transparency (alpha)
data_immigration %>% ggplot(aes(x=caseOutcome, y=D2Agree)) + geom_point(position = position_jitter(heig
                                                                         shape = 16,
                                                                         alpha=.01) +
  geom_pointrange(aes(x=x, y=predicted, ymin=conf.low, ymax=conf.high), data=ate_margins, size=.2) +
  theme_bw() +
  scale_x_continuous("Outcome of the vignette",breaks=c(0,1), limits=c(-.25,1.25), labels=c("Deported",
```




```r
# Compare to just plotting marginal means
ate_margins %>% ggplot() + geom_pointrange(aes(x=x, y=predicted, ymin=conf.low, ymax=conf.high), size=.4
  theme_bw() +
  scale_x_continuous("Outcome of the vignette",breaks=c(0,1), limits=c(-.25,1.25), labels=c("Deported",
```



We can compare plotting on top of the raw data with just plotting the marginal means – some advantages and disadvantages to both – for example, it's hard to see the confidence intervals with the x-axis scaled from 0 to 1 since we have very high precision, but we do get a clearer sense of the magnitude of the effect. Zooming in makes it easier to see the CIs but narrows the X-axis.

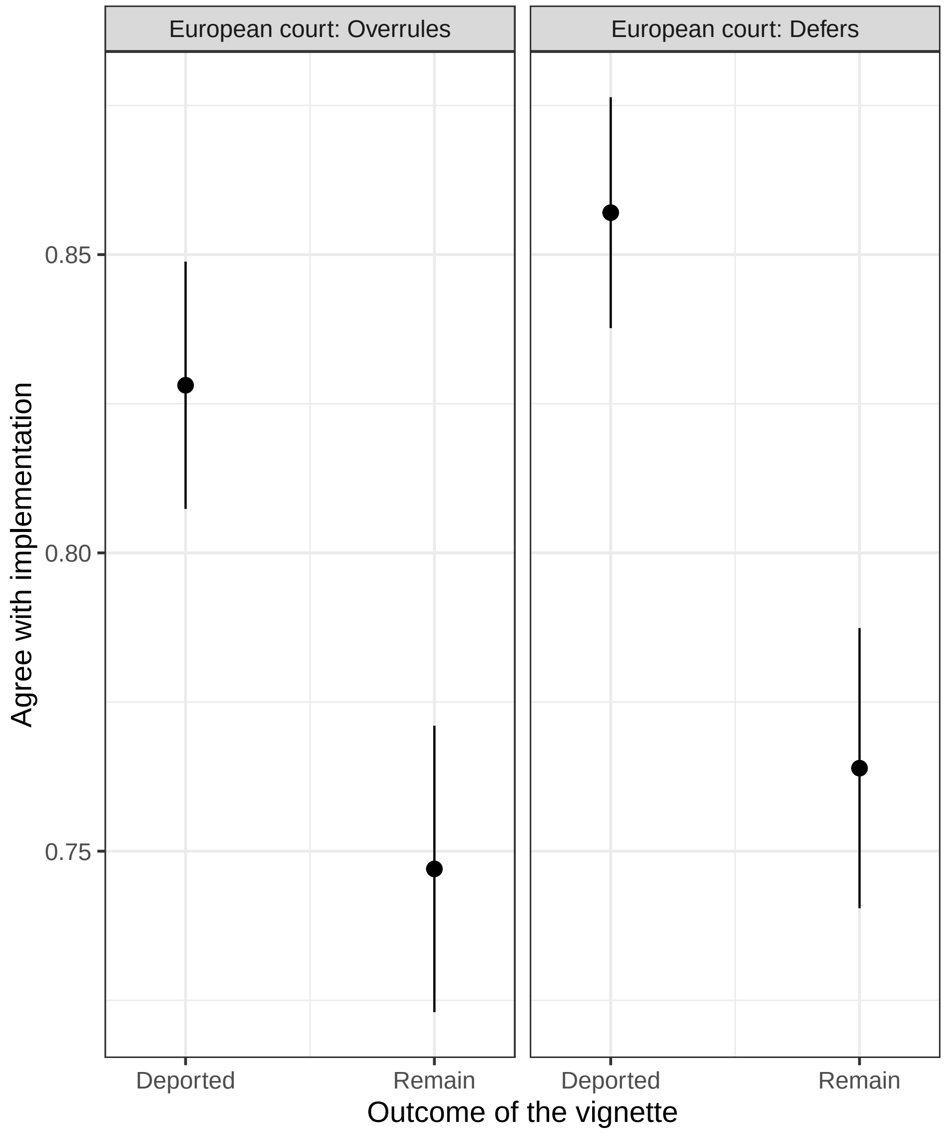
How about the joint treatment effects? Let's plot them using faceting

```
# Plot the marginal means and the data
joint_effect <- lm_robust(D2Agree ~ caseOutcome*caseDefer, data=data_immigration)
joint_margins <- ggeffects::ggeffect(joint_effect, c("caseOutcome","caseDefer"))
joint_margins
```

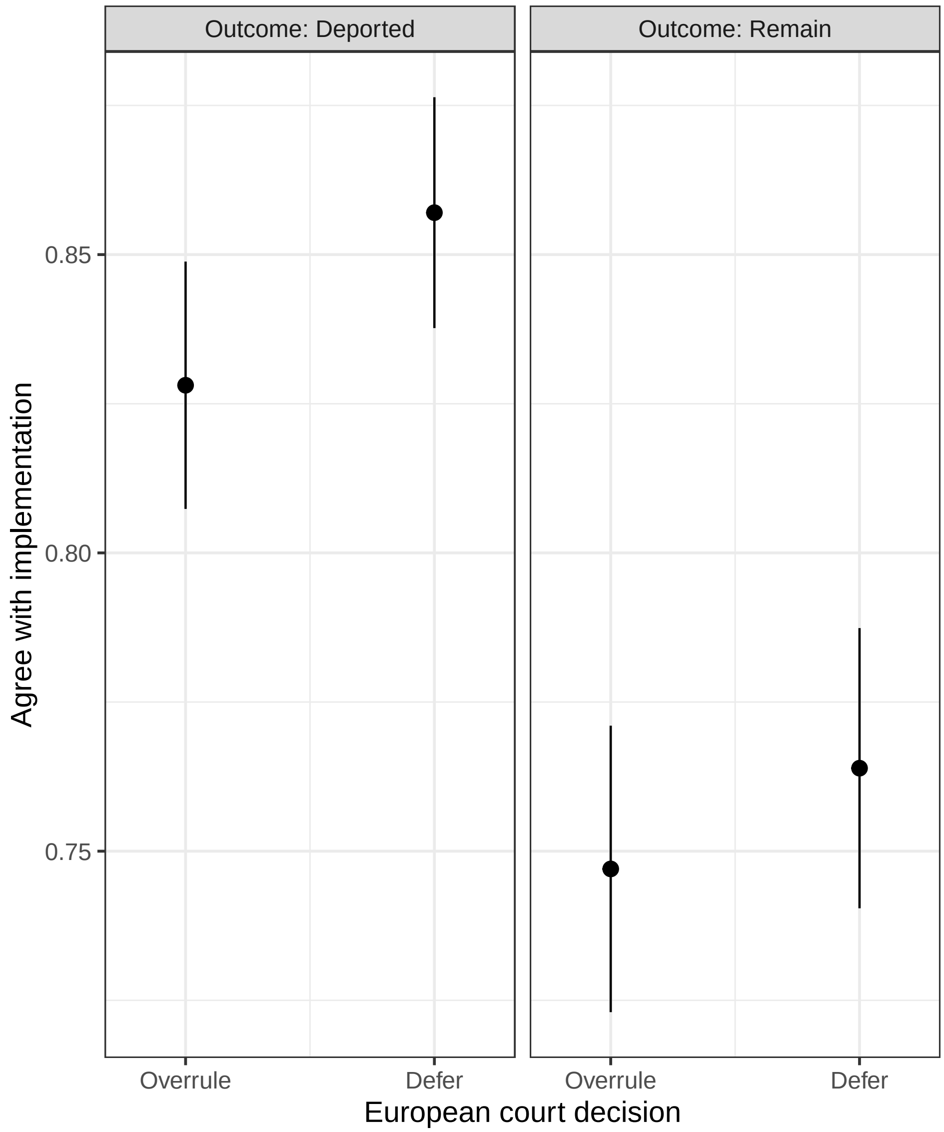
```
## # Predicted values of D2Agree
##
## # caseDefer = 0
##
## caseOutcome | Predicted |       95% CI
## --------------------------------------
##           0 |      0.83 | [0.81, 0.85]
##           1 |      0.75 | [0.72, 0.77]
##
## # caseDefer = 1
##
```

```
## caseOutcome | Predicted |       95% CI
## --------------------------------------
##           0 |      0.86 | [0.84, 0.88]
##           1 |      0.76 | [0.74, 0.79]
```

```r
# Make into a table
joint_margins_tbl <- as_tibble(joint_margins)
joint_margins_tbl <- joint_margins_tbl %>% mutate(caseOutcome = x) %>% mutate(caseDefer = as.numeric(gr

# Let's make a "labeler"
outcome_names = c(`0` = "Outcome: Deported",
                  `1` = "Outcome: Remain")

defer_names = c(`0` = "European court: Overrules",
                `1` = "European court: Defers")

# Which do we facet on?
joint_margins_tbl %>% ggplot() + geom_pointrange(aes(x=caseOutcome, y=predicted, ymin=conf.low, ymax=co
  theme_bw() + facet_wrap(~caseDefer, labeller = as_labeller(defer_names)) +
  scale_x_continuous("Outcome of the vignette",breaks=c(0,1), limits=c(-.25,1.25), labels=c("Deported",
```




```
joint_margins_tbl %>% ggplot() + geom_pointrange(aes(x=caseDefer, y=predicted, ymin=conf.low, ymax=conf
  theme_bw() + facet_wrap(~caseOutcome, labeller = as_labeller(outcome_names)) +
  scale_x_continuous("European court decision",breaks=c(0,1), limits=c(-.25,1.25), labels=c("Overrule",
```



Compare the two figures above – they display the same information, but what does each figure focus on? Which conditional ATEs does the first emphasize vs. the second?

Challenge

Visualize the ATEs on the 6-point approval scale `D2AgreeC` instead of the binary indicator – plot the marginal means on top of the raw data (use jitter). What do you find?