

Visualizing Regression Results with dotwhisker: Homework

Yeiyoung Choo

Spring 2023

For homework, we will practice how to generate dot-and-whisker plots with regression results stored in tidy dataframes. This is the second approach we learned in class. Throughout this exercise, we will use tidy dataframes as input for `dwplot()`.

Load Packages

```
library(tidyverse)
library(dotwhisker)
library(broom) #for tidying results
```

Load Data

Load the RDS file you saved in groupwork (`reg_asklaksen2010.rds`).

```
as2010reg <- readRDS("reg_asklaksen2010.rds")
```

1 Running Linear Regression

Referring to your group work, run all three regression models (Models 1-3) and store them as objects (`n1`, `n2` and `n3`). Again, we use `pr_lead` as the outcome variable. Check regression results using `summary()`.

```
n1 <- lm(pr_lead ~ pr + oilshare, data = as2010reg)
summary(n1)
```

```
##
## Call:
## lm(formula = pr_lead ~ pr + oilshare, data = as2010reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98518 -0.01966 -0.00455  0.01476  0.81061
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0265457  0.0027657   9.598 < 2e-16 ***
## pr           0.9586945  0.0042674 224.657 < 2e-16 ***
## oilshare    -0.0003052  0.0001051  -2.903  0.00371 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1006 on 4394 degrees of freedom
```

```

## (1238 observations deleted due to missingness)
## Multiple R-squared: 0.9243, Adjusted R-squared: 0.9243
## F-statistic: 2.684e+04 on 2 and 4394 DF, p-value: < 2.2e-16
n2 <- lm(pr_lead ~ pr + oilshare + lrgdppc + lpop +
educ, data = as2010reg)
summary(n2)

##
## Call:
## lm(formula = pr_lead ~ pr + oilshare + lrgdppc + lpop + educ,
## data = as2010reg)
##
## Residuals:
## Min 1Q Median 3Q Max
## -0.97566 -0.02452 -0.00315 0.01593 0.78783
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.652e-02 3.105e-02 -0.854 0.392980
## pr 9.102e-01 7.147e-03 127.359 < 2e-16 ***
## oilshare -5.976e-04 1.685e-04 -3.546 0.000397 ***
## lrgdppc 7.761e-03 3.135e-03 2.475 0.013360 *
## lpop -8.157e-05 1.242e-03 -0.066 0.947655
## educ 4.612e-03 1.128e-03 4.089 4.43e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1025 on 3144 degrees of freedom
## (2485 observations deleted due to missingness)
## Multiple R-squared: 0.9187, Adjusted R-squared: 0.9185
## F-statistic: 7102 on 5 and 3144 DF, p-value: < 2.2e-16
n3 <- lm(pr_lead ~ pr + oilshare + lrgdppc + lpop +
educ + open, data = as2010reg)
summary(n3)

##
## Call:
## lm(formula = pr_lead ~ pr + oilshare + lrgdppc + lpop + educ +
## open, data = as2010reg)
##
## Residuals:
## Min 1Q Median 3Q Max
## -0.97980 -0.02469 -0.00333 0.01592 0.78150
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.290e-03 3.348e-02 0.098 0.921727
## pr 9.076e-01 7.223e-03 125.648 < 2e-16 ***
## oilshare -5.974e-04 1.684e-04 -3.547 0.000395 ***
## lrgdppc 8.368e-03 3.143e-03 2.662 0.007804 **
## lpop -1.754e-03 1.429e-03 -1.228 0.219599
## educ 4.818e-03 1.130e-03 4.263 2.08e-05 ***
## open -1.119e-04 4.727e-05 -2.366 0.018030 *

```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1024 on 3143 degrees of freedom
## (2485 observations deleted due to missingness)
## Multiple R-squared:  0.9188, Adjusted R-squared:  0.9186
## F-statistic: 5928 on 6 and 3143 DF,  p-value: < 2.2e-16
```

2 Tidying Regression Results

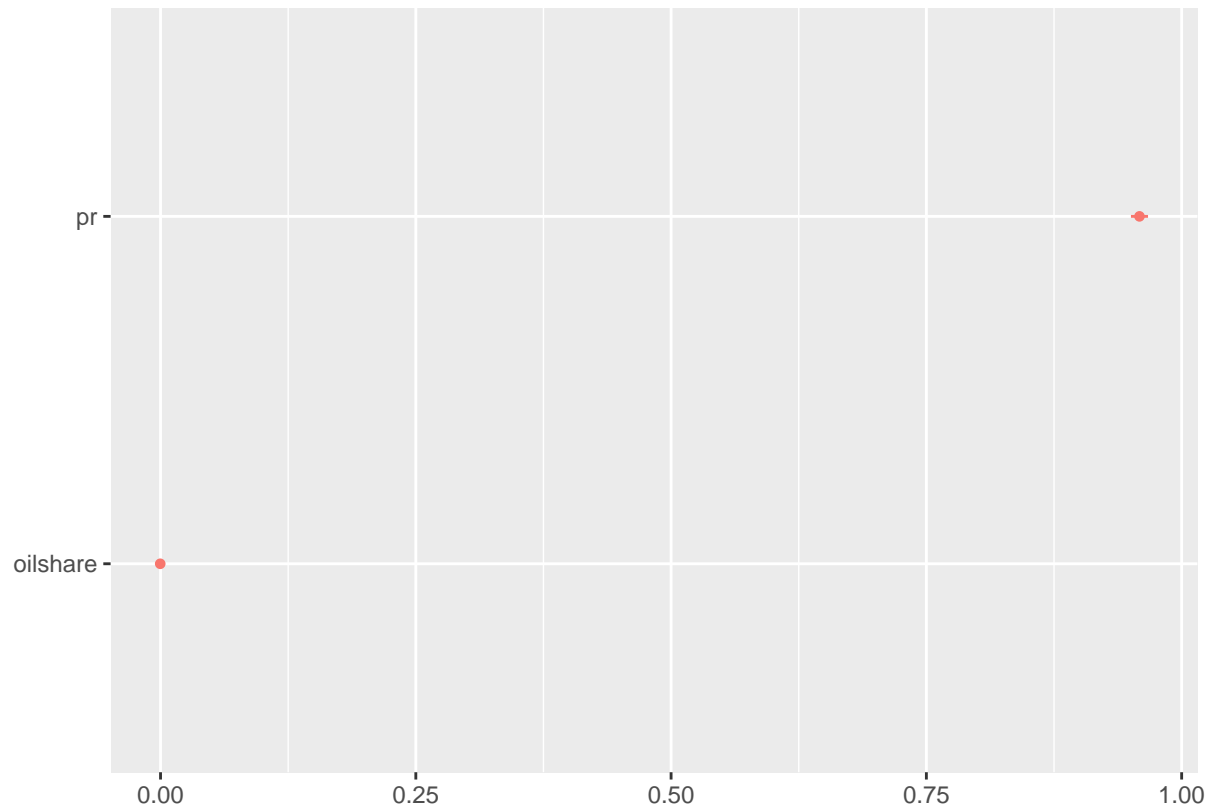
Using `tidy()` in *broom* package, store your regression results as tidy dataframes. Let's call them `n1df`, `n2df` and `n3df`.

```
n1df <- tidy(n1)
n2df <- tidy(n2)
n3df <- tidy(n3)
```

3 Generating a Dot-and-Whisker Plot

- a) Generate a dot-and-whisker plot for Model 1, using the tidy dataframe as input.

```
dwplot(n1df)
```



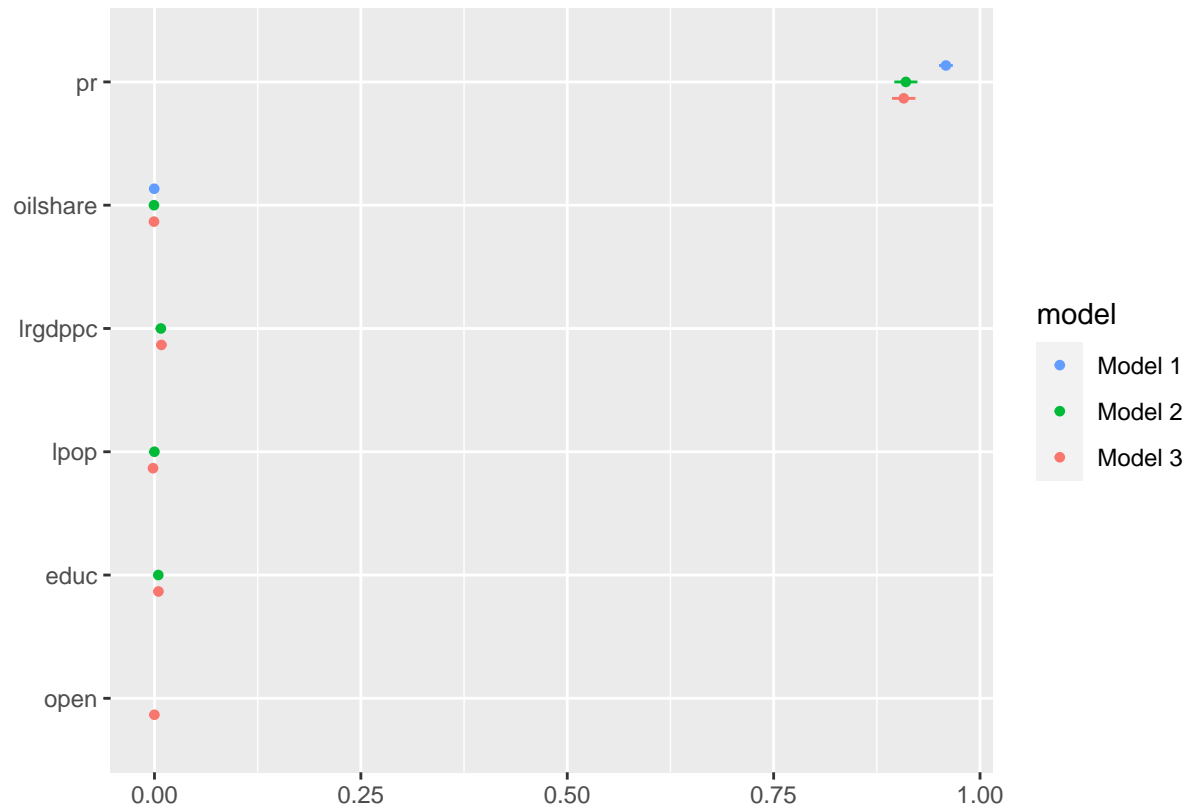
- b) Display regression results from all three models in one plot. Make sure to add a column called `model` to label your tidy dataframes as `Model 1`, `Model 2`, and `Model 3`. When merged, this column will help identify each model. Merge the dataframes into one, calling it `models`. Use it as input.

```
n1df <- tidy(n1) %>%
  mutate(model = "Model 1")
```

```
n2df <- tidy(n2) %>%
  mutate(model = "Model 2")
n3df <- tidy(n3) %>%
  mutate(model = "Model 3")

models <- rbind(n1df, n2df, n3df) #use rbind() to merge dataframes

dwplot(models)
```



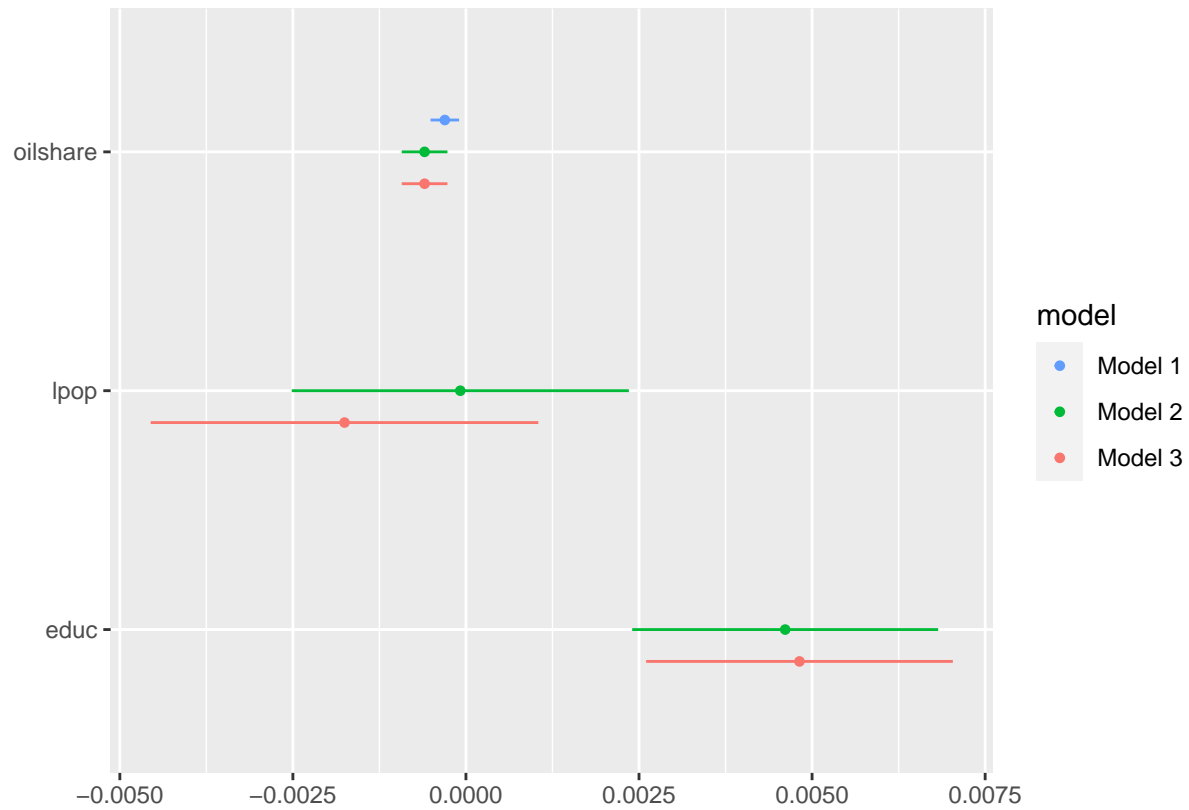
4 Customizing the Plot

- a) Generate a dot-and-whisker plot that shows coefficient estimates from all models, for the following variables: `oilshare`, `lpop`, `educ`. Starting again with `tidy()` and regression objects, prep your tidy dataframes first. Merge the dataframes, and then generate the plot.

```
n1df <- tidy(n1) %>%
  filter(term == "oilshare") %>%
  mutate(model = "Model 1")
n2df <- tidy(n2) %>%
  filter(term == "oilshare" | term == "lpop" | term ==
    "educ") %>%
  mutate(model = "Model 2")
n3df <- tidy(n3) %>%
  filter(term == "oilshare" | term == "lpop" | term ==
    "educ") %>%
  mutate(model = "Model 3")
```

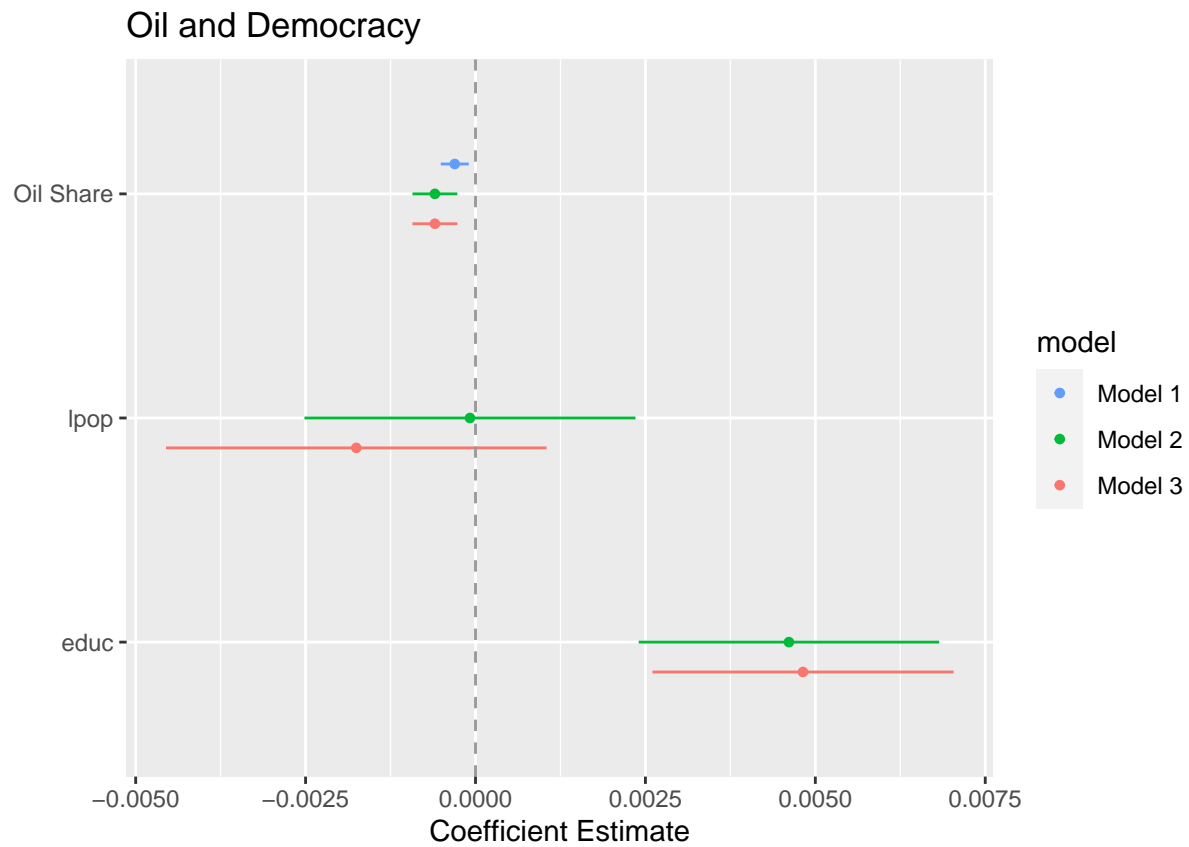
```
models <- rbind(n1df, n2df, n3df)
```

```
dwplot(models)
```



b) Referring to the PNG file you saved in group work, customize your plot to create a polished version of it. Relabel predictors as follows: Oil Share, Population (logged), Education.

```
dwplot(models, model_order = c("Model 1", "Model 2",  
  "Model 3"), vars_order = c("oilshare", "lpop",  
  "educ")) %>%  
  relabel_predictors(c(oilshare = "Oil Share", lrgdppc = "Population (logged)",  
    open = "Education")) + xlab("Coefficient Estimate") +  
  ylab("") + geom_vline(xintercept = 0, colour = "grey60",  
    linetype = 2) + ggtitle("Oil and Democracy")
```



5 Saving Your Work

Save your plot as a PNG file using `ggsave`, setting dimensions as needed.