

Problem Set #3

IRIII.2 - Quantitative Methods in the Study of International Relations

Steven V. Miller

This homework is designed to get students comfortable with making assessments about bivariate relationships. The data used will come by way of `{stevedata}` and assumes the use of `{tidyverse}` for answering the prompts included in the assignment. The student should be explicit about the R packages used and a failure to do this may result in zero-credit for the assignment.

Cross-National Rates of Trade Union Density

This homework assignment will use the `uniondensity` data included in `{stevedata}`. This data set is a relatively famous pedagogical data set for the instruction of Bayesian methods because it includes a well-known case of collinearity among two predictors in a regression model. Substantively, its use by Western and Jackman (1994) attempts to settle an empirical debate between Wallerstein (1989) and Stephens (1991) about what explains the percentage of the work force in advanced states that are unionized (aka “union density”). Wallerstein (1989) says that the primary driver is the size of the civilian labor force, arguing that collective action is harder in states with a lot of workers. Thus, the higher the labor force size (`size` in the data), the lower the percentage of the work force that is unionized (`union` in the data). Stephens (1991) instead argues we can better understand union density as a function of the concentration of the economy in heavy industry (`concen` in the data), given the unique role that industrialization had as a historical process leading to unionization.¹

```
?uniondensity
```

Here are the data.

```
# library(stevedata) # <- load this library first
uniondensity
```

```
## # A tibble: 20 x 5
##   country      union  left  size concen
##   <chr>      <dbl> <dbl> <dbl> <dbl>
## 1 Sweden      82.4 112.   8.28  1.55
## 2 Israel       80   73.2  6.90  1.71
## 3 Iceland     74.3  17.2  4.39  2.06
## 4 Finland     73.3  59.3  7.62  1.56
## 5 Belgium     71.9  43.2  8.12  1.52
```

¹You can find this discussion summarized here: Western, Bruce and Simon Jackman. 1994. “Bayesian Inference for Comparative Research.” *American Political Science Review* 88:412-423. You can find my more modern extension of it here: <http://svmiller.com/blog/2019/08/what-explains-union-density-brms-replication/>

## 6 Denmark	69.8	90.2	7.71	1.52
## 7 Ireland	68.1	0	6.79	1.75
## 8 Austria	65.6	48.7	7.81	1.53
## 9 NZ	59.4	60	6.96	1.64
## 10 Norway	58.9	83.1	7.41	1.58
## 11 Australia	51.4	33.7	8.60	1.37
## 12 Italy	50.6	0	9.67	0.86
## 13 UK	48	43.7	10.2	1.13
## 14 Germany	39.6	35.3	10.0	0.92
## 15 Netherlands	37.7	31.5	8.41	1.25
## 16 Switzerland	35.4	11.9	7.81	1.68
## 17 Canada	31.2	0	9.26	1.35
## 18 Japan	31	1.92	10.6	1.11
## 19 France	28.2	8.67	9.84	0.95
## 20 USA	24.5	0	11.4	1

Answer these questions/do these commands. A successful answer of these question must include the R code you used to help you answer the question. Each question is worth a point.

1. Western and Jackman (1994) are particularly interested in this debate because Wallerstein (1989) and Stephens (1991) are arguing for conceptually distinct phenomena (industrial concentration, labor force size) as explanations of union density, but the two phenomena correlate so highly that it is almost impossible to distinguish the two. What is the correlation between industrial concentration and labor force size? Describe it in at least two complete sentences, but show the code that generated the result.
2. Let's focus on the Wallerstein (1989) variable of labor force size. What is the correlation between labor force size and union density? Describe it in at least two complete sentences, but show the code that generated the result.
3. Create a scatterplot that communicates the bivariate relationship between the labor force size variable (on the x -axis) and the union density variable (on the y -axis) and draw a line of best fit through it. Note this line of best fit should be OLS line of best fit and not the LOESS smoother.
4. Draw a horizontal line on this scatter plot on the y -axis at the mean of the union density variable and draw a vertical line on this scatter plot that is equal to the mean of the labor force size variable.
5. This plot identifies one observation that is firmly in an "off-correlation" quadrant. Which observation is it?

Extra Credit (1 Point)

6. Answer question 5 with some kind of R code—to your discretion—that helped you identify what exactly this observation is.