Module 5 - Replicating Washington (2008)

Oliver Engist

April 3, 2020

In this assignment you will replicate the main findings from the paper "Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's". Along the way you will also replicate a figure.

- Deadline: April 29, 2020, 9 am.

Instructions: You should submit your solutions in an *RMarkdown* document, converted into a PDF. The code chunks should display the code and contain **ALL** steps of the data cleaning process and analysis. Don't show the output of intermediary steps, like regressions or data transformations. I must be able to reproduce everything using your code and the original dataset. log files, screenshots etc. will not be accepted.

- You are supposed to hand in your solutions individually.
- Start early! The exercises build on each other and if you get stuck you might not be able to continue.
- Collaborate. You will learn a lot from explaining concepts to each other. But don't just blindly copy code.

Important: The **README** file contains a description of the variables in the dataset.

Setup

- Load the usual packages (*tidyverse*, *rio*, *magrittr*). Install *stargazer* to conveniently produce regression output tables.
- Load the *basic.dta* data.

Replicate Figure 1

Figure 1 shows the mean NOW score (a score that measures how closely a politician's votes align with recommendations of the National Organization of Women) by party and number of female children.

- Make sure the relevant variables have the correct data type.
- Recode all the independent politicians to Democrat (there's only one).
- Remove observations where the number of daughters is missing.
- NOW scores are only available for the 105^{th} congress. Select this subset.

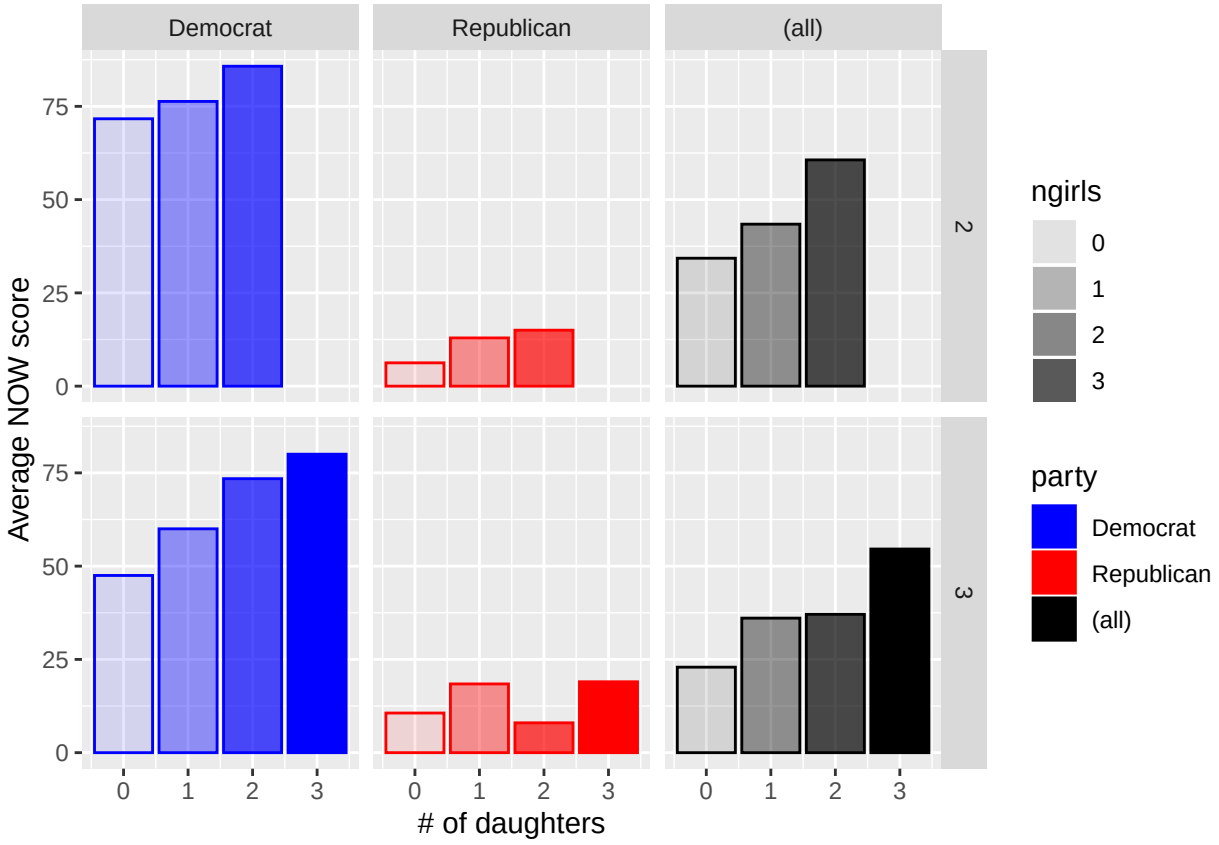
Figure 1 contains interesting information but it is not very pretty. Let's make it look a bit more colorful.

- Use *ggplot2* to replicate figure 1 (with slightly different coloring)
  1. Filter the data for the subset of observation from the 105^{th} congress with all politicians that have either 2 or 3 children
  2. Create a ggplot object with that data set.
  3. Add an *aes* layer with the x dimension being the number of daughters and the y dimension being the total NOW score. Set the *col* and *fill* option to *party* because we want Republicans and Democrats colored differently.

4. Add a geometry layer: use stat_summary, inside which you choose the function "mean" and the geometry "bar". Also set *inherit.aes* to TRUE, such that the filling and shading parameters are taken over from the *aes* layer.
5. Add a facet grid that splits the chart along *party* (column) and *totchi* (rows). Use the *margins* option to generate the bar chart that combines democrats and republicans.
6. Add a color layer using *scale_color_manual* where the breaks are the parties (incl. "(all)") and the values are the colors you want to assign each party. Use the *aesthetics* option to specify that you want to change both the colour and the filling.

- Display the figure.

Here is an example of how this could look like: Feel free to play around with the parameters to make the plot prettier.



Replicate Table 2

Data Preparation

- Create all extra variables that you need for the regression. Don't forget the two fixed effects for *region* and *number of children*. (Fixed effects in this context means simply coding these variables as factors)
- Check the data type of the variables that you will need. Convert them if necessary.
- Recode the Religion as a factor variable that it aligns with the regressors in the table. Use *relevel* to set *protestant* as the reference group.

Regressions

Table 2 shows the coefficients from the model $Score = \beta_0 + \beta_1 ngirls + \beta_2 female + \beta_3 White + ... + \beta_1 3 Dem.vote.share$ where score is either the NOW for the 105^{th} Congress or the AAUW score for each congress from the 105^{th} to the 108^{th}.

- Run the regression from the first column where the dependent variable is the total NOW score using the $lm()$ function.
  - Hint: if you selected only the relevant variables beforehand, you can just type $nowtot \sim .$ as the formula.
- Save the regression output as an object.
- Write a loop (or use $map()$)that runs the regression with the AAUW score as the dependent variable for each congress separately.
- Combine the NOW score regression and the AAUW regressions into one table using *stargazer*.
  - omit the region and number of children fixed effects from the output
  - keep the number of observations as the only statistic
  - set the type to latex
  - *Hint:* set *results = 'asis'* in the code chunk header to properly knit the latex code.

Table 4

The NOW score is composed of several sub-scores for different women's rights issues, such as reproductive rights, equal pay or violence against women. For each issue, politicians are rated either zero (does not agree with NOW on this issue) or one (does agree).

Take a look at Table 4. In each row we run the regression from above but we replace the dependent variable *nowtot* with a sub component $now1$ to $now20$. Each row then shows the coefficient of $ngirls$ from that regression.

- Use the subset from the 105^{th} congress.
- Write a loop that runs the regressions from Table 4. Each regression should be the same model but with a different now subscore on the left hand side (now1 - now20).
- Extract the coefficients of $ngirls$ from the regressions. You can use $tidy()$, which allows you to extract the estimates from each model object.
- Extract the standard errors of $ngirls$ from the regressions. You can again use $tidy()$ but this time get the $std.errors$.
- Bind the coefficients and standard errors into one table as in Table 4 in the paper. Round the coefficients and standard errors to three digits after the decimal.
- Add a column where you just number each score (unfortunately, the data does not contain labels for the now scores)
- Calculate the significance of each coefficient using a standard t-test $\frac{coefficient}{st.error}$. Add a column where you print one asterisk for significance at the 10 % level, two for significance at the 5 % level and three asterisks for 1% significance.
- Print the table using *stargazer*. Set *summary = FALSE* to print the data frame instead of the summary statistics.