PS630 Homework 5

Haohan Chen

Due October 23, 2018

Overview of Your Tasks and the Data

In this homework, you will analyze a small dataset to answer the question how political parties' clientelistic effort (i.e. "buying votes") is associated with some political economy variables. The main programming skills practiced include data cleaning with `dplyr` and `tidyr`, data visualization with `ggplot2`, and fitting linear models with `lm`.

The data of this homework come from multiple sources, including the Democratic Accountability and Linkages Project (DALP), the Polity Project and the World Bank. I understand that this data may not be relevant to your substantive interest. Hence, for this homework you are *not* required to look up for further background of the data beyond the description below. Note all data are collected around 2008.

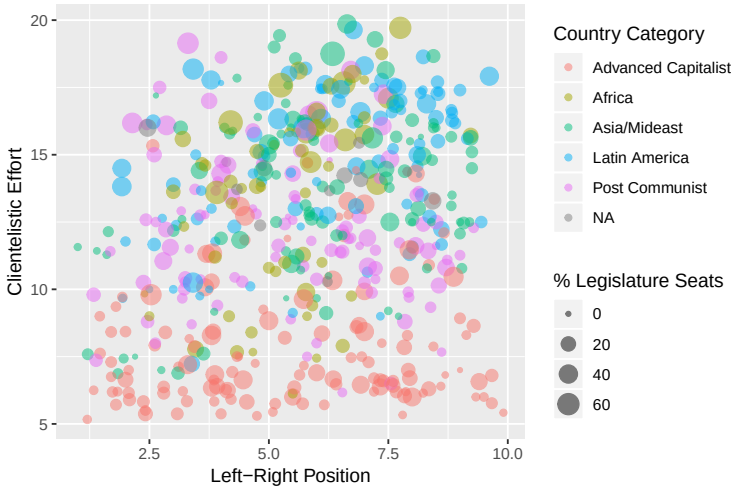
- `country_name`: Names of countries
- `country_code`: Three-letter codes of countries, aka the "iso3c" code.
- `local_presence`: The degree to which parties or their individual candidates maintain offices and paid staff at the local or municipal-level. Larger value means a party has higher level of local presence. (originally `a1` in the DALP dataset)
- `client_effort`: The overall effort of political party in spend clientelistic political exchange (e.g. "buying votes" with consumer goods or preferential treatment). A higher value means a party pays more effort in clientelistic exchange. (originally `b15` in the DALP dataset)
- `left_right`: Overall Left-Right Placement. Small value is associated with a party's left-leaning position. Large value is associated with a party's right-leaning position
- `partysize`: The proportion of seats a party occupies in the legislature
- `GDP_pcap_ppp`: GDP per capita (adjusted by purshasing power parity) of the country where a party locates
- `polity2_score`: Polity score of the country where a party locates. The variable which captures political regime authority spectrum on a 21-pont scale ranging from -10 (hereditary monarchy) to +10 (consolidated democracy)
- `country_category`: A categorical variable classifying countries into various types where a party locates

Data Cleaning (3pt)

0. Load the dataset using `load("data.Rdata")`. (0pt)
1. Print a vector of countries the parties are located, each element being a *unique* country name. (Hint: the total length of the vector should be 88). (0.5pt)
2. Create a table (or tables) of the summary statistics of these variables: `client_effort`, `local_presence`, `left_right` and `partysize`. (0.5pt)
3. Answer the following questions: (hint: `arrange()` or `top_n()`) (0.5pt)
   a. Find 5 parties with the *highest* clientelistic effort.
   b. Find 5 parties with the *lowest* level of local presence.
4. Create a dummy variable `size_10` indicating a party has over 10% of the seats in the legislature (hint: `mutate()`) (0.5pt)
5. To study the prevalence of clientelism at the country level, we subset and aggregate the data.
   a. Create a dataframe with 2 variables `country_code` and `client_effort`. (hint: `select()`) (0.5pt)
   b. With the subset of data, generate a dataframe of the *average* clientelistic effort of political parties in each country (hint: `group_by()`, `summarise()`, `mean()`) (0.5pt)

Data Visualization with `ggplot` (4pt)

6. Suppose you are interested in whether right-wing parties are more likely to be clientelistic, the first step is drawing a scatter plot between the variables `left_right` and `client_effort`. We would also like to have know whether party size and the the type of countries the parties are located have an impact.
   a. Reproduce the following plot (hint: `ggplot`, `geom_point`) (1pt)



6. (con'd)
   b. Discuss: is the graph above informative? Does it reveal any pattern? How would you improve the visualization? Note 1: Plot at least 3 alternative figures and discuss what information they reveal (either pattern or non-pattern). Note 2: Focus on the four variables used in the above figure. Don't bring in more variables (1.5pt)
   c. Discuss: What other variables can you find correlation with the clientelistic effort of political parties? Plot and discuss. (Use any variable in this dataset.) (1.5pt)

Note: For 5(b), (c), you can use whatever visualization tool in R, though `ggplot` suffices. Plus, these are two open-ended questions. Search online what `ggplot` can do.

Linear Model (3pt)

7. Continuing the inquiry in Question 5, we will fit the following models with `data`, all using parties' clientelistic effort as the dependent variable
   a. Fit a simple linear model with parties' left-right position as the independent variable. Interpret the result. (0.5pt)
   b. Fit a linear model with parties' left-right position as the independent variable, controlling parties' size and local presence. Interpret the result. (0.5pt)
   c. Fit a linear model with parties' left-right positions as the independent variable, controlling the dummy variable `size_10` and its interaction with left-right positions. Interpret the result (0.5pt)
   d. With all the given data, fit your *favorite* model, show (in a graph or a table), interpret and discuss the results. Note: You need not go beyond models we have learned in class. (1.5pt)

Bonus (2pt Max.)

Bonus points will be awarded to answers with good discussion or data visualization.