

# Lab 12: Voting in Georgia

David Sichinava, Rati Shubladze

December 27, 2017

## Instruction:

Follow the assignment step-by-step. Name your *.rmd* file in a following format: *your\_surname\_lab12.rmd*. For example,

```
sichinava_lab12.Rmd
```

## Background:

In this exercise, we will explore the patterns of voting of Georgia's population. Specifically, we will analyze the how Georgians voted for the United National Movement (UNM) in the national elections held between 2008 and 2016. We will check how various socio-demographic characteristics affect election outcomes in Georgia's electoral districts.

For this exercise, download the dataset from this [link](https://goo.gl/QoCMrD) or directly copy and paste the link to your browser: <https://goo.gl/QoCMrD>

## Data Description

First of all, create a new notebook and read the dataset to R. Do not forget to indicate the working directory. Name the data frame *geo*.

Table 1 shows the variables in election dataset. Note that each case denotes a separate electoral district and its characteristics. Each district is represented with five records corresponding to each election year. Overall, we present the data for 2008 presidential and parliamentary, 2012 parliamentary, 2013 presidential and 2016 parliamentary elections.

## Data Exploration

Let's explore the dataset. Summarize UNM vote share and Turnout proportions across elections. What is the mean vote share in each elections? In which elections were on average the highest turnout values?

Now analyze other variables. Is there any association between the proportion of people with higher education and the white collar workers? How could you explain this pattern?

Table 1: My caption

Variable	Definition
UNMProp	Proportion of UNM votes per district
TurnoutProp	Total electoral turnout
Urban	Proportion of urban population
Georgian	Proportion of Georgian speakers
Orthodox	Proportion of orthodox population
HiEd	Proportion of population with higher education
WHC	Proportion of white collar workers among working age population
District	District ID
Elections	Election wave

It is always a good idea to visualize the relations in your dataset. Here we will use ‘ggplot2’ library in order to build a diagram displaying the relationship between education and employment. First, take a close look to the example below:

```
ggplot(geo, aes(x=HiEd, y=WHC))+ # This loads the dataset and assigns to axes
  geom_point(aes(colour=Elections))+ # This visualizes data with
# points and assigns colors based on election wave
  geom_smooth(method="lm")+ # This adds a line visualizing linear
# relationship between the variables
  labs(title="Higher Education and Employment", # This assigns the main title to the plot
       x="Proportion of people with higher education", # This gives title to the x axis
       y="Proportion of people with white collar jobs") # This gives title to the y axis
```

Now, let’s visualize the relationship between electoral turnout and UNM vote share per district. Plot turnout values on x axis and vote share on y axis. Bonus points will be awarded if you manage putting subtitle to the plot - use Google and your imagination!

## Data Analysis

Finally, run three separate regression models explaining UNM vote share with turnout, the proportion of higher education holders and the proportion Georgian speakers in the district. For each model run regression diagnostics (the magic plot from meeting 11 and 10) and describe, how well does these models fit to the regression assumptions. For each model, give a substantial explanation of the observed pattern. How do each of these variables influence UNM vote share? To your mind, what would be the explanation of these results?

Another bonus question: if we only focus on turnout as an explanatory variable, do the results differ across different election waves? Give a substantial explanation of your findings.

## Where is the Dropbox link?

Zip the whole folder for the seventh lab. Name the file according to the following format: *surname.lab7.zip*. Upload the file to [this link](#) or here: <https://goo.gl/6SyMxL> before the start of our next meeting. Disregard the message which says that it’s lab 11 )