

Homework #4

Le Wang

Instruction: Do all the following empirical exercises using R. Turn in your R markdown file with answers and supporting tables and graphs, if any. Refer to the R output whenever appropriate when discussing your results. It does not matter which method you use to plot a time series variable.

Question 1: [Employment Rates]

I would like you to calculate unemployment rates (yourself!) in the U.S. for Sept 2018 based on the microdata. Let me walk you through this process. You may also find this [article](#) helpful.

1. Go to IPUMS CPS Website log into your account first.
2. Under **Data**, click **BROWSE AND SELECT DATA**
3. Click **PERSON** → **CORE** → **WORK**
4. Add the variable **EMPSTAT** to cart. Change the plus symbol to a check symbol.
5. Note that there is a brownish box on the top called **SELECT SAMPLES**, click it.
6. Click the tab **BASIC MONTHLY (33 of 513 selected)**. Unclick all the selected sample and choose only the box **Sep** in the row **2018**.
7. Note that there is a brownish box on the top called **SUBMIT SAMPLE SELECTIONS**, click it.
8. Note that there is a brownish box on the top called **VIEW CART**, click it. And then click **CREATE DATA EXTRACT** → **SUBMIT EXTRACT**
9. Once the data are ready for download, download the following files and put them in the **same folder**. And set this folder as your working directory.
 1. R file under **Command Files**
 2. DDI file under **Codebook**
 3. Data file under **DATA**
10. Unzip the data file in the format .gz.

Now, answer the following questions

1. First declare **EMPSTAT** as a factor variable and include it in the dataset. Also remove labels that do not appear in the data
2. Report the possible categories for this variable
3. Sort this variable based on its frequency and save it as **EMPSTAT2**
4. Report the frequency counts of each category of employment status.
5. What is the **percentage** of people who are at work?
6. Create a new variable called **EMPSTAT_new**. This new variable is defined as follows: (1) **NIU** and **Armed Forces** together is combined into the category **NIU**; (2) **At work** and **Has job, not at work last week** combined into **employed**; (3) **Unemployed, Unemployed, experienced worker**, and **Unemployed, new worker** are combined into **unemployed**; (4) the rest is combined into **Not in labor force**.
7. What is the **percentage** of people who are unemployed (i.e., unemployment rates)? **Hint:** Please pay attention to the definition of **unemployment rates** before you use this variable to calculate.
8. Plot the distribution for employment status. And make it as good looking as Austin would like!

Question 2: [Google Searches]

Stock investment often depends not only on fundamentals about a firm, but also on consumer interest in the firm. One way to gauge the latter is probably see how many people have searched for this firm online. Let's use Google Trends to study google stock.

1. Lets download the google searches of **Google Stock** for the past 30 days. You have to modify my code to achieve this. See manual for instructions here.
2. Plot your time series of “interest”.
3. Obtain and plot stock prices on Google using the **quantmod** package for the same time period. Do you see any obvious relationship between goolge searches and stock prices?
4. Lets focus on the searches within U.S. Which state has most interest in Google Stock? Plot the frequency by states.

Question 3: [Twitter and Sentiment Analysis]

Follow the instructions to register an app to use use Twitter API. Click here for instructions. Save your credentials and then

1. dev.twitter.com
2. Roll down to “Manage Your Apps”
3. Log into your account
4. Obtain Keys and Access Tokens

Then use my example code on Github and fill in your own credentials.

1. Find the coordinates for your hometown. Read this article to find out how in Google Maps.
2. Specify a topic or person that you care about (for example, a candidate for an upcoming election) and would like to find out how your neighbors (within 30 miles) feel about this topic/person.
3. Follow my example code and use **nrc** lexicons. Conduct a naive sentiment analysis with the Twitter data that you download.
4. Is the result consistent with what you think before this analysis?