

Econ/Poli 5 Homework 4

Setup

Today we will study what factors are associated with voting for Brexit in the UK. The dataset we will study comes from the paper “Global Competition and Brexit” by Colantone and Stanig (2018). The unit of observation in the dataset is a given region in the U.K., similar to a county in the U.S. (council_id in the dataframe contains the unique identifier for a given region). The main outcome variable we will look at is the percent of voters in the region that voted for the United Kingdom to leave the European Union (i.e. Brexit). On the Canvas page for the assignment, you will find the entire codebook for this dataset.

In the data, we have a number of variables that may be interesting to explore concerning how they relate to Brexit votes.

To get started, we first need to **load the package dplyr into your workspace using the library function**. If you have not installed dplyr yet you will need to install it first.

There are two datasets we will be using (later in the homework we will combine them). The first dataset we will load is called “brexit_df1.csv”. Load this into Rstudio and name it “df”.

Question 1

(1a) The variable “leave_share” computes the fraction of individuals in the region that voted to leave the EU. In other words, a value of 60 indicates that 60 percent of the residents voted to Brexit. What is the average value of “leave_share” in the dataframe. **Paste your code below and report the statistics requested. (3 points)**

```
mean(df$leave_share)
```

```
[1] 53.79917
```

(1b) What is the minimum and maximum value of leave_share across regions in the dataframe? **Paste your code below and report the statistics requested. (3 points)**

```
summary(df$leave_share)
```

```
Min.   Max.
```

```
25.56  72.28
```

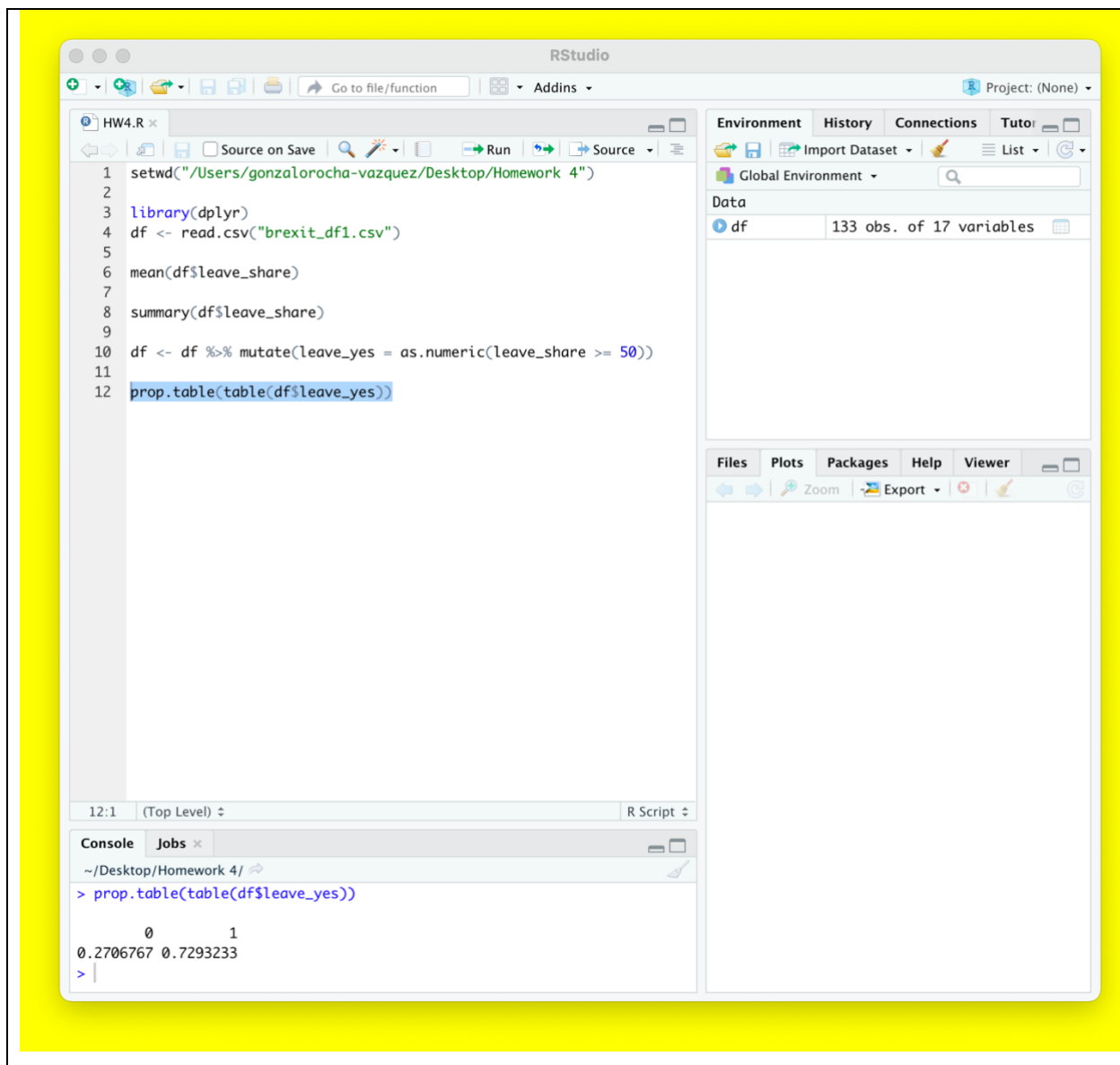
Question 2 (6 points)

(2a) Create a numeric variable named “leave_yes” that is equal to one if the fraction of individuals voting to leave the EU within the region is greater than or equal to 50 percent. **Paste your code below. (4 points)**

```
df$leave_yes <- ifelse(df$leave_share >= 50, 1, 0)
```

(2b) Create a table of proportions that displays the fraction of regions that voted to leave the EU and the fraction of regions that voted to stay in the EU?. **Paste your code below and insert a screenshot of the requested table. (4 points)**

```
prop.table(table(df$leave_yes))
```



Question 3 (6 points)

(3a) Next we will explore whether unemployment rates were related to whether a region voted to Brexit. Compute the average unemployment rates for regions that voted to **leave** the EU and save the average as an object called "m1". **Paste your code below and report the statistic requested. (4 points)**

```
m1 <- mean(df$unemployment[df$leave_yes == 1])
```

```
[1] 5.653608
```

(3b) Compute the average unemployment rate for regions that voted to **stay** the EU and save the average as an object called “m2”. **Paste your code below and report the statistic requested. (4 points)**

```
m2 <- mean(df$unemployment[df$leave_yes == 0])
```

```
[1] 5.447222
```

(3c) Create a vector that stores the average unemployment rate for regions that voted to leave as the first entry and the average unemployment for regions that voted to stay in the second entry. **Paste your code below. (4 points)**

```
vec <- c(m1,m2)
```

(3d) Create a barplot that plots average unemployment rates for regions that voted to leave and regions that voted to stay. Make sure to label the bar plot. Additionally, use the help file to change the default color of the bars. **Paste your bar plot below. (5 points)**



Question 4 (6 points)

Based on these results, does it appear that unemployment might be a potential driver of voting to leave the EU in the U.K? Briefly explain your answer (1-2 sentences). **(5 points)**

Based on the results, we can conclude unemployment might be a potential driver of voting to leave the EU in the U.K. The regions that voted to leave have a higher average unemployment rate than those that chose to stay.

Question 5 (6 points)

(5a) Compute the **median** value for “immigrant_share”. Save this median as an object in R. **Paste your code below and report the requested statistic. (4 points)**

```
med_is <- median(df$immigrant_share)
```

```
[1] 7.692308
```

(5b) Create two new dataframes that are subset of your current dataframe named df. Name the first “df_above” and make this dataframe restricted to regions with “immigrant_share” greater than or equal to the median value of “immigrant_share” in the data. Name the second “df_below” and make this dataframe restricted to regions with “immigrant_share” less than the median value of “immigrant_share”. **Paste your code below. (8 points)**

```
df_above <- df %>% filter(df$immigrant_share >= med_is)
```

```
df_below <- df %>% filter(df$immigrant_share < med_is)
```

(5c) What is the average of “leave_share” for regions with above median immigrant shares and regions below median immigrant shares. **Paste your code below and report the requested statistics. (5 points)**

```
mean(df_above$leave_share)
```

```
[1] 53.8281
```

```
mean(df_below$leave_share)
```

```
[1] 53.7689
```

Question 6 (6 points)

Load in the data frame “brexit_df2.csv” and name it “df2”. This dataframe has a few additional variables on regions. The one we will explore is “share_above_60” which reports the fraction of residents in the region that are above the age of 60.

Use the “dplyr” package to join these two dataframes together by the variable “council_id”. We want observations to be kept if they are in **both** dataframes. **Paste the code to join the dataframes together below. (7 points)**

```
df3 <- df %>% inner_join(df2, by = "council_id")
```

Question 7 (6 points)

(7a) Use the “quantile” function to calculate (and save as objects) the 75th percentile of share_above_60 and the 25th percentile of share_above_60. You may need to search for help to understand the quantile function as we did not cover it in class. **(4 points)**

```
p75 <- quantile(df3$share_above_60, .75)
```

```
p75
```

```
75%
```

```
22.12592
```

```
p25 <- quantile(df3$share_above_60, .25)
```

```
p25
```

```
25%
```

```
18.67732
```

(7b) Briefly explain the interpretation of these percentiles (1-2 sentences) **(4 points)**

These values give us the 1st quartile and 3rd quartile of a boxplot of the fraction of residents in the regions that are above the age of 60; we expect the majority of the regions to have between 18.68% and 22.13% of its population be 60+ years of age.

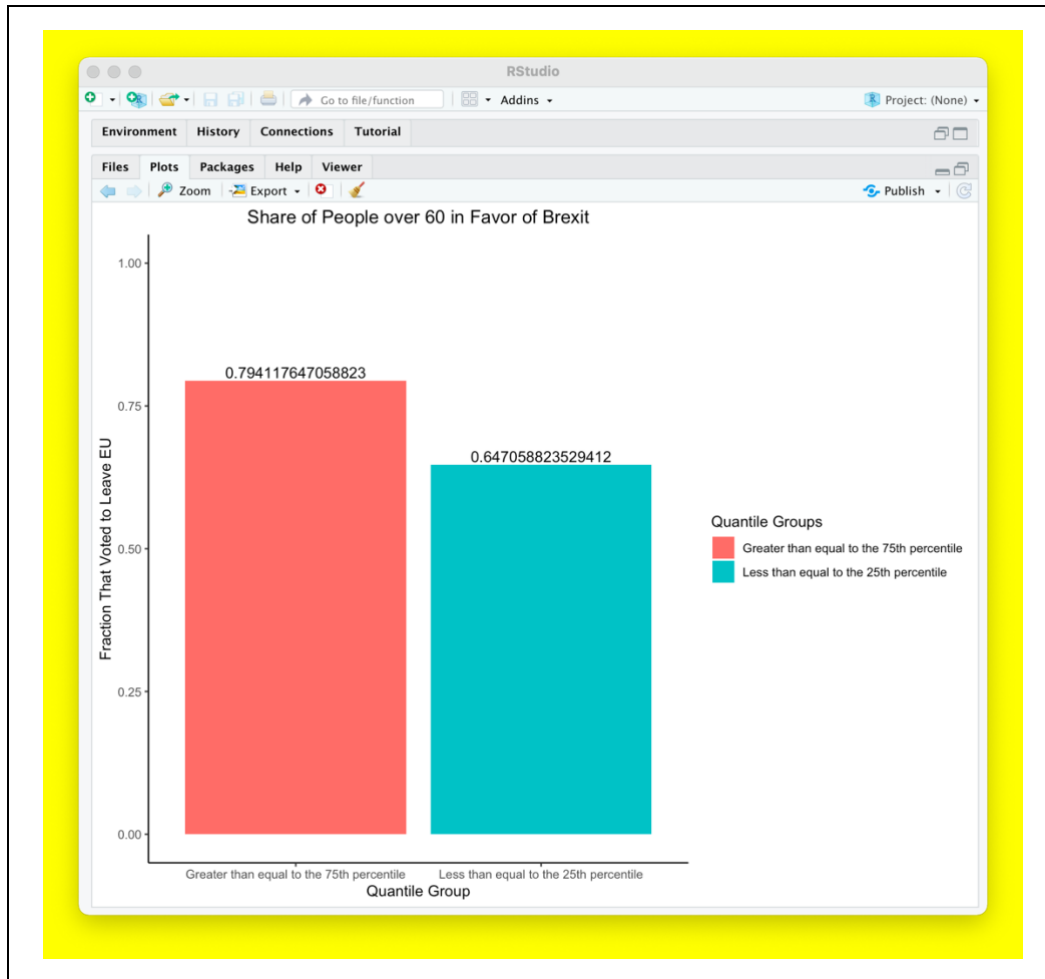
(7c) For regions with a value of “share_above_60” greater than equal to the 75th percentile, what fraction of these regions voted to leave the EU? **Paste the code below and report the statistic of interest. (3 points)**

```
m3 <- prop.table(table(df3$leave_yes[which(df3$ share_above_60 >= p75)]))  
m3 <- m3[2]  
m3  
1  
0.7941176
```

(7d) For regions with a value of “share_above_60” less than equal to the 25th percentile, what fraction of these regions voted to leave the EU? **Paste the code below and report the statistic of interest. (3 points)**

```
m4 <- prop.table(table(df3$leave_yes[which(df3$ share_above_60 <= p25)]))  
m4 <- m4[2]  
m4  
1  
0.6470588
```

(7e) Make a bar plot from the results in (7c) and (7d), labelling as appropriate as usual. Make the bar plot so that the y-axis starts at 0 and ends at 1. **(6 points)**



Question 8 (10 points)

Write a paragraph that synthesizes the results of this problem set. First, state the question we explore, then describe the analysis, and finally make some conclusions. What variables to you seem important in understanding whether a region votes to leave the EU or not? Justify your answers using results from previous questions.

The questions we explore is “What factors influence a regions’ vote on leaving the EU?”. To do this we created a binary variable to see which regions had a majority vote to leave, where 1 = Yes & 0 = No. Afterwards, we use this to see how unemployment rates affect the regions’ vote by comparing the average unemployment rates of the regions that voted “Yes” to those who voted “No”. Next, we joined the brexit_df1 & brexit_df2 to get the share of people over 60 who voted yes. We divide this into quantiles and look at the fraction of people over 60 in the “greater than equal to the 75th percentile” & “less than equal to the 25th percentile” categories who voted yes; we do this to see how the older the population is affects their vote. From these two bar plots we can conclude regions with higher unemployment and an older population are more likely to vote Yes to leaving the EU. I believe that the motivation for this is concern for economic stability, so I suggest looking at the influence of “median_wage_growth”, “share_high_skill”, “fiscal_cuts”, & “eu_eco_dependence” since these are all related to the economy.

Question 9 – Final Project Update (10 points Graded Based on Completeness)

(9a) In one or two sentences, what is the question you will address in your final project? (I will not hold you to keeping this question if you change your mind later, but choosing a final project question soon is recommended).

I plan to use the ANES (1948-2016) dataset to measure how approval for protests varies across the Hispanic community by age.

(9b) What will be the main dependent variable in your project (Y variable, or outcome variable)? What will be your main independent variable in your project (X variable)?

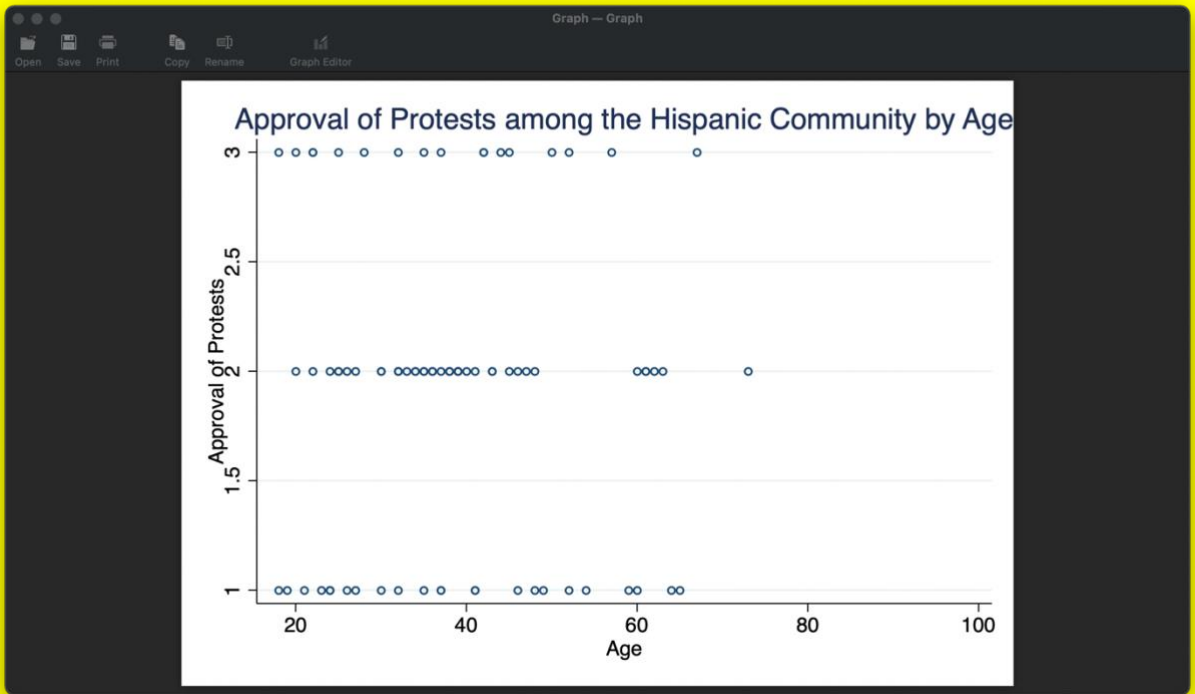
For example, in this homework, our main dependent variable was “share of people who voted to Brexit” and our main independent variable was “percent of region over the age of 60”.

You may explore multiple dependent and independent variables in your project, but just **list what you think of as your two primary variables here**.

Dependent Variable(s) = VCF0601 (Approve Participation in Protests)

Independent Variables = VCF0105a (Race-ethnicity summary, 7 categories) & VCF0101 (Respondent - Age)

(9c) Create one figure or table or summary statistic for your final project and **paste it here** (again, you can change this at a later date if you change your mind).



It's very clear that I may need to reconsider my DV, maybe to something that isn't coded into categories.