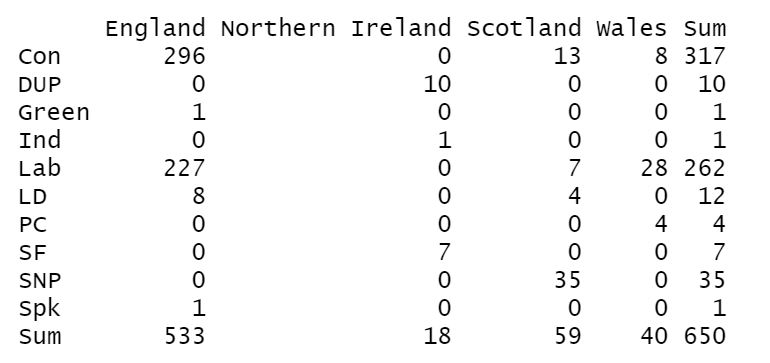
**Problem Set 2**

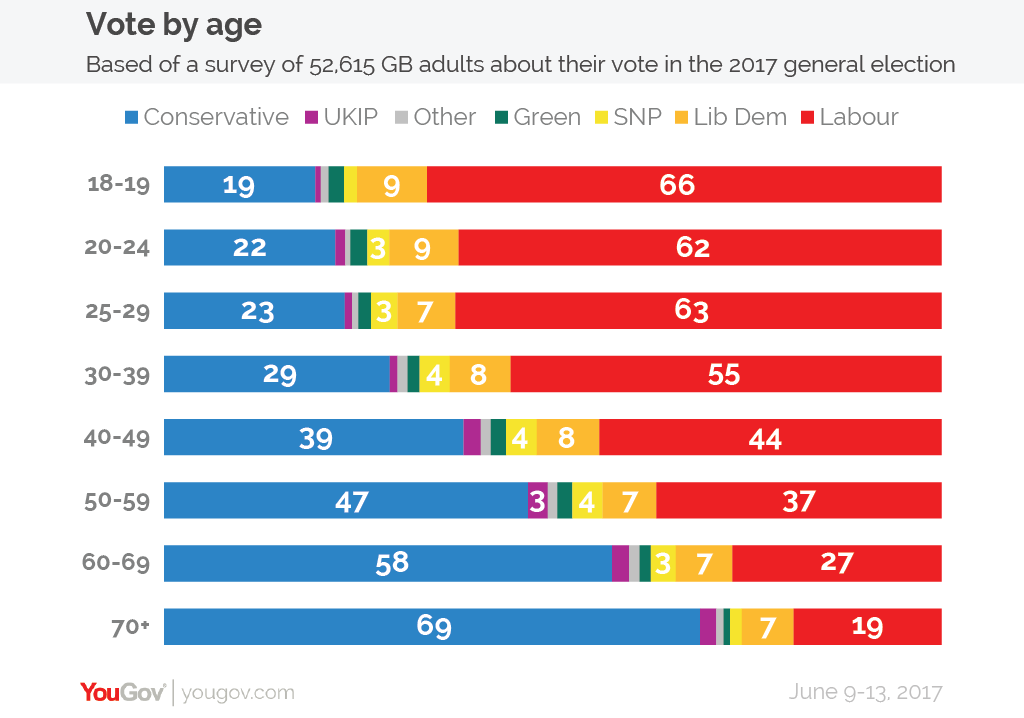
This problem set covers material from classes 4 and 5, including material from OIS section 2.2 and our class discussion on ggplot2 and dplyr. It makes use similar data to that we discussed in class. Partial credit may be given for answers that are correct in part, but not in full. This problem set is due on Gradescope by Wednesday September 8, 2021 at 11:59 PM. Please note- some of the visualizations here are in color; if these are difficult to read, please contact me individually so that I can provide a more accessible version.

**Part I: The 2017 UK Elections (16 points)**

Unlike the 2019 UK Election, the 2017 UK Election featured a close finish between the Conservatives (also commonly called the Tories) and Labour. This section makes use of data from that election. (Note: Con=Conservatives, Green=Green, Lab=Labour, LD=Liberal Democrats, Spk=Speaker).

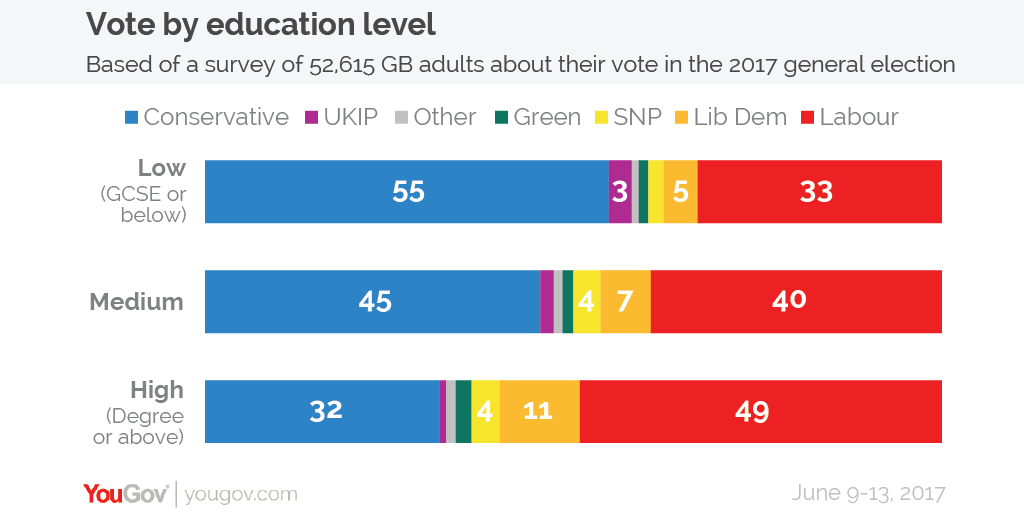


1. Below is a contingency table showing the number of seats won by each party by country. (16 points)
   1. Please calculate the row proportions for the Lib Dems (LD) for each country. You may use a calculator to divide large numbers, but please show how you get the numbers (8 pts.)
   2. Please calculate the column proportions for Wales for the Conservatives, Labour, the Lib Dems, and Plaid Cymru (PC). The above rules about calculators apply; you only have to calculate for parties that received at least one seat. (8 pts.)
2. Following the 2017 UK Election, the polling firm YouGov did a [massive survey](https://yougov.co.uk/topics/politics/articles-reports/2017/06/13/how-britain-voted-2017-general-election) to look at how various demographics affected voting patterns. Below is a bar plot they made looking at how different age groups voted. (12 pts.)



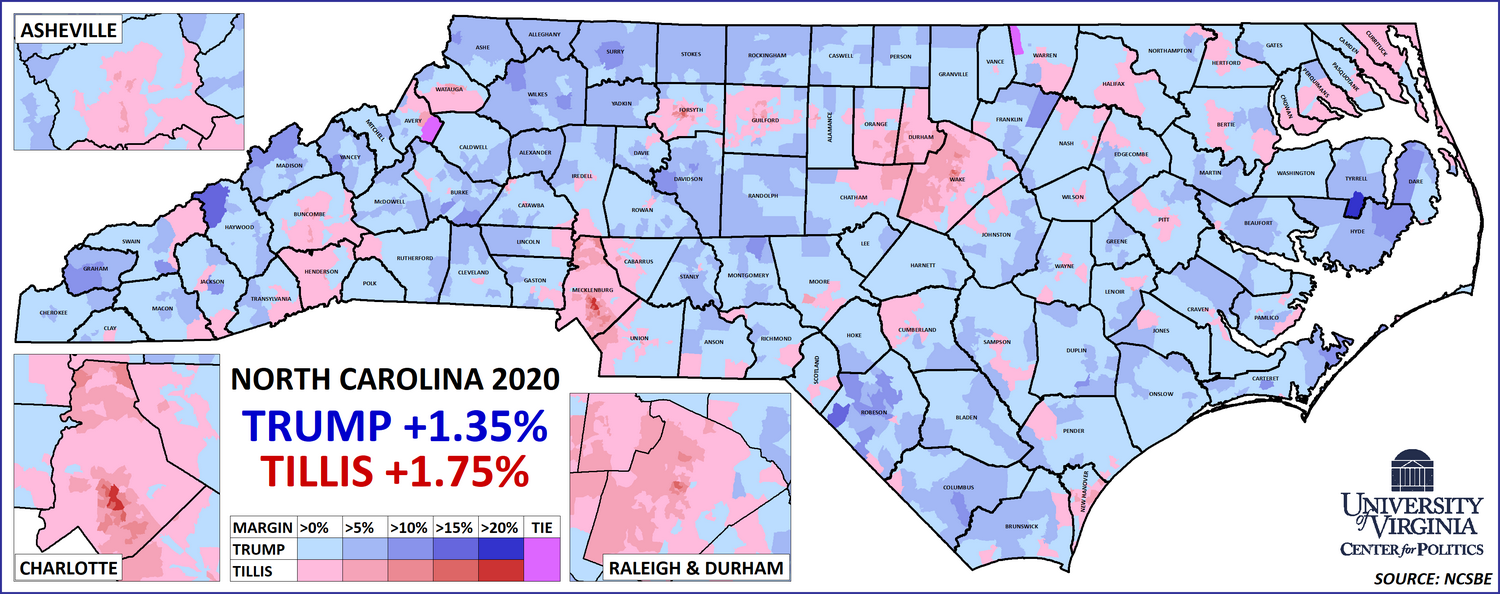
* 1. Which age group saw the most support for Labour? Which saw the Tories do the best? (6 pts.)
  2. As age group increases, does support always go up for the Tories? Does it always go down for Labour? Please provide evidence to support your conclusion. (6 pts.)

1. YouGov also looked at how education related to voting patterns. (17 pts.)



* 1. Which age group saw Labour do the best? Which saw the Tories do the best? (4 pts.)
  2. Not everyone in the UK voted for either the Tories or Labour. Which group saw the greatest amount of third party voting (i.e., parties other than the Tories or Labour) and which saw the least? (8 pts.)
  3. Do we know from this bar plot how many Britons are in each group? If not, what is a type of visualization might allow us to learn this? (5 pts.)

**Part II: The 2020 Election in North Carolina (10 pts.)**



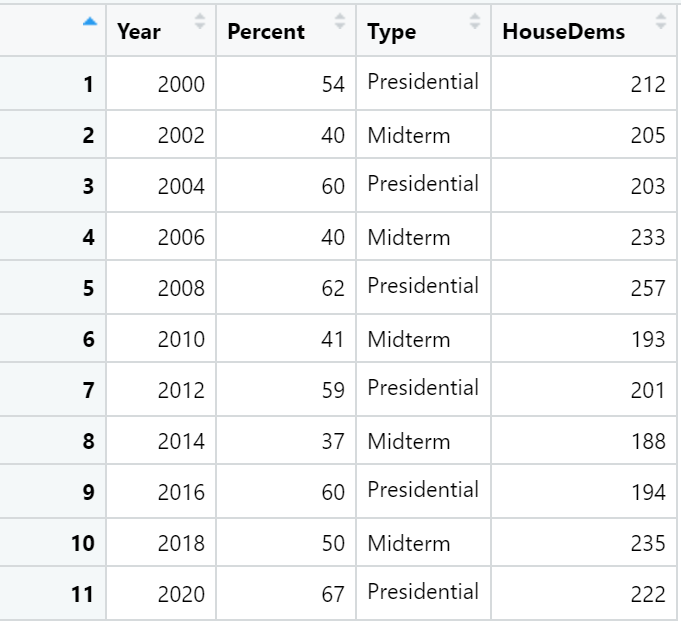
1. Last year, both President Donald Trump and Republican Senator Thom Tillis won North Carolina in their efforts to win a second term. While both won by similar margins, their coalitions for victory were different. Above is a map published in an article from April by [Sabato's Crystal Ball](https://centerforpolitics.org/crystalball/articles/did-scandal-cost-north-carolina-democrats-a-senate-seat/) comparing their performances. Precincts where Trump did better in his contest than Tillis did in his are shaded blue, while precincts where Tillis did better are shaded red. (10 pts.)
   1. What parts of North Carolina did Trump tend to outperform Tillis? Where did Tillis tend to do better than Trump? How do you know? (You can be broad here, I am not expecting you to be an expert on North Carolina geography; 5 pts.)
   2. Assuming that you are currently on the Duke campus in Durham, which Republican did better in the precinct in which you are currently located? How do you know? (5 pts.)

**Part III: Find (and describe) a visualization! (20 pts.)**

1. Please go to a website that produces public-focused statistical research such as [FiveThirtyEight](https://fivethirtyeight.com/) or [the Upshot](https://www.nytimes.com/section/upshot). Find an article that uses at least one visualization that we have learned so far in class this semester. Then, please write a short paragraph where you a) summarize the article’s purpose and findings in two or three sentences (6 pts.), b) describe the visualization that they use (6 pts.), and c) discuss whether you think it is an effective visualization and why you think that is the case (6 pts.) Please include a web link to the article with your submission (2 pts; total question is worth 20 pts.)

**Part IV: ggplot and dplyr (25 pts.)**

For the last problem set, we worked with data on voter turnout in elections since 2000. In this section, I am going to present some scenarios to you and ask how you would use ggplot and/or dplyr in order to carry out that task. For reference, this is what the data looks like in the viewer. The data set is named “Turnout.”

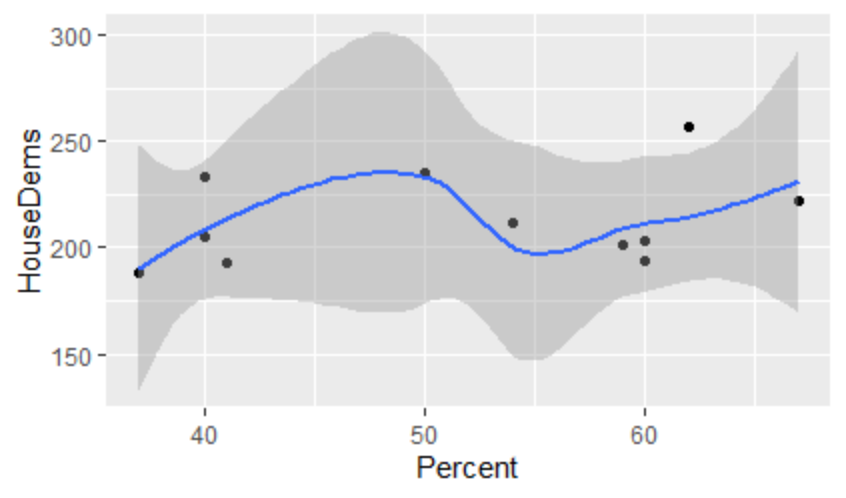


1. GGplot (10 pts.)
   1. Note the below code. What would you add to these three blanks in order to make a scatter plot with a loess regression line going through it? I’ve attached what the scatterplot would look like for reference. (6 pts.)

ggplot(data = Turnout, mapping = aes (x = \_\_\_\_\_\_, y = HouseDems)) +

\_\_\_\_\_\_ +

\_\_\_\_\_\_\_



* 1. Let’s say that you wanted to add a label to this graph with “Voter Turnout” on the x-axis and “Number of House Democrats” on the y-axis. How would you do that? (4 pts.)

1. Next, let’s consider how you might use some dplyr commands in order to work with this data. (15 pts.)
   1. Imagine that you want to determine the average voter turnout by type of election. What two dplyr commands could you use here to calculate this? How might this code look, beginning with the name of the dataset? Write this out here please. How would the code change if we had some missing values for turnout? (8 pts.)
   2. Imagine that you want to make a histogram of voter turnout with bins that are 5 percentage points wide, but only want the histogram to show presidential election years. How might you use a combination of dplyr and ggplot to do this? You don’t need to change axis labels, just show how you would go about making the initial plot. (7 pts.)