Lab Week 6 – More data wrangling

1. Using the same datasets as last week (candidates, constituency, count), answer the following questions:
    a. What does the Count.Number attribute signify in the Candidate dataset?
    b. What does the Count.Number attribute signify in the count dataset?
2. Check that you have the correct combination of attributes to uniquely identify the candidates and the constituencies.
3. Merge these tables to get a new dataframe (dfa). If this is a 1:many, which is the 1 and which is the many? How many rows should there be in the resultant dataframe? When you merge, did you lose any data?
4. Look at the count dataset. How many rows are there? Do you know why?
5. Choose columns for merging with the candidate-constituency dataset dfa. Merge the datasets to get a full dataset df. Did you lose any data?
6. Write a function that takes a parameter of the constituency name and returns a dataframe with the candidate's first and surnames, the constituency name, gender, party and party abbreviation for all elected candidates in that constituency.
7. Using df, create plots to show the solutions to the following:
    a. In a constituency (pick one), show the spread of elected candidates across the parties.
    b. Treating the top 5 parties as individual parties and the remaining parties as 'other', repeat the exercise from part a.
    c. Across all constituencies, show the number of candidates from each party that was elected.

8. Explore the following conundrum. Irish elections use proportional representation to determine who is elected. The results here show how proportional representation works. Some other countries use 'first past the post' systems. Using the datasets provided, given the number of seats in each constituency, generate the results that would occur if each set of seats was filled by the first votes each candidate got – i.e. if it is a 5-seat constituency, which candidates got the top 5 number of votes?
    a. Write a function that will:
        i. Take in a constituency number
        ii. Retrieve the number of seats for that constituency (n)
        iii. Retrieve the first count (count==1) rows from count details for that constituency.
        iv. Take the top n rows (what will you order by?)
    b. Create an empty dataframe for the first past post winners (fppdf) with the columns you need.
    c. For every constituency, use your function to retrieve the rows of elected officials and append them to the dataframe.
    d. To check if your numbers are right:
        i. Sum the number of seats over the constituencies
        ii. Count the number of elected officials.
9. Visualise the difference between the PR result and the FPP result in terms of party.
    a. Using a stacked bar chart
    b. Using a side-by-side bar chart.