FIT5147 Project: Predictors of Australian Election Results

Name: Lachlan Moody

Student Number: 27809951

Tutor Name: Bruno Luis Mendivez Vasquez & Ying Yang

Introduction ~ .5 page

Data Wrangling ~ 1 page

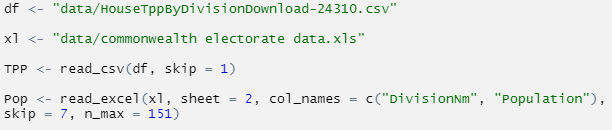


Figure 1: Initial import of tabular data

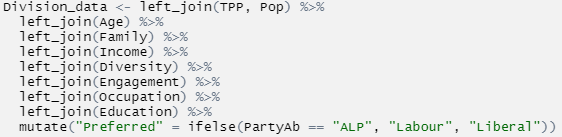


Figure 2: Tabular data wrangling



Figure 3: Initial import of spatial data



Figure 4: Combination of tabular and spatial data

Data Checking ~ 1 page

Once the appropriate data wrangling and checking was completed in R Studio, the shapefile and combined csv file (saved as Division\_data in Figure 1, were imported to Tableau public and were joined using the variables ‘Elect\_div’ and ‘DivisionNm’.

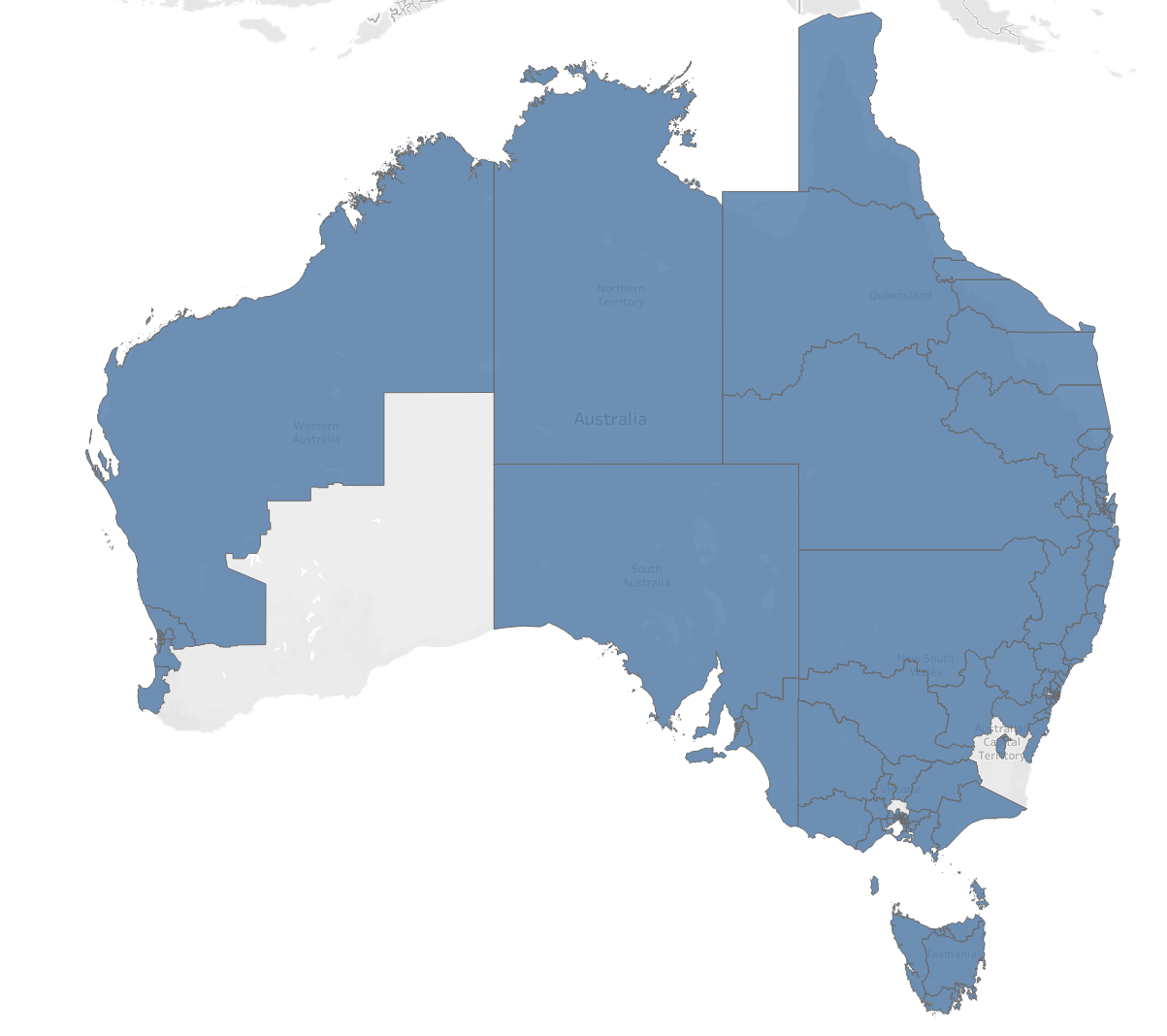


Figure 5: Initial data mapping

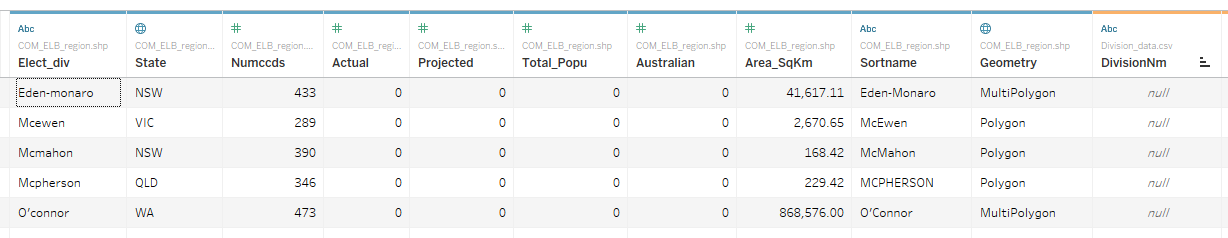


Figure 6: Identifying missing values

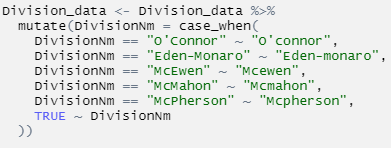


Figure 7: Changing division names

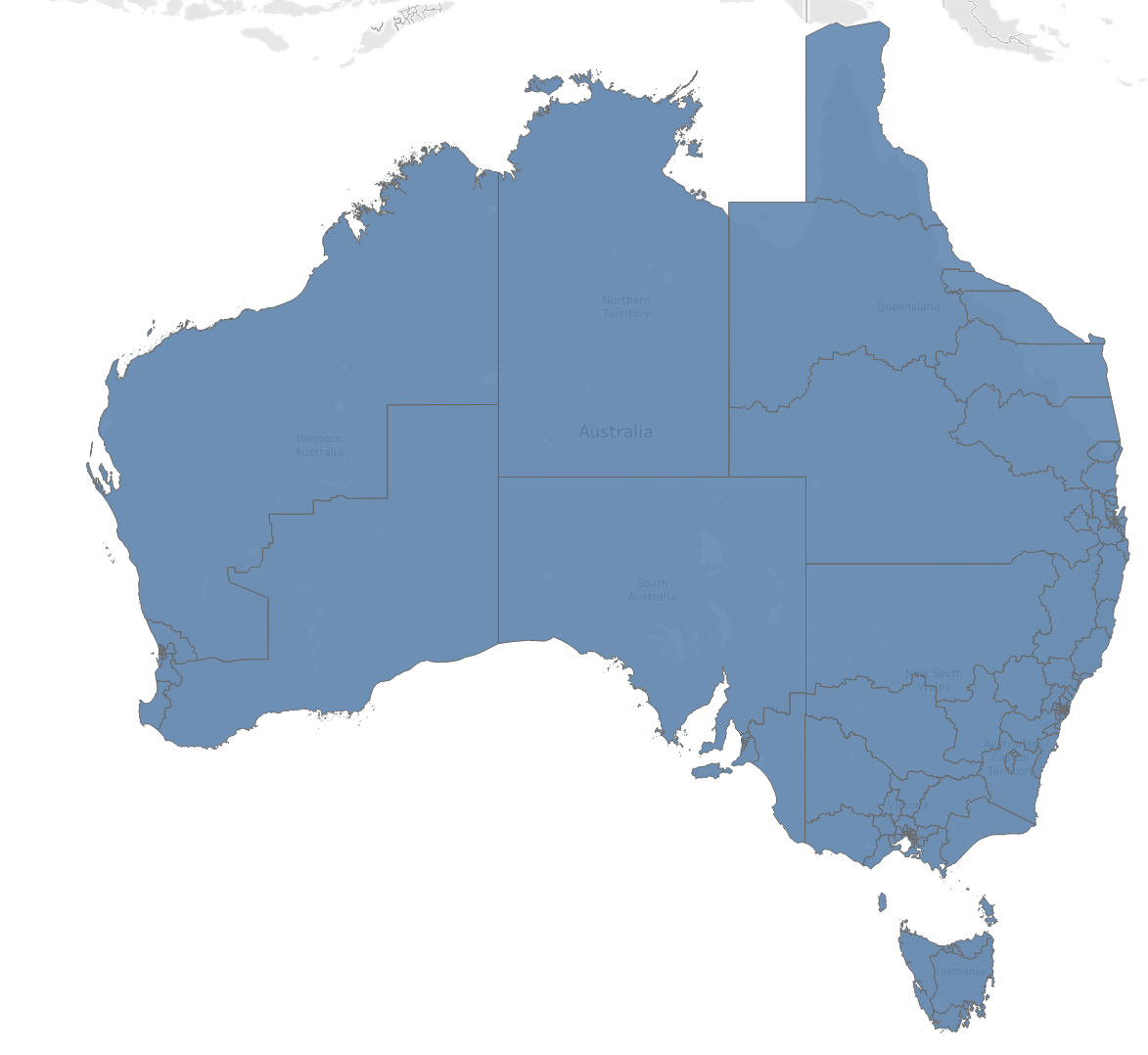


Figure 8: Data mapping with changed names

Data Exploration

Q1: How were Australia’s national election results distributed in 2019 on a two-party preferred basis?

To first examine the 2019 election results, a broad overview of the data distribution was required. The shapefile data was first plotted over a map of Australia in Tableau, with each electoral division coloured according to the party that won the preferred vote. Alongside this a pie chart was created comparing the total number of results received from each party. In both instances Labour was coloured red and Liberal blue to align with the party’s respective colours. The output is shown below in Figure 2.

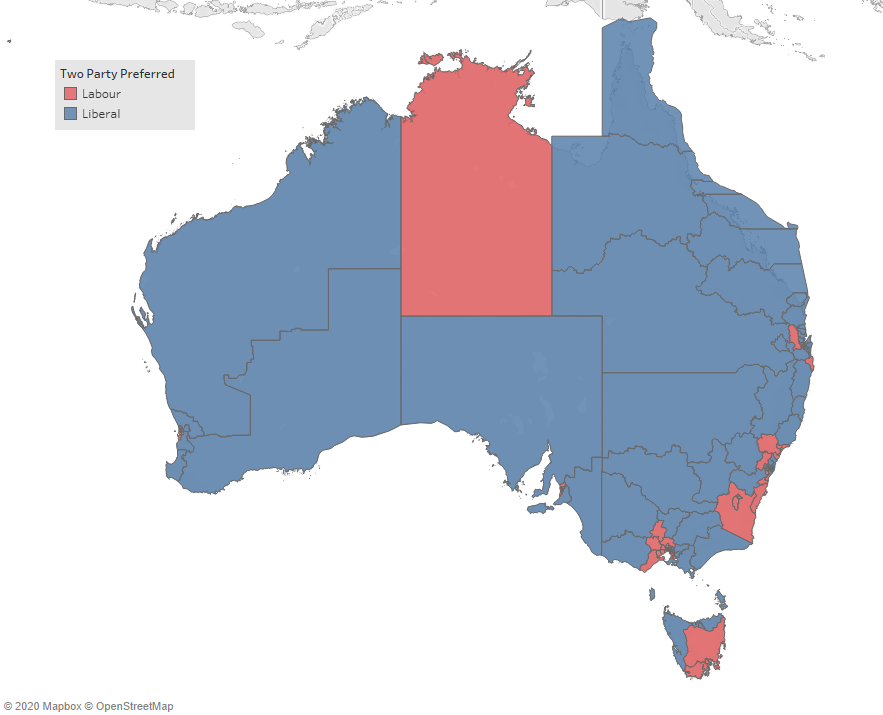
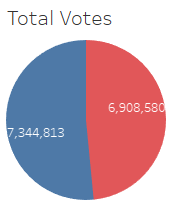


Figure 9: Distribution of TPP votes in Australia for 2019

Looking at these two outputs paints two very different pictures. Examing the chloropleth map, it would first appear that the Liberal party took almost all of the available seats in the country except for some small pockets in New South Wales around Sydney and Victoria in the Melbourne metro area. The majority of the Northern Territory and Tasmania also swung towards Labour. However, when looking at the number of votes received by each party, the gap appears much smaller at approximately 400,000 votes (or around a 3% swing). So what is the explanation for this apparent visual discrepency? To explore this area further, information relating to each seat was analysed, the outputs of which can be seen below in Figure 3, Figure 4 & Figure 5.

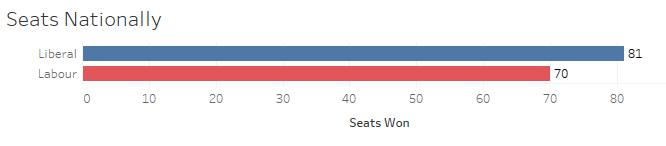


Figure 10: Number of seats won Nationally

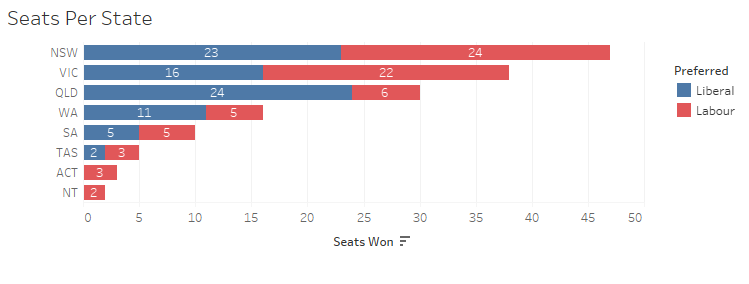


Figure 11: Number of seats won by state

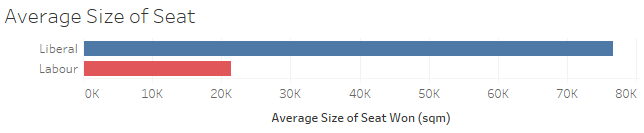


Figure 12: Size of seat won by party

The first plot was created by counting the number of seats won by each party across the entire country. While the overall outcome may be the same, again, the results paint a much tighter picture than one initially may get from examining the map only, as at first glance it would appear that there is a much greater discrepancy than only 11 seats.

Furthermore, examining the second plot which groups the seats won by state, it appears that Labour actually won in New South Wales (barely), Victoria and tied in South Australia. This led to a theory that there were many smaller seats in the metropolitan regions that Labour were winning, while the Liberal party was performing better in larger regional seats. This was confirmed in Figure 5 which plotted the average size of the seats won by each party. As can be seen above, on average, the Liberal party were winning seats almost 4 times larger than that of Labour. This explained the visual discrepancy seen earlier in Figure 2.

The final variable of interest from this data set related to swing – that is the percentage change in votes for each electorate compared to the previous election in 2016. This required more robust visual analysis than available in Tableau, so returning to R Studio, two violin plots were created using ggplot2 (Wickham, 2016) comparing the swing percentage for each party overall and within each individual state, the output is displayed below.

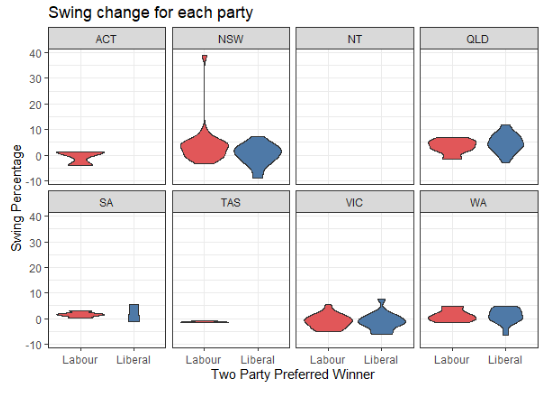
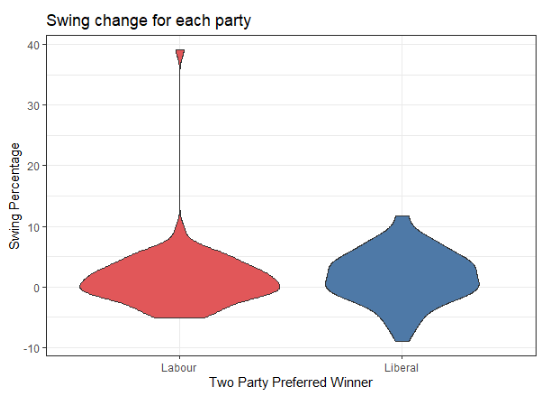


Figure 13: Distribution of swing percentage for each party

Figure 6 provides several key insights. Firstly, both distributions are centred closely above zero, with the Liberals polling slightly higher, supporting the reported swing on election night (Beaumont, 2019). Furthermore, the Labour distribution is thicker around zero which would indicate the level of support they received didn’t change much from the previous election. Noticeably, there is one value which appears to be a positive outlier in the Labour distribution. On further investigation it was found to be the seat of Whitlam in New South Wales and it experienced a 39.09% swing in favour of Labour. Initially this was thought to be a data entry error, however further searching online provided the reason for this drastic change from 2016 to 2019. This occurred because the Liberal Party did not even contest the seat in the latest election, with the opposition candidate coming from The Nationals (Green 2019). Thus, the entire difference in result, which was reported as 39.1% was credited as a swing to the Labour Party. Finally, examining on a state-by-state basis, most notably in Queensland, the Labour party recorded barely any positive swing, with most seats remaining the same or losing votes to the Liberals. This is supported by the findings in Figure 4 which has the state as the Liberal’s biggest win at 24 seats to 6. The differences in demographics within these divisions will be examined in the next section.

Q2: How were Australia’s national demographic statistics distributed in 2019 for each electorate?

The second and third stages of this analysis were conducted in R Studio. To construct the demographic choropleth maps, the maptools package (Bivand & Nicholas Lewin-Koh, 2020) in combination with ggplot2 (Wickham, 2016) were used. This package was also used to produce the density plots. A function was created to automate this process and is shown in Appendix A.3. For each of the demographic statistics discussed, the data was first wrangled to produce the variable of interest before being joined to the spatial data. This will be covered briefly, however, the entire code is available in Appendix A.4 through A.10.

Beginning with age, this variable required the greatest amount of wrangling as originally, the data was provided as proportions of the population within certain age ranges. These were ‘0-17’, ’18-34’, ’35-49’, ’50-64-, ’65-79’ and ‘80+’. To arrive at an average age, the proportions were multiplied by the population of each division to get the counts of each group and the ranges were computed to the midpoint of each group (e.g. 8.5 for ‘0-17’). As the final value was open ended, the average life expectancy was used as the upper limit for the bin (Australian Institute of Health and Welfare, 2020). These counts were then multiplied by the midpoint. The sum of this was then divided by the overall population to arrive at a mean. Note that the range ‘0-17’ was excluded as the voting age in Australia is 18. The revised data set has been plotted below in Figure 14.

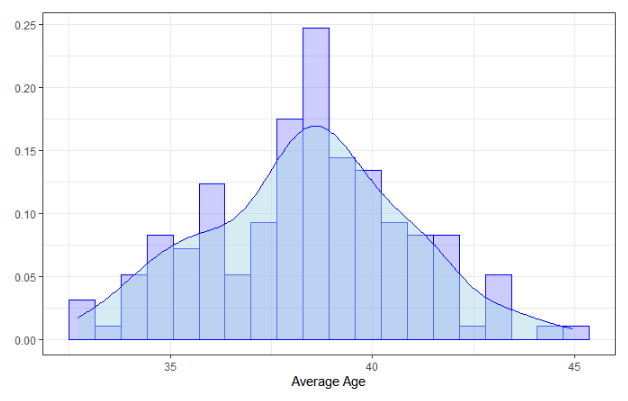
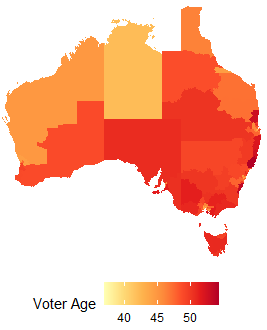


Figure 14: Distribution of voter age across divisions

From these results, it appears that average voter age in Australia is quite normally distributed across divisions, centred around 37-38 years old with quite a reasonable spread from below 35 to over 45. When comparing the map produced to the TPP map in Figure 9, the Northern Territory, an area where Labour performed quite well, is significantly younger than the rest of the country and additionally, many of the metro area have a younger overall age compared to many of the more regional seats. This suggests that there may be some relationship between voter age and preferred party.

The second variable of interest was family status. To evaluate this, the highest proportional family type for each division was calculated. Interestingly, as can be seen in Figure 15 on the following page, only ‘Couple with children’ and ‘Couple without children’ were the most popular type for any division (this data has 4 separate values), and that the former is almost twice as frequent as the latter. Furthermore, the pockets around Sydney, Melbourne, Brisbane & Perth, areas that voted Labour have predominantly ‘Couples with children’ yet are surrounded by large areas of ‘Couple without children’. Again, this may point towards another variable that predicts the election outcome.

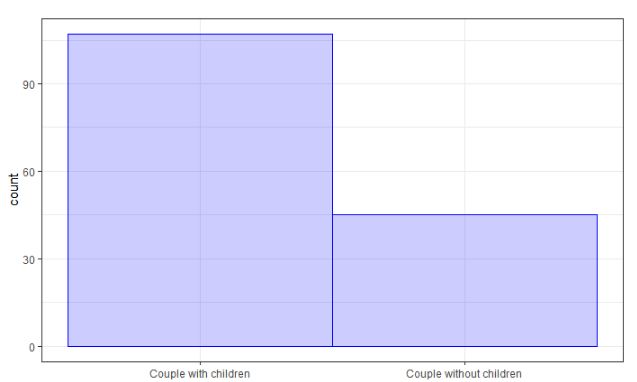
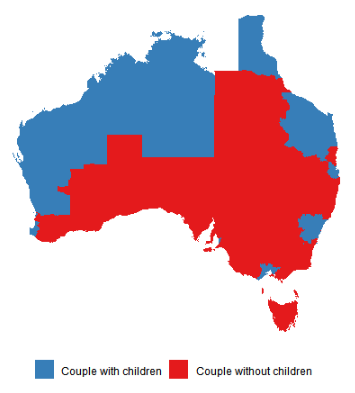


Figure 15: Distribution of family status across divisions

The third variable analysed, income, required little wrangling other than selecting the indicator of interest. For this purpose, ‘Median household income’ was selected. Looking at the density plot in Figure 16, there is quite a noticeable right skew in the data, with a long tail out up to $2,500 per week, much higher than the average of around $1,300. This indicates there is quite a lot of discrepancy in the data with many divisions having low relative incomes. Examining the map, it appears that the areas that voted Liberal generally had a lower weekly income than the metropolitan regions, except for Western Australia which had quite a higher income. It should be noted that in general regional areas have a lower cost of living relative to the metro, so it does not represent the overall financial stability of a division.

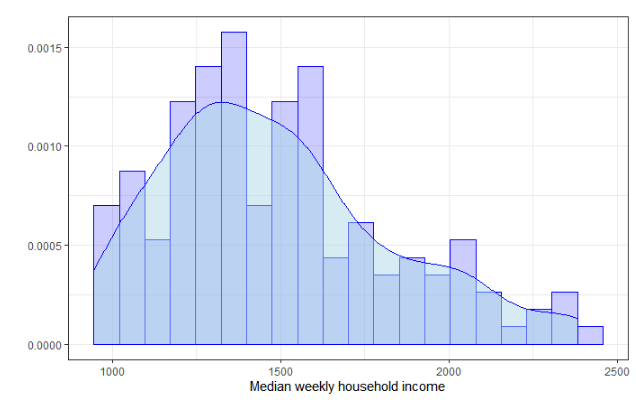
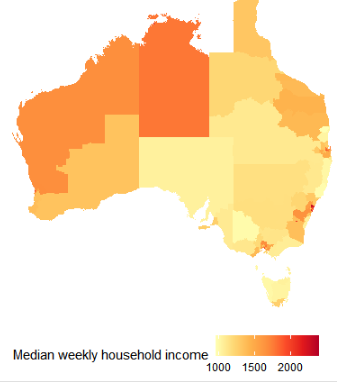


Figure 16: Distribution of median income across divisions

Figure 17 below displays the distribution of diversity within each division. This was produced by taking the proportion of people born overseas within each division as a lead indicator. From the density plot, it appears there is some multi-modality occurring within the distribution, with a peak first at around 10% and a second at a little over 30%. This suggests that most municipalities have at least some diversity, but many have a significant amount. Also, there is quite a range in this variable with some divisions having almost no diversity while some have over 50% of their population born overseas. From the map it appears that many of these regions are metropolitan, suggesting that an increase in this variable may increase the likelihood of a labour division.

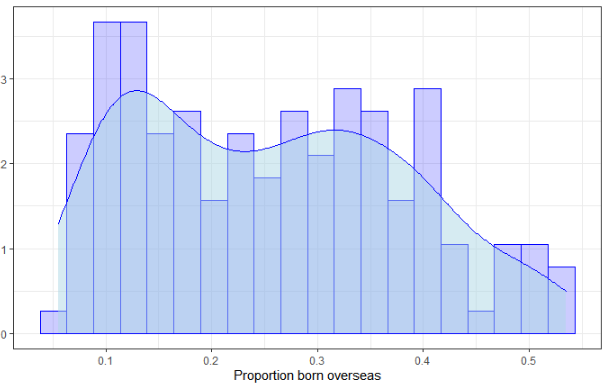
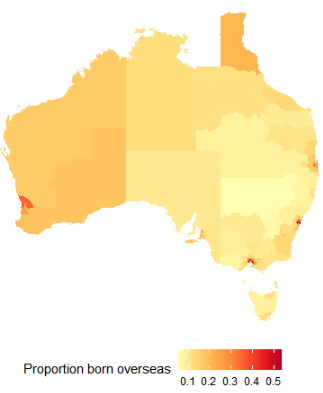


Figure 17: Distribution of diversity across divisions

The next variable of interest was engagement. Engagement refers to the proportion of the population that is engaged in employment, education and training. This value was calculated by subtracting the inverse proportion of those considered ‘Not engaged’. From Figure 18, it appears that the distribution of engagement is approximately normal with a slight left skew and is centred around 0.8. This suggests that over 80% of the population is currently engaged. Meanwhile looking at the map, engagement appears to be more evenly distributed between Labour and Liberal seats.

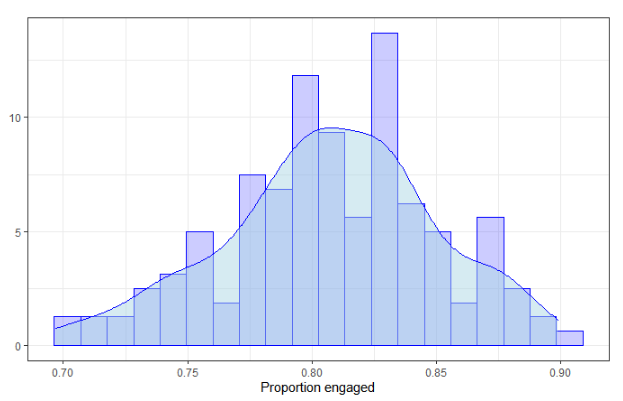
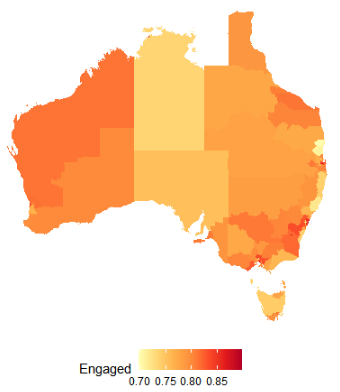


Figure 18: Distribution of engagement level across divisions

Related to engagement, education level was measured as the proportion of the population that had a Certificate III or higher qualification. The distribution in Figure 19 has an opposite direction to that of engagement with significant right skew. From the map it appears that most divisions have quite a low level of further education with much higher recording in metropolitan areas. This suggests that a lower proportion of Certificate III qualifications within a division may be correlated with a Liberal victory.

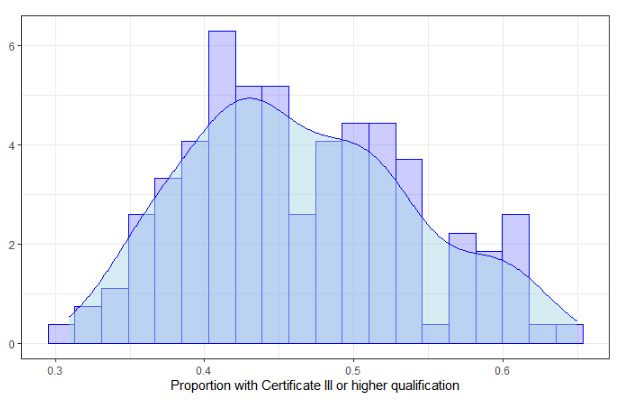
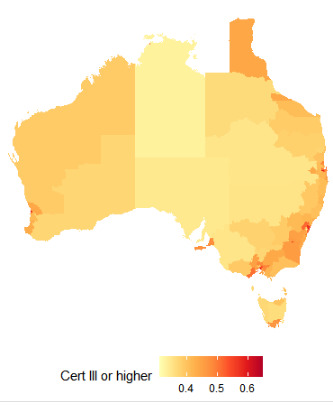


Figure 19: Distribution of education status across divisions

The final demographic variable provided related to the proportion of the population employed in different occupations. From this, the most common occupation was calculated for each division. The resulting plot outputs of this data are displayed in Figure 20. It appears that ‘Professionals’ are frequently the most common profession across all divisions in Australia, however looking at the map it appears to be mostly the case on the East coast between Brisbane and Melbourne and also the Northern Territory. Meanwhile a large swath of central Australia has ‘Managers’ as the most common type whereas in Western Australia and Northern Queensland predominantly have ‘Technicians and Trades Workers’. This is to be expected as these regions are known as mining hubs in Australia. Comparing this map once again to Figure 9 there is possibly some correlation between this variable and TPP, particularly for ‘Professionals’ in Labour areas and ‘Managers’ and ‘Technicians and Trades Workers’ in Liberal seats. This relationship, and those discussed above, will be examined using statistical analysis in the following section.

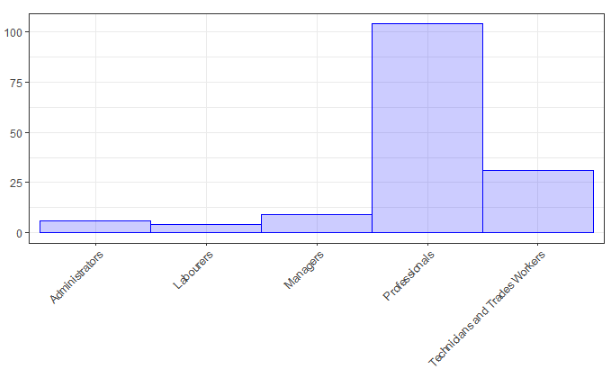
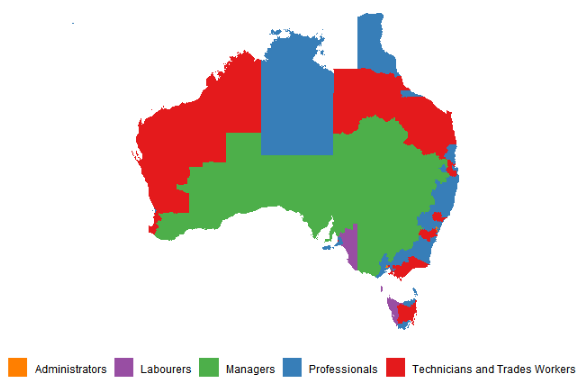


Figure : Distribution of occupation type across divisions

Q3: Is there any interaction or explanatory power between demographic statistics and election results in Australia?

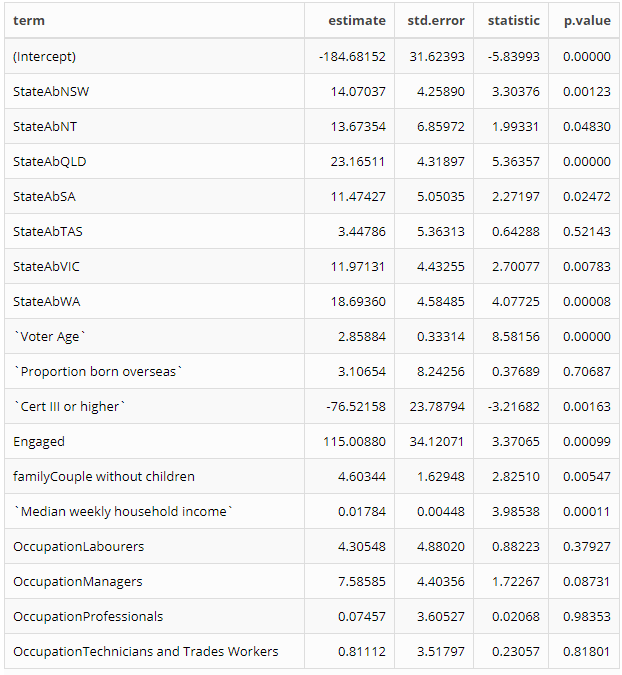


Table 1: Overall model statistics

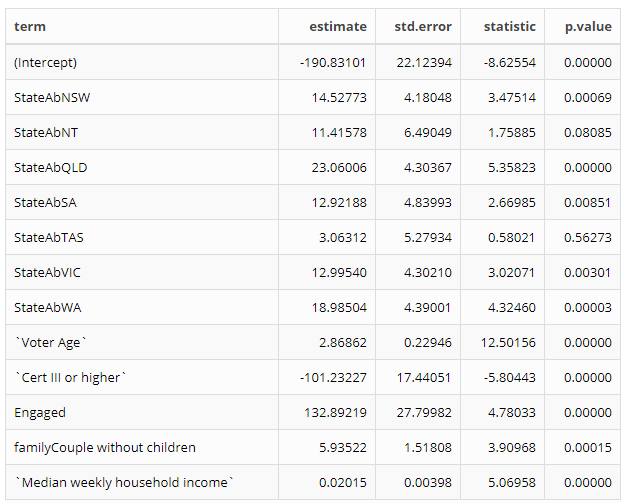


Table 2: Stepwise model results

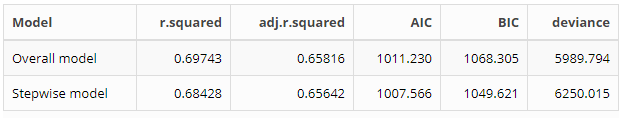


Table 3: Model fit comparisons

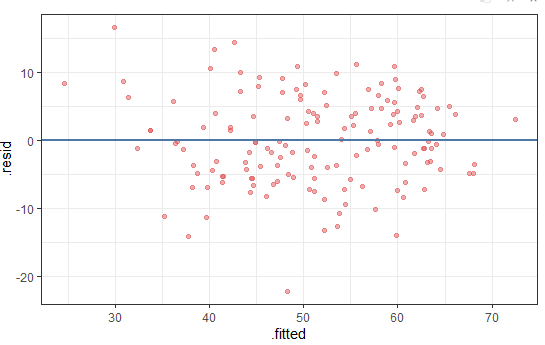


Figure 21: Residual plot for stepwise model

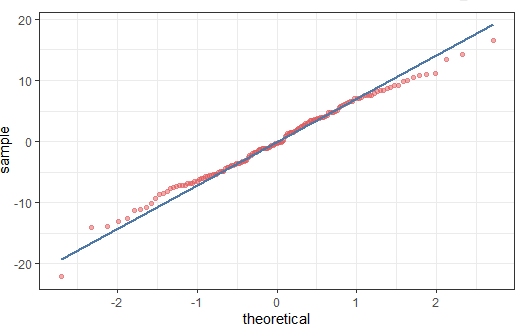
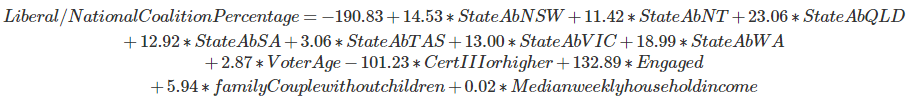


Figure 22: Q-Q plot for stepwise model



Equation 1: Formula for stepwise model

Conclusion ~ .5 page

Reflection ~ .5 page

Bibliography ~.5 page

<https://theconversation.com/final-2019-election-results-education-divide-explains-the-coalitions-upset-victory-118601>

https://www.abc.net.au/news/elections/federal/2019/guide/whit

<https://www.aihw.gov.au/reports/life-expectancy-death/deaths-in-australia/contents/life-expectancy>

<https://results.aec.gov.au/24310/Website/HouseDefault-24310.htm>

<https://www.aec.gov.au/Electorates/gis/gis_datadownload.htm>

<https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2082.02019?OpenDocume>

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Roger Bivand and Nicholas Lewin-Koh (2020). maptools: Tools for Handling Spatial Objects. R package version 1.0-2. <https://CRAN.R-project.org/package=maptools>

https://www.mathsisfun.com/data/frequency-grouped-mean-median-mode.html

Appendix

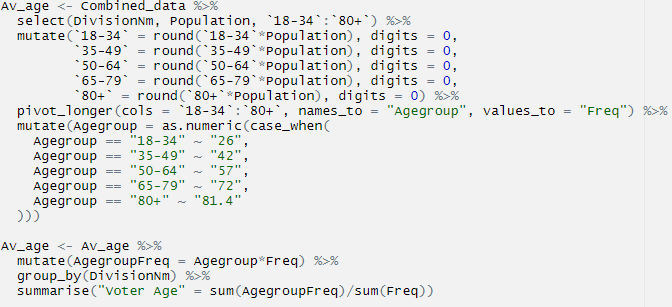


|  |  |
| --- | --- |
| Variable Name | Description |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

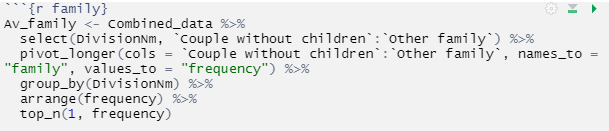
Appendix A.3: Demographic plot functions



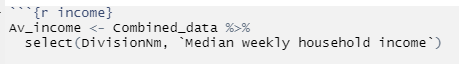
Appendix A.4: Age data wrangling



Appendix A.5: Family status data wrangling



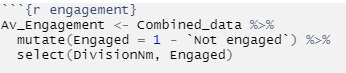
Appendix A.6: Income data wrangling



Appendix A.7: Diversity data wrangling



Appendix A.8: Engagement data wrangling



Appendix A.9: Education data wrangling



Appendix A.10: Occupation data wrangling

