ASSESSMENT COVER SHEET

Student's name	(Surname) Moody	(Given names) Lachlan Thomas	
ID number	27809951	**Phone**	0417 131 326
Unit name	Data exploration and visualisation	**Unit code**	FIT5147

Note: If this is a group assignment, please include the names of all other group members.

Project Title	Predictors of Australian Election Results	**Word Count**	15 Pages
Lecturer/tutor	Bruno Luis Mendivez Vasquez & Ying Yang	**Lab No.**	03-P1
Is this an authorised group assignment? ☐ **Yes** ☒ **No**			
Has any part of this assignment been previously submitted as part of another unit/course? ☐ **Yes** ☒ **No**			
Tutorial/laboratory day & time	Monday 8-10am		
Due date: 16/11/2020	**Date submitted:** 15/11/2020		

Please note that it is your responsibility to retain copies of your assessments

Intentional plagiarism or collusion amounts to cheating under Part 7 of the Monash University (Council) Regulations

Plagiarism: Plagiarism means taking and using another person's ideas or manner of expressing them and passing them off as one's own. For example, by failing to give appropriate acknowledgement. The material used can be from any source (staff, students or the internet, published and unpublished works).

Collusion: Collusion means unauthorised collaboration with another person on assessable written, oral or practical work and includes paying another person to complete all or part of the work.

Where there are reasonable grounds for believing that intentional plagiarism or collusion has occurred, this will be reported to the Associate Dean (Education) or delegate, who may disallow the work concerned by prohibiting assessment or refer the matter to the Faculty Discipline Panel for a hearing.

Student Statement:

- I have read the university's Student Academic Integrity Policy and Procedures.
- I understand the consequences of engaging in plagiarism and collusion as described in Part 7 of the Monash University (Council) Regulations http://adm.monash.edu/legal/legislation/statutes
- I have taken proper care to safeguard this work and made all reasonable efforts to ensure it could not be copied.
- No part of this assignment has been previously submitted as part of another unit/course.
- I acknowledge and agree that the assessor of this assignment may for the purposes of assessment, reproduce the assignment and:
  - i. provide to another member of faculty and any external marker; and/or
  - ii. submit it to a text matching software; and/or
  - iii. submit it to a text matching software which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking.
- I certify that I have not plagiarised the work of others or participated in unauthorised collaboration when preparing this assignment.

Signature: Lachlan Thomas Moody *Date: 14/11/2020*

* delete (iii) if not applicable

Privacy Statement

The information on this form is collected for the primary purpose of assessing your assignment and ensuring the academic integrity requirements of the University are met. Other purposes of collection include recording your plagiarism and collusion declaration, attending to course and administrative matters and statistical analyses. If you choose not to complete all the questions on this form it may not be possible for Monash University to assess your assignment. You have a right to access personal information that Monash University holds about you, subject to any exceptions in relevant legislation. If you wish to seek access to your personal information or inquire about the handling of your personal information, please contact the University Privacy Officer: privacyofficer@adm.monash.edu.au

Name: Lachlan Moody
Student Number: 27809951

Predictors of Australian Election Results

Contents

Introduction

This visualisation project, titled Predictors of Australian Election Results, was designed to explore the results of the 2019 Australian Federal election and whether the distribution of various demographic statistics could be modelled and used to predict outcomes at the electorate level. In order to formally address this topic, the following three research questions were proposed:

1. How were Australia's national election results distributed in 2019 on a two-party preferred basis?
2. How were Australia's national demographic statistics distributed in 2019 for each electorate?
3. Is there any interaction or explanatory power between demographic statistics and election results in Australia?

These questions, analysed in depth during the previous *Data Exploration Project*, were used to form the overarching structure and content of the interactive narrative visualisation which is discussed within this report.

With the above research questions and findings in mind, several key messages have been intentionally conveyed by the final visualisation. Relating to the first question, users are able to identify several important characteristics of Australia's election results (Australian Electoral Commission, 2019) and how they were distributed across the country. This includes that the Liberal Party won the election mainly due to strong results in Queensland and Western Australia and also, the average size of the electorates won by the two parties was drastically different with the Liberal Party securing many seats in large regional areas while Labor performed better in metropolitan centres.

Secondly, question two related to the distribution of various demographic variables amongst these 151 electorates. Due to the interactive nature of the designed visualisation,

all of the demographic variables contained within the Australian Bureau of Statistics (2019) relating to age, income, family status, housing status, employment status, education status and more can all be viewed by users on a map of the country. In addition, the distribution and averages of these variables nationally and within Liberal and Labor electorates visually can be seen. While more exploratory in nature, users can select the variables that are the most interesting or important to them to gain a better understanding of how they are represented across Australia. For example, for the Voter Age statistic, users can see that nationally the average is 47 whereas for Liberal electorates it is 48.21 and for Labor ones it is 45.64. This helps convey to the user the difference in demographic trends between seats that lean towards Liberal or Labor. Similar findings can be deduced for all other variables.

The last research question is based on a combination of the two preceding ones as it relates to a possible interaction between the two. The inclusion of a visualisation relating to this conveys to users that election results and demographic statistics are indeed related and that these different variables can quite accurately predict the outcome of the election. The inclusion of several input sliders in the design illustrates which variables are of particular importance in modelling the outcome. Additionally, as users interact with the sliders, they are able to see precisely how it affects the predicted outcome. For example, if Voter Age (discussed above) were to decrease to 46, only 1 year below the current average, the predicted outcome would flip in the Labor Party's favour. This also helps covey to audiences the message about which variables are more important to each party's performance in an election.

Finally, regarding the intended audience of the narrative visualisation, as election results are off interest and impact on everybody in the country, the design was created for the Australian general public. As such, in general basic and easy-to-understand plots and graphics were chosen including choropleths, bar charts and histograms.

Design

The visualisation design process for this project was conducted in accordance with the Five Design Sheet Methodology for Visualisation (Roberts, Headleand & Ritsos, 2015). This approach is a formalised way of sketching and prototyping designs without needing to engage with the technical aspects or requirements (Wang, 2020). As the name implies the process is divided into five separate sheets with the first focusing on brainstorming ideas, the second through the fourth containing the initial design ideas, and the fifth sheet containing the final realisation design. The complete set of sheets pertaining to this visualisation project are provided in Appendix 1 through to Appendix 5.

Firstly, various plot types were considered during the brainstorming stage that could possibly answer each of the proposed research questions. This included those used in the final visualisation, such as choropleth maps and bar charts, and those that weren't, such as pie charts and model diagnostic plots. A section of this sheet is shown in the figure on the following page and can be seen in full in Appendix 1.

Figure 1: 5DS Sheet 1 Ideas

Regarding the pie chart and model diagnostic plots, these were ultimately not included as human's are not well equipped to judge areas, and a general audience is unlikely to have an understanding of model statistics such as residuals or r squared. Additionally, tables were considered as a viable option due to their ability to provide numerical precision and condense a large amount of data. The next part of this stage involved filtering the above ideas based upon what would best suit the final design. In addition to the eliminations discussed above this included identifying that both histograms and density plots display similar information, as well as considerations relating to the placement and orientation of a bar versus column graph within the visualisation. The ideas considered appropriate for use were then categories by data type as this aligned with the proposed research questions. These were spatial, tabular and model-based respectively. Finally, for this sheet, suitable ideas were then combined in order to provide more refined and data-rich visualisations. The combination of the histogram and density plot discussed previously is one such example of this.

Secondly, the next three sheets contained individual, distinct and fully realised designs that could fully answer all of the research questions. The relative benefits and drawbacks of each sheet would then each be used to weigh in on the final design decision.

The first of these sheets was titled ‘Spatial representation of data using one page’ and centred around three choropleth maps. The design covered in Appendix 2 is provided in the figure below.

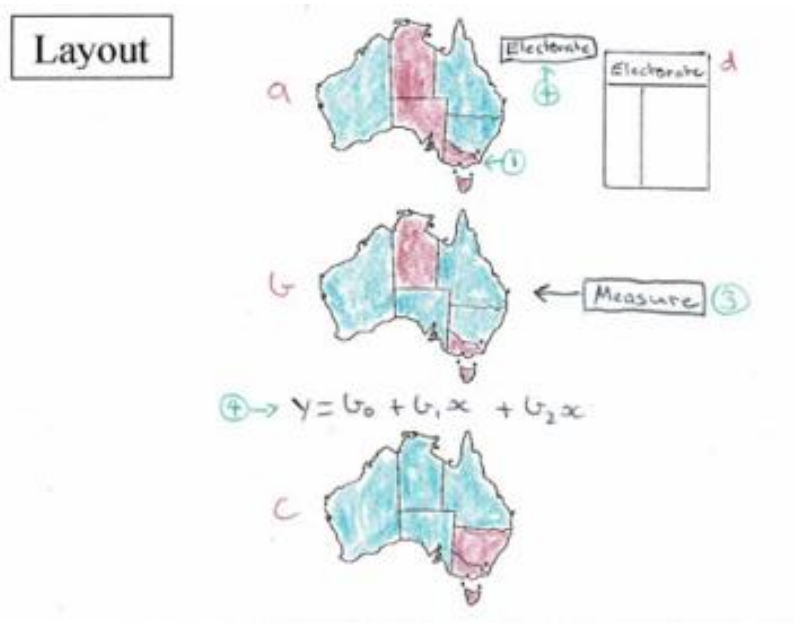


Figure 2: 5DS Sheet 2 Layout

Each choropleth map has been designed to display the election results, demographic statistics and model projected outcome respectively. In addition, a table is included on the right containing detailed information on electorates. This table can be updated both by clicking on the corresponding electorate on the first map or by selecting one from the dropdown list. Any number of electorates could be chosen for comparison. Similarly, the demographic map can also be able to be updated from a dropdown list containing the various variables. Additional interaction is provided in the third map by allowing users to adjust the real values used by the formula to see how the results would change if Australia’s average age or income were to change for example. This design would allow users to easily compare outcomes across the three questions on a single page, ensuring that the connection between election results and demographic variables is clear and maintaining a logical flow. Additionally, there are numerous opportunities for interaction increasing engagement and the exploratory power of the visualisation. However, as all the visualisations shown are quite similar issues may arise relating to user attention. Furthermore, will all visualisations shown at once the user may also become overwhelmed due to visual clutter on the screen.

The second of the initial designs was titled ‘Spatial representation of data using multiple pages’. This design was very similar to the one just discussed and contained identical interaction opportunities except that rather than having all graphics displayed on the one page, each was presented in its own tab. Each of these tabs thus aligned with one research question. The layout of this design is shown on the following page and the full sheet can be viewed in Appendix 3.

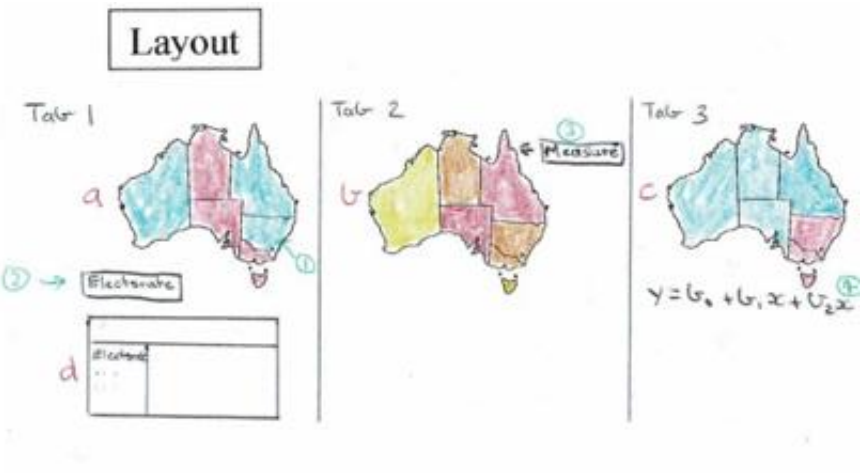


Figure 3: 5DS Sheet 3 Layout

While this did improve on the first design somewhat as each tab had a clear focus on a specific question, there was a significant drawback. According to psychologists, human visual working memory is in fact quite limited. When concentrating most people tend to only remember 3-5 objects at a time and when they are not this number falls to only 1 or 2. With the tab based format, comparisons across questions are made harder as users would have to switch back and forth to check previous findings. In addition, the link between ideas may be lost in the process. This concern is especially relevant considering the chosen audience, the general public, may not be fully engaged when viewing the visualisation thus reducing the number of items they can hold in their memory.

The last of the initial designs moved away from the map-based visualisations and instead focused on presenting the data using a wider-variety of graphics. This sheet was thus titled 'Graphical and model-based representation of data' and a sample from Appendix 4 is shown below.

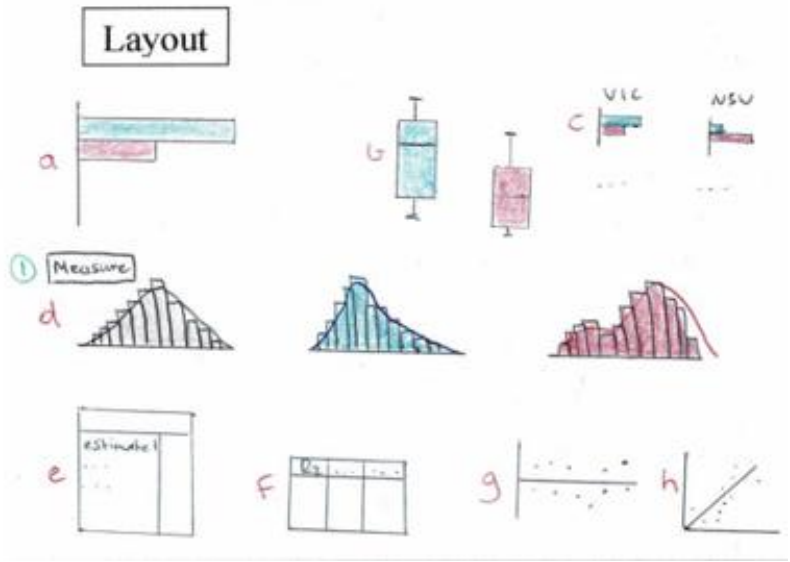


Figure 4: 5DS Sheet 4 Layout

This design was broken into three rows, similar to the first design, with each row created to answer the first, second and third research question respectively. The first contained a bar plot comparing the number of seats won by each party, a box plot comparing the size of

electorates won by each party and a bar chart comparing the number of seats won for each party by state. The second contained three combined histogram and density plots for a chosen demographic measure nationally, in Liberal seats, and in Labor seats. As before the displayed measure could be updated through a dropdown list. The third row, as it pertained to the interaction between the two, contained model-based output. This included a table containing a summary of the estimates, a table containing the model's performance, a residual plot and a quantile-quantile (q-q) plot. While this design contained a diverse and rich range of visualisations providing detailed information for each measure there were several identified limitations. This included failing to capture the geographic distribution of the collected data, limited opportunities for interactivity outside of the demographic statistics, and perhaps most importantly, failing to correctly tailor the design to the chosen audience. Two of the important factors when designing such visualisations is to ensure that the information communicated is relevant and that the audience has the appropriate knowledge to understand what is being shown to them. With the broad audience this visualisation is designed for, and as discussed earlier, it is unlikely that the majority of users will be able to accurately interpret the model output as these concepts are generally taught at the university level in statistics classes. Thus, while such plots and tables are useful, they are likely to be unnecessary in this situation.

Finally, with the entire process in mind, the last sheet relates to final realised design and can be seen in Appendix 1.5. It also contains details about the algorithms, methods, technology dependencies and estimates of cost and time necessary to implement the design. Based on the previous sheets discussed above, this final design was labelled `Spatial representation of data using iterative design' and incorporated various positive elements from each. The layout of this design is shown below.

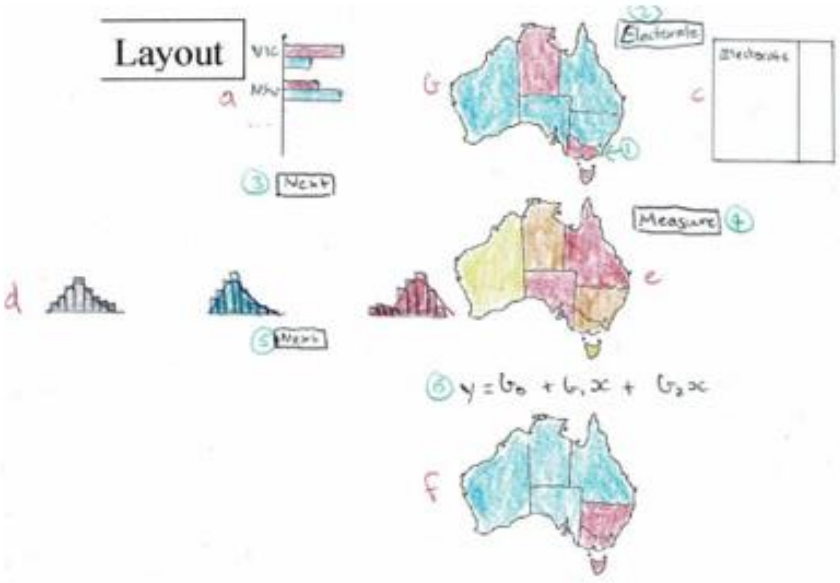


Figure 5: 5DS Sheet 5 Layout

Organised similarly to the first design, the visualisation is centred around three different choropleth maps again with one showing the geographic distribution of the election results, one for the distribution of demographic statistics and a final one for the distribution of the model projected outcome. The table containing information on individual electorates in also

incorporated. However, this design builds on the first by adding some of the plots from the third design, such as the bar chart comparing results by state as well as the histogram and density plots for the selected demographic variable. All interactivity discussed previously is also incorporated into the design. These changes were made to provide some additional variety for visual stimulus but also to more concisely communicate the desired messages – such as the breakdown by states and the difference between Labor and Liberal electorates. Furthermore, this helped overcome some of the limitations of the choropleth graphic in that densely populated regions are poorly represented. This is especially the case in Australia where each electorate is designed to have approximately the same number of people. Because of this and the way Australia's population is distributed, there are many electorates packed tightly around capital city's such as Melbourne and Sydney while in the centre of the country some electorates take up almost the entire state. Thus, these additional graphics help convey the full findings for each of these questions.

Returning to the initial name of the design, it has been called 'iterative' due to the way in which the data is to be initially displayed. As discussed with the first design, having all the visualisations appear at once would create clutter on screen distracting some users. However, as identified with the second separating them into individual tabs tests the limits of the visual system and working memory. To address these issues, initially only the first row would be shown containing information relating to the first research question. Once the user has read and explored the information on screen they can then click on a 'View' button to unveil the second row containing the demographic data and subsequently a second 'View' button to unveil the model projection. Hence the 'iterative' naming convention. This method was chosen for several reasons. First it ensures that the user has read through all the on-screen information for each section before revealing the graphics. This allows them to prepare and familiarise themselves with the information that is to be shown and the different interactive elements that they can use to explore the data. This limits the opportunity for the narrative visualisation to be incorrectly interpreted or misused by ensuring the audience has the appropriate knowledge to navigate it correctly. Second, it reduces the amount of visual stimulus on screen initially and allows the user to become familiar with an individual section before moving on to the next. Third, in conjunction with this, it addresses the limitations in working memory as users are motivated to engage with and understand the content at each stage of the narrative and each has a clear focus on one research question. However, unlike the tab-based layout discussed earlier, users are better able to navigate between sections to compare the findings from each and follow the logical flow from one point to the next. Finally, from an implementation perspective, the efficiency and speed of the code on start-up is improved as only the first section needs to be rendered initially. Then, as the user moves throughout the visualisation each section can be loaded individually as required.

Regarding implementation, which will be discussed in detail in the following section, this chosen visualisation will be produced using R Studio and the programming language R. The primary packages required to incorporate the graphics and interactivity discussed are shiny, shinydashboard, ggplot2, and leaflet. At the time of design, it was estimated 10-12 hours would be required to code the solution at no monetary cost.

Name: Lachlan Moody
Student Number: 27809951

Implementation

Prior to coding the graphics necessary for the visualisation, first some initial data wrangling and cleaning had to be performed. This process is detailed in full the previous *Data Exploration Project* and was much the same except for some small additional changes. This included modifying the existing data to fit the desired table output and transforming the existing demographic information so that it could respond to user interaction. This data manipulation was performed in R Studio and was primarily dependent on the dplyr (2020) and janitor (2020) packages for tabular data and the broom package (2020) for the model output. Additionally, in the interest of optimising the visualisations performance, all of the cleaning was performed in a separate R Markdown file and the resulting outputs were saved as RDS objects. These files were then loaded into the application file to save computing time.

As the visualisation is centred around three choropleth maps, the package rmapshaper (2020) was essential during this phase and was used to simply the polygons stored and drawn by R. During initial testing using just the shapefile provided by the Australian Electoral Commission (2019), the visualisation would often take 5-10 seconds to load initially, and then again, each time the user input was changed. This was not ideal from an implementation perspective as one of the key challenges of effective communication is directing and holding attention. The longer users are forced to wait for the visualisation to react to their inputs, the more likely it is that they will become frustrated or distracted, thus limiting the effective communication of the desired message. Thus, ensuring the visualisation responded swiftly to these inputs was essential and would not have been possible without the rmapshaper package (2020).

With the pre-processing complete, the communication medium chosen was a web-application produced using the packages shiny (2020) and its extension shinydashboard (2018). The reason for this was that shiny allows for the integration of R generated graphics with interactive user input which helps drive engagement and exploration. Whereas, shinydashboard, as an extension, helps present the information in a segmented and stylised manner. The initial view of the visualisation is shown below.

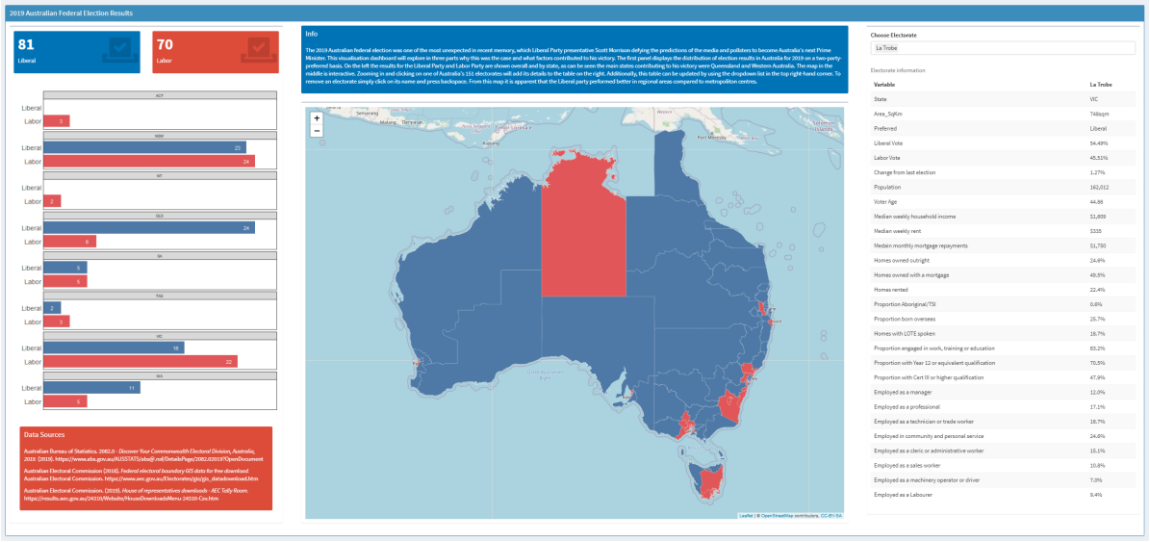


Figure 6: Narrative Visualisation Row 1

This section contains information relevant to the 2019 Australian federal election results and corresponds to the first row of the layout shown in Figure 5. Central to this design is a choropleth map of the election results by electorate. This graphic was chosen as it captures the geographic distribution of each party's support base and contains key information relating to the size and location of the seats each won.

This, and all further maps, were produced using the R package leaflet (2019) as it can be used to create interactive visualisations using spatial data. Through this software, users of the visualisation are able to zoom in on specific areas relevant to them and identify how the election swung within their local area. Additionally, maps produced using leaflet also have hover and click functionality which has been utilised to show the electorate name and Liberal vote percentage when hovered and to update the table on the right with the chosen electorate when clicked. The table contains a combination of geographic, voting and demographic data for each electorate that has been selected and was created using the kableExtra package (2020). This was done to provide a broad overview of all possible variables of interest do a degree of numerical precision not possible from most plots or charts. Looking to the left side of Figure 6, the bar chart of election results by state has also been included and was created using ggplot2 (2016) as it allows for highly customisable graphics. Inside each bar, the exact number each represents has also been included to ensure no information is incorrectly interpreted. The reasons for this graphics use have been discussed in the previous design section.

Unlike the initial design, however, several additional elements have been included to improve the narrative power of the visualisation. First of these is the two value boxes shown in the top left-hand corner displaying the total number of seats won by each party. This helps give a quick high-level view of who won the election and how close it was. Also included are the blue box at the top of the screen and the red box in the bottom left. The top one contains information providing context, signposted messages, and user instructions that ensuring each user has the appropriate knowledge to uncover all the visualisation is attempting to convey. Whereas the bottom one contains the data sources which helps provide credibility to what is shown. These boxes are exclusive to shinydashboard. The changes discussed were made in the implementation stage and not the design stage as the existence and features of the package were not known at the time. Any future design process conducted will benefit from enhanced knowledge in this area.

As a final point for this first view, the colour scheme chosen is based on the traditional colours for each political party in Australia – blue for the Liberal's and red for Labor. This allows users who are familiar with the political system in Australia to more easily identify differences between the two and reduces the amount of new information shown to them. The palette used is kept consistent throughout the whole page and is used for the value boxes, the bar graph and the choropleth map.

As discussed previously, initially what is shown in Figure 6 is all that can be seen on screen. However, as the user scrolls down the page the next section appears as shown on the next page.

Name: Lachlan Moody
Student Number: 27809951

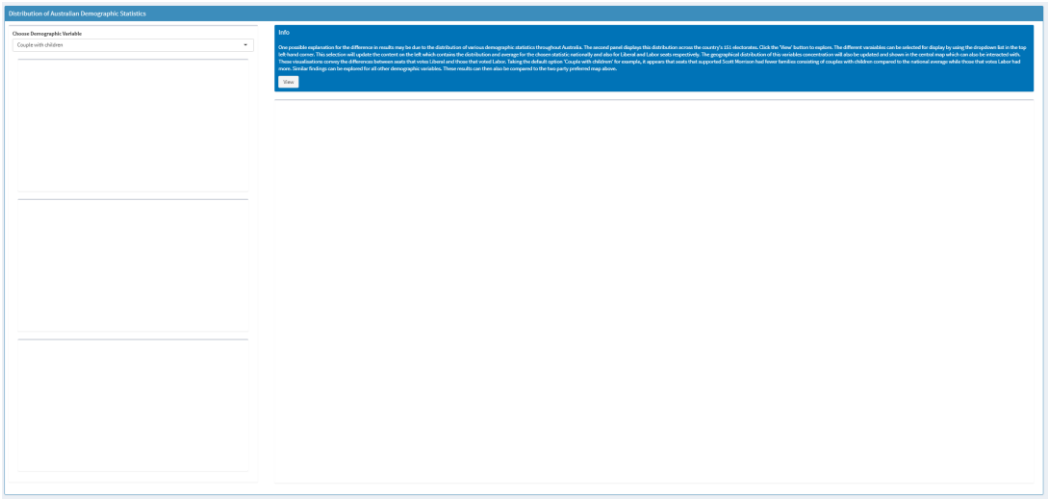


Figure 7: Narrative Visualisation Row 2 Initial View

With the blue text box being the only object of interest on screen, users will naturally read through this information thus preparing them for what is to be shown on the page. Additionally, included in this box is a button labelled 'View' which reveals the full design shown below and incorporates the 'iterative' functionality discussed during the design stage.

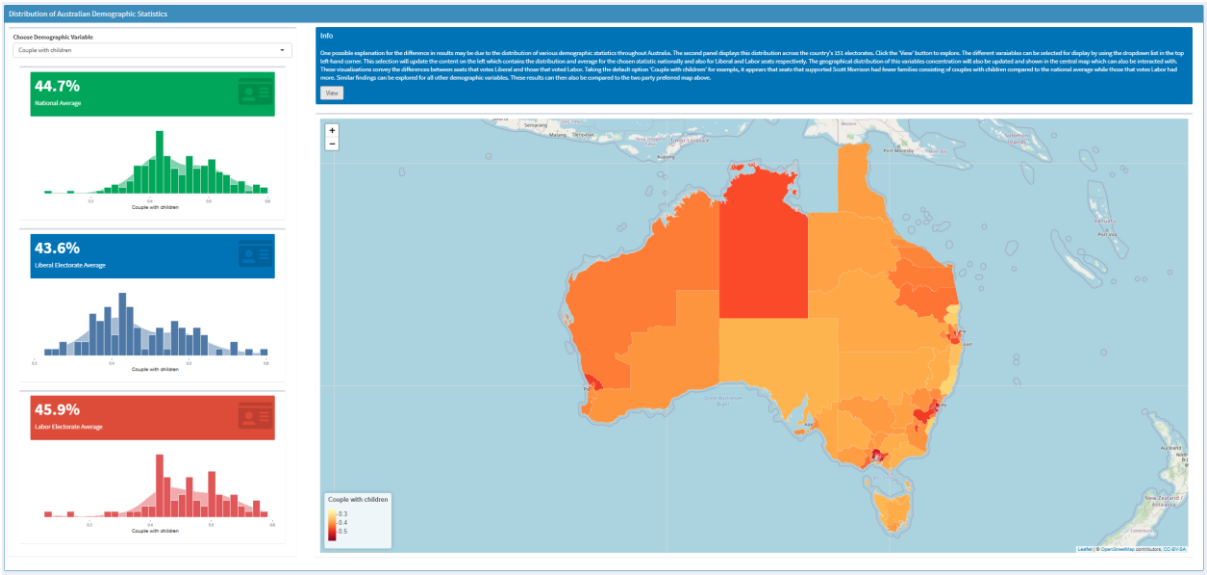


Figure 8: Narrative Visualisation Row 2 Full View

Once clicked, the layout transforms to align with the initial design for the second row from the realisation design in Figure 5. What has remained the same is the dropdown list for the various demographic variables and histogram plus density plots on the left, and also the choropleth map in the middle. The alignment of the plots on the left however has changed. While initially designed to be laid out horizontal, during implementation it was found that this created too small an area for the map and that the dimensions of the histograms became warped. Thus, to resolve this issue, the plots were rearranged vertically down the left-hand side. This recentred the focus of the section on the choropleth map of Australia and the geographic distribution of the chosen variable which better aligned with the research question.

In addition to this, as before, several elements were incorporated into the design during the implementation phase. This includes the value boxes on the left containing the averages nationally and for the Labor and Liberal electorates pertaining to the demographic variable selected. This decision was made to give users of the application a quick reference point to compare differences between the three, thus highlighting one of the key narrative messages – that different demographic variables are associated with areas that vote Liberal or Labor. The colour scheme used for these boxes and histograms matches the one used above ensuring a cohesive flow to the design. Additionally, green was chosen for the national statistics to differentiate it as a new data group not shown above. Similarly, this reasoning was carried over to the choropleth map which used a sequential colour palette rather than a categorical one to represent the different levels of each value chosen. This way the saturation, in addition to the colour, can also be used to identify where each variable is concentrated, thus helping communicate where heightened levels of these different values are most commonly seen throughout Australia. Further to this purpose, this is the only one of the three maps to include a legend displaying the levels and direction of the values shown. Also, a new addition is the info box at the top of the page. Kept consistent with the one used in the first section this box provides the user context for what is being shown and provides details on how to explore the visualisation.

The last section of the visualisation relates to the third research question and attempts to project the election outcome using electorate results predicted from a linear model of demographic variables. As with the second section, this sheet is also blank when loaded except for a blue info box (see Figure 7 for indication). Once the 'View' button is clicked the visualisation appears as below.

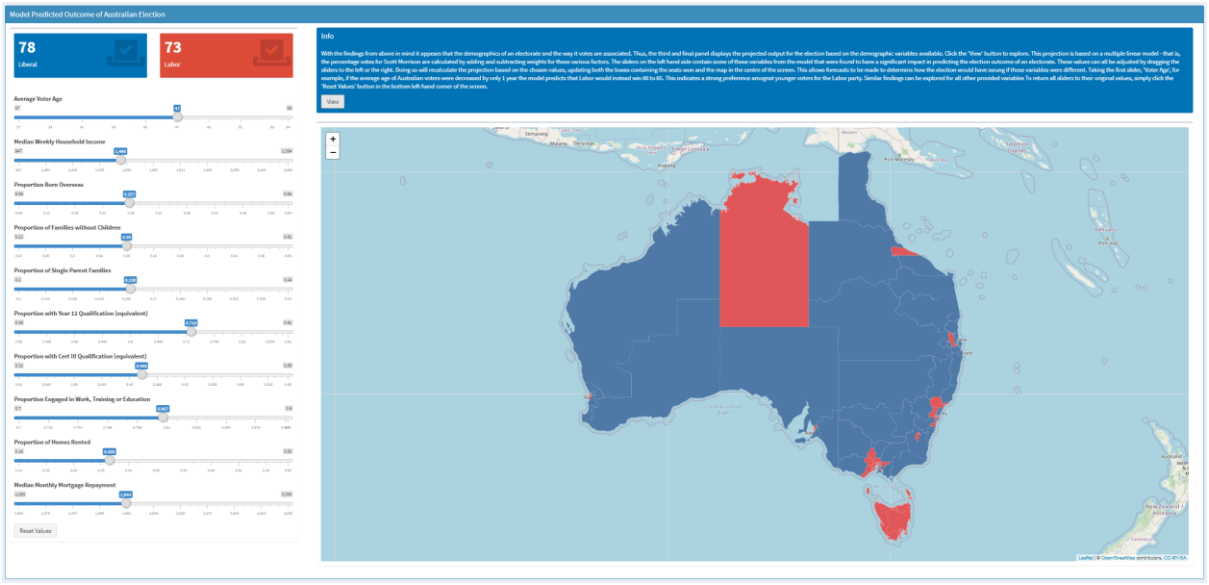


Figure 9: Narrative Visualisation Row 3 Full View

The final row contains quite a simple layout, with again the choropleth map being the central focus of the section – coloured using the same palette as the first. This implementation decision was done to help the user connect what is on screen with what they have seen previously. Additionally, as also seen before, the left-hand side contains two value boxes containing the currently projected model outcome for each of the Liberal and Labor party. This was not included in the design phase but has been used here to allow for a

more precise comparison between the actual and predicted election results than a simple visual comparison.

Also, somewhat different from the earlier design is the inclusion of the sliders on the left-hand side. The initial design idea was to have the full model equation shown with users able to input values for each regressor. This idea was adjusted however to the approach now shown for several reasons. Firstly, the number of estimators for the equation is quite high at 26. Thus, including the full model equation would most likely clutter the screen and be difficult to read. Secondly, it was more difficult to restrict what values could be entered. This would affect the accuracy of the model and open up more opportunities for error as potentially values more extreme than the observed ones may be chosen. With a slider input however, it is easier to set the bounds on the upper and lower end. Finally, and most importantly, as the intended audience is more general in nature, expectations relating to prior knowledge and understanding must be considered. While most people have some working knowledge on regression and linear modelling it is certainly possible that its implementation may not be the most appropriate in this context. Presenting these explanatory variables in a slider format however is more easily understood by most people, ensuring the visualisation can reach as many people as possible. While it does limit the full power of the model from being explored, all of the variables considered most significant, both by the model output and the author, have been included.

These sliders, in conjunction with the map and value boxes, allow users to quickly and easily identify that different factors are associated with an electorate's voting patterns - one of the primary narrative aims of the visualisation.

Incorporated underneath these sliders is an additional important feature not considered during the design process. One to reset the values. This issue was only identified during the implementation process as it was found to be difficult to remember the original location of each of various slider positions. However, with the power of shiny these original values are able to be stored and recalled at the click of a button. This improves the exploratory power of the design as users are further encouraged to adjust the sliders without needing to remember their original locations and are thus able to focus more of their mental focus viewing the outcome. This leads to more values being altered and adjusted helping convey one of the primary messages of the narrative that certain values are positively correlated with a Liberal or Labor-preferred outcome.

A final addition made to the original design in Figure 5 is the info box at the top of the screen. As the calculations being done for this section are slightly more complex than the rest, more space is dedicated to explaining the process in full and how to interpret the various possible findings.

User Guide

The code for running the application is contained within the file named app.R and can be viewed by clicking on the 'Run App' button in the top right-hand corner as highlighted in the image on the following page.

Name: Lachlan Moody
Student Number: 27809951

Figure 10: Viewing the Visualisation

Additionally, to ensure the correct libraries are installed to render the output, the code on line 3 can be uncommented by removing the hashtag symbol at the start of the line and running the code. To improve the load time of the application, it is better to replace the hashtag symbol before launching as this will prevent the code from running each time the button is pressed.

Once loaded the page appears as shown in Figure 6. This section has two points of interaction as shown below.

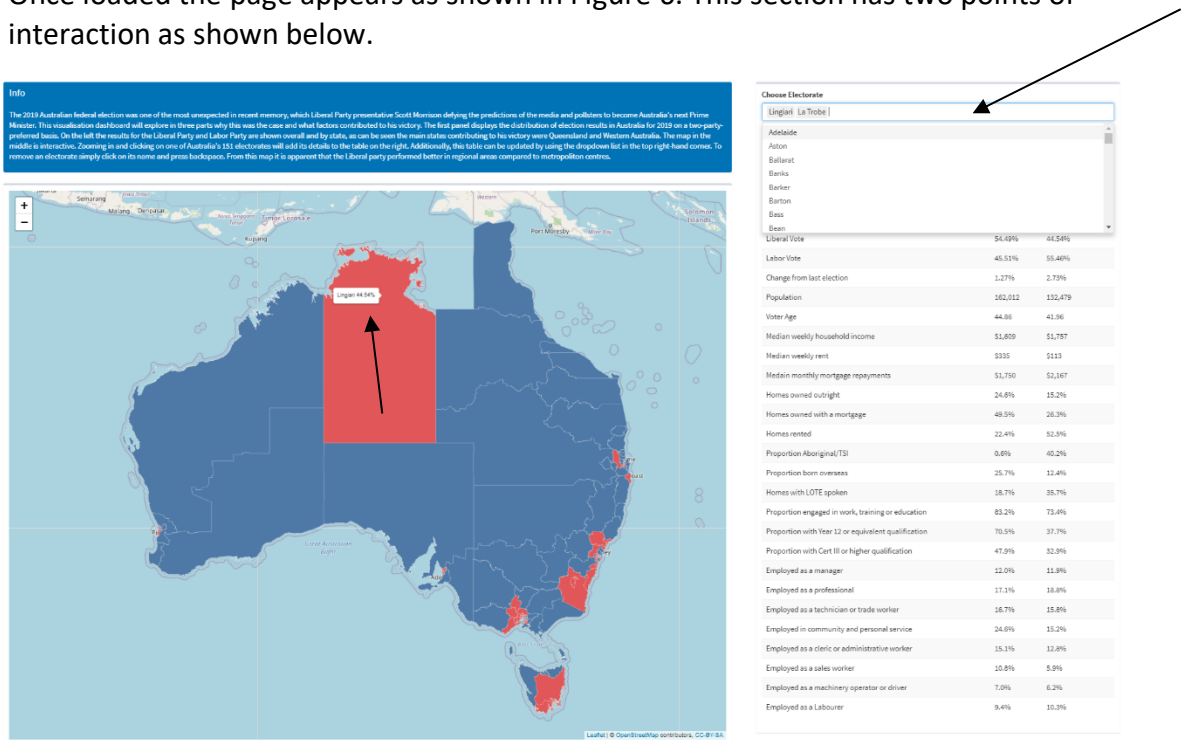


Figure 11: Row 1 Interactions

To view the percentage of votes won by Scott Morrison and the Liberal party, users can hover over each electorate. This will also provide the electorates name and, if desired, can be clicked to add it to the dropdown list in the top-right corner. Additionally, this list can be updated by also scrolling through the selections which are provided in alphabetical order. Either method will add an extra column to the table containing information for the electorate relating to the demographic, geographic and political statistics down the first column. Any number or combination of states can be added in this way. In order to remove states, users are required to click on its name in the selection window and press the delete or backspace key. These instructions, all further ones, are also provided in the blue info box in each row.

The second page has similar interaction opportunities as shown in Figure 12. However, to have these options appear on screen, users must first click the 'View' button which is located within the blue info box.

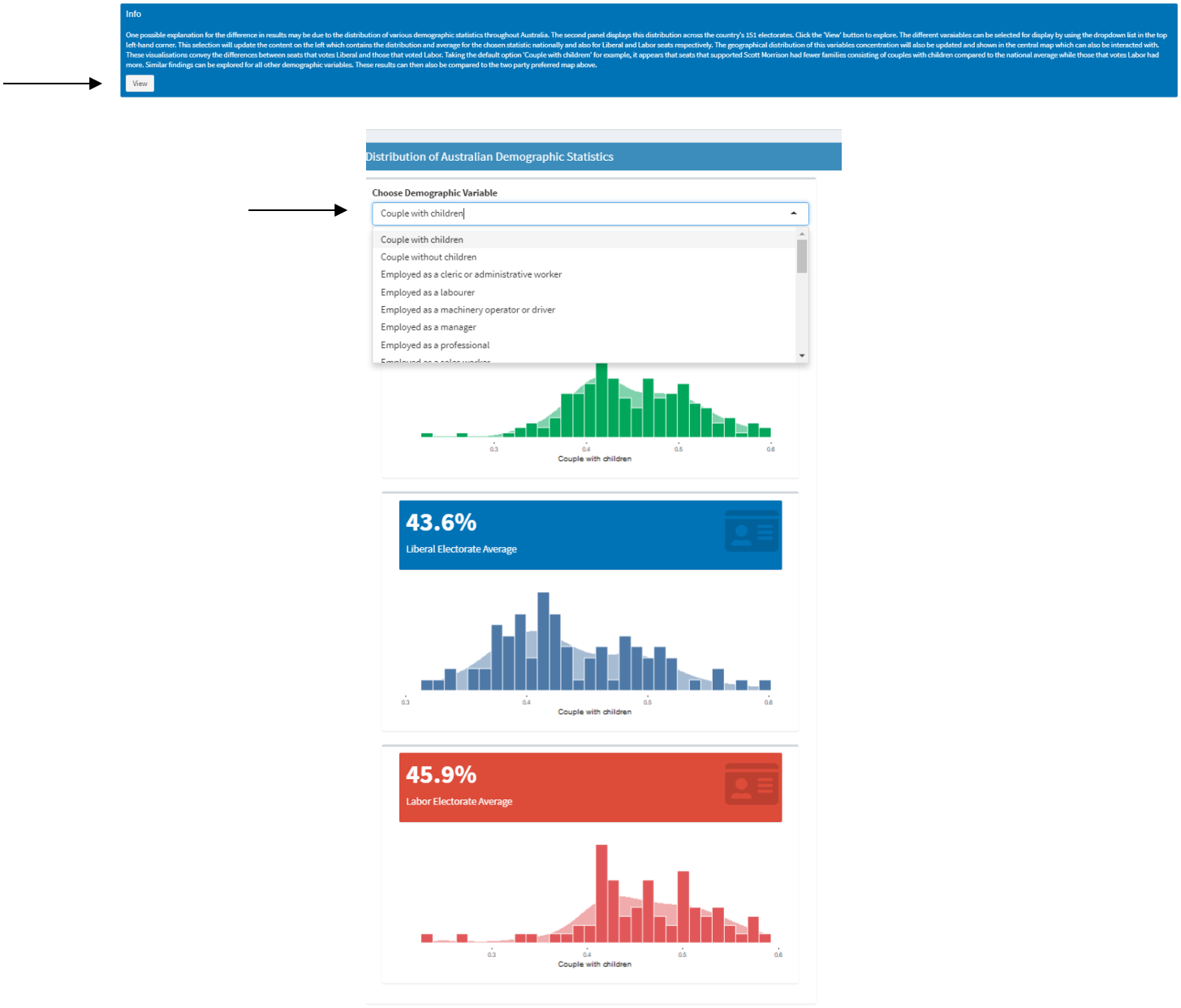


Figure 12: Row 2 Interactions

This dropdown list contains all demographic statistics contained within the ABS data (2019). Like the previous dropdown list, the output can be updated to match the desired demographic variable by selecting from the options, which are also provided in alphabetical order. However, in this case only one can be selected at a time. This will update both the plots shown in Figure 12 and the choropleth map in the centre of the screen as displayed in Figure 8. This map can also be hovered over to get the electorate name as well the value for the corresponding demographic.

The final row has the most opportunity for interaction and exploration. In addition to the same 'View' functionality as mentioned above, there are 10 sliders that can be used to adjust the observed averages for different demographics. Doing so will adjust the projected

outcome of each electorate. To adjust these sliders, simply drag the circle to the left or the right to explore how the projected results change as variables are adjusted. In addition, below these sliders, is a button to reset these sliders to their original values. To view how this changes the election outcome, users can check the value boxes at the top of the screen or hover over the election map to get each electorates name and new outcome.

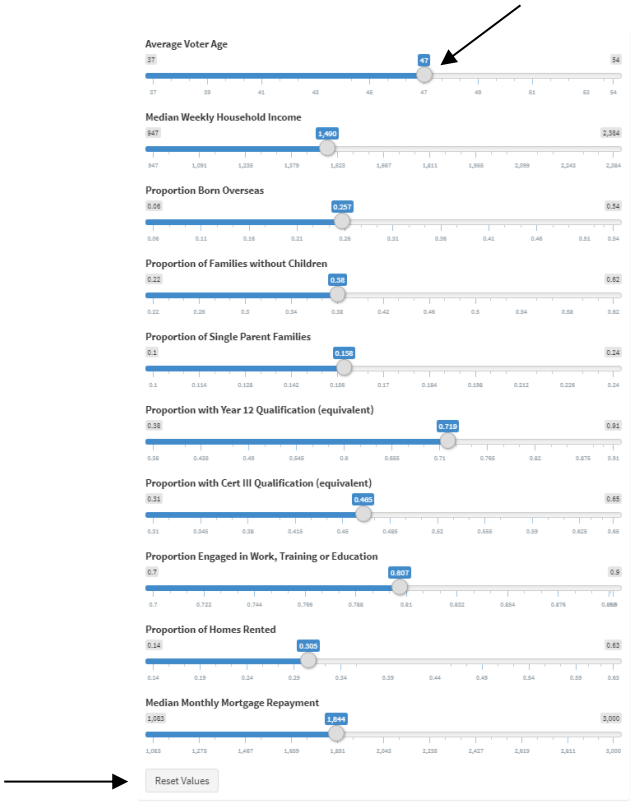


Figure 13: Row 3 Interaction

Conclusion

This entire process has highlighted several key findings relating to the research questions proposed during the introduction. Firstly, it was found that Australia's election results were distributed very differently throughout the country, with Scott Morrison and the Liberal party's victory being attributed primarily to results in Queensland and Western Australia and their preference in regional areas. Secondly, several of Australia's demographics appear to play an pivotal role in determining the electoral outcome of an electorate with some such as voter age and couples without children being positively associated with the percentage of votes received by the Liberal's and others, such as the proportion of recent migrants being positively associated with Labor.

In addition to these empirical findings, a number of technical operations have been learnt. This included developing spatial representations of data using leaflet, designing interactive visualisations using shiny and shiny dashboard, and applying the Five Design Sheet methodology to create a narrative visualisation. In hindsight, the programmed result could have been improved by exploring different designs with less user interaction in order to communicate a more structured narrative. Additionally, different visualisation tools such as D3 could be tested to compare the effectiveness of each.

Bibliography

Australian Bureau of Statistics. 2082.0 - Discover Your Commonwealth Electoral Division, Australia, 2019. (2019). https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2082.02019?OpenDocument

Australian Electoral Commission (2018). Federal electoral boundary GIS data for free download. Australian Electoral Commission. https://www.aec.gov.au/Electorates/gis/gis_datadownload.htm

Australian Electoral Commission. (2019). House of representatives downloads - AEC Tally Room. https://results.aec.gov.au/24310/Website/HouseDownloadsMenu-24310-Csv.htm

Winston Chang and Barbara Borges Ribeiro (2018). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1. https://CRAN.R-project.org/package=shinydashboard

Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2020). shiny: Web Application Framework for R. R package version 1.5.0. https://CRAN.R-project.org/package=shiny

Joe Cheng, Bhaskar Karambelkar and Yihui Xie (2019). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.3. https://CRAN.R-project.org/package=leaflet

Sam Firke (2020). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.0.1. https://CRAN.R-project.org/package=janitor

Roberts, J. C., Headleand, C., & Ritsos, P. D. (2015). Sketching designs using the five design-sheet methodology. *IEEE transactions on visualization and computer graphics*, *22*(1), 419-428.

David Robinson, Alex Hayes and Simon Couch (2020). broom: Convert Statistical Objects into Tidy Tibbles. R package version 0.7.2. https://CRAN.R-project.org/package=broom

Andy Teucher and Kenton Russell (2020). rmapshaper: Client for 'mapshaper' for 'Geospatial' Operations. R package version 0.4.4. https://CRAN.R-project.org/package=rmapshaper

Wang, K. T. (2020, September 29). Five Design-Sheet Methodology Approach to Data Visualisation. Retrieved from https://towardsdatascience.com/five-design-sheet-methodology-approach-to-data-visualisation-603d760f2418

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

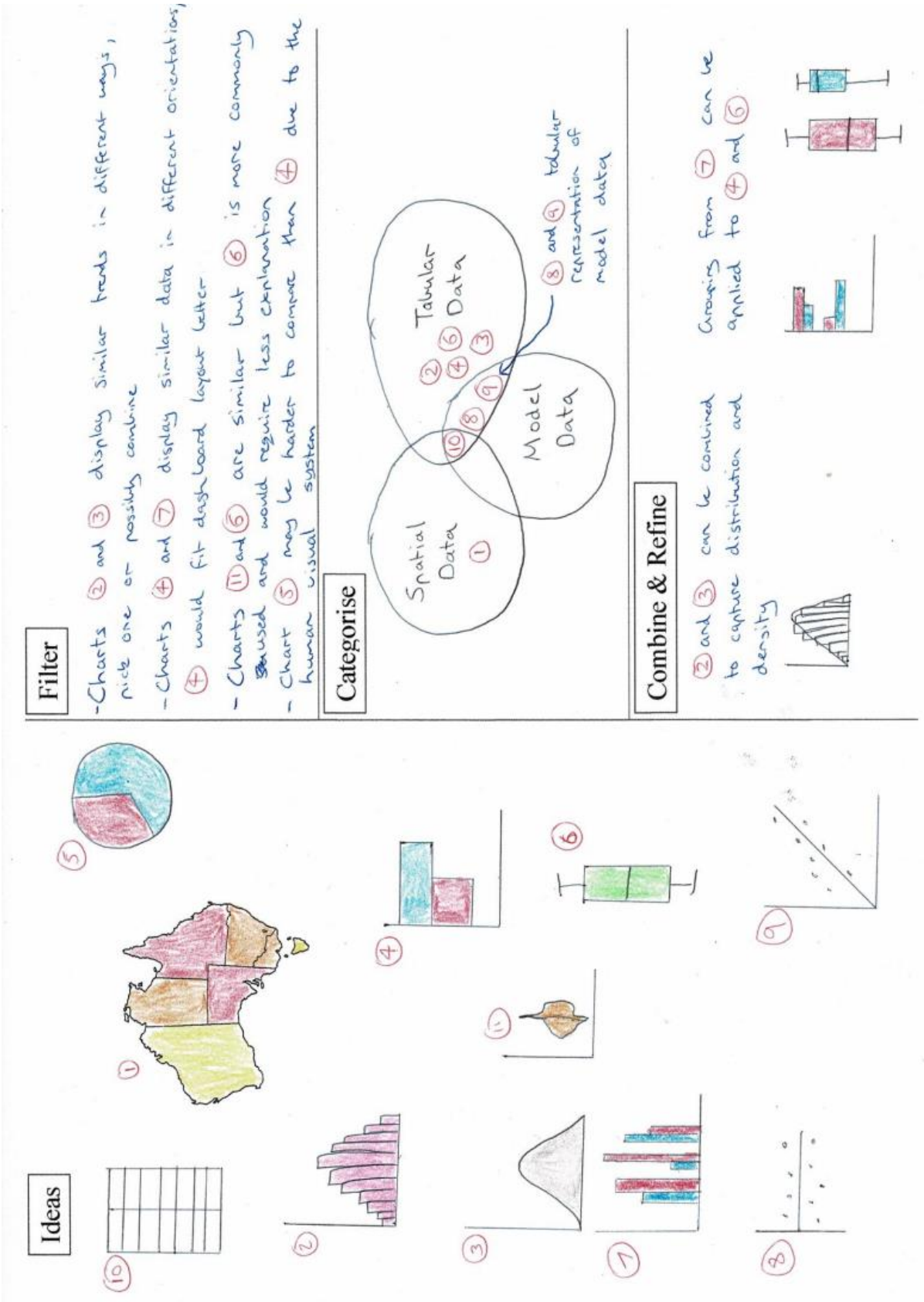
Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.1. https://CRAN.R-project.org/package=dplyr

Hao Zhu (2020). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.1. https://CRAN.R-project.org/package=kableExtra

Name: Lachlan Moody
Student Number: 27809951

Appendix

Appendix 1: 5-Design Sheet 1

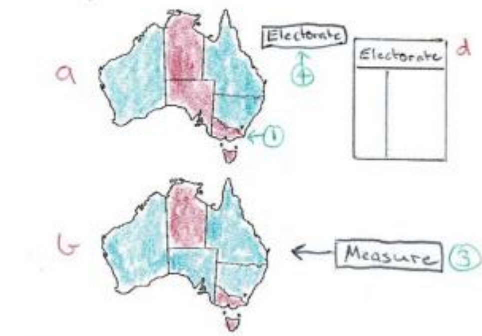


Name: Lachlan Moody
Student Number: 27809951

Appendix 2: 5-Design Sheet 2

Title	Predictors of Australian Election Results		
Author	Lachlan Moody		
Date	21/10/20	Sheet	2
Task	Spatial representation of data using one page		

Layout



④ → $y = b_0 + b_1 x + b_2 x$

Operations

① Click on an electorate to update table d

② Dropdown list of electorates to update table. Will also update via ①

③ Dropdown list of demographic measures for each electorate. Will update output of b

④ Update formula to project model election outcome. Will default to R generated model.

Focus/Zoom

a: choropleth map of election results by electorate

b: choropleth map of demographic statistics by electorate

c: choropleth map of election results by electorate based on model

d: table containing detailed summary statistics for each electorate chosen

Discussion

Positives:

- Easily compare election results, demographic statistics and model predictions on one page
- Detailed information available based on user interest / input

Negatives:

- Similar visualisations for all measures
- User may be overwhelmed due to visual clutter

Name: Lachlan Moody
Student Number: 27809951

Appendix 3: 5-Design Sheet 3

Title	Predictors of Australian Election Results
Author	Lachlan Moody
Date	21/10/20 — Sheet 3
Task	Spatial representation of data using multiple pages

Operations

1. Click on electorate to update table d
2. [Electorate] Dropdown list of electorates to update table. Will also update via 1
3. [Measure] Dropdown list of demographic measures for each electorate. Will update output of b
4. Update formula to project model election outcome. Will default to R generated model.

Focus/Zoom

a: choropleth map of election results by electorate

b: choropleth map of demographic statistics by electorate

c: choropleth map of election results by electorate based on model

d: table containing detailed summary statistics for each electorate chosen

Discussion

Positives:

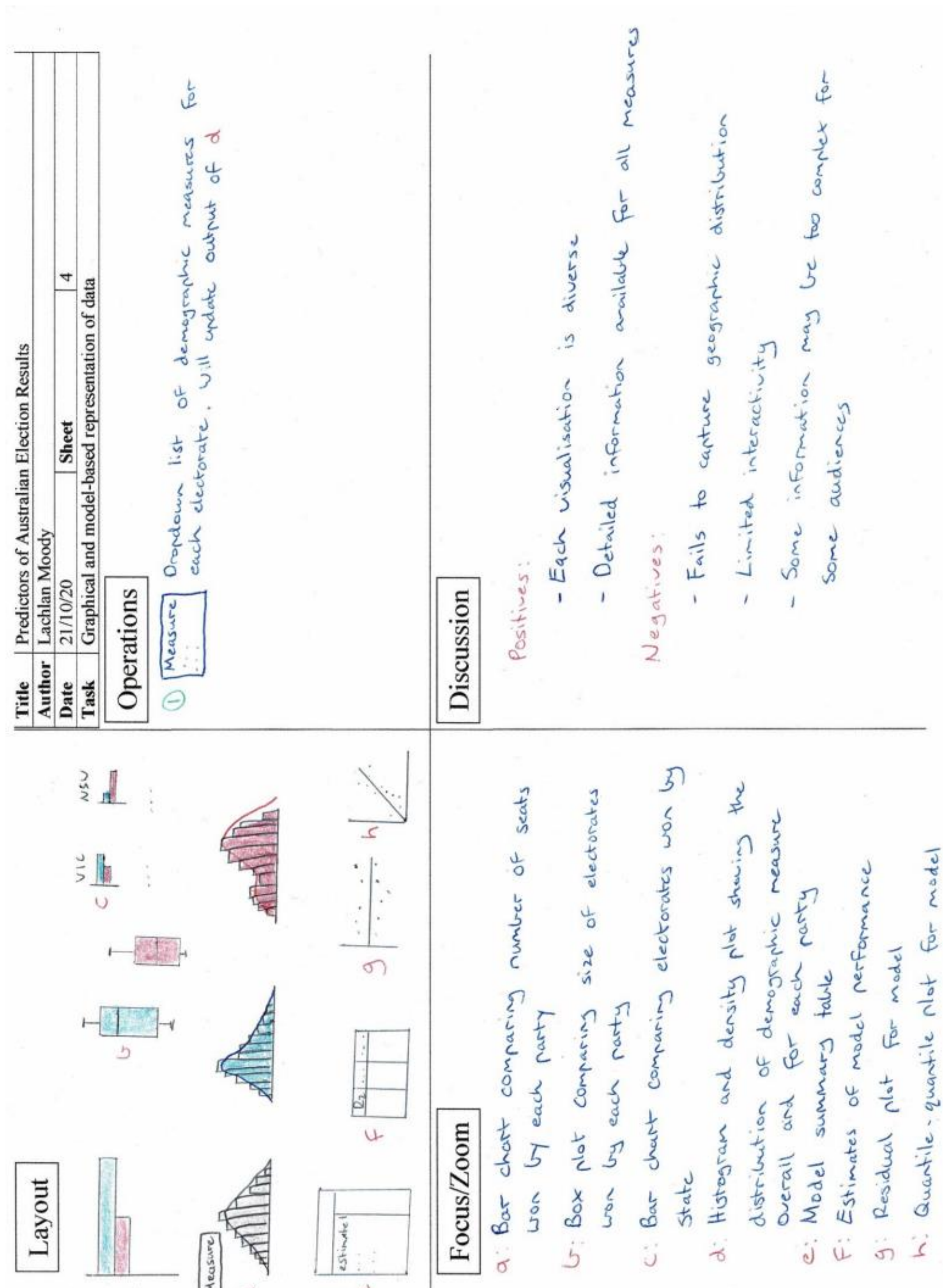
- Each tab has a clear focus on a specific question
- Detailed information and comparisons available based on user interest/input

Negatives:

- Hard to make direct comparisons across sheets/tabs
- Similar visualisations for all measures

Name: Lachlan Moody
Student Number: 27809951

Appendix 4: 5-Design Sheet 4

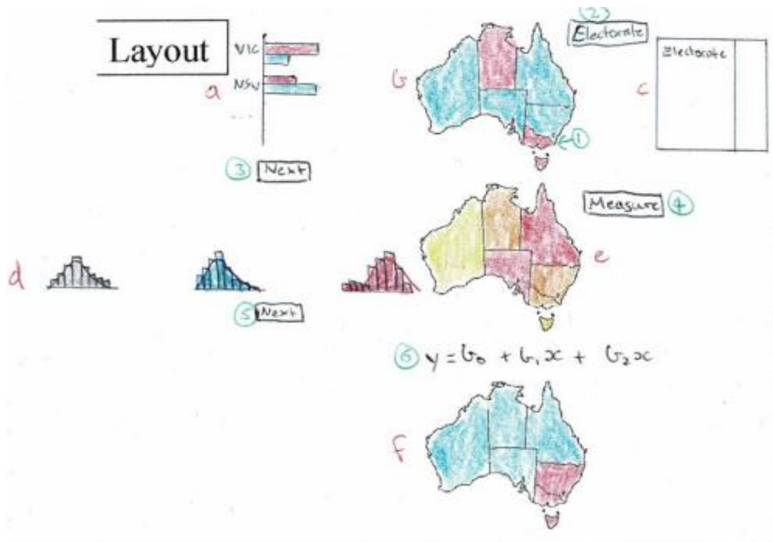


Name: Lachlan Moody
Student Number: 27809951

Appendix 5: 5-Design Sheet 5

Title	Predictors of Australian Election Results		
Author	Lachlan Moody		
Date	21/10/20	**Sheet**	5
Task	Spatial representation of data using iterative design		

Layout



Operations

(1) Click on an electorate to update table c

(2) [Electorate] Dropdown list of electorates to update table. Will also update via (1)

(3) Unveils everything below button up until (5)

(4) [Measure] Dropdown list of demographic measures for each electorate. Will update output of e

(5) Unveils remainder of visualisation

(6) Update Formula to project model election outcome. Will default to R generated model.

Focus/Zoom

a: Grouped bar chart comparing electorates won by state

b: choropleth map showing the election results by electorate

c: table containing detailed summary statistics for each electorate chosen

d: histogram and density plot showing the distribution of demographic measure overall and for each party

e: choropleth map of demographic statistics by electorate

f: Choropleth map of election results by electorate. (based on model

Detail

Algorithms/Methods: The visualisation will be produced using R Studio

Dependencies: The primary packages required to build the visualisation are shiny, ggplot2 and leaflet

Estimates of Cost: Approximately 10-12 hours will be required to implement the chosen design.