Narrative Visualisation Project

Australian newborn naming analysis

Shihan Zhang
31268102
Tutorial 24
Shirin Ghaffarian Maghool

Introduction:

In this visualisation project, I created a Shiny app with R to display several interactive plots to answer the research question from the "Data Exploration project". In this project, my audiences will not only include the groups from the last project but also published to the public.

From the app, Tab1 ("Name Ranking of Each Year") and Tab2("Compare Single Years") will answer the first research question "From 2013 to 2019 which 10 names were most used for males/females".

Tab3("CAGR of Names") will answer the second research question" Which name for females and male was a sudden increase and in which state it happened".

Design:

To provide an excellent Shiny application with multiple plots using the navigation in the "Tabs" window would be the right solution.

## Tab1 "Name Ranking of Each Year":

To answer the research question "From 2013 to 2019, which 10 names were most used for males/females".

The first key elements are the timeline from 2013 to 2019. The second key elements are the ten names were most used. The third elements are gender (for males or females). Combine the above features we have a final idea which is "the ranking of top 10 female or males' names between 2013 and 2019".

human perceptual system and human communication assumptions:

There are many plot styles can use for display ranking and comparison, but the ordered proportional symbol is well meet the conditions of this case. So, I use the bubble chart to represent the counts of each name and ordered by ranking (continuous size changes can speed up information acquisition). Also, use the colour with significant warmth gaps to represent names and the different size of the shape to represent the counts. For more abundant information, we also include the top 3 and top 5 options for the user to choose. All selections are placed in the selection panel and placed on the left side of the page to make the page look neat and orderly.

Therefore, we need to have a selection to switch the plot for different gender and slider for year selection and the selection for different ranking. For plots, the x-axis is the counts of the name, and the y-axis is the male or female name.
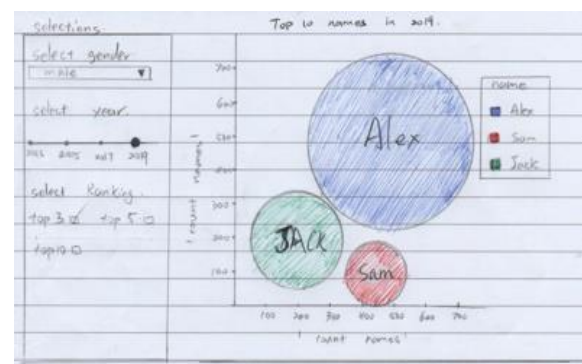
Tab2 "Compare Single Years":

This tab is the idea extends from the Tab1.

human perceptual system and human communication assumptions:

This plot aims to show the changes in the counts of each name with two single years. The columns work well for showing change over time and with the paired columns user can easily see the changed with two different years. The paired bar chart can also give a good comparison. In this case, we have more than ten names to display, and the unordered column will be tricky to read so we must order the column. By using des-order, the plots can be clean and comfortable to read. For the different year, I use contrasting colours (blue and red) to give users a decent view.

Therefore, we need to have a selection to switch gender and the selection for switch plot style and the selection for select year1 and year2. For plots, the x-axis is the name, and the y-axis is the counts of name, the fill for the column or bar is the year.
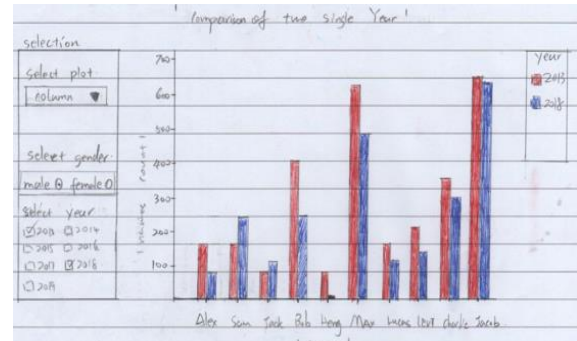
Tab3 "CAGR of Names":

To answer this research question "Which name for females and male was a sudden increase and in which state it happened".

The first key element is to find increases in the name for each year. The second key element is to find the location of the name. The third key element is to see gender. The most important thing is to calculate the compound annual growth rate (CAGR).

human perceptual system and human communication assumptions:

this plot aims to show the CAGR for each year so we can find which one has suddenly increased.

In the design sheet sample, I use the line char to shows the CAGR changes. But when it comes to real plotting, there are more than 20 lines and make the chart unreadable. So, I choose to use the stacked bar plot, which is suitable for this situation. Make bar plot ordered by CAGR can give a clear view and speed up the understanding. Added a vertical line to represent the mean value of CAGR, it can provide users with quick comparison to understand the size of the data expressed by each bar. To show the location of the name with considerable CAGR, we use leaflet map. We filter out the target group and add a circle marked on the map with high contrast colour and size. The colour represented a range of CAGR (e.g." blue 0-20"). The map is zoomed and focus on the region where the name belongs. The map is placed under the plot to

show the connections of both plots. Because not everyone knows and understands what is "CAGR".
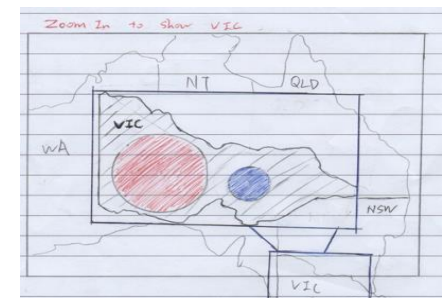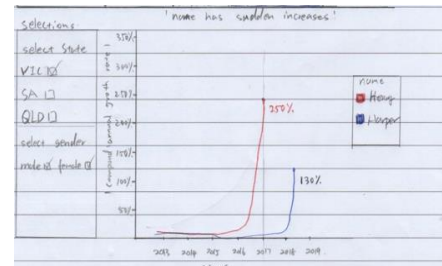
So below the plot, I need to add text explaining "CAGR". The compound annual growth rate (CAGR) is the rate of return required for an investment to grow from its initial balance to its final balance (MURPHY, 2020). But No investment is involved in our case. We only use CAGR to reflect the growth of the data.

Therefore, we have a selection for switch gender, the selection for the select region and the selection switch select maps(gender). For plots, the x-axis is the CAGR, and the Y-axis is the name. For the leaflet map, CAGR for marker size.

Implementation:

Libraries:

ui.R: Purpose:

- library(shiny)   for build Shiny App
- library(leaflet)   for build leaflet map
- library(plotly)   for build plots

server.R:

- library(shiny)   for build Shiny App
- library(leaflet)   for build leaflet map
- library(tidyverse)   for Data frame processing (subset(), unique(), head())
- library(ggplot2)   for build plots
- library(plotly)   for build plots with more interactive functions works with ggplot2
- library(forcats)   for calculations(mean(), median())

reading data source:

reading data source:

In this project, we required to use the data source file from the last project.

To read the data source we use "*au_baby <- read.csv("au_baby.csv")*"

Build UI frames:

For the overall layout, we need a page that automatically adjusts the layout, so we use "fluidPage".

The "tabPanel" provide the tab window to show the different plots relative to our research question.

To make the structure of the page look neat and orderly, we use "sideBarPanel" and place all the selection panels in "sideBarPanel" and set on the left side of the page. The selection group has a title to indicate the propose. We want to arrange all the sub-options vertically, so we use the "column" attribute here. Use "selectInput" (drop-down selection) to set the switch gender (switch plot). Use "sliderInput" (2013-2019) to set the year. Use "radioButtons" to set the target ranking group. Below the option group, we set up some panels that display data sources and author information.


```
    # third tab bar
    tabPanel("CAGR of Names",
        fluid = TRUE,
################# T3 selector panel
        sidebarLayout(
            sidebarPanel(
                # panel title
                titlePanel("Selection Panel"),
                fluidRow(column(
                    8,
                    # panel for select gender
                    selectInput("T3gender",
                        label = p("Select State"),
                        choices = list("Female" = 0, "Male" = 1),
                        selected = 1
                ),
                    # select state
                    radioButtons("state", "Select State:",
                        c(
                            "victoria" = "Victoria",
                            "Queensland" = "Queensland",
                            "South Australia" = "South Australia"
                        ),
                        selected = "victoria"
                ),

                    # swwitch map
                    radioButtons("map", "Select Map:",
                        c(
                            "Male Name" = "MN",
                            "Female Name" = "FN"
                        ),
                        selected = "MN"
                ),
```

Use the main panel to display our plots and add help text to display timely information to help users quickly master the use of plots.

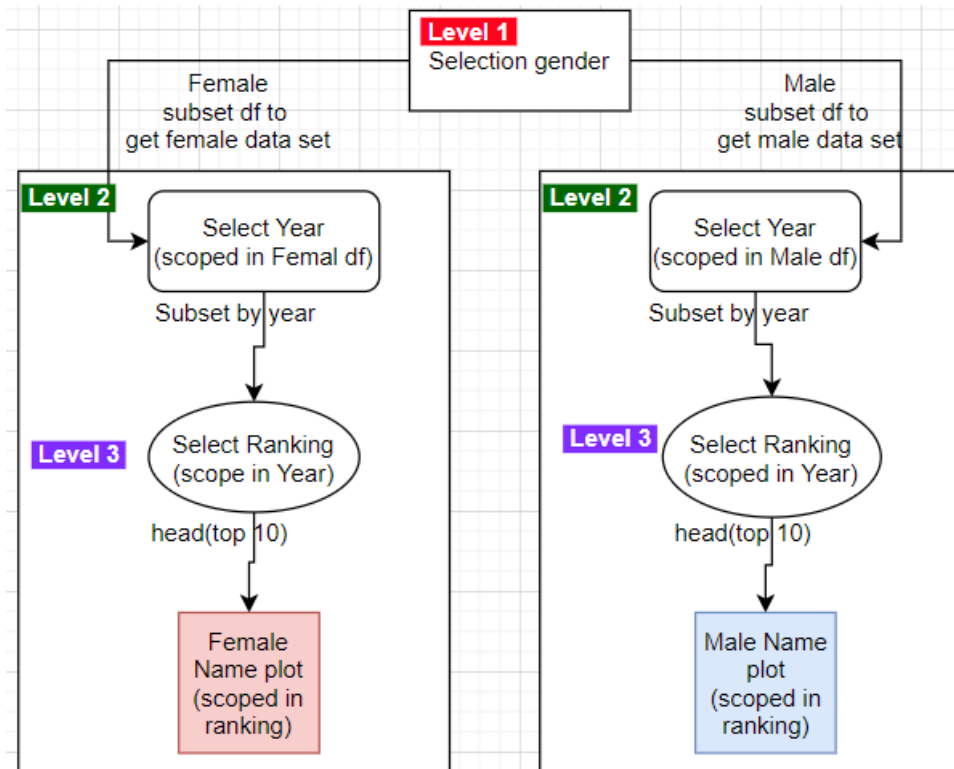Tab 2 Window and Tab 3 Window use the same layout as Tab 1.

Selection and plotting (Major workload):

Why use the 3-level structure?

The advantage of this method is that you can make plots for each theme, and the options can react quickly to reduce delays.

Tab 1 Structure

Based on our data frame, I use a 3-level structure to design the relationship between selection and plot. At the first level, we use the select gender to subset the data frame base on selected gender. In the second level, we use the select year to subset the data frame (scoped in gender) by the selected year. In the third level, we use the select ranking to get the first 10 rows in the data frame (e.g. scoped in year = 2013).

Finally, we do the plotting base on third level data frame then render to the page.

This 3-level structure is used for Tab1 to Tab 3.

Tab 2 Structure
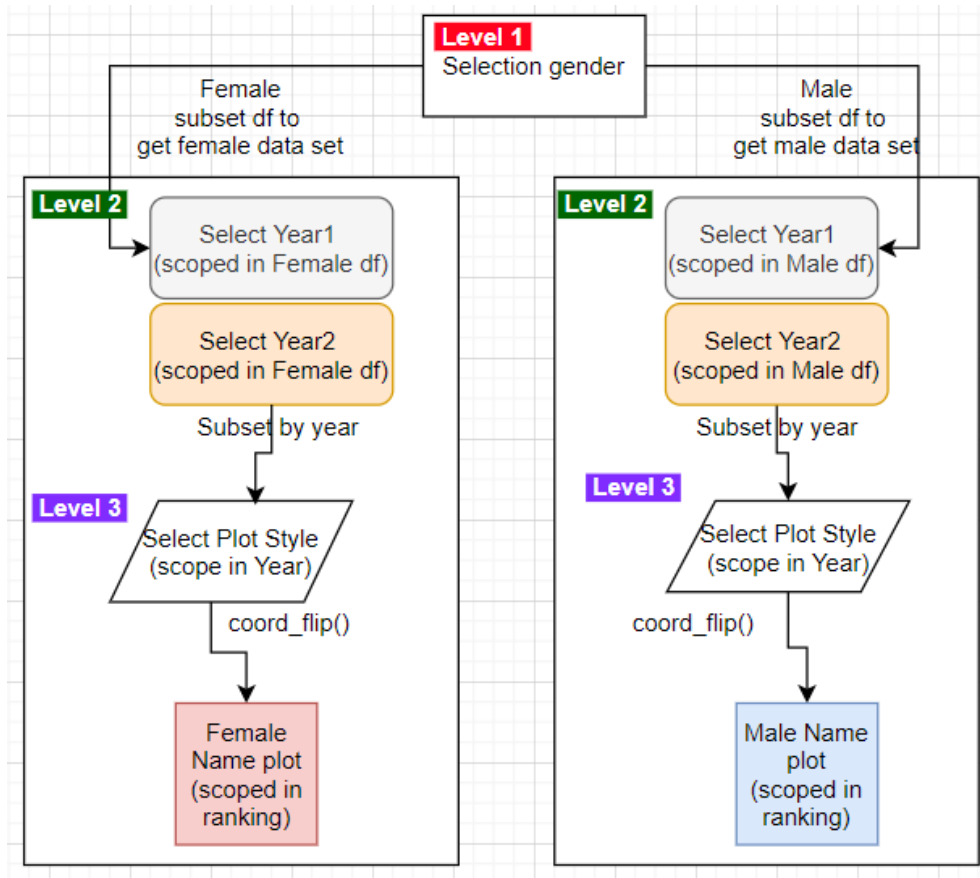
In this tab select Year has 2 value to filter the data frame. The date is subset by condition (Year = 1 and Year 2). The plot style switches the between bar plot and column plot.
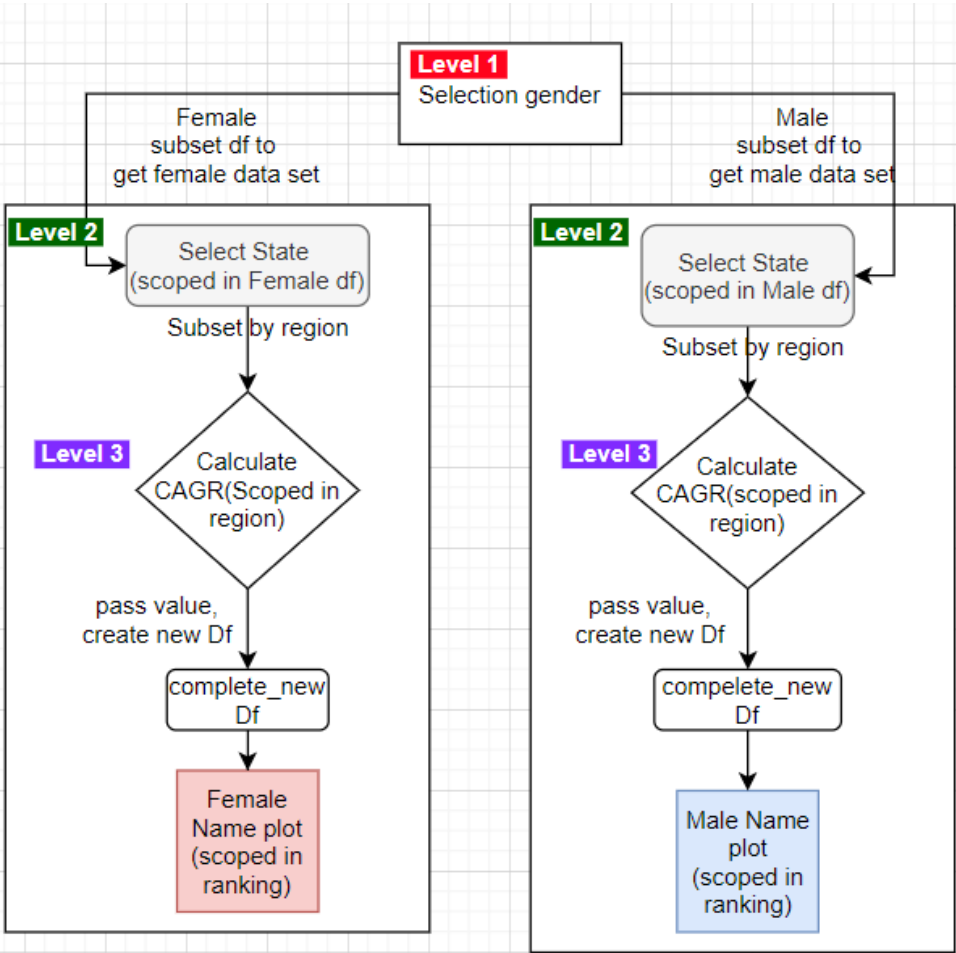
Tab 3 Structure

Process to create "complete_new_DF"

Create complete_new_DFs:

Frist, Calculate CAGR (percentage):

CAGR = (FV/IV) ^(1/N) – 1

For example, the CAGR for 2014:

1. Use the first year's count as the initial value IV (15)
2. Use the second year's count as the final value FV (20).
3. Use the duration as N (2).

CAGR = (((20/15) ^ (1/2))-1) *100 = 15.47%

Second, create new Df:

1. Get latitude, longitude, State, Name from original data frame then create a new data fame.
2. Append the CAGR to new data frame
3. Merge all new data frame to create an entirely new data frame.
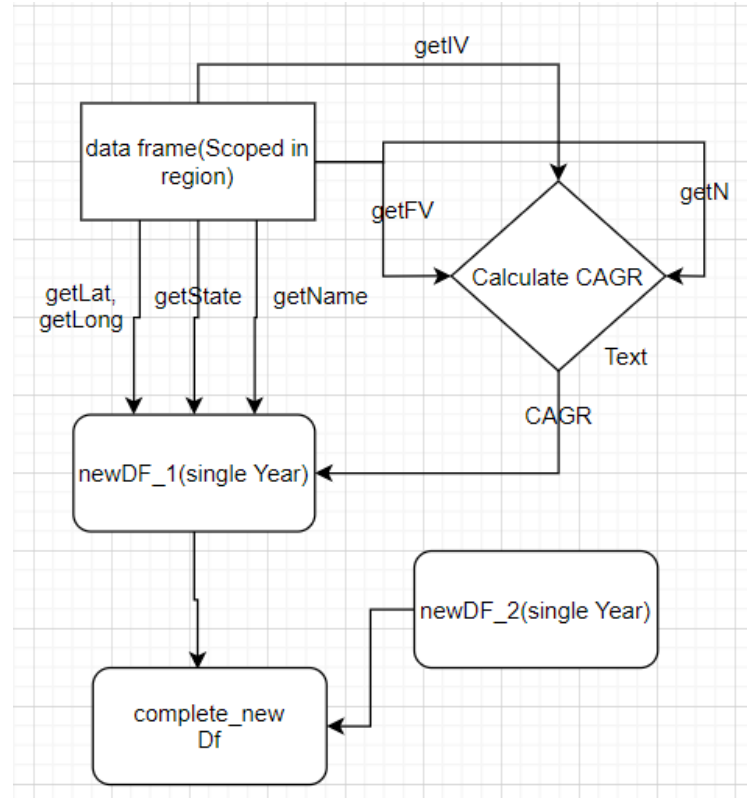
The leaflet map uses the same process as the plots to create DFs and use the new DF to generate map and marks.

In FDS, some plots use checkbox selection but why not use checkbox selection in the real product?One feature of the checkbox is that it can work without any conditions is selected, which will cause the plot not to be displayed. In the product, there is no need for multiple selection situations (Tab2 year 1 and year 2 are single selection). Therefore, the checkbox is replaced by "radioButton" to avoid unnecessary errors and to optimise the user's experience.

User guide

Tab1 "Name Ranking of Each Year":
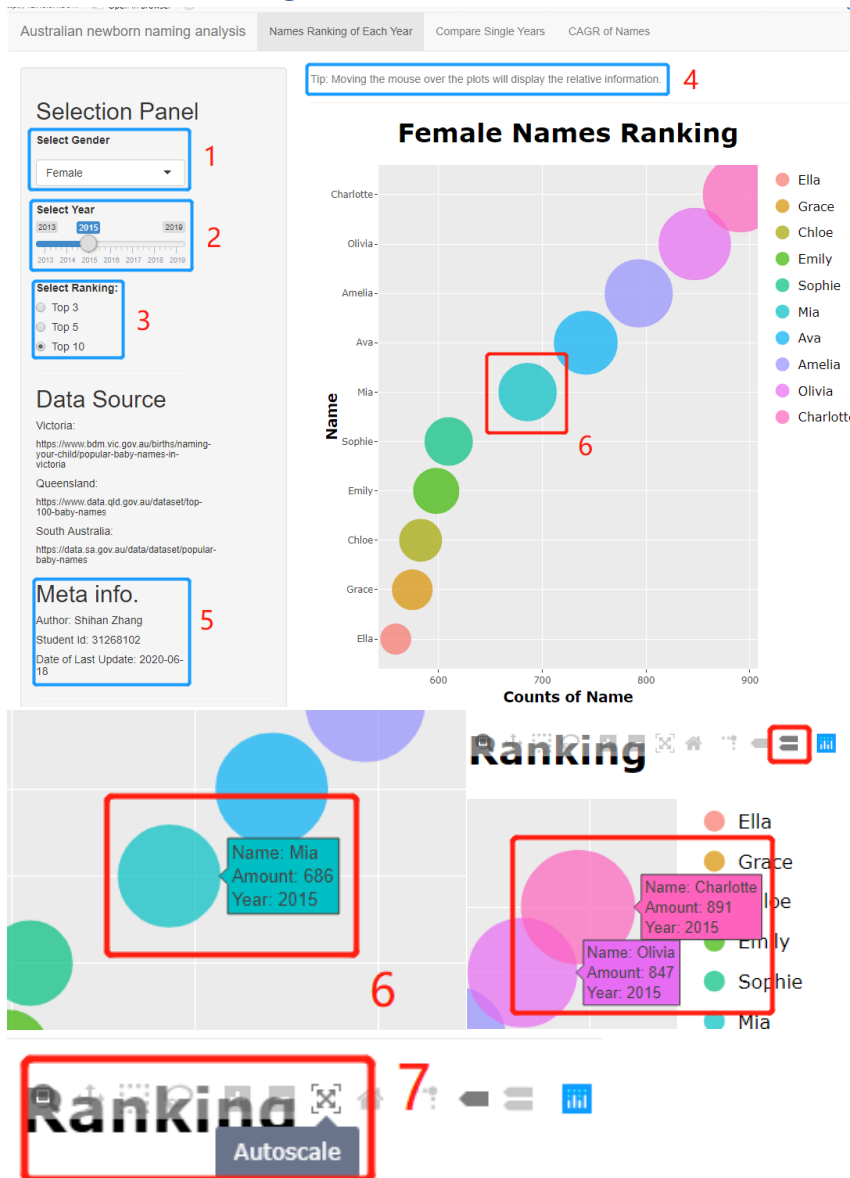
- Section 1: by clicking the drop-down box, you can select the gender to switch the plot.
- Section 2: by slide the bar, you can switch the particular Year you want to look.
- Section 3: by select the options, you can switch the plot based on ranking.
- Section 4: Tips to indicate interactive information.
- Section 5: Meta information about author and Date of the last update
- Section 6: Mover over the shape can show tooltips
- Section 7: interactive functions (Autoscale, Zoom in, Zoom out. Compare data on hover.)

Tab2 "Compare Single Years":
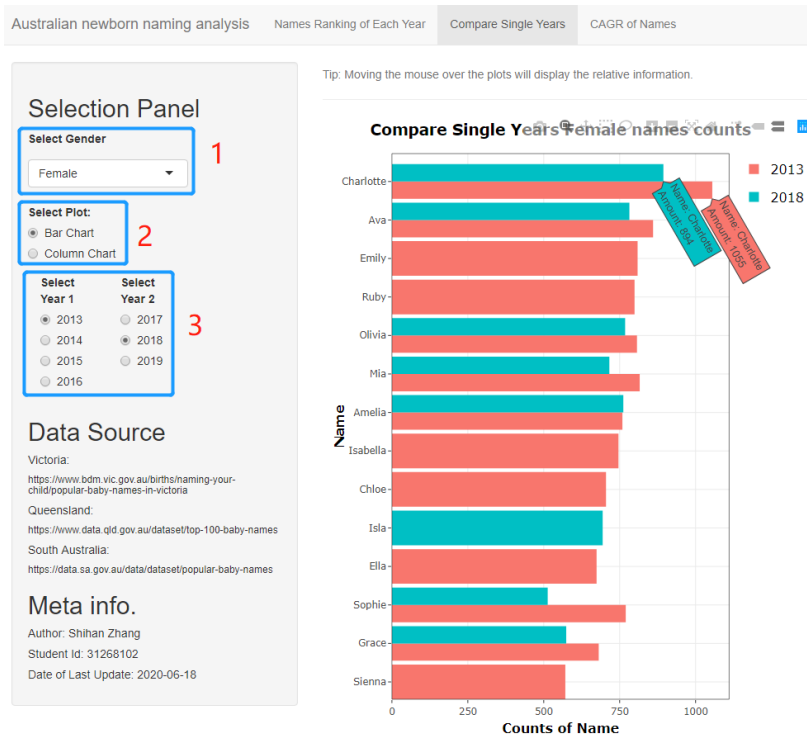
- Section 1: by clicking the drop-down box, you can select the gender to switch the plot.
- Section 2: by selecting the option to change plot style
- Section 3: by selecting the options to change the year1 and year2

Tab3 "CAGR of Names":

- Section 1: by clicking the drop-down box, you can select the gender to switch the plot.
- Section 2: by selecting the option to change region
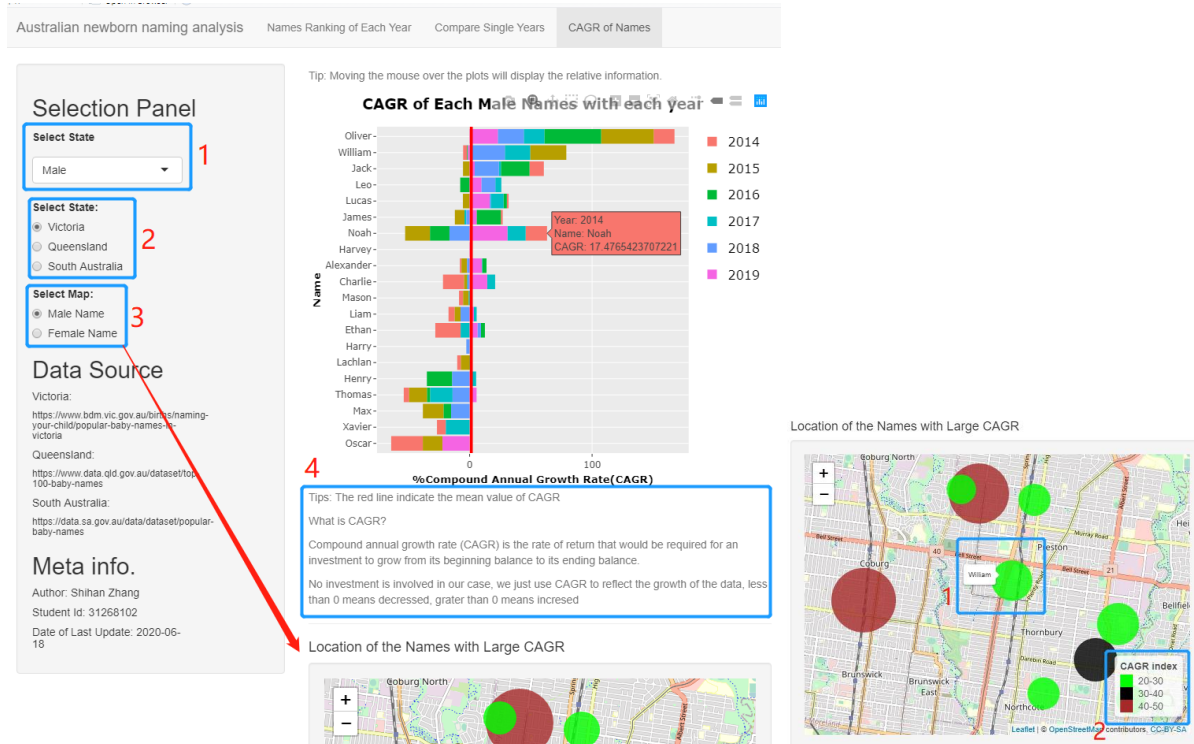- Section 3: by select the option to change the map
- Section 4: the explanation of the term "CAGR."

Map:

- Section 1: mouse over the maker will show the relative information
- Section 2: legend to indicate the colour relationship with the CAGR index

Conclusion

Thought tab1 users could see which ten names are the most used for both male and female. Bubble charts can simply and intuitively display this ranking result.

Through Tab2 users can freely compare changes in the use of names in two years. This helps users understand which names are no longer accessible and which are more prevalent under the influence of time and society.

Through Tab3 users can see the annual growth rate of the name. To understand which titles persistently hot, which names are only popular in specific years, and which names have never been used again.

Through this project, I learned that five design sheets are a beneficial and efficient design step. Using FDS allows me to think more quickly and sharply when designing drafts, which significantly improves my work efficiency and avoids the use of many unrealisable designs.

After learning about the human perceptual system, I also confirmed the importance of colour graphics and layout design in practice. Following the relevant design principles can not only speed up the reader's understanding but also improve product quality.

About what can be improved:

In this project, the use of data can be improved. The code architecture in "server.R" can be redesigned and optimised. In this project, I did not intuitively give a name that suddenly increased but chose to show the growth rate of all names. There may also be a better explanation for the exploratory method of growth rate to optimise Tab 3 results.

Bibliography

1. MURPHY, C. (2020). *Understanding the Compound Annual Growth Rate – CAGR*. Retrieved 19 June 2020, from https://www.investopedia.com/terms/c/cagr.asp

Appendix

Ideas Sheet
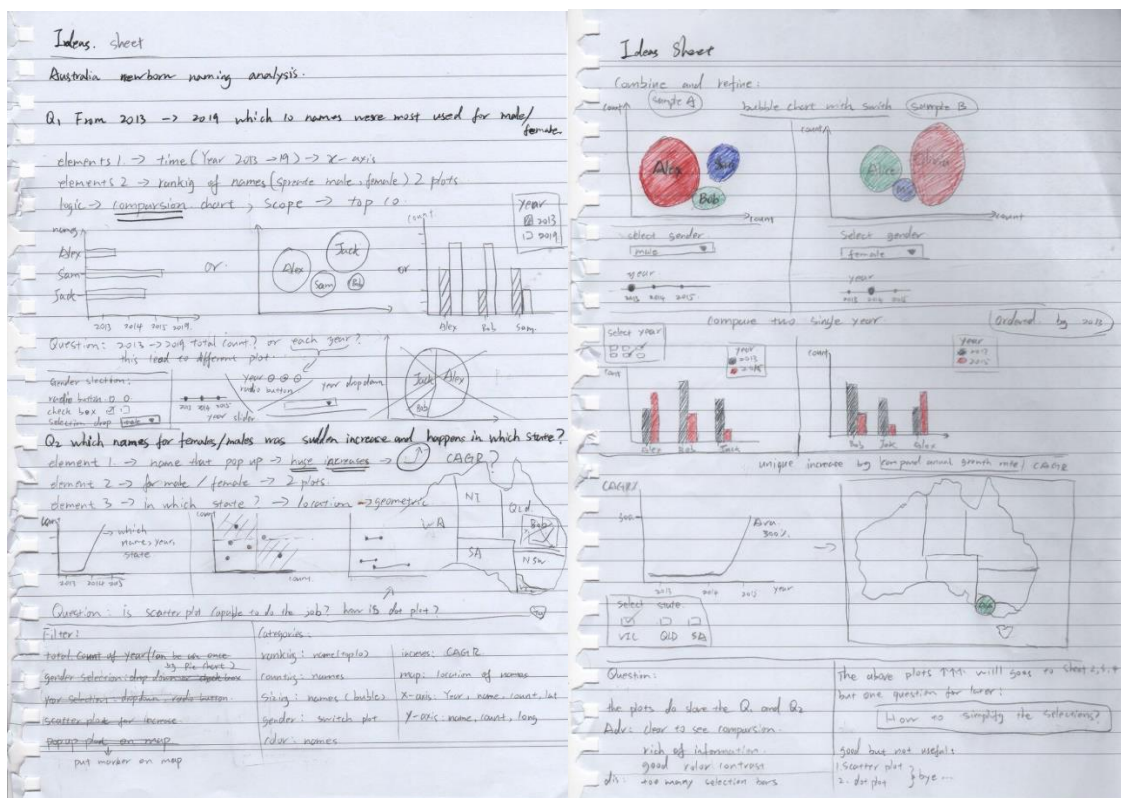
Alternative design

Sheet 2

Title: Top 10 names in 2019   Date: 2020/06/01
authors: Shihan Zhang   Sheet: 2   Task: Australian newborn naming analysis

Alternative Designs

ANNA | Each Year | Compare single Year | Sudden increases

Top 10 names in 2019

selections
Select gender: male
Select year: 2013 2015 2017 2019
Select Ranking: top 3, top 5, top 10

name: Alex, Sam, Jack

'count names'

Selections   'Top 10 female names in 2019'
Select gender: female
Select year: 2013 2015 2017 2019
Select Ranking: top 3, top 5, top 10

name: Ava, Mia, Olivia, Zoe, Lucy

'count names'

Action:
1. Select gender → render plot (subset to male)
2. Select Year → render plot (subset to selection year)
3. Select Ranking Top3 → render plot (subset to Top 3 counts)

Discussion:
Adv: rich information. Clear comparison. good color contrast.
Dis: only one plot the top page. (large gap)

Sheet 3

Title: name has sudden increases   Date: 2020/06/01
authors: Shihan Zhang   Sheet: 3   Task: Australian newborn naming analysis

ANNA | Each Year | Compare single Year | Sudden increases

'name has sudden increases'

selections
select state:
VIC, SA, QLD
select gender: male or female

250%
130%

name: Henry, Harper

Year: 2008 2010 2013 2016 2017 2018 2019

Zoom In to Show VIC

Circle radius = Percentage value from the above map

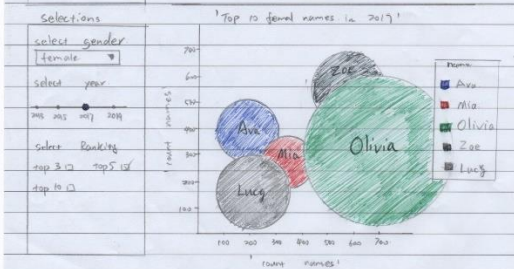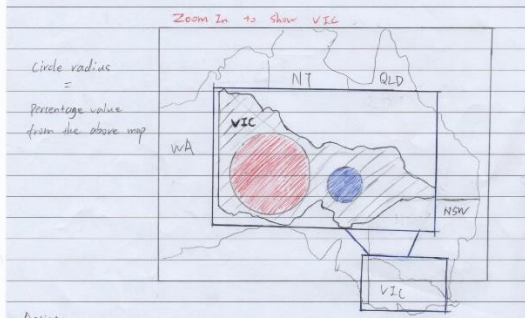NT, QLD, VIC, WA, NSW, VIC
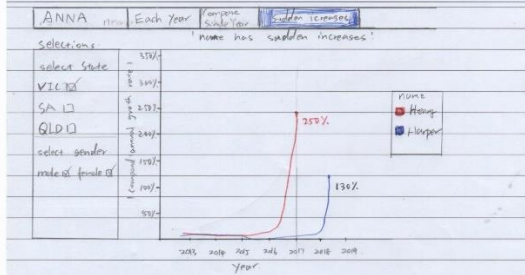
Action:
1. Select State → render plot (subset to selected state → VIC)
2. select gender → render plot (subset to female name, or include all gender)
3. add circle Marker on leaflet map (name counts as the circle radius)
Adv: point out the name on map.
Dis: inefficient use of map (single marker)

Sheet 4

Title: Comparison of two single Year   Date: 2020/06/01
authors: Shihan Zhang   Sheet: 4   Task: Australian newborn naming analysis

ANNA | Each Year | Single Year | Sudden increases

'Comparison of two single Year'

selection
Select plot: column
Select gender: male / female
Select year: 2013 2014 2015 2016 2017 2018 2019

Year: 2013, 2018

Alex Sam Jack Bob Henry Max Lucas Levi Charlie Jacob
'names'

'comparison of two single year'

selection
Select plot: bar
Select gender: male / female
Select year: 2013 2014 2015 2016 2017 2018 2019

Count: 420

Year: 2013, 2018

Zoe Emily Lucy Ella Ruby Mia Grace Evie Lily Ivy
'count names'

Action:
1. Select plot style → render plot (column plot / bar plot)
2. select gender → render plot (subset to female)
3. Select year → render plot (subset to two different year)

Adv: rich information
Dis: unordered bar / column

Realisation

Shiny APP   ⬅ current tab

Australian newborn naming analysis | Top 10 names Ranking | Compare single years | Sudden increase

Selections

Select Gender(s):
☑ Male
☐ Female

Select Ranking
☑ Top 3
☐ Top 5
☐ Top 10

Select Year
2013 2015 2017 2019

2019 male name Ranking

name
🟩 Oliver
🟦 william
🟥 Jack

Oliver
Jack
William

count of name
700
600
500
400
300
200
100

100 200 300 400 500 600 700
'count of name'

male name Ranking (6 Years total counts).

name
🟩 Oliver
🟦 william
🟥 Jack

Oliver
Jack
William

count of name
700
600
500
400
300
200
100

100 200 300 400 500 600 700
'count of name'

Section A   change detail by selection:
1. select Ranking
2. select gender · 3. select year

Section B   change detail by selection.
1. select Ranking
2. select gender.

Page configuration:
3 tab pages. with fluid page.
Selection panel (4 column → 2).
Ranking plots (column → 8)
Map plots (column → 6)
Software requirement: R studio.
Estimate: producing need 40 hrs.
Operating System: windows
                   Linux.
                   OSX.
                   Android.
                   IOS.

Ranking plots:
ggplot (df, aes(count-male, count-male))
  + geom-count (color = name_male)

map plots
leaflet (df) %>%
  add Tiles ( ) %>%
  add Circlemarkers(
      ~ lng,
      ~ lat,
      radius = ~ counts-male, color = name)

compound annual growth rate (CAGR):
Power_Start = count_2013
count-finish = count_2019
functional- year = 6,
type = 'geometric',
CAGR = (count_2013, count_2019, 6, type = "geometric")