Visualizing Urban Travel Behavior in VIC

1. Introduction

The aim of this visualization project is to explore urban travel patterns in Victoria, with a focus on household characteristics, such as household income, household size and total number of vehicles each household owns, and their relationships with different travel modes. By analyzing the travel behavior of individuals across various income groups, time periods, and household sizes, the visualization aims to convey insights into how these factors influence the choice of travel modes.

In the project, there are two tabs, which enables users switch different tabs to see different dimensions. The interactive elements within the application allow users to filter data by travel mode, income group, and year to observe the shifts in travel behavior over time. The stacked bar chart in the "Overview" tab provides a temporal analysis, illustrating the variation in travel modes during different time periods of the day on weekdays and weekends. The "Household" tab explores deeper into the household composition, using a ridge plot to show the distribution of income, household size, and vehicle ownership for each travel mode. Additionally, you can select the sankey diagram to show how household characteristics relate to travel modes.

This narrative visualization is intended for urban planners and policy makers, who seek to understand the dynamics of urban travel behavior in relation to social-economic factors. Additionally, it also serves as a tool for researchers and people who is interested in studying the impact of household characteristics on travel patterns, particularly in the context of improving urban mobility. The findings aim to highlight potential areas for optimizing public transport services, road usage, and promoting sustainable travel behaviors.

Through clear visualizations and interactive exploration, the intended message is to emphasize the diverse factors that drive travel decisions in urban settings, and how understanding these patterns can lead to more effective transportation policies.

2. Design Process

The design process is based on the “Five Sheet Design” methodology, including three parts and five sheets. Three parts are consist of “Brainstorm(Sheet 1)”, “Initial Designs(Sheet 2-4)” and “Realization(Sheet 5)”. Each stage of development follows the principle of data visualization, user needs and effective communication. In addition, my decisions on visualizations are also based on Munzner’s What-Why-How framework, principles of human visual perception, and narrative design approaches, as well as practical considerations for accessibility and usability.

According to Munzner's theory, the design process begins by answering three basic questions: What data are we visualizing, Why is it being visualized, and How will it be represented with interactions?[1]

The data for this project was downloaded from VISTA from 2012 to 2020, including multiple variables such as household income, household size, travel mode and numbers of vehicles each family owns after data wrangling and cleaning. These variables are abstracted into ordinal and categorical data. Therefore, my ideas in Sheet1 are all based on the ordinal and categorical data I have.

The goal of the narrative visualization is to allow readers identify the most prevalent modes of transport, particularly for different income groups, and have an overview on whether time periods will effect people's choices of travel in order to select appropriate travel modes in specific times. What's more, the visualizations tend to allow readers explore the relationship between household demographics and travel choices. The project focuses on both temporal trend analysis(e.g. changes in travel pattern in different time periods over year) and comparison(e.g. between different income groups).

Moving to how to represent the data interactively, it is a brain storm. There are many feasible ways to implement. Therefore, I can use Five Design Sheet Methodology to process my design. There are five sheets in total, sheet1 contains my ideas to display the visualizations and filter out plots which is not helpful or duplicate. In addition, I need to categorize the rest plots and combine some of them together and put them into different tabs for example. Sheet2 to Sheet4 are different designs, containing totally different plots with interactive gadgets. Plus, there are some pros and cons in the designs I need to keep and fix. Sheet 5 is my final implementation with some details. Next, I will explain the details of each sheet.

In sheet1, I have put lots of ideas of visualization, including bar plots, line plots, box plots and so on. But some of them are similar, for example, polar charts and pie charts are similar to stack bar charts, while stack bar charts convey more information to readers than pie charts. Compared to polar plots, stacked bar plots use key visual variables effectively: length, position and color. While in polar plots, the visual variable of angle is introduced, which is less accurate than length or position. According to Gestalt principles, human brains tend to recognize patterns and make sense of complexity through grouping, so it is important to categorize my ideas. After that, to improve user experience and enhance storytelling, I combine plots and refine them into different tabs. Tabs minimize cognitive load by chunking content into pieces according to an article on NNGroup[3]. Additionally, I have added a new diagram which is Sankey diagram in my idea because after seeing peers' presentations, I think a sankey diagram is useful in my project to emphasize movement/change/choice from one source to another[4].

In sheet2, I have a stack bar chart in tab1, and a bubble chart in tab2. I decided to use different color hues to represent each kind of travel pattern(e.g. red represents vehicle) because the human visual system is sensitive to color contrast and can quickly identify different color hues. Additionally, color hue is used as a categorical variable to represent distinct travel modes, making it easier for readers to identify and differentiate between categories. About the color palette I use(see Figure1), I use the colors that are color blind friendly.

Figure1: The Color Palette used for Stack Bar Chart

According to the article, colorblind-friendly palettes usually uses two main colors like blue and red, along with shades from these two hues. Examples include blue/orange, blue/red, or blue/brown[5]. For interaction used in stack bar charts, I decided to use a year filter that readers can select the particular year they want, and a text box to show details when you hover on the bins because the average person can only keep 7(plus or minus 2) items in their working memory according to Miller's Law[6]. Moving to tab2, I designed a bubble chart to show relationship between household income, household size, number of vehicles and travel modes each household chose. A bubble chart is an extension of the scatter plot to look at the relationship between three numeric variables[7]. I use different size to distinguish different bubbles and a filter to select year to show. There are pros and cons in both plots, the visual variable can effectively display the categorical data and relationships between variables, which is good. However, the bins may overlap in stack bar charts because the magnitude of each pattern differs.

In sheet3, I changed both plots in both tabs, which is different from the one in my presentation because according to the Five Sheet Design Methodology, sheet2 to sheet4 should be completely designs to tell a story through visualization. However, my sheet3 and sheet4 are only the improvements of the previous sheet, which actually does not follow the methodology. In tab1, I replaced stack bar charts with a line plot. Line plots are good to show temporal trend. But the line plot, in my opinion, is a little bit simple because it cannot show many dimensions and information to the readers. And the interactions are also limit. In tab2,

I use a ridge plot to show different distributions under various household backgrounds instead of bubble plots. However, the density lines may overlap, so I need to figure a way to separate them.

In sheet4, I also make multiple changes to distinguish the design completely. In tab1, I tried to use an area chart because it typically allow us to visualize how a measure, observed through multiple category values, changes over time[8]. It is a good plot because it shows both temporal changes and differences between categorical variables. But there are some problems, if the temporal trends are not obvious, it is not decent to use a stack area chart. In tab2, I decided to use a sankey diagram. I defined income groups, household size groups, vehicles ownership groups and travel patterns as different sources, and the flows between them perfectly illustrate the relationships. It helps locate the most important contributions to a flow[4]. But due to the variation in magnitude, some source bins with little amounts may not be seen clearly, which is a problem I need to fix.

Moving to sheet5, coming to the final design, it must be a complete layout and may be different from the original 3 sheets, or a combination. Therefore, I changed it a little from my presentation, and gave more design details.

In the final design sheet, I tend to have two tabs and three plots totally. There is a stack bar chart in tab1 called “Overview” and two plots, a ridge plot and a sankey diagram in tab2 called “Household”. As you can see, finally I decided to use a stack bar chart rather than a stack area chart, and combine two different graphs in one tab.

In the “Overview” tab, rather than line graph and stack area chart, I chose the stack bar chart because it conveys more information to readers such as total counts and the proportion each travel mode contributes to along with the temporal change. According to Gestalt Principle of Continuity and Cleveland and McGill’s Graphical Perception Theory, stack area charts are particularly suitable for showing changes over continuous variables such as time, while stack bar charts are ideal for comparing discrete variables because position and length are some of the most accurately perceived visual variables. For the task, it focuses more about the category(travel patterns), the total count and the proportion of it rather than temporal change, although time is another aspect. I can create a year slider, which enables readers to select the year they want and slide to see the change across years.

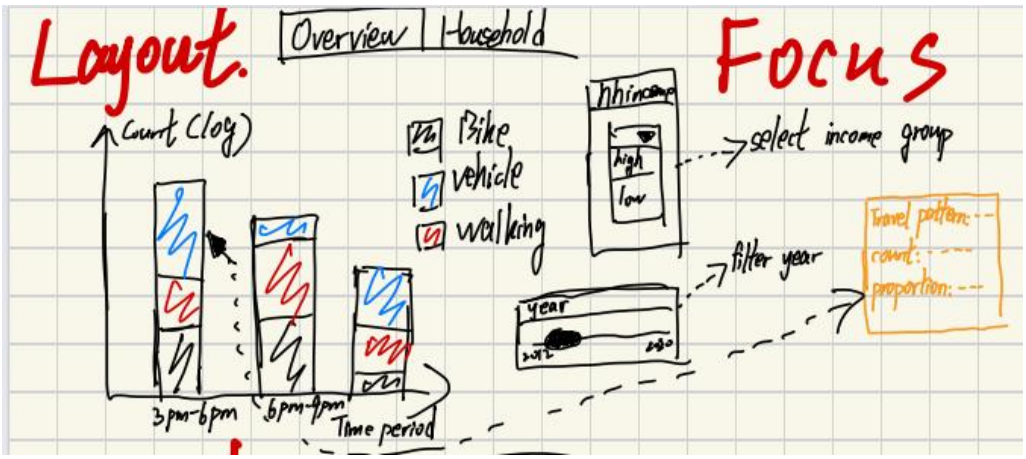


Figure 2: Layout of the "Overview" tab

Through Figure2, you can see that, I use different color hues as the main visual variable to represent each travel mode. The color palette is color-blind friendly(Figure1). The narrative visualization genre in this tab is mainly the annotated chart with a text box showing details of the visualized data. According to Miller's Law[6], the information that people can remember is limited, so the checkbox enables readers check the information of each bin. Additionally, to solve the visual problem caused by the variation in magnitude, I decide to use logarithm to narrow down the gap so that each travel mode could be seen clearly.

In the "Household" tab, there are many items and elements in it. (See through Figure3) First, readers are able to select the specific plot they want(ridge plot or sankey diagram). To keep the consistency of the design(e.g. in tab1, I only have one plot and some interactive gadgets) and enable users to avoid information overload, I design a tickbox, enabling users to interact only with the plot they are interested in.

Let's focus on the ridge plot, I add a "Household size selection", allowing readers to select the necessary information in case to avoid the overlapping problem mentioned before. The mode filter enables readers to explore the relationship between three household characteristic and each choice of travel.

Plus, I refined the year filter with adding a action button. In my opinion, the narrative visualization genre of animation not only shows the temporal change but also makes the plot more vivid and better for storytelling.

The typography of each tab is designed to be similar to ensure the consistency of the project. The layouts are supposed to be one plot and some interactive gadgets next to the plot, with a ordered arrangement.

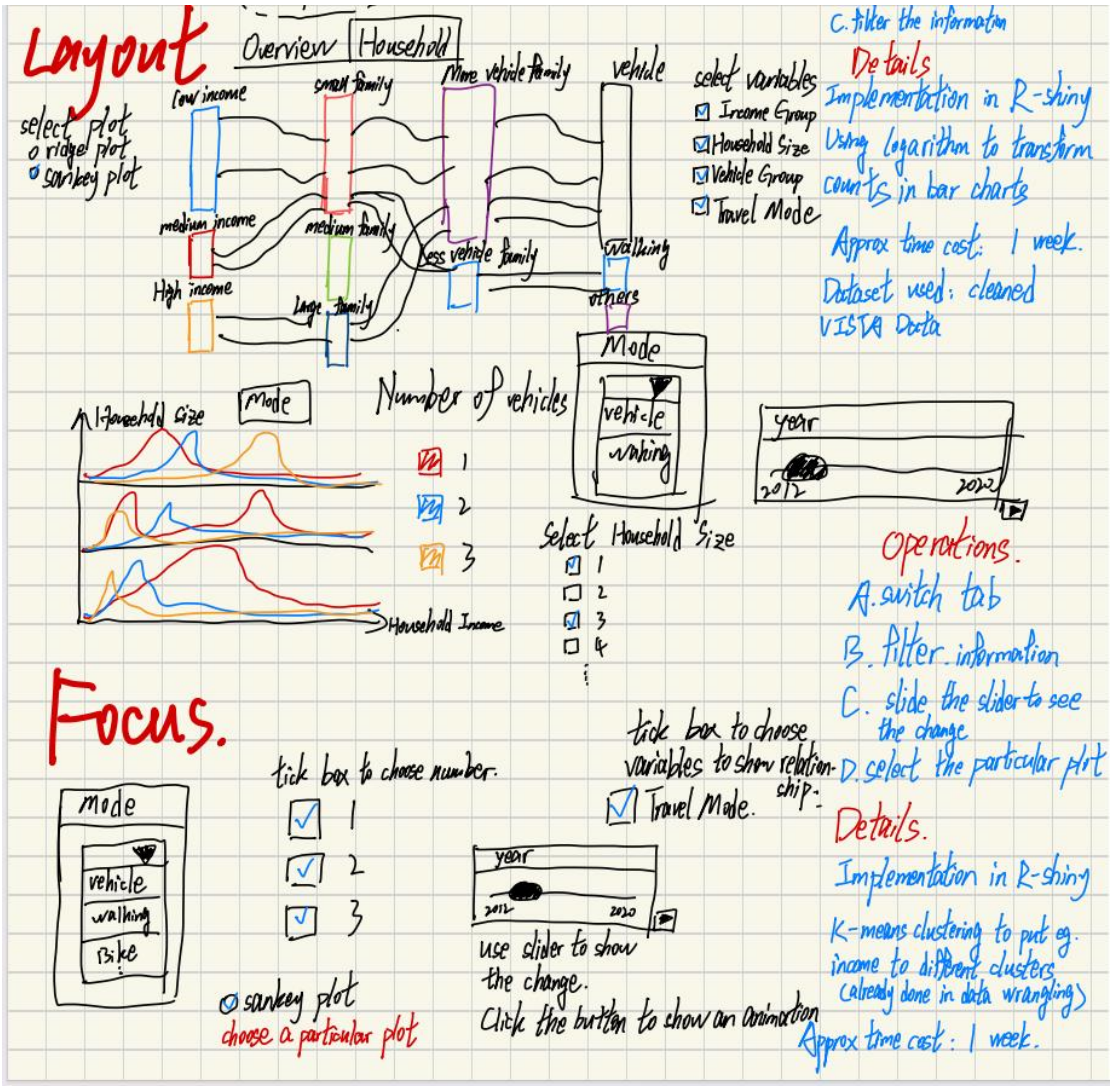


Figure 3: Layout of the "Household" tab

3. Implementation

3.1 Technical Implementation

In this visualization project, I mainly use R-shiny[9] to build the interactive application, along with the R-package such as "ggplot2"[10] for creating static plots, and "plotly"[11], which enables us to create an interactive plot, particularly for the sankey diagram. "ggridges"[12] is used to create ridge plots, which is very helpful to visualize the distribution of data across multiple dimensions. The dataset used for this project was downloaded from VISTA, which required significant data wrangling and cleaning. These have been already done in the DEP part, including K-means clustering to mutate new variables like "income group", which is a necessary advanced data wrangling step.

To make the layout look organized and clear, in addition, to create multiple tabs and other interactive gadgets needed in the dashboard, some external templates and codes are used in the project, such as resources from Shiny Gallery examples[13][14][15][16][17][18]. For

example, I have searched for many articles and codes to find out how to do an animation to my plots in R-shiny.

Furthermore, several adjustments were taken in the final implementation. Therefore, it may be a little bit different from the final design(sheet 5).

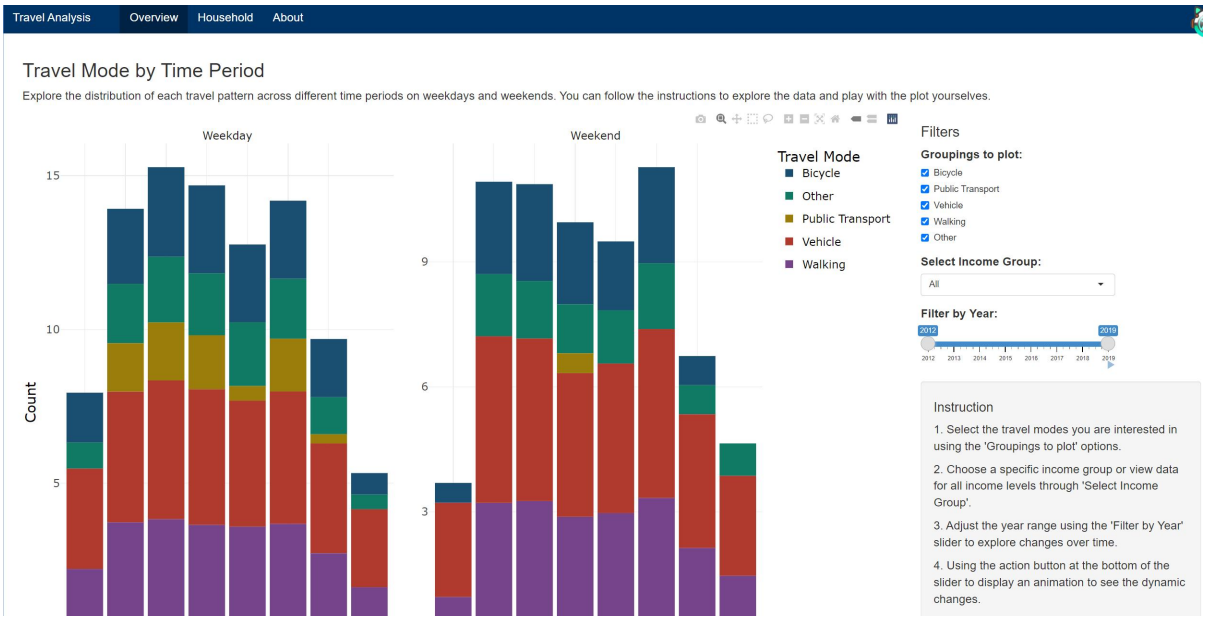


Figure 4: Part of the final implementation

Figure 4 displays the final implementation I will make. As you can see, I have added a new tab called “About”, which contains some important information and details about this dashboard, including what data I use, features and technologies I use, etc. I reckon it is important for readers know the background of the data and the application, plus, how to use the application through interactions. I want my application to be mainly reader-driven, which means readers can freely choose the information they want to know, they can use different interactive methods to explore the data. For example, they can choose a specific income group to see the distribution of each travel pattern across time periods. What’s more, I added a instruction box to teach users how to interact with the plots.

And for the stack bar chart, I have added another filter called “Groupings to plot”, enabling readers to select the combination they want to see. For example, maybe someone would like to see the distribution of “Vehicle” and “Bike”, they could tick these two modes and drop others. It is also a way to solve the overlapping problem(the bins may overlap so that some of them may not clear to see) I mentioned before.

The project might be challenging because the realization of animation could cause several errors in coding. In addition, the Sankey diagram is particularly challenging to implement as it required careful data restructuring and choice to ensure proper flow visualization. Another challenge is the checkbox to select a specific plot in the “Household” tab. I have found that the programming logic behind it is much more complex than I thought. If the logic goes wrong, it will return to an error. Overall, the implementation focuses on balancing complexity with user-friendly interaction to convey the necessary insights effectively.

3.2 Interactive Narrative Visualization Implementation

The final visualization implementation consists of two main interactive components, which is the “Overview” tab and the “Household” tab. Each tab presents different visual insights based on the VISTA dataset, allowing readers to explore the data in an interactive way on their own. Plus, readers can firstly take a look at the “About” tab to have a basic understanding of this application.

The “Overview” tab provides readers with a stack bar chart(see Figure 5) that displays the distribution of travel modes over different time periods on weekdays and weekends across different years. Ideally, it could answer the question “Do people choose different ways to travel depending on when they want to?” Users are allowed to filter the chart by selecting different travel modes, income groups and specific year ranges. Additionally, at the right-bottom of the year filter, there is an action button, which enables readers to observe dynamic changes in the distribution over time.

If users do not know how to use these interactive ways to explore the plot, or they are familiar with the interactive gadgets, they can have a look at the instruction box and follow it step by step. It is a tutorial to teach you what these gadgets are and how to interact with them.

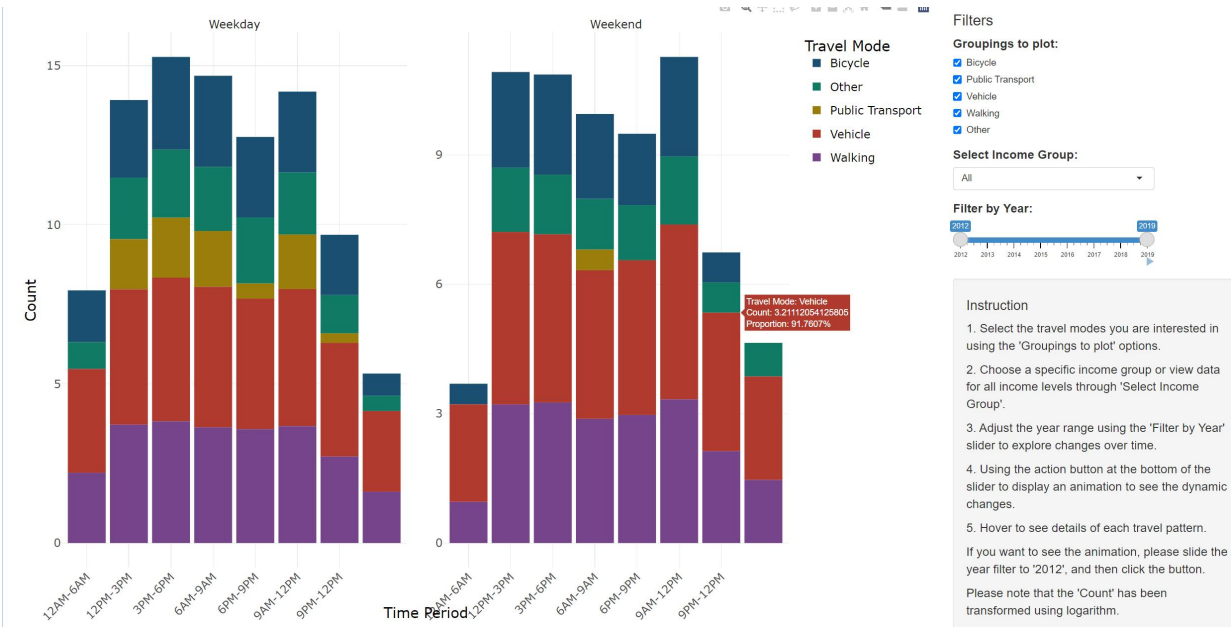


Figure 5: Final Implementation in the “Overview” tab

The interactive features included in the plot are travel mode selection, household selection, year filter and a text box showing details(see Figure 5). Travel mode selection enables readers select one or more travel modes(Bike, vehicle, walking, etc.) to visualize specific distributions. The drop-down menu allows users to filter the data by different income groups, focusing on how income affects travel behavior. The slider lets users filter the data by year, from 2012 to 2020, either in a manual way or through the action button, which will show you an animation when you click on it. It provides insights on how travel patterns evolve across years. Moreover, when users hover their mouse on bins, the chart will show a checkbox to see details like counts, travel mode and proportion.

The “Household” tab includes two visualizations, which is a ridge plot and a sankey diagram. Users can shift plots using the radio button. These visualizations help readers explore relationships between household characteristics(household income, household size, vehicle ownership) and travel pattern choices. Ideally, it will illustrate the question “How do travel patterns vary in different household situations?”

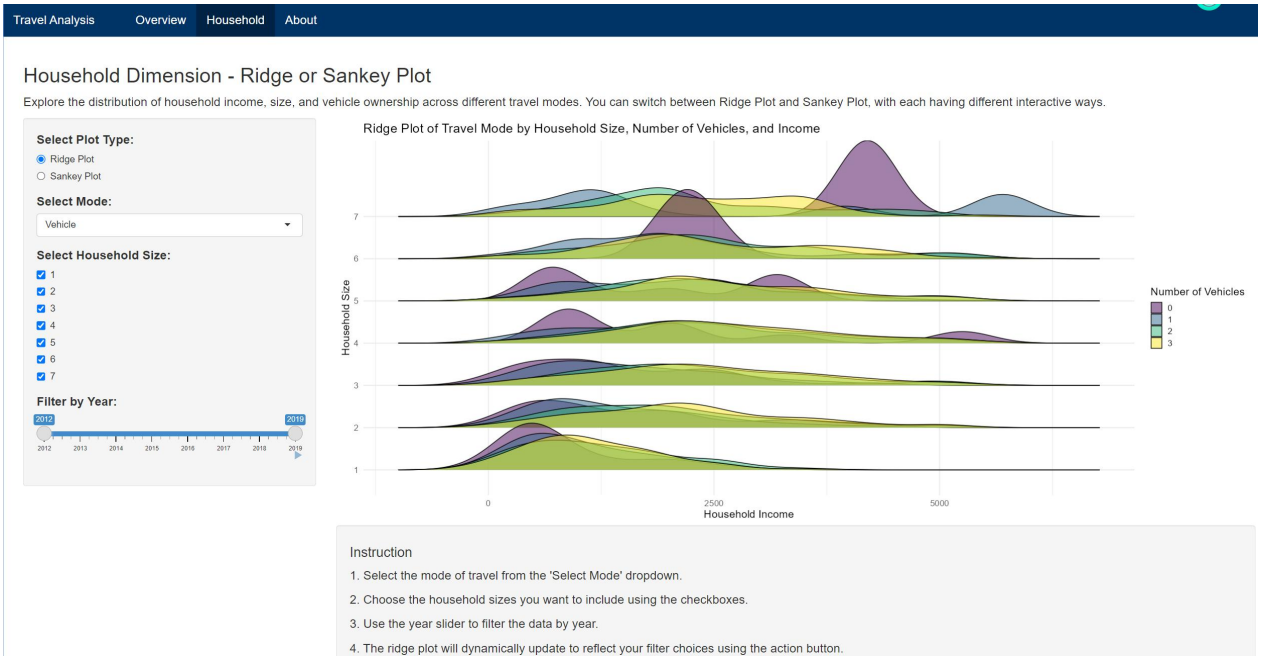


Figure 6: Ridge Plot in the “Household” tab

The ridge plot(Figure 6) visualizes the distribution of household income and size, with an layer showing the numbers of vehicles each household owns, presented through different colors. Users can select the specific mode using mode selection. Plus, users are able to select particular household size to display by ticking the checkbox. Besides, users can see the temporal change using the year filter either manually slide it or through an animation.

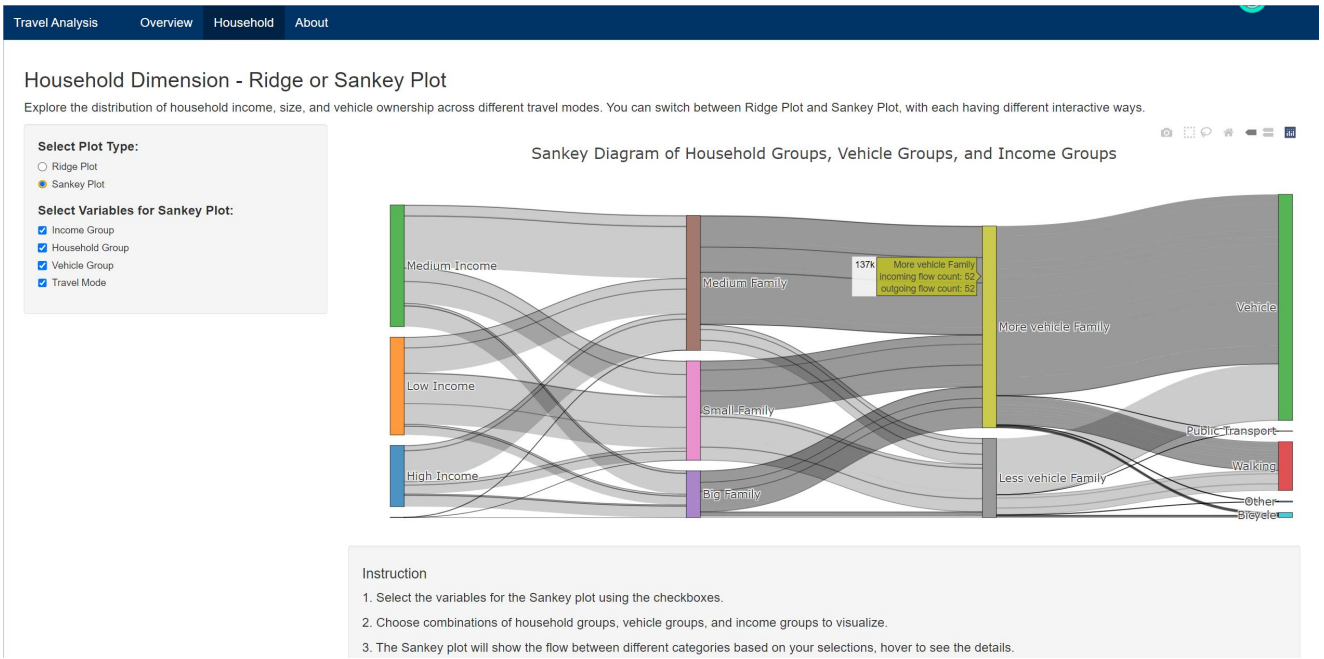


Figure 7: Sankey Diagram in the “Household” tab

The sankey diagram visualizes the flow of income groups, vehicle ownership groups, household size groups and travel modes.By selecting different variables using variable selection, users can see how these categories relate with each other. Additionally, you can hover your mouse on a specific source like “vehicle”, or hover it on a specific flow, either way will show you the details through hover tool tips including source name, incoming flow count and outgoing flow count, making it easy to understand the broader patterns in the data.

Overall, the narrative structure of this visualization application is mainly reader-driven, offering readers the flexibility to explore the data and interact with the plots at their own pace. Through instruction box in each tab and the “About” tab(Figure 8), readers will have a basic understanding the interaction methods and the background of the dataset and the application.

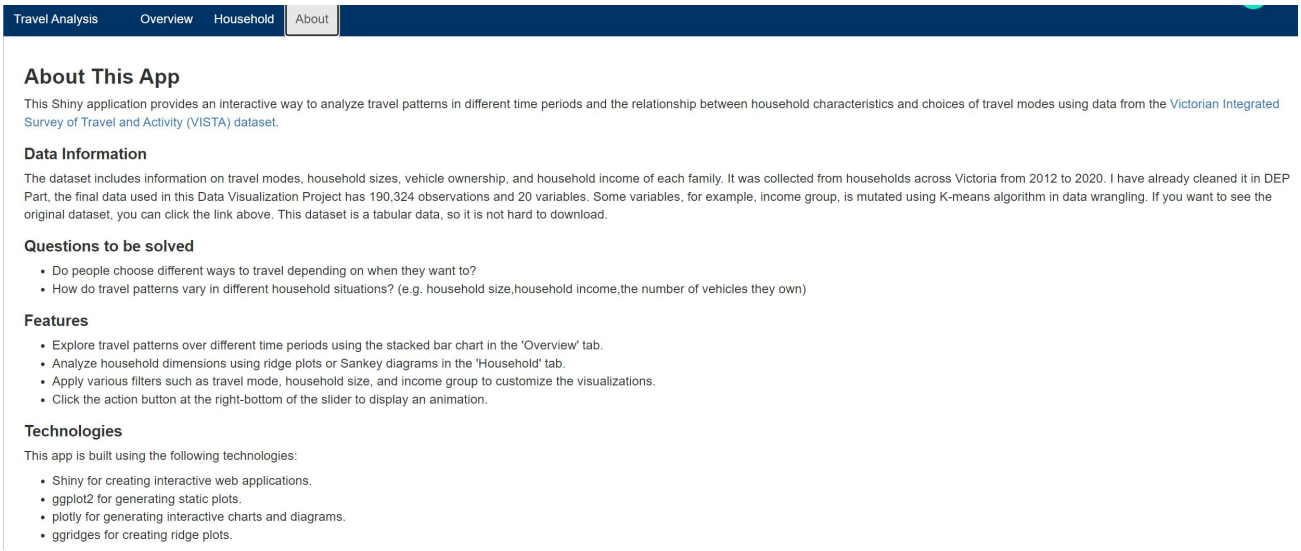


Figure 8 : The “About” tab

The “Overview” tab introduces a high-level comparison of travel mode choices across time periods in different years, allowing users to see general trends. When users switch to the “Household” tab, the focus shifts to more detailed, household-level insights, providing a deeper understanding of how household characteristics influence travel behaviors.

3.3 Using the Implementation

The application is in RStudio, it can be run locally after installing all the packages mentioned in the “Load Libraries” part. You can open the R script and click the button “Run App” to launch the application.

This visualization application is a dashboard with three main tabs, which is the “Overview” tab, “Household” tab, and “About” tab. These can be accessed from the navigation bar at the top of the interface. If you are not familiar with the application, I strongly recommend you to take a look at the “About” tab first, it contains the background and some main features of this app.

After running the app, it will jump to the “Overview” tab first. The plot is on the left side and some interactive gadgets listed on the right side. Users can select which travel modes to include by ticking or unticking boxes. This allows readers to customize the bar plots based on their interest.

Readers can filter the data by selecting an income group using the drop down menu. It will allow you to compare travel patterns between different income groups. The slider below is the year filter, it enables users adjust the year range to display manually. The action button at the bottom right of the slider offers you an easy way to see the temporal change through an animation. What’s more, hovering over the bar segments reveals more details, including total count(log transformed) and proportion.

Moving to the “Household” tab, it will show the ridge plot by default. Users can select the particular plot type by using the radio buttons in “Select Plot Type” selection. It will enable readers shift between the ridge plot and the sankey diagram.

In the ridge plot, readers can select a particular travel mode from the “Select Mode” drop down menu. In addition, readers are allowed to select which household size to show by ticking the boxes in “Select Household Size”. Similar to the year filter in the “Overview” tab, this one in the ridge plot also enables readers to see the temporal trends and changes through a manual or an animated way.

In the sankey diagram, readers can select variables to show different flows and relationships by ticking the checkbox in “Select Variables for Sankey Plot” section. If you want to see some detail information, hover over the source bar or each flow.

Readers may miss some key interactions. For example, the hover tooltips, readers may not immediately realize that hovering reveals additional details such as proportion, total counts. Besides, the action button in the year filter, due to its small size, readers might not notice that. The year slider in both “Overview” and “Household” tabs contains an action button that animates changes over time. Click it, it will show an animation, click it again, it will pause, then click it again, it will continue showing the animation. Please mention that, if you want to see the animation, you should slide the button in the year filter from right to the left, which means the animation should start from year 2012. Additionally, if you are still not familiar with the interaction, you can check the steps in instruction boxes in both “Overview” and “Household” tabs.

4. Conclusion

In this project, the narrative visualization successfully discovered key insights into urban travel patterns across different time periods and household characteristics. For example, through stacked bar charts, we can see that people's choices on travel modes differ in different time periods, especially 6AM-9AM and 3PM-6PM, the travel volume is huge. Additionally, different income groups prefer different patterns of traveling. Through ridge plots, and Sankey diagrams, I was able to effectively communicate how travel behavior is influenced by variables like household income, vehicle ownership, and travel modes. The interactive features allowed users to explore the data dynamically, filtering by various standards to gain a more comprehensive understanding of the trends over time.

Throughout the process, I have learned a lot about designing reader-driven visualizations. Implementing interactivity was particularly challenging but worthy, as it added depth to the narrative and enabled users to explore the dataset on their own. In addition, I have learned how to use visual theory to design a decent visualization, also, how to use proper visual variables to make the plot informative and clear. Besides, the use of various interactive features taught me how to make interactions with the plot to make them more vivid. One aspect that could be improved is the level of detail provided in hover tooltips. Though they are informative, they could be expanded to include additional context to make the insights more intuitive for users.

I would explore additional layers of complexity, such as fitting predictive models to show potential future trends in travel behavior if I had more time. I would also consider refining the color palette or using more interactive features to make the visualizations more accessible to a wider audience and give more information to the readers. Overall, this project provided a strong foundation in narrative visualization, and I look forward to building on these insights in future work.

Reference

[1] Rocks, T. @ D. (2023, April 13). *Visualisation Analysis & Design by Tamara Munzner or the What-Why-How of data viz. | Data Viz Bookshelf | Data Viz Bookshelf*. Data Rocks NZ. https://www.datarocks.co.nz/post/data-viz-bookshelf_visualization-analysis-design-tamara-munzner.

[2] Interaction Design Foundation. (2016, August 30). *What Are Gestalt Principles?* The Interaction Design Foundation. https://www.interaction-design.org/literature/topics/gestalt-principles

[3] Experience, W. L. in R.-B. U. (n.d.). *Tabs, Used Right*. Nielsen Norman Group. https://www.nngroup.com/articles/tabs-used-right/

[4] *Sankey diagram*. (2020, September 28). Wikipedia. https://en.wikipedia.org/wiki/Sankey_diagram

[5] Cravit, R. (2019, August 21). *How to Use Color Blind Friendly Palettes to Make Your Charts Accessible - Venngage*. Venngage. https://venngage.com/blog/color-blind-friendly-palette/

[6] Yablonski, J. (2022). *Miller's Law*. Laws of UX. https://lawsofux.com/millers-law/

[7] Atlassian. (n.d.). *A Complete Guide to Bubble Charts*. Atlassian. https://www.atlassian.com/data/charts/bubble-chart-complete-guide

[8] *Stacked area charts: When to use them and when to avoid them?* (n.d.). Inforiver. https://inforiver.com/insights/stacked-area-charts-when-to-use-and-avoid/

[9] Web Application Framework for R [R package shiny version 1.6.0]. (2021, January 25). Cran.r-Project.org. https://cran.r-project.org/package=shiny

[10] Wickham, H. (2019). *Create Elegant Data Visualisations Using the Grammar of Graphics*. Tidyverse.org. https://ggplot2.tidyverse.org

[11] Sievert, C. (2020). Interactive web-based data visualization with R, plotly, and shiny. In *plotly-r.com*. Chapman and Hall. https://plotly-r.com

[12] Wilke, C. O. (2024). Ridgeline Plots in "ggplot2" [R package ggridges version 0.5.6]. *R-Project.org*. https://cran.r-project.org/package=ggridges

[13] *The Masters Data Viz*. (2014). Shinyapps.io. https://cjteeter.shinyapps.io/MastersGolf/.

[14]*Shiny - Using Action Buttons*. (2015, March 26). Shiny.posit.co. https://shiny.posit.co/r/articles/build/action-buttons/

[15]cjteeter. (2017). *ShinyTeeter/3_MastersGolf at master · cjteeter/ShinyTeeter*. GitHub. https://github.com/cjteeter/ShinyTeeter/tree/master/3_MastersGolf

[16]Tang, T. (2019, May 31). *Shiny - Animated NYC metro traffic*. Shiny. https://shiny.posit.co/r/gallery/government-public-sector/nyc-metro/

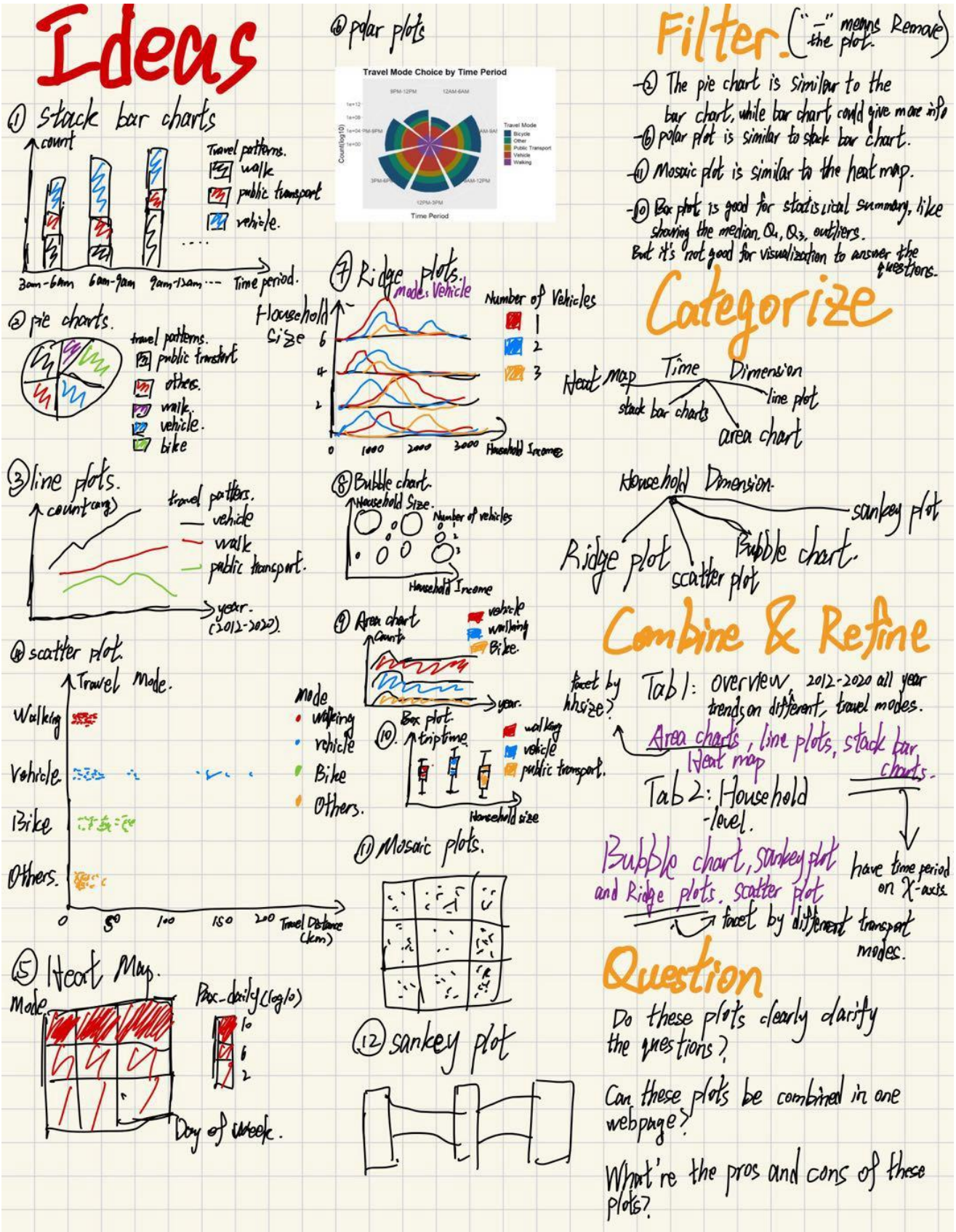
[17]rstudio. (2020). *shiny-gallery/nyc-metro-vis at master · rstudio/shiny-gallery*. GitHub. https://github.com/rstudio/shiny-gallery/tree/master/nyc-metro-vis

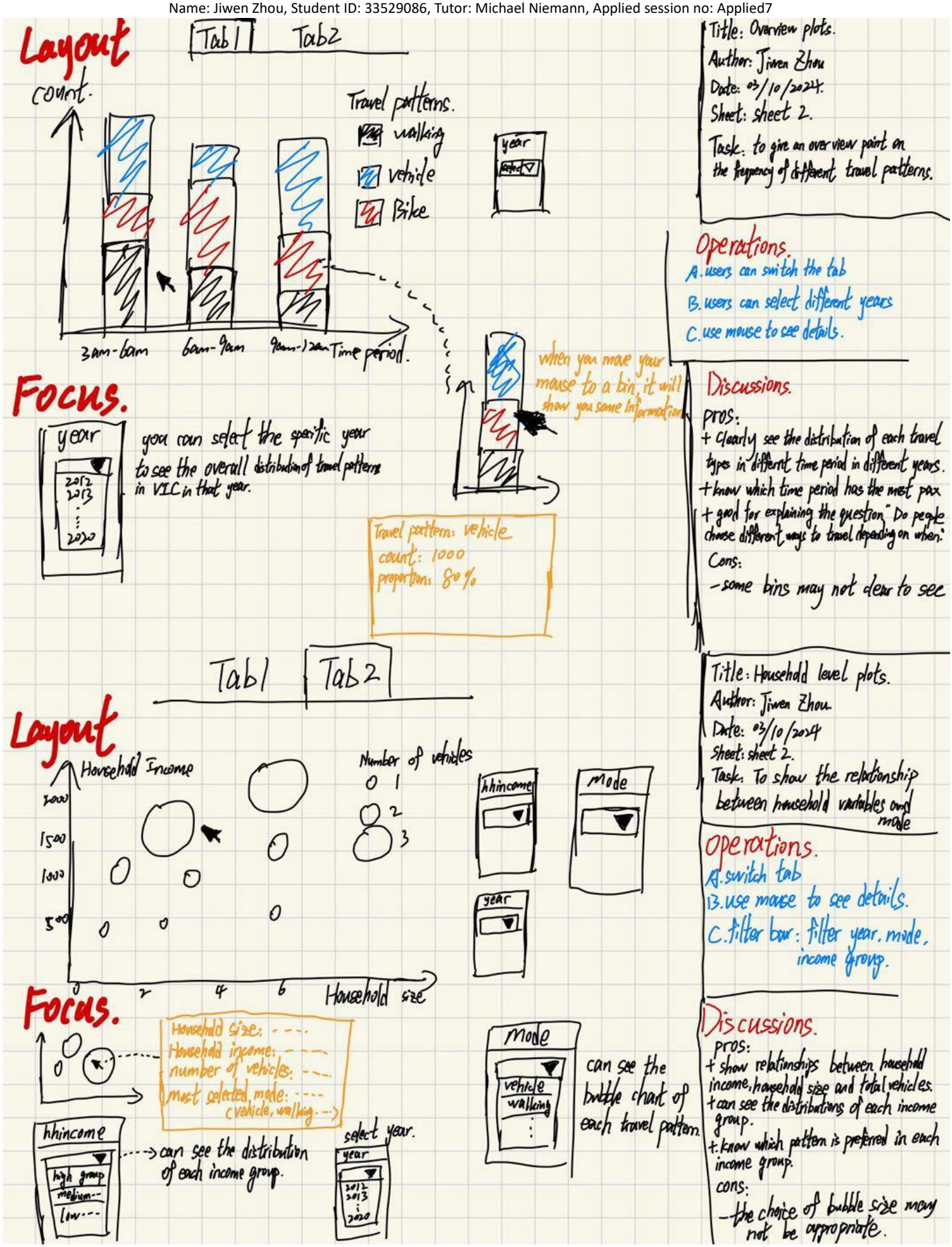
[18]*R Shiny Highcharts - How to Create Interactive and Animated Shiny Dashboards*. (2024). Appsilon.com. https://www.appsilon.com/post/r-shiny-highcharts

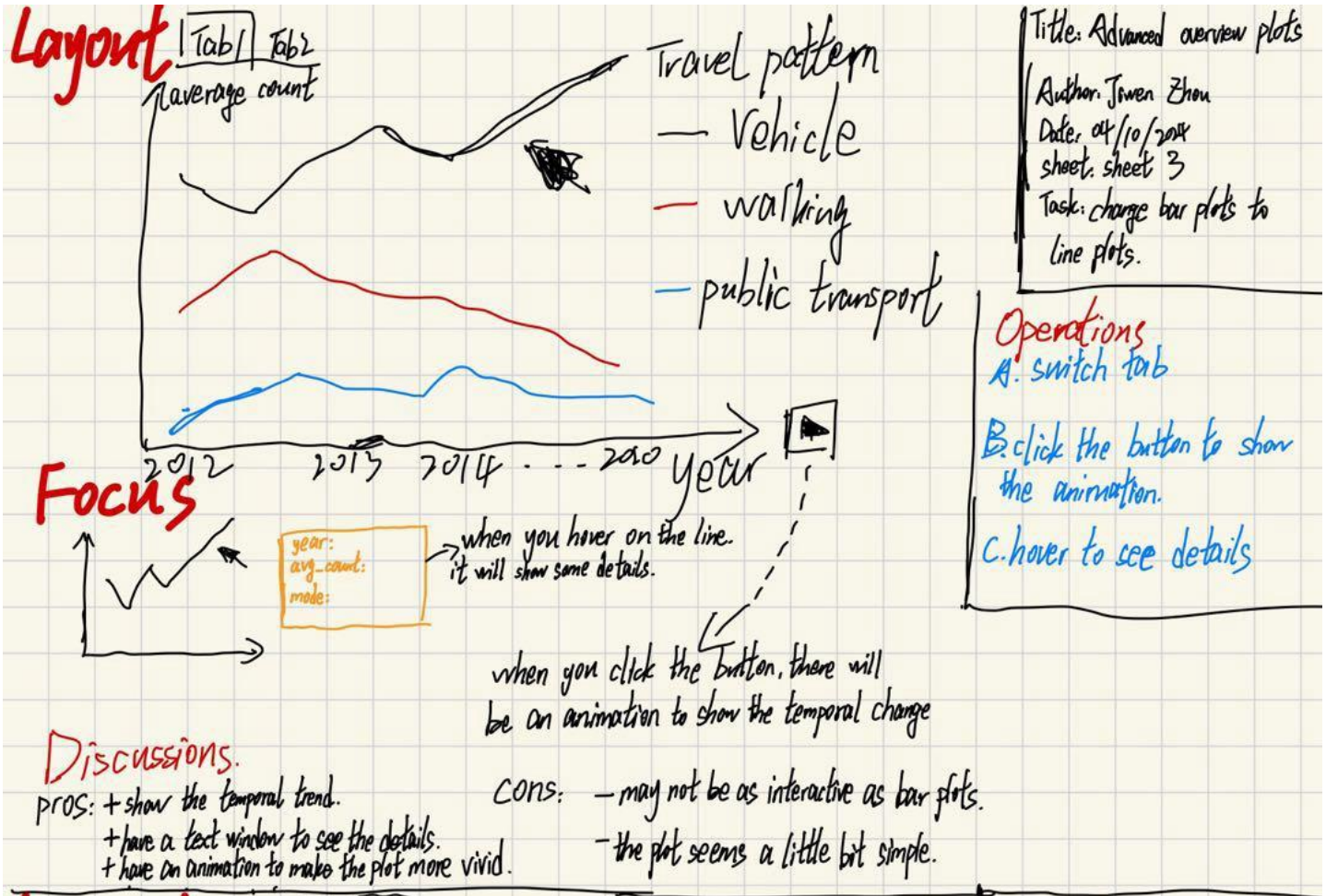
[19]aagarw30. (2022, April 13). *Choose plots and display using radio button*. Gist. https://gist.github.com/aagarw30/fb054a04cc34526b9dddad4981ef2eb1

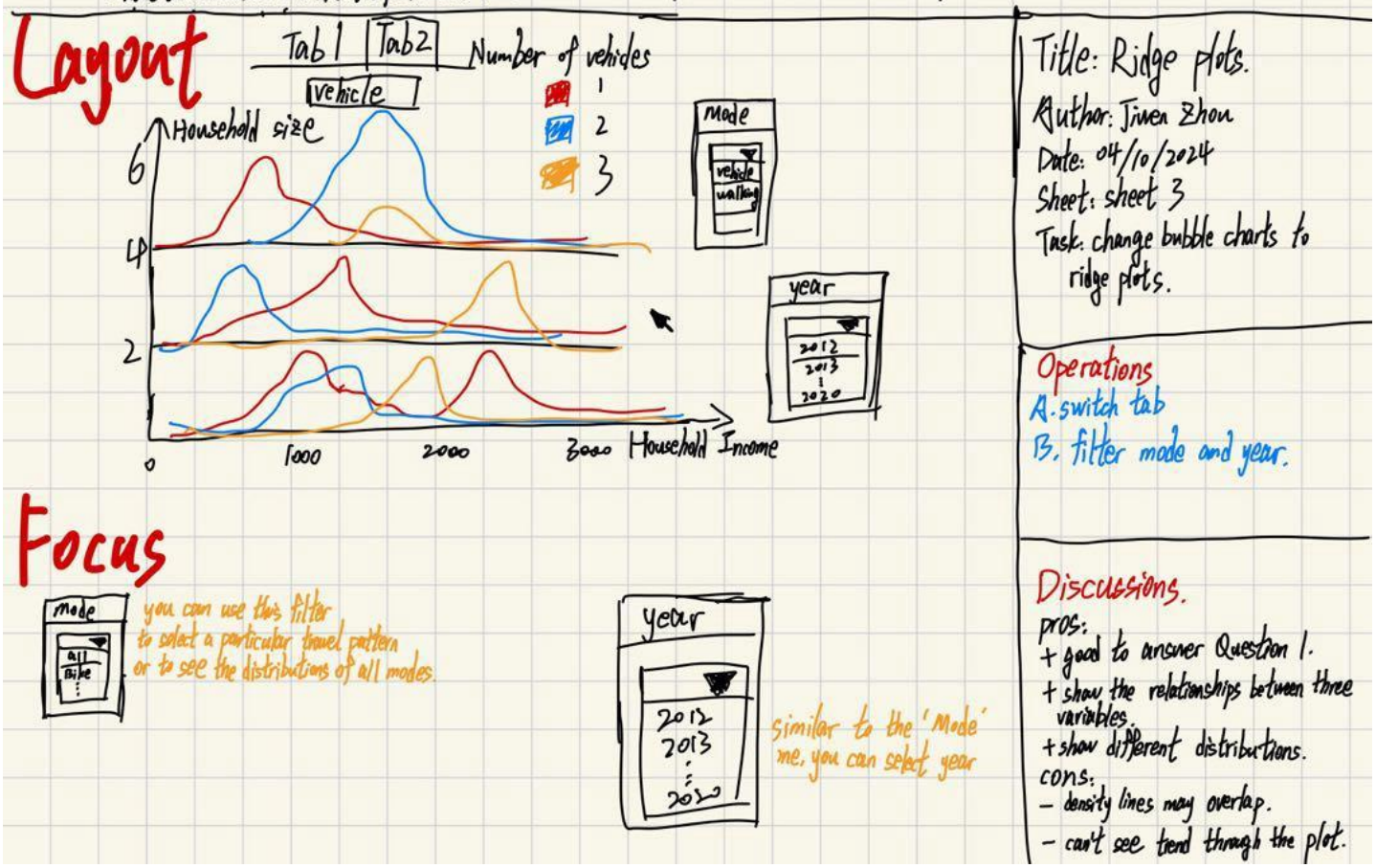
[20]*Victorian Integrated Survey of Travel and Activity (VISTA) - Victorian Government Data Directory*. (n.d.). Discover.data.vic.gov.au. https://discover.data.vic.gov.au/dataset/victorian-integrated-survey-of-travel-and-activity

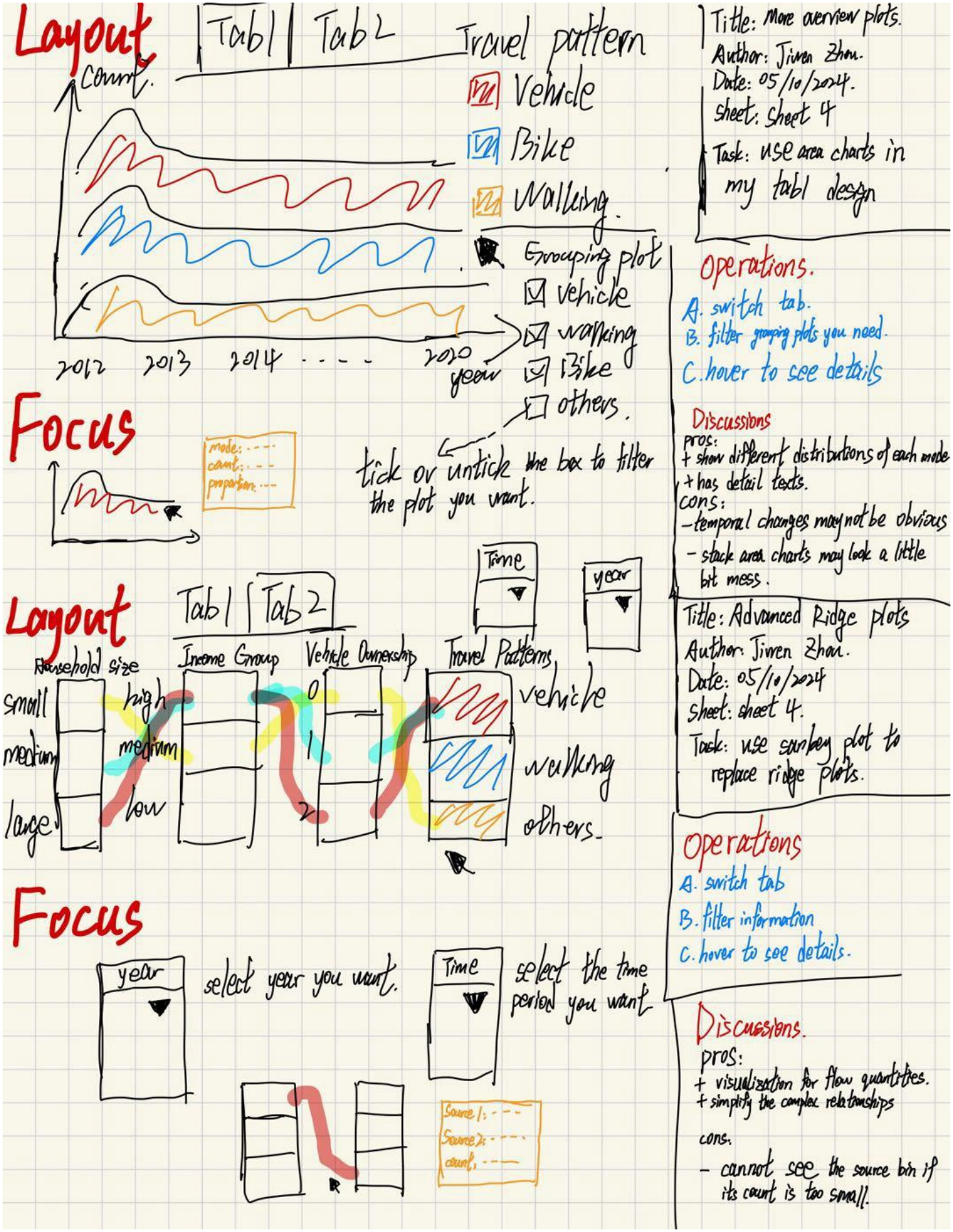
Appendix

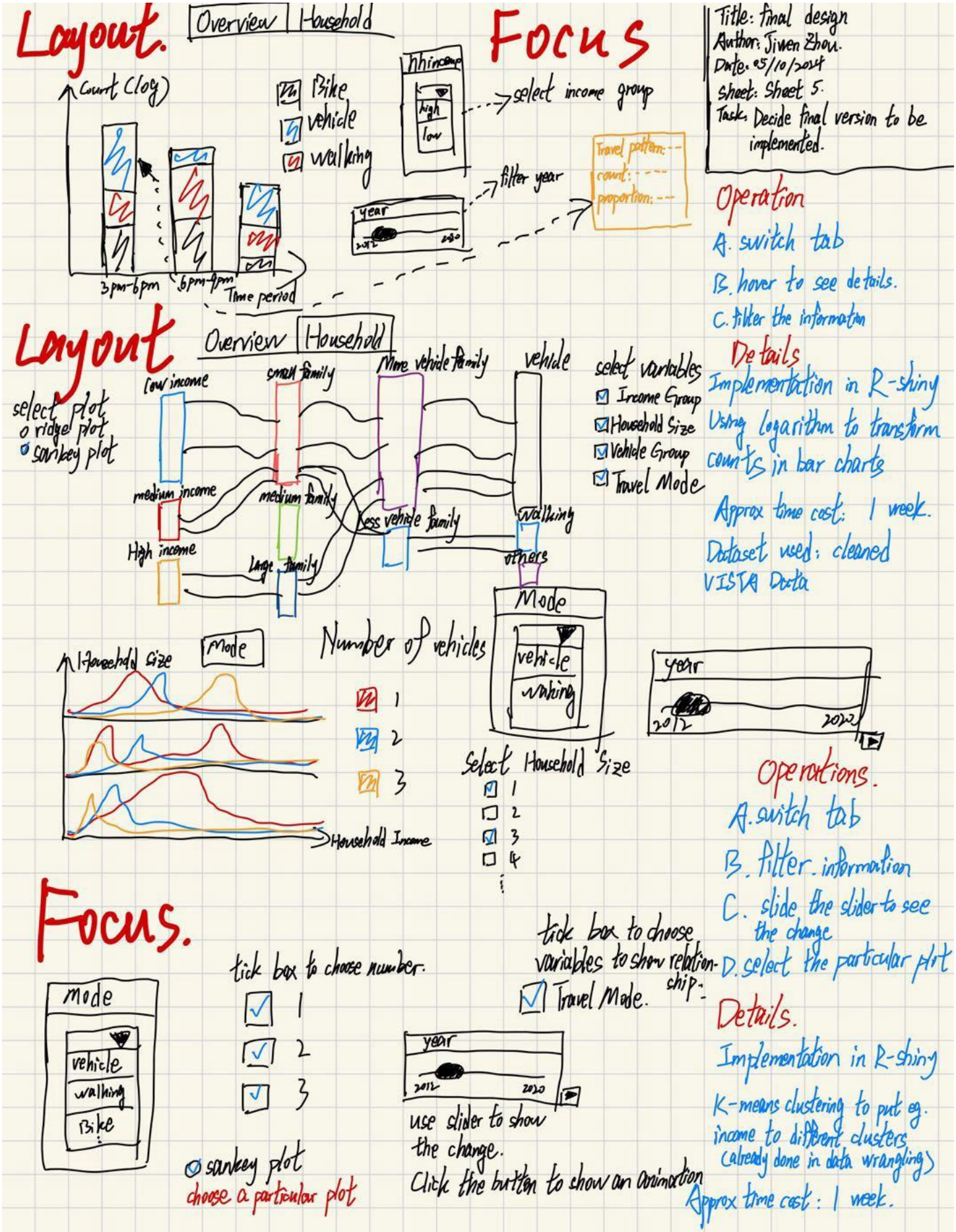

Ideas
① stack bar charts
count
Travel patterns.
walk
public transport
vehicle.
Time period.
② pie charts.
travel patterns.
public transport
others.
walk.
vehicle.
bike
③ line plots.
count(avg)
travel patterns.
vehicle
walk
public transport.
year.
④ scatter plot.
Travel mode.
Walking
Vehicle
Bike
Others.
Travel Distance (km)
⑤ Heat Map.
Mode
Day of week.
⑥ polar plots
Travel Mode Choice by Time Period
Time Period
⑦ Ridge plots.
mode: Vehicle
Household Size
Number of Vehicles
Household Income
⑧ Bubble chart.
Household Size.
Number of vehicles
Household Income
⑨ Area chart
Count
year.
⑩ Box plot.
triptime
Household size
⑪ Mosaic plots.
⑫ sankey plot
Filter. ("-" means Remove the plot.
-② The pie chart is similar to the bar chart, while bar chart could give more info
-⑥ polar plot is similar to stack bar chart.
-⑪ Mosaic plot is similar to the heat map.
-⑩ Box plot is good for statistical summary, like showing the median, Q1, Q3, outliers. But it's not good for visualization to answer the questions.
Categorize
Heat map
Time Dimension
line plot
stack bar charts
area chart
Household Dimension
sankey plot
Ridge plot
Bubble chart.
scatter plot
Combine & Refine
facet by hhsize?
Tab1: overview, 2012-2020 all year trends on different travel modes.
Area charts, line plots, stack bar charts.
Heat map
Tab2: Household -level.
Bubble chart, sankey plot and Ridge plots. scatter plot
have time period on X-axis.
facet by different transport modes.
Question
Do these plots clearly clarify the questions?
Can these plots be combined in one webpage?
What're the pros and cons of these plots?


Layout
Tab1
Tab2
count.
Travel patterns.
walking
vehicle
Bike
year
3am-6am
6am-9am
9am-12am
Time period.
Title: Overview plots.
Author: Jiwen Zhou
Date: 03/10/2024.
Sheet: sheet 2.
Task: to give an overview point on the frequency of different travel patterns.
Operations.
A. users can switch the tab
B. users can select different years
C. use mouse to see details.
Focus.
year
2012
2013
2020
you can select the spesific year to see the overall distribution of travel patterns in VIC in that year.
when you mave your mouse to a bin, it will show you some information.
Travel pattern: vehicle
count: 1000
proportion: 80%
Discussions.
pros:
+ Clearly see the distribution of each travel types in different time period in different years.
+ know which time period has the most pax
+ good for explaining the question "Do people choose different ways to travel depending on when."
Cons:
- some bins may not clear to see
Tab1
Tab2
Layout
Household Income
1500
1000
500
Number of vehicles
0 1
2
3
Household size
hhincome
mode
year
Title: Household level plots.
Author: Jiwen Zhou.
Date: 03/10/2024
Sheet: sheet 2.
Task: To show the relationship between household variables and mode
operations.
A. switch tab
B. use mouse to see details.
C. filter bar: filter year, mode, income group.
Focus.
Household size: ----
Household income: ----
number of vehicles: ----
most selected mode: ----
(vehicle, walking...)
hhincome
high group
medium
low
can see the distribution of each income group.
select year.
year
2012
2013
2020
mode
vehicle
walking
can see the bubble chart of each travel pattern.
Discussions.
pros:
+ show relationships between household income, household size and total vehicles.
+ can see the distributions of each income group.
+ know which pattern is preferred in each income group.
cons:
- the choice of bubble size may not be appropriate.


Layout
Tab1 Tab2
average count
Travel pattern
— vehicle
— walking
— public transport
2012 2013 2014 ... 2020 year
Focus
year:
avg_count:
mode:
when you hover on the line. it will show some details.
when you click the button, there will be an animation to show the temporal change
Discussions.
pros: + show the temporal trend.
+ have a text window to see the details.
+ have an animation to make the plot more vivid.
cons: - may not be as interactive as bar plots.
- the plot seems a little bit simple.
Title: Advanced overview plots
Author: Jiwen Zhou
Date: 04/10/2024
sheet: sheet 3
Task: change bar plots to line plots.
Operations
A. switch tab
B. click the button to show the animation.
C. hover to see details


Layout
Tab1 Tab2
Number of vehicles
1
2
3
vehicle
Household size
6
4
2
0
1000
2000
3000 Household Income
mode
vehicle
walking
year
2012
2013
2020
Focus
mode
all
bike
you can use this filter to select a particular travel pattern or to see the distributions of all modes.
year
2012
2013
2020
similar to the 'Mode' one, you can select year
Title: Ridge plots.
Author: Jiwen Zhou
Date: 04/10/2024
Sheet: sheet 3
Task: change bubble charts to ridge plots.
Operations
A. switch tab
B. filter mode and year.
Discussions.
pros:
+ good to answer Question 1.
+ show the relationships between three variables.
+ show different distributions.
cons:
- density lines may overlap.
- can't see trend through the plot.


Layout
Tab1
Tab2
Travel pattern
Count.
Vehicle
Bike
Walking.
Grouping plot
vehicle
walking
Bike
others.
2012
2013
2014
2020
year
Title: More overview plots.
Author: Jiwen Zhou.
Date: 05/10/2024.
Sheet: Sheet 4
Task: use area charts in my tabl design
Operations.
A. switch tab.
B. filter grouping plots you need.
C. hover to see details
Discussions
pros:
+ show different distributions of each mode
+ has detail tests.
cons:
- temporal changes may not be obvious
- stack area charts may look a little bit mess.
Focus
mode:
count:
proportion:
tick or untick the box to filter the plot you want.
Layout
Tab1
Tab2
Household size
small
medium
large
Income Group
high
medium
low
Vehicle Ownership
0
1
2
Travel Patterns
vehicle
walking
others.
Time
year
Title: Advanced Ridge plots
Author: Jiwen Zhou.
Date: 05/10/2024
Sheet: sheet 4.
Task: use sankey plot to replace ridge plots.
Operations
A. switch tab
B. filter information
C. hover to see details.
Focus
year
select year you want.
Time
select the time period you want
Source 1:
Source 2:
count:
Discussions.
pros:
+ visualization for flow quantities.
+ simplify the complex relationships
cons.
- cannot see the source bin if its count is too small.


Layout.
Overview
Household
Count (log)
Bike
vehicle
walking
3pm-6pm
6pm-9pm
Time period
Focus
select income group
filter year
Travel pattern:
count:
proportion:
Title: final design
Author: Jiwen Zhou.
Date: 25/10/2024
Sheet: Sheet 5.
Task: Decide final version to be implemented.
Operation
A. switch tab
B. hover to see details.
C. filter the information
Details
Implementation in R-shiny
Using logarithm to transform counts in bar charts
Approx time cost: 1 week.
Dataset used: cleaned VISTA Data
Layout
Overview
Household
select plot
ridge plot
sankey plot
Low income
small family
More vehicle family
vehicle
medium income
medium family
Less vehicle family
walking
others
High income
Large family
select variables
Income Group
Household Size
Vehicle Group
Travel Mode
Household Size
Mode
Number of vehicles
1
2
3
Household Income
Mode
vehicle
walking
Select Household Size
1
2
3
4
year
2012
2022
Operations.
A. switch tab
B. filter information
C. slide the slider to see the change
D. select the particular plot
Focus.
Mode
vehicle
walking
Bike
tick box to choose number.
1
2
3
tick box to choose variables to show relationship.
Travel Mode.
year
2012
2020
use slider to show the change.
Click the button to show an animation
sankey plot
choose a particular plot
Details.
Implementation in R-shiny
K-means clustering to put eg. income to different clusters (already done in data wrangling)
Approx time cost: 1 week.