COS30045 – Data Visualisation

Visualisation Design Book

SWINBURNE UNIVERSITY OF TECHNOLOGY

Topic's title: Data Visualisation Project Reflection (Individual)

Student Name and ID: Serena Truong – 104480538

Facilitator: Andrew Guthrie

Class: Thursday 12:30- 2:30pm

Semester 1, 2025, Swinburne University of Technology

Word count: 2724 (excluded Appendix, first page, references, table of content)

Contents

INTRODUCTION 3

Understanding and Application of Data Visualisation Concepts 3

Design process and decisions in ideation and initial 3

Final design decision: 4

Chart Selection and Explanation 4

Data Handling, Programming and Technical Problem Solving 8

Data Processing and Handling with KNIME: 8

Data-Related Challenges I faced: 12

Programming Challenges 12

Technical Problem Solving 14

Team Work and Peer Assessment 16

Use of GenAI Tools 17

Conclusion 17

References: 17

APPENDIX 18

Adherence to Good Design Principles 18

INTRODUCTION

The purpose of this reflection is used to reflect my journey through the Data Visualisation unit, focusing on my design process, technical and conceptual development, and the key lessons learned. Throughout the unit, I worked on a project of Data Visualisation Project which aimed to visualise mobile phone use while driving dashboard in Australia using BITRE 2023 data based on the last Critique assignment, providing insights for group of target users. This report reflects on the challenges, key modification, and learning outcomes from the project. This report outlines my design thinking, technical choices, teamwork experiences, analyse visualisations and lessons learned.

Understanding and Application of Data Visualisation Concepts

Design process and decisions in ideation and initial

At the beginning of the project, my idea for the dashboard was quite ambitious because we had to explore the datasets related to “BITRE” in 2023. At first, I planned to do a multi-page interactive website that would include a main dashboard and a comparison section, especially the idea of combining or comparing 2 visualisations on a single line in one main dashboard like figure 1 and 2 below.

Figure 1. Initial main dashboard in Figma prototype

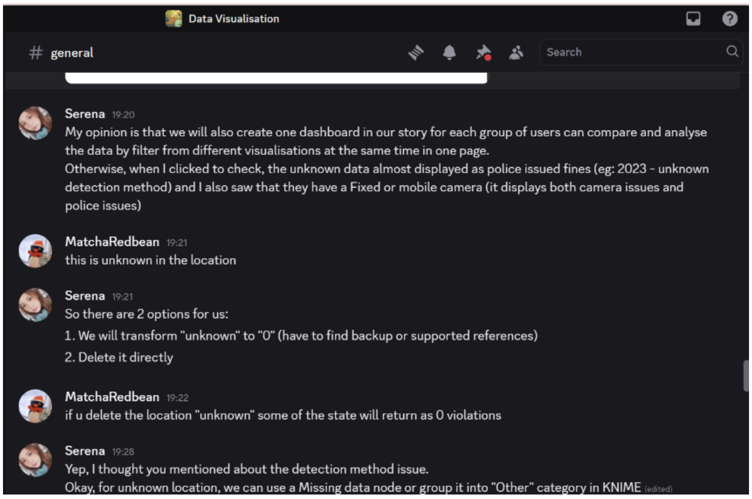


Figure 2. One of the evidence of design decisions

However, at the nearly end of the project (Week 13), I changed the plan and created About Us page, known as data story, was designed to explain the purpose of the website dashboard and help users identify which target audience they belong to. The purpose was to provide insights to users' needs and requirements, including students and young drivers, the general public, government officials, transportation safety analysts, and researchers.

In my early mockups and sketches in Figma, I created 2 pages such as main dashboard and comparison page which provided a basic layout for displaying Annual fines and Monthly violation trends with basic navigation bar and compared 2 charts I the same line. I aimed these mockups would offer an informative layout that would allow users to scroll down into specific data segments.

Final design decision:

Our final decision includes a main dashboard displaying four visualisations on mobile phone use while driving in 2023 and an About Us page showing the dashboard's purpose. Each chart was selected based on data type and clarity in showing trends across Australian states. Design choices focused on readability, accessibility and I learned that applying feedback from previous critiques assignment and cleaned dataset in KNIME improved the project design process.

Chart Selection and Explanation

Line Chart for displaying Monthly Fines by Detection Method

The line chart is designed for displaying trends over time, making it suitable for analysing "Whether phone use while driving is increasing." By plotting monthly fines against detection methods, users can easily observe fluctuations and patterns in data.

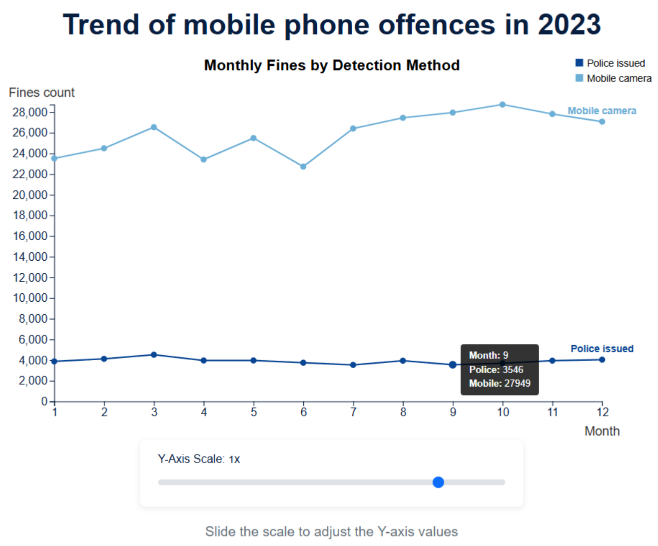


Figure 3. Line chart displays trend of mobile phone offences in 2023

The challenge of for this specific project was the disparity in scale between police-issued fines and other detection methods in the original visualisation. The volume of fines issued by police was much lower than that of other methods, making it difficult to visualise trends effectively To address this, we adjusted the scale and made by using a logarithmic scale or dual axes, allowing for a clearer representation of both datasets without losing the visibility of the police-issued fines line.

I understand that the bar chart shows total violations by jurisdictions because of the huge gap between states, several low metric states like ACT and NT bars could not be drawn. Then our team believes that it would be more appropriate to display the normalised ratio of fines reported in each state per 10,000 issued licenses. This strategy not only improved the visualisation of the bar representing each state but also allows for fair comparisons between states regardless of population size.

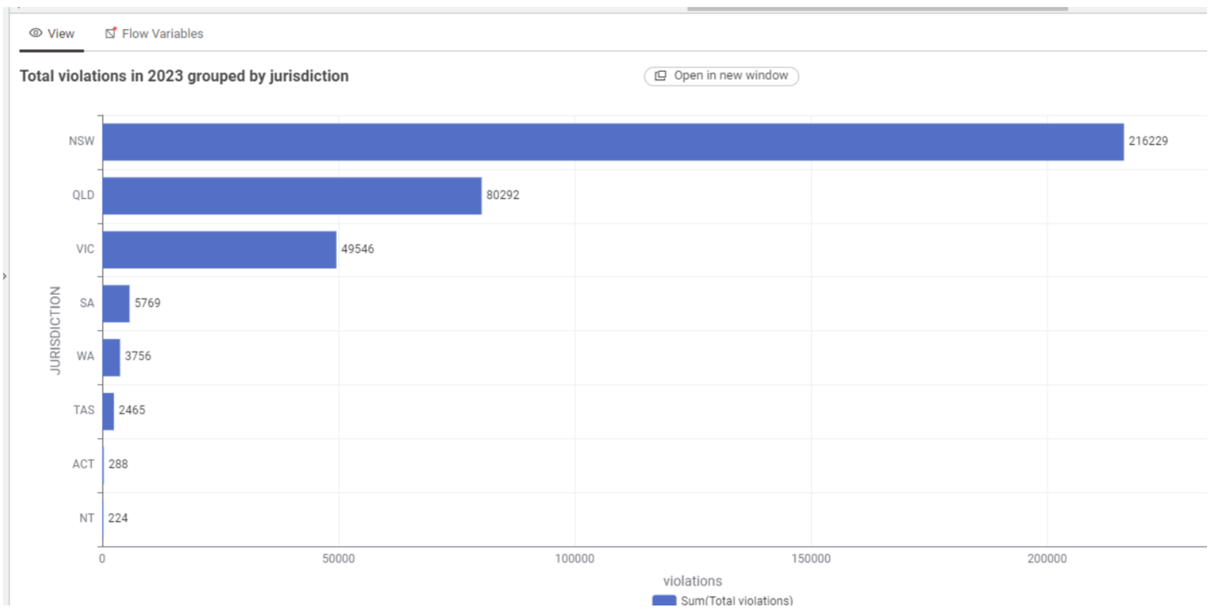


Figure 4. KNIME's Total violations in 2023 grouped by jurisdiction

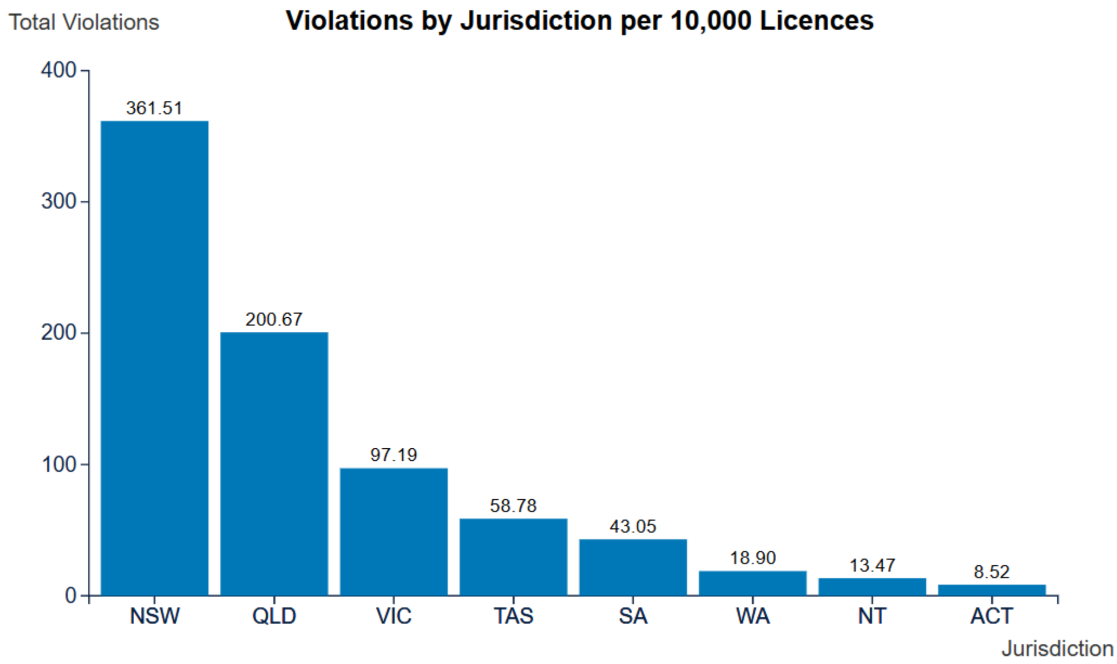


Figure 5. Total Violations in 2023 by jurisdiction per 10,000 licenses

Bar Chart for displaying Fines by Jurisdiction per 10,000 Licences

The initial design concept of dashboard comparing two charts side by side which using absolute fine numbers led to misleading comparisons between jurisdictions due to differences in state sizes. We decided to normalise the data by presenting fines per 10,000 licenses. This approach allowed for fairer comparisons across jurisdictions, providing a clearer picture of enforcement relative to the number of drivers.

Bar Chart/Histogram (Total Fines by Age group):

The data for the 0-16 age group posed a significant challenge, as the bar could not be drawn due to the massive gap in fines compared to other age groups My initial reaction was one of concern on how I can display data well, as it indicated a potential oversight in data representation. But throughout discussions with my teammate and feedbacks from supervisors, we decided that the 0–16 age group was not relevant to our target audience even though we have target group of young driver and student (But this type of user can be known as age group of above 16 year olds). So, removing it contributed to the integrity of the data display. This experience helped me to recognise the value of data accuracy and ensuring that the information presented is appropriate for the audience's needs.

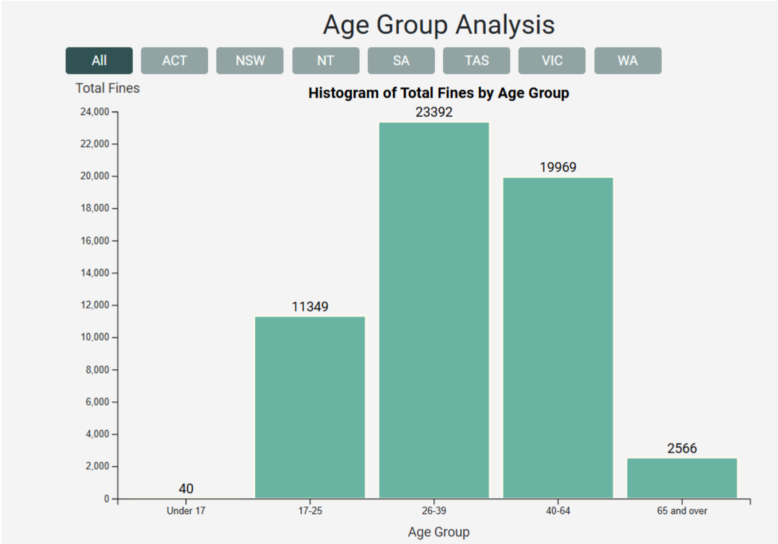


Figure 6. Removed age group of 0-16

Donut Chart for displaying Location breakdown of Fines by Jurisdiction:

The donut chart is chosen to display for this location breakdown which presenting on click (*The Location Insights Dashboard*, 2022), to improve user engagement and avoid overwhelming users with too much information at once. This method enables users to discover or find out the data at their own time or space.

Otherwise, this chart was designed to display the Location breakdown with more detail in some areas like Major cities, Inner Regional, Remote area, etc to an attribute for visualisation.

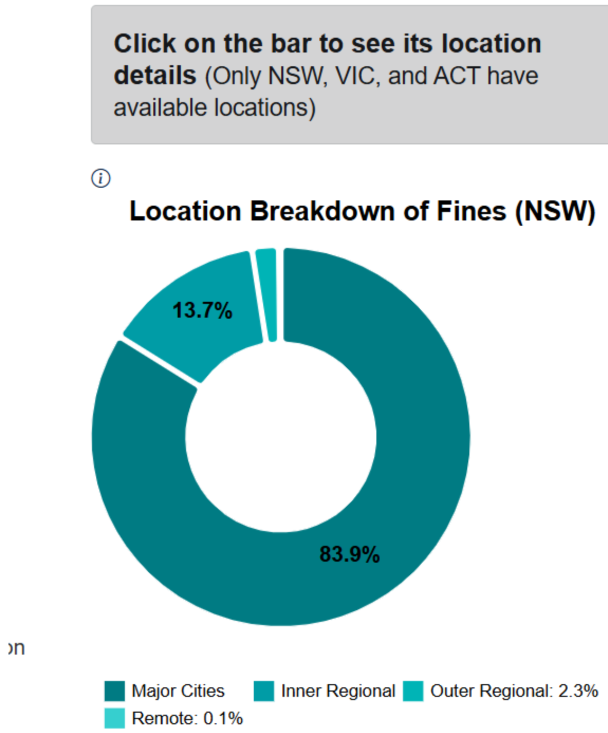


Figure 7. Donut Chart/Pie Chart of Location Breakdown of Fines in specific Jurisdiction (NSW)

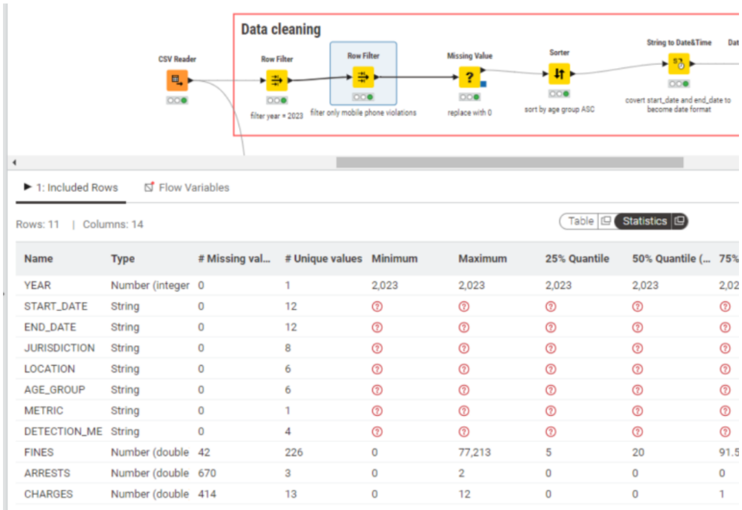
Data Handling, Programming and Technical Problem Solving

This project required important information preparation, modification, and processing with KNIME, D3.js for interactive data visualisation and bootstrap grid framework for styling website. The data came from a government dataset on mobile phone use while driving across different Australian jurisdictions in 2023.

Data Processing and Handling with KNIME:

KNIME is a tool which used process the data. As the requirement of this project was focused on data in 2023, we first used Row Filter node to filter out only records of mobile phone use in 2023 with criteria are “Year = 2023” and “metric = mobile_phone_use”.

Handle Missing Values: To check if the data has missing values, we used “missing values” node to determine For these missing numbers, the detection algorithms were inapplicable in some states, area or the data was not recorded enough. Then, we decided to replace them with 0 (for arithmetic calculations) and add a small note in our designed visualisation for users to know that the metrics are not available in some areas to avoid confusing.



Name	Type	# Missing val...	# Unique values	Minimum	Maximum	25% Quantile	50% Quantile (...	75%
YEAR	Number (integer	0	1	2,023	2,023	2,023	2,023	2,02
START_DATE	String	0	12	?	?	?	?	?
END_DATE	String	0	12	?	?	?	?	?
JURISDICTION	String	0	8	?	?	?	?	?
LOCATION	String	0	6	?	?	?	?	?
AGE_GROUP	String	0	6	?	?	?	?	?
METRIC	String	0	1	?	?	?	?	?
DETECTION_ME	String	0	4	?	?	?	?	?
FINES	Number (double	42	226	0	77,213	5	20	91.5
ARRESTS	Number (double	670	3	0	2	0	0	0
CHARGES	Number (double	414	13	0	12	0	0	1

Figure 8. Statistic view node showed Missing values in Fines, Charges and Arrests

Normalisation and Transformation: Convert String to Date and Time Part Extractor In the original dataset, the Start and End dates were in string format. We changed them to date format by utilising the String to Date&Time node, which made month calculation easier. To determine which record represents which months, we utilised the Date and Time Part Extractor to extract the End_date property into Month.

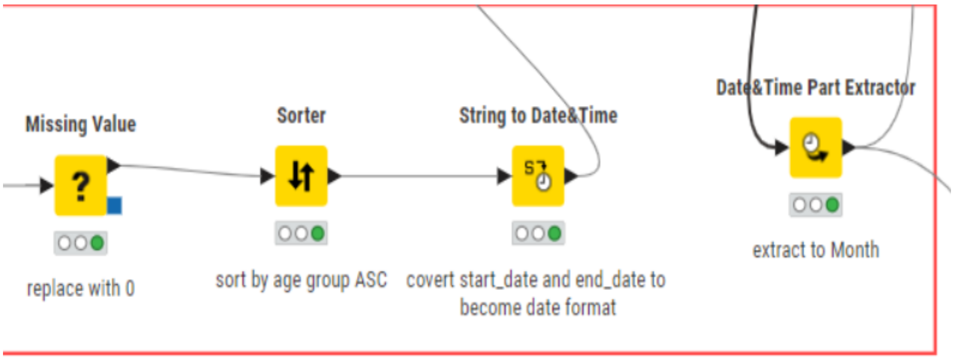


Figure 9. Convert date type and extract to Month

QLD does not have monthly data: Some entries have a “state_date” of 1-1-2023 but “end_date” of 31-12-2023. Without the monthly measures, we normalised the QLD data by dividing the total over 12 months by the average monthly fine count. To analyse the QLD monthly data, the average fines count for each detection method is determined with the Math formula node and rounded to the nearest integer. Then, the data for QLD is then exported to an Excel file (.cvs) using the Excel Writer node. To keep things simple, the month attribute was manually entered into the Excel spreadsheet. After transforming data, we used the Concatenate node to combine the processed QLD data into the dataset.

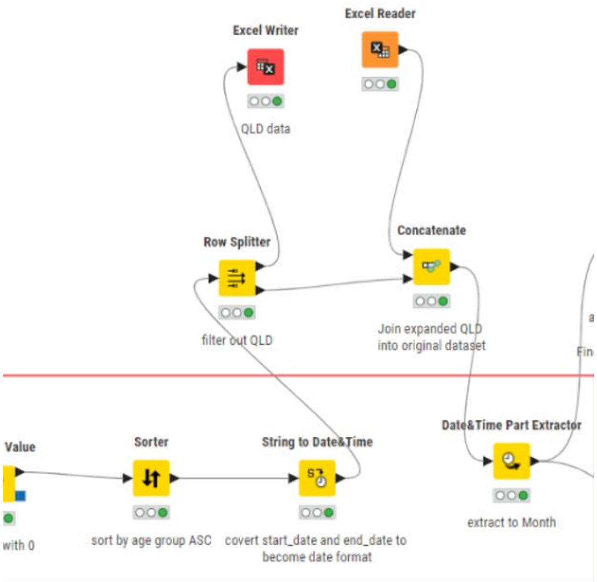


Figure 10. Transforming QLD data

▶ 1: Included Rows ▶ 2: Excluded Rows Flow Variables

Rows: 2 | Columns: 11 Table Statistics

#	RowID	YEAR Number (integer)	START_DATE Local Date	END_DATE Local Date	JURISDICTI... String	LOCATION String	AGE_GROUP String	METRIC String	DETECTION_... String	FINES Number (double)	ARRESTS Number (double)	CHARGES Number (double)
1	Row18	2023	2023-01-01	2023-12-31	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile cai	77,213	0	0
2	Row18	2023	2023-01-01	2023-12-31	QLD	Unknown	Unknown	mobile_phone_use	Police issued	3,079	0	0

Figure 11. Before normalisation of QLD data

Rows: 780 | Columns: 11 Table Statistics

#	RowID	YEAR Number (integer)	START_DATE Local Date	END_DATE Local Date	JURISDICTI... String	LOCATION String	AGE_GROUP String	METRIC String	DETECTION... String	FINES Number (double)
1	Row0	2023	2023-01-01	2023-01-31	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
2	Row1	2023	2023-01-02	2023-02-28	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
3	Row2	2023	2023-01-03	2023-03-31	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
4	Row3	2023	2023-01-04	2023-04-30	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
5	Row4	2023	2023-01-05	2023-05-31	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
6	Row5	2023	2023-01-06	2023-06-30	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
7	Row6	2023	2023-01-07	2023-07-31	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
8	Row7	2023	2023-01-08	2023-08-31	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
9	Row8	2023	2023-01-09	2023-09-30	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
10	Row9	2023	2023-01-10	2023-10-31	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
11	Row10	2023	2023-01-11	2023-11-30	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
12	Row11	2023	2023-01-12	2023-12-31	QLD	Unknown	Unknown	mobile_phone_use	Fixed or mobile ca	6,434
13	Row12	2023	2023-01-01	2023-01-31	QLD	Unknown	Unknown	mobile_phone_use	Police issued	257
14	Row13	2023	2023-01-02	2023-02-28	QLD	Unknown	Unknown	mobile_phone_use	Police issued	257
15	Row14	2023	2023-01-03	2023-03-31	QLD	Unknown	Unknown	mobile_phone_use	Police issued	257
16	Row15	2023	2023-01-04	2023-04-30	QLD	Unknown	Unknown	mobile_phone_use	Police issued	257
17	Row16	2023	2023-01-05	2023-05-31	QLD	Unknown	Unknown	mobile_phone_use	Police issued	257
18	Row17	2023	2023-01-06	2023-06-30	QLD	Unknown	Unknown	mobile_phone_use	Police issued	257
19	Row18	2023	2023-01-07	2023-07-31	QLD	Unknown	Unknown	mobile_phone_use	Police issued	257
20	Row19	2023	2023-01-08	2023-08-31	QLD	Unknown	Unknown	mobile_phone_use	Police issued	257

Figure 12. After normalisation of QLD data

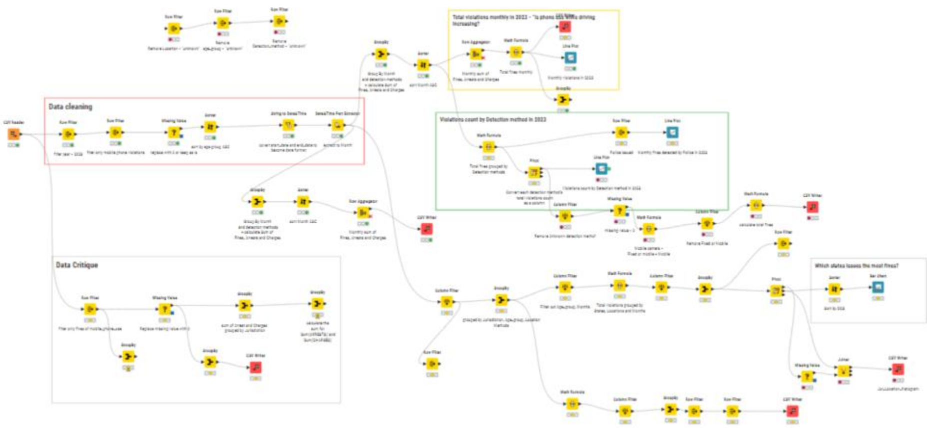


Figure 13. KNIME Workflow in the design process

Data filtering: After processing the data, we removed "Unknown" values and unneeded columns based on the requirements of each visualisation. Finally, use the CSV Writer node to export the data. We used these CSV files to read into D3 and display useful and meaningful visualisations for users in our dashboard.

1: Included Rows   Flow Variables

Rows: 11 | Columns: 14

Name	Type	# Missing val...	# Unique values	Minimum	Maximum	25% Q
YEAR	Number (integer	0	1	2,023	2,023	2,023
START_DATE	String	0	12	?	?	?
END_DATE	String	0	12	?	?	?
JURISDICTION	String	0	8	?	?	?
LOCATION	String	0	6	?	?	?
AGE_GROUP	String	24	6	?	?	?
METRIC	String	0	4	?	?	?
DETECTION_ME	String	0	8	?	?	?
FINES	Number (double	181	1250	0	469,785	4
ARRESTS	Number (double	3913	64	0	106	0
CHARGES	Number (double	2537	192	0	800	0

Figure 14. Data Exploration

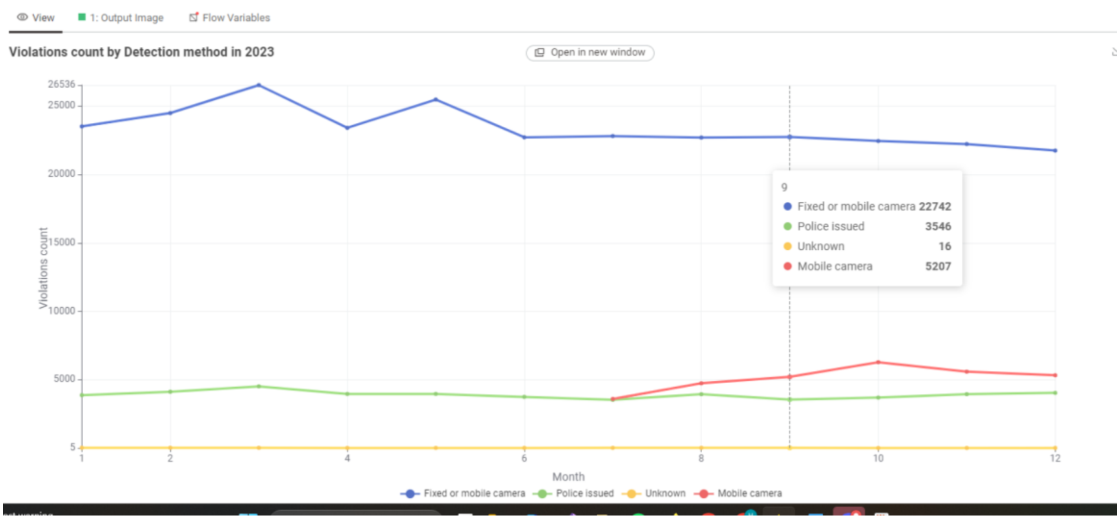


Figure 15. Line chart is designed by KNIME node

Based on the figure 10 and 11, the KNIME chart designed clearly how the basic visualisation like line chart of Violations count by Detection method in 2023 (displayed by each month) to confirm that the data processed is correct and understandable to users. The chart shows a “Unknown” category with a low number of counts. The supervisor recommended that we remove the “Unknown” data because they are informative to users. We then filtered out the “Unknown” and combined “Fix and mobile camera” into “Mobile Camera” (We did what the tutor suggested in Stand up 2 or 3), combining the taction methods to two categories (police issued and mobile camera). My thought on this issue was to remove directly because it does not influence much of the dataset.

Data-Related Challenges I faced:

We were unable to generate additional rows or attributes using KNIME nodes when transforming QLD records into 12 records for 12 months. As a result, we had to manually add the data to an Excel file. Additionally, we struggled to convert each location detail such as Major cities, Inner Regional, Remote area, etc to an attribute for visualisation. Hence, to create the pie chart of location breakdowns for each state, we had to read each location as a category variable. After some investigation on KNIME tutorials, we discovered that the Pivot node can convert columns to rows and vice versa (KNIME AG, 2024). Pivot node also allows you to aggregate specified data.

Difficulty to turn each location in detail like major cities, inner regional, remote area as attributes in the geographic column. To create the pie chart of location breakdowns for each state, we had to read each location as a category variable. After some investigation on KNIME tutorials, we discovered that the Pivot node can convert columns to rows and vice versa. Pivot node also allows you to aggregate specified data and use D3 to finalise these charts below (figure 11).

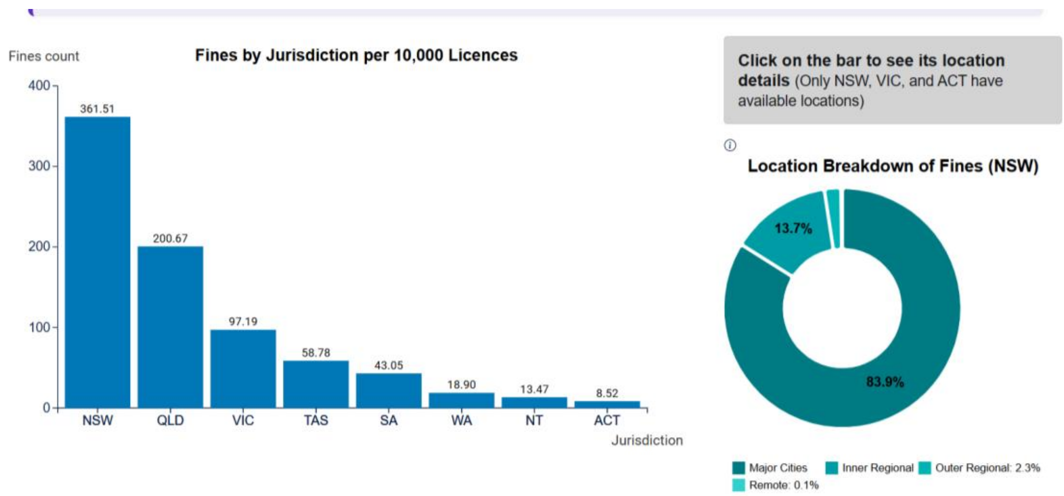


Figure 16. Fines by Jurisdiction by bar chart and turn each location in detail by pie chart

Programming Challenges

A D3 technical challenge came while trying to display "Fines, Arrests, and Charges" together using a stacked bar chart. My code, which used d3.stack(), showed that smaller values were unrecognisable, segments for “arrests” and “charges” were visually lost within larger “fines” bars, making them indistinguishable.

```
const container = svg.append("g")
  .attr("transform", `translate(${margin.left},${margin.top})`);

// Create stacked data
const stack = d3.stack()
  .keys(["fines", "arrests", "charges"]);
const stackedData = stack(data);

// Draw the bars
container.selectAll("g")
  .data(stackedData)
  .join("g")
  .attr("fill", d => colorScale(d.key)) // Assign color based on the key
  .selectAll("rect")
  .data(d => d)
  .join("rect")
  .attr("x", d => xScale(d.data.jurisdiction))
  .attr("y", d => yScale(d[1]))
  .attr("height", d => yScale(d[0]) - yScale(d[1]))
  .attr("width", xScale.bandwidth());
```

Figure 17. Stacked bar code to display 3 items

data > annual_offences.csv

```
"JURISDICTION","Sum(ARRESTS)","Sum(FINES)","Sum(CHARGES)"
"ACT",0,288,0
"NSW",0,216042,187
"NT",5,205,14
"QLD",0,80292,0
"SA",0,5769,0
"TAS",0,2428,37
"VIC",0,49546,0
"WA",7,3592,157
```

Figure 18. Data file of annual offences

d3js.org/d3-shape/stack

Menu

```
svg.append("g")
  .selectAll("g")
  .data(series)
  .join("g")
    .attr("fill", d => color(d.key))
  .selectAll("rect")
  .data(D => D)
  .join("rect")
    .attr("x", d => x(d.data[0]))
    .attr("y", d => y(d[1]))
    .attr("height", d => y(d[0]) - y(d[1]))
    .attr("width", x.bandwidth());
```

Figure 19. D3js example of stacked bar code

code-files / chapter_05 / 5.2-Stack_layout / end / js / **stacked-bars.js**

Code Blame 84 lines (65 loc) · 2.27 KB

```
const drawStackedBars = (data) => {
  /*    Declare the vertical scale    */
  /***********************************/
  // Find the lower and upper boundary of the domain based on the annotated data
  const minLowerBoundaries = [];
  const maxUpperBoundaries = [];
  annotatedData.forEach(series => {
    minLowerBoundaries.push(d3.min(series, d => d[0]));
    maxUpperBoundaries.push(d3.max(series, d => d[1]));
  });
  const minDomain = d3.min(minLowerBoundaries);
  const maxDomain = d3.max(maxUpperBoundaries);

  // Declare the vertical scale
  const yScale = d3.scaleLinear()
    .domain([minDomain, maxDomain])
    .range([innerHeight, 0]);



  /*******************************/
  /*    Append the rectangles    */
  /*******************************/
  annotatedData.forEach(series => {
    innerChart
      .selectAll(`.bar-${series.key}`)
      .data(series)
      .join("rect")
        .attr("class", d => `bar-${series.key}`)
        .attr("x", d => xScale(d.data.year))
        .attr("y", d => yScale(d[1]))
        .attr("width", xScale.bandwidth())
        .attr("height", d => yScale(d[0]) - yScale(d[1]))
        .attr("fill", colorScale(series.key));
  });

```

Figure 20. Stacked-bar.js from D3js in action third edition code file

I researched and learnt the code from the source code named D3js in action third edition which is suggested during the unit from convenor's lecture and tutorial workshop.

Technical Problem Solving

For both data processing and programming:

Remove unknown values: In our visualisation of violations recorded by locations, several jurisdictions did not have the location details or data available. As a result, many values are labelled as "Unknown". During standup 2, we put this question into the "Blocker and Challenge" and aimed to receive the answers to address this issue. Since the location is "Unknown", it cannot provide users with valuable information. Then I provided two options while presenting like removing the "unknown" out of the records because it is not the unnecessary information and also not a big effect to the provided dataset or separate it to new category column and find more data from different reliable source references for analysis. The facilitator also proposed deleting these "unknown" values as they are unhelpful to users while adjusting for most of the analysis dataset and building up visualisations in the programming process.

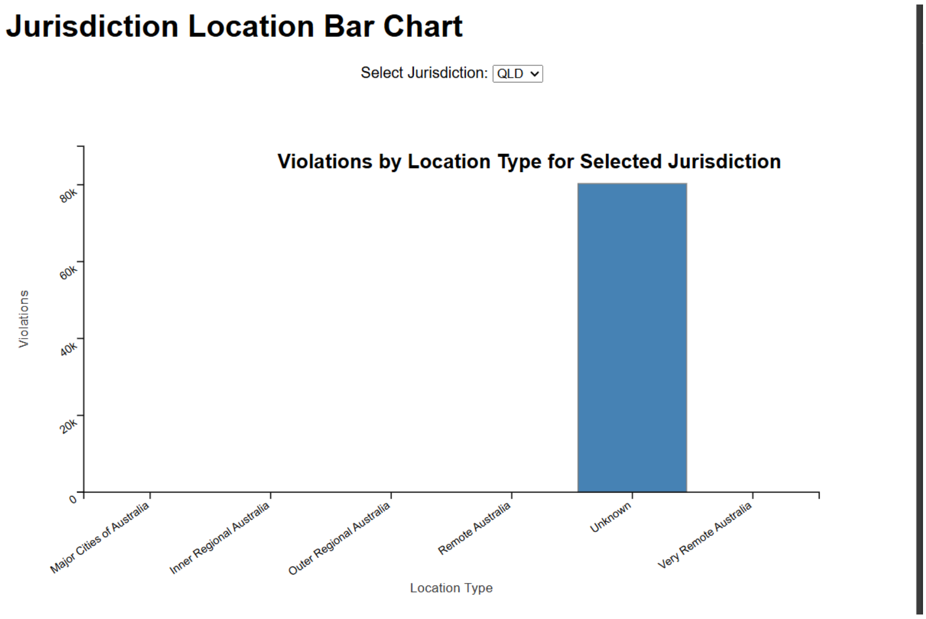


Figure 21. Location "unknown" has the many violations found

Due to limited availability of location information in only three states (NSW, VIC, and ACT), I and my teammate decided to discontinue using the bar chart to track infractions in specific jurisdictions. In contrast, we include a bar chart showing number of fines per 10,000 driving licenses per jurisdiction, as well as a pie chart displaying the fraction of places where violations were recorded, if applicable.

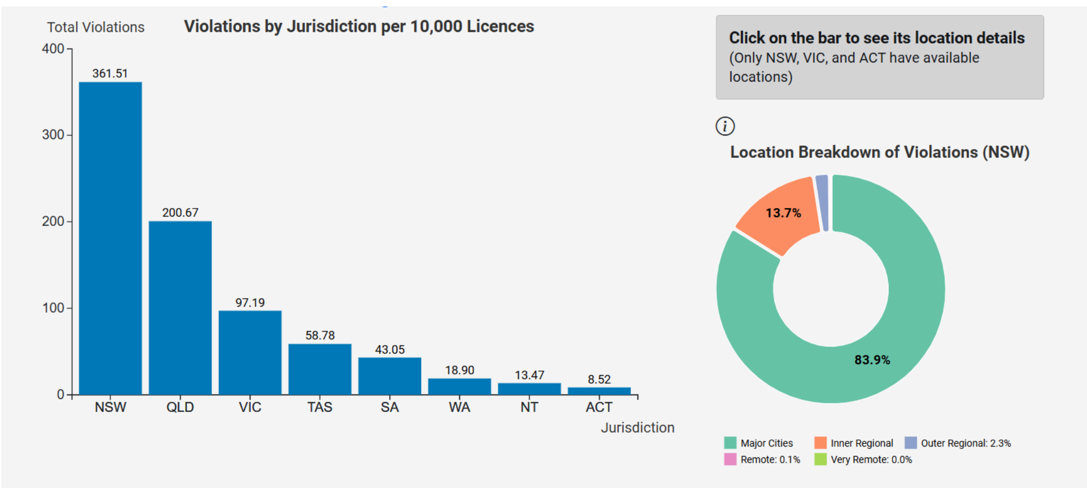


Figure 22. The visualisation is created after removing the unknown records.

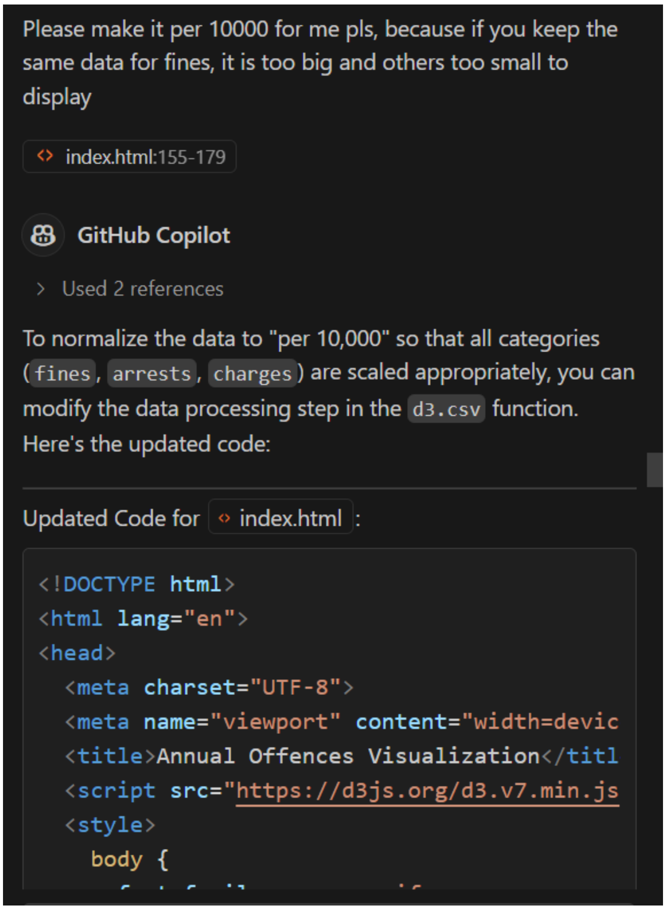


Figure 23. Copilot helped to modify the code based on my idea

Team Work and Peer Assessment

Reflect on my role and my contribution toward the project:

In our team, I was responsible for data cleaning and processing in KNIME, designing three D3 visualisations, and creating the interactive "About Us" page to explain the dashboard's purpose and users. I also implemented visual interactivity and contributed weekly to the design book, including finalising and formatting it. Additionally, I prepared key points for our PowerPoint and submitted our work. Our team used Discord to communicate, share files for weekly submissions, and collaborate on website wireframes, which helped us stay organised and assign tasks.

Reflect on group dynamics, challenges and solutions:

One challenge was differing opinions on charts designs and website layout. For the stacked bar I thought the stacked bar chart showing 3 fine types by state would not display well due to large data gaps or data discrepancies, while my teammate thought it was a limitation of my D3 skills. However, she was also unable to provide a clear alternative, so we agreed to delete the chart and replace it with another effective visualisation, which improved our project's development.

One conflict we faced was my initial plan to include a comparison dashboard with side-by-side charts but in week 13 my teammate removed the idea without informing or consulting me. We resolved this in Discord by proving its pros and cons since she already placed 3–4 visualisations in the main dashboard, adding more could cause redundancy and distract users.

While she created most of the D3 visuals, I contributed ideas on layout and appropriate chart types. This experience taught me that teamwork depends not just on technical skills but also on clear communication, mutual accountability, and adaptability. Drawing from past group work, I proposed using shared tools like Google Docs to assign tasks fairly. Overall, this reinforced the value of proactive problem-solving and collaboration. Based on previous group work, I suggested using a document with using tool like Google Docs to define roles or assign tasks based on each member's strength and ensure equal workload. Overall, by this experience reinforces importance of clear communication, proactive problem solving and flexibility in collaborative data projects.

Peer assessment:

I completed the peer assessment, providing my constructive feedback to team member and reflecting on our collective efforts throughout the unit.

Use of GenAI Tools

I used GENAI to refine my sentences if we used it multiple times in the report. It is so helpful if the users know how to use it correctly. It suggested me a template structure for the report and supported me on how to implement interactive features or build up visualisations with D3. For example, while GENAI might suggest code, I often had to modify it to meet my project requirement dataset and design objectives. However, I also recognised that when using GENAI, the tone was totally different from the human tone, or they cannot understand full sentence meaning when I asked refining the provided words or sentences.

Conclusion

Overall, this unit has improved my knowledge of data visualisation, including both design concepts and technical skills. I developed the ability to balance visual appeal, usability, and accessibility while working with real-world datasets. I further improved my skill in data preparation like cleaning data and enhanced my skills in D3 coding. As a result, this unit has boosted me with both the technical tools and critical thinking needed to create effective, engaging data visualisations in my future professional work.

References:

KNIME AG. (2024). *How to make data in rows as columns?* KNIME Community Forum. https://forum.knime.com/t/how-to-make-data-in-rows-as-columns/77914

The Location Insights Dashboard. (2022). ActivTrak Help Center. https://support.activtrak.com/hc/en-us/articles/8298409034267-The-Location-Insights-Dashboard

Munzner, T. (2014). *Visualization analysis and design.* A K Peters/CRC Press.

Holtz, Y. (2025). *Horizontal barplot in d3.js*. D3-Graph-Gallery.com. https://d3-graph-gallery.com/graph/barplot_horizontal.html

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Wiley.

Doughnut and Pie Charts | Chart.js. (2025). Chartjs.org. https://www.chartjs.org/docs/latest/charts/doughnut.html

APPENDIX

Adherence to Good Design Principles

Graphical Integrity: As shown in the charts, all bar charts start at zero on the Y-axis, ensuring bar lengths accurately reflect data and avoid exaggeration. We picked appropriate chart types for each dataset: line charts for monthly trends, bar charts for discrete comparisons, and donut charts for proportions.

Accessibility: Accessibility was a top priority, especially when it came to colour. We tested our visuals with “Coblis” or Colour Blindness Simulator to ensure they were accessible to colourblind people. Our main theme is Australian Government navy blue, and for categories, we selected blue colours with various saturation. During this test, I learnt that the strategy helped everyone, including those with monochromacy, in distinguishing elements, and was especially useful for the 'Location breakdown of Fines by Jurisdiction' donut chart, which contains several categories.

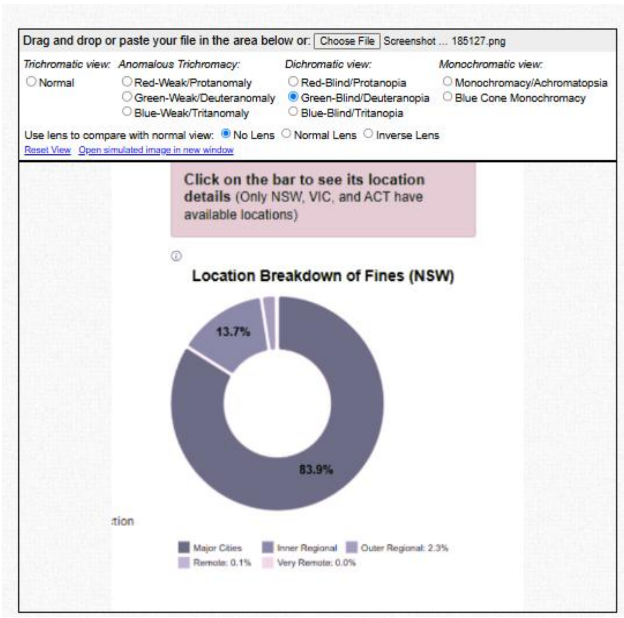


Figure 24. Donut chart was tested against a colour blindness simulator

Scalability: Our team decided to use Bootstrap grid framework to design the D3 & website layout. My key learning that I am achieving the by using D3.js to proactively update SVG dimensions on window resize events, then changing scales and redrawing all chart components accordingly.