COS30045-Data Visualization
August2025

Project Title:

Enforcing Fairness: Visualizing Road Safety Enforcement accross Australia (2008-2024)

Website Link (Mercury hosted):

Group Name: Group 48

Group members:

- Chan Trung Dinh – 105060533
- Lim Zhi Xuan – 105728462
- Chia Choon Yu – 105951129

Semester & Year: August 2025

Wordcount:

Table of Contents

1. Introduction

1.1 Background and Motivation

This project focuses on the enforcement of road safety laws across Australia between 2008 and 2024, using data provided by the Bureau of Infrastructure and Transport Research Economics (BITRE) Road Safety Enforcement dataset. The visualization aims to explore whether enforcement activities such as fines, charges, and arrests are applied fairly across different states and territories over time.

The motivation behind this topic arises from the importance of transparency and accountability in road safety enforcement. Understanding whether variations in enforcement are due to differences in population size, road conditions, or stricter policing practices can help ensure that road safety policies are equitable and evident.

The target users of this visualization include policymakers, researchers, journalist, and members of the public. Policymakers can use the visualization to evaluate the fairness and effectiveness of enforcement efforts across jurisdictions. Researchers may find it valuable for studying long-term trends, demographic patterns, and enforcement intensity. Journalists and the general public can gain insight into when, where, and how road safety enforcement occurs, which fosters public understanding and trust.

Ultimately, the project contributes to promoting transparency in road safety enforcement and supporting data-driven decision-making in policy development.

1.2 Visualization Purpose

The primary purpose of this visualization is to reveal how road safety enforcement varies by time, jurisdiction, and demographic group. It enables fair comparisons between states and territories by normalizing penalties per 10,000 driver licences.

Through this visualization, users will be able to explore several key questions:

- How has road safety enforcement changed over time?

- Which jurisdictions enforce most strictly?

- Which offence types lead to the highest penalties?

- Which detection methods are most common?

- How do penalties differ among age groups?

By addressing these questions, the visualization help users identify patterns, detect potential enforcement disparities, and gain a deeper understanding of how enforcement practices evolve across Australia. The insights gained can support more equitable, transparent, and effective road safety policies.

1.3 Project Scheduale

Week	Task	Expected Outcome
Week 8	Team formation Project brief discussion	Define team roles Understand project requirement
Week 9	Create GitHub team Draft introduction for Design Book Collect data Process data in KNIME	Understand how to work together via GitHub team Complete draft of Background and Purpose section. Collect BITRE and driver licences. Create initial KNIME workflow
Week 10	Draft the storyboard for the website Plan how many visualizations must have to tell the story for project website Second draft for Design Book	Think the project must have how many website and each website to tell a specific story. Think about which visualization must have for each webpage. Complete the Data Source.
Week 11	Draw visualizations Third draft for Design Book	Complete the static visualization for each webpage. Complete the Requirement and Visualization Design for Design Book
Week 12	Add interactions to visualization Finalize the Design Book	Add filter, tooltip, and the interactions to all visualizations. Complete the remaining parts of Design Book.

2. Data

2.1 Data Source

This project uses two datasets to analyze the enforcement of road safety laws across Australian jurisdictions.

The primary dataset is Police Enforcement Fines (2008-2024), which is obtained from the Bureau of Infrastructure and Transport Research Economics (BITRE)data catalogue (BITRE, 2024a)

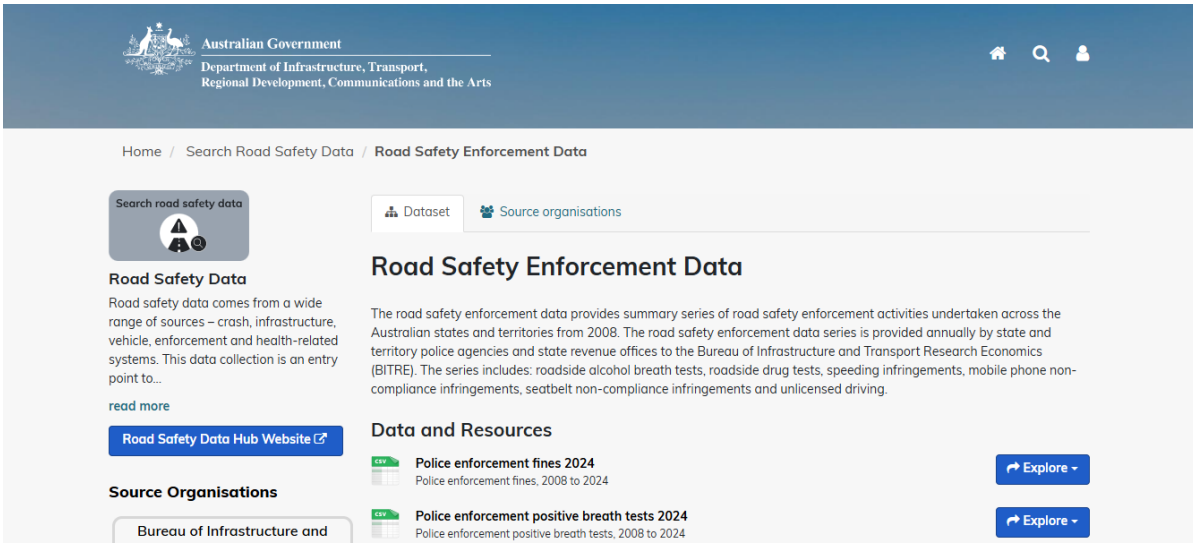


Figure 1: Website data source for Police Enforce Fines

This dataset was downloaded in CSV format and contains 11 attributes: year, start date, end date, location, jurisdiction, age group, metric, detection method, fines, arrests, and charges. It is a tabular, time-series dataset, describing the number of penalties (fines, charges, and arrests) recorded each year across Australian states and territories.

A second dataset was manually compiled to include the total number of driver licences by state and year from 2010 to 2024. The data were sourced from a Police Enforcement Dashboard (BITRE, 2024b). This dataset contains three columns: year, jurisdiction, and licences.

Some attribute from the primary dataset were excluded due to incomplete or missing values. Specifically, start date, end date, and location were removed because they contained significant gaps and would overcomplicate the visualization without adding meaningful insights. Excluding these fields ensures a cleaner, more balanced dataset suitable for comparative analysis across jurisdictions.

2.2 Data Processing

All data cleaning and transformation were performed using a KNIME Analytics Platform workflow.

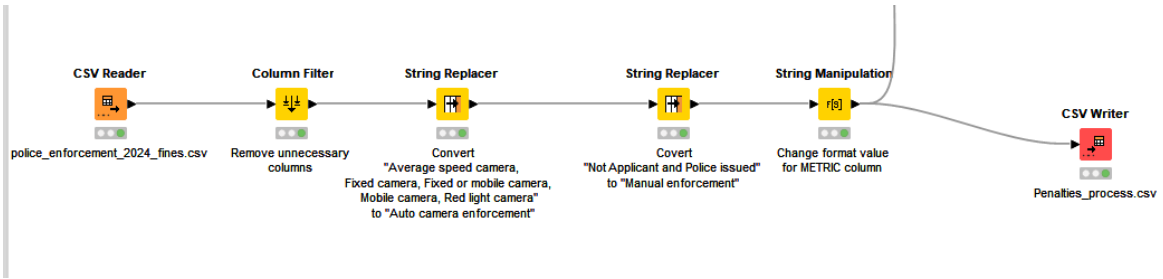


Figure 2: Configuration for data processing of Police Enforcement Fines dataset

The main dataset used, “Police_enforcement_2024_fines.csv”, was preprocessed through a series of KNIME nodes to remove redundant columns, standardize categorical values, and produce a refined dataset suitable for analysis and visualization.

Step 1: Remove unnecessary columns

A Column Filter node was used to remove irrelevant attributes such at Start Date, End Date, and Location, as these fields contained large portion of missing or inconsistent values that were not required for analysis. This step helped simplify the dataset and improve processing efficiency.

Step 2: Standardizing detection method categories

Two String Replacer nodes were applied to unify detection method values:

- The first replacer converted “Average speed camera”, “Fixed camera”, “Fixed or mobile camera”, “Mobile camera”, and “Red light camera” into “Auto camera enforcement”.
- The second replacer converted “Not applicable” and “Police issued” into the category “Manual enforcement”.

This categorization ensures clearer comparison between automated and manual enforcement methods across jurisdictions.

Step 3: Formatting metric names

A String Manipulation node was then used to format the Metric column values by removing underscores and capitalizing each word using the following transformation: capitalize(replace($METRIC$, "_", " "))

This made metric names more readable and presentation-ready for visualization (e.g. mobile_phone_use to Mobile Phone Use).

Step 4: Exporting processed data

Finally, a CSV Write node exported the cleaned and formatted data into a new file named “Penalties_process.csv”.

This processed dataset serves as the standardized input for all subsequent stages of data integration and visualization in the project.

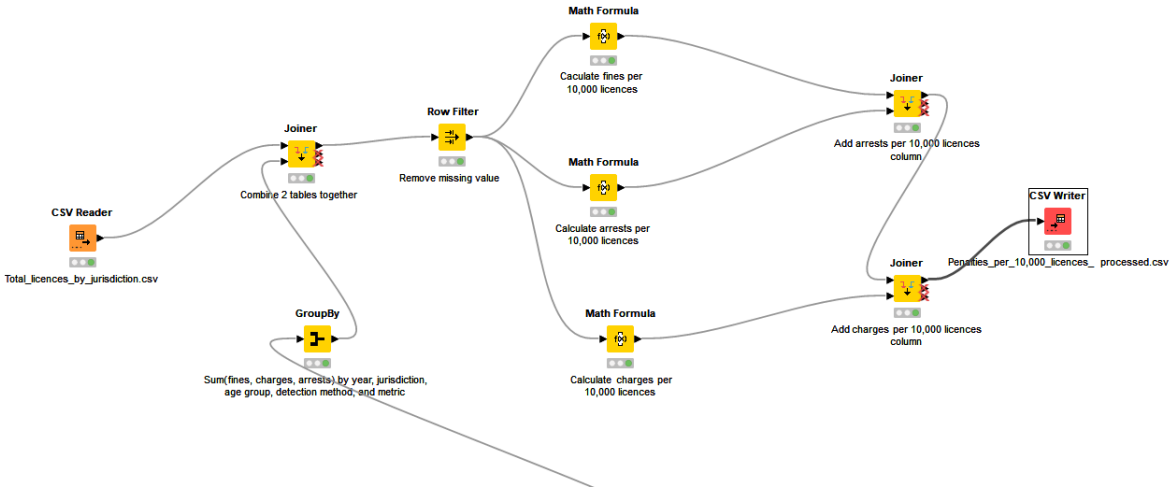


Figure 3: Configuration for data processing of Total Licences dataset

A second dataset, “Total_licensed_by_jurisdction.csv”, was manually compiled and contains three key attributes: Year, Jurisdiction, and Licences. This dataset represents the total number of driver licences across Australian jurisdictions between 2010 and 2024.

Step 1: Aggregating penalty data

Before joining both datasets, the Penalties_process.csv file was aggregated using a GroupBy node to compute the total sum of Fines, Charges, and Arrests by the following fields: Year, Jurisdiction, Age Group, Detection Method, and Metric.

This ensures that the enforcement dataset was grouped to the same structural level as the licence dataset, making them compatible for integration

Step 2: Joining licence data

A Joiner node was then used to merge Total_licensed_by_jurisdiction.csv with the above grouped table.

Because the licence dataset only contains data from 2010 onward, missing values occurred for 2008 and 2009. These incomplete records were removed using a Row Filter node to maintain accuracy and consistency across all years used in visualization.

Step 3: Calculating penalties per 10,000 licences

Three Math Formula nodes were applied to compute normalized values, Fines, Charges, and Arrests per 10,000 driver licences, using the following formula:

- Round(($Sum(FINES)$ / $LICENCES$) * 10,000)
- Round(($Sum(CHARGES)$ / $LICENCES$) * 10,000)
- Round(($Sum(ARRESTS)$ / $LICENCES$) * 10,000)

This calculation allows fair comparisons between jurisdictions with different population sizes.

Step 4: Combining normalized datasets

Two additional Joiner nodes were then used to merge the calculated outputs into a single, consolidated dataset. The joining criteria were Year, Jurisdiction, Licences, and Metric, ensuring that each record contained both total and normalized penalty values.

Step 5: Exporting the final dataset

Finally, a CSV Writer node was used to export the processed dataset as:

“Penalties_per_10,000_licences_processed.csv.”
This file serves as the main input for the fairness comparison visualization, enabling population-adjusted analysis of road safety enforcement across Australia.

3.Requirement

3.1 Must-Have Features

The must-have features represent the essential components required for the successful implementation of this data visualization project. They ensure that users can effectively explore and compare road safety enforcement activities across Australian jurisdictions from 2008 to 2024.

A key requirement of this project is the interactive filtering system implemented within each visualization page. It enables users to explore specific aspects of the dataset based on their interests, such as offence type, detection method, or age group. This feature empowers users to answer a wider range of analytical questions and gain detailed insights into the specific data groups they are most interested in.

Equally important are the interactive tooltip and hover effects, which dynamically display relevant information as users interact with different data points. This allows users to simultaneously observe overall patterns and access exact numerical details, improving both data comprehension and engagement.

Another significant feature is clear legends, descriptive labels, and properly annotated axes, which help users understand exactly what each visualization represents. This ensures accuracy and interpretability, reducing confusion when comparing different data dimensions.

Lastly, consistent and well-contrasted color schemes are also crucial requirement. The design uses clear and uniform color encoding to distinguish between categories and intensities of enforcement. High contrast enhances readability and accessibility, which ensures that visual differences are immediately perceivable across all charts and maps.

Together, these core features establish a visualization system that functional, user-friendly, and analytically meaningful. Without these elements, the project would fail to provide clarity, interactivity, or insight.

3.2 Optional Features

While the must-have features ensure that the visualization performs its analytical function, the optional features are designed to enhance user experience and evaluate the overall presentation quality.

A responsive layout would allow users to access the visualization seamlessly across different devices and screen sizes which ensures consistent usability and visual quality.

Animated transition would make the visualization smoother and more visually engaging, helping users better perceive changes when switching between filters or time periods.

A dark and light mode toggle would provide customization for different viewing preferences and improve accessibility for users with visual sensitivities.

Lastly, a narrative or storytelling view could guide users through the key insights and relationships hidden within the dataset. Beyond displaying graphs or maps, this feature would allow the visualization to communicate a coherent data story, helping users understand not just what the data shows, but why those patterns matter.

These optional features are not essential for the functionality of the project, but they would significantly enhance its overall usability, professionalism, and communication value.

4.Visualization Design

4.1 Week 10 (27 – 31 Novemeber 2025)

At the initial stage of the project, our team created several sketches to explore potential visualization types suitable for our selected datasets. These sketches served as the foundation for our website design, allowing us to experiment with layout, interaction ideas, and how each visualization could best represent road safety enforcement data across Australia.

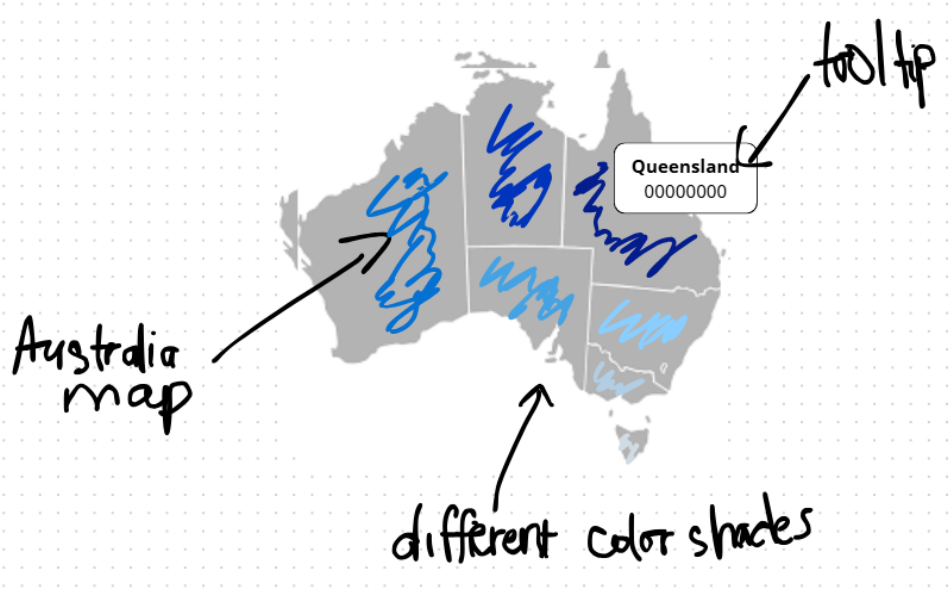


Figure 4: Sketch of Heat Map for National Trend

This sketch represents a heat map of Australia displaying the number of penalties by jurisdictions. Each state or territory is shaded in different color intensities, with darker colors indicating higher numbers of penalties. A tooltip appears when hovering over a region, showing the jurisdiction name and exact value.

This visualization was chosen because it provides users an immediate understanding of regional enforcement and is visually intuitive for comparing jurisdictions.

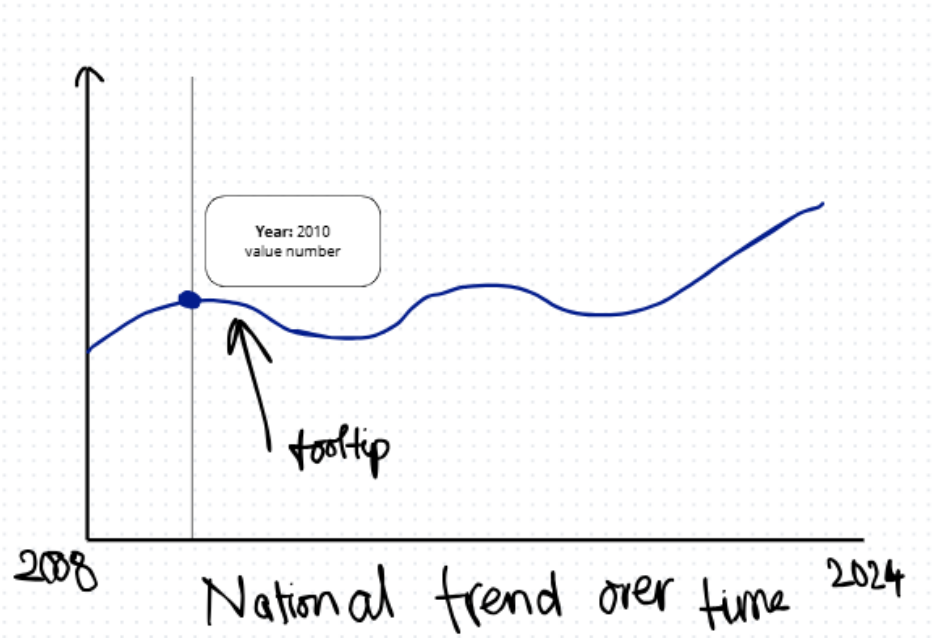


Figure 5: Sketch of National Trend Line Chart

This line chart sketch illustrates the national trend of penalties from 2008 to 2024. When hovering over a data point, a tooltip displays the year and corresponding value. The line chart

helps users observe changes and trends in enforcement activities over time, complementing the heat map by adding a temporal perspective to the national overview.

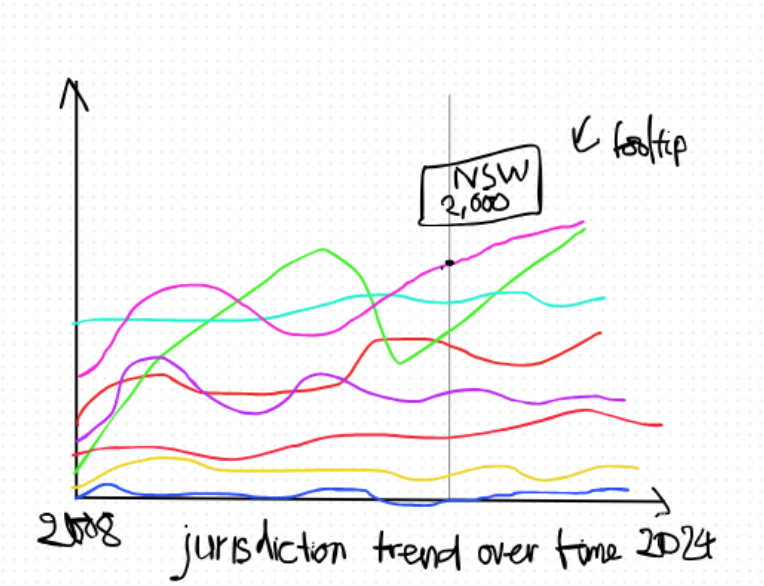


Figure 6: Sketch of Multi-Line Chart

This sketch depicts multi-lines representing penalties across eight Australian jurisdiction from 2008 to 2024. Each line has a unique color to differentiate jurisdictions, and tooltips reveal the state name and penalty count.

The visualization was selected because it enables users to directly compare enforcement trends between states, making it easier to identify variations or patterns across jurisdictions.

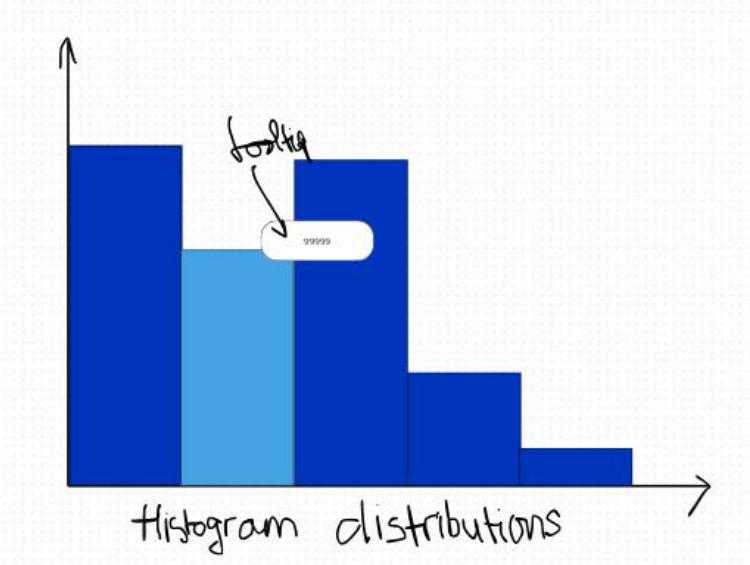


Figure 7: Sketch of Histogram for penalty distribution

This sketch shows a histogram used to visualize the distribution of penalties based on a selected attribute (such as offence type or age group).

Tooltips appear on hover to display the exact value for each bar. The histogram was chosen for its ability to summarize the distribution of enforcement outcomes, helping users understand how penalties are spread across different categories.

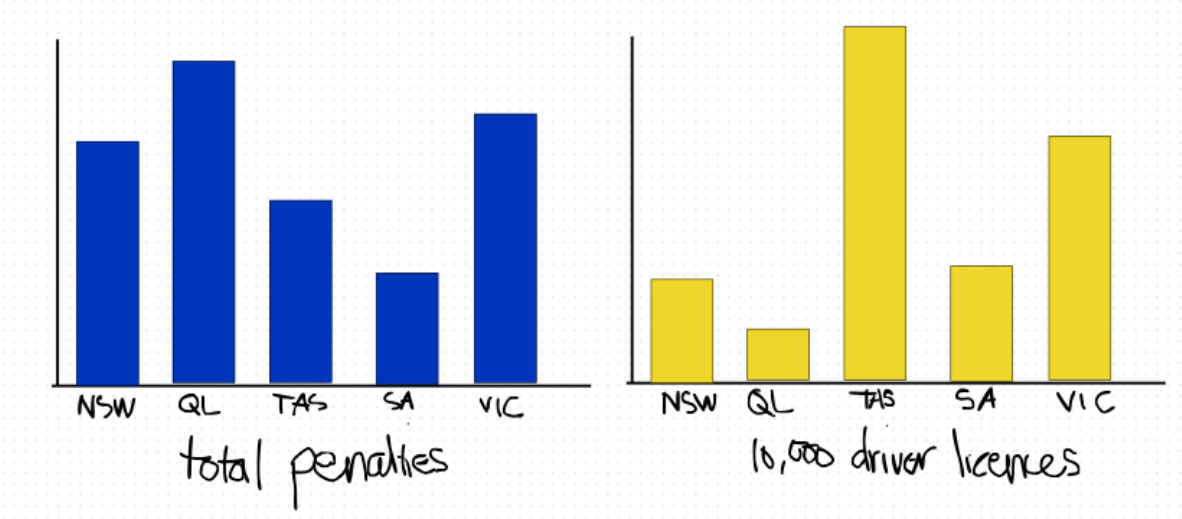


Figure 8: Sketch of Bar Chart for fair comparison

This sketch presents two bar charts. The first compares the total penalties across jurisdictions, while the second adjusts these figures per 10,000 driver licences for fairer comparison. The side-by-side design allows users to easily distinguish between raw totals and population-normalized values, highlighting whether enforcement differences arise from stricter policies or demographic factors.

Overall, these sketches played a crucial role in shaping the visualization strategy for the project. They allowed the team to identify which chart types would best communicate the data's story and provided a clear reference before implementing the visualizations using D3.js.

4.2 Week 11 (3 – 7 November 2025)

At this stage, the data visualization project reached a complete structural prototype with the main layout, navigation components, and defined the storyboard. The website consists of 5 main pages, including Home page, three visualization pages (National Overview, State Trends, and Fairness Comparison), and an About page. The Index page is intended to introduce the project's background and purpose, while the About page will later include group information, data governance statements, AI declaration, and limitation notes.

The design during Week 11 focused primarily on implementing the common layout and came up with ideas for drawing which type of visualization is suitable for 3 visualization pages. Both header and footer sections were finalized during this week. The navigation bar appears consistently across all pages, with the team name "DV Group 48" displayed at the top-left corner, and menu links to the other side. The navigation bar is fully responsive, adapting smoothly to different screen sizes and device types.

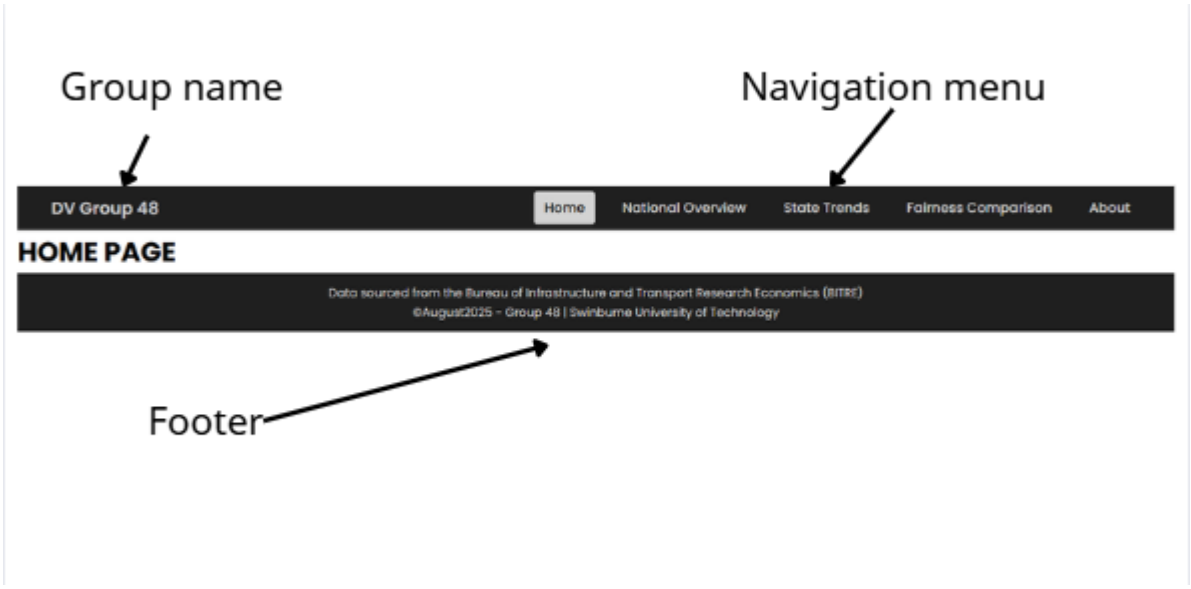


Figure 9: Website navigation bar with header and footer design

The National Overview page serves as the starting point for data exploration, providing a broad insight into enforcement patterns across all Australian jurisdictions. This page integrates both spatial and temporal visualizations to communicate nationwide enforcement intensity and long-term trends effectively.

Figure 10: National Overview Map

The map of Australia displays the eight jurisdictions, with color gradients representing the number of penalties recorded in each state or territory. Darker shades indicate the higher enforcement counts, while lighter shades represent lower counts. Users can interact with the map through zoom and drag actions, as well as reset the map to its default state using a reset

button. This interactive choropleth map design allows viewers to visually identify regions with higher enforcement activity quickly, supporting spatial patterns recognition.

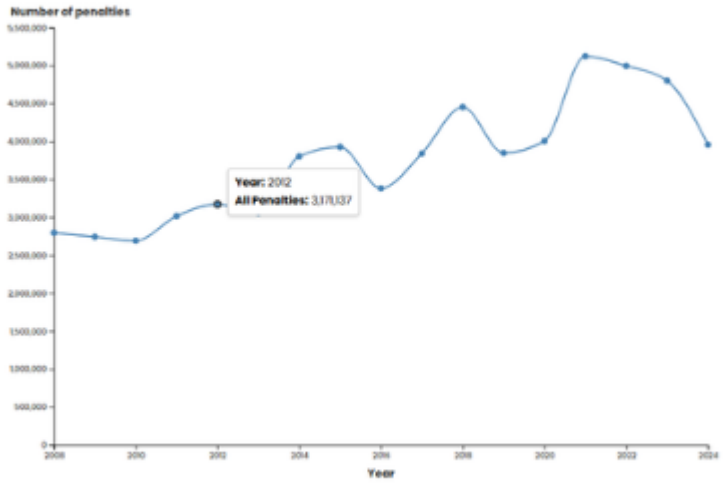


Figure 11: National Line Chart

Below the map, a line chart illustrates national enforcement trends from 2008 to 2024. The line chart captures a fluctuation in penalties over time, giving users an overview of how enforcement levels have evolved. A dynamic tooltip appears when hovering over a point on the line, displaying the exact year and corresponding penalty value. This interaction makes it easier for users to observe both the overall trend and the precise quantitative data simultaneously.

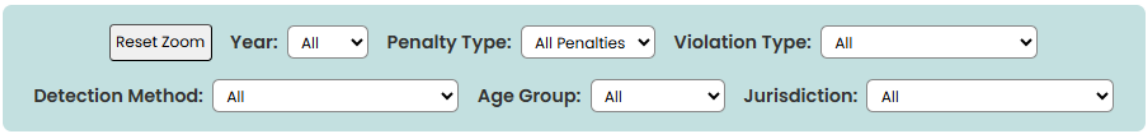


Figure 12: Filter Panel for National Overview

A comprehensive filtering system has been developed to enable deeper data exploration. Users can refine the visualization by combining multiple filters, including year, jurisdiction, penalty type, violation type, detection method, and age group. The filter is implemented to update both the map and the line chart dynamically. This feature enhances user engagement and supports customized analysis across different demographic and temporal categories.

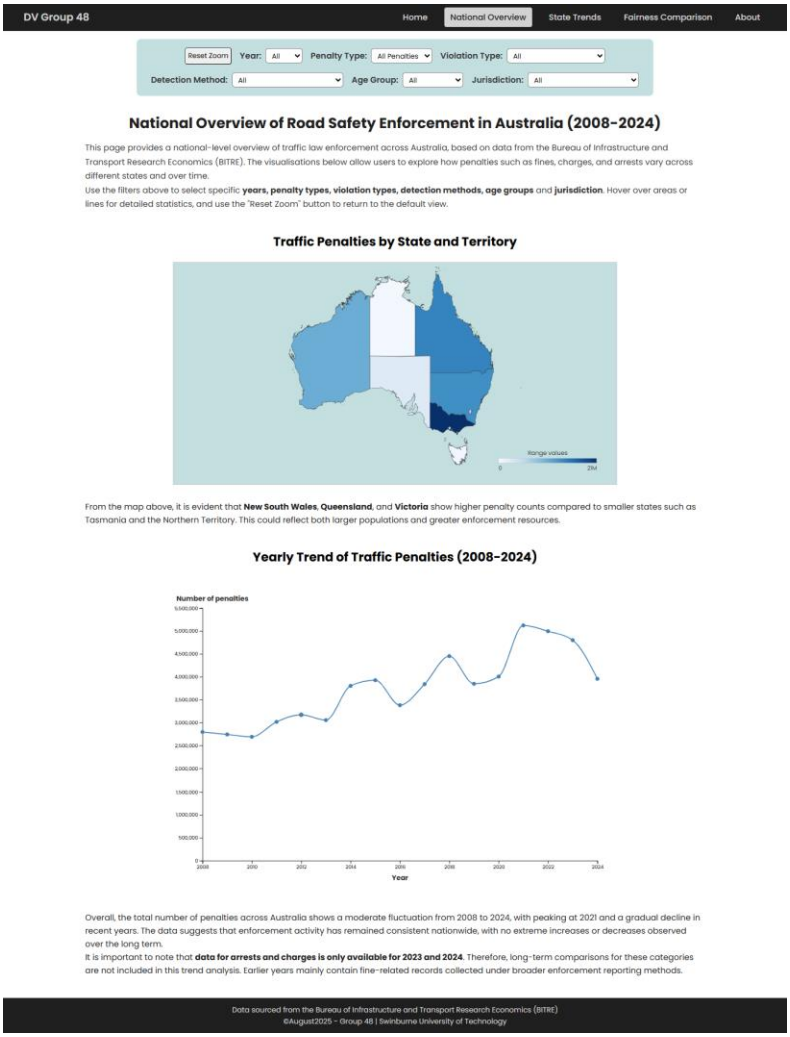


Figure 13: Prototype for National Overview Page in Week 11

Another component developed during week 11 was the State Trends visualization page. This page was designed to allow users to explore enforcement trends across all eight Australian jurisdictions from 2008 to 2024.

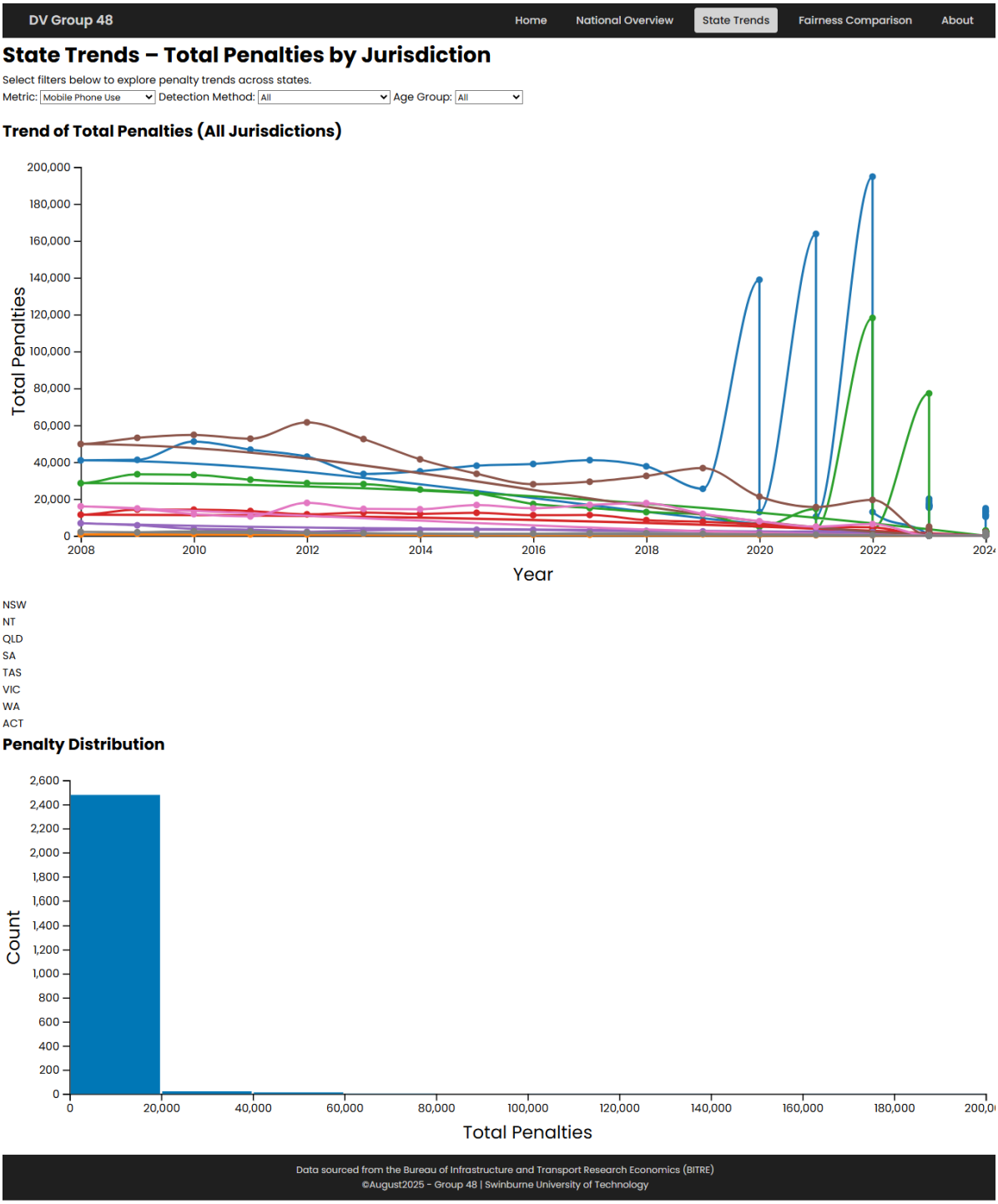


Figure 14: State Trends page

The page currently includes two main visualizations. The first is a multi-line chart, which displays penalty values for each jurisdiction over time. Each line represents a different state or territory, providing a direct comparison of enforcement activity across Australia. Although the visualization appears visually cluttered at this stage due to overlapping data points, it serves as an initial functional prototype. It effectively demonstrates how the design will support multi-state trend analysis once line clarity and colour encoding are refined in later iterations.

Below the line chart, a histogram visualizes the overall penalty distribution across different jurisdictions, offering an immediate sense of how penalty counts vary in scale. Together, these charts provide both temporal and distributional perspectives of enforcement data.

Interactive filter controls were also introduced for this page, allowing users to refine their analysis by metric, detection method, and age group.

Future improvements will target clearer data encoding, optimized colour palettes, and enhanced filtering responsiveness to make the visualization more accessible and informative.

Lastly, the Fairness Comparison page was also developed during Week 11 to highlight differences in enforcement intensity across jurisdictions. This visualization aims to provide an equitable comparison between Australian states by representing both absolute and population-adjusted penalty data.

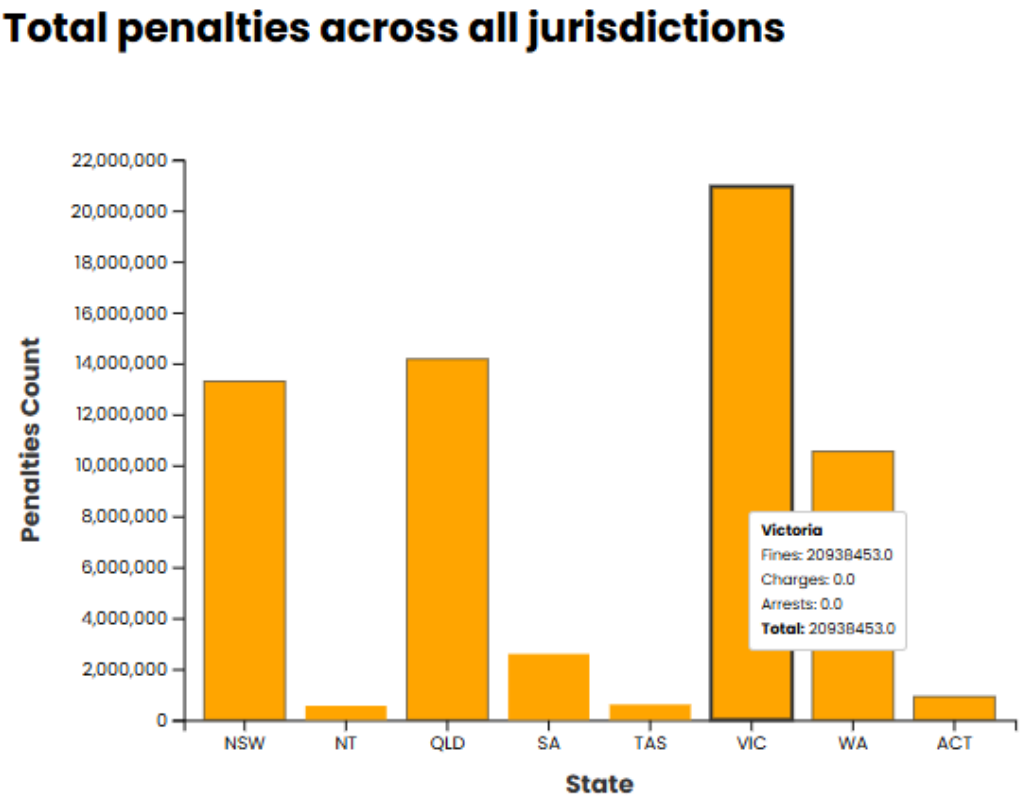


Figure 15: Total penalties by Jurisdiction

The first bar chart displays the total number of penalties issued across all eight jurisdictions. Each bar represents one state or territory, enabling users to visually compare the magnitude of enforcement activity nationwide. A tooltip feature has been added so that when users hover over each bar, the exact penalty value and corresponding jurisdiction are displayed.

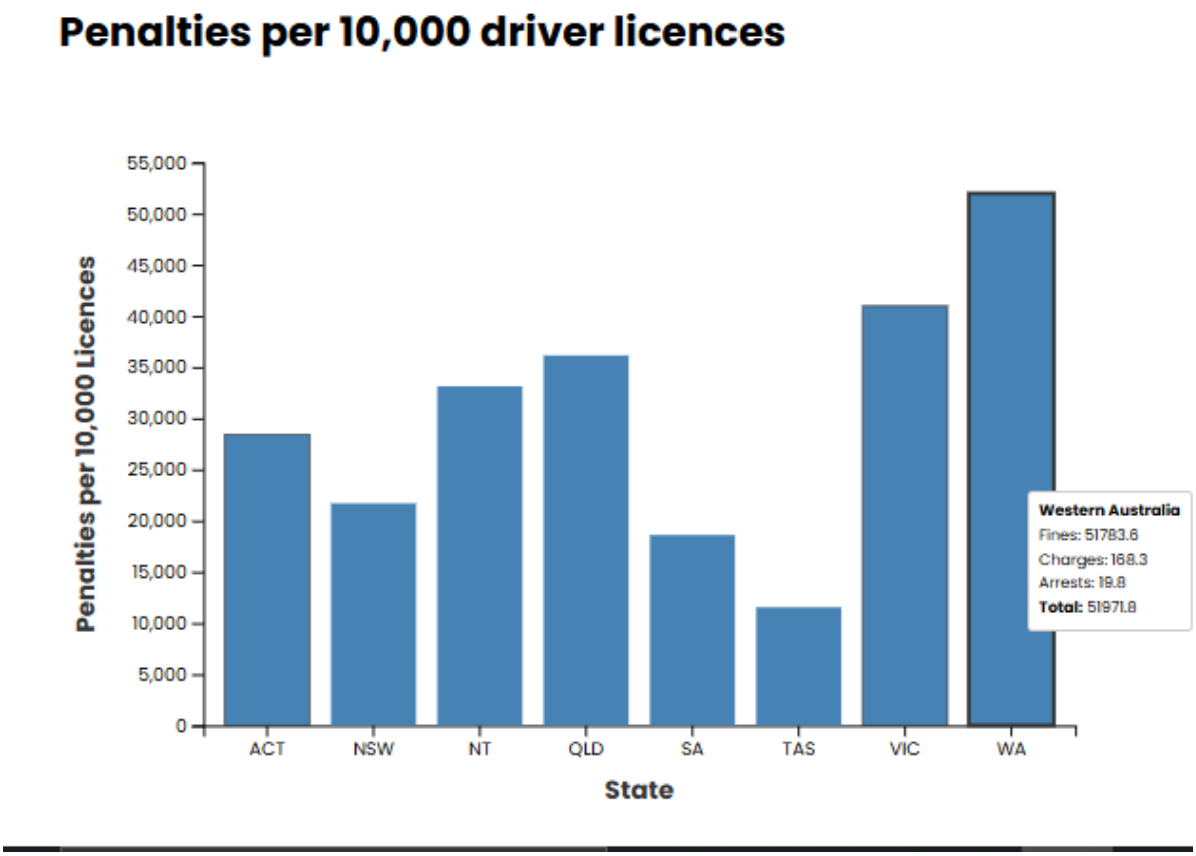


Figure 16: Penalties per 10,000 Driver Licences

The second chart provides a normalized comparison, illustrating penalties per 10,000 driver licences. This approach adjusts for population size and allows users to evaluate whether enforcement differences are due to stricter policing or simply to larger driving populations. Both bar charts use a consistent colour palette to maintain readability and visual harmony with the other pages.

A filter panel was also introduced, allowing users to refine the visualization by year, violation type, detection method, and age group. While the filtering logic is not yet functional, its structure is in place to support future interactivity. Once completed, this system will enable users to view how fairness trends shift across specific demographic and offence categories.

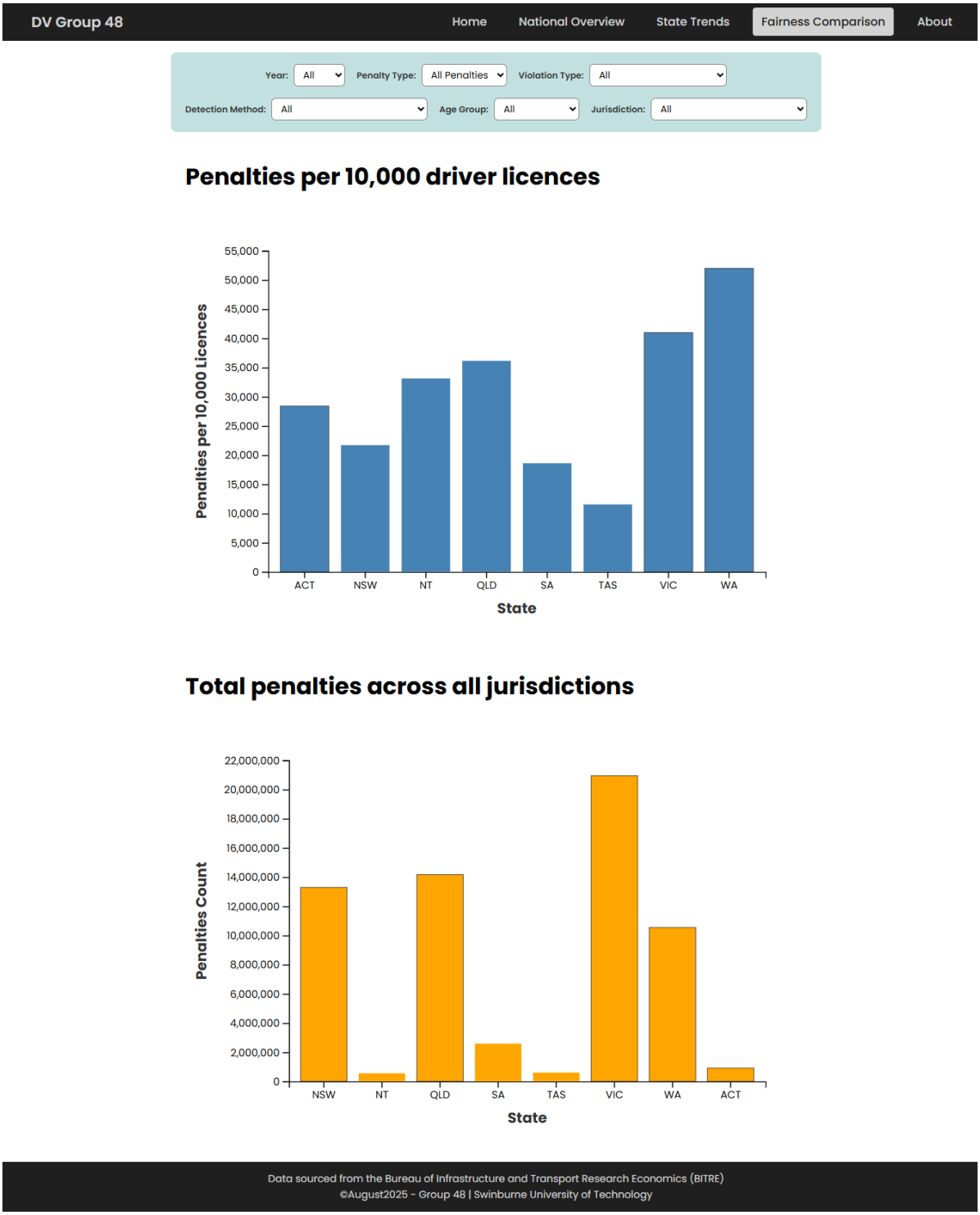


Figure 17: Prototype for Fairness Comparion page in Week 11

This early prototype serves as clear evidence of progress in implementing the fairness-focused analytical view of the project. Future work will focus on enhancing interactivity, refining axis labels, and optimizing colour contrast to ensure that the normalized and total comparisons can be interpreted quickly and intuitively.

4.3 Week 12 (10 – 14 Novemeber 2025)

This week focused on refining the layout, improving interaction design, completing two remaining webpage (Home and About pages), and implementing all must-have features across all three main visualization pages. The refinements were made based on insight from 50% usability test as well as internal feedback within the group.

The National Overview page underwent a layout redesign to improve clarity, spacing, and user flow. The updated layout introduces:

- Increased spacing between the heat map and line chart.
- A cleaner alignment between the narrative text and visual components.
- A more prominent positioning of the filter area (sticking on the page when scrolling down).
- Clear typographic hierarchy for titles and section labels.

These improvements allow users to interpret the map the national trend line more easily without visual crowding.

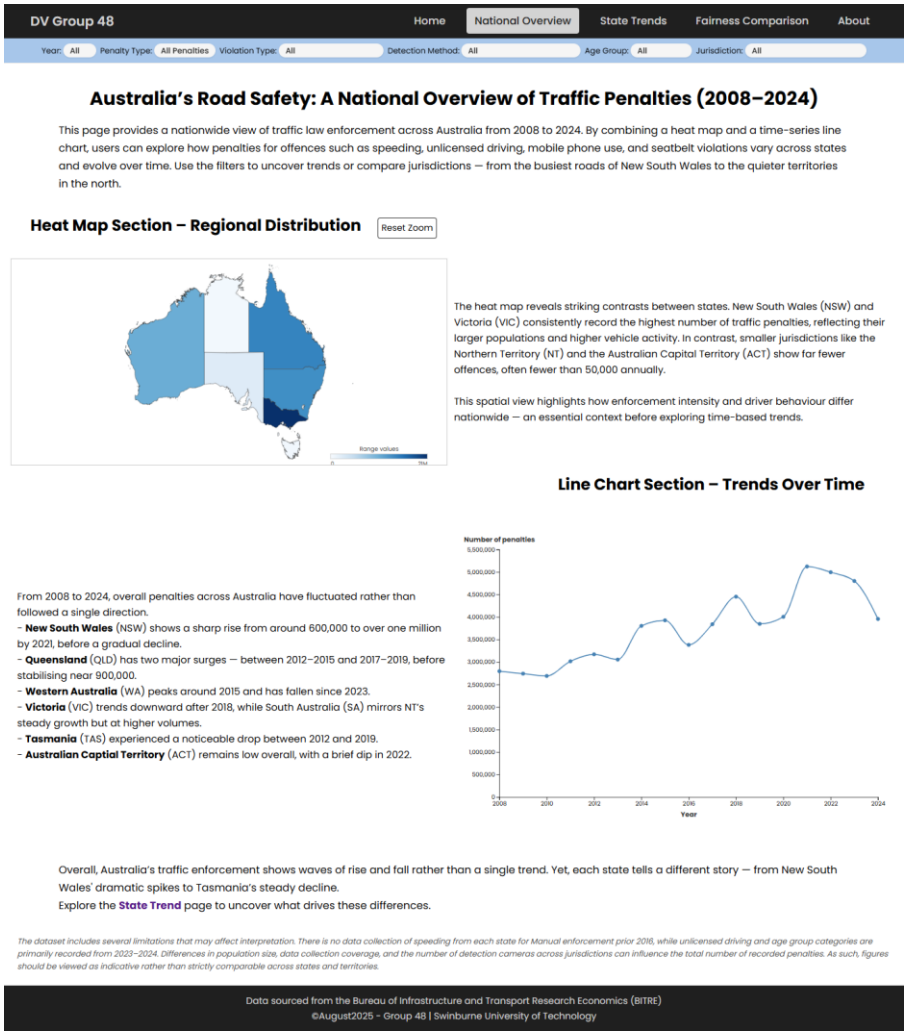


Figure 18: Update National Overview Page Layout

For State Trends page, significant improvements were made, especially targeting interaction issues identified during usability testing.

The previous multi-line chart suffered from severe line overlap, making it difficult to distinguish between jurisdictions. This week, the team implemented:

- Distinct color assignments for each jurisdiction.
- A hover tooltip overlap that displays exact values for the selected year.
- A vertical guideline that follows the mouse position.

These enhancements greatly improved readability and user interaction.

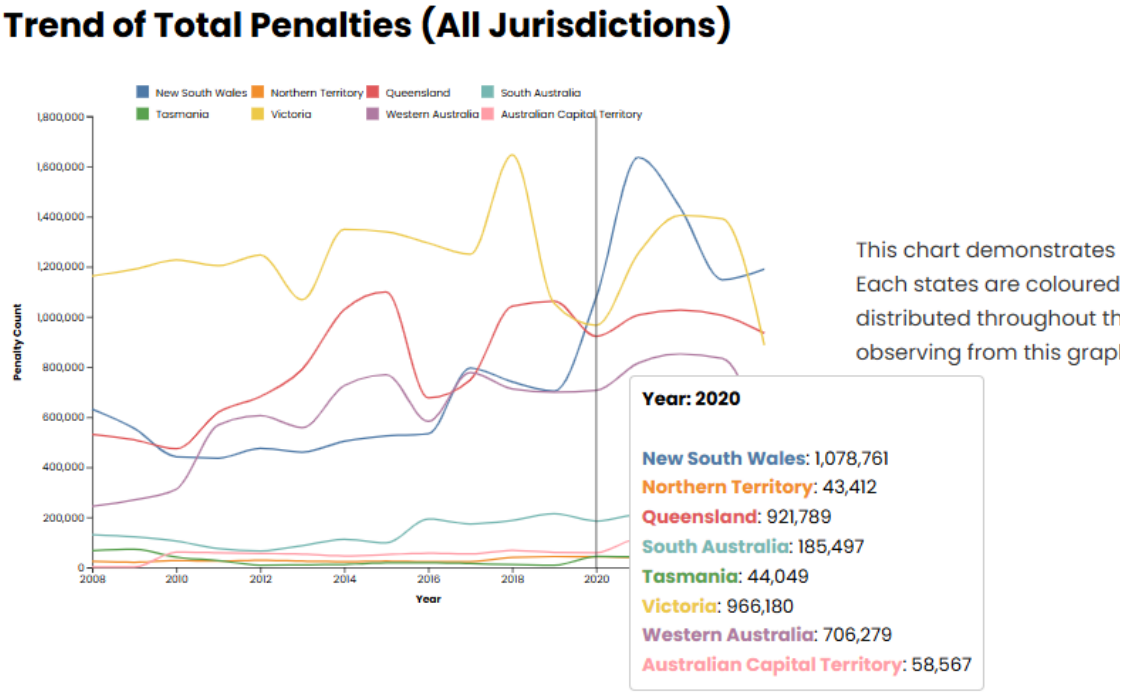


Figure 19: Improved Multi-line chart with tooltip overlay

The histogram, previously lacking clarity, was updated with:

- Correct bin-based tooltip behavior.
- Improved hover interaction, highlighting only the selected bin.
- Clearer axis labels and a more consistent color palette.

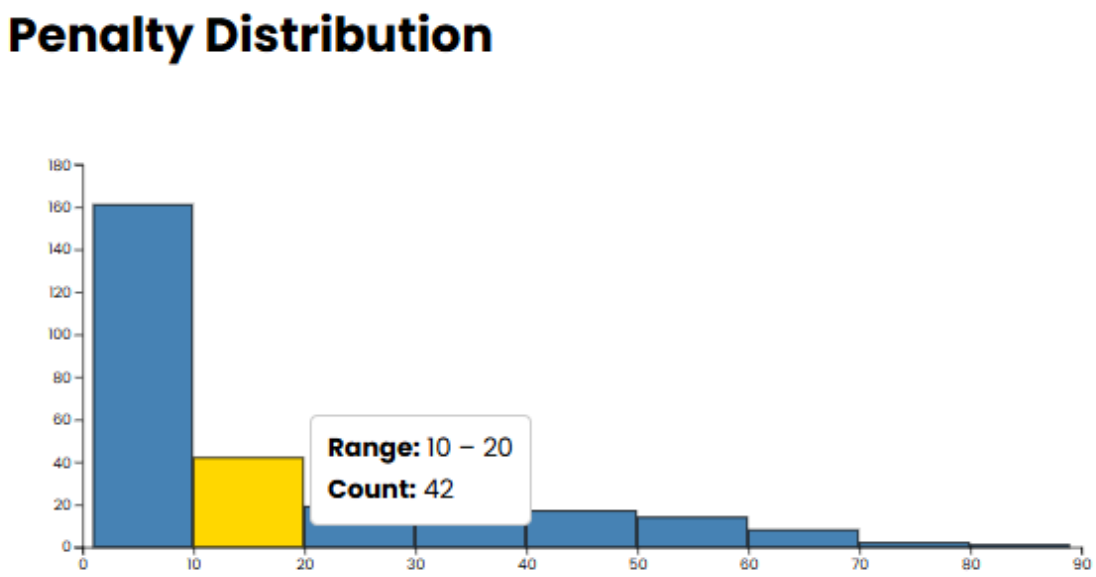
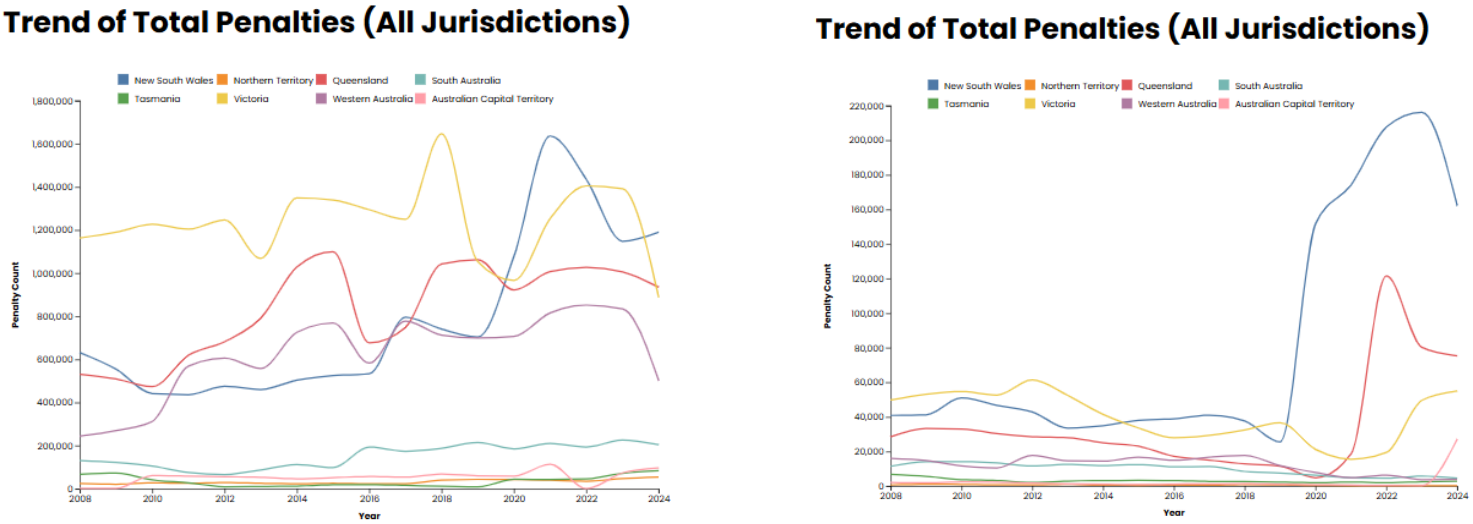


Figure 20: Updated Histogram with Clear bin hover

One of the key improvements for State Trends page was making the filter panel truly dynamic:

- When filtering, the y-axis scales recalibrate automatically.
- This ensures that trends and distribution differences become immediately noticeable without users having to interpret flattened graphs.



y-axis: 0 – 1,800,000 Re-scale: 0 – 220,000

Figure 21: Dynamic Scale Update after filtering

A narrative paragraph was added into the webpage to help users understand:

- What the visualization shows.
- Why the pattern matters.
- What insights they can derive.

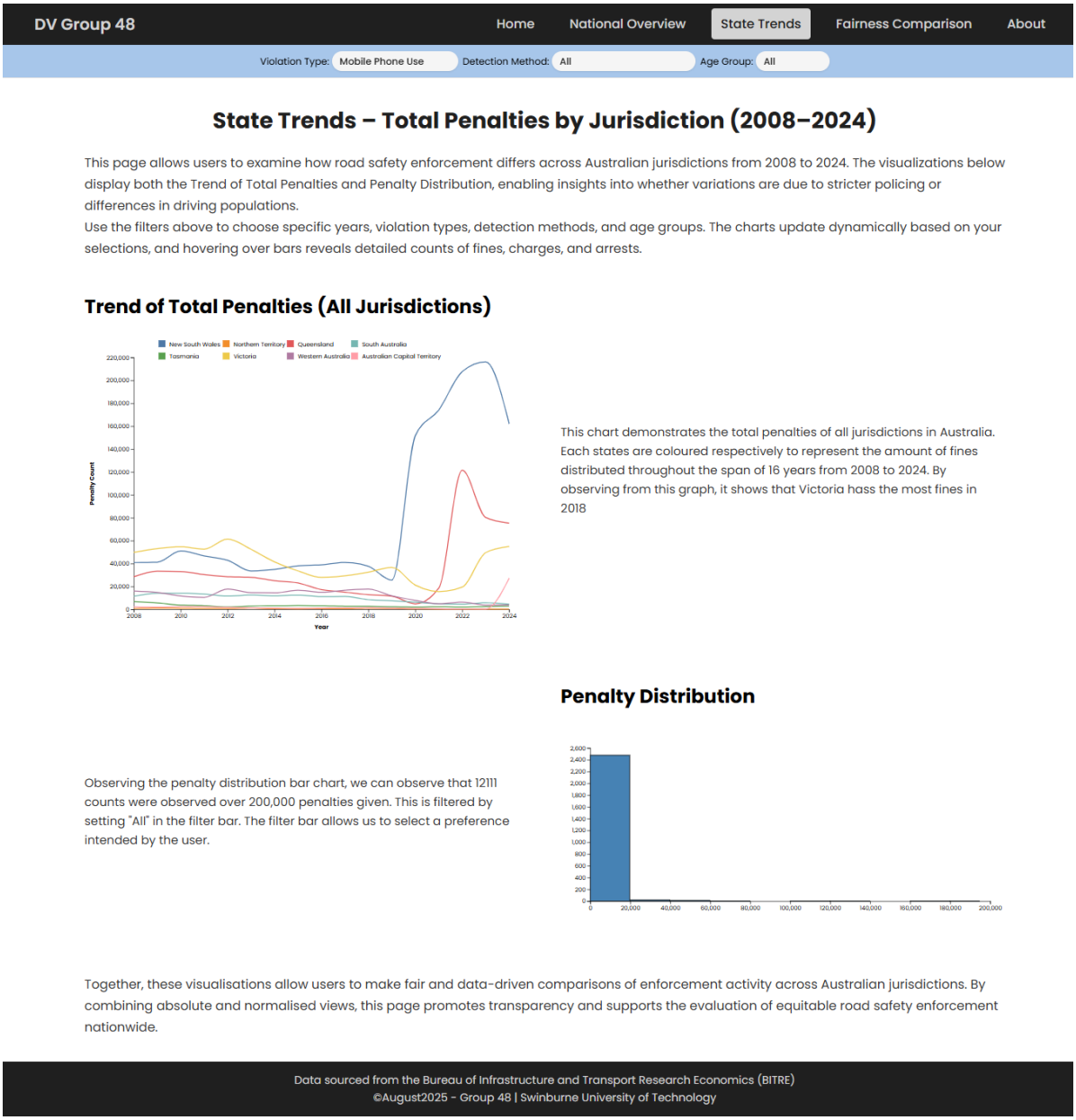


Figure 22: Revised Pape Layout with Narrative Text

The Fairness Comparison page also received several improvements based on usability test.

It was also refined the layout with adding narrative text, making it more linking to the visualizations displayed on the webpage.

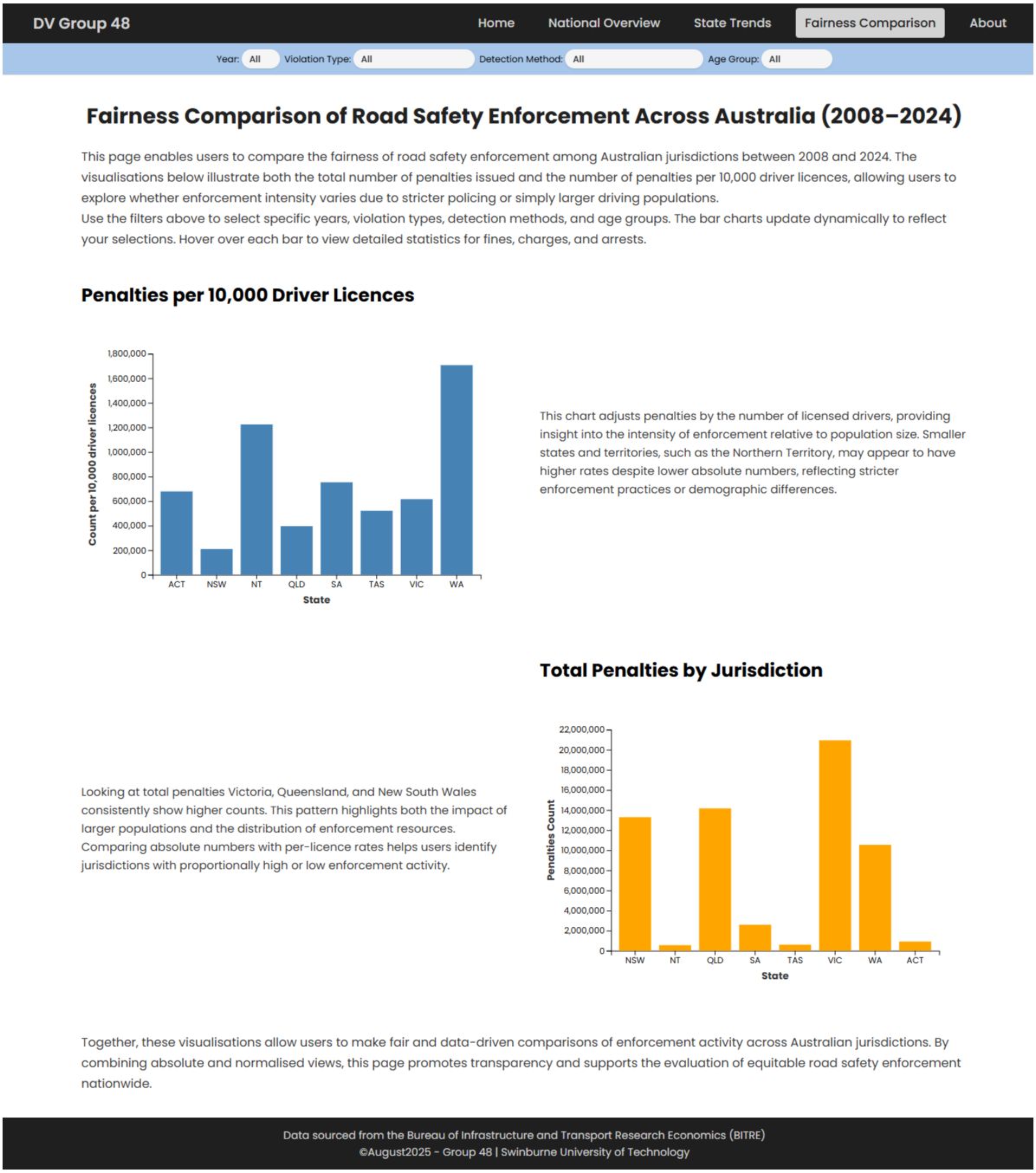


Figure 23: Update Fairness Comparison Layout

Similar to the State Trends page, the filter of this page was also enhanced more dynamically with flexibility of y-axis when filtering, which highlight state differences more clearly.

y-axis: 0-1,800,000 Re-scale: 0-3,500

Figure 24: Dynamic Scale Updater after filtering

The Home page was fully implemented in this week with 3 main sections:

- Background & Motivation: describing road safety enforcement context.
- Purpose & Key Questions: outlining the analytical goals of the project.
- Visualization Overview: introducing the three main pages and what insights they provide.

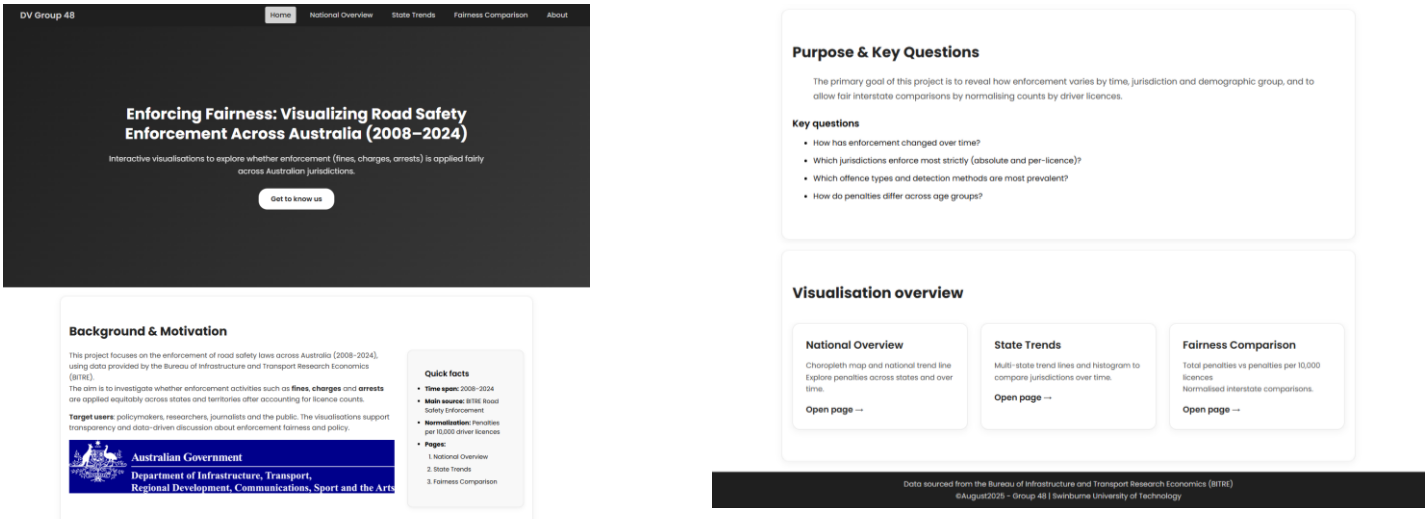


Figure 25: Home Page (Full View)

The About page also was fully completed with the following features:

- Group information and member roles.
- Data governance documentation.
- Academic declaration.

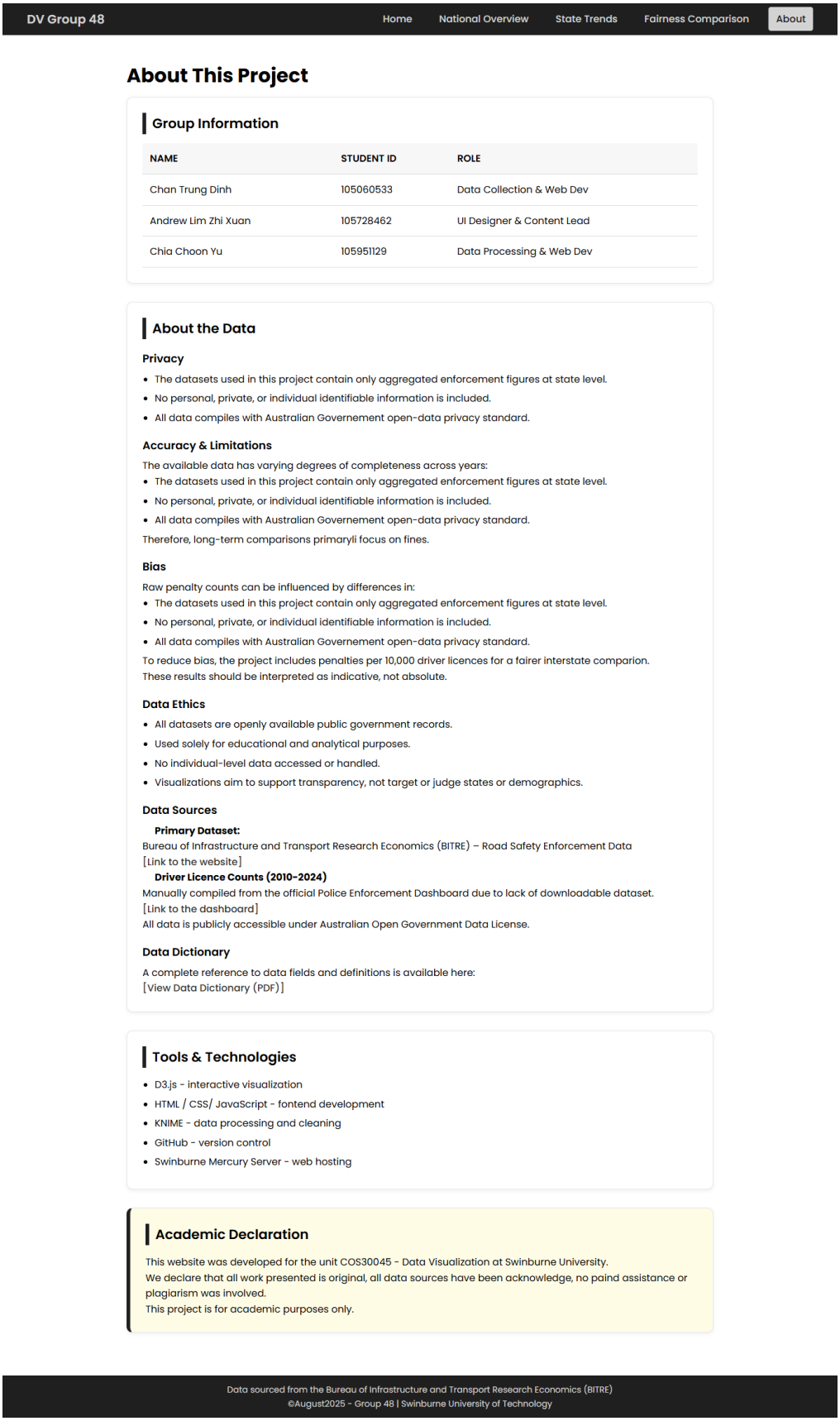


Figure 26: About Page (Full View)

4.4 Week 13 (17 – 21 November 2025)

5.Usability & Validation

This section presents the usability testing and validation processes conducted at different stages of the project. Validation activities were split into two phases:

- Prototype Usability Testing at 50% completion.

- Final Usability Testing at 100% completion.

5.1 Usability Testing (50% completion)

At the mid-project milestone, the group conducted a formative usability test to evaluate the usability of the three visualisation pages at their 50% completion stage. The purpose of this test was not to evaluate the correctness of the user's interpretation but to identify interaction issues, readability problems, layout inconsistencies, and early barriers that could affect user experience.

The Usability Test was developed using three separate Microsoft Forms and one for each webpage. Each team member was responsible for assessing the form for the page they did not develop, ensuring neutral evaluation. In each form, the tester received a link to the webpage and was instructed to complete a set of tasks followed by feedback questions regarding the clarity, usability, and completeness of the page.

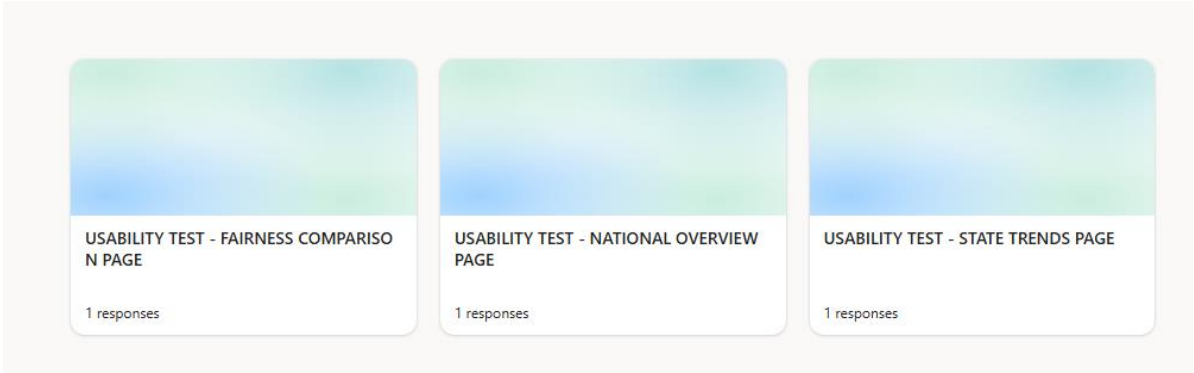


Figure: Interface of the Microsoft Forms usability test

5.1.1 National Overview Page

Evaluator: Chia Choon Yu
Page Developer: Chan Trung Dinh

a/ Task Section:

Task 1: Using the heat map, identify which jurisdiction has the highest penalties in 2024.

1. Using the heat map, identify which jurisdiction has the highest penalties in 2024.

1 Responses

ID ↑	Name	Responses
1	anonymous	New South Wales

Figure: Response for Task 1

Task 2: Hover over two different jurisdictions and describe the information presented in the tooltips.

2. Hover over two different states and describe the tooltip information.

1 Responses

ID ↑	Name	Responses
1	anonymous	When you hover your mouse over Victoria (VIC), the tooltip displays the state name and the total number of fines for the selected year. Hovering your mouse over Queensland (QLD) will display similar information: the state name and the exact number of fines for the current year.

Figure: Response for Task 2

Task 3: Review the national trend line chart and describe how penalties have changed from 2008 to 2024.

3. Look at the national trend line chart. Describe how penalties changed from 2008 to 2024.

1 Responses

ID ↑	Name	Responses
1	anonymous	From 2008 to 2013, the number of fines was relatively even. After 2013, the number of fines increased significantly, followed by large fluctuations until 2020. 2021 saw the highest number of fines. Since 2021, the number of fines has begun to decline.

Figure: Response for Task 3

Task 4: Use the filter panel to select the following:

- **Violation Type:** Speed Fines
- **Detection Method:** Manual Enforcement

- **Age Group:** 17–25

Explain how the heat map and line chart changed based on these selections.

4. Use the filter panel to select: - Violation Type: Speed Fines - Detection Method: Manual enforcemnt - Age Group: 17 - 25 Explain how the heat map and line chart changed.

1 Responses

ID ↑	Name	Responses
1	anonymous	Heatmap.Due to the filtering, the colors changed significantly. The darkest colors shifted to other areas because of the filters. Many areas also became lighter because I narrowed the scope. Line Chart. Because the filters only represent a portion of all penalties, the line chart's lines are lower and more stable than the overall national trend. The lines also became very small.

Figure: Response for Task 4

b/ Feedback Questions Section:

Q1: How easy was it to understand the heat map?

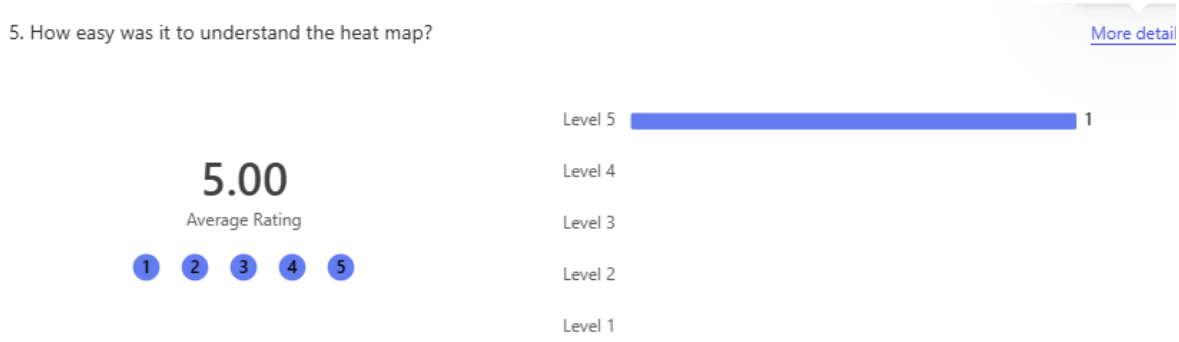


Figure: User rating for Q1

Q2: Was the tooltip information clear?

Figure: User rating for Q2

Q3: Was the national trend line chart easy to understand?

Figure: User rating for Q3

Q4: Were the filters easy to locate and use?

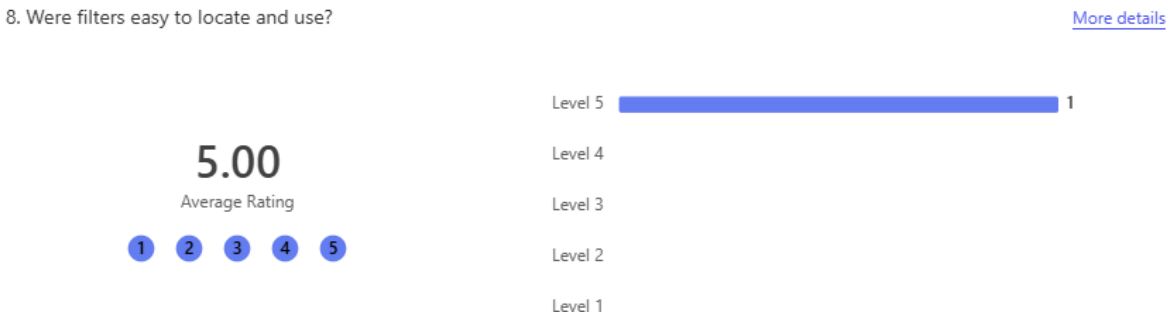


Figure: User rating for Q4

Q5: What confused you on this page?

Figure: Response for Q5

Q6: What improvements do you suggest?

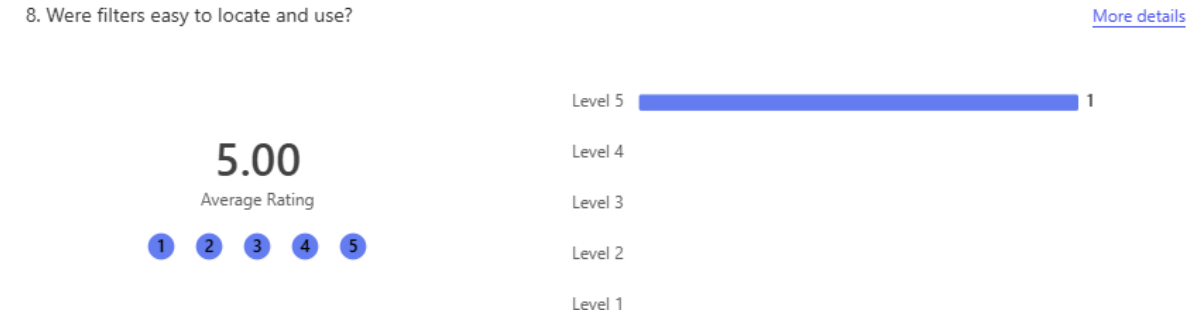


Figure: Response for Q6

c/ Key Findings:

The National Overview page is nearly complete, with all major features functioning correctly. The heat map and national trend line operate smoothly, and filters allow users to explore the data effectively. Colors, labels, and legends are clear and consistent.

However, minor improvements can enhance user experience:

- The layout between the visualisation and descriptive text can be adjusted to create a stronger narrative connection.
- Filter spacing could be improved to reduce visual density.
- Tooltip placement can be refined to prevent overlap on smaller viewports.

5.1.2 State Trends Page

Evaluator: Chan Trung Dinh
Page Developer: Lim Zhi Xuan

a/ Task Section:

Task 1: Describe the penalty trend for VIC (2010–2020) using the multi-line chart.

1. Look at the multi-line chart. Describe the penalty trend for VIC (2010-2020).

1 Responses

ID ↑	Name	Responses
1	anonymous	I cannot define the trend because the lines is overlapping too much. And it seemly doesn't link to each point on the line.

Figure: Response for Task 1

Task 2: Apply the following filters:

- **Violation Type:** Speeding
- **Detection Method:** Auto Camera Enforcement

Describe how the line chart changed.

2. Use the filter to select: Metric: Speeding Detection: Auto camera enforcement Describe what changed.

1 Responses

ID ↑	Name	Responses
1	anonymous	The filter still works efficiently, however, the issue is still the overlapping, making hard to determine between lines.

Figure: Response for Task 2

Task 3: Identify any difficulties when reading the 8-line chart.

3. Identify any difficulties when reading the lines on the chart.

1 Responses

ID ↑	Name	Responses
1	anonymous	It shows overlapping too much, really hard to differentiate them.

Figure: Response for Task 3

Task 4: Review the histogram below the line chart and describe any visible distribution pattern.

4. Look at the histogram below the chart. What distribution or pattern can you observe?

1 Responses

ID ↑	Name	Responses
1	anonymous	This histogram is much better. I can easily understand the penalty distribution with majority in 200,000 and recorded more than 3,000.

Figure: Response for Task 4

b/ Feedback Questions Section:

Q1: How difficult was it to distinguish between the 8 lines?

Figure: User Rating for Q1

Q2: Were the line colors sufficiently different?

Figure: User Rating for Q2

Q3: Did the filter work correctly for the line chart?

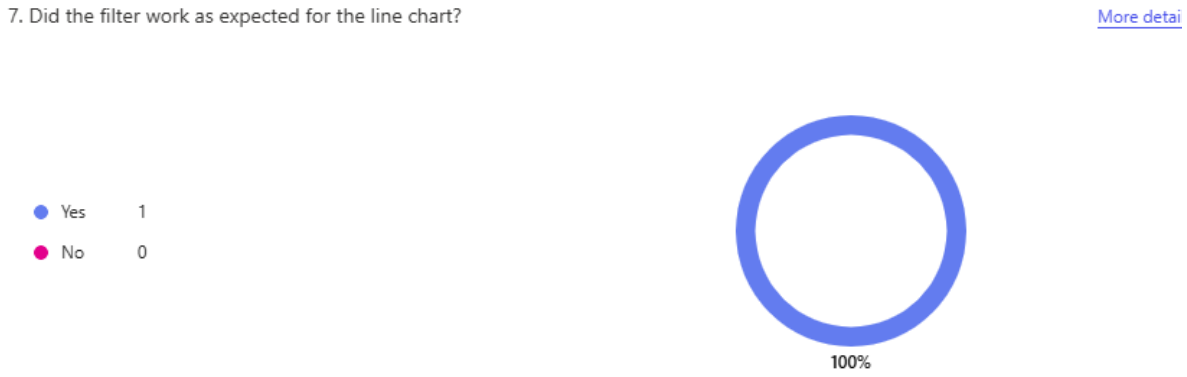


Figure: User Rating for Q3

Q4: Was the histogram easy to understand?

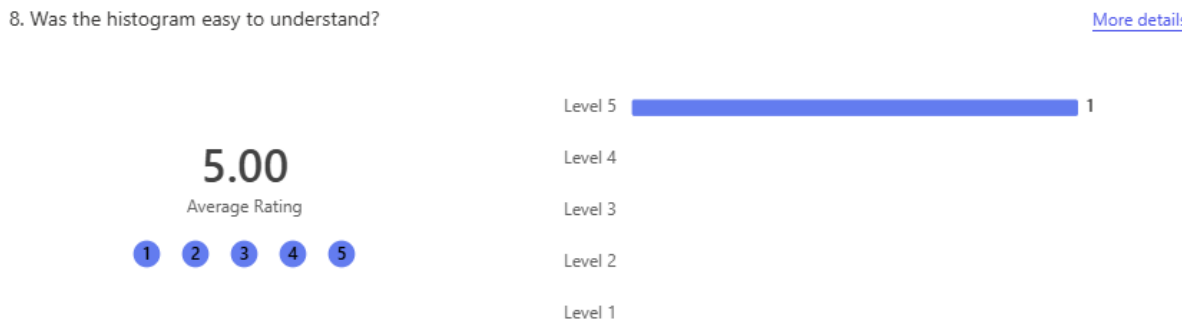


Figure: User Rating for Q4

Q5: What improvements should be made to this page?

9. What improvements should be made to this page? Yes - give your suggesstion No - write "No" only

1 Responses

ID ↑	Name	Responses
1	anonymous	Yes, the line graph shows too messy between lines. I think there is the issues when coding the line chart. This page must add some narrative to tell the story of what data represent.

Figure: User Suggestions

c/ Key Findings:

The State Trends page succeeds in visualizing penalty trends across all eight jurisdictions; however, several usability issues were identified during testing. Participant reported that many lines overlapped heavily, making it difficult to distinguish individual jurisdictions, especially those with lower penalty counts. The lack of tooltips also limited users' ability to identify exact values or verify which line represented which state. In addition, the visual

hierarchy between the line chart and the histogram was unclear, causing users to feel uncertain about how the two visualizations were connected. These findings indicate a need for improved line separation, interactive highlighting, clearer colour mapping, and a more structured layout.

5.1.3 Fairness Comparison Page

Evaluator: Lim Zhi Xuan
Page Developer: Chia Choon Yu

a/ Task Section:

Task 1: Identify which jurisdiction has the highest total penalties.

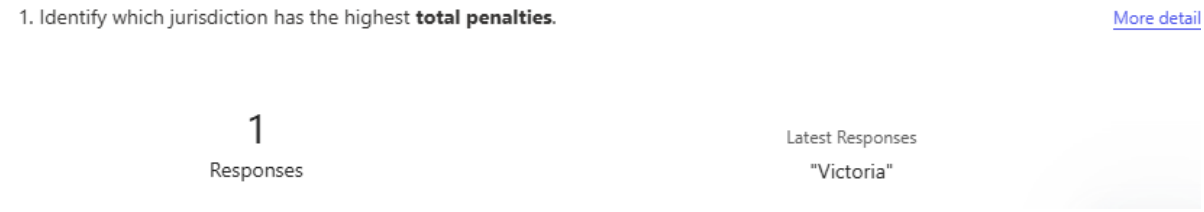


Figure: Response for Task 1

Task 2: Compare NT and VIC in penalties per 10,000 licences.

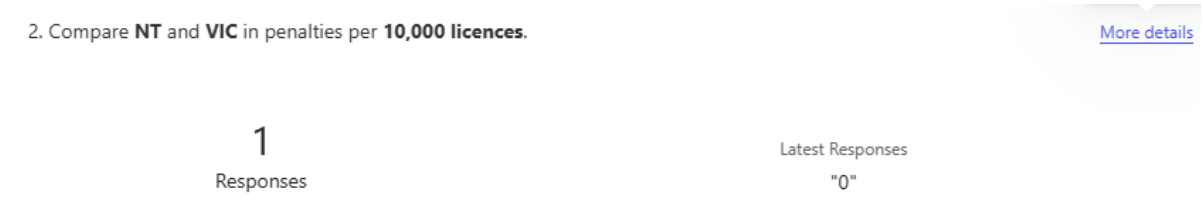


Figure: Response for Task 2

Task 3: Hover over two bars (one in each chart). Describe what the tooltip shows.

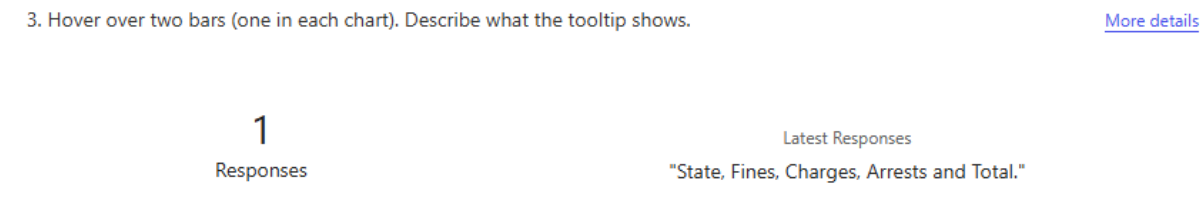


Figure: Response for Task 3

Task 4: What story the two bar charts communicate when viewed together?

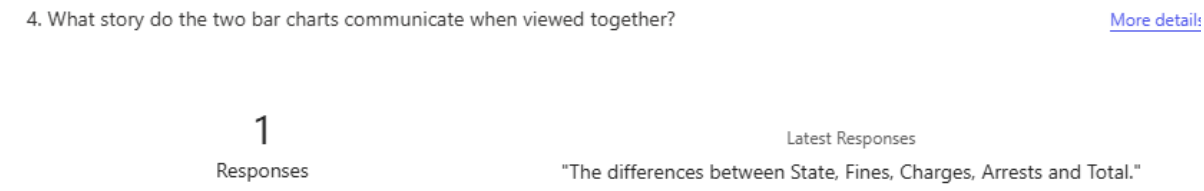


Figure: Response for Task 4

b/ Feedback Questions Section:

Q1: Were the bar charts easy to understand?

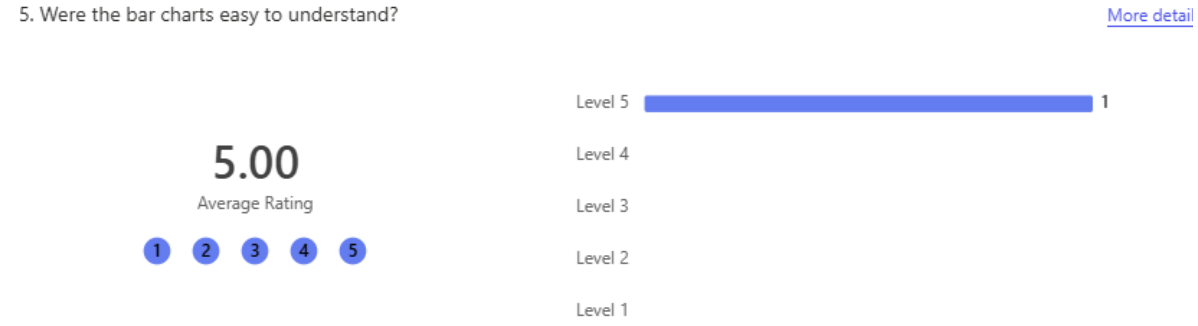


Figure: User Rating for Q1

Q2: Was the tooltip useful?

Figure: User Rating for Q2

Q3: Was it easy to identify which state each bar represented?

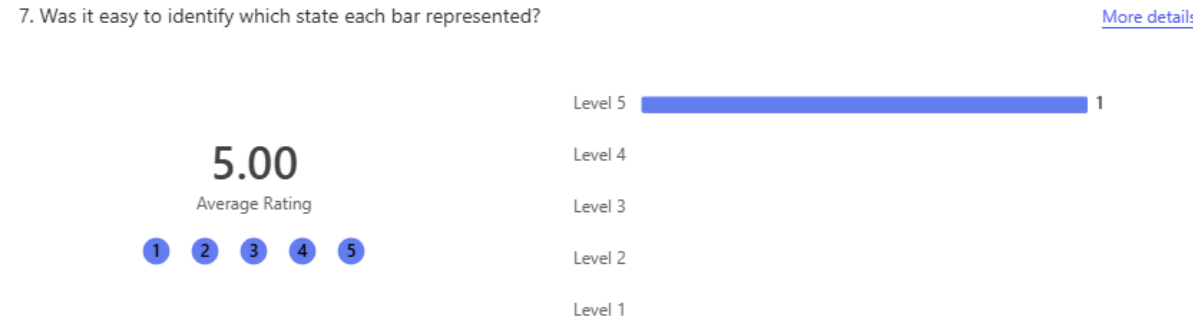


Figure: User Rating for Q3

Q4: Was the layout clear and organized?

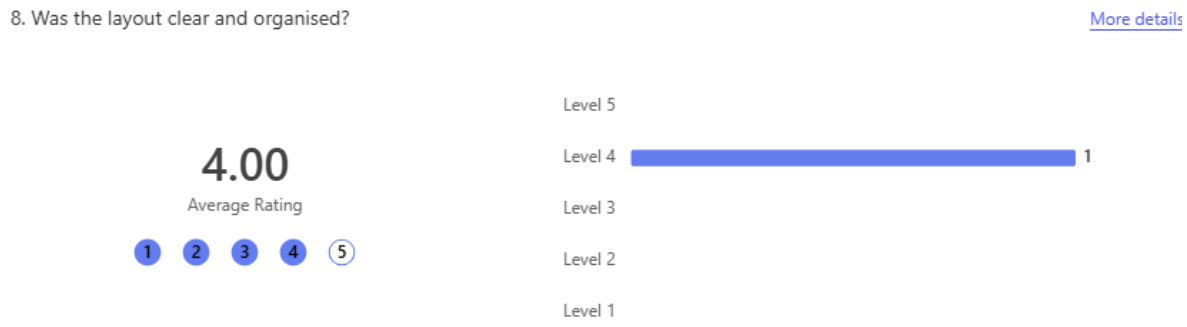


Figure: User Rating for Q4

Q5: What improvements should be made?

Figure: User Suggestions

c/ Key Findings:

The Fairness Comparison page is functional at the 50% completion stage, with both bar charts rendering correctly and tooltips displaying numerical values as expected. Users were able to interpret differences between jurisdictions and understood the narrative contrast between total penalties and population-adjusted penalties. However, several limitations were highlighted during testing. The absence of filters significantly restricted user exploration, making it difficult to view results for specific offence types, detection methods, or age groups. Participant also noted that the bar labels were too small on certain screen sizes, reducing readability and requiring manual zooming to interpret the results. These findings indicate the need to implement page-specific filters and improve label sizing, spacing, and alignment to enhance clarity across devices.

5.2 Usability Testing (100% completion)

5.3 Data Validation

Data validation was conducted to ensure that all datasets used in the project were accurate, consistent, complete, and suitable for visualization. Because this project relies on two separate datasets (1) the BITRE Road Safety Enforcement dataset and (2) the manually collect driver licence dataset from BITRE dashboard. Additional checks were necessary to verify structural alignment and numerical correctness after processing.

5.3.1 Pre-processing validation

Pre-processing validation was conducted to ensure that the raw datasets were structurally complete, consistently formatted, and suitable for subsequent transformations. This step confirmed that all required fields were present, correctly typed, and logically compatible before cleaning, merging, and aggregating the data in KNIME

a/Schema Validation

Schema validation involved checking the structure of both datasets to confirm they contained all required columns and no structural inconsistencies.

For the BITRE Road Safety Enforcement dataset, all elven expected fields were verified to be present, including:

YEAR, JURISDICTION, METRIC, DETECTION_METHOD, FINES, CHARGES, ARRESTS, AGE_GROUP, START_DATE, END_DATE, LOCATION.

Row ID	YEAR	START_DATE	END_DATE	JURISDICTION	LOCATION	AGE_GROUP	METRIC	DETECTION_METHOD	FINES	ARRESTS	CHARGES
Row0	2024	1/1/2024	1/31/2024	NSW	Major Cities of Australia	40-64	non_wearing_seatbelts	Police issued	151	0	2
Row1	2024	1/1/2024	1/31/2024	NSW	Major Cities of Australia	40-64	speed_fines	Police issued	1888	0	10
Row2	2024	1/1/2024	1/31/2024	NSW	Major Cities of Australia	40-64	unlicensed_driving	Not applicable	109	0	431
Row3	2024	1/1/2024	1/31/2024	NSW	Major Cities of Australia	65 and over	mobile_phone_use	Police issued	31	0	0
Row4	2024	1/1/2024	1/31/2024	NSW	Major Cities of Australia	65 and over	non_wearing_seatbelts	Police issued	24	0	0
Row5	2024	1/1/2024	1/31/2024	NSW	Major Cities of Australia	65 and over	speed_fines	Police issued	285	0	0
Row6	2024	1/1/2024	1/31/2024	NSW	Major Cities of Australia	65 and over	unlicensed_driving	Not applicable	9	0	29
Row7	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	0-16	non_wearing_seatbelts	Police issued	1	0	1
Row8	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	0-16	speed_fines	Police issued	2	0	0
Row9	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	0-16	unlicensed_driving	Not applicable	0	0	2
Row10	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	17-25	mobile_phone_use	Police issued	6	0	0
Row11	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	17-25	non_wearing_seatbelts	Police issued	17	0	0
Row12	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	17-25	speed_fines	Police issued	347	0	1
Row13	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	17-25	unlicensed_driving	Not applicable	14	0	47
Row14	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	26-39	mobile_phone_use	Police issued	11	0	0
Row15	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	26-39	non_wearing_seatbelts	Police issued	43	0	0
Row16	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	26-39	speed_fines	Police issued	639	0	1
Row17	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	26-39	unlicensed_driving	Not applicable	24	0	58
Row18	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	40-64	mobile_phone_use	Police issued	3	0	0
Row19	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	40-64	non_wearing_seatbelts	Police issued	20	0	6
Row20	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	40-64	speed_fines	Police issued	732	0	1
Row21	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	40-64	unlicensed_driving	Not applicable	13	0	36
Row22	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	65 and over	mobile_phone_use	Police issued	2	0	0
Row23	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	65 and over	non_wearing_seatbelts	Police issued	8	0	0
Row24	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	65 and over	speed_fines	Police issued	194	0	0
Row25	2024	1/1/2024	1/31/2024	NSW	Outer Regional Australia	65 and over	unlicensed_driving	Not applicable	1	0	3
Row26	2024	1/1/2024	1/31/2024	NSW	Remote Australia	0-16	speed_fines	Police issued	1	0	0

Figure: KNIME Preview BITRE Dataset

Unnecessary fields such as START_DATE, END_DATE, and LOCATION were identified during this stage because they did not contribute to the project's visualization goals.

The total driver licence dataset was also validated to contain the required three attributes: YEAR, JURISDICTION, LICENCES

Row ID	YEAR	JURISD...	LICENCES
Row0	2024	ACT	347892
Row1	2024	NSW	6166226
Row2	2024	NT	163416
Row3	2024	QLD	4154952
Row4	2024	SA	1362548
Row5	2024	TAS	429865
Row6	2024	VIC	5260408
Row7	2024	WA	2194816
Row8	2023	ACT	341420
Row9	2023	NSW	6040358
Row10	2023	NT	161361
Row11	2023	QLD	4051724
Row12	2023	SA	1340060
Row13	2023	TAS	426528
Row14	2023	VIC	5097911
Row15	2023	WA	2132257
Row16	2022	ACT	334900
Row17	2022	NSW	5960216
Row18	2022	NT	159063
Row19	2022	QLD	3973138
Row20	2022	SA	1329139
Row21	2022	TAS	423676

Figure: KNIME Preview total licence dataset

b/Data Type Validation

Data type validation was conducted to ensure that every column in both datasets stored values in the correct format required for processing, aggregation, mathematical calculations, and visualization. Incorrect data types can cause calculation errors, failed joins, unexpected NULL values, or broken visualizations on the website. Therefore, validating the type and consistency of each attribute was a critical step before cleaning and transformation.

police_enforcement_2024_fines.csv

This dataset contains a mixture of numeric attributes and categorical (string) attributes. Through KNIME's table preview, the following data type checks were performed:

- YEAR was confirmed to be an integer, with no decimal or string inconsistencies.
- FINES, CHARGES, and ARRESTS were verified as integer values across all rows. No non-numeric symbols (e.g. "N/A", "-", "null", or commas as "1,200") were detected.
- AGE_GROUP, JURISDICION, METRIC, and DETECTION_METHOD were confirmed to be stored as string fields.
- Minor inconsistencies in the METRIC field format were observed. Metrics appeared in snake_case format (e.g. "mobile_phone_use", "non_wearing_seatbelt"), which required later conversion to readable labels using the String Manipulation node.
- Detection method values displayed mixed formatting (e.g. "Red light camera", "Fixed Camera", "Police issued"), which were later consolidated into two consistent categories.

Columns: 11	Column Type
YEAR	Number (Integer)
START_DATE	String
END_DATE	String
JURISDICTION	String
LOCATION	String
AGE_GROUP	String
METRIC	String
DETECTION_METHOD	String
FINES	Number (Integer)
ARRESTS	Number (Integer)
CHARGES	Number (Integer)

Figure: Data type of police_enforcement_2024_fines.csv

Total_licences_by_jurisdiction.csv

The manually collected driver licence dataset show a cleaner structurer, containing:

- YEAR was confirmed to be an integer.
- LICENCES was confirmed to be an integer.
- JURISDICTION was stored as a string field.

All numeric values were formatted correctly, with no missing entities or irregular character symbols.

Columns: 3	Column Type
YEAR	Number (Integer)
JURISDICTION	String
LICENCES	Number (Integer)

Figure: Data type of Total_licences_by_jurisdiction.csv

Cross-Type Compatibility Check

To ensure both datasets could later be joined, a final compatibility check was conducted:

- YEAR values matched expected integer format in both datasets.
- JURISDICTION values were consistent as strings and shared the same naming conventions across datasets (e.g. “NSW”, “VIC”, “WA”), allowing clean joining without creating duplicate keys or NULL matches.

c/Missing Value Check

Missing value validation was performed to identify any incomplete, null, or inconsistent entities that could affect the accuracy of processing and visualization. The BITRE dataset contained several fields with significant missing values, while the licence dataset also had structural gaps that required careful handling before merging.

police_enforcement_2024_fines.csv

A full scan of the dataset showed that essential fields required for analysis such as YEAR, JURISDICTION, DETECTION_METHOD, AGE_GROUP, FIES, CHARGES, ARRESTS did not contain missing entities. However, specific columns were identified as inappropriate or irrelevant for the project's visualization goals:

(A) START_DATE and END_DATE

Although these fields did not contain missing values, they lacked meaningful analytical purpose for this project. The dataset provides START and END dates for enforcement periods, but the information is incomplete in contextual meaning and does not contribute to cross-state comparison or temporal trend analysis.

(B) LOCATION

The LOCATION field contained descriptive geographic classification such as "Major Cities of Australia", "Outer Regional Australia", "Remote Australia", "Very Remote Australia", and "Unknown".

These values are not missing, but they are not sufficiently precise for state-level spatial analysis and would create analytical inconsistency when visualizing data by jurisdiction. Therefore, LOCATION was identified as contextually irrelevant despite having valid entities.

(C) AGE_GROUP and DETECTION_METHOD

A small number of records were labels as "Unknown". Validation confirmed these were legitimate data categories reported by BITRE, not errors or missing values. Therefore, they were considered valid and did not interfere with processing or analysis.

(D) CHARGES and ARRESTS

They were found to have meaningful values only in 2023-2024. The earlier years showed zero but not missing values.

Validation identified this as structural sparsity, not data loss. Therefore, these zero values were retained because they accurately represent reporting patterns.

Total_licences_by_jurisdiction.csv

The driver licence dataset did not contain missing values within its defined range (2010-2024). However, validation identified a structural gap:

- The dataset does not include licence counts for 2008 and 2009.
- This is not classified as missing data. It is simply not part of dataset's coverage.

During validation, these gaps were noted because they would influence the join with the BITRE dataset. These findings informed later decisions made in the processing stage.

5.3.2 Processing Validation

Processing validation was conducted to ensure that all transformation steps in KNIME such as column filtering, string recoding, aggregation, joining, and calculation produced correct and

consistent outputs. This stage focused on validating how the data was changed, not just that it was changed.

a/ Detection Method Transformation Check

The original BITRE dataset contained multiple detection method labels, such as:

- Average speed camera
- Fixed camera
- Fixed or mobile camera
- Mobile camera
- Red light camera
- Police issued
- Not applicable

Two String Replacer nodes were applied to consolidate these into 2 main categories:

- Auto camera enforcement: for all camera-based methods.
- Manual enforcement: for Police issued and Not applicable.

Processing validation for this step included:

- Generating a frequency table of DETECTION_METHOD before and after transformation to confirm that all camera-related labels were removed and replaced with “Auto camera enforcement”.
- Checking that all “Police issued” and “Not applicable” records appeared as “Manual enforcement” in the processed dataset.
- Verifying that no unexpected third category was created (e.g. due to spelling mistakes or trailing spaces).

This ensured that the detection method dimension was simplified in a controlled and logically consistent way.

b/ Metric Formatting Validation

The METRIC field originally used snake_case formatting (e.g. mobile_phone_use, non_wearing_seatbelt). A String Manipulation node was used to:

- Replace underscores “_” with space “ ”.
- Convert values to capitalized labels for readability.

For example:

- mobile_phone_use = Mobile phone use.
- non_wearing_seatbelt = Non wearing seatbelt.

Processing validation for this transformation included:

- Comparing the list of unique METRIC values before and after transformation to ensure that the number of distinct categories did not change.
- Manually inspecting several rows to confirm that the meaning of each metric was preserved and that only formatting, not semantics, was altered.
- Verifying that no residual snake_case values remained in the final output.

This validation step ensured that metrics were represented consistently across the visualizations without introducing new categories.

c/ Aggregation Validation (GroupBy)

To prepare the enforcement dataset for joining and visualization, penalties were aggregated using a GroupBy node with the group keys:

- YEAR
- JURISDICTION
- METRIC
- AGE_GROUP
- DETECTION_METHOD

For each group, the following aggregations were computed:

- Sum(FINES)
- Sum(CHARGES)
- Sum(ARRESTS)

Processing validation of these aggregation involved:

- Selecting several random YEAR-JURISDICTION-METRIC combinations and manually summing the corresponding rows in the raw data to confirm that the GroupBy output matched these totals.
- Checking that no groups were duplicated and that the number of aggregated rows matched the expected number of unique key combinations.
- Confirming that zero values for CHARGES and ARRESTS in earlier years (before 2023) were preserved and not treated as missing or removed.

This confirmed that aggregation logic was correctly implemented and did not distort the original penalty counts.

d/ Join Validation

The processed enforcement dataset was then joined with the licence dataset using YEAR and JURISDICTION as join keys vi a Joiner node. Processing validation for the join focused on:

- Verifying that each YEAR-JURISDICTION pair in the aggregated enforcement table matched either zero or one corresponding licence record, with no unintended one-to-many matches.
- Checking row counts before and after joining to ensure that the number of rows behaved as expected, and that the join did not unexpectedly duplicate records.

- Inspecting samples of joined rows to confirm that the LICENCES field matched the correct YEAR-JURISDICTION pair and that there were no misaligned states or years.
- Confirming that only years 2010-2024 had valid LICENCES values, consistent with the known coverage of the licence dataset.

These checks ensured that the merged dataset used for visualization was structurally correct and analytically sound.

5.3.3 Numerical Validation

The final step in processing involved calculating penalty rates per 10,000 driver licences using three Math Formula nodes in KNIME. The formals used the Round() function to ensure that the results were clean, readable, and appropriate for public-facing visualization.

The formulas implemented were:

- $\text{FINES PER 10000 LICENCES} = \text{Round}(\frac{Sum(FINES)}{LICENCES} \times 10{,}000)$
- $\text{ARRESTS PER 10000 LICENCES} = \text{Round}(\frac{Sum(ARRESTS)}{LICENCES} \times 10{,}000)$
- $\text{CHARGES PER 10000 LICENCES} = \text{Round}(\frac{Sum(CHARGES)}{LICENCES} \times 10{,}000)$

Numerical validation included:

- Random samples of YEAR-JURISDICTION rows were manually recalculated using a calculator and compared with KNINE output. All values matched correctly after applying the round rule.
- The same LICENCES count was used for fines, charges, and arrests in each row. Rounding behavior was consistent across all three formulas. There was no unintended decimal values remained.
- Because the licence dataset only covers 2010-2024, all rows for 2008-2009 were filtered out earlier in processing. So, there was no division-by-zero occurred and all calculations were based on valid denominators.
- The resulting rate columns were examined for: Extremely large values, negative values, unexpected zeros, and sudden spikes due to join or aggregation issues.
- The rounded outputs were tested inside the D3.js bar charts for the Fairness Comparison page to ensure: no NaN values, no undefined rates, and no formatting issues when displayed in tooltips.

5.3.4 Output File Validation

The final stage of data validation focused on verifying that the processed datasets exported from KNIME were structurally and semantically correct for direct use in the web visualization.

Two main CSV output were validated:

Penalties_process.csv

It contains aggregated enforcement data with: YEAR, JURISDICTION, METRIC, AGE_GROUP, DETECTION_METHOD, Sum(FINES), Sum(CHARGES), and Sum(ARRESTS).

Penalties_per_10,000_licences_processed.csv

It contained the same grouping key plus: FINES_PER_10000_LICENCES, CHARGES_PER_10000_LICENCES, and ARRESTS_PER_10000_LICENCES.

Output validation included:

- Checking that column names followed consistent naming conventions and matched those referenced in the D3.js code used for the website (to prevent undefined field errors).
- Confirming that there were no empty rows, corrupted characters, or misaligned delimiters in the CSV files by previewing them in a text editor and spreadsheet software.
- Verifying that the total number of rows and unique YEAR-JURISDICTION-METRIC combinations remained consistent with the final KNIME table.
- Loading each file into the visualization prototypes to ensure that map, line charts, histogram, and bar charts all rendered without NaN values, missing labels, or unexpected gaps.

These checks ensured that the final exported files were reliable, machine-readable, and fully compatible with the font-end visualization pipeline.

5.4 Visualization Validation

5.5 Accessibility Validation (WAVE)

5.6 Summary of Validation Actions

6.Conclusion

7.References

Bureau of Infrastructure and Transport Research Economics. (2024a). *Road safety enforcement data* [Data set]. Australian Government Department of Infrastructure, Transport, Regional Development, Communications and the Arts. https://catalogue.data.infrastructure.gov.au/dataset/road-safety-enforcement-data

Bureau of Infrastructure and Transport Research Economics. (2024b). *Driver licences by jurisdiction (2010–2024)* [Power BI dashboard]. Australian Government Department of Infrastructure, Transport, Regional Development, Communications and the Arts. https://app.powerbi.com/view?r=eyJrIjoiZTAxY2EyOTAtYTdhMS00NTRiLWI0NDktM2U1ZDI0NzY0ZTU5IiwidCI6ImFhMjFiNjQwLWJhYzItNDU2ZC04NTA1LWYyY2MwN2Y1MTc4NCJ9