Exploring Digital Scams in Australia
Socioeconomic Drivers & Trends

Name: Riddhi Tawde

Student ID: 34757783

Teaching Associate: Bruno Vasquez & Michael Niemann

Applied Session: 08

1. Introduction

In the period following the COVID-19 pandemic in Australia, the rate of digital fraud has dramatically increased. Scamwatch reported that Australians lost more than $3 billion to scams in 2023 alone (ACCC, 2024). The drivers of these scams will be evaluated in Australia between 2021 to 2023, examining different types of scams, the relationship between scam activity and regional socioeconomic position, and the available digital access.
This interactive narrative visualisation builds on the Data Exploration Project (DEP), which analysed Scamwatch, the Australian Bureau of Statistics (ABS) Population data, and the Socio-Economic Indexes for Areas (SEIFA) .

Research Questions
This visualisation is designed to answer the three critical research questions from the DEP:

1. What are the most financially damaging scam types in Australia, and how have their impacts evolved over time?
2. Do regions with lower digital access or socioeconomic status (disadvantage) experience higher scam losses or scam incident rates?
3. How does regional digital inclusion (advantage) relate to the number of individuals reporting scams or fraud?

Key Findings to Communicate
The visualisation is designed to communicate the key findings discovered in the DEP:

1. Dominant Scam Type: Investment scams resulted in the most financially damaging loss by a considerable margin than all other types of scams combined.
2. Temporal Trend: Total financial losses showed a clear and concerning upward trend from 2021 to 2023.
3. Demographic Vulnerability: People aged 65 and older had the highest per capita total dollar loss, indicating that this is a demographic that is vulnerable.
4. The Reporting Paradox: Analysis shows that socioeconomic disadvantage (using IRSD score) does not appear to strongly predict reported scams, but there is a very strong positive correlation (r=0.89) of higher levels of digital inclusion (using IEO score) correlating to higher reporting rates, per capita. This key finding suggests that official records are underreporting the actual numbers of scams occurring to disadvantaged communities as victims are less likely to report.

Targeted Audience
This visualisation is intended for policy makers, government departments (like the ACCC), and community agencies. It is assumed that the audience is informed - but they may not be technical data analysts. For the audience, this is intended to be an exploratory tool for understanding their opportunities for resource allocation, developing intervention strategies, and designing public education campaigns targeted to better protect the most vulnerable groups in our society.

2. Design Process

In accordance with the unit requirements, the design process used the Five Design-Sheet (FdS) approach to explore, iterate, and develop a visualisation solution systematically. In this process, designers will generate three different design options (Sheets 2 - 4), and then synthesise the strongest ideas, concepts or components into a final design/development (Sheet 5).

2.1 Sheet 1: Brainstorming and Ideation

The initial brainstorming phase resulted in 16 individual visualisations. The visualisations were organized into three, themed categories to address the project's RQs:

- Loss & Report Data: Line charts, area charts and scatter plots to show trends and correlations.
- Scam Type: Treemap, donut, and bar charts to show part-to-whole relationships.
- Demographics: Line, pie, bar, sankey, and waffle charts to show demographic breakdowns and flows.

Filtering and Refinement - In the filtering process too some charts which were too complicated or similar (e.g., waterfall, sunburst, funnel) were removed. This filtering process resulted in three possible dashboard structures, (1) Single dashboard view, (2) Story Telling view, and (3) Tab view which was used to create the three design alternatives.

The Sheet 1 brainstorming was guided by Munzner's (2014) 'what-why-how' framework. The **'why'** (user tasks) was defined by the project's RQs, and the **'what'** was the data (losses, scam types). The 16 generated concepts were an exploration of the **'how'** (visual encodings) to connect the 'what' to the 'why'. The filtering process then eliminated the less effective 'how' solutions.

2.2 Sheets 2, 3, & 4 : Three Design Alternatives

Alternative Design	Primary Focus	Central Visual Elements	Key Advantages	Key Disadvantages
Sheet 2: Single Page Dashboard	**What:** Quantitative & categorical data. **Why:** To identify summaries (KPIs), compare magnitudes (bars), & explore correlations (scatter). **How:** Uses key-value (KPIs),	• KPIs (Total Loss, Total Reports) • Waffle Chart (Gender/Age) • Vertical Bar Chart (Scam Type) • Scatter Plot (Factors)	• Provides a fast, comprehensive summary in one view. • Easy filters & bar graphs allow for simple and quick	• High risk of visual clutter and information overload • Limited space forces charts small, with a lack of detail. • Flat visual hierarchy

	bar chart (length), & scatter plot (position) idioms.	• Horizontal Bar Chart (Contact)	comparisons & data slicing.	prevents narrative flow.
Sheet 3: Storytelling View	**What:** Trends over time, categorical breakdowns, & relationships. **Why:** A task that is guided. The user is led into identifying patterns, comparing trends, & correlating metrics. **How:** Uses a guided narrative (scrolling), a stacked area chart, a Sankey chart, & decomposed filter actions (slider, map).	• Vertical scrolling layout • Global Year Slider • Clickable State Map (filter) • Stacked Area Chart (Loss v/s Time) • Donut/Pie Chart (Loss by Type) • Sankey Chart (Contact Method)	• Guided narrative flow makes complex data highly accessible. • Highly intuitive controls (slider, clickable map) provide visual control ease.	• Rigid, linear constructs limit exploration as it is not considered for analytical users. • Scrolling breaks the context between the charts, forcing people to rely on their memory.
Sheet 4: Tab-Viewed Dashboard	**What:** Data is separated out into Trends (Tab 1) & Breakdowns (Tab 2). **Why:** Data is separated out into Trends (Tab 1) & Breakdowns (Tab 2). **How:** Uses a tabbed interface idiom to separate views, then uses line/area charts, scatter plots, & treemaps/pie charts for specific tasks.	• Tabbed interface (Tab 1, Tab 2) • Global Filters (State, Year, Age) • Tab 1: Stacked Area Chart, Multi-Line Chart, Scatter Plot • Tab 2: Treemap, Pie Chart, Multi-Line Chart	• Tabs clean up the interface and significantly reduce clutter. • Separation supports larger more prepared visualizations (scatter plots, line graphs).	• Complex filtering may cause lag in loading. • The separation may make it more difficult to observe connections when the charts are in separate tabs.

Through the analysis of the design process, the main limitation highlighted in all three alternatives (Sheets 2, 3 and 4) was the inability to assess the temporal, geographic and relational dimensions together without significant cognitive switching costs.

2.2 Sheet 5 : Final Realization - Integrated Dashboard

The final design is a synthesis that avoids the information overload of Sheet 2 and the rigid exploration of Sheet 3. It uses the tabbed layout from Sheet 4 to create a hybrid narrative structure, balancing a guided introduction with a reader-driven exploratory tool for a mixed audience.

A. Layout Structure and Narrative Genre

The design utilizes a dual-mode narrative that begins with the "Guided Introduction" (home page).When the user click on "Go to Dashboard," you will enter the "Exploratory Dashboard," a reader-driven, two-tiered interface:

1. Top Tier (Global Controls): Global filters (Year Range and Metric Toggle) and the main KPIs (Total Losses and Total Reports) are always in the header (Top Tier) so that readers can interact with the main KPIs and global filters throughout their use of the dashboard to review the findings in the Project Summary.
2. Lower Tier (Tabbed Views): The dashboard separates the analysis through thematic tabs, the "Geo Analysis" tab on one page and the "Demographic Analysis" on another page, to limit the scope of the dashboard to thematic analysis. Also utilized in the lower tier are larger, more detailed chart views of the data, which may facilitate deeper knowledge and exploration of the data set.

B. Visualization Components and Rationale

The final charts were chosen to best answer the project's RQs, aligning with Munzner's (2014) framework.

View	Encoding (Visual Variables)	Rationale & Justification
Home: KPIs	Text + Size (Large Typography)	**Identify** key summary figures; **preattentive size** establishes an immediate overview.
Tab 1: Geo Analysis	• **Map:** Geography + Sequential Colour • **Scatter:** X/Y Position + Categorical Colour • **Line:** Position + Slope	**Identify** spatial patterns (Map). **Discover** correlations (Scatter). **Compare** filtered vs. national trends (Line).

Tab 2: Demographic Analysis	• **Treemap:** Area + Categorical Colour • **Sankey:** Link Thickness (Quantitative Flow) • **Line:** Position + Slope + Categorical Colour	**Compare** part-to-whole magnitudes (Treemap). **Reveal** flow patterns (Sankey). **Compare** trends across demographic groups (Line).

C. Interaction Design and Consistency

- Global Filters That Are Consistent: The filters across the top bar (Years, Age and Metric) are consistent and always propagate to all charts across all tabs at once keeping the data context consistent.
- Cross Filtering (drill down): The dashboard is fully interactive. On the map, when the user clicks a state, or on the treemap it essentially filters all the other charts. This perfectly implements overview first, zoom and filter, then details-on-demand mantra.
- Hover methods: All charts have an ability to receive detailed data when hovered on. This is also a part of the details on demand component of the dashboard that can provide exact numbers without cluttering.
- Clear State: The Clear Filters button, and the Active filter/text provide a clear way for the user to understand the state of their view, and be able to reset their view.

D. Improvements Over Alternatives

- Versus Sheet 2 (Single Page): It resolves the information overload and small, cluttered charts by using the tabbed layout to give each visualization space.
- Versus Sheet 3 (Storytelling): It fixes the rigid and limited exploration issues because it retains the story concept (the Home Page); however, the story is now at the beginning of the experience, allowing that the main dashboard can be a fully-interactive and reader-driven tool to use.
- Versus Sheet 4 (Tabs): It resolves the disconnected views risk by adding persistent global filtering, making cross filtering between charts easier, and adding named, thematic tabs to give some theme to the tool while still functioning as a single tool.

2.3 Key Changes Made Post-Presentation:

- **Expanded Brainstorming:** The brainstorming was increased to include 16 different possible visualisation types (Sheet 1) which were then explicitly filtered to determine which would work best for the project.

- **Re-conceptualized Alternatives:** The design alternatives were significantly altered. Each sheet originally answered only one research question. Revised sheets now each offer a complete dashboard (Single Page, Storytelling, Tabbed) each answering all research questions.

- **Added Narrative Landing Page:** The final design (Sheet 5) loaded directly into the dashboard. Included a Guided Introduction page to informatively ground the user in essential author-driven context before the user bears witness to the data, as noted in the illustration (Sheet 5).
- **Upgraded Key Visualisations:** Several charts were improved from basic bar charts (Sheet 3) for greater insight, including a Treemap to show part-to-whole, a Sankey Diagram for flow, and a Multi-Line Chart to show demographic trends over time (Sheet 1, 3, 4).
- **Refined Global Controls:** The controls were also updated. The dropdown 'Year' (Sheet 5) was changed into a clearer Year Range Slider (Sheet 3) and also gained an animation feature (Sheet 5).
- **Enhanced Cross-Filtering:** Interaction was also improved from simple filtering by map (Sheet 4) to complete cross-filtering where all charts can be used as a filter (Sheet 5).
- **Specified Design Details:** Design notes changed from non-specific to explicit statements of accessible colour scheme and clear, legible sans-serif font(Sheet 5 ; Sheet 5) to enhance legibility.

3. Implementation

3.1 Technical Implementation

Technologies and Libraries

The visualisation was implemented as an R Shiny application, key libraries include.

1. **`shiny`**: The core application framework (Chang et al., 2024).
2. **`plotly`**: Applied for all interactive charts.
3. **`leaflet`**: Used for the interactive choropleth map of Australia (map_geo) (Cheng et al., 2023).
4. **`dplyr (Wickham et al., 2023)`** and **`tidyr (Wickham & Vaughan, 2024)`**: Used for all of the data manipulation, filtering and summarization in the reactive expressions.
5. **`sf`** and **`rnaturalearth (South, 2017)`**: Used to obtain the Australian state geospatial data and manage geospatial data for the leaflet map.
6. **`shinyjs`**: Used to set up the modal landing page and manage the show/hide logic of the main dashboard.

Data Processing

- Initial Load: The application opens and cleans the Final1.csv and rnaturalearth spatial data. It also engineers date columns and computes static KPIs for the landing screen.
- Reactive Filtering: A central data_all_filtered expression is filtering the data based on all user input. It also dynamically creates a Metric column for the Loss/Report toggle.
- Dynamic Aggregation: All charts are listening to this data, aggregating (group_by/summarise) in real-time. Complex transformations such as lm() fits for the

scatterplots and creating a node/link structure for the Sankey diagram are conducted dynamically.

Technical Challenges

- Persistent Cross-Tab Filtering: Ensuring filters stayed active when switching tabs was a challenge. This was solved using central reactiveVal objects (e.g., shared_year, shared_state) as a single source of truth.
- Universal Cross-Filtering: Implementing click-to-filter across different libraries (Plotly and Leaflet) was complex. This required separate observeEvent listeners for event_data and map_geo_shape_click. These listeners update the shared reactiveVals, which in turn trigger the single data_all_filtered expression to re-calculate all charts.
- Sankey Diagram Data Structure: Plotly's Sankey diagram requires a specific node/link structure with zero-indexed IDs. Custom logic was required to dynamically transform the filtered data into this format, mapping Contact_Mode and Scam_Category strings to a nodes list and building the links dataframe.

3.2 Interactive Narrative Visualisation Implementation

The final implementation follows a two-phase narrative structure, combining author-driven context with reader-driven exploration.

Phase 1: Narrative Landing Page (Author-Driven)

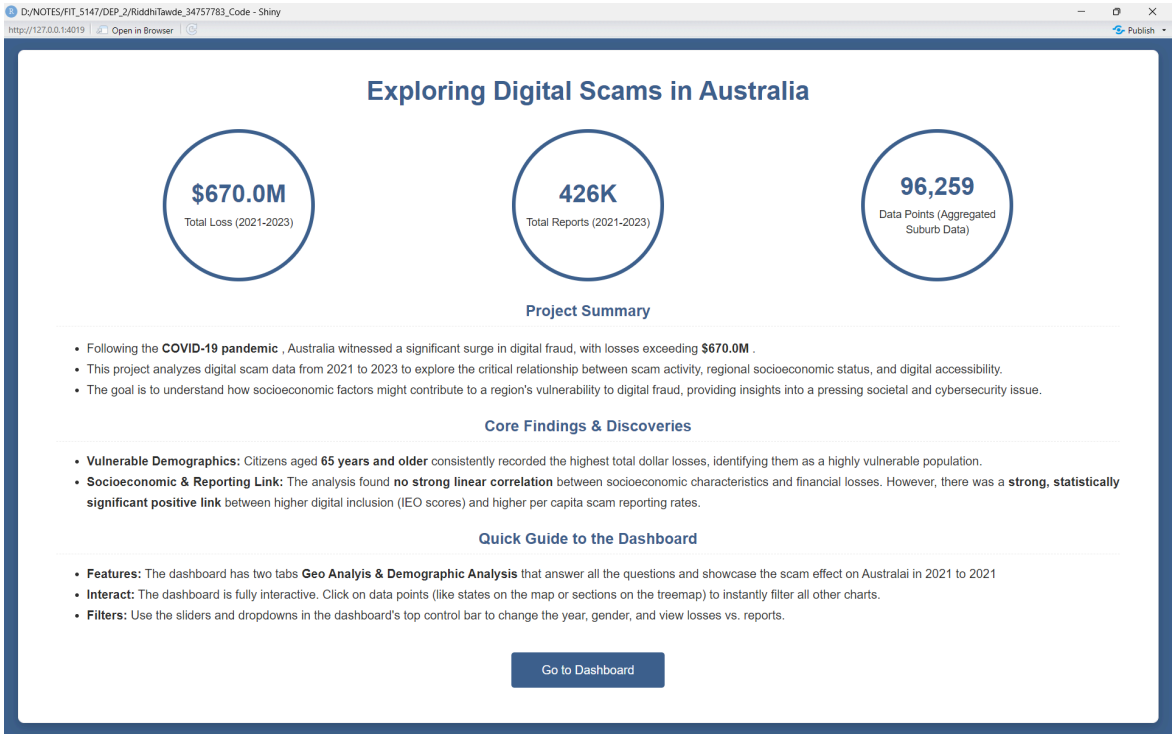

Fig. 1. Narrative Landing Page (Author-Driven)

Upon launch, the user is presented with a modal landing screen (description_page_ui). This landing page is designed to tell the narrative, aligned to a DVP, which tells a data story:

- Project Title: Exploring Digital Scams in Australia.
- Three Key Statistics: Displayed in large KPIs in circles to frame the scale of the problem immediately - Total Loss (e.g., $670.0M), Total Reports (e.g., 426K), Total Data Points
- Narrative Text: Two sections, Project Summary and Core Findings & Discoveries,that prepare the user for what they will go to explore. The text highlights key information like the susceptibility of the 65+ age demographic and how digital inclusion (IEO) is positively related to reporting rates.
- Quick Guide: A quick section that explains the dashboard features, including the tabbed framework and user-activated filters.
- Call-to-Action Button: A large Go to Dashboard button that transitions the user away from the author-narrative to the reader-initiated exploration.

Phase 2: Interactive Dashboard (Reader-Driven)

The primary dashboard has global controls at the top and is divided into two tabs. The global controls include the following filters:

Global Controls & KPIs:

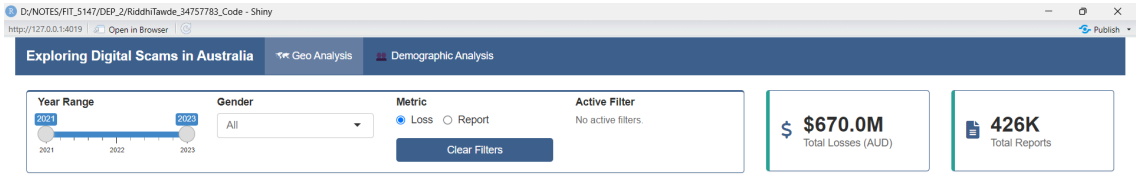

Fig. 2. Filter board

- Global Controls & KPIs: There is a persistent control panel (controls-box) and KPI section at the top of each tab.
- Year Range Slider: Allows users to filter data down to a specific range, between 2021 and 2023.
- Gender Dropdown: Filters all charts by the selected gender.
- Metric Toggle: Radio buttons toggle the entire dashboard, to present data as Loss($) or Report(count). This toggle dynamically changes which Metric column is used by all visualisations.
- Active Filter: Display shows, in a text box, all active non-standard filters that are currently applied (e.g., State: New South Wales, Scam Type: Investment scam).
- Dynamic KPIs: The two kpi-box elements show the Total Losses and Total Reports for the currently filtered data, updating instantly with any interaction.

Tab 1: Geo Analysis

This tab focuses on spatial patterns and socioeconomic correlations.

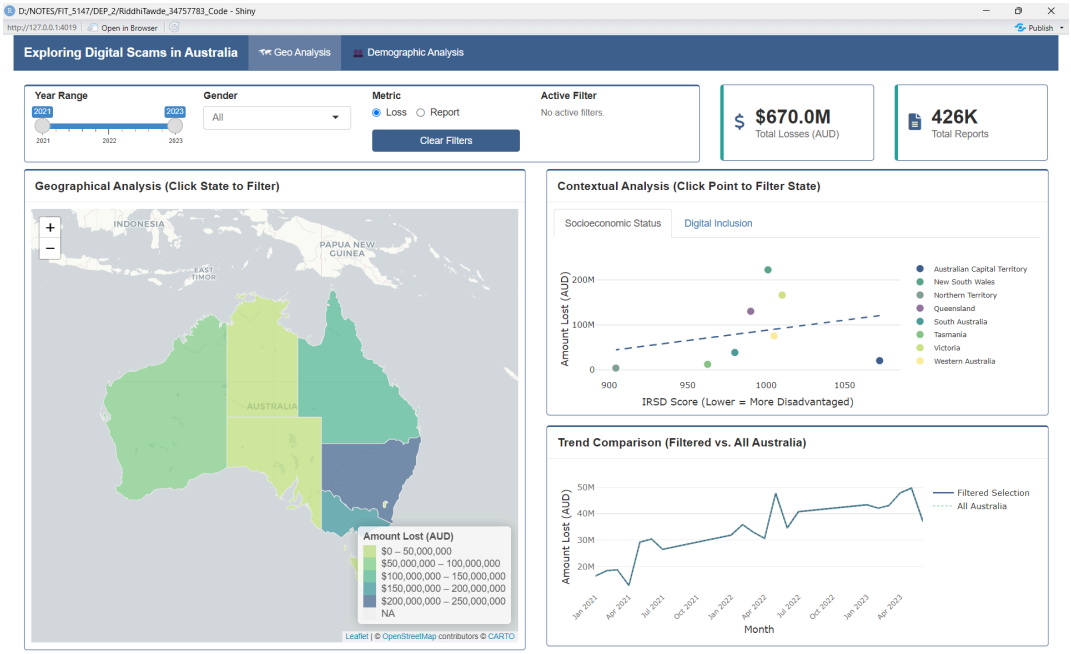

Fig. 3. Tab 1: Geo Analysis

<u>View 1.1: Geographical Analysis Map</u>

The choropleth leaflet map answers the question: How does scam impact vary by region?

- <u>Visual Design:</u>
  - Choropleth Encoding: The Australian states are given color based on the selected Metric.
  - Sequential Palette: It's a green-to-blue sequential color scheme with darker intensity representing greater value.
- <u>Interaction:</u>
  - Hover Tooltips: If you mouse over a state, you get a label with the state's name, the value of the selected Metric, total lost, total reported, and an average score on a mapping of the IRSD index.
  - Linked Filtering: By clicking on a state polygon the whole dashboard (across both tabs) is filtered to that state.
- <u>Design Rationale:</u> A choropleth map is the most effective idiom for showing how a quantitative value (like $Loss) is distributed over geographical regions. It leverages the preattentive visual variable of color intensity to provide an overview first of spatial patterns.

View 1.2: Contextual Analysis (Scatter Plots)

These tabbed scatter plots address: What is the relationship between scam impact and socioeconomic factors?

- Visual Design:
  - Socioeconomic Status: A scatter plot (scatter_socio) created using Plotly of the Metric vs. the IRSD Score (Disadvantage).
  - Digital Inclusion: A Plotly scatter plot (scatter_digital) of the Metric vs. the IEO Score (Digital Inclusion).
  - Trend Line: A trend line (dashed) with R^2 value, the strength of the relationship, will be calculated automatically and overlaid on both plots.
- Interaction:
  - Hover Tooltips: When hovering on each point in the graphics, the state, its socioeconomic score and the Metric value selected will show up.
- Design Rationale: A scatter plot is the standard, most effective idiom for exploring the relationship and correlation between two quantitative variables (Munzner, 2014). Using position (X and Y) allows a user to instantly spot patterns, clusters, and outliers. The on-the-fly trend line calculation is critical for answering the research question.
- Data Insight Revealed: This displays the Core Finding: weak or no relationship with socioeconomic disadvantage (IRSD) but a strong positive relationship with digital inclusion (IEO), which means that areas with more digital inclusion report scams more often.

View 1.3: Trend Comparison (Line Chart)

This line chart addresses: How does a selected region's trend compare to the national average?

- Visual Design:
  - Dual Lines: Plots two series over time: the Filtered Selection(Dark Blue) and the All Australian Baseline (Light Green, Dashed).
- Interaction:
  - Hover Tooltips: When hovering over the lines, the user will see the exact Metric value for that month.
- Design Rationale: Using a line chart is the best way to show trends in quantitative data over time (Munzner, 2014). The alternate line Filtered vs. All Australia gives important context so users can compare the filtered state design (i.e., between states) against the all Australia national baseline, which is a much deeper conclusion than simply seeing the filtered trend on its own.
- Data Insights Revealed: Users will immediately know if the filtered selection (e.g., particular state) is trending higher or lower than the national average trend.

Tab 2: Demographic Analysis

This tab focuses on scam types and the who & how of demographic impact.

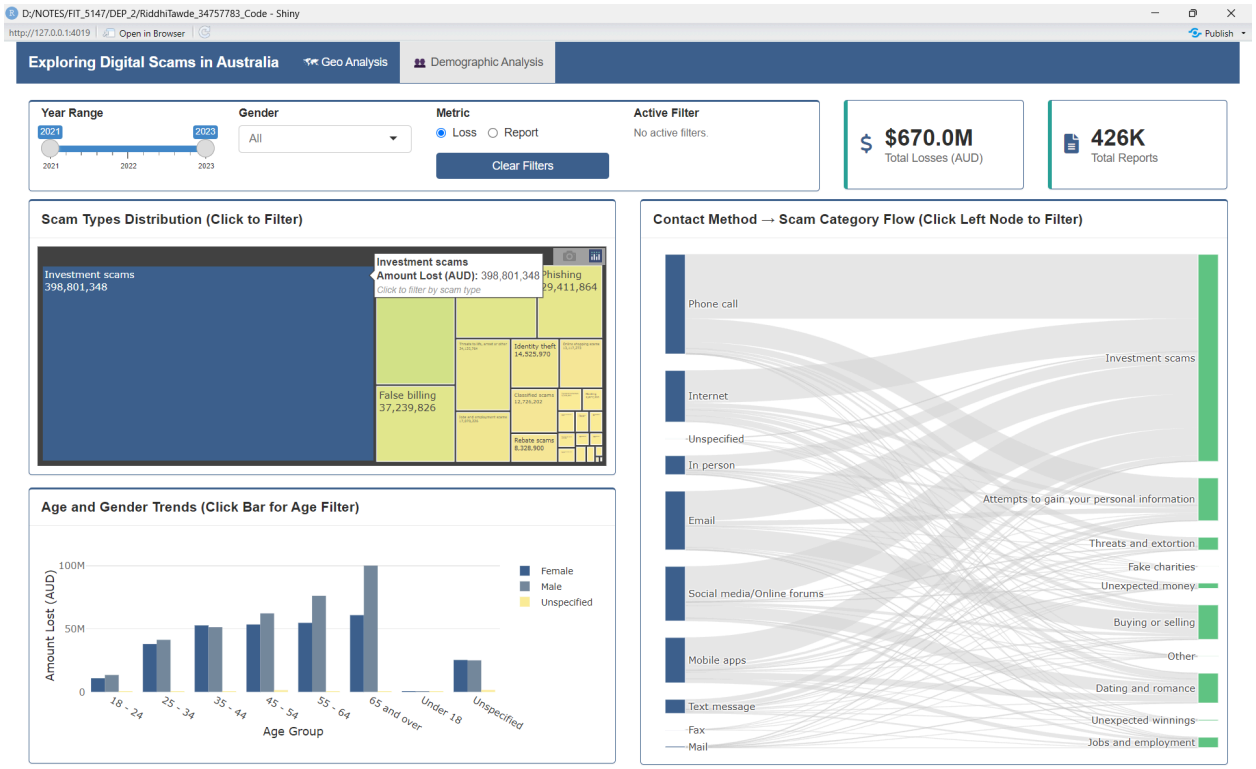

Fig. 4. Tab 2: Demographic Analysis

<u>View 2.1: Scam Types Distribution (Treemap)</u>

The treemap addresses: Which scam types are the most damaging?

- <u>Visual Design:</u>
  - Treemap Encoding: Uses plotly to visualize each Scam_Type. The area of the rectangle is proportional to the Metric you choose to select one of the Scam_Types.
  - Color Encoding: A sequential color scheme is used to distinguish in the blocks as well.
- <u>Interaction:</u>
  - Hover Tooltips: Hovering over a block provides the scam type and its Metric value as an exact number.
  - Linked Filtering: Clicking on the rectangle filters the entire dashboard to that Scam_Type.
- <u>Design Rationale:</u> A treemap is more space efficient and easier to interpret than a pie chart when comparing part-to-whole data across many categories. It employs area as a preattentive visual variable for the user to observe which 'slice' is most important.

- Data Insights Uncovered: Immediately answers a key question by showing Investment scams are the biggest block representing the most financial loss.

View 2.2: Age and Gender Trends (Bar Chart)
This grouped bar chart addresses: Which demographic groups are most vulnerable?

- Visual Design:
  - Grouped Bar Chart: Plots the Metric (y-axis) against Age (x-axis) .
  - Color Encoding: Bars are grouped and colored by Gender in the qual_palette.
- Interaction:
  - Hover Tooltips: Provides detailed figures for the specific age/gender group.
  - Linked Filtering: Clicking a bar filters the entire dashboard by the corresponding Age group .
- Design Rationale: A grouped bar chart is a very good option to compare a quantitative value (Metric) across two different categorical variables (Age and Gender). Length (the height of the bar) is very precise and it is preattentive in this context when comparing values to a categorical sort. Grouping by age on the x-axis also indicates a clear progression.
- Data Insights Revealed: This chart clearly tells the Core Finding that the 65+ age group suffers the greatest financial losses consistently.

View 2.3: Contact Method → Scam Category Flow (Sankey)
The Sankey diagram addresses: How are people being contacted for specific scam categories?

- Visual Design:
  - Sankey Flow Diagram: A plotly diagram visually displaying flow from the Contact_Mode (left nodes) to the Scam_Category (right nodes).
  - Link Encoding: The width (thickness) increases or decreases based on the selected Metric to visually show how large a flow connection is.
  - Color Coding: The nodes will be in separate colors to differentiate Contact_Mode (blue) versus Scam_Category (green).
- Interaction:
  - Hover Tooltips: When hovering on a flow you will see the source, target, and exact value. When hovering on a node it will display information and a prompt to filter.
  - Linked Filtering: Clicking a Contact_Mode (left node) filters the entire dashboard by that method.
- Design Rationale: It is a key visualisation as it demonstrates relationships, not just totals. A simple bar chart (like some of the older designs) could demonstrate, Top Contact Methods.A Sankey diagram is particularly effective in demonstrating flow from one categorical set to another, with the thickness of the links providing effective quantitative encoding.
- Data Insights Revealed: This reveals the how. For example, it might show that Phone callis the primary vector for Investment scams,while Emailis the primary vector for Phishing.This provides actionable insights for targeted interventions.

Layout and Visual Hierarchy
- Primary Level: The modal Home Page(description_page_ui) .

- Secondary Level: The persistent navbarPage header and the global controls-box and kpi-box (high-visibility, top-of-page) .
- Tertiary Level: The main content tabs (Geo Analysis and Demographic Analysis).

Typography

- Font Family: 'Open Sans', sans-serif, is used for all text to ensure readability.
- Project Title: 34px, 700 weight (bold) .
- KPI Circle Value: 32px, 700 weight (bold) .
- Section Headers (Home): 20px, 600 weight .
- Chart Box Titles: 17px, 600 weight .
- KPI Box Value: 28px, 700 weight (bold) .
- Body Text (Home): 16px, 1.7 line-height for comfortable reading .
- KPI Circle Label: 14px, regular weight.

3.3 Using the Implementation

A. Running the Application

- Prerequisites: R and RStudio must be installed, along with the required R packages (shiny, plotly, leaflet, dplyr, etc.) .
- Data Files: The Final1.csv file must be in the same directory as the app.R script.
- Steps to Run**:**
  - Open the app.R script in RStudio.
  - Click the Run App button in the RStudio interface.
  - The application will launch, displaying the Home Page first.

B. Navigation

- Home Page: The application opens on the Home Page(description_page_ui).
- Read the Project Summary,Core Findings,and Quick Guide to understand the project context .
- Click the Go to Dashboard button to enter the main interactive tool .
- Main Dashboard: The main dashboard is a navbarPage with two tabs:
  - Geo Analysis: The first tab, containing the map and socioeconomic scatter plots .
  - Demographic Analysis: The second tab, containing the treemap, Sankey diagram, and age/gender bar chart .

C. Using the Dashboard

Global Controls (At the top of each tab):

- Year Range: Drag the Year Range slider to select a period between 2021 and 2023.
- Metric: Click the Metric toggle to switch the entire dashboard (all charts and map) to display either Loss($) or Report(count) data .
- Gender: Select Male,Female,or Unspecified from the Gender/dropdown to filter all charts.

Tab 1: Geo Analysis

- Overview: Start with the Geographical Analysis map. The color of each state shows the total value of your selected Metric.
- Hover (Details-on-Demand): Hover over a state on the map to see its name, total loss/reports, and average IRSD score in a tooltip.
- Cross-Filter (Map): Click on a state. This will filter the *entire dashboard* (both tabs) to show data for only that state. The Active Filterbox will update to show your selection.
- Explore Context: Look at the Contextual Analysis scatter plots to see where that state sits in relation to socioeconomic status (IRSD) and digital inclusion (IEO) .
- Compare Trends: Check the Trend Comparison chart to see that state's trend (blue line) compared to the All Australia baseline (green line).

Tab 2: Demographic Analysis

- Identify Top Scams: The size of the rectangles shows which scams are largest for your selected Metric.
- Cross-Filter (Treemap): Click a rectangle to filter the entire dashboard by that scam type.
- Find Vulnerable Groups: Examine the Age and Gender Trendsbar chart. This shows the impact by age and gender. Click a bar to filter the dashboard by that age group.
- Analyze Flows: Use the Contact Method → Scam Category FlowSankey diagram. This shows how people are contacted (left) for different scams (right). Click a contact method on the left to filter the dashboard by that method.

D. Resetting

- At any time, click the Clear Filters button in the global controls to reset all filters (State, Age, Gender, Scam Type, etc.) and return to the default view.
- Clicking a selected item a second time (e.g., clicking a state on the map that is already filtered) will also clear that specific filter.

4. Conclusion

The project successfully created an interactive narrative visualisation that conveys complex relationships in the digital scams landscape in Australia. Continuing the work of the Data Exploration Project, the final R Shiny application offers a multi-layered tool for use by policy makers and researchers to explore the interaction of digital scams over time and space. It directly tackles the three main research queries by allowing coordinated views of geographic, socioeconomic and demographic perspectives.

Key Achievements:

- Integrated Multi-Source Data: The dashboard visualises a merged dataset from Scamwatch, ABS and SEIFA, creating a rich multi-dimensional proposition for analysis.

- Three-Question Framework: The visualisation answers all three research questions, through a two-tab approach to enable the audience to engage with the temporal aspects (RQ1), geographic/socioeconomic aspect (RQ2), and digital inclusion "reporting paradox" (RQ3).

- Co-ordinated Interaction: The implementation of sophisticated, persistent form of cross-filtering (reactiveVal and observeEvent) allows a user to fluidly develop an understanding of the relationships between the varying aspects of the data (e.g. when depending from one aspect of data, a user can understand the relationship to the geography of the scam type).

- Narrative Scaffolding: The landing page provides context and key findings in the main area of focus, essentially contributing to the "narrative visualisation" requirement and guiding the audience before them becoming engaged in their own reader-driven engagement.

The dashboard visually confirms the key DEP findings: that "Investment Scams" are the most financially damaging, total losses are trending upwards, and the 65+ age group is the most financially vulnerable.

Reflecting on the work on this project, I learned how difficult, but essential it is to have follow-through across interactions that are complicated and coordinated. Getting the persistent cross-filtering to work helped teach me how to build a truly fluid experience for the user, rather than just building a bunch of standalone charts. If I had a do-over, I would have just made a few more iterations on the data structure on the more complex charts, particularly before diving into the implementation, because building that data structure "on-the-fly" was very taxing. For future work, it would be worth slowly building out this tool overtime to include data from 2024 and onwards to see newly emerging trends, or perhaps, a new tab with visualizations of scam outcomes (i.e recovery of funds or other outcome metrics) to create a more complete tool for policy decision making.

5. Bibliography

Australian Bureau of Statistics. (2018). ***Socio-Economic Indexes for Areas (SEIFA), Australia, 2016*** (Catalogue No. 2033.0.55.001) [Data set]. https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001

Australian Bureau of Statistics. (2023). Regional population by age and sex, 2023 [Data set]. https://www.abs.gov.au/statistics/people/population/regional-population-age-and-sex/2023

Australian Competition and Consumer Commission. (n.d.). Scam statistics [Data set]. https://www.scamwatch.gov.au/research-and-resources/scam-statistics

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). shiny: Web application framework for R (Version 1.8.1.1) [Computer software]. RStudio. https://shiny.rstudio.com/

Cheng, J., Karambelkar, B., & Xie, Y. (2023). leaflet: Create interactive web maps with the Leaflet JavaScript library (Version 2.2.1) [Computer software]. RStudio. https://rstudio.github.io/leaflet/

Munzner, T. (2014). Visualization analysis and design. CRC Press.

R Core Team. (2023). R: A language and environment for statistical computing (Version 4.3.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

South, A. (2017). rnaturalearth: World map data from Natural Earth (Version 1.0.1) [Computer software]. Comprehensive R Archive Network. https://CRAN.R-project.org/package=rnaturalearth

Tufte, E. R. (2001). The visual display of quantitative information (2nd ed.). Graphics Press.

Wickham, H., François, R., Henry, L., & Müller, K. (2023). dplyr: A grammar of data manipulation (Version 1.1.4) [Computer software]. Comprehensive R Archive Network. https://dplyr.tidyverse.org

Wickham, H., & Vaughan, D. (2024). tidyr: Tidy messy data (Version 1.3.1) [Computer software]. Comprehensive R Archive Network. https://tidyr.tidyverse.org

6. Appendix

Sheet 1 - Brainstorming and Ideation

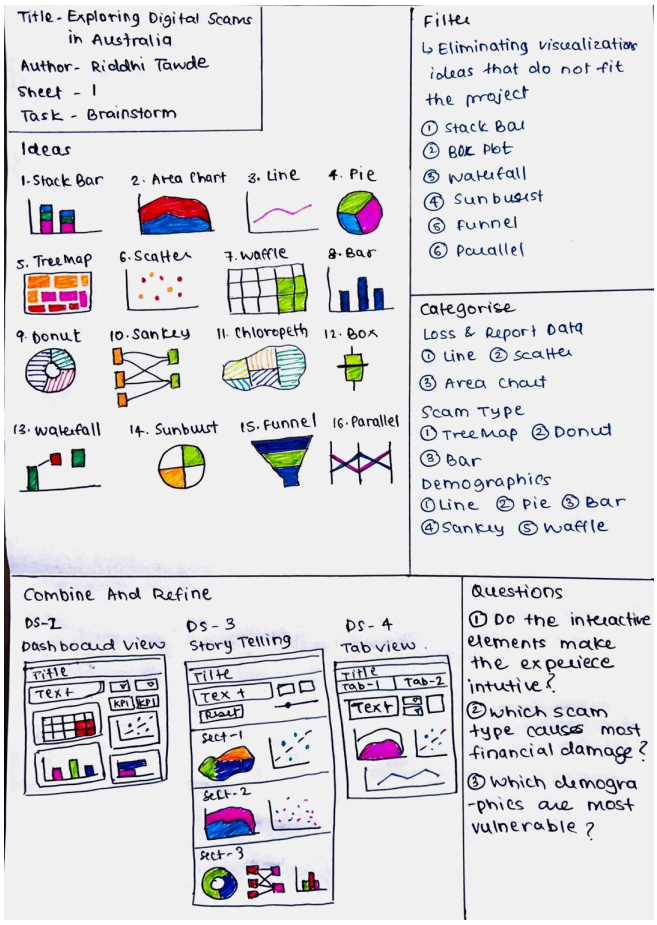

Figure A1: Design Sheet 1 - brainstorming, filtering, categorizing, refining ideas and information collection including data sources, task identification, and preliminary visualization sketches.

1. Title and author of the project.
2. Initial ideation and brainstorming of visualizations,16 chart types were considered.
3. Filtering out the charts that do not satisfy the project aims.
4. Categorising the charts based on three main ideas of the project:- Loss & Reports, Scam Type and Demographics.
5. Combine and refine these ideas to solve the project aim in three different visualizations.
6. Initial Guiding Questions: The design process was guided by key questions:
   - "Do the interactive elements make the experience intuitive?"
   - "Which scam type causes the most financial damage?"
   - "Which demographics are most vulnerable?"

Sheet 2 - (Design Alternative - 1) Single Page Dashboard View

Figure A2: Design Sheet 2 - layout for a single-page dashboard. The design features global filters for State and Year, KPIs at the top, and all charts (Loss by Gender/Age, Loss by Scam Type, Contact methods) on one screen.

1. Layout sketch: A single page, “Dashboard view” - top KPIs, filter, and chart area.
2. Filter panel design: dropdown filters for State, Year, and a "Metric Toggle" for either Loss/ Report, and a Reset button.
3. Charts specifications:
4. KPIs - Total Loss, Total no. of Reports.
5. Charts - Loss by Gender and Age, Loss by Scam Type and Methods of Contact.
6. Interaction notes: filter data (the filters will affect every chart), "Hover for Details" (with tool tips), and Reset ALL (to clear all filters).
7. Rationale (Discussion):
    - Advantages: simple, clear filter, bar charts are great for simple comparison, fast, broad overview.
    - Disadvantages: can lead to information overload, still limited detail smaller charts, although still limited space overall.

Sheet 3 - (Design Alternative - 2) Storytelling View

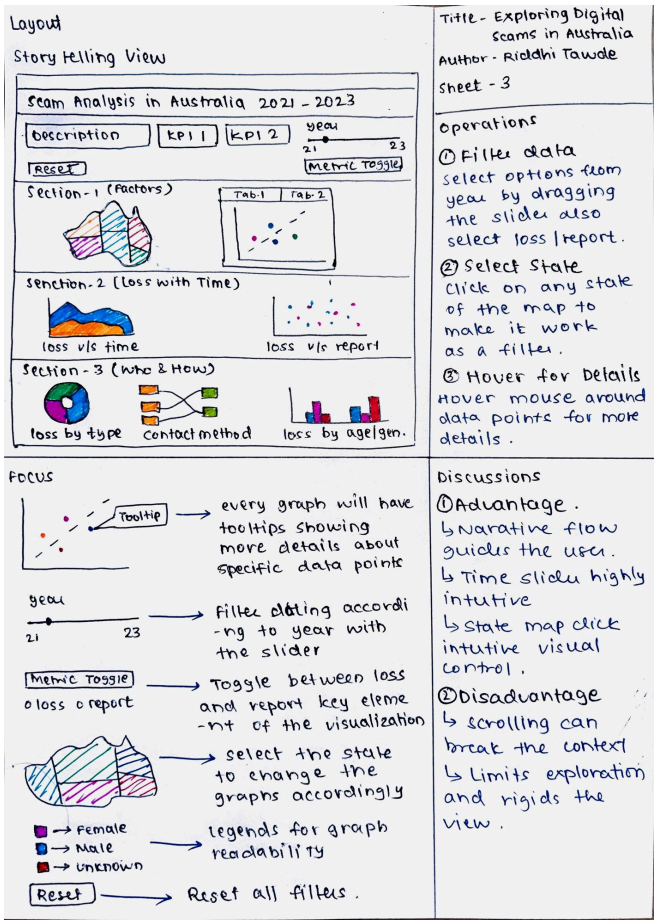

Figure A3: Design Sheet 3 illustrates a vertical, narrative-driven layout. It's divided into thematic sections ("Factors", "Loss with Time", "Who & How") and features interactive controls like a year slider and a clickable map for filtering.

1. Layout sketch: A vertical, section-based layout divided into "Section-1 (Factors)" , "Section-2 (Loss with Time)" , and "Section-3 (Who & How)".
2. Filter panel design: An interactive slider for Year and a clickable map to filter by state.
3. Metric toggle: A toggle between "loss vis time" and "loss vis report" .
4. Interaction notes: Filter by dragging the year slider, filter by clicking on the map, "Hover for Details" on data points and a Reset button.
5. Rationale (Discussion):
    - Advantages: Narrative flow guides the user , the time slider is "highly intuitive" , and the clickable map is an "intuitive visual control".
    - Disadvantages: Scrolling can "break the context", the guided view "Limits exploration", and presents a "rigid" view.

Sheet 4 - (Design Alternative - 3) Tab-Viewed Dashboard

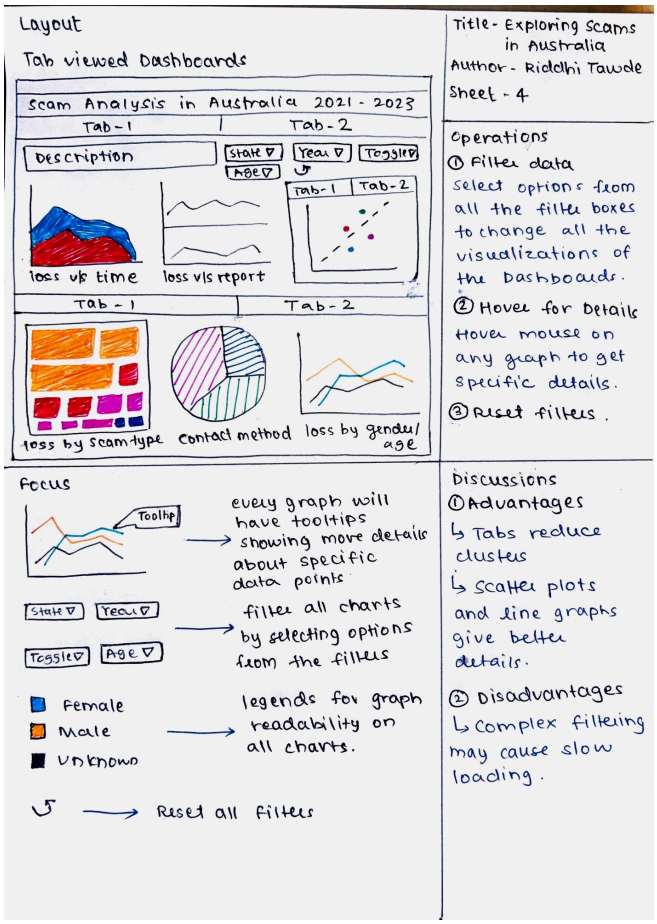

Figure A4: Design Sheet 4 shows a dashboard design that uses tabs to organize visualizations. It features global filters for State, Age, and Year, with charts separated into two tabs to reduce clutter.

1. Layout sketch: The dashboard will be set up with a clear division in content between "Tab-1" and "Tab-2" sections.
2. Filter panel design: The global filters used will be: State, Age, and Year.
3. Chart Design: The tab content will each have primary charts that are organized by 'loss vis time' and 'loss vis report', while other charts will be guided by visually sorting for scam type, contact method, and demographics.
4. Interaction notes: All filters will update and any time the information changes, we will add"Hover for Details", with a Reset option too.
5. Rational (Discussion)
   - Advantages: Tabs "decrease clutter" (the cluster of content) and this layout will allow for "better details" for scatter plots and line graphs.
   - Disadvantages: "Complex filtering" could slow loading times.

Sheet 5 - Final Refined Design - Realisation

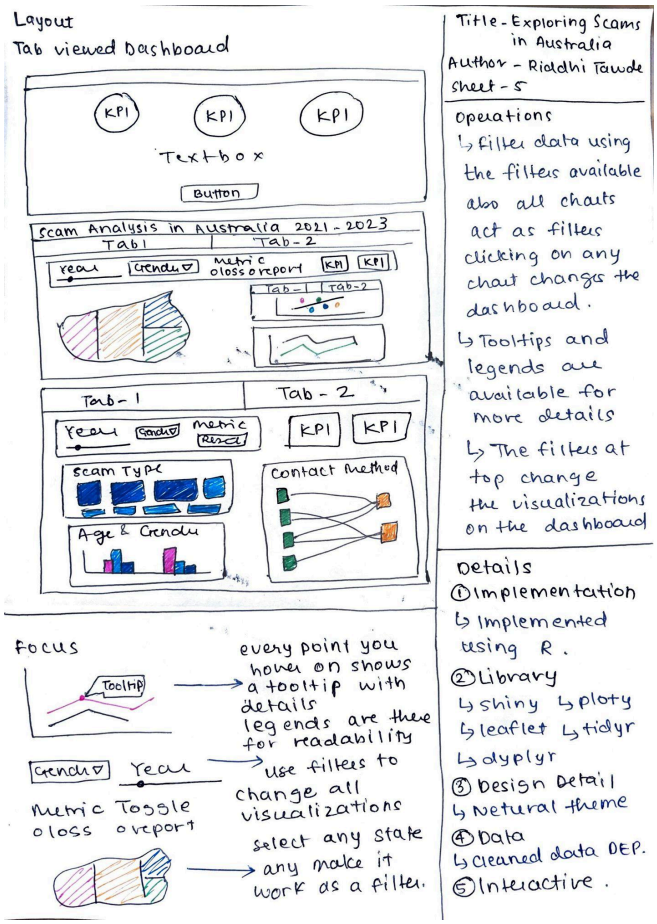

Figure A5: Final Refined Design and Implementation Plan

Dashboard Layout:

- Overall Structure: A "Tab viewed Dashboard" titled "Scam Analysis in Australia 2021-2023".
- Header: A top-level section is sketched with three "KPI" boxes, a "Textbox", and a "Button".
- Tab 1:
    - Filters: Includes dropdowns for "Year" and "Gender", plus a "Metric oloss oreport" toggle.
    - Visualizations: Contains a "Scam Type" (bar chart) and an "Age & Gender" chart.
- Tab 2:
    - KPIs: Includes two "KPI" boxes.
    - Visualizations: Contains a "Contact Method" chart (sketched as a parallel coordinate plot or Sankey diagram).

Interaction Specifications:

- Linked Filtering: All charts "act as filters"; clicking on any chart will update the entire dashboard (cross-filtering).
- Global Filters: Dropdown filters at the top change all visualizations on the dashboard.
- State Filtering: A user can "select any state" to make it "work as a filter" (implying a map, like the one mentioned for leaflet).
- Hover Tooltips: "every point you hover on shows a tooltip with details".
- Metric Toggle: A toggle allows the user to switch the view between "loss" and "report" data.
- Legends: "legends are there for readability".

Visual Design Details:

- Color Scheme: A "Neutral theme" is specified for the design.
- Typography: A clear, legible sans serif.

Implementation Notes:

- Framework: Implementation will be "using R".
- Libraries: The project will use shiny, ploty, leaflet, tidyr, and dyplyr.
- Data: The project will use "Cleaned data".
- Interactivity: The plan is "Interactive", reiterating the use of tooltips and filters.