Massey University

158741

Location Data: Mapping, Analysis and Visualisation

Which parts of the New Zealand road network are worst affected by traffic accidents, and what other attributes are accidents correlated with?

Luis Vieira

23012096

June, 2024

Index

Literature Review

Traffic accidents remain a critical global public health and safety challenge, with the World Health Organisation (WHO, 2023) reporting 1.19 million road fatalities (2023). Achieving the goal of halving road traffic deaths by 2030 requires urgent action. In New Zealand "average social cost is estimated at NZD$4.916 million per fatal crash", highlight the economic impact (Te Manatū Waka, 2022). This literature review examines how Geographic Information Systems (GIS) can enhance road safety through geospatial analysis of traffic accidents.

GIS and Traffic Accidents

The application of GIS in analysing traffic accidents is based on two key perspectives: spatial and temporal aspects. Time and space are fundamental in GIS analyses of traffic accidents. Spatial factors such as road type, weather, population density, culture, and the facilities distribution significantly influence traffic accidents (Budiawan & Purwanggono, 2018). GIS integrates various layers of information and digital maps, allowing for comprehensive spatial analysis and mapping of transportation-related events. GIS can visualise traffic accidents spatial distribution through various maps, which is an effective method for spatial-temporal analysis.

Spatiotemporal Analysis

GIS plays a vital role in traffic accident analysis by facilitating the examination of both spatial and temporal aspects of accidents (Steenberghen et al., 2004). Spatiotemporal analyses identify the causes and consequences of traffic accidents by examining where and when accidents occur. These analyses often reveal non-random patterns, essential for hotspot detection and preventive measures (Kaygisiz et al., 2015).

Spatial Analysis

The spatial distribution of accidents is typically defined using GPS data from traffic databases (Gundogdu, 2010). Various studies have employed methods like Kernel Density Estimation (KDE) and Nearest Neighbourhood Distance for spatial analysis using GIS software (Liu & Zhu, 2004). GIS enables spatial integration, analysis, visualisation, and other functions essential for traffic analysis and urban planning. Parameters like traffic speed, volume, density, and socioeconomic elements can be joined via GIS to demonstrate their relationship with traffic accidents (Ng et al., 2002).

Spatial analysis techniques have been used to identify accident hotspots and contributing factors. For instance, Satria and Santoso (2018) applied GIS-based spatial analysis to identify accident hotspots in Jakarta, using KDE to visualise accident density and assess the impact of road conditions, traffic volume, and environmental factors. Afolayan et al. (2022) identified accident hotspots on a highway in Nigeria, emphasising the importance of targeted preventive measures. Anderson (2009) used GIS and KDE to study spatial patterns in London, providing insights for road safety improvements.

The effectiveness of various GIS-based methods for identifying and ranking road traffic accident hotspots has been evaluated. Zahran et al. (2019) compared network KDE+, Getis-Ord Gi*, and a risk-based method accounting for RTA frequency, severity, and socioeconomic costs (STAA). They found that KDE+ is suitable for network-restricted incidents but has limitations at intersections, while Getis-Ord Gi* is useful for area-wide incidents but less effective for linear road networks, whereas STAA is thorough but can overestimate risk levels.

Temporal Analysis

Temporal analysis of accidents provides a long-term evaluation of accident trends, linking accident time with other influencing factors like traffic flow and weather conditions (Bíl et al., 2013). Temporal analysis is conducted on various scales, including daily, weekly, monthly, or seasonally, depending on the study's purpose (Rodríguez-Morales et al., 2013). Visualisation approaches for temporal analysis include time series, graphs, and spider plots. Nguyen and Ho (2023) identified road traffic accident hotspots in Hanoi and Prasannakumar et al. (2011) in India, using GIS-based temporal-spatial statistical analysis techniques. They focused on the spatial distribution of accidents and contributing factors, recommending targeted traffic management strategies.

GIS Methods for Traffic Accidents Analysis

GIS significantly simplifies the detection of traffic accident hotspots in transportation networks. Integration with GPS enhances GIS capabilities, providing a robust tool for complex multicriteria analysis (Zahran et al., 2019).

Hotspots Analysis Methods

Various statistical methods have been developed to identify traffic accident hotspots, concisely reviewed by Lord and Mannering (2010), their benefits and limitations. Multiple linear regression, Poisson regression, and negative binomial regression models are commonly used to establish accident event models and identify significant contributing factors (Ng et al., 2002). Poisson and negative binomial regression models often perform better in predicting accident patterns and developing accident event models (Abdel-Aty & Radwan, 2000).

Dereli and Erdogan (2017) applied Poisson regression, Negative Binomial regression, and Empirical Bayes methods to identify the relationship between influential features and the number of events occurrences. Erdogan et al. (2008) combined Poisson regression with kernel density analysis to find accident clusters. Pourroostaei Ardakani et al. (2023) used machine learning techniques to propose a predictive model for road car accidents, identifying key predictive features such as weather conditions and road types.

Multiple case studies demonstrate the practical applications of GIS in traffic accident analysis. Kraft et al. (2023) modelled the risk levels of road network sections for motorcycle transport in the Czech Republic, introducing the Road Network Hazard Index (RNHI) combining accident probability and EMS accessibility. Haynes et al. (2008) investigated the relationship between road curvature and fatal crashes in New Zealand, finding no significant evidence that curved roads had higher crash rates. Flahaut et al. (2003) compared local spatial autocorrelation (Moran's I) and KDE in identifying accident black spots, recommending combined methods for road safety analysis. Yannis et al. (2013) examined the impact of GDP changes on road traffic fatalities across various countries, highlighting the correlation between economic fluctuations and accident rates.

The "Road to Zero" strategy outlines New Zealand's comprehensive approach to reducing road deaths and serious injuries by 40% over the next decade, emphasising a holistic approach to road safety (Waka Kotahi NZ Transport Agency, 2020). It highlights the importance of reducing road deaths through continuous monitoring and strategic safety planning. The Waka Kotahi NZ Transport Agency (2017) provides an in-depth analysis of serious injuries resulting from road accidents in New Zealand. The study compiled and analysed serious injury data from various sources, identifying trends and patterns in serious injury occurrences, and evaluating contributing factors and risk profiles.

Sagar and Stamatiadis (2022) using spatial analysis of crashes in Kentucky state, analysed the relationship between safety and socioeconomic characteristics. Their findings revealed higher crash propensity in low income areas, and among certain age and gender driver groups, highlighting the importance of socioeconomic factors in traffic accident analysis.

GIS significantly improves analysis of traffic accidents, providing important insights for identifying hotspots and developing effective road safety programs.

Project Scope

The main objective of this project is to analyse Auckland Region's road network to identify areas that experience the highest frequency of traffic accidents from 2018 onwards, and to investigate the attributes most correlated with these accidents. The geographic area of the study is the administrative Auckland Region, including all its administrative boundaries within the area. This approach ensures that a clear understanding of the traffic accidents is provided.

In this project, key variables will be analysed to understand their relationship and correlation with traffic accidents. These include the frequency, severity, and locations of accidents, all sourced from the Waka Kotahi NZTA Crash Analysis System (CAS). Additional variables include weather conditions at the time of accidents, road conditions, daily average traffic volume and land use, focusing on points of interest like schools and commercial areas. Socio-economic factors such as population density, number of vehicles per household, and income levels from census data will also be considered.

The analysis employs geospatial and statistical techniques to identify accident hotspots and assess correlations with various attributes. The main tool QGIS being used for geospatial analysis and visualisation, while statistical analysis and data wrangling is conducted in a Jupyter notebook in Python. These tools help uncover significant relationships between traffic accidents and correlation levels with attributes.

By exploring these factors, the project aims to provide insights that could inform targeted interventions to improve road safety in the Auckland region.

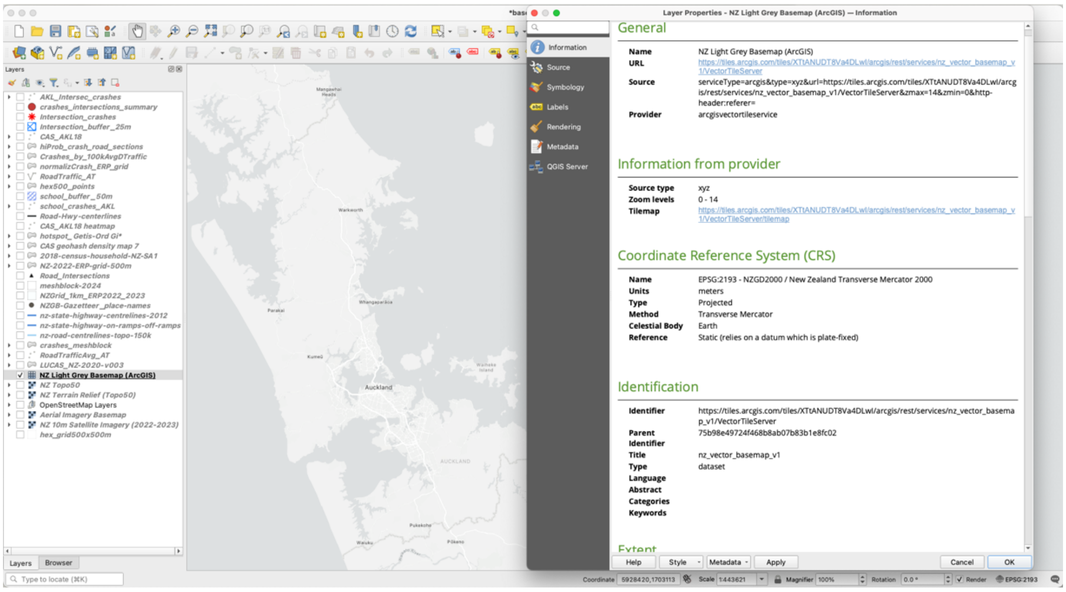
Data Needs

In order to thoroughly investigate the research question, I used various datasets from multiple sources and in different data formats (shapefiles, csv, text, WMS, WMTS) were used. The following datasets were used to provide a detailed analysis of traffic accidents and their correlating attributes.

Basemaps (ArcGIS, WMTS/WMS web services using Koordinates plugin)

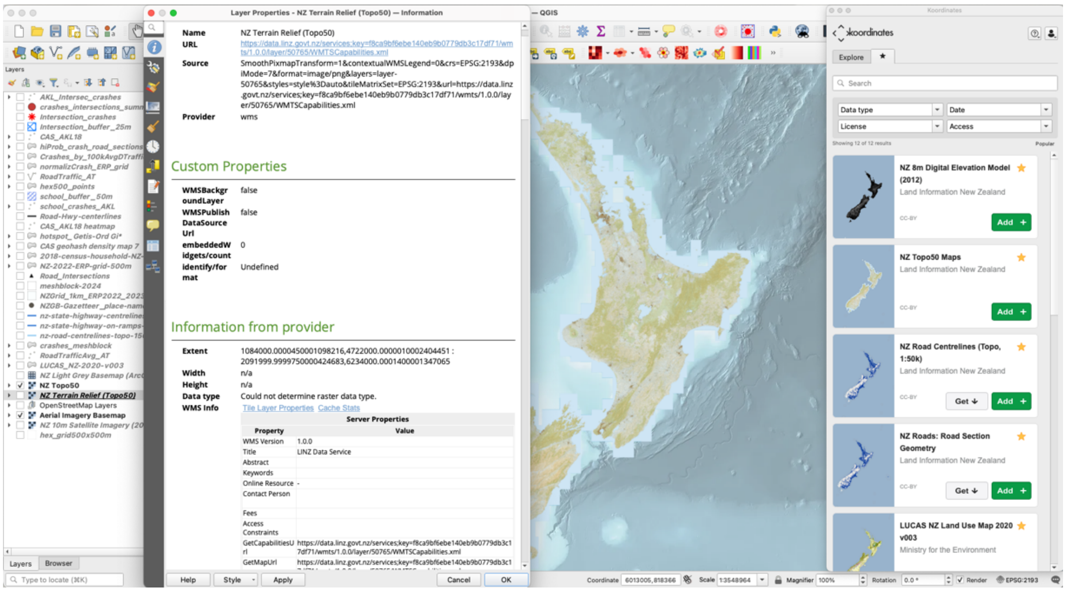
- NZ Light Grey Basemap: downloaded from ArcGIS vector tile server, added as a Vector Tile Layer.

Figure 1 - NZ Light Grey Basemap downloaded from ArcGIS updated with NZGB Place Names



- LINZ NZ Topo50 WMTS, NZ Terrain Relief (Topo50) WMTS, NZ Aerial Imagery Basemap WMS, and NZ 10m Satellite Imagery (2022-2023) WMS were all loaded using the Koordinates plugin/API.

Figure 2 – Basemaps NZ Topo50 and Aerial Imagery overlaid layers and Koordinates plugin API (right side)



Road Accidents Data (NZTA)

- Waka Kotahi (NZTA) Crash Analysis System (CAS): downloaded in CSV and includes detailed information on traffic accidents reported by NZ Police. It was preprocessed in a Python notebook before being loaded as a Delimited Text (point) Layer into QGIS, keeping only 54807 crashes from 2018 onwards.

Figure 3 – NZTA CAS dataset being uploaded to QGIS as Delimited Text

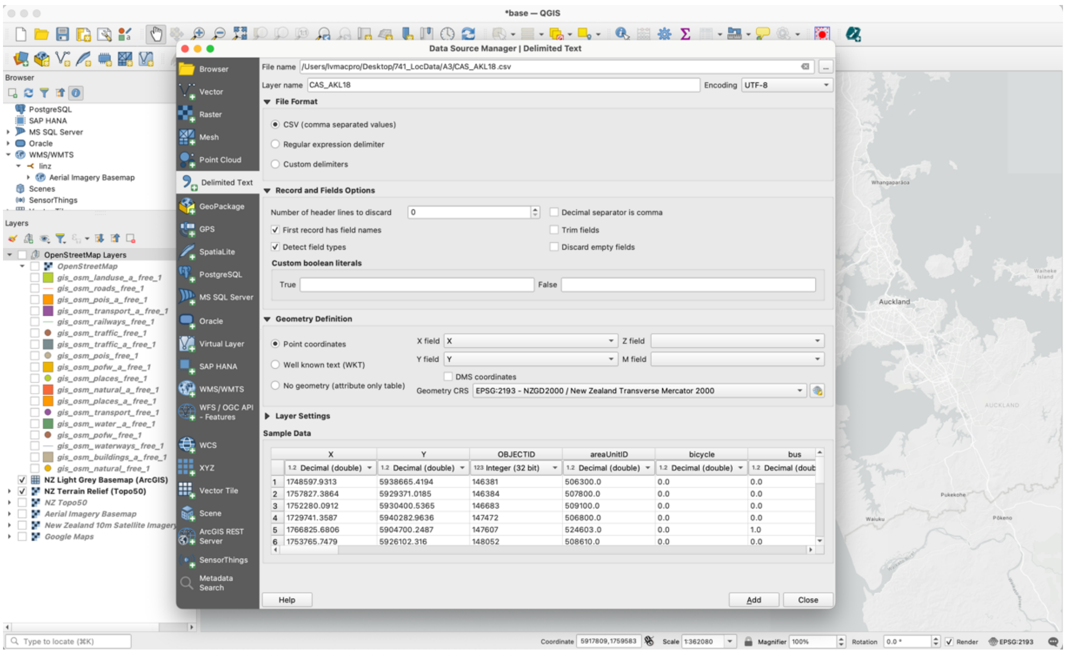
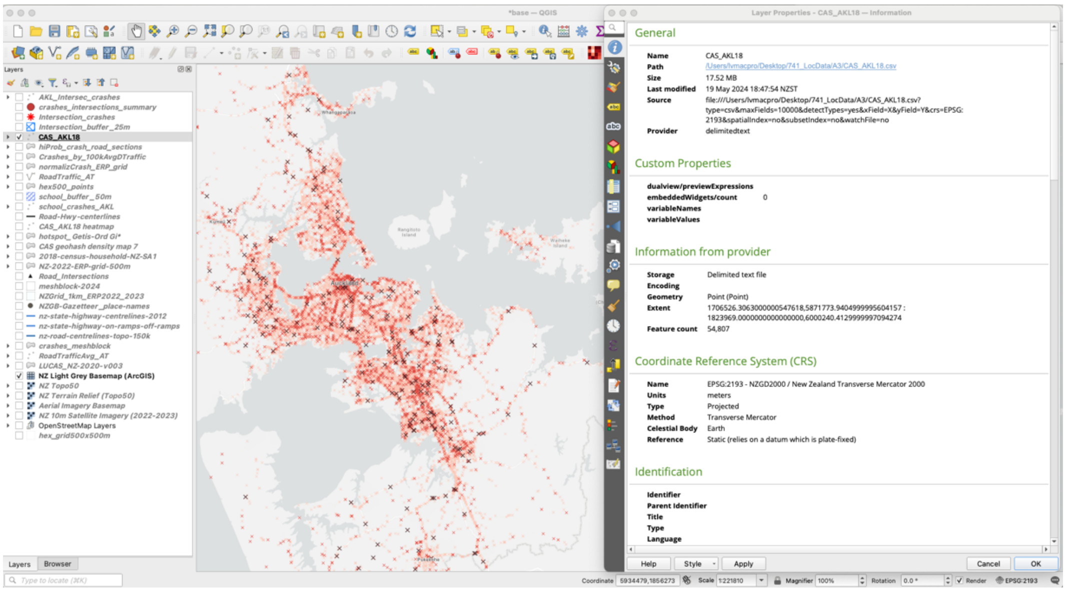


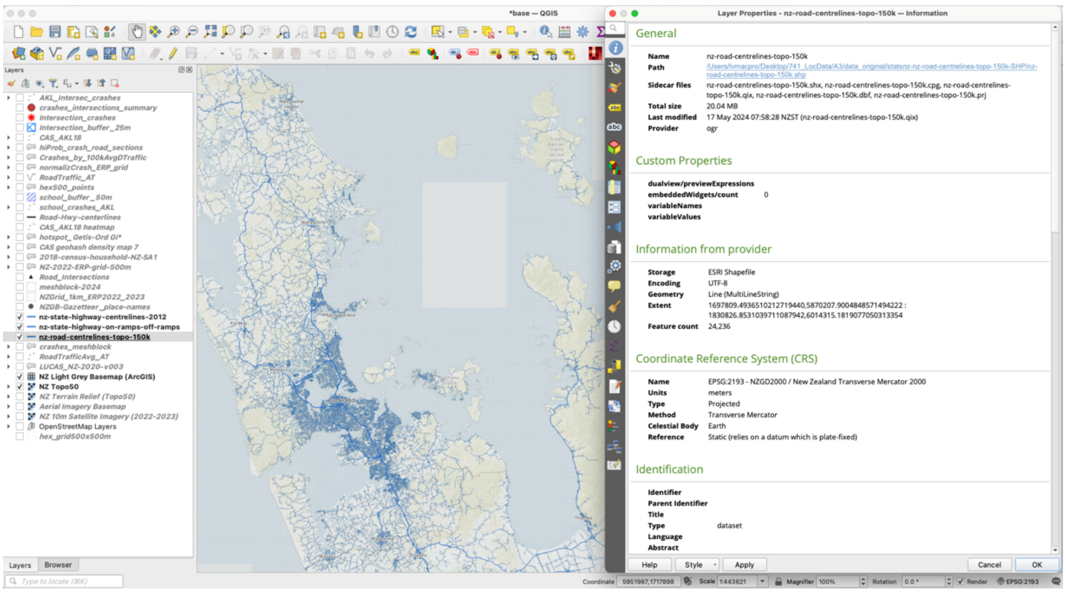
Figure 4 – NZTA CAS location points mapped as graduated color crosses based on Severity Index



Road Network Data and Place Names (LINZ)

- NZ State Highway Centreline (2012): downloaded as a shapefile, includes the central lines of state highways for mapping the primary road network.
- Highway On/Off-ramp Centreline: complements the highway centerlines.
- Road Centreline Data (Topo 150K): shapefile with road network centerlines.

Figure 5 – LINZ Highway, Ramps and Road Network Centerlines

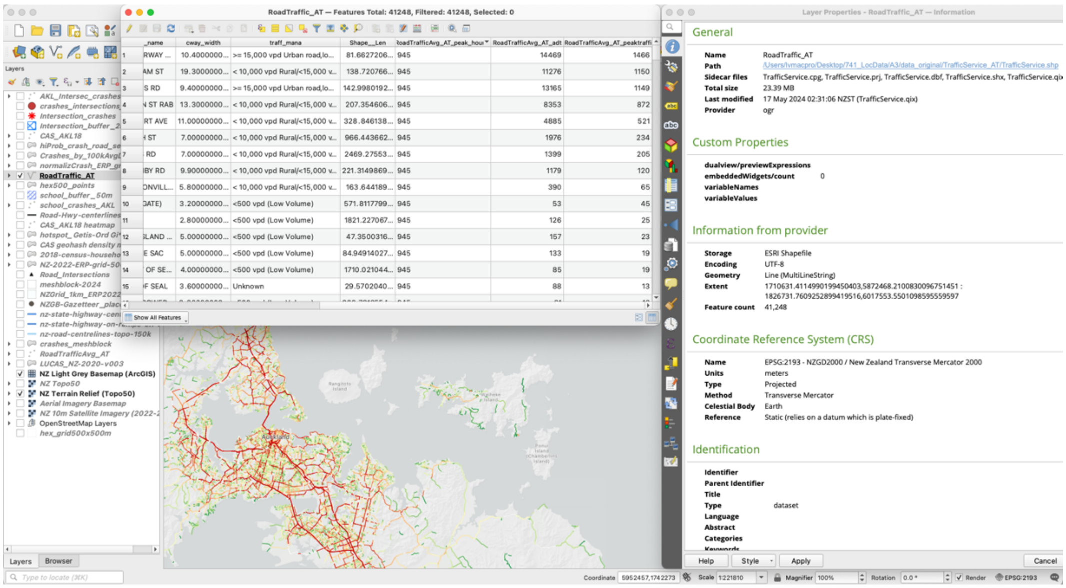


- NZGB Gazetteer Place Names: Loaded to manually update place names in the NZ Light Grey Basemap from ArcGIS.

Traffic Volume Data (Auckland Transport)

- Traffic Service Layer: line geometry layer includes road sections.
- Traffic Service Layer: this point data includes daily traffic averages, for traffic density insights.

Figure 6 – Auckland Transport Road Traffic Service Joined Datasets

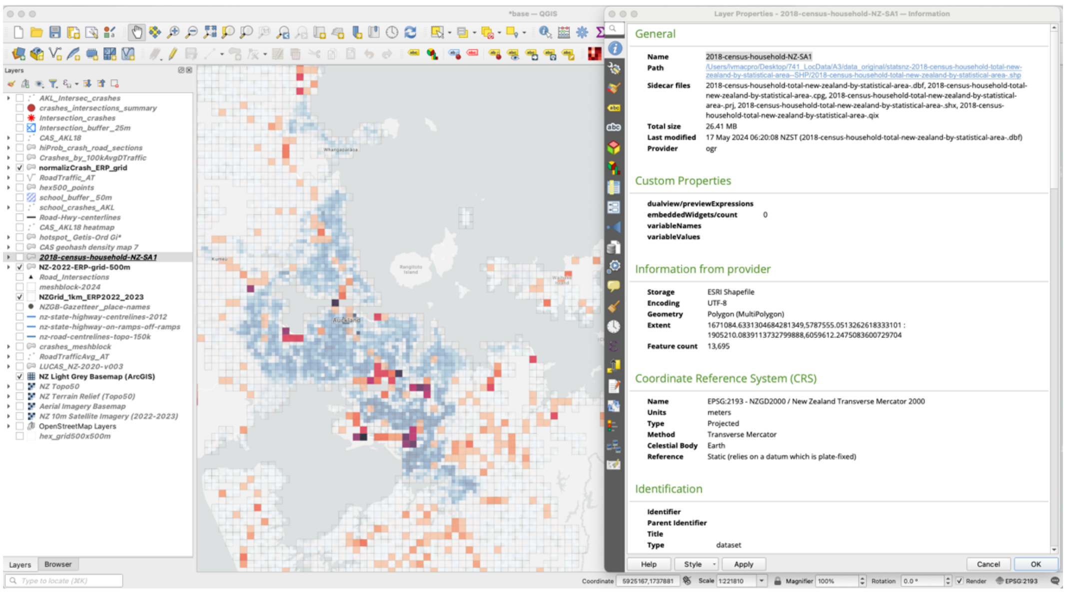


Note. The road network is incomplete, thus color graduated road sections (green, yellow, red based on daily average traffic volumes) are only shown for present data.

Census and Population Data (StatsNZ)

- New Zealand 2022 Estimated Resident Population Grid (500 metre).
- NZ Grid 1 km ERP (2022-2023): population estimates at a 1 kilometer resolution.

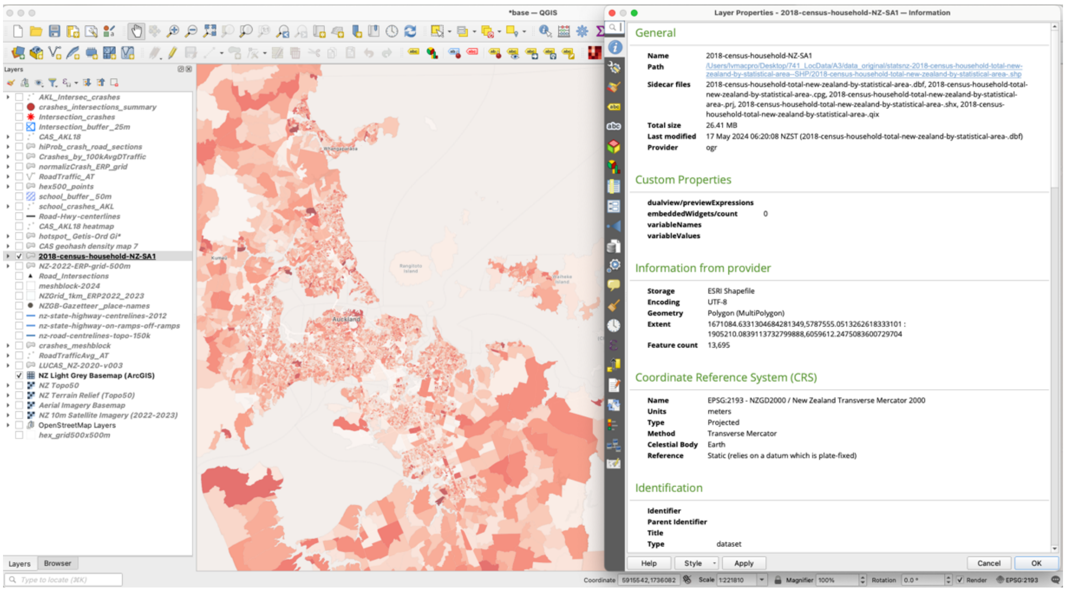
Figure 7 – StatsNZ ERP Grid (500m) and CAS Severity Normalised by ERP Grid (1Km)



Note. Normalising accidents frequency and severity by population density does not produce interesting results as darker red grid blocks appear to be industrial areas and highway bridges. Blue gradient 500m grid shows the ERP, thus high density populated areas

- 2018 Census Household NZ SA1: detailed census data, including socioeconomic and n. vehicles per household.

Figure 8 – StatsNZ Census Vehicles per Household Data (darker red means higher number of vehicles)

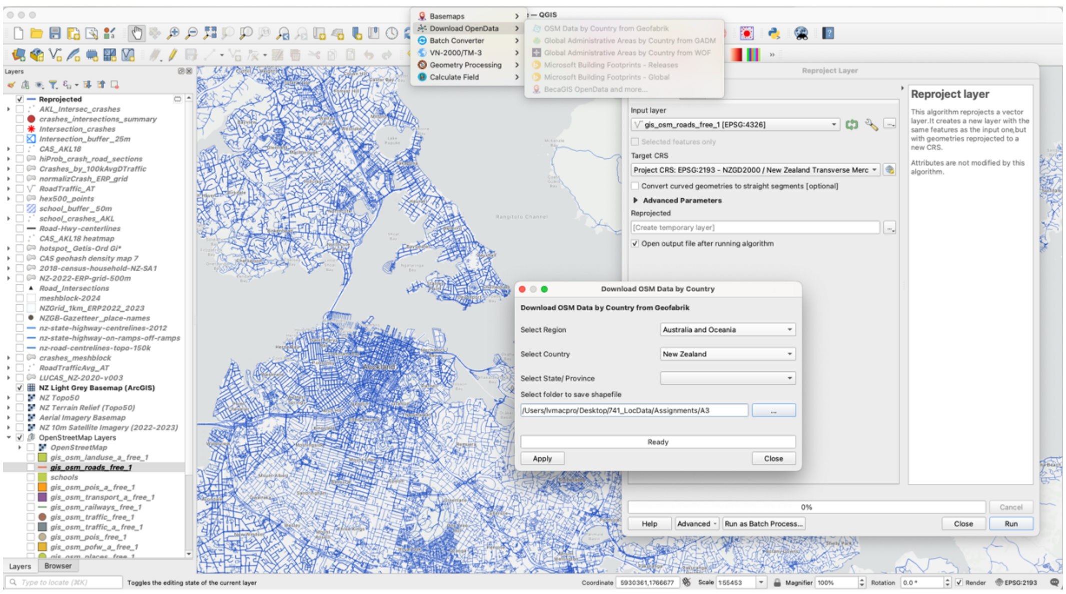


- 2024 Meshblock: polygon layer of NZ's smallest geographic unit for statistical data.

Road Networks and POIs (Open Street Map (OSM))

- Road Networks and POIs: Downloaded using the HCMGIS plugin, these 2 layers include detailed road networks and points of interest (POIs). The data was reprojected from EPSG:4326 - WGS 84 to EPSG:2193 to match the project CRS, ensuring consistency with other spatial datasets.

Figure 9 – OSM Road Network Layer Loaded via Plugin and CRS Reprojection



Data Quality

Data quality is crucial to the integrity of the analysis. Several factors are considered:

Completeness: Acknowledging the presence of missing attributes data. For example, the rich CAS dataset includes fields for the location and severity of accidents. However, it is limited in temporal data, only having the year available. Details about drivers' age, race, gender, or socio-economic information are not included, likely due to privacy reasons. The dataset also has attributes with missing data and some input errors. To address some of these issues, initial data wrangling and cleaning was performed in a Python Jupyter notebook. Having no temporal data besides the year I will use the attribute "Light" to extrapolate the time of day of the accident, whether it was during the day or at night.

Consistency: Checking for uniformity in data formats and units. All spatial datasets are or reprojected to the same CRS (EPSG:2193) to ensure consistency.

Accuracy: Recognising that layers are abstractions and simplified representations of reality. There may be measurement errors, outdated information, and missing or distorted attributes due to privacy and ethical considerations. Police reports may introduce location distortions for privacy reasons, which can affect spatial accuracy. For example, road centreline data may not always match reality, with many minor roads missing. Using data from multiple entities collected by different people with various tools inevitably leads to some distortions and approximations.

Figure 10 – Centerlines mismatch, and missing for smaller roads

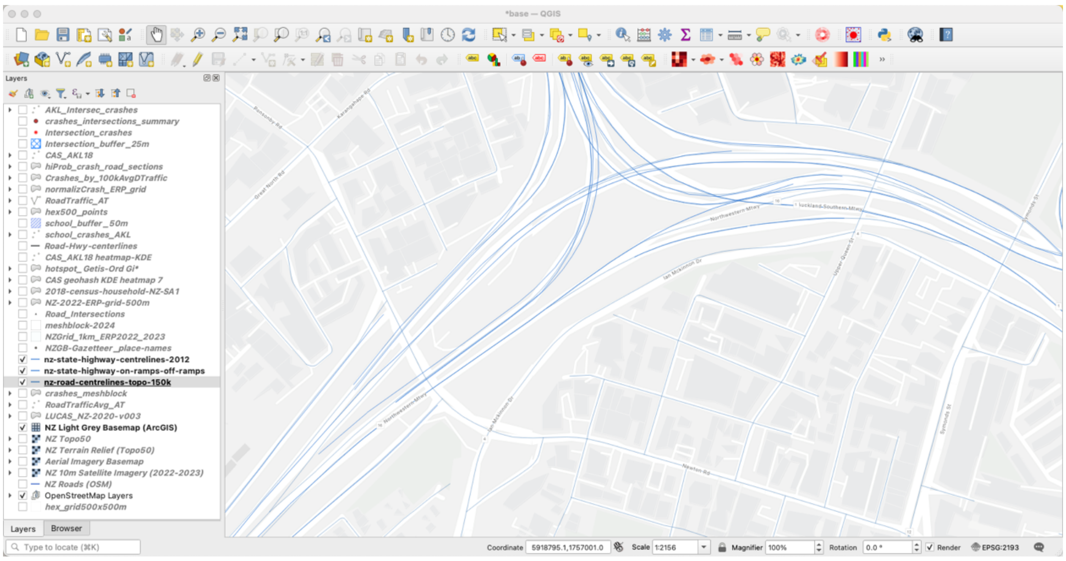
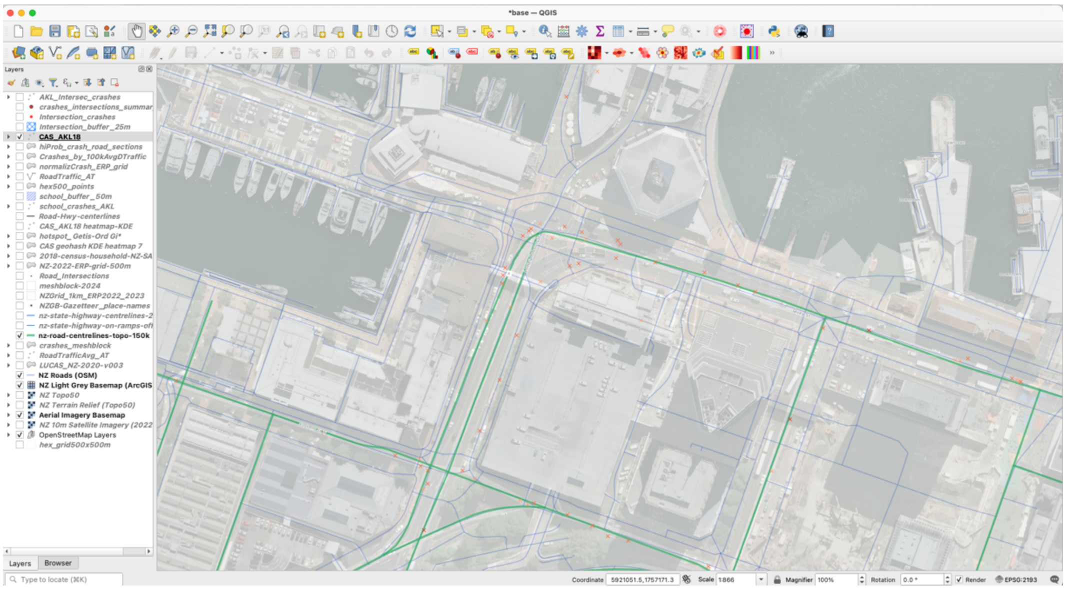


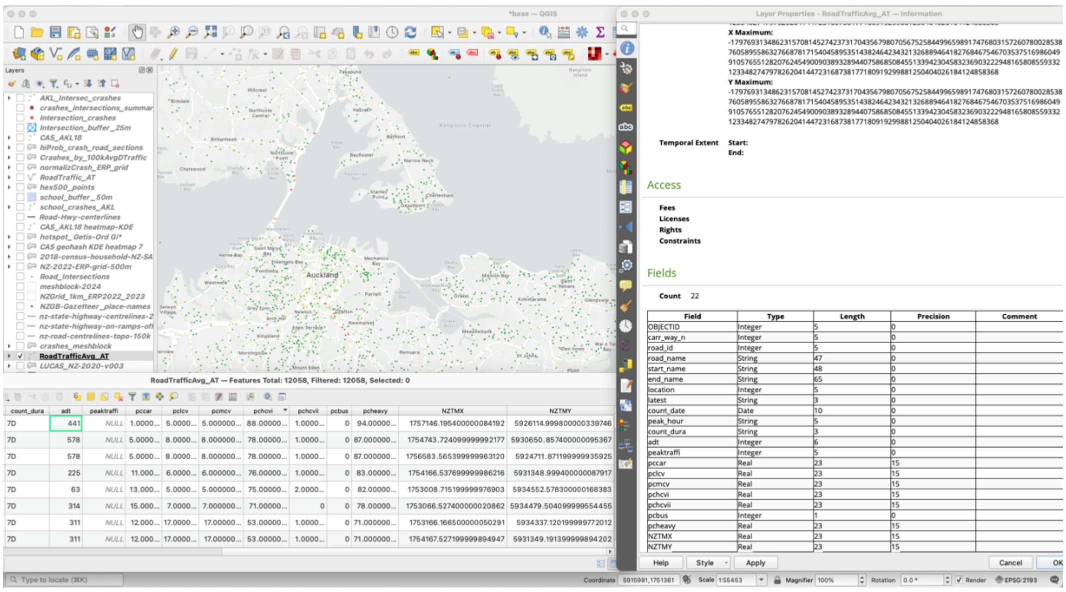
Figure 11 – Auckland CBD – Ferry Terminal Intersection Data Quality



Note. In green are LINZ Road Centerlines clearly simplified in the intersection, contrasting with OSM roads that include more detailed data such as walking paths but show some issues around the piers. Traffic accidents identified as crosses colored by severity (red gradient).

Precision: The reported precision in layers may not truly represent the actual precision. For instance, the high precision of road sections' coordinates from AT contrasts with rounded average daily traffic. There is a disparity between accident locations and road layers; LINZ road centreline data misses some intersections and minor roads, while OSM roads are richer in detail but raise reliability concerns as data is often collected by volunteers using non-standard measurements.

Figure 12 – Transit Service Precision Differences Between Location and Traffic Volumes and Averages



Timeliness: Ensuring data is up to date, the CAS dataset includes data from 2018 onwards latest being from early 2023 as complete reporting often take up to 7 months for non-injury accidents.

Privacy and Ethical Issues

Privacy and ethical considerations are important when dealing with sensitive information such as accident data. To maintain privacy, specific details about accident victims, like age and race are omitted ensure that individual identities are protected.

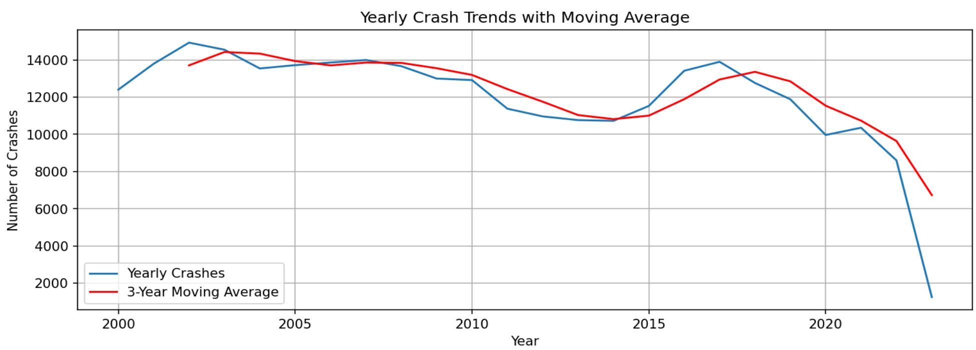
Geospatial Analysis

In this project, I employed several geospatial analysis methods to answer the research questions regarding traffic accidents and attributes correlated in Auckland Region from 2018 onwards.

Data Wrangling & EDA

I began by loading and preparing all relevant datasets into QGIS, ensuring each layer was projected to the same CRS (EPSG:2193). The CAS data required preprocessing due to missing values and inconsistent entries. Done in Python, I performed EDA and correlation analysis of attributes with accidents.

Figure 13 – Whole CAS data Accidents' Trend over the Years



Note. Gradual decrease of accidents over the past decades with a rebound before the 2018 Financial Crisis. Data is mostly incomplete from 2013 onwards.

Figure 14 – Severity of Accidents split from 2018 onwards

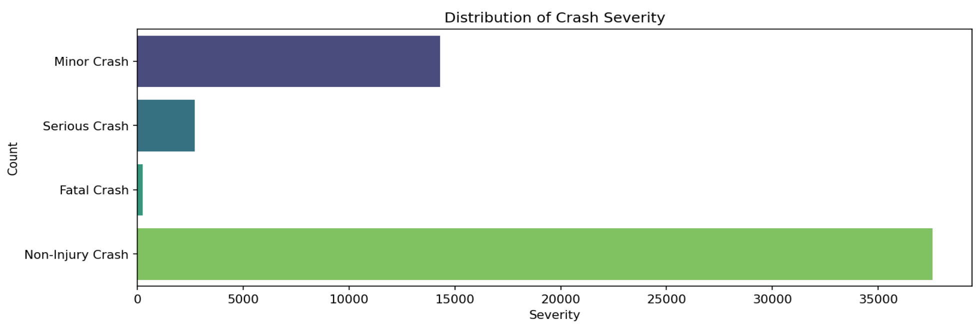
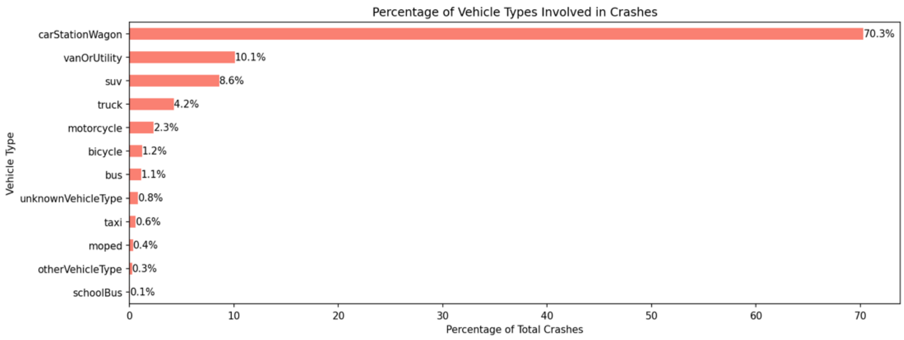


Figure 15 –Vehicle Types Involved in Accidents (percentage)



Large majority of accidents do not involve injuries and include cars/wagons, but I noticed an increase of utility vehicles and SUVs in more recent years.

Severity Index

Then, following literature I engineered a new severity index for better understanding of the accident data, using the frequency and severity of accidents. I considered using the Belgian severity index (Van Raemdonck & Macharis, 2014), but given the different split in our data (Fatal Crash, Serious Crash, Minor Crash, Non-Injury Crash) I opted for the Australian crash severity index (NSW Road Safety, 1999). However, I took a novel approach combining both the Belgian and Australian methods, incorporating crash weights to give higher importance to crashes with fatalities and serious injuries:

Crash Weights = 1 + (5 * fatalCount + 3 * seriousInjuryCount + 1 * minorInjuryCount) / 100

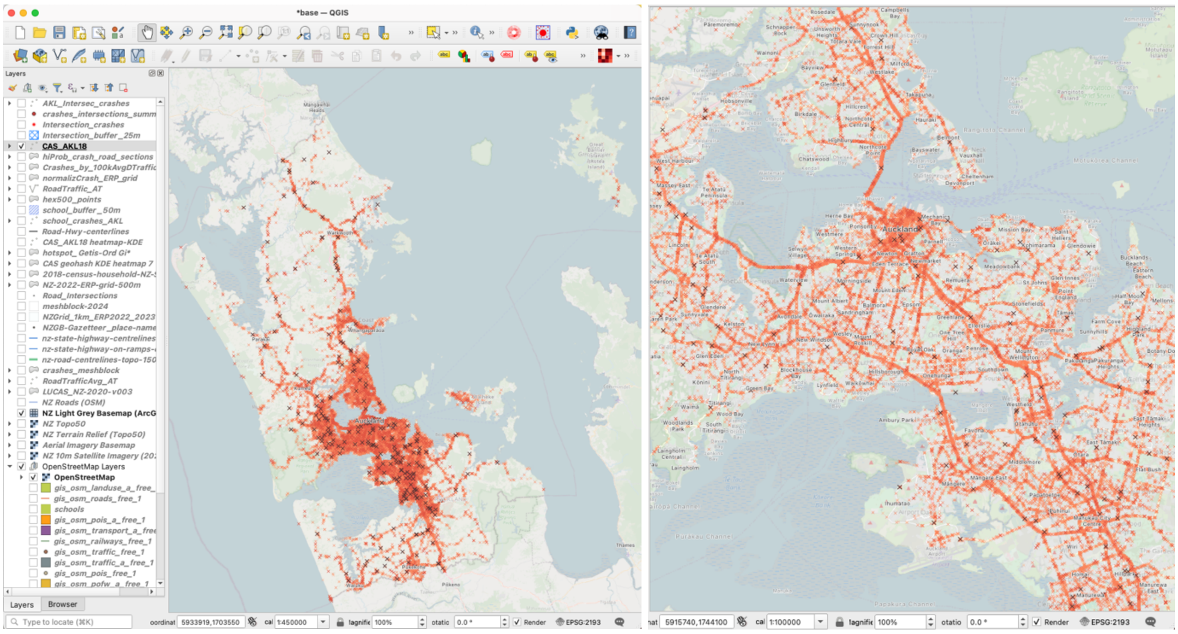
Severity Index = 3 * Fatal Crash * Crash Weights + 1.8 * Serious Crash * Crash Weights + 1.3 * Minor Crash + 1 * Non-Injury Crash

This ensures the severity index reflects both the type of crash and the number of injuries and fatalities involved, providing a more accurate measure of crash severity. Although there might be entries with missing correct inputs for the number of fatalities and injuries was assumed to be accurate.

Exploratory Data Analysis in QGIS

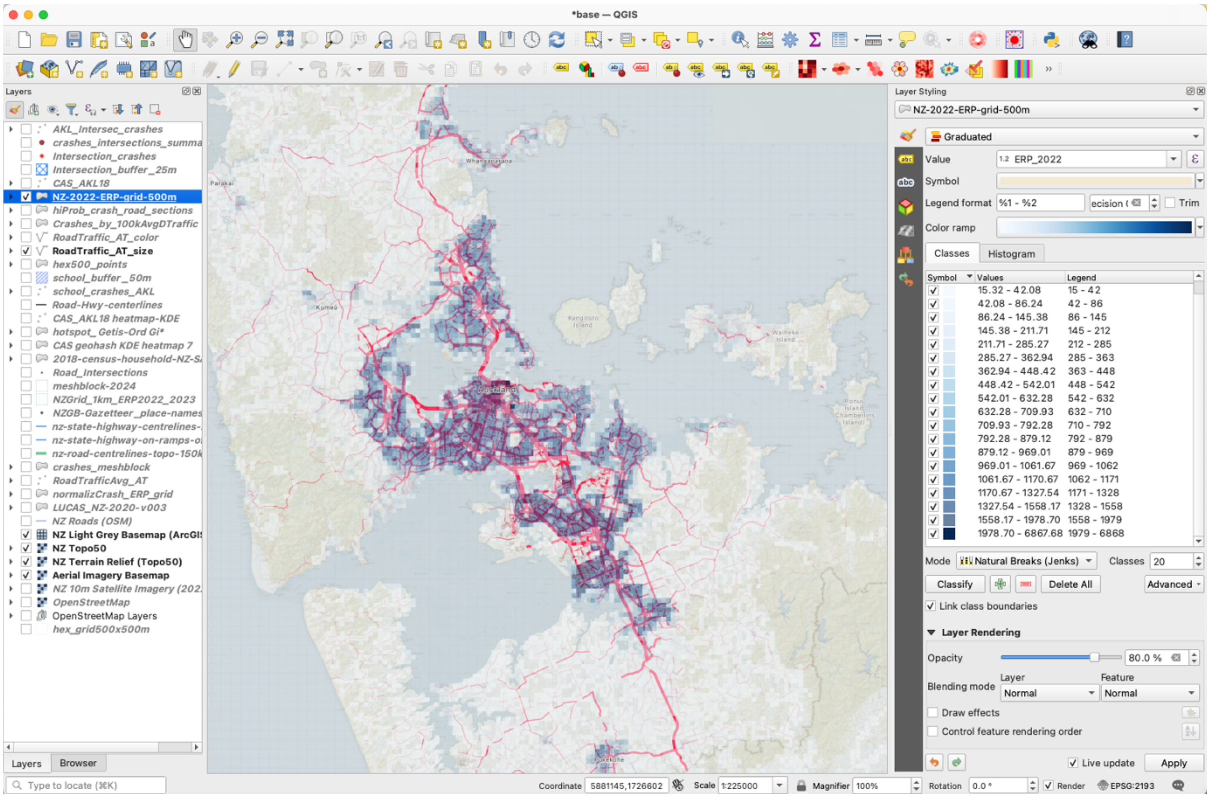
In QGIS, I visualised the CAS accident locations and created choropleth maps to get an initial sense of the distribution of accidents. The severity index was used to colour-code the accident points, highlighting areas with higher severity.

Figure 16 – Side by Side Accidents Location and Severity for Auckland Region and Close Up Partial city map



Note. Accidents are graduated based on severity Index (light orange to red, and black for fatal crashes) and plotted as crosses with size increasing based on crash severity.

Figure 17 – Traffic (red) and Population Density (blue) for Auckland Region (Partial)



Note. The blue gradient represents population density (darker means higher density) and in red the Traffic Density (thicker lines represent higher average daily traffic for that road section).

The main arteries (particularly highways) of Auckland clearly visible and mostly overlapping higher residential areas, closely matching the geographic distribution and density of the accidents in Figure 17.

Hexagon Grid

Then, I first created a hexagon grid 500m and joined this with the CAS data to map accident frequency. The hexagon grid method provides a clear representation of spatial distribution, allowing for better comparison across different areas. This method is effective in managing large datasets and providing a uniform spatial resolution for analysis (Satria & Santoso, 2018).

Figure 18 – Counting Accident Points 500m Hexagon Map of Crashes for Auckland Region (graduated by frequency and location)

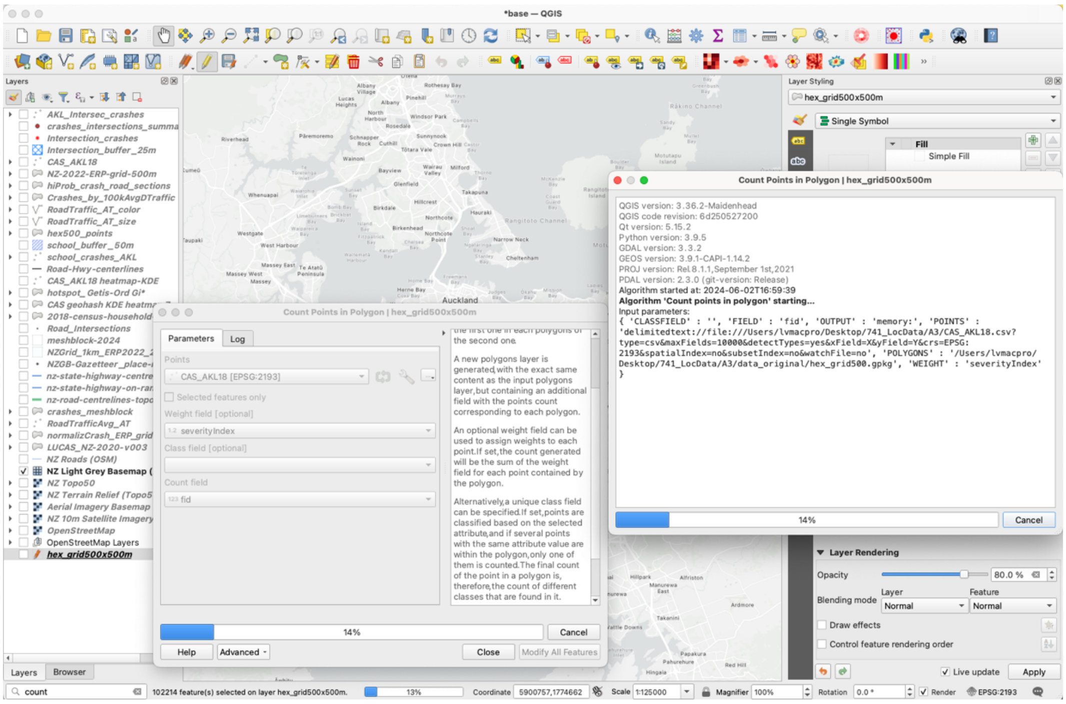
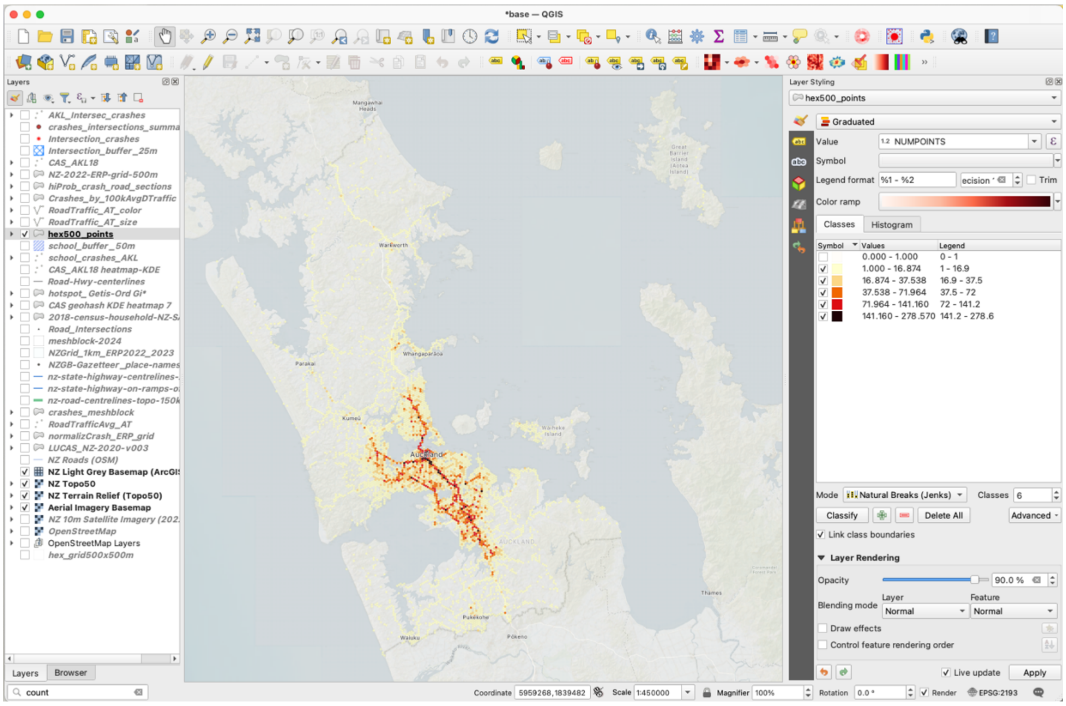


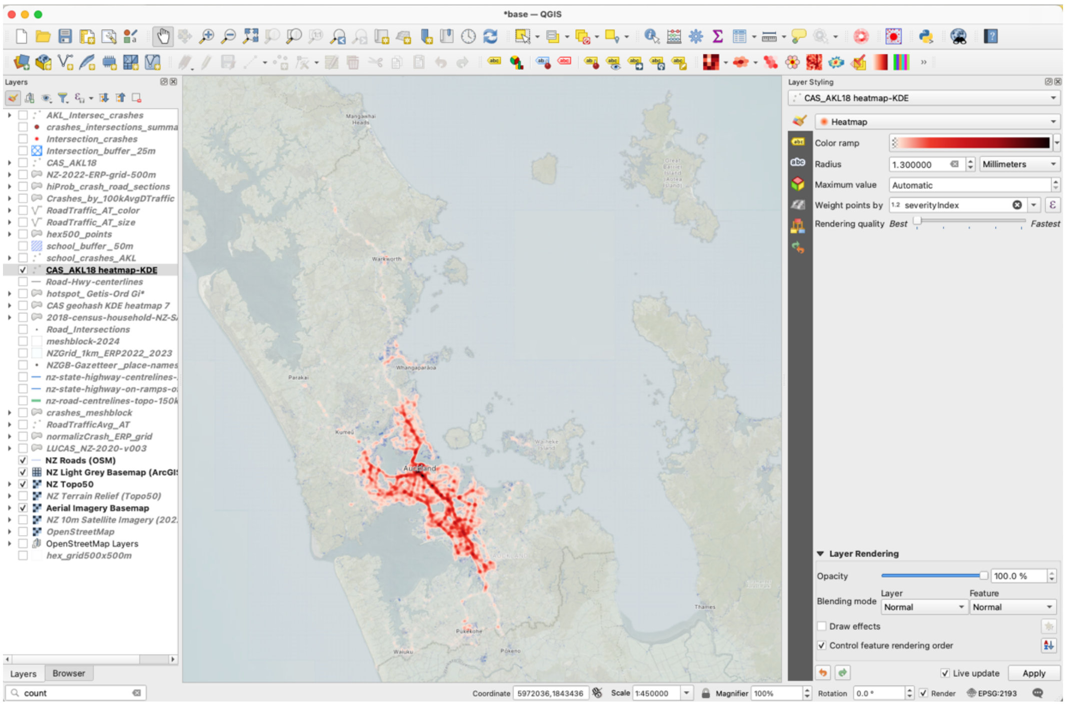
Figure 19 – 500m Hexagon Map of Crashes for Auckland Region (graduated by frequency and location)



Kernel Density Estimation (KDE)

Next, I performed a KDE to create a heatmap of accident density. KDE is commonly used in traffic safety studies (Anderson, 2009; Liu & Zhu, 2004) to identify high-risk areas. Alternatives such as simple point mapping were considered but were less effective in illustrating density and spatial patterns.

Figure 20 – KDE-Heatmap of Accident Density in Auckland Region

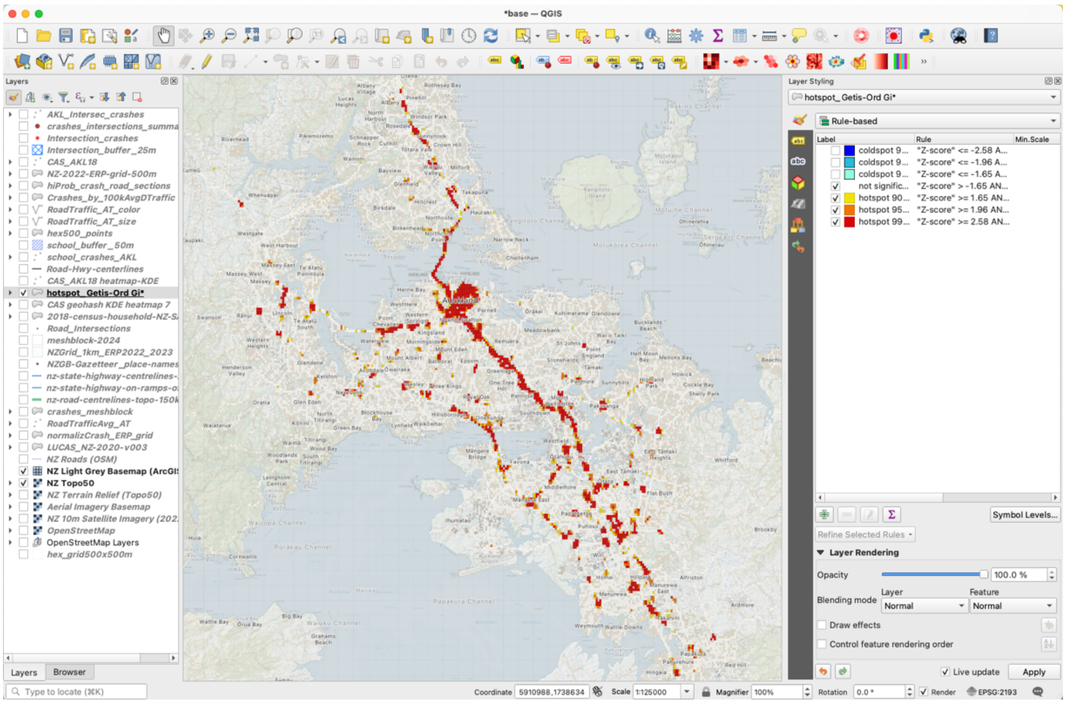


Note. Darker red meaning higher accident density areas.

Getis-Ord Gi* Hotspot Analysis

For more statistically robust identification of hotspots, and complement KDE, I used the Getis-Ord Gi* tool (Hotspot Plugin). This method identifies statistically significant clusters of high or low values, for understanding accident patterns beyond visual density (Erdogan et al., 2008). Getis-Ord Gi* was preferred over Local Moran's I because its more suitable identifying spatial clusters in road networks.

Figure 21 – Accidents Hotspots for Auckland Region (Partial)

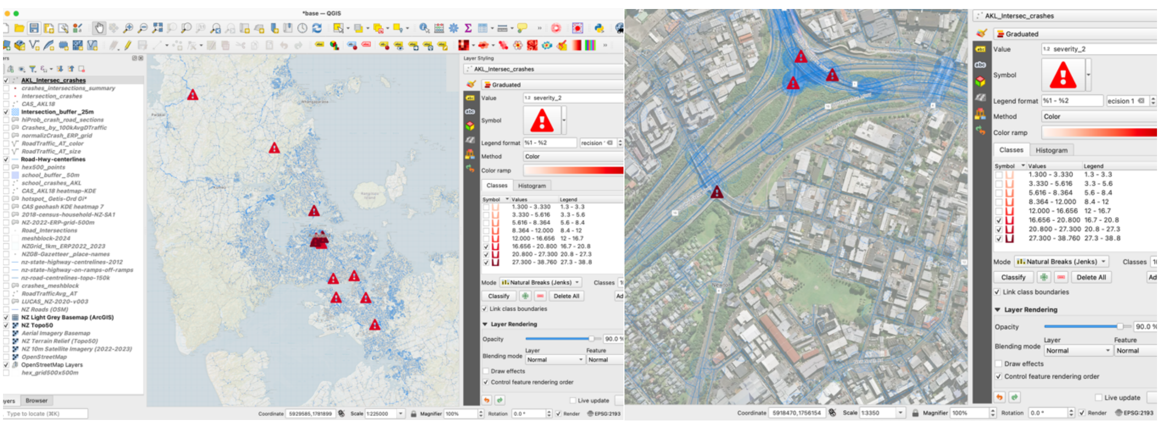


Note. Highways showing high statistical significance as hotspots for road accidents, most Auckland CBD being a major hotspot and some scattered hotspots around the South of Auckland.

Intersection Analysis

Intersections are critical points in any road network, after visually noticing accidents often clustered around intersections, particularly around highway on/off-ramps, I analysed these. First, merging road centerlines, dissolving them to create a unified network, and extracting intersection points. Buffers of 25m were created around intersections, and accident data was joined to identify high-risk intersections. Having all roads on a flat plain and multiple centerlines created redundancies as seen below, nonetheless being a high risk intersection.

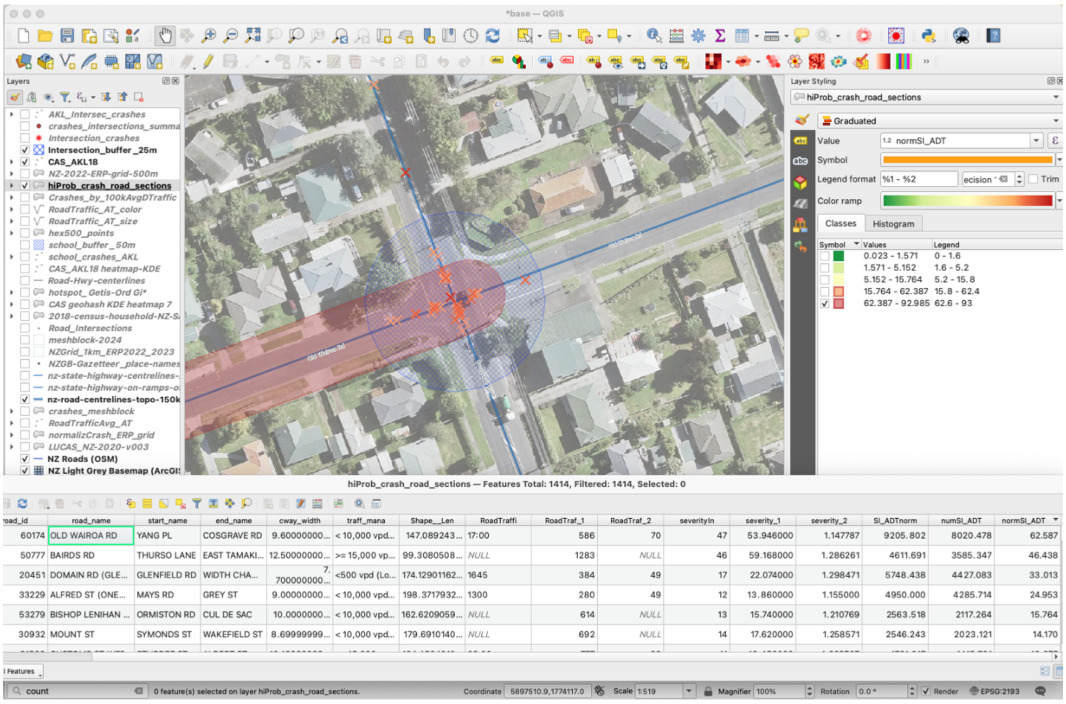
Figure 22 – High Risk Intersections around Auckland Region (Partial) on the left and CBD (particularly around Newton Rd and Grafton) on the right



Normalisation by Traffic Volume

To determine the road sections with the highest probability of accidents, I normalised the accident severity index by average daily traffic and road section length. This step involved creating buffers around road sections and spatially joining traffic volume data. Only sections with more than five accidents were considered to avoid outliers. This method show road sections with a higher probability of accidents relative to their traffic volume/length, with Old Wairoa Rd intersection emerging as a particularly high-risk area especially during peak hours (around 17:00).

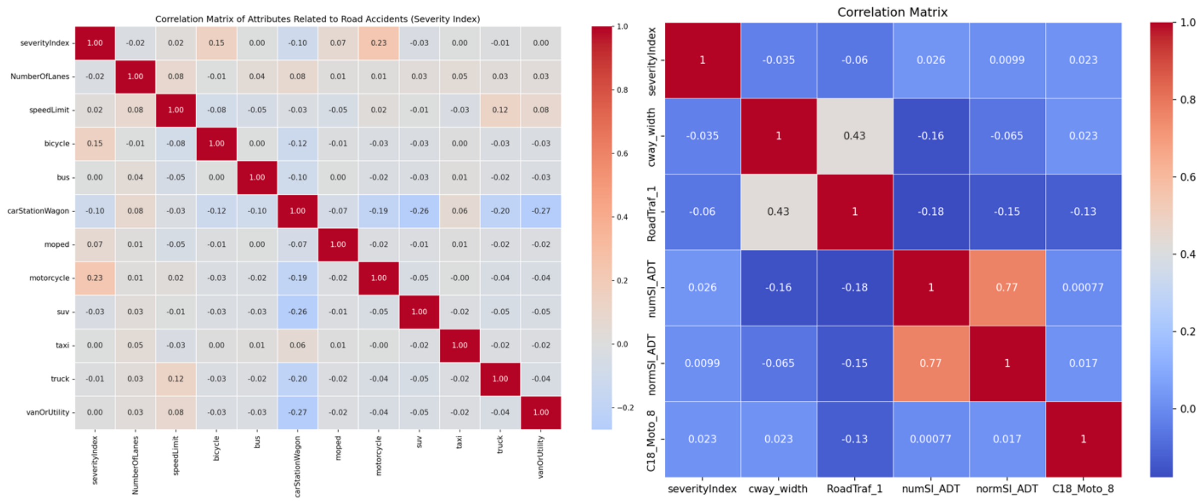
Figure 21 – High Probability of Accidents Road Section (Severity Index normalised by Avg. Daily Traffic and Road Section Length)



Correlation Analysis of Attributes with Accidents

To identify attributes most correlated with traffic accidents, I performed a correlation analysis in Python. Using correlation matrices explored relationships between road attributes and vehicle types. The correlation matrix revealed that motorcycle had the highest positive correlation with accident severity (23.2%), followed by bicycles (14.5%). Conversely, car station wagons showed a negative correlation (-10.1%), indicating they are less likely to be involved in severe accidents. Normalising the severity index by traffic volume and section length highlighted the impact of traffic flow on accident severity, although showing weak correlations.

Figure 22 – Correlation Matrices for Attributes Related to Road Accidents (Severity Index)



Note. On the left correlation among attributes of the CAS dataset. On the right, correlation of Carriage Width, Avg. Daily Traffic (ADT), Normalised Severity Index by ADT and by ADT and Road Section Length (AT), and finally by Number of Vehicles per Household (Census 2018).

The correlation analysis faced challenges such as missing data in key attributes like speed limit and number of lanes. Despite these challenges, the analytical approach is supported by the methods discussed by Anderson (2009) and Zahran et al. (2019), which demonstrate the effectiveness of spatial and statistical analysis in understanding traffic accident patterns.

Geovisualisation

Visualising the results was crucial to effectively communicate the findings. Each visualisation method was selected for its ability to highlight specific aspects of the data, ensuring a comprehensive understanding of traffic accident patterns.

Road Traffic Density Maps

To visualise road traffic density, I compared two methods of the same road traffic density around Auckland region (partial views), in Figure 23 using a colour-coded gradient Green-Yellow-Red as traffic intensity increases. In Figure 24 the exact same data but using a size of road sections and colour red to give a more immediate understanding of the road network traffic density. Both methods were selected for their ability to enhance readability and visual impact.

Figure 23 – Traffic Density by Road Section (colour gradient Green-Yellow-Red)

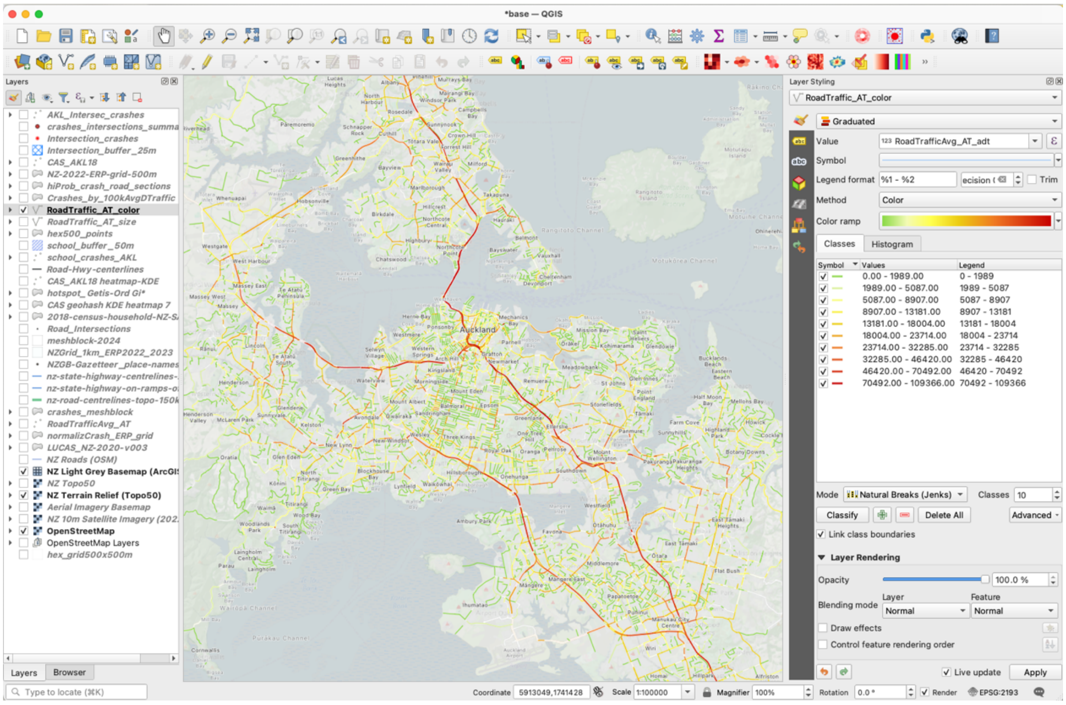
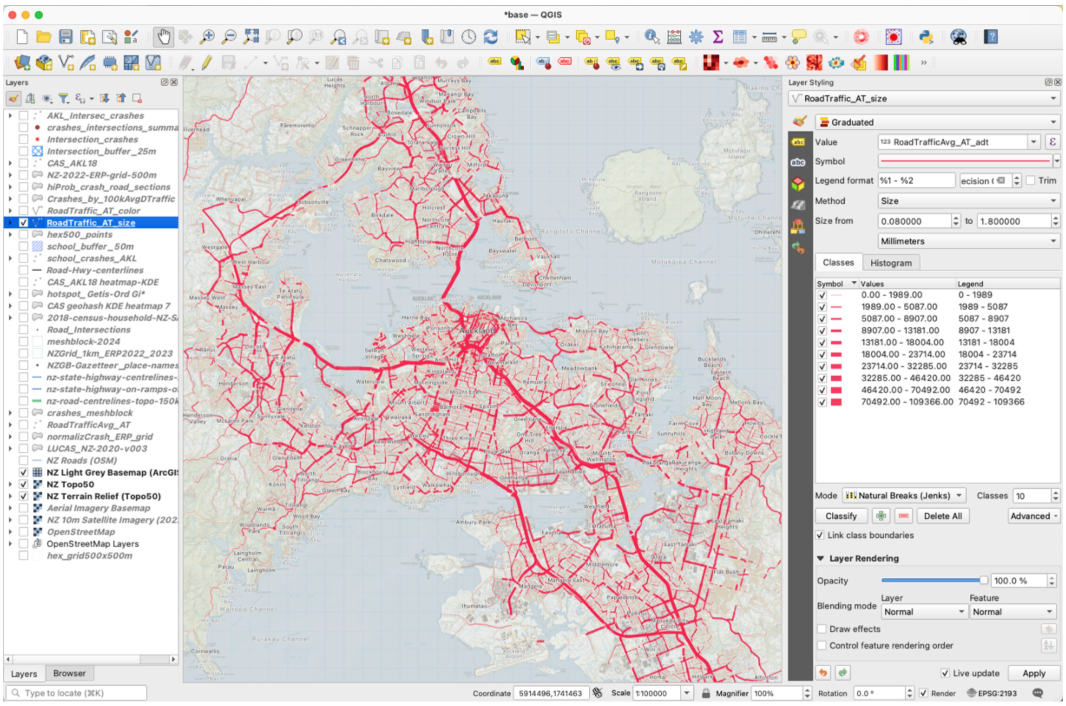


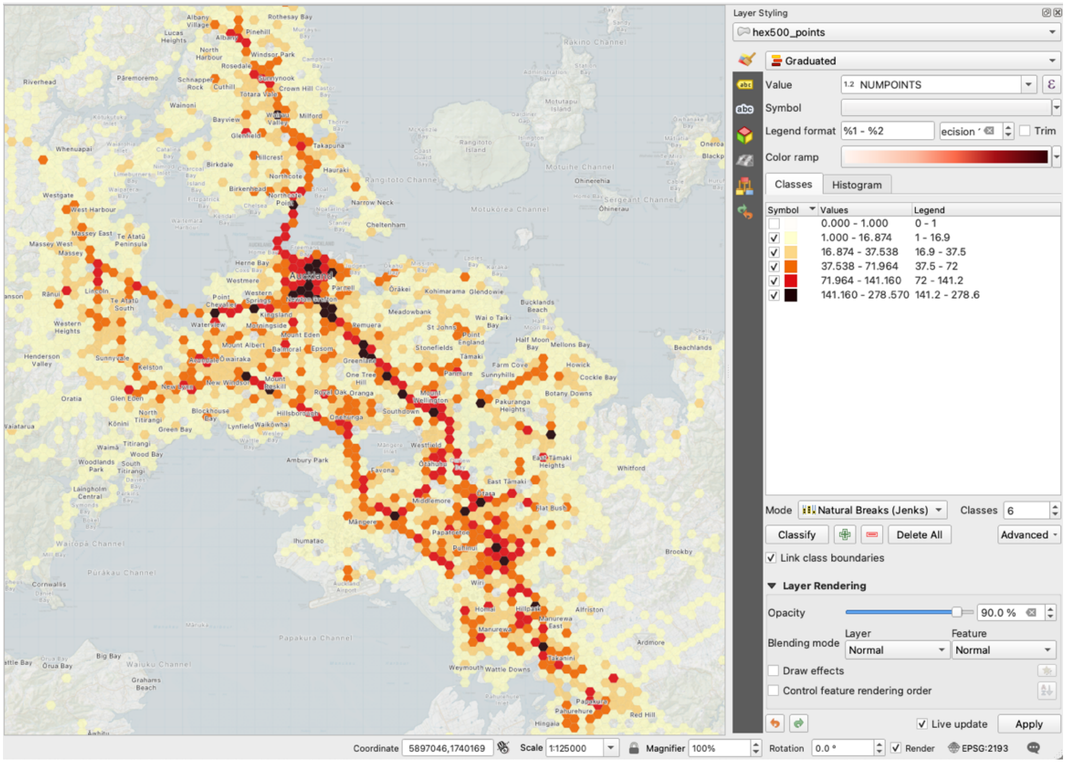
Figure 24 – Traffic Density by Road Section (using size of road sections to display traffic volumes)



Hexagon Grid – Accident Frequencies Map

To display accident frequencies, I used a hexagon grid map (Figure 25). This method provides a uniform spatial resolution, which helps for comparing accident zones across different areas. Hexagon grids were chosen over point density maps because they offer a more structured representation of spatial data, so it is easier to identify and compare accident hotspots effectively (Satria & Santoso, 2018). This method ensures that each hexagon represents an equal area, facilitating fair comparisons between different locations.

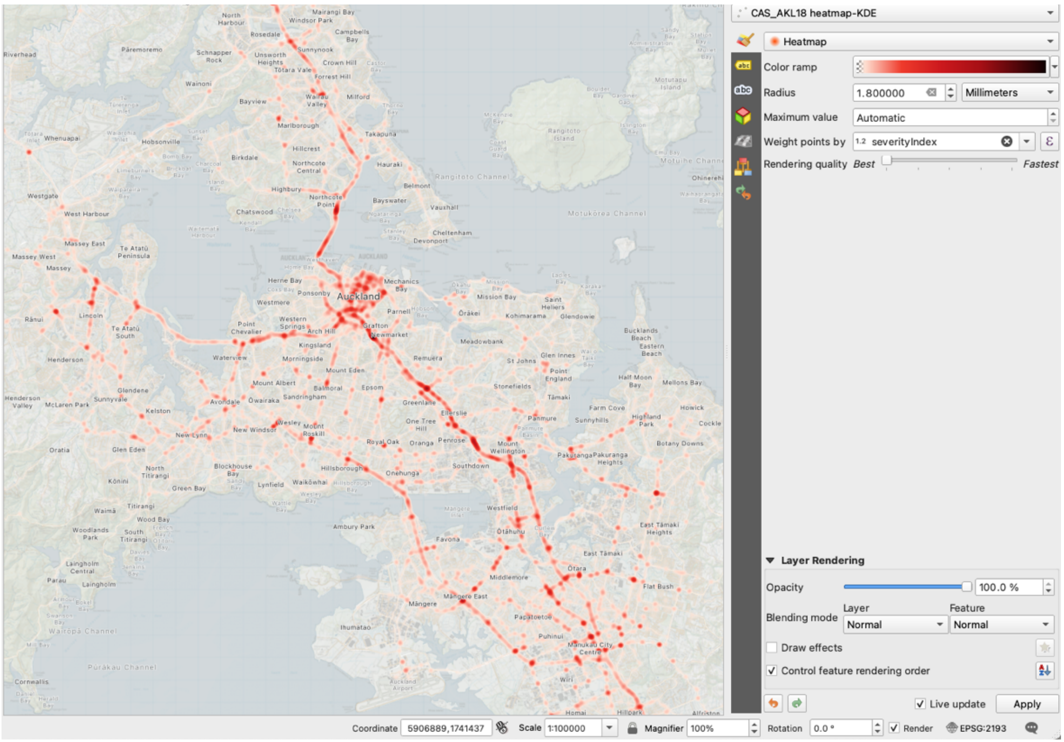
Figure 25 – 500m Hexagon Map of Crashes for Auckland City (partial view, graduated by frequency and location)



KDE Heatmap

The KDE heatmap was chosen for its ability to provide a smooth and continuous representation of accident densities. This visualisation helps identify general patterns and trends in accident occurrences, highlighting high-density areas (Anderson, 2009; Liu & Zhu, 2004). The heatmap uses a red gradient colour scheme, where darker shades mean higher accident densities. This visualisation method was chosen because it clearly displays areas with high accident concentrations, providing a visual guide to critical zones that require attention.

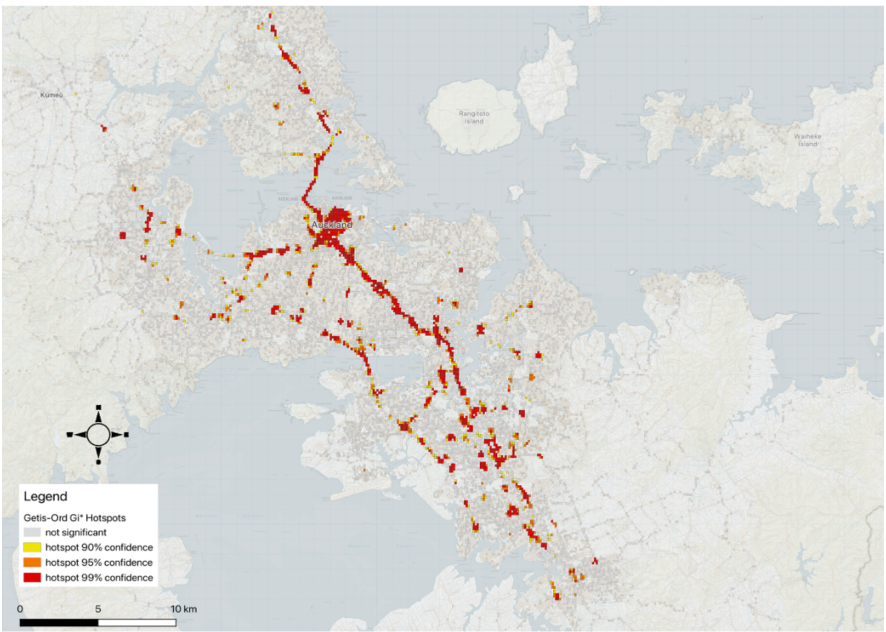
Figure 26 – KDE-Heatmap of Accident Density in Auckland City (Partial)



Getis-Ord Gi* Hotspot

The Getis-Ord Gi* hotspot analysis (Figure 27) was used to identify statistically significant clusters of high and low accident densities. This method is essential for understanding accident patterns beyond simple visuals, providing a statistical relevance for identifying critical areas. The rule-based colour scheme highlights areas with significant clustering, making this method highly valuable for spatial analysis of road safety (Erdogan et al., 2008). Local Moran's I is not as effective as Getis-Ord Gi* identifying spatial clusters such as in road networks.

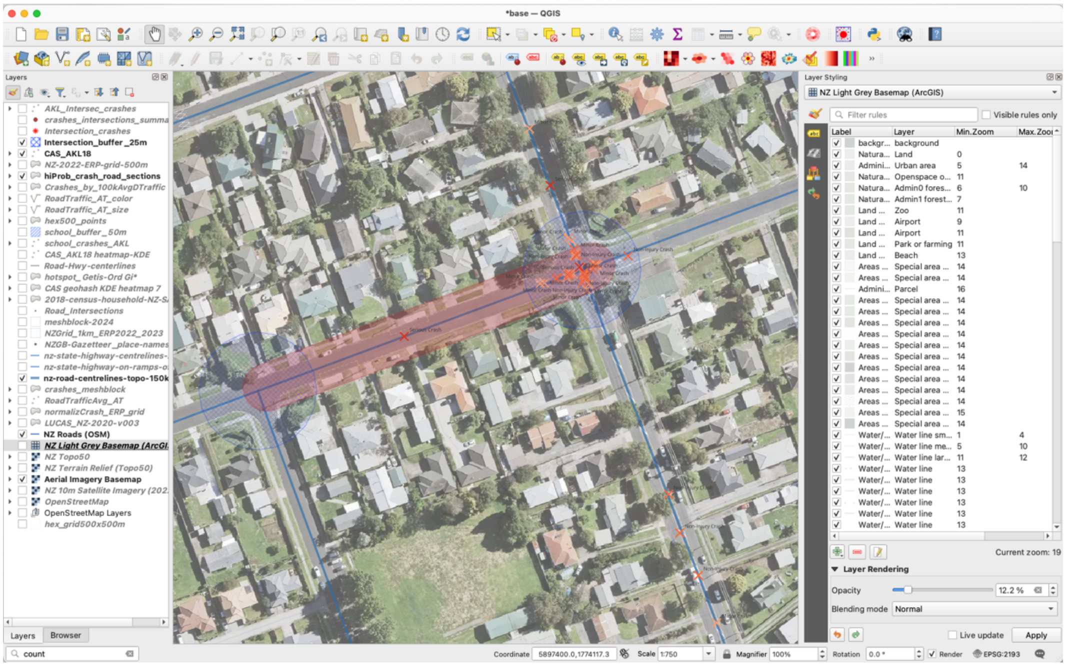
Figure 27 – Getis-Ord Gi* Accident Hostspot Analysis of Auckland region (Partial)



Black Spot – Traffic Volume Normalisation

To visualise the probability of accidents relative to traffic volume, I created map showing normalised accident severity by traffic volume. The map used a combination of colour and size graduations to indicate the likelihood of accidents on different road sections. This visualisation highlighted road sections with a higher probability of accidents relative to their traffic volume, providing actionable insights for traffic management.

Figure 28 – Black Spot – Road Section with High Probability of Accidents



Visualisation Design Considerations

When designing the visualisations, several approaches were considered to ensure clarity and effectiveness, as discussed in our lectures. Colours were selected to clearly distinguish between different data values. For instance, the red gradient in the KDE heatmap effectively highlights areas with high accident density. Varying symbol sizes and colour intensities were used to create a visual hierarchy, making it easier to identify key areas quickly. Tried minimising symbol occlusion, ensuring most important details remained visible. The hexagon grid map, for example, offers a structured layout that reduces overlap and occlusion. The visualisations were made to be informative rather than persuasive, presenting data without manipulation, to ensure they represent the underlying data.

Balanced levels of generalisation were considered to keep the visualisations clear and informative without oversimplifying the data. The hexagon grid map, for instance, balances detail and clarity and provids an overview while maintaining essential spatial relationships.

Alternative Methods and Challenges

While the selected visualisation methods were effective, alternatives such as Local Moran's I for cluster analysis and traditional choropleth maps for visualising accident density were considered. Local Moran's I is useful for identifying local clusters, but Getis-Ord Gi* was chosen for its ability to handle linear networks. Traditional choropleth maps, although useful, were not as effective in representing the continuous nature of accident densities as KDE and hexagon grids.

Challenges included the lack of temporal data in the CAS dataset limited the ability to perform time-based analyses. Additionally, missing data and data quality issues, multiple software crashes and the experimental nature of some plugins, which required modifications to the code.

By integrating these analyses and visualisations, the project aims to provide a comprehensive understanding of traffic accident patterns in Auckland, illustrating that highways large traffic volumes and intersections play an important role in road safety. The maps created were not only focus on being visually compelling but also backed by strong spatial analysis, ensuring their reliability and usefulness for road safety planning and interventions.

Outcome and Issues

Outcome

The analysis identified areas within the Auckland region most affected by traffic accidents, revealing several high-risk intersections and road segments.

Intersections around Newton and Grafton in Auckland CBD were some of the identified as particularly hazardous, with high accident frequencies and severities. These findings align with the Getis-Ord Gi* hotspot analysis, indicating statistically significant clusters of accidents.

Road Segments with High Accident Probability: By normalising accident severity indices by traffic volume and road section length, the study pinpointed specific road sections, such as the Old Wairoa Rd intersection, as having a notably high probability of accidents, particularly during peak traffic hours.

Correlation Analysis: Motorcycles and bicycles are more likely to be involved in severe accidents compared to cars and station wagons, suggesting a need for targeted safety measures for these vehicle types.

Traffic Volume and Accident Severity: Areas with high traffic volumes, such as major highways and arterial roads around the CBD, corresponded with higher accident severities, highlighting the importance of traffic management strategies to mitigate accident risks.

Spatial Patterns and Accident Density: The KDE heatmap and hexagon grid analysis provided a clear visual representation of accident densities, reinforcing the identification of accident hotspots and enabling a focused approach to road safety interventions.

Issues

The project faced several challenges that impacted the analysis. The CAS dataset lacked detailed temporal information and had missing values in important attributes such as driver demographics and socioeconomic data. These gaps limited the ability to perform proper temporal analyses and understand socioeconomic factors contributing to accidents.

Inconsistencies several datasets such as in road centerline data and missing minor roads introduced challenges in accurately mapping the road network. Differences in data collection methods and sources also affected the spatial precision of the analysis.

The analysis encountered multiple software crashes and issues with experimental plugins, requiring modifications to the code and workarounds to complete the visualizations and analyses.

Recommendations

Acquiring more detailed temporal data and demographic information would enhance the depth of the analysis. Collaborating with data providers to fill gaps in the dataset could provide a more comprehensive understanding of accident patterns.

Standardising data collection methods and ensuring consistency across different data sources would improve the accuracy of spatial analyses. Enhancing the precision of road network data, particularly for minor roads and intersections, would provide a more accurate representation of accident locations.

Incorporating machine learning models to predict accident occurrences and identify contributing factors could provide deeper insights. Utilising advanced spatial statistical methods, such as Bayesian hierarchical models, could improve the robustness of hotspot detection.

Developing interactive geovisualisation tools, such as web-based GIS platforms, would allow for more dynamic exploration of accident data and enable stakeholders to interactively identify and analyse high-risk areas.

Reflection

Project Success and Insights

The project demonstrated the powerful capabilities of GIS in analyzing and visualizing traffic accident data, providing valuable insights into road safety in the Auckland region. The integration of spatial and statistical analyses enabled the identification of critical accident hotspots and high-risk road segments, which can inform targeted interventions to improve road safety. The visualizations created, including hexagon grids, KDE heatmaps, and Getis-Ord Gi* hotspot maps, I believe to have effectively communicated the findings and highlighted areas requiring attention.

Challenges and Solutions

One of the primary challenges was dealing with data quality and completeness issues. The CAS dataset's lack of detailed temporal data and missing values in key attributes required extensive data wrangling and preprocessing. There seems to have also several variables with wrong inputs. Future projects could benefit from more comprehensive and detailed datasets, enhancing the granularity and accuracy of the analysis.

Technical issues with software and plugins also posed frustrating delays, using more stable software with improved capabilities might open the door for improved visualisations.

Future projects could explore ways to incorporate anonymized demographic data to gain a better understanding of the socioeconomic factors contributing to traffic accidents.

Future Directions

Building on the findings of this project, future research could explore the integration of machine learning models to predict accident occurrences and identify contributing factors. Advanced spatial statistical methods, such as Bayesian hierarchical models, could improve the robustness of hotspot detection and provide deeper insights into the spatial distribution of accidents.

Developing interactive geovisualization tools, such as web-based GIS app platforms, would enable more dynamic exploration of accident data and facilitate stakeholder engagement. These tools could allow users to interactively identify and analyse high-risk areas, improving the practical application of the findings.

Overall, this project has demonstrated the significant potential of GIS in improving road safety through the analysis and visualization of traffic accident data. Addressing the challenges encountered such as lack of temporal data, and building on the insights gained, future research can continue to contribute to the development of effective road safety strategies and interventions.

References

Abdel-Aty, M., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, *32*(5), 633-642.

Afolayan, A., Abiola, O. S., Easa, S. M., Alayaki, F. M., & Folorunso, O. (2022). GIS-based spatial analysis of accident hotspots: A Nigerian case study. *Infrastructures 7*, 103.

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention, 41*(3), 359-364.

Bíl, M., Andrášik, R., & Janoška, Z. (2013). Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident; Analysis & Prevention, 55*, 265–273.

Budiawan, W., & Purwanggono, B. (2018). Clustering Analysis of Traffic Accident in Semarang City. *E3S Web of Conferences, 73*, 12001.

Chen, H. (2012). Black Spot Determination of Traffic Accident Locations and Its Spatial Association Characteristic Analysis Based on GIS. *Journal of Geographic Information System, 04*(06), 608–617.

Dereli, M. A., & Erdogan, S. (2017). A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. *Transport*, 32(3), 319-331.

Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. *Accident Analysis & Prevention*, *40*(1), 174-181.

Flahaut, B., Mouchart, M., San Martin, E., & Thomas, I. (2003). The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach. *Accident Analysis & Prevention*, 35(6), 991-1004.

Gundogdu, I. B. (2010). Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya. *Safety Science*, 48(6), 763-769.

Haynes, R., Lake, I. R., Kingham, S., Sabel, C. E., Pearce, J., & Barnett, R. (2008). The influence of road curvature on fatal crashes in New Zealand. *Accident Analysis & Prevention*, 40(3), 843-850.

Kaygisiz, Ö., Düzgün, Ş., Yildiz, A., & Senbil, M. (2015). Spatio-temporal accident analysis for accident prevention in relation to behavioral factors in driving: The case of South Anatolian Motorway. *Transportation Research Part F: Traffic Psychology and Behaviour, 33*, 128–140.

Kraft, S., Mrkvička, T., Petříček, J., & Blažek, V. (2023). Modelling the risk levels of road network sections for motorcycle transport in the Czech Republic. *Traffic Injury Prevention*, 24(1), 9-13.

Liu, S., & Zhu, X. (2004). Accessibility Analyst: An Integrated GIS Tool for Accessibility Analysis in Urban Transportation Planning. *Environment and Planning B: Planning and Design, 31*(1), 105–124.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291-305.

Nguyen, H. T., & Ho, L. T. (2023). Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam. *Heliyon*, 9(3), e03317.

Ng, K., Hung, W., & Wong, W. (2002). An algorithm for assessing the risk of traffic accident. *Journal of Safety Research, 33*(3), 387–410.

Pourroostaei Ardakani, S., Liang, X., Mengistu, K. T., So, R. S., Wei, X., He, B., & Cheshmehzangi, A. (2023). Road car accident prediction using a machine-learning-enabled data analysis. *Sustainability*, *15*(7), 5939.

Prasannakumar, V., Vijitha, H., Charutha, R., & Geetha, N. (2011). Spatio-temporal clustering of road accidents: GIS-based analysis and assessment. *Procedia - Social and Behavioral Sciences*, 21, 317-325.

Rodríguez-Morales, B., Díaz-Varela, E. R., & Marey-Pérez, M. F. (2013). Spatiotemporal analysis of vehicle collisions involving wild boar and roe deer in NW Spain. *Accident Analysis & Prevention, 60*, 121–133.

Sagar, S., & Stamatiadis, N. (2022). GSIS techniques to analyze factors associated with crash occurrence. *Traffic Safety Research, 2*, 000005.

Satria, R. A., & Santoso, A. (2018). Traffic accident analysis and prevention using GIS: A case study in Jakarta. Journal of Transportation and Safety, 5(2), 123-135.

Steenberghen, T., Dufays, T., Thomas, I., & Flahaut, B. (2004). Intra-urban location and clustering of road accidents using GIS: A Belgian example. *Networks and Spatial Economics*, 4(2), 135-142.

Stephen, L., Kelakom, G. G., Sojan, J. M., & Sreelakshmi, K. S. (2018). Identification and analysis of accident blackspots using GIS. International Research Journal of Engineering and Technology (IRJET), 5(3), 3455-3459.

Te Manatū Waka. (2022). Te Marutau - Ngā tatauranga ā-tau Safety - Annual statistics. Retrieved from https://www.transport.govt.nz/statistics-and-insights/safety- annual-statistics/sheet/social-cost-of-road-crashes/

Van Raemdonck, K., & Macharis, C. (2014). The road accident analyzer: a tool to identify high-risk road locations. *Journal of Transportation Safety & Security, 6*(2), 130-151.

NSW Road Safety, Traffic Management Directorate. (1999). Road traffic accidents *in New South Wales–1997: Statistical statement: Year ended 31 December 1997*. Roads and Traffic Authority of NSW.

Waka Kotahi NZ Transport Agency. (2020). Road to Zero: New Zealand's road safety strategy 2020-2030. Retrieved from https://www.nzta.govt.nz/assets/resources/road-to-zero/road-to-zero-strategy.pdf

Waka Kotahi NZ Transport Agency. (2017). Serious injuries final report. Retrieved from https://www.nzta.govt.nz/assets/resources/serious-injuries/serious-injuries-final-report.pdf

World Health Organization. (2023). Global status report on road safety 2023. World Health Organization. Retrieved from https://www.who.int/publications/i/item/9789240086517/

Zahran, E.-S. M. M., Tan, S. J., Putra, N. A., Tan, E. H. A., Yap, Y. H., & Rahman, E. K. A. (2019). Evaluation of various GIS-based methods for the analysis of road traffic accident hotspots. *MATEC Web of Conferences, 258*, 03008.