RMIT University

Applied Research Project
Data Analysis for Transportation Equity in Australia

November 2024

RMIT Supervisor Name: Laleh Tafakori

Integrated Travel Supervisor: Vern Raincock

Students:

Victorino Marquez

ID:3852887

Priya Ningappa

ID: 3959738

Arka Dasgupta

ID:3980896

Disclaimer

1. I/we have read and understood the Declaration and Statement of Authorship.
2. If I/we do not agree to the Declaration and Statement of Authorship in this context and a signature is not included below, the assessment outcome is not valid for assessment purposes and will not be included in my final result for this course.

Declaration and Statement of Authorship can be found from assessment declaration

By signing below, I/we confirm acceptance of the above statements:

Name	Signature	Date
Victorino Marquez	*Victorino Marquez*	10/11/2024
Priya Ningappa	*Priya Ningappa*	10/11/2024
Arka Dasgupta	*Arka Dasgupta*	10/11/2024

Acknowledgement

We would like to thank Mr. Vern Raincock and the company at Integrated Travel for supervision and guidance throughout this project. Their ideas and advice were invaluable in helping the project team make consistent progress and keep the scope manageable for this area of research.

We would also like to thank Associate Professor Laleh Tafakori and the Applied Research teaching team for providing feedback after each assignment, ensuring that we could improve the quality of the project going forward. We also thank Dr. Mali Abdollahian for providing advice for the project kick-off meeting and for being available as an academic supervisor.

Last, but not least, we would like to thank the Industry Engagement Coordinator | WIL & Engagement team in RMIT STEM College and Integrated Travel for working together to provide this valuable Work Integrated Learning project. It was a great opportunity for students like us to see how the theories and skills learned in the classroom could be applied to real-world problems. May such projects continue to produce interesting findings in the future.

Executive Summary

Based on the analysis conducted in the report on the transportation equity strategy for Australia, there are some insightful points when analyse the different transport systems and routes.

The research focused on analysing infrastructure investments in road and rail systems to understand their impact on public transport patronage, vehicle usage, and efficiency across various Australian states. The key findings indicate that while road infrastructure investments significantly correlate with vehicle usage in certain states like the Northern Territory, the correlation varies widely across regions, suggesting that other factors like demographics, urbanisation, and public transport availability influence the outcomes.

On the rail side, Victoria showcased a positive relationship between rail investment and increased public transport patronage, especially post-pandemic, highlighting the effectiveness of targeted investments in boosting public transport utilization.

Furthermore, the analysis of specific routes, such as the Sydney-Melbourne corridor, underscored its high interaction level, likely driven by economic and population growth. The study also revealed that investing in key transport corridors could significantly improve connectivity and economic productivity, especially in highly populated regions. These findings provide valuable insights to optimize infrastructure investments and align them with demographic and economic needs for sustainable transport development.

Contents

Table of Figures

Problem Statement

In Australia, significant investments have been made in both rail and road infrastructure with the goal of improving public transport systems and road networks, ultimately enhancing economic activity, mobility, and safety. However, the exact relationship between these expenditures and their impact on public transport patronage, freight movements, vehicle usage, and vehicle registrations remains ambiguous. The COVID-19 pandemic has further complicated these dynamics, introducing external shocks that have affected travel behaviour 1 and infrastructure usage across the country.

This study aims to address these gaps by analysing the relationship between infrastructure investments and usage patterns across both rail and road systems in Australian states, with a focus on Victoria. Specifically, the research will explore the effects of rail infrastructure spending on public transport patronage (trains, trams, and buses) and freight movement, while also investigating the link between road-related expenditures and vehicle usage, such as vehicle kilometers traveled (VKT) and vehicle registrations. Additionally, the study seeks to develop predictive models to forecast future transport demand and vehicle trends, providing policymakers with evidence-based insights to optimize infrastructure investments and support sustainable, resilient mobility systems.

Objectives:

1. Rail Infrastructure Expenditures and Public Transport:

- Analyse the relationship between infrastructure engineering construction work (total value) and public transport patronage in Australian capital cities (Sydney, Melbourne, Brisbane, Adelaide, Perth), evaluating the effect of rail-related expenditure on heavy rail patronage across these cities.
- Investigate trends in rail infrastructure expenditure and assess the relationship between rail-related investments and public transport patronage, with a focus on developing predictive models to forecast future public transport demand based on historical infrastructure spending.
- Estimate passenger flow between major cities (e.g., Sydney, Melbourne, Brisbane) using the Gravity Model, incorporating population, intercapital rail distances, and generalized travel costs. Additionally, examine the correlation between intercapital rail distances and interstate non-bulk rail freight volumes to understand how distance influences freight movement.
- Assess the effectiveness of rail investments in supporting sustainable urban mobility, considering both passenger and freight transport, and the impact of external factors such as economic growth and disruptions like the COVID-19 pandemic.

2. Road Infrastructure Expenditures and Vehicle Usage:

- Analyse the relationship between road-related expenditures and vehicle kilometres travelled (VKT) across Australian states to determine whether higher investments in road infrastructure result in increased road usage.
- Investigate the correlation between road-related expenditures and vehicle registrations to understand whether states with higher road investments experience a rise in the number of registered vehicles.
- Examine state-wise vehicle registrations in relation to road-related expenditures, comparing state-specific trends to assess if road investments coincide with an increase in vehicle registrations across various vehicle categories.

- Forecast vehicle-related data for the next five years by developing a predictive model using historical data to project trends in kilometres travelled and vehicle registrations, aiding in future infrastructure planning.

This comprehensive approach integrates the analysis of both rail and road infrastructure expenditures, providing valuable insights into how investments affect public transport patronage, vehicle usage, and future mobility trends across Australian states and cities.

Key Research Questions:

1. Does an increase in Road Expenditure across Australian States result in a proportional rise in Vehicle Kilometres travelled (VKT)
   - Do states with higher road infrastructure investments see a corresponding increase in vehicle registrations?
   - Can vehicle registrations and usage trends be accurately forecasted based on historical road infrastructure expenditure data?
2. Is investment in rail infrastructure effectively increasing rail patronage.
   - How does public transport patronage in Victoria (trains, trams, buses) relate to rail infrastructure spending?
3. Which Rail Corridor in Australia should be prioritized for Investment?

Literature Review

Infrastructure investment plays a critical role in enhancing the efficiency and capacity of transport systems in urbanized regions. Studies have indicated that well-planned and adequately funded infrastructure contributes to economic growth, reduced congestion, and improved public transport systems (Button & Lall, 1999; Banister & Berechman, 2000). Public transport networks, especially in capital cities, have shown varied responses to infrastructure investments, with outcomes largely dependent on the types and scale of projects (Currie & Phung, 2007). For example, significant investments in rail systems across global cities like London, New York, and Tokyo have led to increased ridership, indicating a positive correlation between investment and public transport usage (Glaeser, 2011).

In Australia, capital cities such as Sydney and Melbourne have experienced population growth and urban expansion, necessitating substantial infrastructure investments to maintain transport efficiency (BITRE, 2018). Recent investments in public transport infrastructure, particularly in rail networks, have aimed at increasing the system's capacity and efficiency. However, the impact of these investments on actual patronage levels remains unclear. Studies by BITRE (2019) and the ABS (2023) highlight trends in infrastructure spending but offer limited insight into the direct effects on public transport usage and freight movement.

Previous research has explored the application of the Gravity Model to forecast passenger flows between major cities, emphasizing the influence of factors like population size, travel cost, and intercity distances (Bovy & Stern, 1990). This model has proven effective in understanding travel demand patterns across urban regions. Similarly, rail freight movement has been analyzed in terms of rail distances and logistics infrastructure, with studies showing that greater distances typically correlate with lower freight volumes (Woodburn, 2003).

This study contributes to the existing literature by examining the relationship between infrastructure investments and public transport patronage in Australian capital cities, using both traditional statistical analysis and predictive modeling approaches.

A lot of work has been done previously on the relationship between infrastructure investment and public rail patronage. Scholars across the world have tested in extreme detail how funding, user perceptions and other such factors can affect the public transport system. Louviere, Hensher and Swait (2000) present a key framework of commuter choice by means of stated choice methods. It can pave way for one to be able to model how the public might be making transport mode choices and decisions based on factors such as change in service quality, infrastructure investment, to name a few. Their findings have been quite helpful to get a glimpse of how rail infrastructure investments can affect the public perception and usage of public transport.

In a study by Andersson (2017), further emphasis has been shed on how the funding for public transportation should be allocated based on the habit of infrastructure investments and the benefits received by the users to subsequently encourage the public to use public transportation. The market approach laid out by the study stresses on how any mode of public transport requires appropriate funding policies tailored to meet the user expectations. The study also aligns quite well with the Victorian government's railway infrastructure plan, which effectively aims to maximize the patronage.

Chen and Hall (2012) assess the economic effects of rail investments, with a focus on high speed trains in the UK. Since there has not yet been a significant development of Victoria's rail infrastructure based on high speed systems, this study is extremely relevant. Improved rail infrastructure paves way for increased levels of accessibility and can thereby bring about more patronage.

In the current context, it is also quite pivotal to comprehend the benefits and costs associated with transport infrastructural investments. Litman (2016) sheds light on this through his Best Practices Guidebook which goes through ways to evaluate both direct and indirect benefits of public transport investments. Litman emphasizes that apart from reduced traffic, better accessibility there are other external benefits too. The framework is quite useful for our research in determining whether the railway expenses in Victoria are yielding expected benefits by means of increased patronage

Investments in road infrastructure are seen as vital for improving transportation efficiency and economic productivity. Previous studies have highlighted the importance of road-related expenditure in facilitating mobility and supporting economic growth (Button & Lall, 1999; Banister & Berechman, 2000). However, the direct impact of road investments on vehicle usage and registrations remains underexplored.

In Australia, road infrastructure investments have been significant, but their effect on road usage, especially vehicle kilometres travelled (VKT) and registrations, is unclear. Historical data on road expenditures and vehicle movement, along with the number of registered vehicles, provide a basis for understanding these relationships (BITRE, 2018). Studies from the Australian Bureau of Statistics (ABS) show infrastructure spending trends but offer limited insights into the impact on vehicle activity.

This study aims to fill this gap by analysing the correlation between road expenditures and vehicle usage, as well as forecasting future vehicle movement trends using predictive modelling approaches.

Several studies have investigated green infrastructure within public and road transport systems, mainly with regards to carbon emission reductions and their contributions to more sustainable urban growth. Ewing and Cervero (2010) in their study observe that compact, transit-oriented developments can reduce private vehicle use and increase public transport use in a manner consistent with the climate change goals set for most parts of the world. This focus is starting to take center stage in Australia, especially as Melbourne and Sydney's urban planning puts

greater emphasis on sustainable transport options. Infrastructural investment that includes environmental concerns helps resolve economic and congestion issues and is effective for long-term environmental sustainability. This could contribute to the drawing of infrastructure policies and changing transport behaviors in the future, while integrating electric buses and energy-efficient rail systems.

Beyond this, the embedding of emerging technologies within transport infrastructure is an area of constant evolution that continues to push the boundaries of both efficiency and user experience. Smart transportation systems-linked real-time analytics of data, intelligent traffic management, and autonomous vehicles-are increasingly being pursued as solutions to urban mobility. Smart infrastructure investments can vastly reduce congestion, optimize traffic flow, and improve public transportation through sensor networks and communication systems. In Australia, initiatives in smart ticketing and integrated transport apps has made a huge leap in reshaping public transportation for more precise and user-friendly travel information to commuters. As urban centres continue to expand, these technologies-enclosed by more traditional investment might just prove to be the key for creating far more adaptive, responsive, and sustainable transport systems.

The policies within the government of Melbourne include supporting this transition, and cities across the world are now focusing on greener and more sustainable transport networks. This will result in increased overall efficiency and sustainability within the transport system. Large projects, such as Australia's Suburban Rail Loop in Melbourne, strive to address the barriers in transport by providing improved connections in the outer suburbs, which are essentially the areas far away from the from the availability of public transport. Focusing on infrastructure investment ensures not only economic efficiency but also equal access to all the people in need of it.

Data Description

Economic – Infrastructure

The BITRE infrastructure dataset consists of several key tables that provide detailed insights into Australia's transport infrastructure, including investments, economic indicators, and employment statistics across multiple sectors. Here are some of the key tables we considered for analysis. For each table used, here is a detailed breakdown:

Producer Price Index for Transport Sectors: Contains the PPI values for rail, road, and water freight indexed to a base of 2011–12.

- Variables:
  - Year: Time series data from 1997 to 2023.
  - PPI for Rail Freight, Road Freight, and Water Freight: Indexed values representing the producer prices for each sector.

Employment in Major Transport Sectors: Employment data across various transport sectors from 2019 to 2023.

- Variables:
  - Year: Time series data from 2019 to 2023.
  - Employment (Thousands): Road, rail, water, and air transport employment figures.

Average Weekly Earnings (Adjusted by CPI): Average weekly earnings for different transport sectors adjusted to 2020–21 prices.

- Variables:
  - Year: Time series data from 1996 to 2021.
  - Wages for Road, Rail, Water, Air, and Other Transport: Adjusted earnings figures for each sector.

Road Data Description

The data for this project was sourced from multiple public datasets provided by the Bureau of Infrastructure and Transport Research Economics (BITRE). These datasets provide detailed statistics on road usage, infrastructure construction, and government expenditure related to transportation in Australia.

Road Dataset:

This dataset contains comprehensive information on road usage across Australia, including vehicle kilometers traveled (VKT) by different vehicle types and across various regions. It also includes data on the number of registered vehicles over time, allowing for a detailed analysis of road traffic trends and vehicle type dominance

Road Related Expenditure Dataset:

This dataset includes detailed government spending on road infrastructure, both regionally and nationally, from the Australian government. The data captures road-related revenue and expenditure, covering investments in road maintenance, construction, and upgrades. The dataset is key to understanding the relationship between infrastructure investments and road usage.

Infrastructure Construction Dataset:

This dataset provides data on the construction work done by both the public and private sectors for transportation infrastructure. It covers various infrastructure projects such as roads, bridges, and railways, and breaks down the construction activities over time. The dataset also highlights the growing role of the private sector in public infrastructure projects.

Each of these datasets was processed using R for analysis. The data provides a solid foundation to analyze the correlation between road-related expenditures and kilometers traveled, forecast future road usage, and evaluate the impact of public-private partnerships on transportation infrastructure development.

Rail

The dataset for rail analysis contains historical data on public transport patronage across different transport modes in Australian capital cities, specifically focusing on rail infrastructure. The data is organized by financial year and includes the following key columns:

Financial Year: This column specifies the financial year for the recorded data, ranging from

1976-77 to more recent years.

Heavy Rail: Represents the number of heavy rail (train) patronage in millions for the specified financial year. This metric is indicative of the usage levels of the heavy rail network in the capital city.

Light Rail: Shows the light rail patronage figures in millions. In some years, this value is zero, indicating that light rail services were either not present or not operational during those periods

UPT (Urban Public Transport): Captures the overall urban public transport usage in millions, combining figures from trains, buses, ferries, and other public transport modes.

Bus: Provides the number of bus patronages in millions for each financial year, highlighting the usage of bus services in the respective capital city.

Ferry: Details ferry patronage numbers in millions, indicating how frequently ferries were used for public transport in the specified financial year.

Total: The total number of public transport patronages across all modes (Heavy Rail, Light

Rail, Bus, and Ferry) in millions, providing an aggregated view of transport usage.

Capital City: Specifies the capital city for which the data is recorded, such as Sydney. This allows for comparison between different cities and their public transport usage patterns.

State/Territory Columns (NSW, VIC, QLD, SA, WA, TAS, NT, ACT): These columns represent individual states and territories in Australia. The values in these columns indicate the number of public transport users (in millions) for each state/territory. The dataset allows for a comparison of patronage across different states, highlighting regional variations in public transport usage.

Public Corporations: This column shows the public sector's share of transport infrastructure investment, in millions of dollars. It provides insight into the role of public corporations in funding transport infrastructure over the years.

Total Government: The total government expenditure on public transport infrastructure, combining investments from both state and federal governments. It serves as an indicator of the overall public sector commitment to developing transport systems.

Total Public Sector: The combined expenditure from public corporations and government, showing the total investment in public transport infrastructure by the public sector for each financial year.

This dataset is valuable for analyzing trends in public transport usage over time, understanding the impact of rail infrastructure investments on patronage, and comparing usage levels across different transport modes in Australian cities.

Proposed Methodology

Economic Analysis

To conduct a comprehensive analysis, a variety of methodologies will be applied based on the specific nature of each research question. For the investment trend analysis, time series analysis will be employed to identify historical patterns and growth rates in the road and rail transport sectors. This involves plotting cumulative investment over time and calculating the year-over-year growth rates to detect periods of significant investment changes. Additionally, linear or polynomial regression models will be used to forecast future trends in both sectors, allowing us to project whether current investment trajectories are sustainable or likely to shift.

For the Producer Price Index (PPI) trends, descriptive statistics will be coupled with time series forecasting models like ARIMA or exponential smoothing to predict future price movements. By plotting the historical PPI values for road, rail, and water freight, we can analyze trends and assess price volatility across these sectors. Understanding PPI trends is crucial for anticipating cost pressures on businesses reliant on transport services. The projected price trends will help identify whether particular sectors, such as water freight,are likely to see continued price inflation, which may impact supply chains and transportation costs.

In terms of employment distribution, sector-wise trends will be analyzed using bar plots and time series methods. The goal is to assess how employment has evolved in road, rail, water, and air transport from 2019 to 2023. Stacked bar plots will visually represent the employment distribution across these sectors, providing a clear understanding of which areas dominate in terms of workforce size. Further analysis will be conducted to explore correlations between employment trends and investment levels, allowing us to assess how infrastructure spending impacts job creation in the transport sector.

For the wage growth analysis, we will use line plots and volatility analysis to explore wage trends across different transport sectors from 1996 to 2021. This analysis will help identify which sector has experienced the highest wage volatility, as well as the factors driving wage growth in each sector. For example, technological advancements, economic cycles, or regulatory changes may explain why some sectors have seen more dramatic shifts in earnings over the years. Understanding wage dynamics is key for workforce planning and policy decisions.

Road and Rail Methodology

The analysis was conducted using secondary data sourced from publicly available datasets provided by the Bureau of Infrastructure and Transport Research Economics (BITRE), which include information on road usage, vehicle registrations, road-related expenditures, and infrastructure construction in Australia. The data was processed and analyzed using R, a programming language suited for statistical analysis and data visualization.

The first step involved data cleaning and preprocessing. Multiple datasets were merged by year to align road-related expenditures with vehicle kilometers traveled (VKT) acrossdifferent Australian regions. The data was then filtered to select key variables, such as state- wise expenditures and VKT, for further analysis.

Once the data was prepared, correlation analysis is performed to assess the relationship between road-related expenditures and VKT across the regions. The correlation coefficients were calculated for each state to understand how investments in infrastructure influenced road usage. In addition, hypothesis testing is conducted to explore whether higher road investments are associated with increased vehicle registrations, focusing on statistically significant relationships.

A 5-year forecast of VKT is generated using time series analysis, leveraging the ARIMA model to predict future road usage trends based on historical data. Visualization of the data is carried out using the ggplot2 package in R, which facilitated the creation of scatter plots and trend lines to present the relationship between expenditures and VKT, as well as forecasting results. The use of 3D visualizations and advanced statistical models was incorporated as per the feedback received from the industry supervisor, enhancing the clarity and depth of the analysis.

Similarly to Road Analysis, Rail and Economic dataset will have visualization, correlation, comparison analysis and forecasting. Once all analysis are completed then we will write a conclusion whether the investment strategy used in different Australian states is highliy correlated with the population growth and needs. First, the data is cleaned and preprocessed to match investment figures for infrastructure in roads and railways with relevant economic indicators such as the PPI, employment, and usage of public transport. We then make use of R to create various visualizations such as trend lines, year-over-year growth rate comparisons, and cumulative population growth charts, to analyze the long- term investment pattern

Data Analysis

Economic Infrastructure Analysis

The dataset we are analysing in our case study offers a comprehensive insight into Australia's transport and infrastructure sectors, with a special focus on economic contributions, employment, earning trends to name a few. The data includes critical metrics such as gross value added by industry, employment statistics, average weekly earnings, and producer price indexes, sourced **from Table 1.1**, **Table 1.2a, Table 1.5, and Table 1.6.** By analysing this data, we aim to uncover key insights that could provide pivotal for future rail and road developments.

The goal of Economic Infrastructure study is to primarily identify and interpret trends lying under the transport and infrastructure sectors. Investment patterns, employment distributions across sectors, average weekly earnings, and freight cost dynamics are the topics of interest here. We intend to derive meaningful insights related to economic contributions, labour requirements, income levels, and operational costs across the various transport sectors across Australia.

Infrastructure Construction

Infrastructure Construction dataset provides information on Australian measures of infrastructure engineering construction. These are classified as: transport (roads, rail, ports, etc), energy (electricity and gas transmission networks, etc), telecommunications networks, and water supply and distribution networks. Data is sourced from the Australian Bureau of Statistics Engineering Construction Activity, Australia publication with some adjustments, as well as the Electric Vehicle Council.

Investment Trend for Road vs Rail Year-Wise

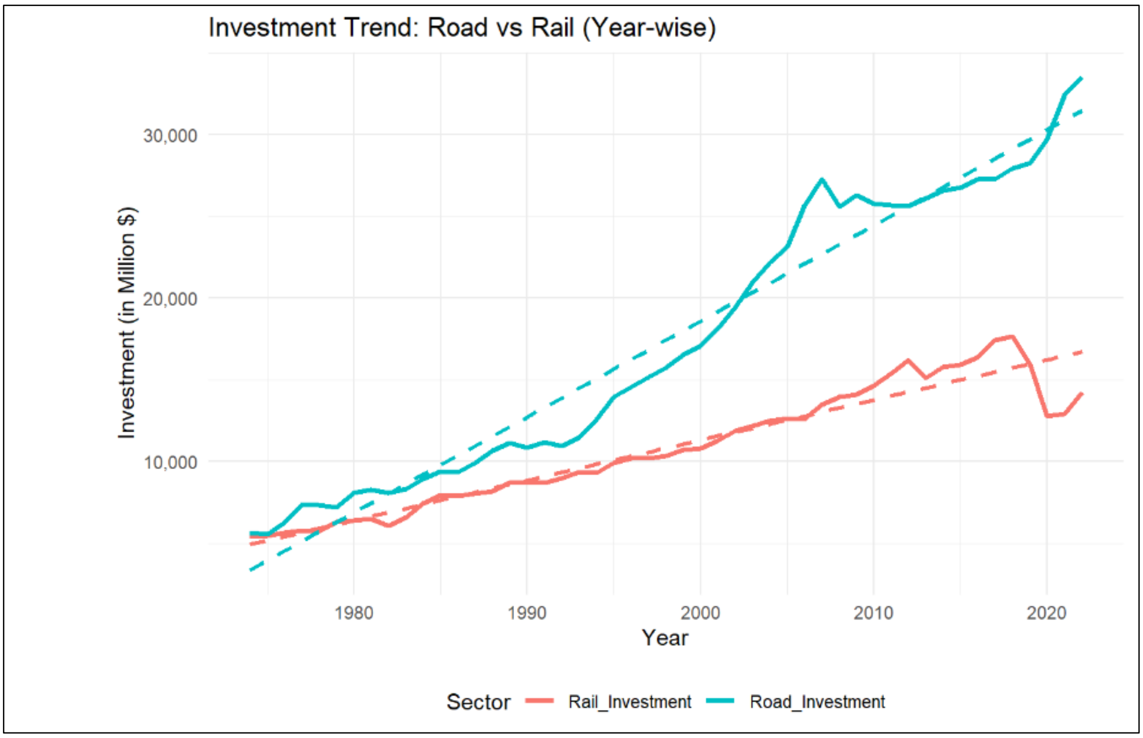

Figure 1: Road vs Rail Investment Trend

Figure 1 depicts the investment trends in road and rail infrastructure in a year wise fashion. With the increasing investment in rail, it is quite evident that there has been a change in funding preference. This could potentially be attributed to the rising sustainability needs and demand for efficient public transport systems. The forecasted trend line suggests that rail investment is going to continue outshining road investments in the near future.

Sector-Wise Employment Distribution

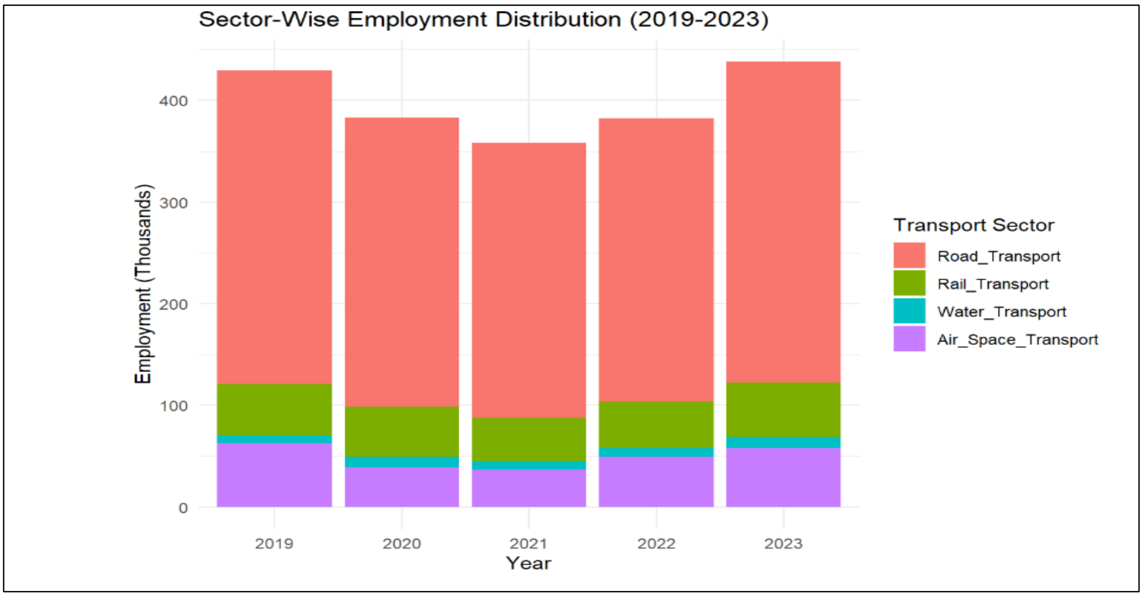

Figure 2: Sector-Wise Employment

In figure 2, we have a stacked bar chart depicting employment distribution across transport sectors from 2019 to 2023. Road transport clearly has the largest share of employment, while air, rail, and water sectors contribute smaller shares. This pattern suggests a steady labour demand within road transport and highlights potential areas for employment growth in the rail and air sectors.

Average Weekly Earnings by Sector

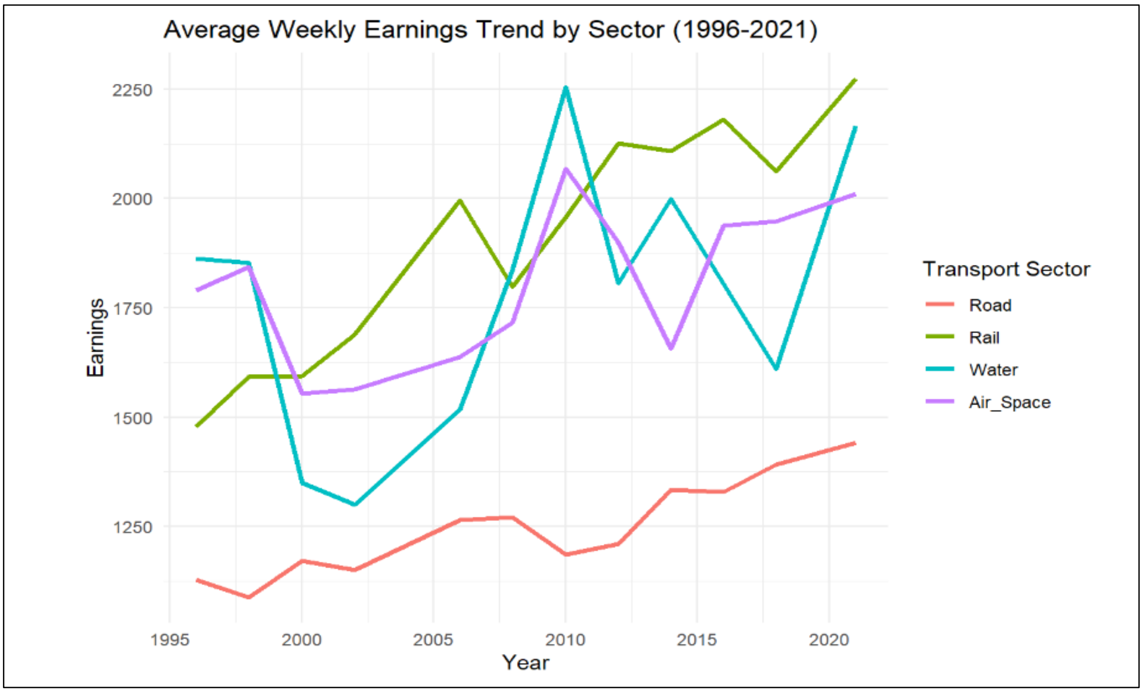

Figure 3: Average Weekly Earnings

Figure 3 presents a line chart showing the trends in average weekly earnings in transport sectors over time. A clear pattern is observed in wage distribution, highlighting the disparities in earnings across these sectors. Air and space transport consistently has higher earnings, which could be due to the fact that these jobs require specialized skills and qualifications to be pilots, engineers, technical staff that eventually gets reflected in the wages.

On the other hand, road transport shows the lowest average weekly earnings among the four sectors throughout. Some reasons for this could be relatively lower skill requirement and lower industry profit margins. Although the plot shows a steady growth in the rail earnings, it is clearly less pronounced as compared to the other sectors.

The gap between the highest and lowest paid sectors illustrates a significant income disparity in the transport industry. Making sense of these trends is crucial for policymakers and industry leaders who oversee this domain.

Forecast for Rail and Road Freight

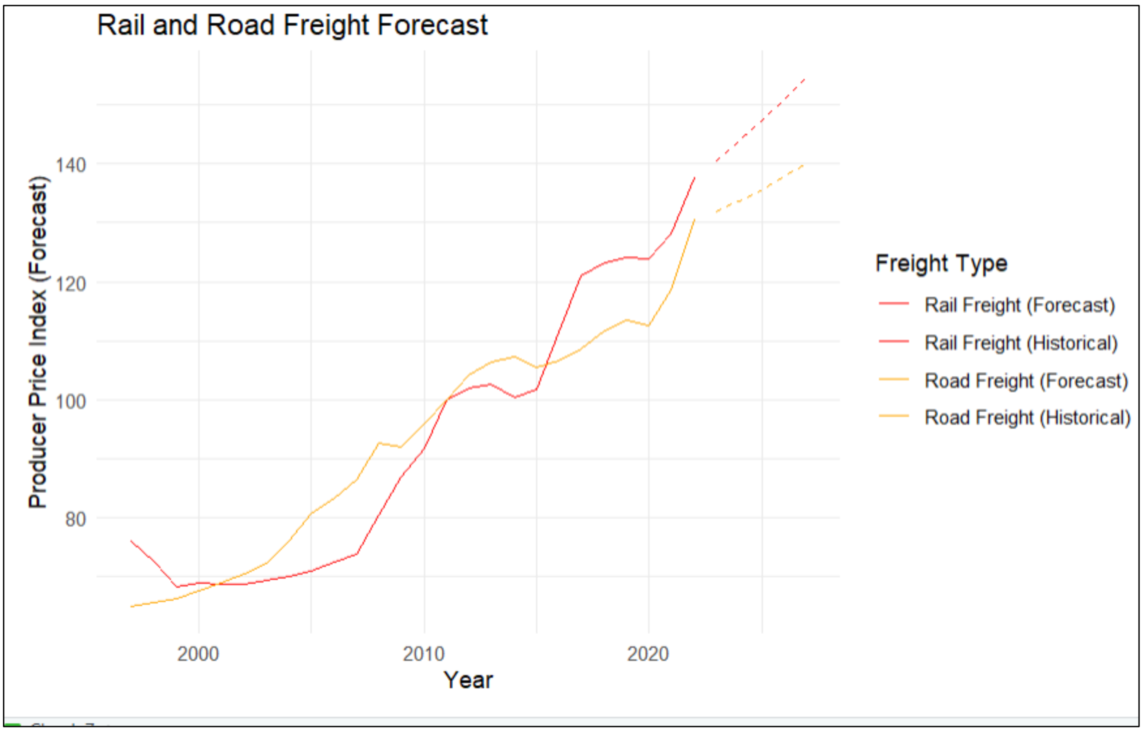

Figure 4: Rail and Road Freight Forecast

Figure 4 is a plot depicting historical and forecasted Producer Price index(PPI) for rail and road freight services. PPIs primarily serve as a tracker that records changes in selling prices received by domestic products. This can help us in understanding the cost trends withing the two industries.

We can observe how in the past both road and rail freight have exponentially increase, denoting increasing costs for running these services. The underlying reasons for this could be fluctuations in fuel rates, wage growth, increased demand for freight transport due to domestic trade and logistics to name a few.

The forecasted data indicates that the upward trend in cost will continue, with rail freight denoting a steeper rise when compared to road freight. This could possibly be due to the higher operational costs when it comes to rail freight, increased infrastructure investment, maintenance costs and so on.

These insights can aid transport and logistics stakeholders to have a better idea about cost trends and make necessary changes to their operations. Freight companies can look forward to improving fuel efficiency and adopt cost effective technologies.

Road Data Analysis

State-Wise Road Spending Analysis

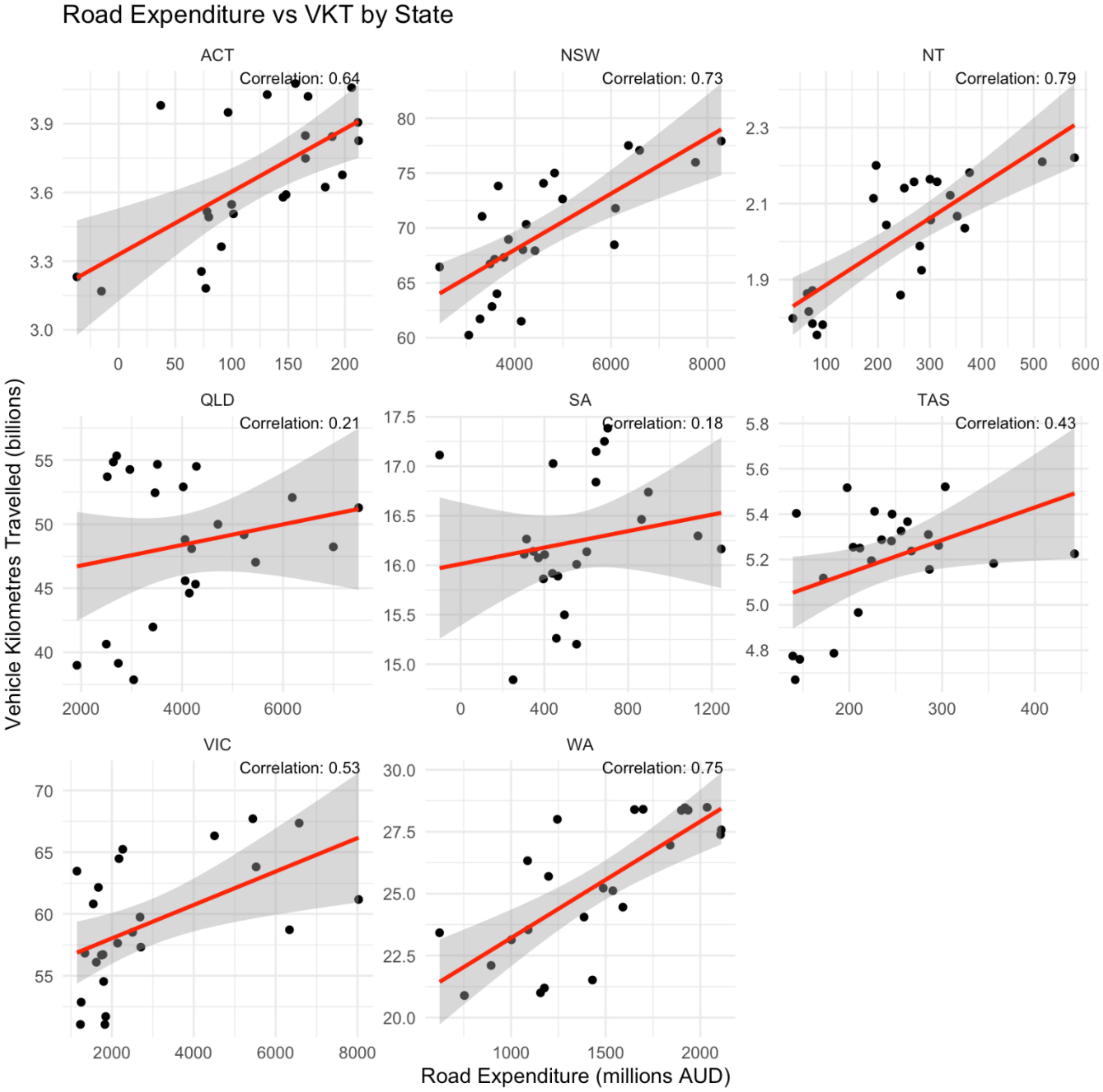

Figure 5 Road Expenditure and Vehicles Usage on Road

```
## State correlation
## <chr>  <dbl>
## 1 ACT   0.644
## 2 NSW   0.733
## 3 NT    0.790
## 4 QLD   0.206
## 5 SA    0.179
## 6 TAS   0.435
## 7 VIC   0.530
```

```
## 8 WA        0.749
```

Figure 5 presents the correlation between road expenditure (in millions AUD) and vehicle kilometres travelled (VKT) across Australian states. This analysis demonstrates how infrastructure investment correlates with road usage, highlighting the disparities in correlation strength across states.

From the correlation analysis, Northern Territory (NT) exhibits the highest correlation (0.79), suggesting a strong relationship between expenditure and vehicle usage. This indicates that increased investment in road infrastructure in NT is closely tied to higher road usage. South Australia (SA) shows the lowest correlation (0.18), suggesting that road investments do not significantly impact vehicle usage in this region. Similar low correlations are observed in Queensland (0.21) and Tasmania (0.43), where infrastructure investment appears less connected to vehicle usage trends.

High Expenditure, Moderate Usage: States like New South Wales (NSW), Queensland (QLD), and Victoria (VIC) show substantial road investments, exceeding 8,000 million AUD, yet vehicle usage does not proportionally increase. For instance, Queensland allocated over 4,383 million AUD to roads in 2021-22, but the VKT was only 54.5 billion, indicating that other factors may limit road usage growth in these areas.

Low Expenditure, Low usage Trends: Lower correlation in states such as Tasmania, Queensland, and South Australia may be attributed to demographic trends, with populations in these regions being smaller or shifting towards urban centres like NSW and Victoria, where job opportunities and infrastructure are more robust.

Trends in Kilometres Travelled by Vehicle Type

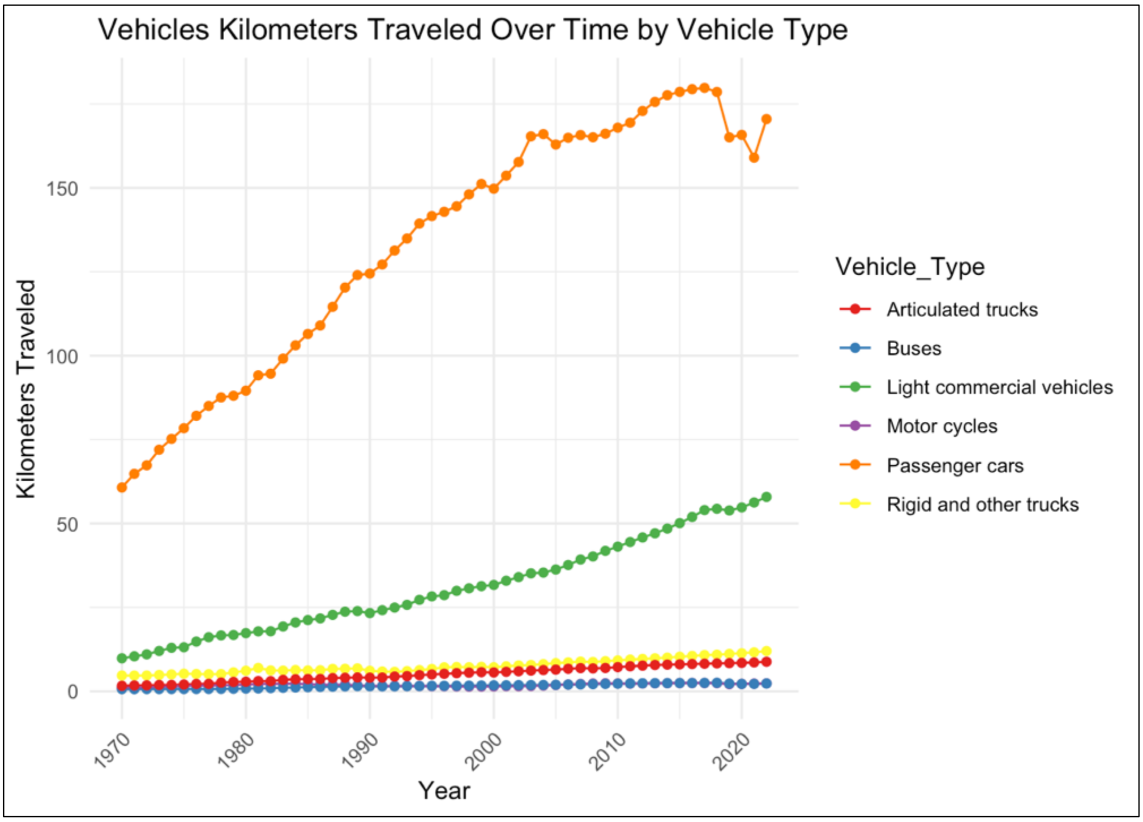

Figure 6 Trends of kilometres travelled over time for each vehicle type

Figure 6 illustrates the trends in vehicle kilometres travelled (VKT) from 1970 to 2023 across different vehicle types.

Passenger Cars: The orange line shows that passenger cars contribute the most to total VKT, with a consistent upward trend from 1970 through 2023. This reflects the widespread reliance on passenger cars as the primary mode of transportation in Australia.

Light Commercial Vehicles: The green line, representing light commercial vehicles (e.g., pickup trucks, vans, SUVs), shows steady growth over time. This increase could be linked to the rise of e-commerce and the growing demand for delivery services, which rely heavily on light commercial vehicles.

Other Vehicle Types: Other vehicle types, such as buses, rigid and articulated trucks, and motorcycles, display relatively stable VKT over the years. However, the impact of trucks on road infrastructure is notable due to their size and weight, which can have significant implications for road maintenance and wear, even if their VKT is lower than that of passenger cars.

Expenditure on Roads and vehicle registration in AU regions

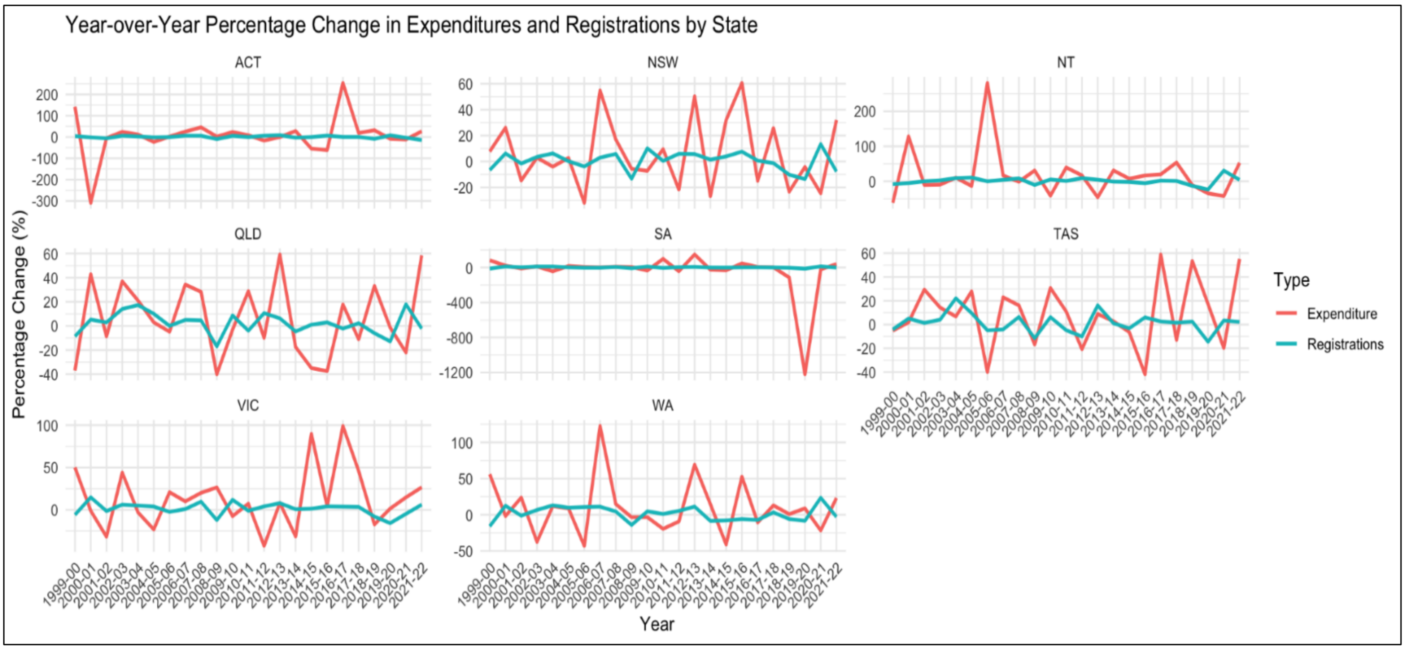

Figure 7 Year to Year Percentage Change in Expenditures and Registrations

This chart compares the trends of road related revenue and expenditure with vehicle registrations across regions of Australia.

The purpose is to see if an increase in road spending is followed by an increase in vehicle registrations, which would suggest that infrastructure investment drives vehicle ownership and road usage.

Red line refers to expenditure and blue refers to vehicle registrations. We can see that for Northern Territory, ACT and Western Australia show strong alignment in expenditure and vehicle registrations and weak alignments for regions like Tasmania and Queensland. This supports our analysis from road expenditure and Vehicle Kilometres Travelled(billions) earlier.

This suggests that road investments alone may not be enough to drive vehicle ownership in certain regions, and other factors such as population growth, public transportation availability and economic conditions, likely play a significant role in influencing vehicle registrations.

5-Year Forecast of Vehicle Usage (ARIMA Model)

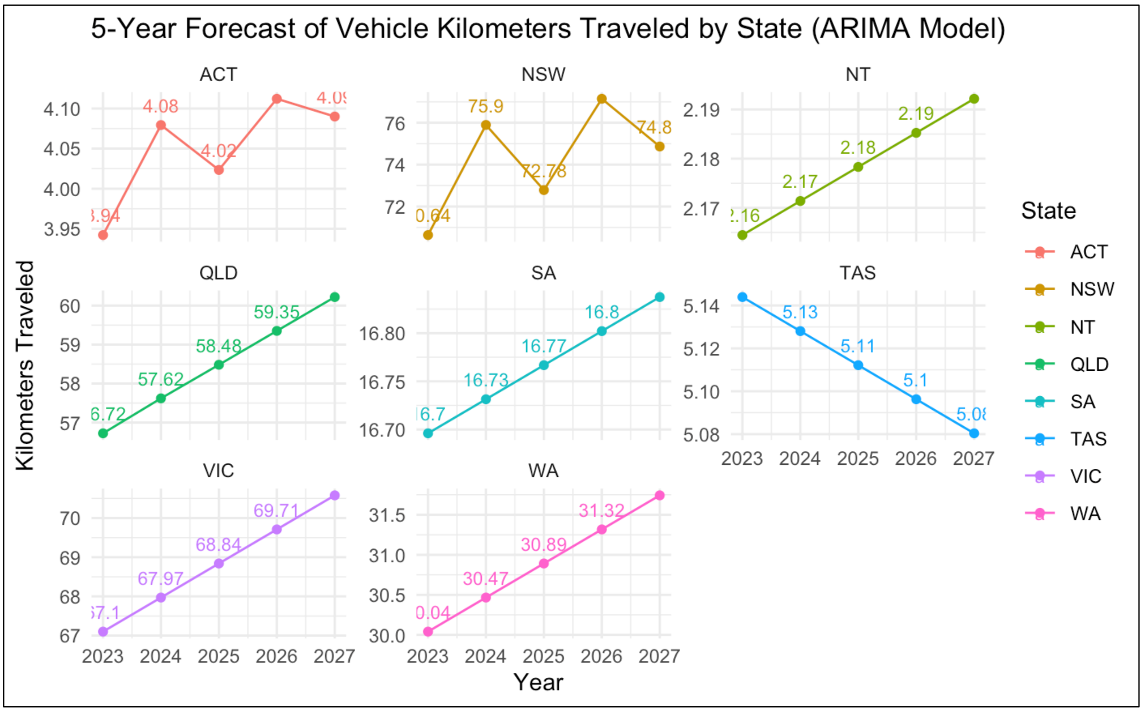

Figure 8: Forecast of 5 years ahead vehicle usage

Figure 8 shows the projected vehicle kilometres travelled (VKT) across Australian states over the next five years, based on ARIMA modelling.

Most regions, including QLD, SA, VIC, and WA, display steady upward trends in vehicle usage, indicating a forecasted increase in road utilization. ACT and NSW also show an upward trend but with notable fluctuations, especially a projected dip around 2025. Tasmania is the exception, with a slight downward trend over the forecast period, suggesting a potential decline in vehicle usage. This forecast highlights regional differences in projected vehicle usage, with most states expecting growth, while Tasmania may see reduced road activity.

Transport Infrastructure Engineering Work by Public and Private Sector

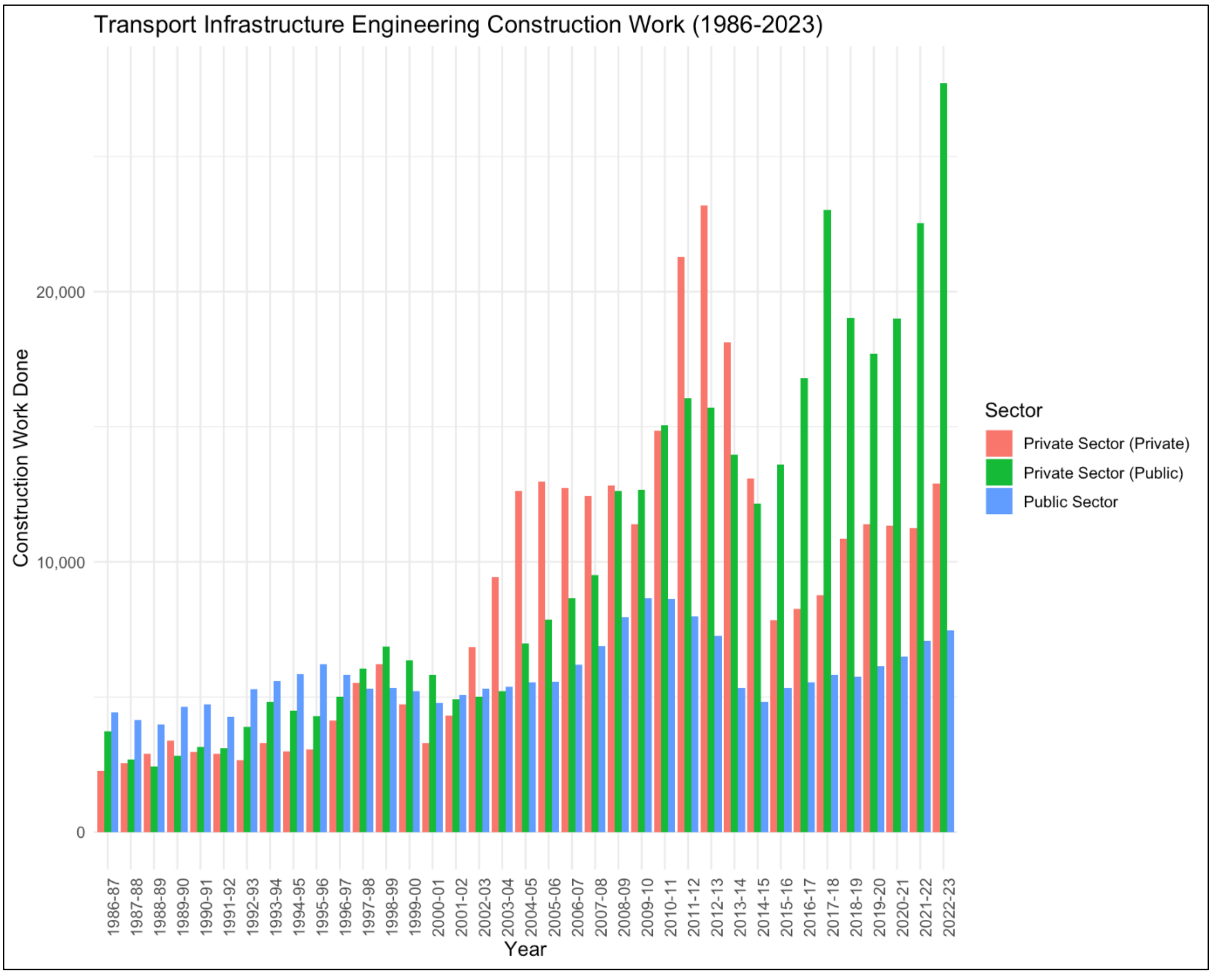

Figure 9: Transport Infrastructure Engineering Construction by Private and Public Sector

The chart illustrates the contributions of the public and private sectors to transport infrastructure construction in Australia from 1986 to 2023, measured in million dollars. These are infrastructure projects that are privately/govt. funded and executed, such as roads, bridges, or rail projects built for private/public use or ownership.

The orange line depicts value of transport infrastructure engineering construction work done, by the private sector for the private sector, adjusted by chain volume index, 2022-23 prices, the green line refers to value of transport infrastructure engineering construction work done by the private sector for the public sector, adjusted by chain volume index, 2022-23 prices and the blue line refers to value of transport infrastructure engineering construction work done by the public sector, adjusted by chain volume index, 2022-23 prices.

Over time, we see a marked shift in public infrastructure work being outsourced to the private sector. This is evident as the green bars (private sector's public work) grow significantly compared to the public sector's contributions. This shift towards increased private sector involvement, especially in public-private partnerships for large infrastructure projects, can be attributed to several policy changes and economic strategies implemented by the Australian government over recent decades. The public sector (blue bars) remains a consistent, though smaller, player in construction work, with less volatility than the private sector.

Rail

Infrastructure Spending Trends

This section illustrates the trend in infrastructure spending across various sectors, including transport, energy, telecommunications, and water, over several financial years. It is evident that transport consistently receives the highest level of investment in Australia.

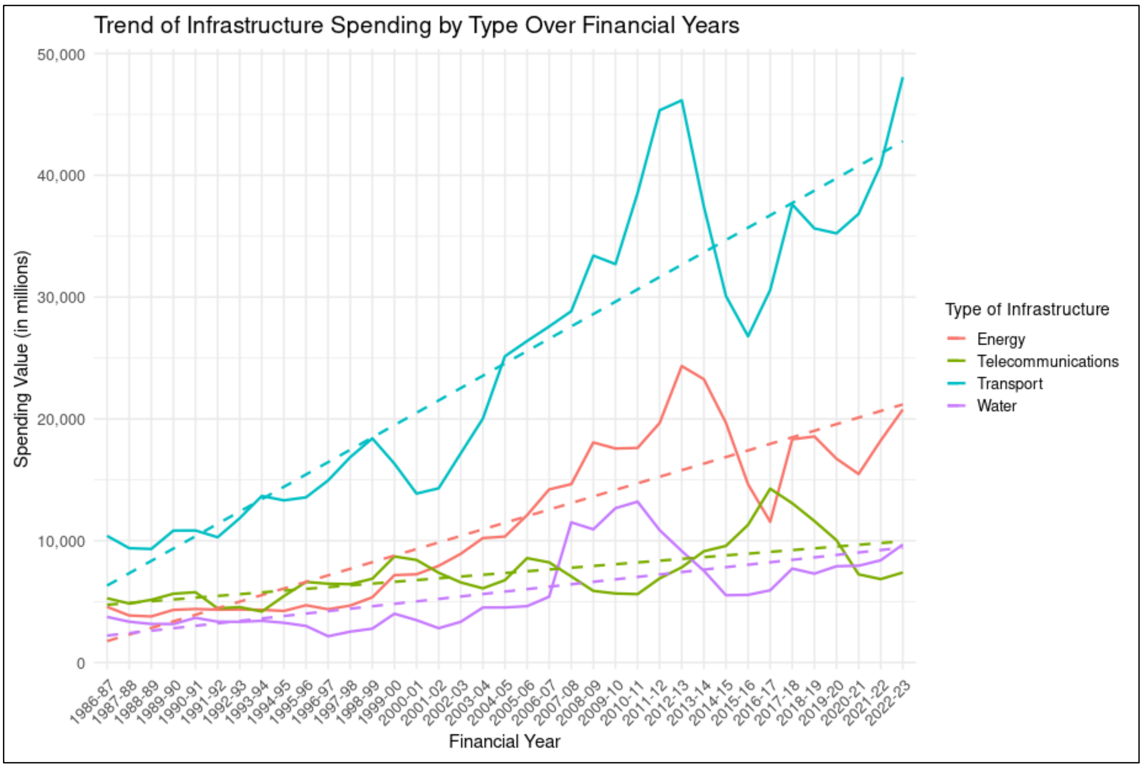

Figure 10 Trend of Infrastructure Spending

Public Transport Patronage Analysis

Plots in this section visualize the patronage for public transport modes, including heavy rail and light rail. Data is presented with distinct lines for each type and city, focusing on how usage has evolved over time. Additionally, the combined analysis of both heavy and light rail allows for comparison to observe patterns in public transportation growth.

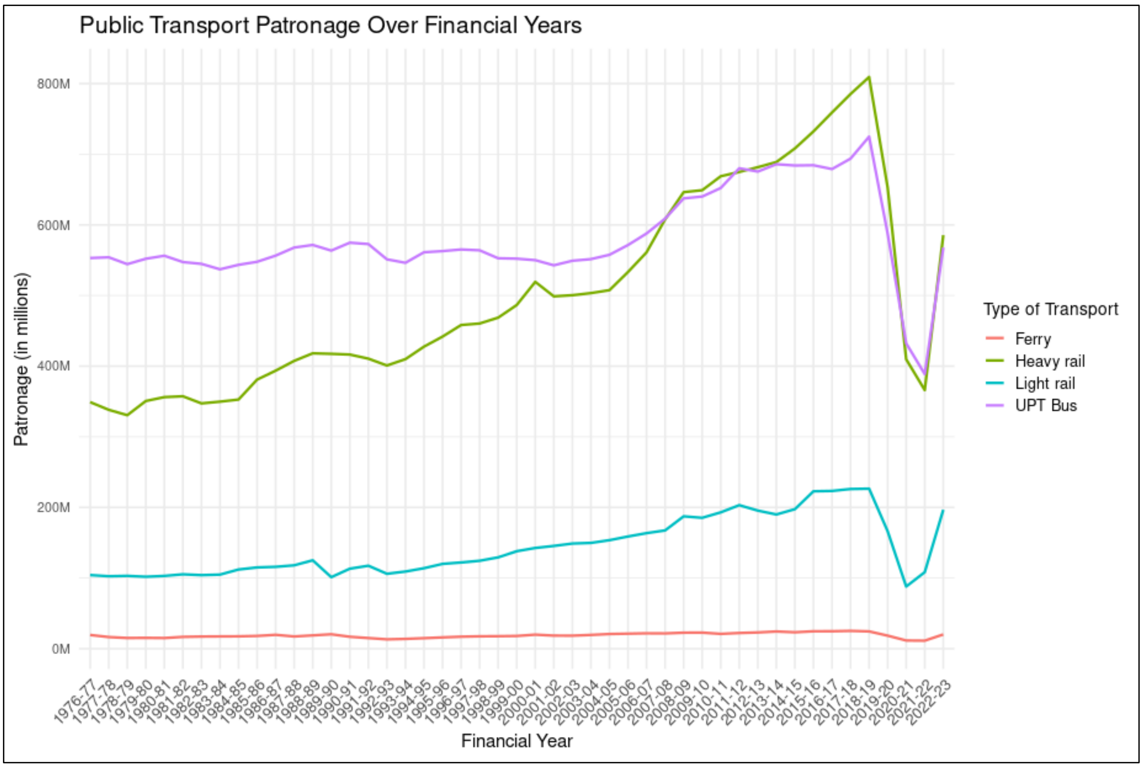

Figure 11 Patronage Over Financial Years

Heavy rail and UPT bus have the highest patronage, with both showing steady growth until a noticeable drop in 2019-2020 financial year. Light rail shows a consistent increase over the years, while ferry usage remains relatively low and stable compared to other modes.

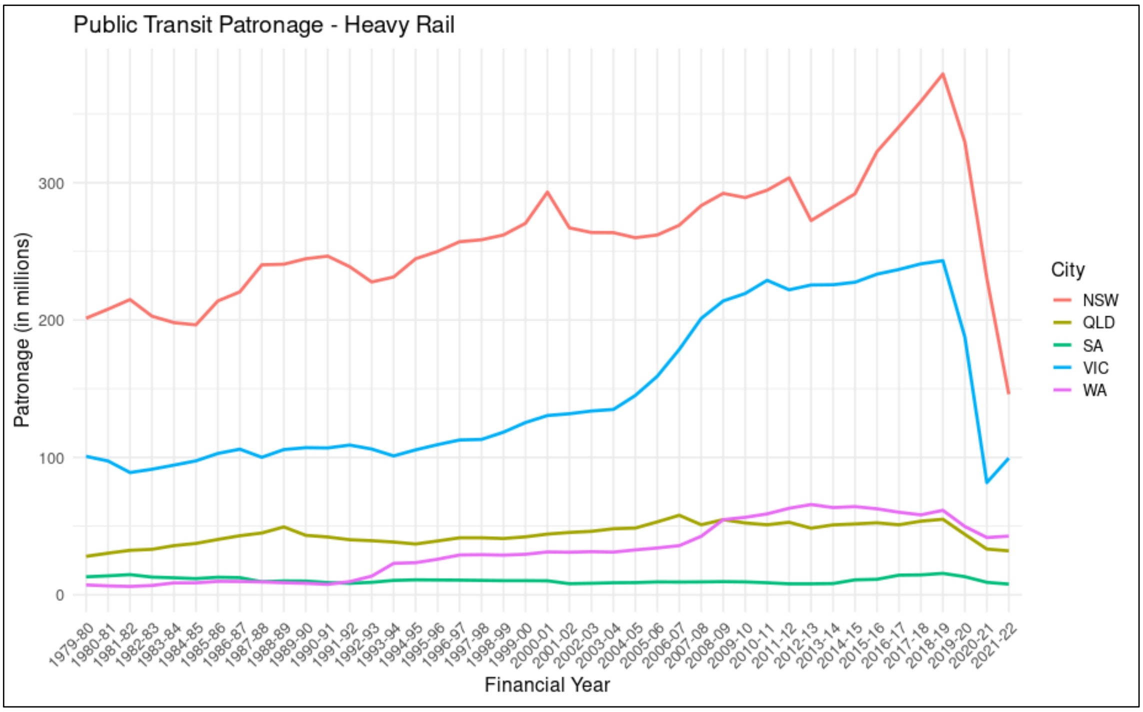

Figure 12 Patronage – Heavy Rail by States

Heavy rail is a high-capacity transit system typically used for long-distance.

Examples in Australia are:

- Melbourne Metro (https://www.metrotrains.com.au/)
- Victoria Vline (https://www.vline.com.au/)
- Sydney Trains (https://transportnsw.info/)

Figure 12 shows a drop in patronage for NSW and VIC indicating a major disruption, potentially due to the pandemic's impact on public transport usage. VIC and WA show a slight rebound at the very end, hinting at a partial recovery or stabilization post-decline. QLD, SA and WA have consistently lower patronage levels compared to NSW and VIC.

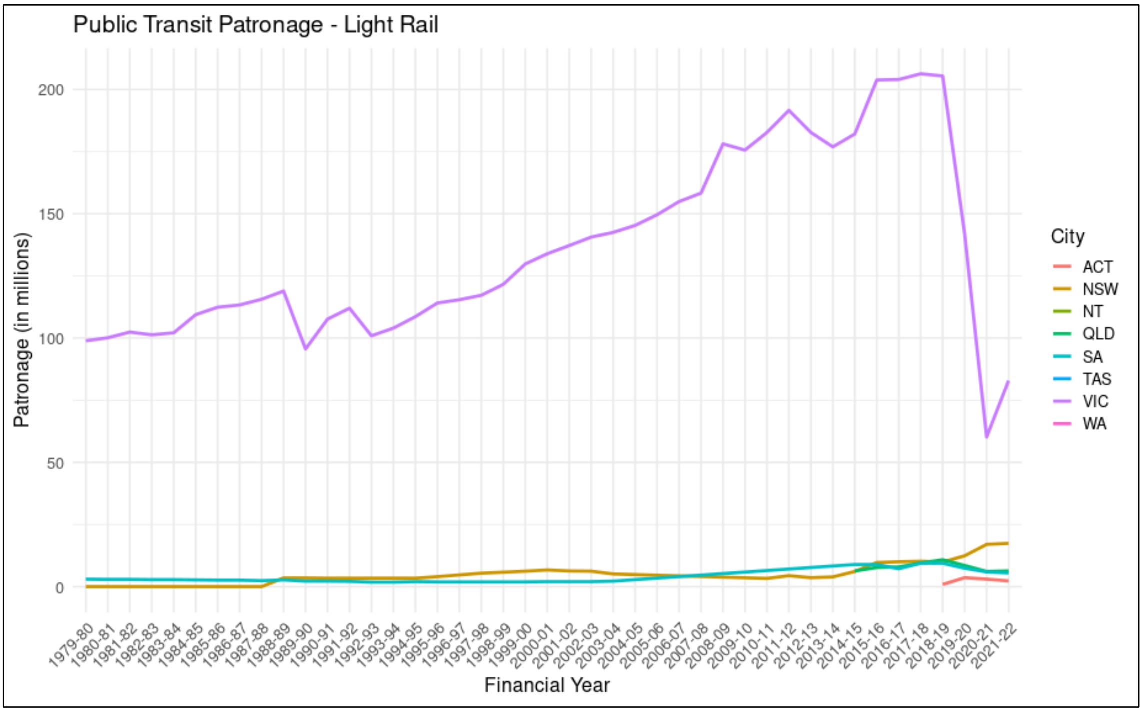

Figure 13 Patronage – Light Rail

Light rail is a medium capacity transit system, typically lighter and slower than heavy rail. It runs on a combination of exclusive tracks and shared roadways. Examples include Melbourne Tram or Sydney Light Rail.

Figure 13 shows that Victoria has the highest light rail patronage, showing a steady increase until a sharp decline around 2019–20. This substantial decrease likely reflects the impact of external factors, such as the COVID-19 pandemic.

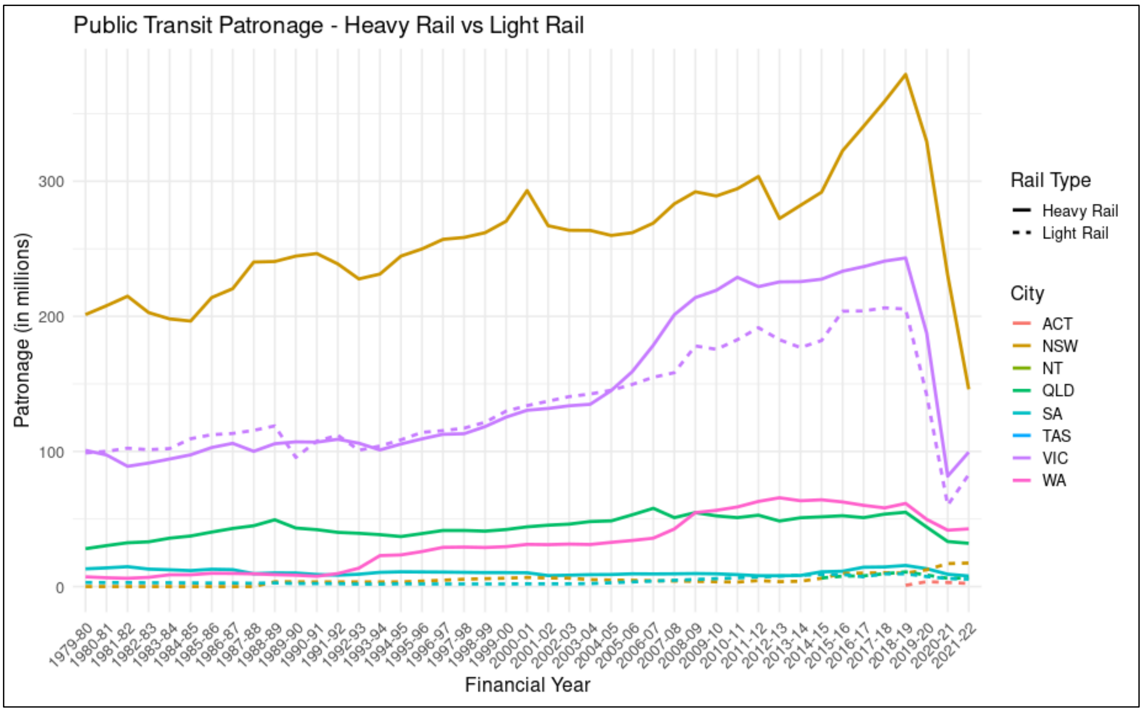

Figure 14 Patronage – Heavy vs Light Rail

VIC Victoria) shows high patronage for both heavy and light rail, with light rail increasing steadily over the years and heavy rail showing strong growth followed by a sharp decline around 2019–20, likely due to external disruptions such as the COVID-19 pandemic.

NSW (New South Wales) displays consistent dominance in heavy rail usage, with a marked rise over the years and a significant drop toward the end, reflecting the same external factor.

WA (Western Australia), heavy rail patronage remains relatively steady with some growth but at a lower magnitude compared to VIC and NSW.

ACT (Australian Capital Territory) shows limited but growing use of light rail, indicating newer developments or expansions in their light rail system.

Impact of External Factors:

The sharp decline seen across both heavy and light rail around 2019–20 points to major disruptions, aligning with global patterns of reduced public transport use during the COVID-19 pandemic.

The partial rebound seen at the end may indicate early recovery trends in public transport use.

Rail Expenditure by State

This section breaks down rail expenditure over time for different states (NSW, VIC, QLD, SA, etc.). **Figure 15** presents a line char that emphasizes regional variations in spending and identifies which states consistently lead in rail infrastructure investments. NSW typically shows the highest investment levels.

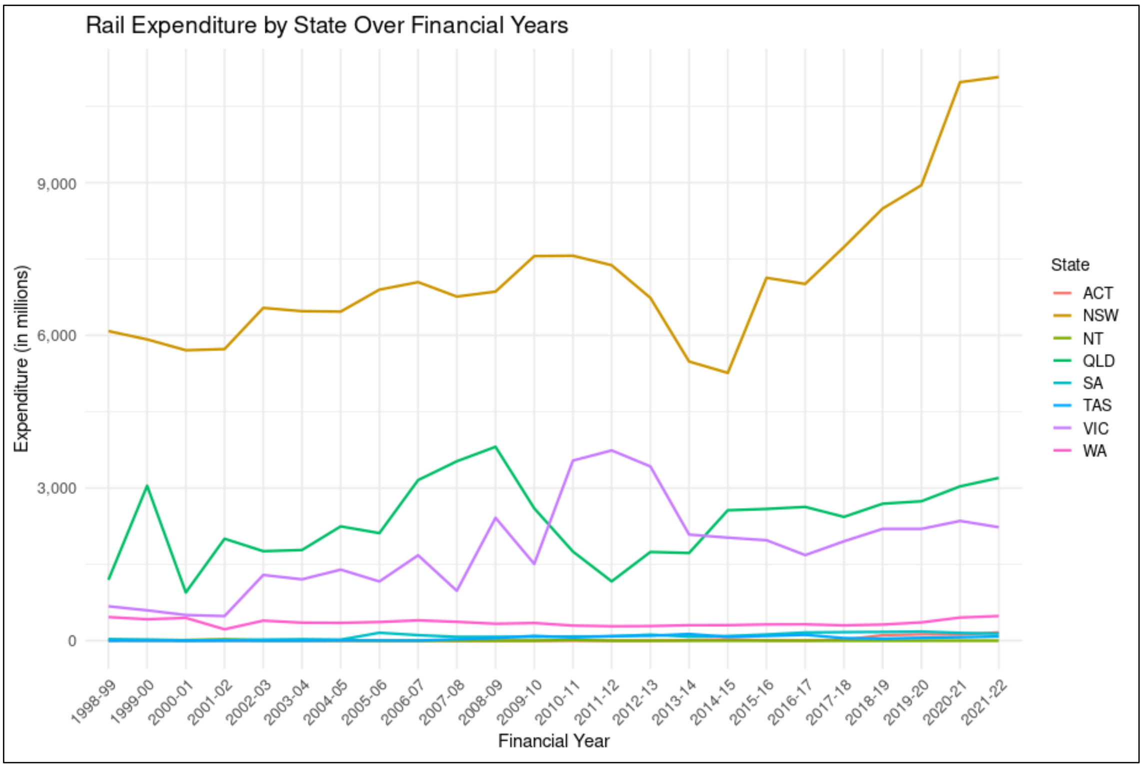

Figure 15 Rail Expenditure by States over the years

Non-Bulk Rail Freight Analysis

The non-bulk rail freight data illustrates state-wise freight volumes across financial years. A line plot with points highlights the trends and any significant fluctuations, offering insight into the movement of freight across states. Victoria and Western Australia show a significant increase in freight rail transport while Australian Capital Territory and Northern Territory have the lowest freight volumes.

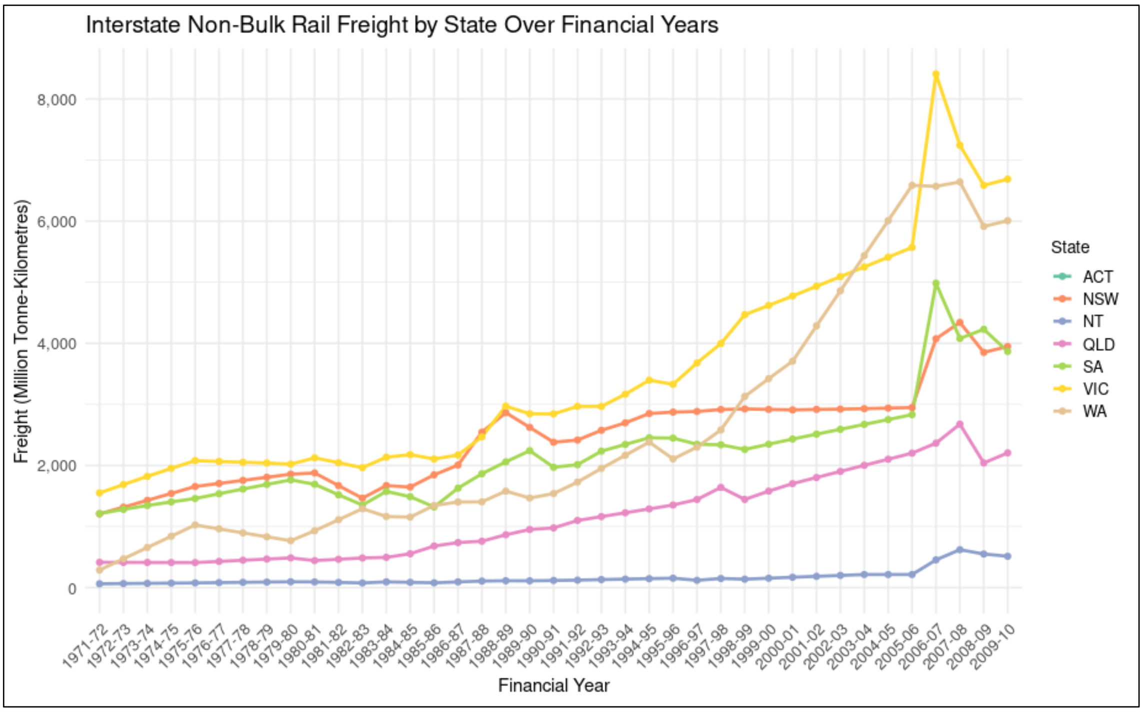

Figure 16 Rail Freight by State

Correlation Analysis Between Rail Expenditure and Patronage

This section includes scatter plots and statistical correlation analysis to assess the relationship between rail expenditure and patronage for selected states. The analysis aims to reveal whether increased spending correlates with higher patronage levels, with correlation coefficients provided for clarity.

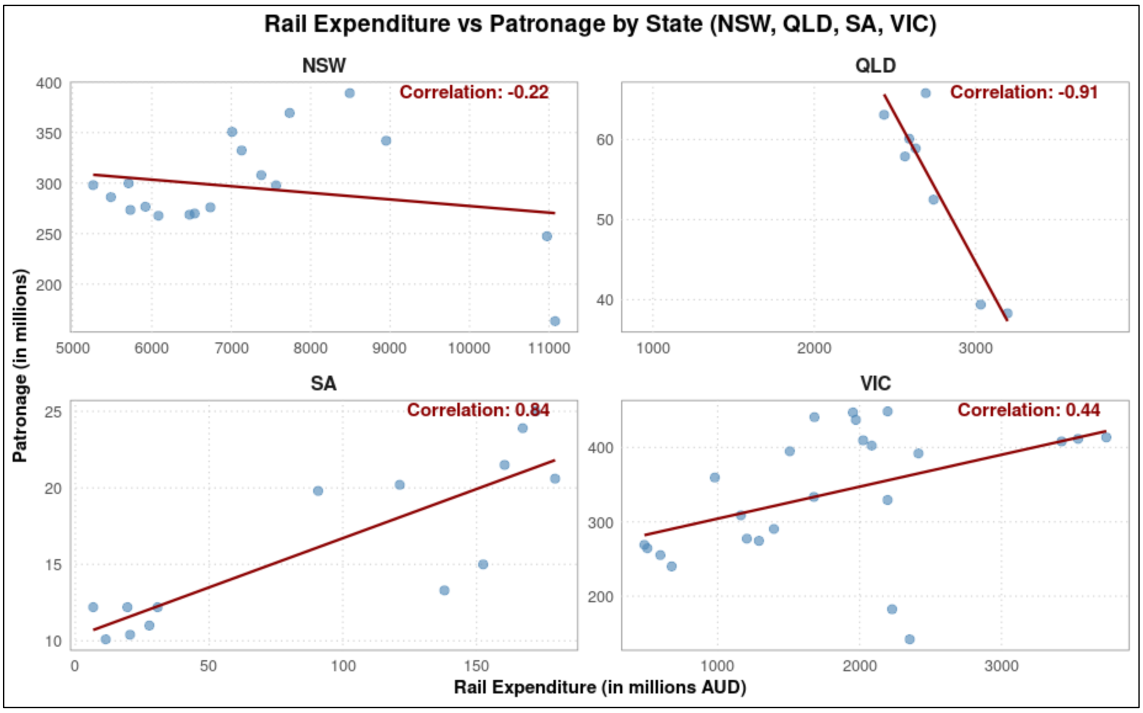

Figure 17 Correlation Rail Expenditure vs Patronage

The provided visualization displays scatter plots of rail expenditure versus patronage for four Australian states (NSW, QLD, SA, VIC), each accompanied by a fitted trend line and correlation coefficient. Below are detailed observations and comments:

NSW (New South Wales)

- Correlation Coefficient: -0.22
- Interpretation: The weak negative correlation suggests that higher rail expenditure does not significantly impact patronage levels. This could indicate that spending is not directly contributing to increased usage or that other influencing factors are at play.

QLD (Queensland)

- Correlation Coefficient: -0.91
- Interpretation: The strong negative correlation points to an inverse relationship where higher rail expenditure is associated with decreased patronage. This may imply external factors.

SA (South Australia)

- Correlation Coefficient: 0.84
- Interpretation: The strong positive correlation suggests a close relationship between rail expenditure and increased patronage. This indicates that investments in rail infrastructure are effectively contributing to higher usage in SA.

VIC (Victoria)

- Correlation Coefficient: 0.44

- Interpretation: The moderate positive correlation implies a somewhat positive relationship between rail expenditure and patronage. While higher expenditure appears to contribute to increased patronage, the correlation is not strong enough to suggest it is the sole factor.

General Observations

- Variability in Correlations: The states demonstrate diverse relationships between rail expenditure and patronage, highlighting that the effectiveness of investments can differ significantly across regions.
- Policy Implications: States such as SA, which show a strong positive correlation, may benefit from continued investment in rail infrastructure. Conversely, states like QLD should reassess their expenditure strategy to address the observed negative correlation.
- Further Analysis: To gain a deeper understanding, it would be useful to explore additional factors influencing these correlations, such as population density, availability of alternative transportation options, or the condition and age of existing rail infrastructure.

Trend Analysis of Movements and Interactions in Different Routes

The Interaction formula calculates the ratio of Population to distance (km), which might be used to represent an efficiency or impact measure, such as population influence per kilometer of distance. It shows how much of a certain 'route' is present per unit distance, which could be relevant in fields such as geography, logistics, or demographic studies.

Interaction =[@[Pop_Product]]/[@[distance (km)]]

"Pop_Product" is the multiplication of *Population 1* and *Population 2*

"Distance" is in kms between *Population 1* and *Population 2*

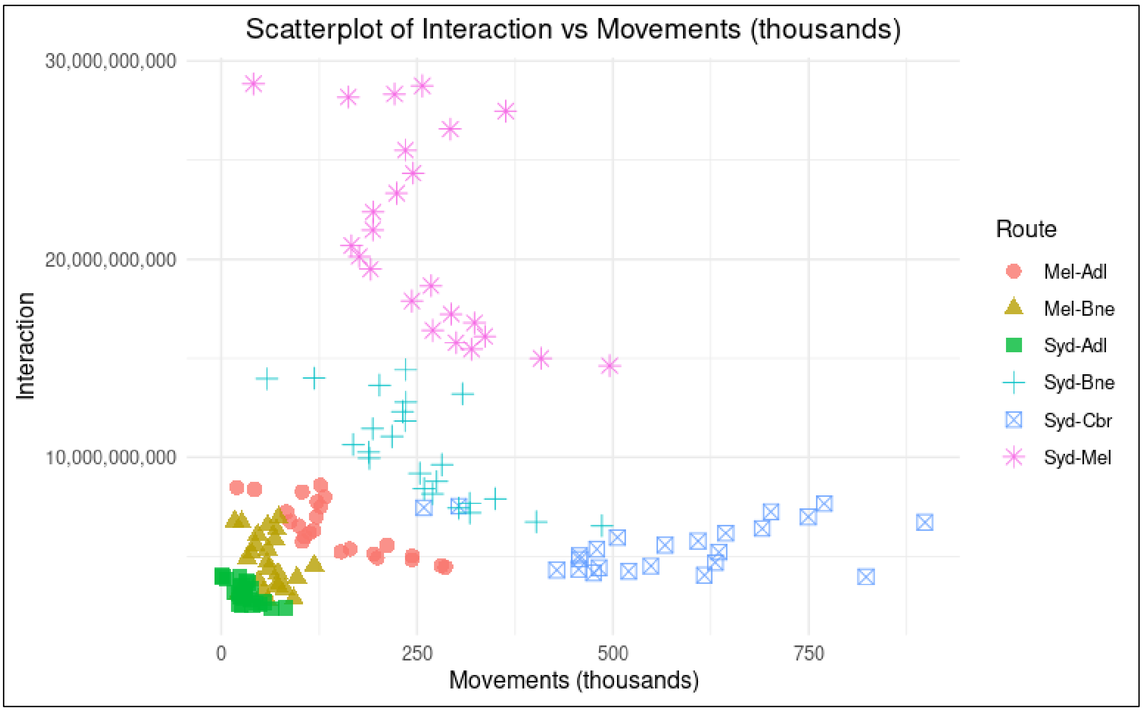

Figure 18 Correlation Rail Expenditure vs Patronage

The scatter plot displays the relationships between Movements and Interaction, with each route represented by a distinct color and shape. This helps in quickly identifying trends specific to different routes.

The **Syd-Mel** route (shown with asterisks) has significantly **higher Interaction** values compared to other routes. We can also see that movements is not correlated with the Interaction.

Syd-Cbr even with low Interaction, it possesses the **highest movements.**

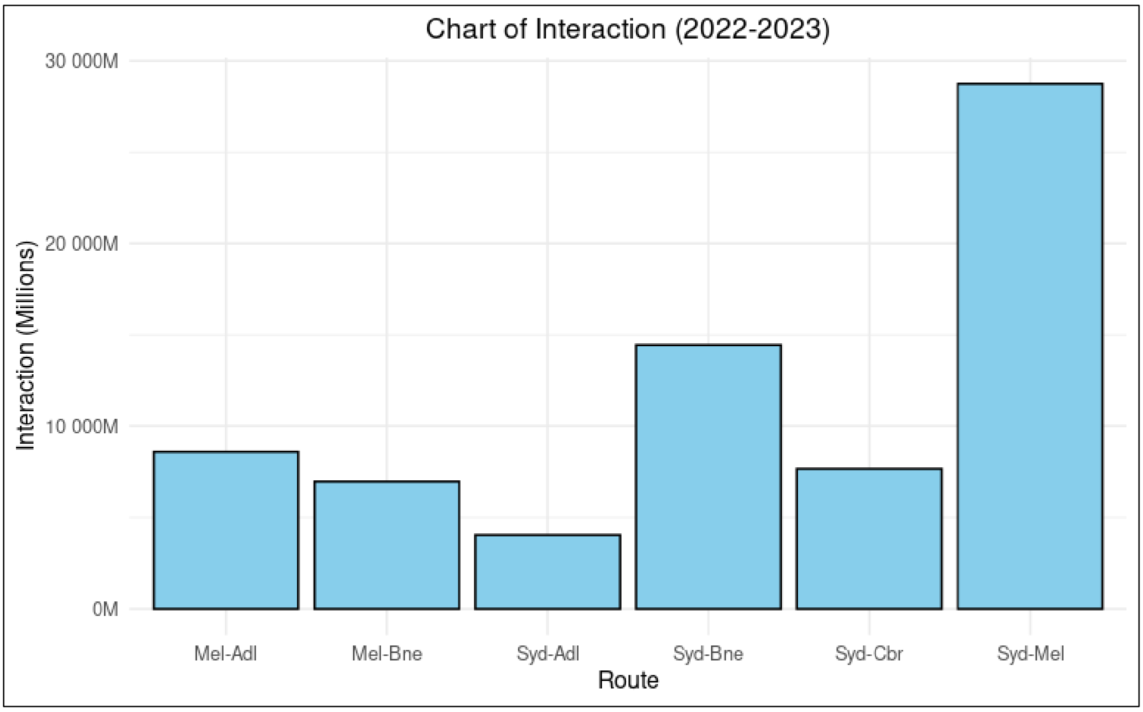

Figure 19 Bar Chart of Interaction

Syd-Mel (Sydney to Melbourne):

- High Interaction Volume: The interaction on the Syd-Mel route is notably the highest among all the routes shown, exceeding 25 million interactions. This suggests that this route is highly significant, potentially due to strong economic, business, or population growth between the two largest cities in Australia.

Syd-Cbr (Sydney to Canberra):

- Lower Interaction Volume: In contrast, the Syd-Cbr route shows a relatively modest interaction volume, possibly under 10 million. This may be due to lower population in Canberra.
- Contextual Factors: This route may primarily serve governmental and administrative purposes, given Canberra's status as the capital. The lower interaction could suggest a less robust transportation demand compared to larger, commercially active routes like Syd-Mel.

These observations highlight the significant difference in interaction levels between the major economic route of Syd-Mel and the more regionally focused Syd-Cbr route

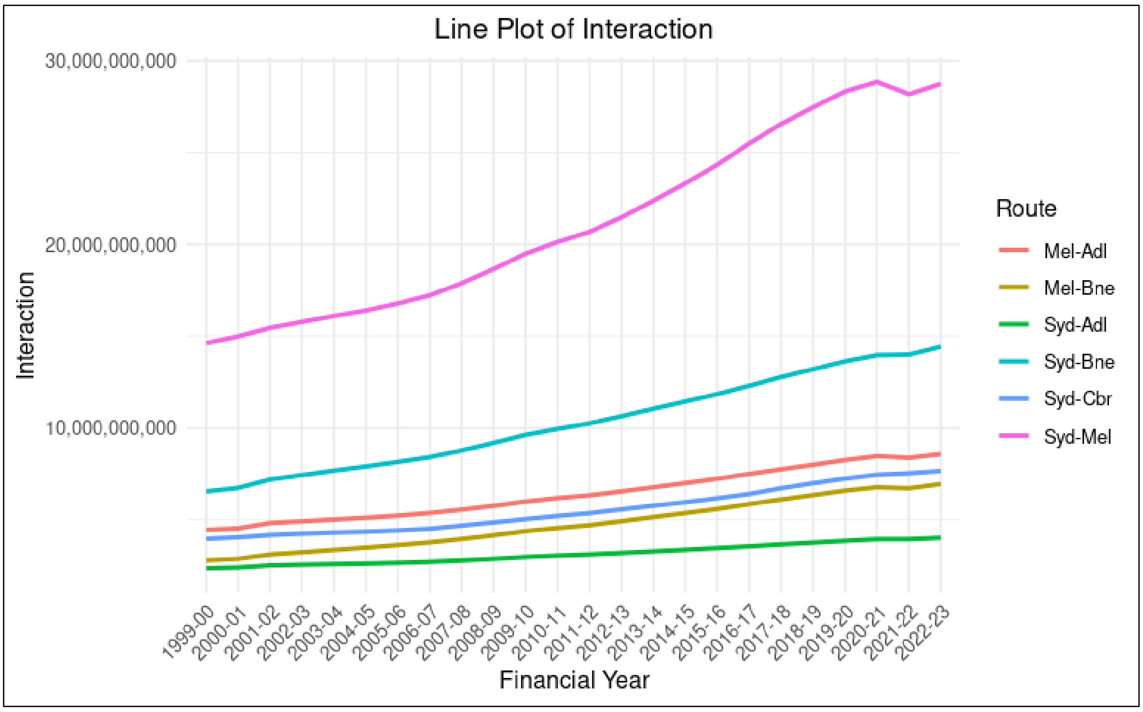

Figure 20 Line Plot of Interaction

The Syd-Mel (Sydney to Melbourne) route clearly has the highest volume of interaction, significantly higher than all other routes.

Syd-Bne (Sydney to Brisbane) routes show consistent growth over the years, suggesting a sustained increase in demand or interaction, potentially due to population growth and economic factors.

The growth in Syd-Bne is notable but still maintains a substantial gap from Syd-Mel, highlighting the comparative importance of the Syd-Mel route.

The Syd-Cbr (Sydney to Canberra) route shows the lowest interaction volume among the listed routes, which might be due to the shorter distance and limited economic or commuting factors compared to larger city pairs. The steady increase may still indicate stable growth but with a lower overall importance in interaction terms.

Growth Rate by Route Over Time

Figure 21 shows the growth rates for different routes, represented through line plots that detail year-over-year changes. The visualization helps in identifying routes with significant growth or decline over the years.

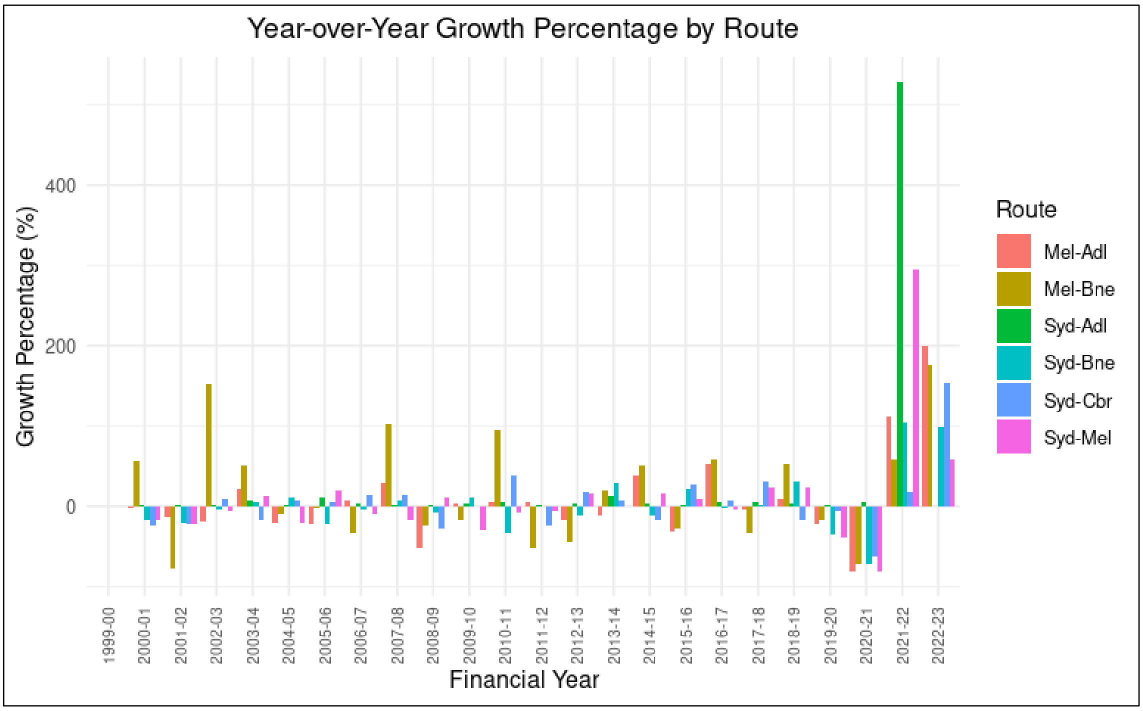

Figure 21 Movements Growth Year to Year

Syd-Mel (Sydney to Melbourne) appears to maintain relatively steady year-over-year growth with fewer extreme fluctuations compared to other routes. This stability underscores its importance and possibly more consistent demand.

Syd-Bne (Sydney to Brisbane) and Mel-Bne (Melbourne to Brisbane) show more variability in their growth percentages, suggesting they may be more sensitive to external factors or shifts in regional travel dynamics.

Route Efficiency Analysis

A bar chart highlights the efficiency of different routes in terms of movements per kilometer for the financial year 2022-23. This analysis identifies which routes achieve the highest efficiency.

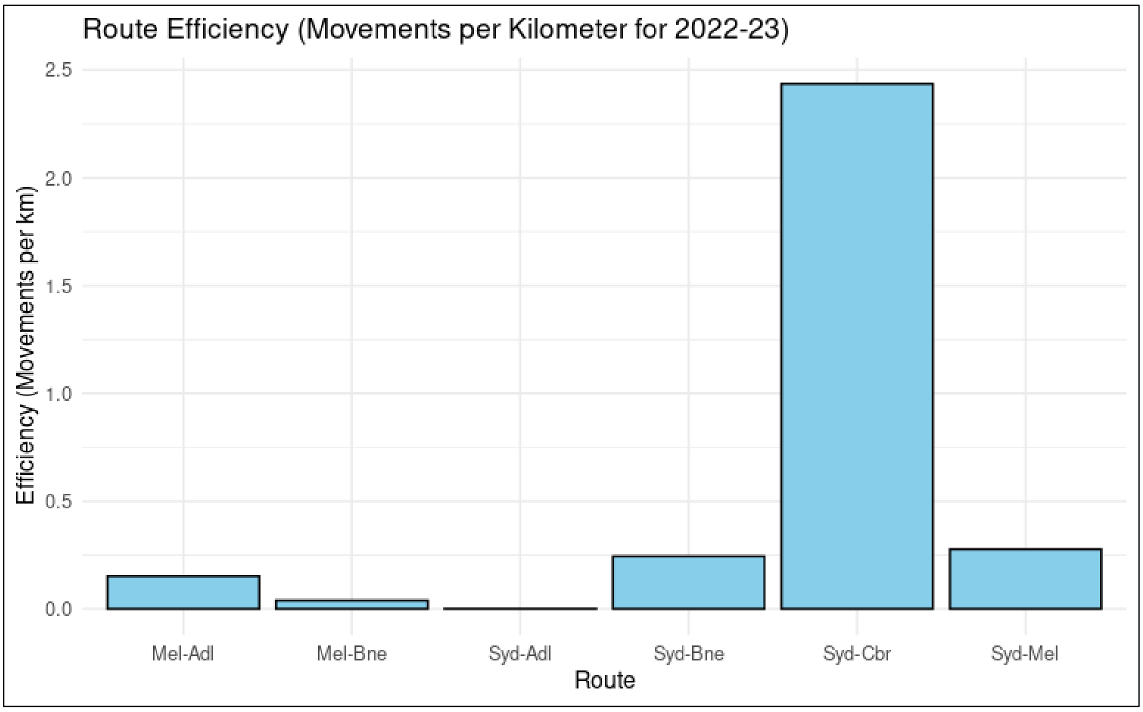

Figure 22 Route Efficiency

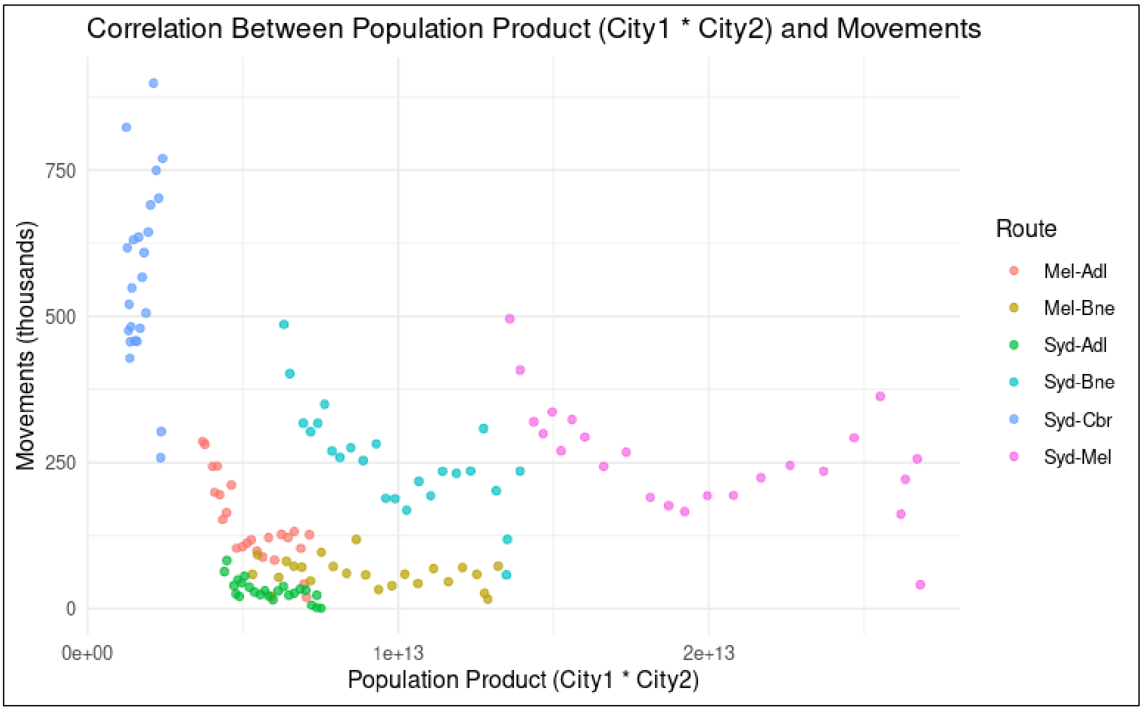

Figure 23 Correlation Population vs Movements

The scatter plot displays a relationship between the product of populations of two cities (x-axis) and the number of movements (y-axis). Syd-Cbr (Sydney to Canberra) shows the highest movement counts with data points extending well above 750 thousand, suggesting a high level of interaction despite a smaller population product compared to other routes.

Syd-Mel (Sydney to Melbourne) displays more distributed data points across a broader range of the population product, suggesting consistent movement patterns that align with the population interaction between two large cities

3D visualisation – Population, Interaction and Movements.

The 3D plot provides an effective way to visualize the multidimensional relationship between population product, movements, and interaction across different city routes.

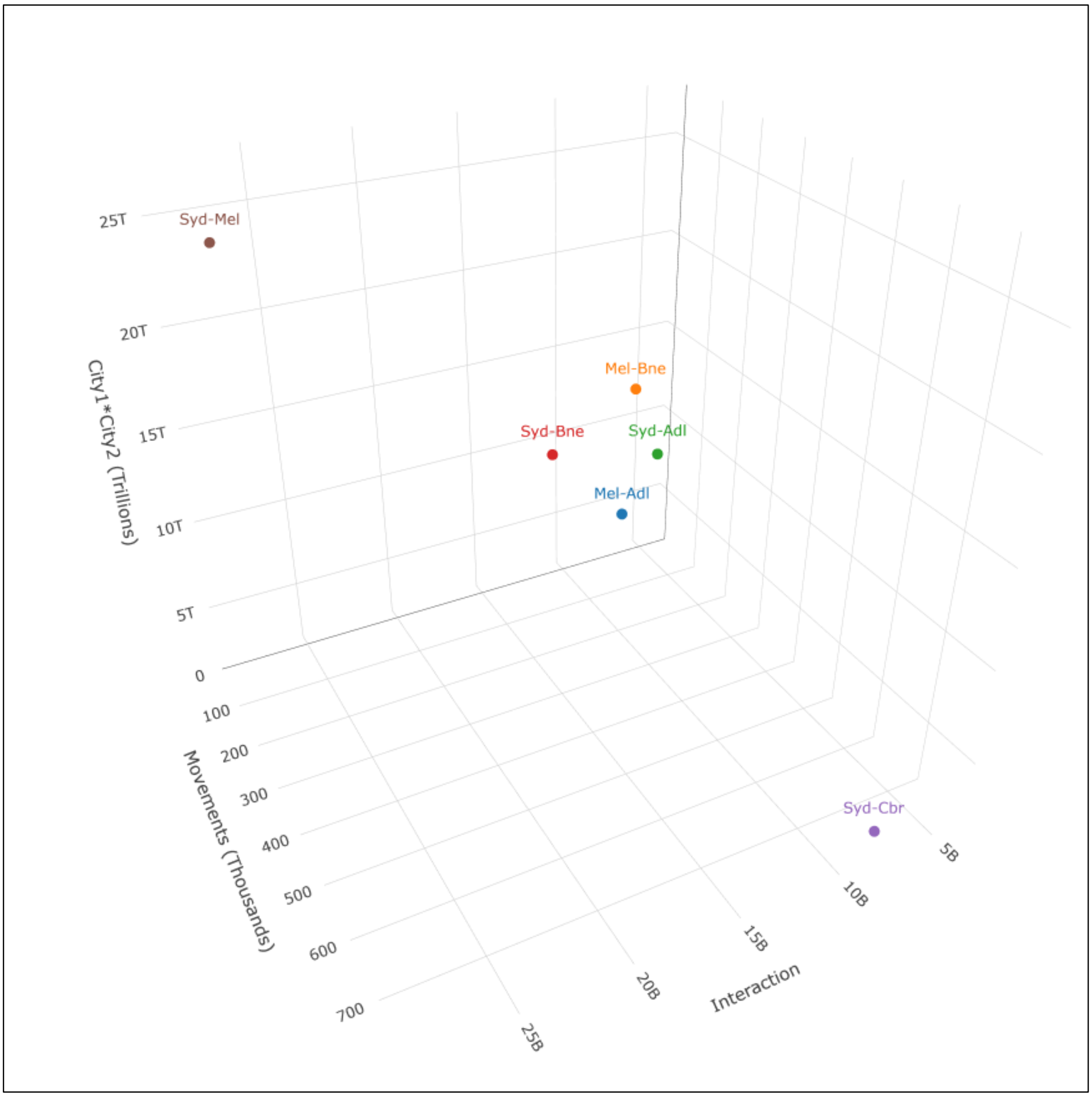

Figure 24 3D Scatter Plot Relationship Between Population, Movements and Interaction

Syd-Mel is a clear outlier, indicating that this route not only has a large population product but also higher interaction values. This suggests that beyond population, there are strong economic or strategic reasons for such high interaction.

Syd-Cbr being positioned separately could imply that despite a lower population product, the interactions are still notable due to factors like political or governmental relevance.

Conclusions

Economic Infrastructure

The analysis we performed reveals some key trends in investment, employment, wages and costs in the transport infrastructure revolving around road and rail domains. While road transport is employing the largest workforce, rail investment is having higher overall growth signalling a major shift towards sustainable transport. There are observable wage disparities between the various transport sectors, which could be contributed to specialized skills and qualifications required for air and space transport which offer the highest earnings. The rising costs in rail and road freight indicate a need to effectively handle cost management.

Road

An increase in road expenditure across Australian states does not universally result in a proportional rise in VKT. While some states experience a positive relationship, others show minimal change, indicating that road spending alone may not be sufficient to drive vehicle usage uniformly. Instead, VKT is influenced by a combination of demographic, economic, and infrastructure factors unique to each state.

This analysis suggests that policymakers should consider these regional differences when planning infrastructure investments, potentially complementing road spending with public transport initiatives or targeted investments in areas with growing demand.

Rail

Based on the analysis of rail data, particularly examining the correlation between expenditure and patronage, it is evident that Victoria, as the second-largest city in Australia, has experienced a significant positive impact on patronage growth aligned with the level of investment in infrastructure. The increased funding has directly contributed to greater utilization of rail services in the region.

Regarding specific routes, the Sydney-Melbourne corridor stands out as having the highest interaction level, largely due to the substantial distance between these two major cities (as illustrated in Figure 19). Although recent shifts in travel patterns might not directly correlate with population growth—particularly influenced by external factors like the COVID-19 pandemic—the Sydney-Melbourne route demonstrates strong potential for continued growth. This potential is supported by recent positive trends in rail movements observed post-pandemic (refer to Figure 21).

Furthermore, the Melbourne-Canberra-Sydney route presents a promising investment opportunity. The high level of interaction along this corridor suggests that it could serve as a critical transportation link, particularly for government employees and commuters traveling between these key cities. Investing in this route could not only enhance accessibility but also address the growing transportation needs in this highly interconnected region.

Data Sources:

1. **Table 2.2**: Total value of infrastructure engineering construction work done, adjusted by chain volume index (2022-23 prices). Source: ABS, 2023, Engineering Construction Activity. https://www.abs.gov.au/statistics/industry/building-and- construction/engineering-construction-activity-australia/latest-release.

2. **Table 2.3**: Total value of transport infrastructure engineering work done, adjusted by chain volume index (2022-23 prices). Source: ABS, 2023, Engineering Construction Activity

.   https://www.abs.gov.au/statistics/industry/building-and- construction/engineering-construction-activity-australia/latest-release.

3. **Table 5.5a**: Total public transport patronage by capital city (heavy rail, light rail, bus, etc.). Source: BITRE estimates (based on various transit authority reporting of patronage levels).

4. **Table 7.1**: Intercapital rail distances (passenger and freight terminals). Source: BITRE estimates, Australian Infrastructure and Transport Statistics Yearbook 2023-rail.https://www.bitre.gov.au/publications/2023/australian-infrastructure-and-transport-statistics-yearbook-2023/rail.

5. **Table 7.4**: Interstate non-bulk rail freight by state/territory of origin. Source: BITRE,Yearbook 2023-rail . https://www.bitre.gov.au/publications/2023/australian-infrastructure-and-transport-statistics-yearbook-2023/rail.

6. **Table 7.5** : Public transit patronage on heavy rail in Australian capital cities. Source: BITRE,Yearbook 2023-rail . https://www.bitre.gov.au/publications/2023/australian-infrastructure-and-transport-statistics-yearbook-2023/rail.

7. **Table 7.6**: Rail-related expenditure by Commonwealth and state governments. Sources: ABS, 2023, Consumer Price Index, Australia ; BITRE, Yearbook 2023-rail . https://www.bitre.gov.au/publications/2023/australian-infrastructure-and-transport-statistics-yearbook-2023/rail.

8. **Table 6.4**: Total vehicle kilometers traveled by state and financial year. Source: BITRE, Yearbook 2023-road .

9. **Table 6.8**: Number of registered road vehicles by year of manufacture and vehicle type. Source: BITRE, Yearbook 2023-road .

10. **Table 6.11**: Number of vehicles registered by state and financial year. Source: BITRE, Yearbook 2023-road .

11. **Table 3.1**: Road-related expenditure by state and financial year. Source: BITRE, Yearbook 2023-road .

12. **Table 1.1a** - Gross Value Added by Major Australian Infrastructure Industries: This table presents the annual gross value added (GVA) in millions of Australian dollars, adjusted to 2022-23 prices, for various infrastructure-related industries.

14. **Table 1.2a** - Australian Employment in Major Infrastructure Industries (Transport and Storage): This table contains employment figures (in thousands) for different sectors within the transport and storage industry.

15. **Table 1.3a and 1.3b** - Employment in Major Australian Transport Industries, by Gender: These tables show employment numbers (in thousands) for major transport sectors, broken down by gender (Male and Female).

16. **Table 1.6a** - Australian Producer Price Indexes, Transport Industry:This table presents annual producer price indexes for various sub-sectors of the transport industry such as road freight, rail freight, pipeline transport, and water freight.

17. **Table 1.7a** - Australian Population by State/Territory and Capital City:This table shows the estimated population of Australia's states and territories, along with their major capital cities, from 1973 to 2004.

References

Infrastructure Australia. (2022). 2022 Infrastructure Market Capacity Report. Retrieved from https://www.infrastructureaustralia.gov.au/publications/2022-market-capacity-report

Bureau of Infrastructure and Transport Research Economics (BITRE). (2023a). Australian Infrastructure and Transport Statistics Yearbook 2023: Road. Retrieved from https://www.bitre.gov.au/publications/2023/australian-infrastructure-and-transport-statistics-yearbook-2023/road

Bureau of Infrastructure and Transport Research Economics (BITRE). (2023b). Australian Infrastructure and Transport Statistics Yearbook 2023: Road-related Revenue and Expenditure. Retrieved from https://www.bitre.gov.au/publications/2023/australian-infrastructure-and-transport-statistics-yearbook-2023/road-related-revenue-expenditure

Bureau of Infrastructure and Transport Research Economics (BITRE). (2023c). Australian Infrastructure and Transport Statistics Yearbook 2023: Infrastructure Construction. Retrieved from https://www.bitre.gov.au/publications/2023/australian-infrastructure-and-transport-statistics-yearbook-2023/infrastructure-construction

Student Contribution

All group members contributed equally to the creation and development of this project. Arka Dasgupta focused on the Economics Infrastructure Analysis and provided the R code (see Appendix A). Priya Ningappa analyzed the dataset related to Roads and supplied the corresponding R code (see Appendix B). Victorino Marquez served as the project lead, organizing meetings, providing updates, and conducting the analysis for Rail (see Appendix C).

Additionally, all members collaborated on the literature review, contributed feedback, and participated in drafting the conclusions. The entire group was involved in writing and presenting the final report.

Appendix A – R code_Economics Analysis

Appendix Code A – Economic Analysis

Libraries used: readr, knitr, dplyr, tidyr, readxl, zoo, tseries, ggplot2, forecast, and reshape2

```
file_path <- "C:/Users/ARKA/Sem 4/ARP/bitre infra.xlsx"

df <- read_excel(file_path, sheet = "Table 1.1", skip = 65)

colnames(df)

# Select the columns for Year, Road transport, and Rail transport
df_cleaned <- df %>%
  select(`Financial year`, `Road transport`, `Rail, pipeline and other transport (3)`) %>%
  filter(!is.na(`Financial year`))  # Remove rows without year data

# Rename columns
colnames(df_cleaned) <- c("Year", "Road_Investment", "Rail_Investment")

# Extract the first 4 digits from the financial year
df_cleaned$Year <- as.numeric(substr(df_cleaned$Year, 1, 4))

# Convert Investment columns to numeric
df_cleaned$Road_Investment <- as.numeric(df_cleaned$Road_Investment)
df_cleaned$Rail_Investment <- as.numeric(df_cleaned$Rail_Investment)

# Remove rows with NA values
df_cleaned <- df_cleaned %>%
  filter(!is.na(Year) & !is.na(Road_Investment) & !is.na(Rail_Investment))

# Sort data by year
df_cleaned <- df_cleaned %>%
  arrange(Year)

df_long <- df_cleaned %>%
  gather(key = "Sector", value = "Investment", -Year)

# Check for duplicate years
duplicates <- df_long %>%
  group_by(Year, Sector) %>%
  filter(n() > 1)

print(duplicates)

# Remove duplicates by keeping only the first occurrence of each Year and Sector combination
df_long_cleaned <- df_long %>%
  distinct(Year, Sector, .keep_all = TRUE)
print(df_long_cleaned)

# Plot the investment trends for Road and Rail over time
ggplot(df_long_cleaned, aes(x = Year, y = Investment, color = Sector, group = Sector)) +
  geom_line(size = 1.2) +  # Line plot for investment
  geom_smooth(method = "lm", se = FALSE, linetype = "dashed") +  # Add a trend line (linear regression)
  labs(title = "Investment Trend: Road vs Rail (Year-wise)",
       x = "Year",
       y = "Investment (in Million $)",
       color = "Sector") +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma) +  # Format y-axis with commas
  theme(legend.position = "bottom")

# Filter the data for the years 2019 to 2023
table_filtered <- subset(table_1_2a_d_cleaned, format(Collection_Month, "%Y") %in% c("2019", "2020", "2021", "2022", "2023"))
```

```
# Check for duplicates by year and Collection_Month
print(table_filtered)

table_filtered <- table_filtered[!duplicated(table_filtered$Collection_Month), ]


# Remove rows with missing data or extreme values
table_filtered <- table_filtered[complete.cases(table_filtered[, c("Road_Transport", "Rail_Transport", "Water_Transport",
"Air_Space_Transport")]), ]
table_filtered_long <- melt(table_filtered, id.vars = "Collection_Month",
                      measure.vars = c("Road_Transport", "Rail_Transport", "Water_Transport", "Air_Space_Transport"),
                      variable.name = "Transport_Sector", value.name = "Employment")

# Create the line plot
ggplot(table_filtered_long, aes(x = Collection_Month, y = Employment, color = Transport_Sector)) +
  geom_line(size = 1.2) +
  labs(title = "Sector-Wise Employment Trends (2019-2023)",
       x = "Year",
       y = "Employment (Thousands)",
       color = "Transport Sector") +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y") +  # Show yearly labels on x-axis
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Rotate x-axis labels for readability
table_1_5_cleaned <- table_1_5_cleaned %>%
  mutate(across(c(Road, Rail, Water, Air_Space, All_Industries), ~ na.approx(., na.rm = FALSE)))

# Convert the Year column to numeric
table_1_5_cleaned$Year <- as.numeric(table_1_5_cleaned$Year)

summary(table_1_5_cleaned)
table_1_5_long <- melt(table_1_5_cleaned, id.vars = "Year",
                  measure.vars = c("Road", "Rail", "Water", "Air_Space"),
                  variable.name = "Transport_Sector", value.name = "Earnings")

# Plot the earnings trend over time for all years
ggplot(table_1_5_long, aes(x = Year, y = Earnings, color = Transport_Sector)) +
  geom_line(size = 1.2) +
  labs(title = "Average Weekly Earnings Trend by Sector (1996-2021)",
       x = "Year",
       y = "Earnings",
       color = "Transport Sector") +
  theme_minimal()
rail_road_ts <- ts(transport_data_cleaned[, c("Rail_Freight", "Road_Freight")],
               start = min(transport_data_cleaned$Year),
               frequency = 1)

# Fit a VAR model for the time series data (Rail and Road Freight)
var_model <- VAR(rail_road_ts, p = 1, type = "both")

# Summary of the model
summary(var_model)

# Forecast both Rail and Road Freight for the next 5 periods
var_forecast <- predict(var_model, n.ahead = 5)

rail_forecast <- var_forecast$fcst$Rail_Freight[, "fcst"]
road_forecast <- var_forecast$fcst$Road_Freight[, "fcst"]
forecast_data <- data.frame(
  Year = seq(max(transport_data_cleaned$Year) + 1, by = 1, length.out = 5),
  Rail_Freight_Forecast = rail_forecast,
  Road_Freight_Forecast = road_forecast
)

# Plot the forecasted trends
```

```
ggplot() +
  geom_line(data = transport_data_cleaned, aes(x = Year, y = Rail_Freight, color = "Rail Freight (Historical)")) +
  geom_line(data = transport_data_cleaned, aes(x = Year, y = Road_Freight, color = "Road Freight (Historical)")) +
  geom_line(data = forecast_data, aes(x = Year, y = Rail_Freight_Forecast, color = "Rail Freight (Forecast)"), linetype = "dashed") +
  geom_line(data = forecast_data, aes(x = Year, y = Road_Freight_Forecast, color = "Road Freight (Forecast)"), linetype = "dashed") +
  labs(title = "Rail and Road Freight Forecast", x = "Year", y = "Producer Price Index (Forecast)") +
  scale_color_manual(name = "Freight Type", values = c("Rail Freight (Historical)" = "red", "Road Freight (Historical)" = "orange",
                                        "Rail Freight (Forecast)" = "red", "Road Freight (Forecast)" = "orange")) +
  theme_minimal()
```

Appendix B – R code_Road Analysis

<u>Appendix Code B – Road Analysis</u>

Libraries used: readr, knitr, dplyr, tidyr, readxl, ggplot2, forecast, and reshape2

```
sheet1 <- read_excel("clean2.xlsx", sheet = "Sheet1")
sheet2 <- read_excel("clean2.xlsx", sheet = "Sheet2")
sheet3_data <- read_excel("clean2.xlsx", sheet = "Sheet3")
sheet4 <- read_excel("clean2.xlsx", sheet = "Sheet4")
sheet5 <- read_excel("clean2.xlsx", sheet = "Sheet5")


merged_data <- merge(sheet1, sheet2, by = "Year", suffixes = c("_expenditure", "_VKT"))

head(merged_data)
```

1. State-Wise Road Spending Analysis

````
```{r}


selected_columns <- c("Year", "NSW_expenditure", "VIC_expenditure", "QLD_expenditure",
                     "SA_expenditure", "WA_expenditure", "TAS_expenditure",
                     "NT_expenditure", "ACT_expenditure", "NSW_VKT", "VIC_VKT",
                     "QLD_VKT", "SA_VKT", "WA_VKT", "TAS_VKT", "NT_VKT", "ACT_VKT")

filtered_data <- merged_data %>%
  select(all_of(selected_columns))


long_data <- filtered_data %>%
  pivot_longer(cols = starts_with(c("NSW", "VIC", "QLD", "SA", "WA", "TAS", "NT", "ACT")),
               names_to = c("State", ".value"), names_sep = "_")


correlations <- long_data %>%
  group_by(State) %>%
  summarize(correlation = cor(expenditure, VKT, use = "complete.obs"))

print(correlations)


ggplot(long_data, aes(x = expenditure, y = VKT)) +
  geom_point() +
  geom_smooth(method = "lm", col = "red") +
  facet_wrap(~ State, scales = "free") +
  labs(title = "Road Expenditure vs VKT by State",
       x = "Road Expenditure (millions AUD)",
       y = "Vehicle Kilometres Travelled (billions)") +
  theme_minimal() +
  geom_text(data = correlations, aes(label = paste("Correlation:", round(correlation, 2))),
            x = Inf, y = Inf, hjust = 1.1, vjust = 1.1, size = 3, inherit.aes = FALSE)


```
````

2. Trends in Kilometres Travelled by Vehicle Type

````
```{r}



sheet5$Year <- as.numeric(sub("-.*", "", sheet5$Year))


sheet5_long <- sheet5 %>%
  select(-Total) %>%
````
````

pivot_longer(cols = -Year, names_to = "Vehicle_Type", values_to = "Kilometers_Traveled")

ggplot(sheet5_long, aes(x = Year, y = Kilometers_Traveled, color = Vehicle_Type, group = Vehicle_Type)) +
  geom_line() +
  geom_point() +
  labs(title = " Vehicles Kilometers Traveled Over Time by Vehicle Type",
       x = "Year", y = "Kilometers Traveled") +
  theme_minimal() +
  scale_color_brewer(palette = "Set1") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

3. Expenditure on Roads and vehicle registration in AU regions

```{r}


merged_data3 <- merge(sheet1, sheet4, by = "Year", suffixes = c("_expenditure", "_registrations"))

merged_data3


data_long <- melt(merged_data3, id.vars = "Year",
                  measure.vars = c("NSW_expenditure", "VIC_expenditure", "QLD_expenditure", "SA_expenditure", "WA_expenditure",
                                   "TAS_expenditure", "NT_expenditure", "ACT_expenditure",
                                   "NSW_registrations", "VIC_registrations", "QLD_registrations", "SA_registrations",
                                   "WA_registrations", "TAS_registrations", "NT_registrations", "ACT_registrations"),
                  variable.name = "State_Type", value.name = "Value")


data_long$State <- sub("_.*", "", data_long$State_Type)
data_long$Type <- ifelse(grepl("expenditure", data_long$State_Type), "Expenditure", "Registrations")


data_long$Year <- factor(data_long$Year)


expenditure_data <- filter(data_long, Type == "Expenditure")
registration_data <- filter(data_long, Type == "Registrations")


data_normalized <- data_long %>%
  group_by(Type) %>%
  mutate(Normalized_Value = (Value - min(Value)) / (max(Value) - min(Value)))

```
```{r}


data_normalized <- data_long %>%
  group_by(Type) %>%
  mutate(Normalized_Value = (Value - min(Value)) / (max(Value) - min(Value)))


data_pct_change <- data_long %>%
  group_by(State, Type) %>%
  arrange(Year) %>%
  mutate(Pct_Change = (Value - lag(Value)) / lag(Value) * 100)


data_pct_change <- na.omit(data_pct_change)


ggplot(data_pct_change, aes(x = Year, y = Pct_Change, color = Type, group = interaction(State, Type))) +

```
  geom_line(size = 1) +
  facet_wrap(~ State, scales = "free_y") +
  theme_minimal() +
  labs(title = "Year-over-Year Percentage Change in Expenditures and Registrations by State",
       x = "Year", y = "Percentage Change (%)", color = "Type") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


```

4. 5-Year Forecast of Vehicle Usage (ARIMA Model)

```{r}

sheet4_clean <- sheet4 %>%
  mutate(Total = as.numeric(gsub(" ", "", Total))) %>%
  filter(!is.na(Total))

sheet4_clean

sheet4_clean$Year <- as.numeric(sub("-.*", "", sheet4_clean$Year))

vehicle_ts <- ts(sheet4_clean$Total, start = min(sheet4_clean$Year), frequency = 1)

plot(vehicle_ts, ylab = "Registrations", xlab = "Year", type = "o", main = "Time Series Plot of Vehicle Registrations")

```

```{r}


sheet2$Year <- as.numeric(sub("-.*", "", sheet2$Year))

forecast_kilometers_by_state <- function(state_column, state_name) {

  state_ts <- ts(sheet2[[state_column]], start = min(sheet2$Year), frequency = 1)

  fit_arima <- auto.arima(state_ts)

  forecast_arima <- forecast(fit_arima, h = 5)

  data.frame(Year = as.numeric(time(forecast_arima$mean)),
             Kilometers = as.numeric(forecast_arima$mean),
             State = state_name)
}


states <- list(
  forecast_kilometers_by_state("NSW", "NSW"),
  forecast_kilometers_by_state("VIC", "VIC"),
  forecast_kilometers_by_state("QLD", "QLD"),
  forecast_kilometers_by_state("SA", "SA"),
  forecast_kilometers_by_state("WA", "WA"),
  forecast_kilometers_by_state("TAS", "TAS"),
  forecast_kilometers_by_state("NT", "NT"),
  forecast_kilometers_by_state("ACT", "ACT")
)


forecast_data <- do.call(rbind, states)


ggplot(forecast_data, aes(x = Year, y = Kilometers, color = State)) +
  geom_line() +
  geom_point() +
  geom_text(aes(label = round(Kilometers, 2)), vjust = -1, size = 3) +
  facet_wrap(~ State, scales = "free_y") +
  labs(title = "5-Year Forecast of Vehicle Kilometers Traveled by State (ARIMA Model)",
       x = "Year", y = "Kilometers Traveled") +
  theme_minimal()
```

````
```
````

6. Transport Infrastructure Engineering Work

````
```{r}

private_sector_private <- read_excel("clean2.xlsx", sheet = "Priv-Priv")
private_sector_public <- read_excel("clean2.xlsx", sheet = "Priv-Pub")
public_sector <- read_excel("clean2.xlsx", sheet = "Public")

head(private_sector_private)
head(private_sector_public)
head(public_sector)

combined_data <- bind_rows(
  private_sector_private %>% mutate(Sector = "Private Sector (Private)"),
  private_sector_public %>% mutate(Sector = "Private Sector (Public)"),
  public_sector %>% mutate(Sector = "Public Sector")
)

combined_data <- combined_data %>%
  mutate(Year = factor(Year, levels = unique(Year)))

str(combined_data)


```

```{r}
# Plot without bar labels for better readability
ggplot(combined_data, aes(x = Year, y = `Transport infrastructure engineering construction work done`, fill = Sector)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Transport Infrastructure Engineering Construction Work (1986-2023)",
       x = "Year",
       y = "Construction Work Done",
       fill = "Sector") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        text = element_text(size = 12)) +
  scale_y_continuous(labels = scales::comma)
```
````
````

Appendix C – R code_Rail Analysis

```
# Load the required libraries
library(ggplot2)
library(dplyr)
library(tidyr)
library(ggplot2)
library(dplyr)
library(tidyr)
library(readxl)
library(readxl)
library(dplyr)

# Reading the CSV file with the header located at row 11

# Then read the file
file_path <- "dataset4.xlsx"
Workdone <- read_excel(file_path, sheet = "Table-2.2")
patronage <- read_excel(file_path, sheet = "Table-5.5a")
rail_distances <- read_excel(file_path, sheet = "Table-7.1-rail-distances")
non_bulk_rail_freight <- read_excel(file_path, sheet = "Table-7.4-non-bulk")
rail_expenditure <- read_excel(file_path, sheet = "Table-7.6-expenditure")
patronage_heavy <- read_excel(file_path, sheet = "Table7.5HeavyRail")
patronage_light <- read_excel(file_path, sheet = "Table7.6LightRail")

head(Workdone)
head (patronage)
head(non_bulk_rail_freight)
head(rail_expenditure)
head(patronage_heavy)
head(patronage_light)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
Workdone_long <- Workdone %>%
  select(`Financial Year`, Transport, Energy, Telecommunications, Water) %>%  # Select relevant columns
  gather(key = "Type", value = "Value", -`Financial Year`)  # Gather columns into 'Type' and 'Value'

# Print the reshaped data to verify
print(Workdone_long)
```

Construcion Done

```
# Ensure that the Financial Year column is treated as a factor
Workdone_long$`Financial Year` <- factor(Workdone_long$`Financial Year`, levels = unique(Workdone_long$`F
inancial Year`))

# Plot the data again
ggplot(Workdone_long, aes(x = `Financial Year`, y = Value, color = Type, group = Type)) +
  geom_line(linewidth = 1) +  # Use `linewidth` for line thickness
  geom_smooth(method = "lm", se = FALSE, linetype = "dashed") +  # Add a linear trend line without confid
ence intervals
  labs(title = "Trend of Infrastructure Spending by Type Over Financial Years",
       x = "Financial Year", y = "Spending Value (in millions)",
       color = "Type of Infrastructure") +
  theme_minimal(base_size = 15) +  # Increase overall base font size
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 12),  # Adjust x-axis label font size and an
gle
    axis.title.x = element_text(size = 14),  # Adjust x-axis title font size
    axis.title.y = element_text(size = 14),  # Adjust y-axis title font size
    legend.title = element_text(size = 13),  # Adjust legend title font size
    legend.text = element_text(size = 12)  # Adjust legend text font size
  ) +
```

```r
  scale_x_discrete(drop = FALSE) +  # Ensure all levels of the factor are displayed on the x-axis
  scale_y_continuous(labels = scales::comma)  # Format y-axis for readability
```

```r
# Print the reshaped data to verify
print(patronage)
str(patronage)
```

Patronage

```r
# Assuming the necessary libraries are already loaded

# Filter the patronage dataset based on Capital City = "CapitalCities"
patronage_filtered <- patronage %>%
  filter(`Capital City` == "CapitalCities") %>%
  select(`Financial Year`, `Heavy rail`, `Light rail`, `UPT Bus`, `Ferry`)

# Reshape the data to long format for plotting multiple transport types
patronage_long <- patronage_filtered %>%
  gather(key = "Type", value = "Value", -`Financial Year`)

# Convert Value column to numeric to avoid scale_y_continuous() error
patronage_long$Value <- as.numeric(patronage_long$Value)

# Scale values to millions for cleaner y-axis
patronage_long$Value <- patronage_long$Value / 1e6

# Convert the `Financial Year` to a factor to retain the original format
patronage_long$`Financial Year` <- factor(patronage_long$`Financial Year`, levels = unique(patronage_long
$`Financial Year`))

# Plot the data for Heavy rail, Light rail, UPT bus, and Ferry over Financial Years
ggplot(patronage_long, aes(x = `Financial Year`, y = Value, color = Type, group = Type)) +
  geom_line(linewidth = 1) +  # Use `linewidth` instead of `size` for line thickness
  labs(title = "Public Transport Patronage Over Financial Years",
       x = "Financial Year", y = "Patronage (in millions)",
       color = "Type of Transport") +
  theme_minimal(base_size = 15) +  # Set a base font size for the plot
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 12),  # Rotate x-axis labels and adjust font
size
    axis.title.x = element_text(size = 14),  # Increase font size of x-axis title
    axis.title.y = element_text(size = 14),  # Increase font size of y-axis title
    axis.text.y = element_text(size = 10),   # Adjust y-axis text font size
    legend.title = element_text(size = 13),  # Increase legend title font size
    legend.text = element_text(size = 12)    # Increase legend text font size
  ) +
  scale_x_discrete(drop = FALSE) +  # Ensure all levels of the Financial Year are shown on x-axis
  scale_y_continuous(labels = scales::comma_format(scale = 1e6, suffix = "M"))  # Format y-axis for reada
bility and scale to millions
```

Rail Expenditure over the time

```r
# Rail Expenditure over the time

# Reshape the rail_expenditure dataset to long format for plotting
rail_expenditure_long <- rail_expenditure %>%
  select(`Financial Year`, NSW, VIC, QLD, SA, WA, TAS, NT, ACT) %>%  # Select relevant columns
  gather(key = "State", value = "Expenditure", -`Financial Year`)  # Convert to long format

# Convert `Financial Year` to a factor to retain the original format (YYYY-YY)
rail_expenditure_long$`Financial Year` <- factor(rail_expenditure_long$`Financial Year`, levels = unique(
rail_expenditure_long$`Financial Year`))

# Convert Expenditure column to numeric
rail_expenditure_long$Expenditure <- as.numeric(rail_expenditure_long$Expenditure)
```

```r
# Plot the rail expenditure data for each state over Financial Years
ggplot(rail_expenditure_long, aes(x = `Financial Year`, y = Expenditure, color = State, group = State)) +
  geom_line(linewidth = 1) +  # Use `linewidth` instead of `size` for line thickness
  labs(title = "Rail Expenditure by State Over Financial Years",
       x = "Financial Year", y = "Expenditure (in millions)",
       color = "State") +
  theme_minimal(base_size = 15) +  # Set a base font size for the plot
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 12),  # Rotate x-axis labels for readability
    axis.title.x = element_text(size = 14),  # Adjust font size of x-axis title
    axis.title.y = element_text(size = 14),  # Adjust font size of y-axis title
    legend.title = element_text(size = 13),  # Adjust font size of legend title
    legend.text = element_text(size = 12)    # Adjust font size of legend text
  ) +
  scale_x_discrete(drop = FALSE) +  # Ensure all levels of the Financial Year are shown on x-axis
  scale_y_continuous(labels = scales::comma)  # Format y-axis for readability
```

NSW consistently leads in rail investment, with a significant gap compared to the other states. QLD and VIC receive moderate levels of investment, while ACT, NT, and TAS see very little expenditure in comparison.

```r
print(rail_expenditure)

print(rail_expenditure_long)
```

Non-Bulk Freight

```r
# Load necessary libraries
library(tidyverse)
library(readxl)

# Read the dataset from Excel (assuming the file is named 'non_bulk_rail_freight.xlsx')
# Replace 'file_path' with the actual path to your Excel file
# non_bulk_rail_freight <- read_excel("path_to_your_file.xlsx")

# Assuming non_bulk_rail_freight is already loaded as a data frame, let's modify the ACT column
# Replace 'NA' values in the 'ACT' column with 0
non_bulk_rail_freight$ACT[is.na(non_bulk_rail_freight$ACT)] <- 0

# Reshape the data into a long format for plotting
non_bulk_rail_freight_long <- non_bulk_rail_freight %>%
  select(`Financial Year`, NSW, VIC, QLD, SA, WA, NT, ACT) %>%  # Select relevant columns
  gather(key = "State", value = "Freight", -`Financial Year`)  # Convert to long format

# Convert `Financial Year` to a factor to retain the original format (YYYY-YY)
non_bulk_rail_freight_long$`Financial Year` <- factor(non_bulk_rail_freight_long$`Financial Year`, levels
= unique(non_bulk_rail_freight_long$`Financial Year`))

# Convert Freight column to numeric
non_bulk_rail_freight_long$Freight <- as.numeric(gsub(",", "", non_bulk_rail_freight_long$Freight))  # Re
move commas and convert to numeric

# Plot the trend line for rail freight data over Financial Years by state
ggplot(non_bulk_rail_freight_long, aes(x = `Financial Year`, y = Freight, color = State, group = State))
+
  geom_line(linewidth = 1.2) +  # Use `linewidth` for line thickness
  geom_point(size = 2) +  # Add data points to highlight observations
  labs(title = "Interstate Non-Bulk Rail Freight by State Over Financial Years",
       x = "Financial Year", y = "Freight (Million Tonne-Kilometres)",
       color = "State") +
  theme_minimal(base_size = 15) +  # Set a base font size for the plot
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 12),  # Rotate x-axis labels for readability
    axis.title.x = element_text(size = 14),  # Adjust font size of x-axis title
    axis.title.y = element_text(size = 14),  # Adjust font size of y-axis title
```

```
    legend.title = element_text(size = 13),  # Adjust font size of legend title
    legend.text = element_text(size = 12)    # Adjust font size of legend text
  ) +
  scale_x_discrete(drop = FALSE) +  # Ensure all levels of the Financial Year are shown on x-axis
  scale_y_continuous(labels = scales::comma_format(), limits = c(0, NA)) +  # Ensure y-axis starts at 0 a
nd format labels
  scale_color_brewer(palette = "Set2")  # Use a color palette for better distinction
```

Patronage in Heavy and Light Rail (Exclusing Metropolitan)

```
# Load necessary libraries
library(readxl)
library(tidyverse)

# Read the datasets from the Excel file (replace 'file_path' with the actual file path)
# patronage_heavy <- read_excel("path_to_your_excel_file.xlsx", sheet = "Table7.5HeavyRail")
# patronage_light <- read_excel("path_to_your_excel_file.xlsx", sheet = "Table7.6LightRail")

# Convert financial Year to a factor to maintain the original format
patronage_heavy$`Financial Year` <- factor(patronage_heavy$`Financial Year`, levels = unique(patronage_he
avy$`Financial Year`))
patronage_light$`Financial Year` <- factor(patronage_light$`Financial Year`, levels = unique(patronage_li
ght$`Financial Year`))

# Reshape heavy rail data to long format for plotting and exclude 'Metropolitan'
patronage_heavy_long <- patronage_heavy %>%
  gather(key = "City", value = "Patronage", -`Financial Year`) %>%
  filter(!is.na(Patronage)) %>%  # Remove missing values
  filter(City != "Metropolitan")  # Exclude 'Metropolitan'

# Reshape light rail data to long format for plotting and exclude 'Metropolitan'
patronage_light_long <- patronage_light %>%
  gather(key = "City", value = "Patronage", -`Financial Year`) %>%
  filter(!is.na(Patronage)) %>%
  filter(City != "Metropolitan")  # Exclude 'Metropolitan'

# Plot for Heavy Rail Patronage
ggplot(patronage_heavy_long, aes(x = `Financial Year`, y = Patronage, color = City, group = City)) +  # A
dd `group = City`
  geom_line(size = 1.2) +
  labs(title = "Public Transit Patronage - Heavy Rail",
       x = "Financial Year", y = "Patronage (in millions)",
       color = "City") +
  theme_minimal(base_size = 15) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(drop = FALSE)  # Use scale_x_discrete to maintain categorical x-axis
```

```
# Plot for Light Rail Patronage
ggplot(patronage_light_long, aes(x = `Financial Year`, y = Patronage, color = City, group = City)) +  # A
dd `group = City`
  geom_line(size = 1.2) +
  labs(title = "Public Transit Patronage - Light Rail",
       x = "Financial Year", y = "Patronage (in millions)",
       color = "City") +
  theme_minimal(base_size = 15) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(drop = FALSE)  # Use scale_x_discrete to maintain categorical x-axis
```

```
# Combine both Heavy Rail and Light Rail in one plot for comparison
# Adding a column to indicate if the data is from Heavy Rail or Light Rail
patronage_heavy_long <- patronage_heavy_long %>% mutate(Type = "Heavy Rail")
patronage_light_long <- patronage_light_long %>% mutate(Type = "Light Rail")

# Combine the datasets, ensuring 'Metropolitan' is excluded
patronage_combined <- bind_rows(patronage_heavy_long, patronage_light_long)
```

```r
# Plot combined Heavy Rail and Light Rail excluding 'Metropolitan'
ggplot(patronage_combined, aes(x = `Financial Year`, y = Patronage, color = City, group = interaction(City, Type), linetype = Type)) +  # Add `group = interaction(City, Type)`
  geom_line(size = 1.2) +
  labs(title = "Public Transit Patronage - Heavy Rail vs Light Rail",
       x = "Financial Year", y = "Patronage (in millions)",
       color = "City", linetype = "Rail Type") +
  theme_minimal(base_size = 15) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(drop = FALSE)  # Use scale_x_discrete to maintain categorical x-axis
```

Patronage in Heavy and Light Rail (Exclusing Metropolitan)

```r
# Load necessary libraries
library(readxl)
library(tidyverse)
library(openxlsx)

# Assuming 'patronage_light' and 'patronage_heavy' are already loaded as data frames
# Uncomment and modify the file path if reading from an Excel file:
# patronage_light <- read_excel("path_to_your_file.xlsx", sheet = "LightRailSheetName")
# patronage_heavy <- read_excel("path_to_your_file.xlsx", sheet = "HeavyRailSheetName")

# Select only the relevant columns (NSW, VIC, QLD, SA, WA, NT, ACT) and 'Financial Year'
patronage_light_selected <- patronage_light %>%
  select(`Financial Year`, NSW, VIC, QLD, SA, WA, NT, ACT)

patronage_heavy_selected <- patronage_heavy %>%
  select(`Financial Year`, NSW, VIC, QLD, SA, WA, NT, ACT)

# Sum corresponding state values for each 'Financial Year'
combined_patronage_sum <- patronage_light_selected %>%
  full_join(patronage_heavy_selected, by = "Financial Year", suffix = c("_Light", "_Heavy")) %>%
  mutate(
    NSW = NSW_Light + NSW_Heavy,
    VIC = VIC_Light + VIC_Heavy,
    QLD = QLD_Light + QLD_Heavy,
    SA = SA_Light + SA_Heavy,
    WA = WA_Light + WA_Heavy,
    NT = NT_Light + NT_Heavy,
    ACT = ACT_Light + ACT_Heavy
  ) %>%
  select(`Financial Year`, NSW, VIC, QLD, SA, WA, NT, ACT)  # Keep only the combined columns

# Write the resulting data frame to a new Excel file
write.xlsx(combined_patronage_sum, "combined_patronage_sum.xlsx", sheetName = "Patronage_Summary", overwrite = TRUE)

# Print the resulting data frame to view
print(combined_patronage_sum)
```

Convert to long format

```r
# Load necessary libraries
library(readxl)
library(tidyverse)
library(openxlsx)

# Assuming 'combined_patronage_sum' is already loaded as a data frame
# Uncomment and modify the file path if reading from an Excel file:
# combined_patronage_sum <- read_excel("path_to_your_file.xlsx", sheet = "SheetName")

# Convert wide format to long format using pivot_longer
patronage_long <- combined_patronage_sum %>%
  pivot_longer(cols = NSW:ACT,  # Select columns to reshape
               names_to = "State",  # Create a 'State' column
```

```r
                values_to = "Patronage")  # Create a 'Patronage' column

# View the resulting long format data
print(patronage_long)

# Write the resulting long format data to a new Excel file
#write.xlsx(patronage_long, "patronage_long_format.xlsx", sheetName = "LongFormatPatronage", overwrite =
TRUE)
```

Merge Rail Expenditure and Rail Patronage

```r
# Load necessary libraries
library(readxl)
library(tidyverse)
library(openxlsx)

# Assuming rail_expenditure_long and patronage_long are already loaded as data frames

# Convert 'Financial Year' columns to factors in both datasets to ensure consistency
rail_expenditure_long$`Financial Year` <- factor(rail_expenditure_long$`Financial Year`, levels = unique(
rail_expenditure_long$`Financial Year`))
patronage_long$`Financial Year` <- factor(patronage_long$`Financial Year`, levels = unique(patronage_long
$`Financial Year`))

# Merge rail_expenditure_long and patronage_long by 'Financial Year' and 'State'
merged_data <- rail_expenditure_long %>%
  inner_join(patronage_long, by = c("Financial Year", "State"))

# View the resulting merged dataset
print(merged_data)

# Write the resulting merged data frame to a new Excel file
#write.xlsx(merged_data, "merged_rail_expenditure_patronage.xlsx", sheetName = "MergedData", overwrite =
TRUE)
```

Correlation Expenditure vs Patronage

```r
# Load necessary libraries
library(readxl)
library(tidyverse)
library(openxlsx)

# Assuming rail_expenditure_long and patronage_long are already loaded as data frames

# Convert 'Financial Year' columns to factors in both datasets to ensure consistency
rail_expenditure_long$`Financial Year` <- factor(rail_expenditure_long$`Financial Year`, levels = unique(
rail_expenditure_long$`Financial Year`))
patronage_long$`Financial Year` <- factor(patronage_long$`Financial Year`, levels = unique(patronage_long
$`Financial Year`))

# Merge rail_expenditure_long and patronage_long by 'Financial Year' and 'State'
merged_data <- rail_expenditure_long %>%
  inner_join(patronage_long, by = c("Financial Year", "State"))

# Filter merged data to include only the specified states
states_to_include <- c("NSW", "QLD", "SA", "VIC")
filtered_data <- merged_data %>%
  filter(State %in% states_to_include)

# Calculate correlation for each of the specified states and store in a new data frame
correlations <- filtered_data %>%
  group_by(State) %>%
  summarize(
    correlation = tryCatch(
      cor(Expenditure, Patronage, use = "complete.obs"),
      error = function(e) NA  # Return NA if there's an error in calculating correlation
    )
  )
```

```r
# Create an enhanced scatter plot for Expenditure vs Patronage with improvements
ggplot(filtered_data, aes(x = Expenditure, y = Patronage)) +
  geom_point(alpha = 0.6, color = "steelblue", size = 3) +  # Change point color and size
  geom_smooth(method = "lm", col = "darkred", se = FALSE, linetype = "solid", size = 1) +  # Change line
style and size
  facet_wrap(~ State, scales = "free") +  # Create separate plots for each state
  labs(title = "Rail Expenditure vs Patronage by State (NSW, QLD, SA, VIC)",
       x = "Rail Expenditure (in millions AUD)",
       y = "Patronage (in millions)") +
  theme_minimal(base_size = 14) +  # Set a minimal theme with larger base font size
  theme(
    strip.text = element_text(size = 15, face = "bold"),  # Adjust facet titles
    plot.title = element_text(size = 18, face = "bold", hjust = 0.5),  # Center align and adjust title si
ze
    axis.title = element_text(size = 14, face = "bold"),  # Adjust axis titles
    axis.text = element_text(size = 12),  # Adjust axis text
    panel.grid.major = element_line(color = "grey80", size = 0.5, linetype = "dotted"),  # Adjust major g
ridlines
    panel.grid.minor = element_blank(),  # Remove minor gridlines
    panel.background = element_rect(fill = "white", color = "grey50")  # Set panel background
  ) +
  geom_text(data = correlations, aes(label = paste("Correlation:", round(correlation, 2))),
            x = Inf, y = Inf, hjust = 1.2, vjust = 1.2, size = 5, inherit.aes = FALSE, color = "darkred",
fontface = "bold") +  # Adjust correlation label
  scale_color_manual(values = c("NSW" = "darkgreen", "QLD" = "purple", "SA" = "orange", "VIC" = "blue"))
# Customize colors if needed
```

```r
library(readxl)
interaction<- read_excel("dataset5.xlsx")

str(interaction)

names(interaction) <- c("Financial_Year", "Route", "Movements", "Distance_km", "Growth_Percent",
                        "Pop_Sydney", "Pop_Melbourne", "Pop_Canberra", "Pop_Brisbane", "Pop_Adelaide",
                        "Pop_Product", "Inv_Distance", "Population_Ratio", "Interaction")
```

##Correlation Analysis

```r
# Correlation between Movements and Pop_Product
cor(interaction$Movements, interaction$Pop_Product, use = "complete.obs")

# Correlation between Movements and Interaction
cor(interaction$Movements, interaction$Interaction, use = "complete.obs")

library(ggplot2)



# Create a scatter plot of Interaction vs. Movements colored by Route
ggplot(interaction, aes(x = Movements, y = Interaction, color = Route, shape = Route)) +
  geom_point(size = 3, alpha = 0.8) +
  scale_y_continuous(labels = scales::comma) +  # Removed dollar formatting
  labs(
    title = "Scatterplot of Interaction vs Movements (thousands)",
    x = "Movements (thousands)",
    y = "Interaction"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5)
  )
```

```r
library(ggplot2)
library(scales)
library(dplyr)
```

```r
# Filter the dataset to include only Financial Year 2022-23
interaction_filtered <- interaction %>%
  filter(Financial_Year == "2022-23")

# Create a bar chart of Interaction by Route with y-axis in Millions
ggplot(interaction_filtered, aes(x = Route, y = Interaction / 1e6)) +
  geom_bar(stat = "identity", fill = "skyblue", color = "black") +
  scale_y_continuous(labels = label_number(suffix = "M", scale = 1)) +
  labs(
    title = "Chart of Interaction (2022-2023)",
    x = "Route",
    y = "Interaction (Millions)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5)
  )
```

```r
ggplot(interaction, aes(x = Financial_Year, y = Interaction, color = Route, group = Route)) +
  geom_line(size = 1) +
  scale_y_continuous(labels = scales::comma) +  # Use comma for better readability or remove for plain nu
mbers
  labs(
    title = "Line Plot of Interaction",
    x = "Financial Year",
    y = "Interaction"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
```

Trend Analysis: Movements Over time

```r
library(ggplot2)

# Create a line plot of Movements over time, grouped by Route
ggplot(interaction, aes(x = Financial_Year, y = Movements, color = Route, group = Route)) +
  geom_line(size = 1) +
  labs(
    title = "Trend of Movements Over Time by Route",
    x = "Financial Year",
    y = "Movements (thousands)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1, size = 8)  # Adjusting x-axis labels
  ) +
  scale_x_discrete(breaks = function(x) x[seq(1, length(x), by = 2)])  # Display every 2nd label for clar
ity
```

##Growth Comparision Rate

```r
library(ggplot2)

# Create a bar plot to show year-over-year growth percentage by route and year
ggplot(interaction, aes(x = Financial_Year, y = Growth_Percent * 100, fill = Route)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Year-over-Year Growth Percentage by Route",
```

```r
    x = "Financial Year",
    y = "Growth Percentage (%)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5, size = 7)  # Rotating and resizing x-a
xis labels
  ) +
  scale_x_discrete(breaks = interaction$Financial_Year)  # Display all financial years
```

```r
library(ggplot2)

ggplot(interaction, aes(x = Financial_Year, y = Growth_Percent * 100, color = Route, group = Route)) +
  geom_line(size = 1) +
  facet_wrap(~ Route, scales = "free_y") +
  labs(
    title = "Growth Rate by Route Over Time",
    x = "Financial Year",
    y = "Growth Percentage (%)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
```

##Correlatio between Movements and Populations

```r
library(ggplot2)
ggplot(interaction, aes(x = Pop_Product, y = Movements, color = Route)) +
  geom_point(alpha = 0.7) +
  labs(
    title = "Correlation Between Population Product (City1 * City2) and Movements",
    x = "Population Product (City1 * City2)",
    y = "Movements (thousands)"
  ) +
  theme_minimal()
```

```r
library(dplyr)
library(ggplot2)

# Filter the data for the financial year 2022-23
interaction_filtered <- interaction %>%
  filter(Financial_Year == "2022-23")

# Calculate efficiency based on the filtered data
interaction_filtered$Efficiency <- interaction_filtered$Movements / interaction_filtered$Distance_km

# Plot the data
ggplot(interaction_filtered, aes(x = Route, y = Efficiency)) +
  geom_bar(stat = "identity", fill = "skyblue", color = "black") +
  labs(
    title = "Route Efficiency (Movements per Kilometer for 2022-23)",
    x = "Route",
    y = "Efficiency (Movements per km)"
  ) +
  theme_minimal()
```

###Population Growth Impact

```r
# Install and load necessary packages if not already installed
# install.packages("plotly")
library(plotly)
library(dplyr)

# Assuming your full dataset is stored in a data frame named 'interaction'
# Filter the data for the Financial Year 2022-2023
data_2022_2023 <- interaction %>%
  filter(Financial_Year == "2022-23")

# Define a custom color palette for the routes
custom_colors <- c(
  "Mel-Adl" = "#1f77b4",  # Blue
  "Mel-Bne" = "#ff7f0e",  # Orange
  "Syd-Adl" = "#2ca02c",  # Green
  "Syd-Bne" = "#d62728",  # Red
  "Syd-Cbr" = "#9467bd",  # Purple
  "Syd-Mel" = "#8c564b"   # Brown
)

# Create the 3D scatter plot with custom colors
plot <- plot_ly(
  data_2022_2023,
  x = ~Interaction,
  y = ~Movements,
  z = ~Pop_Product,
  color = ~Route,
  colors = custom_colors,
  type = "scatter3d",
  mode = "markers+text",
  text = ~Route,
  marker = list(size = 5)
) %>%
  layout(
    title = "3D Chart of Interaction, Movements, and Pop_Product by Route (2022-2023)",
    scene = list(
      xaxis = list(title = "Interaction"),
      yaxis = list(title = "Movements (Thousands)"),
      zaxis = list(title = "City1*City2 (Trillions)")
    )
  )

# Display the plot
plot
```