Registration number: 250125052

Word count: 3,000

15th January, 2026

Topic: COMPOSITE VISUALISATION OF ANALYZING LONG-TERM TRENDS IN BROKEN AND BUCKLED RAIL IN THE UK MAINLINE RAIL NETWORK.

SECTION 1

KNOWLEDGE BUILDING

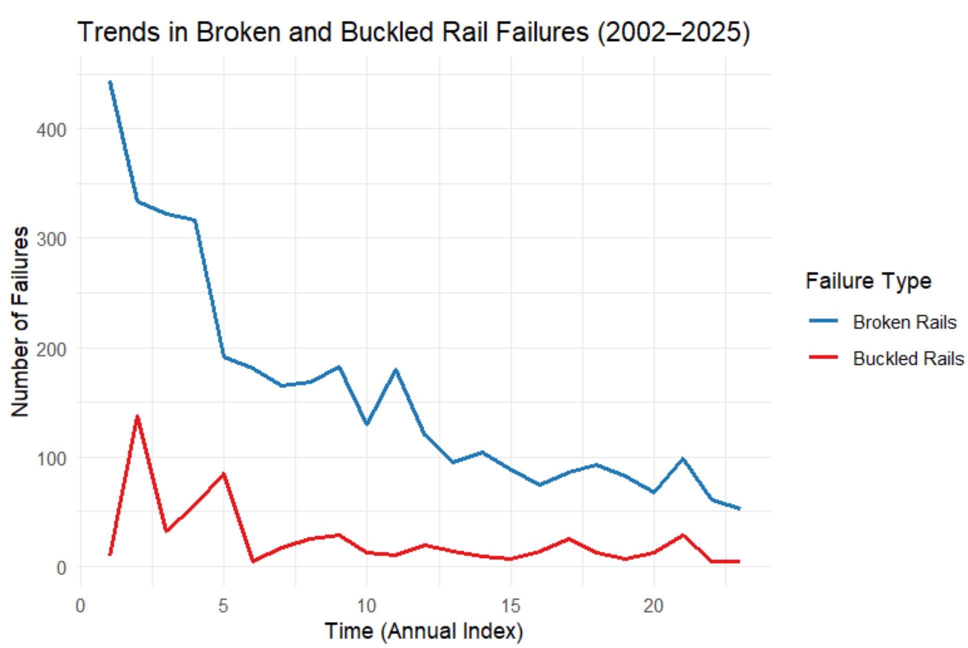
1.1 Topic and Importance

This project examines long-term trends and patterns in mainline track failures, specifically focusing on broken and buckled rails within the UK rail network. These types of failures serve as critical indicators of the condition of rail infrastructure and are closely linked to safety risks and network reliability. Undetected broken rails can result in derailments, while buckled rails, often caused by thermal stress, create significant operational hazards during high-temperature periods.

The importance of this topic is underscored in the UK context, where the rail network is extensively used and faces challenges from ageing infrastructure, climate variability, and rising performance expectations. Prior research indicates that systematic monitoring of track failures offers valuable insights into the effectiveness of maintenance strategies and long-term infrastructure renewal programs implemented by Network Rail (RSSB, 2023; Network Rail, 2024).

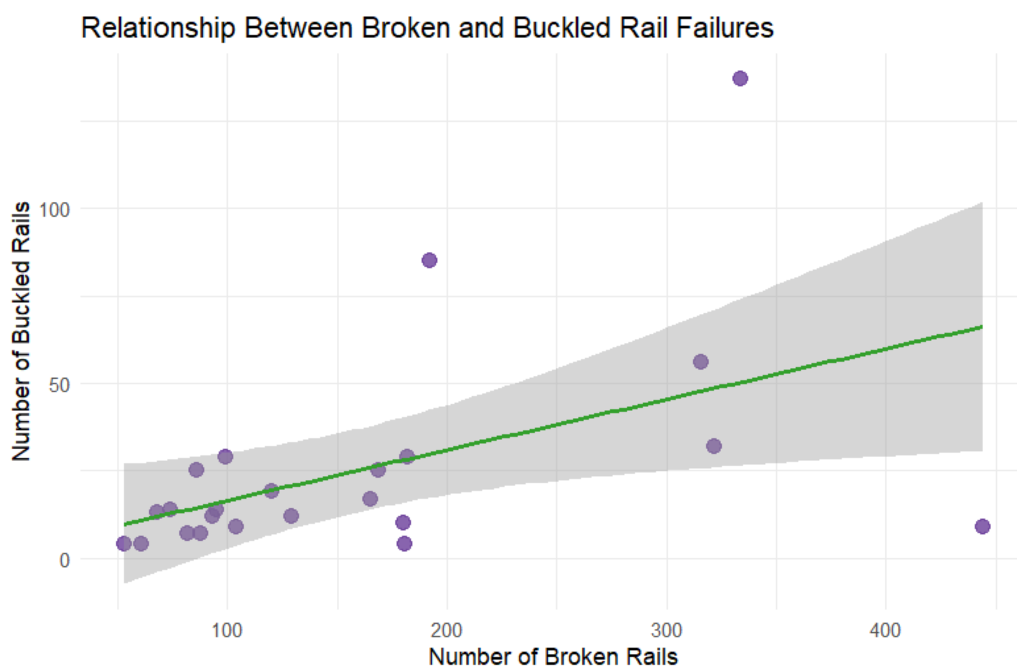
1.2 Explanation of individual visualization

- Time series plot



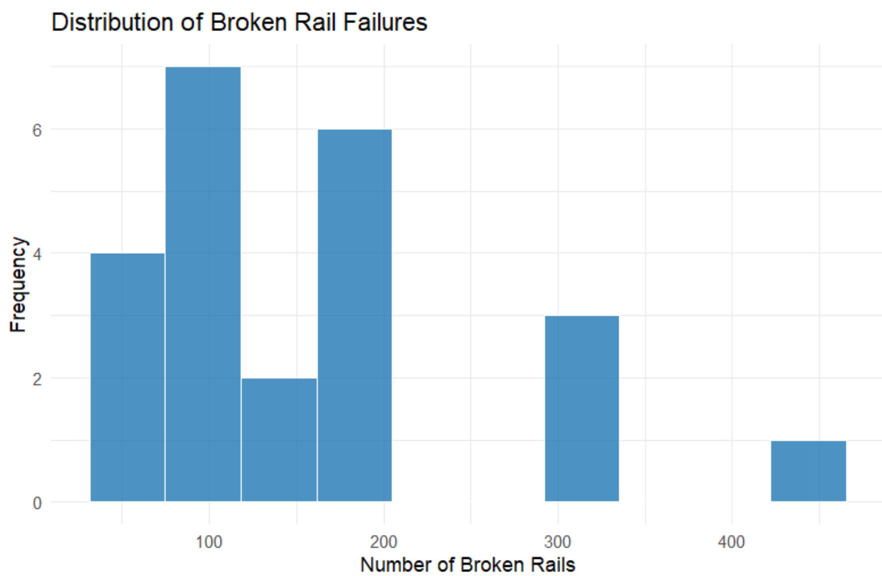
The time series visualization illustrates a distinct long-term decline in both broken and buckled rail failures from the early 2000s through 2024–2025. Broken rail failures demonstrate a significant reduction, particularly after the mid-2000s, while buckled rail failures display greater year-to-year variability yet maintain an overall downward trend. This pattern indicates sustained enhancements in inspection practices, material quality, and maintenance planning over the years (Network Rail, 2024).

- Scatter plot



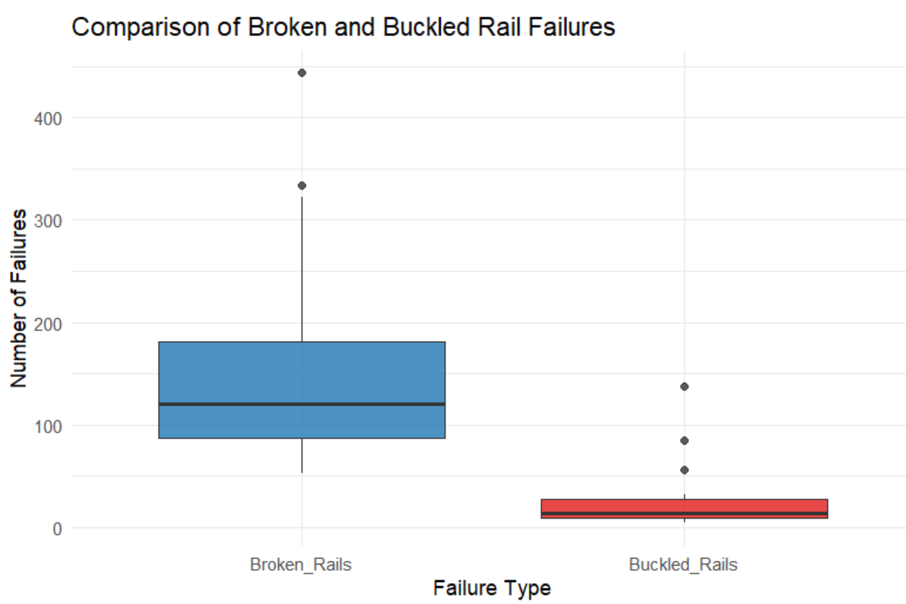
The scatter plot investigates the correlation between broken and buckled rail failures. Although the relationship is not strictly linear, it demonstrates a positive trend, indicating that years with a higher incidence of broken rails frequently correspond to increased counts of buckled rails. This observation aligns with existing engineering literature, which highlights that factors such as track condition, environmental exposure, and the effectiveness of maintenance can simultaneously impact various failure mechanisms (Esveld, 2001).

- Histogram



The histogram illustrating broken rail failures displays a right-skewed distribution, characterized by noticeable gaps at higher failure ranges. This pattern suggests that extreme values were predominantly recorded in earlier years, whereas more recent observations tend to cluster around lower counts. Such skewness is a common feature of infrastructure failure data and underscores the non-normal characteristics of rail defect occurrences (Cameron & Trivedi, 2013).

- Boxplot



The boxplot comparison reveals significant differences in both central tendency and dispersion between broken and buckled rails. Broken rails demonstrate a higher median and greater variability, accompanied by notable outliers, while buckled rails exhibit a more compact distribution. These differences underscore distinct failure mechanisms and emphasize the importance of treating these two failure types separately in analysis and modeling (RSSB, 2023).

1.3 Composite insight from the visualization set

When considered together, the four visualizations offer new, integrated insights that would not be evident from any single chart alone. The composite view illustrates that:

- UK rail track failures have shown a consistent long-term decline, rather than exhibiting random fluctuations.
- Although broken rails and buckled rails are related, they display differing distribution characteristics and levels of variability.
- The presence of skewness, outliers, and temporal clustering indicates that rail failures are influenced by structural and systemic factors rather than short-term noise.

This comprehensive perspective reinforces the notion that sustained investment in infrastructure and risk-based maintenance have produced tangible safety improvements across the UK rail network, aligning with findings in the fields of transport safety and rail engineering literature (Esveld, 2001; Network Rail, 2024).

The findings from the visual analysis strongly corroborate existing research that demonstrates how proactive inspection technologies, continuous welded rail management, and thermal stress mitigation can significantly lower track failure rates over time (Esveld, 2001; RSSB, 2023). Additionally, the observed skewed distributions and extreme values from the early period are consistent with statistical studies of rail infrastructure data, highlighting the critical role of exploratory visualization prior to formal modeling (Cameron & Trivedi, 2013).

SECTION 2

THEORETICAL FRAMEWORK

2.1 Visualisation Focus

The visualization chosen for theoretical discussion is a time series line chart depicting the annual counts of broken and buckled rail failures within the UK rail network. The central analytical question addressed by this visualization is: How have broken and buckled rail failures evolved over time, and what long-term patterns can be identified in the condition of UK rail infrastructure? This question is particularly suited for a time series representation, as the aim is to uncover temporal trends, persistence, and structural changes rather than merely making isolated comparisons (Tufte, 2001).

2.2 Application of the Assert framework

The ASSERT framework offers a structured methodology for visual analytics, guiding the process from problem definition to interpretation (Keim et al., 2010). Each stage was systematically adhered to in the development of the time series visualization.

A - Ask the Question: The initial stage requires a clear definition of the analytical objective. The focus is on understanding temporal trends in two critical types of track failures that have significant implications for safety and reliability. Formulating a time-oriented question ensures that the visualization is crafted to uncover patterns of improvement, persistence, or change over time (Few, 2012).

S - Search for Information: Relevant information was sourced from annual datasets on rail infrastructure failures, specifically concerning broken and buckled rails. These datasets are routinely utilized in UK rail safety monitoring and reflect long-term asset conditions rather than short-term operational fluctuations (RSSB, 2023). By selecting authoritative, longitudinal data, the visualization effectively addresses system-level safety performance.

S - Structure the Data: The data were organized into the following components:

- A temporal variable (year)
- Two quantitative count variables (broken rails and buckled rails)

Annual aggregation was preserved to ensure consistency across the time series and minimize noise. This structured format facilitates direct comparisons between failure types and supports trend analysis, which is essential for evaluating infrastructure safety (Network Rail, 2024).

E - Envision the answer: Prior to creating the plot, the anticipated outcome was envisioned as a trend-based comparison to illustrate whether failure counts had decreased, remained constant, or exhibited fluctuations. A line chart was considered the most effective method for visualizing continuous changes over time, in line with visual perception research indicating that line graphs are optimal for detecting trends (Cleveland & McGill, 1984).

R - Represent the Visualization: The proposed answer was illustrated using a time series line chart. Two lines were plotted on the same axes, utilizing color to differentiate between broken and buckled rail failures. A minimalist design approach was employed to minimize cognitive load and focus attention on the data patterns rather than on decorative elements (Tufte, 2001).

T - Tell a Story: The final stage involves interpreting and communicating the findings. The visualization conveys a compelling narrative of a sustained long-term decline in both broken and buckled rail failures, with broken rails demonstrating a notably strong reduction over time. This narrative supports the view that infrastructure renewal, enhanced inspection technologies, and proactive maintenance strategies have significantly contributed to improved rail safety outcomes in the UK (Esveld, 2001; Network Rail, 2024).

2.3 Grammar of Graphics

The visualisation follows the Grammar of Graphics framework, which defines how data are mapped to visual elements (Wilkinson, 2005).

- Data: Annual counts of broken and buckled rail failures
- Aesthetics:
  - x-axis → Year
  - y-axis → Number of failures
  - colour → Failure type
- Geometries:
  - geom_line was used to represent ordered temporal data, emphasising continuity and trend
- Coordinate System:
  - Cartesian coordinates, appropriate for time-series analysis
- Scales:

    - Linear scales applied to both axes to allow accurate comparison of magnitudes
- Theme:
    - Minimal theme to reduce non-data ink and improve clarity

These choices are supported by visualisation literature, which emphasises clarity, perceptual accuracy, and alignment between data type and graphical form (Few, 2012; Wilkinson, 2005).

2.4 Justification of Design Choices

The integration of the ASSERT framework with the Grammar of Graphics ensures that the visualization is driven by specific questions, grounded in theory, and easily interpretable. The time series line chart effectively illustrates long-term safety trends, enabling viewers to quickly understand variations in infrastructure performance. This makes it especially well-suited for contexts related to rail safety and asset management (Esveld, 2001; RSSB, 2023).

SECTION 3

ACCESSIBILITY

3.1 Accessibility in Data visualization

Accessibility in data visualization refers to enabling users with various visual, cognitive, and perceptual abilities to effectively access, interpret, and understand visual information. An accessible visualization minimizes dependence on a single visual cue, such as color alone, employs clear labeling, and avoids unnecessary complexity that could impede comprehension (Few, 2012; W3C, 2018). This consideration is especially critical in safety-related contexts, where misunderstandings of data can result in erroneous conclusions or poor decision-making.

3.2 Visualisation Focus

This section evaluates the histogram showing the distribution of broken rail failures. The purpose of this visualisation is to explore the shape, spread, and skewness of the data rather than temporal change or relationships. Histograms are commonly used in exploratory data analysis to assess whether data follow normal or skewed distributions, which is essential for guiding subsequent statistical modelling choices (Cameron & Trivedi, 2013).

3.3 Accessibility strengths of histogram

The histogram showcases several features that enhance accessibility:

First, it utilizes a single fill color, which eliminates the need for users to differentiate between multiple categories based on color. This design choice improves accessibility for individuals with color vision deficiencies, as understanding the chart does not depend on distinguishing hues (Ware, 2013).

Second, the chart includes clear axis labels and a descriptive title, allowing users to quickly grasp what the bars represent. This design aids individuals with limited statistical knowledge by minimizing cognitive load (Few, 2012).

Third, histograms are built on simple geometric shapes (rectangular bars) and a Cartesian coordinate system, making them widely understood and perceptually easy to interpret. This simplicity enhances accessibility by steering clear of complex encodings that may confuse viewers (Tufte, 2001).

3.4 Accessibility limitation and critique

While histograms have numerous strengths, they also present certain accessibility challenges. One notable limitation is the presence of gaps between bins, which less experienced users might misinterpret as missing data, rather than an absence of observations within those ranges. Without proper explanation, this can lead to misunderstandings regarding the underlying distribution.

Moreover, effectively interpreting histograms relies on an understanding of binning choices, including bin width and frequency counts. Users who are not familiar with these concepts may find it challenging to accurately interpret the shape of the distribution, highlighting a cognitive accessibility limitation (Few, 2012).

Additionally, as a static visualization, histograms are not automatically accessible to screen-reader users unless they are accompanied by a textual description that summarizes key insights, such as skewness and the existence of extreme values (W3C, 2018)

Overall, the histogram is a largely accessible visualisation due to its simplicity, limited reliance on colour, and clear structure. While some interpretive challenges remain, particularly for users unfamiliar with statistical distributions, these can be mitigated through supporting text. The critique demonstrates that accessibility is an iterative design process rather than a fixed property of a visualisation.

SECTION 4

VISUALISATION CHOICE

4.1 Goal of Visualisation

The primary goal of the selected visualisation is to explore the relationship between two types of rail track failures, broken rails and buckled rails, and assess whether higher levels of one failure type tend to occur alongside the other. This exploratory goal focuses on identifying association, clustering, and variability, rather than temporal trends or distributional properties.

4.2 Visualisation focus and Justification

A scatter plot was selected as it is among the most suitable types of visualizations for examining the relationships between two quantitative variables. Scatter plots allow each observation to be represented independently, making them highly effective for identifying patterns such as positive or negative associations, outliers, and heterogeneity within the data (Cleveland & McGill, 1984).

In this context, plotting broken rail failures against buckled rail failures facilitates a visual assessment of whether years with high counts of broken rails are also linked to increased instances of buckled rails. This approach aligns directly with the exploratory goal of understanding potential interdependence between different failure mechanisms. Additionally, the inclusion of a fitted trend line enhances interpretation by emphasizing the overall direction of the relationship while preserving the visibility of individual data points (Few, 2012).

4.3 Discussion of Alternative visualisation type

Alternative 1: Line Chart (Time Series)

A line chart can effectively visualize broken and buckled rail failures, especially if the objective is to analyze changes over time. Line charts excel at highlighting trends, cycles, and long-term patterns in temporally ordered data (Tufte, 2001).

Strengths:

- Excellent for identifying long-term trends
- Intuitive for time-based data

- Supports comparison across years

Limitations:

- Less effective for analysing direct relationships between variables
- Can obscure variability when multiple series overlap

Given that the goal of this visualisation is relational rather than temporal, a line chart would be less suitable than a scatter plot.

Alternative 2: Bar Chart

A bar chart is suitable for comparing the magnitude of broken and buckled rail failures across different years or categories. While bar charts are effective for categorical comparisons, they are not designed for analyzing relationships between two continuous variables (Few, 2012).

Strengths:

- Easy to interpret
- Suitable for discrete comparisons

Limitations:

- Poor at showing relationships between two quantitative variables
- Can become cluttered with many categories (years)

Using a bar chart would limit the ability to explore correlation or dispersion, making it inappropriate for the stated goal.

4.4 Appropriateness Based on Visualisation Taxonomies

According to visualization taxonomies, such as those developed by Munzner (2014), scatter plots are particularly well-suited for tasks that involve identifying relationships and conducting exploratory analysis. The scatter plot is aligned with the objectives of "discovering correlations" and "identifying patterns," while line charts are more appropriate for "trend detection," and histograms are best for "distribution analysis." By choosing a scatter plot, the visualization method is effectively matched with the data type (quantitative–quantitative) and the analytical task of exploring relationships.

In summary, the scatter plot was chosen because it best supports the analytical goal of exploring the relationship between broken and buckled rail failures. While alternative visualisations such as line charts and bar charts offer strengths in other contexts, they are less effective for relational analysis. The choice demonstrates awareness that effective visualisation depends not only on the data available but also on the specific question being asked and the insights sought.

SECTION 5

IMPLICATIONS AND IMPROVEMENT

5.1 Ethical implications of using Boxplot visualization

The boxplot comparing broken and buckled rail failures raises significant ethical considerations as it condenses safety-critical infrastructure data into a highly compact format. In the context of rail safety, such visualizations may be utilized by policymakers, infrastructure managers, or the public to evaluate relative risks and prioritize maintenance efforts. Although boxplots are effective for comparing distributions, their abstract nature can lead to misunderstandings or oversimplifications if not carefully clarified (Few, 2012).

A notable ethical concern is that users might interpret the higher median and broader spread of broken rail failures as an indication that broken rails are inherently more hazardous than buckled rails, without taking into account variations in detection frequency, reporting practices, or exposure conditions. This could result in misguided conclusions about relative safety risks if the visualization is presented without sufficient contextual explanation (Tufte, 2001).

5.2 Risk of Misinterpretation and Use in Practice

Boxplots inherently suppress temporal and contextual information. In this instance, the visualization fails to convey that numerous extreme values for broken rails predominantly occurred in earlier years, which could mislead viewers into believing that high failure rates continue to be prevalent. Similar issues have been highlighted in transport safety reporting, where aggregated graphics can obscure progress over time and unintentionally amplify perceptions of current risk levels (Rail Safety and Standards Board [RSSB], 2023).

In the broader literature on data ethics, it is emphasized that summary statistics presented without adequate explanatory context may inadvertently mislead audiences who are not familiar with statistical conventions such as quartiles and outliers (Cairo, 2016). In a public-facing context related to rail safety, this could impact trust or perceptions regarding the performance of infrastructure.

5.3 Proposed Improvements to the Visualisation

Several enhancements could bolster the ethical and analytical quality of the boxplot:

Data Enhancements

The boxplot could be augmented by incorporating temporal stratification, such as presenting separate boxplots for early and late periods. This approach would help capture any changes over time, thereby minimizing the risk of misrepresenting historical extremes as current conditions.

Visual Design Improvements

Overlaying individual data points, such as jittered points, would enhance transparency regarding the underlying observations and lessen the abstraction typically associated with boxplots. Additionally, including brief annotations to clarify medians, interquartile ranges, and outliers would significantly improve interpretability (Few, 2012).

Accessibility Improvements

Utilizing patterned fills or labels alongside color would enhance accessibility for users with color vision deficiencies. Furthermore, providing a concise textual explanation alongside the visualization would aid screen-reader users and promote inclusive access (W3C, 2018).

5.4 Reflection

The boxplot serves as a valuable tool for comparing the spread and central tendency of broken and buckled rail failures; however, its ethical application relies on appropriate contextualization and transparency. Without adequate explanation, it risks oversimplifying complex safety data. By incorporating temporal context, explanatory annotations, and design choices that prioritize accessibility, this visualization can more responsibly facilitate discussions about rail infrastructure safety, rather than inadvertently leading to misconceptions.

5.5 Conclusion

This project exemplified the effective use of data visualization to communicate significant safety trends within the UK rail network. By employing the ASSERT framework alongside the Grammar of Graphics, the visualizations were meticulously crafted to explore, structure, and present data on broken and buckled rail failures in a clear and accessible manner. The integration of time series, scatter plots, histograms, and box plots facilitated an easy understanding of both long-term trends and distribution patterns.

Overall, the visualizations indicated a notable decline in rail track failures over time, reinforcing the notion that enhancements in maintenance and infrastructure management have contributed to improved rail safety. This project underscores the importance of thoughtful visual design, accessibility considerations, and ethical communication when conveying safety-critical data, ensuring that insights are accurate, comprehensible, and informative for both decision-makers and the public.

REFERENCES

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. New Riders.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (2nd ed.). Cambridge University Press.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531–554.

Esveld, C. (2001). *Modern railway track* (2nd ed.). MRT-Productions.

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Analytics Press.

Keim, D. A., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the information age: Solving problems with visual analytics*. Eurographics.

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

Network Rail. (2024). *Infrastructure safety and asset management annual report*.

Rail Safety and Standards Board (RSSB). (2023). *Annual safety performance report*.

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

Ware, C. (2013). *Information visualization: Perception for design* (3rd ed.). Morgan Kaufmann.

Wilkinson, L. (2005). *The grammar of graphics* (2nd ed.). Springer.

World Wide Web Consortium (W3C). (2018). *Web Content Accessibility Guidelines (WCAG) 2.1*.

APPENDIX- R Script

```
library(readr)
broken_rails_and_buckled_rails <- read_csv("broken rails and buckled rails.csv")
View(broken_rails_and_buckled_rails)
rail_data <- data.frame(
  Year = c("2002-03","2003-04","2004-05","2005-06","2006-07","2007-08",
          "2008-09","2009-10","2010-11","2011-12","2012-13","2013-14",
          "2014-15","2015-16","2016-17","2017-18","2018-19","2019-20",
          "2020-21","2021-22","2022-23","2023-24","2024-25"),
  Time_Index = 1:23,
  Broken_Rails = c(444,334,322,316,192,181,165,169,182,129,180,120,
                   95,104,88,74,86,93,82,68,99,61,53),
  Buckled_Rails = c(9,137,32,56,85,4,17,25,29,12,10,19,
                    14,9,7,14,25,12,7,13,29,4,4)
)
head(rail_data)

library(ggplot2)
ggplot(rail_data, aes(x = Time_Index)) +
  geom_line(aes(y = Broken_Rails, colour = "Broken Rails"), linewidth = 1) +
  geom_line(aes(y = Buckled_Rails, colour = "Buckled Rails"), linewidth = 1) +
  scale_colour_manual(
    values = c("Broken Rails" = "#1f78b4",
            "Buckled Rails" = "#e31a1c")
  ) +
  labs(
```

```
    title = "Trends in Broken and Buckled Rail Failures (2002–2025)",
    x = "Time (Annual Index)",
    y = "Number of Failures",
    colour = "Failure Type"
  ) +
  theme_minimal()

library(ggplot2)
ggplot(rail_data, aes(x = Broken_Rails, y = Buckled_Rails)) +
  geom_point(colour = "#6a3d9a", size = 3, alpha = 0.8) +
  geom_smooth(method = "lm", se = TRUE, colour = "#33a02c") +
  labs(
    title = "Relationship Between Broken and Buckled Rail Failures",
    x = "Number of Broken Rails",
    y = "Number of Buckled Rails"
  ) +
  theme_minimal()

library(ggplot2)

ggplot(rail_data, aes(x = Broken_Rails)) +
  geom_histogram(
    bins = 10,
    fill = "#1f78b4",
    colour = "white",
    alpha = 0.8
```

```r
 ) +
 labs(
  title = "Distribution of Broken Rail Failures",
  x = "Number of Broken Rails",
  y = "Frequency"
 ) +
 theme_minimal()


library(ggplot2)
library(tidyr)
rail_long <- rail_data %>%
 pivot_longer(
  cols = c(Broken_Rails, Buckled_Rails),
  names_to = "Failure_Type",
  values_to = "Failures"
 )
ggplot(rail_long, aes(x = Failure_Type, y = Failures, fill = Failure_Type)) +
 geom_boxplot(alpha = 0.8) +
 scale_fill_manual(
  values = c("Broken_Rails" = "#1f78b4",
        "Buckled_Rails" = "#e31a1c")
 ) +
 labs(
  title = "Comparison of Broken and Buckled Rail Failures",
  x = "Failure Type",
  y = "Number of Failures"
```

) +

theme_minimal() +

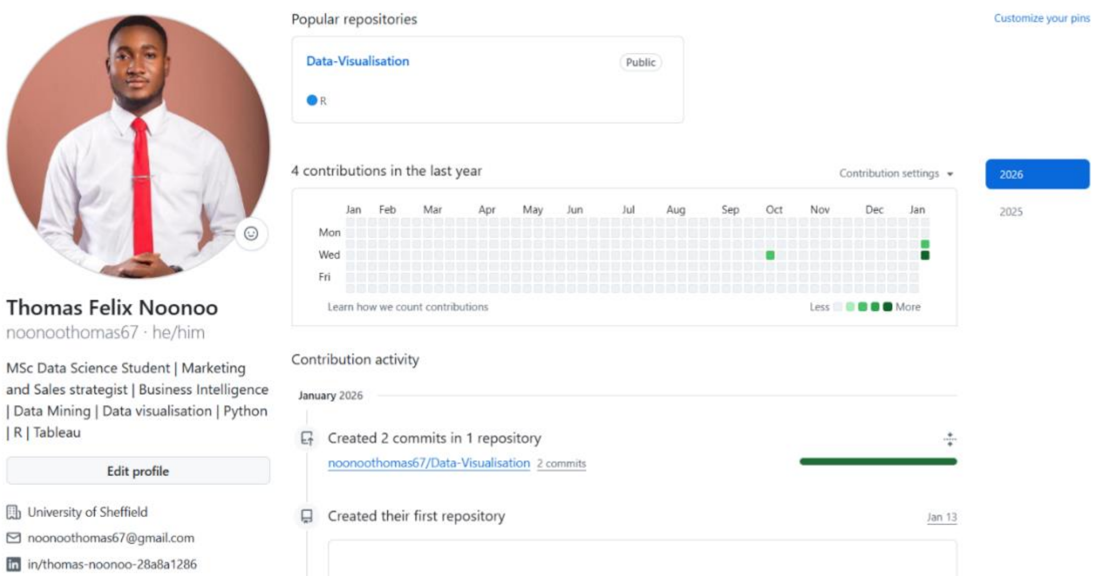
theme(legend.position = "none")

GITHUB

GITHUB profile

https://github.com/noonoothomas67

The profile README provides an overview of my academic background, technical skills, and featured data science projects. It highlights my interests in data analysis, visualisation, and statistical modelling, with direct links to my coursework projects.



Project page

The IJC445 project repository is available at:

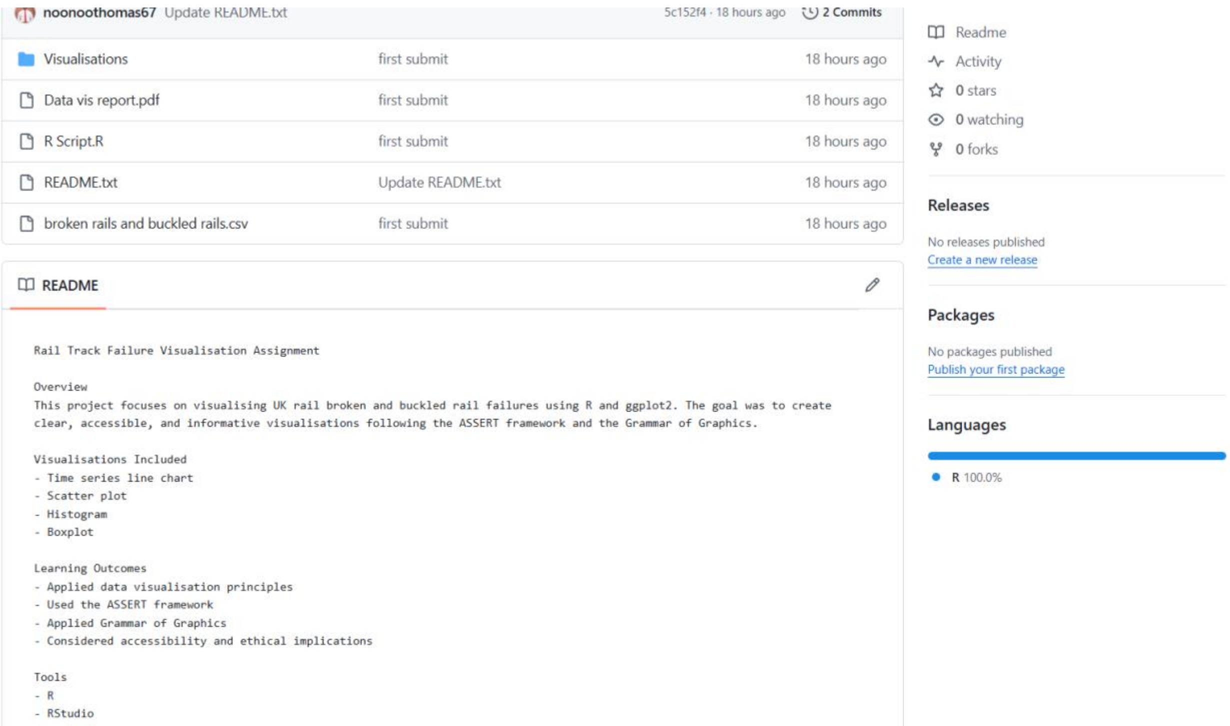
https://github.com/noonoothomas67/Data-Visualisation

his repository includes:

- The visualisation R script (visualisations.R)
- Generated plot images

- The visualisation report
- A README describing the ASSERT framework, Grammar of Graphics, and accessibility considerations

The repository demonstrates how data visualisation theory was applied in practice to communicate rail safety trends.

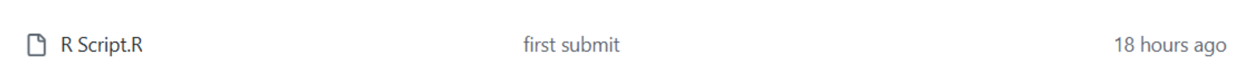


Code

All code used in project is stored in clearly labelled R scripts

R Script.R

GitHub allows the code to be viewed directly in the browser, making it easy for users to review the methods and reproduce the analysis.



Instructions on Running the Code

Each repository includes a How to Run section in the README file. These instructions explain:

1. How to download or clone the repository
2. How to open the R script in RStudio
3. How to load the dataset
4. How to run the code to reproduce the analysis and visualisations

```
How to Run the Code

1. Download or clone this repository
2. Open the file `visualisations.R` in RStudio
3. Ensure the dataset `broken rails and buckled rails.csv` is in the same folder as the script
4. Install required packages if needed
5. Run the script to generate the visualisations

Required packages:
- ggplot2
```