

## Data Analytics Implementation

Wales is a relatively small country that has access to transport by land, sea and air. In Wales the dominant form of transport in terms of distance travelled and amount of journeys undertaken is via personal use of cars. The road network experiences congestion frequently, with the most prominent sites located on or directly connected to the M4 motorway which causes far reaching consequences on accessibility as this is the main route connecting the major locations in the country (Welsh Government, 2021).

Wales has one of the lowest per capita spend on public transport in the UK, which decreased from £74.7 million in 2012-13 to £45.4 million in 2016-17 (The Future Generations Report, 2020). In recent years there has been an increase in the use of train travel, however public satisfaction remains a major issue (The Future Generations Report, 2020). The rail system within Wales frequently experiences overcrowding leading to passengers being unable to get onto trains or enduring cramped situations. The rail network is limited by its east-west nature, greatly restricting travel to the north or south which may leave more rural areas isolated (Fry et al., 2013).

The National Survey for Wales (2019) showed that people felt that using public transport after dark felt less safe than any other mode of transport. Crime around public transport is on the increase with 13.3% more offences occurring in 2019 compared to 2018, the majority of which were public order offences, violence against the person and theft (Welsh Government, 2021). Patients in highly populated areas within Wales can reach medical facilities within an hour via public transport. In more rural areas where service availability is greatly reduced there is a significant increase in travel times (Fry et al., 2013). Travel via bus routes has seen a decrease between 2003-2017.

Multiple pandas datasets were created to store and manipulate the data. Pandas allows for simple methods of preprocessing and transforming data as well as allowing easy integration with other libraries which will allow flexibility and a wide range of analysis (Kupferschmidt, 2020). To reduce the amount of data preparation required, an individual data frame was created for each table used. Individual data frames allowed for greater control of what data is included in the visualisation as the parameters allows individual columns and rows to be collected from the data reducing the need to remove null and drop unnecessary columns. After the individual table data frames were created selected tables were concatenated to create data frames with more in depth data that could be further analysed.

To ensure that seaborn would label the x-axis with the independent variable for each table the index was set to contain the variable that would be plotted on the x-axis.

Columns that contained positive responses to feelings of ease and safety have been combined to create columns containing the total percentages of responses that were positive for comparison to other variables.

Correlation matrices were utilised alongside visualisations of the data to show the statistical strength of the relationships between variables. The method was left as the Pearson's method which is the default as this can be applied to non-quantitative variables with a rank order to compute a Spearman rho correlation coefficient (Hadd & Rodgers, 2021). To ensure that repeated data was not displayed a mask was created which removed the upper right half of the matrix, this ensures that the results are easier to interpret. A red to blue colour map was selected to provide a visual correlation between positive and negative correlations. To create correlation matrices all values in the data frame were set as floats. For variables such as feelings of safety, a divergent colour palette was used with intuitive colours were selected that are instantly associated with the results they are displaying. This allows positive and negative outcomes to be identified immediately based on colour alone (Kashyap, 2020).

The isnull and sum methods were used to detect how many null values there were in a data frame. Within the provided data tables missing values are caused by a small sample size, which also leads to low confidence values in other values. Therefore instead of replacing the missing values, the columns which contain any missing values were omitted from the dataframe. Seaborn was selected over matplotlib to create visualisations of the data. The Seaborn library is more efficient at handling pandas data frames whilst using comparatively simple syntax that offers more functionality than Matplotlib (Hacksight, 2022). A limitation of seaborn is that multiple figures may lead to out of memory related issues, however due to the limited amount of figures being created this should not be an issue (Hacksight, 2022).

Tables 2 and 8 were changed from being displayed as a line and bar graph respectively to a box plot. After initially creating the charts, the graphs did not display the range of data the distribution of the data in as a box plot could. Bar charts are useful for showing counts, whereas box plots can be utilised to display the characteristics of the distribution of data (Streit & Gehlenborg, 2014). A limitation of box plots is that the exact values of the points of data are not retained, meaning that box plots are unable to show detailed analysis of the data (Ladkin, 2019).

For all visualisations that displayed satisfaction against other variables, the satisfaction scores were altered to display on a percentage scale of 0-100% rather than 0-10 for easier comparison with other variables.

Members of the public that do not have use of a car have higher levels of satisfaction (mean: 6.4) than the public that do have use of a car (mean: 5.8). This suggests that

there is a factor(s) related to the roads that are having a detrimental effect on the satisfaction level of users of the transport system.

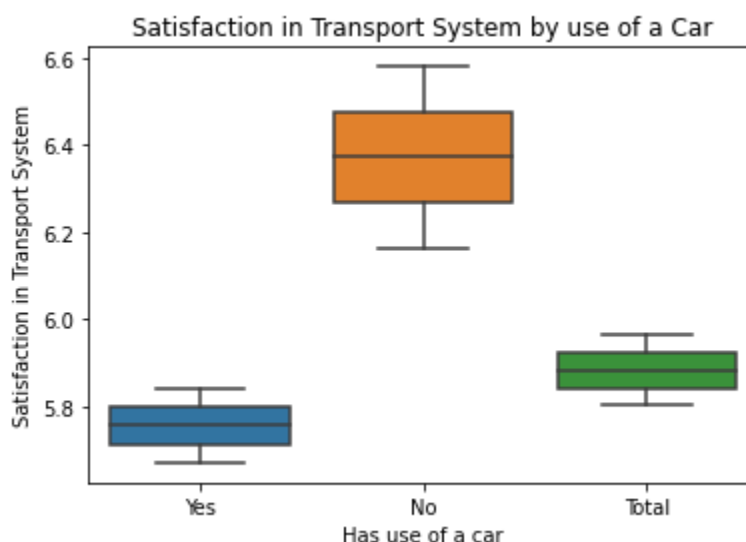


Figure 1: Satisfaction in Transport System by use of a car

North Wales (6.2) is the service that has the highest mean satisfaction levels with Gwent (6.1) and Cardiff & the Vale (6.0) rating above the total average (5.9). The Mid & West area (5.4) had significantly lower satisfaction levels than the other services.

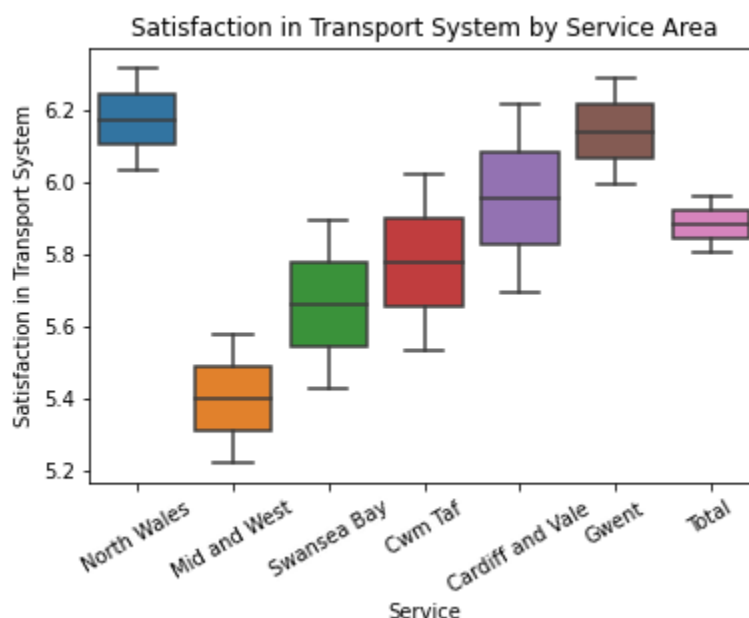


Figure 2: Satisfaction in Transport System by Service Area

The 20% most deprived from access to services section of the public reported the lowest mean level of satisfaction in the transport system. The highest satisfaction levels

were recorded in the 20% least deprived category. As levels of accessibility to services increased the reported rate of satisfaction in the transport system increased.

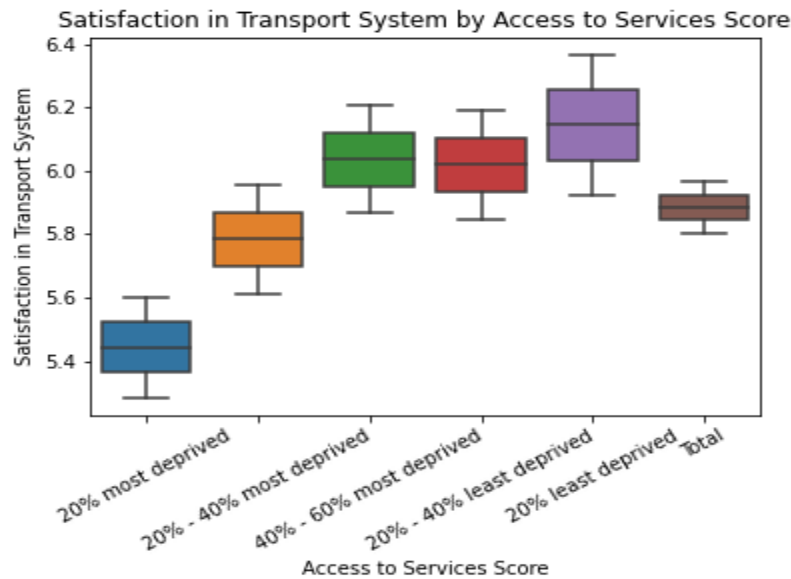


Figure 3: Satisfaction in Transport System by Access to Service Score

Around 80% of people feel safe or very safe using public transport after dark. This means that one in five people (20%) do not feel safe using public transport after dark.

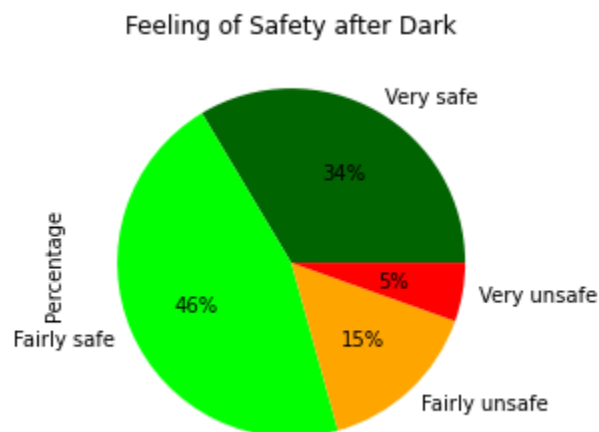


Figure 4: Feeling of Safety to use Public Transport after Dark

As members of the public get older they feel less safe and find it harder to use public transport which sees satisfaction decrease up to the age of 64 years old after which satisfaction levels start to increase again even though feelings of ease and safety continue to decrease. There is a strong positive correlation between how easy it is to get to medical facilities and how safe the public feel using public transport after dark. There is a weak negative correlation between how safe people feel using public

transport after dark and their satisfaction levels. There is a strong negative correlation between how easy it is to get to medical facilities and the public's satisfaction levels.

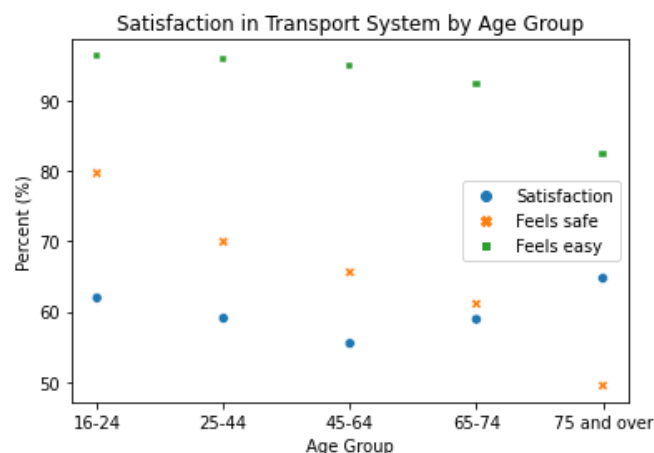


Figure 5: Public scores categorised by age groups

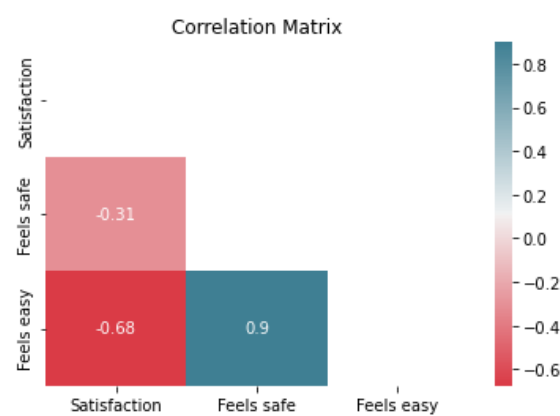


Figure 6: Age groups matrix table

There is a strong negative correlation between the overall satisfaction and the percentage of people that have use of a car. There is a strong negative correlation between the overall satisfaction and the percentage of people that feel safe using public transport at night. There is a strong positive correlation between the percentage of people that have the use of a car and the percentage of people that feel safe using public transport at night. As levels of deprivation decrease there is a significant increase in the percentage of the public that has access to a car and an increase in the feeling of safety using public transport in the dark. However, overall satisfaction in the transport network decreases which suggests that another factor is causing the satisfaction rating to decrease.

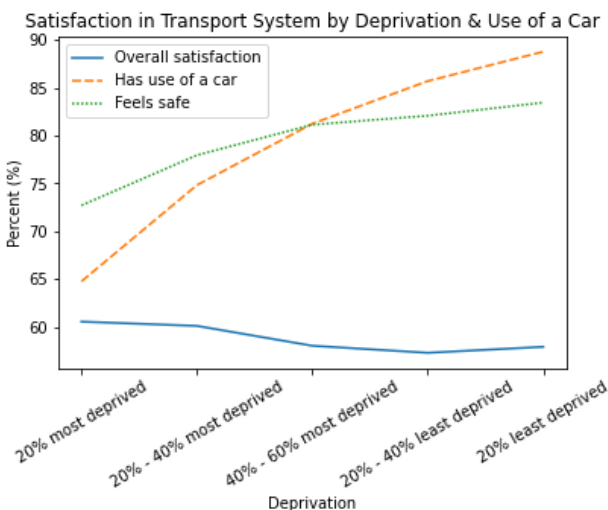


Figure 7: Public scores categorised by deprivation groups

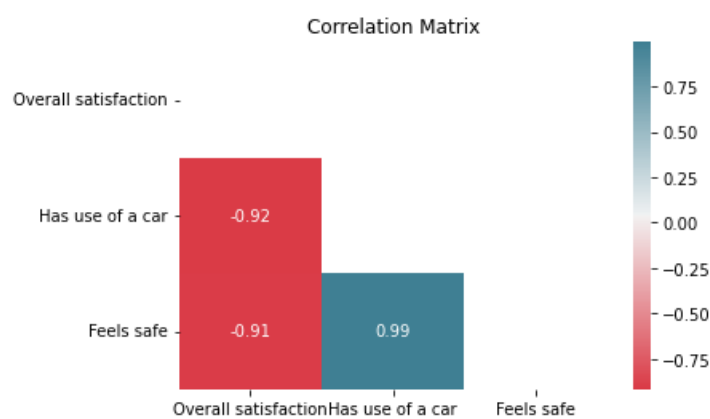


Figure 8: Deprivation groups matrix table

A high percentage of employed people have cars but have lower satisfaction in the transport system even though they find it easier to get to medical facilities and feel safer

travelling in the dark. Therefore, they could be less satisfied with the road network even though it is perceived to be easier and safer to use.

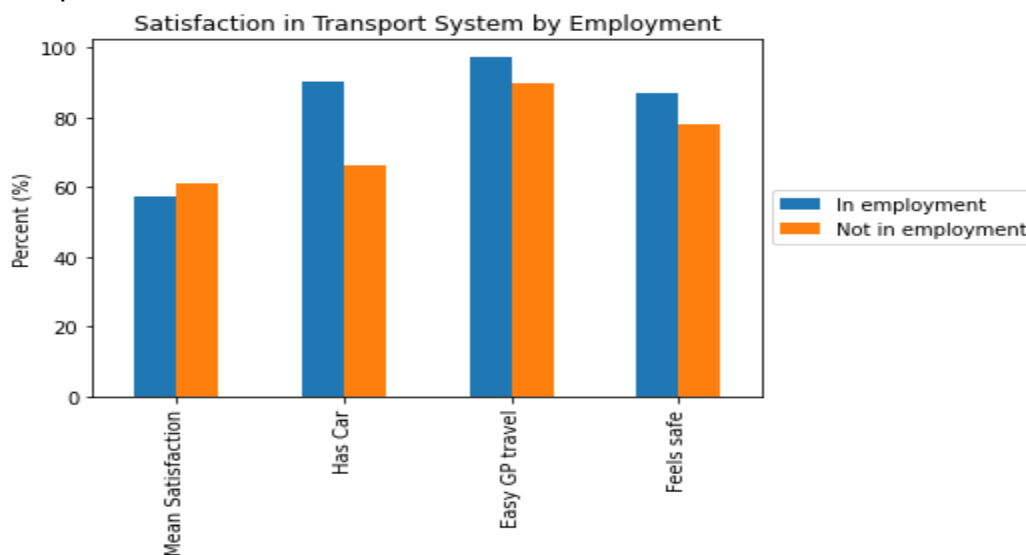


Figure 9: Satisfaction in transport system by employment status

Even though there is no significant difference in satisfaction scores. There is a significant difference between the percentage of female users of transport that feel safe compared to male users. Around 31% of females reported feeling unsafe compared to 11% of males.

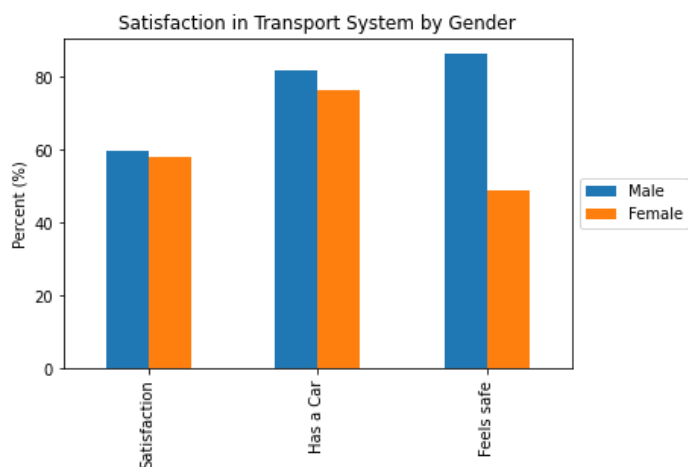


Figure 10: Satisfaction in transport system by Gender

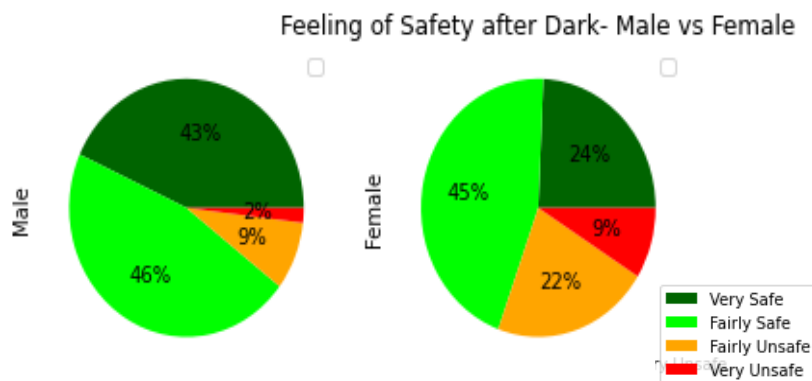


Figure 11: Feeling of Safety after Dark- Male vs Female

The factor that had the strongest correlation to affecting the satisfaction with the transport system was the accessibility to the transport system. The areas that have the least access to public transport reported the lowest levels of satisfaction. To increase overall satisfaction levels in the transport system the accessibility of public transport should be improved. The Mid & West service was rated as being the service with the lowest perceived satisfaction score. This score was similar to the satisfaction score of the 20% most deprived to access services. Therefore an issue with access to services

within the Mid & West service area may be present. Further research should be conducted to discover what factors are causing this low level of satisfaction. To improve public satisfaction in the transport system public transport should be made more accessible to the public. Accessibility could be limited by a high percentage of Wales being rural outside of the major settlements.

The main factor that affected satisfaction scores across multiple demographics of the public was the percentage of the population that had access to a car. As access to cars increased the perceived satisfaction in the transport system decreased, even as feelings of safety and ease increased. Therefore, further research should be conducted to discover what factor(s) related to travel by car are negatively affecting satisfaction scores. To improve satisfaction in the transport service the public could be encouraged to use public transport, however this will be limited by how accessible public transport is within the region. The Future Generations Report (2020) found around 57% of people surveyed said that they would be willing to use their cars less and public transport more, only if there was an improvement in the quality of public transport. Therefore investing in public transport systems may greatly reduce traffic problems. This will be limited by Wales having a limited amount of per capita spending on transport as reported by The Future Generations Report (2020) and may not see much financial return on the investment.

A significantly higher percentage of female passengers of public transport felt unsafe while using public transport in the dark. As the perceived feelings of ease in the use of public transport increased the feeling of safety while using public transport at night also increased. Therefore, to increase how safe the public feel while using public transport in the dark, the transport system should be improved to make it easier and more accessible to the public. Joewono and Kubota (2006) found three areas of security that needs to be addressed for security on public transport:

1. Safety from crime: lighting, monitoring and help points.
2. Safety from accidents: avoidance of hazards and safeguarding.
3. Perception of security: Conspicuousness of safety measures and press relations.

Improving these three areas will help the public feel more secure using public transport in the dark, However this may be very costly with a limited effect on satisfaction in public transport.

This study is limited by only containing data from one year, therefore any change that occurs over time can not be observed and the results can not be compared to other years to check for anomalies in the data. Future research should incorporate multiple years of data. Another limitation of the data is that individual modes of transport (rail,

bus etc.) were not accounted for so analysis could not be conducted on the difference between modes of transport.

## References:

Biderbost, P. and Carrasquero, G. (2021) "Learn to create bar charts in RStudio with data from a study of sleep patterns for college students (2012)." Available at: <https://doi.org/10.4135/9781529777819>.

DeFranzo, S.E. (2022) Advantages and disadvantages of surveys, Snap Surveys Blog. Available at: <https://www.snapsurveys.com/blog/advantages-disadvantages-surveys/> (Accessed: November 11, 2022).

Fry, R., Rodgers, S. and Lyons, R. (2013) Public Transport Overview, ABMU Changing for the Better: Public Transport . Centre for Health Information, Research and evaluation (CHIRAL), Swansea University. Available at: <https://sbuhb.nhs.wales/files/freedom-of-information-disclosure-log-2020/august/20-h-041-public-transport-report-2013/> (Accessed: November 3, 2022).

Hackshaw, A. (2008) "Small studies: Strengths and Limitations," European Respiratory Journal, 32(5), pp. 1141–1143. Available at: <https://doi.org/10.1183/09031936.00136408>.

Hacksight (2022) *Difference between matplotlib vs Seaborn*, GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/difference-between-matplotlib-vs-seaborn/> (Accessed: November 17, 2022).

Hadd, A.R. and Rodgers, J.L. (2021) *Understanding correlation matrices*. Los Angeles: SAGE.



Jahedi, S. and Méndez, F. (2014) “On the advantages and disadvantages of subjective measures,” *Journal of Economic Behavior & Organization*, 98, pp. 97–114. Available at: <https://doi.org/10.1016/j.jebo.2013.12.016>.

Joewono, T.B. and Kubota, H. (2006) “Safety and security improvement in public transportation based on public perception in developing countries,” *IATSS Research*, 30(1), pp. 86–100. Available at: [https://doi.org/10.1016/s0386-1112\(14\)60159-x](https://doi.org/10.1016/s0386-1112(14)60159-x).

Kashyap, A. (2020) *8 rules for optimal use of color in Data Visualization*, Medium. Towards Data Science. Available at: <https://towardsdatascience.com/8-rules-for-optimal-use-of-color-in-data-visualization-b283ae1fc1e2> (Accessed: December 7, 2022).

Krzywinski, M. and Altman, N. (2014) “Visualizing samples with box plots,” *Nature Methods*, 11(2), pp. 119–120. Available at: <https://doi.org/10.1038/nmeth.2813>.

Kupferschmidt, K. (2020) *Spark vs Pandas, part 1 — Pandas - towardsdatascience.com*, *Spark vs Pandas, part 1 — Pandas*. Available at: <https://towardsdatascience.com/spark-vs-pandas-part-1-pandas-10d768b979f5> (Accessed: November 14, 2022).

Ladkin, A. (2019) How to create a box-plot chart, Sciencing. Available at: <https://sciencing.com/create-boxplot-chart-5858878.html> (Accessed: November 14, 2022).

National Survey for Wales (2019) National Survey for Wales 2019, GOV.WALES. National Survey for wales. Available at: <https://gov.wales/national-survey-wales-results-viewer> (Accessed: November 3, 2022).

Royal Geographical Society (nd) “A Guide to Scatter and Line Graphs.” Available at: <https://www.rgs.org/CMSPages/GetFile.aspx?nodeguid=227c9dd8-b2e5-4081-a67f-8a59641207c2&lang=en-GB>

Statistics Kingdom (nd) Box plot maker. Available at:  
<https://www.statskingdom.com/boxplot-maker.html> (Accessed: November 14, 2022).

Streit, M. and Gehlenborg, N. (2014) “Bar charts and Box Plots,” *Nature Methods*, 11(2), pp. 117–117. Available at: <https://doi.org/10.1038/nmeth.2807>.

The Future Generations Report (2020) Executive Summaries Transport - exec summary - futuregenerations.wales. Available at:  
<https://www.futuregenerations.wales/wp-content/uploads/2020/07/Transport-Exec-Summary.pdf> (Accessed: November 3, 2022).

Welsh Government (2021) A New Wales Transport Strategy Consultation Draft , GOV.WALES. Available at:  
<https://gov.wales/sites/default/files/consultations/2020-11/supporting-information-transport-data-and-trends.pdf> (Accessed: November 3, 2022).