



City Research Online

City St George's, University of London

Citation: Slingsby, A. & Hyde, J. (2024). Zoomable Level-of-Detail ChartTables for Interpreting Probabilistic Model Outputs for Reactionary Train Delays. 2024 IEEE Visualization and Visual Analytics (VIS), pp. 241-245. doi: 10.1109/VIS55277.2024.00056 ISSN 2771-9537 doi: 10.1109/VIS55277.2024.00056

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://city-ref-test.eprints-hosting.org/id/eprint/33874/>

Link to published version: <https://doi.org/10.1109/VIS55277.2024.00056>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Zoomable Level-of-Detail ChartTables for Interpreting Probabilistic Model Outputs for Reactionary Train Delays

Aidan Slingsby*
City, University of London

Jonathan Hyde†
Risk Solutions

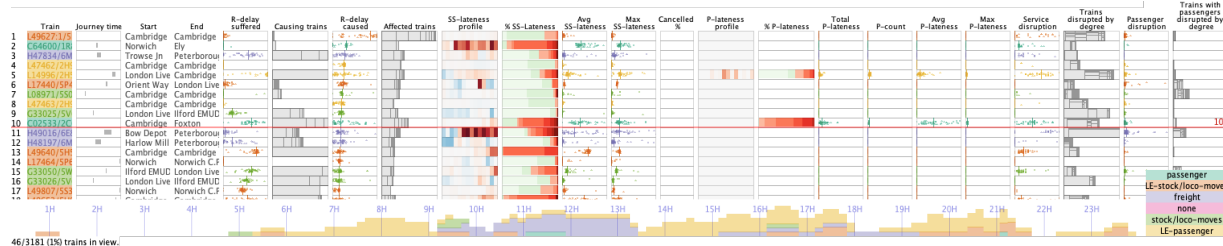


Figure 1: A Zoomable Level-of-Detail ChartTable, in which train delay metrics (columns) are represented as mini-charts for each train (row). The time histogram at the bottom indicates the number of trains operating over a 24h period and can be used to interactively filter trains by hour of day.

ABSTRACT

“Reactionary delay” is a result of the accumulated cascading effects of knock-on train delays which is increasing on UK railways due to increasing utilisation of the railway infrastructure. The chaotic nature of its effects on train lateness is notoriously hard to predict. We use a stochastic Monte-Carlo-style simulation of reactionary delay that produces whole distributions of likely reactionary delay and delays this causes. We demonstrate how Zoomable Level-of-Detail ChartTables – case-by-variable tables where cases are rows, variables are columns, variables are complex composite metrics that incorporate distributions, and cells contain *mini-charts* that depict these as different *levels of detail* through zoom interaction – help interpret whole *distributions* of model outputs to help understand the causes and effects of reactionary delay, how they inform timetable robustness testing, and how they could be used in other contexts.

Index Terms: Level-of-detail, mini-charts, distributions, stochastic modelling.

1 INTRODUCTION

We demonstrate how Zoomable Level-of-Detail ChartTables can help interpret probabilistic Monte-Carlo-style simulations of reactionary delay and its effect on train lateness. “Reactionary delay” is the result of the accumulated cascading effects of knock-on train delays [10]. Its interdependent “knock-on” nature makes its effects hard to predict, with resulting delays often longer than the original delays. Reactionary delay is an increasing problem on UK railways as the number of scheduled train services are increasing [13]. Our approach to helping understand the impact of reactionary delay is to use a stochastic “Monte-Carlo” style Agent-Based Model (“SaviRPM”) that simulates trains running to a fixed timetable. Each model run incorporates randomly generated primary delays (e.g. late departure due to passenger overcrowding) based on historically-derived probabilities, where each run represents a possible alternative “day”, detailing primary delays, resulting reactionary delays and their impacts on other trains. We can use

this for comparing timetables for robustness and resilience.

Key to our approach is to enable consideration of the *distribution* stochastic outputs using interactive visualisation. Agent-Based Models (ABMs; particularly Individual-Based Models) are often used to simulate individual behaviour to predict resulting larger-scale processes [14]. We are modelling individual trains to help understand the accumulation of multiple knock-on effects at a system level. Understanding mechanisms within such models are so challenging that rich probabilistic data are *often reduced to high-level summaries* [6]. Information Visualisation and Visual Analytics are often cited solutions to enable analysts to consider and interpret more detail and nuance than through high-level summaries alone.

In our Zoomable Level-of-Detail ChartTables, rows are *trains*, columns are *composite metrics* that quantify aspects of reactionary delay, visually-represented as *mini-charts* within cells that represent the metric for each train across all model runs. Four types of mini-chart (section 2.1) for four different metric types have visual representations at two levels-of-detail depending on the zoom level (Fig. 2). Vertical zooming changes the height of rows. Where rows are narrow and numerous, low level-of-detail mini-charts summarise across all trains. Where zoomed rows are wide/tall, high level-of-detail mini-charts depict the distribution of model runs.

Whilst many of these design characteristics are not new, we show how our design and interactions enable interpretation of probabilistic Monte-Carlo-style simulation for identifying the effects of reactionary delay. We show how they are helping the UK’s railway industry understand reactionary delay and inform timetable redesign to reduce the effects of reactionary delay. We reflect more widely on how they can be used in other contexts to facilitate comparison and interpretation of complex metrics and their distributions.

Our contributions are to: (a) present Zoomable Level-of-Detail ChartTables, (b) describe a set of composite metrics and visual representations (mini-charts) that capture different aspects of probabilistic reactionary delay at different levels of detail, (c) demonstrate their use for interpreting stochastic Monte-Carlo-style simulation results, and (d) reflect on their potential wider use.

2 ZOOMABLE LEVEL-OF-DETAIL CHARTTABLES

This work arose from a series of workshops and projects starting in 2018. A workshop run by Rail Safety & Standards Board (RSSB) (which funds applied research for the UK rail industry) established a need for a better understanding of reactionary delay to help improve punctuality. They published a call for applied research. We

*e-mail: a.slingsby@city.ac.uk

†e-mail: jonathan.hyde@risksol.co.uk

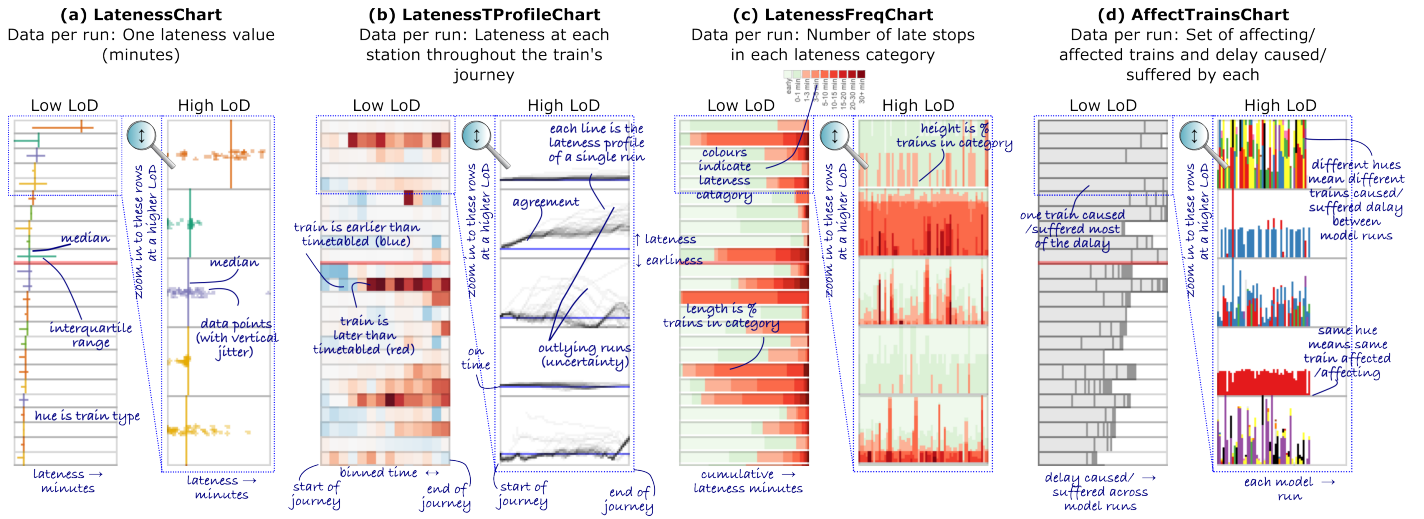


Figure 2: Four metric types depicted by four different mini-charts. Each has both low and high level-of-detail (LoD) variants. In all cases here, the high LoD variant on the right shows the detail of the top five rows of the low LoD on the left (outlined in a blue dotted line). An appropriate threshold then vertically zooming rows determines which variant is used.

partnered with two UK Train Operating Companies ([Great Western Railways](#) and [Greater Anglia](#)) and our research proposal – to investigate the feasibility of combining stochastic Agent-Based Modelling with highly interactive visualisation – was accepted and funded by RSSB. Our approach was shown to be feasible and further funded projects (to develop the ideas) and private consultancy (see acknowledgements; to apply and assess the feasibility, robustness and resilience of alternative timetables). Throughout, interactive visualisation was key for validating and interpreting the modelling. The work has been deeply embedded in industry needs. Early stages of the work used Koh *et al*'s user-based approach of workshops, followed by iterative prototyping loosely based on the AGILE principles of short development/feedback cycles and continual re-prioritisation. This work now underpins consultancy and the software (“SaviRPM”) as well as is being licensed to RSSB and a Train Operating Company for their use. In this short paper, we specifically focus on the Zoomable Level-of-Detail ChartTable technique from this work.

We established some requirements to drive our designs:

- R1: **Depict multiple metrics.** Multiple aspects of lateness help us quantify delays, establish consequences, and identify problematic trains. Metrics include those that relate to delays caused by the train, delays suffered by the train, many of which cause train arrival lateness at stations and subsequent passenger arrival lateness ([listed here](#)).
- R2: **Summarise composite metrics.** Although some metrics are single values, many are composite and cannot be expressed with a single number. See section 2.1 the metric types we used.
- R3: **Depict distribution of metrics across trains.** This is to help rank trains based on their delay metrics; i.e. their different contributions to delays or vulnerability to delay.
- R4: **Depict distribution of metrics across model runs.** This is to help use consider the consistency of different types of delays between model runs and to help where there are plausibly likely serious worse cases. This aspect of probabilistic modelling is often neglected, yet considered important.

For R1, we used the commonly-used case-by-variable table [15] where cases (trains) are rows and variables (metrics) are columns,

enabling multiple metrics to be considered ([listed here](#)).

R2, R3 and R4 call for a solution that can depict complex metrics. Our solution is to use mini-charts that are embedded within cells. These depict composite metric values using different chart types for different metric types. Bertin demonstrated the value of depicting data using visual variables [3]. Mini-charts in tables are widely used for both single value metrics [12], for multiple levels of abstraction [2] and for more complex metrics that summarise various types of aggregations [5, 7, 8]. Mini-charts are also used in multivariate geographical mapping [18, 16, 17].

R3 and R4 call for a solution that succinctly summarises both the metric value per train (R3) and also its distribution across model runs (R4) for that train. Our solution uses mini-charts with two levels-of-detail variants that correspond to these options representing two different abstractions [2]. R3 also calls for identifying the most problematic trains and considering the distribution of metric values across all trains. Our solution is to sort trains based on the median or variance values of the specific metric. This helps identify problematic trains in terms of different metrics and the distribution trains (Fig. 3). Red horizontal decile lines delineate the cumulative delay causes/suffered by trains in 10% chunks (Fig. 3). Tooltips provide the numbers where required (e.g. Fig. 6).

To summarise the technique: in Zoomable Level-of-Detail ChartTables, rows are *trains*, columns are *composite metrics* that quantify aspects of reactionary delay, visually-represented as *mini-charts* within cells that represent the metric for each train across all model runs. Four types of mini-chart (section 2.1) for different metric types have visual representations at two levels-of-detail depending on zoom level (Fig. 2). This semantic zoom (variant of zoom that is not purely geometric [1]) applies only on the y-axis, changing the height of rows but leaving their widths intact. Where rows are narrow and numerous, low level-of-detail mini-charts summarise across all model runs. Where zoomed rows are wide/tall, high level-of-detail mini-chart depict the details of the whole distribution of model runs within that train. See the [video](#).

This is a similar approach to Rao and Card's “Table Lens” [15] in which they demonstrate single value metrics where bar (charts) are used for low level-of-detail compact summaries, directly expressed numbers for high level-of-detail, with level-of-detail zooming that applies to subsets of rows and/or column to give focus+context. Our use differs in three ways: (a) we have more complex met-

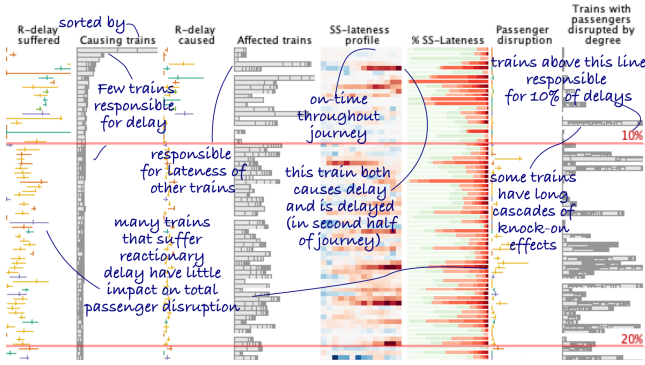


Figure 3: Excerpt zoomed out to the top 20% of delay-causing trains, sorted by delay caused showing the “drop-off”. Only a few trains are suffering significant reactionary delay. This is caused by only a few other trains, but they are not causing much reactionary delay to other services and are not themselves delayed much nor causing passenger journey disruption. Trains suffering 10-20% of the total reactionary delay go on to cause delays to other services.



Figure 4: Four examples of high LoD “LatenessTProfileCharts” (Fig. 2b) showing different lateness consistency between model runs.

rics; (b) low and high levels-of-detail are used to summarise across model runs and to show distributions between model runs respectively (instead of charts vs numbers); and (c) and we apply the level-of-detail zooming on rows only (instead of rows and column) across the whole table (instead of subsets of rows). Our emphasis is their use for helping interpreting probabilistic model outputs.

2.1 Mini-chart designs

Our metrics (section 5) fall into four types, each with a different graphical representation for both level-of-details, as shown in Fig. 2:

(a) **LatenessChart** (Fig. 2a) is used where there is a single value per model run; e.g. reactionary delay minutes caused. The *low LoD* variant summarises this across the model runs as marks depicting the median and standard deviation. The *high LoD* variant is a dot-plot, with a dot for each model run. Animated jitter [4] reduces misleading effects of overplotting.

(b) **LatenessTProfileChart** (Fig. 2b) depicts the lateness of a train over the course of its journey. The *low LoD* variant is a temporally binned heatmap coloured by average lateness throughout its journey with a diverging blue/red colour scheme for early/late station stop lateness. In Fig. 2b, the first (at the top) train is on-time throughout its whole journey, whereas the second train starts off late and does not recover after a sharp increase in lateness early in its journey, perhaps indicating a problematic station. The **high LoD** variant is a line graph where each line represents a model run. In Fig. 2b, the on-time nature of the first train is reflected in all model runs, but for the second train, the model run delays diverge after the aforementioned sharp increase in lateness. Fig. 4 shows some alternative high LoD LatenessTProfileCharts with observations.

(c) **LatenessFreqChart** (Fig. 2c) shows the average frequency of late station stops within lateness categories (using the UK railway industry’s categorisation of “early”, “0-1min”, “1-3min”, etc [11]; see Fig. 2c). The lateness categories are green for <1min late categories and red for >1min late categories, with darker colours

for later lateness categories. The *low LoD* variant uses proportional bar length to indicate the average number of station stops in each lateness category (the greater the proportion of dark red, the later the train). In the *high LoD* variant, this is rotated vertically and one per model run is displayed side-by-side indicating whether there is consistency across model runs. In the *high LoD* variant in Fig. 2c, the second train is consistently late at station stops whereas the third train is sometimes late and sometimes not.

(d) **AffectTrainsChart** (Fig. 2d) shows the trains that either *affect* or *are affected* by the row’s train and **how much delay** each train is affected or suffered. For this description, let us *assume these are the trains causing the delay*. The *low LoD* variant is a stacked barchart where each sub-bar represents a train that causes delay to this row’s train, where its length indicates the magnitude of delay this caused to this train. In the *high LoD* variant, this is rotated vertically and there is a bar for each model run. Each sub-bar coloured by the train number/identifier so that *consistent hues* indicate that *exactly the same trains are involved* in the different multiple model runs. The first train in Fig. 2d for the low LoD variant is only significantly delayed by two trains (one much more than the other). However the inconsistent colours in the high LoD variant indicates that these are different trains in different model runs. For the fourth train, the low LoD variant indicates that one one train causes the delay (one bar) and the high LoD variant confirmed that this is the same train in all cases. Such insights help decide which trains to focus on for mitigating actions.

In general, the low LoD variants use the x-axis to facilitate comparison *between trains* (rows), and high LoD variants use of the y-axis to facilitate comparison *between simulation runs*. In the high LoD variants, model runs are ordered by model run number/identifier so that the y-position of the bar across charts corresponds to the same model run. This is important as it help enable a “bad day” to be identified across charts, especially with interactions that highlight identical model runs over the whole table (section 2.2). Model runs can also be sorted by the value for the metric in help indicate the numerical distribution between model runs.

2.2 Interactions

Semantic zoom. A key contribution of this paper, this triggers mini-charts to be viewed in different level-of-detail modes – shown side-by-side in Fig. 2 – depending on a zoom level appropriate for the mini-chart. In Fig. 2, the high LoD variant is zoomed such that only the top four rows remain visible. In Fig. 3, the ChartTable is zoomed out to trains that cause 20% of the overall delay, enabling the distribution of a given metric between trains to be assessed.

Sorting. Trains can be sorted by metric based on the median or dispersion across model runs. This enables the most or least problematic trains to be identified and the consistency of that metric between the simulation runs and the relationship to other metrics can be determined. See Fig. 3 and Fig. 5.

Scaling. All the cells contain mini-charts. Each metric/column has its own x and/or y axes, which are scaled between zero and the 95th percentile by default. Each column can have its range easily adjusted. The mouseovered cell will always show the entire mini-chart, overlapping its surroundings if necessary.

Tooltips are providing details on demand (example in Fig. 6).

Train filtering allows filtering by train identifier and category.

Temporal filtering. The histogram at the bottom Fig. 1 shows the number of trains by category throughout the day. These can be filtered by time-of-day mouse, preserving sorting and scaling.

Highlighting. Highlighting is an interactive technique for associating data points across the table. Fig. 6 illustrates that selecting a train and an “AffectTrainChart” metric results all trains/rows in view involved in the metric being highlighted (actually others “lo-lighted”) making it easier to access the metric values for this train. The same technique is also used for identifying the same model

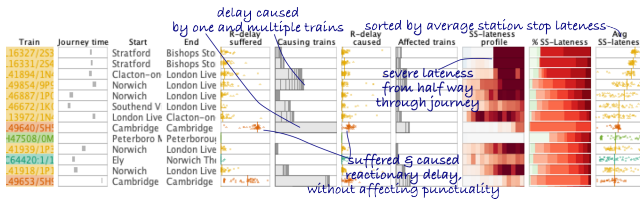


Figure 5: Example sorted by “average station stop lateness” with observations made on other aspects of lateness for these trains.

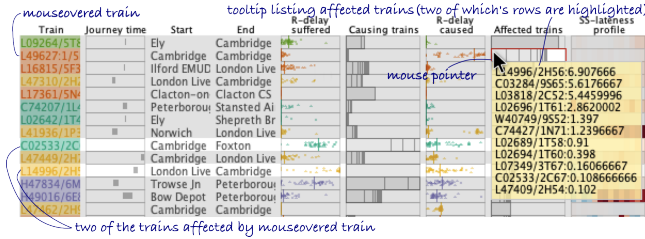


Figure 6: “Highlighting” interaction that identifies two of the trains involved with the train identified with the mouse cursor.

runs across the table to help assess the impact of “bad day” model runs across the dataset.

3 EXPERIENCES, DISCUSSION AND FURTHER WORK

3.1 Reflections on our use of the technique

Alternative ‘cases’. The examples presented here have cases (rows) as trains. By summarising the model output by station stops, we have also used *station stops as cases/rows* where the metrics characterise the delays at these locations. This important complementary information enables us to identify problematic locations. For example, identifying key “pinch-point” stations at which reactionary delays to many trains lead to huge impacts on other parts of the network gives clues to possible solutions.

Adding metrics. It was easy to add additional metrics as the work progressed. This included adding a set of metrics where delays are weighted passenger numbers by highlighting the trains/stations in terms of impact to passengers.

Visual scalability. We have been able to deal with hundreds of simulation runs and thousands of trains, including the ability to zoom-out to all trains. A limitation of our implementation is the requirement for at least one pixel per train and model run (the latter with high LoD variants). We can improve the visual scalability by appropriately sampling trains and model runs where not enough pixels are available to depict them all. This may need to happen dynamically, depending on zoom level.

Sorting. In our implementation, sorting is limited to simple point/dispersion value to summarise the metric by train. We are implementing sorting based on different points on the numerical distribution. For example, 80th percentiles are sometimes used for identifying plausibly likely “bad days”.

Historical data. Train lateness data is routinely collected and used to assess past performance. We can use these data in ChartTables, where real days replace model runs and where the distribution of lateness is based on what actually happened. Since only the effects of delays are recorded, there is a lack of detail on the *mechanism of the delays* afforded by our model. However, the Train Operating Companies we worked with found that visual analysis of historical data was valuable, albeit with fewer metrics.

3.2 Potential uses in other contexts

Interpreting probabilistic model outputs. Stochastic Monte-Carlo-style ABMs are in widespread use, generating large amounts of data, usually too much to consider in detail [6]. In most cases, these can be summarised in a case-by-variable structure. The “alternative case” point above illustrates the flexibility of summarising model outputs by different case-types and interpreting them within ChartTables. Since our four metric types are specific to our application, it is likely that additional mini-chart designs will be needed.

Other data and other mini-charts. Using other model outputs or datasets will likely necessitate additional mini-charts. Fortunately, the design space of such mini-charts is enormous [9] and can be based on existing visualisation idioms (as most of ours were) or customised designs. Whether variables and cases are rows or columns influences the effectiveness of the x - and y -axes. Where rows are cases, using x -axis for low LoD variant facilitates comparison across rows. For high LoD variants, using the y -axis facilitates comparison between model-runs within cases.

Other levels-of-detail. Although two LoDs were appropriate for our use, different conceptualisations of LoD will be appropriate in other contexts; for example, corresponding to levels of abstraction [2], aggregation and/or hierarchical data [5, 7, 8].

3.3 Use in the UK railway industry

The original purpose of our work with the UK railway industry was to develop and determine the feasibility of an ABM/visualisation approach to understanding reactionary delay. We expected the value to industry to be strategic, such as identifying general guidelines for timetable design. However, we found that its use more operationally was more important to the Train Operating Companies we worked with. They were interested in investigating specific timetable proposals and identifying problematic trains with reasons, and using these to tweak then test these modified timetables. Although much of the analytical work was been carried out by Risk Solutions, Zoomable LoD ChartTables play an important narrative role when presenting modelling results to clients.

Feedback has been positive. Simon Greenwood (Performance Manager, GWR) finds them “*really helpful, being able to see where reactionary delays were likely to be, and where they were coming from, delivered new insights and helped us see where to focus our attention*” and Marc Ware (Performance Manager, Greater Anglia) stating that “*it helped us work through a very complex system to understand the individual levers we have to use to deliver change*”. Actionable evidence is also valued, with Mark Walker (Performance Manager, ARL) citing the “*useful evidence to support our discussions with Network Rail about designing a better timetables*” and Marc Ware saying “*this is giving us the evidence we need... and focus on building a more robust timetable*”.

We are now licensing software that incorporates the ABM and visualisation, so some Train Operating Companies will now be using them analytically. Training sessions and the feedback this generates is informing further development.

4 CONCLUSION

Studying reactionary delay in trains is a good example of how visual analytics can help analysts consider nuance and variation in data that is normally averaged out. Tabular layouts are a simple, familiar and effective means to represent variables and cases. Mini-charts provides a huge range of possibilities for representing multivariate composite metrics and data distributions, including qualifications of uncertainty. Zoom-based level-of-detail variants of these mini-charts enabled us to consider distributions of summaries by case and distributions within cases of Monte-Carlo-style stochastic model outputs. There is plenty of scope to generalise to other application areas.

5 SUPPLEMENTAL MATERIAL INSTRUCTIONS

Supplementary materials are available at <https://osf.io/u2ykd/> including a video and a list of our train metrics. The computer code is not available as portions of it are owned commercially, however it is relatively easy to implement and the authors are happy provide further information.

ACKNOWLEDGMENTS

The authors wish to thank the Rail Safety and Standards Board (RSSB), Network Rail, Great Western Railways, Greater Anglia, the West Coast Partnership Development, Arriva Rail London and CrossRail for their funding and other support with this work.

REFERENCES

- [1] B. B. Bederson, J. D. Hollan, K. Perlin, J. Meyer, D. Bacon, and G. Furnas. Pad++: A zoomable graphical sketchpad for exploring alternate interface physics. *Journal of Visual Languages & Computing*, 7(1):3–32, 1996. 2
- [2] R. Beecham, C. Rooney, S. Meier, J. Dykes, A. Slingsby, C. Turkay, J. Wood, and B. W. Wong. Faceted views of varying emphasis (favves): a framework for visualising multi-perspective small multiples. In *Computer Graphics Forum*, vol. 35, pp. 241–249. Wiley Online Library, 2016. 2, 4
- [3] J. Bertin. *Semiology of graphics*. University of Wisconsin press, 1983. 2
- [4] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE transactions on visualization and computer graphics*, 13(6):1216–1223, 2007. 3
- [5] K. Furmanova, S. Gratzl, H. Stitz, T. Zichner, M. Jaresova, A. Lex, and M. Streit. Taggle: Combining overview and details in tabular data visualizations. *Information Visualization*, 19(2):114–136, 2020. 2, 4
- [6] S. Grainger, F. Mao, and W. Buytaert. Environmental Modelling & Software Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling and Software*, 85:299–318, 2016. doi: 10.1016/j.envsoft.2016.09.004 1, 4
- [7] G. Li, R. Li, Y. Feng, Y. Zhang, Y. Luo, and C. H. Liu. Coinsight: Visual storytelling for hierarchical tables with connected insights. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2, 4
- [8] G. Li, R. Li, Z. Wang, C. H. Liu, M. Lu, and G. Wang. Hitailor: Interactive transformation and visualization for hierarchical tabular data. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):139–148, 2022. 2, 4
- [9] T. Munzner. *Visualization analysis and design*. CRC press, 2014. 4
- [10] National Rail. How a delay to services in one area can affect trains elsewhere in the country. Technical report, 2019. 1
- [11] Network Rail. From monday 1 april 2019, the uk rail industry introduced a new, more detailed and precise set of measures to better understand how timely our trains are – from station to station, and from minute to minute. Technical report, 2024. 3
- [12] C. Perin, P. Dragicevic, and J.-D. Fekete. Revisiting bertin matrices: New interactions for crafting tabular visualizations. *IEEE transactions on visualization and computer graphics*, 20(12):2082–2091, 2014. 2
- [13] Rail Safety and Standards Board (RSSB). Call for research: Data sandbox: Improving network performance. Technical report, 2017. 1
- [14] S. F. Railsback and V. Grimm. *Agent-based and individual-based modeling: a practical introduction*. Princeton university press, 2011. 1
- [15] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 318–322, 1994. 2
- [16] A. Slingsby. Tilemaps for summarising multivariate geographical variation. *Workshop on Visual Summarization and Report Generation at VIS 2018*, 2018. 2
- [17] A. Slingsby, R. Reeve, and C. Harris. Gridded glyphmaps for supporting spatial covid-19 modelling. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 1–5. IEEE, 2023. 2
- [18] H. Wickham, H. Hofmann, C. Wickham, and D. Cook. Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics*, 23(5):382–393, 2012. 2