Narrative Visualization Essay

Xinhao Liu (xinhaol2)

Messaging

This narrative visualization aims to introduce viewers to the causes of railway derailments. It adopts an interactive slide show design schema, consisting of eight slides. Beyond the first title slide and the last reference slide, there are six main slides that gradually guide viewers from understanding the broad categories of derailment causes to comprehending these causes from both frequency and severity perspectives. By the end, viewers will have a general understanding of derailment causes and will specifically learn which derailment cause is the most concerning in the current North American railway system from both frequency and severity perspectives.

Narrative Structure

My narrative visualization follows the interactive slide show structure, guiding users through a series of slides that each focus on different aspects of railway derailment causes, ensuring a logical flow of information. Viewers can simply scroll up and down to progress to the next slide or return to the previous one, and the transition between each two slides is smoothed with animation. The use of a page navigation bar on the right-hand side of each slide also allows users to easily move between slides, enabling them to follow the narrative or skip to sections of interest. Slides with major plots also contains interactive visualizations, such as the callout tooltips on both scatter plots and bar charts and interactive legend on the side, allowing users to engage with the data and explore details as they progress through the story. The visual and textual elements are consistently styled across slides, creating a cohesive experience. Annotations and explanations are provided on the visualizations to highlight key points, making the narrative clear and informative.

Visual Structure

Each informative slide is designed with a railway background image, while the middle three slides contain interactive plots with a consistent light-colored background to minimize distractions and allow viewers to focus on the plot and its information. The transition between slides features smooth up and down flow animations as viewers scroll, enhancing the narrative

flow. The navigation bar on the right side of each slide indicates the current page by enlarging the corresponding icon, and it also shows how many slides remain, helping viewers navigate the presentation. This design also facilitates transitions between scenes, helping viewers see how the data in one scene connects to the data in the next.

In each scene, textual content is enclosed in semi-transparent boxes. For the three major scenes containing interactive visualizations, the title is consistently placed at the top, the main plot in the center, and the legend to the left. The size, layout, and font of each component of the charts are uniform across these major scenes. This consistent layout helps viewers quickly understand where to find explanatory text and key information, ensuring they focus on the important parts of the data. Additionally, the axis labels are bolded and in a larger font to draw attention to the information they provide, while annotation text within the plots is lighter gray and smaller to minimize distraction but still convey messages.

For the second and third plots, a secondary y-axis is included on the right in purple, showing the cumulative percentage. The data points for this secondary axis are also in purple, with the axis title and tick markers in the same color, ensuring consistency and making it easier for users to make connections.

Moreover, each main plot appears slightly skewed due to the inclusion of an outlier data point: the cause group "broken rails or welds," which remains the most frequent and severe derailment cause among all causes. This high frequency and severity skew the plots intentionally to emphasize that "broken rails or welds" is a critical issue the railway industry needs to address.

In summary, the visual structure ensures viewers can easily understand the data, navigate the scenes, and focus on the important aspects, while also facilitating smooth transitions between scenes to understand the data connections across the narrative.

Scenes

There are a total of eight slides (scenes) in my narrative visualization. The overall structure includes one cover page (slide 1), one introductory page (slide 2), five body slides (slides 3-7), and one reference page (slide 8).

Slide 1: Cover Page

This slide features the title and caption of the narrative visualization at the top, with a background image of a Canadian National Railway train driving in front of a mountain, signaling to viewers that the visualization is about railway operations.

Slide 2: Introductory Page

This slide includes explanatory paragraphs about the visualization in general, details about the dataset used, the template scene, and descriptions of the annotations, parameters, and triggers used in the narrative visualization.

Slide 3: Introduction to Derailment Categories

This scene introduces the five major derailment categories to viewers. Each category is described with explanatory text on the left and a relevant image on the right. A consistent color code is introduced for each cause category: track (red), signal (black), human factor (blue), equipment (orange), and miscellaneous (magenta pink). From this slide onward, any mention of a derailment sub-cause will be highlighted in its corresponding category color to reinforce viewers' understanding.

Slide 4: Cause Frequency vs. Cause Severity Plot

This is the first interactive visualization slide. It guides viewers through understanding derailment causes by comparing frequency and severity. The x-axis represents the frequency (number of accidents), and the y-axis represents the severity (average number of cars derailed). Each cause is color-coded by category, with a legend on the left. Viewers can select or deselect categories by clicking on the legend. By default, all categories are selected. Deselected categories turn light gray. The plot is divided into four quadrants by average severity and frequency lines, with light gray annotations indicating the properties of each quadrant. Hovering over data points displays detailed derailment information.

Slide 5: Total Number of Cars Derailed by Cause

This slide contains the second interactive visualization. It focuses on the total number of cars derailed for each cause group. The structure is similar to the previous slide, but the x-axis now lists cause groups ranked by severity. The primary y-axis shows the total number of cars derailed, while a secondary y-axis on the right shows cumulative percentage. A draggable callout box highlights that "Broken Rails or Welds" is the most severe cause, derailing 20% of all cars.

Slide 6: Total Monetary Damage by Cause

The third interactive visualization slide introduces total monetary damage as another way to measure severity. The plot structure remains the same, with the primary y-axis showing total damage cost in millions of dollars, and causes ranked accordingly on the x-axis. This slide reinforces that "Broken Rails or Welds" is the most frequent and severe cause, while also demonstrating how severity rankings can change based on different metrics.

Slide 7: Summary and Conclusions

This slide provides a summary and conclusions from the previous slides in a findings box and a reminder box, serving as the final body slide.

Slide 8: References

The final slide contains references and resources used to prepare the narrative visualization.

Annotations

Firstly, a consistent color code is introduced in slide 3 for each cause category: track (red), signal (black), human factor (blue), equipment (orange), and miscellaneous (magenta pink). From this slide onward, any mention of a derailment sub-cause is highlighted in its corresponding category color to reinforce viewers' understanding, including in the legend box of each interactive chart.

Secondly, in slide 4, several annotations are marked on the chart. The average severity and average frequency lines are added to separate the plot into four quadrants, and each quadrant has a light gray annotation indicating its properties. For example, the upper-right quadrant means any cause groups in this area will be "more frequent and more severe," and the lower-left quadrant means any cause groups in this area will be "less frequent and less severe."

Lastly, in slides 5 and 6, a permanent and draggable callout box is located to the right of the bar for "Broken Rails or Welds" in the plot to remind viewers that this is the most severe derailment cause among all. It derails 20% of all cars and causes 19.3% of all monetary damage caused by all derailment causes.

These templates and annotations are used to highlight key data points and trends, making the narrative clear and informative for viewers. The annotations do not change within a single scene but are strategically placed to emphasize the most important information, ensuring that viewers understand the severity and frequency of each derailment cause.

Parameters

The parameters used in the narrative visualization are the derailment cause categories, which are built into the interactive legends. There are five categories: track (red), signal (black), human factor (blue), equipment (orange), and miscellaneous (magenta pink).

For the scenes that include interactive legends (the three body slides with interactive charts), the initial state is that all categories are selected, and this is because I want the viewers to learn the overall picture at first and also the chart is not overly crowded when all data points are highlighted. Users can change the state of the scene by selecting or deselecting the categories they want to see and focus. Deselected categories will turn light gray in both the legend label and the data points in the chart. This functionality is consistent across all three interactive plots.

These parameters and states allow users to filter and explore the data interactively, enhancing their understanding of the severity and frequency of each derailment cause. The interactive legends help define the state of each scene by dynamically updating the visualization based on user selections.

Triggers

Various triggers in the narrative visualization connect user actions to changes of state in the visualization.

Viewers can hover their mouse over the data points on the scatterplot to display detailed information about a specific cause group, including the exact frequency and severity of the cause and the category it belongs to. Tooltips are also available for the bar charts. By hovering over any part of a bar, users can see derailment information for that cause group. For the cumulative percentage curve, the tooltip contains the name of the cause group and the cumulative percentage, along with a dashed line connecting the point to the cause name on the x-axis. When the tooltip for a data point or bar is activated, the bounding box of that data point or bar will be highlighted, making it thicker.

The interactive legend serves as another trigger. Viewers can click on it to highlight or unhighlight the data points belonging to a category. Affordances are provided to make user interactions easier. When users hover their mouse over any part of the cause category text in the legend, the cursor changes to a hand shape, and the text color changes to purple. Additionally, there is a checkbox in front of each category to convey that it is clickable. When a category is selected, a checkmark appears in the box, and it disappears when the category is deselected.

Lastly, a callout box appears when the viewer hovers the mouse over the dots inside the navigation sidebar, displaying the title of each page.