MCSDS – CS498 – Data Visualization Assignment
Narrative visualization on the Caltrain's north bound ridership
Data – 08/02/2019
Author – Manas Mukherjee

This narrative visualization uses the Interactive Slideshow approach to visualize how the number Caltrain commuters varies across different time-segments of any given weekday. It consists of multiple scenes where each scene has a bar chart that shows 'number of average daily commuters' who commute from different 'South-Bay'/'Peninsula' stations to the San Francisco city (4th and King street). End-user can hover their mouse on any bar (related to a station) and explore (from the tooltip) more contextual information like 'zone information', 'avg. number of commuters', "distance from San Francisco" etc. This visualization can help to answer some of the questions like -

1. Which all zones and time-segment are important (w.r.t # of passengers, importance of the public transport etc.)?
2. Can public(cab) transport companies effectively leverage this data to take the passengers from SF station to their destination?
3. What is the preferred time for someone who wants to avoid the rush hours?
4. Can the train-authority consider arranging more trains (or galloping local train) to help the passengers? etc.
Data source
The dataset is available in the Caltrain's' official website.

Reference - 2016 Annual Passenger Counts by Train - Weekdays
For visualization purpose, a master file is created (reference - 4thKingDestination.csv) by joining all the tabs available in the original data file. It contains time-segment wise ridership data between each suburb station and San Francisco.
Scenes
This visualization uses the Interactive Slideshow structure to visualize how the number Caltrain commuters varies across different time-segments of any given weekday. Here, each time-segment (like 'early morning', 'mid-day' etc.) is considered as a scene. In each scene, the bar-chart shows the 'number of average daily commuters' who commute from different 'South-Bay'/'Peninsula' stations to San Francisco city(4th and King street). In each scene, the X-axis shows the list of stations from Gilroy to San Francisco via San Jose. Since the commuter-numbers vary significantly throughout the day, the Y-axis is plotted in the logarithmic scale(base=2).

Annotations
In each scene, a 'red-colored line' indicates the average (across all stations) number of passengers going from each South Bay/Peninsula station to San Francisco. Annotations are used whenever the number of commuters of a specific station or zone is significantly higher or lower compared to the 'average number passengers across all stations'(red ref line). For example - Zone 3(Sunnyvale to Menlo Park) remains very busy compared to the other South bay stations/zones.

Parameters
Different time-segments of an average weekday are treated as a parameter of this narrative visualization. Based on the selected parameter value (for ex - 'Mid-day (12-3 pm'), the high-level as well the drill-content of a given scene gets changed.

Triggers
There are five buttons just below the chart of each slide. These buttons are triggers which connect user actions to changes of state in the narrative visualization. The button values like EA, AM, MD etc. are parameters which are used to determine the customized content of each scene. Each button shows a different look and feel for their active (the color gets darker) and inactive state. This affordance helps the end-user to explore the options in the right way.