MC1: Entry and Exit Data

Project Title: Traffic Impact on Rose-Crested Blue Pipit Nesting Numbers
Team: Pippa Hodgkins (priscilla.hodgkins@tufts.edu), Yinan Qian (yqian03@tufts.edu), Andreas Moe (andreas.moe@tufts.edu)

Abstract: The goals of this project are to determine if changes (or the lack thereof) in traffic patterns in the Lekagul Nature Preserve over a 1-year timeframe have impacted the nesting habits of the Rose-Crested Blue Pipits, specifically the decline in nesting rates. The system should be able to take the traffic sensor data and provide a visualization of the traffic patterns over the timeframe. Ideally, the system will use a map of the preserve to visualize the data and focus on the changes in traffic patterns of specific vehicle and gate categories. The visualization should be able to provide an easy interpretation of the traffic patterns of the park using colored heatmaps at each sensor to easily distinguish between vehicle types. In addition, the visualization should be able to filter timeframes and track progression in order to identify traffic patterns.

Goals:

- The overarching goal of this project is to determine if traffic within the Boonsong Lekagul Nature Preserve has any effect on the decreasing number of Rose-Crested Blue Pipit nesting pairs.
    - Could depend on a variety of traffic pattern changes such as increased truck traffic, increased unauthorized access to restricted areas, or more overnight campers arriving at the park.
    - It is up to us to use the results from both our data preprocessing/analysis and our visualization system to determine a comprehensive result.
- We need to build a comprehensive visualization system to adequately display our results
    - This system should display the dataset in a way that reflects what the data is trying to model in the real world. It should also enable us to interact with the data and make it easier to glean any insight on the problem at large.
    - Our visualization should provide a clear interpretation of our data and results that do not rely on previous understanding of the data set.
    - The system will likely depend on JavaScript and the d3 library for the majority of the visualizations.
    - Ideally, we would like to visualize the data using density plots on the map of the park. Since we do not know how realistic this is to build, we will first use histograms to visualize the data.
- We also need to build a data pre-processing system to filter and evaluate the given data in a meaningful way.
    - We will likely filter the data based on vehicle and gate types in order to investigate subsets of the data.

    - We will also need to determine if any data cleaning is necessary; i.e. missing months of data or sensors with extremely limited data
    - The exact technology that we will use is not clear yet. A possible tool to handle and filter the data could be the Pandas library in Python.
- Interact with our system to test our hypotheses.
    - Once we have a system that allows us to visualize and interact with the data, we need to filter the data in accordance with our hypotheses.
    - Our data filtering should help to provide insight into the problem, not muddle the bigger picture by unnecessarily segmenting the data.
    - The interactions should be intuitive and easy-to-use, enhancing the results
- Create a professional presentation of our results and building process for this project
    - Keeping track of changes that our system goes through over the course of this project is an important step in the development of the system
    - Our presentation should be unambiguous in its results and carefully provide interpretations of our findings based solely on the data and our visualization system

Data Description: The provided data for this project is a supplied .csv file containing the sensor data for vehicles at 5 locations in and around the park: entrances, general-gates, gates, ranger-stops, and camping. For each vehicle which passes through a sensor, the vehicle's category, id, gate-name, time, and date are recorded. Vehicles can be 1 of 6 categories: 2 axle cars (or motorcycle), 2 axle trucks, 3 axle trucks, 4+ axle trucks, 2 axle buses, and 3 axle buses. Park service vehicles are denoted with the letter "P," though all park service vehicles fall into the 2 axle truck category. The sensors in the park do not observe daylight savings time, and there are certain locations that only vehicles tagged as park service vehicles can pass through. The provided data also includes a map of the park which is 12 miles by 12 miles, with the roadways indicated by white pixels and each sensor indicated by a specific color depending on what type of sensor it is (green for entrances, blue for general-gates, red for gates, yellow for ranger-stops, and orange for camping). The data was collected over a 1-year span between May 1 2015 and May 31 2016. Other relevant information about the dataset is that traffic either passes through the preserve, stays as day campers, or stays as extended campers.

3 Hypotheses:

1. The increase in traffic, especially larger trucks and buses, is driving nesting numbers down.
2. Non-park vehicles getting past gates and to unauthorized areas are driving nesting numbers down.
3. The increase in overnight campers is driving nesting numbers down.

Timeline:

- Data pre-processing (basic statistics run) (April 5th)
- Preliminary HTML page with dataset upload and preliminary data filtering (April 5th)
- Early bar chart visualization (April 8th)
    - No timeline yet, just filtered statistics over full timeframe
- Add timeline (April 15th)
- Implement interactive map in place of preliminary bar chart (May 5th)
- Final analysis of results (May 11th)

Feature List:

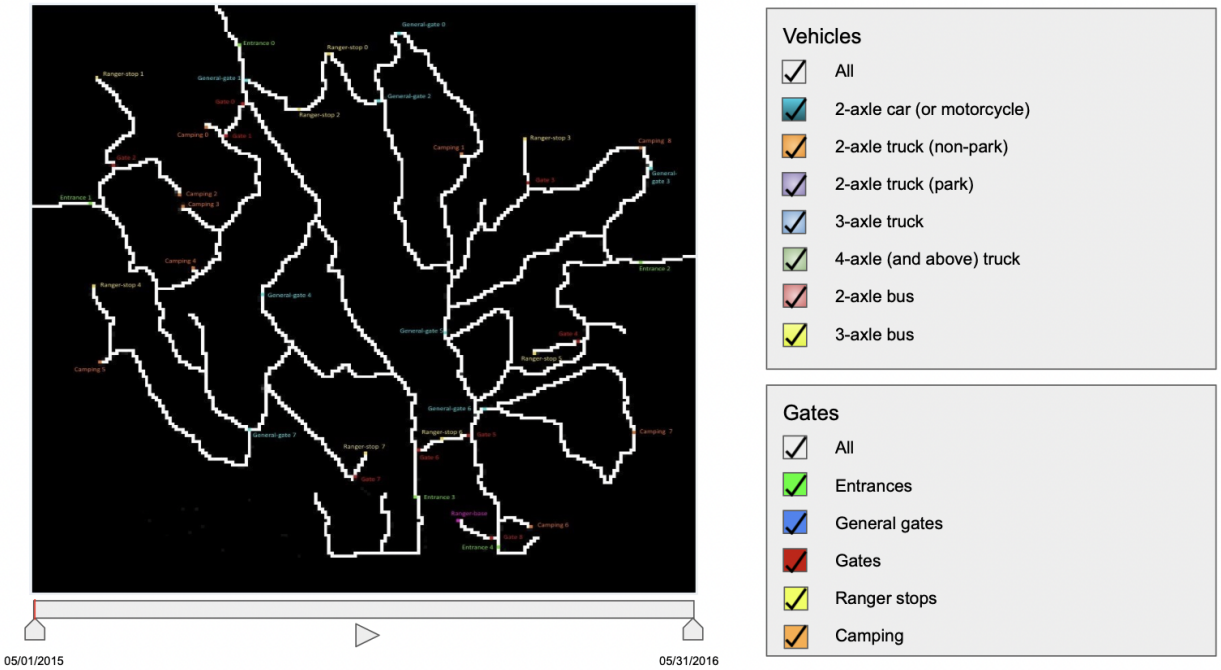
- Interactive map of the preserve based on the provided map in the dataset
    - Serves as a traffic heatmap: circles at each sensor location with width determined by traffic density (nice-to-have)
        - If this seems unfeasible, use histograms or other appropriate charts to display the filtered data.
    - Play button to cycle through the selected timeline. (must-have)
    - Filter by vehicle type- can view the heatmap for 2 axle cars or 3 axle buses depending on selection (must-have) (differentiate vehicle type by color)
    - Filter by time-of-day (must-have)
    - Also filter by sensor type- can only view camping sensors or general-gates depending on preference (must-have)

Team Roles:

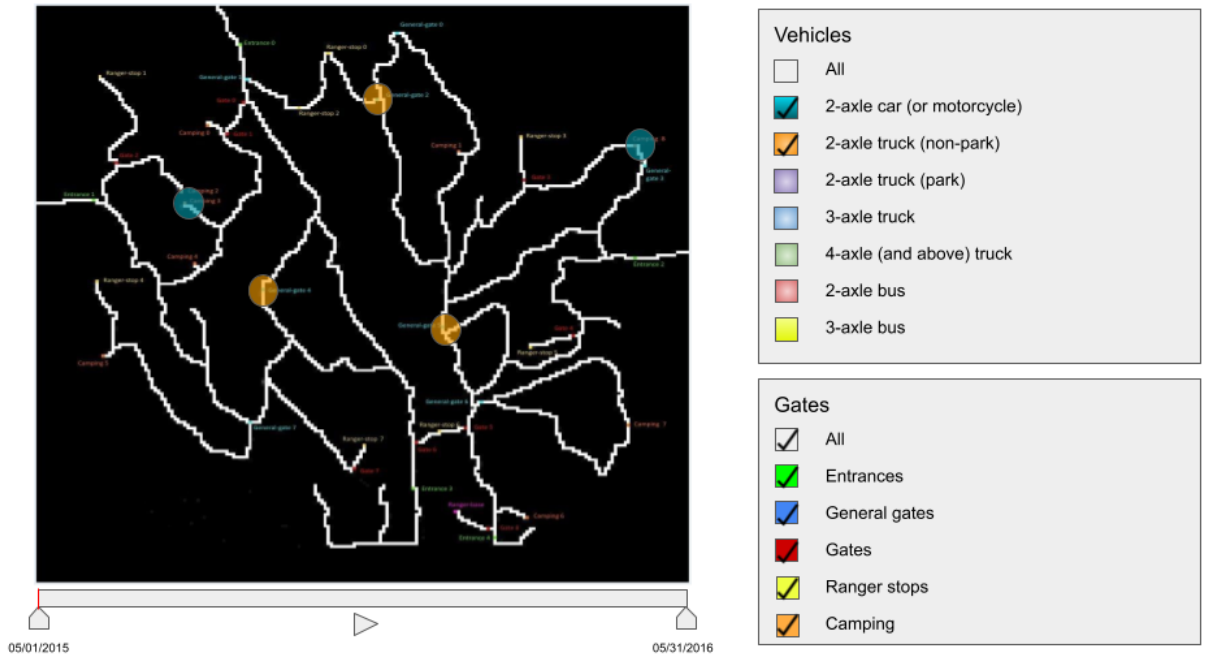
- We plan on splitting the work equally amongst ourselves and working together on the whole project.

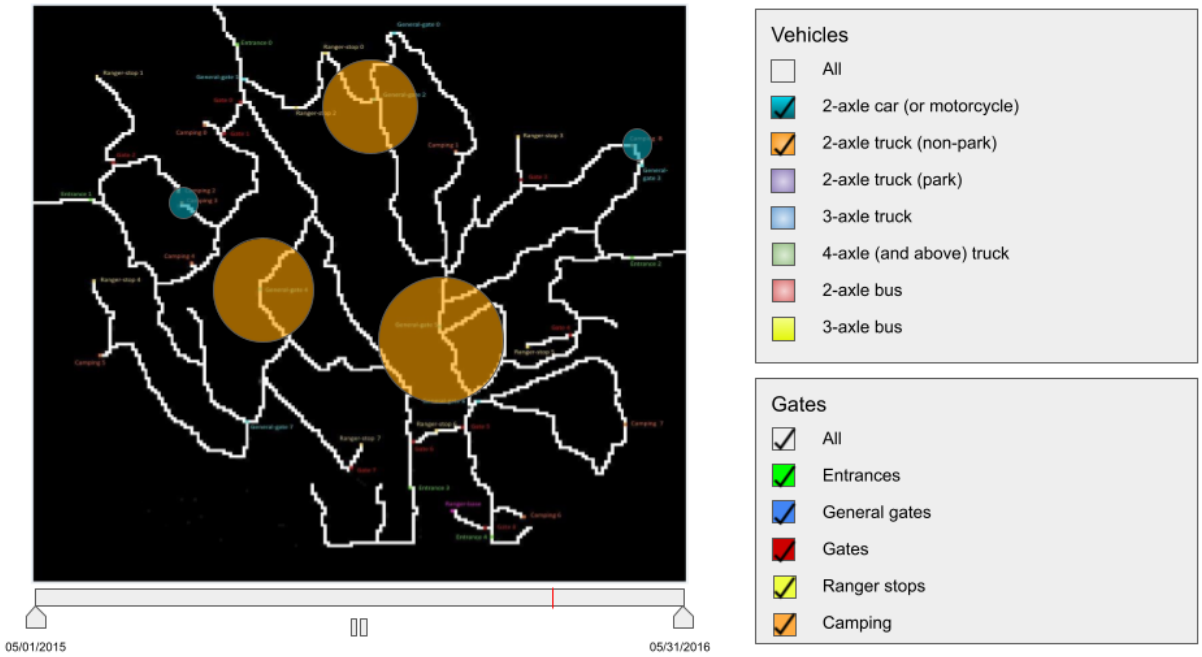
Storyboards:

- Basic Data Visualization

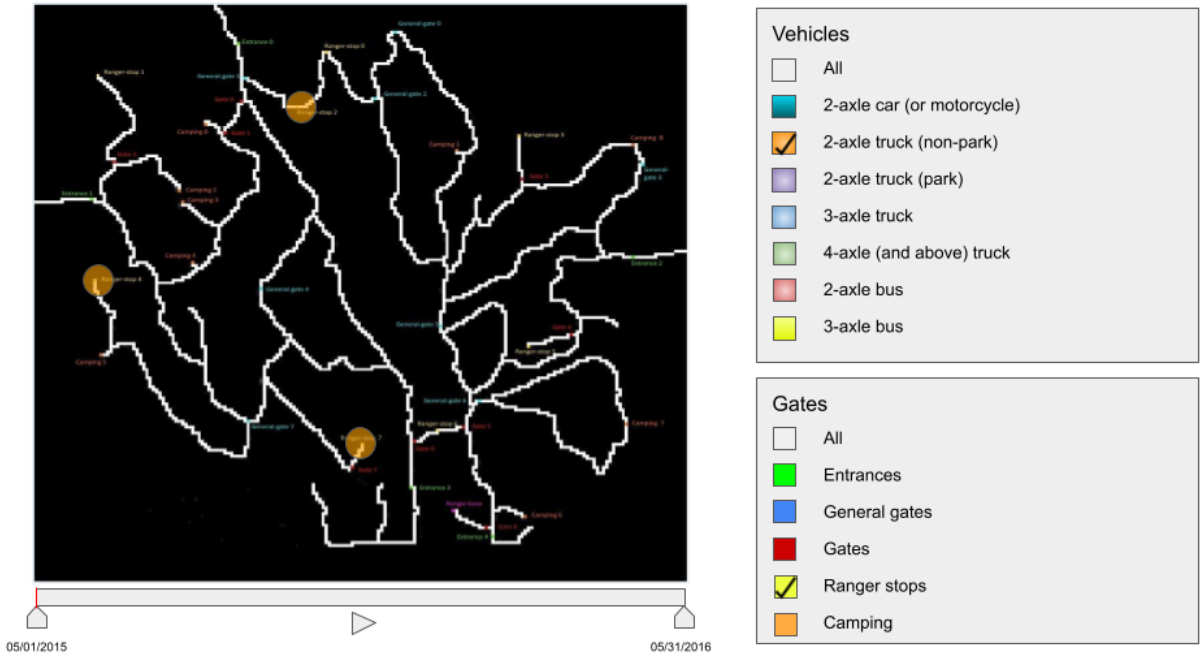


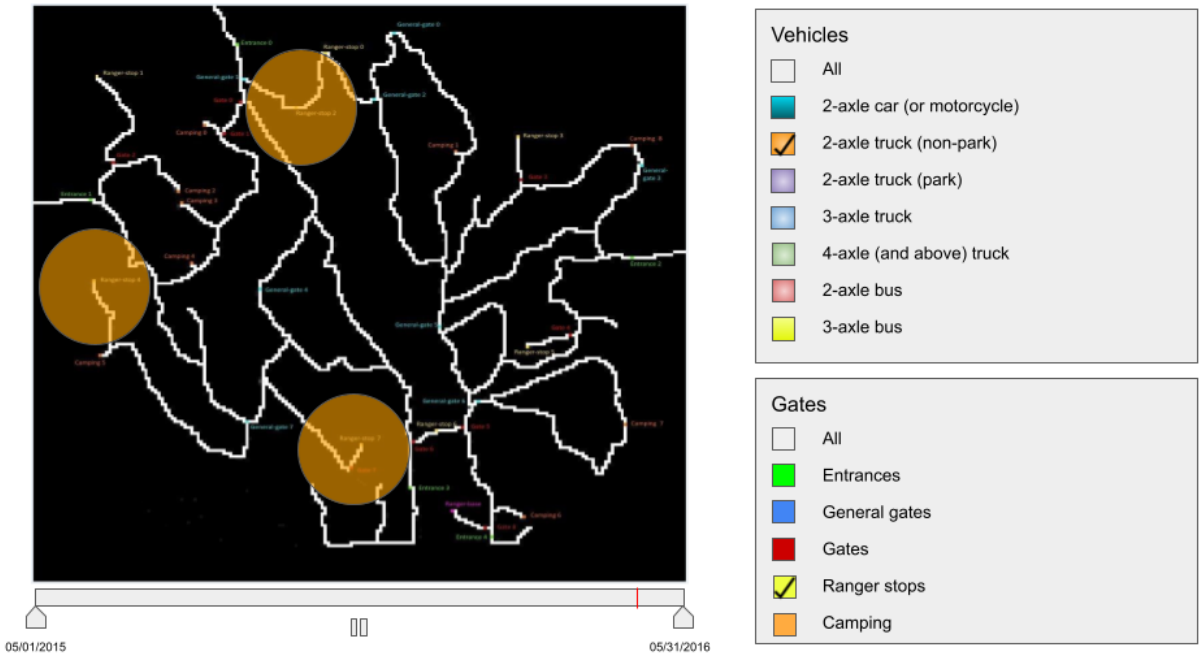
- Using the system to explore Hypothesis 1:



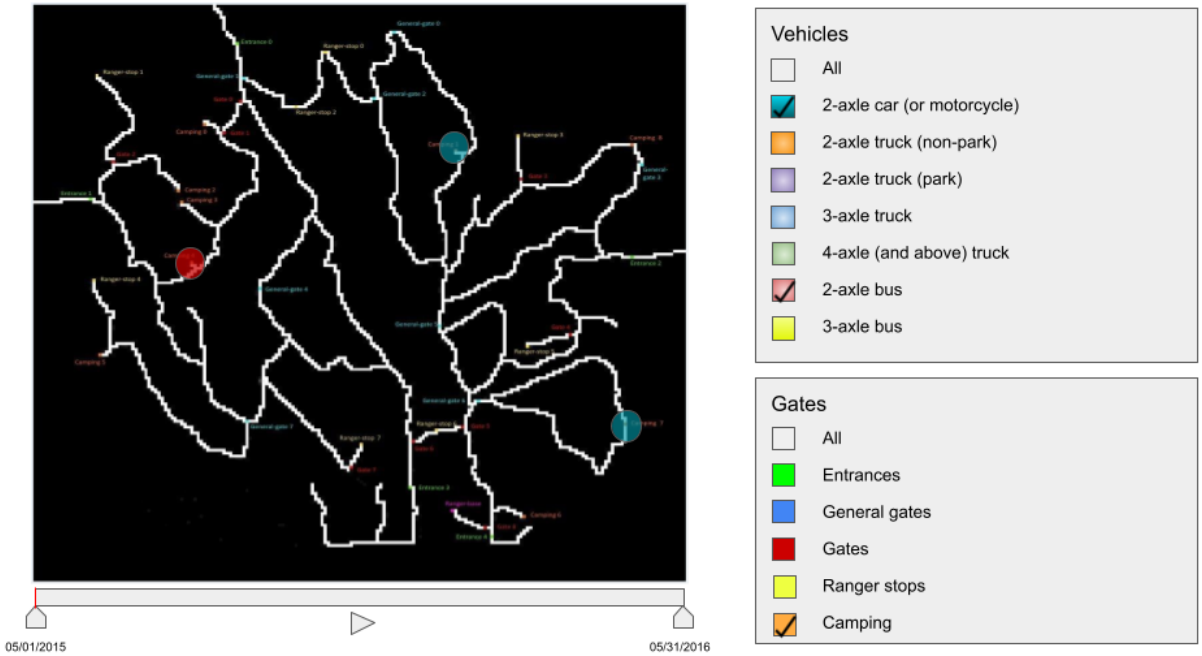


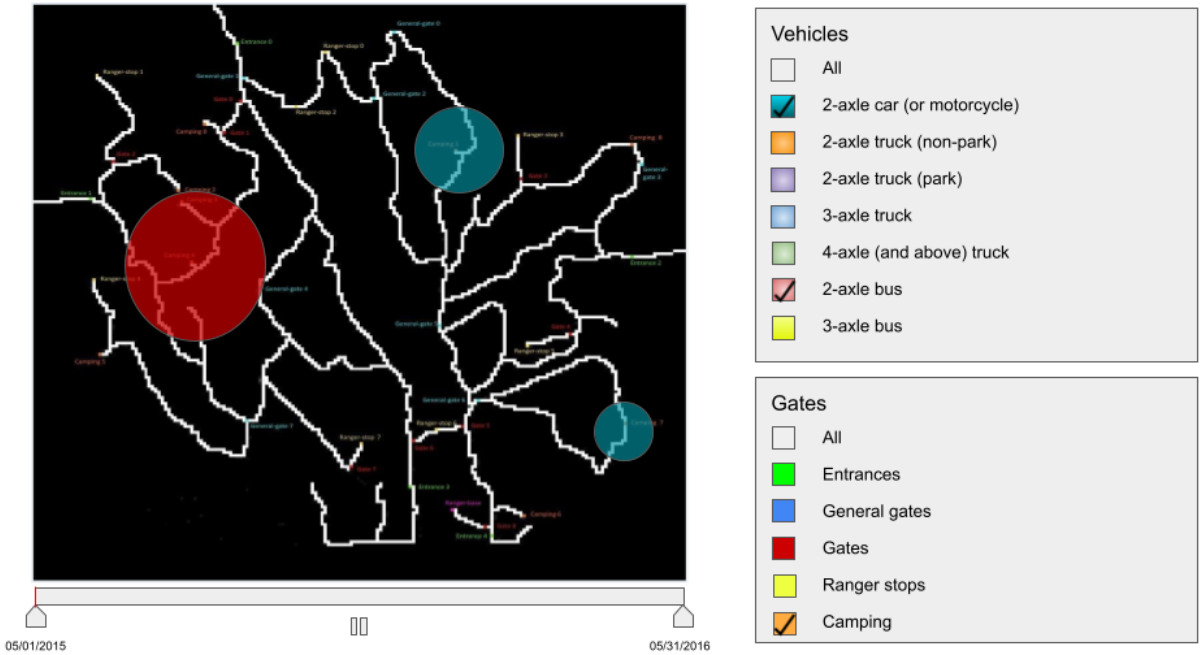
- Using the system to explore Hypothesis 2:



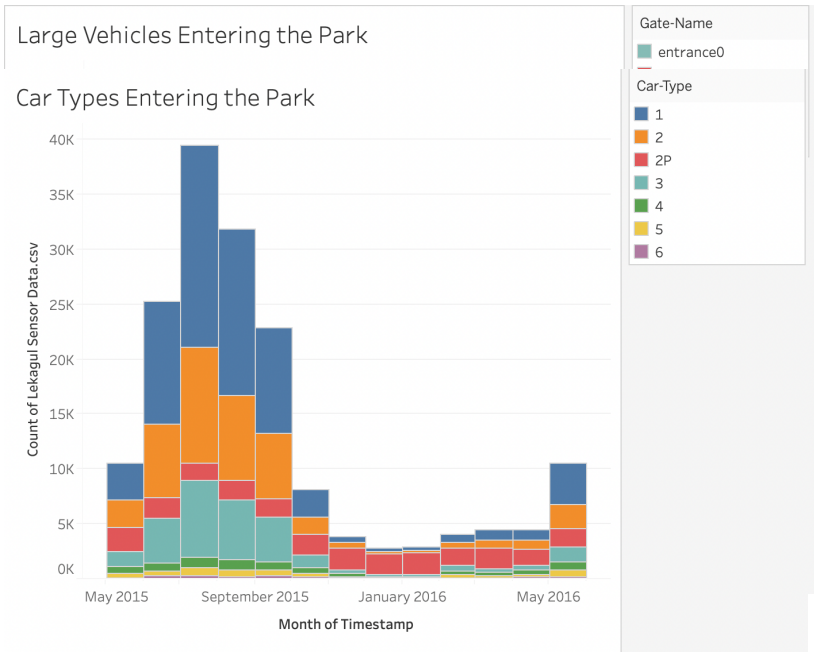


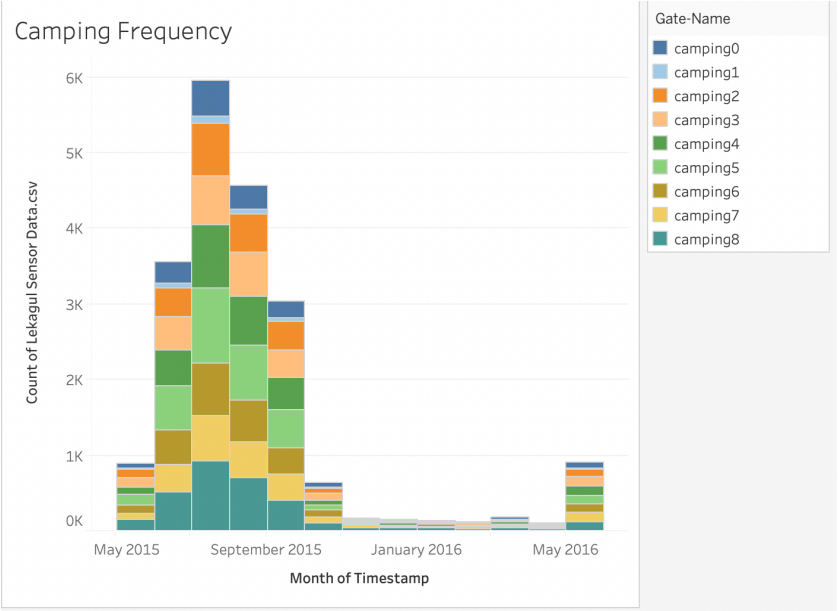
- Using the system to explore Hypothesis 3:

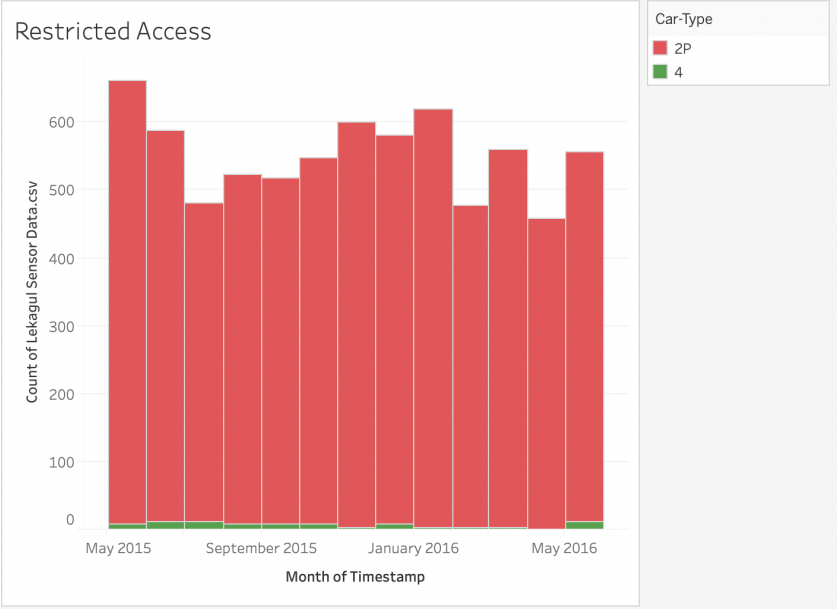




Preliminary Visualizations:







Updated Timeframe:

- Early bar chart visualization (April 25th)
    - No timeline yet, just filtered statistics over full timeframe
- Add timeline (April 30th)
- Implement interactive map in place of preliminary bar chart (May 5th)
- Final analysis of results (May 11th)

Process:

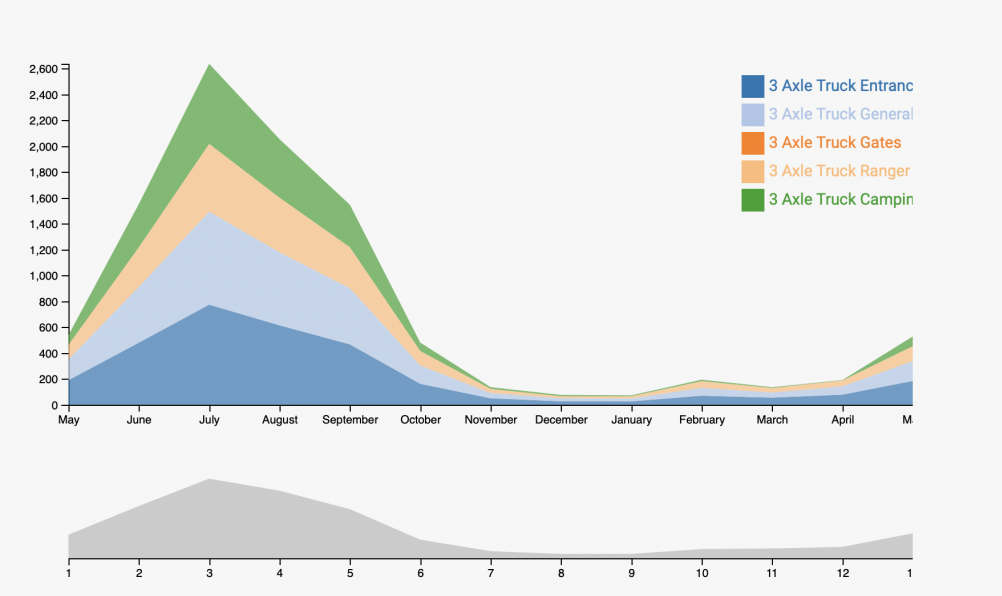
- Originally all data was kept in the same csv with just some added coordinates for gates, but the volume was too high
- Changed to monthly aggregate totals - much easier to manage
- Separated specific gate data from the monthly data and the car type data
    - For gates for the bubble map, it was much easier for the gates to be the rows and the monthly totals to be the columns, with columns for the coordinates of gate x and gate y
- Also decided to be able to filter car type by data as well as our timeline
- Made the timeline over months for readability purposes
- Decided that a dual-visualization system was the best course of action
- Made a mistake and aggregated data by count distinct car id, rather than just count car id, which then didn't count any cars which returned to the park or passed the same sensor multiple times
    - Severely limited the patterns in the data if each car is only pinged once; therefore we were only counting the number of individual cars the visited in a month, rather than the total traffic recorded by each sensor
- Originally uploaded a .csv file to the html page and then filtered from there, but that caused severe problems so we switched to hard-coded json files
    - Had issues with trying to load local files only

Design Revisions:

- Originally only going to have a bubble map overlaid over the map of the park, but realized that comparing bubble sizes is difficult for the human eye to distinguish when the bubble radius ranges from 0 to 5,000
- Added a stacked area chart to demonstrate traffic patterns over time, with a timeline for easy scrolling
    - Originally were going to have a 1-way timeline that would progress over each month, but realized that it was better for the user to have control over the visualization rather than just letting it play out of the user's control
    - Can also go backwards in time and not just forwards without resetting
- Was not able to figure out how to get a custom background on an SVG element from a local file, and could not upload the map to somewhere where the html element could access it
- Changed from buttons for cars and gates to buttons for gates and a drop-down menu for cars, so car type became select-one rather than select all
- Tried to filter the bubble plot by gates but couldn't find a way to skip only certain rows in the json file when building the visualization

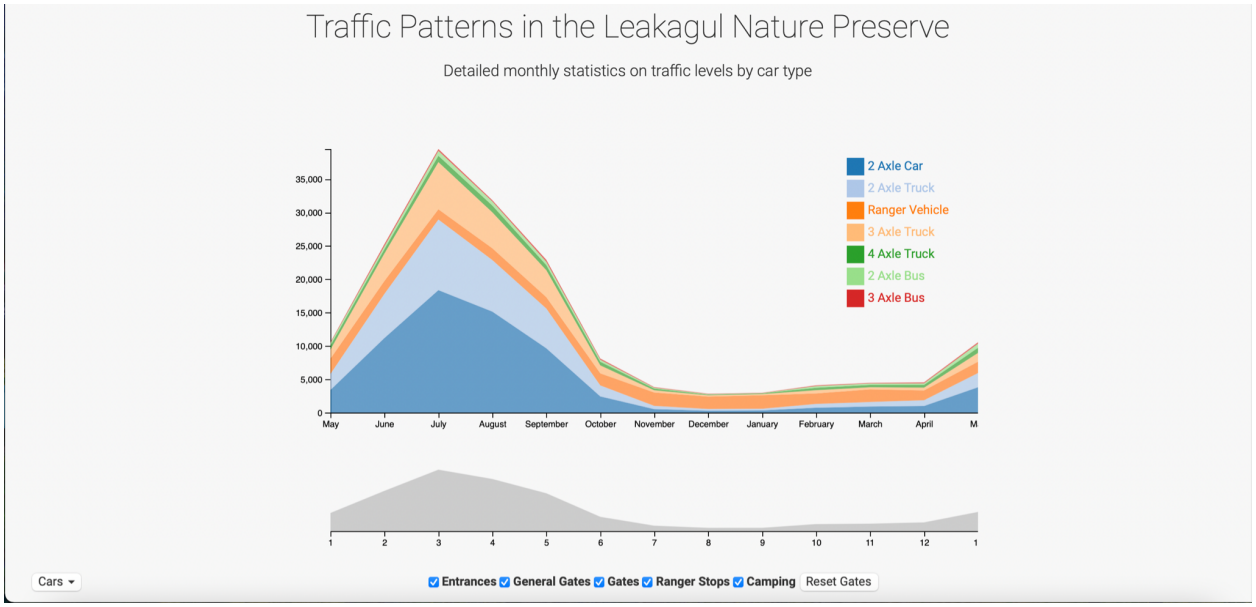
Visual Process

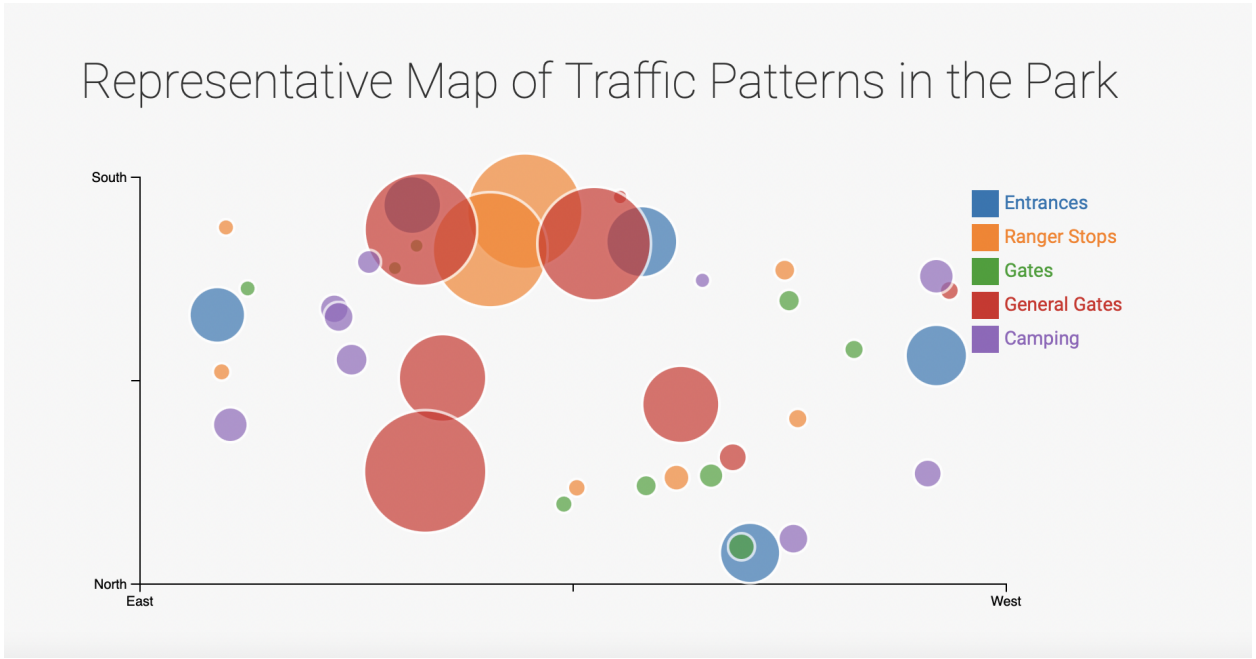
- Realized that the key to this problem was change over time, rather than specific daily patterns
    - It was most important to have a filter over time so we could observe changes
- Also, after implementing the bubble map in conjunction with the timeline, realized that there were no significant geographical changes in traffic over time (traffic didn't increase or decrease in a specific area of the park) so we focused on traffic patterns based on car type instead
- Also added legends to the visualizations rather than tool-tips that appear only on hover because it provided a quick visual key of what's being compared, as the stacked area chart switches from stacking by car type to stacking by gate type, so it's important to have a legend to distinguish what's being stacked
    - Ex // filtering only by 3 axle trucks
    -



- Made sure that the colorScale for the stacked area chart was different from the bubble plot so that there wouldn't be any accidental correlation between unrelated items

Final Visualization:

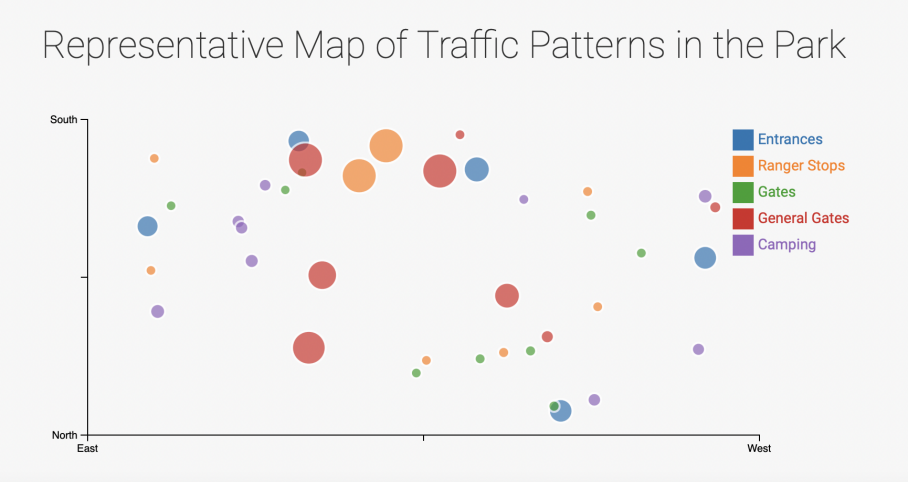




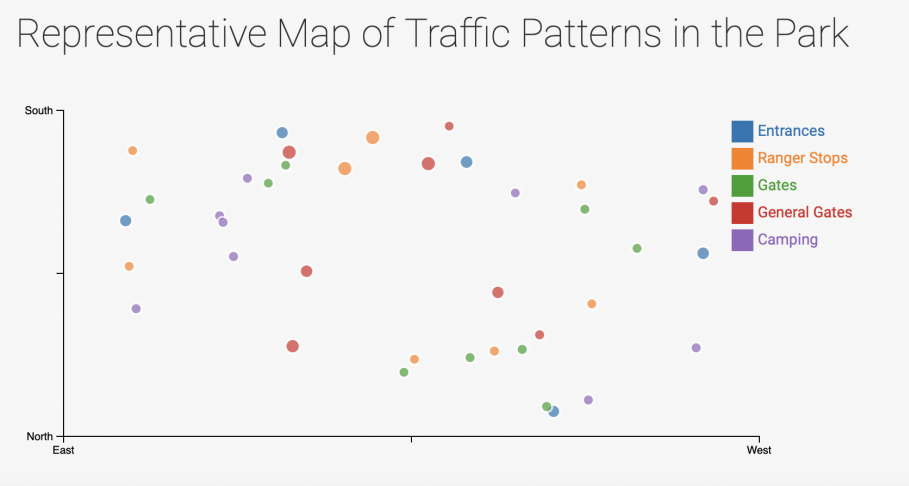
(Split into two screenshots because you have to scroll down)

Findings:

- There appears to be no significant geographical pattern change over the timeframe specified - cars are not traversing areas of the park they hadn't before
    - An original hypothesis was that cars were accessing a part of the park they hadn't before, but as seen here the May/June 2015 heatmap and the April/May 2016 heatmap follow the same geographic distribution:

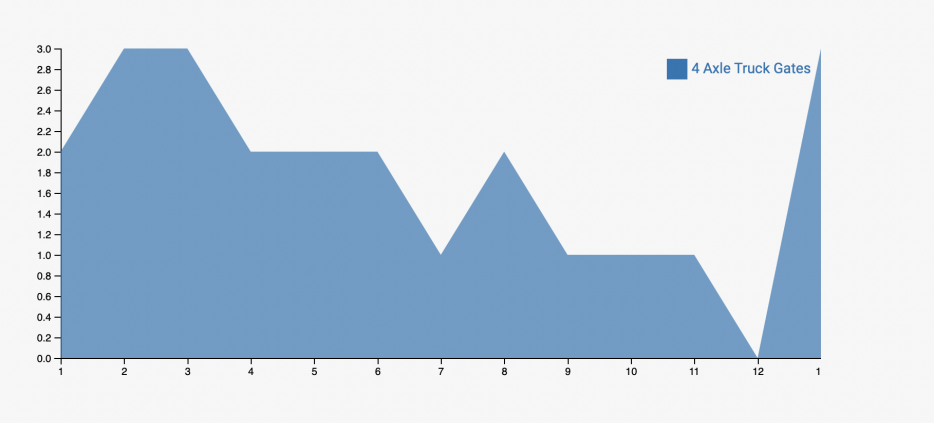


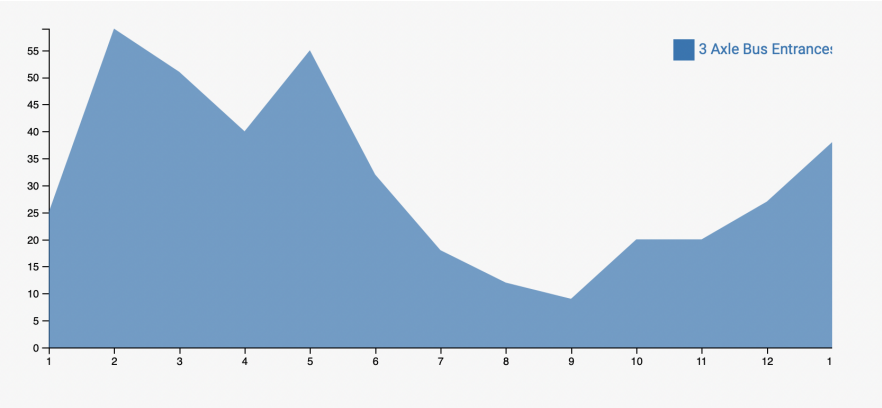
- (May/June 2015)

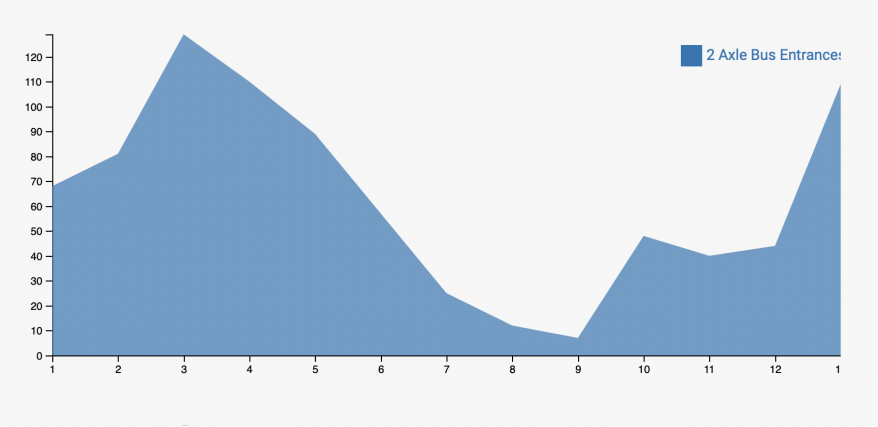


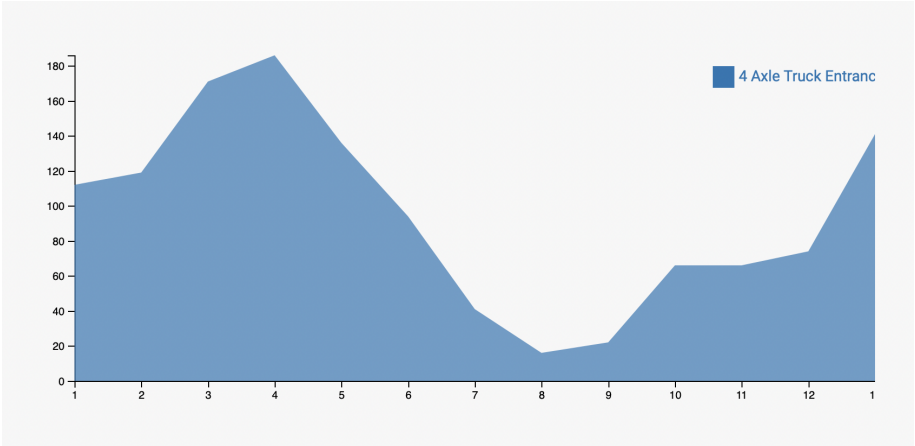
- (April/May 2016)

- Another original hypothesis was that some vehicles were accessing areas of the park they did not have access to (like Gates, which should only be accessed by ranger vehicles). We found that some 4 axle trucks were accessing the restricted areas; however the number that did so over the course of an entire month was too small to be significant in terms of the traffic seen at those gates as a whole

- 

- However, we did discover that the traffic of 4 axle trucks, 2 axle buses, and 3 axle buses had significantly increased entering the park in May of 2016 compared to their May of 2015 levels, suggesting a large change in traffic pattern that could impact the nesting numbers:
    - 

    - 

- 

Answer Sheet:

Entry Name: **"Comp152-AMYQPH-MC1"**

VAST Challenge 2017

Mini-Challenge 1

Team Members:

Student Team: *Please enter YES or NO*
Pippa Hodgkins (priscilla.hodgkins@tufts.edu), Yinan Qian (yqian03@tufts.edu), Andreas Moe (andreas.moe@tufts.edu)

Tools Used:
Python (data preprocessing)
Tableau (preliminary explorations)
D3 (visualization building)
HTML (visualization building)

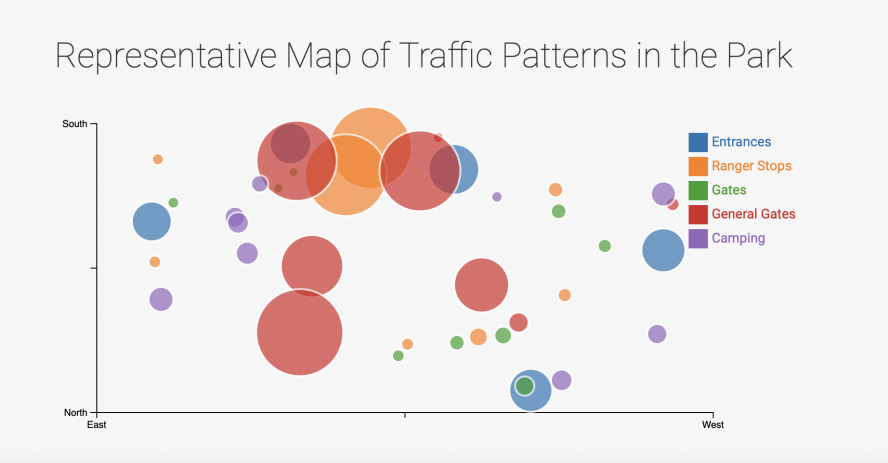
Approximately how many hours were spent working on this submission in total?
About 75 Hours

Questions

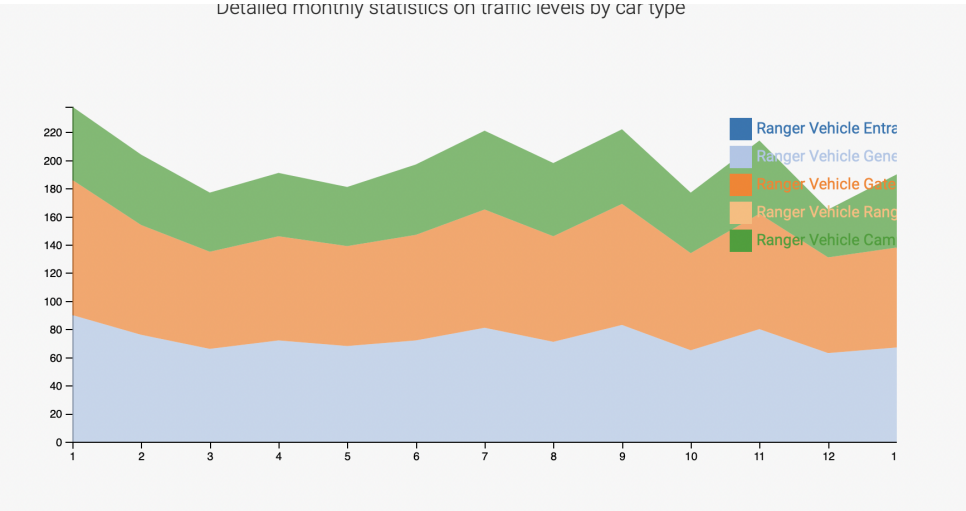
1 – "Patterns of Life" analyses depend on recognizing repeating patterns of activities by individuals or groups. Describe up to six daily patterns of life by vehicles traveling through and within the park.

Characterize the patterns by describing the kinds of vehicles participating, their spatial activities (where do they go?), their temporal activities (when does the pattern happen?), and provide a hypothesis of what the pattern represents (for example, if I drove to a coffee house every morning, but did not stay for long, you might hypothesize I'm getting coffee "to-go").

1) The northern half of the park is much more popular than the southern part of the park (more cars enter through north entrances than south entrances every day, and the general gates and ranger stops in the north are far more trafficked, suggesting rangers also spend most of their time in the north)

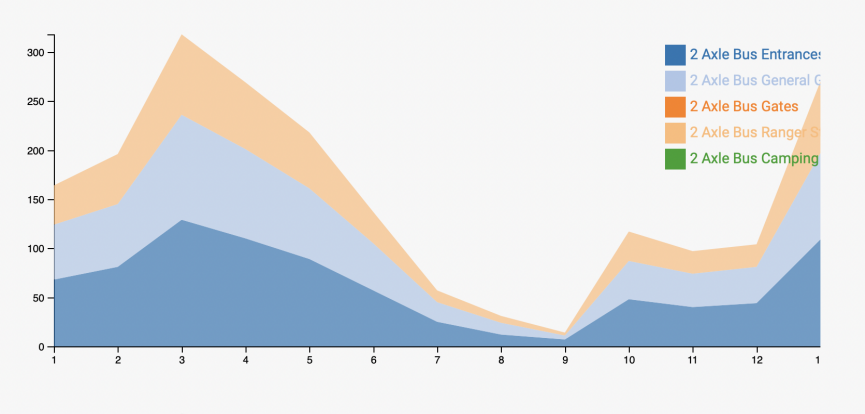


2) Ranger vehicles never leave the park - they are never seen on entrances

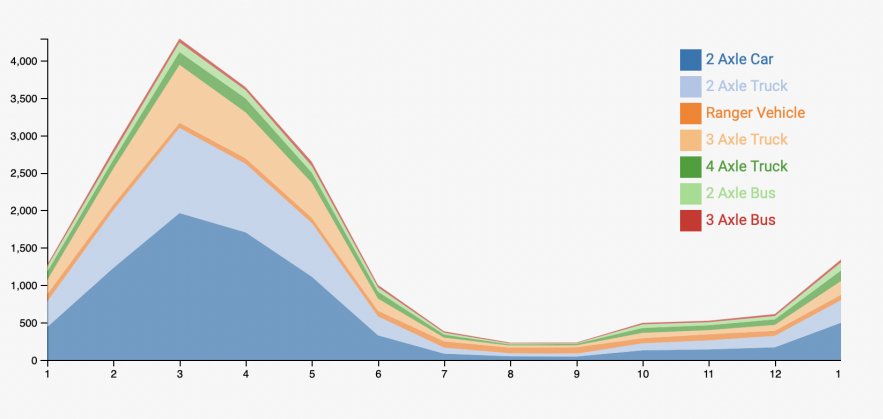


3) Certain gates, such as Gate 2, lead to dead ends and are therefore trafficked twice by every car which passes them

4) Camping gates are rarely, if ever, traversed by buses



5) The most popular type of car seen in the park are 2 axle cars, followed by 2 axle trucks



2 – Patterns of Life analyses may also depend on understanding what patterns appear over longer periods of time (in this case, over multiple days). Describe up to six patterns of life that occur over multiple days (including across the entire data set) by vehicles traveling through and within the park. Characterize the patterns by describing the kinds of vehicles participating, their spatial activities (where do they go?), their temporal activities (when does the pattern happen?), and provide a hypothesis of what the pattern represents (for example, many vehicles showing up at the same location each Saturday at the same time may suggest some activity occurring there each Saturday).

1) The winter months are far less popular that the summer months in the park
2) Ranger vehicle traffic remains consistent over time, however all other vehicle types decrease
3) Large vehicles (3+ axles) appear much more infrequently in the park that 2 axle vehicles
4) However, large vehicle traffic does not decrease in the winter months as much as 2 axle vehicles
5) Ranger vehicles appear to stay in the more popular areas of the park

3 – Unusual patterns may be patterns of activity that changes from an established pattern, or are just difficult to explain from what you know of a situation. Describe up to six unusual patterns (either single day or multiple days) and highlight why you find them unusual.

1) There appears to be a car which entered the park but never left it
2) 4 axle trucks sometimes access ranger vehicle only gates
3) Camping gate 1 is apparently very unpopular with the people
4) Likewise, ranger stop 1 is also not very trafficked

4 — What are the top 3 patterns you discovered that you suspect could be most impactful to bird life in the nature preserve?

1) The increase in large vehicles in the park over the timeframe
2) The large vehicles accessing ranger vehicle only gates
3) The popularity of the northern half of the park