

Using Visual Analytics Methods to Identify Causes of the Rose-crested Blue Pipit's Decline in Nesting

Created on: 2021-03-25
Updated on: 2021-03-25

Abstract

This project is our group's solution to the VAST Challenge 2017 Mini-Challenge 1. The fictional case-study involves an ontology student, Mitch Vogel, who is tasked with analyzing traffic data in the Boonsong Lekagul Nature Preserve to identify the cause of a decrease in nesting of the Rose-crested Blue Pipit. The challenge requires us to create a data visualization so that we can attempt to solve this problem. Along with documentation material and instructions, we are provided with the following materials:

- A map of the Boonsong Lekagul Nature Preserve containing the locations of the various gates throughout the park (in both bitmap and jpeg format);
- A CSV containing identifiers of each car, the type of car, and the timestamp and applicable location name of the gates the car passed.

The goal of our project is to create a data visualization that will allow the user (in this case, Mitch), to easily identify likely reasons for the decrease in nesting. Ideally, our data visualization will be novel in some way, be it a new data-visualization method, a new software developed for our visualization, or a novel application of an existing visualization method or software; this will allow researchers in the field to apply our system to other problems in the field beyond the VAST Challenge 2017 problem.

Team members

- Hao Zhu Hao.Zhu@tufts.edu
- Gerard Frasca Gerard.Frasca@tufts.edu
- Hongjie Wang Hongjie.Wang@tufts.edu

Goals and Tasks

The major goal of this project is to create a visualization that is predominantly confirmatory rather than exploratory; we are most importantly trying to identify the reason for the decrease in nesting. Another goal is to create a novel visualization application or system that could be applied to other situations outside this mini-challenge.

The first phase of the project (which we are currently completing) is the exploratory phase. Some members of our team are looking into ways to parse image data from the BMP to coordinates using python. Other members are exploring the data using excel and tableau to become familiar with the data and spot any major trends. This will allow us to gain the needed familiarity with the dataset to be successful in our project.

The second phase is the design phase. Based on the data, we will formulate a design for our visualization with the key features we hope to include. This will include ideas for visualizations, along with possible visualization interactions on a dashboard. In this phase, we will be predominantly sketching, storyboarding, and possibly even prototyping in UI prototyping tools like Figma.

The third phase is the building phase. Here, we plan to use analysis tools like Python and R for data analysis and cleaning. We will likely be using HTML, CSS, and Javascript (specifically the D3 library) to build our web application and visualize data. We may split up tasks based on our strengths. We will be using Github to compile code, write our project book, and collaborate. We will rigorously test our code throughout the building process, and make sure to cross-check our results by using other data visualization software like Tableau.

We will then present our findings to the class on April 15 as part of the progress updates and in-class presentations. In order to prepare for the presentation, we will have a slide deck and potentially a wireframe or prototype in HTML or Figma (maybe using a mock dataset) to help present the design of our dashboard and intended interaction. If we have any initial findings in this stage, we will also present those.

The fourth phase is the presentation phase. Throughout the process, we will be adding to our project book, and in this phase, we will compile our project book and prepare our final submission. We will also be putting “final touches” on our dashboard, testing for edge cases, and preparing our demonstration.

As previously noted, the major goal of this project is to create a dashboard that visualizes data to help identify the reason for the decrease in nesting habits of the Rose-crested Blue Pipit. However, we also hope to create a dashboard/desktop app with either or both of the following:

- Replicability (meaning our dashboard/desktop app can be used for other situations beyond our mini-challenge)

- Novelty (meaning our dashboard/desktop app uses new software to the field of data visualization, features a new visualization, or features a new application of an existing visualization)

Description of Dataset

- 171,477 Observations
- 2015/5/1 – 2016/5/31 Recorded time
- 18,708 Car id recorded
- 7 Car types
 - 1: 2 axle cars
 - 2: 2 axle trucks
 - 2P: 2 axle trucks belonging to park services
 - 3: 3 axle trucks
 - 4: 4+ axle trucks
 - 5: 2 axle bus
 - 6: 3 axle bus
- 6 Sensor types
 - 9 camping
 - 5 entrance
 - 9 gate
 - 4 general-gate
 - 1 ranger-base
 - 8 ranger-stop

Here we also performed some exploratory data analysis and we summarized the results in the visualizations below in Figure 1, 2, and 3.

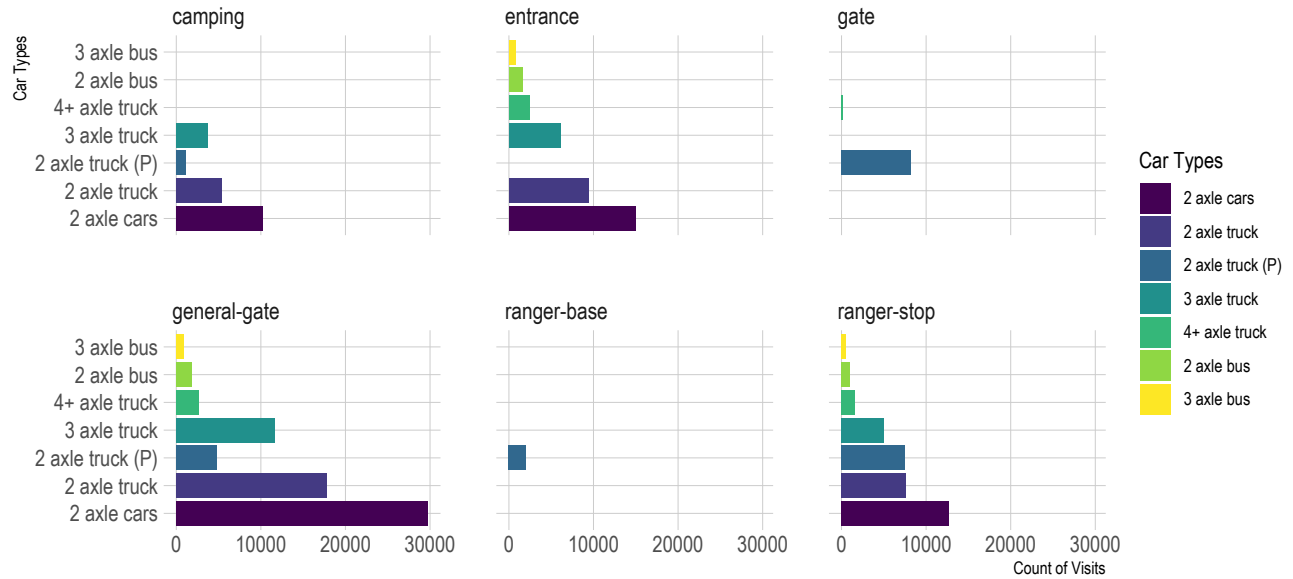


Figure 1: Distribution of visits stratified by Car Types and Gate Types.

Hypothesis

- Certain visitors are irregularly active in certain regions of the park, causing the decreased nesting. We can formulate this problem as an outlier detection problem.
- There are some illegal transpassing of gates by some 4+ axle trucks early in the morning (see the screenshots below). This kind of suspicious behavior might link to some activity that causes the decreased nesting.
- The park rangers are causing the decreased nesting since they are the only ones traveling in certain parts of the park and they drive trucks.

2016/5/3 3:54	20163710053757-851	4 gate5
2016/5/3 4:39	20163712023718-192	4 gate5
2015/6/2 5:03	20150418050406-878	4 gate3
2015/6/2 5:19	20150418070414-573	4 gate3

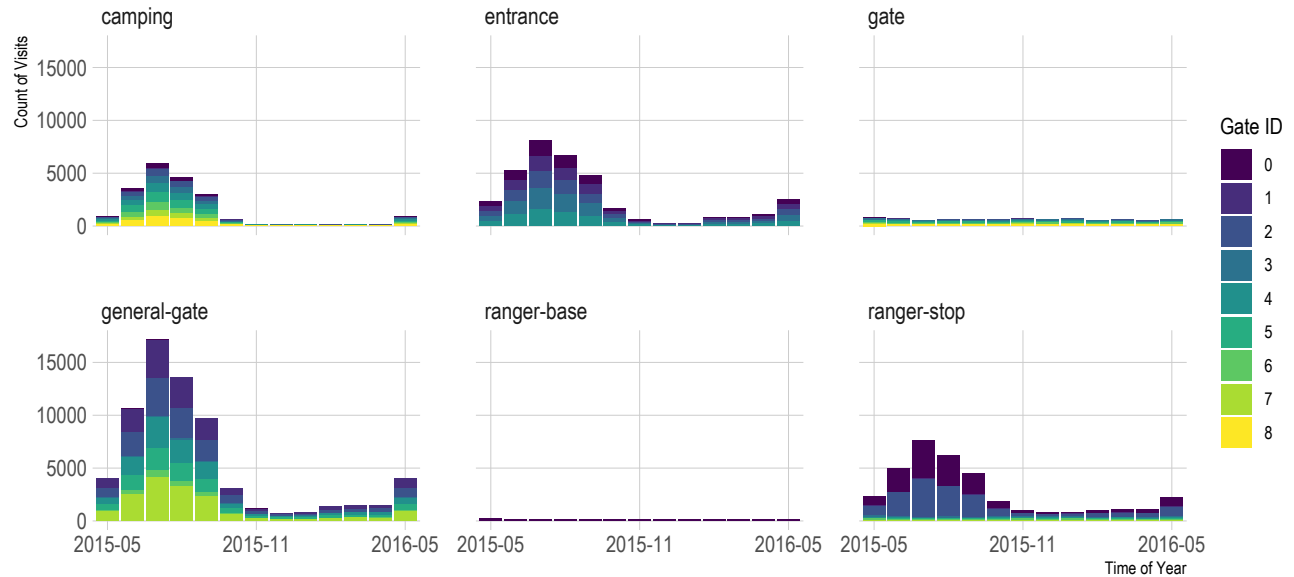


Figure 2: Annual distribution of visits at different types of gates stratified by gate ID. It's obvious that within the same category, some locations are more popular than others. This trend carries out throughout the season but is more obvious in the summertime.

Project Timeline

Phase 1

- March 19: Group meeting, exploratory data analysis
- March 23: Roughout sketches, making hypothesis, feature lists are decided
- March 25: Submission project plan

Phase 2

- March 23 - April 14: Planning and Developing
 - More detailed sketches
 - Wireframes and (maybe) Prototypes
 - Pseudocode (if needed, for more complicated portions)
 - Decisions on which tools to use on features of the dashboard
 - (R shiny, D3, Tableau, etc.)
 - Developing
- April 13 Submission project prototype version1, record feedback
- April 14: Intermediate Milestone Class Presentation

Phase 3

- April 14 - May 11: Designing
 - More development and edits based on feedback from class - presentation

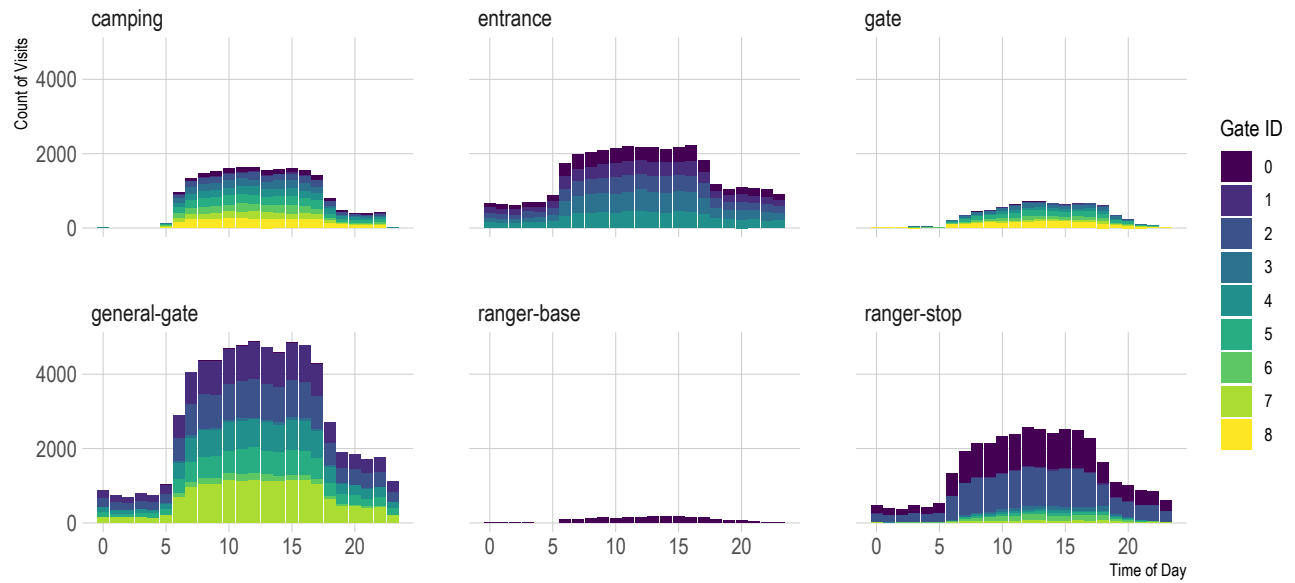
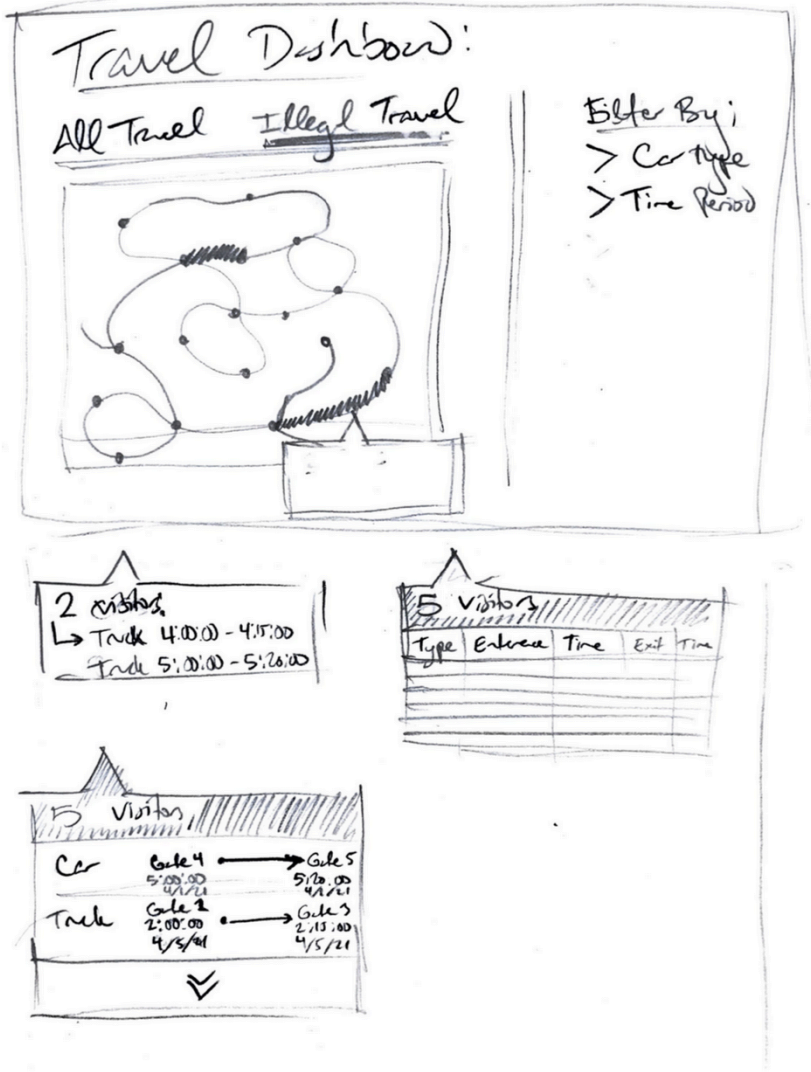


Figure 3: Hourly distribution of visits at different types of gates stratified by gate ID.

- Testing
- Finishing touches and visual design
- May 11: Submission final project
- May 11: Final Project Demo



Feature List

Must-have

- Graphs with the distribution of time spent per segment of the park such that users could use these tools to capture any irregular visitors. It should come with:
- Time/Car/Ranger filter
- Ability to search car ID to allow users to check activity and confirm hypotheses.

Good-to-have

- Interactive map (see Appendix A) with:
- Ability to toggle between all travel and illegal travel
- Tooltip with more detailed info
- Time/Car/Ranger filter
- Color to indicate traffic density

Optional

- Stratified distribution charts ranked by traffic density or time span.

Team Roles

Gerard Frasca will be the designer. He will oversee the general design of the project (meaning the information architecture of the dashboard, the types of visualizations to include, and the visual design of the dashboard). He will also help with wireframing and prototyping if needed. With experience in HTML, CSS, and Javascript, he will help style and format the dashboard. He will also be in charge of writing and compiling the project book.

Hongjie Wang will be the developer. She will use tools like Tableau to analyze data, make hypotheses based on the data and also help to build the interactive dashboard (which will likely be in Javascript).

Hao Zhu will be the data scientist. He will dive into the data and try to discover useful patterns. With experience in R, python, and js, he will also develop visualization as needed to explain the story. He will also take the responsibility of rendering the documents in a professional pdf format and once the web-based visualization is developed, he will help set up the webspace to host the static page.