

DinoFun World: Crime Mystery - Trends & Insights

Introduction:

In celebration of a hometown hero and international soccer star, the Scott Jones celebration was hosted at Dinofun World from June 6th to 8th in 2014. However, the exhibition with Scott Jones awards and memorabilia disappeared and/or was vandalized by unknown individuals. The exhibition, which is located in Creighton Pavilion (Coaster Alley), had to be closed off in order to investigate what may have happened.

Hypothesis:

After a preliminary reading of the DinoFun website and the data presented, we hypothesize that the culprits who vandalized Scott Jones' exhibition could be some of his childhood friends or a group comprised of security staff members. Jones' childhood friends—specifically those he played soccer with in high school, could have a motive of jealousy of Jones' career after one did not receive a soccer scholarship for college. At the time of crime, there were openings for security personnel for DinoFun, giving a malicious group the opportunity to get security clearances and access they would need for this crime. We also acknowledge the fact that the childhood friends and the security group might be the same entities.

Methodology:

While approaching this challenge, we recognized the extensive quality of all the datasets and the context we had from the amusement park's website, the challenge's description, and the article describing the crime incident. While the texts were all imperative to understanding what happened, it did not hold the information needed to determine exactly who the perpetrators were, when exactly it happened, as well as other essential details. It was determined that any interesting patterns or outliers we can find in the communication or movement data will be important to identifying all the unknowns as well as addressing the

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

typical trends at the amusement park. All three days were consistently compared for any important outliers and patterns.

Analysis:

Identifying Trends of Entries and Group Size & Behavior:

We analyzed the movement data to get an overview of visitors' movement patterns, group size, and group behaviors. In order to examine the movement patterns over time across three days, for each ID, we identified their earliest and latest timestamp as their entry time and exit time, and then plotted the count of entries/exits at every 10-minute interval. In Figure 3 and 4, we can see the trends of entries and exits are generally consistent across three days. The majority of people entered the park before 9 am and left after 7 pm, and the total number of visitors increased day by day. To determine group sizes, we identified groups based on the assumption that people who were in the same group might not always move together, but would enter and leave the park at the same time, which also means they would stay in the park for the same length of time. In our analysis, people who had stayed in the park for the same duration of time were considered a group; the time duration was estimated with 6 decimal places using the unit of hours. In Figure 5, we plotted time duration and group size in order to know how common a kind of group was and for how long groups of people stayed. Again, the overall patterns were similar across all three days. From the clusters of points, we could see the most common group size is under five people. The majority of groups spent more than 11 hours in the park, and some groups with small sizes stayed for less than 10 hours. Nearly all groups with more than 20 people stay in the park for more than 11 hours as we could see from the upper right region of each graph. (See interactive plots in "movement_patterns_EY.html" for figures 3-5)

To confirm our group strategy and get a closer look at specific groups, we randomly picked a time duration to select a group. Figure 6 shows the movement of a group of 3 people over time on Saturday; we can see their movement patterns almost perfectly overlap, which lends support to our grouping approach. Figure 7 shows the movement of a group of 14 people.

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

We can see their overall paths were very similar, but they were not always moving together. For example, compared to other people, 1770487, 701095, 827800 did not go to Coaster Alley or Kiddie Land but went to the shopping area. It would be reasonable to assume that there were three adults with 11 kids in this group, and they could be a large family or a school group. Further investigations are needed to see if the patterns are generalizable to other groups. Also, at this point we only identify patterns, but could not find any anomaly from the analysis of movement data. Future steps should be taken to examine movement in subsets of time or specific areas based on what we thought could be suspicious from our hypothesis and the communication data.

Communication Trends:

Our first look into the communication data involved identifying the key variables at hand: sender IDs, receiver IDs (which included external phones), timestamps of each message sent, and the general location of the sender. We grouped the IDs by sender and receiver, summarizing the count of each occurrence to obtain the frequency of individuals communicating amongst each other. By combining this information with the movement trends identified previously, we were able to more solidly identify individuals in groups.

The timestamps allowed us to keep track of total communications activity over the day as well as notice generally where they are often sent from. Additionally, the ability to track unique IDs allows us to take note of unusual activity or spikes from a single person. As shown in Figure 1, each day had relatively similar patterns of communications with messages from the Wet Land being the most common. Friday had lower frequency of messages while Saturday and Sunday generally had more. Messages also tended to drop as it got closer to the later hours (21:00 to 23:59) indicating that this may be around when the amusement park closes. It was believed that many of the messages sent during these hours are employees or security guards who are handling behind the scenes work for the day.

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

Sunday Anomaly:

From this initial analysis, we noticed that there were two large spikes in messages sent on Sunday. These two large increases came from the two areas connected to Creighton Pavilion, Wet Land peaking at 12:00 and Coaster Alley peaking at 14:00 as shown in Figure 2. We decided to further investigate two IDs with the most messages sent on that day: 1278894 and 839736.

Looking through both the movement and communications data, we concluded that 1278894 may have been a staff member or some automated internal system at DinoFun World, especially since this ID appeared over all three days. ID 839736 may be a security guard or staff due to their stationary position at the Entry Corridor throughout all three days. Each had unusually high spikes which indicated that these IDs were contacted by multiple people throughout the park, notably in Wet Land and Coaster Alley, which is where the Scott Jones exhibition is located in/near. However, it is difficult to discern if ID 839736 was one of the main security guards who may have just discovered the vandalism of the exhibit or one of the perpetrators of the crime. It is noted however, that we suspect that the vandalism may have been discovered around 11:00-12:00 due to the unusually high spikes of activity. (Source: Communication Data Files & VAST Challenge 2015 Data Description)

The Wet Land Activity Anomaly :

The strange communication spike that took place between 11:00 to 13:00 on Sunday in the Wet Land area motivated us to look into the movement data. Based on the frequency of message of 839736 (whom we suspect is a staff member because this ID did not appear in any of the park movement data from Friday to Sunday), the time frame and locations that the communication spike occurred, we filtered the Sunday movement data using timestamp and coordinates. Although we couldn't tie the excessive messaging to the movement inside the park on Sunday, we noticed that the frequency of check-in at location (50,57) was much lower than the check-in frequency for other activities in the Wet Land area during the 11-13 time frame. Using the park maps as a reference, (50,57) roughly corresponds to rides 24. To further examine whether this was a pattern, we calculated the frequency of check-ins for that particular

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

coordinate for the entire Sunday and across all three days. The graphs indicated that the check-in anomaly at Wet Land holds true. Though it was inconclusive as to whether the below average check-in frequency in the Wet Land at rides 24 had anything to do with the crime that took place, this observation might be useful for the park to improve its future path design, making rides 24 more accessible to people.

Conclusion:

Overall, we were able to use similarities in movement data and frequent communication between IDs to identify groups traveling in DinoFun world. From this, we were able to rule out groups that were not responsible for the crime at hand, but our analysis was inconclusive as to finding who the actual culprits were. Through investigating the data that broke the standard trends in both movement and communication, we found some potential suspects in IDs 1278894 and 839736 as well as any of the other IDs that were in constant contact with these IDs.

Further research into these suspects' movement and communication at the time of the crime is necessary to make any conclusions about their role in the vandalism. Additionally, next important steps for narrowing down on what groups are involved and who the perpetrator(s) may be is to check for additional unusual occurrences in movement specifically at the crime scene within the timeframe we have suspicion upon. It will also be beneficial in the future to look into additional methods/programs that will allow us to more efficiently analyze our movement data.

(Page count: 5 for text)

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

Data References:

1. DinoFun World: Main Website. <https://jcrouser.github.io/CSC235/dinofunworld/>
2. Mayhem at DinoFun World: News Article. In Auxiliary Files
3. VAST Challenge 2015 Data Description. In Auxiliary Files
4. Communication Data Files
5. Movement Data Files

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

Appendix:

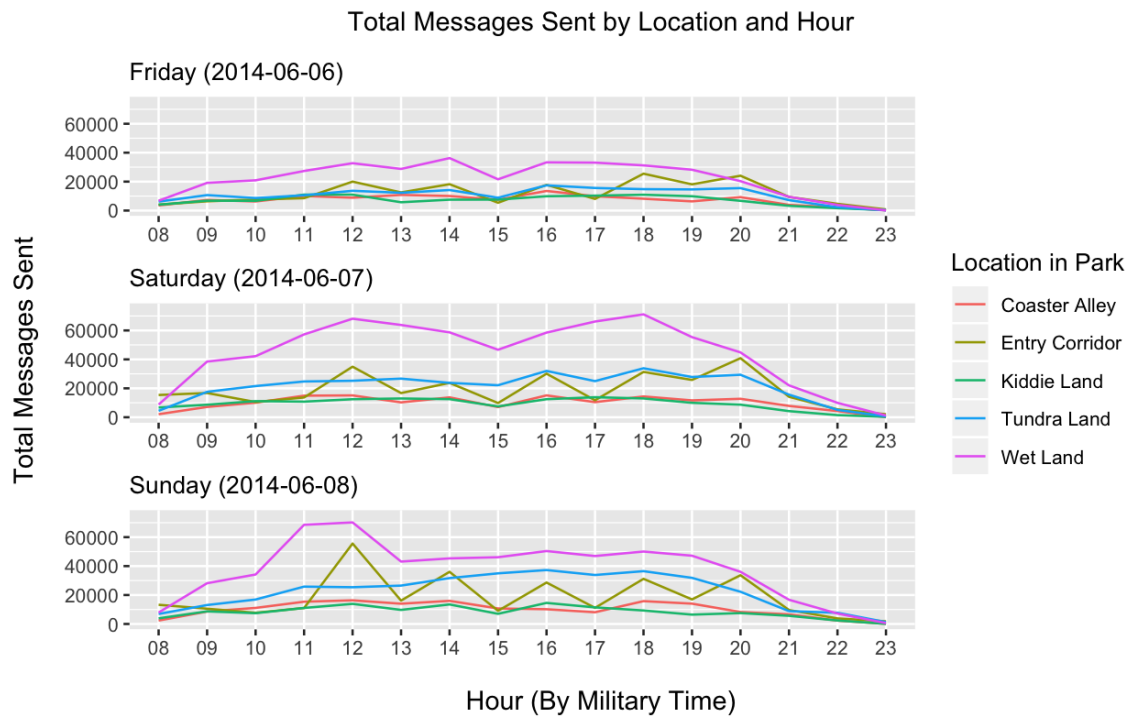


Figure 1: Total Messages Sent in Amusement Park by Hour and Location

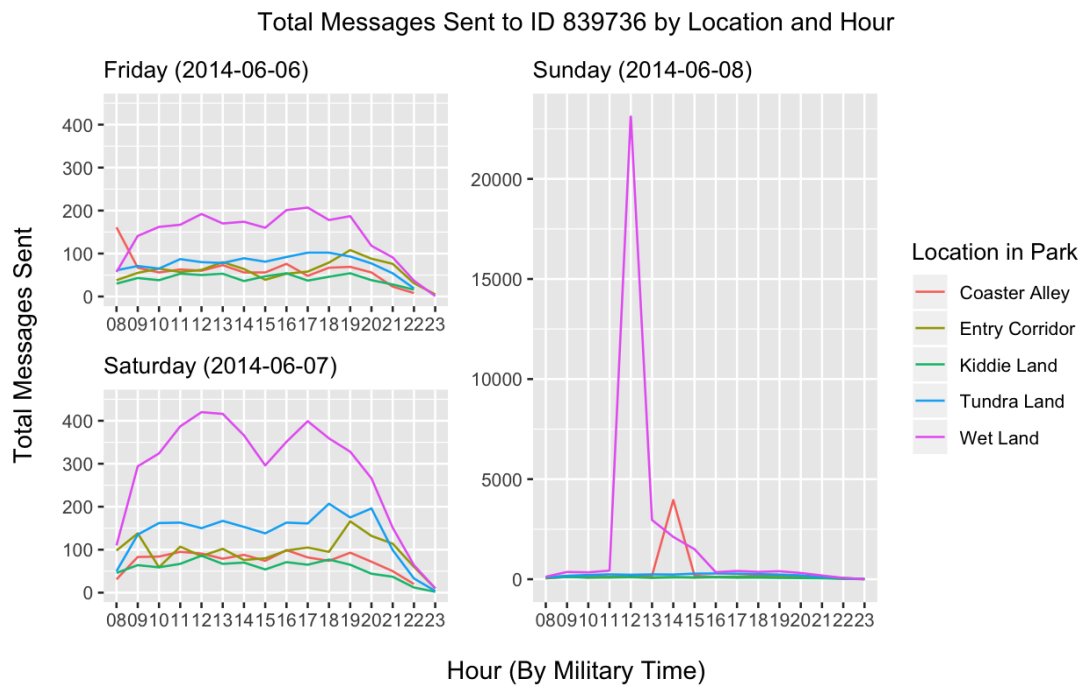


Figure 2: Total Messages Sent to ID 839736 by Hour and Location

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

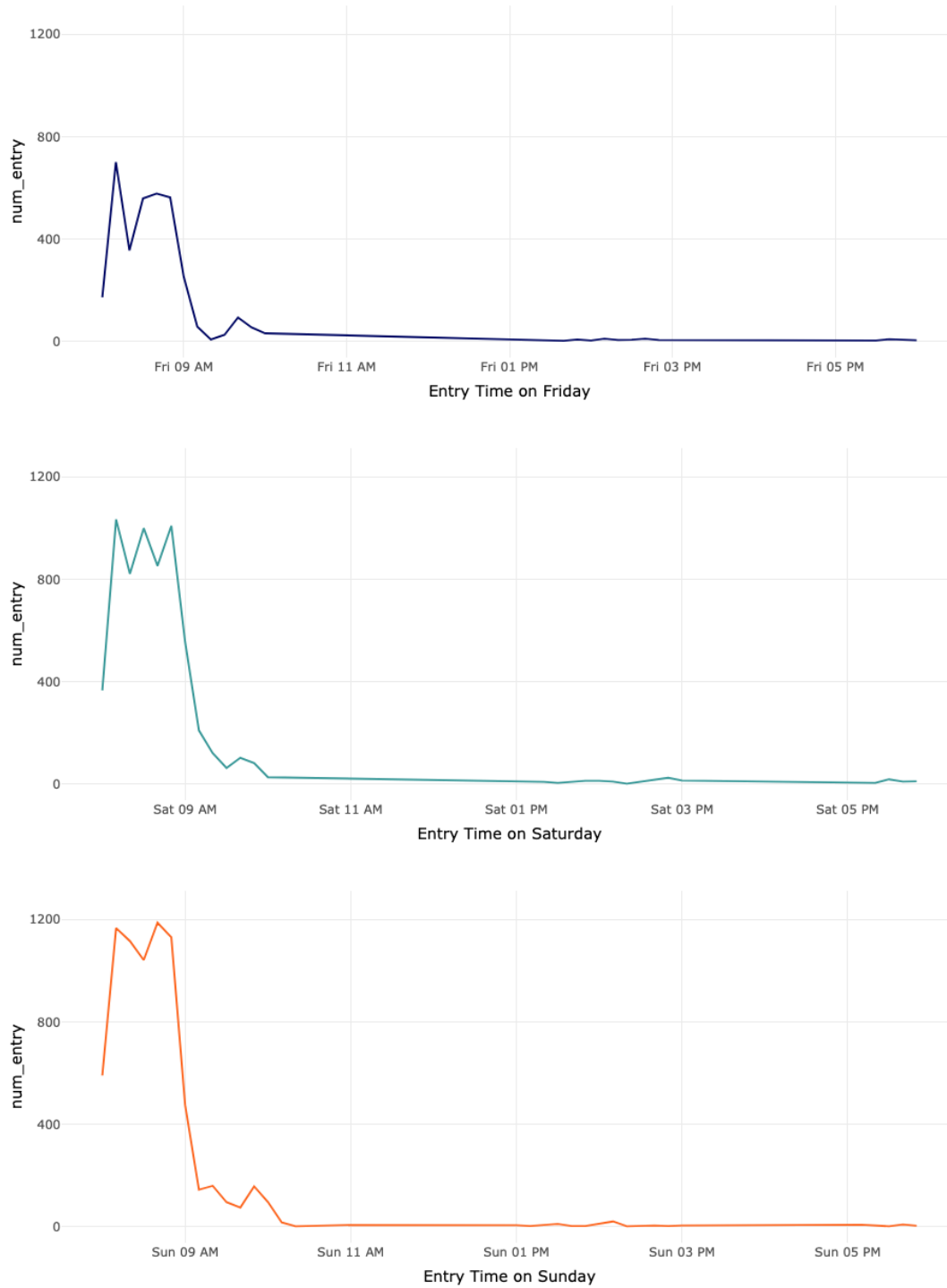


Figure 3: Patterns of Entry Counts over Time.

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

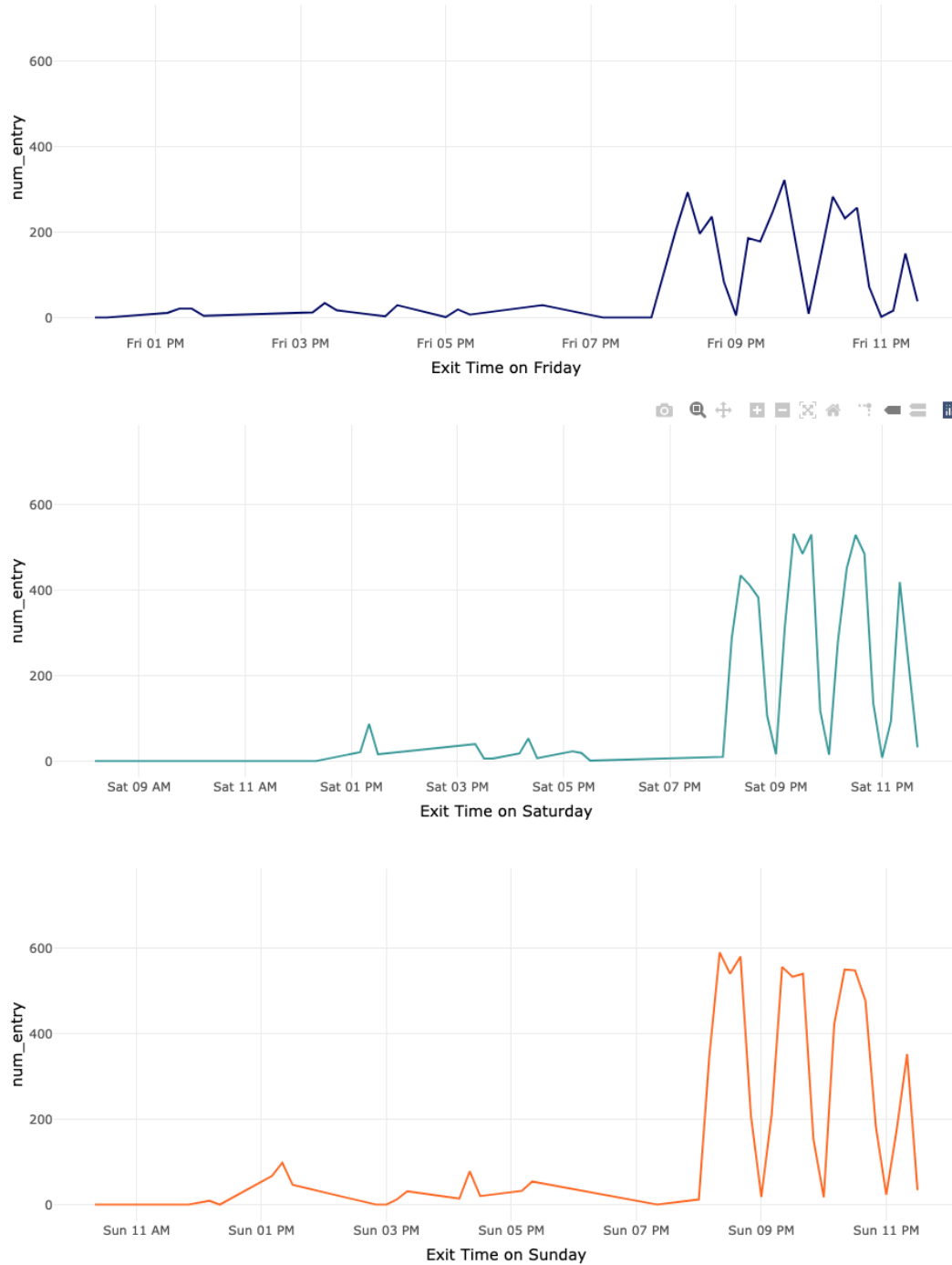


Figure 4: Patterns of Exit Counts over Time.

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

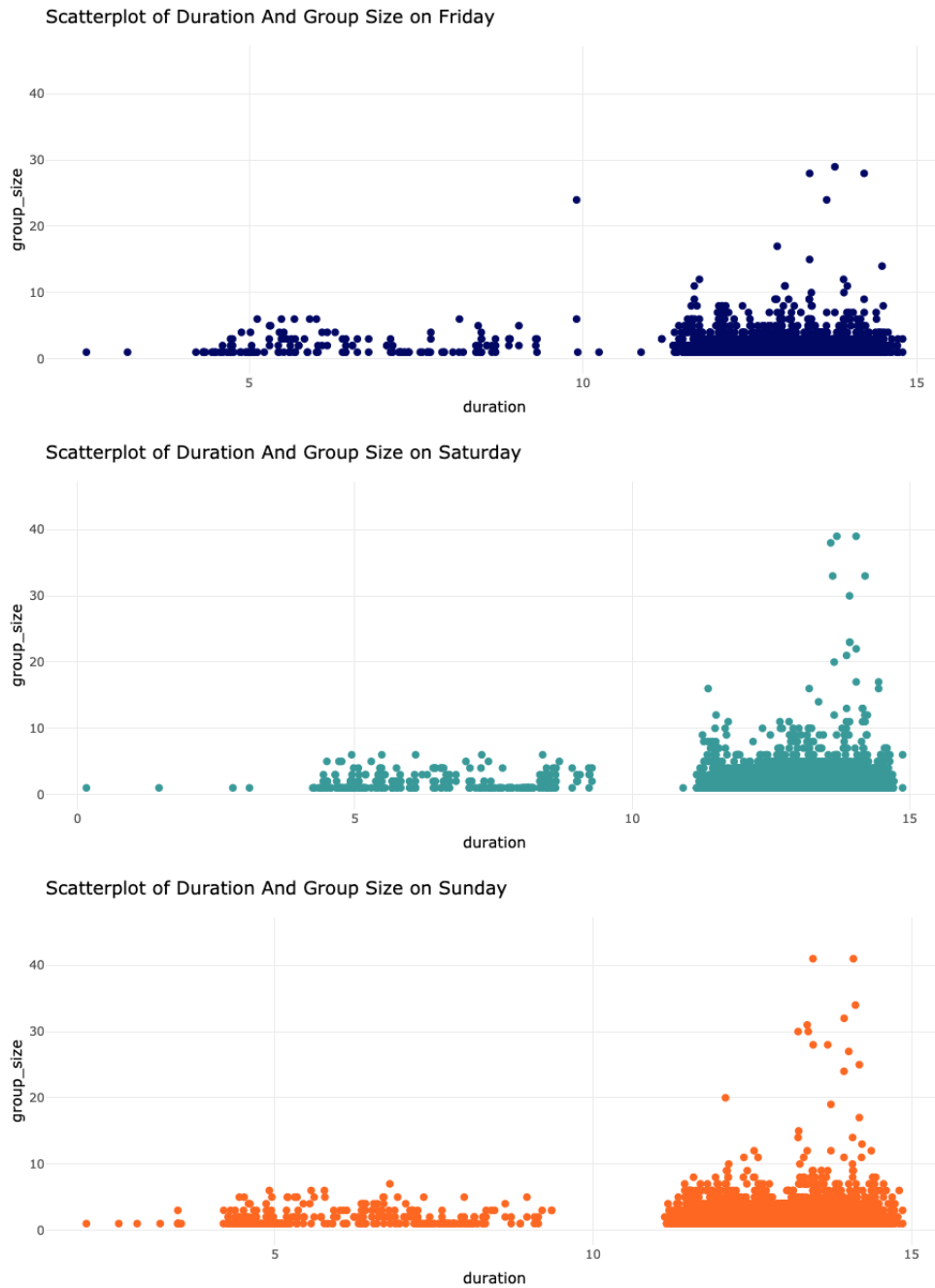


Figure 5: Duration of Time in the Park vs Group Size

DC1: Exploratory Data Analysis

Group: Adriana Beltran Andrade, Catherine Park, Casey Perez, Rachel Yan, Elaine Ye

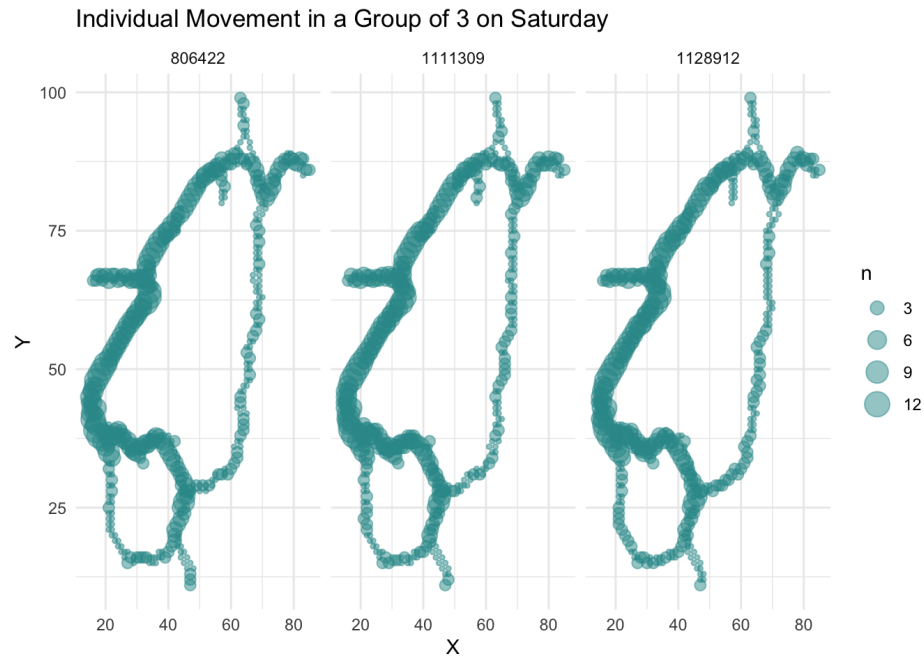


Figure 6: Individual Movement in a Group of 3 on Saturday

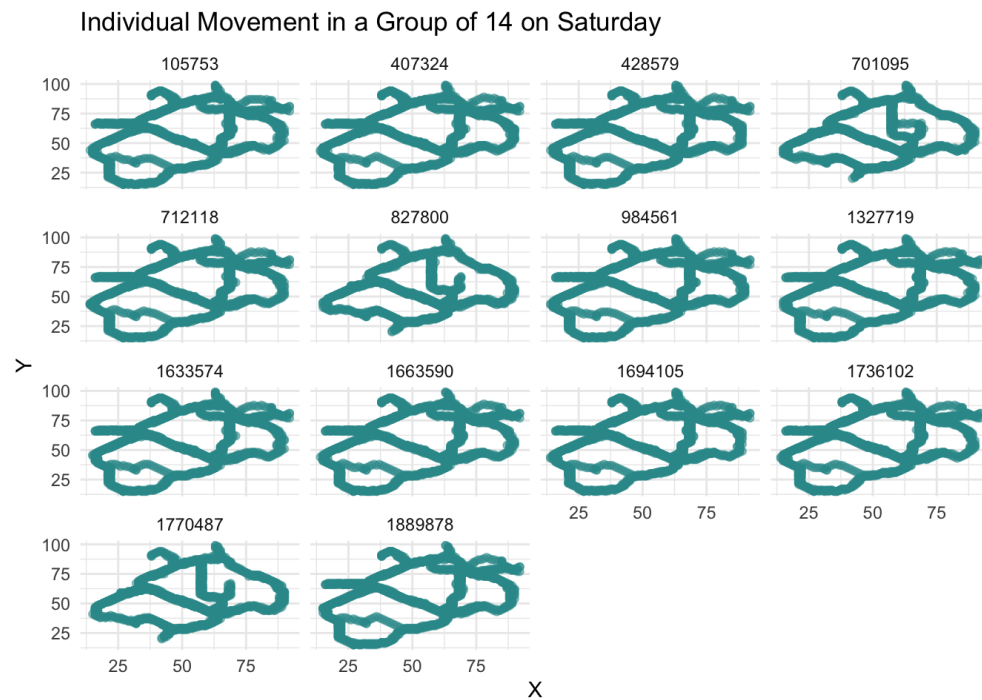


Figure 7: Individual Movement in a Group of 14 on Saturday

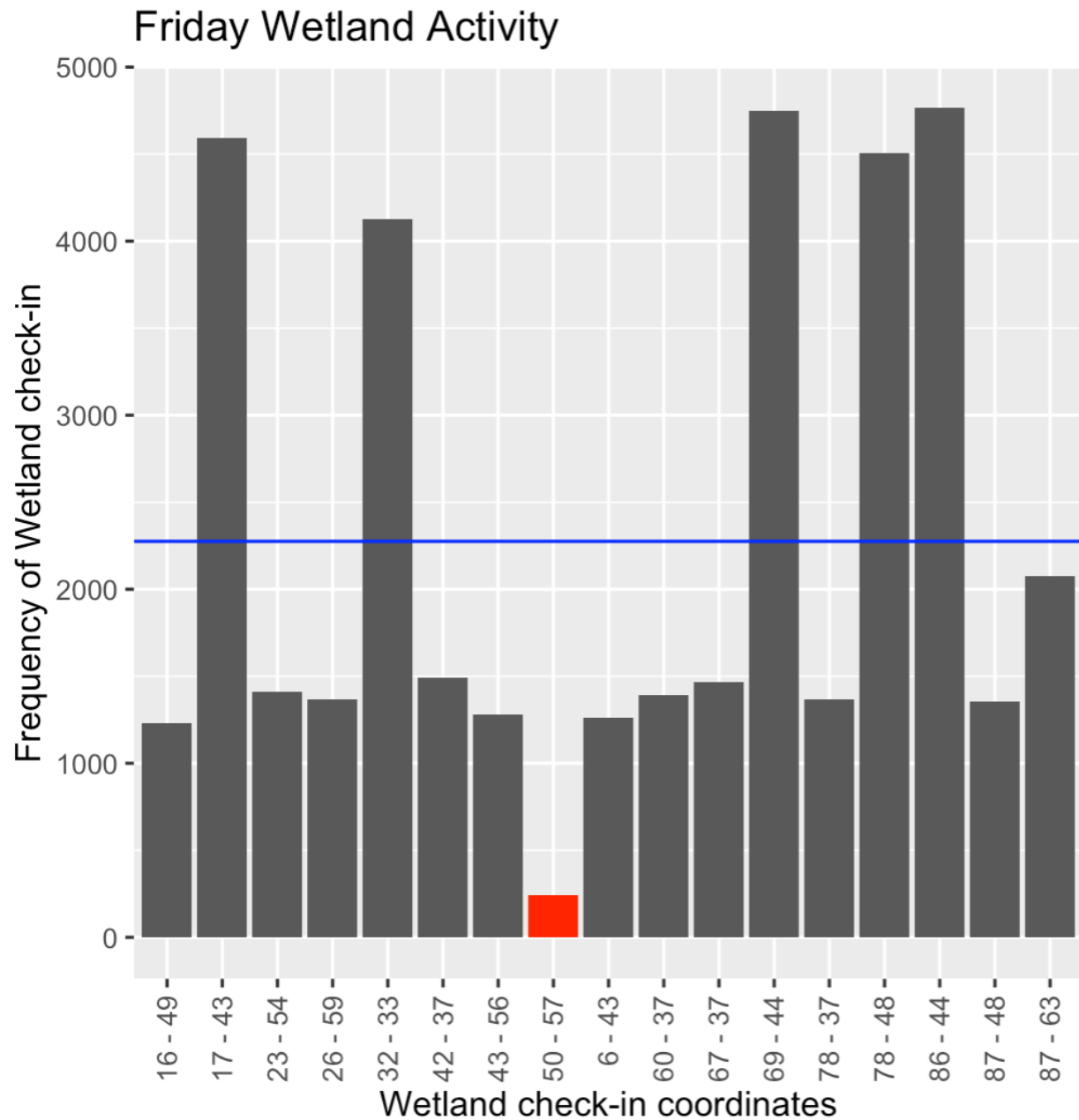


Figure 8: Friday Wet Land activity

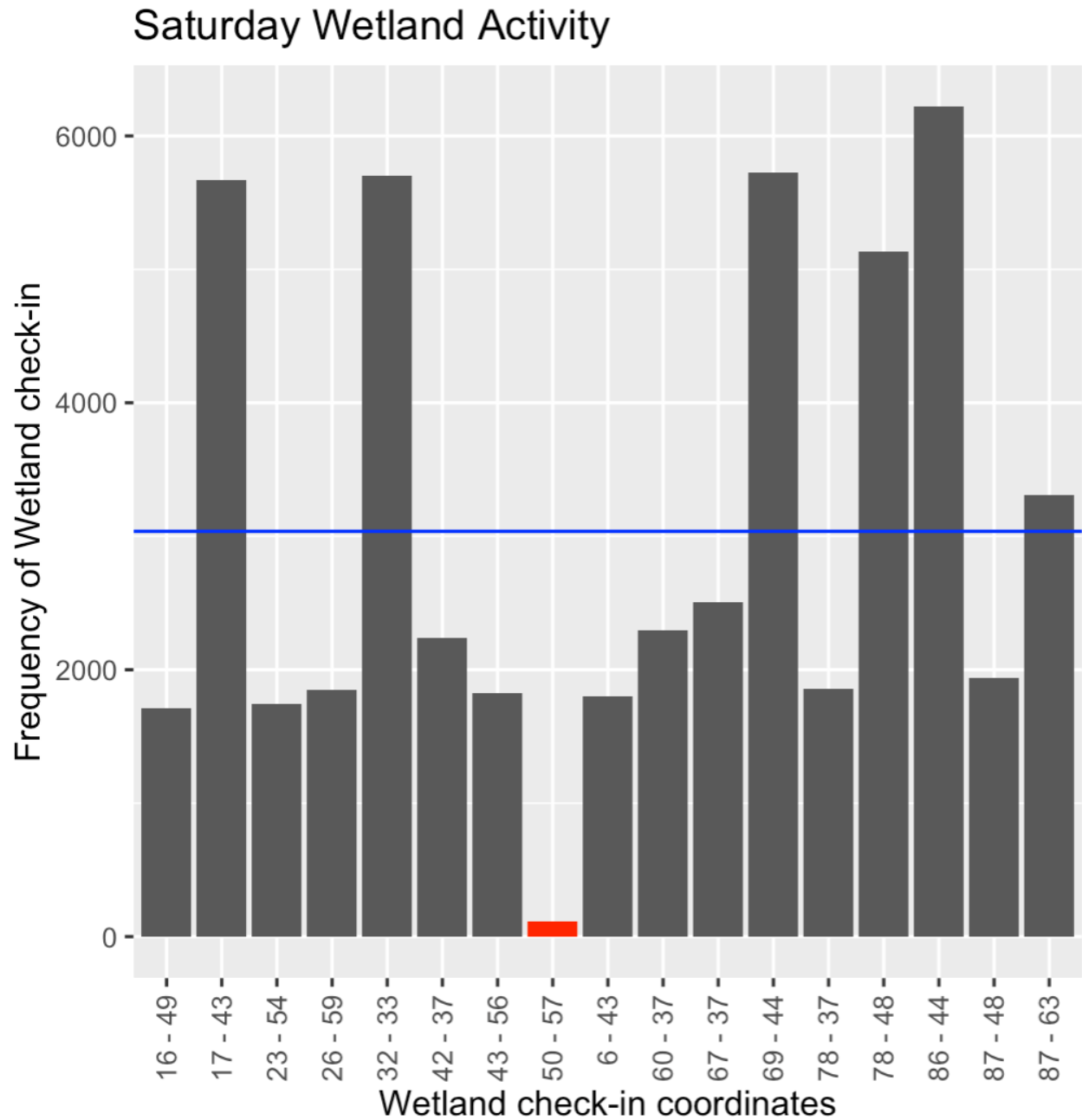


Figure 9. Wet Land Saturday Activity

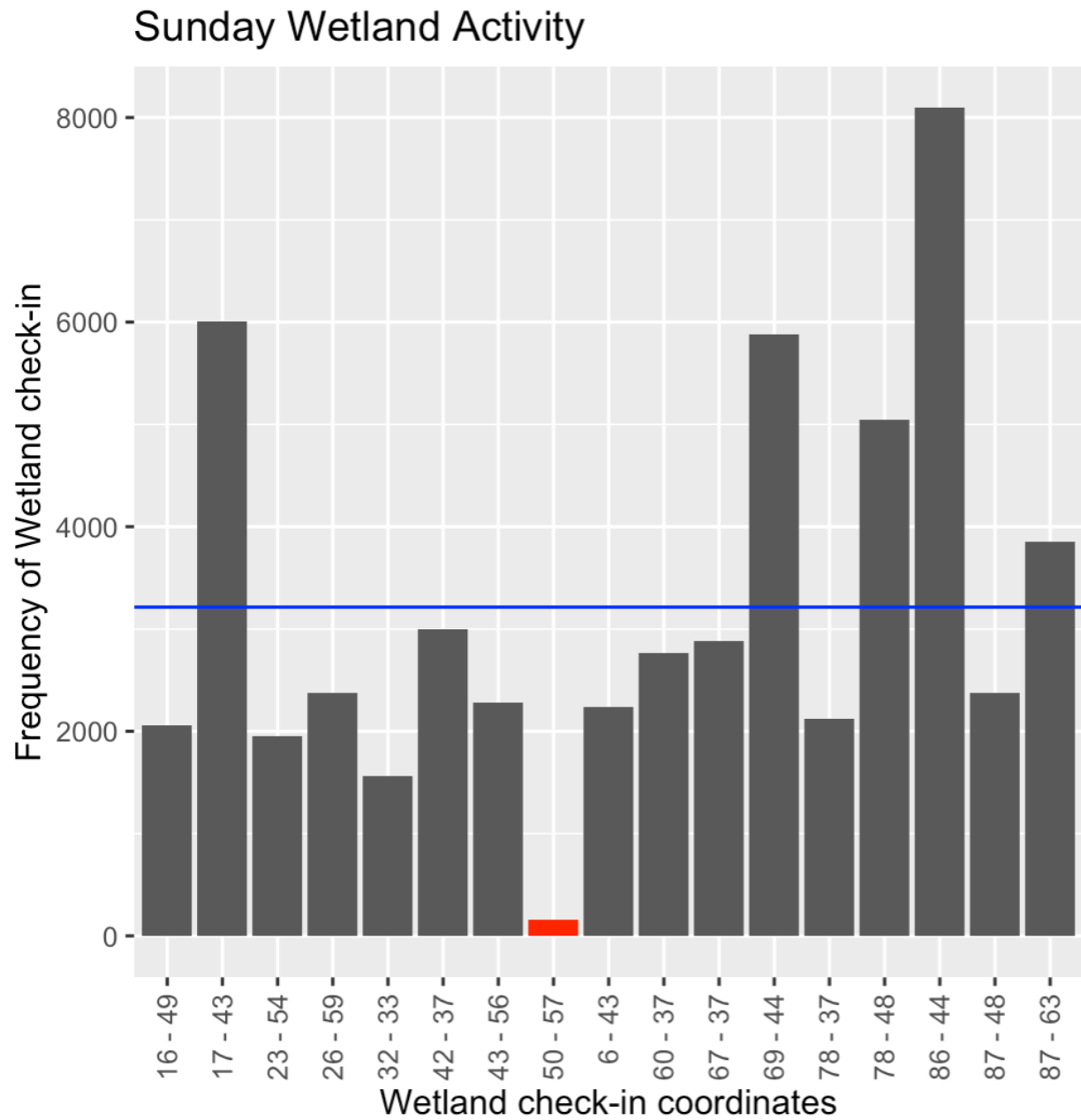


Figure 10. Wet Land Sunday Activity.