On the move at DinoFun World

Heike Hofmann, *Member, IEEE*, Dianne Cook, Andee Kaplan, Eric Hare, Vianey Leos-Barajas, Carson Sievert and Samantha Tyner

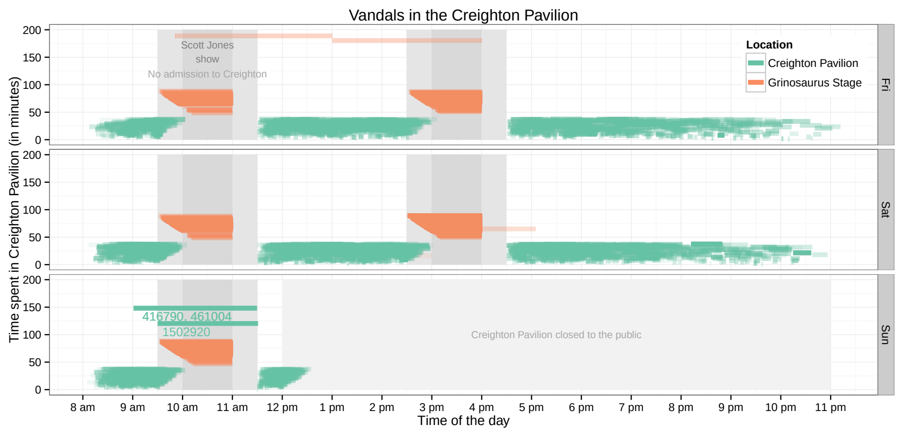

Fig. 1: On Sunday, June 8 2014 a trio of suspects (marked in green) checks in to Creighton Pavilion at 9:01 am and stays there for more than 2 hours, in particular during the time that the Pavilion is closed to the public. What are they doing there? Our working theory is that some vandalism took place during this time, forcing park authorities to close the pavilion for the rest of the day.

Index Terms—Time series plots, Droplet plots, Barcode plots.

1 Data Manipulation

The data set under consideration comes from the VAST 2015 Mini-Challenge 1 consisting of movements of park-goers at DinoFun World, a fictitious medium-sized amusement park, for the weekend of June 8-10, 2014. Each record in the data consists of an identifier, timestamp, X-Y location, and the type of movement, distinguishing check-ins into an attraction and movements within 5m $\times$ 5m grid cells measured by sensors along the park's pathways.

Our results are based on data augmented by more detailed location information and timelines derived from individuals' movements through the park.

For the time spent in rides (section 2.1) we focus on just the check-in events enhanced by location-based auxiliary information given on the DinoFun World website, which allows us to matchup each check-in with the corresponding attraction; its location, name and classification as Kiddie Ride, Thrill Ride or Ride for Everybody. We further derive the time that park-goers spend in an attraction as the time between check-in and first recorded movement afterwards.

The time of park-goers' first movement after a check-in event is also important in identifying when rides at each attraction occur (see section 2.2). We use the time of the first movement as a potential end time of a ride. For most rides, these end times show huge frequency spikes at regular intervals, indicating both the end of a ride and giving some indication how long each ride is lasting.

- *Heike Hofmann, Andee Kaplan, Eric Hare, Vianey Leos-Barajas, Carson Sievert and Samatha Tyner are with the Department of Statistics, Iowa State University. E-mail: {hofmann, ajkaplan, erichare, vianey, csievert, sctyner}@iastate.edu.*
- *Dianne Cook is with the Department of Econometrics and Business Statistics, Monash University. E-mail: dicook@monash.edu.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

2 Results

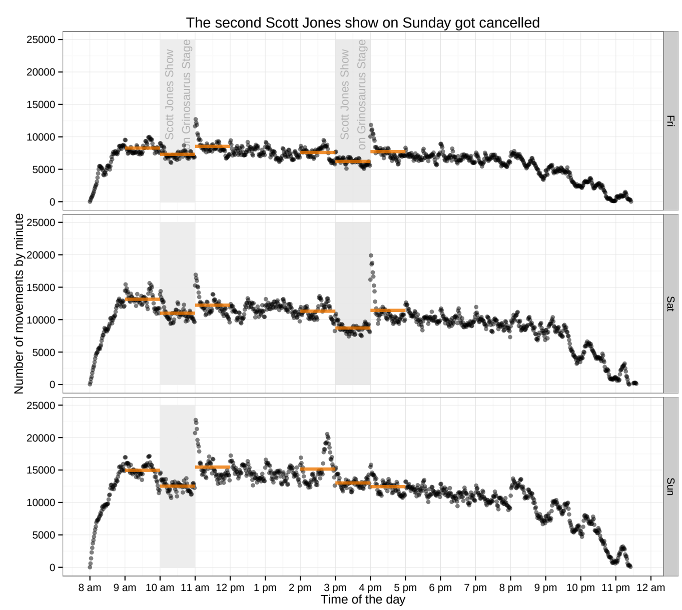

Fig. 2: Number of movements in DinoFun World by minute. The Scott Jones show is responsible for a lot of the movement in the park over the weekend.

Overall attendance at DinoFun World is characterized in figure 2. The number of moves park-goers make is charted for each minute of the day along a horizontal time axis. We can learn a couple of things from this plot:

(1) the Scott Jones show was held from 10 to 11 during all mornings and from 3 to 4 in the afternoons of Friday and Saturday. We can see this from the dip in movements (orange lines are average number of moves during one hour) during this time, and the spikes immediately

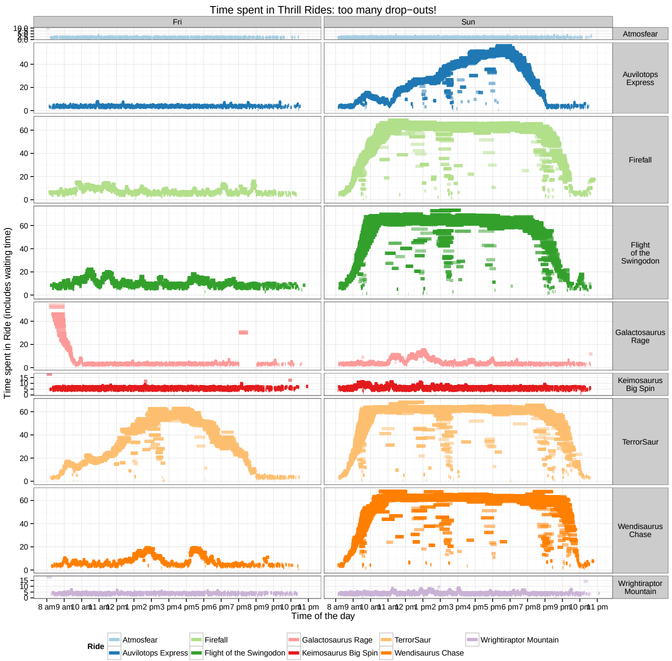

Fig. 3: Time spent in each of the Thrill Rides on Friday and Sunday. Each park-goer is a line, the length on the line and its vertical position indicate how much time was spend in each ride.

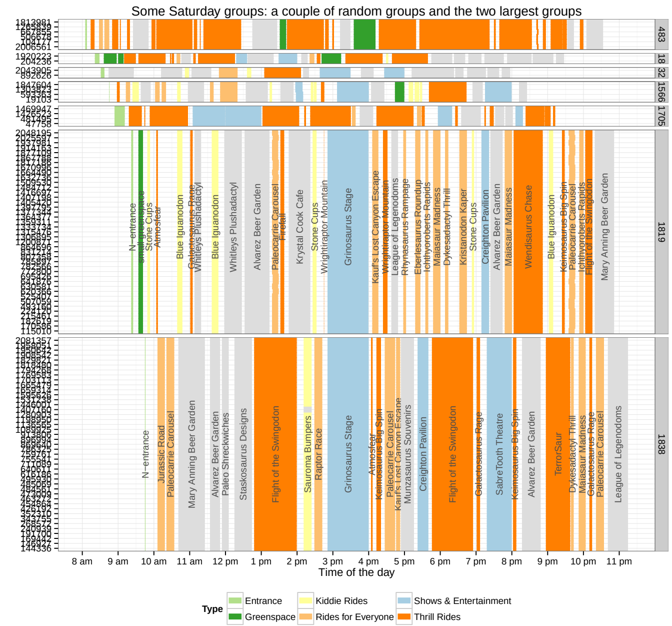

Fig. 4: Each group is uniquely identified by the activities they do together. For each individual coloured horizontal lines are drawn from the time of check-in to the time he/she leaves an attraction. Groups of individuals doing the same activities then result in vertical stripes. Stripes with light grey backgrounds correspond to breaks from the rides, that don't require checking in.

at the end of the show (when a lot of people move out of the area). The second show on Sunday was cancelled: the dip in movements is missing on Sunday afternoon. This is also visible in figure 2, showing check-ins by area: people check into the Grinosaurus Stage, where the Scott Jones show is hosted, before 10 am for all days but before 3 pm only on Friday and Saturday.

(2) there is a spike in movements on Sunday at around 2:30 pm - judging from the movement pattern these are people on their way to the Scott Jones show who get turned away (because there are no check-ins to the Grinosaurus Stage), see also figure 3.

2.1 Time in Rides

Fig 3 shows operating times of each of the thrill rides: on the right time of day is shown, along the y-axis, the amount of time spent in a ride. Operating time is shown as a line, each line corresponds to one individual in a ride. The longer the line, the longer an individual spent in the ride. This includes waiting times after checking into the ride. It is obvious, that rides are much busier on Sunday. On weekend days park-goers have to calculate to take one hour for one ride of the Flight of the Swingodon, whereas they can manage about 3-4 rides an hour of the Flight of the Swingodon on Friday. The pattern of droplets under each of the thick lines indicate people who have checked into a ride, but leave before it is their turn to take it. Just before 3pm on Sunday, a lot of these droplets occur across all the backed up thrill rides - maybe these people want to go to the (by now cancelled) Scott Jones show? DinoFun needs to work on their general announcement system!

2.2 Barcode of activities

We define a group as a set of people who ride the same rides together.

We exclude any of the rides that are potentially walk-throughs: Atmosfear, Galactosaurus Rage and Dykesadactyl Thrill. Beyond that we include all of the Thrill Rides, Kiddie Rides and Rides for Everyone.

For each of the days of the time frame we create a binary matrix of park-goer id and ride id, where a 1 in cell (i, j) indicating that park-goer i was on ride j. We then use a binary metric to calculate pairwise distances between two park-goers as the proportion of rides that only one of the two individuals took and the total number of rides that at least one of them did. This yields a distance between 0 and 1, where 0 means that all rides where done jointly, and 1 that none of the rides was done together.

For example, if individual k took 23 rides, and ℓ did 25 rides, and they did 20 rides of these together, the ride based distance between them is (28 - 20)/28 = 0.29.

Using Ward's method for a hierarchical clustering based on these participant-participant distances, we identify groups of park-goers.

Figure 4 shows a sample of several groups going on rides together on Saturday. Just before 3 pm, Group 1819 seems to realize that taking the Wrightiraptor Mountain ride would make them late for the Scott Jones show, and they file back out after a short wait in the line.

2.3 Vandals in Creighton Park

Figure 1 is constructed similarly to figure 3, but focuses on the time that park-goers spend in the Creighton Pavilion, which hosts the Scott Jones exhibit, and at the Scott Jones show on the Grinosaurus Stage. Besides a group of four individuals, who spend almost all day at the Grinosaurus Stage (security personnel should really check on these guys to see that they are OK), there are three suspicious individuals who check into the Creighton Pavilion and spend almost two hours all by themselves.

Acknowledgments

The authors wish to thank everybody who makes the software R [1] possible, and in particular the authors of the packages knitr [5, 4], ggplot2 [2], and dplyr [3].

References

[1] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[2] H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.

[3] H. Wickham and R. Francois. *dplyr: A Grammar of Data Manipulation*, 2015. R package version 0.4.1.

[4] Y. Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2013. ISBN 978-1482203530.

[5] Y. Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2015. R package version 1.10.5.