Vandals on the move at DinoFun World

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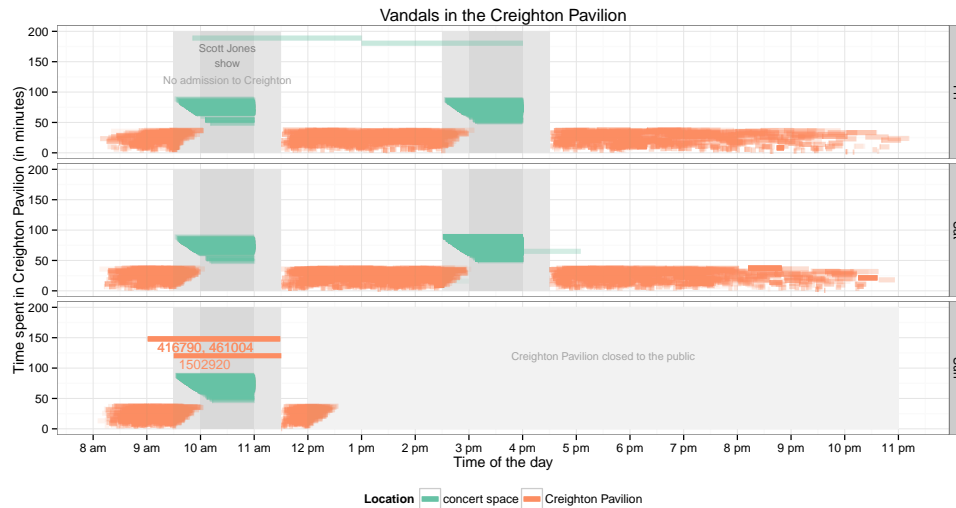

Fig. 1. On Sunday, June 8 2014 a trio of suspects (marked in orange) checks in to Creighton Pavilion at 9:01 am and stays there for more than 2 hour, in particular during the time that the Pavilion is closed to the public. What are they doing there? Our working theory is that the vandalism takes place during this time. Two of the suspects check into the pavilion at 11:30 again and leave within seconds, only to check back into the pavilion at 11:46.

Abstract—Here goes the abstract.

Index Terms—Keywords that describe your work. Will show as 'Index Terms' in journal please capitalize first letter and insert punctuation after last keyword

❖

1 DATA MANIPULATION

here we need to talk about a few things: construction of park.csv and time spent in attractions.

2 RESULTS

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 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 concert space before 10 am for all days but before 3 pm only on Friday and Saturday.

- *Heike Hofmann, Andee Kaplan, Eric Hare, and Vianey Leos-Barajas are with the Department of Statistics, Iowa State University. E-mail: {hofmann, ajkaplan, erichare, vianey}@iastate.edu.*
- *Dianne Cook is with the Department of Econometrics and Business Statistics, Monash University. E-mail: dicook@monash.edu.*
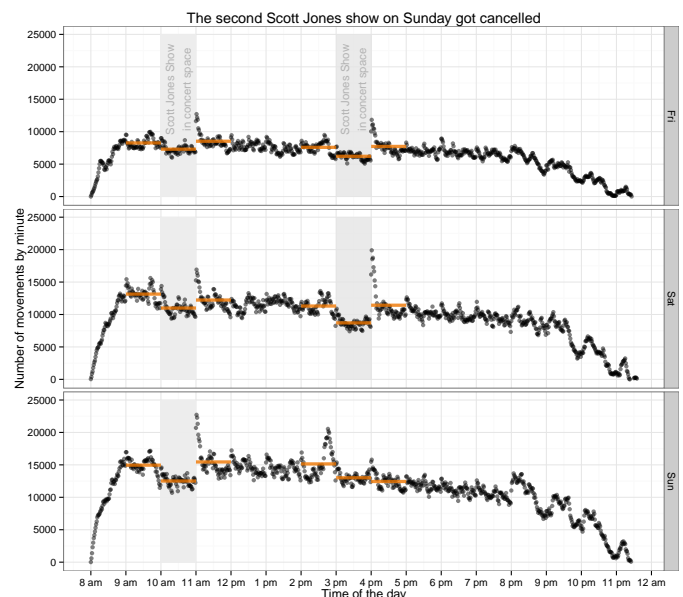
Fig. 2. Number of movements in DinoFun World by minute. The Scott Jones show is responsible for a lot of the movement in the park over the weekend.

Fig. 3. Time spent in each of the Thrill Rides.

(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 concert space), see also figure 3.

2.1 Time in Rides

We focus on just the check-in events in the movement data. Each check-in event is coupled with coordinates. Based on the auxiliary information given on the DinoFun World website (`get.the.sites.url`), we derive for each attraction its location, name and classification as Kiddie Ride, Thrill Ride or Ride for Everybody. This additional information is incorporated in the data of check-ins.

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 (defined as the time between check-in and first recorded movement afterwards). 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 Saturday and Sunday. On Saturday or Sunday 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.

2.2 Barcode of activities

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 matrix of park goer id and ride id, that consists of 0s and 1s, with 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 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 i took 23 rides, and j did 25 rides, and they did 20 rides of these together (i.e. in the same 'batch'), 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 get groups of park-goers on Friday, Saturday and Sunday.

2.3 Vandals in Creighton Park

Figure **??** is constructed similarly to figure 3, but focuses on the time that park goers spend in Creighton Pavilion, which hosts the Scott Jones exhibit, and at the Scott Jones show in the concert space. Besides the occasional individual, who spend almost all day in the concert space (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.

Technical Details   Thanks to R [1], and in particular the packages knitr [6, 4, 5], ggplot2 [2], dplyr [3], and XXX.

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, 2013.

[5] Y. Xie. knitr: A Comprehensive Tool for Reproducible Research in R, 2014.

[6] Y. Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R.* RStudio, 2015.