submission to the MC1 data challenge

Heike Hofmann, *Member, IEEE*, Dianne Cook, Andee Kaplan, Eric Hare and Vianey Leos-Barajas
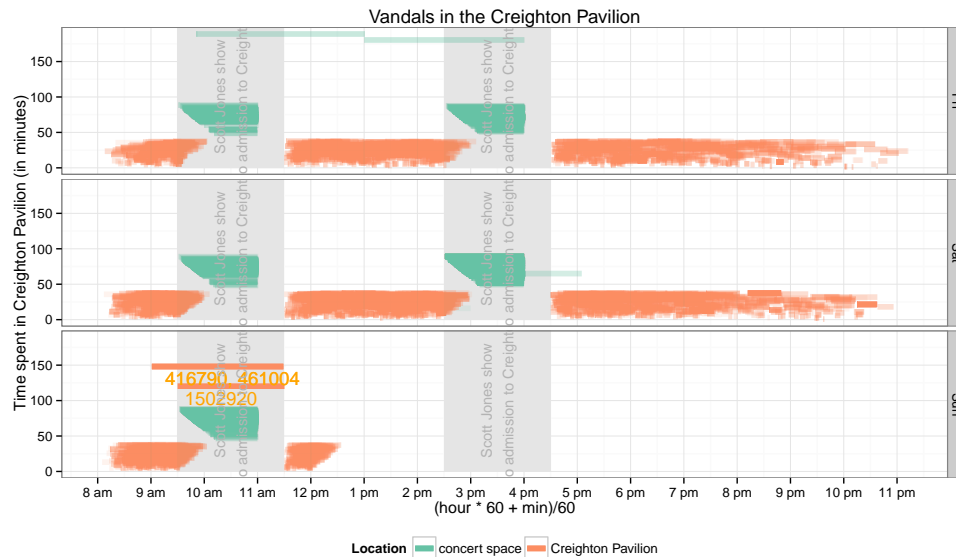
Fig. 1. Vandals!!!

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: park.csv

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.

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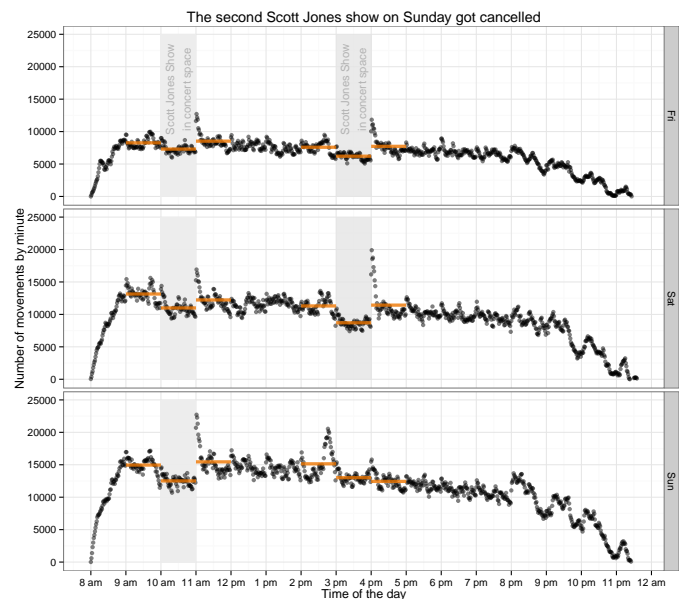
- *Heike Hofmann, Andee Kaplan, Eric Hare, and Vianey Leos-Barajas are with the Department of Statistics, Iowa State University. E-mail: {hofmann, ajkaplan, erichare, vianey}@iastate.edu.*
- *Dianne Cook is with the Department of . E-mail: dicook@monash.edu.*

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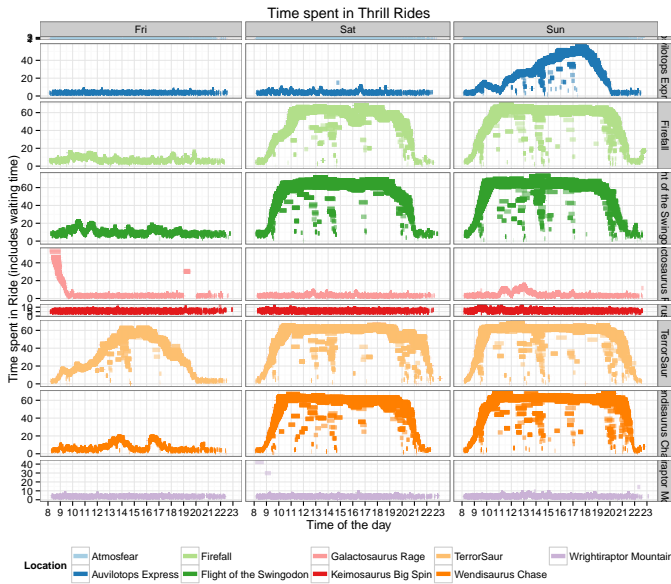
Fig. 3. Time spent in each of the Thrill Rides.

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 define a group as a set of park goers who do rides together.

2.3 Vandals in Creighton Park
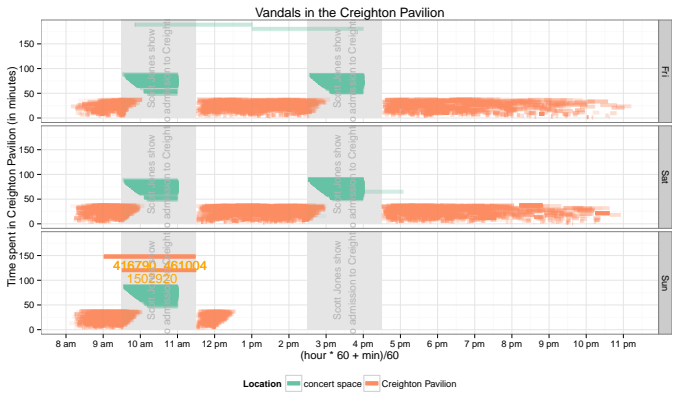
Technical Details Thanks to R [1], and in particular the packages knitr [4, 2, 3], ggplot2, dplyr, and XXX.

REFERENCES

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

[2] Y. Xie. *Dynamic Documents with R and knitr.* Chapman and Hall/CRC, 2013.

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

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