1 Introduction

Capital Bikeshare is a bike sharing system that has been available in the DC metropolitan area since 2008 serving downtown DC, Alexandria, Arlington, and Silver Spring. The service is operated by Alta Bicycle Share, a company that operates several bike sharing systems, including Citibike in New York City and Hubway in Boston. The company makes most per ride data available through their website. We collected a dataset for the full year of 2012 for inspection and analysis.

2 Data Sources

Capital Bikeshare’s open data set includes
• Start and end date and time
• Start and end locations, naming the initial and final station
• Membership type (Casual or Registered)
• Ride duration

Casual users are those that purchased one or three-day memberships while registered users purchased monthly or yearly passes.

3 Highlights and noteworthy findings

After browsing the available data we found that the most interesting discoveries were connected to comparing how casual and registered users make use of the system, and how this is linked to the costs of their rentals. We also discovered that there is a noticeable difference regarding the stations usage, some of them mostly serving as starting point for trips, and others mainly as destinations.

3.1 Registered user ride frequently but pay little variable costs

Registered users accounted for 82% of bike rentals, but only paid 12% of hourly rental fees. This is made possible by Capital Bikeshare’s pricing scheme in which the first 30 mintues of a rental is free, regardless of the user’s membership type. However, each additional 30 minutes incurs fees,
which are higher for casual users. Casual users also tended to rent for longer periods of time, which compounds the effect we see that they pay the lion’s share of the variable costs per rental. If Capital Bikeshare were to try to use this information to increase revenue, they might try lowering the fixed cost (membership fee) for casual riders to encourage more casual rentals.

The following graphics highlight the differences between number of rentals (left) and share of variable costs paid (right) by each of the members types.

Below (left) we see the portion of time that each membership contributed to the total time bikes were rented. That these portions are closer to 50/50 than the portion of rides taken speaks to the high average trip duration for casual riders. We can see that casual rider trip duration in the (right) scatter plot. Notice that the ride duration gap between the two groups persists across every month of the year.
3.2 Capital Bikeshare raked in over half a million dollars in "overdue" bike fees in 2012

As mentioned above, Capital Bikeshare charges its riders by the half-hour. It is in riders best interest to return their bike before the half hour threshold is breached since every 30 minutes longer than one hour is charge an extra $6 or $8 depending on the membership type. (Recall the first 30 minutes is free, and the second 30 minutes is relatively inexpensive). Riders then get the most value from a rental if they return a bike at the very end of a 30 minutes period. Over the course of the year however, millions of rides were just barely late and incurred an extra half hour’s worth of fees for not nearly that much riding time.

We found that over a million rentals were return during the first ten minutes of that rider’s next half-hour, meaning those riders left two thirds of a half-hour’s worth of time on the table. Granted, many of these overages were in the first or second half hour, so to visualize this we filtered out rides lasting less than one hour from the chart below. You can see that around 16,000 riders missed their cutoff by 5 minutes or less, most of them casual riders.
To alleviate the injustice of paying for 30 minutes of riding for only 5 minutes overage, we suggest the Capital Bikeshare program either give a grace period of 5 to 10 minutes or charge users by the minute (13 cents per minute roughly) or by 5 minute intervals (40 cents per 5 minutes).

### 3.3 DC NW has the highest portion of stations with more bikes coming in or going out

For this analysis, we computed a table of the number of bikes that were taken from and returned to each station for all days in 2012.

As the "Bikes in - Bikes out for 2012" chart shows, there is a large discrepancy between the amount of incoming and outgoing bikes at a number of stations. The uppermost stations are those most in demand for borrowing, while the those at the bottom are receiving more bikes. This is a fundamen-
tal issue for the system reliability and its administration, and it shows that some form of load balancing on the part of the Capital Bikeshare program is required in order to maintain the system's stability and reliability.

From this visualization we managed to get two coordinated views showing the top 20 unbalanced and the top 20 balanced stations. Although both coordinated views seem similar one to each other, it is important to notice that all the stations are different and that range of the $x$ axes are quite different; one ranges from -45 to 45 and the other from -4,000 to 12,000. From the address names of the stations, we see that nearly all of the most imbalanced stations are located in DC NW while the most balanced stations are less concentrated, though include a number of SE and NE.
4 Spotfire Critique

Overall, we found that Spotfire is a very powerful tool to examine relatively large sets of data, and to generate complex and rich visualization from them. Some of its highlights are:

- The software UI is highly responsive. Each time a section of a visualization is selected or a filter is applied, all the relevant visualizations on the screen are instantly updated. This allows and even encourages the analysts to try experimental approaches when exploring the data.

- The marking feature proved to be very useful, despite taking some time to discover. We were particularly impressed with the ability to nest or
use multiple markings and among multiple visualizations.

- It was very convenient for us to be able to quickly switch to different chart types. Again, this fluidity supported exploration in a way that slower tools would not.

- Since we updated our data set several times, even adding new columns to it, it was very practical to be able to replace it while maintaining all the previously created visualizations.

We did encounter some usability issues with Spotfire however. They were:

- If a subset of data is selected in the heat map view, it is difficult to unslect it especially if the heat map is in ful screen mode.

- Spotfire attempted to guess the data type of one of our columns to be integer, presumably based on the first 100 or so records. This was incorrect and led to some confusion before we realized the issue.

- Many times when adding a dimension to a chart, Spotfire would apply the SUM() filter to it, even when this was not a frequently chosen option.

- Our workflow tended to be pre-processing using standard SQL queries, visualize in Spotfire, determine other useful pre-processing, repeat. While we later found the "add calulated column" and "add binned column" features, they were still not sufficient for some tasks, or were too cumbersome compared to add new columns or even whole new tables through the database pre-processing alternative.
5 Weather Data

In our proposal we mentioned connecting the bike sharing data with weather data for the same time period. We attempted to find interesting patterns in usage or avoidance but the patterns we found were fairly obvious. For instance, the number of rentals is lowest for the coldest or the rainiest days and is highest when the weather is in the goldilocks zone; not too hot and not too cold. For this reason, we omitted discussion of this data set in our final report.