

Citibike User Behavior

Corentin Llorca (cl3783)

24 November 2018

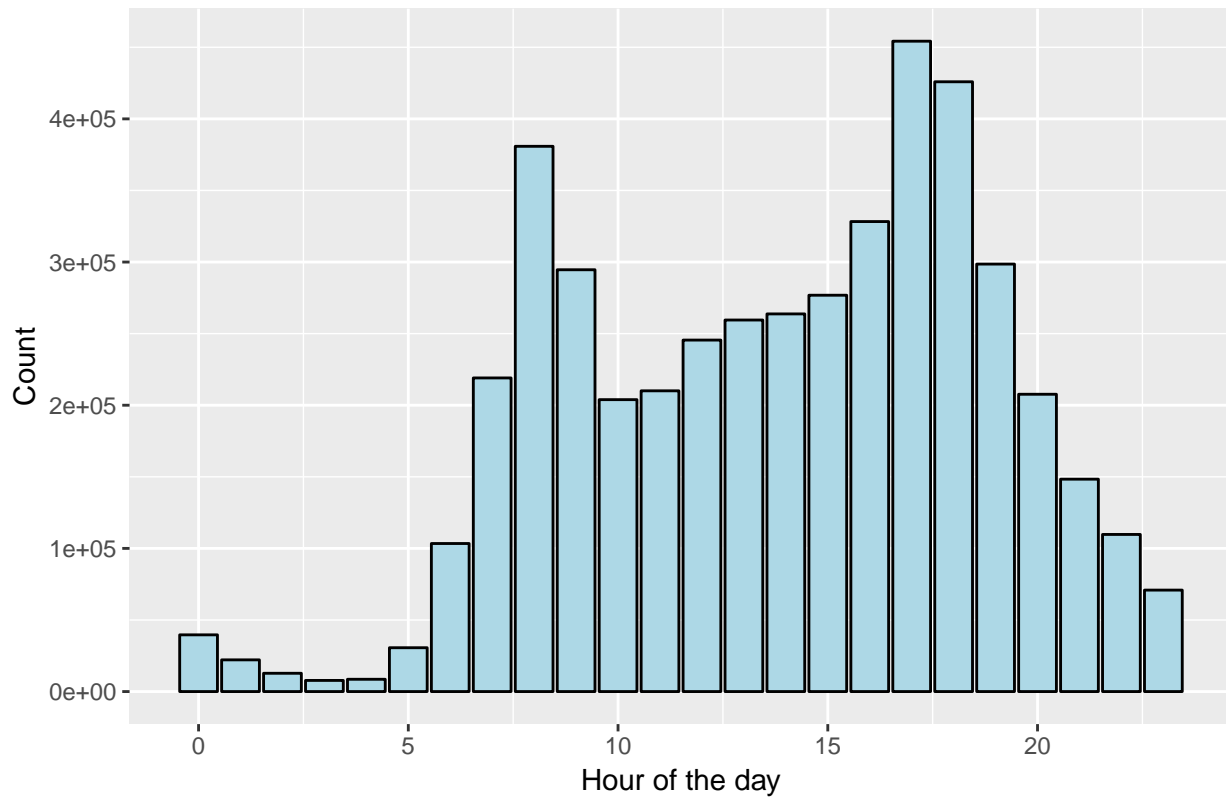
Daily patterns

Here, we want to observe a few daily behavioral patterns in Citi Bike users.

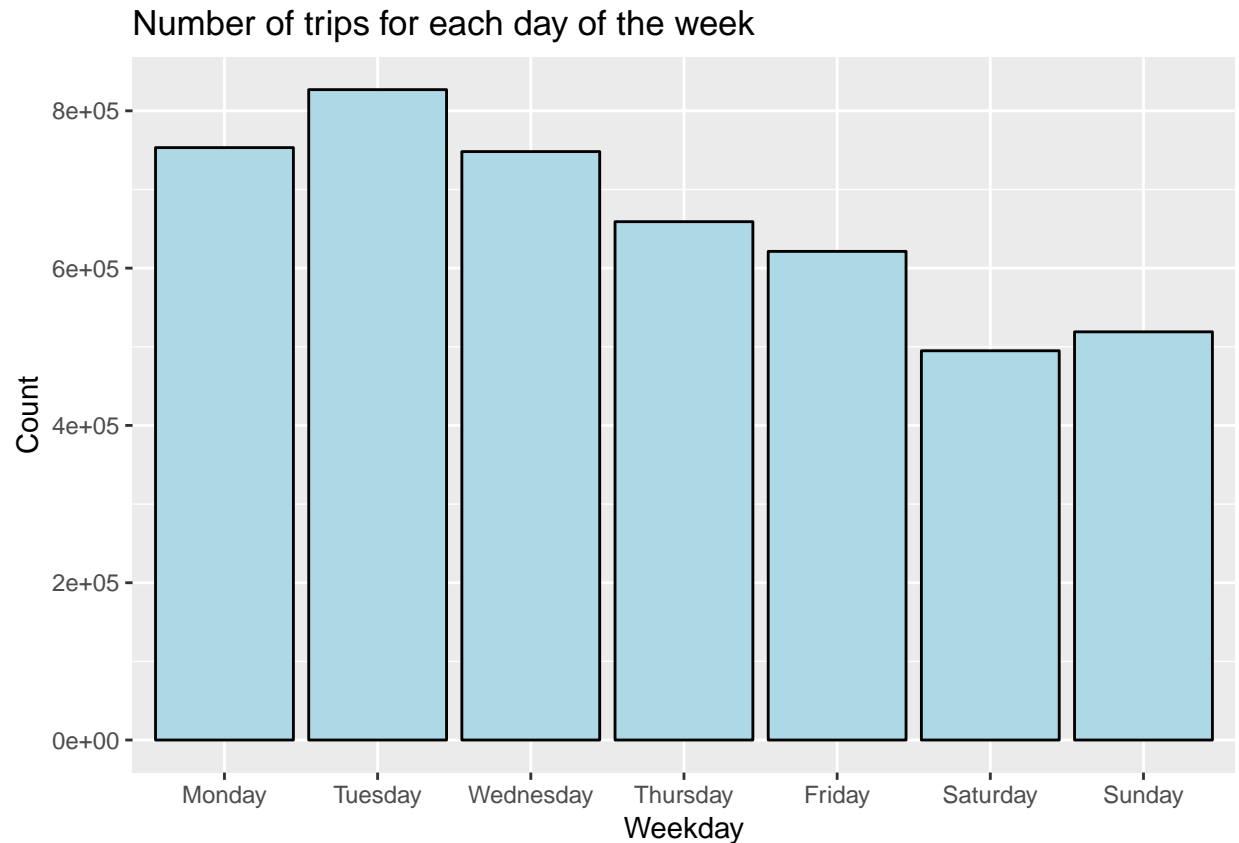
Do people commute with Citi Bike ?

Naturally, the first question to ask would be the following : at which time of day do people travel ? We try to answer this by plotting the distribution of the 'starttime' variable across 24 hours. We use the starting time of the trip to approximate the time at which the person is traveling.

Repartition of daily trip times



Here, we notice something that we were expecting : we observe two modes around 8 a.m. and 6 p.m., which correspond quite precisely to working hours. This seems to indicate that people do use Citi Bike in order to commute. In order to explore the question further, we might want to plot the number of trips taken per day of the week.

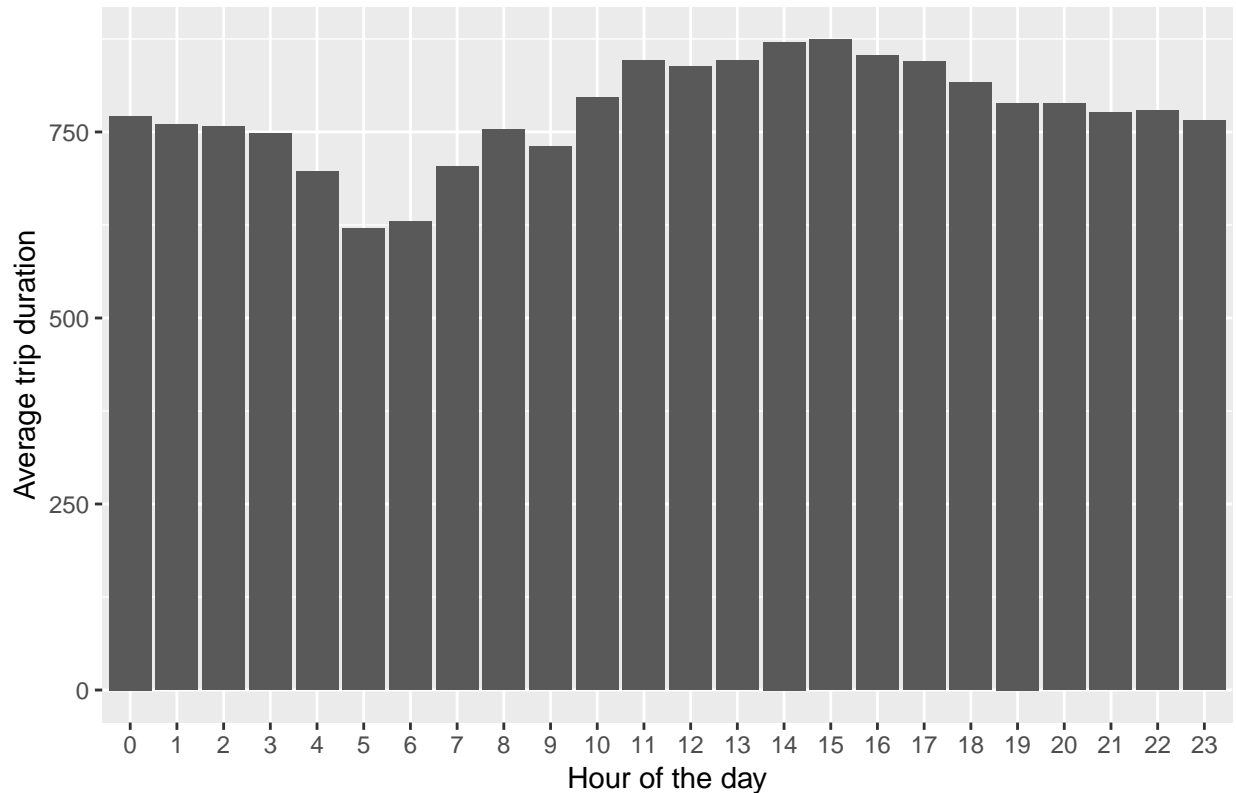


Here, we see that there are less trips being taken during the weekend, thus strenghtening the hypothesis that most people use Citi Bike to commute. However, both plots did not show a significant difference between working hours / days and the rest of the hours / days, meaning that there is still a large part of people who might use Citi Bike for leisure.

What kind of trips do people take ?

With that question, our aim is to understand if people take trips to go to work, to visit friends, or even simply to bike around the city. In order to draw conclusions, we'll first be plotting the average trip duration for each hour of the day.

Average trip duration per hour of day



Unfortunately, there isn't much of a pattern that we can observe here. We do notice a dip in the average trip duration around 5 and 6 a.m., but that is probably due to fluctuation since it's also around the time of day at which we observe the least traffic. We also see that the average trip duration is slightly higher during the start of the afternoon, which would correspond to trips taken for leisure and not for commuting. This checks out, if we estimate that trips made for commuting would be shorter because people would be in more of a hurry.

Who are customers and who are subscribers ?

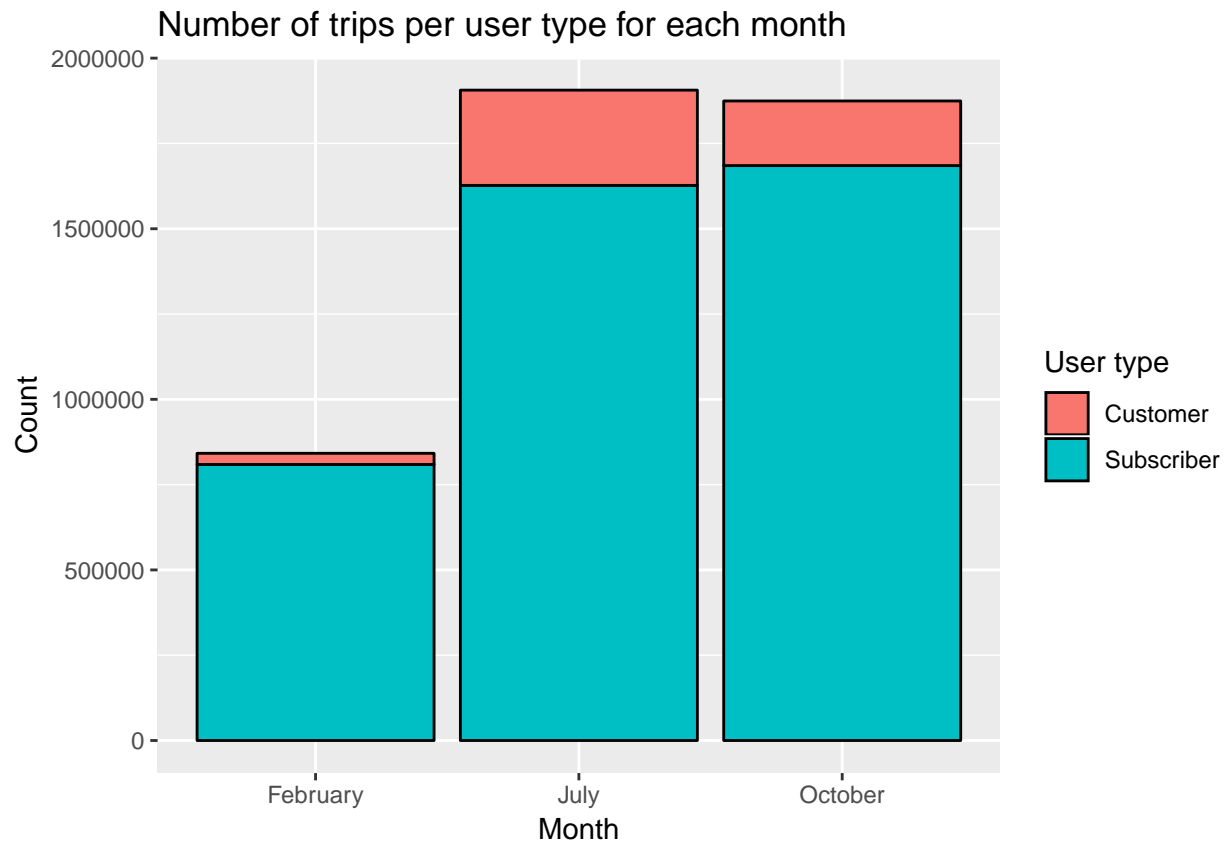
Here, we will be looking at the "Customer Type" variable in the trip database. It is an unordered factor corresponding to the type of customer having taken the trip. The different levels are "Customer" and "Subscriber". The Customers are defined as the people using a 1 or 3 day pass, thus more likely to be occasional users, and the Subscribers are those with the annual pass, thus more likely to be regular users.

The two levels are highly unbalanced, as we can see below :

```
## # A tibble: 2 x 2
##   usertype      n
##   <chr>      <int>
## 1 Customer   501303
## 2 Subscriber 4121147
```

We would expect a service like Citi Bike to be used by subscribers much more than by occasional users, since the annual subscription (170 dollars) is much more cost-efficient than the 1-day pass (12 dollars). We observe a roughly 10%/90% repartition. However, the size of our data (more than 4.5 million trips) makes it so that we can still safely assume that the trends observed among customers are actually meaningful.

Differences in seasonal usage

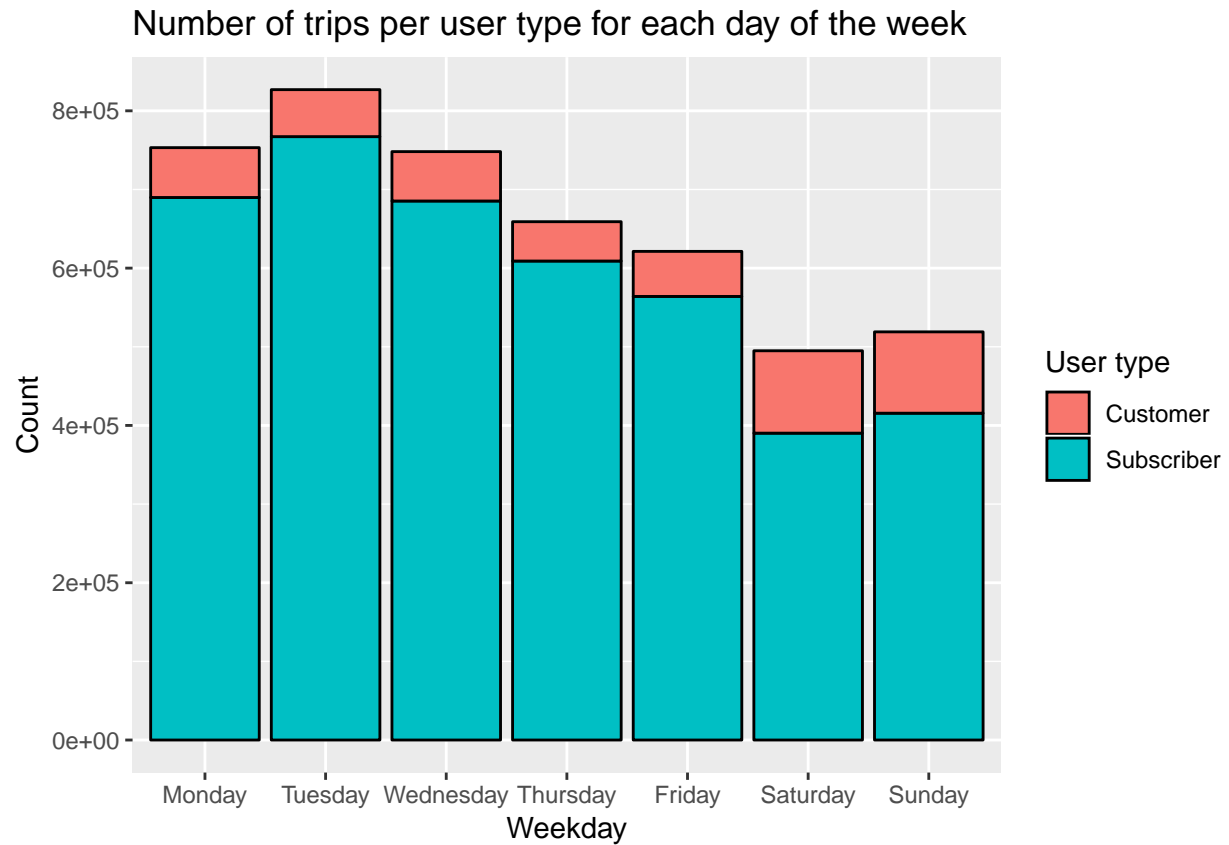


First of all, this confirms that an overwhelming majority of trips are taken by subscribers with the annual pass. As for the variation in the repartition between customers and subscribers, we can clearly see an evolution. We can see that there is a much lower proportion of customers compared to subscribers in the winter than in the summer. Moreover, between July and October, we see the number of customers decrease while the number of subscribers increases: this is a clear indicator that there is a lower proportion of customers in October than in July, although that proportion is still much higher than in February.

This seems logical, as we would expect occasional users to use Citi Bike when the weather is more forgiving. Also, we might emit the hypothesis that Citi Bike is used by tourists - and tourists would in all likelihood be customers rather than subscribers. Then, the increasing number of tourists in the summer compared to the winter might be another cause for that observed difference.

Differences in weekly usage

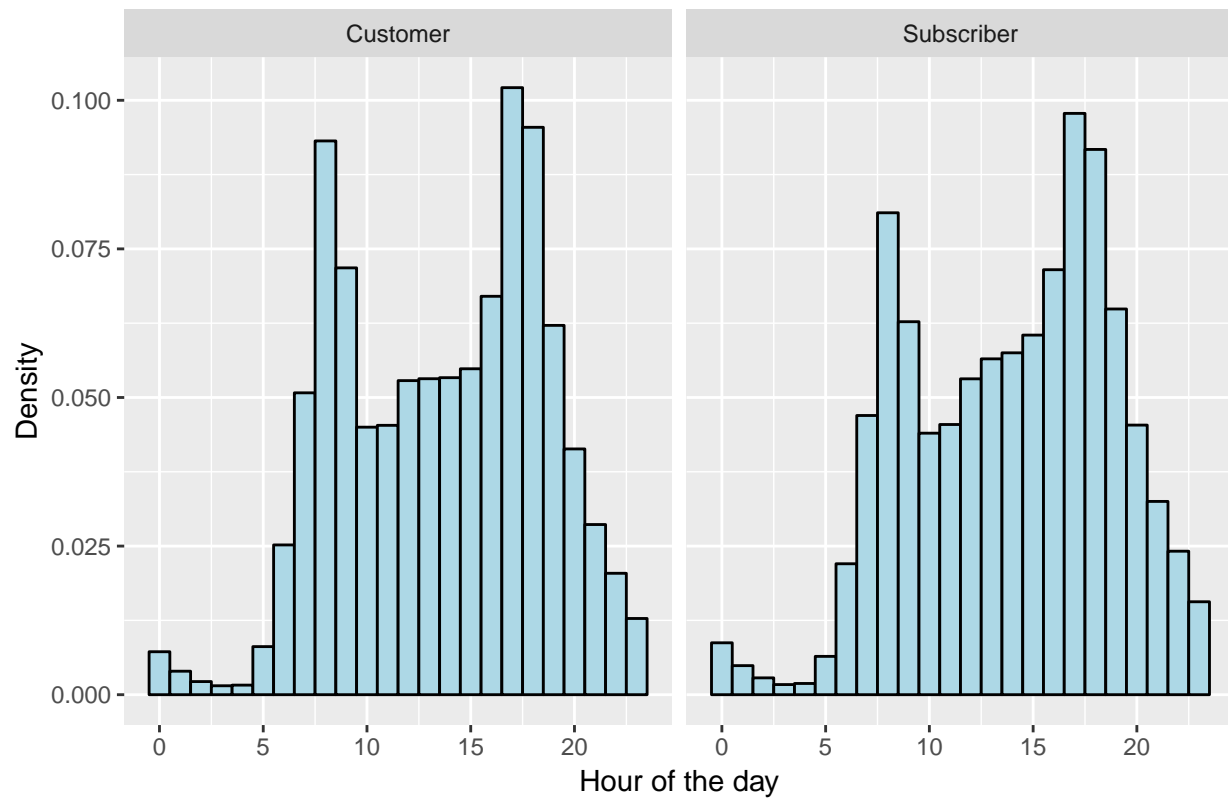
We've already seen that Citi Bikes were being used less on weekends. Let's now see whether the weekly usage differs between the two categories of customers.



We see here that customers actually use Citi Bike on weekends more than they do during the week, while it's the opposite for subscribers. This suggests that subscribers might use Citi Bike to commute to work, while customers use it more for leisure, hence the important use on weekends.

Differences in daily usage

Repartition of trip times for each user type



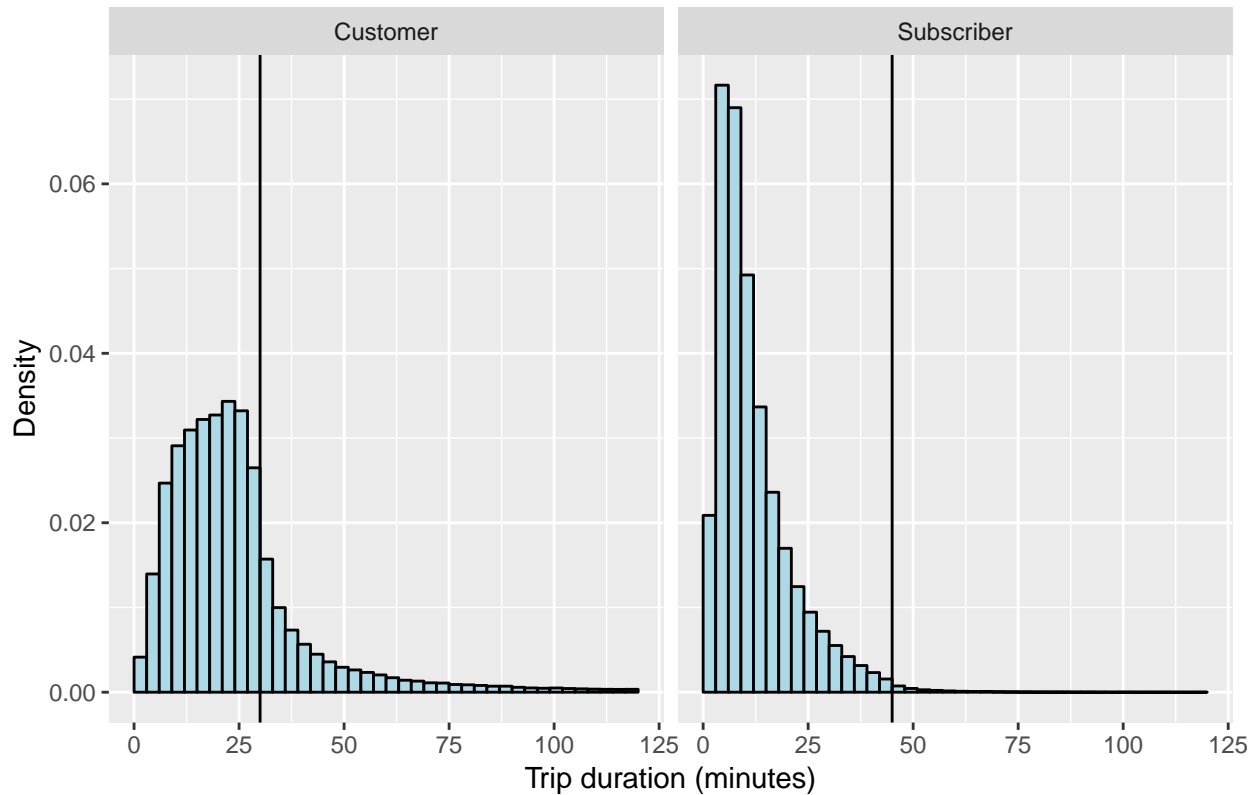
We can see, surprisingly enough, that there is little to no difference in the distributions of the trip hours between subscribers and customers. Although we did expect the subscriber trip times to have two modes around the morning and the late afternoon corresponding to working hours, we would have expected for customer trip times to be more spread out during the day since we assumed that customers mostly used Citi Bike for leisure.

This might for instance mean that customers are actually mostly residents who sometimes purchase short-term passes, but who don't bike enough for a \$170 annual subscription to be worth it. The hypothesis might then be that they purchase passes to commute and not necessarily for leisure, but only during the summer when the weather is nice.

Differences in trip duration

Here, we plot vertical lines for both plots that signal the end of the time included in the plan, which is 30 minutes for customers and 45 for subscribers.

Repartition of trip durations for each user type



This time, both repartitions can be seen clearly, and we notice that they are very different from one another. Indeed, subscribers tend to favor shorter trips, with the overwhelming majority of trips taken being around 10 to 15 minutes long - that might correspond to the time for commuting. Customers, on the other hand, take longer trips, with the mode being around 25 minutes. This can be explained by the fact that we consider some customers to be tourists, and tourists would explore the city much more than residents, thus taking longer-lasting trips.

Moreover, we can notice that customers are generally much keener on taking long trips, going as far as to make them longer than the time that is included in the plan. Indeed, even though we do notice a significant dropoff between the 25-30 and 30-35 minutes bins, a significant part of customer trips are still longer than 30 minutes, with some of them going well past 1 hour. As for subscribers, they usually take very few trips longer than 45 minutes, which is their allotted time limit.

Something we could expand upon later in the project would be a similar type of plot, but also faceted by month, which would reveal the behavioral patterns between different user types in the different seasons.

Geographical differences