

Analysis Change of the Velopass data: the Bird's Eye View

Introduction

We look at the higher level properties of the Velopass data.

Data

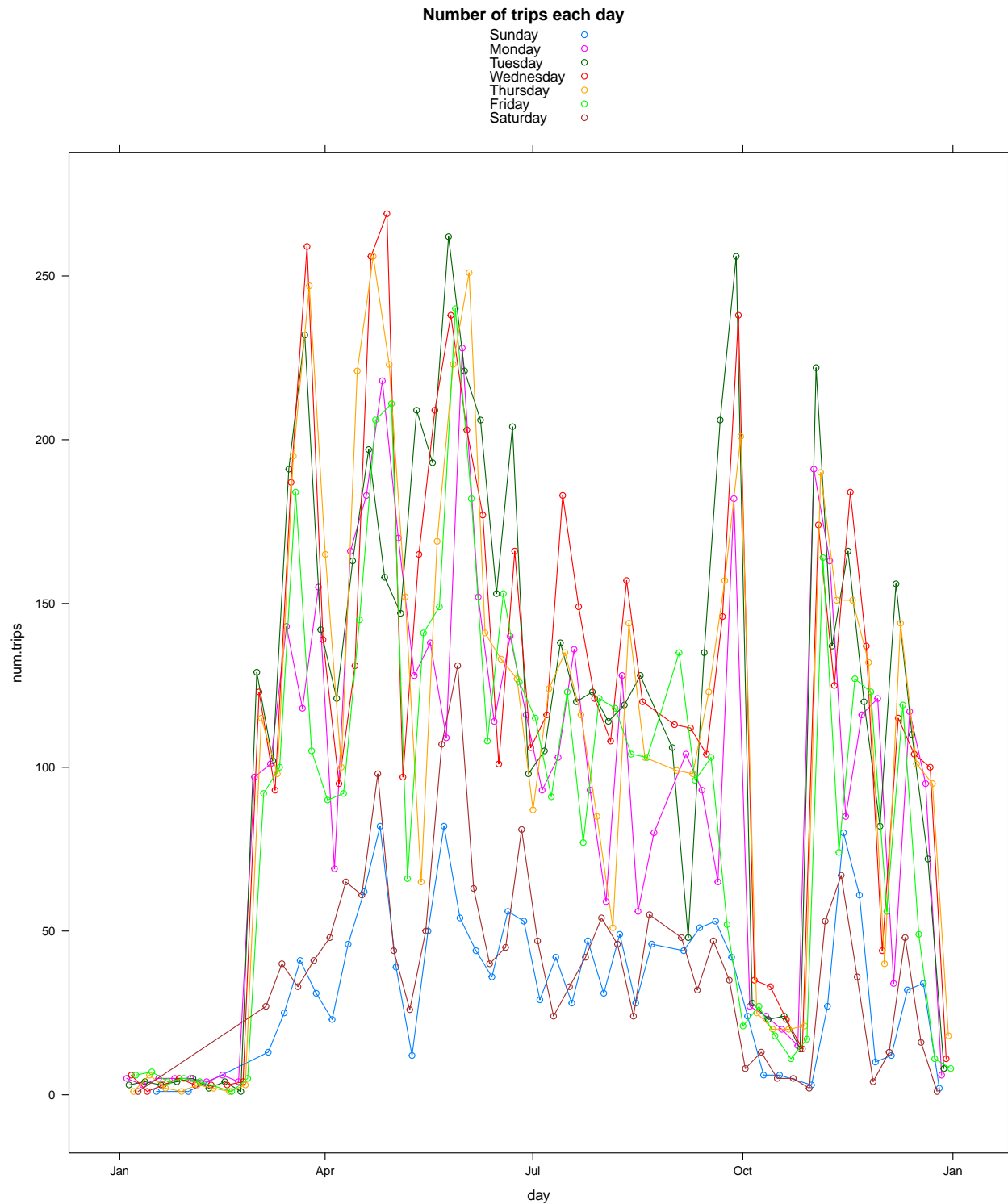
We have written the data into files on the disk, and read them from there,
Once read, we have to convert the dates to date/time objects.

Usage over time

How has the usage changed over time? Lets count the number of trips made every day.
The data frame *tripsByDay* contains three columns,

##	day	num.trips	dayOfWeek	weekdayOrEnd
## 1	2010-01-04	5	Monday	weekday
## 2	2010-01-05	3	Tuesday	weekday
## 3	2010-01-06	6	Wednesday	weekday
## 4	2010-01-07	1	Thursday	weekday
## 5	2010-01-08	6	Friday	weekday
## 6	2010-01-09	1	Saturday	weekend

We can plot the number of trips made on each day of the week, over the duration of the data set



Since people's activity patterns change between the weekdays and weekends, we can aggregate a week's days into these two categories,

We can add a column for the week during which the trip was made, and look at the number of trips by week of the year.

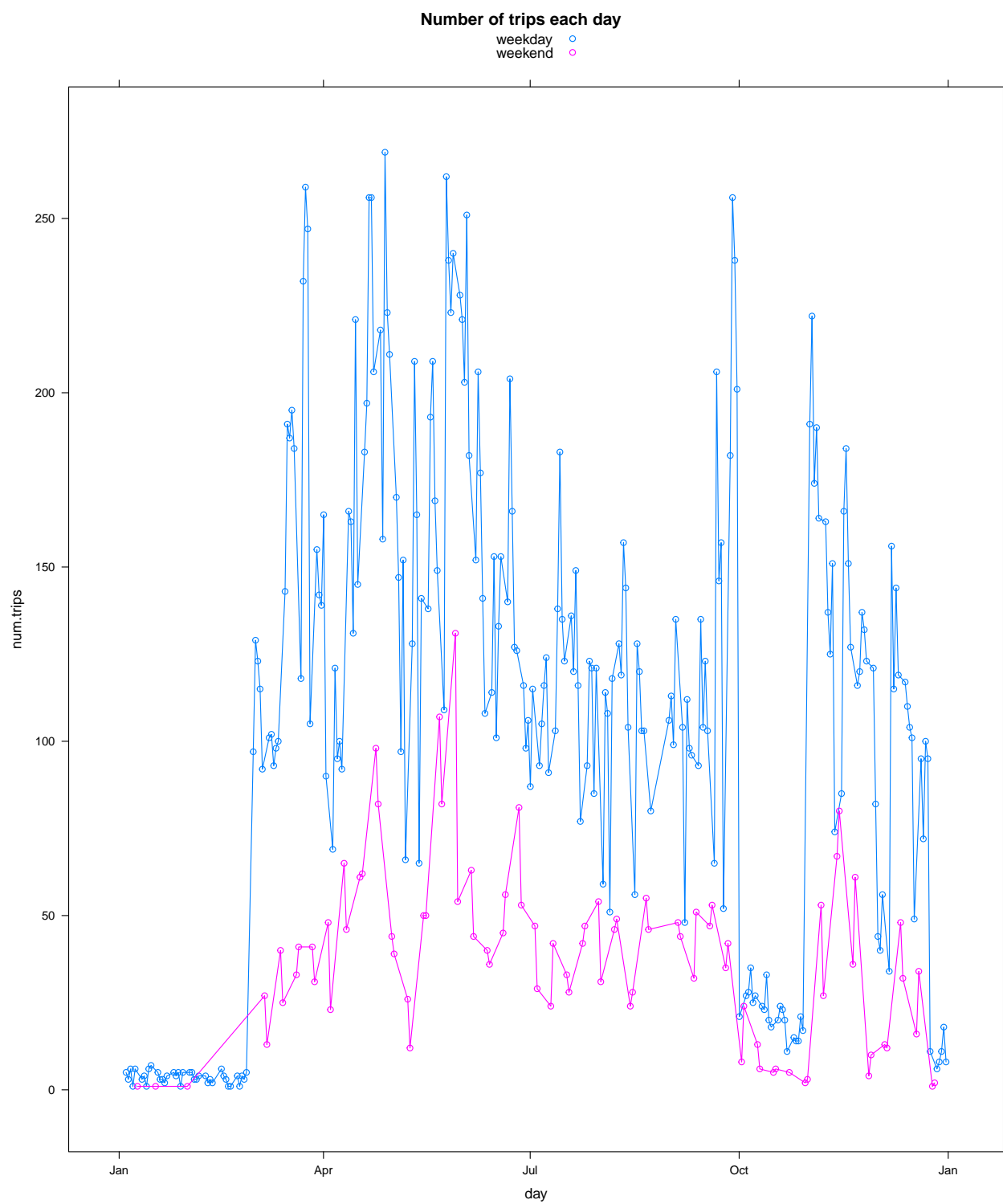
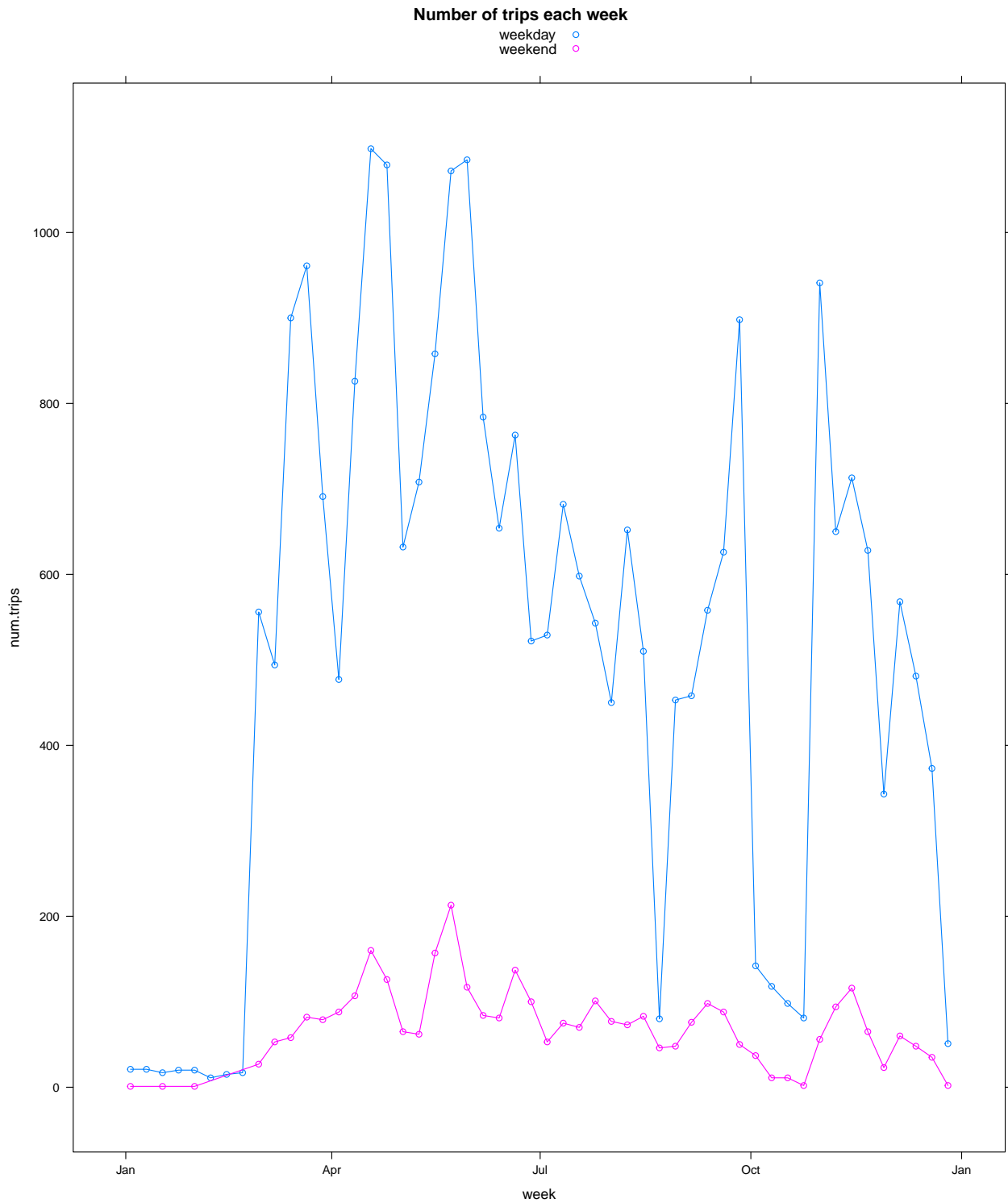


Figure 1:



And do the same by month.

We see significant variation in usage over the months, but there may be a pattern of weekly usage independent of the month. One pattern is that the usage on weekdays is much larger than that on weekends. The usage for individual days is hidden under noise, which we can handle by summing up all days in a month.

We have assumed a normal distribution for the number of trips during a day to plot the error bars as the standard deviation from the mean. However we should expect number of trips in a day to be Poisson

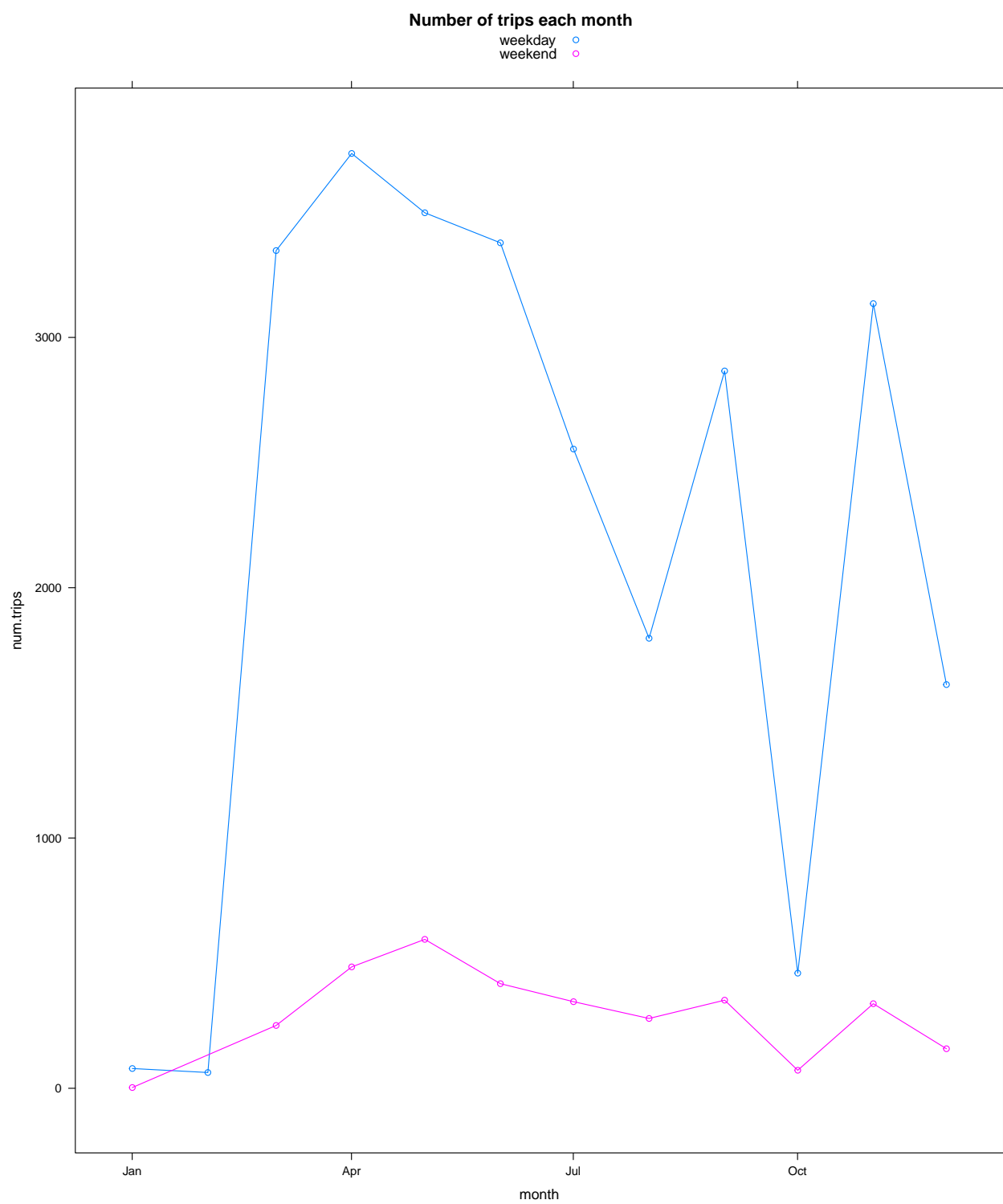


Figure 2:

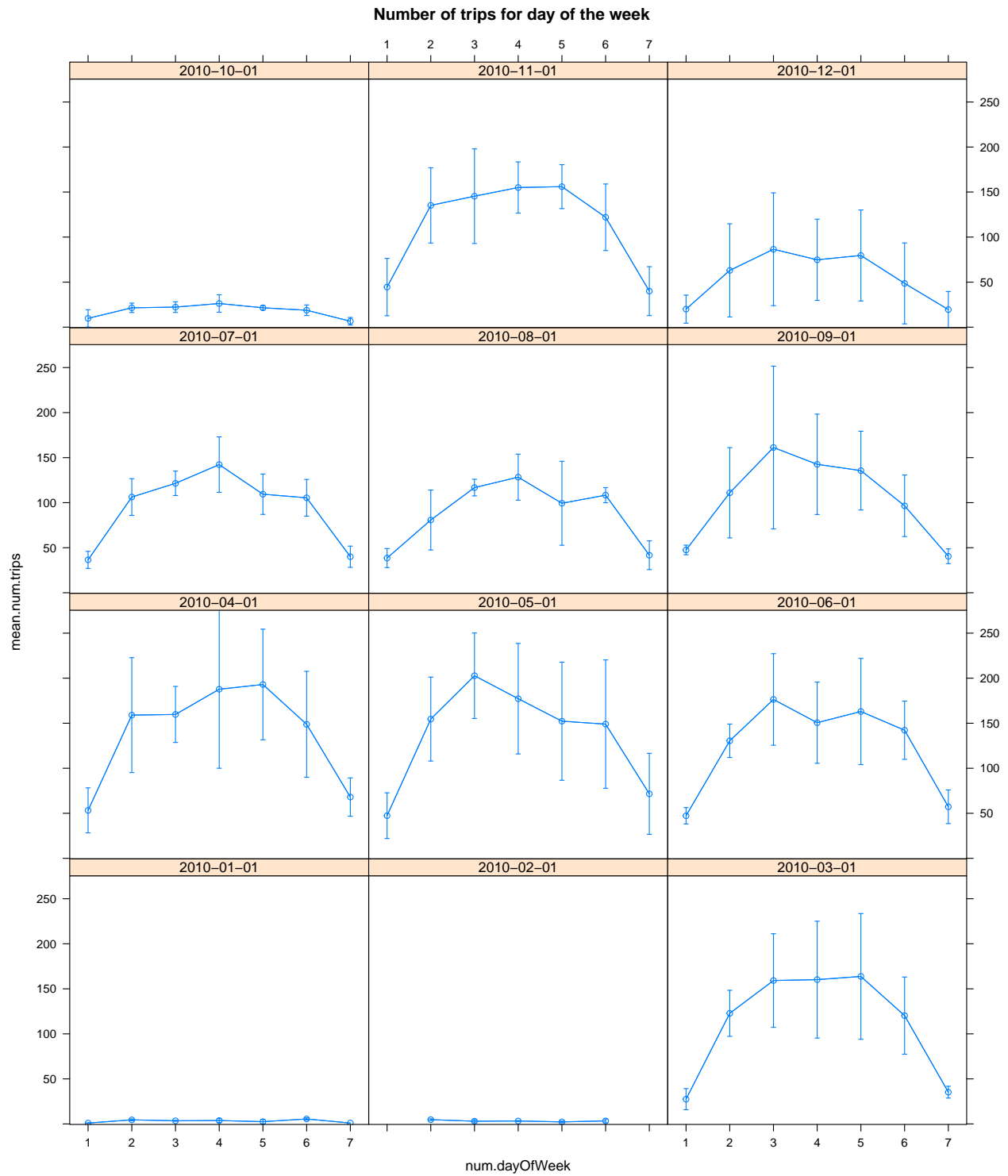
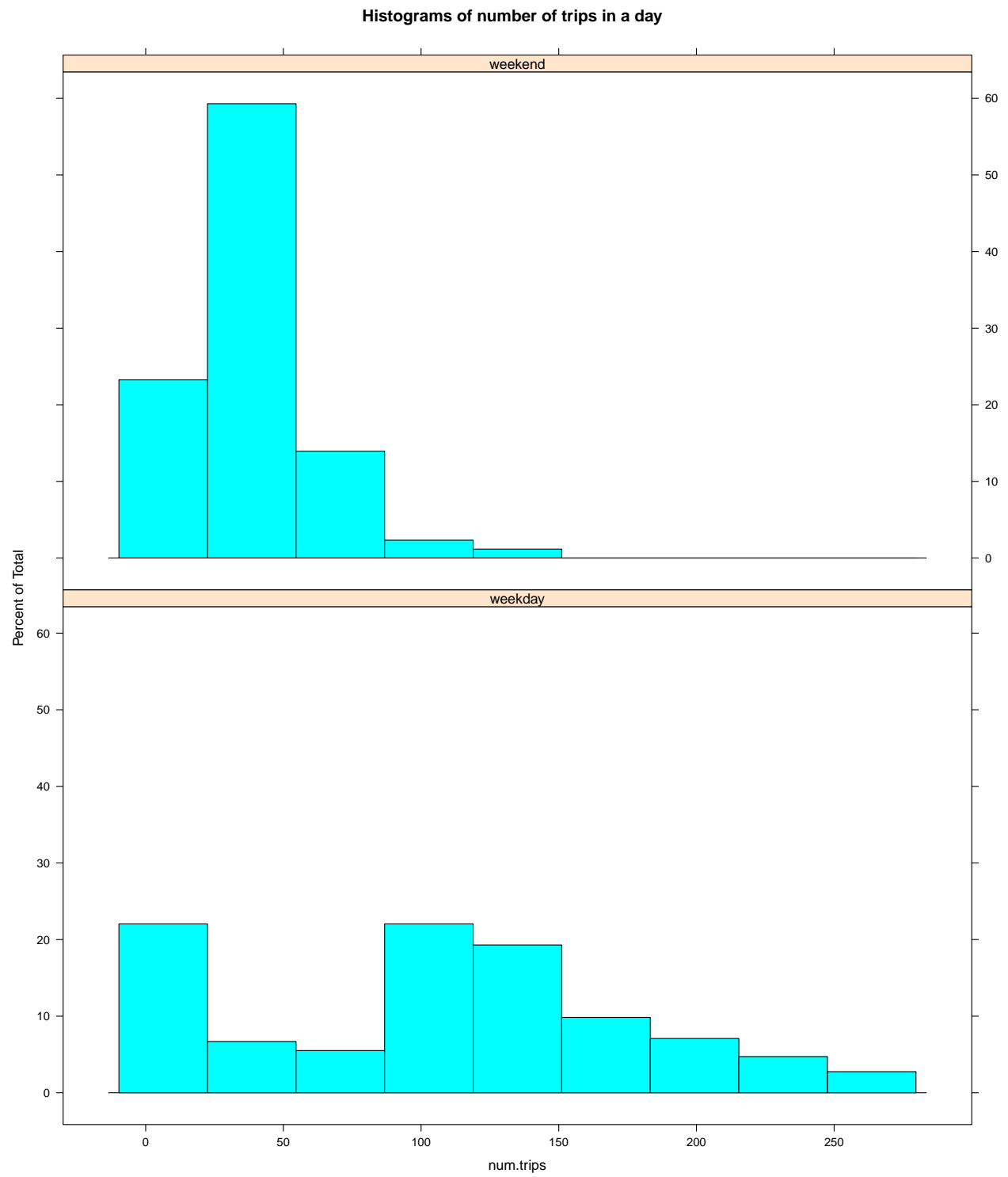
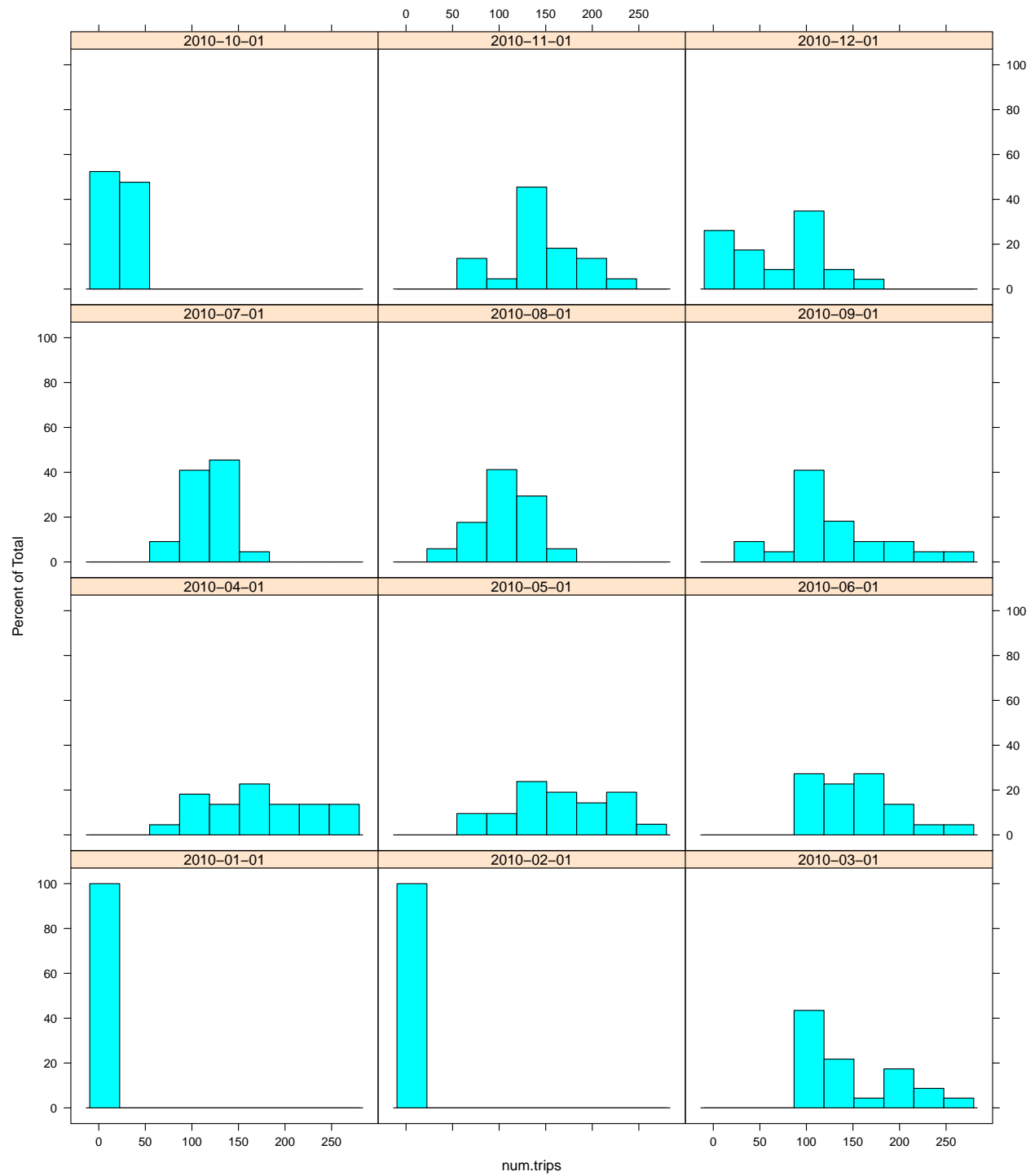


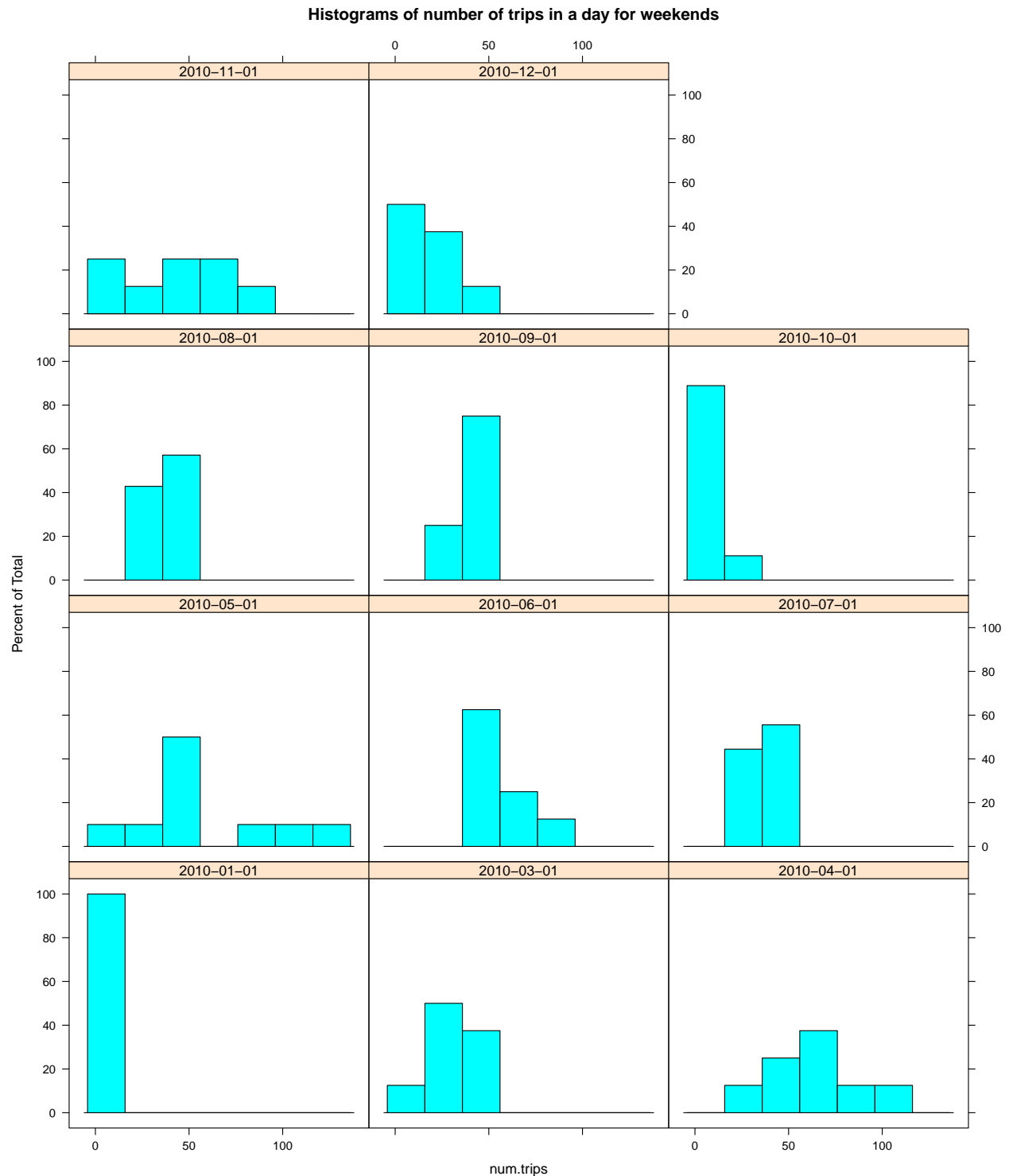
Figure 3:

distributed.



Histograms of number of trips in a day for weekdays



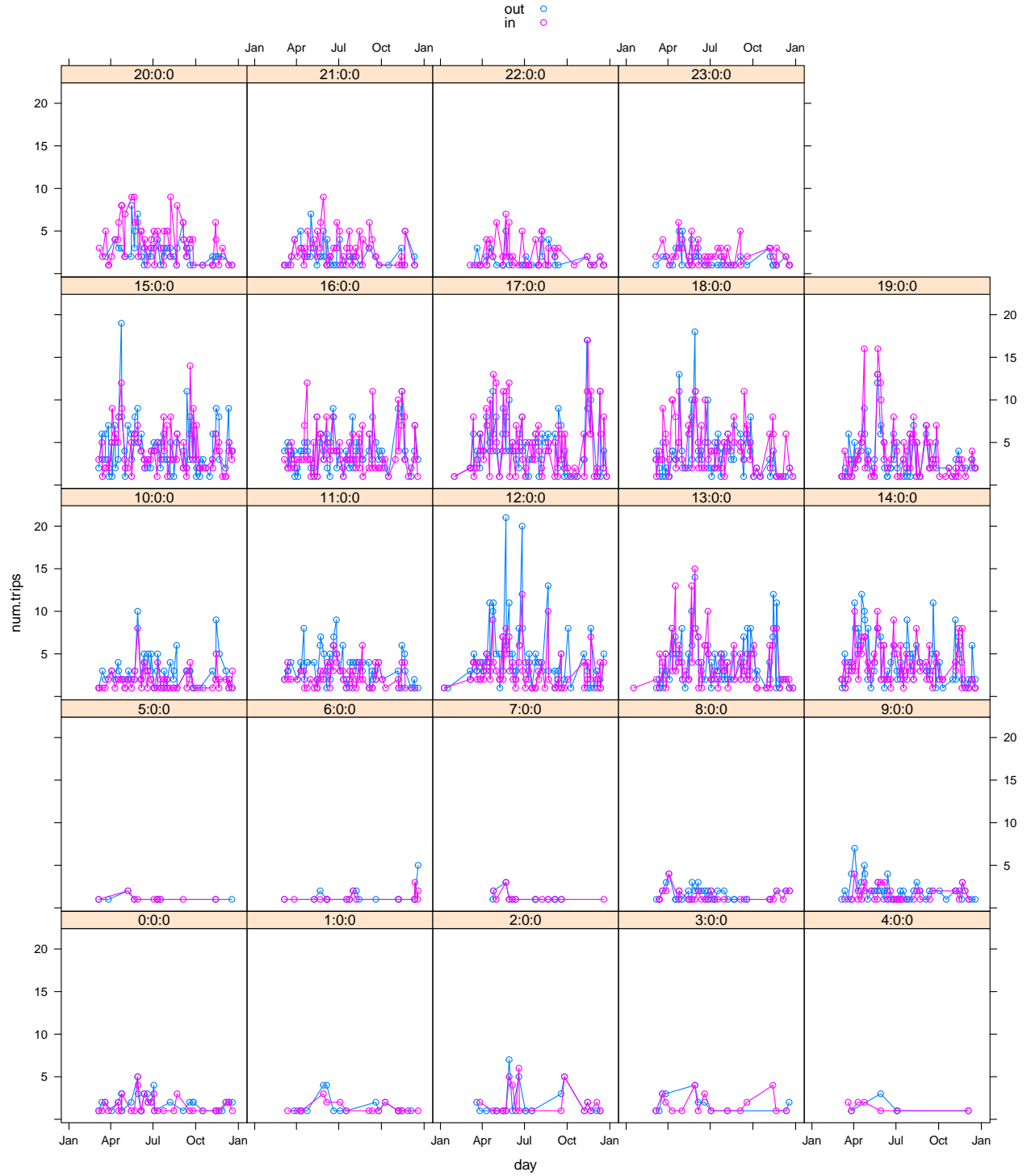


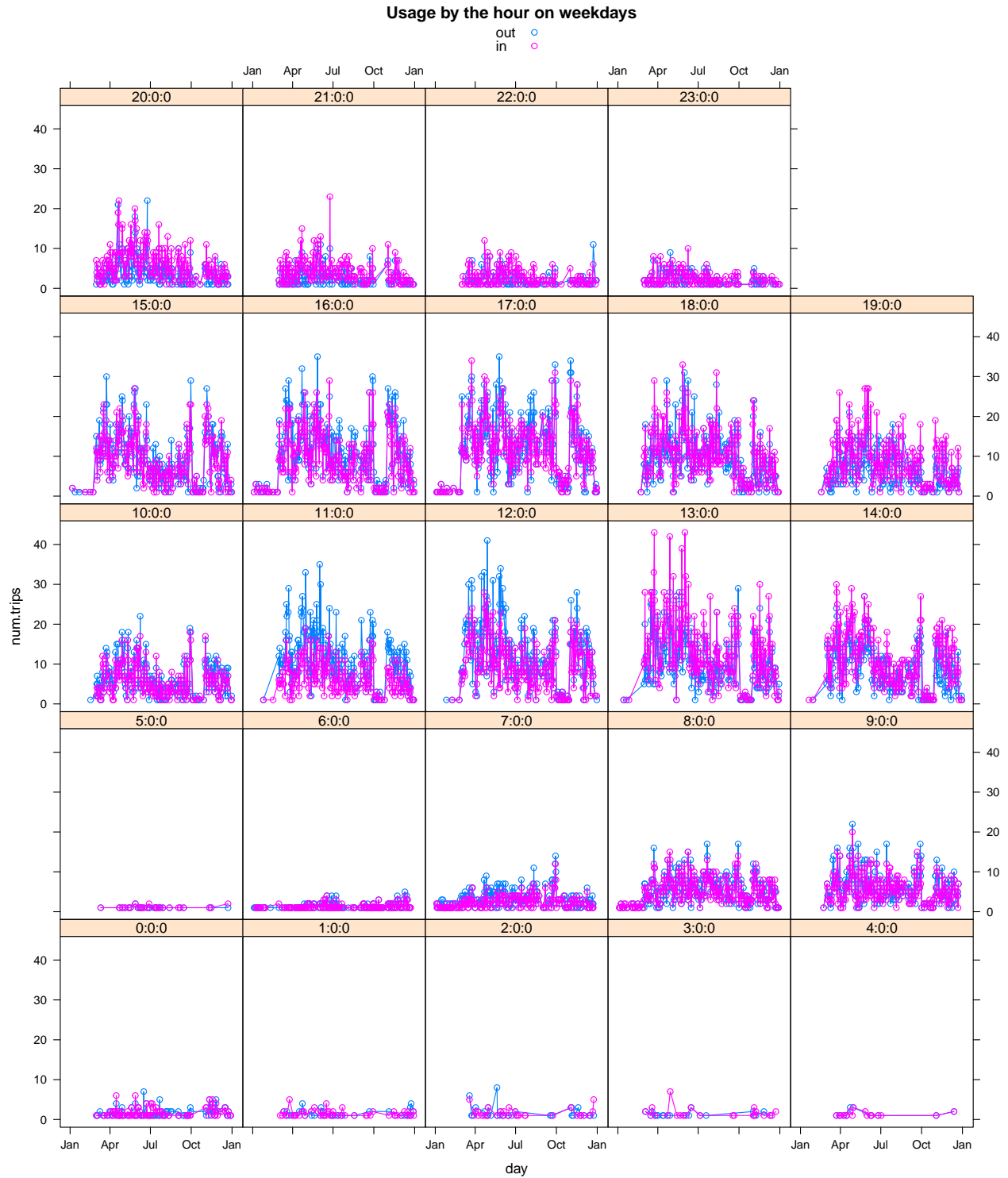
We can look for number of trips during certain hours of the day, summed/averaged over days/weeks/months/years. We first need a function that converts time of day to seconds.

We can use the functions to count trips to find trips by the hour. First we write a function that will get the counts for every hour of the day, and include an offset in minutes.

Now we can make some plots for usage by the hour over the entire time duration of the data.

Usage by the hour on weekends





So far we have looked at the behavior of number of trips in a day. What we would also want to see how many trips are made over a certain period of the day. We can use the data produced by *countTripsHourly* to make these plots as well,

Above we saw total hourly usage for the whole year (all the data in *vlp2010*), and we can aggregate the trips for each month, or for each week as well.

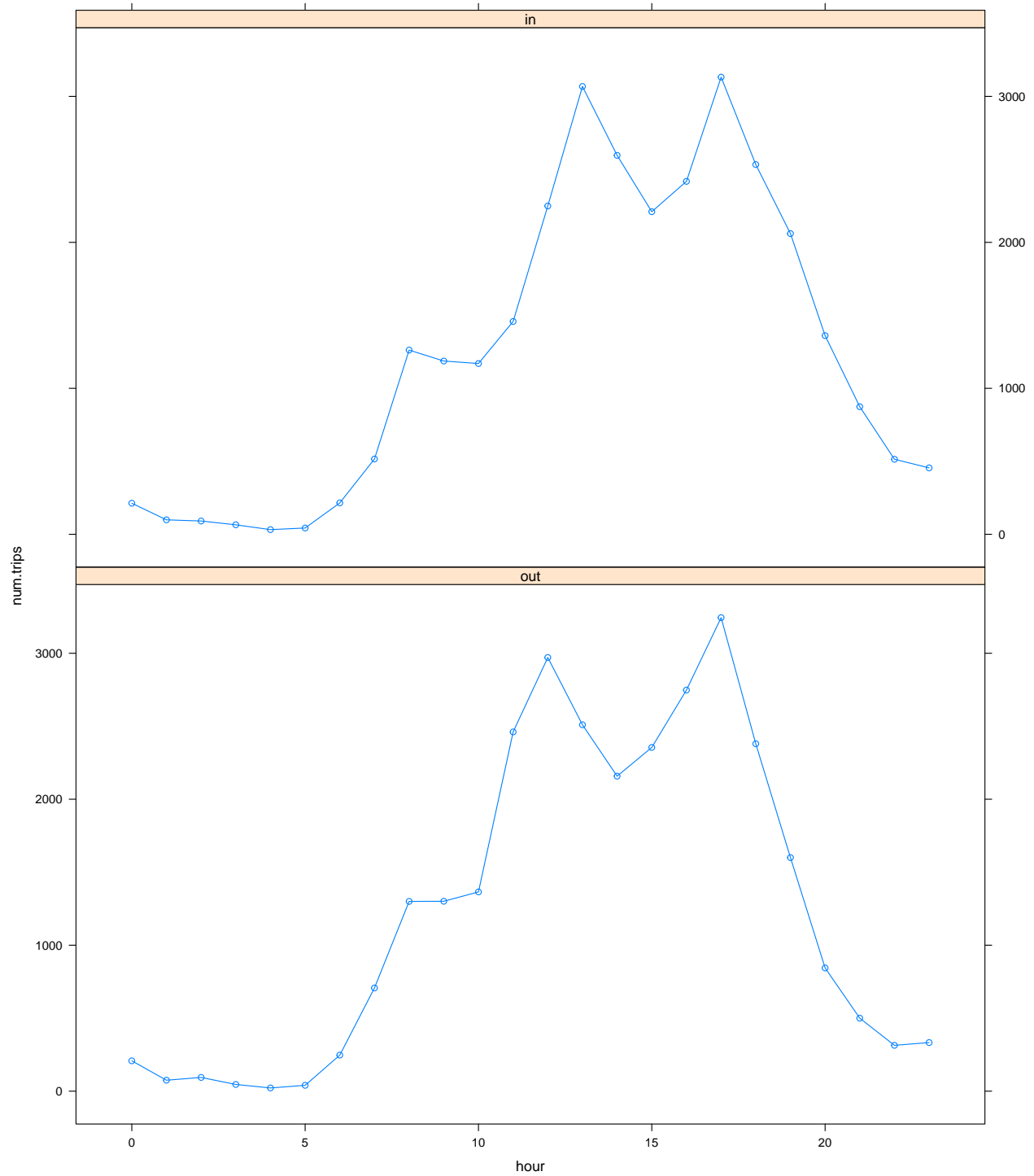
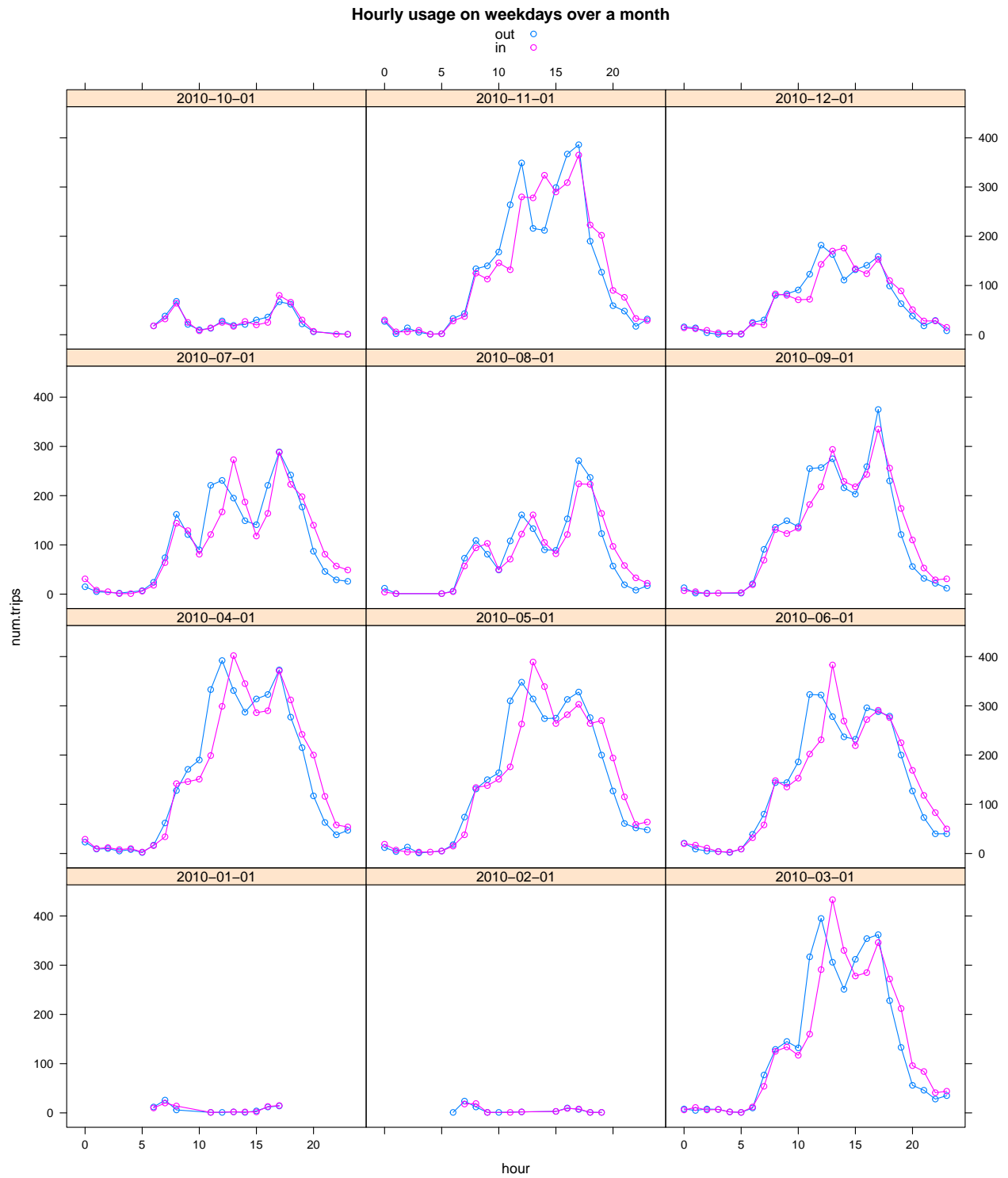
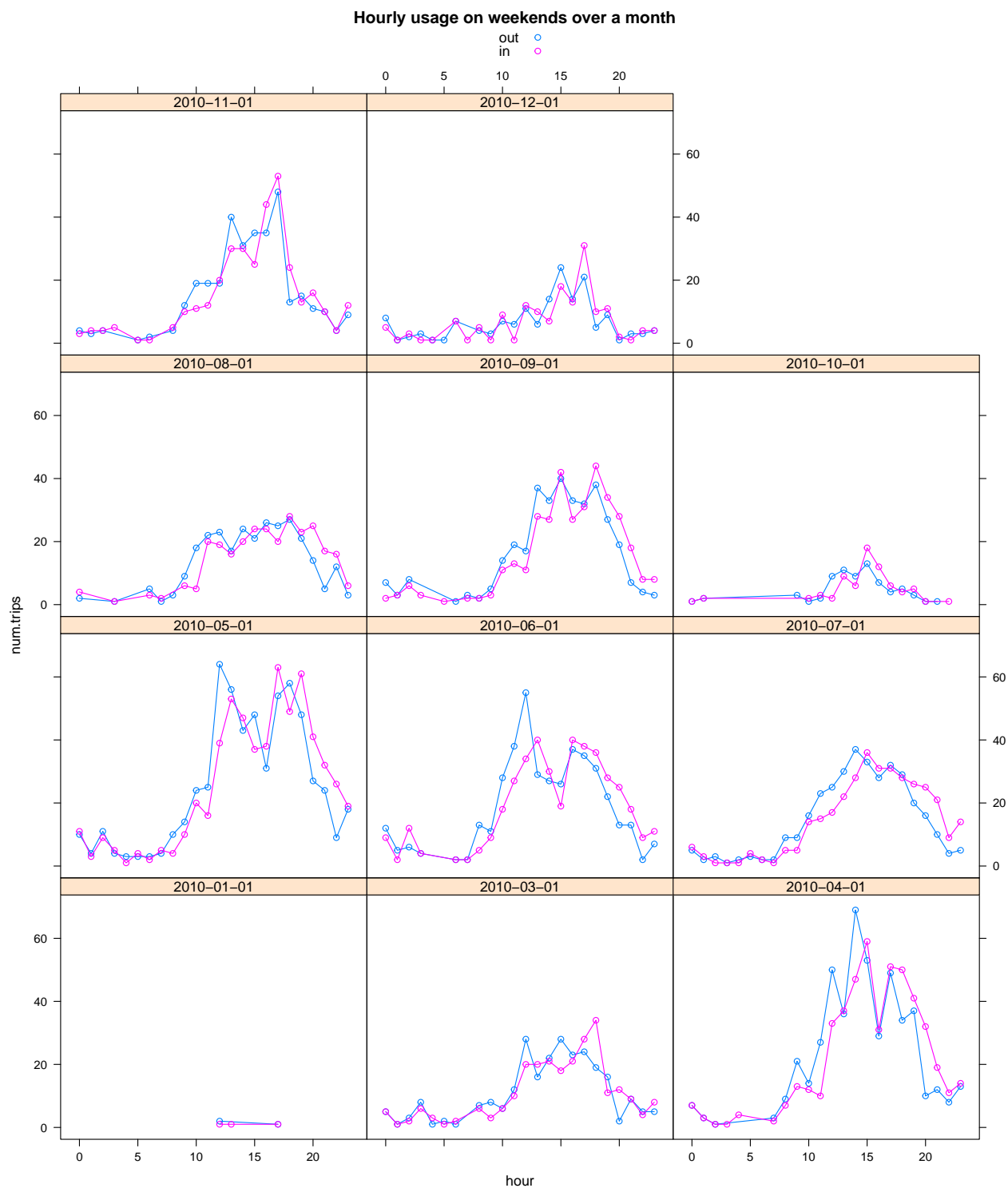
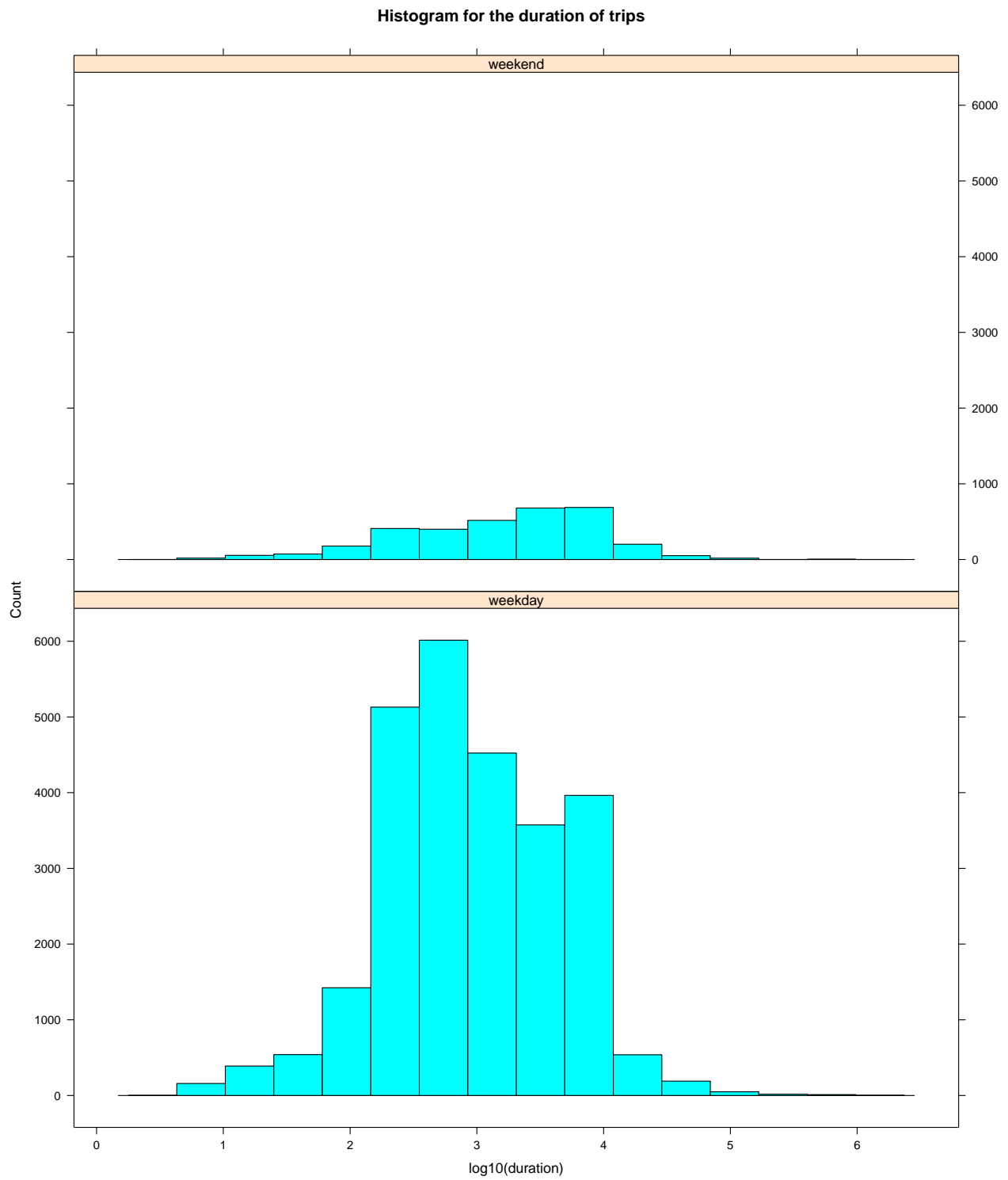


Figure 4:

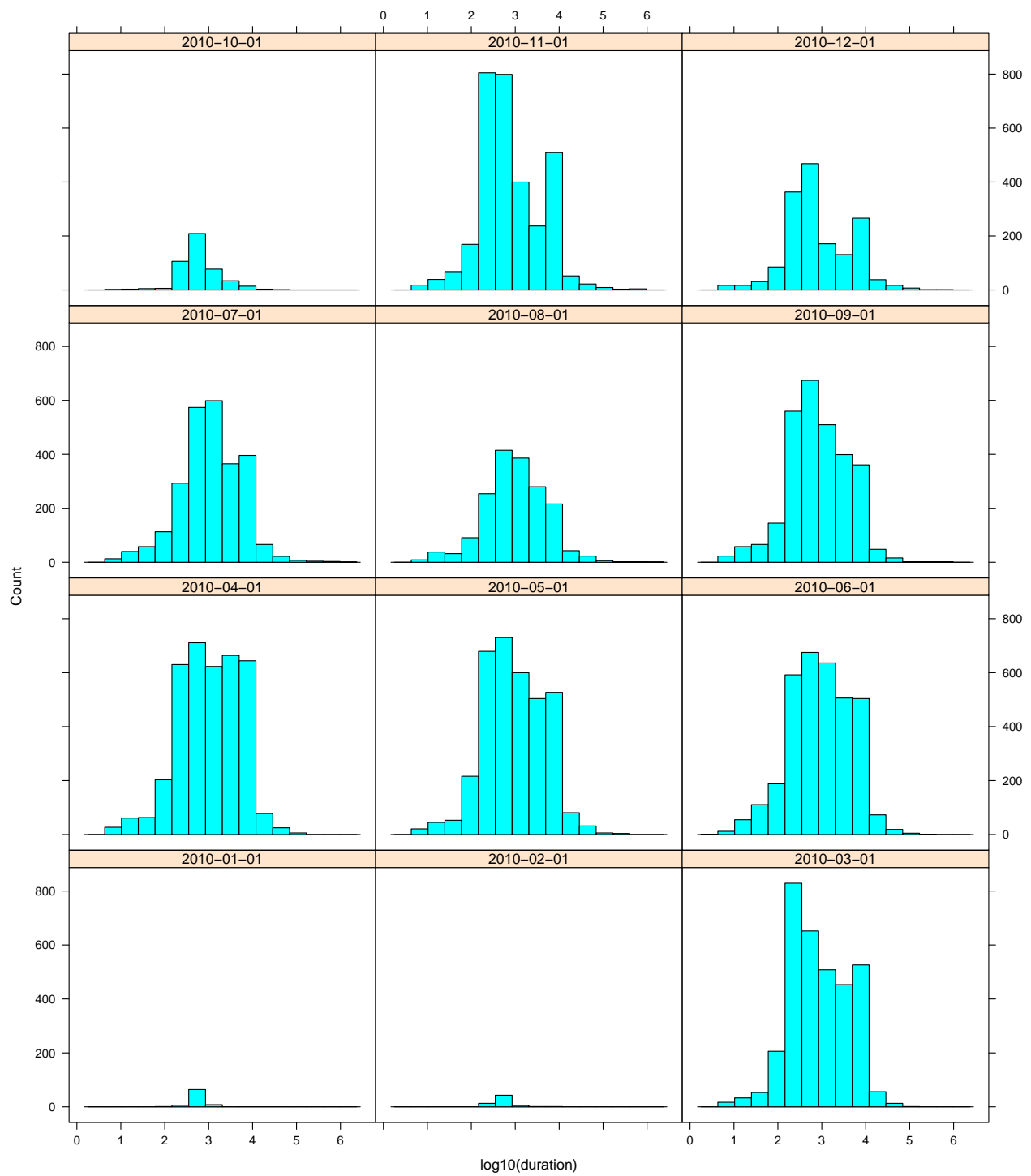




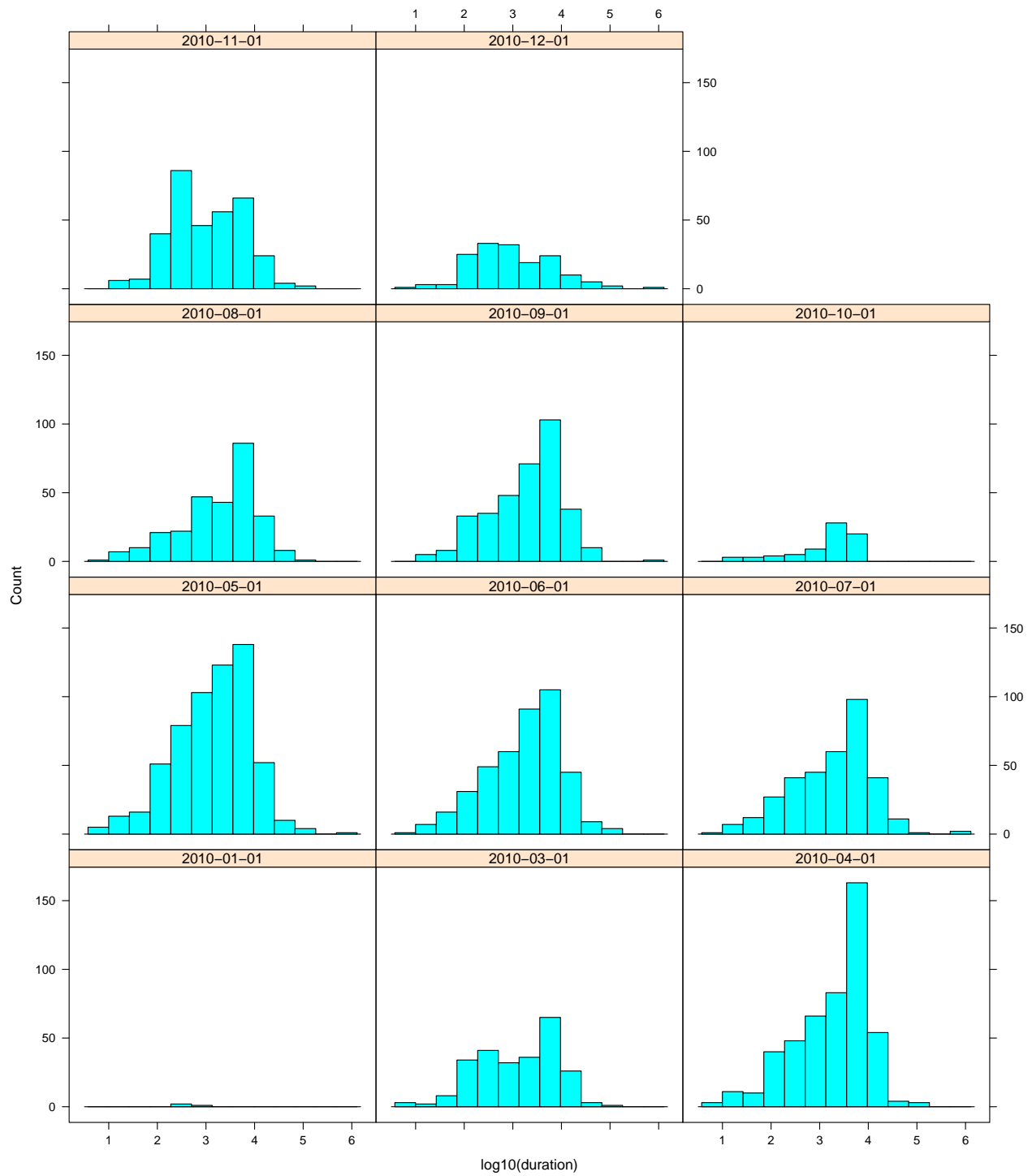
Trip durations



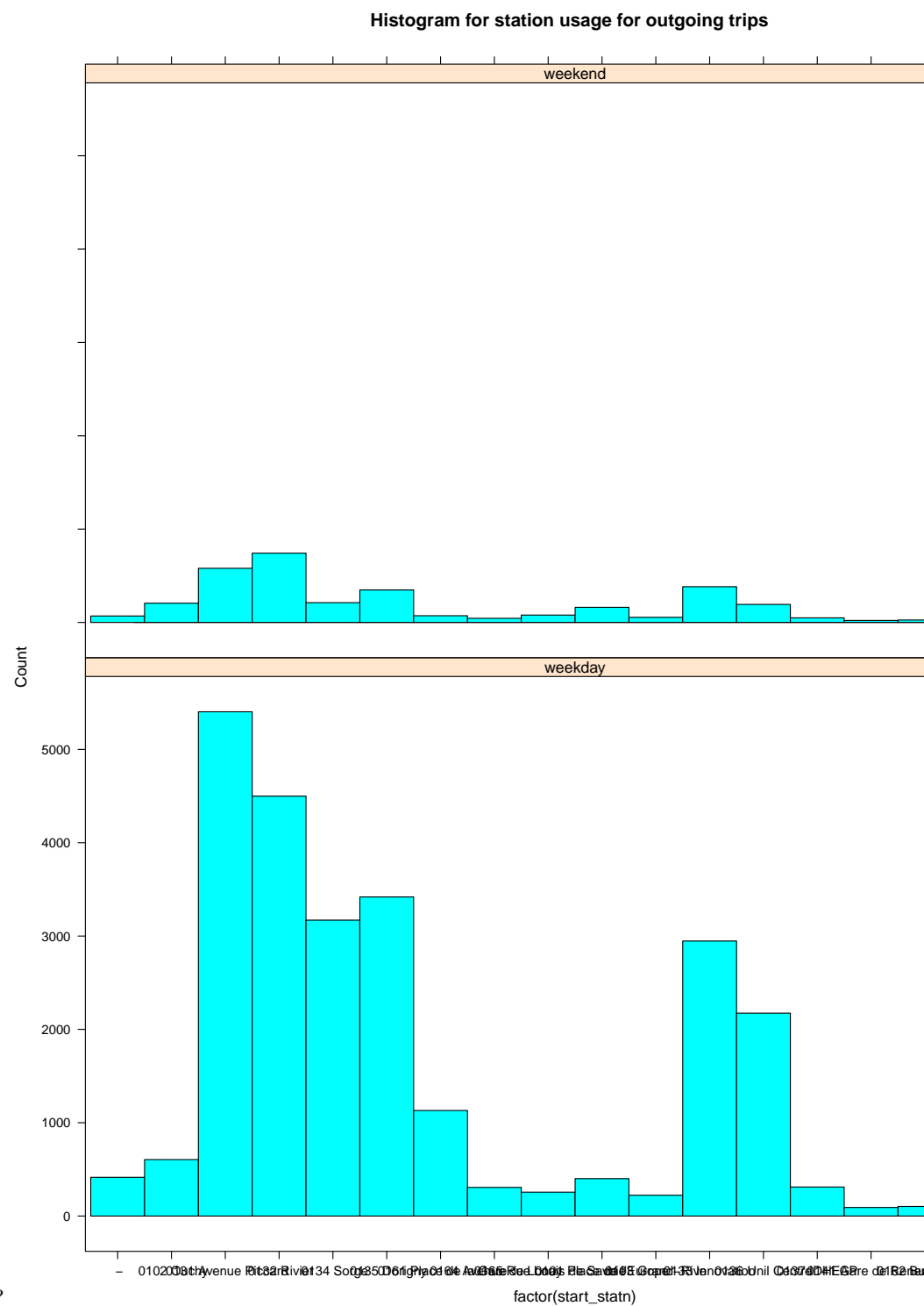
Histogram for the duration of trips on weekdays



Histogram for the duration of trips on weekends

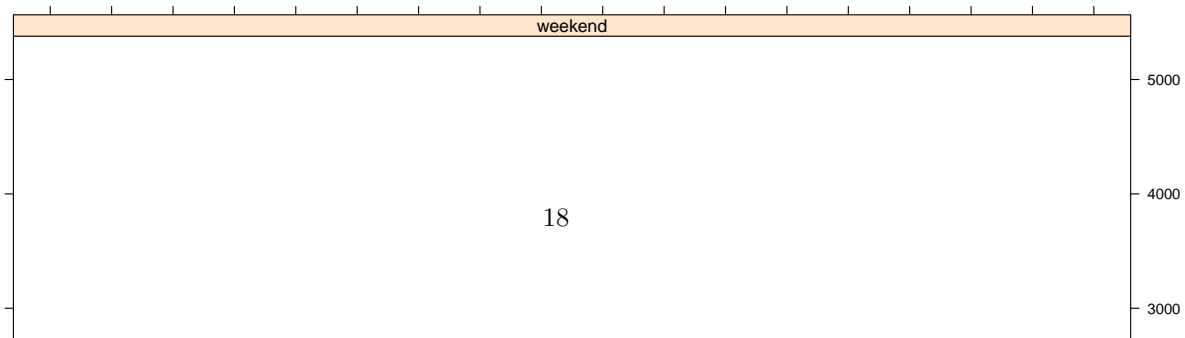


Statistics for stations



What stations are used most often?

Histogram for station usage for incoming trips



Individual station usage statistics is easy. We should also count the number of stations between each pair of stations. For this we can make a table.