Question 1.
Uber has several different products (varying by city); for instance, in New York City, it offers uberX, uberXL, UberBLACK, and UberSUV.

In some cases Uber practices "cross-dispatch"; e.g., generally a driver/vehicle for uberXL can also accept uberX trips. In specific cities, UberBLACK (or equivalent) drivers will be sent uberX requests.

Think about ways in which cross dispatch might make querying Uber's data complicated. Write a paragraph or two about common types of analyses that might fail if the researcher didn't think carefully enough about cross dispatch.

[If it matters, assume Vertica 9.0; feel free to look up the documentation. Don't worry about syntax errors in the queries below, as these queries are known to run]

It is a problem when we measure only by what car was dispatched, we don't see the cars that were called and did NOT pick up a ride. This is something that can be solved for if we create a database that captures this type of data.

Question 2
SELECT driver_id, EXTRACT(week from request_at) as week, count(*)
FROM trips
INNER JOIN cities on cities.id = trips.city_id
WHERE name='New York' AND EXTRACT(month FROM requests_at) = 3 AND
EXTRACT(year FROM requests_at) = 2014 AND status = 'completed
GROUPBY driver_id, EXTRACT(week FROM request_at)

Question 3

Group By 1,2 - it is counting the same driver multiple times. You can have a driver that is technically an "uberblack" and "uberx" and if they receive two requests simultaneously it would be counted erroneously twice.

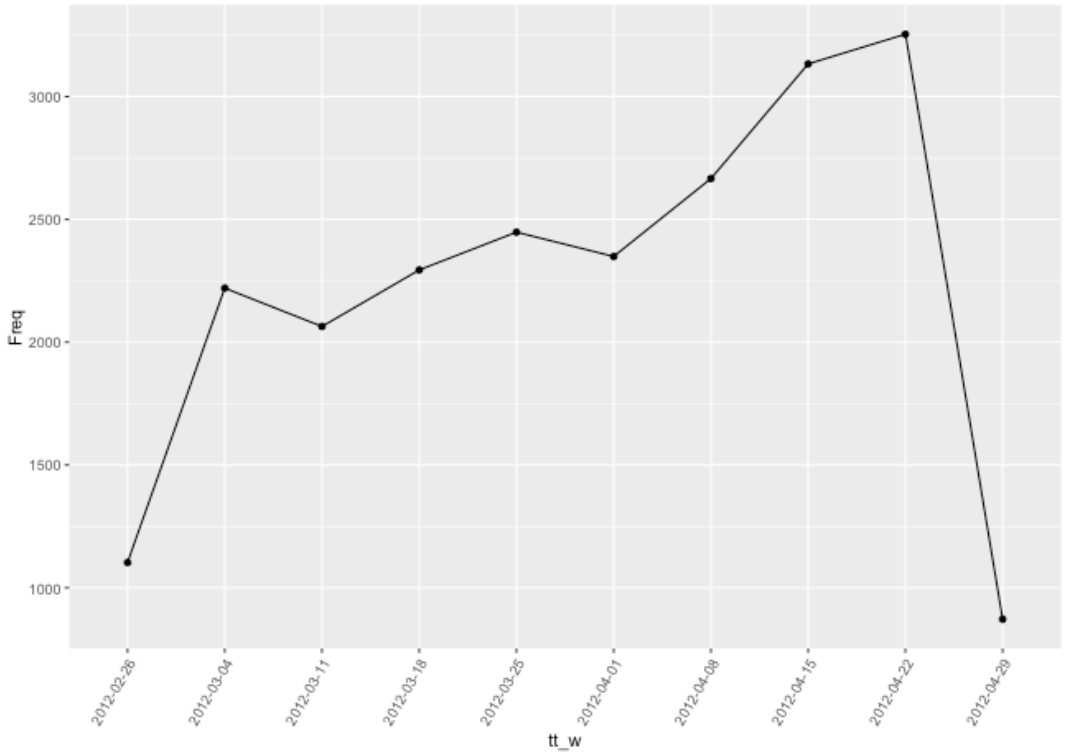
The other error is the amount of time counted for each driver. For every driver we counted, we are increasing their time by the INNER JOIN -

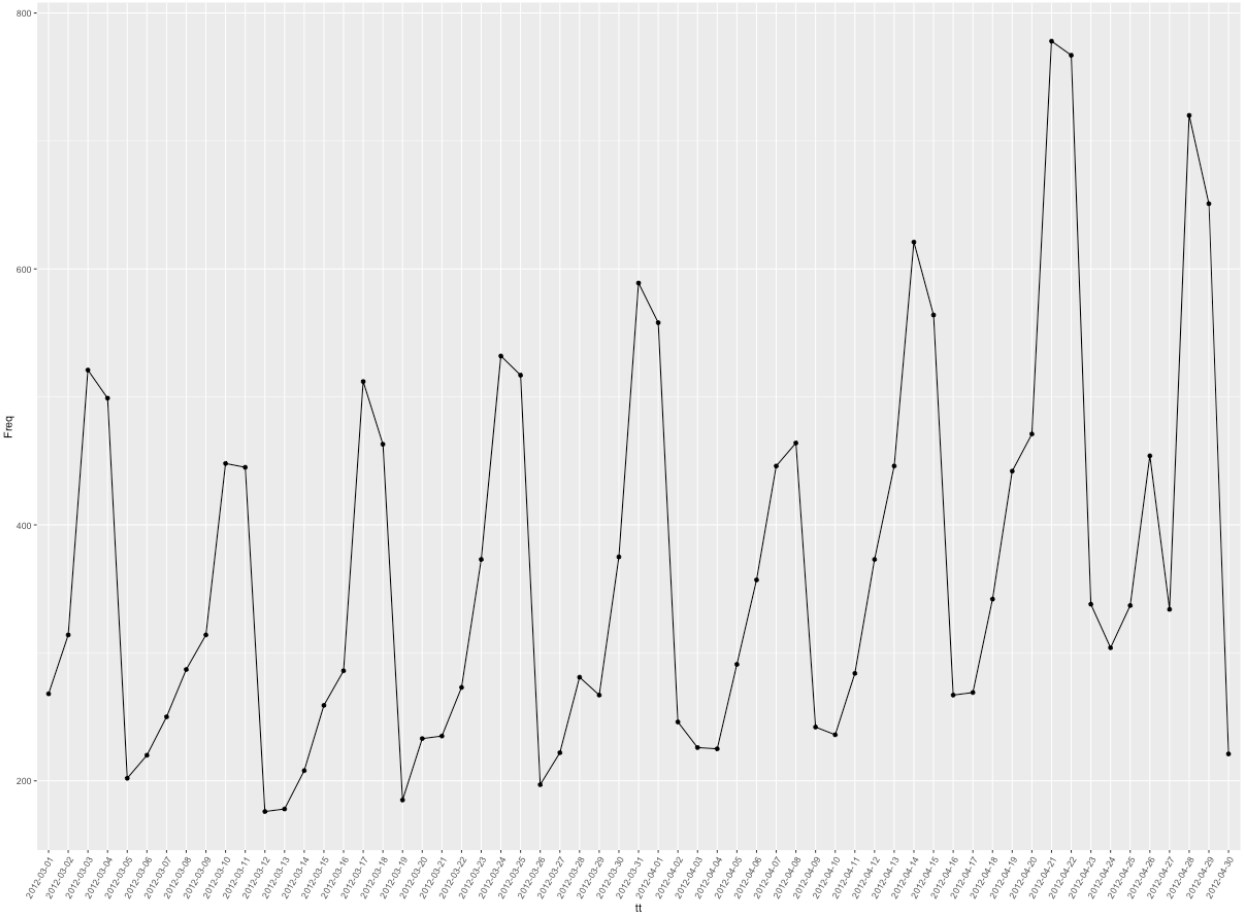
If we have:
A 70 uberX
A 30 uberBlack

And
A 8 hours
Then when we join
A 70 uberX 8
A 30 uberBlack 8 < this is not right

Analysis Time
Plotted in R:

We see on a weekly and monthly level, there is a steady increase in the number of requests, however when we look at the daily level, we see that there are more requests on the weekend(see next page for graphs)


Freq
3000
2500
2000
1500
1000
2012-02-26
2012-03-04
2012-03-11
2012-03-18
2012-03-25
2012-04-01
2012-04-08
2012-04-15
2012-04-22
2012-04-29
tt_w


Freq
800
600
400
200
tt