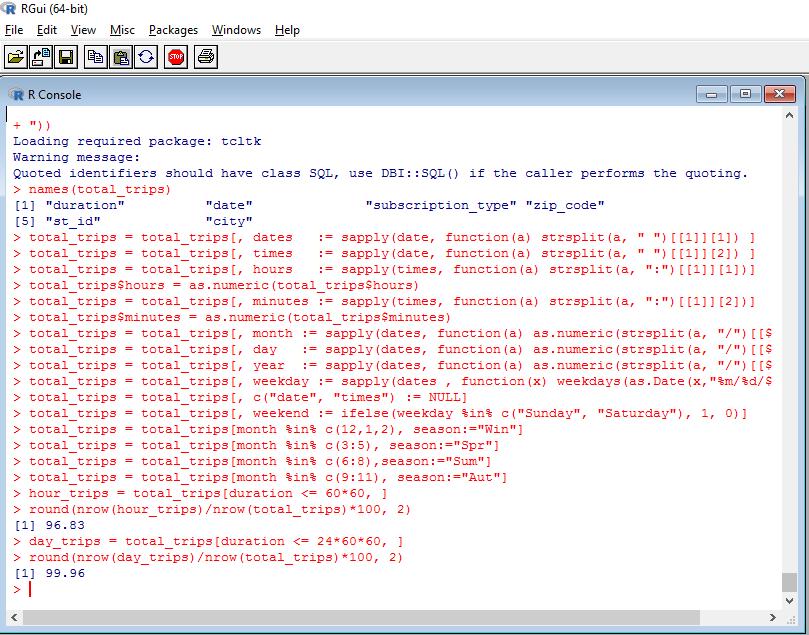
**MODEL CREATION & RESIDUAL ANALYSIS**

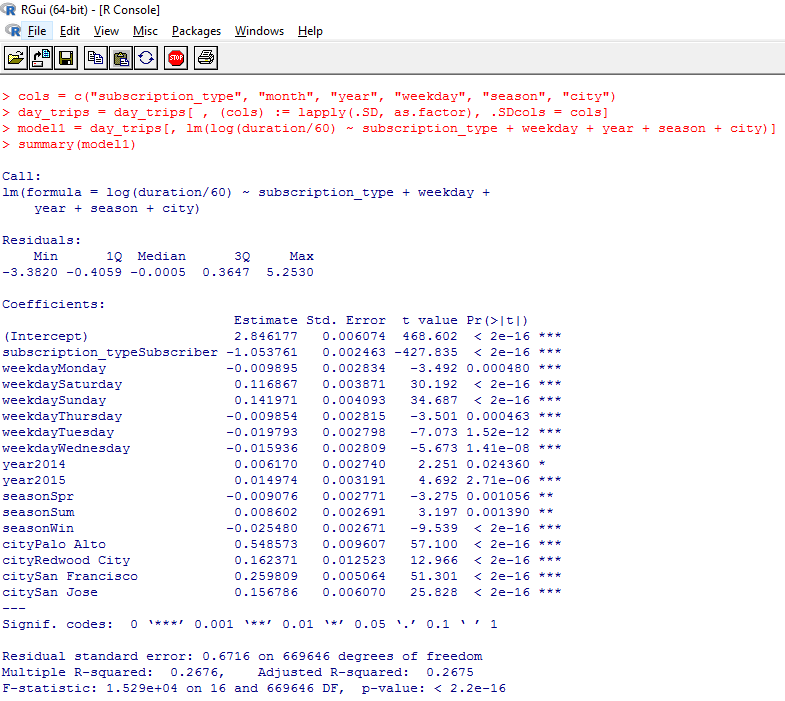
We create a model after we have split our data in terms of data and time.



We can see that 96.8% of the trips are less than an hour or an hour trip.

Also, 99.96% trips are lesser than one day trips only 0.04% trips are full day trips. It means people don’t prefer to take long trips.

We fit a model, using dependent variable duration and other independent variables



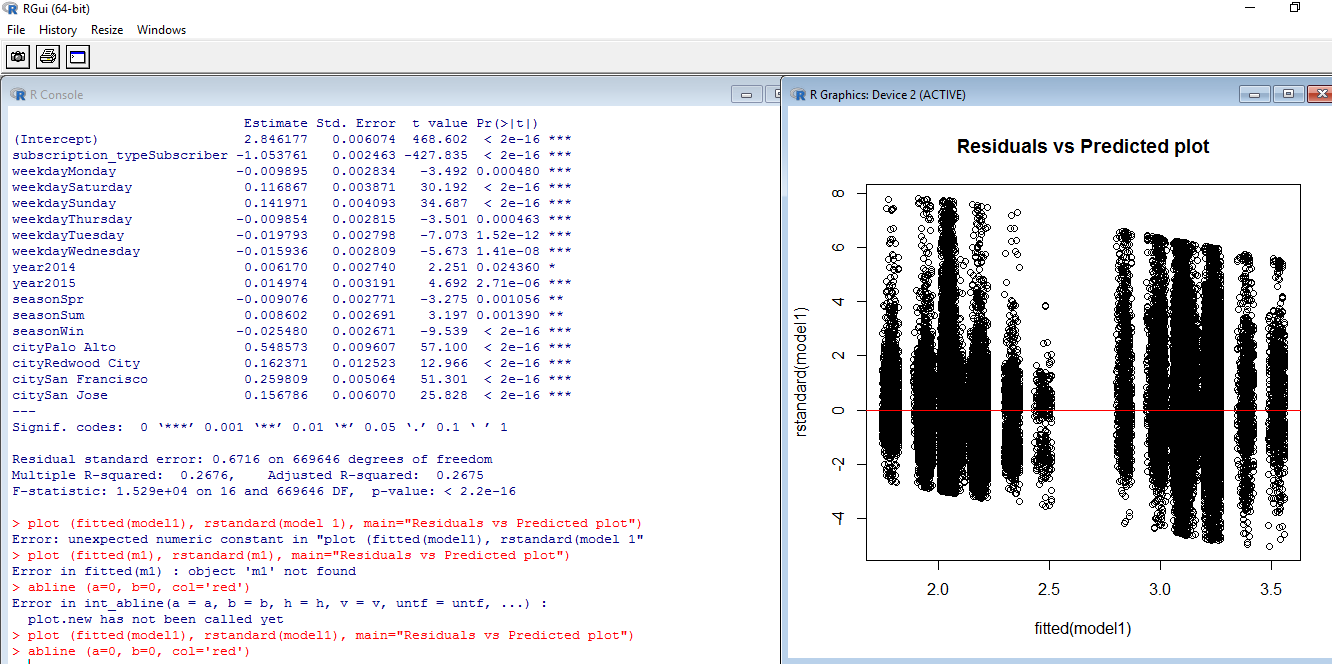
The model explains 26.7% of the variation in data. We test the goodness of fit of model at 5% significant level, from the p-value we can see that p-value is smaller than 0.05.

Subscriber type, weekend, season - winter and city have the most significant impact on the duration calculation.

**We do the residual analysis:**

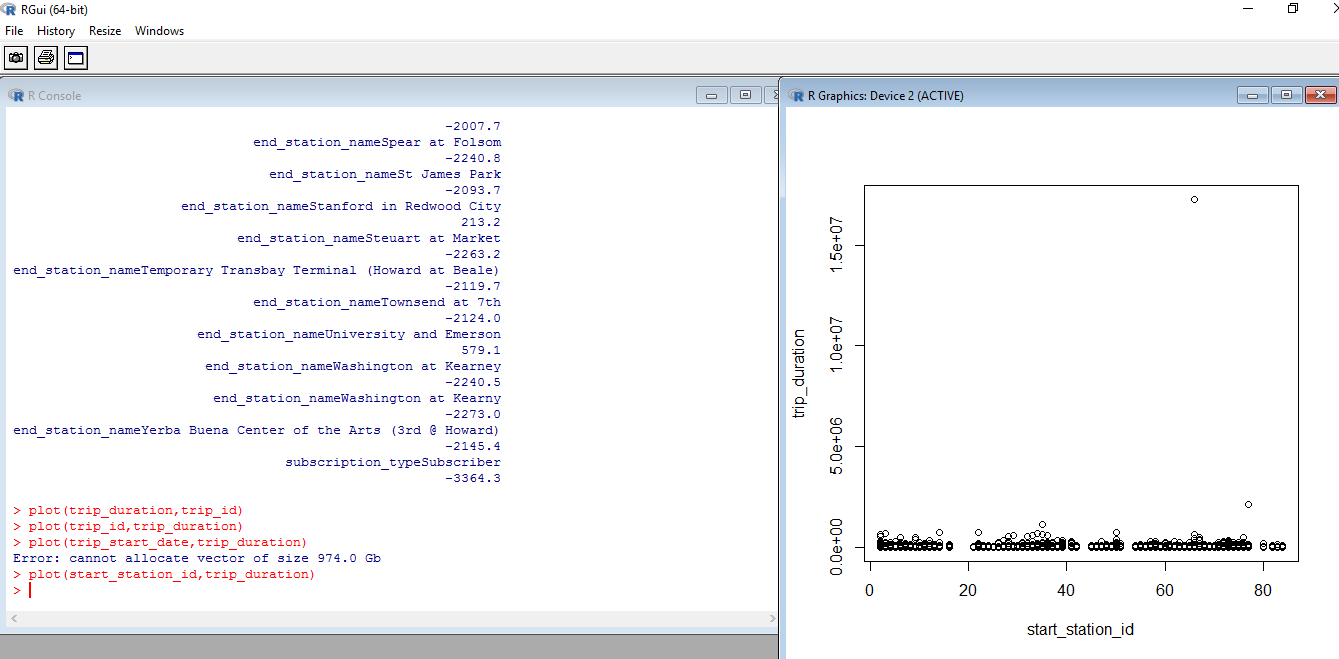
plot (fitted(model1), rstandard(model1), main="Residuals vs Predicted plot")

abline (a=0, b=0, col='red')

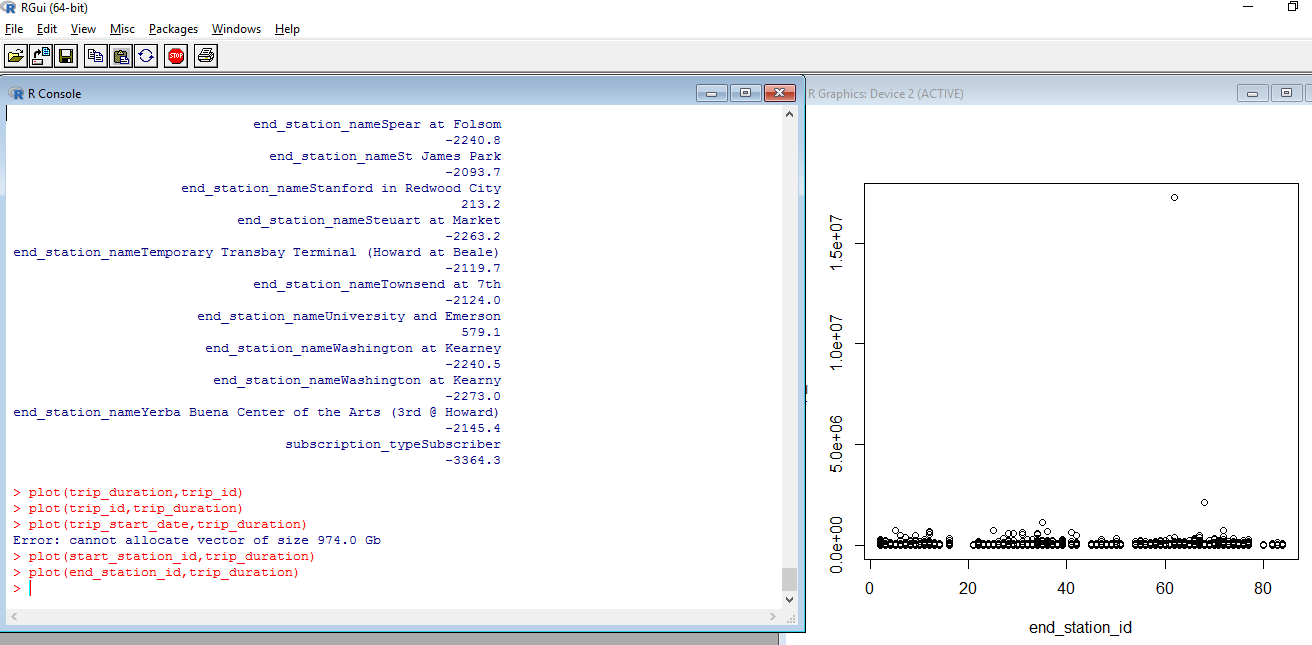


**Independent variable residual analysis**

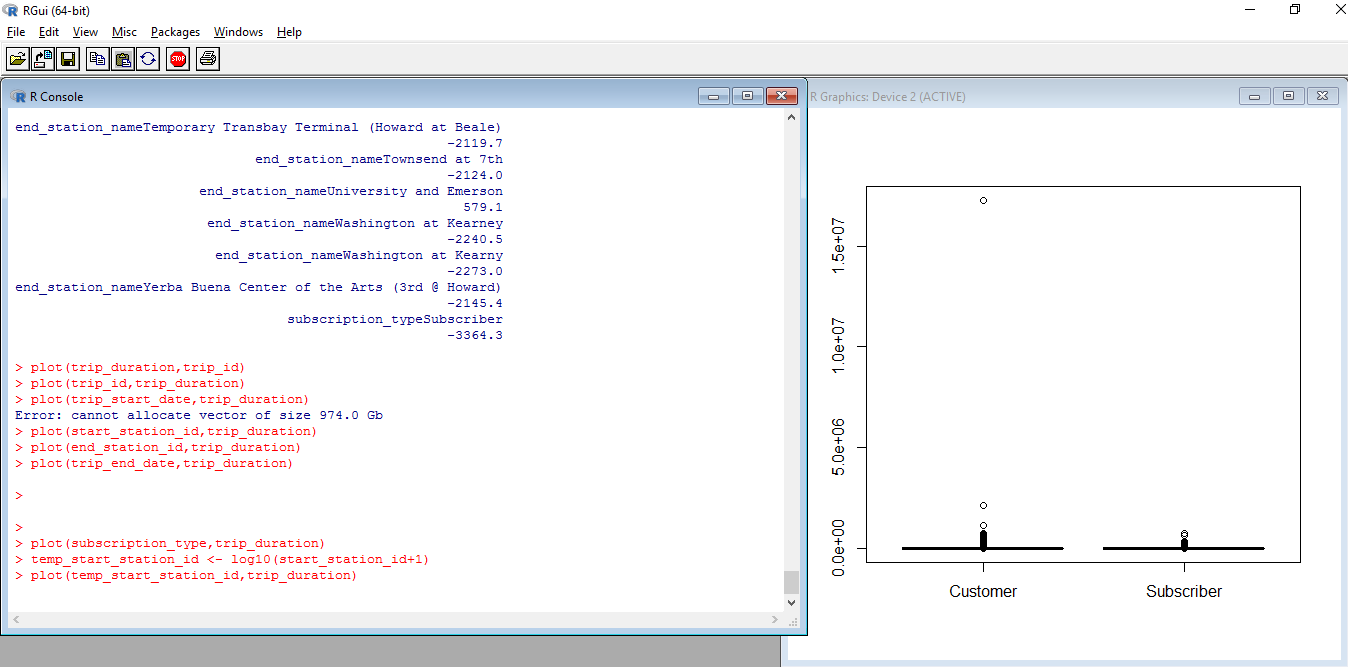
Start\_station\_id, trip\_duration



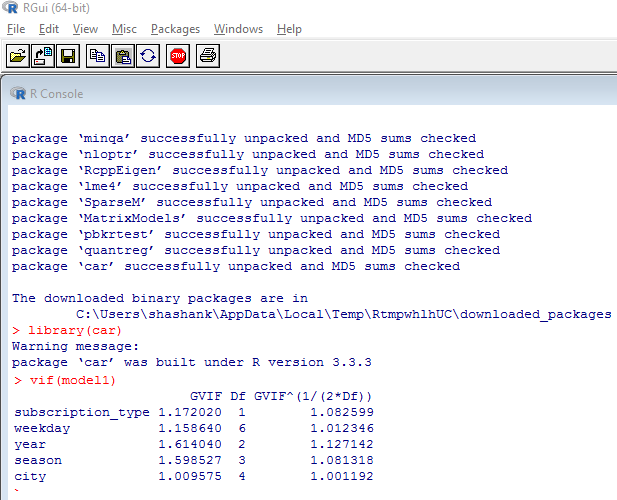
End\_station\_id, trip\_duration



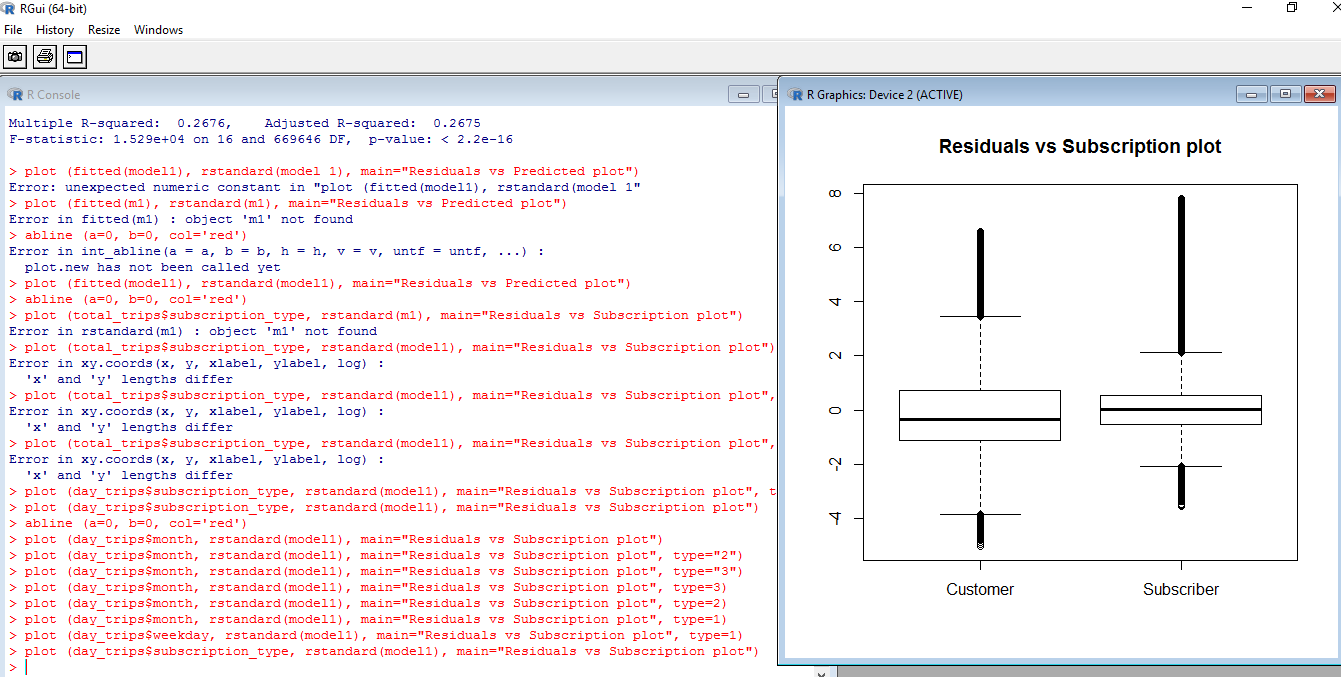
Subscription\_type, duration



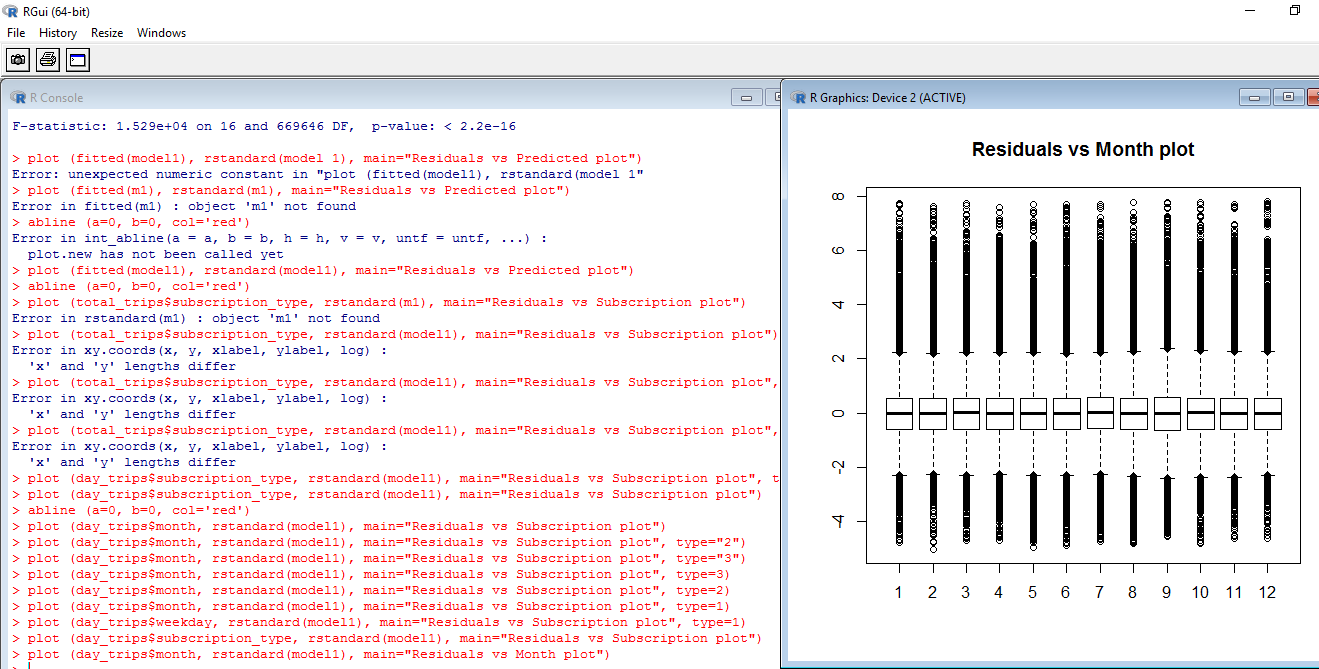
We check VIF value for the model1 and all the values are < 5 which means we don’t have multicollinearity problem.



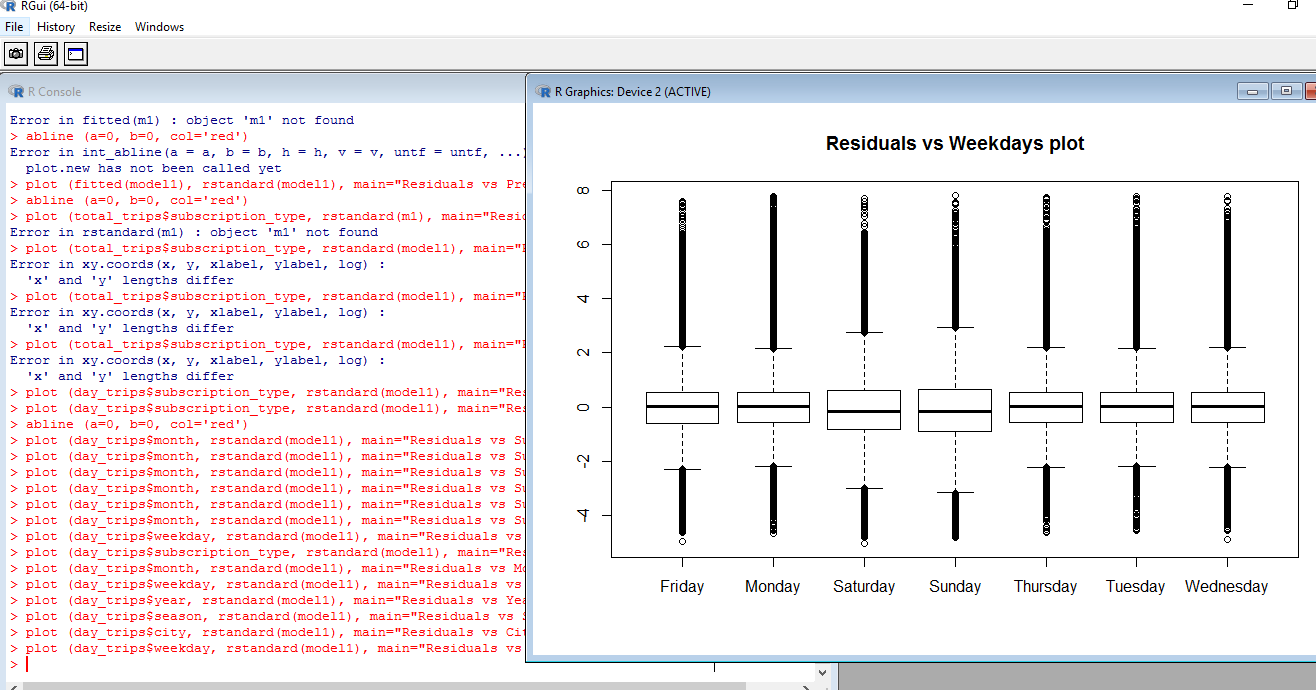
Subscription plot



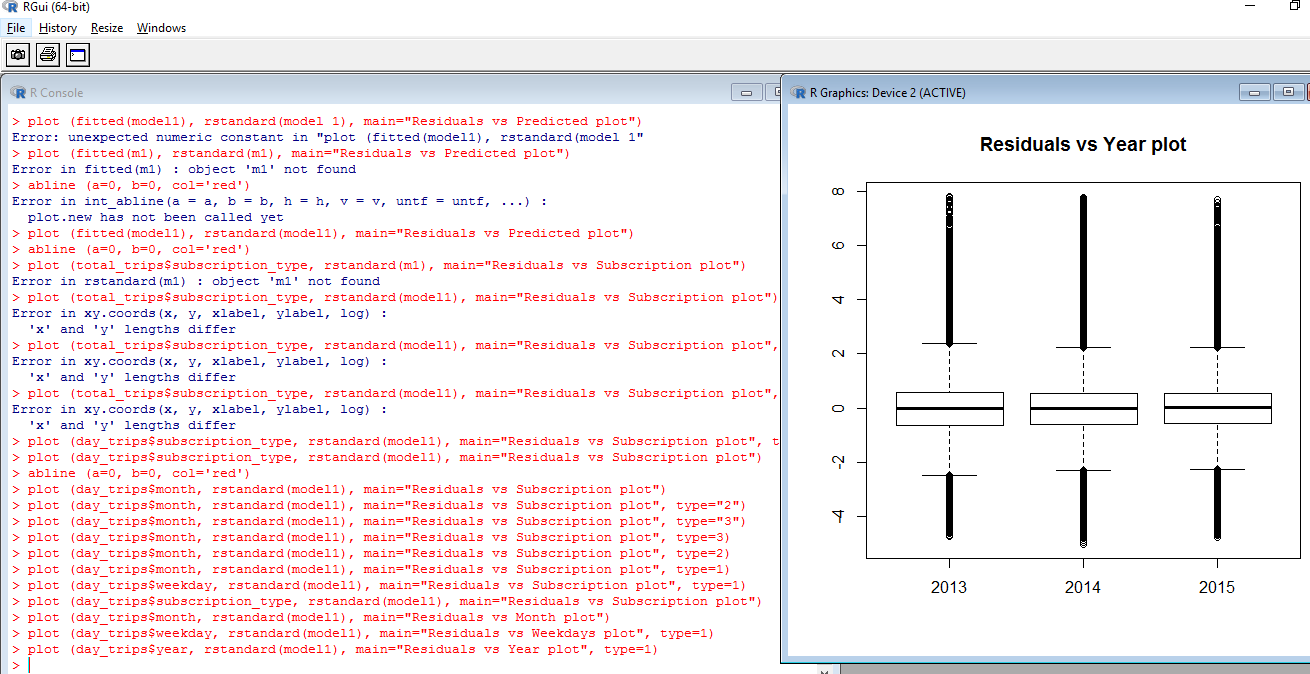
Month plot



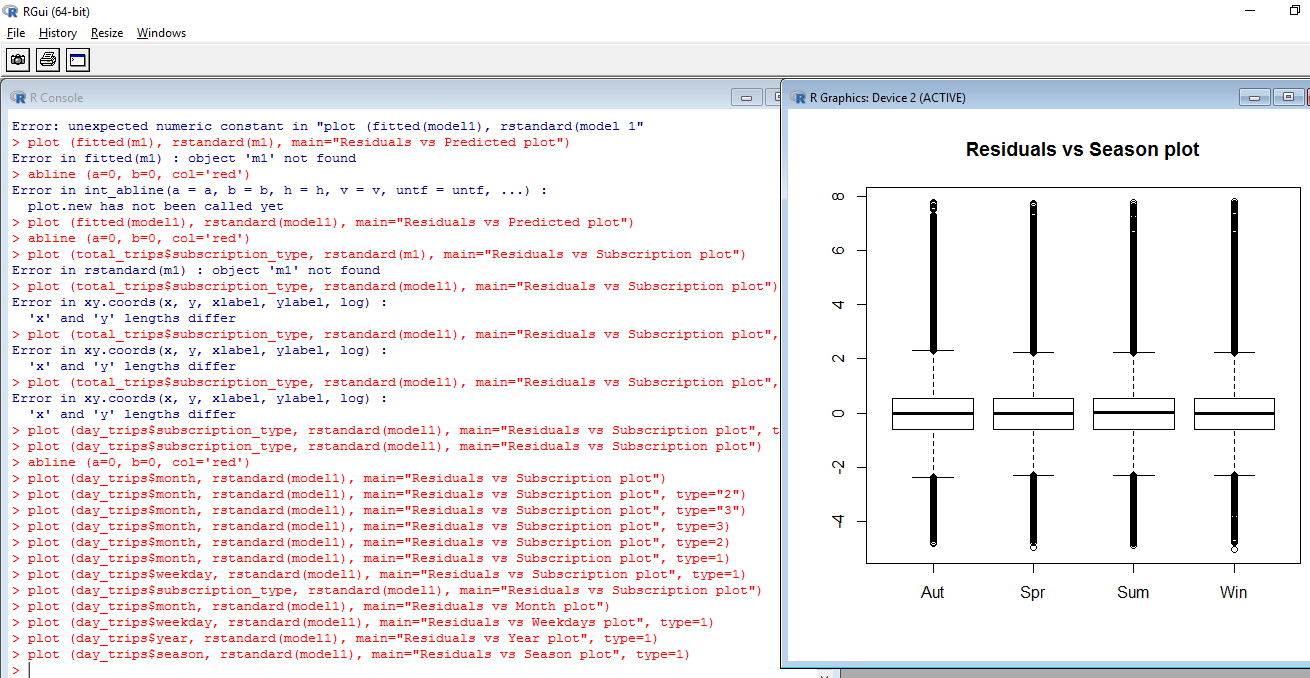
Weekday plot



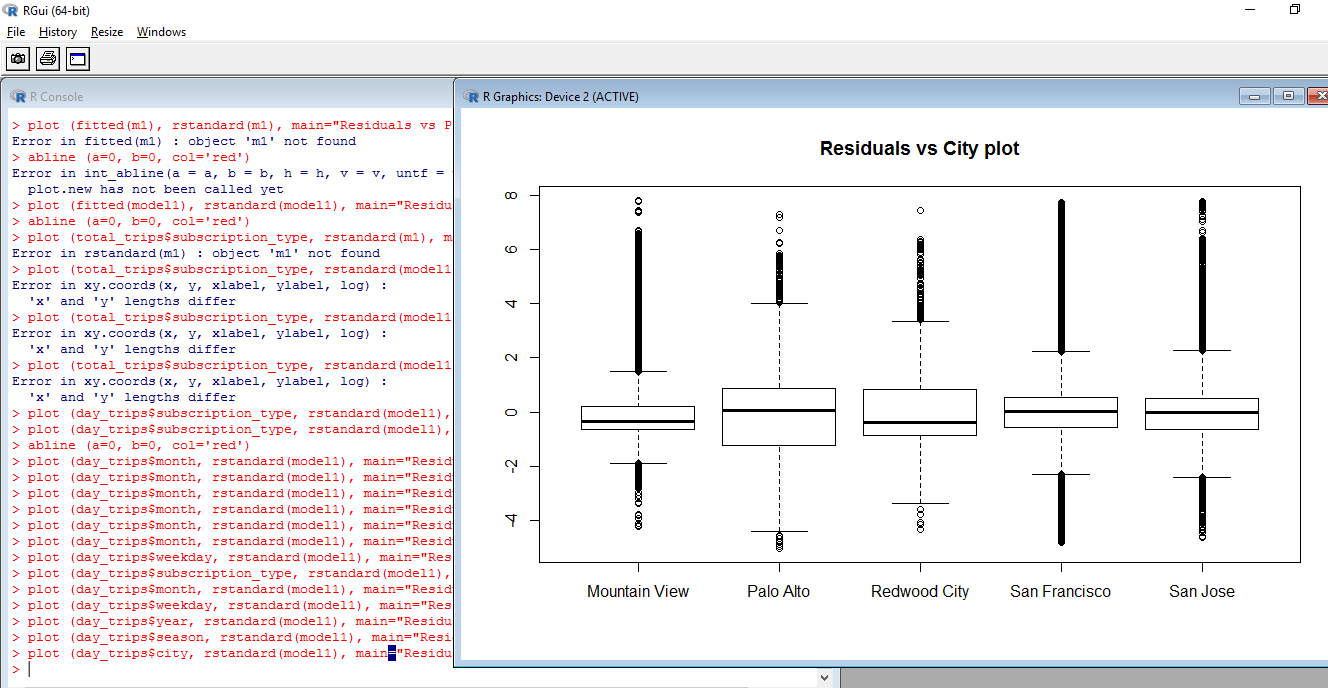
Year plot



Season plot



City plot



QQPLot

