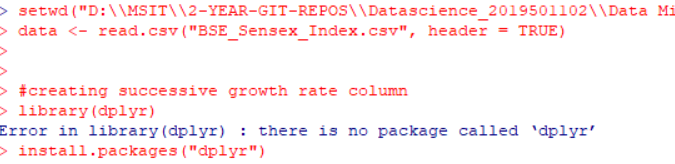
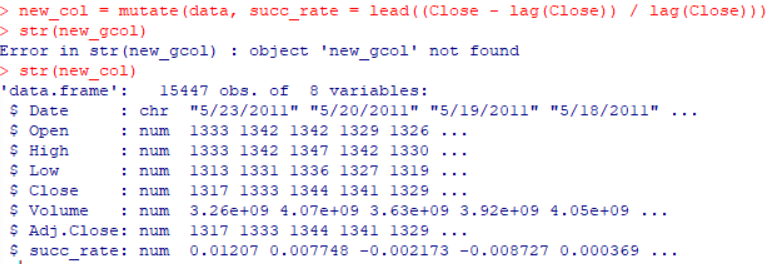
Q8. Consider the dataset BSE\_Sensex\_Index.csv. Create an extra column of successive growth rate for column close where the successive growth rate is defined as

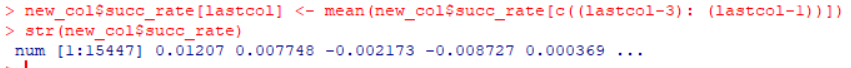
(value of day x- value of day x-1)/value of day x-1. Use a z score cut off of 3 to identify any outliers. List the respective dates from the csv file on which day these outliers fall. (10M)



* Setting environment.
* Reading csv file BSE\_Sensex\_index file.
* And installing and importing dplyr package.
* This package is used for handling data manipulation challenges. Hence we are creating an extra column of successive growth rate, we need dplyr package.



* Creating a new successive growth column using some methods like mutate, lead and lag
* Mutate is for creating new variable from same dataset.
* Lead is for going forward in a data.
* Lag is for going for previous value.
* As in the formula for growth\_rate states 🡪**growth rate** can be computed by taking the current population size and subtracting the previous population size. Divide that amount by the previous size.



* Now replacing the last column of new column with mean of above all rows.

#calculating z-scores

growth\_rate\_mean <- mean(new\_col$succ\_rate, na.rm=TRUE)

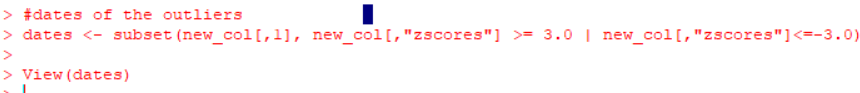
growth\_rate\_sd <- sd(new\_col$succ\_rate,na.rm=TRUE)

z<-(new\_col[,8] – growth\_rate\_mean) / growth\_rate\_sd

sort(z)

Here calculating z scores.

* Formula for zscore -> Find the mean of the vector using function mean().
* Find the standard deviation using function sd().
* Subtract the mean value from the observation and divide the resultant with standard deviation.
* The vector obtained will have the required **Z**-**score** values.



* Finding the outliers using which falls in cut off of 3 of zscores.

