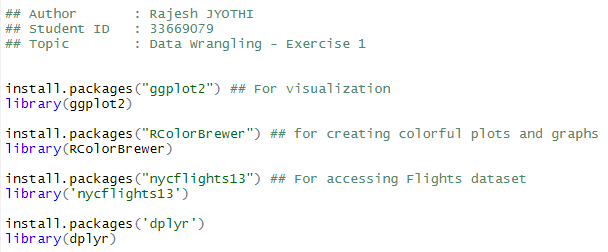
**Data Science Application**

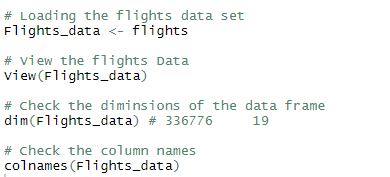
**Exercise 1: Data Wrangling**

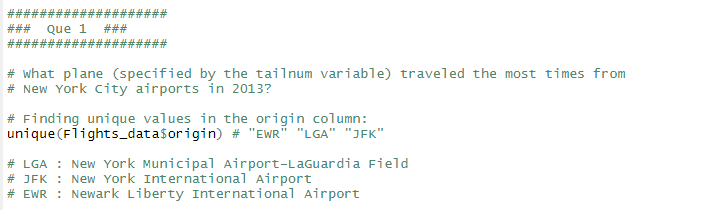
**Question 1:**

Installing and importing required packages and



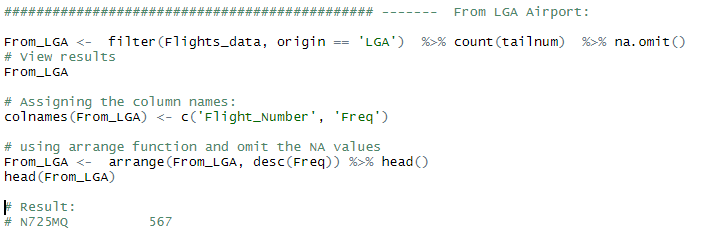
Understand the data set:

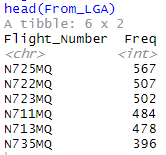




In New York City, we have three major Airport:

1. LGA: New York Municipal Airport-LaGuardia Field
2. JFK: New York International Airport
3. EWR: Newark Liberty International Airport.
4. **Analysing the Flights from LGA Airport:**

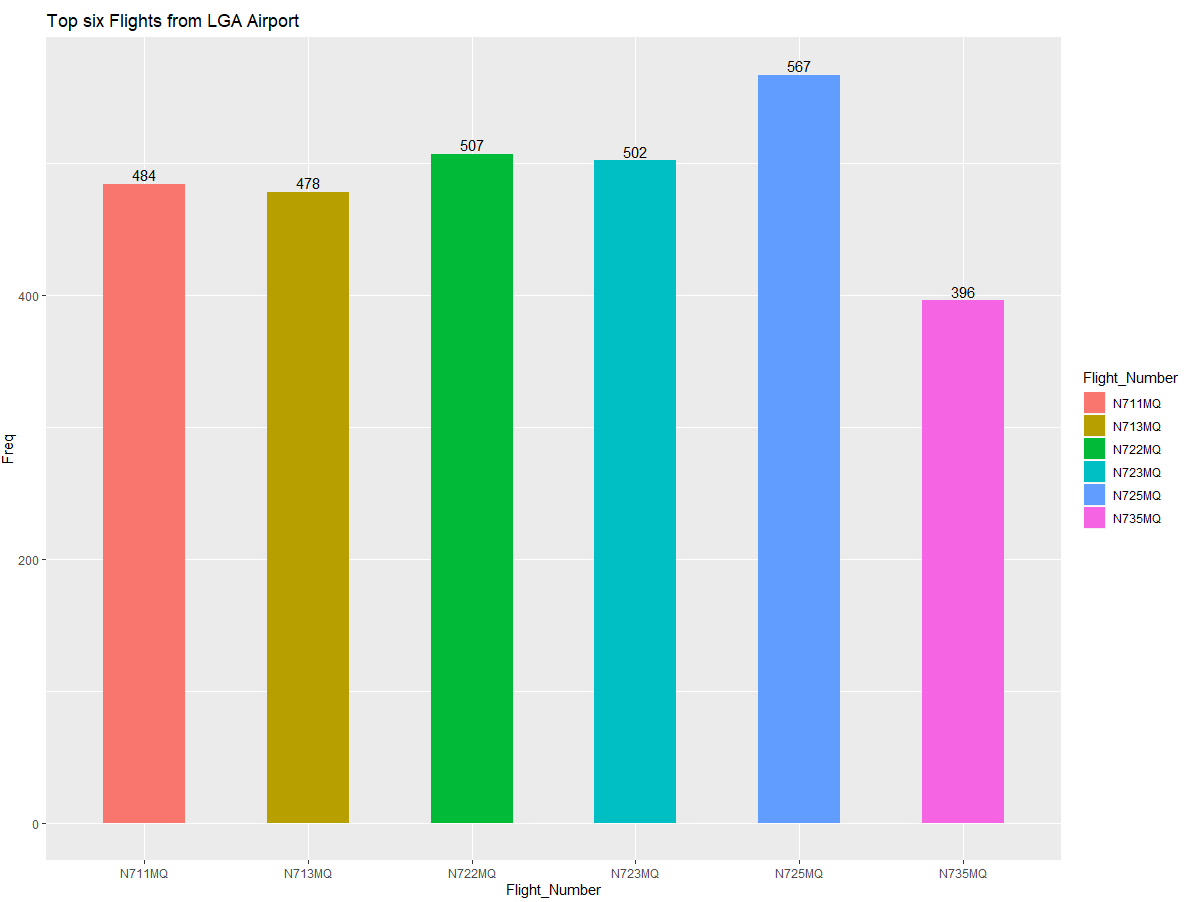




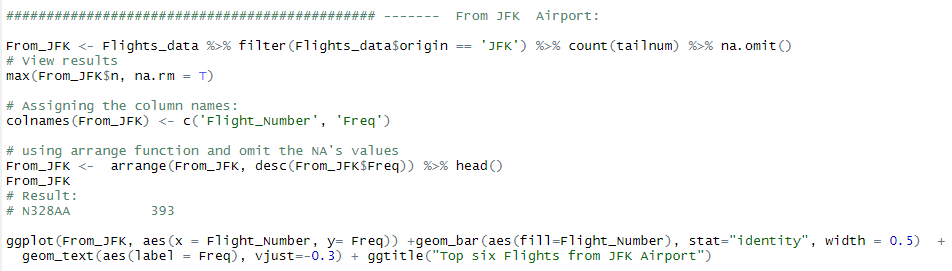
From the above results, we found that Flight number "N725MQ" had flown the most number of times (567) from LGA Airport.

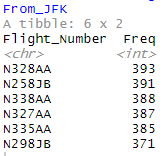
Top 6 Flights from LGA Airport shown below:





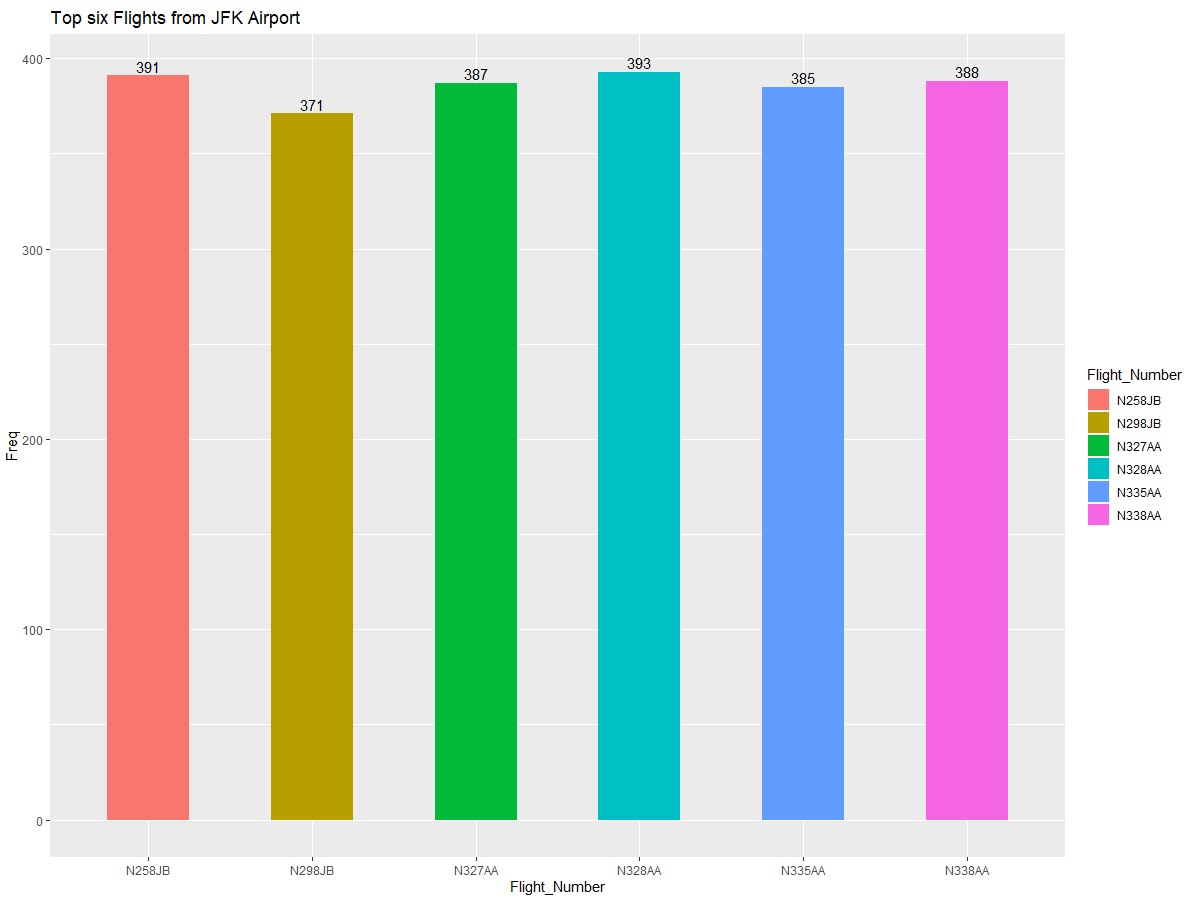
1. **Analysing the Flights from JFK Airport:**



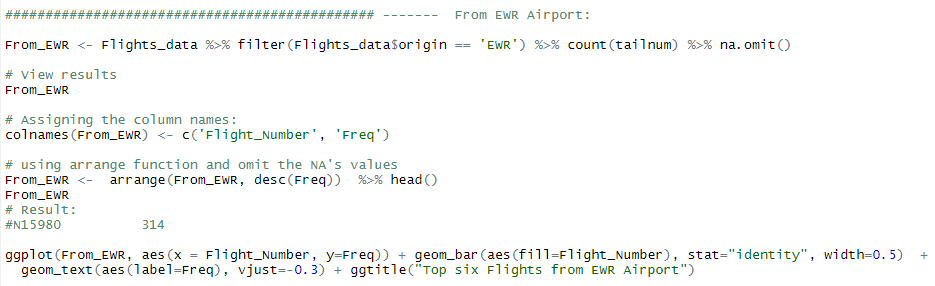


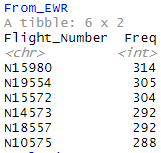
From the above results, we found that Flight number "N328AA" had flown the most number of times (393) from JFK Airport.

Top 6 Flights from JFK Airport shown below:



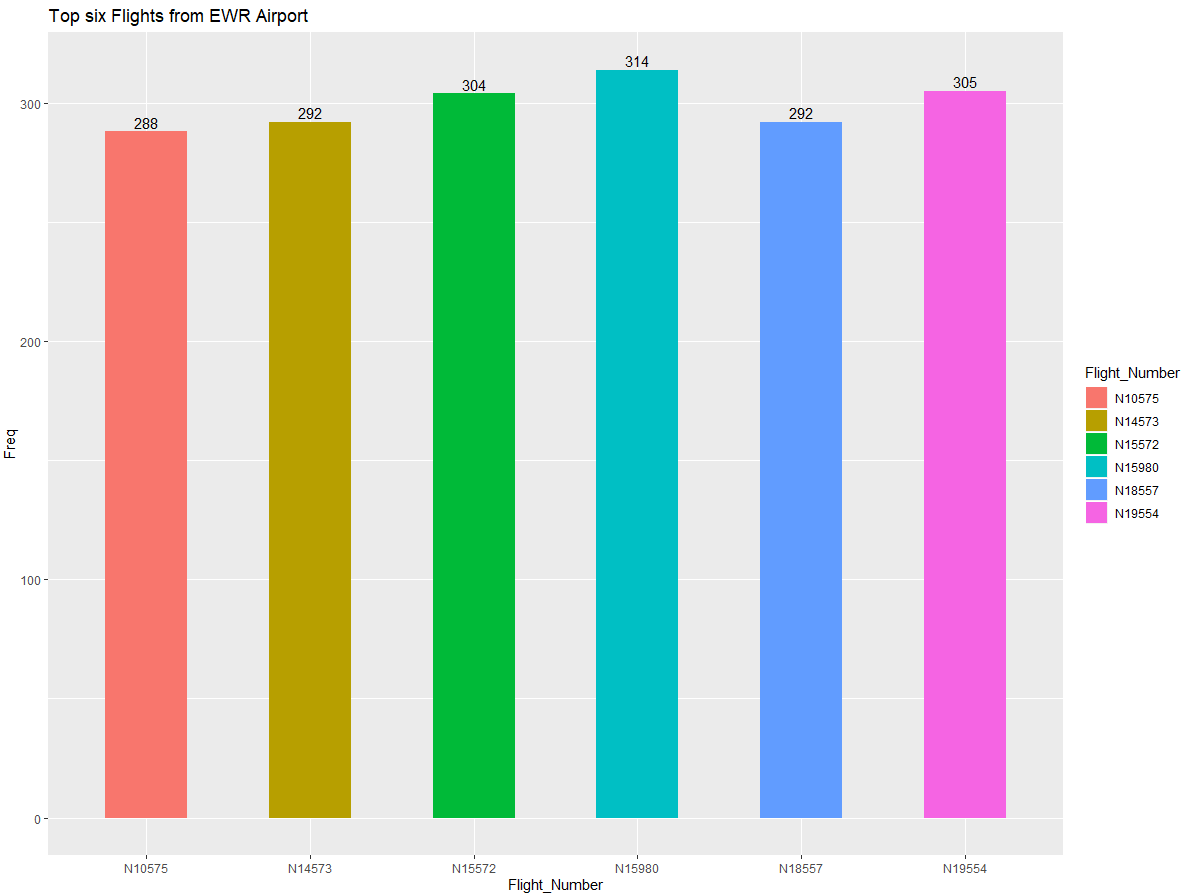
1. **Analysing the Flights from EWR Airport:**



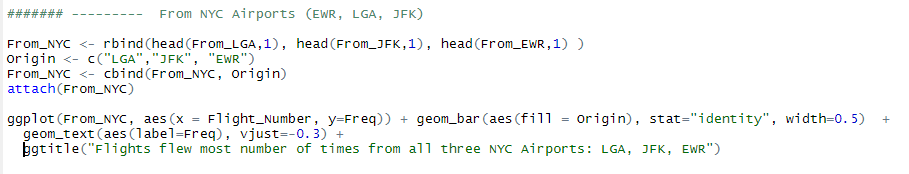


From the above results, we found that Flight number "N15980" had flown the most number of times (314) from EWR Airport.

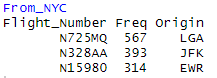
Top 6 Flights from EWR Airport shown below:



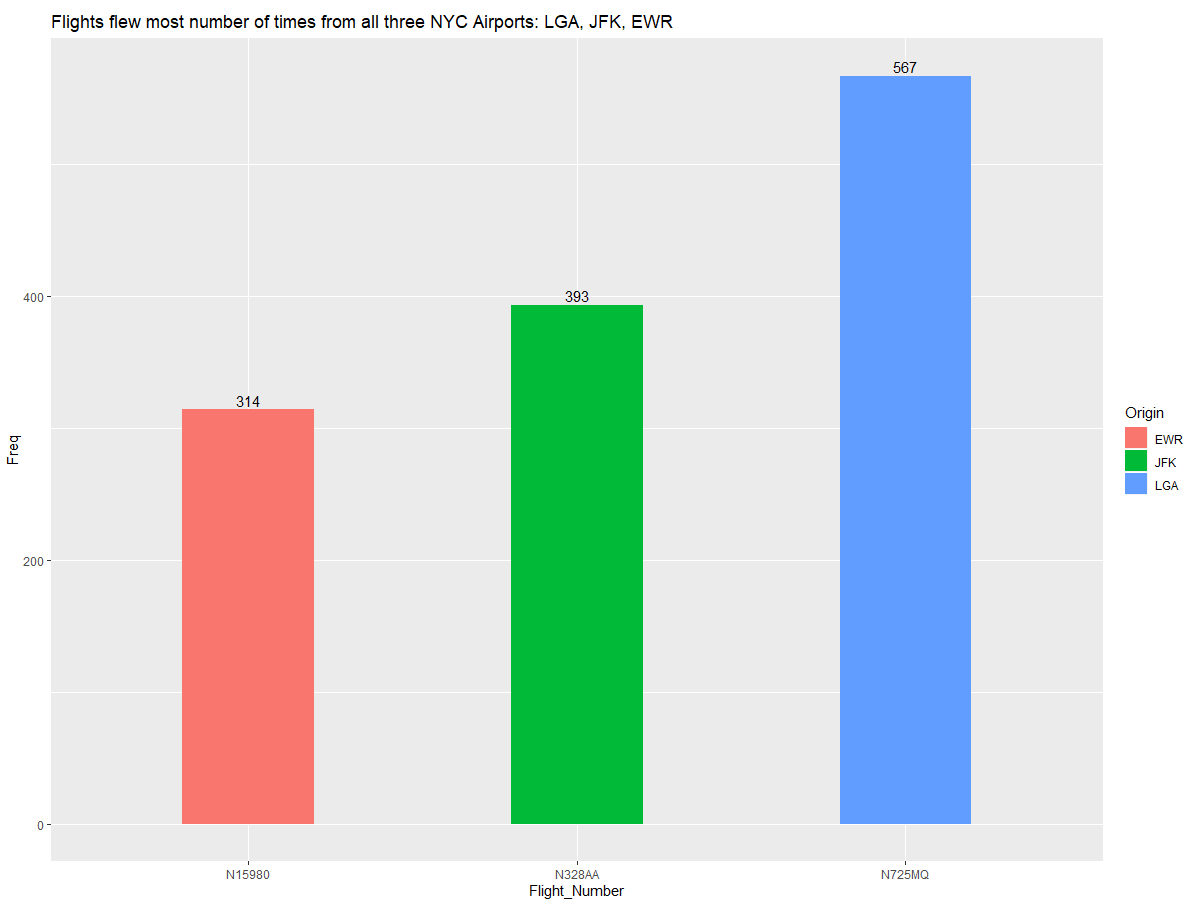
1. **Analysing the Flights from LGA, JFK, EWR Airports:**



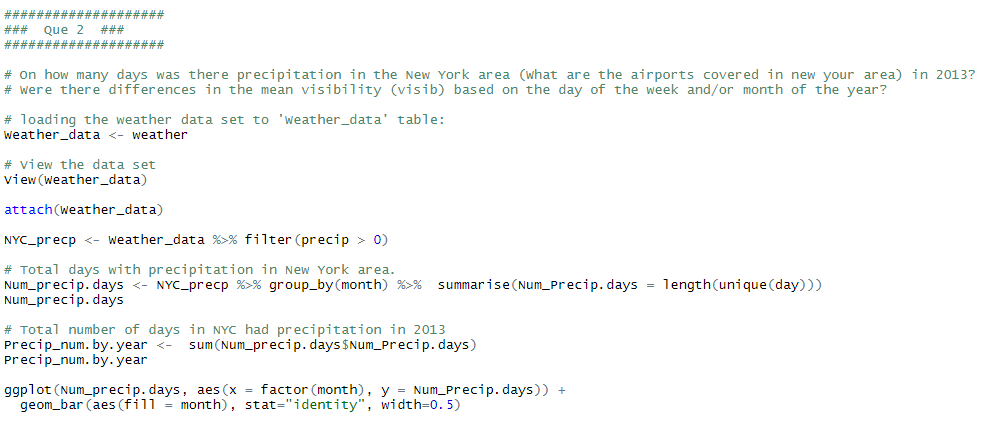
The results shows, the flights that flew the most number of times from each of the Airports.



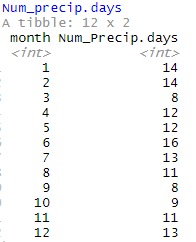
A clear visualization is provided below for a clear understanding of the results.



**Questions 2:**



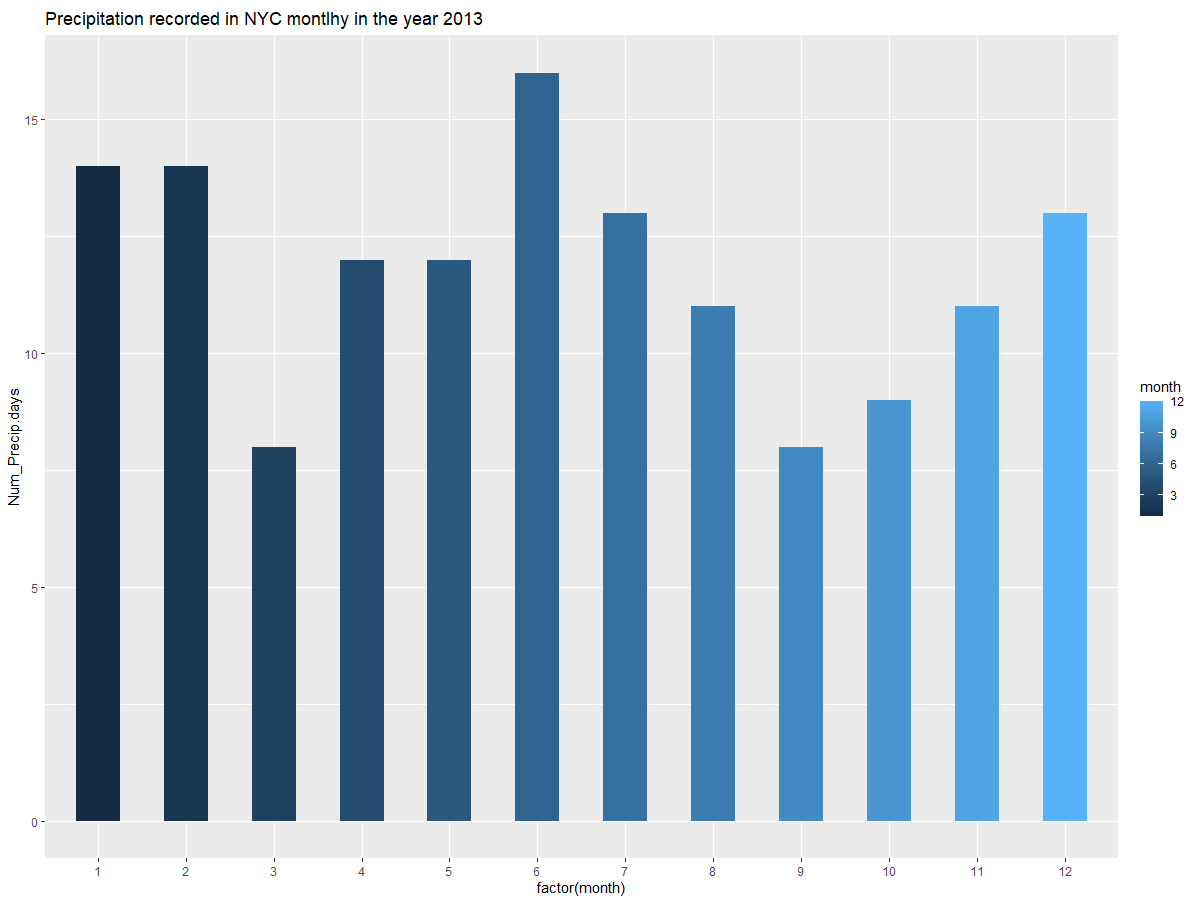
The results for precipitation in New York City In the year 2013 are shown below. The results show's that, January, February, June and December have the most number of Precipitation days. On the other hand, March, September and October have the least number of days with precipitation.



In the year 2013, we have 141 days with precipitation in New York City.

That is, 38.63% precipitation registered in 2013.



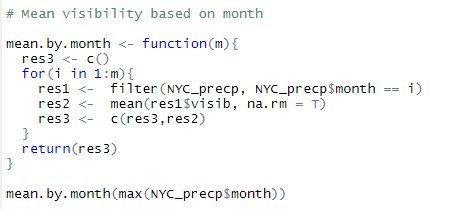


**Mean Visibility:**

Mean Visibility based on the month of a year:

Mean Visibility for a month of a year is calculated in 2 different approaches:

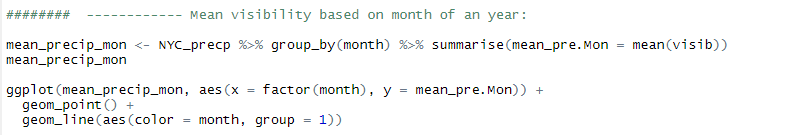
**Method 1: Using a user-defined function**



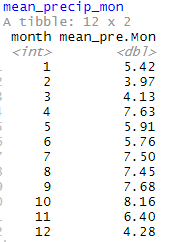
**Result 1:**



**Method 2: Using predefined functions and methods**



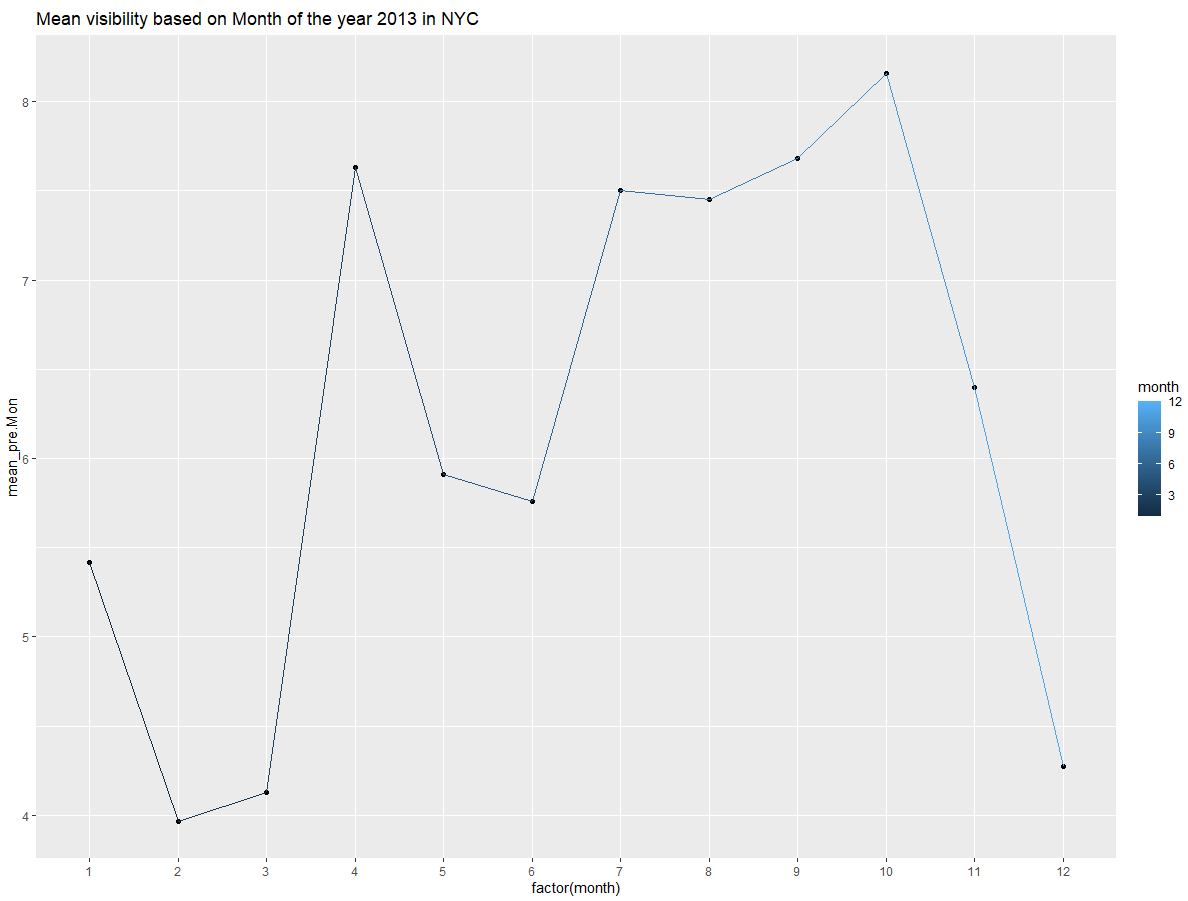
**Result 2:**

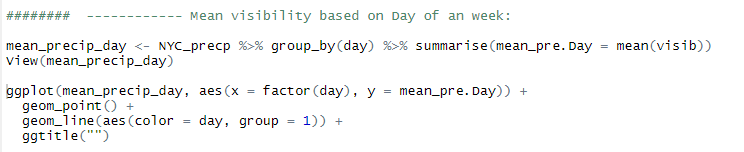


For a clear understanding of the results, let us view the gg line plot.

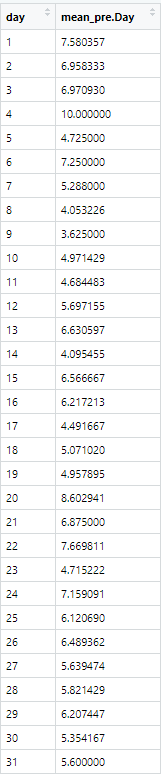
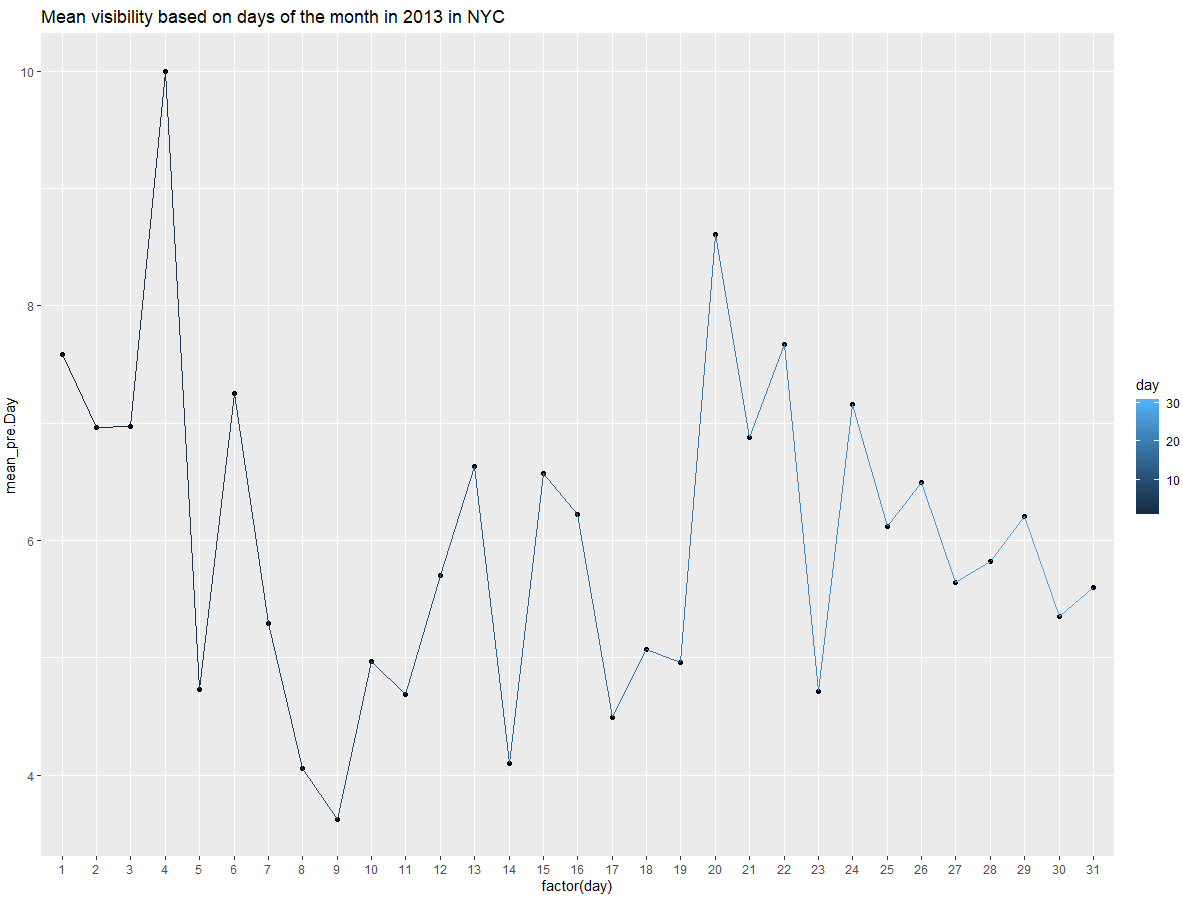
It can be seen that the visibility decreases from Jan to March, but then increased from March to April exponentially. Some fluctuations are seen in from May to October, then decreased drastically till December.

When we compare this with the monthly precipitation graph, we can observe that, as the precipitation increases, the mean visibility decrease at initial stages, but increase with increasing precipitation. From June to December, the Visibility increased with decreasing precipitation.



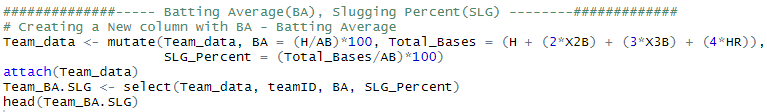


The mean Visibility based on days is given below, The graph shows that the mean visibility is high in the initial stage of the week, and fluctuated from the 2nd the 3rd weeks, eventually decreased from week 3 to week 4.



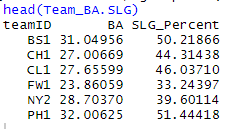
**Question 3:**

**3.1**

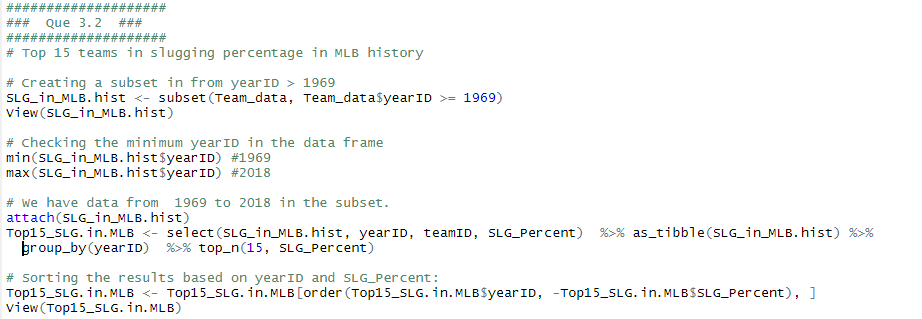


Batting average and Slugging percentage were calculated using the formulae stated in the code.

So easy visibility ” **TeamID**”, “**BA**" and “**SLG\_precent** “are selected suing the above R Code and displayed below.



3.2:



Top 15 teams with highest Slugging percentage (High to Low) from **1969** are displayed below.



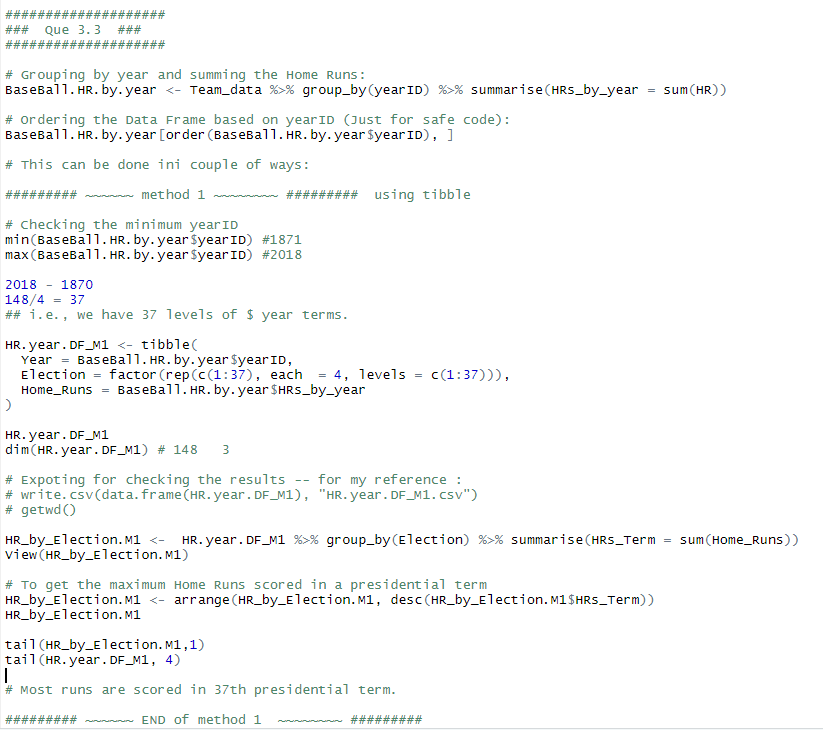




3.3:

I have used two different methods to get the results:

**Method 1: using “tibble()” method.**



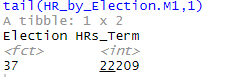
Years from 1871 to 2018 were divided into 4 years of US Presidential years, using the following formulae.

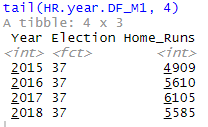
**2018-1870 = 148 years**

**148/4 gives 37 terms.**

From the code results, it is observed that, the highest Home Runs of 22209.

Output:





From the above results, we can see that, in the 37th Presidential year i.e., 2015 to 2018, the highest Home Runs were reordered by Base Ball teams.

**Method 2: using “mutate ()”.**

