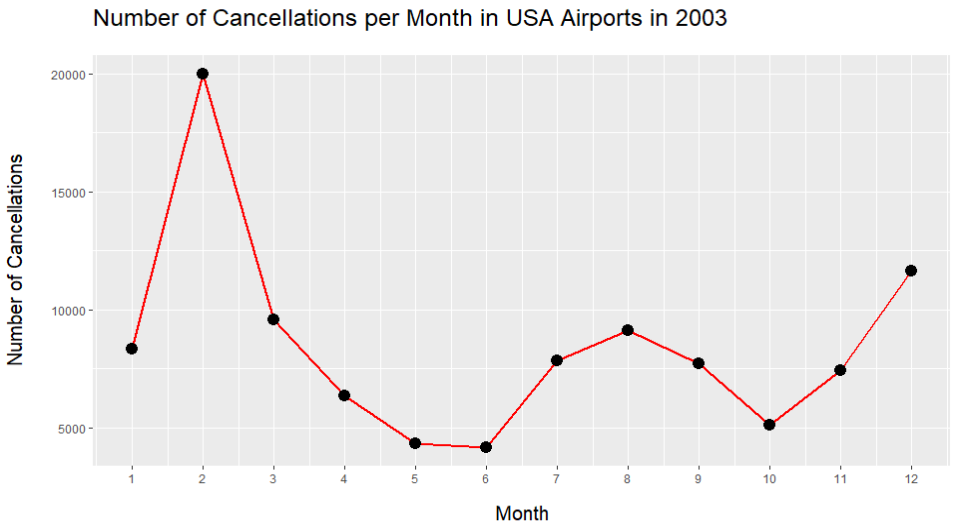
Various Visualizations based on data for all Flights between US Airports in 2003 & 2004

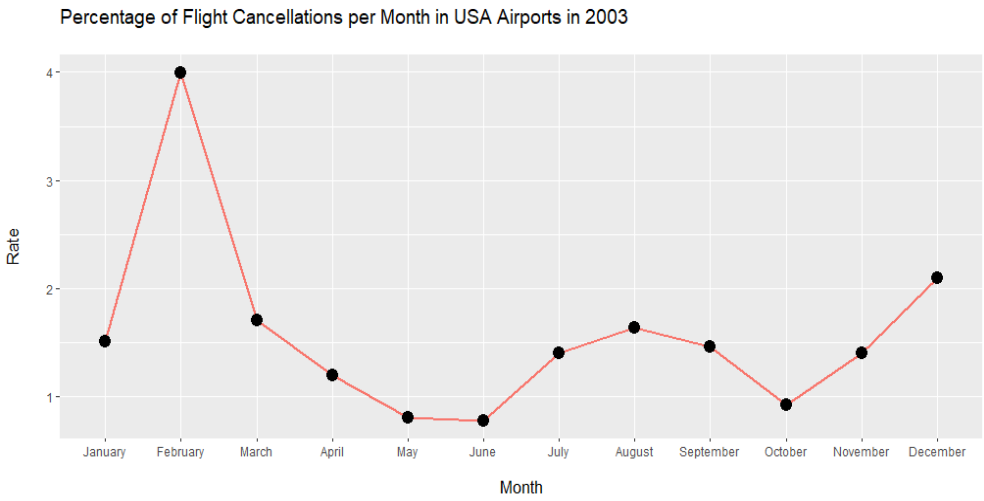
Stavros Alfieris

1) Firstly, we are exploring the total number of Cancellations per month in all US Airports in 2003. We are doing so by creating a line graph connecting the respective number of Cancellations, as follows.



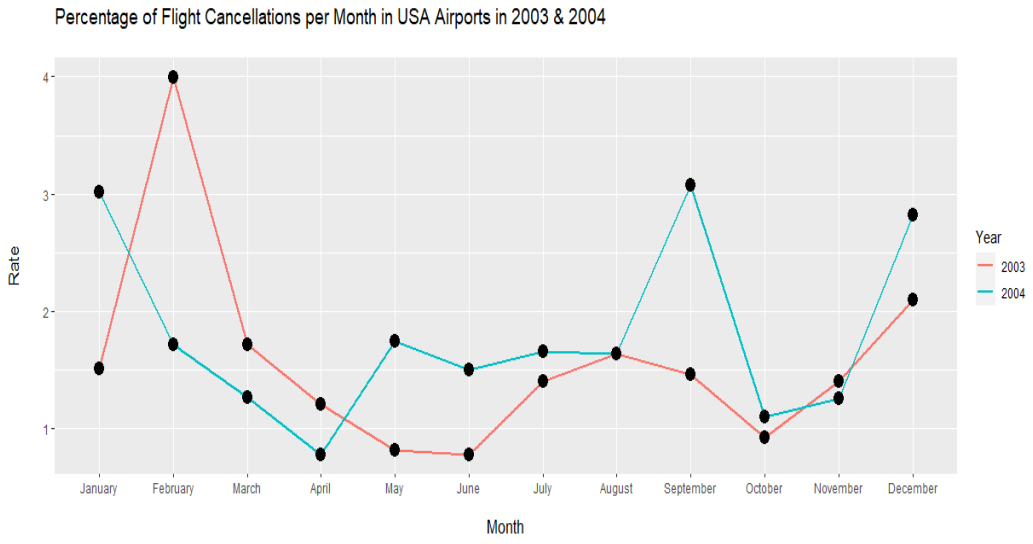
From the graph above, we observe that the number of cancellations is significantly higher in the month of February (20.000 cancellations) than in any other month throughout the year of 2003. However, as the number of flights may be greater during the month of February, it would be more interesting to check the rate of flight cancellation per month.

2) So, now we generate a line graph for the rate of cancelled flights per month, aiming to examine whether the number of flights per month has any significant impact in the number cancellations independently from the month.



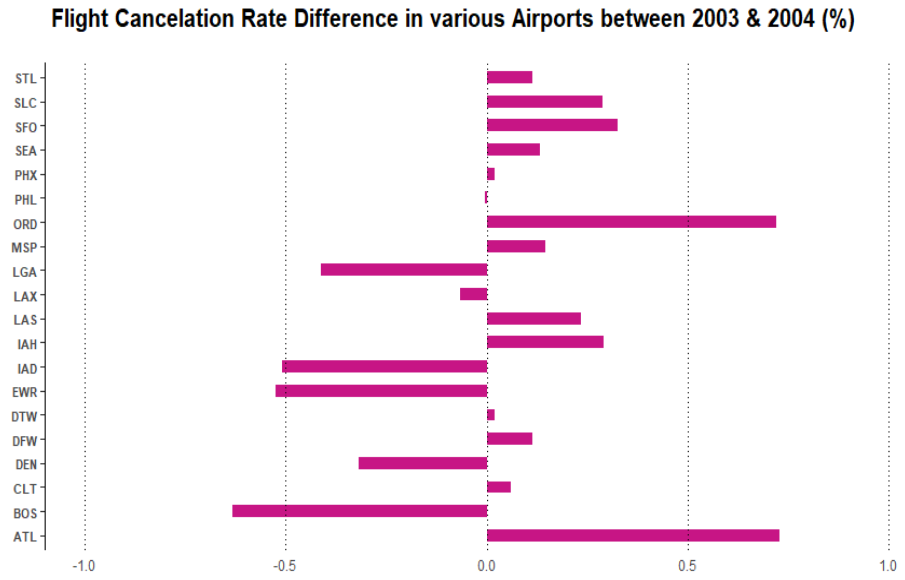
Hence, we conclude that indeed in February of the year 2003, it was way more likely to have your flight cancelled compared to, for example, June where the minimum rate of cancellation is attained.

3) Now, let us check what is the difference in cancellation rates per month between the years 2003 and 2004. To do so, we again create a figure with two lines representing the two years, respectively.



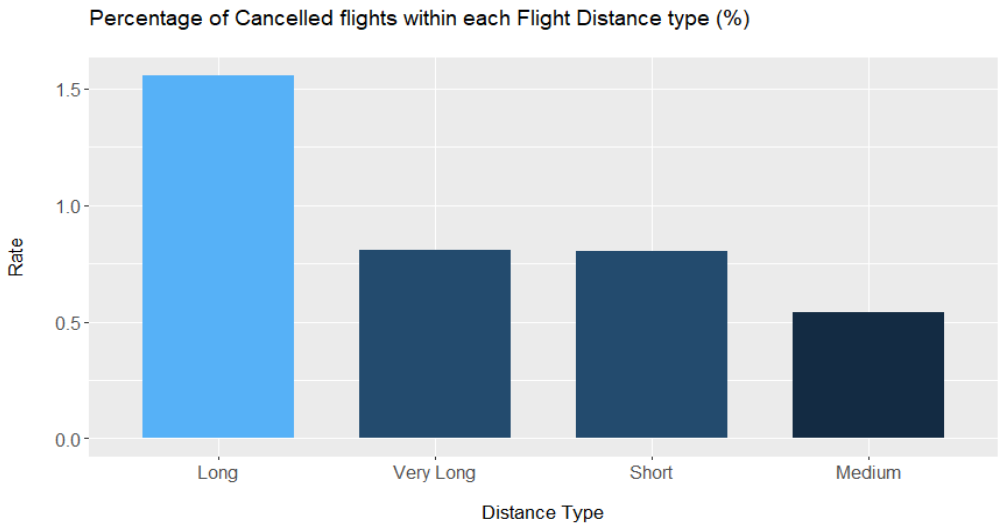
From this graph, we observe that there is remarkable change in the rate of cancellations between 2003 and 2004. Namely, while 4 out of 100 flights on average were cancelled in February 2003, the following the respective rate for the same month went down to less than two cancellations per 100 flights. On the other hand, we see that between May and December of 2004 the rate of cancelled flights was almost everywhere higher or equal to the respective rate of 2003.

4) Finally, we create a barplot indicating the Cancellation Rate Difference between 2003 and 2004 in many major (based on the number of flights they operate) US Airports.



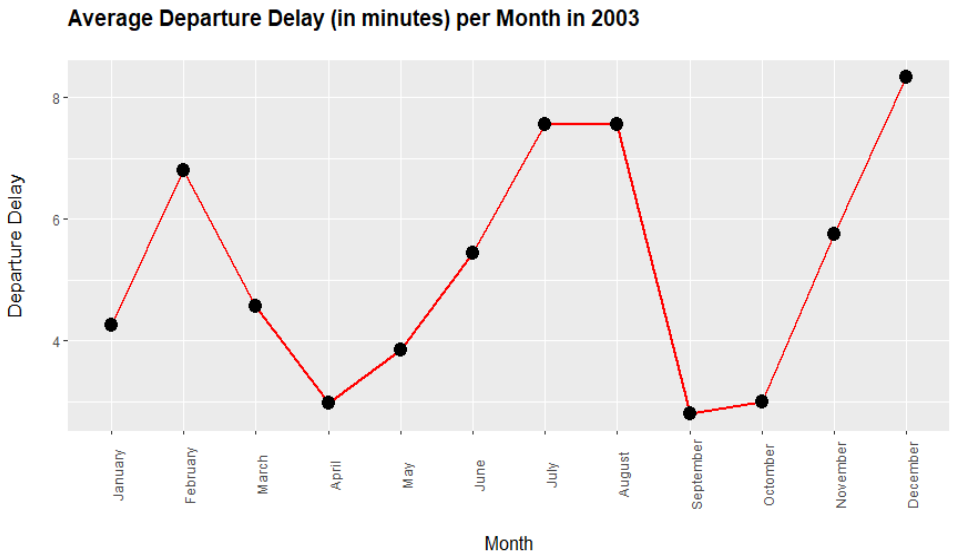
Hence, it is clear that the ORD and ATL airports had a significant rise in the percentage of flights that got cancelled between 2003 and 2004. On the other side, the BOS, EWR and IAD Airports manage to reduce considerably the rate of cancelled flights throughout the same years. In general, we can say that the possibility for a flight to be cancelled increased in 2004 compared to 2003 in the major US airports.

5) Keeping our focus to flight cancellations attribute, we now proceed to check if there is any connection between the flight distance and the possibility for a flight to get cancelled in 2003. To do so, we classify each flight to four different categories based on the range of the flight, employing the descriptive statistics for the flight distance at hand, and then we examine the number of cancellations for each category. The constructed barplot is given below.



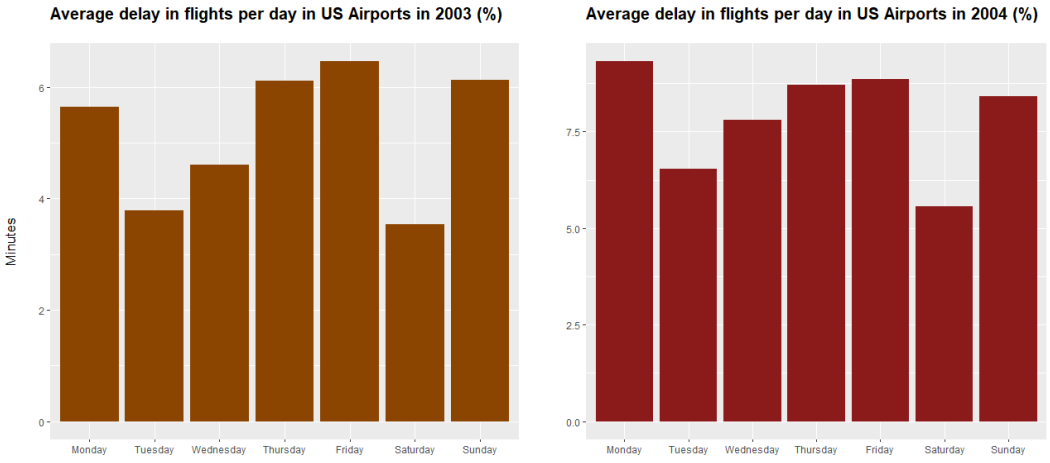
From the barplot above, we can see that there is a pattern between the distance type and cancellations, as a long flight was almost three times more possible to get cancelled than a medium range flight. However, we also observe that between the very long flights and the short flights the cancellation prospect was equally possible, but still higher than for a medium range flight.

6) We are now moving our focus over the flight departure delay time attribute. Firstly, we want to examine whether the average departure delay time is connected to the month of departure. For this we generate a line graph with the months on the x-axis and the respective average Departure Delay Time on the y- axis, as follows.



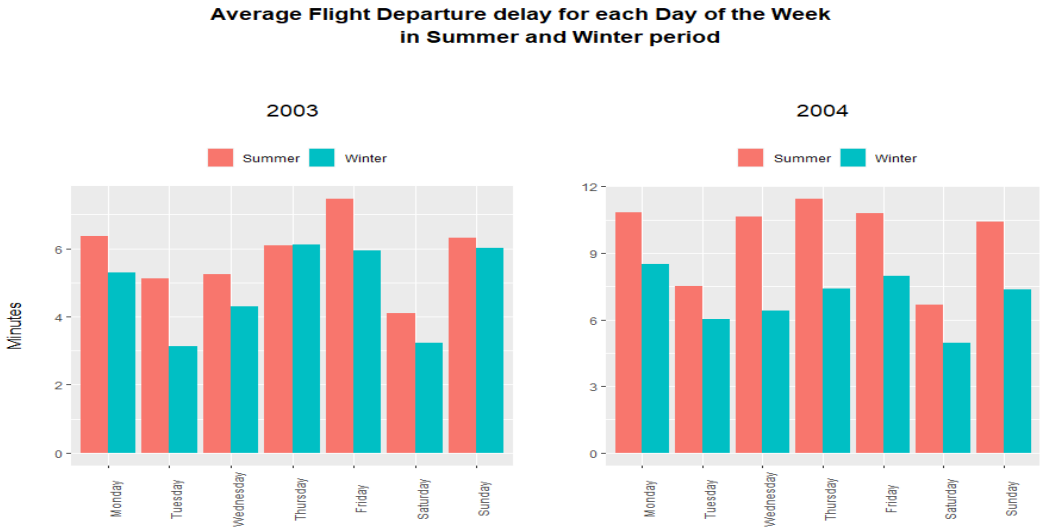
From this graph we get that for a flight scheduled for December we had to expect almost three times greater departure delay compared to April. Also, we see that as we are going towards the summer period, the expected departure delay time in all US airports is increasing, while after August the figure is plunging abruptly. From then on, the figure is increasing again, as we are entering the Christmas Period. Hence, if someone wanted to avoid any significant departure delay time, we would suggest him to travel just before or right after summer period.

7) In sequence, we focus on both years and we aim to examine the relation between flight delay time and the day of the flights departure. A barplot showing the average flight departure delay per day of the week, follows.



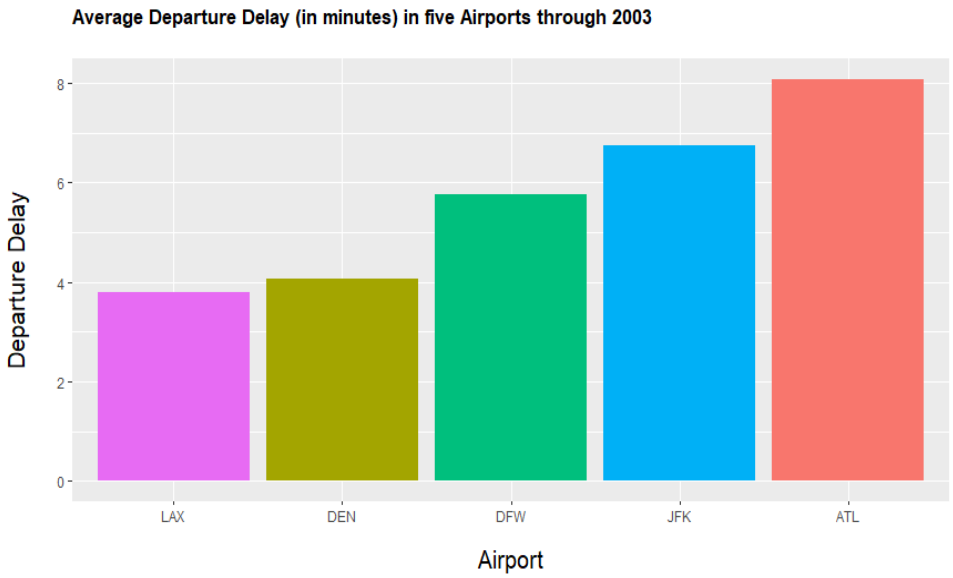
Thus, Tuesday and Saturday are the two days where less flight departure delay is expected compared to the other days of departures, in both years 2003 and 2004. However, we also see that the departure day with the highest average flight delay in 2003 was Friday, whereas Monday is the leading day in terms of flight departure in 2004.

8) Let us moreover separate the days between Winter and Summer period, aiming to find the best day and season selection for reducing the chance of flight delay in 2003 and 2004 respectively.



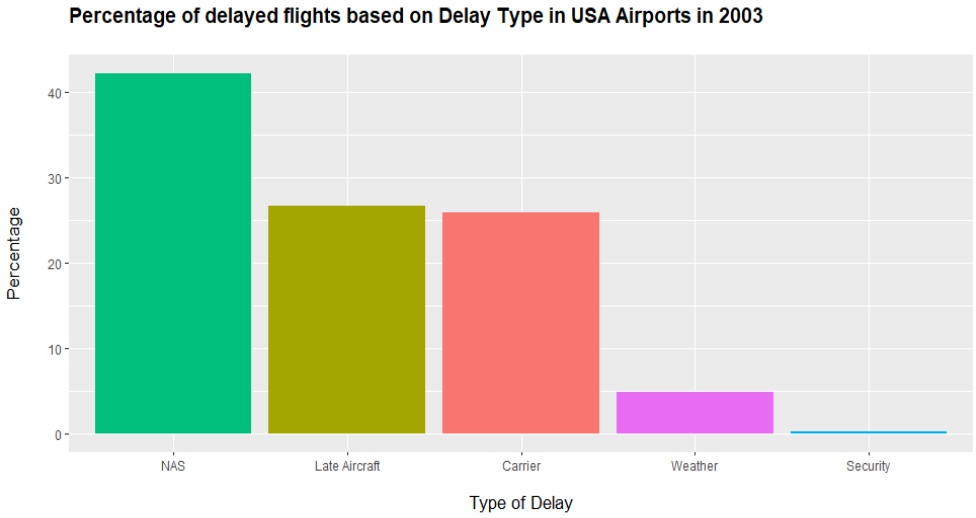
In general, we observe that in both years there are considerably more flight delays during summer than in winter. Also, we see that even if the worst day in terms of flight delays in 2004 was Monday, Thursday in summer period is over Mondays' summer figure (over 11 minutes of expected delay). On the other hand, Tuesday and Saturday are the best day selections for reducing the chance of delay, in both years.

9) Moreover, while keeping our interest in Departure Delay Time, we now emphasize on five different big US Airports. More precisely, we create a barplot indicating the average Departure Delay in these airports under investigation. The airports that have been selected are among the busiest airports in USA.



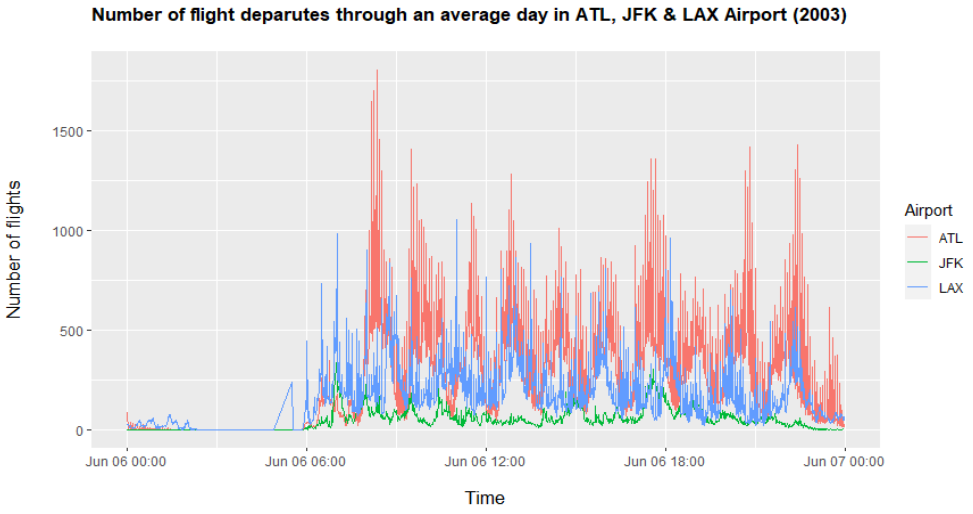
So, someone taking a flight from Atlanta Airport, must expect around twice as more departure delay as in Los Angeles Airport or Denver Airport. On the other hand, passengers in JFK and DFW Airport were likely to wait around 6 minutes on average for their flights to depart in 2003.

10) Since there are many reasons that could yield to a flight delay, in the following we are going to classify all flights in US Airports in 2003 and 2004 based on the type of delay (when is considerable). The histogram below shows the percentage for all the different types of flight delays.



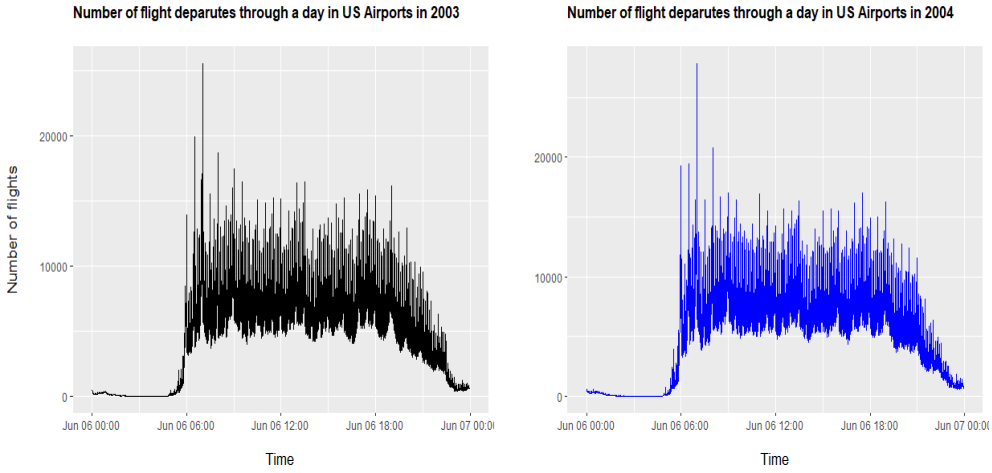
From the created barplot, we note that the most common reason for a delayed flight (either in arrival or in departure) is the weather delay that could be reduced with corrective action by the airports or the Federal Aviation Administration (NAS) which acounts for more than 40% of the total flight delays. Moreover, delays due to a Late Aircraft or due to the Carrier are almost equally possible and combined are responsible for more than 50% of the total delayed flights. Finally, flight delays because of a Security issue is the rarest type of delay among all Airports under investigation.

11) Now, we proceed to find the busiest hours in the Airports of ATL, JFK and LAX within an average day. Busy hours in an average day are determined based on the number of flights departing during these hours.



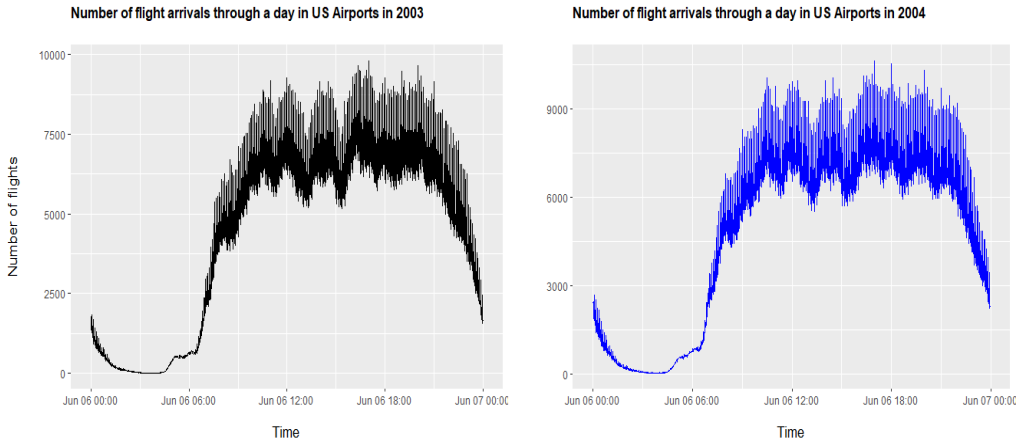
From the Time Series above, we see that all three airports are quite empty between 00:00 and 06:00, while between 06:00 and 09:00 they are getting considerably crowded. Namely, during these hours ATL Airport has its busiest hours, while in LAX Airport there is again considerable number of flights, even though it reaches its maximum between 09:00 and 12:00. Also, we see that the LAX Airport is getting more crowded earlier than ATL Airport, in contrast to the end of the day, where the ATL Airport is still significantly busy compared to the other two Airports. Finally, we observe that JFK Airport is busier during early in the morning or late in the afternoon.

12) Let us now generalize our investigation for the busiest hours in US Airports, by taking under consideration all of US Airports in both 2003 and 2004. The constructed Time Series are as follows.



Hence, we conclude that the distribution of flights, and in turn the busiest hours throughout all US Airports between 2003 and 2004 are identical. Also, in both years the busiest hours are attained between 06:00 and 09:00, while the least busy hours are those between 00:00 and 06:00. Finally, we see that after 21:00 the number of flights is decreasing significantly throughout US Airports.

13) Moreover, we perform the same analysis throughout a day in all US Airports, but this time for the arriving flights. Then, the figures for 2003 and 2004 look as follows.



From the generated Time-Series we get that the most flights are arriving between 15:00 and 18:00 in both years. Also, the minimum number of arrivals is attained between 02:00 and 04:00, where there were almost zero arrivals throughout US Airports in both years.