***Tutorial 10– Data Manipulation (2)***

*flights* data set in the *nycflights13* package contains on-time data for all flights that departed NYC in 2013. The description of columns found in *flights* data set is given below. Use this dataset and *dplyr* package to find answers for the following questions:

|  |  |
| --- | --- |
| Column name(s) | Description |
| *year, month, day* | Date of departure. |
| *dep\_time, arr\_time* | Actual departure and arrival times (format HHMM or HMM), local tz. |
| *sched\_dep\_time, sched\_arr\_time* | Scheduled departure and arrival times (format HHMM or HMM), local tz. |
| *dep\_delay, arr\_delay* | Departure and arrival delays, in minutes. Negative times represent early departures/arrivals. |
| *carrier* | Two letter carrier abbreviation. |
| *flight* | Flight number. |
| *tailnum* | Plane tail number. |
| *origin, dest* | Origin and destination. |
| *air\_time* | Amount of time spent in the air, in minutes. |
| *distance* | Distance between airports, in miles. |
| *hour, minute* | Time of scheduled departure broken into hour and minutes. |
| *time\_hour* | Scheduled date and hour of the flight as a POSIXct date. |

* 1. The values stored in *dep\_time* are in HHMM format. Create a new variable to store these values in minutes since midnight. For example, 525 means 5:25 so it equals to 5\*60+25 =325 minutes.
  2. Increase the arrival delay for flight number 88 by 46 minutes.
  3. Summarize the number of flights, average of time spent in the air, farthest and shortest travelling for each carrier. Order the output by the number of flights in descending order.
  4. Summarize the number of flights, average of time spent in the air, farthest and shortest travelling for each origin and destination for flights travelled between June and September. Order the output by the number of flights in descending order.
  5. Find all destinations that are flown by at least five carriers. **Hint**: use *n\_distinct* function in summarise function.
  6. Find destinations that have average arrival delay more than 10. Order the output in descending order by average arrival delay.
  7. Find the carrier that has the least departure delays on average.
  8. Find the carrier with the highest number of flights.