

Exercise 3.2

Answer the following questions

- **RQ1:** In baby2.R, there is an outlier. What is the data table Rcode to recode it to a correct value? Use full path to csv file.
- **RQ2:** Import "flights14.csv" by using fread() and store it as flight.dt. Use full path to csv file. Remove column "year" from flight.dt as it is all 2014.
- **RQ3:** In flight.dt, create a "major delay" indicator column (0: arrival delay + departure delay < 60, 1: Otherwise). Find a way to verify that the above is done correctly.
- **RQ4:** Identify which (origin, dest, and carrier) has the highest number of major delays?
- **RQ5:** Which (origin, dest, carrier) is the worst in terms of total delay time?
- **RQ6:** Find the top 10 total delay list of (origin, dest, carrier). Note that total delay = arrival delay + departure delay.

The most interesting is RQ3. Can you find a way to verify that the Rcode executed correctly as intended? Be clear on the meaning of "verification".

Coordinator's solution to all exercises will be posted on main site after all classes complete lessons at end of week.