Data Analysis

We would like to analyse how far players of a game progress as a function of age. The age is measured as the number of days since game installation. The progress is measured as a player's level number at the end of a day. If the player does not complete any levels on a given day, then her progress stays constant in the previous available value until she completes some levels again. Note that, in some situations, a player may have to return back to an earlier level and continue from there.

More specifically, we would like to know the average progress by age for two groups of players
1. Those players who actually played the game at a certain age (active players)
2. All players including those who happened not to play at a certain age

There are two data sets available to conduct the task
1. Installation events with fields: player_id, timestamp
2. Level completion events with fields: player_id, timestamp, level_number

The outcome of the task should include
1. An SQL script with queries that calculate the average progress by age and insert the results into a table with fields: age, avg_level_active, avg_level_all
2. An R or Python script that reads data from the results table, plots a chart with the two progress curves and saves the chart as an image file
3. The image file