Assignment 2 Report

I used pandas and dfply in this project. I first load the two data sets into two data frames. To merge the datasets, I processed both by adding 3 new variables – month, day, and hour (to be joined on latter) – using the mutate() function in dfply. To get an aggregate hourly electricity consumption for the second data set, I grouped its values by the just created (month, day, hour), and applied the summarized() function with an aggregate function sum() in it. Then, I joined the two datasets using function left_join() so that the new.app data set with overlapped date times can merge with the first date set. Lastly, a new total variable was created in the merged data set.

By plotting the total consumption, we can observe that the appliance's electricity consumption is huge compared to other from June to September. This may be due to the reason that that period is the summer period. We can also observe that there is cyclicity in the pattern and the consumption level is higher at the beginning and end of the year.