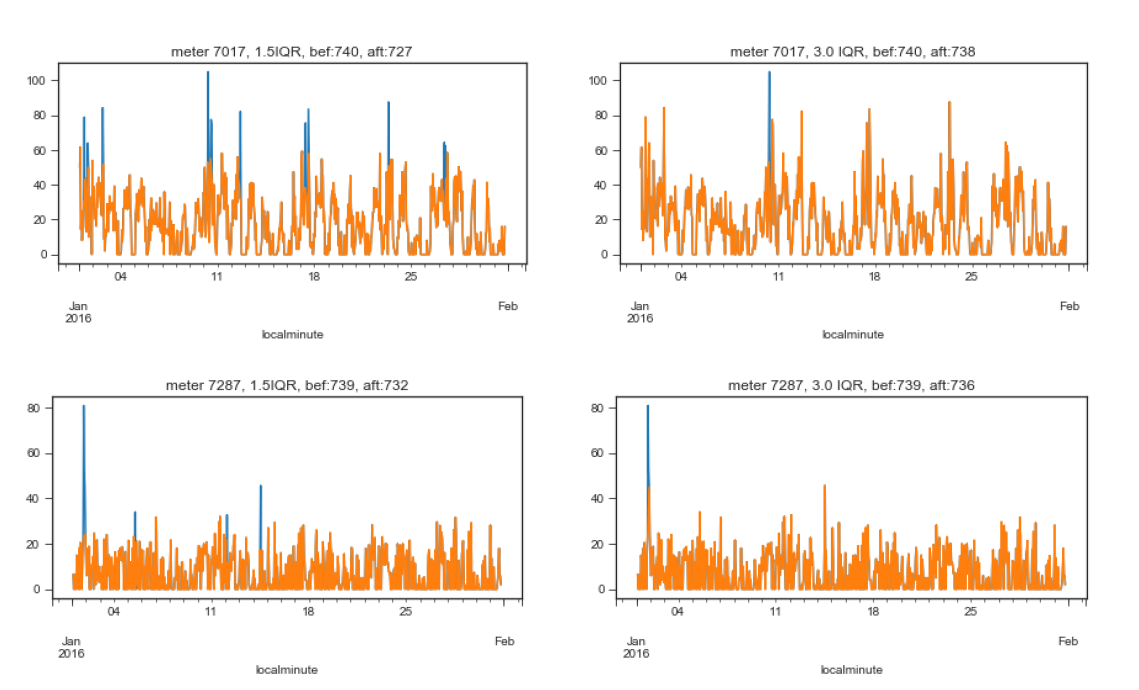
**Q1.2**

We start with selecting the month of January 2016. This is because there were meters that didn’t have readings prior to 2015.

For this part, we have resampled the data hourly and took the mean of the readings as that hour’s new value. If there was no data within that hour, then the reading was simply forward-filled (the empty cell takes the value of the previous cell).

Once resampling was done, we computed the marginal differences hour on hour. We have thus plotted marginal hourly gas usage against time (in hours).

As part of this step, we attempted to do some data cleaning and outlier removal. We removed samples where the marginal value was negative (we expect marginal value to be zero or greater). We also employed the 1.5 IQR rule to elimate samples where the reading was considered an outlier. This means any readings below (Q1 - 1.5IQR) and (Q3+1.5IQR) were considered outliers and removed from the data. The orange plot shows the values after the filter was applied, while the blue plot shows the values before the filter was applied. In addition, on the right column, I plotted the case where a 3.0 IQR rule was used to see how much of the data contains possibly extreme outliers.



. Meter IDs 7017 and 7287. Marginal Hourly Reading against Hour, for Jan 2016.  
Left column is with 1.5 IQR-filtering, right column is with 3.0 IQR filtering.

The process was repeated for all other meters that were not suspected to be faulty or that had sufficient readings for the month of January (greater than 2). The graphs for each valid meter can be seen in Appendix X.