

IOT Sensor Data Analysis

1. Introduction

Monitoring indoor air quality is essential for people's well-being because breathing in a space where the air quality is poor may cause them long-term health problem. The IOT sensors used to collect environmental parameters are meant to measure air quality for a centralized HVAC (heating, ventilation and air conditioning) system.

2. Data Preparation

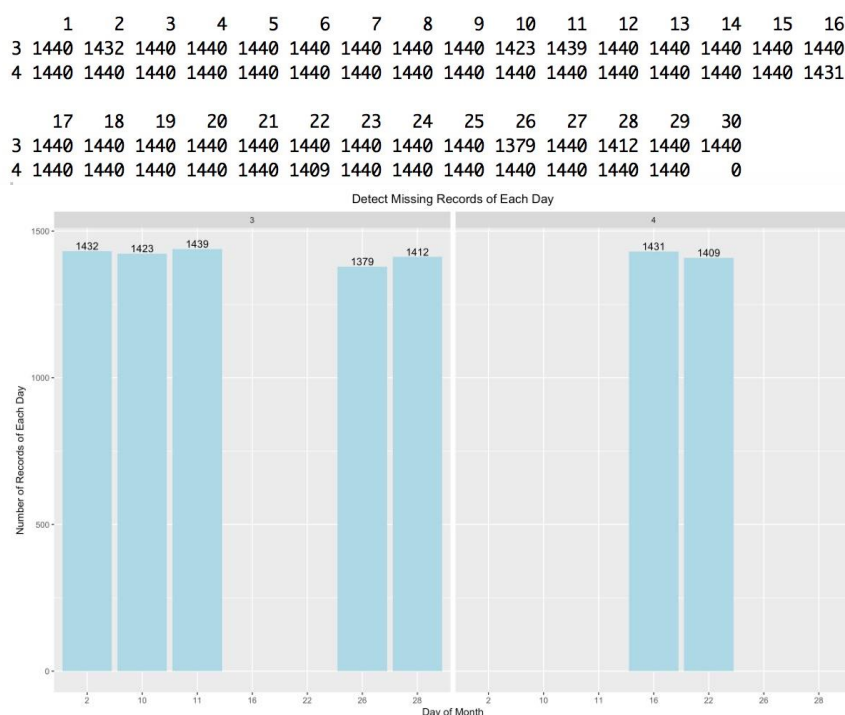
a) Duplication

No duplicated values are detected using R. The sensors did not post the same data point more than once across the datasets.

b) Missing values

There are no missing values in both datasets but there exist some missing records.

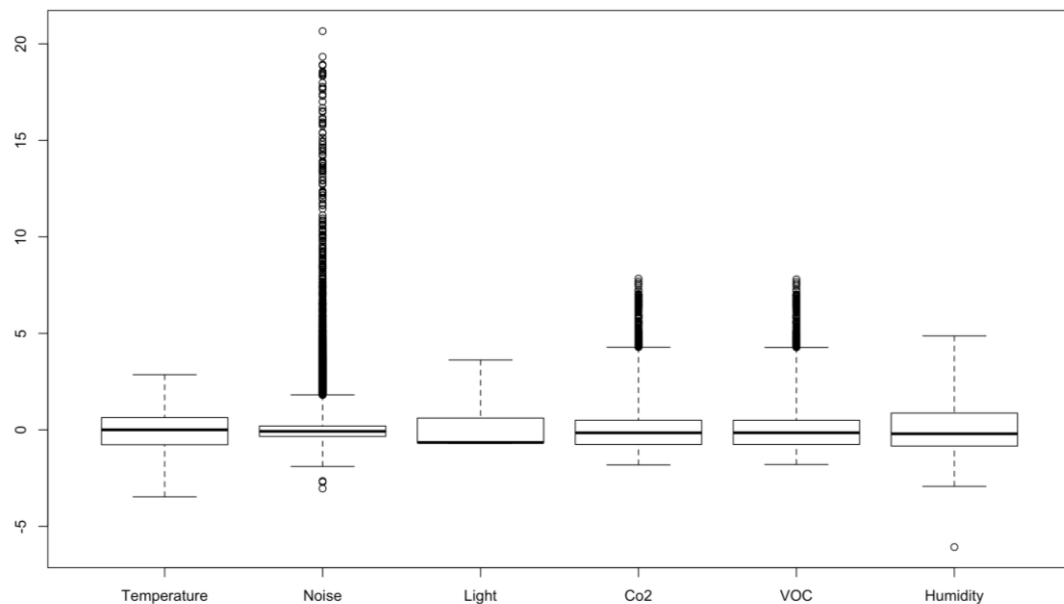
According to the frequency of data posting, which is one record per minute, in one day, there should be $60(\text{minutes}) \times 24(\text{hours}) = 1440$ records; hence, if the number of records is less than 1440 per day or 60 per hour, some records are missing as a result of disconnection with the network.



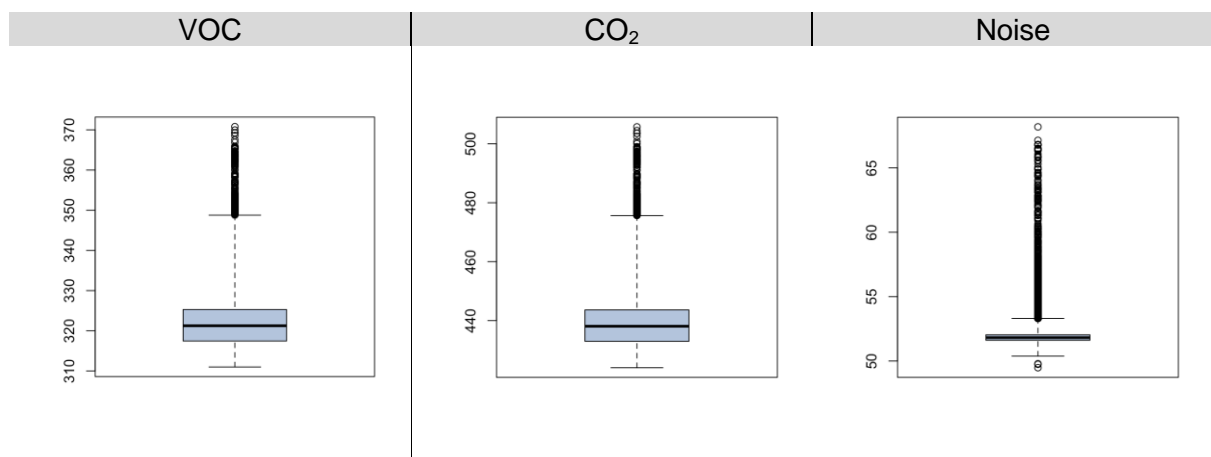
The record began on March 1st, 2017 and ended on April 29th, 2017. Since the missing records are only 0.18% of the datasets, we will not deal with them.

c) Outliers

We apply the Interquartile Rule (IQR) to detect the presence of outliers; most of them are in the range of $Q1 - 1.5IQR$ and $Q3 + 1.5IQR$. Some data points land outside of this range, but they are still considered normal and acceptable for the certain feature. We cannot label them as outliers. Therefore, we expand our detection range to $Q1 - 3IQR$ and $Q3 + 3IQR$.



The magnitude of the variables varies greatly from each other. Scaled boxplot shows that there is only one obvious outlier for Humidity and two for Noise. We will remove the record that contains the Humidity's outlier. For Noise, CO2 and VOC, the "outliers" are all reasonable values, so we will leave them in for the analysis.



3. General Findings

- 1) The sensors are placed in an office with regular working time
- 2) There is a centralized HVAC (heating, ventilation and air conditioning) system in this space
- 3) The timestamp is 7h15m ahead of real time
- 4) There is a curtained window in the wall
- 5) The system never completely shuts down during off-office hours
- 6) Sensors are placed on the ceiling or on the top of the walls
- 7) The air processing module will be automatically turned on at around 8 a.m. and turned off at around 6 p.m. on weekdays, and it will stay off on weekends
- 8) All of the sensors record data at the same time but only one sensor will upload data
- 9) The sensor system backs up or downloads data on the last day of each month
- 10) The sensors probably calibrated measured VOC concentrations to CO₂ - equivalent ppm-values
- 11) Power outage or network disconnection happened a couple times during the 2 months, each lasting around 15 minutes to 1 hour
- 12) More findings related to light

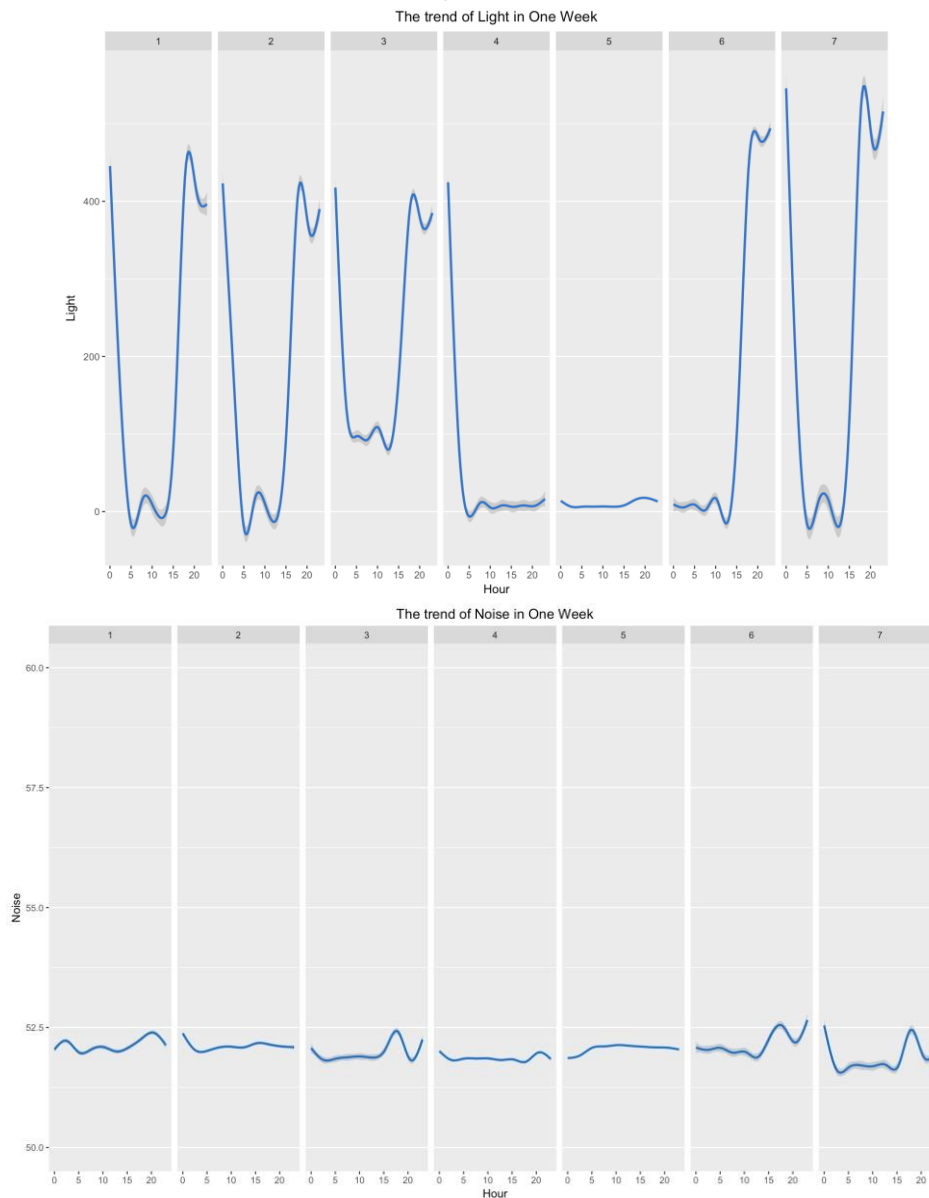
4. Created Features

To verify our assumption and to explore further, we create several new features including:

Datetime_New	The adjusted date and time, 7h and 15ms before the posted time
Month & Month_New	The original month The month after time adjustment
Day & Day_New	The original day The day after time adjustment
Hour & Hour_New	The original hour The hour after time adjustment
Minute & Minute_New	The original minute The minute after time adjustment
Weekday & Weekday_New	The original day of week The day of week after time adjustment
Weekday_Weekend	Group each day into weekday and weekend
NextTime	The time the next record gets posted
Interval_Minutes	The interval minute between 2 records

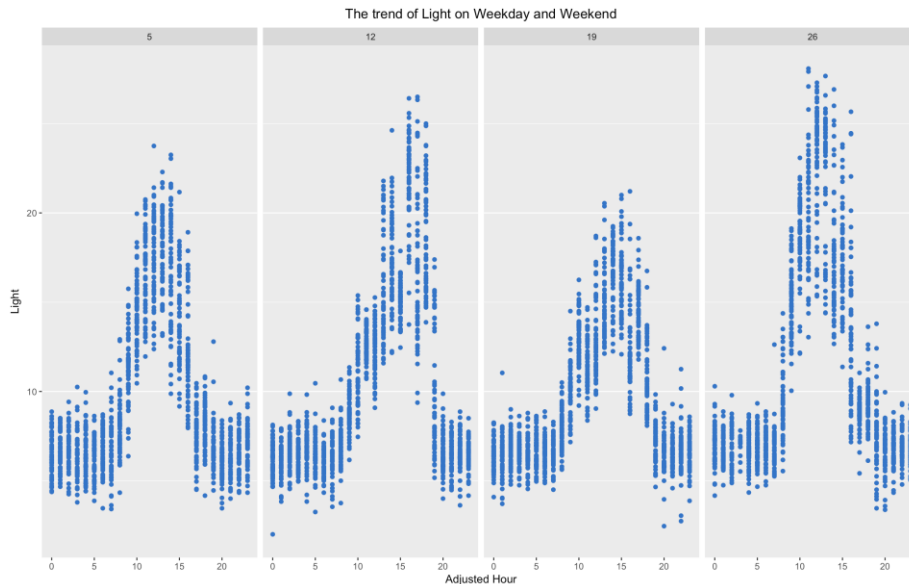
5. Details of the Findings

- 1) The sensors are placed in an office with regular working time
 - 2) There is a centralized HVAC (heating, ventilation and air conditioning) system in this space
 - 3) The timestamp is 7h15m ahead of real time
- First 7 days of March shows light regularity at strange time: high values of light only from 16:15 to 02:15, indicating office-in-use time, which is usually during the day
 - Stable temperature, VOC, CO₂ and humidity, possibly controlled by HVAC system
 - Relatively higher noise, but still conversational level, occurs this period, while noise remains stable at other times
 - The timestamp may be the sensor's unadjusted system time instead of local time since office work usually starts at 9 a.m.
 - Office environment with occasional quiet conversation

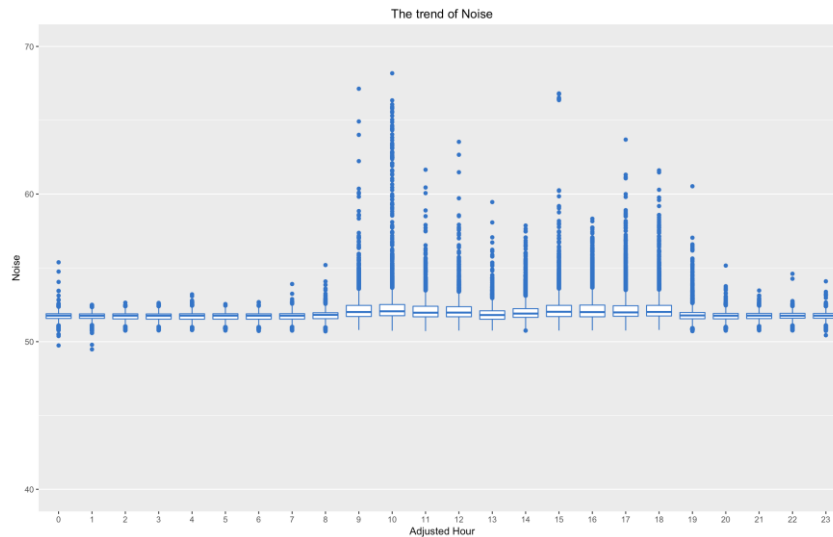


4) There is a curtained window in the wall

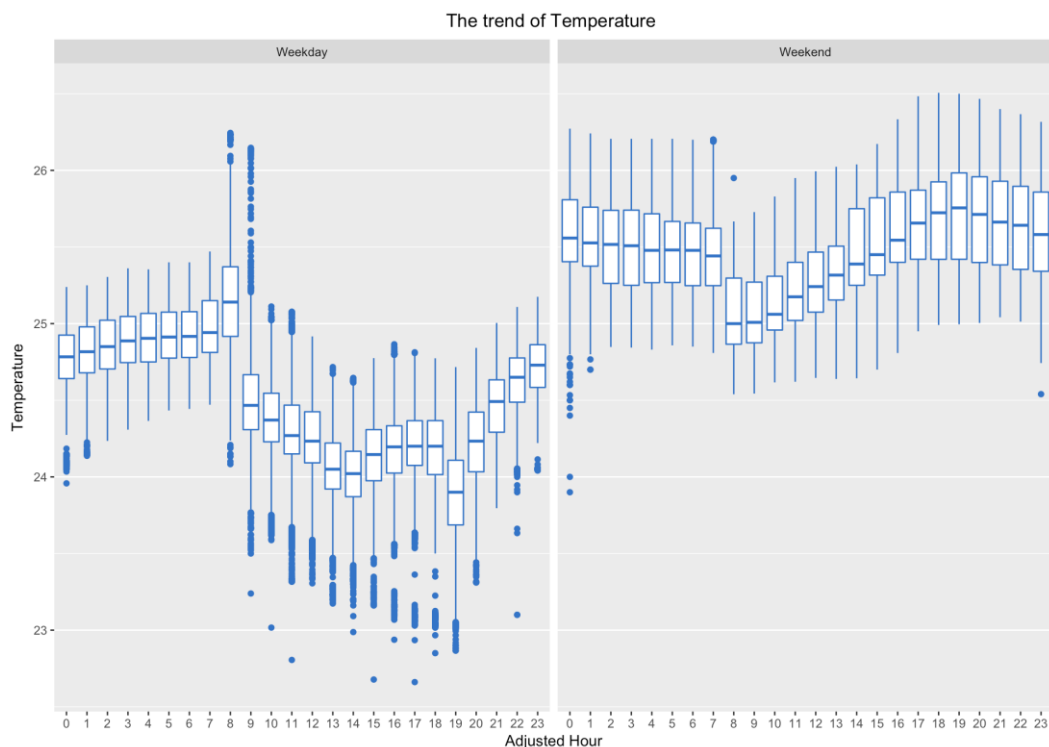
- Comparing to the same periods on other days, light values on Sundays in March (5, 12, 19, 26) are very low (less than 30 Lux)
- Room light is not turned on during these days, indicating no one come in for work
- Light value is highest at noon, which is when the sun shines at its brightest, achieving the highest level of detected light on that day
- The room may have a curtained window that light from outside can penetrate through



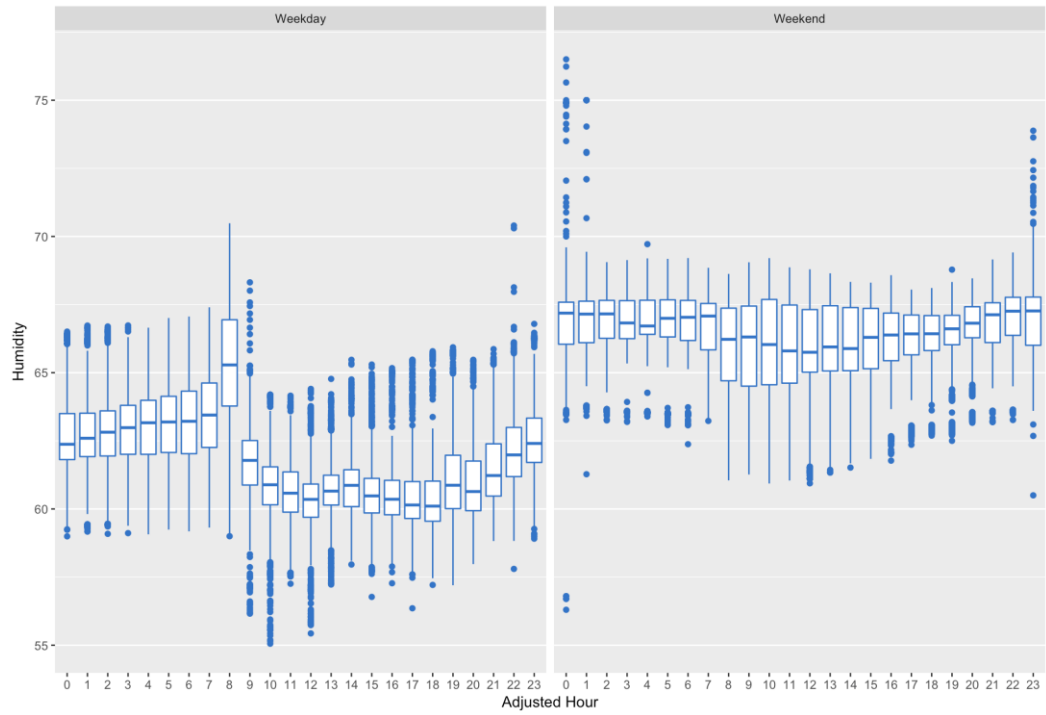
- 5) The HVAC system never completely shuts down during off-office hours
- 6) Sensors are placed on the ceiling or on the top of the walls
- 7) The air processing module will be automatically turned on at around 8 a.m. and turned off at around 6 p.m. on weekdays, and it will stay off on weekends
- Ventilation system in use at all time: stable noise level at ~52db at night time is environmental noise caused by the system stationed near the sensors
 - Sensors are placed on the ceiling or on top of the walls: not close to floor level to capture loud noise and conversations from people



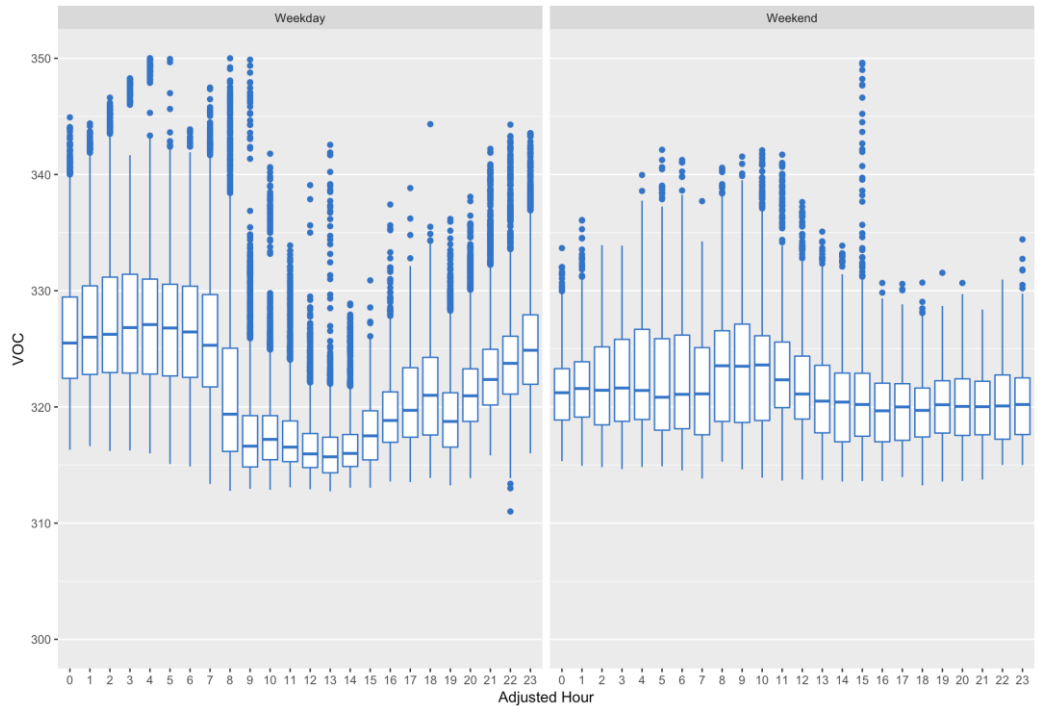
- Temperature, humidity, VOC and CO₂ are all maintained within a narrow range
- Sharp peak observed at around 8 a.m. on weekdays, but there is no obvious peak on weekends: the air processing module automatically turns on at around 8 a.m. to adjust temperature, humidity, VOC and CO₂ to a pre-set level to accommodate the increasing number of room occupants
- At the end of the work day, usually 6 p.m., the module automatically turns off. VOC and CO₂ gradually return to the previous levels

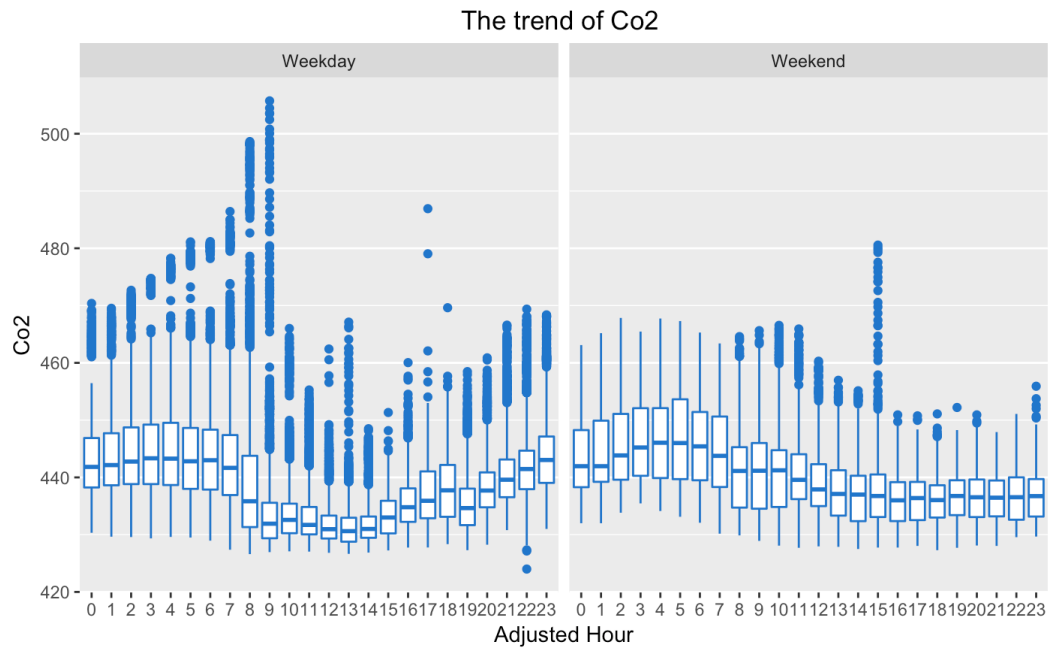


The trend of Humidity



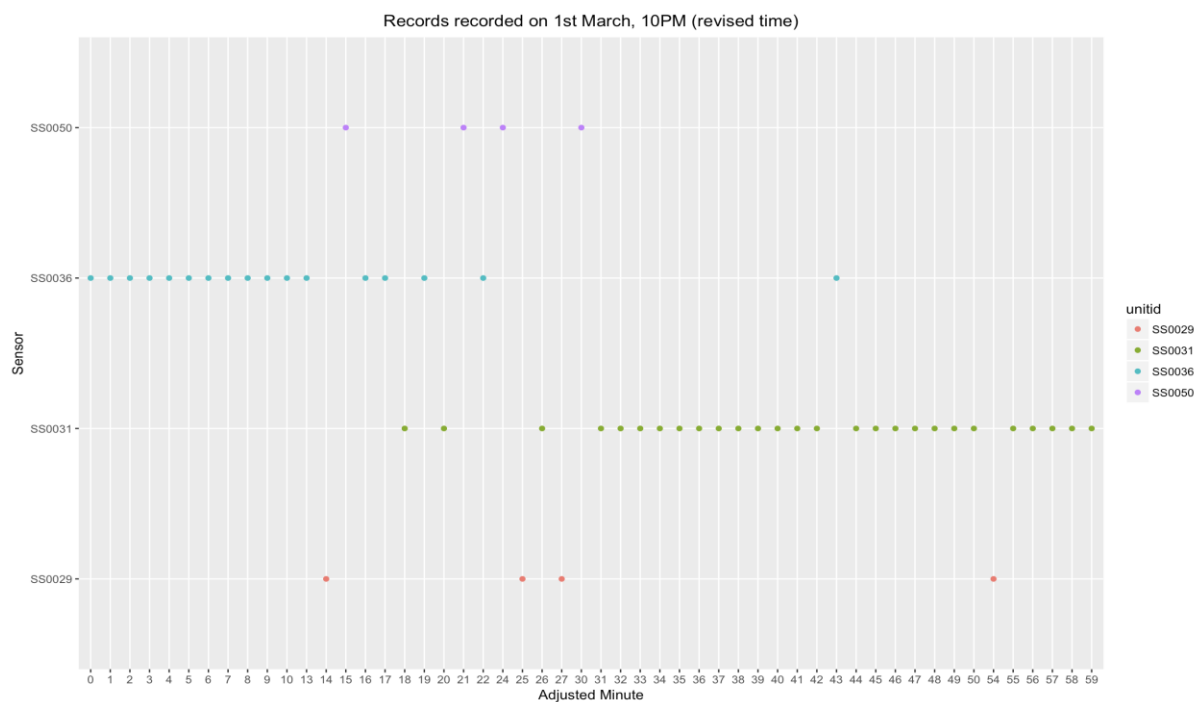
The trend of VOC



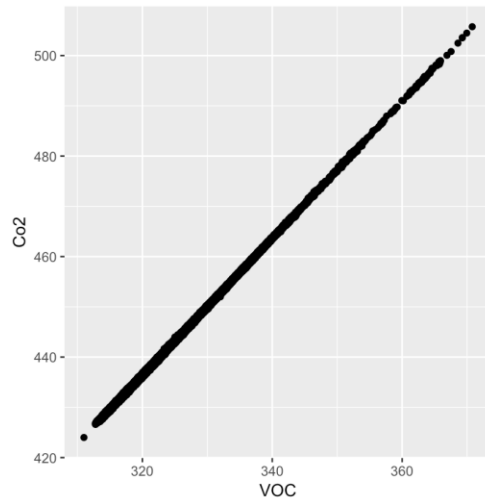
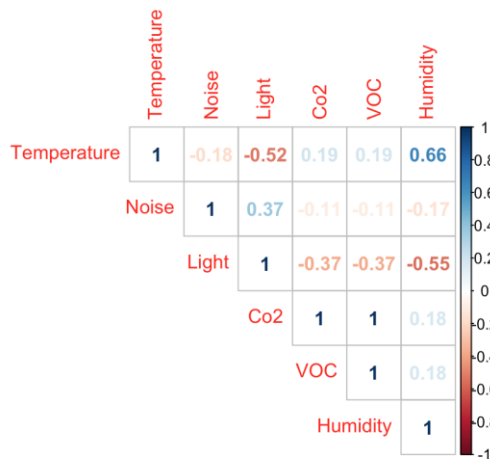


- 8) All of the sensors record data at the same time but only one sensor will upload data
- 9) The sensor system backs up or downloads data on the last day of each month

- 4 sensors are working simultaneously to upload the detected data to the database
- Competitive mechanism: each minute, only the first arrived record will be stored in the database
- Missing records occur whenever the sensors disconnect
- No record was found on March 31th and April 30th when detecting missing records, indicates that the sensor system does not receive new data on the last day of each month



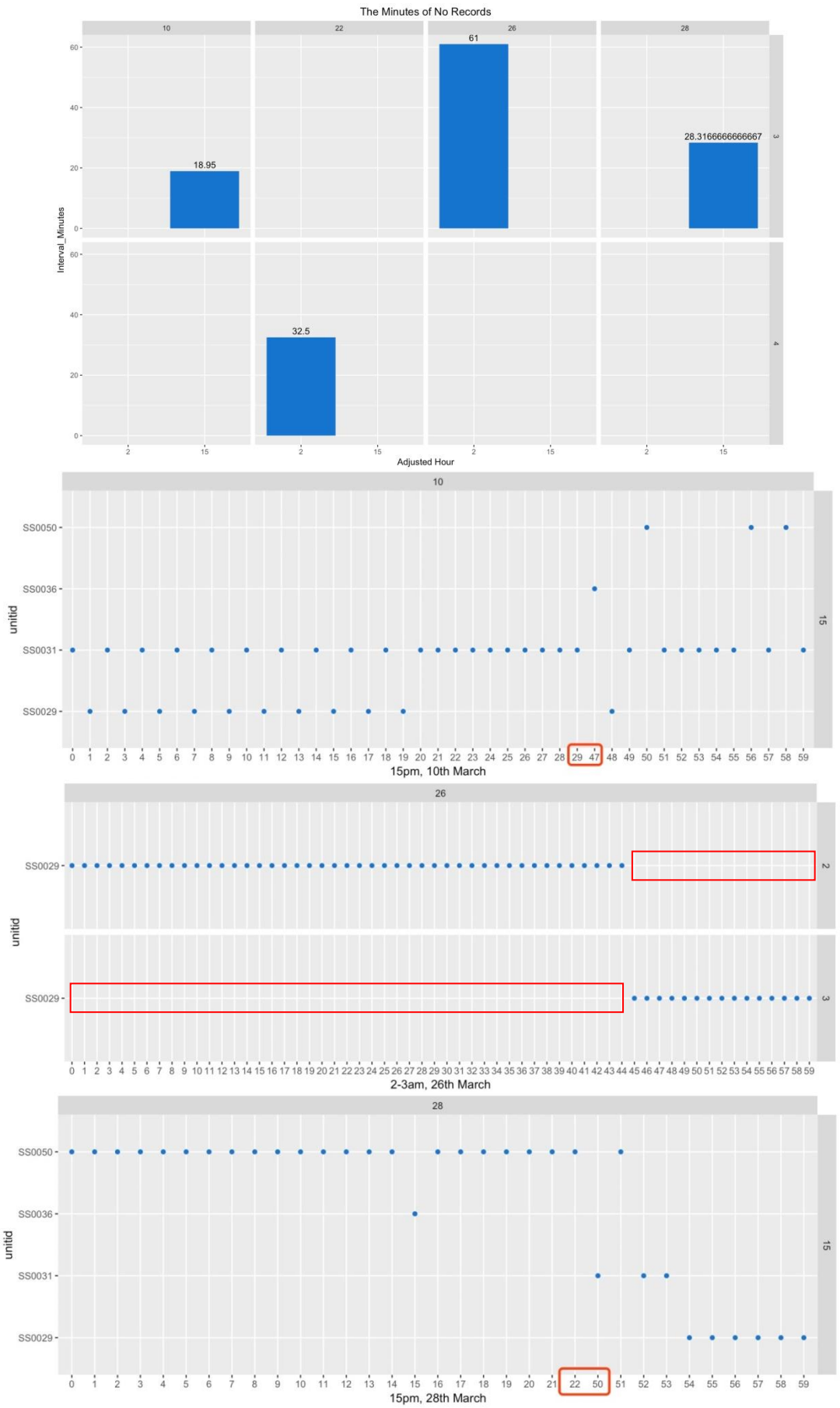
10) The sensors probably calibrated measured VOC concentrations to CO₂ - equivalent ppm-values

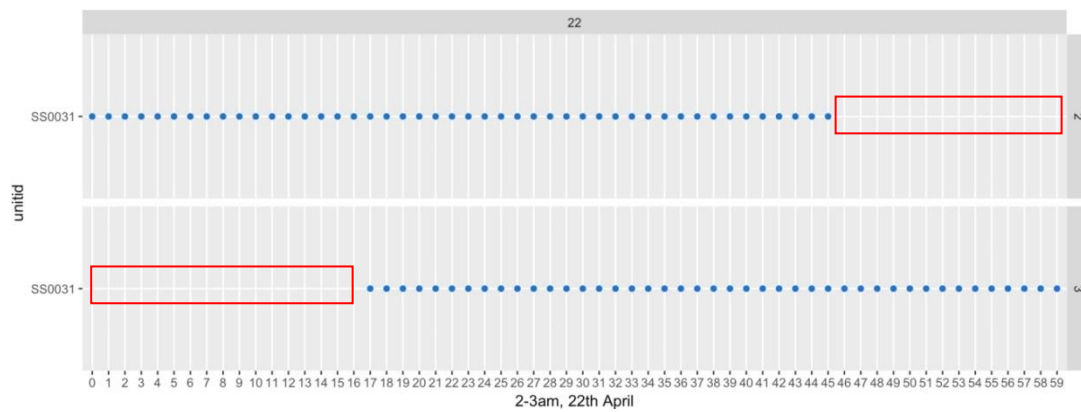


- The correlation between CO₂ and VOC is 1. A perfect positive correlation in this case is questionable
- Human exhales both VOC and CO₂; a rise in one value indicates a rise in the other, corresponding to an increase in room occupancy
- Comparing to CO₂, VOC is easier to detect and be detected more accurately; sensors using VOC sensing only
- The sensors detect a value of VOC, then systematically calculates the value of CO₂ by a certain algorithm based on detected VOC value
- Stable environmental noise (~52db) may be caused by VOC sensor attempting to maintain ventilation rates during unoccupied hours

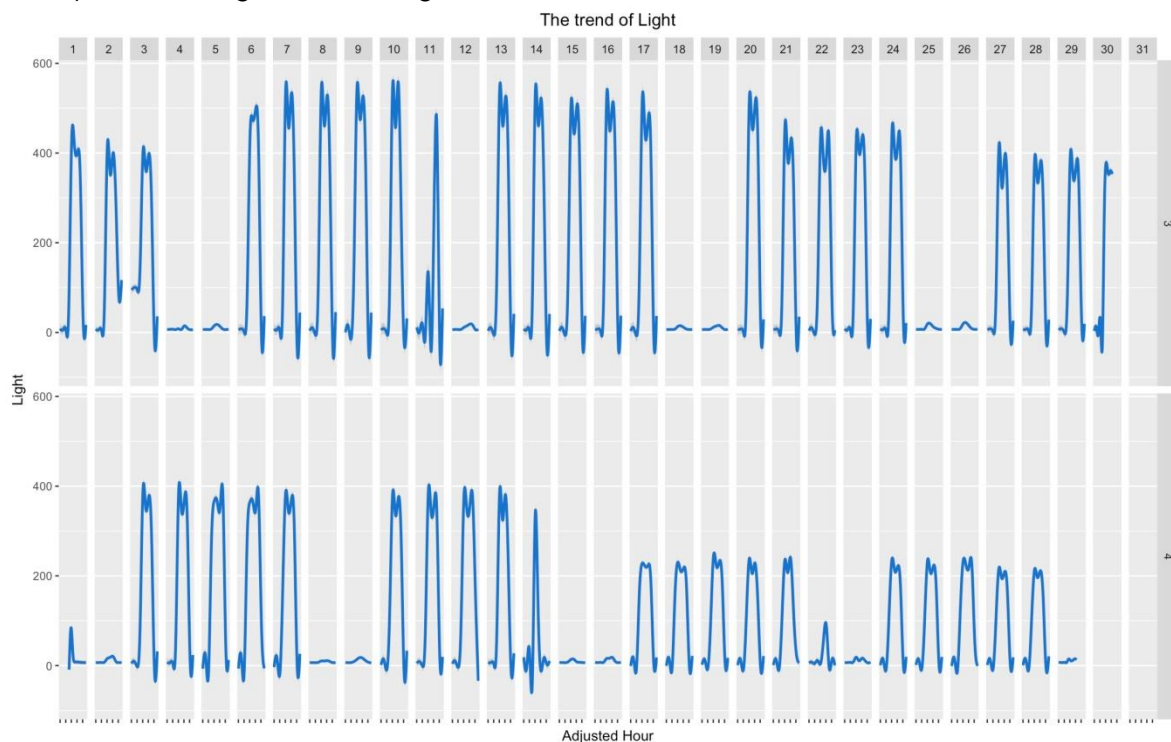
11) Power outage or network disconnection occurred a couple times during the 2 months, each lasting around 15 minutes to 1 hour

- On March 10th, 26th and 28th, and April 22nd, the sensors did not post records for a short period of time
- The missing records are continuous, for example, from 15:30 to 15:46 on March 10, and 2:45 a.m. to 3:44 a.m. on March 26th
- There may be a problem with the power or network or the sensors themselves



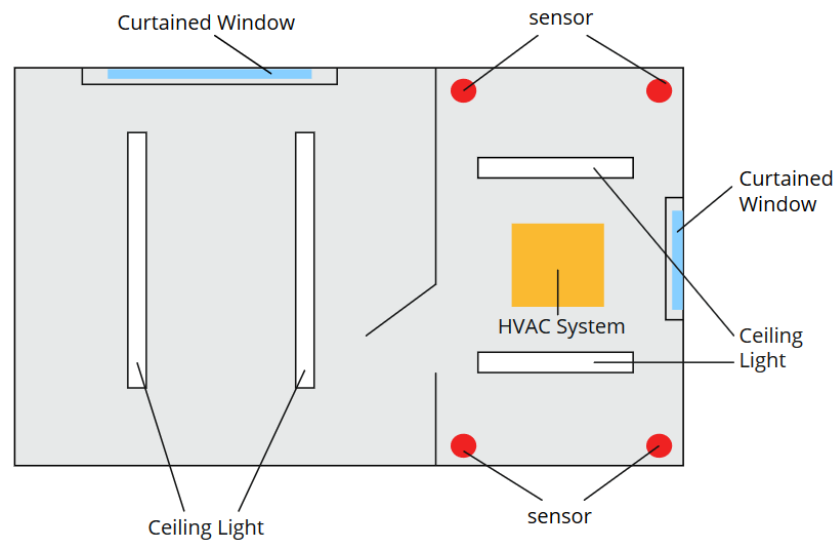


12) More findings related to light



- High values of light occur infrequently during daytime on weekends: office is occupied and light is switched on for a couple hours, possibly for overtime work
- April 14th, Good Friday, is a public holiday: only work half-day and light is switched off in the afternoon
- Usually light value is around 10 Lux at night time. An average value of 100 Lux on the night of March 2nd may result from a table lamp that a staff forgot to switch off
- The values of light for late April are lower than that of other dates, but the rest of the data points are similar
- 2 assumptions of the layout of this office:
 - a) 2 rooms with all the sensors stationed in the smaller one. The smaller room is divided from the bigger room by a glass divider. When all the lights, including the table lamp, are switched on, the light value will reach 500 Lux. On March 2nd, the person working in the smaller room left without switching off the table lamp. From

April 17th to 28th, the person working in the small room is on leave, so the lightness recorded by the sensors is coming from the bigger room.



- b) An open space office with 2 rows of lights. The sensors are placed on the ceiling in between the lights. The office is not as occupied as before after April 14, so the staff decided to switch on only one row of lights for energy saving.

