Exploratory Analysis Report for

Celltraq

Prepared by Quadrant 4 System Corporation

Nov, 2016

**Table of Contents**

[1 Objective 3](#_Toc469059307)

[2 Data Sources 3](#_Toc469059308)

[3 Data Cleaning 4](#_Toc469059309)

[4 Analyses 4](#_Toc469059310)

[4.1 Identification of failed set of batteries. 4](#_Toc469059311)

[4.2 Analysis of samples with high spread in conductance values 5](#_Toc469059312)

[4.2.1 Bin 1 : Standard Deviation > 700 5](#_Toc469059313)

[4.2.2 Bin 2 : Standard Deviation > 500 & < 700 5](#_Toc469059314)

[4.2.3 Similar analysis of other bins 6](#_Toc469059315)

[4.2.4 Intuitions from the initial set of exploration. 6](#_Toc469059316)

[4.3 Analysis of samples in conjunction with discharge voltage 7](#_Toc469059317)

[4.3.1 Trend I: Cup Shaped Profile. 7](#_Toc469059318)

[4.3.2 Trend II: “W” Shaped Profile. 8](#_Toc469059319)

[4.4 Analysis of sample of 1000 batteries 10](#_Toc469059320)

[4.4.1 Analysis of samples with DOD < .65 (52 cases) 10](#_Toc469059321)

[4.4.2 Analysis of samples with DOD > .65 (948 cases) 11](#_Toc469059322)

[4.5 Analysis of new sample of 1000 batteries 14](#_Toc469059323)

[4.5.1 Bin 1: Conductance Drop < 0.2 15](#_Toc469059324)

[4.5.2 Bin 2: Conductance Drop > 0.2 < 0.5 (7 cases) 16](#_Toc469059325)

[4.5.3 Bin 3: Conductance Drop > 0.5 < 0.65 (4 cases) 16](#_Toc469059326)

[4.5.4 Bin 4: Conductance Drop > 0.65 < 0.8 (100 cases,30 sampled) 17](#_Toc469059327)

[4.5.5 Bin 5: Conductance Drop > 0.80 < 0.9 (492 cases,30 sampled) 18](#_Toc469059328)

[4.5.6 Bin 6: Conductance Drop > 0.9 (394 cases,30 sampled) 19](#_Toc469059329)

[4.5.7 Intuitions and inferences from the analysis. 20](#_Toc469059330)

[5 Intuitions and inferences 22](#_Toc469059331)

# Objective

The objective of the of the data Discovery phase is to understand the data that is available from Celltraq data sources and evaluate the sufficiency of data for building models for predicting potential failure of batteries.

# Data Sources

The data was provided in the form of csv files. There were 5 types of data sources which were various variables which would be useful for indicating the state of health or state of charge of batteries. The key data file types were the following

|  |  |
| --- | --- |
| Data Type | Variables Involved |
| Battery Discharge Data | Battery ID, Measurement Timestamp, Voltage, Current, Site Name, Plant Name, String Name, Manufacturer, Model, Battery type, Battery voltage |
| Battery Conductance Data | Battery ID, Measurement Timestamp, Conductance, Conductance high alarm, Conductance high warning, Conductance low warning, Conductance low alarm, Site Name, Plant Name, String Name, Manufacturer, Model, Battery type, Battery voltage |
| Battery Voltage-Temperature | Battery ID, Measurement Timestamp, Voltage, Voltage high alarm, Voltage high warning, Voltage low warning, Voltage low alarm, Temperature, Temperature high alarm, Temperature high warning, Temperature low warning, Temperature low alarm, Site Name, Plant Name, String Name, Manufacturer, Model, Battery type, Battery voltage. |
| String Discharge Data | Battery ID, Measurement Timestamp, Voltage, Current, Site Name, Plant Name, String Name, Manufacturer, Model, Battery type, Battery voltage |
| String Voltage-Temperature | Battery ID, Measurement Timestamp, Voltage, Voltage high alarm, Voltage high warning, Voltage low warning, Voltage low alarm, Temperature, Temperature high alarm, Temperature high warning, Temperature low warning, Temperature low alarm, Site Name, Plant Name, String Name, Manufacturer, Model, Battery type, Battery voltage. |

For all the analyses the information related to discharge data and conductance data were predominantly used.

# Data Cleaning

There was completeness in terms of data points available in all the data sources. However, the major task was to bring data in multiple files into a single data frame format. The major cleaning work which was involved was to bring a particular batteries all important variables into a single data frame so that it was easy for further analysis.

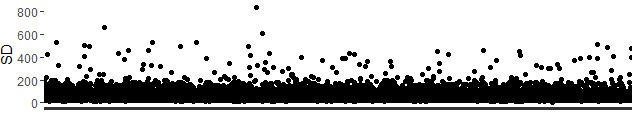
# Analyses

The objective of the analysis was two fold

1. Identify the failed set of batteries by looking at trends from the data
2. Initial study on the feasibility of building predictive models for battery failure

## Identification of failed set of batteries.

Based on discussions with Celltraq business team, a prominent trend which needed to be pursued was sharp drop in conductance. Accordingly all batteries were selected and the standard deviation of conductance values for each battery was used to identify those cases who had large drop in conductance to those which had smaller drop in conductance. The scatter plot of all the batteries with their respective spread of conductance values are as shown in the plot below.



From the scatter plot, it is quite evident that there is a clear cut demarcation at the standard deviation of around 200. Most of the batteries fall under the 200 level which is evident from the dark shades. The dots above the 200 mark represents those batteries who potentially would have higher drops in conductance. To observe and study the behavior of batteries with high drop in conductance, the cut off value of standard deviation of 200 was used and a smaller sample of batteries were selected.

## Analysis of samples with high spread in conductance values

There were 131 batteries who had high spread in conductance (Figure-2). The ranges of standard deviation values for conductance ranged from 200 to around 840. These batteries were separated into various bins as per their

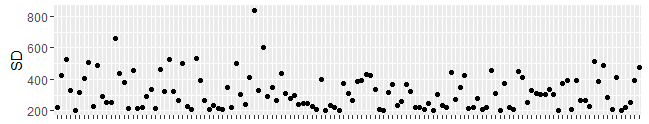


Figure – 2

standard deviation ranges and were further observed for studying their characteristics.

### Bin 1 : Standard Deviation > 700

In this bin there was only one battery. The plot of the battery is as show below in Figure-3 .

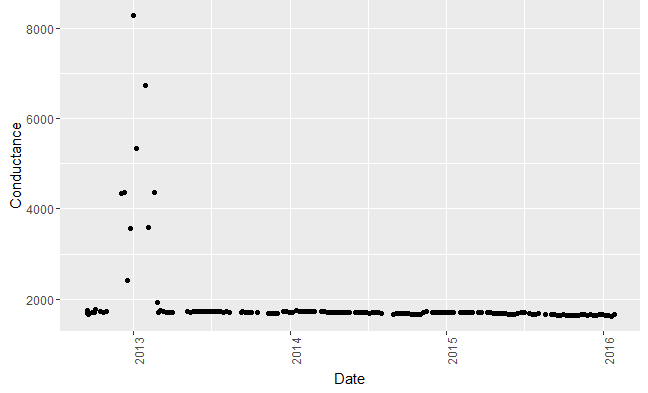


Figure-3

This battery is more of an outlier than an indicator of failing batteries, because this battery only shows a spike in values at a certain point in time and then drop from the spike. For the rest of the time frame, the trend shows constant values.

### Bin 2 : Standard Deviation > 500 & < 700

There were 2 batteries within this bin between 500 and 700 ranges of standard deviation of conductance. The plots of these batteries are as shown in figure-4.

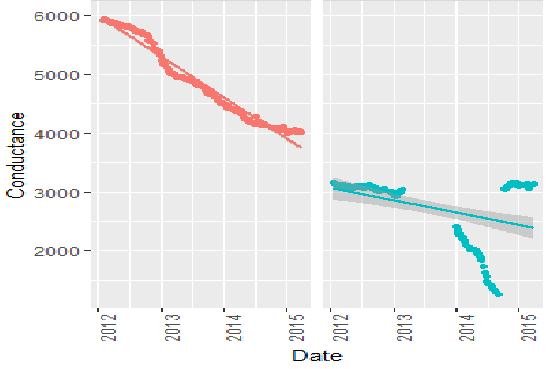


Figure-4

As seen, these batteries showed some sharp falling trends in conductance values over time. Another distinct feature which was observed was the one as seen for the battery 2 (right side of the plot). This battery showed a sharp downward trend followed by a rise to its original level.

### Similar analysis of other bins

Similar to the above analysis all other bins, up to standard deviation value of 200, were studied. A sample set of these batteries is as show in figure-5

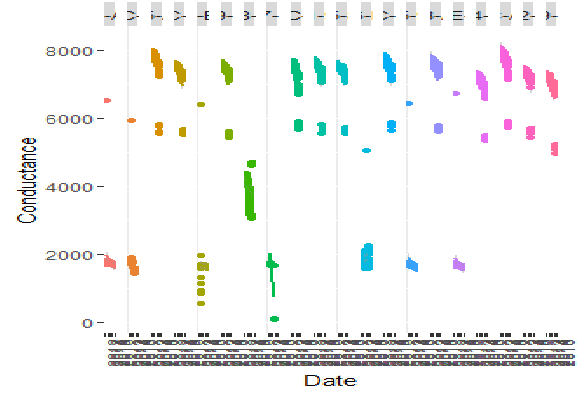


Figure-5

Most of the batteries analysed within the range showed similar trends in conductance. Many of them also showed trend where sharp drop was followed by rise to the original level, similar to battery 2 in section 4.2.2.

### Intuitions from the initial set of exploration.

The early trend in conductance as observed from the above explained analysis were reviewed with the business team of Celltraq. The main action points and further steps derived from the discussions are as elucidated below.

1. Sharp drop in conductance would be a necessary condition which indicates failing batteries. However, this is not a sufficient condition.
2. Conductance values have to be observed in conjunction with discharge voltages to pin point potential failure cases.
3. Some key variables that needs to be introduced would be depth of discharge values (from voltage perspective) to help in further analysis.

## Analysis of samples in conjunction with discharge voltage

Based on intuitions derived after the initial set of explorations, further analysis was carried out by associating current and discharge voltage trends with the conductance values. Trends were observed over time for various samples of batteries. From the samples observed two distinct trends were observed. These trends are as described below.

### Trend I: Cup Shaped Profile.

One distinct trend which showed up is the “Cup Shaped Profile” as shown below in figure-6

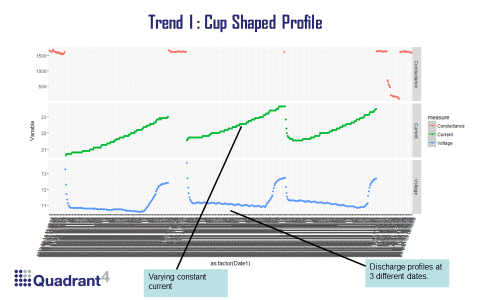


Figure-6

The above figure, is a scatter plot for a sample battery. This figure shows three different variables i.e conductance, temperature and discharge voltage, over time for the selected battery. The top most block within the scatter plot is the conductance, the middle block corresponds to current and the bottom one shows the discharge voltage. Looking at the discharge voltage, one can observe three large cup shaped profiles. Each cup shaped profile is for a specific date and across multiple time periods for the same date. Each of the cup shaped discharge voltage profile would also have a corresponding trend in current.

A closer look at the discharge profile is shown in figure 7 below

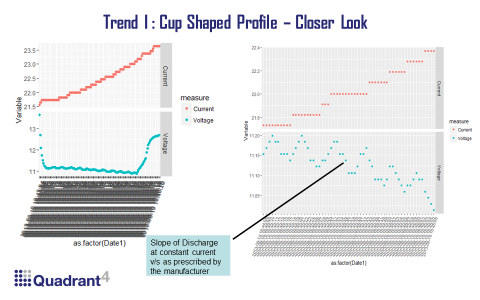


Figure-7

The left side plot in figure-7 is the values of current and voltage corresponding to a particular date where the readings were taken. The right side plot shows a closer look at the cup shaped profile at some instance of time within the same date.

From the right hand side profile it can be seen that the “Cup Shaped Profile” is comprised of smaller cyclical discharge profiles corresponding to some constant current.

**Intuition from the Discharge Profile**

One intuition we can take from the discharge profile (right side of figure 7) is the significance of the cyclical trends in discharge voltage. A discharge voltage in itself wouldn’t provide any significant insights. However, what would be significant is when the observed discharge profile is compared with the manufacturer prescribed acceptable standards of discharge drops. The slope of the observed discharge voltage has to be compared with the manufacturer’s prescribed value to arrive at a conclusion if the trends represent normal ranges or whether it indicates potential failure.

### Trend II: “W” Shaped Profile.

The second overall trend which was observed was the “W Profile” which is as shown in figure-8 below.

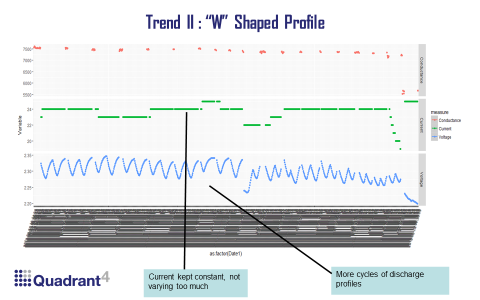


Figure-8

As the figure indicates the defining characteristic of this profile is larger frequencies of discharge over time. In addition there is no indication of any large drop as seen in the cup shaped profile. Each of the “W” profile is a set of readings corresponding to a single day when the reading was taken. A closer perspective of each of the “W” profile is shown in figure-9 below.

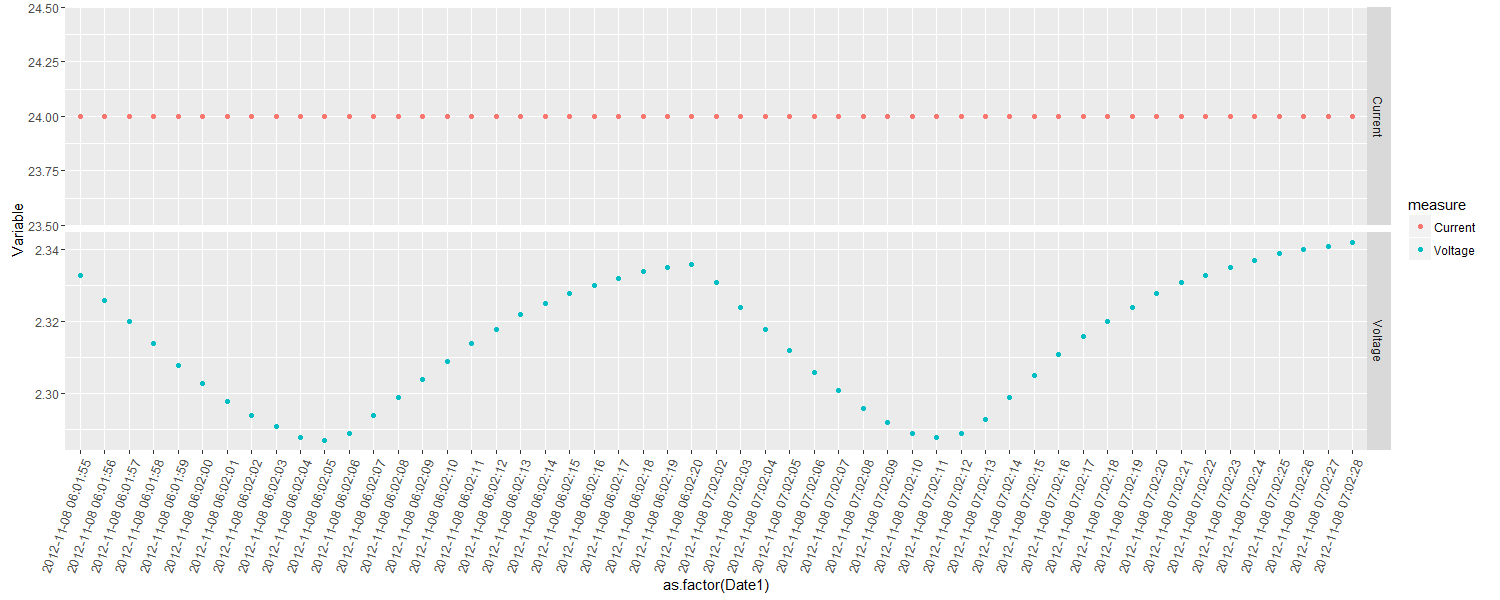


Figure – 9

A closer look at this profile will reveal that the “W” profile is in fact a combination of two “Coup-de-Fouet” effects separated by a time interval of approximately 1 minute.

The two distinct profiles per-se wouldn’t be an indicator of any significant features. However, what would be more significant are features derived from two features derived from these profiles namely,

1. Depth of discharge ( from voltage perspective)
2. Slope of the discharge profile.

The pictorial representation of these two features are as shown in figure-10 below.



Slope of the discharge profile

Depth of discharge(DOD) from a voltage perspective

Calculated the slope of each of these discharge and charge profiles

Calculated the DOD for each discharge cycle

Figure – 10

For further analysis of the data, the DOD and slope of discharge profile were calculated and included as new features. A new sample of 1000 batteries were taken with these two additional features and analyzed further. The details of the analysis are as described below.

## Analysis of sample of 1000 batteries

A random sample of 1000 batteries were taken which had two additional features of depth of discharge(DOD) and slope of discharged. These 1000 batteries were divided into two bins with respect to the DOD.

Bin1: Batteries which had DOD < .65 (52 cases)

Bin2: Rest of the batteries (948 cases).

Further trends observed in these two bins are as described below.

### Analysis of samples with DOD < .65 (52 cases)

The analysis of the 52 cases which had DOD < .65 showed two distinct trends. These are as shown in figure-11 below.

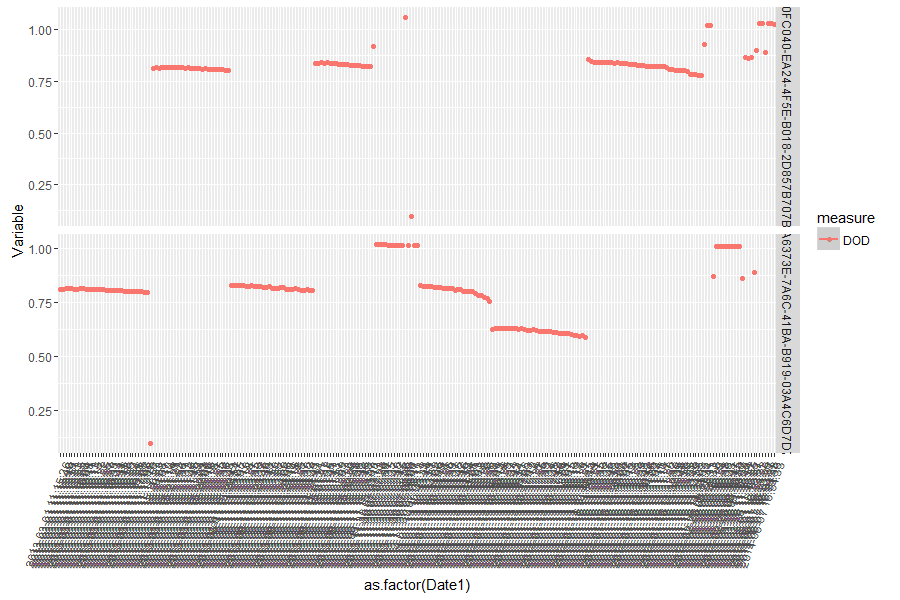


Figure-11

As seen, figure-11 consists of two rows of plots i.e top row and the bottom row. The top row shows a sample plot of one distinct trend observed and the bottom row shows the second trend observed.

**Trend as represented in top row.**

There are 46 batteries, from the sample of 52 batteries, which showed similar trend as observed in the top row. The top row is the plot of one representative battery from this list of 46 batteries. The left side of the top row shows the conductance and the right side shows the trend in DOD of the same representative battery. The distinct feature about this set of 46 batteries was that there were only one of instances where the depth of discharge showed values less than .65. This can be seen as the one off point below the .25 DOD level in the right block of the top row.

**Trend as represented in bottom row.**

There are 6 batteries which showed trends similar to the ones seen in the bottom row. The distinct feature of this trend is the conductance profile which shows a large dip and then retraction. All these 6 cases show multiple values which were below the DOD level of .65.

**Intuitions from the analysis of 52 batteries with DOD < .65**

The following are the intuitions which could be built from the analysis so far.

* One of values of large drop in DOD might not be a potential case for failing battery.
* Multiple values of large drop in DOD in combination with sharp drop in conductance values might be a potential case.
* However to substantiate such inferences, further cases have to be observed and trends studied.

### Analysis of samples with DOD > .65 (948 cases)

In line with the intuitions derived, the balance 948 cases were also analyzed to observe trends. The analysis was done for a random sample taken from this set of 948 batteries. A total sample of 165 batteries were taken and they were divided into 3 tranches of 50, 50 and 65 batteries each. Each tranche was further divided into 3 bins each based on the DOD values of these tranches. The 3 bins taken are as below

Bin1: DOD < 0.8

Bin2: DOD between 0.8 and 0.9

Bin3: DOD > 0.9

A representative analysis of on tranche of 50 batteries is as described below.

#### Bin1: DOD < 0.8

The representative trend for Bin1 is as shown below in figure-12

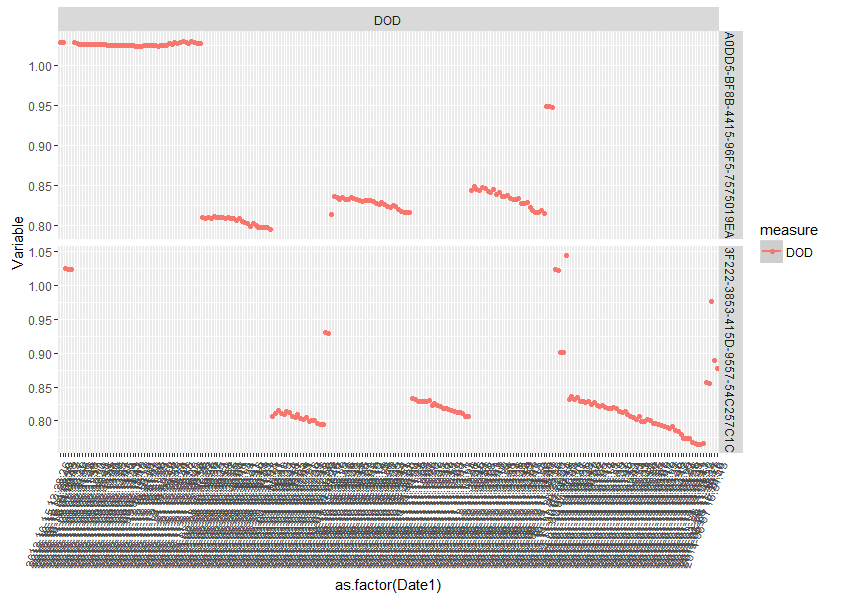
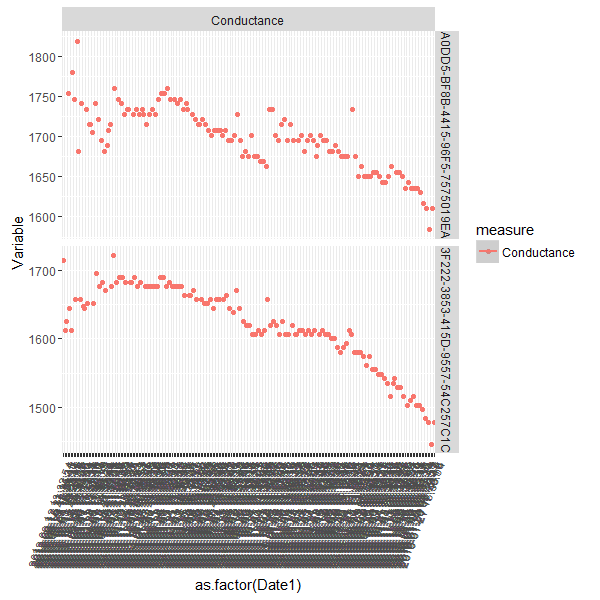


Figure-12

The top row and bottom row represents representative cases of two batteries in contrast. For the first row the number of points which were below the 0.8 level of DOD were far lesser than the bottom battery. The bottom battery showed far more points which were below the DOD of 0.8. Correspondingly, the percentage drop in conductance values were much higher for the bottom row than the top row. So the trend which can be taken out of these representative plots is the following

* Batteries with higher proportion of values falling below the DOD level of 0.8 showed sharper percentage drop in conductance values.

This specific trend was observed in all the batteries in Bin1.

#### Bin2: DOD between 0.8 and 0.9

The representative trends for two batteries within this bin is as shown in figure-13 below.

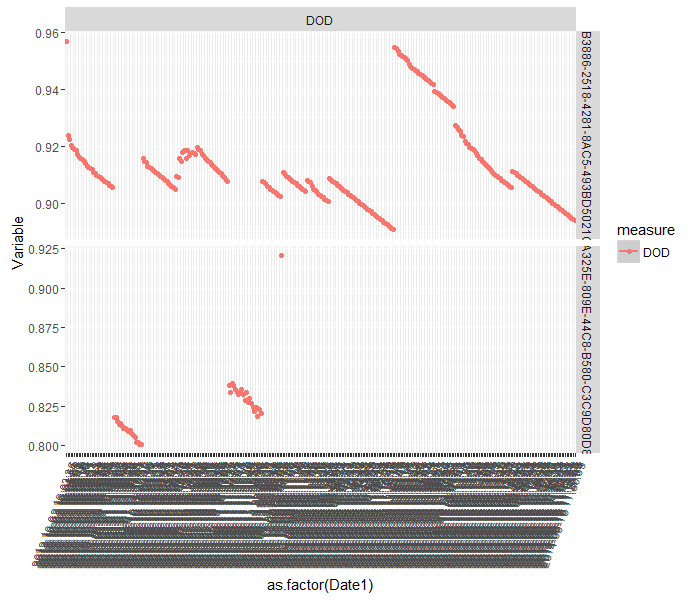
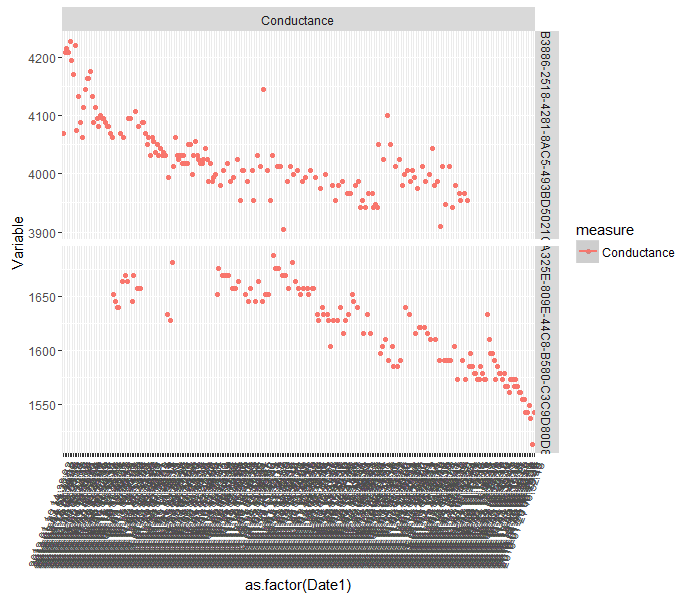


Figure-13

General trends which were observed within the bin are as follows.

* The conductance slopes were generally milder
* Those cases where conductance slope showed sharper drops also had their corresponding DOD values falling in the lower ranges of the bin i.e. near 0.8

#### Bin3: DOD greater than 0.9

The trend for a representative set of batteries within this bin is as shown below in figure-14

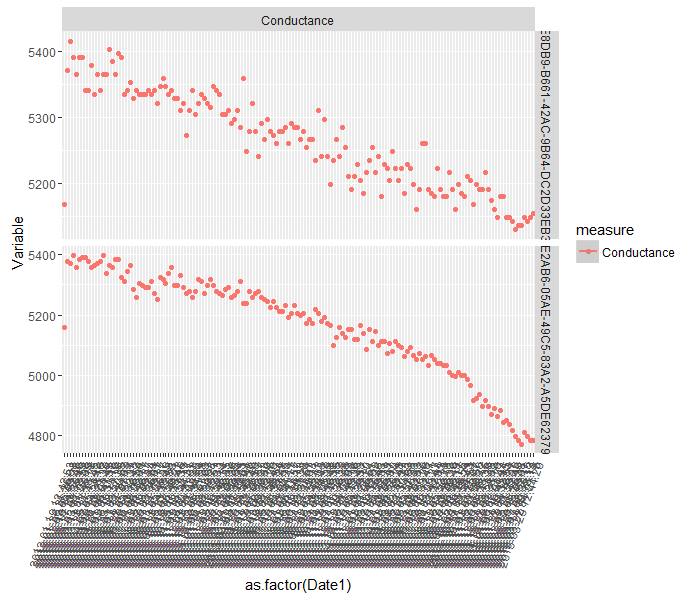


Figure – 14

As seen from the figure, the conductance slopes were generally milder and the no specific association with the DOD values were observed for this bin.

#### Consolidated observations from all 165 cases.

Batteries: 20

DOD: Mostly between 0.8 – 0.85

DOD < 0.8 + Sharpest Conductance drop association: All

Batteries: 17

DOD: Mostly between 0.8 – 0.85

DOD < 0.8 + Sharpest Conductance drop association: All

Batteries: 28

DOD: Mostly between 0.8 – 0.85

DOD < 0.8 + Sharpest Conductance drop association:3/5

Batteries: 22

DOD: Mostly between 0.8 – 0.85

Sharpest Conductance drop associated with lower DOD values:5/5

Batteries: 26

DOD: Mostly between 0.8 – 0.85

Sharpest Conductance drop associated with lower DOD values: 3/5

Batteries: 31

DOD: Mostly between 0.8 – 0.85

Sharpest Conductance drop associated with lower DOD values: 6/6

Batteries: 8

Conductance: Gradual slopes

No specific association noticed between DOD and conductance

Batteries: 7

Conductance: Gradual slopes

No specific association noticed between DOD and conductance

Batteries: 6

Conductance: Gradual slopes

No specific association noticed between DOD and conductance

Tranche 1: 50 Batteries

Tranche 2: 50 Batteries

Tranche 3: 65 Batteries

Bin1:

DOD

< 0.8

Bin2:

DOD > 0.8 < 0.9

Bin1:

DOD

> 0.9

**General observations from the analysis:**

The major observations that can be summarized from the analysis of these 165 sets of batteries are the following

1. DOD values for most of the batteries lie between the ranges of 0.8 and 0.85
2. The percentage of points lower than 0.8 is associated with sharper drops in conductance i.e. more the points which are below 0.8 more the drop in conductance.
3. Even for those bins whose DOD were higher than 0.8, the sharpest drop in conductance was associated with more values at the lower ranges of the DOD

Based on the intuitions derived from the above analysis a new feature related to the percentage of points below DOD of 0.8 was included to the existing features. A new sample of 1000 batteries were further analyzed. The results and trends for this new set is as described in the section below.

## Analysis of new sample of 1000 batteries

The new sample of 1000 batteries were divided into multiple bins with respect to the percentage drop in conductance from the peak value of conductance. The bins which were used for analysis are the following

Bin1: Percentage drop in conductance to < 20% of peak value

Bin2: Percentage drop in conductance between: 20% - 50% of peak value

Bin3: Percentage drop in conductance between: 50% - 65% of peak value

Bin4: Percentage drop in conductance between: 65% - 80% of peak value

Bin5: Percentage drop in conductance between: 80% - 90% of peak value

Bin6: Percentage drop in conductance > 90% of peak value

### Bin 1: Conductance Drop < 0.2

There are 2 batteries which belonged to this bin. The profiles of these batteries are as shown in the figure below.

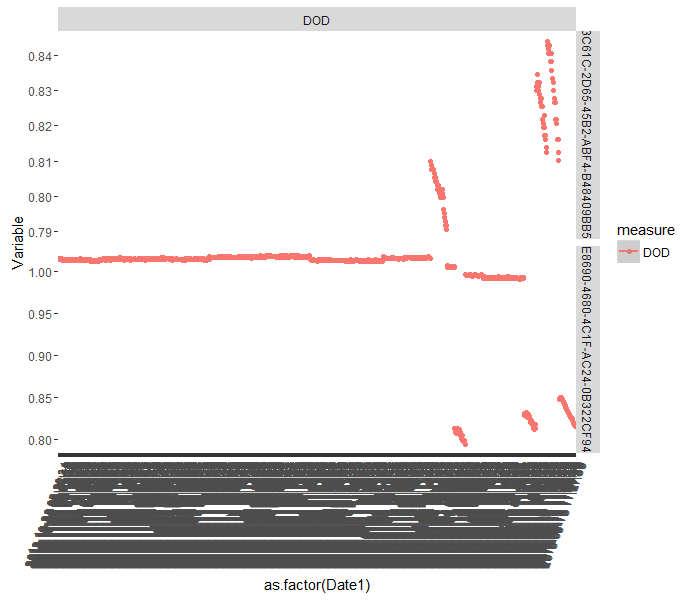
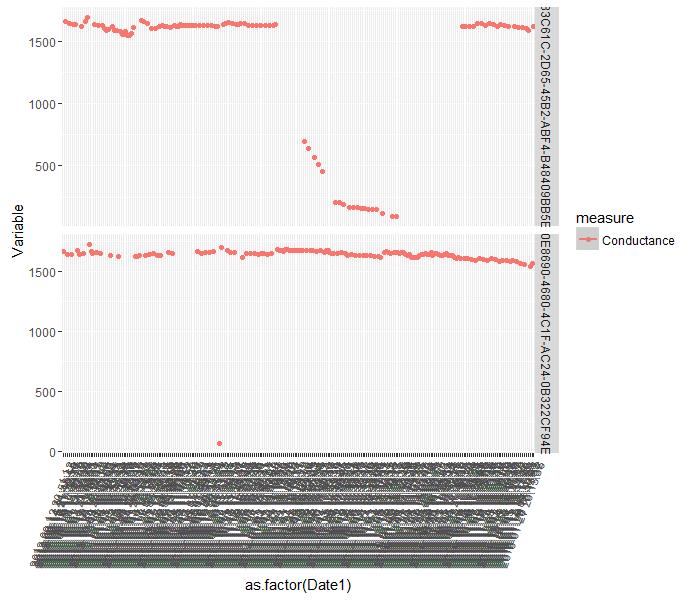


Figure-15

The observations are as follows

* The first battery, has a dropping and retracting trend, with about 10% values below DOD of 0.8
* The second case is more because of an outlier in conductance value. The percentage of values below 0.8 DOD is miniscule.

### Bin 2: Conductance Drop > 0.2 < 0.5 (7 cases)

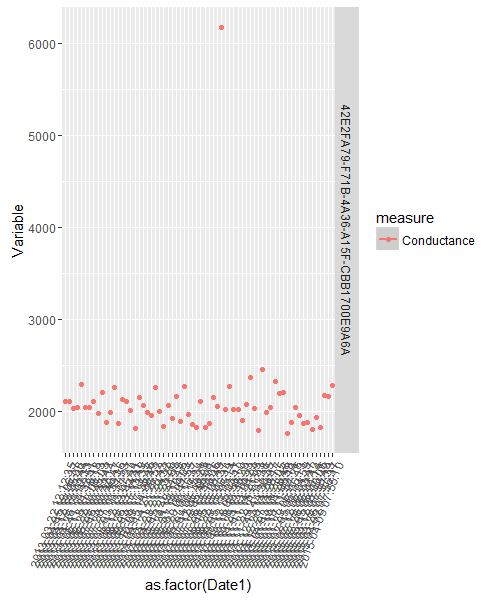
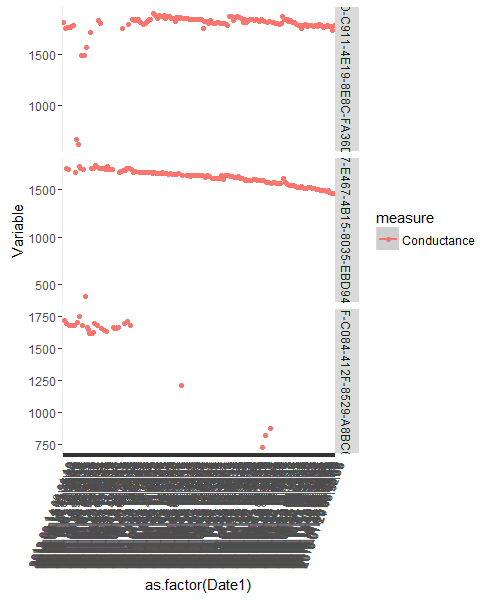
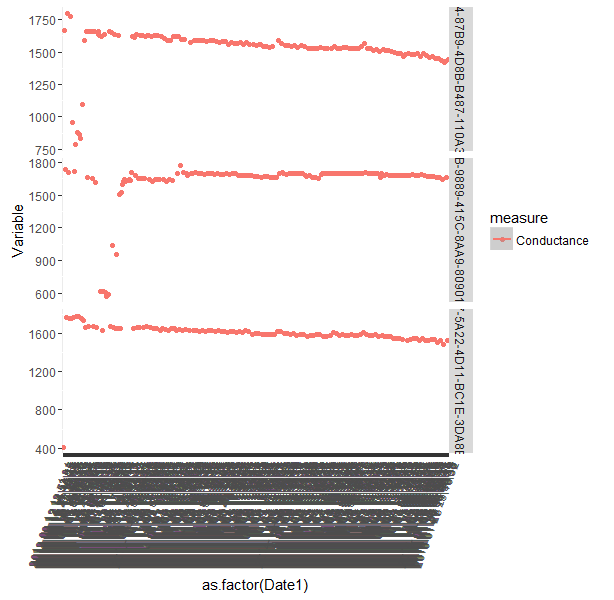


Figure-16

* Batteries 1 and 2 have greater than 10 % points where DOD values are less than 0.8 .
* Battery 6: has around 5 % points where DOD is less than 0.8
* Batteries 3,4,5 have less than 5% points where DOD is less than 0.8
* Battery 7 has no point less than 0.8

### Bin 3: Conductance Drop > 0.5 < 0.65 (4 cases)

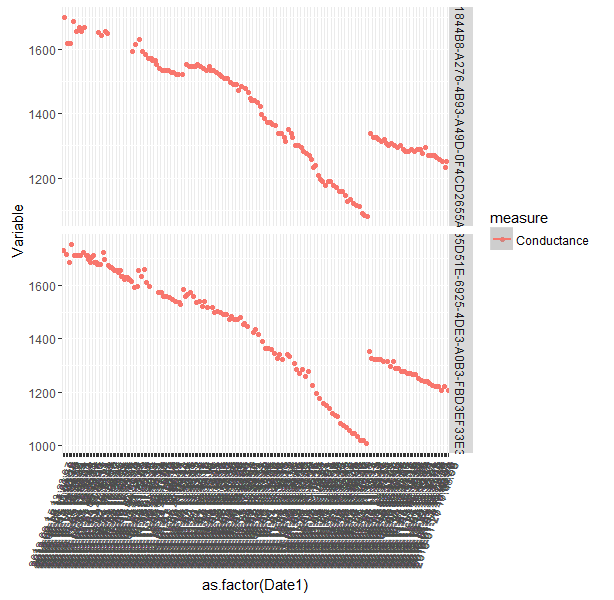
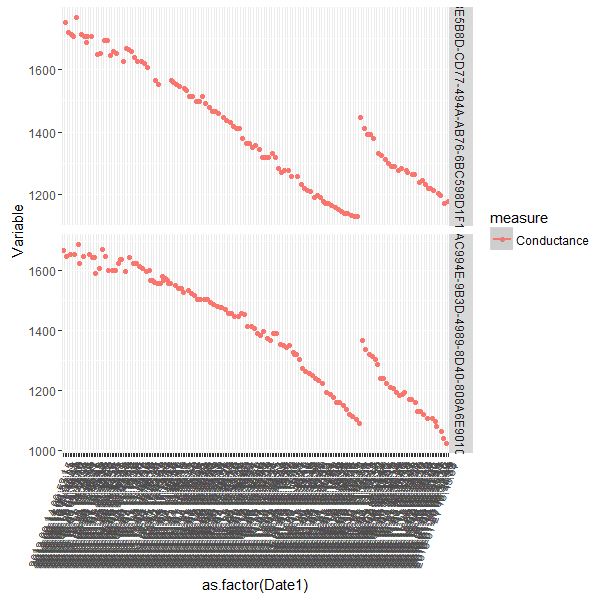


Figure-17

* All batteries have similar trends of dropping and retracting
* Batteries 1& 2 have more than 10% cases where DOD is less than 0.8
* Battery 4 has around 8% and 3 has around 6 %

### Bin 4: Conductance Drop > 0.65 < 0.8 (100 cases,30 sampled)

This bin was divided into two groups based on the percentage of values below the threshold limit of 0.8. The plots of the first group are as follows.

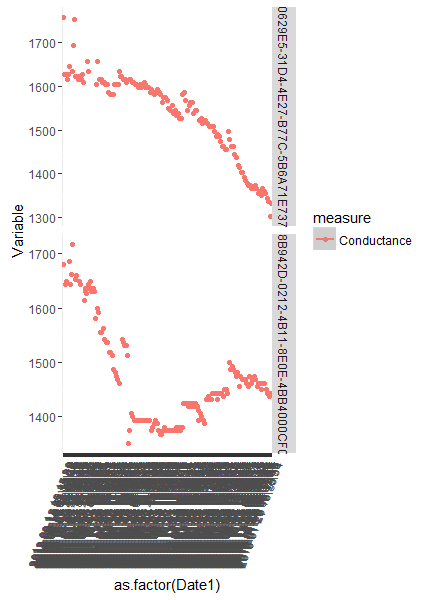
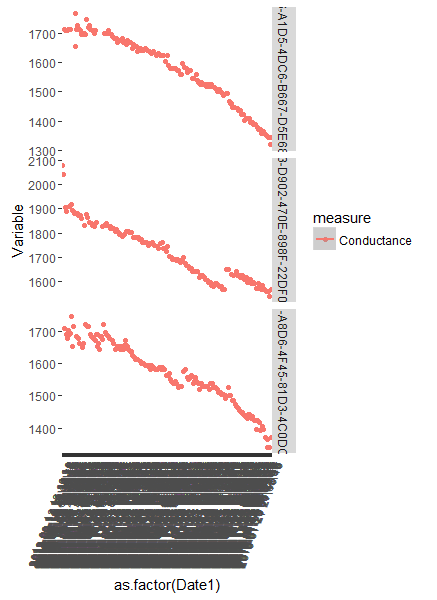
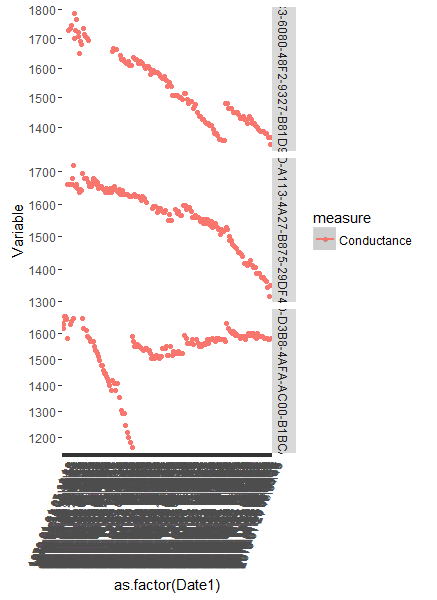


Figure-18

* There are 8 Batteries in this group
* 4 of them shows a trend of drop and retracting values of conductance. Others have very sharp drop in conductance
* All these batteries have more than 10% of their values below DOD 0.8 ( Ranges from 10% to 30%)

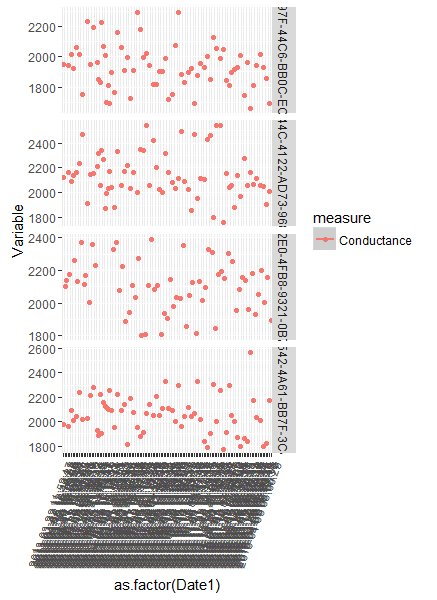
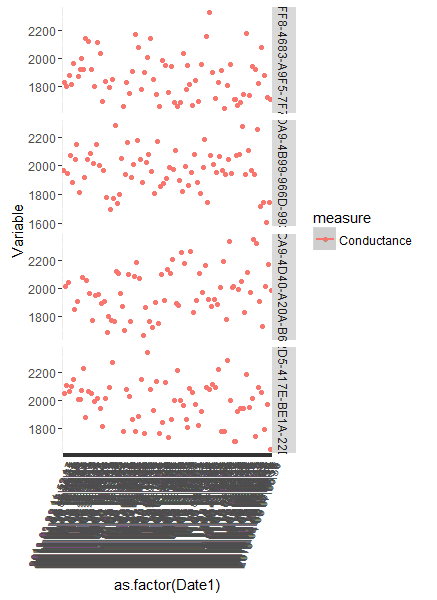
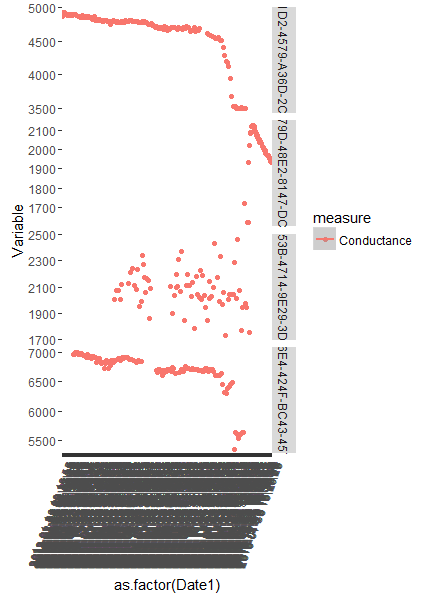
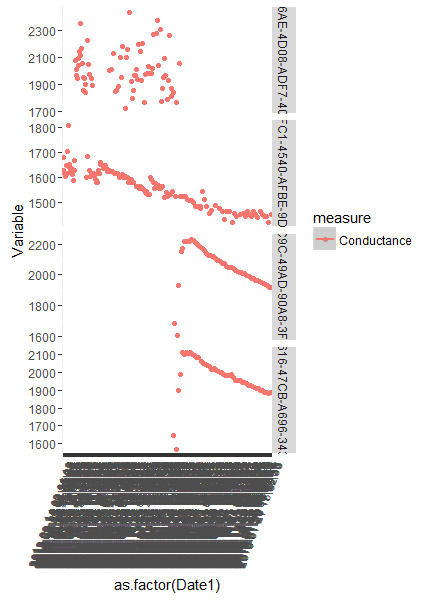


Figure-19

* Group II have 22 batteries in total
* 13 of them showed very high variability in its conductance profile.
* Almost 19 of the total cases have their DOD values above 0.8 threshold value

### Bin 5: Conductance Drop > 0.80 < 0.9 (492 cases,30 sampled)

This bin also divided into two groups. Group 1 profile is shown as below

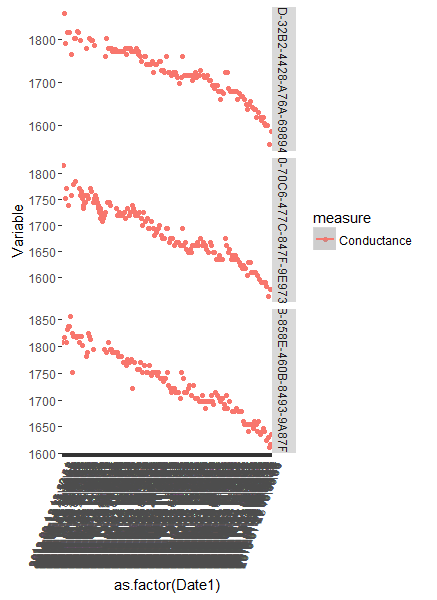
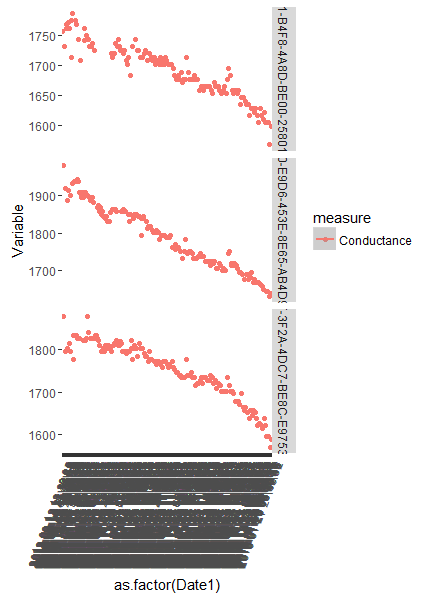
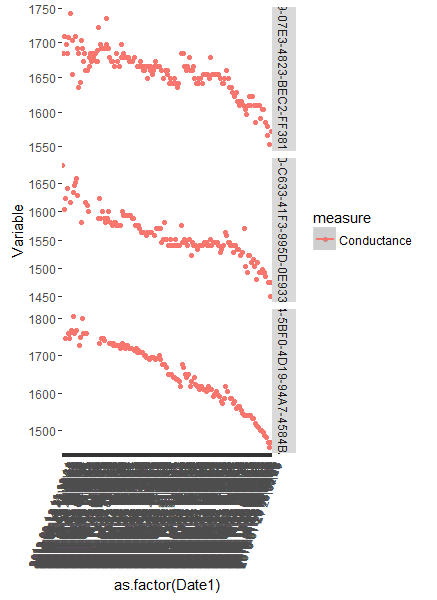


Figure-20

* Total of 9 batteries in group 1
* All of them showing gradual drops. The drops are more muted than the previous bins
* All these batteries have more than 10% of their values below DOD 0.8 (Ranges from 12% to 27%)

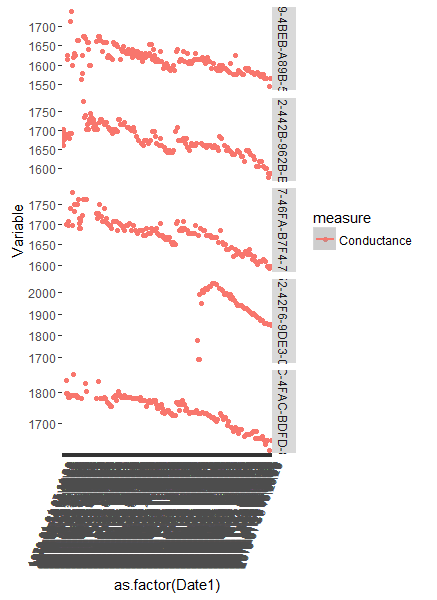
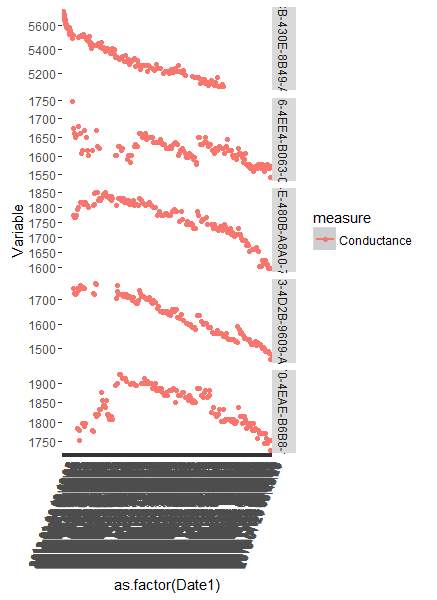
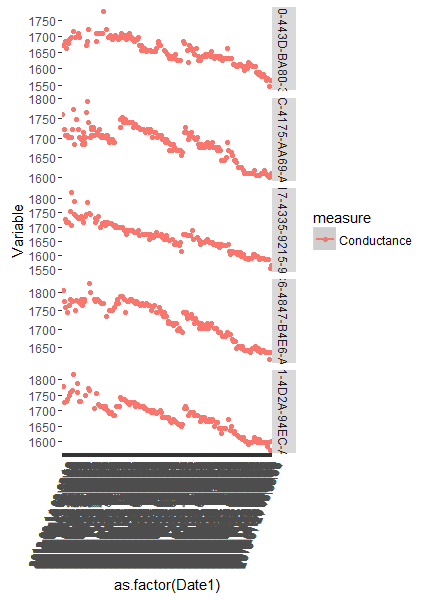
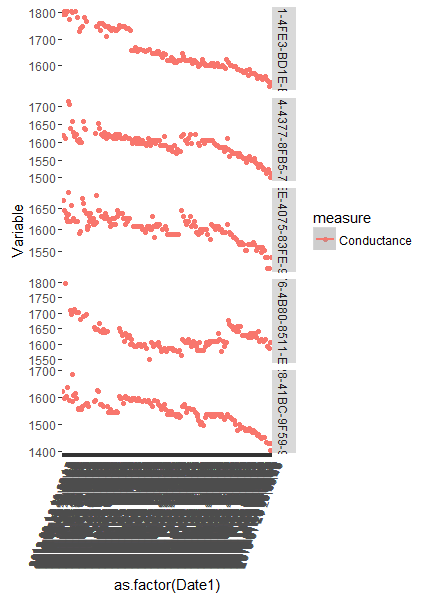


Figure-21

* Total of around 21 batteries in group II.
* Most of the batteries show gradual or flatter conductance drops.
* None of these batteries have more than 10% of DOD values below 0.8 threshold value.

### Bin 6: Conductance Drop > 0.9 (394 cases,30 sampled)

This bin was also divided into two groups based on the percentage of values below DOD 0.8. The group 1 batteries are as shown below.

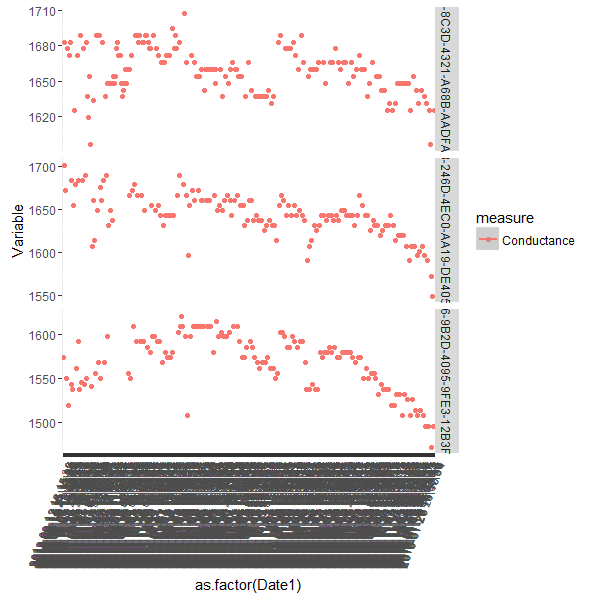


Figure-22

* Total of 3 batteries in group 1
* All batteries show flat conductance profiles.
* All these batteries have more than 10% of their values below DOD 0.8 (Ranges from 41% to 15%)

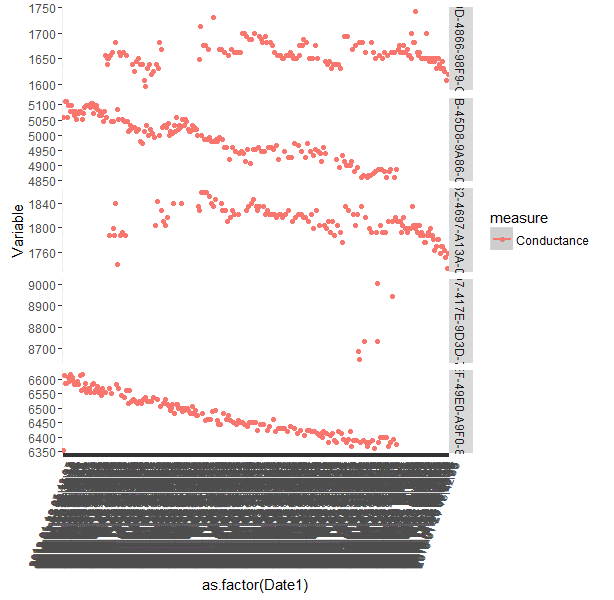
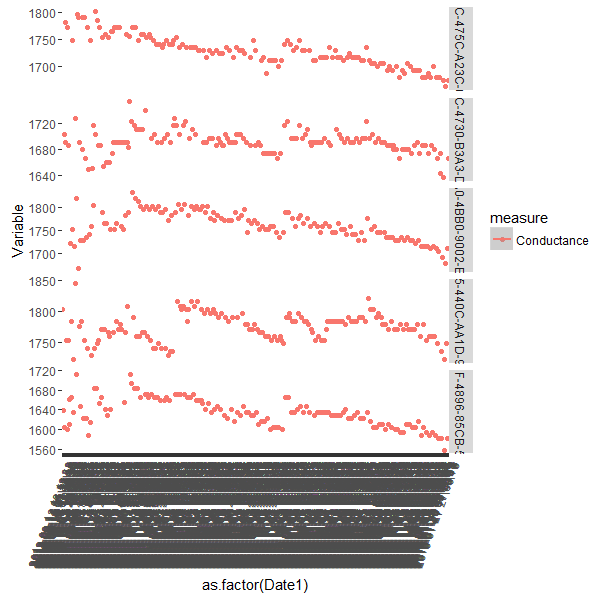
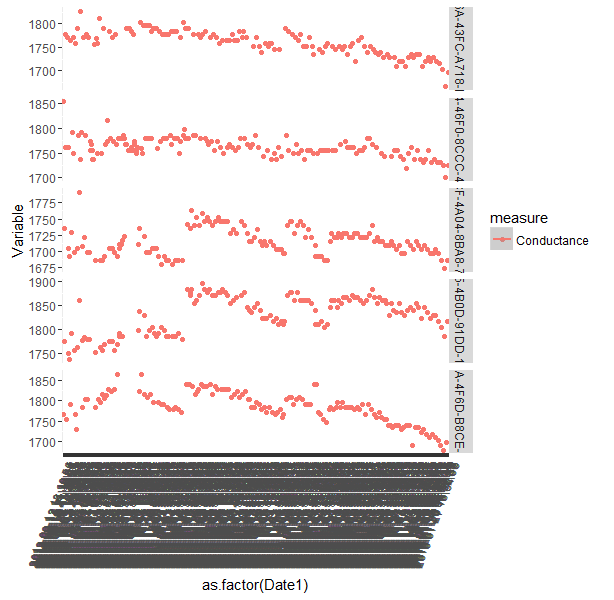


Figure-23

* Total of 27 batteries in this group
* All the batteries predominantly show flat conductance drops
* 16 of these batteries have 100% of their DOD values above 0.8. 3 of them have less than 3% and 2 less than 8% of values less than .8 respectively.

### Intuitions and inferences from the analysis.

Based on the analysis done for the second sample of 1000 batteries, below are the major inferences that can be derived.

* A general trend which was observed was, the DOD (from voltage perspective) for most of the batteries were predominantly between the ranges: 0.85 – 0.80.
* Association between percentage of values having DOD values less than 0.8 and percentage drop in conductance was observed.
* Batteries which had sharper drop in conductance had relatively larger percentage of values below the 0.8 DOD levels.
* A good filter for identifying potential cases of failing batteries could be the following
  + - Drop in conductance > 80% with respect to the highest conductance level
    - 10% or more values with larger than 0.8 DOD
* On the above filter, another level of filter for percentage of conductance points below 80% could be added
  + - 3% or more values with larger than 80% of conductance level

Based on the intuitions derived from the observations, a composite score with all the factors incorporated in it were developed. The composite score was included as a new feature for both samples of 1000 batteries and then it was further ranked. The snap shot of the analysis is as described below.

**Sample 1: 1000 Batteries**

* 21 cases were filtered with the above mentioned filter
* Most cases in top 10 have falling and retracting trend
* Rest of the cases have sharp conductance drops
* All of them have substantial proportion of values above the 80% DOD range (Average 23%)

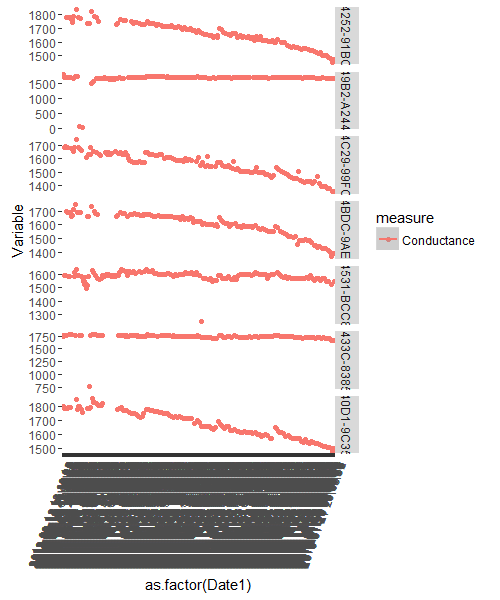
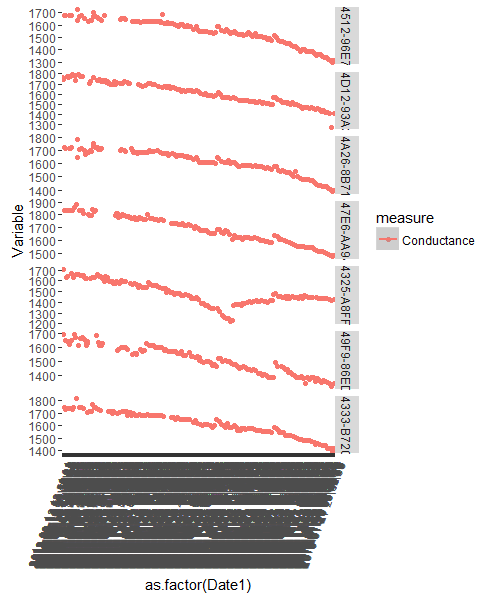
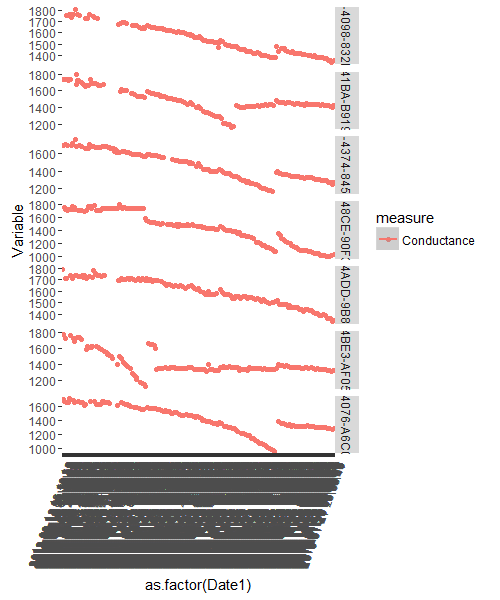


Figure-24

**Sample 2: 1000 Batteries**

* 29 cases were filtered with the above filter
* Most cases in top 14 have falling and retracting trend
* Rest of the cases have sharp conductance drops

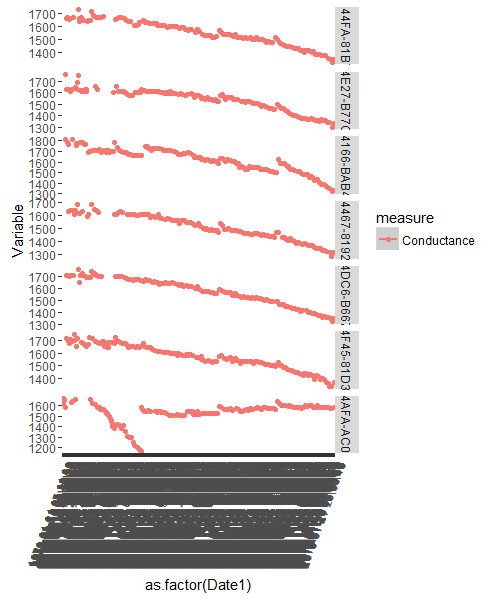
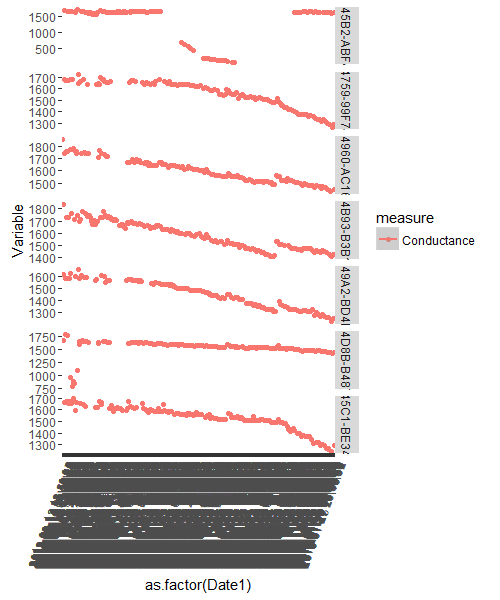
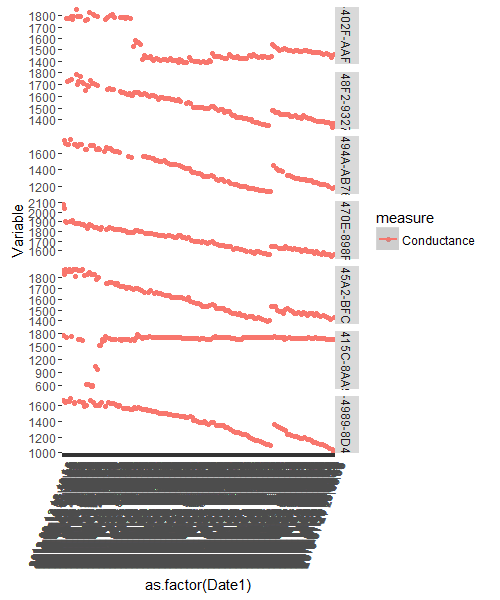


Figure-25

# Intuitions and inferences

Based on the complete set of analysis carried out so far, we can summarize the findings as below.