**HORIZON GRAPH**

Horizon graph was part of the anomaly visualization because it is space-efficient time series visualization technique to estimate percentage of changes between several sensors over time.

**PURPOSE OF PLOT**

* To explore the percentage of changes for sensor readings over time.
* To highlight potentially abnormal sensor readings with highest percentage of changes.
* To highlight potentially abnormal sensor readings with persistent positive/negative trend.

**HOW TO CREATE THE PLOT**

**R Packages**

|  |  |
| --- | --- |
| lubridate | For functions to work with date-times especially extraction of components of months, days, hours, etc. |
| latticeExtra | For horizonplot() function to plot horizon graph |
| DataCombine | For change() function to calculate percentage change from a specified lag, including within groups |
| xts | For as.xts() function to convert xts objects into environments for horizon graph. |
| DT | For DataTables function to display R data objects (matrices or data frames) as tables. |
| data.table | For transform function to perform data.table transformation. |

**Methodology**

Create R-Shiny Dashboard using Shiny package

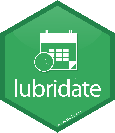
Select only required columns for horizon graph to reduce lag

Calculate percentage of changes within the group of readings by each sensor using change() function in DataCombine package.

Take sensor ID as keys and percentage of changes as values, and spreads into multiple columns.

Create horizon graph using horizonplot() function in latticeExtra package.

Create data tables using DT and data.table package.

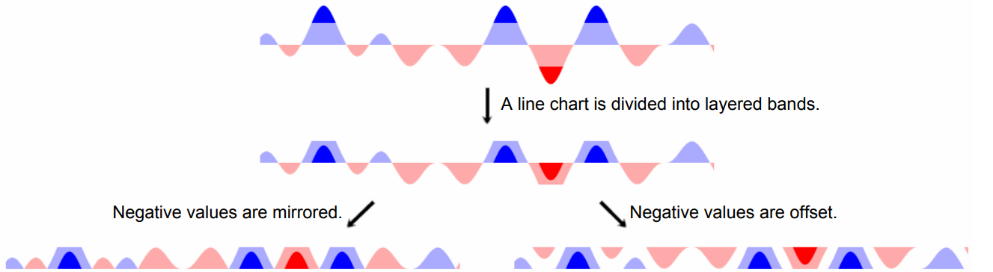
Data Processing

Parallel

Control

**HOW TO READ THE PLOT**

* Sensors identified by its their respective ID are arranged by row, and timestamp from earliest to latest are arranged horizontally from left to right respectively.
* Based on the mirrored chart as described in the paper [1], the negative values are flipped such that both positive and negative values are on the same axis. The blue represents positive percentage of changes between current and last sensor readings, whereas the red represents negative percentage of changes. The magnitude of these values is distinguished by the color hue. Darker shades represent larger values, whereas lighter shades represent smaller values.



(source: <http://vis.stanford.edu/files/2009-TimeSeries-CHI.pdf>)

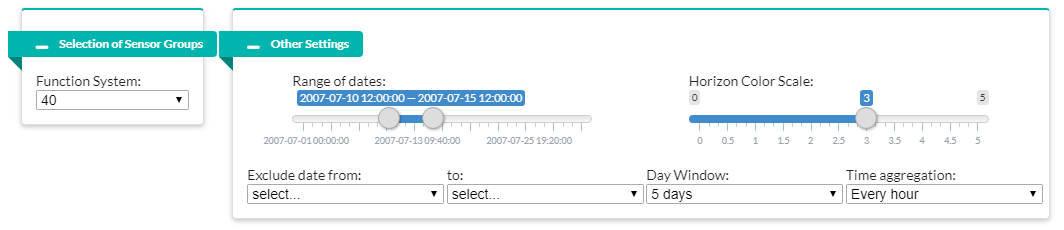
Reference:

[1] Heer, Jeffrey & Kong, Nicholas & Agrawala, Maneesh. (2009). Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. Conference on Human Factors in Computing Systems - Proceedings. 1303-1312. 10.1145/1518701.1518897.

**HOW TO USE THE PLOT**

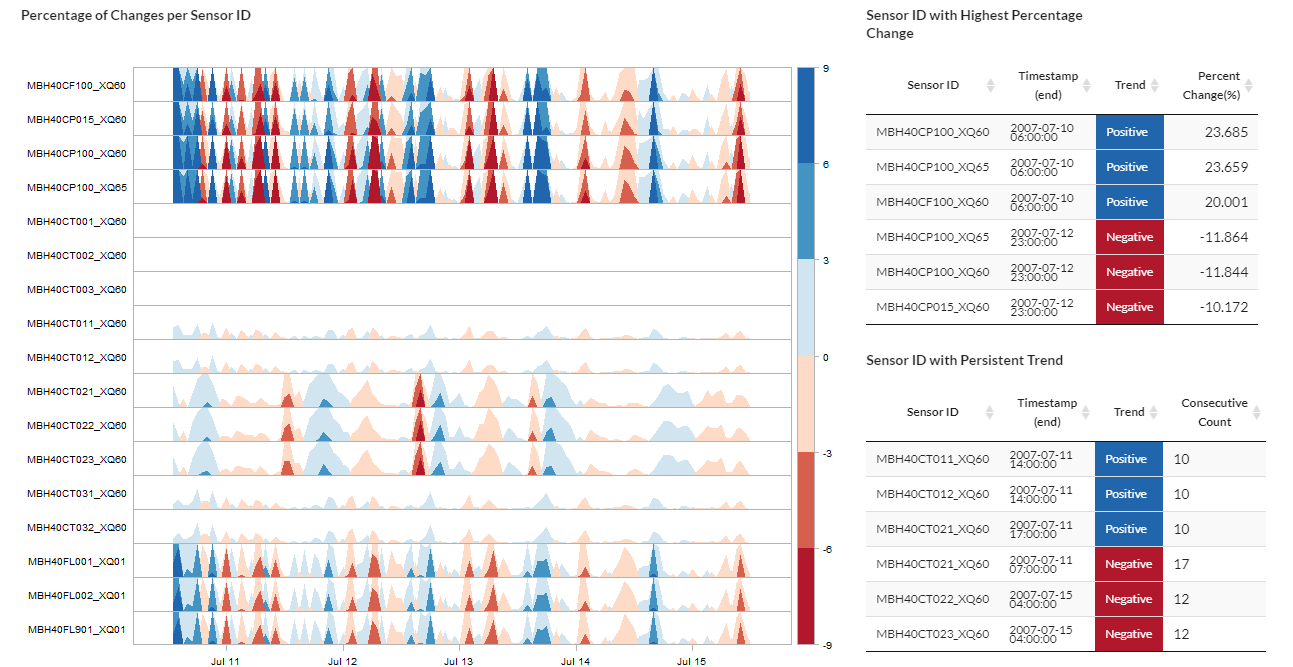
**Allowable selections:**

* Function system
* Range of dates (minimum, maximum)
* Exclude dates
* Day window
* Time aggregation
* Horizon color scale



**EDA:**

1. The range of dates slider can be adjusted accordingly to achieve the desired observation period. Upon selecting full range, percentage of changes can be observed holistically in attempt to pick up sensors with prominent positive or negative value or trend within the same function system.
2. The day window input specifies the number of days for the observation period in order to zoom in on a portion of the graph. The range of dates slider will be reactively updated and can be drag left or right to observe anomalies with greater details for across the full range. (Horizon color scale slider is to make adjustment to the scale of each color segment to enhance the readability.)
3. Within the observation period, top few sensors with highest percentage of changes and with persistent trends, from both positive and negative sections, are displayed together with their timestamp on data tables. They are potentially abnormal sensor readings.
4. These abnormalities suggest that potentially abnormal sensor readings and timestamp that should go for more in-depth investigation.



**Sensors spotted with unusual peak**

**Sensors spotted with persistent trend**

**LESSON LEARNT**

There are few approaches to create horizon graph in R shiny. Here is the overview of the approaches, in which our team tried, along with their respective pros and cons.

1. **Horizon graph via ggTimeSeries package**

(source: <https://cran.r-project.org/web/packages/ggTimeSeries/vignettes/ggTimeSeries.html>)

This horizon graph package is based on ggplot2.

|  |  |
| --- | --- |
| **Cons** | **Cons** |
| **Convenient coding with layering system** | **Bad readability at both big and small scales** |

1. **Horizon graph via base graphics**

(source: <https://www.r-bloggers.com/horizon-plots-in-base-graphics/>)

The original code for this horizon plot was taken from R-bloggers.com by klr, and our team further modified it so that each sensor’s horizon graph can be stacked vertically.

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| **Good readability at small scale** | **Bad readability at big scale** |

1. **Horizon graph via cubism.js**

(source: <https://github.com/kbroman/horizon>)

This R/horizon package is a htmlwidgets-based package using cubism.js but modified by Karl W Broman to handle static data set instead of real-time. The difference with the usual horizon graph is that negative values are offset instead of mirrored. The positive values in blue are placed at bottom axis, while negative values in red are at top axis.

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| **Good interactive overlay** | **Unable to zoom in to a portion of graph as each pixel encodes a distinct point in time** |

1. **Horizon chart via latticeExtra package**

(source: <https://www.rdocumentation.org/packages/latticeExtra/versions/0.6-28/topics/horizonplot>)

This horizon graph package is the final one used by our team. The horizonplot() function is taken from latticeExtra package which was built on the infrastructure provided by the lattice package. While there is a slight bug in the package that causes some overlap between positive and negative areas, it provides good readability at both large and small scales to estimate the percentage of changes between sensors.

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| **Good readability at both big and small scales** | **Positive and negative areas overlap** |