



# MULTINAV

Exploratory Data Analysis  
For Multivariate Data

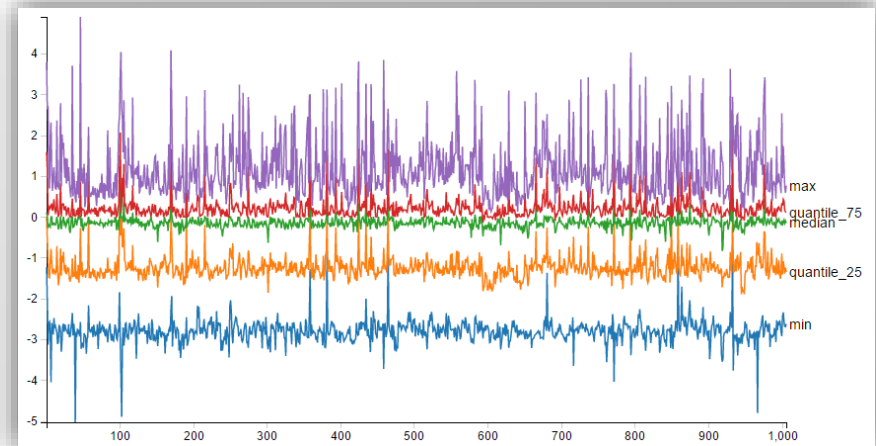
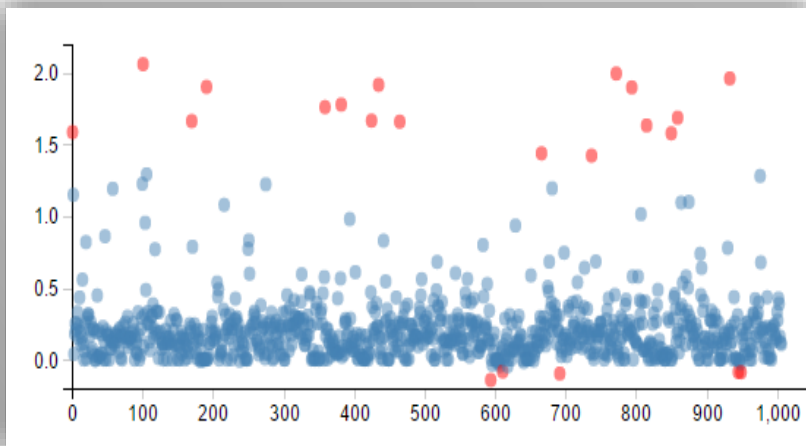
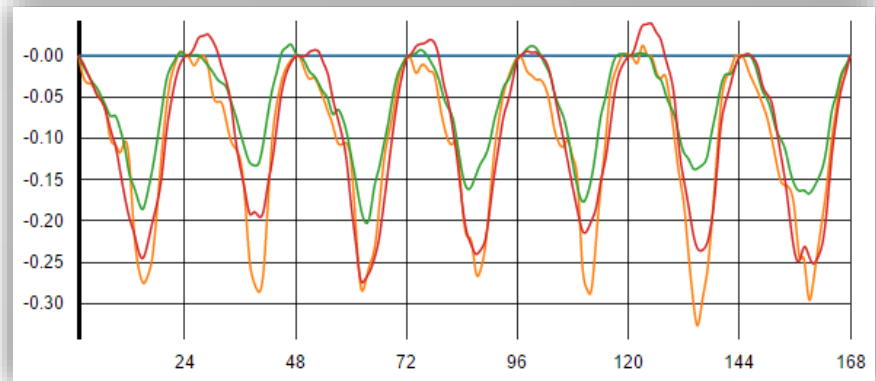
May 2017  
Efrat Vilenski  
Jonathan Rosenblatt

# Exploratory Data Analysis - EDA

A term coined by Tukey 1977

Set of techniques to display data in a way, that interesting features will become apparent.

EDA is a basic step before modeling or hypothesis testing.

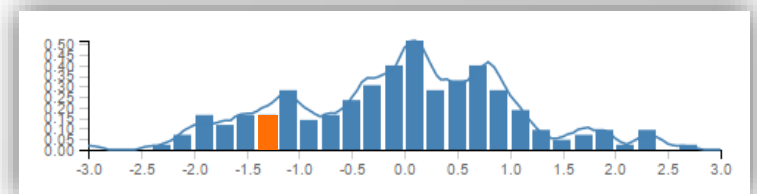
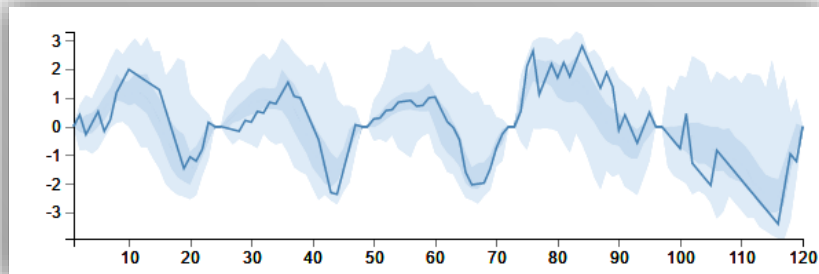


# Introducing MultiNav

## Goal:

Interactive tool for improving (continuous) multivariate EDA process

- **Gain more insights**
- **In less time**



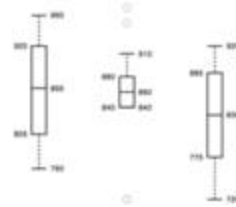
# MultiNav Capabilities

- **Vision:** visual analytics system for multivariate data analysis with various capabilities:
  - **Data preprocessing** with statistical algorithms:
    - Process monitoring.
    - Outlier detection.
    - Clustering.
    - Dimension reduction
    - Robust statistics
    - Social network analysis
  - **Interactive charts**
    - Linked views
    - Brushing
    - Zoom in /out
    - Smart filters

# D3

## Data Driven Documents

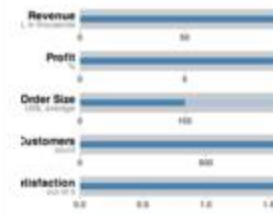
Box Plots



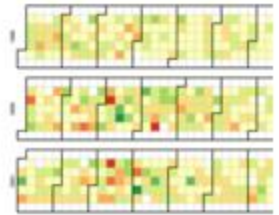
Bubble Chart



Bullet Charts



Calendar View



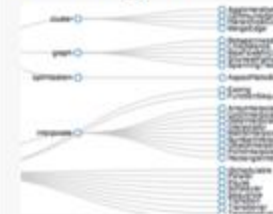
Non-contiguous  
Cartogram



Chord Diagram



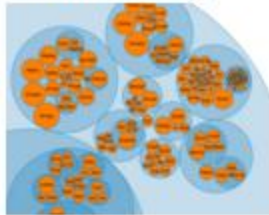
Dendrogram



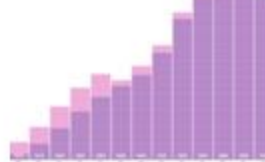
Force-Directed  
Graph



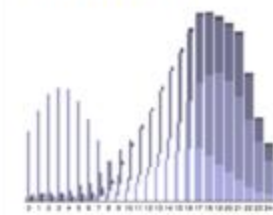
Circle Packing



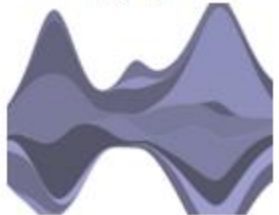
Population  
Pyramid  
2000



Stacked Bars



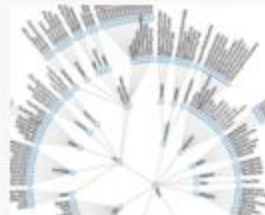
Streamgraph



Sunburst



Node-Link Tree



Treemap



Voronoi Diagram



# MultiNav Development Methodology

Not presuming to be complete EDA suitable to all data scenarios.

Building **general tool** based on **real use cases**

- Detection of outlier sensors in order to make irrigation decisions based only on quality data.
- Predictive maintenance: identification of engine malfunction based on high frequency vibration monitoring.

## Tested on

- Continuous variables
- Variables on same scale
- $n > p$
- $p > 50$  and  $p < 2500$
- Linear relationships

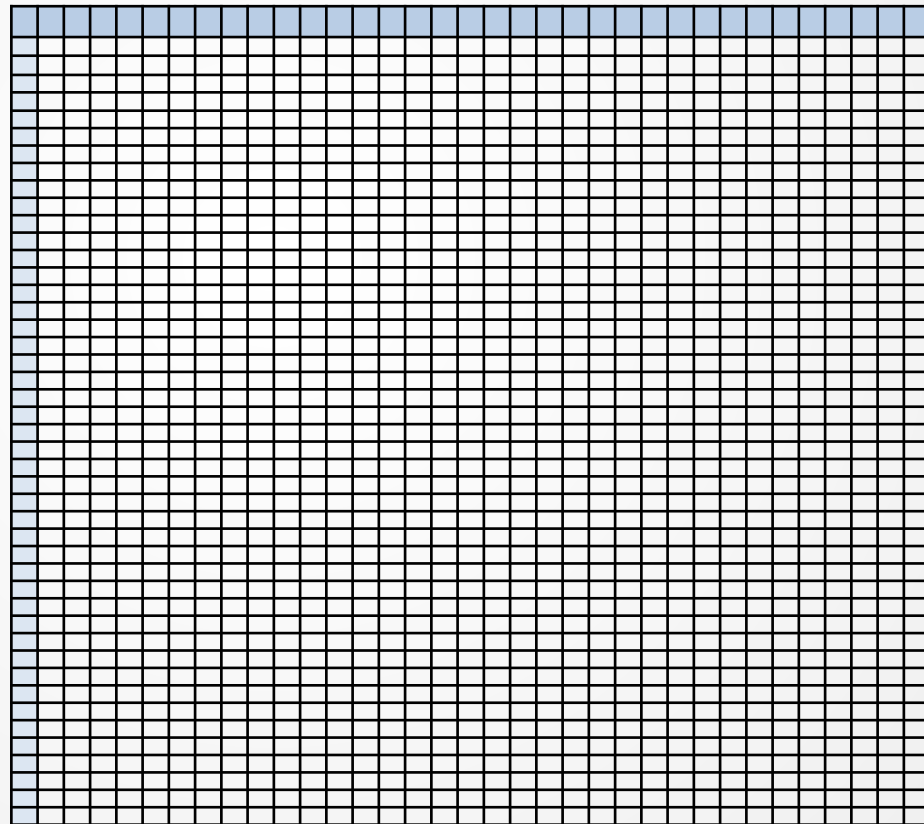
# MultiNav Concept

## Input -

Continuous  
multivariate  
data set

n

p



## Out of scope

- Handling missing data
- Transformations



# MultiNav Concept

# n

[illegible]

p

[illegible]

## Optional

## Input -

# Variables and observations features



# MultiNav Concept (cont)

## Preprocessing

R package

- Data processing functions

## Output

Web Dashboard

- Interactive Dashboard

R package

- Charts (via htmlwidgets)

# Data processing Functions

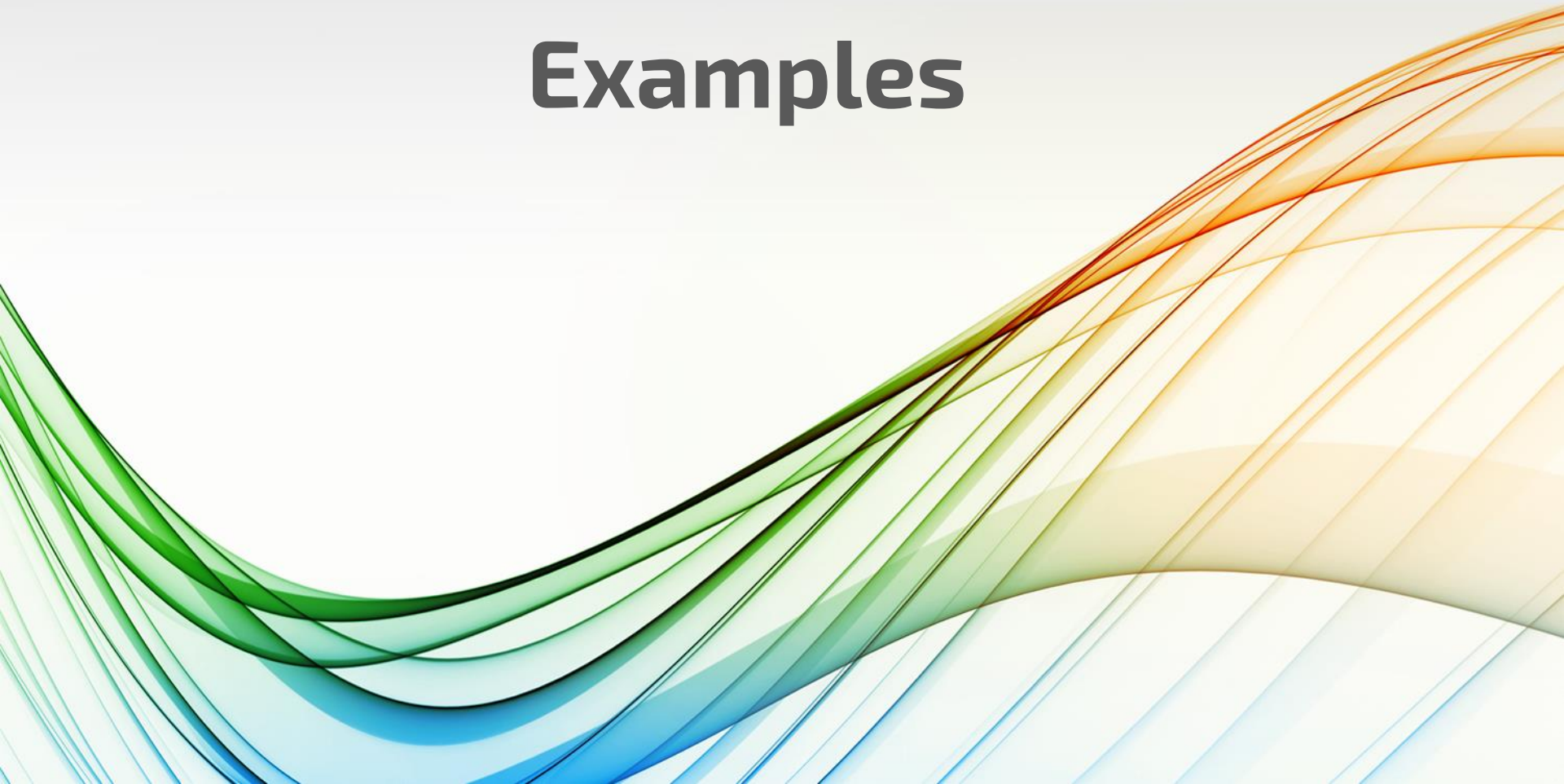
Back-end processing to ensure interactive front end with good response time.

Data processing output stored in specific format, allowing also flexibility for custom processing.

**Processing  
examples:**

- Data summaries
- Variables scores
- Observations scores

# Examples



# Dashboard – univariate analysis Example

EDA100

Multivariate EDA

[About](#)

## Data Set

Eng3 - CP4\_AXI ▾

Correlation

Mass Univariate

Quantiles

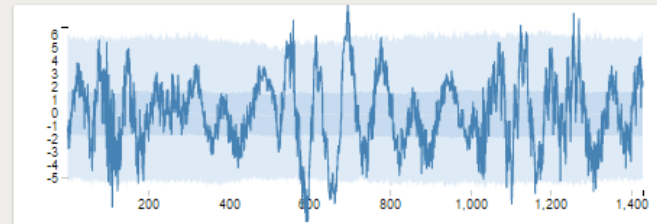
Outliers Matrix

Id:  
967

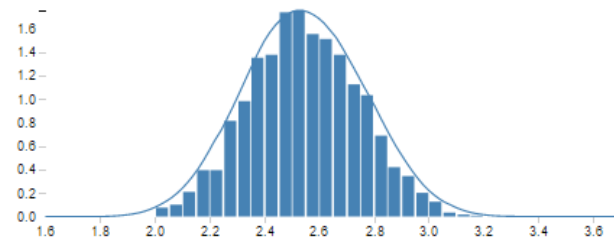
Q 25:

Median:

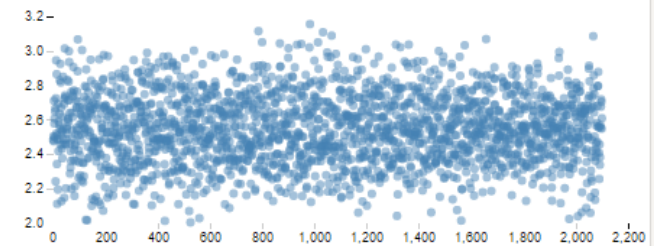
Q 75:



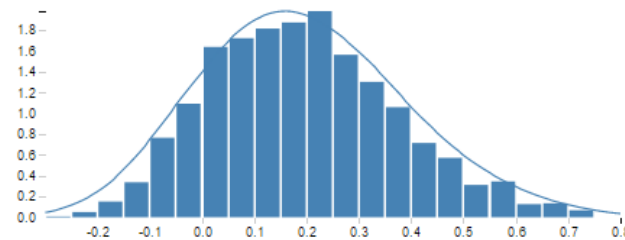
mad



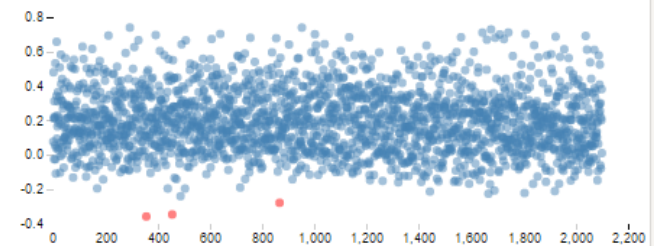
mad



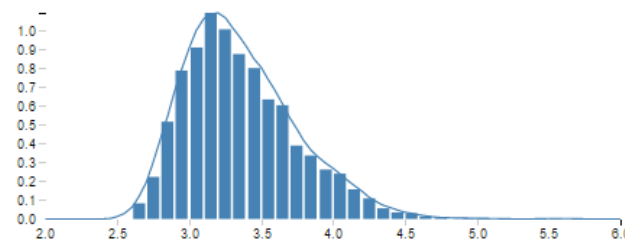
skness



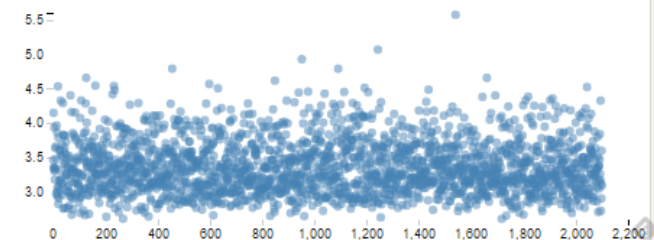
skness



kurtos



kurtos



# Dashboard – multivariate scores example

Demo

PhySod – Outlier Scores View

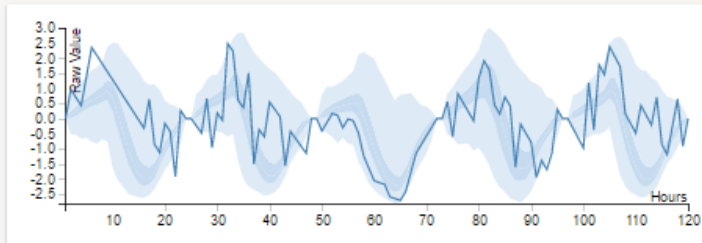
About

High Level View

Outliers View

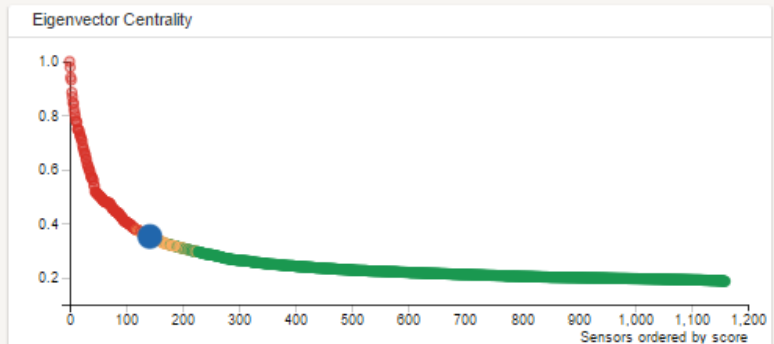
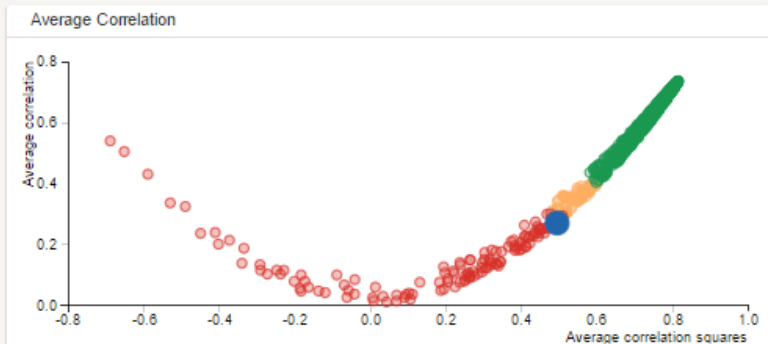
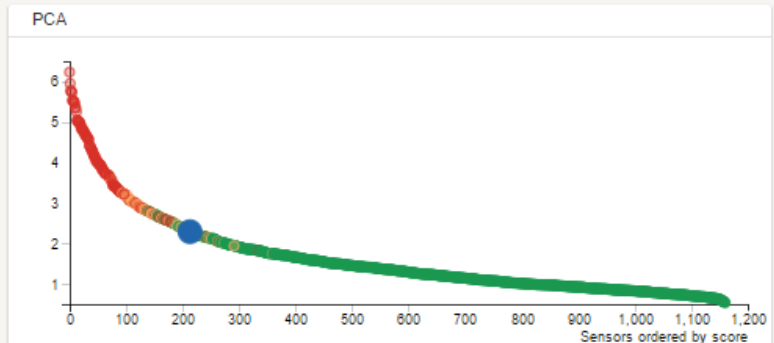
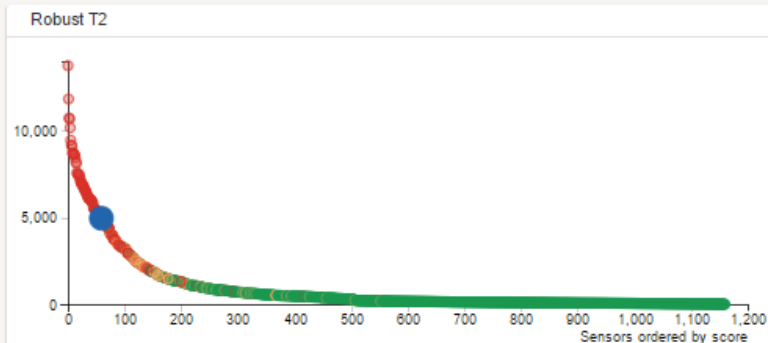
Class What If

Class Performance



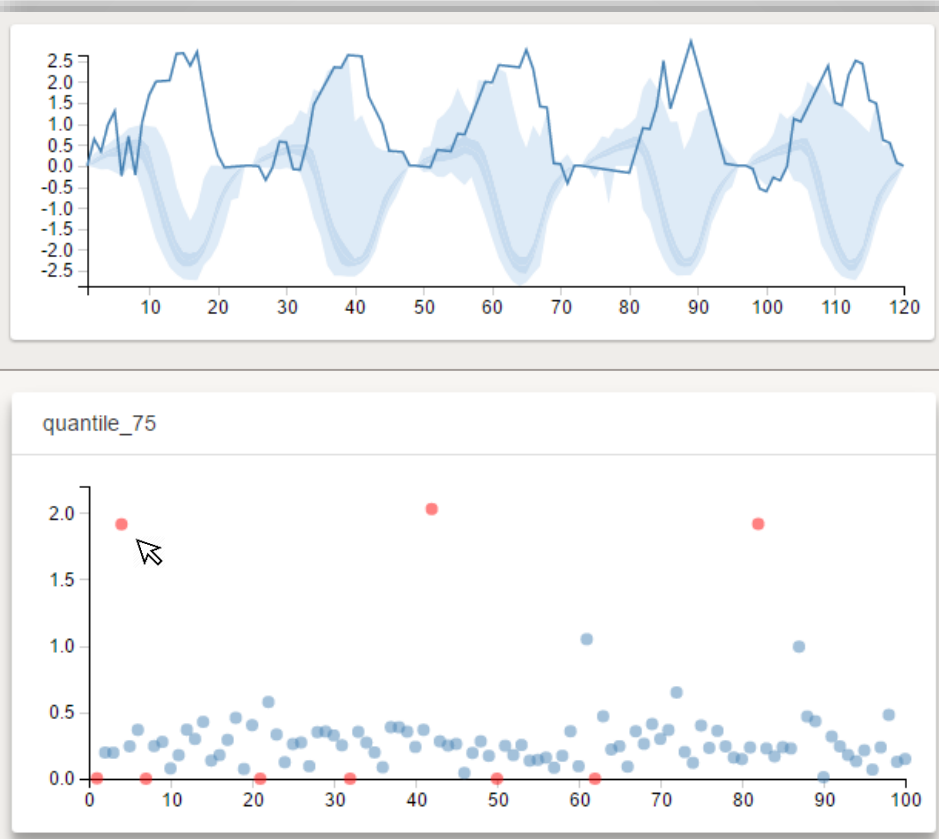
**Sensor ID:**  
32408

**Scores:**  
Robust T2: 4962.02  
PCA: 2.3  
Average Correlation: 0.5  
Eigenvector Centrality: 0.35





## Scatter plot with linked functional box plot





## Distance matrix with linked functional box plot





# Test and Give feedback

MultiNav R package (initial version)  
available soon for download.

Subscribe for MultiNav updates:  
<http://efratvil.github.io/MultiNav/Subscribe.html>