Time series

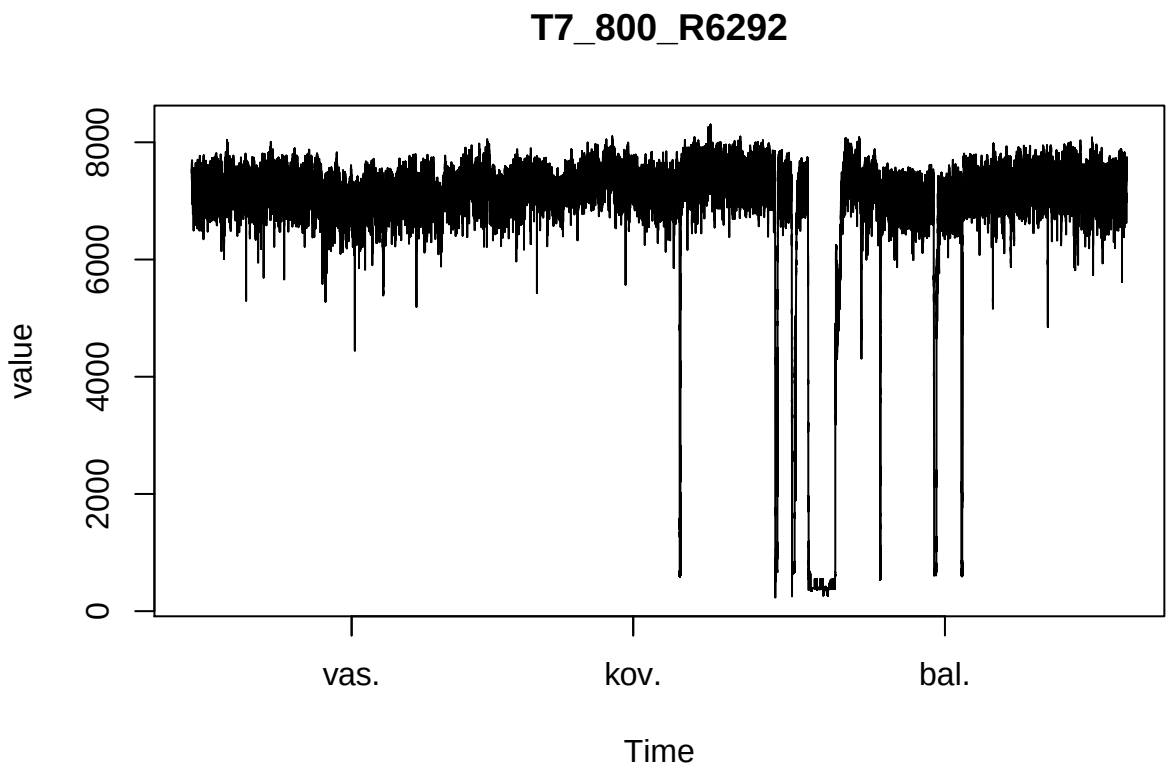
Robert Tarasevic

2020-04-06

Contents

```r
library("tidyverse")
library("caret")
library("dplyr")
library("rhdf5")
library("zoo")
library("knitr")
library("kableExtra")
library("mgcv")
```

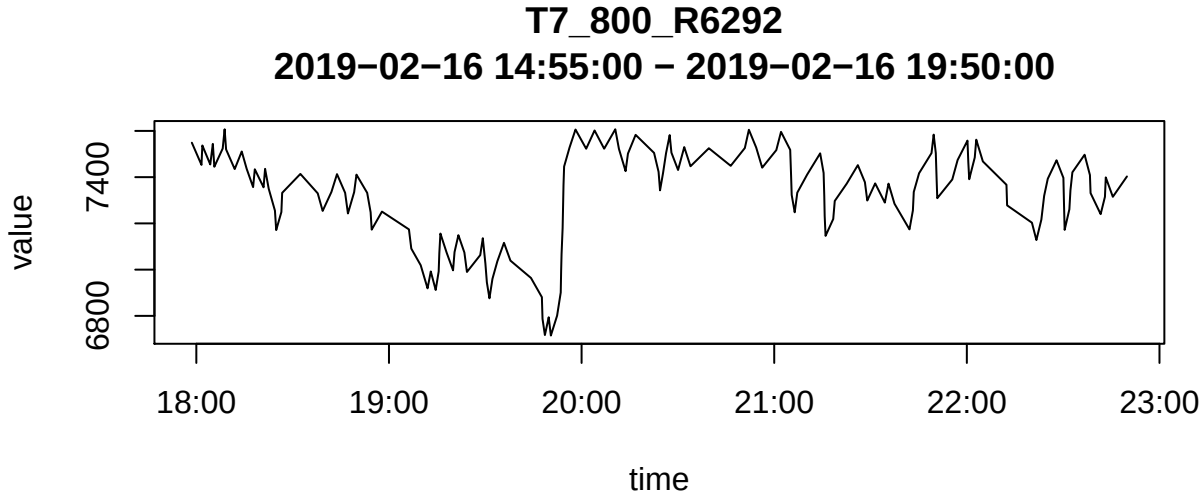
Load data




```
## [1] "Only one graph showed because same sensor chosen"
```

Chosen time period for signals

```
## First signal subset:  2019-02-16 14:55:00 - 2019-02-16 19:50:00
```




```
## Second signal subset:  2019-02-16 19:50:00 - 2019-02-17 01:02:00
```

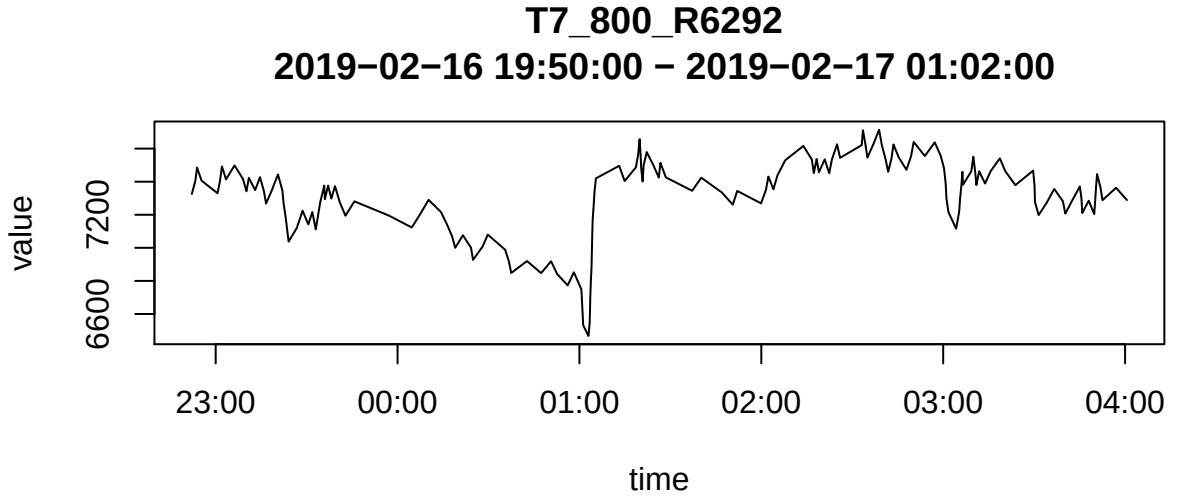


Table 1: Statistics table

	length	t_mode	v_mode	sd	mean	kurt	skew	rms	entropy	cf	sf	cross_r
sub 1	160	28	7332	206.86	7310.76	0.04	-0.84	7310.76	5.07	1.04	1.00	26
sub 2	170	12	7404	250.48	7300.81	1.07	-1.12	7300.81	5.14	1.06	1.00	25
raw 1	115711	4	384	2055.88	6472.54	4.08	-2.40	6472.54	11.58	1.22	1.05	698
raw 2	115711	4	384	2055.88	6472.54	4.08	-2.40	6472.54	11.58	1.22	1.05	698

Statistics table

Some statistics explained:

Signal length - *Data points number in signal*
Time mode - *Mode of time change frequency*
Value mode - *Most frequent value*
Sd - *Standard deviation*
Mean - *Average value*
n5, n25, n75, n95 - *Quantile values*
Kurtosis - *Kurtosis quantifies the peak value of the PDF (positive kurtosis tells that there is lot of data in tails, negative - little data in tails*
Skewness - *Skewness quantifies the asymmetry behavior of vibration signal through its PDF*
RMS - *Root mean square value changes faster then mean*
Entropy - *Amount of uncertainty in an entire probability distribution*
Crest factor (cf) - *Ratio of the instantaneous peak amplitude of a waveform, to its root mean square RMS value*
Shape factor (sf) - *Shape factor is a value that is affected by an object's shape but is independent of its dimensions*
Mean crossing - *Mean value crossing count*

Correlation

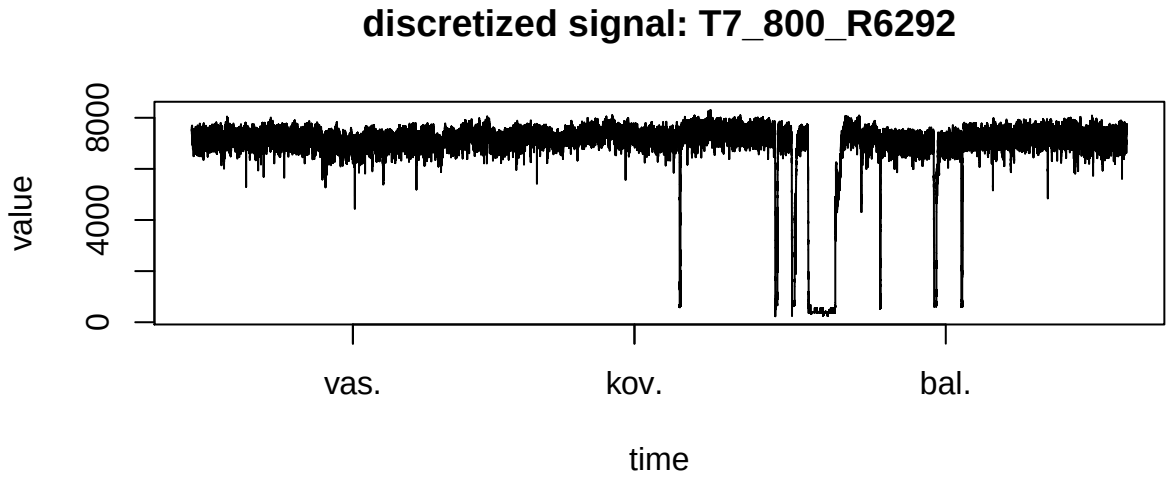
Correlation between sub-sets

```
## [1] 0.6683277
```

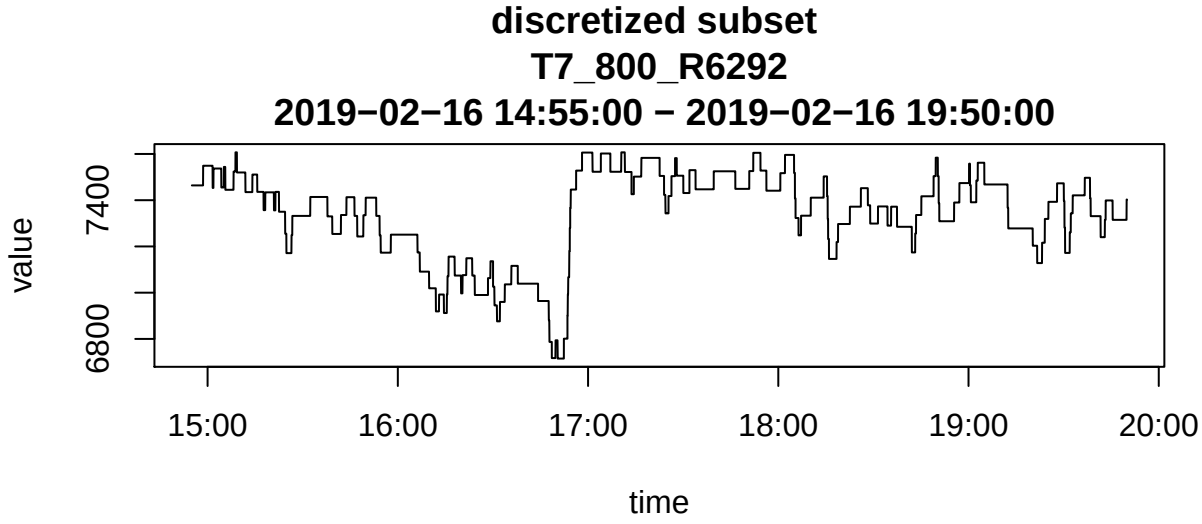
Correlation between signals

```
## [1] 1
```

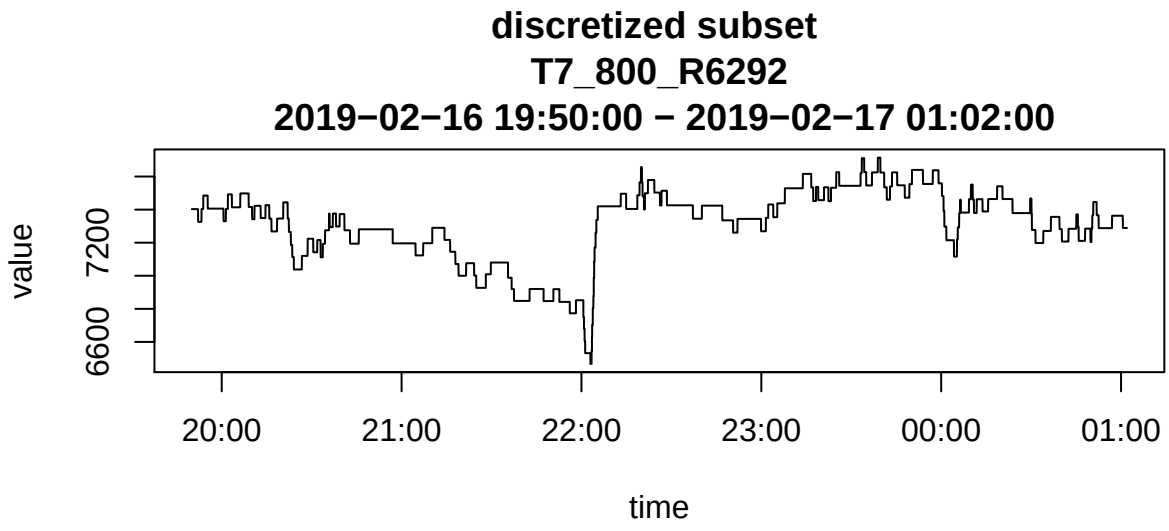
Data discretization




```
## [1] "Only one graph showed because same sensor chosen"
## First signal subset:  2019-02-16 14:55:00 - 2019-02-16 19:50:00
```

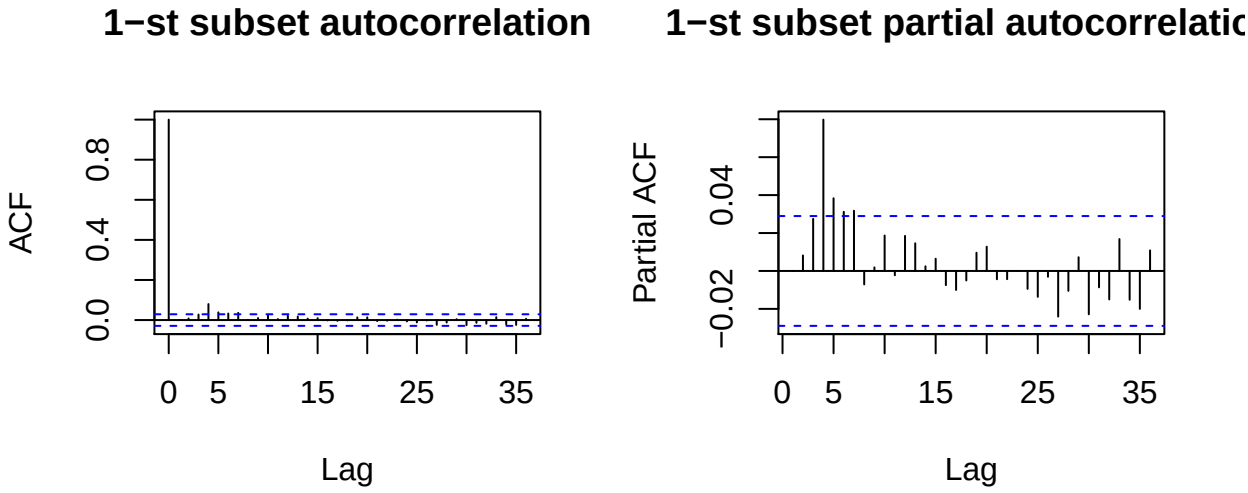


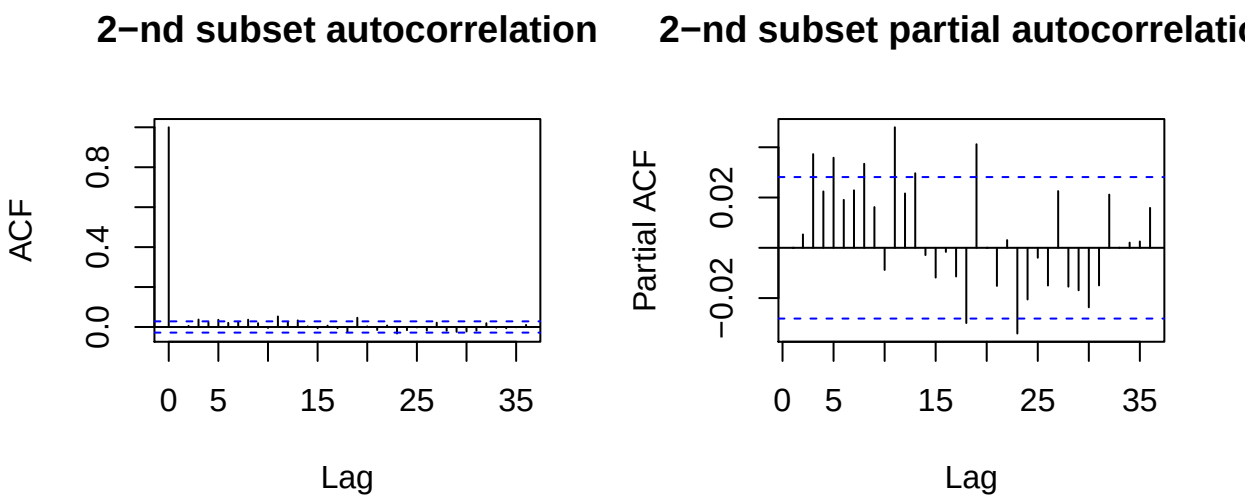

```
## Second signal subset:  2019-02-16 19:50:00 - 2019-02-17 01:02:00
```


discretized subset
T7_800_R6292
2019−02−16 19:50:00 − 2019−02−17 01:02:00
value
6600
7200
20:00
21:00
22:00
23:00
00:00
01:00
time

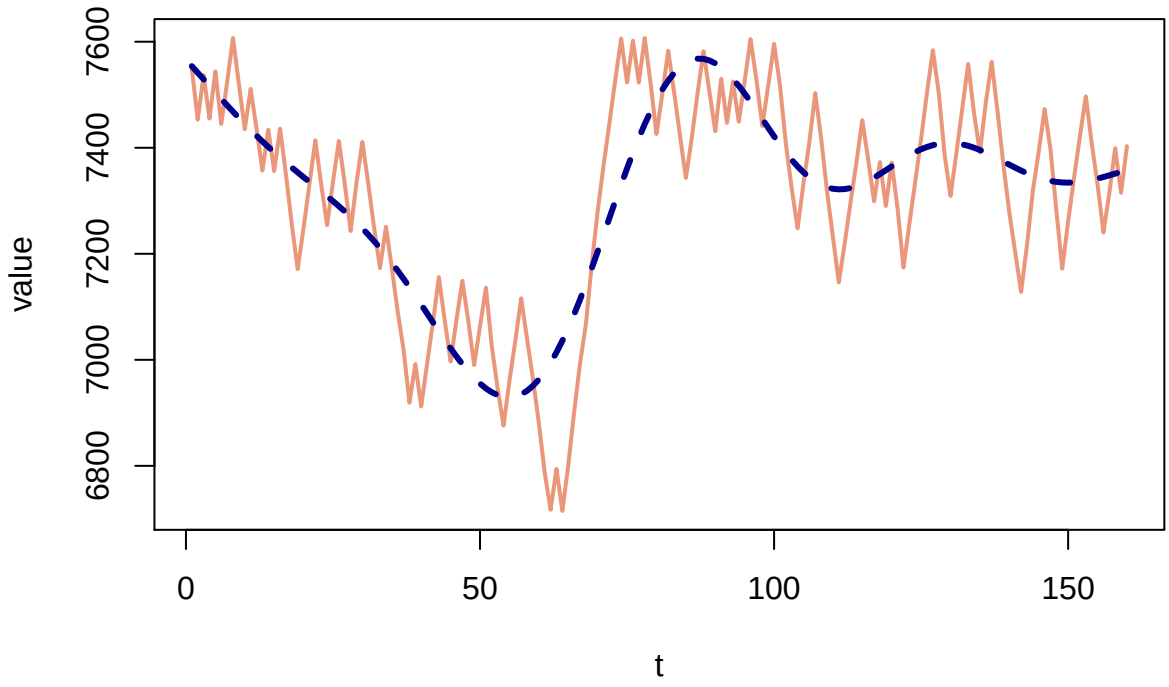
Auto-correlation

Auto correlation is the correlation of a time series with its own past and future values
if data is strongly seasonal - peaks will apear with seasonality period
Partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags





Generalized Additive Model



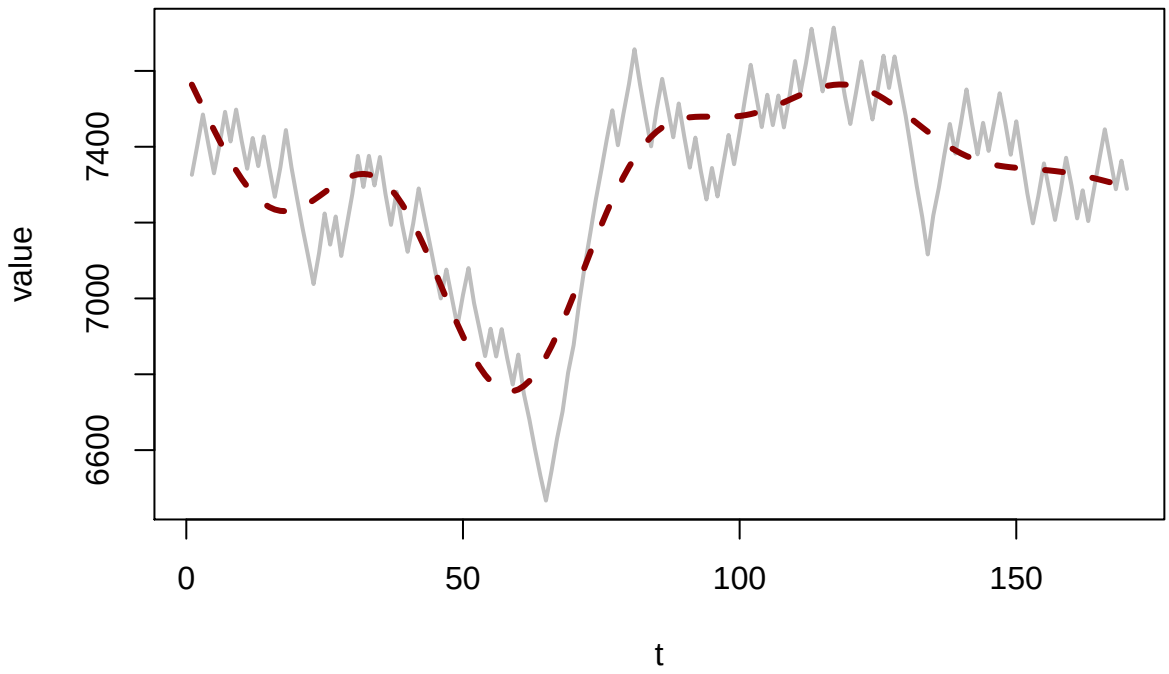
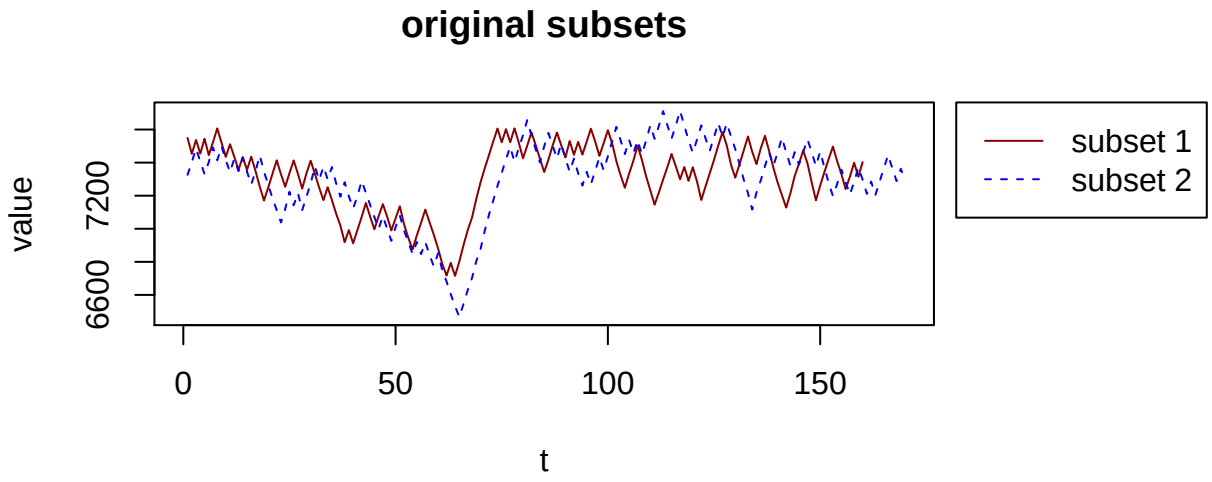


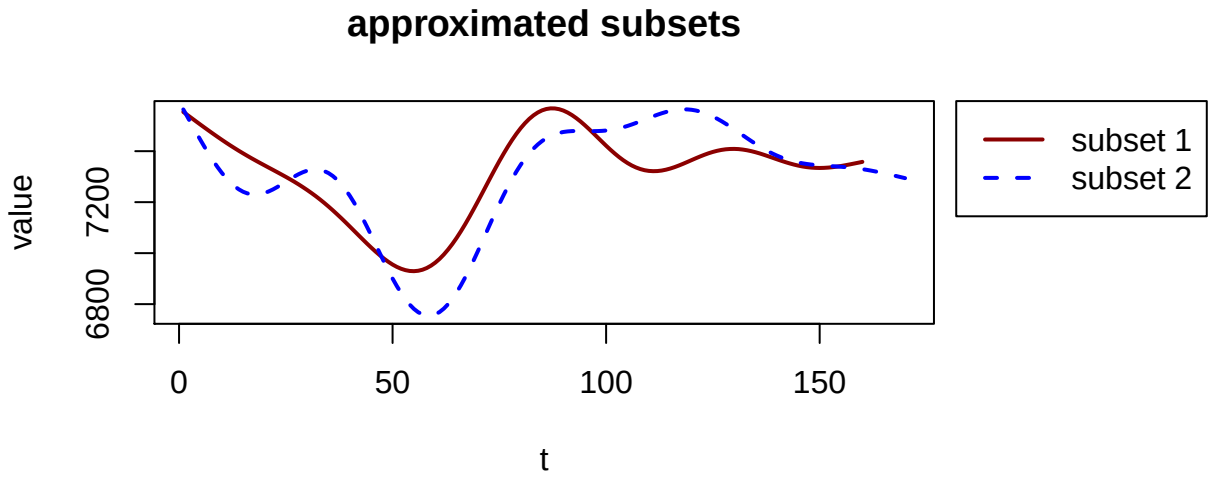
Table 2: Generalized additive method coefficients

	model.coefficients	model2.coefficients
(Intercept)	7311	7301
s(t).1	-420	-741
s(t).2	630	900
s(t).3	-303	-526
s(t).4	272	214
s(t).5	-53	72
s(t).6	-568	-1011
s(t).7	156	-307
s(t).8	-734	-1501
s(t).9	-208	-839

Generalized additive model formula:

- $g(E(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_m(x_m) + \varepsilon$
- $x_1, x_2, \ldots, x_m$ - are independent variables
- β_0 - is an intercept
- $f_1, f_2, \ldots, f_m$ - are unknow smooth functions (splines)
- ϵ - is an random error





Correlation between approximated signals

```
## [1] 0.8334519
```

We can make further signal predictions based on this models.