Projekt z analizy danych

Tomasz Kazmierczak

1/21/2018

Podsumowanie analizy

Przedmiotem analizy są pomiary zgromadzone z 17 czujników umieszczonych przy panelach fotowoltaicznych. // to do

Wykorzystane biblioteki

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(caret)
```

Wczytanie danych

```
elektrownie <- read.csv("elektrownie.csv")
```

Podsumowanie zbioru danych

Rozmiar zbioru

```
dimensions <- dim(elektrownie)
names(dimensions) <- c("kolumny", "wiersze")
dimensions
```

```
## kolumny wiersze
##  235790      51
```

Statystyki zbioru danych

```
summary(elektrownie)
```

```
##        id              idsito          idmodel          idbrand
##  Min.   :     1   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 99646   1st Qu.:0.1000   1st Qu.:0.1670   1st Qu.:0.0830
##  Median :158594   Median :0.2250   Median :0.2080   Median :0.1670
##  Mean   :152703   Mean   :0.2147   Mean   :0.2426   Mean   :0.1519
##  3rd Qu.:217541   3rd Qu.:0.3250   3rd Qu.:0.2920   3rd Qu.:0.1670
##  Max.   :276488   Max.   :0.4250   Max.   :0.7500   Max.   :0.4170
##
##       lat              lon           ageinmonths          anno
##  Min.   :0.4150   Min.   :0.1540   Min.   :0.0000   Min.   :2012
##  1st Qu.:0.4370   1st Qu.:0.6200   1st Qu.:0.0000   1st Qu.:2012
##  Median :0.4370   Median :0.6240   Median :0.1250   Median :2012
##  Mean   :0.4495   Mean   :0.5711   Mean   :0.3145   Mean   :2012
##  3rd Qu.:0.4390   3rd Qu.:0.6300   3rd Qu.:0.7190   3rd Qu.:2013
##  Max.   :0.5530   Max.   :0.6910   Max.   :1.0000   Max.   :2013
```

```
##
##       day              ora                    data
##  Min.   :0.0000   Min.   :0.000   1/1/2013 10:00:    17
##  1st Qu.:0.2520   1st Qu.:0.222   1/1/2013 11:00:    17
##  Median :0.4770   Median :0.500   1/1/2013 12:00:    17
##  Mean   :0.4812   Mean   :0.500   1/1/2013 13:00:    17
##  3rd Qu.:0.7100   3rd Qu.:0.778   1/1/2013 14:00:    17
##  Max.   :1.0000   Max.   :1.000   1/1/2013 15:00:    17
##                                   (Other)       :235688
##  temperatura_ambiente  irradiamento       pressure        windspeed
##  Min.   :0.0450       Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.2120       1st Qu.:0.0000   1st Qu.:0.7480   1st Qu.:0.04200
##  Median :0.3480       Median :0.0350   Median :0.7530   Median :0.06600
##  Mean   :0.3734       Mean   :0.1091   Mean   :0.6504   Mean   :0.07622
##  3rd Qu.:0.5300       3rd Qu.:0.2040   3rd Qu.:0.7550   3rd Qu.:0.10200
##  Max.   :0.8180       Max.   :0.7100   Max.   :0.7690   Max.   :0.69600
##
##     humidity           icon           dewpoint       windbearing
##  Min.   :0.1600   Min.   :0.0000   Min.   :0.1390   Min.   :0.0000
##  1st Qu.:0.5400   1st Qu.:0.0830   1st Qu.:0.5350   1st Qu.:0.3000
##  Median :0.7000   Median :0.6670   Median :0.6190   Median :0.4780
##  Mean   :0.6844   Mean   :0.4623   Mean   :0.6055   Mean   :0.4512
##  3rd Qu.:0.8400   3rd Qu.:0.6670   3rd Qu.:0.6830   3rd Qu.:0.6600
##  Max.   :1.0000   Max.   :0.7500   Max.   :0.8650   Max.   :0.7690
##
##    cloudcover        tempi             irri         pressurei
##  Min.   :0.000   Min.   :0.0090   Min.   :0.108   Min.   :0.000000
##  1st Qu.:0.230   1st Qu.:0.0730   1st Qu.:0.216   1st Qu.:0.000000
##  Median :0.310   Median :0.1110   Median :0.220   Median :0.000000
##  Mean   :0.359   Mean   :0.1225   Mean   :0.222   Mean   :0.000237
##  3rd Qu.:0.510   3rd Qu.:0.1260   3rd Qu.:0.222   3rd Qu.:0.000000
##  Max.   :1.000   Max.   :0.9830   Max.   :1.000   Max.   :1.000000
##
##    windspeedi        humidityi         dewpointi       windbearingi
##  Min.   :0.00000   Min.   :0.03400   Min.   :0.0630   Min.   :0.0000
##  1st Qu.:0.03700   1st Qu.:0.04400   1st Qu.:0.1140   1st Qu.:0.3360
##  Median :0.03800   Median :0.04400   Median :0.1140   Median :0.3360
##  Mean   :0.03852   Mean   :0.06384   Mean   :0.1194   Mean   :0.3455
##  3rd Qu.:0.03900   3rd Qu.:0.06200   3rd Qu.:0.1180   3rd Qu.:0.3390
##  Max.   :1.00000   Max.   :0.57900   Max.   :0.4150   Max.   :1.0000
##
##   cloudcoveri          dist           altitude         azimuth
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.1110   Min.   :0.1280
##  1st Qu.:0.1960   1st Qu.:0.1913   1st Qu.:0.4190   1st Qu.:0.2950
##  Median :0.1960   Median :0.4590   Median :0.5640   Median :0.4250
##  Mean   :0.2062   Mean   :0.4686   Mean   :0.5464   Mean   :0.4546
##  3rd Qu.:0.1980   3rd Qu.:0.7268   3rd Qu.:0.6810   3rd Qu.:0.6350
##  Max.   :1.0000   Max.   :1.0000   Max.   :0.8840   Max.   :0.8180
##
##    altitudei         azimuthi          pcnm1            pcnm2
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0960   1st Qu.:0.2090   1st Qu.:0.3770   1st Qu.:0.2500
##  Median :0.1360   Median :0.2880   Median :0.3780   Median :0.3770
##  Mean   :0.2055   Mean   :0.3653   Mean   :0.4224   Mean   :0.3538
```

```
##  3rd Qu.:0.2660   3rd Qu.:0.4820   3rd Qu.:0.3800   3rd Qu.:0.4220
##  Max.   :0.9820   Max.   :1.0000   Max.   :1.0000   Max.   :0.9720
##
##      pcnm3            pcnm4            pcnm5            pcnm6
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.5510   1st Qu.:0.3630   1st Qu.:0.3310   1st Qu.:0.3390
##  Median :0.6050   Median :0.5310   Median :0.4270   Median :0.4930
##  Mean   :0.6045   Mean   :0.5189   Mean   :0.4165   Mean   :0.4941
##  3rd Qu.:0.7300   3rd Qu.:0.6340   3rd Qu.:0.4620   3rd Qu.:0.4930
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##      pcnm7            pcnm8            pcnm9            pcnm10
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0310   1st Qu.:0.2040   1st Qu.:0.5270   1st Qu.:0.5530
##  Median :0.0520   Median :0.4120   Median :0.5320   Median :0.6190
##  Mean   :0.1142   Mean   :0.4034   Mean   :0.5371   Mean   :0.6276
##  3rd Qu.:0.1140   3rd Qu.:0.5110   3rd Qu.:0.6000   3rd Qu.:0.7170
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##      pcnm11           pcnm12           pcnm13           pcnm14
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.1370   Min.   :0.0000
##  1st Qu.:0.2570   1st Qu.:0.7480   1st Qu.:0.6140   1st Qu.:0.4320
##  Median :0.3270   Median :0.7600   Median :0.6140   Median :0.4730
##  Mean   :0.3236   Mean   :0.7568   Mean   :0.6501   Mean   :0.4893
##  3rd Qu.:0.3270   3rd Qu.:0.8840   3rd Qu.:0.7380   3rd Qu.:0.5300
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##      pcnm15       irr_pvgis_mod    irri_pvgis_mod         kwh
##  Min.   :0.0000   Min.   :0.0000   Min.   :-0.0250   Min.   :0.0000
##  1st Qu.:0.6120   1st Qu.:0.0000   1st Qu.: 0.1580   1st Qu.:0.0000
##  Median :0.6140   Median :0.0560   Median : 0.1940   Median :0.0490
##  Mean   :0.5709   Mean   :0.1767   Mean   : 0.1967   Mean   :0.1688
##  3rd Qu.:0.6150   3rd Qu.:0.3250   3rd Qu.: 0.2130   3rd Qu.:0.3320
##  Max.   :1.0000   Max.   :1.0000   Max.   : 1.0060   Max.   :1.0000
##
```

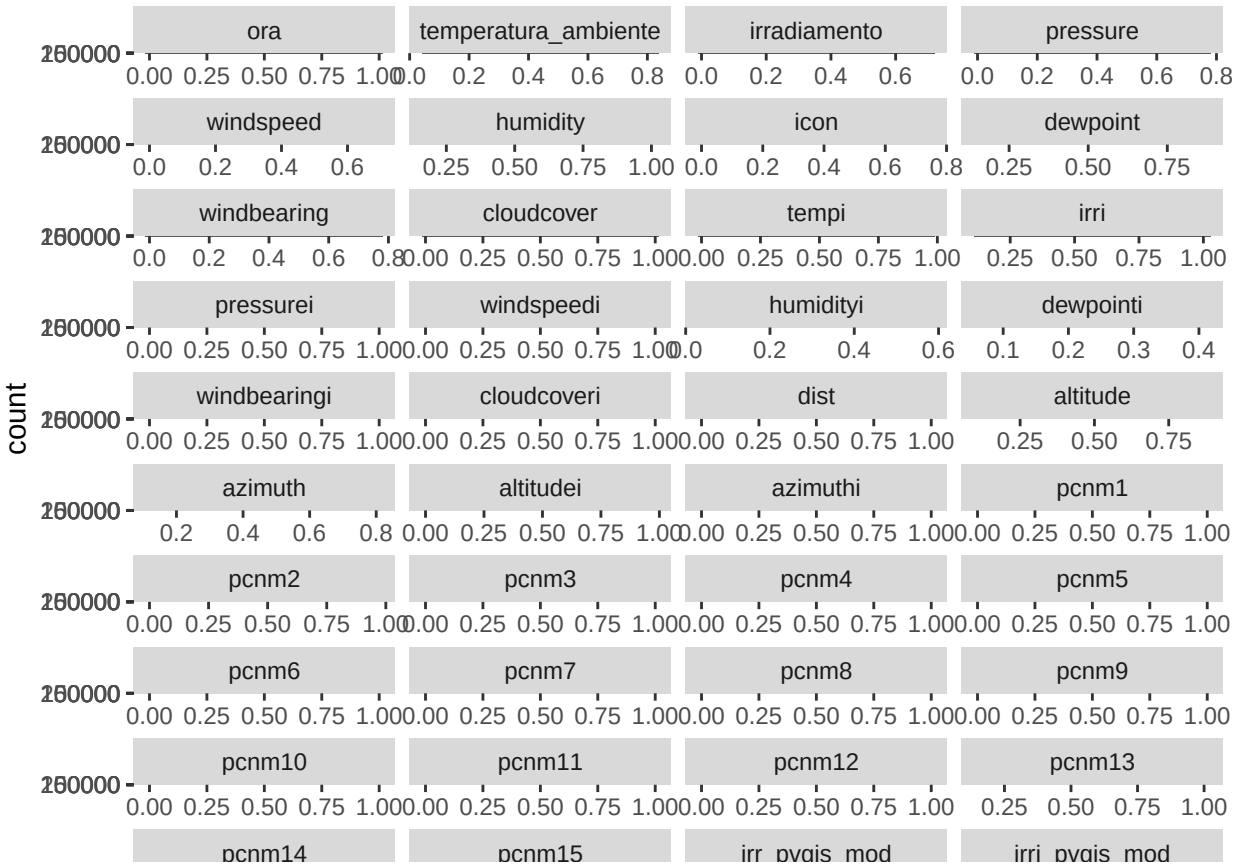
Sprawdzenie czy w zbiorze występują wartości NA

```
any(is.na(elektrownie))
```

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

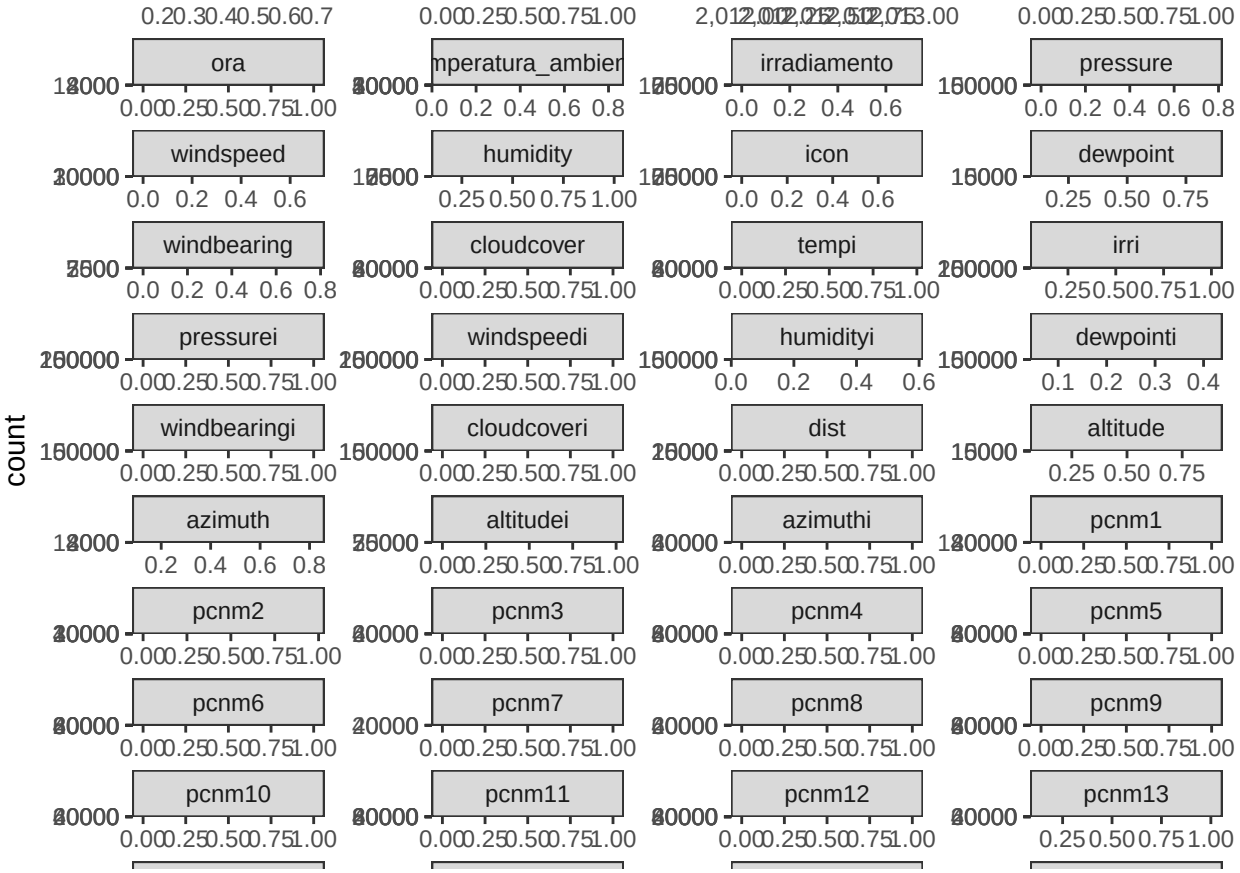
```
ggplot(melt(elektrownie[,-c(1)]),aes(x = value)) +
    facet_wrap(~variable,ncol=4,scales = "free_x") + geom_histogram(bins = 30) + scale_x_continuous(labe
```

```
## Using data as id variables
```




```
ggplot(data = melt(elektrownie[,-c(1)]), mapping = aes(x = value)) +
  geom_histogram(bins=50) +
  labs(title = "Rozkłady wartości atrybutów") +
  facet_wrap(~variable, ncol=4, scales = 'free') +
  scale_x_continuous(labels = scales::comma) +
  theme_bw()
```

```
## Using data as id variables
```


count
ora
temperatura_ambient
irradiamento
pressure
windspeed
humidity
icon
dewpoint
windbearing
cloudcover
tempi
irri
pressurei
windspeedi
humidityi
dewpointi
windbearingi
cloudcoveri
dist
altitude
azimuth
altitudei
azimuthi
pcnm1
pcnm2
pcnm3
pcnm4
pcnm5
pcnm6
pcnm7
pcnm8
pcnm9
pcnm10
pcnm11
pcnm12
pcnm13