Spis treści

Posumowanie analizy zbioru danych

W danych znajdowało się bardzo dużo wartości NA. W obliczeniach w zależności co było liczone zostały one zamienione lub pominięte - szczególnie podczas obliczania korelacji (sposob obsługi wartości NA podczas obliczania korelacji został opisany w punkcie 5).

Zauważono także, że kolumny fo_col fc_col zawierają tylko 1 wartość. Powinny one zostać usunięte ze zbioru przy próbie utworzenia klasyfikatora.

Podczas rysowania wykresów dla kolumn zaczynających się od part_01 zauważono, że dużo wartości jest skupionych w okolicy zera.

Została podjęta próba wykonania klasyfikatora, ale niestety nie zakończyła się powodzeniem. W ostatnim punkcie zostały opisane podjęte kroki, które zostały podjęte oraz napotkane błędy.

Wykorzystane biblioteki

```
library(knitr)
library(ggplot2)
library(dplyr)
library(ggExtra)
library(caret)
library(corrplot)
library(plotly)
library(randomForest)
```

Kod zapewniający powtarzalność

Wczytanie danych z pliku

Wczytanie danych z pliku.

```
rawData <- data.table::fread("all_summary.csv", header="auto",sep="auto")
```

```
## Warning in data.table::fread("all_summary.csv", header = "auto", sep
## = "auto"): Bumped column 6 to type character on data row 13920, field
## contains '260G'. Coercing previously read values in this column from
## logical, integer or numeric back to character which may not be lossless;
## e.g., if '00' and '000' occurred before they will now be just '0', and
## there may be inconsistencies with treatment of ',,' and ',NA,' too (if they
## occurred in this column before the bump). If this matters please rerun and
## set 'colClasses' to 'character' for this column. Please note that column
## type detection uses a sample of 1,000 rows (100 rows at 10 points) so
## hopefully this message should be very rare. If reporting to datatable-help,
## please rerun and include the output from verbose=TRUE.
```

Załadowano 591042 wierszy, które mają 412 zmiennych.

Zliczenie 50 najpopularniejszych

```
popular <- rawData[ , .N, by = res_name]
popular <- popular[order(-N)]
popular <- popular[1:50]
pop_names <- popular$res_name
rawData <- select(filter(rawData, res_name %in% pop_names),matches("*"))
```

Pozostało 382720 wierszy z 50 najpopularniejszych grup.

Usuwanie z danych wiersze posiadające wartość zmiennej res_name równą: “UNK”, “UNX”, “UNL”, “DUM”, “N”, “BLOB”, “ALA”, “ARG” ,“ASN”, “ASP”, “CYS”, “GLN”, “GLU”, “GLY”, “HIS”, “ILE”, “LEU”, “LYS”, “MET”, “MSE”, “PHE”, “PRO”, “SEC”, “SER”, “THR”, “TRP”, “TYR”, “VAL”, “DA”, “DG”, “DT”, “DC”, “DU”, “A”, “G”, “T”, “C”, “U”, “HOH”, “H20”, “WAT” lub “NAN # # Wstępne czyszczenie danych

```
selectedData  <-  selectedData  <-  rawData %>% filter(!(res_name %in% c("UNK", "UNX", "UNL", "DUM", "N", "BLOB", "ALA", "ARG" ,"ASN", "ASP", "CYS", "GLN", "GLU", "GLY", "HIS", "ILE", "LEU", "LYS", "MET", "MSE", "PHE", "PRO", "SEC", "SER", "THR", "TRP", "TYR", "VAL", "DA", "DG", "DT", "DC", "DU", "A", "G", "T", "C", "U", "HOH", "H20", "WAT", "NAN", "", "NA", NA)))
```

Podsumowanie wartości w kolumnach

blob_coverage	res_coverage	title	pdb_code	res_name	res_id	chain_id	blob_volume_coverage	blob_volume_coverage_second	res_vo
Length:382720	Length:382720	Length:382720	Length:382720	Length:382720	Length:382720	Length:382720	Min. :0.02004	Min. :0.00000	Min. :0.
Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	1st Qu.:0.44038	1st Qu.:0.00000	1st Qu.
Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Median :0.69809	Median :0.00000	Median
NA	NA	NA	NA	NA	NA	NA	Mean :0.64142	Mean :0.02088	Mean :
NA	NA	NA	NA	NA	NA	NA	3rd Qu.:0.85816	3rd Qu.:0.00000	3rd Qu
NA	NA	NA	NA	NA	NA	NA	Max. :1.00000	Max. :0.99625	Max. :1
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

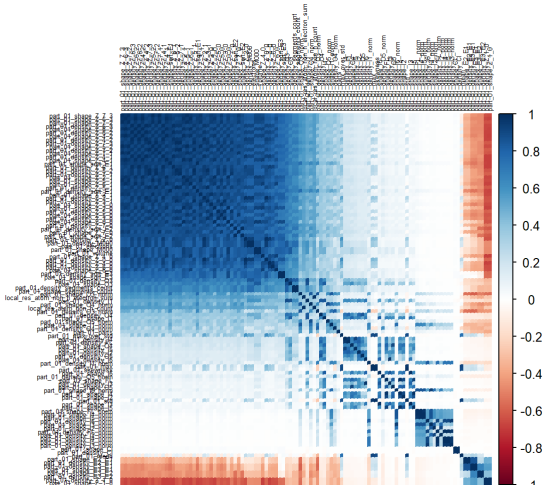
Korelacje między zmiennymi

Do obliczenia korelacji użytko funkcji >cor< z parametrem >use = “pairwise.complete.obs<, który ignoruje w obliczeniach korelacji dla danej pary wartości NA.

Podczas obliczania korelacji zauważono, że dla wszystkich kolumn part_XX jest ona bardzo podobna. W przedstawionej graficznej reprezentacji korelacji zabrano zmienne z part_01, by ograniczyć liczbę danych. Poza kolumnami part_01 w macierzy widzimy kolumy local_res_atom_non_h_electron_sum, local_res_atom_non_h_count, solvent_mask_count, void_mask_count, modeled_mask_count, solvent_ratio. Wybrano takie kolumny, ponieważ te kolumny będą brały w wyznaczaniu klasyfikatora.

Zmienne zostały posortowane wg algorytmu “FPC” (First Principal Component).

```
correlation_data  <- cor(data, use = "pairwise.complete.obs")
corrplot(correlation_data, method = "color", tl.cex = 0.4, order = "FPC", tl.col="black")
```



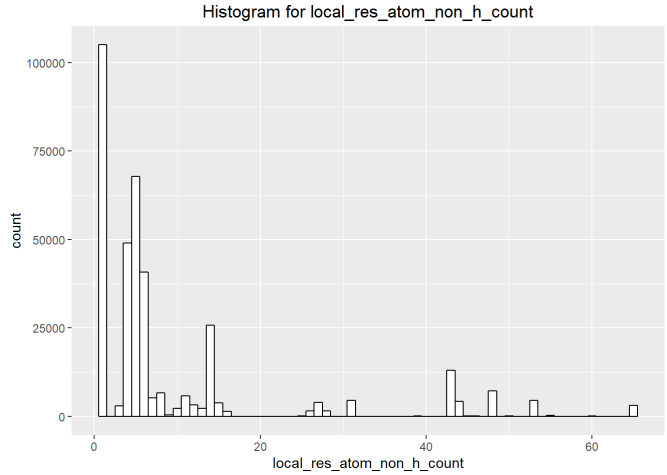
Liczba przypadków dla każdej z klas

count	class
1587	SAH
1589	GDP
1594	PLP
1596	NO3
1602	FE
1609	ACY
1637	NI
1647	SF4
1656	TRS
1905	PGE
1917	HEC
1933	EPE
2084	FMN
2106	NDP
2127	BR
2136	1PE
2183	COA
2296	ATP
2353	CU
2697	MES
2768	PG4
2841	MAN
2918	FMT
3221	MPD
3242	CD
3505	NAP
3509	MLY
3819	ADP
4215	MN
4501	NAD
4555	FAD
4706	K
4784	CLA
4987	PEG
6317	IOD
6633	DMS
8096	ACT
11090	PO4
11192	HEM

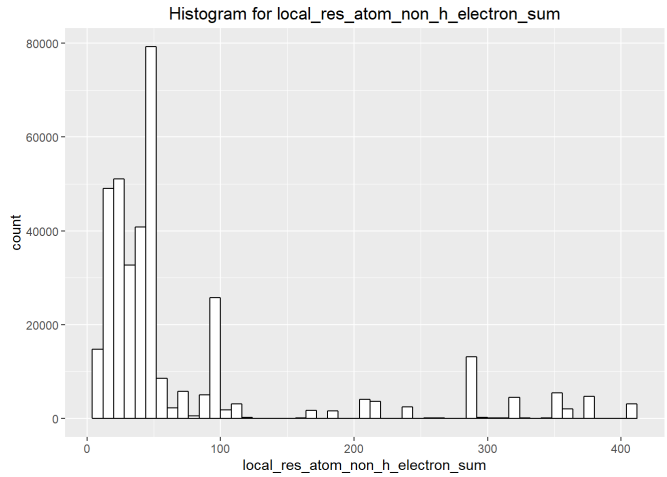
count	class
14779	MG
19826	ZN
21038	CA
23223	CL
26360	NAG
30825	EDO
40606	GOL
56572	SO4

Wykresy rozkładu liczby atomów i elektronów

Rozkład atomów



Rozkład elektronów



Klasy z największą niezgodnością liczby atomów i elektronów

Klasy z największą niezgodnością liczby atomów

res_name	local_res_atom_non_h_count	dict_atom_non_h_count	odds
NAG	369564	395400	25836
CLA	285867	310960	25093
1PE	28487	34176	5689
MLY	37453	42108	4655
NAP	163742	168240	4498
COA	100779	104784	4005
NAD	194945	198044	3099
PG4	33094	35984	2890
MAN	31551	34092	2541
NDP	99007	101088	2081

Klasy z największą niezgodnością liczby elektronów

res_name	local_res_atom_non_h_electron_sum	dict_atom_non_h_electron_sum	odds
NAG	2508452	2715080	206628
CLA	1809084	1961440	152356

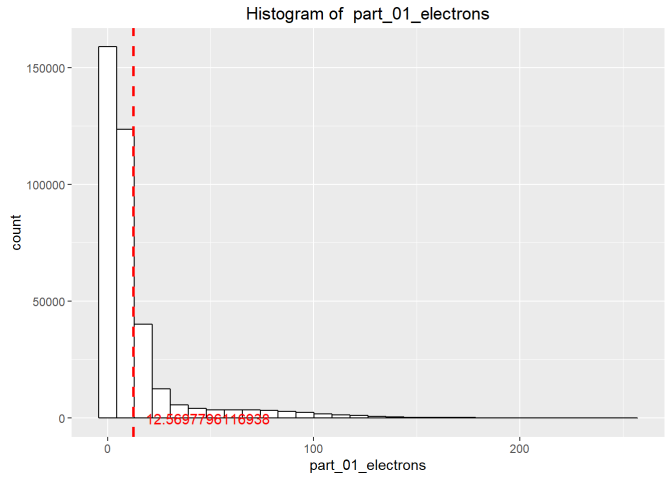
res_name	local_res_atom_non_h_electron_sum	dict_atom_non_h_electron_sum	odds
1PE	191894	230688	38794
MLY	238696	273702	35006
NAP	1217531	1247780	30249
COA	764523	794612	30089
NAD	1406030	1426817	20787
MAN	218316	238644	20328
PG4	224042	243584	19542
NDP	735538	749736	14198

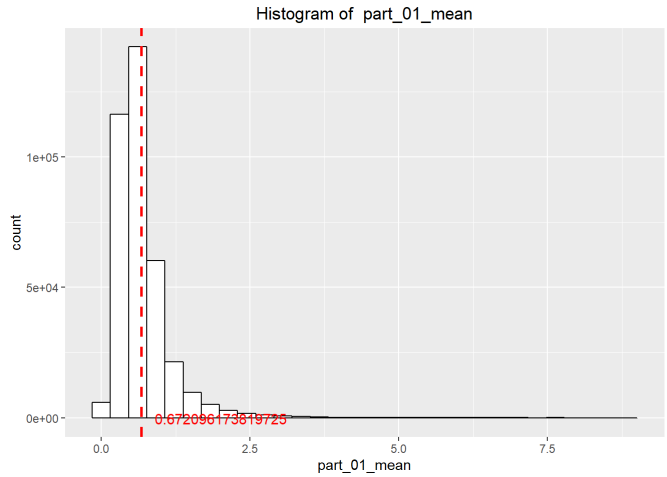
Rozkład wartości kolumn part_01

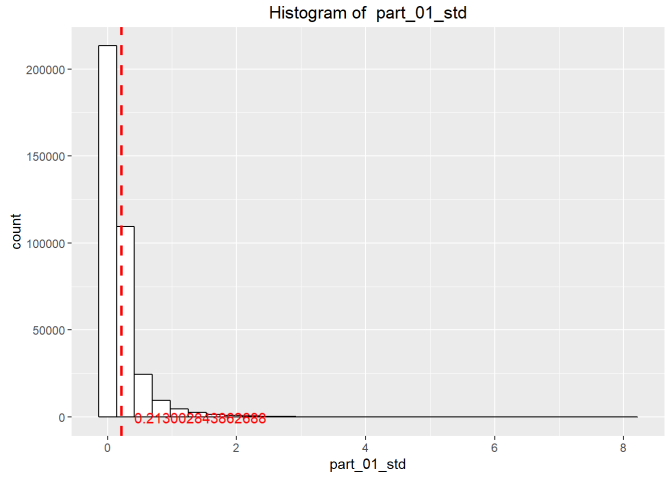
Sekcja przedstawia rozkład wartości wszystkich kolumn zaczynających się od part_01

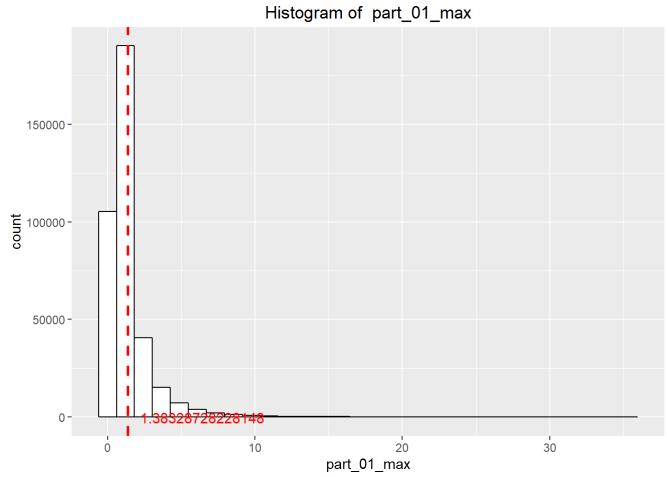
Usunięto wartości NA dla każdej kolumny z osobna. Nie zostały zamienione na wartość 0 by nie zaburzać rozkładu zmiennych.

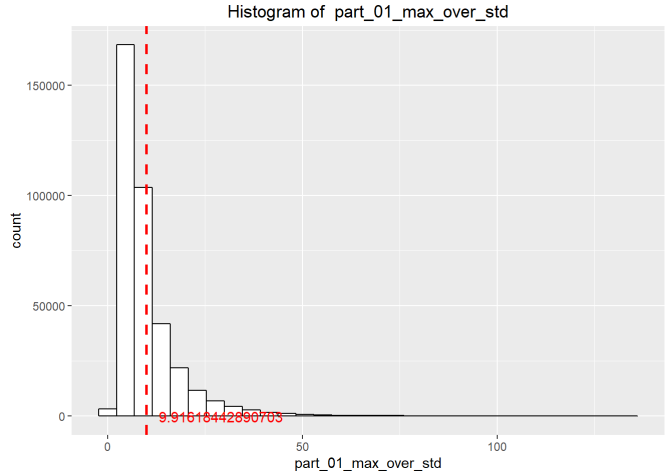
Na wykresach zaznaczono średnią wartość zmiennej (w formie graficznej oraz liczbowej).

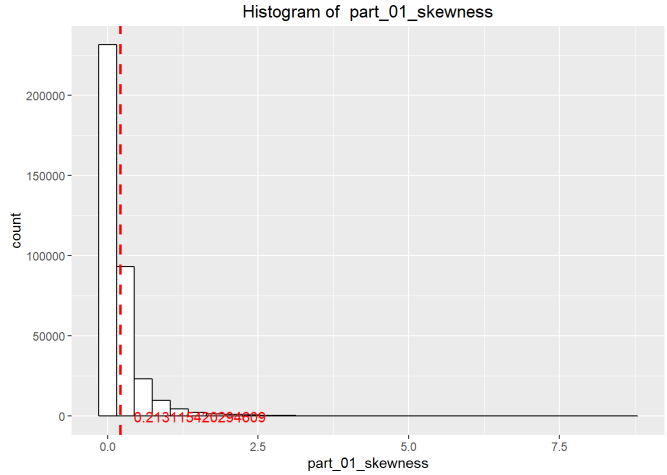


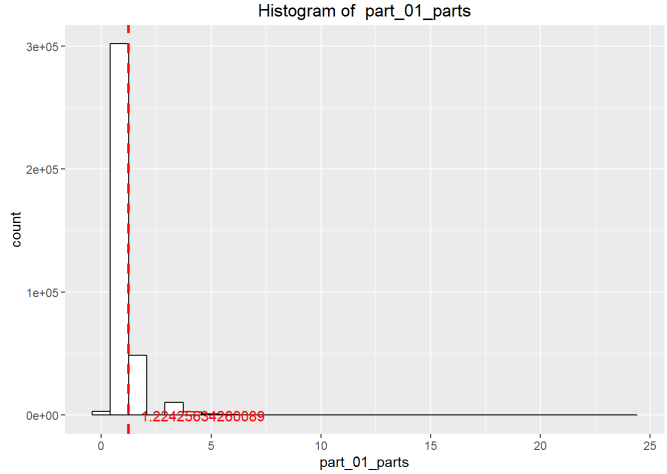


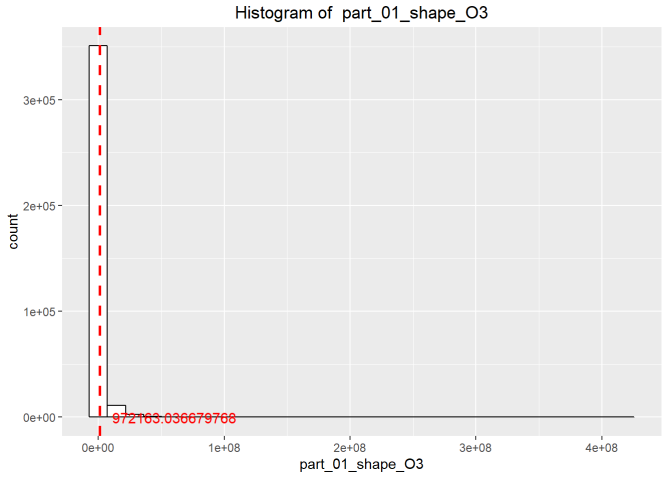


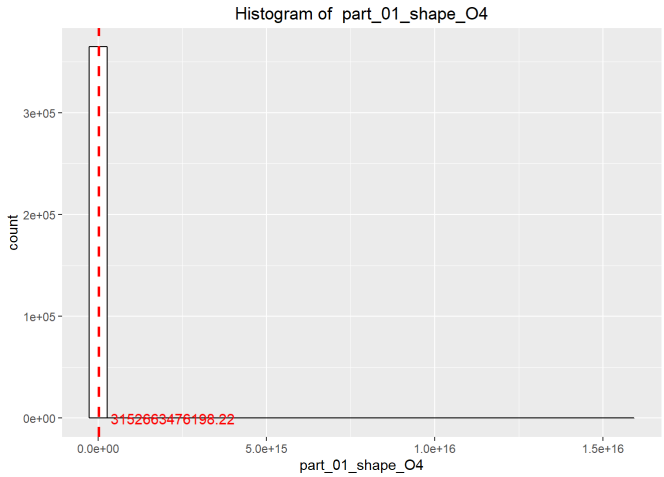

Histogram of part_01_max
count
150000
100000
50000
0
1.3852672822261148
0
10
20
30
part_01_max

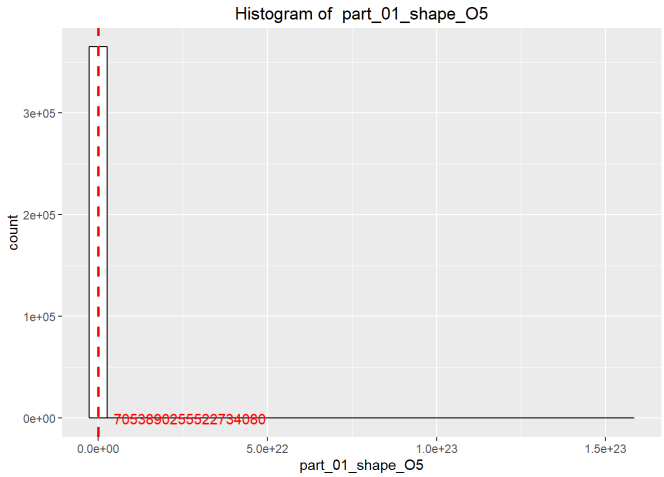

Histogram of part_01_max_over_std
count
150000
100000
50000
0
9.918184428907703
0
50
100
part_01_max_over_std

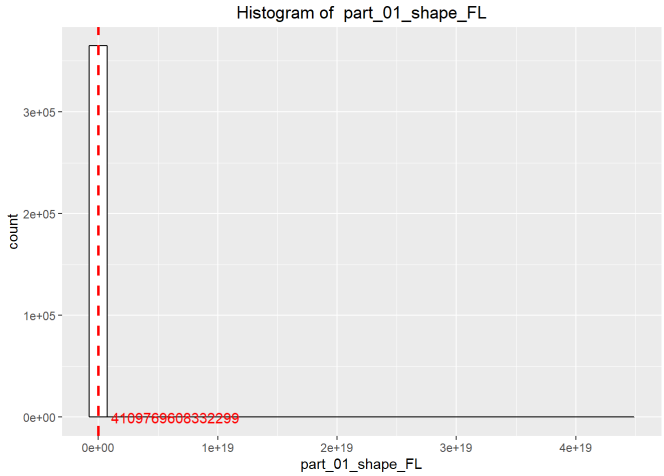

Histogram of part_01_skewness
count
200000
150000
100000
50000
0
0.21311542029468
0.0
2.5
5.0
7.5
part_01_skewness

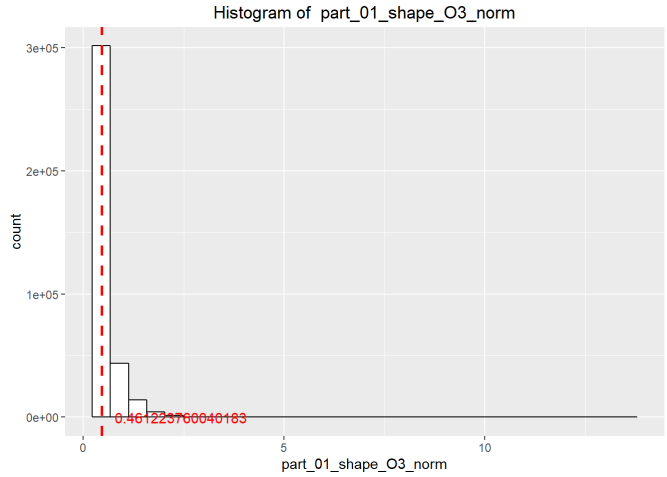

Histogram of part_01_parts
count
3e+05
2e+05
1e+05
0e+00
1.2242563426008
0
5
10
15
20
25
part_01_parts

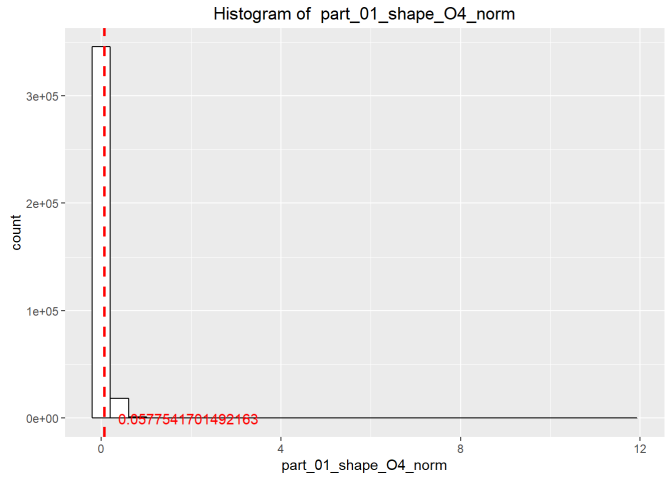

Histogram of part_01_shape_O3
count
3e+05
2e+05
1e+05
0e+00
972783.036679768
0e+00
1e+08
2e+08
3e+08
4e+08
part_01_shape_O3

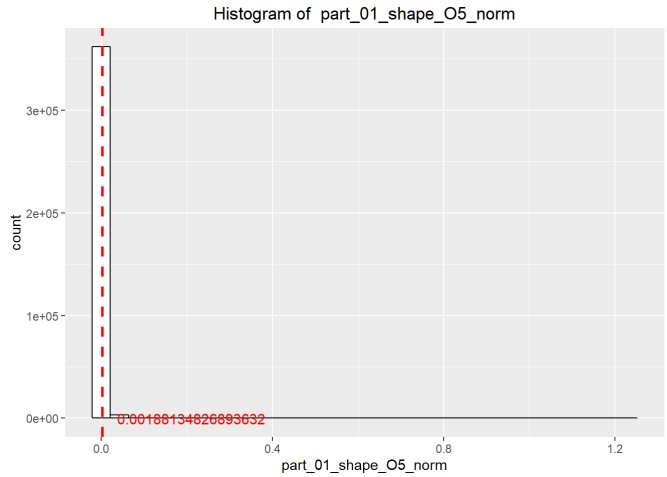

Histogram of part_01_shape_O4
count
3e+05
2e+05
1e+05
0e+00
3152663476198.22
0.0e+00
5.0e+15
1.0e+16
1.5e+16
part_01_shape_O4

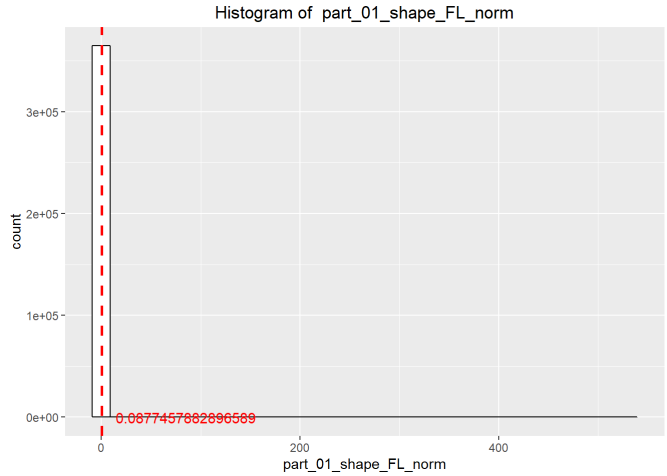

Histogram of part_01_shape_O5
count
3e+05
2e+05
1e+05
0e+00
7053890255522734080
0.0e+00
5.0e+22
1.0e+23
1.5e+23
part_01_shape_O5

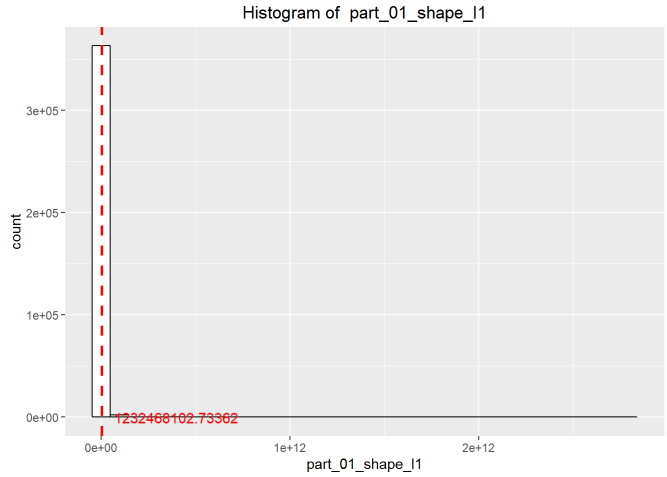

Histogram of part_01_shape_FL
count
3e+05
2e+05
1e+05
0e+00
4109769008332299
0e+00
1e+19
2e+19
3e+19
4e+19
part_01_shape_FL


Histogram of part_01_shape_O3_norm
count
3e+05
2e+05
1e+05
0e+00
0.461223760040183
0
5
10
part_01_shape_O3_norm

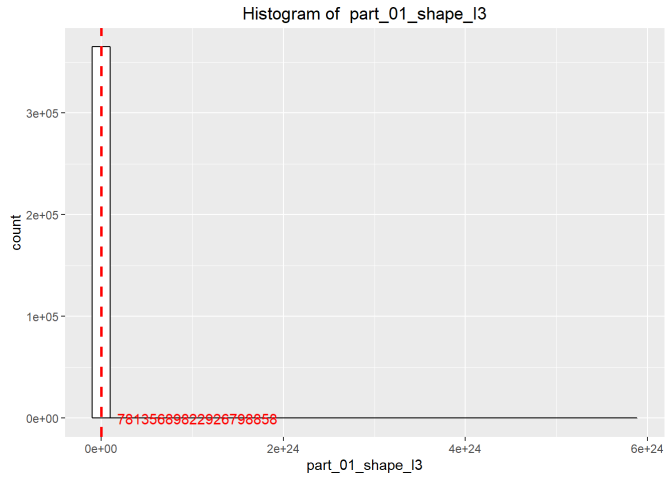

Histogram of part_01_shape_O4_norm
count
3e+05
2e+05
1e+05
0e+00
0.0577541701492163
0
4
8
12
part_01_shape_O4_norm

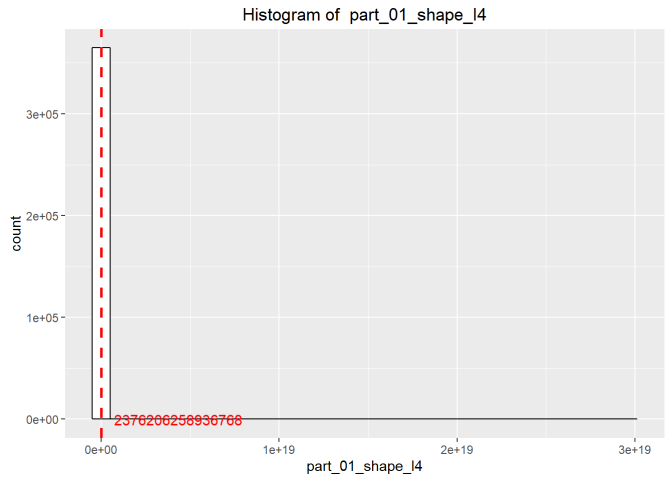

Histogram of part_01_shape_O5_norm
count
3e+05
2e+05
1e+05
0e+00
0.00188134826893632
0.0
0.4
0.8
1.2
part_01_shape_O5_norm

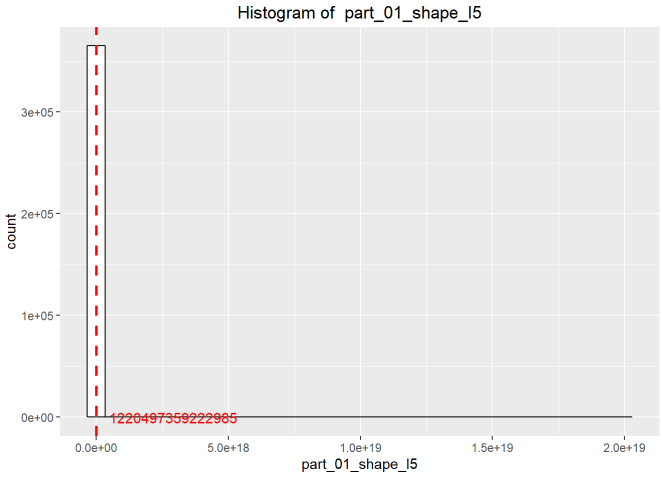

Histogram of part_01_shape_FL_norm
count
3e+05
2e+05
1e+05
0e+00
0.0877457882896589
0
200
400
part_01_shape_FL_norm

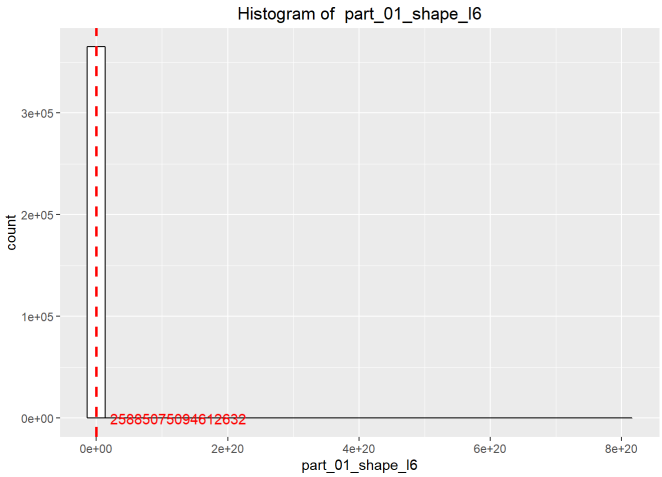

Histogram of part_01_shape_l1
count
3e+05
2e+05
1e+05
0e+00
1232468102.73362
0e+00
1e+12
2e+12
part_01_shape_l1

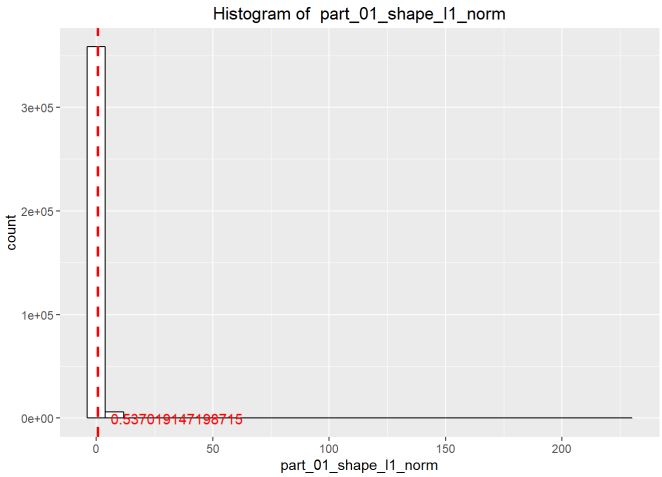

Histogram of part_01_shape_l2
count
3e+05
2e+05
1e+05
0e+00
897233401810970419 2
0.0e+00
5.0e+22
1.0e+23
1.5e+23
part_01_shape_l2

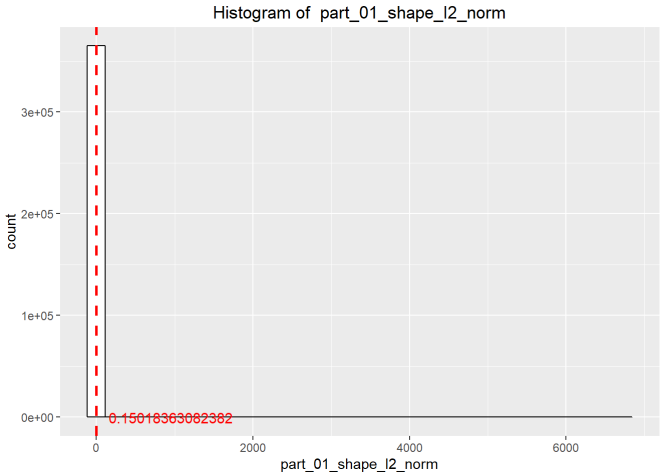

Histogram of part_01_shape_l3
count
3e+05
2e+05
1e+05
0e+00
7813568982292679 8656
0e+00
2e+24
4e+24
6e+24
part_01_shape_l3

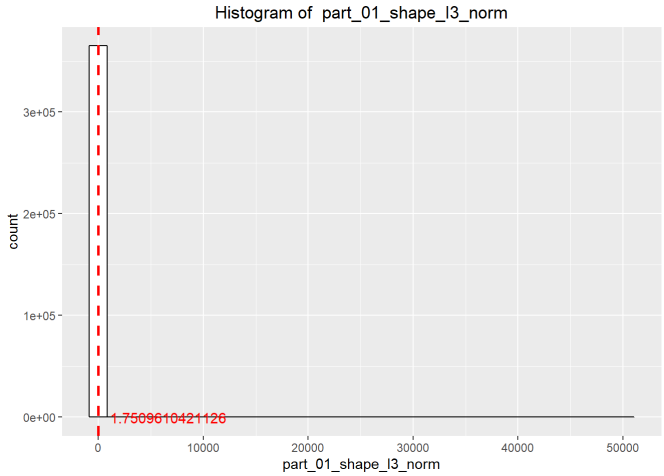

Histogram of part_01_shape_l4
count
3e+05
2e+05
1e+05
0e+00
2376206258936768
0e+00
1e+19
2e+19
3e+19
part_01_shape_l4

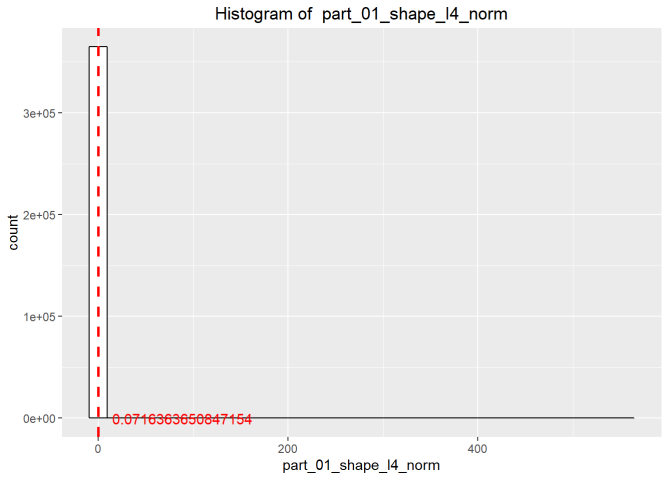

Histogram of part_01_shape_l5
count
3e+05
2e+05
1e+05
0e+00
1220497359222965
0.0e+00
5.0e+18
1.0e+19
1.5e+19
2.0e+19
part_01_shape_l5

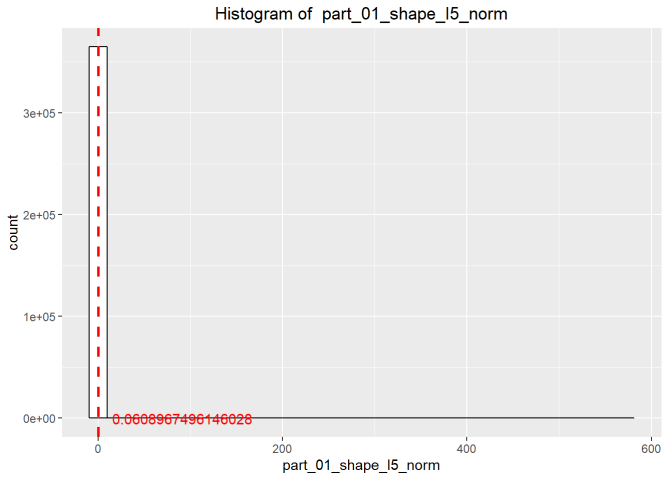

Histogram of part_01_shape_l6
count
3e+05
2e+05
1e+05
0e+00
25885075094612632
0e+00
2e+20
4e+20
6e+20
8e+20
part_01_shape_l6

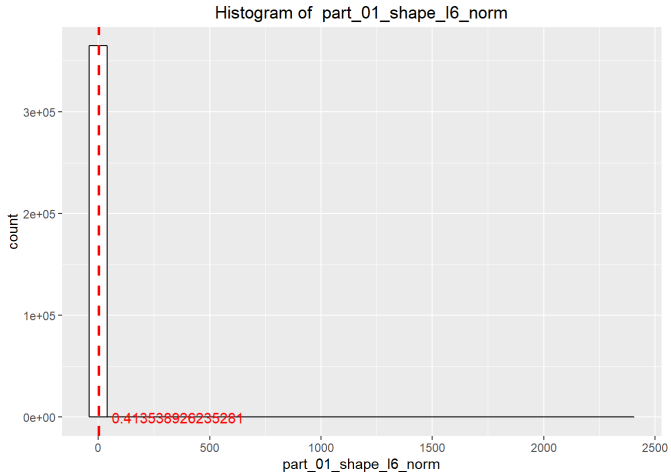

Histogram of part_01_shape_l1_norm
count
3e+05
2e+05
1e+05
0e+00
0.537019147198715
0
50
100
150
200
part_01_shape_l1_norm

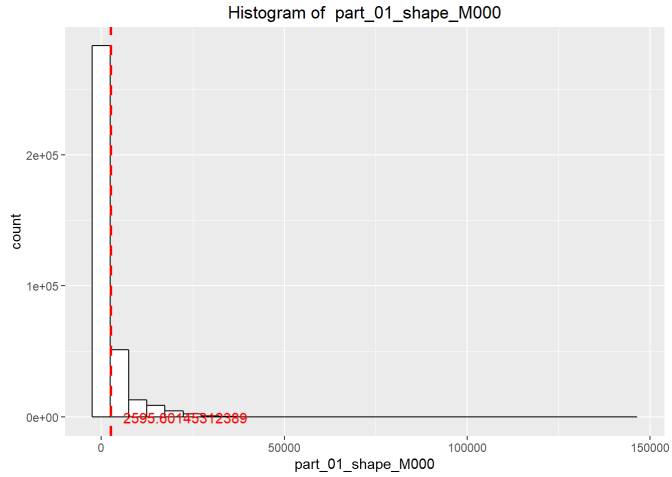

Histogram of part_01_shape_l2_norm
count
3e+05
2e+05
1e+05
0e+00
0.150183663082382
0
2000
4000
6000
part_01_shape_l2_norm

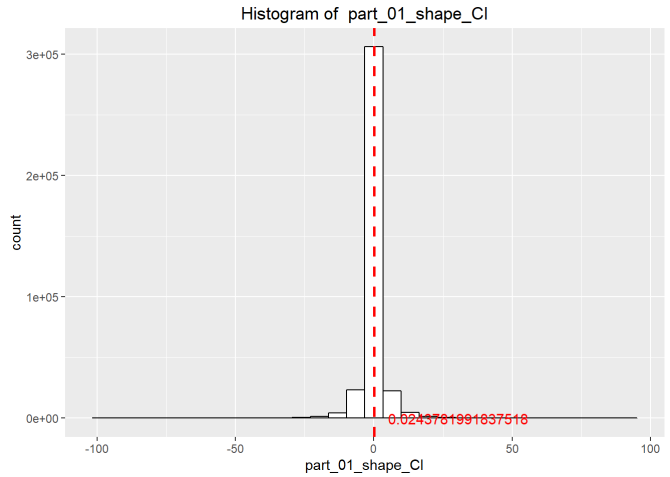

Histogram of part_01_shape_l3_norm
count
3e+05
2e+05
1e+05
0e+00
1.7509610421126
0
10000
20000
30000
40000
50000
part_01_shape_l3_norm

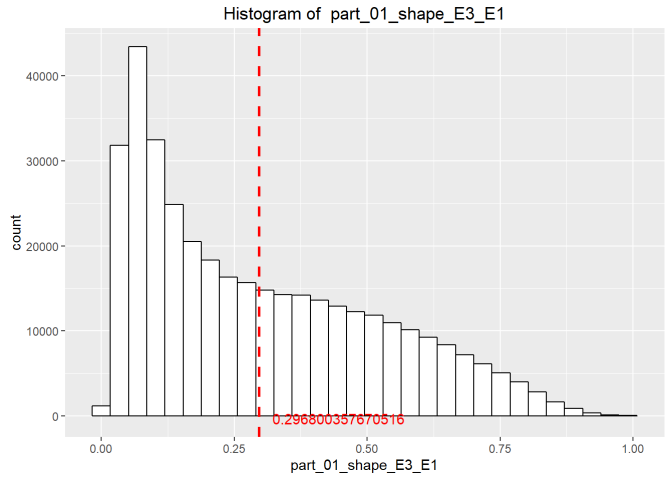

Histogram of part_01_shape_l4_norm
count
3e+05
2e+05
1e+05
0e+00
0.0716363650847154
0
200
400
part_01_shape_l4_norm

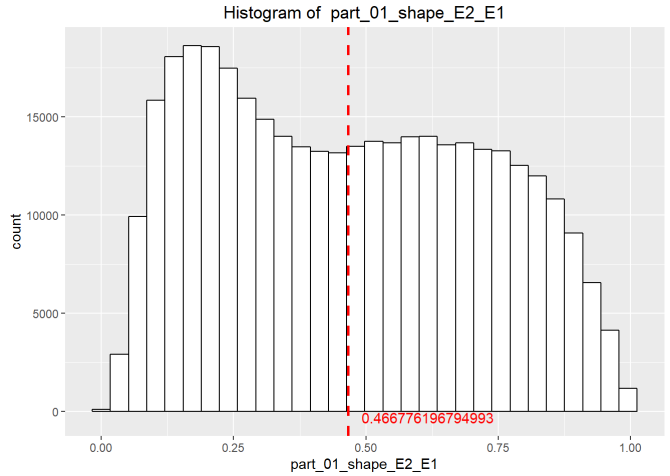

Histogram of part_01_shape_l5_norm
count
3e+05
2e+05
1e+05
0e+00
0.0608967496146028
0
200
400
600
part_01_shape_l5_norm

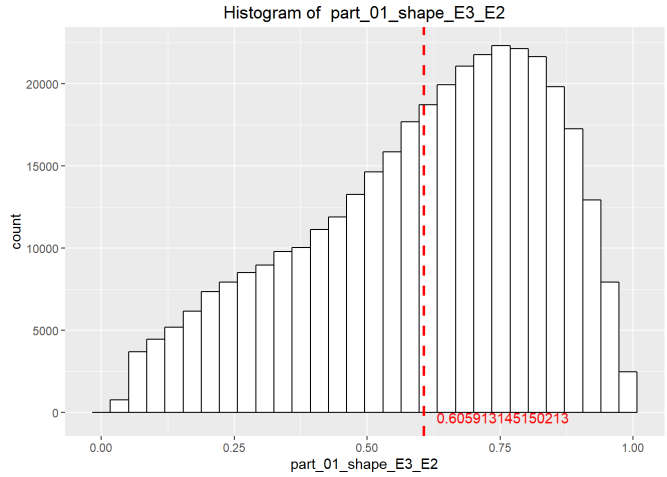

Histogram of part_01_shape_l6_norm
count
3e+05
2e+05
1e+05
0e+00
0.413538926235281
0
500
1000
1500
2000
2500
part_01_shape_l6_norm

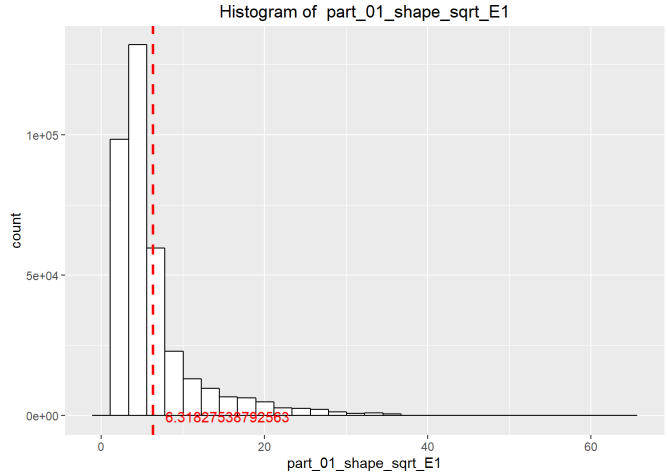

Histogram of part_01_shape_M000
count
2e+05
1e+05
0e+00
2595.0014531238 9
0
50000
100000
150000
part_01_shape_M000

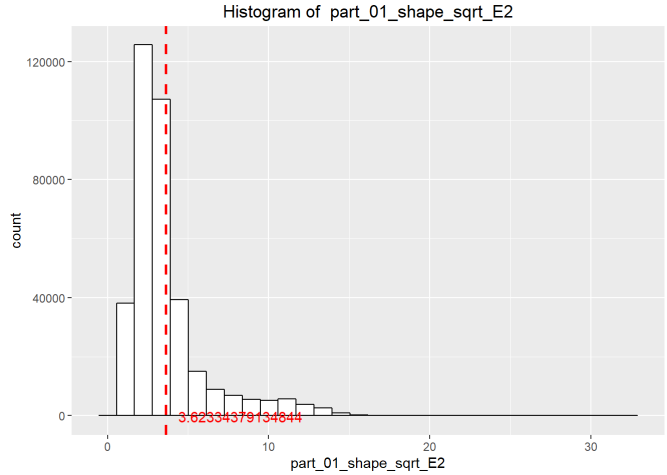

Histogram of part_01_shape_CI
count
3e+05
2e+05
1e+05
0e+00
0.02437819918375 18
-100
-50
0
50
100
part_01_shape_CI

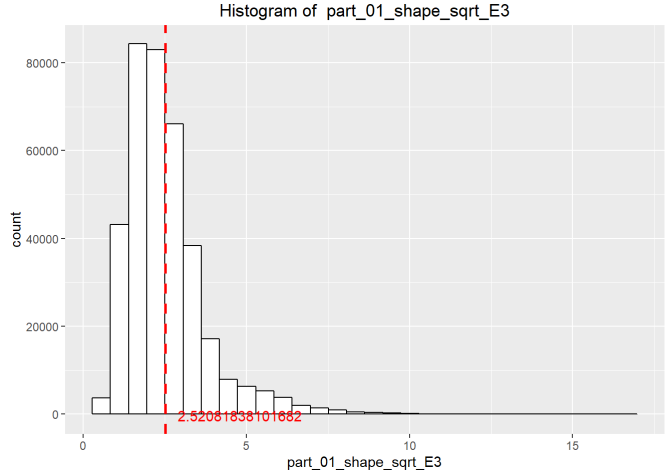

Histogram of part_01_shape_E3_E1
count
40000
30000
20000
10000
0
0.296800357670516
0.00
0.25
0.50
0.75
1.00
part_01_shape_E3_E1

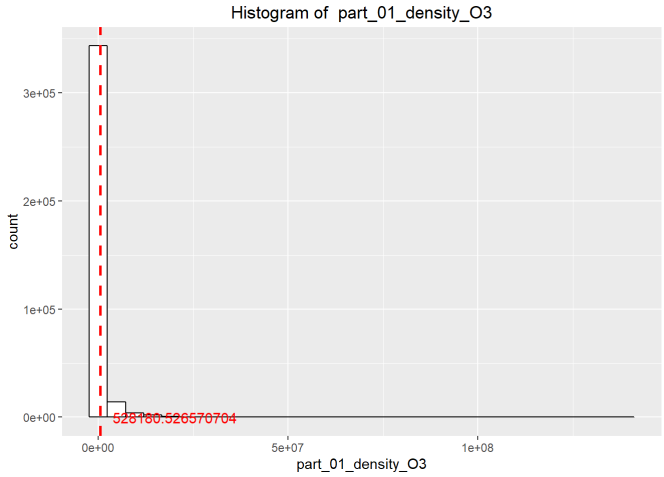

Histogram of part_01_shape_E2_E1
count
15000
10000
5000
0
0.466776196794993
0.00
0.25
0.50
0.75
1.00
part_01_shape_E2_E1

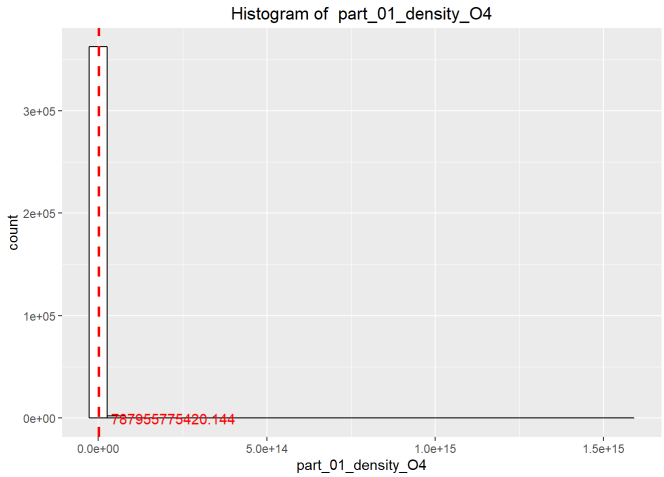

Histogram of part_01_shape_E3_E2
count
20000
15000
10000
5000
0
0.605913145150213
0.00
0.25
0.50
0.75
1.00
part_01_shape_E3_E2


Histogram of part_01_shape_sqrt_E1
count
1e+05
5e+04
0e+00
6.31827538792563
0
20
40
60
part_01_shape_sqrt_E1

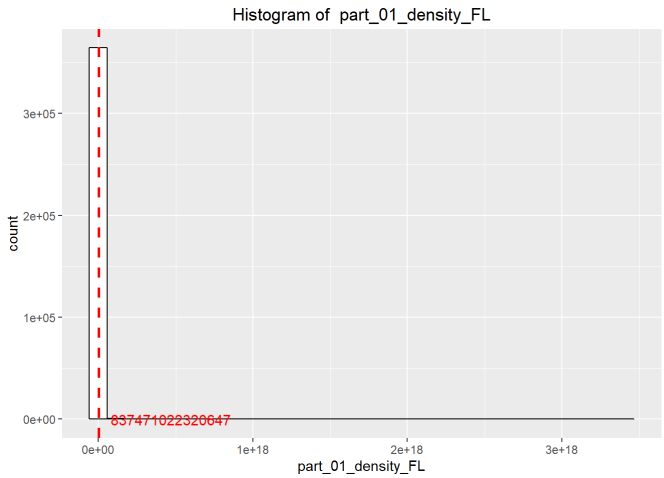

Histogram of part_01_shape_sqrt_E2
count
120000
80000
40000
0
3.62334379134844
0
10
20
30
part_01_shape_sqrt_E2

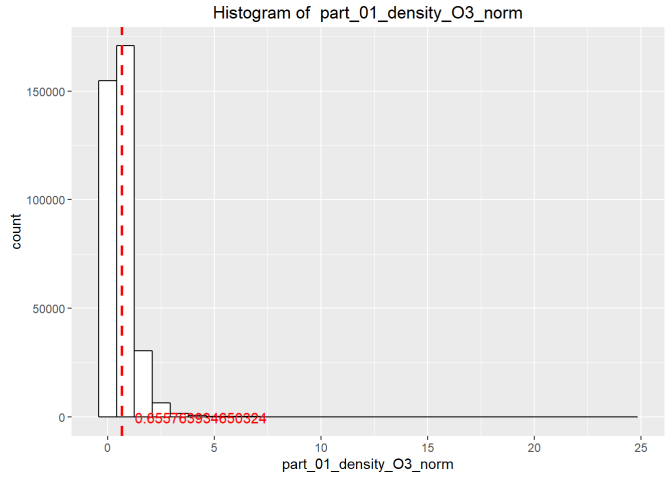

Histogram of part_01_shape_sqrt_E3
count
80000
60000
40000
20000
0
2.52081838101682
0
5
10
15
part_01_shape_sqrt_E3

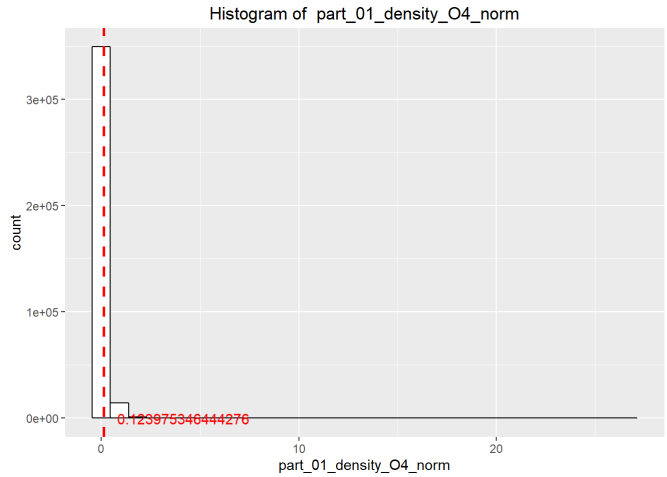

Histogram of part_01_density_O3
count
3e+05
2e+05
1e+05
0e+00
528180.526570704
0e+00
5e+07
1e+08
part_01_density_O3

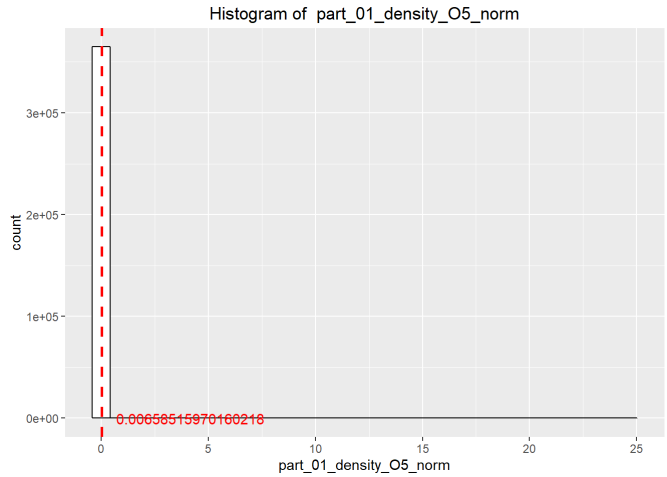

Histogram of part_01_density_O4
count
3e+05
2e+05
1e+05
0e+00
787955775420.144
0.0e+00
5.0e+14
1.0e+15
1.5e+15
part_01_density_O4

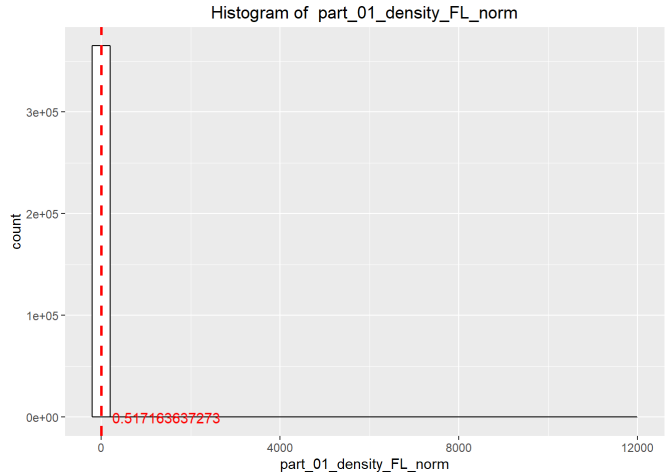

Histogram of part_01_density_O5
count
3e+05
2e+05
1e+05
0e+00
448645157480836416
0e+00
1e+21
2e+21
3e+21
4e+21
part_01_density_O5

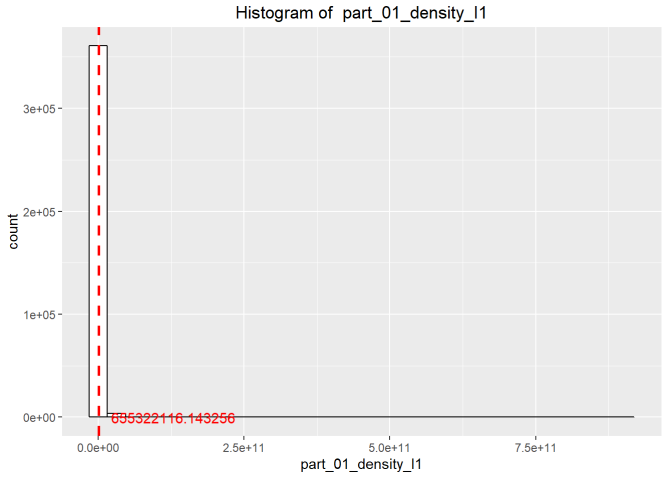

Histogram of part_01_density_FL
count
3e+05
2e+05
1e+05
0e+00
837471022320647
0e+00
1e+18
2e+18
3e+18
part_01_density_FL

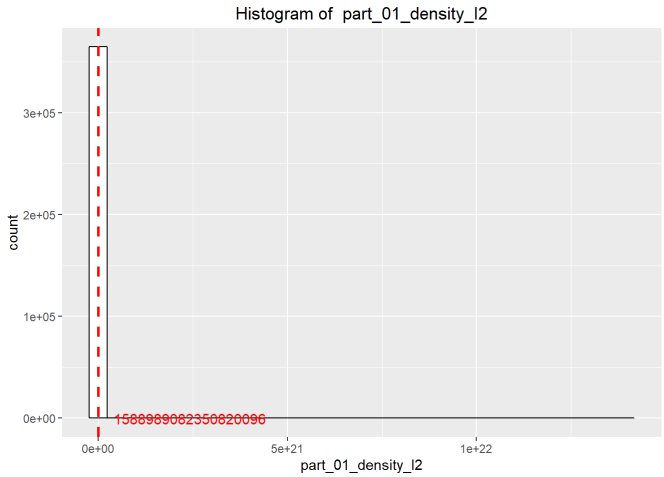

Histogram of part_01_density_O3_norm
count
150000
100000
50000
0
0.855763934656324
0
5
10
15
20
25
part_01_density_O3_norm

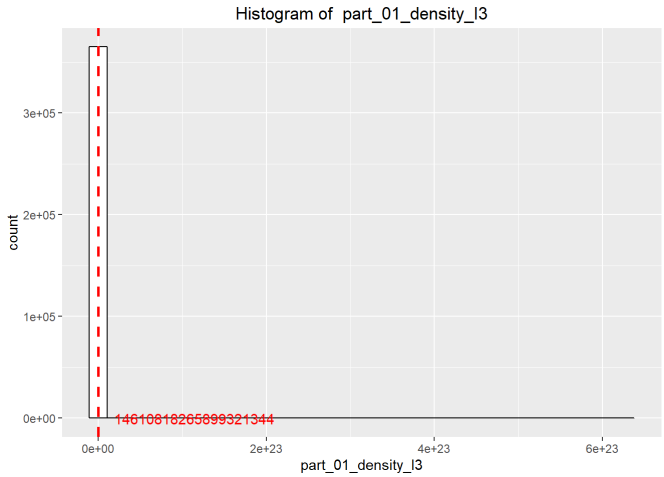

Histogram of part_01_density_O4_norm
count
3e+05
2e+05
1e+05
0e+00
0.123975346444276
0
10
20
part_01_density_O4_norm

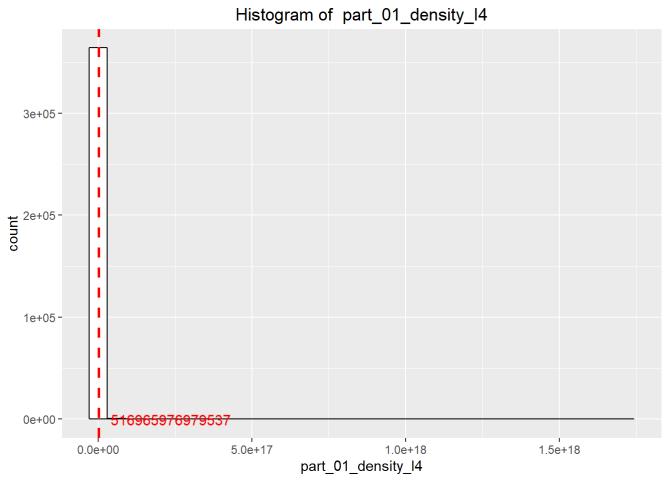

Histogram of part_01_density_O5_norm
count
3e+05
2e+05
1e+05
0e+00
0.00658515970160218
0
5
10
15
20
25
part_01_density_O5_norm

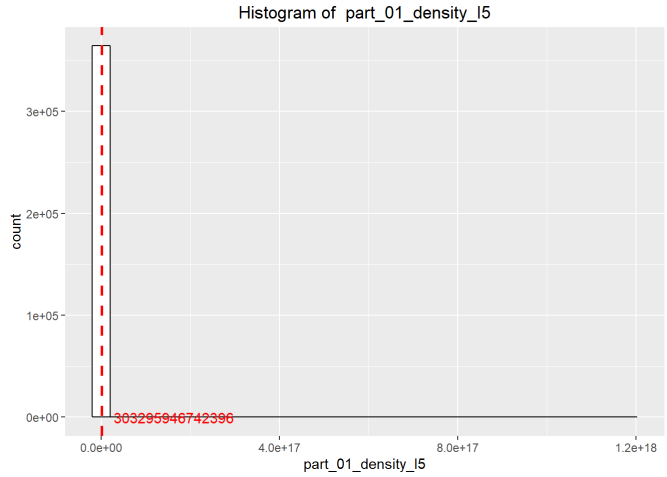

Histogram of part_01_density_FL_norm
count
3e+05
2e+05
1e+05
0e+00
0.517163637273
0
4000
8000
12000
part_01_density_FL_norm

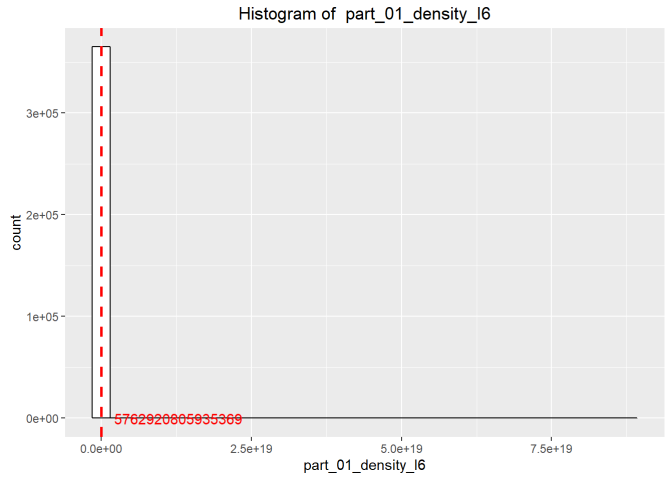

Histogram of part_01_density_l1
count
3e+05
2e+05
1e+05
0e+00
635322116.143256
0.0e+00
2.5e+11
5.0e+11
7.5e+11
part_01_density_l1

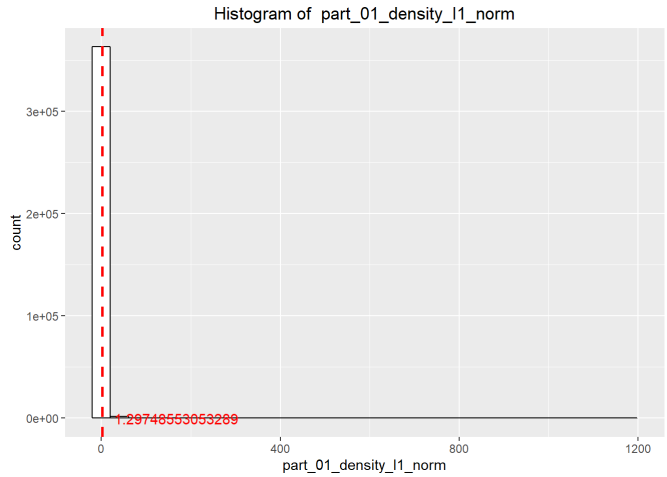

Histogram of part_01_density_l2
count
3e+05
2e+05
1e+05
0e+00
1588989082350820096
0e+00
5e+21
1e+22
part_01_density_l2

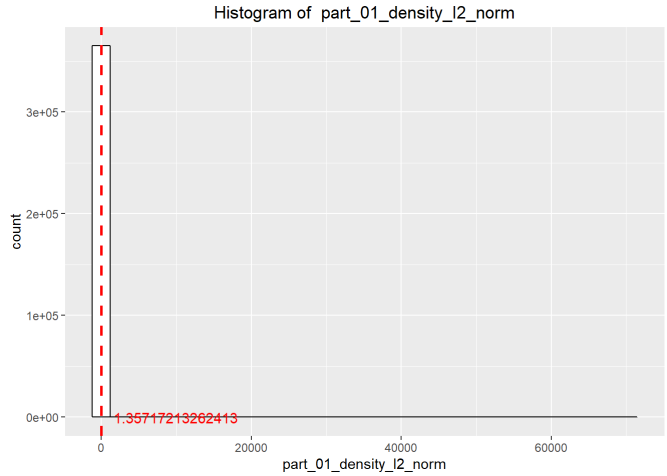

Histogram of part_01_density_l3
count
3e+05
2e+05
1e+05
0e+00
1461081826589932134 4
0e+00
2e+23
4e+23
6e+23
part_01_density_l3

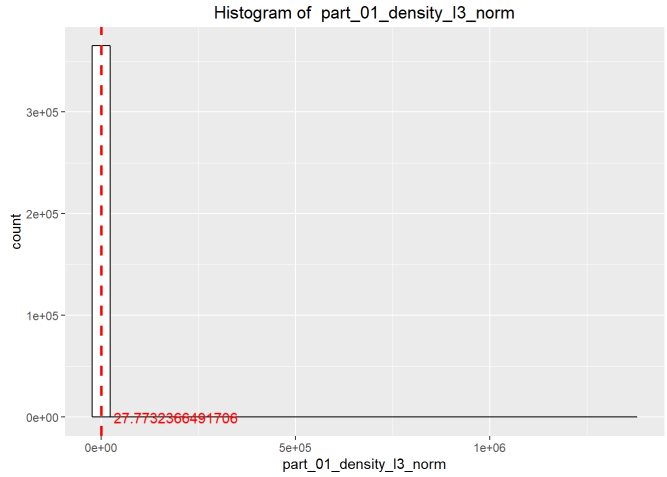

Histogram of part_01_density_l4
count
3e+05
2e+05
1e+05
0e+00
516965976979537
0.0e+00
5.0e+17
1.0e+18
1.5e+18
part_01_density_l4

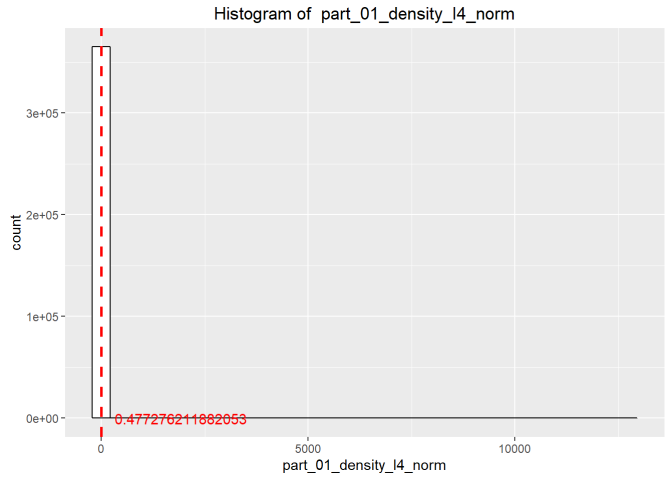

Histogram of part_01_density_l5
count
3e+05
2e+05
1e+05
0e+00
303295946742396
0.0e+00
4.0e+17
8.0e+17
1.2e+18
part_01_density_l5

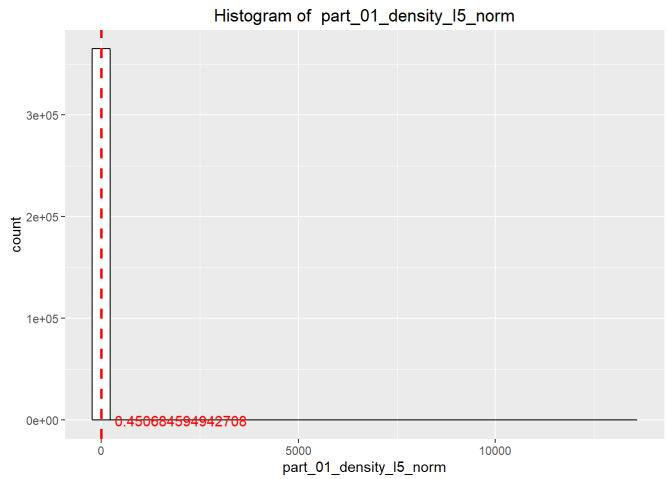

Histogram of part_01_density_l6
count
3e+05
2e+05
1e+05
0e+00
5762920865935369
0.0e+00
2.5e+19
5.0e+19
7.5e+19
part_01_density_l6

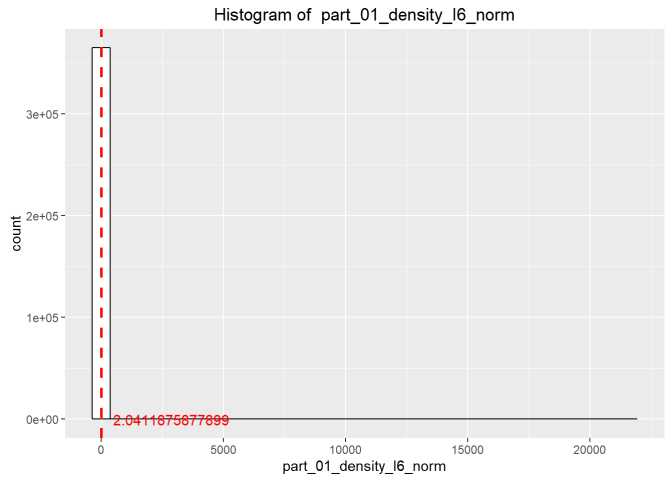

Histogram of part_01_density_l1_norm
count
3e+05
2e+05
1e+05
0e+00
1.29748553053289
0
400
800
1200
part_01_density_l1_norm

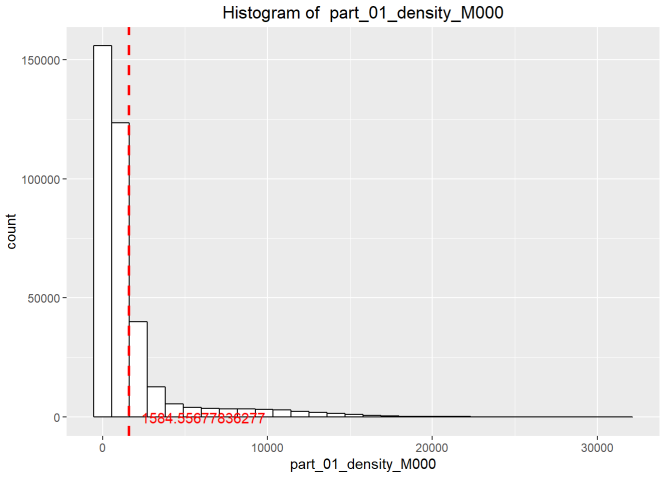

Histogram of part_01_density_l2_norm
count
3e+05
2e+05
1e+05
0e+00
1.35717213262413
0
20000
40000
60000
part_01_density_l2_norm

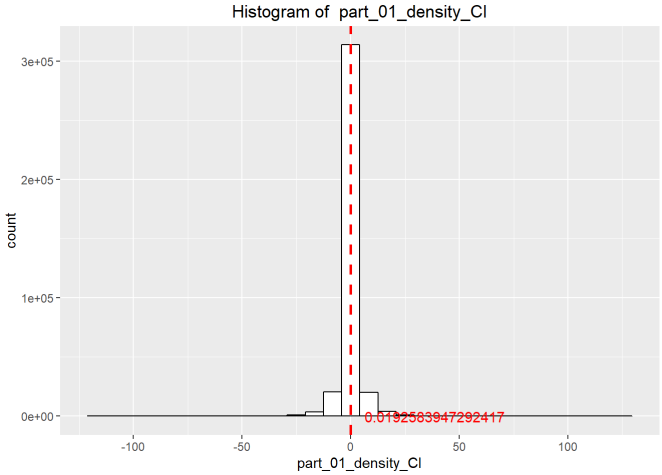

Histogram of part_01_density_l3_norm
count
3e+05
2e+05
1e+05
0e+00
27.7732366491706
0e+00
5e+05
1e+06
part_01_density_l3_norm

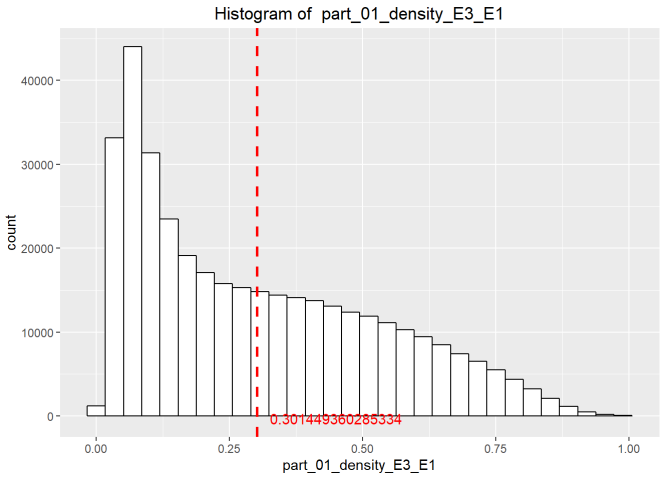

Histogram of part_01_density_l4_norm
count
3e+05
2e+05
1e+05
0e+00
0.477276211882053
0
5000
10000
part_01_density_l4_norm

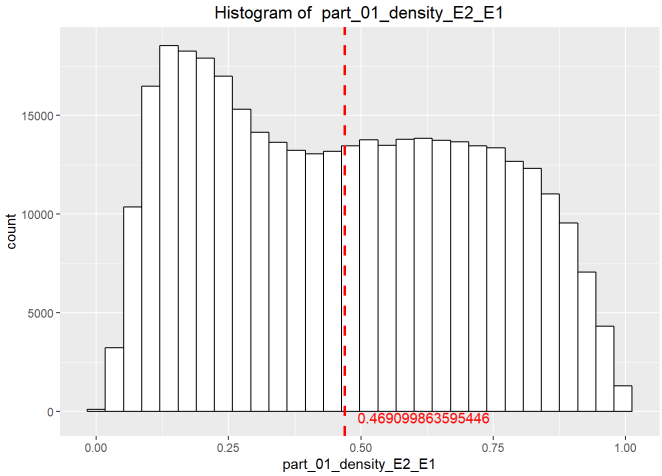

Histogram of part_01_density_l5_norm
count
3e+05
2e+05
1e+05
0e+00
0.450684594942708
0
5000
10000
part_01_density_l5_norm

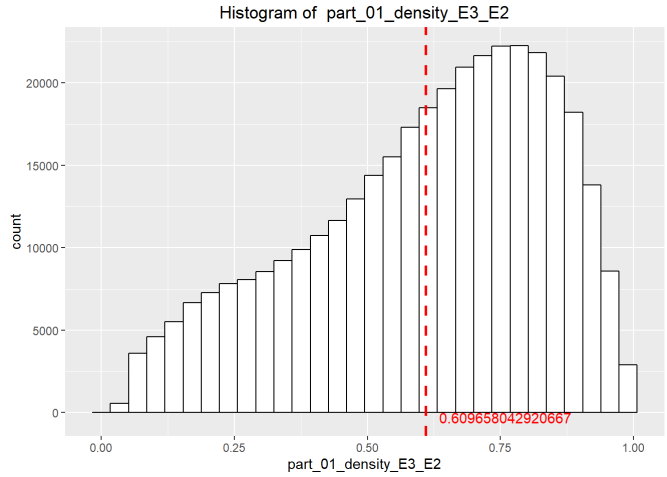

Histogram of part_01_density_l6_norm
count
3e+05
2e+05
1e+05
0e+00
2.0411875877899
0
5000
10000
15000
20000
part_01_density_l6_norm

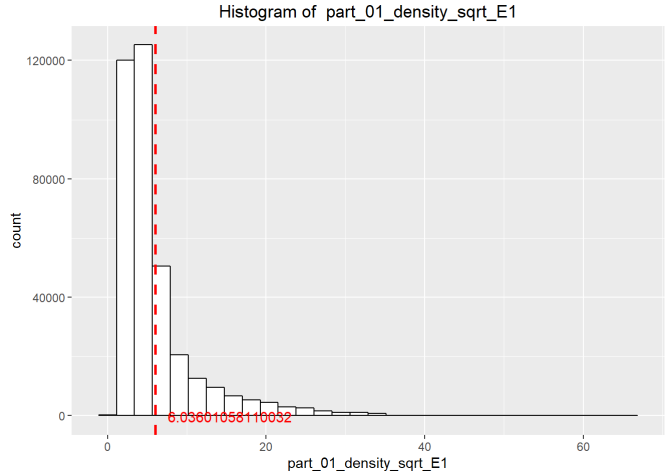

Histogram of part_01_density_M000
150000
100000
50000
0
count
1584.35677836277
0
10000
20000
30000
part_01_density_M000

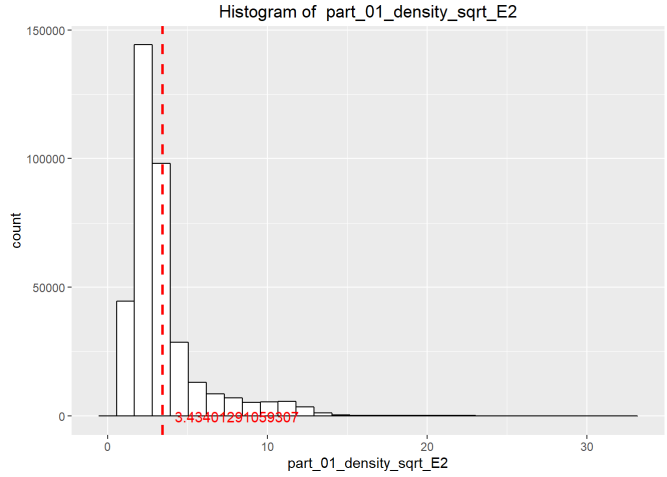

Histogram of part_01_density_CI
3e+05
2e+05
1e+05
0e+00
count
0.0192583947292417
-100
-50
0
50
100
part_01_density_CI

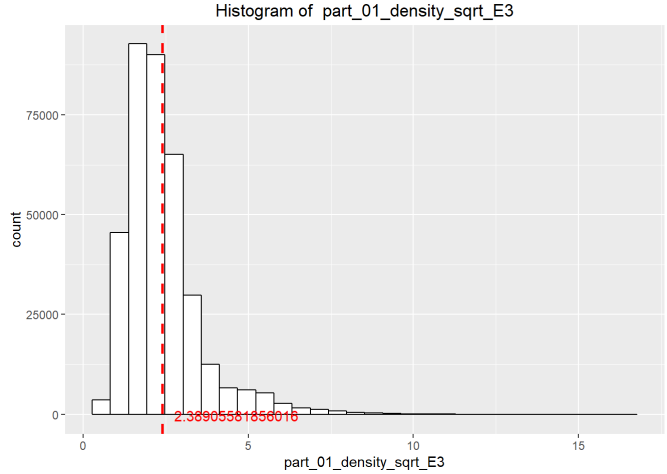

Histogram of part_01_density_E3_E1
40000
30000
20000
10000
0
count
0.301449360285334
0.00
0.25
0.50
0.75
1.00
part_01_density_E3_E1


Histogram of part_01_density_E2_E1
15000
10000
5000
0
count
0.469099863595446
0.00
0.25
0.50
0.75
1.00
part_01_density_E2_E1









Wykres interaktywny

```
ggplotly(qplot(local_res_atom_non_h_electron_sum, data=pdb_code_res_name))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Przewidywanie liczby elektronów i atomów na podstawie innych kolumn

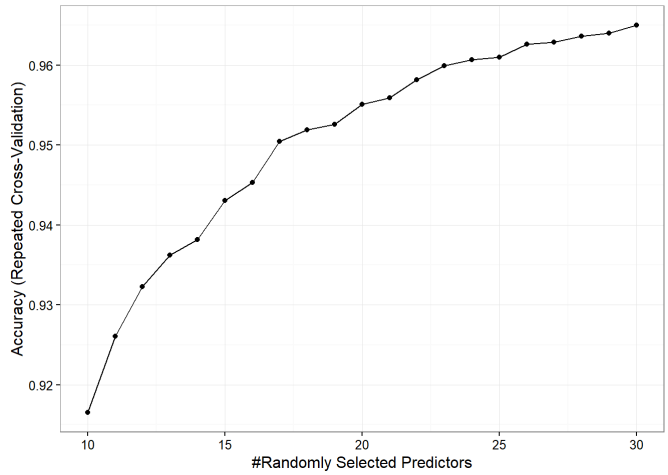
```
lm_data <- rawData
lm_data[is.na(lm_data)] <- 0
lm_data <- dplyr::select_if(lm_data, is.numeric)

lm_atom_model <- lm(local_res_atom_non_h_count ~ ., lm_data)
lm_atom_summary <- summary(lm_atom_model)

lm_electron_model <- lm(local_res_atom_non_h_electron_sum ~ ., lm_data)
lm_electron_summary <- summary(lm_electron_model)

pdb_code_res_name <- pdb_code_res_name[ , -which(names(pdb_code_res_name) %in% c("blob_coverage","res_coverage","
pdb_code","res_id","chain_id","skeleton_data","fc_col", "fo_col", "weight_col", "title"))]

pdb_code_res_name$res_name <- as.character(pdb_code_res_name$res_name)
pdb_code_res_name$res_name <- as.factor(pdb_code_res_name$res_name)
pdb_code_res_name[is.na(pdb_code_res_name)] <- -1000000
```

Miary dla liczby atomów:
R^2: 0.9999915
RMSEL 0.0390963

Miary dla liczby elektronów:
R^2: 0.9999877
RMSEL 0.3168246

Klasyfikator

```
## Random Forest
##
## 276303 samples
##    401 predictor
##     47 classes: '1PE', 'ACT', 'ACY', 'ADP', 'ATP', 'BR', 'CA', 'CD', 'CL', 'CLA', 'COA', 'CU', 'DMS', 'EDO', '
EPE', 'FAD', 'FE', 'FMN', 'FMT', 'GDP', 'GOL', 'HEC', 'HEM', 'IOD', 'K', 'MAN', 'MES', 'MG', 'MLY', 'MN', 'MPD',
'NAD', 'NAG', 'NAP', 'NDP', 'NI', 'NO3', 'PEG', 'PG4', 'PGE', 'PLP', 'PO4', 'SAH', 'SF4', 'SO4', 'TRS', 'ZN'
##
## No pre-processing
## Resampling: Cross-Validated (2 fold, repeated 3 times)
## Summary of sample sizes: 138151, 138152, 138148, 138155, 138148, 138155, ...
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa
##   10    0.9164987  0.9104989
##   11    0.9260666  0.9207726
##   12    0.9322905  0.9274545
##   13    0.9361860  0.9316358
##   14    0.9381850  0.9337804
##   15    0.9430553  0.9390061
##   16    0.9453149  0.9414305
##   17    0.9504288  0.9469149
##   18    0.9518898  0.9484816
##   19    0.9525991  0.9492433
##   20    0.9551097  0.9519343
##   21    0.9559228  0.9528072
##   22    0.9581884  0.9552360
##   23    0.9599112  0.9570827
##   24    0.9606736  0.9578992
##   25    0.9609788  0.9582267
##   26    0.9625870  0.9599505
##   27    0.9628958  0.9602815
##   28    0.9636402  0.9610791
##   29    0.9640166  0.9614827
##   30    0.9649913  0.9625272
##
## Accuracy was used to select the optimal model using  the largest value.
## The final value used for the model was mtry = 30.
```




```
##       Accuracy          Kappa  AccuracyLower  AccuracyUpper   AccuracyNull
##      0.9691461      0.9669763      0.9680088      0.9702536      0.1535964
## AccuracyPValue  McnemarPValue
##      0.0000000            NaN
```

	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision	Recall	F1	Prevalence	Detection Rate	Detection Prevalence	Balanced Accuracy
Class: 1PE	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0057994	0.0057994	0.0057994	1.0000000
Class: ACT	0.3843874	0.9957248	0.6689596	0.9862950	0.6689596	0.3843874	0.4882334	0.0219811	0.0084493	0.0126305	0.6900561
Class: ACY	0.0199005	0.9998691	0.4000000	0.9957201	0.4000000	0.0199005	0.0379147	0.0043658	0.0000869	0.0002172	0.5098848
Class: ADP	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0103607	0.0103607	0.0103607	1.0000000
Class: ATP	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0062338	0.0062338	0.0062338	1.0000000
Class: BR	0.9604520	0.9998580	0.9751434	0.9997706	0.9751434	0.9604520	0.9677419	0.0057668	0.0055387	0.0056799	0.9801550
Class: CA	1.0000000	0.9996545	0.9943279	1.0000000	0.9943279	1.0000000	0.9971559	0.0571140	0.0571140	0.0574398	0.9998272
Class: CD	0.9938272	0.9999343	0.9926017	0.9999452	0.9926017	0.9938272	0.9932141	0.0087968	0.0087425	0.0088077	0.9968807
Class: CL	1.0000000	0.9999884	0.9998278	1.0000000	0.9998278	1.0000000	0.9999139	0.0630437	0.0630437	0.0630546	0.9999942
Class: CLA	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0129888	0.0129888	0.0129888	1.0000000
Class: COA	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0059188	0.0059188	0.0059188	1.0000000
Class: CU	0.9319728	0.9998142	0.9699115	0.9995629	0.9699115	0.9319728	0.9505637	0.0063858	0.0059514	0.0061360	0.9658935
Class: DMS	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0180063	0.0180063	0.0180063	1.0000000
Class: EDO	0.9667791	0.9822692	0.8327744	0.9969206	0.8327744	0.9667791	0.8947874	0.0836890	0.0809088	0.0971557	0.9745242
Class: EPE	1.0000000	0.9999891	0.9979339	1.0000000	0.9979339	1.0000000	0.9989659	0.0052455	0.0052455	0.0052564	0.9999945
Class: FAD	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0123590	0.0123590	0.0123590	1.0000000
Class: FE	0.8900000	0.9998364	0.9595687	0.9995202	0.9595687	0.8900000	0.9234760	0.0043441	0.0038662	0.0040291	0.9449182
Class: FMN	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0056582	0.0056582	0.0056582	1.0000000
Class: FMT	1.0000000	0.9999672	0.9959016	1.0000000	0.9959016	1.0000000	0.9979466	0.0079171	0.0079171	0.0079497	0.9999836
Class: GDP	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0043115	0.0043115	0.0043115	1.0000000
Class: GOL	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.1102423	0.1102423	0.1102423	1.0000000
Class: HEC	0.3820459	0.9997489	0.8883495	0.9967782	0.8883495	0.3820459	0.5343066	0.0052021	0.0019874	0.0022372	0.6908974
Class: HEM	0.9917798	0.9966846	0.9036145	0.9997416	0.9036145	0.9917798	0.9456466	0.0303870	0.0301372	0.0333518	0.9942322
Class: IOD	1.0000000	0.9999779	0.9987350	1.0000000	0.9987350	1.0000000	0.9993671	0.0171483	0.0171483	0.0171700	0.9999890
Class: K	0.9940476	0.9999890	0.9991453	0.9999230	0.9991453	0.9940476	0.9965899	0.0127716	0.0126956	0.0127065	0.9970183
Class: MAN	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0077108	0.0077108	0.0077108	1.0000000
Class: MES	0.9985163	1.0000000	1.0000000	0.9999891	1.0000000	0.9985163	0.9992576	0.0073198	0.0073089	0.0073089	0.9992582
Class: MG	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0401177	0.0401177	0.0401177	1.0000000

	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision	Recall	F1	Prevalence	Detection Rate	Detection Prevalence	Balanced Accuracy
Class: MLY	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0095244	0.0095244	0.0095244	1.0000000
Class: MN	0.9810066	0.9996924	0.9736098	0.9997803	0.9736098	0.9810066	0.9772942	0.0114358	0.0112186	0.0115227	0.9903495
Class: MPD	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0087425	0.0087425	0.0087425	1.0000000
Class: NAD	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0122178	0.0122178	0.0122178	1.0000000
Class: NAG	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0715690	0.0715690	0.0715690	1.0000000
Class: NAP	0.8436073	0.9968312	0.7188716	0.9984953	0.7188716	0.8436073	0.7762605	0.0095136	0.0080257	0.0111643	0.9202193
Class: NDP	0.4505703	0.9985036	0.6336898	0.9968486	0.6336898	0.4505703	0.5266667	0.0057125	0.0025739	0.0040617	0.7245370
Class: NI	0.8753056	0.9998582	0.9649596	0.9994439	0.9649596	0.8753056	0.9179487	0.0044418	0.0038880	0.0040291	0.9375819
Class: NO3	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0043332	0.0043332	0.0043332	1.0000000
Class: PEG	1.0000000	0.9999780	0.9983974	1.0000000	0.9983974	1.0000000	0.9991981	0.0135319	0.0135319	0.0135536	0.9999890
Class: PG4	0.9971098	1.0000000	1.0000000	0.9999781	1.0000000	0.9971098	0.9985528	0.0075153	0.0074936	0.0074936	0.9985549
Class: PGE	1.0000000	0.9999782	0.9958159	1.0000000	0.9958159	1.0000000	0.9979036	0.0051695	0.0051695	0.0051912	0.9999891
Class: PLP	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0043224	0.0043224	0.0043224	1.0000000
Class: PO4	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0301046	0.0301046	0.0301046	1.0000000
Class: SAH	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0043007	0.0043007	0.0043007	1.0000000
Class: SF4	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.0044636	0.0044636	0.0044636	1.0000000
Class: SO4	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.1535964	0.1535964	0.1535964	1.0000000
Class: TRS	0.9951691	1.0000000	1.0000000	0.9999782	1.0000000	0.9951691	0.9975787	0.0044961	0.0044744	0.0044744	0.9975845
Class: ZN	0.9985876	0.9992080	0.9862495	0.9999196	0.9862495	0.9985876	0.9923802	0.0538233	0.0537473	0.0544967	0.9988978