Homework6

YikunHan42

目录

1 导入第三方库

```
library(tidyverse)
```

```
## Warning: 程辑包'tidyverse'是用R版本4.1.3 来建造的
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(mlr)
```

```
## Warning: 程辑包'mlr'是用R版本4.1.3 来建造的
```

```
## 载入需要的程辑包: ParamHelpers
```

```
## Warning: 程辑包'ParamHelpers'是用R版本4.1.3 来建造的
```

```
## Warning message: 'mlr' is in 'maintenance-only' mode since July 2019.
## Future development will only happen in 'mlr3'
## (<https://mlr3.mlr-org.com>). Due to the focus on 'mlr3' there might be
## uncaught bugs meanwhile in {mlr} - please consider switching.
```

```
library(rpart)
```

```
## Warning: 程辑包'rpart'是用R版本4.1.3 来建造的
```

```
library(randomForest)
```

```
## Warning: 程辑包'randomForest'是用R版本4.1.3 来建造的
```

```
## randomForest 4.7-1.1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## 载入程辑包: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(parallel)
library(parallelMap)
```

```
## Warning: 程辑包'parallelMap'是用R版本4.1.3 来建造的
```

```
library(car)
```

```
## Warning: 程辑包'car'是用R版本4.1.3 来建造的
```

```
## 载入需要的程辑包: carData
```

```
## Warning: 程辑包'carData'是用R版本4.1.3 来建造的
```

```
##
## 载入程辑包: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(GGally)
```

```
## Warning: 程辑包'GGally'是用R版本4.1.3 来建造的
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

2 读入数据

```
file <- "D:/Study/DSBI/Task7/solubility_data.csv"
solubility_data <- read.csv(file)
```

3 探索性数据分析

```
summary(solubility_data)
```

```
##      FP001            FP002            FP003            FP004
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :1.0000   Median :0.0000   Median :1.0000
##  Mean   :0.4932   Mean   :0.5394   Mean   :0.4364   Mean   :0.5846
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP005            FP006            FP007            FP008
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :1.0000   Median :0.0000   Median :0.0000   Median :0.000
##  Mean   :0.5794   Mean   :0.4006   Mean   :0.3638   Mean   :0.326
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##      FP009            FP010            FP011            FP012
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.2797   Mean   :0.1788   Mean   :0.2145   Mean   :0.1767
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP013            FP014            FP015            FP016
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :1.0000   Median :0.0000
##  Mean   :0.1661   Mean   :0.1609   Mean   :0.8601   Mean   :0.1462
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP017            FP018            FP019           FP020
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.000   Median :0.0000
##  Mean   :0.1441   Mean   :0.1314   Mean   :0.122   Mean   :0.1199
```

```
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##      FP021            FP022            FP023           FP024
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.000   Median :0.0000
##  Mean   :0.1209   Mean   :0.1041   Mean   :0.123   Mean   :0.1125
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##      FP025            FP026             FP027             FP028
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.1157   Mean   :0.08412   Mean   :0.09779   Mean   :0.1062
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.0000
##      FP029           FP030            FP031            FP032
##  Min.   :0.000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.102   Mean   :0.09359   Mean   :0.08938   Mean   :0.07361
##  3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP033            FP034             FP035             FP036
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0694   Mean   :0.07992   Mean   :0.07256   Mean   :0.07571
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP037             FP038             FP039             FP040
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
```

```
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.07045   Mean   :0.08622   Mean   :0.07466   Mean   :0.06835
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP041             FP042             FP043             FP044
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.06309   Mean   :0.05678   Mean   :0.06625   Mean   :0.05994
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP045             FP046            FP047            FP048
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.000   Median :0.0000
##  Mean   :0.05573   Mean   :0.3155   Mean   :0.266   Mean   :0.1241
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##      FP049           FP050             FP051             FP052
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.000   Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   :0.122   Mean   :0.1125   Mean   :0.1094   Mean   :0.09148
##  3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##      FP053             FP054             FP055             FP056
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.09359   Mean   :0.07571   Mean   :0.05363   Mean   :0.06519
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP057            FP058            FP059             FP060
```

```
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.1199   Mean   :0.1136   Mean   :0.05468   Mean   :0.4816
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##      FP061            FP062            FP063            FP064
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.4469   Mean   :0.4374   Mean   :0.4259   Mean   :0.4164
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP065            FP066            FP067            FP068
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :1.0000   Median :0.0000   Median :0.0000
##  Mean   :0.5931   Mean   :0.6099   Mean   :0.3796   Mean   :0.3617
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP069            FP070            FP071           FP072
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.000   Median :1.0000
##  Mean   :0.3617   Mean   :0.3554   Mean   :0.327   Mean   :0.6583
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##      FP073            FP074            FP075            FP076
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.3102   Mean   :0.3249   Mean   :0.3386   Mean   :0.3281
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
```

```
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP077            FP078            FP079            FP080
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :1.0000   Median :0.0000
##  Mean   :0.3207   Mean   :0.3039   Mean   :0.6898   Mean   :0.3028
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP081            FP082           FP083            FP084
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :1.000   Median :0.0000   Median :0.000
##  Mean   :0.2787   Mean   :0.714   Mean   :0.2734   Mean   :0.286
##  3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.000
##      FP085            FP086            FP087            FP088
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :1.0000   Median :0.0000
##  Mean   :0.2555   Mean   :0.2692   Mean   :0.7266   Mean   :0.2629
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP089            FP090            FP091           FP092
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000
##  Median :0.0000   Median :0.0000   Median :0.000   Median :0.000
##  Mean   :0.2471   Mean   :0.2492   Mean   :0.225   Mean   :0.244
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.000
##      FP093           FP094            FP095            FP096
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.000   Median :0.0000   Median :0.0000   Median :0.0000
```

```
##  Mean   :0.244   Mean   :0.2313   Mean   :0.2198   Mean   :0.2177
##  3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP097            FP098            FP099            FP100
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.2355   Mean   :0.2376   Mean   :0.2271   Mean   :0.2313
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP101            FP102            FP103            FP104
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.2366   Mean   :0.2019   Mean   :0.2187   Mean   :0.2229
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP105            FP106            FP107            FP108
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.000
##  Mean   :0.2156   Mean   :0.1914   Mean   :0.2114   Mean   :0.205
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##      FP109            FP110            FP111            FP112
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.1767   Mean   :0.2061   Mean   :0.1966   Mean   :0.1945
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP113            FP114            FP115            FP116
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
```

```
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.1956   Mean   :0.1556   Mean   :0.1788   Mean   :0.1924
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP117            FP118            FP119            FP120
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.000   Median :0.0000
##  Mean   :0.1788   Mean   :0.1924   Mean   :0.163   Mean   :0.1661
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##      FP121            FP122            FP123            FP124
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.000   Median :0.0000   Median :0.0000
##  Mean   :0.1399   Mean   :0.164   Mean   :0.1672   Mean   :0.1619
##  3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##      FP125            FP126            FP127            FP128
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.1556   Mean   :0.1483   Mean   :0.1399   Mean   :0.1483
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP129            FP130            FP131            FP132
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.1388   Mean   :0.1052   Mean   :0.1262   Mean   :0.1251
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
```

```
##      FP133            FP134            FP135            FP136
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.1262   Mean   :0.1272   Mean   :0.1262   Mean   :0.1209
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP137            FP138            FP139             FP140
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.1157   Mean   :0.1115   Mean   :0.08202   Mean   :0.1115
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##      FP141            FP142            FP143             FP144
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.1167   Mean   :0.1094   Mean   :0.08097   Mean   :0.1041
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##      FP145            FP146            FP147             FP148
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.0000   Median :0.000   Median :0.0000   Median :0.00000
##  Mean   :0.1041   Mean   :0.103   Mean   :0.1052   Mean   :0.08728
##  3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.00000
##      FP149             FP150             FP151             FP152
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.09043   Mean   :0.07886   Mean   :0.05573   Mean   :0.08202
```

```
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP153             FP154             FP155             FP156
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.07781   Mean   :0.03785   Mean   :0.0694   Mean   :0.07045
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000
##      FP157             FP158             FP159             FP160
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.06204   Mean   :0.05363   Mean   :0.07045   Mean   :0.06835
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP161             FP162            FP163            FP164
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.0000   Median :1.0000
##  Mean   :0.06625   Mean   :0.4953   Mean   :0.4763   Mean   :0.6278
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP165            FP166            FP167            FP168
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :1.0000
##  Mean   :0.3491   Mean   :0.3312   Mean   :0.3281   Mean   :0.6656
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      FP169            FP170            FP171            FP172
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
```

```
##  Median :0.0000   Median :0.000   Median :0.0000   Median :0.0000
##  Mean   :0.1861   Mean   :0.184   Mean   :0.1693   Mean   :0.1514
##  3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##      FP173            FP174            FP175            FP176
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.000   Median :0.0000   Median :0.0000   Median :0.000
##  Mean   :0.142   Mean   :0.1304   Mean   :0.1346   Mean   :0.122
##  3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##      FP177            FP178            FP179            FP180
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.1209   Mean   :0.1209   Mean   :0.09779   Mean   :0.1073
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##      FP181            FP182            FP183            FP184
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.09359   Mean   :0.09884   Mean   :0.07571   Mean   :0.08412
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP185            FP186            FP187            FP188
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.08517   Mean   :0.07676   Mean   :0.07256   Mean   :0.06835
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP189            FP190            FP191            FP192
```

```
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.07676   Mean   :0.07256   Mean   :0.07045   Mean   :0.06099
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP193             FP194             FP195             FP196
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.06204   Mean   :0.05889   Mean   :0.06099   Mean   :0.05678
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP197             FP198             FP199             FP200
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.05258   Mean   :0.05678   Mean   :0.04732   Mean   :0.04942
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##      FP201             FP202            FP203             FP204
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   :0.05258   Mean   :0.2576   Mean   :0.1146   Mean   :0.09884
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##      FP205             FP206             FP207             FP208
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.07781   Mean   :0.05994   Mean   :0.05678   Mean   :0.1125
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
```

```
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.0000
##    MolWeight        NumAtoms      NumNonHAtoms      NumBonds
##  Min.   :3.852   Min.   :1.792   Min.   :1.099   Min.   :1.609
##  1st Qu.:4.817   1st Qu.:2.890   1st Qu.:2.197   1st Qu.:2.890
##  Median :5.194   Median :3.135   Median :2.565   Median :3.178
##  Mean   :5.199   Mean   :3.174   Mean   :2.549   Mean   :3.176
##  3rd Qu.:5.581   3rd Qu.:3.466   3rd Qu.:2.890   3rd Qu.:3.481
##  Max.   :6.503   Max.   :4.554   Max.   :3.871   Max.   :4.585
##   NumNonHBonds     NumMultBonds     NumRotBonds      NumDblBonds
##  Min.   :0.7435   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:2.7592   1st Qu.:0.7988   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :3.3514   Median :2.9448   Median :1.0986   Median :0.5671
##  Mean   :3.3623   Mean   :2.5791   Mean   :0.9256   Mean   :0.3981
##  3rd Qu.:4.0099   3rd Qu.:4.0237   3rd Qu.:1.4979   3rd Qu.:0.8045
##  Max.   :5.9770   Max.   :6.7030   Max.   :2.8332   Max.   :1.1880
##  NumAromaticBonds  NumHydrogen      NumCarbon       NumNitrogen
##  Min.   :0.000    Min.   :0.000   Min.   :0.7705   Min.   :0.0000
##  1st Qu.:0.000    1st Qu.:2.887   1st Qu.:2.6426   1st Qu.:0.0000
##  Median :1.946    Median :3.691   Median :3.3175   Median :0.0000
##  Mean   :1.287    Mean   :3.696   Mean   :3.3240   Mean   :0.2308
##  3rd Qu.:1.946    3rd Qu.:4.465   3rd Qu.:3.8622   3rd Qu.:0.4568
##  Max.   :3.258    Max.   :7.314   Max.   :6.2678   Max.   :0.7079
##    NumOxygen        NumSulfer        NumChlorine        NumHalogen
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.6931   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.7470   Mean   :0.04975   Mean   :0.09098   Mean   :0.1201
##  3rd Qu.:1.0986   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.3750
##  Max.   :2.6391   Max.   :0.48000   Max.   :0.49587   Max.   :0.4959
##     NumRings      HydrophilicFactor  SurfaceArea1     SurfaceArea2
##  Min.   :0.0000   Min.   :-2.8413   Min.   : 0.000   Min.   : 0.000
##  1st Qu.:0.0000   1st Qu.:-1.2510   1st Qu.: 3.837   1st Qu.: 4.159
##  Median :0.6931   Median :-0.3630   Median : 7.258   Median : 7.760
```

```
##  Mean   :0.7341   Mean   :-0.4528   Mean   : 6.708   Mean   : 7.081
##  3rd Qu.:1.0986   3rd Qu.: 0.2799   3rd Qu.: 9.854   3rd Qu.:10.500
##  Max.   :2.0794   Max.   : 3.5338   Max.   :23.020   Max.   :23.020
##    Solubility
##  Min.   :-11.620
##  1st Qu.: -3.955
##  Median : -2.510
##  Mean   : -2.719
##  3rd Qu.: -1.360
##  Max.   :  1.580
```

```
str(solubility_data)
```

```
## 'data.frame':    951 obs. of  229 variables:
##  $ FP001            : int  0 0 1 0 0 1 0 1 1 1 ...
##  $ FP002            : int  1 1 1 0 0 0 1 0 0 1 ...
##  $ FP003            : int  0 0 1 1 1 1 0 1 1 1 ...
##  $ FP004            : int  0 1 1 0 1 1 1 1 1 1 ...
##  $ FP005            : int  1 1 1 0 1 0 1 0 0 1 ...
##  $ FP006            : int  0 1 0 0 1 0 0 0 1 1 ...
##  $ FP007            : int  0 1 0 1 0 0 0 1 1 1 ...
##  $ FP008            : int  1 1 1 0 0 0 1 0 0 0 ...
##  $ FP009            : int  0 0 0 0 1 1 1 0 1 0 ...
##  $ FP010            : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ FP011            : int  0 1 0 0 0 0 0 0 1 0 ...
##  $ FP012            : int  0 0 0 0 0 1 0 1 0 0 ...
##  $ FP013            : int  0 0 0 0 1 0 1 0 0 0 ...
##  $ FP014            : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ FP015            : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ FP016            : int  0 1 0 0 1 1 0 1 0 0 ...
##  $ FP017            : int  0 0 1 1 0 0 0 0 1 1 ...
##  $ FP018            : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ FP019            : int  1 0 0 0 1 0 1 0 0 0 ...
##  $ FP020            : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ FP021           : int  0 0 0 0 0 1 0 0 1 0 ...
##  $ FP022           : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ FP023           : int  0 0 0 1 0 0 0 0 1 0 ...
##  $ FP024           : int  1 0 0 0 1 0 0 0 0 0 ...
##  $ FP025           : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ FP026           : int  1 0 0 0 0 0 1 0 0 0 ...
##  $ FP027           : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ FP028           : int  0 1 0 0 0 0 0 0 1 1 ...
##  $ FP029           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FP030           : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ FP031           : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ FP032           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FP033           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FP034           : int  0 0 0 0 1 0 0 0 0 1 ...
##  $ FP035           : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ FP036           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FP037           : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ FP038           : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ FP039           : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ FP040           : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ FP041           : int  0 0 0 1 0 0 0 0 1 0 ...
##  $ FP042           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FP043           : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ FP044           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FP045           : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ FP046           : int  0 1 0 0 0 0 1 0 0 1 ...
##  $ FP047           : int  0 1 1 0 0 0 1 0 0 0 ...
##  $ FP048           : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ FP049           : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ FP050           : int  0 0 0 0 0 0 0 1 0 1 ...
##  $ FP051           : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ FP052           : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ FP053           : int  0 0 0 0 0 0 1 0 0 0 ...
```

```
##  $ FP054           : int  0 0 0 1 0 0 0 0 1 1 ...
##  $ FP055           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FP056           : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ FP057           : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ FP058           : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ FP059           : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ FP060           : int  0 1 1 0 0 0 0 1 1 0 ...
##  $ FP061           : int  0 0 1 0 0 0 0 1 1 0 ...
##  $ FP062           : int  0 0 1 0 0 1 0 1 1 1 ...
##  $ FP063           : int  1 1 0 0 1 1 1 0 0 1 ...
##  $ FP064           : int  0 1 1 0 1 1 0 1 0 0 ...
##  $ FP065           : int  1 1 0 0 1 0 1 0 1 1 ...
##  $ FP066           : int  1 0 1 1 1 1 1 1 1 1 ...
##  $ FP067           : int  1 1 0 0 1 1 1 0 0 1 ...
##  $ FP068           : int  0 1 0 0 1 1 1 0 0 1 ...
##  $ FP069           : int  1 0 1 1 1 1 0 1 1 0 ...
##  $ FP070           : int  1 1 0 1 0 0 1 0 1 0 ...
##  $ FP071           : int  0 0 0 0 0 0 1 0 1 1 ...
##  $ FP072           : int  0 1 1 0 0 1 0 1 1 1 ...
##  $ FP073           : int  0 1 1 0 0 0 0 0 1 0 ...
##  $ FP074           : int  0 1 0 0 0 0 0 0 1 0 ...
##  $ FP075           : int  0 1 0 0 1 1 1 0 0 1 ...
##  $ FP076           : int  1 1 0 0 0 0 1 0 1 1 ...
##  $ FP077           : int  0 1 0 1 0 0 0 1 1 1 ...
##  $ FP078           : int  0 1 0 0 0 0 0 0 1 0 ...
##  $ FP079           : int  1 1 1 1 1 0 1 0 1 1 ...
##  $ FP080           : int  0 1 0 0 1 1 1 1 0 0 ...
##  $ FP081           : int  0 0 1 1 0 0 0 1 1 1 ...
##  $ FP082           : int  1 1 1 0 1 1 1 0 1 1 ...
##  $ FP083           : int  0 0 0 0 1 0 0 0 0 1 ...
##  $ FP084           : int  1 1 0 0 1 0 1 0 0 0 ...
##  $ FP085           : int  0 1 0 0 0 0 1 0 0 0 ...
##  $ FP086           : int  0 0 0 1 1 0 0 1 1 1 ...
```

```
##  $ FP087            : int  1 1 1 1 1 0 1 0 1 1 ...
##  $ FP088            : int  0 1 0 0 0 0 0 1 1 0 ...
##  $ FP089            : int  1 1 0 0 0 0 1 0 0 0 ...
##  $ FP090            : int  0 1 0 1 0 0 0 1 1 1 ...
##  $ FP091            : int  1 1 0 0 1 0 1 0 0 1 ...
##  $ FP092            : int  0 0 0 0 1 1 1 0 1 0 ...
##  $ FP093            : int  0 1 0 1 0 0 0 1 1 1 ...
##  $ FP094            : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ FP095            : int  0 0 0 0 0 0 0 0 1 1 ...
##  $ FP096            : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ FP097            : int  1 1 0 0 0 0 1 0 1 0 ...
##  $ FP098            : int  0 0 1 0 0 0 0 1 0 0 ...
##  $ FP099            : int  0 0 0 0 0 0 0 0 1 0 ...
##   [list output truncated]
```

```
ncol(solubility_data)
```

```
## [1] 229
```

4 数据可视化

4.1 箱线图 + 小提琴图

```
p <- ggplot(data = solubility_data, mapping = aes(x = 0, y = Solubility), fill = attrib
p + geom_boxplot(width = 1,position = position_dodge(0.9), color = " green") + geom_vio
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥'出错: <e5>代替了dot
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥'出错: <8f>代替了dot
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <8a>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <9b>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <be>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <8f>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <8a>代替了dot
```

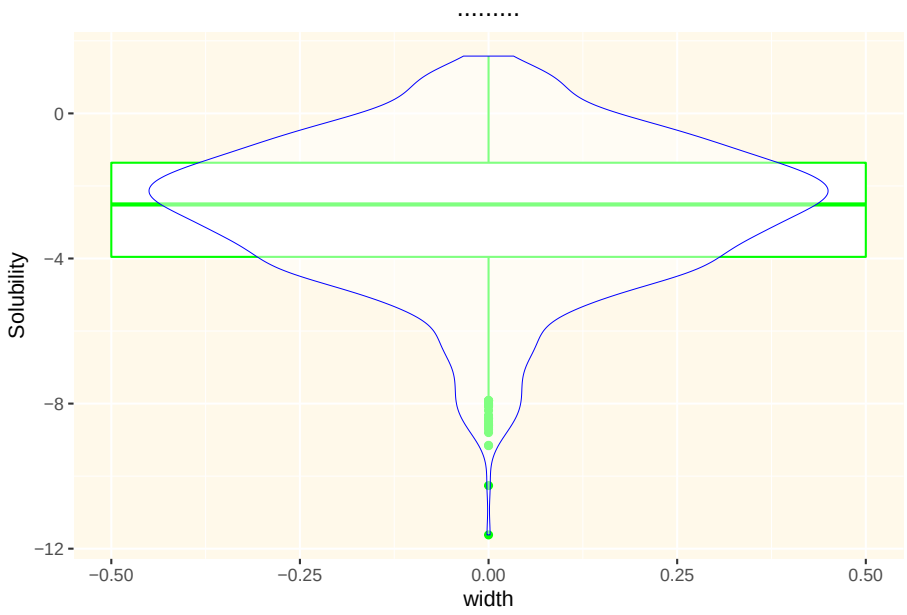
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <9b>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <be>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8f>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8a>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <9b>代替了dot
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <be>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <8f>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <8a>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <9b>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <be>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊豺姑鍥 '出错: <8f>代替了dot
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8a>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <9b>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <be>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8f>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8a>代替了dot
```

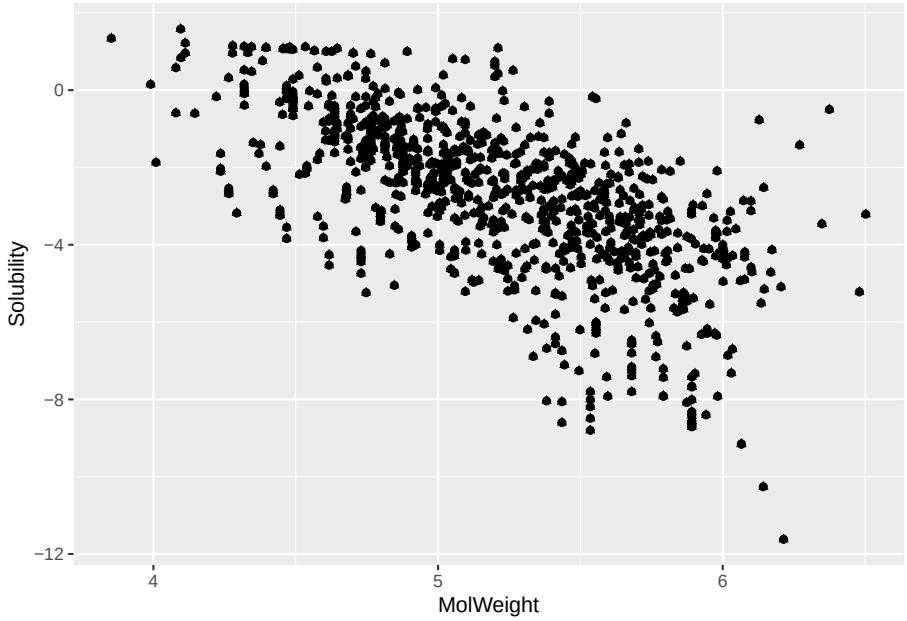
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <9b>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <be>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8f>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8a>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <9b>代替了dot
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <be>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8f>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <8a>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <a0>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <e5>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <9b>代替了dot

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## 'mbcsToSbcs'里转换'鏊犴姑鍥 '出错: <be>代替了dot
```

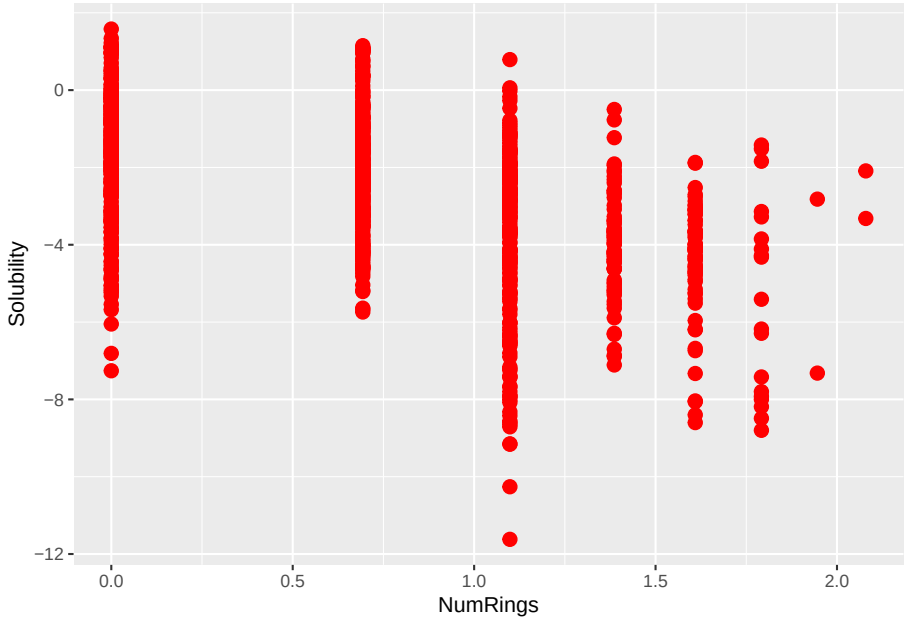


4.2 散点图

```
ggplot(data = solubility_data, aes(x = MolWeight, y = Solubility)) +
  geom_point()+
  geom_point(shape = 17)
```



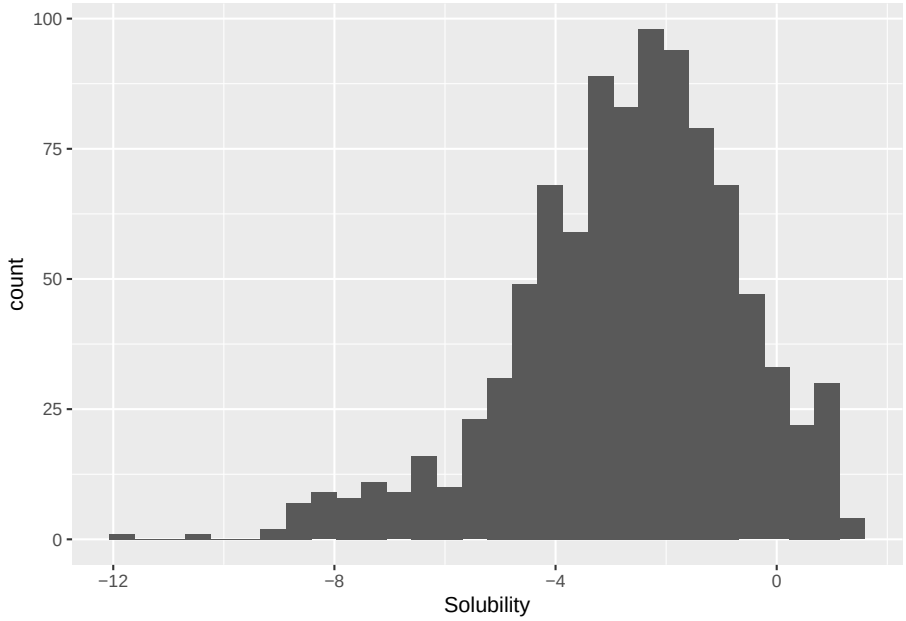

```
ggplot(data = solubility_data, aes(x = NumRings, y = Solubility)) +
  geom_point()+
  geom_point(size = 3, color = "red")
```



4.3 柱状图

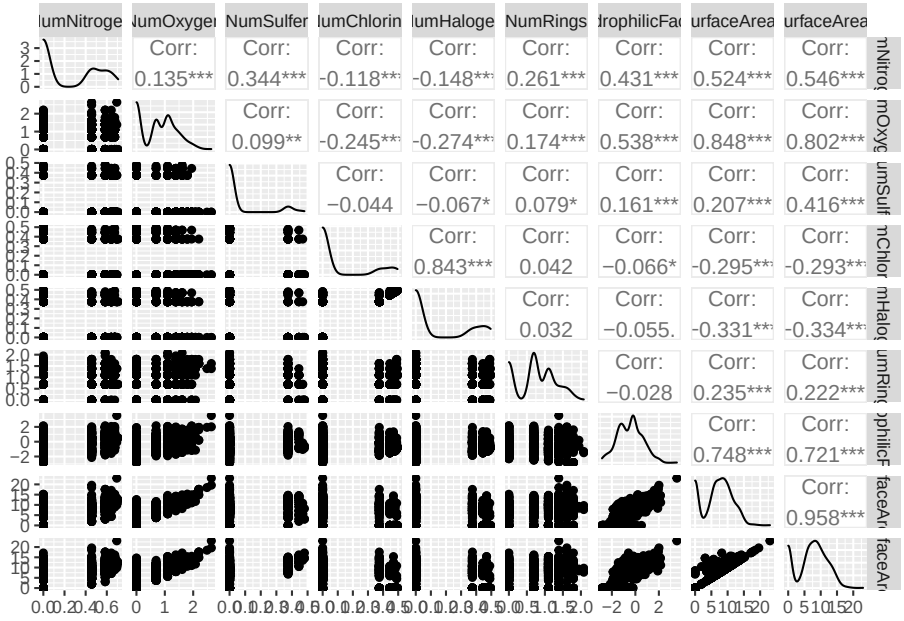
```
ggplot(data = solubility_data, aes(x = Solubility)) +
  geom_histogram()
```

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



4.4 散点图矩阵

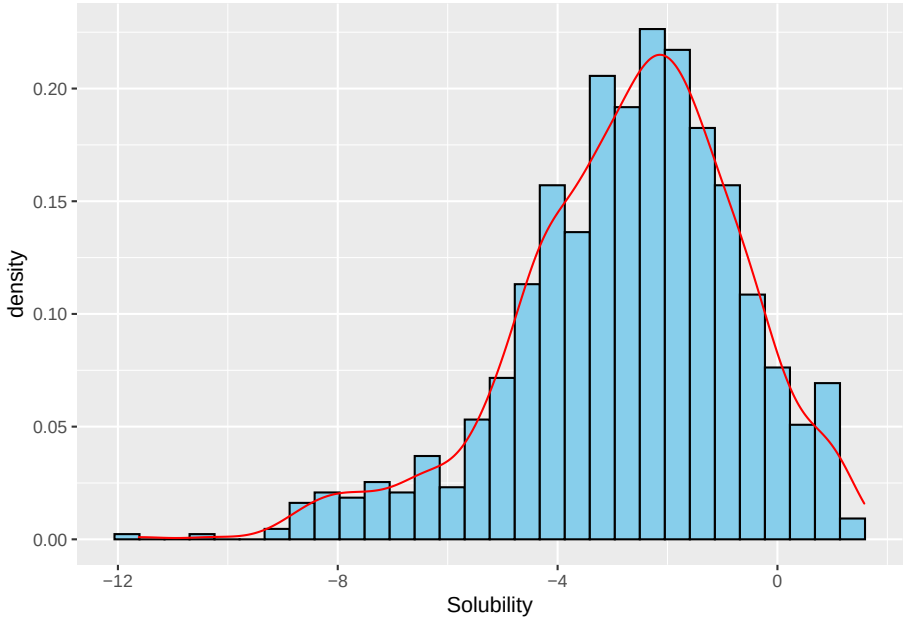
```
ggpairs(solubility_data[,220:228], showStrips = F)
```



4.5 核密度

```
p <- ggplot(data=solubility_data,aes(x=Solubility))
p + geom_histogram(aes(y=..density..),fill="skyblue",color="black")+
  geom_density(color="red")
```

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



5 回归分析

5.1 定义回归任务

```
## define regression task
task <- makeRegrTask(data = solubility_data, target = "Solubility")
```

5.2 决策树

```
## Decision Tree -----
treeLearner <- makeLearner("regr.rpart")
treeParamSpace <- makeParamSet( makeIntegerParam("minsplit", lower = 5, upper = 20),
                                makeIntegerParam("minbucket", lower = 3, upper = 10),
                                makeNumericParam("cp", lower = 0.01, upper = 0.1),
                                makeIntegerParam("maxdepth", lower = 3, upper = 10))
```

```
randSearch <- makeTuneControlRandom(maxit = 100)
cvForTuning <- makeResampleDesc("CV", iters = 5)
library(parallel); library(parallelMap)
parallelStartSocket(cpus = detectCores())
```

```
## Starting parallelization in mode=socket with cpus=12.
```

```
tunedTreePars <- tuneParams(treeLearner, task = task,
                            resampling = cvForTuning,
                            par.set = treeParamSpace, control = randSearch)
```

```
## [Tune] Started tuning learner regr.rpart for parameter set:
```

```
##              Type len Def      Constr Req Tunable Trafo
## minsplit  integer   -   -     5 to 20   -    TRUE     -
## minbucket integer   -   -     3 to 10   -    TRUE     -
## cp        numeric   -   - 0.01 to 0.1   -    TRUE     -
## maxdepth  integer   -   -     3 to 10   -    TRUE     -
```

```
## With control class: TuneControlRandom
```

```
## Imputation value: Inf
```

```
## Exporting objects to slaves for mode socket: .mlr.slave.options
```

```
## Mapping in parallel: mode = socket; level = mlr.tuneParams; cpus = 12; elements = 10
```

```
## [Tune] Result: minsplit=10; minbucket=10; cp=0.0101; maxdepth=5 : mse.test.mean=1.15
```

```
parallelStop()
```

```
## Stopped parallelization. All cleaned up.
```

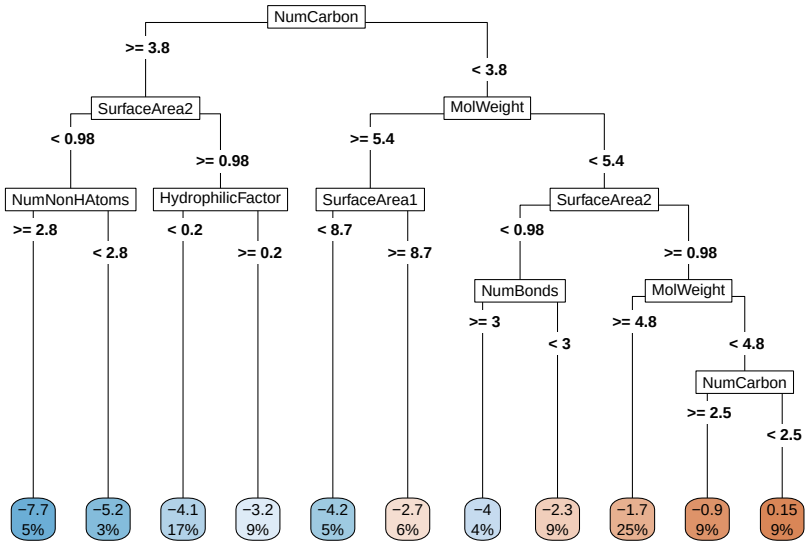
```
tunedTreePars
```

```
## Tune result:
## Op. pars: minsplit=10; minbucket=10; cp=0.0101; maxdepth=5
## mse.test.mean=1.1585897
```

```
tunedTree <- setHyperPars(treeLearner, par.vals = tunedTreePars$x)
tunedTreeModel <- train(tunedTree, task)
library(rpart.plot)
```

```
## Warning: 程辑包'rpart.plot'是用R版本4.1.3 来建造的
```

```
treeModelData <- getLearnerModel(tunedTreeModel)
rpart.plot(treeModelData, roundint = FALSE, box.palette = "BuBn", type = 5)
```



5.3 随机森林

```
## Random Foreast -----
rfLearner <- makeLearner("regr.randomForest")
rfParamSpace <- makeParamSet( makeIntegerParam("ntree", lower = 100, upper = 100),
                              makeIntegerParam("mtry", lower = 6, upper = 10),
                              makeIntegerParam("nodesize", lower = 3, upper = 10),
                              makeIntegerParam("maxnodes", lower = 5, upper = 20))
randSearch <- makeTuneControlRandom(maxit = 100)
cvForTuning <- makeResampleDesc("CV", iters = 5)
parallelStartSocket(cpus = detectCores())
```

```
## Starting parallelization in mode=socket with cpus=12.
```

```
tunedRFPars <- tuneParams( rfLearner, task = task,
                           resampling = cvForTuning,
                           par.set = rfParamSpace, control = randSearch)
```

```
## [Tune] Started tuning learner regr.randomForest for parameter set:
```

```
##             Type len Def     Constr Req Tunable Trafo
## ntree    integer   -   - 100 to 100   -    TRUE     -
## mtry     integer   -   -    6 to 10   -    TRUE     -
## nodesize integer   -   -    3 to 10   -    TRUE     -
## maxnodes integer   -   -    5 to 20   -    TRUE     -
```

```
## With control class: TuneControlRandom
```

```
## Imputation value: Inf
```

```
## Exporting objects to slaves for mode socket: .mlr.slave.options
```

```
## Mapping in parallel: mode = socket; level = mlr.tuneParams; cpus = 12; elements = 10
```

```
## [Tune] Result: ntree=100; mtry=10; nodesize=8; maxnodes=20 : mse.test.mean=1.0922036
```

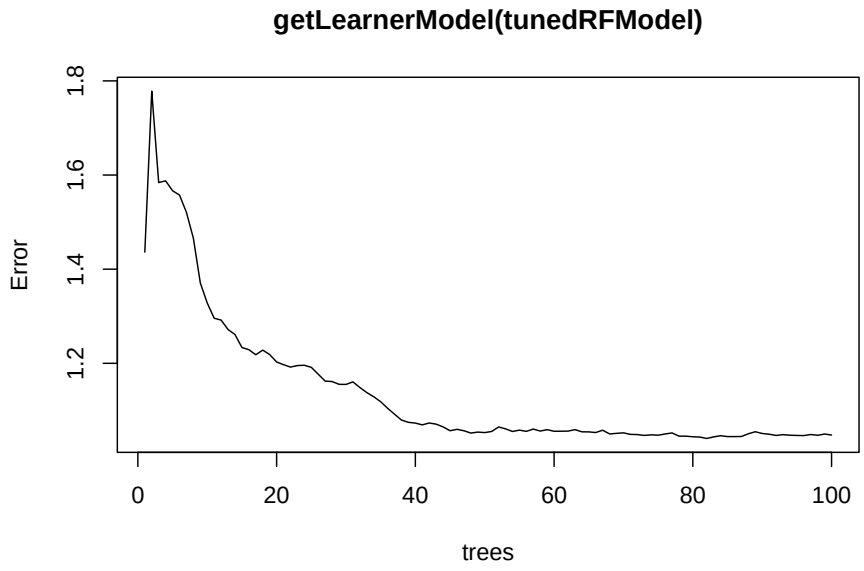
```
parallelStop()
```

```
## Stopped parallelization. All cleaned up.
```

```
tunedRFPars
```

```
## Tune result:
## Op. pars: ntree=100; mtry=10; nodesize=8; maxnodes=20
## mse.test.mean=1.0922036
```

```
tunedRF <- setHyperPars(rfLearner, par.vals = tunedRFPars$x)
tunedRFModel <- train(tunedRF, task)

plot(getLearnerModel(tunedRFModel))
```



5.4 XGBoost

```
## XGBoost -----
xgbLearner <- makeLearner("regr.xgboost")
getParamSet(xgbLearner)
```

```
##                                  Type  len             Def
## booster                       discrete    -          gbtree
## watchlist                      untyped    -          <NULL>
## eta                            numeric    -             0.3
## gamma                          numeric    -               0
## max_depth                      integer    -               6
## min_child_weight               numeric    -               1
## subsample                      numeric    -               1
## colsample_bytree               numeric    -               1
## colsample_bylevel              numeric    -               1
## colsample_bynode               numeric    -               1
## num_parallel_tree              integer    -               1
## lambda                         numeric    -               1
## lambda_bias                    numeric    -               0
## alpha                          numeric    -               0
## objective                      untyped    - reg:squarede...
## eval_metric                    untyped    -            rmse
## base_score                     numeric    -             0.5
## max_delta_step                 numeric    -               0
## missing                        numeric    -
## monotone_constraints   integervector <NA>               0
## tweedie_variance_power         numeric    -             1.5
## nthread                        integer    -               -
## nrounds                        integer    -               -
## feval                          untyped    -          <NULL>
## verbose                        integer    -               1
## print_every_n                  integer    -               1
```

```
## early_stopping_rounds       integer    -          <NULL>
## maximize                    logical    -          <NULL>
## sample_type                discrete    -         uniform
## normalize_type             discrete    -            tree
## rate_drop                   numeric    -               0
## skip_drop                   numeric    -               0
## scale_pos_weight            numeric    -               1
## refresh_leaf                logical    -            TRUE
## feature_selector           discrete    -          cyclic
## top_k                       integer    -               0
## predictor                  discrete    -   cpu_predictor
## updater                     untyped    -               -
## sketch_eps                  numeric    -            0.03
## one_drop                    logical    -           FALSE
## tree_method                discrete    -            auto
## grow_policy                discrete    -       depthwise
## max_leaves                  integer    -               0
## max_bin                     integer    -             256
## callbacks                   untyped    -          list()
##                                                    Constr Req Tunable Trafo
## booster                                gbtree,gblinear,dart   -    TRUE     -
## watchlist                                                 -   -   FALSE     -
## eta                                                 0 to 1   -    TRUE     -
## gamma                                             0 to Inf   -    TRUE     -
## max_depth                                         0 to Inf   -    TRUE     -
## min_child_weight                                  0 to Inf   -    TRUE     -
## subsample                                           0 to 1   -    TRUE     -
## colsample_bytree                                    0 to 1   -    TRUE     -
## colsample_bylevel                                   0 to 1   -    TRUE     -
## colsample_bynode                                    0 to 1   -    TRUE     -
## num_parallel_tree                                 1 to Inf   -    TRUE     -
## lambda                                            0 to Inf   -    TRUE     -
## lambda_bias                                       0 to Inf   -    TRUE     -
```

```
## alpha                                        0 to Inf   -    TRUE     -
## objective                                           -   -   FALSE     -
## eval_metric                                         -   -   FALSE     -
## base_score                                 -Inf to Inf   -   FALSE     -
## max_delta_step                               0 to Inf   -    TRUE     -
## missing                                    -Inf to Inf   -   FALSE     -
## monotone_constraints                          -1 to 1   -    TRUE     -
## tweedie_variance_power                          1 to 2   Y    TRUE     -
## nthread                                      1 to Inf   -   FALSE     -
## nrounds                                      1 to Inf   -    TRUE     -
## feval                                               -   -   FALSE     -
## verbose                                        0 to 2   -   FALSE     -
## print_every_n                                1 to Inf   Y   FALSE     -
## early_stopping_rounds                        1 to Inf   -   FALSE     -
## maximize                                            -   -   FALSE     -
## sample_type                        uniform,weighted   Y    TRUE     -
## normalize_type                          tree,forest   Y    TRUE     -
## rate_drop                                    0 to 1   Y    TRUE     -
## skip_drop                                    0 to 1   Y    TRUE     -
## scale_pos_weight                           -Inf to Inf   -    TRUE     -
## refresh_leaf                                        -   -    TRUE     -
## feature_selector       cyclic,shuffle,random,greedy,thrifty   -    TRUE     -
## top_k                                        0 to Inf   -    TRUE     -
## predictor                cpu_predictor,gpu_predictor   -    TRUE     -
## updater                                             -   -    TRUE     -
## sketch_eps                                   0 to 1   -    TRUE     -
## one_drop                                            -   Y    TRUE     -
## tree_method          auto,exact,approx,hist,gpu_hist   Y    TRUE     -
## grow_policy                    depthwise,lossguide   Y    TRUE     -
## max_leaves                                   0 to Inf   Y    TRUE     -
## max_bin                                      2 to Inf   Y    TRUE     -
## callbacks                                           -   -   FALSE     -
```

```
xgbParamSpace <- makeParamSet( makeNumericParam("eta", lower = 0, upper = 1),
                               makeNumericParam("gamma", lower = 0, upper = 5),
                               makeIntegerParam("max_depth", lower = 1, upper = 5),
                               makeNumericParam("min_child_weight", lower = 1, upper =
                               makeNumericParam("subsample", lower = 0.5, upper = 1),
                               makeNumericParam("colsample_bytree", lower = 0.5, upper
                               makeIntegerParam("nrounds", lower = 20, upper = 20))
randSearch <- makeTuneControlRandom(maxit = 100)
cvForTuning <- makeResampleDesc("CV", iters = 5)
tunedXgbPars <- tuneParams(xgbLearner, task = task,
                           resampling = cvForTuning,
                           par.set = xgbParamSpace, control = randSearch)
```

```
## [Tune] Started tuning learner regr.xgboost for parameter set:
```

```
##                     Type len Def   Constr Req Tunable Trafo
## eta              numeric   -   -   0 to 1   -    TRUE     -
## gamma            numeric   -   -   0 to 5   -    TRUE     -
## max_depth        integer   -   -   1 to 5   -    TRUE     -
## min_child_weight numeric   -   -  1 to 10   -    TRUE     -
## subsample        numeric   -   - 0.5 to 1   -    TRUE     -
## colsample_bytree numeric   -   - 0.5 to 1   -    TRUE     -
## nrounds          integer   -   - 20 to 20   -    TRUE     -
```

```
## With control class: TuneControlRandom
```

```
## Imputation value: Inf
```

```
## [Tune-x] 1: eta=0.979; gamma=1.87; max_depth=3; min_child_weight=2.96; subsample=0.7
```

```
## [Tune-y] 1: mse.test.mean=0.7744634; time: 0.0 min
```

```
## [Tune-x] 2: eta=0.523; gamma=1.77; max_depth=4; min_child_weight=6.54; subsample=0.7
```

```
## [Tune-y] 2: mse.test.mean=0.5665814; time: 0.0 min

## [Tune-x] 3: eta=0.342; gamma=0.839; max_depth=4; min_child_weight=8.98; subsample=0.

## [Tune-y] 3: mse.test.mean=0.4673753; time: 0.0 min

## [Tune-x] 4: eta=0.671; gamma=1.56; max_depth=1; min_child_weight=2.15; subsample=0.9

## [Tune-y] 4: mse.test.mean=0.8691468; time: 0.0 min

## [Tune-x] 5: eta=0.532; gamma=3.57; max_depth=5; min_child_weight=9.42; subsample=0.7

## [Tune-y] 5: mse.test.mean=0.5914411; time: 0.0 min

## [Tune-x] 6: eta=0.901; gamma=3.28; max_depth=5; min_child_weight=1.3; subsample=0.51

## [Tune-y] 6: mse.test.mean=1.0175523; time: 0.0 min

## [Tune-x] 7: eta=0.594; gamma=3.78; max_depth=5; min_child_weight=7.64; subsample=0.8

## [Tune-y] 7: mse.test.mean=0.6257518; time: 0.0 min

## [Tune-x] 8: eta=0.54; gamma=1.1; max_depth=5; min_child_weight=6.27; subsample=0.752

## [Tune-y] 8: mse.test.mean=0.6002412; time: 0.0 min

## [Tune-x] 9: eta=0.818; gamma=2.08; max_depth=1; min_child_weight=5.85; subsample=0.7

## [Tune-y] 9: mse.test.mean=0.8860011; time: 0.0 min

## [Tune-x] 10: eta=0.535; gamma=0.603; max_depth=2; min_child_weight=6.99; subsample=0

## [Tune-y] 10: mse.test.mean=0.6586332; time: 0.0 min

## [Tune-x] 11: eta=0.808; gamma=1.76; max_depth=3; min_child_weight=5.01; subsample=0.
```

```
## [Tune-y] 11: mse.test.mean=0.6589334; time: 0.0 min

## [Tune-x] 12: eta=0.357; gamma=4.57; max_depth=2; min_child_weight=4.07; subsample=0.

## [Tune-y] 12: mse.test.mean=0.6568810; time: 0.0 min

## [Tune-x] 13: eta=0.00369; gamma=1.19; max_depth=3; min_child_weight=5.59; subsample=

## [Tune-y] 13: mse.test.mean=12.7664200; time: 0.0 min

## [Tune-x] 14: eta=0.823; gamma=1.52; max_depth=4; min_child_weight=8.39; subsample=0.

## [Tune-y] 14: mse.test.mean=0.7822538; time: 0.0 min

## [Tune-x] 15: eta=0.998; gamma=0.858; max_depth=3; min_child_weight=7.44; subsample=0

## [Tune-y] 15: mse.test.mean=0.9420068; time: 0.0 min

## [Tune-x] 16: eta=0.186; gamma=3.38; max_depth=5; min_child_weight=6.18; subsample=0.

## [Tune-y] 16: mse.test.mean=0.5336803; time: 0.0 min

## [Tune-x] 17: eta=0.639; gamma=3.61; max_depth=3; min_child_weight=9.38; subsample=0.

## [Tune-y] 17: mse.test.mean=0.6300923; time: 0.0 min

## [Tune-x] 18: eta=0.462; gamma=4.2; max_depth=4; min_child_weight=7.89; subsample=0.6

## [Tune-y] 18: mse.test.mean=0.5854921; time: 0.0 min

## [Tune-x] 19: eta=0.417; gamma=4.72; max_depth=1; min_child_weight=5.82; subsample=0.

## [Tune-y] 19: mse.test.mean=0.8718688; time: 0.0 min

## [Tune-x] 20: eta=0.416; gamma=3.89; max_depth=3; min_child_weight=1.88; subsample=0.
```

```
## [Tune-y] 20: mse.test.mean=0.5845345; time: 0.0 min

## [Tune-x] 21: eta=0.583; gamma=0.867; max_depth=3; min_child_weight=8.41; subsample=0

## [Tune-y] 21: mse.test.mean=0.5716593; time: 0.0 min

## [Tune-x] 22: eta=0.861; gamma=3.45; max_depth=5; min_child_weight=1.33; subsample=0.

## [Tune-y] 22: mse.test.mean=0.7716797; time: 0.0 min

## [Tune-x] 23: eta=0.0335; gamma=2.75; max_depth=4; min_child_weight=5.91; subsample=0

## [Tune-y] 23: mse.test.mean=4.6257778; time: 0.0 min

## [Tune-x] 24: eta=0.244; gamma=4.55; max_depth=1; min_child_weight=2.15; subsample=0.

## [Tune-y] 24: mse.test.mean=1.0902649; time: 0.0 min

## [Tune-x] 25: eta=0.712; gamma=3.05; max_depth=1; min_child_weight=1.47; subsample=0.

## [Tune-y] 25: mse.test.mean=0.8508911; time: 0.0 min

## [Tune-x] 26: eta=0.584; gamma=0.964; max_depth=1; min_child_weight=6.9; subsample=0.

## [Tune-y] 26: mse.test.mean=0.8779755; time: 0.0 min

## [Tune-x] 27: eta=0.273; gamma=1.7; max_depth=2; min_child_weight=8.14; subsample=0.8

## [Tune-y] 27: mse.test.mean=0.6931201; time: 0.0 min

## [Tune-x] 28: eta=0.471; gamma=0.64; max_depth=3; min_child_weight=4.02; subsample=0.

## [Tune-y] 28: mse.test.mean=0.5534129; time: 0.0 min

## [Tune-x] 29: eta=0.728; gamma=4.55; max_depth=1; min_child_weight=6.39; subsample=0.
```

```
## [Tune-y] 29: mse.test.mean=0.8785849; time: 0.0 min

## [Tune-x] 30: eta=0.409; gamma=0.158; max_depth=4; min_child_weight=9.54; subsample=0

## [Tune-y] 30: mse.test.mean=0.5110228; time: 0.0 min

## [Tune-x] 31: eta=0.454; gamma=2.92; max_depth=3; min_child_weight=5.63; subsample=0.

## [Tune-y] 31: mse.test.mean=0.5379262; time: 0.0 min

## [Tune-x] 32: eta=0.217; gamma=0.899; max_depth=3; min_child_weight=9.79; subsample=0

## [Tune-y] 32: mse.test.mean=0.5845813; time: 0.0 min

## [Tune-x] 33: eta=0.627; gamma=1.05; max_depth=5; min_child_weight=3.08; subsample=0.

## [Tune-y] 33: mse.test.mean=0.6581594; time: 0.0 min

## [Tune-x] 34: eta=0.902; gamma=2.05; max_depth=1; min_child_weight=2.29; subsample=0.

## [Tune-y] 34: mse.test.mean=0.9036214; time: 0.0 min

## [Tune-x] 35: eta=0.104; gamma=1.15; max_depth=4; min_child_weight=4.22; subsample=0.

## [Tune-y] 35: mse.test.mean=0.8411709; time: 0.0 min

## [Tune-x] 36: eta=0.499; gamma=3.05; max_depth=3; min_child_weight=3.96; subsample=0.

## [Tune-y] 36: mse.test.mean=0.5923829; time: 0.0 min

## [Tune-x] 37: eta=0.612; gamma=4.18; max_depth=3; min_child_weight=1.81; subsample=0.

## [Tune-y] 37: mse.test.mean=0.6343750; time: 0.0 min

## [Tune-x] 38: eta=0.542; gamma=0.691; max_depth=4; min_child_weight=5.05; subsample=0
```

```
## [Tune-y] 38: mse.test.mean=0.5193695; time: 0.0 min

## [Tune-x] 39: eta=0.869; gamma=3.26; max_depth=2; min_child_weight=8.05; subsample=0.

## [Tune-y] 39: mse.test.mean=0.6776593; time: 0.0 min

## [Tune-x] 40: eta=0.592; gamma=2.09; max_depth=2; min_child_weight=1.66; subsample=0.

## [Tune-y] 40: mse.test.mean=0.6543516; time: 0.0 min

## [Tune-x] 41: eta=0.509; gamma=2.84; max_depth=3; min_child_weight=4.56; subsample=0.

## [Tune-y] 41: mse.test.mean=0.6215488; time: 0.0 min

## [Tune-x] 42: eta=0.198; gamma=1.25; max_depth=5; min_child_weight=1.29; subsample=0.

## [Tune-y] 42: mse.test.mean=0.5047306; time: 0.0 min

## [Tune-x] 43: eta=0.0463; gamma=4.87; max_depth=3; min_child_weight=7.44; subsample=0

## [Tune-y] 43: mse.test.mean=3.1842500; time: 0.0 min

## [Tune-x] 44: eta=0.437; gamma=1.96; max_depth=4; min_child_weight=2.49; subsample=0.

## [Tune-y] 44: mse.test.mean=0.5707890; time: 0.0 min

## [Tune-x] 45: eta=0.205; gamma=3.4; max_depth=5; min_child_weight=7.07; subsample=0.6

## [Tune-y] 45: mse.test.mean=0.5349752; time: 0.0 min

## [Tune-x] 46: eta=0.0625; gamma=3.08; max_depth=2; min_child_weight=5.94; subsample=0

## [Tune-y] 46: mse.test.mean=2.3762497; time: 0.0 min

## [Tune-x] 47: eta=0.992; gamma=4.02; max_depth=4; min_child_weight=5.51; subsample=0.
```

```
## [Tune-y] 47: mse.test.mean=0.8846706; time: 0.0 min

## [Tune-x] 48: eta=0.125; gamma=1.25; max_depth=5; min_child_weight=6.91; subsample=0.

## [Tune-y] 48: mse.test.mean=0.6281202; time: 0.0 min

## [Tune-x] 49: eta=0.565; gamma=3.71; max_depth=5; min_child_weight=7.92; subsample=0.

## [Tune-y] 49: mse.test.mean=0.5977611; time: 0.0 min

## [Tune-x] 50: eta=0.721; gamma=0.652; max_depth=4; min_child_weight=2.89; subsample=0

## [Tune-y] 50: mse.test.mean=0.6103261; time: 0.0 min

## [Tune-x] 51: eta=0.636; gamma=0.214; max_depth=1; min_child_weight=5.43; subsample=0

## [Tune-y] 51: mse.test.mean=0.8767903; time: 0.0 min

## [Tune-x] 52: eta=0.525; gamma=0.126; max_depth=5; min_child_weight=5.64; subsample=0

## [Tune-y] 52: mse.test.mean=0.5822375; time: 0.0 min

## [Tune-x] 53: eta=0.629; gamma=1.15; max_depth=5; min_child_weight=7.7; subsample=0.8

## [Tune-y] 53: mse.test.mean=0.5968220; time: 0.0 min

## [Tune-x] 54: eta=0.574; gamma=4.46; max_depth=4; min_child_weight=2.71; subsample=0.

## [Tune-y] 54: mse.test.mean=0.6403676; time: 0.0 min

## [Tune-x] 55: eta=0.108; gamma=0.914; max_depth=4; min_child_weight=8.76; subsample=0

## [Tune-y] 55: mse.test.mean=0.8085014; time: 0.0 min

## [Tune-x] 56: eta=0.387; gamma=3.52; max_depth=4; min_child_weight=6.31; subsample=0.
```

```
## [Tune-y] 56: mse.test.mean=0.5602168; time: 0.0 min

## [Tune-x] 57: eta=0.643; gamma=0.241; max_depth=1; min_child_weight=1.47; subsample=0

## [Tune-y] 57: mse.test.mean=0.8708133; time: 0.0 min

## [Tune-x] 58: eta=0.013; gamma=2.32; max_depth=2; min_child_weight=1.68; subsample=0.

## [Tune-y] 58: mse.test.mean=9.3719818; time: 0.0 min

## [Tune-x] 59: eta=0.996; gamma=0.383; max_depth=5; min_child_weight=2.67; subsample=0

## [Tune-y] 59: mse.test.mean=1.2259959; time: 0.0 min

## [Tune-x] 60: eta=0.539; gamma=1.74; max_depth=5; min_child_weight=7.26; subsample=0.

## [Tune-y] 60: mse.test.mean=0.5666969; time: 0.0 min

## [Tune-x] 61: eta=0.075; gamma=4.98; max_depth=1; min_child_weight=3.31; subsample=0.

## [Tune-y] 61: mse.test.mean=2.5970627; time: 0.0 min

## [Tune-x] 62: eta=0.0729; gamma=4.18; max_depth=2; min_child_weight=8.92; subsample=0

## [Tune-y] 62: mse.test.mean=1.9305385; time: 0.0 min

## [Tune-x] 63: eta=0.112; gamma=4.35; max_depth=4; min_child_weight=7.04; subsample=0.

## [Tune-y] 63: mse.test.mean=0.7992497; time: 0.0 min

## [Tune-x] 64: eta=0.647; gamma=0.215; max_depth=2; min_child_weight=4.57; subsample=0

## [Tune-y] 64: mse.test.mean=0.6638292; time: 0.0 min

## [Tune-x] 65: eta=0.883; gamma=2.37; max_depth=1; min_child_weight=2.33; subsample=0.
```

```
## [Tune-y] 65: mse.test.mean=0.8463553; time: 0.0 min

## [Tune-x] 66: eta=0.482; gamma=3.46; max_depth=5; min_child_weight=7.54; subsample=0.

## [Tune-y] 66: mse.test.mean=0.5538379; time: 0.0 min

## [Tune-x] 67: eta=0.039; gamma=3.39; max_depth=1; min_child_weight=8.35; subsample=0.

## [Tune-y] 67: mse.test.mean=4.8948406; time: 0.0 min

## [Tune-x] 68: eta=0.829; gamma=0.755; max_depth=2; min_child_weight=8.33; subsample=0

## [Tune-y] 68: mse.test.mean=0.6822841; time: 0.0 min

## [Tune-x] 69: eta=0.989; gamma=4.75; max_depth=2; min_child_weight=8.41; subsample=0.

## [Tune-y] 69: mse.test.mean=0.8038256; time: 0.0 min

## [Tune-x] 70: eta=0.203; gamma=2.52; max_depth=3; min_child_weight=8.63; subsample=0.

## [Tune-y] 70: mse.test.mean=0.6009508; time: 0.0 min

## [Tune-x] 71: eta=0.00071; gamma=1.83; max_depth=1; min_child_weight=2.34; subsample=

## [Tune-y] 71: mse.test.mean=14.2128111; time: 0.0 min

## [Tune-x] 72: eta=0.904; gamma=3.43; max_depth=3; min_child_weight=2.98; subsample=0.

## [Tune-y] 72: mse.test.mean=0.8663598; time: 0.0 min

## [Tune-x] 73: eta=0.164; gamma=1.97; max_depth=1; min_child_weight=4.96; subsample=0.

## [Tune-y] 73: mse.test.mean=1.3732722; time: 0.0 min

## [Tune-x] 74: eta=0.855; gamma=3.99; max_depth=5; min_child_weight=3.44; subsample=0.
```

```
## [Tune-y] 74: mse.test.mean=0.7971487; time: 0.0 min

## [Tune-x] 75: eta=0.247; gamma=3.96; max_depth=4; min_child_weight=8.34; subsample=0.

## [Tune-y] 75: mse.test.mean=0.5110311; time: 0.0 min

## [Tune-x] 76: eta=0.843; gamma=4.12; max_depth=2; min_child_weight=8.21; subsample=0.

## [Tune-y] 76: mse.test.mean=0.6793318; time: 0.0 min

## [Tune-x] 77: eta=0.0682; gamma=3.03; max_depth=3; min_child_weight=1.53; subsample=0

## [Tune-y] 77: mse.test.mean=1.8122322; time: 0.0 min

## [Tune-x] 78: eta=0.588; gamma=0.141; max_depth=4; min_child_weight=5.89; subsample=0

## [Tune-y] 78: mse.test.mean=0.6022628; time: 0.0 min

## [Tune-x] 79: eta=0.481; gamma=0.971; max_depth=5; min_child_weight=5.84; subsample=0

## [Tune-y] 79: mse.test.mean=0.5005877; time: 0.0 min

## [Tune-x] 80: eta=0.772; gamma=0.68; max_depth=5; min_child_weight=1.99; subsample=0.

## [Tune-y] 80: mse.test.mean=0.7581125; time: 0.0 min

## [Tune-x] 81: eta=0.364; gamma=1.46; max_depth=5; min_child_weight=1.91; subsample=0.

## [Tune-y] 81: mse.test.mean=0.5057299; time: 0.0 min

## [Tune-x] 82: eta=0.883; gamma=1.89; max_depth=4; min_child_weight=4.95; subsample=0.

## [Tune-y] 82: mse.test.mean=0.6985856; time: 0.0 min

## [Tune-x] 83: eta=0.113; gamma=0.56; max_depth=4; min_child_weight=4.61; subsample=0.
```

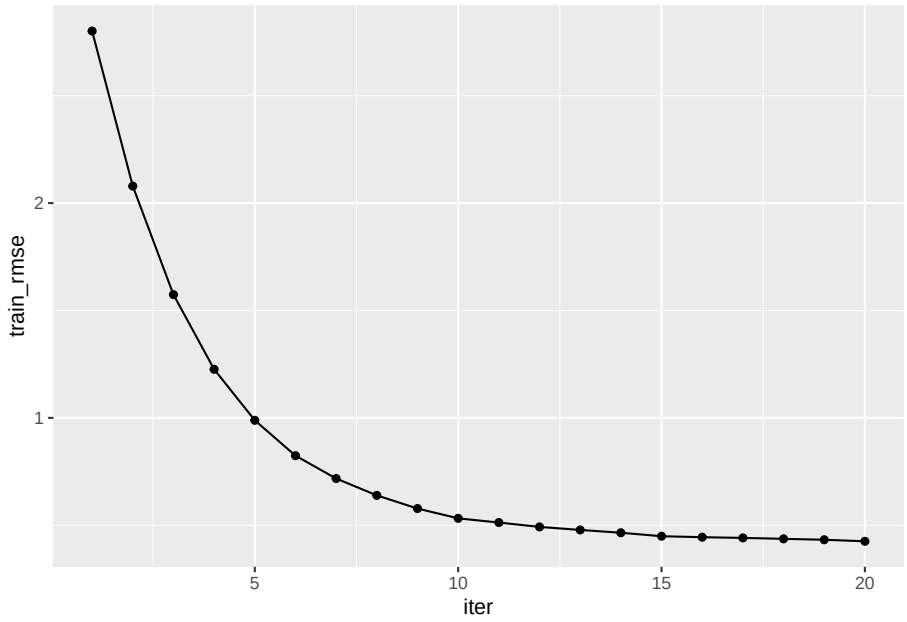
```
## [Tune-y] 83: mse.test.mean=0.7554341; time: 0.0 min

## [Tune-x] 84: eta=0.292; gamma=1.48; max_depth=5; min_child_weight=7.3; subsample=0.9

## [Tune-y] 84: mse.test.mean=0.4471577; time: 0.0 min

## [Tune-x] 85: eta=0.832; gamma=0.181; max_depth=4; min_child_weight=4.4; subsample=0.

## [Tune-y] 85: mse.test.mean=0.7869675; time: 0.0 min

## [Tune-x] 86: eta=0.132; gamma=4.61; max_depth=3; min_child_weight=3.05; subsample=0.

## [Tune-y] 86: mse.test.mean=0.7889893; time: 0.0 min

## [Tune-x] 87: eta=0.427; gamma=1.31; max_depth=3; min_child_weight=3.24; subsample=0.

## [Tune-y] 87: mse.test.mean=0.5047390; time: 0.0 min

## [Tune-x] 88: eta=0.12; gamma=3.05; max_depth=5; min_child_weight=1.97; subsample=0.9

## [Tune-y] 88: mse.test.mean=0.6760414; time: 0.0 min

## [Tune-x] 89: eta=0.757; gamma=2.03; max_depth=1; min_child_weight=8.85; subsample=0.

## [Tune-y] 89: mse.test.mean=0.8291031; time: 0.0 min

## [Tune-x] 90: eta=0.79; gamma=0.374; max_depth=3; min_child_weight=6.74; subsample=0.

## [Tune-y] 90: mse.test.mean=0.6571913; time: 0.0 min

## [Tune-x] 91: eta=0.692; gamma=0.431; max_depth=5; min_child_weight=6.57; subsample=0

## [Tune-y] 91: mse.test.mean=0.7600307; time: 0.0 min

## [Tune-x] 92: eta=0.651; gamma=4.18; max_depth=3; min_child_weight=6.52; subsample=0.
```

```
## [Tune-y] 92: mse.test.mean=0.7165721; time: 0.0 min

## [Tune-x] 93: eta=0.0188; gamma=1.88; max_depth=4; min_child_weight=7.6; subsample=0.

## [Tune-y] 93: mse.test.mean=7.5028757; time: 0.0 min

## [Tune-x] 94: eta=0.984; gamma=3.94; max_depth=4; min_child_weight=1.98; subsample=0.

## [Tune-y] 94: mse.test.mean=0.9118783; time: 0.0 min

## [Tune-x] 95: eta=0.627; gamma=2.54; max_depth=1; min_child_weight=1.39; subsample=0.

## [Tune-y] 95: mse.test.mean=0.8549547; time: 0.0 min

## [Tune-x] 96: eta=0.984; gamma=1.63; max_depth=4; min_child_weight=8.42; subsample=0.

## [Tune-y] 96: mse.test.mean=0.7418318; time: 0.0 min

## [Tune-x] 97: eta=0.211; gamma=1.74; max_depth=1; min_child_weight=8.17; subsample=0.

## [Tune-y] 97: mse.test.mean=1.1824821; time: 0.0 min

## [Tune-x] 98: eta=0.991; gamma=4.38; max_depth=1; min_child_weight=6.1; subsample=0.5

## [Tune-y] 98: mse.test.mean=0.9592601; time: 0.0 min

## [Tune-x] 99: eta=0.73; gamma=1.05; max_depth=1; min_child_weight=4.93; subsample=0.8

## [Tune-y] 99: mse.test.mean=0.8309914; time: 0.0 min

## [Tune-x] 100: eta=0.173; gamma=3.97; max_depth=2; min_child_weight=2.77; subsample=0

## [Tune-y] 100: mse.test.mean=0.8238041; time: 0.0 min

## [Tune] Result: eta=0.292; gamma=1.48; max_depth=5; min_child_weight=7.3; subsample=0
```

```
tunedXgbPars
```

```
## Tune result:
## Op. pars: eta=0.292; gamma=1.48; max_depth=5; min_child_weight=7.3; subsample=0.916;
## mse.test.mean=0.4471577
```

```
tunedXgb <- setHyperPars(xgbLearner, par.vals = tunedXgbPars$x)
tunedXgbModel <- train(tunedXgb, task)
xgbModelData <- getLearnerModel(tunedXgbModel)

ggplot(xgbModelData$evaluation_log,aes(iter,train_rmse)) +
  geom_point()+geom_line()
```



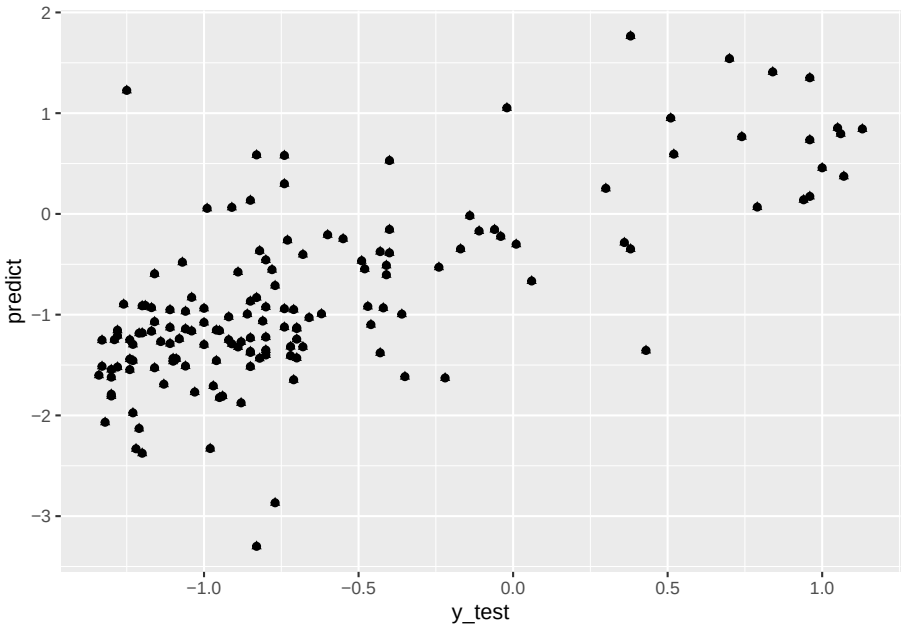

```
#install.packages("DiagrammeR")
library(DiagrammeR)
```

```
## Warning: 程辑包'DiagrammeR'是用R版本4.1.3 来建造的
```

```
xgboost::xgb.plot.tree(model = xgbModelData, trees = 1:2)
```

5.5 线性回归

```
## Linear Regression-----
x_train<-solubility_data[1:800,1:228]
y_train<-solubility_data[1:800,229]
x_test<-solubility_data[801:951,1:228]
y_test<-solubility_data[801:951,229]
x<-cbind(x_train,y_train)
linear<-lm(y_train~.,data=x)
model_sum<-summary(linear)
predict=predict(linear,x_test)
linear_mse=mean((y_test-predict)^2)
y_test=as.data.frame(y_test)
predict=as.data.frame(predict)
z<-cbind(y_test,predict)
ggplot(data = z, aes(x = y_test, y = predict)) +
  geom_point()+
  geom_point(shape = 17)
```




```
print(linear_mse)
```

```
## [1] 0.4303537
```

6 PCA 主成分分析

```
rownames(solubility_data) <- paste("sample",1:nrow(solubility_data),sep = "") # 设置样本
head(solubility_data) # 查看数据集前几行
```

```
##         FP001 FP002 FP003 FP004 FP005 FP006 FP007 FP008 FP009 FP010 FP011 FP012
## sample1     0     1     0     0     1     0     0     1     0     0     0     0
## sample2     0     1     0     1     1     1     1     1     0     0     1     0
## sample3     1     1     1     1     1     0     0     1     0     1     0     0
## sample4     0     0     1     0     0     0     1     0     0     0     0     0
## sample5     0     0     1     1     1     1     0     0     1     0     0     0
## sample6     1     0     1     1     0     0     0     0     1     0     0     1
```

```
##         FP013 FP014 FP015 FP016 FP017 FP018 FP019 FP020 FP021 FP022 FP023 FP024
## sample1     0     0     1     0     0     0     1     0     0     0     0     1
## sample2     0     0     1     1     0     1     0     0     0     0     0     0
## sample3     0     0     1     0     1     0     0     0     0     0     0     0
## sample4     0     0     1     0     1     0     0     0     0     0     1     0
## sample5     1     0     1     1     0     0     1     0     0     0     0     1
## sample6     0     0     1     1     0     0     0     0     1     0     0     0
##         FP025 FP026 FP027 FP028 FP029 FP030 FP031 FP032 FP033 FP034 FP035 FP036
## sample1     0     1     0     0     0     0     0     0     0     0     0     0
## sample2     0     0     0     1     0     0     0     0     0     0     0     0
## sample3     1     0     0     0     0     0     0     0     0     0     0     0
## sample4     0     0     0     0     0     0     0     0     0     0     0     0
## sample5     0     0     0     0     0     1     0     0     0     1     0     0
## sample6     0     0     0     0     0     0     0     0     0     0     0     0
##         FP037 FP038 FP039 FP040 FP041 FP042 FP043 FP044 FP045 FP046 FP047 FP048
## sample1     0     0     1     1     0     0     0     0     0     0     0     0
## sample2     0     0     0     0     0     0     1     0     0     1     1     0
## sample3     0     1     0     0     0     0     0     0     1     0     1     0
## sample4     0     0     0     0     1     0     0     0     0     0     0     0
## sample5     0     0     0     0     0     0     0     0     0     0     0     0
## sample6     0     0     0     0     0     0     0     0     0     0     0     0
##         FP049 FP050 FP051 FP052 FP053 FP054 FP055 FP056 FP057 FP058 FP059 FP060
## sample1     0     0     0     0     0     0     0     1     0     0     0     0
## sample2     0     0     1     0     0     0     0     0     0     0     0     1
## sample3     0     0     0     0     0     0     0     0     0     0     0     1
## sample4     0     0     0     0     0     1     0     0     0     0     0     0
## sample5     0     0     0     0     0     0     0     0     0     0     0     0
## sample6     0     0     0     0     0     0     0     0     0     0     0     0
##         FP061 FP062 FP063 FP064 FP065 FP066 FP067 FP068 FP069 FP070 FP071 FP072
## sample1     0     0     1     0     1     1     1     0     1     1     0     0
## sample2     0     0     1     1     1     0     1     1     0     1     0     1
## sample3     1     1     0     1     0     1     0     0     1     0     0     1
## sample4     0     0     0     0     0     1     0     0     1     1     0     0
```

```
## sample5     0     0     1     1     1     1     1     1     1     0     0     0
## sample6     0     1     1     1     0     1     1     1     1     0     0     1
##         FP073 FP074 FP075 FP076 FP077 FP078 FP079 FP080 FP081 FP082 FP083 FP084
## sample1     0     0     0     1     0     0     1     0     0     1     0     1
## sample2     1     1     1     1     1     1     1     1     0     1     0     1
## sample3     1     0     0     0     0     0     1     0     1     1     0     0
## sample4     0     0     0     0     1     0     1     0     1     0     0     0
## sample5     0     0     1     0     0     0     1     1     0     1     1     1
## sample6     0     0     1     0     0     0     0     1     0     1     0     0
##         FP085 FP086 FP087 FP088 FP089 FP090 FP091 FP092 FP093 FP094 FP095 FP096
## sample1     0     0     1     0     1     0     1     0     0     0     0     0
## sample2     1     0     1     1     1     1     1     0     1     0     0     0
## sample3     0     0     1     0     0     0     0     0     0     0     0     0
## sample4     0     1     1     0     0     1     0     0     1     0     0     0
## sample5     0     1     1     0     0     0     1     1     0     1     0     0
## sample6     0     0     0     0     0     0     0     1     0     0     0     0
##         FP097 FP098 FP099 FP100 FP101 FP102 FP103 FP104 FP105 FP106 FP107 FP108
## sample1     1     0     0     0     1     0     0     1     0     0     0     0
## sample2     1     0     0     0     1     1     0     1     1     1     0     0
## sample3     0     1     0     0     0     0     0     0     0     0     0     0
## sample4     0     0     0     0     0     0     1     0     0     0     0     0
## sample5     0     0     0     0     1     1     0     1     0     1     1     0
## sample6     0     0     0     1     0     1     0     0     0     0     1     0
##         FP109 FP110 FP111 FP112 FP113 FP114 FP115 FP116 FP117 FP118 FP119 FP120
## sample1     0     0     0     0     0     0     0     1     0     0     0     0
## sample2     1     0     0     0     1     0     0     1     0     1     0     0
## sample3     0     0     0     0     0     0     0     0     0     0     0     0
## sample4     0     0     0     0     0     0     0     0     0     0     0     0
## sample5     1     0     1     1     0     1     0     1     0     0     0     0
## sample6     0     1     1     0     0     0     0     0     0     0     0     0
##         FP121 FP122 FP123 FP124 FP125 FP126 FP127 FP128 FP129 FP130 FP131 FP132
## sample1     0     0     0     0     0     0     0     0     0     0     0     0
## sample2     1     0     0     0     0     1     0     0     0     0     1     0
```

```
## sample3      0      0      0      0      0      0      0      0      0      0      0      0
## sample4      0      0      1      0      0      0      0      0      0      0      0      0
## sample5      1      0      0      0      0      0      1      0      0      0      1      0
## sample6      1      0      1      0      0      0      0      0      0      0      0      0
##          FP133  FP134  FP135  FP136  FP137  FP138  FP139  FP140  FP141  FP142  FP143  FP144
## sample1      0      0      0      1      0      0      0      0      0      0      0      0
## sample2      0      1      1      0      1      1      0      0      1      1      0      0
## sample3      1      0      0      0      0      0      0      0      0      0      0      0
## sample4      0      0      0      0      0      0      0      0      0      0      0      0
## sample5      0      0      0      1      1      1      0      0      0      1      0      1
## sample6      1      1      0      0      1      0      0      0      1      0      0      0
##          FP145  FP146  FP147  FP148  FP149  FP150  FP151  FP152  FP153  FP154  FP155  FP156
## sample1      0      0      0      0      0      1      0      0      0      0      0      0
## sample2      0      0      0      0      0      0      0      0      1      1      0      0
## sample3      0      0      0      0      0      0      0      0      0      0      0      0
## sample4      0      1      0      0      0      0      0      0      0      0      1      0
## sample5      0      0      0      0      0      0      0      0      0      0      0      0
## sample6      0      1      0      0      0      0      0      0      0      0      0      0
##          FP157  FP158  FP159  FP160  FP161  FP162  FP163  FP164  FP165  FP166  FP167  FP168
## sample1      0      0      0      0      0      1      0      1      0      0      0      1
## sample2      0      0      1      0      0      1      0      1      0      1      0      1
## sample3      0      0      0      0      0      1      1      1      1      0      1      1
## sample4      0      0      0      0      0      0      0      1      1      0      0      1
## sample5      0      0      0      0      0      0      0      0      0      0      0      1
## sample6      0      0      0      0      0      0      1      0      1      0      1      0
##          FP169  FP170  FP171  FP172  FP173  FP174  FP175  FP176  FP177  FP178  FP179  FP180
## sample1      0      0      0      0      0      0      0      1      0      0      1      0
## sample2      1      1      0      0      1      0      0      0      0      1      0      0
## sample3      0      0      1      0      0      0      0      0      0      0      0      1
## sample4      0      1      0      0      1      0      0      0      0      0      1      0
## sample5      0      0      0      1      0      0      0      1      1      0      0      0
## sample6      0      0      1      0      0      0      0      0      0      0      0      0
##          FP181  FP182  FP183  FP184  FP185  FP186  FP187  FP188  FP189  FP190  FP191  FP192
```

```
## sample1     0     0     0     0     0     0     0     0     0     0     0     0
## sample2     0     0     0     0     1     0     0     0     1     0     0     0
## sample3     0     1     0     0     0     0     0     0     0     0     0     0
## sample4     0     0     0     0     1     0     0     1     0     0     0     0
## sample5     0     0     0     0     0     0     0     0     0     0     0     1
## sample6     0     1     0     0     0     0     0     0     0     0     1     1
##         FP193 FP194 FP195 FP196 FP197 FP198 FP199 FP200 FP201 FP202 FP203 FP204
## sample1     0     0     0     0     0     0     0     0     1     0     0     0
## sample2     0     0     0     0     0     0     0     0     0     1     0     0
## sample3     0     0     0     0     0     0     0     0     0     0     0     0
## sample4     0     0     0     0     1     0     0     0     0     0     0     0
## sample5     0     0     0     0     0     0     0     0     0     0     0     0
## sample6     0     0     0     0     0     0     0     0     0     0     0     0
##         FP205 FP206 FP207 FP208 MolWeight NumAtoms NumNonHAtoms NumBonds
## sample1     0     0     0     0  5.343673 3.367296     2.833213 3.433987
## sample2     0     0     0     0  5.904108 3.912023     3.295837 3.970292
## sample3     0     0     0     0  5.334215 3.526361     2.772589 3.526361
## sample4     0     0     0     0  4.921877 3.295837     2.397895 3.295837
## sample5     0     0     0     0  5.441335 3.465736     2.772589 3.465736
## sample6     0     0     0     0  5.603041 3.496508     2.772589 3.465736
##         NumNonHBonds NumMultBonds NumRotBonds NumDblBonds NumAromaticBonds
## sample1     4.009916     5.264609   0.0000000   0.0000000         2.833213
## sample2     4.871752     4.684412   1.6094379   0.0000000         2.564949
## sample3     3.705506     3.243492   1.6094379   0.5670767         1.945910
## sample4     3.076971     1.379614   0.6931472   0.8045302         0.000000
## sample5     3.705506     2.944766   1.7917595   0.0000000         1.945910
## sample6     3.593860     1.379614   1.7917595   0.8045302         0.000000
##         NumHydrogen NumCarbon NumNitrogen NumOxygen NumSulfer NumChlorine
## sample1    3.862179  4.177811   0.5848146 0.0000000     0.000   0.0000000
## sample2    5.315193  5.092358   0.6423550 0.6931472     0.375   0.0000000
## sample3    4.729818  4.023944   0.0000000 1.0986123     0.000   0.0000000
## sample4    4.465209  3.510455   0.0000000 0.0000000     0.000   0.0000000
## sample5    4.465209  3.317541   0.6943345 0.0000000     0.000   0.3750000
```

```
## sample6    4.600088  3.510455   0.4568260 0.6931472     0.375   0.4444444
##         NumHalogen  NumRings HydrophilicFactor SurfaceArea1 SurfaceArea2
## sample1  0.0000000 1.3862944       -1.60654181     6.812456     6.812456
## sample2  0.0000000 1.6094379       -0.44133043     9.753834    12.029604
## sample3  0.0000000 0.6931472       -0.38485910     8.245324     8.245324
## sample4  0.0000000 0.6931472       -2.37347220     0.000000     0.000000
## sample5  0.3750000 0.6931472       -0.07098726     9.913535     9.913535
## sample6  0.4444444 0.0000000       -0.94925327     5.999109     9.123359
##         Solubility
## sample1      -3.97
## sample2      -3.98
## sample3      -3.99
## sample4      -4.00
## sample5      -4.06
## sample6      -4.08
```

```
solubility_scale <- scale(solubility_data[,-ncol(solubility_data)]) # 标准化原始数据
head(solubility_scale)
```

```
##              FP001      FP002      FP003      FP004      FP005      FP006
## sample1 -0.9859036  0.9235277 -0.8794541 -1.1857971  0.8515806 -0.8171403
## sample2 -0.9859036  0.9235277 -0.8794541  0.8424278  0.8515806  1.2224933
## sample3  1.0132314  0.9235277  1.1358733  0.8424278  0.8515806 -0.8171403
## sample4 -0.9859036 -1.0816660  1.1358733 -1.1857971 -1.1730522 -0.8171403
## sample5 -0.9859036 -1.0816660  1.1358733  0.8424278  0.8515806  1.2224933
## sample6  1.0132314 -1.0816660  1.1358733  0.8424278 -1.1730522 -0.8171403
##              FP007      FP008      FP009      FP010      FP011     FP012
## sample1 -0.7558435  1.4372083 -0.6228265 -0.4663054 -0.5223076 -0.462962
## sample2  1.3216339  1.4372083 -0.6228265 -0.4663054  1.9125675 -0.462962
## sample3 -0.7558435  1.4372083 -0.6228265  2.1422621 -0.5223076 -0.462962
## sample4  1.3216339 -0.6950617 -0.6228265 -0.4663054 -0.5223076 -0.462962
## sample5 -0.7558435 -0.6950617  1.6038953 -0.4663054 -0.5223076 -0.462962
## sample6 -0.7558435 -0.6950617  1.6038953 -0.4663054 -0.5223076  2.157733
##              FP013      FP014     FP015      FP016      FP017      FP018
```

```
## sample1 -0.4461321 -0.4376388 0.4030145 -0.4135242 -0.4100338 -0.3888092
## sample2 -0.4461321 -0.4376388 0.4030145  2.4156953 -0.4100338  2.5692512
## sample3 -0.4461321 -0.4376388 0.4030145 -0.4135242  2.4362590 -0.3888092
## sample4 -0.4461321 -0.4376388 0.4030145 -0.4135242  2.4362590 -0.3888092
## sample5  2.2391315 -0.4376388 0.4030145  2.4156953 -0.4100338 -0.3888092
## sample6 -0.4461321 -0.4376388 0.4030145  2.4156953 -0.4100338 -0.3888092
##              FP019      FP020      FP021      FP022      FP023      FP024
## sample1  2.6815493 -0.3688597 -0.3706955 -0.3406978 -0.3743531  2.8070547
## sample2 -0.3725266 -0.3688597 -0.3706955 -0.3406978 -0.3743531 -0.3558707
## sample3 -0.3725266 -0.3688597 -0.3706955 -0.3406978 -0.3743531 -0.3558707
## sample4 -0.3725266 -0.3688597 -0.3706955 -0.3406978  2.6684658 -0.3558707
## sample5  2.6815493 -0.3688597 -0.3706955 -0.3406978 -0.3743531  2.8070547
## sample6 -0.3725266 -0.3688597  2.6947952 -0.3406978 -0.3743531 -0.3558707
##              FP025      FP026      FP027      FP028      FP029      FP030
## sample1 -0.361468  3.2978859 -0.3290557 -0.3445266 -0.3368437 -0.3211537
## sample2 -0.361468 -0.3029057 -0.3290557  2.8994814 -0.3368437 -0.3211537
## sample3  2.763587 -0.3029057 -0.3290557 -0.3445266 -0.3368437 -0.3211537
## sample4 -0.361468 -0.3029057 -0.3290557 -0.3445266 -0.3368437 -0.3211537
## sample5 -0.361468 -0.3029057 -0.3290557 -0.3445266 -0.3368437  3.1104998
## sample6 -0.361468 -0.3029057 -0.3290557 -0.3445266 -0.3368437 -0.3211537
##              FP031      FP032      FP033      FP034      FP035      FP036
## sample1 -0.3131281 -0.2817297 -0.2729429 -0.2945604 -0.2795515 -0.2860509
## sample2 -0.3131281 -0.2817297 -0.2729429 -0.2945604 -0.2795515 -0.2860509
## sample3 -0.3131281 -0.2817297 -0.2729429 -0.2945604 -0.2795515 -0.2860509
## sample4 -0.3131281 -0.2817297 -0.2729429 -0.2945604 -0.2795515 -0.2860509
## sample5 -0.3131281 -0.2817297 -0.2729429  3.3913200 -0.2795515 -0.2860509
## sample6 -0.3131281 -0.2817297 -0.2729429 -0.2945604 -0.2795515 -0.2860509
##              FP037      FP038      FP039      FP040      FP041      FP042
## sample1 -0.2751584 -0.3070214  3.5187119  3.6900455 -0.2593632 -0.2452292
## sample2 -0.2751584 -0.3070214 -0.2838961 -0.2707144 -0.2593632 -0.2452292
## sample3 -0.2751584  3.2536775 -0.2838961 -0.2707144 -0.2593632 -0.2452292
## sample4 -0.2751584 -0.3070214 -0.2838961 -0.2707144  3.8515432 -0.2452292
## sample5 -0.2751584 -0.3070214 -0.2838961 -0.2707144 -0.2593632 -0.2452292
```

```
## sample6 -0.2751584 -0.3070214 -0.2838961 -0.2707144 -0.2593632 -0.2452292
##              FP043      FP044      FP045      FP046      FP047      FP048
## sample1 -0.2662167 -0.2523714 -0.2428127 -0.6784872 -0.6017334 -0.3761751
## sample2  3.7523881 -0.2523714 -0.2428127  1.4723173  1.6601181 -0.3761751
## sample3 -0.2662167 -0.2523714  4.1140710 -0.6784872  1.6601181 -0.3761751
## sample4 -0.2662167 -0.2523714 -0.2428127 -0.6784872 -0.6017334 -0.3761751
## sample5 -0.2662167 -0.2523714 -0.2428127 -0.6784872 -0.6017334 -0.3761751
## sample6 -0.2662167 -0.2523714 -0.2428127 -0.6784872 -0.6017334 -0.3761751
##              FP049      FP050      FP051     FP052      FP053      FP054
## sample1 -0.3725266 -0.3558707 -0.3502245 -0.317157 -0.3211537 -0.2860509
## sample2 -0.3725266 -0.3558707  2.8523090 -0.317157 -0.3211537 -0.2860509
## sample3 -0.3725266 -0.3558707 -0.3502245 -0.317157 -0.3211537 -0.2860509
## sample4 -0.3725266 -0.3558707 -0.3502245 -0.317157 -0.3211537  3.4922050
## sample5 -0.3725266 -0.3558707 -0.3502245 -0.317157 -0.3211537 -0.2860509
## sample6 -0.3725266 -0.3558707 -0.3502245 -0.317157 -0.3211537 -0.2860509
##              FP055      FP056      FP057      FP058      FP059      FP060
## sample1 -0.2379224  3.7846575 -0.3688597 -0.3577418 -0.2403773 -0.9633427
## sample2 -0.2379224 -0.2639469 -0.3688597 -0.3577418 -0.2403773  1.0369606
## sample3 -0.2379224 -0.2639469 -0.3688597 -0.3577418 -0.2403773  1.0369606
## sample4 -0.2379224 -0.2639469 -0.3688597 -0.3577418 -0.2403773 -0.9633427
## sample5 -0.2379224 -0.2639469 -0.3688597 -0.3577418 -0.2403773 -0.9633427
## sample6 -0.2379224 -0.2639469 -0.3688597 -0.3577418 -0.2403773 -0.9633427
##             FP061      FP062      FP063      FP064      FP065      FP066
## sample1 -0.898407 -0.8813356  1.1604872 -0.8442531  0.8279183  0.7993638
## sample2 -0.898407 -0.8813356  1.1604872  1.1832335  0.8279183 -1.2496793
## sample3  1.111911  1.1334484 -0.8608009  1.1832335 -1.2065786  0.7993638
## sample4 -0.898407 -0.8813356 -0.8608009 -0.8442531 -1.2065786  0.7993638
## sample5 -0.898407 -0.8813356  1.1604872  1.1832335  0.8279183  0.7993638
## sample6 -0.898407  1.1334484  1.1604872  1.1832335 -1.2065786  0.7993638
##              FP067      FP068      FP069      FP070      FP071      FP072
## sample1  1.2777443 -0.7524132  1.3276594  1.3459950 -0.6967256 -1.3871296
## sample2  1.2777443  1.3276594 -0.7524132  1.3459950 -0.6967256  0.7201551
## sample3 -0.7818063 -0.7524132  1.3276594 -0.7421636 -0.6967256  0.7201551
```

```
## sample4 -0.7818063 -0.7524132  1.3276594  1.3459950 -0.6967256 -1.3871296
## sample5  1.2777443  1.3276594  1.3276594 -0.7421636 -0.6967256 -1.3871296
## sample6  1.2777443  1.3276594  1.3276594 -0.7421636 -0.6967256  0.7201551
##              FP073      FP074      FP075      FP076      FP077      FP078
## sample1 -0.6702404 -0.6933991 -0.7151122  1.4303578 -0.6867603 -0.6603763
## sample2  1.4904330  1.4406544  1.3969116  1.4303578  1.4545809  1.5126959
## sample3  1.4904330 -0.6933991 -0.7151122 -0.6983906 -0.6867603 -0.6603763
## sample4 -0.6702404 -0.6933991 -0.7151122 -0.6983906  1.4545809 -0.6603763
## sample5 -0.6702404 -0.6933991  1.3969116 -0.6983906 -0.6867603 -0.6603763
## sample6 -0.6702404 -0.6933991  1.3969116 -0.6983906 -0.6867603 -0.6603763
##              FP079      FP080      FP081      FP082      FP083      FP084
## sample1  0.6702404 -0.6587354 -0.6212014  0.6325882 -0.6130829  1.5791449
## sample2  0.6702404  1.5164639 -0.6212014  0.6325882 -0.6130829  1.5791449
## sample3  0.6702404 -0.6587354  1.6080912  0.6325882 -0.6130829 -0.6325882
## sample4  0.6702404 -0.6587354  1.6080912 -1.5791449 -0.6130829 -0.6325882
## sample5  0.6702404  1.5164639 -0.6212014  0.6325882  1.6293857  1.5791449
## sample6 -1.4904330  1.5164639 -0.6212014  0.6325882 -0.6130829 -0.6325882
##             FP085      FP086      FP087      FP088      FP089      FP090
## sample1 -0.585542 -0.6065954  0.6130829 -0.5968738  1.7445930 -0.5758331
## sample2  1.706024 -0.6065954  0.6130829  1.6736342  1.7445930  1.7347883
## sample3 -0.585542 -0.6065954  0.6130829 -0.5968738 -0.5725969 -0.5758331
## sample4 -0.585542  1.6468118  0.6130829 -0.5968738 -0.5725969  1.7347883
## sample5 -0.585542  1.6468118  0.6130829 -0.5968738 -0.5725969 -0.5758331
## sample6 -0.585542 -0.6065954 -1.6293857 -0.5968738 -0.5725969 -0.5758331
##              FP091      FP092      FP093      FP094      FP095      FP096
## sample1  1.8548056 -0.5677421 -0.5677421 -0.5483075 -0.5304479 -0.5271937
## sample2  1.8548056 -0.5677421  1.7595110 -0.5483075 -0.5304479 -0.5271937
## sample3 -0.5385731 -0.5677421 -0.5677421 -0.5483075 -0.5304479 -0.5271937
## sample4 -0.5385731 -0.5677421  1.7595110 -0.5483075 -0.5304479 -0.5271937
## sample5  1.8548056  1.7595110 -0.5677421  1.8218763 -0.5304479 -0.5271937
## sample6 -0.5385731  1.7595110 -0.5677421 -0.5483075 -0.5304479 -0.5271937
##              FP097     FP098      FP099      FP100      FP101      FP102
## sample1  1.8005896 -0.558029 -0.5418196 -0.5483075  1.7953478 -0.5026912
```

```
## sample2  1.8005896 -0.558029 -0.5418196 -0.5483075  1.7953478  1.9872011
## sample3 -0.5547896  1.790137 -0.5418196 -0.5483075 -0.5564094 -0.5026912
## sample4 -0.5547896 -0.558029 -0.5418196 -0.5483075 -0.5564094 -0.5026912
## sample5 -0.5547896 -0.558029 -0.5418196 -0.5483075  1.7953478  1.9872011
## sample6 -0.5547896 -0.558029 -0.5418196  1.8218763 -0.5564094  1.9872011
##              FP103      FP104     FP105      FP106      FP107      FP108
## sample1 -0.5288211  1.8660610 -0.523937 -0.4862326 -0.5174149 -0.5076074
## sample2 -0.5288211  1.8660610  1.906619  2.0544663 -0.5174149 -0.5076074
## sample3 -0.5288211 -0.5353247 -0.523937 -0.4862326 -0.5174149 -0.5076074
## sample4  1.8890101 -0.5353247 -0.523937 -0.4862326 -0.5174149 -0.5076074
## sample5 -0.5288211  1.8660610 -0.523937  2.0544663  1.9306527 -0.5076074
## sample6 -0.5288211 -0.5353247 -0.523937 -0.4862326  1.9306527 -0.5076074
##             FP109      FP110      FP111      FP112      FP113      FP114
## sample1 -0.462962 -0.5092442 -0.4944765 -0.4911827 -0.4928302 -0.4290863
## sample2  2.157733 -0.5092442 -0.4944765 -0.4911827  2.0269629 -0.4290863
## sample3 -0.462962 -0.5092442 -0.4944765 -0.4911827 -0.4928302 -0.4290863
## sample4 -0.462962 -0.5092442 -0.4944765 -0.4911827 -0.4928302 -0.4290863
## sample5  2.157733 -0.5092442  2.0202142  2.0337617 -0.4928302  2.3280830
## sample6 -0.462962  1.9616295  2.0202142 -0.4911827 -0.4928302 -0.4290863
##              FP115      FP116      FP117      FP118      FP119      FP120
## sample1 -0.4663054  2.0475127 -0.4663054 -0.4878839 -0.4410429 -0.4461321
## sample2 -0.4663054  2.0475127 -0.4663054  2.0475127 -0.4410429 -0.4461321
## sample3 -0.4663054 -0.4878839 -0.4663054 -0.4878839 -0.4410429 -0.4461321
## sample4 -0.4663054 -0.4878839 -0.4663054 -0.4878839 -0.4410429 -0.4461321
## sample5 -0.4663054  2.0475127 -0.4663054 -0.4878839 -0.4410429 -0.4461321
## sample6 -0.4663054 -0.4878839 -0.4663054 -0.4878839 -0.4410429 -0.4461321
##              FP121      FP122      FP123     FP124      FP125      FP126
## sample1 -0.4030145 -0.4427415 -0.4478242 -0.439342 -0.4290863 -0.4170024
## sample2  2.4786909 -0.4427415 -0.4478242 -0.439342 -0.4290863  2.3955459
## sample3 -0.4030145 -0.4427415 -0.4478242 -0.439342 -0.4290863 -0.4170024
## sample4 -0.4030145 -0.4427415  2.2306713 -0.439342 -0.4290863 -0.4170024
## sample5  2.4786909 -0.4427415 -0.4478242 -0.439342 -0.4290863 -0.4170024
## sample6  2.4786909 -0.4427415  2.2306713 -0.439342 -0.4290863 -0.4170024
```

```
##              FP127      FP128      FP129      FP130      FP131      FP132
## sample1 -0.4030145 -0.4170024 -0.4012514 -0.3426153 -0.3798059 -0.3779927
## sample2 -0.4030145 -0.4170024 -0.4012514 -0.3426153  2.6301555 -0.3779927
## sample3 -0.4030145 -0.4170024 -0.4012514 -0.3426153 -0.3798059 -0.3779927
## sample4 -0.4030145 -0.4170024 -0.4012514 -0.3426153 -0.3798059 -0.3779927
## sample5  2.4786909 -0.4170024 -0.4012514 -0.3426153  2.6301555 -0.3779927
## sample6 -0.4030145 -0.4170024 -0.4012514 -0.3426153 -0.3798059 -0.3779927
##              FP133      FP134      FP135      FP136     FP137      FP138
## sample1 -0.3798059 -0.3816148 -0.3798059  2.6947952 -0.361468 -0.3539942
## sample2 -0.3798059  2.6176881  2.6301555 -0.3706955  2.763587  2.8219347
## sample3  2.6301555 -0.3816148 -0.3798059 -0.3706955 -0.361468 -0.3539942
## sample4 -0.3798059 -0.3816148 -0.3798059 -0.3706955 -0.361468 -0.3539942
## sample5 -0.3798059 -0.3816148 -0.3798059  2.6947952  2.763587  2.8219347
## sample6  2.6301555  2.6176881 -0.3798059 -0.3706955  2.763587 -0.3539942
##              FP139      FP140      FP141      FP142      FP143      FP144
## sample1 -0.2987526 -0.3539942 -0.3633234 -0.3502245 -0.2966615 -0.3406978
## sample2 -0.2987526 -0.3539942  2.7494745  2.8523090 -0.2966615 -0.3406978
## sample3 -0.2987526 -0.3539942 -0.3633234 -0.3502245 -0.2966615 -0.3406978
## sample4 -0.2987526 -0.3539942 -0.3633234 -0.3502245 -0.2966615 -0.3406978
## sample5 -0.2987526 -0.3539942 -0.3633234  2.8523090 -0.2966615  2.9320660
## sample6 -0.2987526 -0.3539942  2.7494745 -0.3502245 -0.2966615 -0.3406978
##              FP145     FP146      FP147      FP148      FP149     FP150
## sample1 -0.3406978 -0.338774 -0.3426153 -0.3090656 -0.3151467  3.415804
## sample2 -0.3406978 -0.338774 -0.3426153 -0.3090656 -0.3151467 -0.292449
## sample3 -0.3406978 -0.338774 -0.3426153 -0.3090656 -0.3151467 -0.292449
## sample4 -0.3406978  2.948717 -0.3426153 -0.3090656 -0.3151467 -0.292449
## sample5 -0.3406978 -0.338774 -0.3426153 -0.3090656 -0.3151467 -0.292449
## sample6 -0.3406978  2.948717 -0.3426153 -0.3090656 -0.3151467 -0.292449
##              FP151      FP152      FP153      FP154      FP155      FP156
## sample1 -0.2428127 -0.2987526 -0.2903271 -0.1982496 -0.2729429 -0.2751584
## sample2 -0.2428127 -0.2987526  3.4407687  5.0388432 -0.2729429 -0.2751584
## sample3 -0.2428127 -0.2987526 -0.2903271 -0.1982496 -0.2729429 -0.2751584
## sample4 -0.2428127 -0.2987526 -0.2903271 -0.1982496  3.6599168 -0.2751584
```

```
## sample5 -0.2428127 -0.2987526 -0.2903271 -0.1982496 -0.2729429 -0.2751584
## sample6 -0.2428127 -0.2987526 -0.2903271 -0.1982496 -0.2729429 -0.2751584
##              FP157      FP158      FP159      FP160      FP161      FP162
## sample1 -0.2570485 -0.2379224 -0.2751584 -0.2707144 -0.2662167  1.0089780
## sample2 -0.2570485 -0.2379224  3.6304485 -0.2707144 -0.2662167  1.0089780
## sample3 -0.2570485 -0.2379224 -0.2751584 -0.2707144 -0.2662167  1.0089780
## sample4 -0.2570485 -0.2379224 -0.2751584 -0.2707144 -0.2662167 -0.9900597
## sample5 -0.2570485 -0.2379224 -0.2751584 -0.2707144 -0.2662167 -0.9900597
## sample6 -0.2570485 -0.2379224 -0.2751584 -0.2707144 -0.2662167 -0.9900597
##              FP163      FP164      FP165      FP166      FP167      FP168
## sample1 -0.9532482  0.7696371 -0.7319733 -0.7033934 -0.6983906  0.7084077
## sample2 -0.9532482  0.7696371 -0.7319733  1.4201847 -0.6983906  0.7084077
## sample3  1.0479417  0.7696371  1.3647334 -0.7033934  1.4303578  0.7084077
## sample4 -0.9532482  0.7696371  1.3647334 -0.7033934 -0.6983906  0.7084077
## sample5 -0.9532482 -1.2979474 -0.7319733 -0.7033934 -0.6983906  0.7084077
## sample6  1.0479417 -1.2979474  1.3647334 -0.7033934  1.4303578 -1.4101322
##              FP169      FP170      FP171      FP172      FP173      FP174
## sample1 -0.4779558 -0.4746349 -0.4512019 -0.4221979 -0.4065307 -0.3870166
## sample2  2.0900438  2.1046669 -0.4512019 -0.4221979  2.4572522 -0.3870166
## sample3 -0.4779558 -0.4746349  2.2139721 -0.4221979 -0.4065307 -0.3870166
## sample4 -0.4779558  2.1046669 -0.4512019 -0.4221979  2.4572522 -0.3870166
## sample5 -0.4779558 -0.4746349 -0.4512019  2.3660671 -0.4065307 -0.3870166
## sample6 -0.4779558 -0.4746349  2.2139721 -0.4221979 -0.4065307 -0.3870166
##              FP175      FP176      FP177      FP178      FP179      FP180
## sample1 -0.3941637  2.6815493 -0.3706955 -0.3706955  3.0358039 -0.3464318
## sample2 -0.3941637 -0.3725266 -0.3706955  2.6947952 -0.3290557 -0.3464318
## sample3 -0.3941637 -0.3725266 -0.3706955 -0.3706955 -0.3290557  2.8835355
## sample4 -0.3941637 -0.3725266 -0.3706955 -0.3706955  3.0358039 -0.3464318
## sample5 -0.3941637  2.6815493  2.6947952 -0.3706955 -0.3290557 -0.3464318
## sample6 -0.3941637 -0.3725266 -0.3706955 -0.3706955 -0.3290557 -0.3464318
##              FP181      FP182      FP183      FP184      FP185      FP186
## sample1 -0.3211537 -0.331013 -0.2860509 -0.3029057 -0.3049681 -0.2881945
## sample2 -0.3211537 -0.331013 -0.2860509 -0.3029057  3.2755834 -0.2881945
```

```
## sample3 -0.3211537  3.017853 -0.2860509 -0.3029057 -0.3049681 -0.2881945
## sample4 -0.3211537 -0.331013 -0.2860509 -0.3029057  3.2755834 -0.2881945
## sample5 -0.3211537 -0.331013 -0.2860509 -0.3029057 -0.3049681 -0.2881945
## sample6 -0.3211537  3.017853 -0.2860509 -0.3029057 -0.3049681 -0.2881945
##              FP187      FP188      FP189      FP190      FP191      FP192
## sample1 -0.2795515 -0.2707144 -0.2881945 -0.2795515 -0.2751584 -0.2547181
## sample2 -0.2795515 -0.2707144  3.4662300 -0.2795515 -0.2751584 -0.2547181
## sample3 -0.2795515 -0.2707144 -0.2881945 -0.2795515 -0.2751584 -0.2547181
## sample4 -0.2795515  3.6900455 -0.2881945 -0.2795515 -0.2751584 -0.2547181
## sample5 -0.2795515 -0.2707144 -0.2881945 -0.2795515 -0.2751584  3.9217804
## sample6 -0.2795515 -0.2707144 -0.2881945 -0.2795515  3.6304485  3.9217804
##              FP193      FP194      FP195      FP196      FP197      FP198
## sample1 -0.2570485 -0.2500081 -0.2547181 -0.2452292 -0.2354475 -0.2452292
## sample2 -0.2570485 -0.2500081 -0.2547181 -0.2452292 -0.2354475 -0.2452292
## sample3 -0.2570485 -0.2500081 -0.2547181 -0.2452292 -0.2354475 -0.2452292
## sample4 -0.2570485 -0.2500081 -0.2547181 -0.2452292  4.2427646 -0.2452292
## sample5 -0.2570485 -0.2500081 -0.2547181 -0.2452292 -0.2354475 -0.2452292
## sample6 -0.2570485 -0.2500081 -0.2547181 -0.2452292 -0.2354475 -0.2452292
##              FP199      FP200      FP201      FP202      FP203     FP204
## sample1 -0.2227479 -0.2278958  4.2427646 -0.5887789 -0.3596075 -0.331013
## sample2 -0.2227479 -0.2278958 -0.2354475  1.6966445 -0.3596075 -0.331013
## sample3 -0.2227479 -0.2278958 -0.2354475 -0.5887789 -0.3596075 -0.331013
## sample4 -0.2227479 -0.2278958 -0.2354475 -0.5887789 -0.3596075 -0.331013
## sample5 -0.2227479 -0.2278958 -0.2354475 -0.5887789 -0.3596075 -0.331013
## sample6 -0.2227479 -0.2278958 -0.2354475 -0.5887789 -0.3596075 -0.331013
##              FP205      FP206      FP207      FP208  MolWeight  NumAtoms
## sample1 -0.2903271 -0.2523714 -0.2452292 -0.3558707  0.3029057 0.4231289
## sample2 -0.2903271 -0.2523714 -0.2452292 -0.3558707  1.4764480 1.6150444
## sample3 -0.2903271 -0.2523714 -0.2452292 -0.3558707  0.2831012 0.7711777
## sample4 -0.2903271 -0.2523714 -0.2452292 -0.3558707 -0.5803295 0.2667698
## sample5 -0.2903271 -0.2523714 -0.2452292 -0.3558707  0.5074082 0.6385252
## sample6 -0.2903271 -0.2523714 -0.2452292 -0.3558707  0.8460184 0.7058566
##         NumNonHAtoms  NumBonds NumNonHBonds NumMultBonds NumRotBonds
```

```
## sample1    0.6249705 0.5279098    0.7350798    1.5526037  -1.2878683
## sample2    1.6412036 1.6263559    1.7132483    1.2171732   0.9514166
## sample3    0.4917979 0.7171066    0.3895802    0.3841309   0.9514166
## sample4   -0.3312814 0.2449537   -0.3237944   -0.6934370  -0.3234608
## sample5    0.4917979 0.5929367    0.3895802    0.2114279   1.2050890
## sample6    0.4917979 0.5929367    0.2628641   -0.6934370   1.2050890
##         NumDblBonds NumAromaticBonds NumHydrogen    NumCarbon NumNitrogen
## sample1  -1.0420314        1.3699017   0.1408985  0.865919125   1.2921297
## sample2  -1.0420314        1.1322224   1.3698737  1.793411403   1.5021712
## sample3   0.4422601        0.5837596   0.8747571  0.709874352  -0.8426365
## sample4   1.0637813       -1.1402982   0.6509474  0.189116098  -0.8426365
## sample5  -1.0420314        0.5837596   0.6509474 -0.006528268   1.6919136
## sample6   1.0637813       -1.1402982   0.7650294  0.189116098   0.8249291
##           NumOxygen  NumSulfer NumChlorine NumHalogen    NumRings
## sample1 -1.19031762 -0.3801587   -0.513403 -0.6177483  1.21087398
## sample2 -0.08583615  2.4851927   -0.513403 -0.6177483  1.62515973
## sample3  0.56024408 -0.3801587   -0.513403 -0.6177483 -0.07601512
## sample4 -1.19031762 -0.3801587   -0.513403 -0.6177483 -0.07601512
## sample5 -1.19031762 -0.3801587    1.602691  1.3110648 -0.07601512
## sample6 -0.08583615  2.4851927    1.994560  1.6682524 -1.36290421
##         HydrophilicFactor SurfaceArea1 SurfaceArea2
## sample1       -1.10799299    0.0232306  -0.05757619
## sample2        0.01101117    0.6756205   1.05909379
## sample3        0.06524309    0.3410369   0.24911257
## sample4       -1.84451025   -1.4877541  -1.51570334
## sample5        0.36666813    0.7110417   0.60617374
## sample6       -0.47676967   -0.1571677   0.43704574
```

```
cor_mat <- cor(solubility_scale) # 计算相关系数矩阵
```

```
rs_mat <- eigen(cor_mat) # 特征分解
```

```
val <- rs_mat$values # 提取特征值，即各主成分的方差
standard_deviation <- sqrt(val) # 换算成标准差
```

```
proportion_of_variance <- val/sum(val) # 计算方差贡献率
```

```
cumulative_proportion <- cumsum(proportion_of_variance) # 计算累积贡献率
```

```
load_mat <- as.matrix(rs_mat$vectors) # 提取特征向量，即载荷矩阵 (loadings)
PC <- solubility_scale %*% load_mat # 计算主成分得分
colnames(PC) <- paste("PC",1:ncol(PC),sep = "")
df2 <- as.data.frame(PC) # 转换成数据框，否则直接用于绘图会报错
head(df2)
```

```
##                PC1        PC2        PC3        PC4        PC5        PC6
## sample1  1.2694206 -5.3657924  0.6175658  5.2399136 -2.911619  6.0059470
## sample2  8.3874611  0.6970092 -1.3084024  7.8380490 -5.236612 -2.1872463
## sample3 -0.9198851  1.5990612 -1.5430451 -5.1541645 -2.861326  0.2374496
## sample4 -3.6829643  4.6154670 -3.1335564  4.3047875  2.123944  2.3759472
## sample5  3.5515346 -2.8464702  4.2713495  5.3030045  1.444973 -3.2543831
## sample6 -0.1537242  1.6594926  3.0966348  0.1969316  2.078825 -4.5919788
##                PC7        PC8        PC9       PC10       PC11       PC12
## sample1 -1.0522250 -1.5002023 -2.8554682  4.6542856  2.2131367  1.8153286
## sample2  2.7739096 -3.6812494 -0.3110280 -4.1421323 -0.5032044  1.1481701
## sample3 -0.7202474  0.6296565 -0.9335428  0.4977264  1.3306148  2.4245263
## sample4 -3.1529114  4.1477724  1.2442982  1.5278381  1.4157515 -1.4342115
## sample5  3.5555396 -2.7536109 -0.6409781 -0.3444174  5.3577477 -0.3645779
## sample6  2.5725830  5.2997728 -1.0649177 -2.2590438  2.0861903  0.4478980
##               PC13       PC14       PC15       PC16       PC17       PC18
## sample1 -0.2345638 -2.111890 -2.2882797  1.6131464 -1.4998239  2.0882291
## sample2  4.3078723  3.030211  1.3710004 -0.5328720 -0.3046771 -0.9306152
## sample3  4.7320509 -3.031260 -1.2384777 -0.2828171 -0.1764809  1.9079025
## sample4  1.1416009  2.002426  0.1152025 -1.3798823 -1.6109361  2.6688729
## sample5 -1.2729930 -2.510307  1.4404151  1.5655452 -3.2530991  3.5123799
```

```
## sample6  2.4928129 -2.616129  0.8180277  4.9494715 -1.6270164  1.8581708
##               PC19        PC20       PC21       PC22       PC23       PC24
## sample1 -2.1977777 -0.11151933 -1.0451753  0.2796904  0.3236148 -0.8123828
## sample2  2.0827005 -0.05304995  3.3961330 -1.1407604 -2.0320293 -0.9164224
## sample3 -0.5877576  1.38445238  1.9027526 -0.7228859 -0.1537378 -0.3569579
## sample4  0.9270959  1.27145731 -1.8355022 -2.2834133  0.2198965 -1.0567022
## sample5  0.7080496 -0.05880086 -0.6266060 -0.8906590  3.3536115  2.5430382
## sample6  1.9249279 -1.93326103 -0.3735039 -1.3662382  1.2527806 -1.1909781
##                PC25      PC26       PC27       PC28        PC29       PC30
## sample1 -0.72951592  1.801825 -0.4579442 -0.2348385  0.01006871 -0.5666166
## sample2 -3.30851200  1.422790 -1.7475194  2.8356572 -3.09619288 -2.6474152
## sample3  1.73174205 -1.067207 -0.8555088 -0.6605529  2.69665399  0.3765504
## sample4 -0.06666869  1.154089 -1.9238374 -0.2780025  1.87804194 -1.0825738
## sample5  0.82685445 -1.038052  1.7338718  0.4927458  1.47361504  0.7131333
## sample6  0.69660515  1.167668 -0.5049861 -0.6056061 -0.46764390  1.9524159
##               PC31       PC32        PC33       PC34       PC35        PC36
## sample1 -1.5047710 -0.5291281 -0.13394933 -0.1351282 -0.8280683 -0.52754495
## sample2 -1.1893183  1.1482798  0.03944673 -1.5697685 -0.5082639  2.34753534
## sample3  1.8605183  0.6069951  0.01966110  0.5197459 -0.1699445 -0.36133199
## sample4 -0.1175672 -1.3410878  1.01424221 -0.1441377  2.7637686  0.86736688
## sample5  1.3126875  0.3489088  0.27018026 -2.6906358 -0.6139737 -1.25848161
## sample6  2.5701904 -0.6632748 -2.32196254  3.6856783 -2.8955529  0.03434873
##               PC37       PC38       PC39       PC40        PC41       PC42
## sample1 -0.2563998  0.2928151 -0.6455751 -2.1968023  0.47716784  0.4233815
## sample2  0.8665074 -1.7186716 -0.8980348  0.1497987  0.33089231 -1.0467270
## sample3 -2.6822698  0.7422047 -1.4009577 -0.9988700 -1.85171838 -1.2433960
## sample4 -0.4477682 -0.5674026  0.2576623  1.0890120  2.50626143 -0.8458283
## sample5 -0.3563773 -0.3209598 -1.5245979 -0.2262821  1.88792157  0.1636425
## sample6  1.3071588 -0.2705068 -1.8766115 -0.3441314 -0.08368922 -1.2622181
##                PC43       PC44        PC45        PC46       PC47       PC48
## sample1 -0.371607668 -1.0521128  1.05736871  0.45183828 -2.3605530  0.2403872
## sample2  0.003616634  0.7128949 -0.59041007  0.55428957  1.4445878  1.2243371
## sample3  1.527003133  2.1365443 -0.07840723 -0.02403677  0.3628552 -0.6492186
```

```
## sample4  0.577232473  2.5901335  1.19768458  0.46046964  0.2574770 -2.4520553
## sample5  0.595296761 -1.1732357 -3.19703021 -1.11730395  1.5724897  0.1855018
## sample6 -0.465989475  0.4688734  0.83282272 -0.51640658 -0.3187173  0.3519745
##                PC49        PC50       PC51       PC52       PC53       PC54
## sample1 -0.4141325  0.29635240 -0.1202531  0.3673250  0.2830145 -0.6014879
## sample2  0.4513531 -1.01671789  0.8185976 -1.5074816  0.1636388  1.4532025
## sample3 -1.1114677  0.07784489  0.4701327  0.6656309 -0.7929400 -0.9681848
## sample4  0.8141405  1.34486751  0.7331720  0.1953666 -0.2995166  0.2929549
## sample5 -0.6263616 -0.46372807 -0.6093917  0.9917942  0.8720498  2.4382058
## sample6  1.5412128 -0.69482714 -0.2029589  0.1786412 -0.9950408  0.1598344
##                  PC55       PC56       PC57       PC58        PC59       PC60
## sample1  0.3131913921 -0.4304507  0.6193920 -0.1447021  0.08261858 -0.1395405
## sample2 -0.4667972661  2.7507289 -1.7737673 -0.3640741 -0.48702908  1.1294474
## sample3  0.0738410796  0.5312596 -0.1060027  0.8569724  0.78703972  0.3837932
## sample4 -0.0336069041 -0.3414716  1.0785059  0.8146205  0.85680557 -1.2214237
## sample5 -0.0007358528 -1.7756858  0.3568896  0.5638811 -1.11591731 -2.3331858
## sample6  0.3462825648  1.3529406  1.0156382  0.1043114 -1.07709821 -0.2637875
##                PC61        PC62        PC63        PC64        PC65        PC66
## sample1 -0.1331146 -0.08577211 -0.14979144 -0.23222272  0.09340513 -0.9293223
## sample2  0.7226627  2.56038730  0.38438991 -1.01272474 -0.35369354 -1.0616550
## sample3  0.3381337  0.59711672  0.03957629 -0.35256922 -0.59246115 -0.7685623
## sample4  0.3257030  0.06098446  0.75721858 -0.48748005 -0.39677863  0.6029204
## sample5 -0.4115536 -0.44657344  1.30671198  0.21719579 -0.69311907 -0.3894978
## sample6 -0.4923590  0.30771357  1.05288550  0.06851074  0.91790649 -0.2523001
##                 PC67        PC68       PC69       PC70       PC71       PC72
## sample1  0.58916176  1.39537019  0.2581953 -0.3364049  0.1875460  0.6075556
## sample2 -0.68247433 -0.02058298 -0.4529908 -0.4428880 -0.5626102 -0.4371086
## sample3  0.06880027  0.74146368  0.1564194  0.3816624  0.3690125 -0.5768268
## sample4 -0.75905144 -0.39626992 -0.4344205 -0.6482295 -0.8953195  1.1096970
## sample5 -0.29778523 -0.67641516 -0.1802073  1.2817679  0.2888770 -0.7922433
## sample6 -0.03613555  0.62836130 -0.7132409 -0.4714374 -0.5471414 -0.4269103
##                PC73        PC74       PC75        PC76       PC77       PC78
## sample1  0.3453766  0.08869562 -0.4251439  0.18094133 -1.1669354 -0.8496820
```

```
## sample2  0.4929003 -0.83348420  0.4332404 -1.21237766 -0.0164705  1.0500717
## sample3  0.1073487  0.06291839 -0.5332181 -0.59438806 -0.3134421  0.2353629
## sample4  0.2373750  0.43550649 -1.1140238  1.13942690 -0.4560507 -0.2027549
## sample5 -0.1524531 -0.12721422 -0.5812511  1.24837027  1.4976735  0.9729479
## sample6  1.3257088 -1.72047840 -0.1771071 -0.08635329  0.1514961  0.2342661
##                PC79         PC80        PC81       PC82        PC83       PC84
## sample1  0.580862454 -0.654548687  0.61339272  0.1361594  0.39002890  0.7559264
## sample2 -0.003451639  0.003810793 -0.09303312 -0.9728017 -0.21127214 -0.3615314
## sample3  0.276624716 -0.490195216 -0.54075524 -0.6691073 -0.01033889 -0.1205954
## sample4  1.212208004 -1.027021874  0.46059485 -0.8546908  0.52443526  0.3990114
## sample5  0.666248171  0.525202891 -0.45764224 -0.1548452 -0.96865975  0.6198562
## sample6  0.153418809  0.663256033  0.73831099  0.2365774 -0.04845816 -1.1733536
##               PC85        PC86       PC87        PC88        PC89       PC90
## sample1 -0.9723245  1.07178120 -0.1621062 -0.31580292 -0.08174485 -0.8529082
## sample2  0.1307960 -0.90435920  0.4709655  0.26897354 -0.44134871  0.7929090
## sample3  0.2048896  0.14998956  0.4418199  0.40015937  0.16577868  0.3372637
## sample4  0.6111901  0.62098534 -0.8811867  0.52827786  0.43981084  0.7332558
## sample5 -1.5172713 -0.70919450  0.1015315 -0.06823913 -0.36373410  1.0184285
## sample6 -0.2114733  0.01217484  0.7820721  0.43080506  1.22406597 -0.1080414
##                PC91        PC92        PC93       PC94        PC95       PC96
## sample1 -0.42837530  0.23213647 -0.14275894 -0.2668030  0.20277139 -0.8865963
## sample2 -0.73797560  0.01936754  0.31990212  0.6814386 -0.71897856  0.2804439
## sample3 -0.42530707 -0.07286227 -0.03311724  0.6431753  0.23796884  0.2406017
## sample4 -0.06953859 -0.38980792  0.94589911 -1.0271559 -0.05318057 -0.1866804
## sample5 -0.36117025 -0.04325648  0.34150813  0.7153412 -0.08744305  0.6675815
## sample6  1.05917105  0.63331208 -0.05189331  0.6315124  0.35777018 -0.2169947
##                PC97        PC98        PC99       PC100       PC101       PC102
## sample1  0.21888935 -0.25543202  0.06499145  0.79800928  0.82020248 -0.24118610
## sample2 -0.41230295  0.65651613 -0.59350192  1.48946632  0.54355221 -0.44430922
## sample3  0.06298349 -0.07656307  0.36904988 -0.06169366 -0.09524296  0.17333661
## sample4  0.13820151  0.72602966 -0.03043626  0.42022095  0.01215718 -0.35440298
## sample5 -0.12096271 -0.34127951 -0.12804478  0.37456695 -0.44522473  0.07308107
## sample6  0.20663385  0.06876612  0.48103040 -0.19688752  0.91851748 -0.51069504
```

```
##               PC103       PC104       PC105      PC106      PC107       PC108
## sample1  0.08716340 -0.60711727 -0.15542510 -0.2351459  1.1679299  0.59535144
## sample2  0.29284268 -0.30681669 -0.79877119 -0.2335290 -0.1281978 -0.30893277
## sample3 -0.25494683 -0.08320065  0.04012999 -0.4015807 -0.1497268 -0.04601995
## sample4  0.05958842 -0.73682788 -0.35028824 -0.5340745 -0.2717294 -0.36722530
## sample5 -0.30353296  0.17850942 -0.37723733  0.1263347 -0.1084617  0.50348237
## sample6  0.64152137  0.32555848 -0.75991235  0.5453686 -0.5508419  0.01323722
##               PC109      PC110       PC111       PC112       PC113       PC114
## sample1 -0.2297822  0.1035967 -0.31949836 -0.62955446  0.03404321 -0.69295877
## sample2  0.1894283  0.6899519 -0.66586929 -0.52605000 -0.53308297  0.72645772
## sample3 -0.1839721 -0.2100392 -0.03046837  0.14289129 -0.05410907 -0.04327117
## sample4 -0.1430421 -0.3111750 -0.59851472 -1.36574658 -0.11161047  0.46284430
## sample5  0.4106587  0.1386746 -0.81624859 -0.03924318  0.87967000 -0.36697526
## sample6 -0.4052429 -0.3544536  0.11940692  0.21007742  0.48024329 -0.03875377
##               PC115      PC116       PC117       PC118        PC119       PC120
## sample1 -0.5478781 -0.3474568 -0.49626490  0.05810936 -0.410644655  0.06364514
## sample2  0.3247837  0.2974791  0.32930140  0.03925291 -0.516375973 -0.02125443
## sample3 -0.1772355  0.1898332 -0.18524570 -0.10828229  0.008482775  0.31958202
## sample4  0.2823670  0.2700825  0.56207528  0.60638115 -0.915308168  0.03468492
## sample5  0.5628660 -0.2412158  0.02334069 -0.39723268  0.033259493 -0.14392268
## sample6  0.4957346 -0.8548941 -0.54577143 -0.63852582 -0.341793733 -0.16714598
##               PC121       PC122      PC123        PC124       PC125       PC126
## sample1 -0.1386420 -0.08403621 -0.4741827  0.499059110 -0.34127859 -0.06909438
## sample2 -0.3266692  0.05488123 -0.1214767  0.740198699 -0.30588988  0.26250090
## sample3  0.2998225 -0.25718387  0.1970852 -0.391186995  0.18889597  0.11230358
## sample4 -0.1609227  0.11357609 -0.5001921 -0.004898104  0.02504779  0.01849996
## sample5  0.4652347  0.27320984  0.4332709 -0.647046757  0.31238009  0.28445250
## sample6 -0.1913016 -0.36894255 -0.2775792  0.055507622 -0.73537611  0.17693780
##               PC127      PC128       PC129       PC130       PC131       PC132
## sample1  0.4203525  0.3654229 -0.01220597 -0.27677659  0.41349732  0.08191633
## sample2  0.4555153 -0.1729656  0.10587851 -0.34294888 -0.59721170 -0.33905129
## sample3 -0.1183804 -0.3970052  0.13448301 -0.24917886  0.33761175  0.03352511
## sample4  0.2988233 -0.4026436  0.37089071 -0.13077134 -0.07608759 -0.28676829
```

```
## sample5  0.2482501 -0.5315186  0.68312809 -0.07696628 -0.33859587  0.38730533
## sample6 -0.3657624  0.1672607  0.17170299  0.22333922 -0.32908254 -0.01564231
##                PC133      PC134       PC135      PC136       PC137        PC138
## sample1 -0.203745184  0.5485440  0.05776016  0.3510750 -0.10567119  0.205633677
## sample2  0.082066353 -0.1526878 -0.44236308  0.1866747 -0.30067961  0.197198160
## sample3  0.203279450  0.6289860 -0.04013041 -0.1582930  0.14461695 -0.008672187
## sample4  0.442853656 -0.5285597 -0.21899920  0.2930086 -0.10166378 -0.327516846
## sample5 -0.459439082  0.1887271 -0.11687116 -0.1873848  0.18145523  0.504645439
## sample6 -0.009790492 -0.2043078 -0.06594385 -0.1853478  0.06944065 -0.337417975
##                PC139       PC140       PC141      PC142      PC143       PC144
## sample1 -0.002511209 -0.12812992  0.17145762  0.1565169 -0.3496628 -0.23080698
## sample2  0.289512114  0.16700794 -0.27572041 -0.2050758  0.5654675  0.12606580
## sample3 -0.185973887  0.20845288  0.22920457  0.2871368 -0.1864359  0.26528375
## sample4 -0.042840378 -0.06134026 -0.32990957 -0.1806655  0.1682958 -0.22999285
## sample5 -0.397071195  0.44512253 -0.48980505  0.5142559  0.1551446  0.21279949
## sample6  0.150637286  0.13199452  0.06606232 -0.1355257  0.4818915 -0.02910166
##                PC145      PC146       PC147       PC148       PC149        PC150
## sample1 -0.09388036  0.4038884 -0.00728924 -0.15747098  0.09026213 -0.007330468
## sample2 -0.24476108 -0.1020893 -0.38716447 -0.31218401 -0.22437126  0.240181789
## sample3  0.01771809 -0.1306763  0.08735626  0.25818163  0.30376026 -0.314038357
## sample4  0.19405496  0.3573481  0.34457010 -0.27889139 -0.02320762  0.236630030
## sample5  0.22706322 -0.1046214  0.27798596  0.08775175  0.50265687  0.165148830
## sample6  0.70595037 -0.2894139  0.01284034  0.30627252  0.14218488  0.093512278
##                PC151       PC152       PC153       PC154       PC155        PC156
## sample1 -0.4152325  0.07938604 -0.06075683 -0.44737376  0.05639599 -0.03094608
## sample2 -0.3088162 -0.30550341  0.47992301  0.06612236 -0.13217211  0.37420671
## sample3  0.1121047  0.10502580  0.19410320 -0.29629075 -0.06630255  0.05367717
## sample4 -0.2388197  0.48890475 -0.37125360  0.50717835 -0.19558648 -0.69519398
## sample5  0.1611495 -0.28098127 -0.03131864  0.18911946 -0.14205833  0.01586010
## sample6  0.2651905  0.02982794  0.23533065 -0.24430382 -0.03270207  0.08452278
##                PC157       PC158       PC159       PC160       PC161        PC162
## sample1 -0.25513551 -0.28496456  0.01678875  0.05211229 -0.23588670  0.16346719
## sample2 -0.47147112  0.07924018 -0.20705538 -0.45357834 -0.11266073 -0.03010016
```

```
## sample3  0.02536806  0.04245514  0.13884590 -0.24484260  0.40733242  0.04470447
## sample4  0.05728494 -0.13356263 -0.80210738 -0.28275391  0.03512959 -0.11374008
## sample5  0.22803058 -0.47415303 -0.44590238 -0.08456046 -0.43799346  0.28060987
## sample6 -0.15152106 -0.21668937  0.17907459  0.22256306 -0.59568419  0.23813775
##              PC163       PC164        PC165       PC166       PC167       PC168
## sample1 -0.1521696 -0.32751964 -0.009092855  0.02569953 -0.18130562 -0.13250310
## sample2 -0.1272290 -0.02039715  0.357217063 -0.30112659 -0.04347569 -0.11575280
## sample3  0.2188163  0.09886630  0.054081046  0.01257485 -0.24021094 -0.08155547
## sample4 -0.1848070  0.25135489 -0.332524926 -0.62975228  0.24190180 -0.09598397
## sample5  0.0706700 -0.01764518  0.193866222 -0.24985140  0.05058413  0.16833406
## sample6 -0.2369042 -0.44115339 -0.083090914  0.04353588 -0.11349305  0.29800226
##               PC169       PC170       PC171       PC172       PC173       PC174
## sample1 -0.02693455  0.06130118 -0.03043352 -0.05634289 -0.32176981  0.15203368
## sample2  0.25437968 -0.18794419 -0.10227530  0.06671138 -0.09446323  0.19598364
## sample3  0.17169921 -0.08937660 -0.11389644  0.17645049 -0.12298920 -0.02087733
## sample4  0.33263119 -0.33327852 -0.06923204  0.35798739  0.35716607 -0.24701546
## sample5 -0.02443868 -0.31717314 -0.21211838 -0.06181074 -0.26730654 -0.01753560
## sample6 -0.16950961  0.17173967 -0.04565571 -0.04869302  0.07997684  0.13920049
##                PC175       PC176        PC177        PC178       PC179
## sample1 -0.042858212  0.13939583 -0.074577029 -0.027797844  0.03193647
## sample2 -0.232334278  0.42221839  0.199748802  0.595255126  0.23314195
## sample3 -0.053825398 -0.04309746  0.004696872  0.026612280  0.14128085
## sample4 -0.114343622  0.11417564  0.407023386 -0.176651414  0.08450117
## sample5 -0.008159745 -0.15201120 -0.436277455  0.009435339 -0.19912316
## sample6 -0.119312959 -0.21801103  0.167396684 -0.279565428 -0.19756522
##               PC180        PC181      PC182        PC183         PC184
## sample1  0.21413738  0.043179944  0.1223794 -0.132564017 -0.1737264053
## sample2 -0.24062545  0.073943623  0.2266229 -0.394471611 -0.2702946458
## sample3 -0.07380298 -0.003802155  0.1774031 -0.007362027 -0.2932696758
## sample4 -0.02847869 -0.126203803 -0.3964707  0.207143665 -0.0003630136
## sample5  0.07889897 -0.022998777  0.1507797  0.008803125 -0.2874081112
## sample6 -0.24164986 -0.064153673 -0.2277039  0.097324572  0.0717200703
##               PC185       PC186       PC187        PC188       PC189
```

```
## sample1 -0.11126798 -0.14243071 -0.07169008 -0.215277076 -0.06821280
## sample2  0.25042336 -0.14007100 -0.03941041  0.055170946  0.13489633
## sample3 -0.06549300  0.29838930  0.01518037 -0.006230371  0.01172052
## sample4 -0.12616250  0.48854401 -0.01666230  0.284454304  0.14032323
## sample5  0.04438640 -0.31636550 -0.36107749  0.203084010  0.31446811
## sample6 -0.04768487  0.07375501  0.02812817  0.038596635 -0.03977939
##               PC190       PC191       PC192       PC193        PC194
## sample1  0.15383406 -0.01638648 -0.36669701 -0.03576649 -0.002951877
## sample2  0.06484606 -0.14050753  0.09026978  0.24025998 -0.202728799
## sample3 -0.03257070  0.03860802  0.09111611 -0.17809095  0.019912569
## sample4  0.01637805  0.15031327 -0.31782815  0.43800270  0.166813090
## sample5  0.21369368 -0.08998693  0.03267441 -0.13689029  0.296844584
## sample6 -0.08172560 -0.09446870  0.05536772 -0.19119877 -0.050037371
##               PC195       PC196        PC197       PC198        PC199
## sample1  0.05790646  0.11612686 -0.084962132 -0.06347505 -0.195395911
## sample2  0.15462269 -0.23893176  0.190442131 -0.12054892 -0.149201247
## sample3 -0.04244743  0.08840142 -0.028189326  0.07198567  0.032788615
## sample4 -0.20121370  0.25857632 -0.046253840  0.12680826 -0.017684110
## sample5 -0.05734483 -0.03366494  0.020830040 -0.10763552  0.006493751
## sample6 -0.01121805  0.06985552 -0.002304721 -0.03516768  0.034913267
##               PC200       PC201        PC202        PC203       PC204
## sample1 -0.05959361 -0.16654308 -0.097738315  0.007840548 -0.26769294
## sample2  0.03483269  0.08261427  0.187794885 -0.027646447  0.01917525
## sample3 -0.18239804  0.15520693  0.008449394 -0.052864101 -0.06074288
## sample4  0.21775027 -0.13290544  0.182785188 -0.175213172  0.04955471
## sample5 -0.15417981 -0.16523477  0.184293644 -0.169460488  0.04162415
## sample6  0.20000910  0.04919326 -0.224653571 -0.011870777  0.12250642
##               PC205       PC206       PC207       PC208       PC209
## sample1  0.01520703  0.11473576  0.12707922 -0.02675976  0.02163822
## sample2 -0.08842290  0.02124751  0.16956039 -0.04080387  0.09420781
## sample3 -0.11465461  0.03846782 -0.06048750 -0.05129725  0.01242629
## sample4  0.09657110 -0.09796371 -0.04668648 -0.10127333  0.02489432
## sample5  0.13288346 -0.11283395  0.06558841 -0.26701697 -0.02920717
```

```
## sample6 -0.12514981  0.02435888  0.07670774  0.04942246 -0.06963213
##                PC210        PC211       PC212       PC213        PC214
## sample1  0.078219848 -0.099040519  0.01296124  0.03996228  0.064143818
## sample2  0.052555943 -0.154840314  0.01089780  0.10638344  0.034273609
## sample3 -0.012800551 -0.005584196  0.00148453 -0.01027809 -0.008835724
## sample4 -0.065742750  0.163373553 -0.14104914 -0.06847218  0.016623328
## sample5  0.123097930 -0.065551212  0.13412399 -0.03091692  0.350043972
## sample6  0.008012962 -0.124696756  0.02933288 -0.03833646  0.020965817
##                PC215        PC216        PC217        PC218        PC219
## sample1  0.010147743  0.043358011  0.026283676  0.058347433 -0.014297974
## sample2  0.048128812 -0.110895979  0.060021401 -0.010535258  0.043095347
## sample3 -0.002666527 -0.047460606 -0.007161407 -0.002106205 -0.005524306
## sample4  0.042455894 -0.037665615 -0.082250088 -0.010580625 -0.022380615
## sample5  0.273094949 -0.004320428  0.172423148 -0.032269233  0.054417938
## sample6  0.057179321 -0.059681336  0.197469308  0.016341985  0.023250168
##                PC220        PC221       PC222        PC223        PC224
## sample1  0.02080313  0.038964437 -0.00850176 -0.012591028 -0.017411134
## sample2  0.07592316  0.020048056 -0.05306985  0.103759458 -0.023570617
## sample3 -0.00140366 -0.005433660 -0.04912014  0.009357218  0.008755356
## sample4  0.02158322  0.044361492  0.02633523 -0.019211961 -0.027922328
## sample5 -0.06713282  0.003146911  0.01224925  0.022782701 -0.021441115
## sample6 -0.04406821  0.031869501 -0.01959189  0.002353936  0.016423432
##                PC225         PC226        PC227        PC228
## sample1  0.036675315 -0.0134039459 -0.005404331  0.007592886
## sample2 -0.015844640  0.0331464320 -0.006553896 -0.009773145
## sample3  0.005907911 -0.0030999011  0.009775586 -0.005606844
## sample4  0.037110843  0.0193348387 -0.005765235  0.004717541
## sample5 -0.002233977 -0.0004114733  0.005127198 -0.001696175
## sample6 -0.030344456  0.0057095806 -0.015746971 -0.008779756
```

```
# 提取主成分的方差贡献率，生成坐标轴标题
xlab2 <- paste0("PC1(",round(proportion_of_variance[1]*100,2),"%)")
ylab2 <- paste0("PC2(",round(proportion_of_variance[2]*100,2),"%)")
```

```
# 绘制 PCA 得分图
p.pca2 <- ggplot(data = df2,aes(x = PC1,y = PC2,color = solubility_data$Species))+
  geom_point(size = 3)+
  theme_bw()+
  labs(x = xlab2,y = ylab2,color = "Group",title = "Plot of PCA score")+
  stat_ellipse(aes(fill = solubility_data$Species),
               type = "norm",geom = "polygon",alpha = 0.2,color = NA)+
  guides(fill = "none")+
  theme(plot.title = element_text(hjust = 0.5,size = 15),
        axis.text = element_text(size = 11),axis.title = element_text(size = 13),
        legend.text = element_text(size = 11),legend.title = element_text(size = 13),
        plot.margin = unit(c(0.4,0.4,0.4,0.4),'cm'))
ggsave(p.pca2,filename = "PCA.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

7 性能比较

```
## Benchmark -----
learnersBenchmark <- list( tunedXgb,
                           tunedRF,
                           tunedTree
                 )
cvBenchmark <- makeResampleDesc("RepCV", folds=5, reps=5)
parallelStartSocket(cpus = detectCores())
```

```
## Starting parallelization in mode=socket with cpus=12.
```

```
resultBenchmark <- benchmark(learnersBenchmark, task, cvBenchmark)
```

```
## Exporting objects to slaves for mode socket: .mlr.slave.options
```

```
## Mapping in parallel: mode = socket; level = mlr.benchmark; cpus = 12; elements = 3.
```

```
parallelStop()
```

```
## Stopped parallelization. All cleaned up.
```

```
resultBenchmark
```

```
##           task.id        learner.id mse.test.mean
## 1 solubility_data      regr.xgboost     0.4752053
## 2 solubility_data regr.randomForest     1.0690698
## 3 solubility_data        regr.rpart     1.1512908
```

```
## CV embedded with tuning
cvTune <- makeResampleDesc("CV",iters=5)  # inner cv (for tuning)

treeWrapper <- makeTuneWrapper(treeLearner,
                               resampling=cvTune,
                               par.set=treeParamSpace,
                               control=randSearch)
rfWrapper <- makeTuneWrapper(rfLearner,
                               resampling=cvTune,
                               par.set=rfParamSpace,
                               control=randSearch)
xgbWrapper <- makeTuneWrapper(xgbLearner,
                                resampling=cvTune,
                                par.set=xgbParamSpace,
                                control=randSearch)
learnersBen <- list(treeWrapper,rfWrapper,xgbWrapper)
cvBen <- makeResampleDesc("CV",iters=5)  #outer cv

parallelStartSocket(cpus = detectCores())
```

```
## Starting parallelization in mode=socket with cpus=12.
```

```
resBenchmark <- benchmark(learnersBen,task,cvBen)
```

```
## Exporting objects to slaves for mode socket: .mlr.slave.options
## Mapping in parallel: mode = socket; level = mlr.benchmark; cpus = 12; elements = 3.
```

```
parallelStop()
```

```
## Stopped parallelization. All cleaned up.
```

```
print(" 线性回归的 mse 是")
```

```
## [1] "线性回归的mse是"
```

```
print(linear_mse)
```

```
## [1] 0.4303537
```

```
resBenchmark
```

```
##           task.id              learner.id mse.test.mean
## 1 solubility_data        regr.rpart.tuned      1.185682
## 2 solubility_data regr.randomForest.tuned      1.123463
## 3 solubility_data      regr.xgboost.tuned      0.518075
```

经过对比 mse，不难发现线性回归的性能是最好的