EMDA_HW2

r12631055

2024-09-21

HW2

1.

The qnorm function returns the percentiles (quantiles) of a normal distribution. Use the qnorm function to find the 95th percentile of the standard normal distribution. Then, use the qnorm function to find the quartiles of the standard normal distribution (the quartiles are the 25, 50, and 75 percentiles). Hint: Use c(.25, .5, .75) as the first argument to qnorm.

```
qnorm(0.95)
```

```
## [1] 1.644854
```

```
qnorm(c(0.25, 0.5, 0.75))
```

```
## [1] -0.6744898  0.0000000  0.6744898
```

2.

(iris data). The iris data gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of three species of iris. There are four numeric variables corresponding to the sepal and petal measurements and one factor, Species. Display a table of means by Species (means should be computed separately for each of the three Species).

```
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
means_table <- aggregate(. ~ Species, data = iris, FUN = mean)
print(means_table)
```

```
##      Species Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1     setosa        5.006       3.428        1.462       0.246
## 2 versicolor        5.936       2.770        4.260       1.326
## 3  virginica        6.588       2.974        5.552       2.026
```

3

(mtcars data). Display the mtcars data included with R and read the documentation using ?mtcars. Display parallel boxplots of the quantitative variables. Display a pairs plot of the quantitative variables. Does the pairs plot reveal any possible relations between the variables?

```
print(mtcars)
```

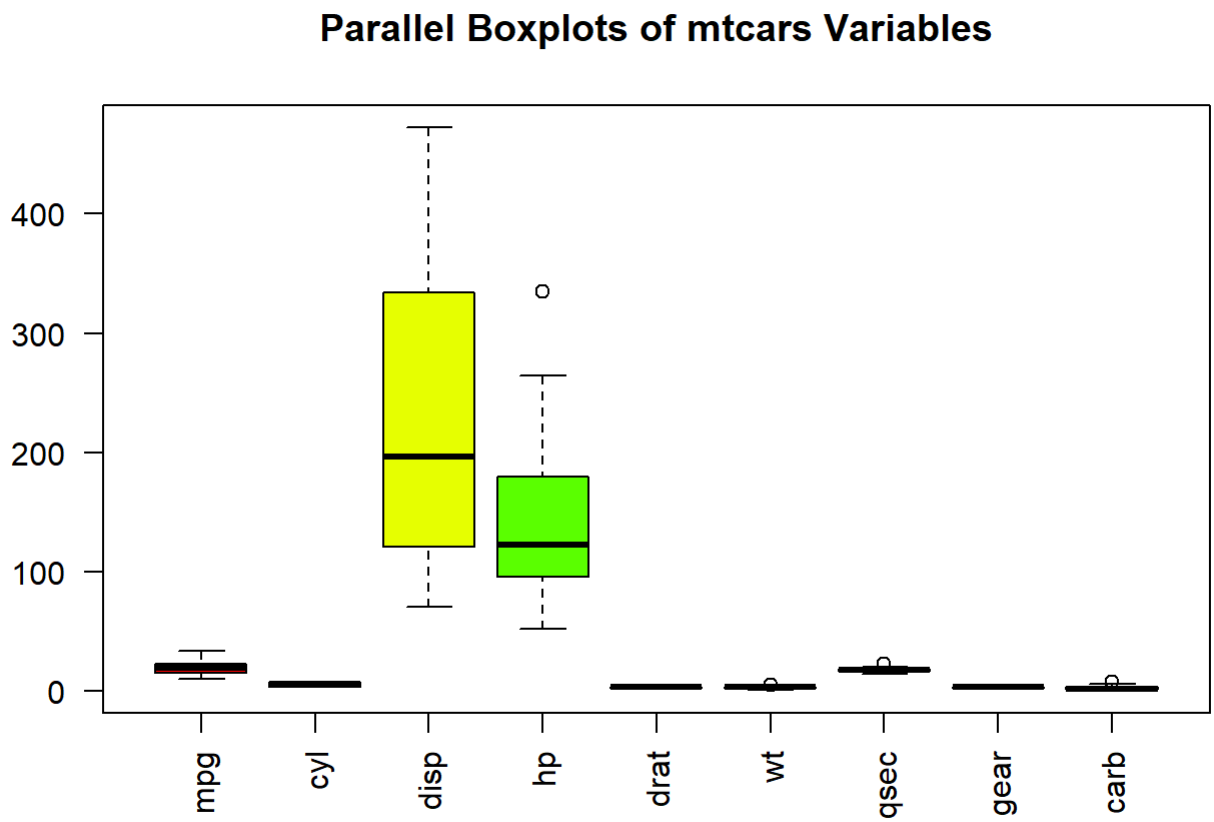
```
##                    mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C         17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE        16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL        17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC       15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128          32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic       30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla    33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona     21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger  15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin       15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28        13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird  19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9         27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2     26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa      30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L    15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino      19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora     15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E        21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

By definition of quantitative variable, where [, 8]vs :Engine(0 = V-shaped, 1 = straight) and [, 9]am Transmission(0 = automatic, 1 = manual) are not quantitative variables.

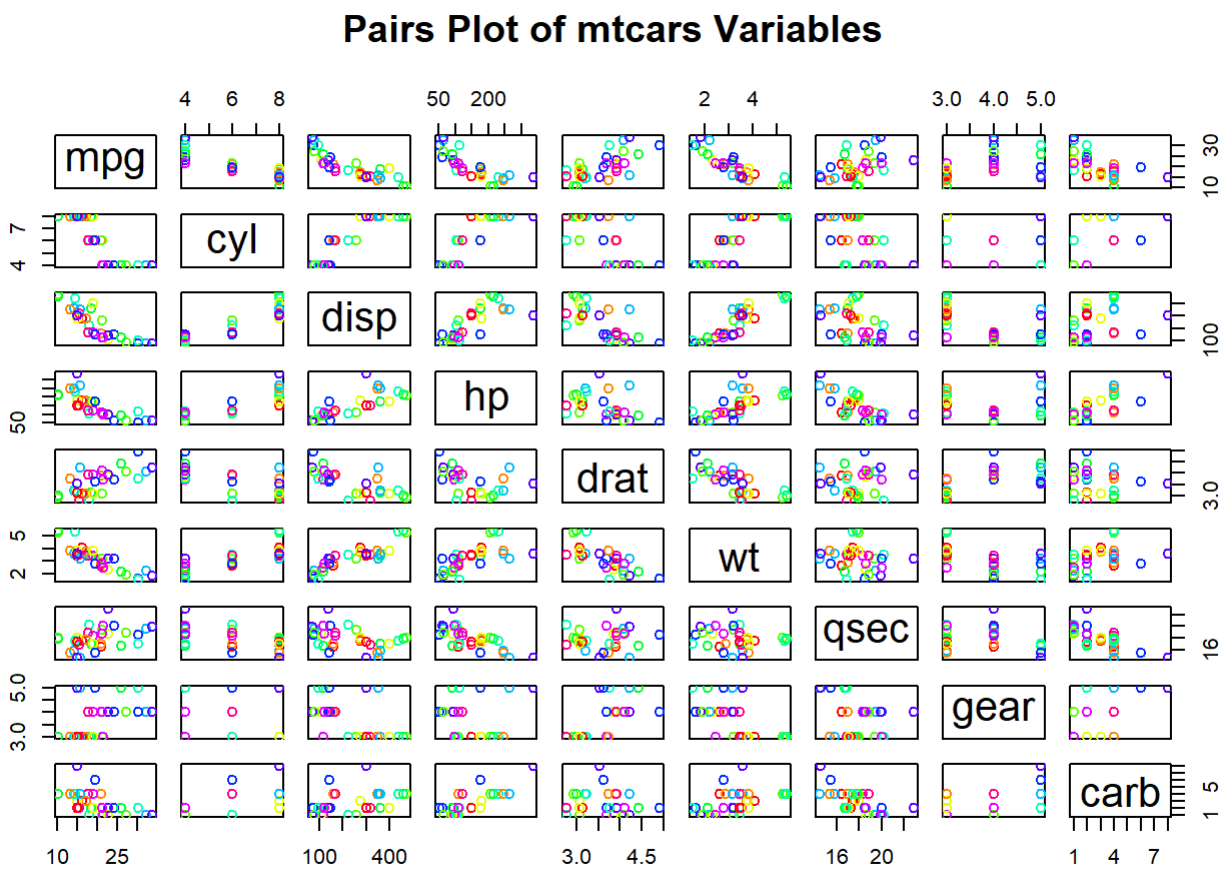
```
mtcars_quantitative <- subset(mtcars, select = -c(vs, am))
head(mtcars_quantitative)
```

```
##                   mpg cyl disp  hp drat    wt  qsec gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22    3    1
```

```
boxplot(mtcars_quantitative, main = "Parallel Boxplots of mtcars Variables",
        las = 2, col = rainbow(ncol(mtcars)))
```




```
pairs(mtcars_quantitative, main = "Pairs Plot of mtcars Variables", col = rainbow(ncol(mtcar
s)))
```



Yes, some pairs seem have some linear pattern with a positive or negative slope, it suggests a possible correlation, also there are clustering in some pairs of variables might suggest distinct groups or relationships, while outliers may indicate unusual data points.

4.

Use the bivariate boxplot on the scatterplot of each pair of variables in the air pollution data to identify any outliers. Calculate the correlation between each pair of variables using all the data and the data with any identified outliers removed. Comment on the results. (USpollution data in MVA package)

```
install.packages("MVA")
```

```
## 將程式套件安載入 'C:/Users/Paul/AppData/Local/R/win-library/4.4'
## (因為 'lib' 沒有被指定)
```

```
## 程式套件 'MVA' 開啟成功, MD5 和檢查也透過
##
## 下載的二進位程式套件在
##  C:\Users\Paul\AppData\Local\Temp\RtmpgrNwFX\downloaded_packages 裡
```

```
library(MVA)
```

```
## Warning: 套件 'MVA' 是用 R 版本 4.4.1 來建造的
```

```
## 載入需要的套件: HSAUR2
```

```
## Warning: 套件 'HSAUR2' 是用 R 版本 4.4.1 來建造的
```

```
## 載入需要的套件: tools
```

```
install.packages("aplpack")
```

```
## 將程式套件安載入 'C:/Users/Paul/AppData/Local/R/win-library/4.4'
## (因為 'lib' 沒有被指定)
```

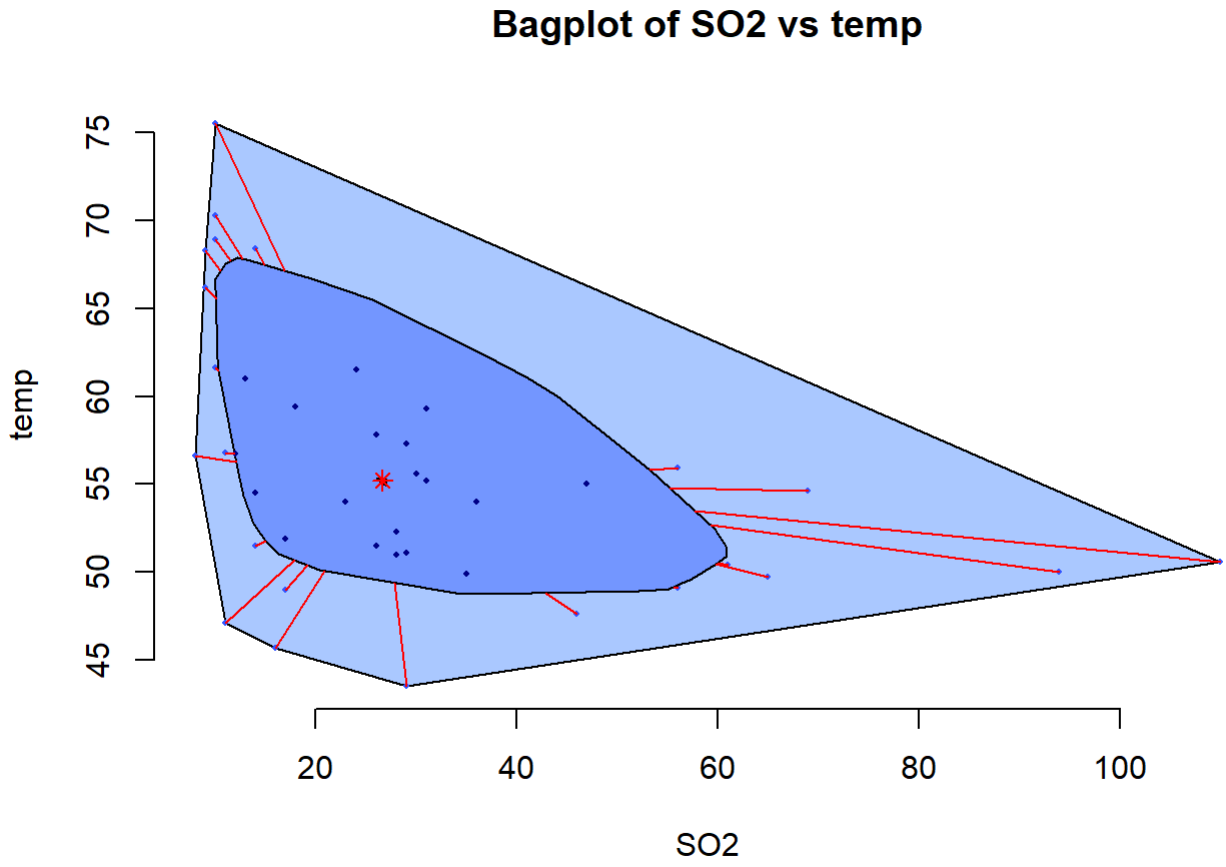
```
## 程式套件 'aplpack' 開啟成功, MD5 和檢查也透過
##
## 下載的二進位程式套件在
##  C:\Users\Paul\AppData\Local\Temp\RtmpgrNwFX\downloaded_packages 裡
```

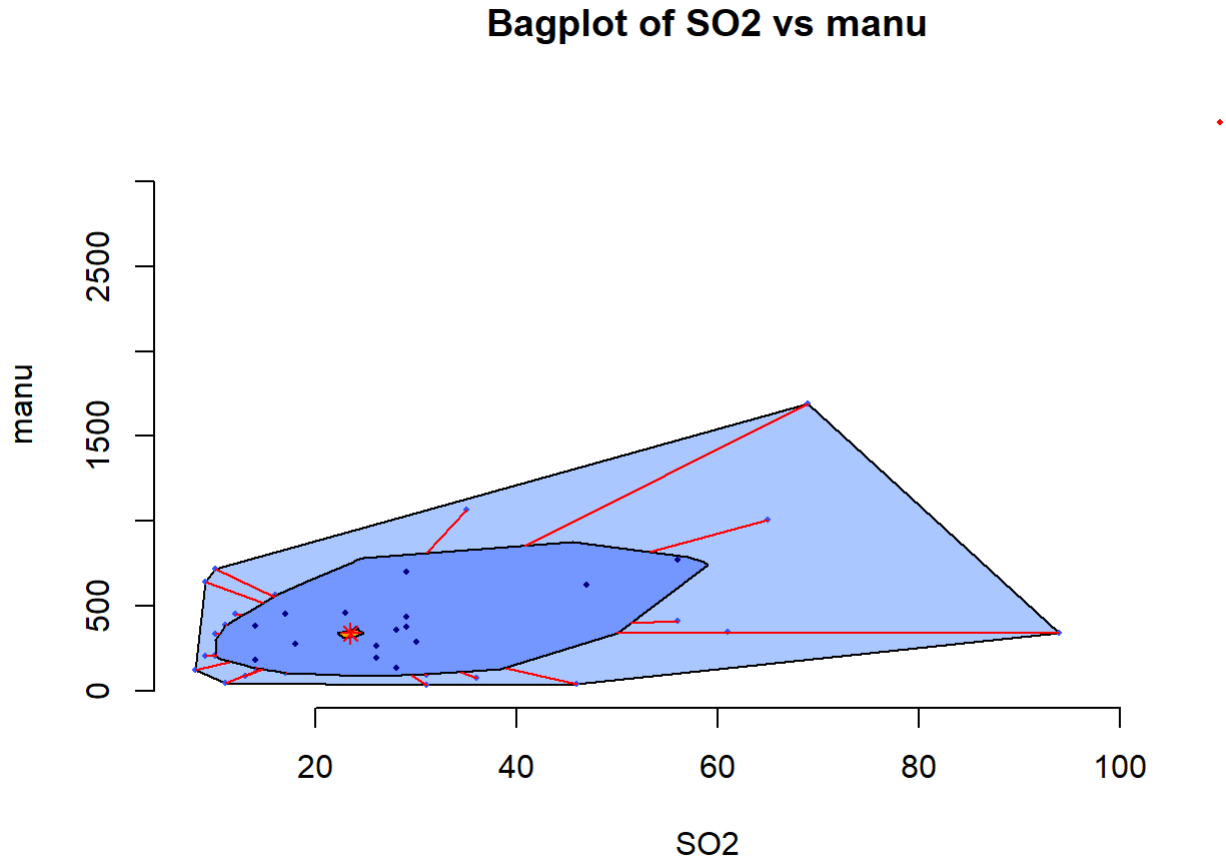
```
library(aplpack)
```

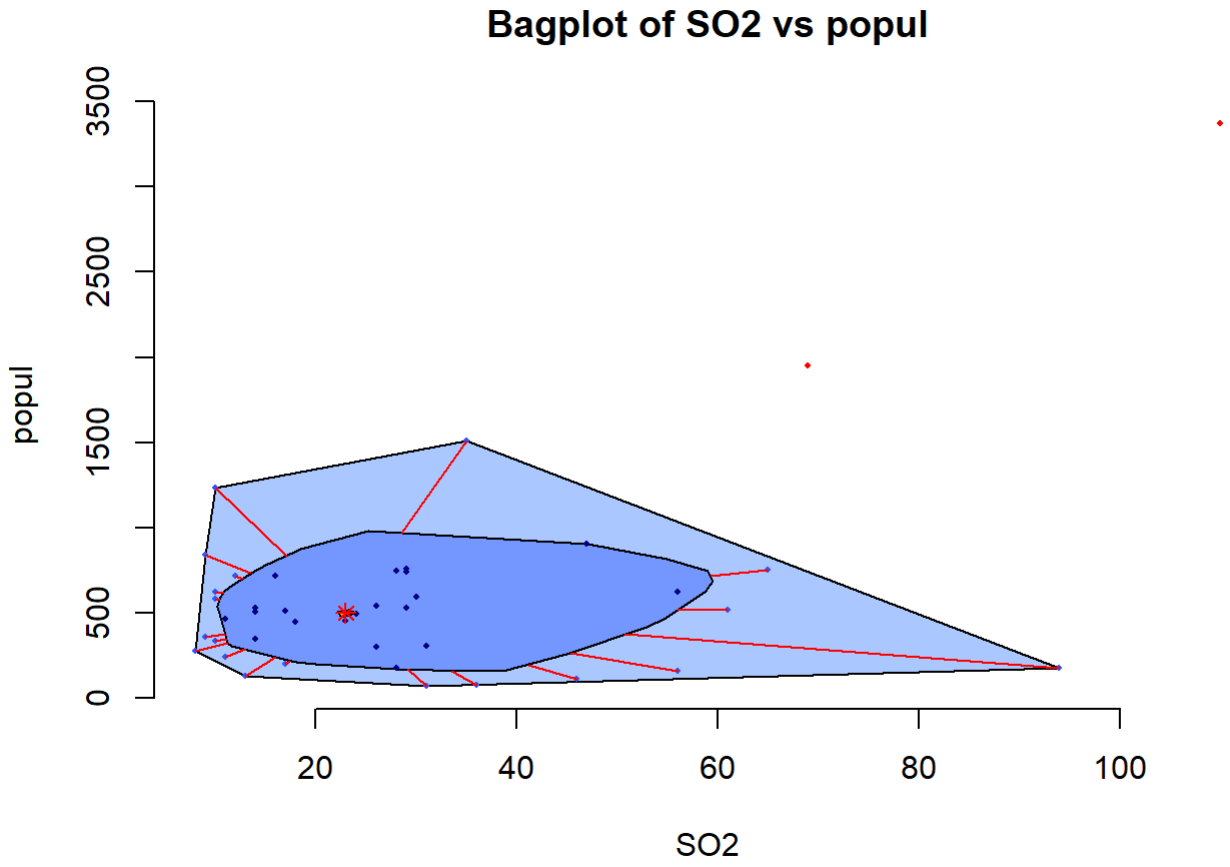
```
head(USairpollution)
```

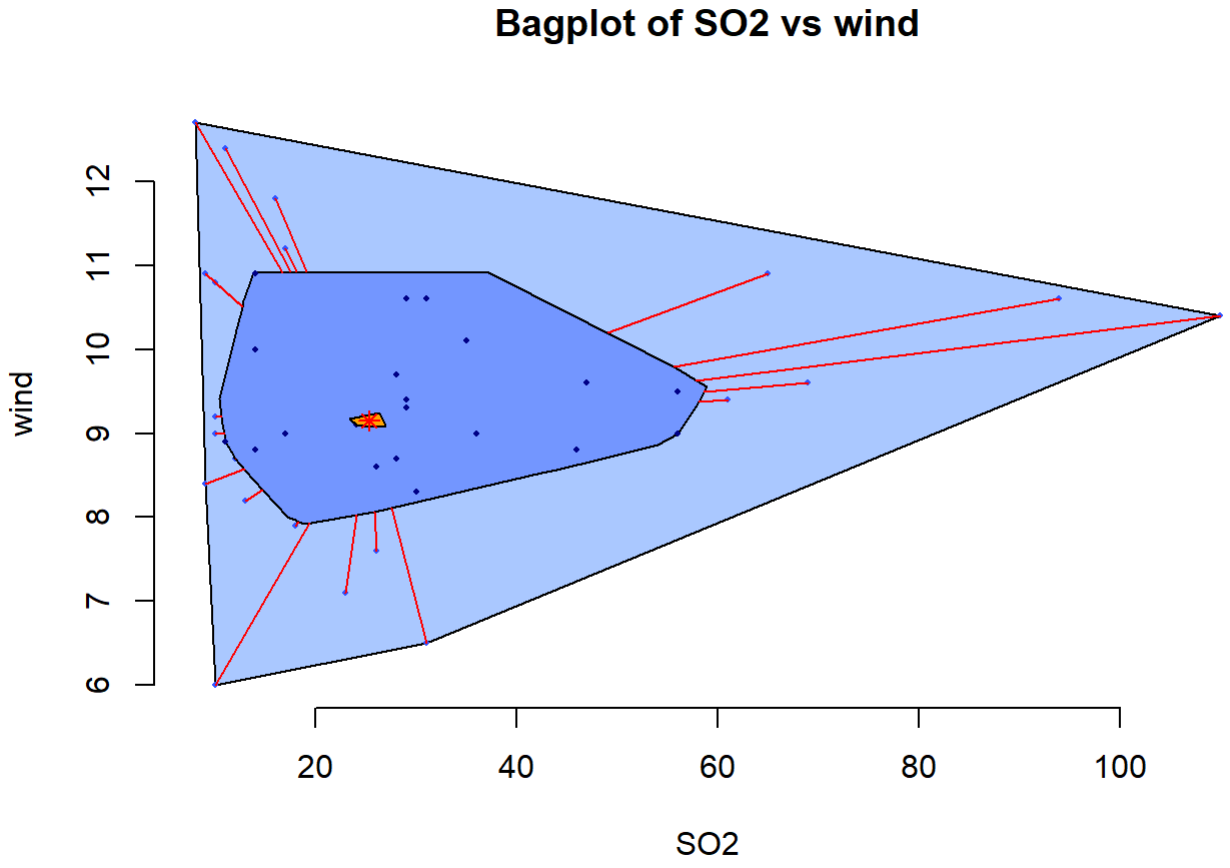
```
##             SO2 temp manu popul wind precip predays
## Albany       46 47.6   44   116  8.8  33.36     135
## Albuquerque  11 56.8   46   244  8.9   7.77      58
## Atlanta      24 61.5  368   497  9.1  48.34     115
## Baltimore    47 55.0  625   905  9.6  41.31     111
## Buffalo      11 47.1  391   463 12.4  36.11     166
## Charleston   31 55.2   35    71  6.5  40.75     148
```

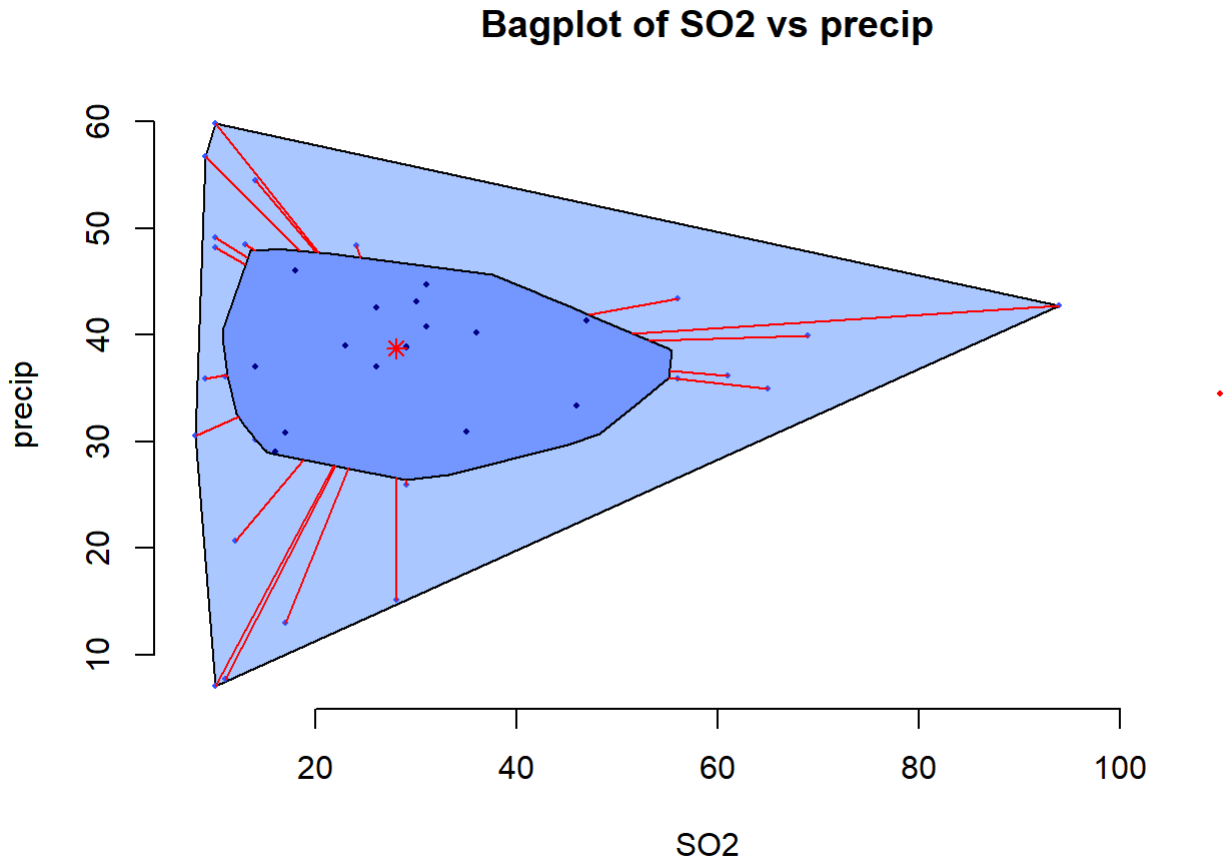
```
for (i in 1:(ncol(USairpollution)-1)) {
  for (j in (i+1):ncol(USairpollution)) {
    bagplot(USairpollution[,i], USairpollution[,j],
            xlab = colnames(USairpollution)[i],
            ylab = colnames(USairpollution)[j],
            main = paste("Bagplot of", colnames(USairpollution)[i], "vs", colnames(USairpollu
tion)[j]))
  }
}
```

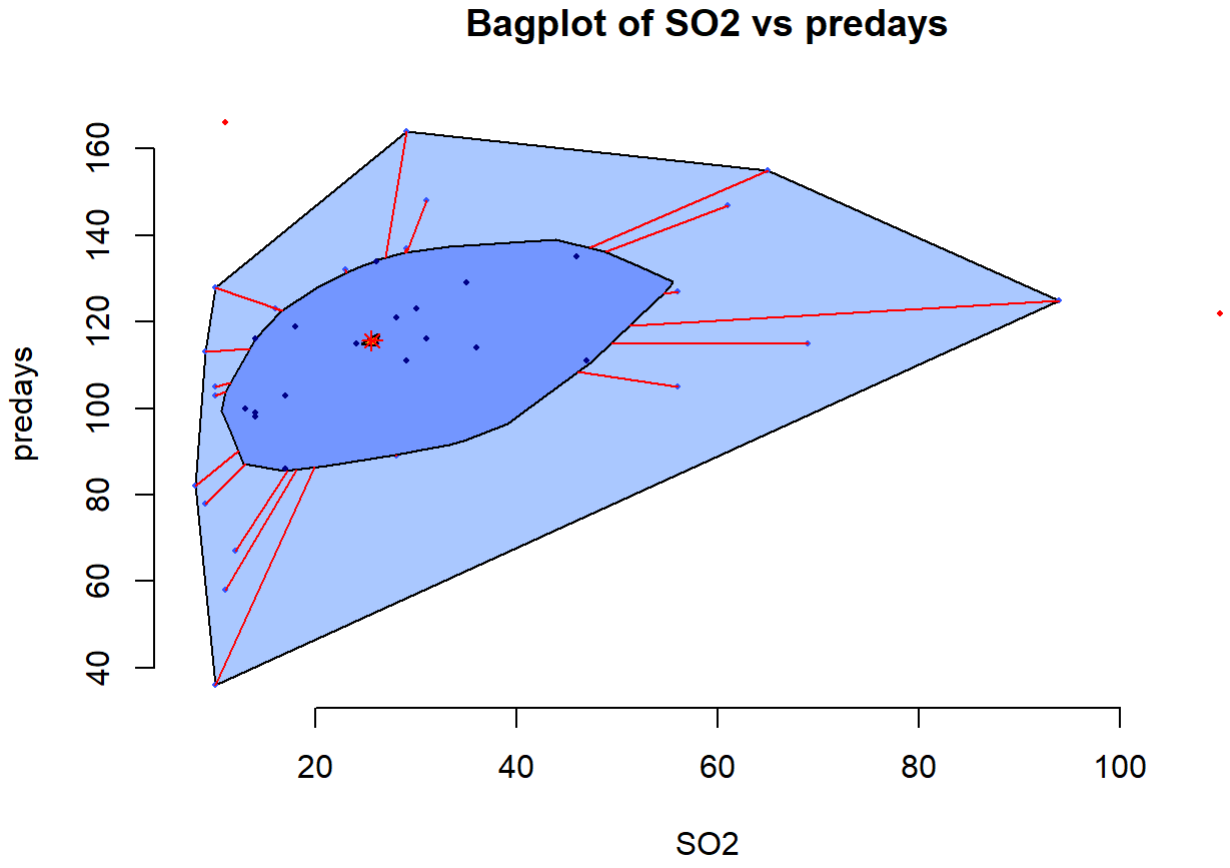

Bagplot of SO2 vs temp
temp
45
50
55
60
65
70
75
20
40
60
80
100
SO2

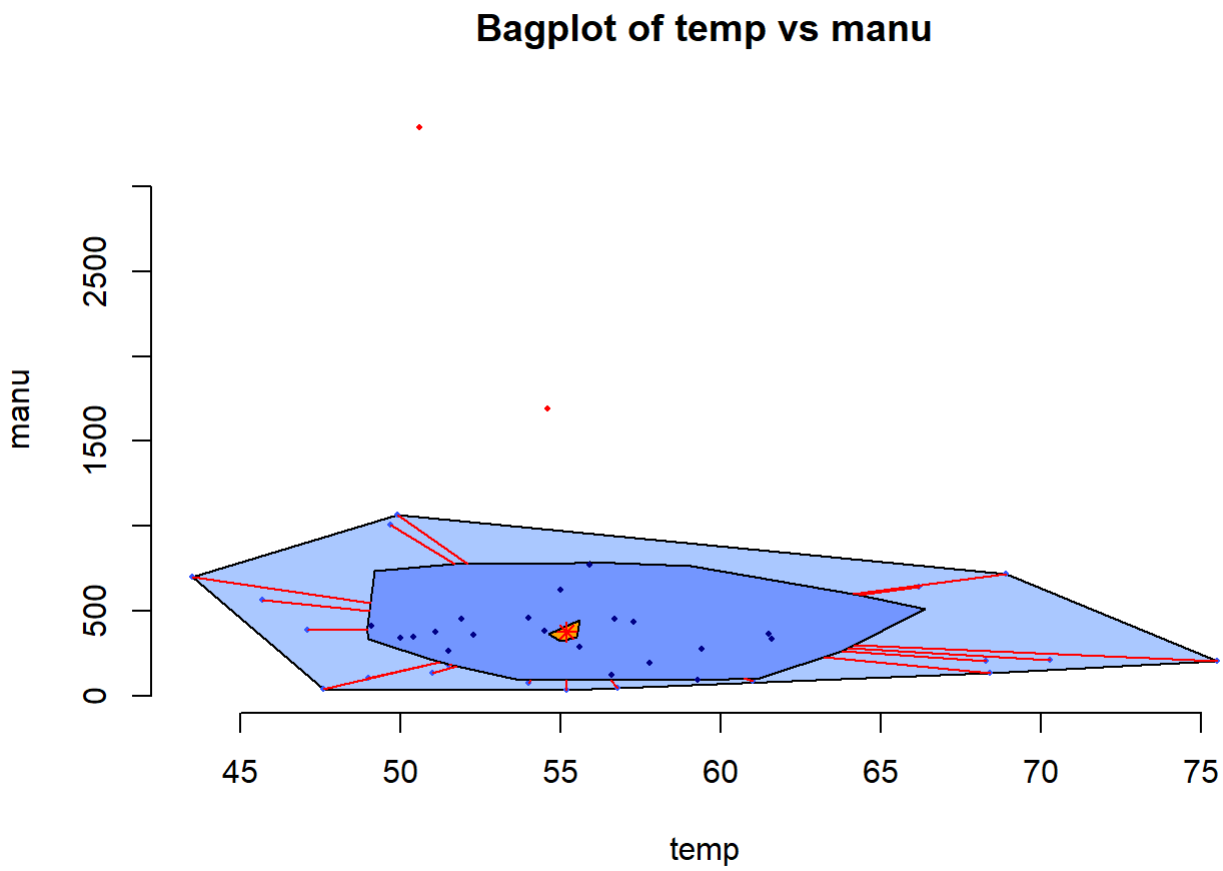

Bagplot of SO2 vs manu
manu
0
500
1500
2500
20
40
60
80
100
SO2

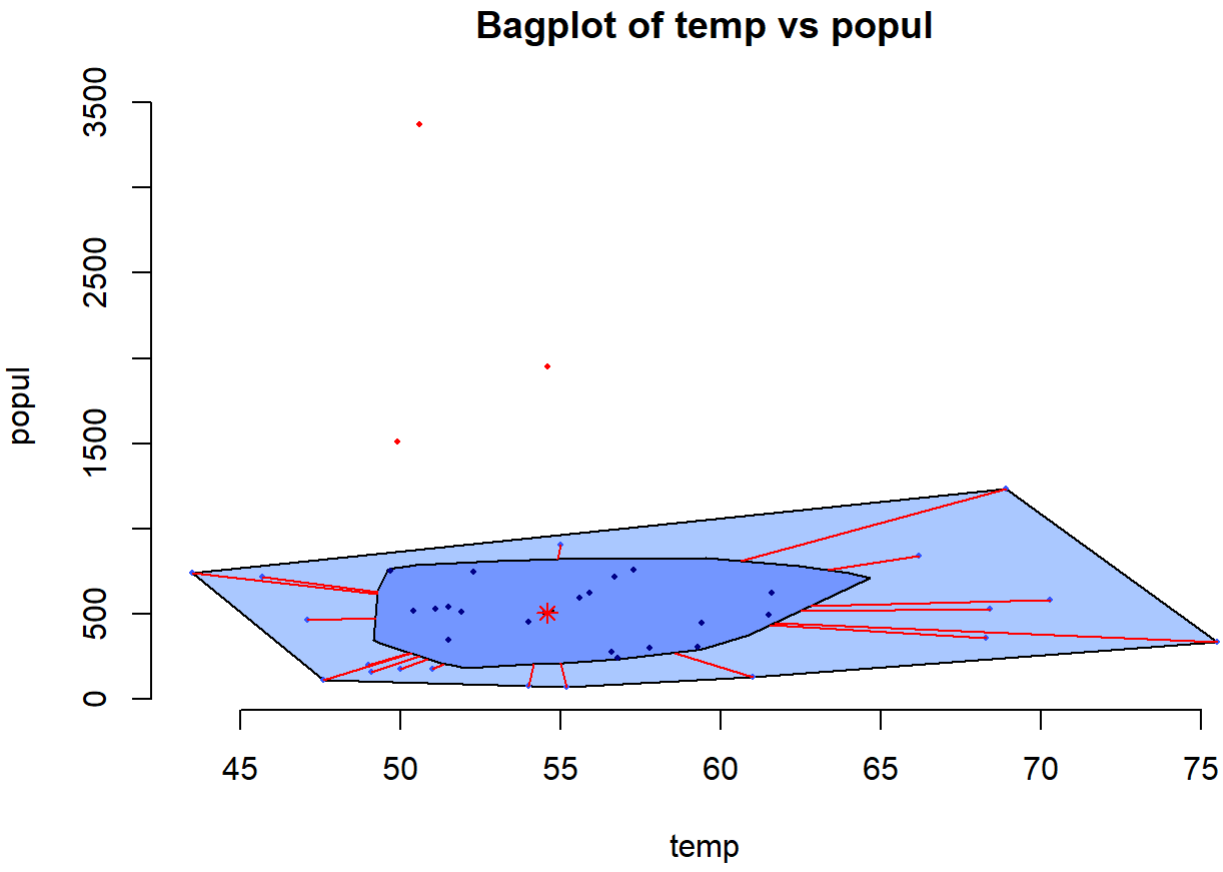

Bagplot of SO2 vs popul
popul
0
500
1500
2500
3500
20
40
60
80
100
SO2

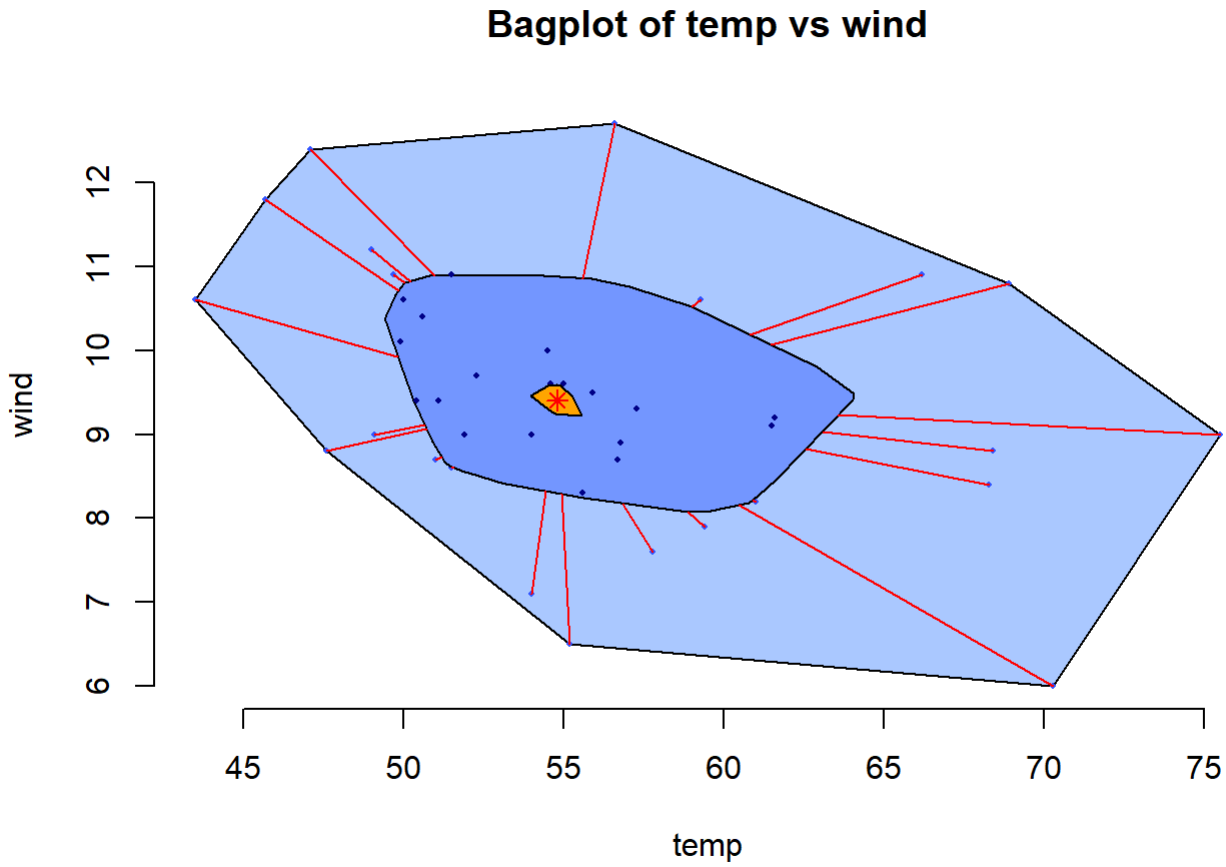

Bagplot of SO2 vs wind
wind
6
7
8
9
10
11
12
20
40
60
80
100
SO2

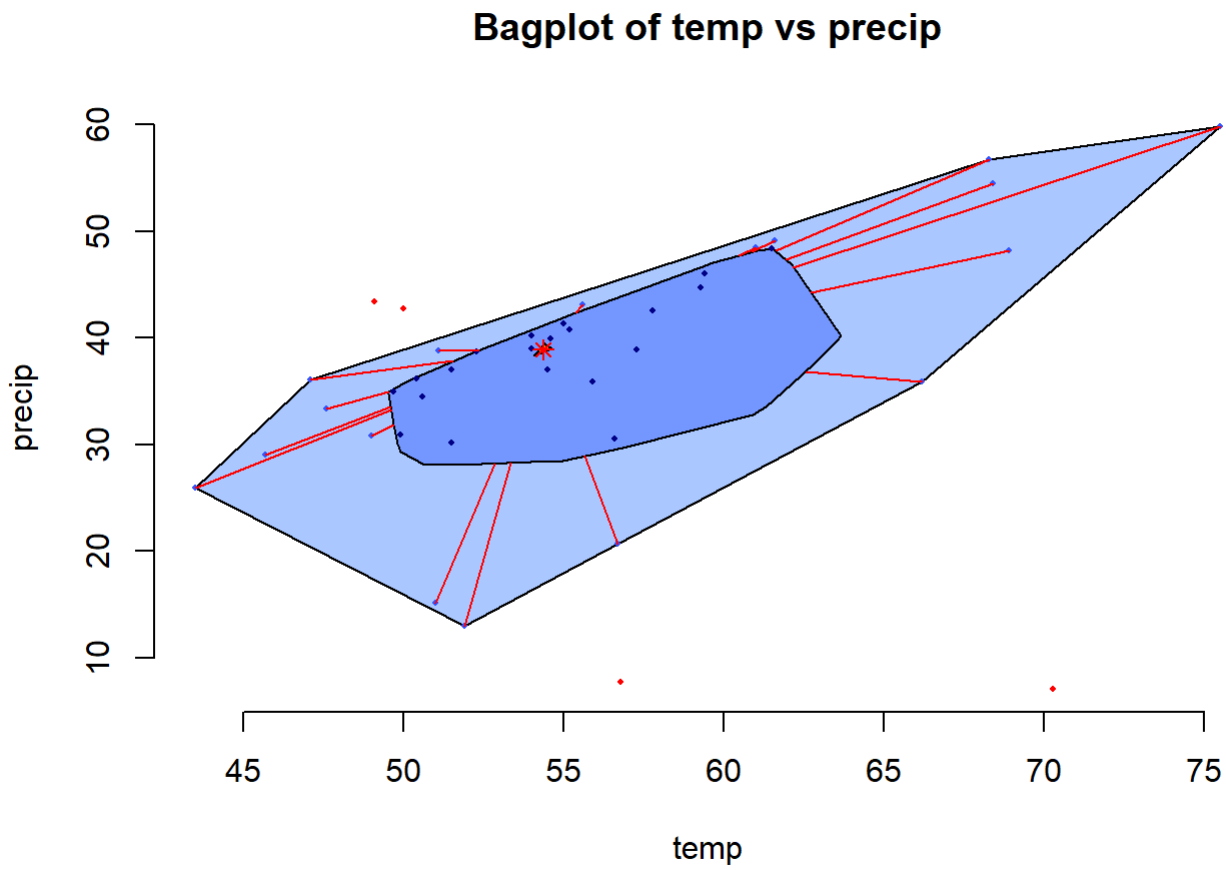

Bagplot of SO2 vs precip
precip
10
20
30
40
50
60
20
40
60
80
100
SO2

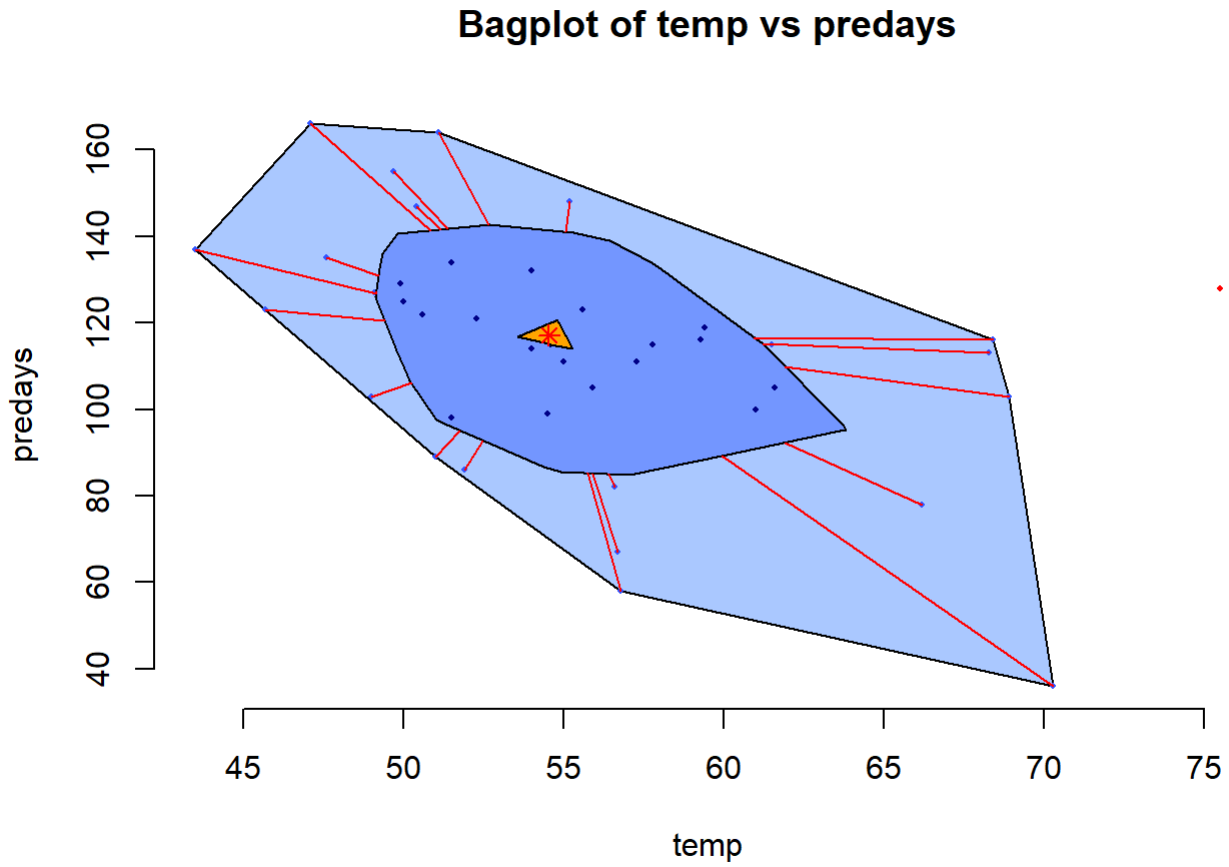

Bagplot of SO2 vs predays
predays
40
60
80
100
120
140
160
20
40
60
80
100
SO2

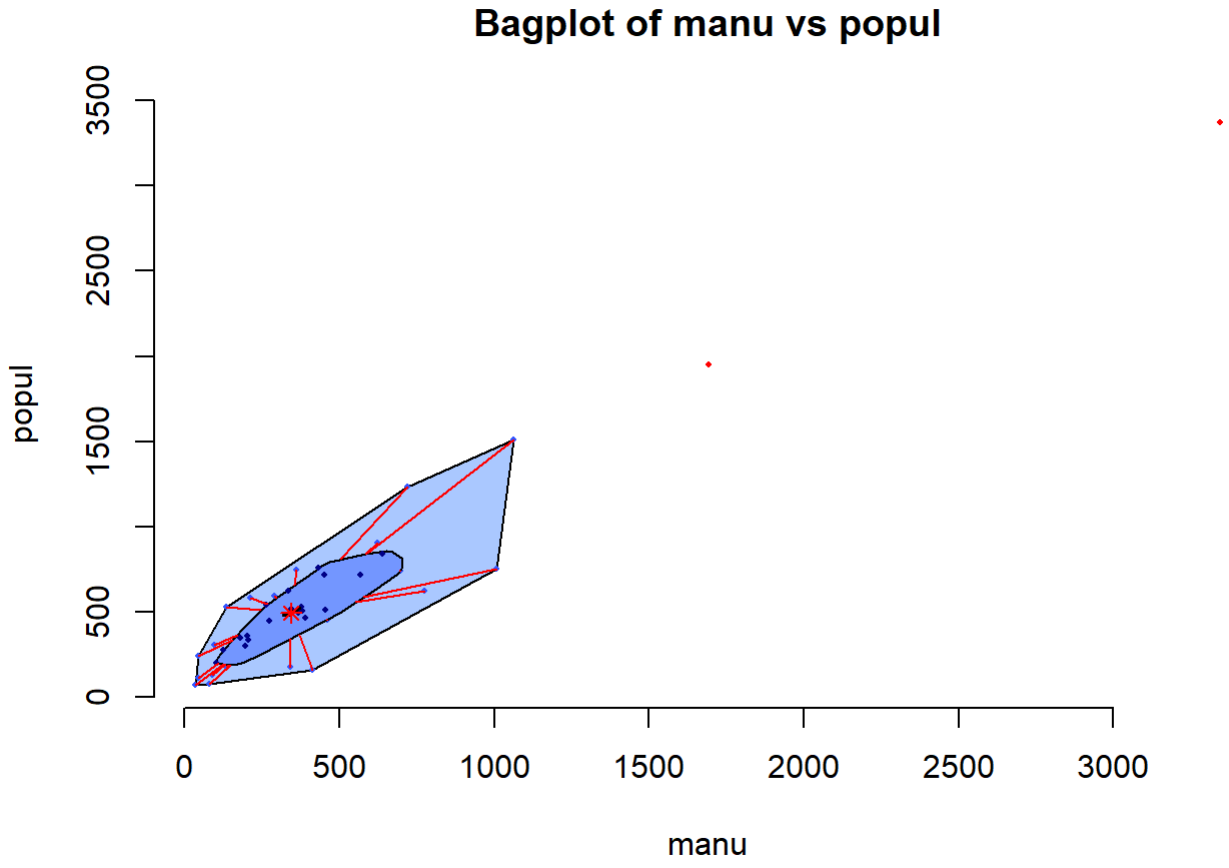

Bagplot of temp vs manu
manu
0
500
1500
2500
45
50
55
60
65
70
75
temp

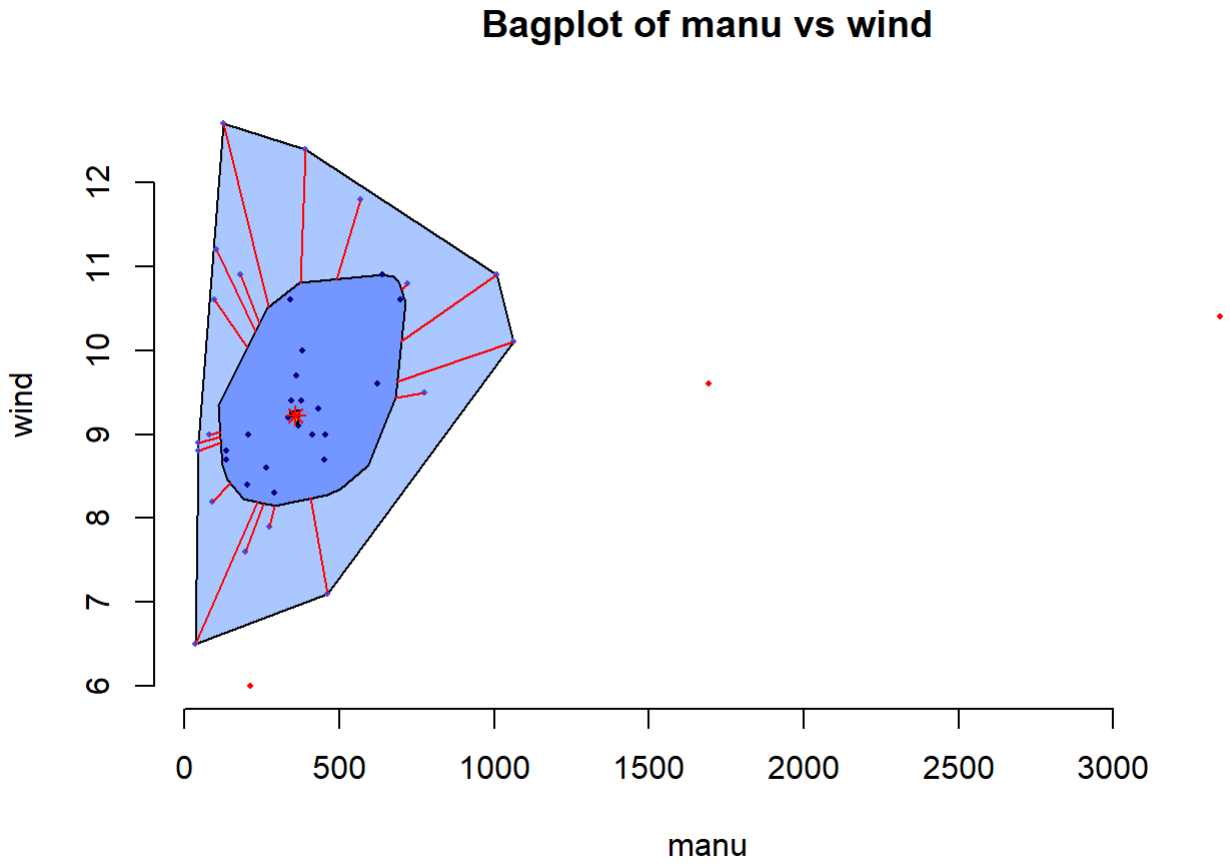

Bagplot of temp vs popul
popul
0
500
1500
2500
3500
45
50
55
60
65
70
75
temp

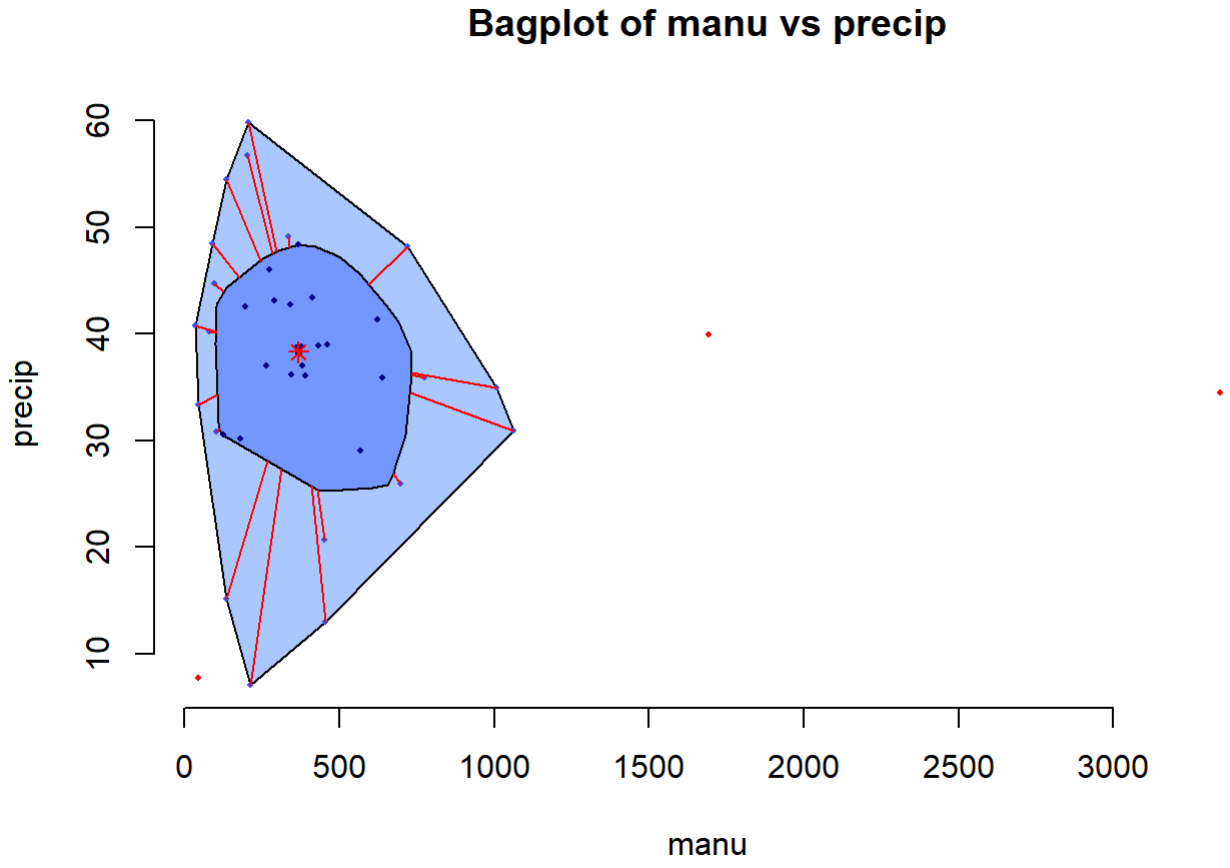

Bagplot of temp vs wind
wind
6
7
8
9
10
11
12
45
50
55
60
65
70
75
temp

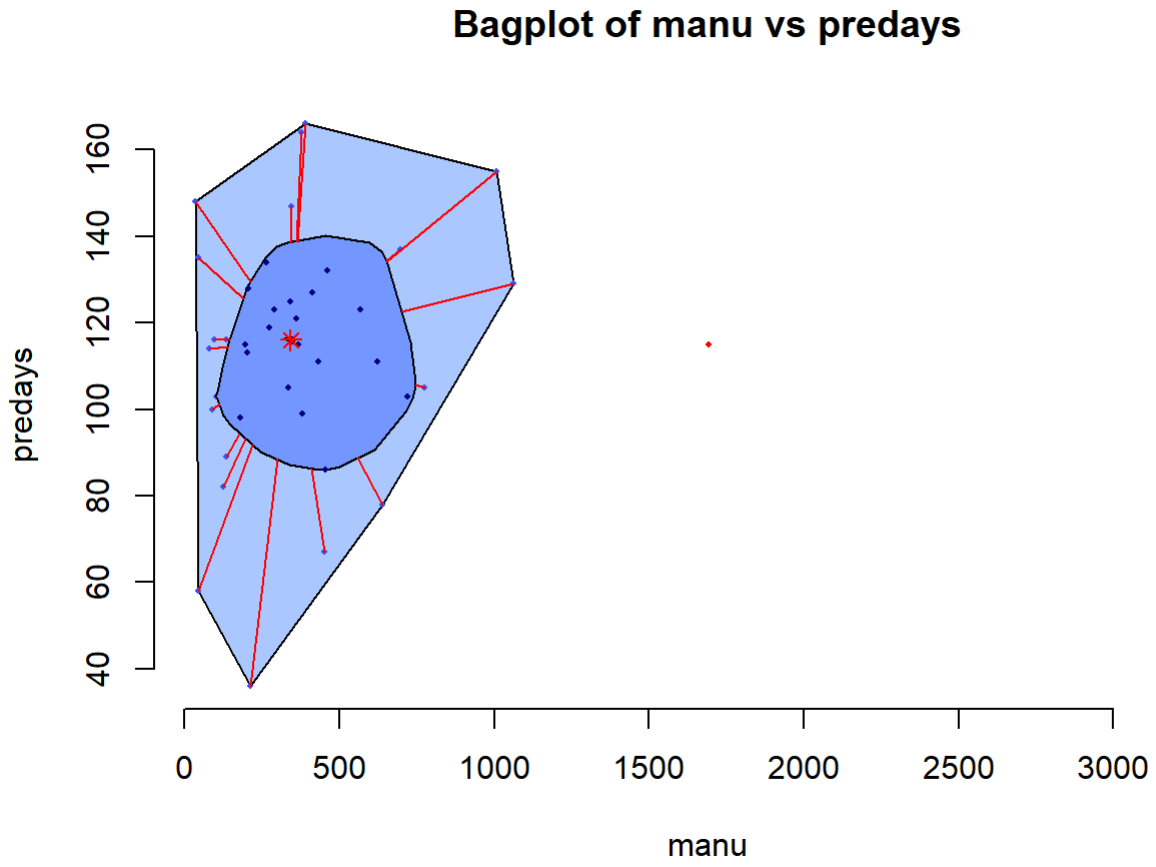

Bagplot of temp vs precip
precip
10
20
30
40
50
60
45
50
55
60
65
70
75
temp

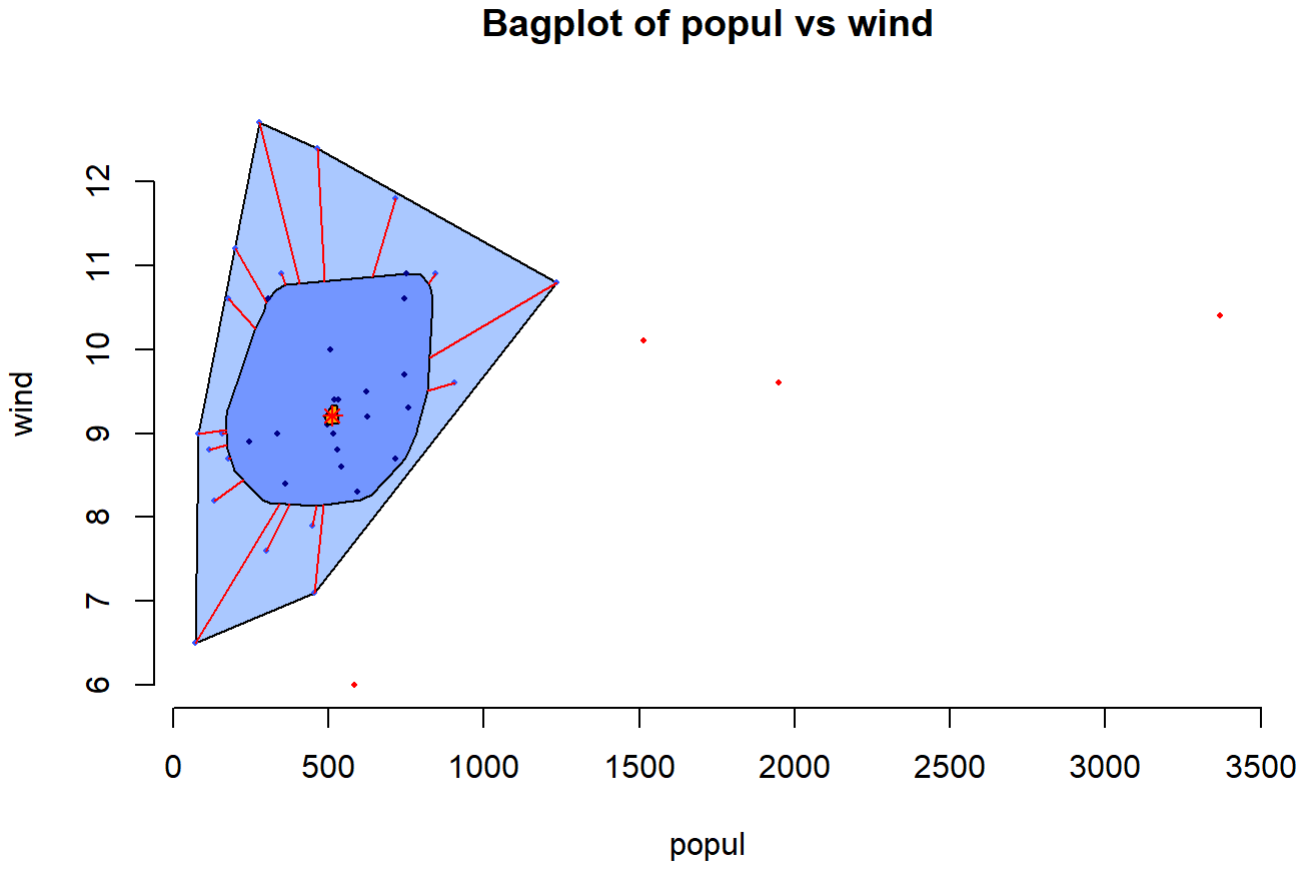

Bagplot of temp vs predays
predays
40
60
80
100
120
140
160
45
50
55
60
65
70
75
temp

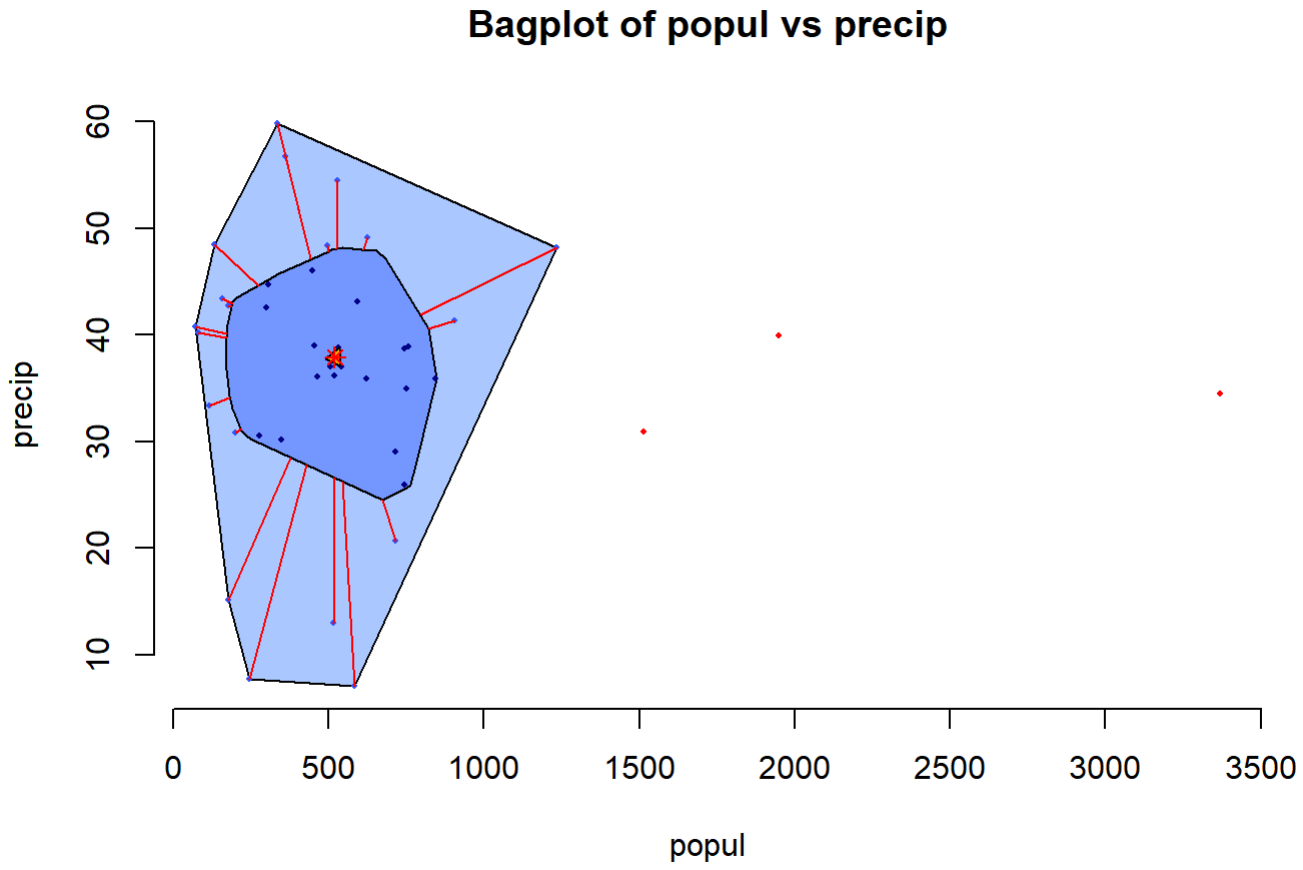

Bagplot of manu vs popul
popul
0
500
1500
2500
3500
0
500
1000
1500
2000
2500
3000
manu

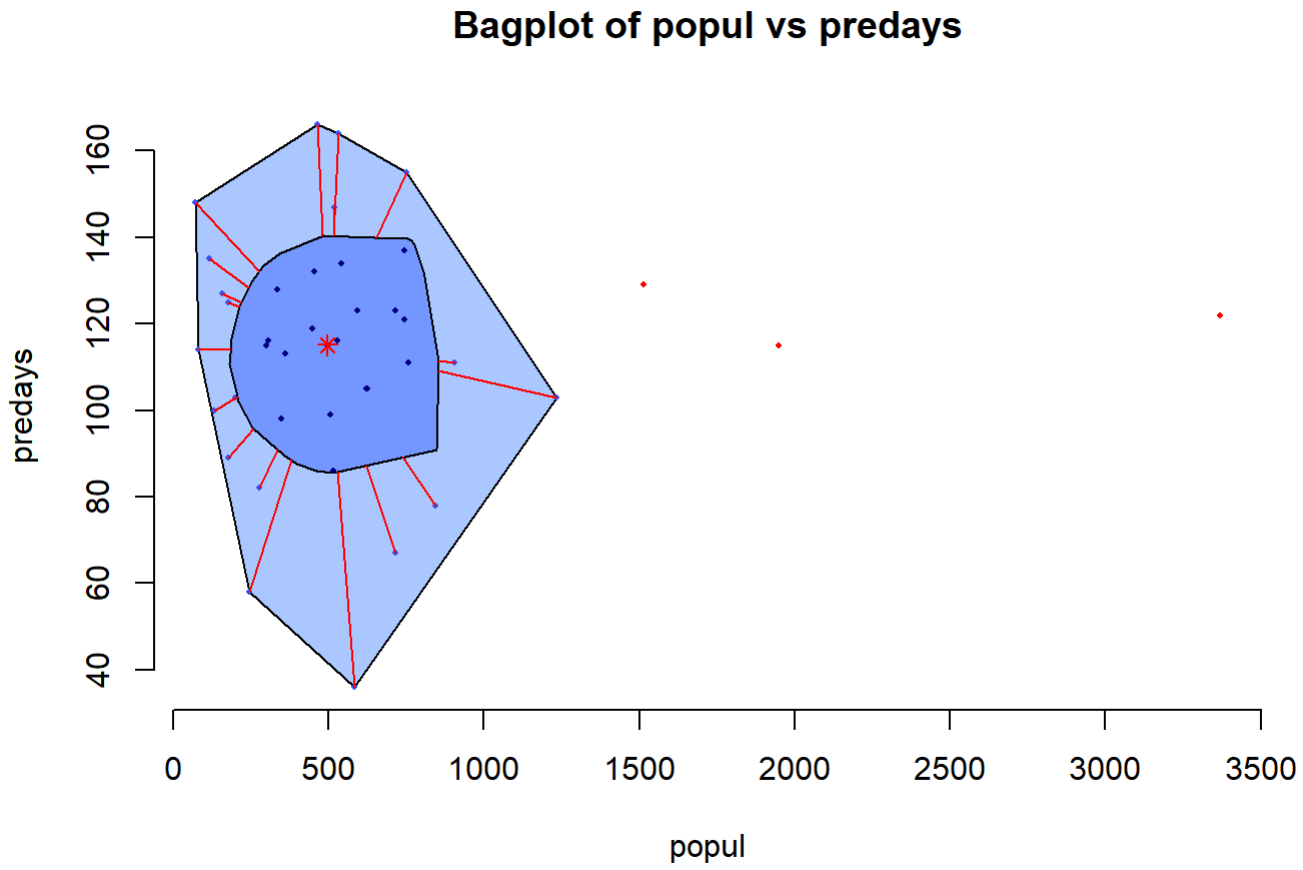

Bagplot of manu vs wind
wind
6
7
8
9
10
11
12
0
500
1000
1500
2000
2500
3000
manu

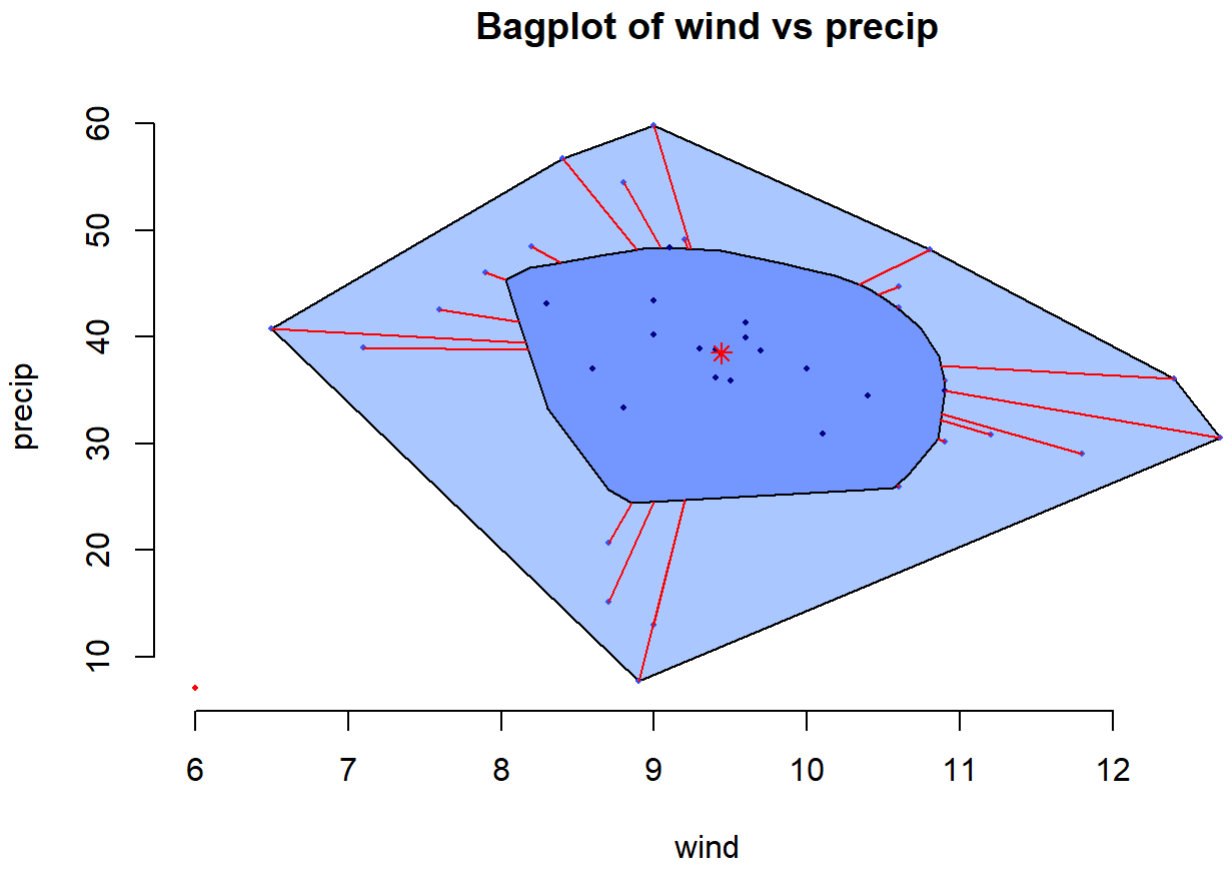

Bagplot of manu vs precip
precip
10
20
30
40
50
60
0
500
1000
1500
2000
2500
3000
manu

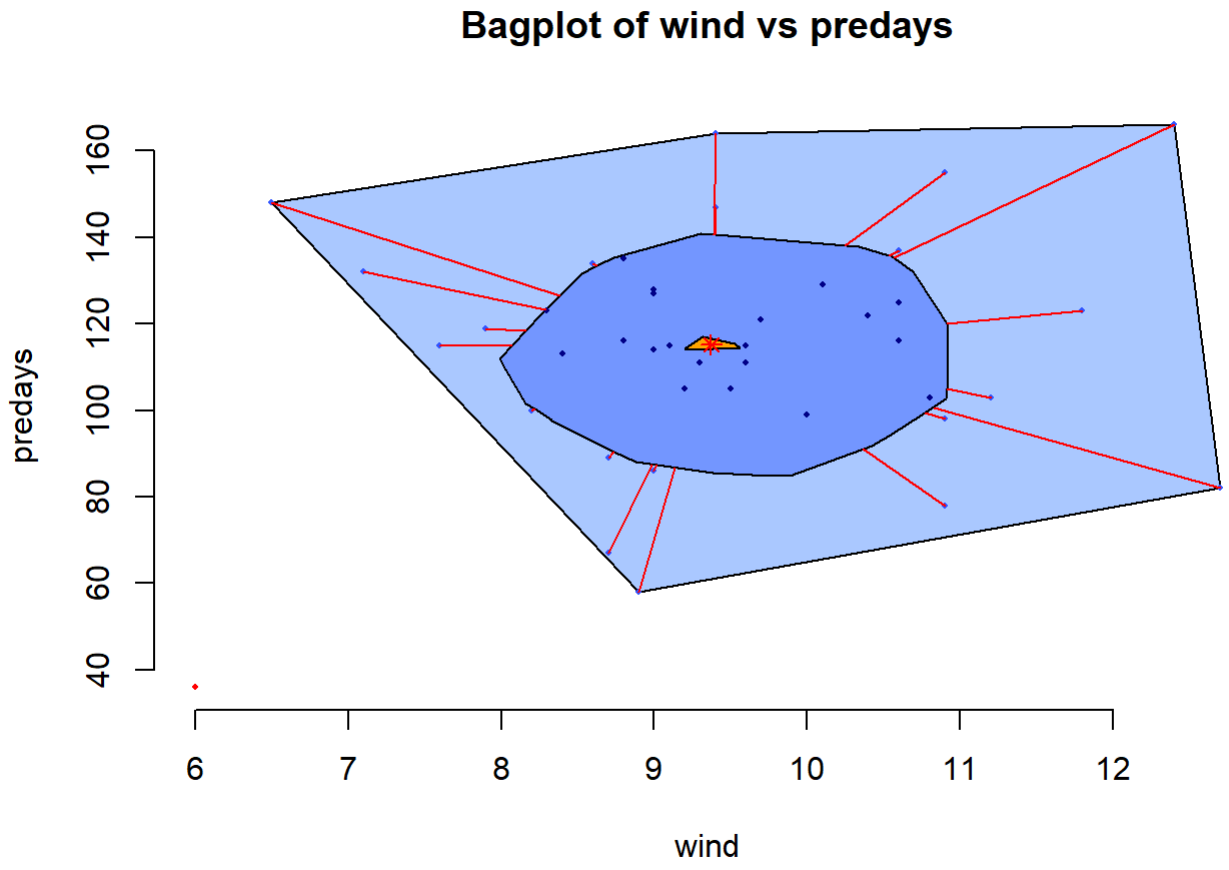

Bagplot of manu vs predays
predays
40
60
80
100
120
140
160
0
500
1000
1500
2000
2500
3000
manu

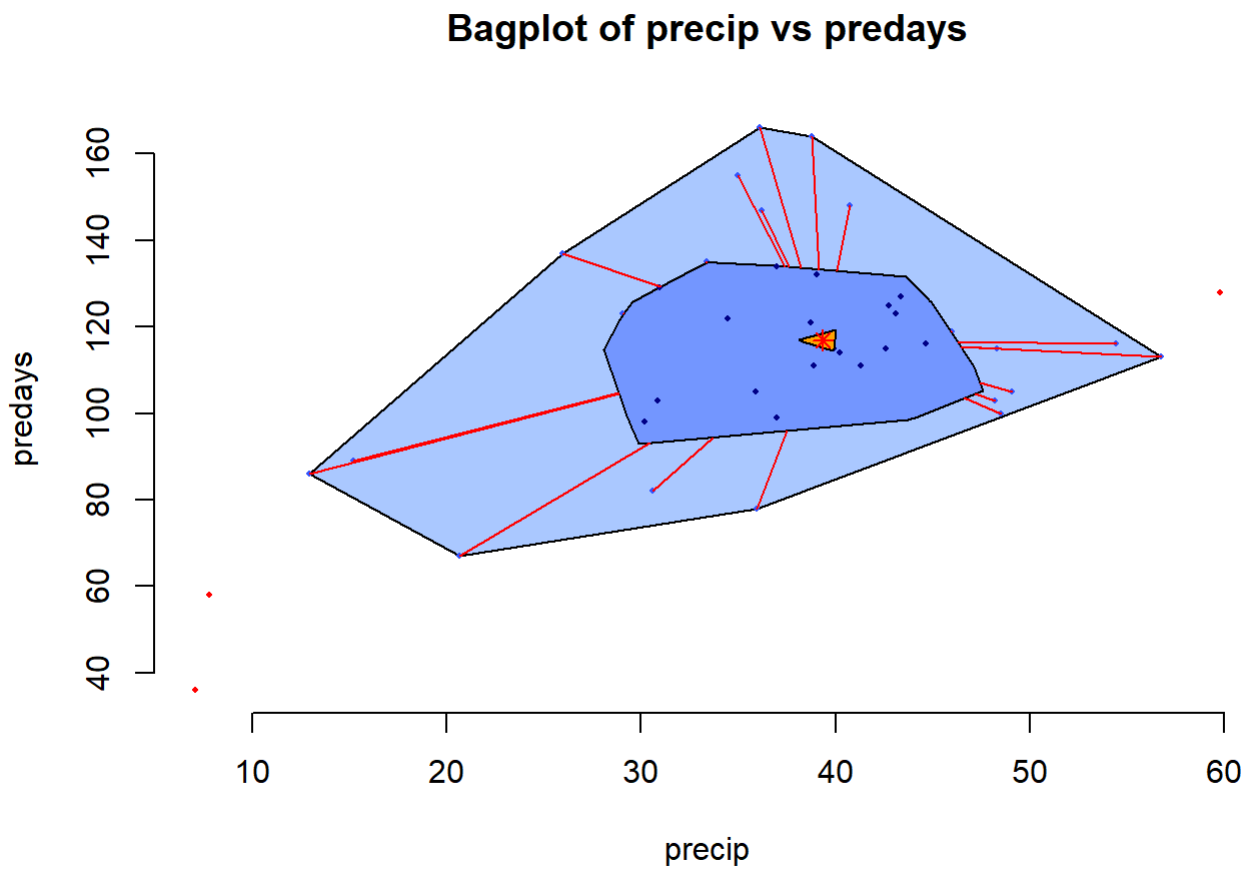

Bagplot of popul vs wind
wind
6
7
8
9
10
11
12
0
500
1000
1500
2000
2500
3000
3500
popul


Bagplot of popul vs precip
precip
10
20
30
40
50
60
0
500
1000
1500
2000
2500
3000
3500
popul


Bagplot of popul vs predays
predays
40
60
80
100
120
140
160
0
500
1000
1500
2000
2500
3000
3500
popul


Bagplot of wind vs precip
precip
10
20
30
40
50
60
6
7
8
9
10
11
12
wind


Bagplot of wind vs predays
predays
40
60
80
100
120
140
160
6
7
8
9
10
11
12
wind




```
cor_all <- cor(USairpollution)
```

```
remove_bagplot_outliers <- function(x, y) {
  bag <- bagplot(x, y, plot = FALSE)
  outliers <- bag$pxy.outlier
  x_clean <- ifelse(paste(x, y) %in% paste(outliers[,1], outliers[,2]), NA, x)
  y_clean <- ifelse(paste(x, y) %in% paste(outliers[,1], outliers[,2]), NA, y)
  return(data.frame(x_clean, y_clean))
}

USairpollution_clean <- USairpollution

for (i in 1:(ncol(USairpollution)-1)) {
  for (j in (i+1):ncol(USairpollution)) {
    cleaned_data <- remove_bagplot_outliers(USairpollution[,i], USairpollution[,j])
    USairpollution_clean[,i] <- cleaned_data$x_clean
    USairpollution_clean[,j] <- cleaned_data$y_clean
  }
}
```

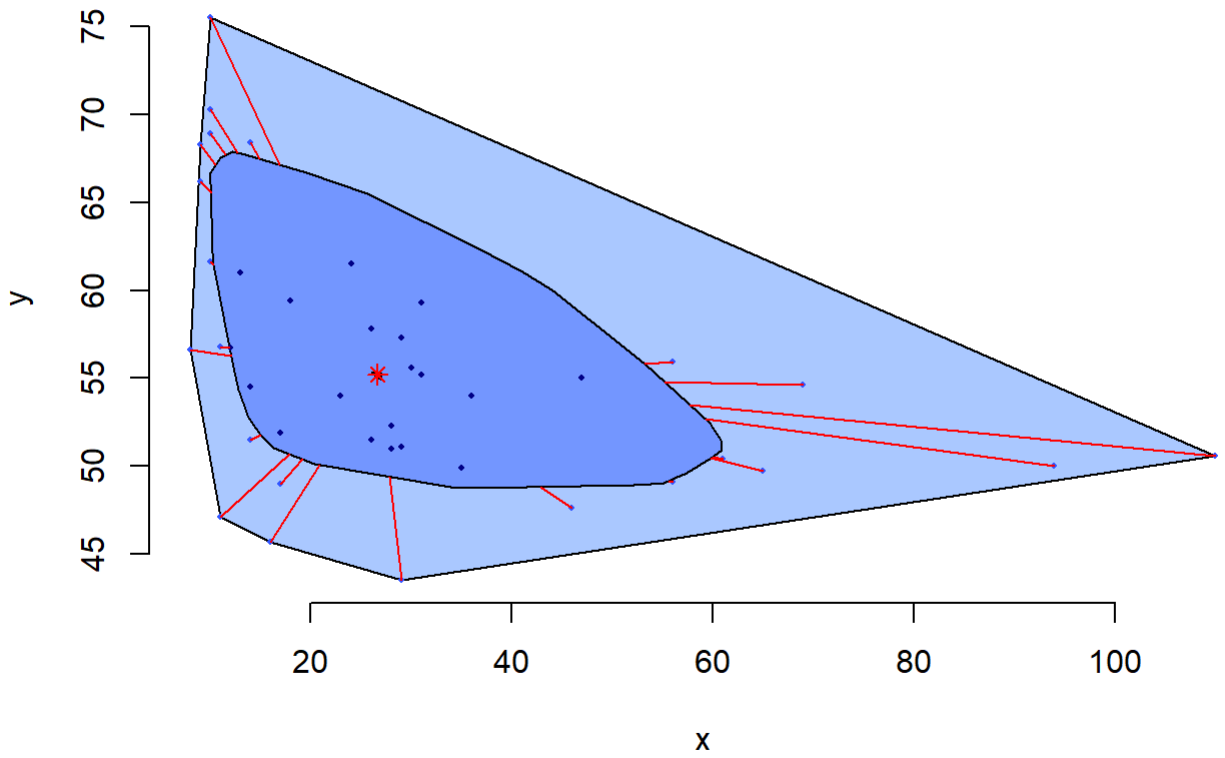
```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



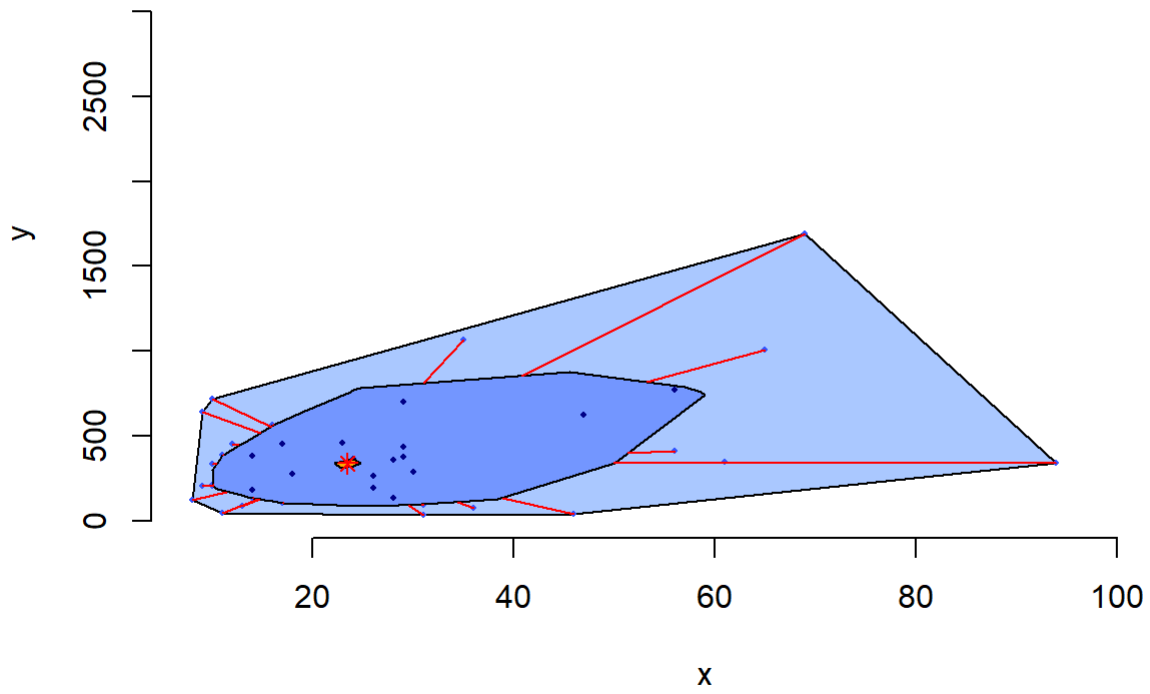

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



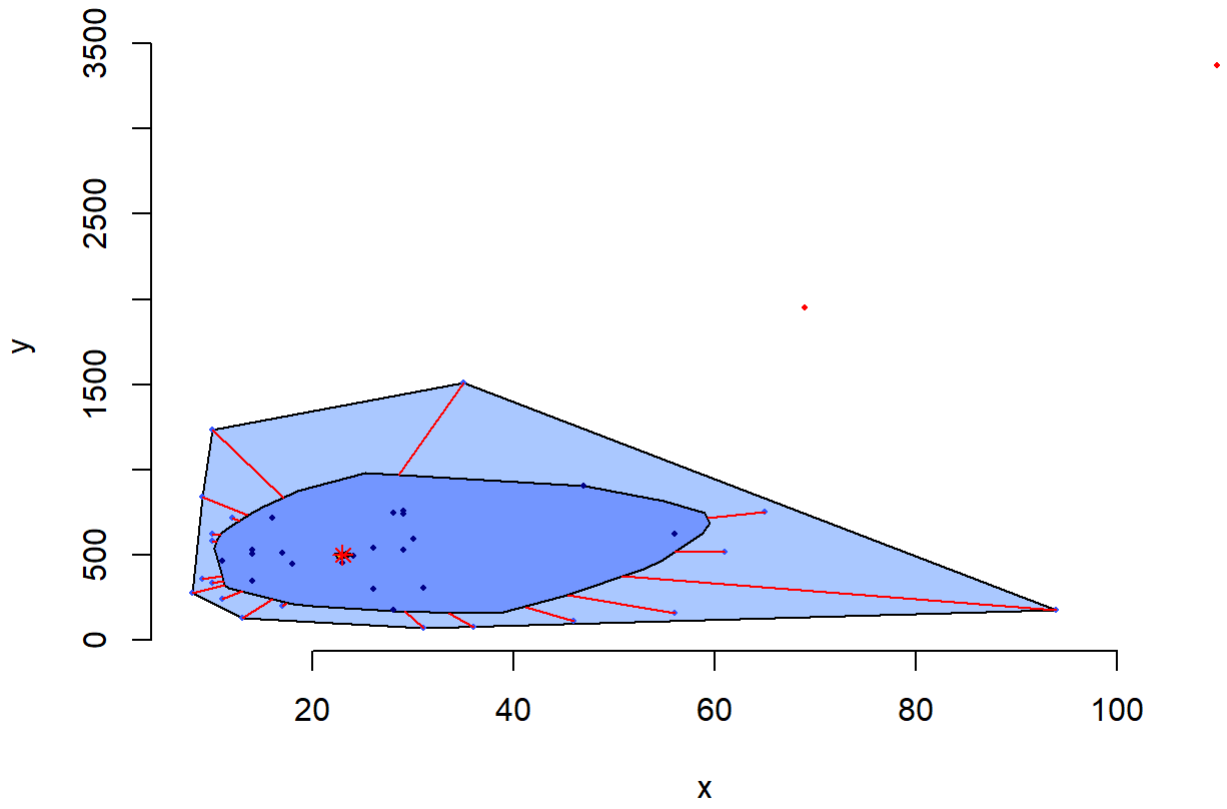

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



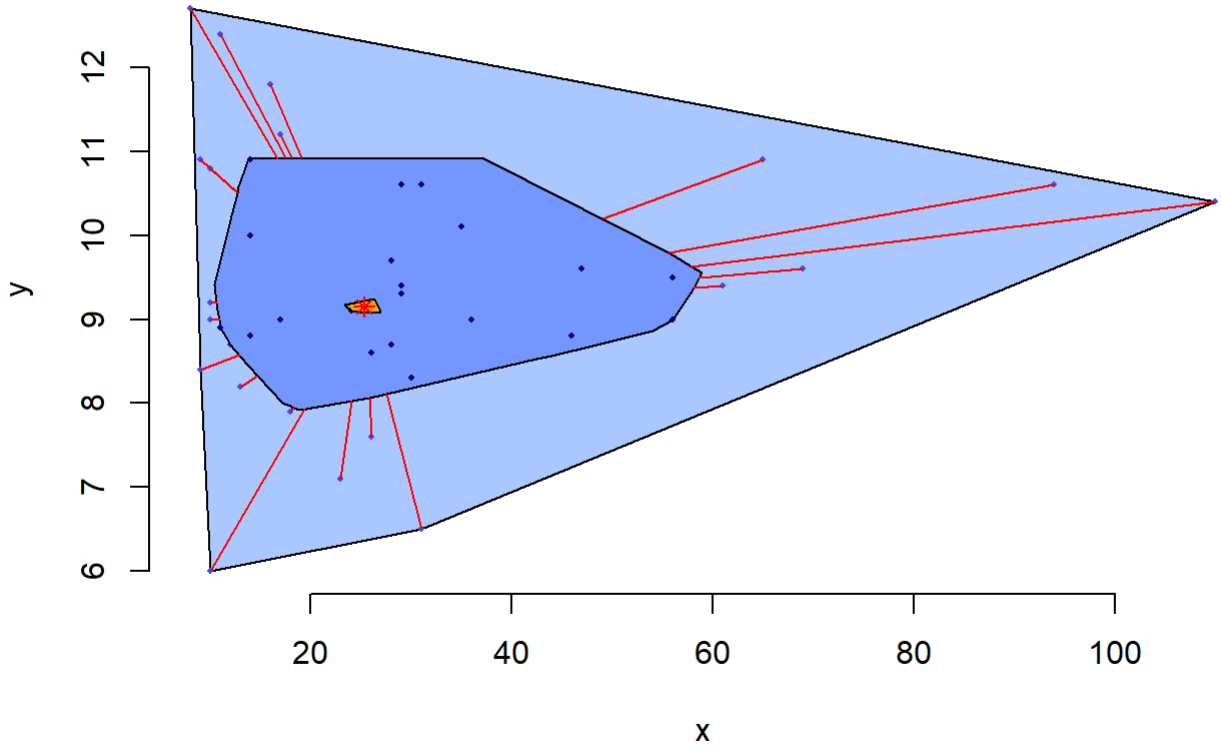

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



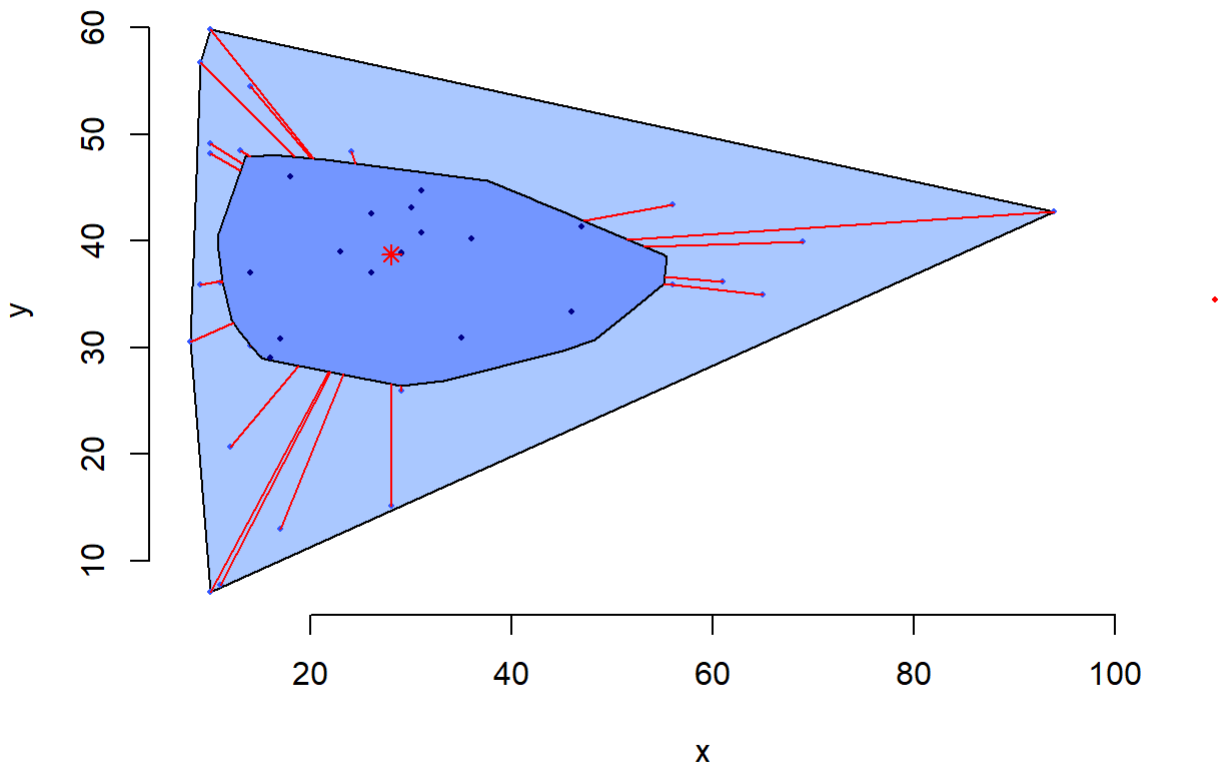

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



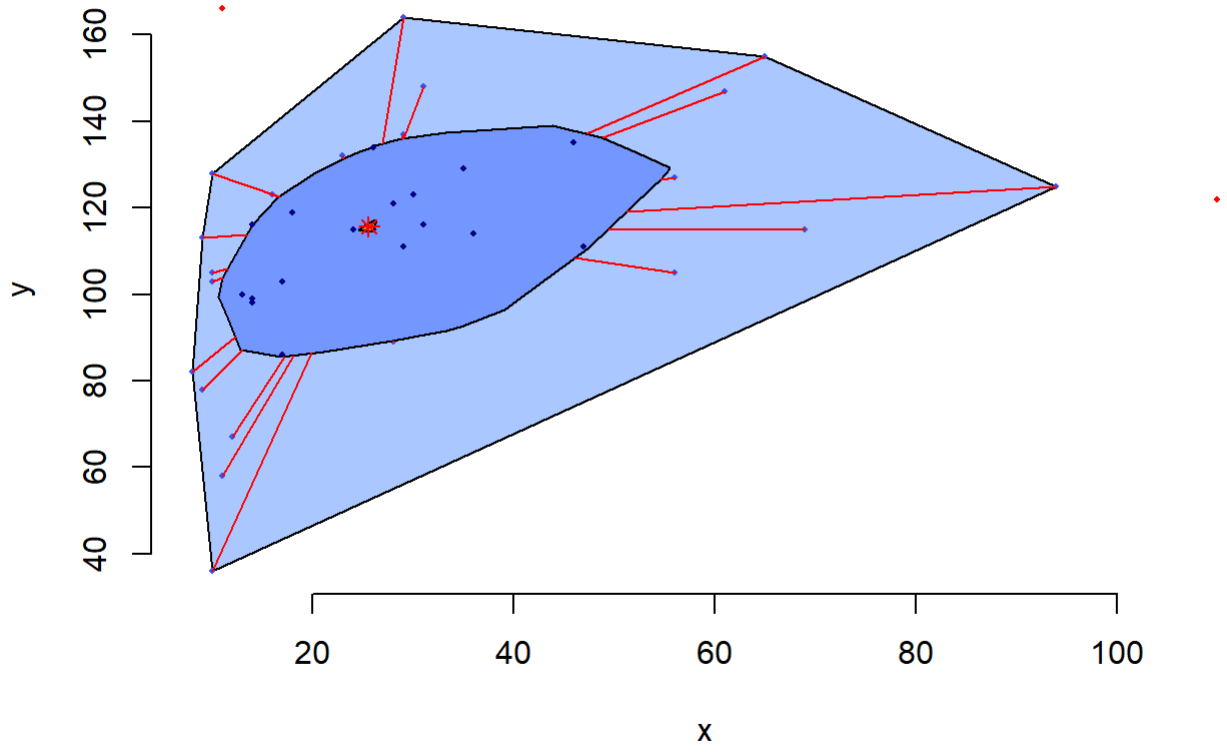

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



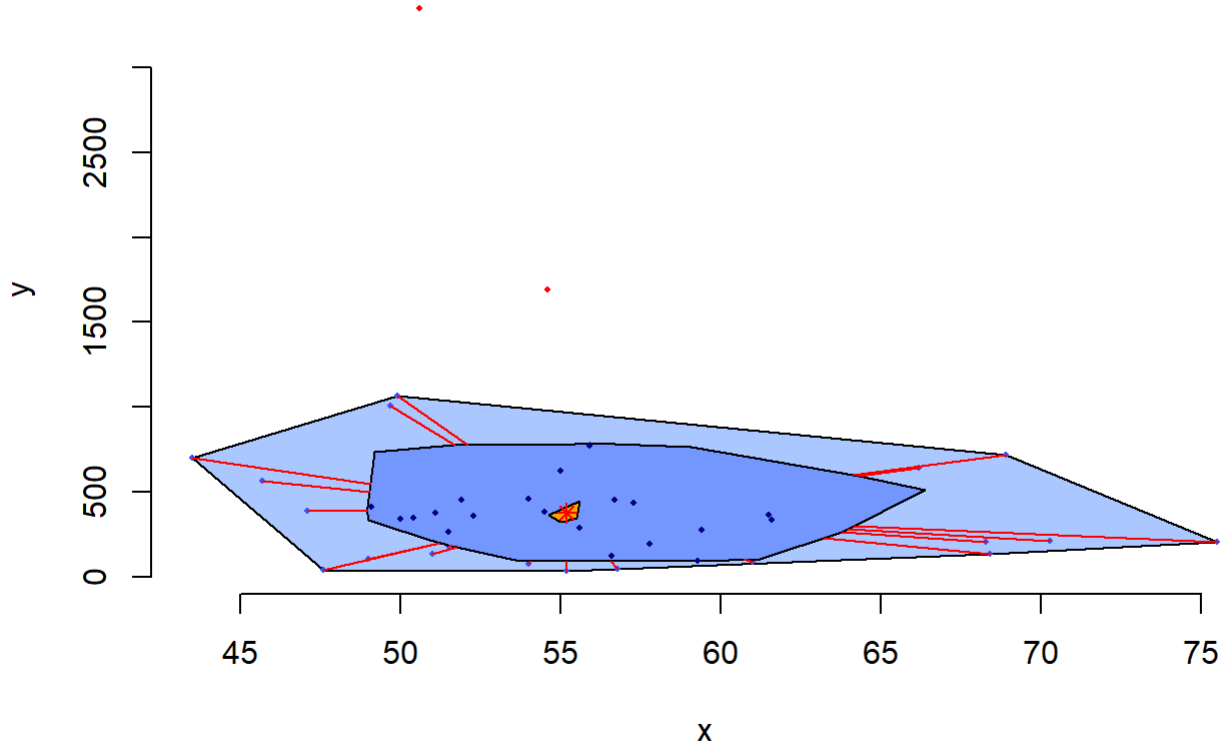

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



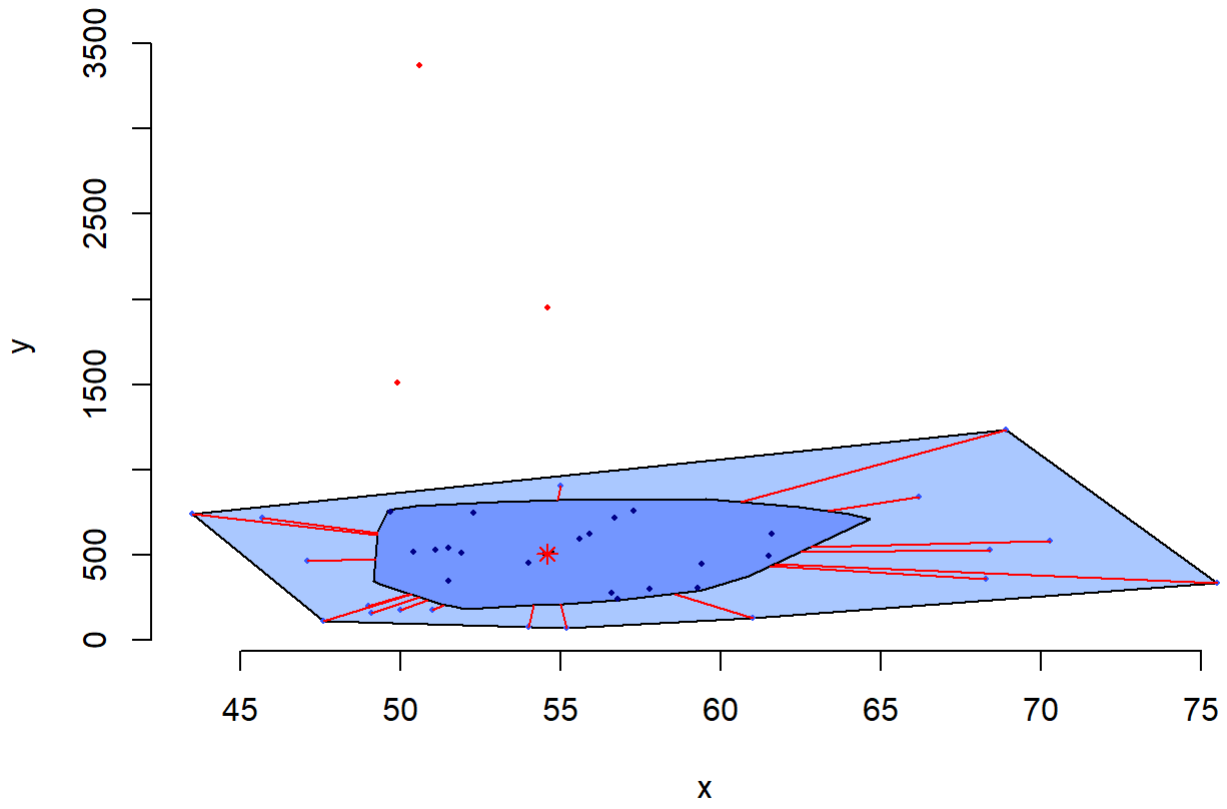

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



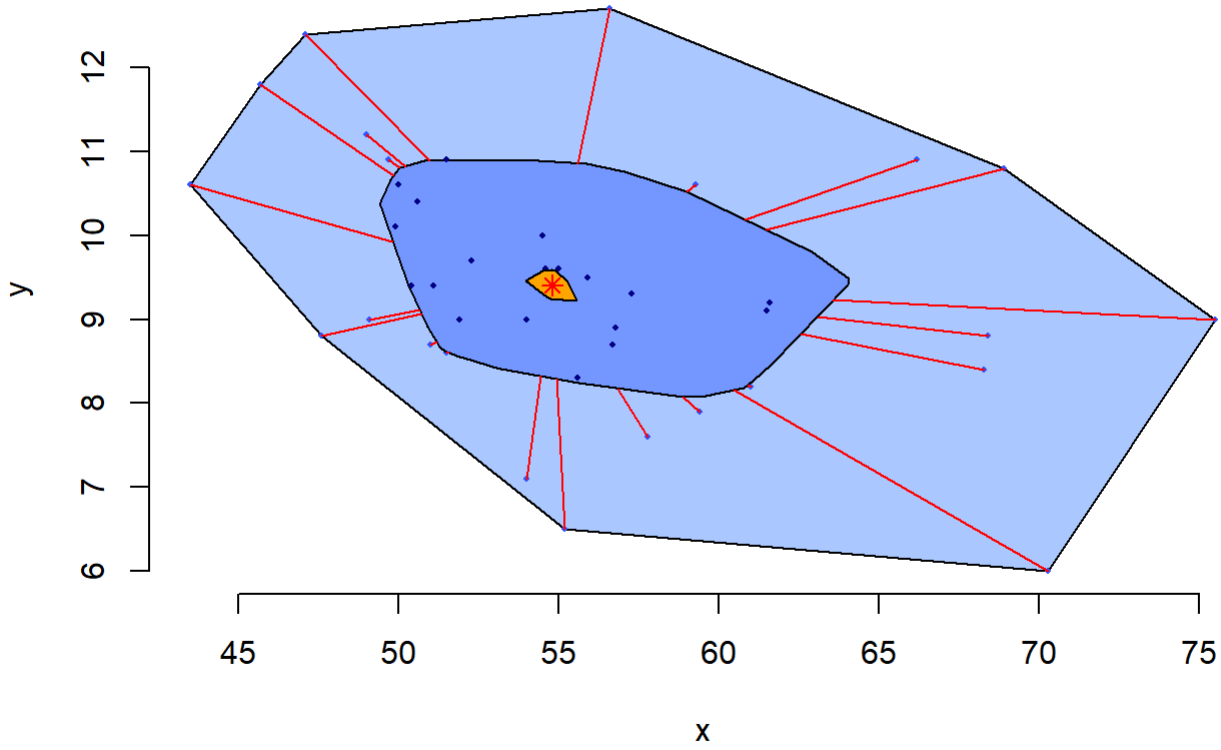

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



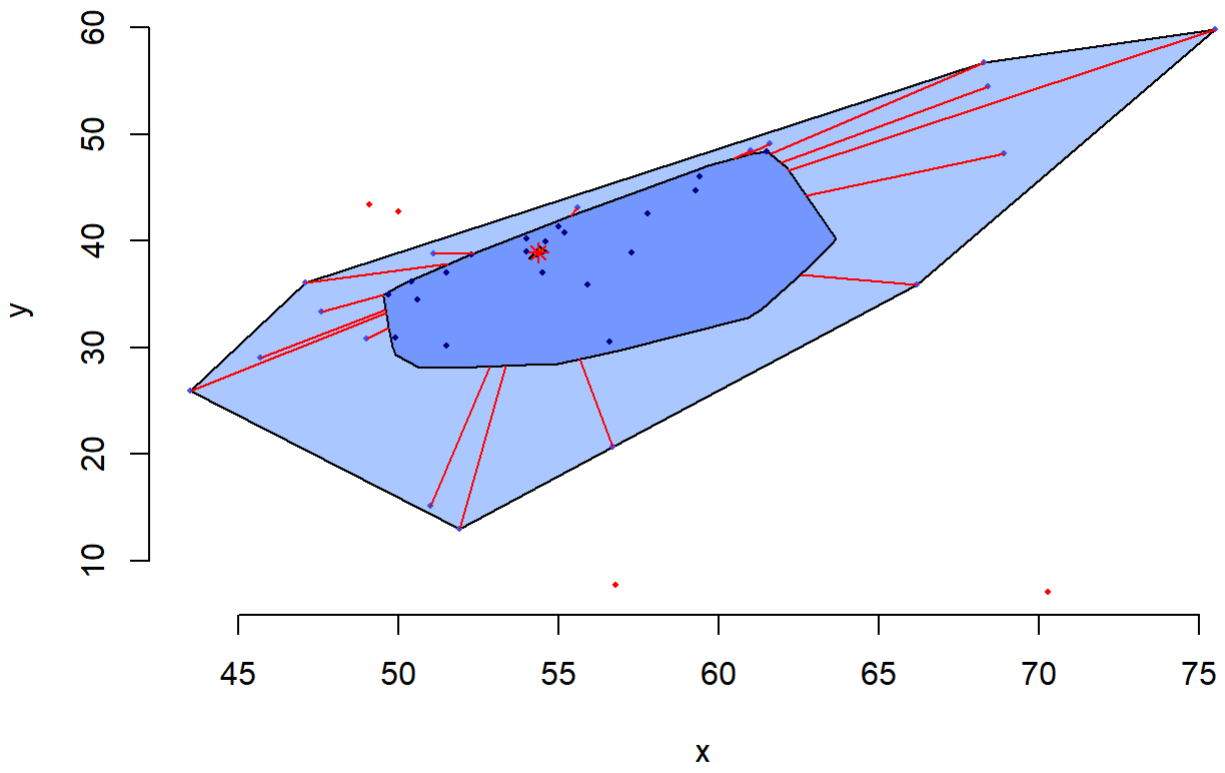

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



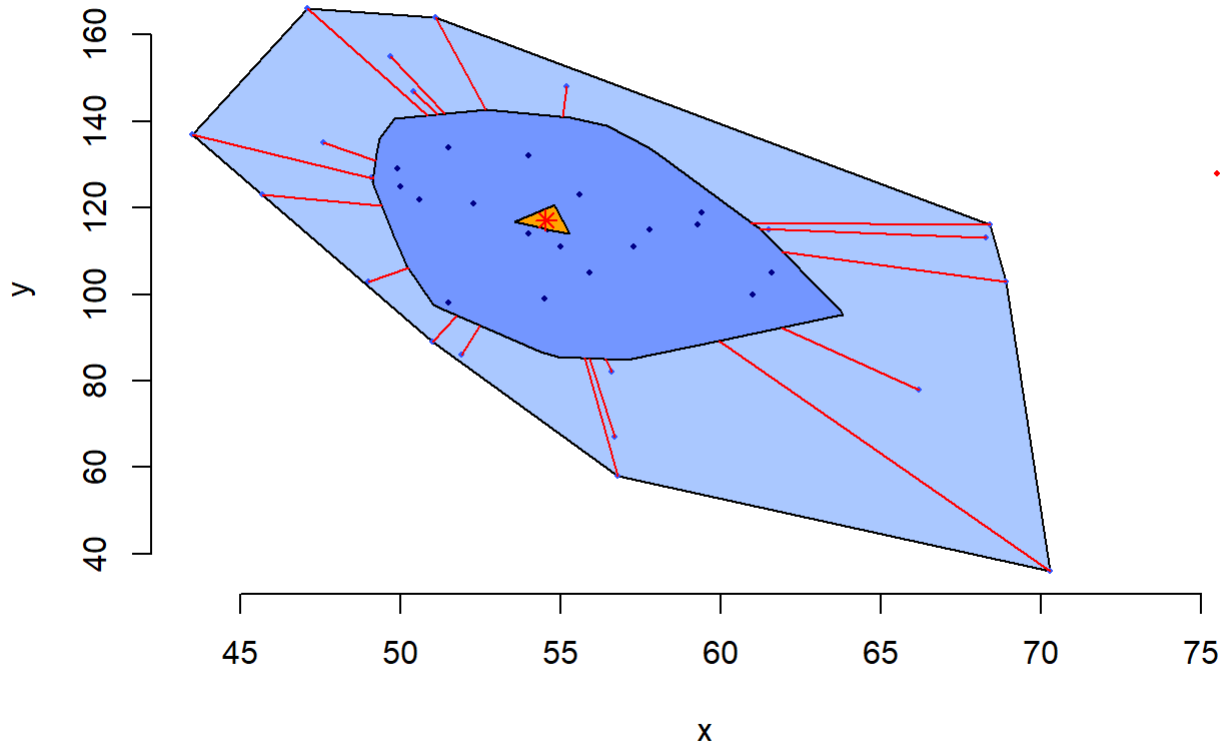

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



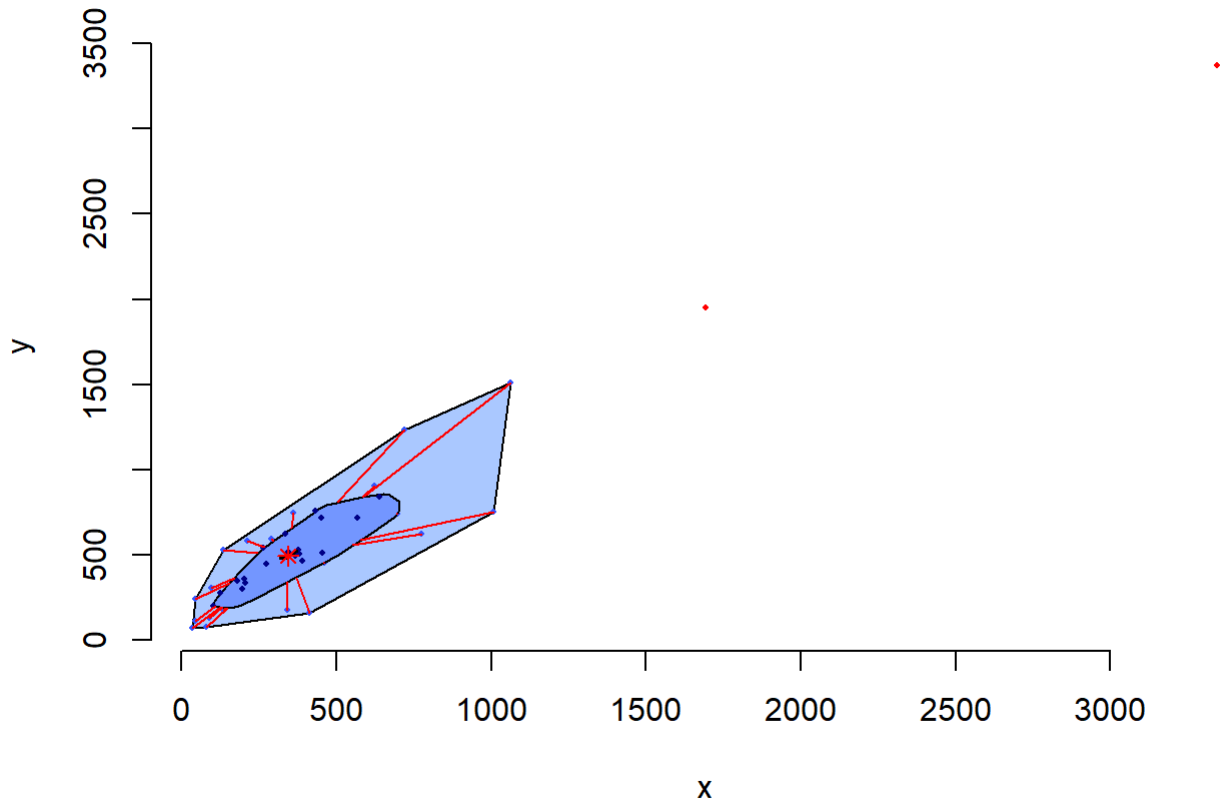

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



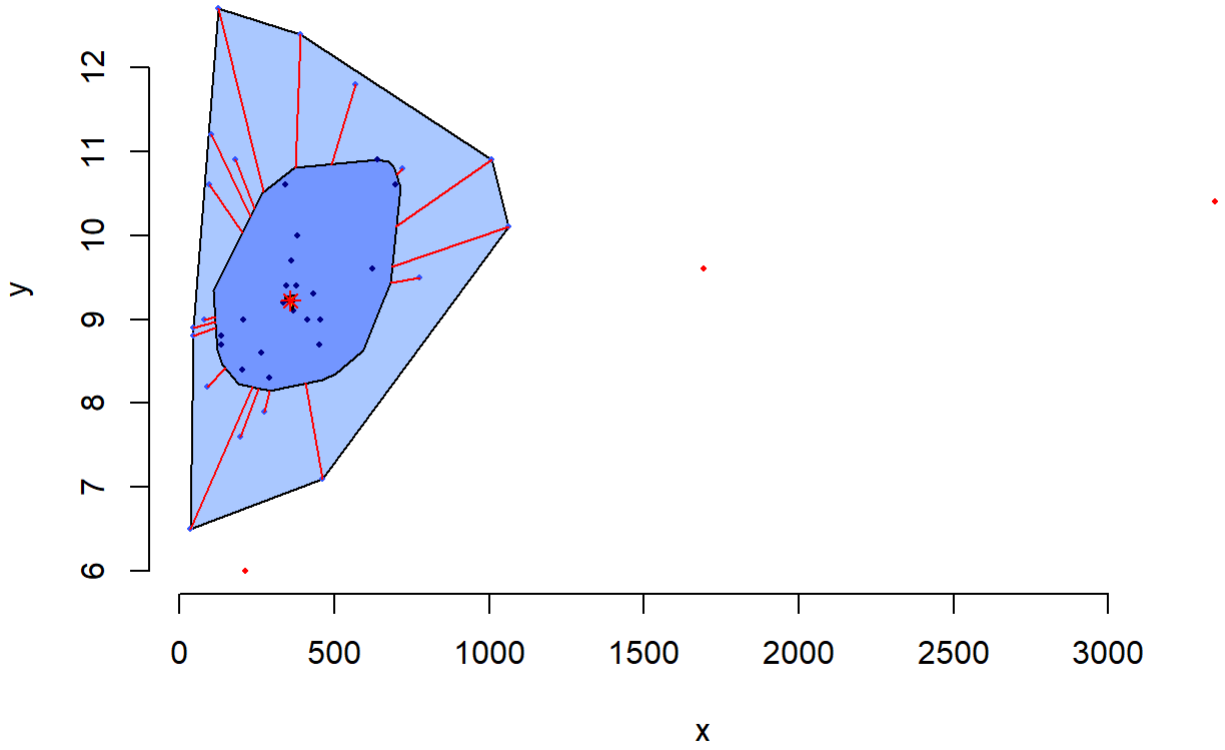

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



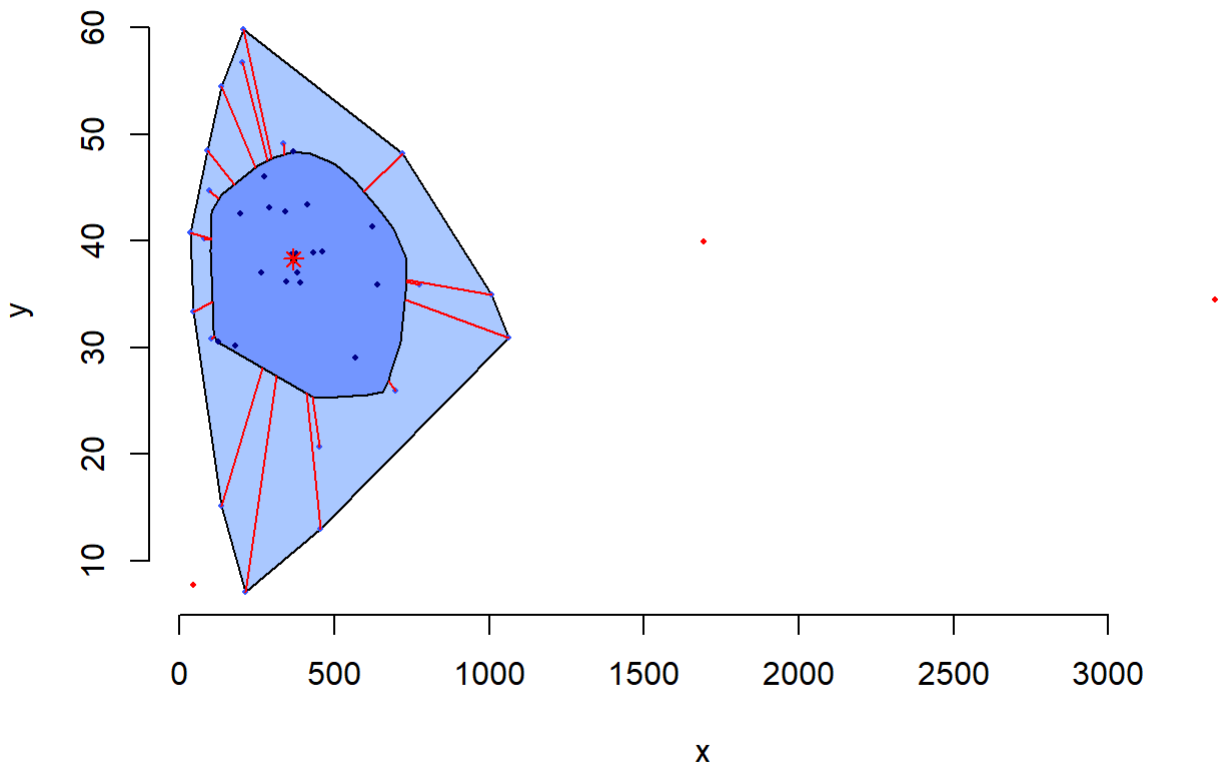

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



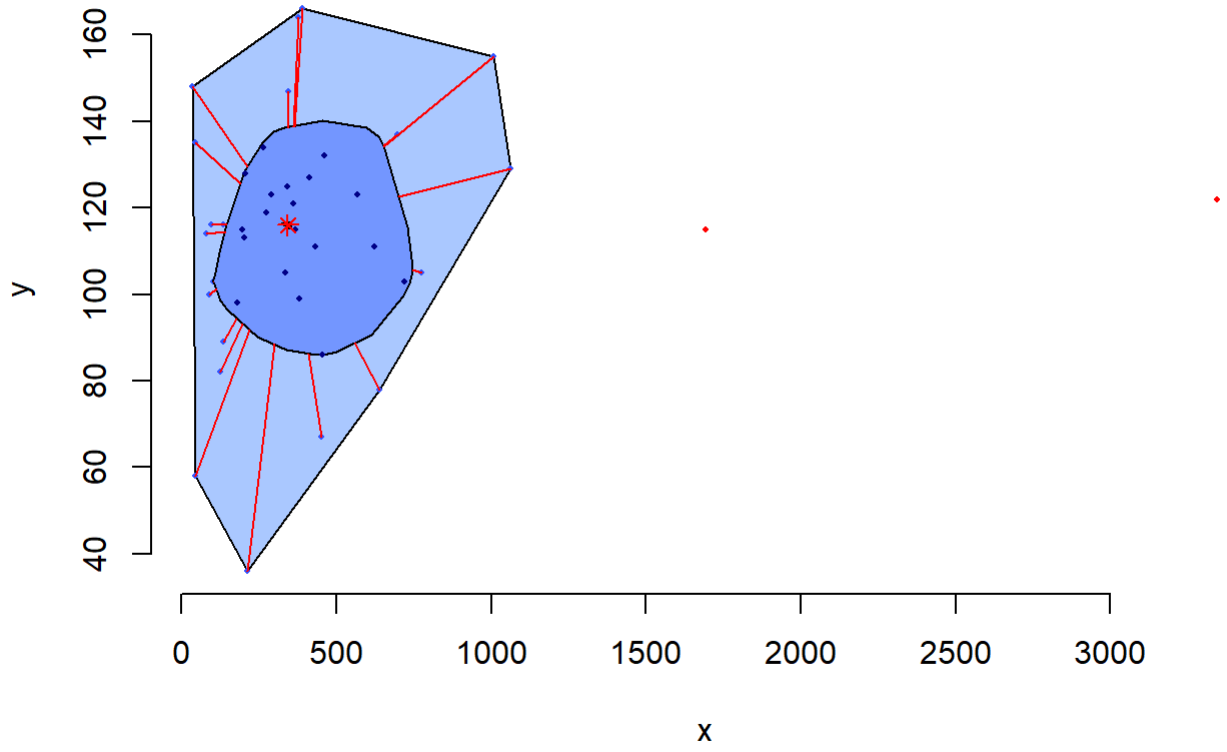

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



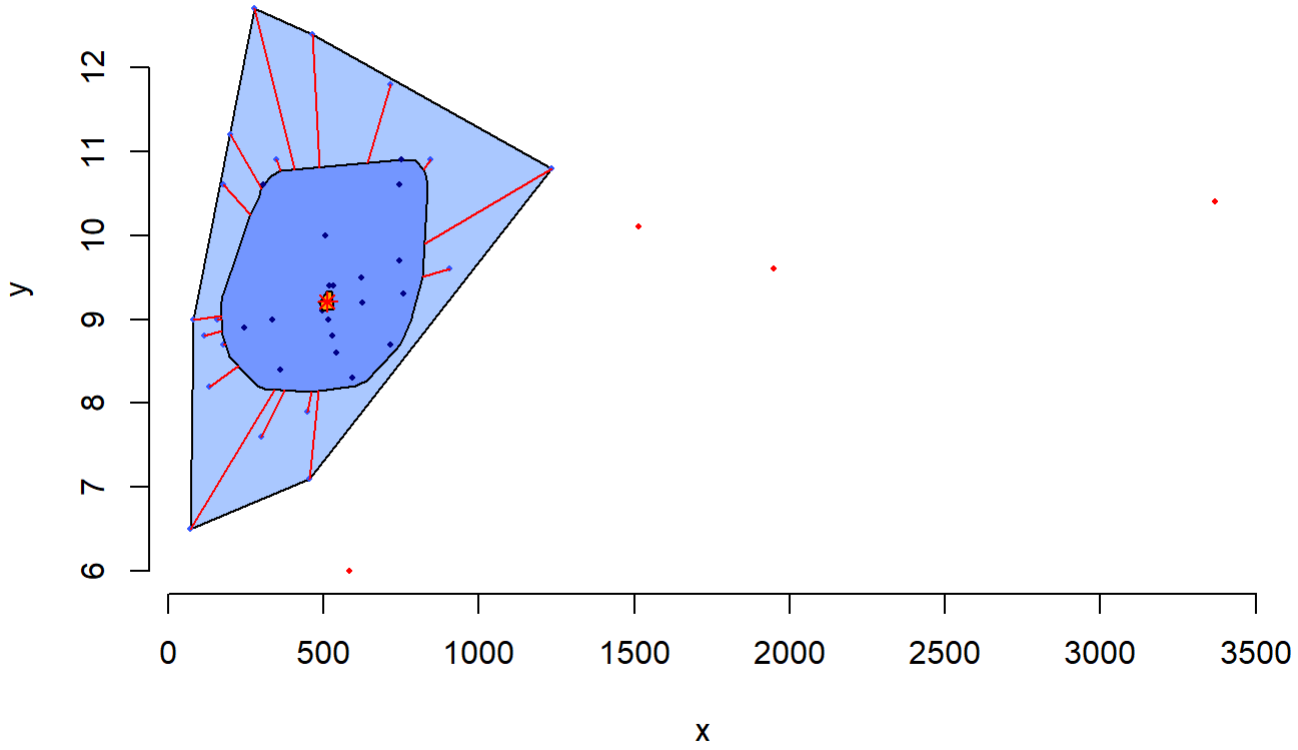

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



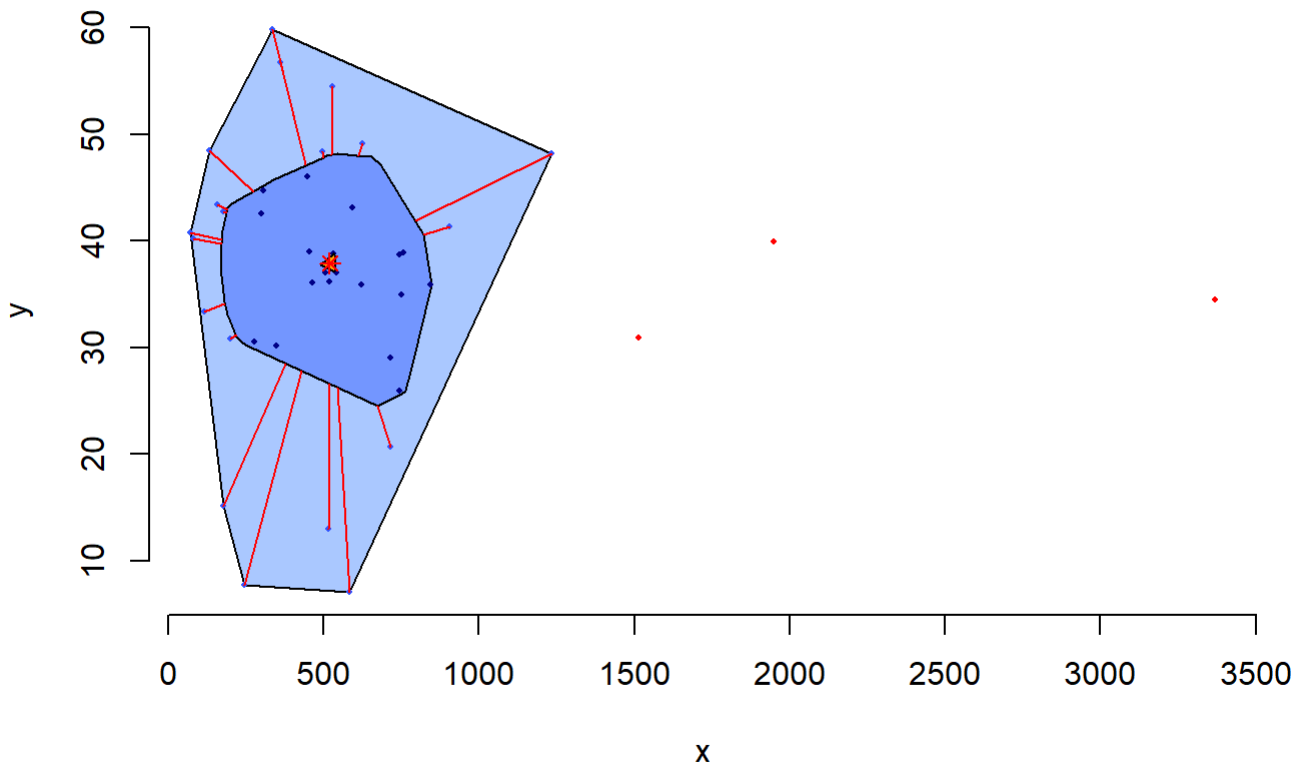

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



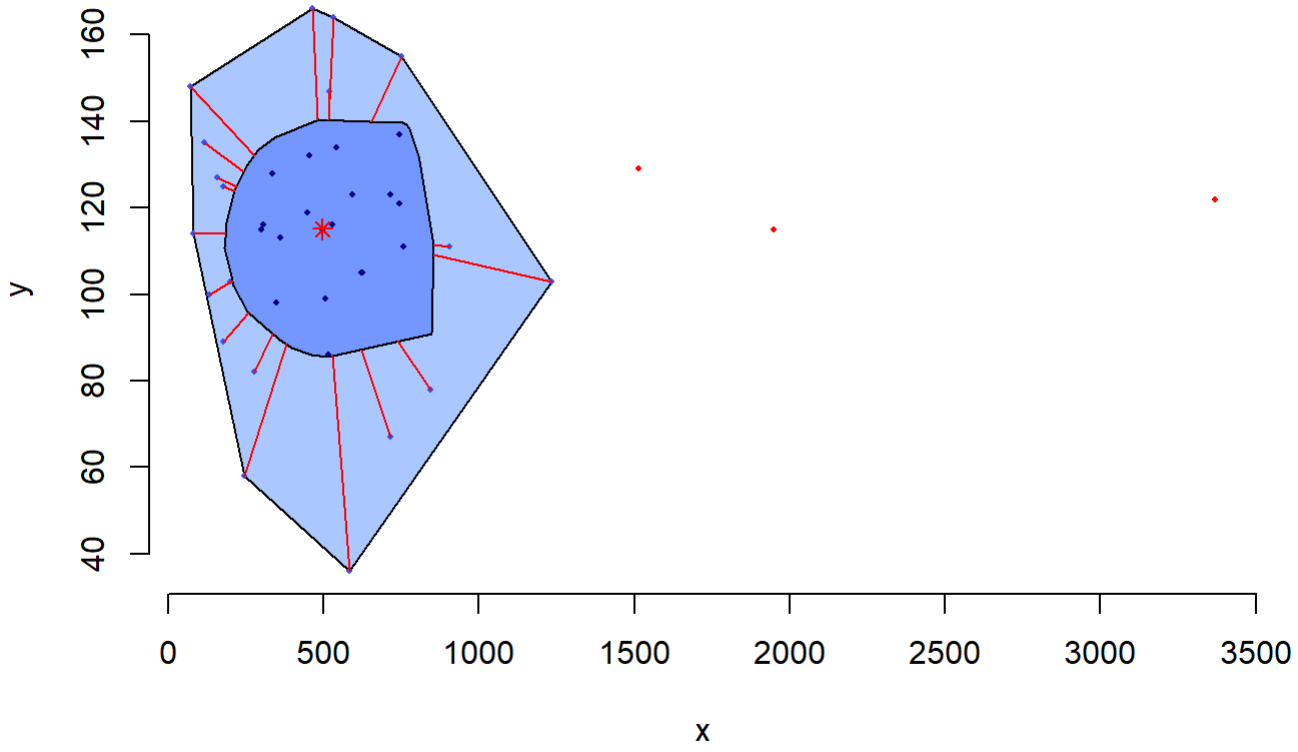

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



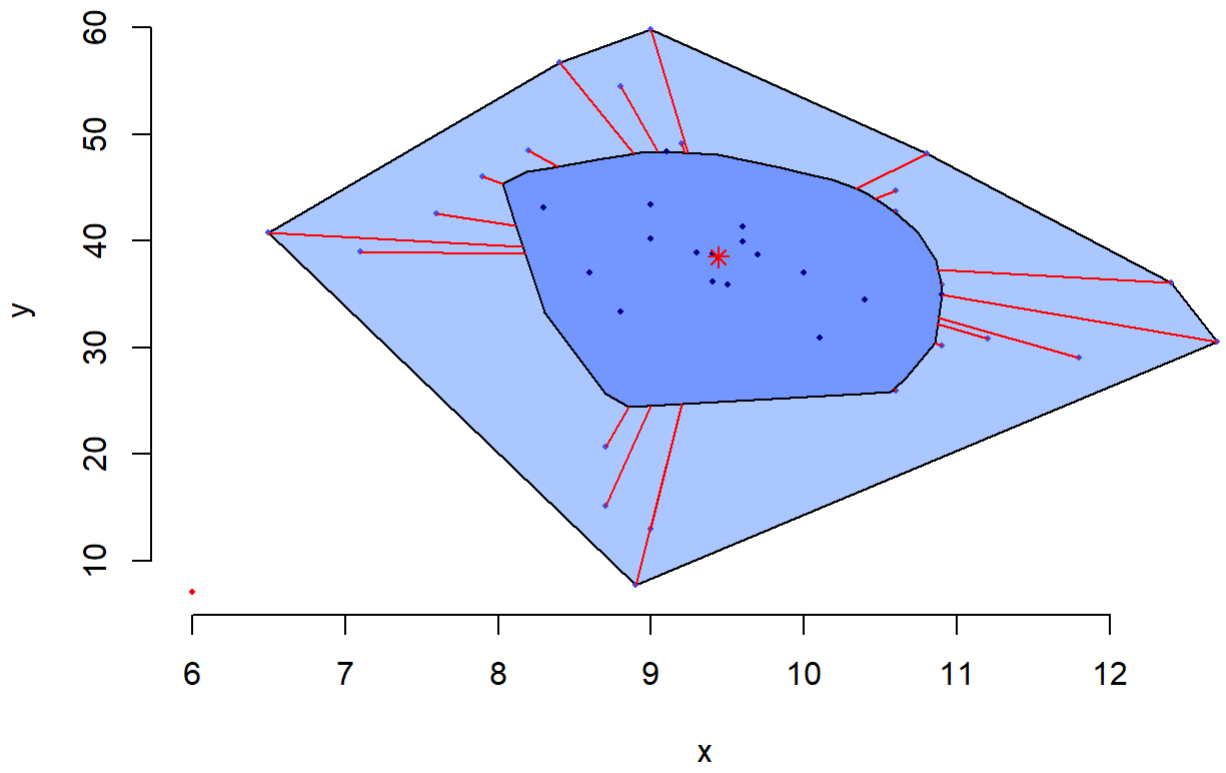

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



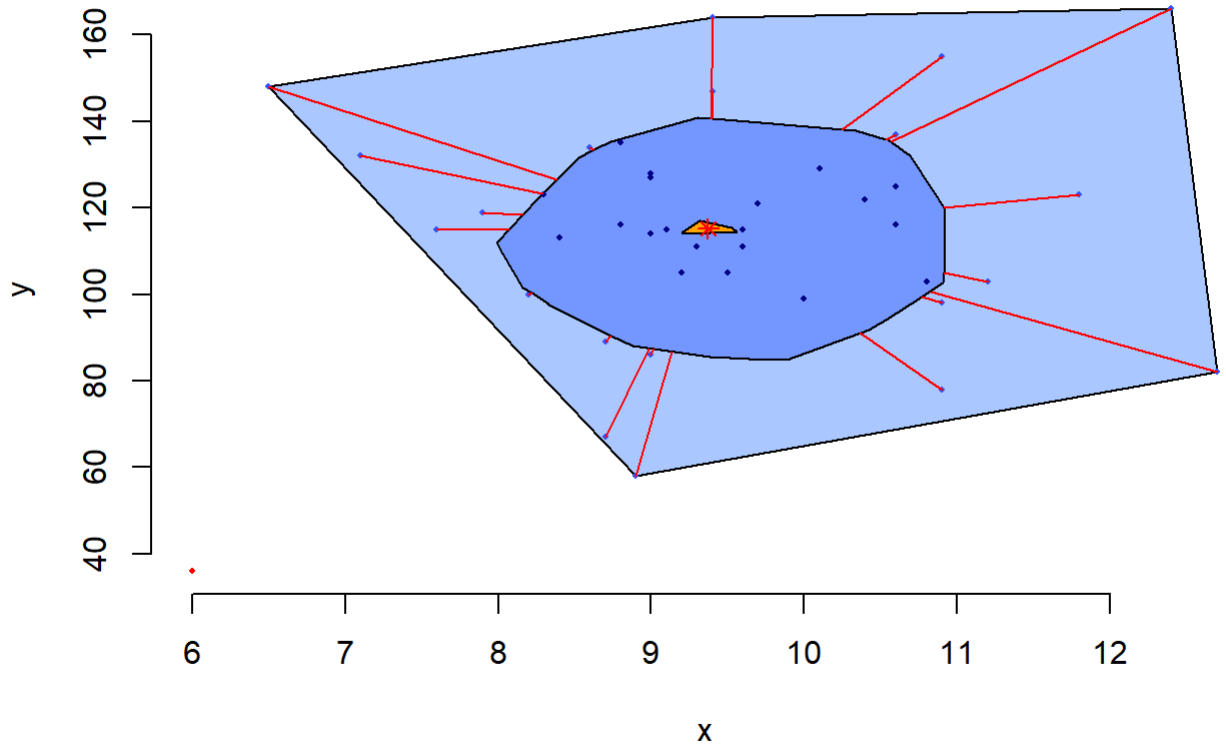

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```



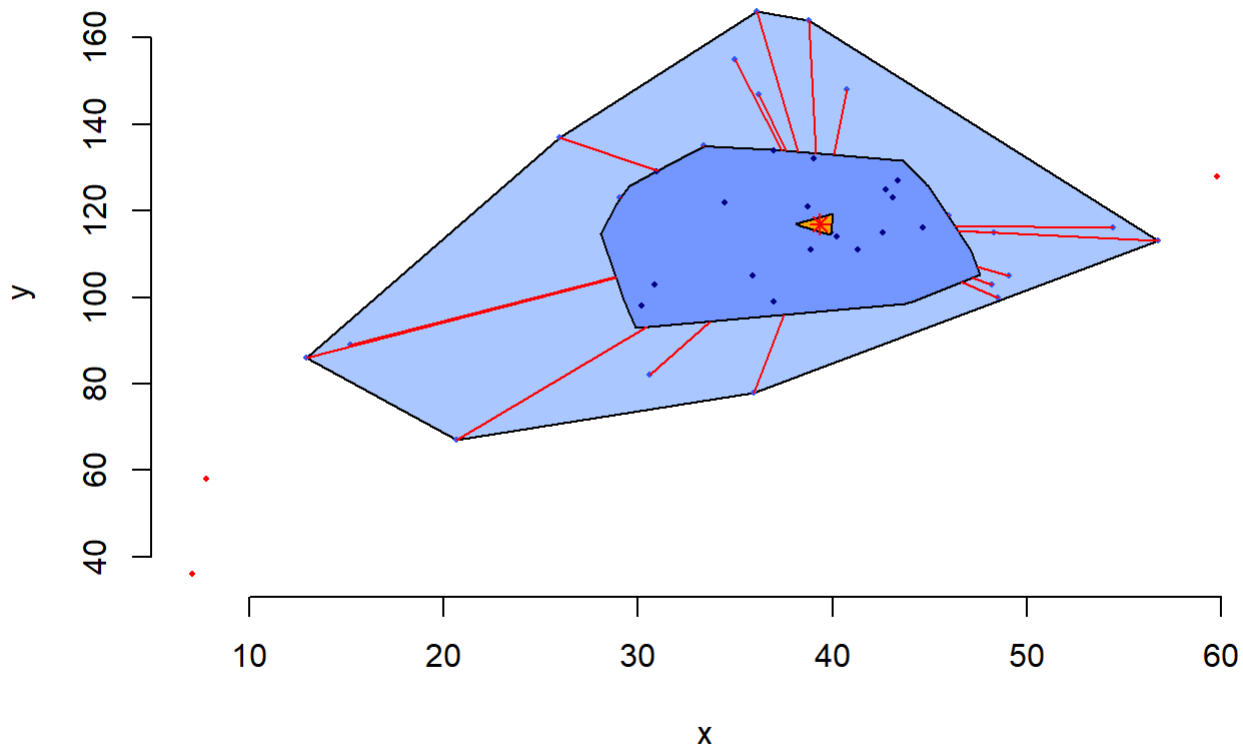

```
## Warning in plot.window(...): "plot" 不是一個繪圖參數
```

```
## Warning in plot.xy(xy, type, ...): "plot" 不是一個繪圖參數
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
## Warning in axis(side = side, at = at, labels = labels, ...): "plot"
## 不是一個繪圖參數
```

```
## Warning in box(...): "plot" 不是一個繪圖參數
```

```
## Warning in title(...): "plot" 不是一個繪圖參數
```




```
USairpollution_clean <- na.omit(USairpollution_clean)
```

```
cor_no_outliers <- cor(USairpollution_clean, use = "complete.obs")
print("Correlation matrix using all data:")
```

```
## [1] "Correlation matrix using all data:"
```

```
print(cor_all)
```

```
##                SO2        temp        manu       popul        wind      precip
## SO2      1.00000000 -0.43360020  0.64476873  0.49377958  0.09469045  0.05429434
## temp    -0.43360020  1.00000000 -0.19004216 -0.06267813 -0.34973963  0.38625342
## manu     0.64476873 -0.19004216  1.00000000  0.95526935  0.23794683 -0.03241688
## popul    0.49377958 -0.06267813  0.95526935  1.00000000  0.21264375 -0.02611873
## wind     0.09469045 -0.34973963  0.23794683  0.21264375  1.00000000 -0.01299438
## precip   0.05429434  0.38625342 -0.03241688 -0.02611873 -0.01299438  1.00000000
## predays  0.36956363 -0.43024212  0.13182930  0.04208319  0.16410559  0.49609671
##             predays
## SO2      0.36956363
## temp    -0.43024212
## manu     0.13182930
## popul    0.04208319
## wind     0.16410559
## precip   0.49609671
## predays  1.00000000
```

```
print("Correlation matrix after removing outliers:")
```

```
## [1] "Correlation matrix after removing outliers:"
```

```
print(cor_no_outliers)
```

```
##                  SO2        temp        manu        popul        wind
## SO2      1.000000000 -0.47129902  0.23535824 -0.173840662  0.01028791
## temp    -0.471299015  1.00000000 -0.08526349  0.240434760 -0.17947144
## manu     0.235358239 -0.08526349  1.00000000  0.774299263  0.31840355
## popul   -0.173840662  0.24043476  0.77429926  1.000000000  0.29836104
## wind     0.010287906 -0.17947144  0.31840355  0.298361037  1.00000000
## precip   0.007658561  0.62387089 -0.15043564  0.002967901 -0.25045584
## predays  0.510183724 -0.38891343  0.08488784 -0.088746234 -0.22375574
##               precip     predays
## SO2      0.007658561  0.51018372
## temp     0.623870894 -0.38891343
## manu    -0.150435642  0.08488784
## popul    0.002967901 -0.08874623
## wind    -0.250455835 -0.22375574
## precip   1.000000000  0.28535590
## predays  0.285355897  1.00000000
```

1. Correlation Matrix with Outliers:

When all data (including outliers) is included in the correlation calculation, the correlation matrix reflects the overall relationships between variables. However, these relationships may be significantly affected by the presence of outliers. Outliers can have a large impact on correlation coefficients, potentially making some variables appear to have higher or lower correlations than they truly do.

2. Correlation Matrix After Removing Outliers:

After removing the outliers, the correlation matrix provides a more accurate reflection of the true relationships between variables. Since the extreme effects of outliers are removed, the correlation coefficients usually become more stable and reliable.

Outliers often exaggerate the correlations between variables. After removing them, the absolute values of correlation coefficients tend to decrease, showing a relationship between variables that is closer to the actual trend in the data.

The correlations calculated without outliers better represent the behavior of the majority of data points, rather than being skewed by a few extreme cases.

3. Impact of Outliers on Correlation:

Outliers can significantly impact the results of a correlation analysis, especially in smaller datasets. Outliers can make variables seem highly correlated or uncorrelated, when in reality most of the data does not support such a relationship.

After removing outliers, the correlation results provide a more accurate overview of the trends in the data, reflecting the true relationships between variables.