Name: Millie Woo
[Student No: s3806940]

Question 1. PCA analysis with 5 plots

i. Prepare the dataset for input for a PCA via SAS.

Step 1: Read in the given excel file into SAS as a table work.readin

SAS Code:

```
PROC IMPORT OUT= work.readin DATAFILE= "/home/u41107333/Multivariate/0.
Project/MATH1309  Drug Bank DATA  set  for Assignment 2.xlsx"
            DBMS=xlsx REPLACE;
     SHEET="final_Drugbank dataset ";
     GETNAMES=YES;
RUN;
```

Step 2: Check if there is any missing value in the dataset.

SAS Code:

```
proc means data = work.readin nmiss;
run;
```

Output:

Variable	Label	N Miss
Drug#Card	Drug#Card	0
MW	MW	0
LogP	LogP	0
LogD	LogD	0
Hdonors	Hdonors	0
Hacceptors	Hacceptors	0
PSA	PSA	0
ROT	ROT	0
NATOM	NATOM	0
NRING	NRING	0
Oral#Corrected	Oral#Corrected	0
Score9_logD	Score9_logD	0
score9_logD_group_1	score9_logD_group_1	0

As there is no missing value, we can proceed with our analysis.

Name: Millie Woo
[Student No: s3806940]

ii. Perform a principal component analysis using SAS on the correlation matrix for the p=9 variables.

iii. Perform the procedures to obtain the following 5 plots:

- Scree Plot
- Profile Plot
- Component Pattern Plots
- Score plots
- Loading plots

These 2 requirements have been undertaken together with the following steps:

Step 1: Create a table work.drug which only contains the 9 molecular variables and score9_logD_group_1 (which is for question 2).

SAS Code:

```
proc sql;
create table work.drug as
select MW, LogP, LogD, Hdonors, Hacceptors, PSA, ROT, NATOM, NRING, score9_logD_group_1
from WORK.readin;
quit;
```

Step 2: Use proc princomp to perform component analysis and generate the scree plot, profile plot, component pattern plots

SAS Code:

```
proc princomp data=work.drug
     out = PCout
     STD /* optional: stdize PC scores to unit variance */
     plots(ncomp=9)=(scree profile pattern score);
     var MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
     ods output Eigenvectors=EV;
run;
```

Step 3: Use proc factor to generate the loading plots

SAS Code:

```
proc factor data=work.drug p=100
     method=principal
     plots(nplots=9)=(initloadings(vector));
     var MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
run;
```

All plots are located at Appendix [A1]

Name: Millie Woo
[Student No: s3806940]

The correlation matrix of the 9 variables are shown below:

Correlation Matrix									
	MW	**LogP**	**LogD**	**Hdonors**	**Hacceptors**	**PSA**	**ROT**	**NATOM**	**NRING**
MW	1.0000	0.0787	0.0077	0.5840	0.7348	0.7014	0.6649	0.9297	0.6450
LogP	0.0787	1.0000	0.8997	-.5151	-.3803	-.5148	0.0908	0.1709	0.2681
LogD	0.0077	0.8997	1.0000	-.5931	-.4201	-.5615	0.0360	0.0928	0.2398
Hdonors	0.5840	-.5151	-.5931	1.0000	0.7575	0.8657	0.4258	0.5285	0.1452
Hacceptors	0.7348	-.3803	-.4201	0.7575	1.0000	0.9005	0.4691	0.6059	0.3779
PSA	0.7014	-.5148	-.5615	0.8657	0.9005	1.0000	0.4969	0.5881	0.2700
ROT	0.6649	0.0908	0.0360	0.4258	0.4691	0.4969	1.0000	0.6929	0.1301
NATOM	0.9297	0.1709	0.0928	0.5285	0.6059	0.5881	0.6929	1.0000	0.6374
NRING	0.6450	0.2681	0.2398	0.1452	0.3779	0.2700	0.1301	0.6374	1.0000

a) Report the eigenvalues and the eigenvectors.

Eigenvalues of the correlation matrix are:

Eigenvalues of the Correlation Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
λ_1	4.73045355	2.12367349	0.5256	0.5256
λ_2	2.60678006	1.76125011	0.2896	0.8152
λ_3	0.84552994	0.58014848	0.0939	0.9092
λ_4	0.26538146	0.02695961	0.0295	0.9387
λ_5	0.23842185	0.12930442	0.0265	0.9652
λ_6	0.10911743	0.01571009	0.0121	0.9773
λ_7	0.09340733	0.02788250	0.0104	0.9877
λ_8	0.06552483	0.02014128	0.0073	0.9950
λ_9	0.04538355		0.0050	1.0000

Name: Millie Woo
[Student No: s3806940]

Eigenvectors of the correlation matrix are:

	Eigenvectors								
	Prin1	**Prin2**	**Prin3**	**Prin4**	**Prin5**	**Prin6**	**Prin7**	**Prin8**	**Prin9**
MW	0.401994	0.265940	-.032637	-.044052	0.001204	-.469565	0.023751	0.222720	-.702858
LogP	-.152728	0.549606	0.111241	0.369102	0.227459	0.050483	-.662800	0.157830	0.086582
LogD	-.184448	0.534108	0.086099	0.388775	-.023836	0.114953	0.712587	-.022743	0.008267
Hdonors	0.392157	-.204210	0.056273	0.220030	0.733720	0.379493	0.087364	-.179267	-.175507
Hacceptors	0.417390	-.069642	-.113475	0.545761	-.476305	-.034340	-.150867	-.507772	0.039562
PSA	0.429203	-.157166	-.015606	0.251764	-.108777	0.044233	0.102064	0.746880	0.381339
ROT	0.296647	0.197127	0.698575	-.327950	-.296005	0.426269	-.055372	-.031554	-.064790
NATOM	0.370392	0.319962	0.039111	-.304304	0.266056	-.443399	0.082460	-.275499	0.562331
NRING	0.208918	0.359791	-.687985	-.317603	-.110657	0.488148	-.040940	0.016552	-.014974

b) What percentage of the total sample variation is accounted for by each of the first PC, 2nd PC to the ninth PC? (5 marks)

Each PC is accounted for the total sample variation as below (in percentage):

	Percentage
PC1	52.561%
PC2	28.964%
PC3	9.395%
PC4	2.949%
PC5	2.649%
PC6	1.212%
PC7	1.038%
PC8	0.728%
PC9	0.504%

c) What percentage of the total sample variation is accounted for by the first PC to the ninth PC? (1 mark)

From first PC to ninth PC the total sample variation should add up to 100%.

d) Write out the formulation for the PCs. (5 marks)

Taking accuracy to 3 decimal places, $PC_i = e_i' X$, where X is the 9 molecular variables, the formulations are listed as below

PC1 = 0.402 * MW - 0.153 * LogP - 0.184 * LogD + 0.392 * Hdonors + 0.417 * Hacceptors

+ 0.429 * PSA + 0.297 * ROT + 0.370 * NATOM + 0.209 * NRING

PC2 = 0.266 * MW + 0.550 * LogP + 0.534 * LogD - 0.204 * Hdonors - 0.070 * Hacceptors

- 0.157 * PSA + 0.197 * ROT + 0.320 * NATOM + 0.360 * NRING

PC3 = - 0.033 * MW + 0.111 * LogP + 0.086 * LogD + 0.056 * Hdonors - 0.113 * Hacceptors

Name: Millie Woo
[Student No: s3806940]

- 0.016 * PSA + 0.699 * ROT + 0.039 * NATOM - 0.688 *NRING

PC4 = - 0.044 * MW + 0.369 * LogP + 0.389 * LogD + 0.220 * Hdonors + 0.546 * Hacceptors

+ 0.252 * PSA - 0.328 * ROT - 0.304 * NATOM - 0.318 * NRING

PC5 = 0.001 * MW + 0.227 * LogP - 0.024 * LogD + 0.734 * Hdonors - 0.476 * Hacceptors

- 0.109 * PSA - 0.296 * ROT + 0.266 * NATOM - 0.110 * NRING

PC6 = - 0.470 * MW + 0.050 * LogP + 0.115 * LogD + 0.380 * Hdonors - 0.034 * Hacceptors

+ 0.044 * PSA + 0.426 * ROT - 0.443 * NATOM + 0.488 * NRING

PC7 = 0.024 * MW - 0.663 * LogP + 0.713 * LogD + 0.087 * Hdonors - 0.150 * Hacceptors

+ 0.102 * PSA - 0.055 * ROT + 0.082 * NATOM - 0.041 * NRING

PC8 = 0.223 * MW - 0.158 * LogP - 0.023 * LogD - 0.179 * Hdonors - 0.508 * Hacceptors

+ 0.747102 * PSA - 0.032 * ROT - 0.275 * NATOM + 0.017 * NRING

PC9 = -0.703 * MW + 0.087 * LogP + 0.008 * LogD - 0.176 * Hdonors + 0.040 * Hacceptors

+ 0.381 * PSA - 0.065 * ROT + 0.562 * NATOM - 0.015 * NRING

e) Interpret the PCs via eigen values. (5 marks)

As we know that

$$Porportion\ of\ total\ variance\ accounted\ by\ PC_i = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

Thus for

$$Porportion\ of\ PC_1 = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{4.73045}{9} = 0.52561$$

$$Porportion\ of\ PC_2 = \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{2.60678}{9} = 0.28964$$

$$Porportion\ of\ PC_3 = \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.84553}{9} = 0.09395$$

$$Porportion\ of\ PC_4 = \frac{\lambda_4}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.26538}{9} = 0.02949$$

$$Porportion\ of\ PC_5 = \frac{\lambda_5}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.23842}{9} = 0.02649$$

Name: Millie Woo
[Student No: s3806940]

$$Porportion\ of\ PC_6 = \frac{\lambda_6}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.10912}{9} = 0.01212$$

$$Porportion\ of\ PC_7 = \frac{\lambda_7}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.09341}{9} = 0.01038$$

$$Porportion\ of\ PC_8 = \frac{\lambda_8}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.06553}{9} = 0.00728$$

$$Porportion\ of\ PC_9 = \frac{\lambda_9}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.04538}{9} = 0.005043$$

Which is basically the answer in part (b).

However, we could also interpret the first 5 significant PCs with eigenvector, in descending order, PSA, Hacceptors, MW, HDonors, PSA and NRING receive positive weight for PC1 (coefficient ranges from 0.429 to 0.209), LogD and LogP receive almost equal amount of negative weight for PC1 (coefficient of 0.184 and 0.153 respectively). The larger the coefficient, the variable would contribute more to PC1.

For PC2, it is contributed by LogP, LogD, NRING, NATOM, MW and PSA (positive weight) in descending order, and Hdonors and PSA (negative weight) in descending order.

For PC3, it is mainly contributed by ROT (positive weight) and NRING (negative weight).

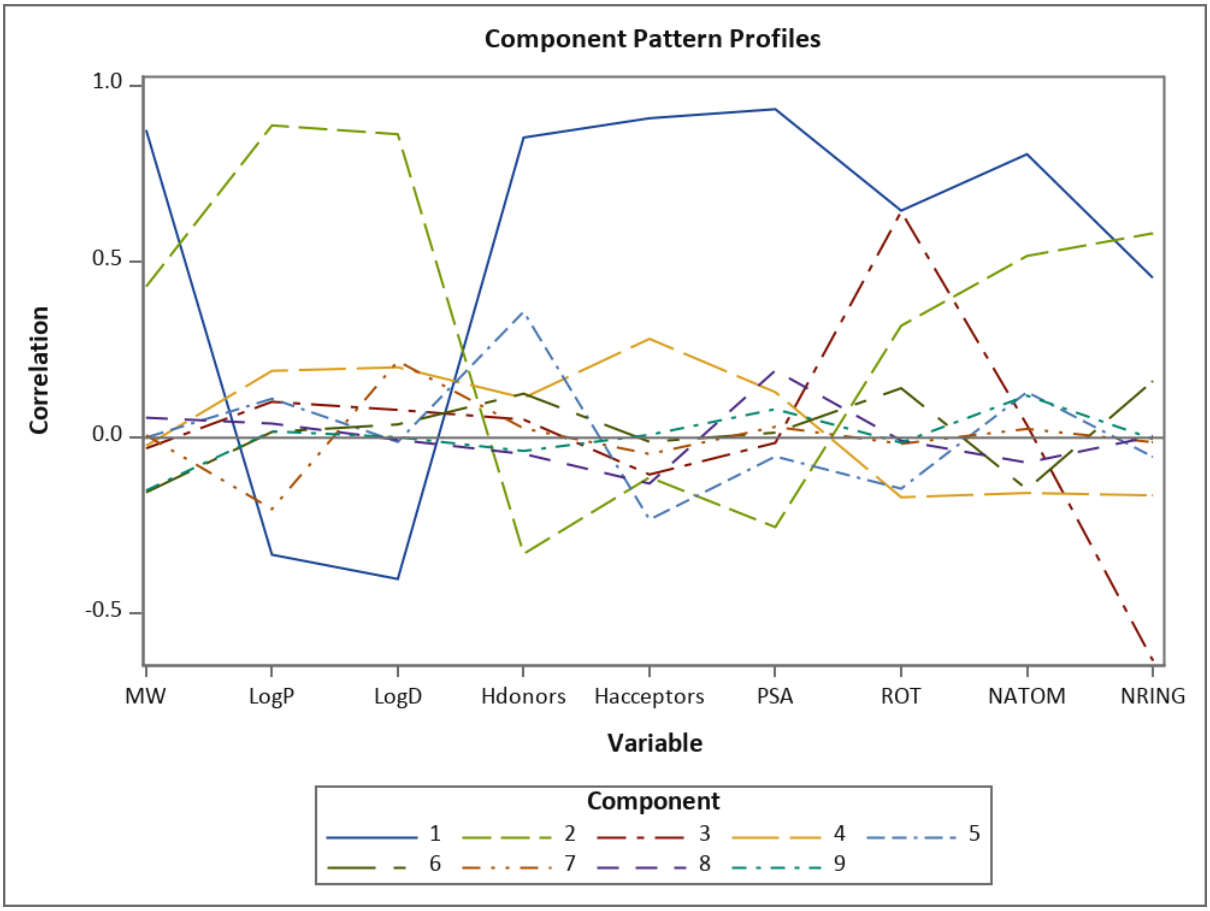
For PC4, it is contributed by Hacceptors, LogD, LogP, PSA and Hdonors (positive weight) in descending order, and ROT, NRING and NATOM (negative weight) in descending order.

For PC5, it is mainly contributed by Hdonors (positive weight) and Hacceptors (negative weight).

f) Interpret the PCs using your component pattern profiles from SAS. (4 marks)

The component profile plot shows the correlations between the original variables and each PC.

Name: Millie Woo
[Student No: s3806940]



For our example with 9 PCs, it shows that:

The first PC (solid blue line) is strongly positively correlated with MW, Hdonors, Hacceptors, PSA, ROT, NATOM and NRING, and it is moderately negatively correlated with LogP and LogD.

The second PC (dashed green line) is strongly positively correlated with LogP and LogD, and moderately positively correlated with ROT, NATOM and NRING. It only has small negative correlation with Hdonors, Hacceptors and PSA.

The third PC (dashed red line) is moderately positively correlated with ROT and moderately negatively correlated with NRING.

The fourth PC (dashed yellow line) is slightly positively correlated with all original variables, except ROT, NATOM and NRING, which is slightly negatively correlated.

The fifth PC (dashed blue line) is slightly positively correlated LogP, Hdonors and NATOM.

PC6 – PC9 only have small correlations with the original variables.

We can also use the below SAS Code to see the actual correlations in table format.

SAS Code:

```
proc corr data=PCout noprob nosimple;
     var MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
     with Prin1-Prin9;
run;
```

Output:

Name: Millie Woo
[Student No: s3806940]

Pearson Correlation Coefficients, N = 1279									
	MW	LogP	LogD	Hdonors	Hacceptors	PSA	ROT	NATOM	NRING
Prin1	0.87432	-0.33218	-0.40117	0.85293	0.90781	0.93350	0.64519	0.80559	0.45439
Prin2	0.42937	0.88737	0.86235	-0.32971	-0.11244	-0.25375	0.31827	0.51660	0.58090
Prin3	-0.03001	0.10229	0.07917	0.05174	-0.10434	-0.01435	0.64236	0.03596	-0.63262
Prin4	-0.02269	0.19014	0.20028	0.11335	0.28115	0.12970	-0.16894	-0.15676	-0.16361
Prin5	0.00059	0.11106	-0.01164	0.35826	-0.23257	-0.05311	-0.14453	0.12991	-0.05403
Prin6	-0.15511	0.01668	0.03797	0.12536	-0.01134	0.01461	0.14081	-0.14647	0.16125
Prin7	0.00726	-0.20257	0.21779	0.02670	-0.04611	0.03119	-0.01692	0.02520	-0.01251
Prin8	0.05701	0.04040	-0.00582	-0.04589	-0.12998	0.19118	-0.00808	-0.07052	0.00424
Prin9	-0.14973	0.01845	0.00176	-0.03739	0.00843	0.08124	-0.01380	0.11980	-0.00319

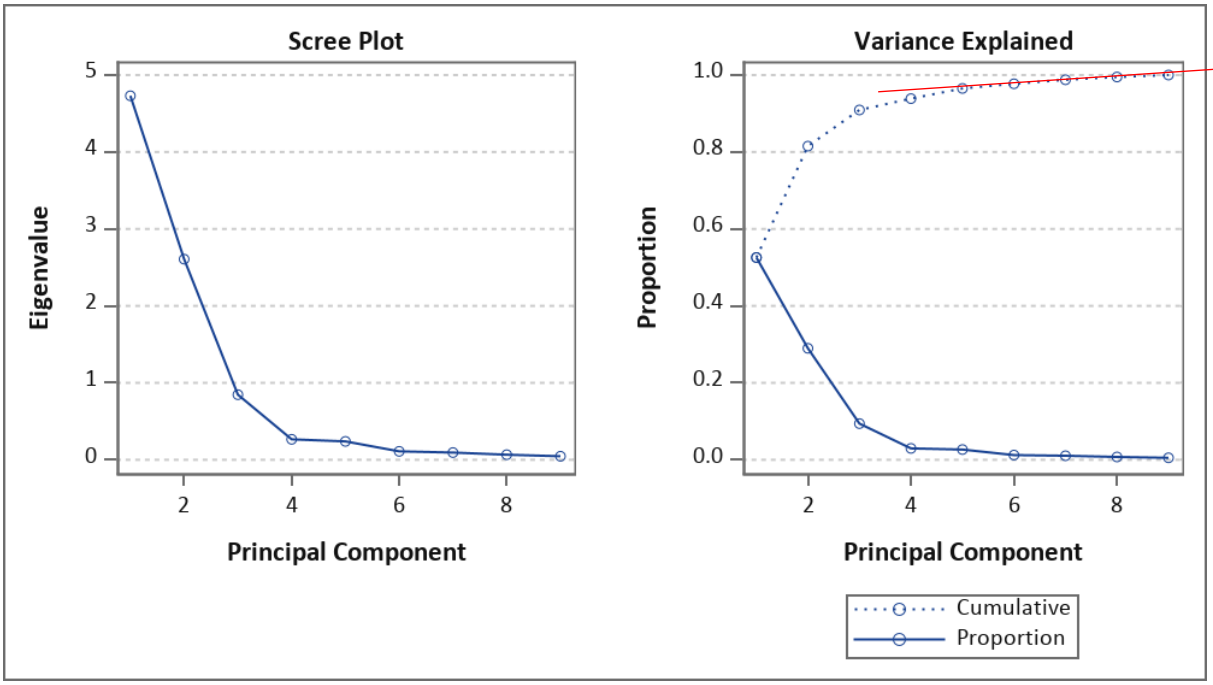
g) Can the data be effectively summarised in fewer than 9 dimensions? Justify your answer using BOTH relevant plots and eigenvalues.

We usually look for an elbow in the scree plot to determine the appropriate number of components for dimension reduction. However, if “detect the elbow” is too imprecise for you, a more precise algorithm is to start at the right-hand side of the scree plot and look at the points that lie (approximately) on a straight line. The leftmost point along the trend line indicates the number of components to retain. (In geology, “scree” is rubble at the base of a cliff; the markers along the linear trend represent the rubble that can be discarded.) For the example data, the markers for components 4–7 are linear, so components 1–4 would be kept. This rule (and the scree plot) was proposed by Cattell (1966) and revised by Cattell and Jaspers (1967). (The above paragraph is quoted from Reference [R1]).

In our scree plot, as shown below, PC5 – PC9 lie on the same straight line, so I would suggest to keep PC1- PC5. From the cumulative proportion of eigenvalues (as shown below the scree plot), 96.5% of the variance can be explained by the first 5 components, and we can instantly reduce the dimensions from 9 to 5.

MATH1309 Multivariate Analysis
Assignment 2

Name: Millie Woo
[Student No: s3806940]



Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
λ_1	4.73045355	2.12367349	0.5256	0.5256
λ_2	2.60678006	1.76125011	0.2896	0.8152
λ_3	0.84552994	0.58014848	0.0939	0.9092
λ_4	0.26538146	0.02695961	0.0295	0.9387
λ_5	0.23842185	0.12930442	0.0265	0.9652
λ_6	0.10911743	0.01571009	0.0121	0.9773
λ_7	0.09340733	0.02788250	0.0104	0.9877
λ_8	0.06552483	0.02014128	0.0073	0.9950
λ_9	0.04538355		0.0050	1.0000

Name: Millie Woo
[Student No: s3806940]

Question 2: PCA with reduced k <p for plots. Rerun the PCA for the violators and the non-violators separately as delineated by score_9 log D using your k from Q 1.

a) Recreate the 5 plots related to PROC PCA for your given k.

SAS Code:

Step 1:

Sort the dataset by score9_logD_group_1.

SAS Code:

```
proc sort data=work.drug out=work.drug_sorted;
   by score9_logD_group_1;
run;
```

Step 2: Use proc princomp to perform component analysis and generate the scree plot, profile plot, component pattern plots for **n=5**

```
proc princomp data=work.drug_sorted n=5
     STD
     out = PCout2
     plots(ncomp=5)=(scree profile pattern score);
     var MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
     by score9_logD_group_1;
     ods output Eigenvectors=EV;
run;
```

Step 3: Use proc factor to generate the loading plots, specify p=96 to generate up to 5 PCs (as cumulative proportion from PC1 to PC5 is 96.52%, for PC1 to PC6 is 97.73%, specify p=96,the plots would not be generated for PC more than 97%)

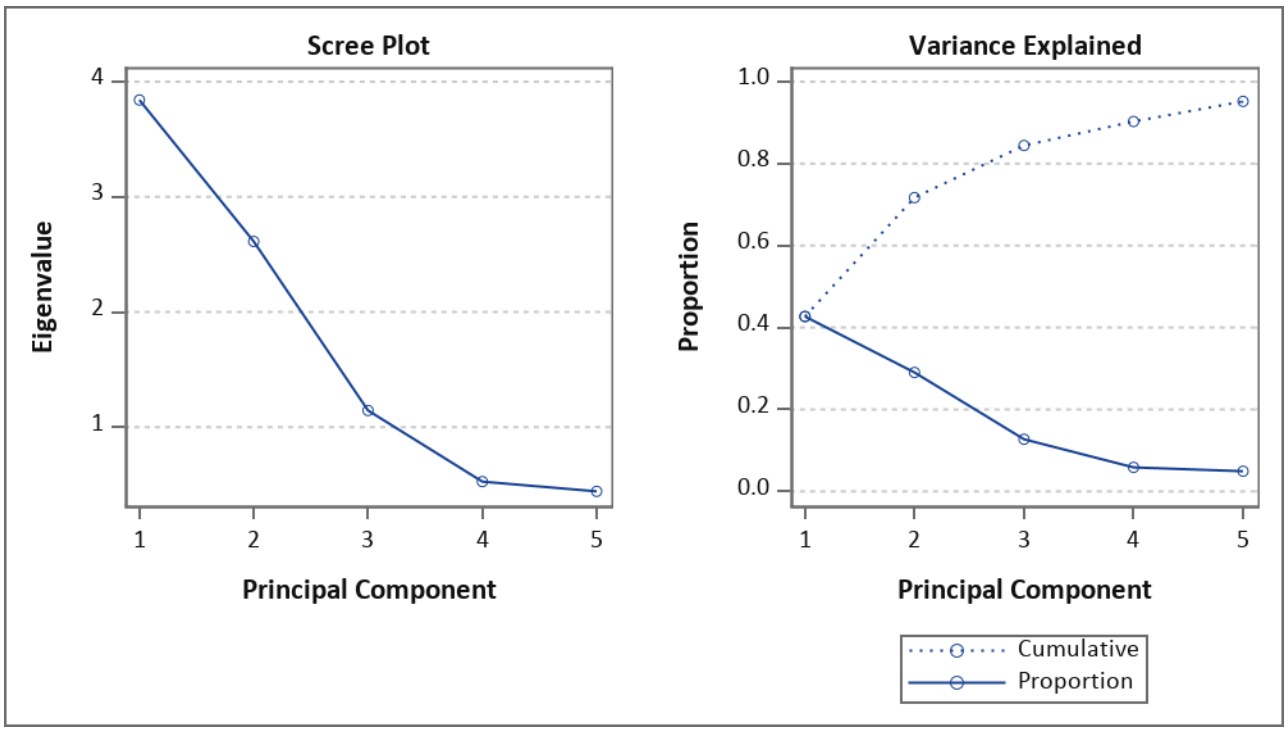
```
proc factor data=work.drug_sorted p=96
     method=principal
     plots(nplots=5)=(initloadings(vector));
     var MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
     by score9_logD_group_1;
run;
```

All plots are located at Appendix [A2]

Name: Millie Woo
[Student No: s3806940]

b) Using the plots based on your reduced dimensionality k from part a) and outputs interpret the first to k PC's via eigenvalues.

For score9_logD_group_1=1, i.e. non-violator, the scree plot and the cumulative proportion of eigenvalues are shown as below:



Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
λ_1	3.84083397	1.22864064	0.4268	0.4268
λ_2	2.61219333	1.46762578	0.2902	0.7170
λ_3	1.14456755	0.61768852	0.1272	0.8442
λ_4	0.52687903	0.08457332	0.0585	0.9027
λ_5	0.44230571		0.0491	0.9519

Each PC can explain the variation of non-violator in the dataset as:

$$Porportion\ of\ PC_1 = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{3.8408}{9} = 0.4268$$

$$Porportion\ of\ PC_2 = \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{2.6122}{9} = 0.2902$$

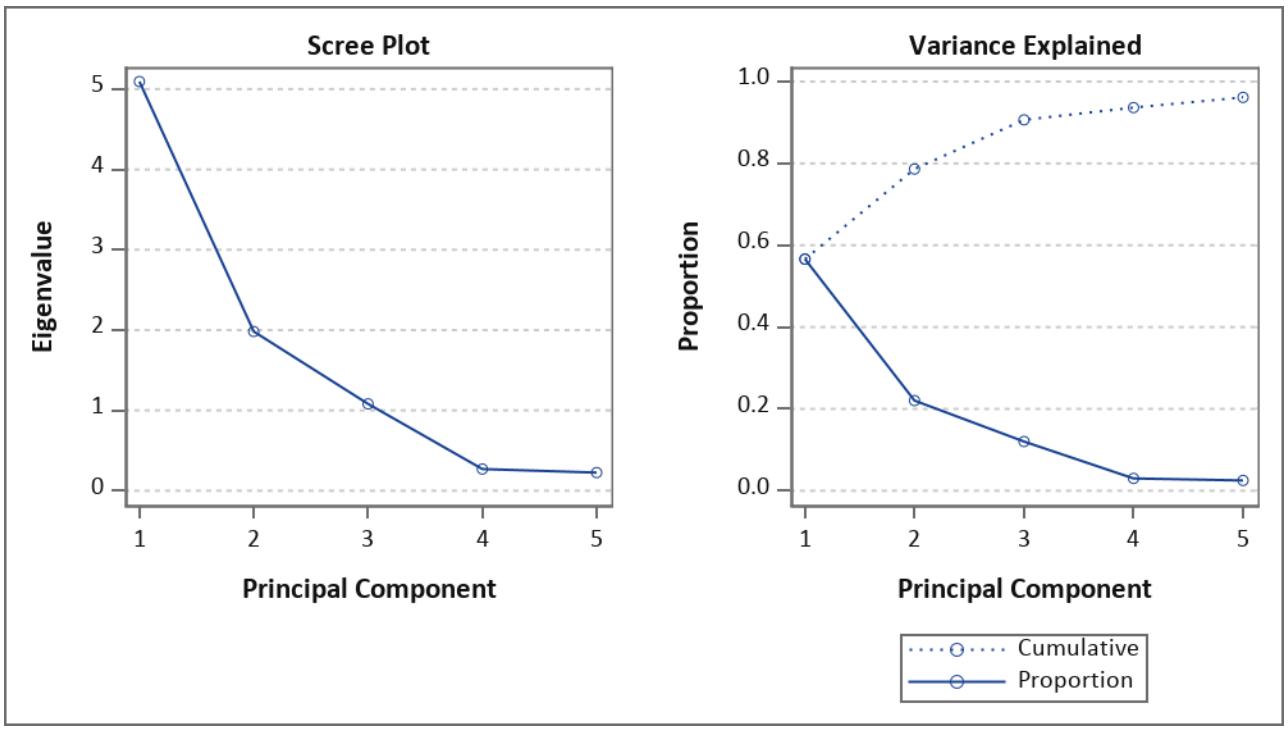
Name: Millie Woo
[Student No: s3806940]

$$Porportion\ of\ PC_3 = \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{1.1446}{9} = 0.1272$$

$$Porportion\ of\ PC_4 = \frac{\lambda_4}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.5269}{9} = 0.0585$$

$$Porportion\ of\ PC_5 = \frac{\lambda_5}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.4423}{9} = 0.0491$$

For score9_logD_group_1=2, i.e. violator, the scree plot and the cumulative proportion of eigenvalues are shown as below:



Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
λ_1	5.09425408	3.11217126	0.5660	0.5660
λ_2	1.98208283	0.90104753	0.2202	0.7863
λ_3	1.08103530	0.81137333	0.1201	0.9064
λ_4	0.26966197	0.04448633	0.0300	0.9363
λ_5	0.22517564		0.0250	0.9614

Name: Millie Woo
[Student No: s3806940]

Each PC can explain the variation of violator in the dataset as:

$$Porportion\ of\ PC_1 = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{5.0943}{9} = 0.5660$$

$$Porportion\ of\ PC_2 = \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{1.9821}{9} = 0.2202$$

$$Porportion\ of\ PC_3 = \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{1.081}{9} = 0.1201$$

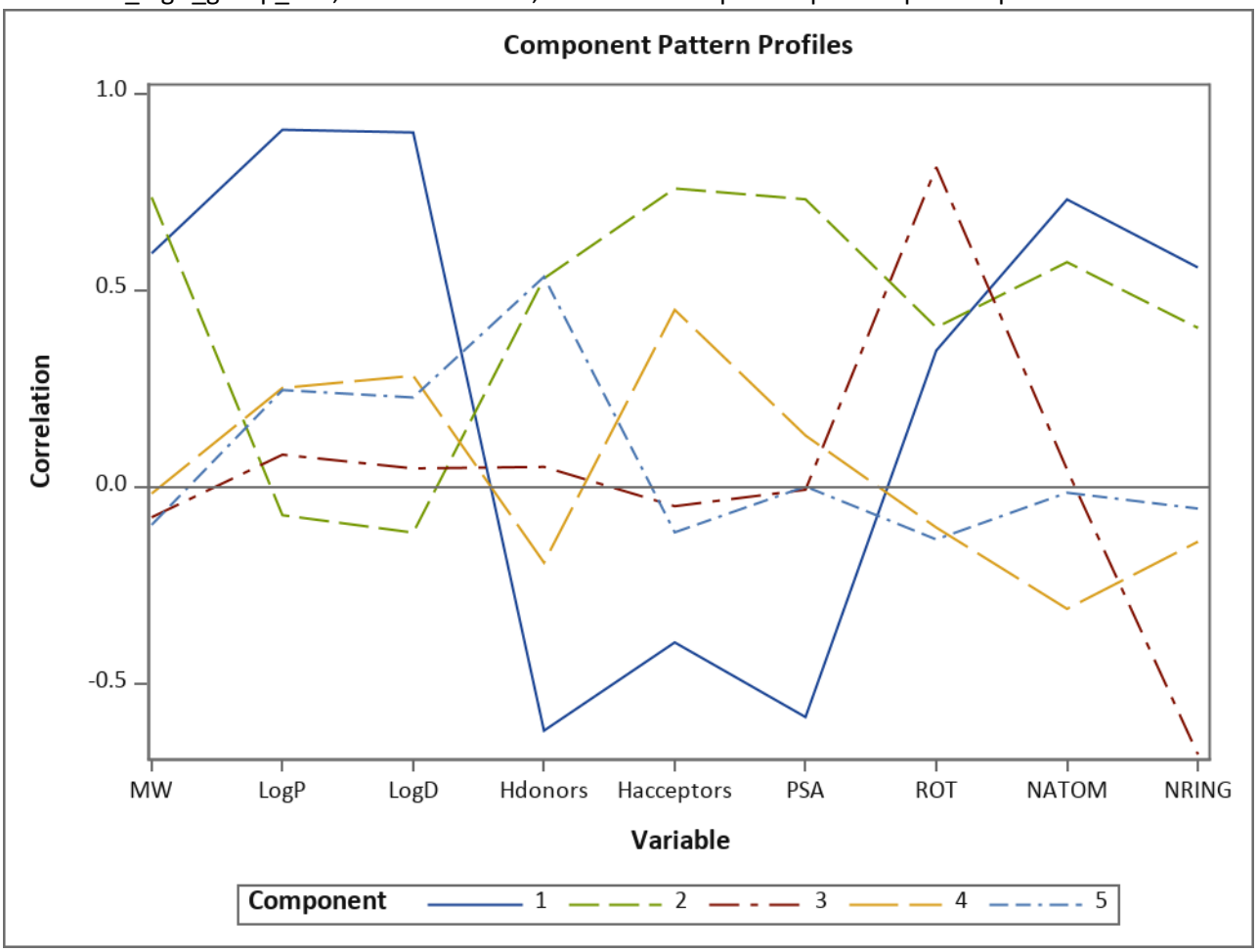
$$Porportion\ of\ PC_4 = \frac{\lambda_4}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.2697}{9} = 0.0300$$

$$Porportion\ of\ PC_5 = \frac{\lambda_5}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7+\lambda_8+\lambda_9} = \frac{0.2252}{9} = 0.0250$$

Name: Millie Woo
[Student No: s3806940]

c) Using the plots based on your reduced dimensionality k from part a) and outputs interpret the first to k PC's via the outputs (you choose the optimal k).

For score9_logD_group_1=1, i.e. non-violator, the below component pattern profiles plot shows:



PC1 (solid blue line) is strongly positively correlated with LogP, LogD and NATOM and moderately positively correlated with MW, ROT and NRING , and it's moderately negatively correlated with Hacceptors, PSA and Hdonors.

PC2(dashed green line) is strongly positively correlated with MW, Hacceptors and PSA, and moderately positively correlated with Hdonors, ROT, NATOM and NRING. It only has small negative correlation with LogP and LogD.

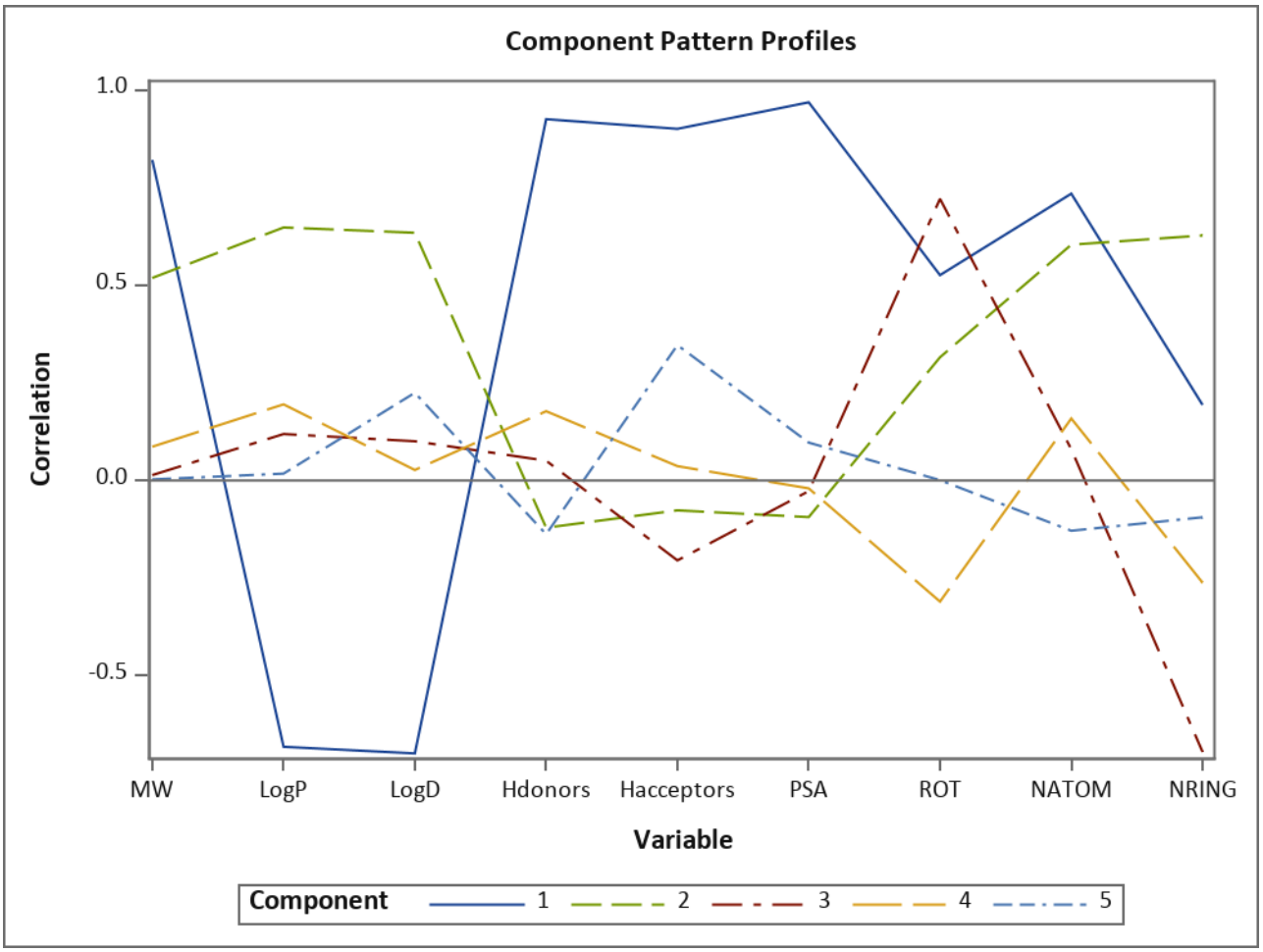
PC3 (dashed red line) is strongly positively correlated with ROT and strongly negatively correlated with NRING.

PC4 (dashed yellow line) is moderately positively correlated with LogD, LogP and Hacceptors, and moderately negatively correlated with NATOM.

PC5 (dashed blue line) is moderately positively correlated Hdonors.

Name: Millie Woo
[Student No: s3806940]

For score9_logD_group_1=2, i.e. violator, the below component pattern profiles plot shows:



PC1 (solid blue line) is strongly positively correlated with Hdonors, Hacceptors, PSA and MW and moderately positively correlated with NATOM and ROT, and it's strongly negatively correlated with LogP and LogD.

PC2(dashed green line) is moderately positively correlated with MW, LogP, LogD, NATOM NRING and ROT.

PC3 (dashed red line) is strongly positively correlated with ROT and strongly negatively correlated with NRING.

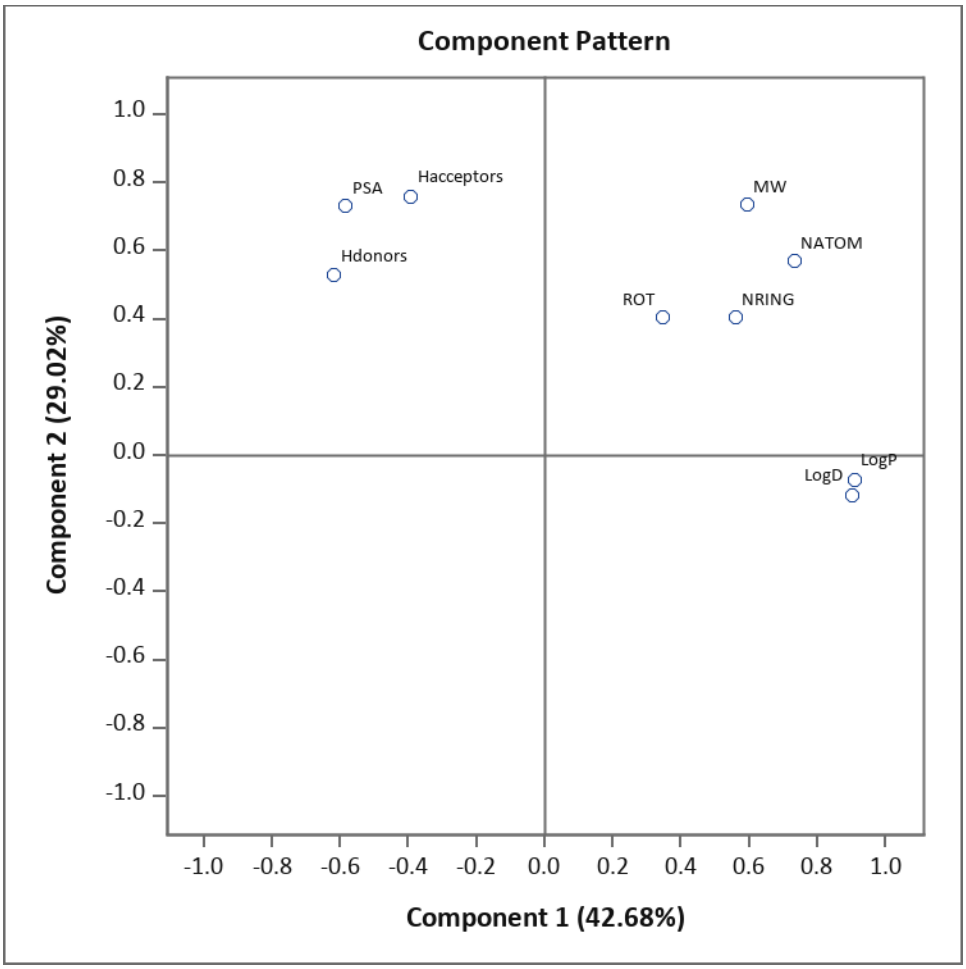
PC4 (dashed yellow line) is slightly positively correlated with all original variables, except ROT, NRING and PSA, which is slightly negatively correlated.

PC5 (dashed blue line) is slightly positively correlated Hacceptors and LogD.

Name: Millie Woo
[Student No: s3806940]

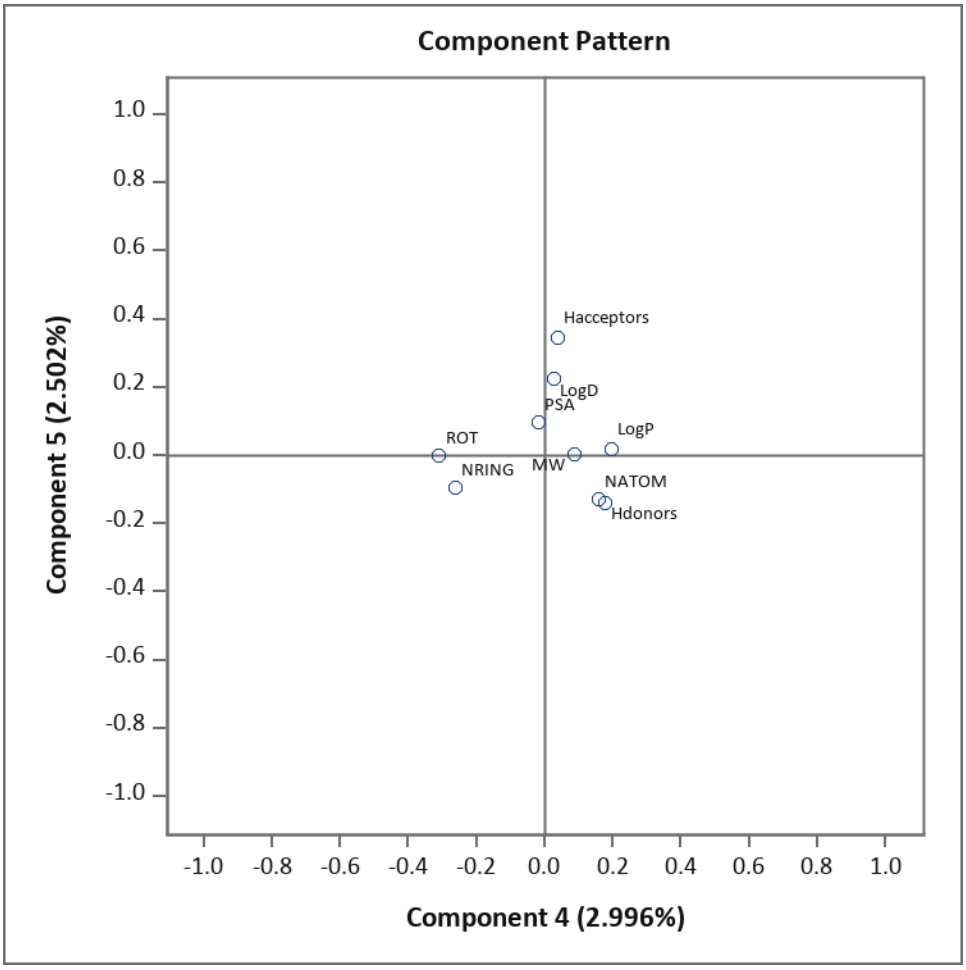
The pattern plot shows the correlations between the original variables and pairwise combination of PCs. As we chose k =5. There would be 5 * (5 – 1) / 2 = 10 plots each for both violator and non-violators.

For example, for non-violator, if we show the correlations between PC1 and PC2 in the below graph, we could clearly notice that NRING, NATOM, MW, ROT are both positively correlated with PC1 and PC2, Hacceptors, PSA, Hdonors are positively correlated with PC2, but negatively correlated with PC1. LogD and LopP is slightly negatively correlated with PC2 and positively correlated with PC1.



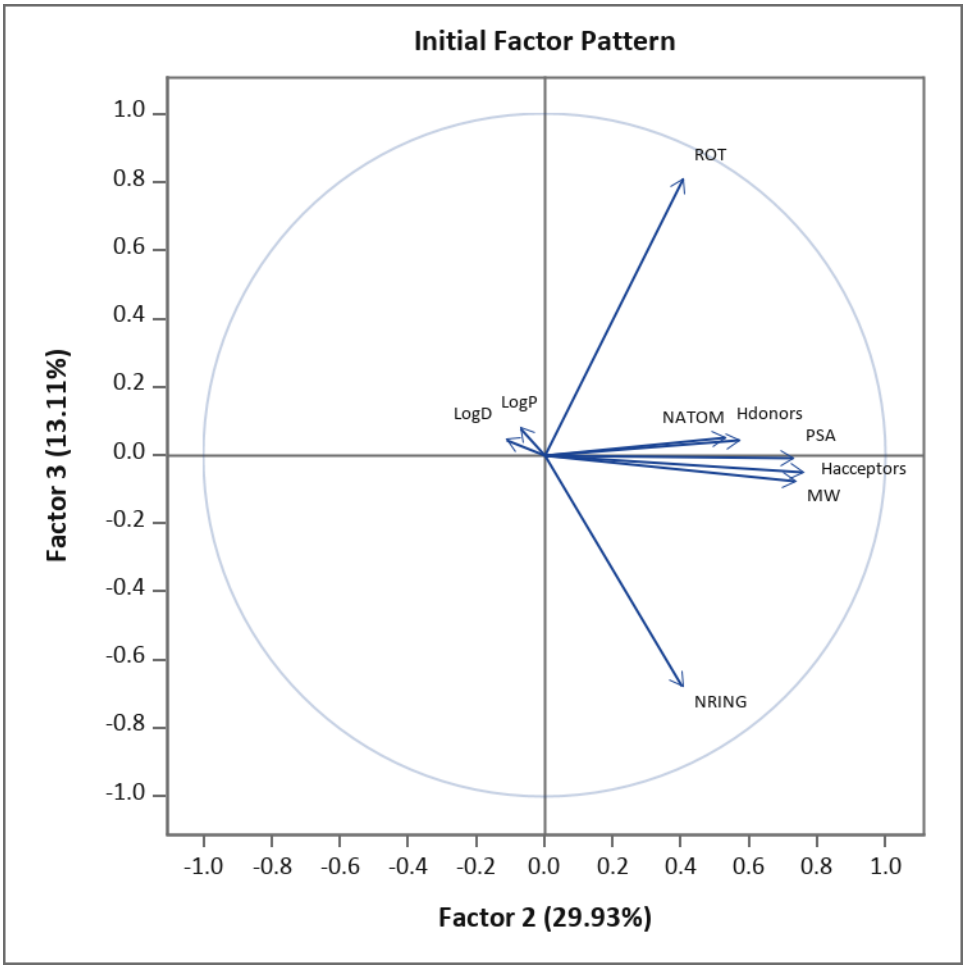
Name: Millie Woo
[Student No: s3806940]

The correlation becomes less spread out when we compare lateral PCs, for example, for violator, in the (PC4, PC5) pair-wise plot, we could see all the points are more centred to the origin as the correlations of original variables decreases in both PC4 and PC5. The below plot only shows Hacceptors has slightly higher correlations with PC5.



The loading plot shows the relationship between the eigenvectors and PCs in pairwise combinations, for example, for non-violator, the below plot shows that NATOM, Hdonors, PSA, Hacceptors and MW have much higher coefficients in PC2 (all positive) than PC3, ROT (positive for both PC2 and PC3) and NRING (positive for PC2 and negative for PC3) have higher coefficient in PC3 than PC2, while LogD and LogP (positive in PC3 and negative in PC2) have both low coefficient in both PC2 and PC3.

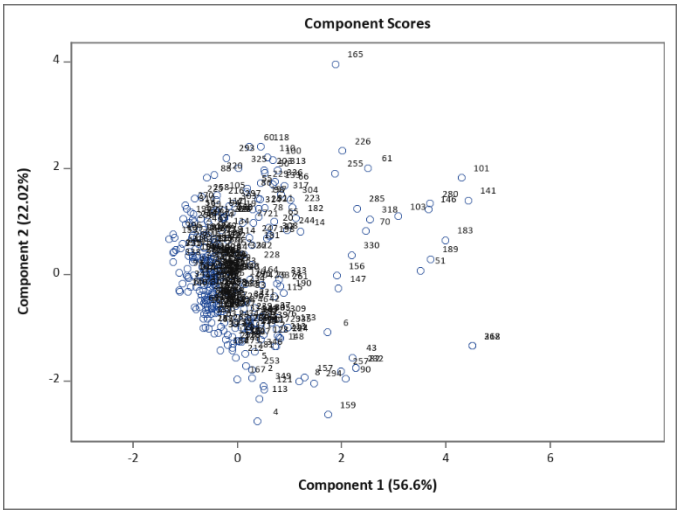
Name: Millie Woo
[Student No: s3806940]

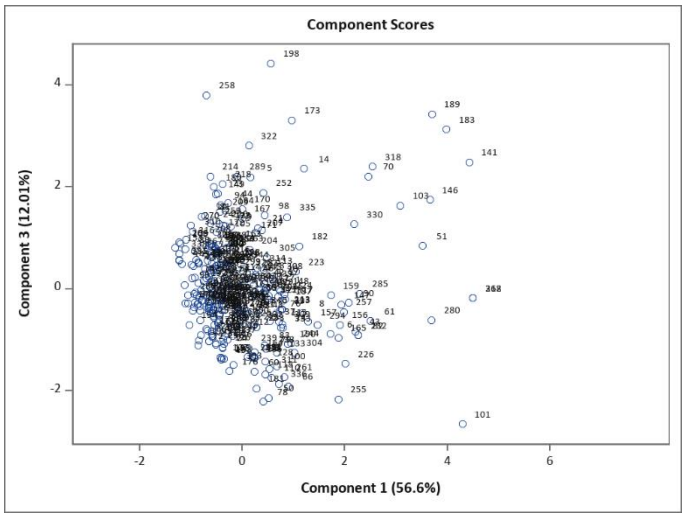


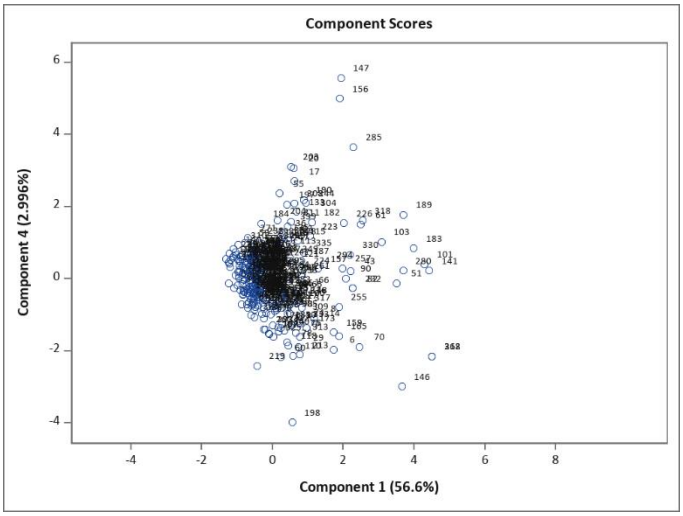
Name: Millie Woo
[Student No: s3806940]

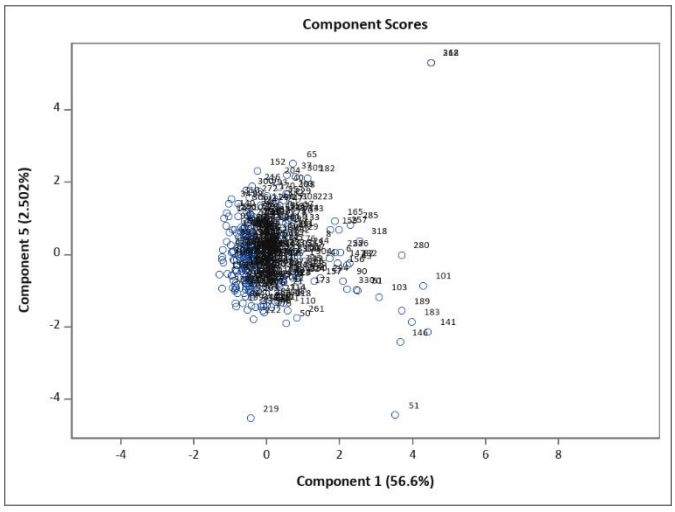
d) Which of the k PCs are skewed? Use your plots to answer this.

The score plots indicate the projection of data onto the span of PCs in pair-wise comparison. We could spot that when PC1 compare to other PCs (below 4 graphs), data are spread positively in horizontal axis for violators



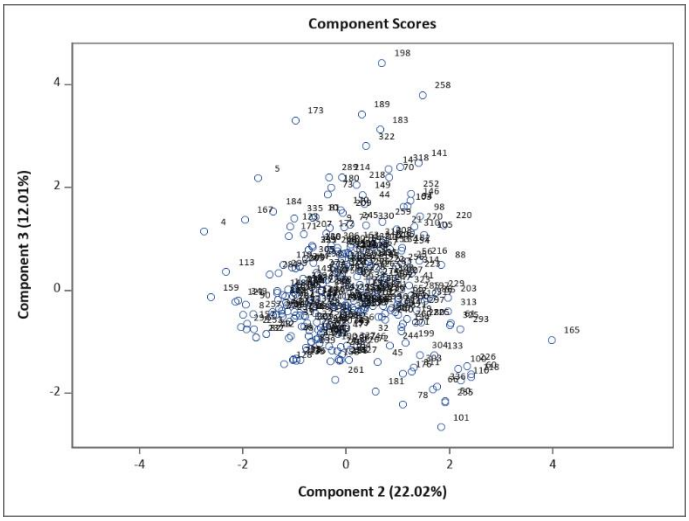


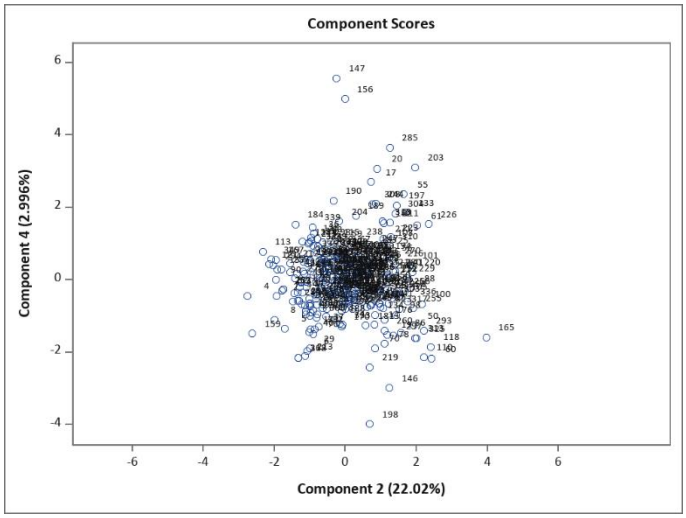


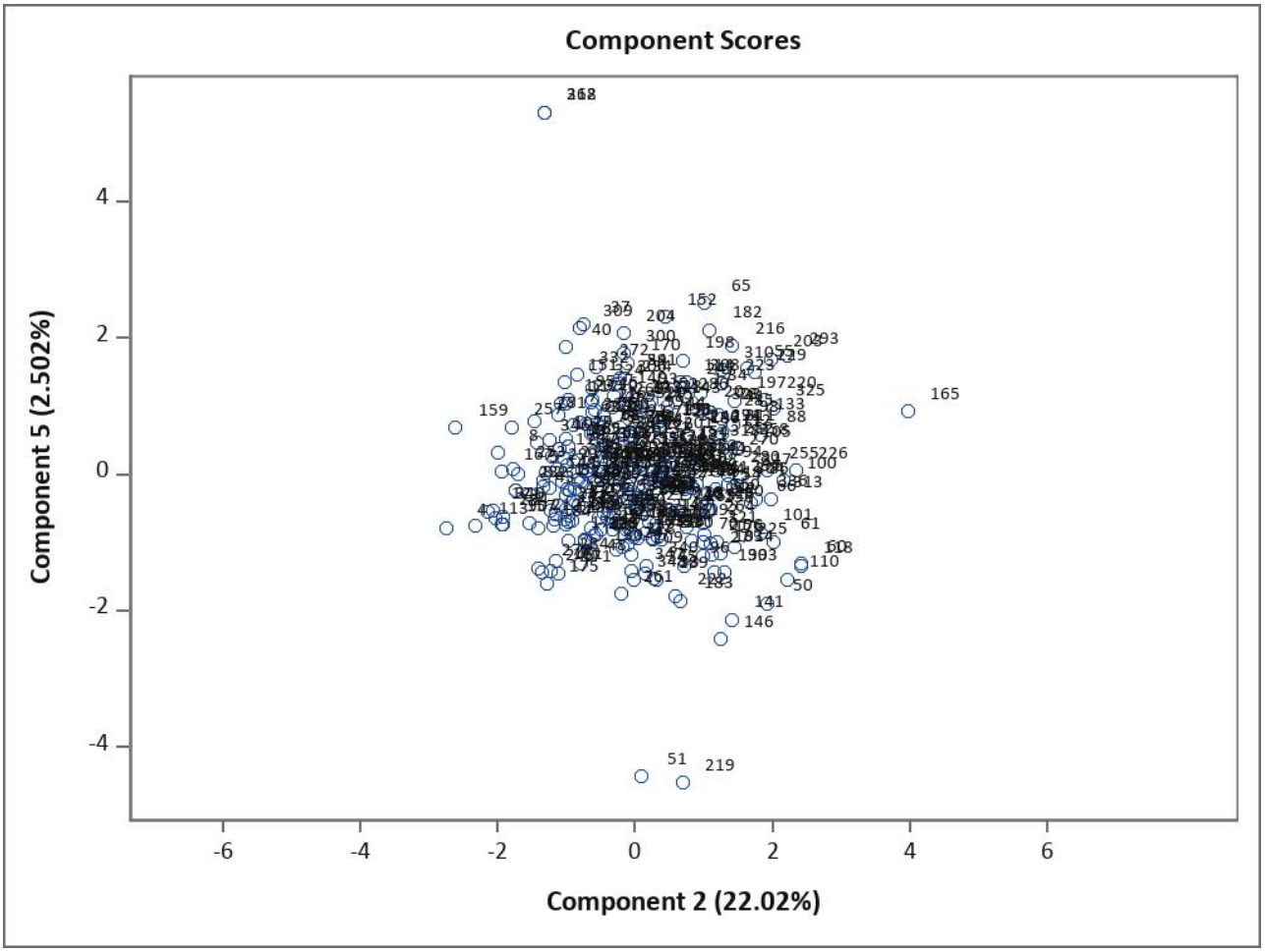


Name: Millie Woo
[Student No: s3806940]

When PC2 compare with other PCs (PC3, PC4 and PC5,below 3 plots), data are centrally located within ±2 in the horizontal axis. PC3 and PC4 do have data spreading out but they are in both directions vertically. There are just few outliners for PC5, also in both directions.

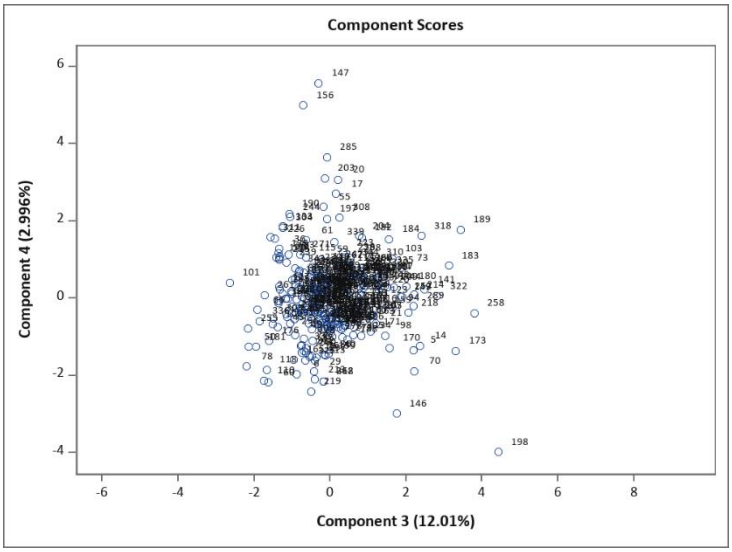


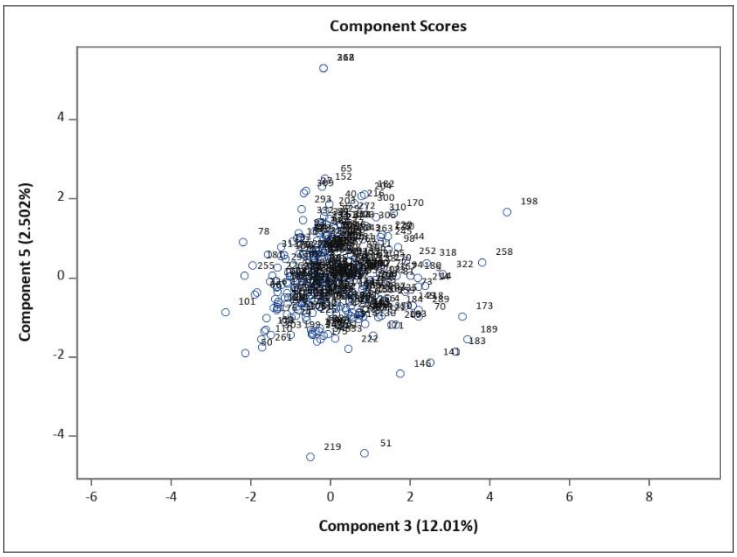




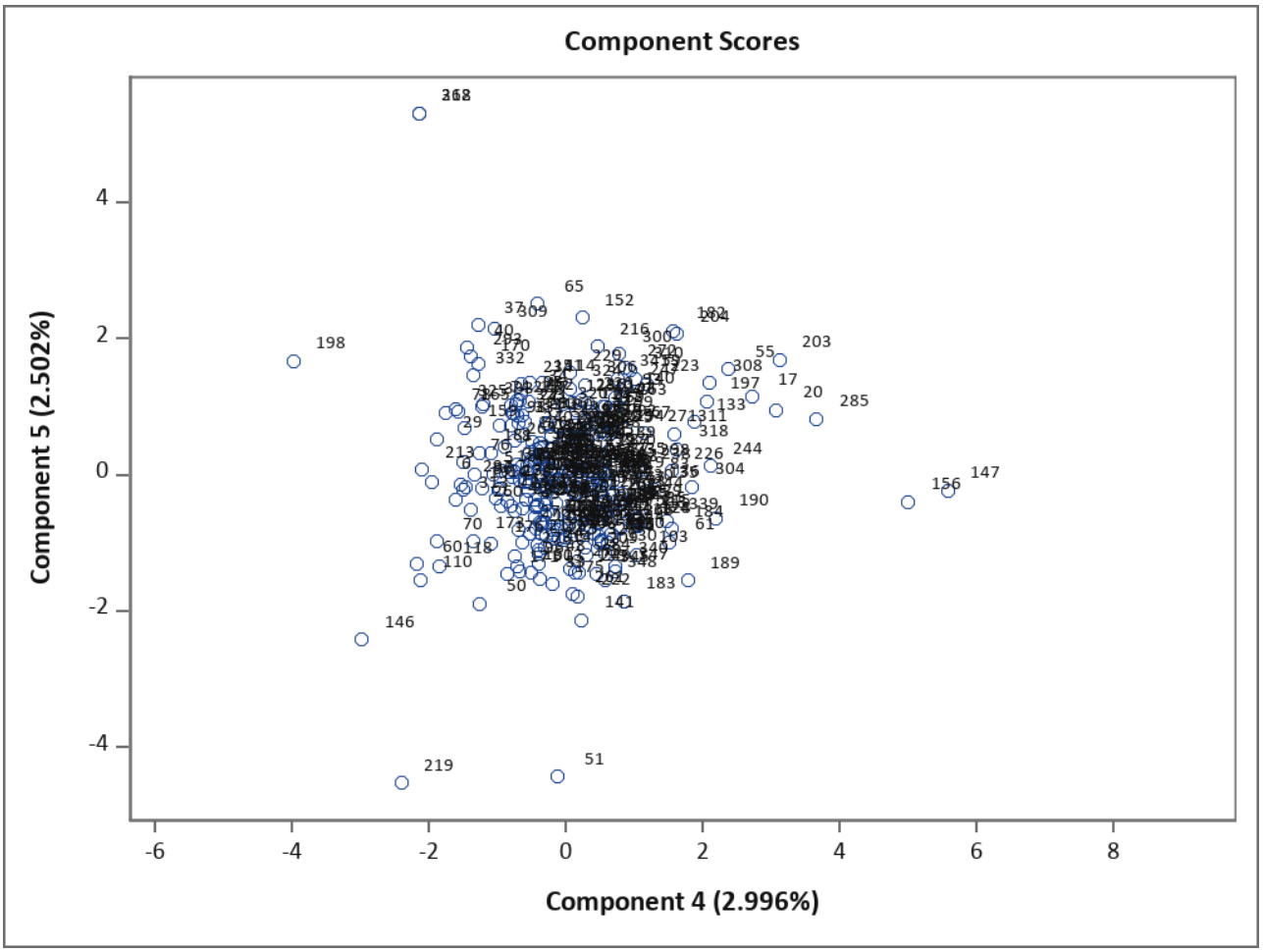
Name: Millie Woo
[Student No: s3806940]

This also applies to PC3 when it compares with PC4 and PC5, just that there are more outliners in both ends, positive side would be slightly more outliners. (horizontal axis)





For PC4 and PC5, we could see, most of the data are still centrally located, just slightly more outliners toward negative side for PC4, and slightly more outliners toward positive side for PC5.



This is less obvious for non-violators. Thus, we could conclude that PC1 is highly positive skewed for violators, while other PCs are not skewed.

Name: Millie Woo
[Student No: s3806940]

Question 3: Discriminant on 2 groups of Molecules

1. Prepare the dataset for input for a Discriminant analysis via SAS.

Step 1: Read in the given excel file into SAS as a table work.readin

SAS Code:

```
PROC IMPORT OUT= work.readin DATAFILE= "/home/u41107333/Multivariate/0.
Project/MATH1309  Drug Bank DATA  set  for Assignment 2.xlsx"
            DBMS=xlsx REPLACE;
     SHEET="final_Drugbank dataset ";
     GETNAMES=YES;
RUN;
```

Step 2: As we have already checked missing values in question 1, we don't need to check that again here, however, for question 3, beside the 9 molecular variables, we also need oral_status and score9_logD_group_1 for the analysis, so we create the table work.drug which contains all these variables to work with.

SAS Code:

```
proc sql;
create table work.drug as
select MW, LogP, LogD, Hdonors, Hacceptors, PSA, ROT, NATOM, NRING, oral_status,
score9_logD_group_1
from WORK.readin;
quit;
```

2. Generate the means, standard deviations and the variance-covariance matrix of the data for the violators.

3. Generate the means, standard deviations and the variance-covariance matrix of the data for the non-violators

4. Produce the correlation matrix and an associated scatterplot of the inputted data for the violators.

5. Produce the correlation matrix and an associated scatterplot of the inputted data for the non-violators.

These 4 requirements have been undertaken together with the following steps:

Step 1:

Sort the dataset by score9_logD_group_1.

SAS Code:

```
proc sort data=work.drug out=work.drug_sorted;
   by score9_logD_group_1;
run;
```

Step 2:

Use proc corr cov to generate the means, standard deviations, variance-covariance matrix, correlation matrix and associated scatterplot of the inputted data for violators and non-violators.

Name: Millie Woo
[Student No: s3806940]

SAS Code:

```
proc corr cov data=work.drug_sorted PLOTS(MAXPOINTS=NONE)=MATRIX(NVAR=ALL histogram) ;
   var MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
   by score9_logD_group_1;
run;
```

Output:

1. means, standard deviations and the variance-covariance matrix of the data for the violators.

score9_logD_group_1=2

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
MW	351	519.48967	207.22440	182341	306.27080	1550	MW
LogP	351	1.84160	3.27632	646.40100	-13.20000	8.90000	LogP
LogD	351	1.38787	3.78731	487.14199	-11.63290	10.40523	LogD
Hdonors	351	3.56980	3.50348	1253	0	19.00000	Hdonors
Hacceptors	351	8.66097	4.22734	3040	0	33.00000	Hacceptors
PSA	351	146.68852	98.83273	51488	0	652.39001	PSA
ROT	351	8.20798	4.91407	2881	1.00000	33.00000	ROT
NATOM	351	68.68376	29.89199	24108	30.00000	203.00000	NATOM
NRING	351	3.72365	1.57316	1307	0	10.00000	NRING

2. means, standard deviations and the variance-covariance matrix of the data for the non-violators.

score9_logD_group_1=1

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
MW	928	262.74342	97.42170	243826	30.00610	776.87000	MW
LogP	928	1.73783	2.26848	1613	-5.50000	10.00000	LogP
LogD	928	1.62423	2.36067	1507	-6.17104	10.50772	LogD
Hdonors	928	1.49784	1.23155	1390	0	6.00000	Hdonors
Hacceptors	928	3.74677	1.85141	3477	0	12.00000	Hacceptors
PSA	928	63.86678	37.08149	59268	0	237.75000	PSA
ROT	928	3.53125	2.80758	3277	0	20.00000	ROT
NATOM	928	35.35237	14.49219	32807	2.00000	98.00000	NATOM
NRING	928	2.08728	1.42634	1937	0	10.00000	NRING

Name: Millie Woo
[Student No: s3806940]

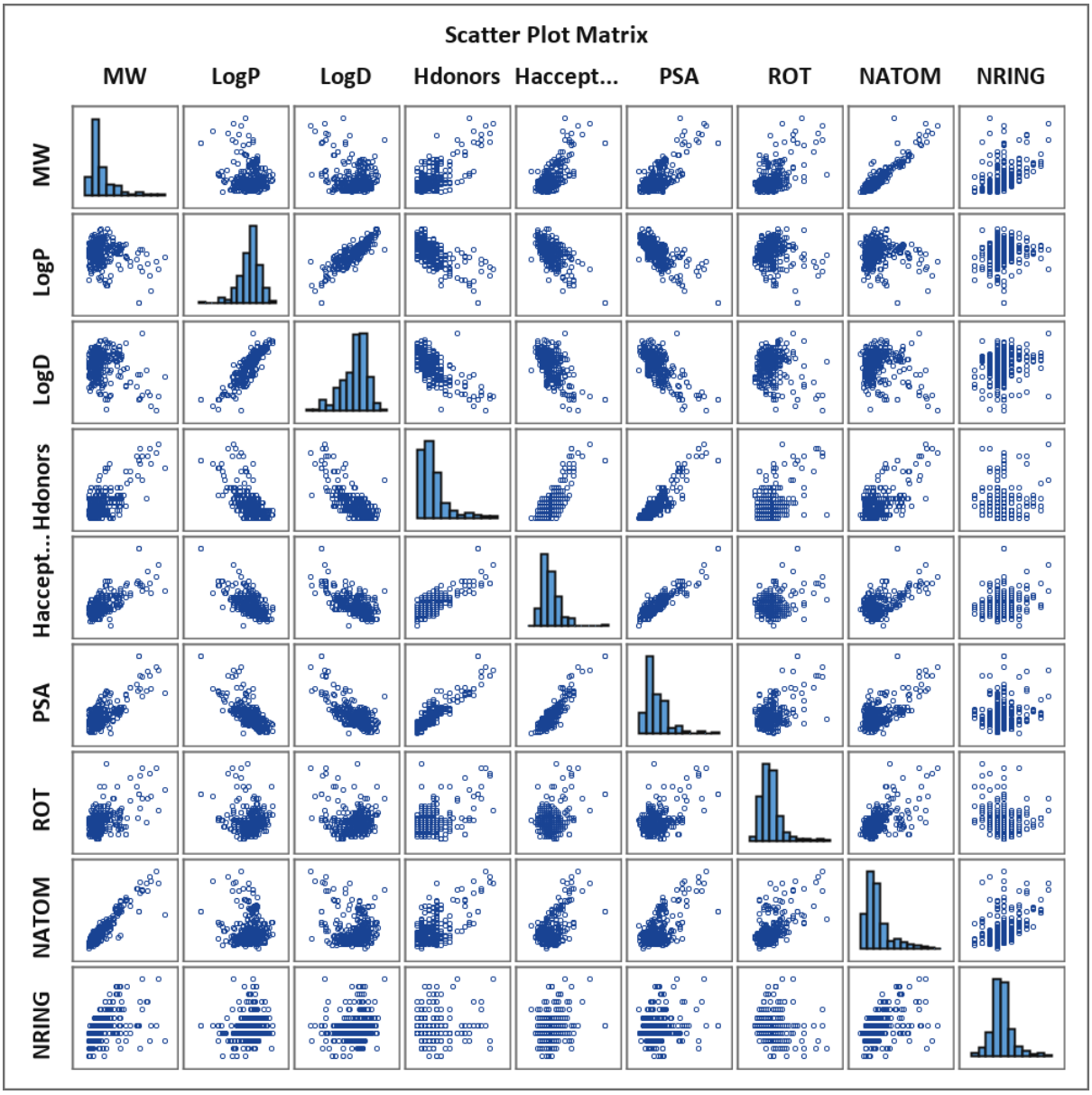
3. correlation matrix and an associated scatterplot of the inputted data for the violators

score9_logD_group_1=2

Pearson Correlation Coefficients, N = 351 Prob > |r| under H0: Rho=0									
	MW	**LogP**	**LogD**	**Hdonors**	**Hacceptors**	**PSA**	**ROT**	**NATOM**	**NRING**
MW	1.00000	-0.21970 <.0001	-0.24783 <.0001	0.68464 <.0001	0.68983 <.0001	0.74296 <.0001	0.57036 <.0001	0.92313 <.0001	0.44107 <.0001
LogP	-0.21970 <.0001	1.00000	0.87510 <.0001	-0.65662 <.0001	-0.66137 <.0001	-0.71390 <.0001	-0.11443 0.0321	-0.10164 0.0571	0.15258 0.0042
LogD	-0.24783 <.0001	0.87510 <.0001	1.00000	-0.71565 <.0001	-0.63670 <.0001	-0.71583 <.0001	-0.10893 0.0414	-0.14568 0.0063	0.16662 0.0017
Hdonors	0.68464 <.0001	-0.65662 <.0001	-0.71565 <.0001	1.00000	0.79419 <.0001	0.89089 <.0001	0.44300 <.0001	0.62250 <.0001	0.05457 0.3080
Hacceptors	0.68983 <.0001	-0.66137 <.0001	-0.63670 <.0001	0.79419 <.0001	1.00000	0.90138 <.0001	0.29897 <.0001	0.56513 <.0001	0.23033 <.0001
PSA	0.74296 <.0001	-0.71390 <.0001	-0.71583 <.0001	0.89089 <.0001	0.90138 <.0001	1.00000	0.46379 <.0001	0.62657 <.0001	0.14532 0.0064
ROT	0.57036 <.0001	-0.11443 0.0321	-0.10893 0.0414	0.44300 <.0001	0.29897 <.0001	0.46379 <.0001	1.00000	0.57872 <.0001	-0.11525 0.0309
NATOM	0.92313 <.0001	-0.10164 0.0571	-0.14568 0.0063	0.62250 <.0001	0.56513 <.0001	0.62657 <.0001	0.57872 <.0001	1.00000	0.42575 <.0001
NRING	0.44107 <.0001	0.15258 0.0042	0.16662 0.0017	0.05457 0.3080	0.23033 <.0001	0.14532 0.0064	-0.11525 0.0309	0.42575 <.0001	1.00000

MATH1309 Multivariate Analysis
Assignment 2

Name: Millie Woo
[Student No: s3806940]



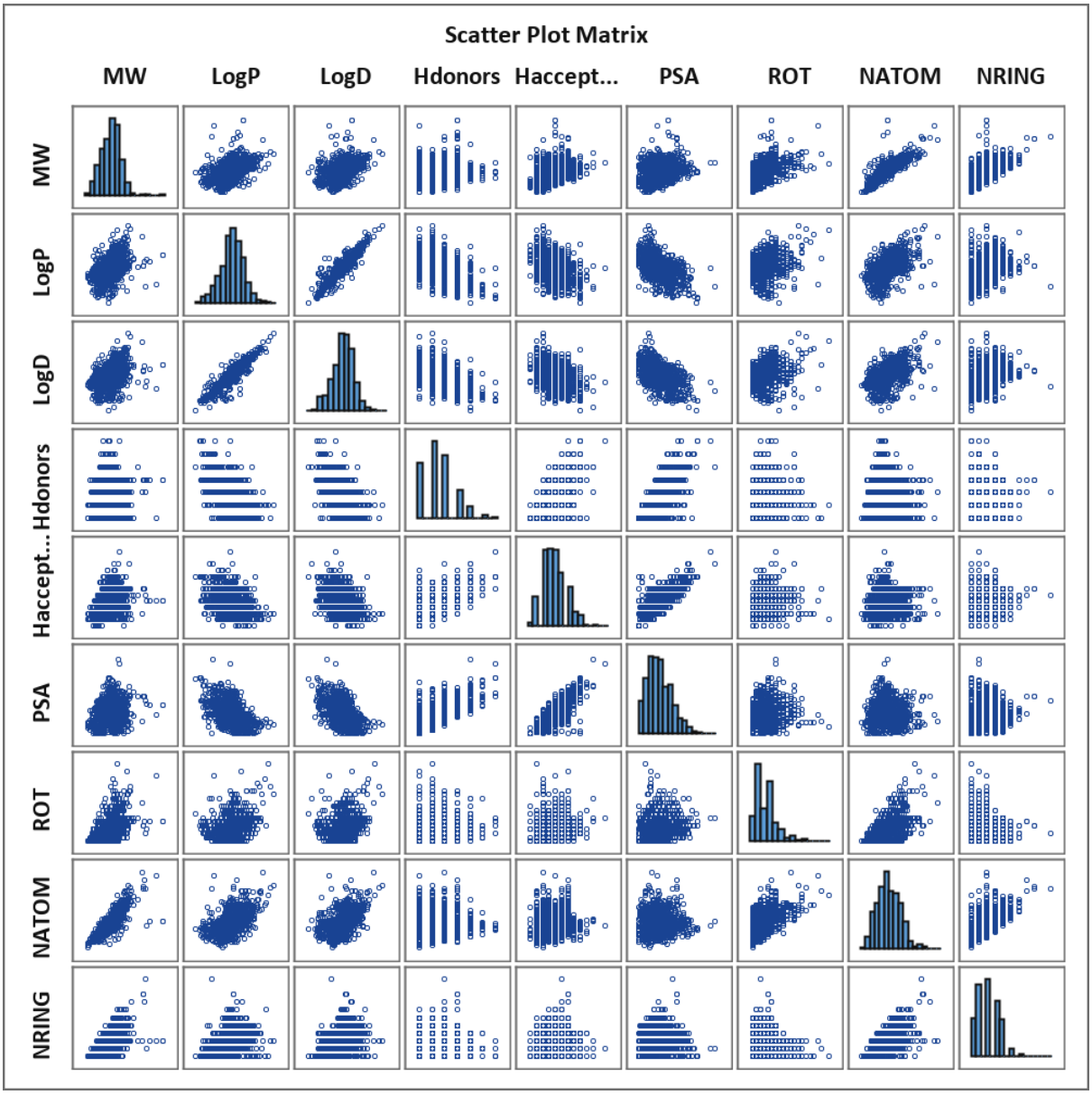
Name: Millie Woo
[Student No: s3806940]

4. correlation matrix and an associated scatterplot of the inputted data for the non-violators

score9_logD_group_1=1

Pearson Correlation Coefficients, N = 928 Prob > |r| under H0: Rho=0									
	MW	**LogP**	**LogD**	**Hdonors**	**Hacceptors**	**PSA**	**ROT**	**NATOM**	**NRING**
MW	1.00000	0.44797	0.41502	-0.02382	0.30944	0.17993	0.43289	0.83011	0.64869
MW		<.0001	<.0001	0.4686	<.0001	<.0001	<.0001	<.0001	<.0001
LogP	0.44797	1.00000	0.92620	-0.51689	-0.33610	-0.53793	0.29580	0.54594	0.37750
LogP	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
LogD	0.41502	0.92620	1.00000	-0.55269	-0.35361	-0.55636	0.24778	0.50801	0.37494
LogD	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Hdonors	-0.02382	-0.51689	-0.55269	1.00000	0.51333	0.69505	-0.00357	-0.10123	-0.16417
Hdonors	0.4686	<.0001	<.0001		<.0001	<.0001	0.9136	0.0020	<.0001
Hacceptors	0.30944	-0.33610	-0.35361	0.51333	1.00000	0.78270	0.10913	0.02122	0.07374
Hacceptors	<.0001	<.0001	<.0001	<.0001		<.0001	0.0009	0.5185	0.0247
PSA	0.17993	-0.53793	-0.55636	0.69505	0.78270	1.00000	0.06942	-0.04843	-0.04626
PSA	<.0001	<.0001	<.0001	<.0001	<.0001		0.0345	0.1404	0.1591
ROT	0.43289	0.29580	0.24778	-0.00357	0.10913	0.06942	1.00000	0.54368	-0.14116
ROT	<.0001	<.0001	<.0001	0.9136	0.0009	0.0345		<.0001	<.0001
NATOM	0.83011	0.54594	0.50801	-0.10123	0.02122	-0.04843	0.54368	1.00000	0.64240
NATOM	<.0001	<.0001	<.0001	0.0020	0.5185	0.1404	<.0001		<.0001
NRING	0.64869	0.37750	0.37494	-0.16417	0.07374	-0.04626	-0.14116	0.64240	1.00000
NRING	<.0001	<.0001	<.0001	<.0001	0.0247	0.1591	<.0001	<.0001	

Name: Millie Woo
[Student No: s3806940]



6. Using the SAS DISCRIM and your resultant outputs answer the following questions. Use priors "violators"=0.30 "non-violators"=0.70.

SAS Code:

```
proc discrim data=work.drug_sorted outstat=work.drug_stat
wcov pcov method=normal pool=test
distance anova manova listerr crosslisterr;
class score9_logD_group_1;
var MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
prior '1'=0.7 '2'=0.3;

run;
```

Name: Millie Woo
[Student No: s3806940]

Output:

Total Sample Size	1279	**DF Total**	1278
Variables	9	**DF Within Classes**	1277
Classes	2	**DF Between Classes**	1

Number of Observations Read	1279
Number of Observations Used	1279

Class Level Information					
score9_logD_group_1	**Variable Name**	**Frequency**	**Weight**	**Proportion**	**Prior Probability**
1	1	928	928.0000	0.725567	0.700000
2	2	351	351.0000	0.274433	0.300000

Within-Class Covariance Matrices

score9_logD_group_1 = 1, DF = 927									
Variable	**MW**	**LogP**	**LogD**	**Hdonors**	**Hacceptors**	**PSA**	**ROT**	**NATOM**	**NRING**
MW	9490.987463	99.000935	95.445763	-2.857872	55.813521	650.012636	118.403820	1171.999690	90.139688
LogP	99.000935	5.145999	4.959945	-1.444051	-1.411589	-45.249920	1.883935	17.947864	1.221433
LogD	95.445763	4.959945	5.572763	-1.606830	-1.545479	-48.702142	1.642228	17.379576	1.262452
Hdonors	-2.857872	-1.444051	-1.606830	1.516716	1.170435	31.741563	-0.012338	-1.806683	-0.288377
Hacceptors	55.813521	-1.411589	-1.545479	1.170435	3.427713	53.734900	0.567253	0.569371	0.194727
PSA	650.012636	-45.249920	-48.702142	31.741563	53.734900	1375.036816	7.227086	-26.026287	-2.446968
ROT	118.403820	1.883935	1.642228	-0.012338	0.567253	7.227086	7.882518	22.121123	-0.565298
NATOM	1171.999690	17.947864	17.379576	-1.806683	0.569371	-26.026287	22.121123	210.023490	13.278811
NRING	90.139688	1.221433	1.262452	-0.288377	0.194727	-2.446968	-0.565298	13.278811	2.034444

Name: Millie Woo
[Student No: s3806940]

score9_logD_group_1 = 2, DF = 350									
Variable	**MW**	**LogP**	**LogD**	**Hdonors**	**Hacceptors**	**PSA**	**ROT**	**NATOM**	**NRING**
MW	42941.95298	-149.16364	-194.49978	497.05281	604.30032	15216.13709	580.80168	5718.16941	143.78699
LogP	-149.16364	10.73427	10.85868	-7.53704	-9.16010	-231.16595	-1.84237	-9.95446	0.78645
LogD	-194.49978	10.85868	14.34372	-9.49579	-10.19377	-267.94172	-2.02739	-16.49255	0.99273
Hdonors	497.05281	-7.53704	-9.49579	12.27440	11.76230	308.47816	7.62687	65.19214	0.30077
Hacceptors	604.30032	-9.16010	-10.19377	11.76230	17.87044	376.59758	6.21071	71.41248	1.53175
PSA	15216.13709	-231.16595	-267.94172	308.47816	376.59758	9767.90931	225.24791	1851.06991	22.59476
ROT	580.80168	-1.84237	-2.02739	7.62687	6.21071	225.24791	24.14805	85.00882	-0.89093
NATOM	5718.16941	-9.95446	-16.49255	65.19214	71.41248	1851.06991	85.00882	893.53114	20.02093
NRING	143.78699	0.78645	0.99273	0.30077	1.53175	22.59476	-0.89093	20.02093	2.47484

Pooled Within-Class Covariance Matrix, DF = 1277									
Variable	**MW**	**LogP**	**LogD**	**Hdonors**	**Hacceptors**	**PSA**	**ROT**	**NATOM**	**NRING**
MW	18659.22390	30.98402	15.97752	134.15758	206.14271	4642.29420	245.13777	2418.01332	104.84334
LogP	30.98402	6.67763	6.57667	-3.11402	-3.53530	-96.20576	0.86263	10.30040	1.10221
LogD	15.97752	6.57667	7.97671	-3.76904	-3.91580	-108.79130	0.63646	8.09591	1.18853
Hdonors	134.15758	-3.11402	-3.76904	4.46518	4.07345	107.58949	2.08142	16.55635	-0.12690
Hacceptors	206.14271	-3.53530	-3.91580	4.07345	7.38618	142.22506	2.11401	19.98604	0.56118
PSA	4642.29420	-96.20576	-108.79130	107.58949	142.22506	3675.35426	66.98221	488.44800	4.41647
ROT	245.13777	0.86263	0.63646	2.08142	2.11401	66.98221	12.34057	39.35737	-0.65455
NATOM	2418.01332	10.30040	8.09591	16.55635	19.98604	488.44800	39.35737	397.35918	15.12669
NRING	104.84334	1.10221	1.18853	-0.12690	0.56118	4.41647	-0.65455	15.12669	2.15515

Within Covariance Matrix Information		
score9_logD_group_1	**Covariance Matrix Rank**	**Natural Log of the Determinant of the Covariance Matrix**
1	9	21.31705
2	9	30.65325
Pooled	9	25.70378

Name: Millie Woo
[Student No: s3806940]

Test of Homogeneity of Within Covariance Matrices

hi-Square	DF	Pr > ChiSq
2311.124852	45	<.0001

Squared Distance to score9_logD_group_1		
From score9_logD_group_1	1	2
1	0	3.60779
2	13.00818	0

Generalized Squared Distance to score9_logD_group_1		
From score9_logD_group_1	1	2
1	22.03040	36.66898
2	35.03859	33.06120

Univariate Test Statistics								
F Statistics, Num DF=1, Den DF=1277								
Variable	Label	Total Standard Deviation	Pooled Standard Deviation	Between Standard Deviation	R-Square	R-Square / (1-RSq)	F Value	Pr > F
MW	MW	178.2711	136.5988	162.0228	0.4133	0.7045	899.70	<.0001
LogP	LogP	2.5835	2.5841	0.0655	0.0003	0.0003	0.41	0.5217
LogD	LogD	2.8252	2.8243	0.1492	0.0014	0.0014	1.78	0.1819
Hdonors	Hdonors	2.3059	2.1131	1.3075	0.1609	0.1917	244.85	<.0001
Hacceptors	Hacceptors	3.4918	2.7178	3.1012	0.3947	0.6520	832.67	<.0001
PSA	PSA	70.9887	60.6247	52.2656	0.2712	0.3722	475.31	<.0001
ROT	ROT	4.0853	3.5129	2.9513	0.2612	0.3535	451.37	<.0001
NATOM	NATOM	24.8684	19.9339	21.0342	0.3580	0.5576	712.05	<.0001
NRING	NRING	1.6392	1.4680	1.0326	0.1986	0.2478	316.42	<.0001

Average R-Square	
Unweighted	0.228844
Weighted by Variance	0.3930344

Name: Millie Woo
[Student No: s3806940]

Multivariate Statistics and Exact F Statistics					
S=1 M=3.5 N=633.5					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.47115928	158.26	9	1269	<.0001
Pillai's Trace	0.52884072	158.26	9	1269	<.0001
Hotelling-Lawley Trace	1.12242450	158.26	9	1269	<.0001
Roy's Greatest Root	1.12242450	158.26	9	1269	<.0001

Misclassified observation

Resubstitution Results using Quadratic Discriminant Function

Name: Millie Woo
[Student No: s3806940]

osterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
17	1	2	*	0.3836	0.6164
18	1	2	*	0.0000	1.0000
31	1	2	*	0.4414	0.5586
32	1	2	*	0.4575	0.5425
39	1	2	*	0.1941	0.8059
68	1	2	*	0.4133	0.5867
75	1	2	*	0.3602	0.6398
77	1	2	*	0.0558	0.9442
108	1	2	*	0.0000	1.0000
112	1	2	*	0.0000	1.0000
159	1	2	*	0.4046	0.5954
178	1	2	*	0.0889	0.9111
208	1	2	*	0.0741	0.9259
220	1	2	*	0.0498	0.9502
231	1	2	*	0.0209	0.9791
233	1	2	*	0.0000	1.0000
242	1	2	*	0.0345	0.9655
253	1	2	*	0.0039	0.9961
272	1	2	*	0.0000	1.0000
276	1	2	*	0.0226	0.9774
293	1	2	*	0.0018	0.9982
303	1	2	*	0.0003	0.9997
308	1	2	*	0.3216	0.6784
345	1	2	*	0.0002	0.9998
365	1	2	*	0.2736	0.7264
383	1	2	*	0.4936	0.5064
403	1	2	*	0.0000	1.0000
405	1	2	*	0.0095	0.9905
406	1	2	*	0.3132	0.6868
426	1	2	*	0.0549	0.9451
429	1	2	*	0.0220	0.9780
435	1	2	*	0.0026	0.9974
436	1	2	*	0.1595	0.8405

Name: Millie Woo
[Student No: s3806940]

osterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
438	1	2	*	0.0004	0.9996
467	1	2	*	0.0880	0.9120
474	1	2	*	0.2408	0.7592
483	1	2	*	0.0178	0.9822
517	1	2	*	0.2574	0.7426
528	1	2	*	0.2669	0.7331
539	1	2	*	0.1489	0.8511
551	1	2	*	0.0822	0.9178
560	1	2	*	0.0000	1.0000
579	1	2	*	0.0332	0.9668
586	1	2	*	0.0040	0.9960
598	1	2	*	0.0474	0.9526
613	1	2	*	0.3265	0.6735
627	1	2	*	0.0064	0.9936
632	1	2	*	0.2662	0.7338
646	1	2	*	0.1996	0.8004
665	1	2	*	0.4428	0.5572
710	1	2	*	0.3042	0.6958
712	1	2	*	0.0485	0.9515
725	1	2	*	0.4085	0.5915
740	1	2	*	0.0029	0.9971
762	1	2	*	0.4512	0.5488
811	1	2	*	0.0238	0.9762
902	1	2	*	0.0676	0.9324
907	1	2	*	0.4851	0.5149
937	2	1	*	0.6256	0.3744
938	2	1	*	0.7560	0.2440
939	2	1	*	0.8203	0.1797
940	2	1	*	0.8598	0.1402
941	2	1	*	0.5801	0.4199
943	2	1	*	0.8785	0.1215
944	2	1	*	0.6297	0.3703
956	2	1	*	0.9764	0.0236

Name: Millie Woo
[Student No: s3806940]

osterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
959	2	1	*	0.8975	0.1025
960	2	1	*	0.9228	0.0772
961	2	1	*	0.9716	0.0284
966	2	1	*	0.9312	0.0688
967	2	1	*	0.5622	0.4378
975	2	1	*	0.6292	0.3708
977	2	1	*	0.8583	0.1417
982	2	1	*	0.9953	0.0047
984	2	1	*	0.8847	0.1153
985	2	1	*	0.9597	0.0403
986	2	1	*	0.7921	0.2079
991	2	1	*	0.8680	0.1320
992	2	1	*	0.8515	0.1485
995	2	1	*	0.8350	0.1650
996	2	1	*	0.7702	0.2298
999	2	1	*	0.9025	0.0975
1000	2	1	*	0.9175	0.0825
1001	2	1	*	0.6678	0.3322
1003	2	1	*	0.9574	0.0426
1005	2	1	*	0.5523	0.4477
1008	2	1	*	0.8382	0.1618
1009	2	1	*	0.9112	0.0888
1013	2	1	*	0.9354	0.0646
1017	2	1	*	0.9087	0.0913
1019	2	1	*	0.7657	0.2343
1020	2	1	*	0.9881	0.0119
1021	2	1	*	0.5929	0.4071
1024	2	1	*	0.9416	0.0584
1025	2	1	*	0.9801	0.0199
1027	2	1	*	0.8252	0.1748
1030	2	1	*	0.8433	0.1567
1032	2	1	*	0.8329	0.1671
1037	2	1	*	0.9384	0.0616

Name: Millie Woo
[Student No: s3806940]

osterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
1039	2	1	*	0.8023	0.1977
1044	2	1	*	0.8059	0.1941
1047	2	1	*	0.8911	0.1089
1048	2	1	*	0.9604	0.0396
1050	2	1	*	0.6219	0.3781
1051	2	1	*	0.8273	0.1727
1052	2	1	*	0.7447	0.2553
1058	2	1	*	0.9073	0.0927
1059	2	1	*	0.8969	0.1031
1060	2	1	*	0.5619	0.4381
1064	2	1	*	0.5045	0.4955
1065	2	1	*	0.8187	0.1813
1066	2	1	*	0.6576	0.3424
1070	2	1	*	0.7895	0.2105
1071	2	1	*	0.8182	0.1818
1072	2	1	*	0.9502	0.0498
1081	2	1	*	0.9769	0.0231
1082	2	1	*	0.6183	0.3817
1083	2	1	*	0.8822	0.1178
1088	2	1	*	0.7799	0.2201
1090	2	1	*	0.9654	0.0346
1094	2	1	*	0.9711	0.0289
1095	2	1	*	0.8149	0.1851
1096	2	1	*	0.9441	0.0559
1097	2	1	*	0.9386	0.0614
1105	2	1	*	0.7379	0.2621
1106	2	1	*	0.8947	0.1053
1108	2	1	*	0.7055	0.2945
1112	2	1	*	0.6854	0.3146
1113	2	1	*	0.9338	0.0662
1120	2	1	*	0.9640	0.0360
1122	2	1	*	0.7399	0.2601
1127	2	1	*	0.8616	0.1384

Name: Millie Woo
[Student No: s3806940]

osterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
1128	2	1	*	0.9196	0.0804
1129	2	1	*	0.7087	0.2913
1130	2	1	*	0.6243	0.3757
1133	2	1	*	0.8725	0.1275
1134	2	1	*	0.9295	0.0705
1136	2	1	*	0.8653	0.1347
1138	2	1	*	0.6178	0.3822
1139	2	1	*	0.9955	0.0045
1142	2	1	*	0.6895	0.3105
1145	2	1	*	0.9892	0.0108
1155	2	1	*	0.9497	0.0503
1159	2	1	*	0.9167	0.0833
1163	2	1	*	0.8861	0.1139
1164	2	1	*	0.7360	0.2640
1165	2	1	*	0.8869	0.1131
1166	2	1	*	0.8359	0.1641
1169	2	1	*	0.8016	0.1984
1170	2	1	*	0.9191	0.0809
1173	2	1	*	0.7038	0.2962
1178	2	1	*	0.7445	0.2555
1179	2	1	*	0.9485	0.0515
1184	2	1	*	0.9313	0.0687
1190	2	1	*	0.9486	0.0514
1192	2	1	*	0.9044	0.0956
1193	2	1	*	0.9215	0.0785
1195	2	1	*	0.9212	0.0788
1200	2	1	*	0.7430	0.2570
1201	2	1	*	0.7513	0.2487
1202	2	1	*	0.5198	0.4802
1203	2	1	*	0.9559	0.0441
1210	2	1	*	0.8827	0.1173
1211	2	1	*	0.6221	0.3779
1216	2	1	*	0.8887	0.1113

Name: Millie Woo
[Student No: s3806940]

osterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
1218	2	1	*	0.5276	0.4724
1220	2	1	*	0.9276	0.0724
1224	2	1	*	0.8989	0.1011
1227	2	1	*	0.7097	0.2903
1229	2	1	*	0.9677	0.0323
1243	2	1	*	0.8368	0.1632
1244	2	1	*	0.9204	0.0796
1255	2	1	*	0.8515	0.1485
1256	2	1	*	0.9809	0.0191
1257	2	1	*	0.9167	0.0833
1266	2	1	*	0.9777	0.0223
1268	2	1	*	0.8958	0.1042
1269	2	1	*	0.5919	0.4081
1270	2	1	*	0.9765	0.0235
1275	2	1	*	0.6063	0.3937
1278	2	1	*	0.8734	0.1266
1279	2	1	*	0.9433	0.0567

Number of Observations and Percent Classified into score9_logD_group_1			
From score9_logD_group_1	1	2	Total
1	870 93.75	58 6.25	928 100.00
2	124 35.33	227 64.67	351 100.00
Total	994 77.72	285 22.28	1279 100.00
Priors	0.7	0.3	

Name: Millie Woo
[Student No: s3806940]

Error Count Estimates for score9_logD_group_1			
	1	2	Total
Rate	0.0625	0.3533	0.1497
Priors	0.7000	0.3000	

From the above 2 tables, we could see that using Quadratic Discriminant analysis, 182 out of 1279 observations would be classified incorrectly. For non-violator (score9_logD_group_1 = 1), 58 out of 928 (6.25%) observations would be classified incorrectly. For violator (score9_logD_group_1 = 2), 124 out of 351 (35.33%) observations would be classified incorrectly.

Misclassified observation

Cross-validation Results using Quadratic Discriminant Function

Name: Millie Woo
[Student No: s3806940]

Posterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
17	1	2	*	0.3374	0.6626
18	1	2	*	0.0000	1.0000
31	1	2	*	0.3790	0.6210
32	1	2	*	0.3770	0.6230
39	1	2	*	0.1285	0.8715
68	1	2	*	0.3431	0.6569
75	1	2	*	0.3178	0.6822
77	1	2	*	0.0328	0.9672
108	1	2	*	0.0000	1.0000
112	1	2	*	0.0000	1.0000
140	1	2	*	0.4902	0.5098
155	1	2	*	0.4782	0.5218
159	1	2	*	0.3633	0.6367
178	1	2	*	0.0579	0.9421
208	1	2	*	0.0548	0.9452
220	1	2	*	0.0274	0.9726
231	1	2	*	0.0139	0.9861
233	1	2	*	0.0000	1.0000
242	1	2	*	0.0038	0.9962
253	1	2	*	0.0015	0.9985
257	1	2	*	0.4877	0.5123
272	1	2	*	0.0000	1.0000
276	1	2	*	0.0034	0.9966
293	1	2	*	0.0007	0.9993
303	1	2	*	0.0001	0.9999
308	1	2	*	0.2765	0.7235
345	1	2	*	0.0000	1.0000
365	1	2	*	0.2343	0.7657
383	1	2	*	0.4101	0.5899
403	1	2	*	0.0000	1.0000
405	1	2	*	0.0040	0.9960
406	1	2	*	0.2121	0.7879
426	1	2	*	0.0322	0.9678

Name: Millie Woo
[Student No: s3806940]

Posterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
429	1	2	*	0.0084	0.9916
435	1	2	*	0.0014	0.9986
436	1	2	*	0.1062	0.8938
438	1	2	*	0.0002	0.9998
450	1	2	*	0.4473	0.5527
467	1	2	*	0.0587	0.9413
474	1	2	*	0.1654	0.8346
483	1	2	*	0.0120	0.9880
517	1	2	*	0.2153	0.7847
528	1	2	*	0.2289	0.7711
539	1	2	*	0.1170	0.8830
551	1	2	*	0.0528	0.9472
560	1	2	*	0.0000	1.0000
579	1	2	*	0.0202	0.9798
586	1	2	*	0.0010	0.9990
598	1	2	*	0.0334	0.9666
613	1	2	*	0.2581	0.7419
627	1	2	*	0.0039	0.9961
632	1	2	*	0.2183	0.7817
638	1	2	*	0.4503	0.5497
646	1	2	*	0.1503	0.8497
665	1	2	*	0.4112	0.5888
668	1	2	*	0.4836	0.5164
710	1	2	*	0.2443	0.7557
712	1	2	*	0.0287	0.9713
725	1	2	*	0.3684	0.6316
740	1	2	*	0.0011	0.9989
762	1	2	*	0.4032	0.5968
811	1	2	*	0.0112	0.9888
871	1	2	*	0.4931	0.5069
902	1	2	*	0.0371	0.9629
907	1	2	*	0.3986	0.6014
937	2	1	*	0.6312	0.3688

Name: Millie Woo
[Student No: s3806940]

Posterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
938	2	1	*	0.7695	0.2305
939	2	1	*	0.8273	0.1727
940	2	1	*	0.8625	0.1375
941	2	1	*	0.5884	0.4116
943	2	1	*	0.8810	0.1190
944	2	1	*	0.6380	0.3620
956	2	1	*	0.9773	0.0227
959	2	1	*	0.9001	0.0999
960	2	1	*	0.9243	0.0757
961	2	1	*	0.9725	0.0275
966	2	1	*	0.9323	0.0677
967	2	1	*	0.5723	0.4277
975	2	1	*	0.6377	0.3623
977	2	1	*	0.8628	0.1372
982	2	1	*	0.9954	0.0046
984	2	1	*	0.8895	0.1105
985	2	1	*	0.9606	0.0394
986	2	1	*	0.7947	0.2053
991	2	1	*	0.8705	0.1295
992	2	1	*	0.8549	0.1451
995	2	1	*	0.8417	0.1583
996	2	1	*	0.7747	0.2253
999	2	1	*	0.9038	0.0962
1000	2	1	*	0.9237	0.0763
1001	2	1	*	0.6829	0.3171
1003	2	1	*	0.9586	0.0414
1005	2	1	*	0.5576	0.4424
1008	2	1	*	0.8407	0.1593
1009	2	1	*	0.9182	0.0818
1013	2	1	*	0.9364	0.0636
1017	2	1	*	0.9118	0.0882
1019	2	1	*	0.7716	0.2284
1020	2	1	*	0.9884	0.0116

Name: Millie Woo
[Student No: s3806940]

Posterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
1021	2	1	*	0.6054	0.3946
1024	2	1	*	0.9437	0.0563
1025	2	1	*	0.9804	0.0196
1027	2	1	*	0.8291	0.1709
1030	2	1	*	0.8462	0.1538
1032	2	1	*	0.8417	0.1583
1037	2	1	*	0.9401	0.0599
1039	2	1	*	0.8054	0.1946
1044	2	1	*	0.8154	0.1846
1047	2	1	*	0.8933	0.1067
1048	2	1	*	0.9616	0.0384
1050	2	1	*	0.6285	0.3715
1051	2	1	*	0.8341	0.1659
1052	2	1	*	0.7532	0.2468
1058	2	1	*	0.9112	0.0888
1059	2	1	*	0.8992	0.1008
1060	2	1	*	0.5673	0.4327
1064	2	1	*	0.5098	0.4902
1065	2	1	*	0.8240	0.1760
1066	2	1	*	0.6622	0.3378
1068	2	1	*	0.5036	0.4964
1070	2	1	*	0.7940	0.2060
1071	2	1	*	0.8263	0.1737
1072	2	1	*	0.9524	0.0476
1081	2	1	*	0.9776	0.0224
1082	2	1	*	0.6285	0.3715
1083	2	1	*	0.8837	0.1163
1088	2	1	*	0.7830	0.2170
1090	2	1	*	0.9660	0.0340
1094	2	1	*	0.9718	0.0282
1095	2	1	*	0.8269	0.1731
1096	2	1	*	0.9456	0.0544
1097	2	1	*	0.9396	0.0604

Name: Millie Woo
[Student No: s3806940]

Posterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
1105	2	1	*	0.7451	0.2549
1106	2	1	*	0.8994	0.1006
1108	2	1	*	0.7190	0.2810
1112	2	1	*	0.7164	0.2836
1113	2	1	*	0.9358	0.0642
1120	2	1	*	0.9647	0.0353
1122	2	1	*	0.7528	0.2472
1127	2	1	*	0.8801	0.1199
1128	2	1	*	0.9207	0.0793
1129	2	1	*	0.7132	0.2868
1130	2	1	*	0.6295	0.3705
1133	2	1	*	0.8746	0.1254
1134	2	1	*	0.9315	0.0685
1136	2	1	*	0.8702	0.1298
1138	2	1	*	0.6251	0.3749
1139	2	1	*	0.9956	0.0044
1142	2	1	*	0.7050	0.2950
1145	2	1	*	0.9895	0.0105
1155	2	1	*	0.9507	0.0493
1159	2	1	*	0.9189	0.0811
1163	2	1	*	0.8902	0.1098
1164	2	1	*	0.7397	0.2603
1165	2	1	*	0.8892	0.1108
1166	2	1	*	0.8409	0.1591
1169	2	1	*	0.8083	0.1917
1170	2	1	*	0.9202	0.0798
1173	2	1	*	0.7115	0.2885
1178	2	1	*	0.7527	0.2473
1179	2	1	*	0.9492	0.0508
1184	2	1	*	0.9331	0.0669
1190	2	1	*	0.9496	0.0504
1192	2	1	*	0.9070	0.0930
1193	2	1	*	0.9228	0.0772

Name: Millie Woo
[Student No: s3806940]

Posterior Probability of Membership in score9_logD_group_1					
Obs	From score9_logD_group_1	Classified into score9_logD_group_1		1	2
1195	2	1	*	0.9228	0.0772
1200	2	1	*	0.7517	0.2483
1201	2	1	*	0.7571	0.2429
1202	2	1	*	0.5256	0.4744
1203	2	1	*	0.9566	0.0434
1210	2	1	*	0.8856	0.1144
1211	2	1	*	0.6286	0.3714
1216	2	1	*	0.8903	0.1097
1218	2	1	*	0.5327	0.4673
1220	2	1	*	0.9289	0.0711
1224	2	1	*	0.9010	0.0990
1227	2	1	*	0.7272	0.2728
1229	2	1	*	0.9682	0.0318
1243	2	1	*	0.8397	0.1603
1244	2	1	*	0.9227	0.0773
1255	2	1	*	0.8546	0.1454
1256	2	1	*	0.9814	0.0186
1257	2	1	*	0.9189	0.0811
1266	2	1	*	0.9783	0.0217
1268	2	1	*	0.9028	0.0972
1269	2	1	*	0.6067	0.3933
1270	2	1	*	0.9769	0.0231
1275	2	1	*	0.6239	0.3761
1278	2	1	*	0.8764	0.1236
1279	2	1	*	0.9451	0.0549

Name: Millie Woo
[Student No: s3806940]

Number of Observations and Percent Classified into score9_logD_group_1			
From score9_logD_group_1	1	2	Total
1	863 93.00	65 7.00	928 100.00
2	125 35.61	226 64.39	351 100.00
Total	988 77.25	291 22.75	1279 100.00
Priors	0.7	0.3	

Error Count Estimates for score9_logD_group_1			
	1	2	Total
Rate	0.0700	0.3561	0.1559
Priors	0.7000	0.3000	

From the above 2 tables, we could see that using Quadratic Discriminant analysis (with cross validation) ,190 out of 1279 observations would be classified incorrectly. For non-violator (score9_logD_group_1 = 1), 65 out of 928 (7.00%) observations would be classified incorrectly. For violator (score9_logD_group_1 = 2), 125 out of 351 (35.61%) observations would be classified incorrectly.

7. Is $\Sigma_1 = \Sigma_2$ Justify your answer.

From the test of homogeneity of within covariance metrices, we found that $p < 0.0001$, which means it is significant to reject the null hypothesis that the covariance metrics for violator and non-violator are equal, thus we are at 90% confidence that $\Sigma_1 \neq \Sigma_2$.

hi-Square	DF	Pr > ChiSq
2311.124852	45	<.0001

8. How is a molecule with $X0^T$ = (MW, LogP, LogD, Hdonors, Hacceptors, PSA, ROT, NATOM, NRING) = (445.429, -2.7, -3.28938, 8, 12, 207.27, 9, 55, 3) allocated? i.e. allocates it to either the violators or the non-violators group.

SAS code:

```
data x0;
input MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
datalines;
445.429 -2.7 -3.28938 8 12 207.27 9 55 3
;
```

Name: Millie Woo
[Student No: s3806940]

```
run;

proc discrim data=work.drug_sorted testdata=x0;
prior '1'=0.7 '2'=0.3;
class score9_logD_group_1;
run;
```

Output:

Observation Profile for Test Data	
Number of Observations Read	1
Number of Observations Used	1

Number of Observations and Percent Classified into score9_logD_group_1			
	1	2	Total
Total	0 0.00	1 100.00	1 100.00
Priors	0.7	0.3	

Thus, SAS would allocate $X0^T$ = (MW, LogP, LogD, Hdonors, Hacceptors, PSA, ROT, NATOM, NRING) = (445.429, -2.7, -3.28938, 8, 12, 207.27, 9, 55, 3) to score9_logD_group_1=2, which is the violator.

9. Write down the resultant confusion matrix.

From the given file, we can spot that $X0^T$ = (MW, LogP, LogD, Hdonors, Hacceptors, PSA, ROT, NATOM, NRING) = (445.429, -2.7, -3.28938, 8, 12, 207.27, 9, 55, 3) belongs to the observation Drug#Card= 116, which has the value of 2 for score9_logD_group_1, that is violator.

Thus, the confusion matrix would be:

	Classification under Quadratic Discriminant Function in SAS		
True Population		Violator	Non-violator
	Violator	1	0
	Non-violator	0	0

Question 4: Stepwise Discriminant on 4 groups of Molecules

1. For Question 4 you will need to create the following variable i.e. an interaction term between oral status and score 9_ Log D violation status at 4 levels as defined below:

SAS code:

```
data work.drug_oral_score;
      set work.drug_sorted;
      length score9_logD_group_1_name $ 15;
```

Name: Millie Woo
[Student No: s3806940]

```
      select;
            when ((score9_logD_group_1 = 2 ) and (oral_status = 'oral'))
                  do;
                  oral_score=1;
                  score9_logD_group_1_name='violator';
                end;
            when ((score9_logD_group_1 = 1) and (oral_status = 'oral'))
                  do;
                  oral_score=2;
                      score9_logD_group_1_name='non-violator';
                  end;
            when ((score9_logD_group_1 = 2) and (oral_status = 'non_oral'))
                  do;
                  oral_score=3;
                  score9_logD_group_1_name='violator';
                end;
            when ((score9_logD_group_1 = 1) and (oral_status = 'non_oral'))
                  do;
                  oral_score=4;
                  score9_logD_group_1_name='non-violator';
                end;
      end;
      keep MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING oral_status oral_score
score9_logD_group_1_name;
run;
```

After executing the above code chunk, a new table work.drug_oral_score will be created with 2 new fields oral_score such that
oral_score = 1 and score9_logD_group_1_name= ‘violator’ when oral_status = ‘oral’ and score9_logD_group_1 = 2
oral_score = 2 and score9_logD_group_1_name= ‘non-violator’ when oral_status = ‘oral’ and score9_logD_group_1 = 1
oral_score = 3 and score9_logD_group_1_name= ‘violator’ when oral_status = ‘non_oral’ and score9_logD_group_1 = 2
oral_score = 4 and score9_logD_group_1_name= ‘non-violator’ when oral_status = ‘non_oral’ and score9_logD_group_1 = 1

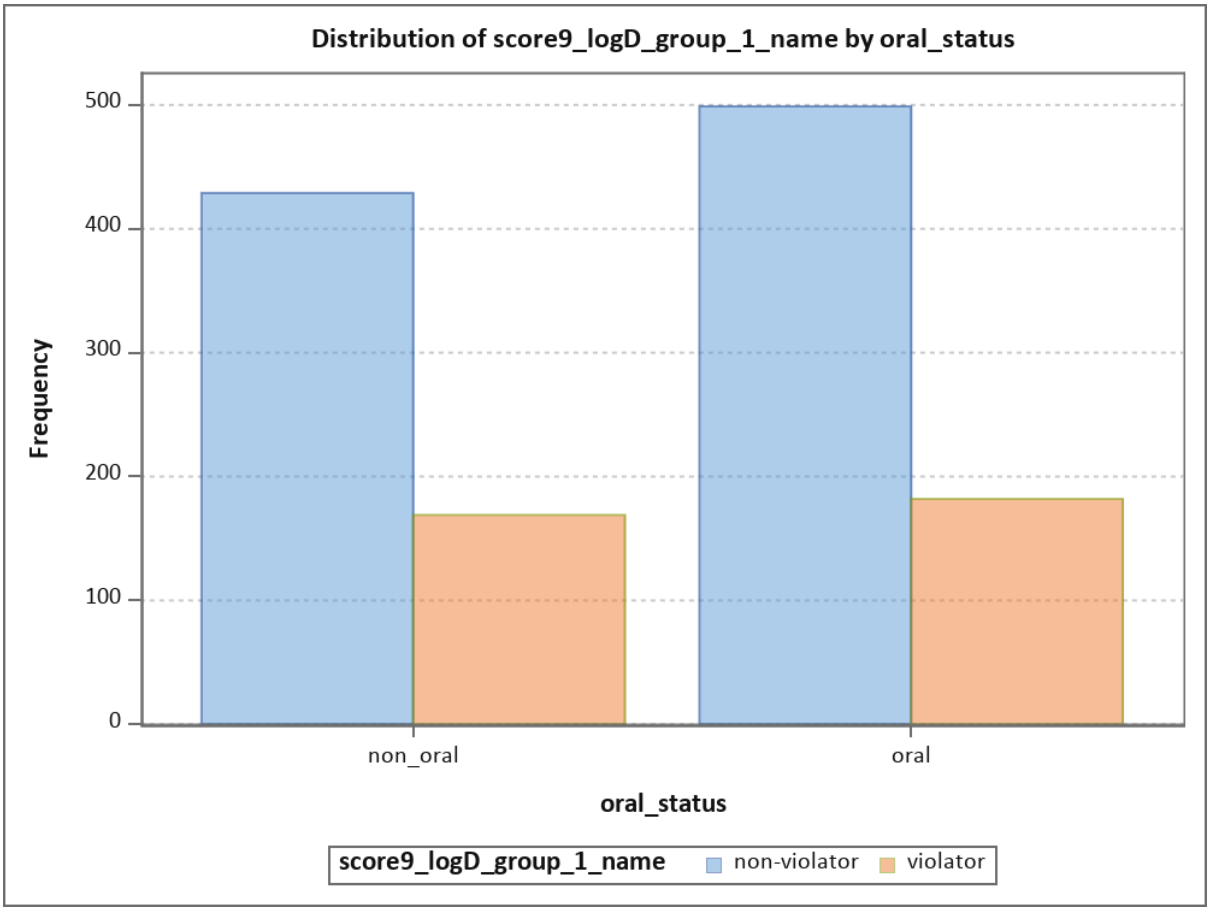
2. Crosstabulate in SAS or otherwise oral by violatory status for the whole group. How many molecules in each of these 4 levels? Create a table or histogram.

SAS code:

```
proc freq data=work.drug_oral_score;
   tables score9_logD_group_1_name* oral_status / plots=freqplot(twoway=cluster);
run;
```

Name: Millie Woo
[Student No: s3806940]

Table of score9_logD_group_1_name by oral_status			
score9_logD_group_1_name	oral_status(oral_status)		
Frequency Percent Row Pct Col Pct	non_oral	oral	Total
non-violator	429 33.54 46.23 71.74	499 39.01 53.77 73.27	928 72.56
violator	169 13.21 48.15 28.26	182 14.23 51.85 26.73	351 27.44
Total	598 46.76	681 53.24	1279 100.00



Name: Millie Woo
[Student No: s3806940]

3. Run a STEPWISE DISCRIM analysis using the above 4 level grouping variable.

SAS code:

```
proc stepdisc data=work.drug_oral_score bsscp tsscp;
class oral_score;
var MW LogP LogD Hdonors Hacceptors PSA ROT NATOM NRING;
run;
```

Output:

The STEPDISC Procedure

The Method for Selecting Variables is STEPWISE			
Total Sample Size	1279	**Variable(s) in the Analysis**	9
Class Levels	4	**Variable(s) Will Be Included**	0
		Significance Level to Enter	0.15
		Significance Level to Stay	0.15

Number of Observations Read	1279
Number of Observations Used	1279

Class Level Information				
oral_score	Variable Name	Frequency	Weight	Proportion
1	1	182	182.0000	0.142299
2	2	499	499.0000	0.390149
3	3	169	169.0000	0.132134
4	4	429	429.0000	0.335418

Name: Millie Woo
[Student No: s3806940]

Between-Class SSCP Matrix									
Variable	**MW**	**LogP**	**LogD**	**Hdonors**	**Hacceptors**	**PSA**	**ROT**	**NATOM**	**NRING**
MW	17958160.35	-4980.70	-24721.99	153100.71	338108.09	5992565.43	322498.28	2328252.41	113323.35
LogP	-4980.70	437.15	401.29	-335.47	-150.06	-8244.83	-123.63	-520.54	72.82
LogD	-24721.99	401.29	400.23	-476.14	-539.29	-14161.18	-492.64	-3090.58	-55.94
Hdonors	153100.71	-335.47	-476.14	1502.26	2920.77	55514.41	2772.73	19765.10	895.93
Hacceptors	338108.09	-150.06	-539.29	2920.77	6430.05	113559.05	6120.45	43815.95	2105.91
PSA	5992565.43	-8244.83	-14161.18	55514.41	113559.05	2099373.68	108042.48	775043.51	36269.81
ROT	322498.28	-123.63	-492.64	2772.73	6120.45	108042.48	5828.44	41799.22	2015.89
NATOM	2328252.41	-520.54	-3090.58	19765.10	43815.95	775043.51	41799.22	301891.08	14722.97
NRING	113323.35	72.82	-55.94	895.93	2105.91	36269.81	2015.89	14722.97	743.61

Total-Sample SSCP Matrix									
Variable	**MW**	**LogP**	**LogD**	**Hdonors**	**Hacceptors**	**PSA**	**ROT**	**NATOM**	**NRING**
MW	40615588.98	46351.67	4948.19	306797.34	584567.09	11343640.41	618836.15	5267228.75	240881.08
LogP	46351.67	8530.08	8392.16	-3921.85	-4384.71	-120666.02	1225.17	14034.46	1450.77
LogD	4948.19	8392.16	10200.48	-4937.78	-5296.29	-143912.03	531.24	8332.06	1419.25
Hdonors	306797.34	-3921.85	-4937.78	6795.35	7794.89	181094.60	5125.75	38730.53	701.41
Hacceptors	584567.09	-4384.71	-5296.29	7794.89	15582.36	285274.40	8552.61	67237.05	2764.56
PSA	11343640.41	-120666.02	-143912.03	181094.60	285274.40	6440348.34	184180.34	1326791.81	40154.86
ROT	618836.15	1225.17	531.24	5125.75	8552.61	184180.34	21329.08	89958.41	1113.12
NATOM	5267228.75	14034.46	8332.06	38730.53	67237.05	1326791.81	89958.41	790365.75	33207.27
NRING	240881.08	1450.77	1419.25	701.41	2764.56	40154.86	1113.12	33207.27	3434.06

Stepwise Selection: Step 1

Step 1: the tolerance is 1.0 for each variable under consideration because no variables have yet entered the model. The variable MW is selected because it's F statistics, 336.85, is largest among all variables.

Name: Millie Woo
[Student No: s3806940]

Statistics for Entry, DF = 3, 1275					
Variable	**Label**	**R-Square**	**F Value**	**Pr > F**	**Tolerance**
MW	MW	0.4421	336.85	<.0001	1.0000
LogP	LogP	0.0512	22.96	<.0001	1.0000
LogD	LogD	0.0392	17.36	<.0001	1.0000
Hdonors	Hdonors	0.2211	120.62	<.0001	1.0000
Hacceptors	Hacceptors	0.4126	298.59	<.0001	1.0000
PSA	PSA	0.3260	205.54	<.0001	1.0000
ROT	ROT	0.2733	159.81	<.0001	1.0000
NATOM	NATOM	0.3820	262.66	<.0001	1.0000
NRING	NRING	0.2165	117.46	<.0001	1.0000

Variable MW will be entered.

Variable(s) That Have Been Entered
MW

Multivariate Statistics					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.557851	336.85	3	1275	<.0001
Pillai's Trace	0.442149	336.85	3	1275	<.0001
Average Squared Canonical Correlation	0.147383				

Stepwise Selection: Step 2

Step 2: with MW already in the model, MW is tested for removal before new variable is selected for entry. Since MW meets the criterion to stay, it is used as a covariate in the analysis of covariance for variable selection.
The variable Hacceptors is selected since it F statistics, 45.77, is the largest among all the variables not in the model and because its associated tolerance, 0.4601 meets the criterion to enter.

Statistics for Removal, DF = 3, 1275				
Variable	**Label**	**R-Square**	**F Value**	**Pr > F**
MW	MW	0.4421	336.85	<.0001

Name: Millie Woo
[Student No: s3806940]

No variables can be removed.

Statistics for Entry, DF = 3, 1274					
Variable	Label	Partial R-Square	F Value	Pr > F	Tolerance
LogP	LogP	0.0590	26.65	<.0001	0.9938
LogD	LogD	0.0430	19.08	<.0001	0.9999
Hdonors	Hdonors	0.0508	22.72	<.0001	0.6590
Hacceptors	Hacceptors	0.0973	45.77	<.0001	0.4601
PSA	PSA	0.0596	26.90	<.0001	0.5081
ROT	ROT	0.0231	10.06	<.0001	0.5579
NATOM	NATOM	0.0003	0.14	0.9340	0.1357
NRING	NRING	0.0165	7.13	<.0001	0.5840

Variable Hacceptors will be entered.

Variable(s) That Have Been Entered	
MW	Hacceptors

Multivariate Statistics					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.503578	173.77	6	2548	<.0001
Pillai's Trace	0.500388	141.81	6	2550	<.0001
Average Squared Canonical Correlation	0.166796				

Stepwise Selection: Step 3

Step 3: with MW and Hacceptors already in the model, MW and Hacceptors are tested for removal before new variable is selected for entry. Since MW and Hacceptors both meet the criterion to stay, they are used as a covariate in the analysis of covariance for variable selection.
The variable LogP is selected since its F statistics, 47.02, is the largest among all the variables not in the model and because its associated tolerance, 0.2669 meets the criterion to enter.

Statistics for Removal, DF = 3, 1274				
Variable	Label	Partial R-Square	F Value	Pr > F
MW	MW	0.1426	70.65	<.0001
Hacceptors	Hacceptors	0.0973	45.77	<.0001

No variables can be removed.

Statistics for Entry, DF = 3, 1273					
Variable	Label	Partial R-Square	F Value	Pr > F	Tolerance
LogP	LogP	0.0998	47.02	<.0001	0.2669
LogD	LogD	0.0659	29.91	<.0001	0.2788
Hdonors	Hdonors	0.0759	34.87	<.0001	0.2964
PSA	PSA	0.0915	42.73	<.0001	0.1681
ROT	ROT	0.0298	13.06	<.0001	0.3286
NATOM	NATOM	0.0111	4.76	0.0026	0.0892
NRING	NRING	0.0240	10.42	<.0001	0.3027

Variable LogP will be entered.

Variable(s) That Have Been Entered		
MW	LogP	Hacceptors

Multivariate Statistics					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.453345	132.23	9	3098.3	<.0001
Pillai's Trace	0.590008	104.05	9	3825	<.0001
Average Squared Canonical Correlation	0.196669				

Stepwise Selection: Step 4

Step 4: with MW, LogP and Hacceptors already in the model, MW, LogP and Hacceptors are tested for removal before new variable is selected for entry. Since MW, LogP and Hacceptors both meet the criterion to stay, they are used as a covariate in the analysis of covariance for variable selection.
The variable ROT is selected since its F statistics, 12.57, is the largest among all the variables not in the model and because its associated tolerance, 0.2482 meets the criterion to enter.

Statistics for Removal, DF = 3, 1273				
Variable	Label	Partial R-Square	F Value	Pr > F
MW	MW	0.0818	37.83	<.0001
LogP	LogP	0.0998	47.02	<.0001
Hacceptors	Hacceptors	0.1363	66.99	<.0001

No variables can be removed.

Statistics for Entry, DF = 3, 1272					
Variable	Label	Partial R-Square	F Value	Pr > F	Tolerance
LogD	LogD	0.0043	1.84	0.1381	0.1744
Hdonors	Hdonors	0.0203	8.81	<.0001	0.2477
PSA	PSA	0.0246	10.68	<.0001	0.1140
ROT	ROT	0.0288	12.57	<.0001	0.2482
NATOM	NATOM	0.0086	3.69	0.0116	0.0849
NRING	NRING	0.0110	4.71	0.0028	0.2524

Variable ROT will be entered.

Variable(s) That Have Been Entered			
MW	LogP	Hacceptors	ROT

Multivariate Statistics					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.440288	101.95	12	3365.7	<.0001
Pillai's Trace	0.605907	80.61	12	3822	<.0001
Average Squared Canonical Correlation	0.201969				

Stepwise Selection: Step 5

Step 5: with MW, LogP, Hacceptors and ROT already in the model, MW, LogP, Hacceptors and ROT are tested for removal before new variable is selected for entry. Since MW, LogP, Hacceptors and ROT both meet the criterion to stay, they are used as a covariate in the analysis of covariance for variable selection.
The variable NRING is selected since its F statistics, 20.1, is the largest among all the variables not in the model and because its associated tolerance, 0.1483 meets the criterion to enter.

Statistics for Removal, DF = 3, 1272				
Variable	Label	Partial R-Square	F Value	Pr > F
MW	MW	0.0458	20.36	<.0001
LogP	LogP	0.0988	46.47	<.0001
Hacceptors	Hacceptors	0.1394	68.66	<.0001
ROT	ROT	0.0288	12.57	<.0001

No variables can be removed.

Statistics for Entry, DF = 3, 1271					
Variable	Label	Partial R-Square	F Value	Pr > F	Tolerance
LogD	LogD	0.0044	1.87	0.1334	0.1743
Hdonors	Hdonors	0.0241	10.47	<.0001	0.2291
PSA	PSA	0.0316	13.85	<.0001	0.1088
NATOM	NATOM	0.0025	1.06	0.3670	0.0848
NRING	NRING	0.0453	20.10	<.0001	0.1483

Variable NRING will be entered.

Variable(s) That Have Been Entered				
MW	LogP	Hacceptors	ROT	NRING

Multivariate Statistics					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.420349	86.28	15	3509.1	<.0001
Pillai's Trace	0.628720	67.50	15	3819	<.0001
Average Squared Canonical Correlation	0.209573				

Stepwise Selection: Step 6

Step 6: with MW, LogP, Hacceptors, ROT and NRING already in the model, MW, LogP, Hacceptors, ROT and NRING are tested for removal before new variable is selected for entry. Since MW, LogP, Hacceptors, ROT and NRING both meet the criterion to stay, they are used as a covariate in the analysis of covariance for variable selection. The variable PSA is selected since its F statistics, 1.24, is the largest among all the variables not in the model and because its associated tolerance, 0.1067 meets the criterion to enter.

Statistics for Removal, DF = 3, 1271				
Variable	**Label**	**Partial R-Square**	**F Value**	**Pr > F**
MW	MW	0.0195	8.41	<.0001
LogP	LogP	0.0755	34.62	<.0001
Hacceptors	Hacceptors	0.1446	71.64	<.0001
ROT	ROT	0.0625	28.24	<.0001
NRING	NRING	0.0453	20.10	<.0001

No variables can be removed.

Statistics for Entry, DF = 3, 1270					
Variable	**Label**	**Partial R-Square**	**F Value**	**Pr > F**	**Tolerance**
LogD	LogD	0.0036	1.53	0.2045	0.1477
Hdonors	Hdonors	0.0171	7.35	<.0001	0.1301
PSA	PSA	0.0281	12.24	<.0001	0.1067
NATOM	NATOM	0.0007	0.29	0.8338	0.0791

Variable PSA will be entered.

Variable(s) That Have Been Entered					
MW	LogP	Hacceptors	PSA	ROT	NRING

Multivariate Statistics					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.408538	74.31	18	3592.6	<.0001
Pillai's Trace	0.653712	59.07	18	3816	<.0001
Average Squared Canonical Correlation	0.217904				

Stepwise Selection: Step 7

Step 7: with MW, LogP, Hacceptors, PSA, ROT and NRING already in the model, MW, LogP, Hacceptors, PSA, ROT and NRING are tested for removal before new variable is selected for entry. Since MW, LogP, Hacceptors, PSA, ROT and NRING both meet the criterion to stay, they are used as a covariate in the analysis of covariance for variable selection.
The variable LogD is selected since its F statistics, 4.55, is the largest among all the variables not in the model and because its associated tolerance, 0.1005 meets the criterion to enter.

Statistics for Removal, DF = 3, 1270				
Variable	**Label**	**Partial R-Square**	**F Value**	**Pr > F**
MW	MW	0.0050	2.15	0.0927
LogP	LogP	0.0239	10.38	<.0001
Hacceptors	Hacceptors	0.1416	69.85	<.0001
PSA	PSA	0.0281	12.24	<.0001
ROT	ROT	0.0657	29.75	<.0001
NRING	NRING	0.0418	18.46	<.0001

No variables can be removed.

Statistics for Entry, DF = 3, 1269					
Variable	**Label**	**Partial R-Square**	**F Value**	**Pr > F**	**Tolerance**
LogD	LogD	0.0106	4.55	0.0035	0.1005
Hdonors	Hdonors	0.0070	2.97	0.0309	0.0839
NATOM	NATOM	0.0006	0.27	0.8438	0.0734

Variable LogD will be entered.

Variable(s) That Have Been Entered						
MW	LogP	LogD	Hacceptors	PSA	ROT	NRING

Multivariate Statistics					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.404188	64.37	21	3644.4	<.0001
Pillai's Trace	0.663483	51.56	21	3813	<.0001
Average Squared Canonical Correlation	0.221161				

Stepwise Selection: Step 8

Step 8: with MW, LogP, LogD, Hacceptors, PSA, ROT and NRING already in the model, MW, LogP, LogD, Hacceptors, PSA, ROT and NRING are tested for removal before new variable is selected for entry. Since MW, LogP, LogD, Hacceptors, PSA, ROT and NRING both meet the criterion to stay, they are used as a covariate in the analysis of covariance for variable selection.
The variable Hdonors is selected since its F statistics, 4.15, is the largest among all the variables not in the model and because its associated tolerance, 0.0828 meets the criterion to enter.

Statistics for Removal, DF = 3, 1269				
Variable	**Label**	**Partial R-Square**	**F Value**	**Pr > F**
MW	MW	0.0047	1.98	0.1145
LogP	LogP	0.0244	10.57	<.0001
LogD	LogD	0.0106	4.55	0.0035
Hacceptors	Hacceptors	0.1440	71.16	<.0001
PSA	PSA	0.0350	15.33	<.0001
ROT	ROT	0.0666	30.18	<.0001
NRING	NRING	0.0416	18.38	<.0001

No variables can be removed.

Statistics for Entry, DF = 3, 1268					
Variable	Label	Partial R-Square	F Value	Pr > F	Tolerance
Hdonors	Hdonors	0.0097	4.15	0.0061	0.0828
NATOM	NATOM	0.0006	0.25	0.8582	0.0730

Variable Hdonors will be entered.

Variable(s) That Have Been Entered							
MW	LogP	LogD	Hdonors	Hacceptors	PSA	ROT	NRING

Multivariate Statistics					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.400255	56.89	24	3678.2	<.0001
Pillai's Trace	0.671810	45.81	24	3810	<.0001
Average Squared Canonical Correlation	0.223937				

Stepwise Selection: Step 9

Step 9: with MW, LogP, LogD, Hdonors, Hacceptors, PSA, ROT and NRING already in the model, MW, LogP, LogD, Hdonors, Hacceptors, PSA, ROT and NRING are tested for removal before new variable is selected for entry. Since MW does not meet the criterion to stay, it is removed. LogP, LogD, Hdonors, Hacceptors, PSA, ROT and NRING both meet the criterion to stay, they are used as a covariate in the analysis of covariance for variable selection.

Statistics for Removal, DF = 3, 1268				
Variable	Label	Partial R-Square	F Value	Pr > F
MW	MW	0.0026	1.09	0.3517
LogP	LogP	0.0265	11.50	<.0001
LogD	LogD	0.0134	5.74	0.0007
Hdonors	Hdonors	0.0097	4.15	0.0061
Hacceptors	Hacceptors	0.1417	69.76	<.0001
PSA	PSA	0.0215	9.28	<.0001
ROT	ROT	0.0659	29.83	<.0001
NRING	NRING	0.0382	16.80	<.0001

Variable MW will be removed.

Variable(s) That Have Been Entered						
LogP	LogD	Hdonors	Hacceptors	PSA	ROT	NRING

Multivariate Statistics					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.401289	64.97	21	3644.4	<.0001
Pillai's Trace	0.669514	52.16	21	3813	<.0001
Average Squared Canonical Correlation	0.223171				

Stepwise Selection: Step 10

Step 10: with LogP, LogD, Hdonors, Hacceptors, PSA, ROT and NRING already in the model, LogP, LogD, Hdonors, Hacceptors, PSA, ROT and NRING are tested for removal before new variable is selected for entry. Since LogP, LogD, Hdonors, Hacceptors, PSA, ROT and NRING both meet the criterion to stay, they are used as a covariate in the analysis of covariance for variable selection.
As NATOM's tolerance is 0.0888, it does not meet the criterion to enter.

Statistics for Removal, DF = 3, 1269				
Variable	**Label**	**Partial R-Square**	**F Value**	**Pr > F**
LogP	LogP	0.0245	10.64	<.0001
LogD	LogD	0.0142	6.11	0.0004
Hdonors	Hdonors	0.0118	5.05	0.0018
Hacceptors	Hacceptors	0.1431	70.62	<.0001
PSA	PSA	0.0260	11.29	<.0001
ROT	ROT	0.0916	42.64	<.0001
NRING	NRING	0.0660	29.89	<.0001

No variables can be removed.

Statistics for Entry, DF = 3, 1268					
Variable	Label	Partial R-Square	F Value	Pr > F	Tolerance
MW	MW	0.0026	1.09	0.3517	0.0828
NATOM	NATOM	0.0005	0.23	0.8752	0.0888

No variables can be entered.

No further steps are possible.

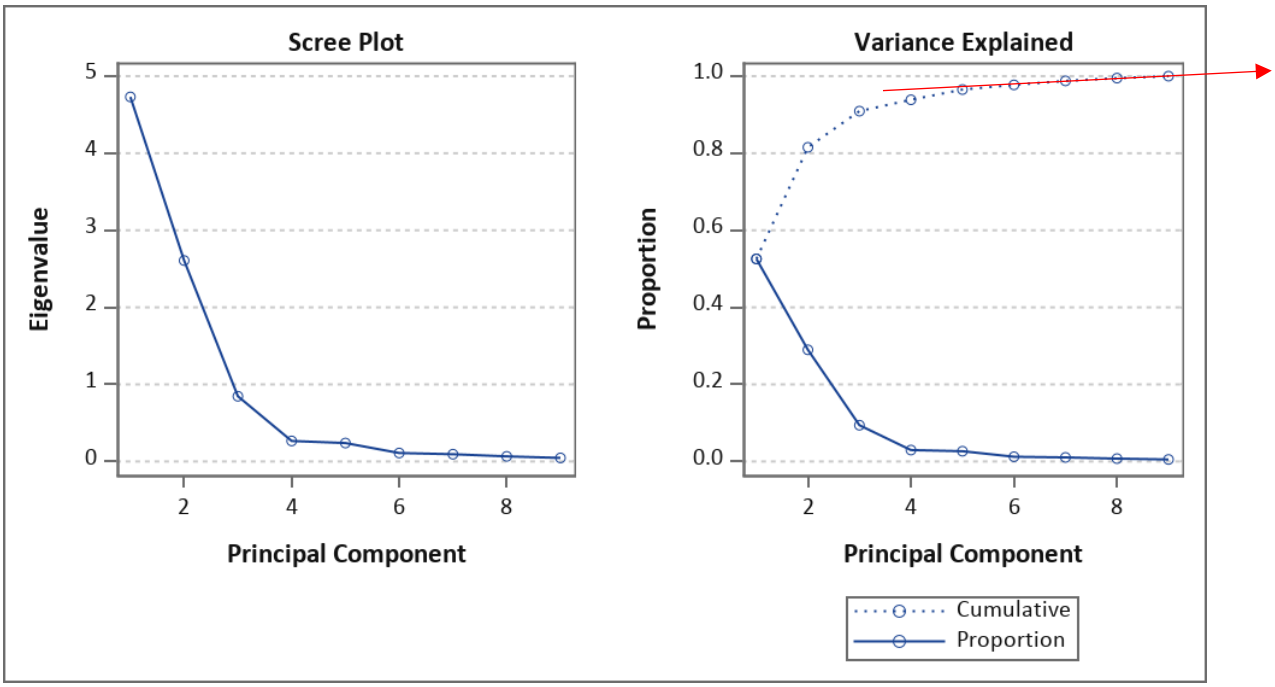
Stepwise Selection Summary										
Step	Number In	Entered	Removed	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Squared Canonical Correlation	Pr > ASCC
1	1	MW		0.4421	336.85	<.0001	0.55785055	<.0001	0.14738315	<.0001
2	2	Hacceptors		0.0973	45.77	<.0001	0.50357785	<.0001	0.16679607	<.0001
3	3	LogP		0.0998	47.02	<.0001	0.45334506	<.0001	0.19666926	<.0001
4	4	ROT		0.0288	12.57	<.0001	0.44028838	<.0001	0.20196915	<.0001
5	5	NRING		0.0453	20.10	<.0001	0.42034939	<.0001	0.20957329	<.0001
6	6	PSA		0.0281	12.24	<.0001	0.40853831	<.0001	0.21790399	<.0001
7	7	LogD		0.0106	4.55	0.0035	0.40418761	<.0001	0.22116106	<.0001
8	8	Hdonors		0.0097	4.15	0.0061	0.40025519	<.0001	0.22393663	<.0001
9	7		MW	0.0026	1.09	0.3517	0.40128866	<.0001	0.22317131	<.0001

4. Which variables best discriminate the 4 oral by violatory groups/classes?

From the above SAS output, in descending order, Hacceptors, LogP, ROT, NRING, PSA, LogD and Hdonors are all best to discriminate the 4 oral by violatory classes.

5. Write a clear description of your conclusions include the SAS code and outputs. (10 marks)

There are 1279 observations in this dataset. From principal analysis, the 9 molecular variables can be reduced to 5 dimensions to effectively summarize 96.52% of its variations.



Out of the 1279 observations, 928 are from the group of non-violators, 351 are from the group of violators. We can summarize the eigenvalues and the variations proportion of the first 5 principals explaining the overall observations, non-violators and violators group in the following table:

Eigenvalues of the Correlation Matrix						
	Overall Observations		Non-Violators		Violators	
	Eigenvalue	Proportion	Eigenvalue	Proportion	Eigenvalue	Proportion
λ_1	4.73045355	0.5256	3.84083397	0.4268	5.09425408	0.5660
λ_2	2.60678006	0.2896	2.61219333	0.2902	1.98208283	0.2202
λ_3	0.84552994	0.0939	1.14456755	0.1272	1.08103530	0.1201
λ_4	0.26538146	0.0295	0.52687903	0.0585	0.26966197	0.0300
λ_5	0.23842185	0.0265	0.44230571	0.0491	0.22517564	0.0250
Cumulative proportion for first 5 PCs		0.9652		0.9519		0.9614

For example, PC1 which explains 52.56% of variation of overall observations, explains 42.68% of the variation of the group of non-violators and 56.6% of variation of the group of violators. When we look at the eigenvectors for overall observations, the group of non-violators and the group of violators, they all differ as follows:

Eigenvectors for overall observations

Eigenvectors					
	Prin1	Prin2	Prin3	Prin4	Prin5
MW	0.401994	0.265940	-.032637	-.044052	0.001204
LogP	-.152728	0.549606	0.111241	0.369102	0.227459
LogD	-.184448	0.534108	0.086099	0.388775	-.023836
Hdonors	0.392157	-.204210	0.056273	0.220030	0.733720
Hacceptors	0.417390	-.069642	-.113475	0.545761	-.476305
PSA	0.429203	-.157166	-.015606	0.251764	-.108777
ROT	0.296647	0.197127	0.698575	-.327950	-.296005
NATOM	0.370392	0.319962	0.039111	-.304304	0.266056
NRING	0.208918	0.359791	-.687985	-.317603	-.110657

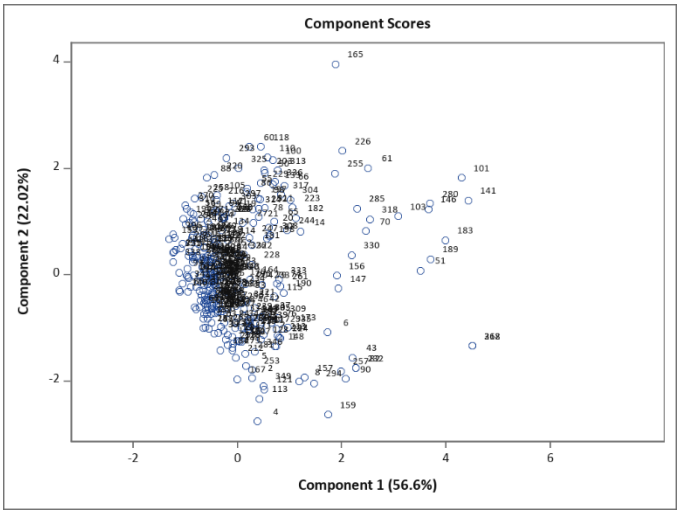
Eigenvectors for the group of non-violators

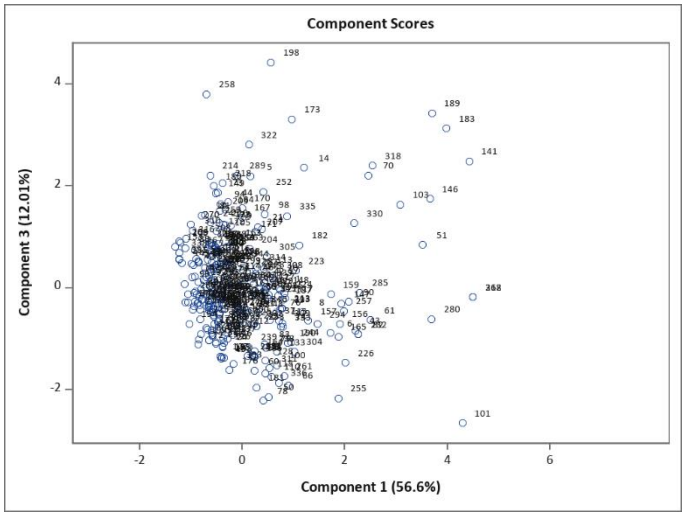
Eigenvectors					
	Prin1	Prin2	Prin3	Prin4	Prin5
MW	0.303393	0.455930	-.071376	-.022216	-.143829
LogP	0.463685	-.043977	0.077318	0.347666	0.371545
LogD	0.460213	-.071356	0.044655	0.390437	0.343032
Hdonors	-.315753	0.328388	0.048364	-.263936	0.804337
Hacceptors	-.201283	0.469904	-.045086	0.621293	-.172408
PSA	-.298101	0.452890	-.005871	0.181217	0.002785
ROT	0.177306	0.251398	0.760530	-.141065	-.199575
NATOM	0.373294	0.353997	0.042064	-.425815	-.020872
NRING	0.285112	0.250657	-.634312	-.190969	-.081493

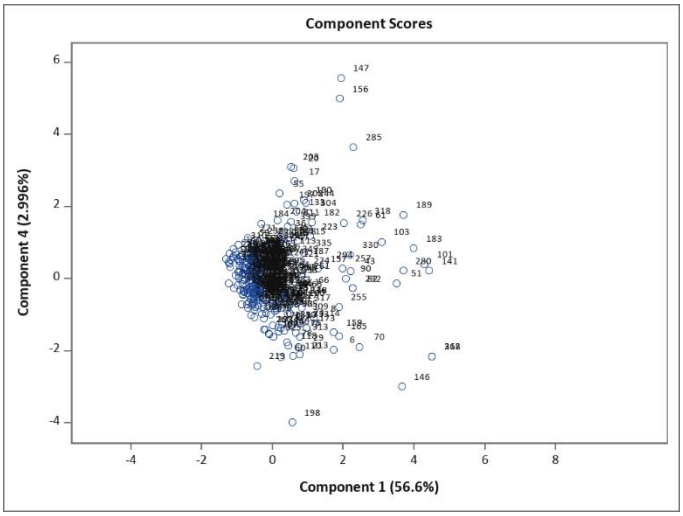
Eigenvectors for the group of violators

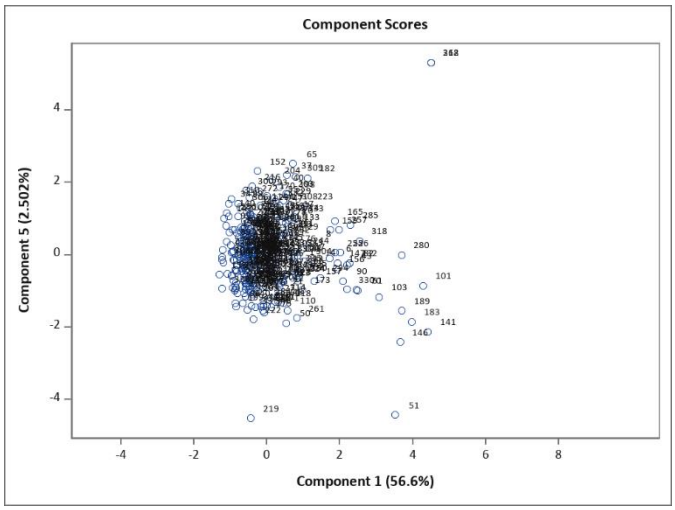
Eigenvectors					
	Prin1	Prin2	Prin3	Prin4	Prin5
MW	0.303393	0.455930	-.071376	-.022216	-.143829
LogP	0.463685	-.043977	0.077318	0.347666	0.371545
LogD	0.460213	-.071356	0.044655	0.390437	0.343032
Hdonors	-.315753	0.328388	0.048364	-.263936	0.804337
Hacceptors	-.201283	0.469904	-.045086	0.621293	-.172408
PSA	-.298101	0.452890	-.005871	0.181217	0.002785
ROT	0.177306	0.251398	0.760530	-.141065	-.199575
NATOM	0.373294	0.353997	0.042064	-.425815	-.020872
NRING	0.285112	0.250657	-.634312	-.190969	-.081493

We also found that PC1 is positively skewed with violators. All others PCs are normally distributed.









From discriminant analysis, with priors violators=0.3, non-violators=0.7 and unequal covariance matrices, using cross validation, 190 out of 1279 observations would be classified incorrectly. For non-violator (score9_logD_group_1 = 1), 65 out of 928 (7.00%) observations would be classified incorrectly. For violator (score9_logD_group_1 = 2), 125 out of 351 (35.61%) observations would be classified incorrectly.

Number of Observations and Percent Classified into score9_logD_group_1			
From score9_logD_group_1	1	2	Total
1	863 93.00	65 7.00	928 100.00
2	125 35.61	226 64.39	351 100.00
Total	988 77.25	291 22.75	1279 100.00
Priors	0.7	0.3	

Error Count Estimates for score9_logD_group_1			
	1	2	Total
Rate	0.0700	0.3561	0.1559
Priors	0.7000	0.3000	

In descending order, Hacceptors, LogP, ROT, NRING, PSA, LogD and Hdonors are all best to discriminate the 4 oral by violatory classes. (oral_violator, oral_non-violator, non-oral_violator, non-oral_non-violator)

Stepwise Selection Summary										
Step	Number In	Entered	Removed	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Squared Canonical Correlation	Pr > ASCC
1	1	MW		0.4421	336.85	<.0001	0.55785055	<.0001	0.14738315	<.0001
2	2	Hacceptors		0.0973	45.77	<.0001	0.50357785	<.0001	0.16679607	<.0001
3	3	LogP		0.0998	47.02	<.0001	0.45334506	<.0001	0.19666926	<.0001
4	4	ROT		0.0288	12.57	<.0001	0.44028838	<.0001	0.20196915	<.0001
5	5	NRING		0.0453	20.10	<.0001	0.42034939	<.0001	0.20957329	<.0001
6	6	PSA		0.0281	12.24	<.0001	0.40853831	<.0001	0.21790399	<.0001
7	7	LogD		0.0106	4.55	0.0035	0.40418761	<.0001	0.22116106	<.0001
8	8	Hdonors		0.0097	4.15	0.0061	0.40025519	<.0001	0.22393663	<.0001
9	7		MW	0.0026	1.09	0.3517	0.40128866	<.0001	0.22317131	<.0001

Reference

[R1]

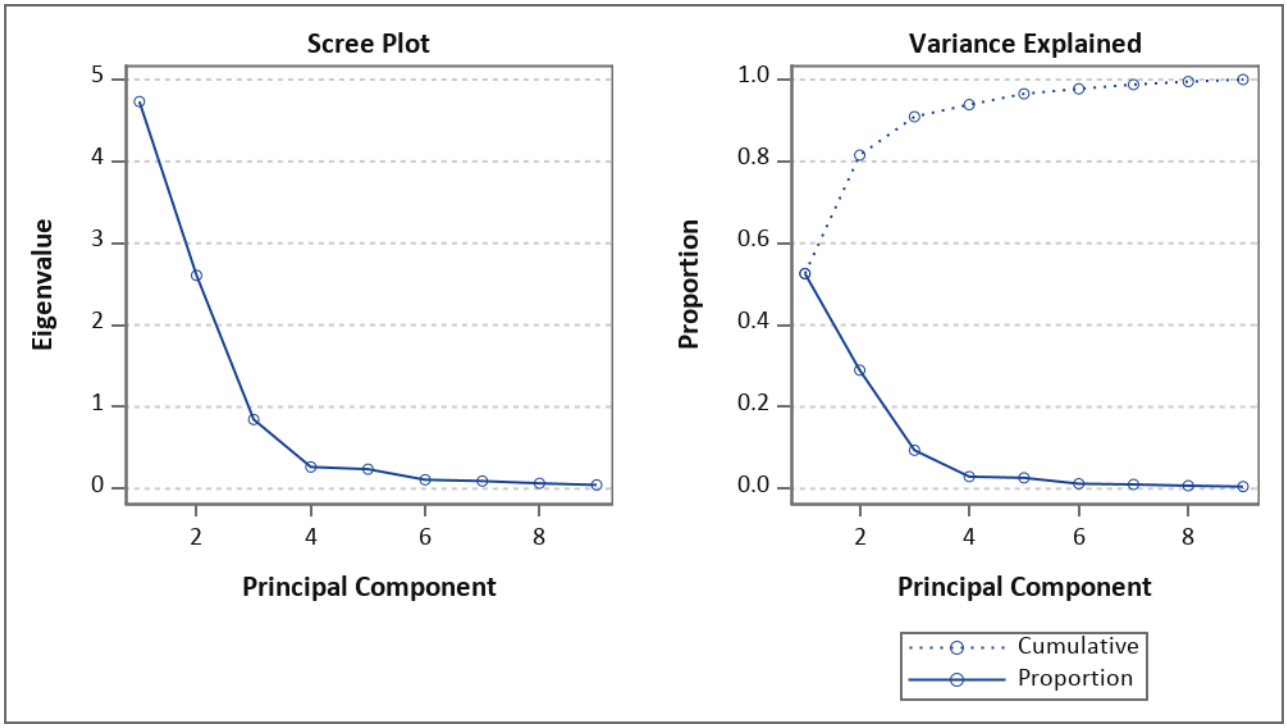
Dimension reduction: Guidelines for retaining principal components

http://proc-x.com/2017/08/dimension-reduction-guidelines-for-retaining-principal-components/. Accessed on 04-06-2020.

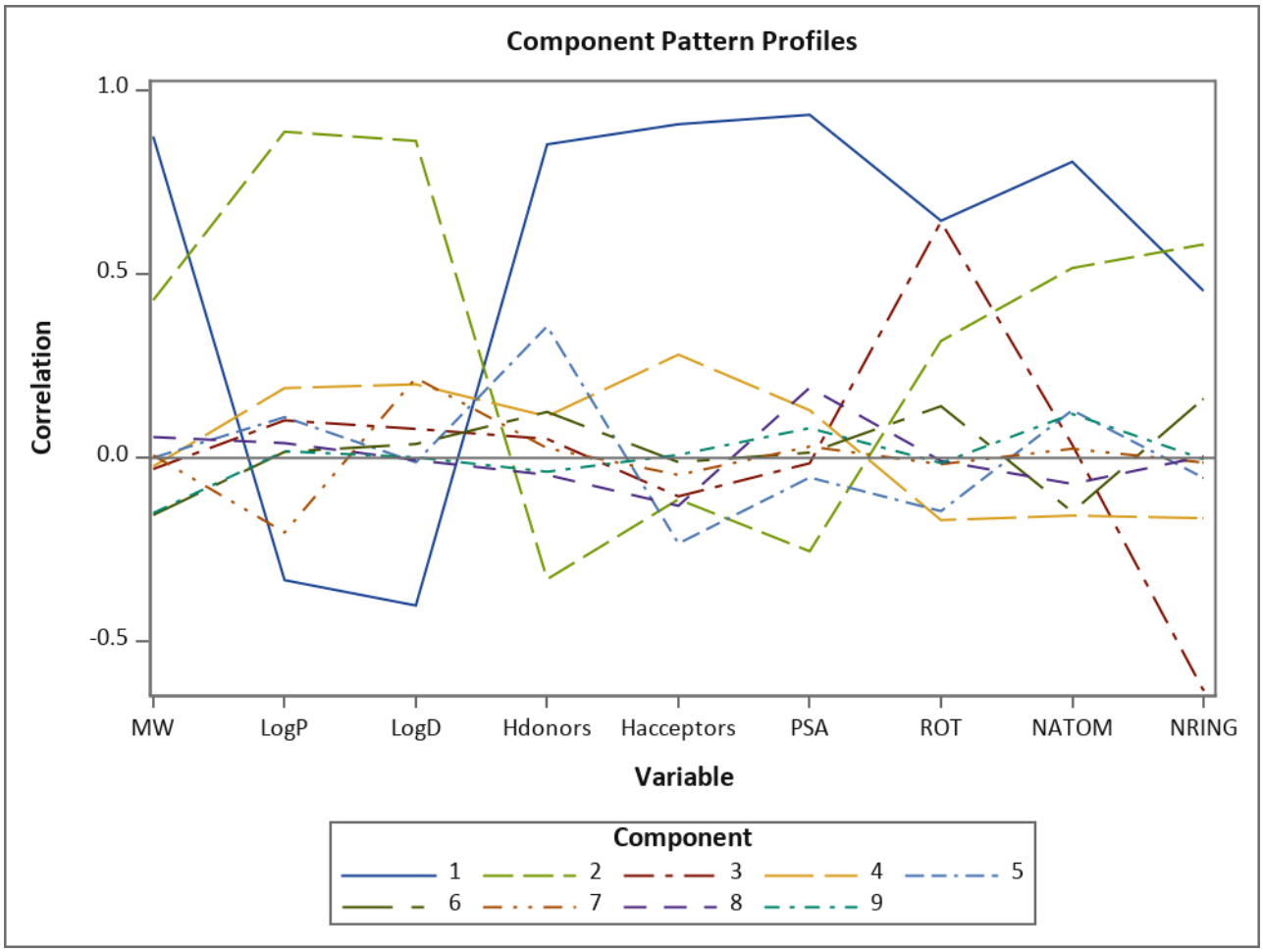
Appendix

[A1]

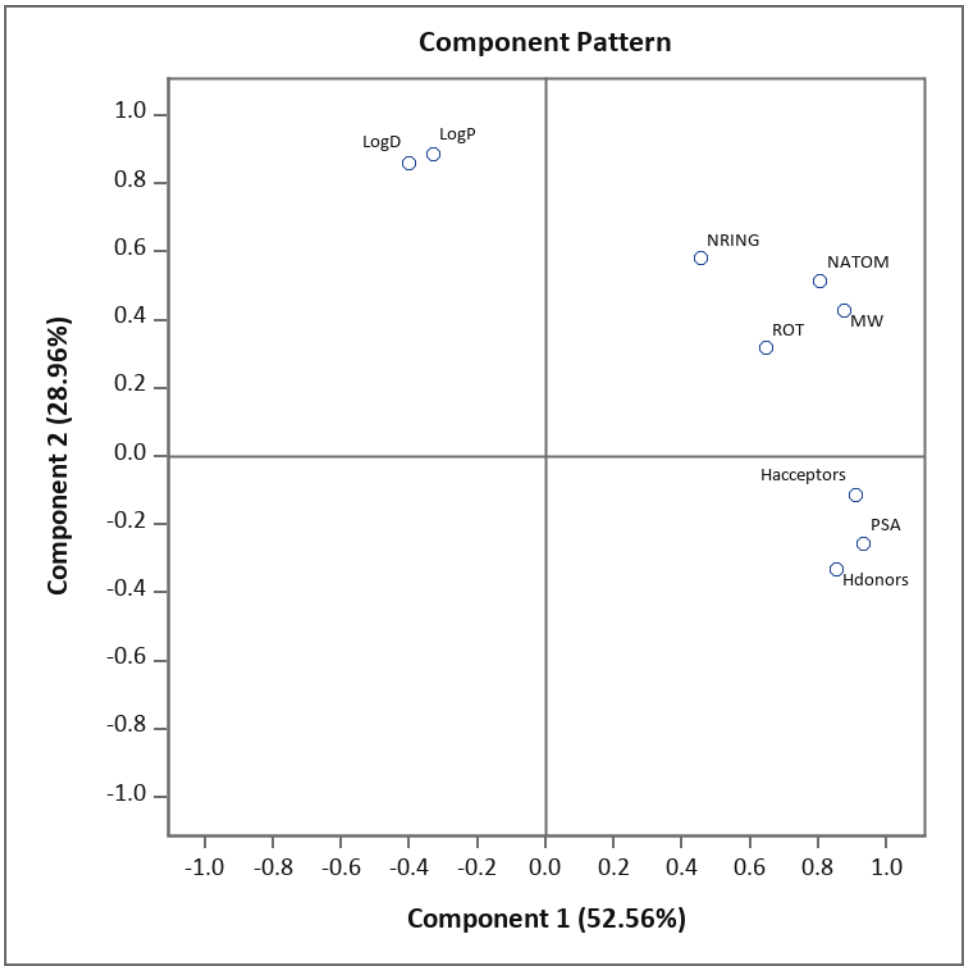
1. Scree plot for all observations

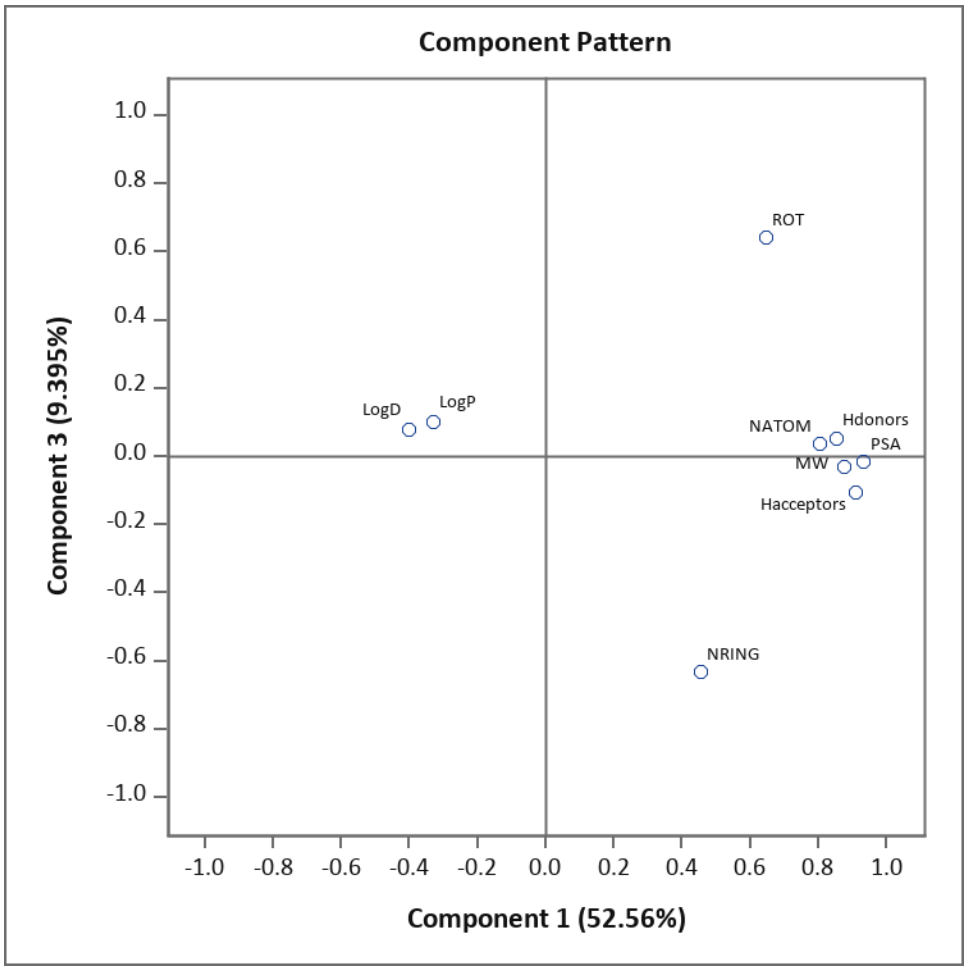


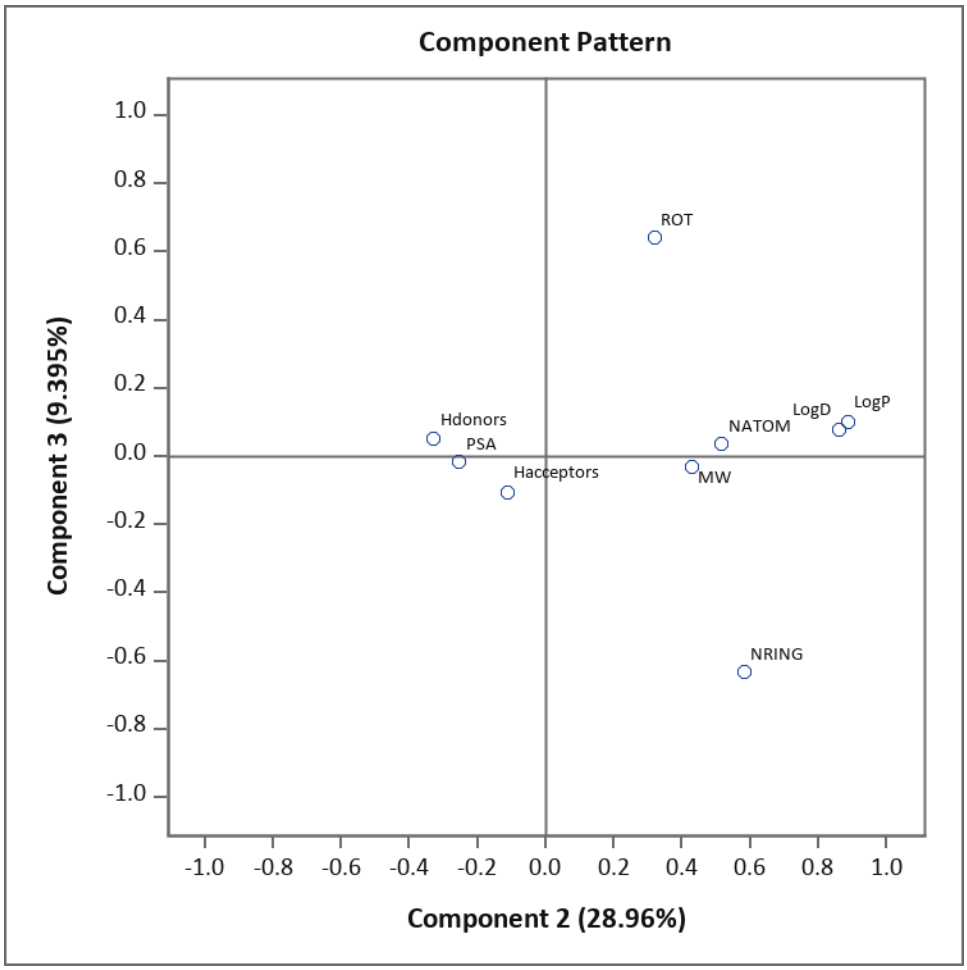
2. Profile plot for all observations

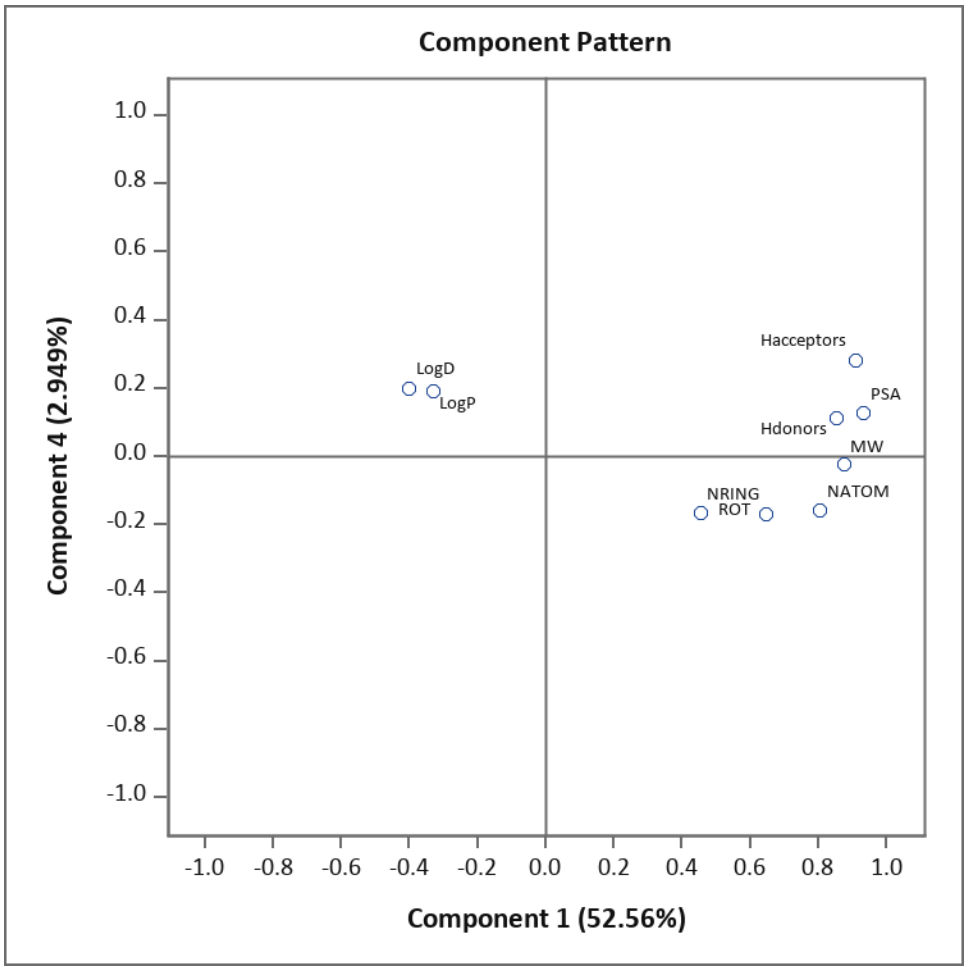


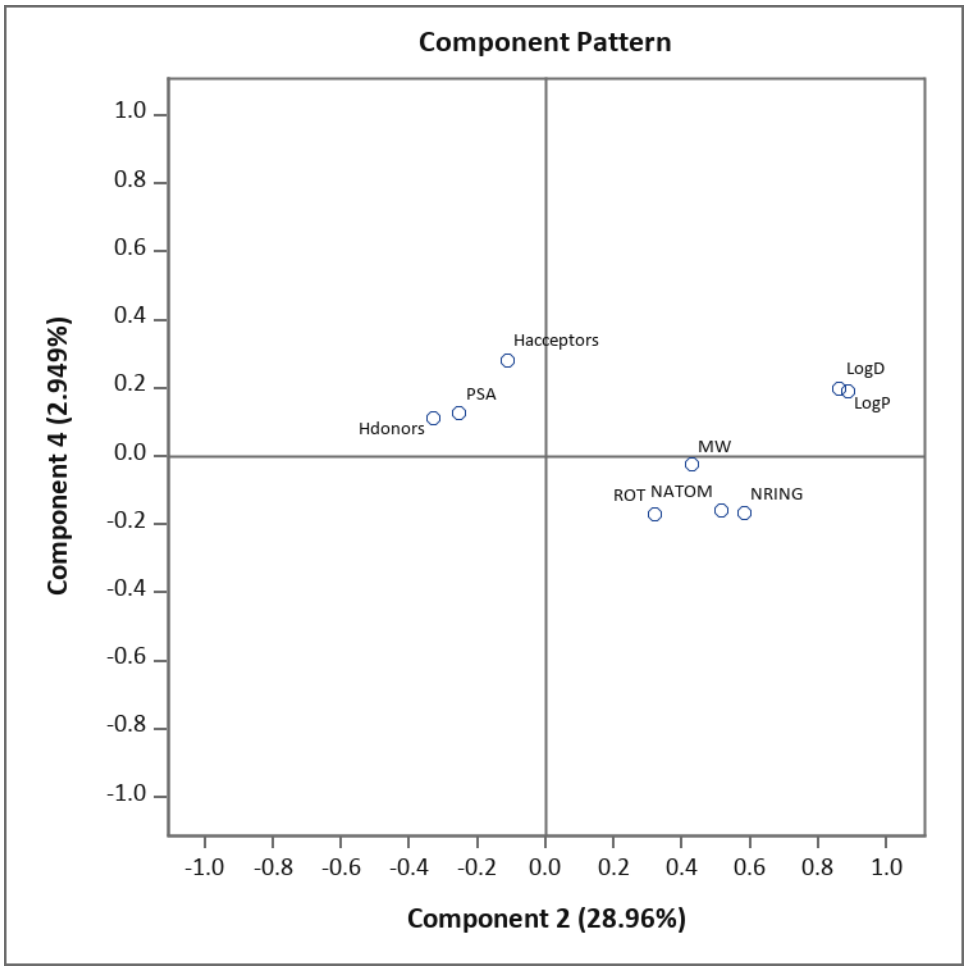
3. Component Pattern plots for all observations

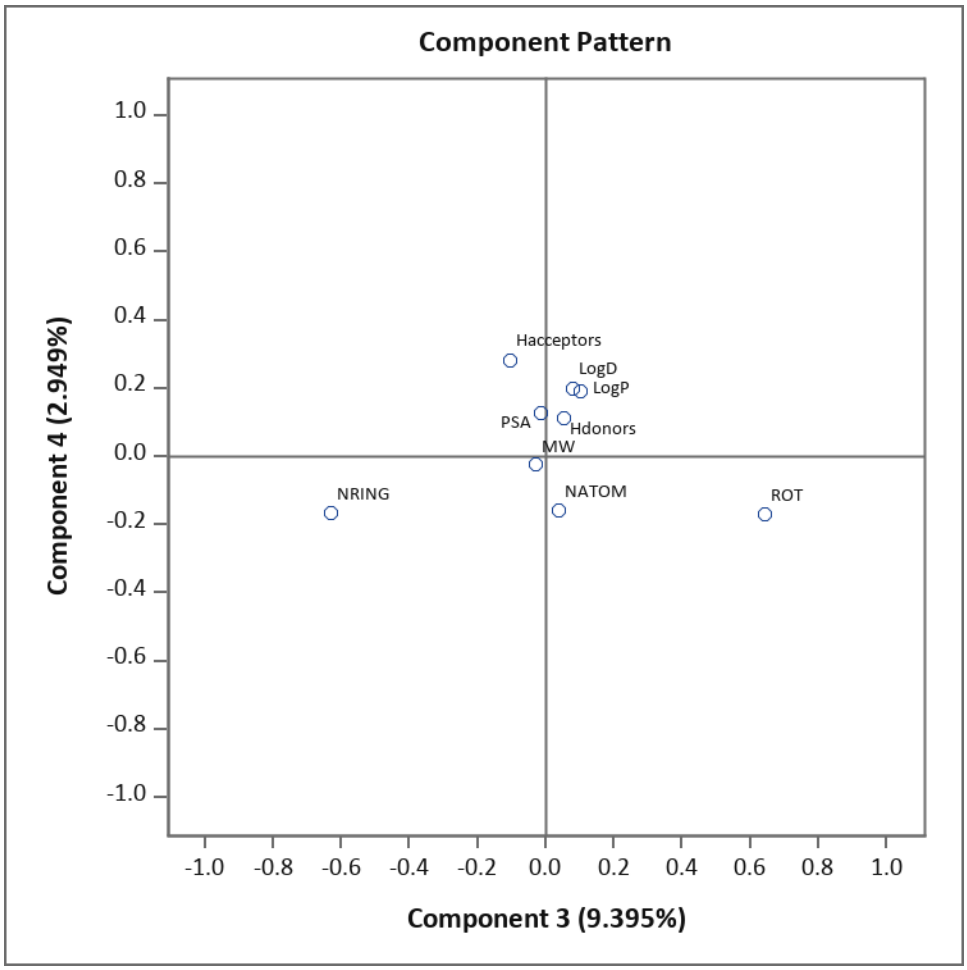


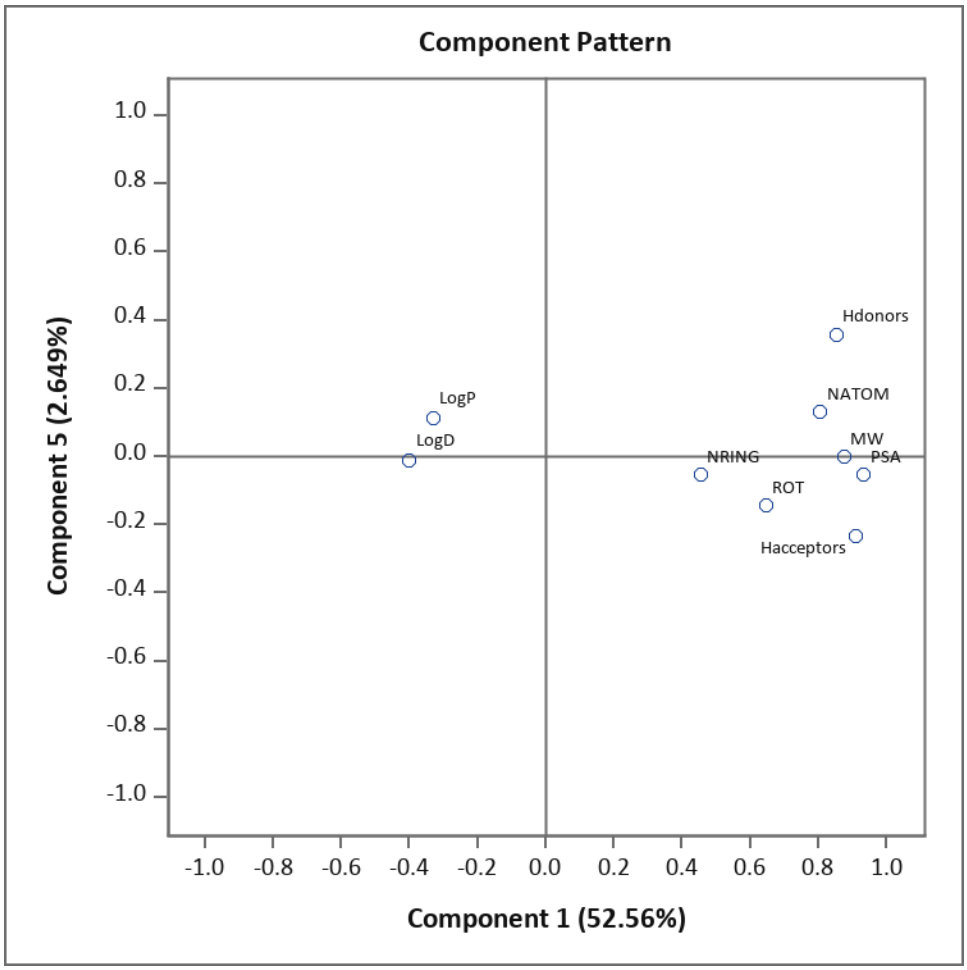

Component Pattern
Component 3 (9.395%)
Component 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
ROT
LogD
LogP
NATOM
Hdonors
PSA
MW
Hacceptors
NRING

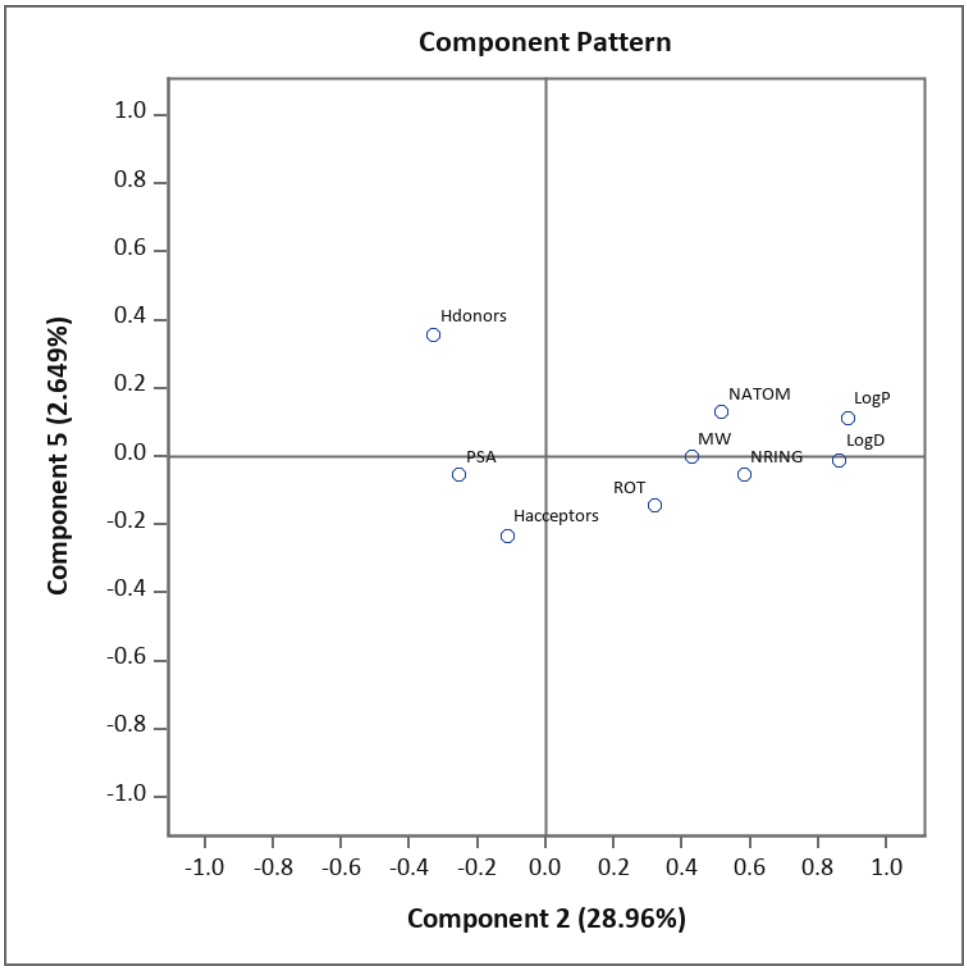

Component Pattern
Component 3 (9.395%)
Component 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
ROT
Hdonors
PSA
Hacceptors
NATOM
LogD
LogP
MW
NRING

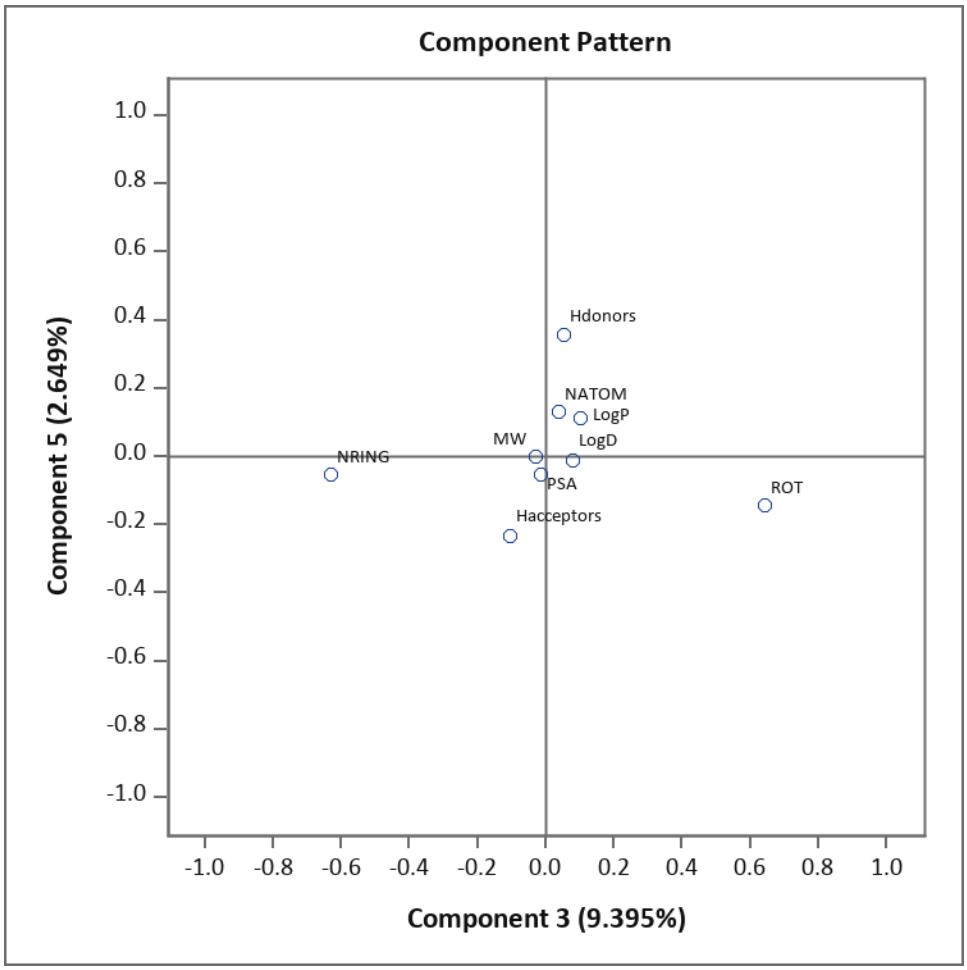

Component Pattern
Component 4 (2.949%)
Component 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
LogP
Hacceptors
PSA
Hdonors
MW
NRING
ROT
NATOM

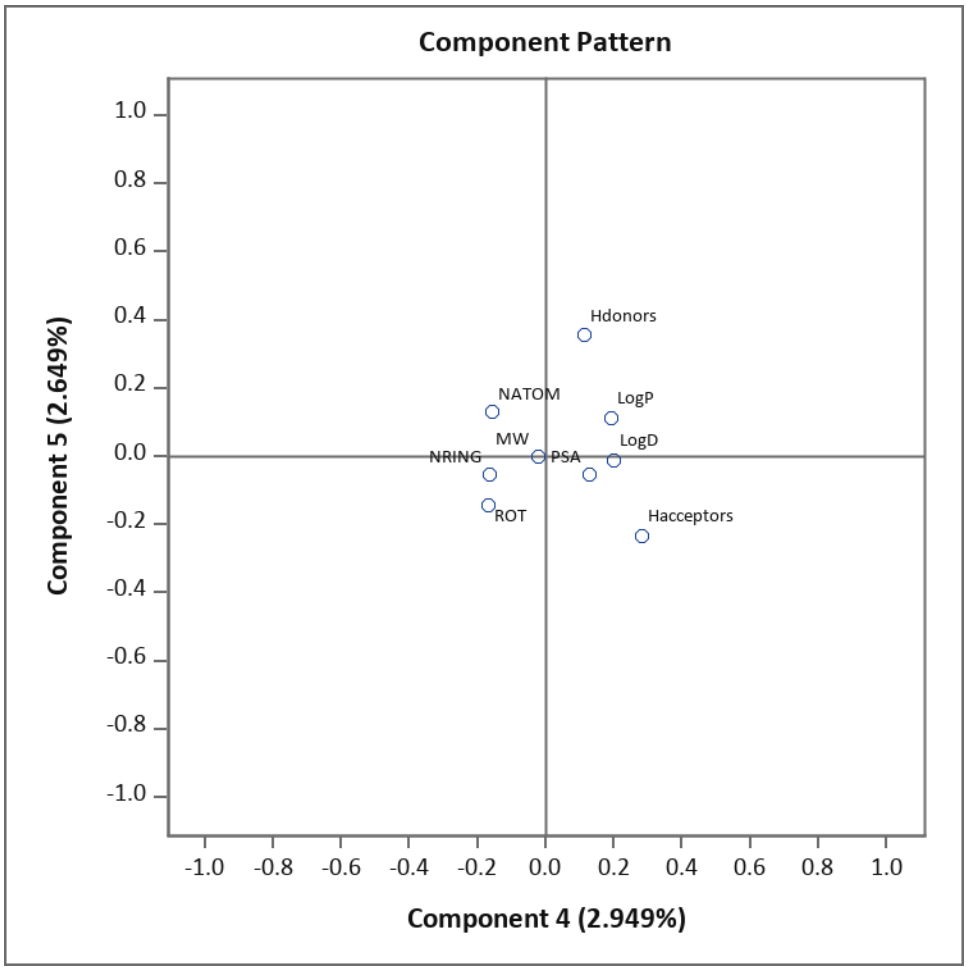

Component Pattern
Component 4 (2.949%)
Component 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
PSA
Hdonors
LogD
LogP
MW
ROT
NATOM
NRING

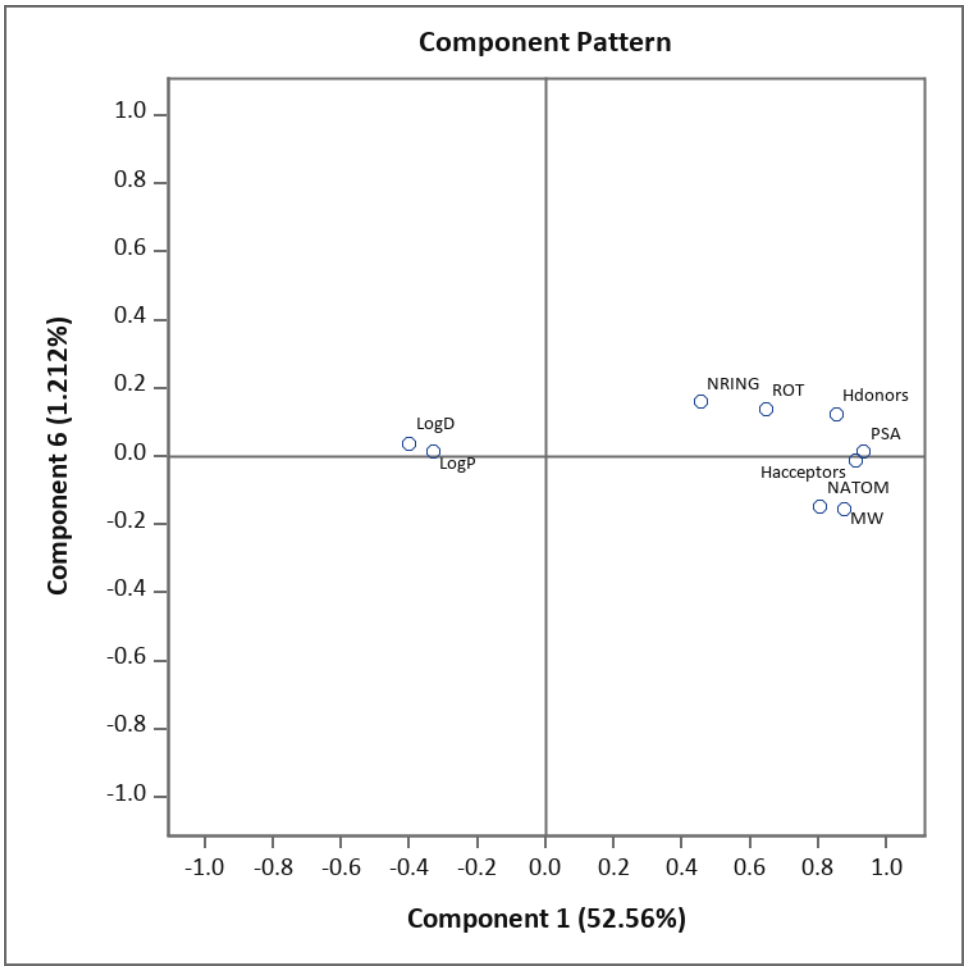

Component Pattern
Component 4 (2.949%)
Component 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
LogD
LogP
PSA
Hdonors
MW
NRING
NATOM
ROT

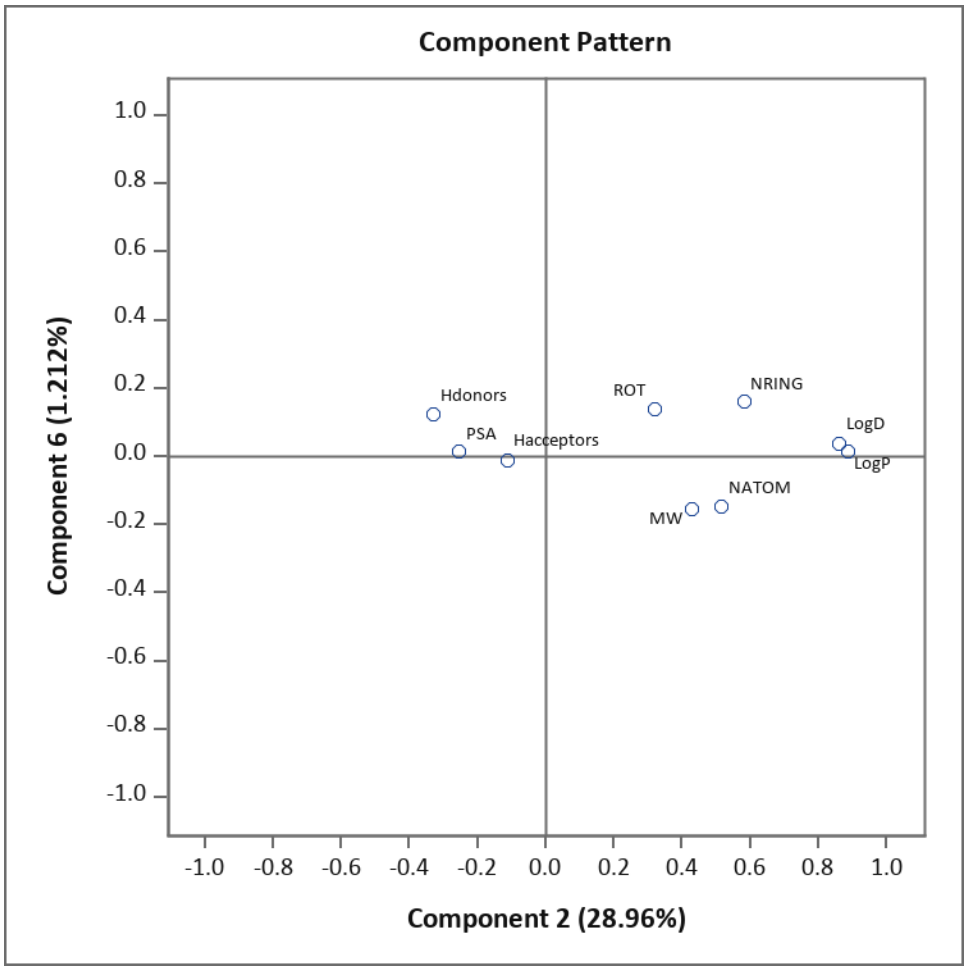

Component Pattern
Component 5 (2.649%)
Component 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
NATOM
LogP
LogD
MW
NRING
PSA
ROT
Hacceptors

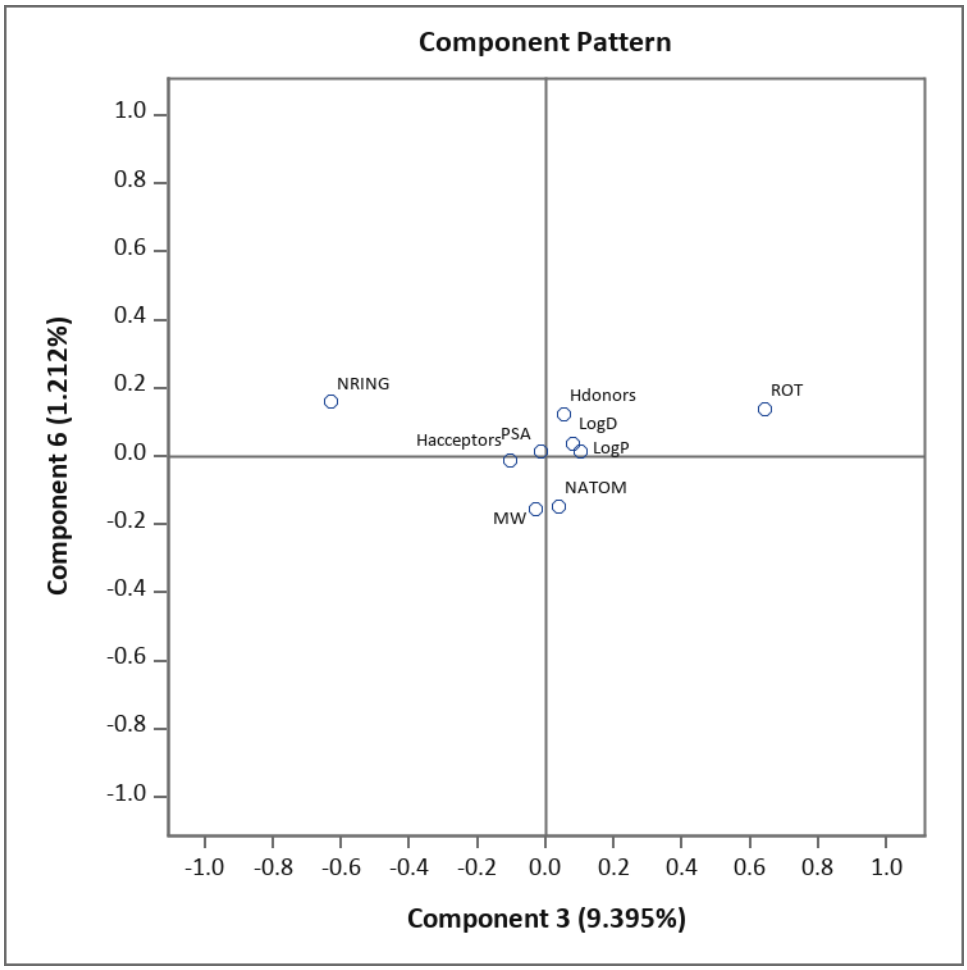

Component Pattern
Component 5 (2.649%)
Component 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
NATOM
LogP
MW
LogD
PSA
NRING
ROT
Hacceptors

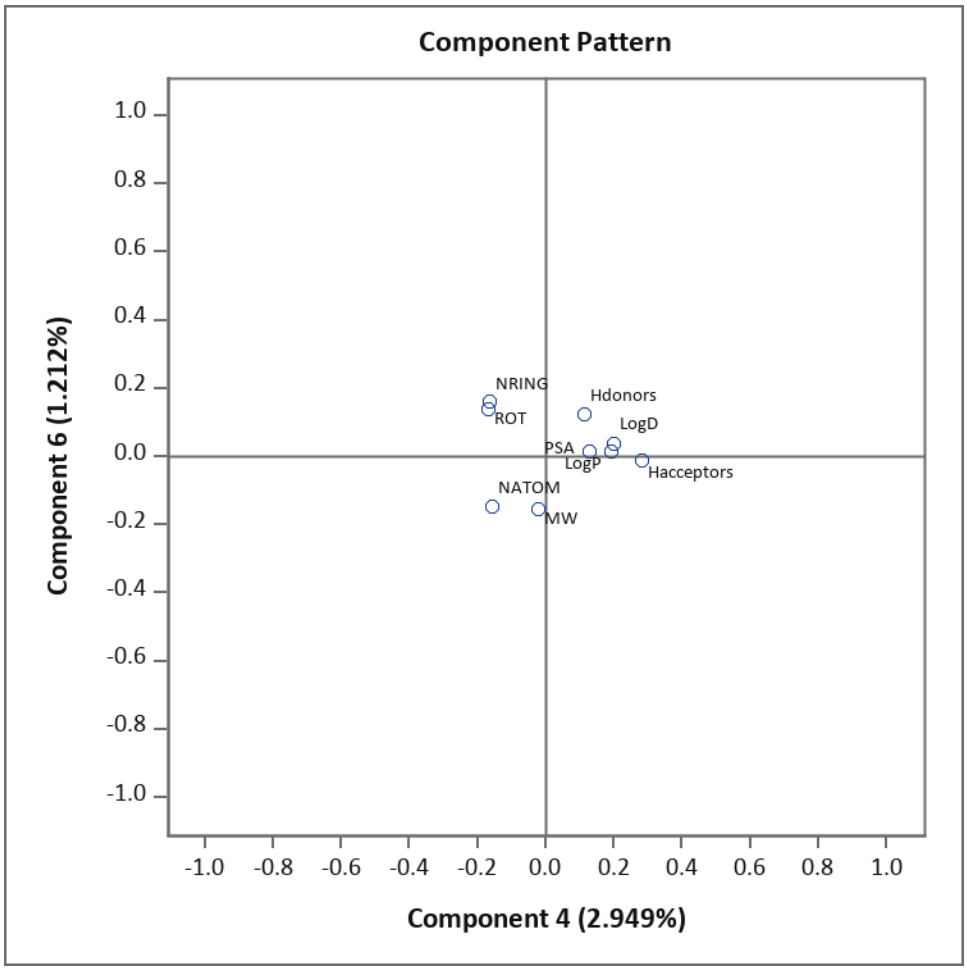

Component Pattern
Component 5 (2.649%)
Component 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
NATOM
LogP
LogD
MW
NRING
PSA
ROT
Hacceptors

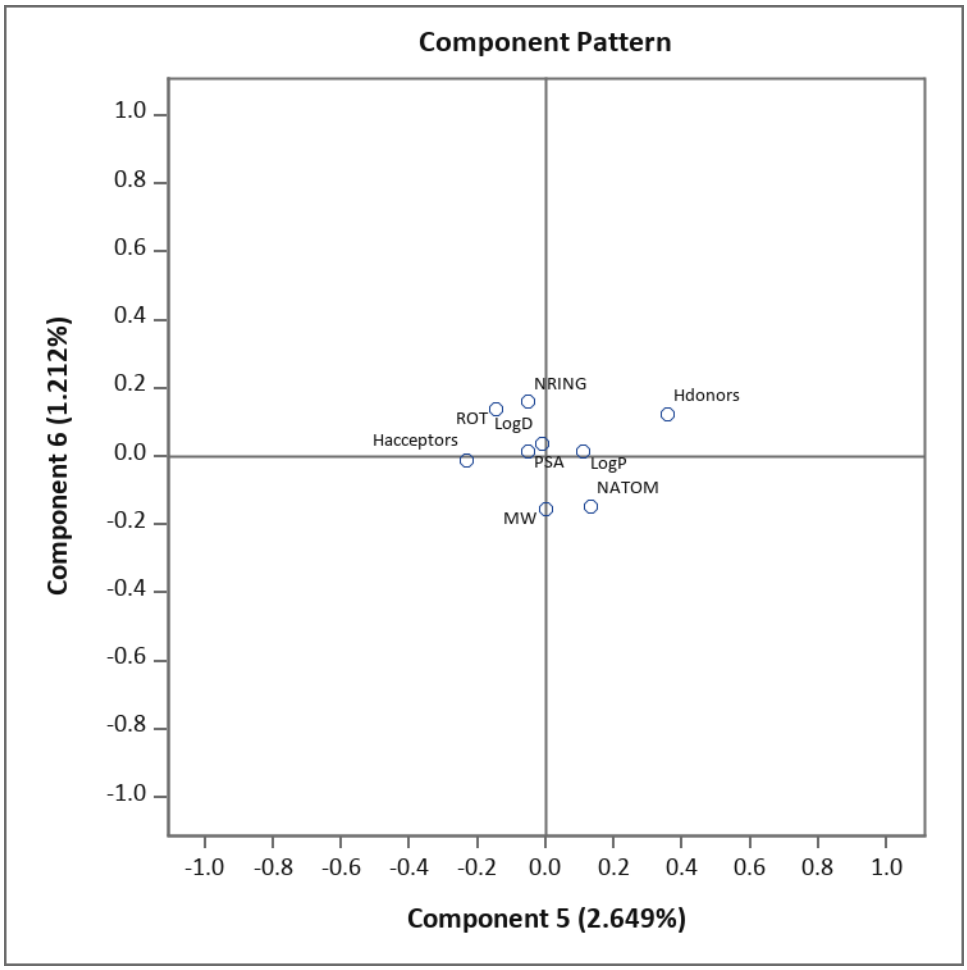

Component Pattern
Component 5 (2.649%)
Component 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
NATOM
LogP
MW
LogD
NRING
PSA
ROT
Hacceptors

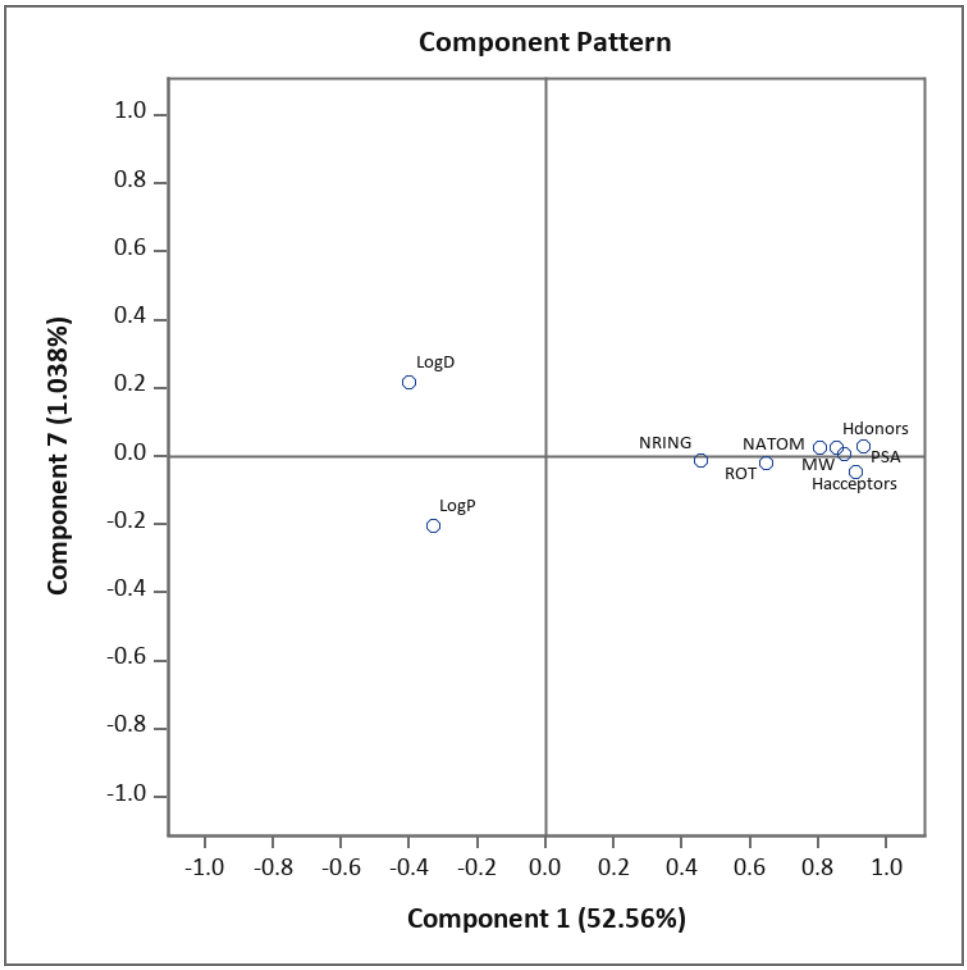

Component Pattern
Component 6 (1.212%)
Component 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NRING
ROT
Hdonors
PSA
LogD
LogP
Hacceptors
NATOM
MW

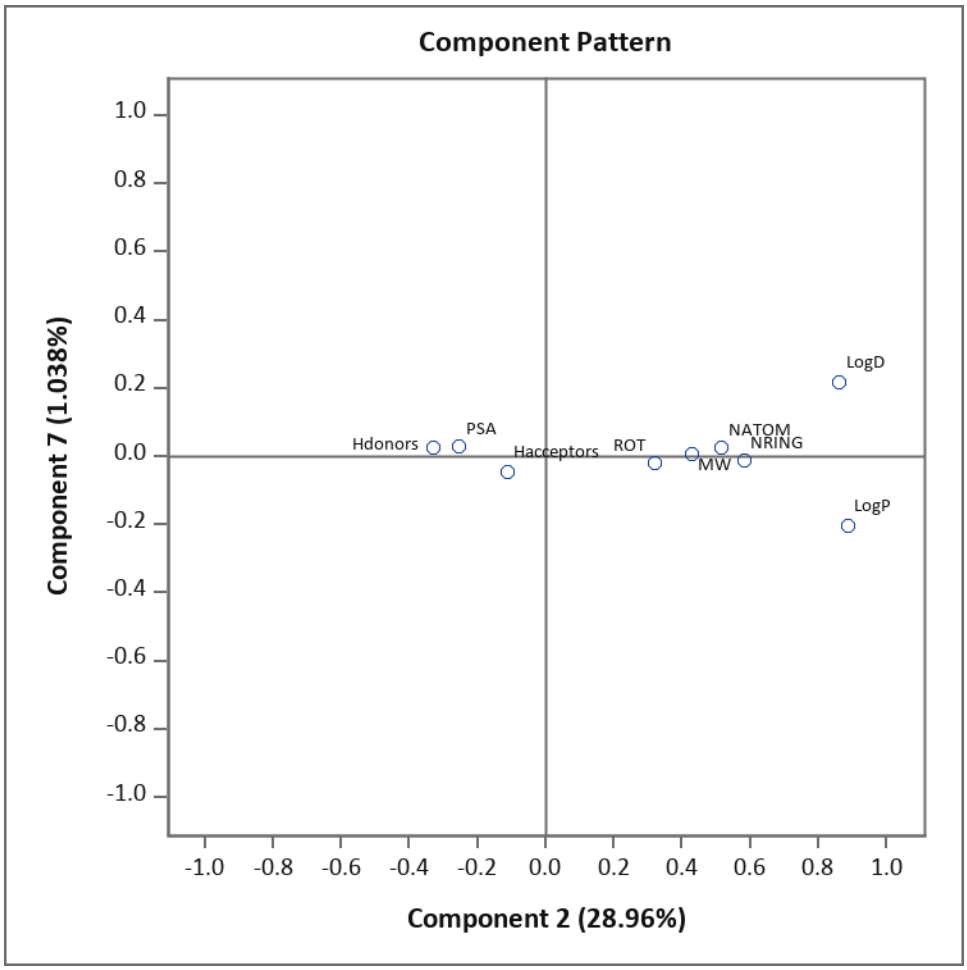

Component Pattern
Component 6 (1.212%)
Component 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
PSA
Hacceptors
ROT
NRING
LogD
LogP
NATOM
MW

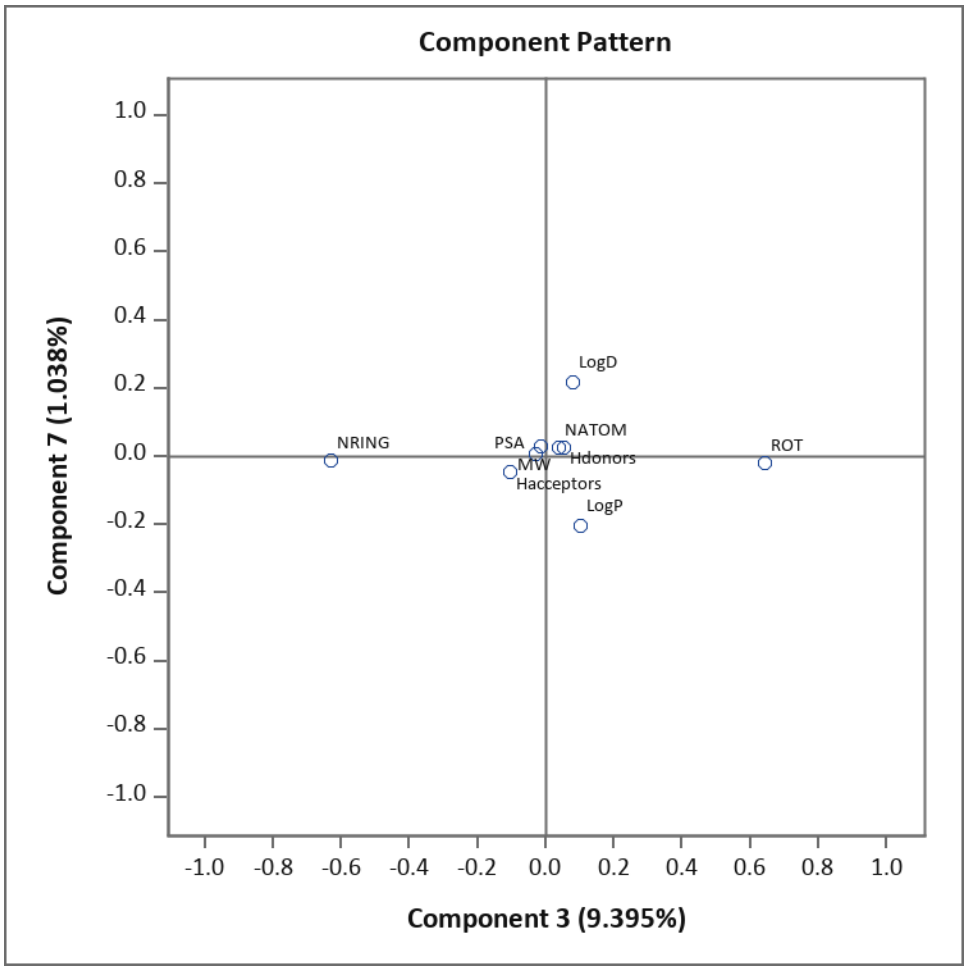

Component Pattern
Component 6 (1.212%)
Component 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NRING
Hdonors
ROT
LogD
Hacceptors
PSA
LogP
NATOM
MW

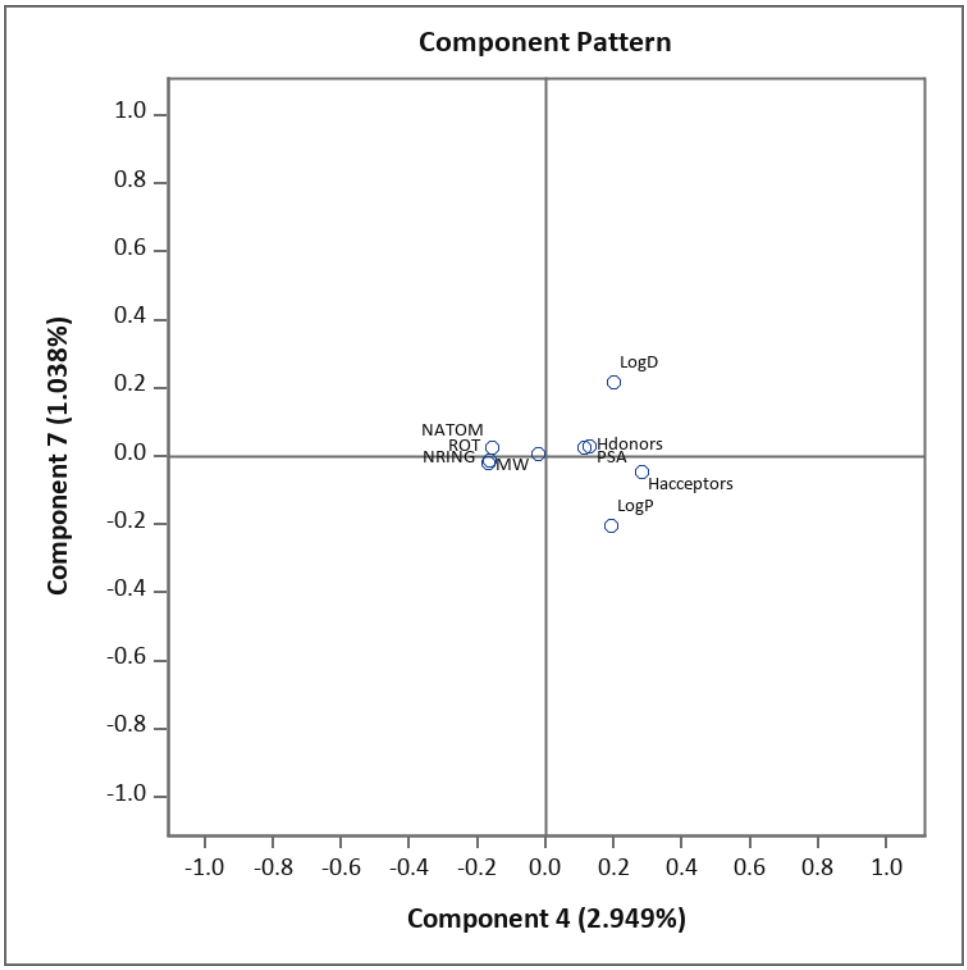

Component Pattern
Component 6 (1.212%)
Component 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NRING
ROT
Hdonors
LogD
PSA
LogP
Hacceptors
NATOM
MW

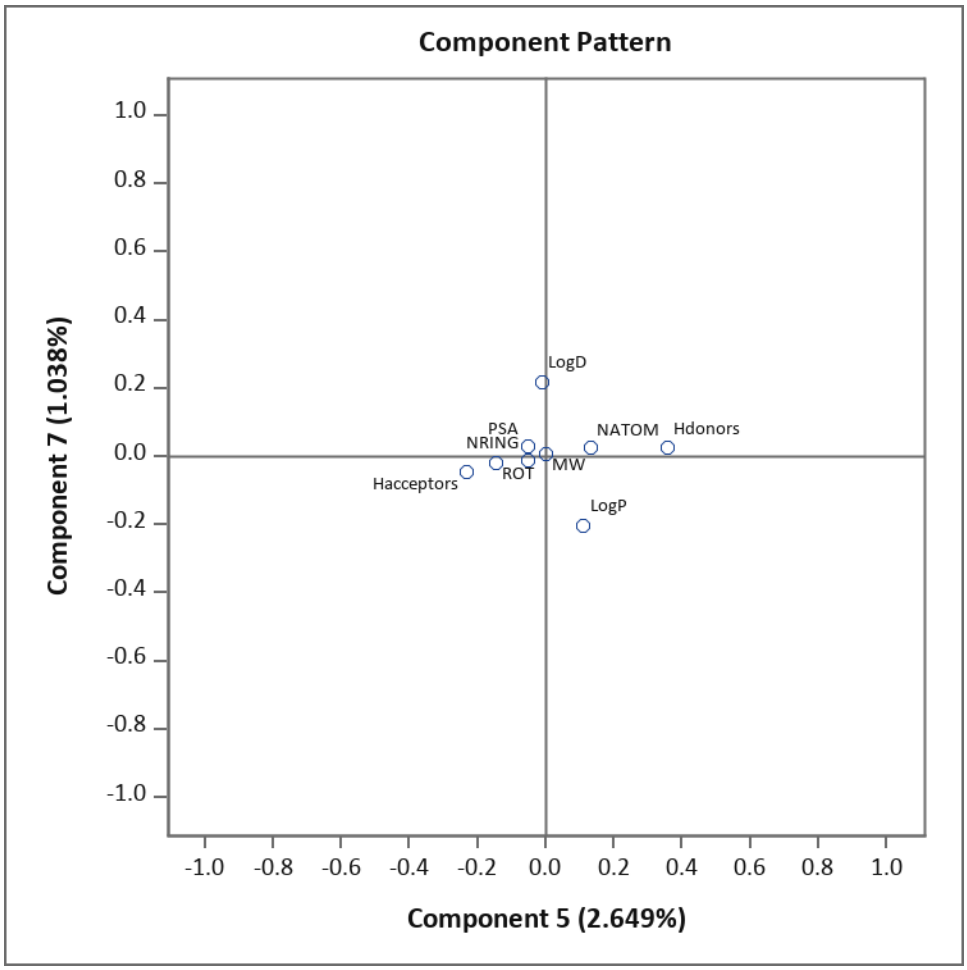

Component Pattern
Component 6 (1.212%)
Component 5 (2.649%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NRING
Hdonors
ROT
LogD
Hacceptors
PSA
LogP
NATOM
MW

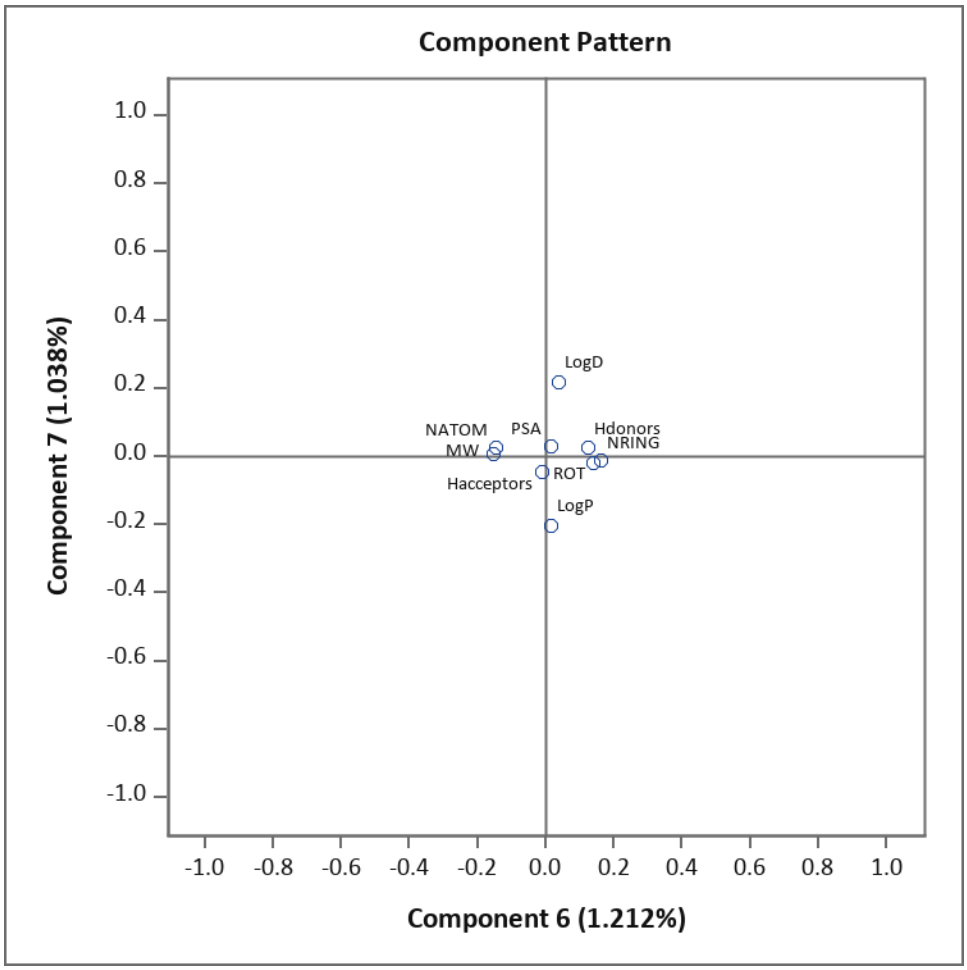

Component Pattern
Component 7 (1.038%)
Component 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
LogP
NRING
NATOM
ROT
MW
Hdonors
PSA
Hacceptors

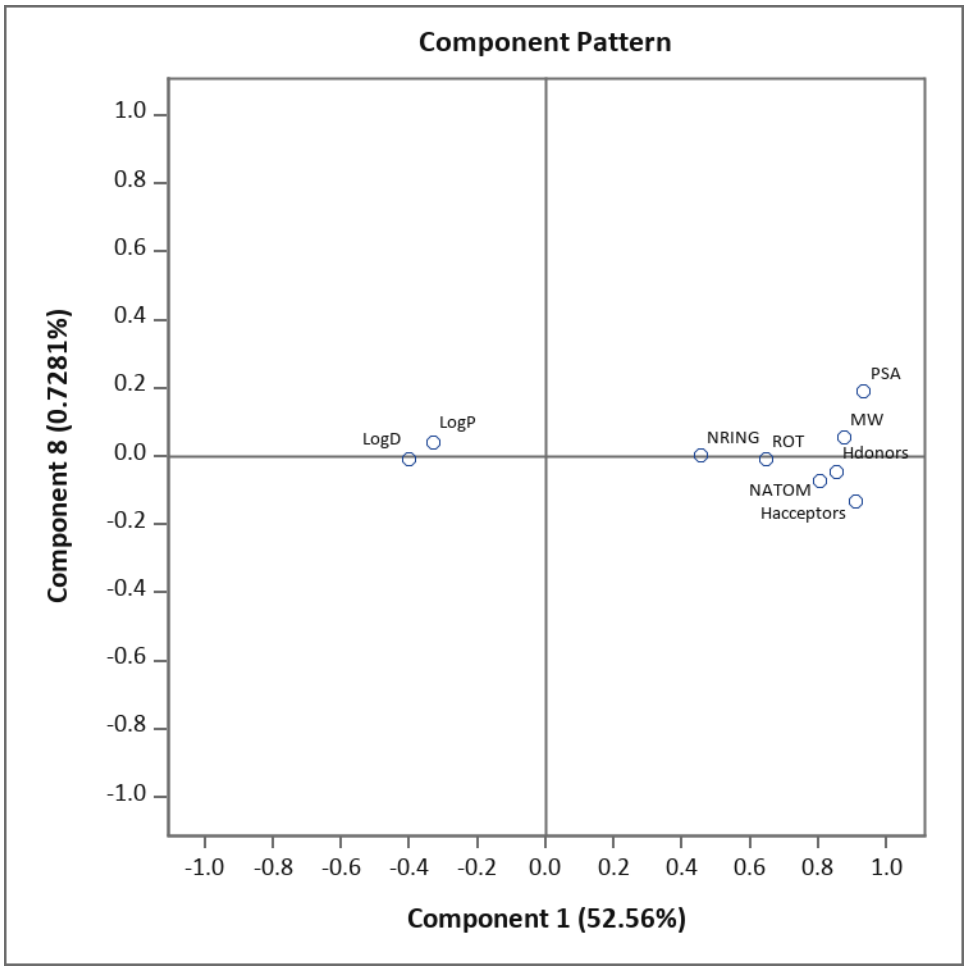

Component Pattern
Component 7 (1.038%)
Component 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
PSA
Hacceptors
ROT
MW
NATOM
NRING
LogD
LogP

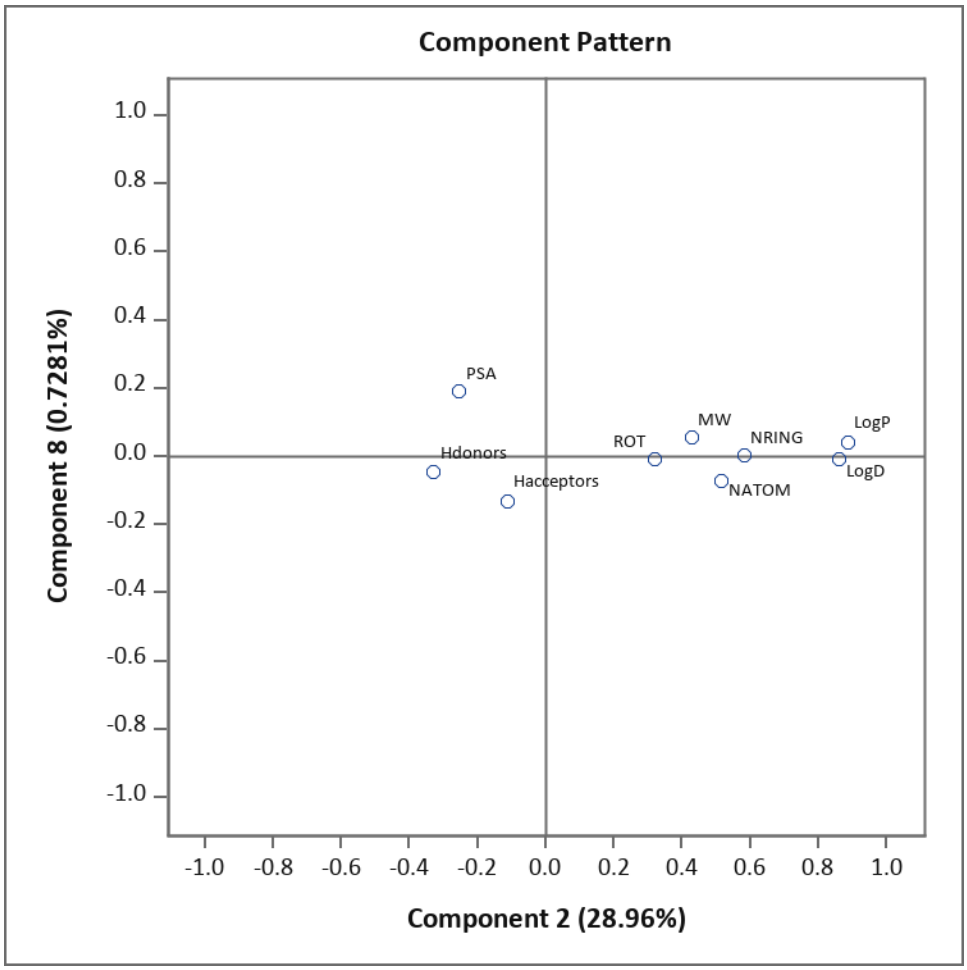

Component Pattern
Component 7 (1.038%)
Component 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
NRING
PSA
NATOM
ROT
Hdonors
MW
Hacceptors
LogP

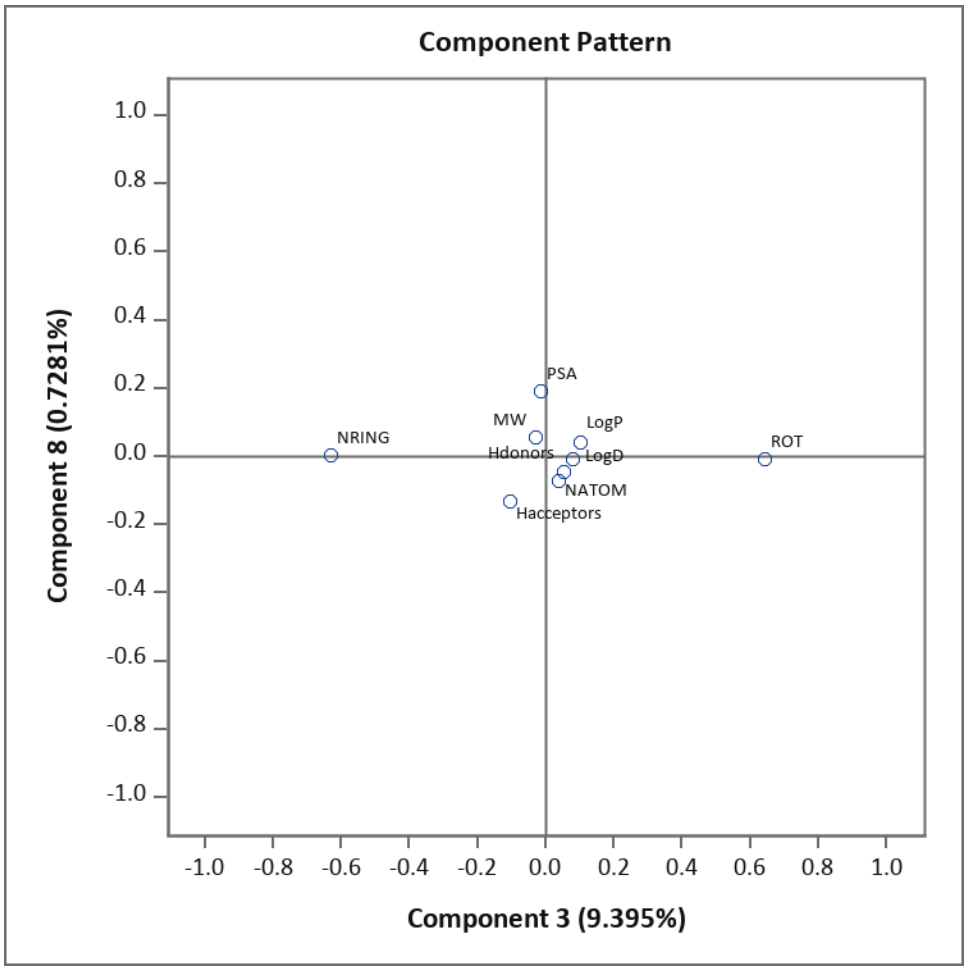

Component Pattern
Component 7 (1.038%)
Component 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
NATOM
ROT
NRING
MW
Hdonors
PSA
Hacceptors
LogP

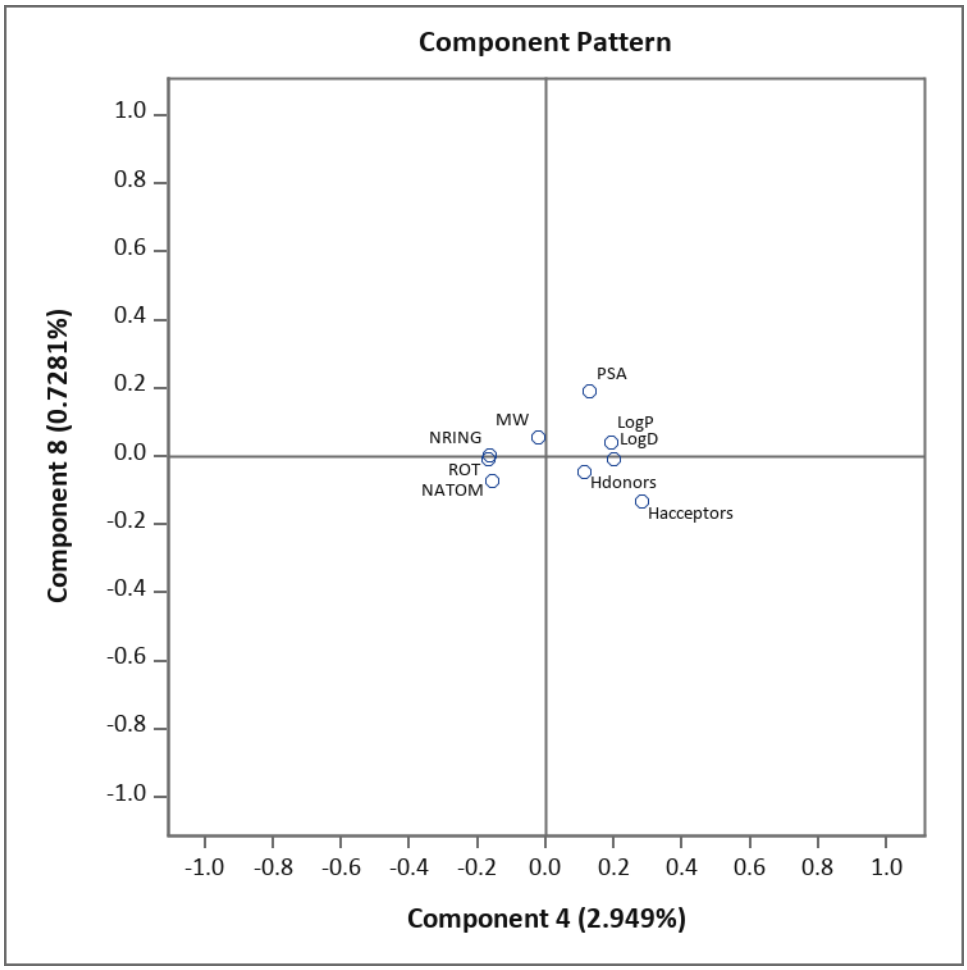

Component Pattern
Component 7 (1.038%)
Component 5 (2.649%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
PSA
NRING
NATOM
Hdonors
MW
ROT
Hacceptors
LogP

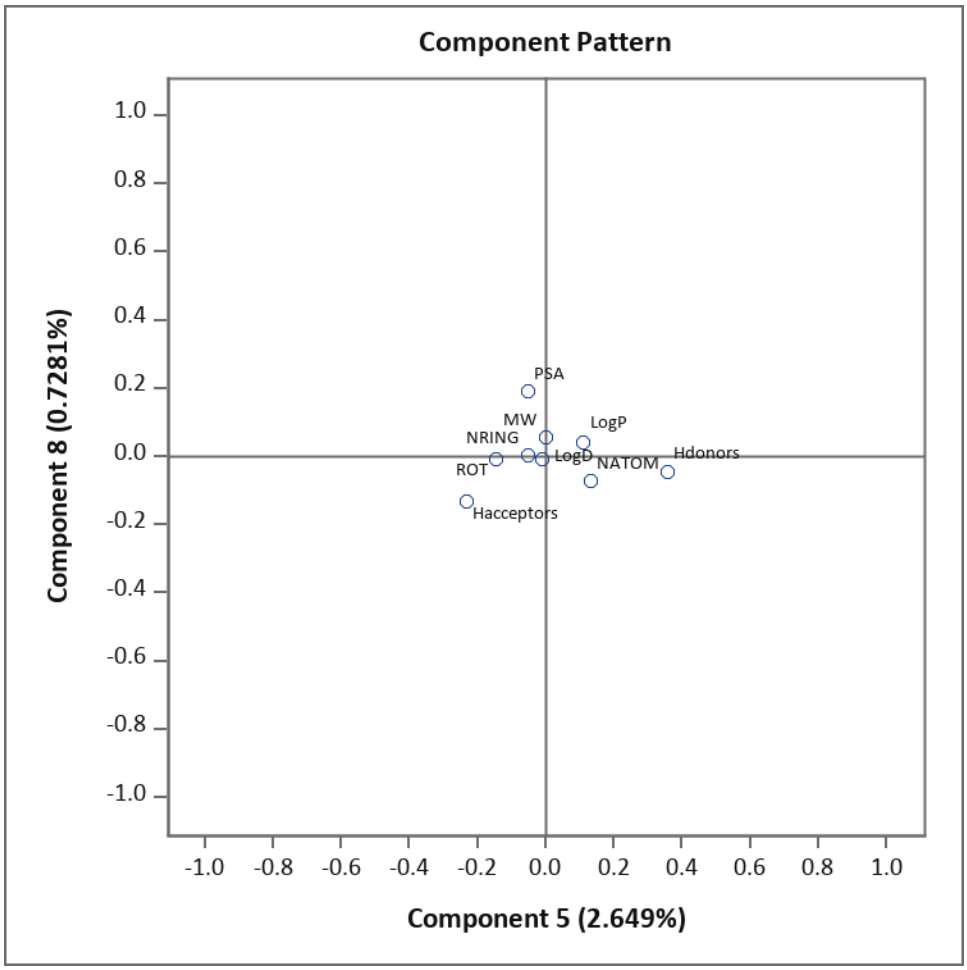

Component Pattern
Component 7 (1.038%)
Component 6 (1.212%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
NATOM
PSA
Hdonors
NRING
MW
ROT
Hacceptors
LogP

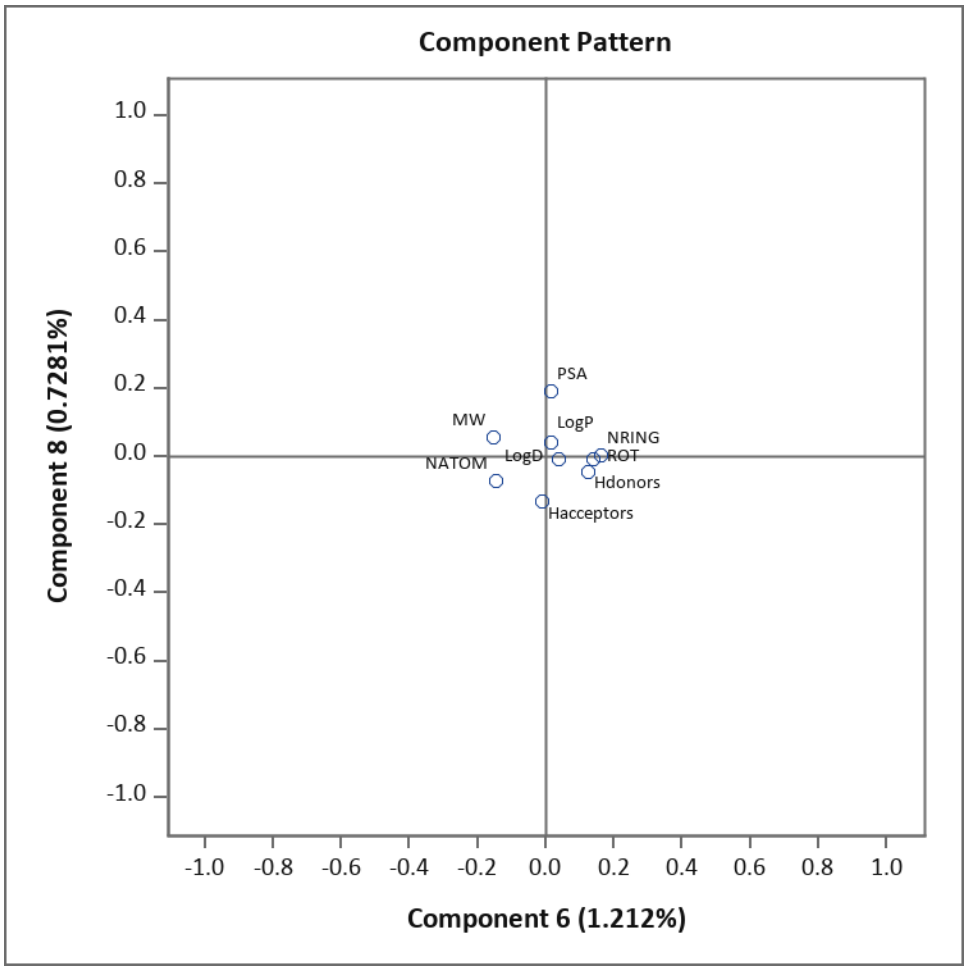

Component Pattern
Component 8 (0.7281%)
Component 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
LogP
NRING
ROT
PSA
MW
Hdonors
NATOM
Hacceptors

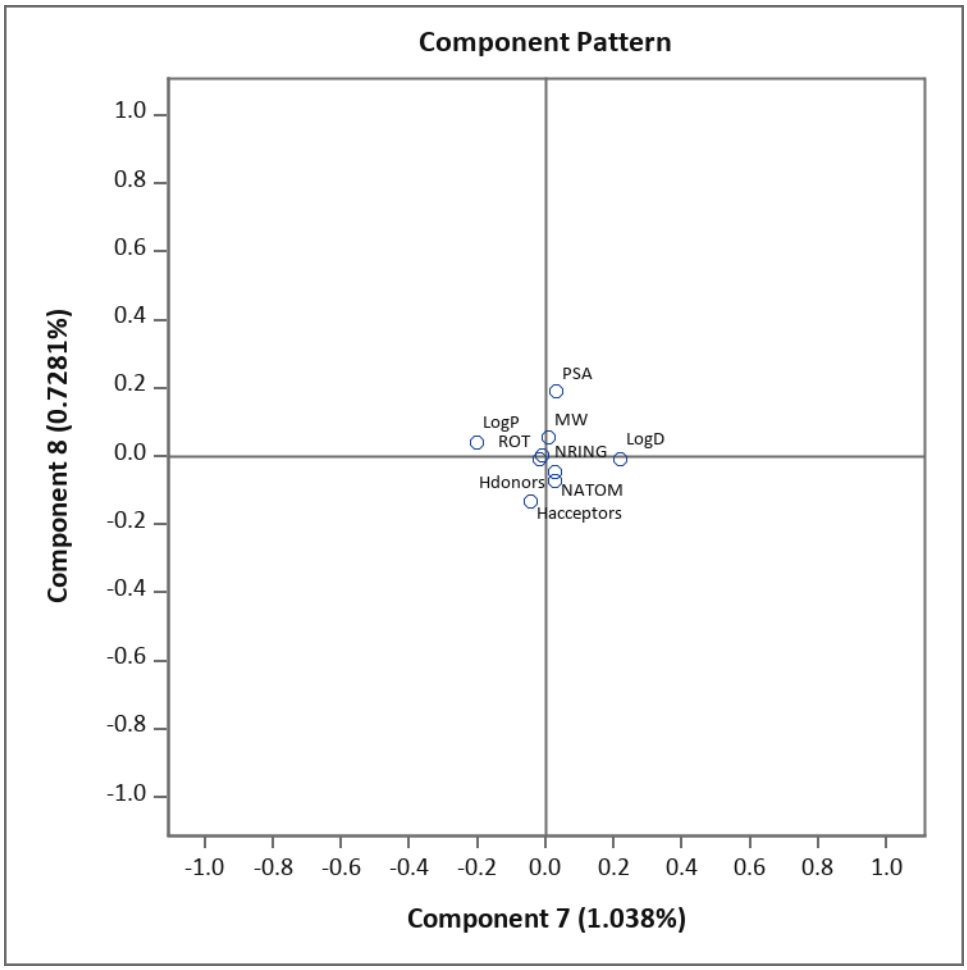

Component Pattern
Component 8 (0.7281%)
Component 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
Hdonors
Hacceptors
MW
ROT
NRING
NATOM
LogP
LogD

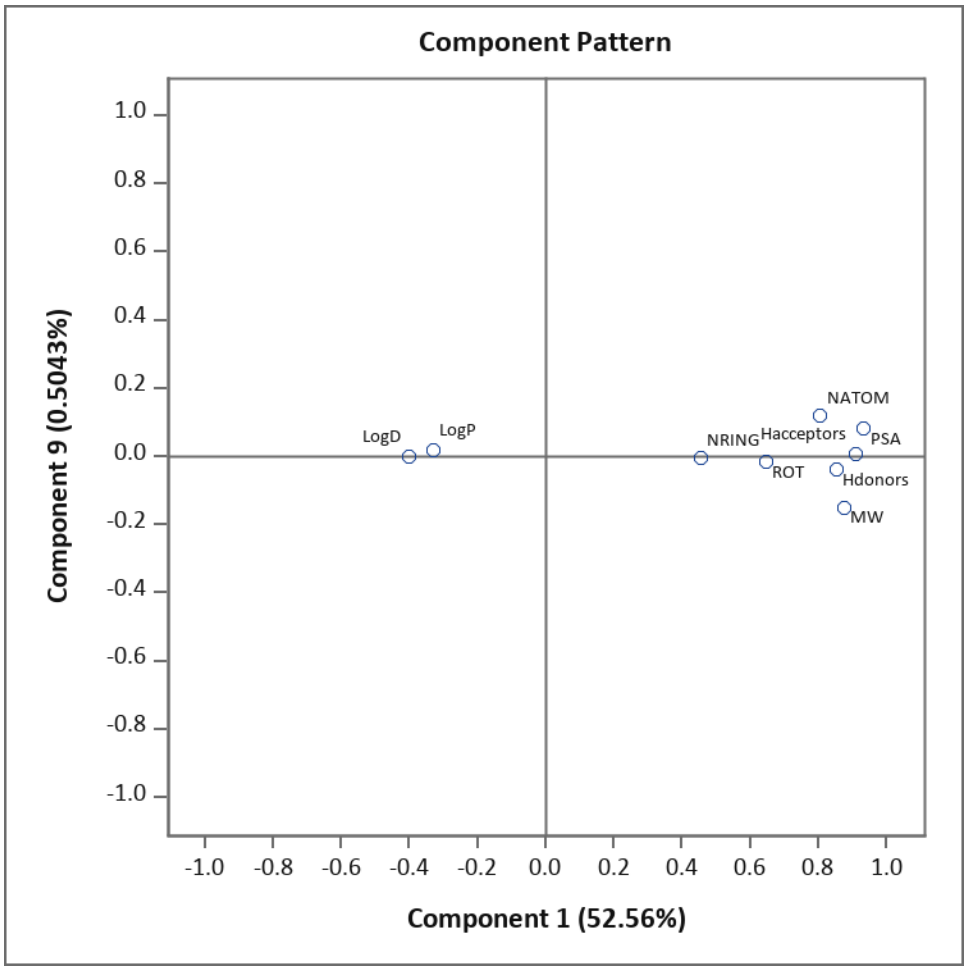

Component Pattern
Component 8 (0.7281%)
Component 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
MW
LogP
NRING
Hdonors
LogD
ROT
NATOM
Hacceptors

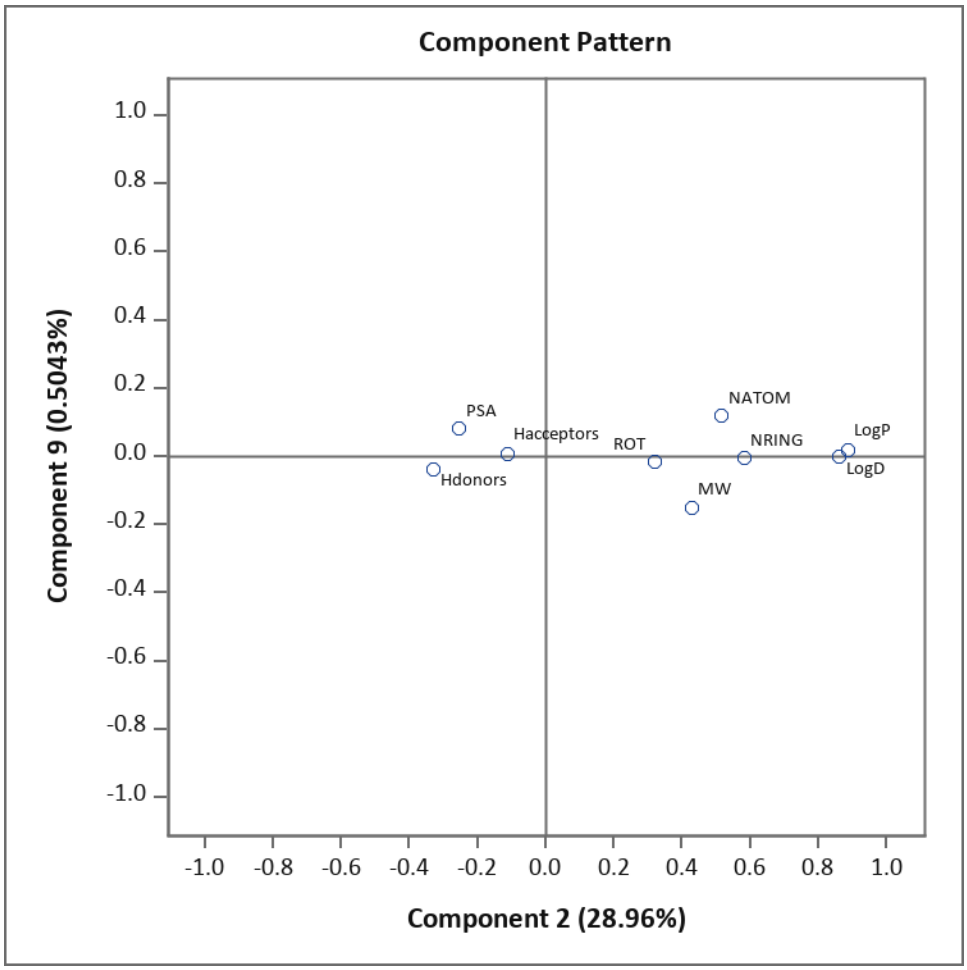

Component Pattern
Component 8 (0.7281%)
Component 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
MW
LogP
LogD
NRING
ROT
NATOM
Hdonors
Hacceptors

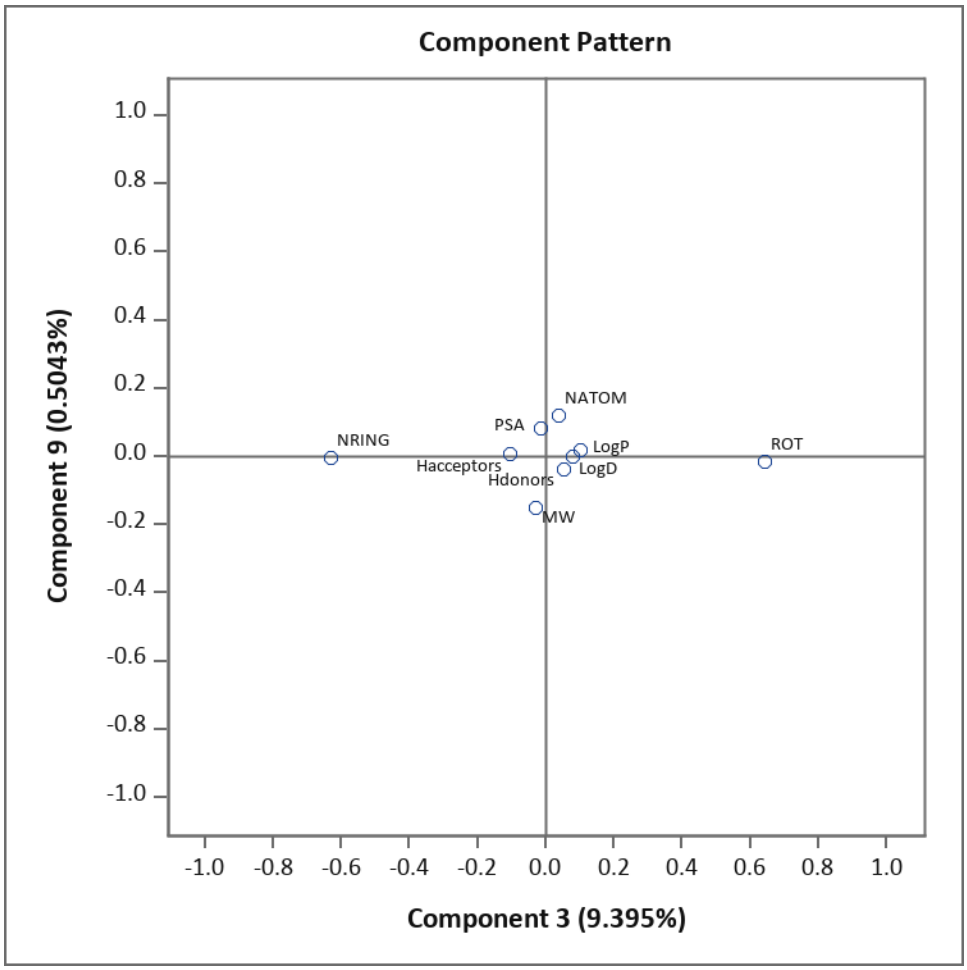

Component Pattern
Component 8 (0.7281%)
Component 5 (2.649%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
MW
LogP
NRING
LogD
NATOM
Hdonors
ROT
Hacceptors

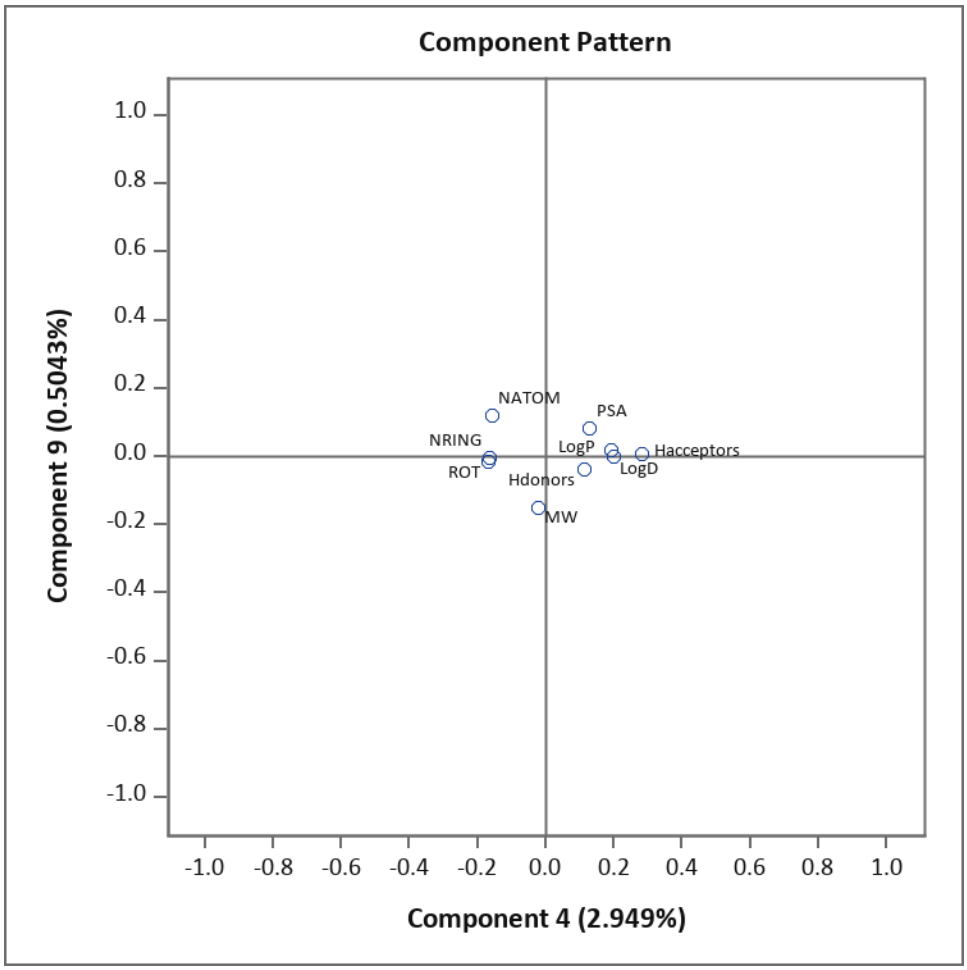

Component Pattern
Component 8 (0.7281%)
Component 6 (1.212%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
MW
LogP
NRING
NROT
NATOM
LogD
Hdonors
Hacceptors

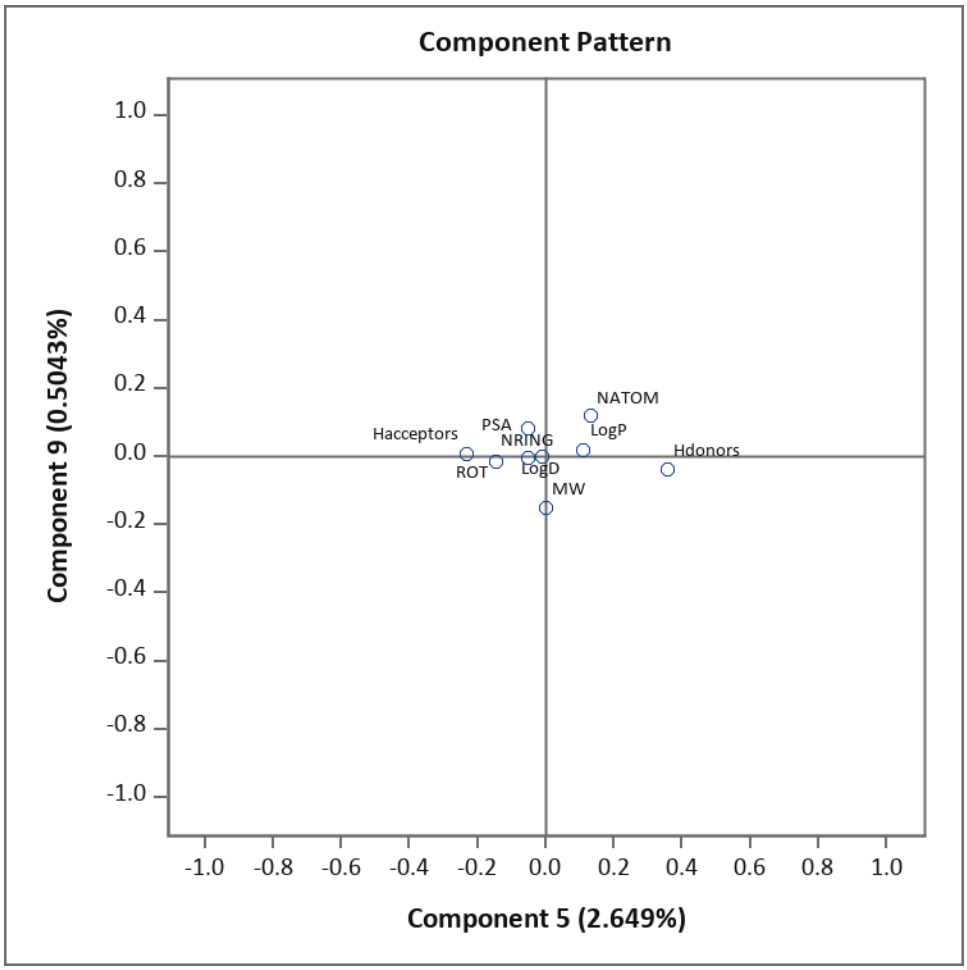

Component Pattern
Component 8 (0.7281%)
Component 7 (1.038%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
LogP
MW
ROT
LogD
NRING
Hdonors
NATOM
Hacceptors

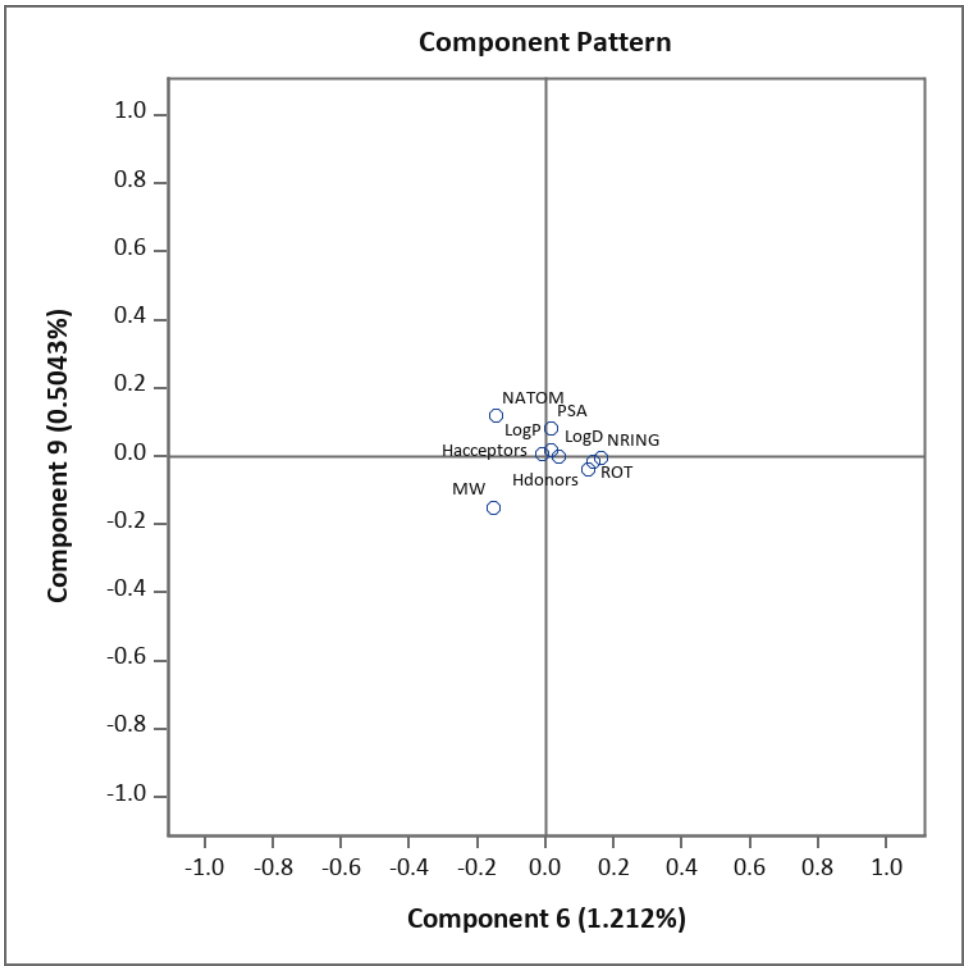

Component Pattern
Component 9 (0.5043%)
Component 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
LogP
NRING
NATOM
Hacceptors
PSA
ROT
Hdonors
MW

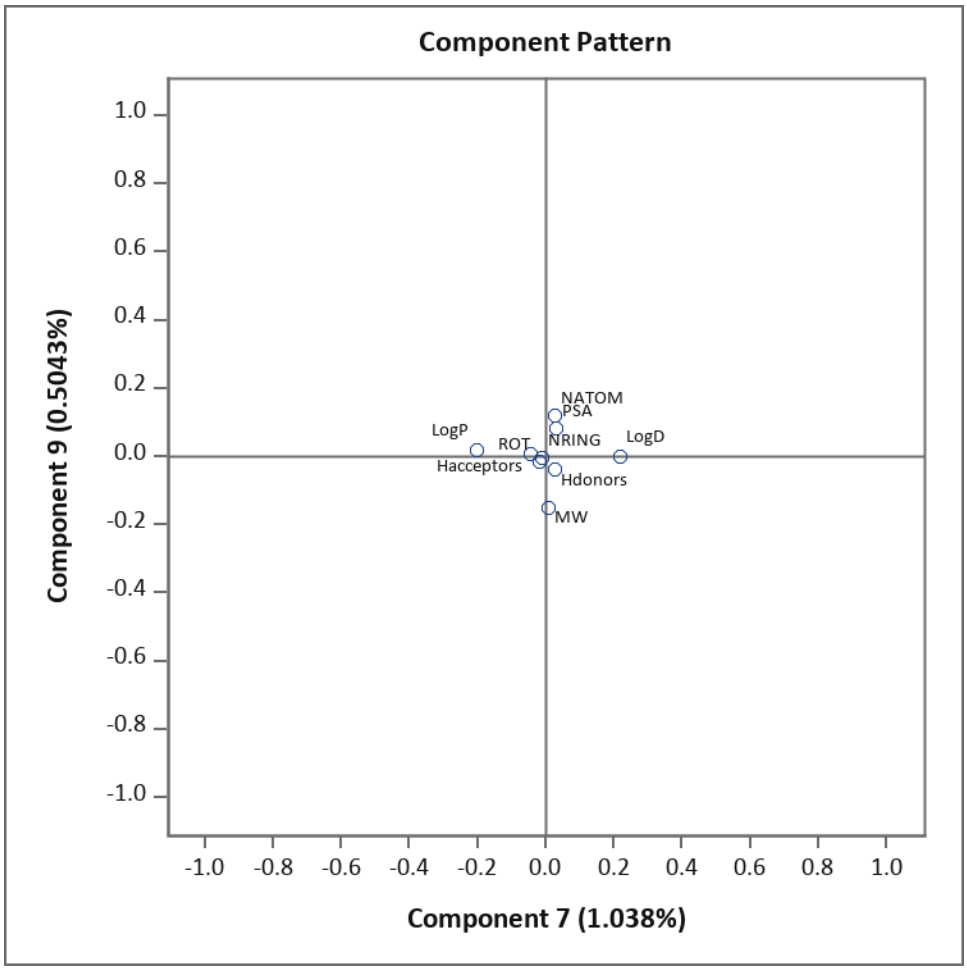

Component Pattern
Component 9 (0.5043%)
Component 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
Hacceptors
Hdonors
ROT
NATOM
NRING
MW
LogP
LogD

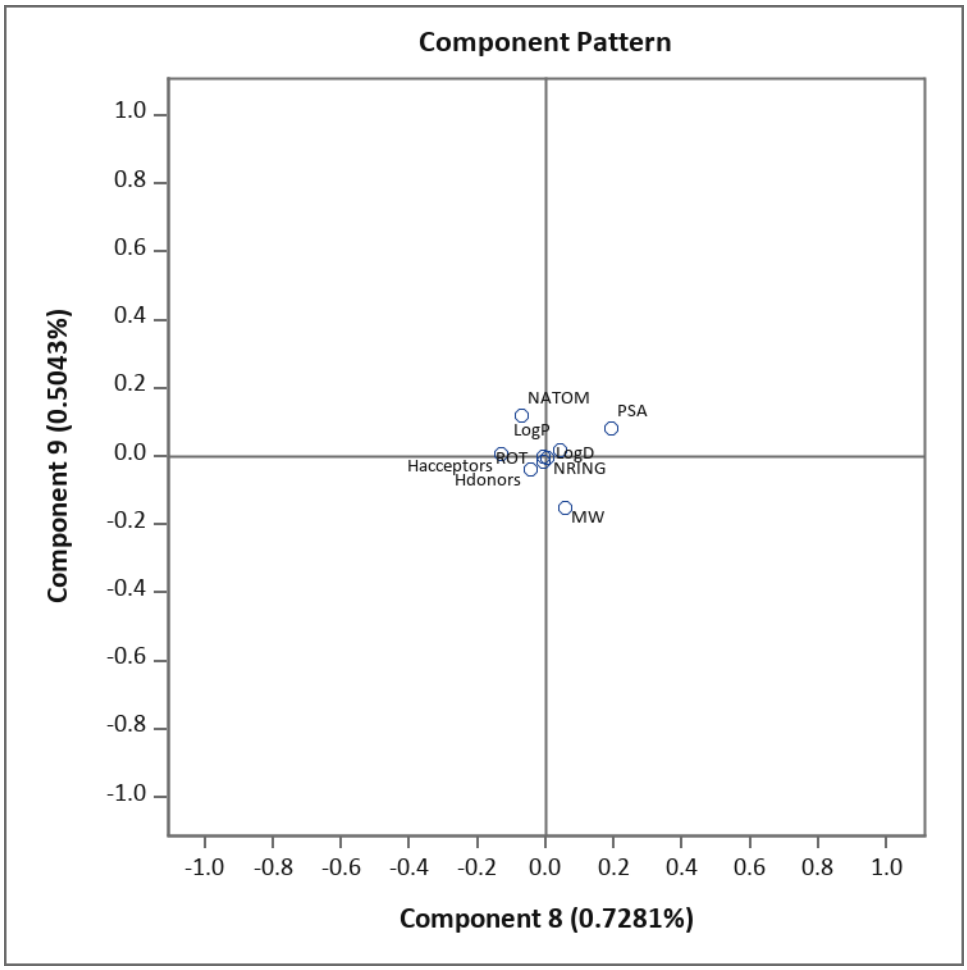

Component Pattern
Component 9 (0.5043%)
Component 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NATOM
PSA
NRING
LogP
ROT
Hacceptors
LogD
Hdonors
MW


Component Pattern
Component 9 (0.5043%)
Component 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NATOM
PSA
NRING
LogP
Hacceptors
ROT
LogD
Hdonors
MW

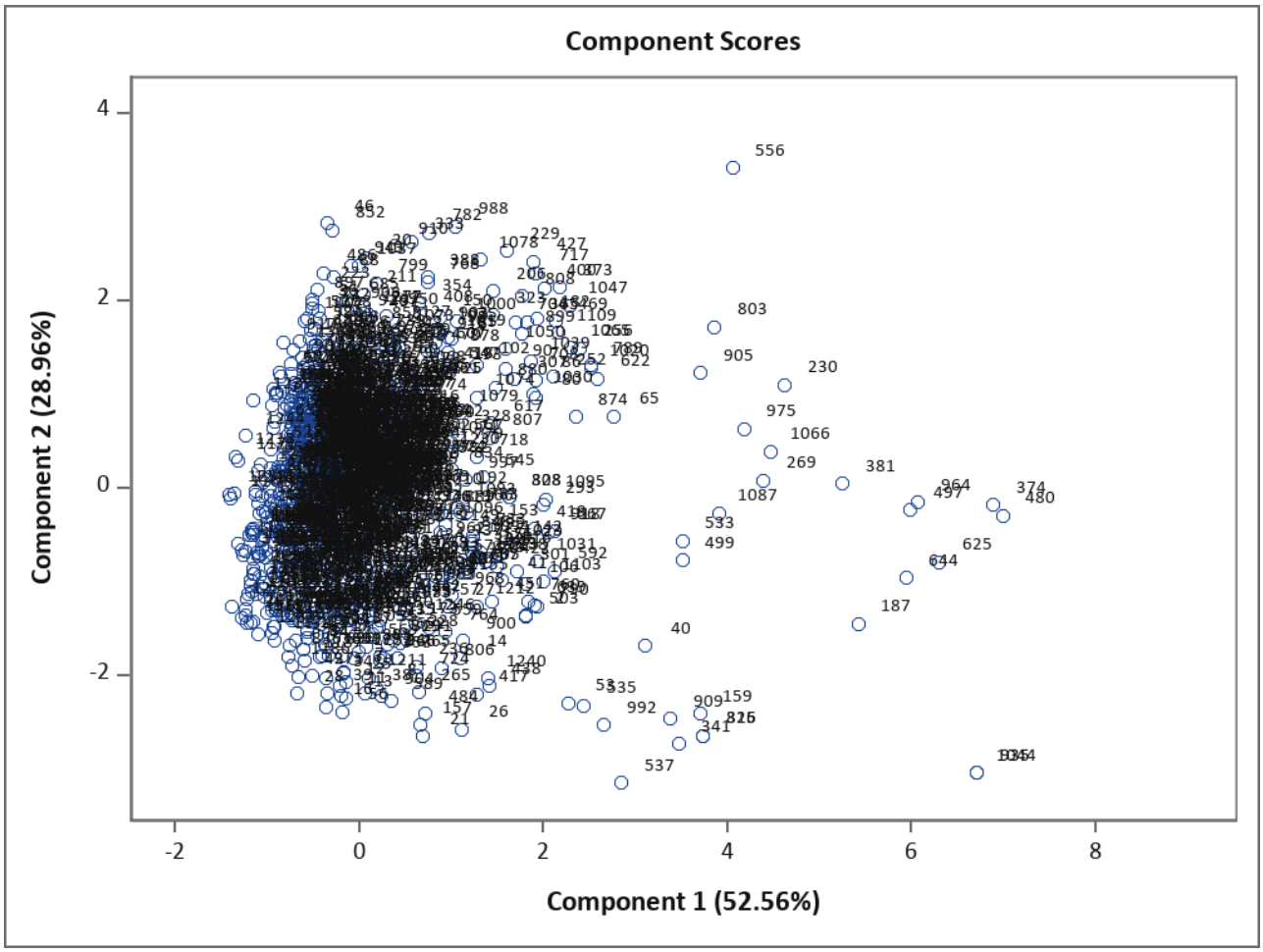

Component Pattern
Component 9 (0.5043%)
Component 5 (2.649%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NATOM
LogP
Hacceptors
PSA
NRING
Hdonors
ROT
LogD
MW

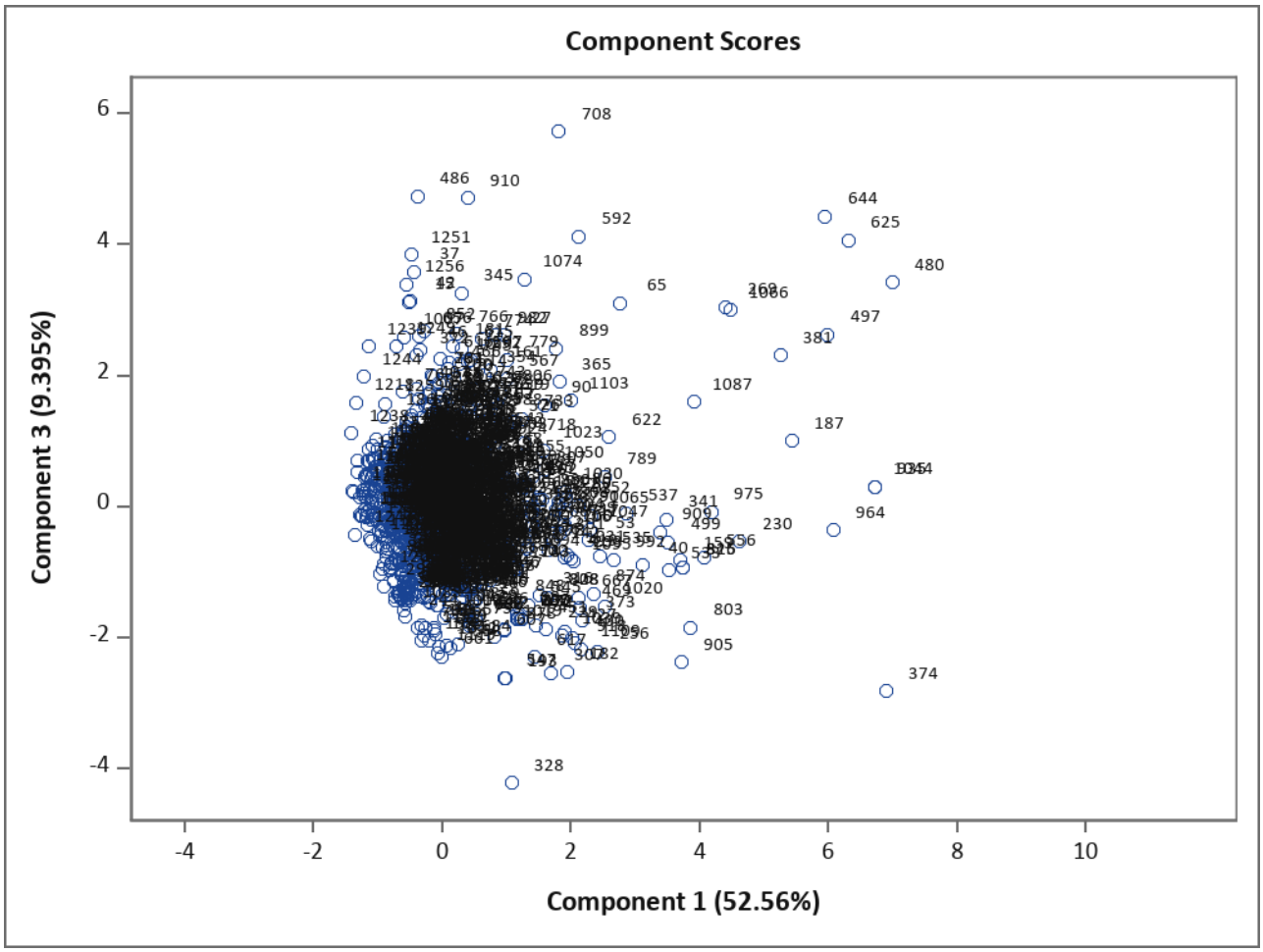

Component Pattern
Component 9 (0.5043%)
Component 6 (1.212%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NATOM
PSA
LogP
LogD
NRING
Hacceptors
ROT
Hdonors
MW

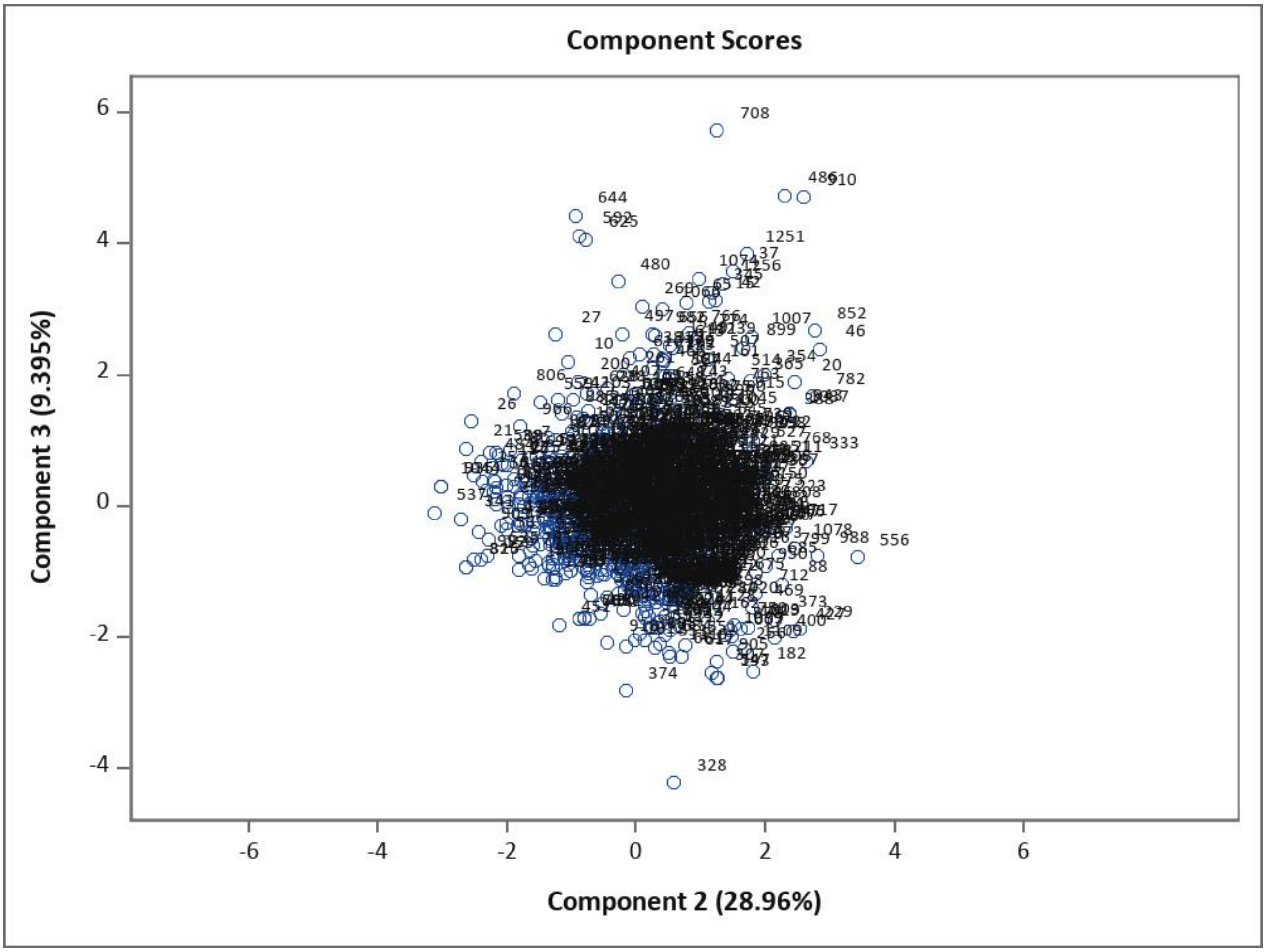

Component Pattern
Component 9 (0.5043%)
Component 7 (1.038%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NATOM
PSA
LogP
ROT
NRING
LogD
Hacceptors
Hdonors
MW

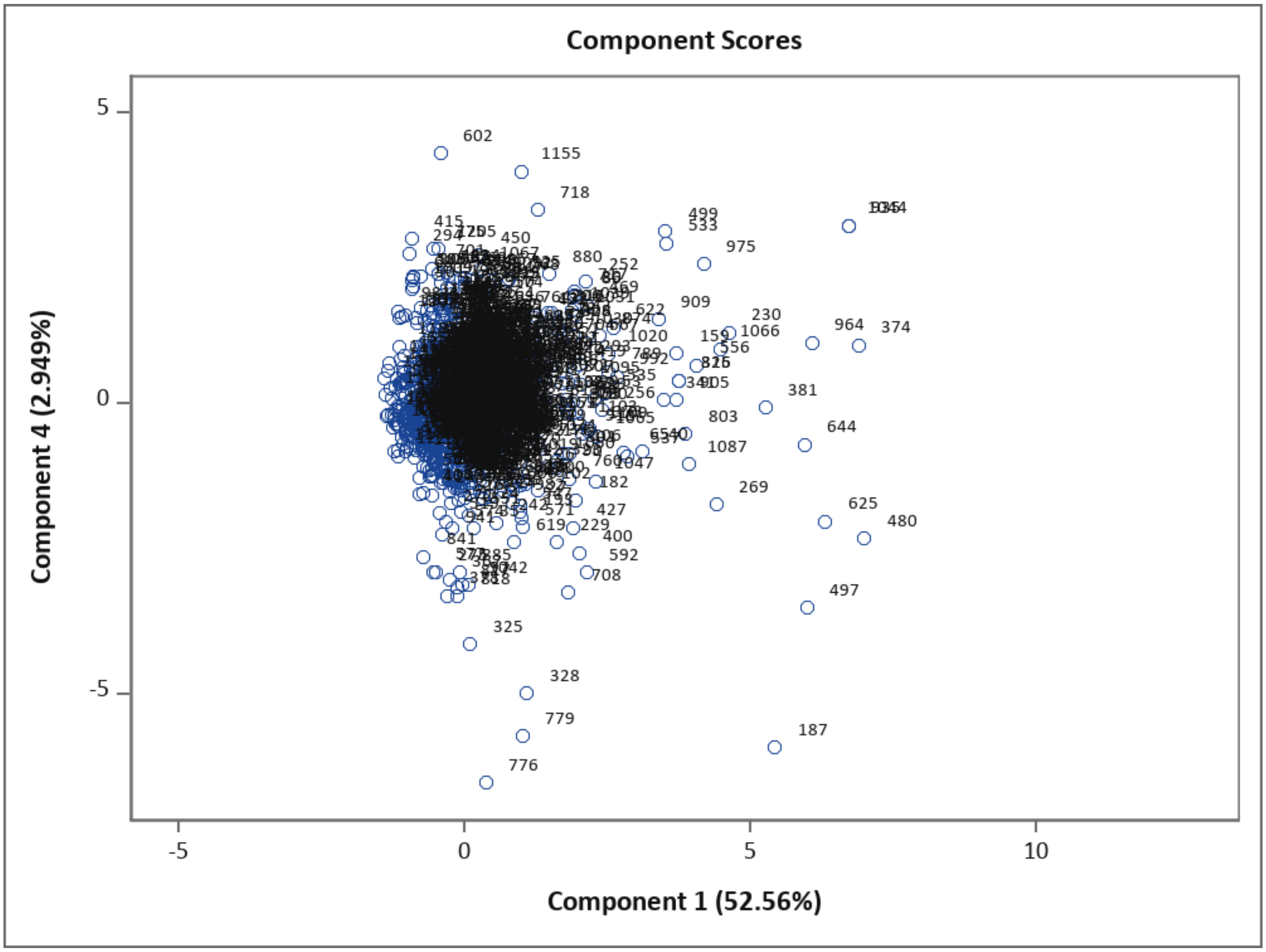

Component Pattern
Component 9 (0.5043%)
Component 8 (0.7281%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NATOM
PSA
LogP
LogD
Hacceptors
ROT
NRING
Hdonors
MW

4. Score plots for all observations

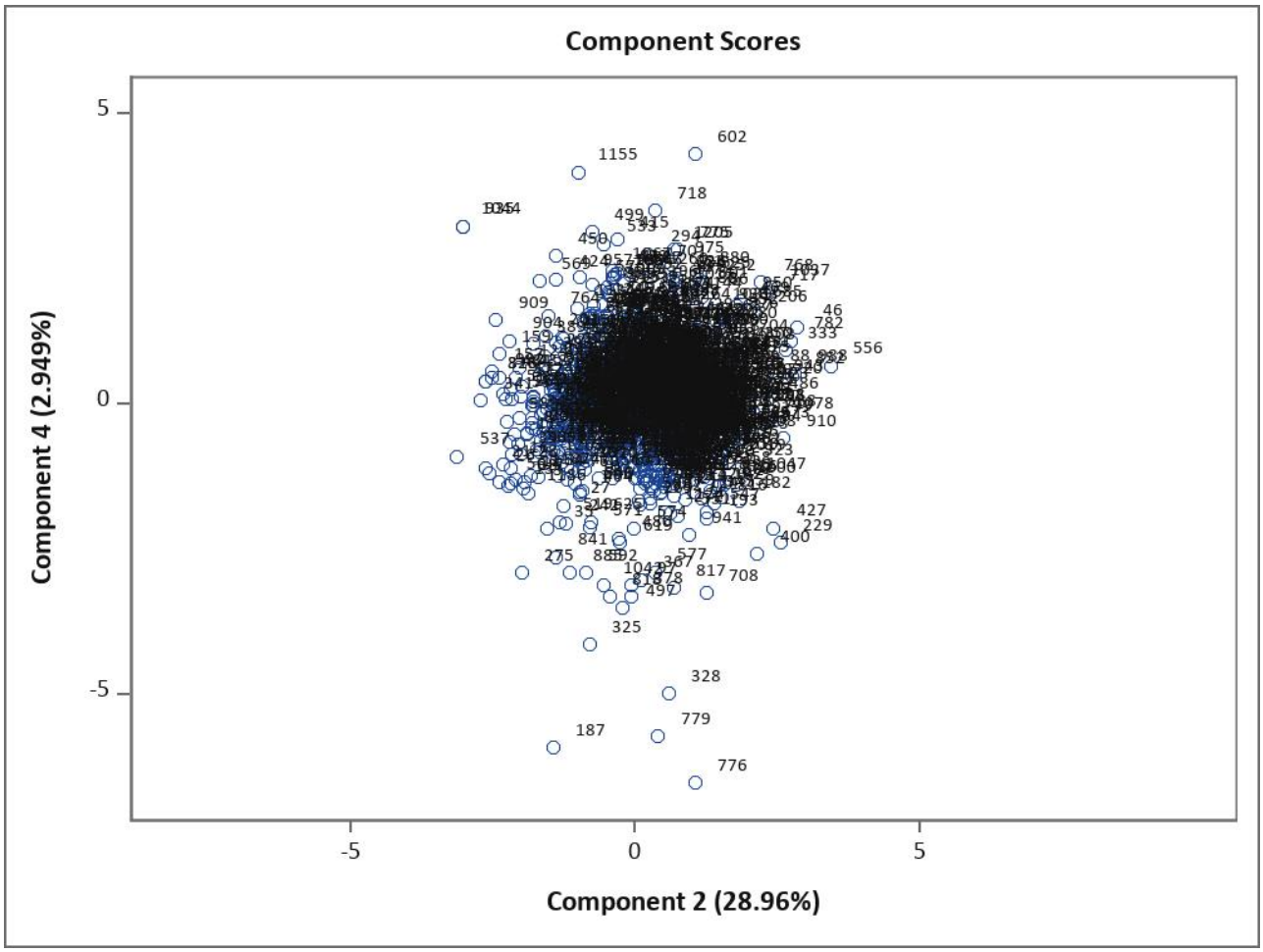


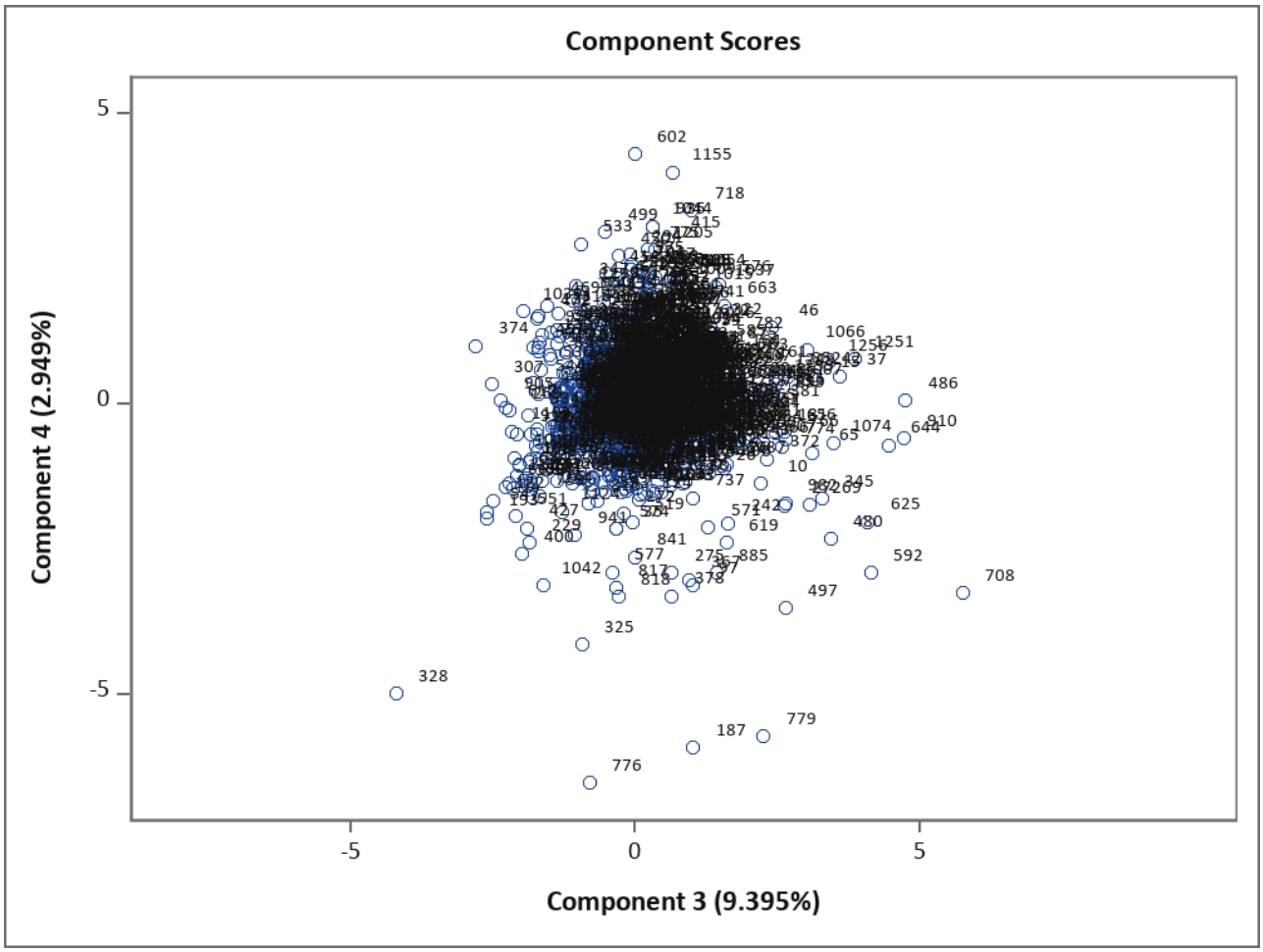

Component Scores
Component 3 (9.395%)
Component 1 (52.56%)
6
4
2
0
-2
-4
-4
-2
0
2
4
6
8
10
708
486
910
644
625
592
1251
37
1256
1074
480
345
65
497
899
381
365
1103
1087
622
187
789
975
964
230
341
909
499
537
803
905
374
328

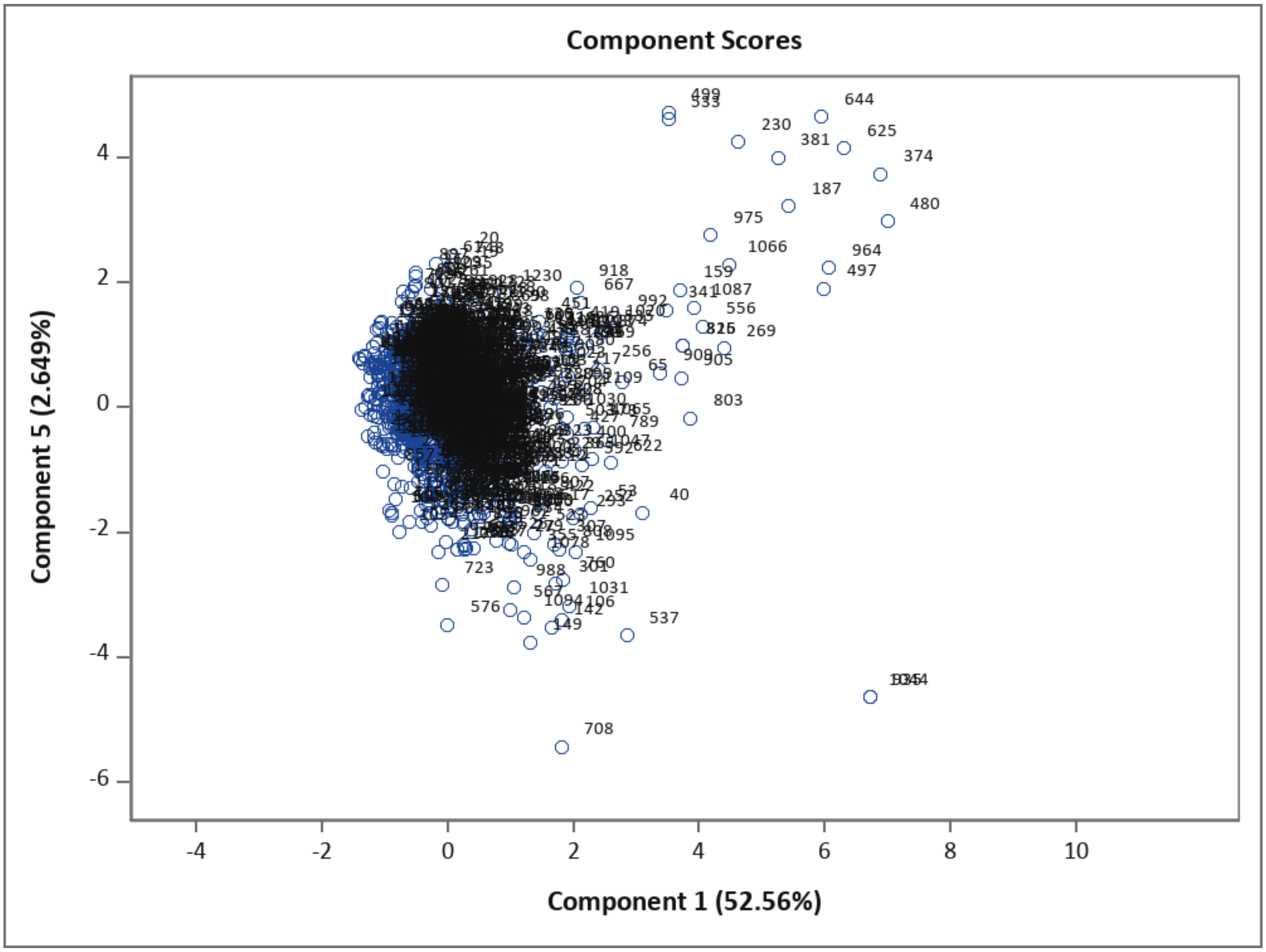

Component Scores
Component 3 (9.395%)
Component 2 (28.96%)
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
708
486
910
644
592
625
1251
480
328
374
556
852
46
782
333
988
1078
182

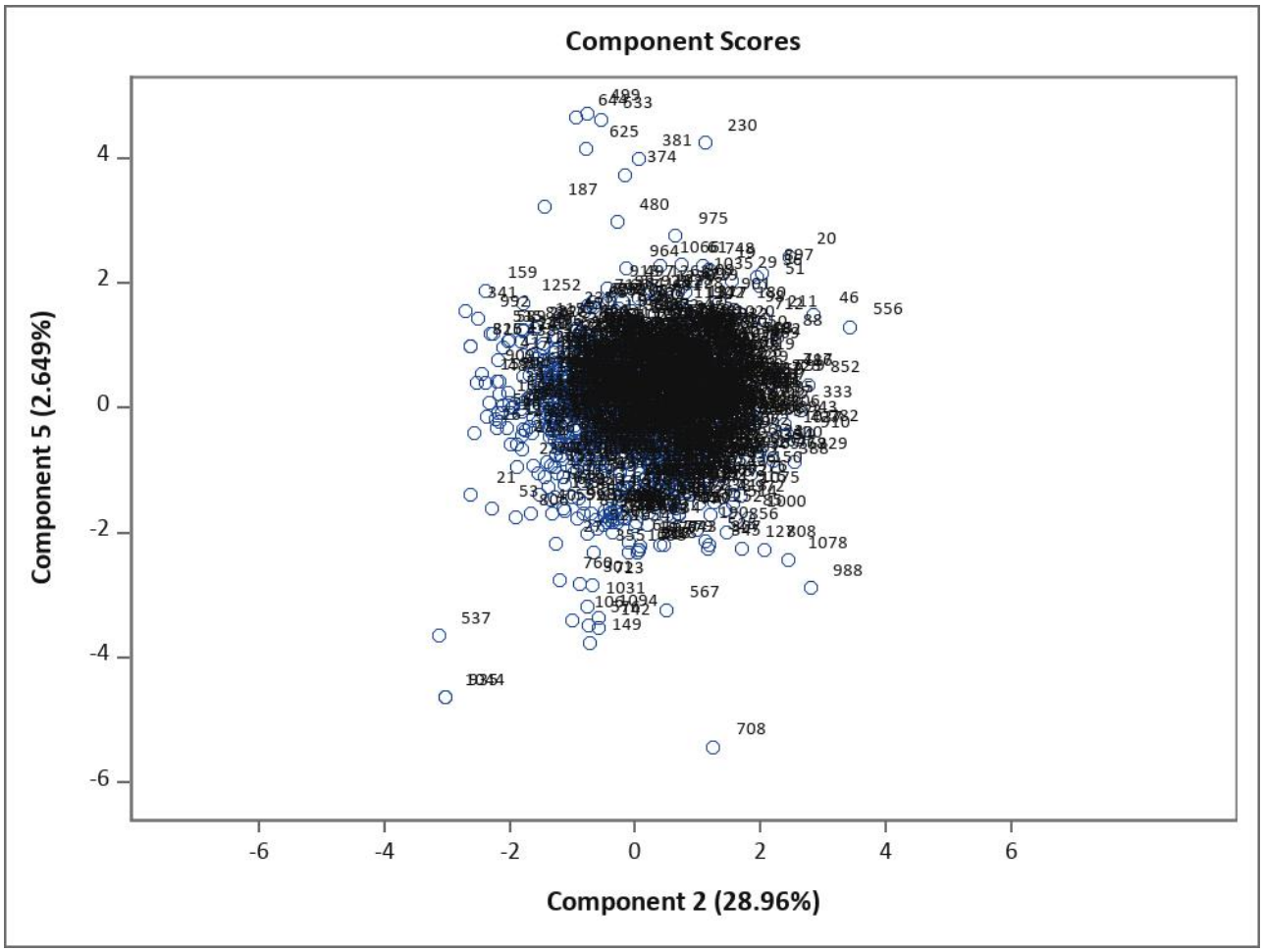


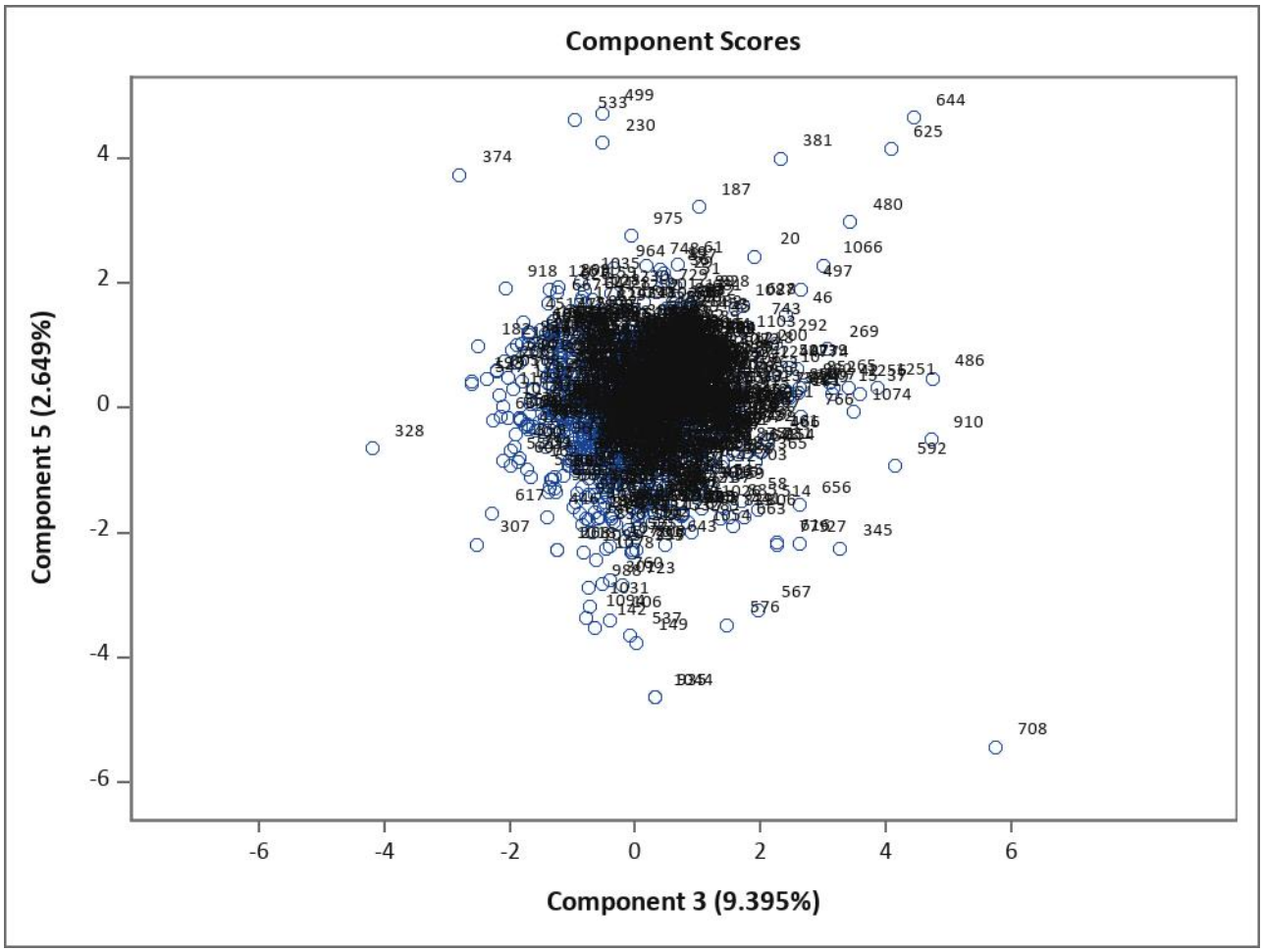
Component Scores
Component 4 (2.949%)
Component 1 (52.56%)

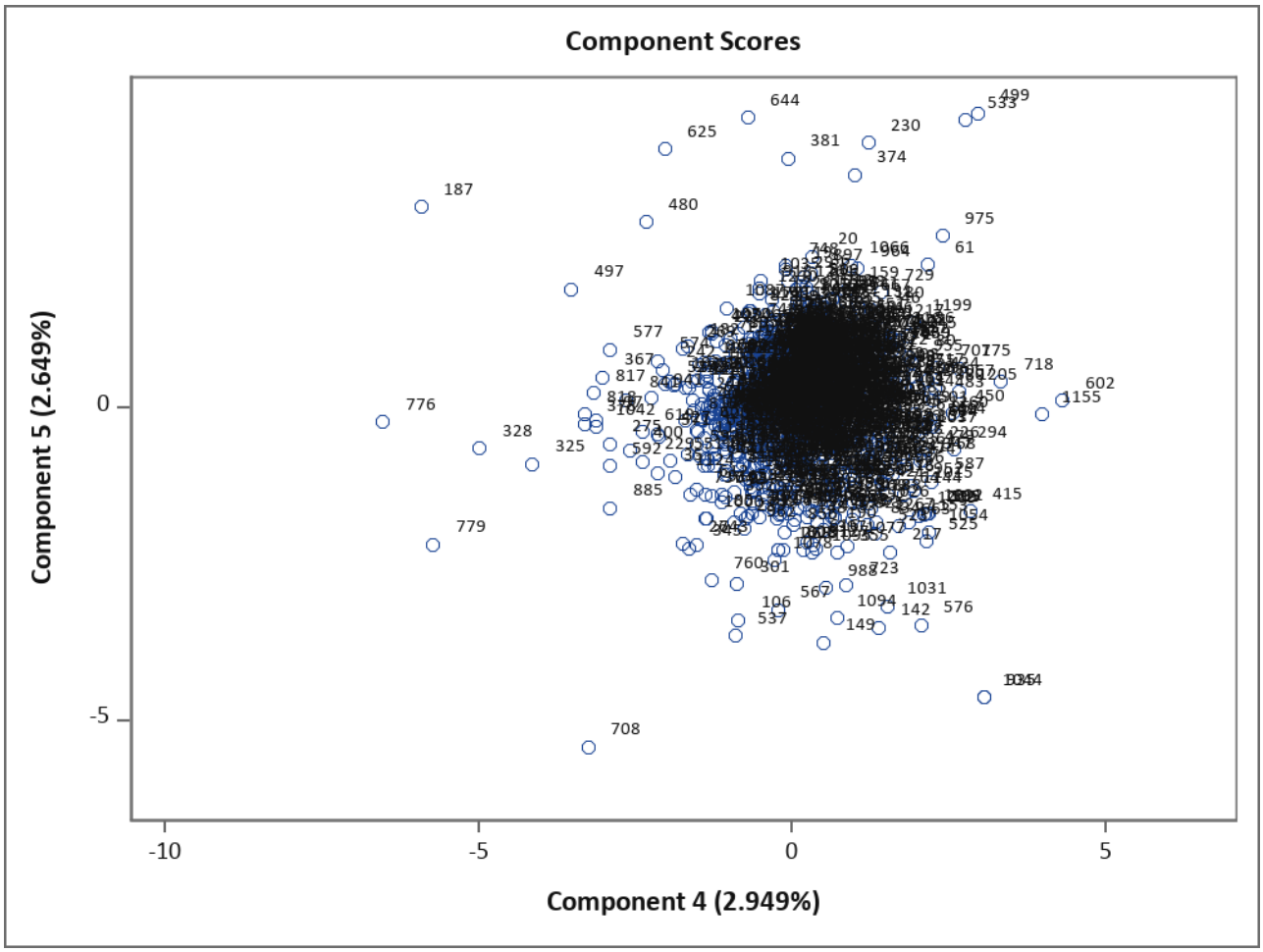

Component Scores
Component 4 (2.949%)
Component 2 (28.96%)
5
0
-5
-5
0
5
602
1155
718
499
450
909
556
782
333
46
910
427
229
400
941
577
817
708
325
328
779
187
776

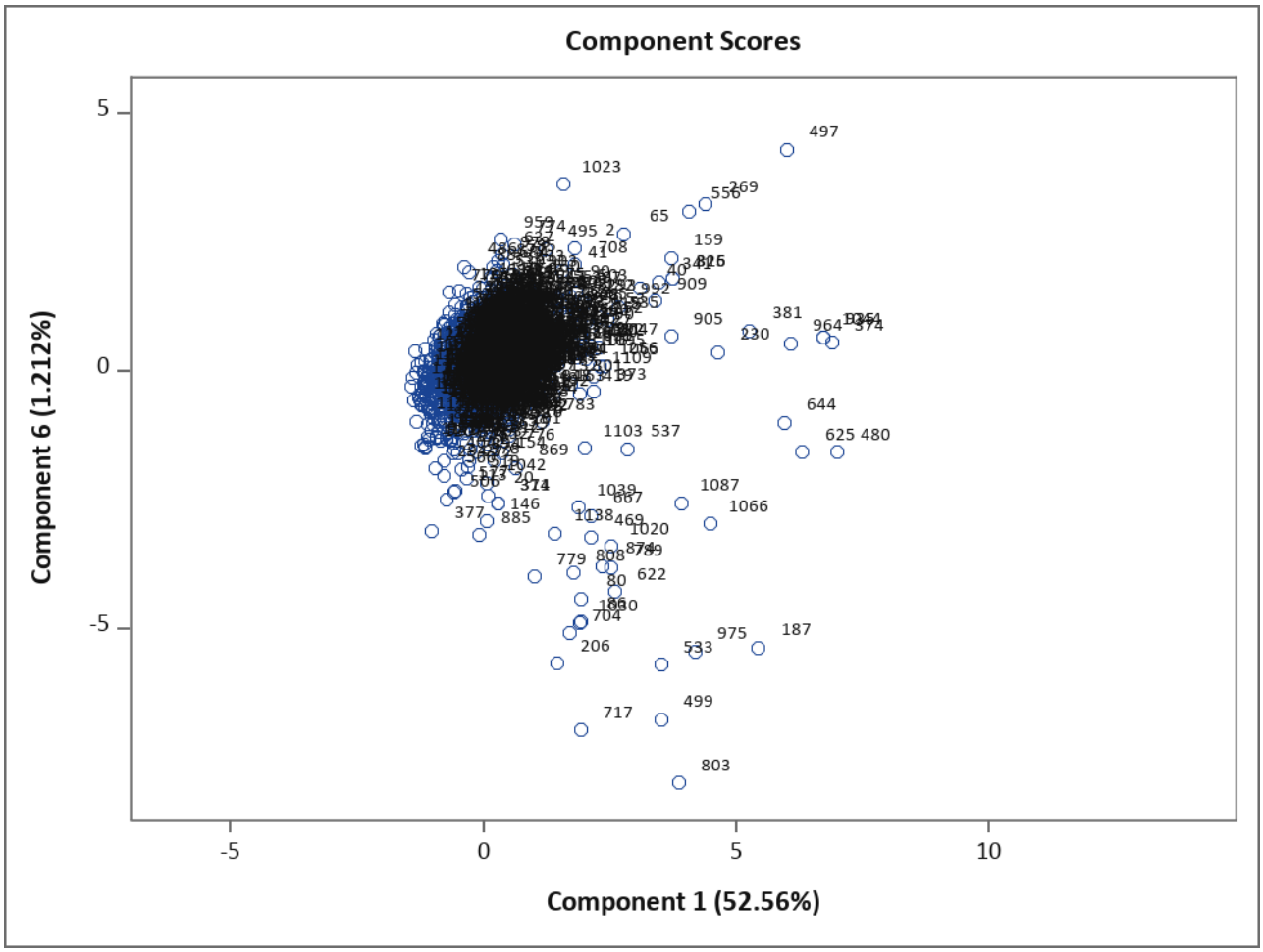

Component Scores
Component 4 (2.949%)
Component 3 (9.395%)
5
0
-5
-5
0
5
602
1155
718
499
533
415
663
46
374
1066
1256
1251
307
37
486
910
644
1074
65
372
10
737
345
625
480
619
571
242
841
577
275
885
592
708
1042
817
818
378
497
325
328
779
187
776

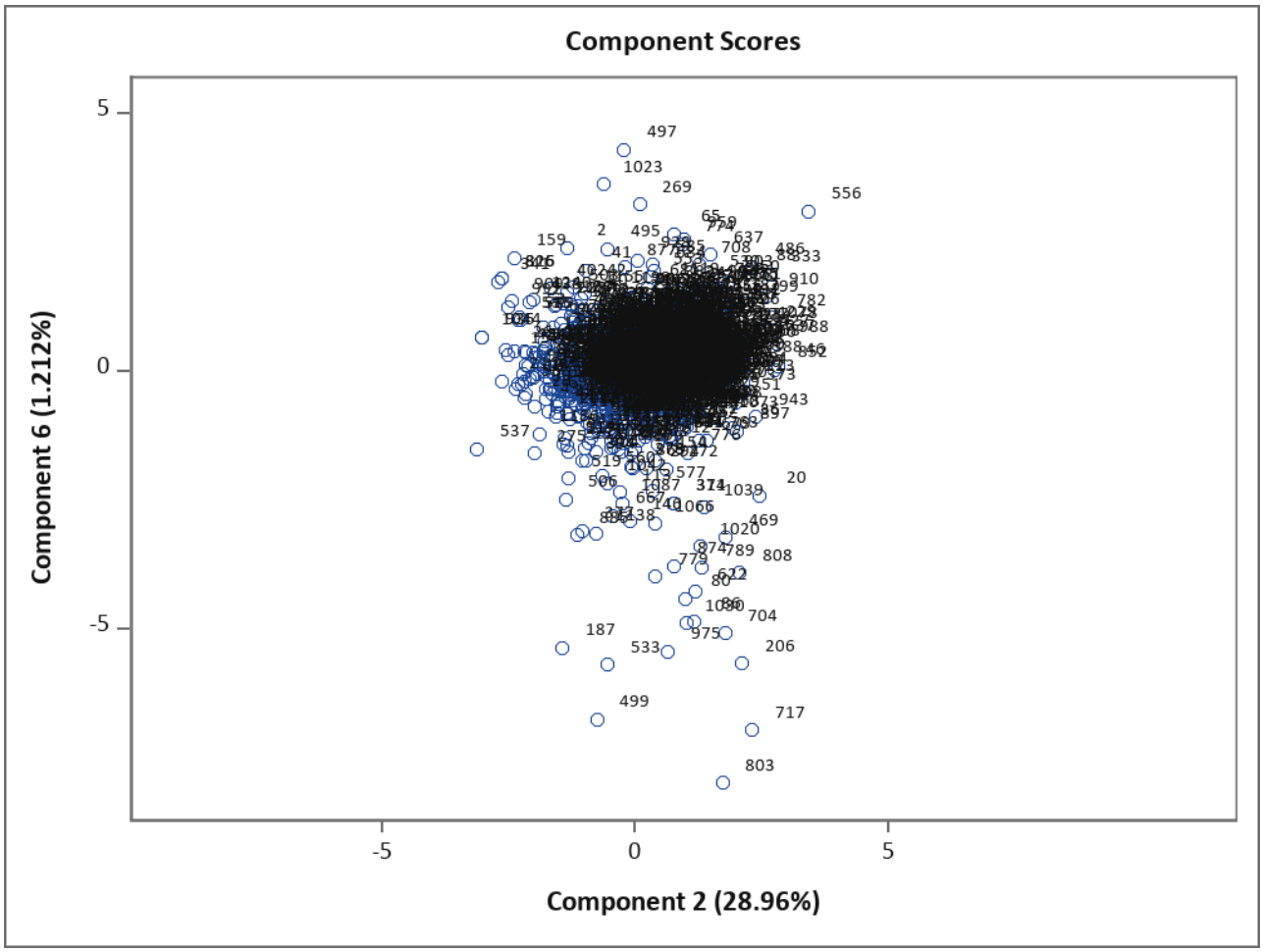

Component Scores
Component 5 (2.649%)
Component 1 (52.56%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8
10
499
533
644
230
381
625
374
187
480
975
1066
964
497
918
159
667
341
1087
992
556
269
256
909
905
65
803
789
40
537
708
723
988
301
760
1031
567
1094
106
142
576
149
1078
1095

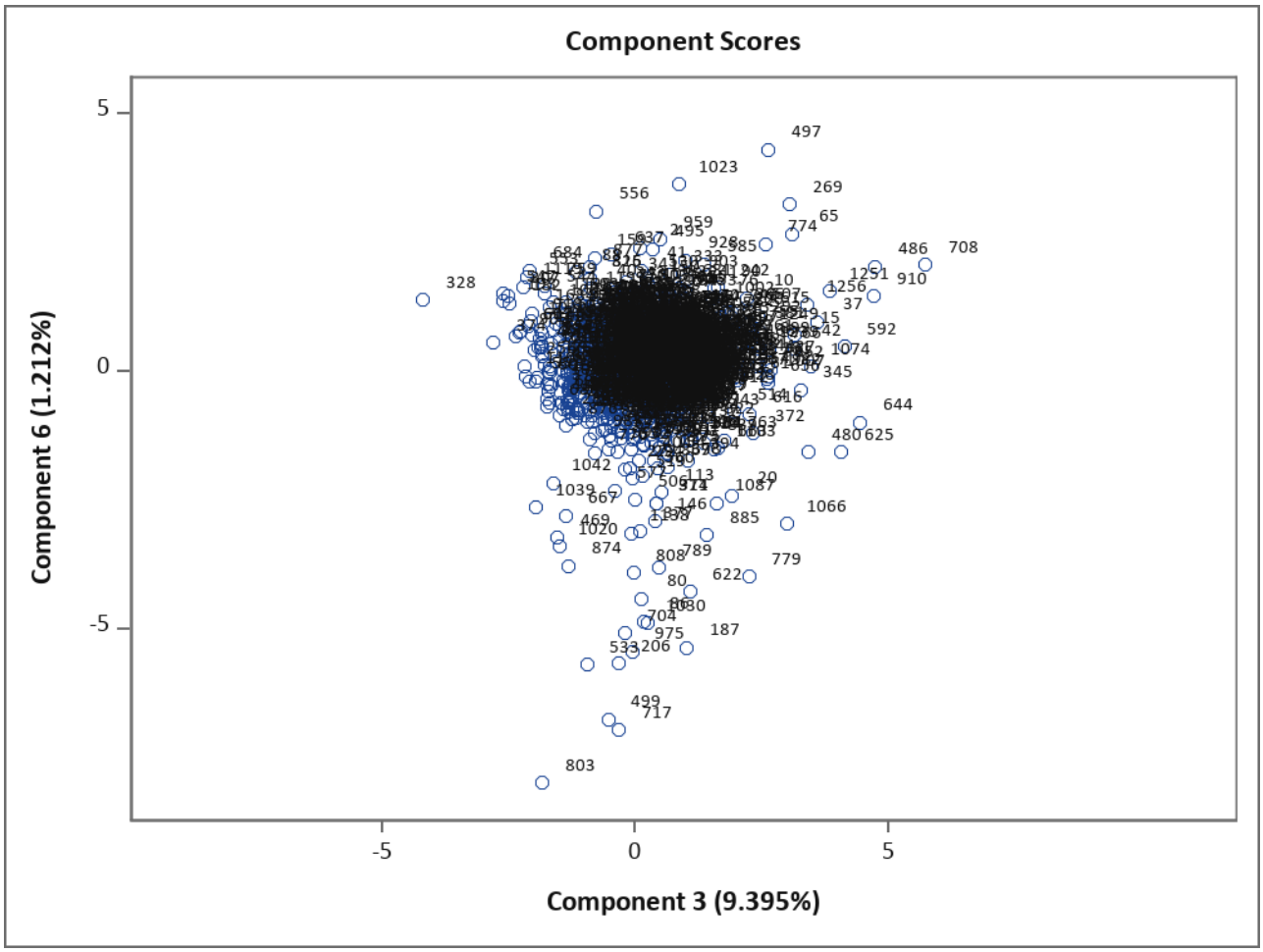

Component Scores
Component 5 (2.649%)
Component 2 (28.96%)
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6
499
533
625
230
381
374
187
480
975
20
159
1252
46
556
852
333
537
708
1078
988
567
1031
149

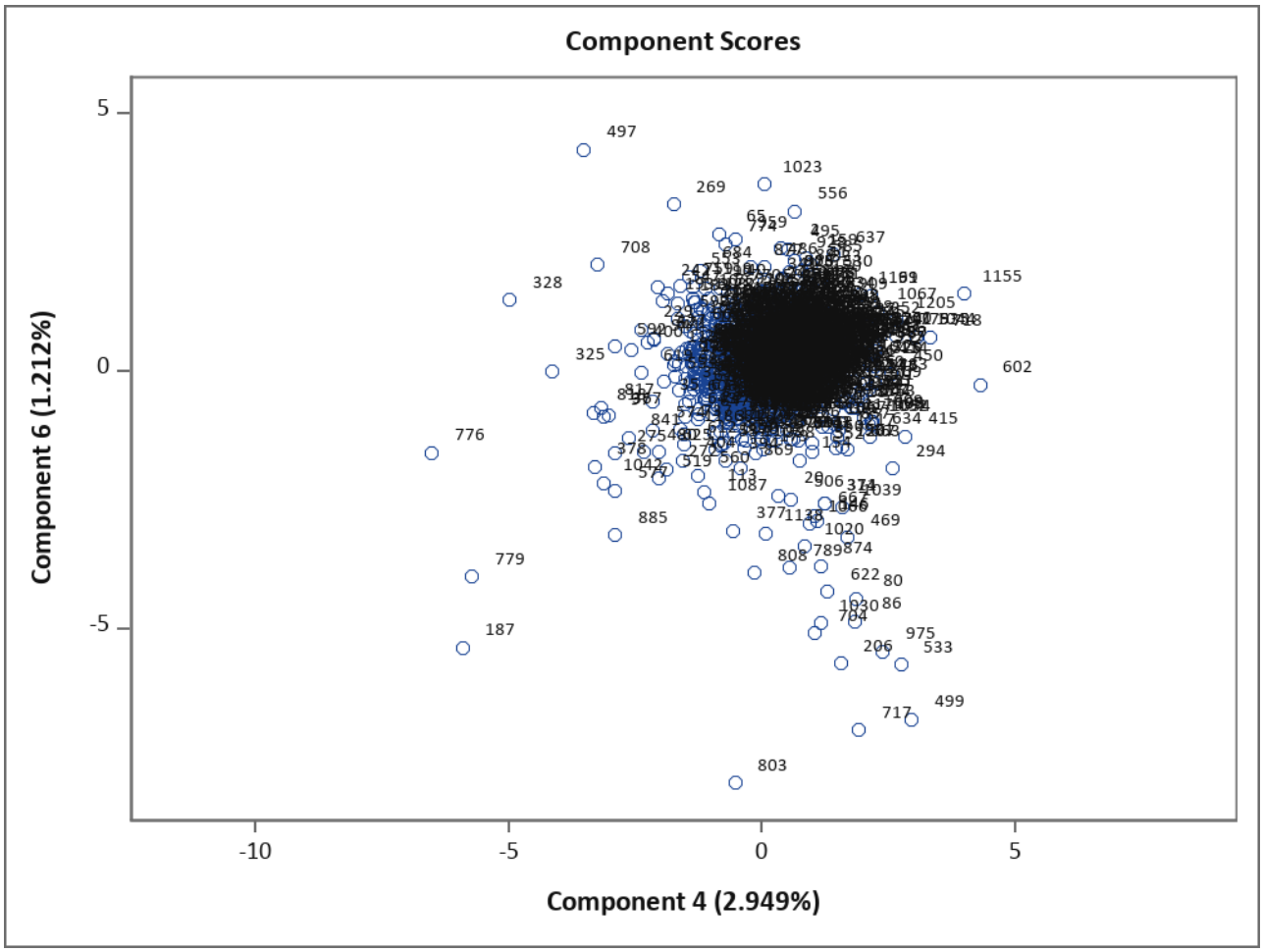

Component Scores
Component 5 (2.649%)
Component 3 (9.395%)
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6
499
533
230
644
625
381
374
187
480
975
20
1066
964
497
46
743
292
269
486
1074
910
328
592
656
307
345
567
576
149
708

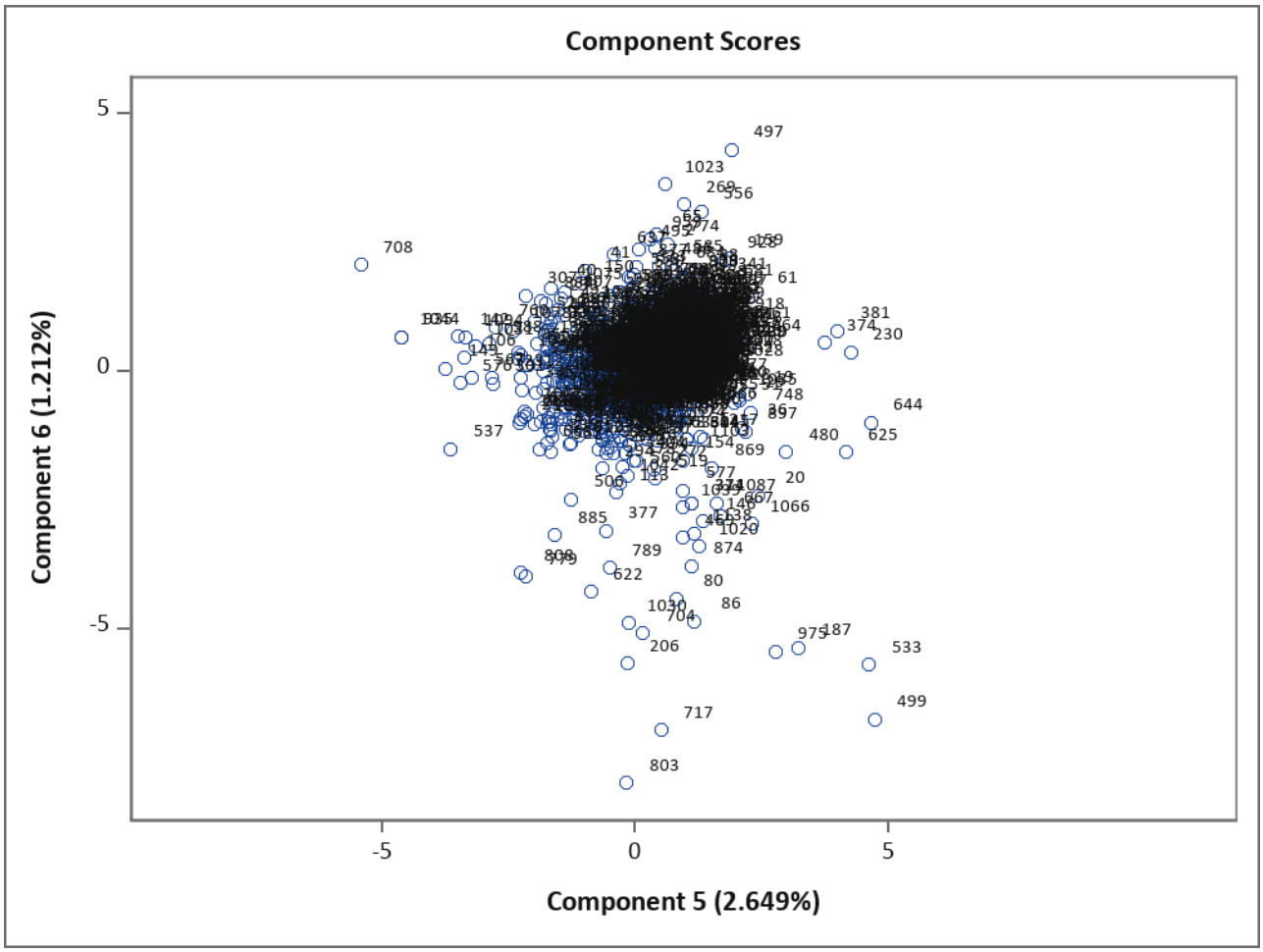

Component Scores
Component 5 (2.649%)
Component 4 (2.949%)
0
-5
-10
-5
0
5
644
625
381
230
374
499
187
480
975
61
497
776
328
325
779
885
577
367
817
718
602
1155
708

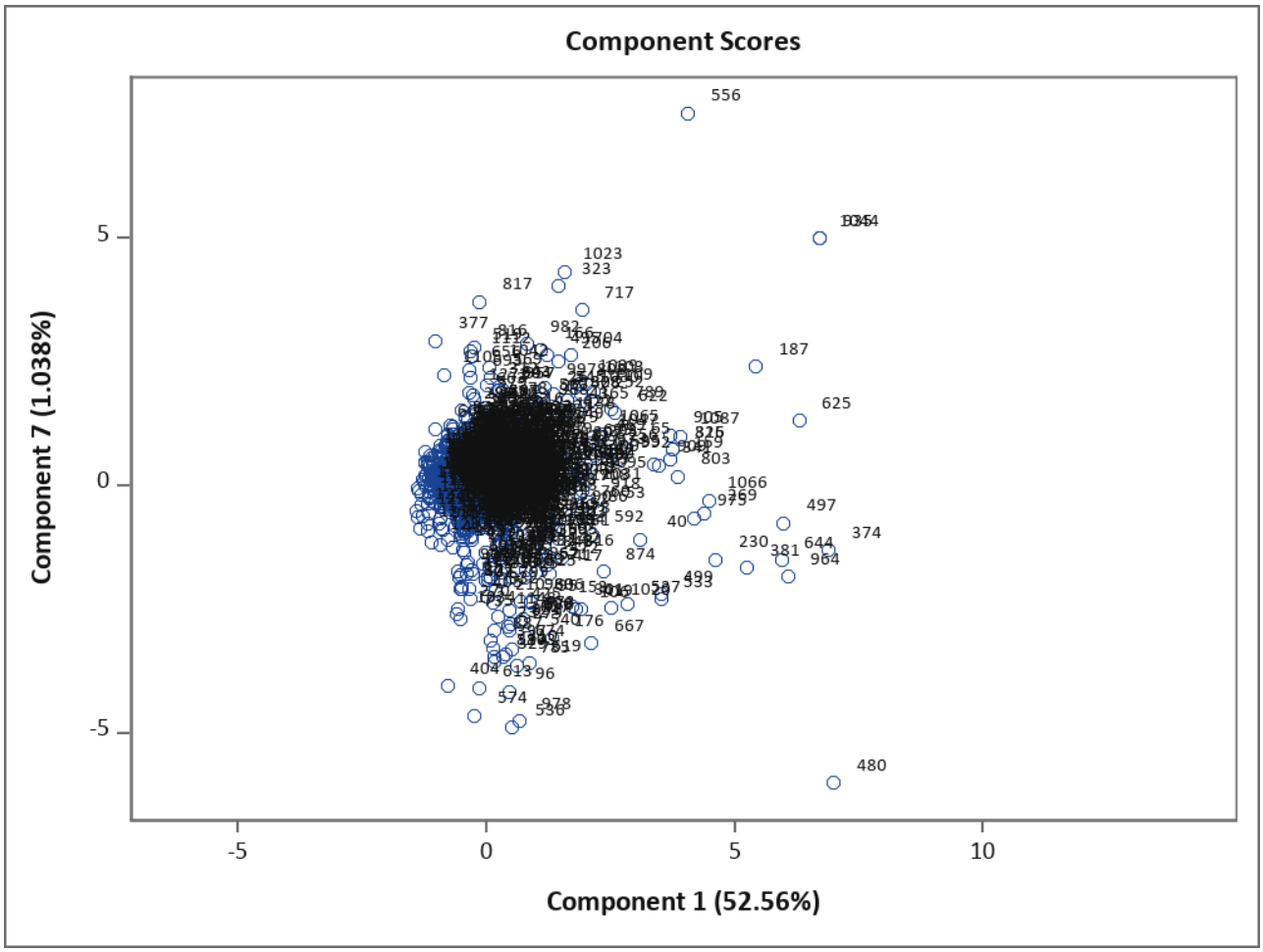

Component Scores
Component 6 (1.212%)
Component 1 (52.56%)
5
0
-5
-5
0
5
10
497
1023
269
556
65
495
2
159
708
41
40
909
905
381
230
964
374
644
625
480
1103
537
869
1087
1066
1039
667
1138
469
1020
779
808
622
80
704
975
187
206
533
499
717
803

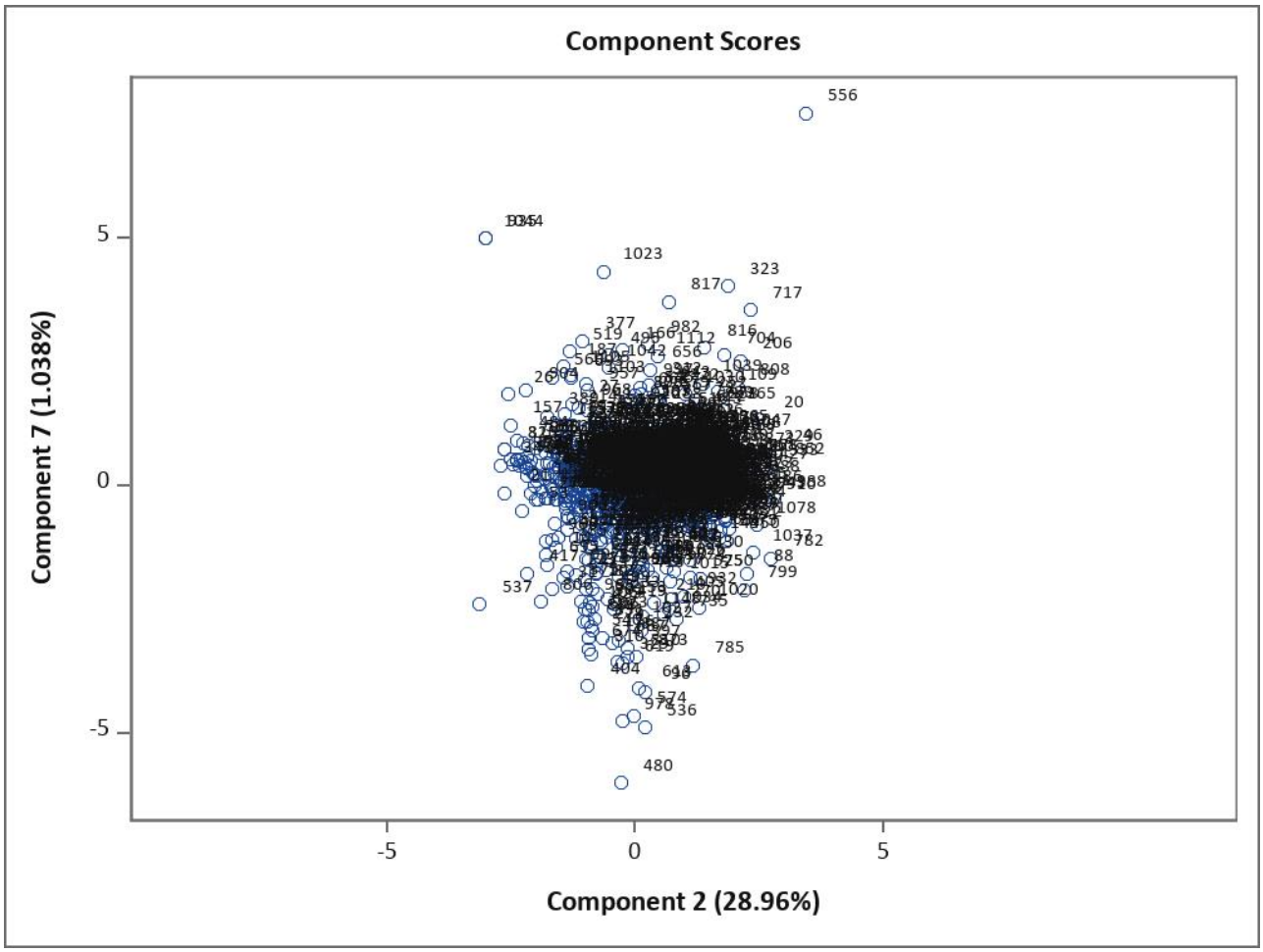

Component Scores
Component 6 (1.212%)
Component 2 (28.96%)
5
0
-5
-5
0
5
497
1023
269
556
65
774
2
495
637
159
41
708
486
833
910
782
537
897
943
20
1039
469
1020
808
789
704
187
533
206
499
717
803

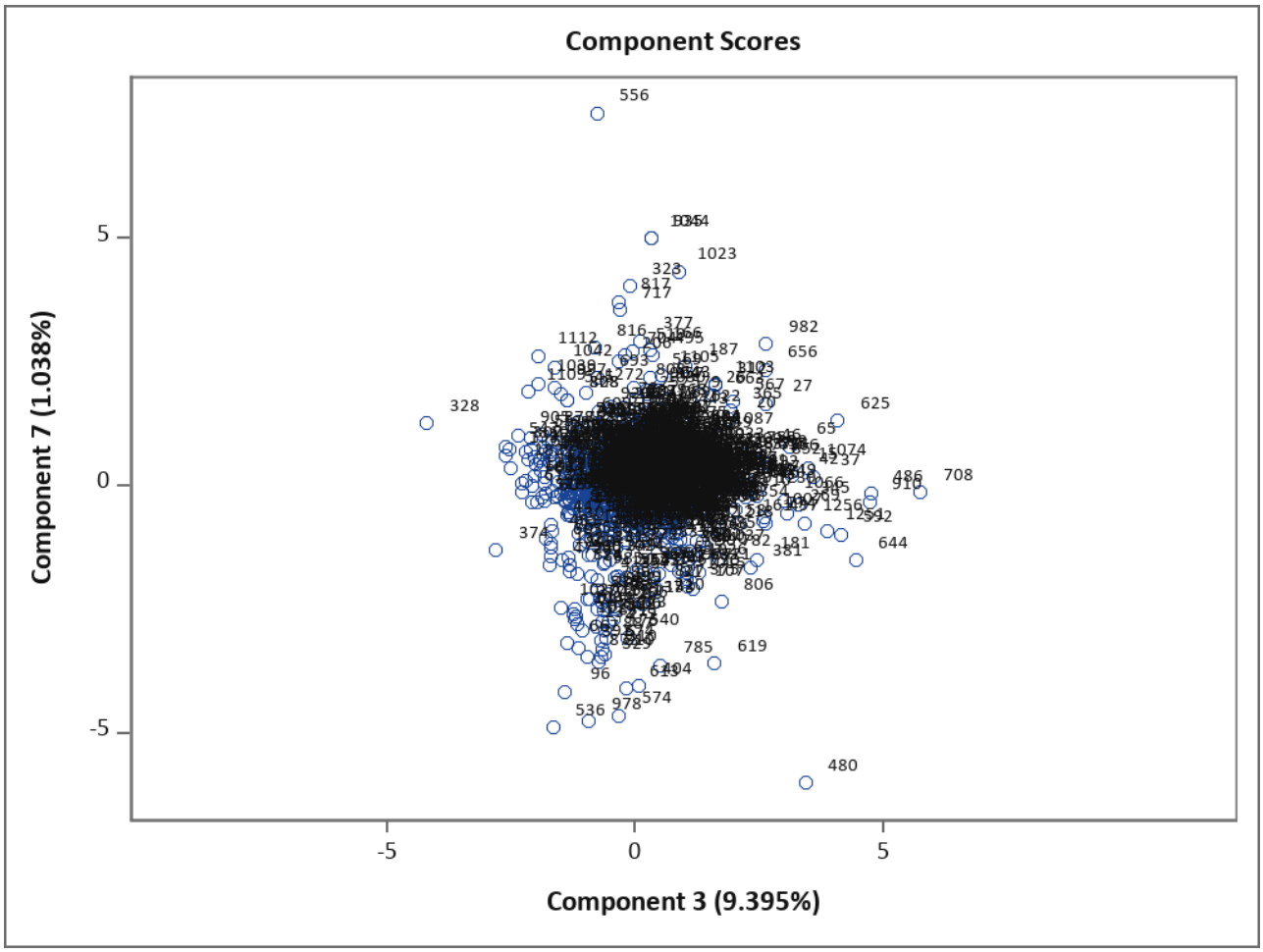

Component Scores
Component 6 (1.212%)
Component 3 (9.395%)
5
0
-5
-5
0
5
497
1023
556
269
65
774
959
486
708
328
1251
910
1256
37
592
1074
345
644
372
480
625
1042
1039
667
1066
885
1020
874
808
789
779
622
80
704
975
187
533
206
499
717
803

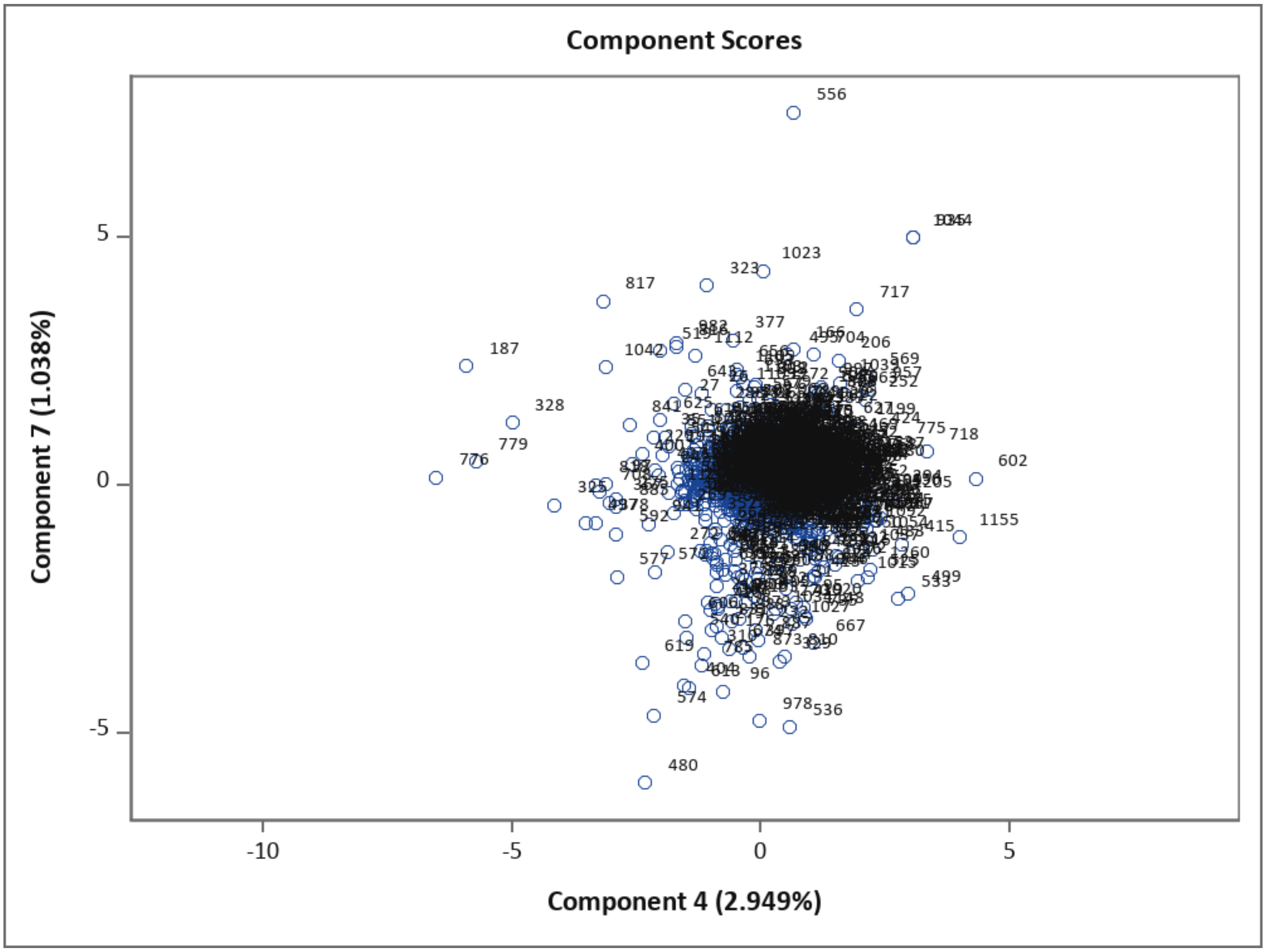

Component Scores
Component 6 (1.212%)
Component 4 (2.949%)
5
0
-5
-10
-5
0
5
497
1023
269
556
708
328
1155
602
325
776
294
885
779
187
975
533
499
717
803

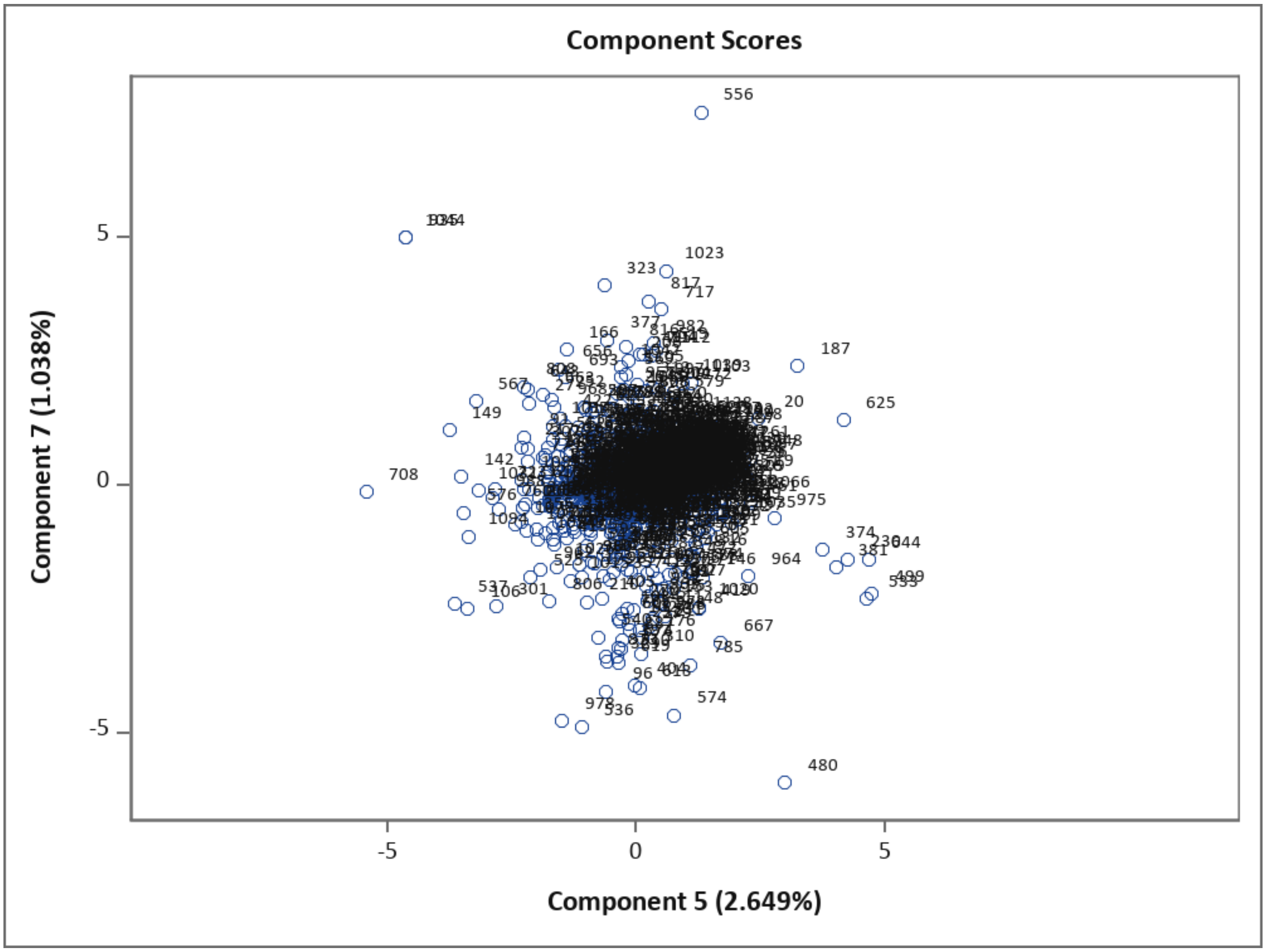

Component Scores
Component 6 (1.212%)
Component 5 (2.649%)
5
0
-5
-5
0
5
497
1023
269
556
708
61
381
374
230
644
537
480
625
869
20
1066
885
377
789
874
622
80
1030
86
704
206
975
187
533
499
717
803


Component Scores
Component 7 (1.038%)
Component 1 (52.56%)
5
0
-5
-5
0
5
10
556
1023
323
817
717
377
187
625
803
1066
497
374
592
40
230
381
644
964
874
667
404
613
96
574
978
530
480

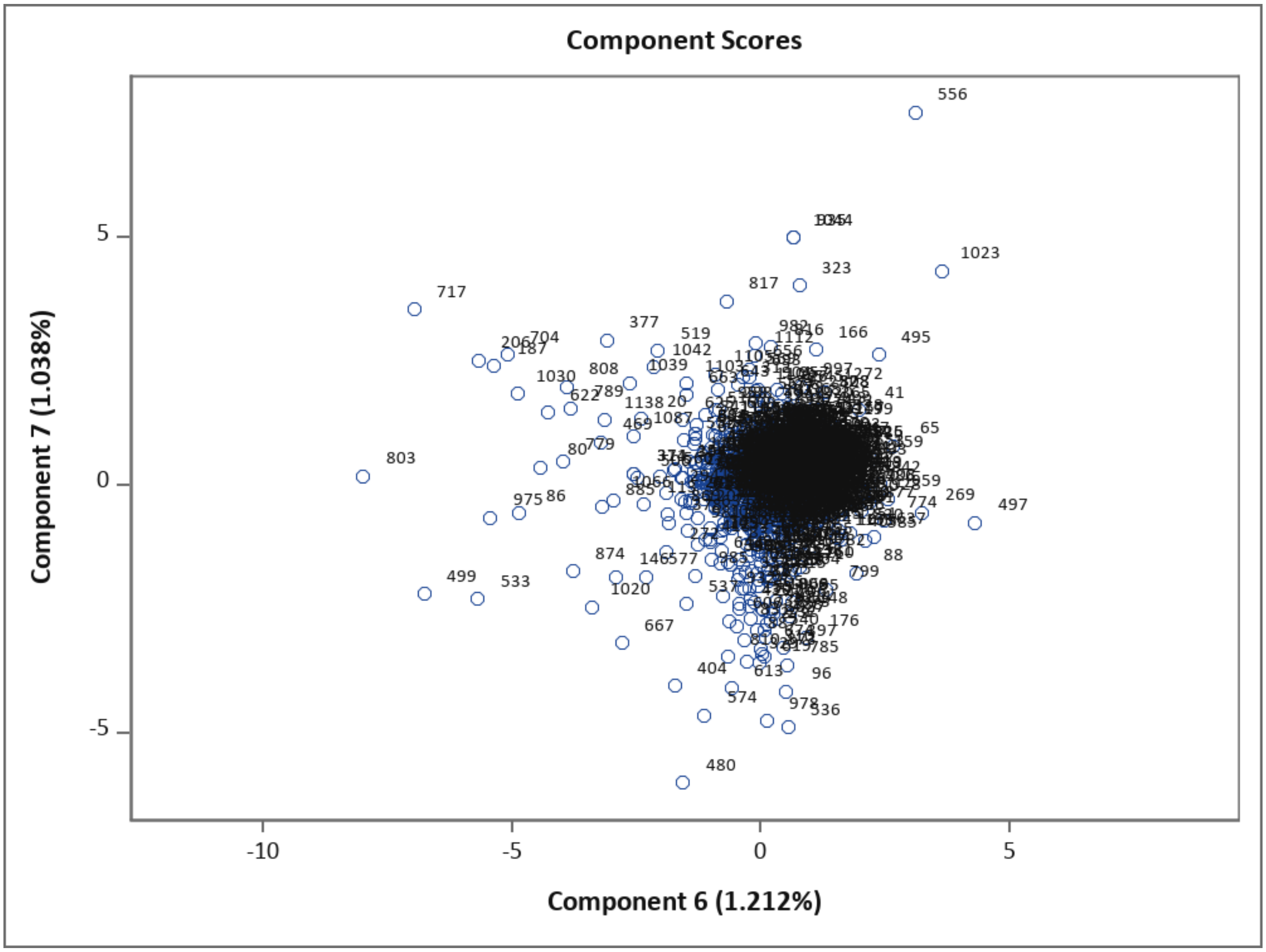

Component Scores
Component 7 (1.038%)
Component 2 (28.96%)
5
0
-5
-5
0
5
556
1023
323
817
717
377
816
704
206
20
1078
1037
782
88
799
537
785
480
574
536

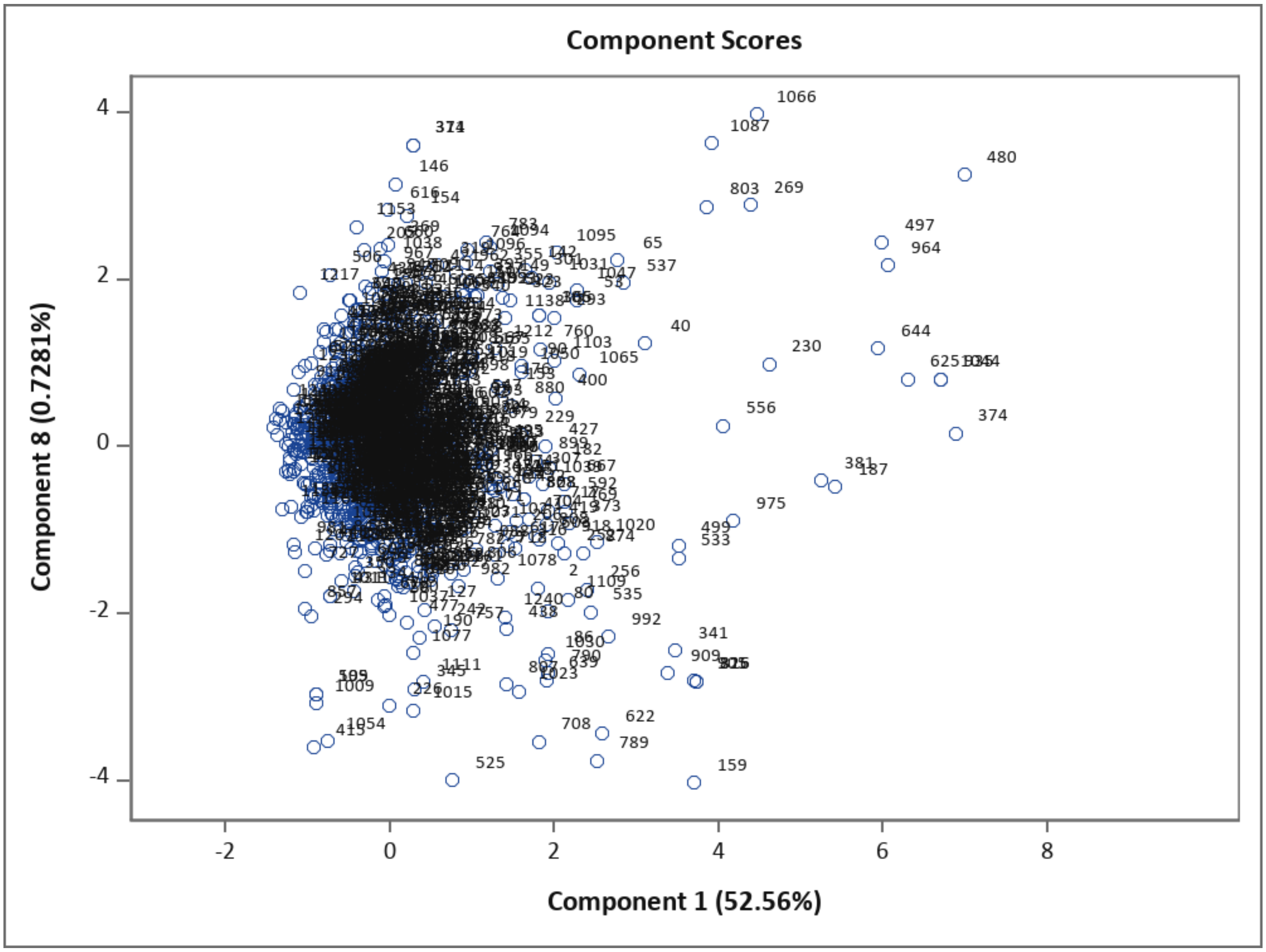

Component Scores
Component 7 (1.038%)
Component 3 (9.395%)
5
0
-5
-5
0
5
556
1023
323
817
717
982
656
187
328
625
65
1074
708
644
181
806
785
619
96
574
536
978
480

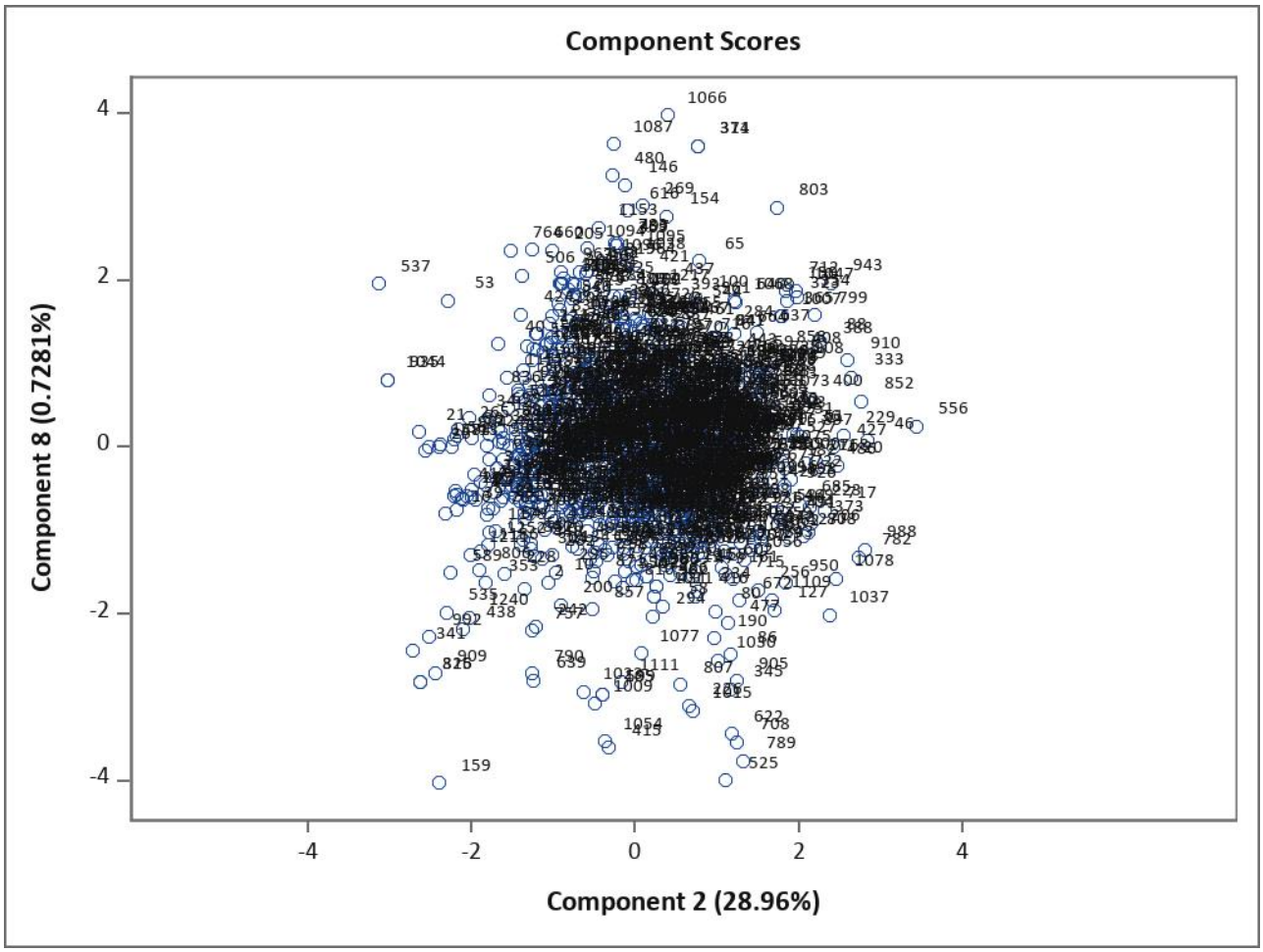

Component Scores
Component 7 (1.038%)
Component 4 (2.949%)
5
0
-5
-10
-5
0
5

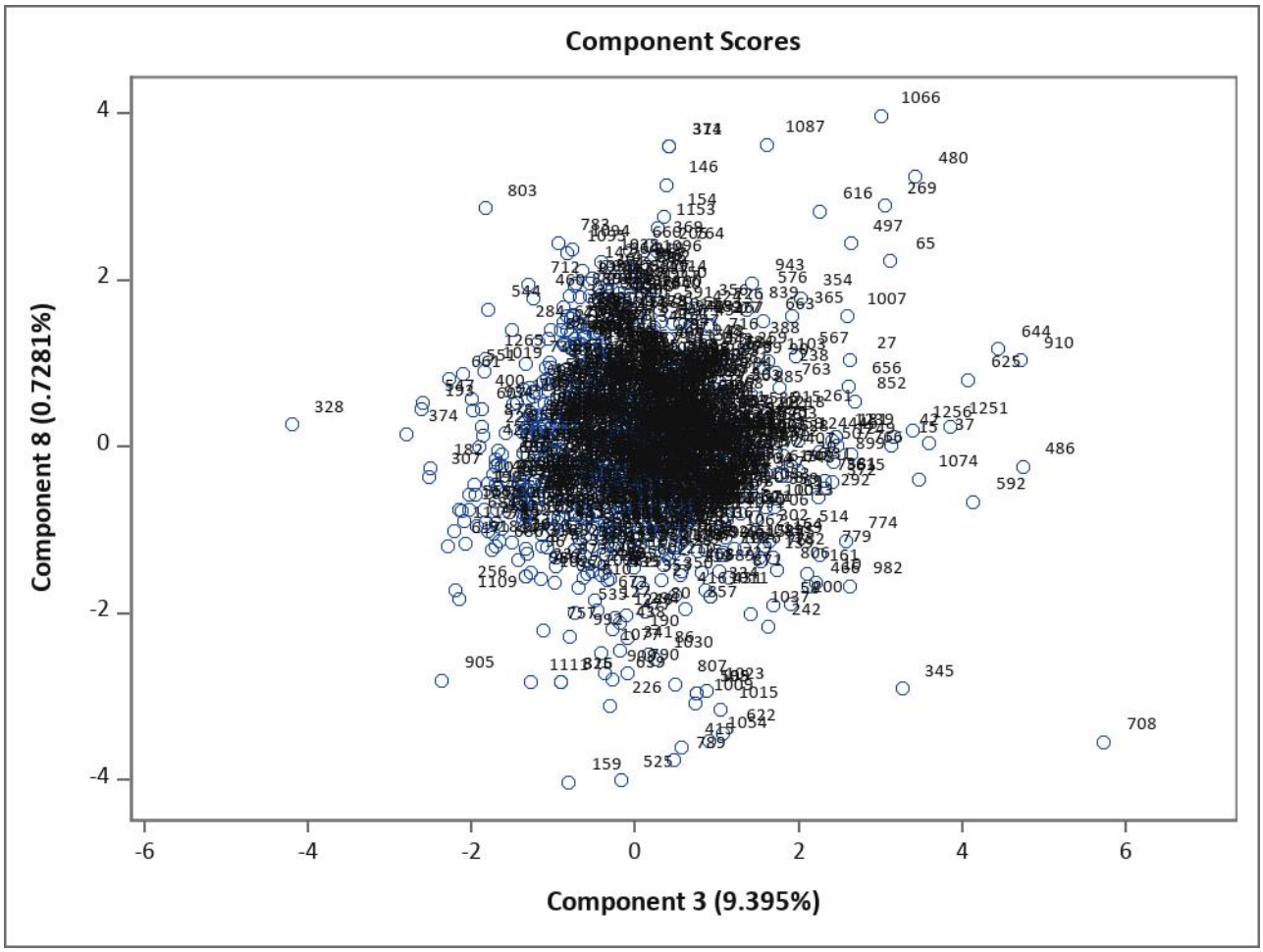


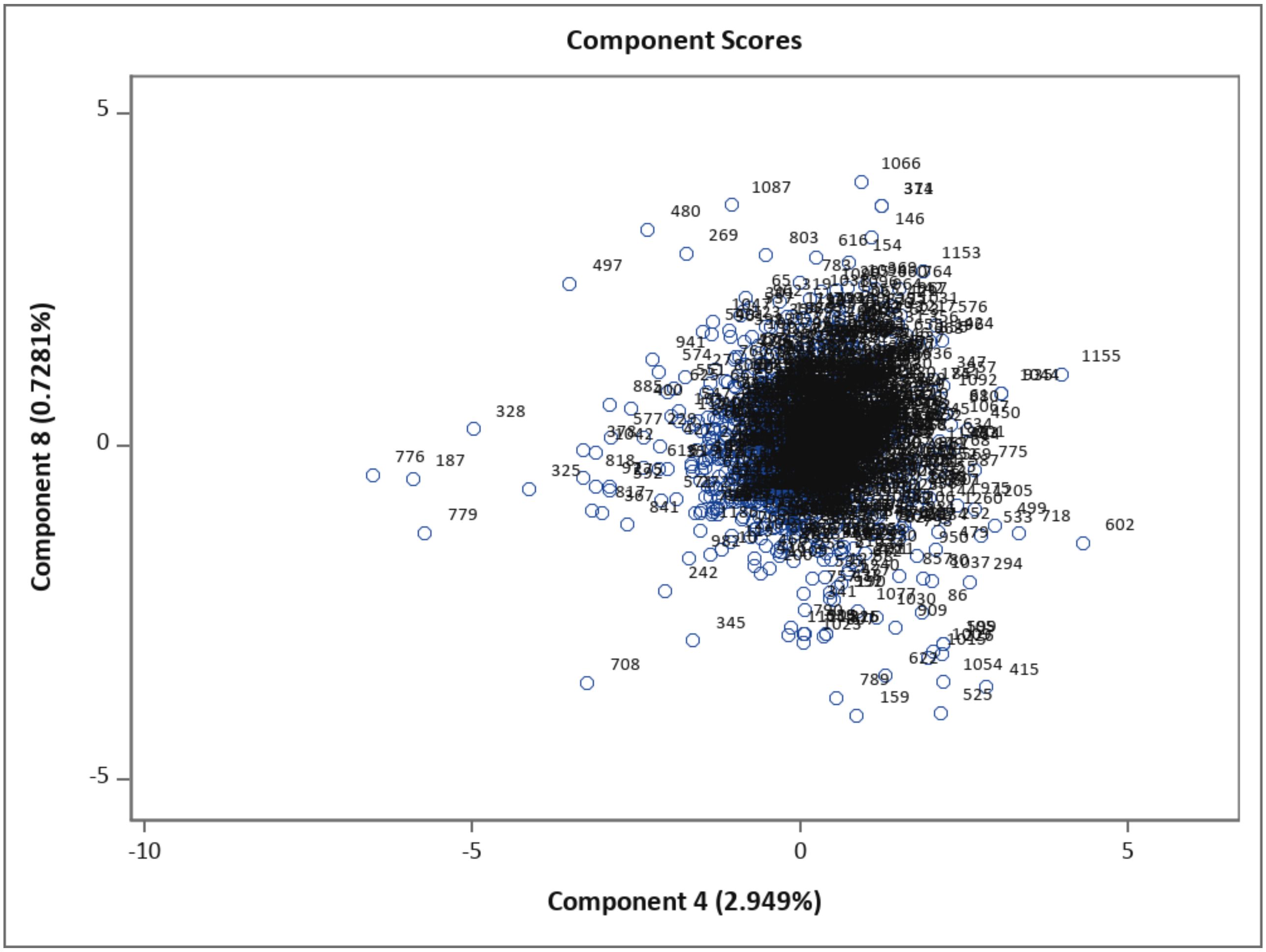
Component Scores
Component 7 (1.038%)
Component 5 (2.649%)
5
0
-5
-5
0
5
556
1023
323
817
717
187
625
20
708
149
142
567
975
374
964
667
785
574
480

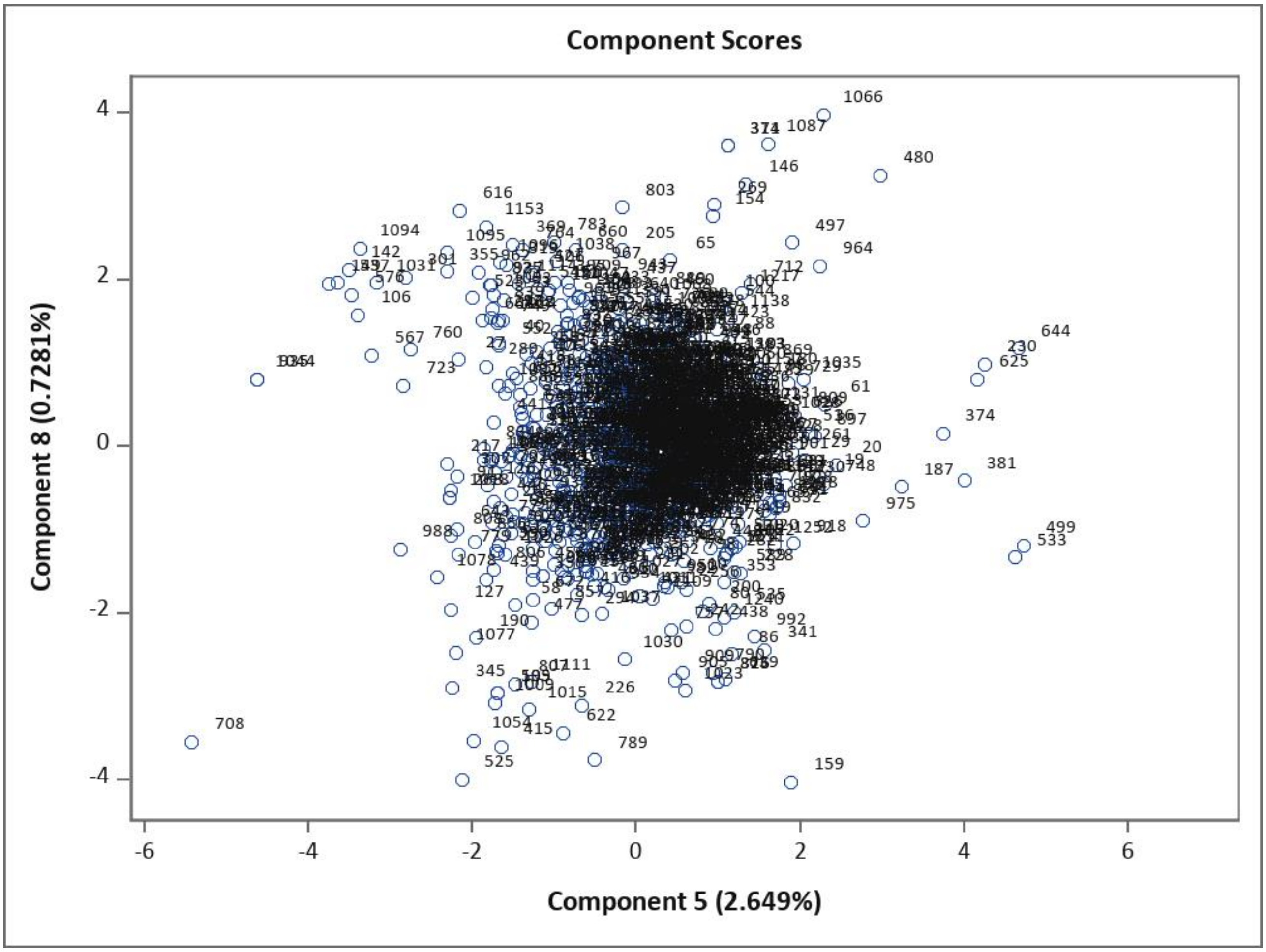

Component Scores
Component 7 (1.038%)
Component 6 (1.212%)
5
0
-5
-10
-5
0
5

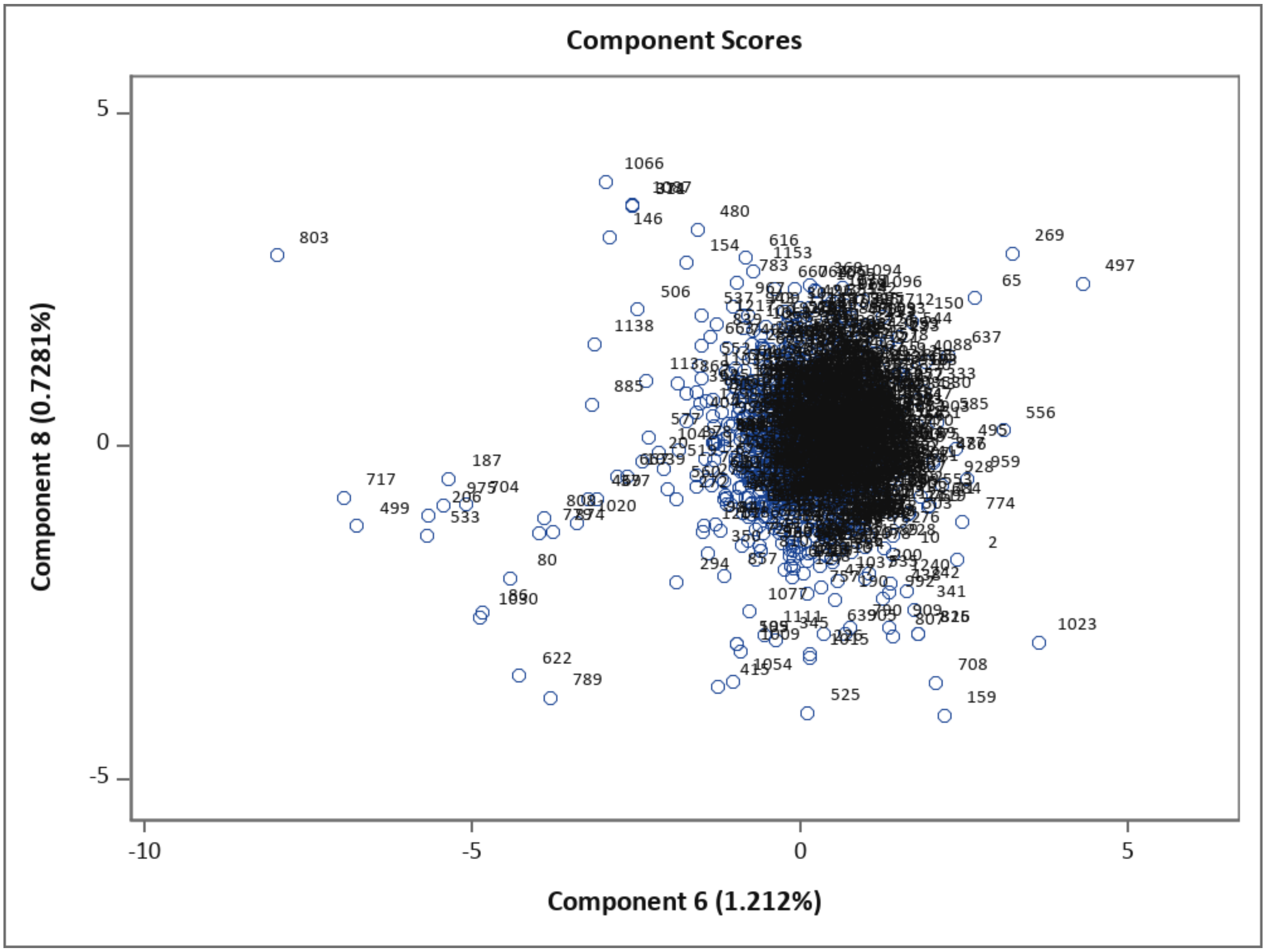

Component Scores
Component 8 (0.7281%)
Component 1 (52.56%)
4
2
0
-2
-4
-2
0
2
4
6
8

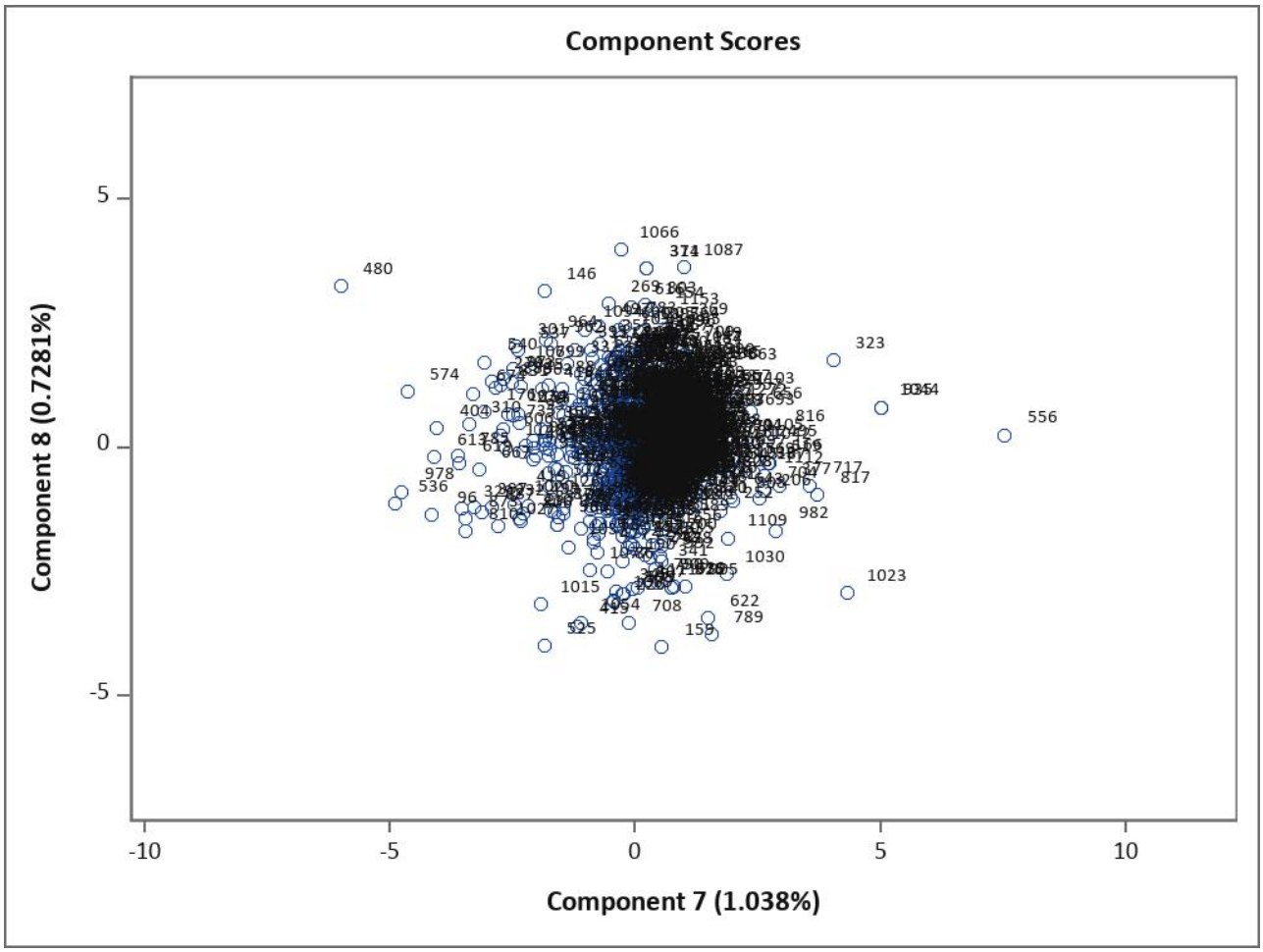

Component Scores
Component 8 (0.7281%)
Component 2 (28.96%)
4
2
0
-2
-4
-4
-2
0
2
4

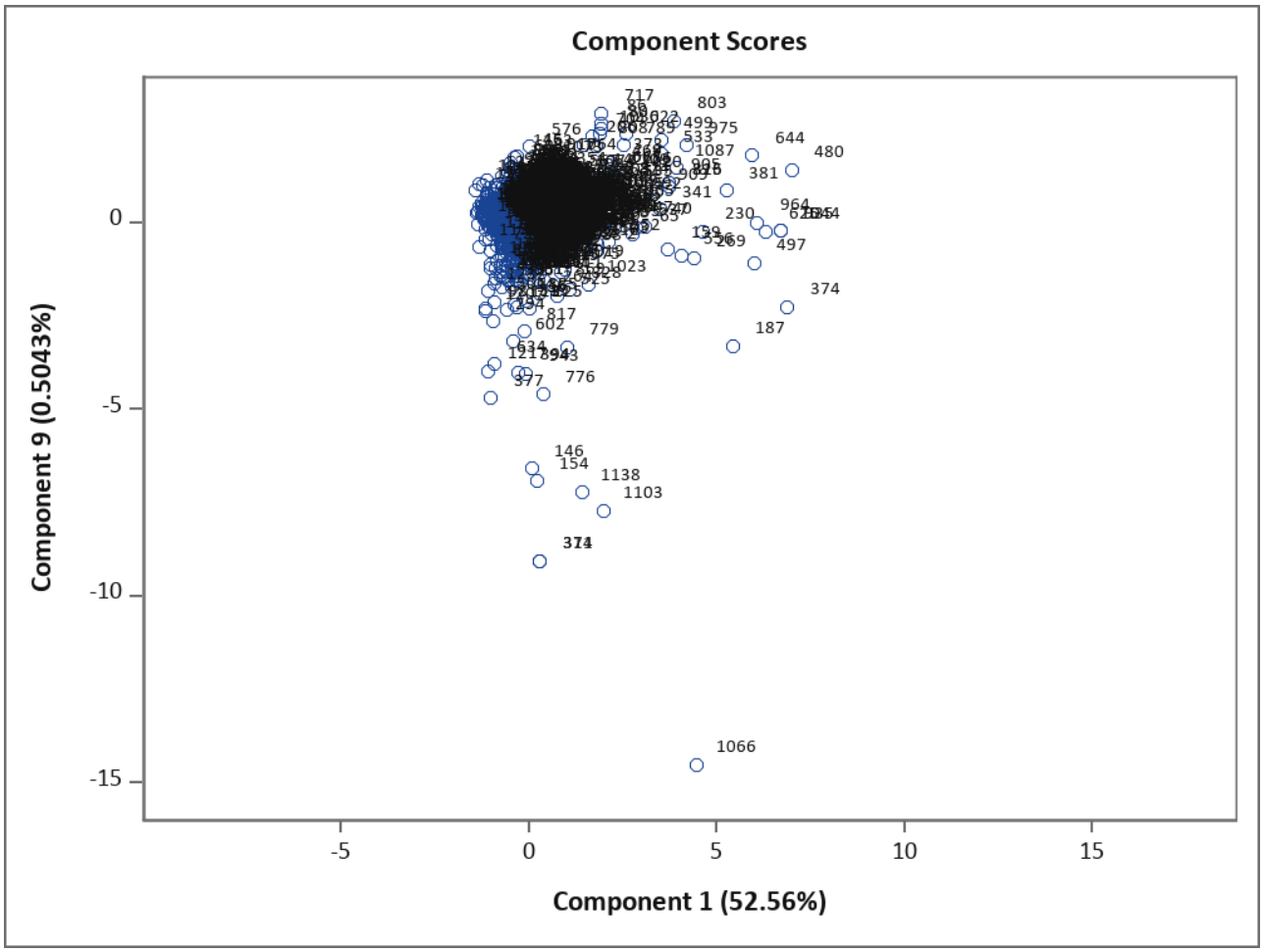

Component Scores
Component 8 (0.7281%)
Component 3 (9.395%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
1066
1087
480
269
616
803
146
154
1153
497
65
354
1007
644
910
625
27
656
852
328
374
1256
1251
486
1074
592
774
982
345
708
905
226
159
525

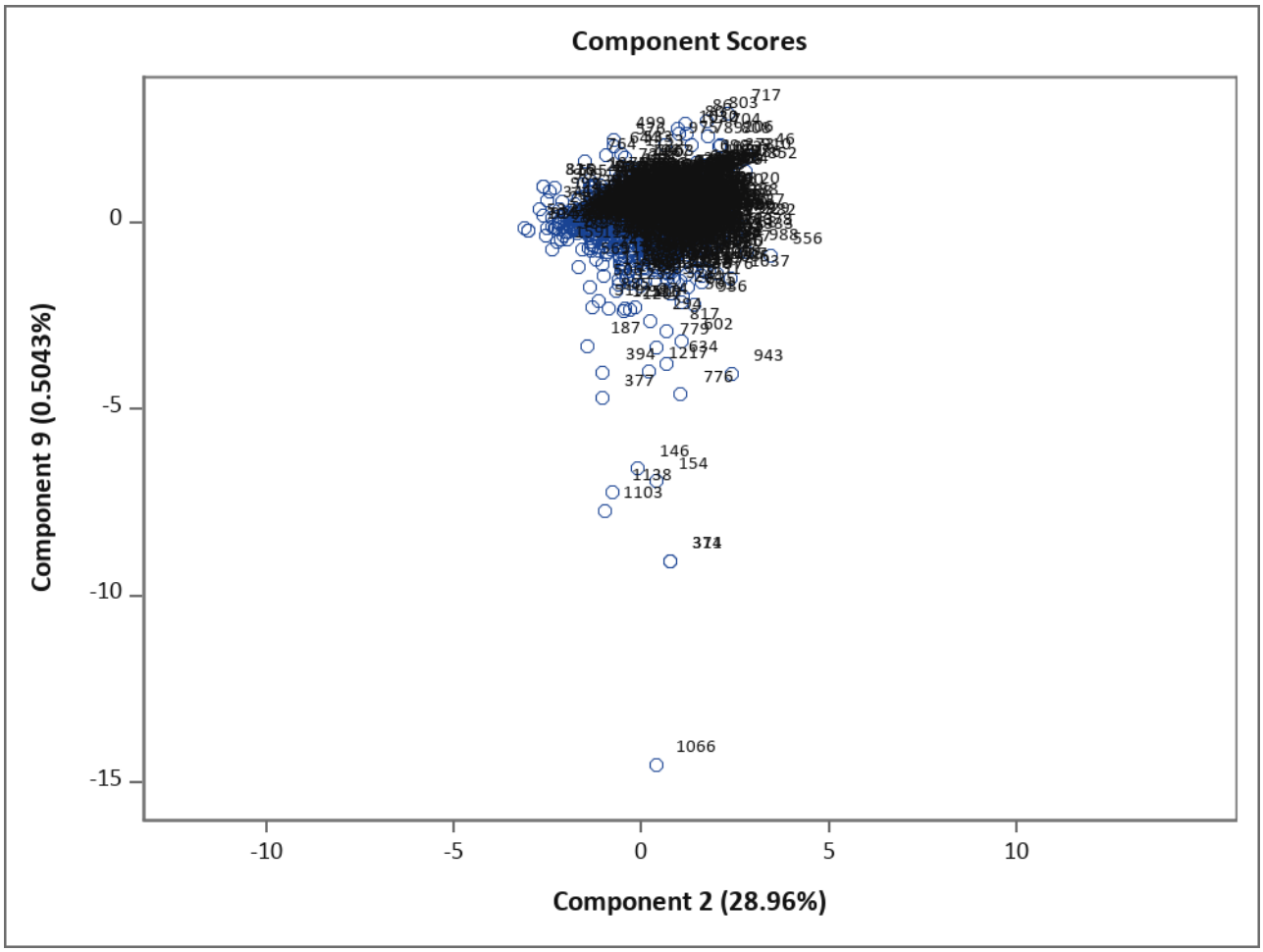

Component Scores
Component 8 (0.7281%)
Component 4 (2.949%)
5
0
-5
-10
-5
0
5

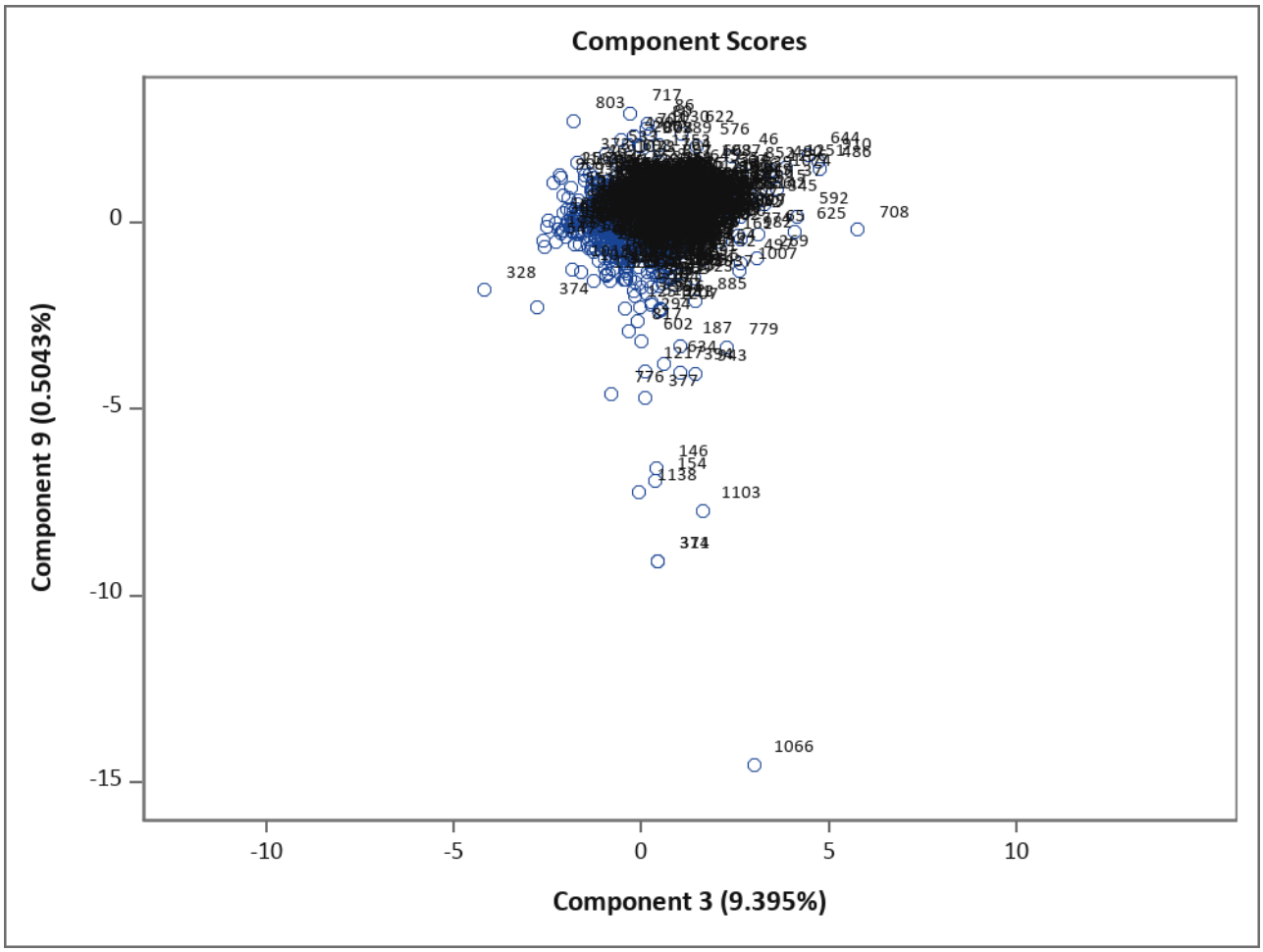

Component Scores
Component 8 (0.7281%)
Component 5 (2.649%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6

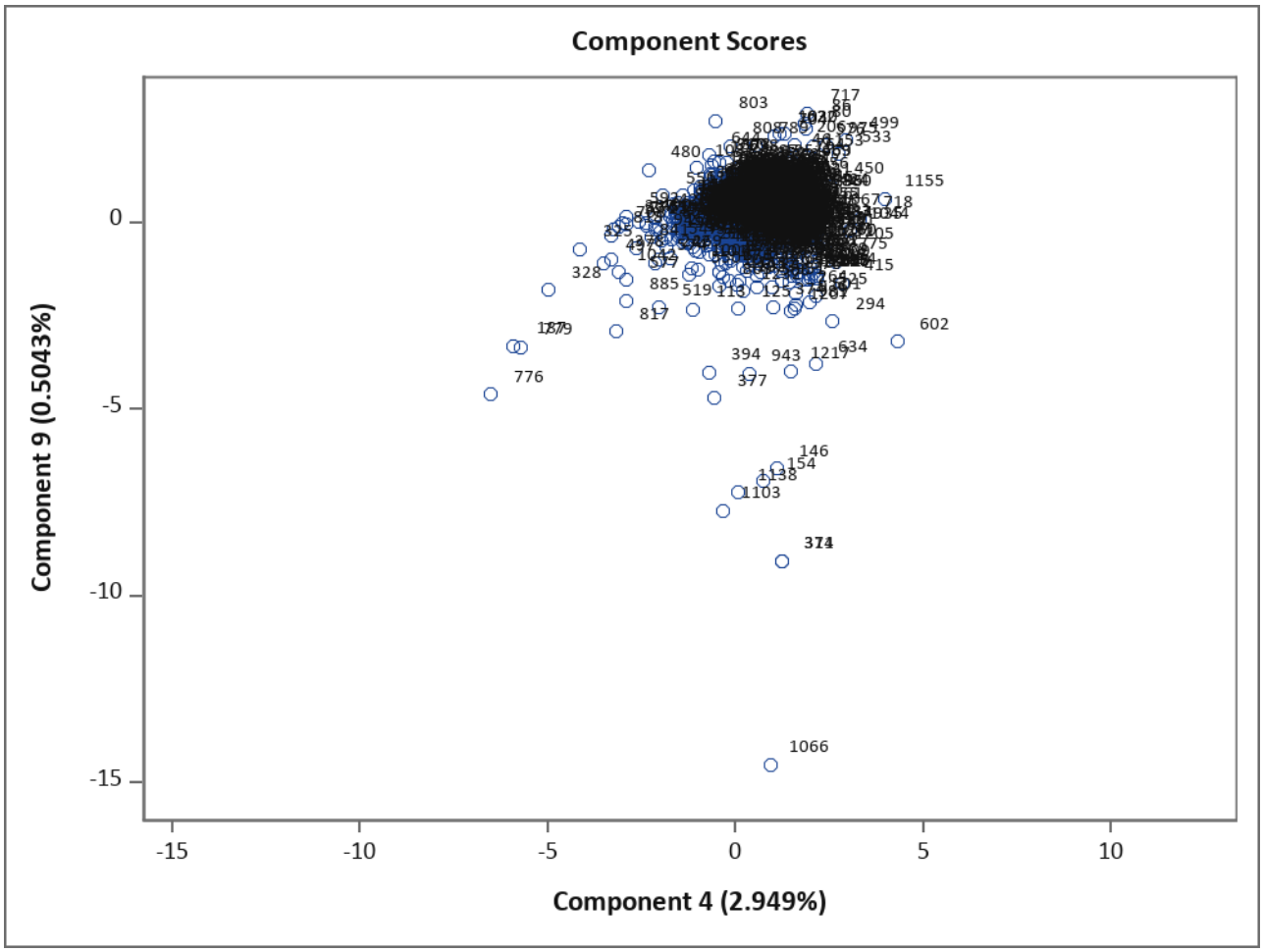

Component Scores
Component 8 (0.7281%)
Component 6 (1.212%)
5
0
-5
-10
-5
0
5

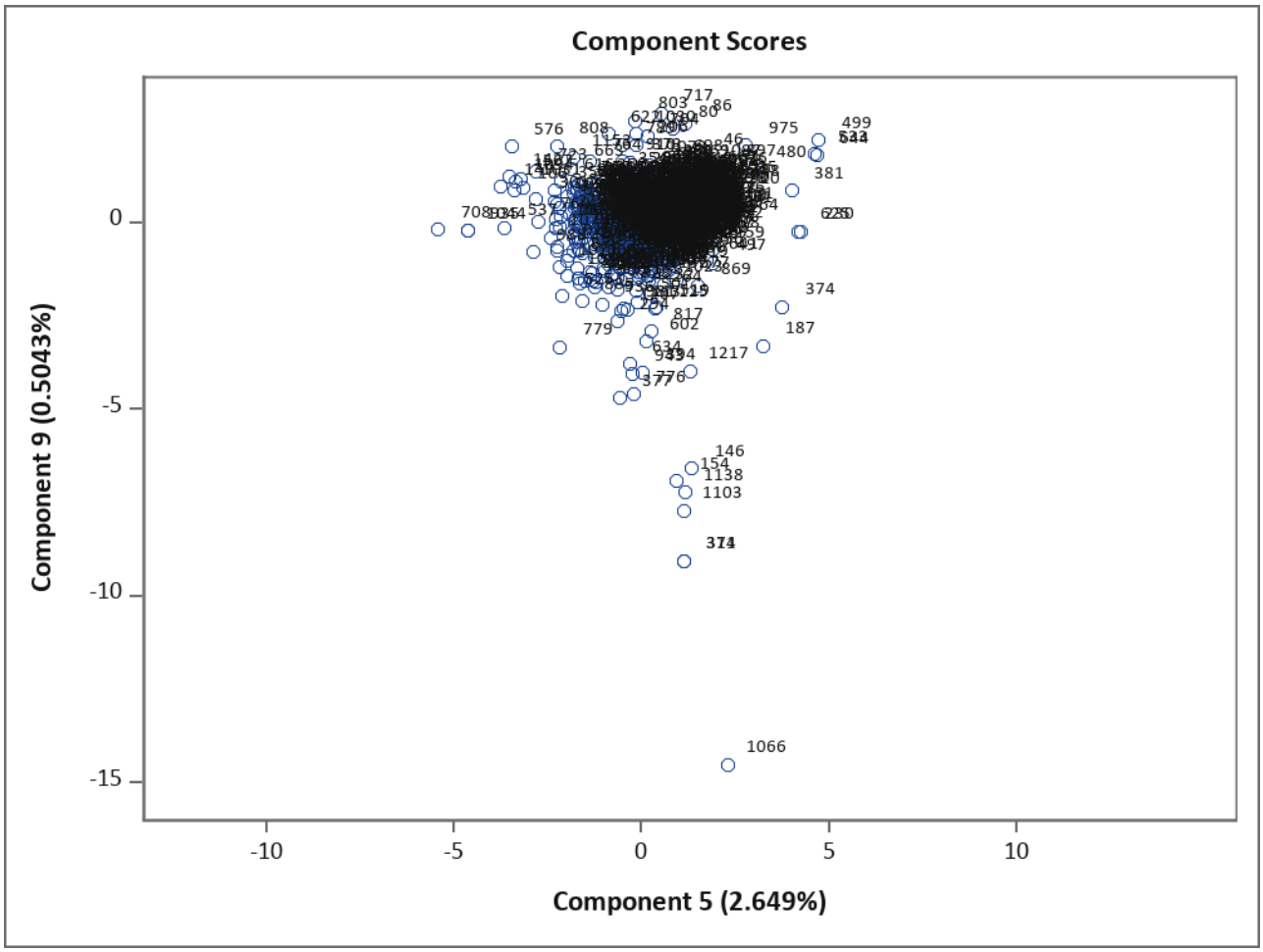

Component Scores
Component 8 (0.7281%)
Component 7 (1.038%)
5
0
-5
-10
-5
0
5
10
1066
1087
480
146
323
574
556
816
978
536
96
982
1109
1030
1023
1015
622
789
708
525
153
817

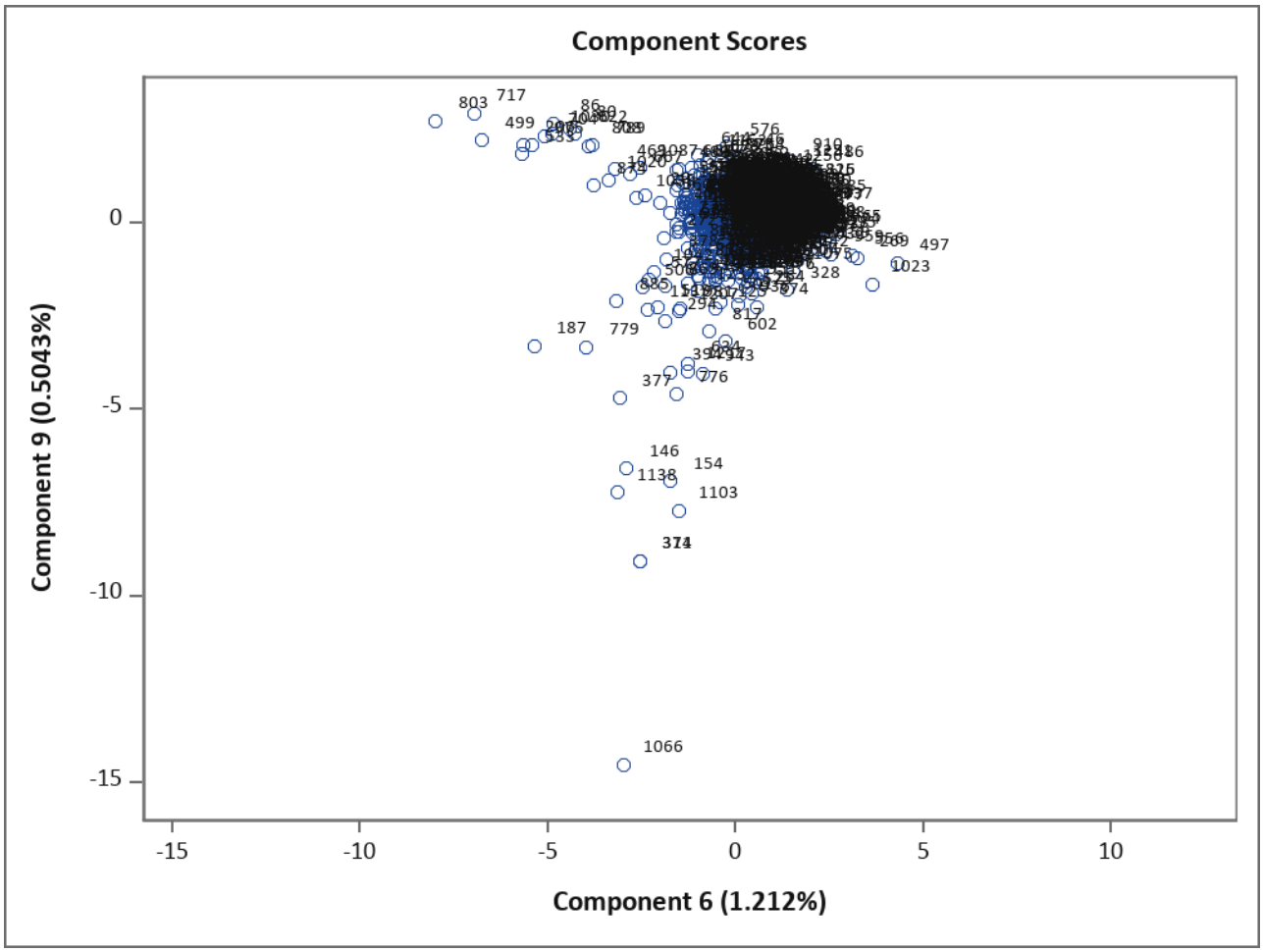

Component Scores
Component 9 (0.5043%)
Component 1 (52.56%)
0
-5
-10
-15
-5
0
5
10
15
717
803
576
499
975
644
533
1087
480
381
341
964
230
497
374
187
779
602
817
776
146
154
1138
1103
1066

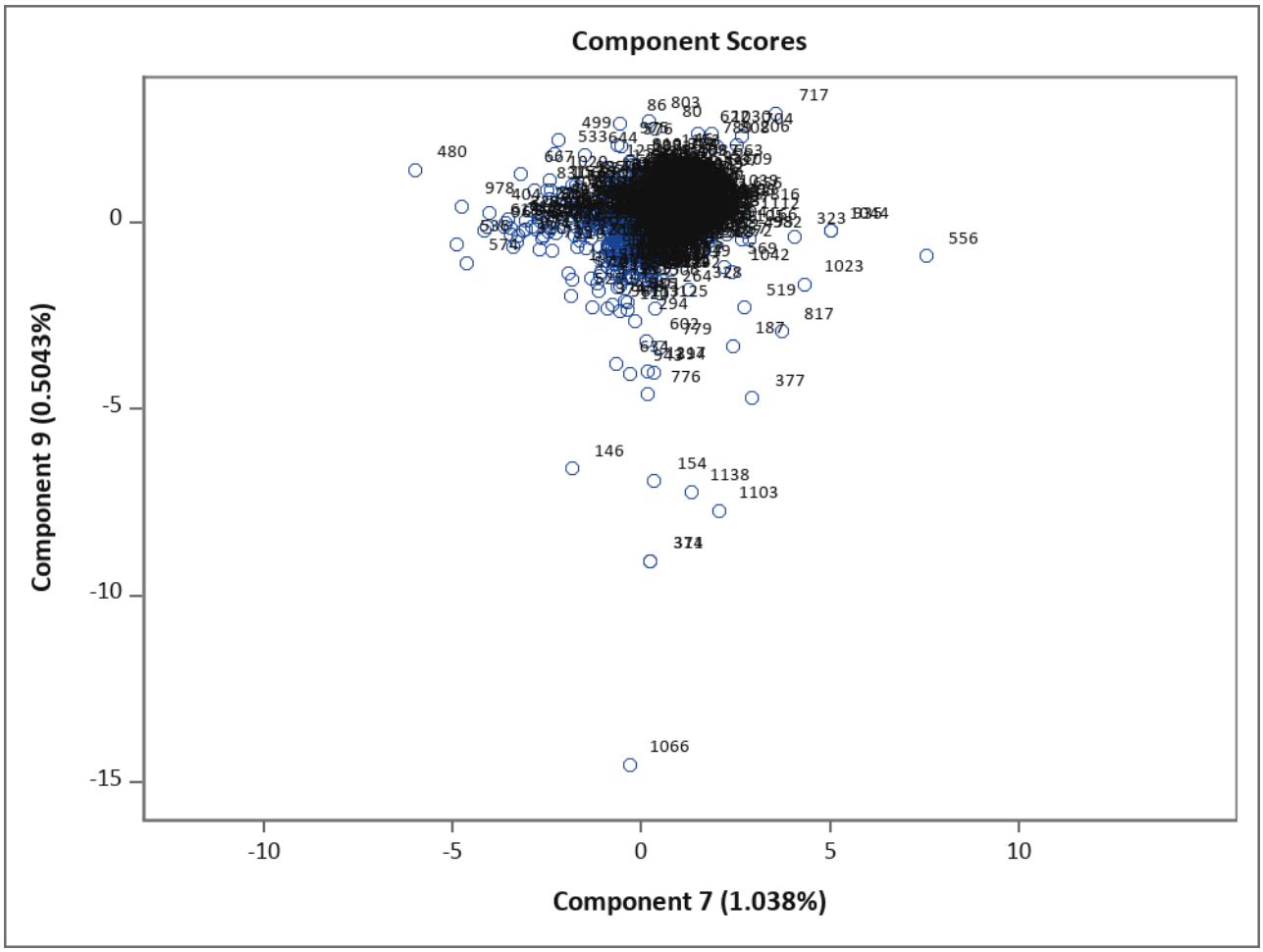

Component Scores
Component 9 (0.5043%)
Component 2 (28.96%)
0
-5
-10
-15
-10
-5
0
5
10
717
803
499
556
1037
187
779
902
1634
394
1217
943
377
776
146
154
1138
1103
374
1066

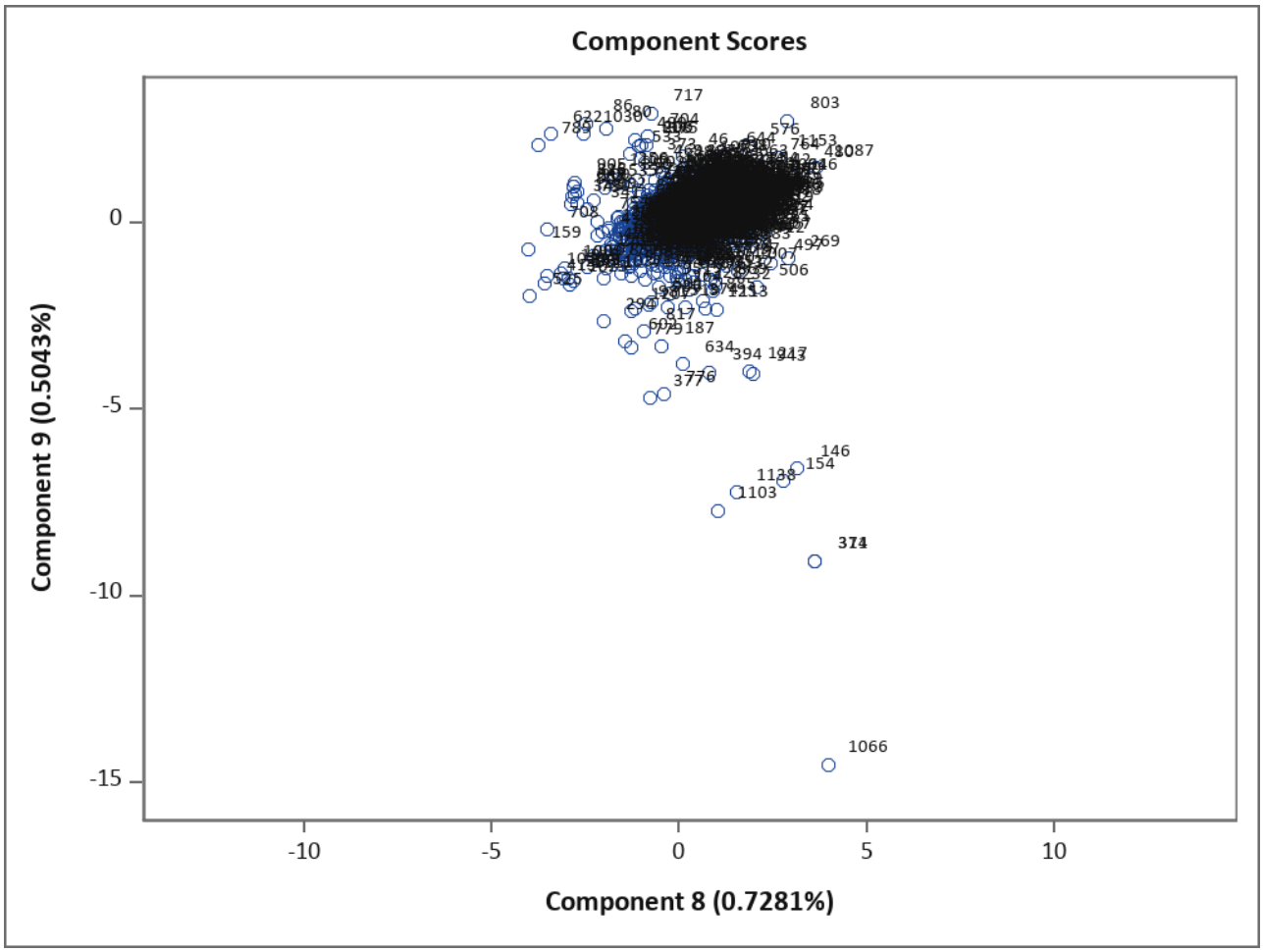

Component Scores
Component 9 (0.5043%)
Component 3 (9.395%)
0
-5
-10
-15
-10
-5
0
5
10
803
717
576
46
644
592
625
708
328
374
885
187
779
602
776
377
146
154
1138
1103
1066

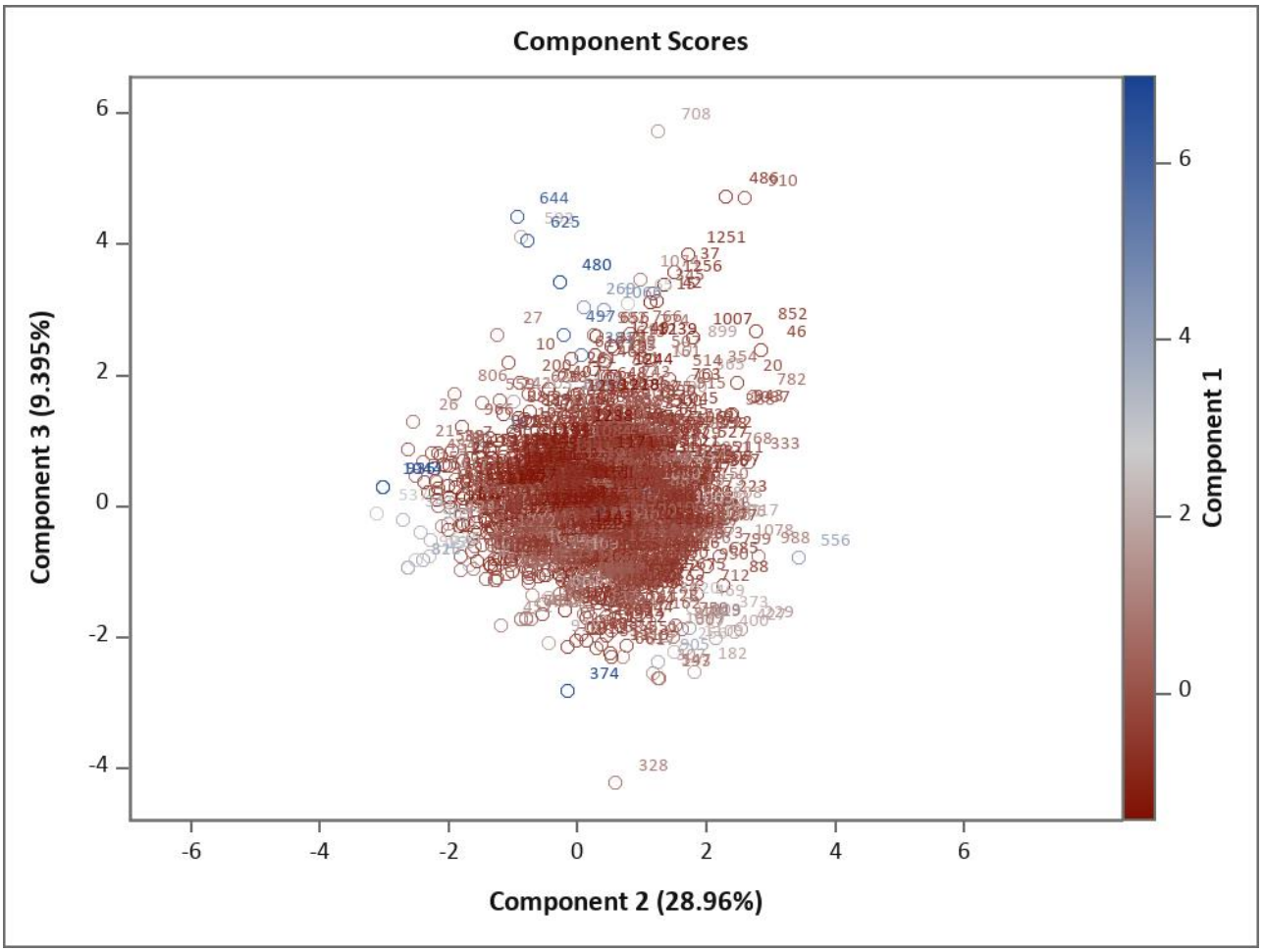

Component Scores
Component 9 (0.5043%)
Component 4 (2.949%)
0
-5
-10
-15
-15
-10
-5
0
5
10
803
717
499
533
480
450
1155
218
328
885
519
817
294
602
634
394
943
1217
776
377
146
154
1138
1103
1066


Component Scores
Component 9 (0.5043%)
Component 5 (2.649%)
0
-5
-10
-15
-10
-5
0
5
10
717
803
86
80
576
808
975
499
381
374
187
1217
779
602
817
146
154
1138
1103
1066

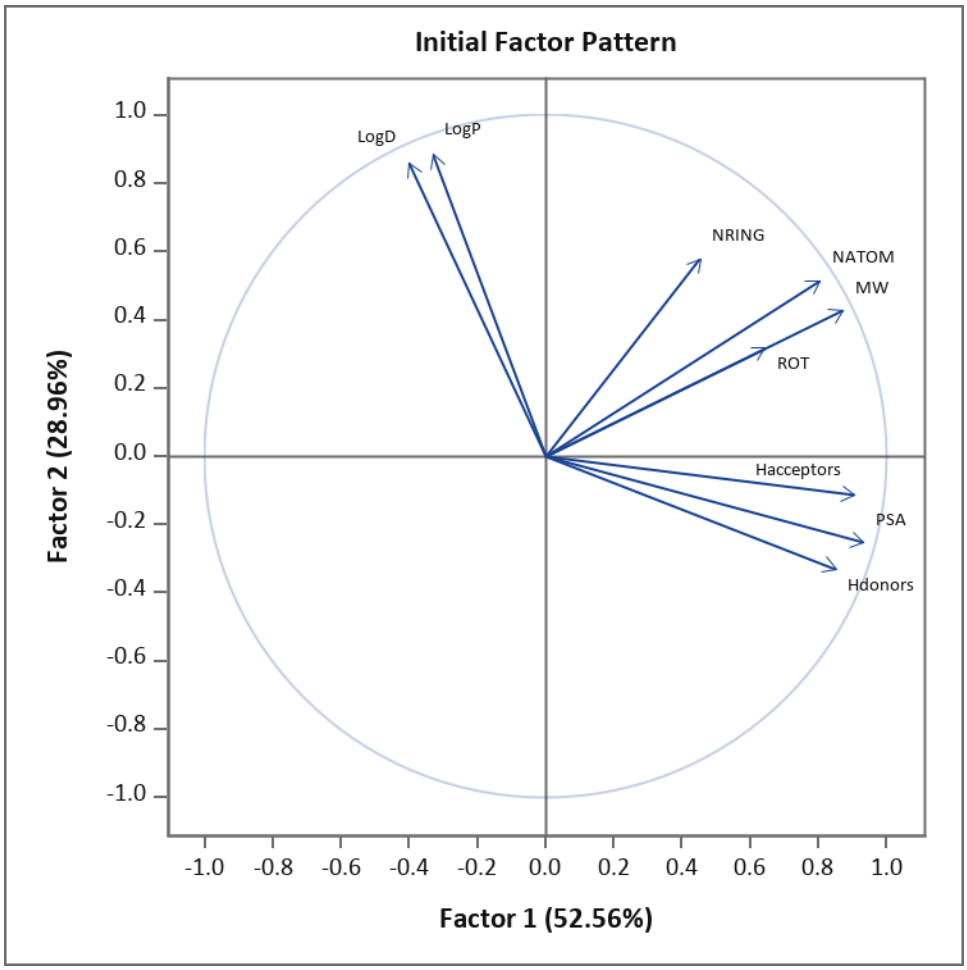

Component Scores
Component 9 (0.5043%)
Component 6 (1.212%)
0
-5
-10
-15
-15
-10
-5
0
5
10
803
717
499
576
497
1023
328
187
779
602
377
776
146
154
1138
1103
374
1066

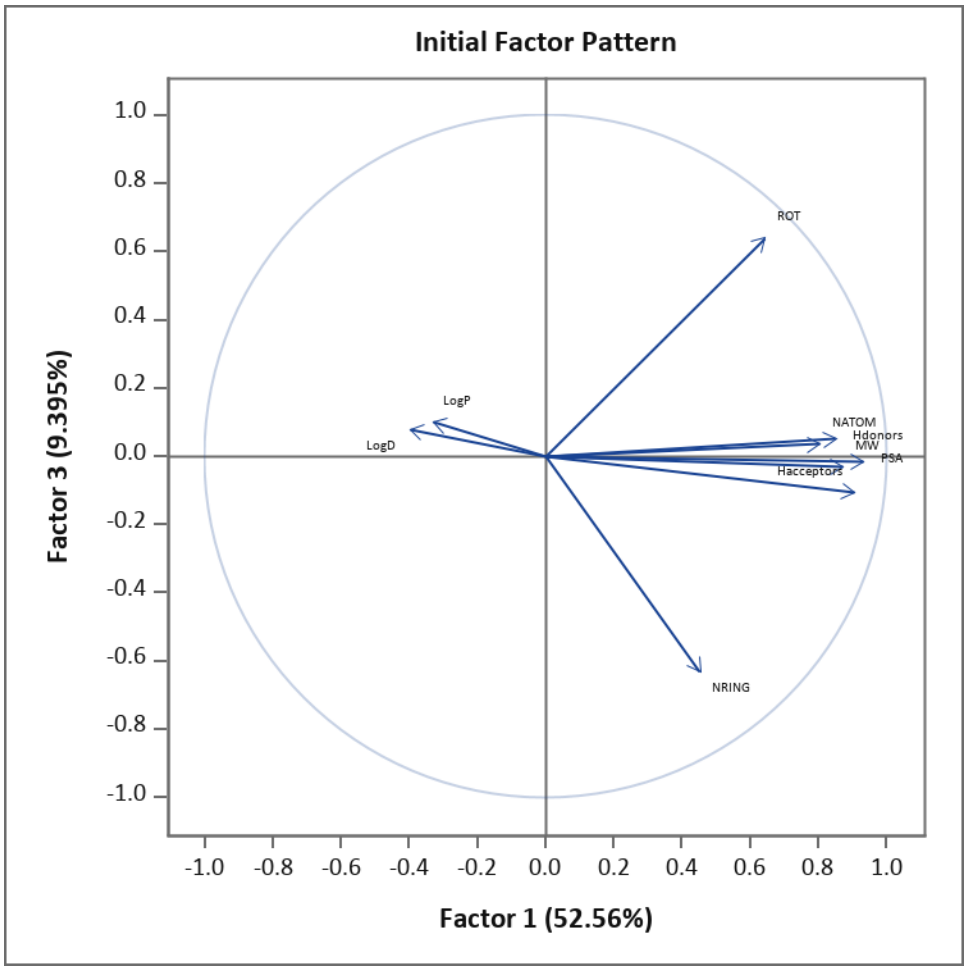

Component Scores
Component 9 (0.5043%)
Component 7 (1.038%)
0
-5
-10
-15
-10
-5
0
5
10
717
86
803
80
499
533
644
480
978
323
556
1023
519
817
187
602
779
776
377
146
154
1138
1103
1066

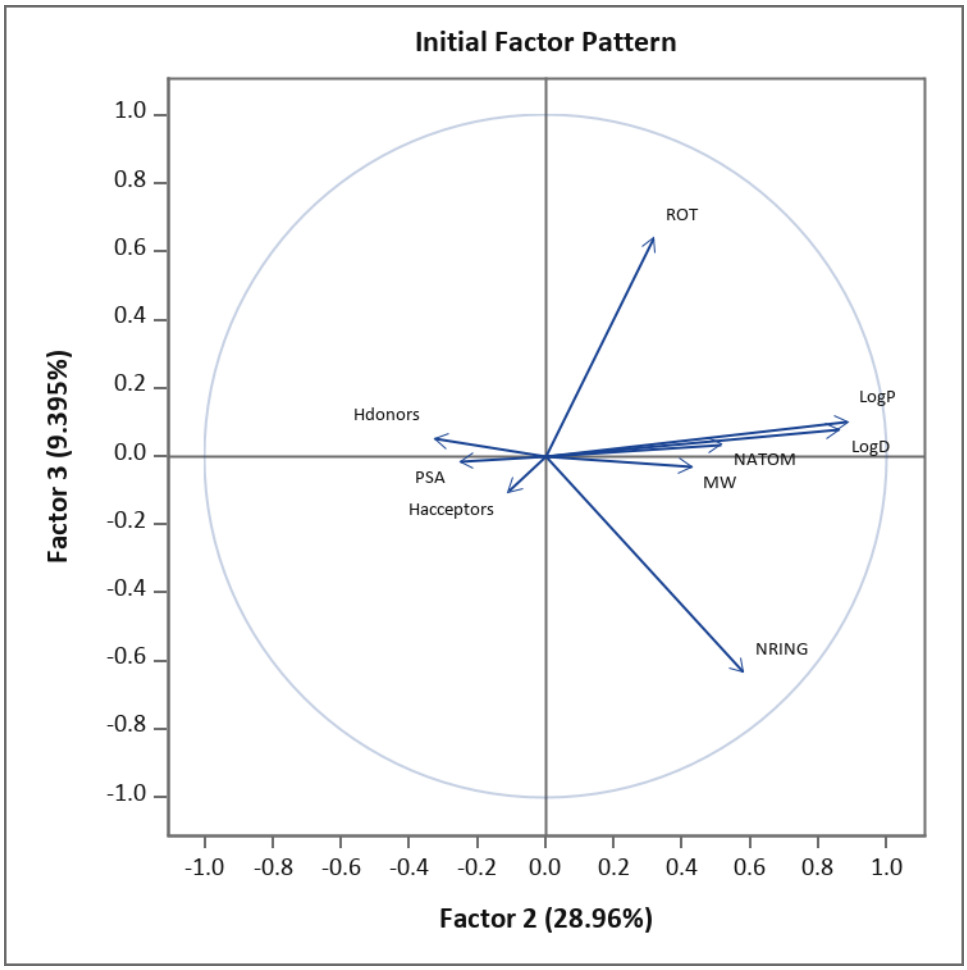

Component Scores
Component 9 (0.5043%)
Component 8 (0.7281%)
0
-5
-10
-15
-10
-5
0
5
10
717
803
576
1066
146
154
1103
374

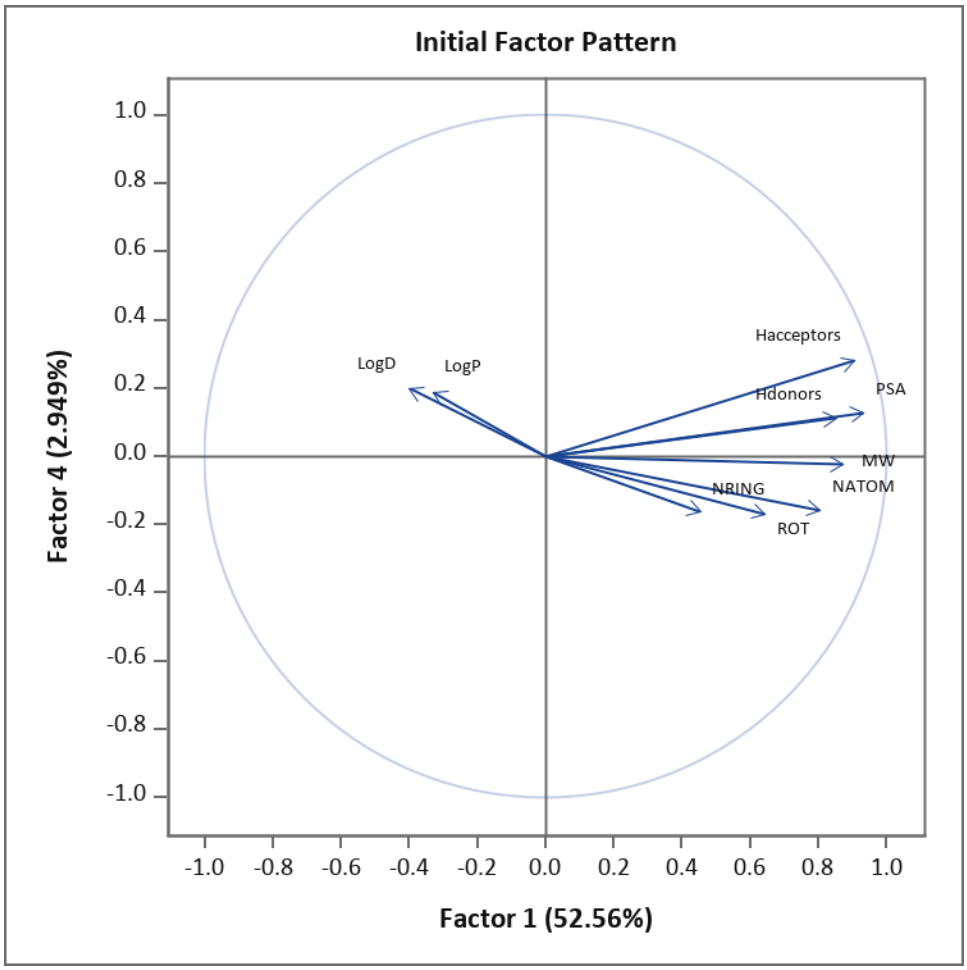

Component Scores
Component 3 (9.395%)
Component 2 (28.96%)
Component 1
6
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6
708
644
625
480
1251
1256
1007
852
46
782
556
374
328

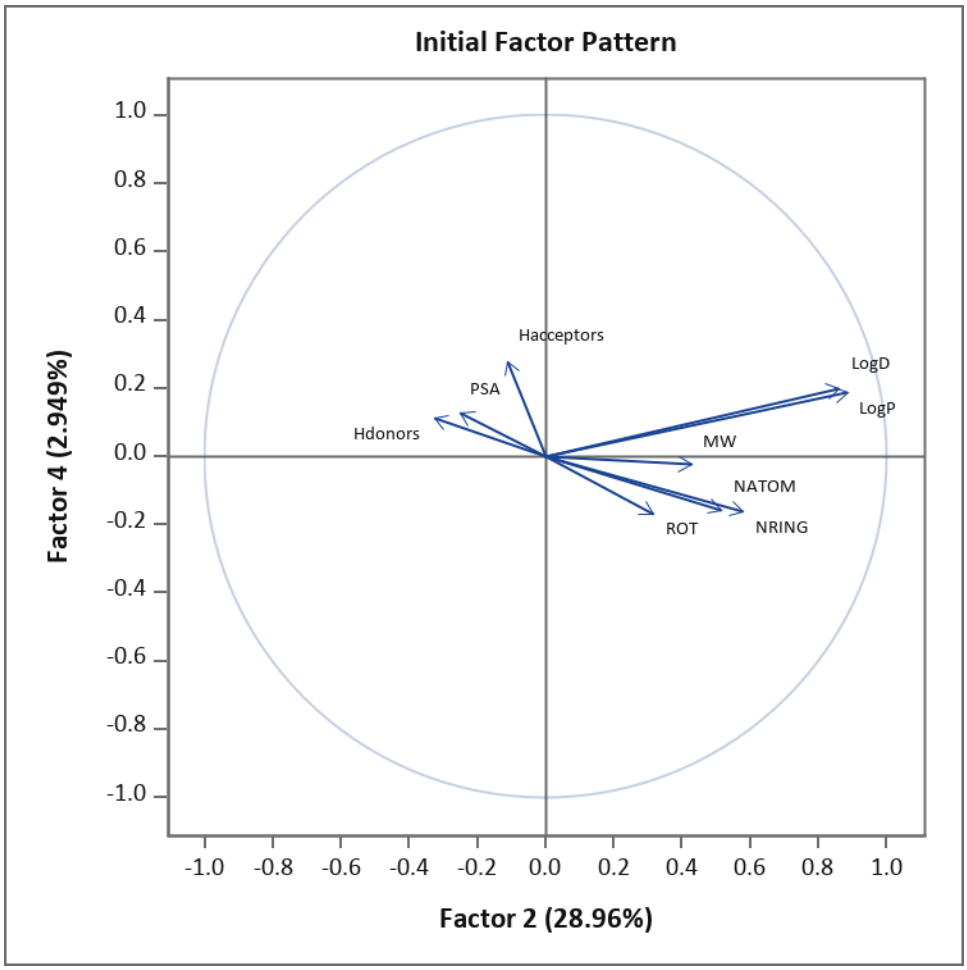
5. Loading plots for all observations

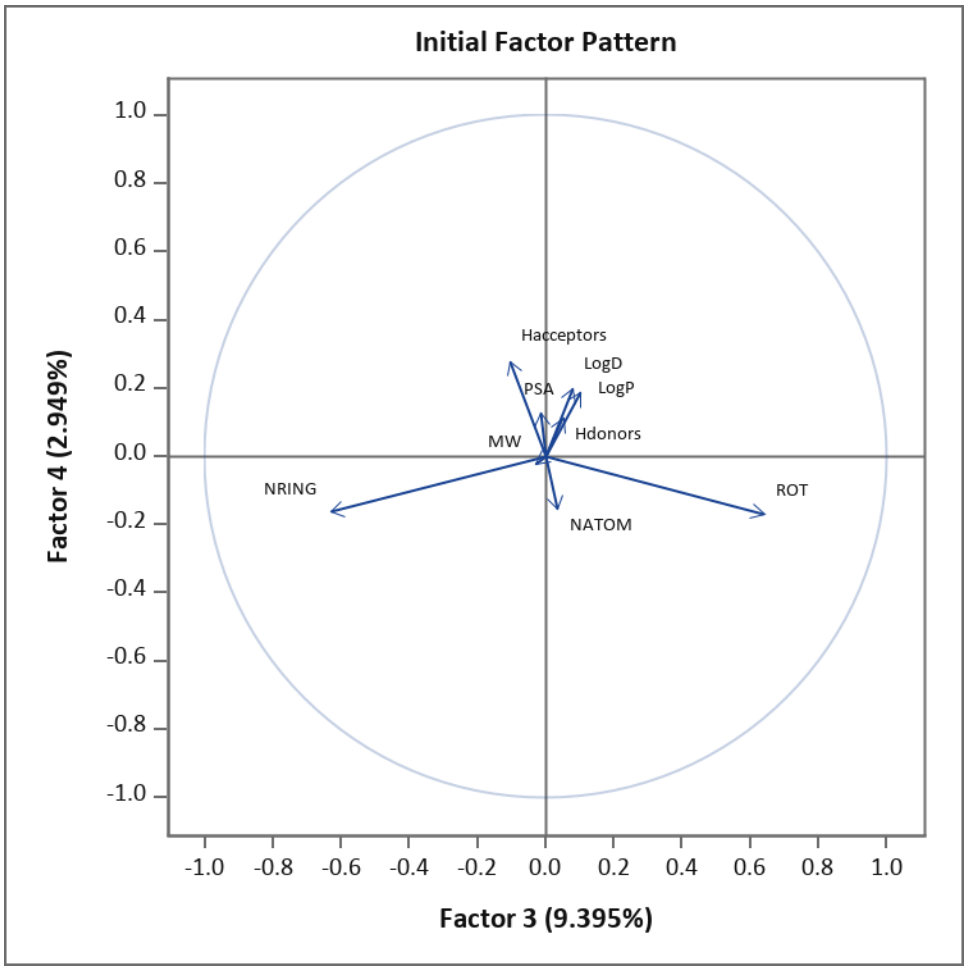


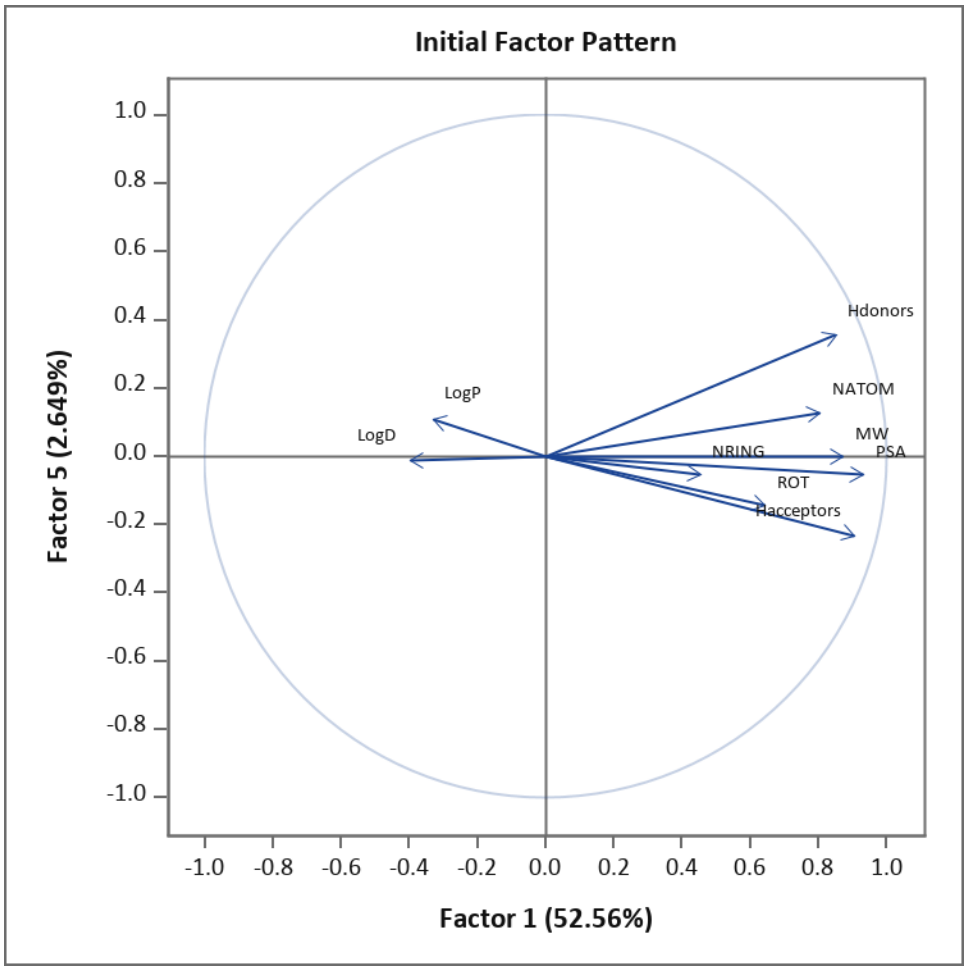

Initial Factor Pattern
Factor 3 (9.395%)
Factor 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
ROT
LogP
LogD
NATOM
Hdonors
MW
PSA
Hacceptors
NRING

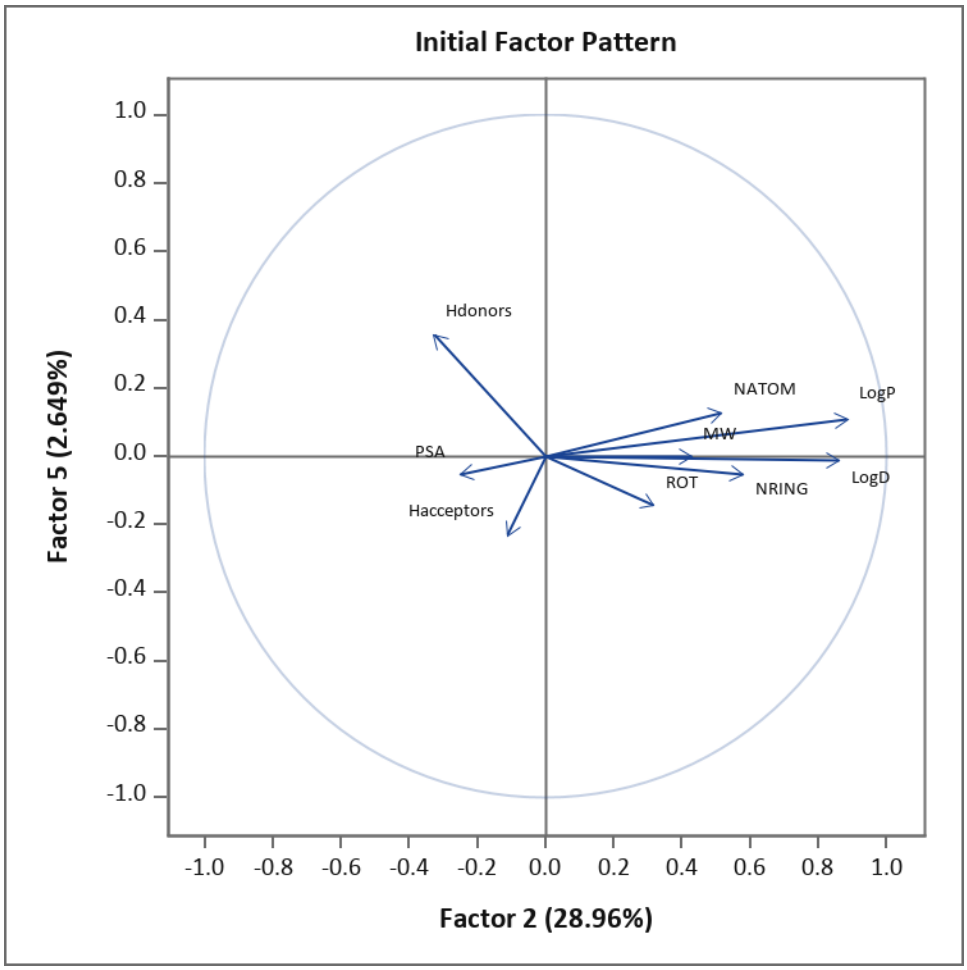

Initial Factor Pattern
Factor 3 (9.395%)
Factor 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
ROT
LogP
LogD
NATOM
MW
NRING
Hdonors
PSA
Hacceptors

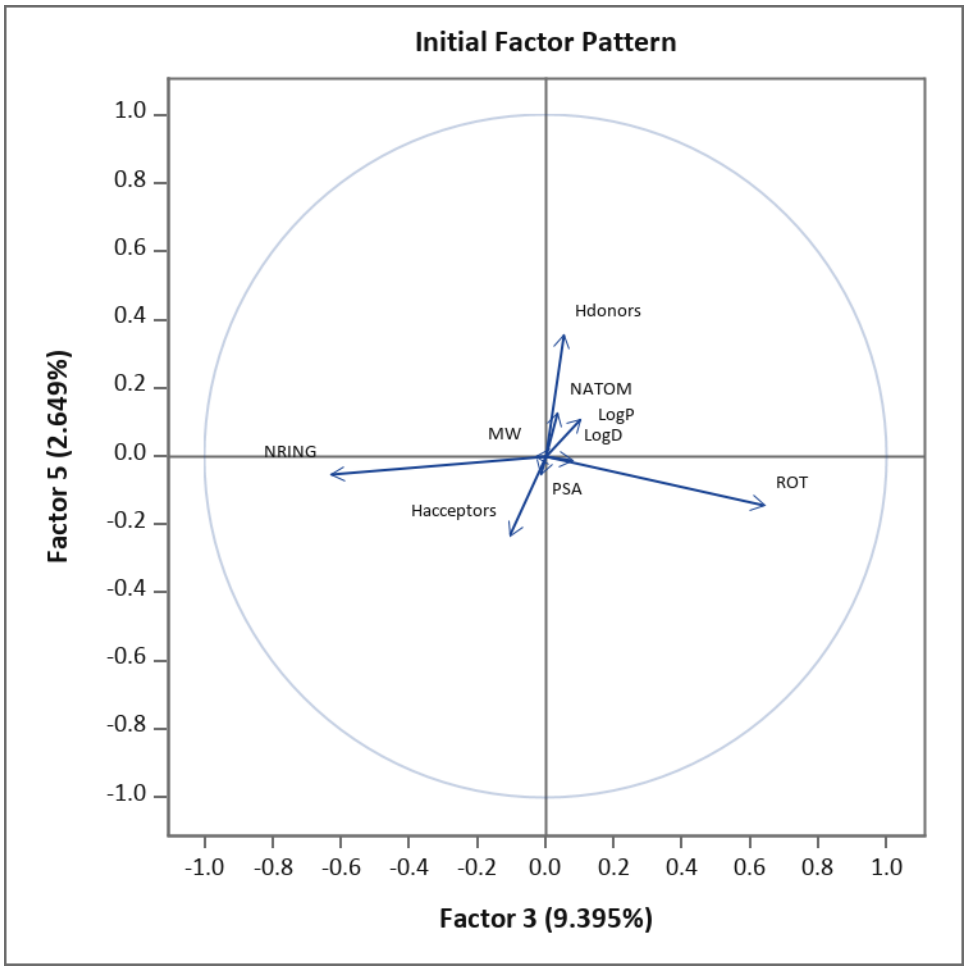

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 4 (2.949%)
LogD
LogP
Hacceptors
Hdonors
PSA
MW
NATOM
NRING
ROT
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 1 (52.56%)

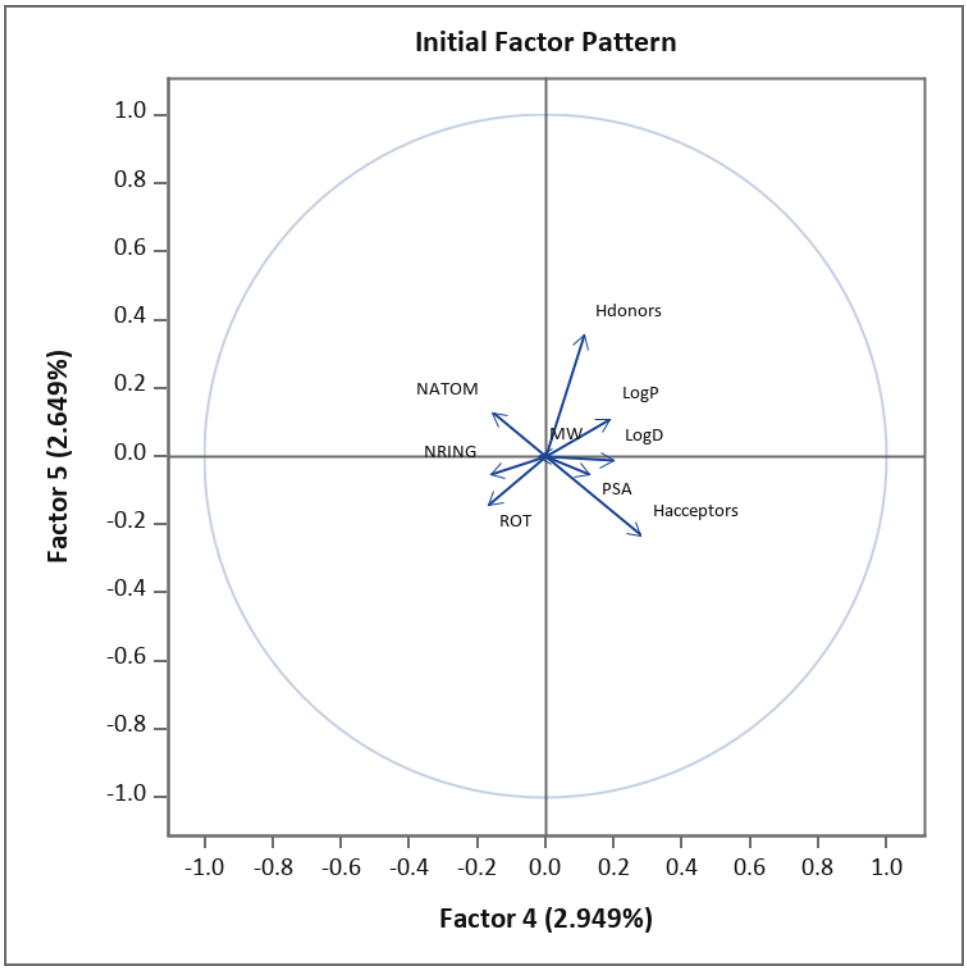

Initial Factor Pattern
Factor 4 (2.949%)
Factor 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hacceptors
PSA
Hdonors
LogD
LogP
MW
NATOM
NRING
ROT

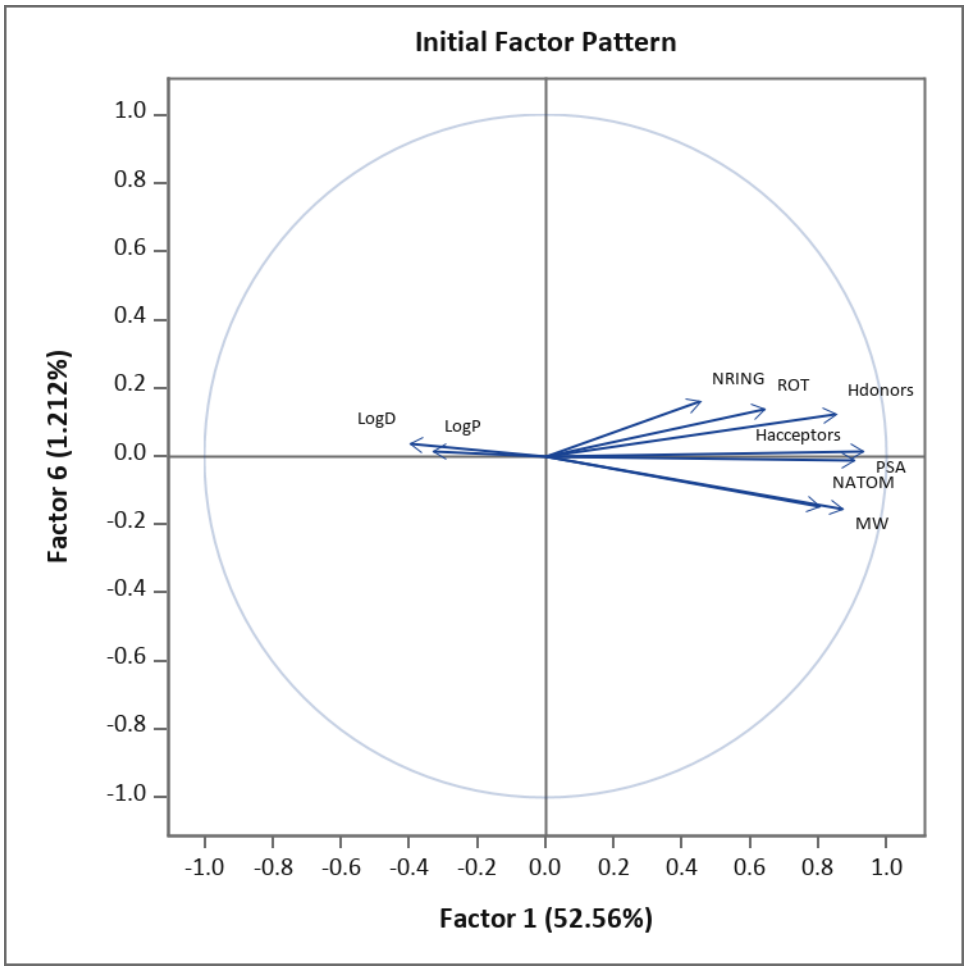

Initial Factor Pattern
Factor 4 (2.949%)
Factor 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
LogD
PSA
LogP
MW
Hdonors
NRING
ROT
NATOM

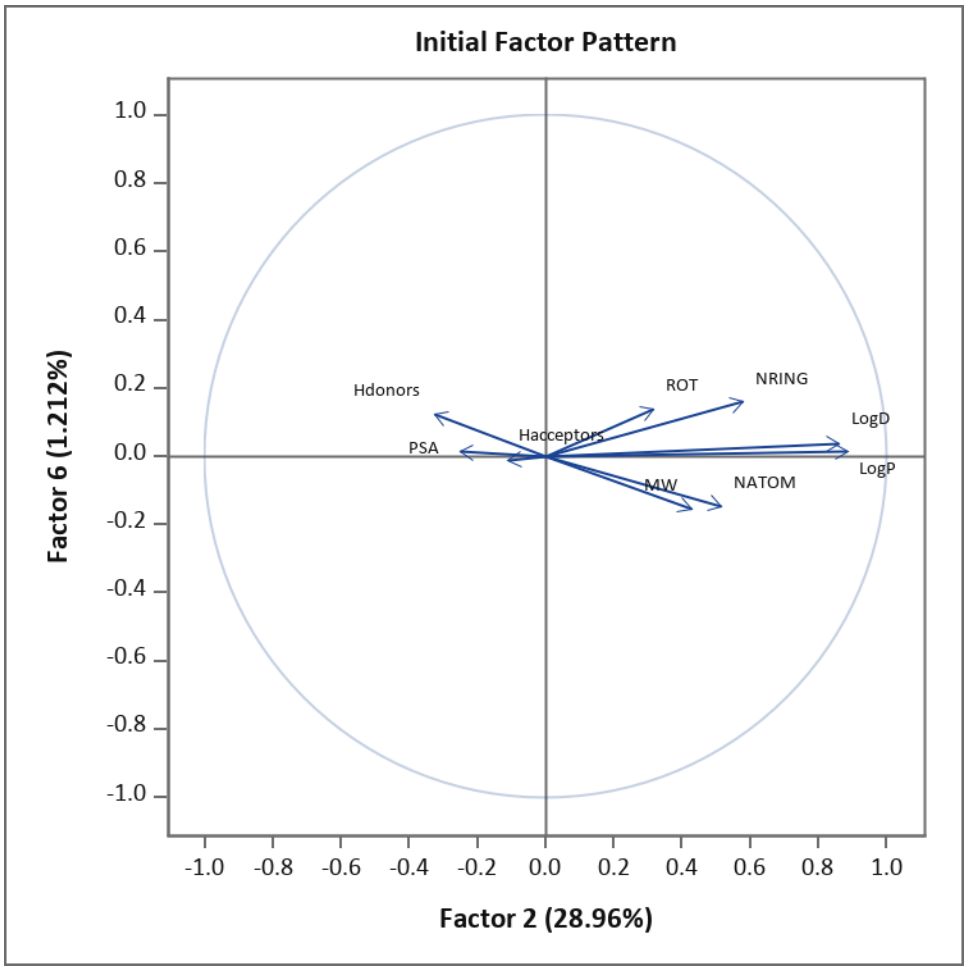

Initial Factor Pattern
Factor 5 (2.649%)
Factor 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
NATOM
LogP
LogD
MW
NRING
PSA
ROT
Hacceptors

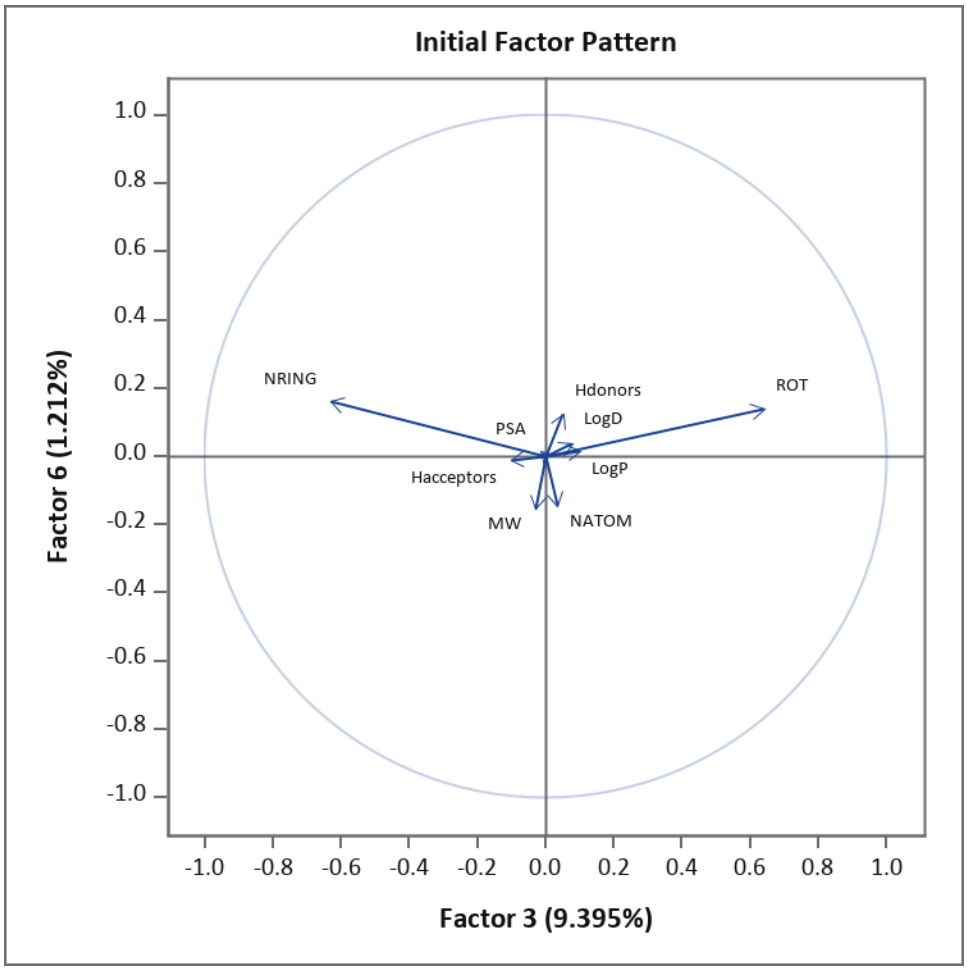

Initial Factor Pattern
Factor 5 (2.649%)
Factor 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
NATOM
LogP
MW
PSA
ROT
NRING
LogD
Hacceptors

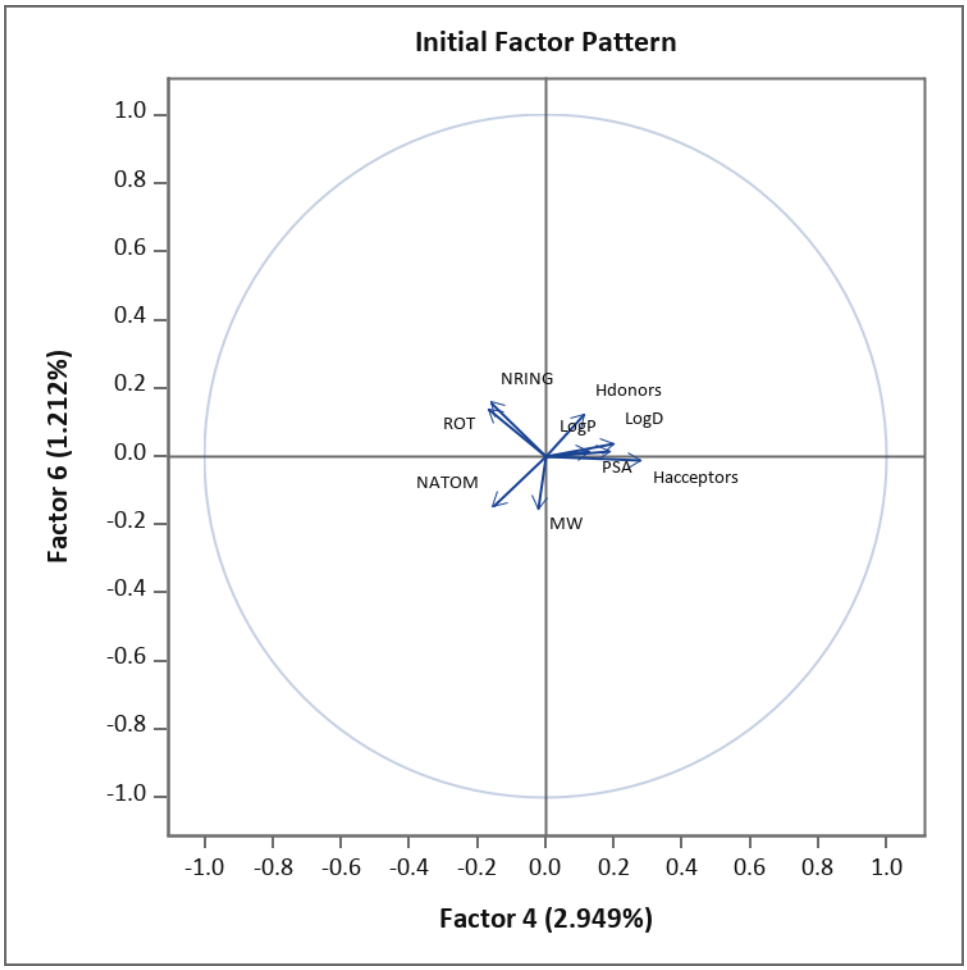

Initial Factor Pattern
Factor 5 (2.649%)
Factor 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hdonors
NATOM
LogP
LogD
MW
NRING
PSA
ROT
Hacceptors

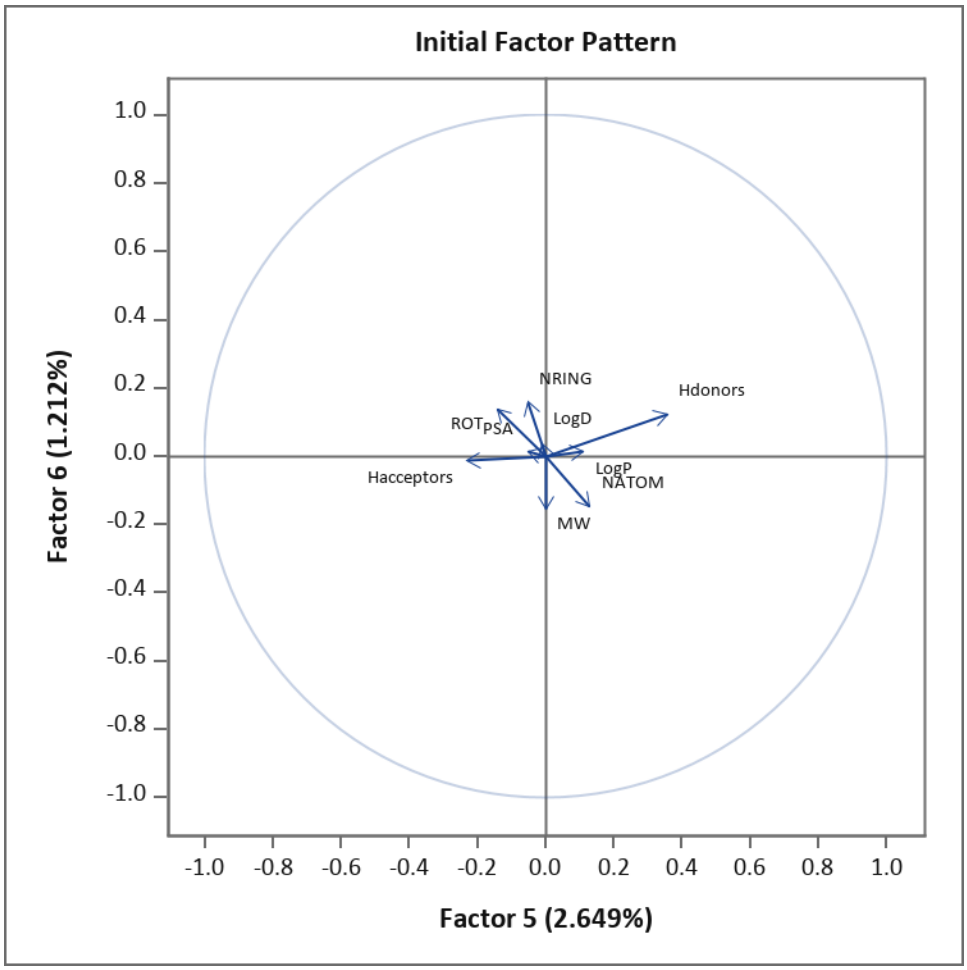

Initial Factor Pattern
Factor 5 (2.649%)
Factor 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
NATOM
LogP
MW
LogD
NRING
PSA
Hacceptors
ROT

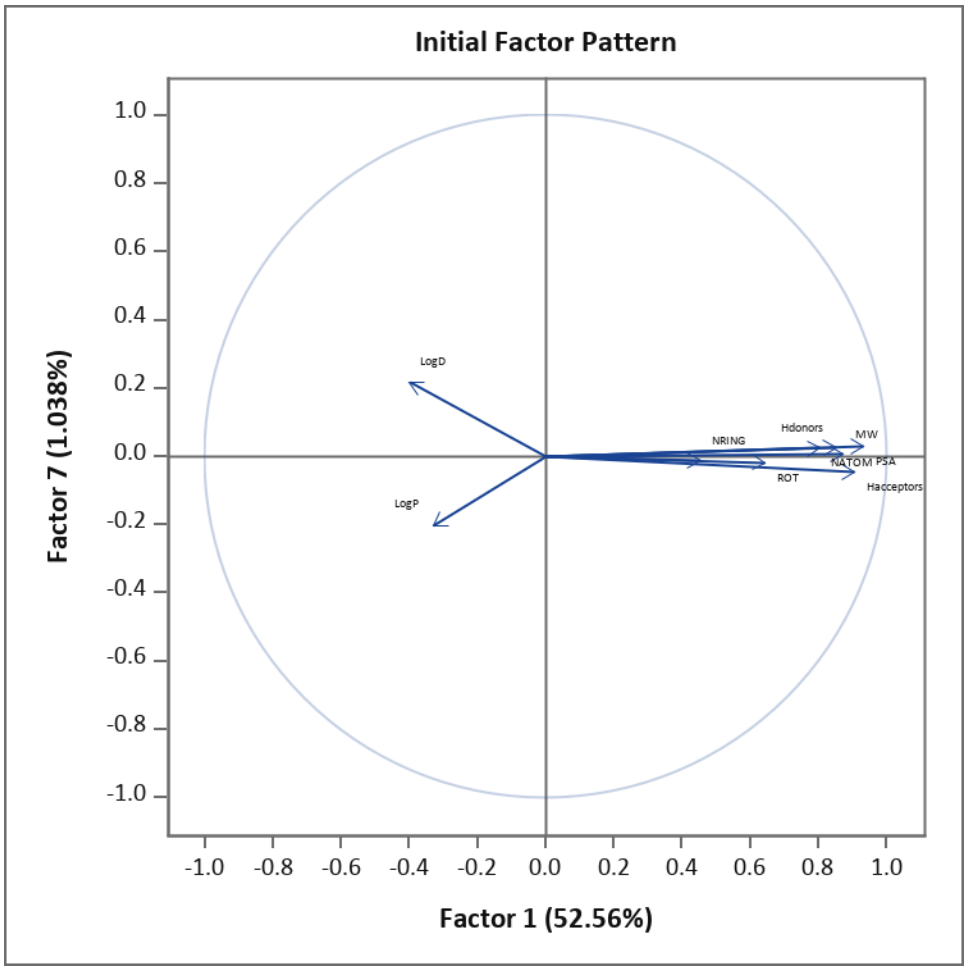

Initial Factor Pattern
Factor 6 (1.212%)
Factor 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NRING
ROT
Hdonors
LogD
LogP
Hacceptors
PSA
NATOM
MW

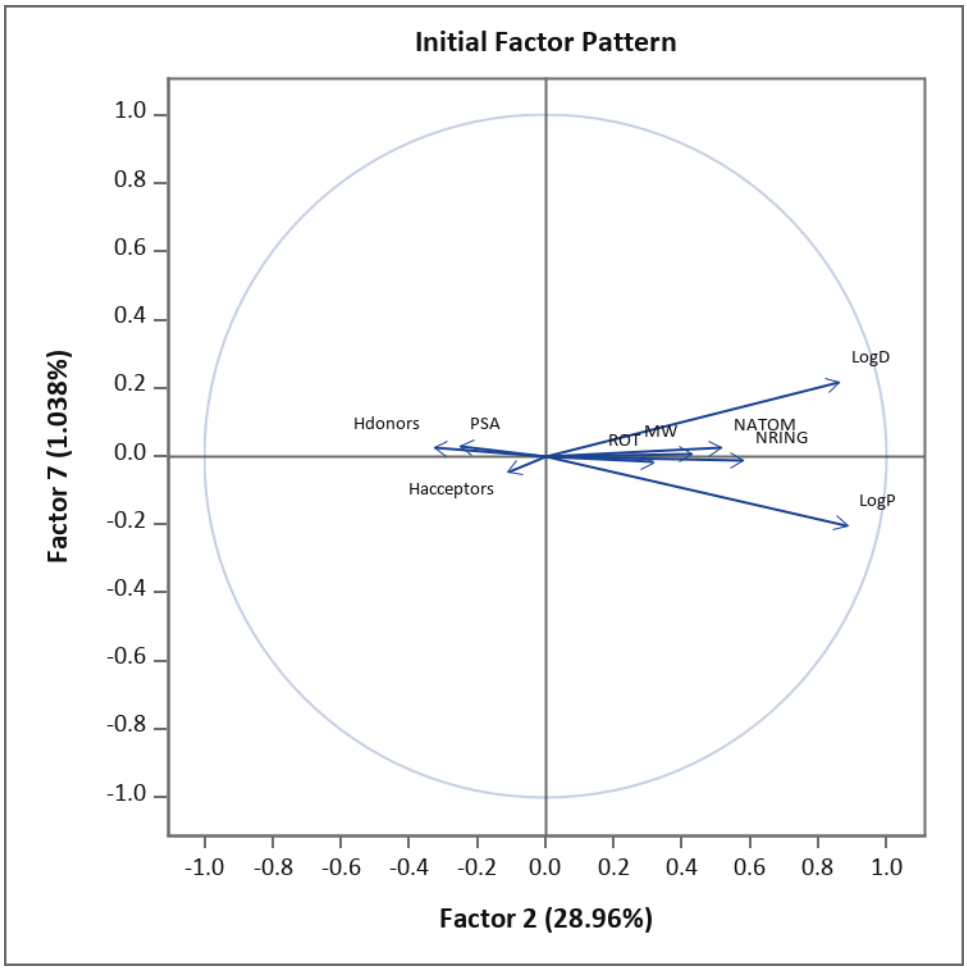

Initial Factor Pattern
Factor 6 (1.212%)
Factor 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
ROT
NRING
LogD
Hacceptors
PSA
LogP
MW
NATOM

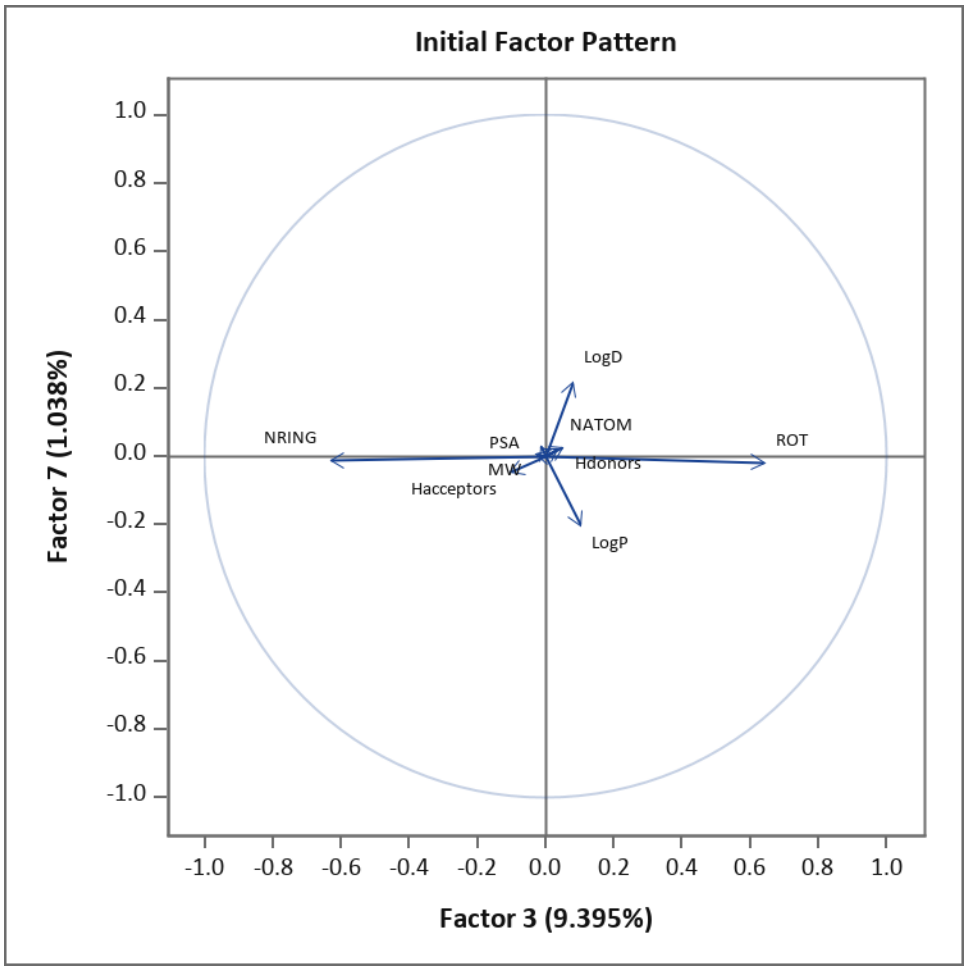

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 6 (1.212%)
NRING
Hdonors
ROT
LogD
PSA
LogP
Hacceptors
MW
NATOM
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 3 (9.395%)

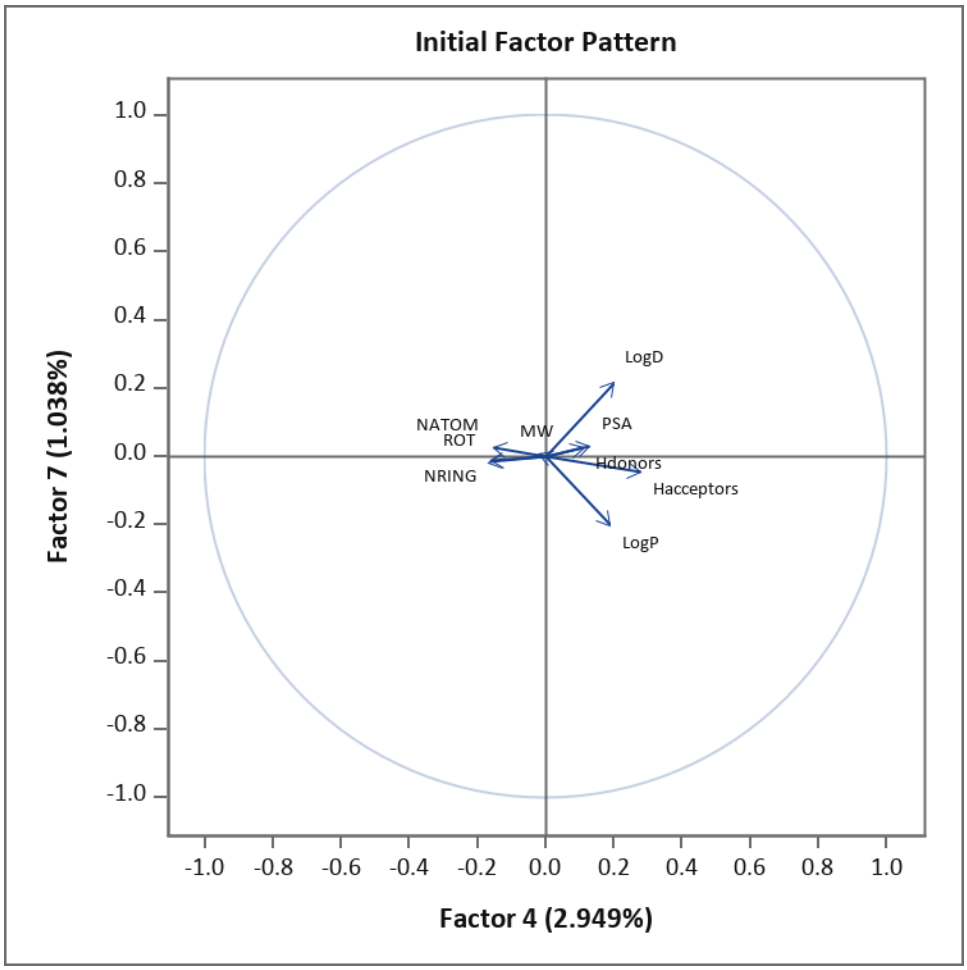

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 6 (1.212%)
NRING
Hdonors
ROT
LogP
LogD
PSA
Hacceptors
NATOM
MW
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 4 (2.949%)

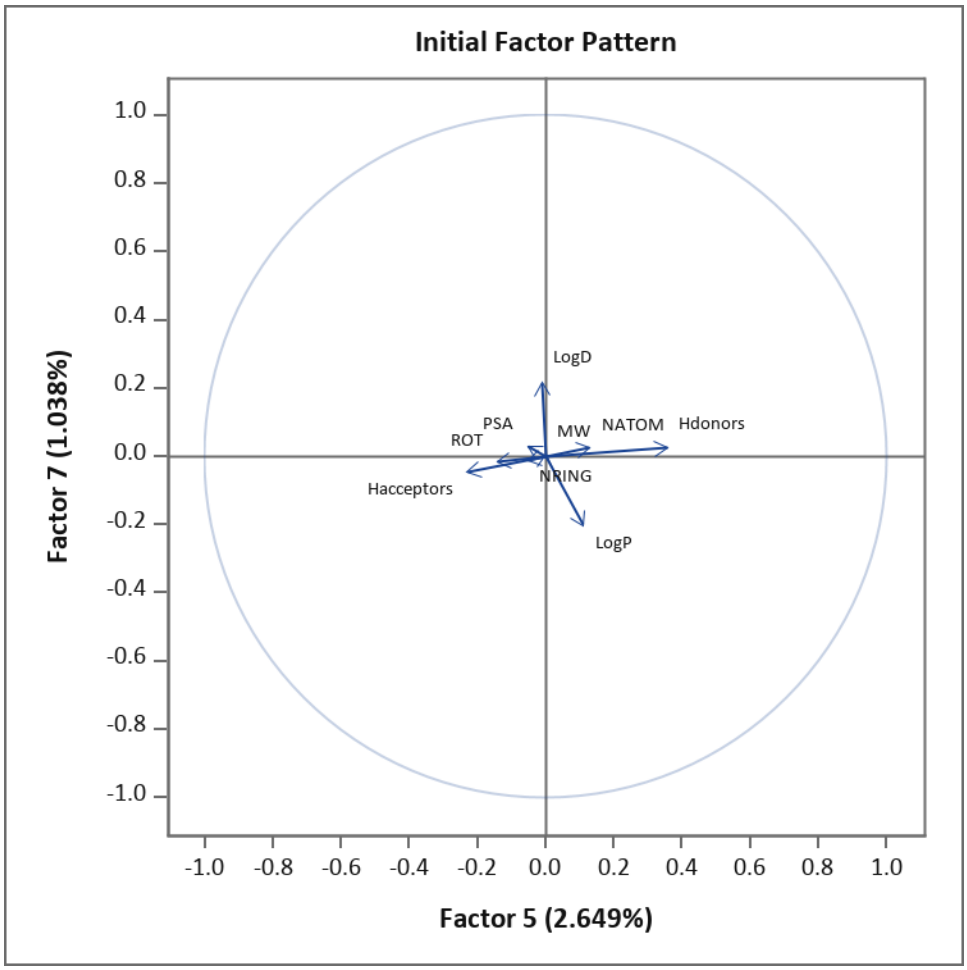

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 6 (1.212%)
NRING
Hdonors
ROTPSA
LogD
Hacceptors
LogP
NATOM
MW
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 5 (2.649%)

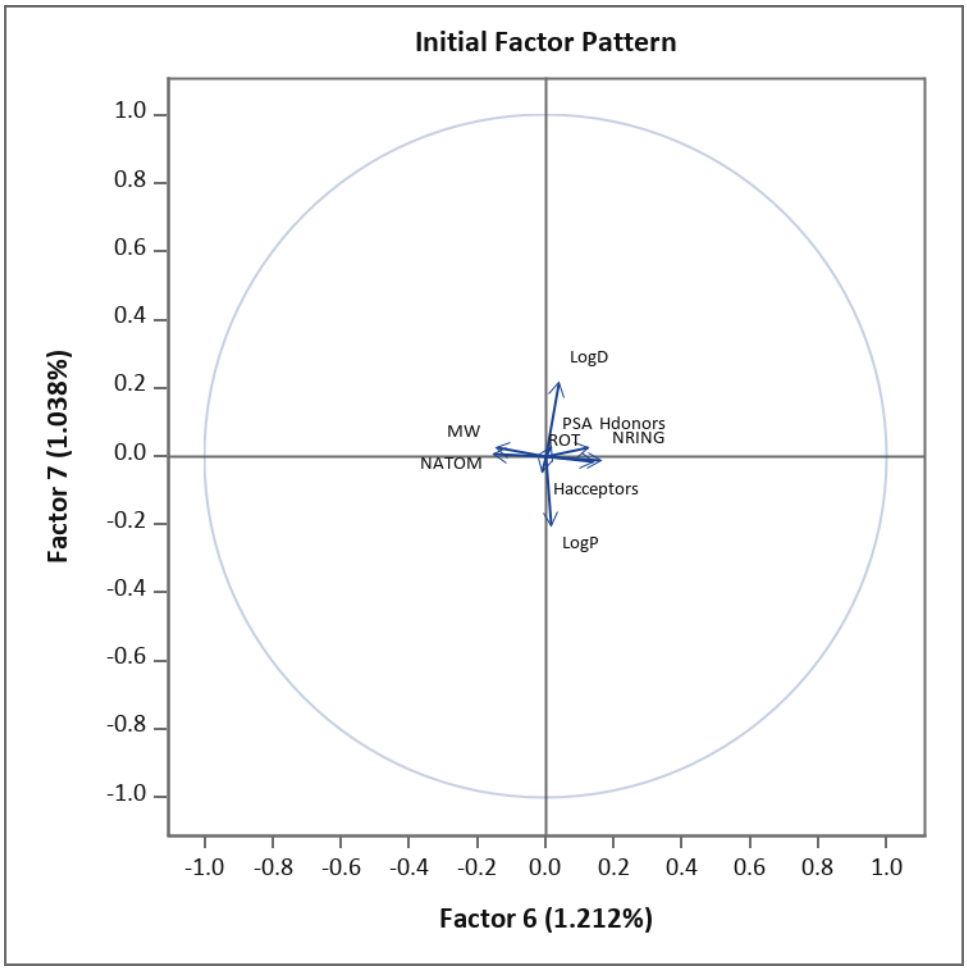

Initial Factor Pattern
Factor 7 (1.038%)
Factor 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
LogP
NRING
Hdonors
MW
NATOM
PSA
ROT
Hacceptors

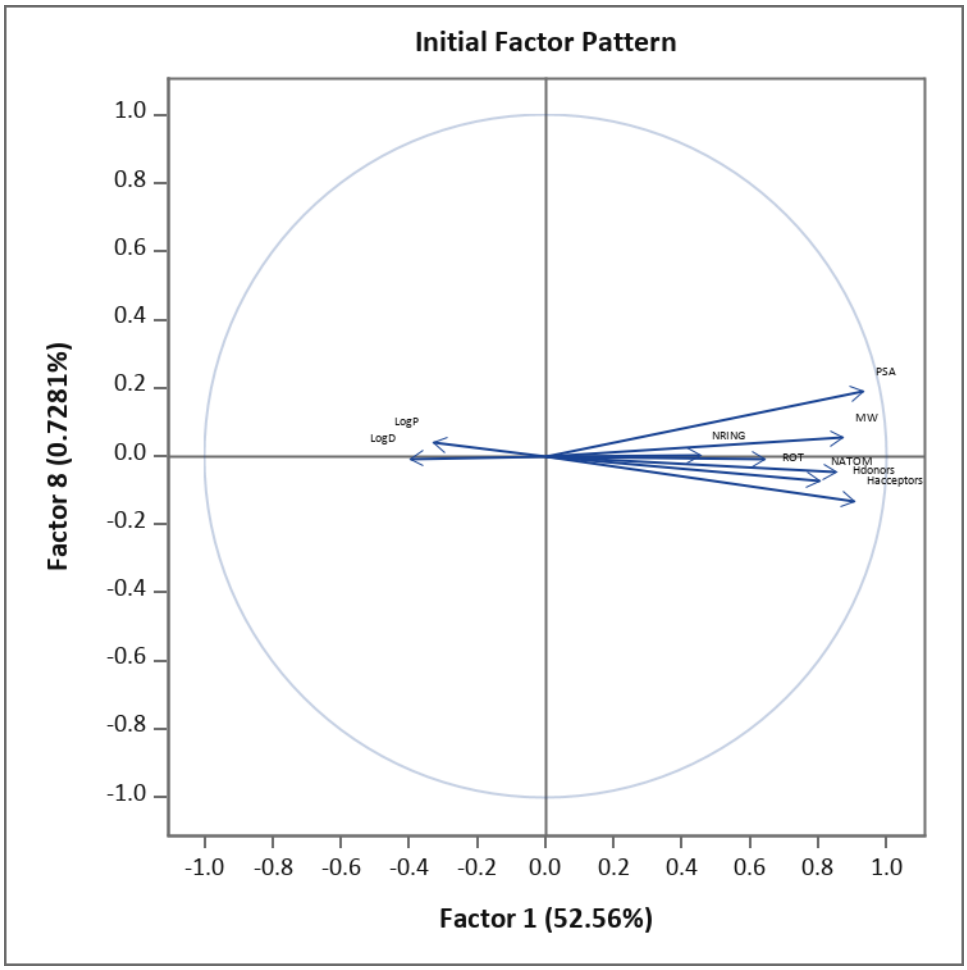

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 7 (1.038%)
LogD
Hdonors
PSA
MW
ROT
NATOM
NRING
Hacceptors
LogP
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 2 (28.96%)

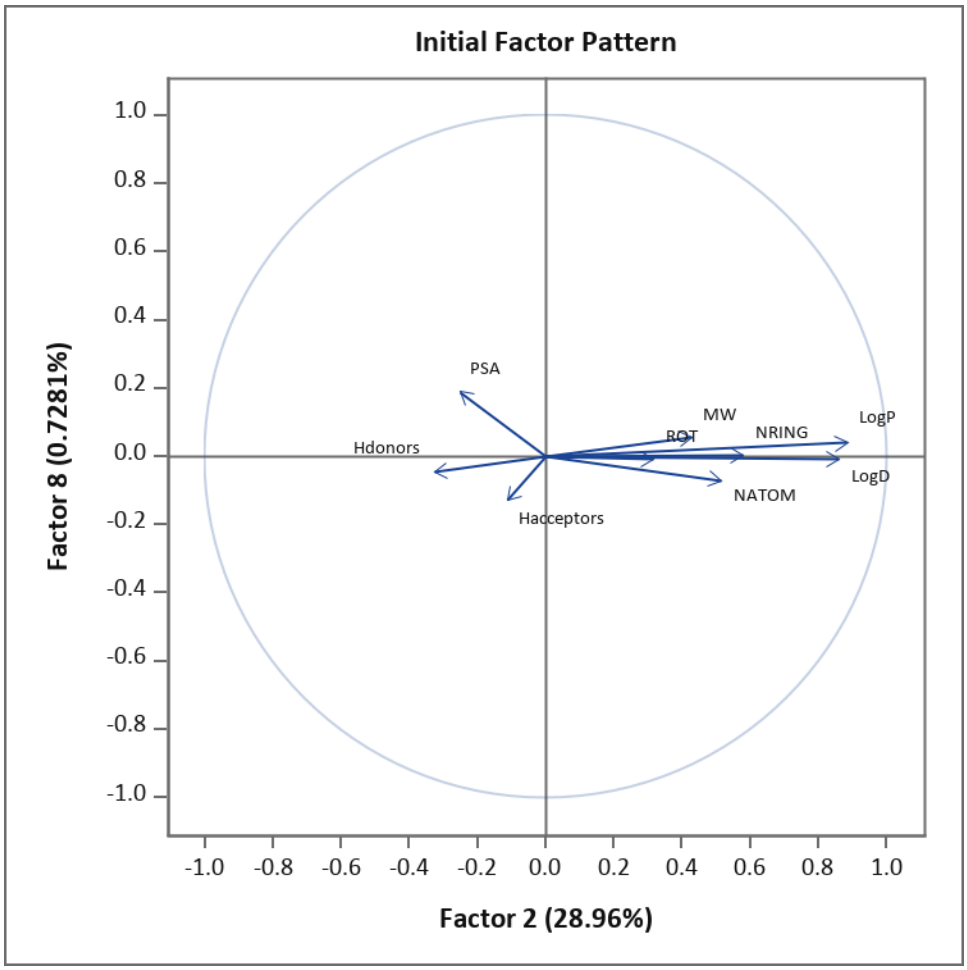

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 7 (1.038%)
LogD
NRING
NATOM
PSA
ROT
MW
Hdonors
Hacceptors
LogP
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 3 (9.395%)

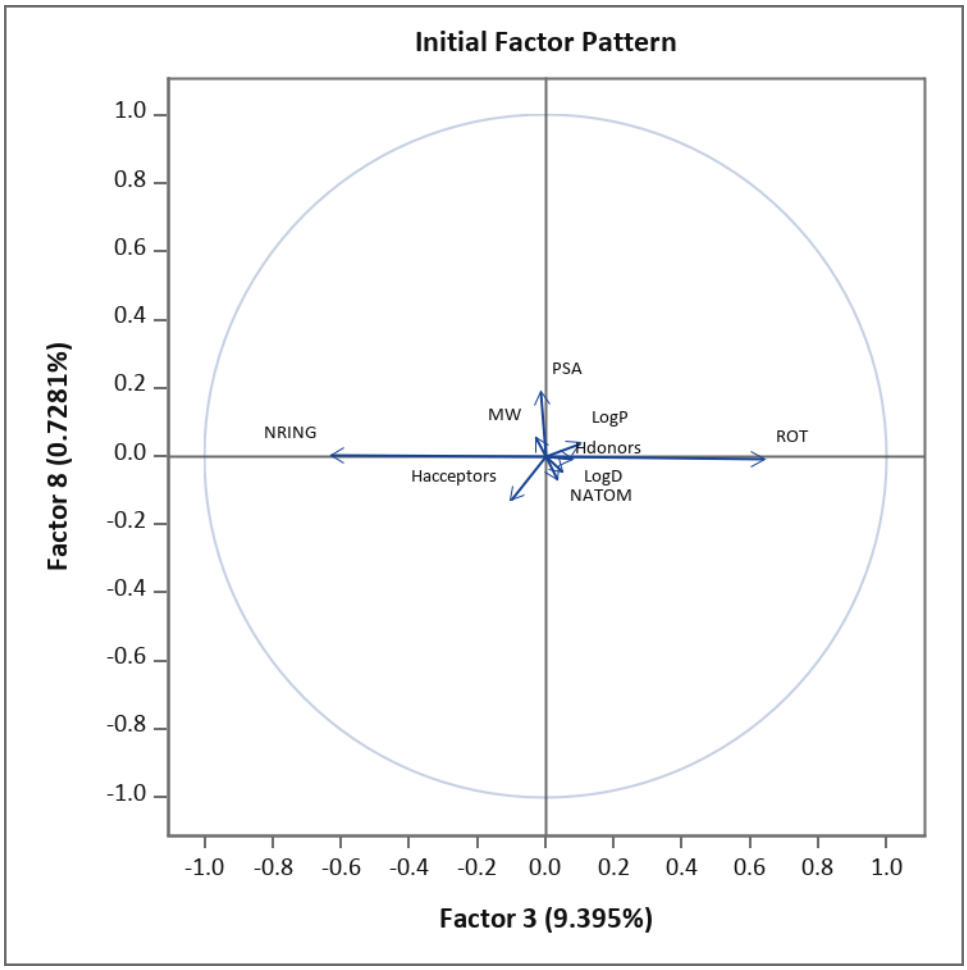

Initial Factor Pattern
Factor 7 (1.038%)
Factor 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
NATOM
ROT
MW
PSA
Hdonors
NRING
Hacceptors
LogP

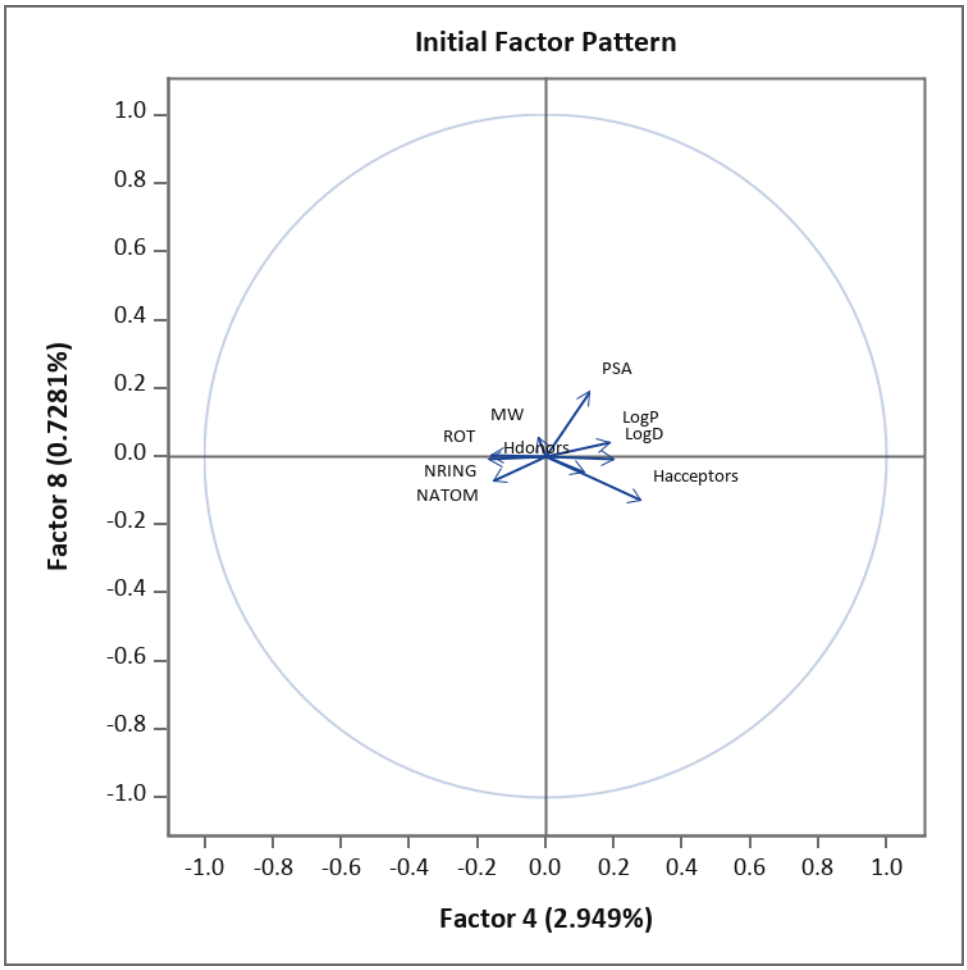

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 7 (1.038%)
LogD
PSA
ROT
MW
NATOM
Hdonors
Hacceptors
NRING
LogP
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 5 (2.649%)

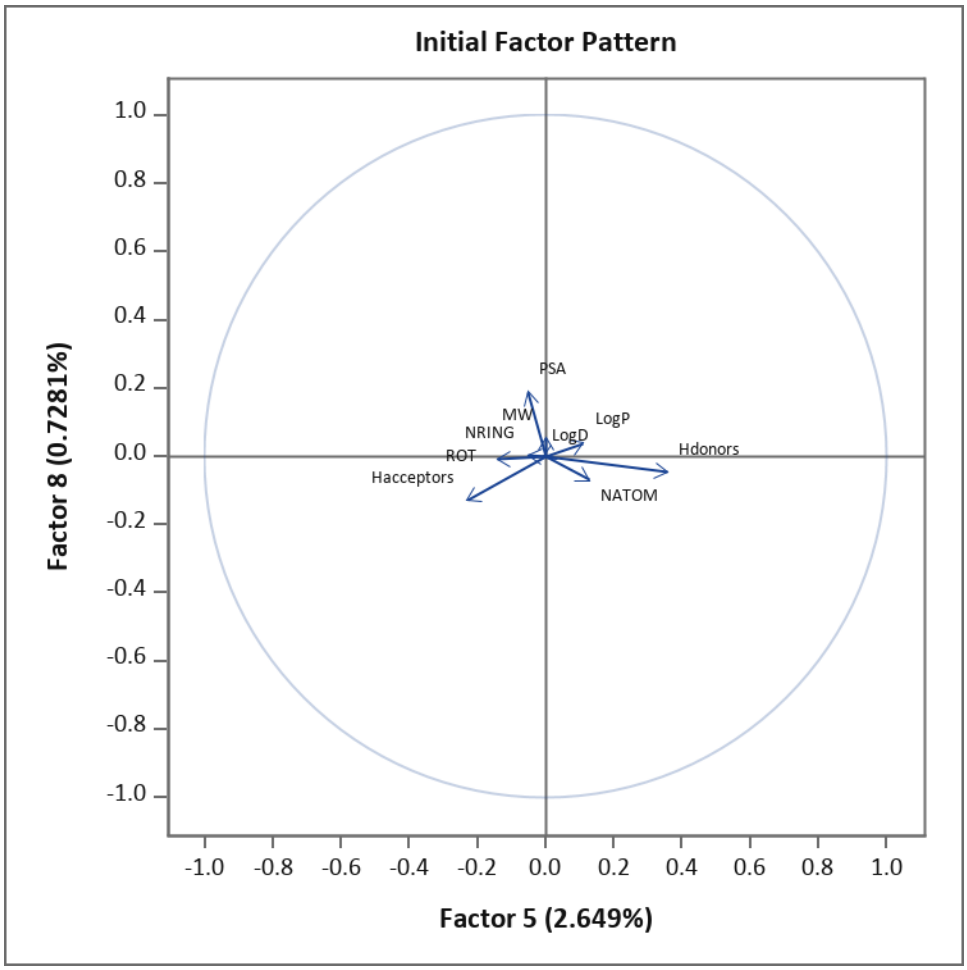

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 7 (1.038%)
LogD
MW
PSA
Hdonors
ROT
NRING
NATOM
Hacceptors
LogP
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 6 (1.212%)

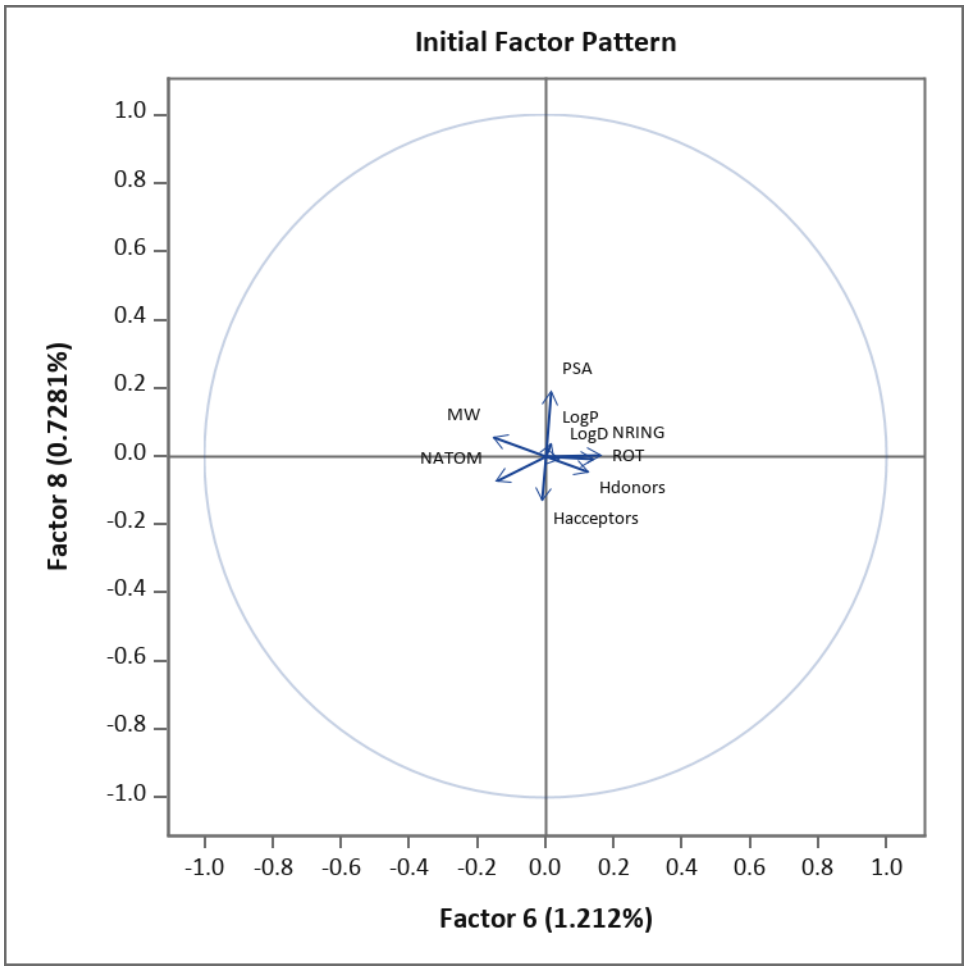

Initial Factor Pattern
Factor 8 (0.7281%)
Factor 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogP
LogD
PSA
MW
NRING
ROT
NATOM
Hdonors
Hacceptors

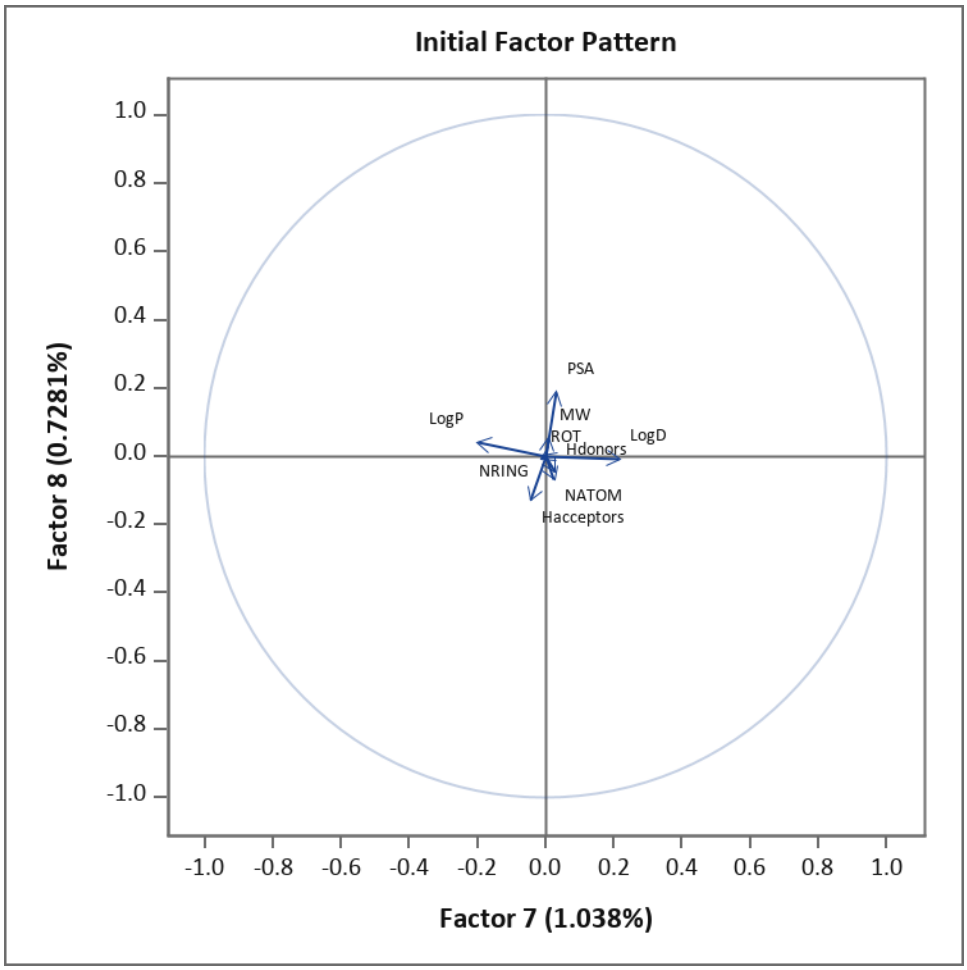

Initial Factor Pattern
Factor 8 (0.7281%)
Factor 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
PSA
MW
NRING
LogP
ROT
Hdonors
LogD
NATOM
Hacceptors

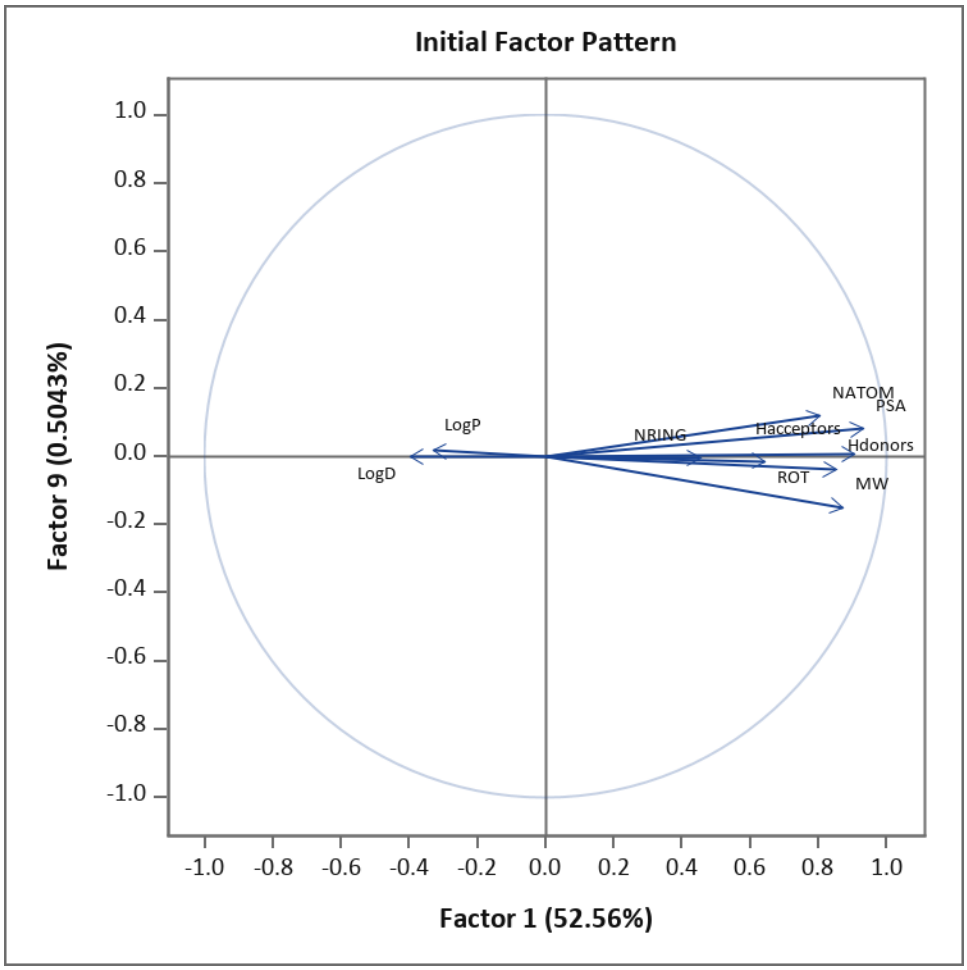

Initial Factor Pattern
Factor 8 (0.7281%)
Factor 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
MW
LogP
NRING
Hdonors
ROT
Hacceptors
LogD
NATOM

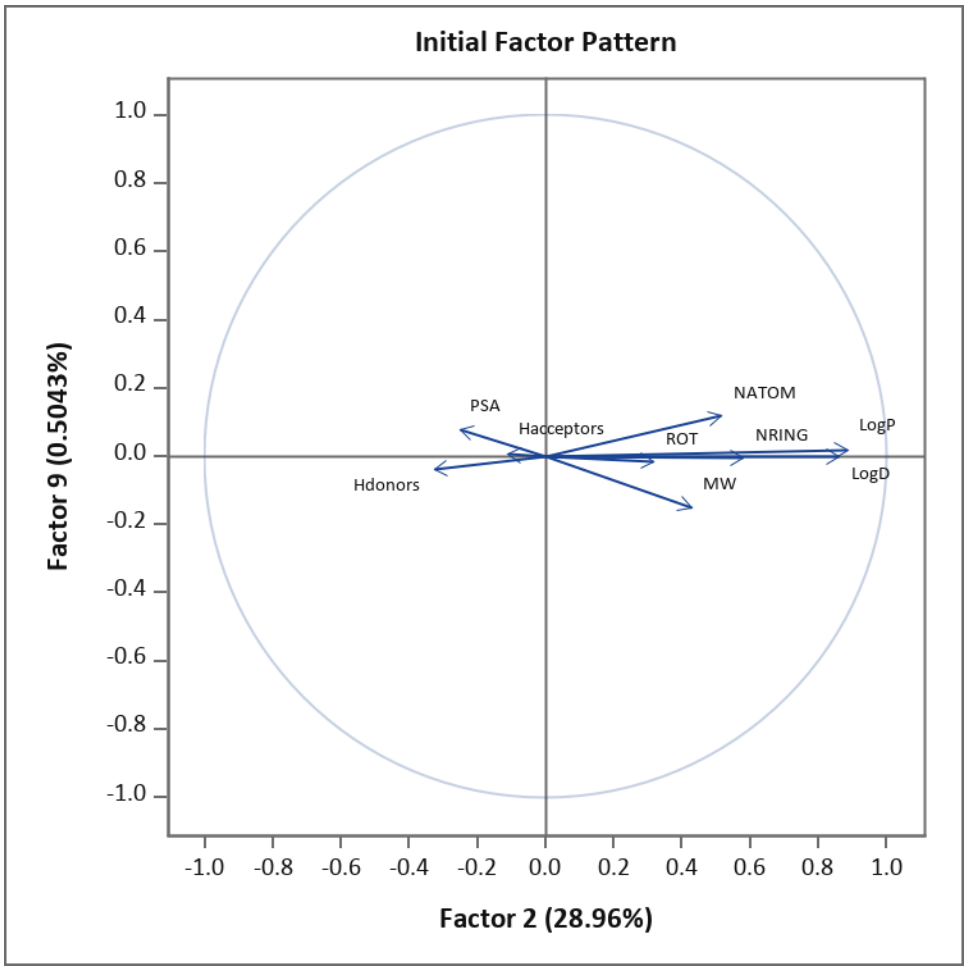

Initial Factor Pattern
Factor 8 (0.7281%)
Factor 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
PSA
MW
LogP
LogD
ROT
Hdonors
NRING
NATOM
Hacceptors

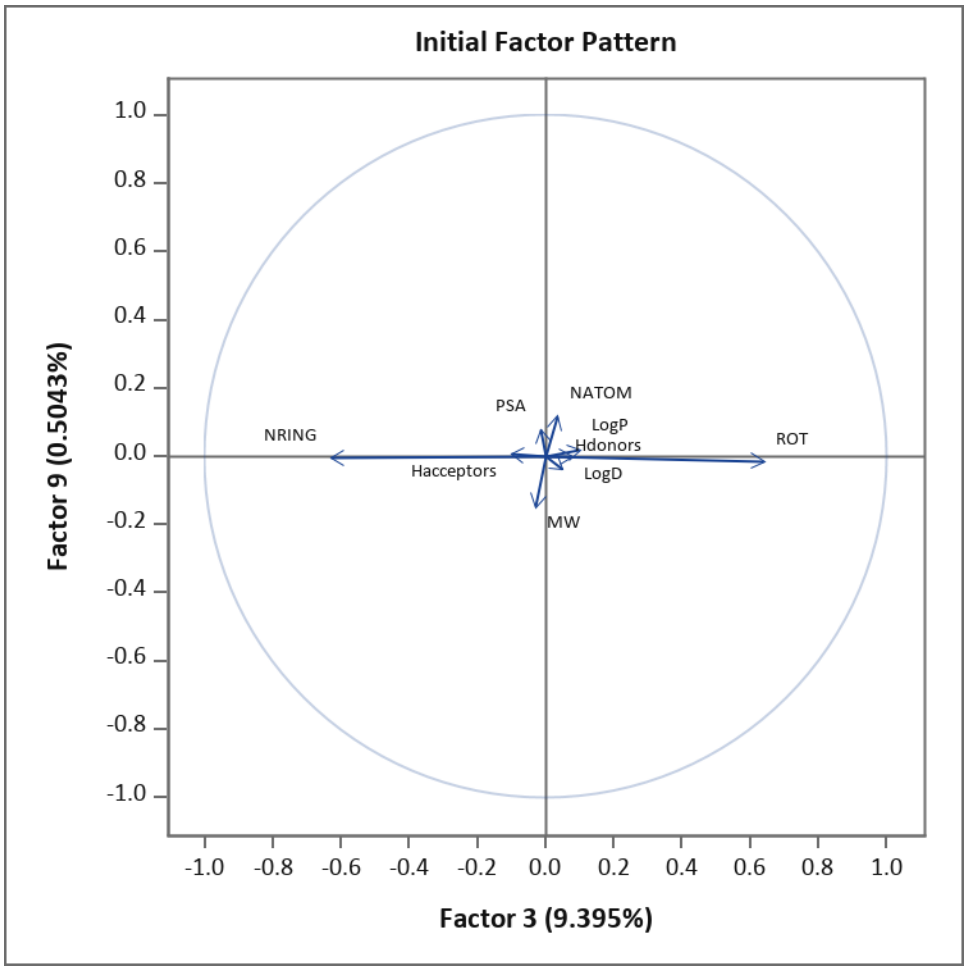

Initial Factor Pattern
Factor 8 (0.7281%)
Factor 5 (2.649%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
PSA
MW
LogP
NRING
LogD
ROT
Hdonors
Hacceptors
NATOM

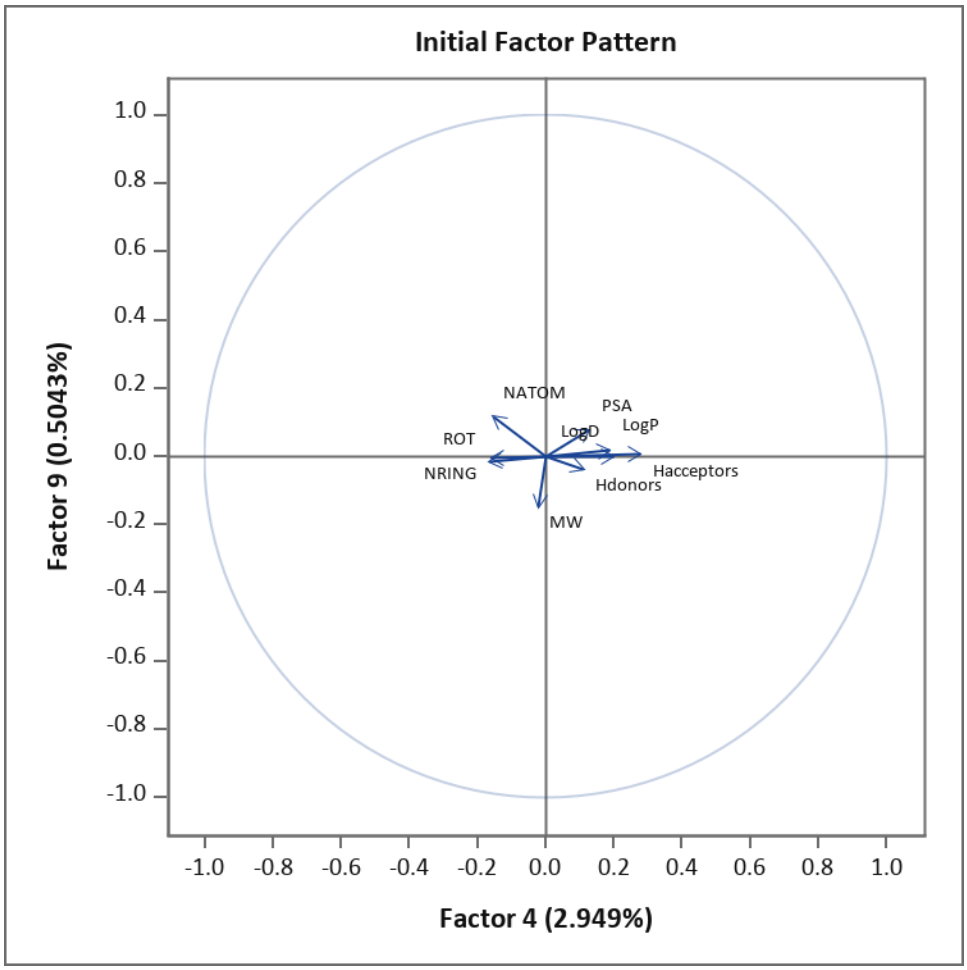

Initial Factor Pattern
Factor 8 (0.7281%)
Factor 6 (1.212%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
PSA
MW
LogP
LogD NRING
NATOM
ROT
Hdonors
Hacceptors

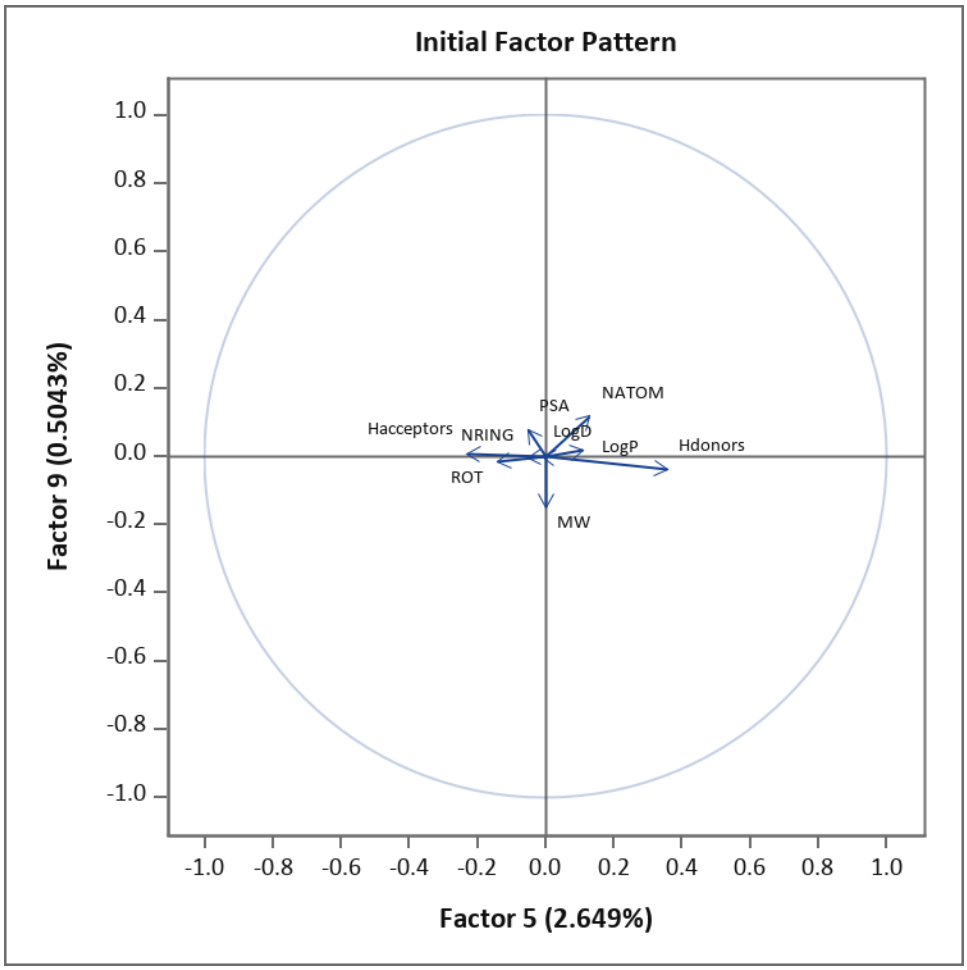

Initial Factor Pattern
Factor 8 (0.7281%)
Factor 7 (1.038%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
PSA
MW
LogP
ROT
LogD
Hdonors
NRING
NATOM
Hacceptors

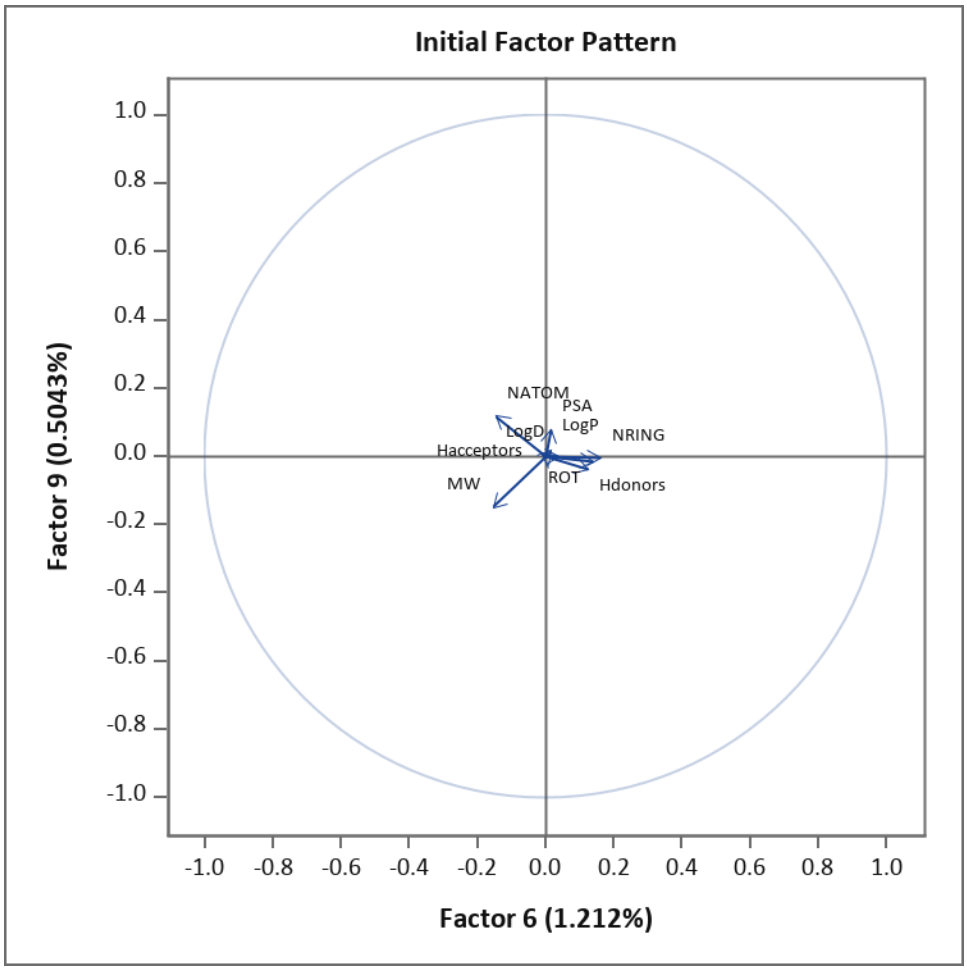

Initial Factor Pattern
Factor 9 (0.5043%)
Factor 1 (52.56%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
LogP
LogD
NRING
NATOM
PSA
Hacceptors
Hdonors
ROT
MW

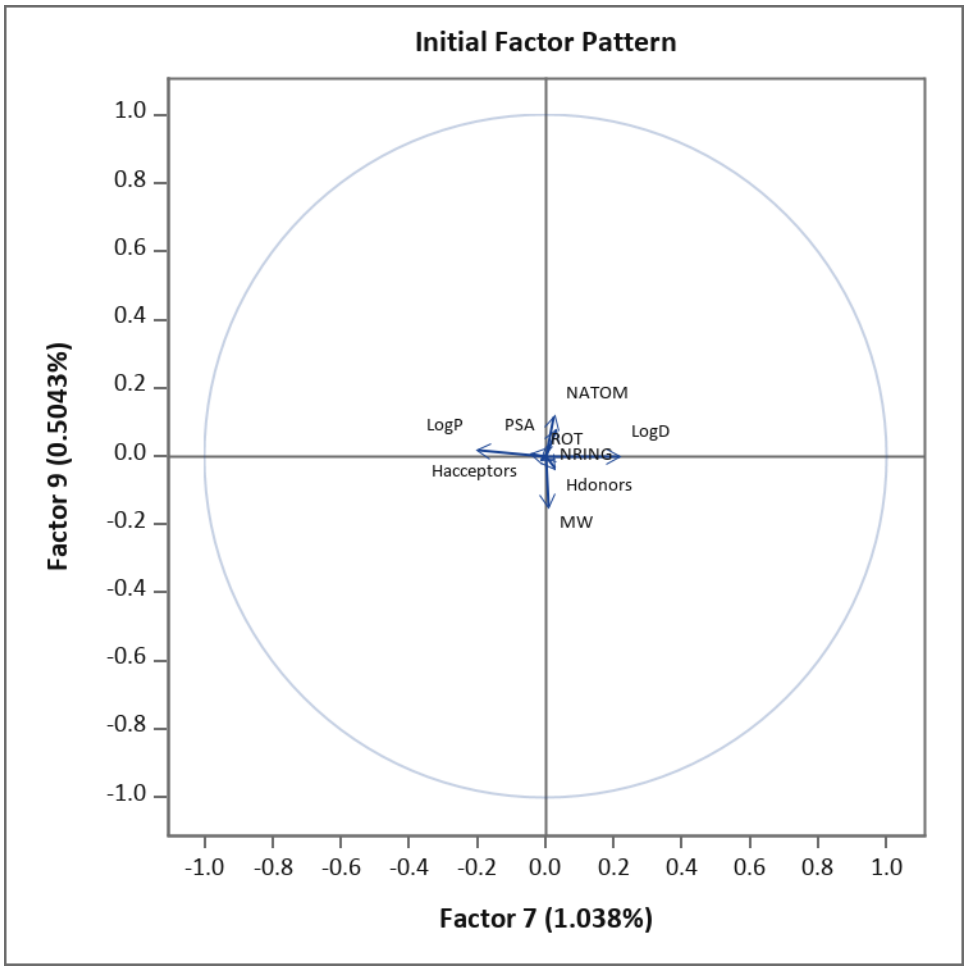

Initial Factor Pattern
Factor 9 (0.5043%)
Factor 2 (28.96%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
PSA
Hacceptors
NATOM
ROT
NRING
LogP
Hdonors
MW
LogD

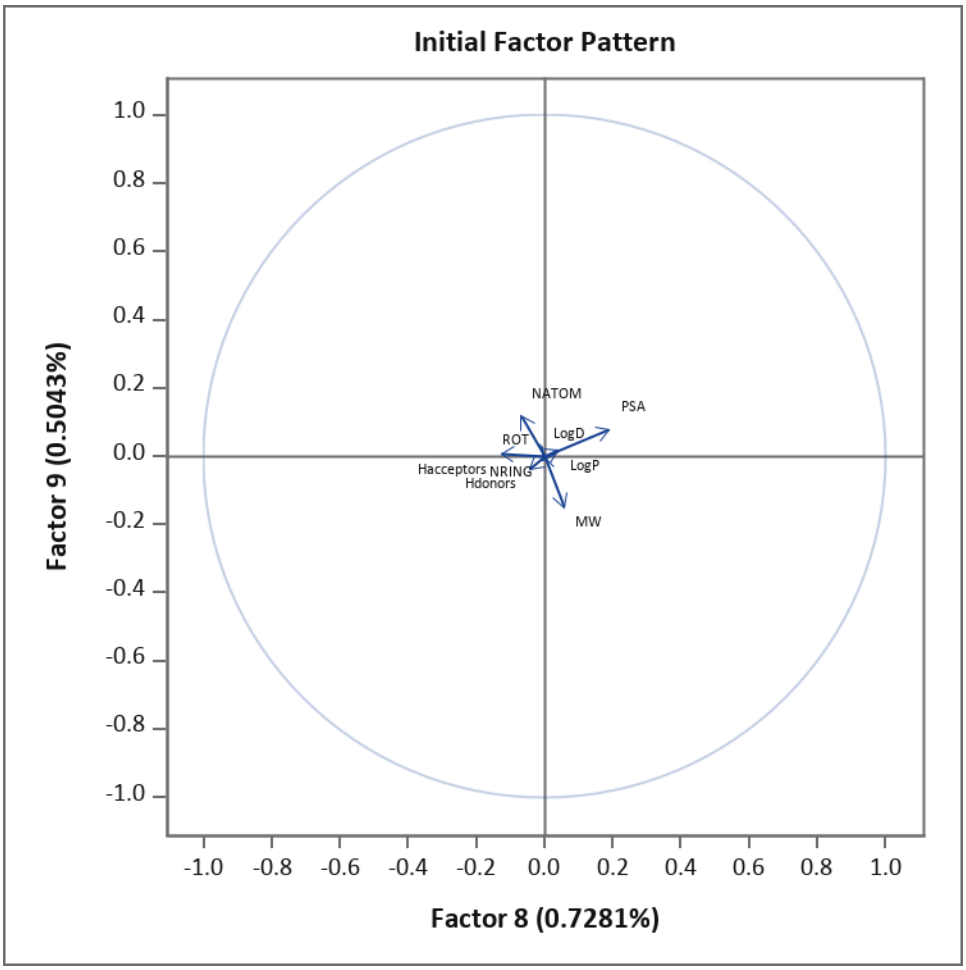

Initial Factor Pattern
Factor 9 (0.5043%)
Factor 3 (9.395%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NRING
PSA
NATOM
LogP
Hdonors
ROT
Hacceptors
LogD
MW


Initial Factor Pattern
Factor 9 (0.5043%)
Factor 4 (2.949%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
NATOM
PSA
LogP
LogD
ROT
NRING
Hacceptors
Hdonors
MW


Initial Factor Pattern
Factor 9 (0.5043%)
Factor 5 (2.649%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
NATOM
PSA
Hacceptors
NRING
LogD
LogP
Hdonors
ROT
MW


Initial Factor Pattern
Factor 9 (0.5043%)
Factor 6 (1.212%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
NATOM
PSA
LogD
LogP
NRING
Hacceptors
MW
ROT
Hdonors

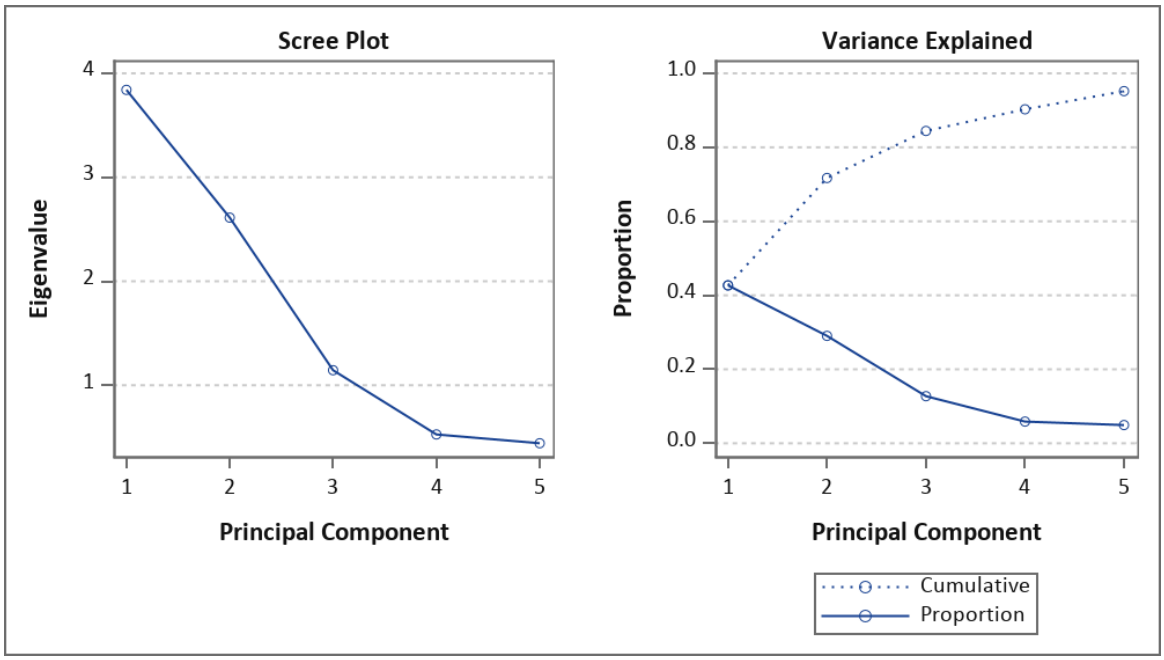

Initial Factor Pattern
Factor 9 (0.5043%)
Factor 7 (1.038%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
NATOM
LogP
PSA
ROT
LogD
NRING
Hacceptors
Hdonors
MW


Initial Factor Pattern
Factor 9 (0.5043%)
Factor 8 (0.7281%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
NATOM
PSA
ROT
LogD
Hacceptors
NRING
Hdonors
LogP
MW

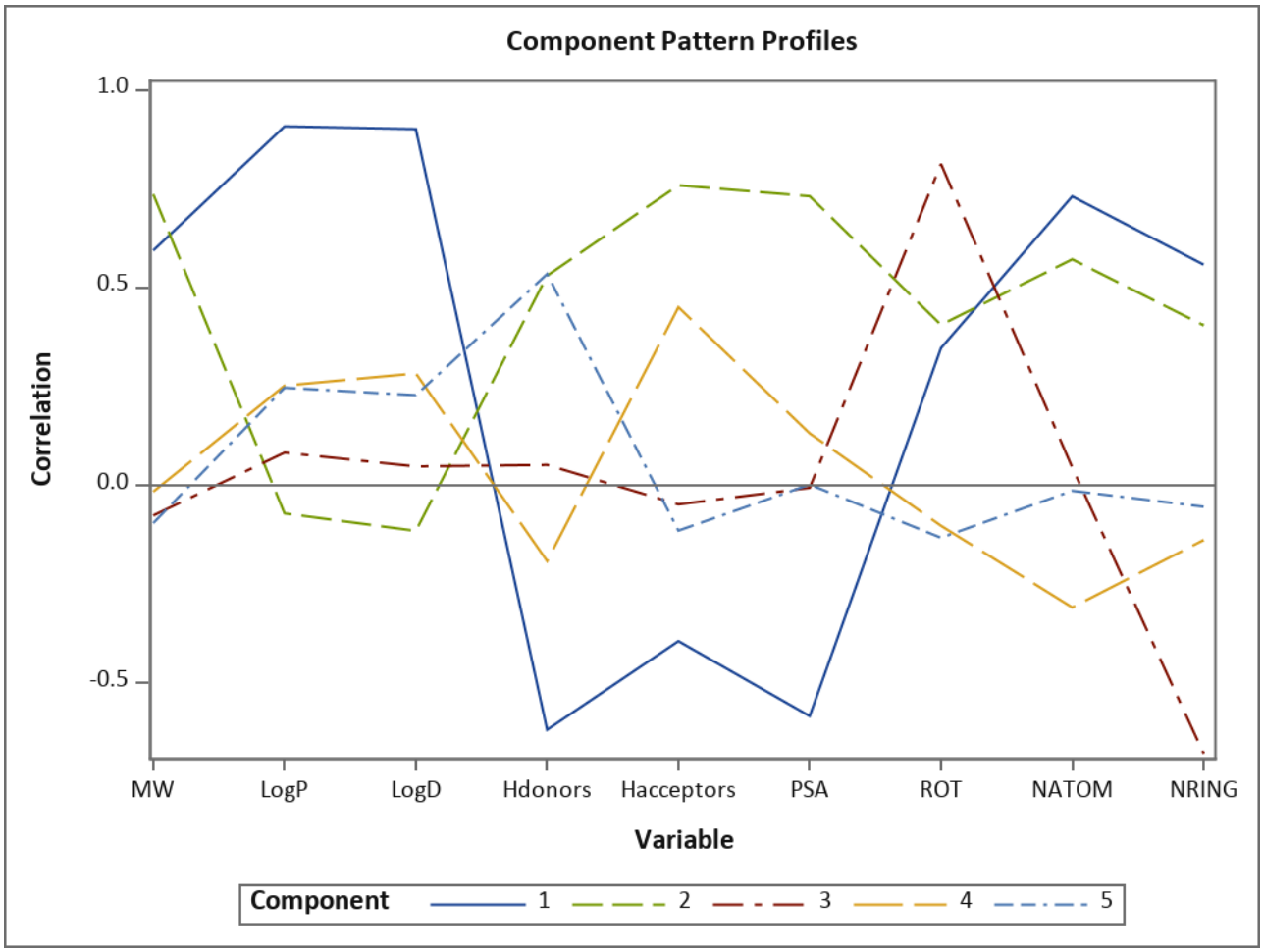
[A2]

score9_logD_group_1=1, non-violators

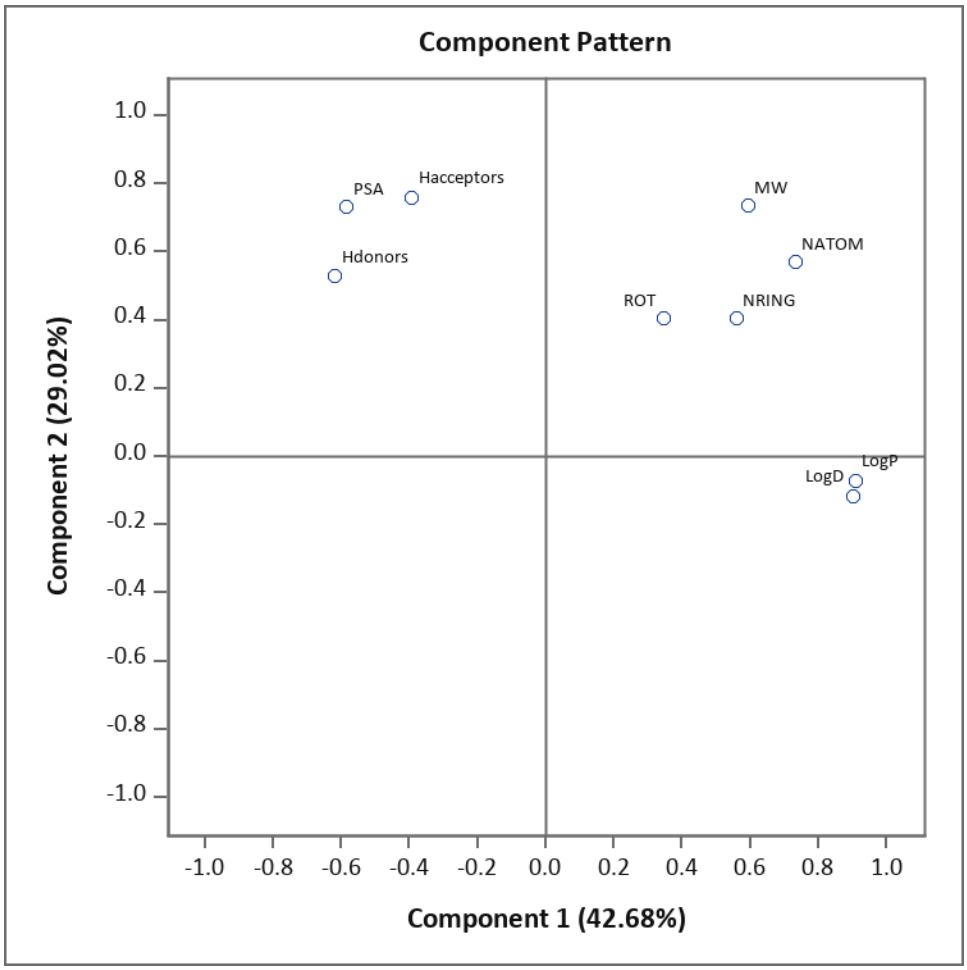
1. Scree plot for non-violators

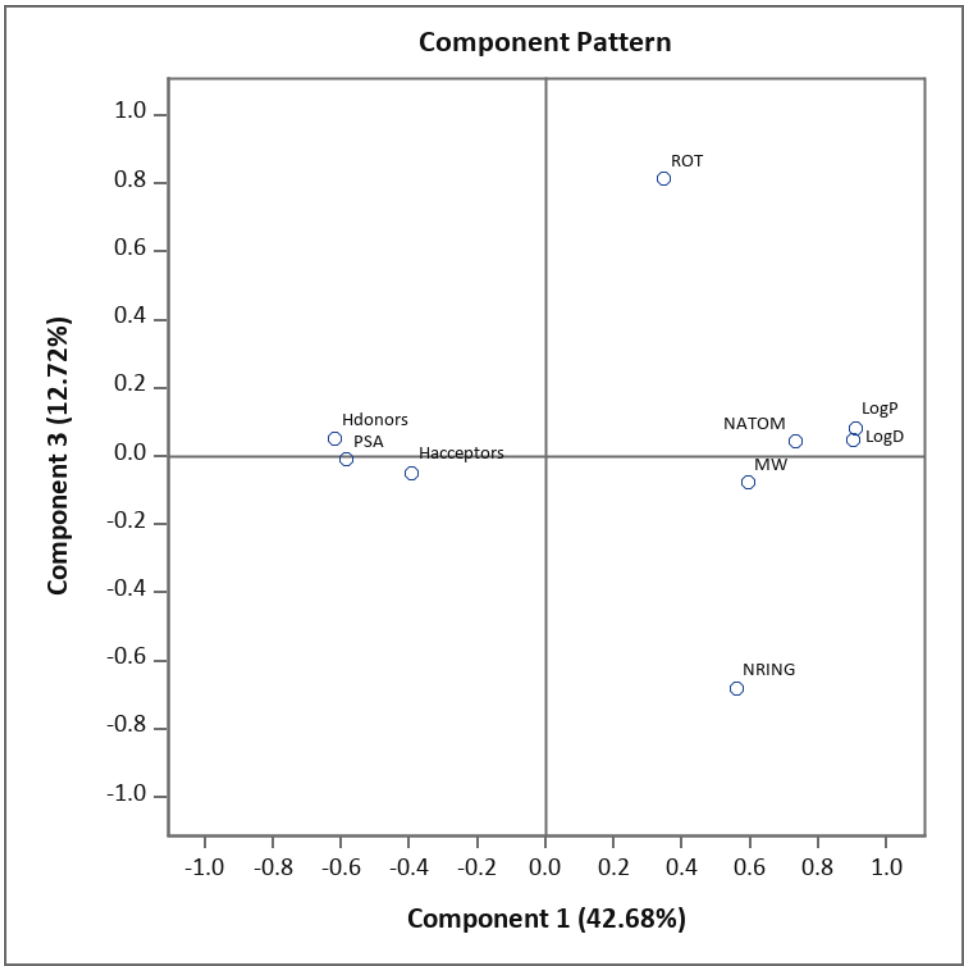


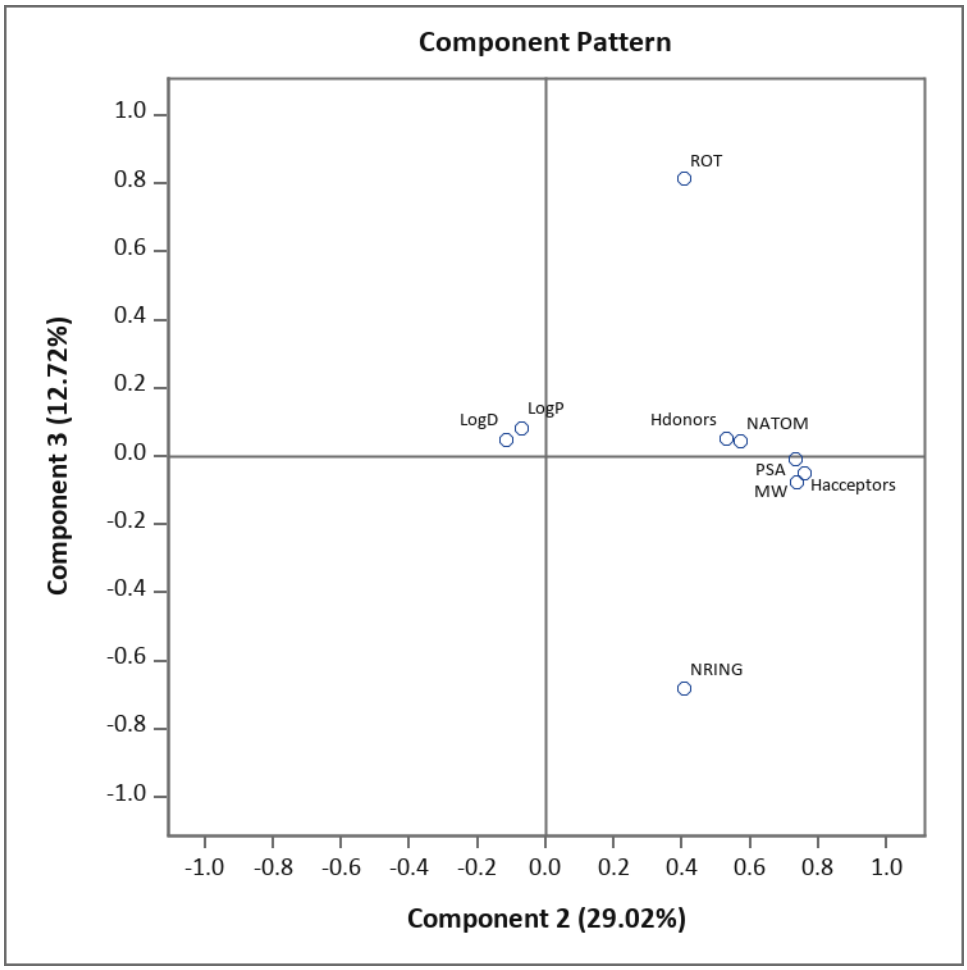
2. Profile plot for non-violators

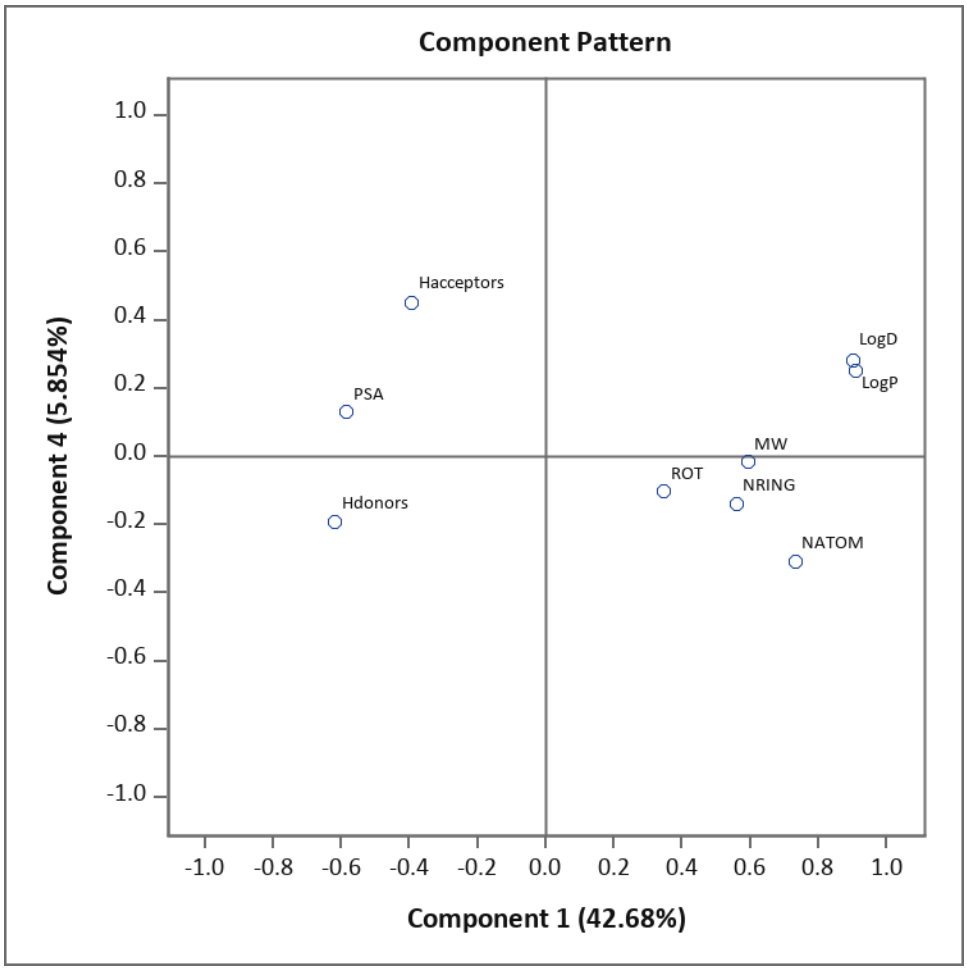


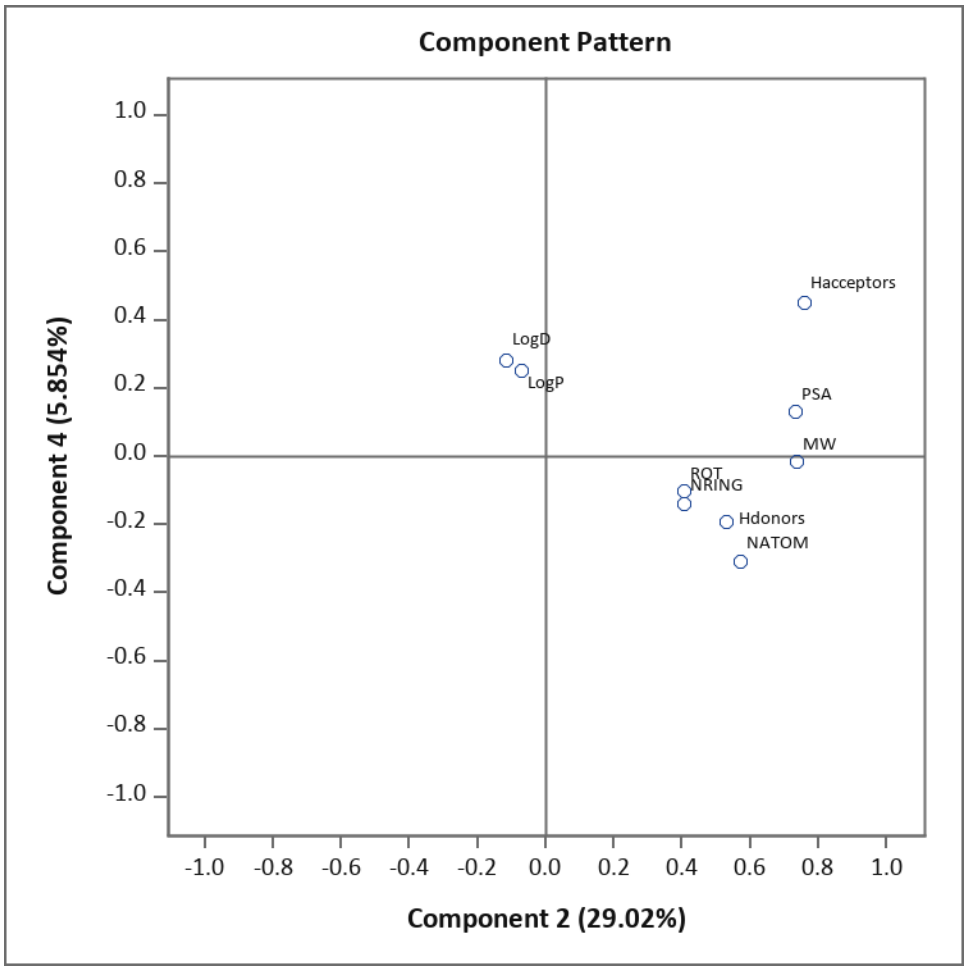
3. Component Pattern plots for non-violators

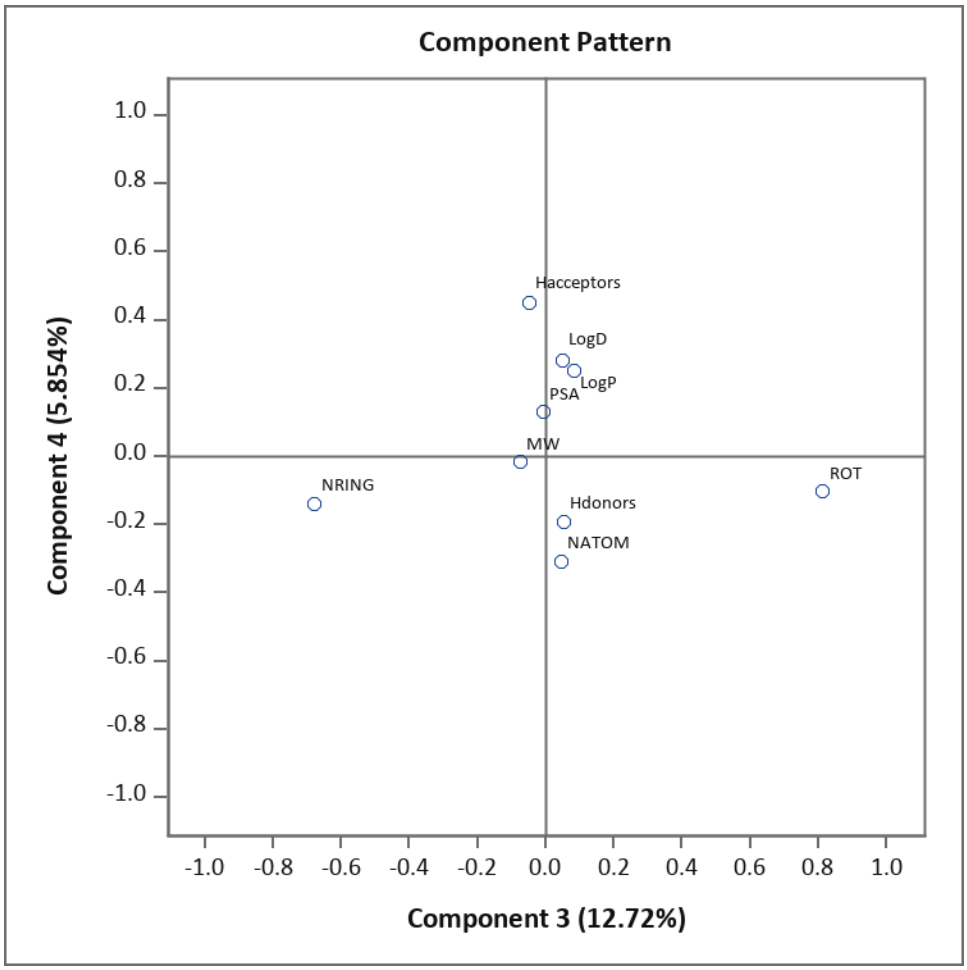


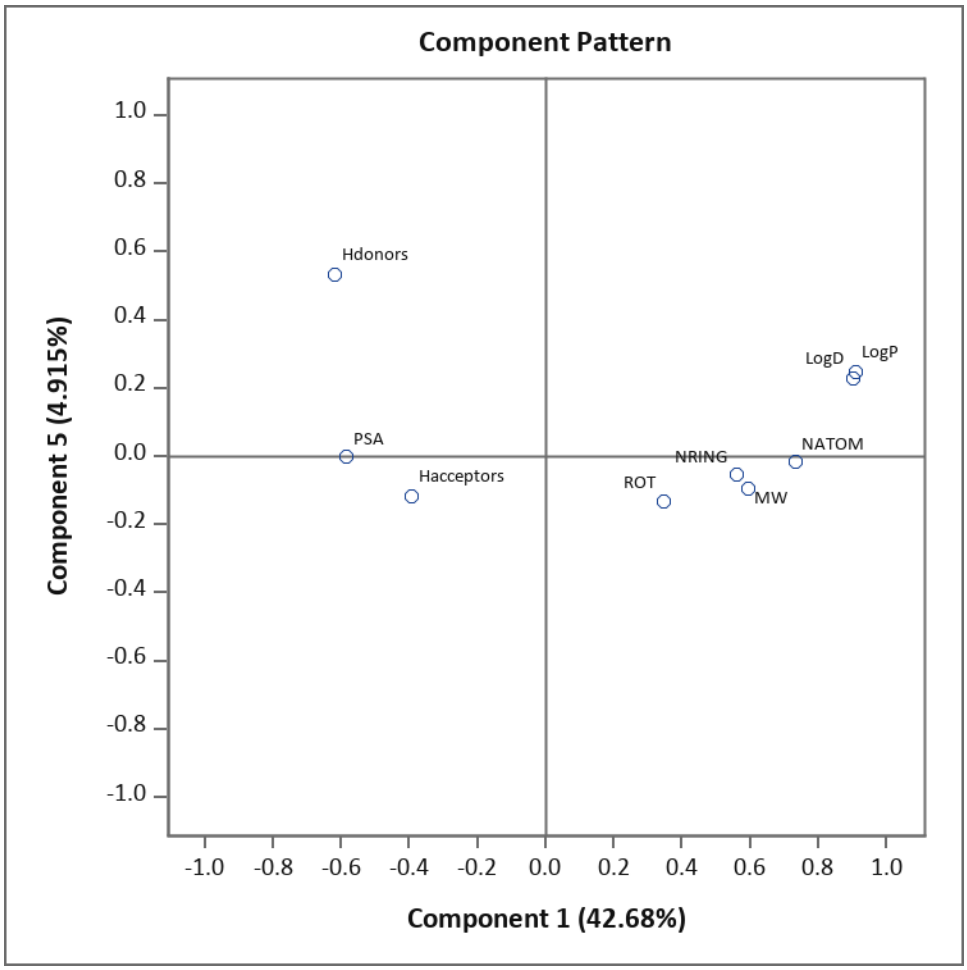

Component Pattern
Component 3 (12.72%)
Component 1 (42.68%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
ROT
Hdonors
PSA
Hacceptors
NATOM
LogP
LogD
MW
NRING

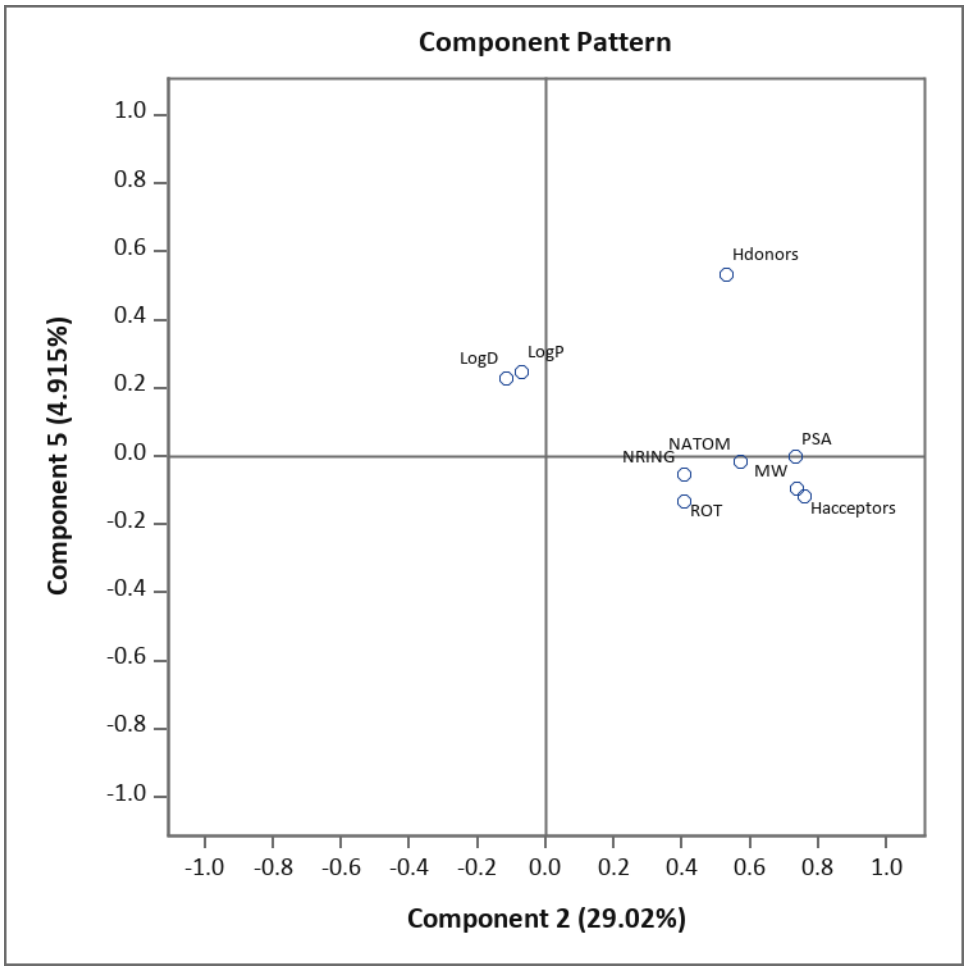

Component Pattern
Component 3 (12.72%)
Component 2 (29.02%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
ROT
LogD
LogP
Hdonors
NATOM
PSA
MW
Hacceptors
NRING

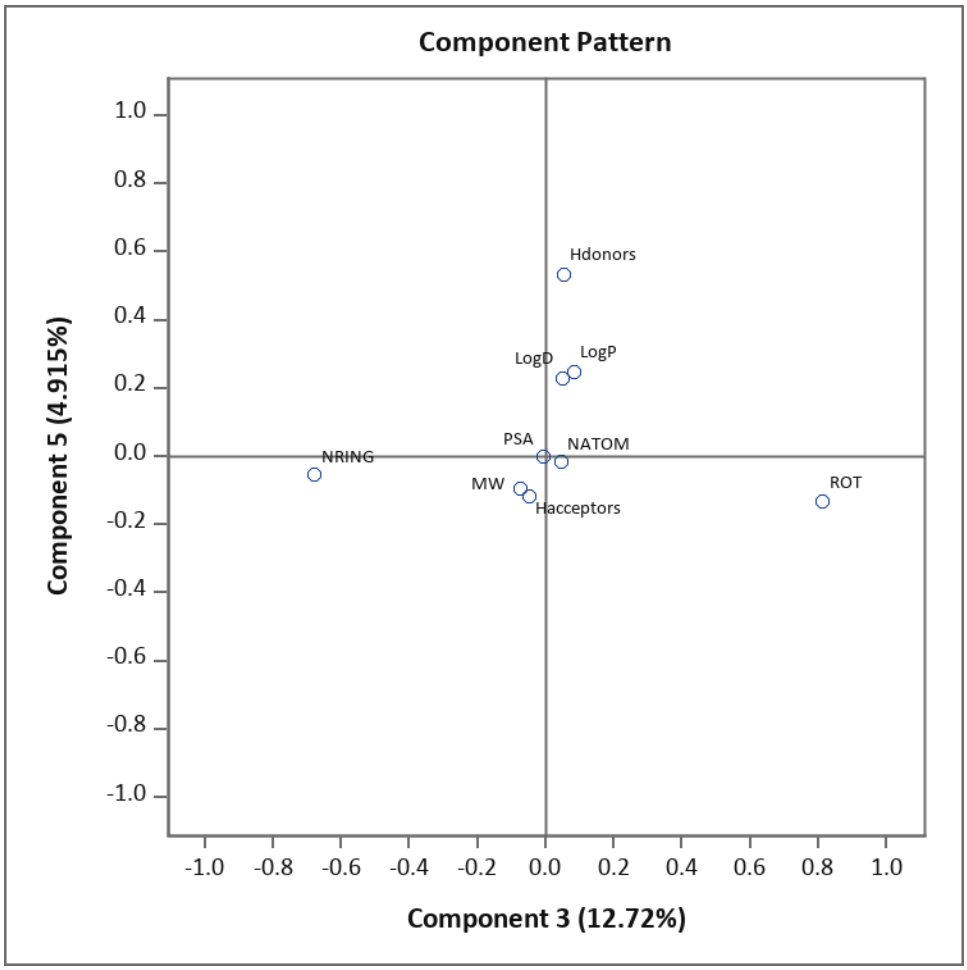

Component Pattern
Component 4 (5.854%)
Component 1 (42.68%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
PSA
Hdonors
LogD
LogP
MW
ROT
NRING
NATOM

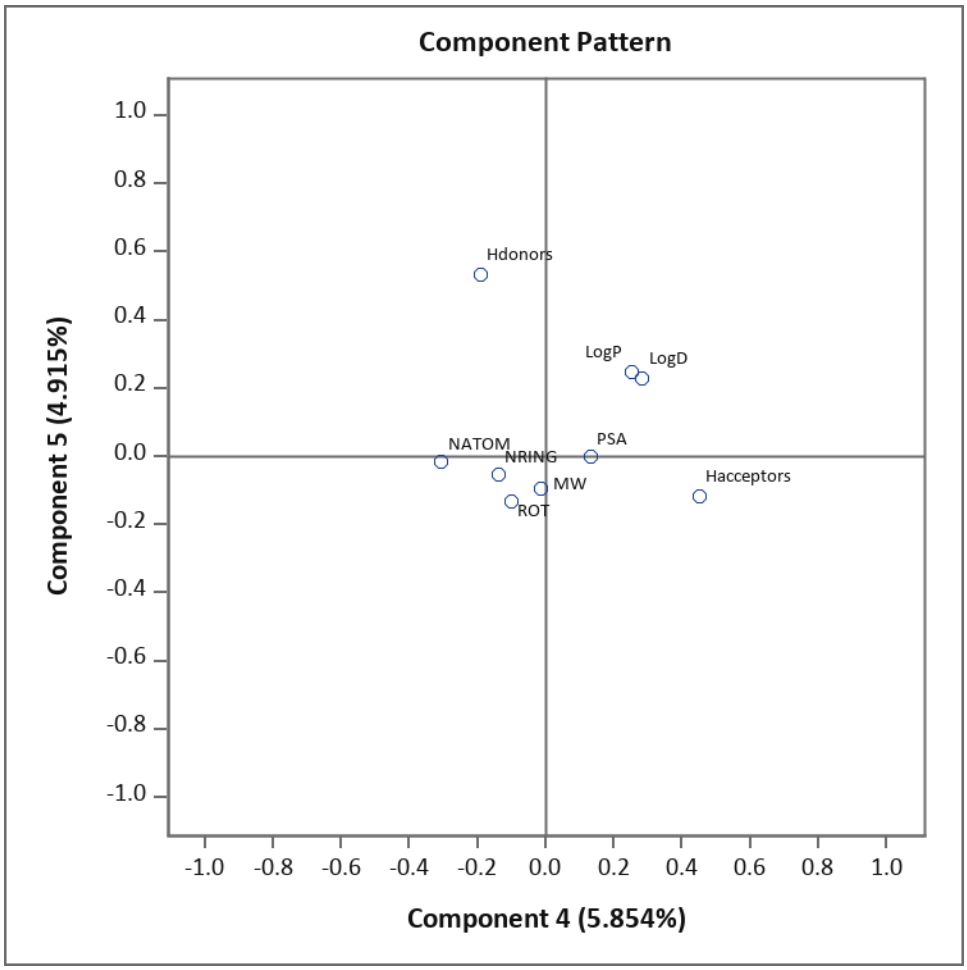

Component Pattern
Component 4 (5.854%)
Component 2 (29.02%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
LogD
LogP
PSA
MW
RQT
NRING
Hdonors
NATOM


Component Pattern
Component 4 (5.854%)
Component 3 (12.72%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
LogD
LogP
PSA
MW
NRING
ROT
Hdonors
NATOM

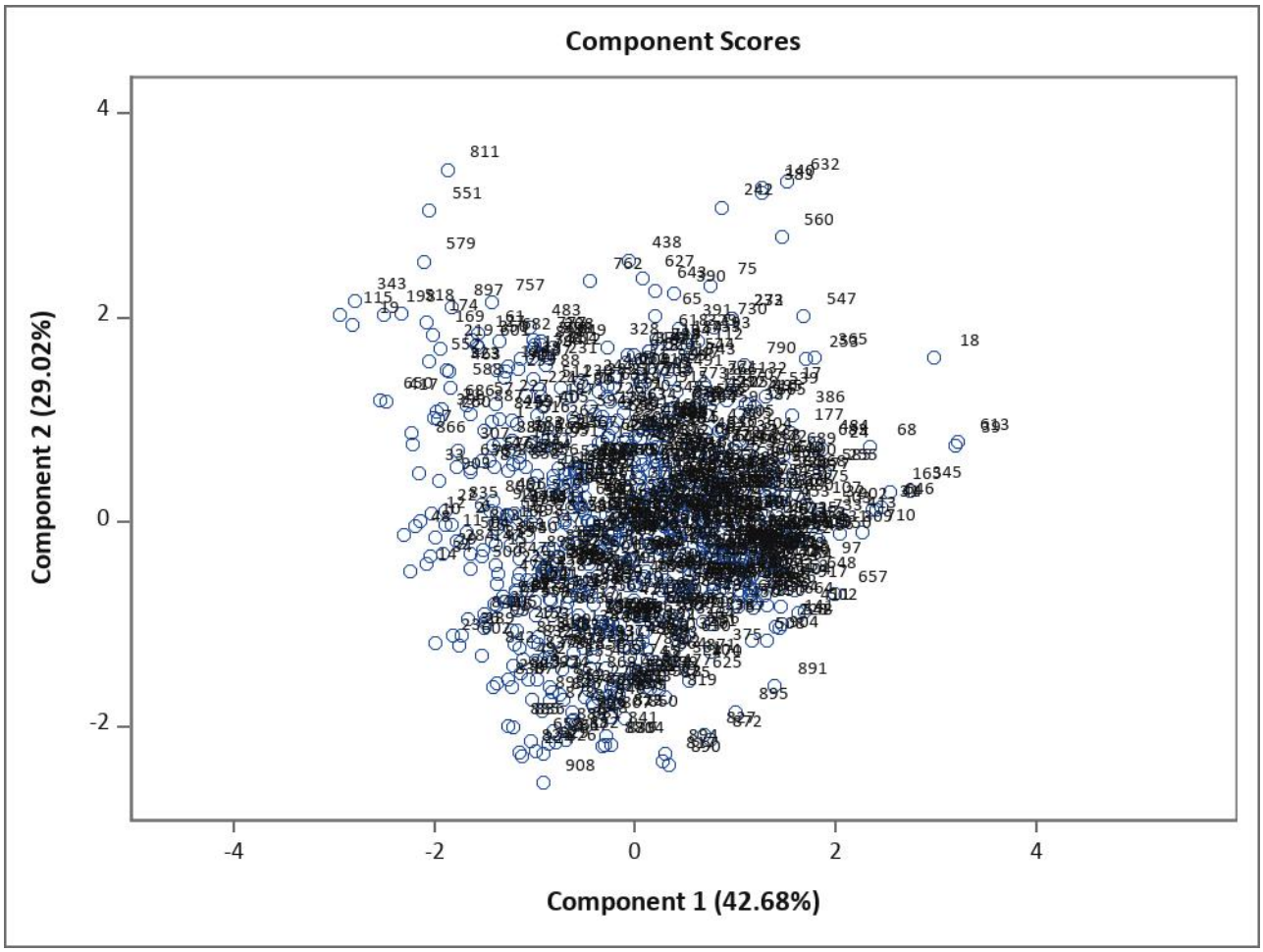

Component Pattern
Component 5 (4.915%)
Component 1 (42.68%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
PSA
Hacceptors
LogD
LogP
NRING
NATOM
ROT
MW

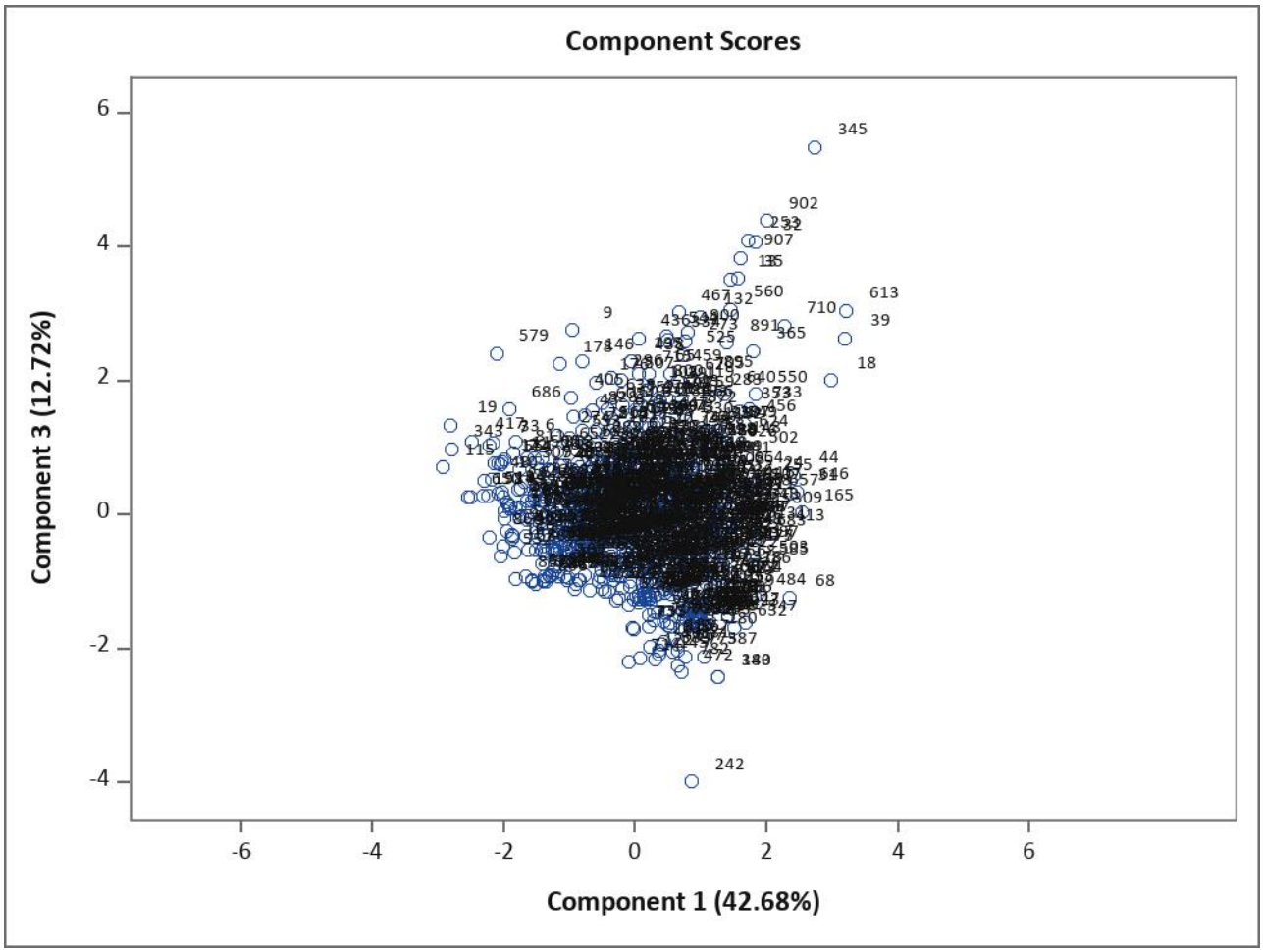

Component Pattern
Component 5 (4.915%)
Component 2 (29.02%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
LogD
LogP
NATOM
PSA
NRING
MW
ROT
Hacceptors

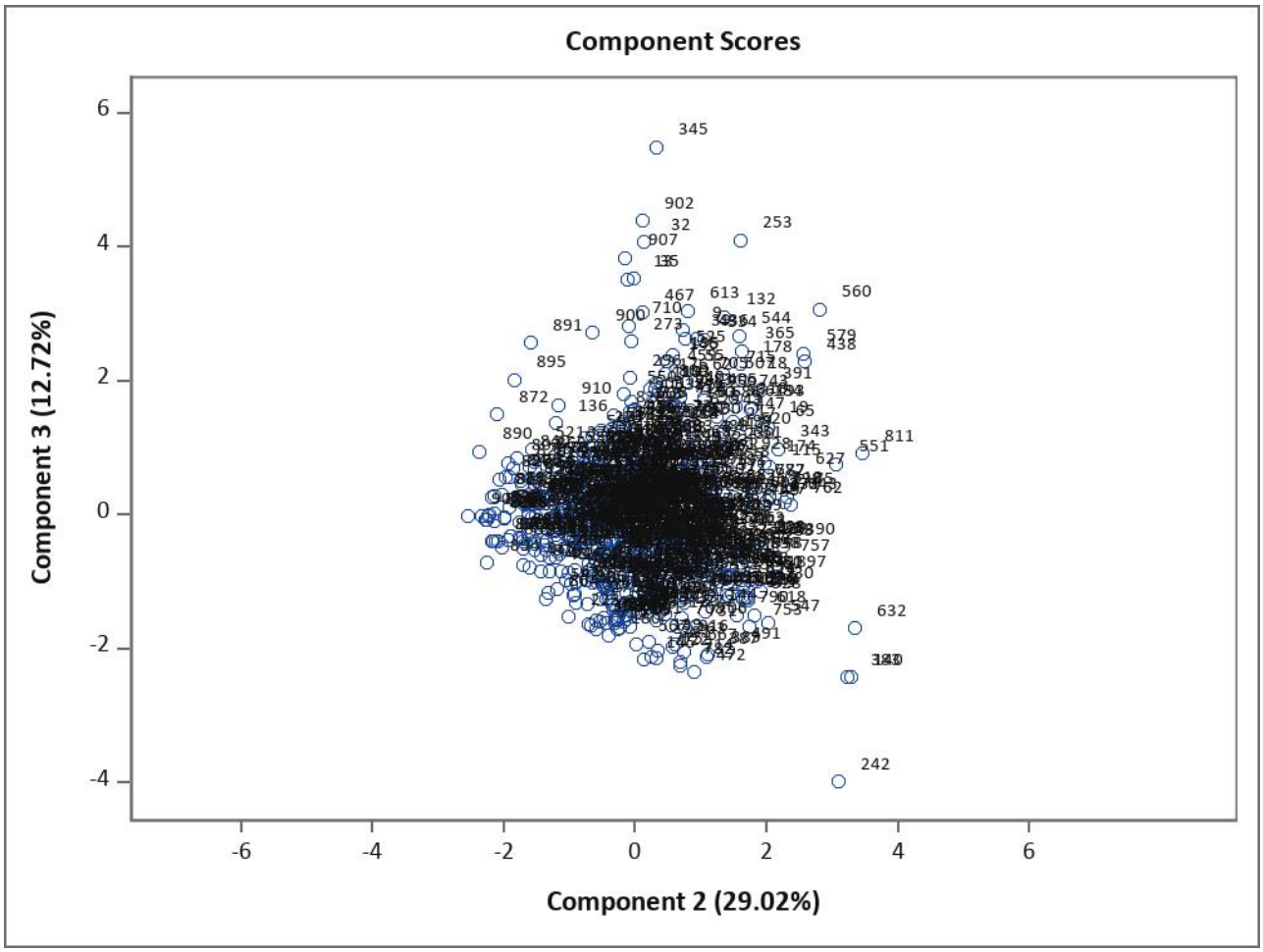

Component Pattern
Component 5 (4.915%)
Component 3 (12.72%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
LogD
LogP
PSA
NATOM
NRING
MW
Hacceptors
ROT

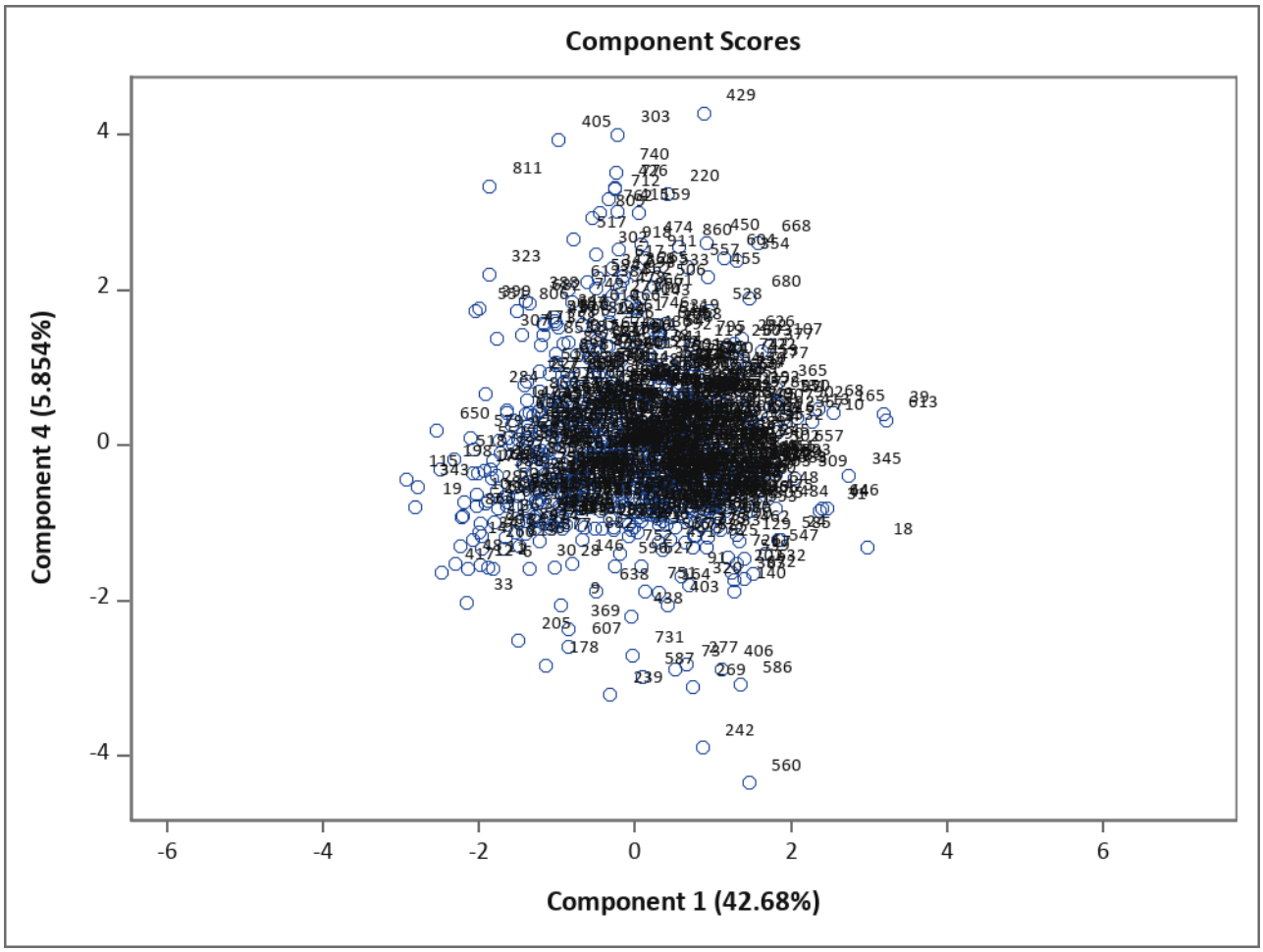

Component Pattern
Component 5 (4.915%)
Component 4 (5.854%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
LogP
LogD
NATOM
NRING
PSA
MW
ROT
Hacceptors

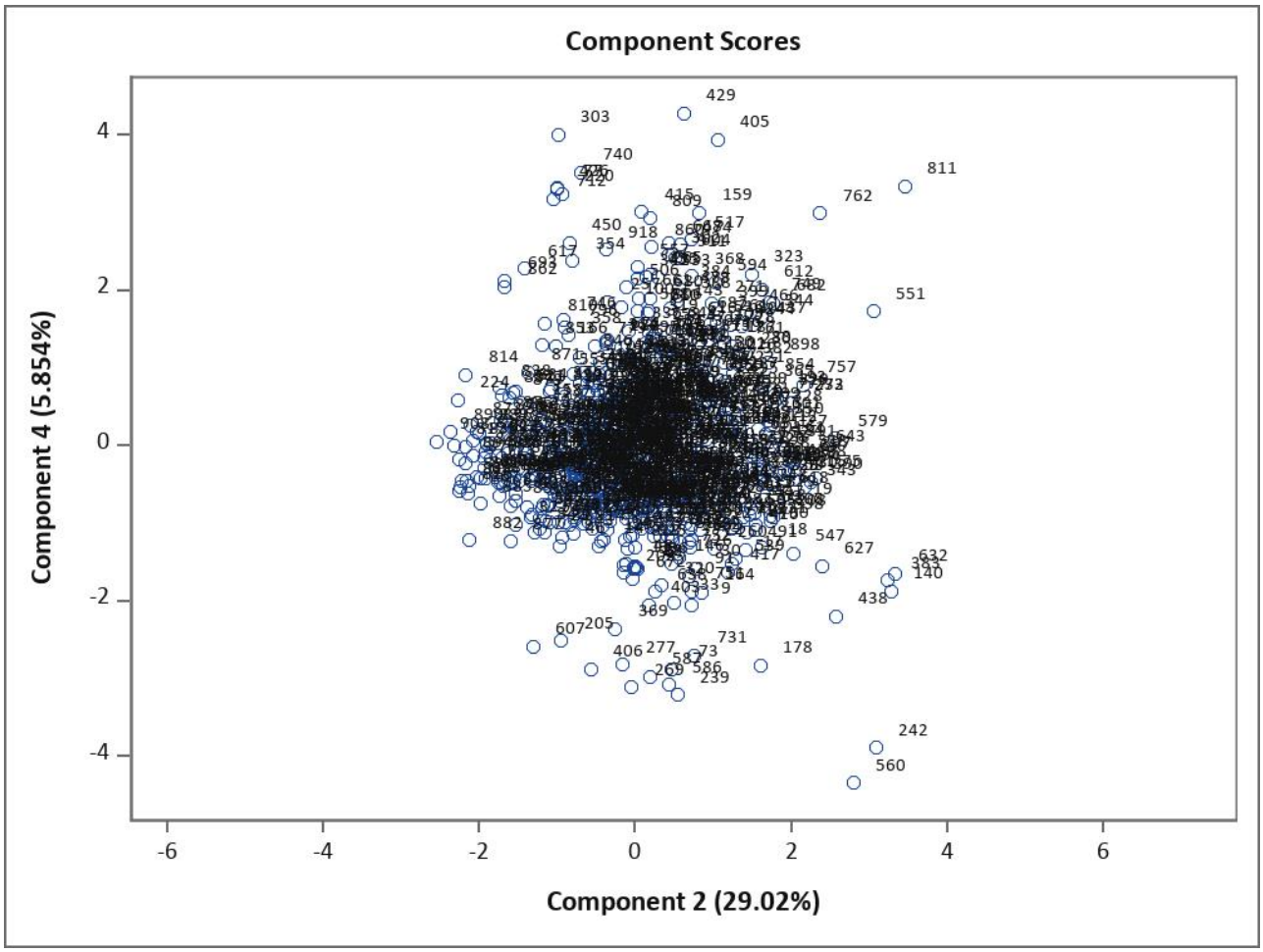
4. Score plots for non-violators

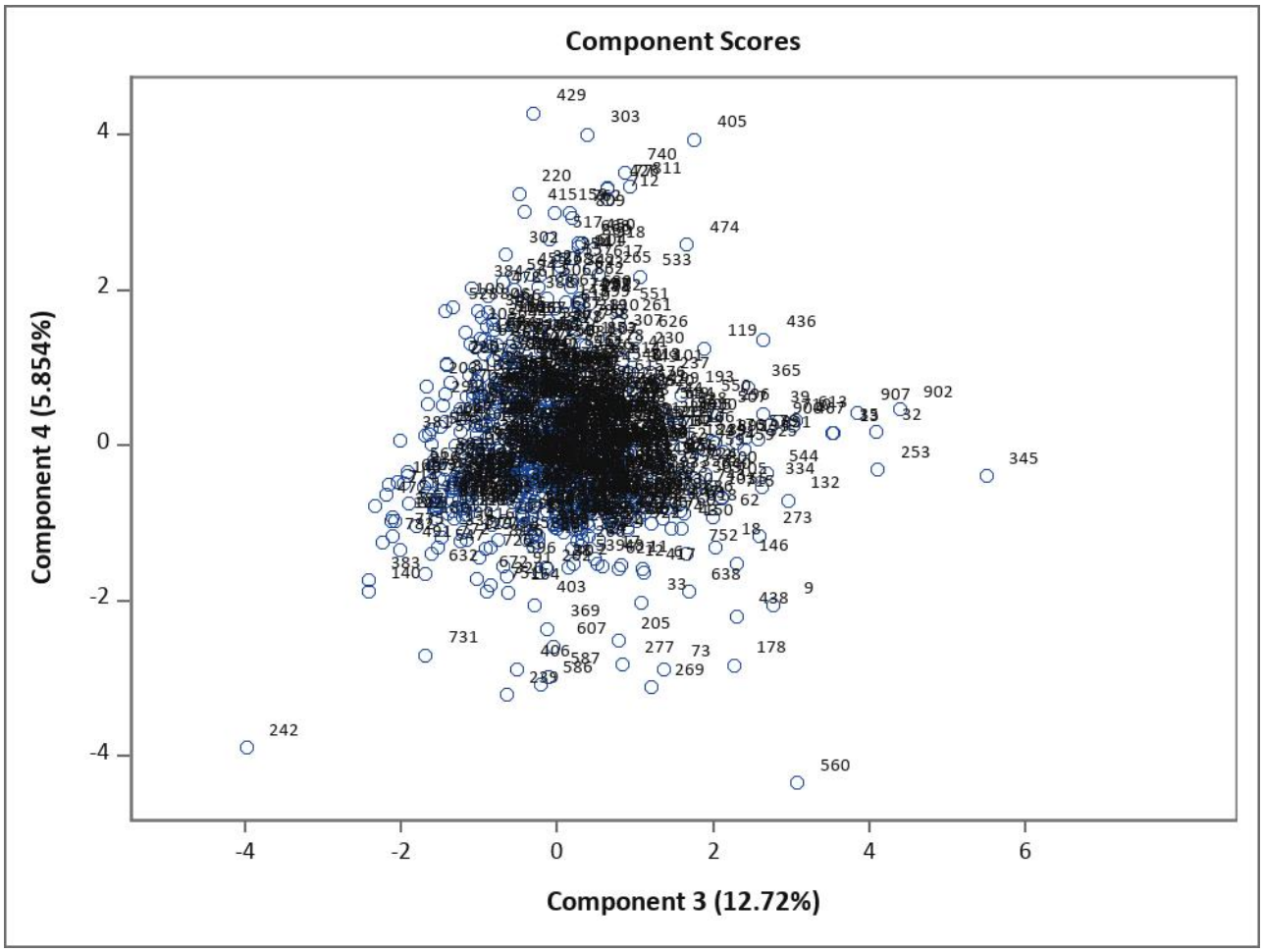


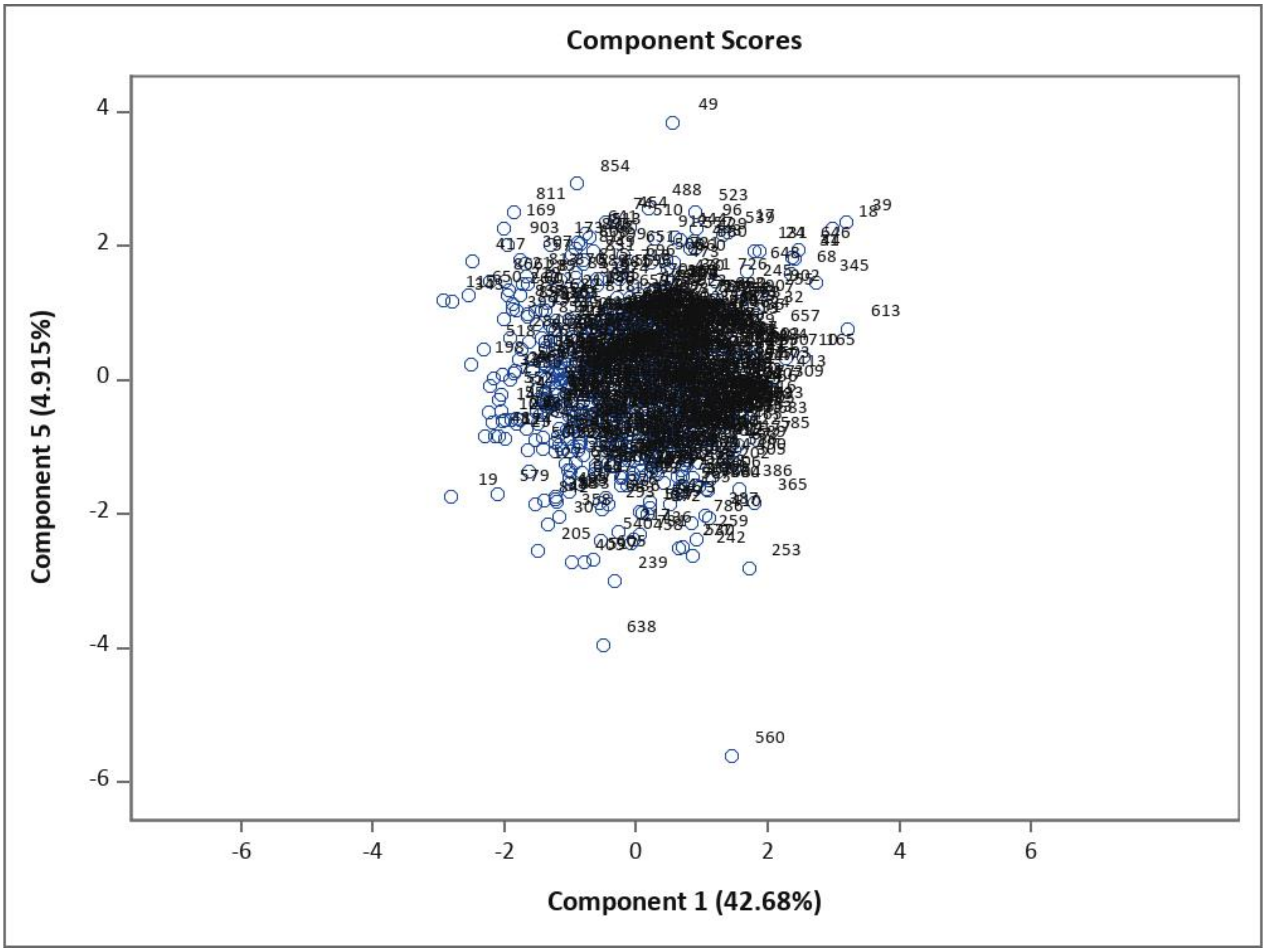

Component Scores
Component 3 (12.72%)
Component 1 (42.68%)
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
345
902
907
560
613
710
39
891
365
579
18
686
19
44
165
68
242

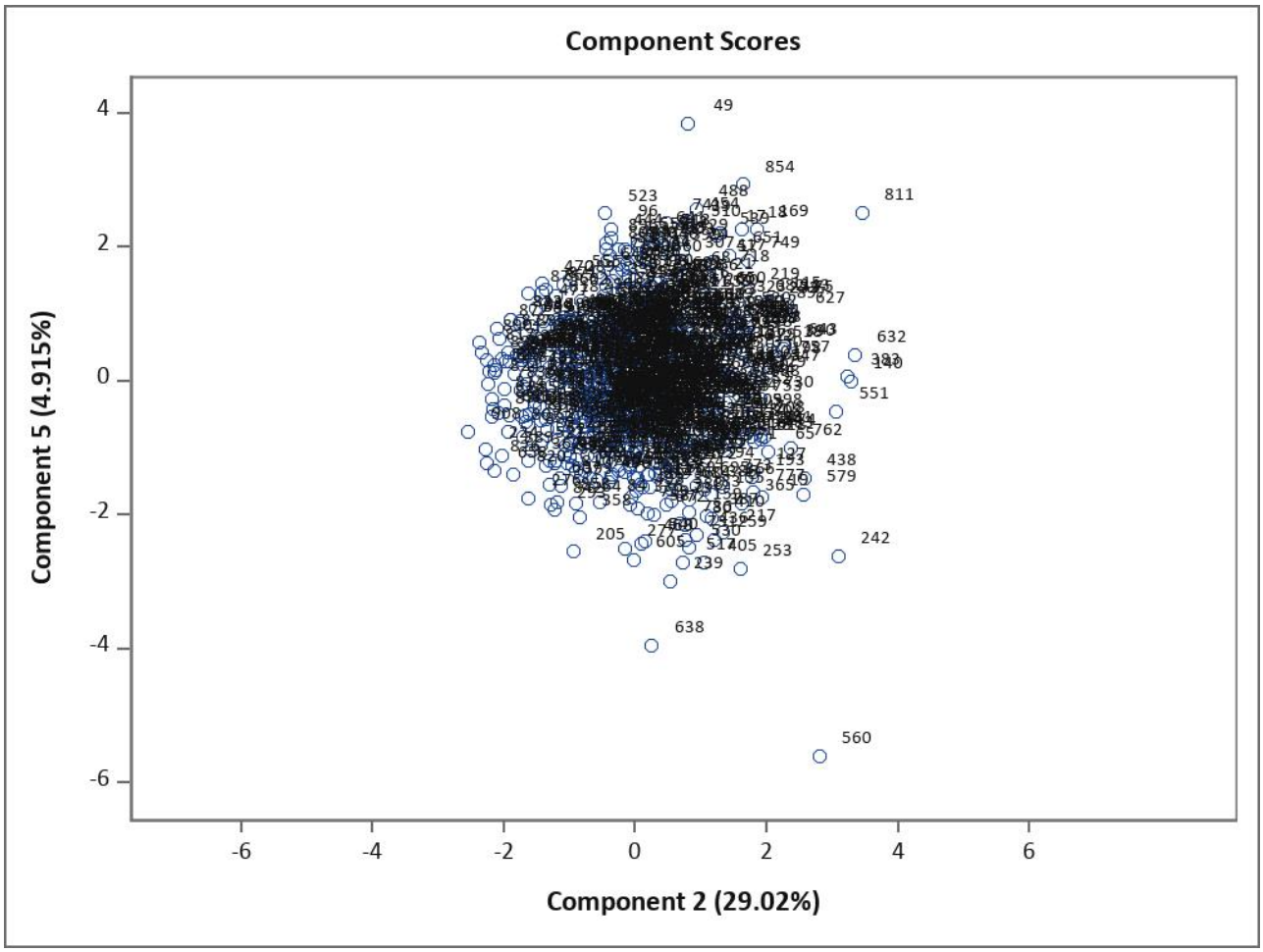

Component Scores
Component 3 (12.72%)
Component 2 (29.02%)
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
345
902
32
253
907
467
613
132
560
710
900
891
273
544
365
579
438
178
895
391
872
910
136
890
343
811
551
627
632
242

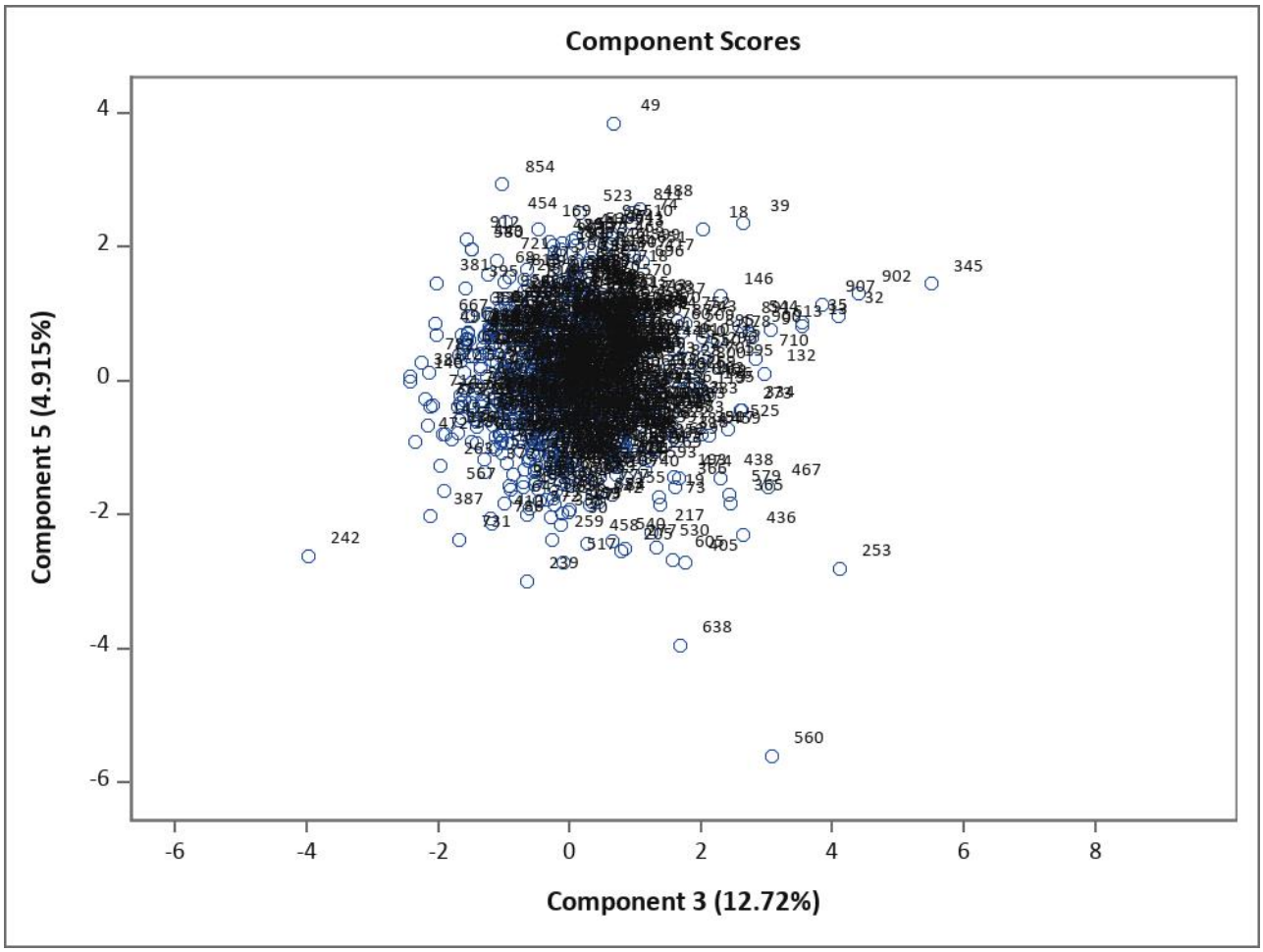

Component Scores
Component 4 (5.854%)
Component 1 (42.68%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6

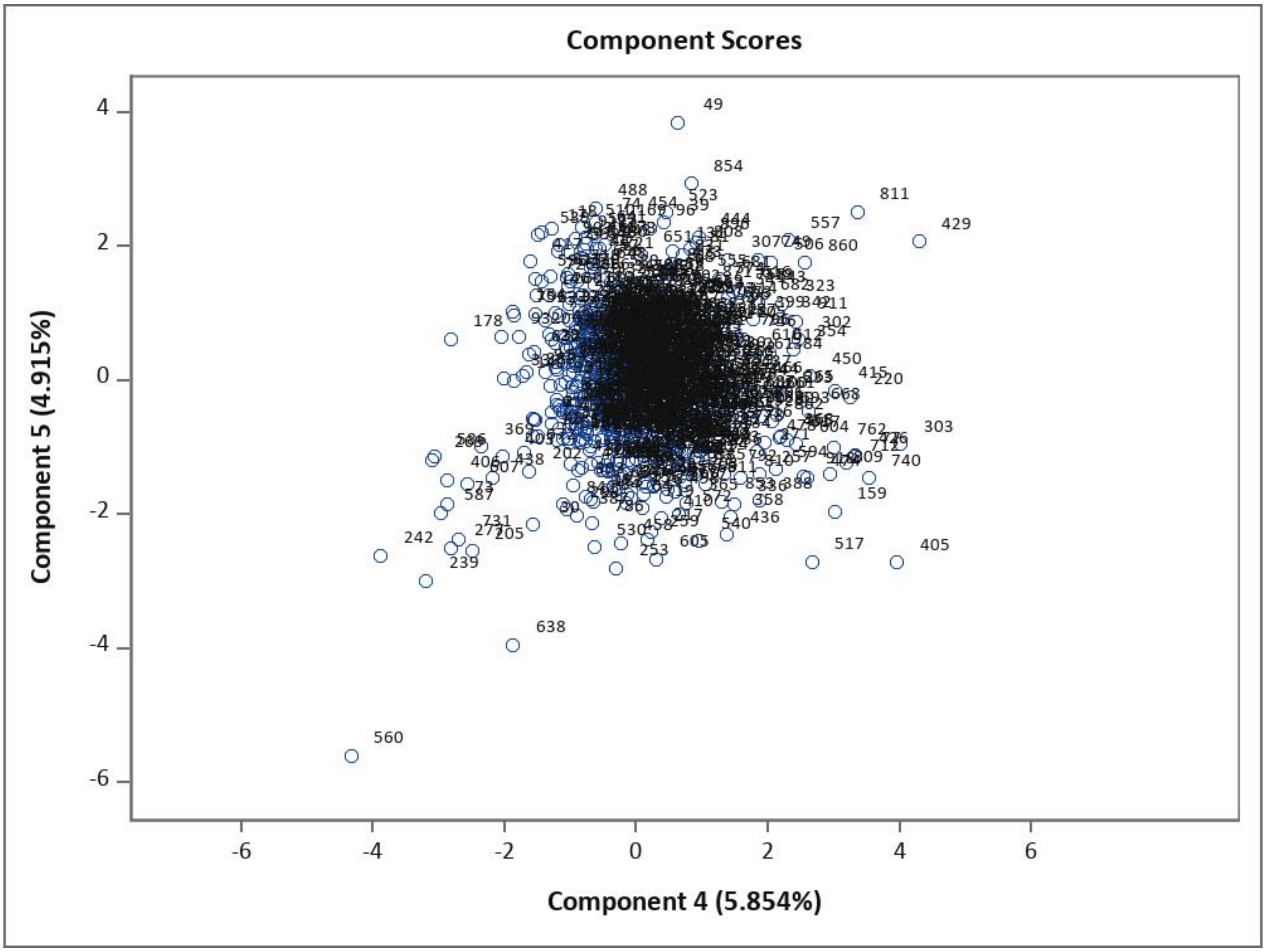

Component Scores
Component 4 (5.854%)
Component 2 (29.02%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
429
303
405
740
811
762
159
450
918
354
551
757
579
547
627
632
140
438
607
205
731
178
406
277
73
586
239
242
560


Component Scores
Component 4 (5.854%)
Component 3 (12.72%)
4
2
0
-2
-4
-4
-2
0
2
4
6
429
303
405
740
220
474
533
436
119
365
907
902
253
345
544
334
132
273
752
146
638
438
9
403
369
607
205
731
406
277
73
178
269
242
560

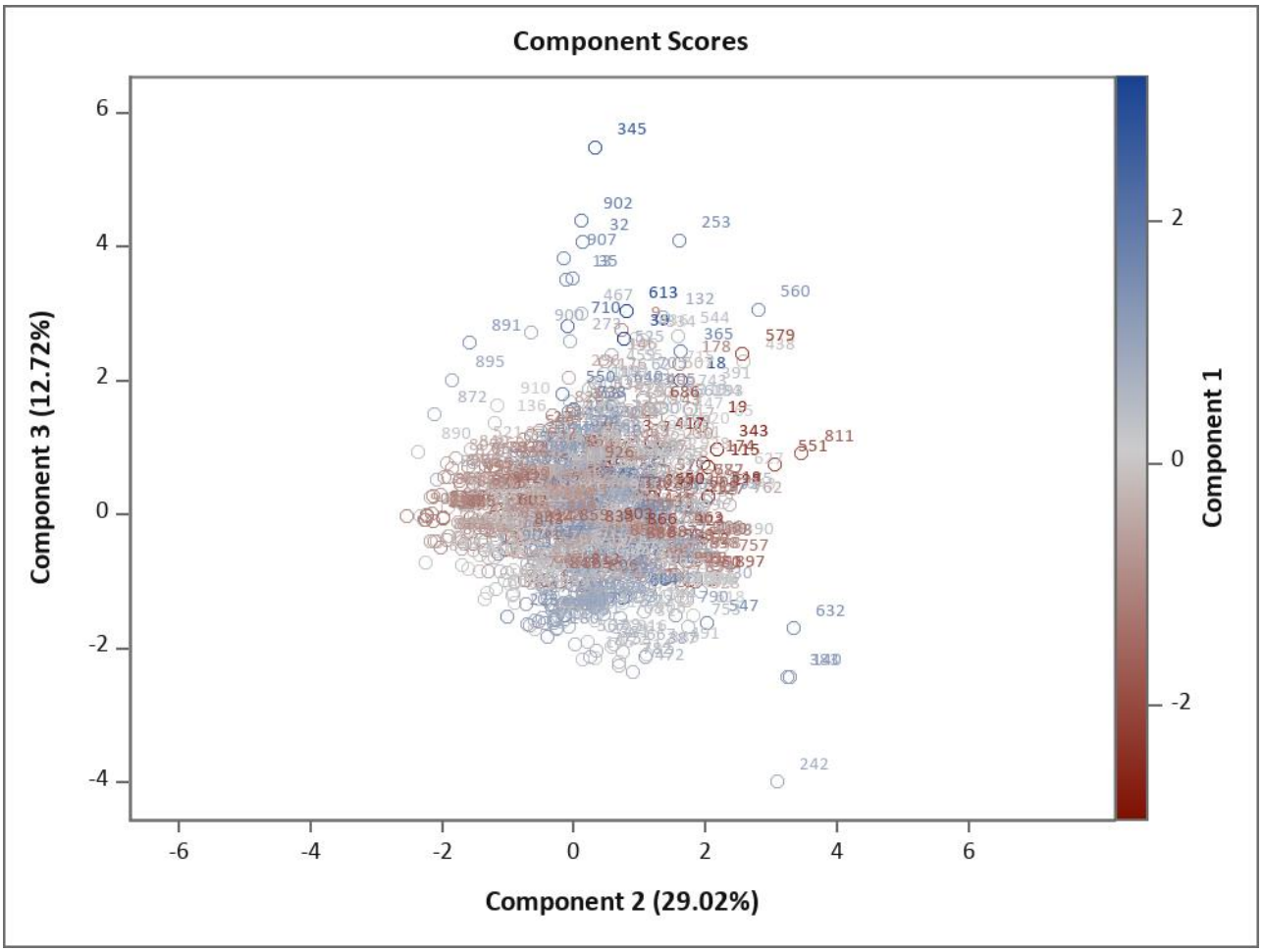

Component Scores
Component 5 (4.915%)
4
2
0
-2
-4
-6
Component 1 (42.68%)
-6
-4
-2
0
2
4
6


Component Scores
Component 5 (4.915%)
Component 2 (29.02%)
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6
49
854
811
523
488
632
551
438
579
242
253
205
638
560

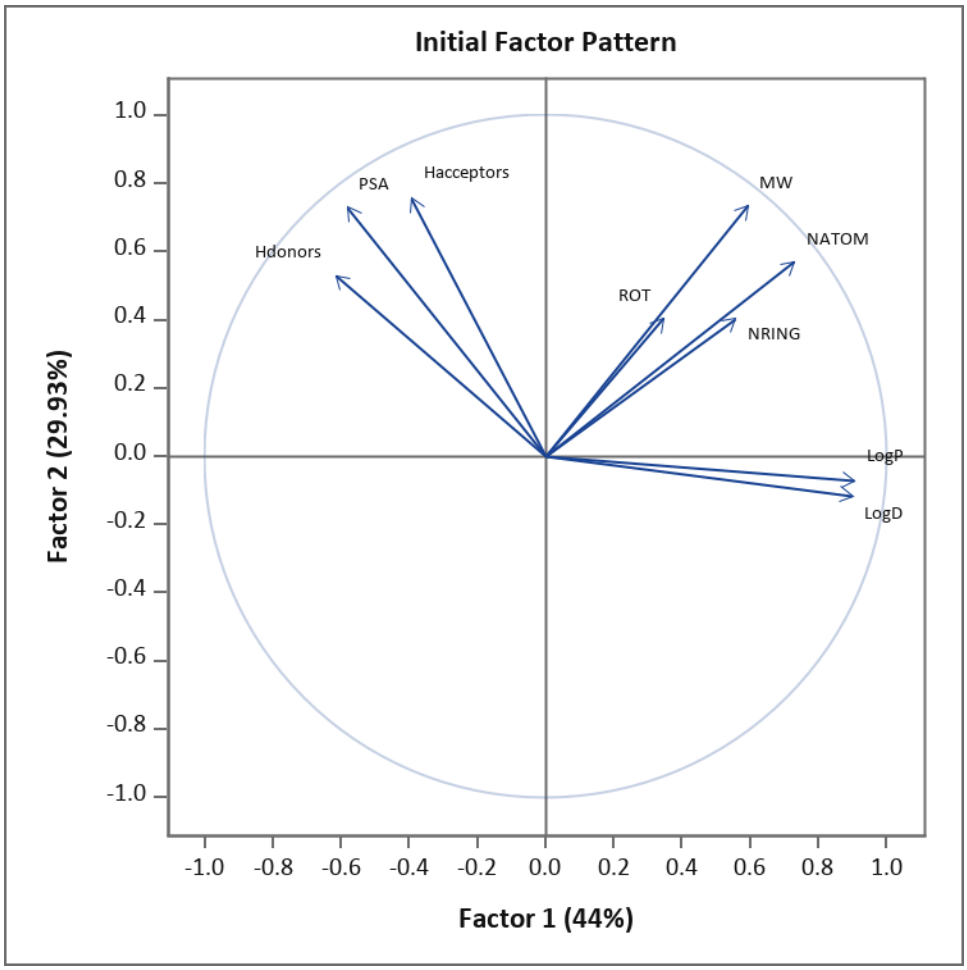

Component Scores
Component 5 (4.915%)
Component 3 (12.72%)
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6
8
49
854
454
169
523
488
18
39
345
902
907
32
146
710
132
525
438
467
579
365
436
253
242
638
560
387
731
259
458
549
217
530
605
405
517
239

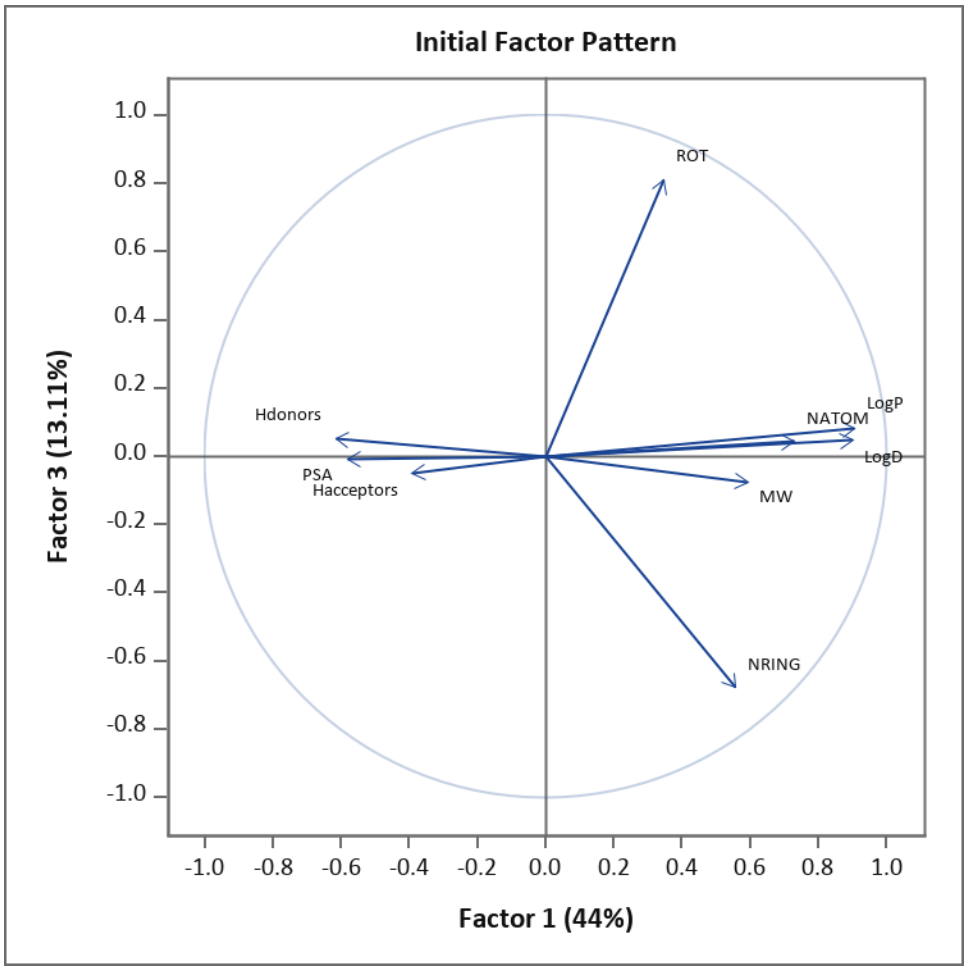


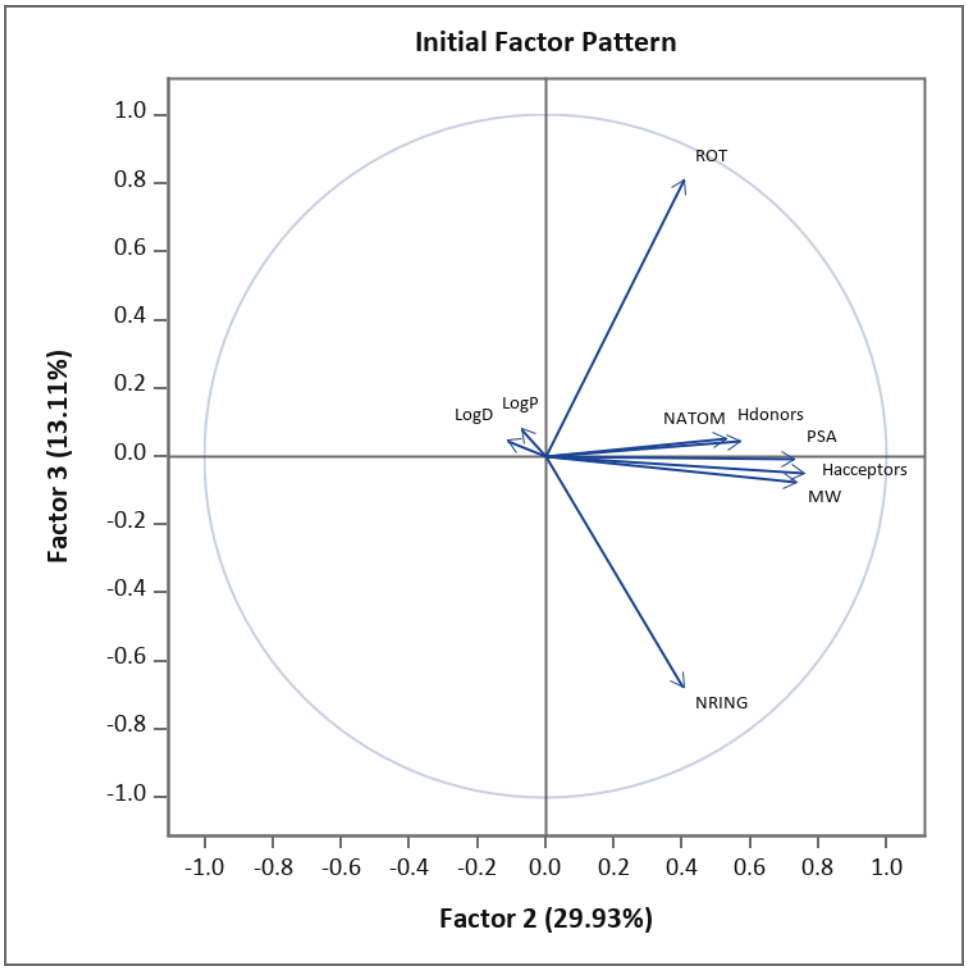
Component Scores
Component 5 (4.915%)
Component 4 (5.854%)

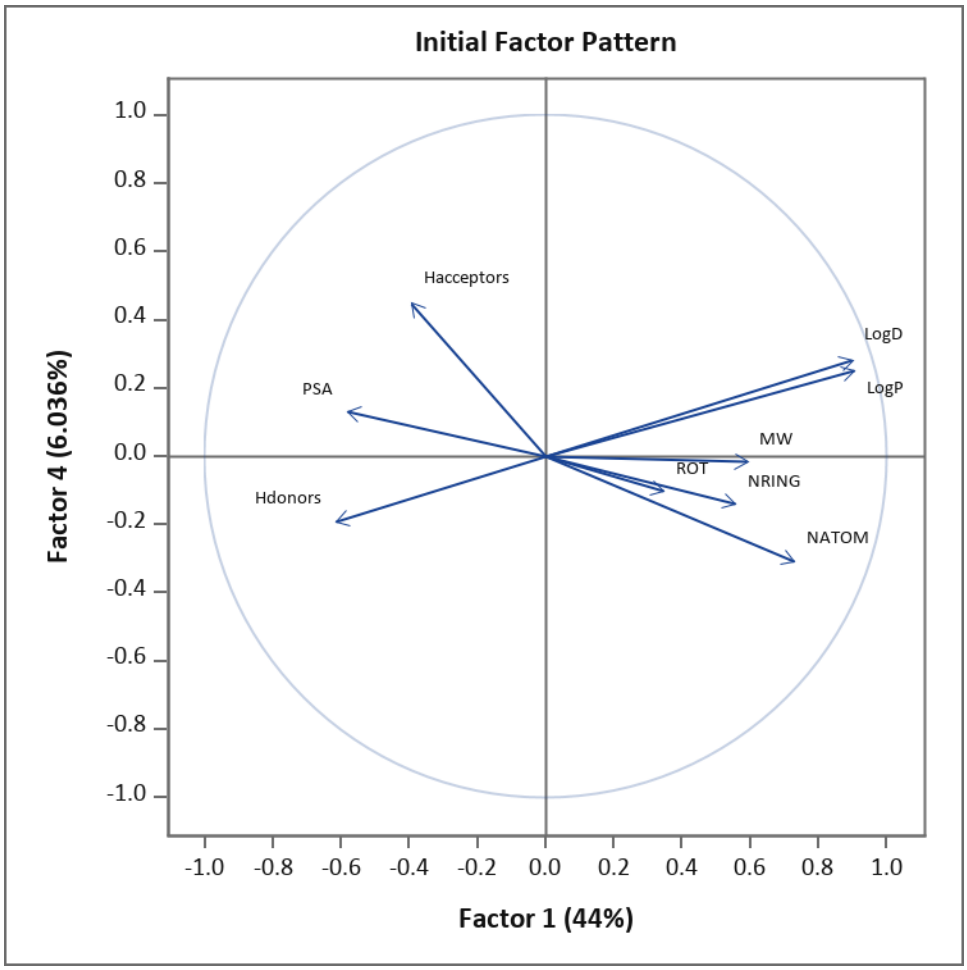

Component Scores
Component 3 (12.72%)
Component 2 (29.02%)
Component 1
6
4
2
0
-2
-4
-6
345
902
32
253
907
613
710
39
560
132
544
891
365
579
178
895
18
550
19
343
811
551
632
242

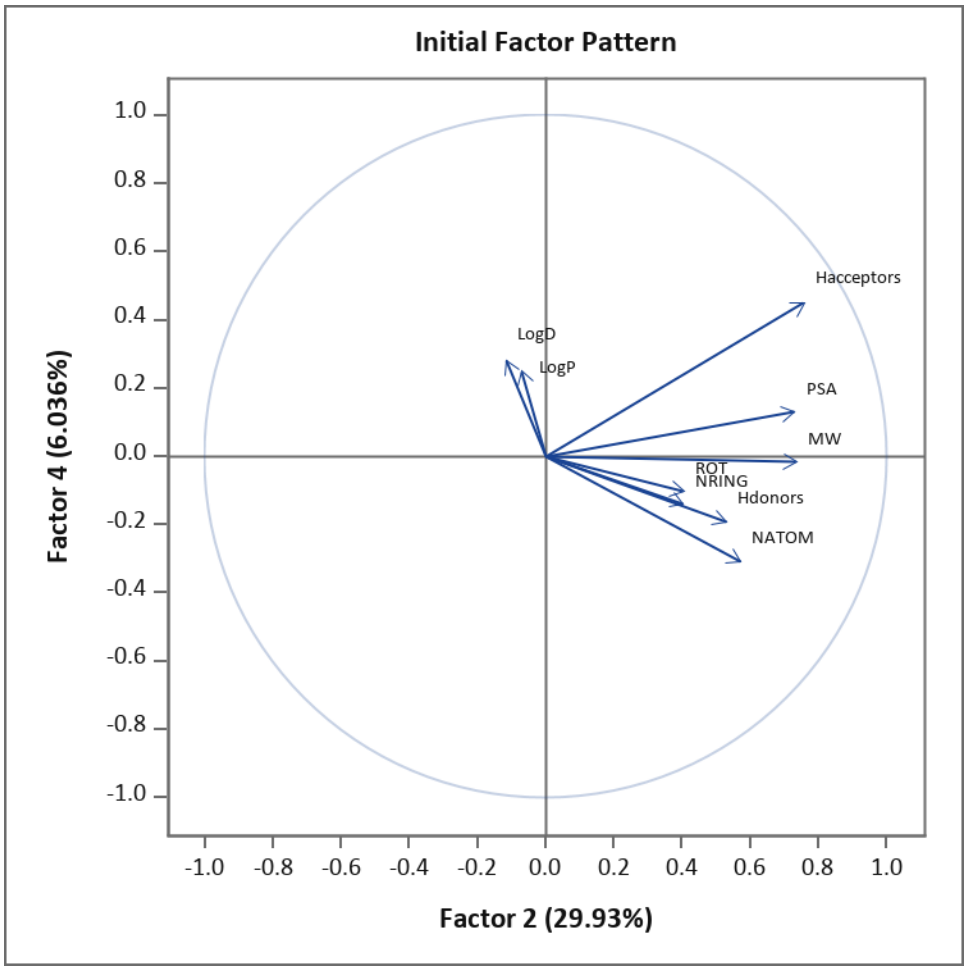
5. Loading plots for non-violators

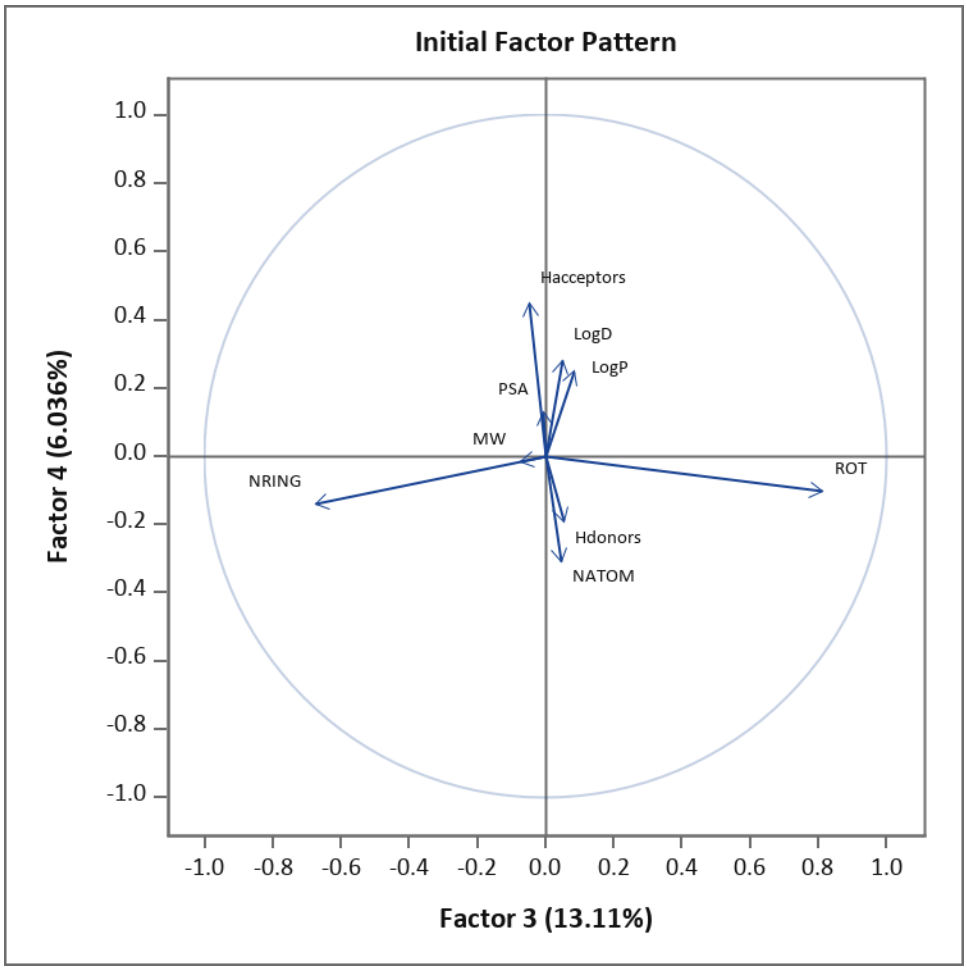


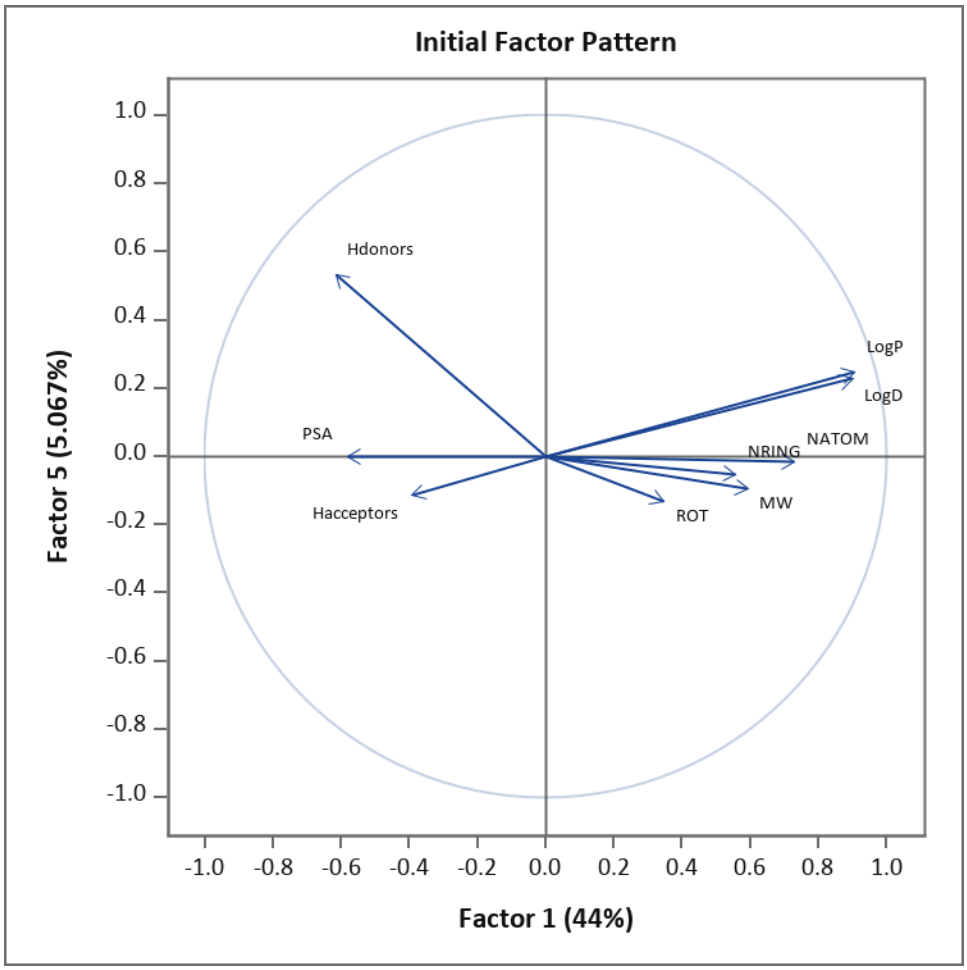

Initial Factor Pattern
Factor 3 (13.11%)
Factor 1 (44%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
ROT
Hdonors
PSA
Hacceptors
NATOM
LogP
LogD
MW
NRING

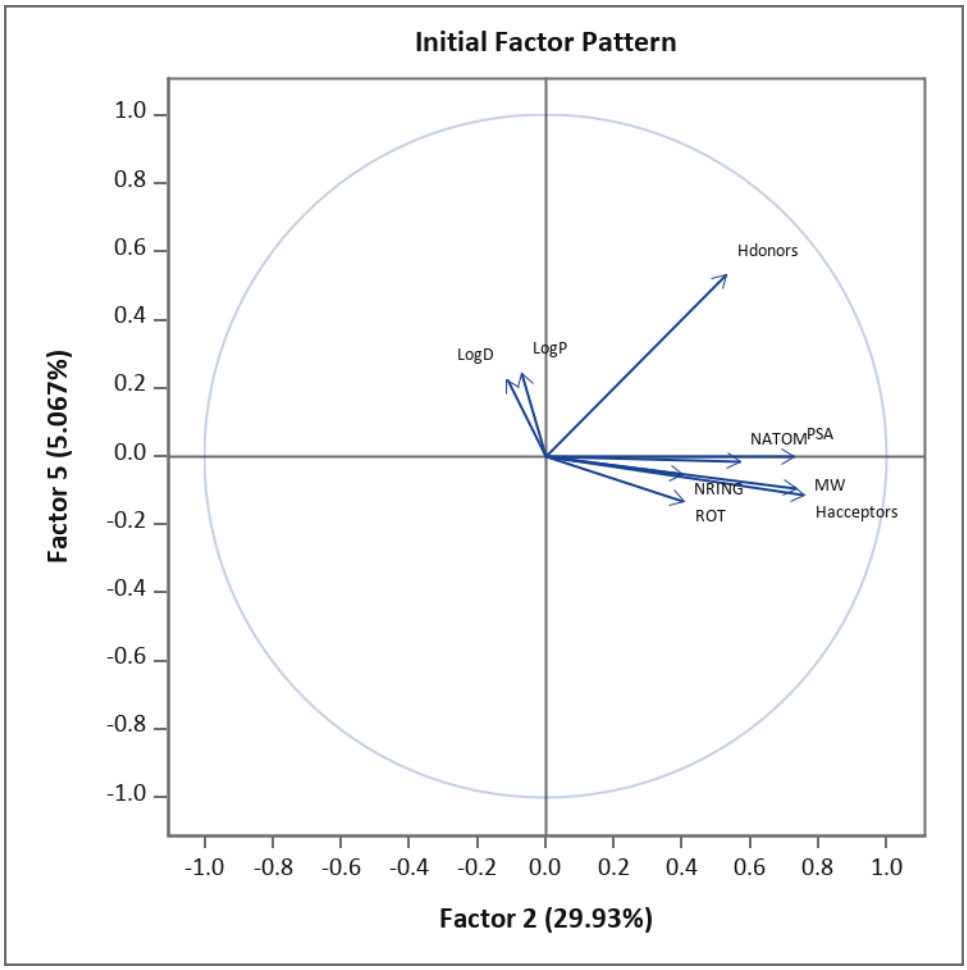

Initial Factor Pattern
Factor 3 (13.11%)
Factor 2 (29.93%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
ROT
LogD
LogP
NATOM
Hdonors
PSA
Hacceptors
MW
NRING

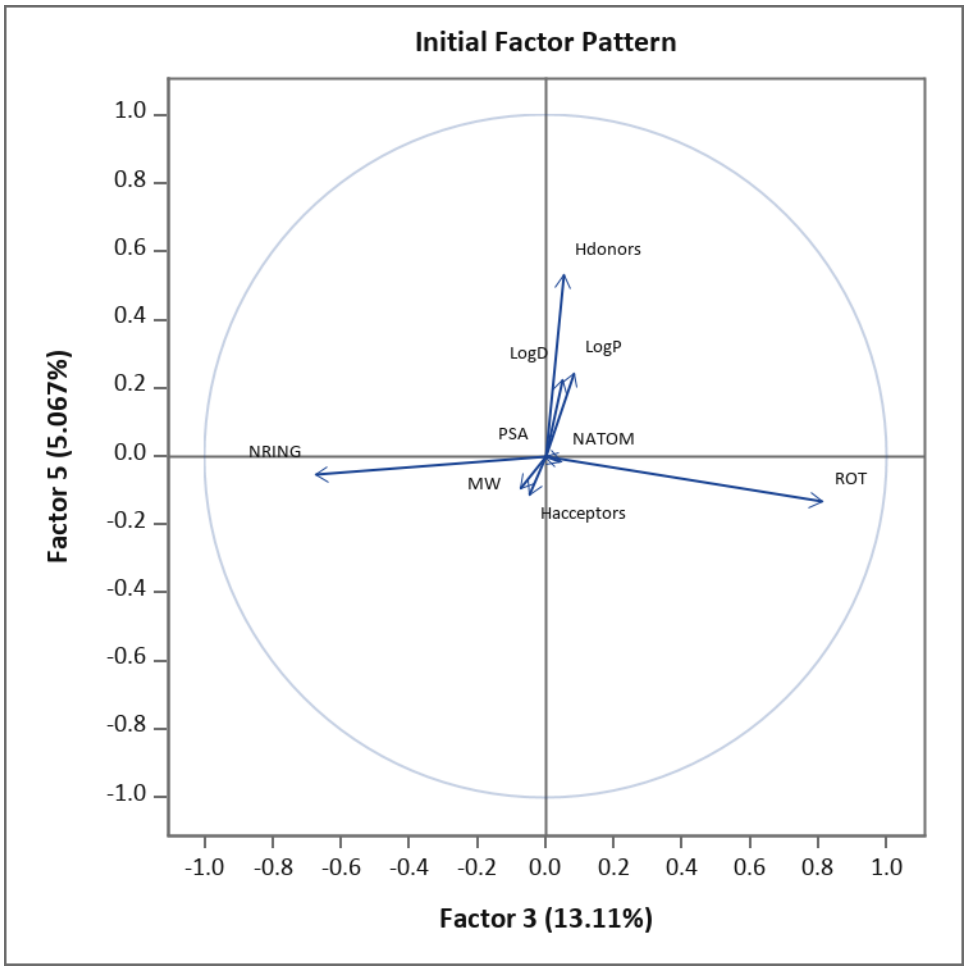

Initial Factor Pattern
Factor 4 (6.036%)
Factor 1 (44%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hacceptors
LogD
LogP
PSA
MW
ROT
NRING
Hdonors
NATOM

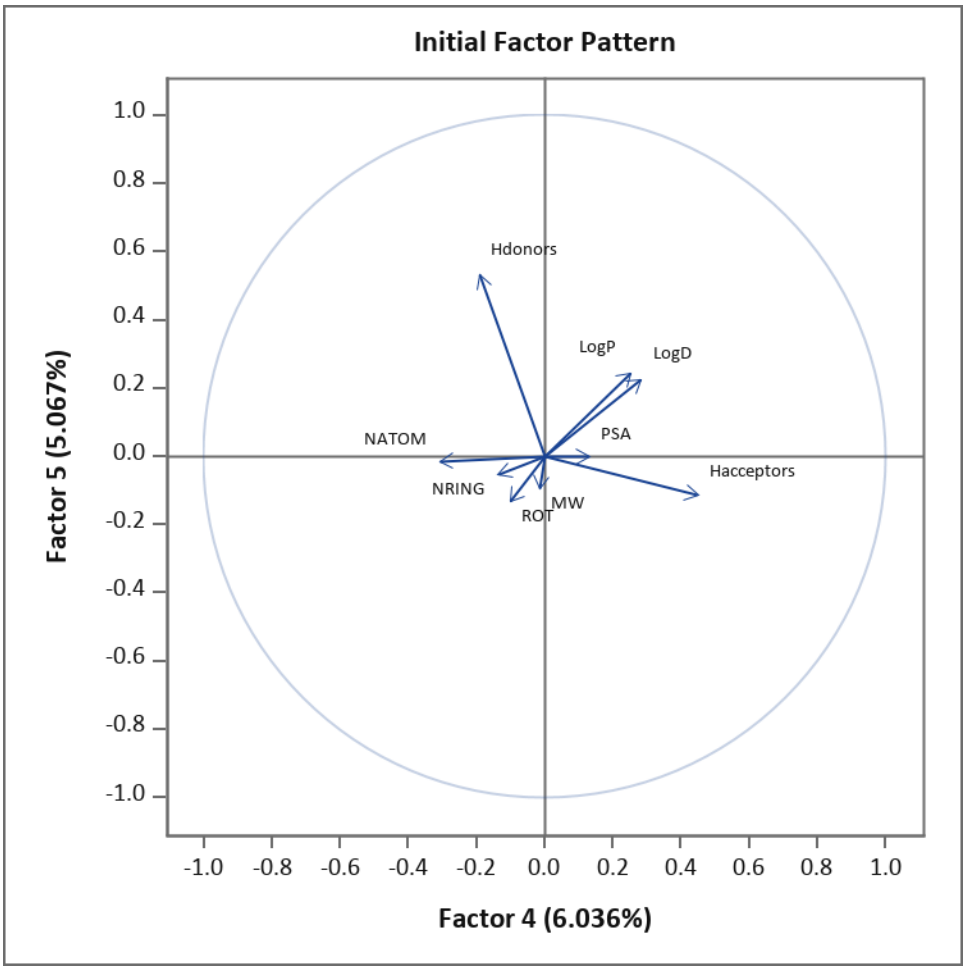

Initial Factor Pattern
Factor 4 (6.036%)
Factor 2 (29.93%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
LogD
LogP
PSA
MW
ROT
NRING
Hdonors
NATOM


Initial Factor Pattern
Factor 4 (6.036%)
Factor 3 (13.11%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hacceptors
LogD
LogP
PSA
MW
NRING
ROT
Hdonors
NATOM


Initial Factor Pattern
Factor 5 (5.067%)
Factor 1 (44%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hdonors
LogP
LogD
PSA
NATOM
NRING
MW
ROT
Hacceptors

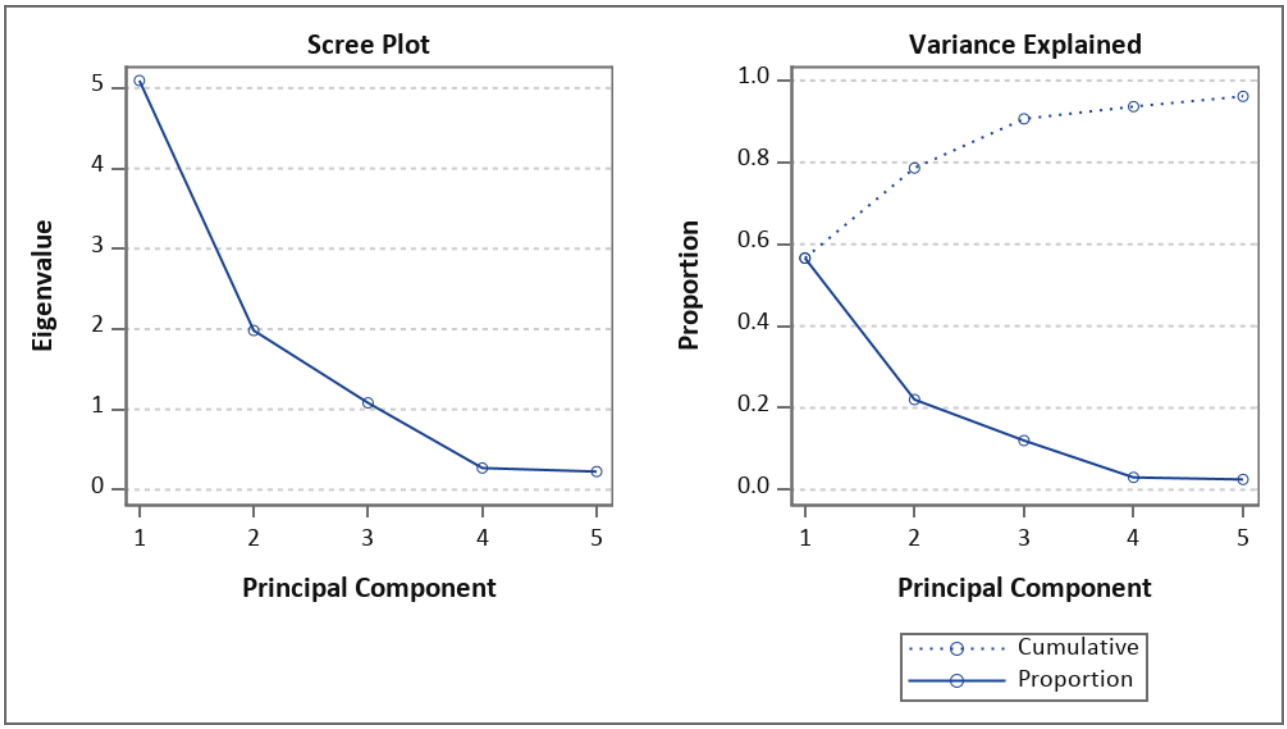

Initial Factor Pattern
Factor 5 (5.067%)
Factor 2 (29.93%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hdonors
LogD
LogP
NATOM
PSA
NRING
MW
ROT
Hacceptors


Initial Factor Pattern
Factor 5 (5.067%)
Factor 3 (13.11%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hdonors
LogD
LogP
PSA
NATOM
NRING
MW
ROT
Hacceptors

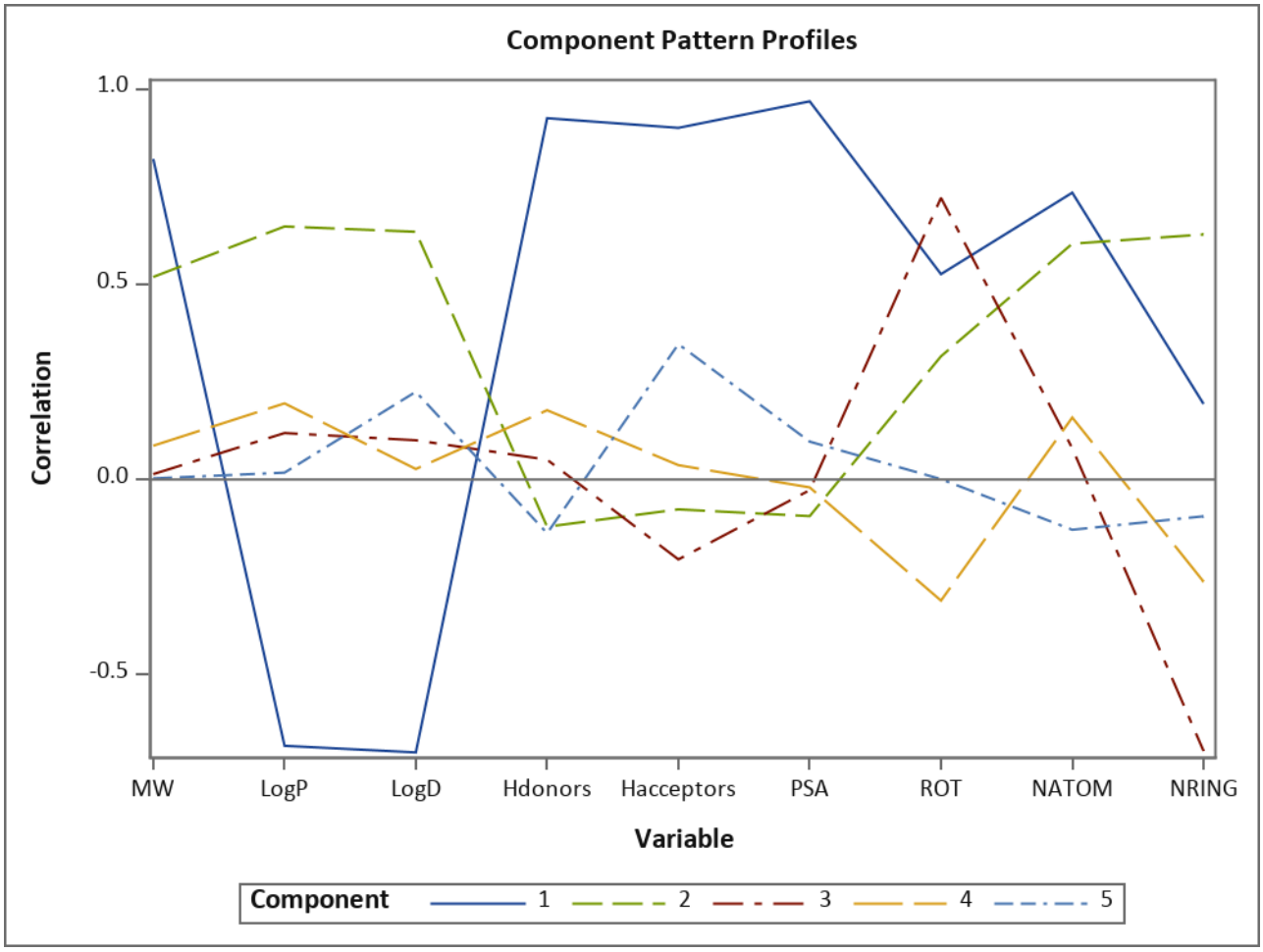

Initial Factor Pattern
Factor 5 (5.067%)
Factor 4 (6.036%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hdonors
LogP
LogD
PSA
NATOM
Hacceptors
NRING
MW
ROT

score9_logD_group_1=2, violators

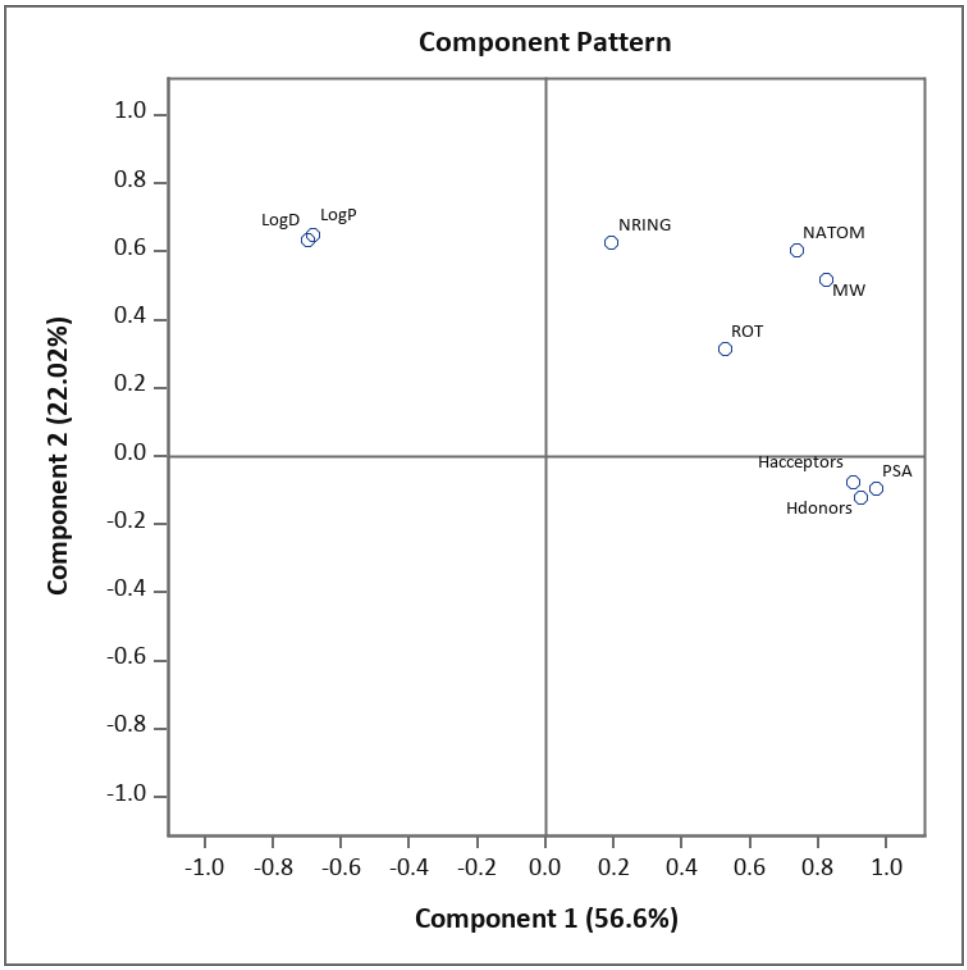
1. Scree plot for violators

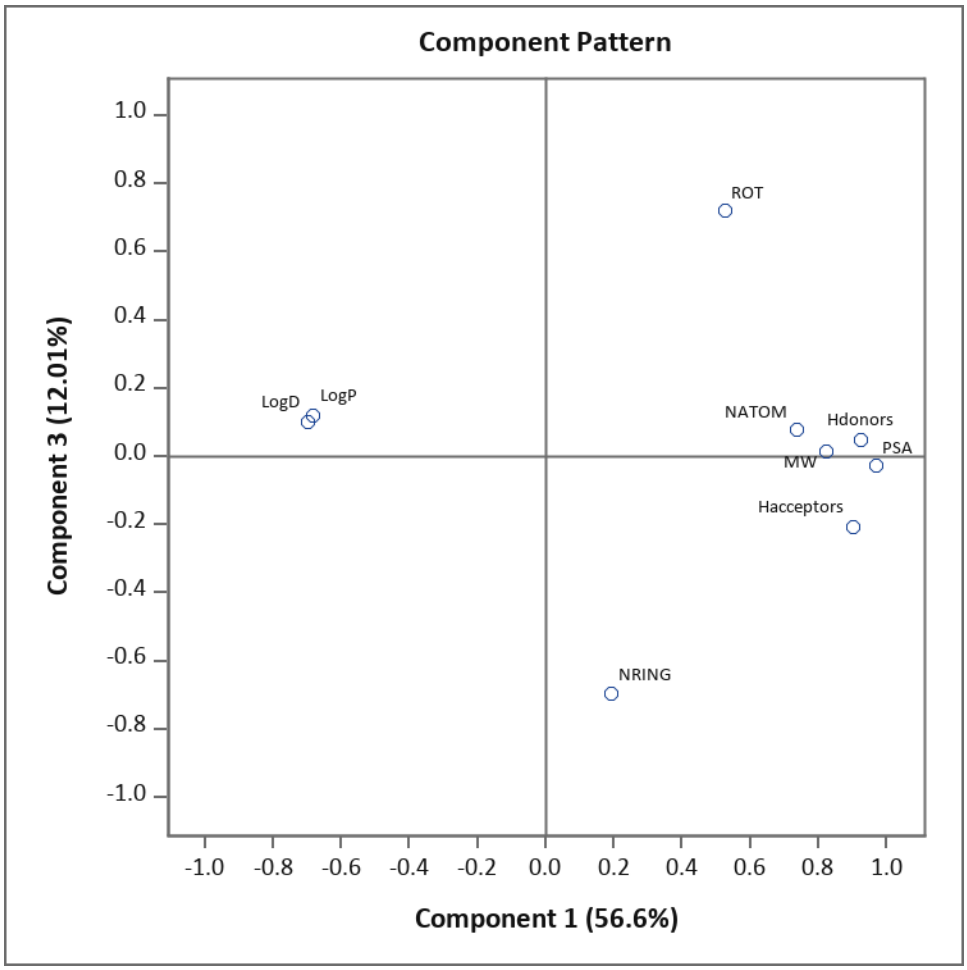


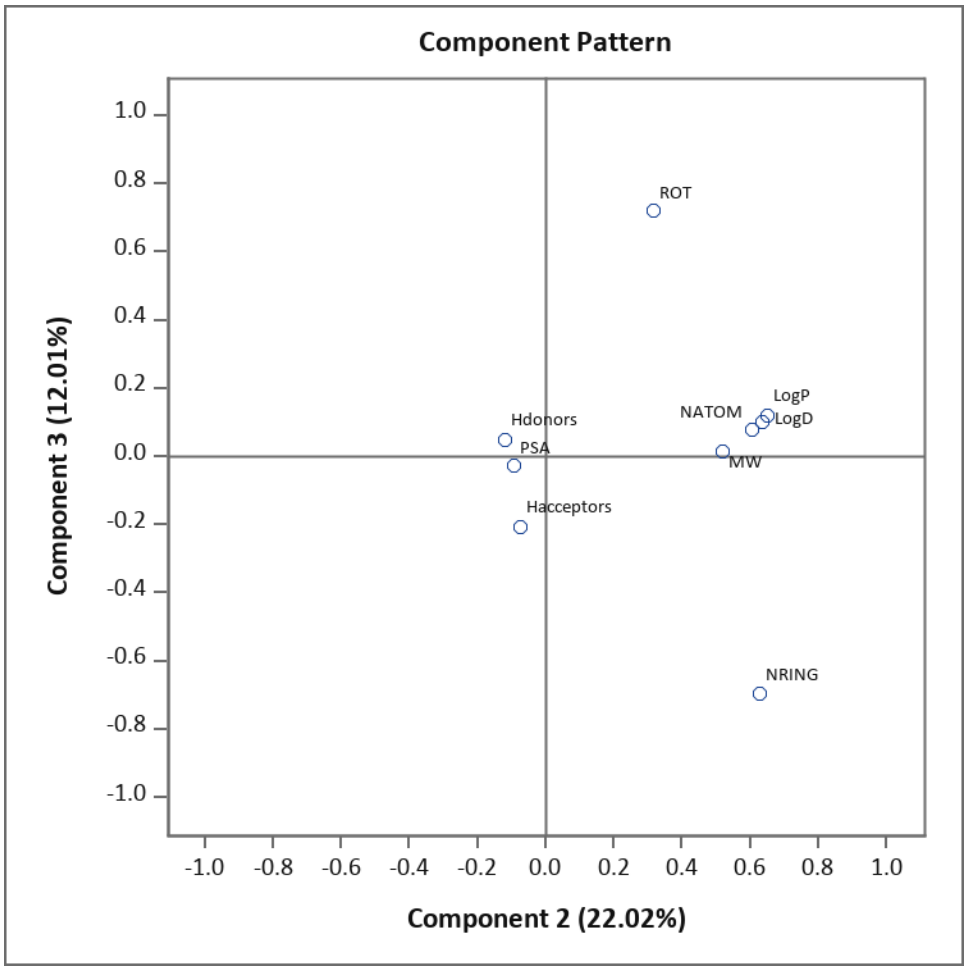
2. Profile plot for violators

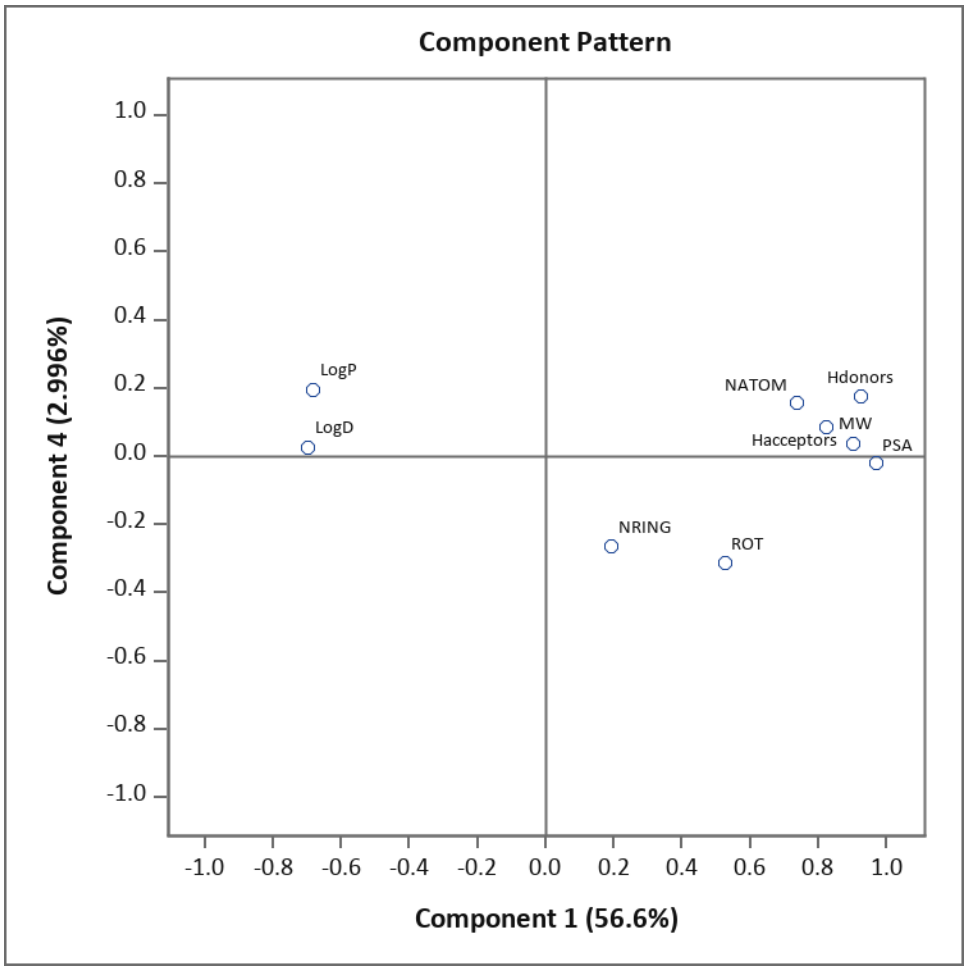


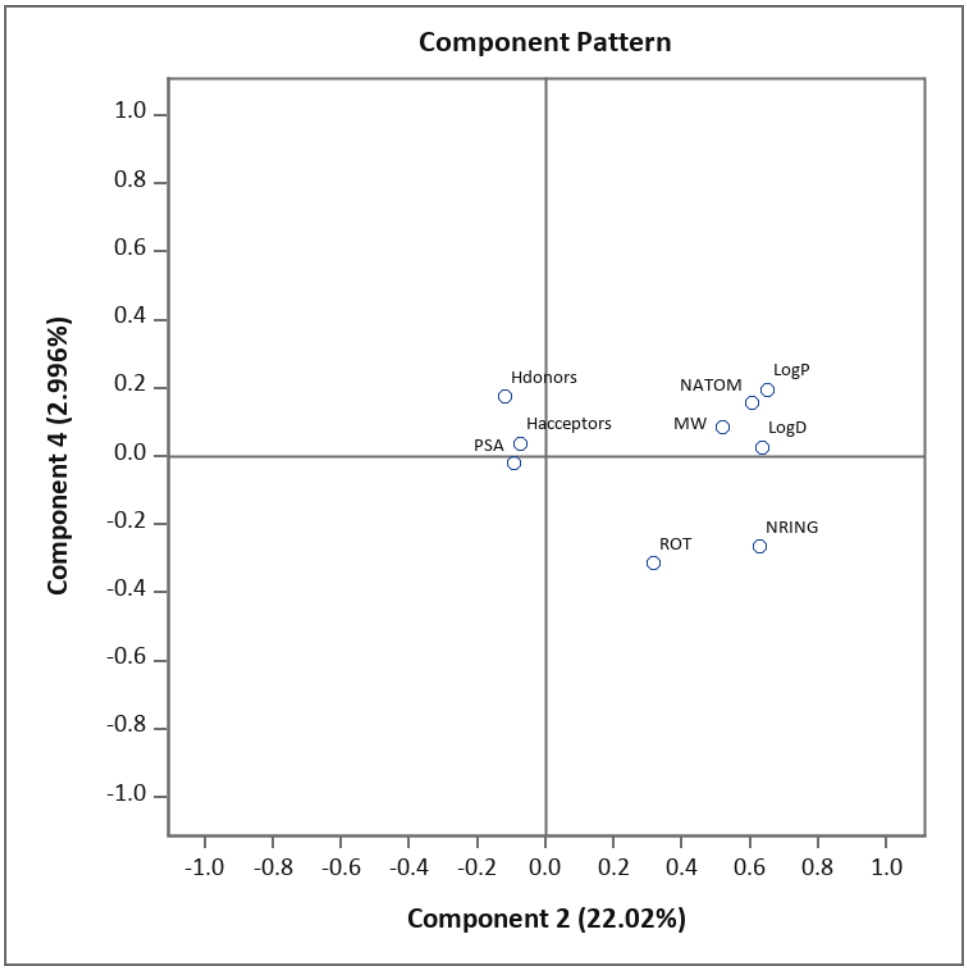
3. Component Pattern plots for violators

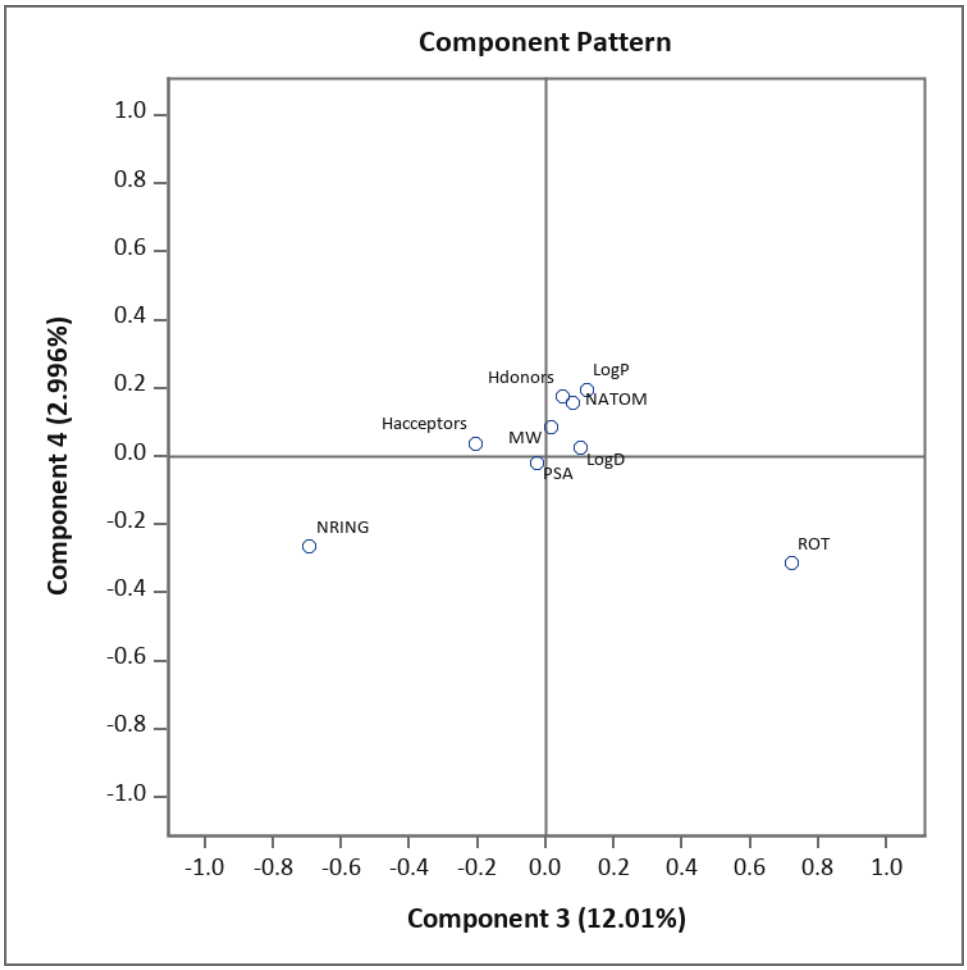


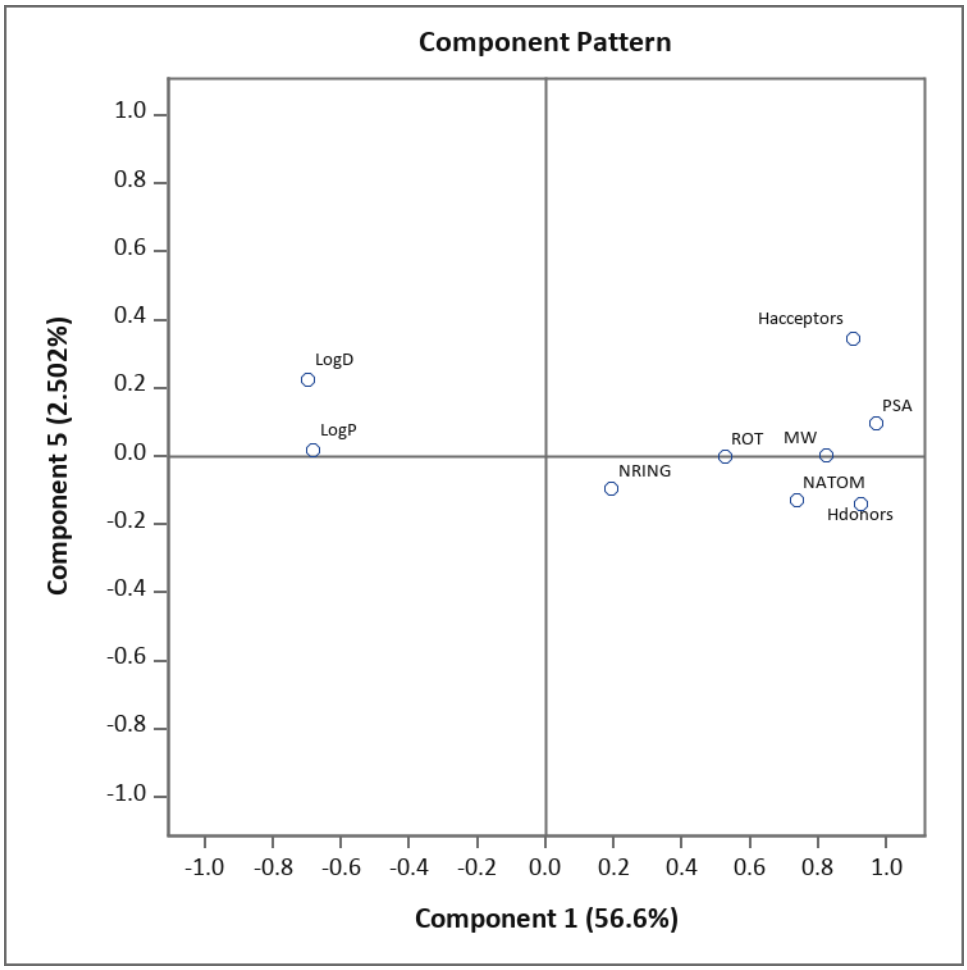

Component Pattern
Component 3 (12.01%)
Component 1 (56.6%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
ROT
LogD
LogP
NATOM
Hdonors
PSA
MW
Hacceptors
NRING

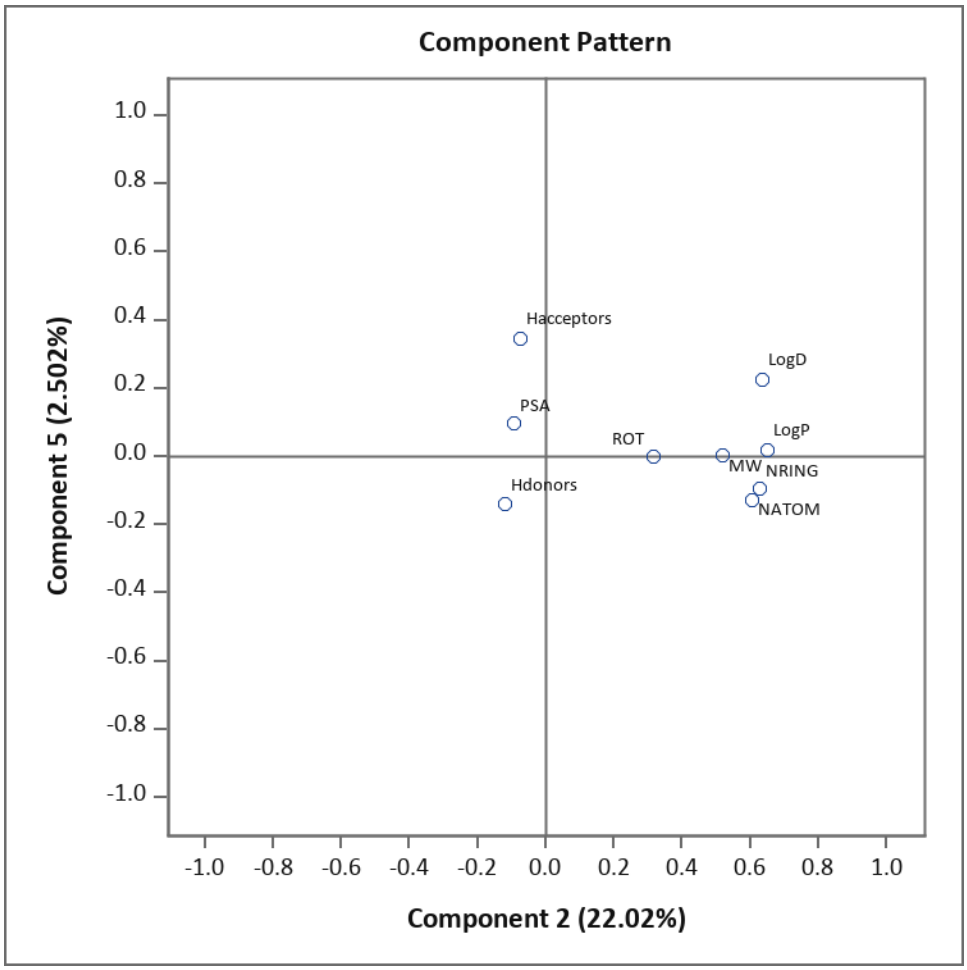

Component Pattern
Component 3 (12.01%)
Component 2 (22.02%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
ROT
LogP
NATOM
LogD
Hdonors
PSA
MW
Hacceptors
NRING

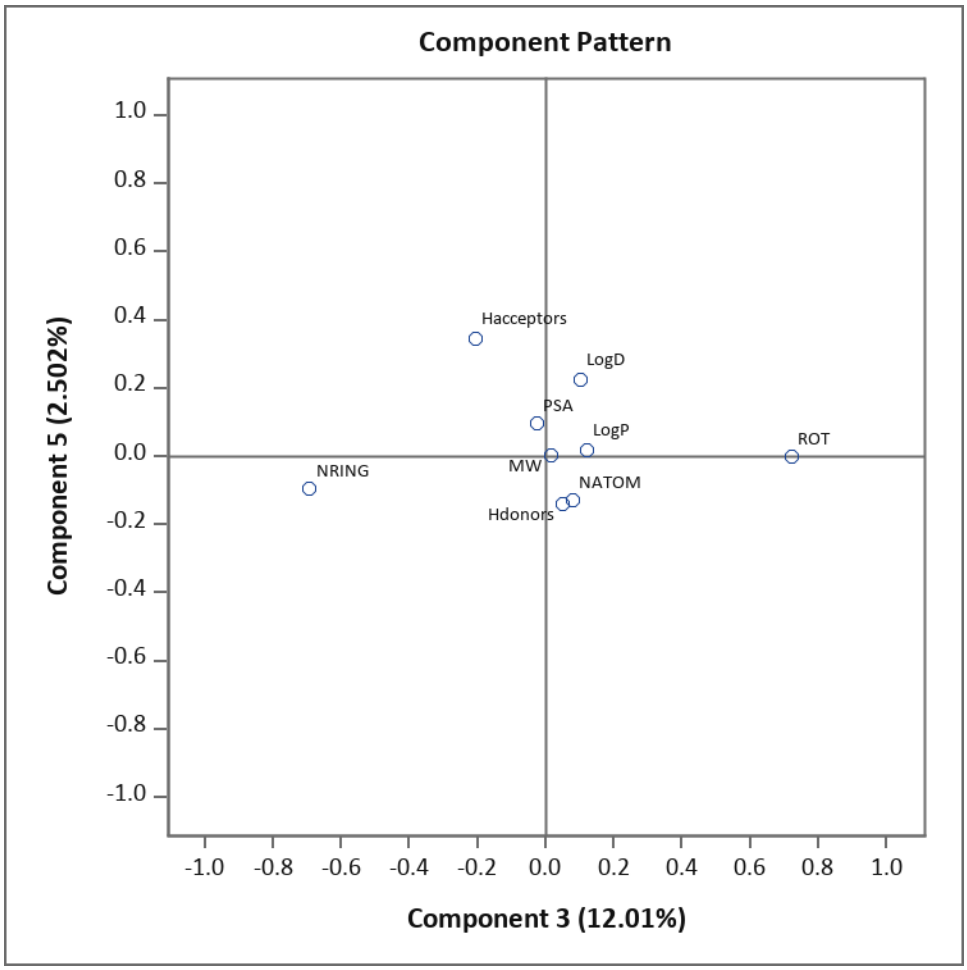

Component Pattern
Component 4 (2.996%)
Component 1 (56.6%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogP
LogD
NATOM
Hdonors
MW
Hacceptors
PSA
NRING
ROT

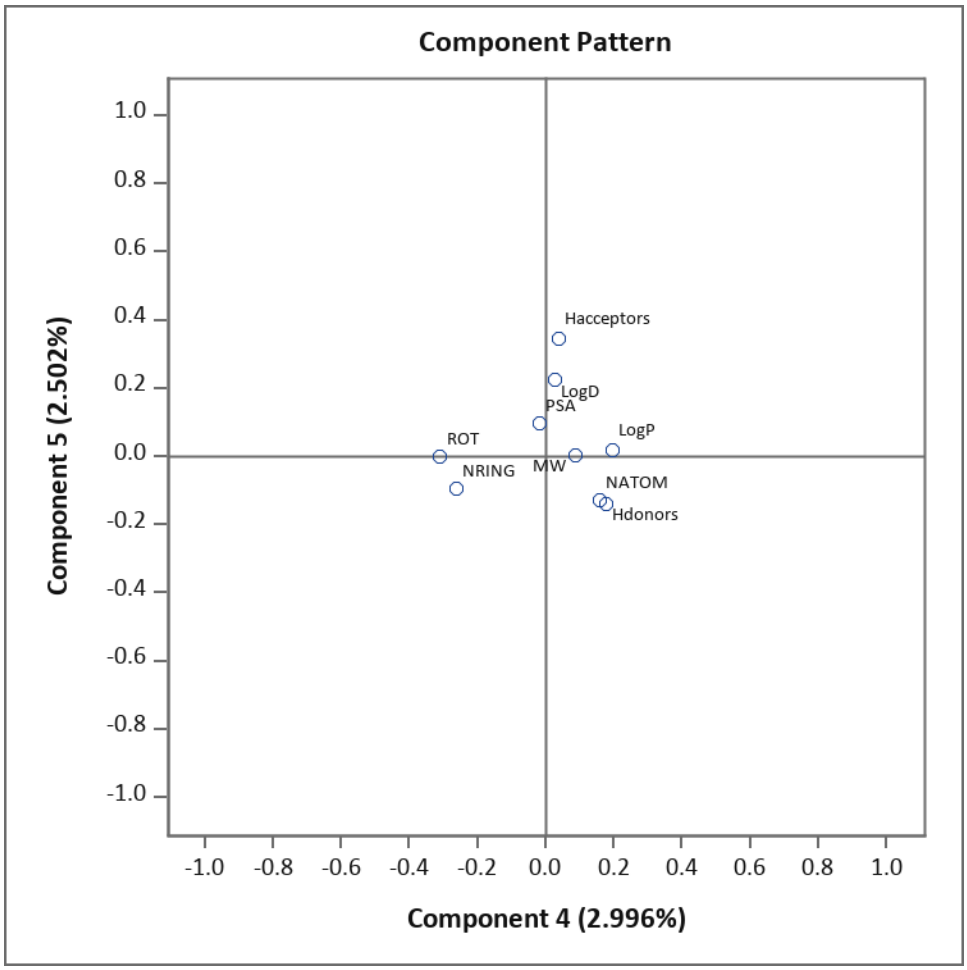

Component Pattern
Component 4 (2.996%)
Component 2 (22.02%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
Hacceptors
PSA
NATOM
LogP
MW
LogD
ROT
NRING


Component Pattern
Component 4 (2.996%)
Component 3 (12.01%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hdonors
LogP
NATOM
Hacceptors
MW
LogD
PSA
NRING
ROT

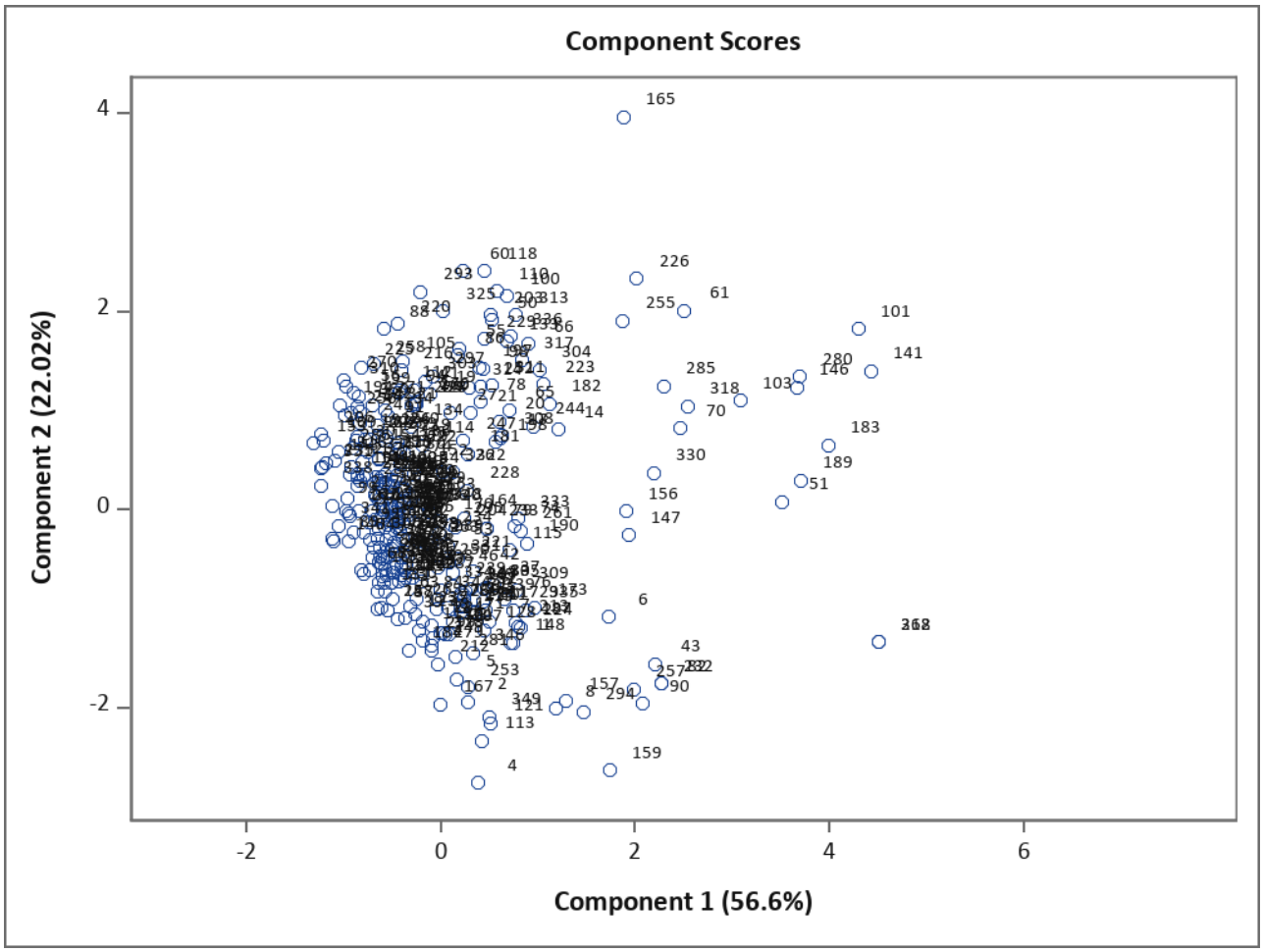

Component Pattern
Component 5 (2.502%)
Component 1 (56.6%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogD
LogP
Hacceptors
PSA
ROT
MW
NRING
NATOM
Hdonors

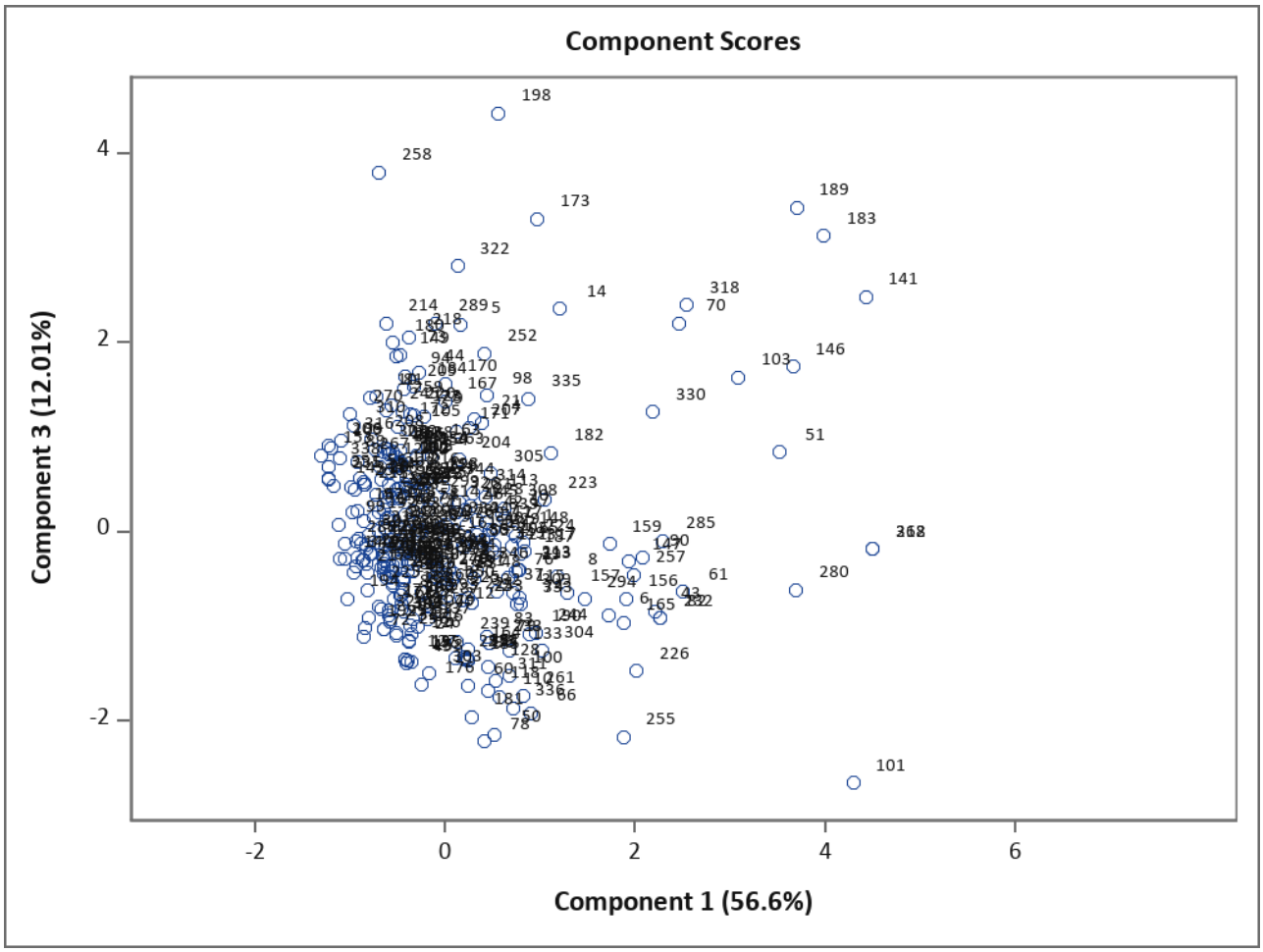

Component Pattern
Component 5 (2.502%)
Component 2 (22.02%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
LogD
PSA
ROT
LogP
MW
NRING
Hdonors
NATOM

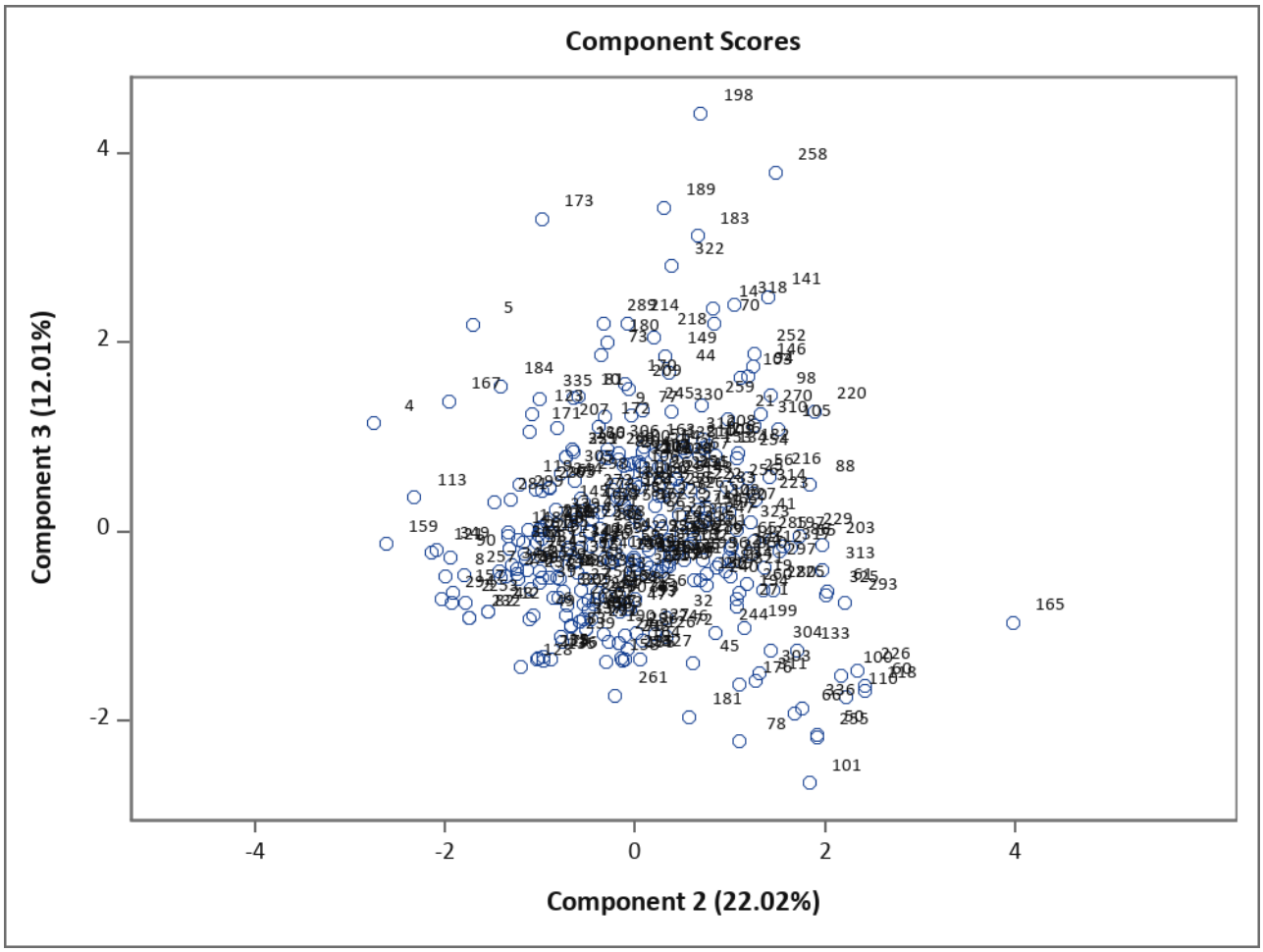

Component Pattern
Component 5 (2.502%)
Component 3 (12.01%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
LogD
PSA
LogP
ROT
NRING
MW
NATOM
Hdonors

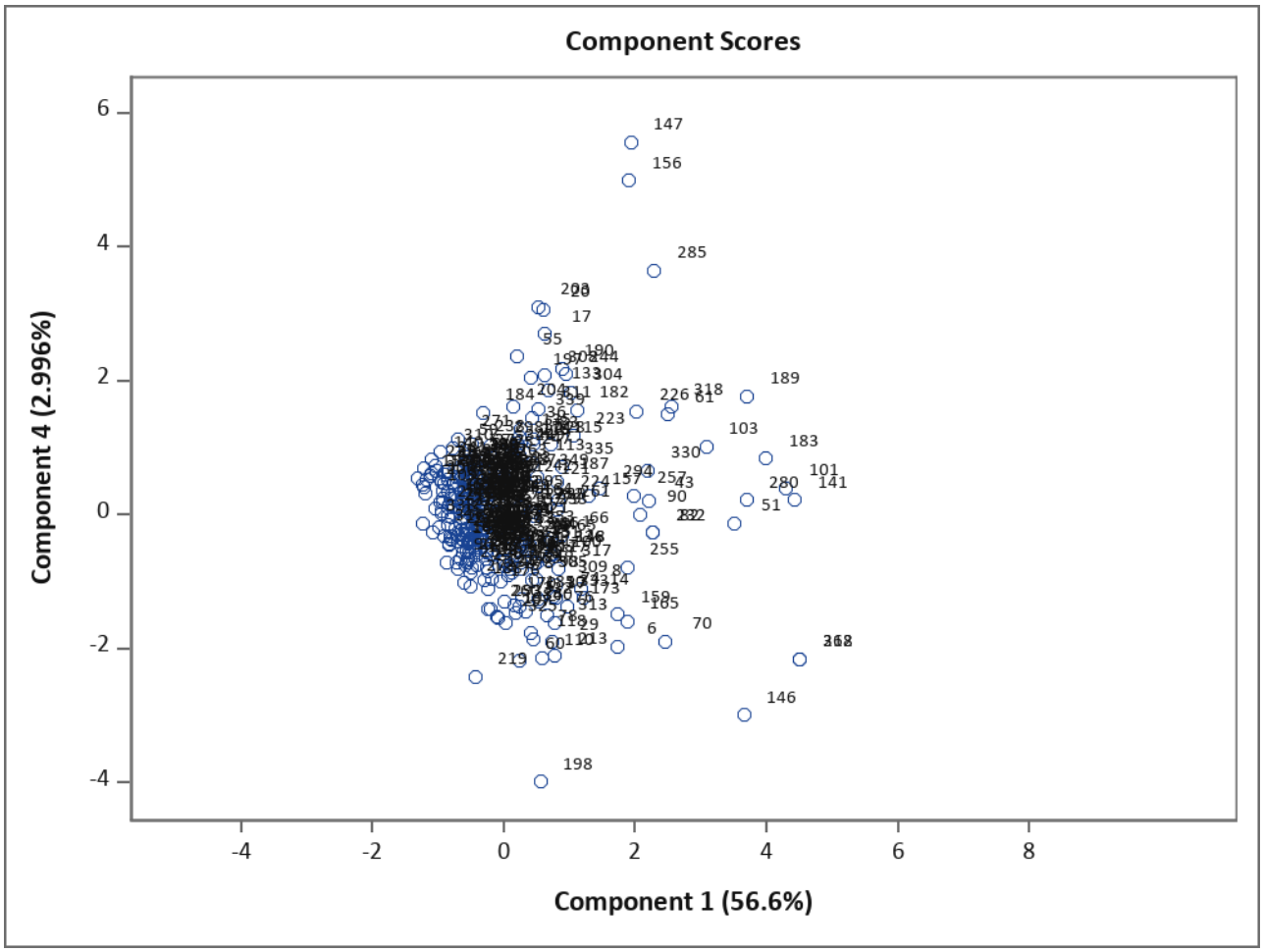

Component Pattern
Component 5 (2.502%)
Component 4 (2.996%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Hacceptors
LogD
PSA
LogP
ROT
MW
NRING
NATOM
Hdonors

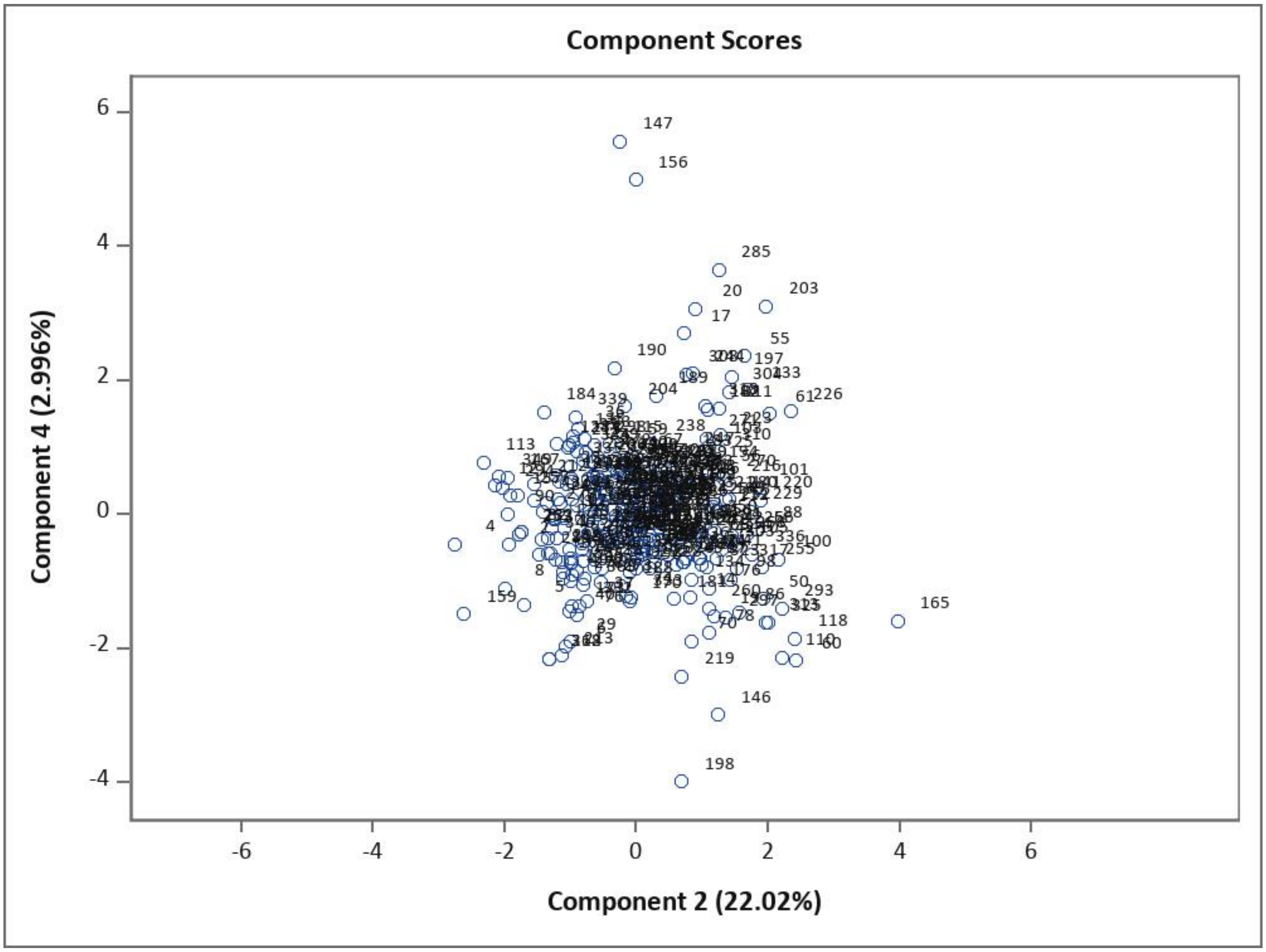
4. Score plots for violators

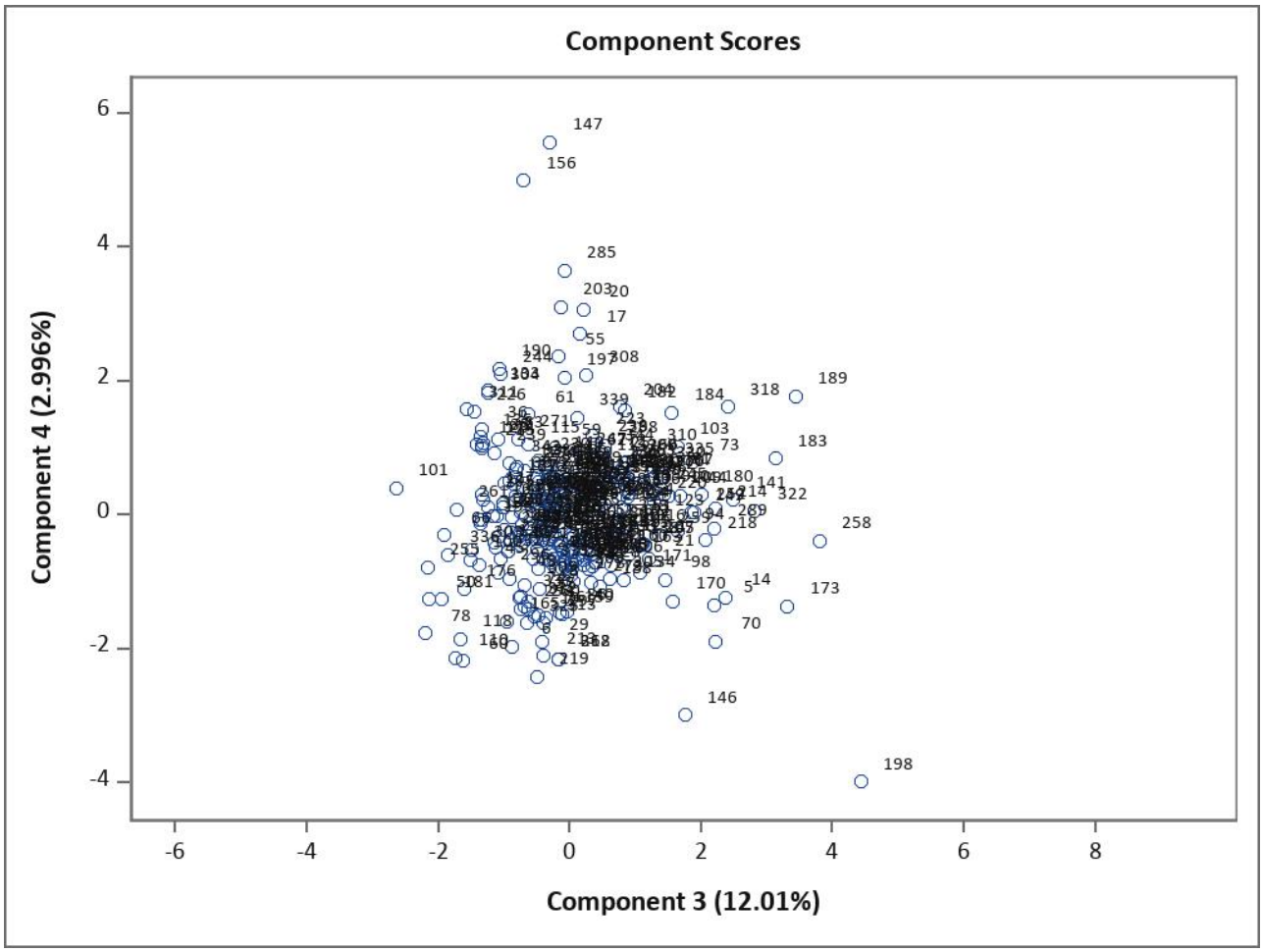


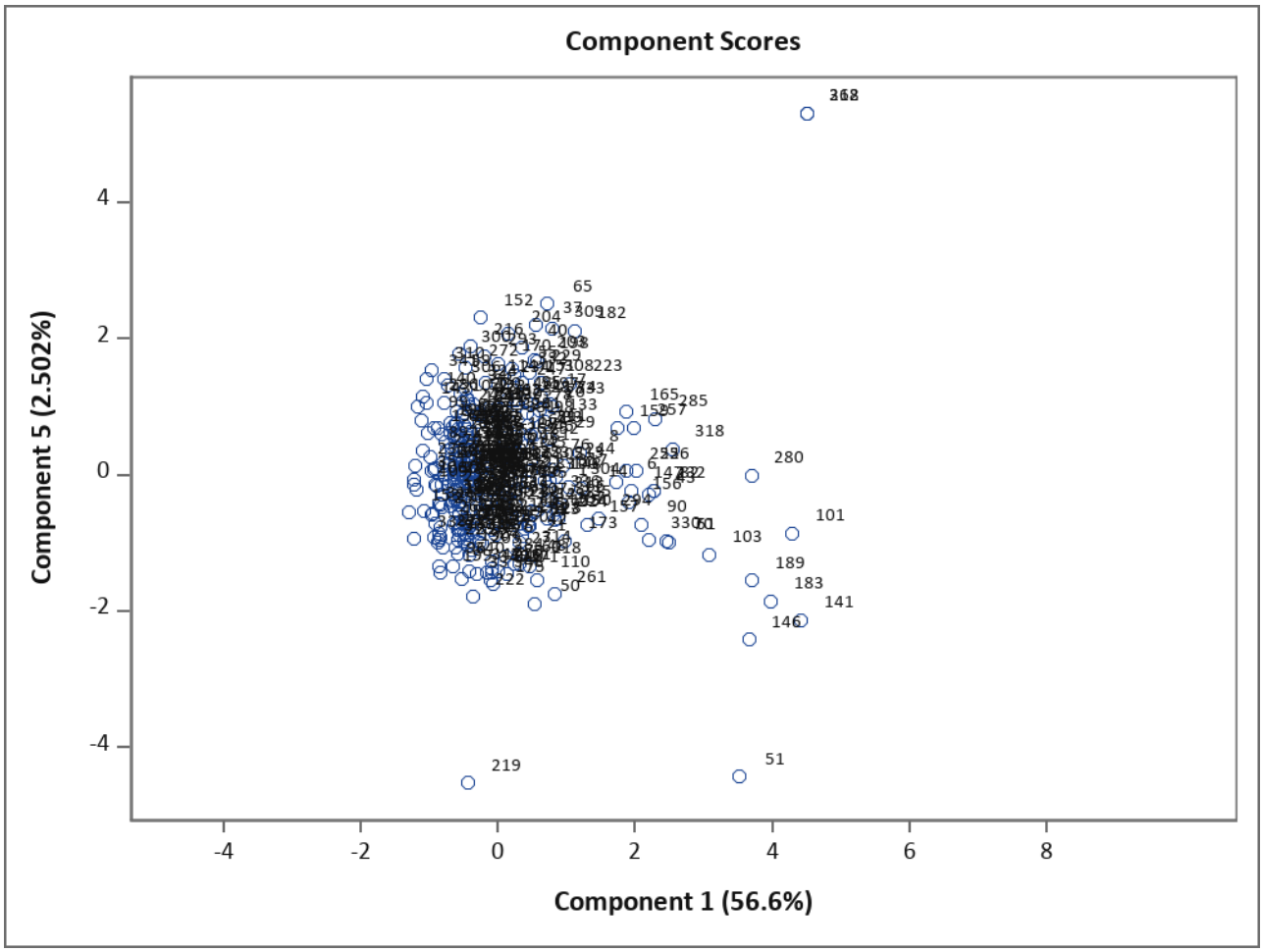

Component Scores
Component 3 (12.01%)
Component 1 (56.6%)
4
2
0
-2
-2
0
2
4
6
198
258
173
189
183
322
141
14
318
70
214
289
5
252
146
103
98
335
330
182
51
305
223
159
285
257
280
61
294
156
165
226
304
255
101

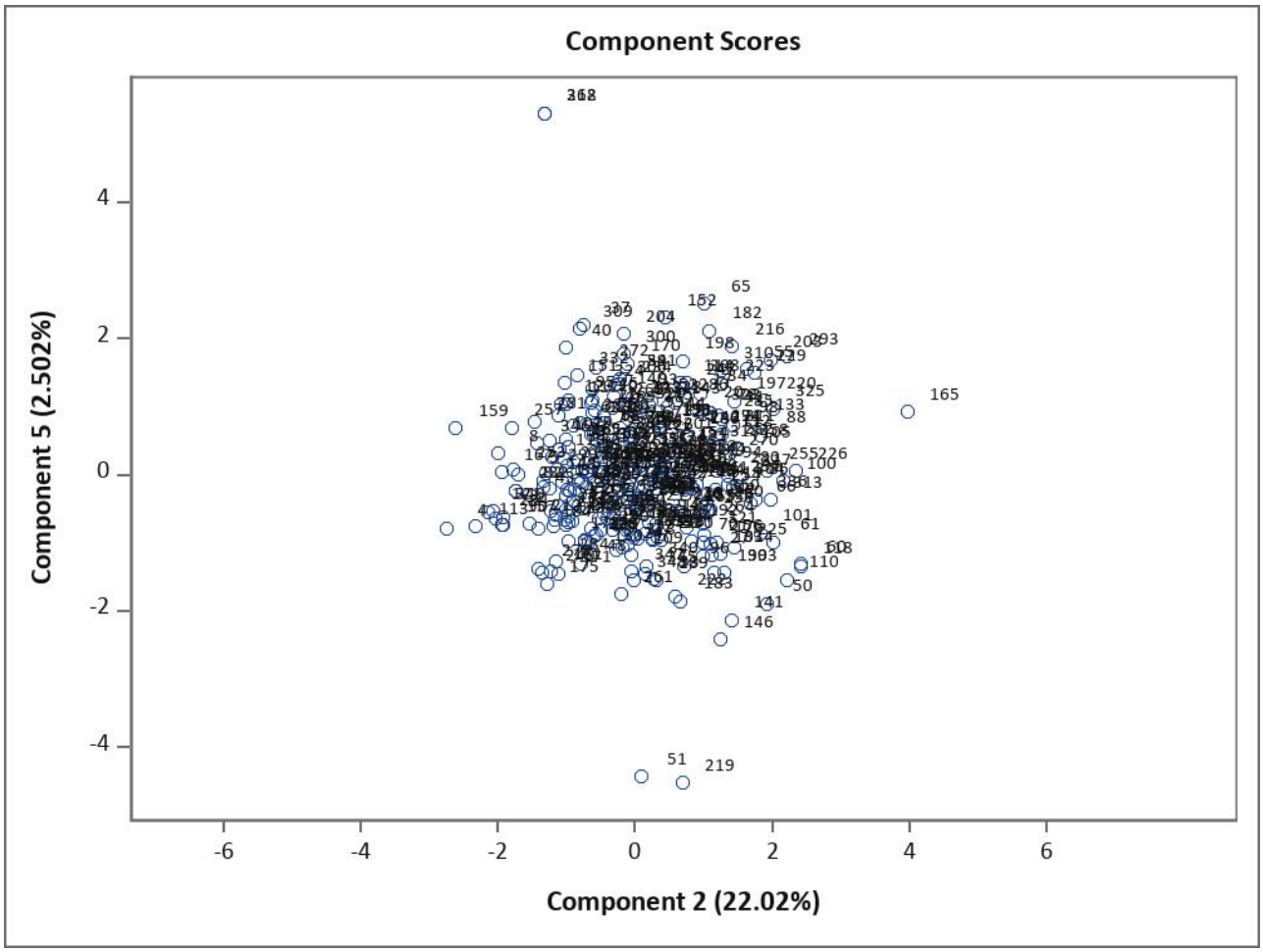

Component Scores
Component 3 (12.01%)
Component 2 (22.02%)
4
2
0
-2
-4
-2
0
2
4

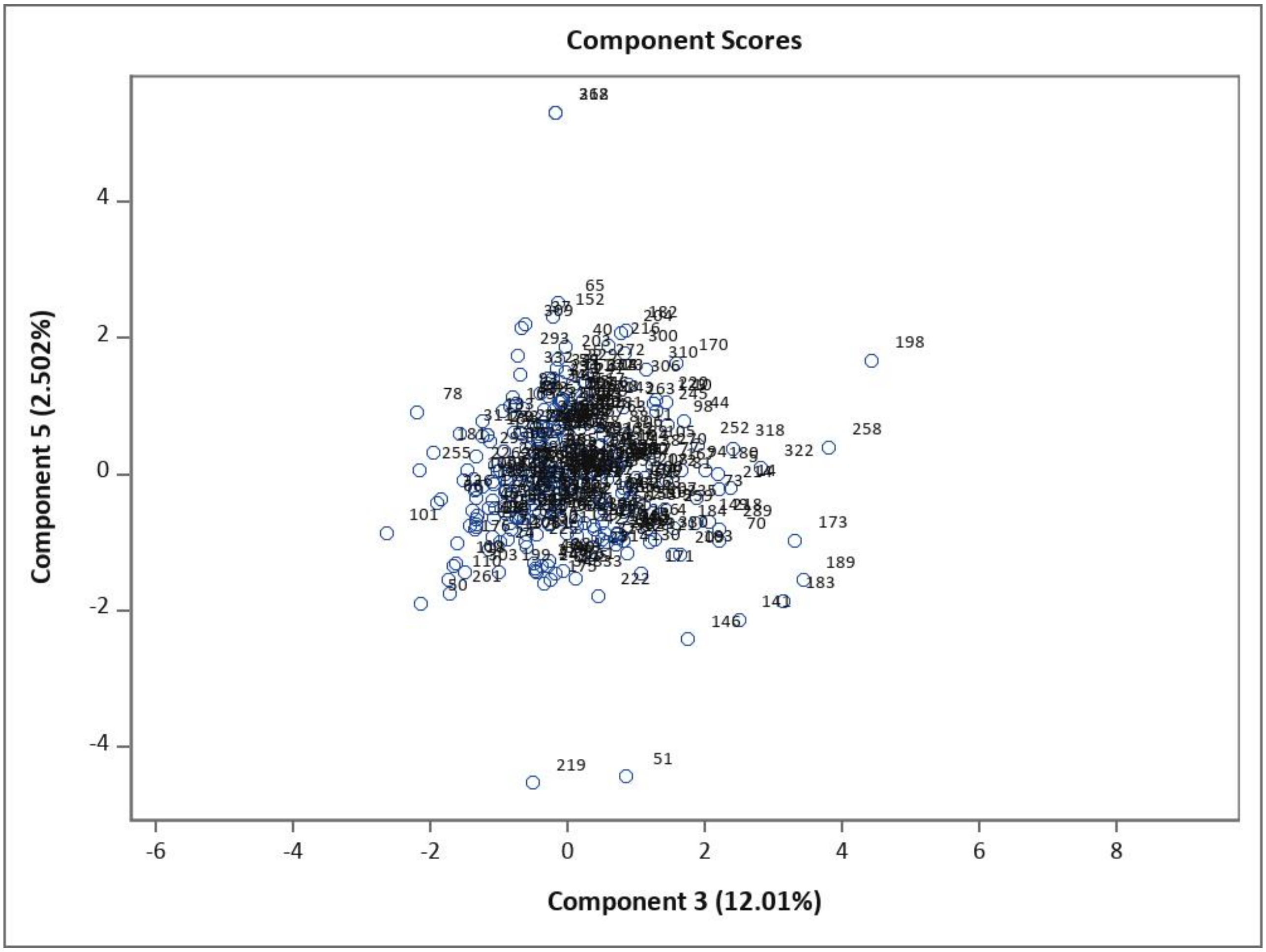

Component Scores
Component 4 (2.996%)
Component 1 (56.6%)
6
4
2
0
-2
-4
-4
-2
0
2
4
6
8
147
156
285
17
189
103
183
330
101
141
280
51
90
255
70
146
198

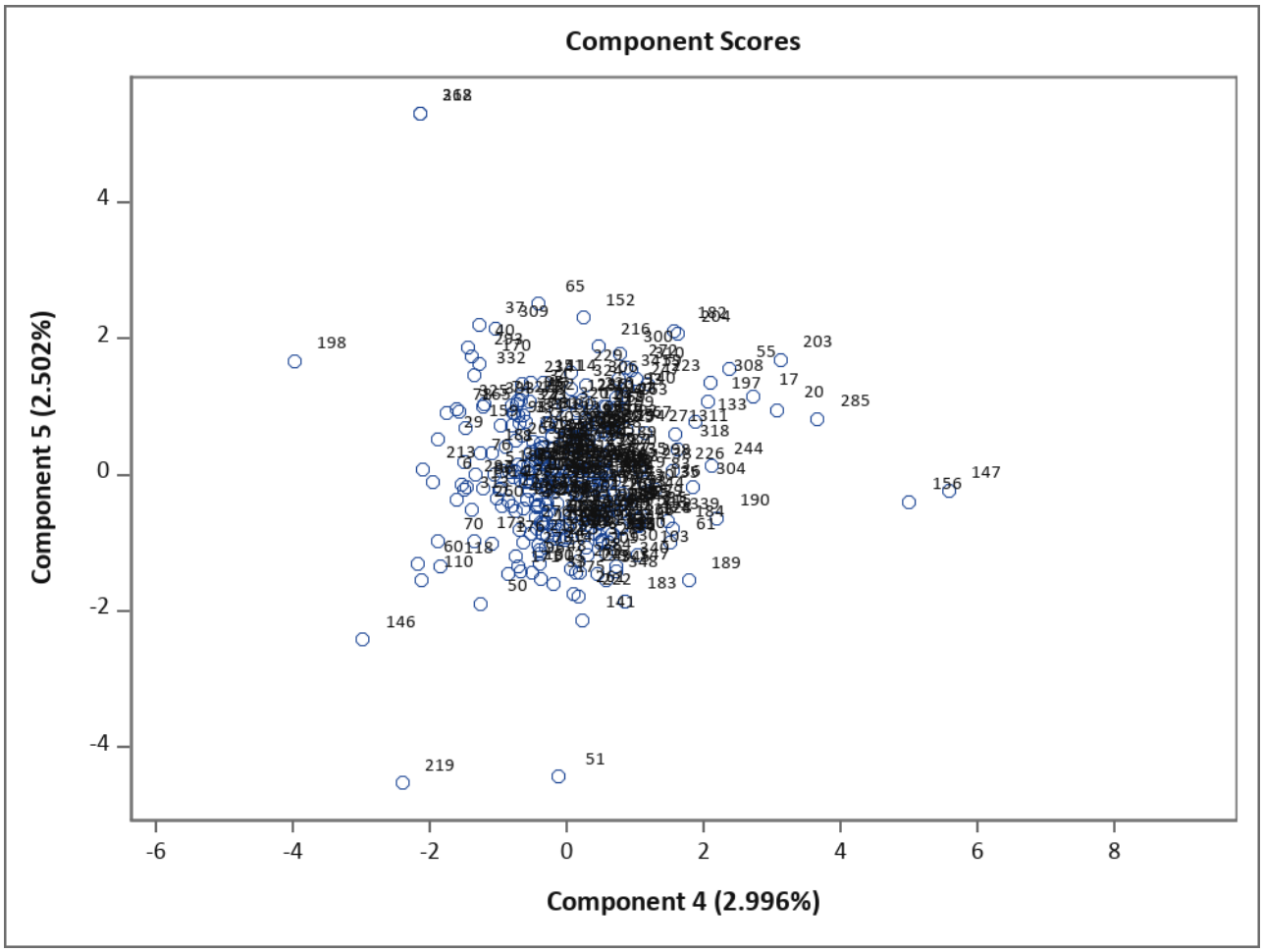

Component Scores
Component 4 (2.996%)
Component 2 (22.02%)
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
147
156
285
203
20
17
55
190
197
226
165
198
146
219
118
159
113
4
8
5
29

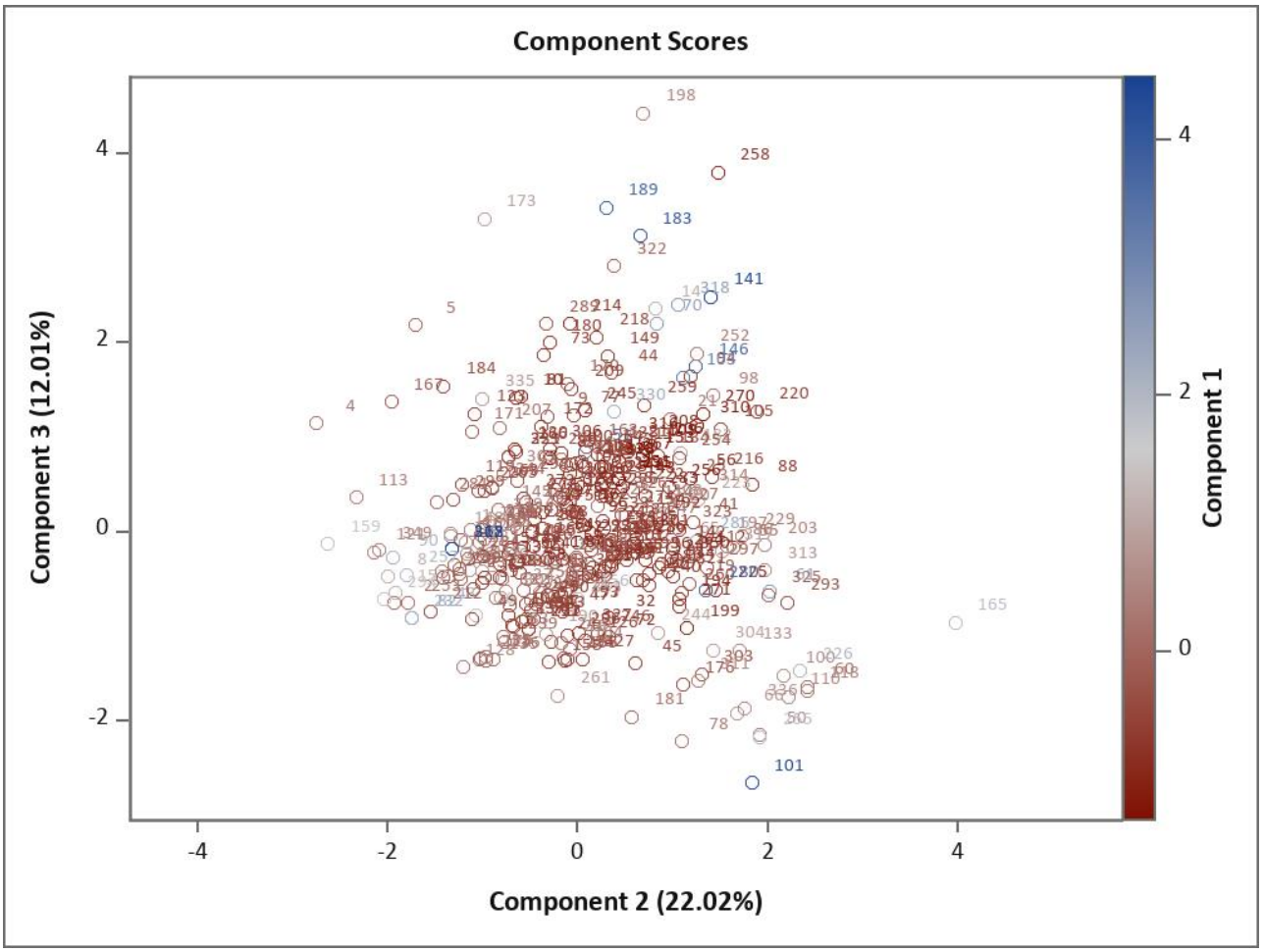

Component Scores
Component 4 (2.996%)
Component 3 (12.01%)
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8
147
156
285
203
20
17
55
190
244
197
308
189
318
184
61
339
103
310
73
183
101
180
141
322
214
269
218
258
98
170
5
14
173
70
146
198


Component Scores
Component 5 (2.502%)
Component 1 (56.6%)
4
2
0
-2
-4
-4
-2
0
2
4
6
8
65
152
182
165
285
318
280
101
103
189
183
141
146
110
261
50
219
51

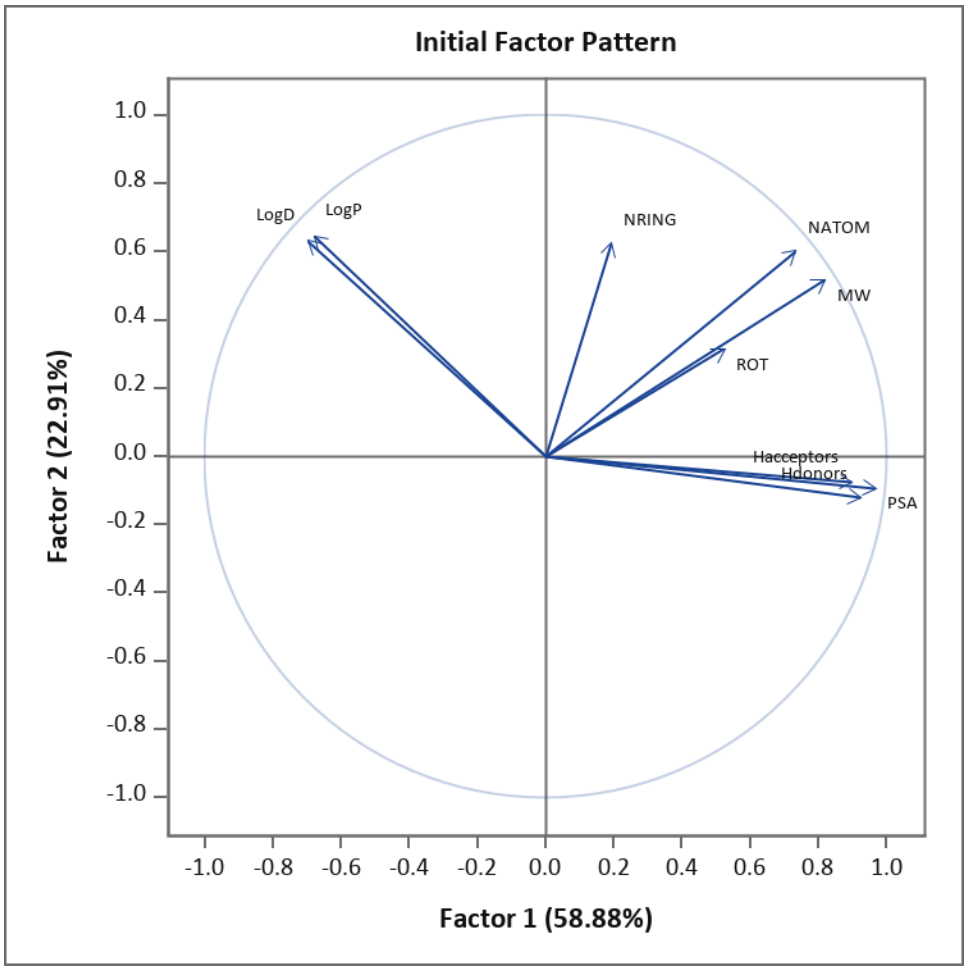

Component Scores
Component 5 (2.502%)
Component 2 (22.02%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
65
182
216
165
159
146
51
219
50
110
61
100
101

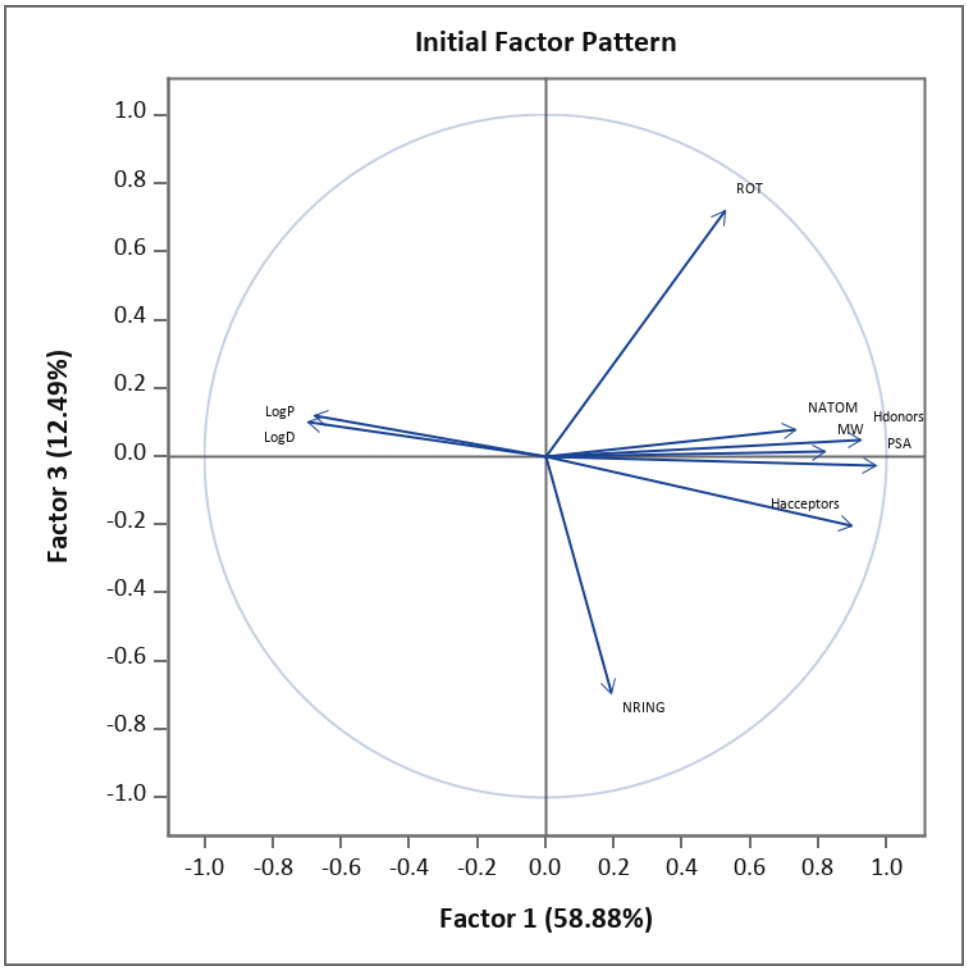

Component Scores
Component 5 (2.502%)
Component 3 (12.01%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8

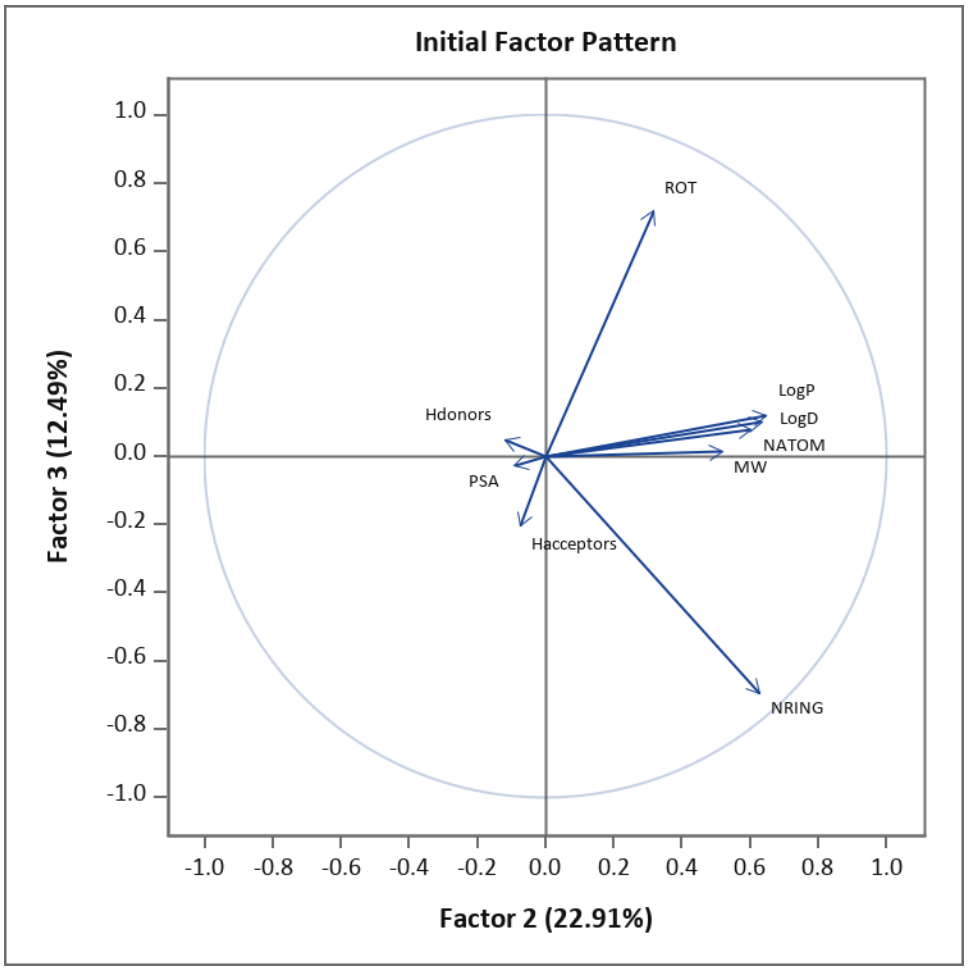

Component Scores
Component 5 (2.502%)
Component 4 (2.996%)
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8
198
65
152
309
216
203
55
308
197
17
20
285
133
318
244
226
304
147
156
190
189
183
146
219
51

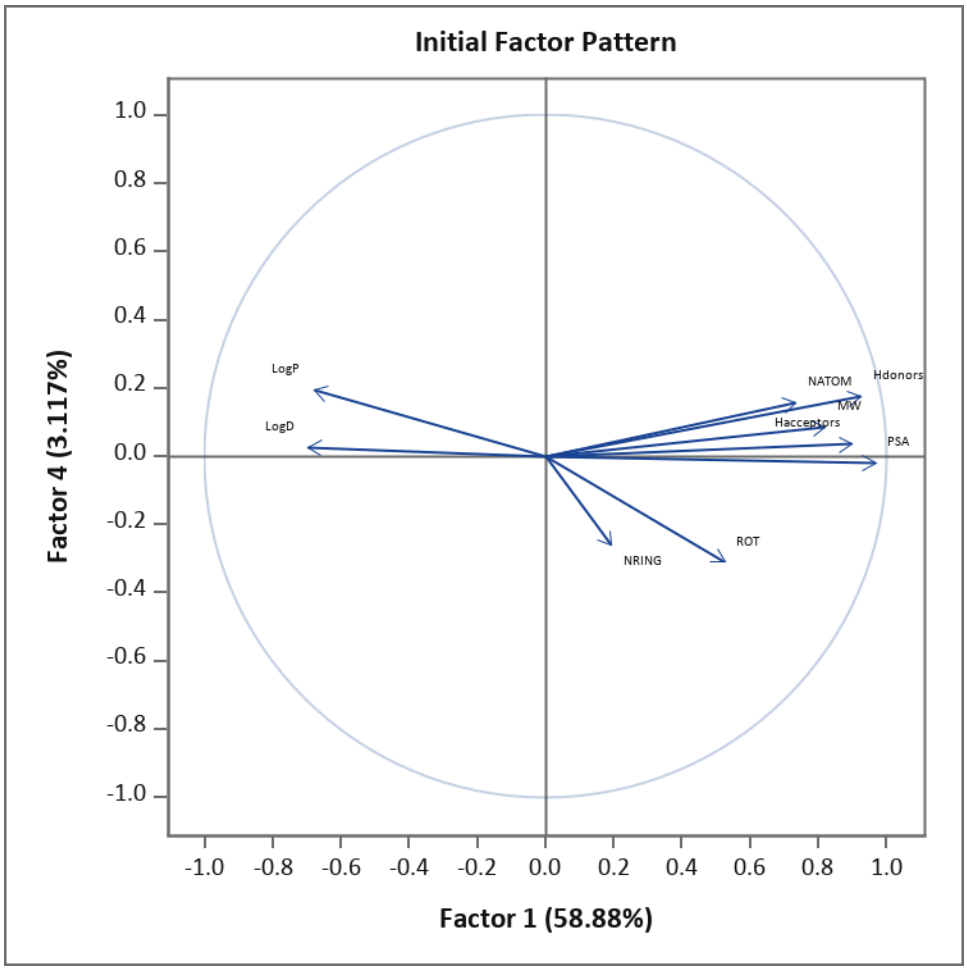

Component Scores
Component 3 (12.01%)
Component 2 (22.02%)
Component 1
4
2
0
-2
-4
-2
0
2
4

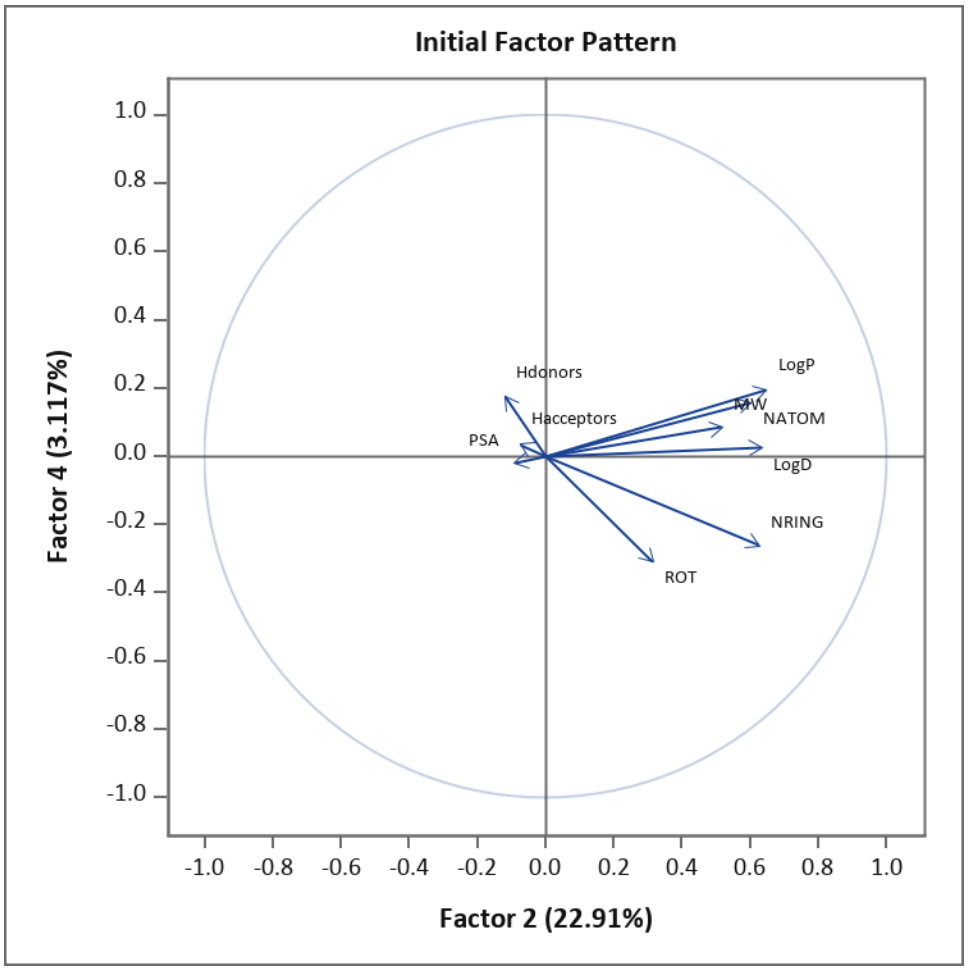
5. Loading plots for violators

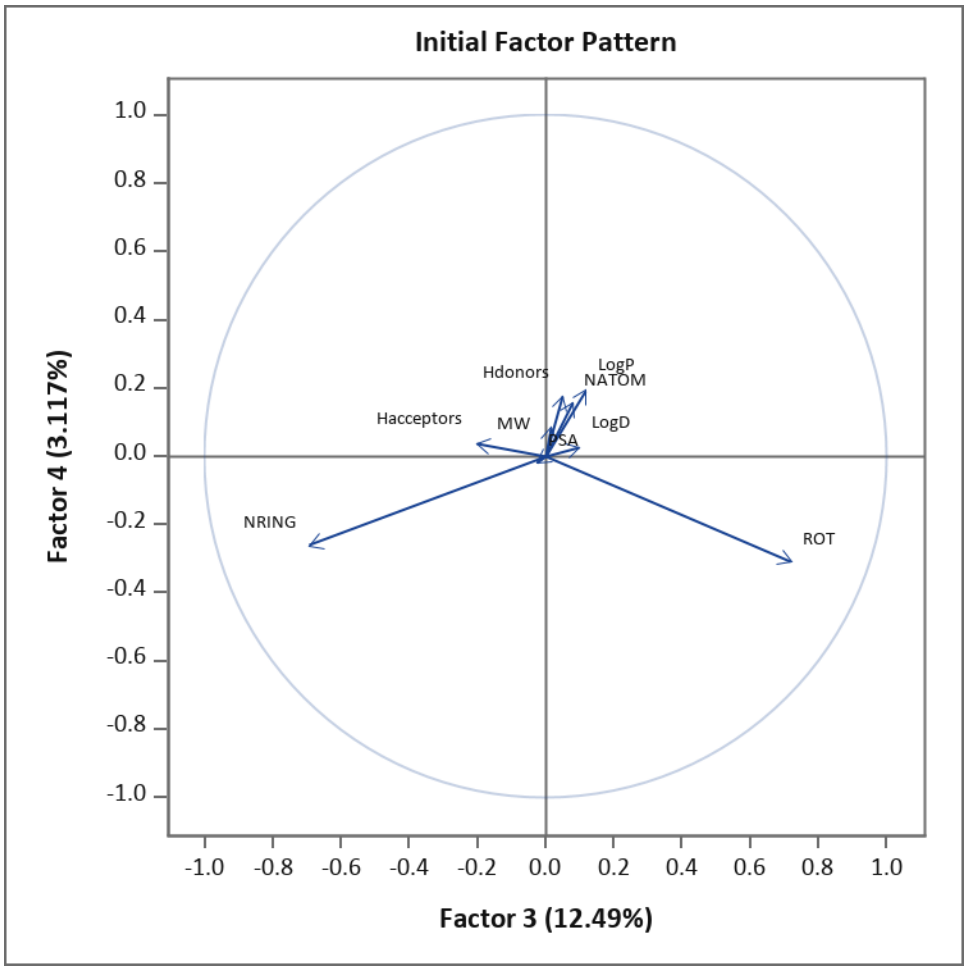


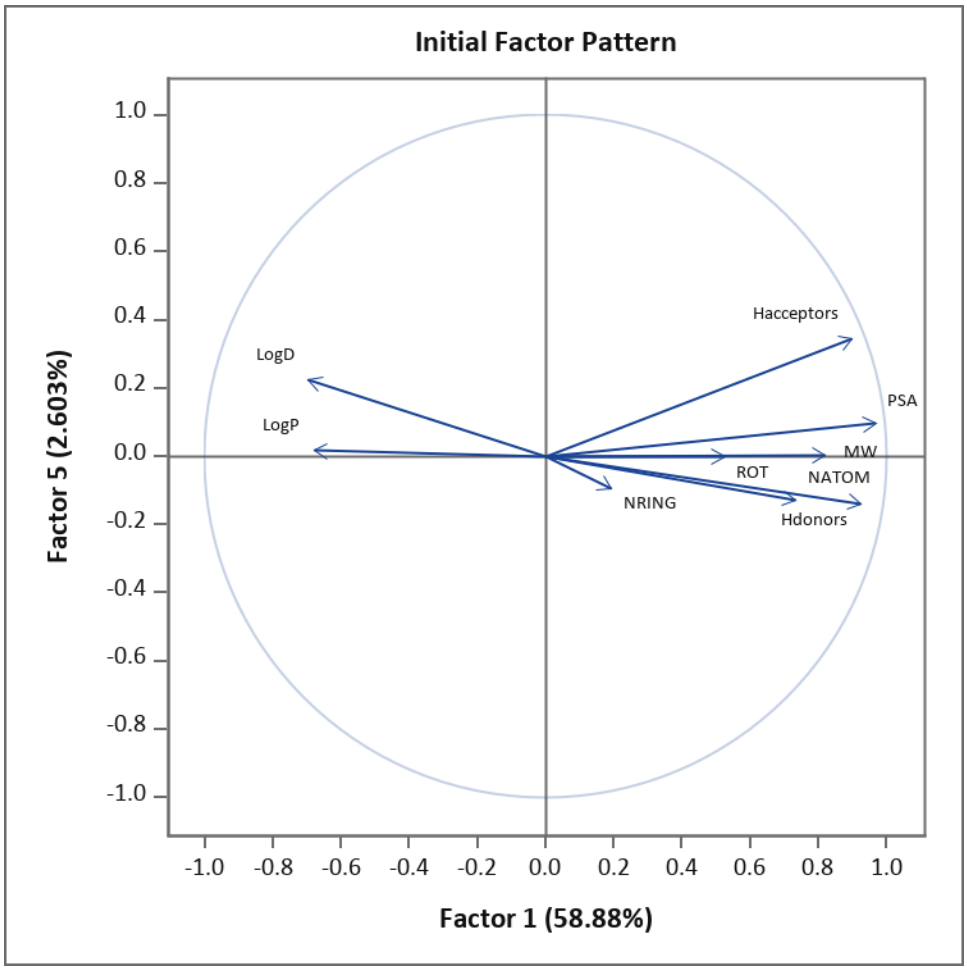

Initial Factor Pattern
Factor 3 (12.49%)
Factor 1 (58.88%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
ROT
LogP
LogD
NATOM
MW
Hdonors
PSA
Hacceptors
NRING

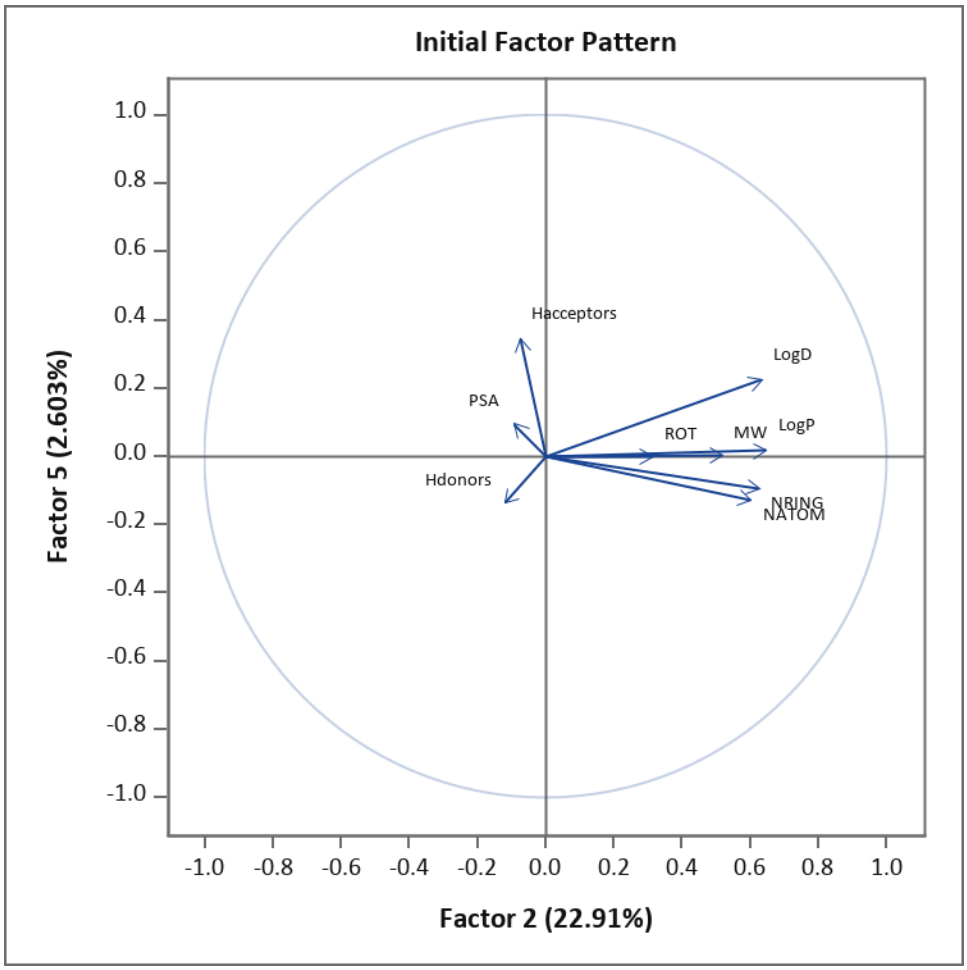

Initial Factor Pattern
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Factor 3 (12.49%)
ROT
LogP
LogD
NATOM
MW
Hdonors
PSA
Hacceptors
NRING
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Factor 2 (22.91%)

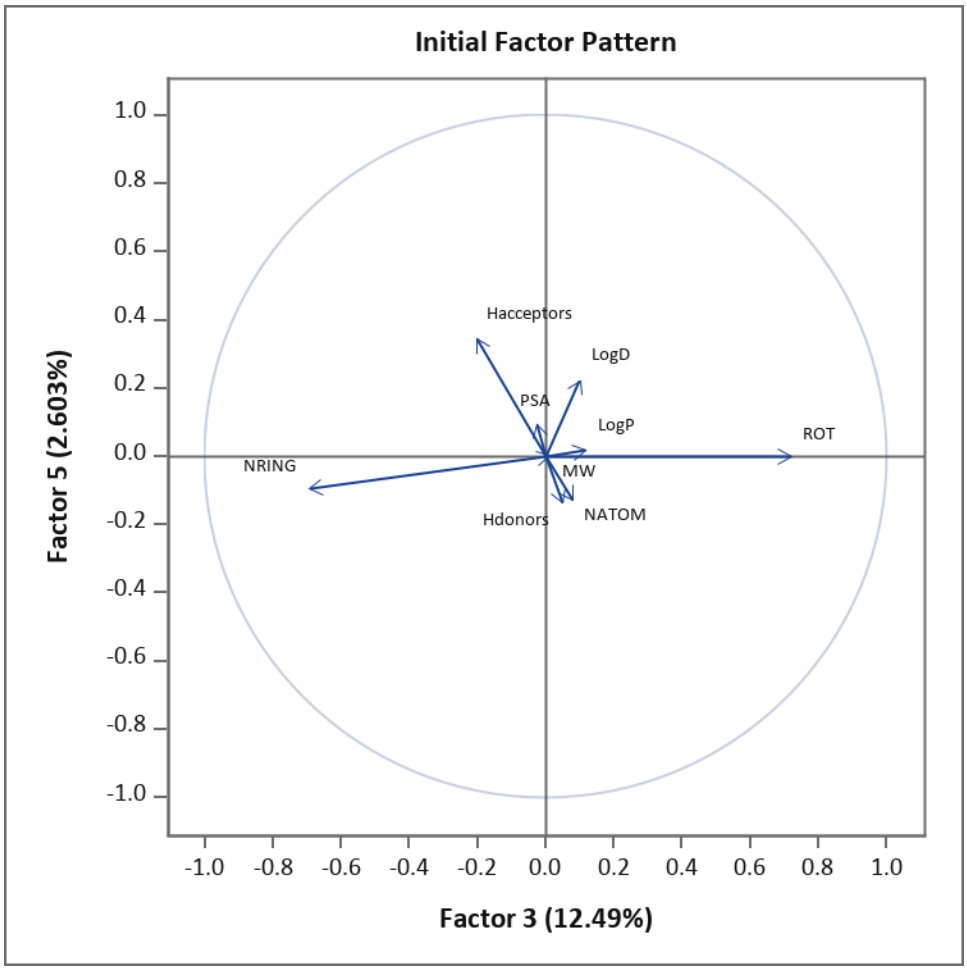

Initial Factor Pattern
Factor 4 (3.117%)
Factor 1 (58.88%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
LogP
LogD
NATOM
Hdonors
MW
Hacceptors
PSA
ROT
NRING

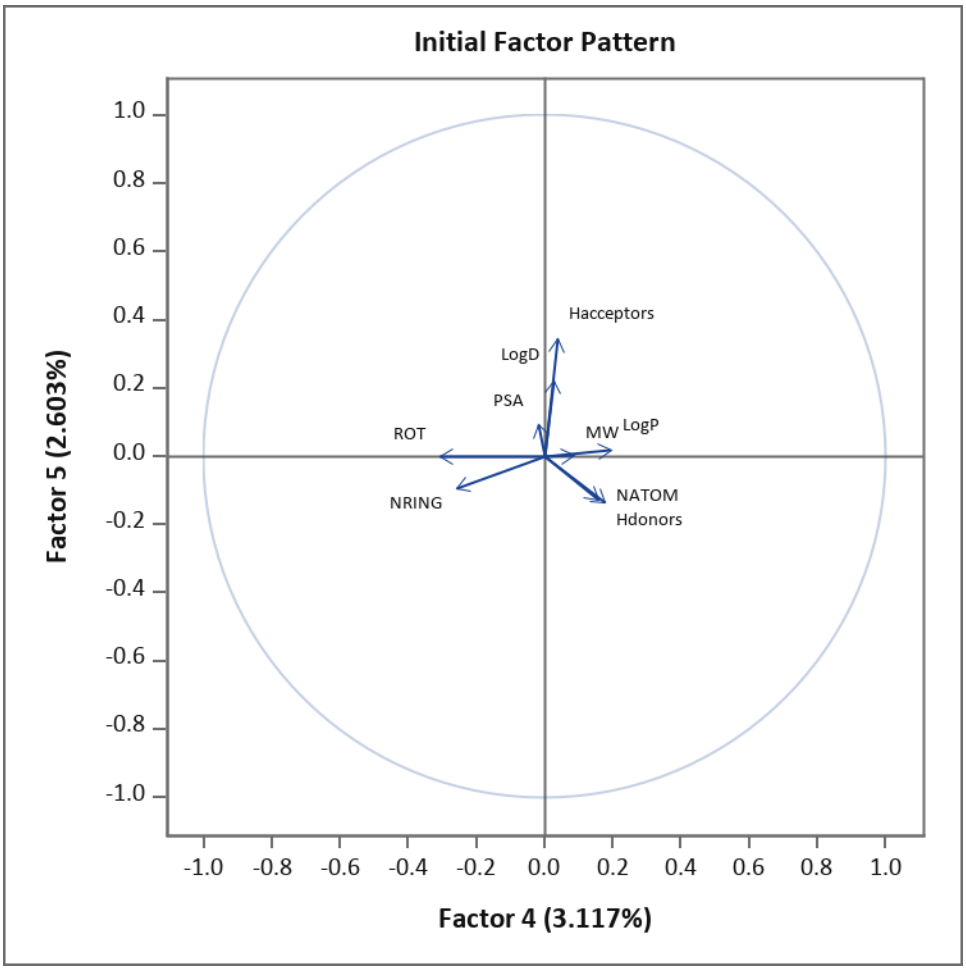

Initial Factor Pattern
Factor 4 (3.117%)
Factor 2 (22.91%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hdonors
Hacceptors
PSA
LogP
MW
NATOM
LogD
NRING
ROT


Initial Factor Pattern
Factor 4 (3.117%)
Factor 3 (12.49%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hdonors
LogP
NATOM
Hacceptors
MW
LogD
PSA
NRING
ROT


Initial Factor Pattern
Factor 5 (2.603%)
Factor 1 (58.88%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hacceptors
LogD
PSA
LogP
MW
ROT
NATOM
NRING
Hdonors


Initial Factor Pattern
Factor 5 (2.603%)
Factor 2 (22.91%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hacceptors
LogD
PSA
ROT
MW
LogP
Hdonors
NRING
NATOM


Initial Factor Pattern
Factor 5 (2.603%)
Factor 3 (12.49%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hacceptors
LogD
PSA
LogP
ROT
NRING
MW
Hdonors
NATOM


Initial Factor Pattern
Factor 5 (2.603%)
Factor 4 (3.117%)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Hacceptors
LogD
PSA
ROT
MW
LogP
NRING
NATOM
Hdonors