



ANALISIS ESTADISTICO EXPLORATORIO DE CAMPAÑA DE PRUEBAS LABORATORIO

OBJETIVO

El presente trabajo consistió en realizar un análisis estadístico exploratorio y de selección/identificación de variables claves del proceso a partir de datos generados en diferentes campañas de pruebas de laboratorio, a fin de encontrar las principales tendencias, relaciones y/o correlaciones entre las variables clave del proceso, que permita identificar las unidades geometalúrgicas que tendrán el mayor impacto en el proceso de planificación de largo plazo de la planta concentradora.

EXPLORACION INICIAL

El resumen estadístico del conjunto de datos original se muestra a continuación:

[1] 60 76

Se reportan 76 variables con 60 registros cada una.

Nombre de Variables

Se cambiaron los nombres de las variables según su naturaleza, las cuales se separaron en tres grupos:

- Entrada (ent),
- Operacionales (op),
- Salida (sal)

Creación de Variable UGG y Selección de Variables

En base a la información del proyecto se usaran algunas variables de entrada para establecer las diferentes UGG's del mineral alimentado, siendo esta última, una variable clave del análisis.

Adicionalmente se seleccionaron todas las variables que describen las propiedades químicas, mineralógicas y metalúrgicas del mineral alimentado a las pruebas, así como las principales variables de respuesta (recuperaciones metalúrgicas, leyes y constantes cinéticas), también se seleccionaron variables operacionales relativas a la adición de reactivos. El resto de las variables, incluidas aquellas que reportaron valores constantes, se dejaron fuera del conjunto de datos.

De esta forma, la estructura del conjunto de datos seleccionado, se muestra a continuación:

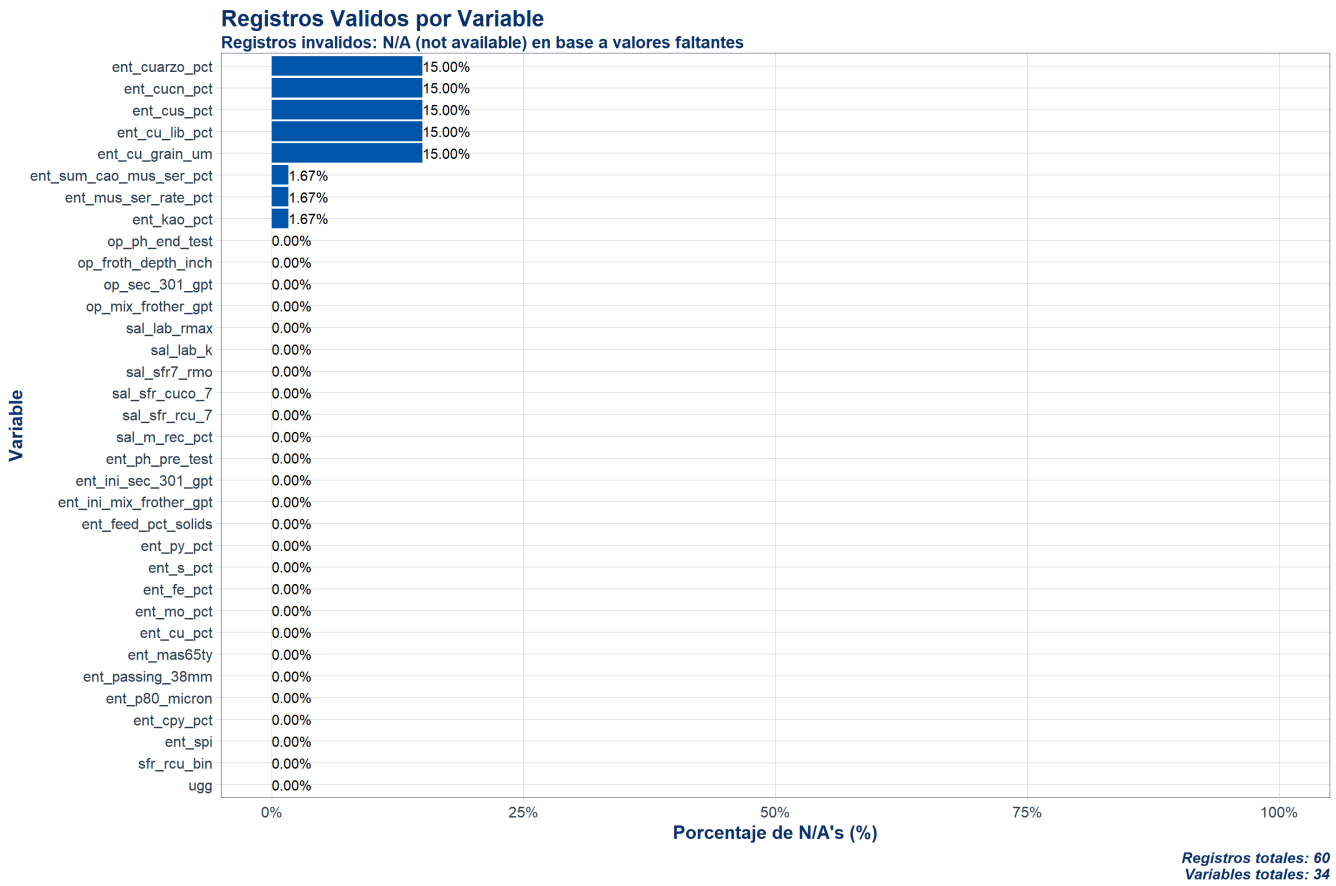
[1]	"ugg"	"sfr_rcu_bin"
[3]	"ent_kao_pct"	"ent_mus_ser_rate_pct"
[5]	"ent_sum_cao_mus_ser_pct"	"ent_cu_grain_um"
[7]	"ent_cu_lib_pct"	"ent_spi"
[9]	"ent_cpy_pct"	"ent_p80_micron"
[11]	"ent_passing_38mm"	"ent_mas65ty"
[13]	"ent_cu_pct"	"ent_mo_pct"
[15]	"ent_fe_pct"	"ent_s_pct"
[17]	"ent_py_pct"	"ent_cus_pct"
[19]	"ent_cucn_pct"	"ent_cuarzo_pct"
[21]	"ent_feed_pct_solids"	"ent_ini_mix_frother_gpt"
[23]	"ent_ini_sec_301_gpt"	"ent_ph_pre_test"
[25]	"sal_m_rec_pct"	"sal_sfr_rcu_7"
[27]	"sal_sfr_cuco_7"	"sal_sfr7_rmo"
[29]	"sal_lab_k"	"sal_lab_rmax"
[31]	"op_mix_frother_gpt"	"op_sec_301_gpt"
[33]	"op_froth_depth_inch"	"op_ph_end_test"

ESTRUCTURA DE REGISTROS

En esta sección se muestra la estructura del conjunto de datos.

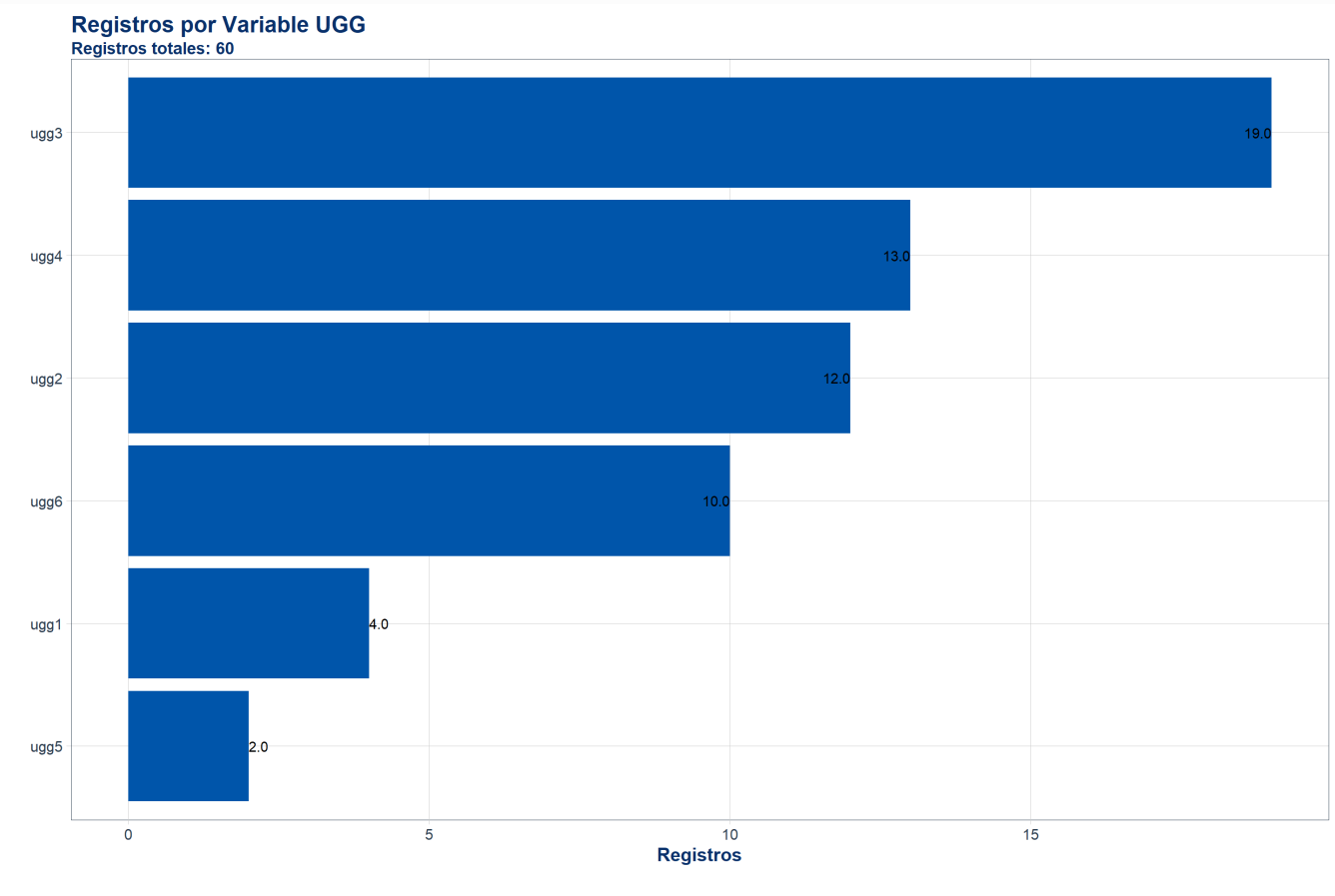
Registros Válidos

A continuación se muestra la proporción de registros válidos por variable:



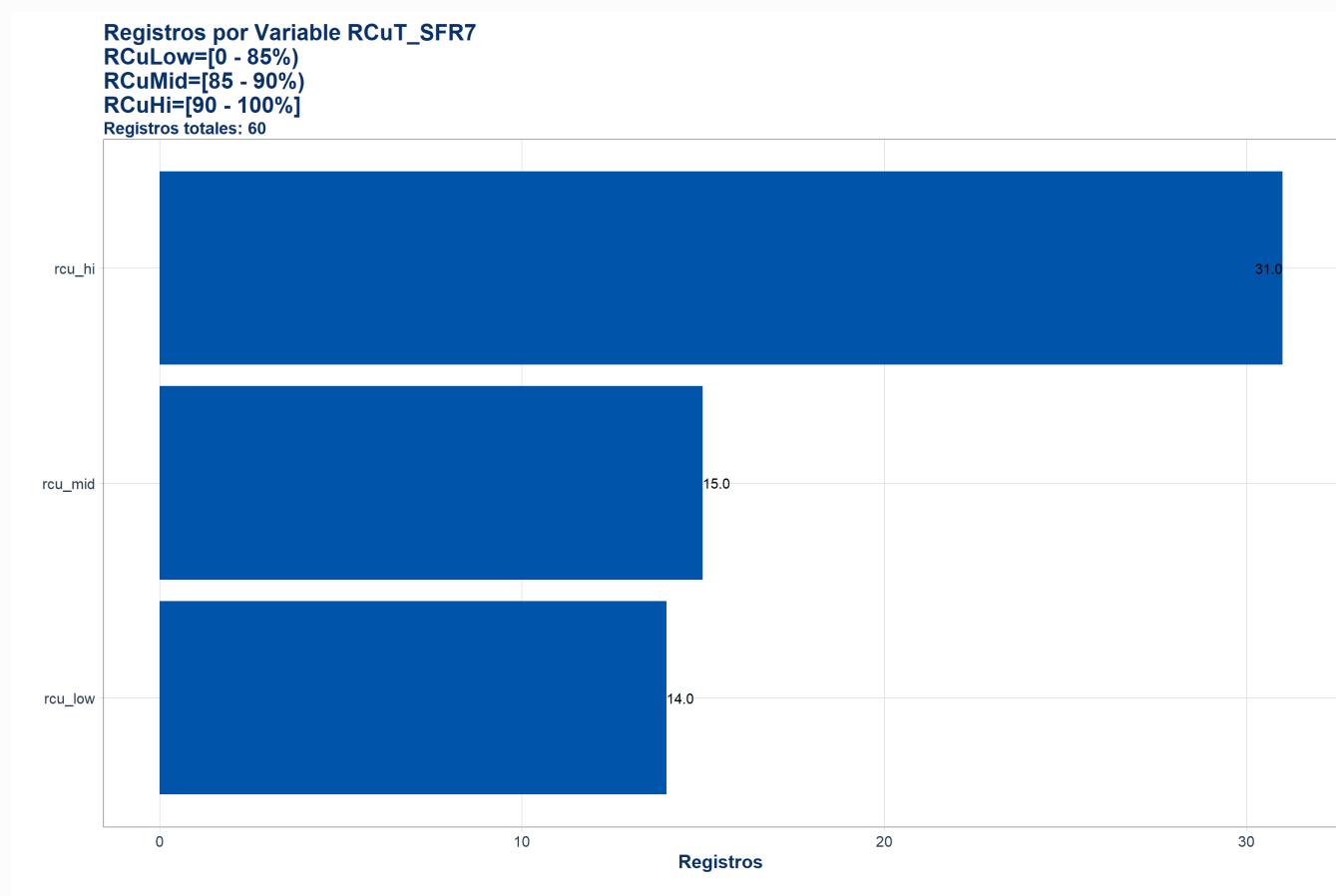
Registros por UGG

A continuación se muestra la proporción de registros por UGG:



Registros por Nivel de RCuT_SFR7

La recuperación de cobre en la celda SFR7 se discretizó en tres niveles a fin de poder visualizar tendencias operacionales.



ESTADISTICOS DE BASE

Los estadísticos considerados se muestran a continuación:

- n: número total de registros.
- na: número de registros inválidos.
- complete: número de registros válidos.
- min: mínimo.
- q1: primer cuartil, 25% de los datos son menores o iguales a este valor.
- avg: promedio aritmético.
- med: mediana, 50% de los datos son menores o iguales a este valor.
- sd: desviación estándar, nivel de dispersión.
- iqr: rango intercuartílico, medida de dispersión del 50% medio de los datos (resistente a “outliers”).
- q3: tercer cuartil, 75% de los datos son menores o iguales a este valor.
- max: máximo.

Estadísticos por Variable

variable	n	na	n_comp	min	q1	median	avg	iqr
ent_cu_grain_um	60	9.00	51.00	11.00	28.50	33.00	34.33	10.50
ent_cu_lib_pct	60	9.00	51.00	0.03	0.16	0.28	0.28	0.20
ent_cus_pct	60	9.00	51.00	0.01	0.03	0.05	0.06	0.03
ent_cucn_pct	60	9.00	51.00	0.08	0.19	0.36	0.45	0.57
ent_cuarzo_pct	60	9.00	51.00	15.65	21.66	30.49	29.65	12.33
ent_kao_pct	60	1.00	59.00	0.02	0.07	0.13	0.13	0.10
ent_mus_ser_rate_pct	60	1.00	59.00	0.06	0.14	0.19	0.21	0.15
ent_sum_cao_mus_ser_pct	60	1.00	59.00	0.13	0.25	0.34	0.34	0.19
ent_spi	60	0.00	60.00	23.64	50.75	60.76	65.64	23.55
ent_cpy_pct	60	0.00	60.00	0.00	0.12	0.60	0.52	0.74

Estadísticos por Variable (en base a n_comp)

Showing 1 to 10 of 32 entries

Estadísticos por UGG y por Variable

variable	ugg	n	na	n_comp	min	q1	median	avg
ent_kao_pct	ugg1	4.00	0.00	4.00	0.03	0.10	0.12	0.11
ent_kao_pct	ugg2	12.00	0.00	12.00	0.05	0.11	0.14	0.14
ent_kao_pct	ugg3	19.00	0.00	19.00	0.02	0.07	0.14	0.12
ent_kao_pct	ugg4	13.00	0.00	13.00	0.05	0.11	0.15	0.16

ent_kao_pct	ugg5	2.00	0.00	2.00	0.07	0.09	0.10	0.10	(
ent_kao_pct	ugg6	10.00	1.00	9.00	0.03	0.04	0.06	0.08	(
ent_mus_ser_rate_pct	ugg1	4.00	0.00	4.00	0.14	0.15	0.19	0.23	(
ent_mus_ser_rate_pct	ugg2	12.00	0.00	12.00	0.16	0.24	0.31	0.29	(
ent_mus_ser_rate_pct	ugg3	19.00	0.00	19.00	0.06	0.12	0.16	0.17	(
ent_mus_ser_rate_pct	ugg4	13.00	0.00	13.00	0.11	0.17	0.28	0.25	(

Estadísticos por UGG y por Variable(en base a n_comp)

Showing 1 to 10 of 192 entries

Previous 1 2 3 4 5 ... 20 Next

IMPUTACION DE DATOS FALTANTES

Las variables se imputaron mediante la técnica de vecinos cercanos (KNN).

Estadísticos de principales variables sin imputación de datos faltantes:

ent_cu_grain_um	ent_cu_lib_pct	ent_cus_pct	ent_cucn_pct
Min. :11.00	Min. :0.0300	Min. :0.01396	Min. :0.07789
1st Qu.:28.50	1st Qu.:0.1600	1st Qu.:0.03264	1st Qu.:0.19113
Median :33.00	Median :0.2800	Median :0.04762	Median :0.36152
Mean :34.33	Mean :0.2786	Mean :0.06348	Mean :0.45424
3rd Qu.:39.00	3rd Qu.:0.3600	3rd Qu.:0.06595	3rd Qu.:0.76501
Max. :61.00	Max. :0.7400	Max. :0.33213	Max. :1.00226
NA's :9	NA's :9	NA's :9	NA's :9
ent_cuarzo_pct	ent_kao_pct	ent_mus_ser_rate_pct	ent_sum_cao_mus_ser_pct
Min. :15.65	Min. :0.01929	Min. :0.05599	Min. :0.1346
1st Qu.:21.66	1st Qu.:0.06605	1st Qu.:0.14456	1st Qu.:0.2471
Median :30.49	Median :0.13253	Median :0.19488	Median :0.3416
Mean :29.65	Mean :0.12776	Mean :0.21390	Mean :0.3417
3rd Qu.:33.99	3rd Qu.:0.16138	3rd Qu.:0.29208	3rd Qu.:0.4348
Max. :51.73	Max. :0.33109	Max. :0.43485	Max. :0.5668
NA's :9	NA's :1	NA's :1	NA's :1

Estadísticos de principales variables con datos imputados via KNN:

ent_cu_grain_um	ent_cu_lib_pct	ent_cus_pct	ent_cucn_pct
Min. :11.00	Min. :0.0300	Min. :0.01396	Min. :0.07789
1st Qu.:29.75	1st Qu.:0.1700	1st Qu.:0.03621	1st Qu.:0.21711
Median :34.00	Median :0.2800	Median :0.05026	Median :0.39915

Mean :34.46	Mean :0.2790	Mean :0.06204	Mean :0.46354
3rd Qu.:39.00	3rd Qu.:0.3515	3rd Qu.:0.06445	3rd Qu.:0.70643
Max. :61.00	Max. :0.7400	Max. :0.33213	Max. :1.00226
ent_cuarzo_pct	ent_kao_pct	ent_mus_ser_rate_pct	ent_sum_cao_mus_ser_pct
Min. :15.65	Min. :0.01929	Min. :0.05599	Min. :0.1346
1st Qu.:23.06	1st Qu.:0.06637	1st Qu.:0.14515	1st Qu.:0.2475
Median :30.75	Median :0.13083	Median :0.19520	Median :0.3413
Mean :30.01	Mean :0.12778	Mean :0.21359	Mean :0.3414
3rd Qu.:34.31	3rd Qu.:0.16038	3rd Qu.:0.29046	3rd Qu.:0.4289
Max. :51.73	Max. :0.33109	Max. :0.43485	Max. :0.5668

Gráfica de Imputación de Variables

0=registro completo, 1=registro imputado

■ 0 ■ 1



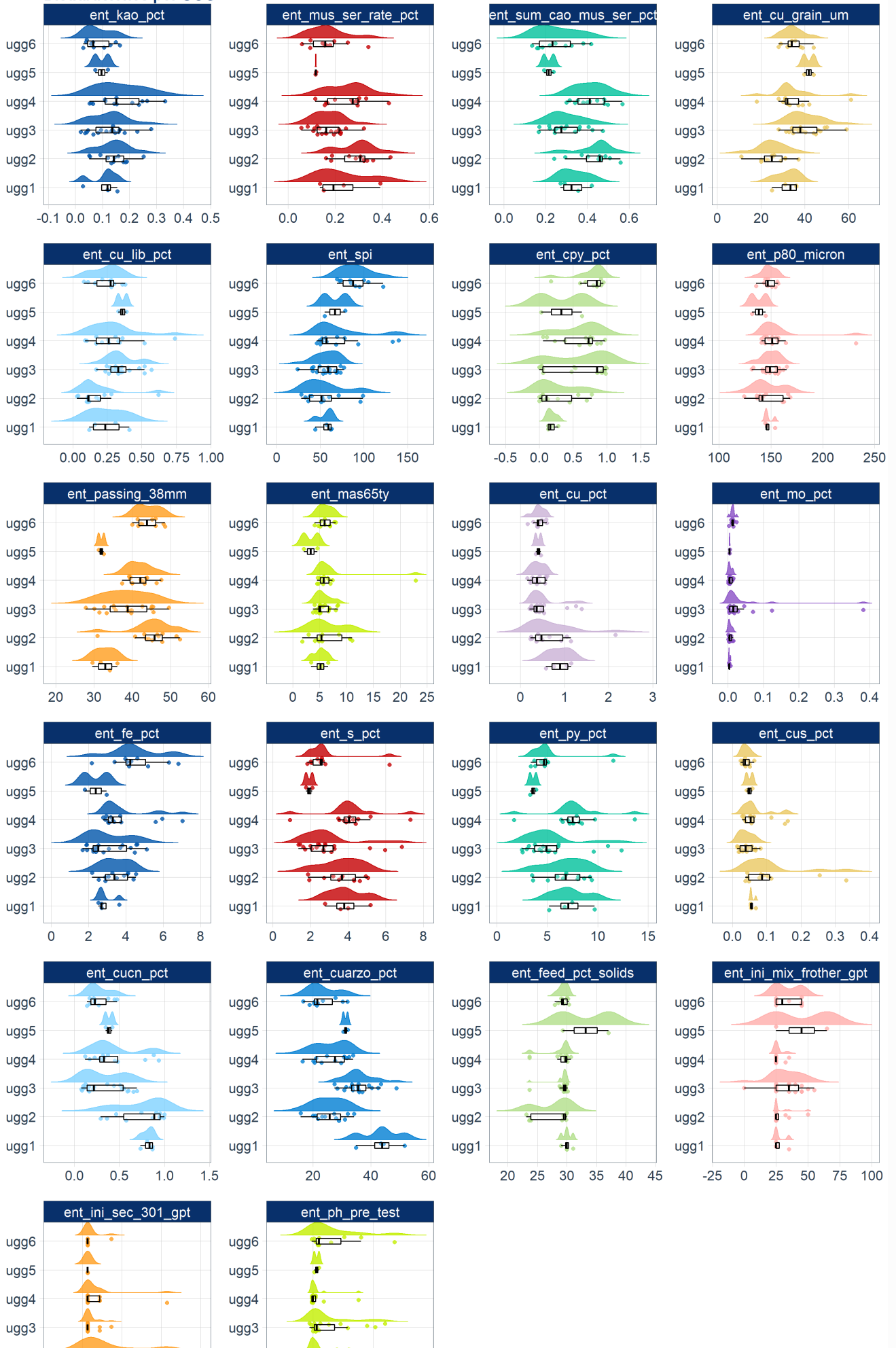
EXPLORACION GRAFICA EXPLORATORIA

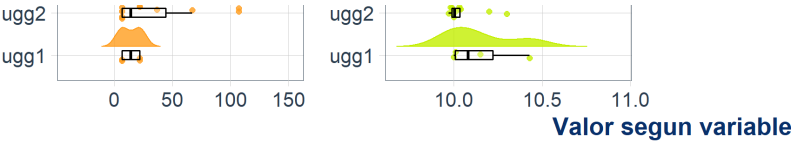
A continuación se muestran gráficas de las diferentes variables disponibles según categoría (ugg y RCuT_SFR7).

Variables de Entrada

Variables de Entrada

Distribuciones por UGG





Variables de Entrada

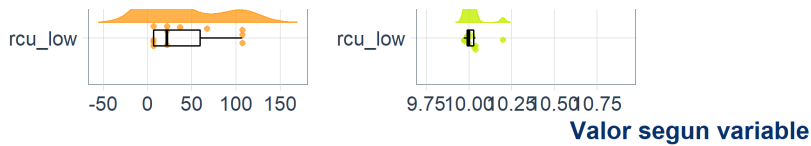
Distribuciones por RCuT

RCuLow=[0 - 85%)

RCuMid=[85 - 90%)

RCuHi=[90 - 100%]

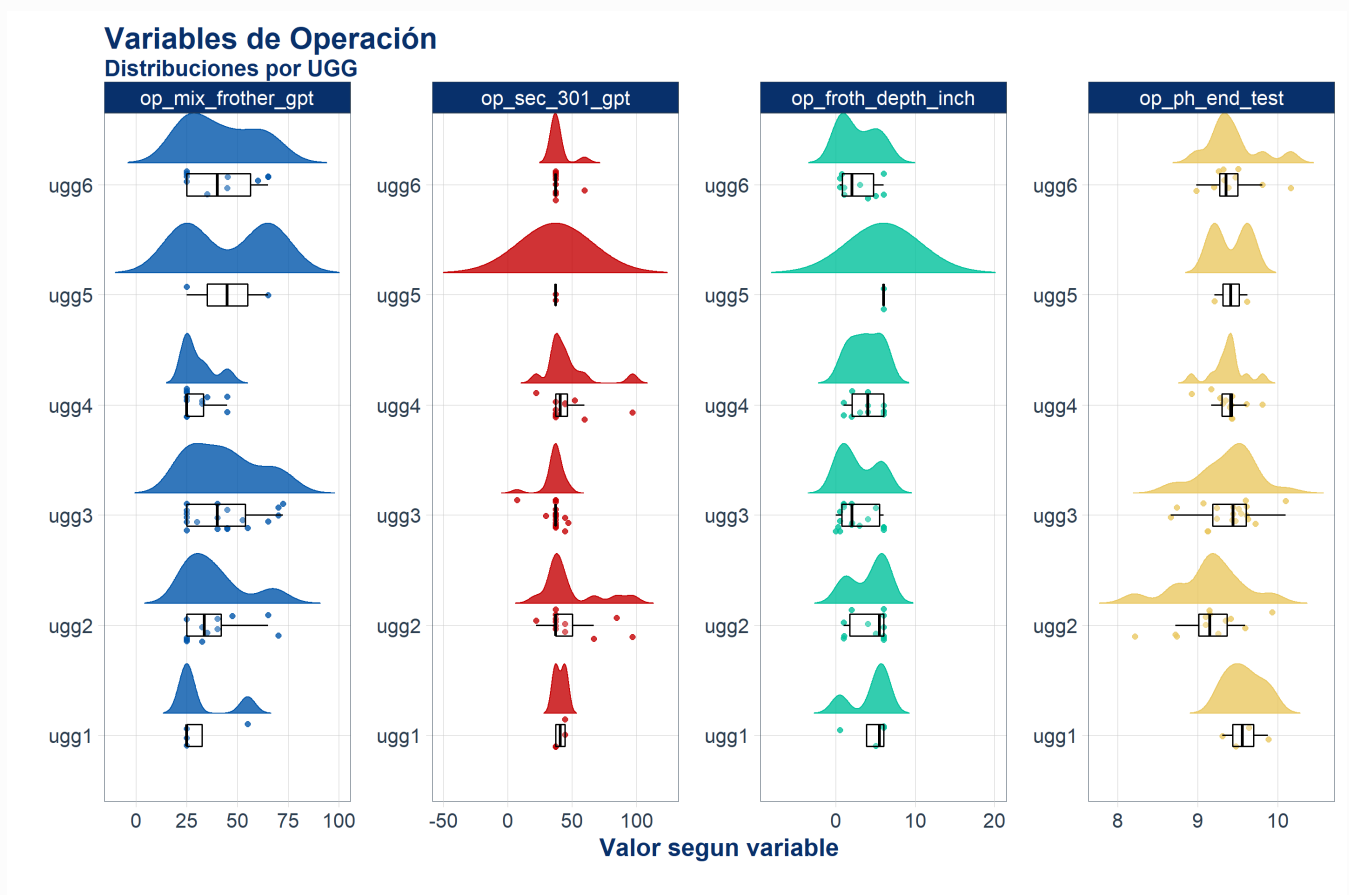




De las variables de entrada se puede indicar principalmente lo siguiente:

- Se observa una correlación inversa entre los diferentes niveles de recuperación metalúrgica de cobre con los contenidos de arcillas, específicamente; el % de caolinita y el % total de arcillas (cao+mus+ser). En este contexto, las UGG's 2 y 4 reportan las mayores niveles de arcillas.
- Se observa una correlación proporcional entre los niveles de recuperación de cobre con el tamaño de grano de cobre y % liberación. En este contexto la UGG 2 muestra los valores mas bajos de tamaño de grano de Cu y % liberación.
- A mayores valores de %-38 um la recuperación reporta valores mas bajos. En este contexto las UGG's 2, 4 y 6 reportan los mayores %-38 um.
- También se observa una relación inversa de la recuperación con el cobre soluble (CuS) y cianurable (CuCN). En este contexto las UGG 2 reporta los mayores contenidos de CuS y CuCN.
- Se observa una relación directa leve del porcentaje de cuarzo con la recuperación de cobre.

Variables de Operación



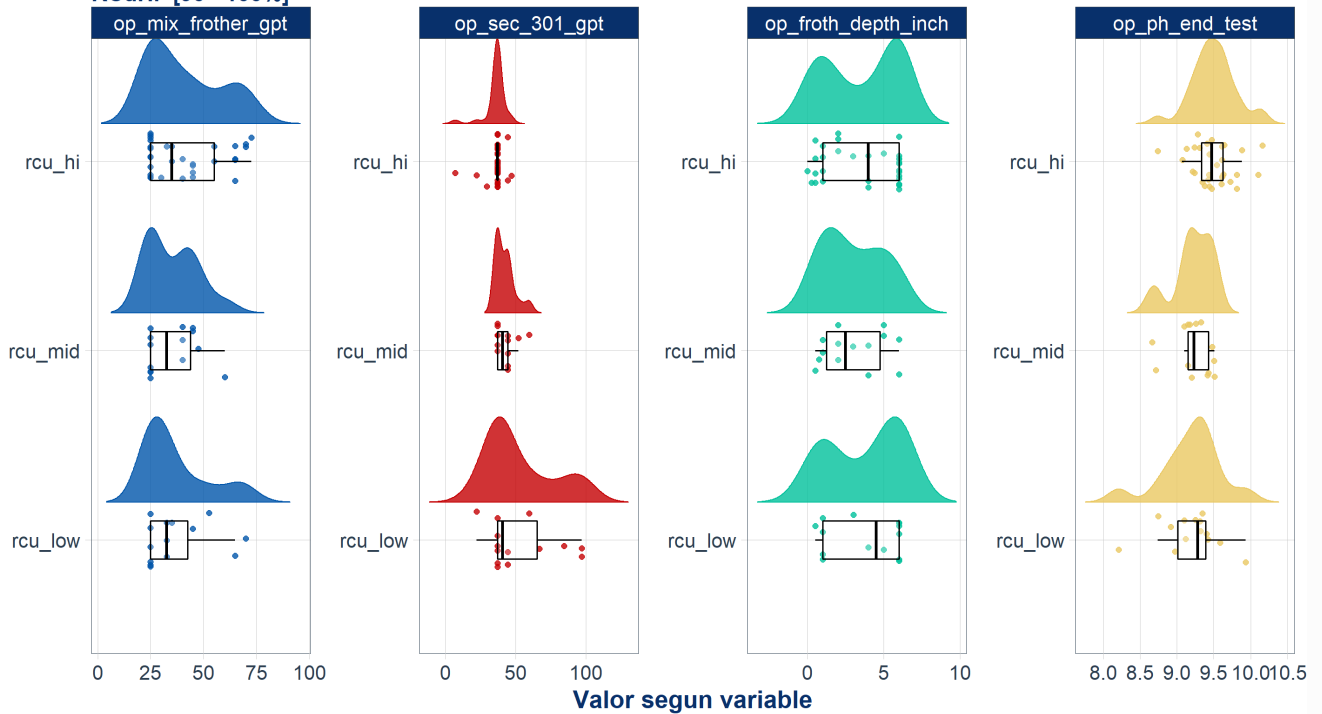
Variables de Operación

Distribuciones por RCuT

RCuLow=[0 - 85%)

RCuMid=[85 - 90%)

RCuHi=[90 - 100%]

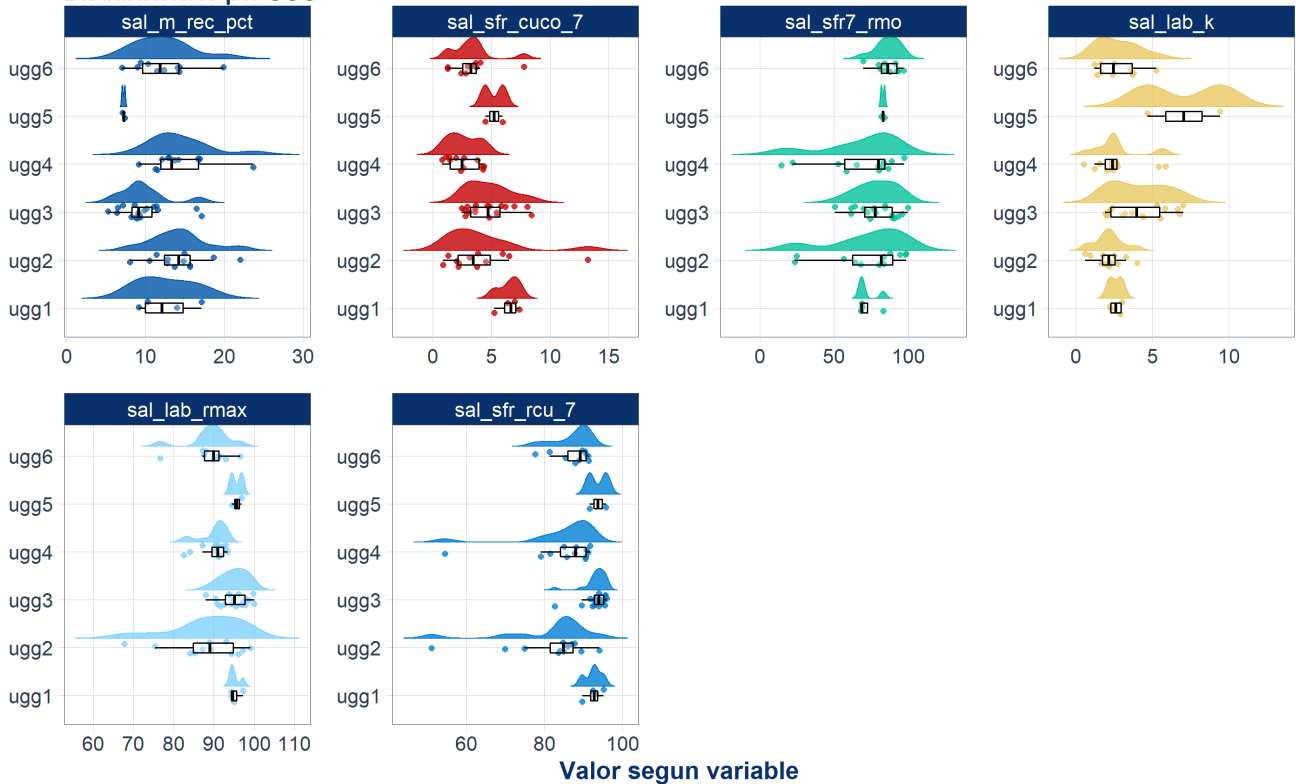


La adición de colector para el rango de recuperaciones bajas fue mayor que en los otros casos (media y alta). Esto debe investigarse en las observaciones de las pruebas experimentales.

Variables de Salida

Variables de Salida

Distribuciones por UGG



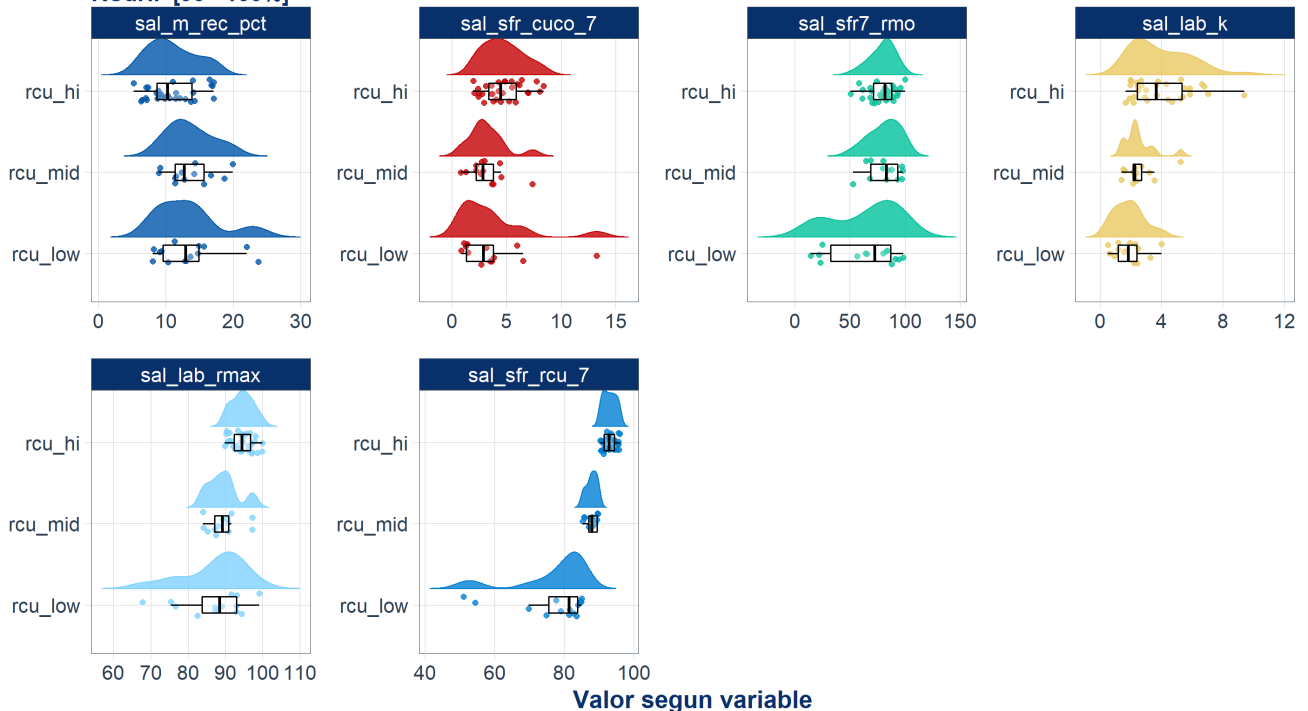
Variables de Salida

Distribuciones por RCuT

RCuLow=[0 - 85%)

RCuMid=[85 - 90%)

RCuHi=[90 - 100%]



De las variables de salida se puede indicar principalmente lo siguiente:

- Como era de esperar los niveles de recuperación están correlacionados con las constantes cinéticas K y RMAX. A su vez UGG's 3 y 5 reportan las mayores K y UGG's 2, 4 y 6 los menores RMAX.
- UGG 2 y 4 reportan las mayores recuperaciones máxicas y a su vez las menores leyes de Cu.

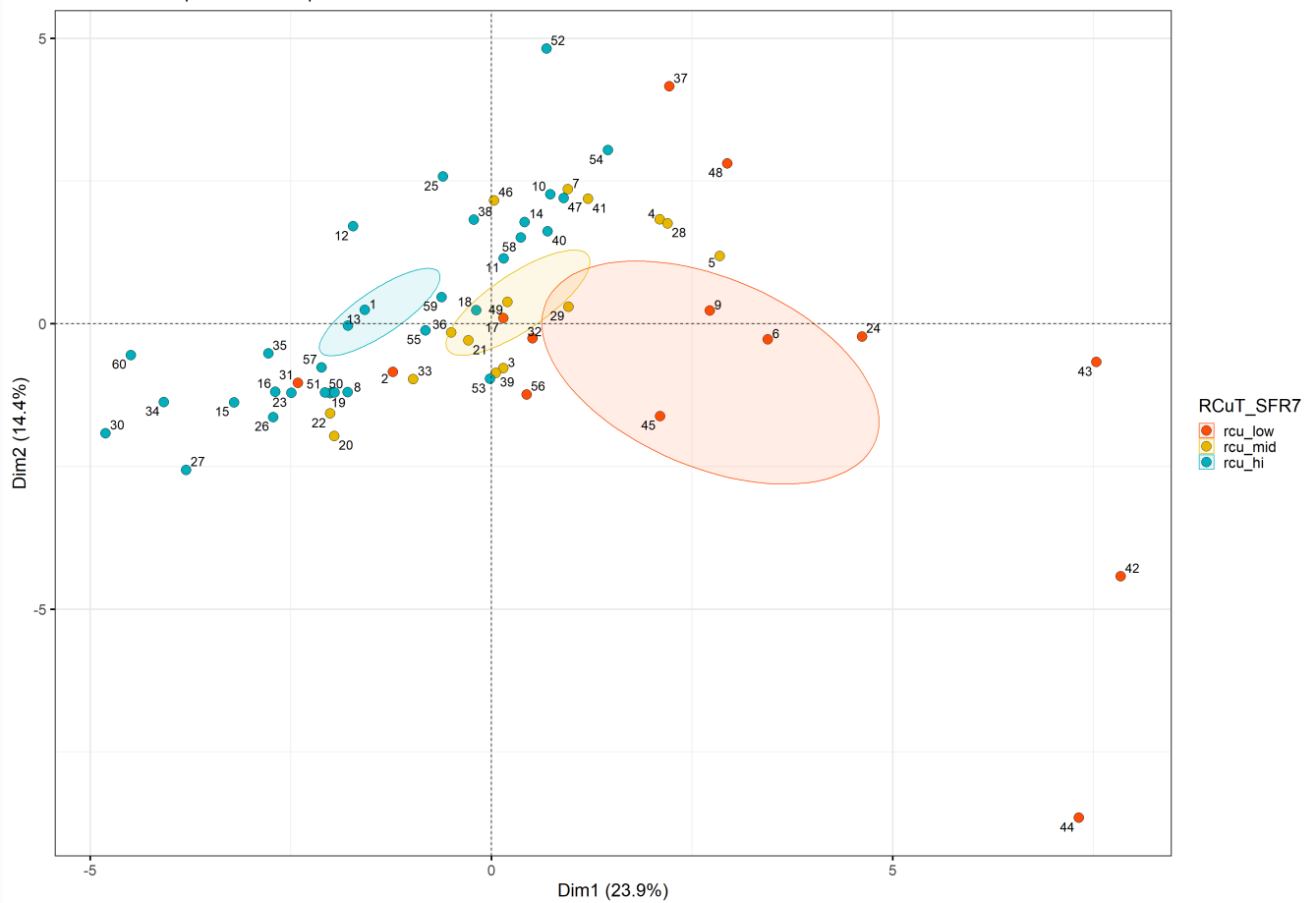
ANALISIS Y SELECCION DE VARIABLES

Componentes Principales

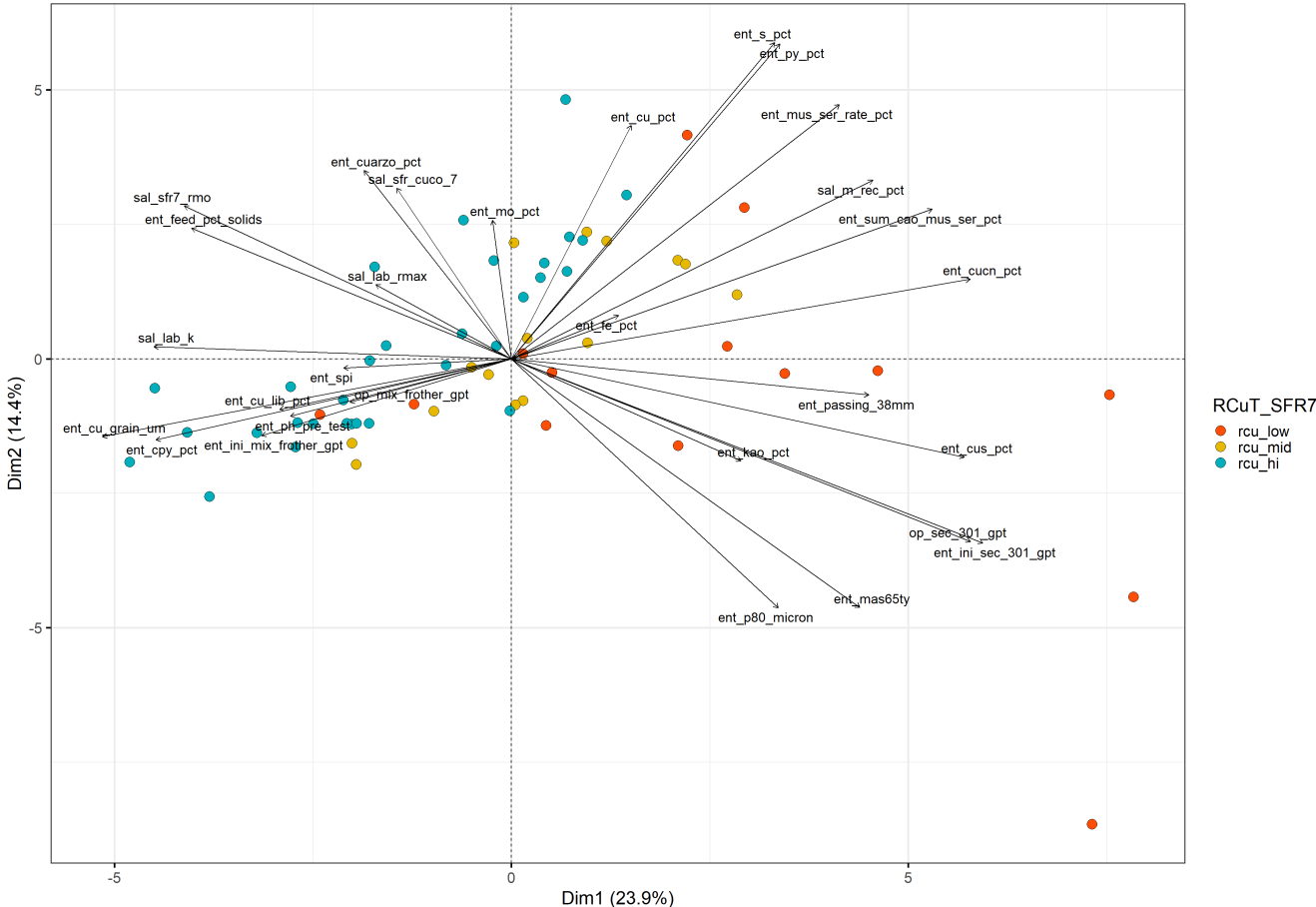
El análisis de componentes principales se utilizó para evaluar el conjunto de datos, de forma de simplificar/reducir sus dimensiones conservando gran parte de la información estadística. Adicionalmente este análisis permite establecer correlaciones entre las variables más importantes del conjunto de datos.

Correlación de Registros Agrupado por RCuT SFR7

Gráfica de Correlación de Registros - Agrupado por RCuT_SFR
Análisis de Componentes Principales

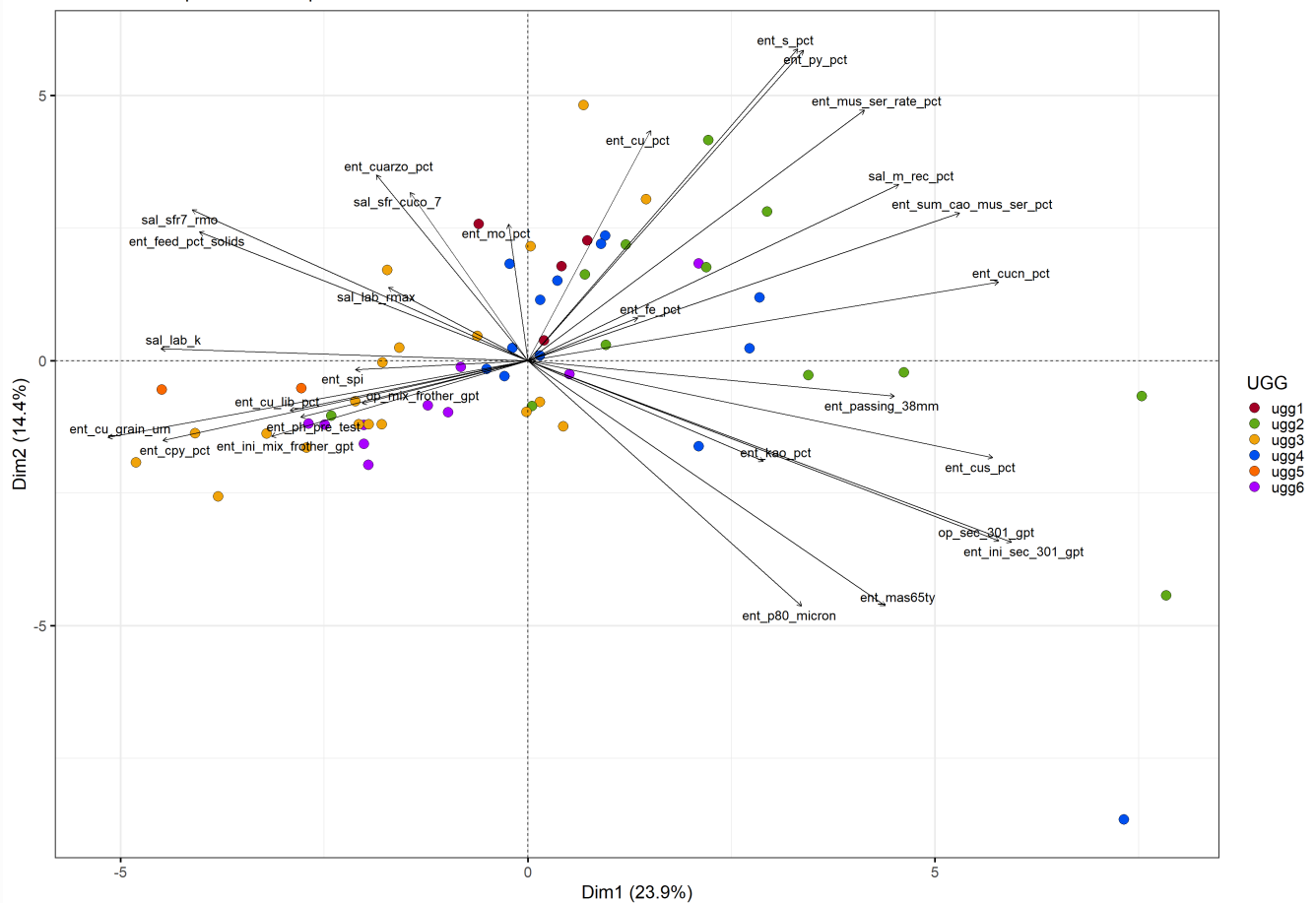


Analisis de Componentes Principales



Correlación de Registros y Variables Agrupado por UGG

Gráfica de Correlación de Registros y Variables - Agrupado por UGG
Análisis de Componentes Principales



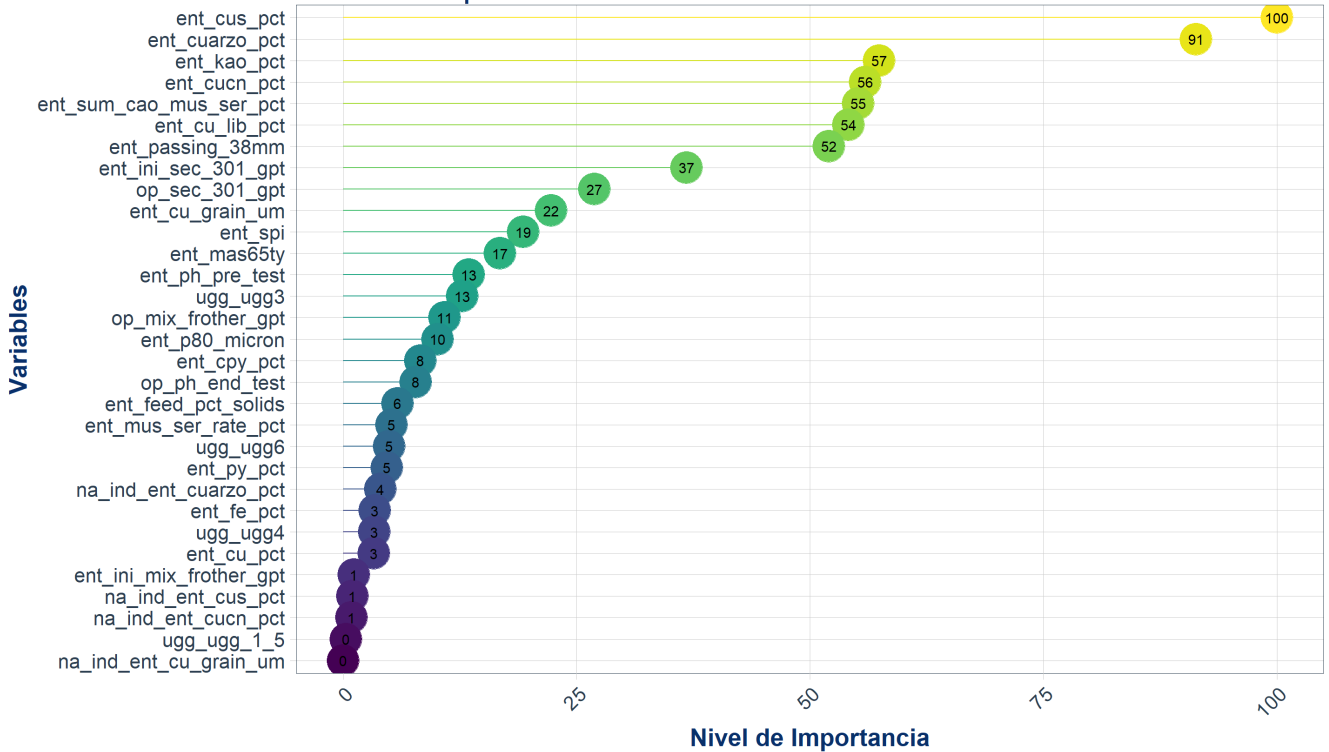
De las gráficas antes mostradas se puede inferir lo siguiente:

1. Registros con perfiles similares tienden a agruparse juntos. En este caso se observa que en general los diferentes niveles de recuperación se agrupan en cuadrantes diferentes, reportando un centro de gravedad distinto como lo muestra la gráfica de correlación de registros.
2. Un registro que esta al mismo lado que una variable (flechas) reporta un valor alto en esa variable. Lo anterior significa que aquellos registros con una recuperación metalurgica de cobre baja y media, tienen valores altos en las variables del 1er y 4to cuadrante como se puede observar en la gráfica de registros y variables.

Metodo de Bosque Aleatorio

Consiste en un conjunto de arboles de desición, los cuales se combinan muchas veces en repeticiones aleatorias con distintas porciones del conjunto de datos a fin de predecir la variable objetivo.

Grafica de Selección de Variables RCuT_SFR7_[%] v/s resto variables Metodo de Bosque Aleatorio



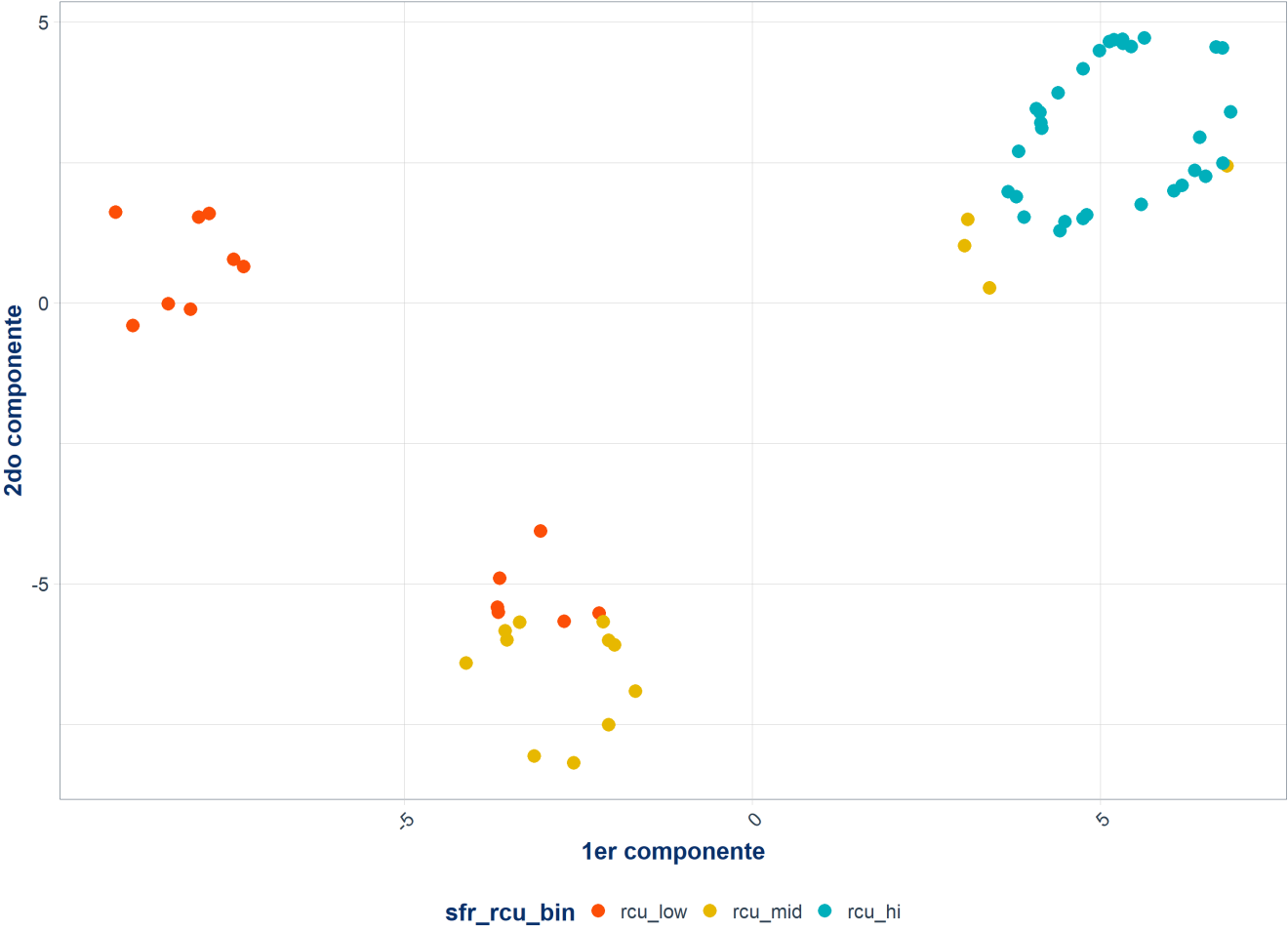
Metodo UMAP (Aproximación y Proyección Uniforme)

Al igual que el análisis de componentes principales, UMAP se usa para reducir la dimensionalidad de un conjunto de datos, de muchas variables/columnas a un espacio 2D o 3D (2 o 3 ejes/variables), las cuales capturan gran parte de la variabilidad del conjunto original. La principal diferencia de UMAP con PCA radica en que el primero es un método no lineal, mientras que el segundo es lineal, lo anterior permitiría obtener un mejor rendimiento de UMAP frente a PCA en estructuras de datos mas complejas.

UMAP para RCuT_SFR

Gráfica de Clusters - Agrupado por RCuT_SFR

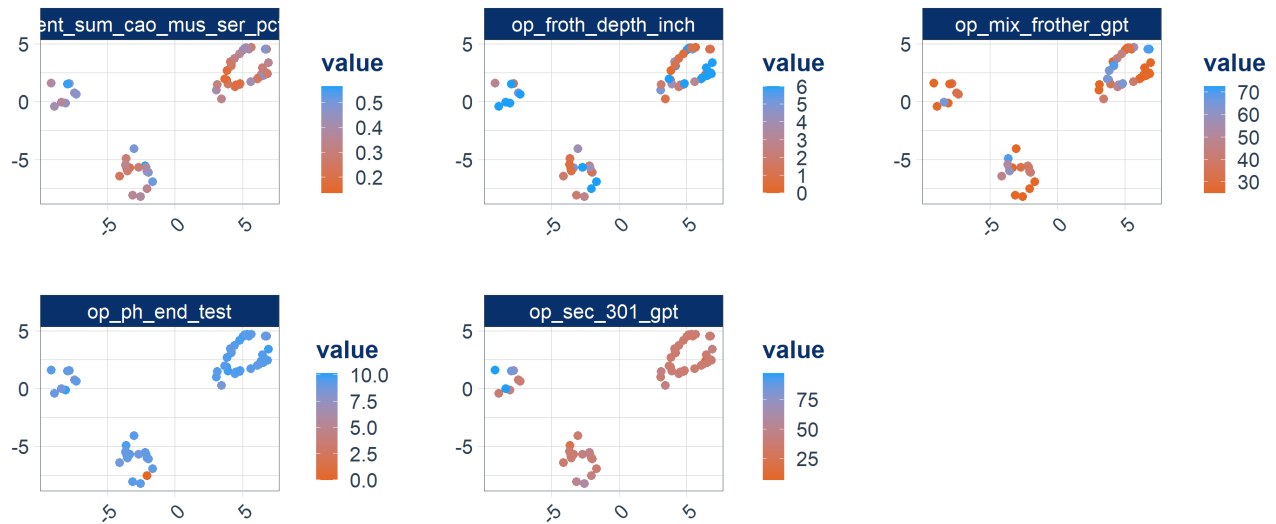
Metodo UMAP



Gráfica de Clusters para Variables

Metodo UMAP



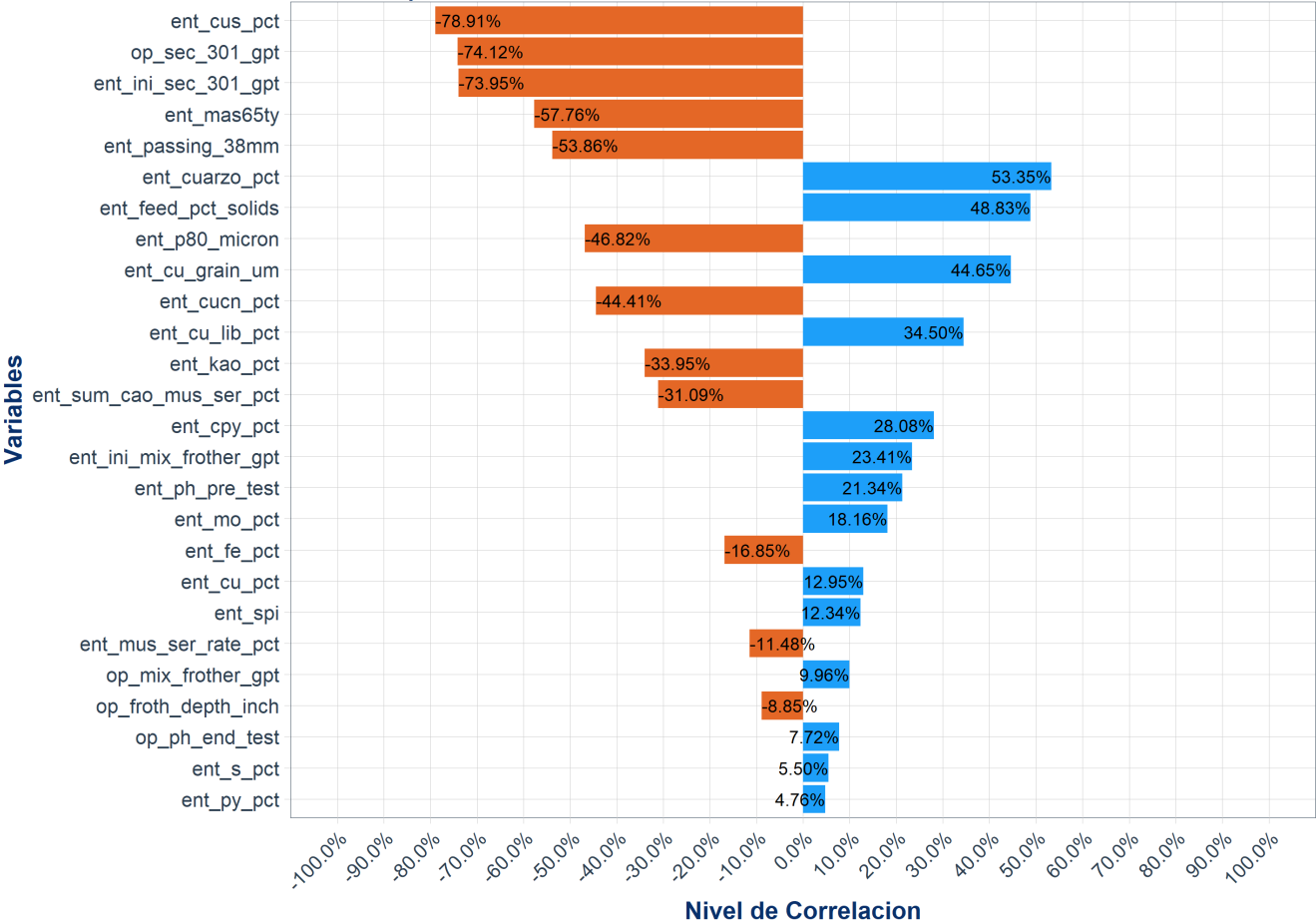


Metodo de Correlación

Consiste en medir el grado en que dos variables tienden a cambiar juntas. El coeficiente describe tanto la fuerza como la dirección de la relación.

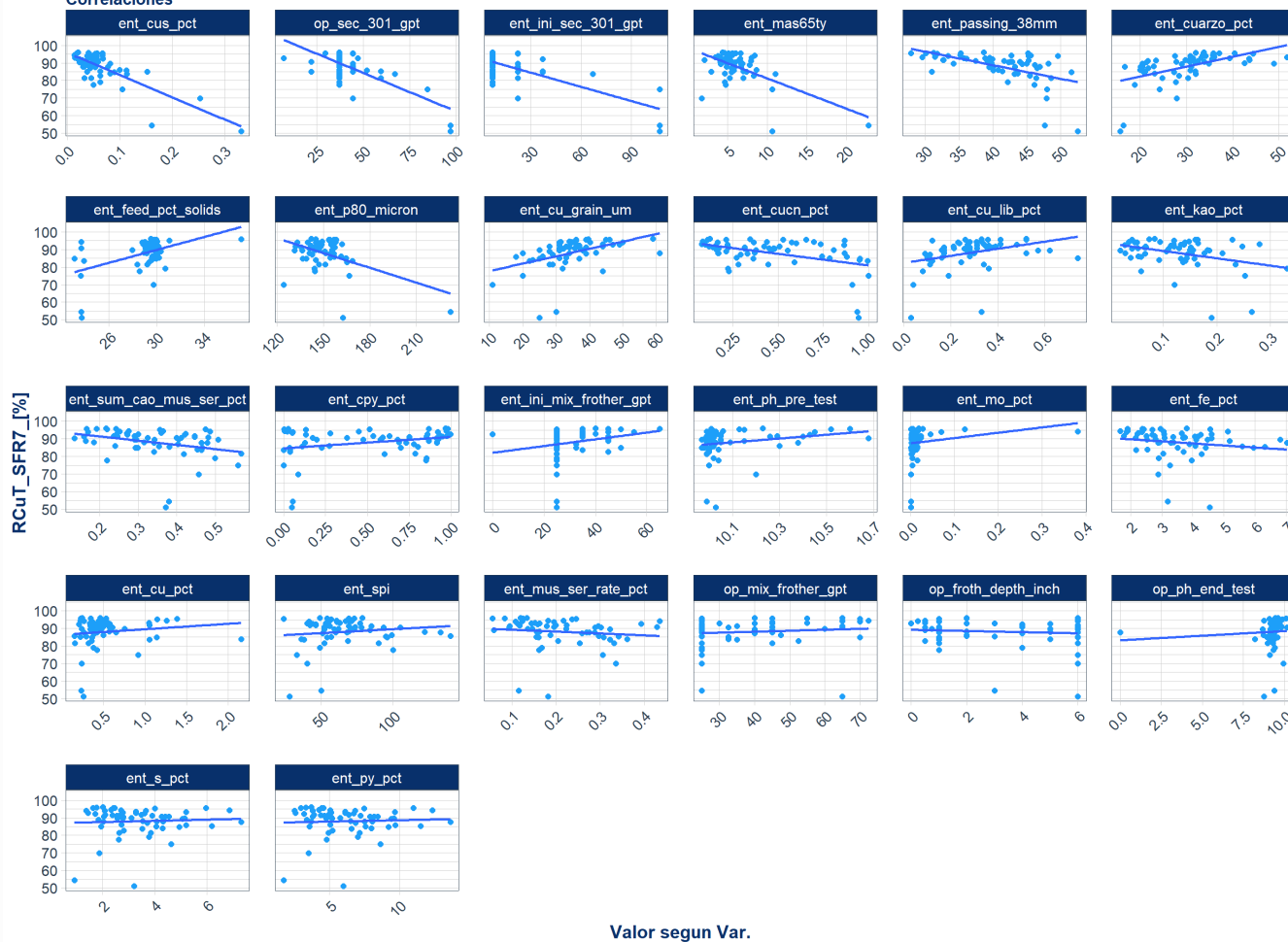
```
[[1]]
```

Grafica de Correlacion
RCuT_SFR7_[%] v/s resto de variables
Metodo de pearson



Grafica de Dispersion RCuT_SFR7_[%] v/s resto de variables

Correlaciones



RCuT_SFR7 v/s PRIN. VAR. POR UGG

RCuT_SFR7 v/s Principales Variables por UGG

RCuT v/s Principales Variables por UGG

