

Contextualización y organización inicial de las BD consolidadas

Etiqueta	Cierre
Fecha	@November 27, 2022
archivos adjuntos	Metadata_5BD.xlsx

Una vez organizada la base de datos, se toma de la metadata las variables necesarias (Solo se omitieron las que eran redundantes en su categorización)

Nota: Solo se seleccionaron registros de los estudiantes de pregrado

Cabe aclarar que las BD se seccionaron para que no haya redundancia en los datos

Los grupos fueron:

- Inscritos
- Admitidos
- Matriculados
- Matriculados en 1er Curso
- Graduados

Entonces teniendo en cuenta lo anterior tenemos **17 variables**

sus nombres originales son los siguientes:

```
> matrix(c(b1,b2,b3,b4,b5), ncol = 5, nrow = 17,  
        dimnames = list(c(), c("INSCRITOS", "ADMITIDOS",  
                                "MATRICULADOS", "MATRI-1ER CURSO", "GRADUADOS"))))
```

INSCRITOS	ADMITIDOS	MATRICULADOS
[1,] "ies_padre"	"ies_padre"	"ies_padre"
[2,] "institucion_de_educacion_superior_ies"	"institucion_de_educacion_superior_ies"	"institucion_de_educacion_superior_ies"
[3,] "ies_sector_ies"	"sector_ies"	"sector_ies"
[4,] "id_caracter"	"id_caracter"	"id_caracter"
[5,] "codigo_del_departamento_ies"	"codigo_del_departamento_ies"	"codigo_del_departamento_ies"
[6,] "departamento_de_domicilio_de_la_ies"	"departamento_de_domicilio_de_la_ies"	"departamento_de_domicilio_de_la_ies"
[7,] "codigo_snies_del_programa"	"codigo_snies_del_programa"	"codigo_snies_del_programa"
[8,] "programa_academico"	"programa_academico"	"programa_academico"
[9,] "nivel_academico"	"nivel_de_formacion"	"nivel_de_formacion"
[10,] "nivel_de_formacion"	"metodologia"	"metodologia"
[11,] "metodologia"	"id_area"	"id_area_de_conocimiento"
[12,] "id_area"	"id_nucleo"	"id_nucleo"
[13,] "id_nucleo"	"nucleo_basico_del_conocimiento_nbc"	"nucleo_basico_del_conocimiento_nbc"
[14,] "nucleo_basico_del_conocimiento_nbc"	"idsexo"	"idsexo"
[15,] "ano"	"ano"	"ano"
[16,] "semestre"	"semestre"	"semestre"

	"inscritos"	"admitidos"	"matriculados"
[17,]	MATRI-1ER CURSO	GRADUADOS	
[1,]	"ies_padre"	"ies_padre"	
[2,]	"institucion_de_educacion_superior_ies"	"institucion_de_educacion_superior_ies"	
[3,]	"sector_ies"	"sector_ies"	
[4,]	"id_caracter"	"id_caracter"	
[5,]	"codigo_del_departamento_ies"	"codigo_del_departamento_ies"	
[6,]	"departamento_de_domicilio_de_la_ies"	"departamento_de_domicilio_de_la_ies"	
[7,]	"codigo_snies_del_programa"	"codigo_snies_del_programa"	
[8,]	"programa_academico"	"programa_academico"	
[9,]	"nivel_de_formacion"	"nivel_de_formacion"	
[10,]	"metodologia"	"metodologia"	
[11,]	"id_area_de_conocimiento"	"id_area"	
[12,]	"id_nucleo"	"id_nucleo"	
[13,]	"nucleo_basico_del_conocimiento_nbc"	"nucleo_basico_del_conocimiento_nbc"	
[14,]	"id_sexo"	"id_sexo"	
[15,]	"ano"	"ano"	
[16,]	"semestre"	"semestre"	
[17,]	"primer_curso"	"graduados"	

Ahora para homogeneizar las variables, creamos un vector con el nuevo nombre de las 17 columnas (renombramiento):

```
> new.nombs
[1] "code_ies"  "nom_ies"  "sector"
[4] "caracter"  "code_dep" "nom_dep"
[7] "code_prog" "prog_aca" "nivel_form"
[10] "metodo"    "area"     "nucleo"
[13] "nbc"       "sexo"     "year"
[16] "sem"       "conteo"
```

Con su respectiva Metadata (archivo adjunto).