

POSTGRADO |

UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

Maestría en Data Science

GESTIÓN DE DATOS

Análisis de matrícula de estudiantes de nivel inicial de Educación Básica Regular del 2020, 2021 y 2022

Integrantes:

Alexander Cubas Francia
Ernesto Igor Laura Mamani
Grabiela Quispe Ochoa
Yerimen Antonio Campos Luyo

Profesor:

Oscar Efraín Ramos Ponce

Fecha de entrega
28 de Octubre de 2022

Índice

1	Introducción	1
1.1	Objetivos	1
1.2	Descripción de la información	2
1.3	Datos utilizados	2
1.4	Importancia de los datos	3
2	Metodología	4
2.1	Metodología utilizada	4
2.2	Herramientas usadas	4
2.3	Trabajo con los datos	5
2.4	Análisis realizados	5
2.4.1	Niveles educativos	5
2.4.2	Evolución mensual de matrículas	5
2.4.3	Discapacidad	6
2.4.4	Nacionalidades	6
2.4.5	Género	6
2.4.6	DNI Validados	6
2.5	Uso de shiny	7
2.6	Fuentes	7
3	Resultados	8
3.1	Nivel educativo	8
3.2	Evolución mensual de matrículas	9
3.3	Discapacidad	9
3.4	Nacionalidad de estudiantes	9
3.5	Distribución de género	10
3.6	DNI Validados	10
3.7	Representación por Shiny	11
3.8	Representaciones de ubicación de instituciones educativas por mapa	11
4	Anexos	15
4.1	Código de conexión a la base de datos	15
4.2	Niveles educativos	15
4.3	Evolución mensual de matrículas	15
4.4	Discapacidad	16
4.5	Nacionalidades	16
4.6	Género	16
4.7	DNI Validados	16
4.8	Shiny	17
4.9	Representaciones de ubicación de instituciones educativas por mapa	19

Bibliografía

20

Capítulo 1

Introducción

En el Sistema de Información de Apoyo a la Gestión de la Institución Educativa – SIAGIE, se realiza el registro administrativo de la trayectoria educativa del estudiante durante su permanencia en la Institución Educativa, basado en la información contenida en las nóminas de matrícula y actas de evaluación emitidas por las Instituciones Educativas o Programas Educativos.

El SIAGIE se configura por servicio educativo y se adecua a la normatividad vigente que regula los procesos de matrícula y evaluación de aprendizajes en las distintas modalidades de la Educación Básica.

1. RM N° 609-2018-MINEDU 09-11-2018
2. RM N° 665-2018-MINEDU 04-12-2018
3. RVM N° 025-2019-MINEDU 10-02-2019
4. RM N° 712-2018-MINEDU 21-12-2018

1.1 Objetivos

Alineados con los objetivos del área de estadística del ministerio de educación, los cuales son:

- Realizar la crítica e interpretación de los datos estadísticos para su difusión y conocimiento de los órganos de la sede regional.
- Consolidar a nivel regional la información preliminar de cada mes, reportadas por las UGEs.
- Recopilar, consolidar, elaborar y difundir la información estadística básica cada mes.
- Realizar la estimación de la matrícula regional captada por muestreo y difundir al inicio del año lectivo, por el método de cifras relativas.

TABLA 1.1: Detalle de los campos de la entidad estudiante

Columna	Descripción	Tipo
ID Año	Identificador del año escolar	ID
Departamento	Nombre del departamento	cualitativa - nominal
Provincia	Nombre de la provincia	cualitativa - nominal
Distrito	Nombre del distrito	cualitativa - nominal
Centro Poblado	Centro poblado	cualitativa - nominal
NLongitud IE	Coordenada de la longitud de la IE	cuantitativa - continua
NLatitud IE	Coordenada de la latitud de la IE	cuantitativa - continua
DRE	Dirección regional de educación	cualitativa - nominal
UGEL	Ugel a la que pertenece la IE	cualitativa - nominal
Cod Mod	Código modular	ID
Anexo	Anexo	cualitativa - nominal
Nombre IE	Nombre de la Institución educativa	cualitativa - nominal
D Gestión	Dirección de gestión	cualitativa - nominal
Dsc Turno	Descripción del turno	cualitativa - nominal
Modalidad	Modalidad	cualitativa - nominal
Nivel Educativo	Nivel en el que se encuentra el estudiante	cualitativa - ordinal
Fecha Matrícula	Fecha de la matrícula	cuantitativa - ordinal
ID Persona	Identificador del estudiante	ID
Documento Identidad	Documento de identidad	ID
Sexo	Sexo del alumno	cualitativa - nominal
Madre vive	Indicador si la madre vive	cualitativa - nominal
Padre vive	Indicador si el padre vive	cualitativa - nominal
Dsc Grado	Descripción del grado	cualitativa - ordinal
Dsc Sección	Descripción de la sección	cualitativa - ordinal
Nacionalidad	Nacionalidad del estudiante	cualitativa - nominal
Dsc Discapacidad	Descripción de discapacidad	cualitativa - nominal
Dirección	Dirección del estudiante	cualitativa - nominal
Lugar	Lugar de la dirección	cualitativa - nominal
Ubigeo	Ubigeo del estudiante	cualitativa - nominal

1.2 Descripción de la información

En el presente trabajo se está analizando las matrículas de los estudiantes de nivel inicial de Educación Básica Regular de los años 2020, 2021, 2022, con la finalidad de conocer diversos factores que involucran una matrícula correctamente realizada, como también los datos de los estudiantes, los cuales deben ser registrados correctamente en la plataforma; logrando un seguimiento mensual de dicho estado, entre otras características.

1.3 Datos utilizados

La información con la que se va a trabajar está conformada por los datos de alumnos de niveles inicial, primaria y secundaria de colegios en todo el Perú, esta información corresponde a los 3 últimos años, el detalle de esta entidad se ve en la tabla 1.1

1.4 Importancia de los datos

Debido a que la información trabajada proviene del ámbito educativo de niveles básicos, se pueden generar estadísticas para poder evaluar el estado actual educativo.

A continuación, se utiliza R Studio y Shiny para implementar un dashboard que permita realizar el seguimiento mensual. Cada uno de los gráficos ayudará a identificar por año el incremento o caída de cada variable mostrada; esta información permitirá tomar cualquier acción preventiva o mejoras frente a la educación inicial.

Capítulo 2

Metodología

2.1 Metodología utilizada

En el presente proyecto se ha utilizado la metodología CRISP DM[1], pero sólo en las etapas iniciales que tienen que ver con el trabajo de los datos. En la imagen 2.1 se pueden ver resaltadas las etapas cubiertas.

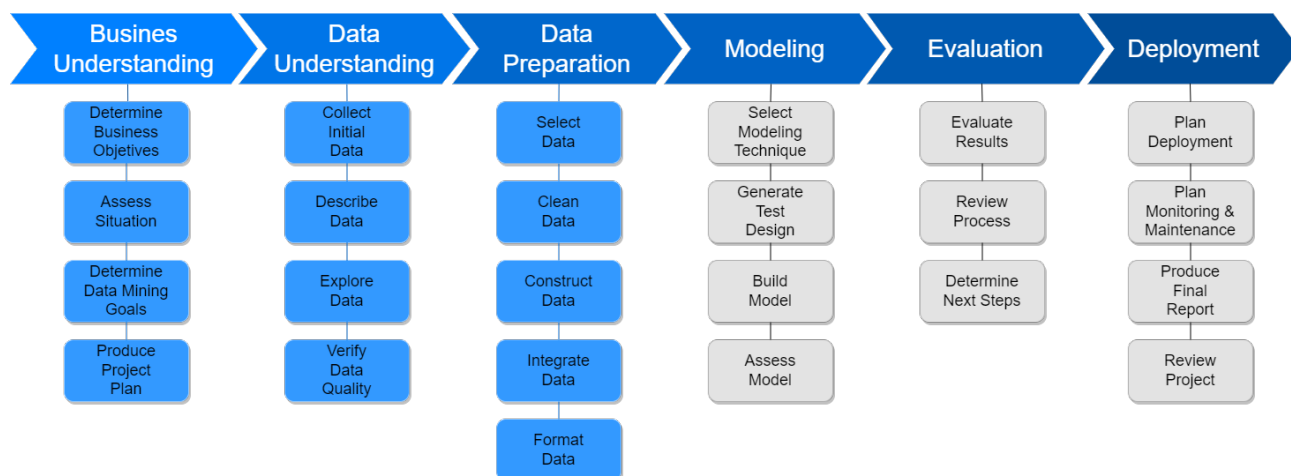


IMAGEN 2.1: Secciones cubiertas en el proyecto

2.2 Herramientas usadas

Se han hecho uso de las siguientes herramientas:

- Base de datos MySql
- Servicio RDS de AWS
- R Studio
- JetBrains DataGrip
- Shiny
- Git (github)

Se obtuvo un archivo CSV con toda la información de estudiantes, este archivo fue subido mediante DataGrip a una BD MySQL, previamente creada como un servicio RDS de AWS, con estos datos cargados se realizaron procesos mediante R para obtener información que posteriormente fueron registrados en otra base de datos y mostrados mediante una aplicación Shiny.

En la imagen 2.2 se puede ver el flujo del proceso en este trabajo de análisis.

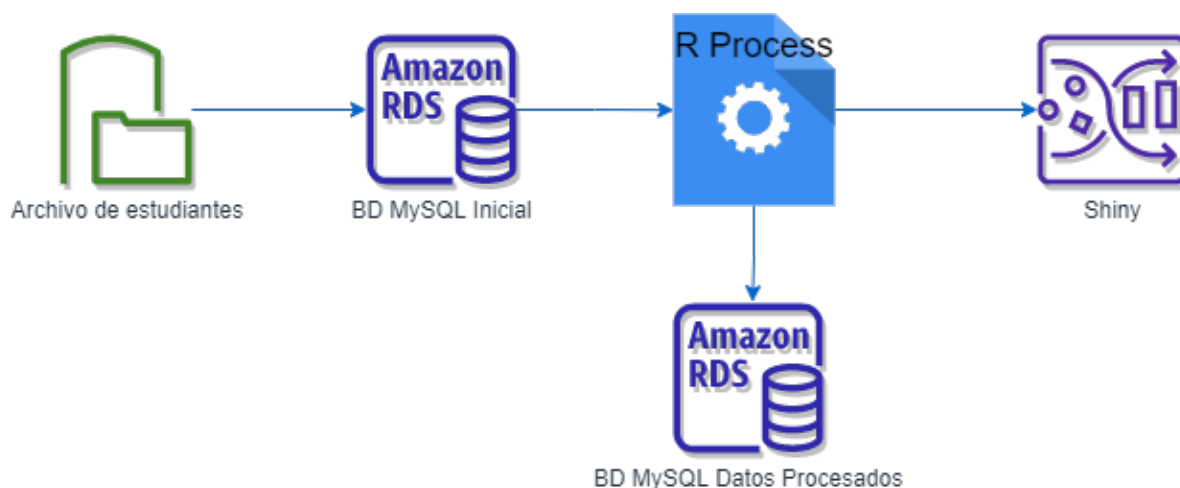


IMAGEN 2.2: Herramientas usadas

2.3 Trabajo con los datos

Para el trabajo con R y MySQL se hizo uso de la biblioteca RMySQL y DBI, el código para la conexión se encuentra en la sección 4.1.

2.4 Análisis realizados

2.4.1 Niveles educativos

Para el caso del nivel educativo, es importante conocer la cantidad de alumnos por año en cada nivel, de esta forma identificamos si existe algún incremento o alguna disminución, para lo que se deberá realizar diversas acciones contrarrestando cualquier inconveniente de materiales y de inmobiliarios.

En el apéndice 4.2 está el código fuente de esto.

2.4.2 Evolución mensual de matrículas

- Se hizo una consulta a la Base de Datos donde se encuentra la tabla estudiante.
- Debido que sólo había una columna con la fecha de la matrícula en formato YYYY-MM-DD, se obtuvo sólo los valores de año y mes para que se puedan agrupar por este campo.
- Luego se agrupó esta información por el nuevo campo que contiene el año y el mes.

- Se categorizó el campo mes.
- Se agregó un campo de año para que pueda ser filtrado posteriormente.

El código para obtener la información para este análisis se encuentra en el anexo 4.3.

El resultado de este análisis se encuentra en la sección 3.2.

2.4.3 Discapacidad

Para el procesamiento de la data se realizó limpieza de datos vacíos, posterior a eso se agrupó y realizó el conteo de estudiantes que tengan alguna discapacidad. Con estos datos se puede posteriormente filtrar por año para que pueda ser visualizada dinámicamente en shiny. En el apéndice 4.4 está el código fuente de esto.

2.4.4 Nacionalidades

Se desea mostrar la cantidad de estudiantes de nacionalidad diferente a la peruana. Para el procesamiento de los datos no se toma en cuenta a los estudiantes con nacionalidad peruana y datos en blanco. En el apéndice 4.5 está el código fuente de esto.

2.4.5 Género

- Se analizó la distribución del género de los estudiantes por año, para esto necesitamos todos los DNIs válidos para proceder con el conteo, por lo cual se realizó una limpieza a la data.
 - Limpieza a los strings de genero ya que se encontraban con saltos de línea:

```
> head(ValidSexo)
  ID_ANIO VALIDACION_DNI SEXO
1  2022   DNI VALIDADO MUJER
2  2022   DNI VALIDADO MUJER
3  2022   DNI VALIDADO HOMBRE
4  2022   DNI VALIDADO MUJER
5  2022   DNI VALIDADO MUJER
6  2022   DNI VALIDADO MUJER
```

IMAGEN 2.3

- Para ello solo extraemos una parte del string (6 primeros caracteres) y está solucionado:
- Se lee la información con un parámetro que quita los registros en blanco, para evitar DNIs que no sean válidos.

El código se encuentra en el apéndice 4.6

2.4.6 DNI Validados

- Se agrupa el dataframa de estudiantes por el campo de validación de DNI.

```
> ValidSexo$SEXO = substr(ValidSexo$SEXO, 1, 6)
> head(ValidSexo)
```

	ID_ANIO	VALIDACION_DNI	SEXO
1	2022	DNI VALIDADO	MUJER
2	2022	DNI VALIDADO	MUJER
3	2022	DNI VALIDADO	HOMBRE
4	2022	DNI VALIDADO	MUJER
5	2022	DNI VALIDADO	MUJER
6	2022	DNI VALIDADO	MUJER

IMAGEN 2.4

- Se crea un nuevo DF que contiene el conteo del campo que representa la validación de un DNI.
- Se obtiene un porcentaje de los valores del paso anterior.

El código se encuentra en el apéndice 4.7

2.5 Uso de shiny

Se usó Shiny [2] para mostrar los análisis de manera interactiva, ya que se tiene información de 3 años (2020, 2021 y 2022).

Los componentes utilizados en el proyecto fueron tablas, gráfico de barras, gráfico circular y gráfico de líneas [3]. El código para realizar esta funcionalidad se encuentra en el anexo 4.8.

Para la publicación se hace uso del servidor que brinda RStudio [4].

2.6 Fuentes

Las fuentes del proyecto se encuentran en el repositorio Git: <https://github.com/Maestria-Data-Science-UPC-G1/gestiondatos.git>

Capítulo 3

Resultados

3.1 Nivel educativo

Para el caso del nivel educativo, es importante conocer la cantidad de alumnos por año en cada nivel, de esta forma identificamos si existe algún incremento o alguna disminución, para lo que se deberá realizar diversas acciones contrarrestando cualquier inconveniente de materiales y de inmobiliarios. Ver la imagen 3.1

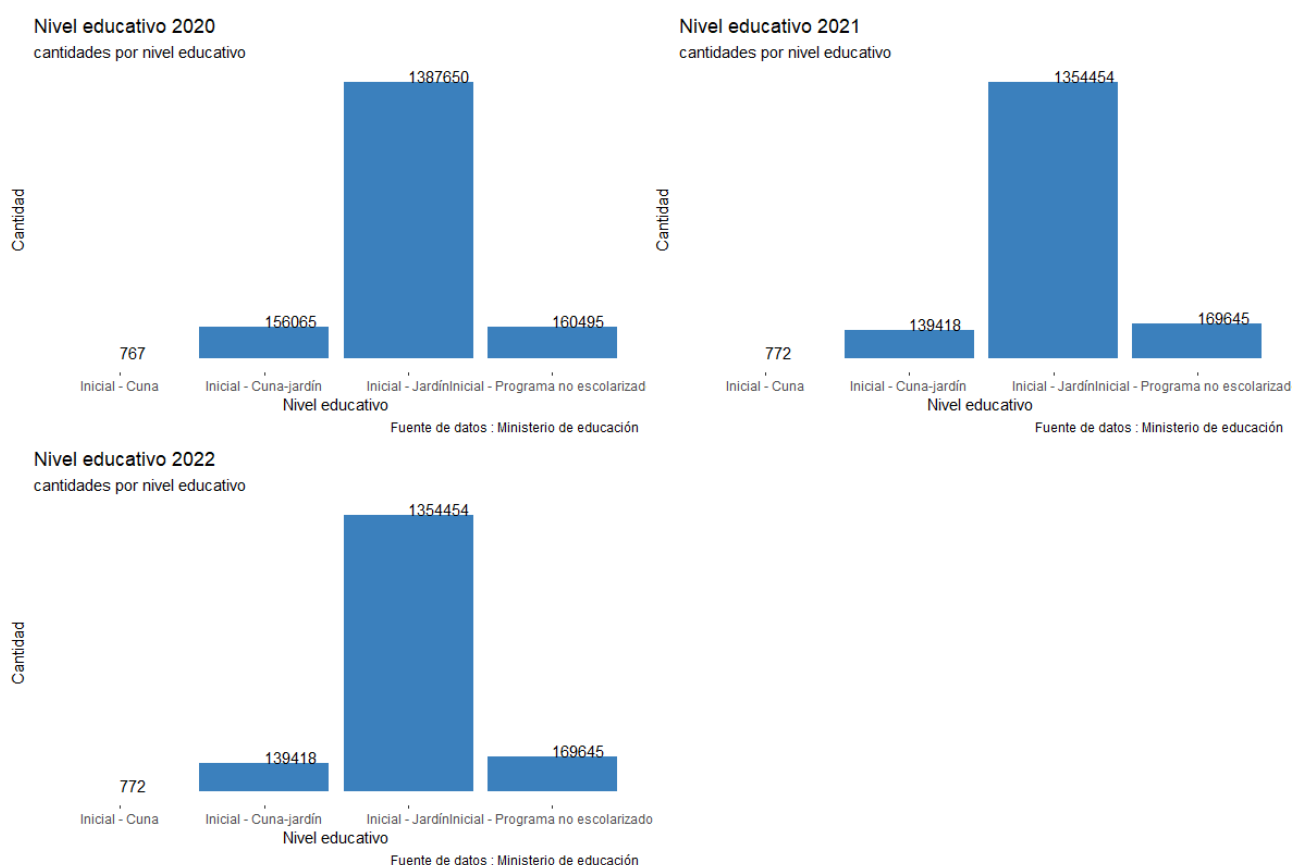


IMAGEN 3.1: Niveles educativos

3.2 Evolución mensual de matrículas

En este análisis se identificó la evolución que han tenido la cantidad de matrículas de cada mes en los últimos 3 años, las siguientes imágenes 3.2 muestran la evolución por año.

Se aprecia una mayor cantidad de incidencias entre Marzo y Abril, esto se puede entender por el inicio del año escolar en el Perú. Y en menor medida hay una cantidad elevada de matrículas desde el inicio de año, durante el resto del año se ve una menor cantidad pero si ocurren, esto podría deberse posiblemente a traslados de colegio.

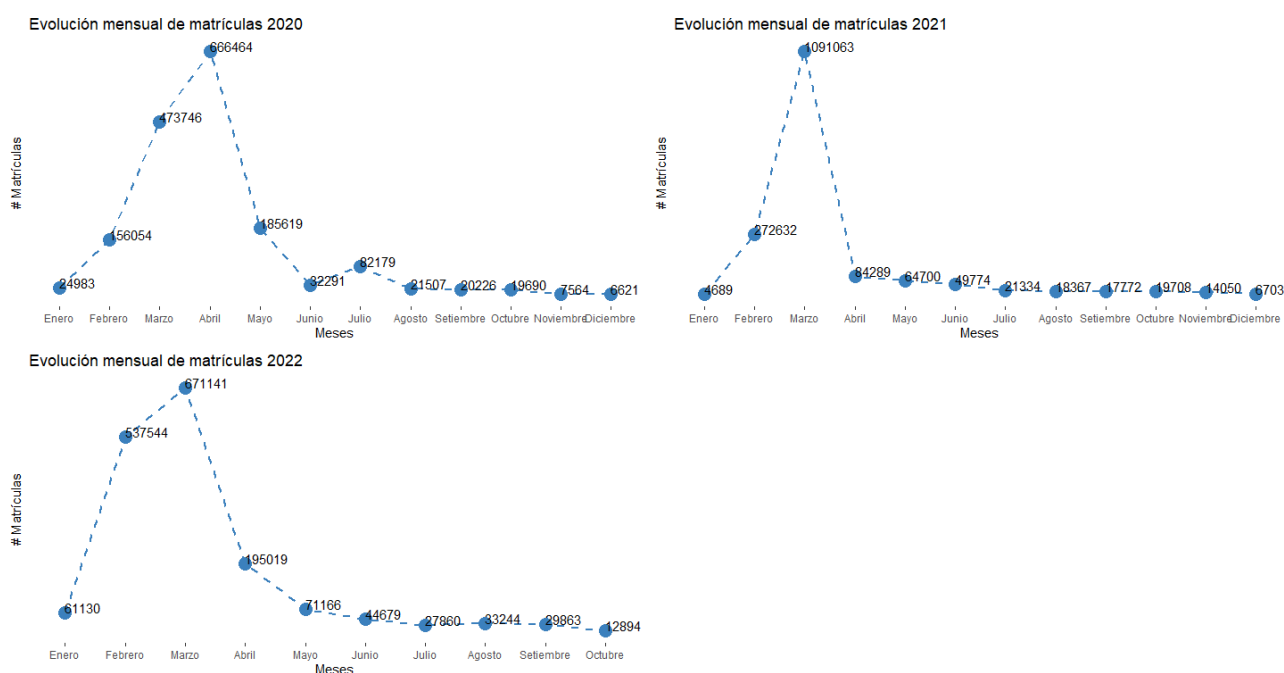


IMAGEN 3.2: Evolución de cantidades por cada mes

3.3 Discapacidad

El presente gráfico 3.3 queremos mostrar la cantidad de estudiantes con alguna discapacidad sea visual, sordoceguera, autismo, auditiva entre otros.

Se puede apreciar que en los últimos 2 años se tenía mayor incidencia en las discapacidades motora, intelectual y autismo, pero en el presente año han disminuído las que tienen que ver con motora e intelectual, pero el autismo se mantiene como un factor predominante.

3.4 Nacionalidad de estudiantes

Se muestra a continuación una imagen 3.4 la distribución de nacionalidades, teniendo como mayor cantidad, luego de la peruana, la nacionalidad Venezolana, debido a que hay muchos países en este análisis, se ve conveniente el uso de una tabla para poder mostrar los datos.

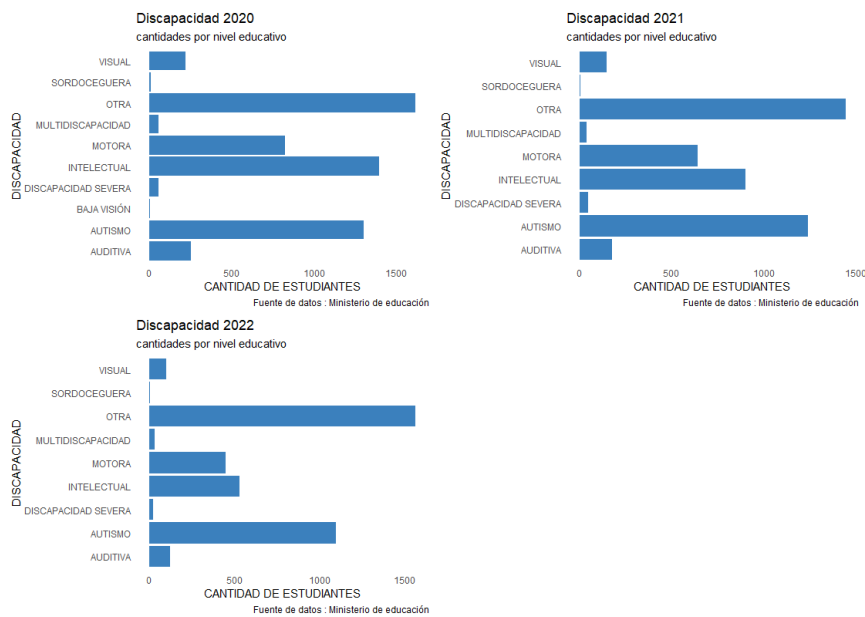


IMAGEN 3.3: Tipos de discapacidad

Estudiantes de otra nacionalidad

Show entries

ID_ANIO	NACIONALIDAD	CANTIDAD
2020	ALEMANIA	28
2020	ALEMANIA REPUBLICA DEMOCRATICA	5
2020	ANGOLA	1
2020	ANTILLAS HOLANDESAS	3
2020	ARABIA SAUDITA	1
2020	ARGELIA	2
2020	Argentina	2195
2020	AUSTRALIA	28
2020	AUSTRIA	5
2020	AZERBAIYAN	1

Showing 1 to 10 of 96 entries

Previous **1** 2 3 4 5 ... 10 Next

IMAGEN 3.4: Nacionalidad de estudiantes

3.5 Distribución de género

En la imagen 3.5 se pueden apreciar distribuciones similares entre hombres y mujeres.

3.6 DNI Validados

En la imagen 3.6 se aprecia que hay una gran mayoría de documentos validados, pero se ve en menor medida estudiantes indocumentados.

Distribucion % por genero

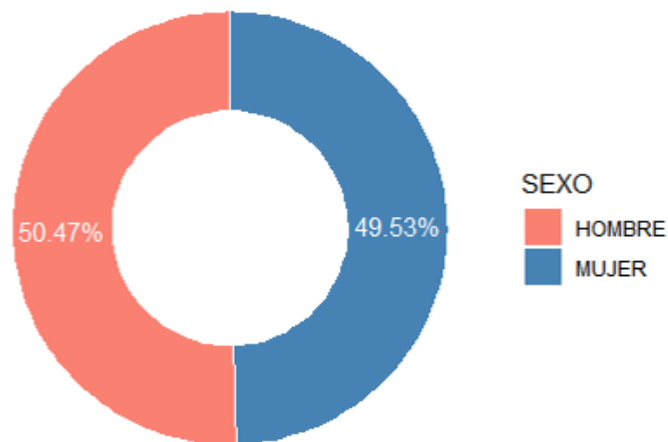


IMAGEN 3.5: Distribuciones de género

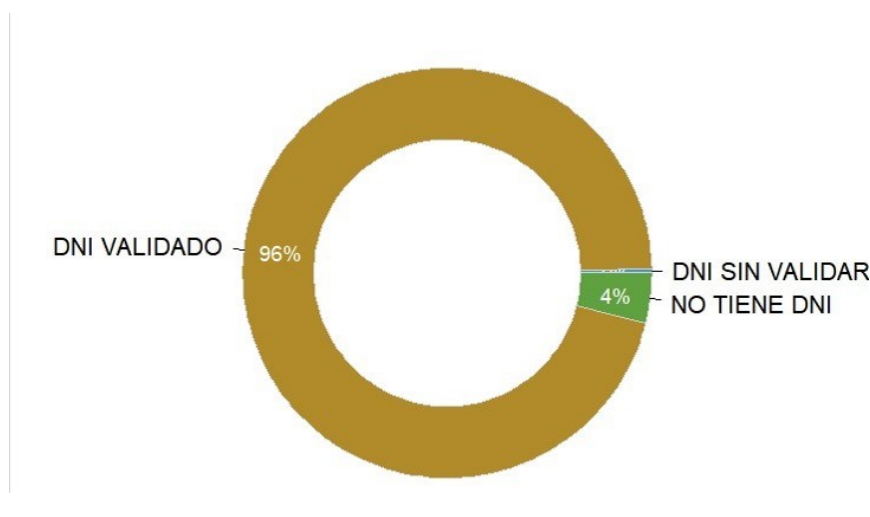


IMAGEN 3.6: Validaciones de DNI

3.7 Representación por Shiny

En shiny se desarrolló un panel que tiene un filtro por año para poder filtrar los gráficos presentados anteriormente según la imagen 3.7, la reactividad ante este campo de entrada permite tener la visualización de lo que ocurre por año.

3.8 Representaciones de ubicación de instituciones educativas por mapa

Ya que es importante conocer la ubicación de los colegios, se realizó un mapa utilizando las librerías ggplot2, tidyverse y mapview en R, esta última librería proporciona funciones para crear visualizaciones interactivas de datos espaciales de manera muy rápida y conveniente. El código de estas representaciones se encuentra en

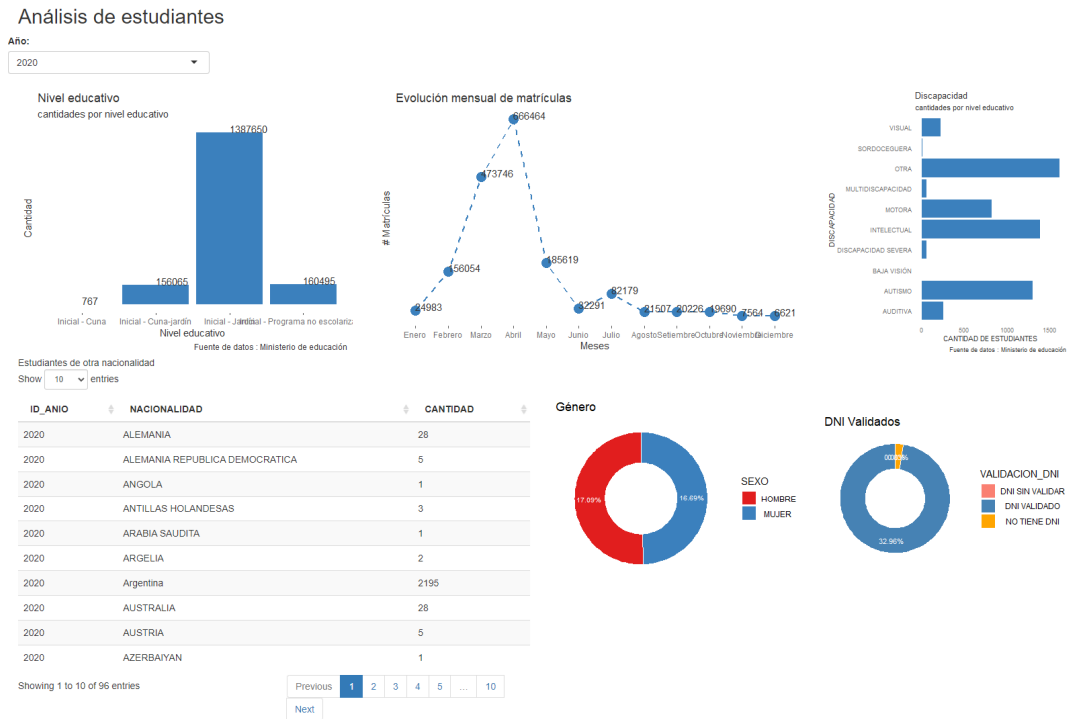


IMAGEN 3.7: Representación mediante Shiny

el anexo 4.9

Diagrama 3.8: Su objetivo principal es llenar el vacío del trazado interactivo rápido (sin grado de presentación) para examinar e investigar visualmente ambos aspectos de los datos espaciales, las geometrías y sus atributos

Diagrama 3.9: También podemos ver un conteo rápido del número de colegios por sector, con la librería leaflet, con una muestra que se extrajo, podemos ver a medida que vamos aumentando el zoom en que zonas están mayor concentrados los colegios.

Diagrama 3.10: Aprovechamos la interactividad del gráfico y cada que acercamos (zoom) nos muestra la cantidad de colegios acumulados en cada zona.

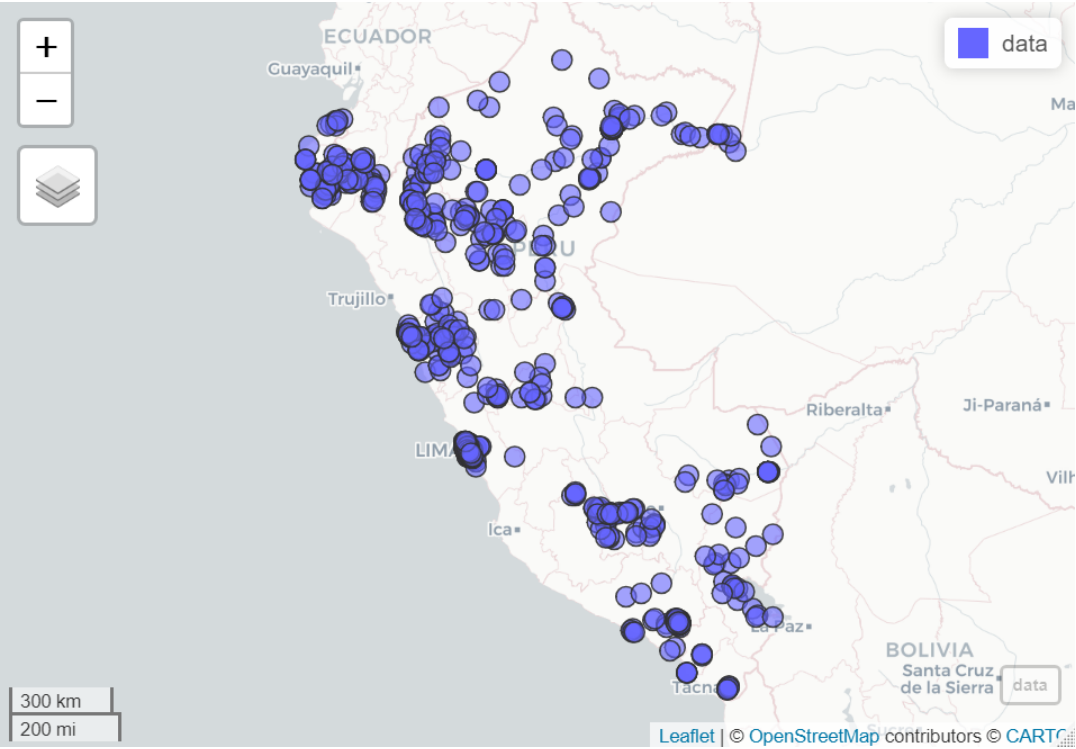


IMAGEN 3.8: Representación geográfica

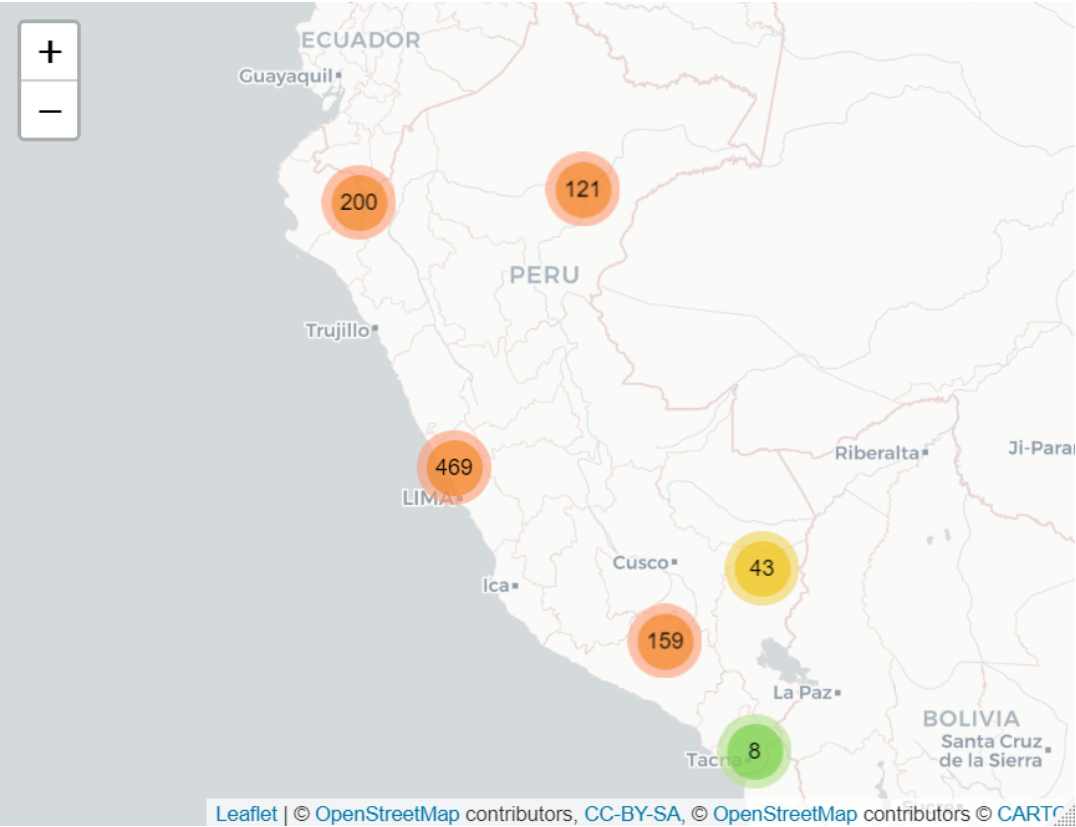


IMAGEN 3.9: Colegios

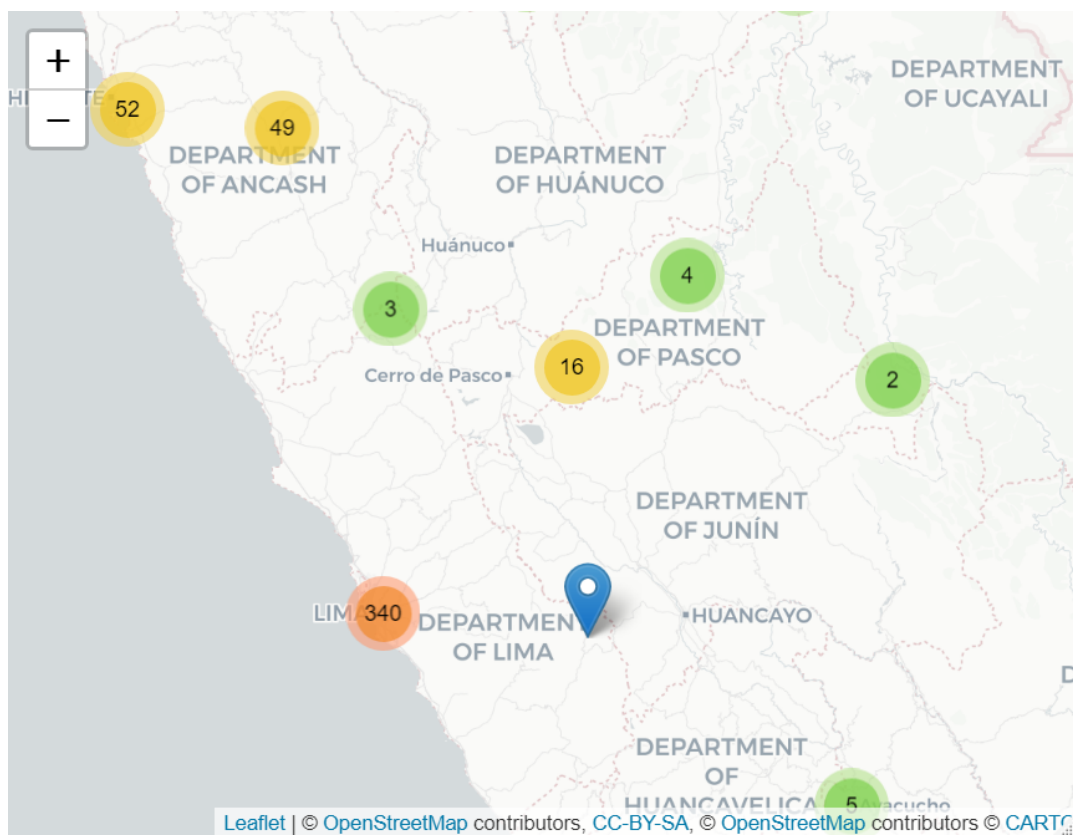


IMAGEN 3.10: Instituciones educativas acumuladas

Capítulo 4

Anexos

4.1 Código de conexión a la base de datos

```

1 library(shiny)
2 library(ggplot2)
3 library(DBI)
4 library(RMySQL)
5 library(reshape2)
6
7 id_host = "dbmysql.ckvpndo3tqd7.us-east-1.rds.amazonaws.com"
8 db <- dbConnect(RMySQL::MySQL(),
9                 dbname = "bd_gestiondatos",
10                 host = id_host,
11                 user = "admin",
12                 password = rstudioapi::askForPassword("DB password"),
13                 Port = 3306)
14
15 df <- dbGetQuery(db, 'SELECT * FROM estudiante')
```

4.2 Niveles educativos

```

1 niveles_educativos = dbGetQuery(db, 'select NIVEL_EDUCATIVO,ID_ANIO,count(
    *)CNT from estudiante group by NIVEL_EDUCATIVO,ID_ANIO')
```

4.3 Evolución mensual de matrículas

```

1 getMonthDay = function(arrayMes) {
2   month = as.Date(arrayMes)
3   fec = format(month, "%Y-%m")
4 }
5
6 getFormattedMonth <- function(arrayMes) {
7   nombre_meses <- c("01"="Enero", "02"="Febrero", "03"="Marzo",
8                     "04"="Abril", "05"="Mayo", "06"="Junio",
9                     "07"="Julio", "08"="Agosto", "09"="Setiembre",
10                    "10"="Octubre", "11"="Noviembre", "12"="Diciembre")
11   mo = nombre_meses[substr(arrayMes, 6, 7)]
12 }
13
14 df2 = df
15
16 df2$month = getMonthDay(df2$FECHA_MATRICULA)
17 df2 = data.frame(df2$FECHA_MATRICULA, df2$ID_PERSONA, df2$month)
18
19 df_graph = aggregate(x = df2.ID_PERSONA ~ df2.month,
```

```

20         data = df2,
21         FUN = length)
22
23 df_graph$df2.month <- factor(df_graph$df2.month, levels = df_graph$df2.
    month, ordered=T)
24
25 df_graph$formated_month = getFormatedMonth(df_graph$df2.month)
26 df_graph$anho = substr(df_graph$df2.month, 1, 4)

```

4.4 Discapacidad

```

1 dfdisc = df[c("ID_ANIO", "DSC_DISCAPACIDAD", "ID_PERSONA")]
2
3 dfdiscFilt <- dfdisc %>% filter(DSC_DISCAPACIDAD != "")
4
5 dfdiscFilt
6
7 grupoDisc <- dfdiscFilt %>%
8   group_by(ID_ANIO, DSC_DISCAPACIDAD) %>%
9   summarise(
10     T_dfdisc = n()
11   )

```

4.5 Nacionalidades

```

1 dfpais <- df[c("ID_ANIO", "NACIONALIDAD", "ID_PERSONA")]
2
3 dfpaisFilt <- dfpais %>%
4   filter(NACIONALIDAD != "Per " & NACIONALIDAD != "PERU" & NACIONALIDAD !
    = " " & NACIONALIDAD != "")
5
6 grupoPais <- dfpaisFilt %>%
7   group_by(ID_ANIO, NACIONALIDAD) %>%
8   summarise(
9     CANTIDAD = n()
10   )

```

4.6 Género

```

1 ValidCampos= df[,c(1, 20, 21)]
2 ValidSexo = ValidCampos
3
4 ValidSexo$SEX0 = substr(ValidSexo$SEX0, 1, 6)
5 ValidSexo <- ValidSexo[!is.na(ValidSexo$VALIDACION_DNI),]
6 gruposSexo <- group_by(ValidSexo, SEX0)
7 agrupSexo = summarise(gruposSexo, num = n())
8
9 agrupSexo2 = agrupSexo %>% as.data.frame()
10 agrupSexo2$porcentaje <- prop.table(agrupSexo2$num)

```

4.7 DNI Validados

```

1 ValidDNI = ValidCampos
2 ValidDNI$SEX0 = substr(ValidDNI$SEX0, 1, 6)
3 ValidDNI <- ValidDNI[!is.na(ValidDNI$VALIDACION_DNI),]
4
5 gruposValDNI <- group_by(ValidDNI, VALIDACION_DNI)

```

```

6 agrupValDNI = summarise(gruposValDNI, num = n())
7 agrupValDNI2 = agrupValDNI %>% as.data.frame()
8 agrupValDNI2$porcentaje <- prop.table(agrupValDNI2$num)

```

4.8 Shiny

```

1 ui <- fluidPage(
2   # theme = bslib::bs_theme(bootswatch = "yeti"),
3   title = "An lisis de estudiantes",
4   titlePanel(
5     # T tulo del app o descripci n
6     "An lisis de estudiantes"
7   ),
8   fluidRow(
9     selectInput(
10      "anho",
11      "A o:",
12      c("2020" = "2020",
13        "2021" = "2021",
14        "2022" = "2022"),
15      selected = "2020"
16    ),
17  ),
18  fluidRow(
19    column(4,
20      plotOutput("histogNivEducativos")
21    ),
22    column(5,
23      plotOutput("histogMatriculas")
24    ),
25    column(3,
26      plotOutput("grupDisc")
27    )
28  ),
29  fluidRow(
30    column(6,
31      "Estudiantes de otra nacionalidad",
32      dataTableOutput("grupPais")
33    ),
34    column(3,
35      plotOutput("histogSexo")
36    ),
37    column(3,
38      plotOutput("histogValidacionDNI")
39    )
40  )
41 )
42
43 server <- function(input, output, session) {
44   # thematic::thematic_shiny()
45   output$histogMatriculas <- renderPlot({
46     # Gr fico
47     subset = df_graph[df_graph$anho == input$anho, ]
48
49     id=factor(subset$formated_month, levels=c("Enero", "Febrero", "Marzo",
50       "Abril", "Mayo", "Junio",
51       "Julio", "Agosto", "
52       Setiembre",
53       "Octubre", "Noviembre", "
54       Diciembre"))

```

```

54   ggplot(subset, aes(x = id, y = df2.ID_PERSONA, label=df2.ID_PERSONA,
55   group = 1)) +
56   geom_line(linetype = 2, colour = "#3B80BD", size = 1) +
57   #geom_point(aes(size = 0.1)) +
58   #geom_point(aes(colour = "red")) +
59   geom_point(colour = "#3B80BD", size = 5) +
60   geom_text(hjust=0, vjust=0) +
61   labs(x = "Meses", y = "# Matr culas",
62   title = "Evoluci n mensual de matr culas") +
63   theme(panel.background = element_blank(),
64   axis.text.y = element_blank(),
65   axis.ticks.y = element_blank())
66   }, res = 96)
67
68   output$histogNivEducativos <- renderPlot({
69   subsetNivEducativos = niveles_educativos[niveles_educativos$ID_ANIO ==
70   input$anho, ]
71
72   ggplot(subsetNivEducativos,
73   aes(x = NIVEL_EDUCATIVO, y = CNT, label=CNT)) +
74   geom_bar(stat = "identity", fill='#3B80BD')+
75   geom_text(hjust=0, vjust=0) +
76   theme(panel.background = element_blank(),
77   axis.text.y = element_blank(),
78   axis.ticks.y = element_blank())+
79   labs(title ='Nivel educativo',
80   x = 'Nivel educativo',
81   y = 'Cantidad',
82   subtitle = 'cantidades por nivel educativo',
83   caption = 'Fuente de datos : Ministerio de educaci n'
84   )
85   }, res = 96)
86
87   output$histogSexo <- renderPlot({
88   # Gr fico
89   ggplot(agrupSexo2, aes(x=2, y=porcentaje, fill=SEX0))+
90   geom_bar(stat = "identity",
91   color="white")+
92   geom_text(aes(label=percent(porcentaje)),
93   position=position_stack(vjust=0.5), color="white", size=3)+
94   coord_polar(theta = "y")+
95   scale_fill_manual(values=c("#E1E1E1", "#3B80BD"))+
96   theme_void()+
97   labs(title="G nero")+
98   xlim(0.5, 2.5)
99   }, res = 96)
100
101   output$histogValidacionDNI <- renderPlot({
102   ggplot(agrupValDNI2, aes(x=2, y=porcentaje, fill=VALIDACION_DNI))+
103   geom_bar(stat = "identity",
104   color="white")+
105   geom_text(aes(label=percent(porcentaje)),
106   position=position_stack(vjust=0.5), color="white", size=3)+
107   coord_polar(theta = "y")+
108   scale_fill_manual(values=c("salmon", "steelblue", "orange", "gray"))+
109   theme_void()+
110   labs(title="DNI Validados")+
111   xlim(0.5, 2.5)
112   }, res = 96)
113
114   output$grupDisc <- renderPlot({
115   subsetDiscapacidad = grupoDisc[grupoDisc$ID_ANIO == input$anho, ]

```

```

115 ggplot(subsetDiscapacidad, aes(x = DSC_DISCAPACIDAD, y=T_dfdisc)) +
116   geom_bar(stat = "identity", fill='#3B80BD') +
117   coord_flip() +
118   labs(title = 'Discapacidad',
119        x = 'DISCAPACIDAD',
120        y = 'CANTIDAD DE ESTUDIANTES',
121        subtitle = 'cantidades por nivel educativo',
122        caption = 'Fuente de datos : Ministerio de educaci n'
123   ) +
124   theme(
125     axis.ticks.x = element_blank(),
126     axis.ticks.y = element_blank(),
127     panel.background = element_blank()
128   )
129 })
130 })
131
132 output$grupPais <- renderDataTable(
133   grupoPais[grupoPais$ID_ANIO == input$anho, ],
134   options = list(pageLength = 10, searching = FALSE)
135 )
136 }
137
138 shinyApp(ui, server)

```

4.9 Representaciones de ubicación de instituciones educativas por mapa

```

1 library(dplyr)
2 library(ggplot2)
3 library(scales)
4 library(tidyverse)
5 library(sf)
6 dfMapas = df
7 dfMapas <- dfMapas[!is.na(dfMapas$NLONG_IE),]
8 # sacar solo una muestra:
9 dfMapMuestra<- dfMapas %>%
10   sample_n(size=1000, replace=FALSE)
11 head(dfMapMuestra)
12 str(dfMapMuestra)
13 points = st_as_sf(dfMapMuestra, coords = c("NLONG_IE", "NLAT_IE"), crs =
14   4326)
15 plot(st_geometry(points), pch=16, col="navy")
16
17 library(mapview)
18 mapview(dfMapMuestra, xcol = "NLONG_IE", ycol = "NLAT_IE", crs = 4269,
19   grid = FALSE)

```

Bibliografía

- [1] P Chapman, J Clinton, and R Kerber. *CRISP-DM 1.0 Step by step guide*. 2000. URL: <ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>.
- [2] RStudio Inc. *Shiny: Web Application Framework for R*. Oct. 2021. URL: <https://cran.r-project.org/web/packages/shiny/index.html>.
- [3] Jimmy Doi et al. *Web Application Teaching Tools for Statistics Using R and Shiny*. 2016. DOI: [10.5070/T591027492](https://doi.org/10.5070/T591027492). URL: <https://escholarship.org/uc/item/00d4q8cp>.
- [4] *shinyapps.io*. May 2022. URL: <https://www.shinyapps.io/>.