**GESTIÓN DE DATOS Y APROXIMACIÓN ESTADÍSTICA**

El análisis estadístico implementado en el presente trabajo fue de tipo descriptivo y exploratorio. Para ello, se hizo uso de diferentes estrategias de tipo gráfico y tabular para aproximarnos al comportamiento observado en las variables de estudio de acuerdo con la tipología de estas. En primer lugar, para el caso de las tendencias/series de tiempo de las poblaciones analizadas, se hizo uso de gráficos de líneas. En segundo lugar, para analizar el comportamiento de las variables desde una perspectiva transversal -en un momento específico del tiempo-, se acudió al uso de varias estrategias de visualización como: gráficos de barras con valores absolutos y relativos, treemaps y tablas de frecuencias. Finalmente, para el caso de variables de naturaleza geográfica, se hizo uso de mapas de cOLOMBIA a nivel departamental y municipal, principalmente. Finalmente, para el caso de variables numéricas, se hizo uso de estadísticos para aproximarnos al comportamiento medio, la dispersión, la simetría, así como el grado de apuntamiento observado en los valores de estas.

Las poblaciones analizadas en este trabajo, al ser de diferentes fuentes, estaban conformadas por estructuras disímiles hecho que implicó un trabajo importante de transformación de datos desde estructuras desordenadas hasta convertirlos en estructura de datos ordenadas -tidy data-. Estandarización de nombres y variables, adaptación y uso de estándares y codificaciones nacionales e internacionales, completitud de variables, manejo de información faltante, cruces de bases de datos, agregación de bases de datos, entre otras estrategias de gestión de datos, fueron empleadas a lo largo del desarrollo del presente trabajo con el fin de organizar los datos en las estructuras requeridas para adelantar los análisis respectivos.

El alto volumen de datos empleados en este trabajo implicó el uso de un software estadístico orientado a la gestión de datos como lo es el lenguaje de programación R y su IDE Rstudio. En el contexto de R, se hizo uso de la base del lenguaje, de varias librerías del entorno tidiverse como: dplyr, tidyr y ggplot2; así mismo, se acudió al uso de algunas librerías orientadas a la importación y exportación de datos como son readxl y writexl, así como a propósitos generales como las librerías scales, cowplot y viridis. Por otra parte, se hizo uso del paquete experimental de R UnalR provisto por la Universidad Nacional de Colombia y orientado a la gestión y visualización de estadísticas e indicadores oficiales. Finalmente, para el caso del estudio de caso de la Universidad Nacional de Colombia, se hizo uso del paquete de R UnalData el cual aloja la información estadística histórica general -más de 15 años- y a nivel de microdatos de las principales poblaciones que hacen parte de esta institución de educación superior pública del país. El anexo 1 contiene el código de R empleado en este trabajo.

Finalmente, con el objetivo de gestionar adecuadamente las versiones de este trabajo, así como su reproducibilidad y disposición en la nube, se hizo uso de la herramienta de control de versiones Git y se construyó un repositorio en la herramienta GitHub para el almacenamiento y disposición del presente trabajo. Este repositorio se encuentra disponible en la siguiente dirección electrónica.

<https://github.com/estadisticaun/Doc_Tatiana/tree/main>

**ANEXO 1**

El presente código/programa de R se encuentra alojado en el repositorio de GitHub empleado para la gestión del presente trabajo**.**

[**https://github.com/estadisticaun/Doc\_Tatiana/blob/main/Scrip%20de%20Analisis.R**](https://github.com/estadisticaun/Doc_Tatiana/blob/main/Scrip%20de%20Analisis.R)

***Programa de R***

##%###############################################%###

#### ANÁLISIS TESIS DE DOCTORADO ####################################################%##

# Librerias ----

library(tidyverse)

library(UnalData)

library(UnalR)

library(viridis)

library(readxl)

library(writexl)

library(scales)

library(cowplot)

# Importar poblaciones ----

Poblacion <- read\_excel("Fuentes/Proy\_Demográfica.xlsx",

sheet = "Demografia",

guess\_max = 1000)

# Importar Base de Datos zonas rurales con

# condiciones de difícil acceso a la educación superior

Mun\_DificilES <- read\_excel("Fuentes/Municipios Dificiles ES.xlsx",

sheet = "Hoja1",

guess\_max = 60000)

# Importar bases de datos - Tendencias

Tendencias <- read\_excel("Fuentes/Asp\_Adm.xlsx",

sheet = "General",

guess\_max = 1000)

Departamentos <- read\_excel("Fuentes/Asp\_Adm.xlsx",

sheet = "Departamentos",

guess\_max = 1000) %>%

mutate(Total = `2010` + `2011` + `2012` + `2013` +

`2014` + `2015` + `2016` + `2017` +

`2018` + `2019` + `2020` + `2021`) %>%

select(c(Poblacion, COD\_DEP, Dpto, Total))

p1721 <- read\_excel("Fuentes/Asp\_Adm.xlsx",

sheet = "Pob17-21",

guess\_max = 1000) %>%

filter(Dpto == "Total Nacional", Ano >= 2010) %>%

select(Year = Ano, Total = Pob1721) %>%

mutate(Poblacion = rep("Población 17 a 21 Años", 11))

# Poblaciones SNIES 2018-2021

Snies1821 <- read\_excel("Fuentes/Asp\_Adm.xlsx",

sheet = "General\_New",

guess\_max = 600000)

# Cruce con municipios dificil acceso a la educación superior

Poblacion <- left\_join(Poblacion, Mun\_DificilES, by = "COD\_MPIO") %>%

mutate(DificilES = ifelse(is.na(DificilES), "No", DificilES))

# Evolución Población General (Serie de Tiempo)

Poblacion\_Total <- Poblacion %>%

summarise(Total = sum(Total), .by = c(YEAR, AREA)) %>%

pivot\_wider(names\_from = c(AREA), values\_from = Total) %>%

rename(Cabecera = `Cabecera Municipal`,

Resto = `Centros Poblados y Rural Disperso`)

# Evolución Población Municipios difíciles ES

Poblacion\_DificilES <- Poblacion %>% filter(AREA != "Total") %>%

summarise(Total = sum(Total), .by = c(YEAR, DificilES)) %>%

pivot\_wider(names\_from = DificilES, values\_from = Total)

# Población General por años

Poblacion\_Total <- left\_join(Poblacion\_Total, Poblacion\_DificilES, by = "YEAR")

# Crear Tabla con tres poblaciones

Tres\_pob <- Tendencias %>% filter(Variable == "Sector") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

pivot\_wider(names\_from = c(Modalidad), values\_from = Total) %>%

mutate(Total = Oficial + Privado,

Poblacion = ifelse(Poblacion == "Mpvez", "Primera Matrícula", Poblacion))

# Parte 1. Análisis Demográfico ----

# Serie General

ggplot(data = Poblacion\_Total, aes(x = YEAR, y = Total, label = Total)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(40000000,60000000))+

ggtitle("Evolución Población Proyectada Para Colombia", subtitle = "Periodo 2020-2035\n")+

ylab("\n Población Proyectada\n ")+

xlab("Año")+

annotate(geom="text", x=2020.5, y=49500000,

label= format(as.numeric(Poblacion\_Total[1, 4]),big.mark=","), color="red")+

annotate(geom="text", x=2034.5, y=57000000,

label=format(as.numeric(Poblacion\_Total[nrow(Poblacion\_Total), 4]),big.mark=","), color="red")+

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Serie General \* Rural y Cabecera

Poblacion\_Grupos <- Poblacion\_Total %>% pivot\_longer(c(Cabecera, Resto) ,names\_to = "Tipo", values\_to = "TotalF")

ggplot(data = Poblacion\_Grupos, aes(x = YEAR, y = TotalF, color = Tipo)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(10000000,45000000))+

ggtitle("Evolución Población Proyectada Para Colombia", subtitle = "Periodo 2020-2035\n")+

ylab("\n Población Proyectada\n ")+

xlab("Año")+

labs(colour = "Ubicación")+

annotate(geom="text", x=2020.7, y=40500000,

label= format(as.numeric(Poblacion\_Grupos[1, 4]),big.mark=","), color="red")+

annotate(geom="text", x=2020.7, y=13500000,

label=format(as.numeric(Poblacion\_Grupos[2, 4]),big.mark=","), color="blue")+

annotate(geom="text", x=2034.5, y=40500000,

label= format(as.numeric(Poblacion\_Grupos[nrow(Poblacion\_Grupos)-1, 4]),big.mark=","), color="red")+

annotate(geom="text", x=2034.5, y=12000000,

label=format(as.numeric(Poblacion\_Grupos[nrow(Poblacion\_Grupos), 4]),big.mark=","), color="blue")+

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Serie General \* Municipios dificil educación superior

Poblacion\_Dificil <- Poblacion\_Total %>% pivot\_longer(c(No, Sí) ,names\_to = "Tipo", values\_to = "TotalF")

ggplot(data = Poblacion\_Dificil, aes(x = YEAR, y = TotalF, color = Tipo)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(10000000,45000000))+

ggtitle("Evolución Población Proyectada Para Municipios con Condiciones \n de Difícil Acceso a la Educación Superior en Colombia\n", subtitle = "Periodo 2020-2035\n")+

ylab("\n Población Proyectada\n ")+

xlab("Año")+

labs(colour = "Ubicación")+

annotate(geom="text", x=2020.7, y=41000000,

label= format(as.numeric(Poblacion\_Dificil[1, 6]),big.mark=","), color="red")+

annotate(geom="text", x=2020.7, y=13500000,

label=format(as.numeric(Poblacion\_Dificil[2, 6]),big.mark=","), color="blue")+

annotate(geom="text", x=2034.3, y=41000000,

label= format(as.numeric(Poblacion\_Dificil[nrow(Poblacion\_Dificil)-1, 6]),big.mark=","), color="red")+

annotate(geom="text", x=2034.3, y=11500000,

label=format(as.numeric(Poblacion\_Dificil[nrow(Poblacion\_Dificil), 6]),big.mark=","), color="blue")+

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Poblacion 2023 por Departamentos

Capital <- c(5001, 8001, 11001, 13001, 15001, 17001, 18001, 19001,

20001, 23001, 27001, 41001, 44001, 47001, 50001, 52001,

54001, 63001, 66001, 68001, 70001, 73001, 76001,81001,

85001, 86001, 88001, 91001, 94001, 95001, 97001, 99001)

Pob\_Dpto\_2023 <- Poblacion %>% filter(YEAR == 2023, AREA == "Total") %>%

mutate(Capital = case\_when(COD\_MPIO %in% Capital ~ "Sí",

TRUE ~ "No")) %>%

summarise(Total = sum(Total), .by = c("COD\_DEP", "DEP", "Capital")) %>%

pivot\_wider(names\_from = Capital, values\_from = Total, values\_fill = 0) %>%

mutate(Total = Sí + No,

Total\_Mil = Total/1000)

# Estadística de poblaciones en capitales y resto

Pob\_Dpto\_2023 %>% summarise(Total\_Capital = sum(Sí),

Total\_Resto = sum(No),

Total\_General = sum(Total))

# Mapa de calor población por Departamentos

Plot.Mapa(

df = Pob\_Dpto\_2023,

depto = COD\_DEP,

variable = Total\_Mil,

agregado = FALSE,

tipo = "Deptos",

titulo = "Población Colombia Proyectada por Departamentos",

cortes = c(0, 100, 500, 2000, 4000, Inf),

colores = c("#FED600", "#02D46E", "#006389", "#FA006E", "red"),

colBorde = "#3A0F2D",

estatico = TRUE,

textSize = 10,

opacidad = 0.6,

estilo = list(

Style = "Intervalo", Theme = 5,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Total de Habitantes", subtitle = "Año 2023\n", caption = "(\*) Por Mil Habitantes")

)

)

# Poblacion 2023 por Municipios

Pob\_Mpio\_2023 <- Poblacion %>%

filter(YEAR == 2023, AREA == "Total") %>%

mutate(Total\_Mil = Total/1000,

Grupos = case\_when(Total <= 10000 ~ "< 10 mil",

between(Total, 10000, 50000) ~ "Entre 10 y 50 mil",

between(Total, 50001, 200000) ~ "Entre 50 y 200 mil",

between(Total, 200000, 1000000) ~ "Entre 200 mil y 1 millón",

TRUE ~ "Más de 1 millón")

)

Pob\_Mpio\_2023 %>% summarise(Total = n(), .by = c(Grupos))

# Mapa de calor población por municipio

Plot.Mapa(

df = Pob\_Mpio\_2023,

mpio = COD\_MPIO,

#variable = Total\_Mil,

variable = Total,

agregado = FALSE,

zoomIslas = TRUE,

tipo = "Mpios",

titulo = "Población Colombia Proyectada por Municipios\nAño 2023",

naTo0 = FALSE,

#centroideMapa = c("ANTIOQUIA"),

#cortes = c(0, 10, 50, 200, 1000, Inf),

cortes = c(0, 10000, 50000, 200000, 1000000, Inf),

colores = c("#ffffcc", "#10F235", "red", "yellow", "blue"),

estatico = TRUE,

estilo = list(

Style = "Intervalo", Theme = 5, anchoBorde = 0.2,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Total de Habitantes"),

Text = list(color = "#011532", size = 0)

)

) -> listMaps

ggdraw() +

draw\_plot(listMaps$M\_COL) +

draw\_plot(listMaps$M\_SanAndres , x = 0.31, y = 0.35, width = 0.060) +

draw\_plot(listMaps$M\_Providencia, x = 0.36, y = 0.38, width = 0.055)

ggsave("X.png", width = 12, height = 12)

# Consolidado Sí/No Población Cabecera vs Resto

Pob\_Resto\_2023 <- Poblacion %>%

filter(YEAR == 2023, AREA != "Total") %>%

pivot\_wider(names\_from = AREA, values\_from = Total) %>%

rename(Cabecera = `Cabecera Municipal`,

Resto = `Centros Poblados y Rural Disperso`) %>%

mutate(Mayor = ifelse(Cabecera < Resto, 1, 0))

# Tabla

table(Pob\_Resto\_2023$Mayor)

# Mapa municipios con Cabecera < Resto

Pob\_Rural\_2023 <- Pob\_Resto\_2023 %>% filter(Mayor == 1)

# Mapa poblaciones rurales

Plot.Mapa(

df = Pob\_Rural\_2023,

depto = COD\_DEP,

mpio = COD\_MPIO,

tipo = "SiNoMpios",

zoomIslas = TRUE,

SiNoLegend = c("Sí", "No"),

# centroideMapa = c("RISARALDA"),

titulo = "Municipios con mayor población en áreas rurales",

colores = c("red", "#10F235"),

opacidad = 1,

textSize = 0,

limpio = FALSE,

estatico = TRUE,

estilo = list(

Style = "SiNo", Theme = 5, anchoBorde = 0.2,

labelX = "",

labelY = "",

Labs = list(fill = "¿Municipio rural?", subtitle = "Año 2023\n"),

Legend = list(legend.position = "bottom",

legend.direction = "vertical")))-> listMaps

ggdraw() +

draw\_plot(listMaps$M\_COL) +

draw\_plot(listMaps$M\_SanAndres , x = 0.31, y = 0.32, width = 0.060) +

draw\_plot(listMaps$M\_Providencia, x = 0.36, y = 0.35, width = 0.055)

# Mapas municipios con condiciones de difícil acceso a la educación superior

# Mapa Sí/No

Poblacion\_Dif\_SiNo <- Poblacion %>% filter(DificilES == "Sí", AREA == "Total", YEAR == 2023)

Plot.Mapa(

df = Poblacion\_Dif\_SiNo,

depto = COD\_DEP,

mpio = COD\_MPIO,

tipo = "SiNoMpios",

zoomIslas = TRUE,

# centroideMapa = c("RISARALDA"),

titulo = "Municipios con condiciones de difícil acceso \na la educación superior",

colores = c("#10F235", "red"),

opacidad = 1,

textSize = 0,

limpio = FALSE,

# centroide = "TOLIMA",

estatico = TRUE,

SiNoLegend = c("Sí", "No"),

estilo = list(

Style = "SiNo", Theme = 5, anchoBorde = 0.2,

labelX = "",

labelY = "",

Labs = list(fill = "¿Acceso dificil?", subtitle = "Año 2023\n"),

Legend = list(legend.position = "bottom",

legend.direction = "horizontal")))-> listMaps

ggdraw() +

draw\_plot(listMaps$M\_COL) +

draw\_plot(listMaps$M\_SanAndres , x = 0.31, y = 0.29, width = 0.060) +

draw\_plot(listMaps$M\_Providencia, x = 0.36, y = 0.32, width = 0.055)

# Mapa calor total de habitantes.

# Parte 2. Transito ----

Transito\_Dpto <- read\_excel("Fuentes/Transito.xlsx",

sheet = "Departamentos",

guess\_max = 50) %>%

mutate(across(.cols = starts\_with("Y"), .fns = ~round(.x\*100, 1)))

Transito\_Mpio <- read\_excel("Fuentes/Transito.xlsx",

sheet = "Municipios",

guess\_max = 50) %>%

mutate(across(.cols = starts\_with("Y"), .fns = ~round(.x\*100, 1))) %>%

mutate(Y2021\_T30 = ifelse(Y2021 <= 30, 1, 0)) %>%

mutate(Capital = case\_when(COD\_MPIO %in% Capital ~ Y2021,

TRUE ~ NA))

Transito\_Capital <- Transito\_Mpio %>% filter(!is.na(Capital)) %>%

select(-c(Y2021\_T30, Capital)) %>%

rename(`2014` = Y2014, `2015` = Y2015, `2016` = Y2016,

`2017` = Y2017, `2018` = Y2018, `2019` = Y2019,

`2020` = Y2020, `2021` = Y2021) %>%

pivot\_longer(cols = c(`2014`:`2021`), names\_to = "YEAR", values\_to = "Tasa") %>%

mutate(YEAR = as.numeric(YEAR))

ggplot(data = Transito\_Capital, aes(x = YEAR, y = Tasa, color = MPO)) +

geom\_point(size = 0)+

geom\_line() +

scale\_y\_continuous(limits = c(0,100))

# Evolución Tasa de Transito Inmediato en Colombia

Evol\_Transito <- Transito\_Dpto %>% filter(DEP == "Colombia") %>%

rename(`2014` = Y2014, `2015` = Y2015, `2016` = Y2016,

`2017` = Y2017, `2018` = Y2018, `2019` = Y2019,

`2020` = Y2020, `2021` = Y2021) %>%

pivot\_longer(cols = `2014`:`2021`, names\_to = "YEAR", values\_to = "Tasa") %>%

mutate(YEAR = as.numeric(YEAR))

ggplot(data = Evol\_Transito, aes(x = YEAR, y = Tasa)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(limits = c(0,100))+

ggtitle("Evolución Tasa de Transito Inmediata en Colombia", subtitle = "Periodo 2014-2021\n")+

ylab("\n Tasa de Transito\n ")+

xlab("Año") +

annotate(geom="text", x=2014, y=42,

label= paste(as.numeric(Evol\_Transito[1, ncol(Evol\_Transito)]), "%"), color="red")+

annotate(geom="text", x=2021, y=44,

label= paste(as.numeric(Evol\_Transito[nrow(Evol\_Transito), ncol(Evol\_Transito)]), "%"), color="red")+

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Tabla transito menor al 30%

table(Transito\_Mpio$Y2021\_T30)

# Mapa de calor por departamentos

Plot.Mapa(

df = Transito\_Dpto,

depto = COD\_DEP,

variable = Y2021,

agregado = FALSE,

tipo = "Deptos",

titulo = "Tasa de Transito Inmmediato por Departamentos",

naTo0 = FALSE,

#centroideMapa = c("ANTIOQUIA"),

cortes = c(0, 30, 40, Inf),

colores = c("red", "yellow", "#10F235"),

estatico = TRUE,

estilo = list(

Style = "Intervalo", Theme = 5, anchoBorde = 0.2,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Tasa de transito (%)", subtitle = "Año 2021\n"),

Text = list(color = "#011532", size = 0)

))

# Mapa de calor por Municipios

Plot.Mapa(

df = Transito\_Mpio,

mpio = COD\_MPIO,

variable = Y2021,

agregado = FALSE,

tipo = "Mpios",

zoomIslas = TRUE,

titulo = "Tasa de Transito Inmmediata por Municipios",

naTo0 = TRUE,

#centroideMapa = c("TOLIMA"),

cortes = c(0, 30, 40, Inf),

colores = c("red", "yellow", "#10F235"),

estatico = TRUE,

estilo = list(

Style = "Intervalo", Theme = 5, anchoBorde = 0.2,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Tasa de transito (%)", subtitle = "Año 2021\n"),

Text = list(color = "#011532", size = 0)

)) -> listMaps

ggdraw() +

draw\_plot(listMaps$M\_COL) +

draw\_plot(listMaps$M\_SanAndres , x = 0.31, y = 0.29, width = 0.060) +

draw\_plot(listMaps$M\_Providencia, x = 0.36, y = 0.32, width = 0.055)

# Mapa de calor por capitales

Plot.Mapa(

df = Transito\_Mpio,

mpio = COD\_MPIO,

variable = Capital,

agregado = FALSE,

tipo = "Mpios",

zoomIslas = TRUE,

titulo = "Tasa de Transito Inmmediata en Capitales de \n Departamentos",

naTo0 = TRUE,

# centroideMapa = c("TOLIMA"),

cortes = c(0, 30, 40, Inf),

colores = c("red", "yellow", "#10F235"),

estatico = TRUE,

estilo = list(

Style = "Intervalo", Theme = 5, anchoBorde = 0.2,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Tasa de transito (%)", subtitle = "Año 2021\n", caption = "NA: No Aplica"),

Text = list(color = "#011532", size = 0)

))-> listMaps

ggdraw() +

draw\_plot(listMaps$M\_COL) +

draw\_plot(listMaps$M\_SanAndres , x = 0.33, y = 0.28, width = 0.060) +

draw\_plot(listMaps$M\_Providencia, x = 0.38, y = 0.31, width = 0.055)

?Plot.Mapa

# Parte 3. Inscritos ----

# Base de datos tendencias por poblaciones

Ten\_Ins <- Tendencias %>% filter(Poblacion == "Aspirantes")

Ten\_Adm <- Tendencias %>% filter(Poblacion == "Admitidos")

Ten\_Mpvez <- Tendencias %>% filter(Poblacion == "Mpvez")

# Base de datos poblaciones por departamentos

Dep\_Ins <- Departamentos %>% filter(Poblacion == "Aspirantes")

Dep\_Adm <- Departamentos %>% filter(Poblacion == "Admitidos")

Dep\_Mpvez <- Departamentos %>% filter(Poblacion == "Mpvez")

# Inicio Gráficos de Tendencias

# Serie Inscritos

Ten\_Ins1 <- Ten\_Ins %>% filter(Variable == "Sector") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

pivot\_wider(names\_from = c(Modalidad), values\_from = Total) %>%

mutate(Total = Oficial + Privado)

ggplot(data = Ten\_Ins1, aes(x = Year, y = Total)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,2500000))+

scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

ggtitle("Evolución Total de Inscritos en Educación Superior en Colombia", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño") +

annotate(geom="text", x=2010.5, y=500000,

label= format(as.numeric(Ten\_Ins1[1, 6]),big.mark=","), color="red") +

annotate(geom="text", x=2020.9, y=2250000,

label= format(as.numeric(Ten\_Ins1[12, 6]),big.mark=","), color="red") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Sector

Ten\_Ins %>% filter(Variable == "Sector") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,1700000))+

scale\_color\_discrete(name = "Sector")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Inscritos en Colombia por Sector", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Sexo

Ten\_Ins %>% filter(Variable == "Sexo") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,1500000))+

scale\_color\_discrete(name = "Sexo")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Inscritos en Colombia por Sexo", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Nivel

Ten\_Ins %>% filter(Variable == "Nivel") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,1100000))+

scale\_color\_discrete(name = "Nivel")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Inscritos en Colombia por Nivel de Formación", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13),

legend.position="bottom")

# Por Área

Ten\_Ins %>% filter(Variable == "Area") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,1000000))+

scale\_color\_discrete(name = "Área")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Inscritos en Colombia por Áreas del Conocimiento", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Departamentos

Plot.Mapa(

df = Dep\_Ins,

depto = COD\_DEP,

variable = Total,

agregado = FALSE,

tipo = "Deptos",

titulo = "Total de Inscritos en Colombia por Departamentos",

cortes = c(0, 10000, 100000, 500000, 1000000, Inf),

colores = c("#FED600", "#02D46E", "#006389", "#FA006E", "red"),

colBorde = "#3A0F2D",

estatico = TRUE,

textSize = 10,

opacidad = 0.6,

estilo = list(

Style = "Intervalo", Theme = 5,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Total Aspirantes", subtitle = "Años 2010 a 2021\n")

)

)

# Parte 4. Admitidos ----

# Inicio Gráficos de Tendencias

# Serie Admitidos

Ten\_Adm1 <- Ten\_Adm %>% filter(Variable == "Sector") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

pivot\_wider(names\_from = c(Modalidad), values\_from = Total) %>%

mutate(Total = Oficial + Privado)

ggplot(data = Ten\_Adm1, aes(x = Year, y = Total)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,1500000))+

scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

ggtitle("Evolución Total de Admitidos en Educación Superior en Colombia", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño") +

annotate(geom="text", x=2010.1, y=270000,

label= format(as.numeric(Ten\_Adm1[1, 6]),big.mark=","), color="red") +

annotate(geom="text", x=2020.9, y=1250000,

label= format(as.numeric(Ten\_Adm1[12, 6]),big.mark=","), color="red") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Sector

Ten\_Adm %>% filter(Variable == "Sector") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,750000))+

scale\_color\_discrete(name = "Sector")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Admitidos en Colombia por Sector", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Sexo

Ten\_Adm %>% filter(Variable == "Sexo") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,750000))+

scale\_color\_discrete(name = "Sexo")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Admitidos en Colombia por Sexo", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Nivel

Ten\_Adm %>% filter(Variable == "Nivel") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,650000))+

scale\_color\_discrete(name = "Nivel")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Admitidos en Colombia por Nivel de Fomración", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13),

legend.position="bottom")

# Por Área

Ten\_Adm %>% filter(Variable == "Area") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,500000))+

scale\_color\_discrete(name = "Área")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Admitidos en Colombia por Áreas del Conocimiento", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Departamentos

Plot.Mapa(

df = Dep\_Adm,

depto = COD\_DEP,

variable = Total,

agregado = FALSE,

tipo = "Deptos",

titulo = "Total de Admitidos en Colombia por Departamentos",

cortes = c(0, 5000, 50000, 200000, 300000, Inf),

colores = c("#FED600", "#02D46E", "#006389", "#FA006E", "red"),

colBorde = "#3A0F2D",

estatico = TRUE,

textSize = 10,

opacidad = 0.6,

estilo = list(

Style = "Intervalo", Theme = 5,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Total Admitidos", subtitle = "Años 2010 a 2021\n")

)

)

# Parte 5. Mpvz ----

# Inicio Gráficos de Tendencias

# Serie Primera Matricula

Ten\_Mpvez1 <- Ten\_Mpvez %>% filter(Variable == "Sector") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

pivot\_wider(names\_from = c(Modalidad), values\_from = Total) %>%

mutate(Total = Oficial + Privado)

ggplot(data = Ten\_Mpvez1, aes(x = Year, y = Total)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,1100000))+

scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

ggtitle("Evolución Total de Matriculados en Primer Curso en Colombia", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total Matriculados\n ")+

xlab("\nAño") +

annotate(geom="text", x=2010.1, y=200000,

label= format(as.numeric(Ten\_Mpvez1[1, 6]),big.mark=","), color="red") +

annotate(geom="text", x=2020.9, y=980000,

label= format(as.numeric(Ten\_Mpvez1[12, 6]),big.mark=","), color="red") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Sector

Ten\_Mpvez %>% filter(Variable == "Sector") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,750000))+

scale\_color\_discrete(name = "Sector")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Matriculados en Primer Curso en Colombia por Sector", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Matriculados en Primer Curso\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Sexo

Ten\_Mpvez %>% filter(Variable == "Sexo") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,600000))+

scale\_color\_discrete(name = "Sexo")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Matriculados en Primer Curso en Colombia por Sexo", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Matriculados en Primer Curso\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Nivel

Ten\_Mpvez %>% filter(Variable == "Nivel") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,550000))+

scale\_color\_discrete(name = "Nivel")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Matriculados en Primer Curso en Colombia \npor Nivel de Formación", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Matriculados en Primer Curso\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13),

legend.position="bottom")

# Por Área

Ten\_Mpvez %>% filter(Variable == "Area") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year)) %>%

ggplot(aes(x = Year, y = Total, color = Modalidad)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,400000))+

scale\_color\_discrete(name = "Área")+

# labs(colour = "Sector")+

ggtitle("Evolución Total de Matriculados en Primer Curso en Colombia \npor Áreas del Conocimiento", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total de Matriculados en Primer Curso\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Por Departamentos

Plot.Mapa(

df = Dep\_Mpvez,

depto = COD\_DEP,

variable = Total,

agregado = FALSE,

tipo = "Deptos",

titulo = "Total de Matriculados en Primer Curso \nen Colombia por Departamentos",

cortes = c(0, 5000, 50000, 200000, 300000, Inf),

colores = c("#FED600", "#02D46E", "#006389", "#FA006E", "red"),

colBorde = "#3A0F2D",

estatico = TRUE,

textSize = 10,

opacidad = 0.6,

estilo = list(

Style = "Intervalo", Theme = 5,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Total Matriculados en Primer Curso", subtitle = "Años 2010 a 2021\n")

)

)

# Parte 6. Mpvz - Universitaria ----

Mpvez\_Univer <- Tendencias %>% filter(Poblacion == "Mpvez",

Modalidad == "Universitario") %>%

pivot\_longer(c(`2010`:`2021`) ,names\_to = "Year", values\_to = "Total") %>%

mutate(Year = as.numeric(Year),

Poblacion = ifelse(Poblacion == "Mpvez", "Primera Matrícula Universitaria", Poblacion)) %>%

select(Poblacion, Year, Total)

ggplot(data = Mpvez\_Univer, aes(x = Year, y = Total)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,600000))+

scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

ggtitle("Evolución Total de Matriculados en Primer Curso de Universidad en Colombia", subtitle = "Periodo 2010-2021\n")+

ylab("\n Total Matriculados \n ")+

xlab("\nAño") +

annotate(geom="text", x=2010, y=130000,

label= format(as.numeric(Mpvez\_Univer[1, 3]),big.mark=","), color="red") +

annotate(geom="text", x=2020.9, y=470000,

label= format(as.numeric(Mpvez\_Univer[12, 3]),big.mark=","), color="red") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Parte 7. Asp-Adm-Mpvz ----

# Cruzar Poblaciones

Tres\_pob <- bind\_rows(Tres\_pob, Mpvez\_Univer)

# Gráfico de tres poblaciones

Tres\_pob %>%

ggplot(aes(x = Year, y = Total, color = Poblacion)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma)+

scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

labs(title = "Evolución Poblaciones en Educación Superior en Colombia",

subtitle = "Periodo 2010-2021\n",

color = "Población",

x = "\nAño",

y = "\n Total \n")+

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13),

legend.position="bottom",

legend.title = element\_text("Población"))

# Tasa de Absorción en Educación Superior

Tasa\_absorcion <- Tres\_pob %>% select(Poblacion, Year, Total) %>%

filter(Poblacion %in% c('Aspirantes', "Admitidos", "Primera Matrícula", "Primera Matrícula Universitaria")) %>%

pivot\_wider(names\_from = Poblacion, values\_from = Total) %>%

mutate(Admitidos = Admitidos / Aspirantes,

`Primera Matrícula` = `Primera Matrícula` / Aspirantes ,

`Primera Matrícula Universitaria` = `Primera Matrícula Universitaria` / Aspirantes ) %>%

select(-Aspirantes) %>%

pivot\_longer(c(Admitidos, `Primera Matrícula`, `Primera Matrícula Universitaria`),

names\_to = "Poblacion",

values\_to = "Total")

ggplot(data = Tasa\_absorcion, aes(x = Year, y = Total, color = Poblacion)) +

geom\_point(size = 3)+

geom\_line()+

labs(title = "Evolución Tasa de Absorción en Educación Superior en Colombia",

subtitle = "Periodo 2010-2021\n",

color = "Población",

x = "\nAño",

y = "\n Tasa de Absorción \n")+

scale\_y\_continuous(limits = c(0, 1),

labels = scales::percent)+

scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

annotate(geom="text", x=2010, y=0.61,

label= scales::percent(as.numeric(Tasa\_absorcion[1, 3])),

color="red") +

annotate(geom="text", x=2010, y=0.37,

label= scales::percent(as.numeric(Tasa\_absorcion[2, 3])),

color="#2d572c") +

annotate(geom="text", x=2010, y=0.20,

label= scales::percent(as.numeric(Tasa\_absorcion[3, 3])),

color="blue") +

annotate(geom="text", x=2021, y=0.61,

label= scales::percent(as.numeric(Tasa\_absorcion[34, 3])),

color="red") +

annotate(geom="text", x=2021, y=0.37,

label= scales::percent(as.numeric(Tasa\_absorcion[35, 3])),

color="#2d572c") +

annotate(geom="text", x=2021, y=0.15,

label= scales::percent(as.numeric(Tasa\_absorcion[36, 3])),

color="blue") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13),

legend.position="bottom",

legend.title = element\_text("Población"))

# Parte 8. Universidad 18-21 ----

# Base de datos de población universitaria

Universidad <- Snies1821 %>% filter(Formacion %in% c("Universitaria", "UNIVERSITARIA"))

Aspirantes <- Universidad %>% filter(Poblacion == "Inscritos")

Admitidos <- Universidad %>% filter(Poblacion == "Admitidos")

Mpvez <- Universidad %>% filter(Poblacion == "Mpvez")

# Aspirantes ----

# Serie de tiempo general

Ten\_Ins <- Aspirantes %>% summarise(Total = sum(Total), .by = c(Ano))

ggplot(data= Ten\_Ins, aes(x = Ano, y = Total)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,1200000))+

# scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

ggtitle("Evolución Total de Inscritos Formación Universitaria en Colombia", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño")+

annotate(geom="text", x=2018.1, y=1100000,

label= format(as.numeric(Ten\_Ins[1, 2]),big.mark=","), color="red") +

annotate(geom="text", x=2020.95, y=1100000,

label= format(as.numeric(Ten\_Ins[4, 2]),big.mark=","), color="red") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Sector

Aspirantes %>%

rename(Sector = `Sector IES`) %>%

summarise(Total = sum(Total), .by = c(Ano, Sector)) %>%

ggplot(aes(x = Ano, y = Total, color = Sector)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,650000))+

scale\_color\_discrete(name = "Sector")+

ggtitle("Evolución Inscritos en Formación Universitaria en Colombia por Sector", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Sexo

Aspirantes %>%

summarise(Total = sum(Total), .by = c(Ano, Sexo)) %>%

ggplot(aes(x = Ano, y = Total, color = Sexo)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,650000))+

scale\_color\_discrete(name = "Sexo")+

ggtitle("Evolución Inscritos en Formación Universitaria en Colombia por Sexo", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Metodología

Aspirantes %>%

summarise(Total = sum(Total), .by = c(Ano, Metodologia)) %>%

ggplot(aes(x = Ano, y = Total, color = Metodologia)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,1000000))+

scale\_color\_discrete(name = "Metodología")+

ggtitle("Evolución Inscritos en Formación Universitaria en Colombia por Metodología", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total de Inscritos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Admitidos ----

# Serie de tiempo general

Ten\_Adm <- Admitidos %>% summarise(Total = sum(Total), .by = c(Ano))

ggplot(data= Ten\_Adm, aes(x = Ano, y = Total)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,650000))+

# scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

ggtitle("Evolución Total de Admitidos a Formación Universitaria en Colombia", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño")+

annotate(geom="text", x=2018.1, y=590000,

label= format(as.numeric(Ten\_Adm[1, 2]),big.mark=","), color="red") +

annotate(geom="text", x=2020.95, y=630000,

label= format(as.numeric(Ten\_Adm[4, 2]),big.mark=","), color="red") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Sector

Admitidos %>%

rename(Sector = `Sector IES`) %>%

summarise(Total = sum(Total), .by = c(Ano, Sector)) %>%

ggplot(aes(x = Ano, y = Total, color = Sector)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,500000))+

scale\_color\_discrete(name = "Sector")+

ggtitle("Evolución Admitidos en Formación Universitaria en Colombia por Sector", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Sexo

Admitidos %>%

mutate(Sexo1 = case\_when(Sexo == "Hombre"~ "Hombre",

Sexo == "Mujer"~ "Mujer",

Sexo == "HOMBRE"~ "Hombre",

Sexo == "MUJER"~ "Mujer",

Sexo == "Sin Información" ~ "Sin Información")) %>%

summarise(Total = sum(Total), .by = c(Ano, Sexo1)) %>%

ggplot(aes(x = Ano, y = Total, color = Sexo1)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,450000))+

scale\_color\_discrete(name = "Sexo")+

ggtitle("Evolución Admitidos en Formación Universitaria en Colombia por Sexo", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Metodología

Admitidos %>%

mutate(Metodologia1 = case\_when(Metodologia == "Presencial" ~ "Presencial",

Metodologia == "PRESENCIAL" ~ "Presencial",

Metodologia == "Presencial-Dual" ~ "Presencial-Dual",

Metodologia == "Presencial-Virtual" ~ "Presencial-Virtual",

Metodologia == "Distancia (tradicional)" ~ "Distancia (tradicional)",

Metodologia == "DISTANCIA (TRADICIONAL)" ~ "Distancia (tradicional)",

Metodologia == "Distancia (virtual)" ~ "Distancia (virtual)",

Metodologia == "DISTANCIA (VIRTUAL)" ~ "Distancia (virtual)")) %>%

summarise(Total = sum(Total), .by = c(Ano, Metodologia1)) %>%

ggplot(aes(x = Ano, y = Total, color = Metodologia1)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,600000))+

scale\_color\_discrete(name = "Metodología")+

ggtitle("Evolución Admitidos en Formación Universitaria en Colombia por Metodología", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total de Admitidos\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Mpvez ----

# Serie de tiempo general

Ten\_Mpvez <- Mpvez %>% summarise(Total = sum(Total), .by = c(Ano))

ggplot(data= Ten\_Mpvez, aes(x = Ano, y = Total)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,650000))+

# scale\_x\_continuous(breaks = c(2010, 2013, 2016, 2019, 2021))+

ggtitle("Evolución Total de Matriculados Primera Vez en Formación\nUniversitaria en Colombia", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total Matriculados\n ")+

xlab("\nAño")+

annotate(geom="text", x=2018.1, y=470000,

label= format(as.numeric(Ten\_Mpvez[1, 2]),big.mark=","), color="red") +

annotate(geom="text", x=2020.95, y=480000,

label= format(as.numeric(Ten\_Mpvez[4, 2]),big.mark=","), color="red") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Sector

Mpvez %>%

rename(Sector = `Sector IES`) %>%

summarise(Total = sum(Total), .by = c(Ano, Sector)) %>%

ggplot(aes(x = Ano, y = Total, color = Sector)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,500000))+

scale\_color\_discrete(name = "Sector")+

ggtitle("Evolución Matriculados Primera Vez en Formación\nUniversitaria en Colombia por Sector", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total Matriculados\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Sexo

Mpvez %>%

mutate(Sexo1 = case\_when(ID\_Sexo == 1 ~ "Hombre",

ID\_Sexo == 2 ~ "Mujer",

ID\_Sexo == 0 ~ "Sin Información")) %>%

summarise(Total = sum(Total), .by = c(Ano, Sexo1)) %>%

ggplot(aes(x = Ano, y = Total, color = Sexo1)) +

geom\_point(size = 3)+

geom\_line()+

scale\_y\_continuous(labels = comma, limits = c(0,450000))+

scale\_color\_discrete(name = "Sexo")+

ggtitle("Evolución Matriculados Primera Vez en Formación\nUniversitaria en Colombia por Sexo", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total Matriculados\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Metodología

Mpvez %>%

mutate(Metodologia1 = case\_when(Metodologia == "Presencial" ~ "Presencial",

Metodologia == "PRESENCIAL" ~ "Presencial",

Metodologia == "Presencial-Dual" ~ "Presencial-Dual",

Metodologia == "Presencial-Virtual" ~ "Presencial-Virtual",

Metodologia == "Distancia (tradicional)" ~ "Distancia (tradicional)",

Metodologia == "DISTANCIA (TRADICIONAL)" ~ "Distancia (tradicional)",

Metodologia == "Distancia (virtual)" ~ "Distancia (virtual)",

Metodologia == "DISTANCIA (VIRTUAL)" ~ "Distancia (virtual)")) %>%

summarise(Total = sum(Total), .by = c(Ano, Metodologia1)) %>%

ggplot(aes(x = Ano, y = Total, color = Metodologia1)) +

geom\_point(size = 3)+

geom\_line() +

scale\_y\_continuous(labels = comma, limits = c(0,450000))+

scale\_color\_discrete(name = "Metodología")+

ggtitle("Evolución Matriculados Primera Vez en Formación\nUniversitaria en Colombia por Metodología", subtitle = "Periodo 2018-2021\n")+

ylab("\n Total Matriculados\n ")+

xlab("\nAño") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13))

# Global ----

Universidad1 <- Universidad %>% summarise(Total = sum(Total), .by = c(Poblacion, Ano))

ggplot(data = Universidad1, aes(x = Ano, y = Total, color = Poblacion))+

geom\_line()+

geom\_point()+

labs(title = "Evolución Total Inscripciones, Admisiones y Matriculas\nPrimera Vez Formación Universitaria en Colombia",

subtitle = "Periodo 2018-2021\n",

color = "Población",

x = "\nAño",

y = "\n Total de Individuos \n")+

scale\_y\_continuous(labels = comma, limits = c(0,1200000))+

scale\_color\_discrete(breaks = c("Inscritos", "Admitidos", "Mpvez"),

labels = c("Inscritos", "Admitidos", "Matriculados Primera Vez"))+

annotate(geom="text", x=2018.1, y=1050000,

label= format(as.numeric(Universidad1[1, 3]),big.mark=","), color="red") +

annotate(geom="text", x=2020.9, y=1050000,

label= format(as.numeric(Universidad1[4, 3]),big.mark=","), color="red") +

annotate(geom="text", x=2018.1, y=600000,

label= format(as.numeric(Universidad1[5, 3]),big.mark=","), color="red") +

annotate(geom="text", x=2020.9, y=650000,

label= format(as.numeric(Universidad1[8, 3]),big.mark=","), color="red") +

annotate(geom="text", x=2018.1, y=350000,

label= format(as.numeric(Universidad1[9, 3]),big.mark=","), color="red") +

annotate(geom="text", x=2020.9, y=350000,

label= format(as.numeric(Universidad1[12, 3]),big.mark=","), color="red") +

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13),

legend.position="bottom",

legend.title = element\_text("Población", face="bold"),

legend.text = element\_text(size = 12))

# Contexto Público vs Privado

Universidad2 <- Universidad %>% rename(Sector = `Sector IES`) %>%

summarise(Total = sum(Total), .by = c(Sector, Poblacion, Ano)) %>%

mutate(Sector1 = case\_when(Sector == "Oficial" ~ "Oficial",

Sector == "OFICIAL" ~ "Oficial",

Sector == "Privada" ~ "Privada",

Sector == "PRIVADA" ~ "Privada"))

# Textos de las facetas

Text\_Uni2 <- Universidad2 %>% filter(Ano %in% c(2018, 2021)) %>%

mutate(Year = ifelse(Ano == 2018, 2017.5, 2021.5))

# Gráfico

Uni2 <- ggplot(data = Universidad2, aes(x = Ano, y = Total, color = Poblacion))+

geom\_line()+

geom\_point()+

labs(title = "Evolución Total Inscripciones, Admisiones y Matriculas Primera Vez Formación\nUniversitaria en Colombia por Sector de la Educación",

subtitle = "Periodo 2018-2021\n",

color = "Población",

x = "\nAño",

y = "\n Total de Individuos \n")+

xlim(2017, 2022)+

# scale\_x\_continuous(breaks = c(2017, 2018, 2019, 2020, 2021, 2022))+

scale\_y\_continuous(labels = comma, limits = c(0,750000))+

scale\_color\_discrete(breaks = c("Inscritos", "Admitidos", "Mpvez"),

labels = c("Inscritos", "Admitidos", "Matriculados Primera Vez"))+

theme(axis.text.y = element\_text(size = 10, face = "bold"),

axis.text.x = element\_text(size = 12, colour = "blue"),

axis.title = element\_text(face="bold", color="black", size=13),

legend.position="bottom",

legend.title = element\_text("Población", size = 12, face="bold"),

legend.text = element\_text(size = 12),

strip.text = element\_text(size = 12))+

facet\_wrap(vars(Sector1))

Uni2 + geom\_text(data = Text\_Uni2,

aes(x = Year, y = Total,

label = format(Total, big.mark=",")))

# Análisis IES ----

# Crear bases de datos

IES <- Snies1821 %>% filter(Poblacion == "Mpvez", Formacion %in% c("Universitaria", "UNIVERSITARIA")) %>%

select(COD\_INTS:Municipio\_IES, Total) %>%

mutate(Caracter = case\_when(Carácter == "Universidad" ~ "Universidad",

Carácter == "Institución Tecnológica" ~ "Institución Tecnológica",

Carácter == "Institución Universitaria/Escuela Tecnológica" ~ "Institución Universitaria/Escuela Tecnológica",

Carácter == "Institución Técnica Profesional" ~ "Institución Técnica Profesional",

Carácter == "UNIVERSIDAD" ~ "Universidad",

Carácter == "INSTITUCIÓN TECNOLÓGICA" ~ "Institución Tecnológica",

Carácter == "INSTITUCIÓN TÉCNICA PROFESIONAL" ~ "Institución Técnica Profesional",

Carácter == "INSTITUCIÓN UNIVERSITARIA/ESCUELA TECNOLÓGICA" ~ "Institución Universitaria/Escuela Tecnológica"

)) %>%

summarise(Institucion = max( Institucion),

COD\_INTS = max(COD\_INTS),

Principal = min(Principal),

ID\_Sector\_IES = max(ID\_Sector\_IES),

`Sector IES` = max(`Sector IES`),

Acredita = max(Acredita, na.rm = TRUE),

ID\_Caracter = max(ID\_Caracter),

Caracter = max(Caracter),

Total = sum(Total),

.by = c(IES\_PADRE))

IES\_Seccional <- Snies1821 %>% filter(Poblacion == "Mpvez", Formacion %in% c("Universitaria", "UNIVERSITARIA")) %>%

select(COD\_INTS:Municipio\_IES, Total) %>%

summarise(Institucion = max( Institucion),

COD\_DEP\_IES = max(COD\_DEP\_IES),

DPTO\_IES = max(DPTO\_IES),

COD\_MPIO\_IES = max(COD\_MPIO\_IES),

Municipio\_IES = max(Municipio\_IES),

.by = c(IES\_PADRE, COD\_INTS))

# Tablas

# IES por Sector

IES\_Sector <- IES %>% summarise(Total = n(), .by = c(`Sector IES`))

IES\_Sector

IES\_Sector\_Mat <- IES %>% summarise(Total = sum(Total), .by = c(`Sector IES`))

IES\_Sector\_Mat

IES\_Acredita <- IES %>% summarise(Total = n(), .by = c(Acredita))

IES\_Acredita

IES\_Acredita\_Mat <- IES %>% summarise(Total = sum(Total), .by = c(Acredita))

IES\_Acredita\_Mat

IES\_Caracter <- IES %>% summarise(Total = n(), .by = c(Caracter))

IES\_Caracter

IES\_Caracter\_Mat <- IES %>% summarise(Total = sum(Total), .by = c(Caracter))

IES\_Caracter\_Mat

IES\_Caracter\_Sector <- IES %>% summarise(Total = n(), .by = c(`Sector IES`, Caracter))

IES\_Caracter\_Sector

# Gráficos

# Mapas IES con seccionales

IES\_Seccional\_Map <- IES\_Seccional %>%

summarise(Total = n(), .by = c(COD\_DEP\_IES)) %>%

add\_row(COD\_DEP\_IES = 81, Total=0) %>%

add\_row(COD\_DEP\_IES = 99, Total=0) %>%

add\_row(COD\_DEP\_IES = 94, Total=0) %>%

add\_row(COD\_DEP\_IES = 95, Total=0) %>%

add\_row(COD\_DEP\_IES = 97, Total=0) %>%

add\_row(COD\_DEP\_IES = 91, Total=0) %>%

add\_row(COD\_DEP\_IES = 88, Total=0)

Plot.Mapa(

df = IES\_Seccional\_Map,

depto = COD\_DEP\_IES,

variable = Total,

agregado = FALSE,

tipo = "Deptos",

naTo0 = TRUE,

zoomIslas = TRUE,

titulo = "Total de Instituciones de Educación Superior\n(IES) en Colombia por Departamentos\n",

cortes = c(-1, 0, 1, 10, 50, Inf),

colores = c("#FED600", "#02D46E", "#006389", "#FA006E", "red"),

colBorde = "#3A0F2D",

estatico = TRUE,

textSize = 10,

opacidad = 0.6,

estilo = list(

Style = "Intervalo", Theme = 5,

labelX = "",

labelY = "",

Legend = list(legend.position = "bottom", legend.direction = "horizontal"),

Labs = list(fill = "Total IES")

)

)-> listMaps

ggdraw() +

draw\_plot(listMaps$M\_COL) +

draw\_plot(listMaps$M\_SanAndres, x = 0.32, y = 0.735, width = 0.4)

# Caracter de las IES

Car\_IES <- IES %>% summarise(Total = n(), .by = c(Caracter)) %>%

rename(Clase = Caracter) %>%

mutate(Variable = "CARACTER",

Clase = factor(Clase))

Gra1 <- Plot.Barras(

datos = Car\_IES,

categoria = "CARACTER",

estatico = TRUE,

vertical = FALSE,

freqRelativa = TRUE,

ordinal = FALSE,

titulo = "Distribución Porcentual Total IES por tipología de las instituciones",

labelEje = "Porcentaje",

colores = RColorBrewer::brewer.pal(4, "Set1"),

estilo = list(gg.Tema = 5,

gg.Bar = list(width = 0.5, color = "#000000")))

Gra1 + scale\_y\_continuous(limits = c(NA, 100))+

theme(plot.title=element\_text(hjust=0, size=12),

axis.title=element\_text(size=14),

plot.caption=element\_text(size=7),

legend.text=element\_text(size=9),

axis.text = element\_text(size=10))

# Caracter de las IES - Matriculados

Car\_IES\_Mat <- IES %>% summarise(Total = sum(Total), .by = c(Caracter)) %>%

rename(Clase = Caracter) %>%

mutate(Variable = "CARACTER",

Clase = factor(Clase))

Gra1 <- Plot.Barras(

datos = Car\_IES\_Mat,

categoria = "CARACTER",

estatico = TRUE,

vertical = FALSE,

freqRelativa = TRUE,

ordinal = FALSE,

titulo = "Distribución Total Matriculados Primera Vez en las IES\npor Tipología de las Instituciones",

labelEje = "Porcentaje",

colores = RColorBrewer::brewer.pal(4, "Set1"),

estilo = list(gg.Tema = 5,

gg.Bar = list(width = 0.5, color = "#000000")))

Gra1 + scale\_y\_continuous(limits = c(NA, 100))+

theme(plot.title=element\_text(hjust=0, size=12),

axis.title=element\_text(size=14),

plot.caption=element\_text(size=7),

legend.text=element\_text(size=9),

axis.text = element\_text(size=10))

names(IES)

# Análisis Programas ----

PROGRAMAS <- Snies1821 %>% filter(Poblacion == "Mpvez", Formacion %in% c("Universitaria", "UNIVERSITARIA")) %>%

select(Institucion, SNIES\_PROGRA:Prog\_Acredita, ID\_Metodologia:NBC, COD\_DEP\_PRO:MUN\_PRO, Total) %>%

filter(!is.na(SNIES\_PROGRA)) %>%

summarise(Programa = max(Programa),

Institucion = max(Institucion),

Prog\_Acredita = max(Prog\_Acredita, na.rm = TRUE),

ID\_Metodologia = max(ID\_Metodologia),

Metodologia = max(Metodologia),

ID\_Area\_Con = max(ID\_Area\_Con),

Area\_Con = max(Area\_Con),

Id\_Nucleo = max(Id\_Nucleo),

NBC = max(NBC),

COD\_DEP\_PRO = max(COD\_DEP\_PRO),

DEP\_Programa = max(DEP\_Programa),

COD\_MUN\_PRO = max(COD\_MUN\_PRO),

MUN\_PRO = max(MUN\_PRO),

Total = sum(Total),

.by = c(SNIES\_PROGRA)) %>%

mutate(Prog\_Acredita = case\_when(Prog\_Acredita == "SI" ~ "Sí",

Prog\_Acredita == "NO" ~ "No",

is.na(Prog\_Acredita) == TRUE ~ "Sin Información"),

Metodologia = case\_when(Metodologia == "Presencial" ~ "Presencial",

Metodologia == "PRESENCIAL" ~ "Presencial",

Metodologia == "Presencial-Dual" ~ "Presencial-Dual",

Metodologia == "Presencial-Virtual" ~ "Presencial-Virtual",

Metodologia == "Distancia (tradicional)" ~ "Distancia (tradicional)",

Metodologia == "DISTANCIA (TRADICIONAL)" ~ "Distancia (tradicional)",

Metodologia == "Distancia (virtual)" ~ "Distancia (virtual)",

Metodologia == "DISTANCIA (VIRTUAL)" ~ "Distancia (virtual)"),

Area\_Con = case\_when(Area\_Con == "INGENIERÍA, ARQUITECTURA, URBANISMO Y AFINES" ~ "Ingeniería, arquitectura, urbanismo y afines",

Area\_Con == "AGRONOMÍA, VETERINARIA Y AFINES" ~ "Agronomía, veterinaria y afines",

Area\_Con == "BELLAS ARTES" ~ "Bellas artes",

Area\_Con == "CIENCIAS DE LA SALUD" ~ "Ciencias de la salud",

Area\_Con == "CIENCIAS SOCIALES Y HUMANAS" ~ "Ciencias sociales y humanas",

Area\_Con == "ECONOMÍA, ADMINISTRACIÓN, CONTADURÍA Y AFINES" ~ "Economía, administración, contaduría y afines",

Area\_Con == "MATEMÁTICAS Y CIENCIAS NATURALES" ~ "Matemáticas y ciencias naturales",

Area\_Con == "CIENCIAS DE LA EDUCACIÓN" ~ "Ciencias de la educación",

Area\_Con == "Matemáticas y ciencias naturales" ~ "Matemáticas y ciencias naturales",

Area\_Con == "Ciencias de la educación" ~ "Ciencias de la educación",

Area\_Con == "Bellas artes" ~ "Bellas artes",

Area\_Con == "Economía, administración, contaduría y afines" ~ "Economía, administración, contaduría y afines",

Area\_Con == "Ingeniería, arquitectura, urbanismo y afines" ~ "Ingeniería, arquitectura, urbanismo y afines",

Area\_Con == "Ciencias sociales y humanas" ~ "Ciencias sociales y humanas",

Area\_Con == "Agronomía, veterinaria y afines" ~ "Agronomía, veterinaria y afines",

Area\_Con == "Ciencias de la salud" ~ "Ciencias de la salud",

Area\_Con == "Sin clasificar" ~ "Sin clasificar",

Area\_Con == "No Aplica" ~ "Sin clasificar"),

NBC = str\_to\_sentence(NBC))

# Tabla Total Matriculados Pvez 18-21

sum(PROGRAMAS$Total)

# Tabla programas acreditados

Progra\_Acredita <- PROGRAMAS %>% summarise(Total = n(), .by = c(Prog\_Acredita))

Progra\_Acredita

Progra\_Acredita\_Mat <- PROGRAMAS %>% summarise(Total = sum(Total), .by = c(Prog\_Acredita))

Progra\_Acredita\_Mat

# Gráficos

# Metodología

Met\_Pro <- PROGRAMAS %>% summarise(Total = n(), .by = c(Metodologia)) %>%

rename(Clase = Metodologia) %>%

mutate(Variable = "METODOLOGIA",

Clase = factor(Clase))

Gra1 <- Plot.Barras(

datos = Met\_Pro,

categoria = "METODOLOGIA",

estatico = TRUE,

#vertical = FALSE,

freqRelativa = TRUE,

ordinal = FALSE,

titulo = "Distribución Total Programas Universitarios por Metodología de Formación",

labelEje = "Porcentaje",

colores = RColorBrewer::brewer.pal(5, "Set1"),

estilo = list(gg.Tema = 5,

gg.Texto = list(subtitle = "Periodo 2018-2021")))

Gra1 + scale\_y\_continuous(limits = c(NA, 100))+

theme(plot.title=element\_text(hjust=0, size=12),

axis.title=element\_text(size=14),

plot.caption=element\_text(size=7),

legend.text=element\_text(size=9),

axis.text = element\_text(size=10))

# Metodología - Matriculados

Met\_Pro\_Mat <- PROGRAMAS %>% summarise(Total = sum(Total), .by = c(Metodologia)) %>%

rename(Clase = Metodologia) %>%

mutate(Variable = "METODOLOGIA",

Clase = factor(Clase))

Gra1 <- Plot.Barras(

datos = Met\_Pro\_Mat,

categoria = "METODOLOGIA",

estatico = TRUE,

#vertical = FALSE,

freqRelativa = TRUE,

ordinal = FALSE,

titulo = "Distribución Matriculados Primera Vez por Metodología de Formación",

labelEje = "Porcentaje",

colores = RColorBrewer::brewer.pal(5, "Set1"),

estilo = list(gg.Tema = 5,

gg.Texto = list(subtitle = "Periodo 2018-2021")))

Gra1 + scale\_y\_continuous(limits = c(NA, 100))+

theme(plot.title=element\_text(hjust=0, size=12),

axis.title=element\_text(size=14),

plot.caption=element\_text(size=7),

legend.text=element\_text(size=9),

axis.text = element\_text(size=10))

# Áreas del Conocimiento

Area\_Prog <- PROGRAMAS %>% summarise(Total = n(), .by = c(Area\_Con)) %>%

rename(Clase = Area\_Con) %>%

mutate(Variable = "AREA")

Gra2 <- Plot.Barras(

datos = Area\_Prog,

categoria = "AREA",

freqRelativa = TRUE,

estatico = TRUE,

vertical = FALSE,

ordinal = FALSE,

titulo = "Distribución de Programas Académicos Universitarios por Áreas del Conocimiento",

labelEje = "Porcentaje",

colores = RColorBrewer::brewer.pal(9, "Set1"),

estilo = list(gg.Tema = 5,

gg.Texto = list(subtitle = "Periodo 2018-2021")))

Gra2 +

scale\_y\_continuous(limits = c(NA, 50))+

theme(plot.title=element\_text(hjust=0, size=12),

axis.title=element\_text(size=14),

plot.caption=element\_text(size=7),

legend.text=element\_text(size=9),

axis.text = element\_text(size=10))

# Áreas del Conocimiento Matriculados

Area\_Prog\_Mat <- PROGRAMAS %>% summarise(Total = sum(Total), .by = c(Area\_Con)) %>%

rename(Clase = Area\_Con) %>%

mutate(Variable = "AREA")

Gra2 <- Plot.Barras(

datos = Area\_Prog\_Mat,

categoria = "AREA",

freqRelativa = TRUE,

estatico = TRUE,

vertical = FALSE,

ordinal = FALSE,

titulo = "Distribución Matriculados Primera Vez Programas Universitarios\npor Áreas del Conocimiento",

labelEje = "Porcentaje",

colores = RColorBrewer::brewer.pal(9, "Set1"),

estilo = list(gg.Tema = 5,

gg.Texto = list(subtitle = "Periodo 2018-2021")))

Gra2 +

scale\_y\_continuous(limits = c(NA, 40))+

theme(plot.title=element\_text(hjust=0, size=12),

axis.title=element\_text(size=14),

plot.caption=element\_text(size=7),

legend.text=element\_text(size=9),

axis.text = element\_text(size=10))

# NBC

NBC\_Prog <- PROGRAMAS %>% summarise(Total = n(), .by = c(NBC)) %>%

rename(Clase = NBC) %>%

mutate(Variable = "NBC") %>% arrange(desc(Total))

NBC\_Prog\_Top10 <- PROGRAMAS %>% summarise(Total = n(), .by = c(NBC)) %>%

rename(Clase = NBC) %>%

mutate(Variable = "NBC") %>%

arrange(desc(Total)) %>%

slice\_max(Total, n= 10)

write\_xlsx(NBC\_Prog\_Top10, "Datos/NBC\_Prog\_Top10.xlsx")

Gra2 <- Plot.Barras(

datos = NBC\_Prog,

categoria = "NBC",

freqRelativa = TRUE,

estatico = TRUE,

vertical = FALSE,

ordinal = TRUE,

titulo = "Distribución de Programas Académicos Universitarios por Núcleos del Conocimiento",

labelEje = "Porcentaje"

#colores = RColorBrewer::brewer.pal(56, "Set1")

)

Gra2 +

scale\_y\_continuous(limits = c(NA, 30))+

theme(plot.title=element\_text(hjust=1, size=12),

axis.title=element\_text(size=14),

plot.caption=element\_text(size=7),

legend.text=element\_text(size=9),

axis.text = element\_text(size=10))

# NBC + Matriculados

NBC\_Prog\_Mat <- PROGRAMAS %>% summarise(Total = sum(Total), .by = c(NBC)) %>%

rename(Clase = NBC) %>%

mutate(Variable = "NBC") %>% arrange(desc(Total))

Gra2 <- Plot.Barras(

datos = NBC\_Prog\_Mat,

categoria = "NBC",

freqRelativa = TRUE,

estatico = TRUE,

vertical = FALSE,

ordinal = TRUE,

titulo = "Distribución Matriculados Primera Vez Programas Universitarios por Núcleos del Conocimiento",

labelEje = "Porcentaje"

#colores = RColorBrewer::brewer.pal(56, "Set1")

)

Gra2 +

scale\_y\_continuous(limits = c(NA, 30))+

theme(plot.title=element\_text(hjust=1, size=12),

axis.title=element\_text(size=14),

plot.caption=element\_text(size=7),

legend.text=element\_text(size=9),

axis.text = element\_text(size=10))

# Áreas y NBC TreeMap

Plot.Treemap(

datos = PROGRAMAS,

variables = vars(Area\_Con, NBC),

estatico = TRUE,

estilo = list(

gg.fontsize.title = 12, gg.fontsize.labels = c(15, 9),

gg.fontcolor.labels = c("#FFFFFF", "#212020"),

gg.border.lwds = c(4, 2), gg.border.col = c("#636363", "white"),

gg.lowerbound.cex.labels = 0.2, gg.overlap.labels = 0.1

)

)

# Top 20 programas académicos - General

Top\_20 <- PROGRAMAS %>% summarise(Total = sum(Total), .by = c(SNIES\_PROGRA, Programa, Prog\_Acredita, Institucion, Metodologia)) %>%

slice\_max(Total, n= 20)

write\_xlsx(Top\_20, "Datos/Top\_20.xlsx")

# Top 20 programas académicos -

Top\_20\_Presencial <- PROGRAMAS %>% filter(Metodologia == "Presencial") %>%

summarise(Total = sum(Total), .by = c(SNIES\_PROGRA, Programa, Prog\_Acredita, Institucion, Metodologia)) %>%

slice\_max(Total, n= 20)

write\_xlsx(Top\_20\_Presencial, "Datos/Top\_20\_Presencial.xlsx")

Top\_n50 <- PROGRAMAS %>% summarise(Total = sum(Total), .by = c(SNIES\_PROGRA, Programa, Prog\_Acredita, Institucion, Metodologia)) %>%

filter(Total <= 50)