

# Investigación sobre las universidades

INTEGRANTES: Camila Gonzalez, Luciana Paralela, Camila Rondon, Mauricio Medina, Lucas Colosi

## Recolección De Datos

## INTRODUCCIÓN

### Opción 1

El trabajo se basa en el análisis de datos provenientes de dos fuentes de datos principales que recopilan y publican estadísticas del sistema universitario argentino:

**1 Anuarios de Estadísticas Universitarias del Ministerio de Educación de la Nación:** Publicados en el portal oficial del Ministerio de Educación de la Nación Argentina ([argentina.gob.ar/educacion/universidades](http://argentina.gob.ar/educacion/universidades)), estos anuarios reúnen información cuantitativa sobre las universidades públicas y privadas del país. Los datos incluyen variables como cantidad de ingresantes, reinscriptos y egresados, área disciplinar, modalidad, género y otros factores relevantes. Esta fuente es una herramienta clave para el análisis y la planificación de políticas educativas a nivel superior.

**2 Encuesta sobre Condiciones de Ingreso a la Universidad (ECIU):** Se utiliza una base de datos derivada de la ECIU, difundida a través de una hoja de cálculo compartida públicamente (Google Sheets). Esta encuesta recopila información cualitativa y cuantitativa sobre los estudiantes al momento de ingresar a la universidad, incluyendo variables como si trabajan, si tienen personas a cargo, el nivel educativo de los padres, entre otras dimensiones socioeconómicas y demográficas. Esta información resulta útil para explorar posibles factores asociados a la retención y egreso de los estudiantes.

Ambas fuentes son de gran relevancia institucional y social, ya que permiten conocer en profundidad la dinámica del sistema universitario argentino. En particular, su análisis contribuye a comprender los **factores que inciden en la tasa de egreso universitaria**, temática central de este trabajo.

### Opción 2 (muy breve)

Este trabajo utiliza datos del **Ministerio de Educación de la Nación Argentina**, extraídos de los Anuarios de Estadísticas Universitarias, y de una base complementaria proveniente de la **Encuesta sobre Condiciones de Ingreso a la Universidad (ECIU)**. Los datasets incluyen información sobre **ingresantes, reinscriptos y egresados** en universidades públicas y privadas del país, desagregada por *carrera, área disciplinar, género, edad, modalidad de estudio y situación laboral*. Estas fuentes permiten analizar los factores que influyen en la *tasa de egreso universitaria*, como el tipo de institución, la disciplina académica, y las condiciones socioeconómicas de los estudiantes

## Definición de los objetivos del análisis

La *problemática* que hemos elegido para el presente trabajo se basa en los factores que hacen que haya menor o mayor tasa de egresados en las universidades de Buenos Aires. En base a esta problemática y a partir de los data sets seleccionados y a su respectivo análisis, hemos elegido los siguientes factores a analizar:

- Cuál es la diferencia en cantidad de graduados en instituciones públicas o privadas (tanto a nivel general y en carreras específicas).
- Cómo varía la cantidad de egresados s/ área disciplinar (Área de ingeniería tiene más egresados que en otras áreas). (Gráfico de torta o barras: eje x: área disciplinar, eje y: cantidad de egresados)
- Carreras que tienen menor/mayor cantidad de egresados. (Abogacía, contador público, enfermería).
- Carreras con mayor y menor retención estudiantil (Con mayor egresados/retención son abogacía, contador público y enfermería; con menor retención (más cantidad de ingresantes pero menor cantidad de egresados).
- Cómo varía la cantidad de graduados según su edad y género. **CAMI G, SOLO GENERO**
- Cómo varía la cantidad de egresados empleados, dependiendo de la cantidad de ingresantes con o sin empleo.

### Análisis exploratorio (algunas preguntas las vamos a sacar porque son muchas)

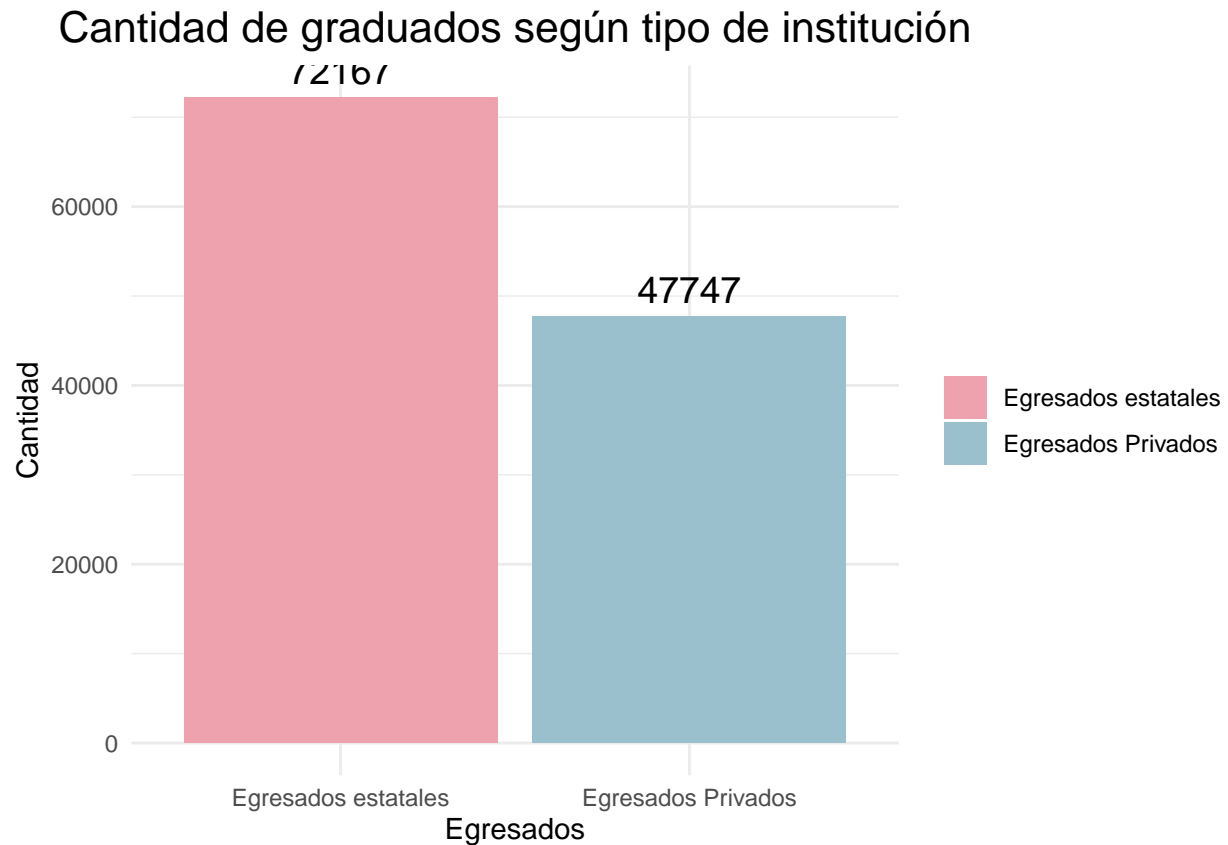
- ¿Qué carreras tienen mayor y menor cantidad de egresados? **CAMI G, SOLO MAYOR CANTIDAD**
- ¿Cómo varía la cantidad de egresados por área disciplinar? **CAMI R (CUADRO Y OJIVA)**
- ¿Existen diferencias significativas en la cantidad de egresados entre instituciones estatales y privadas en la misma carrera? **CAMI G, DIF. ENTRE Q EN PRIV Y PUB, PERO NO SEGUN CARRERAR**
- Cómo cambió el número total de egresados entre 2020, 2021, 2023?
- ¿Crecieron más los egresados en universidades estatales o privadas?
- ¿Qué año tiene mayor cantidad de reinscriptos?
- ¿Existe relación entre duración promedio y tasa de egreso?
- ¿La relación alumno/docente influye en la tasa de egreso?
- ¿Qué modalidad tiene mayor cantidad de ingresantes/graduados?
- ¿Cómo cambió a lo largo del tiempo la modalidad elegida y su cantidad de graduados?
- ¿Qué género tiene mayor cantidad de graduados, son más hombres o mujeres? **CAMI G**
- ¿Qué rango de edad tiene más egresados?
- ¿Cómo afecta en la cantidad de graduados el nivel de academico de los padres? **luca**
- ¿Cómo varía la cantidad que inscriptos que trabajan en universidades privadas y públicas? Puede esto afectar la tasa de egreso en cada una?

### Justificación de la elección del dataset

El dataset de los “Anuarios Estadísticos” del sistema universitario es la fuente más adecuada para estudiar la evolución en diversas áreas, las causas de ello y diferencias en la tasa de egreso universitaria en Buenos Aires. Además la información es confiable porque es oficial, completa, accesible, comparable entre años, y centrada en variables directamente vinculadas al fenómeno de estudio.

### Cantidad de graduados según tipo de institución

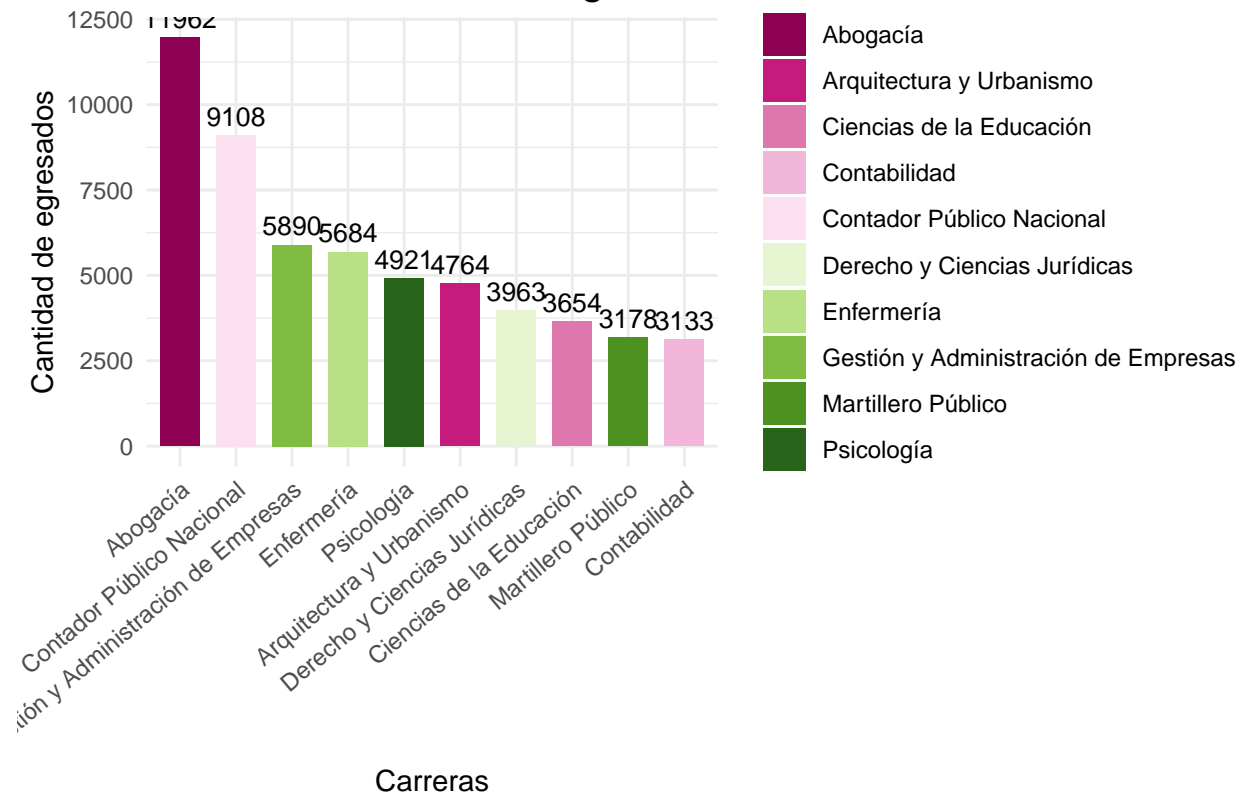
Se analiza la **diferencia entre cantidad de graduados provenientes de universidades públicas y universidades privadas**, el objetivo es ver cuál tipo proporciona una mayor cantidad de graduados a nivel nacional, podremos ver estas diferencias representadas en el siguiente gráfico:



### TOP 10 CARRERAS CON MÁS GRADUADOS

Para ampliar la información, nos hemos preguntado ¿cuáles son las carreras con más cantidad de graduados? para responder a ello, hemos escogido las **top 10 carreras con más cantidad de graduados** y hemos adicionado el total por cada una de ellas, esto es a nivel nacional, según datos del 2020.

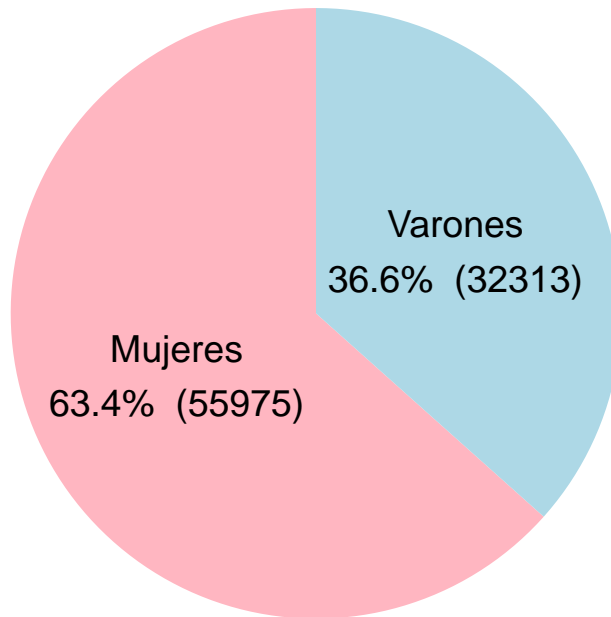
## TOP 10 carreras con mas graduados



## RELACION GÉNERO - CANTIDAD DE GRADUADOS - TIPO DE INSTITUCIÓN

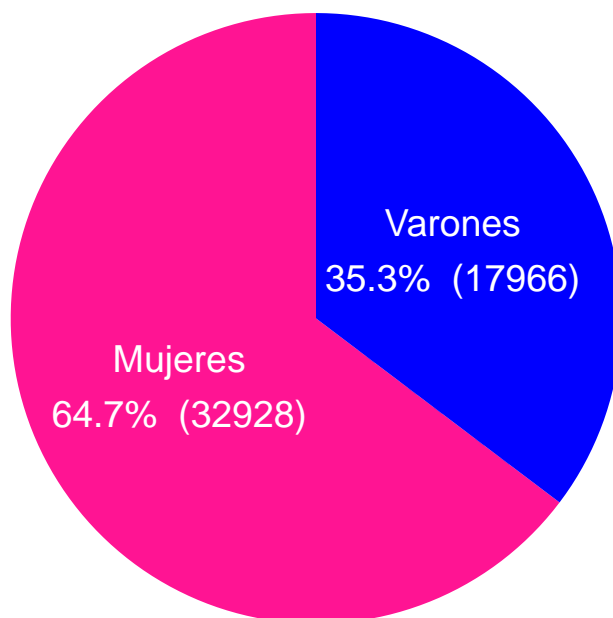
A continuación, realizamos un pequeño análisis respecto a la cantidad de egresados, su género (mujer u hombre) y el tipo de institución de la cual proviene, ya sea universidad pública o privada. El objetivo de esto es ver si hay más egresados del género femenino o masculino. El primer gráfico que veremos será aquel proveniente de las **universidades públicas**:

## Distribución de Egresados por Género Universidades Publicas – Año 2023



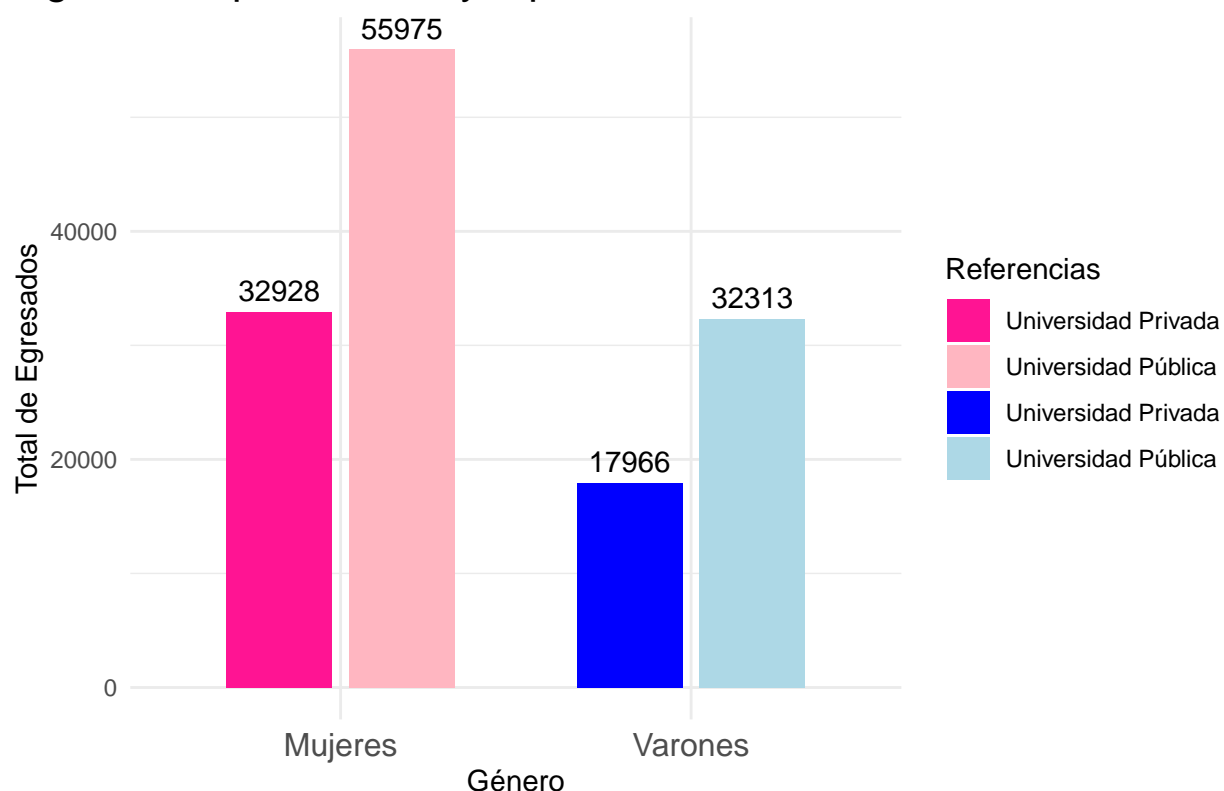
A continuación, el gráfico relacionado a las **universidades privadas**:

## Distribución de Egresados por Género Universidades Privadas – Año 2023



Por último, el siguiente gráfico es una pequeña **comparación de la cantidad de egresados según género y tipo de universidad**, como podemos apreciar, hay una mayor cantidad de graduadas mujeres, y este patrón se da tanto en universidades públicas como privadas:

## Egresados por Género y Tipo de Universidad – 2023



### Reinscriptos en universidades nacionales de gestión estatal

Se analiza la cantidad de reinscriptos en universidades **nacionales de gestión estatal** en los años **2020, 2021 y 2023** a partir de los anuarios estadísticos del sistema universitario.

```
library(readxl)
library(dplyr)
library(ggplot2)

# Leer datos
data_2020 <- read_excel("1.2 Indicadores 2020.xlsx", sheet = "C 1.2.1a", skip = 2)
data_2021 <- read_excel("1.2 Indicadores 2021.xlsx", sheet = "C 1.2.1a", skip = 2)
data_2023 <- read_excel("1.2 Indicadores 2023.xlsx", sheet = "C 1.2.1a", skip = 2)

# Filtrar y extraer los valores de reinscriptos desde la 6ta columna
ri_2020 <- data_2020 %>%
  filter(Institución == "Total Universidades Nacionales") %>%
  pull(6)

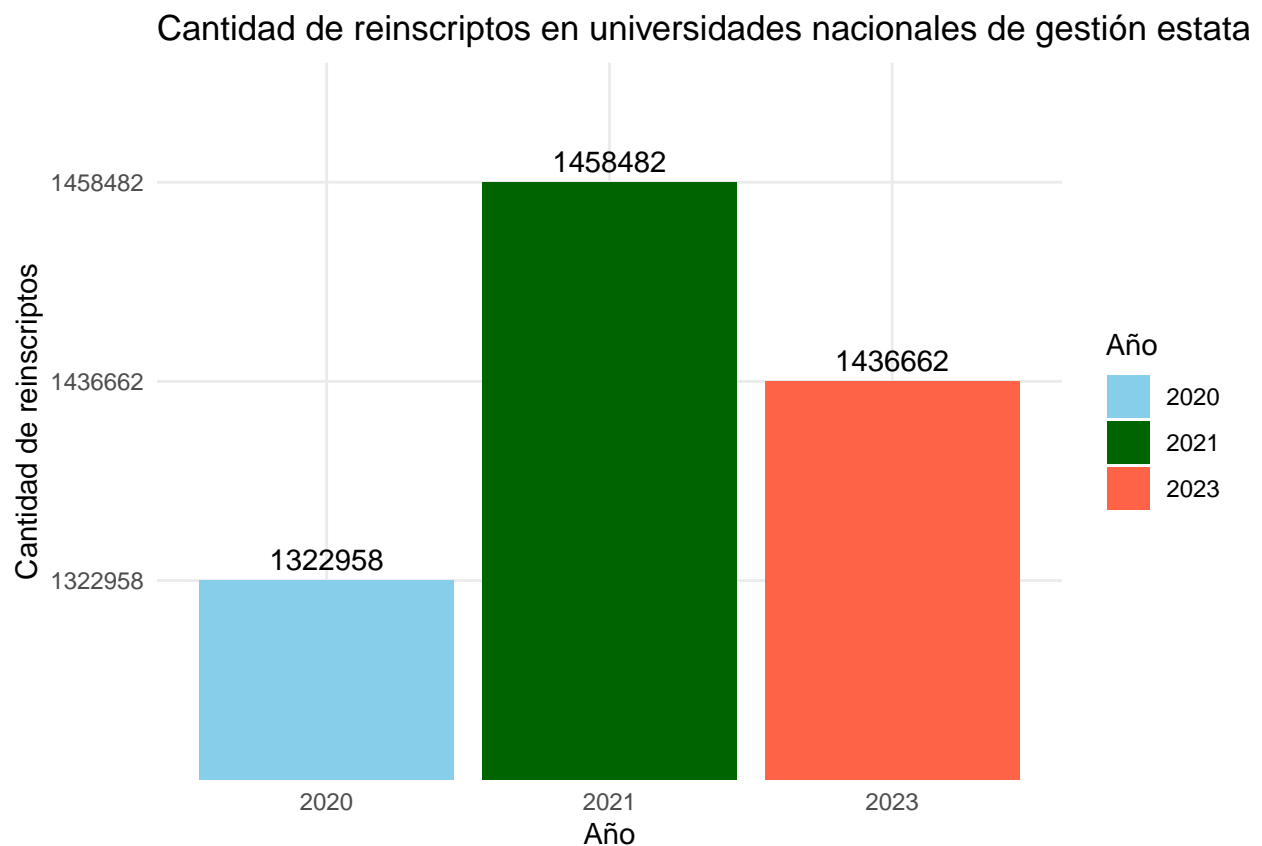
ri_2021 <- data_2021 %>%
  filter(Institución == "Total Universidades Nacionales") %>%
  pull(6)

ri_2023 <- data_2023 %>%
  filter(Institución == "Total Universidades Nacionales") %>%
```

```
pull(6)

# Creo un data frame con los datos
reinscriptos <- data.frame(
  Año = c("2020", "2021", "2023"),
  Reinscriptos = c(ri_2020, ri_2021, ri_2023)
)

# Creo el gráfico
ggplot(reinscriptos, aes(x = Año, y = Reinscriptos, fill = Año)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = Reinscriptos), vjust = -0.5) +
  labs(
    title = "Cantidad de reinscriptos en universidades nacionales de gestión estatal",
    x = "Año",
    y = "Cantidad de reinscriptos"
  ) +
  scale_fill_manual(values = c("2020" = "skyblue", "2021" = "darkgreen", "2023" = "tomato")) +
  theme_minimal()
```





Cantidad de egresados de Universidades Nacionales y Provinciales.

A partir del siguiente gráfico se hace un analisis de la cantidad de egresados de los años 2021 y 2023 en Universidades tanto Nacionales como provinciales.

```
# Cargo paquetes necesarios
library(readxl)
library(dplyr)
library(ggplot2)

# Cargo datos desde la hoja correspondiente
datos_2021 <- read_excel("1.2 Indicadores 2021.xlsx", sheet = "C 1.2.1a", skip = 2)

## New names:
## * `` -> `...3`
## * `` -> `...5`
## * `` -> `...7`
## * `` -> `...9`

datos_2023 <- read_excel("1.2 Indicadores 2023.xlsx", sheet = "C 1.2.1a", skip = 2)

## New names:
## * `` -> `...3`
## * `` -> `...5`
## * `` -> `...7`
## * `` -> `...9`

# Proceso datos de 2021
egresados_2021 <- datos_2021 %>%
  select(Institución, `Egresadas/os (EGR)`) %>%
  filter(!is.na(`Egresadas/os (EGR)`) ) %>%
  mutate(
    `Egresadas/os (EGR)` = as.numeric(gsub("[^0-9]", "", `Egresadas/os (EGR)`)),
    tipo = case_when(
      grepl("Nacional", Institución) ~ "Nacional",
      grepl("Provincial", Institución) ~ "Provincial",
      TRUE ~ NA_character_
    ),
    año = 2021
  ) %>%
  filter(!is.na(tipo)) %>%
  group_by(año, tipo) %>%
  summarise(Egresados = sum(`Egresadas/os (EGR)`, na.rm = TRUE), .groups = "drop")

# Proceso datos de 2023
egresados_2023 <- datos_2023 %>%
  select(Institución, `Egresados (EGR)`) %>%
  filter(!is.na(`Egresados (EGR)`) ) %>%
  mutate(
    `Egresados (EGR)` = as.numeric(gsub("[^0-9]", "", `Egresados (EGR)`)),
    tipo = case_when(
```

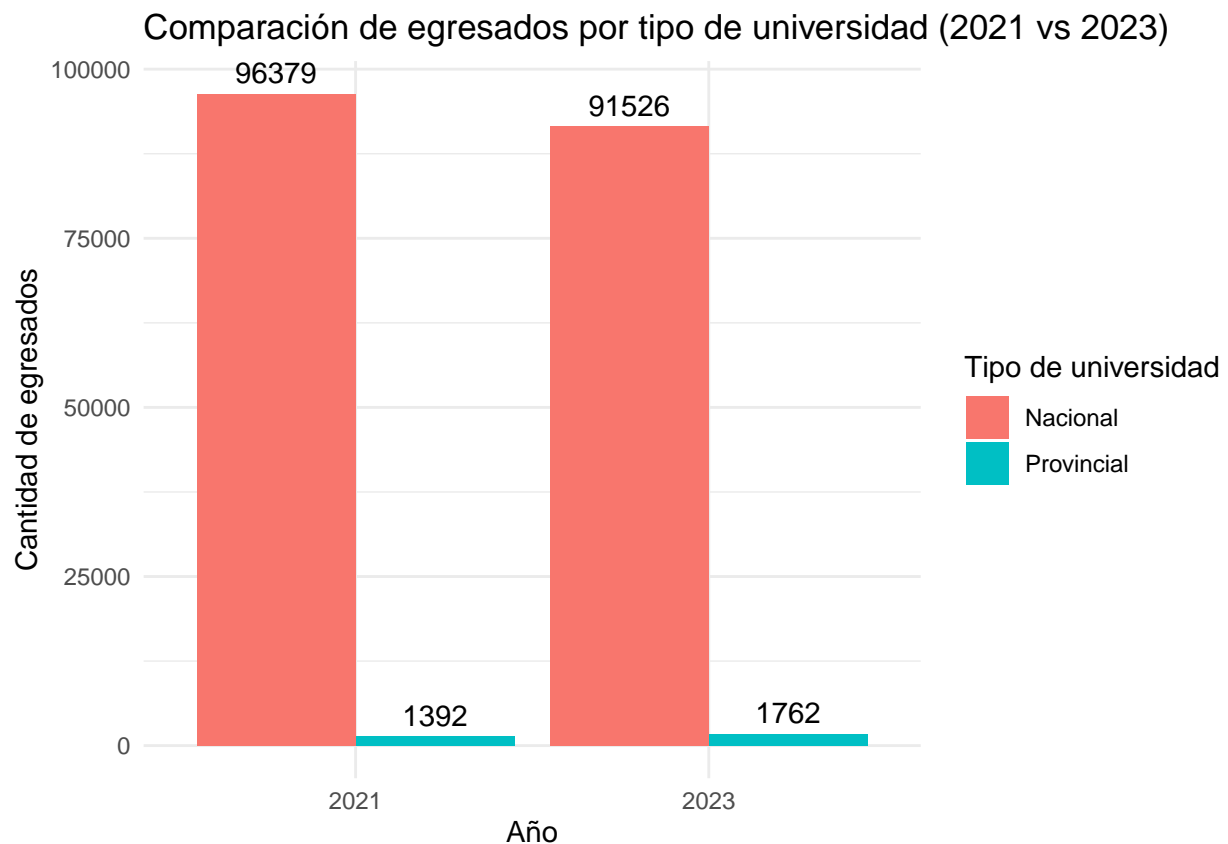
```

grepl("Nacional", Institución) ~ "Nacional",
grepl("Provincial", Institución) ~ "Provincial",
TRUE ~ NA_character_
),
año = 2023
) %>%
filter(!is.na(tipo)) %>%
group_by(año, tipo) %>%
summarise(Egresados = sum(`Egresados (EGR)`, na.rm = TRUE), .groups = "drop")

# Uno todos los datos
egresados_comparados <- bind_rows(egresados_2021, egresados_2023)

# Creo gráfico de columnas agrupadas
ggplot(egresados_comparados, aes(x = factor(año), y = Egresados, fill = tipo)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = Egresados), position = position_dodge(width = 0.9), vjust = -0.5) +
  labs(
    title = "Comparación de egresados por tipo de universidad (2021 vs 2023)",
    x = "Año",
    y = "Cantidad de egresados",
    fill = "Tipo de universidad"
  ) +
  theme_minimal()

```



## ¿Cómo varía la cantidad de egresados por área disciplinar?

A partir de los datos recolectados, se puede observar que la cantidad de egresados varía bastante según el área disciplinar. Algunas áreas, como *Ciencias de la Salud: Medicina*, tienen una cantidad muy alta de egresados tanto en universidades estatales como privadas, lo que indica una fuerte presencia en el sistema educativo. En cambio, áreas como *Arquitectura y Diseño* o *Biología* tienen cifras más bajas.

También se nota que en ciertas disciplinas, como *Economía y Administración*, el número de egresados en la gestión privada es incluso mayor que en la estatal. Este tipo de diferencias nos ayudan a entender cómo se distribuyen los egresos en función del tipo de institución y del interés que despierta cada carrera.

Cantidad de egresados por área disciplinar (2020-2022)

Rama y disciplina

Años

Total gestión estatal

Total gestión privada

Total

Ciencias aplicadas: arquitectura y diseño

2020, 2021, 2023

141, 122, 249

28, 32, 21

593

ciencias basicas: Biología

2020, 2021, 2023

392, 363, 478

13, 24, 23

1293

Ciencias de la salud: Medicina

2020, 2021, 2023

1656, 1762, 4968

658, 778, 747

10569

Ciencias humanas: Educacion

2020, 2021, 2023

1364, 930, 1160

289, 264, 316

4323

Ciencias sociales: Economía y administración

2020, 2021, 2023

638, 930, 1121

1653, 1429, 2056

7827

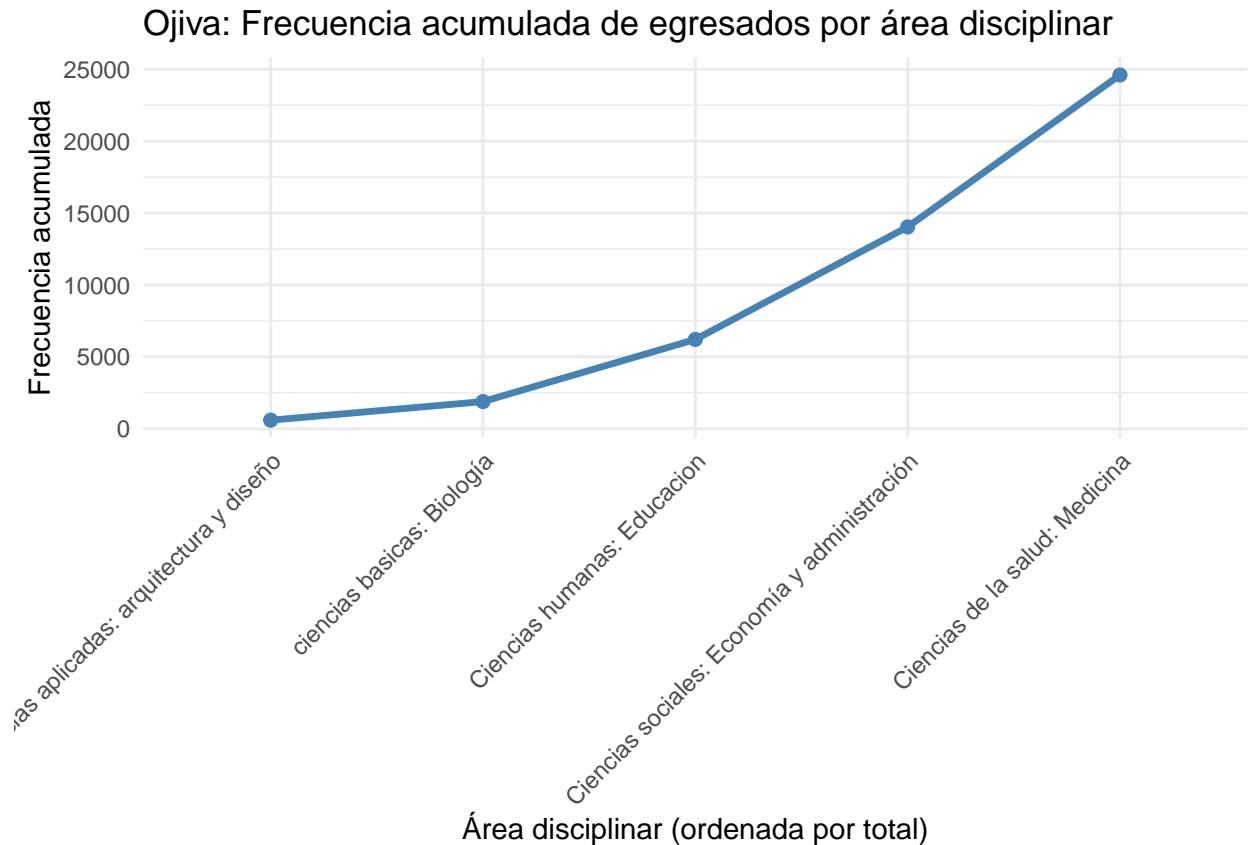
## Ojiva

```
library(dplyr)
library(ggplot2)

# Aseguramos que la columna Total sea numérica
egresados <- egresados %>%
  mutate(Total = as.numeric(Total)) %>%
  arrange(Total) %>%
  mutate(FrecuenciaAcumulada = cumsum(Total))

# Creamos el Ojiva
ggplot(egresados, aes(x = reorder(`Rama y disciplina`, Total), y = FrecuenciaAcumulada, group = 1)) +
  geom_line(color = "steelblue", size = 1.2) +
  geom_point(color = "steelblue", size = 2) +
  labs(
    title = "Ojiva: Frecuencia acumulada de egresados por área disciplinar",
    x = "Área disciplinar (ordenada por total)",
    y = "Frecuencia acumulada"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



### Interpretación de la ojiva

La ojiva permite visualizar cómo se acumula el total de egresados por área disciplinar. Al ordenar las disciplinas de menor a mayor según su cantidad de egresados, observamos que unas pocas áreas concentran una gran parte del total acumulado. Por ejemplo, Ciencias de la Salud: Medicina aporta un incremento considerable en la curva, mientras que disciplinas como Arquitectura y Diseño o Biología generan un aumento más gradual.

Este tipo de representación es útil porque permite identificar rápidamente qué áreas tienen mayor peso en el total de egresados. Desde una perspectiva de Ciencia de Datos, la ojiva ayuda a detectar posibles desequilibrios en la distribución y podría servir como base para futuros análisis más profundos o recomendaciones educativas.

### Conclusión

Podemos observar que la variabilidad no solo depende del área, sino también del año y del tipo de institución. Este tipo de patrones pueden ser claves para entender cómo se comporta el sistema educativo y qué áreas requieren más atención o inversión. Como estudiante de Ciencia de Datos, es interesante pensar cómo estos datos podrían ayudar a construir modelos predictivos o a guiar futuras decisiones en política educativa.

### Datos de egresados

```
egresados <- data.frame(
  Año = c(2020, 2021, 2023),
  Total = c(139182, 142826, 139182)
)
```

```
egresados
```

```
##   Año  Total
## 1 2020 139182
## 2 2021 142826
## 3 2023 139182
```

## Cálculo de cambios entre años

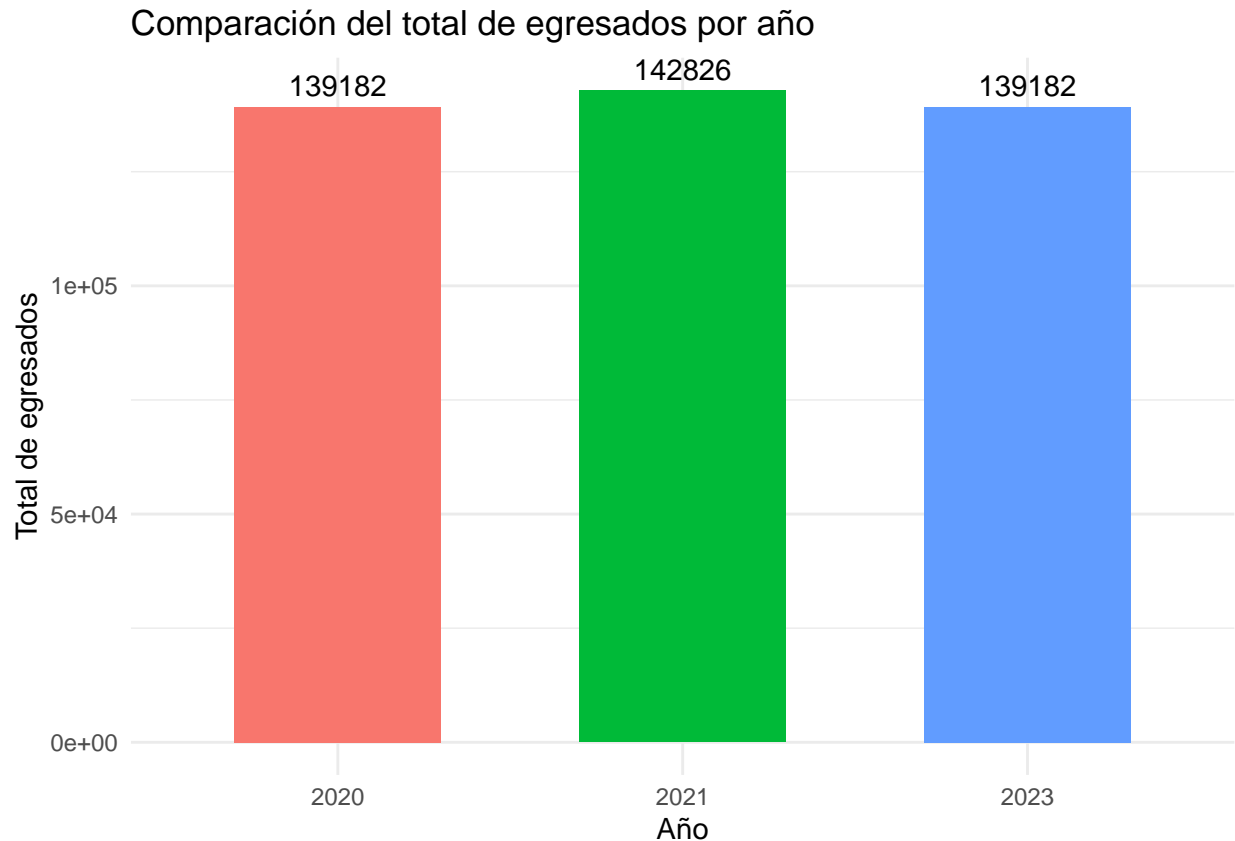
```
egresados$Cambio_Absoluto <- c(NA, diff(egresados$Total))
egresados$Cambio_Porcentaje <- c(NA, round((diff(egresados$Total) / head(egresados$Total, -1)) * 100, 2))
```

```
egresados
```

```
##   Año  Total Cambio_Absoluto Cambio_Porcentaje
## 1 2020 139182             NA              NA
## 2 2021 142826          3644           2.62
## 3 2023 139182         -3644          -2.55
```

## Gráfico comparativo de egresados

```
ggplot(egresados, aes(x = factor(Año), y = Total, fill = factor(Año))) +
  geom_bar(stat = "identity", width = 0.6) +
  labs(title = "Comparación del total de egresados por año",
       x = "Año",
       y = "Total de egresados") +
  theme_minimal() +
  theme(legend.position = "none") +
  geom_text(aes(label = Total), vjust = -0.5)
```



## Interpretación

### Análisis del cambio en el número total de egresados

Entre 2020 y 2021, el total de egresados pasó de 139182 a 142826, lo que representa un aumento de 3644 egresados (2.62%).

Entre 2021 y 2023, el total de egresados pasó de 142826 a 139182, lo que representa una disminución de 3644 egresados (-2.55%).

En resumen, el gráfico de barras nos permite visualizar claramente que el año 2021 fue el de mayor cantidad de egresados, mientras que en 2020 y 2023 el número total fue exactamente el mismo. Esta oscilación puede reflejar variaciones en cohortes, políticas educativas o condiciones externas como la pandemia.

## Cómo afecta a la cantidad de graduados el nivel académico de sus padres ?

```
#llamamos a las librerías que podemos llegar a utilizar
library(readxl)
library(ggplot2)
library(dplyr)
library(tidyr)

# nivel académico de la madre del graduado
```

```

nivel_madre <- read_excel("2.1 Pregrado y Grado Gestión Estatal 2023.xlsx", sheet = "C.2.1.32", skip=4)

#reorganizamos y filtramos el dataset
colnames(nivel_madre)[1] <- "Institucion"
nivel_madre_limpio <- nivel_madre %>%

  filter(Institucion == "Total Universidades Nacionales") %>%
  pivot_longer(
    cols = -Institucion,
    names_to = "Nivel_academico",
    values_to = "Cantidad") %>%

  mutate(
    Cantidad = as.numeric(Cantidad),
    Nivel_academico_sesgo = trimws(tolower(Nivel_academico)),
    Padre_Madre = "Madre") %>%

#limpiamos las columnas que no nos interesan
  filter(!Nivel_academico_sesgo %in% c("total", "no sabe", "ndi"))

library(readxl)
library(dplyr)
library(tidyr)

# nivel academico del Padre del graduado

#primer paso leemos el dataset y llamamos nuestra hoja correspondiente.
nivel_padre <- read_excel("2.1 Pregrado y Grado Gestión Estatal 2023.xlsx", sheet = "C.2.1.33", skip=4)
colnames(nivel_padre)[1] <- "Institucion"
nivel_padre_limpio <- nivel_padre %>%
  filter(Institucion == "Total Universidades Nacionales") %>%
  pivot_longer(
    cols = -Institucion,
    names_to = "Nivel_academico",
    values_to = "Cantidad"
  ) %>%
  mutate(
    Cantidad = as.numeric(Cantidad),
    Nivel_academico_sesgo = trimws(tolower(Nivel_academico)),
    Padre_Madre = "Padre") %>%
  filter(!Nivel_academico_sesgo %in% c("total", "no sabe", "ndi"))

```

Unimos los dos codigos para luego graficar la comparacion ver juntos ambos resultados

```

library(readxl)
library(dplyr)
library(tidyr)

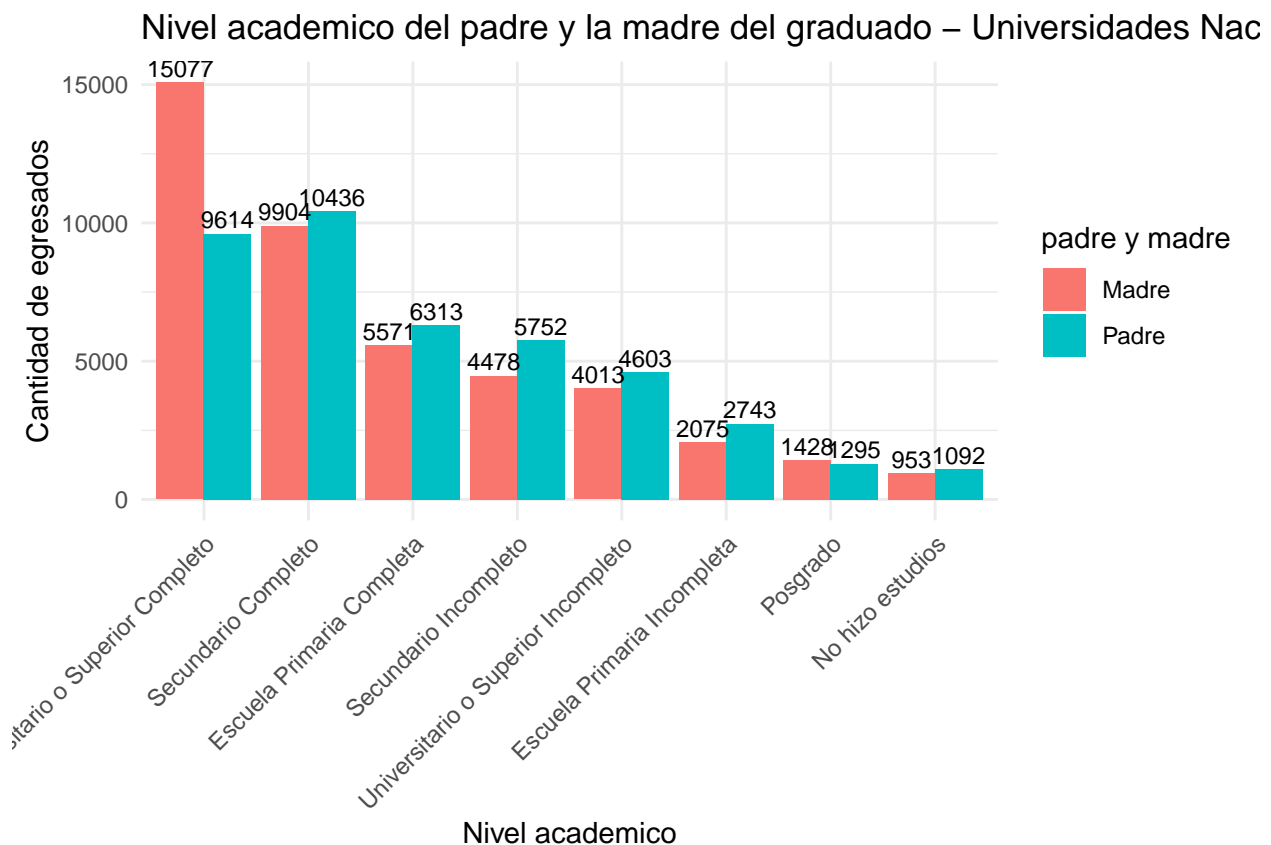
#asignamos la comparación

```



```
comparativo <- bind_rows(nivel_madre_limpio, nivel_padre_limpio)

# Gráficamos
ggplot(comparativo, aes(x = reorder(Nivel_academico, -Cantidad), y = Cantidad, fill = Padre_Madre)) +
  geom_col(position = "dodge") +
  geom_text(aes(label = Cantidad), position = position_dodge(width = 0.9), vjust = -0.3, size = 3) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(
    title = "Nivel academico del padre y la madre del graduado - Universidades Nacionales (2023)",
    x = "Nivel academico",
    y = "Cantidad de egresados",
    fill = "padre y madre"
  )
)
```



Como podemos observar en el grafico de barras el nivel de academico de los padres es crucial en la cantidad de graduados!