



DEPARTAMENTO  
DE COMPUTACION  
Facultad de Ciencias Exactas y Naturales - UBA

# Laboratorio de Datos

## Trabajo Práctico 1 1er. Cuatrimestre - 2025

**Tema:** Escuelas y bibliotecas

**Grupo:** JSG

**Integrantes:** Gastón Arida, Sebastián Ferreiro, Jerónimo Vernay

**Fecha de entrega:** 25 de mayo

## Sección Resumen

El objetivo de este trabajo fue investigar si existe una relación entre la cantidad de establecimientos educativos (EE) y bibliotecas populares (BP) en Argentina.

Para ello se analizaron bases de datos oficiales, evaluando su calidad, normalizando su estructura y comparando su distribución de EE y BP por departamento y provincia ajustados con los datos del censo para evitar sesgos poblacionales.

Luego de completar el trabajo pudimos encontrar una relación entre la cantidad de establecimientos educativos y la cantidad de bibliotecas populares en el país, pero también notamos que esta generalización no se expresa tal cual si miramos cada provincia por separado. Como veremos más adelante, hay provincias donde existe esta relación y otras donde no es así.

## Sección Introducción

El trabajo tiene como objetivo analizar la existencia de una relación entre la cantidad de establecimientos educativos y la cantidad de bibliotecas populares por provincia en Argentina.

Para abordar esto, el grupo realizó un análisis exhaustivo de bases de datos que contienen información sobre los establecimientos educativos, las bibliotecas populares y el padrón poblacional de la Argentina. Algunos de las actividades que realizamos fueron:

- Procesamiento y normalización de datos: Se evaluó la estructura de las tablas para asegurar que estén normalizadas (1FN, 2FN y 3FN). En el proceso se identificaron y resolvieron problemas como incompletitud y redundancias.
- Análisis y calidad de datos: Mediante la técnica de GQM se midieron los problemas asociados a las tablas originales proporcionadas por el enunciado y se midieron los atributos de calidad afectados.
- Unificación de criterios: En base a nuestras decisiones, estandarizamos los códigos de departamento ("id") para permitir comparaciones.
- Armado de Gráficos y reportes: Se diseñaron gráficos, incluyendo algunos interactivos, para comparar la relación entre establecimientos educativos y bibliotecas populares en los departamentos y las provincias, ajustados poblacionalmente para evitar el sesgo previamente mencionado.

Los resultados preliminares muestran que a pesar de que existe una relación moderada entre ambas variables, hay una intuitiva desigualdad territorial, debido a que Buenos Aires, Córdoba y Santa Fe (3 de 24 jurisdicciones) representan la mayor proporción de bibliotecas populares, mientras que las otras concentran una cantidad muy baja.

A continuación el informe continuará con las siguientes secciones:

- Procesamiento de datos : Normalizamos las tablas.
- Análisis de la calidad de datos : Exponemos algunos problemas de las tablas originales y proponemos soluciones de mejora.
- Toma de Decisiones : Explicamos las decisiones que tomamos.
- Análisis de resultados : Interpretamos los hallazgos encontrados.
- Conclusiones : Resumimos las implicaciones de nuestros resultados y su relevancia para nuestro objetivo planteado.

# Sección Procesamiento de Datos

DER (ver *Figura 1*)

**Modelo Relacional** (ver *Figura 2*) :

**Dependencias Funcionales:**

BP

- $\text{nro\_conabip} \rightarrow \text{año, dominio, id\_departamento}$

PADRÓN

- $(\text{id\_departamento, grupo\_etario}) \rightarrow \text{casos}$

EE

- $\text{cue\_anexo} \rightarrow \text{común, primario, jardín, secundario, id\_departamento}$

PROV

- $\text{id} \rightarrow \text{nombre\_provincia}$

DEPTO

- $\text{id} \rightarrow \text{nombre\_departamento, id\_provincia}$

Las dependencias funcionales que fueron planteadas previamente, se hicieron teniendo en cuenta que, para toda instancia “r” de cada uno de los esquemas de relación, no es posible que “r” tenga dos tuplas que coincidan en los atributos de lado izquierdo y no lo hagan en los de lado derecho.

Además, es importante tener en cuenta que para definir las dependencias funcionales, y asegurarnos que se mantengan en todos los estados de relación, utilizamos nuestra comprensión de la semántica de los atributos, es decir, analizamos cómo se relacionaban los atributos entre sí en cada caso.

Para analizar si nuestros esquemas están en 3FN, se debería cumplir lo siguiente:

- Satisfacen 2FN
- Ningún atributo no primo es transitivamente dependiente en la clave principal

Para empezar, verifiquemos que están en 1FN. Como en cada uno de nuestros esquemas los atributos contienen valores atómicos, se cumple.

Luego, al no haber dependencias parciales, ya que todo atributo no primo es completa y funcionalmente dependiente de la PK de su esquema, están en 2FN.

Como no hay dependencias transitivas, sumado a que previamente verificamos que están en 2FN, es válido afirmar que están en 3FN. No sólo esto, sino que también están en BCNF, que al ser más estricta que 3FN, toda relación que está en BCNF lo está también en 3FN. Esto se debe a que para todas las dependencias funcionales, el lado izquierdo siempre es una superclave del esquema.

Analicemos cómo se encuentran las formas normales en las tablas de Establecimientos Educativos y Bibliotecas Populares.

Por un lado, Establecimientos Educativos no cumple con 1FN, ya que en los atributos “Teléfono” y “Mail” por ejemplo, se puede ver como el dominio está incluyendo valores que no son atómicos. La manera más eficiente de resolver este problema y poder lograr que esté en 1FN, sería reubicar estos atributos que violan 1FN y ubicarlos en nuevas relaciones junto con la PK de la tabla original. Podríamos crear las tablas Establecimiento\_Teléfono(PK, Teléfono) y Establecimiento\_Mail(PK, Mail).

Por otro lado, en Bibliotecas Populares el dominio de todos los atributos es atómico. Con lo cual, cumple con 1FN. Luego, para chequear si cumple 2FN, habría que hacer un análisis de los atributos que componen esta tabla para ver cuáles son las CK.

A través de Python, usando los datos que se nos brindaron en Bibliotecas Populares, obtuvimos que agrupando los datos por las siguientes combinaciones minimales de atributos, genera todas filas únicas:

- {'latitud', 'longitud', 'nombre'}
- {'nro\_conabip'}
- {'nombre', 'cp'}
- {'domicilio', 'nombre'}

Sin embargo, ¿sería válido afirmar que son nuestras CK?

Un conjunto de atributos que constituyen una clave debe mantenerse en cada estado de relación válido del esquema. Esta propiedad es fija en el tiempo, por lo tanto, debe mantenerse cuando insertemos nuevas tuplas en la relación.

Entonces, pensando desde el punto de vista del diseñador de este esquema, no tendría sentido decir que la 1, 3 o 4 opción lo son, ya que en cualquiera de ellas, al agregar nuevos datos a la tabla corre riesgo de que se rompa la clave. Mientras que el nro\_conabip, que es un identificador único, evita esos problemas. Además, nos va servir el hecho de que la clave tenga un sólo atributo.

Podemos decir que cumple con 2FN, porque al tener como única CK y por lo tanto PK “nro\_conabip”, que está compuesta por un solo atributo, no hace falta realizar ningún test.

Finalmente, para analizar la 3FN, vamos a usar la definición general de la 3FN. En esta se nos dice :

*“Un esquema de relación  $R$  está en tercera forma normal (3FN) si, siempre que una dependencia funcional no trivial  $X \rightarrow A$  se cumple en  $R$ , ya sea (a)  $X$  una superclave de  $R$ , o (b)  $A$  un atributo primo de  $R$ .”<sup>1</sup>*

Ahora, analizando los atributos de Bibliotecas Populares, podemos detectar la siguiente dependencia funcional:

id\_departamento  $\rightarrow$  provincia

---

<sup>1</sup> Elmasri, R., & Navathe, S. B. (2007). *Fundamentos de Sistemas de Bases de Datos* (5ª ed., p. 308). Pearson Educación

¿Es “id\_departamento” una superclave del esquema? ¿Es “provincia” un atributo primo del esquema?

Dado que la respuesta a ambas preguntas es no, encontramos una dependencia funcional que viola las dos condiciones (a) y (b) de la 3FN y por lo tanto el esquema viola la definición general de 3FN.

Para revisar la calidad de datos de las dos tablas originales, vamos a usar la técnica de GQM

## **Análisis de Calidad de Datos - Establecimientos Educativos**

### **Problema 1: Completitud del campo “Mail”**

- *Atributo de calidad afectado* : Completitud y
- *Problema asociado* : Instancia
- *GQM* :
  - ❖ *Goal* : Evaluar la completitud de la columna “Mail”
  - ❖ *Question*:
    - ¿Qué porcentaje de los establecimientos no tiene correo electrónico?
    - ¿Cuál es la proporción de registros cuyo campo mail no contiene el @, induciendo que el dato está mal cargado?
  - ❖ *Metric* : Porcentaje de registros con “mail” Null + Porcentaje de correos inválidos (sin @)
- *Resultados*:
  - ❖ **18.29%** de los establecimientos no tienen ningun mail registrado
  - ❖ **0.35%** de los correos existentes, son inválidos
  - ❖ Total de Correos válidos: **81.42%**
- *Diagnóstico*: 1 de cada 5 mails tiene una dirección de correo no existente o invalida. La falta de correos electrónicos puede afectar la comunicación institucional y comprometer a dichos establecimientos el acceso a mejores plataformas virtuales
- *Mejoras propuestas*: Una solución viable para resolver el tema de los correos, sería contactar a las jurisdicciones para hacer una actualización de los correos. Luego a la par de esto, generar un nuevo parámetro de entrada para excluir mails inválidos.

### **Problema 2: Estructura del campo “Domicilio”**

- *Atributo de calidad afectado* : Consistencia y Relevancia
- *Problema asociado* : Modelo de Datos
- *GQM* :
  - ❖ *Goal* : Evaluar la calidad estructural del campo “Domicilio”
  - ❖ *Question*: ¿Qué porcentaje de domicilios posee letras y ningún número?  
¿Qué porcentaje tiene caracteres extraños o están vacíos?
  - ❖ *Metric* : Porcentaje de domicilios sin número o vacíos
- *Resultados*:
  - ❖ **26.11%** de los domicilios no tienen número

❖ **1.14%** de los domicilios están vacíos

- *Diagnóstico:* 1 de cada 4 establecimientos educativos no tiene un domicilio preciso y 739 EE's no tienen siquiera un domicilio
- *Mejoras propuestas:* Como propuesta de mejora para el campo "domicilio", sugerimos implementar controles de entrada más estrictos para futuras bases de datos.  
Siendo el problema tan grande no vemos que haya alguna alternativa viable para resolverlo a corto plazo.

### **Otros problemas en EE:**

Evaluando el total de la información brindada. Vemos que hay problemas de completitud con el código Postal ("cp" tiene 2429 Nulls) y en "Ámbito" (tan solo 6 Nulls) también, pero decidimos buscar distintas causas de problemas.

## **Análisis de Calidad de Datos - Bibliotecas Populares**

### **Problema 1: Formato del campo "cp"**

- *Atributo de calidad afectado :* Consistencia
- *Problema asociado :* Proceso (misma información, distintas asunciones)
- *GQM :*
  - ❖ *Goal :* Garantizar que el campo "cp" tenga el formato válido y tipo adecuado
  - ❖ *Question:* ¿Está "cp" definido como string cuando debería ser numérico?  
¿Hay códigos postales con caracteres no numéricos?
  - ❖ *Metric :* Porcentaje de códigos postales con formato incorrecto
- *Resultados:*
  - ❖ **100%** de los CP son strings (deberían ser numéricos)
  - ❖ **12.78%** de los CP (numérico) utilizan formato CPA (alfanumérico)
- *Diagnóstico:* El tipo de dato con el que está indexado es incorrecto, idealmente deberían ser numéricos (consume menos memoria y facilita operaciones sobre el cp, por ejemplo si quisiéramos agrupar por rango). Hay inconsistencia en el armado de datos, a veces se espera el CP, otras el CPA.
- *Mejoras propuestas:* Para solucionar el problema del código postal en "cp", proponemos cambiar la descripción del campo "cp" y decidir si usar CPA o CP para estandarizar la información de la columna.

### **Problema 2: Consistencia geográfica ("localidad", "provincia", "departamento")**

- *Atributo de calidad afectado :* Consistencia estructural
- *Problema asociado :* Instancia
- *GQM :*
  - ❖ *Goal :* Verificar la coherencia de las columnas "localidad", "provincia" y "departamento"
  - ❖ *Question:* ¿Hay localidades con mismo nombre en distintas provincias?

¿Coincide cada localidad con su departamento y provincia?

¿Existen combinaciones provincia + localidad que aparezcan con más de un departamento?

❖ *Metric* : Porcentaje de inconsistencia y casos selectos problemáticos

- *Resultados:*

- ❖ **2.05%** de las localidades repetidas en distintas provincias (“Santa Lucia” esta en 4)

- ❖ **2.31%** inconsistencias de “localidad” en distintos (“departamento”, “provincia”) (“San Justo” tiene 6 combinaciones)

- ❖ **5** provincias tienen localidades asignadas a múltiples departamentos (“San Juan” tiene 3 ciudades con misma localidad y 3 departamentos dentro)

- *Diagnóstico:* Dificulta el análisis geográfico preciso, dando cierta ambigüedad. Es importante notar que a pesar de que el 2.31% es un porcentaje bajo, CABA tiene **15 localidades con inconsistencias geográficas**, lo que distorsiona un análisis territorial.

- *Mejoras propuestas:* Una solución, para el problema de instancia en la consistencia estructural de las columnas mencionadas, puede ser crear claves únicas compuestas. Mejorando entonces la calidad estructural para detectar de manera más simple errores de instancia como los encontrados.

### Problema 3: Columnas Irrelevantes o vacías

- *Atributo de calidad afectado* : Relevancia

- *Problema asociado* : Modelo

- Observación:

- ❖ **5** columnas están completamente vacías (“piso”, “web”, “informacion\_adicional”, “observación” y “subcategoría”)

- ❖ **4** columnas tienen un valor único repetido

- En **tipo\_latitud\_longitud** ( el dato es: “Exacta”). En **fuentes** ( el dato es: “CONABIP”) .

- En **anio\_actualizacion** ( el dato es: “2023”). En **categoría** ( el dato es: “Biblioteca Popular”)

- *Mejoras propuestas:* Eliminar columnas vacías o redundantes en futuras exportaciones

### +Observaciones:

Si continuamos el análisis, vemos claros problemas de *instancia*, afectando a la completitud de columnas como lo son “mail” que tiene 880 Nulls y también como “cod\_tel”(código de teléfono) y “teléfono” que tienen 165 Nulls respectivamente. Como ya hicimos antes el mismo GQM, decidimos resolver otros problemas.

Luego, la única columna no trivial de las columnas repetitivas puede ser “anio\_actualizacion”, ya que si se actualiza por bibliotecas individualmente puede ser útil. Esto refleja un problema de *proceso* dado que se carga el dato con una asunción y se usa con otra, en este caso con ninguna.

En vez de escribir “Fuente consultada en la construcción del dato” para “fuente”, es preferible aclarar que se consultó el CONABIP<sup>2</sup>.

---

<sup>2</sup> Ministerio de Cultura de Argentina. 2023. "Mapa Cultural de Espacios Culturales." datos.gob.ar.  
[https://datos.gob.ar/dataset/cultura-mapa-cultural-espacios-culturales/archivo/cultura\\_01c6c048-dbeb-44e0-8efa-6944f73715d7](https://datos.gob.ar/dataset/cultura-mapa-cultural-espacios-culturales/archivo/cultura_01c6c048-dbeb-44e0-8efa-6944f73715d7).

## Sección Decisiones tomadas

En esta sección vamos a hablar de las decisiones que tuvimos que tomar a la hora de realizar el trabajo. Para empezar, nos vimos obligados a elegir un atributo que en mayor medida se comparta en los tres archivos originales. Nuestra decisión aquí fue elegir los códigos de departamento que aparecen en los tres archivos. Había alrededor de 20 departamentos/comunas en los cuales el código de departamento variaba según la base de datos. Como medida de uniformidad en este caso decidimos tomar los códigos de departamento (`id_departamento`) del archivo de los “establecimientos educativos” y trasladarlos a los demás archivos. Esto se debe a que BP (Bibliotecas populares) tenía más coincidencias con EE (establecimientos educativos) que con la base de datos Población.

A pesar de haber uniformado, seguía habiendo ciertos valores que no estaban coincidiendo. Esto pasaba, porque existían EE's en la Antártida, donde no había habido censo, por lo que tomamos la decisión de borrar tales EE's de la tabla.

Luego, las BP de CABA no tenían ninguna comuna anexada, por lo que decidimos (con la columna “domicilio”) buscar a qué comuna pertenecía cada una. Eso lo hicimos usando el mapa interactivo del gobierno de la ciudad de Buenos Aires<sup>3</sup>.

Nuestra próxima decisión tuvo que ver con el armado de los gráficos. Al armar los gráficos 2 y 4, vimos que las escalas eran imprevistas a lo que estábamos acostumbrados y creímos que un “scatter” solo no era una solución óptima. Por esta razón, decidimos hacer más de un gráfico, uno por provincia, para no perder la calidad del análisis que conseguimos mediante la limpieza y las consultas. Consideramos que la mejor opción hubiera sido armar un gráfico interactivo con slider pero aún no tenemos las habilidades para programarlo.

Decidimos no incluir la consulta 4 (iv) en el análisis, porque no creíamos que aportaba para responder el objetivo.

## Sección de Análisis de datos

Analicemos primero los gráficos.

### Primer gráfico :

Primero tenemos un gráfico de barras que muestra la cantidad de bibliotecas populares por provincia (ver *Figura 3*). De aquí podemos extraer que la cantidad de bibliotecas populares es bastante baja a lo largo del país, ya que sacando a Buenos Aires, Santa Fe y Córdoba, todas las demás provincias tienen menos de 100 bibliotecas populares. Números muy bajos teniendo en cuenta que la población total del país ronda los 45 millones.

Dato de curiosidad (usando consulta 2 SQL): Agarramos las provincias en el mismo orden previo, y nos fijamos para cada una, el porcentaje de bibliotecas populares creadas después de 1950 con respecto al total. (ver *Figura 4*)

---

<sup>3</sup> Gobierno de la Ciudad de Buenos Aires. (s.f.). *Mapa interactivo* [Herramienta digital]. Recuperado de: <https://mapa.buenosaires.gob.ar/comollego/?lat=-34.620000&lng=-58.440000&zl=12&modo=transporte>



Vemos que las provincias con mayor porcentaje son: Neuquén, Formosa y Santa Cruz. Esto se puede deber a los planes que hubo alrededor de los años 80 y 90 con el objetivo de modernizar y expandir la red de bibliotecas, y a su vez garantizar que todos puedan tener acceso a la educación<sup>4</sup>.

Las que tienen menor porcentaje son: CABA y Tucuman. Un motivo de esto podría ser la cantidad de bibliotecas históricas que hay en CABA y en Tucuman, y que hayan priorizado conservar el patrimonio histórico.

## Segundo gráfico:

En este caso tenemos dos versiones con dos tipos de gráficos distintos. Por un lado tenemos un heatmap con la cantidad de estudiantes que *debería* asistir a cada establecimiento educativo separado por nivel educativo (ver *Figura 9*). Mientras que por el otro tenemos gráficos de barras separados por provincia (ver *Figura 10*). *(La totalidad de los gráficos analizados se encuentra en la carpeta Graficos/segundo\_grafico para el heatmap y Graficos/segundo\_grafico\_2.0 para el gráfico de barras, donde allí pueden buscar la provincia que quieran ver si lo desean).*

Lo que podemos ver es que en general, los jardines son los que menor cantidad de niños tienen por establecimiento educativo, seguidos por los primarios y luego los secundarios. Podríamos considerar que esto se debe a que son pocos años que necesita una persona para completar el jardín, sumado al poco espacio que ocupan sus aulas. Lo que tendería a desembocar en una mayor cantidad de jardines al menos comparada con los secundarios. Luego si analizamos la diferencia entre primarios y secundarios, podemos notar que es mucha mayor la cantidad de estudiantes por establecimiento educativo en el nivel secundario que en primario. Si analizamos la consulta 1, qué es la fuente en la que nos basamos para armar estos gráficos, podemos notar que la cantidad de establecimientos educativos para el nivel secundario es mucho menor a la cantidad de nivel primario. Mientras que en cuanto a población están bastante parejos. Podríamos considerar que esto se debe a que la cantidad de profesores capacitados para dictar materias en el nivel secundario es menor a la cantidad de maestros capacitados para dictar materias a nivel primario. Desde ya no tenemos los datos correspondientes para fundamentar esta suposición por lo que lo dejamos como un comentario y no una conclusión.

## Tercer gráfico:

En este gráfico de tipo boxplot podemos ver la cantidad de establecimientos educativos separados por provincia (ver *Figura 11*). Lo que podemos extraer de aquí es que la altura y el tamaño de cada caja es distinto. Esto nos habla tanto de una desigualdad entre provincias como de una desigualdad entre los propios departamentos que componen cada provincia. Por ejemplo hay cajas como la de Buenos Aires, Chubut y Santa Cruz que son bastante amplias mientras que provincias como la Ciudad de Buenos Aires, San Luis y Neuquén poseen cajas más compactas. Otro dato que vale la pena mencionar es que casi todas las provincias poseen outliers, de hecho solo 5 de las 24 no tienen outliers. Nuevamente esto indica que hay una fuerte desigualdad entre departamentos. Podría deberse a diferencia tanto territorial como poblacional, pero también podría estar ligado a desigualdades económicas entre provincias y mismo entre departamentos.

## Cuarto gráfico:

En este caso nuevamente tenemos varias versiones de gráficos que nos ayudarán a comprender distintos aspectos de lo que está pasando. El primero es un gráfico de puntos acerca de la cantidad de

---

<sup>4</sup> Congreso de la Nación Argentina. (1986). *Ley N.º 23.351 de Bibliotecas Populares*. Boletín Oficial de la República Argentina. Recuperado de <https://www.argentina.gob.ar/normativa/nacional/ley-23351-23024/texto>

establecimientos educativos (EE) y bibliotecas populares (BP) cada mil habitantes separado por provincia (ver *Figuras 5-8*). Cuenta con regresión lineal para analizar la tendencia en la relación entre estos dos parámetros. (Si desean ver alguna provincia en particular, los gráficos se encuentran en la carpeta *Graficos/cuarto\_grafico\_2.0*). El segundo es un gráfico de líneas donde podemos ver los mismos parámetros pero con los departamentos ordenados por “EE por mil habitantes” de manera ascendente, y regresión lineal sobre las BP para analizar su tendencia (ver *Figura 12*). Por último tenemos un gráfico con todas las subdivisiones de Argentina (departamentos, comunas y partidos) en escala logarítmica para compactar las distancias reales de un gráfico estándar pero mantener las distancias relativas (ver *Figura 13*).

Si miramos la primera versión, podemos notar que existe cierta relación entre la cantidad de EE y BP ya que en 18 de las 24 provincias la regresión lineal es una recta creciente. Esto nos indica que a medida que hay más EE, podemos encontrar más BP. En los demás casos como por ejemplo Córdoba, Corrientes o Formosa, notamos que las rectas son horizontales o decrecientes. Quiere decir que en ciertas provincias no existe tal relación entre la cantidad de EE y BP, pero siendo una minoría de casos habría que ver que pasa si analizamos todo junto. En cuanto a la segunda versión, podemos extraer algo similar a lo que vimos antes ya que en la mayoría de los casos la tendencia es el que sí existe una relación la cantidad de EE y BP. La diferencia aquí es que podemos ver mejor la disparidad que hay entre departamentos, relacionado a lo que hablamos en el apartado del gráfico 3. Podemos ver que los puntos están muy dispersos entre sí, debido a una notoria desigualdad entre departamentos dentro de las mismas provincias. Por último, en la tercera versión podemos ver que hay una clara relación entre la cantidad de EE y la cantidad de BP ya que en este gráfico podemos ver una especie de nube que tiende a ir del lado inferior izquierdo al lado superior derecho. Sugiriendo que las tendencias que vimos antes por provincia realmente reflejan una relación entre los dos parámetros que comparamos en esta instancia. Si bien es cierto que mirando cada provincia por separado encontramos que la tendencia general era que existe una relación, no es menos importante aclarar que esta generalización no se extiende a cada provincia en particular, como vimos con las dos primeras versiones de este gráfico.

Continuemos con los reportes:

### **Primer Reporte:**

Este reporte contiene información detallada sobre establecimientos educativos por departamento, con datos separados por nivel educativo (Jardín, Primario y Secundario) incluyendo la cantidad de establecimientos educativos y la población estudiantil, ambos por nivel.

Este reporte es fundamental para analizar la distribución de la infraestructura educativa. También nos proporciona datos para comparar con las bases de bibliotecas populares, lo que permite un análisis por un nivel educativo de ser el objetivo.

### **Segundo Reporte:**

Este reporte contiene información sobre la cantidad de bibliotecas públicas fundadas desde 1950 por provincia y departamento.

Este reporte nos proporciona otro componente clave para la pregunta general del trabajo, además de que nos permite estudiar la distribución temporal. De hacerlo podemos ver que 1264 de las 1903 son fundadas del año 1950 para adelante, lo que implica que hubo un incentivo nacional a la creación de estos establecimientos populares. Interesados en ver cierta relación cuantitativa, podemos ver que en ciertas jurisdicciones hay gran desigualdad territorial, como ocurre en Neuquén. Donde en Confluencia fueron creadas 39 BP y en el resto de departamentos la distribución es similar, pero con gran reducción en cantidad, siendo Chos Malal, el departamento que le sigue en número, con tan solo 4 fundadas después de 1950. En otras jurisdicciones, como Jujuy, CABA y Misiones hay una distribución más equitativa

### **Tercer Reporte:**

Este reporte contiene información sobre la población, cantidad de bibliotecas populares y establecimientos educativos por cada departamento de cada provincia.

Para analizar este reporte y poder obtener conclusiones, vamos a realizar un gráfico de la población con respecto a la infraestructura educativa. De esta manera, vamos a poder identificar si se cumple que mientras mayor es la infraestructura educativa, mayor es la población y viceversa. (Ver *Figura 14*)

Analizando el gráfico, lo primero que observamos es que hay una gran cantidad de departamentos cuya población es menor a 250.000 personas. Más allá de esto, se puede ver a partir de la distribución de los puntos en el gráfico, que en general se cumple que los departamentos con mayor población suelen tener una mayor infraestructura educativa.

### **Cuarto Reporte:**

Este reporte contiene información sobre qué dominio de mail se usa más para las bibliotecas populares de cada departamento.

Vemos que no hay un dominio estándar y que varía mucho por departamento, habiendo casos incluso donde no hay un dominio. Esto complica las comunicaciones oficiales porque las hace ineficiente.

Para nuestro caso concreto, este reporte no va a ser de mucha utilidad.

### **Extra:**

Con el fin de evaluar ciertos datos cuantitativos, haciendo uso de lo visto en clase, intentamos cuantificar las relaciones entre las variables involucradas<sup>5</sup>. Si calculamos el ratio (relación) BP/EE podemos llegar a tener una métrica que representa el porcentaje general. Teniendo en cuenta que hay 64721 EE y 1903 BP, tenemos un promedio de 0.03% de promedio nacional BP/EE. Lo que indica que hay 3 bibliotecas populares cada 100 establecimientos educativos. Las provincias en su mayoría poseen un promedio dentro de todo similar a este, quizá exceptuando Chubut (0.06%) con una variabilidad de 0.03% de diferencia

Si tomamos los *cuartos gráficos* la variabilidad es alta dependiendo del departamento en que se encuentren, tal que un departamento puede tener 0.48% de diferencia en BP cada mil habitantes con otro de su misma provincia. Como ocurre con Jujuy en Susques (0.50% BP/1000) y Doctor Manuel Belgrano (0.02% BP/1000) Además, muchas de las jurisdicciones muestran gran diferencia en la distribución territorial. Si tomamos el *tercer reporte* podemos ver que hay departamentos con muchos EE's y pocas BP's relativamente, como ocurre en el caso de la Matanza en Buenos Aires (1201 EE's y 16 BP's) o General Güemes en Chaco (409 EE's y 1 BP). Otras tienen proporciones más equilibradas, como Marcos Juárez en Córdoba (204 EE's y 11 BP's) o Iriondo en Santa fe (143 EE's y 12 BP's).

Los datos extraídos presentan evidencia cuantitativa que puede ser útil a la hora de tomar futuras decisiones en el plano educativo.

---

<sup>5</sup> Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. (2025). *Laboratorio de Datos: Introducción a Modelado* [Presentación de PowerPoint]. Departamento de Computación

## Sección de Conclusiones

Si hacemos uso de los gráficos e informes, podemos ver que hay muy pocas bibliotecas en relación a la cantidad de establecimientos educativos. Esto hace que el acceso a las bibliotecas sea limitado para las comunidades educativas en la mayoría de los departamentos analizados. El hecho de que haya tanta variabilidad en los departamentos refleja desigualdades territoriales, poblacionales y económicas. Pero volviendo a la pregunta original que queríamos responder que decía “*¿Existe alguna relación entre la cantidad de establecimientos educativos en cada una de las provincias y la cantidad de Bibliotecas Populares?*”. La respuesta es *depende*. Como vimos previamente, en la mayoría de las provincias, de hecho en 18 de 24, sí existe una relación entre la cantidad de establecimientos educativos y la cantidad de bibliotecas. Sin embargo, también vimos que en otros casos esto no es así. En Córdoba, La Rioja, Corrientes, Santiago del Estero, Tierra del Fuego y Formosa, no podemos determinar que exista realmente esta relación. Ahora bien, si generalizamos y miramos a todo el país como conjunto, podríamos decir que esta relación existe. Lo que no implica que se traslade a cada provincia en particular. Por lo tanto, consideramos que la respuesta no podría ser un simple “*sí*” o un simple “*no*” ya que en este caso como en muchos otros, la realidad es mucho más compleja.

## Bibliografía:

- **Ministerio de Cultura de Argentina.** 2023. *Mapa Cultural de Espacios Culturales*. datos.gob.ar. Recuperado de: <https://datos.gob.ar/dataset/cultura-mapa-cultural-espacios-culturales>
- **Gobierno de la Ciudad de Buenos Aires.** (s.f.). *Mapa interactivo* [Herramienta digital]. Recuperado de: <https://mapa.buenosaires.gob.ar>
- **Elmasri, R., & Navathe, S. B.** (2007). *Fundamentos de Sistemas de Bases de Datos* (5ª ed.). Pearson Educación
- **Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.** (2025). *Laboratorio de Datos: Introducción a Modelado* [Presentación de PowerPoint]. Departamento de Computación.
- **Congreso de la Nación Argentina.** (1986). *Ley N.º 23.351 de Bibliotecas Populares*. Boletín Oficial de la República Argentina. Recuperado de <https://www.argentina.gob.ar/normativa/nacional/ley-23351-23024/texto>

# Anexo:

Figura 1: Diagrama DER que representa la estructura de datos de un sistema que incluye las entidades: Departamento, Biblioteca Popular, Provincia, Establecimiento Educativo y Poblacion.

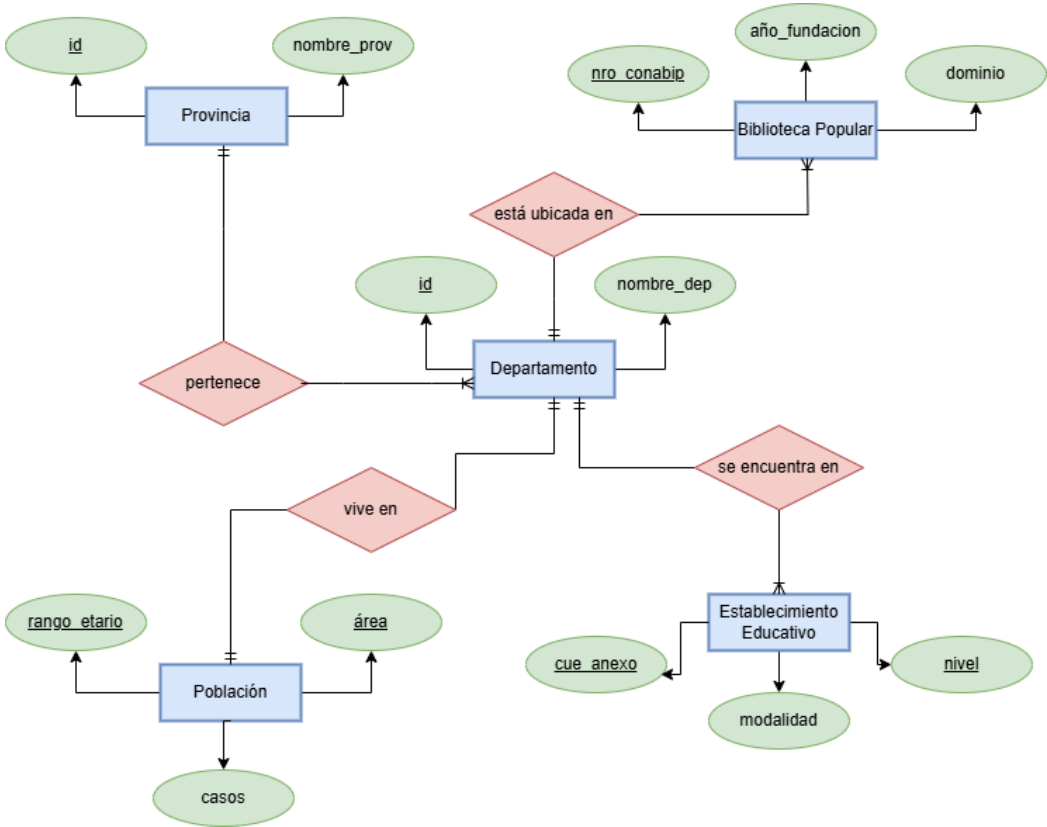
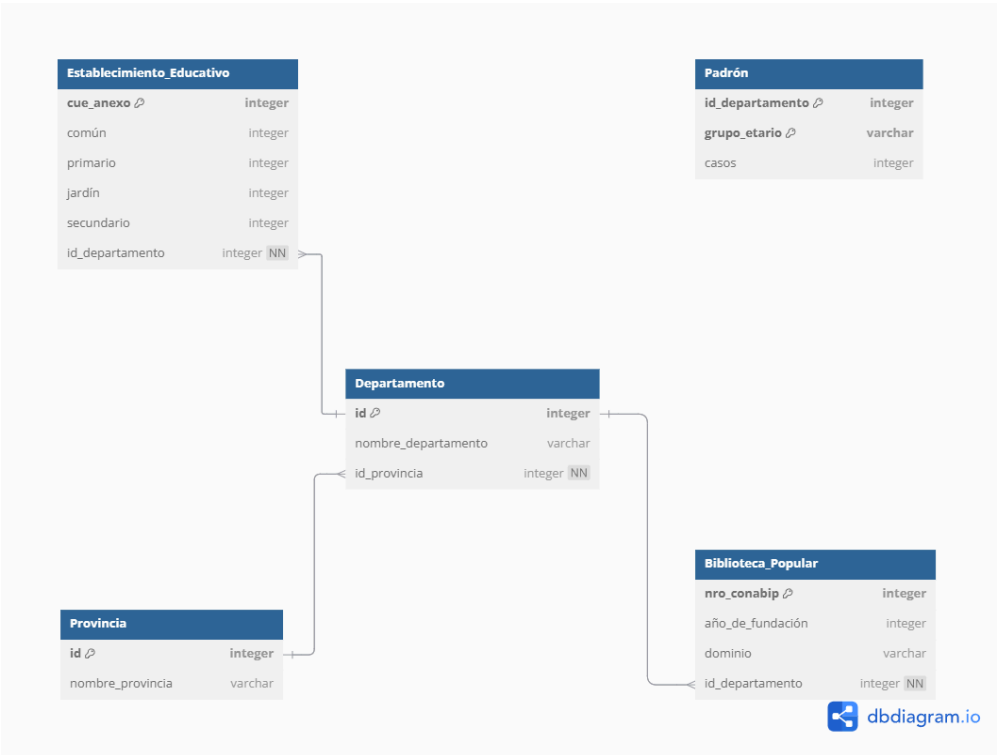
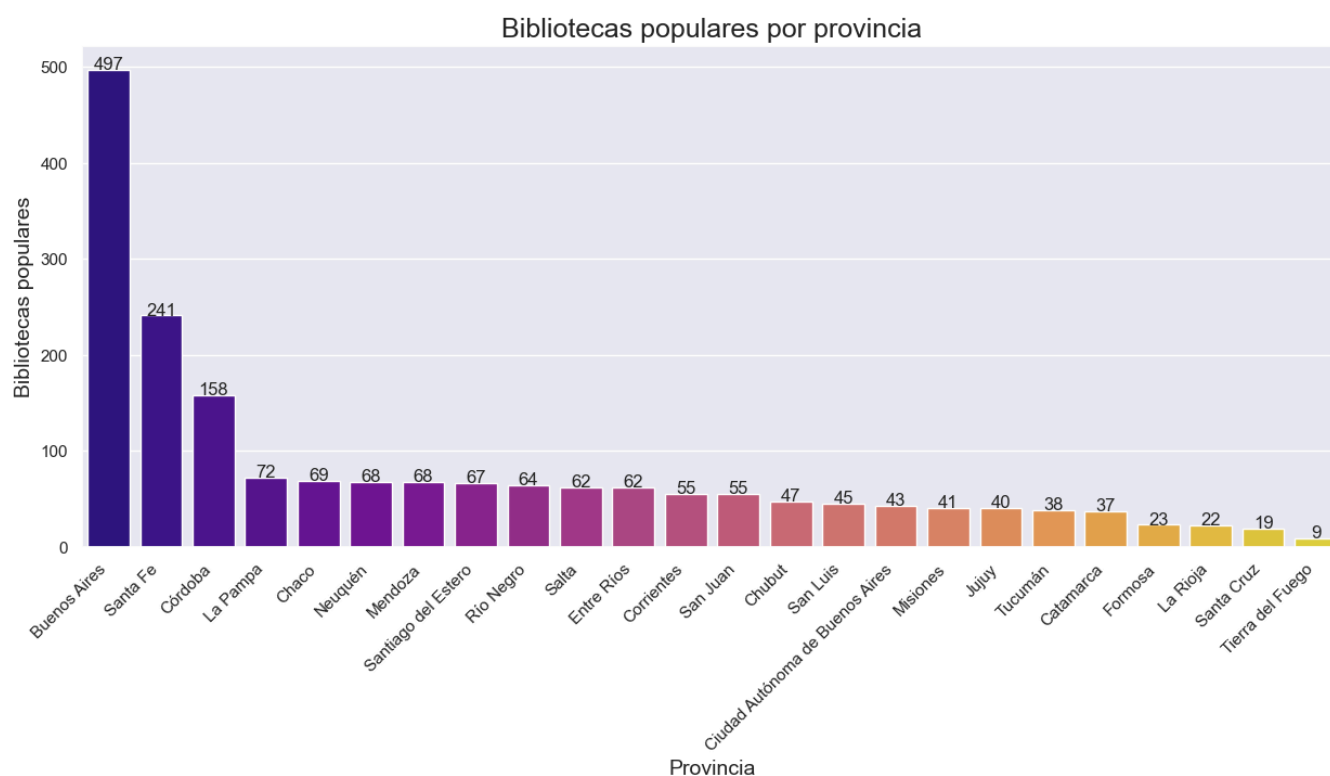


Figura 2: Modelo Relacional que representa las entidades del sistema de la Figura 1. Cada tabla contiene claves primarias y algunas foráneas, que permiten establecer relaciones entre ellas.



**Figura 3 : Gráfico que muestra la cantidad de bibliotecas populares en las distintas provincias utilizando un gráfico de barras y ajustado con una paleta de colores accesible para personas con daltonismo**



**Figura 4: Gráfico que muestra el porcentaje de Bibliotecas populares creadas después de 1950. Se calculó haciendo:  $100 * BP \text{ creadas después de } 1950 \text{ por provincia} / \text{Cantidad Total de BP por provincia}$**

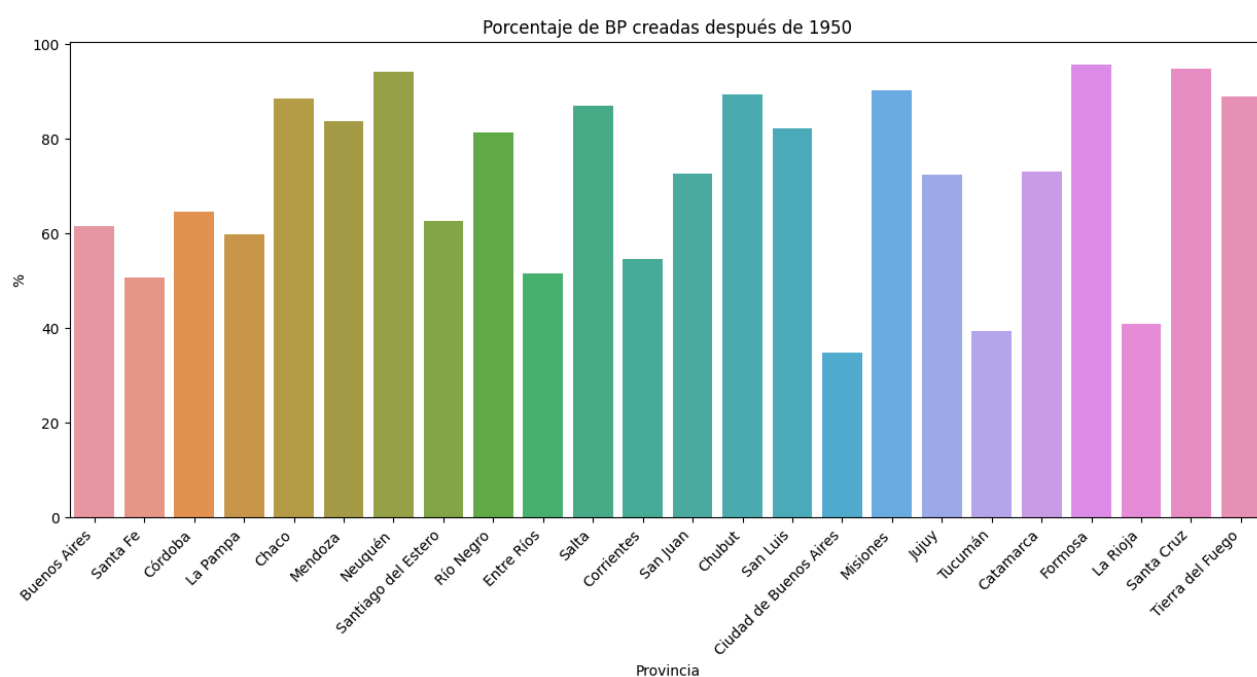


Figura 5 : Gráfico de relación entre bibliotecas populares (BP) y establecimientos educativos (EE) cada 1000 habitantes en La Pampa. Es un diagrama de dispersión con línea de regresión lineal ascendente. Cada punto representa un departamento.

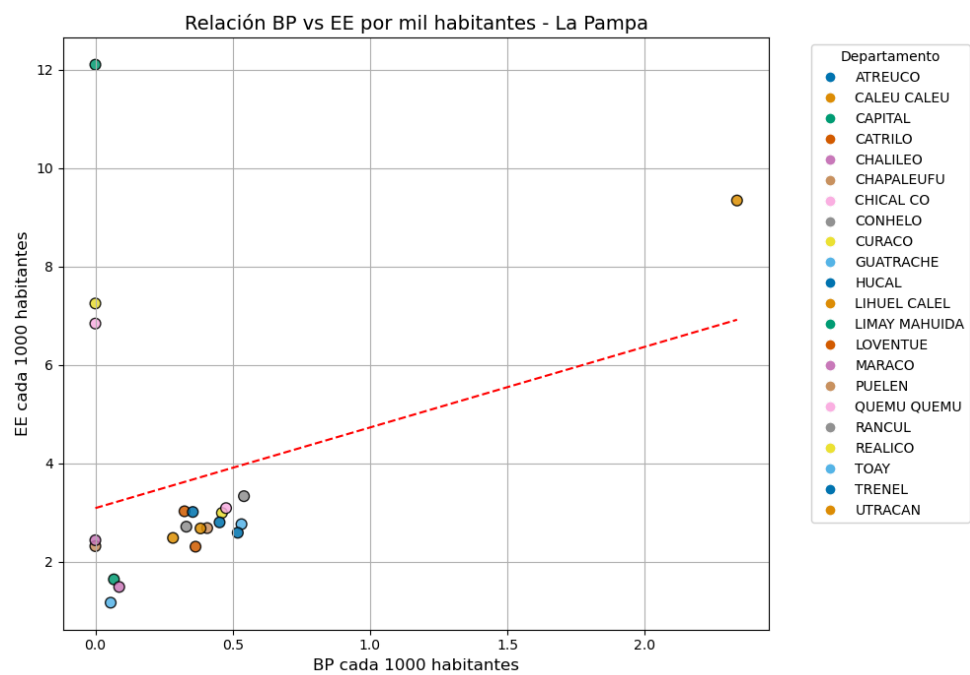


Figura 6: Gráfico de relación entre bibliotecas populares (BP) y establecimientos educativos (EE) cada 1000 habitantes en Salta. Es un diagrama de dispersión con línea de regresión lineal ascendente. Cada punto representa un departamento.

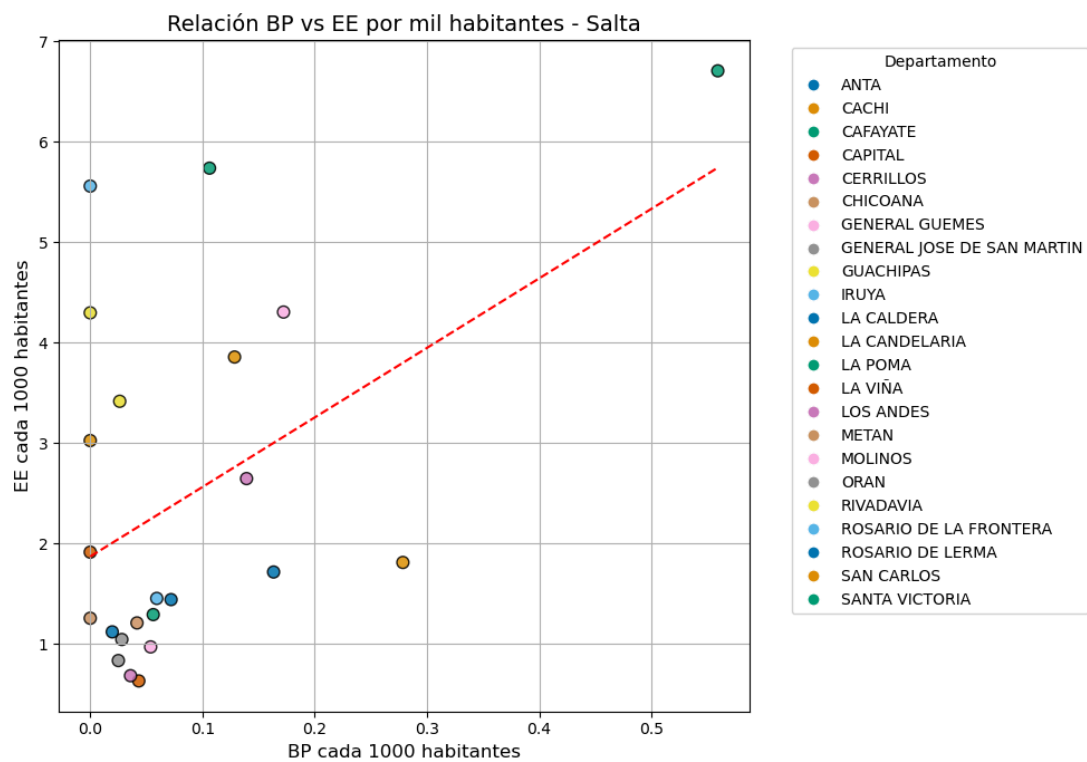




Figura 7: Gráfico de relación entre bibliotecas populares (BP) y establecimientos educativos (EE) cada 1000 habitantes en Mendoza. Es un diagrama de dispersión con línea de regresión lineal ascendente. Cada punto representa un departamento.

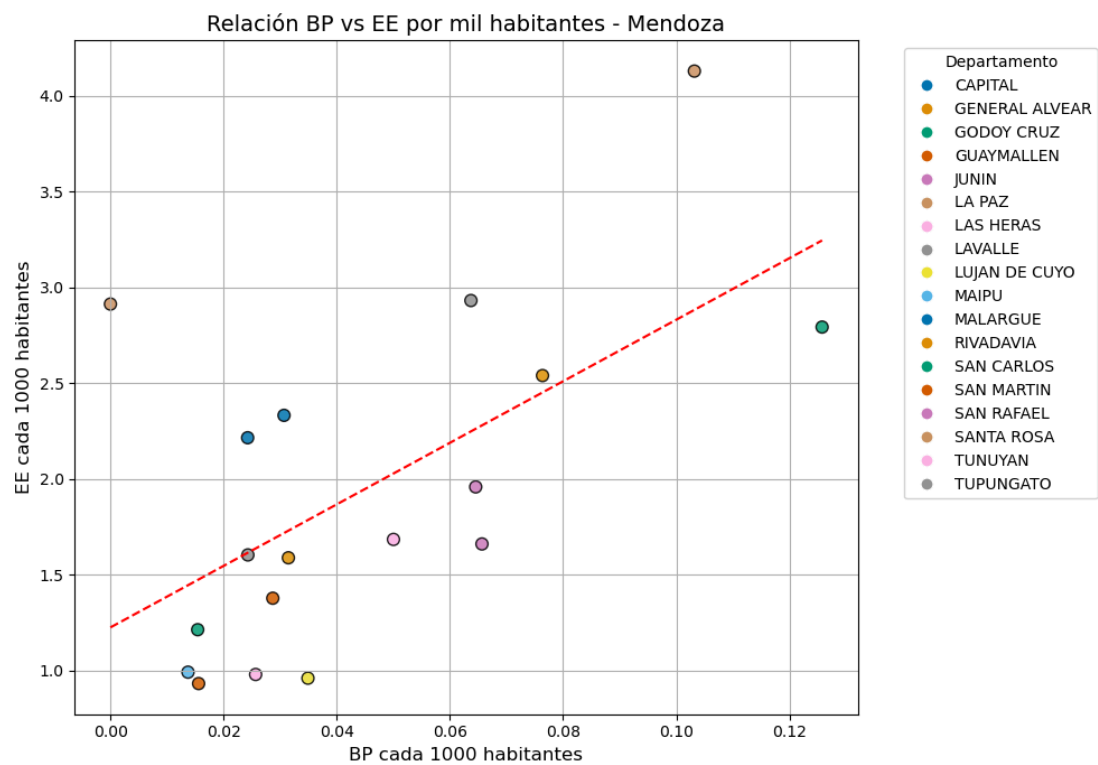


Figura 8: Gráfico de relación entre bibliotecas populares (BP) y establecimientos educativos (EE) cada 1000 habitantes en Chaco. Es un diagrama de dispersión con línea de regresión neutra tendiendo a ascendente. Cada punto representa un departamento.

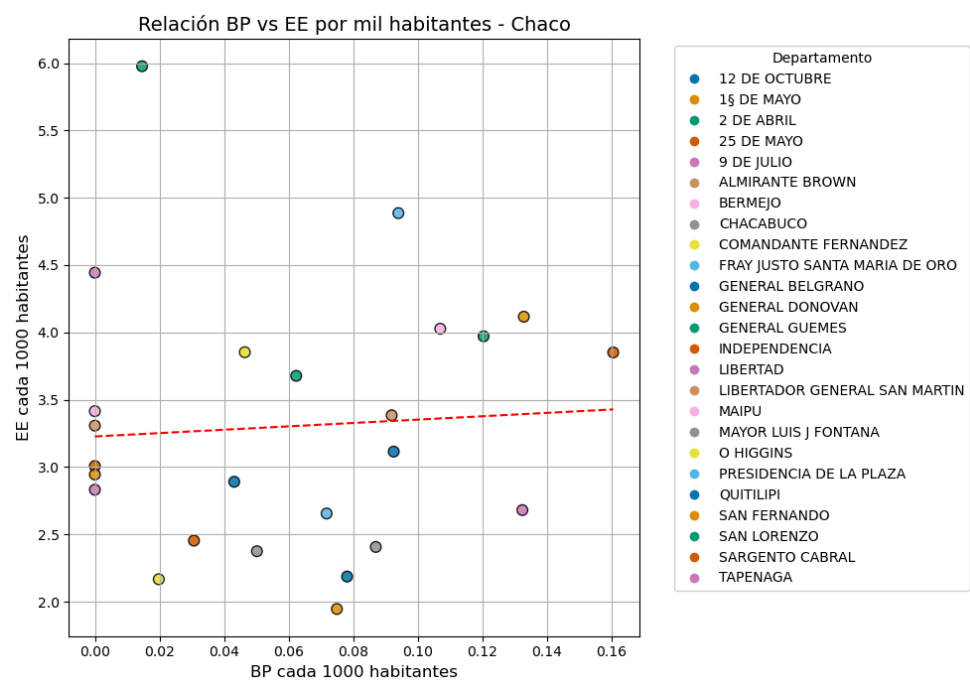


Figura 9: Gráfico de tipo heatmap que representa la cantidad de establecimientos educativos en función de la población según nivel educativo para Córdoba, Entre Ríos y Río Negro.

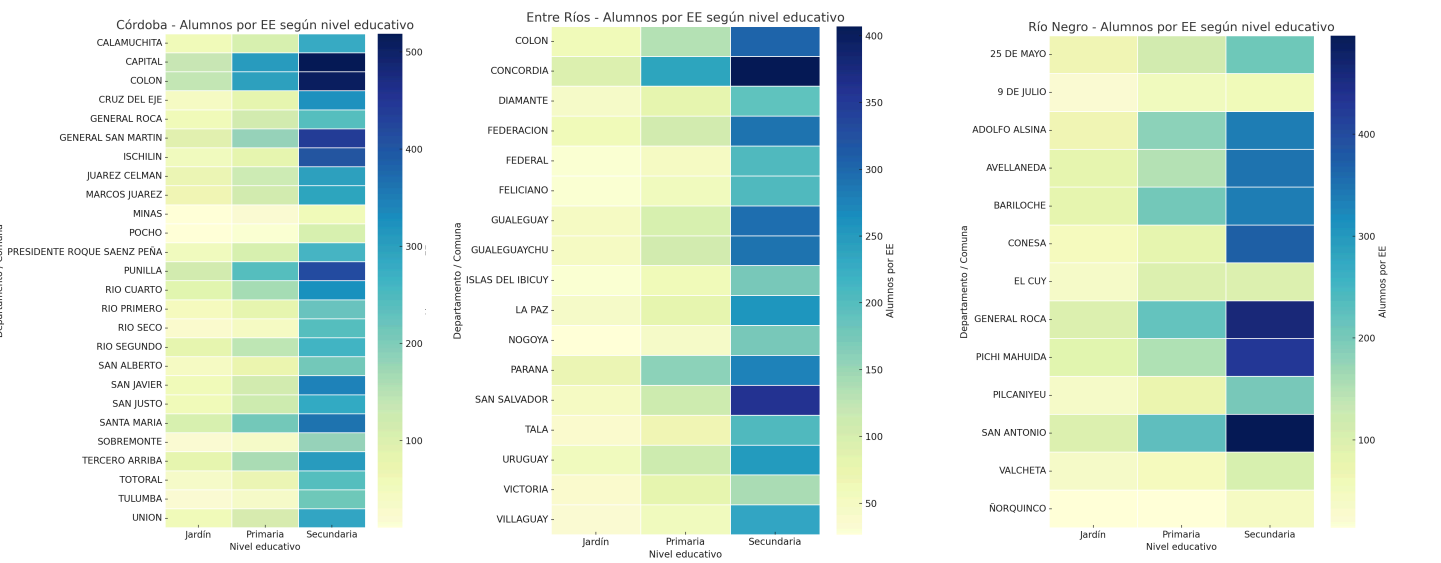


Figura 10: Gráfico de barras que representa la cantidad de establecimientos educativos en función de la población según nivel educativo para Córdoba, Entre Ríos y Río Negro.

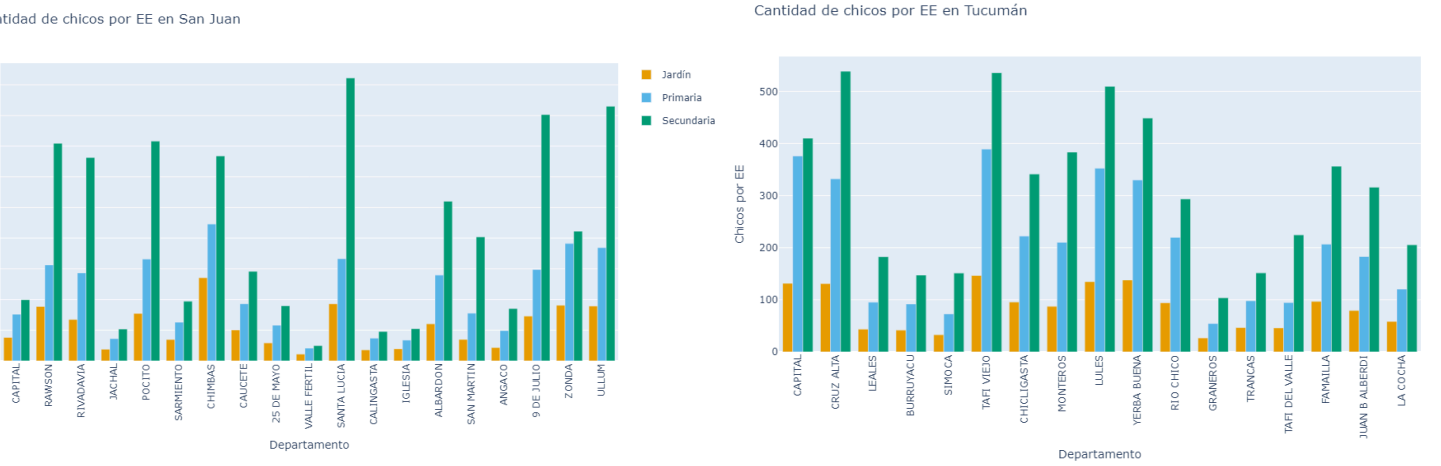


Figura 11: Gráfico de distribución de los establecimientos educativos (EE) por provincia. Es un diagrama de caja con puntos de dispersión que se superponen. Cada punto representa un departamento.

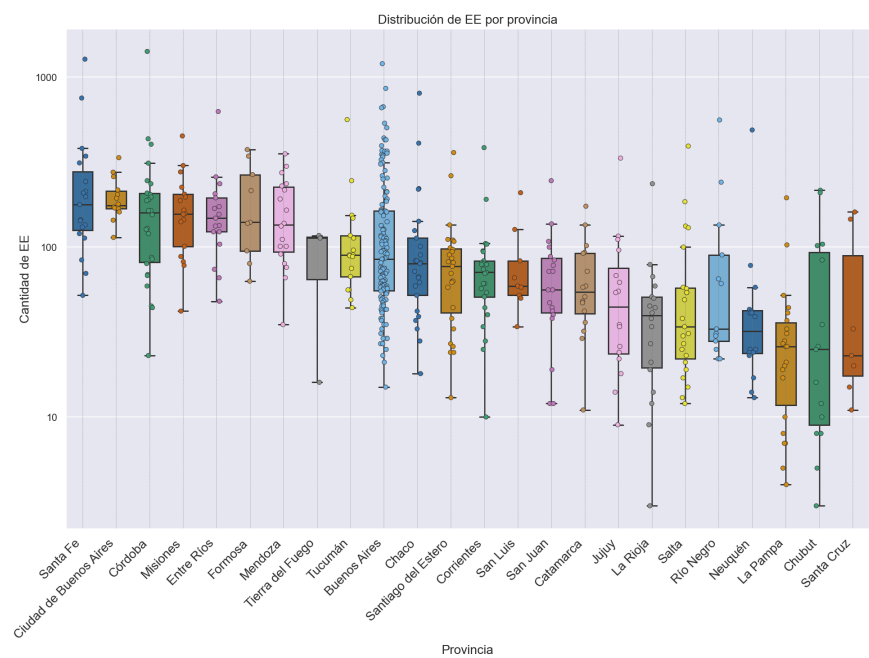


Figura 12: Gráfico de líneas acerca de la relación entre bibliotecas populares (BP) y establecimientos educativos (EE) cada 1000 habitantes de Buenos Aires y Neuquén.

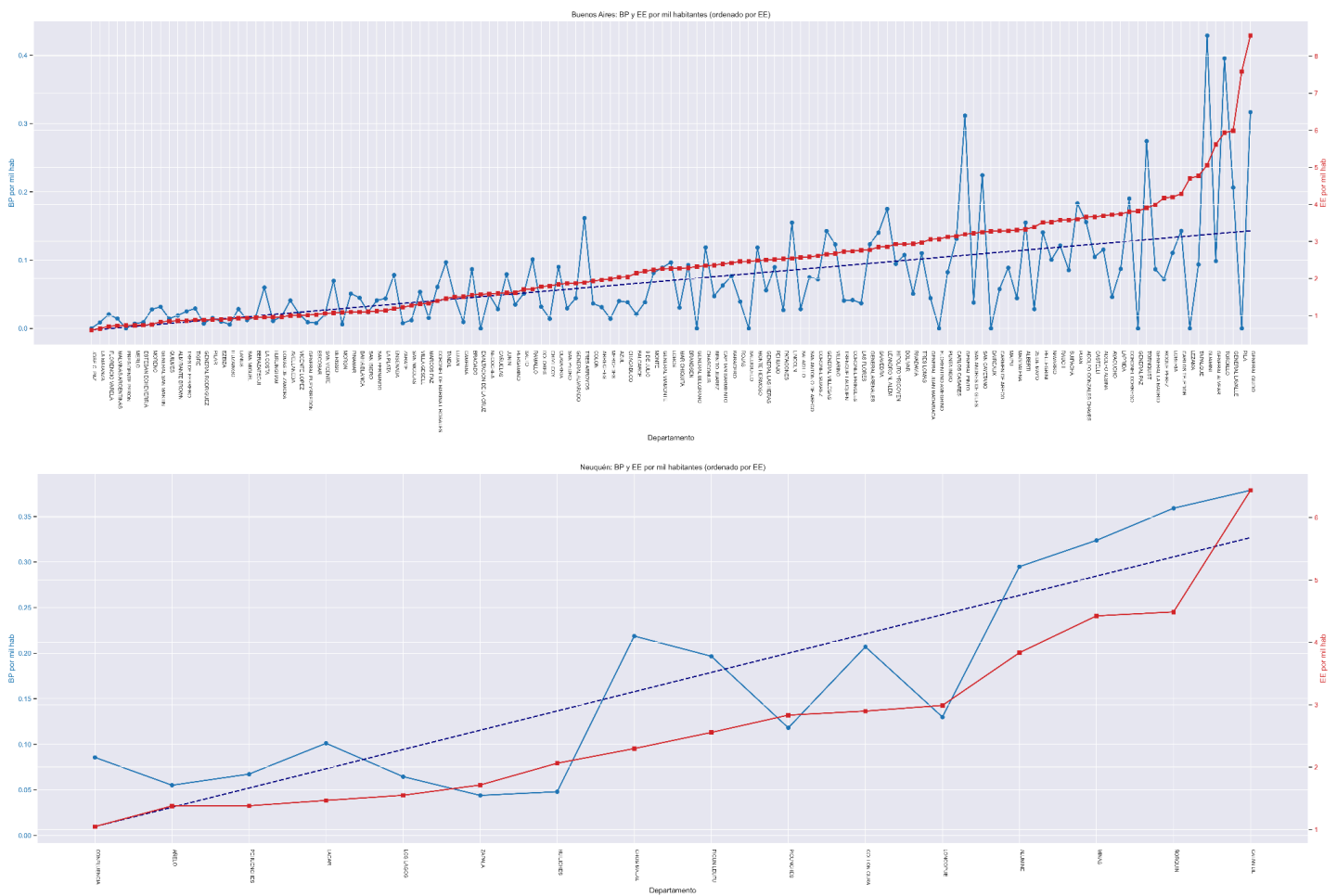


Figura 13: Gráfico de relación entre bibliotecas populares (BP) y establecimientos educativos (EE) cada 1000 habitantes de Argentina con escala logarítmica.

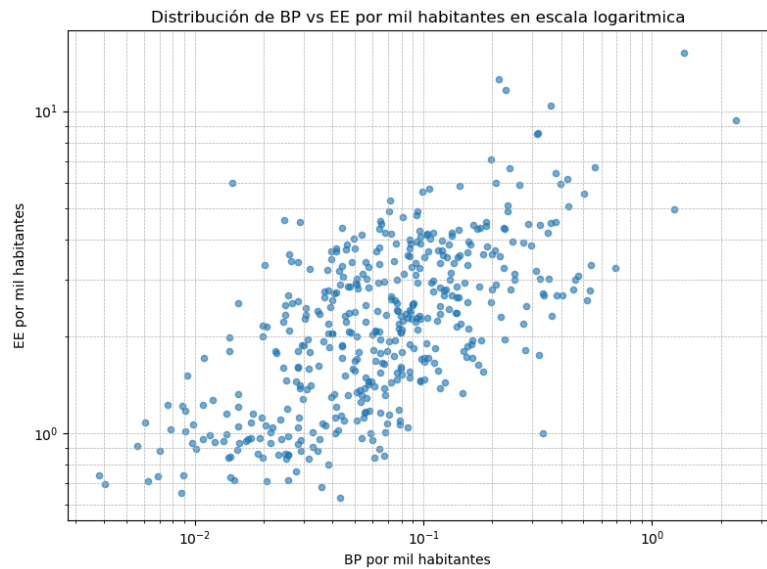


Figura 14: En el eje x se encuentra la suma de Bibliotecas populares y Establecimientos educativos, la llamamos infraestructura educativa, y en el eje y la población total en escala logarítmica. Cada punto representa un departamento.

