

TRABAJO PRÁCTICO 01

Laboratorio de Datos
DC - 1er Cuatrimestre 2025

GRUPO “AMICHIS LABO”

- FRANCISCO MARGITIC
- ALESSIA LOURDES LIMA
- KATERINA LICHTENSZTEIN



Sección Resumen

A partir de la fuente de datos abiertos correspondientes a los Establecimientos Educativos y las Bibliotecas Populares de la República Argentina, surge el interés de analizar una posible relación entre ellas. Para esto consideramos la información que proviene de las mismas más los Datos de población por Departamento obtenidos a partir del Censo de 2022, específicamente en la sección de Estructura por edad de la población.

Para abordar esta cuestión llevamos a cabo un trabajo que consta de distintos reportes, los cuales relacionan las fuentes de datos mencionadas anteriormente. Previamente, realizamos un análisis de calidad, una limpieza y una selección de aquellos datos originales que consideramos relevantes para el objetivo planteado, descartando aquellos que no aportan información significativa sobre la cuestión principal.

La conclusión demuestra que no es posible determinar una relación entre la cantidad de establecimientos educativos y la cantidad de bibliotecas públicas utilizando solamente los datos proporcionados.

Sección Introducción

El punto de partida de este trabajo surge de un problema inicial: contamos con distintas fuentes de datos que, a priori, no evidencian una relación clara entre sí y presentan diversas deficiencias en términos de calidad de datos. Esta situación nos lleva a plantear un objetivo general: determinar si existe o no una relación entre la cantidad de establecimientos educativos y bibliotecas populares en cada provincia y, en caso afirmativo, identificar cuál es esa relación y qué aspectos describe. Este objetivo nos servirá como guía y fundamentará las decisiones tomadas a lo largo del proceso de análisis.

Para alcanzar dicho fin comenzamos explorando el contenido de las fuentes de datos con las que contamos, comprendiendo qué datos aporta cada una y qué información se puede deducir a partir de sus datos en conjunto. Durante este proceso identificamos información innecesaria para el análisis que llevaremos a cabo, así como estructuras con “malos” diseños que permiten falencias en aspectos importantes para la calidad y el análisis de los datos. Para mitigar estos problemas generamos nuestras propias fuentes con una estructura optimizada a partir de un proceso de selección de datos provenientes de las fuentes originales.

La definición de esta nueva estructura la planteamos a través de un diagrama conceptual que organiza los datos en entidades junto con sus atributos correspondientes y la relación que hay entre ellas: Diagrama Entidad Relación (DER), considerando únicamente aquellos datos necesarios para nuestro objetivo. A partir del DER, implementamos un modelo relacional alimentado con los datos ya “limpios”, es decir, aquellos que ya pasaron por un proceso de revisión y análisis de calidad. Este modelo lo diseñamos respetando los principios de la Tercera Forma Normal (3FN), asegurando integridad y coherencia en el almacenamiento de los datos.

Una vez alimentados los esquemas de datos, generamos cuatro reportes clave alineados con nuestro objetivo, utilizando consultas SQL. Luego, empleamos herramientas de visualización para representar de forma clara las relaciones más relevantes, facilitando así el desarrollo de conclusiones en torno al objetivo general.

Sección Procesamiento de Datos

Para el análisis de las fuentes tomaremos en cuenta los siguientes criterios: Primera Forma Normal (1FN), donde el dominio de un atributo debe incluir sólo valores atómicos, Segunda Forma Normal (2FN) en la que todo atributo no primo depende funcionalmente de manera completa de la clave primaria y cumple 1FN, y Tercera Forma Normal (3FN) que respeta las dos anteriores, además de que ningún atributo no primo debe depender transitivamente de la clave primaria. Finalmente, consideramos que las Formas Normales siguen una relación jerárquica ($1FN \subseteq 2FN \subseteq 3FN$) y continuamos con el análisis de las fuentes de datos.

- **Forma normal de las fuentes de datos**

- Análisis de “Establecimientos Educativos”:

A simple vista notamos que ciertos atributos como *teléfono* y *mail* admiten un conjunto de valores (*teléfono* : “4301-2922/4301-1606”, *mail* : “RECTORIA@INSTITUTOWMORRIS.EDU.AR // DIRECCIONPRIMARIA@INSTITUTOWMORRIS@EDU.AR”). Con esto podemos afirmar que su dominio no está compuesto por valores simples e indivisibles, lo cual infringe el principio de **atomicidad**. Además, esta fuente de datos presenta una estructura con relaciones **anidadas** según la modalidad y los niveles del establecimiento. Dado que la estructura no cumple con 1FN, esto implica que tampoco satisface los criterios de 2FN ni los de 3FN.

Ahondando un poco más en la estructura, aún si estuviese en 1FN, hay atributos no primos que dependen funcionalmente de manera **parcial** de una PK. Por ejemplo, de la $PK=\{\text{Código de localidad, Código de departamento, Domicilio, C.P.}\}$ (asumiendo que no hubiese valores nulos y que los domicilios fuesen únicos para dicho código postal en ese departamento) tenemos una DF que sale únicamente de *Código de localidad* y conecta con *Localidad*, o una DF que parte de *Código de departamento* y determina *Departamento*.

- Análisis de forma normal de Bibliotecas Populares:

Lo mismo sucede con esta estructura, en el atributo *mail* de la biblioteca *nro_conabip* 3900 posee dos valores, uno en mayúscula y otro en minúscula (“sanestebanbibliotecapopular@yahoo.com.ar” <SANESTEBANBIBLIOTECAPOPULAR@YAHOO.COM.AR>”). En conclusión, no está en 1FN, no está en 2FN y tampoco está en 3FN.

Si por un momento pretendemos que está en 1FN, notamos al plantear la $PK=\{id_provincia, domicilio, cp, piso\}$ (asumiendo que no hubiese valores nulos y que en una provincia hubiese un único domicilio y piso para el código postal concreto), tenemos DF que serían parciales: *provincia* podría conectarse únicamente con *id_provincia*; y no estaría en 2FN. Aún si estuviese en 1FN y 2FN, podemos demostrar que no cumple 3FN ya que si determinamos la $PK=\{latitud, longitud\}$, tendríamos una DF que conecte cada uno de los otros atributos. Nuevamente al tomar *id_provincia*, este luego podría determinar *provincia*. Habría varias transitividades de atributos no primos.

- **Análisis de calidad de datos**

Principales problemas identificados en las fuentes de datos:

Valores nulos: como problema asociado al modelo de datos, se detectan valores nulos en diversos atributos que afectan a la completitud, tales como *observación*, *subcategoría*, *piso*, *mail*, *web*, *información adicional* en Bibliotecas Populares, o en los distintos tipos de *modalidades* y *niveles* en Establecimientos Educativos;

Atributos con valores repetidos: por un mal manejo de software se han agregado atributos que presentan un mismo valor en todas las tuplas, causando una cantidad inadecuada de datos repetidos que se vuelven irrelevantes. Por ejemplo en Bibliotecas Populares *categoría* (cuyo valor es siempre “Biblioteca Popular”), *tipo_latitud* (“Exacta”), *fuentes* (“CONABIP”), *año_actualización* (“2023”).;

Falta de estandarización: se observaron inconsistencias causadas por un problema de instancia en la codificación y nomenclatura de datos. La comuna 4 posee el ID “02104001” en Establecimientos Educativos y “02028” en el censo; la localidad de Florida* tiene código de localidad “06861004” en Establecimientos Educativos (inexistente en Bibliotecas Populares) y “06861010” en Bibliotecas Populares (correspondiente a Vicente López en Establecimientos Educativos). Por otro lado, vemos variaciones en la escritura de direcciones, en *domicilio* (Establecimientos Educativos) “AVENIDA SANTA FE” y “SANTA FE”, y un uso inconsistente de abreviaturas, como “Almte” y “Almirante” en Bibliotecas Populares. Esto afecta la consistencia de los datos almacenados;

Desactualización de datos: existe una discrepancia temporal entre las fuentes, afectando la vigencia de los datos. En Bibliotecas Populares el último año de actualización es 2023, sin embargo tanto los datos de Establecimientos Educativos como los del censo corresponden al 2022. Esta diferencia impacta el análisis comparativo ya que la cantidad de población y de establecimientos que variaron en el período no contabilizado de 2022 a 2023 quedan por fuera. Es un problema causado por la instancia;

Atributos irrelevantes para el análisis: se identificó un problema de calidad proveniente de datos que podrían considerarse no esenciales para nuestro objetivo, que afectan la relevancia para representar la realidad, tales como *código postal*, *código de área*, *teléfono*, *latitud*, *longitud*, entre otros.;

Para un análisis formal utilizaremos la métrica Goal Question Metric (GQM).

- 1) G1: Verificar que el dato correspondiente a la fecha de fundación de cada biblioteca popular esté completo.

Q1: ¿Cuál es la proporción de bibliotecas populares que tienen el dato *fecha_fundación* vacío?

```
M1: CantFechaFundacionNull = BP["fecha_fundacion"].isna().sum()
    CantFechaFundacionTotal = len(BP)
    print((CantFechaFundacionNull/CantFechaFundacionTotal)*100)
```

Valor obtenido: 2.10%

Diagnóstico y posibles acciones de mejora: problema modelo admite valores nulos que podríamos completar con una nueva búsqueda a partir de otras fuentes, pero al ser un porcentaje bajo decidimos no tomar esta acción ya que llevaría más tiempo

del que contamos. En su lugar, nos enfocaremos únicamente en las bibliotecas populares que muestren una fecha de fundación para el reporte solicitado.

- 2) G2: Verificar que las claves de las localidades sean consistentes en las distintas fuentes de datos.

Q2: ¿El *cod_localidad* de Bibliotecas Populares es el mismo que *Código de localidad* de Establecimientos Educativos para una misma Localidad?

M2:

```
biblio.sort_values(by='localidad', ascending = True)  
escuelas.sort_values(by='Localidad', ascending = True)  
  
for localidad in escuelas['Localidad'] do  
comparamosLocalidad = escuelas['Localidad'].equals(biblio['localidad'])  
comparamosCodigos = escuelas['Código de  
localidad'].equals(biblio['cod_localidad'])  
  
if comparamosLocalidad == True:  
    if comparamosCodigos == False:  
        return False  
    else:  
        res = True  
        return res
```

Valor obtenido: para que el código funcione, tendría que poder identificar localidades que sean la misma por más de que estén escritas de manera distinta (mayús/minús, etc.), o bien primero tendríamos que limpiar ambas fuentes. Sin embargo, omitiremos este proceso por observar a simple vista el caso [Florida*](#) (mencionado anteriormente) que demuestra que no se cumple la igualdad entre el *cod_localidad* de las bibliotecas populares y el *código localidad* de los establecimientos educativos.

Diagnóstico y acción de mejora: solucionamos el problema previamente explicado.

Luego, volvemos a aplicar la métrica a las nuevas tablas con los valores de *localidad* y *Localidad* modificados, y descubrimos que de igual manera nos da False.

Concluimos que los valores de *cod_localidad* y *código localidad* poseen discrepancias que complicarían su uso sin previa corrección, posibilitando la generación de tuplas espúreas. De igual modo, podemos desviar este problema al usar los *id departamento* que poseen tanto padrón Población como padrón Bibliotecas Populares, y luego generar los *id departamento* del padrón Establecimientos Educativos a partir de los valores de código localidad.

3) G3: Verificar la completitud de representación de las provincias en las fuentes de Establecimientos Educativos y Bibliotecas Populares.

Q3: ¿Figuran todas las provincias de Argentina en la columna Jurisdicción de Establecimientos Educativos y en la columna Provincia de Bibliotecas Populares?

M3: `consulta_sql = ""`

```
SELECT DISTINCT "Jurisdicción" FROM escuelas
""
```

```
jurisdicciones_unicas = dd.sql(consulta_sql).df()
```

```
# Convertimos a lista
```

```
lista_jurisdicciones = jurisdicciones_unicas["Jurisdicción"].tolist()
```

```
Provincias = ["Buenos
```

```
Aires", "Catamarca", "Chaco", "Chubut", "Córdoba", "Corrientes", "Entre
```

```
Ríos", "Formosa", "Jujuy", "La Pampa", "La
```

```
Rioja", "Mendoza", "Misiones", "Neuquén", "Río Negro", "Salta", "San Juan", "San
```

```
Luis", "Santa Cruz", "Santa Fe", "Santiago del Estero", "Tierra del
```

```
Fuego", "Tucumán", "Ciudad de Buenos Aires"]
```

```
# Comparar si son iguales (sin importar el orden)
```

```
conjunto_jurisdicciones = set(lista_jurisdicciones)
```

```
conjunto_provincias = set(Provincias)
```

```
if conjunto_jurisdicciones == conjunto_provincias:
```

```
    print("Coinciden exactamente.")
```

```
else:
```

```
    print("Hay diferencias.")
```

```
    print("En Jurisdicción pero no en Provincias:", conjunto_jurisdicciones -
conjunto_provincias)
```

```
    print("En Provincias pero no en Jurisdicción:", conjunto_provincias -
conjunto_jurisdicciones)
```

Nota: (Para bibliotecas es el mismo código modificando los nombres de data frame y atributo, y ajustando el nombre que se usa para la Ciudad de Buenos Aires. El código se encuentra en el archivo .py)

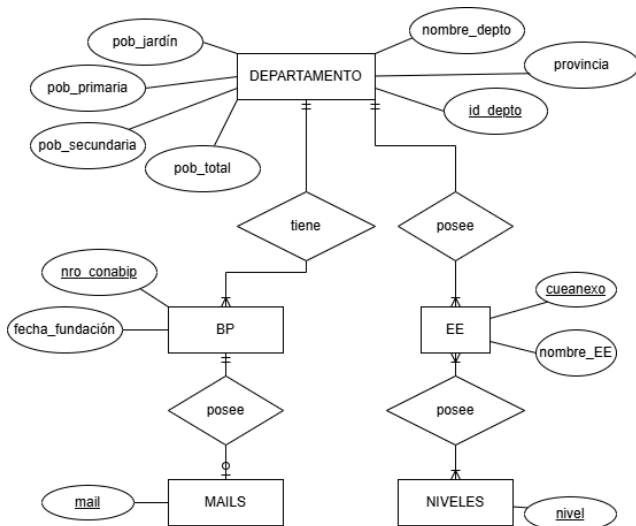
Valor obtenido: **"Coinciden exactamente"**

Diagnóstico y acción de mejora: hay completitud, no es necesario una mejora.

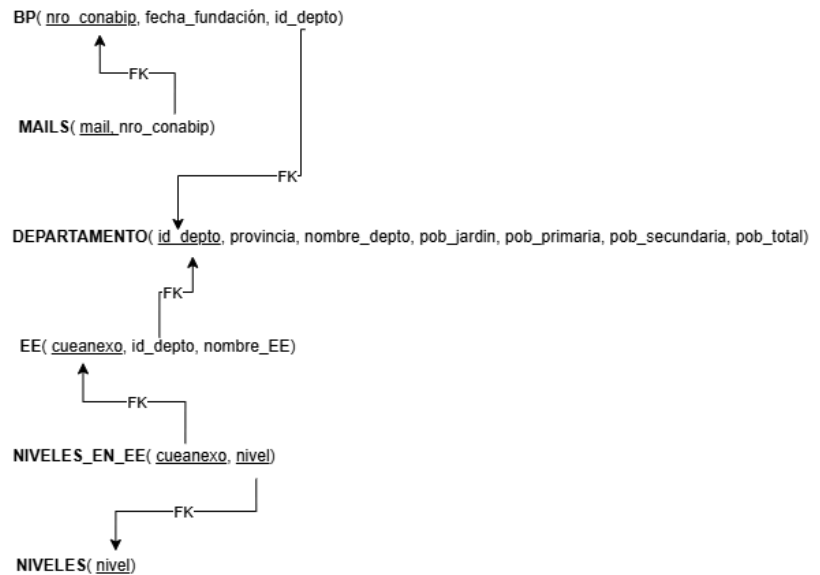
En conclusión, podemos afirmar que los datos disponibles no son de una calidad deseada para nuestro análisis. Por lo tanto, es imprescindible que se realice un proceso de limpieza y depuración antes de proceder con el análisis y la generación de reportes: corregir aquellos formatos que impiden una estructura que respete 3FN (reduciendo todos los datos a valores atómicos, quedándonos solo con un valor como el caso del mail repetido, o agregando filas en caso de ser valores no descartables), reemplazar los nulls por ceros, y cambiar los valores incorrectos o inconsistentes que encontramos (unificar códigos, comunas, etc.).

Luego, importamos los datos con estas características a una nueva estructura que planteamos desde un diseño que nos asegurara el mantenimiento de la calidad deseada, el DER:

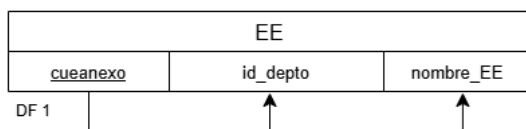
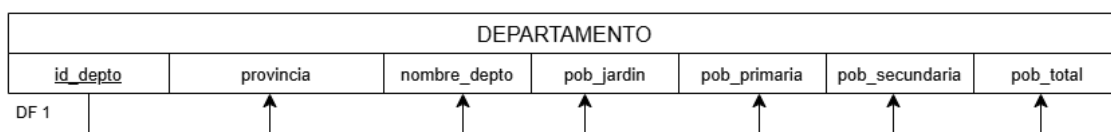
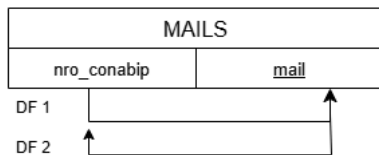
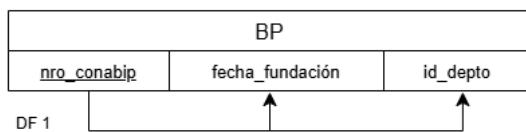
DER



MODELO RELACIONAL



DEPENDENCIAS FUNCIONALES



Nota: La entidad NIVELES y la relación NIVELES_EN_EE no poseen ninguna DF, por eso no figuran en la imagen.

- **Análisis de forma normal de nuestro esquema**

- Claves primarias (PK) definidas para cada entidad/relación del esquema: BP = nro_conabip; DEPARTAMENTO = id_depto; EE = cueanexo; MAILS = mail, nro_conabip; NIVELES = nivel; NIVELES_EN_EE = {cueanexo, nivel}.
Nota: En la entidad MAILS tanto mail como nro_conabip pueden funcionar como claves primarias independientes. En cambio, en NIVELES_EN_EE la clave primaria es una composición de los atributos cueanexo y nivel.
- Las EK están ya representadas en el gráfico a través de las flechas que apuntan a su entidad de origen.
- Primera Forma Normal: Todos los atributos tienen un dominio de valores atómicos indivisibles y no hay relaciones anidadas.
- Segunda Forma Normal: En todos nuestros esquemas que poseen DFs (el caso de NIVELES y NIVELES_EN_EE no es necesario analizarlo), contamos con PKs simples y, por lo tanto, con dependencias completas.
- Tercera Forma Normal: Exceptuando el caso de MAILS*, en cada esquema tenemos una única DF (o nulas) y por ende no es posible generar relaciones de transitividad.
**En cuanto al caso de MAILS, la transitividad que se genera involucra un atributo primo. En conclusión, nuestro esquema respeta la 3FN.*

- **Creación de DataFrames**

BP: Importa las columnas *nro_conabip*, *fecha_fundacion* e *id_depto* del padrón de bibliotecas populares.

DEPARTAMENTO: Tanto las franjas etarias como la población total provienen de los datos de población por departamento. El dato de provincia, del padrón oficial de establecimientos educativos. Esto es debido a que si bien bibliotecas populares también posee provincias, el hecho de que no todos los departamentos poseen bibliotecas populares imposibilitaba unificar departamentos con provincias. Por último los nombres e ids de los departamentos provienen tanto de establecimientos educativos como de población.

EE: Todos los datos son importados del padrón de establecimientos educativos, no sin antes unificar las comunas de CABA en un único departamento.

MAILS: Importa las columnas *nro_conabip* y *mails* del padrón de bibliotecas populares.

NIVELES: generada a partir del padrón de establecimientos educativos.

NIVELES EN EE: generada a partir del padrón de establecimientos educativos.

Sección Decisiones

A continuación presentaremos una serie de situaciones en las que hemos tenido que tomar decisiones, modificando ciertos valores para forzar su solución.

El padrón de bibliotecas públicas lista con id 2000 a todas las comunas de la Ciudad de Buenos Aires, lo que vuelve imposible (con los datos proporcionados para la realización del TP) conocer la cantidad de bibliotecas de cada comuna. Debido a esto tomamos la decisión de unificar Ciudad de Buenos Aires en un único departamento y asignarle el id 2000.

El padrón de población posee un error tanto en el id de Río Grande como en el de Ushuaia, en ambos casos sus ids difieren en 1 dígito con respecto al que les corresponde. Este error no está presente en los padrones de bibliotecas o establecimientos educativos, por lo que finalmente decidimos usar el id correspondiente y corregir el error de padrón de población.

En el año 2017 se legisló la creación del departamento de Tolhuin, antes dependencia de Río Grande. Esta actualización no se ve reflejada en bibliotecas o establecimientos educativos, los cuales siguen usando el id de Río Grande. Tomamos la decisión de obviar Tolhuin como partido independiente y unificar su población con Río Grande.

Para el segundo reporte de la Sección Análisis de Datos omitiremos las bibliotecas públicas que posean un valor nulo en fecha_fundación, siendo conscientes de que podríamos estar desestimando algunas que cumplan lo requerido pero para las cuales no se hizo registro. De igual manera lo haremos con las que tengan un valor nulo en mail, ya que podría no ser preciso nuestro reporte al desconocer si cuentan con un correo o si el mismo no fue ingresado.

Sección de Análisis de datos

Consultas SQL

Todas las tablas se encuentran completas dentro de la carpeta ConsultasSQL en formato csv.

En la primera consulta debemos informar para cada departamento su nombre, la provincia en la que se encuentran y la cantidad de establecimientos educativos según los niveles educativos, enfocándonos en la modalidad común.

Provincia	Departamento	Jardines	Población Jardín	Primarias	Población Primaria	Secundarios	Población Secundaria
Buenos Aires	La Matanza	325.0	87162.0	333.0	193043.0	335.0	184423.0
Buenos Aires	La Plata	215.0	29260.0	199.0	66893.0	206.0	65622.0
Buenos Aires	Lomas de Zamora	160.0	28806.0	178.0	65571.0	191.0	65827.0
Buenos Aires	General Pueyrredón	177.0	23100.0	169.0	53294.0	170.0	56683.0
Buenos Aires	Quilmes	160.0	26975.0	146.0	60261.0	153.0	60966.0

Esta relación nos será de gran utilidad para los gráficos pedidos. Limita la información en tres grupos etarios, y en una única modalidad educativa, la común.

Luego, a partir de la segunda consulta podemos obtener la cantidad de bibliotecas populares de cada departamento fundadas a partir de 1950, además de su respectiva provincia.

Provincia	Departamento	Cantidad de BP fundadas desde 1950
Buenos Aires	Avellaneda	9.0
Buenos Aires	Florencio Varela	9.0
Buenos Aires	Quilmes	8.0
Buenos Aires	Olavarría	8.0
Buenos Aires	Lanús	7.0

La tercera consulta nos será de gran utilidad para la formación de los gráficos, ya que indica la cantidad de bibliotecas populares y establecimientos educativos por departamento, su población y la provincia en la que se encuentran.

Provincia	Departamento	Cant_BP	Cant_EE	Población
Ciudad de Buenos Aires	CABA	43	1552	3095454.0
Córdoba	Capital	20	1054	1498060.0
Buenos Aires	La Matanza	16	957	1837168.0
Santa Fe	Rosario	39	742	1337958.0
Buenos Aires	La Plata	33	604	756074.0

Finalmente, se ha indicado el dominio más frecuente en los mail de las bibliotecas populares por departamento, además de su respectiva provincia, en la consulta número cuatro.

Provincia	Departamento	Dominio más frecuente en BP
Buenos Aires	25 de Mayo	ninguno
Buenos Aires	9 de Julio	ninguno
Buenos Aires	Adolfo Alsina	hotmail
Buenos Aires	Adolfo Gonzales Chaves	yahoo
Buenos Aires	Alberti	live

Gráficos

Para nuestros gráficos contamos con **variables cuantitativas** como las cantidades de establecimientos educativos, bibliotecas populares, población, y **variables categóricas** como los niveles, departamentos y provincias. Más allá de que nuestras fuentes de datos cuentan con una diferencia en el tiempo de actualización, para el análisis que estamos haciendo consideramos a los datos como transversales.

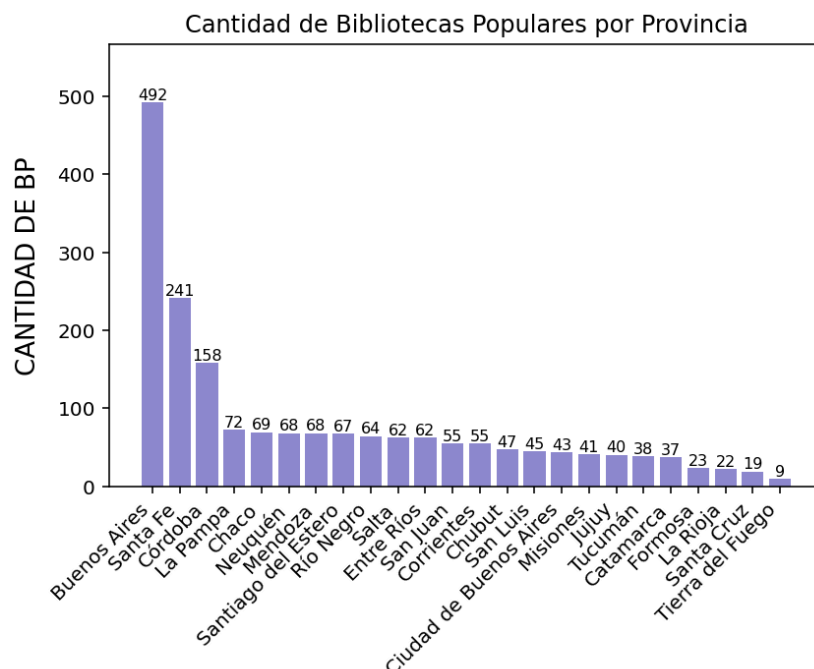


Gráfico que relaciona la cantidad de bibliotecas públicas en cada provincia de manera decreciente.

Al observar el primer gráfico, notamos que existe una marcada diferencia entre la cantidad de bibliotecas populares de las primeras tres provincias y el resto.

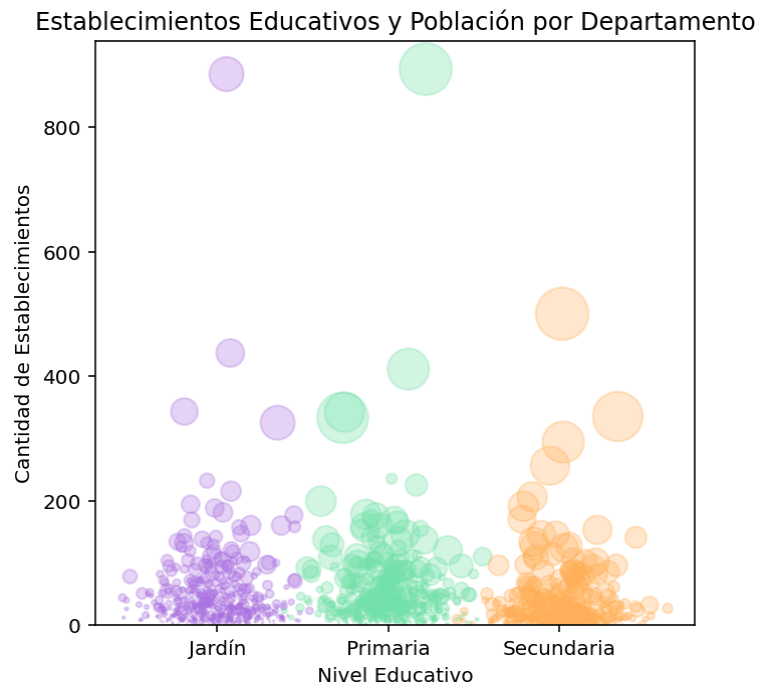
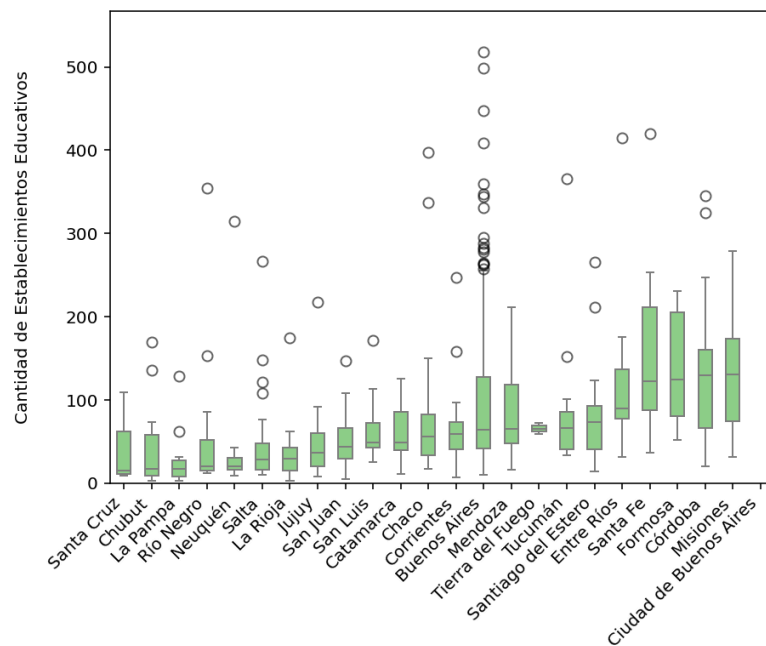


Gráfico en el cual cada burbuja representa un departamento distinto que varía el tamaño acorde a la población de su propio grupo etario separado según el nivel educativo.

Vemos que a medida que las burbujas “suben” estas tienden a ser de mayor tamaño. Intuitivamente, a mayor población se requiere aumentar la cantidad de establecimientos educativos para cubrir la demanda. Las poblaciones de primaria y secundaria son mayores que las de jardín. Además, se destaca una gran concentración de burbujas del nivel secundario cuando las mismas se acercan al eje x, lo que podría indicar una mayor cantidad de departamentos con población adolescente, así como una mayor proporción de población de su grupo etario. De igual modo, a esta escala los tres niveles muestran una distribución medianamente similar.

La capacidad de los establecimientos educativos es un dato que podría ser de utilidad y no está presente. A partir del gráfico notamos que los jardines albergan una menor cantidad de población que los secundarios, aunque la cantidad de establecimientos es similar en los tres niveles. Si este dato estuviese a nuestro alcance, podría explicarse por qué para una cantidad menor de niños se necesita la misma cantidad de establecimientos que para una cantidad mayor de adolescentes. Para mostrar esta relación, realizamos una tabla extra que figura en [segunda posición en el anexo](#), la misma nos permite ver que la cantidad de personas que hay por establecimiento educativo es ascendente según el rango etario.

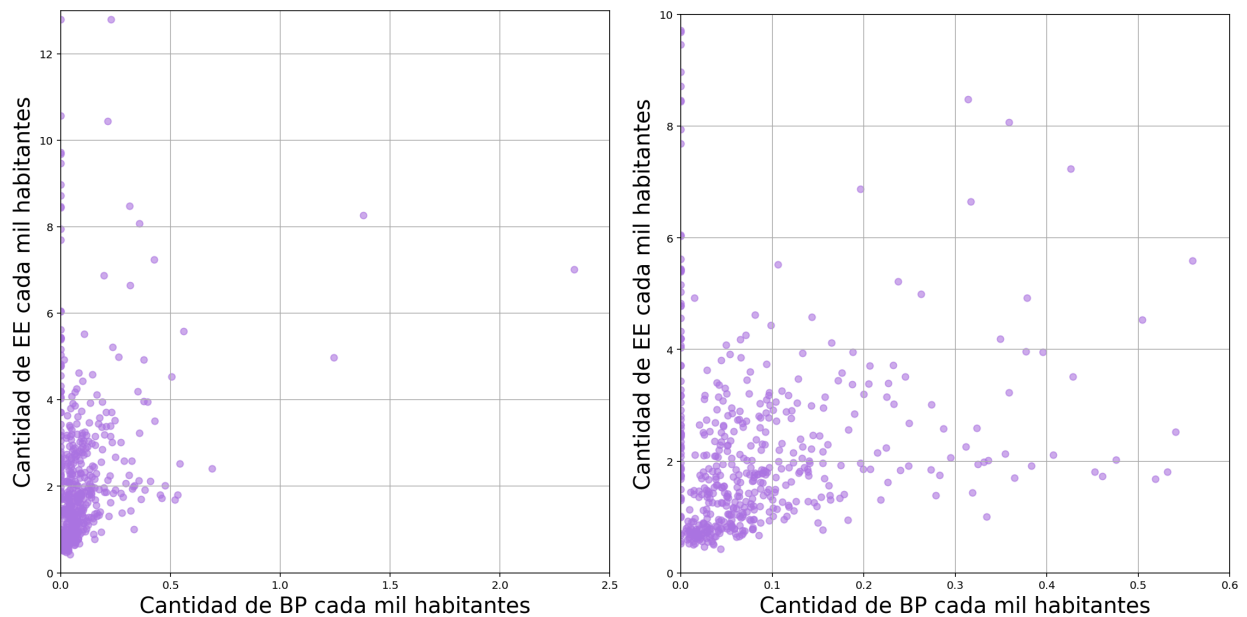
Cantidad de Establecimientos Educativos por cada departamento de las Provincias



Resumen gráfico de la distribución de establecimientos educativos por cada departamento de las provincias ordenado según su mediana de manera ascendente.

La mayor parte de las provincias tiene una distribución bastante concentrada en valores bajos. Si consideramos que una mediana baja en boxplot significa que al menos el 50% de los departamentos tiene una cantidad de establecimientos educativos relativamente baja en comparación con el resto, podemos hablar de una desigualdad en la distribución: existen pocos departamentos con muchos establecimientos y muchos departamentos con pocos establecimientos, lo que obliga a los valores de la distribución a desplazarse hacia valores menores. Además, este gráfico presenta *outliers*, departamentos con cantidades de establecimientos mucho mayores al rango intercuartílico. En el caso de Buenos Aires vemos que posee una distribución más desigual que el resto de las provincias. Tierra del Fuego, en cambio, apenas tiene variabilidad y un rango estrecho.

Relación entre BP y EE cada mil habitantes por departamento



El gráfico de la izquierda muestra la relación entre las bibliotecas públicas y los establecimientos educativos cada mil habitantes por departamento. El de la derecha, es el mismo gráfico en otra escala, utilizado para mostrar la zona con mayor densidad de manera más clara.

Notamos que existe una gran cantidad de casos de departamentos que poseen establecimientos educativos pero no bibliotecas, aunque no hay ningún caso inverso.

La relación claramente no es proporcional. La mayor densidad en el gráfico se da entre las coordenadas (0,4) y (0.2,4). Fuera de esa zona, hay algunos departamentos que suben mucho en cuanto a cantidad de establecimientos educativos, pero se mantienen dentro de los valores 0-0.5 de bibliotecas populares, luego hay casos extraños en los que aumenta significativamente la cantidad de bibliotecas a comparación de la norma, pero no alcanzan para poder llegar a una conclusión distinta.

Sección de Conclusiones

¿Cuál es la relación existente entre la cantidad de bibliotecas populares y establecimientos educativos en los departamentos de la República Argentina? ¿Podemos decir que existe una y describirla? A partir del análisis previo, hemos llegado a unos resultados que describiremos a continuación.

En primer lugar, si hablamos del cuarto gráfico, notamos que organiza la cantidad de establecimientos educativos y bibliotecas populares según los departamentos a pesar de que el planteo principal de este informe busca una relación entre provincias. Además, al no haber identificado de manera única los departamentos del gráfico por el motivo de que hubiera sido ilegible, creemos que hubiera sido útil realizar un gráfico que compare la cantidad de establecimientos educativos y de bibliotecas populares por provincia en su lugar.

Si comparamos el primer gráfico con el tercero, podemos ver que si bien las provincias con más bibliotecas populares y más establecimientos educativos son Buenos Aires, Santa Fe y Córdoba, esta relación no se mantiene para el resto de provincias. Por

ejemplo: la Ciudad de Buenos Aires tiene muchos más establecimientos educativos que La Pampa, pero en el caso de las bibliotecas populares ocurre lo opuesto.

Consideramos importante tener en cuenta que, con el paso del tiempo, la digitalización de contenidos —tanto académicos como de otro tipo— se ha vuelto cada vez más común. Esto podría haber generado una disminución en la cantidad de material impreso en formato físico y, en consecuencia, una menor necesidad de espacios destinados a albergarlo, como las bibliotecas. Además, el creciente acceso de la población a contenidos digitales habría reducido la frecuencia de visitas a espacios como las bibliotecas populares.

Los datos obtenidos no son lo suficientemente concluyentes para poder determinar una relación explícita entre la cantidad de establecimientos educativos y la cantidad de bibliotecas populares por provincia. En la [primera](#) y [tercera](#) tabla del anexo podemos observar que no existe un crecimiento proporcional entre cantidad de habitantes, establecimientos educativos y bibliotecas populares.

Se necesitaría acceder a una mayor cantidad de datos de relevancia, mencionaremos ejemplos de información que podría llegar a establecer una relación más explícita entre ambos establecimientos:

Políticas Públicas: Gobiernos locales que prioricen la cultura, poder acceder al porcentaje de presupuesto destinado a educación y acceso a la cultura.

Geografía y Distribución Poblacional: Provincias con población dispersa o mayor porcentaje de población rural contarán con mayor cantidad de establecimientos educativos para garantizar el acceso a la educación, mientras que provincias con mayor población urbana tendrá más chances de contar con bibliotecas populares.

Factores Económicos: que podrían influir en la necesidad de acceso a libros mediante bibliotecas populares.

Estado de Establecimientos: algunos departamentos pueden tener baja cantidad de bibliotecas populares pero estas ser de gran tamaño y poseer suficiente oferta literaria para la población.

Finalmente, la existente discrepancia temporal entre las fuentes afecta la vigencia de los datos. Como la cantidad de población y de establecimientos que variaron en el período no contabilizado de 2022 a 2023 quedan excluidos de la fuente de datos (bibliotecas populares fue lo único actualizado) utilizada para este informe, hay una desactualización de datos causada por esta misma instancia.

Anexo

1)

Provincia	BP	BP cada 100k habitantes	EE	EE cada 100k habitantes	Población
Buenos Aires	492.0	2.83	15526.0	89.18	17408906.0
Córdoba	158.0	4.14	4324.0	113.43	3812064.0
Santa Fe	241.0	6.85	3346.0	95.08	3519059.0
Ciudad de Buenos Aires	43.0	1.39	1552.0	50.14	3095454.0
Mendoza	68.0	3.35	1515.0	74.6	2030773.0
Tucumán	38.0	2.2	1423.0	82.38	1727337.0
Salta	62.0	4.32	1188.0	82.83	1434225.0
Entre Ríos	62.0	4.38	2012.0	142.18	1415097.0
Misiones	41.0	3.22	2357.0	185.1	1273347.0
Corrientes	55.0	4.55	1603.0	132.52	1209671.0
Chaco	69.0	6.14	2034.0	180.86	1124603.0
Santiago del Estero	67.0	6.33	2079.0	196.55	1057752.0
San Juan	55.0	6.71	977.0	119.23	819445.0
Jujuy	40.0	4.94	824.0	101.81	809364.0
Río Negro	64.0	8.56	816.0	109.14	747697.0
Neuquén	68.0	9.6	645.0	91.03	708578.0
Formosa	23.0	3.8	1254.0	207.1	605507.0
Chubut	47.0	7.97	609.0	103.32	589454.0
San Luis	45.0	8.32	615.0	113.77	540548.0
Catamarca	37.0	8.65	961.0	224.73	427625.0
La Rioja	22.0	5.75	657.0	171.79	382453.0
La Pampa	72.0	20.04	511.0	142.26	359193.0
Santa Cruz	19.0	5.66	278.0	82.82	335677.0
Tierra del Fuego	9.0	4.73	131.0	68.78	190455.0

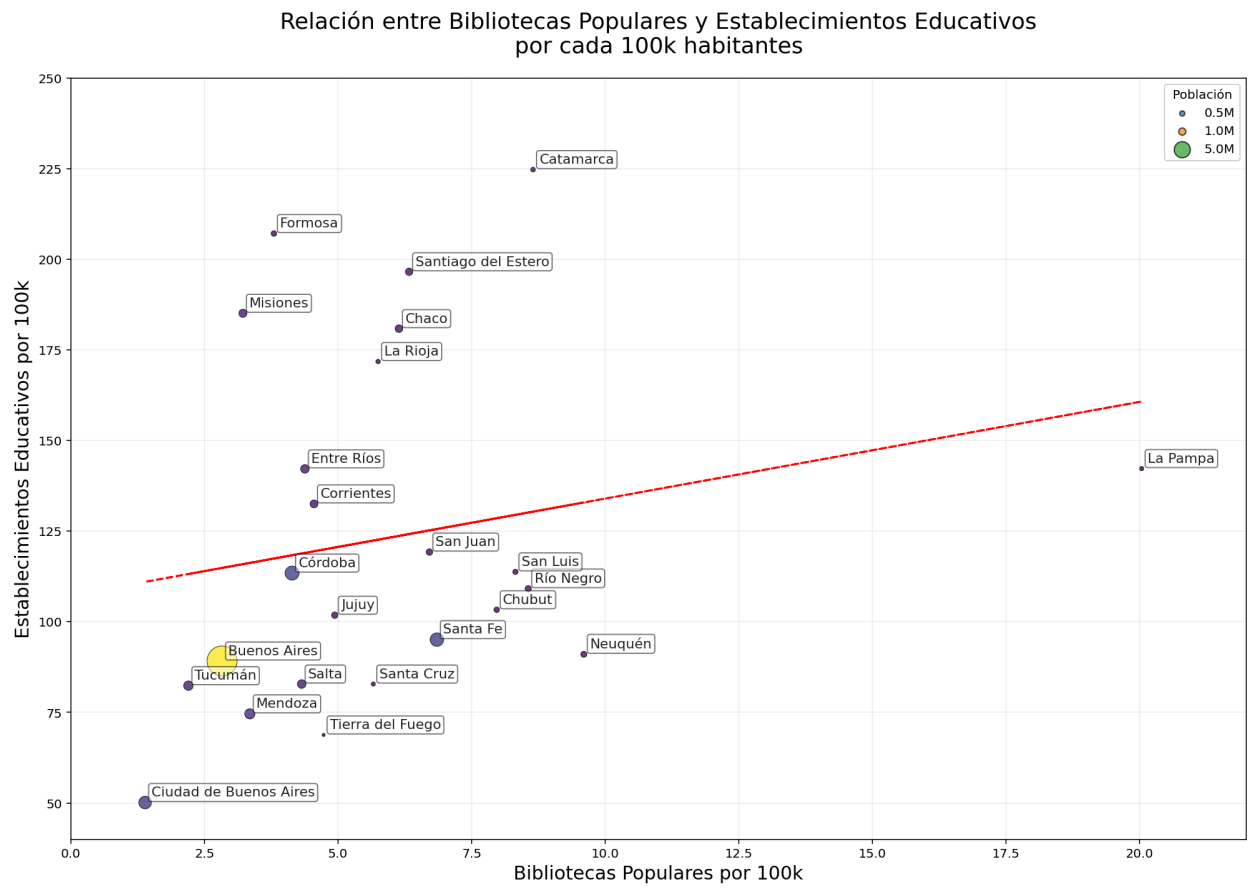
Tabla comparativa con cantidades de BP, EE, sus respectivas cantidades cada 100.000 habitantes y Población separadas por provincia.

2)

Provincia	Departamento	poblacion_por_jardin	poblacion_por_primaria	poblacion_por_secundaria
Buenos Aires	General Alvear	30.6	37.2	155.2
Buenos Aires	Guaminí	30.7	57.1	129.1
Buenos Aires	Pila	31.0	34.1	221.0
Buenos Aires	Tordillo	34.0	51.8	221.0
Buenos Aires	General Guido	34.3	20.4	107.3
Buenos Aires	General Lavalle	35.7	68.6	174.3
Buenos Aires	Roque Pérez	36.2	53.7	225.3
Buenos Aires	Rauch	37.2	65.3	224.1
Buenos Aires	Coronel Dorrego	37.2	68.4	147.1
Buenos Aires	Tapalqué	37.4	44.1	208.6
Buenos Aires	Lobería	38.8	50.2	164.6
Buenos Aires	Adolfo Gonzales Chaves	40.2	74.0	120.4
Buenos Aires	Tornquist	40.7	67.6	111.2
Buenos Aires	Adolfo Alsina	40.7	63.6	121.8
Buenos Aires	Suipacha	41.2	82.6	206.8
Buenos Aires	Puan	41.6	68.7	97.0
Buenos Aires	Carlos Tejedor	41.9	70.1	142.4
Buenos Aires	General Paz	42.6	59.8	231.3
Buenos Aires	25 de Mayo	42.8	67.9	194.4
Buenos Aires	General La Madrid	43.1	67.7	126.3
Buenos Aires	Navarro	43.6	66.5	232.0
Buenos Aires	San Cayetano	44.9	73.1	259.0
Buenos Aires	San Andrés de Giles	47.5	89.6	161.8
Buenos Aires	General Juan Madariaga	47.8	95.3	222.1

Tabla comparativa que relaciona la cantidad de personas del rango etario que hay en cada establecimiento educativo correspondiente a su nivel

3)



Este gráfico muestra la escasa relación que existe entre las Bibliotecas Populares y los Establecimientos Educativos cada 10.000 habitantes. El tamaño de los puntos determina la población, y lo que se demuestra es que no hay ningún tipo de relación entre ambas variables.