Universidad de Buenos Aires

Departamento de computación

Laboratorio de Datos
Trabajo práctico 01

Objetivo

Manejo de datos y visualización

Presentado por

Santiago De Luca - Lautaro Aguilar - Federico Bourrat

Grupo

“La T y la F”

Profesores

Pablo Turjanski - Manuela Cerdeiro - Mateo Guerrero Schmidt

Año

2025

Buenos Aires - Argentina

Sección resumen

En este informe analizamos si existe una relación entre la cantidad de establecimientos educativos y centros culturales distribuidos en Argentina. Para ello, utilizamos datos sobre ambos, junto con información sobre la población total del país, dividida en departamentos.

Para llevar a cabo el análisis, confeccionamos varios diagramas, organizamos las tablas según criterios detallados a lo largo del informe y elaboramos gráficos que ilustran nuestra conclusión final.

El resultado de nuestro estudio indica que no existe una relación extremadamente clara a nivel nacional, esta relación es débil. Sin embargo, al dividir el país en regiones, observamos que en algunas de ellas sí se manifiesta una correlación más evidente. Esto sugiere que la falta de una relación clara a nivel nacional se debe al "ruido" generado por las diferencias entre las distintas regiones.

Sección introducción

El objetivo de este trabajo es analizar la relación entre la cantidad de centros culturales y establecimientos educativos en Argentina. Para responder esta pregunta, realizamos varias actividades clave, como la elaboración de un **Diagrama de Entidad-Relación (DER)**, que nos permitió visualizar mejor las posibles conexiones entre los conjuntos de datos. Consideramos este paso como el objetivo a resolver en el trabajo, ya que fue fundamental para lo realizado luego. Posteriormente, desarrollamos **Esquemas Relacionales** para representar las relaciones entre las distintas tablas de nuestro modelo.

Durante la exploración inicial de las tablas, nos encontramos con la necesidad de tomar decisiones estratégicas para optimizar la importación y manipulación de los datos. Entre estas decisiones se incluyen la eliminación de columnas innecesarias, la selección de valores en casillas con múltiples opciones y otros criterios detallados en el informe.

Para responder a nuestra pregunta principal, generamos gráficos que facilitaron la interpretación de la relación entre los centros culturales y los establecimientos educativos.

Todo este proceso se encuentra detallado y fundamentado en el informe, el cual está dividido en las siguientes secciones (links):

Sección: Procesamiento de datos

Formas normales

La tabla original de establecimientos educativos no se encuentra en 1FN porque los atributos Teléfono y Mail tienen múltiples valores en algunas tuplas (en general, separados por /), es decir, los valores no son atómicos. Como resultado, no está en 2FN ni en 3FN.
De forma similar, la tabla original de centros culturales tampoco se encuentra en 1FN porque el atributo Mail tiene varios valores asignados en algunas tuplas (en general, separados por espacios).

Calidad de datos

Establecimientos educativos:

Esta fuente de datos tiene un problema en su atributo de **completitud** y **consistencia**. El atributo código área tiene varios datos que son NULL y para representar esta ausencia de información a veces se deja vacío el dato (literalmente NULL), pero a veces se escribe "S/N","S/I","0","00", etc. Estos son problemas de instancia, ya que pareciera que dependiendo quién completó los datos, la forma de manejar la ausencia de información fue distinta. Utilizamos el método GQM:
Objetivo: Que el dato de código área esté completo.
Pregunta: ¿Cuál es la proporción de códigos de área que son NULL o alguna variación?
Métrica:
Cantidad de códigos de área que son NULL o alguna variación / Cantidad tuplas = 0.455
Es decir, casi la mitad de los códigos de área son NULL.

Centros culturales:

Esta fuente de datos tiene un problema en el atributo de **relevancia**. Hay tres columnas que tienen información completamente irrelevante porque tienen un valor que se repite en todas las tuplas. Este es el caso de la columna Observaciones, InfoAdicional (siempre vacías) y Categoría (siempre es Centro cultural). Esto es un problema de modelo porque, por ejemplo, en el caso de Categoría la categoría siempre iba a ser centro cultural, ya que de eso se trata la información, no había forma de que ese dato tenga otro valor. Utilizando el método GQM:
Objetivo: Que todas las columnas sean relevantes.
Pregunta: ¿Cuál es la proporción de columnas que no tienen información relevante?
Métrica: Columnas irrelevantes / columnas relevantes = 3/24 = 0,125

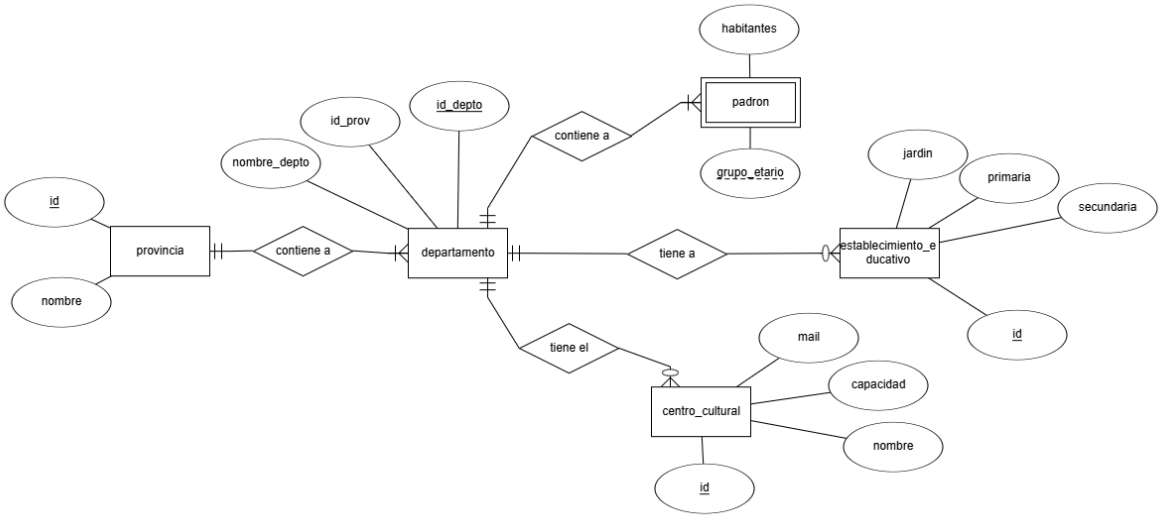
Padrón:

Esta fuente de datos tiene un problema en el atributo de **disponibilidad**. En el principio, las primeras 15 filas tienen todas sus columnas fusionadas. Lo mismo pasa por cada uno de los 513 departamentos que aparecen, para cada uno hay tres filas que están fusionadas y una con el total y una por el encabezado. Esto genera, por ejemplo, que la primera columna en algunas filas tome como valor strings y en otras enteros. Todas estas cosas dificultan el manejo de la información para usuarios que quieran utilizarla y analizarla. El procesamiento necesario involucra no solo eliminar todas las filas fusionadas sino que también borrar el encabezado y las líneas en blanco entre departamentos (513 veces). Esto es un problema de modelo.

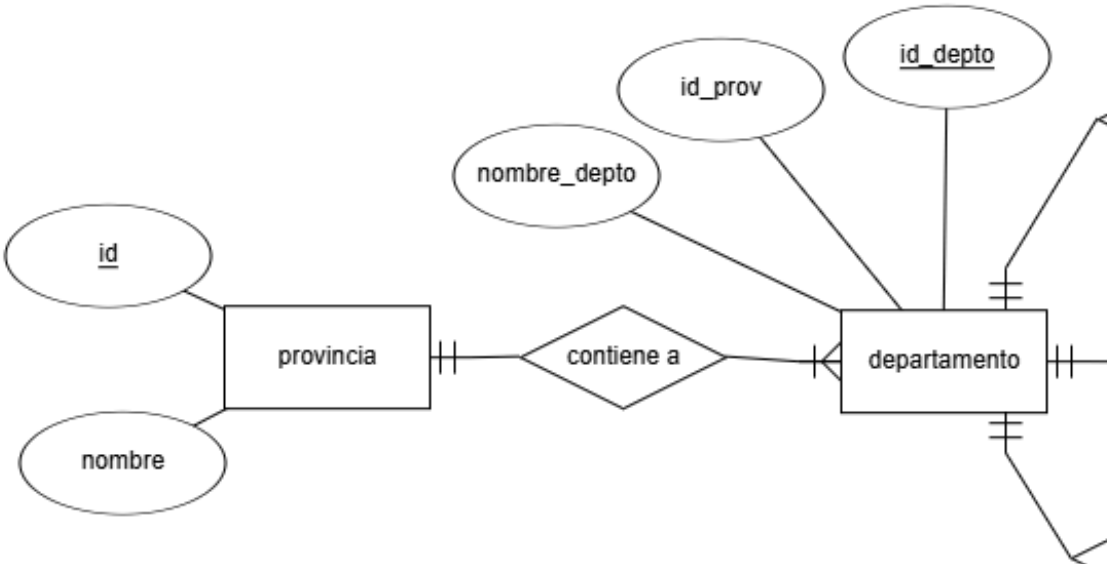
Como objetivo, se podría pedir que toda la información se guarde en una única tabla (con un único encabezado) o que no haya filas fusionadas. En ningún caso la métrica sería positiva.
Objetivo: Que todas las filas tengan la misma cantidad de columnas.
Pregunta: ¿Todas las filas tienen la misma cantidad de columnas?
Métrica: No.

Diagrama Entidad-Relación

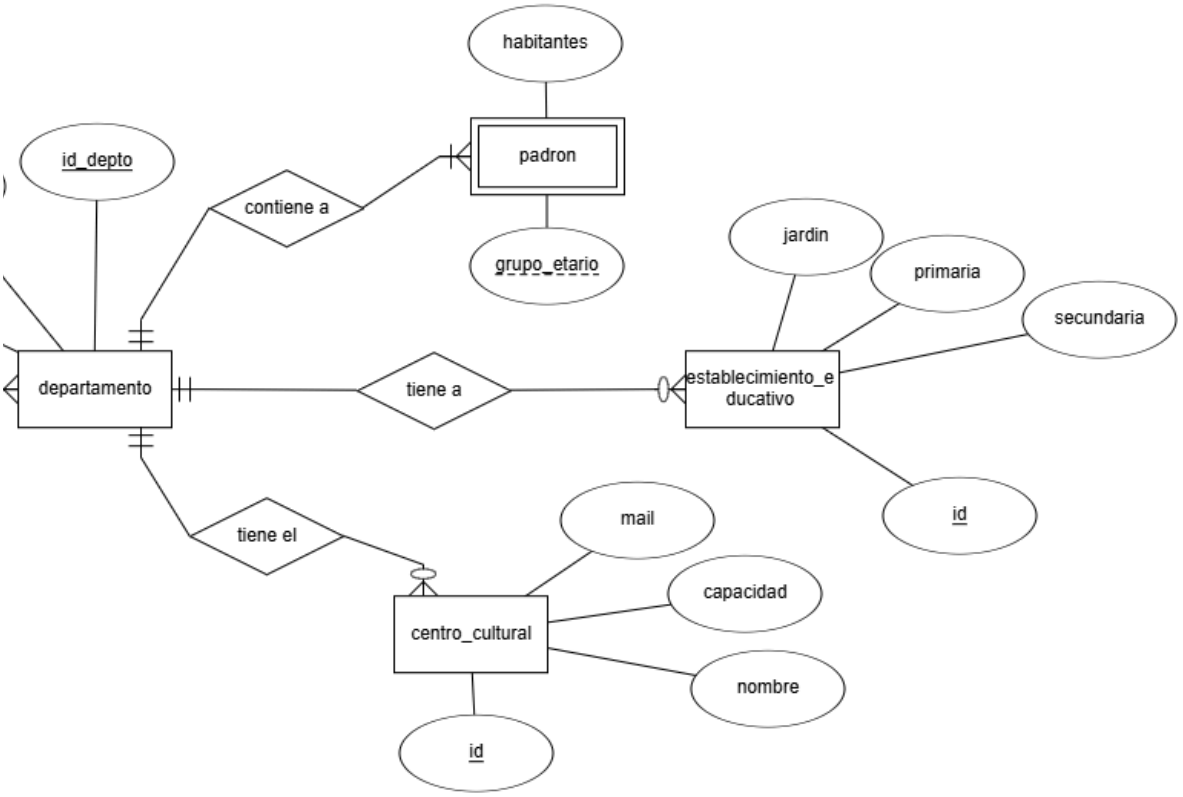
A continuación presentamos el Diagrama de Entidad-Relación (DER) que confeccionamos para representar nuestro modelo. Para armar el DER fue necesario encontrar un atributo que relacione los establecimientos educativos, los centros culturales y el padrón.
Con el objetivo de que las relaciones estén en 3FN, surgió la siguiente disposición, donde decidimos relacionarlos a través del departamento donde residen los establecimientos y centros y donde corresponde el empadronamiento.



Este es el diagrama de nuestro modelo. Está comprendido por cinco entidades, cuatro fuertes y una débil. El punto de unión central del diagrama es la entidad *departamento*, a la que todas las demás entidades están relacionadas.



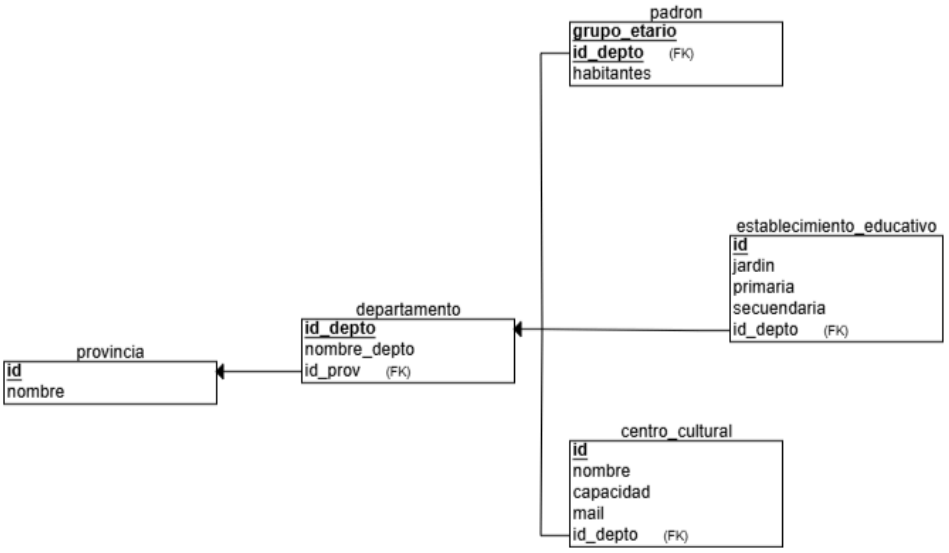
En el lado izquierdo tenemos las entidades *provincia* y *departamento*. El nombre de la provincia no es un atributo más de *departamento* porque sino existiría la dependencia funcional *id_prov -> nombre_prov*, lo cual rompería con la 3FN (ambos serían atributos no primos de departamento).



En el lado derecho tenemos tres relaciones uno-a-muchos entre *departamento* y *padrón*, *establecimiento_educativo* y *centro_cultural*. Para que *padrón* tuviera su propio identificador no basta con saber el grupo etario, ya que por cada departamento hay información sobre cada uno de los grupos etarios. Como resultado, la clave de *departamento* es compuesta, y depende de un atributo foráneo: la clave es *id_depto* junto a *grupo_etario*. Esta combinación sí es única en padrón. Como la clave de *padrón* depende de un atributo foráneo, debe ser considerada entidad débil.

Esquemas Relacionales

Una vez realizado el DER, procedemos al armado del esquema relacional donde debimos definir las relaciones entre las tablas, las claves principales y las claves foráneas. Las claves principales de cada entidad están subrayadas.



Este esquema muestra las relaciones que hay entre las diferentes tablas de nuestro modelo. Algo muy importante es notar que tanto *padrón* como *centro_cultural* como *establecimiento_educativo* tienen la clave foránea *id_depto*. Por otro lado, *departamento* tiene como clave foránea *id_provincia*.

Dependencias funcionales

A la hora de establecer las dependencias funcionales, nos enfocamos en obtener y mantener la 3FN en cada una de las relaciones.

departamento

id_depto	nombre_depto	id_prov(FK)

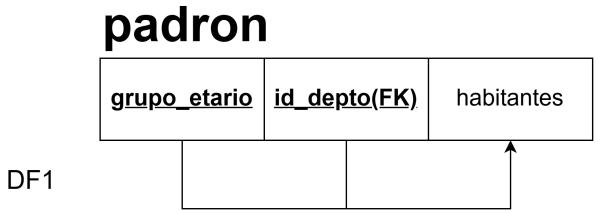
DF1

establecimiento_ecucativo

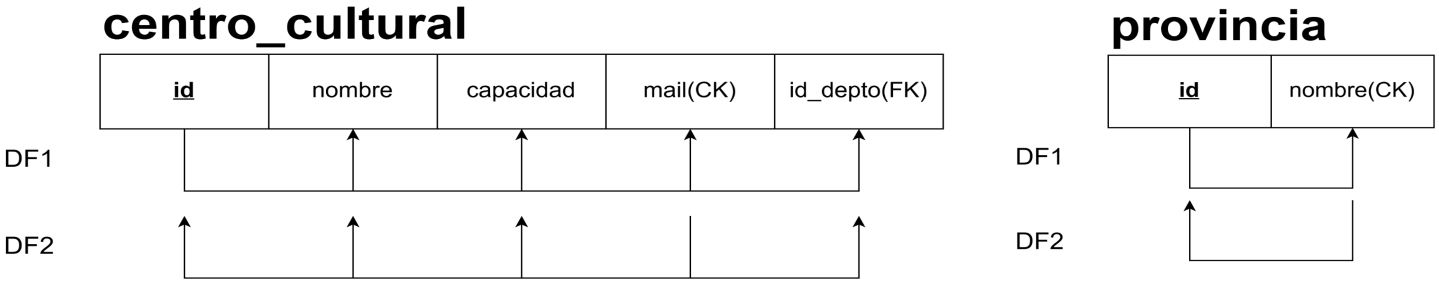
id	jardin	primaria	secundaria	id_depto(FK)

DF1

Es fácil ver que estos dos esquemas están 3FN: no tienen atributos multivaluados (1FN), no tienen ninguna clave candidata que sea compuesta, entonces no pueden tener ningún atributo que dependa parcialmente de otro (2FN) y no hay ninguna dependencia transitiva, la única dependencia es de la clave a los demás atributos (3FN).



En el caso de *padrón*, la única dependencia que hay de la clave (que en este caso es compuesta) es a un atributo no primo. Aquí se cumple 1FN porque no hay atributos multivariados, se cumple 2FN porque, por mucho de que haya una clave compuesta, el único atributo que depende de ella no depende parcialmente, y cumple 3FN porque no hay relaciones transitivas (habitantes no determina ningún atributo).



En el caso de centro cultural, interpretamos que no puede haber mails repetidos, es decir que el mail podría ser una clave candidata.
Este esquema está en 1FN porque no hay atributos multivariados, está en 2FN porque no hay ninguna clave candidata compuesta y está en 3FN porque ambas DF cumplen con alguna de las reglas de 3FN:
Un esquema R está en 3FN si, para toda dependencia funcional no trivial X → A de R, se cumple alguna de las siguientes condiciones:

- X es SK de R: DF1 y DF2 cumplen
- A es atributo primo de R

El caso de provincia es similar. Consideramos que el nombre de la provincia es clave candidata porque en Argentina no existen dos provincias con un mismo nombre. Pero, como ocurre en centro cultural, DF1 y DF2 cumplen con que X es SK de la relación.

Importación de datos

Armado de departamento y provincia:

Decidimos sacar los departamentos de la tabla de establecimientos educativos porque era la tabla con más cantidad de departamentos: Establecimiento educativo tiene 514, padrón 513 y centro cultural 189. No hay departamento que aparezca en centro cultural que no esté entre los 514 de establecimiento educativo, pero sí hay dos de padrón que no están en establecimiento educativo, así que los agregamos manualmente. Estos son:

- 94015: ISLAS DEL ATLÁNTICO SUR
- 94008: RÍO GRANDE

Padrón no tenía asociado los nombres de los departamentos, así que buscamos el nombre de los departamentos por fuera de las tres bases de datos dadas. Finalmente, el número total de departamentos que tenemos en departamento es 516.
Para la tabla provincia, sacamos los datos de la tabla de centros culturales, tanto el nombre como el id. En el caso de provincia es mucho más sencillo verificar que estén presentes todas las tuplas que debe haber en la tabla, ya que sabemos que hay 23 provincias, más la ciudad autónoma.

Armado de establecimiento educativo:

Sacamos la información de la tabla de establecimientos educativos. Aprovechamos el atributo CUEANEXO, que es un identificador único para cada establecimiento educativo del país, y lo asignamos como id.
Por otro lado, en cuanto a los jardines, primarias y secundarias, mantuvimos una estructura similar a la de la tabla original, donde dejamos un 1 en el nivel que el establecimiento educativo tenía. Si un establecimiento educativo no tenía cierto nivel, en vez de dejarlo como NULL, le asignamos un 0. Ya que en la tabla original dentro de la categoría secundario se encuentran distintos tipos de secundario (por ejemplo, SNU, superior no universitario) consideramos que para tener secundario bastaba tener un 1 en cualquiera de las 4 categorías de secundario. Lo mismo hicimos con los jardines.
Los únicos tipos de jardines, primarios y secundarios que consideramos fueron los que estaban debajo de la columna "común".
La tabla original no contenía una columna que sea código de departamento, pero sí tenía el código de localidad, que está conformado por el código de departamento (primeros cinco dígitos) y otros tres dígitos más. De esta forma, tomamos esos primeros cinco dígitos para la columna id_depto.

Armado de centro cultural:

Para esta tabla usamos la base de datos de centros culturales. Esta fue la tabla de armado más directo: ID_DEPTO lo asignamos a id_depto, Nombre a nombre y Capacidad a capacidad. Como esta tabla no tenía un identificador único, lo creamos nosotros con un índice numérico.
Finalmente, en el caso de mail, decidimos que para los centros que tenían dos mails asociados solo dejaríamos el primero, entendiendo que era el principal.

Armado de padrón:

Los datos se tomaron del censo. Asignamos el número de área (que es equivalente al código de departamento) a cada una de las filas dentro de cada lista de poblaciones. Luego agrupamos por grupos etarios las edades, y la suma de la cantidad de personas de esas edades fue el valor que tomó habitantes en cada fila.
En medio de este proceso eliminamos las filas que había entre las listas de poblaciones de cada departamento, que tenían los totales y los encabezados.

Sección: Decisiones tomadas

Sobre el departamento como unidad territorial elegida:

Decidimos elegir el departamento como unidad territorial sobre las provincias o las localidades ya que la mayoría de las consignas apuntaban a realizar el análisis dividiendo el país de esta forma.

Sobre padrón como entidad débil:

Ya que sus atributos (grupo etario y habitantes) podían repetirse, ninguno de ellos podría ser clave. Tampoco podíamos armar una superclave con ellos, porque para distintos departamentos podría existir el mismo grupo etario con la misma cantidad de habitantes. Decidimos conformar una clave compuesta entre *id_depto* y *grupo_etario*. Como resultado, la clave de padrón es una clave compuesta por un atributo primo que es foráneo, convirtiendo la entidad en una entidad débil.

Sobre establecimientos educativos sin modalidad común:

Al armar la tabla *establecimiento_educativo* decidimos no tomar aquellos que no tenían ninguno de los niveles de la modalidad común porque las consignas pedían que trabajemos sobre los establecimientos educativos de esa modalidad.

Sobre CABA:

Dado que CABA era tratado distinto en algunas tablas con respecto a otras, ciertos códigos de las comunas estaban mal y tratar las comunas dentro de CABA como la misma jerarquía de división que los departamentos no nos parecía correcto, decidimos tratar CABA como una provincia con un único departamento. Para eso a todos los códigos de comuna (que empiezan con 02) le asignamos el código 02000 manualmente en el código, en cada una de las tres tablas.

Sobre los departamentos que no tenían ningún centro cultural con capacidad mayor a 100:

Decidimos no ponerlos en la tabla del resultado final de la consulta (ii), ya que eso involucraría agregar casi 400 departamentos que tenían un 0 asociado en la columna "cantidad de centros culturales con capacidad mayor a 100" y se perderían de vista los pocos departamentos a los que sí se les relaciona una cantidad de centros mayor a 0.

Sobre falta de datos en la consulta (i):

En los casos en los que para un departamento no teníamos información sobre sus establecimientos o población, decidimos no incluirlo, para evitar NULLs y porque creemos que el punto central de esta tabla es relacionar población con establecimientos, si uno de esos datos no está pierde sentido mostrar ese departamento.

Sobre falta de datos en la consulta (iii):

La consulta (iii) pide explícitamente no omitir departamentos que no tengan datos asociados, lo que produce que, por ejemplo, todos aquellos departamentos que no aparezcan en la tabla de centros tengan NULL en el campo que indica la cantidad de centros. Decidimos dejar estos NULL porque por consigna no podemos eliminar departamentos solo porque no tengan información, pero cambiar ese NULL a 0 u otro valor sería erróneo porque no tenemos la información necesaria para asegurar que eso sea cierto.

Sobre empate en dominio más frecuente de mail en consulta (iv):

En caso de empate en el dominio más frecuente en un departamento decidimos dejar ambos, ya que para eliminar uno deberíamos tener algún criterio que justifique la decisión y tenga sentido, y no lo tenemos. La otra opción, que consideramos aún peor, sería eliminar ambos dominios. Somos muy conscientes de que esta decisión hace que en la tabla final, un departamento pueda aparecer dos veces con dos dominios distintos, pero no encontramos una salida mejor.
Por otro lado, los casos en los que el dominio de mail era NULL no fueron contados.

Sobre falta de datos en el gráfico (iv):

Para este gráfico se omitieron los departamentos que no aparecían en *centro_cultural,* porque sin esa información no se puede asegurar nada sobre el valor de centros culturales cada 1000 personas (asumir 0 centros culturales sería erróneo). También se omitieron los que no estaban en *padrón* (no sabemos la población y sucede lo mismo).

Sección: Análisis de datos

Consultas a las tablas

Las tablas completas correspondientes a las consultas SQL se guardarán en una carpeta llamada TablasConsultas. En esta sección sólo incluiremos las primeras cinco filas de cada consulta. El proceso de recorte de las tablas está incluido al final del archivo de código entregado.
Consulta (i): Para cada departamento informar la provincia, cantidad de EE de cada nivel educativo, considerando solamente la modalidad común, y cantidad de habitantes por edad según los niveles educativos.

Provincia	Departamento	Jardines	Poblacion jardin	Primarias	Poblacion primaria	Secundarias	Poblacion secundaria
Buenos Aires	LA MATANZA	325.0	157034.0	355.0	193043.0	333.0	214041.0
Buenos Aires	LA PLATA	232.0	52075.0	254.0	66893.0	199.0	78431.0
Buenos Aires	LOMAS DE Z.	162.0	50825.0	219.0	65571.0	178.0	76653.0
Buenos Aires	GRAL. PUEY.	178.0	41427.0	205.0	53294.0	169.0	67001.0
Buenos Aires	QUILMES	162.0	47353.0	167.0	60261.0	146.0	70705.0

Consulta (i): Ordenada provincia(des) y primarias(des)

En esta primera consulta podemos ver una relación entre la cantidad de jardines, primarias, secundarias y sus respectivas poblaciones. Esta tabla está ordenada primero por provincia (y ya que Buenos Aires empieza con B, está primera) y luego por primarias de manera descendente, y se puede ver cómo, a pesar de que el orden descendente es sobre las primarias, el resto de las variables tienden al descenso junto a ella también, lo que indica que existe una relación entre las variables de los niveles y poblaciones.

Consulta (ii): Para cada departamento informar la provincia y la cantidad de CC con capacidad mayor a 100 personas.

Provincia	Departamento	Cantidad de CC con cap >100
Buenos Aires	AVELLANEDA	20
Buenos Aires	LA PLATA	8
Buenos Aires	LOMAS DE ZAMORA	3
Buenos Aires	ALMIRANTE BROWN	2
Buenos Aires	GENERAL PUEYRREDON	2

Consulta (ii): Ordenada provincia(des) y CC(des)

En esta consulta nos resulta llamativa la velocidad a la que decrecen la cantidad de centros culturales con capacidad mayor a cien personas. Esto habla de que en general parecería haber pocos centros culturales por departamento, aunque algunos tengan muchos más que los demás departamentos dentro de su provincia.
Para sumar a la idea de que hay pocos departamentos con esta característica, la consulta entera tiene solo 56 filas, es decir, de los 516 departamentos que consideramos solo 56 tienen algún centro cultural con capacidad mayor a 100 personas.

Consulta (iii): Para cada departamento, indicar provincia, cantidad de CC, cantidad de EE (de modalidad común) y población total. No omitir casos sin CC o EE.

Provincia	Departamento	Cant_EE	Cant_CC	Poblacion
CABA	CABA	1782.0	296.0	3095454.0
Córdoba	CAPITAL	1136.0	30.0	1498060.0
Buenos Aires	LA MATANZA	977.0	2.0	1837168.0
Santa Fe	ROSARIO	817.0	36.0	1337958.0
Buenos Aires	LA PLATA	669.0	72.0	756074.0

Consulta (iii): Ordenada por EE (des), CC (des), prov. (asc) y depto. (asc).

En esta consulta vemos que hay una relación entre la cantidad de establecimientos educativos y la población del departamento, ya que, por mucho de estar ordenada primero por cantidad de establecimientos de manera descendente, la población también sigue una tendencia decreciente a lo largo de la tabla. Esta relación no es tan clara en la cantidad de centros culturales que hay, ya que las excepciones a la baja son mayores.
Por otro lado, en este ejercicio se nos pide no omitir casos sin centros culturales, lo que genera que muchos departamentos tengan NULL asignado en su campo de cantidad de centros. La frecuencia de estos NULLs parecería aumentar a lo largo de la tabla, es decir que a medida que decrece la cantidad de establecimientos educativos y población, se tiene menos información acerca de los centros culturales de ese departamento.

Consulta (iv): Para cada departamento, indicar provincia y qué dominios de mail se usan más para los CC.

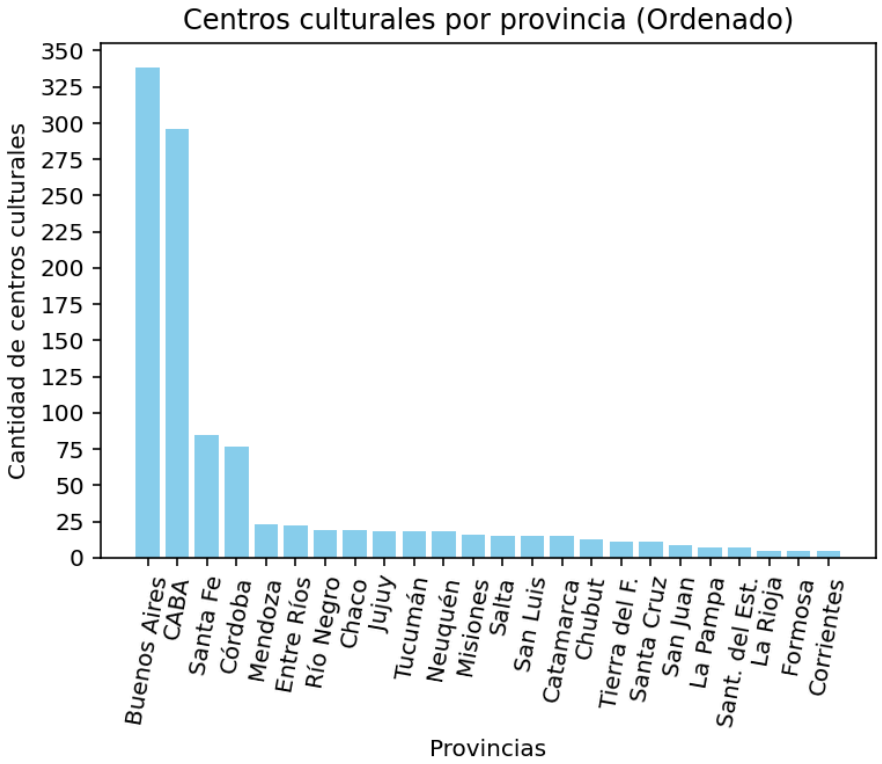
Provincia	Departamento	Dominio más frecuente en CC
Catamarca	VALLE VIEJO	gmail
Santa Fe	CASTELLANOS	gmail
Buenos Aires	GENERAL PUEYRREDON	gmail
Buenos Aires	SAN FERNANDO	yahoo
Buenos Aires	SAN NICOLAS	hotmail

Consulta (iv).

En esta consulta vemos que entre los dominios de mail más usados se encuentran gmail, hotmail y yahoo. Para tener mejor dimensión de cuán comunes son, agrupamos y contamos los dominios. Los resultados fueron que la suma de estos tres dominios da 164 de los 189 departamentos que aparecen en la tabla, con gmail siendo el más usado por mucho, con 112 en los que es el más frecuente.

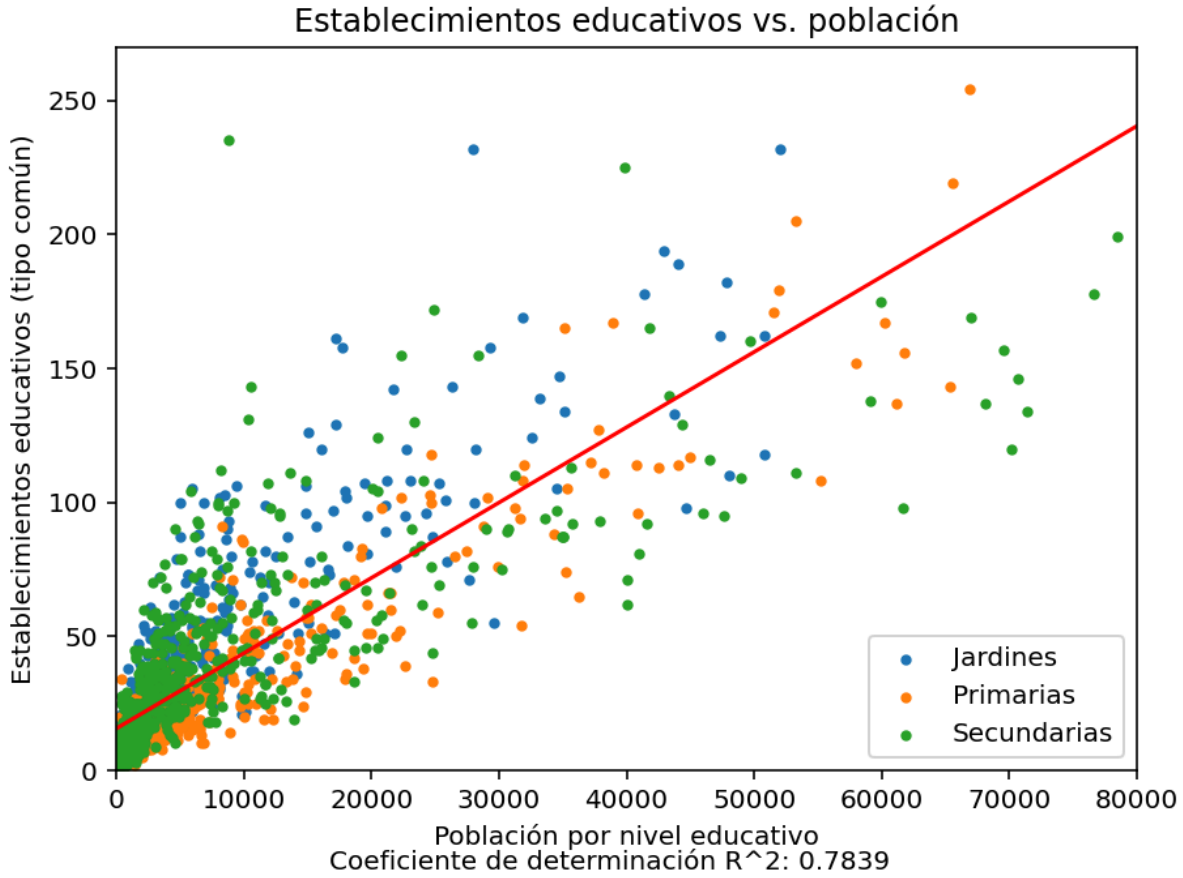
Visualización de datos

Gráfico (i): Cantidad de CC por provincia. Mostrarlos ordenados de manera decreciente por dicha cantidad.



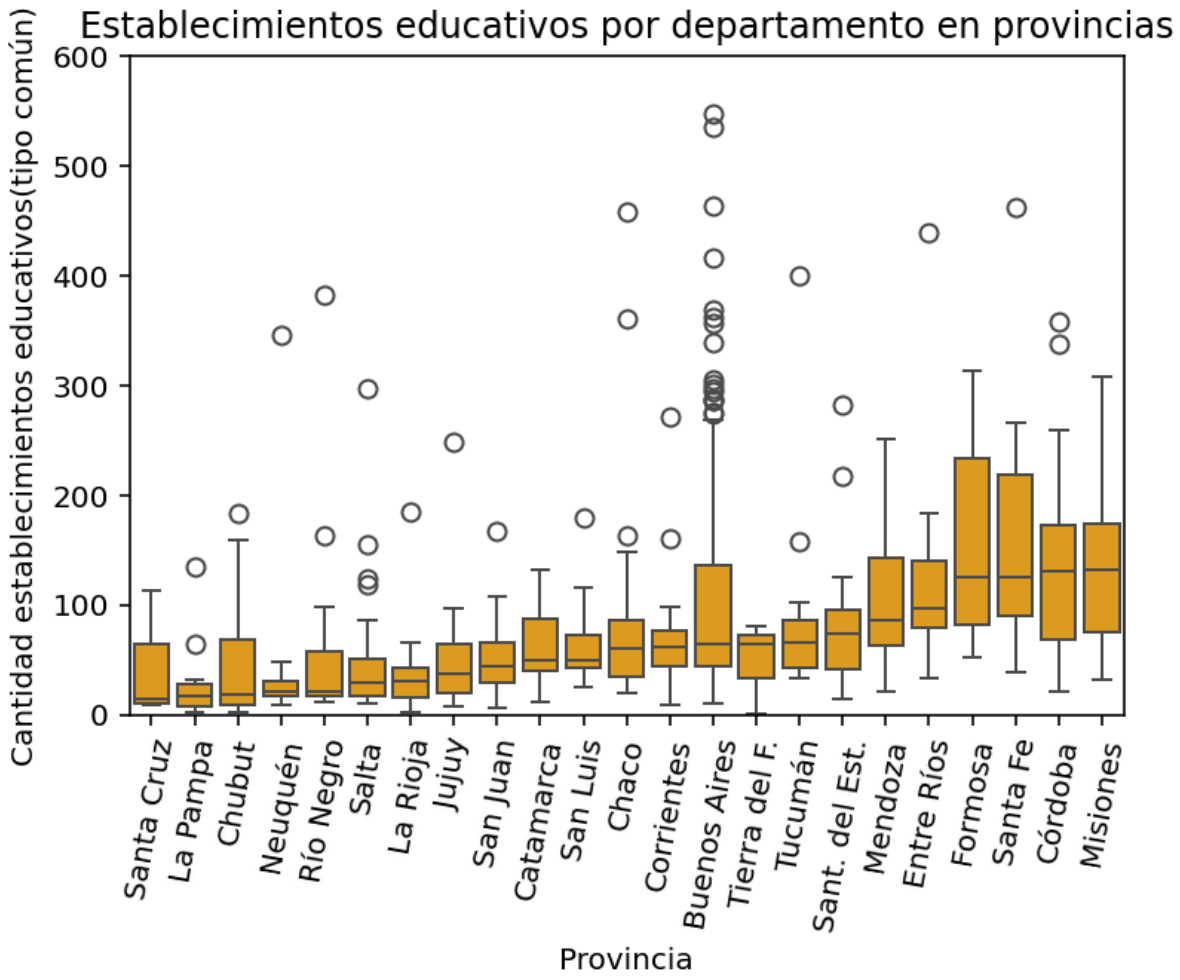
En este gráfico notamos que hay ciertas provincias en el país en las que hay muchos centros culturales más que en otras, y que esa diferencia es muy grande. Si sacamos las primeras cuatro provincias, no queda ninguna con más de 25 centros culturales.
Por otro lado, vemos una relación entre la población de las provincias y la cantidad de centros culturales, ya que las provincias a la izquierda del gráfico son a grandes rasgos las más pobladas del país. Vale aclarar que por mucho de qué CABA no sea una provincia en sí misma, tiene más población que algunas provincias del país.

Gráfico (ii): Graficar la cantidad de EE de los departamentos en función de la población, separando por nivel educativo y su correspondiente grupo etario (identificándolos por colores). Se pueden basar en la primera consulta SQL para realizar este gráfico.



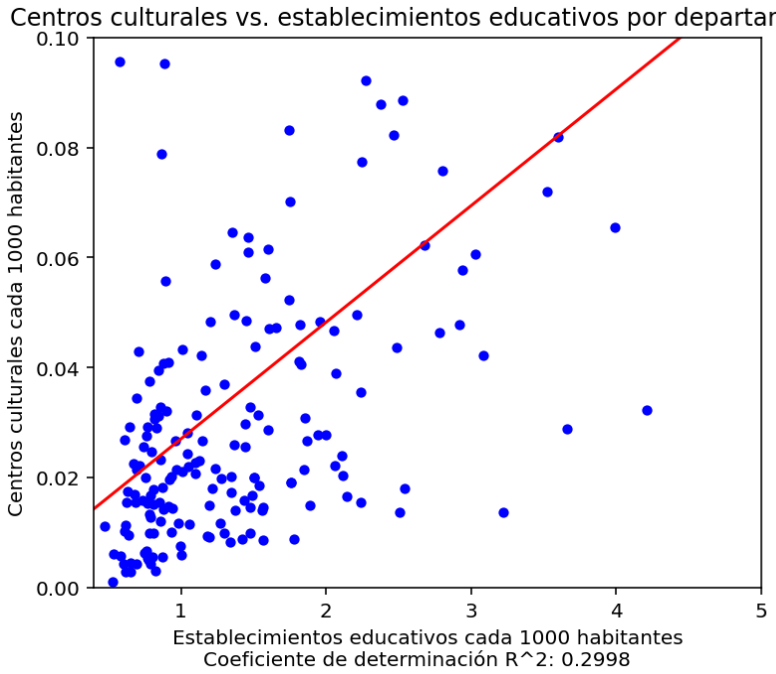
En este gráfico vemos como hay una relación directa y clara entre la cantidad de establecimientos educativos de tipo común y la población. Fuera de pocos *outliers*, el gráfico sigue de manera muy directa una relación: a mayor población, más establecimientos educativos. Esto se puede ver en el alto valor de R^2 (0,78) que indica que la población influye fuertemente la cantidad de establecimientos.
A grandes rasgos, parecería haber más jardines que primarios y secundarios cuando el tamaño de la población de su grupo etario es mediano, esto se puede ver en que los puntos azules están, en general, más arriba que los verdes y naranjas en la zona central del gráfico (entre 30.000 y 50.000). También llama la atención la ausencia de jardines después de los 60.000, que podría estar causado por una cuestión de distribución demográfica (más personas entre 12 y 18 que entre 0 y 5).

Gráfico (iii): Realizar un boxplot por cada provincia, de la cantidad de EE por cada departamento de la provincia. Mostrar todos los boxplots en una misma figura, ordenados por la mediana de cada provincia.

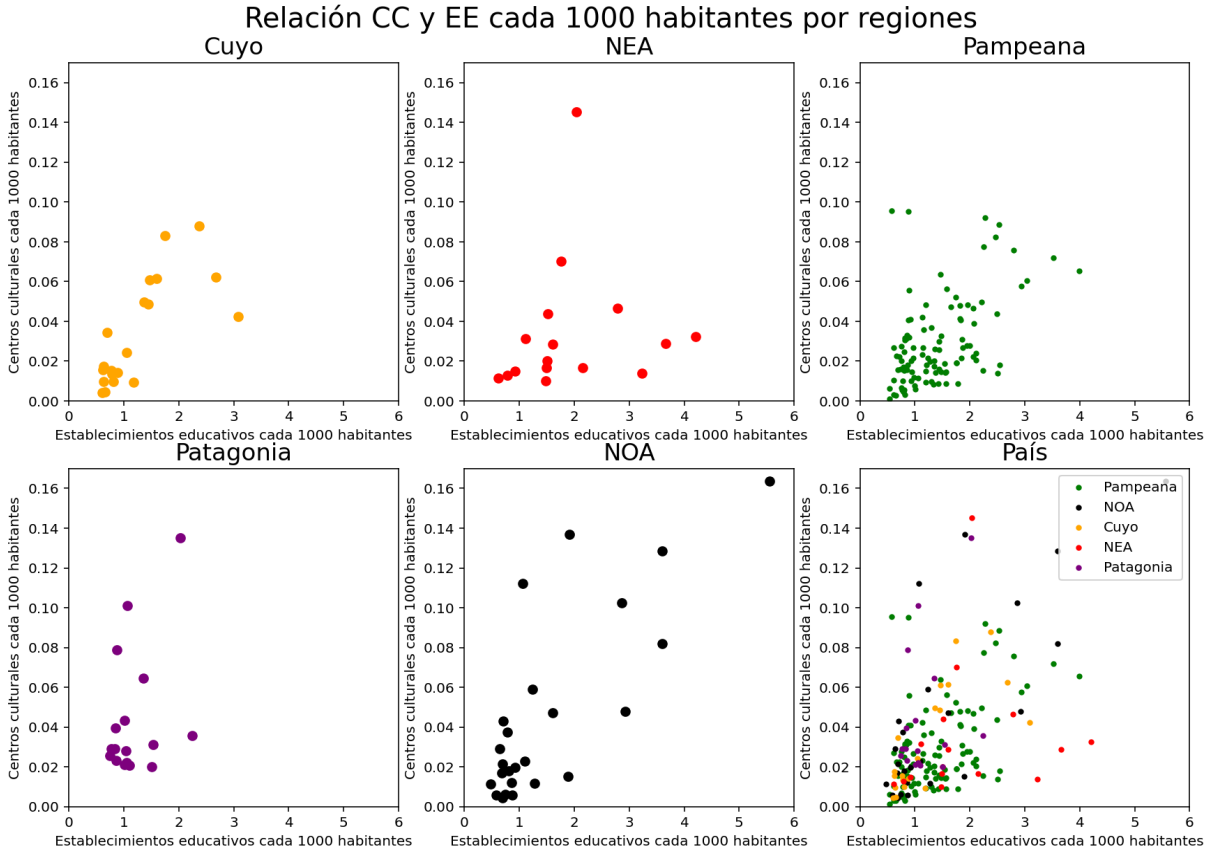


En este gráfico vemos cómo las provincias con mayor mediana en la cantidad de establecimientos educativos también son las que, en general, tienen un cuarto cuartil más largo, lo que quiere decir que tienen más departamentos con muchos establecimientos educativos.
Por otro lado, vale la pena notar la cantidad de *outliers* que tiene la provincia de Buenos Aires, que tiene también uno de los cuartos cuartiles más largos de todas las provincias. Esto quiere decir que tiene muchos departamentos con muchísimos establecimientos educativos, lo que se podría explicar con que es la provincia con mayor población del país. De todas formas, debemos tener en cuenta que Buenos Aires está dividida en muchos departamentos más que otras provincias. Todo esto produce que su distribución sea muy alargada.
También vale la pena notar que, a grandes rasgos, las provincias con mediana más alta tienden a ser las que más población tienen, aunque no es una relación que se vea tan claramente como se veía en el gráfico (ii) entre centros y población de provincia.
Aclaración: la ausencia de CABA en este gráfico se debe a que al considerarla un único departamento, sería solo una línea en la altura de la suma de todos sus establecimientos (1782), lo que está muy por fuera de este gráfico y arruina la visualización de la información del resto de las provincias.

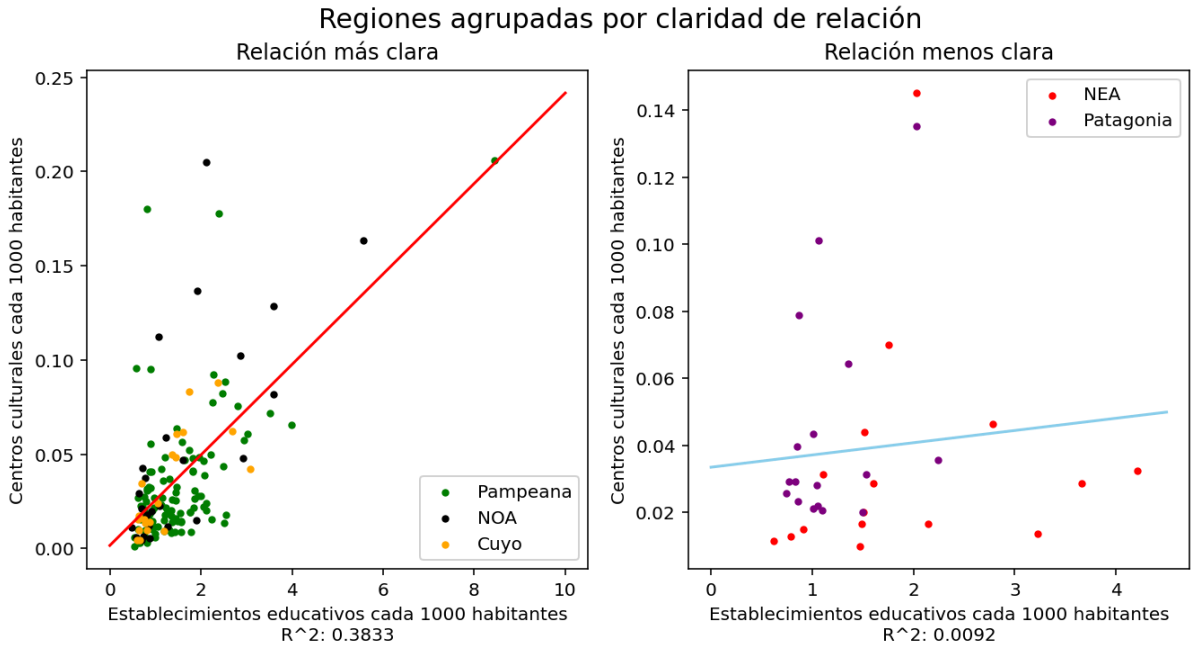
Gráfico (iv): Relación entre la cantidad de CC cada mil habitantes y de EE cada mil habitantes.



Este gráfico representa muy bien nuestra respuesta a la pregunta central del trabajo práctico ya que vemos que existe una relación débil entre la cantidad de establecimientos educativos y centros culturales. Es decir, la relación existe, pero no es clara.
La debilidad de esta relación se ve expresada en su R^2 de apenas 0,3. Con el objetivo de mejorar el estudio de esta relación, decidimos analizar por separado cinco regiones diferentes del país.



Lo que encontramos en estos gráficos fue que algunas regiones tienen una relación más clara entre centros culturales y establecimientos educativos. Cuyo, la región pampeana y NOA tienen una relación mucho más directa mientras que en la patagonia y NEA esta relación es más débil.
Para mostrar esto, agrupamos las regiones según la claridad de la relación que muestran.



Aquí vemos cómo, a nivel nacional, la claridad de la relación se ve afectada negativamente por la falta de correlación en las regiones de NEA y la Patagonia. En el gráfico de la izquierda la relación es más fuerte que lo observado en el gráfico a nivel nacional, ya que R^2 vale 0,38 (a nivel nacional era 0,29). Por otro lado, el bajísimo R^2 (0,009) del gráfico de la derecha muestra la falta de relación. De todas formas, aún en las regiones con mejor relación, el R^2 sigue siendo bajo.

Sección: Conclusiones

A partir de los datos obtenidos y los gráficos generados, especialmente los gráficos adicionales y el gráfico (iv), concluimos que existe una relación parcial entre la cantidad de establecimientos educativos y la cantidad de centros culturales en Argentina. Si bien no podemos afirmar que haya una relación clara y directa, consideramos que sí existe, aunque sea débil.

Observamos que la fortaleza de esta relación varía según la región del país. En algunas zonas, la correlación es más evidente, mientras que en otras es prácticamente inexistente. Esta disparidad hace que, al analizar la relación a nivel nacional, el vínculo entre ambos grupos sea menos claro. Es decir, las regiones donde se observa una relación más marcada quedan opacadas por aquellas en las que no se encuentra ninguna tendencia significativa.

Estas diferencias son comprensibles en un país tan extenso y diverso como Argentina. Por ello, consideramos que el análisis a nivel regional es clave para comprender las características de esta relación y cómo varía según las provincias estudiadas.