

A9: Creación de visualización

Alumno: Juan Manuel Penalta Rodríguez.

Índice

M2.859 – A9: Creación de visualización.....	1
1. Título de la visualización.....	3
2. Descripción de los datos.....	3
3.- Herramientas.....	4
4.- Exploración y procedimiento de generación de los datos.....	4
Financiación.....	5
Recursos TIC.....	5
Capacidades TIC.....	6
6.- Visualización.....	6
Financiación.....	7
Recursos TIC.....	7
Capacidades TIC.....	8
5.- Preguntas y respuestas.....	9

1. Título de la visualización.

“Escuelas PISA 2018”

2. Descripción de los datos

El conjunto de datos que utilicé son los datos recogidos por el **informe PISA relativo a escuelas**.

El informe PISA es un conjunto de pruebas de diagnóstico y encuestas realizado por la OCDE, para analizar el ámbito educativo de los países miembros.

La recogida de datos de los informes se realiza cada tres años, existiendo informes desde el año 2000 hasta el 2018.

Los datos que se usaron corresponde a la encuesta realizada a las encuestas. También los datos resumen que publica la OCDE con los rankings en Lengua, Ciencia y Matemáticas, que obtienen los alumnos en las pruebas.

Estos datos están disponibles en la página de la OCDE, bajo la url: <https://www.oecd.org/pisa/data/>

En concreto los del año 2018 en <https://www.oecd.org/pisa/data/2018database/> en formato SPSS y SAP.

En la misma página está el codebook de los datos y los cuestionarios utilizados para las encuestas.

Casi todos los informes se centran en los **resultados de los estudiantes** y hay pocos (al menos que yo conozca) informes que exploten los datos relativo a los **docentes** y los **centros escolares**, que también son encuestados en este proceso. Por eso mi intención es trabajar con los datos de:

- [School questionnaire data file](#) (21.903 entradas, con 197 columnas)

Que contienen datos sobre las encuestas realizadas a las escuelas, tantos datos categóricos como cuantitativos.

Los datos descargados se pueden usar libremente con la única limitación de citar su origen. Se pueden ver las condiciones de uso de los datos en la Sección I apartado c, de la página de términos y condiciones de la OCDE:

<https://www.oecd.org/termsandconditions/>

Adicionalmente se utiliza un fichero con los nombres de los países y su identificador(código de tres letras indicativo del país), necesario para poder realizar una de las visualizaciones. Este fichero fue de realización propia.

También se utiliza un fichero con la competencia de cada país en Lengua, Ciencias y Matemáticas, descargado a partir de un explorador de datos disponible en <https://pisadataexplorer.oecd.org/ide/idepisa/>

3.- Herramientas

Para realizar el análisis de los datos se utilizó la herramienta Jupyter [<https://jupyter.org/>].

Para la realización de las gráficas la librería D3.js [<https://d3js.org/>].

Y para el formato de la página web se utilizó el framework bootstrap [<https://getbootstrap.com/>].

4.- Exploración y procedimiento de generación de los datos.

Se generan los siguientes ficheros de datos:

- ranking_paises_pisa_2018.csv: Competencia de los países en capacidad lingüística, científica y matemática.
- financiacion.csv: Datos relativos a las formas de financiación de los centros.
- recursostic.csv: Ratio de recursos TIC disponibles por alumno.
- capacidadesttic.csv: Capacidades de la escuela o del profesorado para enseñar usando herramientas TIC.

El primero fichero se generó y descargó directamente con la herramienta “Pisa Data Explorer”, ya comentada anteriormente. Se descarga en formato excel, se realizaron modificaciones para unificar los tres datos en una única hoja, añadir la media y el código de tres caracteres de cada país. Este proceso se hizo de manera manual.

Para la generación de los otros tres ficheros se utilizó la herramienta Jupyter:

Se importo el fichero, en formato SAS, con los datos correspondientes a los datos de las escuelas del año 2018.

Para cada uno de los casos, se selecciono los campos necesarios, se hizo una agrupación por países y se calculó la media para cada campo seleccionado.

Financiación

Los campos utilizados para obtener la **financiación** de los centros fueron: 'SC016Q01TA', 'SC016Q02TA', 'SC016Q03TA', 'SC016Q04TA', que representan las aportaciones por Gobierno, Padres, Benefactores y Otros respectivamente.

El valor de estos campos es el porcentaje del total de financiación del centro que representa. La suma de los cuatro campos debe dar un 100%.

Para la generación de la media nacional en cada campos, se realizó la suma de cada campo y se dividió entre el total de los tres campos.

Recursos TIC

Los campos utilizados para la obtención de los recursos TIC disponibles para alumnos y profesores fueron los siguientes:

'SC018Q01TA01', 'SC018Q01TA02', 'SC004Q01TA', 'SC004Q02TA', 'SC004Q03TA', 'SC004Q04NA', 'SC004Q05NA', 'SC004Q06NA', 'SC004Q07NA'

Los tres primeros representan el total de profesorado del centro, el total de chicos y el total de chicas.

Los campos de recursos TIC aparecen como totales por centro, de cara a mostrarlos en la visualización se realizó la división de cada campo por el número total de alumnos o de profesorado en función de lo que correspondiera. Obteniendo así un ratio de recurso/alumno o recurso/profesor que son los utilizados en la visualización.

Capacidades TIC

Los campos utilizados para obtener las capacidades TIC del alumnado y profesorado son los siguientes: 'SC155Q01HA', 'SC155Q02HA', 'SC155Q03HA', 'SC155Q04HA', 'SC155Q05HA', 'SC155Q06HA', 'SC155Q07HA', 'SC155Q08HA', 'SC155Q09HA', 'SC155Q10HA', 'SC155Q11HA'

Los campos de capacidades son categóricos, en cada uno de ellos se realiza una valoración entre los siguientes cuatro valores: “totalmente en desacuerdo”, “en desacuerdo”, “de acuerdo”, “muy de acuerdo” indicando si el nivel es suficiente.

Lo primero que se realizó fue la conversión de cada categoría en un valor numérico entre 1 y 4. Posteriormente se calculo la media para cada país de cada uno de los campos.

6.- Visualización

Al ser los datos agrupados por países la primera visualización es un mapa sobre el que se muestra la competencia media de los alumnos alcanzada en las pruebas PRISA. Cada país aparece coloreado con un tono distinto de azul indicando una menor o mayor competencia media.

Sobre el mismo mapa se puede realiza la selección de dos países, que van a ser los utilizados como filtro para el resto de las visualizaciones.

Al lado del mapa opté por mostrar un gráfico de barras donde se puede visualizar la competencia alcanzada por los dos países seleccionados en lengua, ciencias y matemáticas.

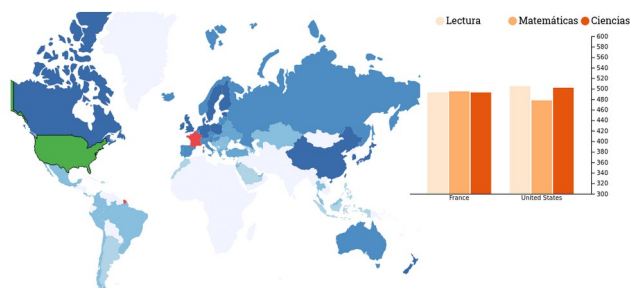


Figura 1: Mapa de competencias con países seleccionados

El resto de las visualizaciones tienen como objetivo la comparación de los datos entre los países seleccionados.

Financiación

Como los datos de visualización son porcentuales decidí la utilización de una gráfica de barras en pila porcentual.

Para cada uno de los dos países seleccionados se muestra una barra con las cuatro categorías de un color distinto. Al pasar el ratón por encima se puede obtener el valor que representa, tal como se muestra en la figura 2.

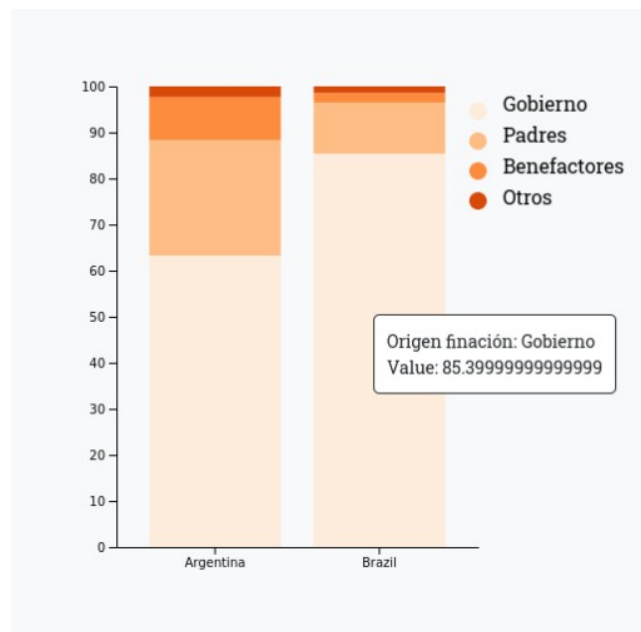


Figura 2: Financiación de escuelas con detalle de leyenda

Recursos TIC

Opté aquí por una gráfica de tipo lollipop que muestra para cada categoría los valores de cada país. Esta gráfica permite apreciar claramente la diferencia existente entre ambos valores, además del propio valor. Se puede apreciar en la figura 2, en la que se ve la diferencia de recursos entre España y Francia.

M2.859 – A9: Creación de visualización

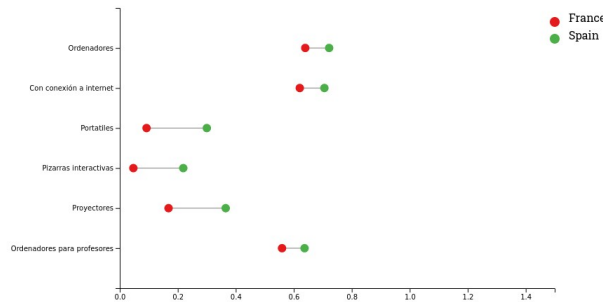


Figura 3: Recursos TIC disponible

Capacidades TIC

En este caso la cantidad de categorías es bastante grande, y me interesaba más mostrar una diferencia global entre los dos países seleccionados. Para ello elegí una gráfica de tipo radar, donde cada categoría se muestra en un radio alrededor de un centro único. Los puntos del mismo país se unen generando un área que representa la capacidad global para el país, más allá de que en una determinada categoría se obtenga un mejor resultado, como se puede ver en la figura 3, comparando China y Argentina.

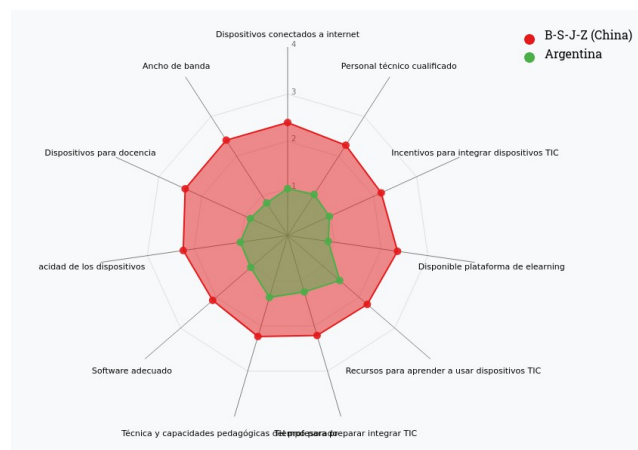


Figura 4: Capacidades para enseñar usando TIC

En un principio pensaba permitir que en cada una de estas gráficas se permitiera seleccionar si mostrar los datos de centros financiados con fondos públicos, fondos privados o ambos, pero por falta de tiempo no pude realizarlo.

La visualización de los datos se puede consultar en:

<https://jpenalta.github.io/uoc2020M859A9/>

5.- Preguntas y respuestas

De los dos conjuntos de datos se podrían analizar las siguientes cuestiones:

- Sobre los centros:
 - Recursos disponibles en los centros para el profesorado y alumnado.
 - Tipos de centros: tamaño, tamaño de la población, nivel de estudios impartido.
 - Alumnado y profesorado del centro. Ratios alumno/profesor

Con la visualización se puede comparar los datos de dos países seleccionados.

Por ejemplo:

- Al comparar China, país con mejor competencia mundial en PISA, con EEUU; China tiene menos recursos TIC disponibles para sus alumnos y profesores, pero en cambio tiene a sus profesores más preparados para utilizar estos recursos que EEUU.

Se podría intuir que una mejor preparación del profesorado tiene un mayor peso en la enseñanza que unos mejores recursos.

- Al comparar España con Italia, dos países con parecidos niveles de competencia, la procedencia de la financiación de los colegios no afecta a esta. Italia tiene un 40% de procedencia de la financiación de manos privadas, frente a un 20% en España. Se puede apreciar este mismo efecto comparando más países con niveles de competencia similares, recursos y capacidades similares que tienen muy diversas combinaciones de financiación de sus escuelas.

6- Visualización sobre los datos.

La visualización los datos se puede consultar en:

<https://jpenalta.github.io/uoc2020M859A9/>

El código utilizado para la generación de los datos, los datos generados, así como el código de la página se puede consultar en:

<https://github.com/jpenalta/uoc2020M859A9>