



Práctica 2: Creación de la visualización y entrega del proyecto

Visualización de datos

Curso 2020-21

Jordi Lladó Valero

1. Descripción y enunciado

En esta segunda parte de la práctica el estudiante tendrá que desarrollar una visualización de datos que demuestre sus conocimientos del ámbito así como el uso de diferentes herramientas y técnicas. Además de la visualización, el estudiante tendrá que librar un informe que describa el proceso de creación seguido y las decisiones de diseño que se hayan tomado a lo largo del desarrollo, así como una pequeña presentación de la visualización, el conjunto de datos que se visualiza, las preguntas que se responden y los elementos interactivos disponibles como parte de la misma visualización.

La visualización debe incluir los elementos siguientes:

1) Título de la visualización donde se presentan la visualización realizada. URL de la visualización y del código. Y descripción corta del documento y del que se presenta

1.1) El código es tiene que publicar en un repositorio público de GitHub con una licencia de código abierto. Tiene que contenga los archivos necesarios para correr la aplicación en local con la ayuda de un servidor web si fuera necesario.

1.2) La URL de la visualización tiene que ser pública y poderse explorar sin login. Por ejemplo GitHub pages (<https://pages.github.com/>), surge.sh o subir la aplicación a un espacio web propio accesible o compartir un url accesible sin usuario y contraseña.

2) Explicar razonadamente qué preguntas responde la visualización presentada y qué uso puede tener por un usuario tipo.

3) Descripción técnica del proyecto: lenguajes, librerías, licencias, descripción técnica del proyecto.

4) La/s visualizaciones realizadas

2. Presentación de la visualización

El siguiente proyecto y visualización lleva por título:

‘Enseñanza superior mundial (2018): titulación, staff y género ‘

La visualización resultante del proyecto se realizó con Tableau y se puede encontrar en el enlace:

Enlace Tableau:

https://public.tableau.com/views/Practica1_16231030632960/Prctica1?:language=en-US&:display_count=n&:origin=viz_share_link

La presente visualización muestra los datos de la enseñanza superior (terciaria) en el mundo en el año 2018. Los datos se han presentado en un mapa mundial y se pueden distinguir tres partes: Porcentaje de la población que tiene un máster, porcentaje de la población que tiene un doctorado y matrícula enseñanza superior, personal trabajador y ratios. Las distintas partes dan una visión global de la enseñanza a nivel mundial y permite comparar distintos países, así como debatir las desigualdades de género que se pueden encontrar en algunos países.

Toda la información técnica del proyecto (documentación, datos, código y licencia) se encuentra disponible en el siguiente repositorio:

https://github.com/JFox83/Proyecto_visualizacion

3. Explicación de la visualización y visualización de datos

El tema global del siguiente proyecto es la educación terciaria o superior en el año 2018, es decir, es la enseñanza o trayectorias formativas postsecundarias. La idea principal es comparar la educación entre los distintos países del mundo, ya sea aportando información referente a masters, doctorado, o bien des del punto de vista de la perspectiva de género. La visualización consta de una secuencia de tres mapas mundiales. El primero de ellos se muestra la ratio de matrícula de enseñanza superior según un degradado de color naranja (valores elevados naranja intenso hasta valores mínimos naranja transparente). A continuación, se muestra una primera imagen (figura 1) sobre la visualización con este atributo:

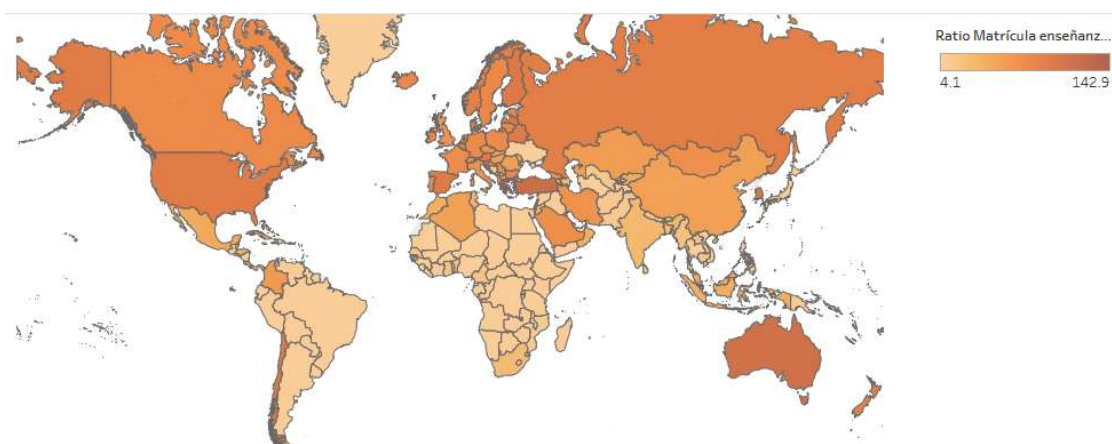


Fig. 1 Visualización principal referente a la ratio de matrícula de enseñanza superior

Con ello nos podemos formular la pregunta: ¿Qué países tienen mayor matrícula en la enseñanza superior?

También se han introducido dos variables más en la visualización (ratio alumnos/profesor y el personal académico en la enseñanza superior (% femenino)) con el objetivo de comparar la calidad de la enseñanza que reciben los alumnos, ya que una ratio muy elevada hace referencia a que un profesor tiene gran cantidad de alumnado. Y por otra parte, comparar el porcentaje mujer – hombre que trabajan en la enseñanza superior con el fin de tener una paridad similar entre géneros. Con esto se pueden formular distintas preguntas cómo ¿Que ratio alumnado/profesorado tiene Algeria?

¿Qué porcentaje de mujeres trabajan en la enseñanza superior en España? ¿Y en Afganistán?

Estas preguntas se pueden resolver utilizando la visualización y el filtro que tienen, poniendo el nombre del país en cuestión o buscándolo en la visualización (se muestra ejemplo en la figura 2).



Fig. 2: Búsqueda información mediante cursor sobre país (ej. Afganistán)

El siguiente mapa hace referencia al porcentaje de población mayor a 25 años que se ha titulado en un master. Para llegar a este mapa, se ha creado una secuencia (en formato historia) donde se puede avanzar a las distintas pestañas (Pestaña 1: matrícula, ratio Alum/prof y personal; Pestaña 2: Máster; Pestaña 3: Doctorado), Figura 3.

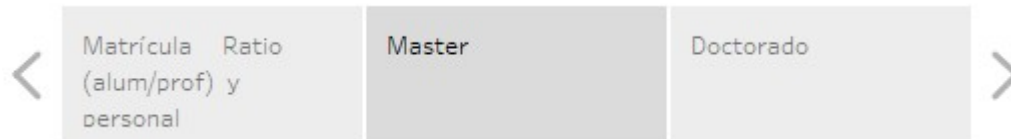


Fig. 3 Pestaña para poder pasar de mapa en la visualización

La información que contiene este mapa es el porcentaje de mujeres que han obtenido un master respecto al total de la población. En este caso se ha utilizado una degradado de azules para mostrar en que países ha más o menos porcentaje. También se ha incluido de forma adicional el porcentaje de hombres que han obtenido un máster con el fin de comparar ambos géneros. En la siguiente figura 4 se muestra el mapa comentado. Las preguntas relacionadas con este tema podrían ser: ¿Qué países tienen mayor cantidad de mujeres con un máster? ¿Qué porcentaje de mujeres con máster tiene Estados Unidos? ¿Y de hombres?

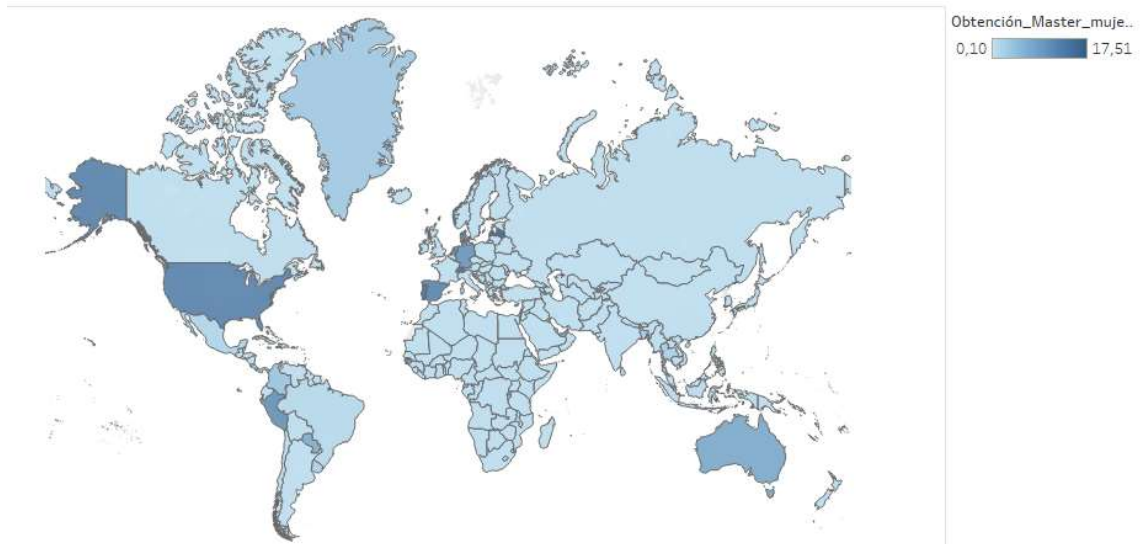


Fig. 4 Porcentaje de mujeres que tienen máster

Por último, se ha realizado una tercera visualización referente al porcentaje de mujeres que tienen doctorado. El proceso seguido es similar al anterior, pero en este caso se ha utilizado un degradado de color verde para mostrar la cantidad en el mapa (figura 5).



Fig. 5 Porcentaje de mujeres que tienen doctorado

Las preguntas que se pueden realizar en este caso son:

¿Qué países tienen mayor cantidad de mujeres con un doctorado? ¿Qué porcentaje de mujeres con doctorado tiene España? ¿Y de hombres?

Por último, se muestra la visualización final (figura 6)

Enseñanza superior mundial (2018): titulación, staff y género

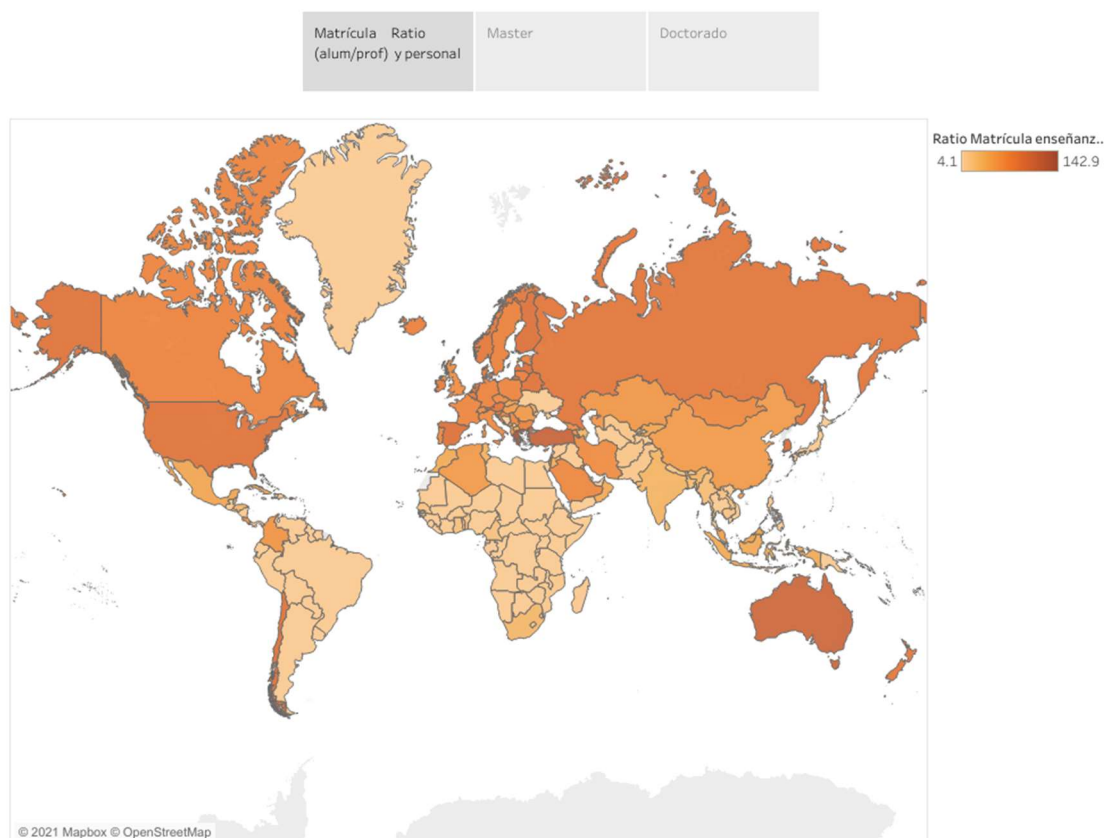


Fig. 6 Visualización final

4. Descripción técnica

El siguiente proyecto se ha utilizado la base de datos 'World bank indicators collection' que se puede encontrar en el siguiente enlace:

<https://www.kaggle.com/ploverbrown/world-bank-indicators-collection>

El dataset incluye distintos indicadores agrupados en temáticas. Estas temáticas son: población, densidad de población, crecimiento de la población, población femenina y masculina, educación, cambio climático, comercio, infraestructura, pobreza, desarrollo social, medio ambiente, agricultura y rural, crecimiento económico, salud, sector privado, sector público, sector financiero, ciencia y tecnología, deuda externa, ... Cada una de estas temáticas se encuentra en una hoja de Excel y contiene varios indicadores o variables.

En este caso, se escogió la temática de educación y se copió los datos de esta hoja de Excel y se guardó en un archivo separado ('Education', Nota: se tuvieron problemas en el momento de abrir con R el archivo concreto). La temática educación contiene de forma general 65 atributos (una serie temporal que va desde el año 1960 hasta el 2020, y 4 atributos más: Country Name, Country Code, Indicator Name y indicator code) con un total de 42768 (representa 264 países con 162 variables cada uno de ellos). La variable indicator Name representa la información temporal de 162 variables distintas relacionadas con el mundo de la educación, ya sea primaria, secundaria y terciaria, género, gasto educación, población, ...

Una vez escogida la hoja y archivo 'Education' con las variables de interés, se procede a abrir el archivo, limpiar y transformar los datos mediante un código en R. La primera tarea en el código es descargar las librerías 'readxl' y 'knitr' que permiten leer archivos Excel y manipular, transformar datos respectivamente. Después se procede a la lectura del archivo 'Education' y se nombran de nuevo los atributos. A continuación, se hace la selección de las variables de nuestro interés para la visualización. Estas variables se resumen en:

- Country: Country Name
- Población (% mayor a 25 años) : Population ages 15-64 (% of total population)
- Gasto 1 (% PIB): Total Government expenditure on education(% of GDP)
- Gasto 2 (% gasto gobierno)": Government expenditure on education, total (% of government expenditure)
- Gasto educación terciaria:"Expenditure on tertiary education (% of government expenditure on education)"

- Gasto total educación (% gasto público): Current education expenditure, total (% of total expenditure in public institutions)
 - Personal académico (%mujer): Tertiary education, academic staff (% female)
 - Matrícula terciaria (% Hombres): School enrollment, tertiary, male (% gross)
 - Matrícula terciaria (%mujeres): School enrollment, tertiary, female (% gross)
 - Ratio matrícula enseñanza superior: School enrollment, tertiary (% gross).
Definición: Ratio matrícula total (Gross enrollment ratio): es la ratio de matriculación, independientemente de la edad, a la población del grupo de edad que corresponda al nivel de estudios que se muestra (en este caso enseñanza superior). En la educación avanzada, con título avanzado o no, requiere una condición mínima de admisión.
-
- "Ratio" # "Pupil-teacher_ratio_tertiary"
 - Obtención Master total: "Educational attainment, at least Master's or equivalent, population 25+, total (%) (cumulative)"
 - Obtención Master Hombres: "Educational attainment, at least Master's or equivalent, population 25+, male (%) (cumulative)"
 - Obtención Master mujeres: "Educational attainment, at least Master's or equivalent, population 25+, female (%) (cumulative)"
 - Obtención doctorado total: "Educational attainment, Doctoral or equivalent, population 25+, total (%) (cumulative)"
 - Obtención doctorado Hombres: "Educational attainment, Doctoral or equivalent, population 25+, male (%) (cumulative)"
 - Obtención doctorado Mujeres: "Educational attainment, Doctoral or equivalent, population 25+, female (%) (cumulative)"

Para escoger estas variables se han realizado dos procesos. El primero se guardó la serie temporal de cada una de las variables (se encuentran en filas en concreto al país) y después se escogió los datos del año 2018 de estas variables que servirán para realizar la visualización general. También en caso de que sea posible o haya más tiempo se podrán recuperar las series temporales para su visualización. Una vez realizada la selección y transformación de los datos se guardan en una tabla y fichero .txt, xls o csv según necesidad (nota: los formatos xls y csv dieron problemas cuando se utilizaron en Tableau, y sólo se muestra como obtener el txt). El código final utilizado se encuentra en el siguiente enlace o en la carpeta comprimida enviada. Enlace:

https://github.com/JFox83/Proyecto_visualizacion/blob/main/Pr%C3%A0ctica%20%20visualitzaci%C3%B3.Rmd

El dataset resultante se ha guardado con el nombre final y en los dos formatos .txt y .xlsx. Este data set resultante contiene 17 atributos con 264 entradas que representan cada uno de los países (la explicación de ellos ya se explicó anteriormente).

Una vez con los datos, se procede a la realización de la visualización mediante 'Tableau'. Esta visualización tiene formado de 'story' permitiendo cambiar de datos según el tema o información que se quiere obtener.

Para finalizar se ha creado una licencia tipo GNU General Public License v3.0 (https://github.com/JFox83/Proyecto_visualizacion/blob/main/LICENSE). Esta licencia permita la distribución, modificación y uso comercial de la visualización siempre que se cite el autor, la fuente original y las modificaciones.

5. Conclusiones y documentos de entrega

En esta práctica se ha realizado la visualización: 'Enseñanza superior mundial (2018): titulación, staff y género ' donde muestra distinta información sobre la enseñanza superior en los distintos países del mundo en el año 2018. Se ha trabajado con la base de datos extraídos del Population Data World Bank, más en concreto con los datos referentes a educación (Education). El dataset original contenía series temporales de datos de educación de los distintos países del mundo. A partir de estos datos se han limpiado y transformado mediante código R, y después se han utilizado Tableau para la realización de la visualización.

La visualización da respuesta a varias preguntas como que porcentaje de mujeres y hombres tienen máster, doctorado, ratios de alumnado/profesorado en la enseñanza superior, ...

Por último se adjunta un archivo.zip con toda la documentación y archivos necesarios para la realización de la práctica

6. Bibliografía

Index mundi (2019) School enrollment, tertiary (% gross), URL:
<https://www.indexmundi.com/facts/indicators/SE.TER.ENRR>