



# PRA2: CREACIÓN DE LA VISUALIZACIÓN Y ENTREGA DEL PROYECTO

VISUALIZACIÓN DE DATOS

PEDRO UCEDA MARTÍNEZ  
MÁSTER EN DATA SCIENCE  
Universitat Oberta de Catalunya



Contenido

Introducción ..... 2

    Título de la visualización..... 2

    Objetivos ..... 2

    Qué preguntas se quieren resolver ..... 2

    Desarrollo de la visualización ..... 2

        Los datos: ETL ..... 2

        Mockups ..... 3

        Tecnología..... 3

La visualización..... 4

    Descripción..... 4

    Principios ..... 4

## Introducción

Durante esta práctica, se ha desarrollado la visualización para la cual se escogieron los datos en la PRA1 de esta misma asignatura. En este caso, se seleccionó el conjunto de datos ofrecido por la Agencia para el Empleo de Madrid, que recogía, entre otras cosas, **los objetivos profesionales de las personas en situación de desempleo de la Villa de Madrid**.

En el presente documento se procede a explicar el **trabajo realizado para llevar a cabo la visualización objeto de esta práctica**.

## Título de la visualización

La visualización se titula: **Personas inscritas en la Agencia para el Empleo de Madrid**.

## Objetivos

Mediante el uso de esta visualización, se pretende que el espectador pueda explorar los datos para ver **cuáles son los intereses mayoritarios de los madrileños que se encuentran en situación de desempleo** y si estos **varían según la parte de la ciudad en la que vivan**. Asimismo, también se pretende **conocer su distribución, tanto en el tiempo como en el espacio**.

## Qué preguntas se quieren resolver

Las principales preguntas a las que **se quiere dar la respuesta** mediante el uso de esta visualización son:

- ¿Se distribuye de la misma manera el número de inscritos a la Agencia para el Empleo de Madrid en todos los distritos de la ciudad? ¿Hay distritos que presentan mayor número de inscritos con una diferencia notable?
- ¿Existe alguna tendencia a lo largo del tiempo según la cual varíe el número de inscritos en la Agencia?
- ¿Existen más extranjeros que españoles en la Agencia?
- ¿Las profesiones más demandadas son las mismas en toda la ciudad o, por el contrario, varían según los distritos?

## Desarrollo de la visualización

Los datos: ETL

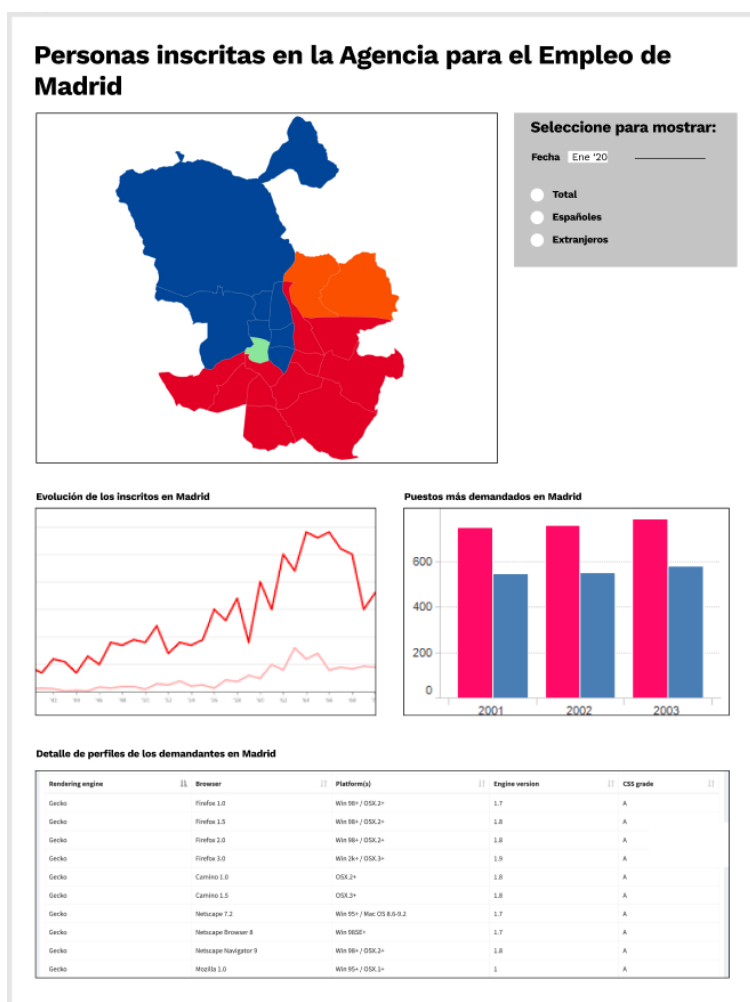
Como ya se ha mencionado, se extrajeron los datos del siguiente enlace, en **formato CSV**:

<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=710e760a581a7610VgnVCM1000001d4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default>.

Utilizando R, y, en concreto, **RMarkdown**, se procedió a **normalizar tales datos**, ya que los del 2018 presentaban distinto formato en la fecha de inscripción. Además, se fusionaron en un único fichero, `agencia_empleo_inscritos_agrupados.csv`, que será el que alimente nuestra futura visualización. Los ficheros que se mencionan en este apartado pueden consultarse en la carpeta procesamiento-datos del repositorio de github <https://github.com/pucedam/pra2-visualizacion-datos>.

## Mockups

De manera similar a cómo se trabajó en la PEC3, se llevó a cabo el **siguiente mockup para hacerse una idea de cómo sería la visualización final.**



Como se puede apreciar en la imagen, se ofrece **un mapa de la ciudad de Madrid en el que se señalan los distritos que la componen**. Asimismo, se mostrará la evolución de la **cantidad de personas inscritas en la Agencia**, las **20 profesiones más demandadas** y una tabla con el **detalle de los registros**.

El mapa tendrá dimensión temporal, y si se sitúa el ratón sobre él, se detallará cuántas personas inscritas están y cuales de ellas son españolas, ciudadanos comunitarios o ciudadanos extracomunitarios. **Si se hace *click* sobre algún distrito, se actualizarán todos los gráficos para reflejar la información sobre el mismo.**

## Tecnología

El deseo de **utilizar un mapa** para reflejar la distribución de los datos ha **influido** notablemente en la **elección de la tecnología a usar**. Prácticamente me he visto obligado a utilizar **D3.js** para poder plasmar los datos sobre un mapa de los distritos de la Villa de Madrid.

Es por ello que se ha optado por utilizar **observablehq** de nuevo **como herramienta para desarrollar la visualización**, utilizando en su interior D3.js para elaborar el mapa

junto a los gráficos de la visualización. Asimismo, también se ha utilizado Markdown para meter texto.

Cabe mencionar que se ha invertido un gran tiempo solamente en poder dibujar este mapa: Normalmente ya existen **desarrollos reusables** con mapas de países o incluso de comunidades autónomas o regiones, pero, sin embargo, **no es el caso para las ciudades**. En concreto, sobre Madrid, **no se encontró ningún mapa que se pudiera aprovechar**. Puesto que el mero hecho de hacer un mapa coroplético con estos distritos parecía un factor decisivo, se investigó. De este modo, se obtuvieron los datos en formato `geojson` (`json`) del siguiente enlace [https://raw.githubusercontent.com/codeforamerica/click\\_that\\_hood/master/public/data/madrid-districts.geojson](https://raw.githubusercontent.com/codeforamerica/click_that_hood/master/public/data/madrid-districts.geojson) y se les dio forma para visualizar el mapa. Una vez realizado este proceso, se procedió a trabajar sobre el mismo con los datos.

La elaboración de las otras gráficas que componen la visualización sí que eran más familiares para el alumno.

## La visualización

La visualización, titulada Personas inscritas en la Agencia para el Empleo de Madrid, puede encontrarse en el **siguiente enlace público**:

<https://observablehq.com/@pedrouceda/intereses-oficina-empleo-madrid>

### Descripción

Tal y como puede verse en **el código fuente de la página**, se ha cargado el dataset anteriormente mencionado en formato **CSV** y se ha realizado las **distintas agrupaciones necesarias para visualizar el mapa y el resto de gráficos**.

Existen un **datepicker** en el que se ha de seleccionar la fecha indicada para que se coloree el mapa. Nótese que, en este caso, **lo que interesa es cambiar de mes, puesto que los datos de los que dispone la visualización son mensuales.**

### Principios

Como visualización interactiva de la que se trata, se han seguido los principios marcados por **Shneiderman**. Se puede destacar:

- **Overview.** Se visualiza, en primera instancia, los datos de la Agencia para todo Madrid, en todas las gráficas disponibles. Asimismo, en el caso del mapa, existe una dimensión temporal que indica cómo han ido evolucionando los inscritos en los distritos. No procede hacer zoom en este caso, ya que la visualización es bastante esquemática.
- **Filtrado o selección de atributos.** Se permite ver los datos totales pero también distinguir entre españoles, comunitarios o extracomunitarios. Todas las gráficas evolucionarán para mostrar los datos seleccionados.
- **Detalle.** En caso de situar el ratón sobre un distrito, se dará detalle de el total de inscritos en un mes dado, si son españoles, comunitarios, o extracomunitarios.

- **Brushing.** Tanto para el selector de tipo de inscritos como cuando se hace *click* sobre un distrito, el resto de gráficas pasará a reflejar la información sobre el distrito apenas seleccionado.

Mencionar que se han seleccionado los colores del siguiente modo:

- El **total de inscritos** con **rojo**.
- Los **españoles** con **naranja**.
- Los **comunitarios**, **azul**.
- Los **extranjeros**, **verde**.

Salvo la decisión de visualizar los datos de los inscritos comunitarios en azul, el resto de colores se ha seleccionado de manera arbitraria. Sin embargo, sí que, al seleccionar un tipo de inscrito en concreto, el color de las gráficas se actualizará, quedando coherente.