GUÍA GITHUB:

Este material documenta el proyecto llamado Cerebros en movimiento. Ha sido un viaje de dos semanas cuya dirección ha ido mutando hasta obtener el resultado final.

Para seguir este viaje, se puede consultar la siguiente guía:

* Una carpeta con las tres bases de datos de las cuales han partido todos los análisis que se han realizado a lo largo de este viaje.
* Una carpeta llamada ‘Análisis de prueba’ que contiene los *excels* con los primeros análisis o las primeras tentativas que posteriormente se descartaron, como los movimientos entre países en vías de desarrollo y los países desarrollados, el recuento de los movimientos realizados por cada investigador y un primer documento que analiza el perfil de los individuos con doctorado.
* Una carpeta llamada ‘Análisis finales’ alberga todas las operaciones que han dado lugar al proyecto final. Estas operaciones son las que muestran cuánto talento, es decir, cuántos científicos salen y reciben los países. Otro *excel* con los movimientos de la gente con doctorado, los movimientos que tienen que ver con España, cuántos regresan al país de origen, es decir, el país donde se han titulado por primera vez, y la frecuencia de cada ruta efectuada por todos los individuos en porcentaje. Y en la sub carpeta llamada ‘Empleo y formación’ que presenta el análisis de los movimientos según si se realizan para estudiar o para trabajar.
* Una carpeta con el código de python, que ha dado lugar a la base de datos de todos los viajes que los individuos han realizado desde que Orcid tiene registro, desde 1913 hasta 2016. Se adjunta un documento donde se explica el funcionamiento del código.
* Una carpeta que contiene la tabla donde se compara la base de datos de Orcid trabajada, con otros estudios que también profundizan en la migración de los científicos.

TEXTO PARA LA WEB:

¡Orcid es una mina de datos! Los 3 millones de usuarios que almacena parecen ser una buena muestra para revelar el comportamiento migratorio de los científicos. Nada más lejos de la realidad. El optimismo inicial, aunque no desapareció, fue mutando.

Antes de ponerse a escarbar, se plantearon varias hipótesis que se esperaban confirmar con el análisis de datos. ¡Caso error! Cayeron una detrás de otra, bien porque los datos decían lo contrario o por falta de tiempo para limpiarlos. Sin embargo, esta es la gracia de trabajar con una hipótesis. El resultado puede ser favorable o no, e incluso dar resultados no esperados. Aquí van algunos de ellos:

EXPLORACIONES SIN SALIDA:

* Fuga de cerebros en España. No hay dato con la nacionalidad de los investigadores

- Analizar el movimiento por el género. No hay dato con el sexo.

- Migración por especialidades. La especialidad de cada científico es un dato que lo aportan voluntariamente y por ello, no es estandarizado. Es un dato muy complicado de limpiar que supone mucho tiempo.

- Las universidades y centros de investigación españoles, ¿pierden o ganan talento? Dato voluntario y difícil de limpiar.

- Ver movimiento entre países en vías de desarrollo y los países desarrollados. Descartado porque la muestra no era representativa. No se mostraban ninguna tendencia clara. El único punto de cambio se marca en 2016, pero no es representativo porque tenemos constancia de que en ese año hay menos datos registrados.

- Qué investigador ha viajado más. Es difícil de determinar porque los usuarios de Orcid suelen registrar también estancias cortas como conferencias o simposios y por tanto no se identifican como un movimiento migratorio.

EXPLORACIONES FÉRTILES:

- Nuestro trabajo necesariamente debe ser una descripción de la base de datos y descartar cualquier intención de probar hechos puesto que es incompleta y sesgada. No están todos los países representados y predominan algunos porque la inscripción de Orcid en eso países es obligatoria.

- Flujos predominantes: China – USA y España – países de habla hispana (colonias)

TOOLS:

- Python 3

- Open Refine

- Tableau

- Excel

- Hojas de Cálculo de Google

- Google Drive

- Raw

- Carto DB

- Ilustrator

- Quadrigram

- Gephi

De primeras, queríamos probar la fuga de cerebros en España. Esta fue la primera hipótesis en caer ya que no disponemos de la nacionalidad de los investigadores registrados en Orcid.

Otra información que inicialmente nos propusimos averiguar era el sexo de cada uno de los científicos. La base de datos de Orcid tampoco proporciona este dato, por ello quisimos averiguarlo comparando los nombres de cada científico con la lista de los nombres más frecuentes del INE. Este propósito se vino abajo porque finalmente consideramos que era una asignación que podría dar muchos errores.

También se propuso distinguir los movimientos migratorios en función de la especialidad de los científicos, para detectar posibles tendencias. Y también ver qué universidades o centros de investigación ganaban o perdían talento. Pero ambas propuestas se descartaron porque son campos que se rellenan aleatoriamente, de modo que son datos difíciles de limpiar y analizar. Requiere tiempo.

Dicho esto, conseguimos extraer el Excel con todas las rutas efectuadas desde el primer año registrado, 1913, hasta 2016. Este nos ha permitido jugar con los datos, analizarlos y ver los polos que más atraen el movimiento de los científicos.

Uno de los análisis que se intentó ver fue el movimiento entre países en desarrollo y desarrollados, pero lo descartamos porque la muestra no era representativa. No se mostraban ninguna tendencia clara. El único punto de cambio se marca en 2016, pero no es representativo porque tenemos constancia de que en ese año hay menos datos registrados.

Dicho esto, se han utilizado varias herramientas de visualización para analizar datos y para mostrarlos. Algunas han sido válidas y las mantenemos pero otras las hemos descartados por no sernos del todo útil.

Para empezar, intentamos analizar los datos desde las hojas de cálculo de Google, porque de este modo podíamos trabajar todos con la misma hoja. Pero Google no soportaba la cantidad de datos y tuvimos que recurrir a Excel.

Provamos Tableau por su capacidad de análisis y visualización a la vez. Pero por cuestiones logísticas lo descartamos como pieza visual final porque es probable que una vez subido a la web, no encaje con las dimensiones que necesitamos.

También intentamos analizar los datos con Open Refine. Pero había ciertas anomalías de la base de datos que no tenía en cuenta y daba resultados erróneos.

Como herramientas puramente de visualización, hemos probado varias para representar los flujos migratorios. Gephi ha sido la herramienta por la que empezamos, pero nos dimos cuenta que no podíamos representar todas las rutas a la vez porque el resultado es una maraña de líneas. Además, es una herramienta que ninguno del equipo conocía previamente y ha costado horas de aprendizaje a la infografista del grupo. Aún así, no lo hemos descartados. Mediante Gephi representamos las rutas que consideramos más destacadas e importantes de reseñar, que son EEUU y China y las rutas entre España y los países sudamericanos.

Para representar lo mismo, también hemos probado Raw, pero estéticamente no nos acababa de gustar. Y Carto, pero no conseguíamos la sensación de movimiento. Aunque nos daba los Tooltips dinámicos que no conseguimos con Gephi.

Los gráficos estáticos más pequeños se han hecho con Ilustrator porque no creemos que sea necesario hacerlos dinámicos.