

# Prácticas de Datos abiertos y Visualización

Fecha de entrega: 7 de marzo

El objetivo de la práctica es desarrollar un *storytelling* o un *dashboard* basado en la información extraída de alguna fuente de datos abiertos de tu elección. Más concretamente, debes:

- **Seleccionar una fuente de datos “abiertos”:** bien un fichero de información estática, bien una API dinámica a la que poder lanzar consultas. Se valorará positivamente la dificultad en la obtención y tratamiento de los datos. Por ejemplo: a la hora de obtener los datos no puntúa igual recurrir a una API o Web Scraping que limitarse a cargar un simple csv.
- **Desarrollar el proceso de carga desde la fuente seleccionada:** se deberá cargar la información desde la fuente seleccionada y hacer los tratamientos necesarios sobre la misma de forma que se disponga de un set de datos “limpio” y correctamente preparado. También se valorará positivamente mezclar y unir datos de más de una fuente de datos.
- **Contar una historia (*storytelling*) o desarrollar un panel de control (*dashboard*) a partir de la información de la fuente de datos seleccionada basada en visualización:** utilizando cualquiera de las herramientas y técnicas de visualización revisadas durante las clases teóricas, se deberá realizar una exploración de los datos preparados de forma que se puedan transmitir algunas conclusiones o *insights* derivados de dicho análisis.

El criterio de evaluación es el siguiente:

- **Carga y tratamiento de información:** 1,5 puntos.
- **Visualización estática:** 3,5 puntos.
- **Visualización dinámica:** 3,5 puntos.
- **Dificultad, interés y coherencia del análisis realizado:** 1,5 puntos.

Es importante que el ejercicio sea fácilmente reproducible, es decir, debe de ejecutarse sin necesidad de hacer descargas de datos, instalaciones de paquetes o librerías, etc. En el caso de ser necesarias instalaciones incorporar un fichero explicando que es necesario instalar antes de ejecutar la práctica. Si el tamaño del fichero de la entrega imposibilita subir todo al aula virtual recuerda eliminar los ficheros de datos e incluir una referencia a algún sistema para compartir archivos desde donde poder descargarlo. Sólo se considerarán visualizaciones estáticas, a efectos de evaluación, aquellas realizadas con los paquetes vistos en las sesiones de visualización estática (ggplot2 en R, matplotlib, seaborn o plotnine en Python), los gráficos hechos con, por ejemplo, plotly no cuentan como visualización estática.

## Contenido de la entrega

La entrega deberá contener:

- Documento en el que se detalle:
  - Proceso llevado a cabo para la extracción, tratamiento y preparación de los datos obtenidos de la fuente de datos abiertos seleccionada.
  - Justificación de las visualizaciones estáticas y dinámicas realizadas como parte de la práctica.
  - Explicación de los *insights*/conclusiones alcanzados y reflejados en las visualizaciones realizadas.
- En el caso de entregar *storytelling* se recomienda entregarlo en un R markdown o Jupyter notebook el cual ya está renderizado.
- En el caso de entregar *dashboard* se recomienda subirlo a <https://www.shinyapps.io/> donde puedes colgar tus aplicaciones de shiny. En el caso de no realizar la entrega mediante un sistema de alojamiento tipo shinyapps, y de requerir que el profesor la ejecute será obligatorio entregar instrucciones para su ejecución (paquetes a instalar, consideraciones previas a la ejecución, etc.).
- **Se penalizará aquellas prácticas que no consten del documento explicativo, que den errores de ejecución o que no tengan instrucciones claras para su puesta en funcionamiento. Si el volumen de errores es tan elevado que imposibilita su corrección la práctica no será evaluada.**
- Ficheros de datos necesarios para poder reproducir las visualizaciones.
- Código necesario para poder reproducir las visualizaciones.
  - Posibilidades para visualizaciones estáticas:
    - Scripts de R básico.
    - R markdown.
    - Scripts de Python básico.
    - Jupyter Notebook.
    - ...
    - R: ggplot2
    - Python: matplotlib, seaborn y plotnine
  - Posibilidades para visualizaciones dinámicas:
    - Código de aplicación Shiny.
    - R markdown.
    - Jupyter Notebook.
    - ...
    - R: Plotly, Highcharter, dygraphs, scatterd3 y Shiny
    - Python: Plotly y Bokeh

Se deberá entregar un ZIP que contenga todos los ficheros asociados a la entrega y que tenga por nombre el nombre y apellidos del alumno (ej. jose\_miguel\_morella.zip)

# Recursos para la práctica

## Ejemplos de *storytelling*

- [What's Really Warming The World?](#)
- [Find a Living Place with the Help of Data Visualization Techniques](#)

## Ejemplos de *dashboard*

- [COVID-19 ASTURIAS](#)

## Fuentes de datos “abiertos”

Aunque la elección de la fuente de datos “abiertos” a utilizar queda completamente a vuestro criterio, a continuación podéis encontrar algunos ejemplos o repositorios de los que podrías obtener la información.

## Repositorios de datos “abiertos” estáticos (ficheros)

- [UCI Machine Learning Repository](#)
- [Datahub](#)
- [Kaggle](#)
- [Portal de datos abiertos de la Ciudad de Madrid](#)
- [Portal de datos abiertos de España](#)
- [Portal de datos abiertos de Zaragoza](#)
- [Portal de datos abiertos del Instituto Nacional de Estadística \(INE\)](#)
- [Portal de datos abiertos de Estados Unidos](#)
- [Portal de datos abiertos del Banco Mundial](#)

## Repositorios de datos “abiertos” dinámicos (APIs)

- [Programmable Web](#)