Actividad. Representación gráfica de datos y análisis de resultados

Objetivos

Por medio de esta actividad, aplicarás lo aprendido a lo largo del tema de la asignatura a la visualización de datos biológicos y al análisis de los resultados obtenidos tras la generación de las gráficas que deberás realizar. El objetivo fundamental de esta actividad es, por tanto, que consigas analizar un *dataset* mediante la representación gráfica de sus datos y responder a las diferentes cuestiones que se te plantean en el siguiente apartado de este documento.

Pautas de elaboración

Esta actividad consistirá en dos partes principales: una relativa a la visualización y análisis de los datos de un *dataset* relacionado con la incidencia registrada para la **COVID-19**.

1. ***dataset* para la COVID-19**:para la realización de esta parte de la actividad, deberás descargar el *dataset* de interés, el cual se corresponde con los casos diarios registrados de COVID-19, así como fallecimientos, desde el inicio de la pandemia hasta el 14 de diciembre de 2020, desde [este enlace](https://opendata.ecdc.europa.eu/covid19/casedistribution/csv/data.csv) o desde el aula virtual. Una vez descargado, deberás importarlo en RStudio de la siguiente manera:

data <- read.csv("data.csv", na.strings = "", fileEncoding = "UTF-8-BOM") %>% mutate(date=as.Date(as.character(dateRep),"%d/%m/%Y"), month=as.factor(month), year=as.factor(year)).

Tras importarlo, deberás responder a las siguientes cuestiones, teniendo en cuenta que todos los gráficos generados mediante ***ggplot2*** deben tener una **leyenda** (si aplica), un **título**, un **subtítulo**, **títulos de eje** y un **tema**:

* + ¿Cuál es la **distribución** global de los casos de COVID-19? Para responder esta cuestión, apóyate en un **histograma** para los casos.
  + ¿Cuántas **olas de casos** quedaron registradas en este *dataset* para España y cuál fue su duración aproximada? Para responder esta cuestión, apóyate en un **diagrama de líneas** en donde se representen los casos diarios frente a la variable de fechas recogidas.
  + De entre España, Francia, Italia, Alemania y el Reino Unido, ¿cuál de estos países registró un **mayor número de casos** durante la **segunda ola** posterior al verano de 2020? Para responder esta cuestión, apóyate en un **diagrama de líneas** en donde se representen los casos diarios frente a la variable de fechas recogidas. Genera este diagrama, en primer lugar, con los datos de todos los países y, luego, por separado, empleando la función facet\_wrap().
  + ¿En qué **continente** se registraron más **casos medios** de COVID-19 durante el periodo temporal recogido en el *dataset*? Para responder esta cuestión, apóyate en un **diagrama de barras**.

Extensión y formato

Para la resolución de esta actividad, deberás entregar dos archivos:

* Un único fichero **R Markdown** (*.Rmd*) con todo el código y texto Markdown que hayas generado. No existe un límite de extensión para este fichero.
* Un fichero **HTML** generado a partir de dicho archivo R Markdown. Asegúrate de que aparecen todas las figuras, todo el texto y todos los cuadros de código R que hayas introducido en el archivo *.Rmd* anterior.

**Rúbrica**

|  |  |  |  |
| --- | --- | --- | --- |
| Representación gráfica de datos y análisis de resultados | Descripción | Puntuación máxima  (puntos) | Peso  % |
| Gráficos | Construcción correcta de los gráficos tal y como se solicitan en el enunciado de la actividad | 3 | 30 % |
| Cuestiones | Resolución de las cuestiones planteadas de forma correcta | 2 | 20 % |
| Originalidad | Generación de un informe R Markdown original, empleando temas o estilos CSS, índices, etc. | 1 | 10 % |
| Orden | Exposición de lo que se pide en el enunciado de forma clara y ordenada | 1 | 10 % |
| Justificación | Justificación detallada en la respuesta a las preguntas planteadas y en el análisis de los gráficos generados | 3 | 30 % |
|  |  | **10** | **100 %** |