

Autoevaluación 2

[0010] Representación y tabulación de datos

Enunciado

Vamos a responder preguntas sobre los datos que se encuentran en el fichero `vuelos.csv` que contiene datos sobre horarios, retrasos, aviones, tiempos y distancias de vuelos que salen de Houston los primeros meses de 2011.

Es obligatorio utilizar las funciones del paquete `dplyr` y recomendable utilizar pipes `%>%`.

1. Carga la librería `dplyr`
2. Descarga el fichero de datos `vuelos.csv` situado en <http://gauss.inf.um.es/datos/> en local o localiza la url donde se encuentra.

Lee el fichero (desde local o desde la url) usando correctamente los argumentos de la función `read.table()`: `header`, `sep`, `dec`. Guarda los datos en una variable llamada `vuelos`.

3. Selecciona los vuelos con destino `SFO` u `OAK` utilizando las funciones del paquete `dplyr`. ¿Con cuántos vuelos nos quedamos?

Hay XX vuelos con esos destinos

4. Selecciona los vuelos que se han retrasado más de una hora. ¿Cuál es el destino que más se retrasa en proporción al número de vuelos?

El destino que más se retrasa es XX

5. Encuentra 4 maneras diferentes de utilizar la función `select` par seleccionar las variables relacionadas con los retrasos (`delay`)
6. Agrupa los vuelos por fecha y calcula: media, mediana y cuartil 75 de los retrasos en los vuelos por hora.
7. Utilizando pipes calcula la media de retraso en los vuelos por día y hora, la cantidad de vuelos por día y hora y luego muestra solo los casos para los cuales haya más de 10.