

Entrega un documento PDF en el que se incluya, para cada pregunta:

1. El enunciado de la pregunta.
2. Los comandos de R utilizados para responder a la pregunta.
3. Los resultados mostrados por R como respuesta a cada comando.
4. Cuando se pida, el análisis de los resultados obtenidos.
5. Copia, en su caso, de las gráficas generadas por R.

Valoración: 50% de la nota final.

Los datos sobre los que se va a trabajar en este proyecto proceden de la página **Berkeley Earth**

<http://berkeleyearth.org/data/>, sección *Quality Controlled*, subsección *TAVG*

que reúne registros de temperaturas en distintos puntos del planeta durante amplios periodos de tiempo.

Archivos de datos para el proyecto (en formato CSV):

data_series.txt	68493 registros con formato: Station_ID, Series_Number, Date, Temperature
data_clustering.txt	5917 registros con formato: Station_ID, Year, T01, T02, ... T11, T12
Whole_years.txt	5235 registros de años completos con formato: Station_ID, Whole_year
Ciudades.docx	documento informativo de las 72 ciudades (Station_ID) de los archivos de datos

La variable *Temperature* es la media mensual del "año.mes" *Date*, y las variables *T01*, *T02*, ... *T11*, *T12* son las medias de los doce meses del año *Year*.

1. Clustering

- (a) Suprimir de la base de datos (data_clustering.txt) los registros en los que haya algún dato indefinido ("NA").
- (b) Para los años 2000 y 2017:
 - i) Crear para cada uno un Frame con las variables Station_ID, T01, T02, ... T11 y T12 incluyendo los registros en los que Year coincide con cada uno de los dos años indicados.
 - ii) Hacer un clustering de las ciudades (Station_ID) usando las variables "T01" a "T12" (usar el método elbow para determinar el valor de k).
- (c) Describir las características de los clústeres obtenidos correspondientes al año 2000.
- (d) Describir los cambios que se observan al pasar del clustering del año 2000 al del año 2017.