

# Examen-1-Parcial-EnunciatsMODEL1...



alucero



Visualització de Dades



3º Grado en Ingeniería de Datos



Escuela de Ingeniería  
Universidad Autónoma de Barcelona

antes



**Descarga sin publi  
con 1 coin**



Después

**WUOLAH**



Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato  
→ Planes pro: más coins

pierdo espacio



Necesito concentración

ali ali ooh  
esto con 1 coin me  
lo quito yo...

WUOLAH

Visualització de dades (Enginyeria de Dades – EE - UAB)  
Examen Primer Parcial – 11 Abril 2023  
ENUNCIATS MODEL 1

Nom i Cognom: \_\_\_\_\_

NIU: \_\_\_\_\_ Grup de Matrícula: \_\_\_\_\_

**PARTE 1 (5 pt)**

Dataset: *PIB\_population\_lifeExpectancy.csv*

1.1. (0.5 pt) Abre el fichero. ¿Qué tipo de atributo son: Country, Region, RegionCode, LifeExpectancy\_2018, PIBperCapita\_2018, y Population\_2018?

RESPOSTA:

1.2. Haz un bubble chart con las variables LifeExpectancy\_2018, PIBperCapita\_2018, y Population\_2018 (puedes controlar el tamaño de los puntos con `aes(..., size)`). Cualquiera de las 3 variables se puede codificar en cualquiera de los canales disponibles: X, Y, y tamaño.

A) (1 pt) ¿Cuál es el mapeo óptimo de esos 3 canales para explorar la correlación entre variables? Justifica brevemente tu respuesta.

RESPOSTA:

B) (1 pt) Sube el código y la gráfica.

RESPOSTA:

1.3. Mejora la gráfica añadiendo transparencia, color, bordes, etc., y las leyendas necesarias para que la gráfica sea comprensible.

A) (1 pt) Haz una gráfica en la que el color represente la variable *Region*. Explica brevemente la ventaja de usar el color para esta variable en relación a las demás, y el tipo de paleta que utilizas.

RESPOSTA:

B) (1 pt) Haz una gráfica en la que el color sea una de las 3 variables que ya has usado en X, Y, o tamaño. Explica brevemente la ventaja de usar el color para esa variable en relación a las demás, y el tipo de paleta que utilizas.

RESPOSTA:

C) **(0.5 pt)** Sube el código y la gráfica de una de las dos opciones, A o B.

RESPOSTA:

## PARTE 2 (5 pt)

Dataframe: iris de R



NOTA: En los ejercicios de esta parte, hacer uso de las *pipes*

**2.1. (0.75 pt) NOTA:** Para este ejercicio no se piden hacer las visualizaciones, solo responder las preguntas.

a) ¿Qué tipo de variables tiene este dataframe? **(0.25 pt)**

b) ¿Qué gráfico de los siguientes podemos hacer que nos permita comparar las variables, conectándolas con líneas que permitan visualizar patrones y relaciones entre ellas? ¿Por qué? **(0.5 pt)**

b.1. Un *parallel coordinate plot* que nos permita comparar las cinco variables de este dataframe juntas.

b.2. Un *parallel coordinate plot* que nos permita comparar algunas variables de este dataframe juntas. ¿Qué variables?

b.3. Ninguna de las anteriores. De hecho, no haríamos un *parallel coordinate plot* con este dataframe, haríamos un *parallel set plot*.

## 2.2. (2.25 pt)

a) Crea dos nuevas variables “Petal.shape” y “Sepal.shape” que corresponda en cada caso a la ratio entre su anchura y su longitud. **(0.25 pt)**

b) Compara la distribución de las nuevas variables en un gráfico multi-panel **(1.25 pt)**.  
Nota: En el caso de no saber hacer el multi-panel y realizar dos figuras distintas se contará solo un máximo de 0.75 pt.

c) Ahora compara en una sola visualización la distribución de la nueva variable Petal.shape para cada especie **(0.75 pt)**

**2.3. (1.5 pt)** Haz un gráfico que permita ver claramente cuáles de las métricas del dataframe tendríamos en cuenta para tener una máxima correlación. ¿Y para tener una mínima correlación? Razona tu respuesta

RESPOSTA:

**2.4. (0.5 pt)** Data massage

**a)** Calcula el valor mínimo de Petal.Length por Species (0.25 pt)

**b)** Crea un dataframe que para todas las flores con una longitud de sépalo menor que 5 cm, contenga como columnas la anchura del sépalo, la longitud del sépalo y las especies. ¿Cuántas observaciones tiene el nuevo dataframe? (0.25 pt)

RESPOSTA: