

2025  
10-13 JULY



MAD  
COOL  
FESTIVAL



## Visualització de dades (Enginyeria de Dades – EE - UAB)

Examen Recuperació Segon Parcial – 3 Juliol 2023

### ENUNCIAT

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

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

Només es permet l'ús d'internet per l'accés al campus virtual en el moment de descarregar el full d'enunciats y d'entregar l'examen.

Sólo se permite el uso de internet para el acceso al campus virtual en el momento de descargar la hoja de enunciados y de entregar el examen.

### PART 1 (4 pt)

*Dataset: US\_Births\_Names\_1910-1950.csv*

*Agafarem aquest dataset recull els noms dels nadons que s'han donat en cada estat i any en Estats Units des del 1910 a 1950. Per a cada registre es guarda el nom i nombre de nadons.*

*Cada registre conté informació del nom donat al nadó, estat, any i sexe. Conté les variables:*

1. State ---> 2-alphanumeric digit state code
2. Sex ---> M=male or F=female
3. Year ---> 4-digit year of birth
4. Name ---> Given name of a persona t birth
5. Count ---> Number of occurrences of the name

**1.1 (2 pts)** Mostra el codi (en ggplotly) i la gràfica de línies de l'evolució temporal en anys del nombre dels noms femenins majors de 4.500 per any en tot els Estats Units (gràfica 1) i fixa't amb els tres noms femenins més posats l'any 1945. Digues quin són aquests noms per ordre decreixent i mostra el codi (en ggplotly o shiny+plotly) i la gràfica de línies de l'evolució temporal d'aquests tres noms (gràfica 2). Per a cada gràfica, explica si fas Data Massaging i quines operacions fas i posa també el codi.

1.2 (0,5 pts) Compara els tres noms femenins de la gràfica 2 segons els següents criteris:

a) (0,2 pts) Anys en que assoleixen el màxim i el mínim.

RESPOSTA:

NOM	MÍNIM: Any - Nombre	MÀXIM: Any - Nombre

b) (0,1 pts) De la taula anterior i les gràfiques 1 i 2, trobes algunes singularitats en les gràfiques?. Sabries dir el per què?.

c) (0,2 pts) Les tres gràfiques de noms tenen creuaments. De cada creuament, digues quin parell de noms creuen i l'interval dels dos anys en que es produeix.

RESPOSTA:

NOMS	Any Anterior	Any Posterior

1.3. (1,5 pts) Mostra el codi i adjunta el fitxer AVI o GIF del Ranking de Barres Animades (*Animated Bar Race Ranking*) sobre els 10 noms femenins més posats entre els anys del 1925 al 1950. Explica si fas Data Massaging i quines operacions fas i posa també el codi.

RESPOSTA:

## PART 2 (4 pt)

*Dataset: US\_Births\_Names\_1910\_1950.csv*

*Continuarem utilitzant el mateix dataset que en la part 1 de l'examen.*

**2.1 (2 pt)** Fes un *bubble map* per representar quants homes nascuts al 1940 van rebre el nom de "John" als estats de: Carolina del Nort "NC", California "CA", Wyoming "WY", Texas "TX", Alabama "AB" i New York "NY".

- a) Indica el datamassage necessari (0,5 pts) .
- b) Fes un *bubble map* dels Estats Units , que doni informació sobre el nombre de rutes per cada ciutat. Pots usar el mapa de *states* de USA que vam usar al seminari. Al dibuixar-lo treu la graella del fons (1,5 pts).

*Indicacions:*

Primer necessites crear una *tibble* amb les latituds i longituds d'aquests estats, les quals et donem:

	State	lat	lon
1	CA	36.77826	-119.41790
2	NC	35.20520	-78.82523
3	WY	42.98329	-107.54867
4	TX	30.77917	-97.80889
5	AL	32.31823	-86.90230
6	NY	39.76595	-75.00323

~ State, ~lat, ~lon,

"CA", 36.77826, -119.4179,

"NC", 35.20520, -78.82523,

"WY", 42.983286, -107.548667,

"TX", 30.779167, -97.808891,

"AL", 32.318230, -86.902298,

"NY", 39.76595, -75.00323

Aquesta *tibble* l'hauràs d'ajuntar amb un *dataframe* construït contenint el "Count" dels homes recent nascuts que van rebre el nom de "John" l'any 1940 en aquests estats.

*Nota: Fixa't que també hi ha dones en alguns estats amb el nom de John.*

Per ajuntar ambdós *dataframes* pots usar la comanda de R:

```
> inner_join(nom_tibble, nom_dataframe, by="nom_variable_compartida")
```

**RESPOSTA:**

a) *Data massaging*

b) *Bubble Map*

2025  
10-13 JULY



MAD  
COOL  
FESTIVAL



2.2 (2 pt) Mostra dos mapes d'arbre (*treemaps*). Un t'ha de permetre saber quins noms de cada sexe eren a la llista dels nou nascuts entre els top 30 l'any 1910. I l'altre t'ha d'ajudar a respondre quins seguien (o no) sent també entre els top 30 l'any 1950

- a) Fes el datamassage necessari (0.25 pt). Pots usar `top_n(30,Variable)`.
- b) Argumenta com és un *treemap* en general, quin tipus de dades tracta normalment i els passos que has de fer per construir les dues visualitzacions d'aquest exercici (és a dir quina variable utilitzes per l'àrea de les graelles, variables d'agrupació, etc). Posa llegenda i títol. I inclou el codi si el fas en R. Mostra els dos mapes d'arbre (1.25 pt).
- c) Argumenta quina/es de les següents comandes no usaries en un codi en R per posar la paleta de color i inclou la que més t'agradi als teus *treemaps* (o una que et sembli bé en Tableau argumentant per què) (0.25 pt) :
- `scale_fill_continuous ()`
  - `scale_fill_discrete ()`
  - `scale_fill_manual(values=c('#87C55F','#9EB9F3'))`
- d) Diques un nom de cada sexe que estigués en la llista dels top 30 l'any 1910 i posteriorment també en el 1950? I un de cada sexe també, que estigués en el top 30 l'any 1910 però desaparegués de la llista dels top 30 l'any 1950 (0.25 pt).

RESPOSTA: