

2025  
10-13 JULY



## 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 person at 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 datassage 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



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 gralles, 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) Digues 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:**

2025  
10-13 JULY



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

Examen Segon Parcial - 21 Juny 2021

MODEL A

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.

### PARTE 1 (2.5 pt)

1.1. (0.5 pt) Marca el tipo de datos que es mejor codificar en cada uno de los componentes del color:

- |    |            |             |               |
|----|------------|-------------|---------------|
| a) | Tono       | categóricos | cuantitativos |
| b) | Saturación | categóricos | cuantitativos |
| c) | Luminancia | categóricos | cuantitativos |

RESPOSTA:

1.2. (0.5 pt) ¿Cuál es el número máximo recomendado de colores distintos en una paleta categórica?

RESPOSTA:



RESPOSTA:

1.4. (1 pt) ¿Qué escala de color usarías para representar la temperatura en un mapa, con valores min=-20°C y max=30°C? Adjunta una imagen del esquema de color y justifica tu respuesta.

**RESPOSTA:**

## PART 2 (3.5 pt)

*Dataset: beers.csv*

Aquest dataset té 7 atributs per 44 cerveses de Minnesota i el farem servir en l'exercici 2.1. Els atributs són: Nom de la cerveseria, cervesa, descripció, estil, alcohol en volum (ABV), unitats internacionals d'amargor (IBU), puntuació que han rebut.

*Dataset: statsNBA2008.csv*

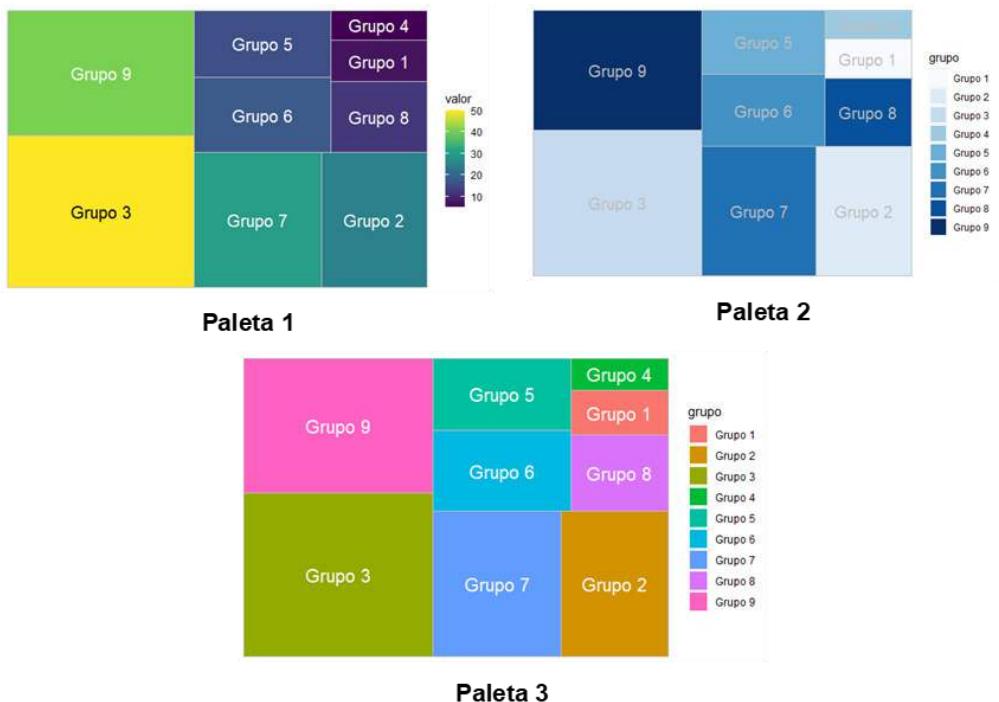
Aquest dataset té 21 atributs de 50 jugadors de la NBA. Recull les estadístiques de la NBA del 2008. I estudiarem la relació entre alguns dels atributs en l'exercici 2.2

2.1 (2.5 pt) És estiu, esteu a Minnesota i voleu anar a prendre una cervesa. Utilitzant el dataset *beers.csv*.

- a) Feu un mapa d'arbre (*treemap*) que us permeti contestar les preguntes de l'apartat b. Podeu fer-lo en R o en qualsevol altre llenguatge, utilitzant les llibreries que us semblin convenient. Ara bé, heu d'explicar-ho bé: tot narrant com és un *treemap* en general - redacteu els passos que heu de fer per construir la visualització triada. (1.5 pt)
- b) Un cop tingueu la visualització, contesteu: Esteu amb una persona que li encanta la cervesa però que no li agrada la cervesa Ale, a quina cerveseria no anirieu? Sabeu que li agrada la cervesa més amarga de totes, a quina de les cerveseries en venen més d'aquestes? (0.25 pt)
- c) Expliqueu en quin cas són òptims el mosaic, el treemap i un paral·lel set (sense necessitat de fer cap visualització) i quines diferències principals hi ha entre ells. (0.25 pt)
- d) Quin tipus d'escala de color, de les de la pàgina següent, seria la més òptima i perquè? (0.5 pt)

- Paleta 1
- Paleta 2
- Paleta 3

Raona la resposta.



**RESPOSTA:**

**2.2. (1 pt)** Voleu estudiar algunes relacions entre el tipus de jugades de basquet i extreure conclusions de si això porta als equips a guanyar o perdre. Per això teniu el dataset *statsNBA2008.csv*. Estudieu la relació entre les variables següents, tot fent una visualització i mostrant només la informació d'interès. Podeu fer-ho en qualsevol llenguatge, però expliqueu perquè heu escollit aquesta visualització, els passos que heu seguit per fer-la i extrieu alguna/es conclusió/ns del vostre gràfic.

- PTS (percentatge de punts de l'equip, punts per joc)
- FGM (cistelles de camp realitzats, percentatge de cistelles de camp)
- FGA (tirs de camp intentats, número d'intencions de cistelles de camp)
- DRB (rebots defensius)
- ORB (rebots ofensius)
- TRB (rebots totals)

**RESPOSTA:**

2025  
10-13 JULY



### PART 3 (4 pt)

Dataset: 25\_noms\_padro\_any\_sexe\_1996\_2019.csv

Agafarem el dataset de noms del padró de naixements de Barcelona. Utilitzeu les llibreries (plotly, gganimate, shiny, etc.) que creieu convenient i dibuixeu els gràfiques que us facin falta.

RESPOSTA:

3.1 (1 pt) Mostra el codi i la gràfica de l'evolució temporal dels noms femenins (gràfica 1) i fixa't amb els tres noms femenins més posats l'any 2015. Digues quin són aquests noms per ordre decreixent i mostra el codi i la gràfica de l'evolució temporal d'aquests tres noms (gràfica 2).

RESPOSTA:

3.2. (1 pt) Compara els tres noms en la gràfica 2 segons els següents criteris:

- Anys en que assoleixen el màxim i el mínim.

RESPOSTA:

NOM	MÍNIM: Any - Nombre	MÀXIM: Any - Nombre

- b) Dos dels noms tenen el mateix nombre de nadons. Diga'm el nombre i l'any en que coincideixen.

**RESPOSTA:**

- c) Dos dels noms tenen tres creuaments. Digues l'interval dels dos anys en que es produeix cada creuament.

**RESPOSTA:**

NOMS	Any Anterior	Any Posterior

4 (1 pt) Classifica les següents preguntes o mesures que s'utilitzen en tests d'Usabilitat o d'Experiència d'Usuari (UX) en les següents dues categories:

- a) Instrumentals / Usabilitat / Pragmàtiques  
 b) No instrumentals / Hedònics / Emocionals

**PREGUNTES:**

1. *I thought the System was wasy to use*
2. *Non-inclusive 1 2 3 4 5 6 7 Inclusive*
3. *I found the various functions in this System very well integrated*
4. *Complicated 1 2 3 4 5 6 7 Simple*
5. *The product is creatively designed*
6. *The product is stylish: Strongly Disagree 2 3 4 5 6 Strongly Agree*
7. *It was easy to find the information I needed*
8. *This System has all the functions and capabilities I expect it to have*
9. *The design looks attractive*
10. *Conservative 1 2 3 4 5 6 7 Innovative*

**RESPOSTA:**

PREGUNTA	CATEGORIA
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

**5 (1 pt)** Defineix 4 dels següents conceptes en animació, interactivitat, usabilitat i Experiència d'Usuari en Visualització de Dades:

- Event
- Usabilitat
- *plotly\_hover*
- *Data linking*
- *Slider*
- *Data Usability*
- *Toggle*
- *Data Reliability*



la app para encontrar curro este verano  
con más de 5000 ofertas de empleo.



descarga la ap

**Visualització de dades (Enginyeria de Dades - EE - UAB)**  
**Examen Segon Parcial – 3 Juny 2022**  
**RESPOSTES MODEL A**

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.

**PARTE 1 (2,5 pt)**

*Dataset: GDP-LifeExp-Population.csv*

\* Lee los tres enunciados de esta parte antes de empezar \*

1.1. (0,5 pt) Filtra las observaciones que tengan valores NAN o null.

Identifica outliers (valores extremos que distorsionen la muestra) en la variable *Population* y filtra las observaciones que no sean comparables al resto.

**RESPOSTA:**

Nulls: Andorra, Eritrea, Gibraltar, N. Korea, Gibraltar, F. Polynesia, S. Sudan, Venezuela  
Outliers: World, OCDE members

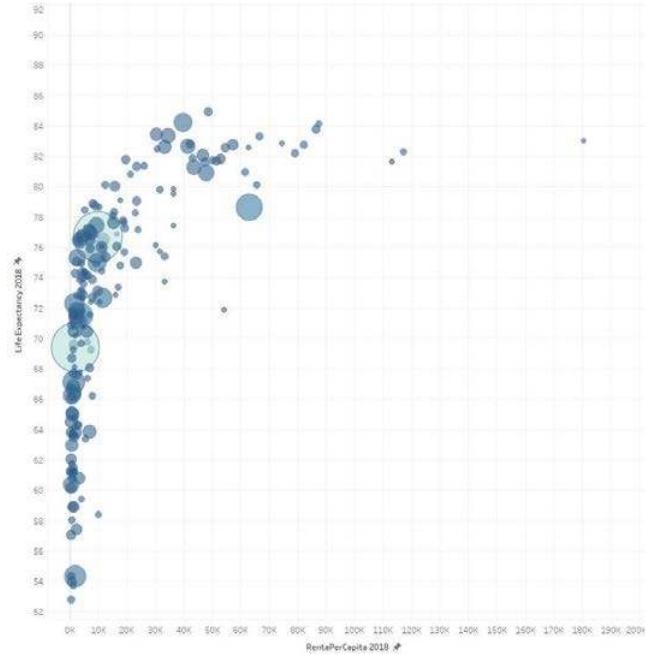
1.2. (1 pt) Haz un **bubble chart** con las variables *Life expectancy 2018*, *PIB per capita USD 2018* y *Population 2018*, donde *Population* sea el tamaño de los círculos.

Los círculos deben estar coloreados por la variable de tu elección que creas que más aporta a la visualización.

Añade las etiquetas, leyendas, y transparencias necesarias para que la gráfica sea legible y comprensible.



**RESPOSTA:**



**1.3. (0,5 pt)** ¿Por qué un *bubblechart* es una visualización apropiada en este caso? Razona brevemente según el tipo y el número de atributos.

**RESPOSTA:**

Tenemos tres atributos cuantitativos y un número alto de observaciones (200). Un bubble chart sirve para graficar la relación entre tres atributos cuantitativos, dos en la posición X e Y, y un tercero en el tamaño de los círculos. Además es un tipo de gráfica que escala bien a cientos de observaciones, a diferencia de otras, como las gráficas de barras.

**1.4. (0,5 pt)** ¿Qué variable usas para colorear los datos y qué escala de color utilizas? Razona brevemente porqué la variable y porqué la escala.

**RESPOSTA:**

Cualquiera de las 3 cuantitativas serviría, aunque la más apropiada será *Population* para reforzar el canal más débil, que es el tamaño de los círculos. La opción incorrecta sería usar el país.

## PART 2 (3,5 pt)

**Dataset:** starwars de R per l'exercici 2.1

**Dataset:** statsNBA2008.csv . Aquest dataset té 21 atributs de 50 jugadors de la NBA. Recull les estadístiques de la NBA del 2008. I estudiarem les components principals d'alguns dels atributs en l'exercici 2.2.

**2.1 (2,5 pt)** Volem representar els pesos dels personatges de starwars (dades incloses en R) en un mapa d'arbre (*treemap*) que us permeti contestar les preguntes de l'apartat d. El mapa d'arbre (*treemap*) podeu fer-lo en R o en qualsevol altre llenguatge, utilitzant les llibreries que us semblin convenientes.

- a) Feu el següent data massage: Descarteu les files del dataframe on la variable massa ('mass') o la variable especie ('species') continguin un valor Nan i reemplaceu els Nans de la columna gènere per 'none'. (0,5 pt)
- b) Feu un mapa d'arbre. Ara bé, heu d'explicar-ho bé: tot narrant com és un treemap en general - redacteu els passos que heu de fer per construir la visualització triada. Poseu llegenda i títol i mostreu el gràfic (1,25 pt)
- c) Canvieu la paleta de color. Si treballau en R podeu posar manualment la paleta de color `c('#DCB0F2', '#87C55F', '#9EB9F3')` o posar una altra paleta adient que us agradi. En tots els casos, si treballau amb R, en un altre llenguatge o amb Tableau, presenteu en l'apartat anterior el gràfic que us ha sortit amb el color per defecte i aquí amb una paleta diferent (0,25 pt)
- d) Un cop tingueu la visualització, contesteu: Quin és el personatge amb més massa? De quina espècie i gènere és? Digueu el nom dels dos robots ('droids') amb menys pes, quin gènere són? (0,25 pt)
- e) Expliqueu en quin cas són òptims el mosaic, el treemap i un paral·lel set (sense necessitat de fer cap visualització) i quines diferències principals hi ha entre ells. (0,25 pt)

#### **RESPOSTA:**

a)

> *Library ("tidyverse")*

```
> data = starwars %>% drop_na(mass) %>% drop_na(species) %>%  
  replace_na(list(gender = "none"))
```

b) Com vam veure a classe, un mapa d'arbre és un dibuix rectangular dividit en caselles, i cada casella representa una sola observació. Vam veure que era una bona manera de mostrar dades jeràrquiques mitjançant rectangles imbricats. I l'àrea relativa de cada casella expressava una variable contínua. També vam veure que era òptim quan hi ha com a màxim dues variables d'agrupació, per tant no en definirem més.

Per tant, una possibilitat amb aquest dataset seria:

- Per definir el color i actuar doncs com un 'grup pare' utilitzaríem la variable "gender" (Agafem aquesta com grup de color, ja que sol té 3 nivells ("masculine", "feminine" i "none" i serà fàcil entendre la llegenda, que si posem "species" que en té més).
- Com a 'subgrup' utilitzaríem la especie (species, que té més nivells que gènere).
- Com a variable que descrigui l'àrea de les caselles triarem "mass". Aquesta tria ve donada perquè ens estan preguntant sobre la massa dels personatges. I de totes les variables que ens pregunten, és l'única numèrica.
- I com a 'label/nivell, escollim el nom ("name").

Si ho fem amb R, el primer que hem de fer és carregar les llibreries necessàries. La llibreria específica aquí és treemapify (abans necessitarem tenir instal·lat el paquet com vam veure al seminari corresponent). També haurem de fer ús de geom\_treemap.

Per fer un treemap basic doncs:

```
> library ("treemapify")
```

d) El personatge amb més pes és en Jabba Desilijic Tiure, masculí i de l'espècie hut (Hutt)

Els robots amb menys massa són el R2-D2 i el R5-D4, són masculins

e) Quan s'especifiquen les proporcions d'acord amb múltiples variables d'agrupació, les representacions de mosaics, els mapes d'arbres o els conjunts paral·lels (paral·lel sets) són aproximacions de visualització útils. Ara bé, els mosaics assumeixen que tots els nivells d'una variable d'agrupament es poden combinar amb tots els nivells d'una altra variable d'agrupació, mentre que els mapes d'arbres no fan aquesta suposició.

Els mapes d'arbres com el que hem fet, funcionen bé fins i tot si les subdivisions d'un grup són completament diferents de les subdivisions d'un altre.

Finalment, els conjunts paral·lels funcionen millor que les trames de mosaic o els mapes d'arbres quan hi ha més de dues variables d'agrupació.

**2.2. (1 pt)** Voleu saber quines són les components principals de les variables numerades a sota del dataset `statsNBA2008.csv`. Podeu fer-ho en qualsevol llenguatge, però expliqueu perquè heu escollit la visualització que heu escollit, els passos que heu seguit per fer-la, mostreu-la, i extreieu alguna/es conclusió/ns del vostre gràfic.

- PTS (percentatge de punts de l'equip, punts per joc)
- FGM (cistelles de camp realitzats, percentatge de cistelles de camp)
- FGA (tirs de camp intentats, número d'intencions de cistelles de camp)
- DRB (rebots defensius)
- ORB (rebots ofensius)
- TRB (rebots totals)

#### **RESPOSTA:**

Llegim dades i carreguem llibreries:

```
> NBA=read_csv('statsNBA2008.csv')
```

Primer ens preparam el dataframe que ens interessa, vam veure al seminari 5 les tibles de R i com utilitzar-les. Una manera que es podria fer servir, seria

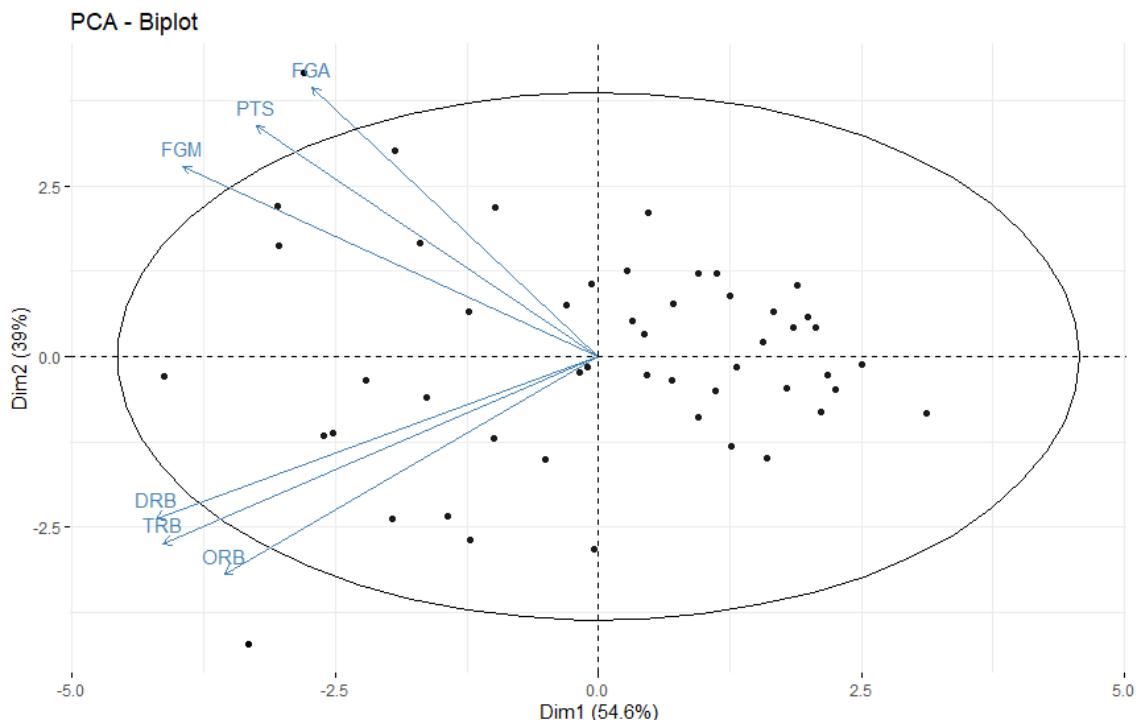
```
> NBA_tb <- as_tibble(NBA)
> NBA_interes<-NBA_tb[,c('PTS', 'FGM', 'FGA', 'DRB', 'ORB', 'TRB')]
```

També ho podríem fer utilitzant `select` de `dplyr` com havíem vist en classes anteriors. Assignem a una nova variable anomenada `NBA_pca` la matriu resultant d'aplicar la funció `prcomp()` vista en la primera part de la classe d'avui. Centrem les variables a zero utilitzant l'argument `'center=TRUE'`, i les escalem com al seminari per a que tinguin variança igual a 1, fent us de l'argument `'scale.=TRUE'`.

```
> NBA_pca<-prcomp(NBA_interes, center=TRUE, scale.=TRUE)
```

Podem per exemple mostrar un biplot

```
> fviz_pca_biplot (NBA_pca, geom.ind="point", geom.var = c("arrow", "text"), addEllipses = TRUE, Legend.tittle="Groups")
```



- PC1 correspon al 54.6% de la variança total.
- PC2 correspon al 39% de la variança total.

Per tant, coneixent la posició d'una mostra en relació amb (només) les components PC1 i PC2, podem tenir una visió molt precisa de la seva ubicació en relació amb altres mostres. Això és degut a que només PC1 i PC2 poden explicar el 93.6% de la variança.

## PART 3 (4 pt)

Dataset: Suicide rates 1985-2016.csv

Agafarem el dataset de noms del nombre de suïcidis en diferents països entre els anys 1985 i 2016 segmentat per diferents variables com sexe, edat, generació, població, etc. Utilitzeu les llibreries (plotly, gganimate, shiny, etc.) que creieu convenientes i dibuixeus les gràfiques que us facin falta.

**RESPOSTA:**

```
> library(tidyverse)
```

estás a un escaneo

de encontrar curro.



descarga la

descarga randstad app y empieza hoy.



```
> library(dplyr)
> library(plotly)
> library(shiny)

> getwd()
> setwd("C:/Users/enric/Documents/R")
> SWorld <- read.csv('./Suicide rates 1985-2016.csv')
> str(SWorld)
# Renombrar alguns camps amb caracters erronis (també es pot utilitzar la comanda rename):
> names(SWorld)[1] <- "country"
> names(SWorld)[10] <- "gdp_for_year"
> names(SWorld)[11] <- "gdp_per_capita"
> str(SWorld)
```

**3.1. (1 pt) Mostra el codi i una gràfica de línies de l'evolució del nombre de suïcidis per anys i generació. Digues quines generacions són capdavanteres i en quin període d'anys ho són.**

**RESPOSTA:**

PAS 1: DATA MASSAGING: Seleccionar variables i agrupar per generació i any

```
> SWorldG <- SWorld %>% select(generation, year, suicides_no, population,
  gdp_per_capita) %>% group_by(generation, year) %>% summarise(numeroS =
  sum(suicides_no), poblacio=sum(population), PIB=mean(gdp_per_capita))
```

O bé...

```
> SWorldG <- SWorld %>% group_by(generation, year) %>% summarise(numeroS =
  sum(suicides_no), poblacio=sum(population), PIB=mean(gdp_per_capita))
```

# PAS 2a: GRAFICA DE LINIES AMB ggplotly

```
> ggplotWG <- ggplot(SWorldG, aes(x=year, y=numeroS, color=generation)) +
  geom_line()
> ggplotly(ggplotWG)
```

# PAS 2b: GRAFICA DE LINIES AMB plotly

```
> plot_ly(SWorldG, x=~year, y=~numeroS, color=~generation) %>% add_lines()
```

# PAS 2c: GRÀFICA DE LINIES AMB shiny

```
ui <- fluidPage(  selectizeInput( inputId = "generat",
  label = "Selecciona una Generació:",
  choices = unique(SWorldG$generation),
  selected = c("Silent"),
  multiple = TRUE
),
  plotlyOutput(outputId = "p")
)
server <- function(input, output, ...)
```

```
{  output$p <- renderPlotly (
  {plot_ly(SWorldG, x = ~year, y = ~numeroS, color=~generation) %>%
```

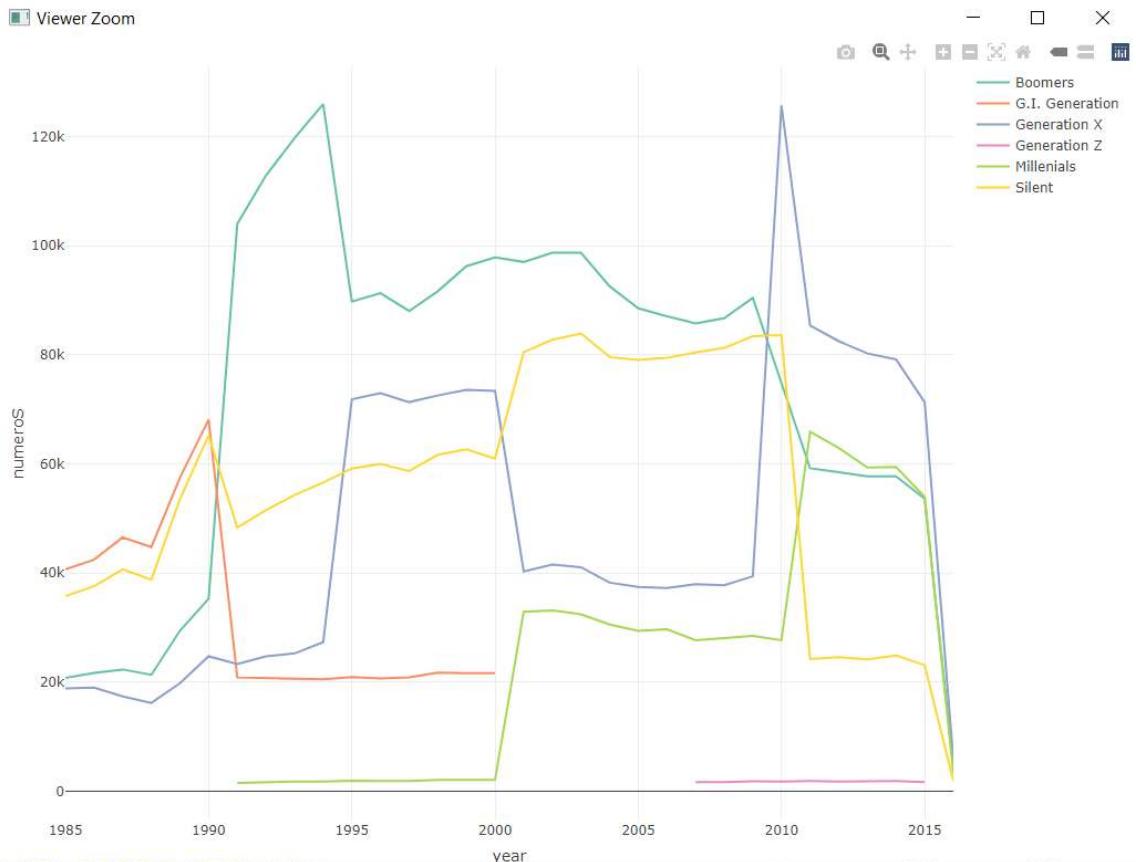
```

      filter(generation %in% input$generat) %>%
      group_by(generation) %>%
      add_lines()
    })
}
shinyApp(ui, server)

```

### GENERACIONS CAPDAVANTERES:

- GI GENERATION: 1985-1990
- BOOMERS: 1991-2009
- GENERATION X: 2010-2016



### 3.2. (0,5 pts) D'aquesta, obteniu la següent informació:

- a) Classificació per generacions l'any 2011.

**RESPOSTA:**

POSICIÓ	GENERACIÓ	Nombre de Suïcidis
1	GENERATION X	85.345
2	MILLENIALS	65.873
3	BOOMERS	59.178
4	SILENT	24.209

5	GENERATION Z	1.879
6		

b) Nombre màxim de suïcidis per generació.

**RESPOSTA:**

GENERACIÓ	ANY	Nombre de Suïcidis
BOOMERS	1994	125.932
GENERATION X	2012	125.681
GENERATION Z	2014	1.882
G.I. GENERATION	1990	68.118
MILLENIALS	2011	65.873
SILENT	2003	83.902

**3.3. (1 pt)** Mostra el codi i una gràfica d'un scatter plot 3D sobre el nombre de suïcidis en DONES per a cada 100k, població i PIB per càpita (gdp\_per\_capita).

**RESPOSTA:**

PAS 1: DATA MASSAGING: Seleccionar variables i agrupar per generació i any

```
> SWorldD <- SWorld %>% filter(sex == "female") %>% select(country, year, suicides.100k.pop, population, gdp_per_capita) %>% group_by(country) %>% summarise(numeroS.100k = sum(suicides.100k.pop), poblacio=sum(population), PIB=mean(gdp_per_capita))
```

O bé...

```
SWorldD <- SWorld %>% filter(sex == "female") %>% group_by(country) %>% summarise(numeroS.100k = sum(suicides.100k.pop), poblacio=sum(population), PIB=mean(gdp_per_capita))
```

PAS 2: GRAFICA DE SCATTER 3D AMB plotly

```
> plot_ly(SWorldD, x=~numeroS.100k, y=~poblacio, z=~PIB, color=~country, type='scatter3d', mode='markers')
```

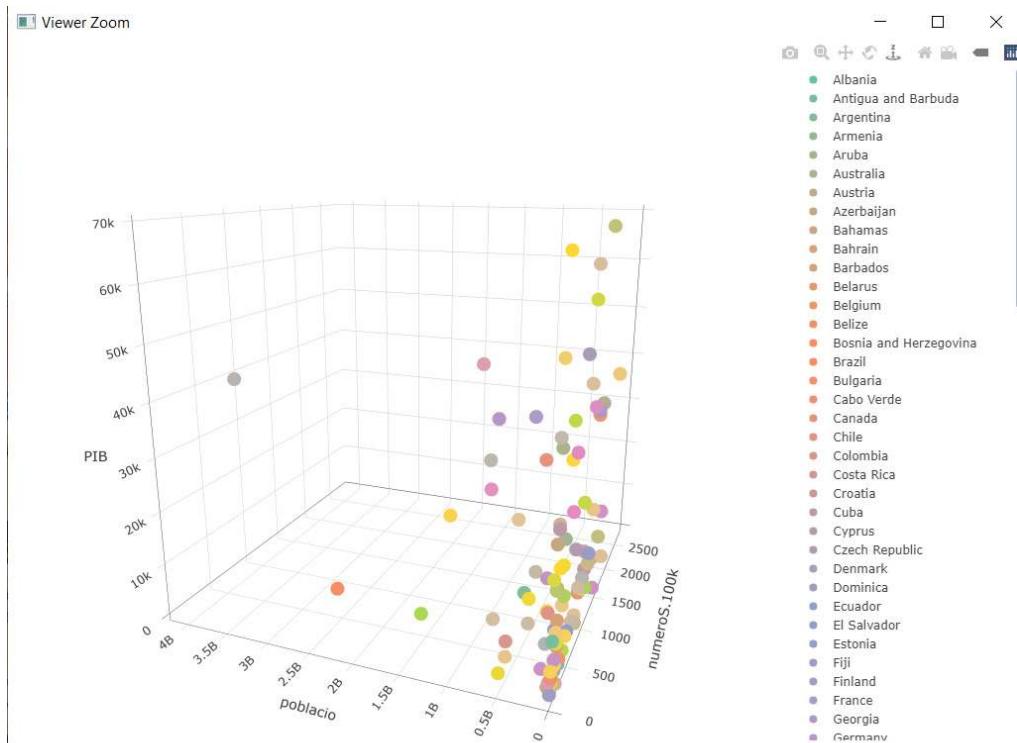
descarga randstad app y empieza hoy.



la app para encontrar curro este verano  
con más de 5000 ofertas de empleo.



descarga la ap



3.4. (0,5 pts): D'aquesta gràfica, obtenuï la següent informació:

- a) Dades (país, suicidis.100k.pop, població i PIB per càpita) dels 4 països amb més població. Trobes a faltar algun país?. Màxim 2.

**RESPOSTA:**

PAÍS	SUICIDIS.100k.pop	Població	PIB per càpita
USA	869,81	41.113.698	39.269,6
Brazil	394,84	24.654.8B	6.091,484
Russian Federation	1879,37	19.80711B	6.518.815
Japan	2546,84	18.847.87B	36.397,55

- b) Dades (país, suicidis.100k.pop, població i PIB per càpita) del païs amb més PIB per càpita.

**RESPOSTA:**

PAÍS	SUICIDIS.100k.pop	Població	PIB per càpita
Luxemburg	1455,66	6.577.184	68.798,39



4 (1 pt) Classifica les següents preguntes o mesures que s'utilitzen en tests d'Usabilitat o d'Experiència d'Usuari (UX) en les següents dues categories:

- a) Instrumentals / Usabilitat / Pragmàtiques
- b) No instrumentals / Hedònics / Emocionals

I digues en quin tipus de test s'inclou cada pregunta o mesura.

**PREGUNTES:**

1. *Complicated 1 2 3 4 5 6 7 Simple*
2. *The design looks attractive*
3. *Non-inclusive 1 2 3 4 5 6 7 Inclusive*
4. *It was easy to find the information I needed*
5. *I thought the System was easy to use*
6. *The product is creatively designed*
7. *Conservative 1 2 3 4 5 6 7 Innovative*
8. *The product is stylish: Strongly Disagree 1 2 3 4 5 6 Strongly Agree*
9. *This System has all the functions and capabilities I expect it to have*
10. *I found the various functions in this System very well integrated*

**RESPOSTA:**

PREGUNTA	CATEGORIA	TIPUS DE TEST
1	a)	Attrakdiff, Pragmatic
2	b)	meCUE, A2, Aesthetic
3	b)	Attrakdiff, Hedonic
4	a)	PSSUQ
5	a)	SUS
6	b)	meCUE, A1, Aesthetic
7	b)	Attrakdiff, Hedonic
8	b)	meCUE, A3, Aesthetic
9	a)	PSSUQ
10	a)	SUS

2025  
10-13 JULY



**Visualització de dades (Enginyeria de Dades - EE - UAB)**  
**Examen Segon Parcial - 19 Juny 2023**  
**RESPOSTES MODEL A**

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 (6 pt)

*Dataset: Australian\_Flight\_dataset\_2003-2022.csv*

*Agafarem aquest dataset que conté totes les rutes (cada ruta conté un o més vols diaris, setmanals o mensuals) d'arribada i/o sortida als aeroports d'Austràlia durant els anys 2003 al 2022. Utilitzeu les llibreries (plotly, gganimate, shiny, etc.) que creieu convenientes i dibuixeu les gràfiques que us facin falta.*

*Cada registre conté informació d'una ruta d'una companyia aèria. Conté les variables:*

1. Month ---> Date in month and day
2. In\_Out ---> Status of flight [Incoming/Outgoing]
3. Australian\_City ---> Australian city name
4. International\_City ---> International city name
5. Airline ---> Airline owning the flight
6. Route ---> Route taken by the flight
7. Port\_Country ---> Port country
8. Port\_Region ---> Port region
9. Service\_Country ---> Service country
10. Service\_Region ---> Service region
11. Stops ---> Number of stops taken by the flight
12. All\_Flights ---> Total number of flights
13. Max\_Seats ---> Total capacity of seats in flight
14. Year ---> Date in year
15. Month\_num ---> Date in month number

*Com a principals variables que utilitzarem en aquesta 1a part, cal ressaltar Airline, Max\_Seats i Year.*

*Càrrega de les llibreries i dataset:*

```
> library(tidyverse)
> library(dplyr)
> library(plotly)
> library(shiny)
> getwd()
> setwd("C:/Users/enric/Documents/R")
```

```
> Australian_FD <- read.csv('./Australian_Flight_Dataset_2003-2022.csv')
```

**1.1 (2 pts)** Mostra el codi (en ggplotly) i la gràfica de línies de l'evolució temporal en anys del nombre de rutes de cada companyia, seguin de sortida o arribada a qualsevol ciutat australiana (gràfica 1) i fixa't amb les tres companyies aèries amb més rutes l'any 2017. Digues quines companyies són per ordre decreixent i mostra el codi (en ggplotly o shiny+plotly) i la gràfica de línies de l'evolució temporal d'aquestes tres companyies (gràfica 2). Especifica si fas Data Massaging.

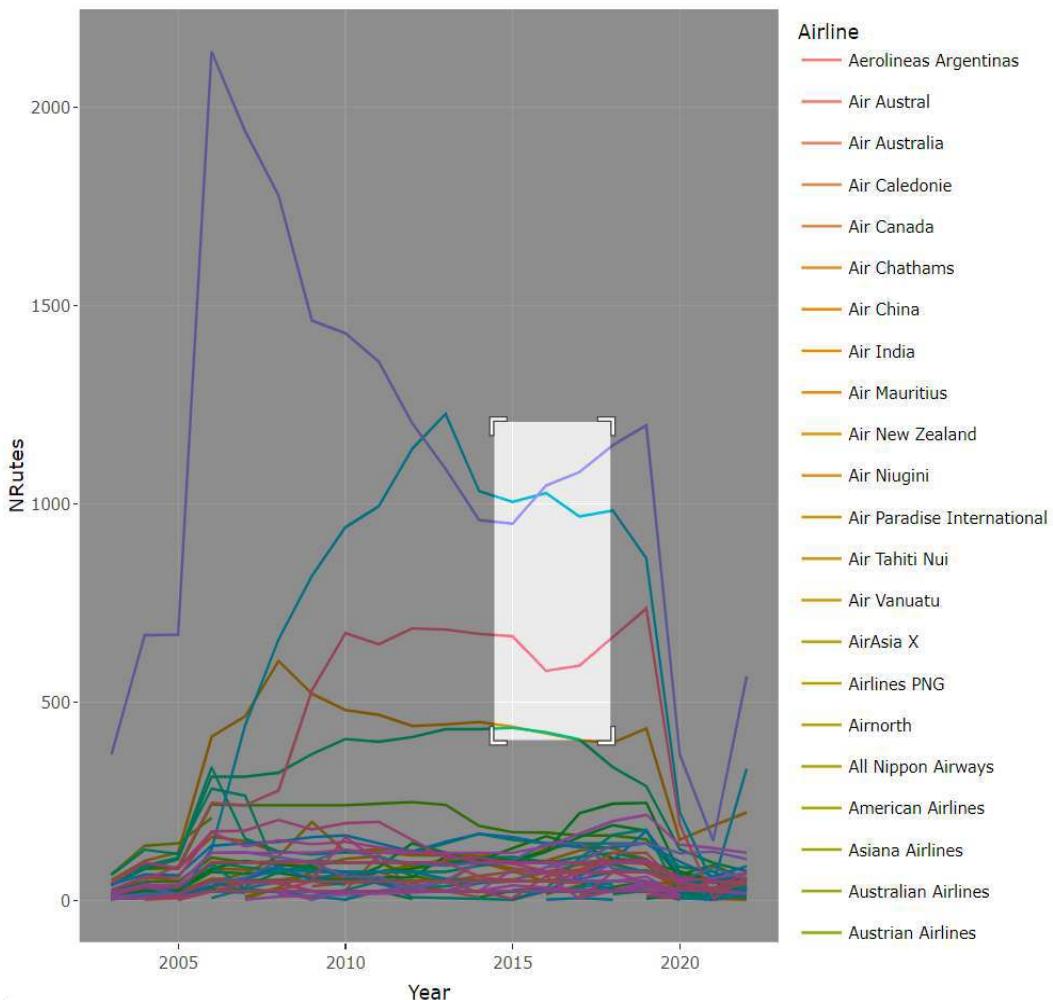
**RESPOSTA:** En primer lloc, hem de fer data massaging sobre el data set original.

**DATA MASSAGING:** Agrupar per (Any, Companyia Aèria, Sexe) calculant el nombre de rutes:

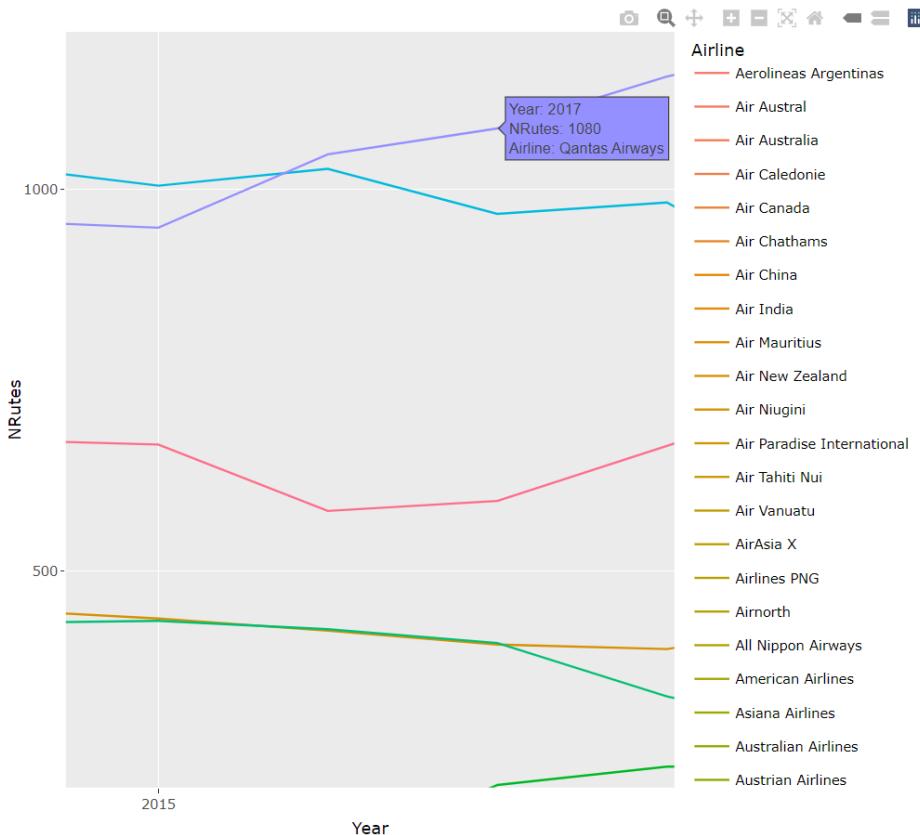
```
> Australian_FD_Rutes <- Australian_FD %>% group_by(Year,Airline) %>%  
summarize(NRoutes = n())
```

**GRÀFICA 1 de línies amb ggplotly:**

```
> ggplotR <- ggplot(Australian_FD_Rutes, aes(x=Year, y=NRoutes,  
color=Airline)) + geom_line()  
> ggplotly(ggplotR)
```



I fer un zoom en la zona propera a l'any 2017:



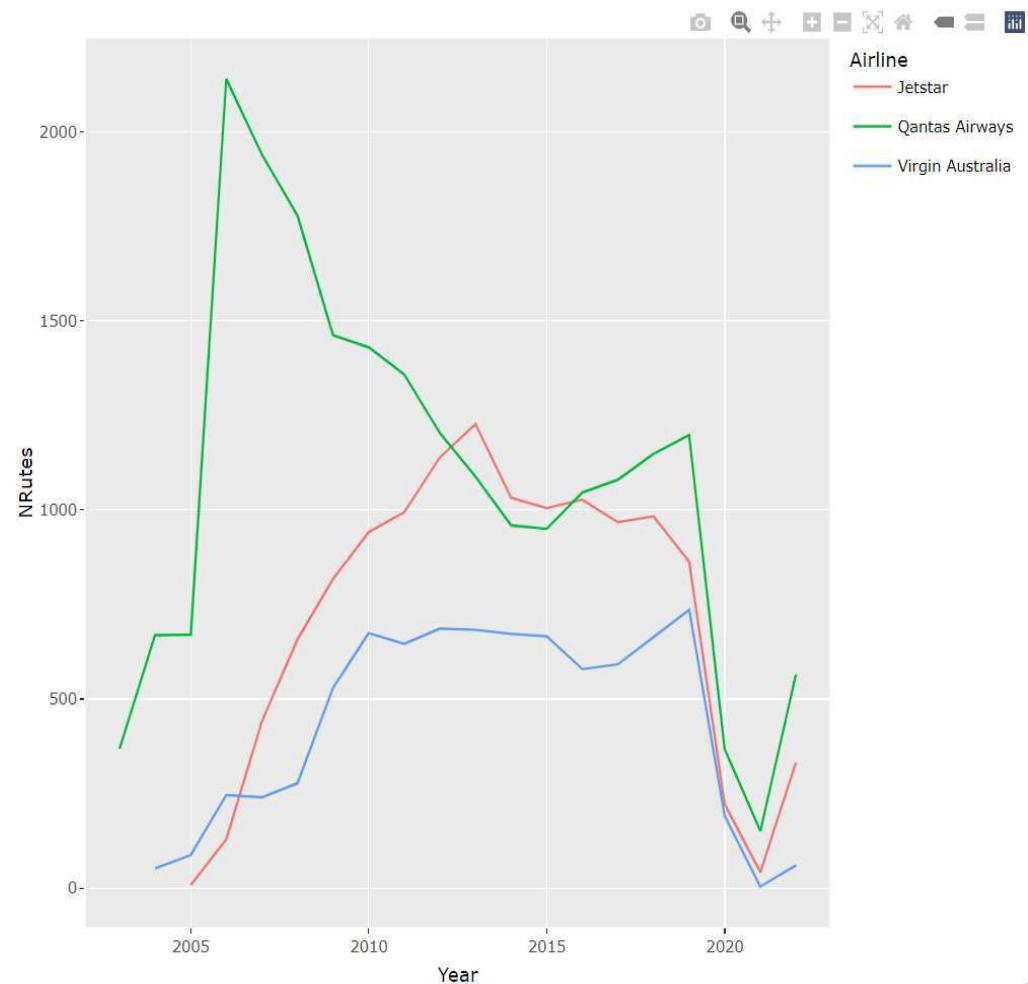
Els noms de les companyies són **Qantas Airways (1080 rutes), Jetstar (968) i Virgin Australia (592)**

**GRÀFICA 2:** Després podem llistar la gràfica de dues formes:

**AMB ggplotly (abans filtrem el data set per les tres companyies aèries):**

```
> Australian_FD_Rutes_QJV <- Australian_FD_Rutes %>% filter(Airline == "Qantas Airways" | Airline == "Jetstar" | Airline == "Virgin Australia")
> ggplotQJV <- ggplot(Australian_FD_Rutes_QJV, aes(x=Year, y=NRoutes,
color=Airline)) + geom_line()
> ggplotly(ggplotQJV)
>
```

2025  
10-13 JULY



### AMB shiny + plotly:

```
ui <- fluidPage(
  selectizeInput(  inputId = "Airline",
  label = "Selecciona una COMPANYIA:",
  choices = unique(Australian_FD_Rutes_QJV$Airline),
  selected=c("Qantas Airways","Jetstar",
            "Virgin Australia"),
  multiple = TRUE
  ),
  plotlyOutput(outputId = "p")
)

server <- function(input, output, ... )
  {output$p <- renderPlotly (
    {plot_ly(Australian_FD_Rutes_QJV,x=~Year,y=~NRutes,color=~Airline)%>%
      filter(Airline %in% input$Airline) %>%
      group_by(Airline) %>%
      add_lines()
    })
}
shinyApp(ui, server)
```

Selecciona una COMPANYIA:

Qantas Airways Virgin Australia Jetstar



1.2 (0,5 pts) Compara les tres companyies en 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
Qantas Airways	2021 - 151	2006 – 2140
Jetstar	2005 – 8	2013 – 1227
Virgin Australia	2021 – 4	2019 – 736

- b) (0,1 pts) Comenta la taula anterior. Trobes alguna singularitat?. Sabries el per què?.

**RESPOSTA:** El mínim nombre de rutes de les companyies coincideix el 2021, amb la pandèmia del Covid-19.

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

**RESPOSTA:**

NOMS	Any Anterior	Any Posterior
Virgin Australia-Jetstar	2006	2007
Jetstar-Qantas Airways	2012	2013
Qantas Airways-Jetstar	2015	2016

**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 les 10 companyies amb més rutes per anys, del 2003 al 2020.**

*Càrrega de les llibreries:*

```
> library(tidyverse)
> library(gganimate)      # Generació de frames i la seva compilació per a
generar fitxer animació
> library(gifski)        # Generació fitxer animació GIF
```

### **PAS 1: DATA MASSAGING:**

- *Filtrar el dataset pels anys 2003 al 2020:*

```
> Australian_FD_Rutes <- Australian_FD_Rutes %>% filter(Year < 2021)
```

- *Definir nova variable rank amb la posició decreixent en nombre de rutes i filtrar les 10 primeres companyies segons rank:*

```
> Australian_FD_Rutes_formatted <- Australian_FD_Rutes %>%
group_by(Year) %>%
# The * 1 makes it possible to have non-integer ranks while sliding
mutate(rank = rank(-NRoutes)) %>%
group_by(Airline) %>%
filter(rank <=10) %>%
ungroup()
```

### **PAS 2: ANIMATED BAR RACE RANKING:**

```
> anim <- ggplot(Australian_FD_Rutes_formatted, aes(rank, group = Airline,
           fill = as.factor(Airline), color = as.factor(Airline))) +
geom_tile(aes(y = NRutes/2,
              height = NRutes,
              #width = 0.9), alpha = 0.8, color = "black") +
width = 0.9), alpha = 0.8, color = NA) +
geom_text(aes(y = 0, label = paste(Airline, " ")), vjust = 0.2, hjust = 1) +
geom_text(aes(y=NRutes,label = NRutes, hjust=0)) +
coord_flip(clip = "off", expand = FALSE) +
scale_y_continuous(labels = scales::comma) +
scale_x_reverse() +
guides(color = "none", fill = "none") +
theme(axis.line=element_blank(),
axis.text.x=element_blank(),
axis.text.y=element_blank(),
axis.ticks=element_blank(),
axis.title.x=element_blank(),
axis.title.y=element_blank(),
legend.position="none",
panel.background=element_blank(),
panel.border=element_blank(),
panel.grid.major=element_blank(),
panel.grid.minor=element_blank(),
panel.grid.major.x = element_line( size=.1, color="grey" ),
panel.grid.minor.x = element_line( size=.1, color="grey" ),
plot.title=element_text(size=25, hjust=0.5, face="bold", colour="grey",
vjust=-1),
```

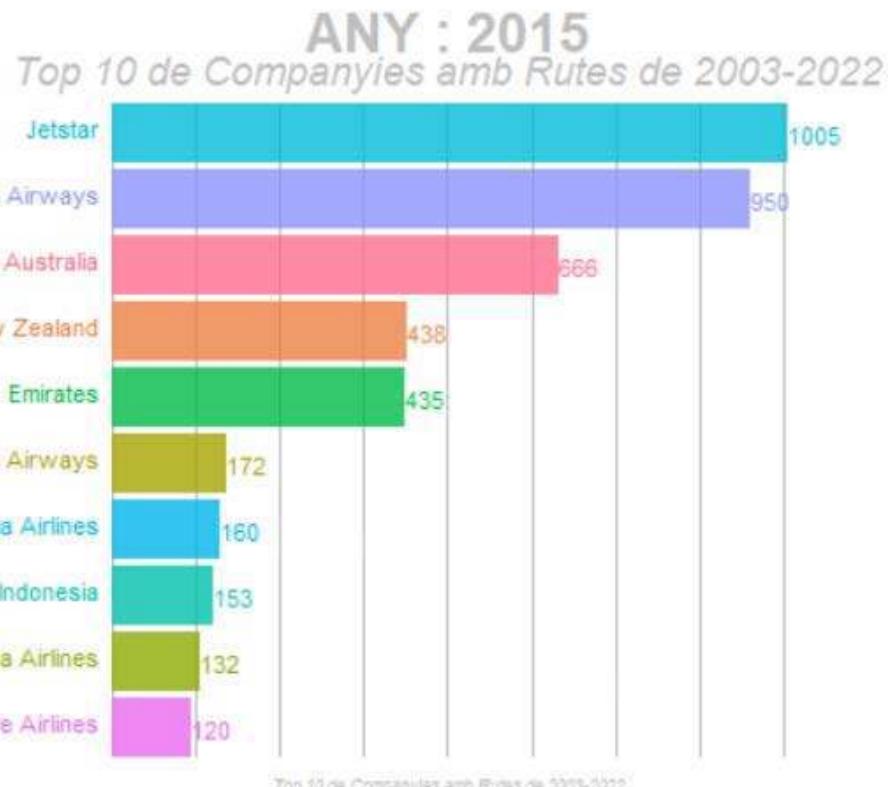
2025  
10-13 JULY



```
plot.subtitle=element_text(size=18,           hjust=0.5,           face="italic",
color="grey"),
plot.caption     =element_text(size=8,           hjust=0.5,           face="italic",
color="grey"),
plot.background=element_blank(),
plot.margin = margin(2,2, 2, 4, "cm")) +
transition_states(Year, transition_length = 4, state_length = 1, wrap = FALSE)
+
view_follow(fixed_x = TRUE) +
labs(title = 'ANY : {closest_state}',
subtitle = "Top 10 de Companies amb Rutes de 2003-2022",
caption = "Top 10 de Companies amb Rutes de 2003-2022")
> anim

# PAS 2. Exportar frames a fitxer GIF
> animate(anim, 200, fps = 20, duration= 30, width = 1200, height = 1000,
          renderer = gifski_renderer("USNamesF.gif"), end_pause = 15, start_pause
= 15)
>

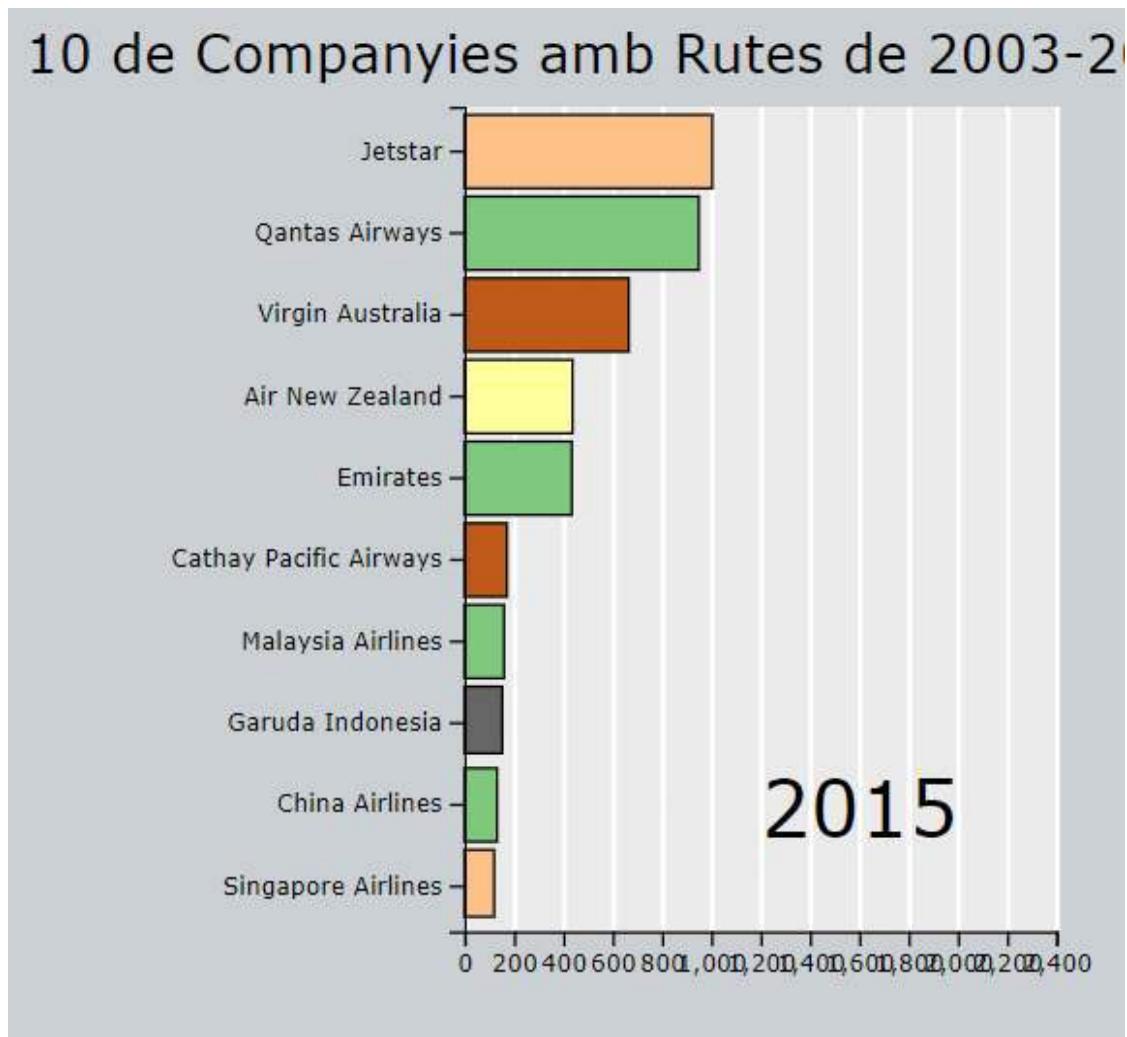
# PAS 3. Exportar frames a fitxer AVI
> animate(anim, 200, fps = 20, duration= 30, width = 1200, height = 1000,
          renderer = av_renderer("USNamesF.avi"), end_pause = 15, start_pause =
15)
```



**ANIMATED BAR RACE RANKING. ALTERNATIVA** (*no available, doncs no s'ha explicat, però es posa per a que ho coneixeu*): Llibreria ddplot que defineix una gràfica de forma més senzilla i clara

```
> install.packages("ddplot") # Càrrega paquet ddplot (primera vegada)
> library(ddplot)           # Generació de Racing Bar Charts

> Australian_FD_Rutes_formatted %>%
  barChartRace(
    x = "NRutes",
    y = "Airline",
    time = "Year",
    title = "Top 10 de Companies amb Rutes de 2003-2022"
  )
```



No hem trobat encara la forma de generar fitxer de vídeo amb aquesta funció.

1.4 (1 pt) Classifica les següents preguntes o mesures que s'utilitzen en tests d'Usabilitat o d'Experiència d'Usuari (UX) en les següents dues categories:

- a) Instrumentals / Usabilitat / Pragmàtiques
- b) No instrumentals / Hedònics / Emocionals

**PREGUNTES:**

1. *Complicated 1 2 3 4 5 6 7 Simple*
2. *Non-inclusive 1 2 3 4 5 6 7 Inclusive*
3. *The design looks attractive*
4. *It was easy to find the information I needed*
5. *I thought the System was easy to use*
6. *Conservative 1 2 3 4 5 6 7 Innovative*
7. *The product is creatively designed*
8. *This System has all the functions and capabilities I expect it to have*
9. *I found the various functions in this System very well integrated*
10. *The product is stylish: Strongly Disagree 2 3 4 5 6 Strongly Agree*

**RESPOSTA:**

PREGUNTA	CATEGORIA	TIPUS DE TEST
1	a)	Attrakdiff, Pragmatic
2	b)	Attrakdiff, Hedonic
3	b)	meCUE, A2, Aesthetic
4	a)	PSSUQ
5	a)	SUS
6	b)	Attrakdiff, Hedonic
7	b)	meCUE, A1, Aesthetic
8	a)	PSSUQ
9	a)	SUS
10	b)	meCUE, A3, Aesthetic

1.5 (1 pt) Defineix 4 dels següents conceptes en animació, interactivitat, usabilitat i Experiència d'Usuari en Visualització de Dades:

- Event
- *plotly\_hover*
- *Data linking*
- *Slider*
- *Learnability*
- *Toggle*

Guardant només les ciutats de les quals us hem donat la longitud i latitud, construïu un dataframe, que contingui el nombre de rutes que surten (*In\_Out=0*) de cada ciutat de l'any 2022. Un cop tingueu el dataframe, ajunteu-lo amb el dataframe creat amb dites latituds i longituds. Per això últim, si ho feu en R, podeu utilitzar la comanda *inner\_join* donada:

```
> Australian_FD_Rutes2022<- Australian_FD%>%filter(Australian_City %in%
c("Adelaide", "Brisbane", "Cairns", "Darwin", "Melbourne", "Perth",
"Sydney")) %>% filter (Year==2022) %>%filter(In_Out=="0")
%>% group_by(Australian_City) %>% summarize(NRutes = n())
```

(No es penalitza si no es fa el *filter(In\_Out=="0")*)

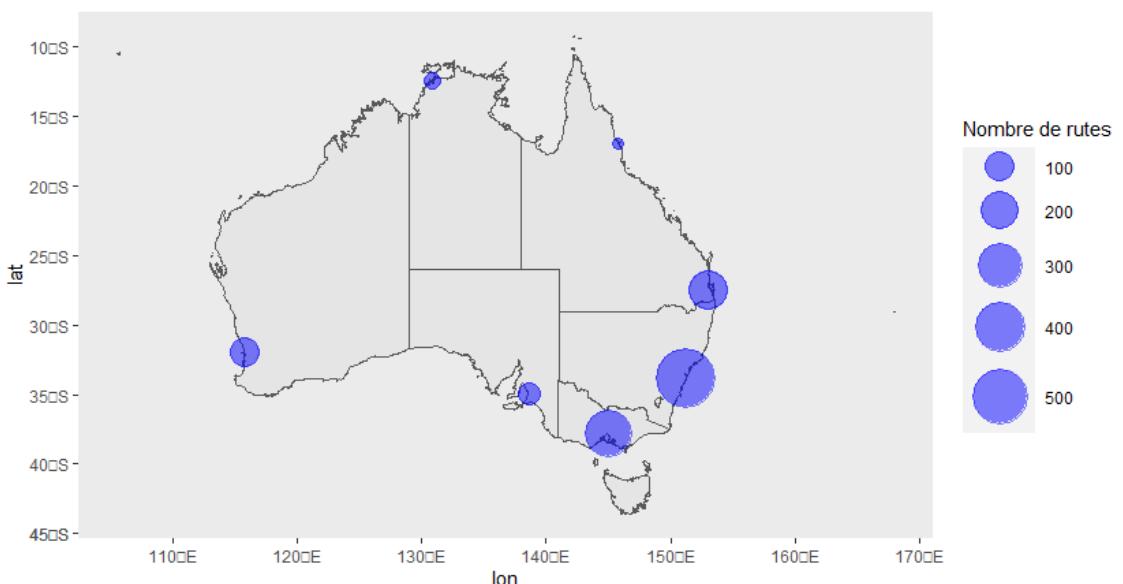
```
> dades<-
inner_join(Australian_FD_Rutes2022, oz_cities, by="Australian_City")
```

(b) Sobre el mapa de oz\_states Dibuixeu tot traient la graella del fons. Feu un bubble map on la talla de les bombolles creix amb el nombre de rutes per ciutat

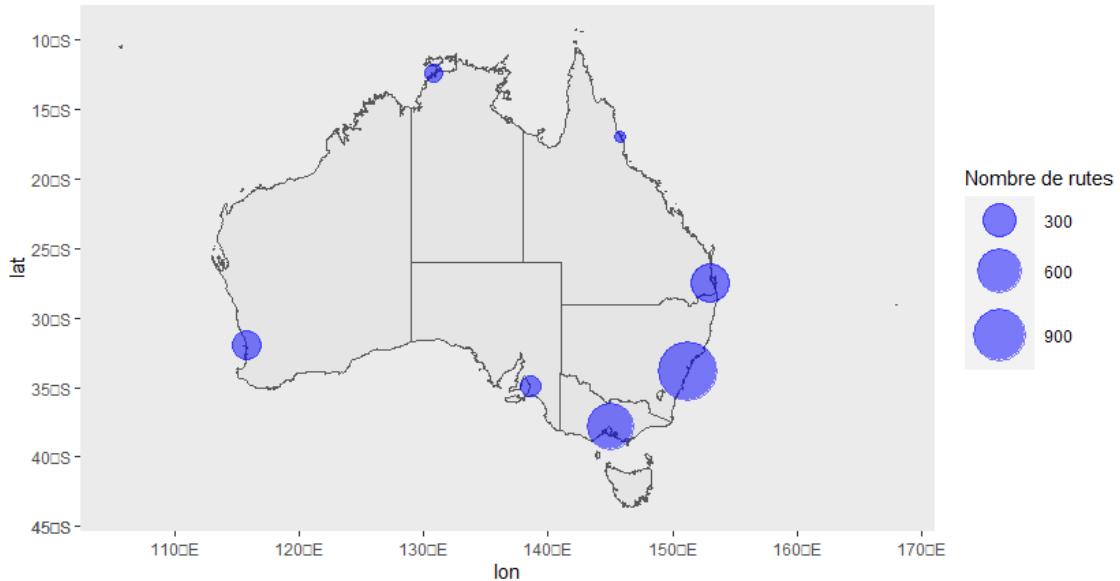
```
> Library (ozmaps)
> Library (sf)

> oz_states <- ozmaps::ozmap_states

> ggplot() + geom_sf(data = oz_states)+geom_point(data = dades, mapping =
aes(x      =     Lon,       y      =     Lat, size=NRutes), alpha=0.5,
color="blue")+coord_sf()+theme(panel.grid.major      =     element_blank(),
panel.grid.minor      =     element_blank())
+scale_size_continuous(range=c(3,15), "Nombre de rutes")
```



També es compta bé si no s'ha especificat el filtre per agafar només les rutes de sortida. En aquest cas el mapa seria:



**2.2 (2 pt)** Mostra un mapa d'arbre (treemap) que et permeti saber des de quina de les dues ciutats, Melbourne o Sidney, van volar més vols l'any 2022 a cada regió de servei (Service\_Region).

- Fes el datamassage necessari (0.5 pt).
- Argumenta com és un *treemap* en general i els passos que has de fer per construir la visualització d'aquest exercici (és a dir quina variable utilitzes per l'àrea de les graelles, variables d'agrupació, etc). Inclou el codi si el fas en R (1 pt).
- Mostra el mapa d'arbre, posa llegenda i títol i posa una paleta de color (en R podeu usar una paleta manual, per exemple amb els colors `c('#87C55F', '#9EB9F3')`, o una altra al vostre gust). Quina és/són la/les regions de destí on l'any 2022 no van volar vols des de Melbourne? (0.5 pt).

#### RESPOSTA:

a)

```
> library ("tidyverse")
> Aust_FD_regiodes2022<- Australian_FD %>% filter(Australian_City %in%
c("Melbourne", "Sydney")) %>% filter (Year==2022) %>% filter(In_Out=="0")
(Com abans, no es penalitza si no es fa el filter(In_Out=="0"))
```

b) Com vam veure a classe, un mapa d'arbre és un dibuix rectangular dividit en caselles, i cada casella representa una sola observació. Vam veure que era una bona manera de mostrar dades jeràrquiques mitjançant rectangles imbricats. L'àrea relativa de cada casella expressava una variable contínua. També vam veure que era òptim quan hi ha com a màxim dues variables d'agrupació, per tant no en definirem més.

Per tant, una possibilitat amb aquest dataset seria:

2025  
10-13 JULY



- Per definir el color i actuar doncs com un 'grup pare' utilitzem la variable Australian\_City (Agafem aquesta com grup de color, ja que sol té 2 nivells ("Melbourne", "Sydney" i serà més fàcil entendre la llegenda, que si posem Service\_Region que en té més).
- Com a 'subgrup' utilitzaríem la regió de servei (Service\_Region, que té més nivells que Australian City).
- Com a variable que descrigui l'àrea de les caselles triarem All\_Flights. Aquesta tria ve donada perquè ens estan preguntant per regions amb més/menys vols des d'una certa ciutat. I de totes les variables que ens pregunten, és l'única numèrica.

*Si ho fem amb R, el primer que hem de fer és carregar les llibreries necessàries. La llibreria específica aquí és treemapify (abans necessitarem tenir instal·lat el paquet com vam veure al seminari corresponent). També haurem de fer ús de geom\_treemap.*

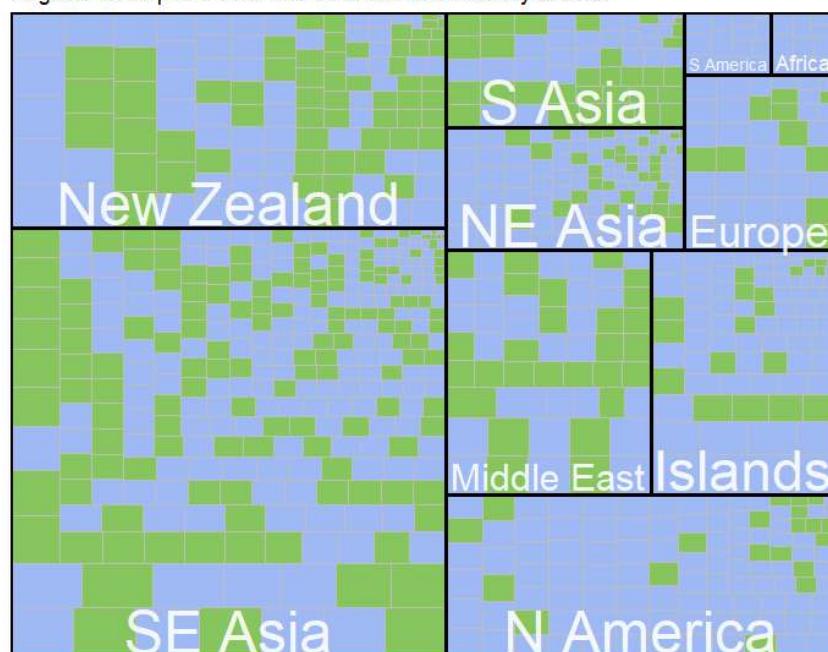
*Per fer un treemap basic doncs:*

```
> library ("treemapify")
> ggplot(Aust_FD_regiodesi2022, aes(area = All_Flights, fill =
  Australian_City, subgroup=Service_Region))+geom_treemap()+
  geom_treemap_subgroup_border(colour = "black",size=3)
  +geom_treemap_subgroup_text(alpha=0.9,colour = "white")+labs(title =
  "Regions on es podia volar des de Melbourne i Sidney al 2022", fill="Ciutat
  Australiana") +scale_fill_manual(values=c('#87C55F','#9EB9F3'))
```

*Fixeu-vos que hem posat el color amb fill:*

```
scale_fill_manual(values=c ('#87C55F', '#9EB9F3'))
```

Regions on es podia volar des de Melbourne i Sidney al 2022

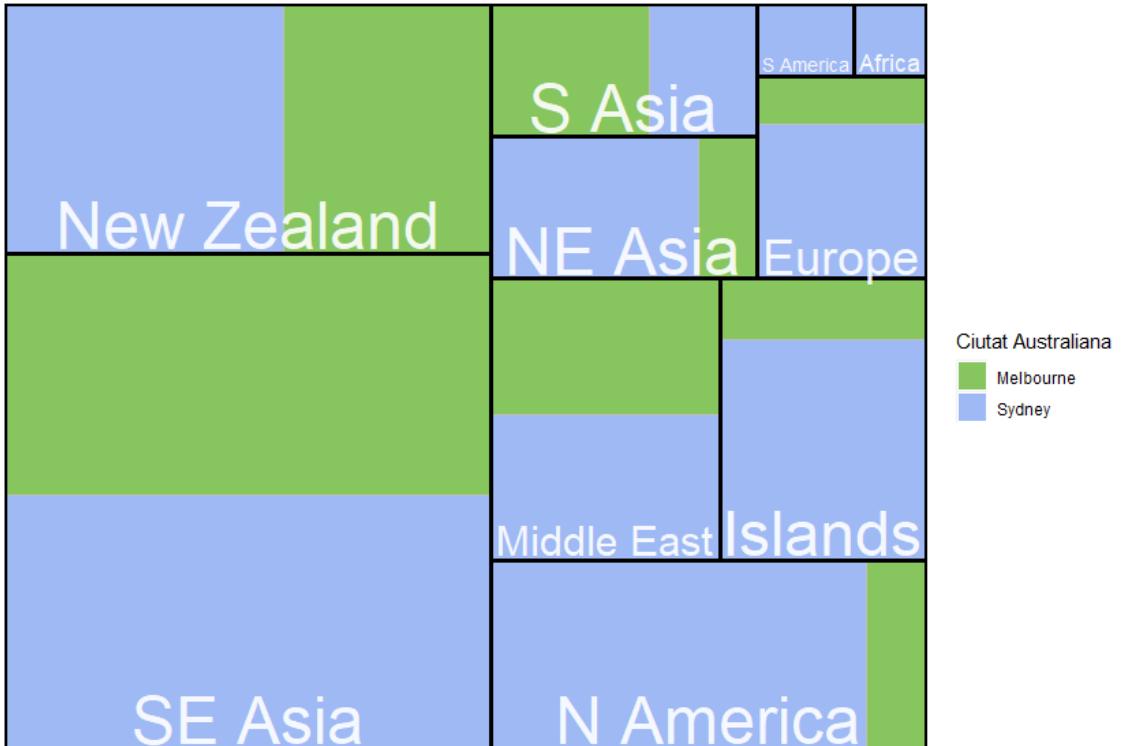


Podeu a més ordenar la visualització:

```
> Aust_FD_regiodes2022<-  
Aust_FD_regiodes2022%>%group_by(Australian_City,  
Service_Region)%>%summarise(All_Flights=sum(All_Flights))
```

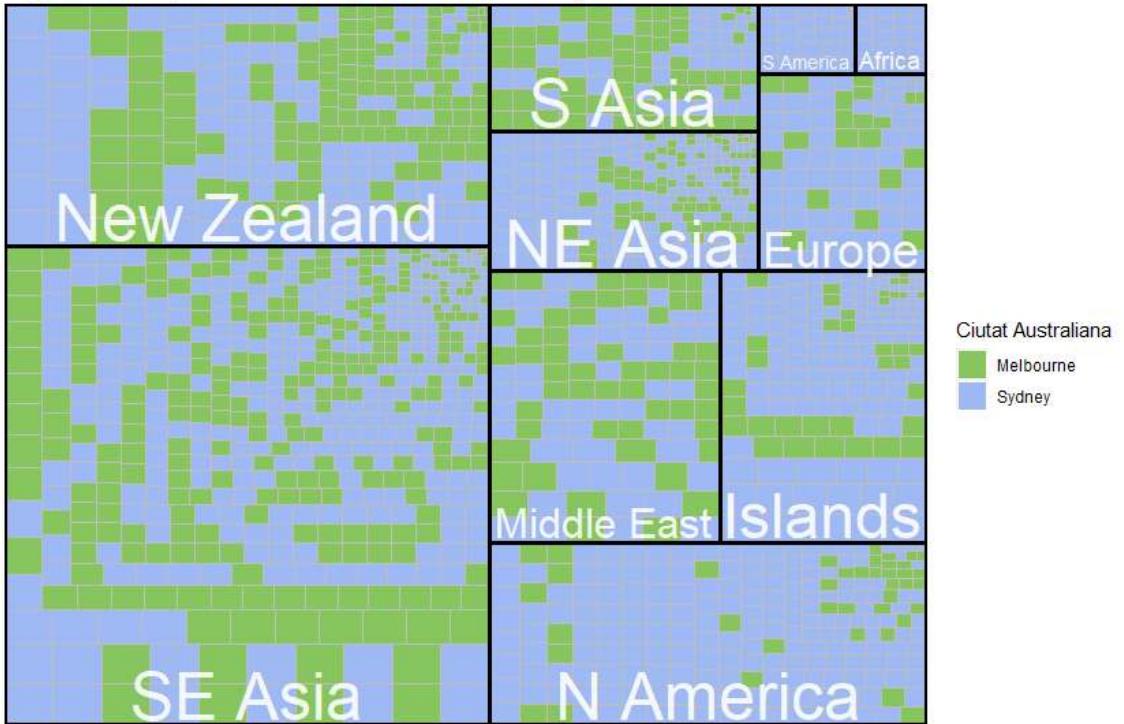
Us sortirà un warning que no té major importància, però que podeu treure (veure: <https://statisticsglobe.com/dplyr-message-summarise-has-grouped-output-r>)

Regions on es podia volar des de Melbourne i Sidney al 2022



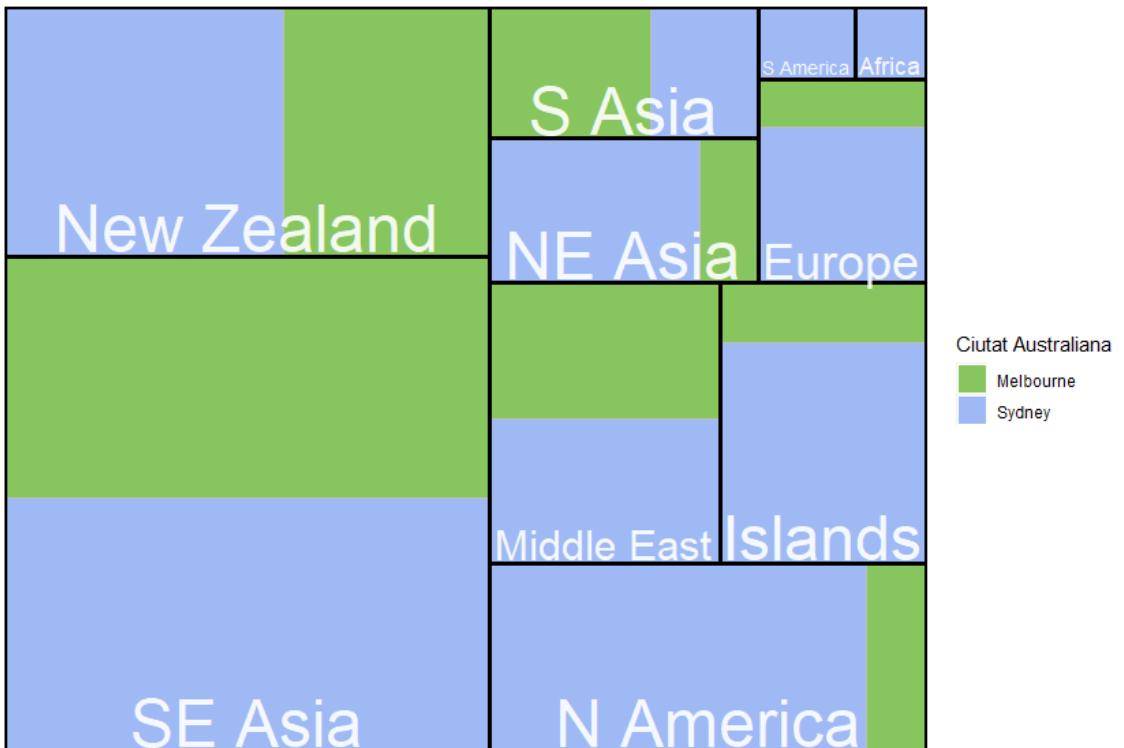
I si no es fa el filter(*In\_Out=="0"*):

Regions on es podia volar des de Melbourne i Sidney al 2022



O ordenant com abans:

Regions on es podia volar des de Melbourne i Sidney al 2022



En qualsevol cas, la resposta és: A Àfrica i Sud Amèrica no hi van volar vols des de Melbourne al 2022.



Preparando un cuestionario con tus apuntes...

**Visualització de dades (Enginyeria de Dades - EE - UAB)**  
**Examen Segon Parcial - 19 Juny 2023**  
**MODEL B**

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 (6 pt)

**Dataset: Australian\_Flight\_dataset\_2003-2022.csv**

Agafarem aquest dataset que conté totes les rutes (cada ruta conté un o més vols diaris, setmanals o mensuals) d'arribada i/o sortida als aeròports d'Austràlia durant els anys 2003 al 2022. Utilitzeu les llibreries (plotly, ganimate, shiny, etc.) que creieu convenient i dibuixeu les gràfiques que us facin falta.

Cada registre conté informació d'una ruta d'una companyia aèria. Conté les variables:

1. Month ---> Date in month and day
2. In\_Out ---> Status of flight [Incoming/Outgoing]
3. Australian\_City ---> Australian city name
4. International\_City ---> International city name
5. Airline ---> Airline owning the flight
6. Route ---> Route taken by the flight
7. Port\_Country ---> Port country
8. Port\_Region ---> Port region
9. Service\_Country ---> Service country
10. Service\_Region ---> Service region
11. Stops ---> Number of stops taken by the flight
12. All\_Flights ---> Total number of flights
13. Max\_Seats ---> Total capacity of seats in flight
14. Year ---> Date in year
15. Month\_num ---> Date in month number

Com a principals variables que utilitzarem en aquesta 1a part, cal ressaltar Airline, Max\_Seats i Year.

**1.1 (2 pts) Mostra el codi (en ggplotly) i la gràfica de línies de l'evolució temporal en anys de la capacitat en seients de cada companyia, siguin de rutes de sortida o arribada a qualsevol ciutat australiana (gràfica 1) i fixa't amb les tres companyies aèries amb més seients l'any 2021. Digues quines companyies són per ordre decreixent i mostra el codi (en ggplotly o shiny+plotly) i la gràfica de línies de l'evolució temporal d'aquestes tres companyies (gràfica 2). Especifica si fas Data Massaging.**

**RESPOSTA:**

**1.2 (0,5 pt) Compara les tres companyies en 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)** Comenta la taula anterior. Trobes alguna singularitat?.

**RESPOSTA:**

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

**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 les 10 companyies amb més seients per anys, del 2003 al 2020.**

**RESPOSTA:**

**1.4 (1 pt) Classifica les següents preguntes o mesures que s'utilitzen en tests d'Usabilitat o d'Experiència d'Usuari (UX) en les següents dues categories:**

- a) Instrumentals / Usabilitat / Pragmàtiques  
b) No instrumentals / Hedònics / Emocionals

**PREGUNTES:**

1. *Conservative 1 2 3 4 5 6 7 Innovative*
2. *Non-inclusive 1 2 3 4 5 6 7 Inclusive*
3. *This System has all the functions and capabilities I expect it to have*
4. *Complicated 1 2 3 4 5 6 7 Simple*
5. *I thought the System was wasy to use*
6. *The product is stylish: Strongly Disagree 2 3 4 5 6 Strongly Agree*
7. *The design looks attractive*
8. *I found the various functions in this System very well integrated*
9. *It was easy to find the information I needed*

**10. The product is creatively designed**

**RESPOSTA:**

PREGUNTA	CATEGORIA	TIPUS DE TEST
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

**1.5 (1 pt) Defineix 4 dels següents conceptes en animació, interactivitat, usabilitat i Experiència d'Usuari en Visualització de Dades:**

- Usabilitat
- *plotly\_click*
- *Utility*
- *Toggle*
- *Efficiency*
- *Control*

**RESPOSTA:**

## **PART 2 (4 pt)**

**Dataset: Australian\_Flight\_dataset\_2003-2022.csv**

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

**2.1 (2 pt) Fes un bubble map per representar el nombre de rutes que van sortir l'any 2018 de cadascuna de les següents ciutats d'Austràlia: "Adelaide", "Brisbane", "Cairns", "Darwin", "Melbourne", "Perth", "Sydney".**

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

	Australian_City	lat	lon
1	Sydney	-33.8688	151.2093
2	Melbourne	-37.8136	144.9631
3	Brisbane	-27.4698	153.0251
4	Adelaide	-34.9285	138.6007
5	Perth	-31.9505	115.8605
6	Cairns	-16.9230	145.7662
7	Darwin	-12.4612	130.8418

~ Australian\_City, ~lat, ~lon,  
"Sydney", -33.8688, 151.2093,  
"Melbourne", -37.8136, 144.9631,  
"Brisbane", -27.4698, 153.0251,  
"Adelaide", -34.9285, 138.6007,  
"Perth", -31.9505, 115.8605,  
"Cairns", -16.9230, 145.7662,  
"Darwin", -12.4612, 130.8418,



la app para encontrar curro este verano  
con más de 5000 ofertas de empleo.



descarga la ap

**Visualització de dades (Enginyeria de Dades - EE - UAB)**  
**Examen Segon Parcial - 3 Juny 2022**  
**RESPOSTES MODEL B**

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.**

### **PARTE 1 (2,5 pt)**

**Dataset: GDP-LifeExp-Population.csv**

**\* Lee los tres enunciados de esta parte antes de empezar \***

**1.1. (0,5 pt)** Filtra las observaciones que tengan valores NAN o null.

Identifica outliers (valores extremos que distorsionen la muestra) en la variable *Population* y filtra las observaciones que no sean comparables al resto.

**RESPOSTA:**

Nulls: Andorra, Eritrea, Gibraltar, N. Korea, Gibraltar, F. Polynesia, S. Sudan, Venezuela

Outliers: World, OCDE members

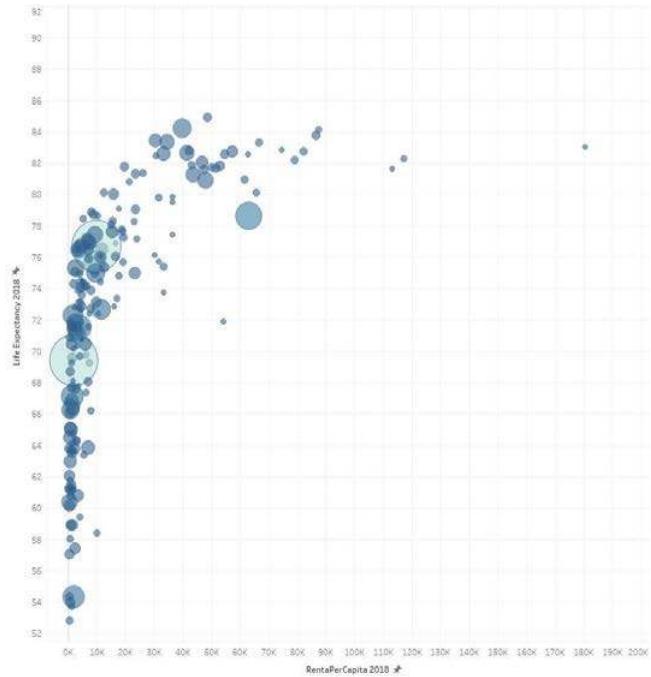
**1.2. (1 pt)** Haz un **bubble chart** con las variables *Life expectancy 2018*, *PIB per capita USD 2018* y *Population 2018*, donde *Population* sea el tamaño de los círculos.

Los círculos deben estar coloreados por la variable de tu elección que creas que más aporta a la visualización.

Añade las etiquetas y leyendas necesarias para que la gráfica sea legible y comprensible.



**RESPOSTA:**



**1.3. (0,5 pt)** ¿Por qué un bubble chart es una visualización apropiada en este caso? Razona brevemente según el tipo y el número de atributos.

**RESPOSTA:**

Tenemos tres atributos cuantitativos y un número alto de observaciones (200). Un bubble chart sirve para graficar la relación entre tres atributos cuantitativos, dos en la posición X e Y, y un tercero en el tamaño de los círculos. Además es un tipo de gráfica que escala bien a cientos de observaciones, a diferencia de otras, como las gráficas de barras.

**1.4. (0,5 pt)** ¿Qué variable usas para colorear los datos y qué escala de color utilizas? Razona brevemente porqué la variable y porqué la escala.

**RESPOSTA:**

Cualquiera de las 3 cuantitativas serviría, aunque la más apropiada será *Population* para reforzar el canal más débil, que es el tamaño de los círculos. La opción incorrecta sería usar el país.

## PART 2 (3,5 pt)

**Dataset:** starwars de R per l'exercici 2.1

**Dataset:** statsNBA2008.csv. Aquest dataset té 21 atributs de 50 jugadors de la NBA. Recull les estadístiques de la NBA del 2008. I estudiarem les components principals d'alguns dels atributs en l'exercici 2.2.

**2.1 (2,5 pt)** Volem representar les alçades dels personatges de starwars (dades incloses en R) en un mapa d'arbre (*treemap*) que us permeti contestar les preguntes de l'apartat d. El mapa d'arbre (*treemap*) podeu fer-lo en R o en qualsevol altre llenguatge, utilitzant les llibreries que us semblin convenientes.

- a) Feu el següent data massage: Descarteu les files del dataframe on la variable alçada ('height') o la variable món natal ('homeworld') continguin un valor Nan i reemplaceu els Nans de la columna sexe per 'none'. (0,5 pt)
- b) Feu un mapa d'arbre. Ara bé, heu d'explicar-ho bé: tot narrant com és un treemap en general - redacteu els passos que heu de fer per construir la visualització triada. Poseu llegenda i títol i mostreu el gràfic (1,25 pt)
- c) Canvieu la paleta de color. Si treballau en R podeu posar manualment la paleta de color `c('#DCB0F2', '#87C55F', '#9EB9F3', '#FE88B1', '#C9DB74')` o posar una altra paleta adient que us agradi. En tots els casos, si treballau amb R, en un altre llenguatge o amb Tableau, presenteu en l'apartat anterior el gràfic que us ha sortit amb el color per defecte i aquí amb una paleta diferent (0,25 pt)
- d) Un cop tingueu la visualització, contesteu: Quin és el personatge més baix nascut a Tatooine i de quin sexe és? Quants personatges hermafrodites us han sortit? Digueu un i qui és el seu món natal? (0,25 pt)
- e) Expliqueu en quin cas són òptims el mosaic, el treemap i un paral·lel set (sense necessitat de fer cap visualització) i quines diferències principals hi ha entre ells. (0,25 pt)

**RESPOSTA:**

a)

> *library ("tidyverse")*

> *data = starwars %>% drop\_na(height) %>% drop\_na(homeworld) %>% replace\_na(list(sex = "none"))*

b) *Com vam veure a classe, un mapa d'arbre és un dibuix rectangular dividit en caselles, i cada casella representa una sola observació. Vam veure que era una bona manera de mostrar dades jeràrquiques mitjançant rectangles imbricats. I l'àrea relativa de cada casella expressava una variable contínua. També vam veure que era òptim quan hi ha com a màxim dues variables d'agrupació, per tant no en definirem més.*

*Per tant, una possibilitat amb aquest dataset seria:*

- *Per definir el color i actuar doncs com un 'grup pare' utilitzaríem la variable "sex" (Agafem aquesta com grup de color, ja que sol té 4 nivells ("male", "female", "hermaphrodite" i "none" i serà fàcil entendre la llegenda, que si posem "homeworld" que en té més).*
- *Com a 'subgrup' utilitzaríem la variable que denota el món natal ("homeworld", que té més nivells que "sex").*
- *Com a variable que descrigui l'àrea de les caselles triarem "height". Aquesta tria ve donada perquè ens estan preguntant sobre la massa dels personatges. I de totes les variables que ens pregunten, és l'única numèrica.*
- *I com a 'label'/nivell, escollim el nom ("name").*

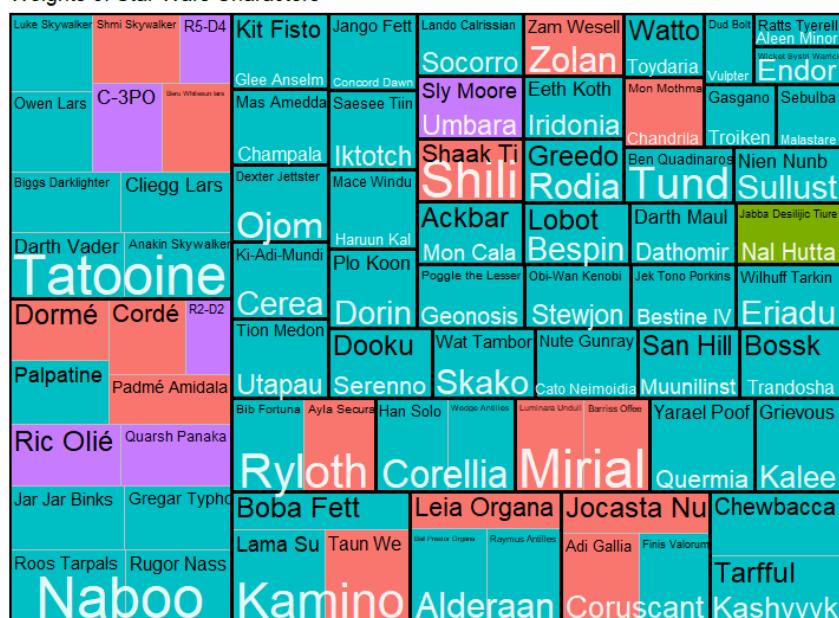
*Si ho fem amb R, el primer que hem de fer és carregar les llibreries necessàries. La llibreria específica aquí és treemapify (abans necessitarem tenir instal·lat el paquet com vam veure al seminari corresponent). També haurem de fer ús de geom\_treemap.*

*Per fer un treemap basic doncs:*

```
> library ("treemapify")
```

```
> visualitzacio= ggplot(data, aes(area=height, fill=sex, label=name,
  subgroup=homeworld))+geom_treemap()+ geom_treemap_subgroup_border(colour =
  "black",size=3) +geom_treemap_subgroup_text(alpha=0.9,colour
  ="white")+geom_treemap_text (aes(label=name))+labs(title = "Weights of Star
  Wars Characters")
```

Weights of Star Wars Characters

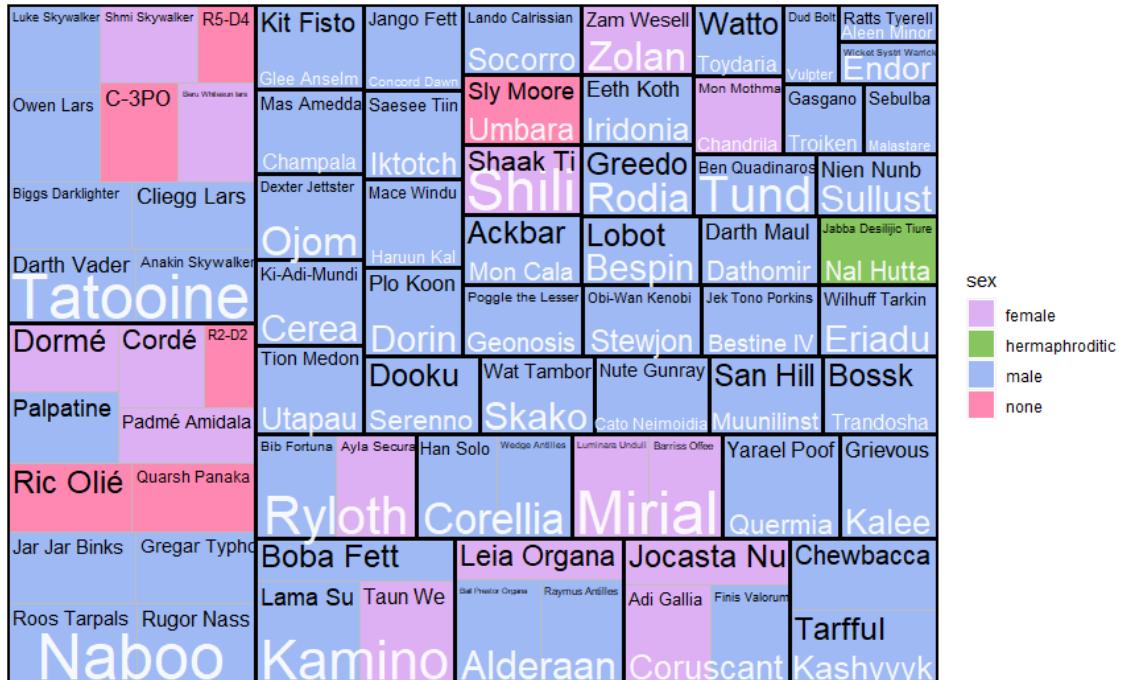


sex  
female  
hermaphroditic  
male  
none

*c) Fixeu-vos que hem posat el color amb fill*

```
>visualitzacio+scale_fill_manual(values=c('#DCB0F2','#87C55F','#9EB9F3','#FE88B1'
,'#C9DB74'))
```

Weights of Star Wars Characters



d) El personatge més baix de Tatooine és el R5-D4 i no té sexe.

Només hi ha un personatge hermafrodita i és de Nal Hutta, és en Jabba Desilijic Tiure

e) Quan s'especifiquen les proporcions d'acord amb múltiples variables d'agrupació, les representacions de mosaics, els mapes d'arbres o els conjunts paral·lels (parallel sets) són aproximacions de visualització útils. Ara bé, els mosaics assumeixen que tots els nivells d'una variable d'agrupament es poden combinar amb tots els nivells d'una altra variable d'agrupació, mentre que els mapes d'arbres no fan aquesta suposició.

Els mapes d'arbres com el que hem fet, funcionen bé fins i tot si les subdivisions d'un grup són completament diferents de les subdivisions d'un altre.

Finalment, els conjunts paral·lels funcionen millor que les trames de mosaic o els mapes d'arbres quan hi ha més de dues variables d'agrupació.

2.2. (1 pt) Voleu saber quines són les components principals de les variables numerades a sota del dataset *statsNBA2008.csv*. Podeu fer-ho en qualsevol llenguatge, però expliqueu perquè heu escollit la visualització que heu escollit, els passos que heu seguit per fer-la, mostreu-la, i extreieu alguna/es conclusió/ns del vostre gràfic.

- PTS (percentatge de punts de l'equip, punts per joc)
- FGM (cistelles de camp realitzats, percentatge de cistelles de camp)
- FTP (percentatge de tirs lliures)3PM (cistelles de camp de 3 punts anotats, tirs de 3 punts)
- 3PA (rebots defensius, tirs de tres punts)

- 3PP (percentatge de tirs de camp de 3 punts)

**RESPOSTA:**

*Llegim dades i carreguem llibreries:*

```
> NBA=read_csv('statsNBA2008.csv')
```

*Primer ens preparam el dataframe que ens interessa, vam veure al seminari 5 les tibles de R i com utilitzar-les. Una manera que es podria fer servir, seria*

```
> NBA_tb <- as_tibble(NBA)
> NBA_interes<- NBA_tb[,c('PTS','FGM','FTP','3PM','3PA','3PP')]
```

*També ho podríem fer utilitzant select de dplyr com havíem vist en classes anteriors. Assignem a una nova variable anomenada NBA\_pca la matriu resultant d'aplicar la funció prcomp() vista en la primera part de la classe d'avui. Centrem les variables a zero utilitzant l'argument 'center=TRUE', i les escalem com al seminari per a que tinguin variança igual a 1, fent us de l'argument 'scale.=TRUE'.*

```
> NBA_pca<-prcomp(NBA_interes, center=TRUE, scale.=TRUE)
```

*Podem per exemple mostrar un biplot*

```
> fviz_pca_biplot (NBA_pca, geom.ind="point", geom.var = c("arrow",
"text"), addEllipses = TRUE, Legend.tittle="Groups")
```

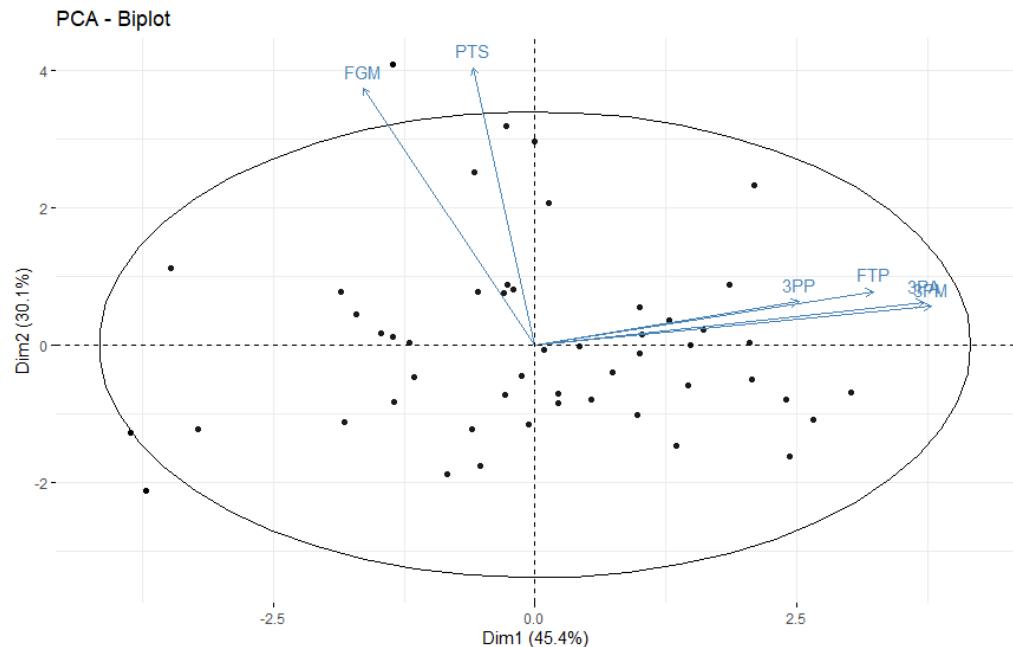


la app para encontrar curro este verano  
con más de 5000 ofertas de empleo.



descarga la ap

descarga randstad app y empieza hoy.



- PC1 correspon al 45.4% de la variança total.
- PC2 correspon al 30.1% de la variança total.

Per tant, coneixent la posició d'una mostra en relació amb (només) les components PC1 i PC2, podem tenir una visió molt precisa de la seva ubicació en relació amb altres mostres. Això és degut a que només PC1 i PC2 poden explicar el 70.5% de la variança.

### PART 3 (4 pt)

Dataset: Suicide rates 1985-2016.csv

Agafarem el dataset de noms del nombre de suïcidis en diferents països entre els anys 1985 i 2016 segmentat per diferents variables com sexe, edat, generació, població, etc. Utilitzeu les llibreries (plotly, gganimate, shiny, etc.) que creieu convenientes i dibuixeu les gràfiques que us facin falta.

**RESPOSTA:**

```
> library(tidyverse)
> library(dplyr)
> library(plotly)
> library(shiny)

> getwd()
> setwd("C:/Users/enric/Documents/R")
> SWorld <- read.csv('./Suicide rates 1985-2016.csv')
> str(SWorld)
# Renombrar alguns camps amb caracters erronis (també es pot utilitzar la comanda rename):
> names(SWorld)[1] <- "country"
> names(SWorld)[10] <- "gdp_for_year"
> names(SWorld)[11] <- "gdp_per_capita"
```



```
> str(SWorld)
```

**3.1. (1 pt) Mostra el codi i una gràfica d'un scatter plot 3D sobre el nombre de suïcidis per a cada 100k, població i PIB per càpita (gdp\_per\_capita).**

**RESPOSTA:**

PAS 1: DATA MASSAGING: Seleccionar variables i agrupar per generació i any

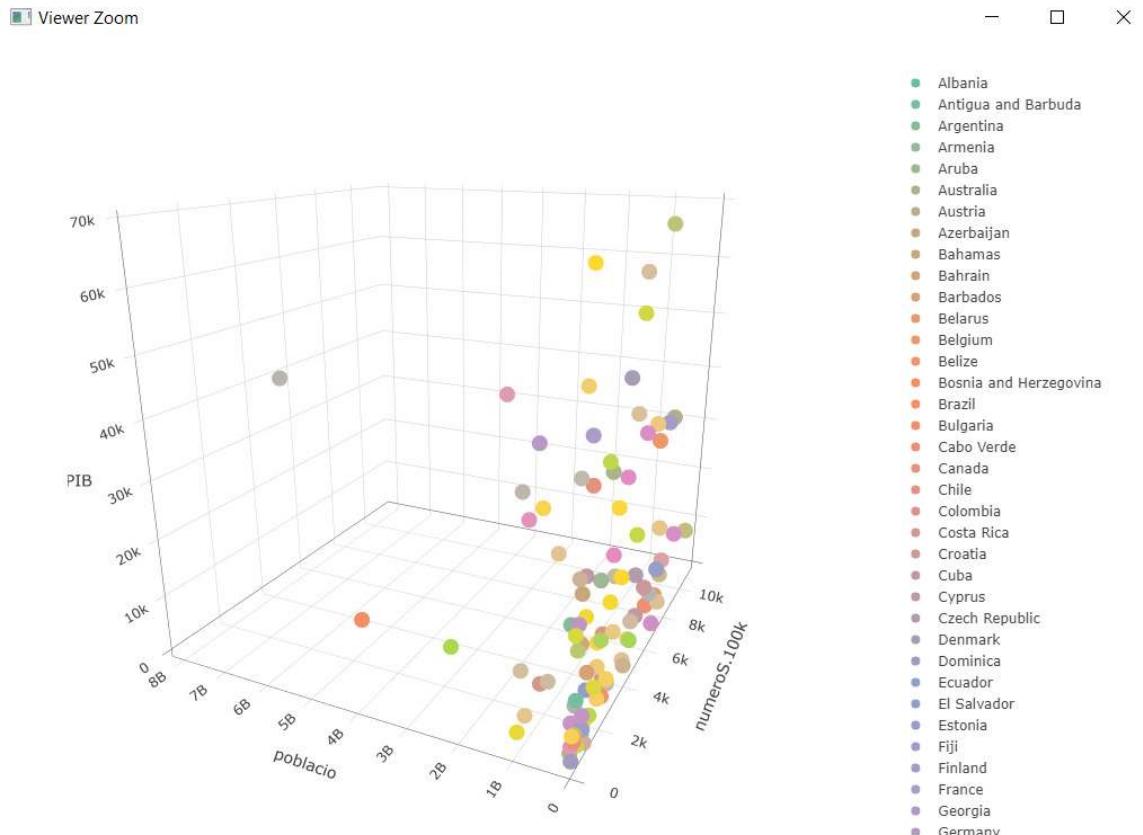
```
> SWorldPIB <- SWorld %>% select(country, year, suicides.100k.pop, population, gdp_per_capita) %>% group_by(country) %>% summarise(numeroS.100k = sum(suicides.100k.pop), poblacio=sum(population), PIB=mean(gdp_per_capita))
```

O bé...

```
SWorldPIB <- SWorld %>% group_by(country) %>% summarise(numeroS.100k = sum(suicides.100k.pop), poblacio=sum(population), PIB=mean(gdp_per_capita))
```

# PAS 2: GRAFICA DE SCATTER 3D AMB plotly

```
> plot_ly(SWorldPIB, x=~numeroS.100k, y=~poblacio, z=~PIB, color=~country, type='scatter3d', mode='markers')
```



**3.2. (0,5 pts): D'aquesta gràfica, obteniu la següent informació:**

- Dades (país, suicidis.100k.pop, població i PIB per càpita) dels 4 països amb més població. Trobes a faltar algun país?. Màxim 2.

**RESPOSTA:**

PAÍS	SUICIDIS.100k.pop	Població	PIB per càpita
USA	5.140,97	80.540.27B	39.269,61
Brazil	2.174,72	48.560.94B	6.091,484
Russian Federation	11.305,13	36.908.03B	6.518,815
Japan	8.025,23	36.810.25B	36.397,55

- b) Dades (país, suicidis.110k.pop, població i PIB per càpita) del país amb més PIB per càpita.

**RESPOSTA:**

PAÍS	SUICIDIS.100k.pop	Població	PIB per càpita
Luxemburg	6.156,56	12.953.16M	68.798,39

**3.3. (1 pt)** Mostra el codi i una gràfica de línies de l'evolució del nombre de suïcidis de DONES per anys i generació. Digues quines generacions són capdavanteres i en quin període d'anys ho són.

**RESPOSTA:**

PAS 1: DATA MASSAGING: Seleccionar variables i agrupar per generació i any

```
> SWorldGD <- SWorld %>% filter(sex == "female") %>% select(generation, year, suicides_no, population, gdp_per_capita) %>% group_by(generation, year) %>% summarise(numeroS = sum(suicides_no), poblacio=sum(population), PIB=mean(gdp_per_capita))
```

O bé...

```
SWorldGD <- SWorld %>% filter(sex == "female") %>% group_by(generation, year) %>% summarise(numeroS = sum(suicides_no), poblacio=sum(population), PIB=mean(gdp_per_capita))
```

PAS 2a: GRAFICA DE LINIES AMB ggplotly

```
> ggplotWGD <- ggplot(SWorldGD, aes(x=year, y=numeroS, color=generation)) + geom_line()
> ggplotly(ggplotWGD)
```

PAS 2b: GRAFICA DE LINIES AMB plotly

```
> plot_ly(SWorldGD, x=~year, y=~numeroS, color=~generation) %>% add_lines()
```

PAS 2c: GRÀFICA DE LINIES AMB shiny

```
ui <- fluidPage( selectizeInput( inputId = "generat",
                                label = "Selecciona una Generació:",
                                choices = unique(SWorldGD$generation),
                                selected = c("Silent"),
```

**RESPOSTA:**

POSICIÓ	GENERACIÓ	Nombre de Suïcidis
1	SILENT	21.620
2	BOOMERS	17.989
3	GENERATION X	7.254
4	MILLENIALS	5.876
5	GENERATION Z	585
6		

b) Nombre màxim de suïcidis per generació.

**RESPOSTA:**

GENERACIÓ	ANY	Nombre de Suïcidis
BOOMERS	1994	23.084
GENERATION X	2010	25.831
GENERATION Z	2014	761
G.I. GENERATION	1990	23.032
MILLENIALS	2011	13.905
SILENT	2002	22.605

4 (1 pt) Classifica les següents preguntes o mesures que s'utilitzen en tests d'Usabilitat o d'Experiència d'Usuari (UX) en les següents dues categories:

- a) Instrumentals / Usabilitat / Pragmàtiques
- b) No instrumentals / Hedònics / Emocionals

I digues en quin tipus de test s'inclou cada pregunta o mesura.

**PREGUNTES:**

1. *Conservative 1 2 3 4 5 6 7 Innovative*
2. *The design looks attractive*
3. *I found the various functions in this System very well integrated*
4. *This System has all the functions and capabilities I expect it to have*
5. *Complicated 1 2 3 4 5 6 7 Simple*
6. *Non-inclusive 1 2 3 4 5 6 7 Inclusive*
7. *I thought the System was easy to use*

8. *The product is stylish: Strongly Disagree 1 2 3 4 5 6 Strongly Agree*
9. *The product is creatively designed*
10. *It was easy to find the information I needed*

**RESPOSTA:**

PREGUNTA	CATEGORIA	TIPUS DE TEST
1	b)	Attrakdiff, Hedonic
2	b)	meCUE, A2, Aesthetic
3	a)	SUS
4	a)	PSSUQ
5	a)	Attrakdiff, Pragmatic
6	b)	Attrakdiff, Hedonic
7	a)	SUS
8	b)	meCUE, A3, Aesthetic
9	b)	meCUE, A1, Aesthetic
10	a)	PSSUQ

2025  
10-13 JULY



**Visualització de dades (Enginyeria de Dades - EE - UAB)**  
**Examen Segon Parcial - 19 Juny 2023**  
**RESPOSTES MODEL B**

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 (6 pt)

*Dataset: Australian\_Flight\_dataset\_2003-2022.csv*

Agafarem aquest dataset que conté totes les rutes (cada ruta conté un o més vols diaris, setmanals o mensuals) d'arribada i/o sortida als aeròports d'Austràlia durant els anys 2003 al 2022. Utilitzeu les llibreries (plotly, gganimate, shiny, etc.) que creieu convenient i dibuixeu les gràfiques que us facin falta.

Cada registre conté informació d'una ruta d'una companyia aèria. Conté les variables:

1. Month ---> Date in month and day
2. In\_Out ---> Status of flight [Incoming/Outgoing]
3. Australian\_City ---> Australian city name
4. International\_City ---> International city name
5. Airline ---> Airline owning the flight
6. Route ---> Route taken by the flight
7. Port\_Country ---> Port country
8. Port\_Region ---> Port region
9. Service\_Country ---> Service country
10. Service\_Region ---> Service region
11. Stops ---> Number of stops taken by the flight
12. All\_Flights ---> Total number of flights
13. Max\_Seats ---> Total capacity of seats in flight
14. Year ---> Date in year
15. Month\_num ---> Date in month number

Com a principals variables que utilitzarem en aquesta 1a part, cal ressaltar Airline, Max\_Seats i Year.

Càrrega de les llibreries i data set:

```
> library(tidyverse)
> library(dplyr)
> library(plotly)
> library(shiny)
> getwd()
> setwd("C:/Users/enric/Documents/R")
```

```
> Australian_FD <- read.csv('./Australian_Flight_Dataset_2003-2022.csv')
```

**1.1 (2 pts)** Mostra el codi (en ggplotly) i la gràfica de línies de l'evolució temporal en anys de la capacitat en seients de cada companyia, seguin de rutes de sortida o arribada a qualsevol ciutat australiana (gràfica 1) i fixa't amb les tres companyies aèries amb més seients l'any 2021. Digues quines companyies són per ordre decreixent i mostra el codi (en ggplotly o shiny+plotly) i la gràfica de línies de l'evolució temporal d'aquestes tres companyies (gràfica 2). Especifica si fas Data Massaging.

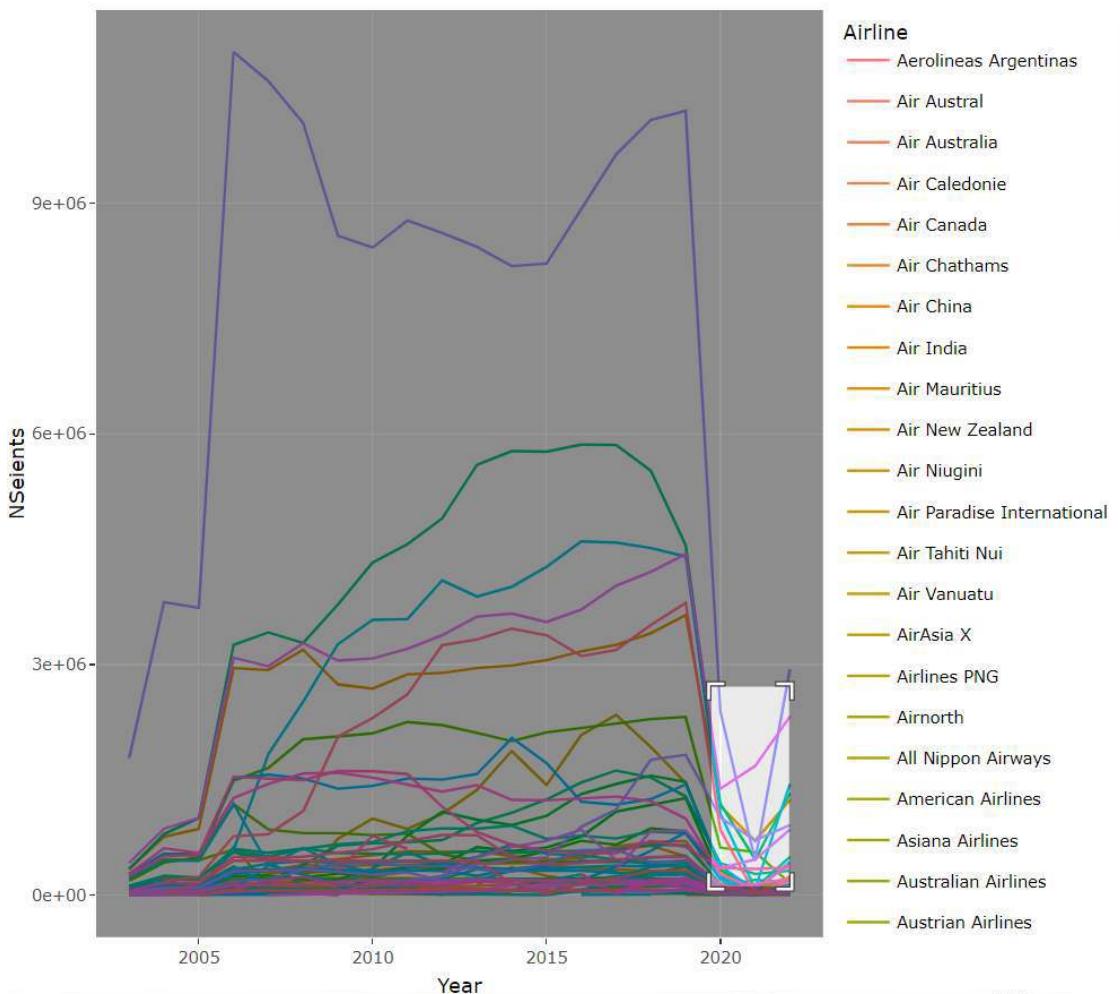
**RESPOSTA:** En primer lloc, hem de fer data massaging sobre el data set original.

**DATA MASSAGING:** Agrupar per (Any, Companyia Aèrea, Sexe) calculant el nombre de seients (Max\_Seats):

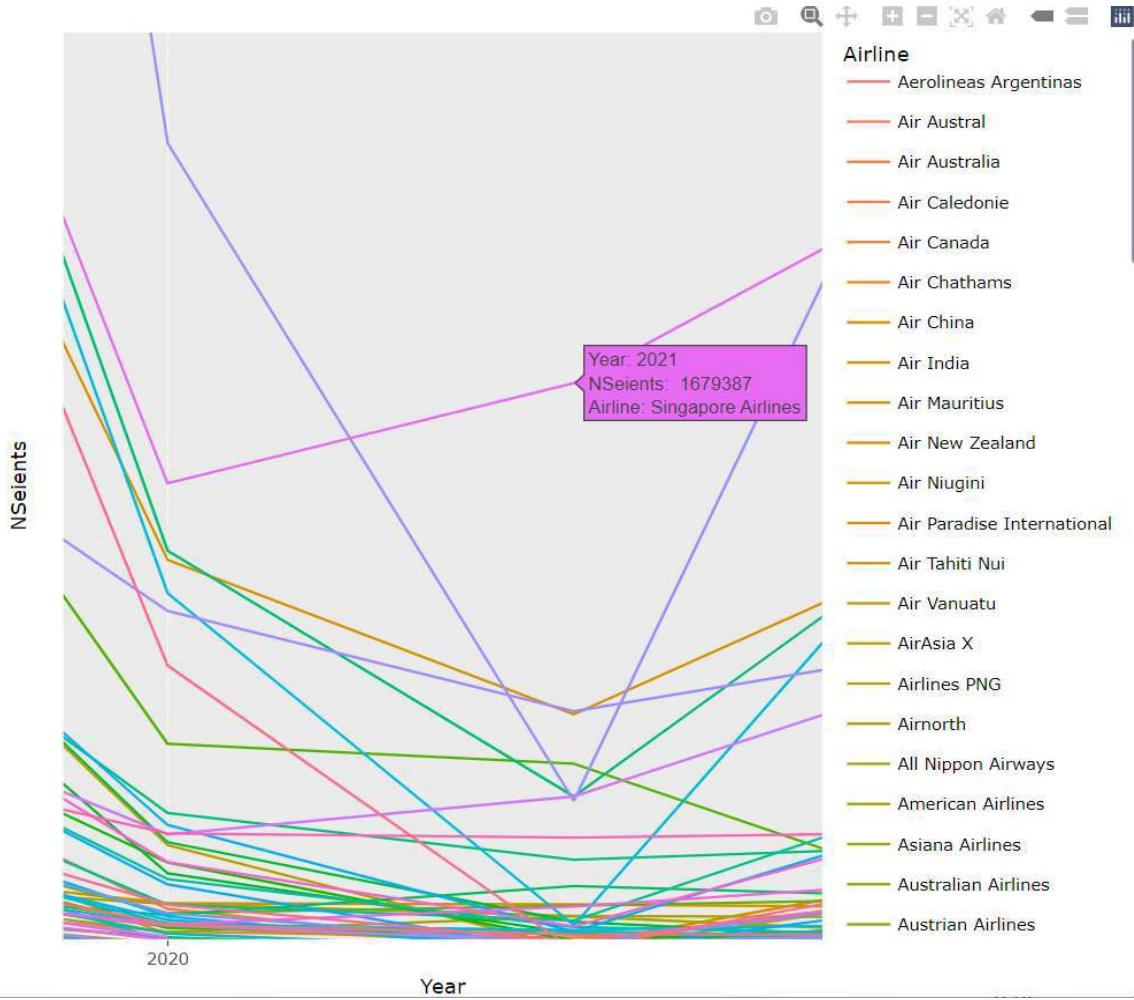
```
> Australian_FD_Rutes <- Australian_FD %>% group_by(Year,Airline) %>%  
summarize(NSeients = sum(Max_Seats))
```

**GRÀFICA 1 de línies amb ggplotly:**

```
> ggplotR <- ggplot(Australian_FD_Rutes, aes(x=Year, y=NSeients,  
color=Airline)) + geom_line()  
> ggplotly(ggplotR)
```



I fer un zoom en la zona propera a l'any 2021:



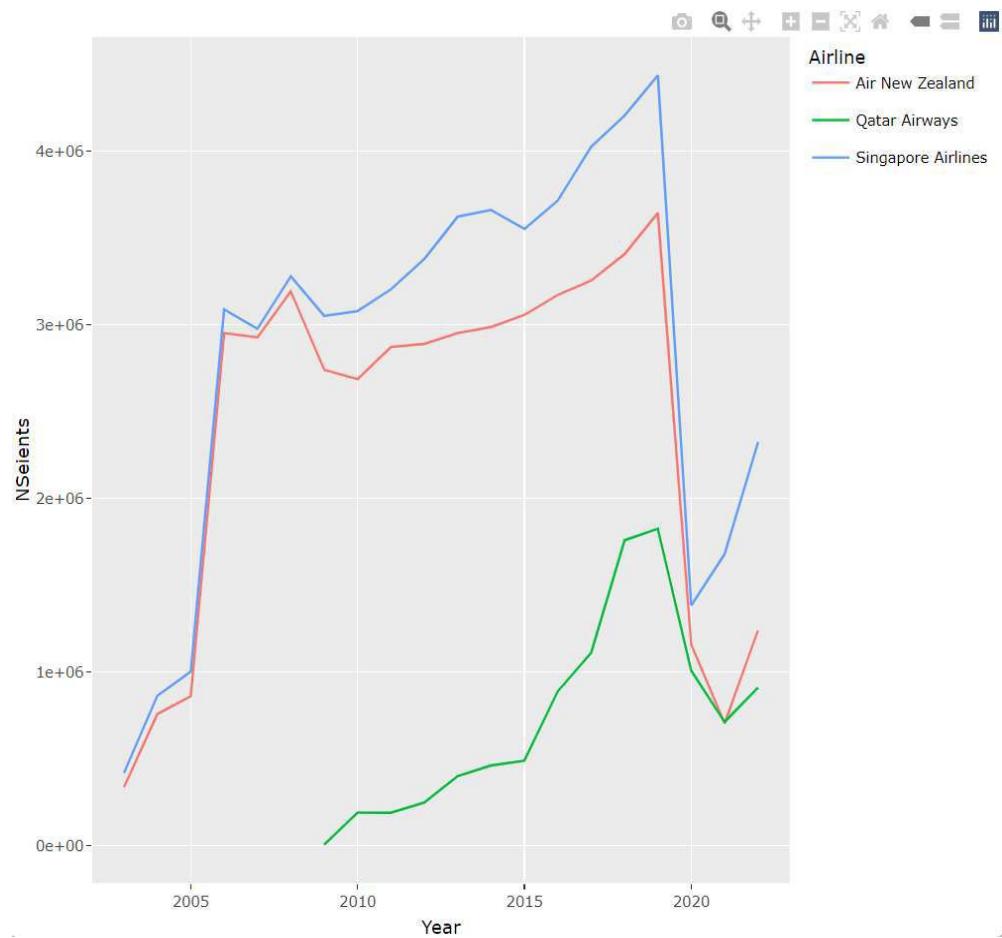
Els noms de les companyies són **Singapore Airlines (1679387 seients), Qatar Airways (712357) i Air New Zealand (703436)**

**GRÀFICA 2:** Després podem llistar la gràfica de dues formes:

**AMB ggplotly (abans filtrem el data set per les tres companyies aèries):**

```
> Australian_FD_Seients_SQA <- Australian_FD_Seients %>% filter(Airline == "Singapore Airlines" | Airline == "Qatar Airways" | Airline == "Air New Zealand")  
  
> ggplotSQA <- ggplot(Australian_FD_Seients_SQA, aes(x=Year, y=NSeients, color=Airline)) + geom_line()  
> ggplotly(ggplotSQA)  
>
```

2025  
10-13 JULY



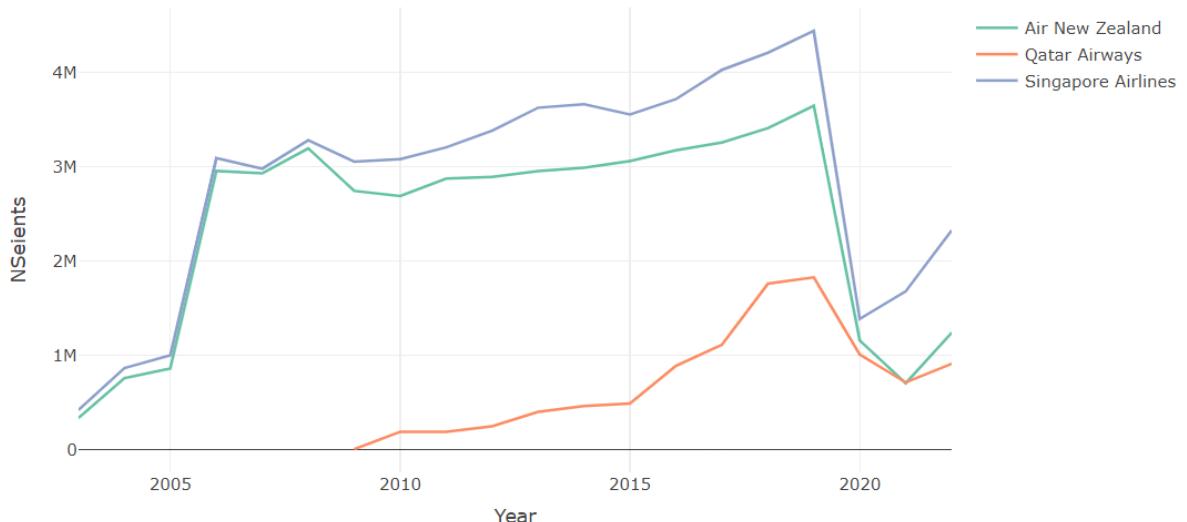
### AMB shiny + plotly:

```
ui <- fluidPage(
  selectizeInput( inputId = "Airline",
  label = "Selecciona una COMPANYIA:",
  choices = unique(Australian_FD_Seients_SQA$Airline),
  selected = c("Singapore Airlines","Qatar Airways",
  "Air New Zealand"),
  multiple = TRUE
  ),
  plotlyOutput(outputId = "p")
)

server <- function(input, output, ...){
  output$p <- renderPlotly (
    plot_ly(Australian_FD_Seients_SQA, x = ~Year, y = ~NSeients,
color=~Airline) %>%
      filter(Airline %in% input$Airline) %>%
      group_by(Airline) %>%
      add_lines()
  )
}
shinyApp(ui, server)
```

Selecciona una COMPANYIA:

Air New Zealand  Singapore Airlines  
 Qatar Airways



1.2 (0,5 pt) Compara les tres companyies en 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
Singapore Airlines	2003 – 418696	2019 – 4436641
Qatar Airways	2009 – 5698	2019 – 1824887
Air New Zealand	2003 – 335642	2019 – 3644470

- b) (0,1 pts) Comenta la taula anterior. Trobes alguna singularitat?.

**RESPOSTA:** El màxim nombre de rutes de les companyies coincideix el 2019, poc abans la pandèmia del Covid-19.

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

**RESPOSTA:**

NOMS	Any Anterior	Any Posterior
Air New Zealand-Qatar Airways	2020	2021
Qatar Airways-Air New Zealand	2021	2022

**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 les 10 companyies amb més seients per anys, del 2003 al 2020.**

**RESPOSTA:**

*Càrrega de les llibreries*

```
> library(tidyverse)
> library(gganimate)      # Generació de frames i la seva compilació per a
generar fitxer animació
> library(gifski)         # Generació fitxer animació GIF
```

**PAS1: DATA MASSAGING:**

- *Filtrar el dataset pels anys 2003 al 2020:*

```
> Australian_FD_Seients <- Australian_FD_Seients %>% filter(Year < 2021)
```

- *Definir nova variable rank amb la posició decreixent en nombre de rutes i filtrar les 10 primeres companyies segons rank:*

```
> Australian_FD_Seients_formatted <- Australian_FD_Seients %>%
group_by(Year) %>%
# The * 1 makes it possible to have non-integer ranks while sliding
mutate(rank = rank(-NSeients)) %>%
group_by(Airline) %>%
filter(rank <=10) %>%
ungroup()
```

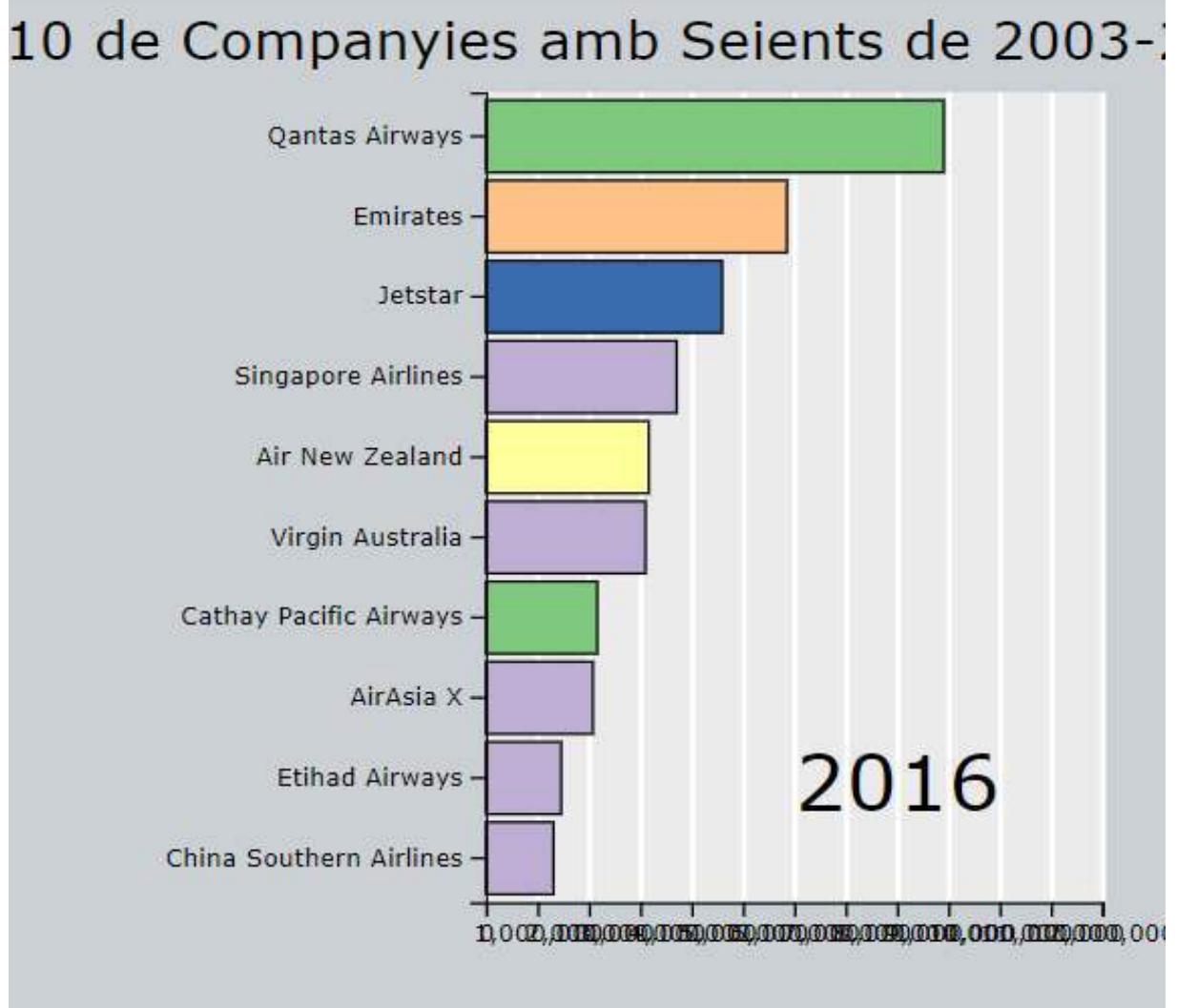
**PAS 2: ANIMATED BAR RACE RANKING:**

```
> anim <- ggplot(Australian_FD_Seients_formatted, aes(rank, group = Airline,
           fill = as.factor(Airline), color = as.factor(Airline))) +
geom_tile(aes(y = NSeients/2,
           height = NSeients,
           width = 0.9), alpha = 0.8, color = "black") +
width = 0.9), alpha = 0.8, color = NA) +
geom_text(aes(y = 0, label = paste(Airline, " ")), vjust = 0.2, hjust = 1) +
geom_text(aes(y=NSeients,label = NSeients, hjust=0)) +
coord_flip(clip = "off", expand = FALSE) +
scale_y_continuous(labels = scales::comma) +
scale_x_reverse() +
guides(color = "none", fill = "none") +
theme(axis.line=element_blank(),
axis.text.x=element_blank(),
axis.text.y=element_blank(),
axis.ticks=element_blank(),
axis.title.x=element_blank(),
axis.title.y=element_blank(),
legend.position="none",
panel.background=element_blank(),
panel.border=element_blank(),
panel.grid.major=element_blank(),
panel.grid.minor=element_blank(),
panel.grid.major.x = element_line( size=.1, color="grey" ),
panel.grid.minor.x = element_line( size=.1, color="grey" ),
```

**ANIMATED BAR RACE RANKING. ALTERNATIVA** (*no available, doncs no s'ha explicat, però es posa per a que ho coneixeu*): Llibreria ddplot que defineix una gràfica de forma més senzilla i clara

```
> install.packages("ddplot") # Càrrega paquet ddplot (primera vegada)
> library(ddplot)           # Generació de Racing Bar Charts

> Australian_FD_Seients_formatted %>%
  barChartRace(
    x = "NSeients",
    y = "Airline",
    time = "Year",
    title = "Top 10 de Companies amb Seients de 2003-2022"
  )
```



No hem trobat encara la forma de generar fitxer de vídeo amb aquesta funció.

1.4 (1 pt) Classifica les següents preguntes o mesures que s'utilitzen en tests d'Usabilitat o d'Experiència d'Usuari (UX) en les següents dues categories:

- a) Instrumentals / Usabilitat / Pragmàtiques
- b) No instrumentals / Hedònics / Emocionals

**PREGUNTES:**

1. *Conservative 1 2 3 4 5 6 7 Innovative*
2. *Non-inclusive 1 2 3 4 5 6 7 Inclusive*
3. *This System has all the functions and capabilities I expect it to have*
4. *Complicated 1 2 3 4 5 6 7 Simple*
5. *I thought the System was easy to use*
6. *The product is stylish: Strongly Disagree 2 3 4 5 6 Strongly Agree*
7. *The design looks attractive*
8. *I found the various functions in this System very well integrated*
9. *It was easy to find the information I needed*
10. *The product is creatively designed*

**RESPOSTA:**

PREGUNTA	CATEGORIA	TIPUS DE TEST
1	b)	Attrakdiff, Hedonic
2	b)	Attrakdiff, Hedonic
3	a)	PSSUQ
4	a)	Attrakdiff, Pragmatic
5	a)	SUS
6	b)	meCUE, A3, Aesthetic
7	b)	meCUE, A2, Aesthetic
8	a)	SUS
9	a)	PSSUQ
10	b)	meCUE, A1, Aesthetic

1.5 (1 pt) Defineix 4 dels següents conceptes en animació, interactivitat, usabilitat i Experiència d'Usuari en Visualització de Dades:

- Usabilitat
- *plotly\_click*
- *Utility*
- *Toggle*
- *Efficiency*

2025  
10-13 JULY



- *Control*

**RESPOSTA:**

## PART 2 (4 pt)

*Dataset: Australian\_Flight\_dataset\_2003-2022.csv*

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

**2.1 (2 pt)** Fes un bubble map per representar el nombre de rutes que van sortir l'any 2018 de cadascuna de les següents ciutats d'Austràlia: "Adelaide", "Brisbane", "Cairns", "Darwin", "Melbourne", "Perth", "Sydney".

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

```
~ Australian_City, ~lat, ~lon,  
"Sydney", -33.8688, 151.2093,  
"Melbourne", -37.8136, 144.9631,  
"Brisbane", -27.4698, 153.0251,  
"Adelaide", -34.9285, 138.6007,  
"Perth", -31.9505, 115.8605,  
"Cairns", -16.9230, 145.7662,  
"Darwin", -12.4612, 130.8418,
```

Aquesta *tibble* l'hauràs d'ajuntar amb un dataframe construït contenint el nombre de rutes per ciutat de l'any 2018. Per ajuntar ambdós dataframes pots usar la comanda de R:

- ```
> inner_join (nom_tibble, nom_dataframe, by="nom_variable_compartida")
```
- Indica el datamassage necessari (0,25 pts) .
  - Feu un bubble map d'Austràlia , que doni informació sobre el nombre de rutes per cada ciutat. Podeu usar el mapa oz\_states que vam usar al seminari. Al dibuixar-lo treu la graella del fons (1,5 pts).

**RESPOSTA:**

**(a) Feu la tibble**

```
> library(tidyverse)  
> library(dplyr)  
> oz_cities <- tribble::tribble(  
~ Australian_City, ~lat, ~lon,  
"Sydney", -33.8688, 151.2093,  
"Melbourne", -37.8136, 144.9631,  
"Brisbane", -27.4698, 153.0251,  
"Adelaide", -34.9285, 138.6007,  
"Perth", -31.9505, 115.8605,  
"Cairns", -16.9230, 145.7662,  
"Darwin", -12.4612, 130.8418, )
```

Guardant només les ciutats de les quals us hem donat la longitud i latitud, construïu un dataframe, que contingui el nombre de rutes que surten (*In\_Out=0*) de cada ciutat de l'any 2018. Un cop el tingueu, ajunteu-lo amb el dataframe creat amb dites latituds i longituds. Per això últim, si ho feu en R, podeu utilitzar la comanda *inner\_join* donada:

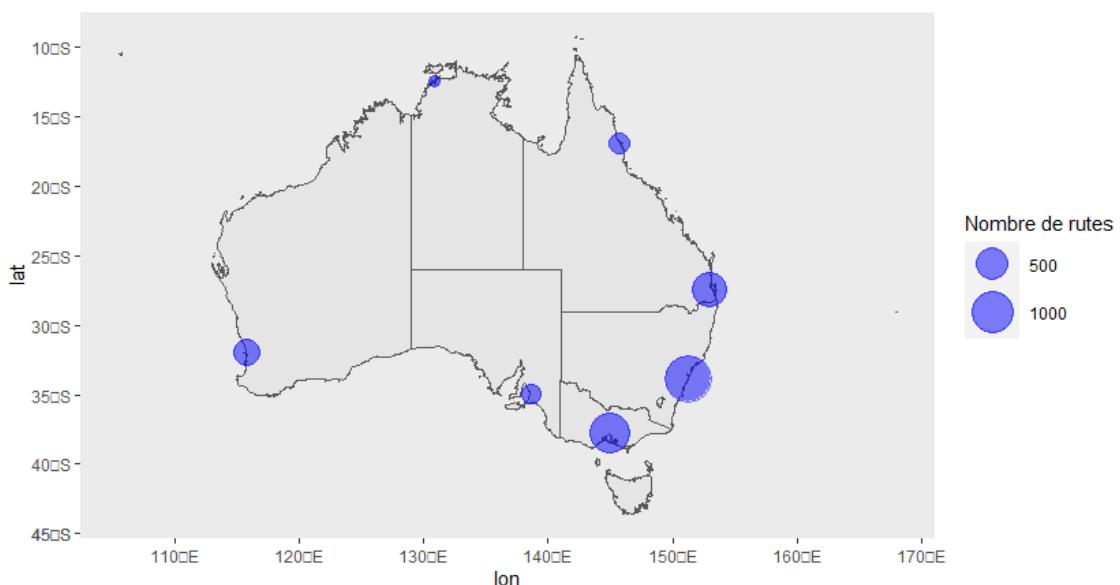
```
> Australian_FD_Rutes2018<- Australian_FD%>%filter(Australian_City %in%
c("Adelaide", "Brisbane", "Cairns", "Darwin", "Melbourne", "Perth", "Sydney")) %>%
filter (Year==2018) %>%filter(In_Out=="0")%>% group_by(Australian_City) %>%
summarize(NRutes = n())

> dades<-inner_join(Australian_FD_Rutes2018,oz_cities,by="Australian_City")
```

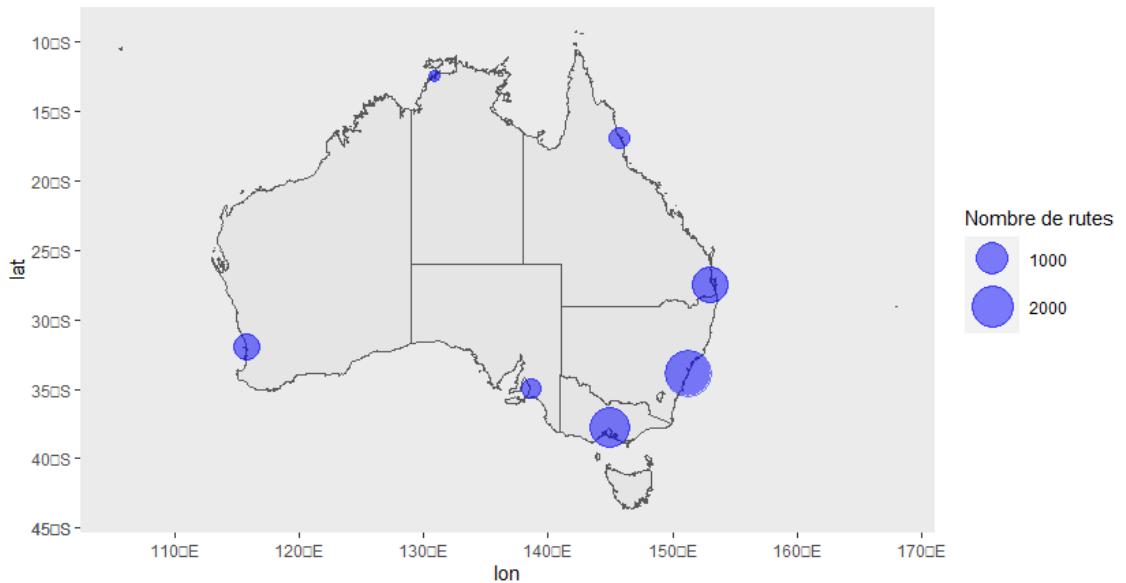
b) Sobre el mapa de oz\_states Dibuixeu tot traient la graella del fons. Feu un bubble map on la talla de les bombolles creix amb el nombre de rutes per ciutat

```
> library(ozmaps)
> library(sf)
> oz_states <- ozmaps::ozmap_states

> ggplot() + geom_sf(data = oz_states)+geom_point(data = dades, mapping =
aes(x = lon, y = lat,size=NRutes), alpha=0.5,
color="blue")+coord_sf()+theme(panel.grid.major = element_blank(),
panel.grid.minor = element_blank()) +scale_size_continuous(range=c(3,15),
"Nombre de rutes")
```



També es compta bé si no s'ha especificat el filtre per agafar només les rutes de sortida. En aquest cas el mapa seria:



**2.2 (2 pt)** Mostra un mapa d'arbre (treemap) que et permeti saber des de quina de les dues ciutats, Perth o Sidney, van volar més vols l'any 2018 a cada regió de servei (Service\_Region).

- Fes el datamassage necessari (0.5 pt).
- Argumenta com és un *treemap* en general i els passos que has de fer per construir la visualització d'aquest exercici (és a dir, quina variable utilitzes per l'àrea de les gralles, variables d'agrupació, etc). Inclou el codi si el fas en R (1 pt).
- Mostra el mapa d'arbre, posa llegenda i títol i posa una paleta de color (en R podeu usar una paleta manual, per exemple amb els colors c(' #87C55F ', '#9EB9F3 ') , o una altra al vostre gust). Quina és/són la/les regions de destí on l'any 2018 no van volar vols des de Perth? (0.5 pt).

**RESPOSTA:**

a)

> *library ("tidyverse")*

```
> Aust_FD_regiodesti2018<- Australian_FD%>%filter(Australian_City %in%
c("Perth", "Sydney")) %>% filter (Year==2018) %>%filter(In_Out=="0")
```

(Com abans, no es penalitza si no es fa el *filter(In\_Out=="0")*)

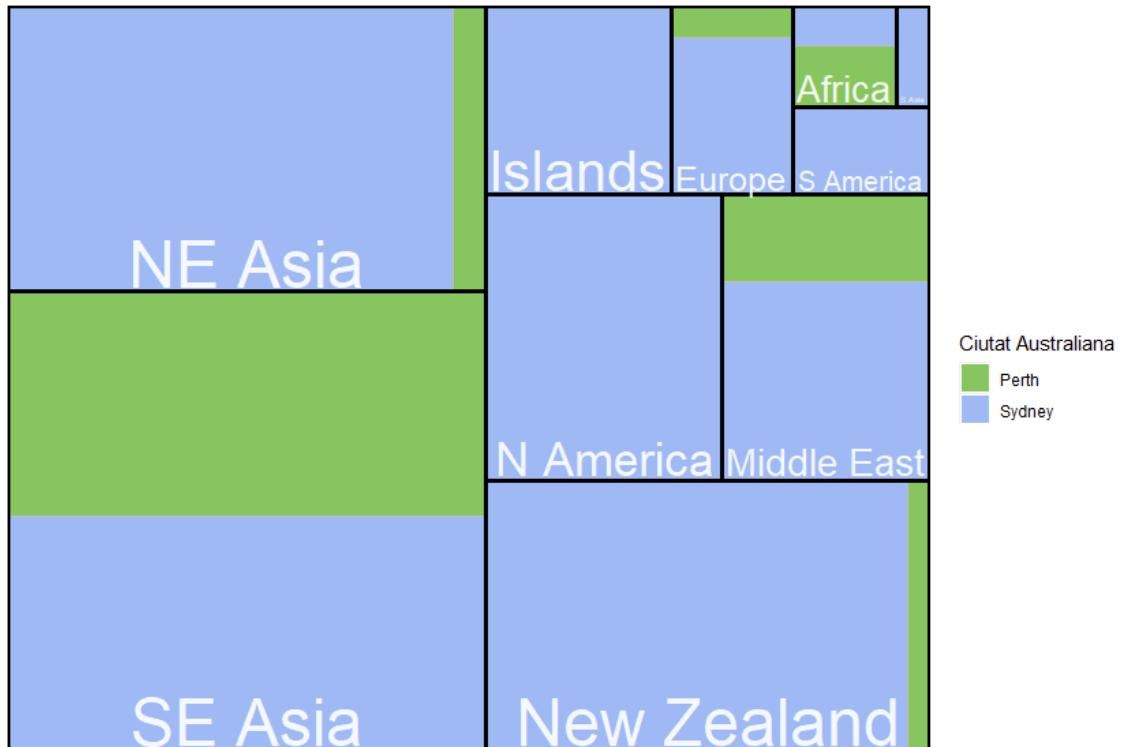
b) Com vam veure a classe, un mapa d'arbre és un dibuix rectangular dividit en caselles, i cada casella representa una sola observació. Vam veure que era una bona manera de mostrar dades jeràrquiques mitjançant rectangles imbricats. I l'àrea relativa de cada casella expressava una variable contínua. També vam veure que era òptim quan hi ha com a màxim dues variables d'agrupació, per tant no en definirem més.

Podeu a més ordenar la visualització:

```
>Aust_FD_regiodesti2018<-Aust_FD_regiodesti2018%>%group_by(Australian_City,  
Service_Region)%>%summarise(All_Flights=sum(ALL_Flights))
```

Us sortirà un warning que no té major importància, però que podeu treure (veure: <https://statisticsglobe.com/dplyr-message-summarise-has-grouped-output-r>)

Regions on es podia volar des de Perth i Sidney al 2018



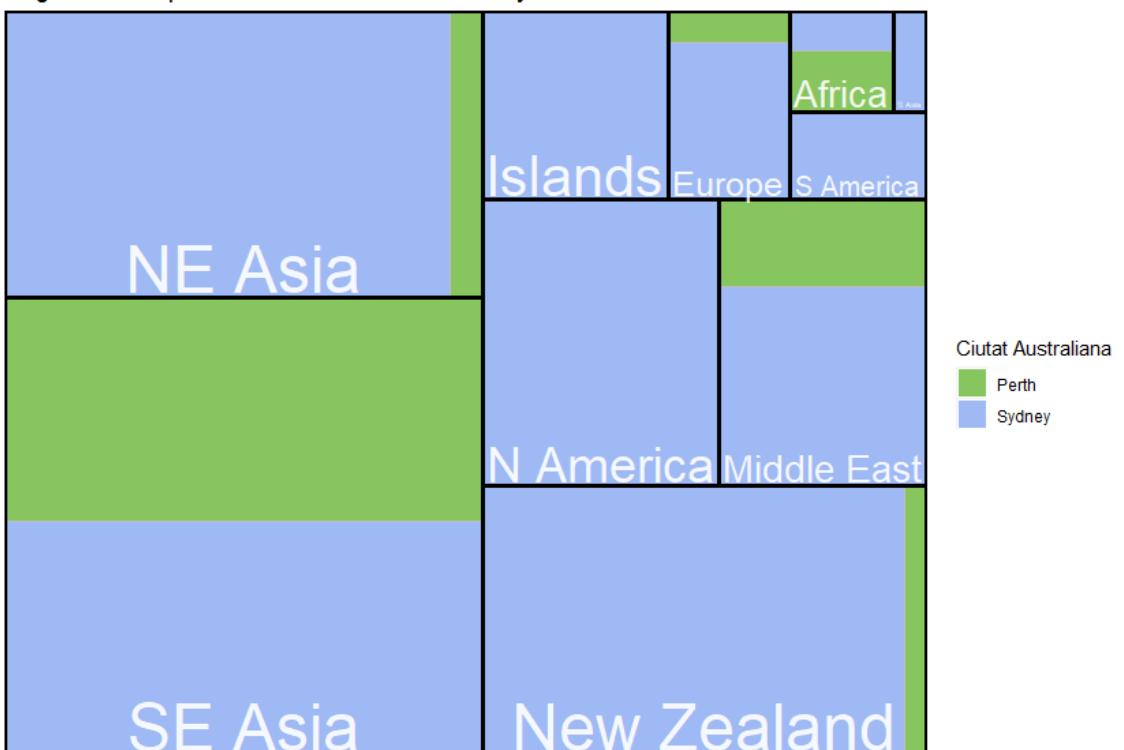
I si no es fa el `filter(In_out=="0")`:

Regions on es podia volar des de Perth i Sidney al 2018



O ordenant com abans:

Regions on es podia volar des de Perth i Sidney al 2018



En qualsevol cas, all 2018 no hi va haver vols des de Perth a les regions: d'Amèrica del Nord (N America), Amèrica del Sud (S America), Islands i Àsia del Sud (S Asia)

2025

10-13 JULY



estás a un escaneo

de encontrar curro.



descarga la



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

Examen de Pràctiques - 16 Abril 2021

MODEL 1

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

NIU: \_\_\_\_\_

**Dataset:** *binge\_drinking.csv*

1 (0,25 pt) Crear un dataframe que solo contenga observaciones a nivel de condado ('county') y asignarlo a una variable ('County\_DATA'). Crear un nuevo dataset que contenga solo la tasa/prevalencia media de consumición excesiva a nivel de estado ('state') de las mujeres durante el año 2010.

**RESPOSTA:** Resposta

descarga randstad app y empieza hoy.



estás a un escaneo

de encontrar curro.



descarga la

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

Examen Recuperació Segon Parcial - 5 Juliol 2021

SOLUCIONS MODEL A

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.

### PARTE 1 (2.5 pt)

1.1. (0.5 pt) Explica brevemente qué es una escala de color perceptualmente correcta y qué ventajas tiene.

#### RESPOSTA:

Son escalas que tienen en cuenta cómo nuestro cerebro procesa el color y en las que los colores varían de forma gradual e uniforme. Una ventaja es que estas variaciones entre colores se asocian mejor a las variaciones graduales e uniformes entre los valores del dataset.

1.2. (1 pt) Imagina que tenemos un dataset de precipitaciones (lluvia) en Europa y queremos visualizarlo en un mapa. La precipitación es un atributo cuantitativo continuo con valores min=0 y max=10 mm/h. ¿Qué escala de color usarías?

Di qué tipo de escala es, justifica la elección de colores, y hazla en R.

#### RESPOSTA:

Una escala secuencial continua. Los colores más adecuados por semántica y significado podrían ser una gama de azules.

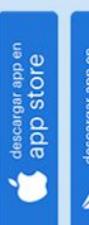
1.3. (1 pt) ¿Qué escala de color usarías para representar la velocidad en X del viento, teniendo en cuenta que tiene valores min=-15 y max=50 m/s?

Di qué tipo de escala es, justifica la elección de colores, y hazla en R o sube una imagen.

#### RESPOSTA:

Escala divergente o cuantitativa divergente. El color neutro (blanco) debería corresponder al 0 en los datos y habrá dos colores distintos en cada extremo.

descarga randstad app y empieza hoy.



## PART 2 (3.5 pt) Dataset:

*beers.csv*

Aquest dataset té 7 atributs per 44 cerveses de Minnesota i el farem servir en l'exercici 2.1. Els atributs són: Nom de la cerveseria, cervesa, descripció, estil, alcohol en volum (ABV), unitats internacionals d'amargor (IBU), puntuació que han rebut

Dataset: *statsNBA2008.csv*

Aquest dataset té 21 atributs de 50 jugadors de la NBA. Recull les estadístiques de la NBA del 2008. I estudiarem la relació entre alguns dels atributs en l'exercici 2.2

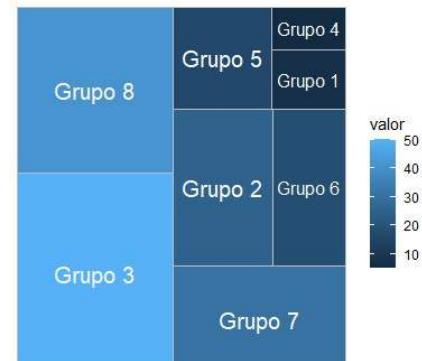
2.1 (2.25 pt) És estiu, esteu a Minnesota i voleu anar a prendre una cervesa. Utilitzant el dataset *beers.csv*.

- a) Feu un mapa d'arbre (*treemap*) que us permeti contestar les preguntes de l'apartat b. Podeu fer-lo en R o en qualsevol altre llenguatge, utilitzant les llibreries que us semblin convenient. Ara bé, heu d'explicar-ho bé: comenceu narrant com és un *treemap* en general i per quin tipus de dades serveix i redacteu els passos que heu de fer per construir la visualització triada. (1.5 pt)
- b) Esteu amb una persona que li puja molt l'alcohol i vol una cervesa amb poc alcohol per volum. Si esteu a la cerveseria Bauhaus, quin tipus de cervesa (Lager, Ale o IPA) li demaneu?. I si esteu en la cerveseria Summit, entre quines tres cerveses podeu triar per demanar-li (doneu el nom de les 3 cerveses)? (0.25 pt)
- c) Expliqueu quan usem un mosaic enlloc d'un gràfic de barres apilades, i quan no podem utilitzar un mosaic i usem un treemap. (0.25 pt)
- d) Quin tipus d'escala de color, de les de la pàgina següent, seria la més òptima i perquè? (0.25 pt) Paleta 1  
Paleta 2  
Paleta 3

Raona la resposta:



Paleta 1



Paleta 2



Paleta 3

### RESPOSTA:

a) Com vam veure a classe, un mapa d'arbre és un dibuix rectangular dividit en caselles, i cada casella representa una sola observació. Vam veure que era una bona manera de mostrar dades jeràrquiques mitjançant rectangles imbricats. I l'àrea relativa de cada casella expressava una variable contínua. També vam veure que era òptim quan hi ha com a màxim dues variables d'agrupació, per tant no en definirem més.

Quan s'especifiquen les proporcions d'acord amb múltiples variables d'agrupació, les representacions els mapes d'arbres són aproximacions de visualització útils. Per tant, una possibilitat amb aquest dataset seria:

- Per definir el color i actuar doncs com un 'grup pare' utilitzaríem l'estil /style (Agafem aquest com grup de color, ja que sol hi ha 3 estils – Lager, Ale, IPA- i serà més fàcil entendre la llegenda de la visualització que si posem el nom de la cervesa, que té molts més nivells).
- Com a 'subgrup' utilitzaríem la cerveseria (variable amb 8 nivells).
- Com a variable que descrigui l'àrea de les caselles triarem per exemple ABV. Aquesta tria ve donada perquè ens estant preguntant sobre el alcohol en volum de les cerveses i l'àrea de les caselles necessita d'una variable numèrica continua.
- Finalment, com a 'label/nivell', escollirem el nom de la cervesa.

estás a un escaneo

de encontrar curro.



descarga la

descarga randstad app y empieza hoy.



b)

En la cervesería Bauhaus, la cerveza IPA sembla tenir més ABV (àrea major) que les Lager, per tant demanaria una Lager.

En la cervesería Summit, les tres cerveses amb menys ABV són la Pilsener, la Oatmeal Stout i la Extra Pale Ale

c) Els mosaics són similars als stacked bar o gràfic de barres apilades, però en el cas dels mosaics, l'amplada i l'alçada de les àrees individuals poden variar. En el gràfic de barres apilades només varia una d'aquestes, normalment l'alçada.

Ara bé, els mosaics assumeixen que tots els nivells d'una variable d'agrupament es poden combinar amb tots els nivells d'una altra variable d'agrupació. El treemap no fa aquesta assumpció.

(d) Estem agrupant variables categòriques. La paleta 2 és per variables continues, la paleta 3 és per variables contínues discretitzades. La paleta 1 en canvi, presenta colors diferents per cada grup/categoría, i és per tant la paleta óptima

**2.2. (1.25 pt)** Voleu estudiar algunes relacions entre el tipus de jugades de basquet i extreure conclusions de si això porta als equips a guanyar o perdre. Per això teniu el dataset `statsNBA2008.csv`.

(a) Feu una tibble amb les 10 variables detallades a sota i estudieu la relació entre elles sense reduir-ne el número, tot fent una visualització. Un cop tingueu la visualització mostreu només aquella part que us mostra la informació d'interès. Podeu fer-ho en qualsevol llenguatge, però expliqueu perquè heu escollit aquesta visualització, si la vostra elecció està relacionada amb el tipus de dades que teniu, els passos que heu seguit per fer-la i extrieu alguna/es conclusió/ns del vostre gràfic.

- PTS (percentatge de punts de l'equip, punts per joc)
- FGM (cistelles de camp realitzats, percentatge de cistelles de camp)
- FGA (tirs de camp intentats, número d'intencions de cistelles de camp)
- FTP (percentatge de tirs lliures)
- DRB (rebots defensius)
- ORB (rebots ofensius)
- TRB (rebots totals)
- 3XPM /3PM (cistelles de camp de 3 punts anotats, tirs de 3 punts)
- 3XPA/3PA (rebots defensius, tirs de tres punts)
- 3XPP/3PA (percentatge de tirs de camp de 3 punts)

Nota: Segons les versions us sortirà la X o no en les últimes 3 variables (és a dir, per exemple, alguns tindreu 3XPM i altres 3PM ) **(1 pt)**

(b) Quin gràfic podríeu fer si volguéssiu reduir el nombre de variables però no fos possible identificar les variables que es podrien eliminar a simple vista, i volguéssiu a la vegada, assegurar que les vostres variables són independents unes de les altres? (no cal fer la visualització, només contestar la pregunta) **(0.25pts)**

**RESPOSTA:** Les matrius de correlació mostren els coeficients de correlació/relacions entre un nombre relativament gran de variables contínues. Com tenim 10 variables, podem fer una matriu de correlació per veure les relacions entre totes elles.

Primer ens preparam el dataframe que ens interessa, vam veure al seminari 7 les tibles de R i com utilitzar-les. Una manera que es podria fer servir, seria

```
> NBA_tb <- as_tibble(NBA)
> NBA_sel<-NBA_tb[,c('PTS','FGM','FGA', 'FTP', 'DRB','ORB','TRB',
  '3PM','3PA','3PP' )]
```

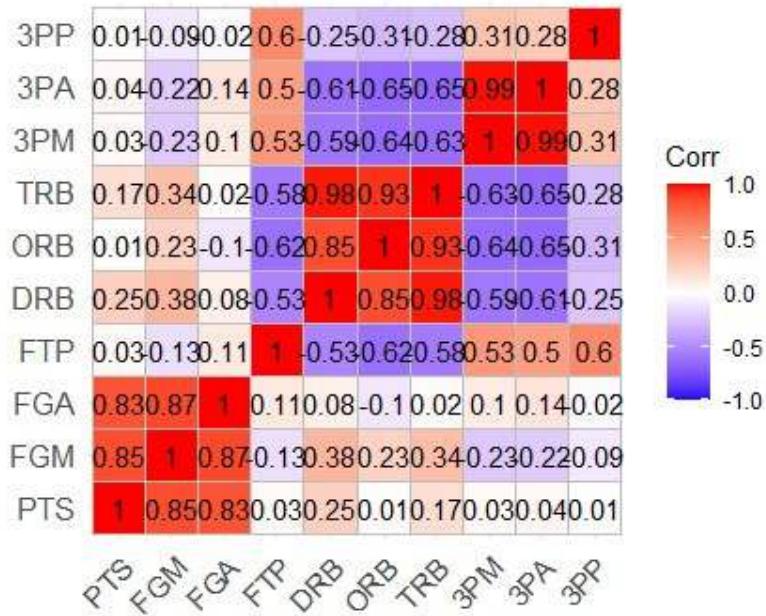
També ho podríem fer utilitzant `select` de `dplyr` com havíem vist en classes anteriors.

Necessitarem a més de la `tidyverse`, la llibreria `ggcorrplot`

Fem la matriu de correlació com vam veure al seminari 7, arrodonint a 2 decimals, per exemple

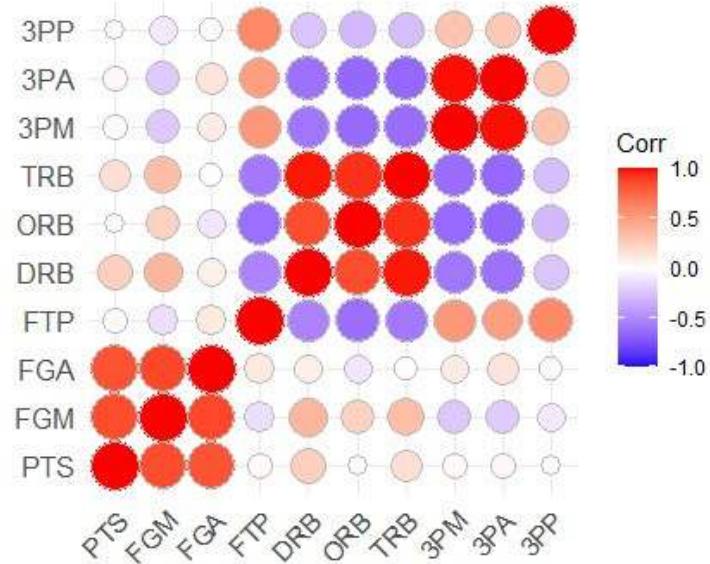
```
> cormat <- round(cor(NBA_sel),2)
```

```
> ggcorrplot(cormat, lab=TRUE)
```



O amb cercles: >

```
ggcorrplot(cormat,method= 'circle')
```



En qualsevol dels casos al ser una matriu simètrica, només mostrem, la part de sobre o sota la diagonal

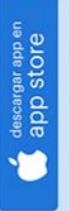
estás a un escaneo

de encontrar curro.



descarga la

descarga randstad app y empieza hoy.

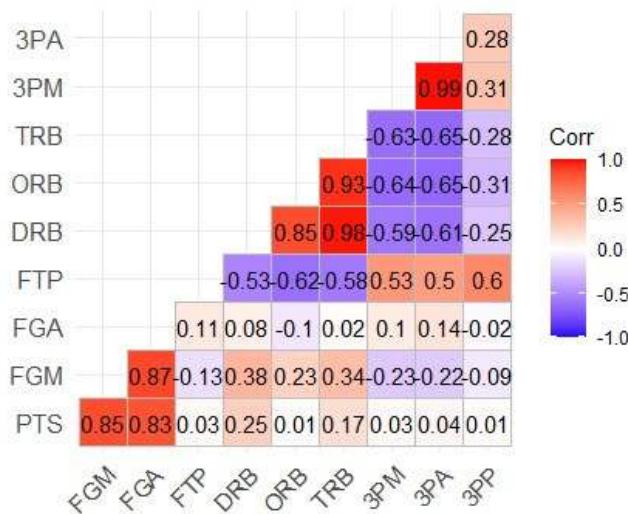


app store

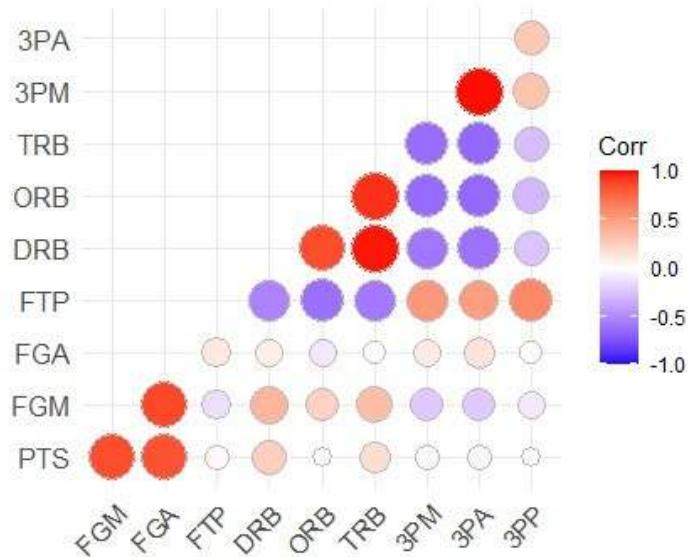


descargar app en  
google play

```
> ggcorrplot(cormat, lab=TRUE, type = "lower")
```



```
> ggcorrplot(cormat, method='circle', type = "lower")
```



#### CONCLUSIONS POSIBLES (n'hi ha moltes)

- Les variables: PTS (percentatge de punts de l'equip, punts per joc), FGM (cistelles de camp realitzats, percentatge de cistelles de camp) i FGA (tirs de camp intentats, número

- d'intencions de cistelles de camp) , tenen una correlació positiva entre ells és forta, sembla que fer més intents resulta en obtenir una puntuació més alta.*
- *Sembla que els DRB (Rebots defensius), ORB (rebots ofensius), també van estar molt correlacionats (negativament) en la NBA 2008.*
  - *Hi ha bastanta correlació també entre FTP i 3PM, 3PA i 3PP*
  - *No hi ha cap correlació entre ORB i PTS. Fer més rebots ofensius no ajudava a guanyar. B) Clarament si volem reduir el nombre de variables, assegurant la seva independència, usarem un PCA com vam veure en Teoria 7 i seminari següent*

## PART 3 (4 pt)

*Dataset: 25\_noms\_padro\_any\_sexe\_1996\_2019.csv*

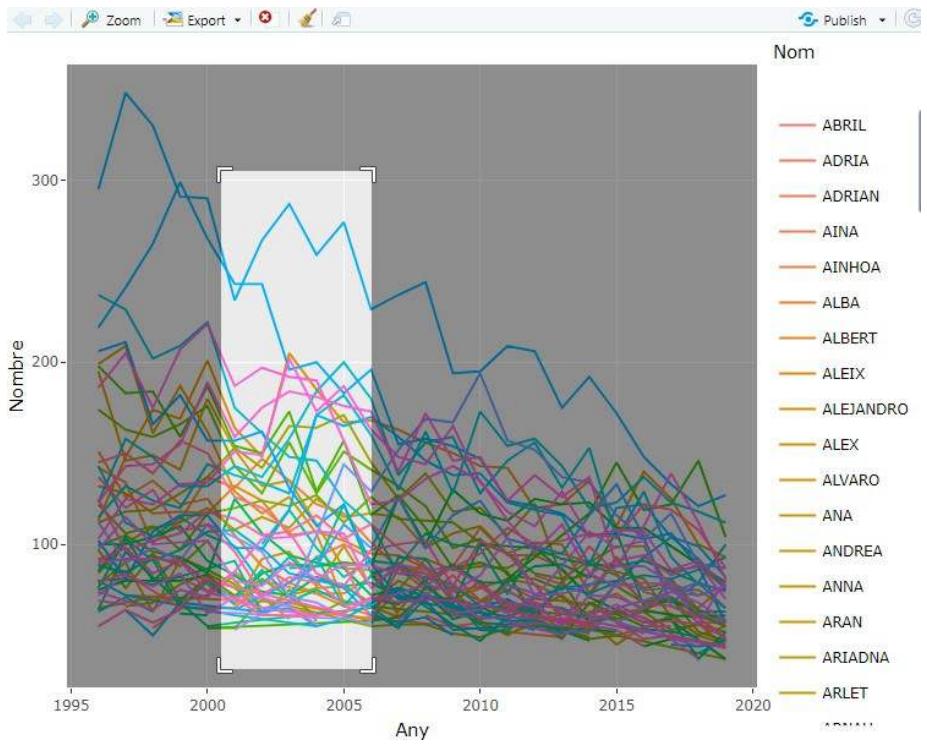
*Agafarem el dataset de noms del padró de naixements de Barcelona. Utilitzeu les llibreries d'interactivitat o animació (plotly, ganimate, shiny, gifski, etc.) que creieu convenient i dibuixeus gràfiques que us facin falta. RESPOSTA: (afegiu en aquesta secció les llibreries que utilitzareu)*

```
> library(tidyverse)
> library(dplyr)
> library(plotly)
> getwd()
> setwd("C:/Users/enric/Documents/R")
> NadonsBCN <- read.csv('./25_noms_padro_any_sexe_1996_2019.csv')
```

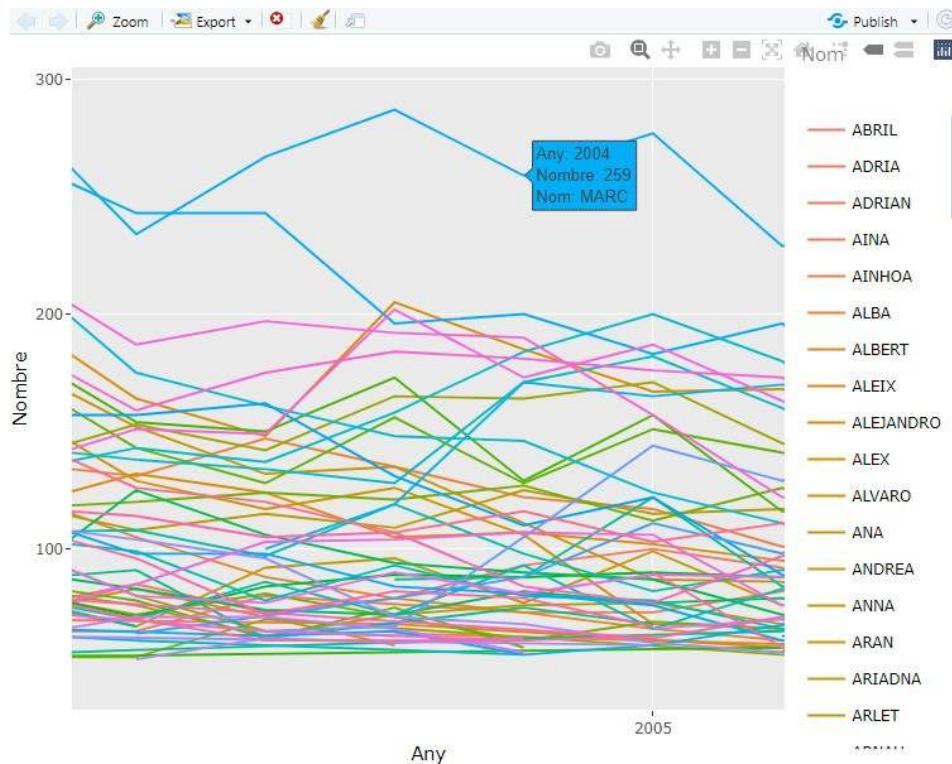
**3.1 (1 pt)** Mostra el codi i la gràfica de l'evolució temporal de tots els noms del dataset (gràfica 1) i enumera els 10 noms femenins o masculins més posats l'any 2004. Digues quin són aquests noms i el nombre de nadons, per ordre decreixent i mostra el codi i la gràfica de l'evolució temporal d'aquests 10 noms (gràfica 2).

**RESPOSTA:** En primer lloc, llistar tots els noms del dataset (gràfica 1):

```
> plotNoms<- ggplot(NadonsBCN,aes(x=Any, y=Nombre, color=Nom))+geom_path()
> ggplotly(plotNoms)
```



*CERCA INTERACTIVA: fer un zoom en la zona propera a l'any 2004.*



*CERCA AMB DATA MASSAGING (igual de bé):*

*Filtrar per l'any 2004, definir un ranking per a obtenir els 10 primers i (opcionalment) ordenar-los en decreixent:*

estás a un escaneo

de encontrar curro.



descarga la

descarga randstad app y empieza hoy.



```
> NadonsBCN_20040 <- NadonsBCN %>% filter(Any==2004) %>% mutate(rank = rank(-Nombre)) %>% group_by(Nom) %>% filter(rank <=10) %>% ungroup() %>% arrange(-Nombre) > NadonsBCN_20040
# A tibble: 10 x 6
  Ordre Nom   Sexe Any Nombre rank
  <int> <chr> <chr> <int> <int> <dbl>
1     1 MARC Home  2004    259     1
2     1 MARIA Dona  2004    200     2
3     2 PAULA Dona  2004    190     3
4     2 ALEX  Home  2004    185     4
5     3 LAIA  Dona  2004    184     5
6     3 POL   Home  2004    181     6
7     4 PAU   Home  2004    173     7
8     4 JULIA Dona  2004    171    8.5
9     5 LUCIA Dona  2004    171    8.5
10    6 CARLA Dona  2004    164    10
```

Els noms són **MARC, MARIA, PAULA, ALEX, LAIA, POL, PAU, JULIA, LUCIA i CARLA**.

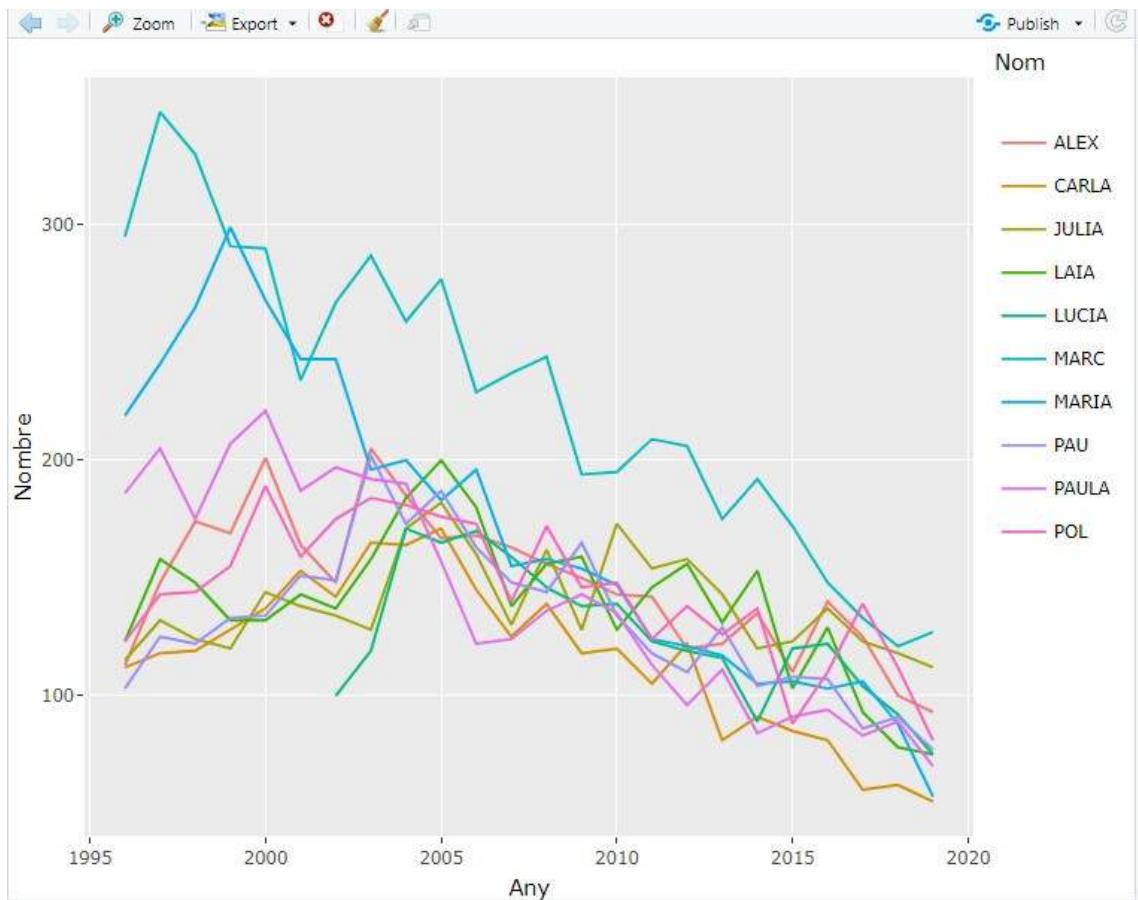
### GRÀFICA 2:

Després llistem les 10 gràfiques, bé per ggplotly:

```
> NadonsBCN_Noms2004 <- NadonsBCN %>% filter(Nom=="MARC" | Nom=="MARIA"
| Nom=="PAULA" | Nom=="ALEX" | Nom=="LAIA" | Nom=="POL" | Nom=="PAU" | Nom=="JULIA"
| Nom=="LUCIA" | Nom=="CARLA")
> plotNoms2004 <- ggplot(NadonsBCN_Noms2004,aes(x=Any, y=Nombre,
color=Nom)) + geom_path() > ggplotly(plotNoms2004)
```

# O bé per plot\_ly:

```
> plot_ly(NadonsBCN_Noms2004,x=~Any, y=~Nombre,color=~Nom) %>%
add_lines()
```



O bé per shiny:

```
> ui <- fluidPage( selectizeInput( inputId = "Noms",
  label = "Selecciona un Nom:", choices =
  unique(NadonsBCN$Nom), selected =
  c("MARC", "MARIA", "PAULA", "ALEX", "LAIA",
    "POL", "PAU", "JULIA", "LUCIA", "CARLA"),
  multiple = TRUE
),
  plotlyOutput(outputId = "p")
)

server <- function(input, output, ...)
{ output$p <- renderPlotly (
  { plot_ly(NadonsBCN, x = ~Any, y = ~Nombre, color=~Nom) %>%
    filter(Nom %in% input$Noms) %>%
    group_by(Nom) %>%
    add_lines()      }
)
shinyApp(ui, server)
```



**3.2. (0,5 pt) Sobre gràfica 2 anterior, quin és l'ordre decreixent d'aquests noms l'any 2010?.**

**RESPOSTA:**

**CERCA INTERACTIVA:** Mirar la gràfica 2 i cercar en ordre decreixent els noms l'any 2010.

| Ordre | Nom   | Sexe | Any  | Nombre |
|-------|-------|------|------|--------|
| 1     | MARC  | Home | 2010 | 195    |
| 2     | JULIA | Dona | 2010 | 173    |
| 3     | POL   | Home | 2010 | 148    |
| 4     | MARIA | Dona | 2010 | 147    |
| 5     | ALEX  | Home | 2010 | 143    |
| 6     | LUCIA | Dona | 2010 | 139    |
| 7     | PAULA | Dona | 2010 | 135    |
| 8     | PAU   | Home | 2010 | 134    |
| 9     | LAIA  | Dona | 2010 | 128    |
| 10    | CARLA | Dona | 2010 | 120    |

**CERCA AMB DATA MASSAGING (igual de bé):**

Sobre el data set inicial (NadonsBCN) filtrar pels 10 noms i per l'any 2010, i ordenar en decreixent:

```
> NadonsBCN_2004_20100 <- NadonsBCN %>% filter(Nom=="MARC" | Nom=="MARIA" | Nom=="PAULA" | Nom=="ALEX" | Nom=="LAIA" | Nom=="POL" | Nom=="PAU" |
```

```

axis.text.x=element_blank(),
axis.text.y=element_blank(),
axis.ticks=element_blank(),
axis.title.x=element_blank(),
axis.title.y=element_blank(),           legend.position="none",
panel.background=element_blank(),
panel.border=element_blank(),
panel.grid.major=element_blank(),
panel.grid.minor=element_blank(),
      panel.grid.major.x = element_line( size=.1, color="grey" ),
panel.grid.minor.x = element_line( size=.1, color="grey" ),
plot.title=element_text(size=25,      hjust=0.5,      face="bold",
colour="grey", vjust=-1),
      plot.subtitle=element_text(size=18, hjust=0.5, face="italic",
color="grey"),
      plot.caption =element_text(size=8, hjust=0.5, face="italic",
color="grey"),
      plot.background=element_blank(),
plot.margin = margin(2,2, 2, 4, "cm")) +
transition_states(Any, transition_length = 4, state_length = 1, wrap
= FALSE) +
view_follow(fixed_x = TRUE) +
labs(title = 'Noms de Nens a BCN Any: {closest_state}',
subtitle = "Top 10 Noms del 2004",
caption = "Nombre de nadons | Data Source: Ajuntament de BCN")
> anim
>

```

*PAS 3a: EXPORTAR FRAMES A FITXER GIF*

```

> animate(anim, 200, fps = 20, width = 1200, height = 1000, renderer
      = gifski_renderer("nadonsBCN_2004.gif"), end_pause =
      15, start_pause = 15)
>

```

*PAS 3b: EXPORTAR FRAMES A FITXER AVI*

```

> animate(anim, 200, fps = 20, width = 1200, height = 1000,
renderer = av_renderer("nadonsBCN_2004.avi"), end_pause = 15,
start_pause = 15) >

```

**4 (1 pt)** Defineix 4 dels següents conceptes en animació, interactivitat, usabilitat i Experiència d'Usuari en Visualització de Dades:

- **SUS**
- **Control**
- **CheckBox**

- **Participar / Col·laborar**
- ***Useful***
- **Qüestionari Attrakdiff**
- **UEQ**
- **Qüestionari SAM**

2025  
10-13 JULY



**Visualització de dades (Enginyeria de Dades - EE - UAB)**  
**Examen Recuperació Segon Parcial - 5 Juliol 2021**  
**SOLUCIONS MODEL A**

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.

### **PARTE 1 (2.5 pt)**

1.1. (0.5 pt) Explica brevemente qué es una escala de color perceptualmente correcta y qué ventajas tiene.

**RESPOSTA:**

Son escalas que tienen en cuenta cómo nuestro cerebro procesa el color y en las que los colores varían de forma gradual e uniforme. Una ventaja es que estas variaciones entre colores se asocian mejor a las variaciones graduales e uniformes entre los valores del dataset.

1.2. (1 pt) Imagina que tenemos un dataset de precipitaciones (lluvia) en Europa y queremos visualizarlo en un mapa. La precipitación es un atributo cuantitativo continuo con valores min=0 y max=10 mm/h. ¿Qué escala de color usarías?

Di qué tipo de escala es, justifica la elección de colores, y hazla en R.

**RESPOSTA:**

Una escala secuencial continua. Los colores más adecuados por semántica y significado podrían ser una gama de azules.

1.3. (1 pt) ¿Qué escala de color usarías para representar la velocidad en X del viento, teniendo en cuenta que tiene valores min=-15 y max=50 m/s?

Di qué tipo de escala es, justifica la elección de colores, y hazla en R o sube una imagen.

**RESPOSTA:**

Escala divergente o cuantitativa divergente. El color neutro (blanco) debería corresponder al 0 en los datos y habrá dos colores distintos en cada extremo.

## PART 2 (3.5 pt)

*Dataset: beers.csv*

Aquest dataset té 7 atributs per 44 cerveses de Minnesota i el farem servir en l'exercici 2.1. Els atributs són: Nom de la cerveseria, cervesa, descripció, estil, alcohol en volum (ABV), unitats internacionals d'amargor (IBU), puntuació que han rebut

*Dataset: statsNBA2008.csv*

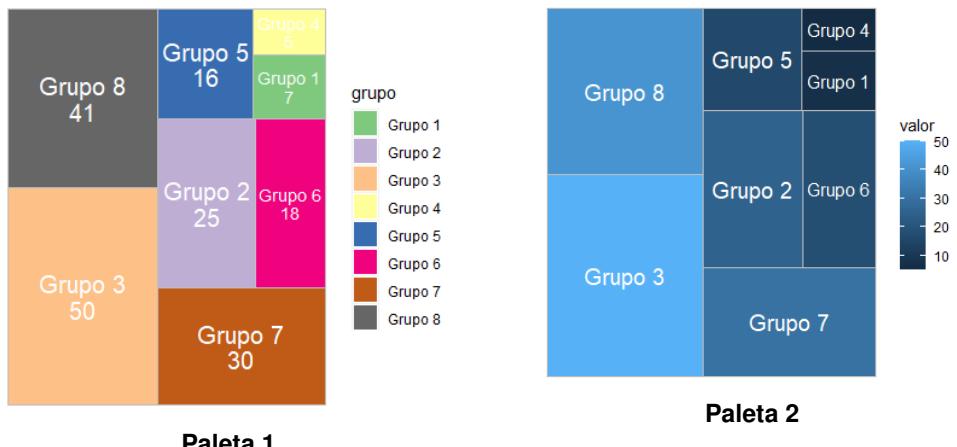
Aquest dataset té 21 atributs de 50 jugadors de la NBA. Recull les estadístiques de la NBA del 2008. I estudiarem la relació entre alguns dels atributs en l'exercici 2.2

**2.1 (2.25 pt)** És estiu, esteu a Minnesota i voleu anar a prendre una cervesa. Utilitzant el dataset *beers.csv*.

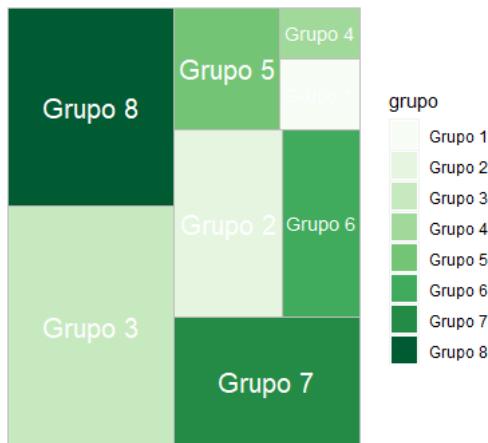
- a) Feu un mapa d'arbre (*treemap*) que us permeti contestar les preguntes de l'apartat b. Podeu fer-lo en R o en qualsevol altre llenguatge, utilitzant les llibreries que us semblin convenient. Ara bé, heu d'explicar-ho bé: comenceu narrant com és un *treemap* en general i per quin tipus de dades serveix i redacteu els passos que heu de fer per construir la visualització triada. (1.5 pt)
- b) Esteu amb una persona que li puja molt l'alcohol i vol una cervesa amb poc alcohol per volum. Si esteu a la cerveseria Bauhaus, quin tipus de cervesa (Lager, Ale o IPA) li demaneu?. I si esteu en la cerveseria Summit, entre quines tres cerveses podeu triar per demanar-li (doneu el nom de les 3 cerveses)? (0.25 pt)
- c) Expliqueu quan usem un mosaic enlloc d'un gràfic de barres apilades, i quan no podem utilitzar un mosaic i usem un *treemap*. (0.25 pt)
- d) Quin tipus d'escala de color, de les de la pàgina següent, seria la més òptima i perquè? (0.25 pt)

- Paleta 1
- Paleta 2
- Paleta 3

Raona la resposta:



**Paleta 1**



**Paleta 3**

### RESPOSTA:

a) Com vam veure a classe, un mapa d'arbre és un dibuix rectangular dividit en caselles, i cada casella representa una sola observació. Vam veure que era una bona manera de mostrar dades jeràrquiques mitjançant rectangles imbricats. I l'àrea relativa de cada casella expressava una variable contínua. També vam veure que era òptim quan hi ha com a màxim dues variables d'agrupació, per tant no en definirem més.

Quan s'especifiquen les proporcions d'acord amb múltiples variables d'agrupació, les representacions els mapes d'arbres són aproximacions de visualització útils.

Per tant, una possibilitat amb aquest dataset seria:

- Per definir el color i actuar doncs com un 'grup pare' utilitzaríem l'estil /style (Agafem aquest com grup de color, ja que sol hi ha 3 estils – Lager, Ale, IPA- i serà més fàcil entendre la llegenda de la visualització que si posem el nom de la cervesa, que té molts més nivells).
- Com a 'subgrup' utilitzaríem la cerveseria (variable amb 8 nivells).
- Com a variable que descrigui l'àrea de les caselles triarem per exemple ABV. Aquesta tria ve donada perquè ens estant preguntant sobre el alcohol en volum de les cerveses i l'àrea de les caselles necessita d'una variable numèrica continua.
- Finalment, com a 'label/nivell', escollirem el nom de la cervesa.

2025  
10-13 JULY



*Si ho fem amb R, el primer que hem de fer és carregar les llibreries necessàries. La llibreria específica aquí és treemapify (abans necessitarem tenir instal·lat el paquet com vam veure al seminari 7). També haurem de fer us de geom\_treemap.*

*Per fer un treemap basic doncs:*

```
> library(ggplot2)
> library(treemapify)
> Beers <-read_csv('C:/DATA/beers.csv')
> ggplot(Beers, aes(area=ABV, fill=Style, label=Beer,
+ subgroup=Brewery))+geom_treemap()+ geom_treemap_subgroup_border(colour =
+ "black",size=3) +geom_treemap_subgroup_text(alpha=0.5,colour =
+ "white")+geom_treemap_text (aes(label=Beer))
```



b)

*En la cerveseria Bauhaus, la cervesa IPA sembla tenir més ABV (àrea major) que les Lager, per tant demanaria una Lager.*

*En la cerveseria Summit, les tres cerveses amb menys ABV són la Pilsener, la Oatmeal Stout i la Extra Pale Ale*

c) Els mosaics són similars als stacked bar o gràfic de barres apilades, però en el cas dels mosaics, l'amplada i l'alçada de les àrees individuals poden variar. En el gràfic de barres apilades només varia una d'aquestes, normalment l'alçada.

*Ara bé, els mosaics assumeixen que tots els nivells d'una variable d'agrupament es poden combinar amb tots els nivells d'una altra variable d'agrupació. El treemap no fa aquesta assumpció.*

(d) *Estem agrupant variables categòriques. La paleta 2 és per variables continues, la paleta 3 és per variables contínues discretitzades. La paleta 1 en canvi, presenta colors diferents per cada grup/categoría, i és per tant la paleta òptima*

**2.2. (1.25 pt)** Voleu estudiar algunes relacions entre el tipus de jugades de basquet i extreure conclusions de si això porta als equips a guanyar o perdre. Per això teniu el dataset `statsNBA2008.csv`.

(a) Feu una tibble amb les 10 variables detallades a sota i estudieu la relació entre elles sense reduir-ne el número, tot fent una visualització. Un cop tingueu la visualització mostreu només aquella part que us mostra la informació d'interès. Podeu fer-ho en qualsevol llenguatge, però expliqueu perquè heu escollit aquesta visualització, si la vostra elecció està relacionada amb el tipus de dades que teniu, els passos que heu seguit per fer-la i extrieu alguna/es conclusió/ns del vostre gràfic.

- PTS (percentatge de punts de l'equip, punts per joc)
- FGM (cistelles de camp realitzats, percentatge de cistelles de camp)
- FGA (tirs de camp intentats, número d'intencions de cistelles de camp)
- FTP (percentatge de tirs lliures)
- DRB (rebots defensius)
- ORB (rebots ofensius)
- TRB (rebots totals)
- 3XPM /3PM (cistelles de camp de 3 punts anotats, tirs de 3 punts)
- 3XPA/3PA (rebots defensius, tirs de tres punts)
- 3XPP/3PA (percentatge de tirs de camp de 3 punts)

Nota: Segons les versions us sortirà la X o no en les últimes 3 variables (és a dir, per exemple, alguns tindreu 3XPM i altres 3PM )

**(1 pt)**

(b) Quin gràfic podríeu fer si volguéssiu reduir el nombre de variables però no fos possible identificar les variables que es podrien eliminar a simple vista, i volguéssiu a la vegada, assegurar que les vostres variables són independents unes de les altres? (no cal fer la visualització, només contestar la pregunta)

**(0.25pts)**

**RESPOSTA:** Les matrius de correlació mostren els coeficients de correlació/relacions entre un nombre relativament gran de variables contínues. Com tenim 10 variables, podem fer una matriu de correlació per veure les relacions entre totes elles.

Primer ens preparam el dataframe que ens interessa, vam veure al seminari 7 les tibles de R i com utilitzar-les. Una manera que es podria fer servir, seria

```
> NBA_tb <- as_tibble(NBA)
> NBA_sel<-NBA_tb[,c('PTS','FGM','FGA', 'FTP', 'DRB','ORB','TRB',
  '3PM','3PA','3PP' )]
```

També ho podríem fer utilitzant `select` de `dplyr` com havíem vist en classes anteriors.

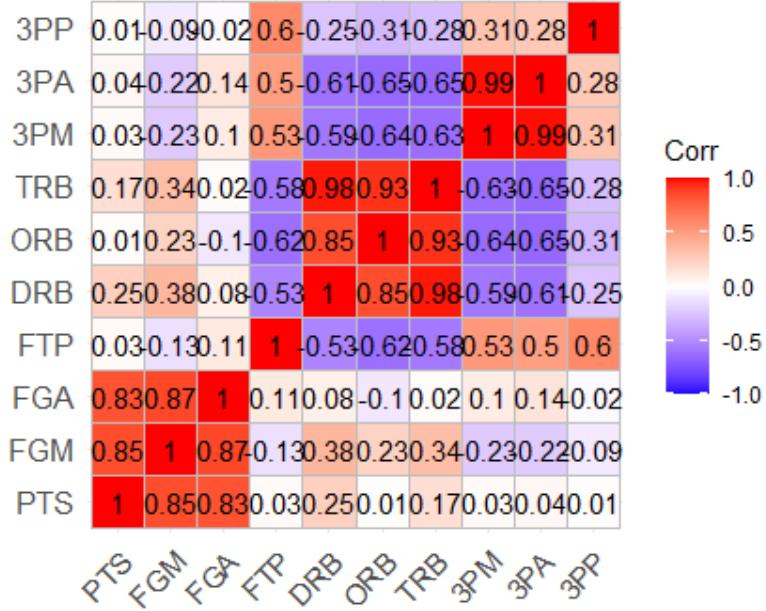
Necessitarem a més de la tidyverse, la llibreria `ggcorrplot`

```
> library(ggcorrplot)
```

Fem la matriu de correlació com vam veure al seminari 7, arrodonint a 2 decimals, per exemple

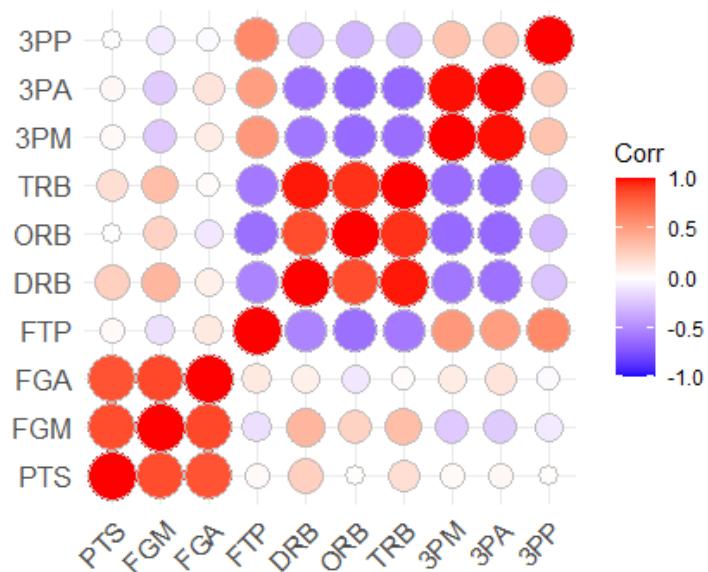
```
> cormat <- round(cor(NBA_sel),2)
```

```
> ggcorrplot(cormat, lab=TRUE)
```



O amb cercles:

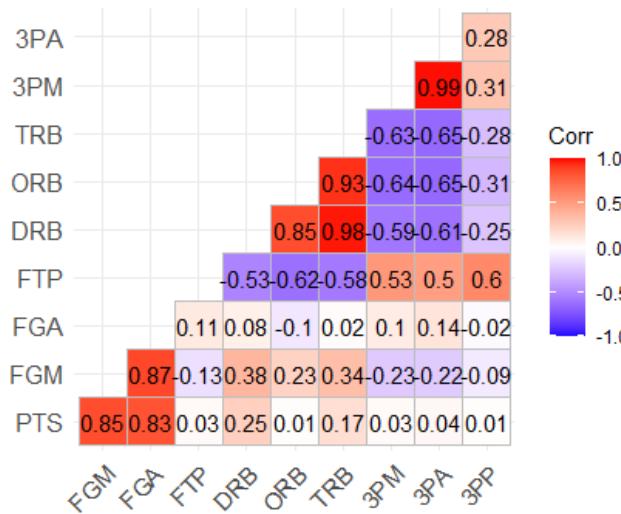
```
> ggcorrplot(cormat, method= 'circle')
```



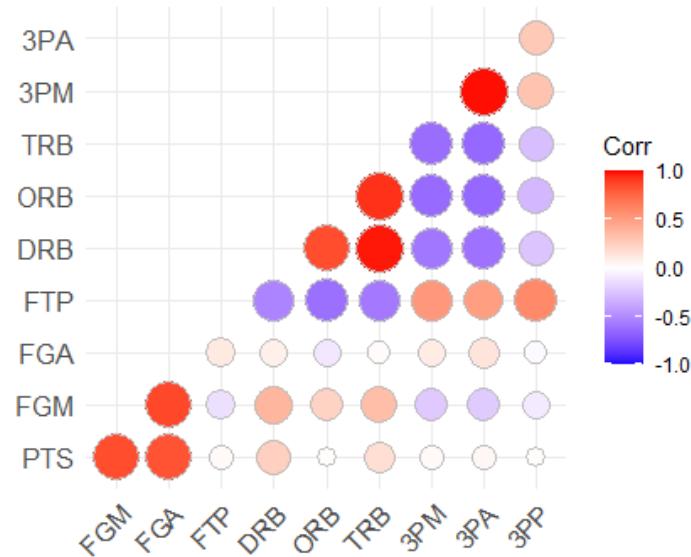
En qualsevol dels casos al ser una matriu simètrica, només mostrem, la part de sobre o sota la diagonal



> `ggcorrplot(cormat, lab=TRUE, type = "lower")`



> `ggcorrplot(cormat, method='circle', type = "lower")`



#### CONCLUSIONS POSIBLES (n'hi ha moltes)

- Les variables: PTS (percentatge de punts de l'equip, punts per joc), FGM (cistelles de camp realitzats, percentatge de cistelles de camp) i FGA (tirs de camp intentats, número d'intencions de cistelles de camp) , tenen una correlació positiva entre ells és forta, sembla que fer més intents resulta en obtenir una puntuació més alta.

- Sembla que els DRB (Rebots defensius), ORB (rebots ofensius), també van estar molt correlacionats (negativament) en la NBA 2008.
  - Hi ha bastanta correlació també entre FTP i 3PM, 3PA i 3PP
  - No hi ha cap correlació entre ORB i PTS. Fer més rebots ofensius no ajudava a guanyar.
- B) Clarament si volem reduir el nombre de variables, assegurant la seva independència, usarem un PCA com vam veure en Teoria 7 i seminari següent

## PART 3 (4 pt)

*Dataset: 25\_noms\_padro\_any\_sexe\_1996\_2019.csv*

*Agafarem el dataset de noms del padró de naixements de Barcelona. Utilitzeu les llibreries d'interactivitat o animació (plotly, gganimate, shiny, gifski, etc.) que creieu convenient i dibuixeus les gràfiques que us facin falta.*

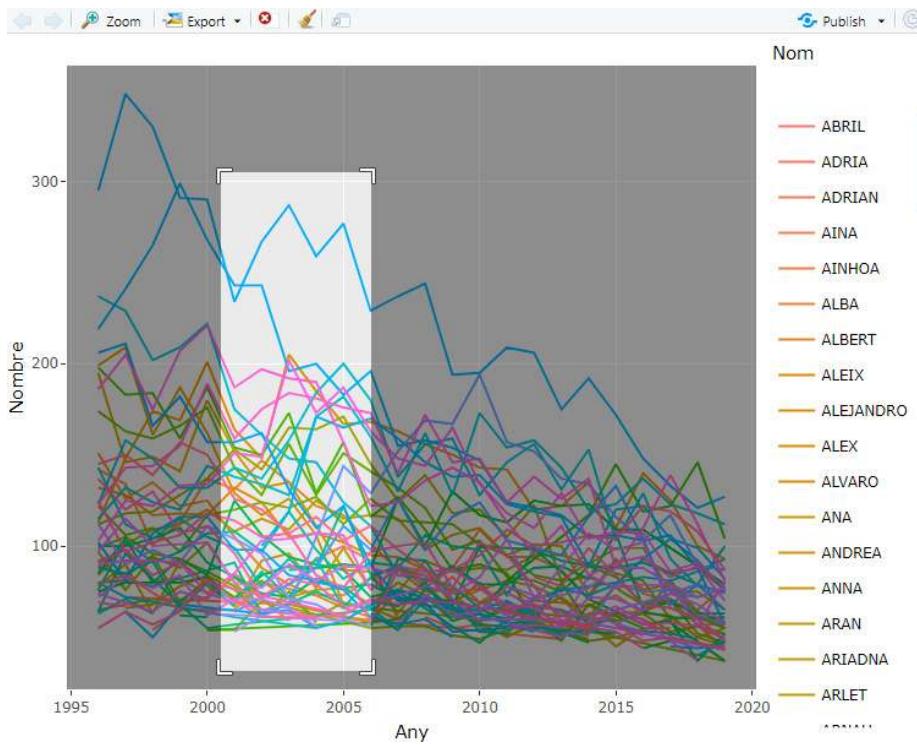
**RESPOSTA:** (afegiu en aquesta secció les llibreries que utilitzareu)

```
> library(tidyverse)
> library(dplyr)
> library(plotly)
> getwd()
> setwd("C:/Users/enric/Documents/R")
> NadonsBCN <- read.csv('./25_noms_padro_any_sexe_1996_2019.csv')
```

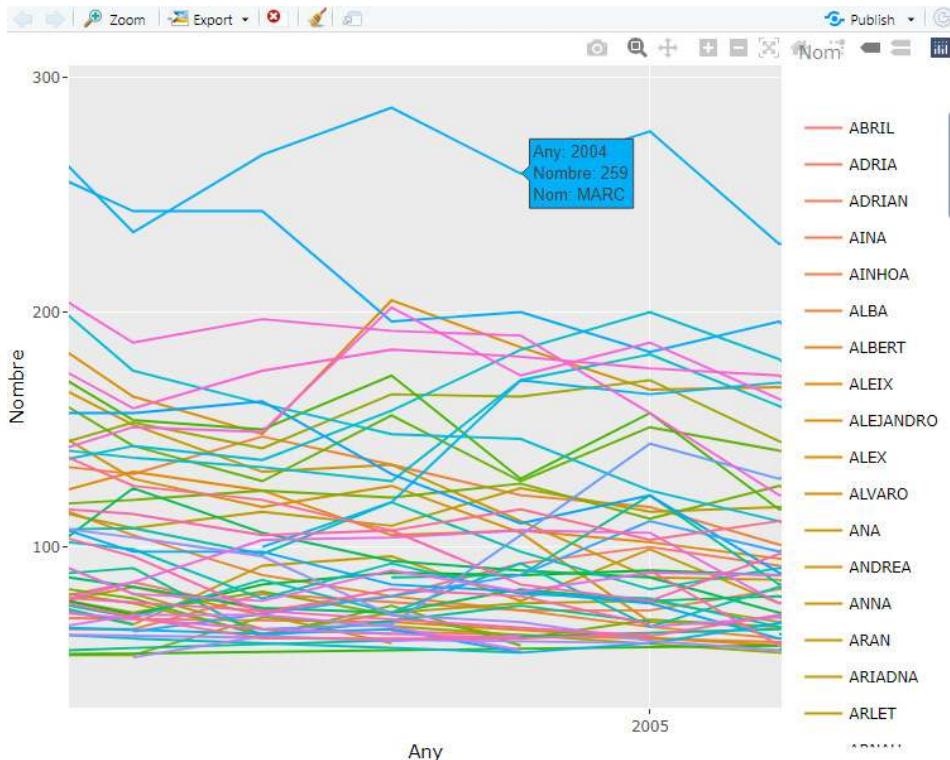
**3.1 (1 pt)** Mostra el codi i la gràfica de l'evolució temporal de tots els noms del dataset (gràfica 1) i enumera els 10 noms femenins o masculins més posats l'any 2004. Digues quin són aquests noms i el nombre de nadons, per ordre decreixent i mostra el codi i la gràfica de l'evolució temporal d'aquests 10 noms (gràfica 2).

**RESPOSTA:** En primer lloc, llistar tots els noms del dataset (gràfica 1):

```
> plotNoms<- ggplot(NadonsBCN,aes(x=Any, y=Nombre,color=Nom))+geom_path()
> ggplotly(plotNoms)
```



CERCA INTERACTIVA: fer un zoom en la zona propera a l'any 2004.



CERCA AMB DATA MASSAGING (igual de bé):

Filtrar per l'any 2004, definir un ranking per a obtenir els 10 primers i (opcionalment) ordenar-los en decreixent:

2025  
10-13 JULY



```
> NadonsBCN_20040 <- NadonsBCN %>% filter(Any==2004) %>% mutate(rank = rank(-Nombre)) %>% group_by(Nom) %>% filter(rank <=10) %>% ungroup() %>% arrange(-Nombre)
> NadonsBCN_20040
# A tibble: 10 x 6
  Ordre Nom   Sexe Any Nombre rank
  <int> <chr> <chr> <int> <int> <dbl>
1     1 MARC   Home  2004    259    1
2     1 MARIA  Dona  2004    200    2
3     2 PAULA  Dona  2004    190    3
4     2 ALEX   Home  2004    185    4
5     3 LAIA   Dona  2004    184    5
6     3 POL    Home  2004    181    6
7     4 PAU    Home  2004    173    7
8     4 JULIA  Dona  2004    171  8.5
9     5 LUCIA  Dona  2004    171  8.5
10    6 CARLA  Dona  2004    164   10
```

Els noms són **MARC, MARIA, PAULA, ALEX, LAIA, POL, PAU, JULIA, LUCIA i CARLA**.

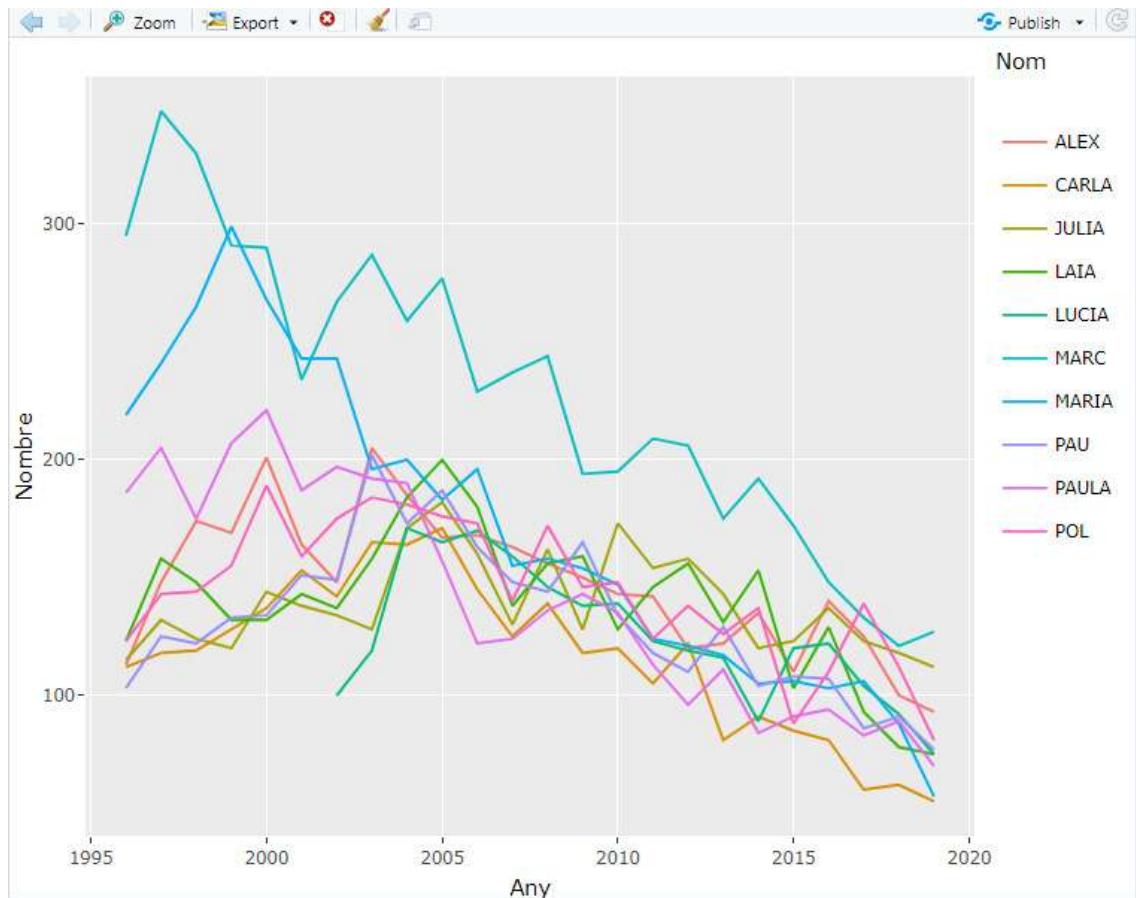
### GRÀFICA 2:

Després llistem les 10 gràfiques, bé per ggplotly:

```
> NadonsBCN_Noms2004 <- NadonsBCN %>% filter(Nom=="MARC" | Nom=="MARIA"
| Nom=="PAULA" | Nom=="ALEX" | Nom=="LAIA" | Nom=="POL" | Nom=="PAU" | Nom=="JULIA"
| Nom=="LUCIA" | Nom=="CARLA")
> plotNoms2004 <- ggplot(NadonsBCN_Noms2004,aes(x=Any,      y=Nombre,
color=Nom)) + geom_path()
> ggplotly(plotNoms2004)
```

# O bé per plot\_ly:

```
> plot_ly(NadonsBCN_Noms2004,x=~Any, y=~Nombre,color=~Nom) %>% add_lines()
```



O bé per shiny:

```
> ui <- fluidPage(selectizeInput( inputId = "Noms",
  label = "Selecciona un Nom:",
  choices = unique(NadonsBCN$Nom),
  selected = c("MARC","MARIA","PAULA","ALEX","LAIA",
              "POL","PAU","JULIA","LUCIA","CARLA"),
  multiple = TRUE
  ),
  plotlyOutput(outputId = "p")
  )

server <- function(input, output, ...)
{  output$p <- renderPlotly (
  {plot_ly(NadonsBCN, x = ~Any, y = ~Nombre, color=~Nom) %>%
    filter(Nom %in% input$Noms) %>%
    group_by(Nom) %>%
    add_lines()
  })
}
shinyApp(ui, server)
```



**3.2. (0,5 pt)** Sobre gràfica 2 anterior, quin és l'ordre decreixent d'aquests noms l'any 2010?

**RESPOSTA:**

*CERCA INTERACTIVA: Mirar la gràfica 2 i cercar en ordre decreixent els noms l'any 2010.*

| Ordre | Nom   | Sexe | Any  | Nombre |
|-------|-------|------|------|--------|
| 1     | MARC  | Home | 2010 | 195    |
| 2     | JULIA | Dona | 2010 | 173    |
| 3     | POL   | Home | 2010 | 148    |
| 4     | MARIA | Dona | 2010 | 147    |
| 5     | ALEX  | Home | 2010 | 143    |
| 6     | LUCIA | Dona | 2010 | 139    |
| 7     | PAULA | Dona | 2010 | 135    |
| 8     | PAU   | Home | 2010 | 134    |
| 9     | LAIA  | Dona | 2010 | 128    |
| 10    | CARLA | Dona | 2010 | 120    |

*CERCA AMB DATA MASSAGING (igual de bé):*

Sobre el data set inicial (NadonsBCN) filtrar pels 10 noms i per l'any 2010, i ordenar en decreixent:

```
> NadonsBCN_2004_20100 <- NadonsBCN %>% filter(Nom=="MARC" | Nom=="MARIA" | Nom=="PAULA" | Nom=="ALEX" | Nom=="LAIA" | Nom=="POL" | Nom=="PAU" | Nom=="JULIA" | Nom=="LUCIA" | Nom=="CARLA") %>% filter(Any==2010) %>% arrange(-Nombre)
```



> NadonsBCN\_2004\_20100

|    | Ordre | Nom   | Sexe | Any  | Nombre |
|----|-------|-------|------|------|--------|
| 1  | 1     | MARC  | Home | 2010 | 195    |
| 2  | 2     | JULIA | Dona | 2010 | 173    |
| 3  | 2     | POL   | Home | 2010 | 148    |
| 4  | 3     | MARIA | Dona | 2010 | 147    |
| 5  | 3     | ALEX  | Home | 2010 | 143    |
| 6  | 4     | LUCIA | Dona | 2010 | 139    |
| 7  | 5     | PAULA | Dona | 2010 | 135    |
| 8  | 4     | PAU   | Home | 2010 | 134    |
| 9  | 6     | LAIA  | Dona | 2010 | 128    |
| 10 | 7     | CARLA | Dona | 2010 | 120    |

3.3. (1,5 pt) Mostra el codi i un frame del Ranking de Barres Animades (*Animated Bar Race Ranking*) temporal amb els 10 noms femenins o masculins més posats l'any 2004. Inclou el fitxer d'animació GIF (*nadonsBCN\_2004.gif*) o AVI (*nadonsBCN\_2004.avi*) en l'entrega. Carrega les llibreries que et facin falta.

**RESPOSTA:**

**PAS 0: CARREGAR LLIBRERIES:**

```
> library(gganimate)
> library(gifski)
> library(av)
>
```

**PAS 1: FILTRAR EL DATASET AMB ELS 10 NOMS FEMENINS O MASCULINS MÉS POSATS L'ANY 2004:**

```
> NadonsBCN_Noms2004 <- NadonsBCN %>% filter(Nom=="MARC" | Nom=="MARIA" |
  Nom=="PAULA" | Nom=="ALEX" | Nom=="LAIA" | Nom=="POL" | Nom=="PAU" | Nom=="JULIA" | Nom=="LUCIA" | Nom=="CARLA")
```

**PAS 2: GENERAR FRAMES DEL RANKING DE BARRES ANIMADES (*Animated Bar Race Ranking*) EL DATASET AMB ELS 10 NOMS FEMENINS O MASCULINS MÉS POSATS L'ANY 2004.**

```
> anim <- ggplot(NadonsBCN_Noms2004, aes(Ordre, group = Nom,
  fill = as.factor(Nom), color = as.factor(Nom))) +
  geom_tile(aes(y = Nombre/2,
    height = Nombre,
    width = 0.9), alpha = 0.8, color = NA) +
  geom_text(aes(y = 0, label = paste(Nom, " ")), vjust = 0.2, hjust =
  1) +
  geom_text(aes(y=Nombre,label = Nombre, hjust=0)) +
  coord_flip(clip = "off", expand = FALSE) +
  scale_x_reverse() +
  guides(color = FALSE, fill = FALSE) +
  theme(axis.line=element_blank(),
  axis.text.x=element_blank(),
  axis.text.y=element_blank(),
```

```

axis.ticks=element_blank(),
axis.title.x=element_blank(),
axis.title.y=element_blank(),
legend.position="none",
panel.background=element_blank(),
panel.border=element_blank(),
panel.grid.major=element_blank(),
panel.grid.minor=element_blank(),
panel.grid.major.x = element_line( size=.1, color="grey" ),
panel.grid.minor.x = element_line( size=.1, color="grey" ),
plot.title=element_text(size=25, hjust=0.5, face="bold",
colour="grey", vjust=-1),
plot.subtitle=element_text(size=18, hjust=0.5, face="italic",
color="grey"),
plot.caption =element_text(size=8, hjust=0.5, face="italic",
color="grey"),
plot.background=element_blank(),
plot.margin = margin(2,2, 2, 4, "cm")) +
transition_states(Any, transition_length = 4, state_length = 1, wrap
= FALSE) +
view_follow(fixed_x = TRUE) +
labs(title = 'Noms de Nens a BCN Any: {closest_state}',
subtitle = "Top 10 Noms del 2004",
caption = "Nombre de nadons | Data Source: Ajuntament de BCN")
> anim
>

```

*PAS 3a: EXPORTAR FRAMES A FITXER GIF*

```

> animate(anim, 200, fps = 20, width = 1200, height = 1000,
    renderer = gifski_renderer("nadonsBCN_2004.gif"), end_pause =
15, start_pause = 15)
>

```

*PAS 3b: EXPORTAR FRAMES A FITXER AVI*

```

> animate(anim, 200, fps = 20, width = 1200, height = 1000,
    renderer = av_renderer("nadonsBCN_2004.avi"), end_pause = 15,
start_pause = 15)
>

```

**4 (1 pt)** Defineix 4 dels següents conceptes en animació, interactivitat, usabilitat i Experiència d'Usuari en Visualització de Dades:

- SUS
- Control
- *CheckBox*
- Participar / Col·laborar

- *Useful*
- Qüestionari Attrakdiff
- UEQ
- Qüestionari SAM

2025  
10-13 JULY



**Visualització de Dades (Enginyeria de Dades - EE - UAB)**  
**Recuperació Examen Segon Parcial - 1 Juliol 2024**  
**SOLUCIONS**

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

NIU: \_\_\_\_\_

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

*Podeu utilitzar Tableau o les llibreries R (plotly, ganimate, shiny, etc.) que creieu convenient i dibuixeu les gràfiques que us facin falta en cada exercici:*

- *Si utilitzeu Tableau, cal incloure una petita explicació del que heu configurat en Tableau i una captura de pantalla de l'aplicació on es vegin les configuracions actives i la gràfica.*
- *Si utilitzeu R, cal incloure les comandes R i una captura de pantalla de la gràfica.*

## PART 1 (3 pts)

*Dataset: Taxa\_Suicidis\_1985-2016.csv*

*Agafarem el dataset de suïcidis en diferents països i franges d'edat entre els anys 1985 i 2016. Cada registre d'aquest dataset conté les següents variables:*

1. pais → Nom del país
2. any → Any del registre
3. sexe → Sexe: Pot agafar els valors: *male, female*
4. edat → Franja d'edat en un interval d'anys
5. num\_suicidis → Nombre de suïcidis per a cada registre
6. poblacio → Nombre d'habitants del país i any
7. num\_suicidis.100k → Índex de Nombre de suïcidis cada 100.000 habitants
8. pib\_per\_any → Producte interior brut del país en dolars
9. pib\_per\_capita → Producte interior brut per capita en dolars
10. generacio → Generació del registre. Variable lligada a la franja d'edat. Pot agafar els valors: *GENERATION X, MILLENNIALS, BOOMERS, SILENT, ...*

*Si necessiteu fer Data Massaging abans de dibuixar la gràfica, expliqueu quines operacions feu. Adjunteu el codi en el cas de R. Per a Tableau adjunteu la captura de tota la pantalla de Tableau, gràfica inclosa.*

### 1.1 (1,75 pt)

En aquest exercici en concret, ens centrarem en les dades de la població entre 25 i 74 anys amb un PIB per capita superior a 60000\$ l'any de la crisi del 2007. Per aquestes dades, mostra un mapa d'arbre (treemap) per la població femenina que et permeti relacionar les diferents generacions amb el seu índex de nombre de suïcidis cada 100.000 habitants i el país d'on són. Fes el mateix per la població masculina (1,25 pt). Un cop tinguis els dos treemaps, contesta les següents preguntes (0,5 pt):

- a) L'índex de nombre de suïcidis cada 100.000 habitants per aquestes dades tenia alguna relació amb la generació de la població?
- b) Quins són els països amb un PIB per capita superior a 60000\$ on diries que l'índex de nombre de suïcidis cada 100.000 habitants era menor?

**RESPOSTA:**

*Càrrega de les llibreries i dataset:*

```
> library(tidyverse)
> library(dplyr)

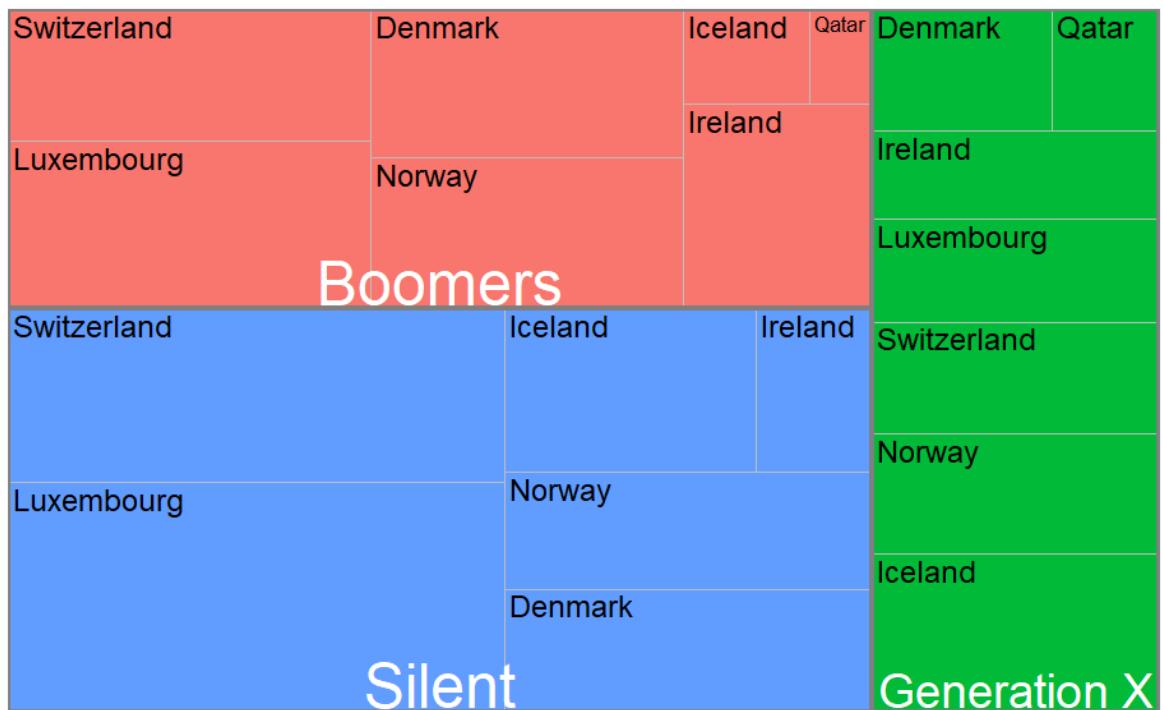
> setwd("C:/Users/...")
> suicidis <- read.csv('./Taxes_Suicidi_1985-2016.csv')
> library (tidyverse)
> library (dplyr)
> library (treemapify)
```

*Datamassage.* Ens quedem amb les dades de l'any 2007 del rang d'edat que ens diuen i pib\_per\_capita que ens indiquen

```
> suicidis07<-suicidis%>%filter(any=='2007')%>%filter(edat == '25-34
years' | edat == '35-54 years' | edat == '55-74
years')%>%filter(pib_per_capita>=60000)
> suicidis07f<-suicidis07%>%filter(sexe=='female')
> suicidis07m<-suicidis07%>%filter(sexe=='male')
(Fins aquí 0,5 pts)
```

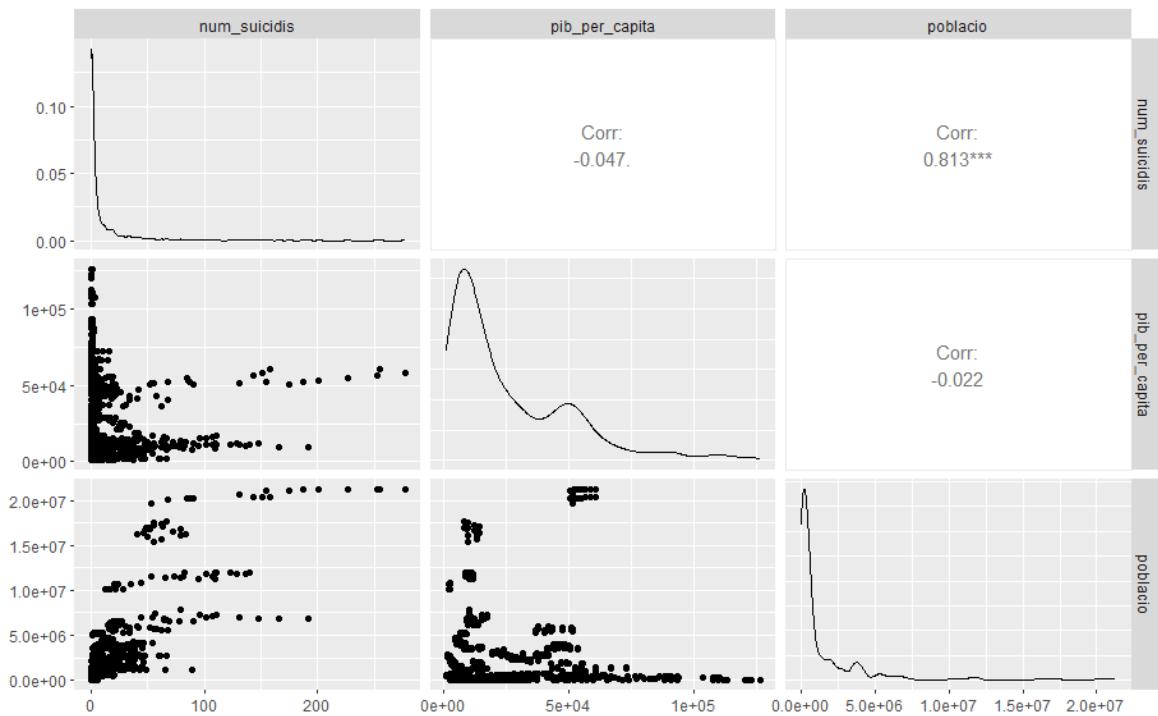
*Treemap.* Ens demanen la informació de tres variables, dues categòriques i una numèrica. La numèrica està clar que l'usarem per fer l'àrea de les caselles del treemap. La informació de les variables categòriques que ens demanen una la donarem a partir de la nostra label i l'altra mitjançant el color. Usem per tant, la variable amb menys categories (generacio) per posar el color (com a variable pare). Finalment posem el país en les labels. **Fer ambdós treemaps val 0,75 pts.**

```
> ggplot (suicidis07f, aes(area=(num_suicidis.100k), fill=generacio,
subgroup=generacio))+geom_treemap()+geom_treemap_subgroup_border()
+geom_treemap_subgroup_text(color='white')+geom_treemap_text
(aes(label=pais))+theme(legend.position = "none")
```



```
> ggplot (suicidis07m, aes(area=(num_suicidis.100k), fill=generacio,
subgroup=generacio))+geom_treemap()+geom_treemap_subgroup_border()
+geom_treemap_subgroup_text(color='white')+geom_treemap_text
(aes(label=pais)) +theme(legend.position = "none")
```





b) La quantitat de població sí que està correlacionada altament i positivament amb el nombre de suïcidis. En canvi la quantitat de població i/o el nombre de suïcidis NO podem dir que estiguin correlacionades amb el PIB\_per\_capita.

## PART 2 (5 pts)

Dataset: Taxa\_Suicidis\_1985-2016.csv

*Agafarem el dataset de suïcidis en diferents països i franges d'edat entre els anys 1985 i 2016. Cada registre d'aquest dataset conté les següents variables:*

1. pais → Nom del país
2. any → Any del registre
3. sexe → Sexe: Pot agafar els valors: *male, female*
4. edat → Franja d'edat en un interval d'anys
5. num\_suicidis → Nombre de suïcidis per a cada registre
6. poblacio → Nombre d'habitants del país i any
7. num\_suicidis.100k → Índex de Nombre de suïcidis cada 100.000 habitants
8. pib\_per\_any → Producte interior brut del país en dolars
9. pib\_per\_capita → Producte interior brut per capita en dolars
10. generacio → Generació del registre. Variable lligada a la franja d'edat. Pot agafar els valors: *GENERATION X, MILLENNIALS, BOOMERS, SILENT, ...*

*Si necessiteu fer Data Massaging abans de dibuixar la gràfica, expliqueu quines operacions feu. Adjunteu el codi en el cas de R. Per a Tableau adjunteu la captura de tota la pantalla de Tableau, gràfica inclosa.*

*Càrrega de les llibreries i dataset:*

```
> library(tidyverse)
> library(dplyr)
> library(plotly)
```

```

> library(shiny)

> setwd("C:/Users/enric/Documents/R")
> SWorld <- read.csv('./Taxes_Suicidi_1985-2016.csv')
> str(SWorld)

```

**2.1. (1 pt)** Mostra el codi i la gràfica de línies de l‘evolució del nombre de suïcidis per anys i generació. Diges quines generacions són capdavanteres i en quin període d’any ho són. En quin any assoleix cada una d’aquestes generacions capdavanteres el màxim de suïcidis i qui és el nombre?

PAS 1: DATA MASSAGING: Seleccionar variables i agrupar per generació i any

```

> SWorldG <- SWorld %>% select(any, generacio, num_suicidis) %>%
  group_by(generacio, any) %>% summarise(numeroS = sum(suicides_no))

```

O bé...

```

> SWorldG <- SWorld %>% group_by(generacio, any) %>% summarise(numeroS =
  sum(suicides_no))

```

# PAS 2a: GRAFICA DE LINIES AMB ggplotly

```

> ggplotG <- ggplot(SWorldG, aes(x=any, y=numeroS, color=generacio)) +
  geom_line()
> ggplotly(ggplotG)

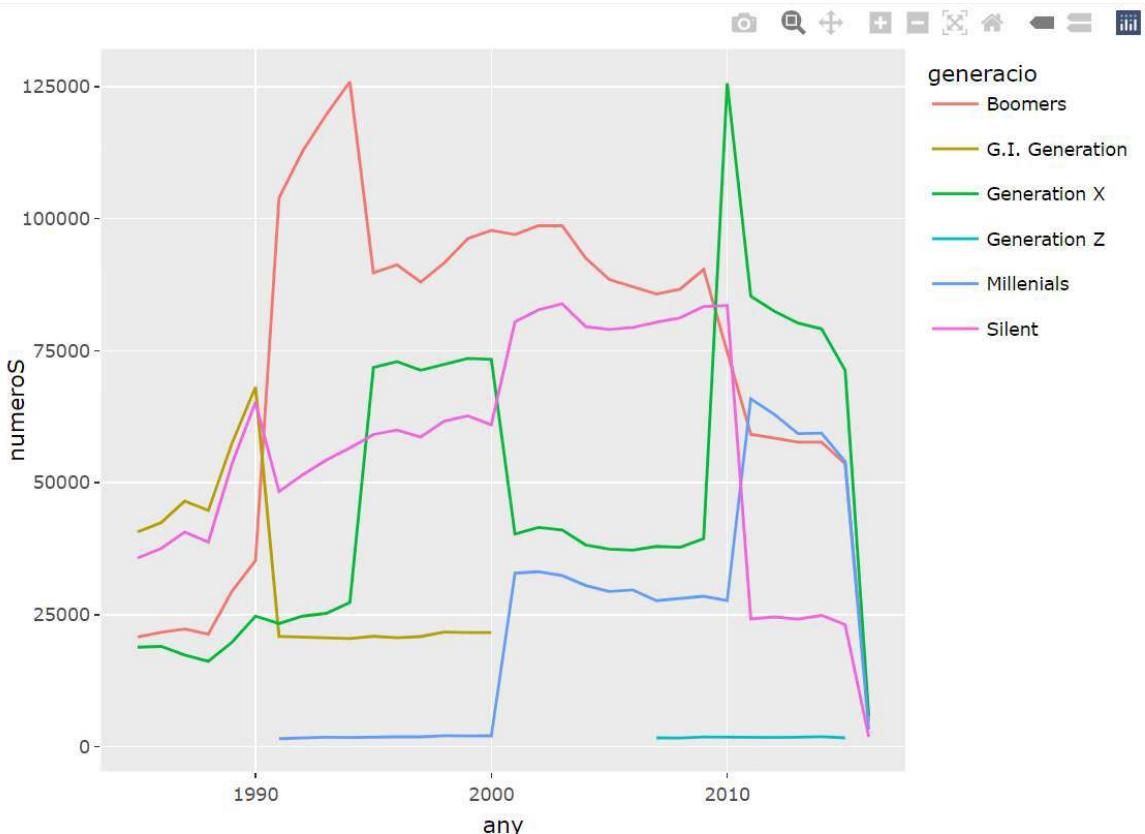
```

# PAS 2b: GRAFICA DE LINIES AMB plotly

```

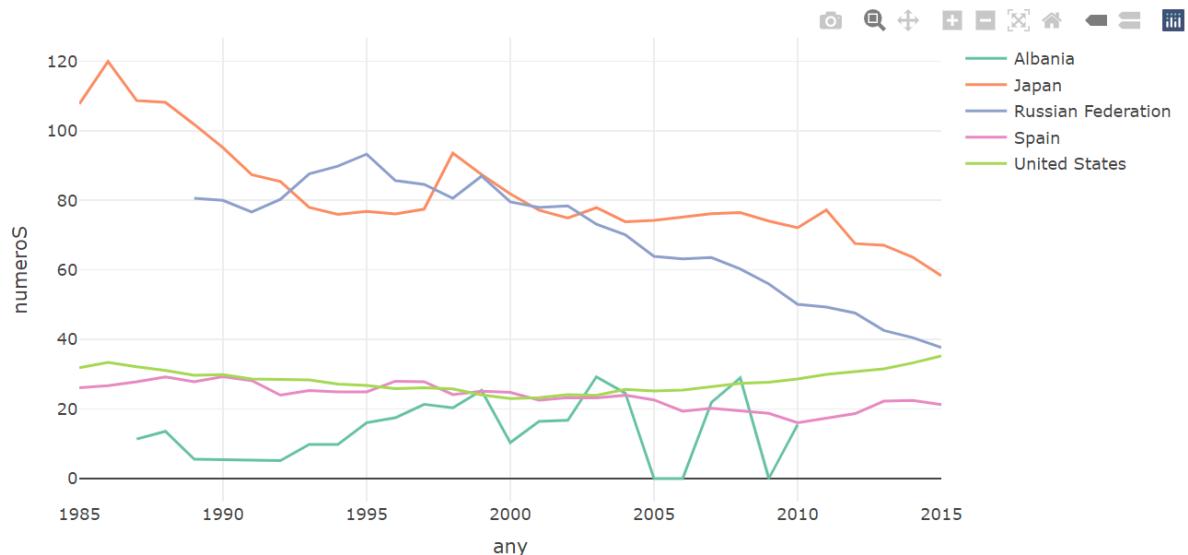
> plot_ly(SWorldG, x=~any, y=~numeroS, color=~generacio) %>% add_lines()

```



Selecciona un País:

Albania Japan Spain United States  
Russian Federation



Per a cada país, busca l'any que ha assolit el màxim índex. Dona l'any i el valor de l'índex:

**RESPOSTA:**

| PAÍS               | ANY  | INDEX SUICIDIS.100K |
|--------------------|------|---------------------|
| ALBANIA            | 2003 | 29,20               |
| SPAIN              | 1990 | 29,27               |
| JAPAN              | 1986 | 119,93              |
| RUSSIAN FEDERATION | 1995 | 93,26               |
| UNITED STATES      | 2015 | 35,23               |

Hi ha algun país que et sorprengui?. Raona la resposta.

**RESPOSTA:**

Japó té un índex alt de suïcidis tenint en compte la seva població respecte Russia o US. Aquest últim té un índex de suïcidis baix respecte Russia.

**2.3. (2 pt)** Mostra el codi per a generar el Ranking de Barres Animades (*Animated Bar Race Ranking*) sobre els 10 països amb més suïcidis de DONES, i el codi per a generar el fitxer GIF. Quina mida té el fitxer generat?. Adjunta en l'examen un parell de captures de pantalla pels anys 1989 i 2014.

# PAS 1: DATA MASSAGING: Seleccionar variables i agrupar per país i any en numero de suicidis

```
> SWorldW <- SWorld %>% filter(sexe == "female") %>% select(any, num_suicidis, pais) %>% group_by(pais, any) %>% summarise(numeroS = sum(num_suicidis))  
>
```

PAS 2: DATA MASSAGING: Seleccionar els 10 països amb nombre / index suicidis més alt i definir nova columna (rank) que ajudara a displaiar

```
> SWorldW_formatted <- SWorldW %>%  
  group_by(any) %>%  
  # The * 1 makes it possible to have non-integer ranks while sliding  
  mutate(rank = rank(-numeroS)) %>%  
  group_by(pais) %>%  
  filter(rank <=10) %>%  
  ungroup()  
>
```

# PAS 3: ANIMATED BAR RACE RANKING

```
> anim <- ggplot(SWorldW_formatted, aes(rank, group = pais,  
   fill = as.factor(pais), color = as.factor(pais))) +  
  geom_tile(aes(y = numeroS/2,  
                height = numeroS,  
                #width = 0.9), alpha = 0.8, color = "black") +  
  width = 0.9), alpha = 0.8, color = NA) +  
  geom_text(aes(y = 0, label = paste(pais, " ")), vjust = 0.2, hjust =  
  1) +  
  geom_text(aes(y=numeroS, label = numeroS, hjust=0)) +  
  coord_flip(clip = "off", expand = FALSE) +  
  scale_y_continuous(labels = scales::comma) +  
  scale_x_reverse() +  
  guides(color = "none", fill = "none") +  
  theme(axis.line=element_blank(),  
        axis.text.x=element_blank(),  
        axis.text.y=element_blank(),  
        axis.ticks=element_blank(),  
        axis.title.x=element_blank(),  
        axis.title.y=element_blank(),  
        legend.position="none",  
        panel.background=element_blank(),  
        panel.border=element_blank(),  
        panel.grid.major=element_blank(),  
        panel.grid.minor=element_blank(),
```

2025  
10-13 JULY



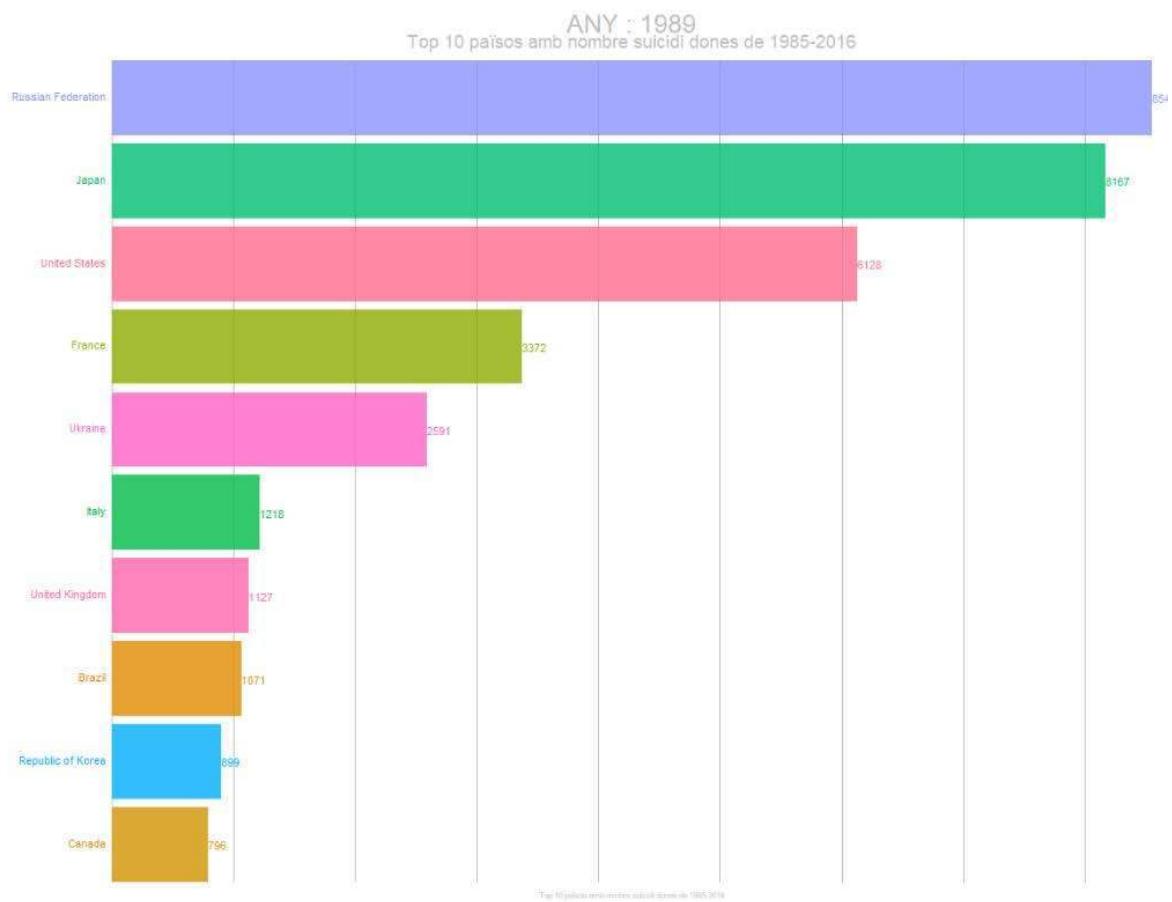
```
panel.grid.major.x = element_line( size=.1, color="grey" ),
panel.grid.minor.x = element_line( size=.1, color="grey" ),
plot.title=element_text(size=25, hjust=0.5, face="bold",
colour="grey", vjust=-1),
plot.subtitle=element_text(size=18, hjust=0.5, face="italic",
color="grey"),
plot.caption =element_text(size=8, hjust=0.5, face="italic",
color="grey"),
plot.background=element_blank(),
plot.margin = margin(2,2, 2, 4, "cm")) +
transition_states(any, transition_length = 4, state_length = 1, wrap
= FALSE) +
view_follow(fixed_x = TRUE) +
labs(title = 'ANY : {closest_state}',
subtitle = "Top 10 països amb nombre suicidi dones de 1985-
2016",
caption = "Top 10 països amb nembre suicidi dones de 1985-2016")
> anim

# PAS 3a. Generar animació (igual que cridar > anim)
> animate(anim, 200, fps = 20, duration= 30, width = 640, height = 480)
# animate(anim, 200, fps = 20, width = 640, height = 480)
>z

# PAS 3b. Exportar frames a fitxer GIF
> animate(anim, 200, fps = 20, duration= 30, width = 1200, height =
1000, renderer = gifski_renderer("Top_10_Suicidis_Dones.gif"),
end_pause = 15, start_pause = 15)
>
```

### RESPOSTA:

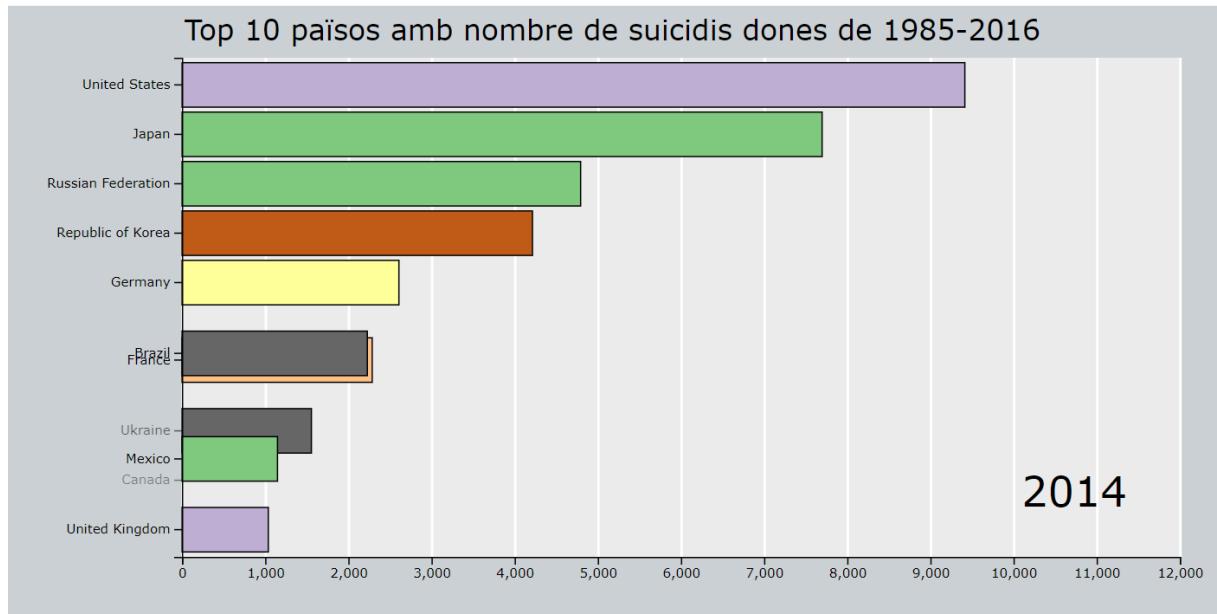
El fitxer ocupa 7,76Kb.



**ANIMATED BAR RACE RANKING. ALTERNATIVA (no available, doncs no s'ha explicat, però es posa per a que ho coneixeu): Llibreria ddplot que defineix una gràfica de forma més senzilla i clara**

```
> install.packages("ddplot") # Càrrega paquet ddplot (primera vegada)
> library(ddplot)           # Generació de Racing Bar Charts

> SWorldW_formatted %>%
  barChartRace(
    x = "numeroS",
    y = "pais",
    time = "any",
    title = "Top 10 països amb index suïcidi dones per 100K 1985-2016"
  )
```



No hem trobat encara la forma de generar fitxer de vídeo amb aquesta funció.

**2.4 (1 pt)** Classifica les següents preguntes o mesures que s'utilitzen en tests d'Usabilitat o d'Experiència d'Usuari (UX) en les següents dues categories:

- a) Instrumentals / Usabilitat / Pragmàtiques
- b) No instrumentals / Hedònics / Emocionals

I digues en quin tipus de test s'inclou cada pregunta o mesura.

#### PREGUNTES:

1. *Conservative 1 2 3 4 5 6 7 Innovative*
2. *I thought the System was easy to use*
3. *The design looks attractive*
4. *Complicated 1 2 3 4 5 6 7 Simple*
5. *This System has all the functions and capabilities I expect it to have*
6. *It was easy to find the information I needed*
7. *Non-inclusive 1 2 3 4 5 6 7 Inclusive*
8. *I found the various functions in this System very well integrated*
9. *The product is creatively designed*
10. *The product is stylish: Strongly Disagree 1 2 3 4 5 6 Strongly Agree*

#### RESPOSTA:

| PREGUNTA | CATEGORIA | TIPUS DE TEST         |
|----------|-----------|-----------------------|
| 1        | b)        | Attrakdiff, Hedonic   |
| 2        | a)        | SUS                   |
| 3        | b)        | meCUE, A2, Aesthetic  |
| 4        | a)        | Attrakdiff, Pragmatic |
| 5        | a)        | PSSUQ                 |
| 6        | a)        | PSSUQ                 |