

2025
10-13 JULY



MAD
COOL
FESTIVAL



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, gganimate, shiny, etc.) que creieu convenientes 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*, *MILLENIALS*, *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)
```

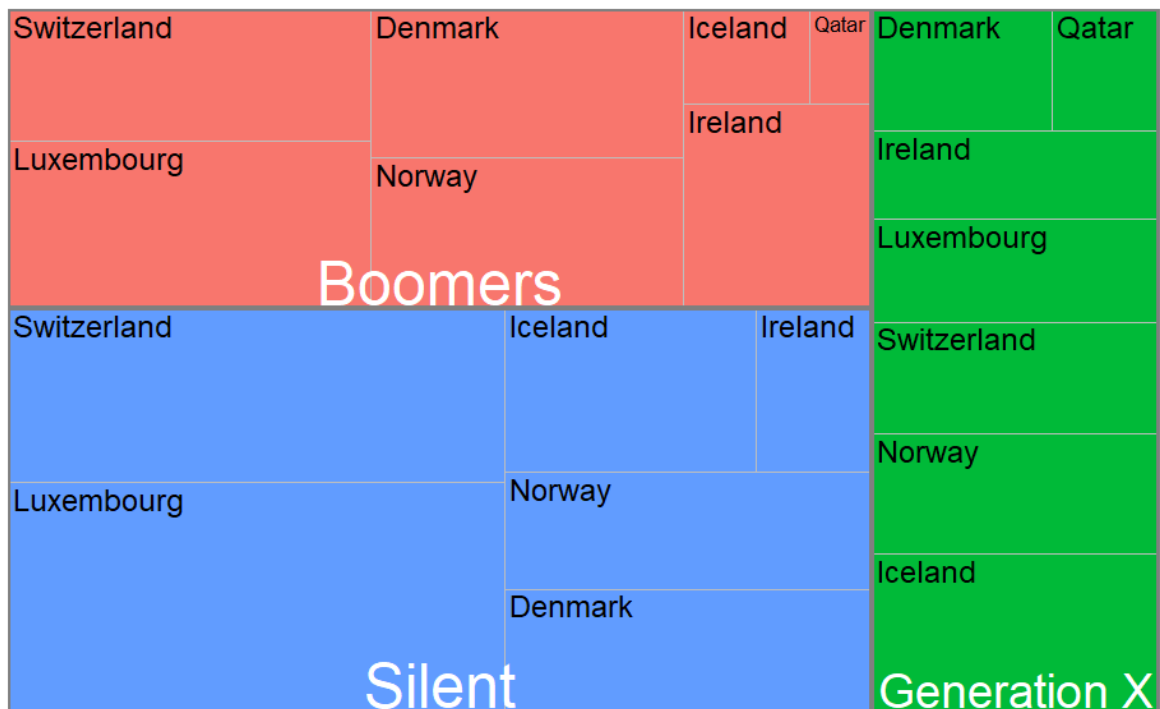
Datamassage. Ens quedem amb les dades de l'any 2007 del rang d'edat que ens diuen i pib_per_capita que ens indiquen

```
> 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')
```

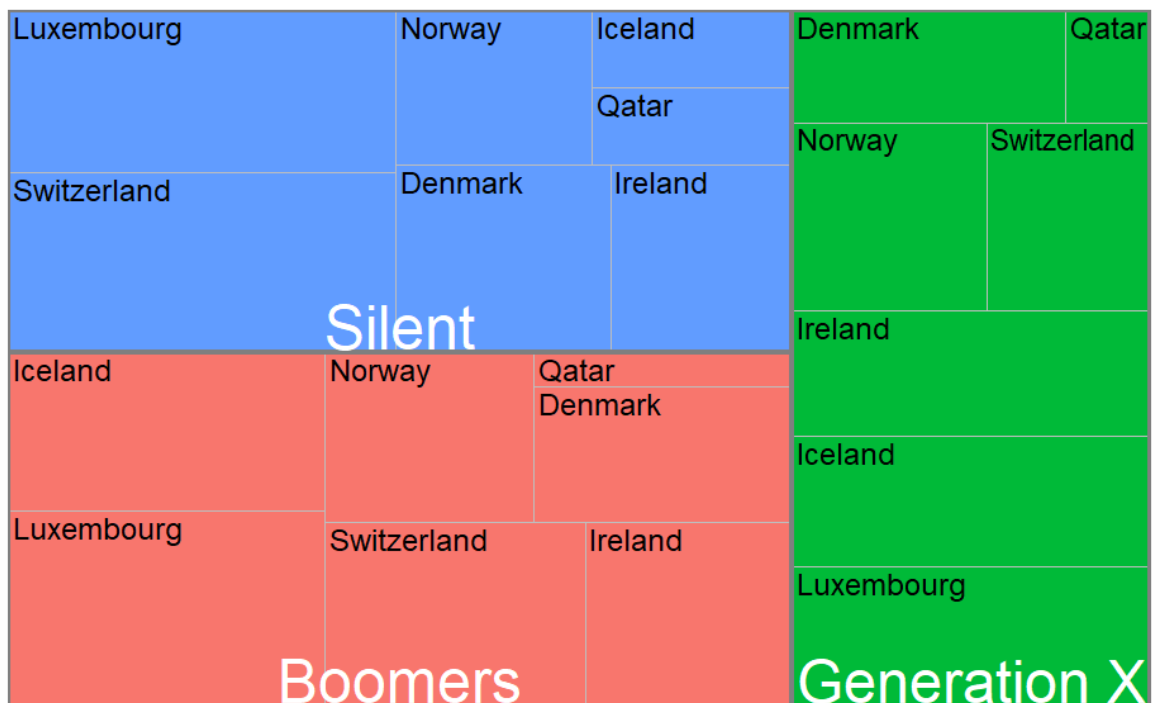
(Fins aquí 0,5 pts)

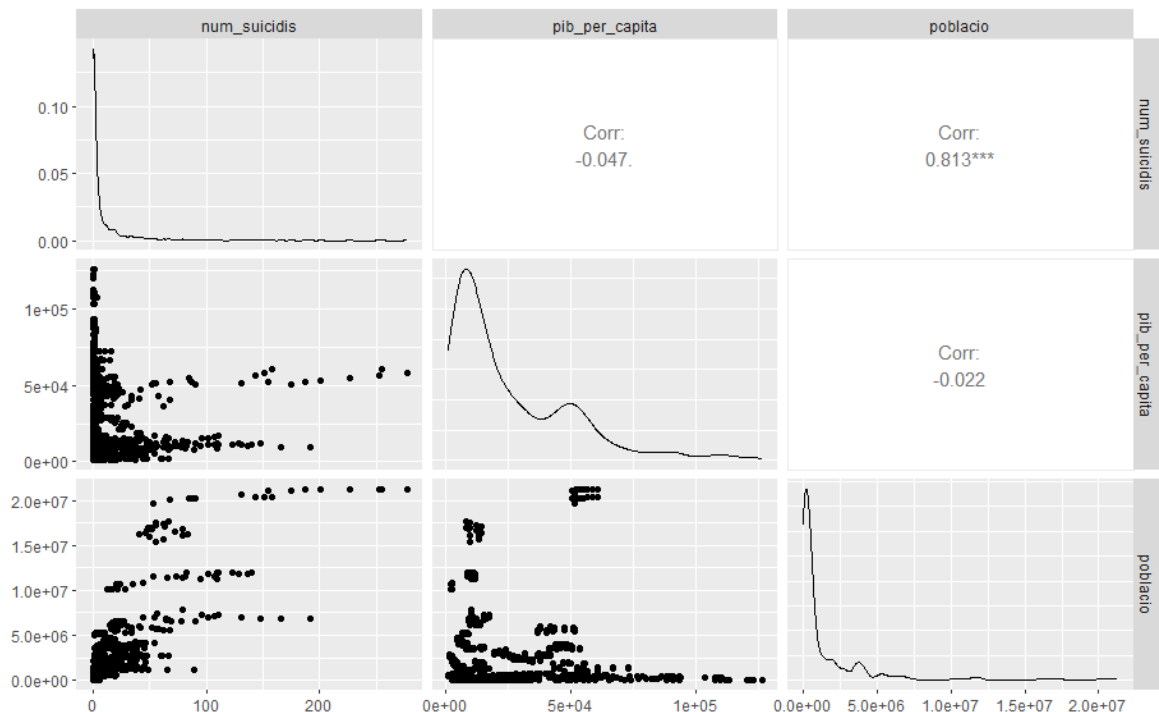
Treemap. Ens demanen la informació de tres variables, dues categòriques i una numèrica. La numèrica està clar que l'usarem per fer l'àrea de les caselles del treemap. La informació de les variables categòriques que ens demanen una la donarem a partir de la nostra label i l'altra mitjançant el color. 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=país))+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. Digues quines generacions són capdavanteres i en quin període d'anys ho són. En quin any assoleix cada una d'aquestes generacions capdavanteres el màxim de suïcidis i quin és el nombre?.

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

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

O bé...

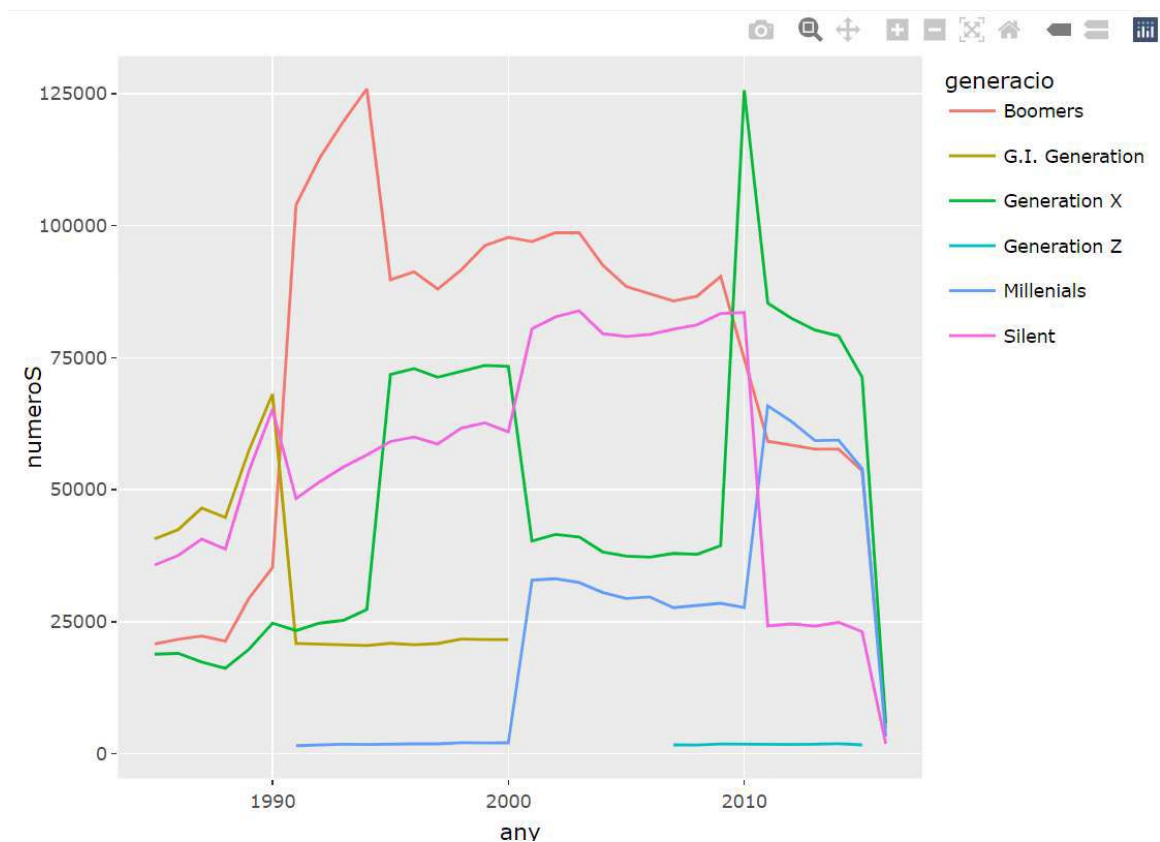
```
> SWorldG <- SWorld %>% group_by(generacio, any) %>% summarise(numeroS =
sum(suicides_no))
```

PAS 2a: GRAFICA DE LINIES AMB ggplotly

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

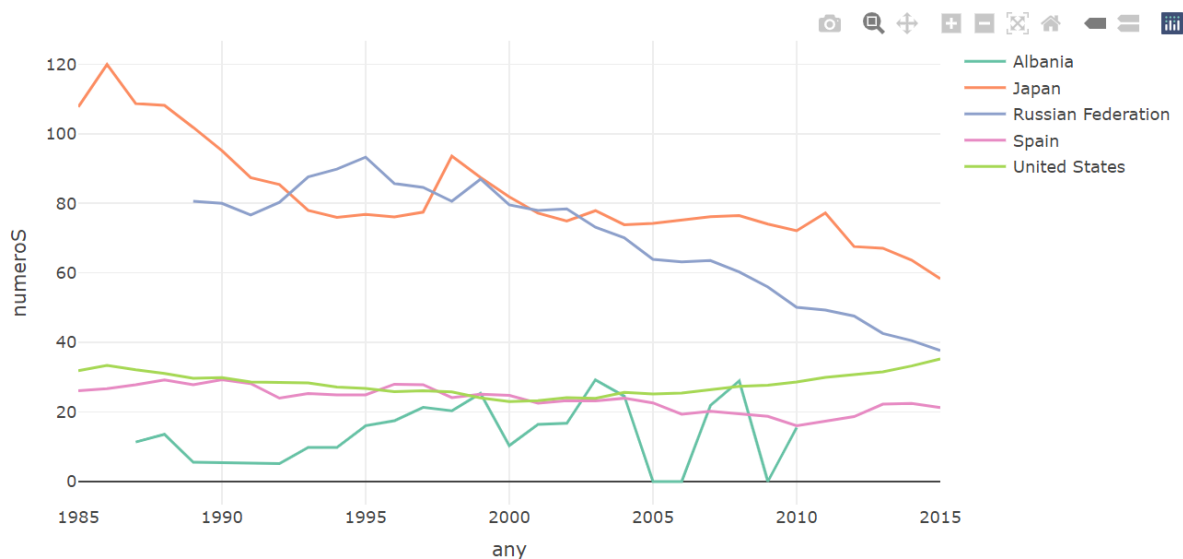
PAS 2b: GRAFICA DE LINIES AMB plotly

```
> 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 Rússia o US. Aquest últim té un índex de suïcidis baix respecte Rússia.

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 suïcidis

```
> 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 suïcidis més alt i definir nova columna (rank) que ajudara a desplaçar

```
> 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(),
```




```

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 suïcidi dones de 1985-
2016",
caption = "Top 10 països amb nombre suïcidi 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

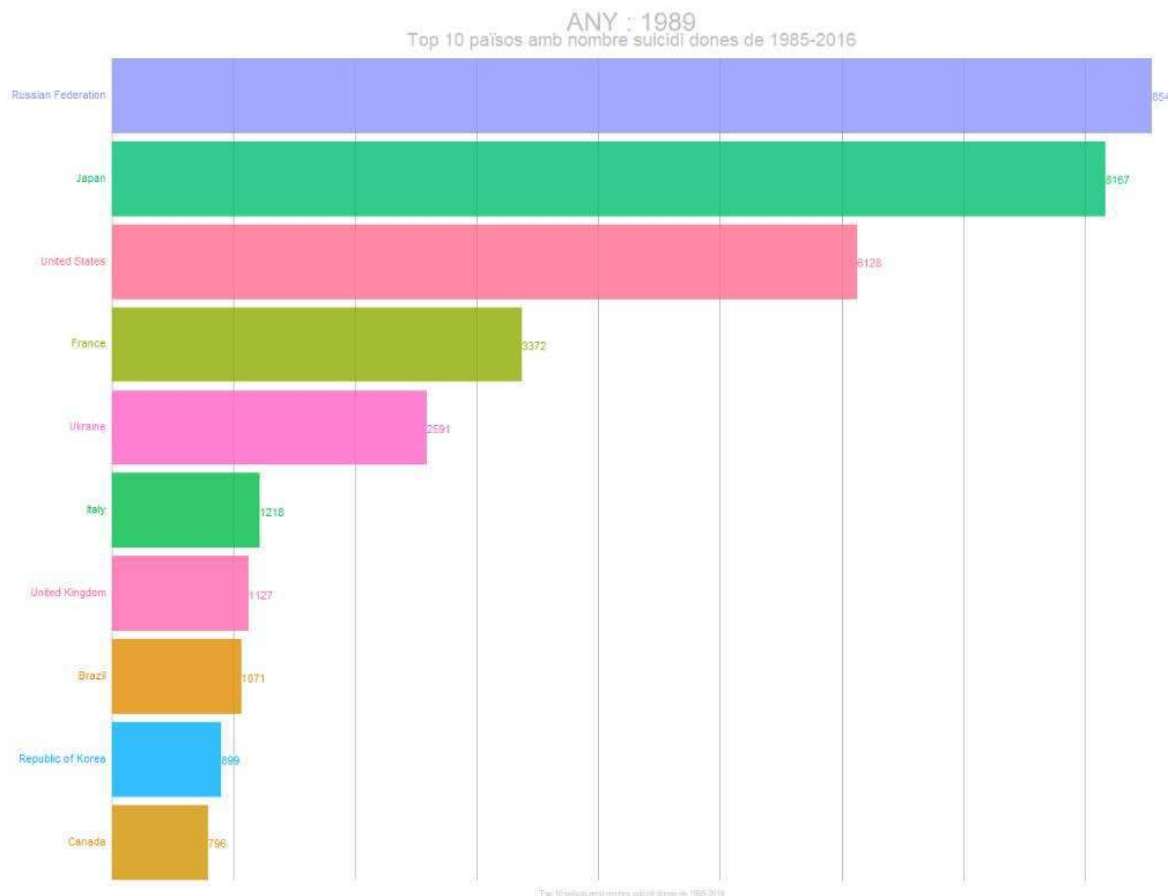
```

> 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:

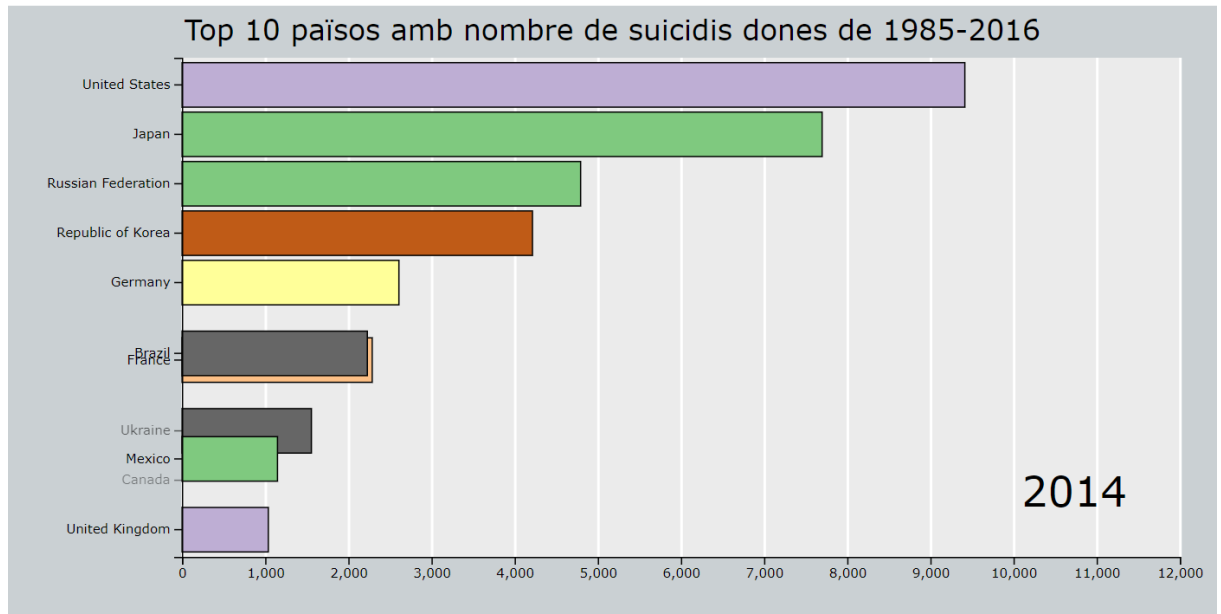
El fitxer ocupa 7,76Kb.



ANIMATED BAR RACE RANKING. ALTERNATIVA (no avaluable, 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