

Solucionario: Recuperatorio (30pts)

Programacion Estadística I

Lic. Alvaro Chirino Gutierrez

Julio, 2020

PRIMER PARCIAL

Pregunta 1 (15 puntos)

Usando la información de www.worldometers.com sobre el COVID con la librería rvest, para los datos mas actuales, generar en un documento R markdown:

- Tabla de países por continente (separadas o juntas)
 - Top 5 casos por millón
 - Top 5 test por millón

```
rm(list=ls())  
library(rvest)
```

```
## Loading required package: xml2
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
covid<-read_html("https://www.worldometers.info/coronavirus/")  
aux<-html_table(covid)  
t<-aux[[1]]  
names(t)[1]<-"id"  
t<-t[!is.na(t$id),]  
names(t)[11]<-"casoM"  
names(t)[14]<-"testM"  
names(t)[12]<-"muerteM"  
names(t)[7]<-"recup"
```

```
t$casoM<-as.numeric(gsub(",","",t$casoM))
```

```
t$testM<-as.numeric(gsub(",","",t$testM))
```

```

t$muerteM<-as.numeric(gsub(",","",t$muerteM))
t$recup<-as.numeric(gsub(",","",t$recup))

## Warning: NAs introducidos por coerción

t <- t %>% filter(16 != "")
#casos
tt1<-t %>% mutate(nn=1) %>% arrange(desc(casoM)) %>% group_by(Continent) %>% mutate(nn=cumsum(nn)) %>%

## Adding missing grouping variables: `Continent`
tt1<-tt1 %>% arrange(Continent)
knitr::kable(tt1,caption = "Top 5 casos por Millon por país y continente")

```

Table 1: Top 5 casos por Millon por país y continente

Continent	Country,Other	casoM	nn
	Diamond Princess	NA	1
	MS Zaandam	NA	2
Africa	Mayotte	10477	1
Africa	South Africa	6875	2
Africa	Djibouti	5088	3
Africa	Cabo Verde	3936	4
Africa	Sao Tome and Principe	3414	5
Asia	Qatar	38551	1
Asia	Bahrain	22293	2
Asia	Kuwait	14476	3
Asia	Oman	14207	4
Asia	Armenia	12202	5
Australia/Oceania	Australia	521	1
Australia/Oceania	New Zealand	311	2
Australia/Oceania	French Polynesia	221	3
Australia/Oceania	New Caledonia	77	4
Australia/Oceania	Fiji	30	5
Europe	San Marino	20598	1
Europe	Vatican City	14981	2
Europe	Andorra	11505	3
Europe	Luxembourg	9499	4
Europe	Sweden	7796	5
North America	Panama	13156	1
North America	USA	12593	2
North America	Dominican Republic	5308	3
North America	Honduras	3642	4
North America	Cayman Islands	3087	5
South America	French Guiana	23692	1
South America	Chile	17712	2
South America	Peru	11246	3
South America	Brazil	10768	4
South America	Bolivia	5490	5

```

#test
tt2<-t %>% mutate(nn=1) %>% arrange(desc(testM)) %>% group_by(Continent) %>% mutate(nn=cumsum(nn)) %>%

## Adding missing grouping variables: `Continent`

```

```
tt2<-tt2 %>% arrange(Continent)
knitr::kable(tt2,caption = "Top 5 test por Millon por país y continente")
```

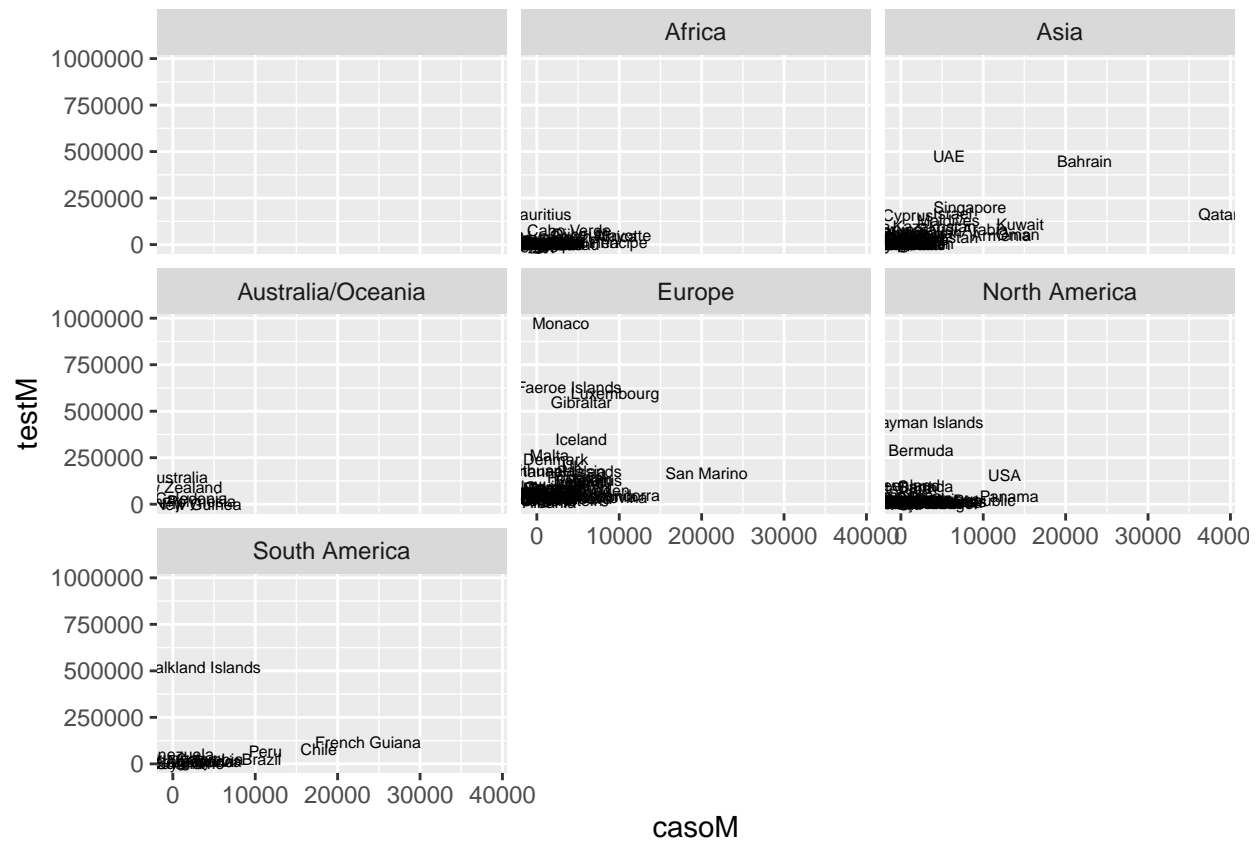
Table 2: Top 5 test por Millon por país y continente

Continent	Country,Other	testM	nn
	Diamond Princess	NA	1
	MS Zaandam	NA	2
Africa	Mauritius	161400	1
Africa	Cabo Verde	74834	2
Africa	Djibouti	56067	3
Africa	Mayotte	47590	4
Africa	South Africa	44347	5
Asia	UAE	472575	1
Asia	Bahrain	446337	2
Asia	Singapore	199900	3
Asia	Israel	166918	4
Asia	Qatar	162332	5
Australia/Oceania	Australia	145408	1
Australia/Oceania	New Zealand	89720	2
Australia/Oceania	New Caledonia	34694	3
Australia/Oceania	French Polynesia	16544	4
Australia/Oceania	Fiji	4460	5
Europe	Monaco	973255	1
Europe	Faeroe Islands	627294	2
Europe	Luxembourg	598913	3
Europe	Gibraltar	550134	4
Europe	Iceland	348886	5
North America	Cayman Islands	438496	1
North America	Bermuda	288422	2
North America	USA	155688	3
North America	Greenland	99426	4
North America	Canada	97910	5
South America	Falkland Islands	521090	1
South America	French Guiana	116350	2
South America	Chile	76578	3
South America	Peru	65420	4
South America	Venezuela	51729	5

- Gráficos de países por continente (incluir en el gráfico los nombres de los países.)
 - Casos por millón vs Test por millón
 - Muertes por millón vs logaritmo de los recuperados

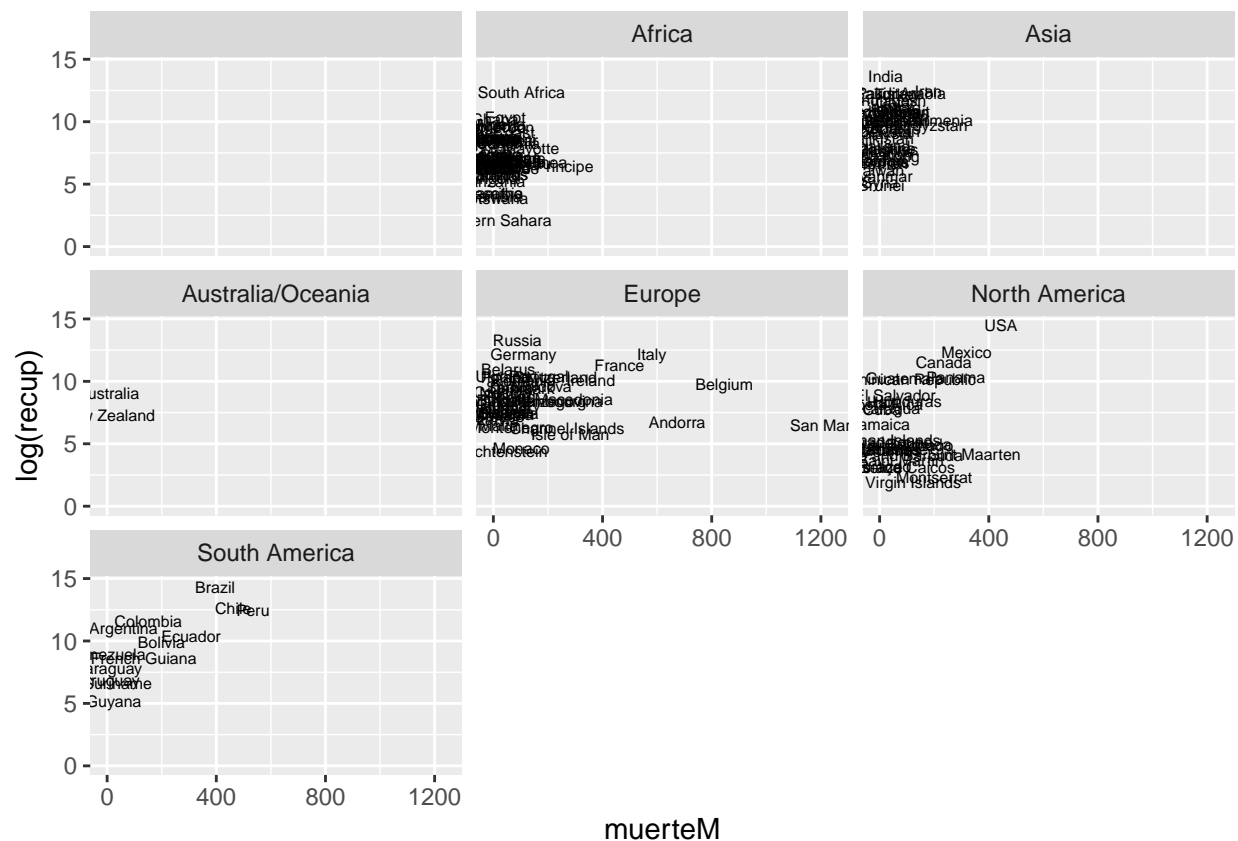
```
library(ggplot2)
#Casos por millón vs Test por millón
ggplot(t,aes(casoM,testM,label=~Country,Other`))+geom_text(size=2)+facet_wrap(~Continent)

## Warning: Removed 20 rows containing missing values (geom_text).
```



```
#Muertes por millón vs logaritmo de los recuperados
ggplot(t,aes(muerteM,log(recup),label=`Country,Other`))+geom_text(size=2)+facet_wrap(~Continent)
```

```
## Warning: Removed 34 rows containing missing values (geom_text).
```



Pregunta 2 (15 puntos)

Realice una función en Shiny empleando la base de datos para la ENDSA que permita ver por departamento y año:

- Tabla que muestre la relación entre sexo y años de educación
- Gráfico que muestre la relación entre “edad a la primera relación sexual” vs “Total de niños nacidos”

SEGUNDO PARCIAL

Pregunta 1 (15 puntos)

Usando la encuesta de hogares 2018 defina una base de datos que contenga las siguientes variables para jefes de hogar mujeres:

- Edad
- Ingreso laboral
- Ingreso no laboral
- Ingreso personal
- Departamento
- Años de educación
- Área
- Tipo de hogar
- Número de miembros en el hogar
- Personas de 5 años o menos en el hogar
- Personas de 60 años o más en el hogar
- Hogar con electricidad
- Acceso a internet en el hogar
- Piso de tierra en la vivienda

(Sugerencia: Use la variable folio y el comando merge para unir bases, para identificar algunas variables se recomienda usar el cuestionario en pdf)

Pregunta 2 (15 puntos)

Para la base alojada en R y la base alojada en Spark (del ejercicio anterior), genere lo siguiente:

- Promedio de años de educación e ingreso personal por departamento, área y acceso a internet
 - Proporción de jefas del hogar con un ingreso personal superior a 3000 Bs. Por departamento y tipo de hogar
 - Promedio de personas de 5 años o menos y personas de 60 años o más, por tipo de hogar, área y piso de tierra.
 - Gráfico sobre el acceso al internet por tipo de hogar (ggplot)
-