

para análisis de datos

Manipulación de datos

Ejercicios

Roxana N. Villafañe | LEMyP |  @data_datum
Florencia D'Andrea | INTA-CONICET |  @cantoflor_87

Slides disponibles en <https://bit.ly/curso-r-fca> 
Página web del curso en <https://curso-r-fca.netlify.com> 

Dataset



Coronavirus 2020

```
library(readr)
data <- read.csv("C://Users/Roxana/cursos-r-analisis-datos/data/2019_nCov_
```

1. Realizamos un análisis exploratorio con las funciones útiles que vimos.



`head()`

`tail()`

`glimpse()`

`str()`

`summary()`

```
head(corona)
```



##	Sno	Date	Province.State	Country	Last.Update	Confir
## 1	1	01/22/2020 12:00:00	Anhui	China	01/22/2020 12:00:00	
## 2	2	01/22/2020 12:00:00	Beijing	China	01/22/2020 12:00:00	
## 3	3	01/22/2020 12:00:00	Chongqing	China	01/22/2020 12:00:00	
## 4	4	01/22/2020 12:00:00	Fujian	China	01/22/2020 12:00:00	
## 5	5	01/22/2020 12:00:00	Gansu	China	01/22/2020 12:00:00	
## 6	6	01/22/2020 12:00:00	Guangdong	China	01/22/2020 12:00:00	
##	Deaths	Recovered				
## 1	0	0				
## 2	0	0				
## 3	0	0				
## 4	0	0				
## 5	0	0				
## 6	0	0				

```
tail(corona)
```



##	Sno	Date	Province.State	Country	Last.Update
##	1414	1414	02/13/2020 21:15:00	Los Angeles, CA	US 2020-02-01 19:53:03
##	1415	1415	02/13/2020 21:15:00	Madison, WI	US 2020-02-05 21:53:02
##	1416	1416	02/13/2020 21:15:00	Orange, CA	US 2020-02-01 19:53:03
##	1417	1417	02/13/2020 21:15:00	San Antonio, TX	US 2020-02-13 18:53:02
##	1418	1418	02/13/2020 21:15:00	Seattle, WA	US 2020-02-09 07:03:04
##	1419	1419	02/13/2020 21:15:00	Tempe, AZ	US 2020-02-01 19:43:03
##		Confirmed	Deaths	Recovered	
##	1414	1	0	0	
##	1415	1	0	0	
##	1416	1	0	0	
##	1417	1	0	0	
##	1418	1	0	1	
##	1419	1	0	0	

```
library(dplyr)
glimpse(corona)
```



```
## Observations: 1,419
## Variables: 8
## $ Sno          <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
## $ Date         <fct> 01/22/2020 12:00:00, 01/22/2020 12:00:00, 01/22/2020
## $ Province.State <fct> Anhui, Beijing, Chongqing, Fujian, Gansu, Guangdong
## $ Country      <fct> China, China, China, China, China, China, China, Ch
## $ Last.Update  <fct> 01/22/2020 12:00:00, 01/22/2020 12:00:00, 01/22/2020
## $ Confirmed    <dbl> 1, 14, 6, 1, 0, 26, 2, 1, 4, 1, 0, 5, 0, 444, 4, 0,
## $ Deaths       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
## $ Recovered    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

```
str(corona)
```



```
## 'data.frame':      1419 obs. of  8 variables:
## $ Sno              : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Date             : Factor w/ 23 levels "01/22/2020 12:00:00",..: 1 1 1 1 1
## $ Province.State  : Factor w/ 62 levels "", "Anhui", "Arizona",..: 2 5 11 14
## $ Country          : Factor w/ 33 levels "Australia", "Belgium",..: 6 6 6 6 6
## $ Last.Update     : Factor w/ 398 levels "01/22/2020 12:00:00",..: 1 1 1 1
## $ Confirmed       : num  1 14 6 1 0 26 2 1 4 1 ...
## $ Deaths          : num  0 0 0 0 0 0 0 0 0 0 ...
## $ Recovered       : num  0 0 0 0 0 0 0 0 0 0 ...
```

summary(corona)



```
##           Sno                Date      Province.State
## Min.      : 1.0   02/13/2020 21:15:00: 74           :374
## 1st Qu.: 355.5   02/11/2020 20:44:00: 73   Anhui       : 23
## Median : 710.0   02/12/2020 22:00:00: 73   Beijing    : 23
## Mean    : 710.0   02/07/2020 20:24:00: 72   Chongqing  : 23
## 3rd Qu.:1064.5   02/08/2020 23:04:00: 72   Fujian     : 23
## Max.    :1419.0   02/09/2020 23:20:00: 72   Gansu      : 23
##                (Other)                :983   (Other)    :930
##           Country                Last.Update      Confirmed
## Mainland China:677   01/31/2020 19:00:00: 63   Min.      : 0.0
## US                 :144   01/30/2020 21:30:00: 59   1st Qu.:  2.0
## Australia          : 68   01/29/2020 21:00:00: 56   Median   : 11.0
## Canada              : 47   01/27/2020 20:30:00: 52   Mean     : 353.1
## China               : 34   01/28/2020 23:00:00: 52   3rd Qu.:  95.0
## Japan               : 23   01/26/2020 23:00:00: 47   Max.    :51986.0
## (Other)            :426   (Other)                :1090
##           Deaths                Recovered
## Min.      : 0.000   Min.      : 0.00
## 1st Qu.: 0.000   1st Qu.: 0.00
## Median   : 0.000   Median   : 0.00
## Mean     : 7.838   Mean     : 24.92
## 3rd Qu.: 0.000   3rd Qu.: 4.00
## Max.    :1426.000   Max.    :4131.00
##
```


2. Pasar todas las variables (columnas) a minúscula, y eliminar la primera columna (Sno)



```
##           date           province.state           country
## 1  01/22/2020 12:00:00           Anhui           China
## 2  01/22/2020 12:00:00           Beijing          China
## 3  01/22/2020 12:00:00      Chongqing          China
## 4  01/22/2020 12:00:00           Fujian           China
## 5  01/22/2020 12:00:00           Gansu           China
## 6  01/22/2020 12:00:00      Guangdong          China
## 7  01/22/2020 12:00:00           Guangxi          China
## 8  01/22/2020 12:00:00           Guizhou          China
## 9  01/22/2020 12:00:00           Hainan          China
## 10 01/22/2020 12:00:00           Hebei           China
## 11 01/22/2020 12:00:00      Heilongjiang          China
## 12 01/22/2020 12:00:00           Henan           China
## 13 01/22/2020 12:00:00           Hong Kong          China
## 14 01/22/2020 12:00:00           Hubei           China
## 15 01/22/2020 12:00:00           Hunan           China
## 16 01/22/2020 12:00:00      Inner Mongolia          China
## 17 01/22/2020 12:00:00           Jiangsu          China
```

```
library(magrittr)
```

```
library(dplyr)
```

```
corona %>%
```

```
  select_all(tolower) %>%
```

```
  select(-sno)
```

##		date	province.state	country
## 1	01/22/2020	12:00:00	Anhui	China
## 2	01/22/2020	12:00:00	Beijing	China
## 3	01/22/2020	12:00:00	Chongqing	China
## 4	01/22/2020	12:00:00	Fujian	China
## 5	01/22/2020	12:00:00	Gansu	China
## 6	01/22/2020	12:00:00	Guangdong	China
## 7	01/22/2020	12:00:00	Guangxi	China
## 8	01/22/2020	12:00:00	Guizhou	China
## 9	01/22/2020	12:00:00	Hainan	China
## 10	01/22/2020	12:00:00	Hebei	China
## 11	01/22/2020	12:00:00	Heilongjiang	China
## 12	01/22/2020	12:00:00	Henan	China
## 13	01/22/2020	12:00:00	Hong Kong	China
## 14	01/22/2020	12:00:00	Hubei	China
## 15	01/22/2020	12:00:00	Hunan	China
## 16	01/22/2020	12:00:00	Inner Mongolia	China
## 17	01/22/2020	12:00:00	Jiangsu	China
## 18	01/22/2020	12:00:00	Jiangxi	China
## 19	01/22/2020	12:00:00	Jilin	China
## 20	01/22/2020	12:00:00	Liaoning	China

3. ¿Cuáles son los lugares con más coronavirus reportados a la fecha?



```
## # A tibble: 74 x 8
## # Groups:   Country [29]
##   Sno Date Province.State Country Last.Update Confirmed Deaths Rec
##   <int> <fct> <fct> <fct> <fct> <dbl> <dbl>
## 1 1346 02/13/~ Hubei Mainlan~ 2020-02-14 ~ 51986 1426
## 2 1347 02/13/~ Guangdong Mainlan~ 2020-02-14 ~ 1261 2
## 3 1348 02/13/~ Henan Mainlan~ 2020-02-14 ~ 1184 11
## 4 1349 02/13/~ Zhejiang Mainlan~ 2020-02-14 ~ 1155 0
## 5 1350 02/13/~ Hunan Mainlan~ 2020-02-14 ~ 988 2
## 6 1351 02/13/~ Anhui Mainlan~ 2020-02-14 ~ 934 6
## 7 1352 02/13/~ Jiangxi Mainlan~ 2020-02-14 ~ 900 1
## 8 1353 02/13/~ Jiangsu Mainlan~ 2020-02-14 ~ 593 0
## 9 1354 02/13/~ Chongqing Mainlan~ 2020-02-13 ~ 529 4
## 10 1355 02/13/~ Shandong Mainlan~ 2020-02-14 ~ 519 2
## # ... with 64 more rows
```

```
corona %>%
  group_by(Country) %>%
  filter(Date=="02/13/2020 21:15:00") %>%
  arrange(desc(Confirmed))
```

```
## # A tibble: 74 x 8
## # Groups:   Country [29]
##   Sno Date Province.State Country Last.Update Confirmed Deaths Rec
##   <int> <fct> <fct> <fct> <fct> <dbl> <dbl>
## 1 1346 02/13/~ Hubei Mainlan~ 2020-02-14 ~ 51986 1426
## 2 1347 02/13/~ Guangdong Mainlan~ 2020-02-14 ~ 1261 2
## 3 1348 02/13/~ Henan Mainlan~ 2020-02-14 ~ 1184 11
## 4 1349 02/13/~ Zhejiang Mainlan~ 2020-02-14 ~ 1155 0
## 5 1350 02/13/~ Hunan Mainlan~ 2020-02-14 ~ 988 2
## 6 1351 02/13/~ Anhui Mainlan~ 2020-02-14 ~ 934 6
## 7 1352 02/13/~ Jiangxi Mainlan~ 2020-02-14 ~ 900 1
## 8 1353 02/13/~ Jiangsu Mainlan~ 2020-02-14 ~ 593 0
## 9 1354 02/13/~ Chongqing Mainlan~ 2020-02-13 ~ 529 4
## 10 1355 02/13/~ Shandong Mainlan~ 2020-02-14 ~ 519 2
## # ... with 64 more rows
```

4. ¿Cuántos casos hay reportados en Taiwan?



##	Sno	Date	Province.State	Country	Last.Update
## 1	67	01/23/2020 12:00:00	Taiwan	Taiwan	01/23/2020 12:00:00
## 2	107	01/24/2020 12:00:00	Taiwan	Taiwan	01/24/2020 12:00:00
## 3	155	01/25/2020 22:00:00	Taiwan	Taiwan	01/25/2020 12:00:00
## 4	200	01/26/2020 23:00:00	Taiwan	Taiwan	01/26/2020 23:00:00
## 5	247	01/27/2020 20:30:00	Taiwan	Taiwan	01/27/2020 20:30:00
## 6	298	01/28/2020 23:00:00	Taiwan	Taiwan	01/28/2020 23:00:00
## 7	350	01/29/2020 21:00:00	Taiwan	Taiwan	01/29/2020 21:00:00
## 8	406	01/30/2020 21:30:00	Taiwan	Taiwan	01/30/2020 21:30:00
## 9	465	01/31/2020 19:00:00	Taiwan	Taiwan	01/31/2020 19:00:00
## 10	532	02/01/2020 23:00:00	Taiwan	Taiwan	1/31/2020 15:20
## 11	601	02/02/2020 21:00:00	Taiwan	Taiwan	1/31/2020 15:20
## 12	668	02/03/2020 21:40:00	Taiwan	Taiwan	1/31/20 15:20
## 13	737	02/04/2020 22:00:00	Taiwan	Taiwan	2020-04-02 14:53:00
## 14	808	02/05/2020 12:20:00	Taiwan	Taiwan	2020-04-02 14:53:00
## 15	876	02/06/2020 20:05:00	Taiwan	Taiwan	2020-06-02 15:03:00
## 16	948	02/07/2020 20:24:00	Taiwan	Taiwan	2020-06-02 15:03:00
## 17	1020	02/08/2020 23:04:00	Taiwan	Taiwan	2020-08-02 05:23:00
## 18	1093	02/09/2020 23:20:00	Taiwan	Taiwan	2020-02-09 09:13:11
## 19	1165	02/10/2020 19:30:00	Taiwan	Taiwan	2020-02-09 09:13:11

4. ¿Cuántos casos hay reportados en Taiwan?



```
corona %>%  
  filter(Country == "Taiwan")
```

##	Sno	Date	Province.State	Country	Last.Update
## 1	67	01/23/2020 12:00:00	Taiwan	Taiwan	01/23/2020 12:00:00
## 2	107	01/24/2020 12:00:00	Taiwan	Taiwan	01/24/2020 12:00:00
## 3	155	01/25/2020 22:00:00	Taiwan	Taiwan	01/25/2020 12:00:00
## 4	200	01/26/2020 23:00:00	Taiwan	Taiwan	01/26/2020 23:00:00
## 5	247	01/27/2020 20:30:00	Taiwan	Taiwan	01/27/2020 20:30:00
## 6	298	01/28/2020 23:00:00	Taiwan	Taiwan	01/28/2020 23:00:00
## 7	350	01/29/2020 21:00:00	Taiwan	Taiwan	01/29/2020 21:00:00
## 8	406	01/30/2020 21:30:00	Taiwan	Taiwan	01/30/2020 21:30:00
## 9	465	01/31/2020 19:00:00	Taiwan	Taiwan	01/31/2020 19:00:00
## 10	532	02/01/2020 23:00:00	Taiwan	Taiwan	1/31/2020 15:20
## 11	601	02/02/2020 21:00:00	Taiwan	Taiwan	1/31/2020 15:20
## 12	668	02/03/2020 21:40:00	Taiwan	Taiwan	1/31/20 15:20
## 13	737	02/04/2020 22:00:00	Taiwan	Taiwan	2020-04-02 14:53:00
## 14	808	02/05/2020 12:20:00	Taiwan	Taiwan	2020-04-02 14:53:00
## 15	876	02/06/2020 20:05:00	Taiwan	Taiwan	2020-06-02 15:03:00
## 16	948	02/07/2020 20:24:00	Taiwan	Taiwan	2020-06-02 15:03:00

5. Después de China, ¿cuál es el segundo país con casos de Coronavirus?



```
## # A tibble: 43 x 8
## # Groups:   Country [28]
##   Sno Date Province.State Country Last.Update Confirmed Deaths Rec
##   <int> <fct> <fct> <fct> <fct> <dbl> <dbl>
## 1 1364 02/13/~ "Diamond Prince~ Others 2020-02-14~ 218 0
## 2 1376 02/13/~ "" Singap~ 2020-02-13~ 58 0
## 3 1377 02/13/~ "Hong Kong" Hong K~ 2020-02-13~ 53 1
## 4 1378 02/13/~ "" Thaila~ 2020-02-13~ 33 0
## 5 1379 02/13/~ "" Japan 2020-02-13~ 28 1
## 6 1380 02/13/~ "" South ~ 2020-02-12~ 28 0
## 7 1381 02/13/~ "" Malays~ 2020-02-13~ 19 0
## 8 1383 02/13/~ "Taiwan" Taiwan 2020-02-09~ 18 0
## 9 1384 02/13/~ "" Germany 2020-02-13~ 16 0
## 10 1385 02/13/~ "" Vietnam 2020-02-13~ 16 0
## # ... with 33 more rows
```

```
corona %>%
```

```
  group_by(Country) %>%
```

```
  filter(Date=="02/13/2020 21:15:00", Country!="Mainland China")%>%
```

```
  arrange(desc(Confirmed))
```

```
## # A tibble: 43 x 8
```

```
## # Groups:   Country [28]
```

##	Sno	Date	Province.State	Country	Last.Update	Confirmed	Deaths	Rec
##	<int>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	
##	1	1364	02/13/~	"Diamond Prince~	Others	2020-02-14~	218	0
##	2	1376	02/13/~	"	Singap~	2020-02-13~	58	0
##	3	1377	02/13/~	"Hong Kong"	Hong K~	2020-02-13~	53	1
##	4	1378	02/13/~	"	Thaila~	2020-02-13~	33	0
##	5	1379	02/13/~	"	Japan	2020-02-13~	28	1
##	6	1380	02/13/~	"	South ~	2020-02-12~	28	0
##	7	1381	02/13/~	"	Malays~	2020-02-13~	19	0
##	8	1383	02/13/~	"Taiwan"	Taiwan	2020-02-09~	18	0
##	9	1384	02/13/~	"	Germany	2020-02-13~	16	0
##	10	1385	02/13/~	"	Vietnam	2020-02-13~	16	0

```
## # ... with 33 more rows
```


6. Después de Hubei, ¿cuáles son las provincias más afectadas?



```
## # A tibble: 74 x 8
## # Groups:   Province.State [54]
##       Sno Date      Province.State Country Last.Update Confirmed Deaths Rec
##   <int> <fct>    <fct>          <fct> <fct>      <dbl> <dbl>
## 1  1346 02/13/~ Hubei          Mainlan~ 2020-02-14 ~ 51986 1426
## 2  1347 02/13/~ Guangdong     Mainlan~ 2020-02-14 ~ 1261 2
## 3  1348 02/13/~ Henan         Mainlan~ 2020-02-14 ~ 1184 11
## 4  1349 02/13/~ Zhejiang     Mainlan~ 2020-02-14 ~ 1155 0
## 5  1350 02/13/~ Hunan        Mainlan~ 2020-02-14 ~ 988 2
## 6  1351 02/13/~ Anhui        Mainlan~ 2020-02-14 ~ 934 6
## 7  1352 02/13/~ Jiangxi      Mainlan~ 2020-02-14 ~ 900 1
## 8  1353 02/13/~ Jiangsu     Mainlan~ 2020-02-14 ~ 593 0
## 9  1354 02/13/~ Chongqing   Mainlan~ 2020-02-13 ~ 529 4
## 10 1355 02/13/~ Shandong    Mainlan~ 2020-02-14 ~ 519 2
## # ... with 64 more rows
```

```
corona %>%
```

```
  group_by(Province.State) %>%
```

```
  filter(Date=="02/13/2020 21:15:00") %>%
```

```
  arrange(desc(Confirmed))
```

```
## # A tibble: 74 x 8
```

```
## # Groups:   Province.State [54]
```

##		Sno	Date	Province.State	Country	Last.Update	Confirmed	Deaths	Rec
##		<int>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	
##	1	1346	02/13/~	Hubei	Mainlan~	2020-02-14 ~	51986	1426	
##	2	1347	02/13/~	Guangdong	Mainlan~	2020-02-14 ~	1261	2	
##	3	1348	02/13/~	Henan	Mainlan~	2020-02-14 ~	1184	11	
##	4	1349	02/13/~	Zhejiang	Mainlan~	2020-02-14 ~	1155	0	
##	5	1350	02/13/~	Hunan	Mainlan~	2020-02-14 ~	988	2	
##	6	1351	02/13/~	Anhui	Mainlan~	2020-02-14 ~	934	6	
##	7	1352	02/13/~	Jiangxi	Mainlan~	2020-02-14 ~	900	1	
##	8	1353	02/13/~	Jiangsu	Mainlan~	2020-02-14 ~	593	0	
##	9	1354	02/13/~	Chongqing	Mainlan~	2020-02-13 ~	529	4	
##	10	1355	02/13/~	Shandong	Mainlan~	2020-02-14 ~	519	2	

```
## # ... with 64 more rows
```

7. ¿Cuántos casos hay en Australia reportados según regiones? Ordenar de manera descendente



```
## # A tibble: 4 x 8
## # Groups:   Country [1]
##   Sno Date      Province.State Country Last.Update Confirmed Deaths Rec
##   <int> <fct>      <fct>          <fct> <fct>          <dbl> <dbl>
## 1  1390 02/13/2~ Queensland     Austr~ 2020-02-09 ~      5      0
## 2  1391 02/13/2~ New South Wales Austr~ 2020-02-13 ~      4      0
## 3  1392 02/13/2~ Victoria       Austr~ 2020-02-13 ~      4      0
## 4  1397 02/13/2~ South Australia Austr~ 2020-02-02 ~      2      0
```

```
corona %>%
  group_by(Province.State) %>%
  filter(Date=="02/13/2020 21:15:00", Country=="Australia") %>%
  arrange(desc(Confirmed))
```

```
## # A tibble: 4 x 8
```

```
## # Groups:   Province.State [4]
```

##	Sno	Date	Province.State	Country	Last.Update	Confirmed	Deaths	Rec
##	<int>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	
## 1	1390	02/13/2~	Queensland	Austra~	2020-02-09 ~	5	0	
## 2	1391	02/13/2~	New South Wales	Austra~	2020-02-13 ~	4	0	
## 3	1392	02/13/2~	Victoria	Austra~	2020-02-13 ~	4	0	
## 4	1397	02/13/2~	South Australia	Austra~	2020-02-02 ~	2	0	