

Series de Tiempo - Actividad 202

Edgar Rios

9/10/2021

Materia: Modelos Avanzados en Ciencia de Datos

Profesor: Dr. Juliho Castillo Colmenares

Alumno: Edgar Rios Linares

Actividad:

1. Analiza el capítulo 2 de nuestro libro de texto. <https://otexts.com/fpp3/graphics.html#graphics>
2. Descarga el archivo international-tourism-...csv <https://ourworldindata.org/grapher/international-tourism-number-of-arrivals?tab=chart&country=~MEX>
3. Calcula el máximo de arribos internacionales anuales por país
4. Determina los cinco países con el indicador anterior más alto y el año en que este ocurrió.

Cargamos las librerías

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.4      v dplyr   1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Leemos los datos *Nota: Del archivo original se modifíco la columna “International tourism, number of arrivals” por “Arrivals” para un manejo más sencillo en las operaciones del dataframe

```
data_tourism <- read.csv("international-tourism-number-of-arrivals.csv")
```

Revisamos los datos

```
head(data_tourism)
```

```
##           Entity Code Year Arrivals
## 1 Africa Eastern and Southern      1995 11583545
## 2 Africa Eastern and Southern      1996 13088654
## 3 Africa Eastern and Southern      1997 13456246
## 4 Africa Eastern and Southern      1998 14403852
## 5 Africa Eastern and Southern      1999 15309378
## 6 Africa Eastern and Southern      2000 15353177
```

Filtramos los registros donde “Code” es nulo Guardamos el resultado en arrivals_country_year

```
arrivals_country_year <- data_tourism %>%
  filter(Code != " ")
```

Filtramos los registros donde “Entity” es igual a “World” Agrupamos por “Entity” (país) Agregamos “my_ranks” para clasificar por “Entity” y “Arrivals” (país y número de arribos) Se obtiene una clasificación por cada país y sus números de arribos. Se realiza una división para pasar a millones los arribos “TotalArrivals”

```
arrivals_country_year <- arrivals_country_year %>%
  filter(Entity != 'World') %>%
  group_by(Entity) %>%
  mutate(TotalArrivals = Arrivals / 1e6) %>%
  mutate(my_ranks = order (order(Arrivals, decreasing=TRUE)) )
```

Revisamos los datos. Por cada país se tiene una clasificación “my_ranks” de los arribos “Arrivals”(de manera descendente) Cada país tiene una clasificación del 1 hasta N, considerando el número de Arribos “Arrivals” 1 representa el año con mayor arribos, 2 el segundo registro con más arribos y así sucesivamente

```
arrivals_country_year
```

```
## # A tibble: 4,541 x 6
## # Groups:   Entity [203]
##   Entity Code   Year Arrivals TotalArrivals my_ranks
##   <chr>   <chr> <int>     <dbl>         <dbl>     <int>
## 1 Albania ALB   1995     304000         0.304         22
## 2 Albania ALB   1996     287000         0.287         23
## 3 Albania ALB   1997     119000         0.119         25
## 4 Albania ALB   1998     184000         0.184         24
## 5 Albania ALB   1999     371000         0.371         19
## 6 Albania ALB   2000     317000         0.317         21
## 7 Albania ALB   2001     354000         0.354         20
## 8 Albania ALB   2002     470000         0.47          18
## 9 Albania ALB   2003     557000         0.557         17
## 10 Albania ALB   2004     645000         0.645         16
## # ... with 4,531 more rows
```

Seleccionamos los registros donde “my_ranks” es igual a 1.

```
arrivals_country_year <- arrivals_country_year %>%
  filter(my_ranks == 1)
```

Revisamos los datos Y observamos que sólo tenemos los registros de cada país (Entity) del año(year) en donde tuvo la mayor cantidad de arribos (Arrivals)

```
arrivals_country_year
```

```
## # A tibble: 203 x 6
## # Groups:   Entity [203]
##   Entity      Code  Year Arrivals TotalArrivals my_ranks
##   <chr>      <chr> <int>    <dbl>      <dbl>    <int>
## 1 Albania    ALB    2019  6406000      6.41        1
## 2 Algeria    DZA    2013  2733000      2.73        1
## 3 American Samoa ASM    2019   52700.      0.0527       1
## 4 Andorra    AND    2004 11668000     11.7        1
## 5 Angola     AGO    2013   650000      0.65        1
## 6 Antigua and Barbuda ATG    2018  1064000      1.06        1
## 7 Argentina  ARG    2019  7399000      7.40        1
## 8 Armenia    ARM    2019  1894000      1.89        1
## 9 Aruba      ABW    2019  1951000      1.95        1
## 10 Australia AUS    2019  9466000      9.47        1
## # ... with 193 more rows
```

Ordenamos de forma descendente los arribos (Arrivals) para obtener los países (Entity) con el año (Year) en donde tuvieron más arribos (Arrivals)

```
arrivals_country_year[order(arrivals_country_year$Arrivals, decreasing = TRUE),] -> arrivals_country_year_order
```

Revisamos los datos

```
arrivals_country_year_order
```

```
## # A tibble: 203 x 6
## # Groups:   Entity [203]
##   Entity      Code  Year Arrivals TotalArrivals my_ranks
##   <chr>      <chr> <int>    <dbl>      <dbl>    <int>
## 1 France    FRA    2018 211998000     212.        1
## 2 United States USA    2006 183178000     183.        1
## 3 China     CHN    2019 162538000     163.        1
## 4 Spain     ESP    2019 126170000     126.        1
## 5 Mexico    MEX    2000 105673000     106.        1
## 6 Italy     ITA    2019  95399000      95.4        1
## 7 Poland    POL    1999  89118000      89.1        1
## 8 Hong Kong HKG    2018  65148000      65.1        1
## 9 Hungary   HUN    2019  61397000      61.4        1
## 10 Croatia  HRV    2019  60021000      60.0        1
## # ... with 193 more rows
```

Generamos el dataframe con las respuestas requeridas Código del país (Code), Número máximo de arribos anuales (TotalArrivals), Año de Ocurrencia (Year) “Entity” se debe agregar para que R no genere error

```
arrivals_country_year_order <- arrivals_country_year_order %>%
  select(Entity, Code, TotalArrivals, Year)
```

Revisamos los datos con nuestra respuesta

```
head(arrivals_country_year_order, n=5)
```

```
## # A tibble: 5 x 4
## # Groups:   Entity [5]
##   Entity      Code TotalArrivals Year
##   <chr>      <chr>         <dbl> <int>
## 1 France      FRA             212.  2018
## 2 United States USA             183.  2006
## 3 China       CHN             163.  2019
## 4 Spain       ESP             126.  2019
## 5 Mexico      MEX             106.  2000
```