

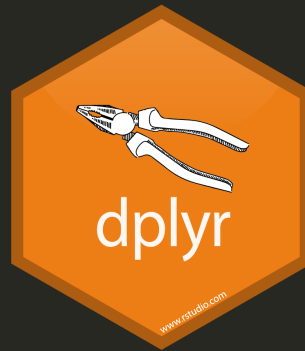
Manipulación de datos


con dplyr

Curso Aprender R desde cero

FACENA UNNE

26 de Agosto de 2019



Ejercicios

Tener los paquetes cargados



```
library(dplyr)
library(magrittr)
library(gapminder)
```

1. De gapminder seleccionar los datos correspondientes a Argentina.



```
## # A tibble: 12 x 6
##   country    continent  year lifeExp      pop gdpPercap
##   <fct>      <fct>      <int>  <dbl>    <int>    <dbl>
## 1 Argentina Americas    1952   62.5  17876956   5911.
## 2 Argentina Americas    1957   64.4  19610538   6857.
## 3 Argentina Americas    1962   65.1  21283783   7133.
## 4 Argentina Americas    1967   65.6  22934225   8053.
## 5 Argentina Americas    1972   67.1  24779799   9443.
## 6 Argentina Americas    1977   68.5  26983828  10079.
## 7 Argentina Americas    1982   69.9  29341374   8998.
## 8 Argentina Americas    1987   70.8  31620918   9140.
## 9 Argentina Americas    1992   71.9  33958947   9308.
## 10 Argentina Americas    1997   73.3  36203463  10967.
## 11 Argentina Americas    2002   74.3  38331121   8798.
## 12 Argentina Americas    2007   75.3  40301927  12779.
```

Solución



```
gapminder %>%  
  filter(country=="Argentina")
```

```
## # A tibble: 12 x 6  
##   country    continent  year lifeExp      pop gdpPercap  
##   <fct>      <fct>      <int>  <dbl>    <int>    <dbl>  
## 1 Argentina Americas    1952   62.5  17876956   5911.  
## 2 Argentina Americas    1957   64.4  19610538   6857.  
## 3 Argentina Americas    1962   65.1  21283783   7133.  
## 4 Argentina Americas    1967   65.6  22934225   8053.  
## 5 Argentina Americas    1972   67.1  24779799   9443.  
## 6 Argentina Americas    1977   68.5  26983828  10079.  
## 7 Argentina Americas    1982   69.9  29341374   8998.  
## 8 Argentina Americas    1987   70.8  31620918   9140.  
## 9 Argentina Americas    1992   71.9  33958947   9308.  
## 10 Argentina Americas    1997   73.3  36203463  10967.  
## 11 Argentina Americas    2002   74.3  38331121   8798.  
## 12 Argentina Americas    2007   75.3  40301927  12779.
```



2. Seleccionar las columnas correspondientes a ingresos per capita y esperanza de vida

```
## # A tibble: 1,704 x 2
##   lifeExp gdpPercap
##   <dbl>    <dbl>
## 1    28.8    779.
## 2    30.3    821.
## 3    32.0    853.
## 4    34.0    836.
## 5    36.1    740.
## 6    38.4    786.
## 7    39.9    978.
## 8    40.8    852.
## 9    41.7    649.
## 10   41.8    635.
## # ... with 1,694 more rows
```

Solución



```
gapminder %>%  
  select(lifeExp, gdpPercap)
```

```
## # A tibble: 1,704 x 2  
##   lifeExp gdpPercap  
##   <dbl>   <dbl>  
## 1    28.8     779.  
## 2    30.3     821.  
## 3    32.0     853.  
## 4    34.0     836.  
## 5    36.1     740.  
## 6    38.4     786.  
## 7    39.9     978.  
## 8    40.8     852.  
## 9    41.7     649.  
## 10   41.8     635.  
## # ... with 1,694 more rows
```

3. Eliminar la columna correspondiente a la población en gapminder



```
## # A tibble: 1,704 x 5
##   country      continent  year lifeExp gdpPercap
##   <fct>         <fct>    <int>   <dbl>    <dbl>
## 1 Afghanistan Asia      1952    28.8     779.
## 2 Afghanistan Asia      1957    30.3     821.
## 3 Afghanistan Asia      1962    32.0     853.
## 4 Afghanistan Asia      1967    34.0     836.
## 5 Afghanistan Asia      1972    36.1     740.
## 6 Afghanistan Asia      1977    38.4     786.
## 7 Afghanistan Asia      1982    39.9     978.
## 8 Afghanistan Asia      1987    40.8     852.
## 9 Afghanistan Asia      1992    41.7     649.
## 10 Afghanistan Asia      1997    41.8     635.
## # ... with 1,694 more rows
```


Solución



```
gapminder %>%  
  select(-pop)
```

```
## # A tibble: 1,704 x 5  
##   country      continent  year lifeExp gdpPercap  
##   <fct>         <fct>    <int>   <dbl>    <dbl>  
## 1 Afghanistan Asia      1952    28.8     779.  
## 2 Afghanistan Asia      1957    30.3     821.  
## 3 Afghanistan Asia      1962    32.0     853.  
## 4 Afghanistan Asia      1967    34.0     836.  
## 5 Afghanistan Asia      1972    36.1     740.  
## 6 Afghanistan Asia      1977    38.4     786.  
## 7 Afghanistan Asia      1982    39.9     978.  
## 8 Afghanistan Asia      1987    40.8     852.  
## 9 Afghanistan Asia      1992    41.7     649.  
## 10 Afghanistan Asia      1997    41.8     635.  
## # ... with 1,694 more rows
```



4. Seleccionar los datos correspondientes al continente americano, correspondientes del año 80 en adelante. Los datos deben estar ordenados según los años más actuales.

```
## # A tibble: 150 x 6
```

```
##   country      continent year lifeExp      pop gdpPercap
##   <fct>        <fct>    <int>   <dbl>    <int>    <dbl>
## 1 Argentina    Americas    2007    75.3   40301927   12779.
## 2 Bolivia      Americas    2007    65.6    9119152    3822.
## 3 Brazil       Americas    2007    72.4  190010647   9066.
## 4 Canada       Americas    2007    80.7   33390141  36319.
## 5 Chile        Americas    2007    78.6   16284741  13172.
## 6 Colombia     Americas    2007    72.9   44227550   7007.
## 7 Costa Rica   Americas    2007    78.8    4133884   9645.
## 8 Cuba         Americas    2007    78.3   11416987   8948.
## 9 Dominican Republic Americas    2007    72.2    9319622   6025.
## 10 Ecuador     Americas    2007    75.0   13755680   6873.
## # ... with 140 more rows
```

Solución



```
gapminder %>%  
  filter(continent=="Americas", year >= 1980) %>%  
  arrange(desc(year))
```

```
## # A tibble: 150 x 6
```

##	country	continent	year	lifeExp	pop	gdpPercap
##	<fct>	<fct>	<int>	<dbl>	<int>	<dbl>
##	1 Argentina	Americas	2007	75.3	40301927	12779.
##	2 Bolivia	Americas	2007	65.6	9119152	3822.
##	3 Brazil	Americas	2007	72.4	190010647	9066.
##	4 Canada	Americas	2007	80.7	33390141	36319.
##	5 Chile	Americas	2007	78.6	16284741	13172.
##	6 Colombia	Americas	2007	72.9	44227550	7007.
##	7 Costa Rica	Americas	2007	78.8	4133884	9645.
##	8 Cuba	Americas	2007	78.3	11416987	8948.
##	9 Dominican Republic	Americas	2007	72.2	9319622	6025.
##	10 Ecuador	Americas	2007	75.0	13755680	6873.
##	# ... with 140 more rows					

5. Calcular la media por continente del ingreso per cápita y la esperanza de vida. Ordenarlos de mayor a menor según la esperanza de vida.



```
## # A tibble: 5 x 3
##   continent      gpd lifeE
##   <fct>         <dbl> <dbl>
## 1 Oceania      18622.   74.3
## 2 Europe       14469.   71.9
## 3 Americas     7136.   64.7
## 4 Asia         7902.   60.1
## 5 Africa       2194.   48.9
```

Solución



```
gapminder %>%  
  group_by(continent) %>%  
  summarize(gpd=mean(gdpPercap), lifeE=mean(lifeExp)) %>%  
  arrange(desc(lifeE))
```

```
## # A tibble: 5 x 3  
##   continent    gpd lifeE  
##   <fct>      <dbl> <dbl>  
## 1 Oceania    18622.  74.3  
## 2 Europe     14469.  71.9  
## 3 Americas   7136.  64.7  
## 4 Asia       7902.  60.1  
## 5 Africa     2194.  48.9
```

6. ¿Cuáles son los 7 países con mayor esperanza de vida en el año 2002?



```
## # A tibble: 7 x 6
```

```
##   country      continent  year lifeExp      pop gdpPercap
##   <fct>        <fct>      <int>  <dbl>    <int>    <dbl>
## 1 Japan        Asia        2002    82   127065841  28605.
## 2 Hong Kong, China Asia        2002   81.5   6762476   30209.
## 3 Switzerland Europe        2002   80.6   7361757   34481.
## 4 Iceland      Europe        2002   80.5    288030   31163.
## 5 Australia    Oceania       2002   80.4  19546792  30688.
## 6 Italy         Europe        2002   80.2  57926999  27968.
## 7 Sweden       Europe        2002   80.0   8954175   29342.
```

Solución



```
gapminder %>%  
  filter(year==2002) %>%  
  arrange(desc(lifeExp)) %>%  
  top_n(7, lifeExp)
```

```
## # A tibble: 7 x 6  
##   country      continent  year lifeExp      pop gdpPercap  
##   <fct>        <fct>    <int>  <dbl>    <int>    <dbl>  
## 1 Japan        Asia      2002    82  127065841  28605.  
## 2 Hong Kong, China Asia      2002   81.5   6762476  30209.  
## 3 Switzerland  Europe    2002   80.6   7361757  34481.  
## 4 Iceland      Europe    2002   80.5    288030  31163.  
## 5 Australia    Oceania   2002   80.4  19546792  30688.  
## 6 Italy         Europe    2002   80.2  57926999  27968.  
## 7 Sweden       Europe    2002   80.0   8954175  29342.
```

Para seguir practicando...



https://garhtarr.github.io/meatR/dplyr_ex1.html

R4DS Capítulo 5: Data Transformation

<https://r4ds.had.co.nz/transform.html>

R4DS Capítulo 5 (en español)

<https://es.r4ds.hadley.nz/transform.html>

Soluciones (en inglés)

<https://jrnold.github.io/r4ds-exercise-solutions/transform.html>