

Variantes de las funciones



dplyr package

Alex Bajaña

Marzo 2019

Buenos días



Empecemos

Datos a usarse

En esta clase vamos a usar el paquete `wbstats` en donde hallaremos datasets interesantes en los cuales aplicar las funciones hasta ahora aprendidas.

```
if(!"wbstats" %in% installed.packages()[,1]){  
  install.packages("wbstats")  
}
```

```
library(wbstats)
```

```
?wbstats
```

Este paquete permite la búsqueda y descarga de datos recolectados por el Banco Mundial.



Funciones básicas de dplyr

- select
- filter
- **mutate**
- group_by
- **summarise**

Funciones auxiliares:

- **Matemáticas:** +, -, log()
- **Temporales:** lead(), lag()
- **Rankings:** dense_rank(), min_rank(), percent_rank(), row_num(), cume_dist()
- **Acumulativas:** cumsum(), cummean(), cummin(), cummax()
- **Condicionales:** if_else(), case_when()

Aprendan más en: [Cran R Project](#)

Extracción de datos petroleros

Vamos a extraer los datos del porcentaje de pobreza a nivel mundial a través del paquete `wbstats`.

```
library(ggplot2)
library(dplyr)

unemploy_vars <- wbsearch(pattern = "unemployment")

country_data <- wb(indicator = "UNEMPSA_")
```

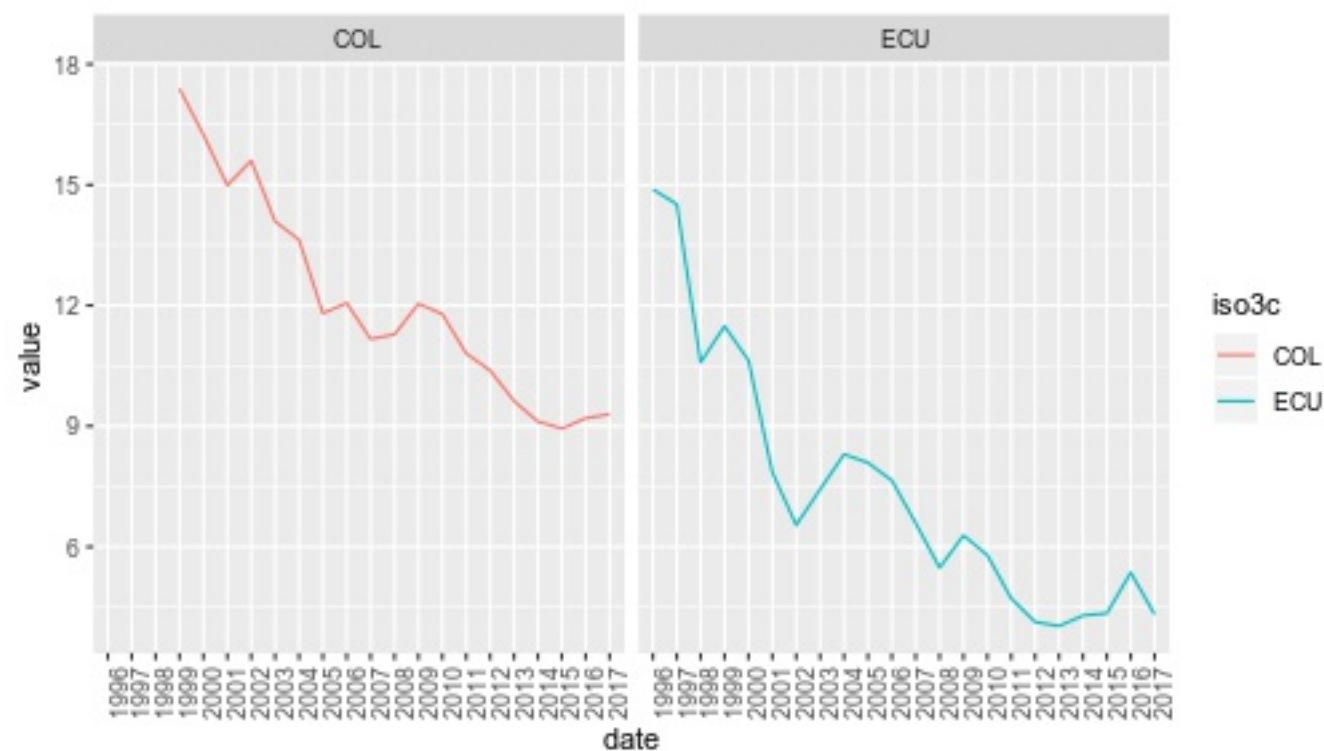
Para ver un listado completo de los indicadores disponibles:

```
wbsearch() %>% head
```

```
##           indicatorID
## 820 SH.UHC.NOP2.ZS
## 821 SH.UHC.NOP2.ZG
## 822 SH.UHC.NOP2.TO
## 823 SH.UHC.NOP2.CG
## 824 SH.UHC.NOP1.ZS
## 825 SH.UHC.NOP1.ZG
##
## 820           Proportion of population pushed below the $3.10 ($ 2011 PPP) poverty line
```

Evolución en la tasa de pobreza

```
country_data %>%  
  filter(iso3c %in% c("ECU", "COL")) %>%  
  ggplot()+  
  geom_line(aes(x= date, y = value, color = iso3c, group=1))+  
  facet_grid(.~iso3c)+  
  theme(axis.text.x = element_text(angle = 90))
```



Funciones que vamos a revisar:

mutate_at

```
library(knitr)
country_data %>%
  mutate_at(.vars = c("iso3c", "country"),
            .funs = funs(tolower)) %>%
  head
```

```
##   iso3c date    value indicatorID          indicator iso2c
## 1  <NA> 2017 6.002274 UNEMPSA_ Unemployment rate,Percent,,, AME
## 2  <NA> 2016 6.470872 UNEMPSA_ Unemployment rate,Percent,,, AME
## 3  <NA> 2015 6.950939 UNEMPSA_ Unemployment rate,Percent,,, AME
## 4  <NA> 2014 7.556938 UNEMPSA_ Unemployment rate,Percent,,, AME
## 5  <NA> 2013 8.177776 UNEMPSA_ Unemployment rate,Percent,,, AME
## 6  <NA> 2012 8.216599 UNEMPSA_ Unemployment rate,Percent,,, AME
##               country
## 1 advanced economies
## 2 advanced economies
## 3 advanced economies
## 4 advanced economies
## 5 advanced economies
## 6 advanced economies
```


Funciones que vamos a revisar:

mutate_if

```
country_data %>%  
  mutate_if(is.numeric,  
            round,  
            digits = 0) %>%  
  head
```

```
##   iso3c date value indicatorID          indicator iso2c  
## 1  <NA> 2017     6  UNEMPSA_ Unemployment rate,Percent,,, AME  
## 2  <NA> 2016     6  UNEMPSA_ Unemployment rate,Percent,,, AME  
## 3  <NA> 2015     7  UNEMPSA_ Unemployment rate,Percent,,, AME  
## 4  <NA> 2014     8  UNEMPSA_ Unemployment rate,Percent,,, AME  
## 5  <NA> 2013     8  UNEMPSA_ Unemployment rate,Percent,,, AME  
## 6  <NA> 2012     8  UNEMPSA_ Unemployment rate,Percent,,, AME  
##               country  
## 1 Advanced Economies  
## 2 Advanced Economies  
## 3 Advanced Economies  
## 4 Advanced Economies  
## 5 Advanced Economies  
## 6 Advanced Economies
```

Funciones que vamos a revisar:

mutate_all

```
airquality %>%  
  mutate_all(funs(cm = .*100)) %>%  
  head
```

##	Ozone	Solar.R	Wind	Temp	Month	Day	Ozone_cm	Solar.R_cm	Wind_cm	Temp_cm
## 1	41	190	7.4	67	5	1	4100	19000	740	6700
## 2	36	118	8.0	72	5	2	3600	11800	800	7200
## 3	12	149	12.6	74	5	3	1200	14900	1260	7400
## 4	18	313	11.5	62	5	4	1800	31300	1150	6200
## 5	NA	NA	14.3	56	5	5	NA	NA	1430	5600
## 6	28	NA	14.9	66	5	6	2800	NA	1490	6600

##	Month_cm	Day_cm
## 1	500	100
## 2	500	200
## 3	500	300
## 4	500	400
## 5	500	500
## 6	500	600