

Clase 1.3

Acomodo de datos

Marcos Rosetti y Luis Pacheco-Cobos

Estadística y Manejo de Datos con R (EMDR) — Virtual

Acomodo de datos: **magrittr**



Acomodo de datos: **magrittr**

- **magrittr** introduce el operador “pipeline” (tubo) `%>%` que sirve para transferir datos entre funciones. Esto resulta en un código más dinámico y fácil de leer que muchas funciones anidadas o consecutivas.
- Esta forma anidada de expresar una función:

```
mi.function(mis.datos, arg = valores)
```

- Es lo mismo que ésta en formato de “tubo”:

```
mis.datos %>% mi.function(arg = valores)
```

Acomodo de datos: `magrittr`

- `magrittr` hace homenaje al pintor surrealista René Magritte. Autor de la serie *La traición de las imágenes* (1928 - 1929) que incluye `Esto no es una pipa`.



Acomodo de datos: **magrittr**

- `magrittr` conectamos varias funciones fácilmente.
- Antes:

```
w3 <- mutate(w2, day = extract_numeric(day))
w4 <- select(w3, id, year, month, day, element, value)
w5 <- arrange(w4, id, year, month, day)
```

- Después:

```
library(magrittr)

w3 <- w2 %>% mutate(day = extract_numeric(day)) %>%
select(id, year, month, day, element, value) %>%
arrange(id, year, month, day)
```

Acomodo de datos: **magrittr**

El estilo tradicional

```
summarize(  
  group_by(  
    filter(  
      mutate(gm, gdp=gdpPercap*pop),  
      continent=="Asia"),  
    year),  
  mean(lifeExp), mean(gdp))
```

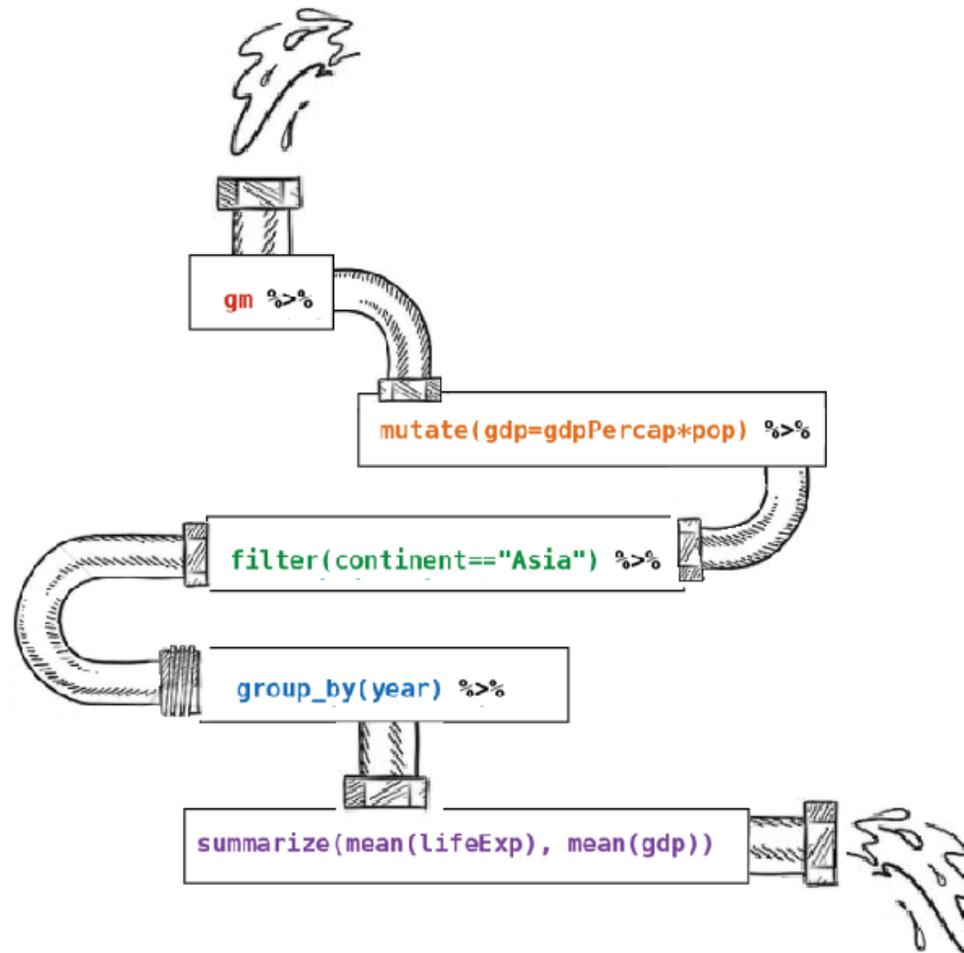
El proceso cognitivo

1. Take the **gm** data, then
2. **Mutate** it to add "gdp" variable, then
3. **Filter** where continent=="Asia", then
4. **Group by** year, then
5. **Summarize** to get mean life exp & GDP

El estilo magrittr

```
gm %>%  
  mutate(gdp=gdpPercap*pop) %>%  
  filter(continent=="Asia") %>%  
  group_by(year) %>%  
  summarize(mean(lifeExp), mean(gdp))
```

Acomodo de datos: `magrittr`



Acomodo de datos: **magrittr**

- Podemos modificar el df original, producto de las operaciones dentro de la pipa, usando el operador `%<>%` en vez de del operador flecha `<-`.

```
w2 %<>%  
mutate(day = extract_numeric(day)) %>%  
select(id, year, month, day, element, value) %>%  
arrange(id, year, month, day)
```

Licencia CC BY



Estadística y Manejo de Datos con R (EMDR) por Marcos F. Rosetti S. y Luis Pacheco-Cobos se distribuye bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).