



Introducción al tidyverse

Una serie de paquetes para Data Science

Miguel A. Porras Naranjo

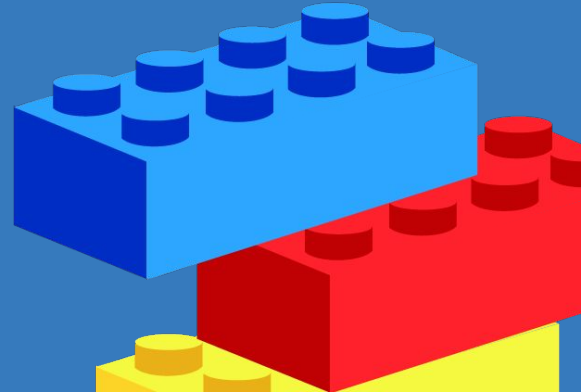
github.com/migpornar/meetup-tidyverse/



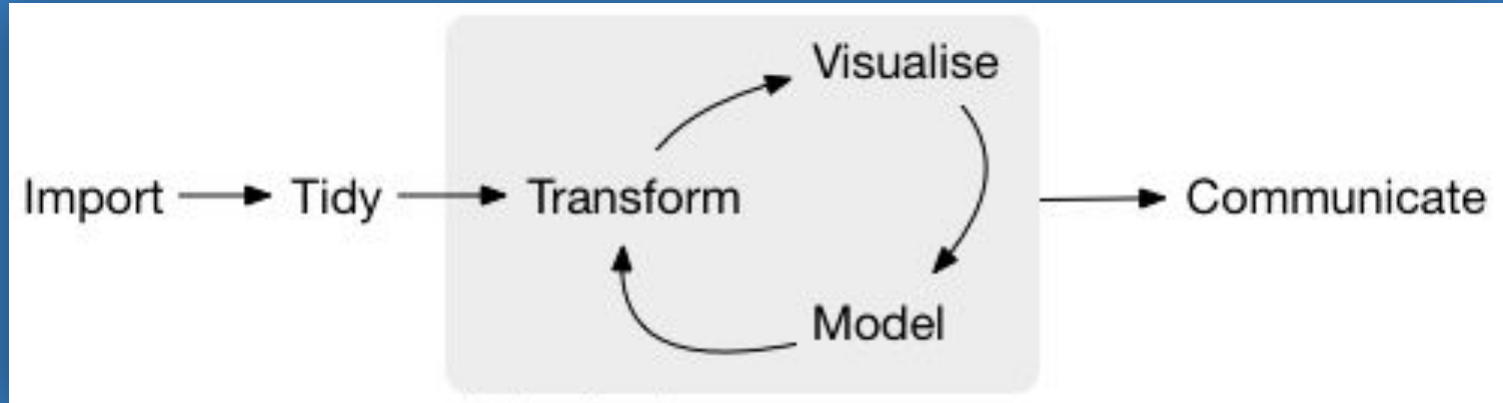
¿Qué narices es tidyverse?

- Es una colección de paquetes diseñados para hacer Data Science.
- Todos los paquetes tienen filosofía y gramática muy similar.
- Permite realizar análisis de manera rápida y fácil.

“Resolver problemas complejos combinando
sencillas piezas”



Workflow en Data Science



Fuente: R for Data Science

¿Qué es el formato tidy?

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20095360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272015272
China	2000	210766	128042583

variables

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20095360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272015272
China	2000	210766	128042583

observations

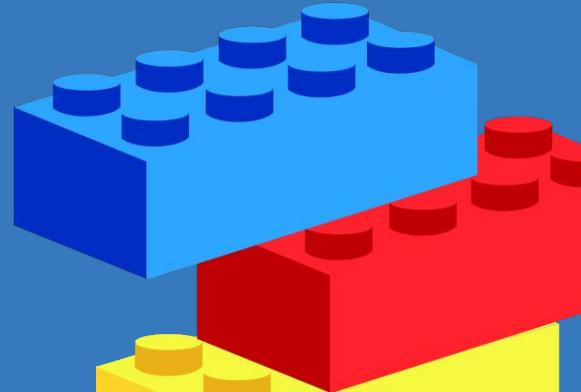
country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20095360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272015272
China	2000	210766	128042583

values

Fuente: R for Data Science

Si tenemos nuestros datos en formato tidy, podremos sacarles el máximo partido

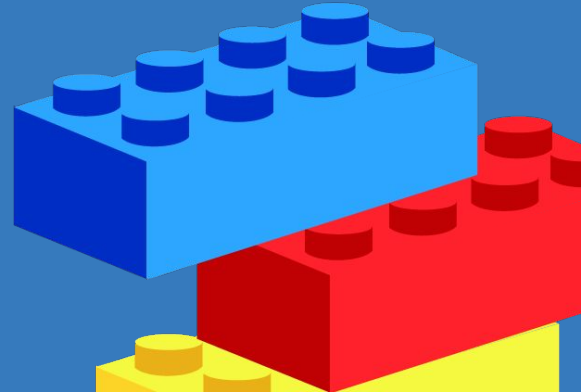
“Resolver problemas complejos combinando
sencillas piezas”



Paquetes del tidyverse

- Importar: **readr** (.csv, .tsv, .fwf), **readxl** (.xls), **haven** (SPSS, Stata, SAS), **jsonlite** (JSON), **rvest** (web scraping), **DBI** (bases de datos) ...
- Limpieza de datos: **tidyr** (tablas dinámicas), **dplyr** (operaciones básicas), **lubridate** (para fechas), **stringr** (para strings), **forcats** (para factores) ...
- Programación: **magrittr** (pipes), **purrr** (programación en listas), **glue** (similar a paste) ...
- Visualización: **ggplot2** (representaciones gráficas) ...

“Resolver problemas complejos combinando
sencillas piezas”



Definiendo operador pipe



$f(x) \longrightarrow x \%>\% f$

$f(x, y) \longrightarrow x \%>\% f(y)$

Ejemplos con el operador pipe

$f(x)$  $x \%>\% f$

$\cos(2)$  $2 \%>\% \cos$

$\text{sen}(\cos(2))$  $2 \%>\% \cos \%>\% \text{sen}$


Código de R

```
data <- read.csv("data.csv")
```

```
data2 <- subset(data, v_a > x)
```

```
data3 <- transform(data2,  
variable_c = variable_a/variable_b)
```

```
data_final <- head(data3, 100)
```

Código de R

```
data <- read.csv("data.csv")

data2 <- subset(data, v_a > x)

data3 <- transform(data2,
  variable_c = variable_a/variable_b)

data_final <- head(data3, 100)
```

Código de R + pipes

```
data <- read.csv("data.csv") %>%
  subset(v_a > x) %>%
  transform(v_c = v_a/v_b) %>%
  head(100)
```

¡Hagamos un análisis!

github.com/migpornar/meetup-tidyverse/