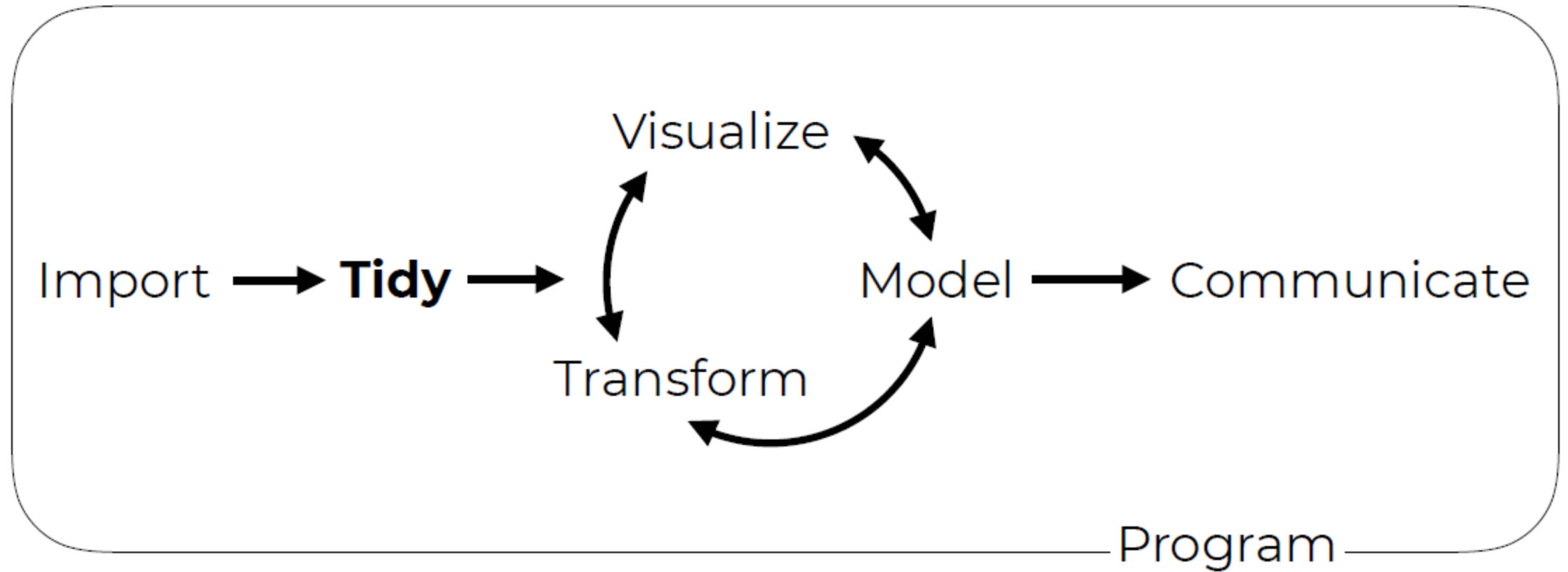


Programación en R

Giovany Babativa



Ciclo de procesamiento y análisis



Reglas de los datos

country	year	cases	population
Afghanistan	1999	3745	19987071
Afghanistan	2000	3666	20595360
Brazil	1999	3737	17206362
Brazil	2000	8488	174504898
China	1999	21258	1272015272
China	2000	21766	128042583

variables

country	year	cases	population
Afghanistan	1999	3745	19987071
Afghanistan	2000	3666	20595360
Brazil	1999	3737	17206362
Brazil	2000	8488	174504898
China	1999	21258	1272015272
China	2000	21766	128042583

observations

country	year	cases	population
Afghanistan	1999	3745	19987071
Afghanistan	2000	3666	20595360
Brazil	1999	3737	17206362
Brazil	2000	8488	174504898
China	1999	21258	1272015272
China	2000	21766	128042583

values

... estructura de los datos

```
head(yourdata)  
str(yourdata)  
length(yourdata)  
glimpse(yourdata)  
names(yourdata)
```


Con librería y sin librerías

```
fa=table(datos$Sexo)
view(fa)
fr=round(100*fa/sum(fa),1)
Tabla=cbind(fa,fr)
```

```
install.packages("packagename")
```

?descr::freq

```
library(packagename)
```

El mundo Tidyverse

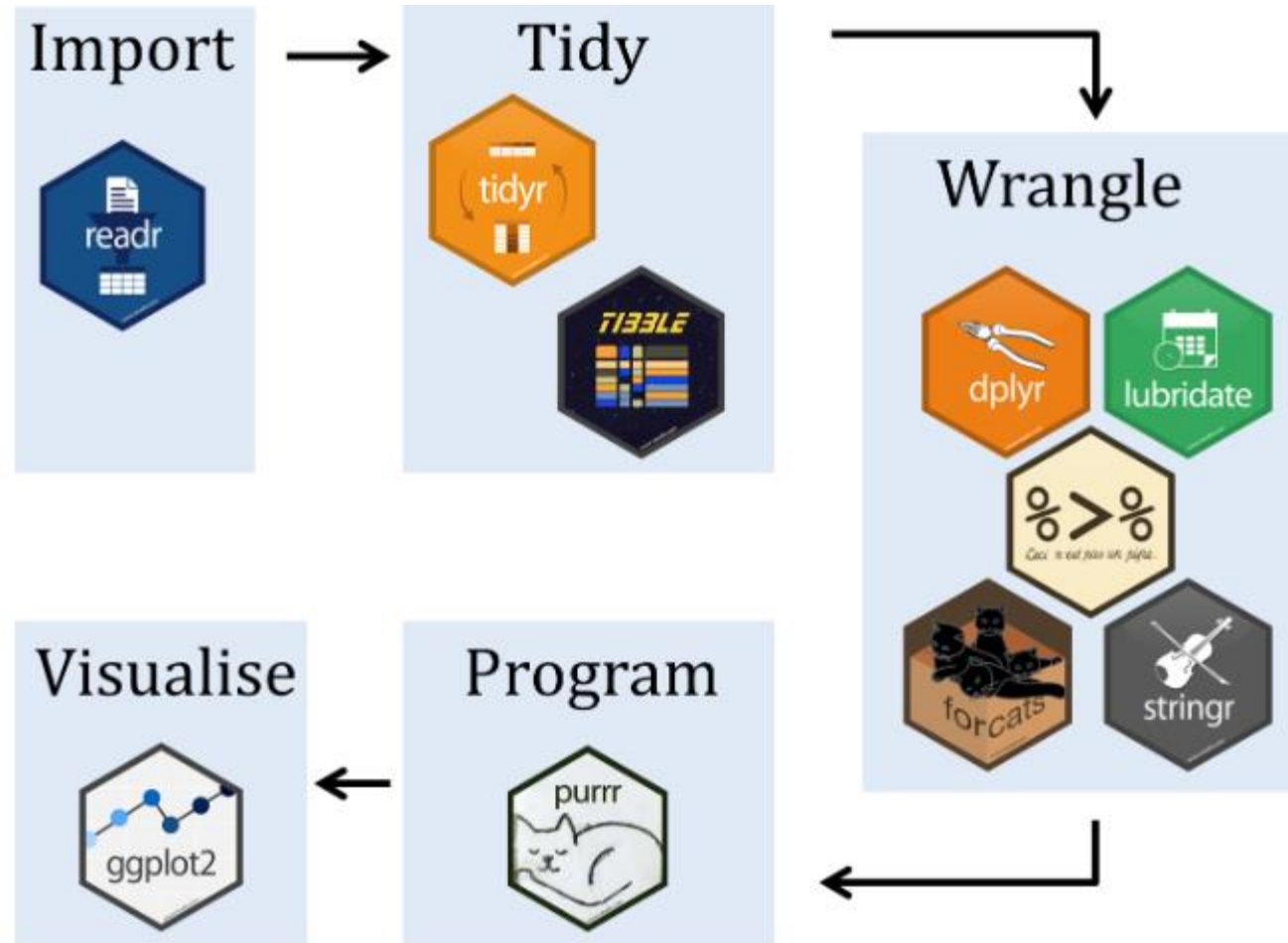


```
library(tidyverse)
```



```
library(readr)  
library(dplyr)  
library(tidyr)  
library(ggplot2)  
library(purrr)  
library(tibble)  
library(stringr)  
library(forcats)
```

El mundo Tidyverse



La gramática de las gráficas con ggplot2



```
ggplot(data, aes(x = __, y = __)) +  
  geom_point()
```

```
data %>%  
  filter(interesting_variable > z) %>%  
  ggplot(aes(x = __, y = __, colour = condition))  
  geom_point() +  
  facet_wrap(~ group)  
  
ggsave("your_first_ggplot.png")
```

Tidy - Simple

filter() - Extraer casos

arrange() - Ordenar casos

group_by() - Agrupar casos

select() - Seleccionar variables

mutate() - Crear nuevas variables

summarize() - Resumir variables o crear casos



country	year	cases	population
Afghanistan	1999	745	15467071
Afghanistan	2000	666	20395360
Brazil	1999	30737	17206362
Brazil	2000	80488	17404898
China	1999	210258	1272015272
China	2000	210366	128043583

variables

country	year	cases	population
Afghanistan	1999	745	15467071
Afghanistan	2000	666	20395360
Brazil	1999	30737	17206362
Brazil	2000	80488	17404898
China	1999	210258	1272015272
China	2000	210366	128043583

observations

country	year	cases	population
Afghanistan	1999	745	15467071
Afghanistan	2000	666	20395360
Brazil	1999	30737	17206362
Brazil	2000	80488	17404898
China	1999	210258	1272015272
China	2000	210366	128043583

values

Ejemplo

Use la base **MUNDO**

- Filtre solo los países de la OCDE
- Filtre los países de la OCDE y de América.
- Ordene los países por su población (A y D)
- Seleccione solo las variables de país, región, población, esperanza de vida femenina y masculina.
- Cree una nueva variable que sea la población sobre 10.000.
- Cree una nueva variable que contenga el valor promedio de la esperanza de vida masculina.
- Cree la variable anterior pero por OCDE y América.
- Calcule una variable que valga 1 si está por encima de la media y 0 en otro caso.

Tidy - Verbos



Base larga a base ancha **spread()**, **pivot_wider()**



Base ancha a base larga **gather()**, **pivot_longer()**



Separar cadena en columnas **separate()**



Unir columnas **unite()**

Ejemplo

```
cases <- tribble(
  ~Country, ~"2011", ~"2012", ~"2013",
  "FR", 7000, 6900, 7000,
  "DE", 5800, 6000, 6200,
  "US", 15000, 14000, 13000
)
```

Quiero graficar por año

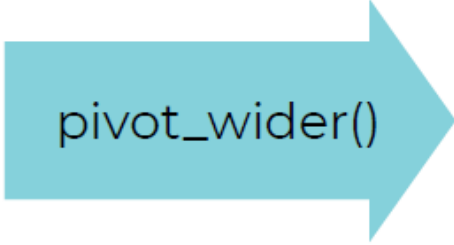


```
cases %>% pivot_longer(-country, names_to = "year", values_to = "n")
```

country	year	n
FR	2011	7000
DE	2011	5800
US	2011	15000
FR	2012	6900
DE	2012	6000
US	2012	14000
FR	2013	7000
DE	2013	6200
US	2013	13000

Ejemplo

city	size	amount
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56



`pivot_wider()`

city	large	small
New York	23	14
London	22	16
Beijing	121	56

```
pollution %>% pivot_wider(names_from = size, values_from = amount)
```