

Akram Méndez

[contacto@solariabiodata.com.mx](mailto:contacto@solariabiodata.com.mx)

WebinR<sup>3</sup>



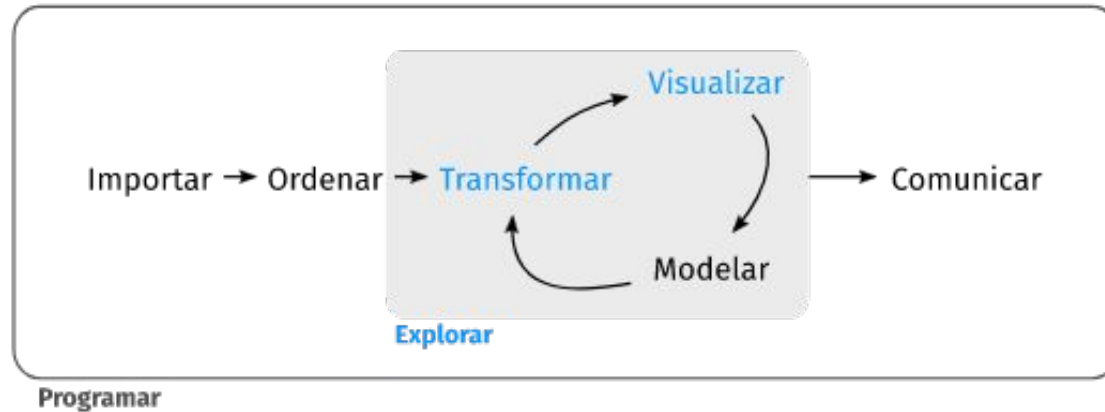
# Tidyverse

- Funciones y paquetes disponibles
- Manejo de datos y programación con Tidyverse

## Visualización de datos

- Gramática de gráficos
- Graficación con ggplot2
- Gráficas interactivas

# Análisis exploratorio de datos con Tidyverse



Grolemund, G., & Wickham, H. (2018). R for data science.  
Versión en español: <https://es.r4ds.hadley.nz/>

# Importar/exportar datos



- `read_[csv,tsv,delim,rds]:` Importar/exportar archivos delimitados
- `write_[csv,tsv,delim,rds]:`
- `readxl:` Lectura de archivos .xls, .xlsx
- `DBI, odbc:` Bases de datos relacionales
- `rvest:` Webscapping HTML
- `jsonlite:` Archivos JSON
- `xml2:` Archivos XML

Explorar la estructura de los datos

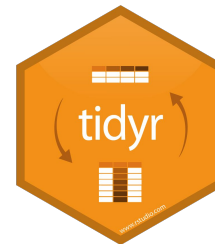
- `glimpse():` Explorar tipos de variables, número de columnas, filas.

# “Tidy” Ordenar los datos



## Estructura estándar:

1. Cada columna es una variable
2. Cada fila es una observación
3. Cada celda tiene sólo un valor



Formato  
largo/  
agregado

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K



country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

key value

## gather(),spread():

Transformar los datos a un formato “largo” o “extendido”

Formato  
extendido

country	year	type	count
A	1999	cases	0.7K
A	1999	pop	19M
A	2000	cases	2K
A	2000	pop	20M
B	1999	cases	37K
B	1999	pop	172M
B	2000	cases	80K
B	2000	pop	174M
C	1999	cases	212K
C	1999	pop	1T
C	2000	cases	213K
C	2000	pop	1T



country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172M
B	2000	80K	174M
C	1999	212K	1T
C	2000	213K	1T

key value

(la última versión de *tidyr* introduce las funciones mejoradas `pivot_longer()` y `pivot_wider()`, respectivamente).

# Transformar los datos



- **arrange():**
- **select():**
- **filter():**
- **summarise():**
- **group\_by():**
- **(left,right,full,anti,semi)\_join():**

Ordenar datos de acuerdo con los valores de una o más columnas

Seleccionar columnas por nombre

Seleccionar casos de acuerdo con un valor o condición

Reduce múltiples valores

Agrupar observaciones de acuerdo con una variable

Combinar y filtrar observaciones entre dos conjuntos de datos



**pipes**

**x %>% f(y)**  
becomes **f(x, y)**

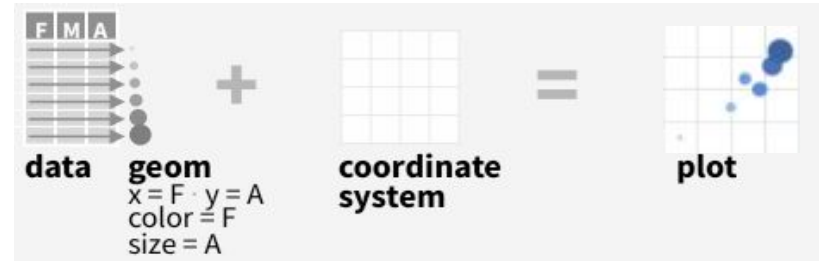
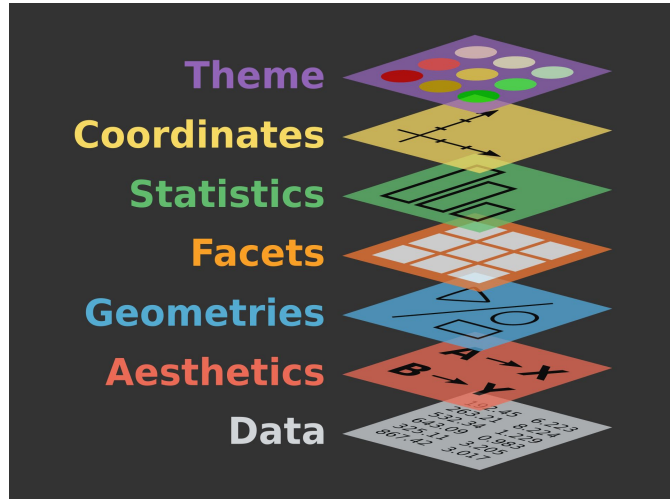
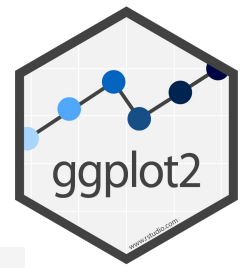
ID	X1
1	a1
2	a2

ID	X2
2	b1
3	b2

inner_join			left_join			right_join			full_join			semi_join		anti_join	
ID	X1	X2	ID	X1	X2	ID	X1	X2	ID	X1	X2	ID	X1	ID	X1
2	a2	b1	1	a1	NA	2	a2	b1	1	a1	NA	2	a2	1	a1
			2	a2	b1	3	NA	b2	2	a2	b1				
									3	NA	b2				

# Visualización de datos

## Gramática de gráficos



```
data %>%  
ggplot(aes(x=F,y=A,color=F)) +  
geom_point(aes(size=A)) +  
theme_bw()
```

## Estéticas

fill  
color, colour  
size  
shape  
alpha

0	1	2	3	4	
5	6	7	8	9	
10	11	12	13	14	
15	16	17	18	19	
20	21	22	23	24	25

## Coordenadas

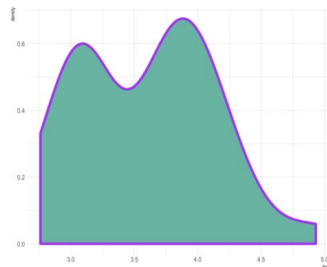
coord\_cartesian()  
coord\_polar()  
coord\_equal()  
coord\_fixed()  
coord\_map()

## Geometrías

	<code>geom_hline()</code> , <code>geom_vline()</code> , <code>geom_abline()</code>
	<code>geom_point()</code>
	<code>geom_col()</code> , <code>geom_bar()</code> , <code>geom_histogram()</code>
	<code>geom_polygon()</code>
	<code>geom_boxplot()</code>
	<code>geom_violin()</code>



# Colores

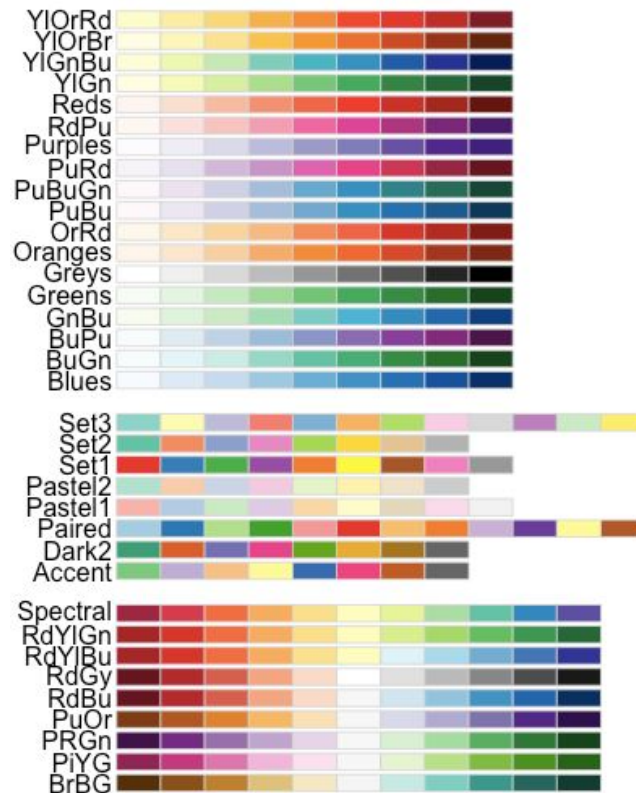


- `scale_fill_continuous()`
- `scale_colour_continuous()`
- `scale_fill_discrete()`
- `scale_fill_manual()`
- `scale_fill_gradient()`

“color” vs “fill”



Las líneas y puntos se mapean a la estética de “color”



- Más paletas de color en R:  
<https://github.com/EmilHvitfeldt/r-color-palettes>

# Gráficas interactivas

- [Plotly](#)



- [Highcharter, htmlwidgets](#)



- [Ggvis](#)



# Sitios de interés

- R para ciencia de datos: <https://es.r4ds.hadley.nz/>
- <http://www.cookbook-r.com/>
- ggplot2: <https://ggplot2-book.org/>
- <https://www.r-graph-gallery.com/index.html>
- Plotly: <https://plotly-r.com/>
- Data to viz: <https://www.data-to-viz.com/>
- <https://rstudio.com/resources/cheatsheets/>

#tidytuesday

#rstats

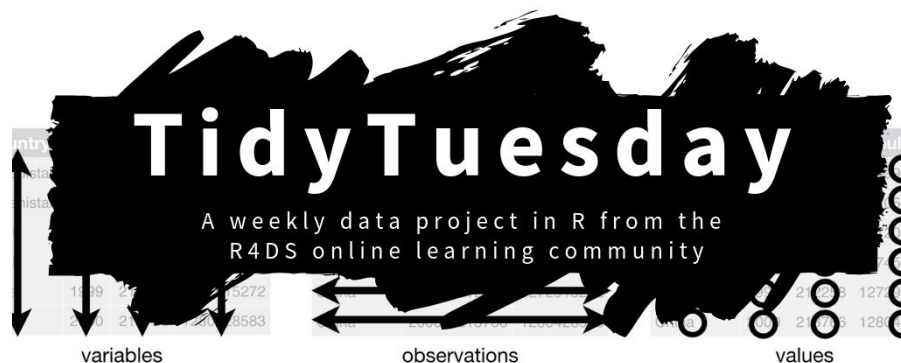
Repo en Github:

<https://solariabiodata.github.io/W3bin-R/>

Contacto

Twitter: @Akrams

akram.mendez@gmail.com





# ¡Gracias!

[contacto@solariabiodata.com.mx](mailto:contacto@solariabiodata.com.mx)

+52 1 55 6331 5961

<http://solariabiodata.com.mx>

[fb.me/SolariaBiodata](https://fb.me/SolariaBiodata)