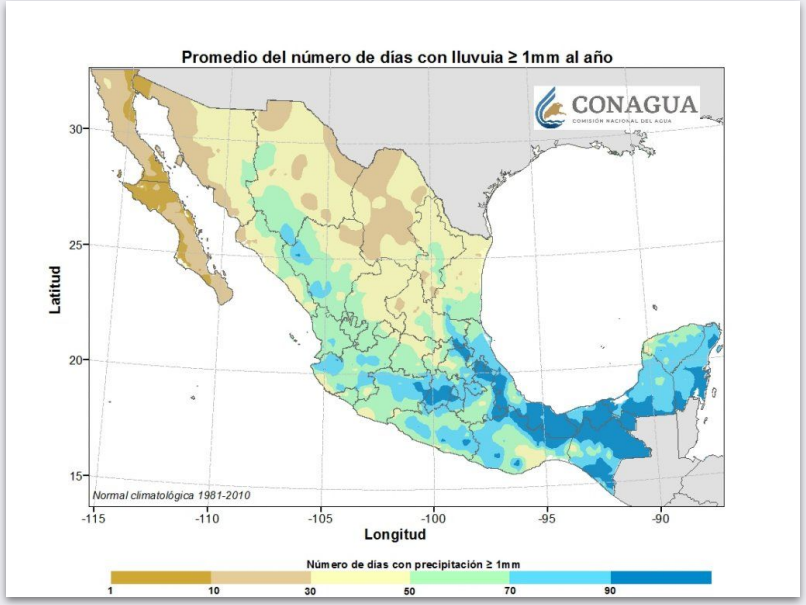
Drop plot en R

BioFreelancer - María Fernanda Mirón

El problema

- La lluvia es un **fenómeno natural** importante para la vida
- La **CONAGUA** registra los datos históricos de la cantidad de lluvia en México

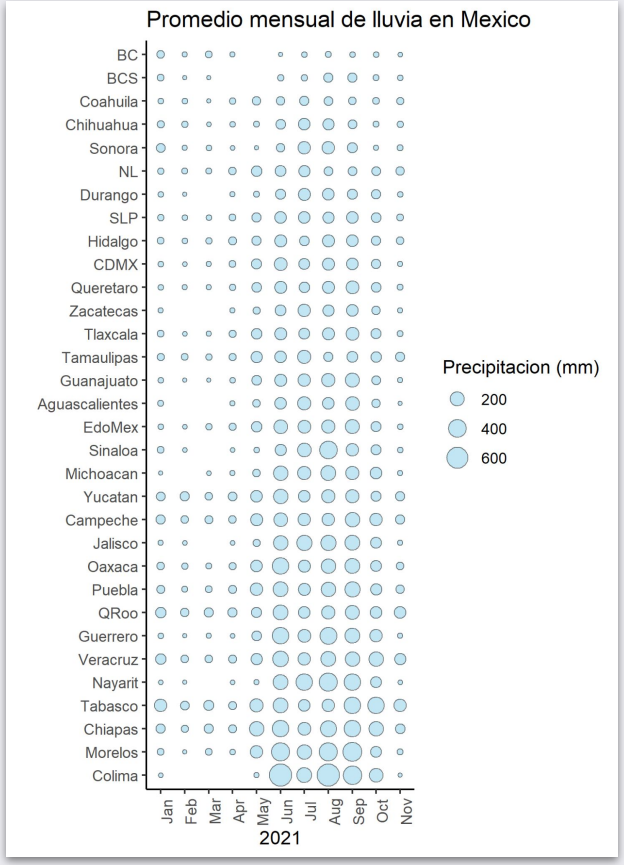


La pregunta

- ¿Cuáles son los **meses** con lluvias más fuertes en México?
- ¿Cómo se ve esto en un **dotplot**?

La respuesta

- Es puro **ggplot**
- Usaremos la geometrìa de **geom_point( )**
- una columna (mes del año) define el eje X
- otra columna (Estado de la República) define las categorías del eje Y
- El tamaño de los puntos muestra la cantidad de lluvia en mm



Instalación de paquetes

Comando

```
install.packages("nombre_paquete")
```

Después de la instalación es importante que lo **carguemos** en nuestra **sesión R** :

```
library(“nombre_paquete”)
```

Tablas largas y anchas:
Reshaping datasets

Tablas anchas

Variable 1	Variable 2	Variable 3	Variable 4
0.91	1.1	3.45	10.2
0.92	1.2	4.53	10.3
0.78	1.4	3.67	10.6

Las tablas donde todas las columnas son variables, se denomina: **tabla ancha**.

Tablas largas

Una tabla larga es aquella en la que aparece **una columna que contienen más de una variable**, y otra columna que contienen los valores de estas **diferentes variables**

Variables	Valores
Variable 1	0.91
Variable 1	0.92
Variable 1	0.78
Variable 2	1.1
Variable 2	1.2
Variable 2	1.4
Variable 3	3.45
Variable 3	4.53
Variable 3	3.67
Variable 4	10.2
Variable 4	10.3
Variable 5	10.6

De ancho a **largo**

```
pivot_longer( data = lluvias,
cols = col_1, names_to = "mes",
values_to = "precipitacion_mm" )
```

Data = datos, **cols =** las columnas que queremos apilar,
names_to=nombre de la nueva columna donde se apilan,
values_to = nombre de la columna donde se apilan los valores

Repaso algunas funciones:
dplyr

mutate dplyr

- **mutate**: permite agregar una nueva columna o transformar una existente.

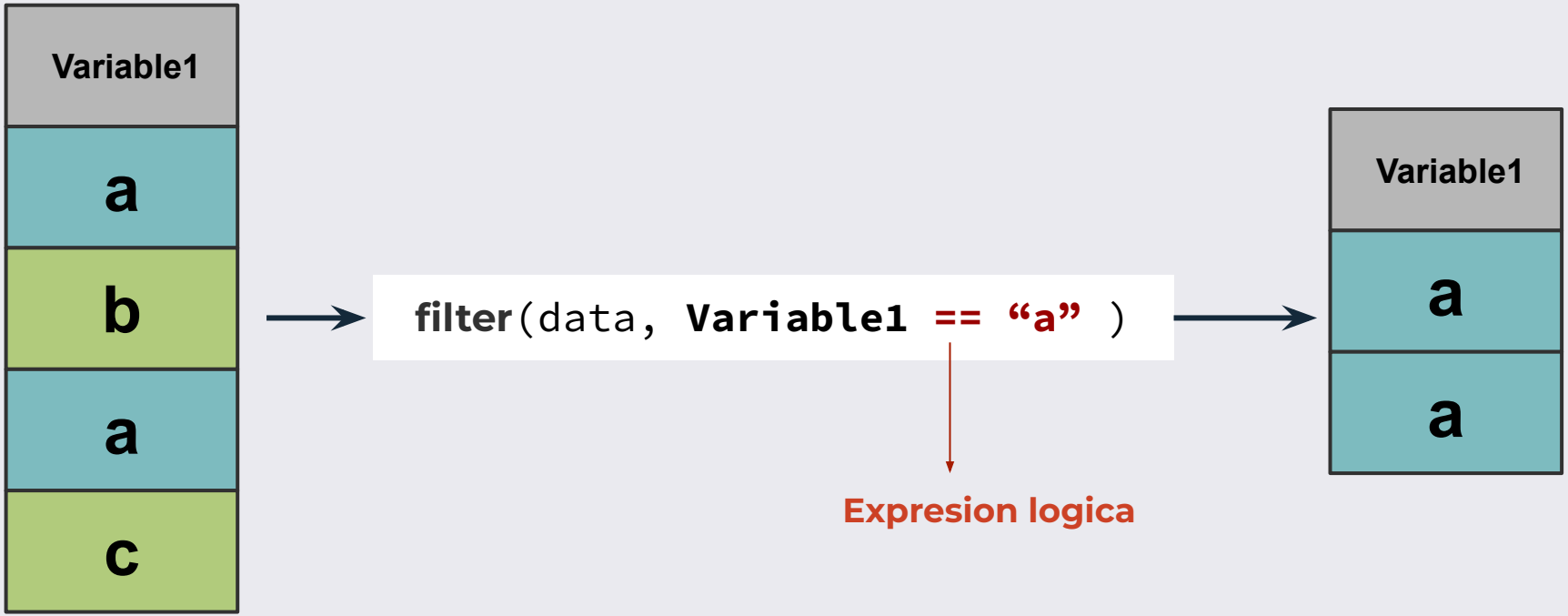
```
mutate(data, Var3 = Var1 + Var2)
```

Var1	Var2
1	2
2	3
2	2
3	3

Var1	Var2	Var3
1	2	3
2	3	5
2	2	4
3	3	6

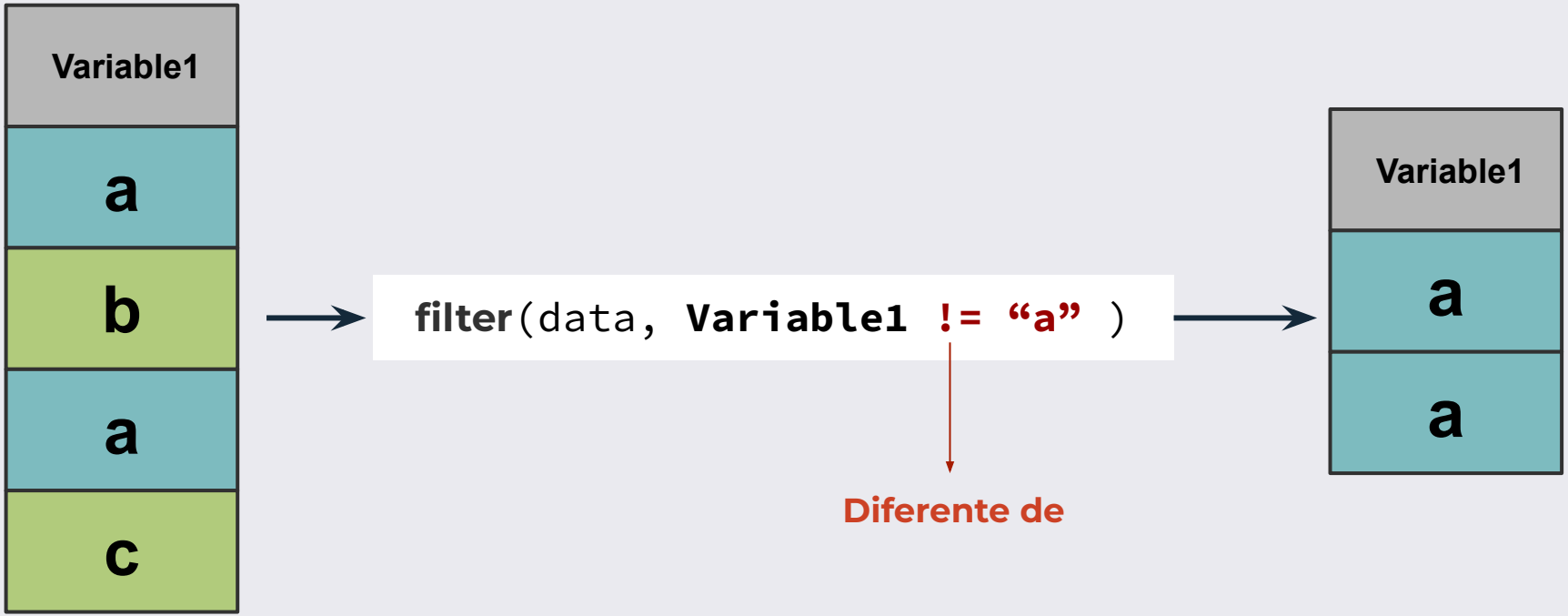
filter dplyr

- **filter**: permite filtrar filas de una data frame según una expresión lógica.



filter dplyr

- **filter**: permite filtrar filas de una data frame según una expresión lógica.



Lubridate: Paquete R

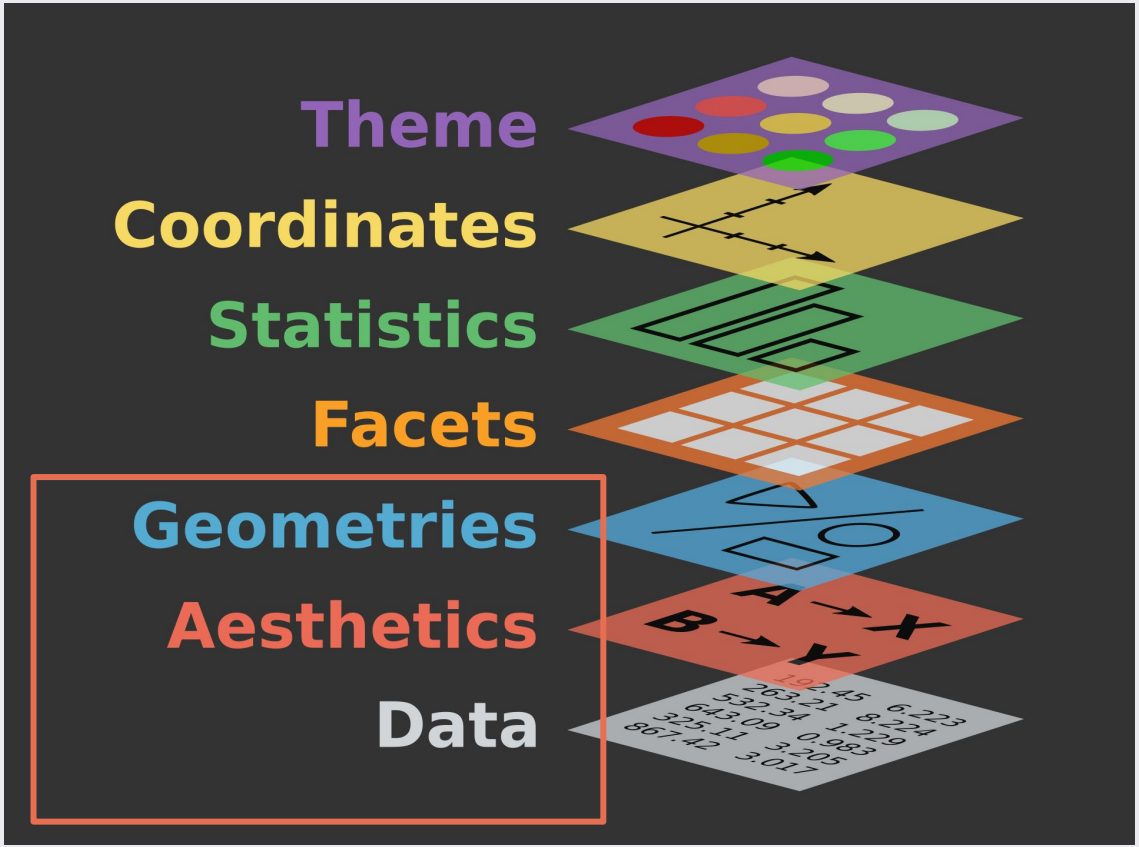
lubridate month

Variable1
1
2
3
4

```
month(data, x=Variable1,label = TRUE)
```

Variable1
ene
feb
mar
abr

Repaso ggplot2: dplyr



https://twitter.com/cedscherer/status/1229392907234402305?lang=eu

Datos

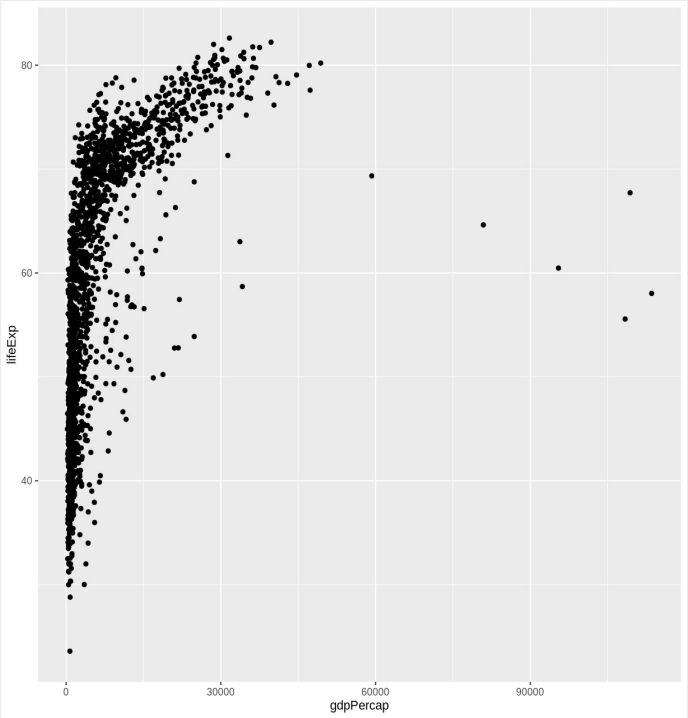
¿Qué datos queremos usar?

Aesthetics

De nuestros datos, ¿qué será x y qué será y?

Geom

¿Qué geometría queremos usar? ¿Qué gráfico quieres?

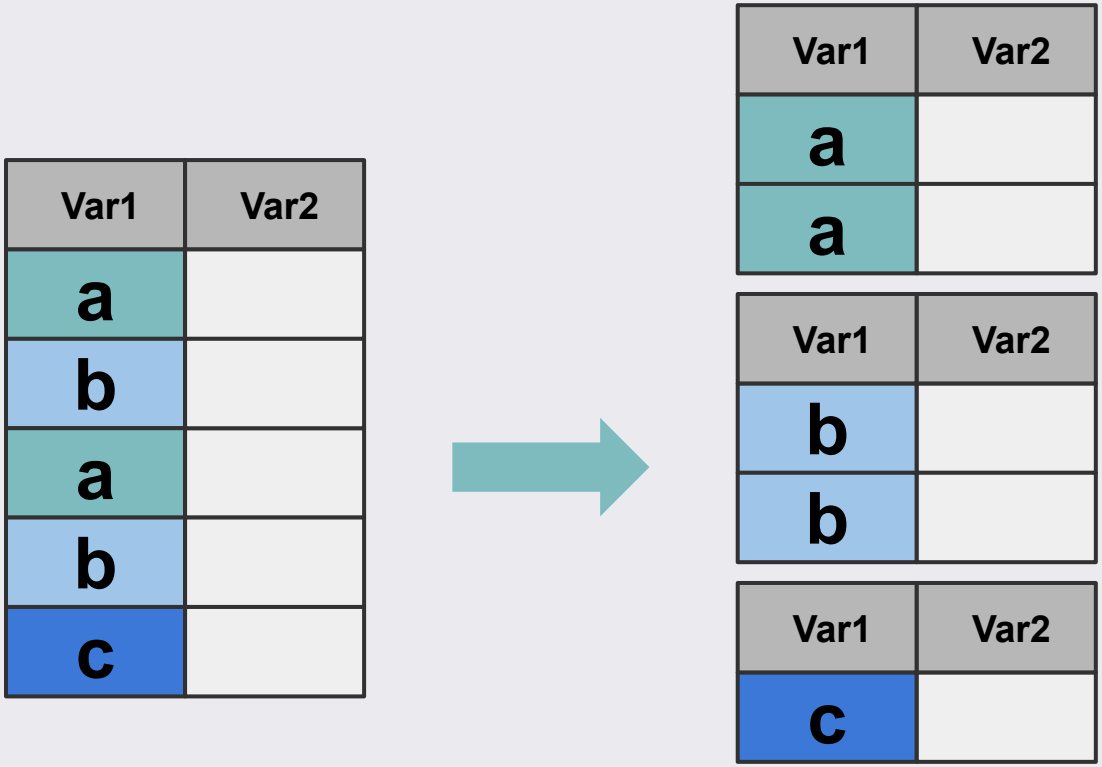



```
ggplot(data = mi_data,
mapping = aes(x = esto_es_x,
                y = esto_es_y)) +
geom_point()
```

Repaso ggplot2: dplyr

group_by dplyr

- **group_by**: agrupa las filas de datos de acuerdo a una o más variables



summarise dplyr

- **sumarise()**: aplica una función estadística a determinada columna