



## *Querido Científico de Datos:*

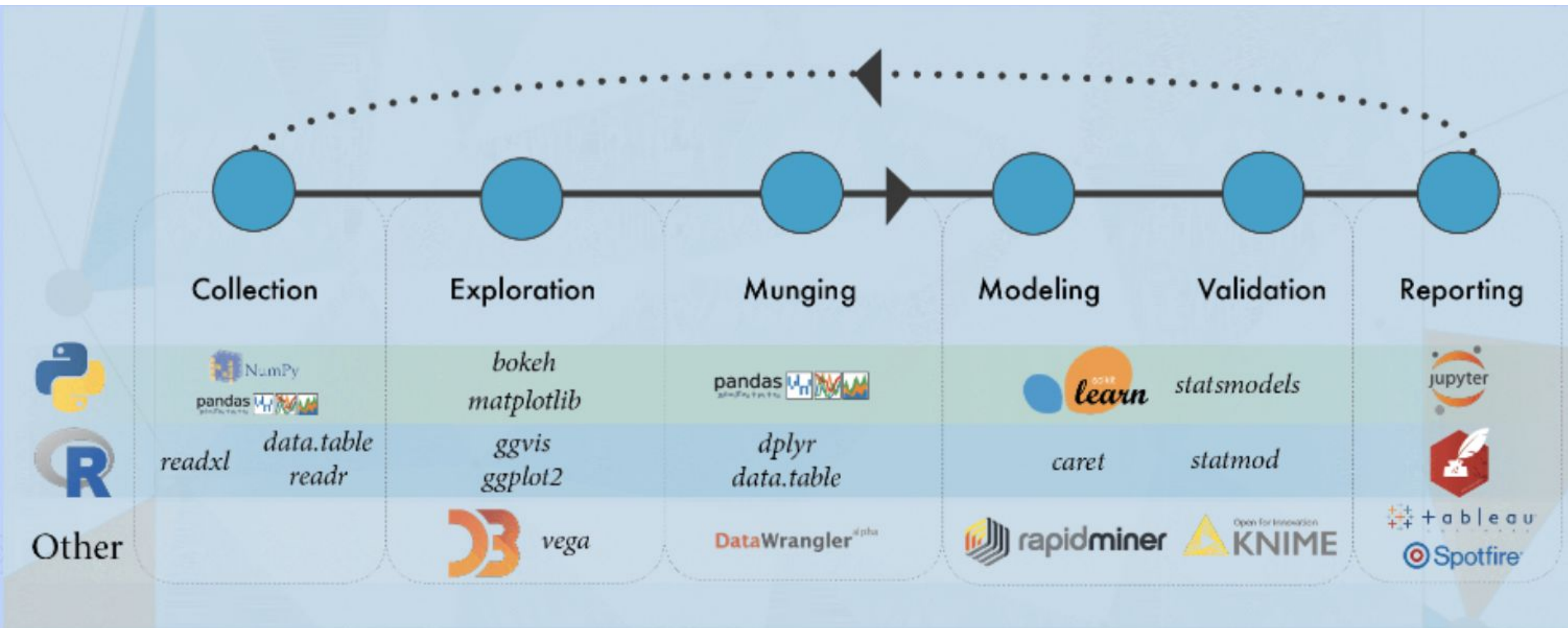
Los datos nunca están limpios

# ¿ POR QUÉ LIMPIAR LOS DATOS?

- Asegura la calidad de los datos que vamos a procesar.
- Evita la información no veraz o errónea.
- Ahorra costes de espacio en disco al eliminarse la información duplicada.
- Agiliza las consultas por la ausencia de datos repetidos o inservibles.
- Ayuda a tomar decisiones estratégicas correctas.

# ¿ CUALES SON LOS PROBLEMAS COMUNES DE DATOS.?

- Nombres inconsistentes de columnas (mayúsculas o caracteres malos)
- Datos perdidos
- Outliers
- Columnas duplicadas
- Desordenados (transformarlos en datasets)
- Necesidad de procesar las columnas



## CONVIRTIENDO A DATOS ORDENADOS

Mejor para informar vs. mejor para el análisis

- Tabla A: Mejor para informar.

	Name	treatment_a	treatment_a
0	Daniel	-	42
1	Jhon	12	31
2	Jane	24	27

- Tabla B: Mejor para el análisis

	Name	treatment_a	Value
0	Daniel	treatment_a	-
1	Jhon	treatment_a	31
2	Jane	treatment_a	27
3	Daniel	treatment_b	42
4	Jhon	treatment_b	31
5	Jhon	treatment_b	27