# Raw dataset

## 330,850 observations

# First elimination based on ‘category of product’

## -181,335 = 149,515 observations

A graph with a bar graph

Description automatically generated

1 bactericida 18

2 otro 321

3 aditivo 1492

4 mejorador de suelo 4983

5 fungicida 5876

6 fertilizante 17607

7 biofertilizante 18789

8 fertilizante foliar 28206

9 insecticida 58816

10 herbicida 84805

11 fertilizante quimico 109937

Majority are fertilizers. Only keep herbicida (84,805), insecticida (58,816), fungicida (5,876), bactericida (18). Total of **149,515 entries**

# Elimination based on unit applied

## -10,715 = 138,800 observations

A graph with numbers and a bar graph

Description automatically generated

NAs are removed, paquetes are removed, piezas are removed, and kg\_planta is removed.

# Make active ingredient the observation unit

## +19,547 = 158,347 observations

Most products contained only 1 active ingredient, some two, etc.

A screenshot of a computer code

Description automatically generated

# Eliminate biological control entries (parasitic wasps)

## -11 = 158,336 observations

# Reasonable quantities applied

Imidacloprid shouldn’t be applied at rates higher than 10 l/ha. Also not sure what a kg means, as it is a liquid, so eliminate instances when kg is the unit (4) and when L applied is higher than 10 (3 instances).