EPM's Assets Maintanance - Descriptive Analysis

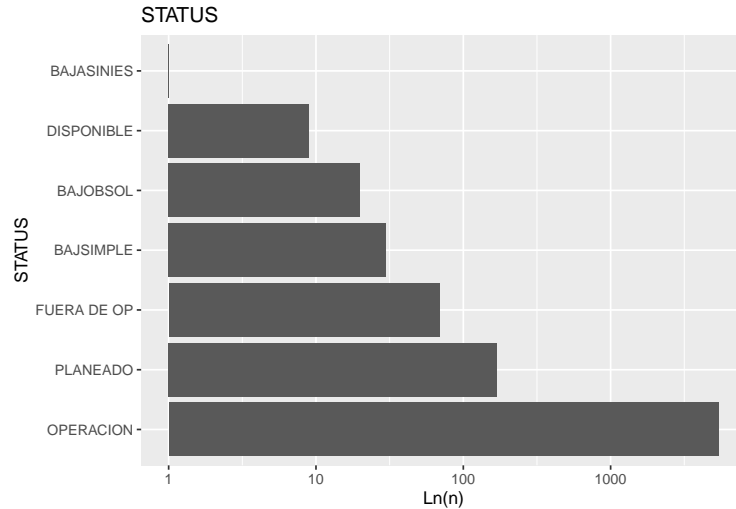## Team 74 - DS4A by Correlation1

epm®

estamos ahí.

Dataset: `activos.csv`

This dataset contains 5699 rows and 11 columns. In order to get a deeper understanding of the variables, it is also imporant to see how many values exist for each column.
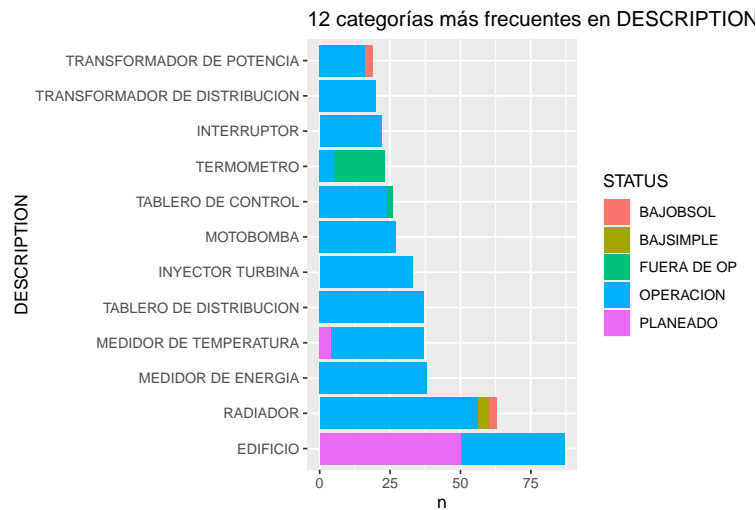
```
## Rows: 1
## Columns: 11
## $ ASSETNUM         <int> 5699
## $ DESCRIPTION      <int> 4675
## $ ASSETTAG         <int> 1792
## $ LOCATION         <int> 1450
## $ SLXJDETECNICO    <int> 5603
## $ INSTALLDATE      <int> 420
## $ STATUS           <int> 7
## $ SITEID           <int> 1
## $ PRIORITY         <int> 4
## $ CLASSSTRUCTUREID <int> 187
## $ CLASSSTRUCTURE   <int> 186
```

Since `ASSETNUM` variable has the same number of distinct values as the whole dataset has rows, it could be taken as an indicator, apparently a number of seven digits, *e.g.* it is not of our analytic interest by now. The Variable `SITUATED` only has 1 value, *i.e.* it is a constant, while the `STATUS` variable has 7 different values, which makes it of particular interest:
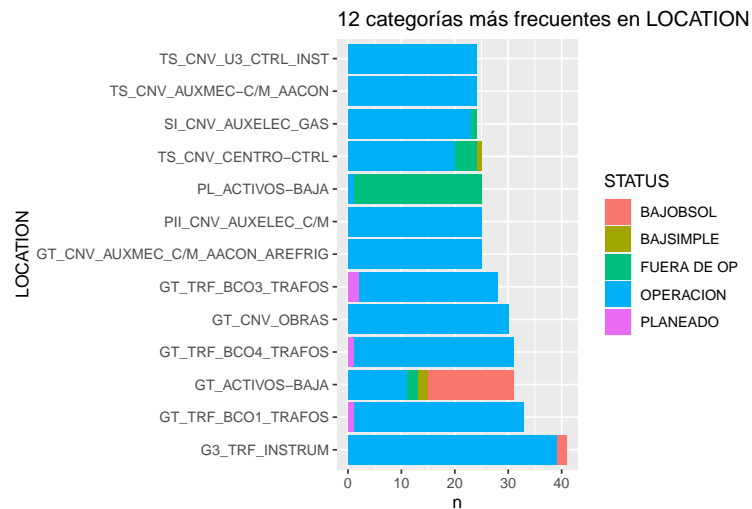
STATUS

In the `STATUS` variable, the category `OPERACION` is the most frequent, followed by `PLANEADO`. It is important to stress that the scale of the frequency have been transformed through a logarithmic function. This was done because the first category was far more frequent than the second one, so the use of a typical scale would make harder to see the remaining categories' relevance.

Despite the high number of unique values in variable `DESCRIPTION`, the highest 12 categories were retrieved for the next plot, and then crossed versus `STATUS`:

12 categorías más frecuentes en DESCRIPTION

Recalling that this graphic is built just with the 12 more important categories of the variable in `DESCRIPTION`, the category `OPERACION` seems to be the most important. The category `PLANEADO` exists only in two categories of the variable `DESCRIPTION`: `EDIFICIO` and barely in `MEDIDOR DE TEMPERATURA`.

12 categorías más frecuentes en LOCATION

According to the data dictionary, `ASSETTAG` is an index variable. Above it can be seen that there are no acummulations.

12 categorías más frecuentes en ASSETTAG de 'activos.csv'

12 categorías más frecuente

DESCRIPTION Vs. PRIORITY

Histórico PRIORITY por mes