

Science toolkit: Data analysis and making graphs

The importance of data analysis

One of the important aspects when conducting an investigation is to convert the collected data into useful information. Data analysis can help us to verify the initial hypotheses of the investigation, as well as to interpret the results of the data collection. One of the most interesting tools is graphic representation, since it allows us to convert the data into easily interpretable information. The graphics are used to illustrate and present a set of data, so as to facilitate their understanding, comparison and analysis. The graphics will depend on the type of variables that are being represented.

Types of variables

The type of graph that can be used depends on the type of variables. The types of variables are:

- **Categorical Variables (or qualitative variables):** Those variables that represent a quality of the individuals studied. For example: Area (North-South), Species (Amphibious, Fish, Bird ...), Part of day (Morning, afternoon, night), ...
- **Continuous Variable (or quantitative variables):** Those variables that represent the number of individual variables studied. For example: PH, COD, Conductivity, number of birds, number of fish, ...

Some variables can be transformed to offer useful information for the interpretation of the data:

- **Dates and hours:** The variables dates and times can be transformed into qualities related to them. For example: Sample 1 taken on 2-12-20, 10.40 a.m. can be transformed into several variables such as: Day of Week: Monday, Month: December, Year: 2020, Season: Winter, Time of day: Morning,...
- **Georeferences:** Georeferenced points can be transformed into qualities such as Area (Pinedo-Júcar), position (North-South),...
- **Continuous variables in qualities:** Some continuous variables may interest us to represent them as qualities. For example: the Ph measured in acid-neutral-basic, the amount of fish in few-enough-many, ...



Data structure

One of the main tasks is to structure the data properly. The simplest way * is to represent in the rows the items studied (species, sample, people, ...) and in the columns the variables.

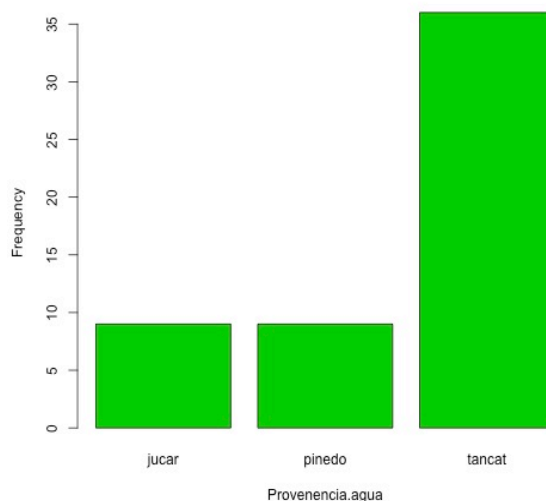
* This format is correct for the use of statistical packages. When spreadsheets are used, calculations must be made in the data to obtain certain graphs.

Individuos	Ubicación	Lugar	Hora	Franja	pH	Oxígeno disuelto	Conductividad	Temperatura
muestra 1	Entrada	Entrada	10:45	Mañana	7,92	117,6	1779	20,4
muestra 2	Centro	Entrada	11:01	Mañana	7,8	117,4	1712	21
muestra 3	Tancat de Mila	Entrada	11:28	Mañana	7,79	114,8	1703	18,3
muestra 4	Oeste	Entrada	12:03	Mediodía	7,78	115	1714	20
muestra 5	G.Perellonet- Salida	Salida	12:44	Mediodía	7,75	115,4	1641	18
muestra 6	G.Puçol- Salida	Salida	13:18	Mediodía	7,86	127,6	1640	18,5
muestra 7	Entrada	Entrada	10:45	Mañana	8,2	116,9	1750	14,5
muestra 8	Centro	Entrada	11:01	Mañana	8,5	117,7	1540	17,5
muestra 9	Tancat de Mila	Entrada	11:28	Mañana	8,5	115,3	1610	17,6
muestra 10	Oeste	Entrada	12:03	Mediodía	8,6	115,8	1560	17,7
muestra 11	G.Perellonet-Salida	Salida	12:44	Mediodía	8,4	115,2	1520	17,6
muestra 12	G.Puçol-Salida	Salida	13:18	Mediodía	8,5	126,9	1520	17,5

Graphs for categorical variables

- Bar Graph

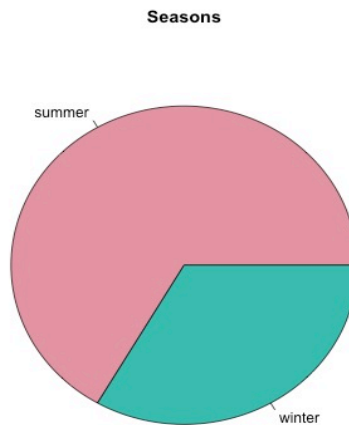
The number of elements or the percentage of each category can be displayed on the bar chart.





- **Pie chart**

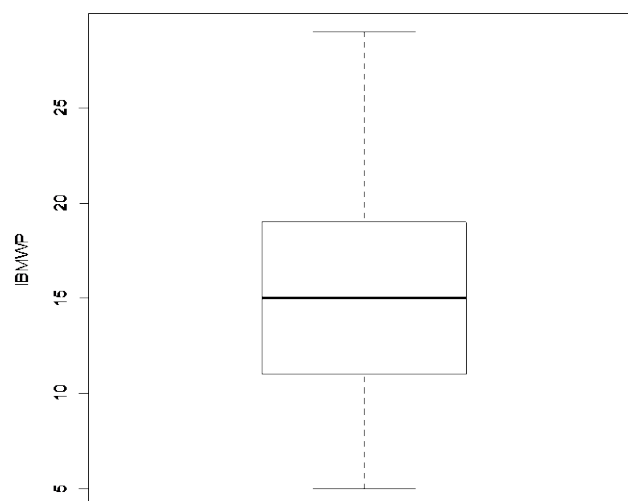
The pie chart is like a bar chart but using polar coordinates.



Graphs for quantitative variables

- **Box and whiskers chart**

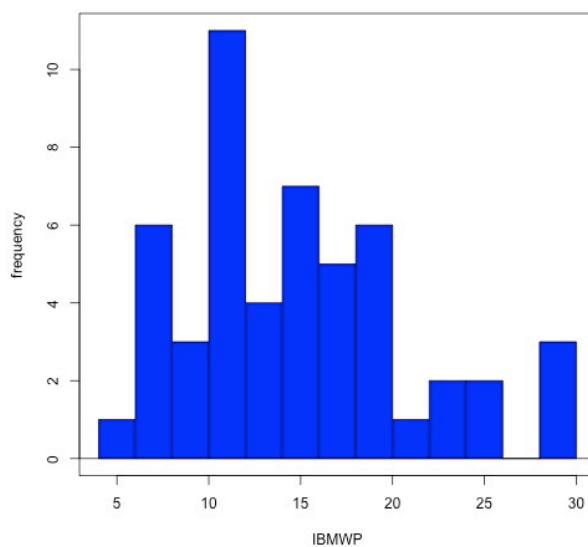
The box plot is a method of graphically representing groups of numerical data through their quartiles and extreme values.





- **Histogram**

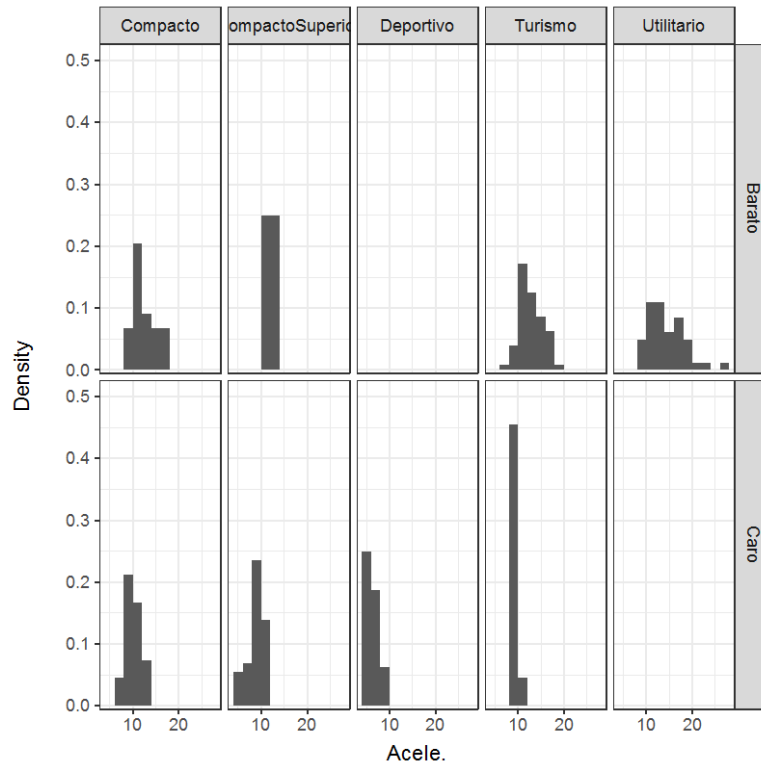
A histogram is a precise representation of the distribution of numerical data.



Graphics with several variables

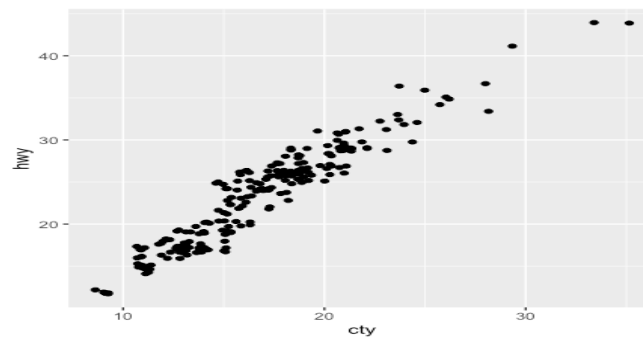
- **Quantitative variable vs. qualitative variable**

In this case, the graphs for quantitative variables (histogram, box graphs) can be made taking into account different factors of some qualitative variable.



- **Quantitative vs. Quantitative Variable**

The scatter plot is the most useful graph to analyze two quantitative variables and study their relationship.





- **Quantitative variables vs. time variable**

Time series are used to analyze the evolution of a variable over time.

