

5. Introduction and objectives

Will be devoted to apply exploratory multidimensional methods to "New York taxis" data.

The main objectives are to understand the relationships between the variable `Total_amount` and the other variables. The multivariate methods are useful to enlighten the relationships between this variable and the others in a very flexible way.

This will allow to establish the different patterns of behaviors of the taxis through using clustering methods.

6. Supplementary coding

Recode the pickup and dropoff variables into factors. Explain the reason for this recoding. It could be interesting to create the variable "actual velocity during the trip" and also the variable computing the part of the amount due to the total distance of the trip.

6.1. Analysis of the variables influencing the total-amount through PCA.

A PCA applied to these variables will provide a synthesis of them. The other variables, available in the file, have to be used as illustrative. `Total_amount` is also used as supplementary. Why?

A trip typology is obtained by clustering the trips from their principal coordinates. An appropriate number of principal components has to be selected in the clustering step.

The clusters will be described from all the variables. First conclusions are drawn on the results obtained through both complementary methods (ACP + clustering).

7. Analysis strategy

After obtaining a first view as described above, you have to define a clear analysis strategy in order to achieve the objectives that you have chosen (among those that are listed at Section 1). This strategy is specific to each group. Different strategies are possible.