Gasca Falcon Monica Gabriela

Flores Sanchez Jose de Jesus

Pernet Diaz Infante Thomas Carlos Cedric

Vazquez Villaseñor Luis Fernando

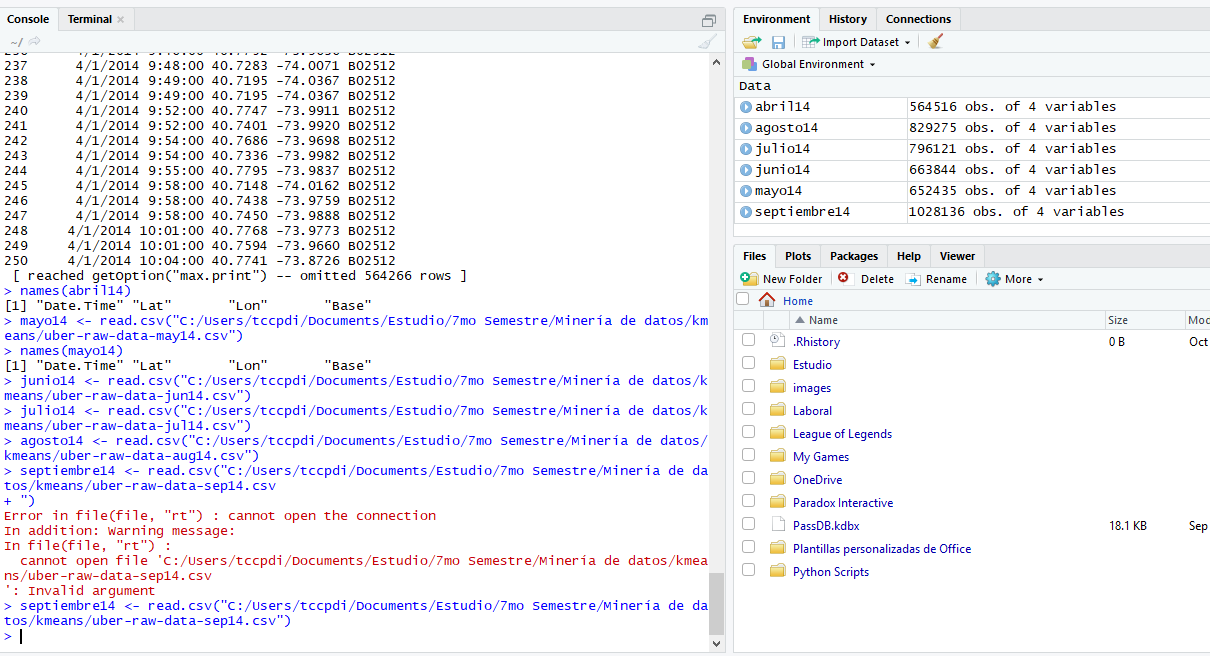
k-means

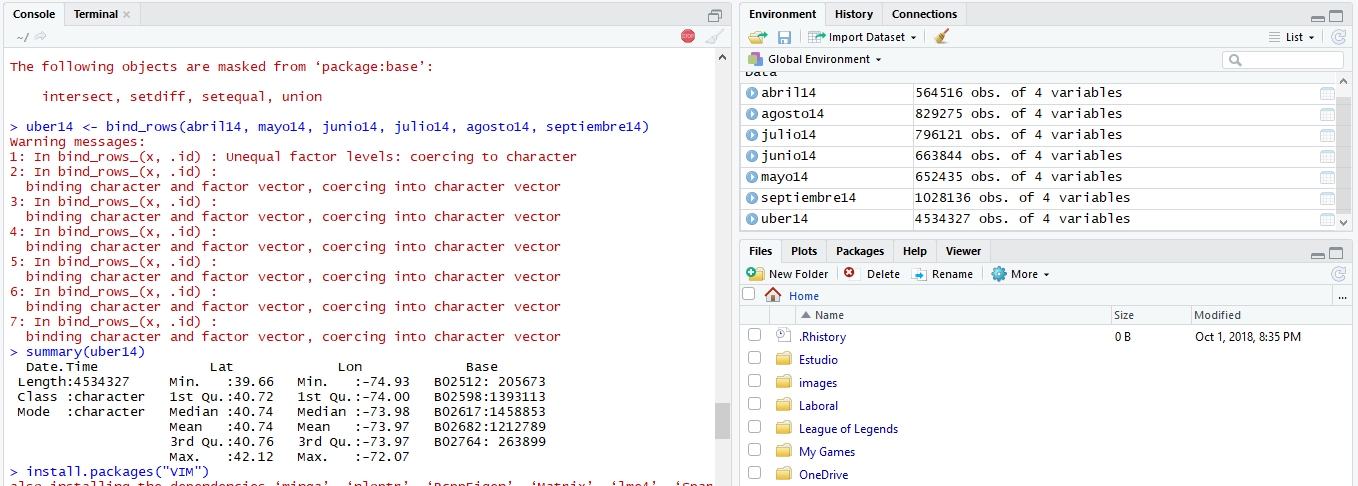
We used the uber database from April to September of the same year in New York and we applied the function of k-means that is included in R, using the maps of Stamen to plot the centroids with the map of New York

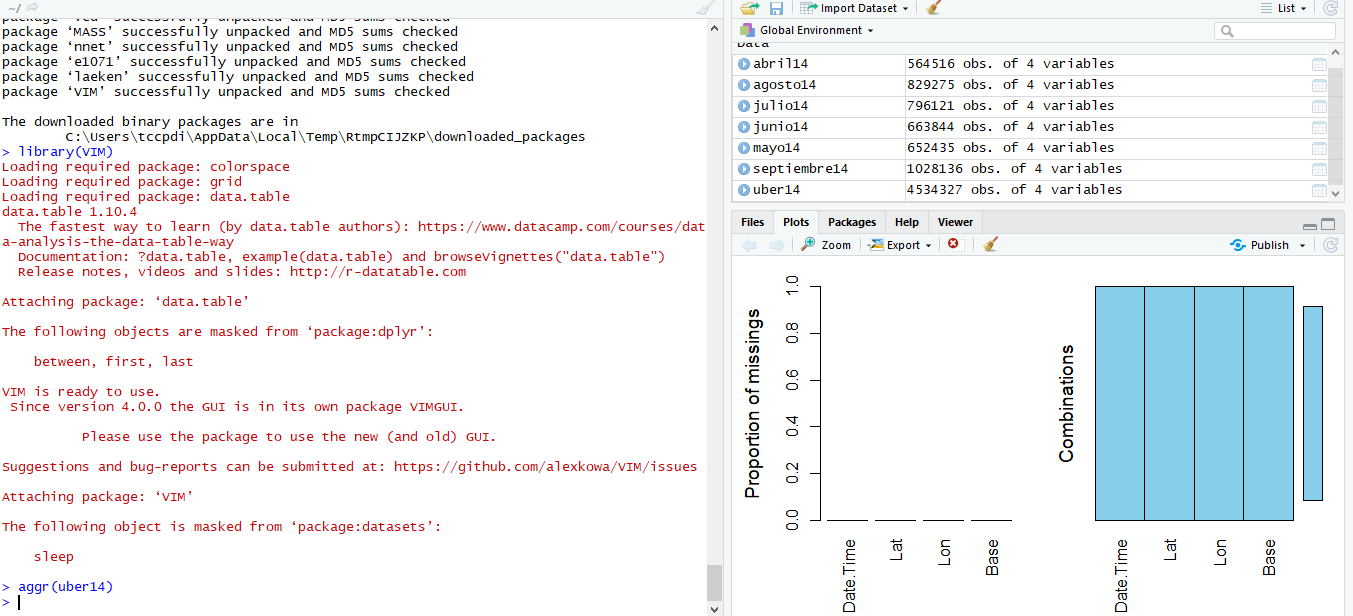
**Pseudocode**

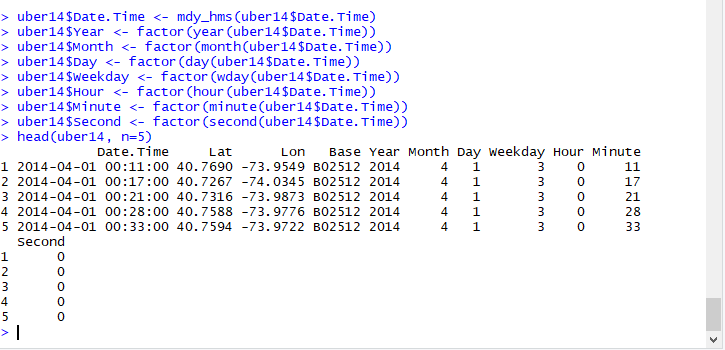
1. Start.
2. The connection to the database is made.
3. The data in the database is read.
4. The number of data in each group is printed.
5. The nearest centroid with the Euclidean square distance is obtained.
6. The centroids are recalculated.
7. The value of k is obtained.
8. It is graphed with the value of k and they are grouped with different color.
9. End

Load the .csv files into manipulable R variables:

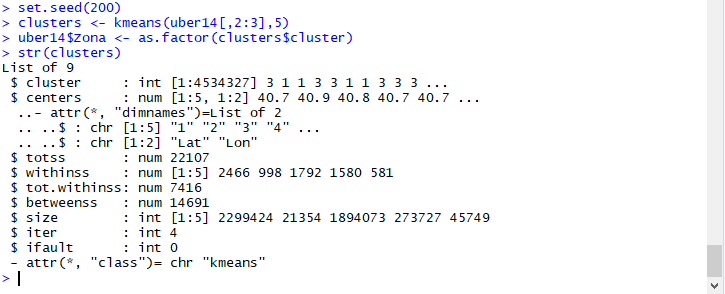


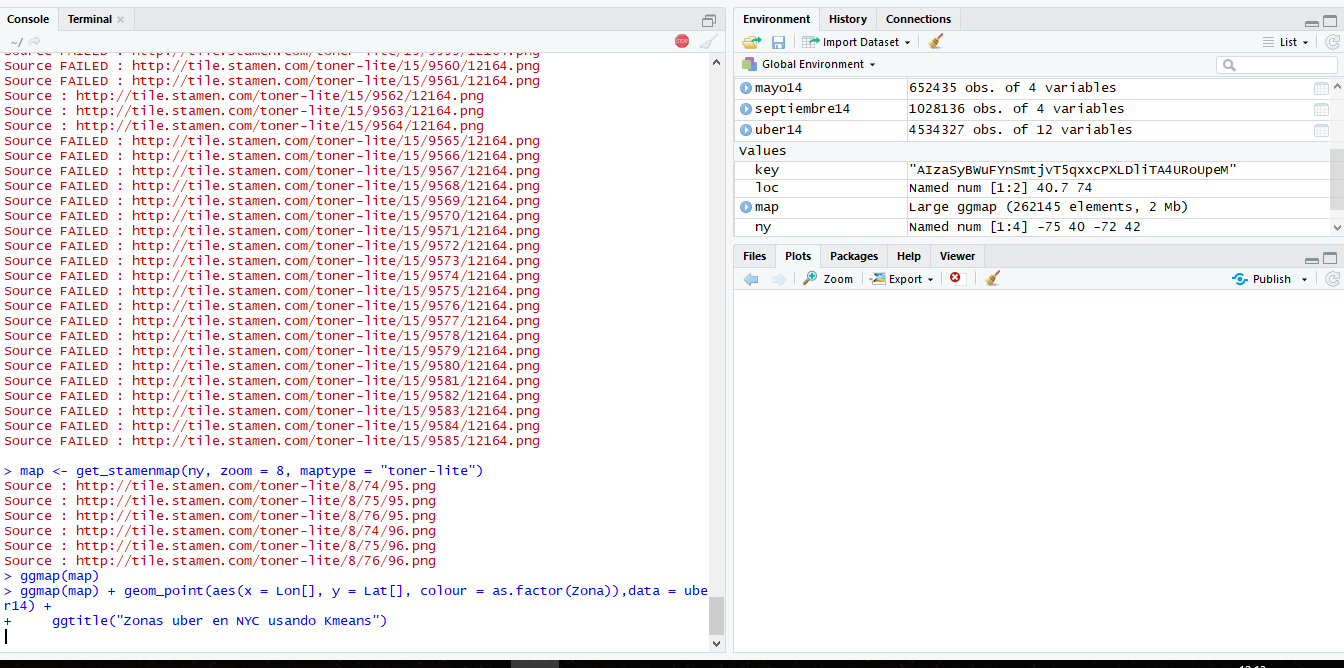
Join all the variables in one

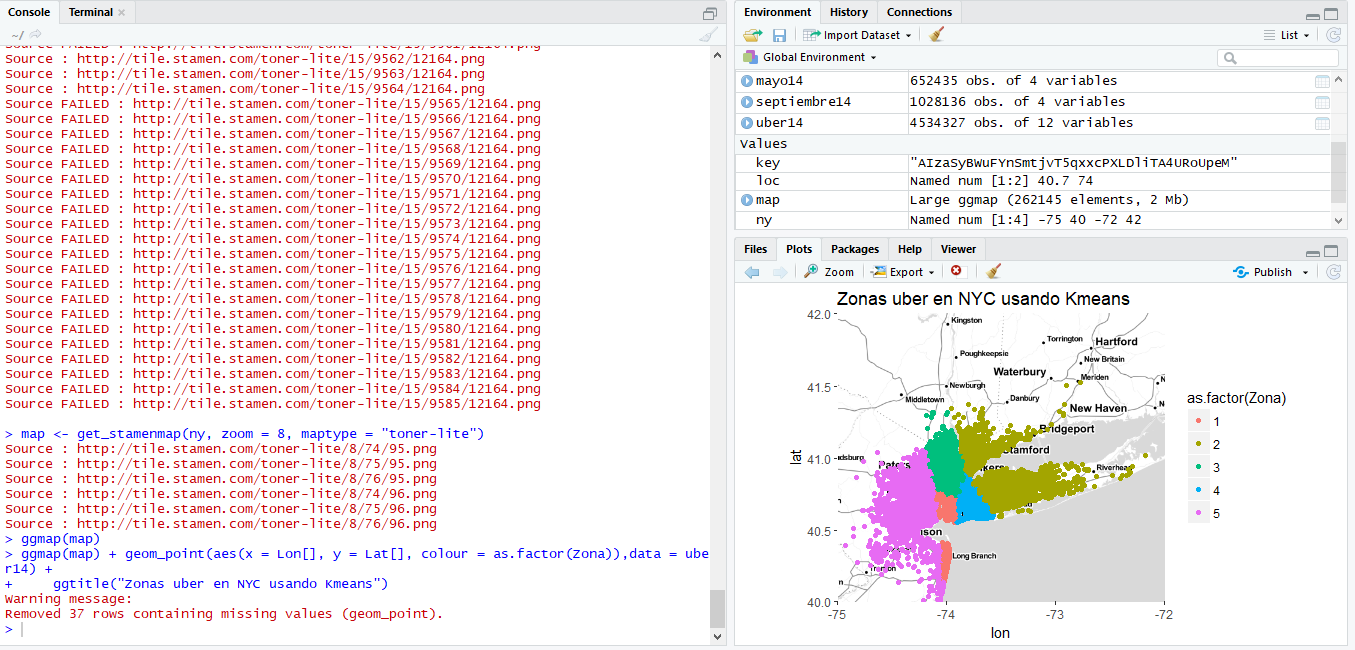
Verify that all information is correct and does not have invalid characters with VIM, by graphing it:

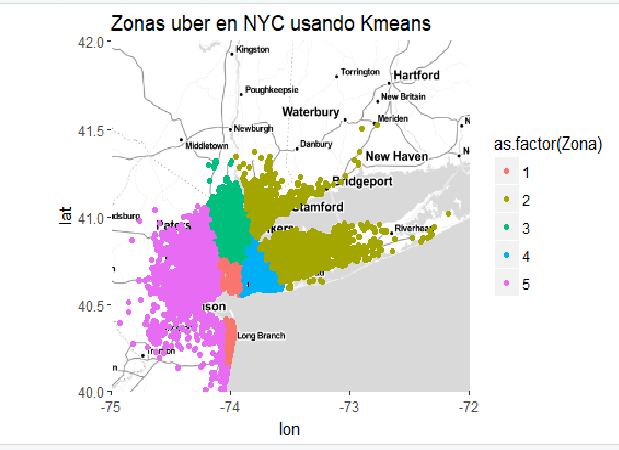
Convert DateTime column to Year, Month, Day, Week Day, Hour, and Minute columns:

Prepare "seed" for kmeans with number 200, to obtain the same results, this is known as "Soft Clustering". (set.seed (200))

Assign variable "clusters" with "kmeans" function, where columns and iterations are specified:

Obtain map of database "stamen" with library "ggmap", specify the coordinates in variable "ny" and obtain all the images from the internet, then plot the map with ggmap and add all the information calculated with kmeans, which will have its Real position on the map and assigned a centroid:



Result:

**Architectural diagram**

