Gasca Falcon Monica Gabriela

Flores Sanchez Jose de Jesus

Pernet Diaz Infante Thomas Carlos Cedric

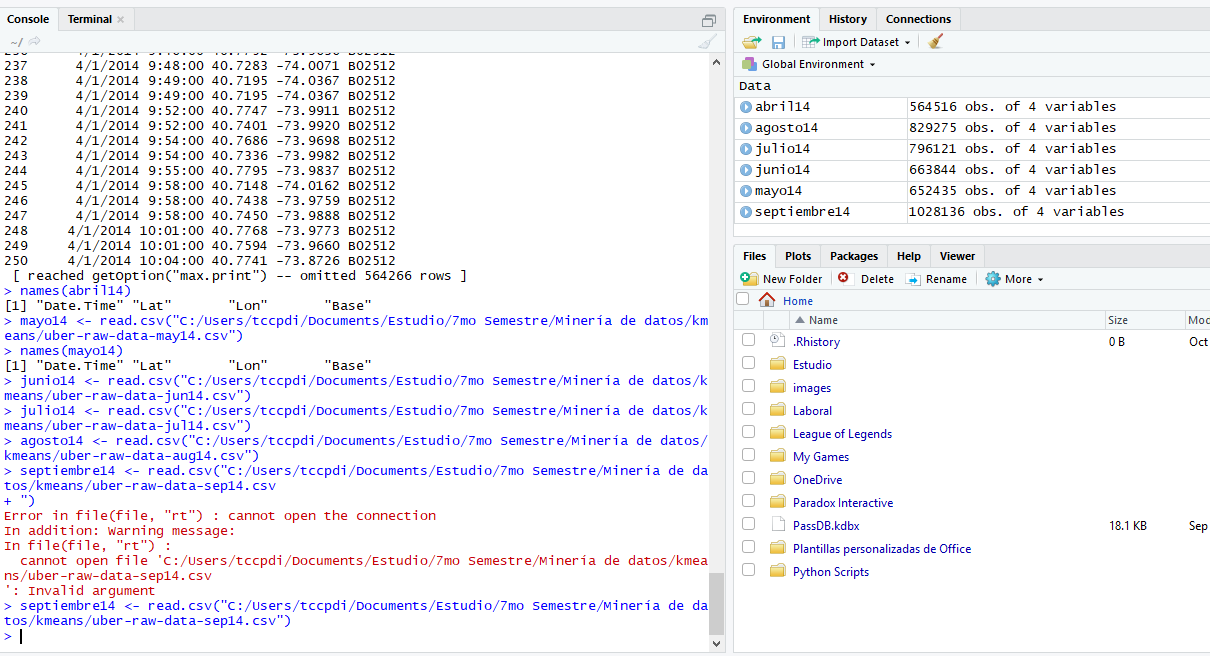
Vazquez Villaseñor Luis Fernando

k-means

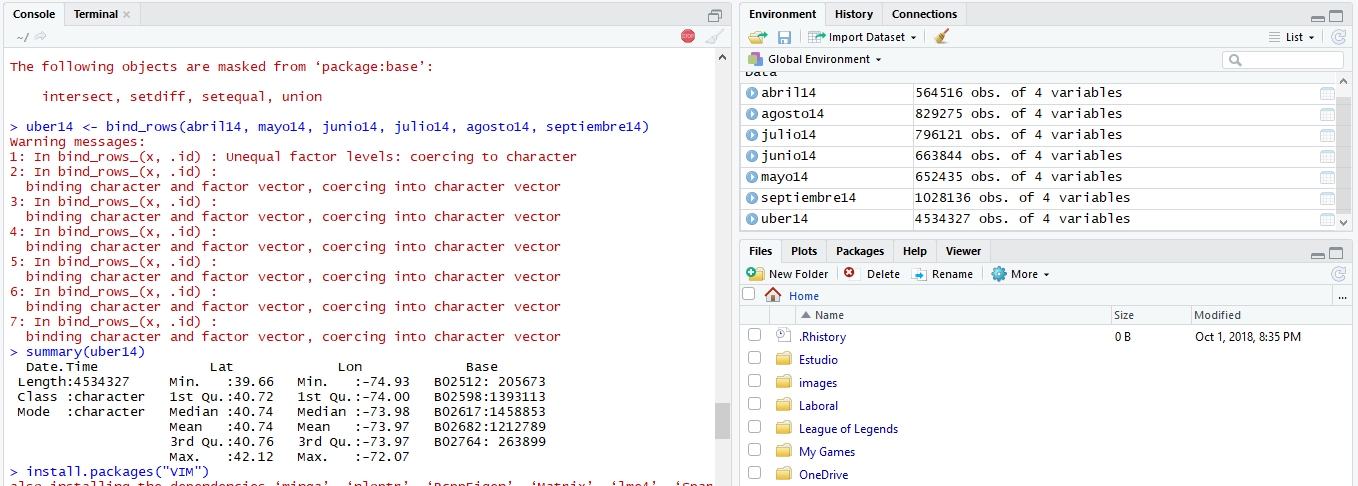
Pseudo Codigo

1. Inicio.
2. Se realiza la conexión con la base de datos.
3. Se leen los datos de la base de datos.
4. Se imprime el numero de datos que se tienen en cada grupo.
5. Se obtiene el centroide mas cercano con la distancia cuadrada Euclideana.
6. Se recalculan los centroides.
7. Se obtiene el valor de k.
8. Se grafica con el valor de k y se agrupan con diferente color.
9. Fin

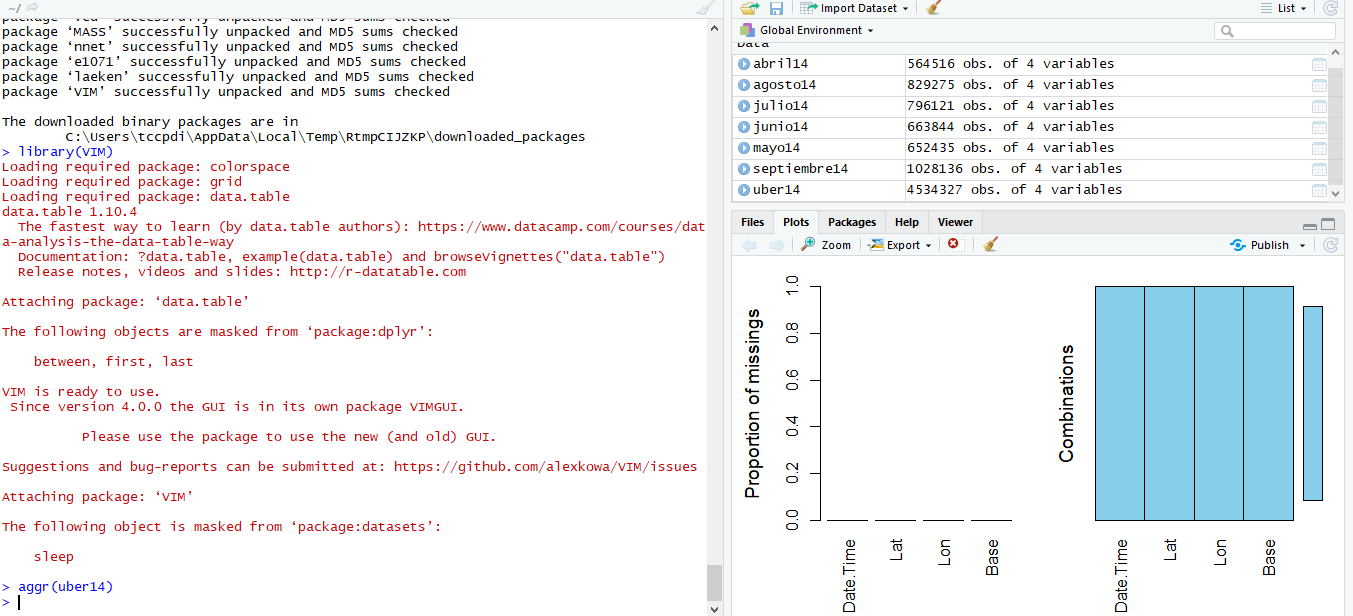
Cargar los archivos .csv en variables de R manipulables:



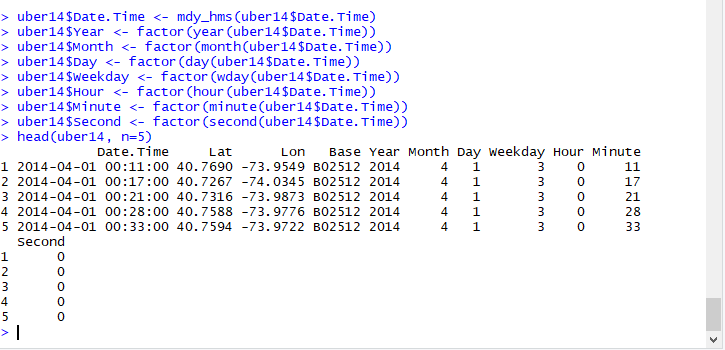
Unir todas las variables en una sola



Verificar que toda la información sea correcta y no tenga caracteres inválidos con VIM, graficándolo:

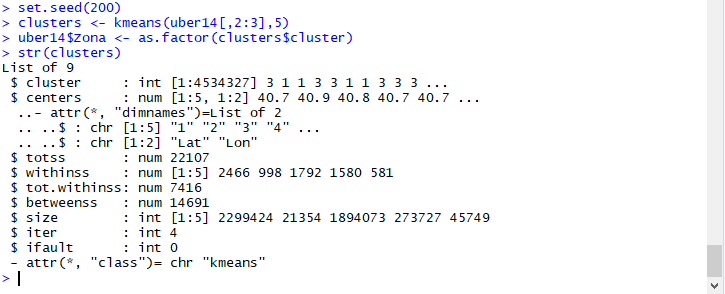


Convertir columna DateTime a columnas de Year, Month, Day, Weekday, Hour y Minute:

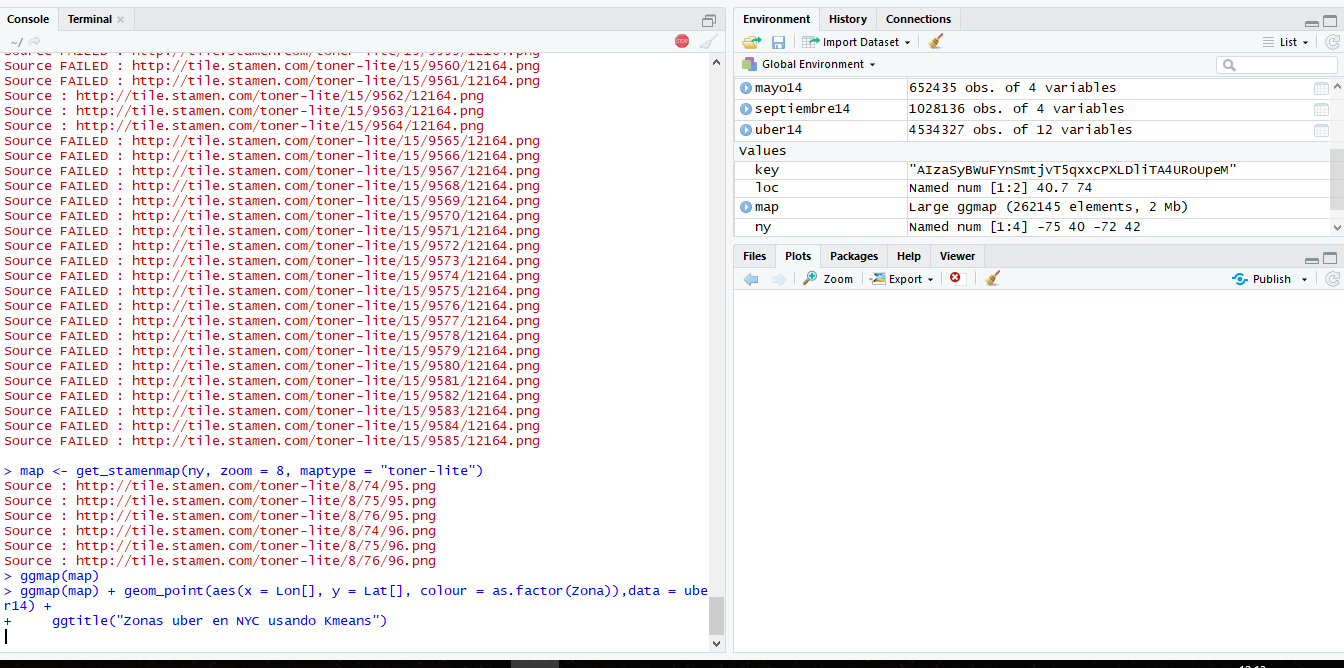


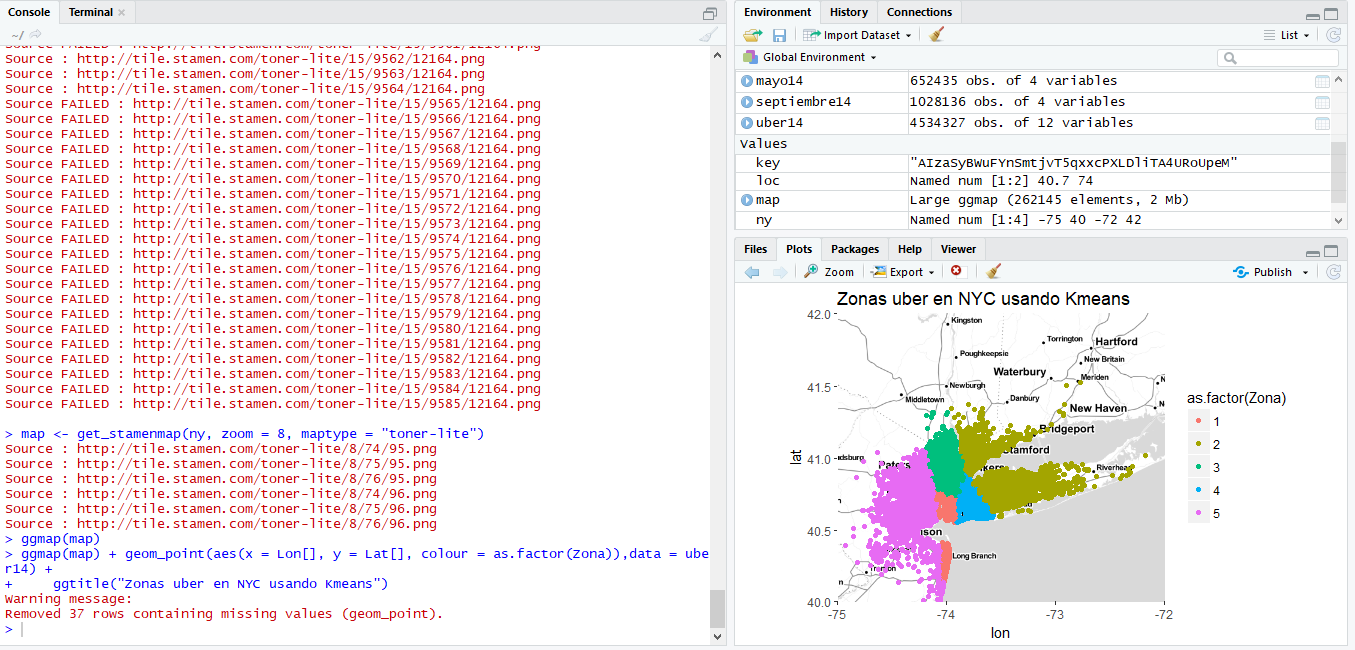
Preparar “semilla” para kmeans con número 200, para obtener los mismos resultados, esto se conoce como “Soft Clustering”. (set.seed(200))

Asignar variable “clusters” con función de “kmeans”, en donde se especifica las columnas y las iteraciones:

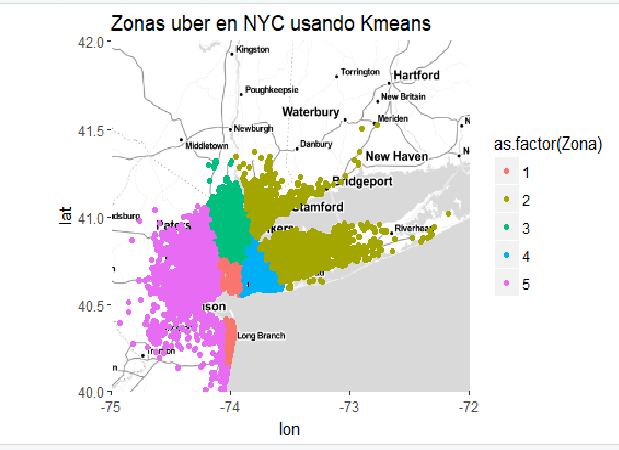


Obtener mapa de base de datos “stamen” con librería “ggmap”, especificamos las coordenadas en variable “ny” y obtiene todas las imágenes de internet, después ploteamos el mapa con ggmap y le sumamos toda la información calculada con kmeans, que tendrán su posición real en el mapa y se les asignara un centroide:





Resultado final:



**Diagrama de Tecnologias**

