

VINCULACIÓN

Trabajo individual

TEMA: Análisis de Clúster

Segmentación de Grupos:

Utilizar técnicas de clustering (como K-means) para identificar segmentos de la población con comportamientos y actitudes similares hacia la gestión de residuos.

WASHINGTON RAMIRO BONILLA VIMOS

Octubre, 2024

El Coca

ANÁLISIS DE CLÚSTER

Segmentación de Grupos: Utilizar técnicas de clustering (como K-means) para identificar segmentos de la población con comportamientos y actitudes similares hacia la gestión de residuos.

El análisis de clúster usando el método de K-means en estos datos. Primero cargaré y revisaré el contenido del archivo.

El archivo contiene una hoja llamada "GRUPO 1" con datos de una encuesta. Sin embargo, parece haber encabezados adicionales y filas vacías, lo que indica que los datos necesitan limpieza antes de ser utilizados para el análisis de clúster.

Para las técnicas de clustering se realizará el siguiente proceso de los datos, así como para realizar el análisis de clúster con K-means. El código incluirá:

1. Carga y limpieza del archivo.
2. Selección de las columnas relevantes para el análisis.
3. Aplicación de la técnica K-means para identificar los clústeres.

Instalar y cargar las librerías necesarias:

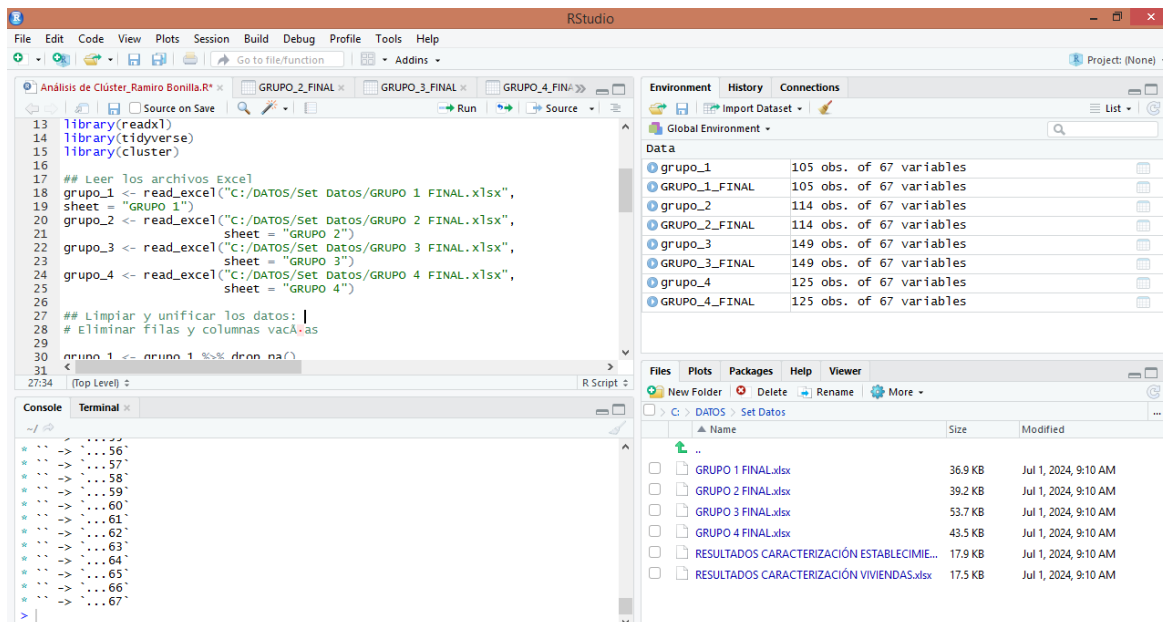
```
if (!require("readxl")) install.packages("readxl")
if (!require("dplyr")) install.packages("dplyr")
if (!require("factoextra")) install.packages("factoextra")
if (!require("magrittr")) install.packages("magrittr")
```

Cargar librerías

```
library(readxl)
library(magrittr)
library(dplyr)
library(factoextra)
```

Leer los archivos Excel

```
grupo_1 <- read_excel("C:/DATOS/Set Datos/GRUPO 1 FINAL.xlsx", sheet = "GRUPO 1")
grupo_2 <- read_excel("C:/DATOS/Set Datos/GRUPO 2 FINAL.xlsx", sheet = "GRUPO 2")
grupo_3 <- read_excel("C:/DATOS/Set Datos/GRUPO 3 FINAL.xlsx", sheet = "GRUPO 3")
grupo_4 <- read_excel("C:/DATOS/Set Datos/GRUPO 4 FINAL.xlsx", sheet = "GRUPO 4")
```



LIMPIAR Y UNIFICAR LOS DATOS:

Eliminar filas y columnas vacías

```
grupo_1 <- grupo_1 %>% drop_na()
grupo_2 <- grupo_2 %>% drop_na()
grupo_3 <- grupo_3 %>% drop_na()
grupo_4 <- grupo_4 %>% drop_na()
```

Cargar el archivo DE DATOS UNIFICADO

```
file_path <- "c:/DATOS/Set Datos/DATOS_FINAL.xlsx"
```

Omitir las primeras filas

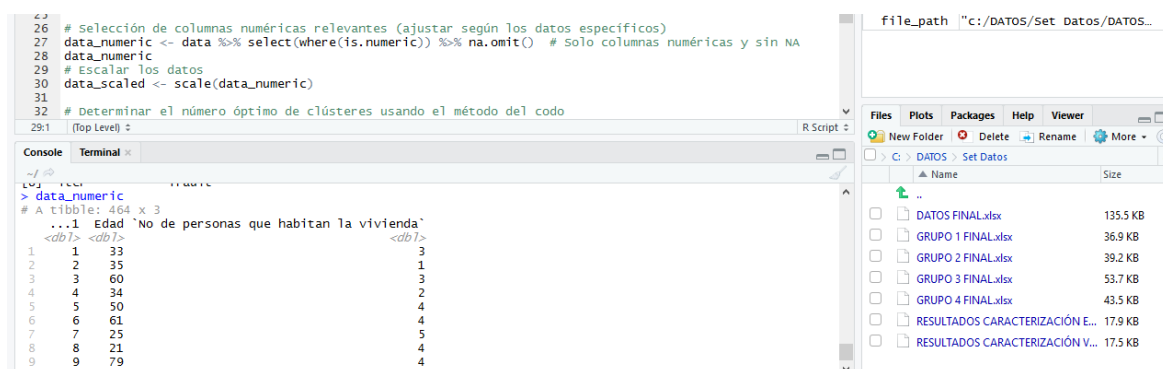
```
data <- read_excel("C:/DATOS/Set Datos/DATOS_FINAL.xlsx", sheet = "GRUPO 1",
skip = 4)
```

SELECCIONAR LAS COLUMNAS RELEVANTES PARA EL ANÁLISIS

Aquí debes ajustar la selección de columnas según las que sean relevantes

Selección de columnas numéricas relevantes (ajustar según los datos específicos)

```
data_numeric <- data %>% select(where(is.numeric)) %>% na.omit() # Solo columnas
numéricas y sin NA
```



REALIZAR LA NORMALIZACIÓN DE LOS DATOS: Normalizar los datos

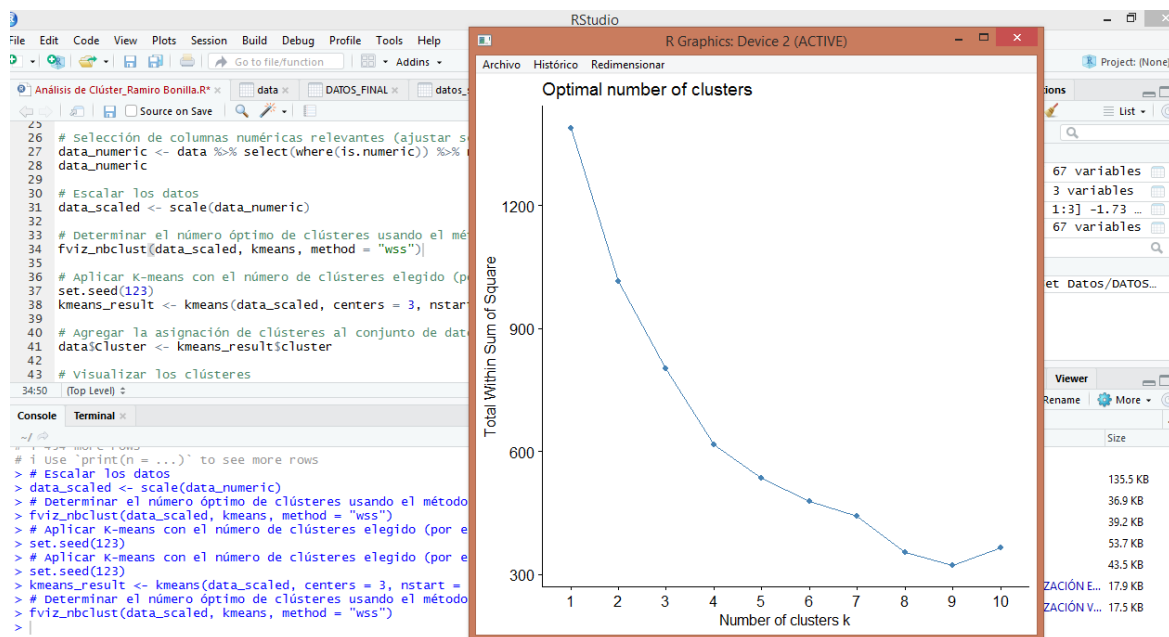
Escalar los datos

```
data_scaled <- scale(data_numeric)
```

DETERMINAR EL NÚMERO ÓPTIMO DE CLÚSTERES: Calcular el número óptimo de clústeres usando el método del codo

Determinar el número óptimo de clústeres usando el método del codo

```
fviz_nbclust(data_scaled, kmeans, method = "wss")
```



APLICAR K-MEANS CLUSTERING:

Aplicar K-means con el número de clústeres elegido (por ejemplo, k = 3)

```
set.seed(123)
```

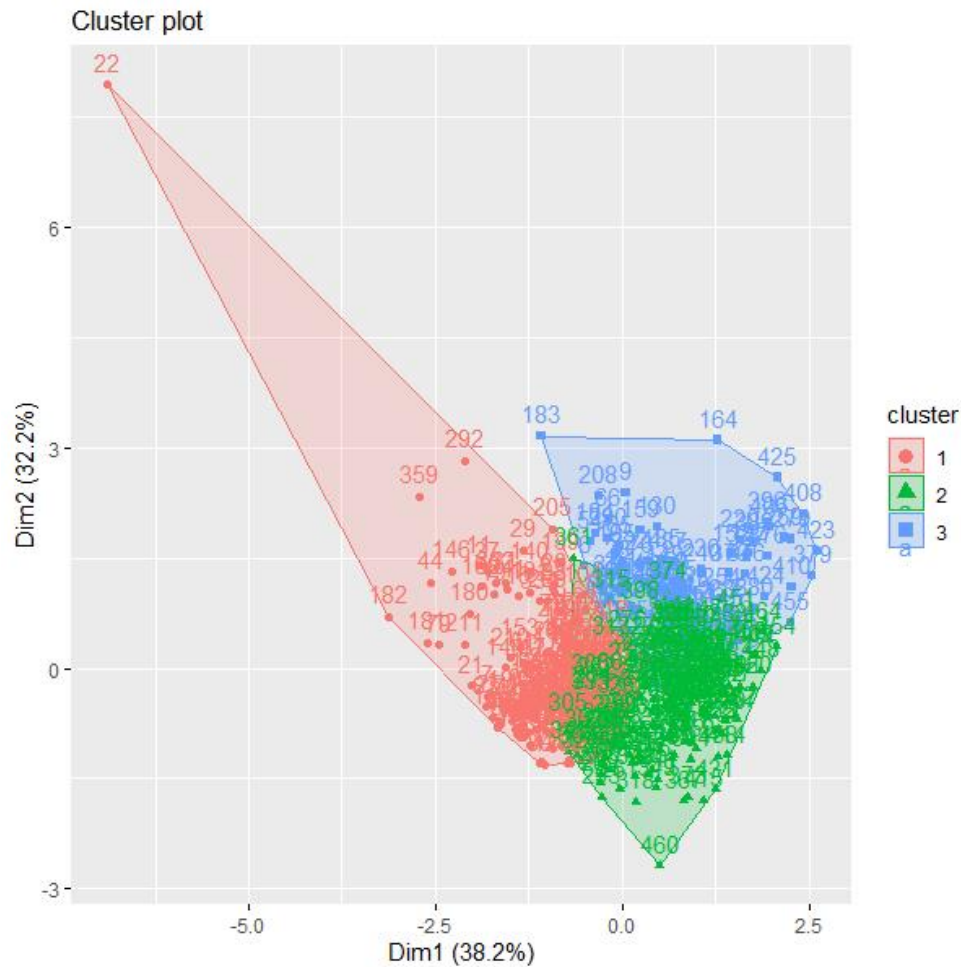
```
kmeans_result <- kmeans(data_scaled, centers = 3, nstart = 25)
```

Agregar la asignación de clústeres al conjunto de datos original

```
data$Cluster <- kmeans_result$cluster
```

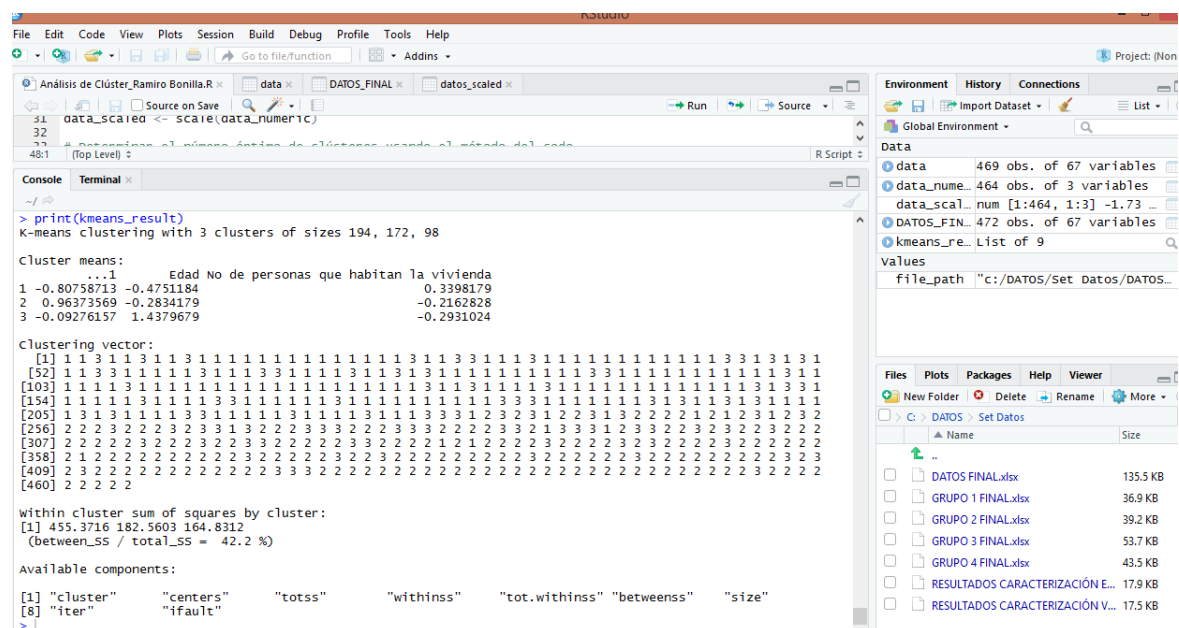
Visualizar los clústeres

```
fviz_cluster(kmeans_result, data = data_scaled)
```



Mostrar el resumen del resultado

```
print(kmeans_result)
```



EXPLICACIÓN DEL CÓDIGO:

- **Librerías:** readxl para leer el archivo Excel, dplyr para el manejo de datos y factoextra para visualización.
- **Carga y limpieza:** Se omiten filas iniciales y se seleccionan solo las columnas numéricas.
- **Escalamiento:** Los datos son escalados para asegurar una adecuada distancia en el análisis de clúster.
- **Método del codo:** Determina el número óptimo de clústeres.
- **K-means:** Se aplica el algoritmo con un número determinado de clústeres (centers = 3 en este caso, pero puedes ajustar).
- **Visualización:** Muestra los clústeres y asigna etiquetas a cada grupo.

SUBIR EVIDENCIA A GITHUB

LINK: https://github.com/SISTEMAS-2024/VINCULACION_MAESTRIA.git