

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

Source Visual B I <> Normal Format Insert Table Outline

```
---
title: "Practica 4 - 1 de 3"
author: "GRUPO 06"
format: html
editor: visual
---
```

GRUPO 6

INTEGRANTES:

- ALACOTE SALAS, Aylin Elisa.
- ESQUIVEL MOLERO, Lourdes Valeria|
- PALACIOS SALINAS, Danna Ariela.
- ROJAS CORREA, Aitana.
- VELASQUEZ ROMAN, Alvaro Fabian.

Instalamos y cargamos los paquetes:

```
{r}
install.packages("factoextra")
install.packages("<cluster")

{r}
library(factoextra)
library(cluster)
library(here)
library(rio)
library(tidyverse)
```

Environment History Connections Tutorial

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12
<input type="checkbox"/> bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

10File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

Source Visual B I <> Normal Format Insert Table Outline

```
library(here)
library(rio)
library(tidyverse)

here() starts at C:/Users/Alvaro/Documents
Some optional R packages were not installed and therefore some file formats are
not supported. Check file support with show_unsupported_formats()
— Attaching core tidyverse packages —
tidyverse 2.0.0 —
✔ dplyr 1.1.4 ✔ readr 2.1.5
✔ forcats 1.0.0 ✔ stringr 1.5.1
✔ lubridate 1.9.4 ✔ tibble 3.2.1
✔ purrr 1.0.4 ✔ tidyr 1.3.1— Conflicts — tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag() masks stats::lag()
ℹ Use the conflicted package to force all conflicts to become errors
```

Importamos los datos:

```
{r}
cancer_data <- import(here("estadistica_upsjb_estudio/Estadistica/data
/conoc_actit_factor_cancer_cervical.csv"))

Registered S3 method overwritten by 'data.table':
  method      from
print.data.table
```

Preparamos los datos - Solo datos numéricos:

```
{r}
cancer_data_1 = cancer_data |>
  select(-procedencia, -e_marital, -n_educacion, -religion, -etnia, -procedencia,
        -ocupacion, -ocupacion_convi, -antec_fam, -met_anticoncep, -antec_ets,
        -conocimiento, -actitud, -practica) |>
  column_to_rownames("paciente_num")
```

Preparamos los datos - Solo datos numéricos:

```
{r}
cancer_data_1 = cancer_data |>
  select(-procedencia, -e_marital, -n_educacion, -religion, -etnia, -procedencia,
        -ocupacion, -ocupacion_convi, -antec_fam, -met_anticoncep, -antec_ets,
        -conocimiento, -actitud, -practica) |>
  column_to_rownames("paciente_num")
```

Standardizamos:

```
{r}
cancer_data_escalado = scale(cancer_data_1)
```

Hechamos un vistazo a los datos antes del escalamiento:

```
{r}
head(cancer_data_1)
```

Description: df [6 x 15]

	edad<chr>	e_marital<chr>	n_educac...<chr>	religion<chr>	etnia<chr>	ocupacion<chr>
1	53	casada	superior	catolico	mestizo	otro
2	54	casada	superior	catolico	mestizo	empleada
3	26	soltera	superior	catolico	blanco	estudiante
4	25	soltera	superior	ninguna	mestizo	sin empleo
5	25	soltera	superior	ninguna	mestizo	estudiante
6	38	soltera	superior	catolico	mestizo	ama de c...

Environment History Connections Tutorial

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12
<input type="checkbox"/> bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

10File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

Source Visual B I <> Normal Format Insert Table Outline

```
library(here)
library(rio)
library(tidyverse)

here() starts at C:/Users/Alvaro/Documents
Some optional R packages were not installed and therefore some file formats are
not supported. Check file support with show_unsupported_formats()
— Attaching core tidyverse packages —
tidyverse 2.0.0 —
✔ dplyr 1.1.4 ✔ readr 2.1.5
✔ forcats 1.0.0 ✔ stringr 1.5.1
✔ lubridate 1.9.4 ✔ tibble 3.2.1
✔ purrr 1.0.4 ✔ tidyr 1.3.1— Conflicts — tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag() masks stats::lag()
ℹ Use the conflicted package to force all conflicts to become errors
```

Importamos los datos:

Preparamos los datos - Solo datos numéricos:

Preparamos los datos - Solo datos numéricos:

Standardizamos:

Hechamos un vistazo a los datos antes del escalamiento:

Description: df [6 x 15]

Environment History Connections Tutorial

Data

Values

Files Plots Packages Help Viewer Presentation

Install Update

EnvironmentHistoryConnectionsTutorial

Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Ahora con el escalado:

```
{r}
head(cancer_data_escalado)
```

	edad	edad_relacion_sexual	parejas_sex	num_hijos
1	0.6716910	1.5824717	-0.2996692	0.71040667
2	0.7601676	-0.5709542	-0.8882088	0.71040667
3	-1.7171751	-0.5709542	0.8774100	-0.04505018
4	-1.8056516	-0.5709542	4.4086475	-1.55596388
5	-1.8056516	-0.5709542	2.0544892	-1.55596388
6	-0.6554568	0.2904162	-0.2996692	-0.04505018

Calculo de distancias:

```
{r}
dist_cancer_data <- dist(cancer_data_escalado, method = "euclidean")
```

EnvironmentHistoryConnectionsTutorial

Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Visualizamos las distancias euclidianas en un mapa de calor:

```
{r}
fviz_dist(dist_cancer_data)
```

Nuestro mapa de calor donde se visualiza la matriz de distancias euclidianas de nuestras muestras de

(Top Level) ↕

EnvironmentHistoryConnectionsTutorial

Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Nuestro mapa de calor donde se visualiza la matriz de distancias euclidianas de nuestras muestras de datos de "dist_cancer_data". Los tonos más claros (naranja/blanco) indican distancias menores (mayor similitud), mientras que los tonos más oscuros (violeta/azul) representan distancias mayores (menor similitud). Se observa una estructura de bloques, sugiriendo la **presencia de clústeres o agrupaciones naturales dentro de los datos**, donde **las muestras dentro de un bloque son más similares entre sí que con las muestras de otros bloques**.

Usamos el metodo de agrupamiento: Funcion de enlace (linkage)

```
{r}
dist_link_cancer_data <- hclust(d = dist_cancer_data, method = "ward.D2")
```

Empleamos los dendogramas par la visualizacion de los patrones:

```
{r}
fviz_dend(dist_link_cancer_data, cex = 0.7)
```

Cluster Dendrogram

20 -

EnvironmentHistoryConnectionsTutorial

Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

FilesPlotsPackagesHelpViewerPresentation

InstallUpdate

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12
<input type="checkbox"/> bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1* cancer_data_escalado cancer_data_1 conoc_actit_factor_cancer_cervical

Render on Save Render Run Publish

Source Visual B I <> Normal Format Insert Table Outline

Empleamos los dendrogramas par la visualizacion de los patrones:

```
{r}
fviz_dend(dist_link_cancer_data, cex = 0.7)
```

Cluster Dendrogram

Height

(Top Level) Quarto

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1* cancer_data_escalado cancer_data_1 conoc_actit_factor_cancer_cervical

Render on Save Render Run Publish

Source Visual B I <> Normal Format Insert Table Outline

En nuestro dendrograma, vemos la representación gráfica del análisis de clúster jerárquico. En el eje Y, "Height", indica la distancia o disimilitud a la que los clústeres se fusionan. Las ramas que se unen a alturas bajas representan muestras o grupos de muestras muy similares. A medida que se asciende en el dendrograma, las uniones ocurren a mayores distancias, indicando la formación de clústeres más grandes y heterogéneos. La estructura ramificada revela la jerarquía de las agrupaciones, permitiendo identificar subgrupos dentro de los datos.

Revisamos los grupos que se formaron en el dendrograma:

```
{r}
fviz_dend(dist_link_cancer_data,
  k = 3,
  cex = 0.5,
  k_colors = c("#e700a9", "#00ce70", "#e70400"),
  color_labels_by_k = TRUE,
  rect = TRUE)
```

Cluster Dendrogram

Height

(Top Level) Quarto

Console

Go to file/function Addins

Untitled1* cancer_data_escalado cancer_data_1 conoc_actit_factor_cancer_cervical

Render on Save Render Run Publish

Source Visual B I <> Normal Format Insert Table Outline

Revisamos los grupos que se formaron en el dendrograma:

```
{r}
fviz_dend(dist_link_cancer_data,
  k = 3,
  cex = 0.5,
  k_colors = c("#e700a9", "#00ce70", "#e70400"),
  color_labels_by_k = TRUE,
  rect = TRUE)
```

Cluster Dendrogram

Height

Project (None)

Environment History Connections Tutorial

Import Dataset 17 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12
<input type="checkbox"/> bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

Project (None)

Environment History Connections Tutorial

Import Dataset 17 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12
<input type="checkbox"/> bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

Environment History Connections Tutorial

Import Dataset 17 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1* cancer_data_escalado cancer_data_1 conoc_actit_factor_cancer_cervical

Render on Save Render Run Publish

Source Visual B I </> Normal Format Insert Table Outline

Se muestran los resultados de un análisis de clúster jerárquico, donde se han identificado y visualizado K=3 grupos distintos, según la instrucción del código (k=3). Cada color (magenta, verde, rojo) representa uno de los tres clústeres. Las líneas punteadas verticales indican los límites de estos clústeres. La "Height" en el eje Y sigue representando la disimilitud. La segmentación visual facilita la interpretación de las agrupaciones de las muestras, revelando patrones y relaciones inherentes dentro del conjunto de datos.

Estimamos el numero optimo de clusters:

1º escalamos los datos:

```
{r}
cancer_data_escalado = scale(cancer_data_1)
```

2º Graficamos la suma de cuadrados dentro de los graficos:

```
{r}
fviz_nbclust(cancer_data_escalado, kmeans, nstart = 25, method = "wss") +
  geom_vline(xintercept = 3, linetype = 2)
```

Optimal number of clusters

(Top Level) Quarto Console

Environment History Connections Tutorial

Import Dataset 17 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12
<input type="checkbox"/> bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0
<input type="checkbox"/> bit64	A S3 Class for Vectors of 64bit Integers	4.6.0-1

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1* cancer_data_escalado cancer_data_1 conoc_actit_factor_cancer_cervical

Render on Save Render Run Publish

Source Visual B I </> Normal Format Insert Table Outline

```
cancer_data_escalado = scale(cancer_data_1)
```

2º Graficamos la suma de cuadrados dentro de los graficos:

```
{r}
fviz_nbclust(cancer_data_escalado, kmeans, nstart = 25, method = "wss") +
  geom_vline(xintercept = 3, linetype = 2)
```

Optimal number of clusters

(Top Level) Quarto Console

Se muestra la "Total Within Sum of Square" (WCSS) en función del número de clústeres (k). Se busca el

Environment History Connections Tutorial

Import Dataset 17 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12
<input type="checkbox"/> bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1* cancer_data_escalado cancer_data_1 conoc_actit_factor_cancer_cervical

Render on Save Render Run Publish

Source Visual B I </> Normal Format Insert Table Outline

Se muestra la "Total Within Sum of Square" (WCSS) en función del número de clústeres (k). Se busca el punto de inflexión, o "codo", donde la disminución marginal de WCSS se estanca. La línea punteada indica k=3 como el número óptimo, sugiriendo que más de tres clústeres no aportan una reducción significativa en la varianza intra-clúster.

Calculamos el agrupamiento de los k-means:

```
{r}
set.seed(123)
km_res <- kmeans(cancer_data_escalado, 3, nstart = 25)
```

```
{r}
km_res
```

K-means clustering with 3 clusters of sizes 163, 49, 6

Cluster means:

	edad	edad_relacion_sexual	parejas_sex	num_hijos
1	0.4735688	0.01298091	-0.1191356	0.3303609
2	-1.4535511	0.03552084	-0.1675481	-1.0626043
3	-0.9946168	-0.64273506	4.6048274	-0.2968691

Clustering_vector:

Environment History Connections Tutorial

Import Dataset 18 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_actit_factor_ca...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12
<input type="checkbox"/> bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
<input type="checkbox"/> bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

Untitled1 * cancer_data_escalado cancer_data_1 conoc_activ_factor_cancer_cervical

Source Visual B I <> Normal Format Insert Table Outline

Calculamos el agrupamiento de los k-means:

```
{r}
set.seed(123)
km_res <- kmeans(cancer_data_escalado, 3, nstart = 25)
```

```
{r}
km_res
```

K-means clustering with 3 clusters of sizes 163, 49, 6

Cluster means:

	edad	edad_relacion_sexual	parejas_sex	num_hijos
1	0.4735688	0.01298091	-0.1191356	0.3303609
2	-1.4535511	0.03552084	-0.1675481	-1.0626043
3	-0.9946168	-0.64273506	4.6048274	-0.2968691

Clustering vector:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
21	22	23	1	1	2	3	2	1	2	1	1	2	2	2	2	2	1	2	1	2	2	2
24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
3	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	2	1	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92
1	2	1	1	1	1	2	1	1	1	2	1	2	1	2	3	2	1	1	1	1	1	1
1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115

Chunk 14

Console

Environment History Connections Tutorial

Import Dataset 19 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_activ_factor_cancer...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
abind	Combine Multidimensional Arrays	1.4-8
askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
base64enc	Tools for base64 encoding	0.1-3
bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
BiasedUrn	Biased Urn Model Distributions	2.0.12
bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

Untitled1 * cancer_data_escalado cancer_data_1 conoc_activ_factor_cancer_cervical

Source Visual B I <> Normal Format Insert Table Outline

Visualizamos los clusters k-means

```
{r}
fviz_cluster(
  km_res,
  data = cancer_data_escalado,
  palette = c("#0064E7", "#E67E22", "#A52A2A"),
  ellipse.type = "euclid",
  repel = TRUE,
  ggtheme = theme_minimal()
)
```

Chunk 14

Console

Within cluster sum of squares by cluster:

[1]	331.35589	142.43927	41.05674
-----	-----------	-----------	----------

(between_SS / total_SS = 40.7 %)

Available components:

[1]	"cluster"	"centers"	"totss"	"withinss"	"tot.withinss"
[6]	"betweenss"	"size"	"iter"	"ifault"	

Environment History Connections Tutorial

Import Dataset 16 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_activ_factor_cancer...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
abind	Combine Multidimensional Arrays	1.4-8
askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
base64enc	Tools for base64 encoding	0.1-3
bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
BiasedUrn	Biased Urn Model Distributions	2.0.12
bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1
bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.6.0

Untitled1 * cancer_data_escalado cancer_data_1 conoc_activ_factor_cancer_cervical

Source Visual B I <> Normal Format Insert Table Outline

Visualizamos los clusters k-means

```
{r}
fviz_cluster(
  km_res,
  data = cancer_data_escalado,
  palette = c("#0064E7", "#E67E22", "#A52A2A"),
  ellipse.type = "euclid",
  repel = TRUE,
  ggtheme = theme_minimal()
)
```

Cluster plot

cluster

- 1
- 2
- 3

Environment History Connections Tutorial

Import Dataset 16 MiB

R Global Environment

Data

cancer_data	218 obs. of 18 variables
cancer_data_1	218 obs. of 4 variables
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...
conoc_activ_factor_cancer...	218 obs. of 18 variables
dist_link_cancer_data	List of 7
km_res	List of 9

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
User Library		
abind	Combine Multidimensional Arrays	1.4-8
askpass	Password Entry Utilities for R, Git, and SSH	1.2.1
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0
base64enc	Tools for base64 encoding	0.1-3
bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0
bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7
BiasedUrn	Biased Urn Model Distributions	2.0.12
bigD	Flexibly Format Dates and Times to a Given Locale	0.3.1

CONCLUSION:

Environment

History

Connections

Tutorial

Import Dataset

15 MiB

Global Environment

List

Data

cancer_data	218 obs. of 18 variables	
cancer_data_1	218 obs. of 4 variables	
cancer_data_escalado	num [1:218, 1:4] 0.672 0.76 -1.717 -1.806 -1.806 ...	
conoc_activ_factor_ca...	218 obs. of 18 variables	
dist_link_cancer_data	List of 7	
km_res	List of 9	

Values

dist_cancer_data	'dist' num [1:23653] 2.23 3.51 6.17 4.63 2 ...
------------------	--

Files

Plots

Packages

Help

Viewer

Presentation

Install

Update

Name

Description

Version

User Library

<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-8	
<input type="checkbox"/> askpass	Password Entry Utilities for R, Git, and SSH	1.2.1	
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.5.0	
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3	
<input type="checkbox"/> bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.16.0	
<input type="checkbox"/> bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-7	
<input type="checkbox"/> BiasedUrn	Biased Urn Model Distributions	2.0.12	