Documentation for Family Code Update Script

Overview

This script updates family codes based on relationships (mother and father names) using graph theory. It creates a graph where individuals are vertices and edges represent shared parent names. The script then finds connected components in the graph and assigns family codes accordingly.

Requirements

R Version

R version 4.3.1 (2023-06-16 ucrt)

Necessary Packages

igraph: This package is used for creating and manipulating graphs.

To install the igraph package, run the following command in your R console:

```
install.packages("igraph")
```

Loading the Package

After installing the package, load it using:

```
library(igraph)
```

Script Description

Function: update_family_codes

This function takes a data frame with columns Family_code, Mother_name, and Father_name, and updates the Family_code based on shared parent names.

Parameters:

data: A data frame containing the columns Family_code, Mother_name, and Father_name.

Returns:

A data frame with updated Family_code values.

Code:

```
update_family_codes <- function(data) {
  # Create a graph
  g <- graph.empty(directed = FALSE)
  # Add vertices (individuals)
  g <- add_vertices(g, nrow(data), name = 1:nrow(data))

  # Add edges based on parent relationships
  for (i in 1:nrow(data)) {
    for (j in 1:nrow(data)) {
      if (i != j && (data$Mother_name[i] == data$Mother_name[j] || data$Father_name[i] == data$Father_name[j])) {
        g <- add_edges(g, c(i, j))
      }
    }
  }

  # Find connected components
  components <- clusters(g)$membership

  # Assign family codes based on connected components
  for (i in 1:length(components)) {
    data$Family_code[i] <- paste0("n", components[i])
  }

  return(data)
}
```

Example Usage

The following examples demonstrate how to use the update_family_codes function with different datasets.

Example 1

```
data1 <- data.frame(
  Family_code = c("n1", "n2", "n3"),
  Mother_name = c("Laura", "Katty", "Katty"),
  Father_name = c("Mourinho", "Ferguson", "Mourinho")
)

updated_data1 <- update_family_codes(data1)
print(updated_data1)
```

Example 2

```
data2 <- data.frame(
  Family_code = c("n1", "n2", "n3", "n4"),
  Mother_name = c("Laura", "Katty", "Laura", "Katty"),
  Father_name = c("Mourinho", "Ferguson", "Ferguson", "Mourinho")
)

updated_data2 <- update_family_codes(data2)
print(updated_data2)
```

Example 3

```
data3 <- data.frame(
  Family_code = c("n1", "n2", "n3", "n4", "n5"),
  Mother_name = c("Laura", "Katty", "Laura", "Katty", "Laura"),
  Father_name = c("Mourinho", "Ferguson", "Ferguson", "Mourinho", "Mourinho")
)

updated_data3 <- update_family_codes(data3)
print(updated_data3)
```

Example 4

```
data4 <- data.frame(
  Family_code = c("n1", "n2", "n3", "n4", "n5"),
  Mother_name = c("Laura", "Katty", "Isabelle", "Carole", "Sandra"),
  Father_name = c("Mourinho", "Ferguson", "Guardiola", "Ancelotti", "Simeone")
)

updated_data4 <- update_family_codes(data4)
print(updated_data4)
```

Example 5

```
data5 <- data.frame(
  Family_code = c("n1", "n2", "n3", "n4", "n5"),
  Mother_name = c("Laura", "Katty", "Isabelle", "Carole", "Sandra"),
  Father_name = c("Mourinho", "Ferguson", "Guardiola", "Ancelotti", "Simeone"),
  Sister_name = c("Anna", "Beth", "Clara", "Diana", "Eva")
)

updated_data5 <- update_family_codes(data5)
print(updated_data5)
```

Example 6

```
data6 <- data.frame(
  Family_code = c("n1", "n2", "n3", "n4", "n5"),
  Mother_name = c("Laura", "Katty", "Isabelle", "Carole", "Katty"),
  Father_name = c("Mourinho", "Mourinho", "Guardiola", "Ancelotti", "Simeone"),
  Sister_name = c("Anna", "Anna", "Clara", "Diana", "Eva")
)

updated_data6 <- update_family_codes(data6)
print(updated_data6)
```

Notes

Ensure that the data frame passed to the update_family_codes function contains the necessary columns (Family_code, Mother_name, and Father_name).

The function assumes that individuals with the same mother or father are part of the same family.