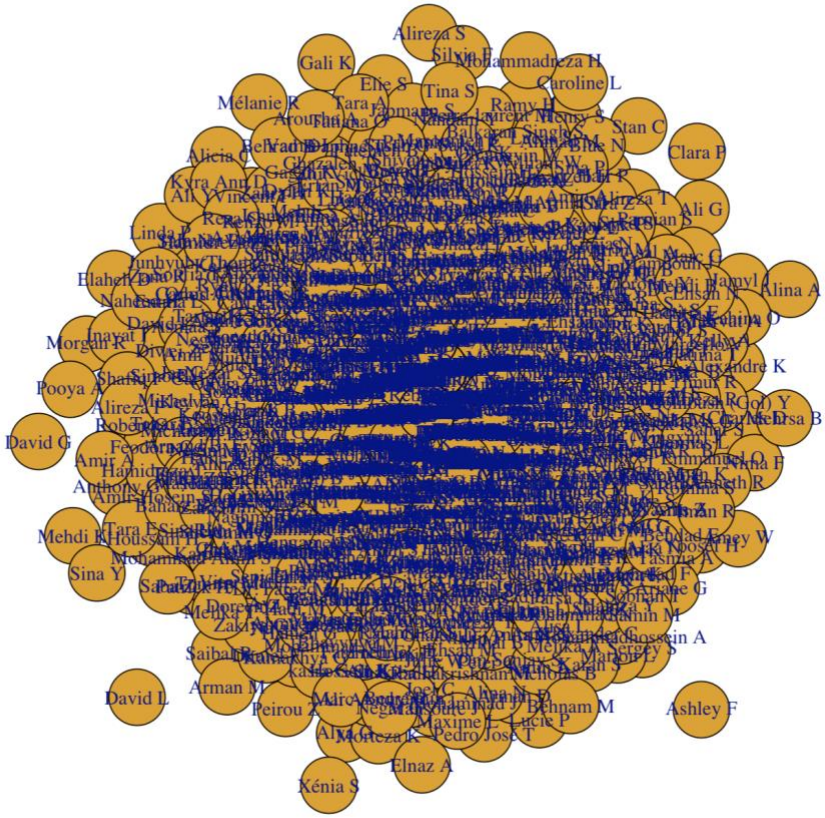
Assignment 1

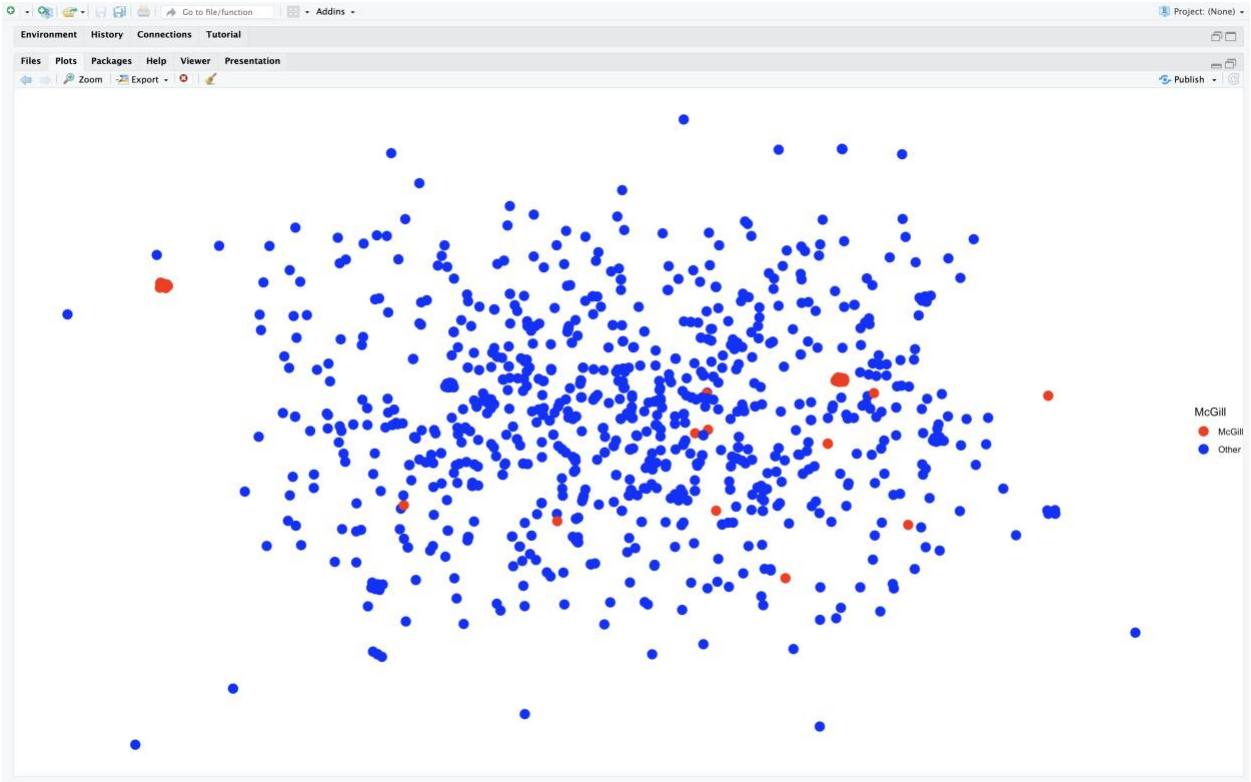
Niki_Mahmoodzadeh_Assignment1.R | Connections.csv

Source on Save

```r
# Load required libraries
library(dplyr)
library(stringr)
library(tidyr)
library(igraph)
library(ggplot2)
library(ggraph)
library(purrr)

# Read the LinkedIn connections CSV file
Connections <- read.csv("/Users/nikimahmoodzadeh/Downloads/Connections.csv", sep = ",", skip = 3)

# Count connections by company, sorted by count
number_of_connections <- Connections %>%
  group_by(Company) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count))

# Print the number of connections by company
print(number_of_connections)

# Total number of connections
total_num <- nrow(Connections)

# Print the total number of connections
print(total_num)

# Rename columns to remove spaces
Connections <- Connections %>%
  rename(FirstName = `First.Name`, LastName = `Last.Name`)

# Create labels for nodes using the first name and the first letter of the last name
Connections$Label <- with(Connections, paste(FirstName, substr(LastName, 1, 1)))

# Assign unique IDs to each connection
Connections <- Connections %>%
  mutate(ID = row_number())

# Create nodes dataframe using ID, Label, and Company
nodes <- Connections %>%
  distinct(ID, Label, Company)

# Join the IDs back to the original data
linkedin_data_with_ids <- Connections %>%
  left_join(nodes, by = c("Label", "Company"))
```

```r

# Create edges based on IDs within the same company
edges <- nodes %>%
  group_by(Company) %>%
  filter(n() > 1) %>%
  summarise(Combo = list(combn(ID, 2, simplify = FALSE))) %>%
  unnest(Combo) %>%
  ungroup() %>%
  mutate(From = sapply(Combo, `[`, 1),
         To = sapply(Combo, `[`, 2)) %>%
  select(From, To)

# View the edges dataframe
print(edges)

# Create graph from edges dataframe, using the updated nodes and labels
g <- graph_from_data_frame(d = edges, vertices = nodes, directed = TRUE)

# Plot the graph
plot(g, vertex.label = V(g)$Label)

# Create a new column 'McGill' to identify contacts affiliated with McGill
nodes <- nodes %>%
  mutate(McGill = ifelse(str_detect(Company, "McGill"), "McGill", "Other"))

nodes$ID <- as.character(nodes$ID)

# Generate layout
g_M <- graph_from_data_frame(d = edges, vertices = nodes, directed = FALSE)
layout <- as.data.frame(layout_with_fr(g_M))
names(layout) <- c("x", "y")
layout$ID <- V(g_M)$name

# Add McGill information to the layout
layout <- layout %>%
  left_join(nodes %>% select(ID, McGill), by = "ID")

edges$From <- as.character(edges$From)
edges$To <- as.character(edges$To)

# Join edge start positions
edges_coords <- edges %>%
  left_join(layout %>% select(ID, x_start = x, y_start = y), by = c("From" = "ID"))

# Join edge end positions
edges_coords <- edges_coords %>%
  left_join(layout %>% select(ID, x_end = x, y_end = y), by = c("To" = "ID"))

# Plotting the network with McGill connections highlighted
ggplot() +
  geom_segment(data = edges_coords, aes(x = x_start, y = y_start, xend = x_end, yend = y_end), color = "gray50") +
  geom_point(data = layout, aes(x = x, y = y, color = McGill), size = 4) +
  scale_color_manual(values = c("McGill" = "red", "Other" = "blue")) +
  theme_void() +
  theme(legend.position = "right") +
  labs(color = "McGill")
```


Alireza S
Gali K
Elie S
Tina S
Caroline L
Mélanie R
Stan C
Clara P
Ali G
Alina A
Pooya A
David G
Sina Y
David L
Arman M
Peirou Z
Behnam M
Ashley F
Maxime F
Lucie P
Pedro José T
Elnaz A
Kénia S


Go to file/function
Addins
Project: (None)
Environment
History
Connections
Tutorial
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
McGill
McGill
Other