

Exercise 1

2024-03-14

```
# Load the necessary libraries  
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
#install.packages("tidygraph")  
#install.packages("tibble")  
library(tibble)
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
library(tidygraph)
```

```
## Warning: package 'tidygraph' was built under R version 4.2.3
```

```
##  
## Attaching package: 'tidygraph'
```

```
## The following object is masked from 'package:stats':  
##  
##   filter
```

```
library(igraph)
```

```
## Warning: package 'igraph' was built under R version 4.2.3
```

```
##  
## Attaching package: 'igraph'
```

```
## The following object is masked from 'package:tidygraph':  
##  
## groups
```

```
## The following object is masked from 'package:tibble':  
##  
## as_data_frame
```

```
## The following objects are masked from 'package:dplyr':  
##  
## as_data_frame, groups, union
```

```
## The following objects are masked from 'package:stats':  
##  
## decompose, spectrum
```

```
## The following object is masked from 'package:base':  
##  
## union
```

```
#install.packages("ggraph")  
library(ggraph)
```

```
## Warning: package 'ggraph' was built under R version 4.2.3
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(purrr)
```

```
## Warning: package 'purrr' was built under R version 4.2.3
```

```
##  
## Attaching package: 'purrr'
```

```
## The following objects are masked from 'package:igraph':  
##  
## compose, simplify
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
##  
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:igraph':  
##  
## crossing
```

```
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 4.2.3
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.2.3
```

```
Connections <- read_csv("D:\\Google Drive\\McGill\\Winter Semester\\W2\\Talent-Analytics-Assignments\\P
```

```
## Rows: 387 Columns: 7
## -- Column specification -----
## Delimiter: ","
## chr (7): First Name, Last Name, URL, Email Address, Company, Position, Conne...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
Connections <- Connections[!is.na(Connections$Company), ]
# Group by 'Company' and count the number of contacts per company
company_counts <- Connections %>%
  group_by(Company) %>%
  summarise(Count = n())

# Calculate the total count of contacts
total_count <- nrow(Connections)

# Print the count of contacts by company and the total count
print(paste("Total count of contacts:", total_count))
```

```
## [1] "Total count of contacts: 377"
```

```
# Viewing the dataset in descending order
company_counts <- arrange(company_counts, desc(Count))
head(company_counts, 10)
```

```
## # A tibble: 10 x 2
##   Company                                Count
##   <chr>                                <int>
## 1 McGill University - Desautels Faculty of Management      18
## 2 McGill University                                         10
## 3 Orta Doğu Teknik Üniversitesi / Middle East Technical University    8
## 4 Arçelik Global                                             7
## 5 CN                                                         5
## 6 Environics Analytics                                       5
## 7 BRP                                                         4
## 8 Baykar Technologies                                       4
## 9 Bell                                                       4
## 10 CAE                                                        4
```

```

# Keep only the first name and the first letter of the last name as a label for the nodes
Connections <- Connections %>%
  mutate(label = paste(`First Name`, substr(`Last Name`, 1, 1)))
Connections <- Connections %>%
  mutate(title = paste(`First Name`, substr(`Last Name`, 1, 1)))

# Create edges dataframe
edges <- Connections %>%
  select(label, Company) %>%
  distinct() %>%
  group_by(Company) %>%
  filter(n() > 1) %>%
  summarise(pair = list(combn(label, 2, simplify = FALSE))) %>%
  unnest(pair) %>%
  mutate(from = map_chr(pair, 1), to = map_chr(pair, 2)) %>%
  select(from, to, Company) # Add Company column to the edges dataframe

# Create the graph
graph_connections <- as_tbl_graph(edges)

# Add title, label, and Company attributes to the nodes
graph_connections <- graph_connections %>%
  activate(nodes) %>%
  mutate(
    title = str_to_title(name),
    label = str_replace_all(title, " ", "\n"),
  )

# Visualize the graph
g <- ggraph(graph_connections, layout = "kk") +
  geom_edge_link(aes(edge_alpha = 0.1, color = Company)) + # Color edges by Company
  geom_node_text(aes(label = label), size = 3) +
  scale_color_viridis_d() +
  theme_void() +
  theme(legend.position = "none") +
  geom_node_point(color = "blue", size = 1)

# Print the graph
print(g)

```

