Exercise1

Read the data

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
## Warning: package 'lubridate' was built under R version 4.3.1
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
df <- read_csv("Connections.csv")
```

```
## Rows: 490 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.
```

Get the count of contancts by their current employer + total count

```
count_by_emp <- df %>%
  group_by(Company) %>%
  summarise(count = n()) %>%
  arrange(desc(count))

# total count
total_count <- df %>%
  summarise(count = n())

print(count_by_emp)
```

```
## # A tibble: 348 x 2
##    Company                                                count
##    <chr>                                                  <int>
##  1 <NA>                                                      24
##  2 McGill University - Desautels Faculty of Management      20
##  3 McGill University                                        14
##  4 Sophia University                                         8
##  5 Deloitte                                                  7
##  6 CN                                                        6
##  7 L'Oréal                                                   6
##  8 Amazon                                                    4
##  9 BOMBARDIER                                                4
## 10 BRP                                                       4
## # i 338 more rows
```

```
print(total_count)
```

```
## # A tibble: 1 x 1
##   count
##   <int>
## 1   490
```

Create nodes and edges dataframe to use with igraph (use tidygraph)

```
library(tidygraph)
```

```
## Warning: package 'tidygraph' was built under R version 4.3.1
```

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

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

```
library(igraph)
```

```
## Warning: package 'igraph' was built under R version 4.3.1
```

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

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

```
## The following objects are masked from 'package:lubridate':
##
##     %--%, union
```

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

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

## The following object is masked from 'package:tidyr':
##
##     crossing

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

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

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

```
# rename columns
df <- df %>%
  rename(First = `First Name`)
df <- df %>%
  rename(Last = `Last Name`)

# create label as a combination of first and first letter of last name
df <- df %>%
  mutate(label = paste(First, substr(Last, 1, 1), sep = " "))
```

If we color "McGill" differently...

```
library(ggraph)
```

```
## Warning: package 'ggraph' was built under R version 4.3.1
```

```
df <- df %>%
        mutate(color = case_when(
                Company == "McGill University" ~ "red",
                Company == "McGill University - Desautels Faculty of Management" ~ "green",
                TRUE ~ "blue"
        ))

nodes <- df %>%
        select(name = label, color)

edges <- expand.grid(from = df$label, to = df$label) %>%
```

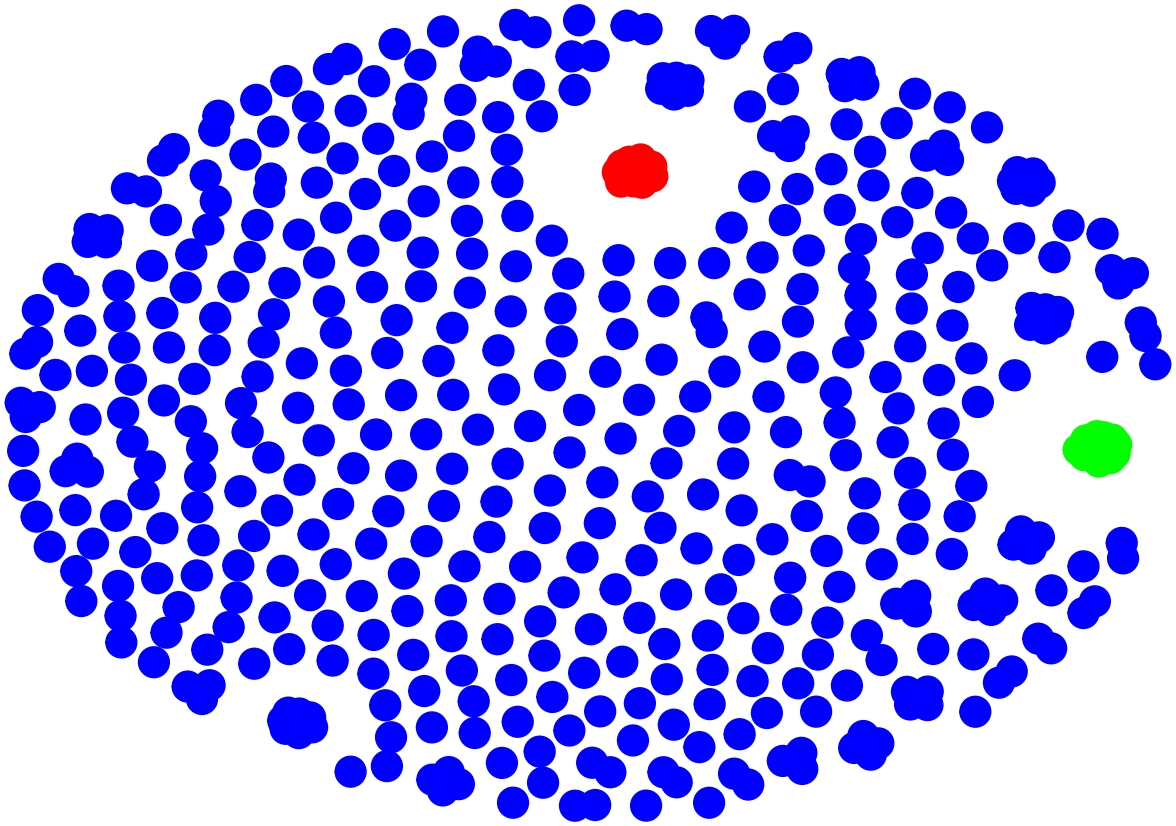
```
        left_join(df, by = c("from" = "label")) %>%
        rename(from_company = Company) %>%
        left_join(df, by = c("to" = "label")) %>%
        rename(to_company = Company) %>%
        filter(from_company == to_company, from != to) %>%
        select(from, to) %>%
        distinct()
```

```
## Warning in left_join(., df, by = c(from = "label")): Detected an unexpected many-to-many relationshi
## i Row 14 of `x` matches multiple rows in `y`.
## i Row 79 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
## Warning in left_join(., df, by = c(to = "label")): Detected an unexpected many-to-many relationship
## i Row 6449 of `x` matches multiple rows in `y`.
## i Row 1 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
# Use the modified nodes dataframe to create the tbl_graph
graph <- tbl_graph(nodes = nodes, edges = edges, directed = FALSE)

# Plot with color differentiation
ggraph(graph, layout = "fr") +
        geom_edge_link() +
        geom_node_point(aes(color = color), size = 5) +
        geom_node_text(aes(label = name, filter = name == "McGill University"), repel = TRUE) +
        scale_color_identity() +
        theme_void()
```



What do we see in the network?

- The network is very dense for McGill and Desautels, with many connections between people. (I did not manually color MMA cohorts but just chose to filter by the McGill name)
- If "Company" columns could contain historical data such as where we have worked, the plot would show much more interesting interactions between us.
- At the same time we have many single node without any connections. This is expected as we receive many random invitations from people we don't know on LinkedIn.