

# Exercise 1

Utkarsh Pal

2023-03-14

## Loading the LinkedIn Data

```
library("tidyverse")
```

```
## Warning: package 'tidyverse' was built under R version 4.2.2
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
```

```
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
## Warning: package 'readr' was built under R version 4.2.2
```

```
## Warning: package 'purrr' was built under R version 4.2.2
```

```
## Warning: package 'dplyr' was built under R version 4.2.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library("igraph")
```

```
## Warning: package 'igraph' was built under R version 4.2.2
```

```
##
```

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

```
##
```

```
## 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
```

```
LinkedIn = read.csv("C:\\Users\\Utkarsh\\Desktop\\Connections.csv")

head(LinkedIn,10)
```

```
##      First.Name  Last.Name Email.Address
## 1      Emma Soukiassian
## 2      Tarek   Cheaito
## 3      Anojan  Gnanendran
## 4      Xénia   Sozonoff
## 5      Lucas   Larbodiére
## 6      Thomas  McNulty
## 7      Patricia Betts
## 8      Liliana Tretyakova
## 9      Michael  Murphy
## 10 Ying-Fang (Ylfa) Liang
```

```
##
## 1      Compar
## 2      Pratt & Whitney Canada
## 3      L'Oréal
## 4      Canadian Tire Corporation
## 5      CAE Inc
## 6      Lud
## 7      Societe Generale Corporate and Investment Banking - SGC
## 8      Memory and History: Transforming the Narrative of the Spanish Civil War and Francoist Dictatorship
## 9      Kruger Inc
## 10     McGill University - Desautels Faculty of Management
##      Pratt & Whitney Canada
##      Position
## 1      Data analyst, Project manager- Academic Consulting Project
## 2      Business Strategist & Data Scientist Team Lead (Academic Consulting Project)
## 3      Associate Brand Manager
## 4      Analytics Academic Consulting Project, Data Analyst
## 5      Associate Product Manager
## 6      Spécialiste principal en attraction de talents / Principal Onsite Sourcing Specialist (via LevelUp)
## 7      Associated Researcher and Translator
## 8      Lead UI/UX Designer and Project Manager - Academic Consulting Project
## 9      Masters in Analytics Studies
```

```
## 10
##   Connected.On
## 1   09-Mar-23
## 2   07-Mar-23
## 3   01-Mar-23
## 4   28-Feb-23
## 5   21-Feb-23
## 6   19-Feb-23
## 7   17-Feb-23
## 8   08-Feb-23
## 9   05-Feb-23
## 10  01-Feb-23
```

```
attach(LinkedIn)
```

```
## The following object is masked from package:ggplot2:
##
##   Position
```

## Pre-Processing the Data and Filtering LinkedIn Connections

```
LinkedIn$Name <- paste(LinkedIn$First.Name, substr(LinkedIn$Last.Name, 1, 1), sep = " ")

count_of_contacts <- LinkedIn %>%
  group_by(Company) %>%
  summarize(count = n()) %>%
  arrange(desc(count))

head(count_of_contacts, 10)
```

```
## # A tibble: 10 x 2
##   Company                count
##   <chr>                  <int>
## 1 KPI Digital Solutions     8
## 2 University of Waterloo    7
## 3 TD                        4
## 4 Pratt & Whitney Canada    3
## 5 Bank of Canada            2
## 6 BRP                       2
## 7 Deloitte Canada           2
## 8 FutureFit AI              2
## 9 Primerica                 2
## 10 Sun Life                 2
```

```
# Filtering my contacts so only the connections that have a company in common remain
Connections <- LinkedIn %>%
  inner_join(count_of_contacts, by = "Company") %>%
  filter(count >= 2) %>%
  select(Name, Company)
```

## Total Count of My Connections

```
total_count <- nrow(LinkedIn)
total_count
```

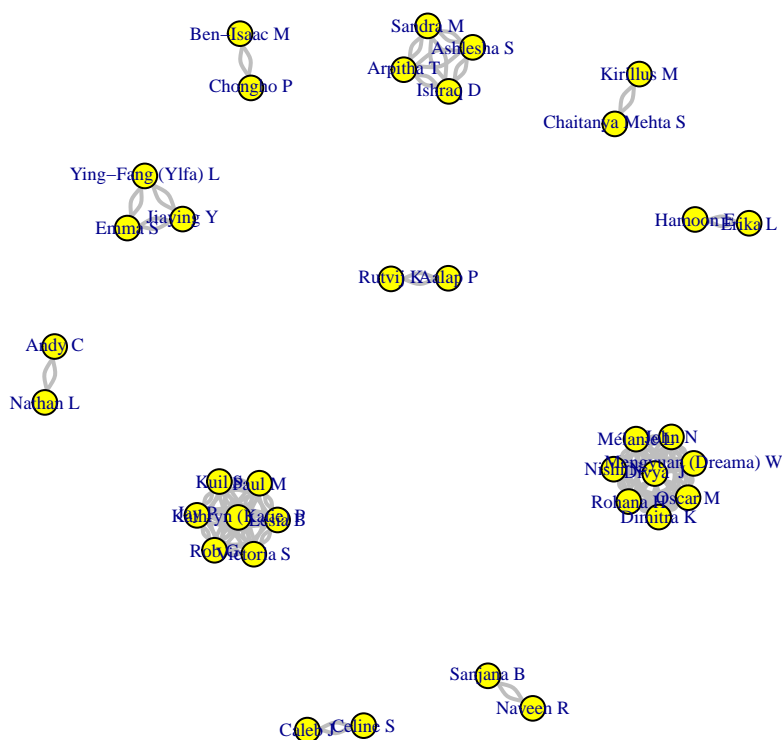
```
## [1] 139
```

## Creating Nodes and Edges for My Connections

```
# Create nodes using tidygraph
Nodes <- Connections %>%
  mutate(label = Name) %>%
  distinct(label) %>%
  as_tibble() %>%
  select(label)
# Create edges using tidygraph
Edges <- Connections %>%
  left_join(Connections, by = "Company") %>%
  filter(Name.x != Name.y) %>%
  mutate(from = Name.x,
         to = Name.y) %>%
  select(from, to)
```

## Creating a Graph using of My Connections

```
Connections_Graph <- graph_from_data_frame(Edges, vertices = Nodes, directed = FALSE)
par(mar = rep(1, 4))
options(repr.plot.width = 10, repr.plot.height = 10)
plot(Connections_Graph, vertex.size = 7, vertex.color = "yellow", vertex.label.cex = 0.6, edge.color = "black")
```



From this graph, I realized that most of my connections that have a common company between them are the ones I formed here at McGill or the ones from my undergraduate university (University of Waterloo). Since we are currently working on the academic consulting project, a large portion of the connections are the capstone groups. The connections from my undergraduate university are my former professors or my former classmates that are pursuing further education. This makes sense for someone like me, who does not have much work experience to have a lot of connections with multiple people from the same company.