

# ETC1010-5510 Project Milestone 3- T11\_Wed\_skimr

## A Report On The LinkedIn-World Bank Data for Development

Hanchen Wang 30704456, Hao Li 32041594, Jiaying Zhang 30930685, Mohammed Faizan 31939872, Karan Garg 32106580

24th May 2021

**TUTOR:** Samantha Dawson

**LECTURER:** Patricia Menéndez

## 1 Introduction: Project and Data Description

### 1.1 The LinkedIn-World Bank Digital Data for Development: Industry Jobs and Skills Trends

"The World Bank Group and LinkedIn have created the Digital Data for Development collaboration to support innovative policy decisions as developing countries grapple with a rapidly changing global economy. With hundreds of millions of members worldwide, LinkedIn has the potential to offer a new, timely, and granular source of data about emerging industries, workers' changing skills composition and how they're engaging with labor markets globally." ~ 1 (<https://linkedindata.worldbank.org/about>)

This collaboration enables government and policy makers to drive better policy implementations, thus creating opportunities to the global work force. The data represents LinkedIn members' data based on four metrics: Industry Employment Shifts, Talent Migration, Industry Skills Needs and Skills Penetration. The records in the data represent over 100 countries having a distribution across six major industry sectors(represents 148 industries): Financial Services, Professional Services, Information & Communication Technology (ICT), the Arts & Creative Industries, Manufacturing, and Mining/Quarrying and possessing skills within the over 50,000 distinct, standardized skills classified by LinkedIn into 249 skill groups, further categorised as: Business Skills, Disruptive Tech Skills, Soft Skills, Specialized Industry Skills and Tech Skills.

The data is referenced below in Section 4.1.

For our project, we will be working on the following questions.

1. Which skill category is most common across all Industry Sections and how does it vary between each section? (Hao Li)
2. What is the average percentage of net migration for each industry over the past five years and Is the growth rate of immigration within the industry related to the growth rate of the industry? (Jiaying Zhang)
3. For each common skill\_category, which industry has the highest penetration rate and what is the change of the common skill penetration rate over the period of time? (Hanchen Wang)
4. Find the industry\_section that is best to each region/continent. (Faizan and Karan)
5. Which industry\_name sees the maximum growth over time in each region/continent, depending on the region's best industry\_section ? (Mohammed Faizan)
6. For each region, which country did the above found industry had had maximum growth? And, what is the income group of that nation? (Karan Garg)

## 2 QUESTIONS

### 2.1 Which skill category is most common across all Industry Sections and how does it vary between each section?

"The Industry Skills Needs metric captures which skills are most likely to be added to a member's profile in one industry compared to other industries. It's calculated using an adapted version of a text mining technique called Term Frequency - Inverse Document Frequency (TF-IDF). This method gives more weight to a skill for an industry if more members in the industry list the skill on their profiles and the skill is more unique to the industry. The skills included are those added while a member holds a particular occupation (i.e. the skill flow approach). While the skill flow approach creates a trade-off whereby long-held basic skills, such as Microsoft Office being given a lesser weight, the approach is shown to be stronger at identifying the latest emerging skills in a specific industry than including all historical skills that are added during prior occupations. On balance, since the objective of this metric is to detect the latest skills needs, a skill flow approach is adopted." ~2 (<https://datacatalog.worldbank.org/dataset/skills-linkedin-data>)

*The most common skill category across different sections is reported as Business Skills.*

Table 2.1: Skills and Industry Section

| skill_group_category        | Arts, entertainment and recreation | Financial and insurance activities | Information and communication | Manufacturing | Mining and quarrying | Professional scientific and technical activities |
|-----------------------------|------------------------------------|------------------------------------|-------------------------------|---------------|----------------------|--|
| Specialized Industry Skills | 266                                | 5                                  | 185                           | 228           | 39                   | 387  |
| Tech Skills                 | 118                                | 26                                 | 307                           | 88            | 10                   | 205  |
| Soft Skills                 | 83                                 | 82                                 | 104                           | 151           | 25                   | 202  |
| Business Skills             | 32                                 | 184                                | 138                           | 215           | 26                   | 273  |

| skill_group_category   | Arts, entertainment and recreation | Financial and insurance activities | Information and communication | Manufacturing | Mining and quarrying | Professional scientific and technical activities |
|------------------------|------------------------------------|------------------------------------|-------------------------------|---------------|----------------------|--|
| Disruptive Tech Skills | 1                                  | 3                                  | 66                            | 18            | NA                   | 33   |

Table 2.1: Top 1 skill category in every industry section

| isic_section_name                                | skill_group_category        | n   |
|--|-----------------------------|-----|
| Arts, entertainment and recreation               | Specialized Industry Skills | 266 |
| Financial and insurance activities               | Business Skills             | 184 |
| Information and communication                    | Tech Skills                 | 307 |
| Manufacturing                                    | Specialized Industry Skills | 228 |
| Mining and quarrying                             | Specialized Industry Skills | 39  |
| Professional scientific and technical activities | Specialized Industry Skills | 387 |

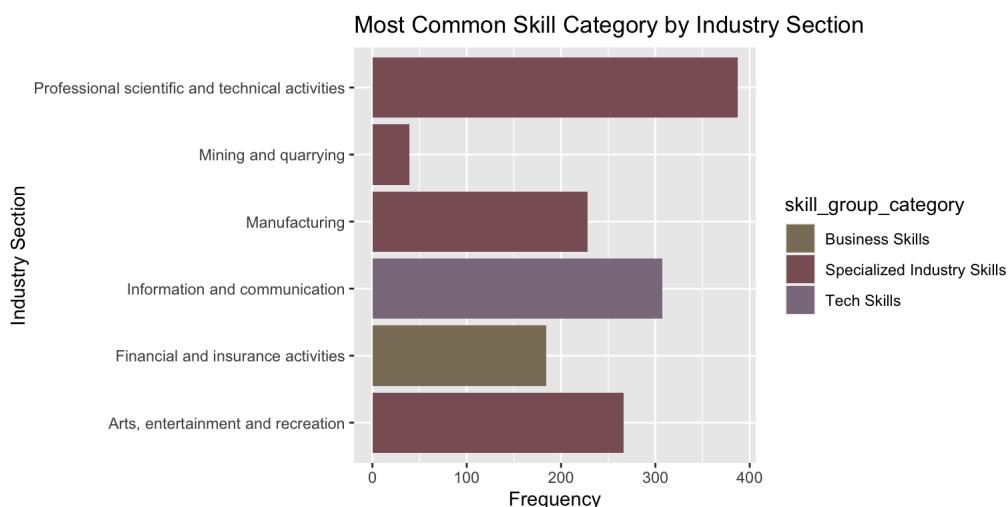


Figure 2.1: count of different skills

The TF-IDF text mining technique used in calculation for the skill rank includes the skills that are more unique to one industry than any other. For this reason, for each industry section, Specialized Industry Skills group count is the highest. Specialized Industry Skills are the most common skill category in 4 out of 6 industry sections. However, Information and Communication, has more of Tech Skills, this is owing to the fact that most of the Tech Skills like Microsoft Office are basic across all industries and hence are not categorized as Specialized Industry Skills for ICT. Business Skills(e.g.,) for financial and insurance activities is reasoned in the same manner as above.

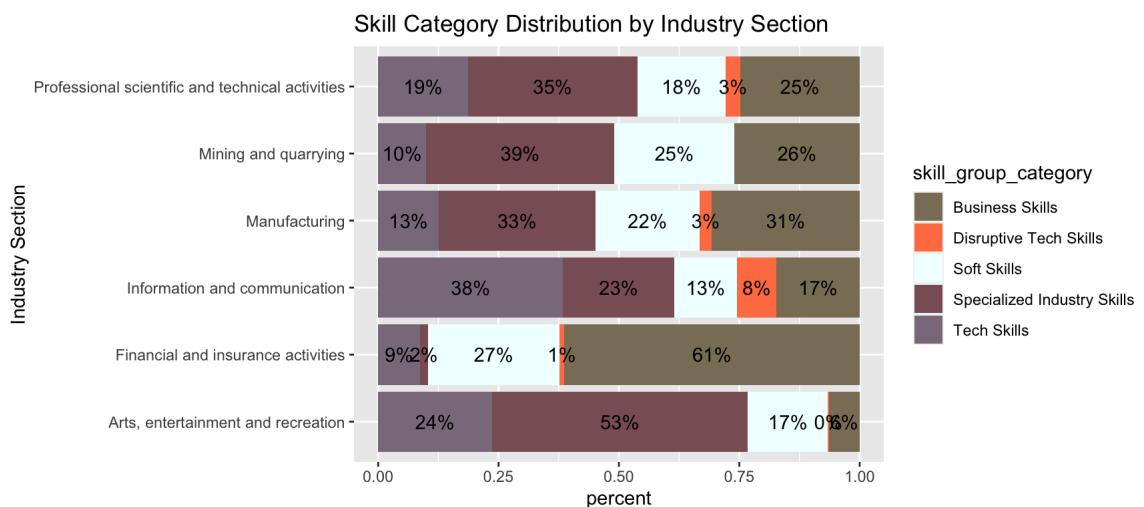


Figure 2.2: percentage of different skills

The above 100% stacked bar chart 2.2 shows the skill category distribution within each industry section. While 4 out of 6 industry sections have a similar skill category distribution, Financial & Insurance Activities and Arts, Entertainment & Recreation have a rather different skill category distribution. This is because Arts, Entertainment & Recreation is a field in which each talent is a skill and thus Specialized Industry Skills(53%)!!! Financial & Insurance Activities commands Soft Skills and Business Skills(61%). Disruptive Tech Skills are not possessed by members belonging to Mining and quarrying.

Overall, specialized Industry Skills are the most common skill in professional scientific and technical activities. While business skills are the most important for people to acquire in financial and insurance activities.

**2.2 What is the average percentage of net migration for each industry over the past five years and Is the growth rate of immigration within the industry related to the growth rate of the industry?**

<sup>3</sup>All the Talent Migration metrics are based on net migration (arrivals minus departures). (https://datacatalog.worldbank.org/dataset/talent-migration-linkedin-data)

### **2.2.1 Average percentage of net migration for each industry Section and industry over the past five years**

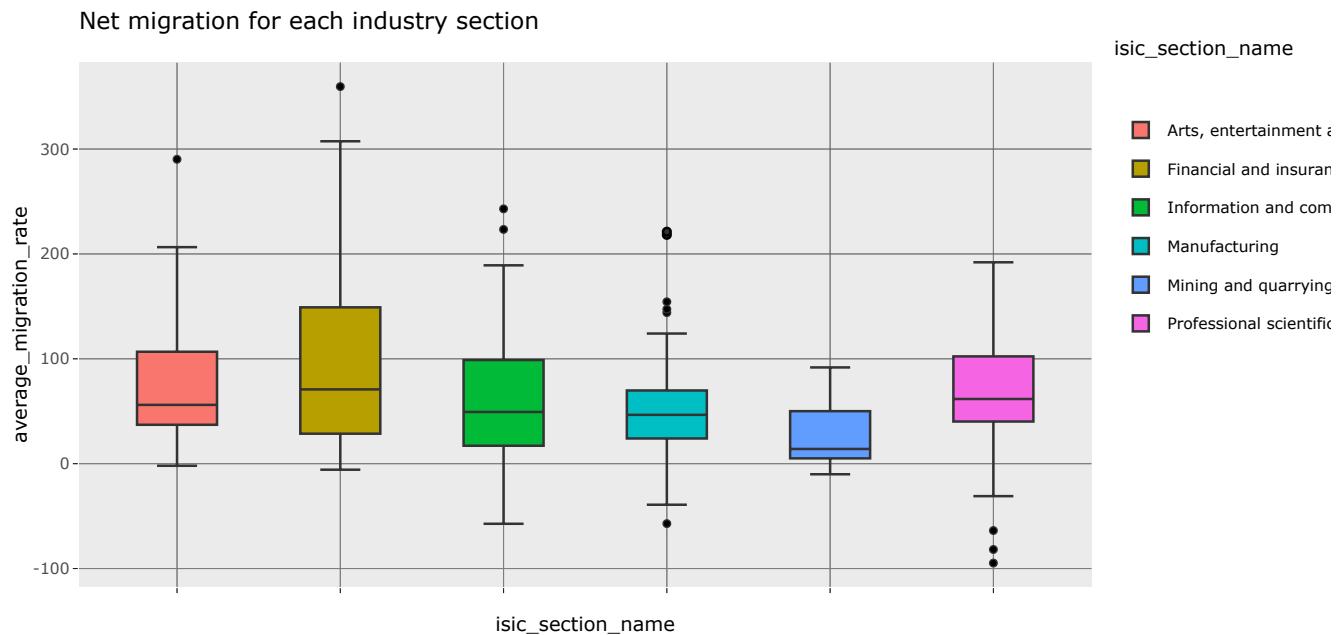


Figure 2.3: Net migration for each industry section

The figure 2.3 shows average percentage of net migration for each industry section:

- There are some differences in each industry, but the differences are not large and all the industry sections' median net migration rate is positive.
  - If we see the median, the growth of financial and insurance activities showed the highest growth, while mining and quarrying industry showed the slowest growth.

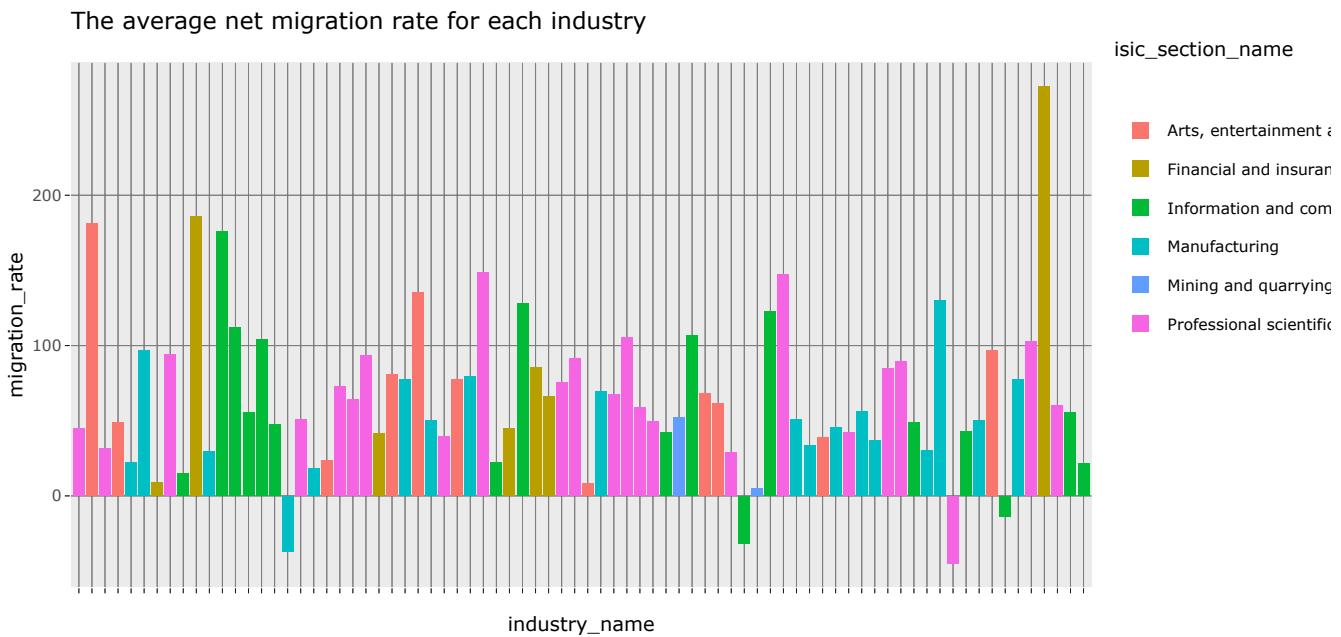


Figure 2.4: The industry average net migration rate for each industry section

The figure 2.4 shows Average percentage of net migration for each industry

- The two highest net migration rates also exist in financial and insurance activities: Venture Capital & Private Equity (about 272) and Capital Markets (about 186).

- The third place comes from the animation industry of Arts, Entertainment and Recreation. The average annual net migration rate reaches about 181.
- The net migration rate of most industries is positive, while the net migration rate of a few industries, such as Research and Defence and Space, is negative.

## 2.2.2 Is the growth rate of immigration within the industry related to the growth rate of the industry?

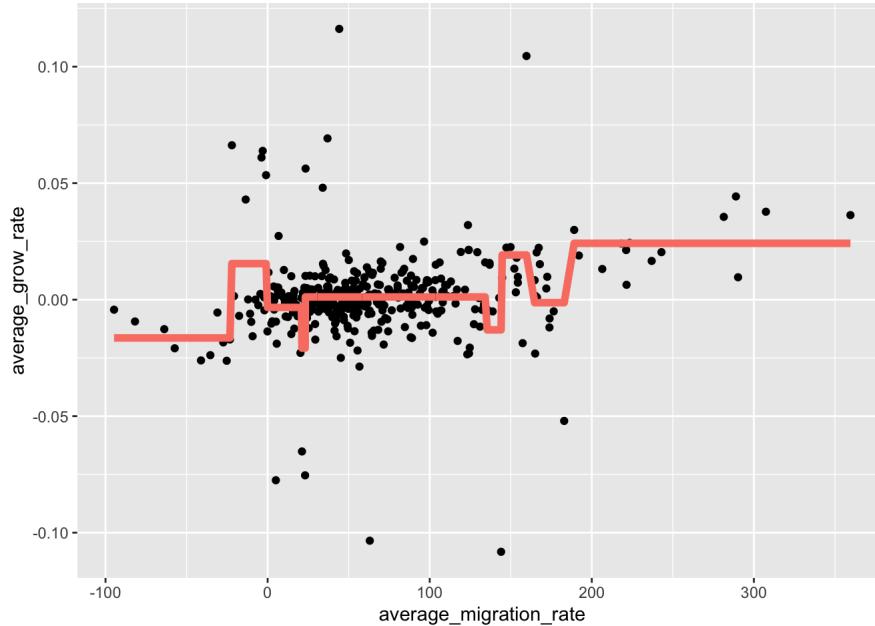


Figure 2.5: Relationship between migration growth and industry growth

The figure 2.5 shows:

- There is no obvious linear relationship between industry growth rate and migration rate.
- Although the migration rate of most industries has increased, the industry growth rate in most industries has not changed.
- However, we can see that the industry growth rate also increases with the large increase in migration rate.

## 2.2.3 Inference:

- Among all the industry sections, Net migration of financial and insurance activities industry section is the highest. The average net migration is positive for all industry sections.
- In the Financial and insurance activities industry section, the Financial and insurance activities industry also has the highest net migration rate among all industries. Net migration rate is positive in most industries.
- In most of the industry, industry growth rate and migration rate, there is no obvious linear relationship. In most of the industries, the growth of migration rate does not significantly promote the growth of the industry.

## 2.3 For each common skill\_category, which industry has the highest penetration rate and what is the change of the common skill penetration rate over the period of time?

"The Skill Penetration metric looks at how many skills from each of LinkedIn's skill groups (see "Notes" tab) appear among the top 30 skills for each occupation in an industry. For example, if 3 of 30 skills for Data Scientists in the Information Services industry fall into the Artificial Intelligence skill group, Artificial Intelligence has a 10% penetration for Data Scientists in Information Services. These penetration rates are averaged across occupations to derive the industry averages reported. It is likely this metric is best at capturing skill penetration across tradable and knowledge-intensive sectors. For example, it may under-estimate the adoption of AI in Manufacturing, since LinkedIn members are less likely to be in this sector compared to others." 2 (<https://datacatalog.worldbank.org/dataset/skills-linkedin-data>)

### 2.3.1 The highest penetration rate for different industry

Table 2.2: Top Industry by Penetration Rate for Each Skill Category

|                             | skill_group_name  | isic_section_name                                | industry_name     |
|-----------------------------|-------------------|--|-------------------|
| Specialized Industry Skills | Music             | Arts, entertainment and recreation               | Music             |
| Tech Skills                 | Graphic Design    | Professional scientific and technical activities | Graphic Design    |
| Business Skills             | Insurance         | Financial and insurance activities               | Insurance         |
| Soft Skills                 | Writing           | Information and communication                    | Writing & Editing |
| Disruptive Tech Skills      | Development Tools | Information and communication                    | Computer Software |

The table 2.2 demonstrates that music industry has the highest skill penetration rate for skill groups among all industries.

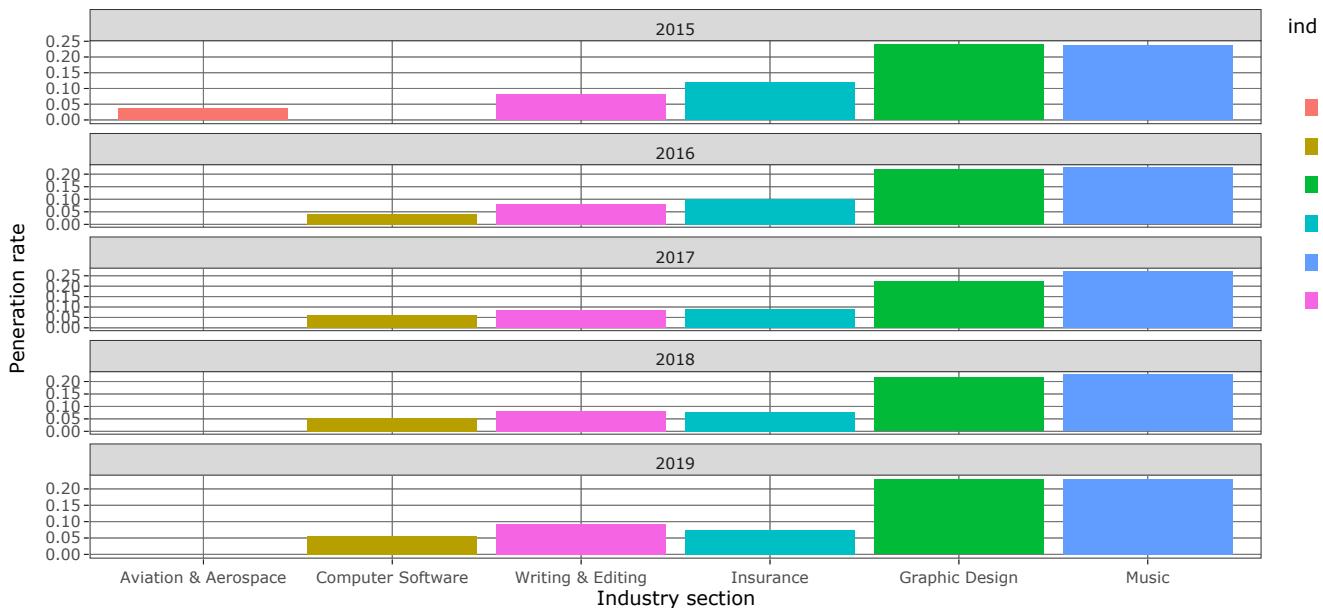


Figure 2.6: The penetration rate for different industry

The figure 2.6 shows that music industry ranks first with the penetration rate of 25% on average in a time span of five years, graphic design ranked the second (22%), insurance ranked the third(9%) and writing & editing ranked the third (8%). In addition, aviation & aerospace ranked last in 2015(3%), but were replaced by computer software(5%) in the following years.

### 2.3.2 The change of the common skill penetration rate

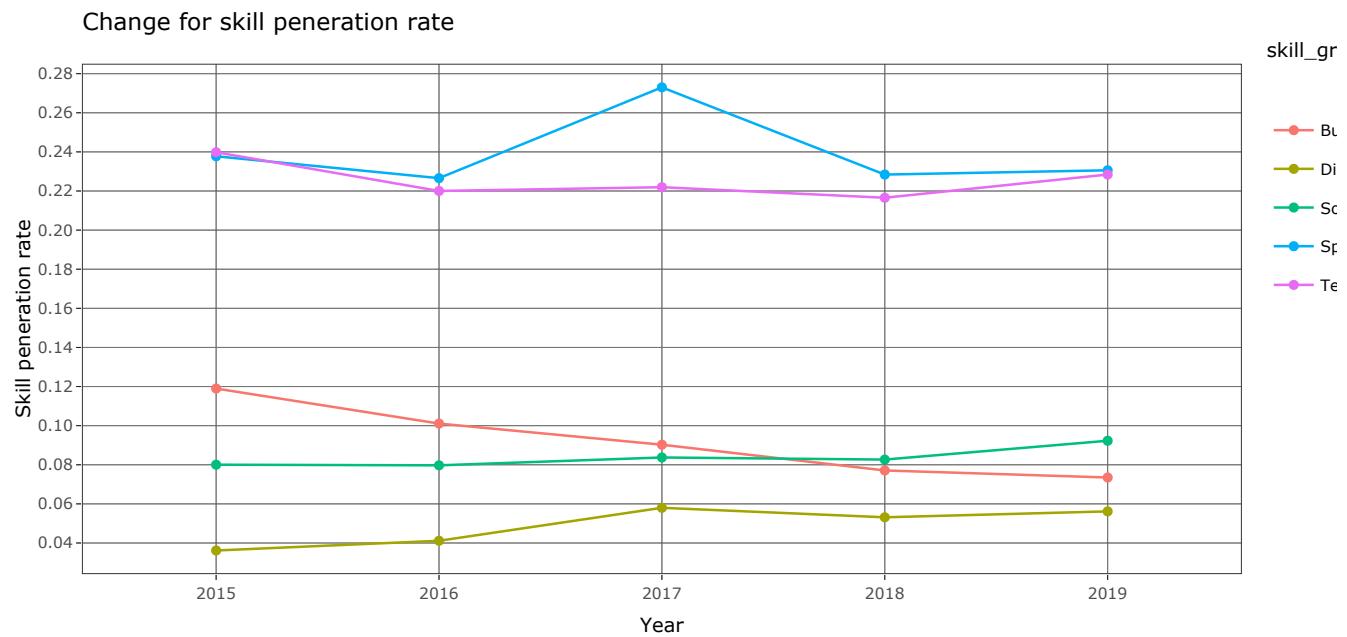


Figure 2.7: Change for skill penetration rate

The figure 2.7 deceits that specialized industry skills ranking first which peaked at 27% in 2017 had fluctuation during the time period while tech skills(2nd, average 22%) keep steadily. And soft skills had a slight climb until exceeding business skills in mid-2017(8%) which had a downward trend. However, disruptive tech skills ranks the last with 5% on average.

### 2.3.3 Inference

From the two figures, it is not hard to find that if workers want to enter an industry with a strong professional field, they are often required to master a unique and high threshold skill, such as music and graphic design. While industries with low technical penetration may require more alternative skills due to the fragmentation of the industry. Therefore it is true that the skill penetration rate of an industry is effected by various factors such as the degree of internal segmentation of the industry and the industry itself.

In general, hundreds of skills can be categorized by five common skills. Undoubtedly the specialized industry skills and tech skills has the higher rate which meet the requirements of industry development. Interestingly, with the advent of the era of big data and technology, the importance of many traditional skills like business skill has gradually declined, as shown in the decreasing penetration rate, which means that they are more replaceable in the industry and therefore no longer unique.

## 2.4 Find the industry\_section that is best to each region/continent.

The employment growth data represents the overall rate of change of employment between a pair of consecutive years for an industry, across 2015 to 2019. This rate of change is called “growth rate” which is measured by the percentage change in the number of employees for that industry .The sample of LinkedIn members is limited to those that have a company registered on LinkedIn on their profile. For any year, the number of employees working in an industry is the cumulative sum of the shift in the employed industry of the LinkedIn members, that is, the sum of the linked profiles with no shift in industry and the difference of the number of employees entering that industry and the number of employees leaving that industry. For example, if an industry has for the year 2014, 1000 employees and 100 employees enter this industry and 50 employees leave this industry in 2015, then the number of employees in this industry for the year 2015 is  $1000 + 100 - 50$  and the growth rate in 2015 is 0.05%. The formula for growth rate is,

$$\text{growthrate} = (\text{membercount}_i - \text{membercount}_j / \text{membercount}_i) * 100$$

The growth is described with respect to variables such as, country\_code, country\_name, wb\_region, wb\_income, isic\_section\_index, isic\_section\_name, industry\_id, industry\_name, year, and growth\_rate. However, the original data is in a wider format with growth rates for different years represented in the same row. As such, in accordance with the tidy data definition, the data was transformed into longer format and hence cleaned. For the question, records are filtered for the appropriate industry section for each region and then analysis is done for industries within that section.

### 2.4.1 Industry Sections: Highest and Lowest Average Growth Rate

The average growth rate is calculated for industry sections in each region and the industry sections with highest and lowest average growth rate over all the years is reported in Table 2.3 and Table 2.4 respectively.

Table 2.3: Industry Sections: Highest Average Growth Rate

| Region                     | Industry_section                   | Avg_growth_rate |
|----------------------------|------------------------------------|-----------------|
| East Asia & Pacific        | Financial and insurance activities | 0.026           |
| Europe & Central Asia      | Financial and insurance activities | 0.013           |
| Latin America & Caribbean  | Financial and insurance activities | -0.003          |
| Middle East & North Africa | Mining and quarrying               | 0.008           |
| North America              | Financial and insurance activities | 0.026           |
| South Asia                 | Manufacturing                      | -0.006          |
| South Asia                 | Mining and quarrying               | -0.006          |
| Sub-Saharan Africa         | Manufacturing                      | 0.006           |

Table 2.4: Industry Sections: Lowest Average Growth Rate

| Region                     | Industry_section                                 | Avg_growth_rate |
|----------------------------|--|-----------------|
| East Asia & Pacific        | Mining and quarrying                             | 0.002           |
| Europe & Central Asia      | Professional scientific and technical activities | 0.002           |
| North America              | Mining and quarrying                             | 0.001           |
| Sub-Saharan Africa         | Information and communication                    | -0.009          |
| Latin America & Caribbean  | Information and communication                    | -0.016          |
| Middle East & North Africa | Information and communication                    | -0.017          |
| South Asia                 | Information and communication                    | -0.019          |

### 2.4.2 Industries and Regions: Analysis

However, owing to several factors such as environmental resources, workforce and infrastructure, the economies of regions are dependent on specific industries. For example, Middle East & North Africa have huge fossil fuel deposits, and hence excel in Oil and Mining, while North America has the infrastructure and workforce for Information and communication industries. The Table 2.5 shows the the industry sections and the regions in which they had highest growth.

**Deeper Insights can be made from 2.9 and 2.10**

Table 2.5: Region: Industry Sections

| Industry_Section                   | Region                | Avg_growth_rate |
|------------------------------------|-----------------------|-----------------|
| Financial and insurance activities | East Asia & Pacific   | 0.026           |
| Arts, entertainment and recreation | East Asia & Pacific   | 0.008           |
| Mining and quarrying               | Europe & Central Asia | 0.008           |

| Industry_Section                                 | Region                     | Avg_growth_rate |
|--|----------------------------|-----------------|
| Mining and quarrying                             | Middle East & North Africa | 0.008           |
| Financial and insurance activities               | North America              | 0.026           |
| Information and communication                    | North America              | 0.022           |
| Professional scientific and technical activities | North America              | 0.014           |
| Manufacturing                                    | North America              | 0.013           |

The presence of industries in the regions is shown in Figure 2.8.

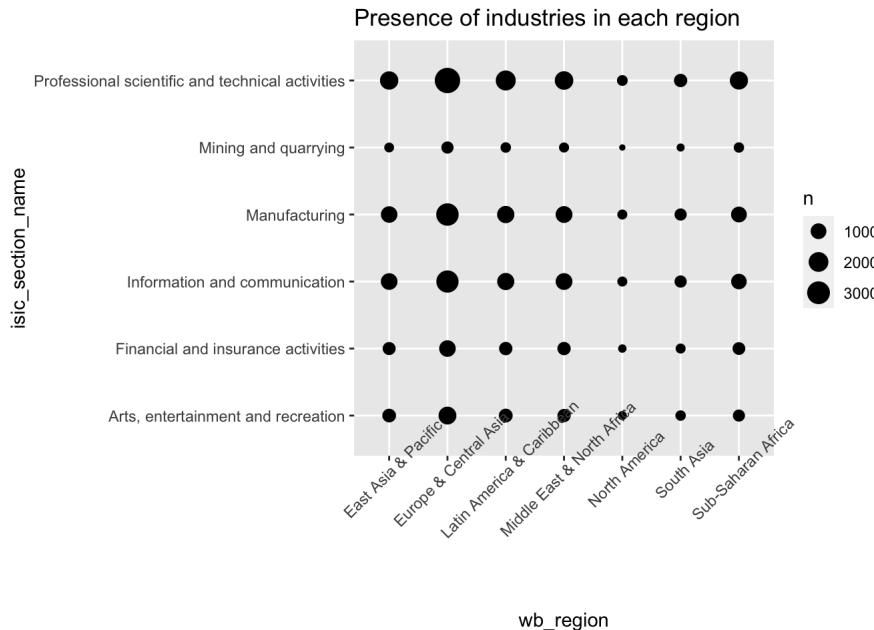


Figure 2.8: Presence of industries in each region

- **East Asia & Pacific, North America and Europe & Central Asia** have been growing in terms of employment with **Financial and insurance activities** being the most significant employer.
- Industries in **South Asia** and **Latin America & Caribbean** had only contraction, with industries under the section **Manufacturing** and **Mining and quarrying** being the least affected. In **Sub-Saharan Africa** other than **Manufacturing** all other industries have been declining in terms of employment.
- **Information and communication** has been contracting in Sub-Saharan Africa, Latin America & Caribbean, Middle East & North Africa and South Asia which otherwise has a tremendous scope in **North America**.
- **North America, East Asia & Pacific, and Europe & Central Asia** are the regions where all industries upgraded.
- **North America** has been the leader in all Financial and insurance activities, Information and communication, Professional scientific and technical activities, Manufacturing whose biggest competitor is **East Asia & Pacific**.
- **Mining and quarrying**, however, retains a strong position in **Middle East & North Africa**.

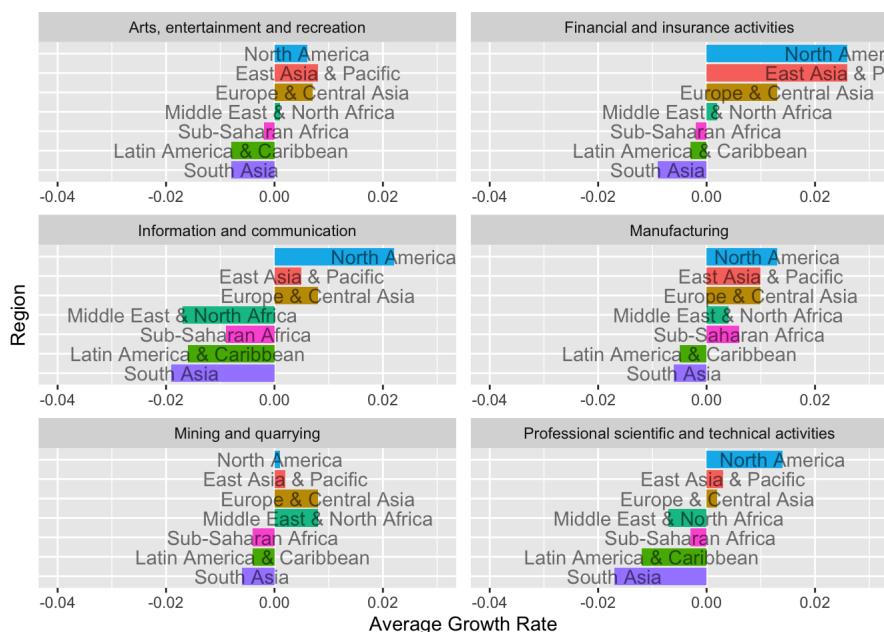


Figure 2.9: Region: Industry Sections

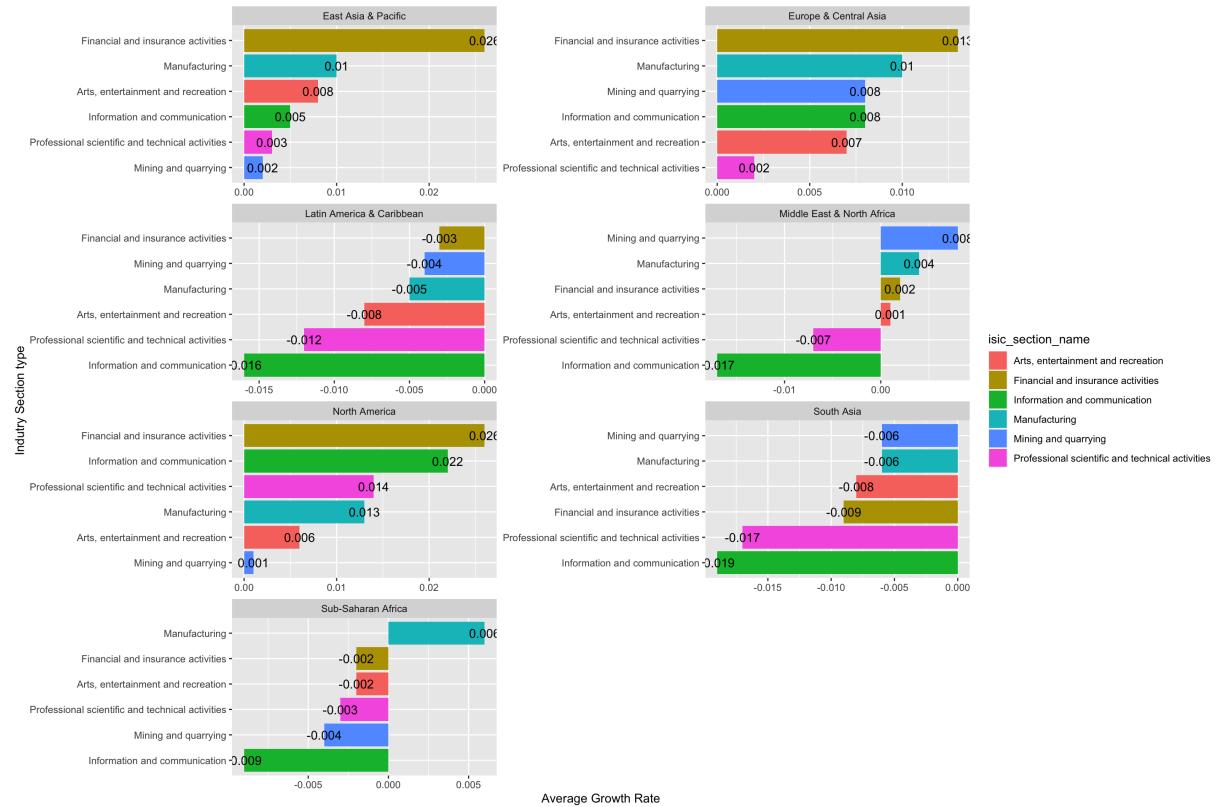


Figure 2.10: Industry Sections: Region

## 2.5 Which industry\_name sees the maximum growth over time in each region/continent , depending on the region's best industry\_section ?

The selection of an industry section for a region is based on the Table 2.3. A comprehensive analysis is made on the industries that fall in these industry sections. The number of industries in each industry section is given in Figure 2.11

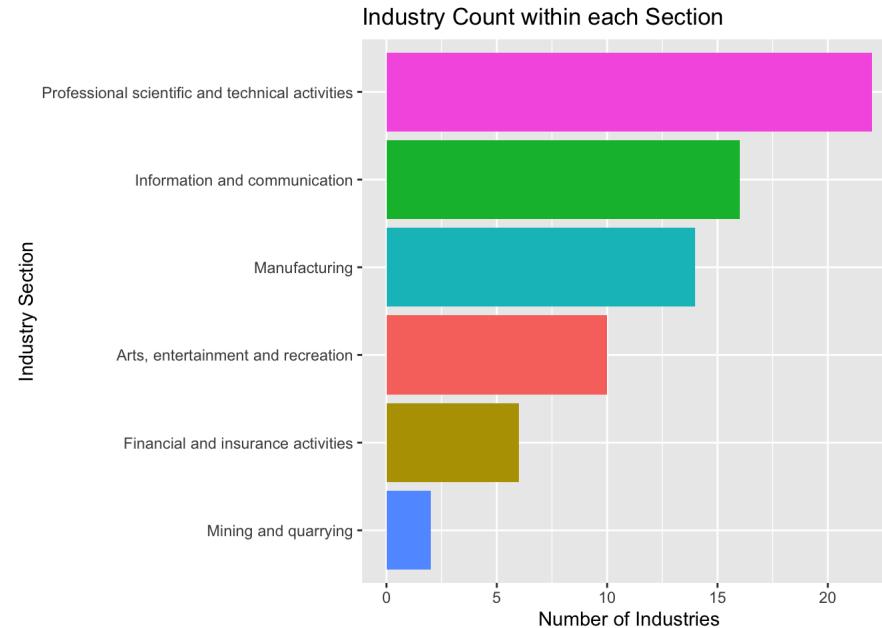


Figure 2.11: Industry Count within each Section

The regions **North America**, **East Asia & Pacific**, and **Europe & Central Asia** have a similar distribution of the growth rates for industries in **Financial and insurance activities**. Industries relating to investments have a growth rate[0.03,0.05] far exceeding other industries within this field. Banking, however remained in place. It is interesting to note that in the Middle East, Oil and Energy saw a decline. These comparisons have been made in Figure 2.12. The aggregated growth rate for each region is made in the time series graphs 2.14 .

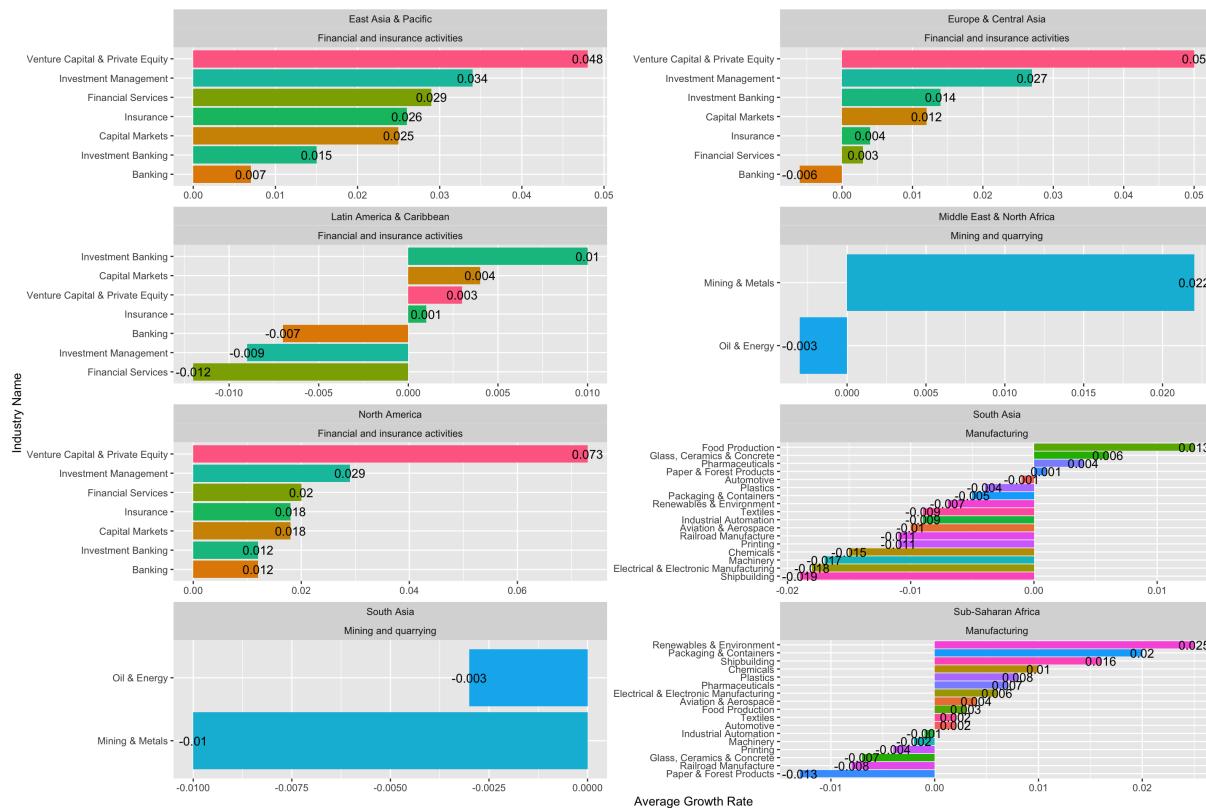


Figure 2.12: Avg. growth of an industry within a region w.r.t best industry section

Each of the time series graphs below represents the cumulative averages for the growth rates of industry sections. The regions having the same industry sections are compared in each graph. The growth rate for Mining and quarrying in South Asia has been declining below whereas in Middle East & North Africa it has seen a steady growth . **North America** and **East Asia & Pacific** are close competitors in Financial and insurance activities with North America beating East Asia & Pacific in the recent times. The growth rate for Manufacturing is a similar trend as the Mining and quarrying where steady growth is observed in Sub-Saharan Africa.

The trend of industries within each section is represented in the Figure 2.15.

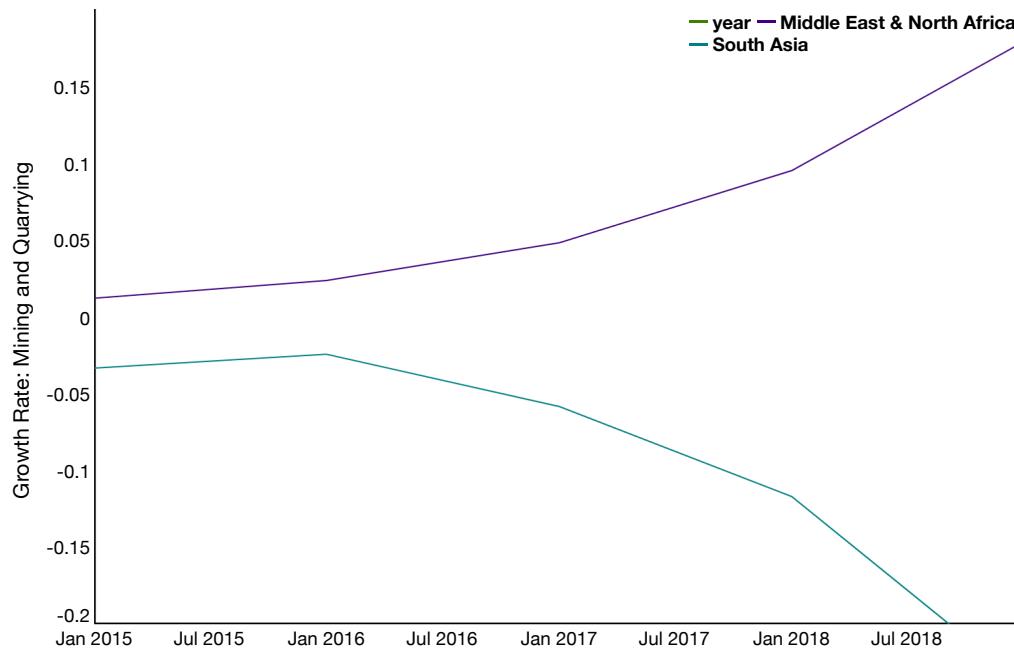
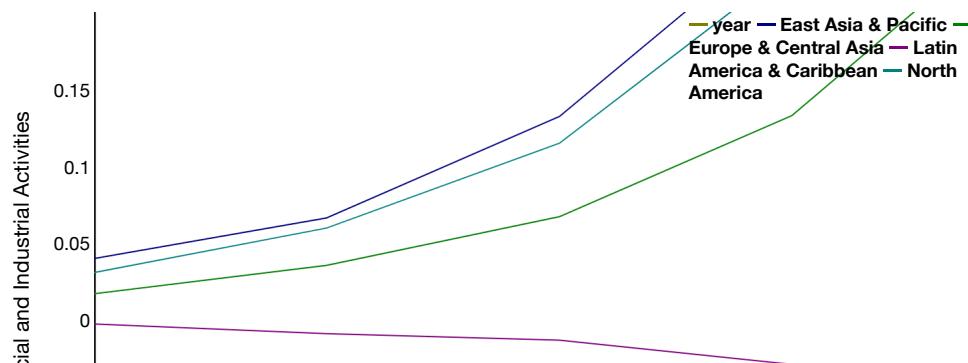


Figure 2.13: Time Series: Aggregated Growth Rate



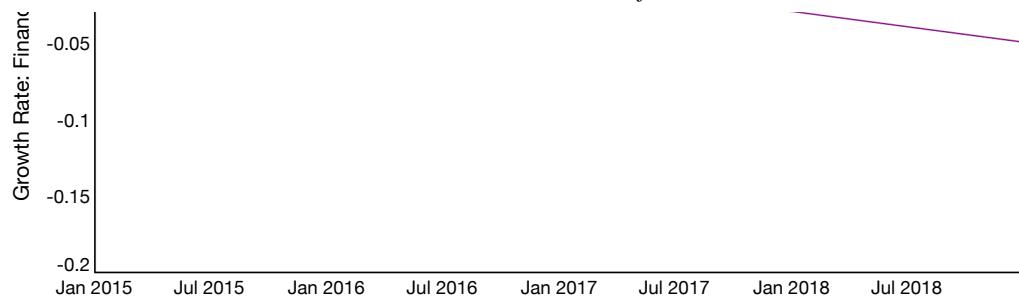


Figure 2.13: Time Series: Aggregated Growth Rate

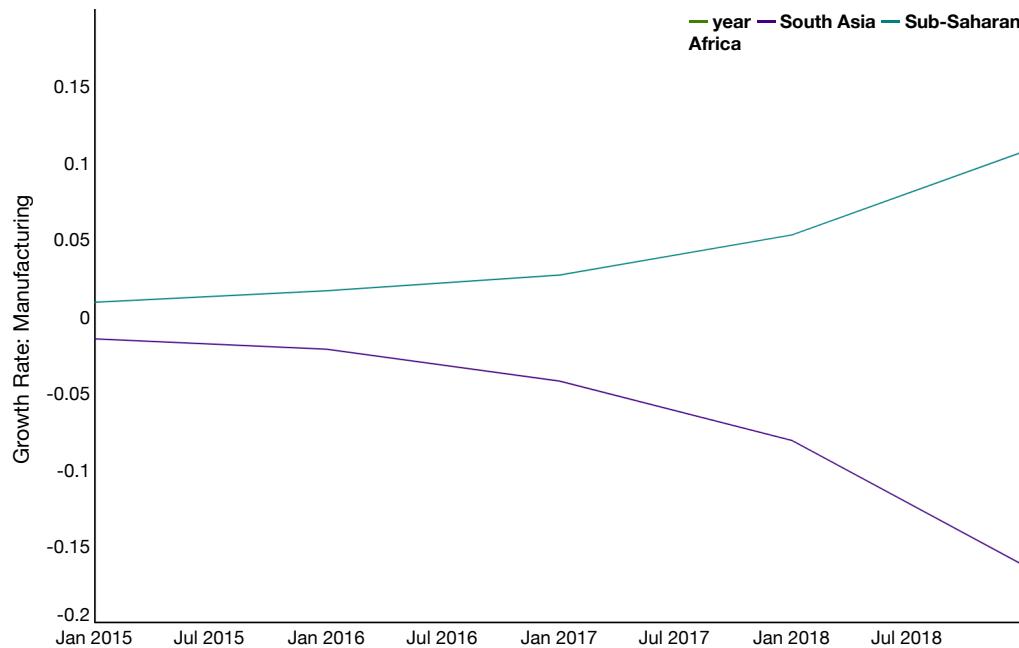


Figure 2.13: Time Series: Aggregated Growth Rate

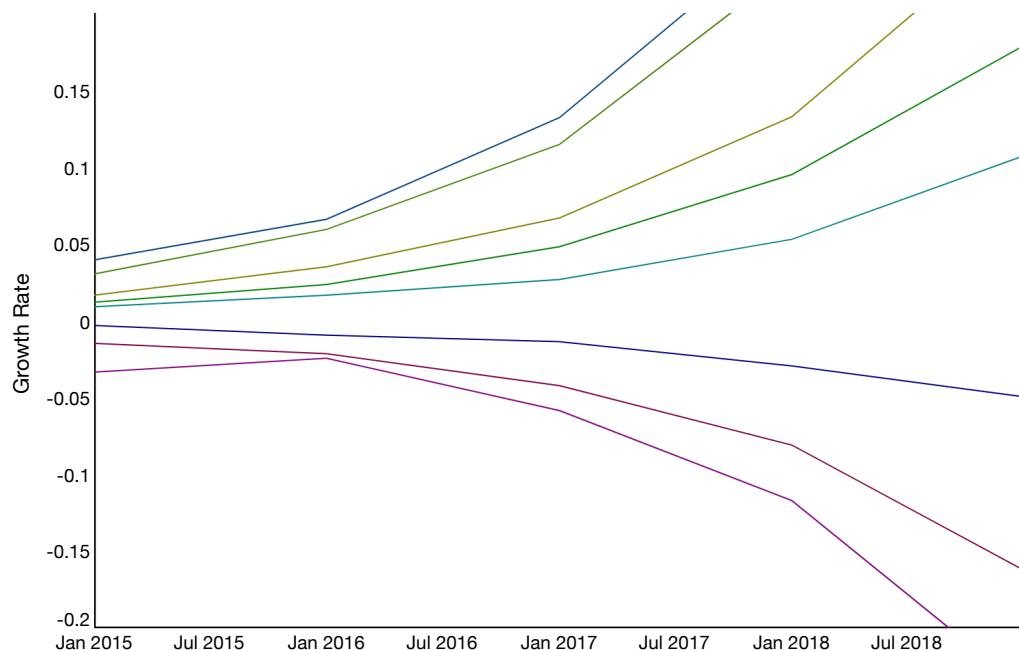


Figure 2.13: Time Series: Aggregated Growth Rate



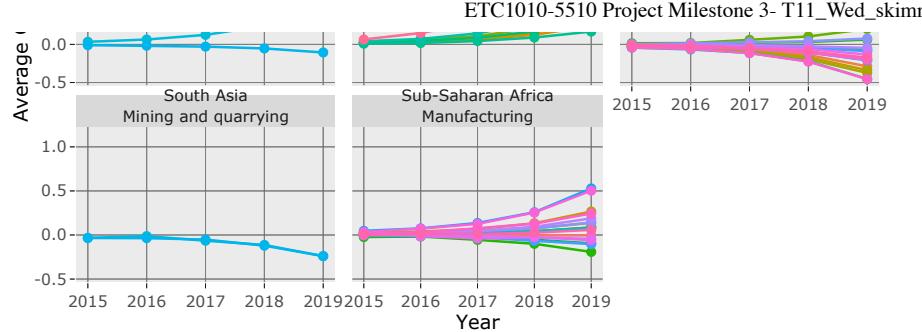


Figure 2.15: Time Series: Industry Growth Rate

## 2.5.1 Discussion on Skills and Industries

The presence of skill categories in industry sections is shown in Figure 2.16 and the count of skills within each skill category is displayed. Specialized Industry skills have a very large difference as compared to other skills owing to the fact that these skill categories are most unique in an industry. Business skills have a presence in many industries.

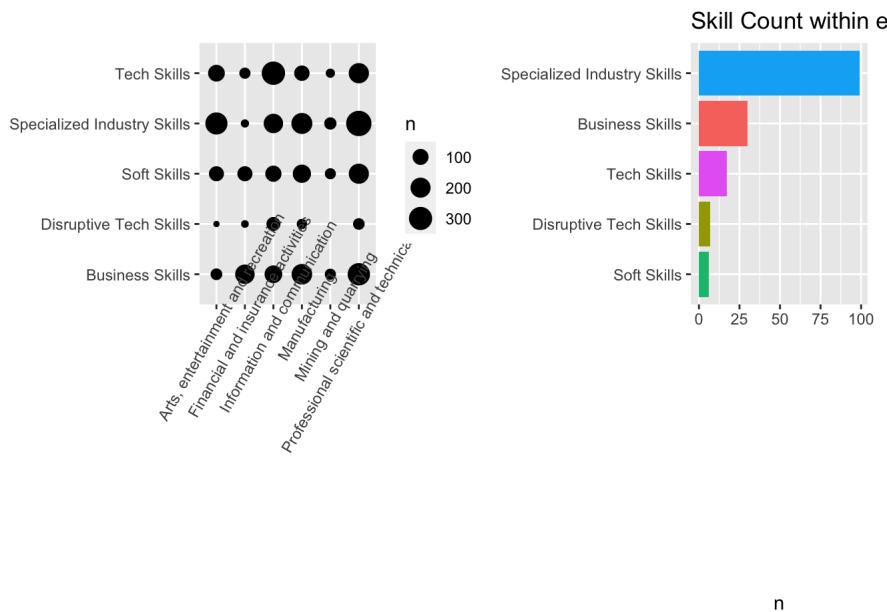


Figure 2.16: skill categories in industry sections

### 2.5.1.1 Relationship between Skill Group Rank, Industry Growth Rate and Skill Group Penetration Rate

There exists no relationship between skill group rank and skill group penetration rate and for some industries, penetration rate is higher where there is no growth or little growth, thus suggesting that employees incorporate more skills. No relationship is determined.

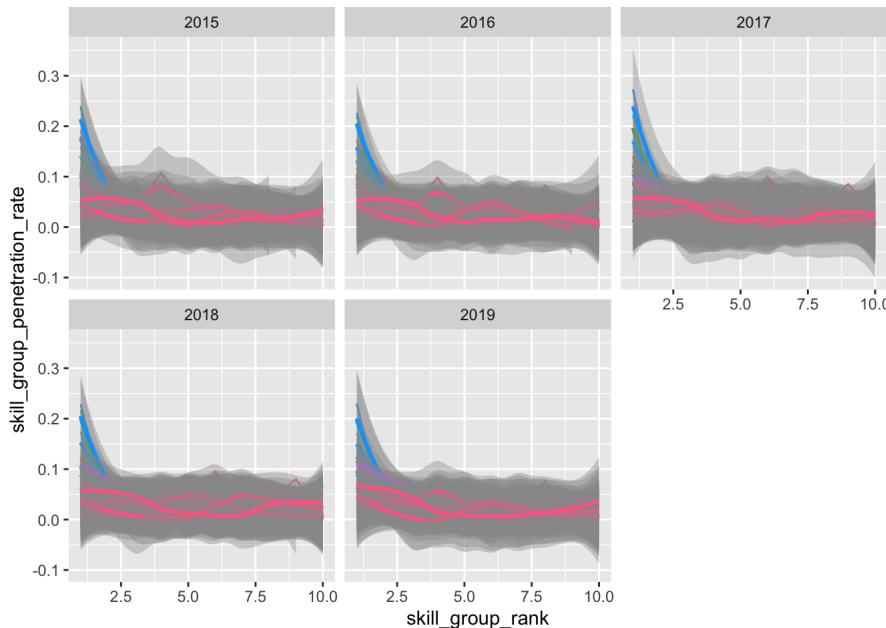


Figure 2.17: Skill Group Rank vs Skill Group Penetration Rate

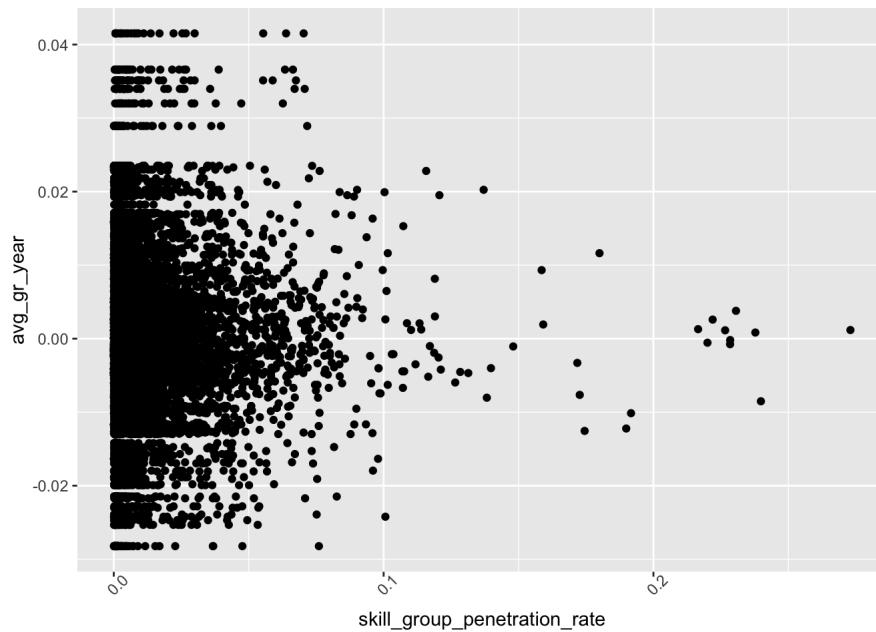


Figure 2.18: Industry Growth Rate and Skill Group Penetration Rate

## 2.5.2 Networks: Industry Sections and Skill Groups

The network below 2.19 shows the relationship between industry sections and skill categories weighted by the mean rank of these skills. Specialized Industry Skills have the highest rank across all industries. However, Financial and insurance activities demand more of Business skills. Business skills have a fair rank across industries. Tech skills and soft skills are ranked well for all industries; tech skills are more important to Information and communication whereas soft skills are important to manufacturing. Disruptive tech skills are however ranked highly only for Information and communication, manufacturing and professional, scientific and technical activities. It is to be noted that the count of combination of industry sections and skill categories in the observations(Figure 2.16) gives a similar result as the ranked network analysis.

**Detailed network of industries and skills weighed by the skill group ranks for every industry section is present in subsequent graphs.**

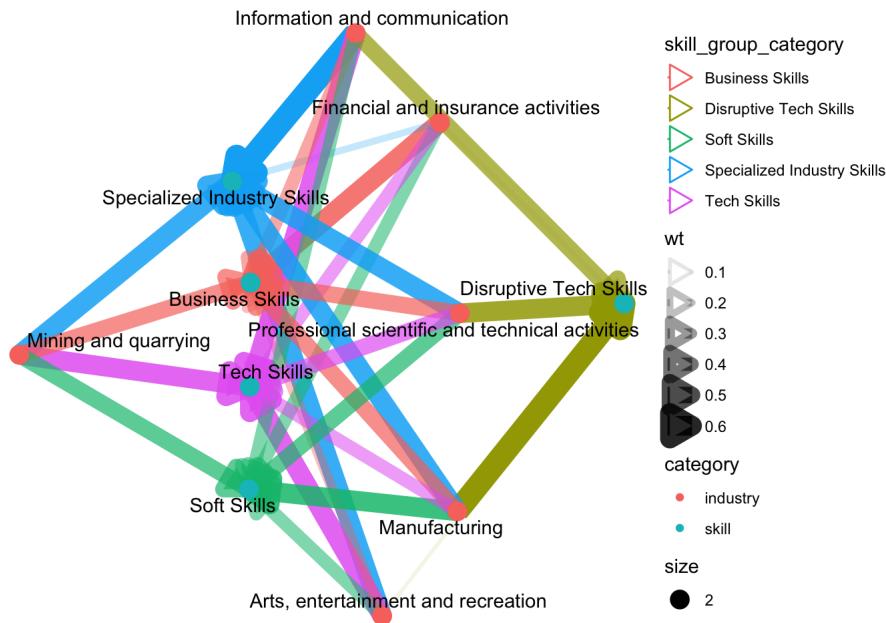


Figure 2.19: Network: Industry Section and Skill Category

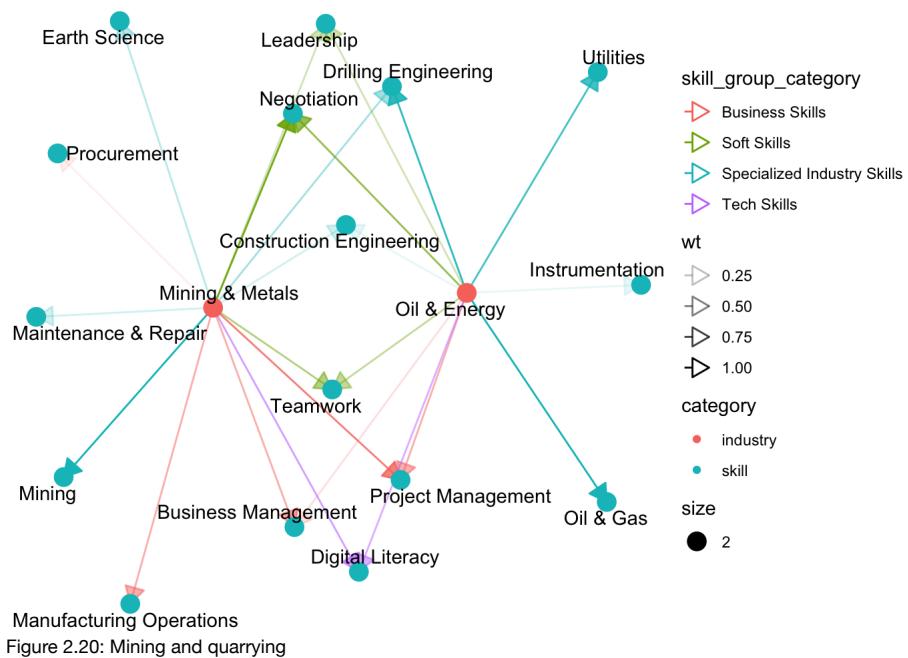


Figure 2.20: Mining and quarrying

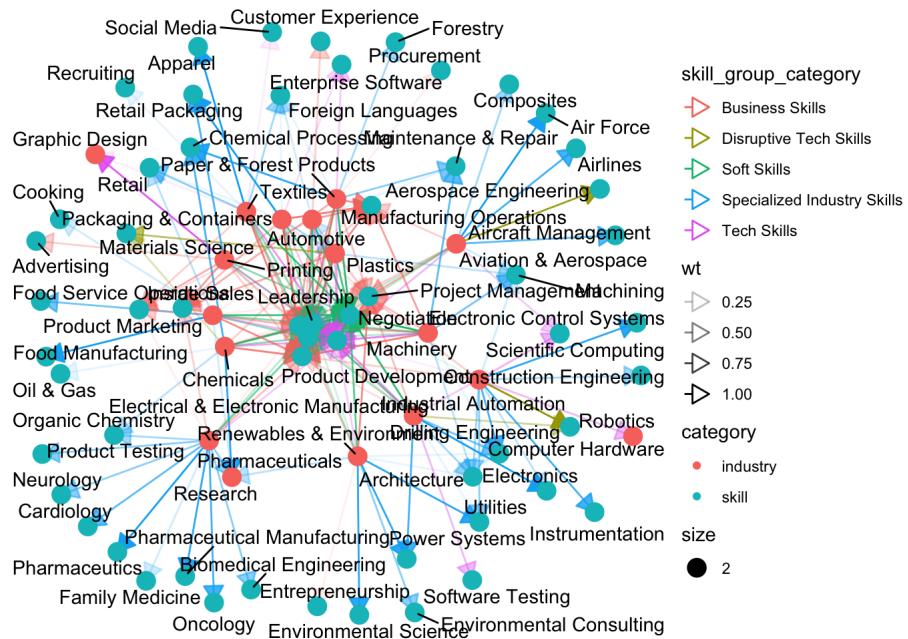


Figure 2.21: Manufacturing

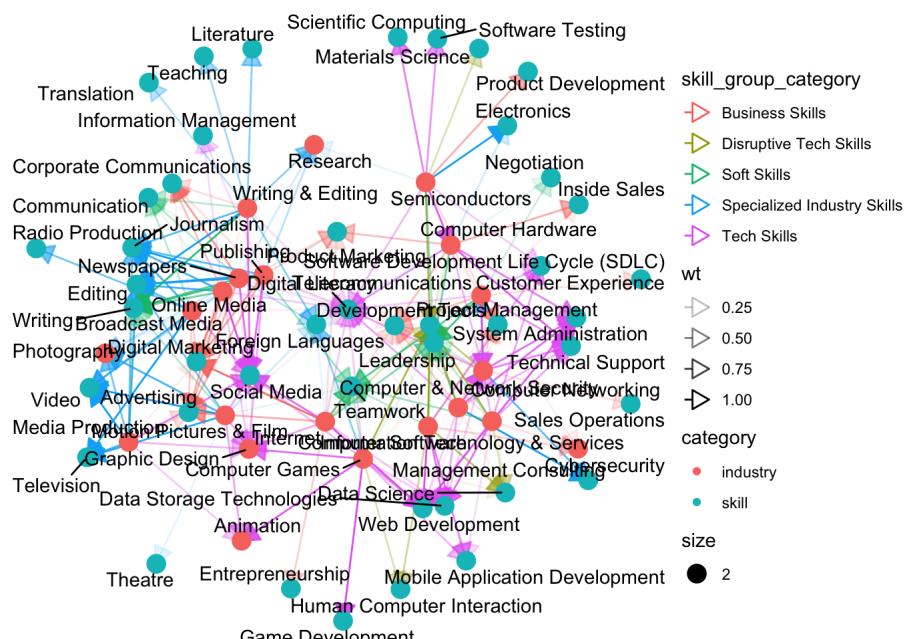


Figure 2.22: Information and communication

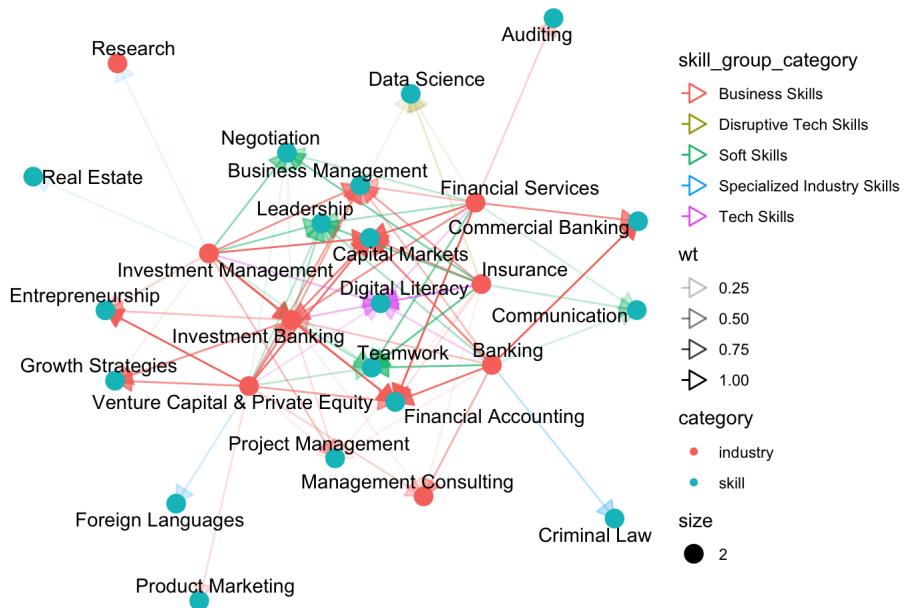


Figure 2.23: Financial and insurance activities

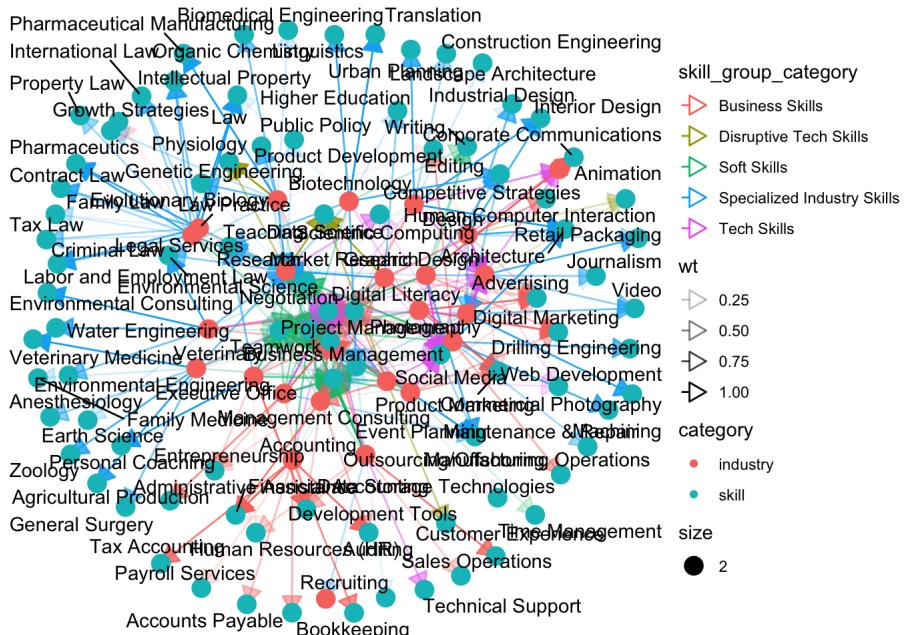


Figure 2.24: Professional scientific and technical activities

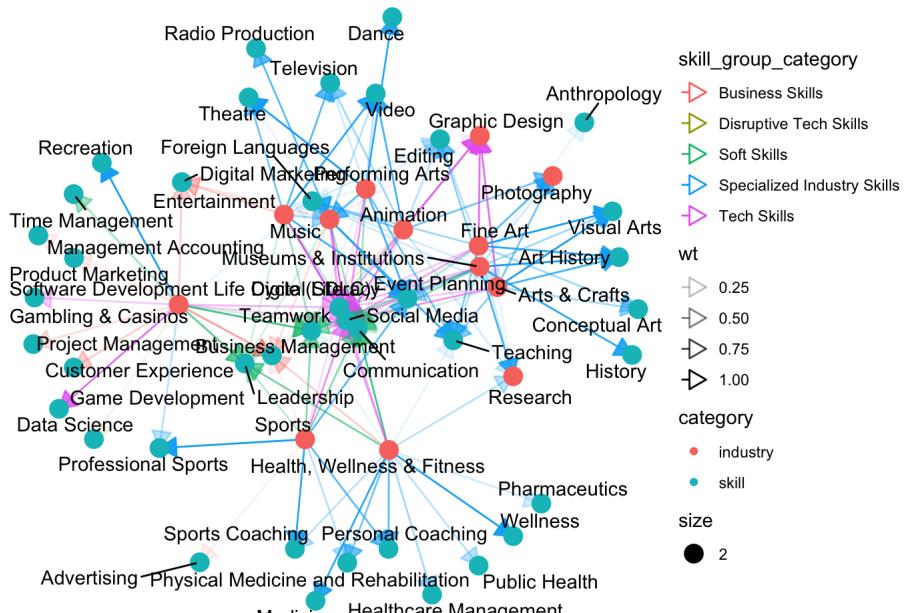


Figure 2.25: Arts, entertainment and recreation

### 2.5.3 Networks: Migration

Migration rate is the net flows(arrivals - departures) normalized based on the member count in the target country multiplied by 10000. A positive migration is when the arrivals are greater than the departures and vice-versa. The migration rate for the countries averaged over all industries and years is shown in the map 2.26

Table 2.6: Top Countries for Migration

| country_name         | average_migration_rate |
|----------------------|------------------------|
| Luxembourg           | 765.3817               |
| United Arab Emirates | 442.7116               |
| Malta                | 396.6229               |
| Estonia              | 347.1595               |
| Cyprus               | 342.0833               |
| Qatar                | 332.0523               |
| Panama               | 283.6780               |
| Myanmar              | 258.0705               |
| Kuwait               | 237.3493               |
| Mali                 | 237.1740               |
| Switzerland          | 233.6345               |
| Burkina Faso         | 220.4540               |
| Saudi Arabia         | 208.4615               |
| New Zealand          | 197.5190               |
| Bahrain              | 195.1179               |
| Ireland              | 182.0494               |
| Singapore            | 178.8927               |
| Rwanda               | 175.4360               |
| Germany              | 171.9108               |
| Papua New Guinea     | 169.6560               |
| Japan                | 168.5637               |
| Congo, Dem. Rep.     | 161.9225               |
| Zambia               | 151.8496               |
| Georgia              | 150.8675               |
| Australia            | 142.9156               |
| Austria              | 137.7601               |
| Canada               | 133.5462               |
| Chile                | 119.2267               |
| Czech Republic       | 118.9811               |
| Thailand             | 115.2884               |

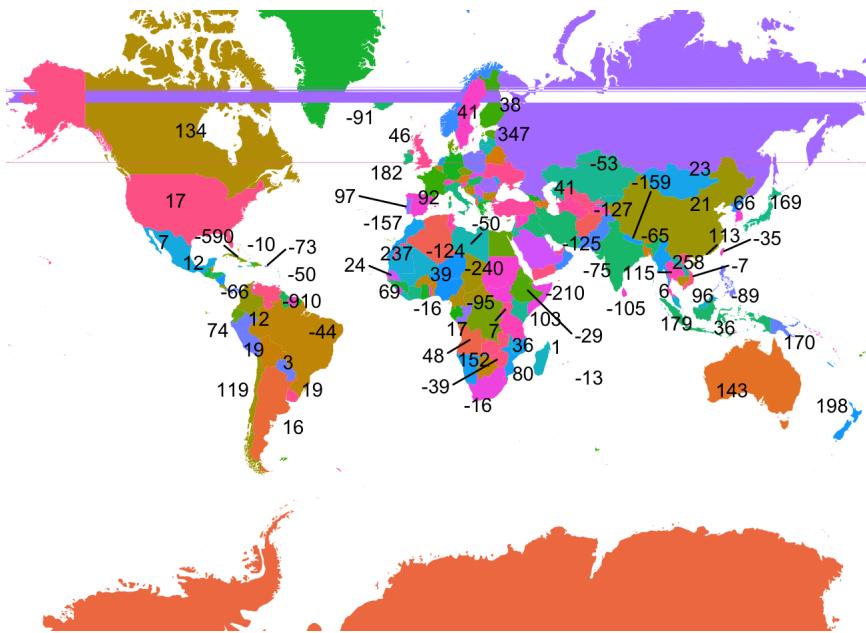


Figure 2.26: Map: Migration Rate of Countries

A network depicting the highest migration rate for a base country is shown below. This means the highest number of people that migrated to a country. The network is weighted on the average migration rate over the years. The two major clusters, the United States and India suggest that most of people from most countries migrate to the United States of America. However, for India these might be the returning people who migrated a few years ago to the base countries. We can also see that the migration linkage is also dependent on the geographical and historical ties of the countries. For example, Venezuela is target country for the countries in Latin America and Caribbean, Hong Kong to China, West Bank and Gaza to Israel.

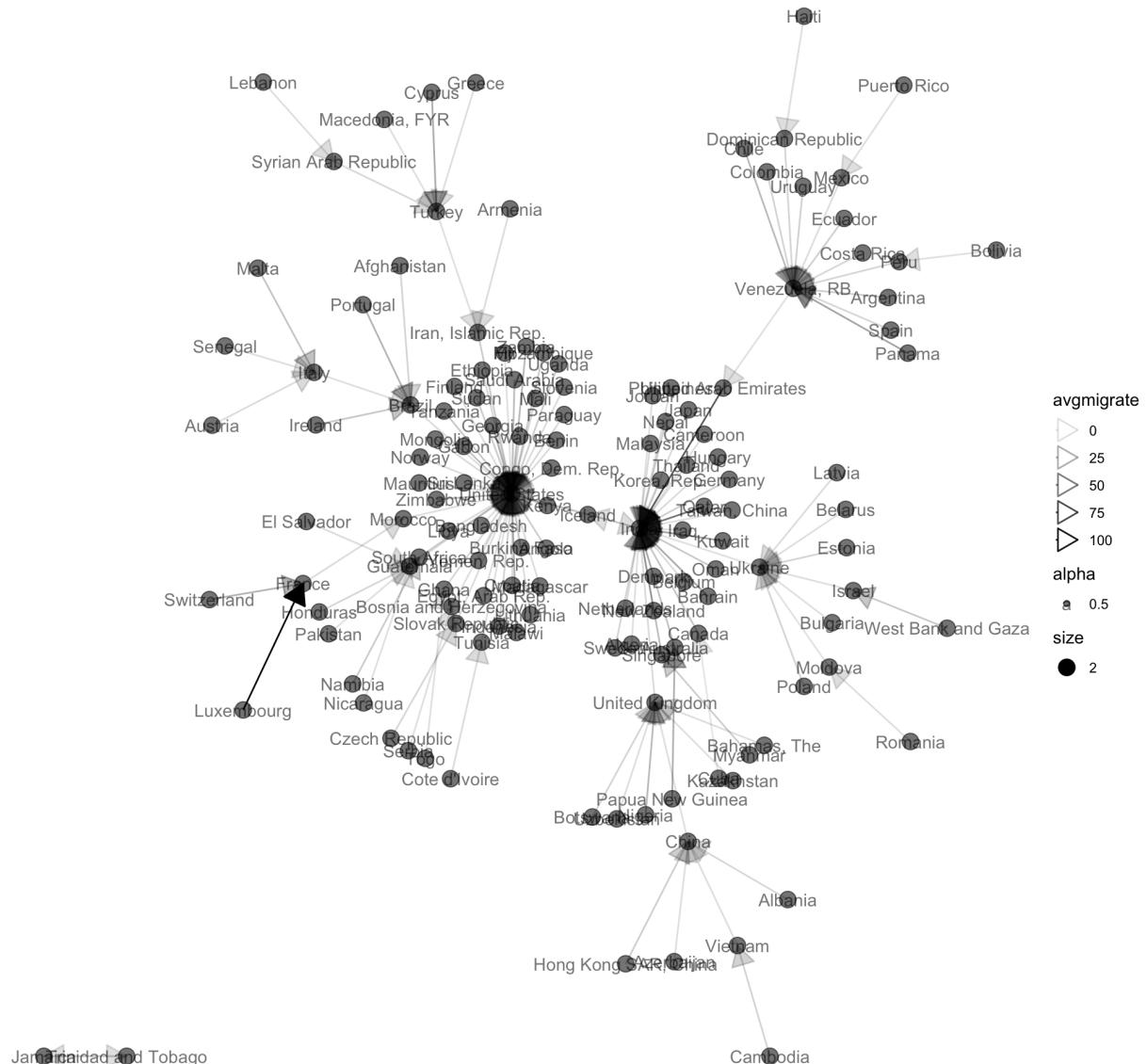
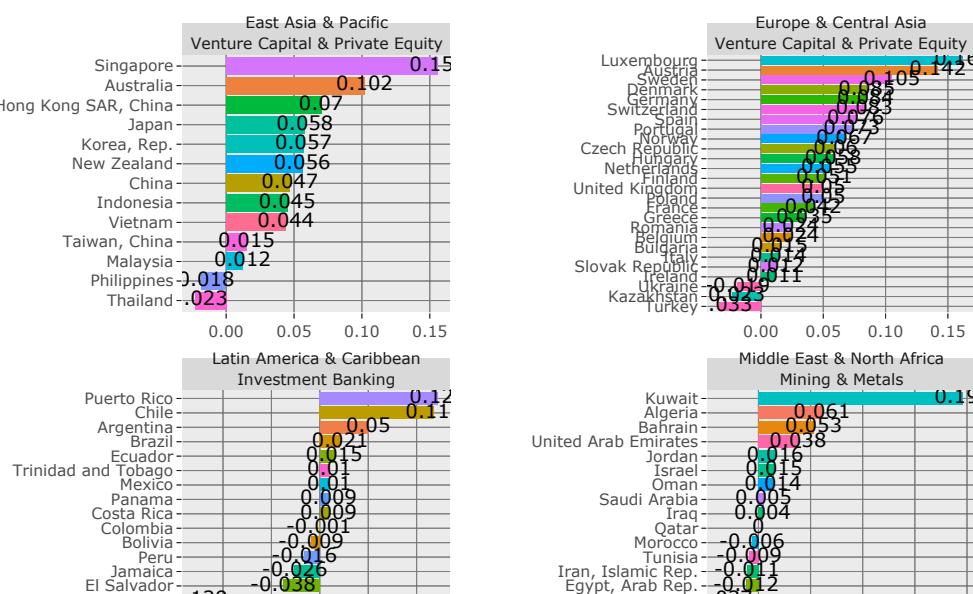


Figure 2.27: Highest Migration Rate Selected: Base Country to Target Country

**2.6 For each region, which country did the above found industry had had maximum growth? And, what is the income group of that nation?**

**2.6.1 What is the Avg. growth of the best industry within a country w.r.t ts best industry section?**



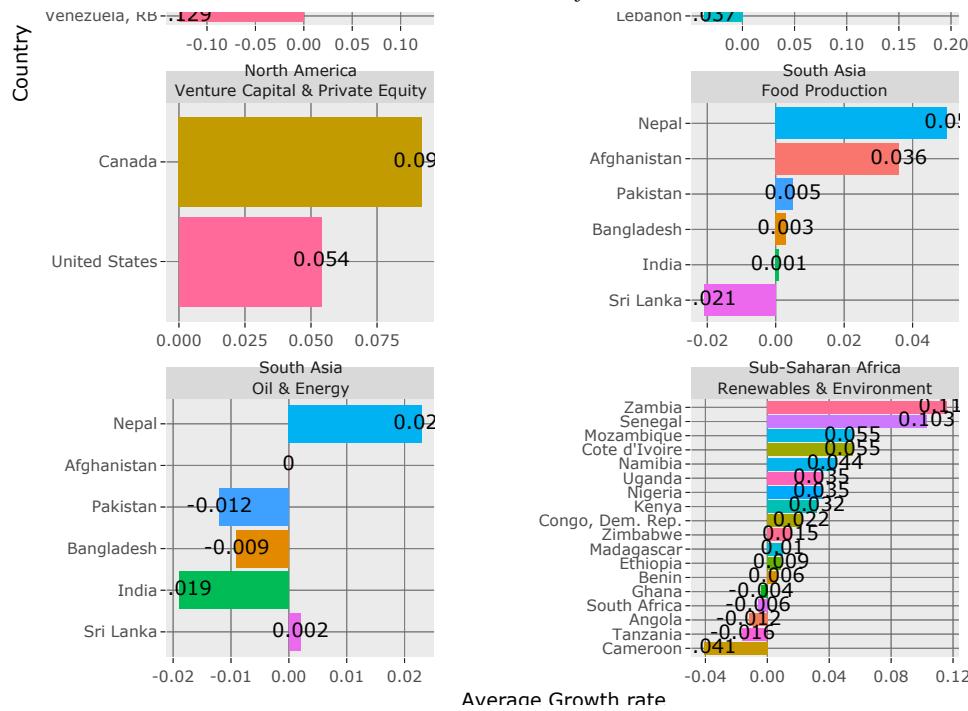


Figure 2.28: Avg growth of the best industry within in a country w.r.t region

The figure 2.28 take us further step by adding another layer of **Country** to the ??.

- Mostly every region had a big top knot Country having the max growth rate of an employee, whereas regions like **South Asia** and **Sub-Saharan Africa** had countries like **Nepal** and **Zambia** having the maximum growth rate even though coming under Low/Lower middle income categories.
- Though overall **North America** had the max growth of employee in the **Venture Capital & Private Equity** (see Figure ??), but when seen country wise, **Luxembourg** in **Europe & Central Asia** region had approximately double the growth than top country **Canada**.

## 2.6.2 What is the trend of the best industry within a country w.r.t ts best industry section?

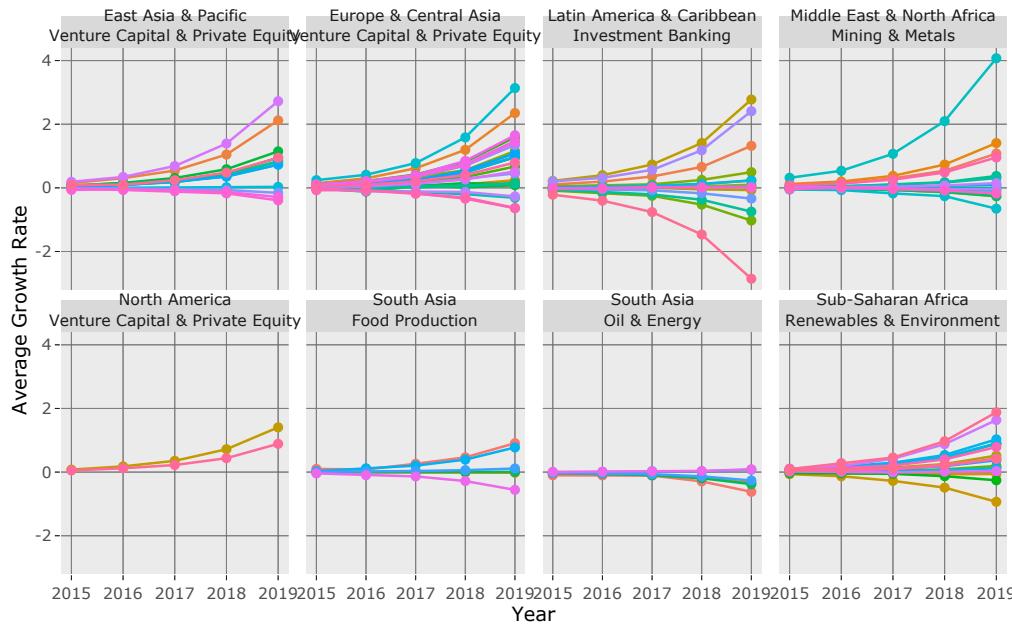


Figure 2.29: Trend of best industry within in a country w.r.t region

The figure 2.29 provides a broader insight to the figure 2.28 by showing the trend of growth in each Country in a span of 5 years from 2015-2019.

Mostly every country showed a steady graph with some exceptions like:

- In **Latin America & Caribbean**, Chile saw a constant drop in its growth rate whereas **Venezuela** showed an opposite behaviour.
- In **Middle East & North Africa** region, **Kuwait** though still topping the Avg.growth rate over the years, yet saw a massive fall in its growth rates.

## 3 Conclusion

This analysis report harnesses the dynamic, fast-growing LinkedIn dataset, which covers more than 100 countries, to derive insights about the metrics: skills, industries and migration trends of this modern world. Linked profiles have data that is valid in real time as the members tend to keep their career profiles updated. This kind of data is unlikely to be reflected in government statistics.

"LinkedIn data have unique strengths in that they enable new insights into the emerging digital sectors and skills, with near real-time updates that are unlikely to be reflected in government statistics. Certain tradable and knowledge-intensive sectors also have good coverage across income levels and geographic locations, which allows for global benchmarking. In this manner, it may from the outset serve as a complementary dataset to other government statistics. With the growing use of LinkedIn, these data can become increasingly relevant for developing countries around the globe." 5 (<https://documents1.worldbank.org/curated/en/827991542143093021/pdf/World-Bank-Group-LinkedIn-Data-Insights-Jobs-Skills-and-Migration-Trends-Methodology-and-Validation-Results.pdf>)

The data provided by The LinkedIn-World Bank Digital Data for Development is a cleaned data set which only requires to be adjusted in the wider or longer format based on the analysis question. In this report a comprehensive analysis was done with respect to these metrics on the higher level of classification: the skill group categories, industry sections and the world bank classified regions to gain an overall knowledge about the shifts in the trends of these metrics. Each question section discussed the shifts in these metrics to bring forward this knowledge and specific details were listed in the tables. Some complex networks were plotted to have a visual representation of the relationship between the skills and the industries to understand the relevance of a skill to an industry. The growth of the industries was studied with respect to the changes in its member population.

Specialized Industry Skills have the highest rank across all industries and Business, Tech skills were found to be common across all industries and were ranked similarly. Industries were categorized depending their growth rates and were mapped to different regions. This mapping summarized that North America leaded in terms of employment in several industries including Financial and insurance activities, Information and communication, Professional scientific and technical activities and Manufacturing and Financial and insurance activities was the highest. Again, the business skills and tech skills were highly ranked for this field.

The migration rates was studied which revealed that the United States is a popular migration destination from all over the world. In general, members possess a diverse set of skills and the common skills, business and tech skills, are applicable to all linked in members. This commonness compromises the rank of these skills. Hundreds of skills are be categorized into five skill categories. Undoubtedly, the specialized industry skills and tech skills have the higher rate which meet the requirements of industry development. Interestingly, with the advent of the era of big data and technology, the importance of many traditional skills has gradually declined, as shown in the decreasing penetration rate, which means that they are more replaceable in the industry and therefore no longer unique. However, these skills are basic and must be possessed in this modern era and other skills categories are industry specific additions.

The LinkedIn data provides data that brings out the generalized patterns and individual characteristics of industries and LinkedIn members in the developed countries, especially in the tradable, technology, and digital sectors.. However, this dataset has a limitation that the population of the developing countries in non-tradable, non-digital is under-represented.

## 4 References

### 4.1 Data Source

1. The LinkedIn-World Bank Digital Data for Development:Industry Jobs and Skills Trends - About (<https://linkedindata.worldbank.org/about>)
2. The World Bank: Industry Skills Needs Dataset(3500 X 7), Skill Penetration Dataset(20780 X 7) (<https://datacatalog.worldbank.org/dataset/skills-linkedin-data>)
3. The World Bank: Talent Migration Dataset(Industry Migration-5295 X 13) (<https://datacatalog.worldbank.org/dataset/talent-migration-linkedin-data>)
4. The World Bank: Industry Employment Shifts Dataset(7335 X 13) (<https://datacatalog.worldbank.org/dataset/employment-growth-linkedin-data>)
5. The World Bank: World-Bank-Group-LinkedIn-Data-Insights-Jobs-Skills-and-Migration-Trends-Methodology-and-Validation-Results (<https://documents1.worldbank.org/curated/en/827991542143093021/pdf/World-Bank-Group-LinkedIn-Data-Insights-Jobs-Skills-and-Migration-Trends-Methodology-and-Validation-Results.pdf>)
6. The World Bank: Terms of Use for Datasets(CC BY 4.0) (<https://www.worldbank.org/en/about/legal/terms-of-use-for-datasets>)

### 4.2 Bibliographies

R Core Team (2021) Xie (2021a) Tierney and Lincoln (2021) Wickham et al. (2021) Vanderkam et al. (2018) Wickham (2021a) Hijmans (2019) Kahle, Wickham, and Jackson (2019) Wickham et al. (2020) Pedersen (2021) Slowikowski (2021) Arnold (2021) Müller (2020) file. (2020) Zhu (2021) Xie (2021b) Spinu, Grolemund, and Wickham (2021) Brownrigg (2018) Tierney et al. (2020) Pedersen (2020) Sievert et al. (2021) Henry and Wickham (2020) Ryan and Ulrich (2020a) Wickham and Hester (2020) Wickham and Bryan (2019) Therneau and Atkinson (2019) Wickham (2019) Müller and Wickham (2021) Wickham (2021b) Robinson and Silge (2021) Wickham (2021c) Ulrich (2020) Tierney (2019) Ryan and Ulrich (2020b) Zeileis, Grothendieck, and Ryan (2021) Xie (2016) Kahle and Wickham (2013) Wickham (2016) Csardi and Nepusz (2006) Xie (2015) Xie (2014) Grolemund and Wickham (2011) Sievert (2020) Silge and Robinson (2016) Wickham et al. (2019) Tierney (2017) Zeileis and Grothendieck (2005)

- Arnold, Jeffrey B. 2021. *Ggthemes: Extra Themes, Scales and Geoms for Ggplot2*. <https://github.com/jrnold/ggthemes> (<https://github.com/jrnold/ggthemes>).
- Brownrigg, Ray. 2018. *Maps: Draw Geographical Maps*. <https://CRAN.R-project.org/package=maps> (<https://CRAN.R-project.org/package=maps>).
- Csardi, Gabor, and Tamás Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal Complex Systems*: 1695. <https://igraph.org> (<https://igraph.org>).
- file., See AUTHORS. 2020. *Igraph: Network Analysis and Visualization*. <https://igraph.org> (<https://igraph.org>).
- Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/> (<https://www.jstatsoft.org/v40/i03/>).
- Henry, Lionel, and Hadley Wickham. 2020. *Purrr: Functional Programming Tools*. <https://CRAN.R-project.org/package=purrr> (<https://CRAN.R-project.org/package=purrr>).
- Hijmans, Robert J. 2019. *Geosphere: Spherical Trigonometry*. <https://CRAN.R-project.org/package=geosphere> (<https://CRAN.R-project.org/package=geosphere>).
- Kahle, David, and Hadley Wickham. 2013. "Ggmap: Spatial Visualization with Ggplot2." *The R Journal* 5 (1): 144–61. <https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf> (<https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>).
- Kahle, David, Hadley Wickham, and Scott Jackson. 2019. *Ggmap: Spatial Visualization with Ggplot2*. <https://github.com/dkahle/ggmap> (<https://github.com/dkahle/ggmap>).
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here> (<https://CRAN.R-project.org/package=here>).

- Müller, Kirill, and Hadley Wickham. 2021. *Tibble: Simple Data Frames*. <https://CRAN.R-project.org/package=tibble> (<https://CRAN.R-project.org/package=tibble>).
- Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork> (<https://CRAN.R-project.org/package=patchwork>).
- — . 2021. *Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. <https://CRAN.R-project.org/package=ggraph> (<https://CRAN.R-project.org/package=ggraph>).
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/> (<https://www.R-project.org/>).
- Robinson, David, and Julia Silge. 2021. *Tidytext: Text Mining Using Dplyr, Ggplot2, and Other Tidy Tools*. <https://github.com/juliasilge/tidytext> (<https://github.com/juliasilge/tidytext>).
- Ryan, Jeffrey A., and Joshua M. Ulrich. 2020a. *Quantmod: Quantitative Financial Modelling Framework*. <https://CRAN.R-project.org/package=quantmod> (<https://CRAN.R-project.org/package=quantmod>).
- — . 2020b. *Xts: eXtensible Time Series*. <https://github.com/joshuaulrich/xts> (<https://github.com/joshuaulrich/xts>).
- Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with r, Plotly, and Shiny*. Chapman; Hall/CRC. <https://plotly-r.com> (<https://plotly-r.com>).
- Sievert, Carson, Chris Parmer, Toby Hocking, Scott Chamberlain, Karthik Ram, Marianne Corvellec, and Pedro Despouy. 2021. *Plotly: Create Interactive Web Graphics via Plotly.js*. <https://CRAN.R-project.org/package=plotly> (<https://CRAN.R-project.org/package=plotly>).
- Silge, Julia, and David Robinson. 2016. “Tidytext: Text Mining and Analysis Using Tidy Data Principles in r.” *JOSS* 1 (3). <https://doi.org/10.21105/joss.00037> (<https://doi.org/10.21105/joss.00037>).
- Slowikowski, Kamil. 2021. *Ggrepel: Automatically Position Non-Overlapping Text Labels with Ggplot2*. <https://github.com/slowkow/ggrepel> (<https://github.com/slowkow/ggrepel>).
- Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. 2021. *Lubridate: Make Dealing with Dates a Little Easier*. <https://CRAN.R-project.org/package=lubridate> (<https://CRAN.R-project.org/package=lubridate>).
- Therneau, Terry, and Beth Atkinson. 2019. *Rpart: Recursive Partitioning and Regression Trees*. <https://CRAN.R-project.org/package=rpart> (<https://CRAN.R-project.org/package=rpart>).
- Tierney, Nicholas. 2017. “Visdat: Visualising Whole Data Frames.” *JOSS* 2 (16): 355. <https://doi.org/10.21105/joss.00355> (<https://doi.org/10.21105/joss.00355>).
- — . 2019. *Visdat: Preliminary Visualisation of Data*. <https://CRAN.R-project.org/package=visdat> (<https://CRAN.R-project.org/package=visdat>).
- Tierney, Nicholas, Di Cook, Miles McBain, and Colin Fay. 2020. *Nanar: Data Structures, Summaries, and Visualisations for Missing Data*. <https://github.com/njtierney/nanar> (<https://github.com/njtierney/nanar>).
- Tierney, Nicholas, and Matthew Lincoln. 2021. *Broomstick: Convert Decision Tree Objects into Tidy Data Frames*.
- Ulrich, Joshua. 2020. *TTR: Technical Trading Rules*. <https://github.com/joshuaulrich/TTR> (<https://github.com/joshuaulrich/TTR>).
- Vanderkam, Dan, JJ Allaire, Jonathan Owen, Daniel Gromer, and Benoit Thieurmel. 2018. *Dygraphs: Interface to Dygraphs Interactive Time Series Charting Library*. <https://github.com/rstudio/dygraphs> (<https://github.com/rstudio/dygraphs>).
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org> (<https://ggplot2.tidyverse.org>).
- — . 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr> (<https://CRAN.R-project.org/package=stringr>).
- — . 2021a. *Forcats: Tools for Working with Categorical Variables (Factors)*. <https://CRAN.R-project.org/package=forcats> (<https://CRAN.R-project.org/package=forcats>).
- — . 2021b. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr> (<https://CRAN.R-project.org/package=tidyr>).
- — . 2021c. *Tidyverse: Easily Install and Load the Tidyverse*. <https://CRAN.R-project.org/package=tidyverse> (<https://CRAN.R-project.org/package=tidyverse>).
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686> (<https://doi.org/10.21105/joss.01686>).
- Wickham, Hadley, and Jennifer Bryan. 2019. *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl> (<https://CRAN.R-project.org/package=readxl>).
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2020. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2> (<https://CRAN.R-project.org/package=ggplot2>).
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr> (<https://CRAN.R-project.org/package=dplyr>).
- Wickham, Hadley, and Jim Hester. 2020. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr> (<https://CRAN.R-project.org/package=readr>).
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595> (<http://www.crcpress.com/product/isbn/9781466561595>).
- — . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/> (<https://yihui.org/knitr/>).
- — . 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/bookdown> (<https://bookdown.org/yihui/bookdown>).
- — . 2021a. *Bookdown: Authoring Books and Technical Documents with r Markdown*. <https://CRAN.R-project.org/package=bookdown> (<https://CRAN.R-project.org/package=bookdown>).
- — . 2021b. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/> (<https://yihui.org/knitr/>).
- Zeileis, Achim, and Gabor Grothendieck. 2005. “Zoo: S3 Infrastructure for Regular and Irregular Time Series.” *Journal of Statistical Software* 14 (6): 1–27. <https://doi.org/10.18637/jss.v014.i06> (<https://doi.org/10.18637/jss.v014.i06>).
- Zeileis, Achim, Gabor Grothendieck, and Jeffrey A. Ryan. 2021. *Zoo: S3 Infrastructure for Regular and Irregular Time Series (z’s Ordered Observations)*. <https://zoo.R-Forge.R-project.org/> (<https://zoo.R-Forge.R-project.org/>).
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with Kable and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra> (<https://CRAN.R-project.org/package=kableExtra>).

## Appendix

Country – countries with 100,000+ LinkedIn members.

World Bank Region – countries as classified given the most recent 6 regional World Bank country categories.

**World Bank Income Group** – countries are classified given the most recent World Bank country classification by GNI into 4 categories: Low Income, Lower Middle Income, Upper Middle Income, and High Income.

**Industry** – Detailed economic activity defined through the LinkedIn industry classification (approximately ISIC Rev. 4 2 digit level), covering approximately 140 industries (industries may be excluded based on data quality considerations) which compose the six ISIC Rev. 4 tradable sectors (ISIC Index: B, C, K, J, M, R). Please see LinkedIn – ISIC industry mapping file <https://datacatalog.worldbank.org/node/144635> (<https://datacatalog.worldbank.org/node/144635>)

**ISIC Section** – The LinkedIn industry taxonomy is mapped to ISIC Rev. 4 Sector (1 digit) categories. Data is limited to 6 tradable sectors (ISIC Index: B, C, K, J, M, R). Please see LinkedIn – ISIC industry mapping file <https://datacatalog.worldbank.org/node/144635> (<https://datacatalog.worldbank.org/node/144635>)

**Tradable and Knowledge-Intensive Sectors** – Six knowledge-intensive and tradable sectors, using ISIC Rev. 4 classification, are: B-mining and quarrying; C-manufacturing; J-information and communication; K-financial and insurance activities; M-professional, scientific, and technical activities; and R-arts, entertainment and recreation.

**Skill Group** – Skill groups categorize the 50,000 detailed individual skills into approximately 250 skills groups (some skill groups may be excluded based data quality considerations). Skill related metrics are presented at the skill group rather than detailed skill level.

**Industry Skills Needs** – Captures the most-distinctive, most-represented skills of LinkedIn members working in a particular industry. Based on the skills section of the LinkedIn profile. It's calculated using an adapted version of a text mining technique called Term Frequency - Inverse Document Frequency (TF-IDF).

**Skill Penetration** – Measures the time trend of a skill across all occupations within an industry. Based on skill addition rates, and the number of times a particular skill appears in the top 30 skills added across all of the occupations within an industry. For example, if 3 of 30 skills for Data Scientists in the Information Services industry fall into the Artificial Intelligence skill group, Artificial Intelligence has a 10% penetration for Data Scientists in Information Services. These penetration rates are averaged across occupations to derive the industry averages reported.

**Migration Overview** – All the metrics are based on net migration (arrivals minus departures). These net migration figures are each normalized differently to enable fairer comparisons across samples. We calculate all on an annual basis, and report an average of the last three years.

**Industry Migration** – Industries gained and lost. Based on the industry associated with a member's company at the time of migration. The net gain or loss of members from another country working in a given industry divided by the number of LinkedIn members working in that industry in the target (or selected) country, multiplied by 10,000.

**Industry Employment Shifts** – Captures the transitions among industries over time by LinkedIn members as a proxy for industry employment growth. Based on the industries declared by the companies in a member's work history.