

Nanodegree

Final project cover sheet

Title: Gender Disparity in the Technology Industry: A Deeper Look at the UK Situation

Team: Betty Abate, Kamile Sudziute, Monaliza Wohlers, Rachel Cunningham and Safia Elmi

Submission date: 19 December 2021

1 INTRODUCTION

We ourselves are participating in an initiative encouraging more women into technology, yet figures suggest that women hold between 10–30% of technology roles and even less in senior positions (Harvey Nash Tech Survey, 2021; Office for National Statistics, 2021). Therefore, further investigation is needed to understand the disparity that exists in the UK and whether there are any regional differences. It is also pertinent that the trends in gender disparity be analysed to assess if changes are taking place over time, rather than simply evaluating the proportion of women in technology roles for a single year. The following will outline the specific aims and objectives of the current project and how we plan to address these.

1.1 Aims

Our primary goal is to investigate gender disparity in the UK technology industry, across all UK nations, in the past five years (2016–2021).

1.2 Objectives

- Investigate to what extent a gender disparity exists between males and females in the technology industry by region, year, and the UK overall.
- Investigate whether gender disparity exists between certain occupations within this industry. This will be analysed by region, UK total and by year.
- Assess trends of count for males and females in the technology using time series over the past five years.
- Predictive modelling will be used to anticipate future trends in the count for males and females in the technology industry overall in the UK.

1.3 Roadmap of the Project

Week 1	Week 2
<ul style="list-style-type: none">Decide on specific topic for projectExplore potential data sets and extract data of interestSWOT analysis of group member's strengthsPrepare documents for project and submission checklistWhat are our aims and objectives/what do we want to know?Do we have enough data for our project topic to proceed?	<ul style="list-style-type: none">Data cleaningPreparing data for analysesDiscuss potential analysesSubmit HW 3 with project plans and outlineDivide tasks between group membersGitHub Repo setup

Week 3	Week 4
Data Analysis	Finalising Work Document
Drafting Work Document	Prepare ppt presentation
Drafting ReadMe file	Finalise notebook (clean, headings, commented, logical order)
Predictive Modelling: trends forecasting	Make sure repo is up to date and contains all necessary files
	Check submission list requirements are met

1.4 Roadmap of the Report

2. [Background](#)
3. [Steps Specifications](#)
4. [Implementation and Execution](#)
5. [Results Reporting](#)
6. [Conclusion](#)
7. [References](#)

2 BACKGROUND

As individuals who are part of a project aimed at getting more women into technology, finding out more about the current state of gender disparity is of particular interest. We chose to look at the UK context as this is of particular relevance to our own experience but also to avoid over-complication of the data by introducing other potential confounding variables. But the purpose of our project is not simply personal interest; it is also important to better understand how things have shifted in recent years and whether the gap is narrowing or not. By analysing this trend, it will be possible to assess whether the number of people in technology roles is increasing overall and whether the number of women in technology roles is increasing relative to this. Our analysis will also look at the gender of those in different occupations within the technology industry to assess whether differences exist depending on the occupation (and where applicable, occupation level, e.g. manager, entry-level etc.) within this industry. Although data is available for employment by sex for the UK, there are few examples of where the technology industry has been looked at specifically, and particularly the trends in this data. The purpose of this analysis is to inform organisations of where gaps still exist in terms of gender disparity within the technology industry and also to highlight if the gap is narrowing. It is also to identify how things have changed over time and therefore allow for predictions on future patterns in this trend.

3 STEPS SPECIFICATIONS

3.1 Question Framing

After deciding that we wanted to analyse the topic of gender in technology, we started to look at potential areas of focus. We began by exploring the available data sources to help inform our question development. We also wanted to apply a variety of the skills we have learned so far in the course and so sought out a diverse range of data types and information. During this whole process, we used informal chat platforms and video calls to refine our project focus. The data we found guided our

decision to focus on gender disparity in the technology industry in the UK. It would also allow us to examine a breakdown by different occupations in this industry, therefore allowing for in-depth analysis and opportunities to apply diverse analyses.

3.2 Data Gathering

Data was sourced from the Annual Population Survey via the ONS data hub. We also took gender pay data from national labour and workforce resources but decided this data was not suitable for our analysis. To gather data on the overall trends of males and females in the technology industry, we used the ONS API. We collated data for the total count in all regions of the UK for 2016–2021. The decision to use this time frame was due to the short time scale of this project and the work involved in extracting the API data. In hindsight, and for the purposes of predictive models, data spanning a broader time frame would have been useful.

Data was also downloaded in CSV format, also from the ONS, detailing information on the count of males and females in seven different occupations within the technology industry (information technology and telecommunications directors; IT specialist managers; IT project and programme managers; IT business analysts, architects and systems designers; programmers and software development professionals; web design and development professionals; information technology and telecommunications professionals, n.e.c.). We kept the date range the same as a matter of consistency between datasets.

3.3 Preprocessing

In order to access the API data, we had to extract specific values and then append them to an empty data frame. We first had to turn the data into a JSON file then access specific values using the `itemgetter` method. This was used because the JSON file contains a list of dictionaries and so required extra steps to access. These values were then assigned to a variable, and a data frame was created. Producing the data frame allowed us to identify which values were required for our analyses. From there, we could then extract specific values for the count of males and females for a specific region and year and assign these to variables. These variables then helped create the final data frame, which contains information on the count of males and females for each region in the UK from 2016 to 2021.

The CSV file containing the raw data for occupation and gender was derived from annual surveys that were carried out yearly by Nomis from 2016 to 2021. The dataset contains the count of employees across seven different sectors within the field of technology, as well as a breakdown of the number of males and females there within those sectors. In order to clean the datasets, we accessed the file and extracted the information that was relevant to the topic at hand. That meant locating all the values for all employees, male employees and female employees within the seven sectors recorded from 2016 to 2021, extracting that information and saving it to a new CSV file. Upon reflection, it was clear to see that there were a lot more missing data values for the females in the dataset than men. However, we were able to construct a dataset for female employees using the values for all employees and male employees. From there, we were able to compile and categorise the data into relevant CSV files: one for all employees, one for all males, and another for all females from 2016 to 2021.

3.4 In-Depth Analysis

For the API data, we were primarily interested in the trends in the number of males and females in technology. The reason for this interest was twofold: Firstly, we wanted to see how the numbers were increasing overall and whether the numbers for males and females were increasing concurrently and therefore reflecting wider workforce changes in the industry. Secondly, to identify the difference in

the count between males and females, and to identify whether this remained fairly stable, narrowing, widening, etc. We attempted some series forecasting on this data to make predictions for future trends in disparity. The final piece of analysis on this data was a choropleth map showing the count of males and females for specific regions in the UK for 2021. The purpose of this was to create a visual of the current count of males and females in the technology industry for each region and to go beyond our current skill set. Given that we are looking at national and regional data, we also felt that a map was an appropriate visual for this data.

The data from the occupation_gender files gave us insight into the trends in specific sectors within the technology field; thus, we were able to derive deeper insight to complement the trends identified from the API data. We were able to visualise the trends within each sector in specific regions such as the East Midlands and the South West for both males and females across five years. We were also able to calculate the percentage change from this data for each sector and region, which allowed us to identify where there seems to be an increase in women in a certain sector and where there still remains a large disparity.

4 IMPLEMENTATION AND EXECUTION

4.1 Development Approach and Team Member Roles

We used a SWOT analysis to identify strengths and areas for development within our team. Tasks were also divided based on areas of particular interest. All group members contributed to the decision-making process, final work document and presentation.

- Rachel: API data extraction and analysis.
- Betty, Kamile, Safia: occupation and region data extraction and analysis.
- Monaliza: ReadMe and GitHub management, PPT slides.

4.2 Tools and Libraries

- Python
- Matplotlib
- Pandas
- Plotly Express
- NumPy
- Seaborn
- Statsmodels
- Sklearn

4.3 Implementation process: Achievements, Challenges, Decision to Change Something

- We were originally going to compare the UK to US data but felt this would make the analysis too broad.
- Extracting the API data was a significant hurdle in obtaining this data, and so when we found a way to do this, it was a significant achievement.
- Another achievement was learning how to create GeoJSON files and use these to map data.

4.4 Agile Development

- We set out tasks using cards into different categories to keep track of our project progress as can be seen in the project roadmap above. We worked in sprints to meet deadlines and split our work into iterations so that we could all be working on it concurrently.

4.5 Implementation Challenges

- One of the biggest challenges for our team was finding a time that we were all available to meet and discuss the project. Due to classes in the evenings, work commitments, and external commitments, this proved to be one of the more challenging aspects of this project.
- Extracting the API was also challenging due to it being a list of dictionaries; this required a few extra steps in data cleaning and preprocessing.

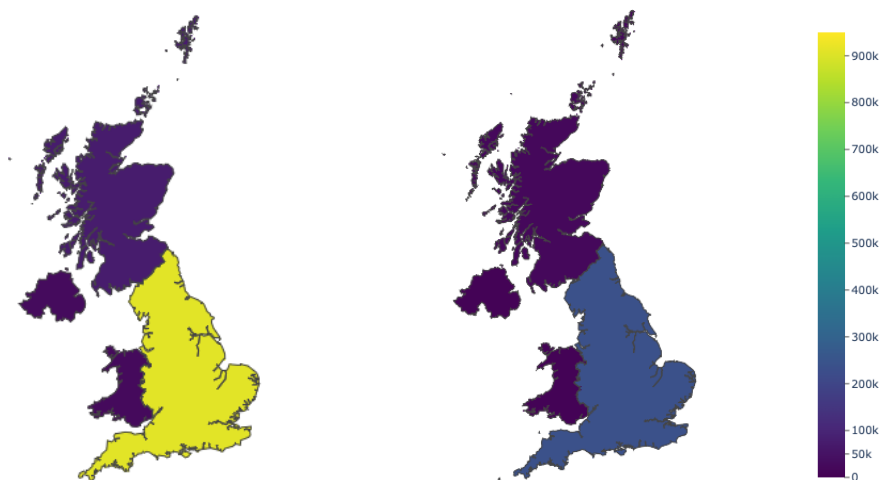
5 RESULTS REPORTING

5.1 Overall Trend in the Count of Males and Females in the Technology Industry in the UK

Our data shows that the technology industry comprises a majority of males relative to females. The gap between males and females saw a small decrease in 2018 and then an upward trend from 2018 to 2021. This difference between genders increases in line with the overall numbers of males and females increasing in technology. This indicates an overall trend in workforce numbers in this industry, but still a larger proportion of males than females. Despite this, looking at the trends in values for the difference in total males and females in the technology industry, the change from year to year is decreasing overall. This suggests that although the total difference is increasing, the change in the difference from year to year is decreasing. When looking at the regional analysis, the increase in total numbers of both males and females occurred mostly in England, contributing to the overall upward trend in UK numbers for both genders. The following maps for 2021 demonstrate an example of this spread across the UK:

Males

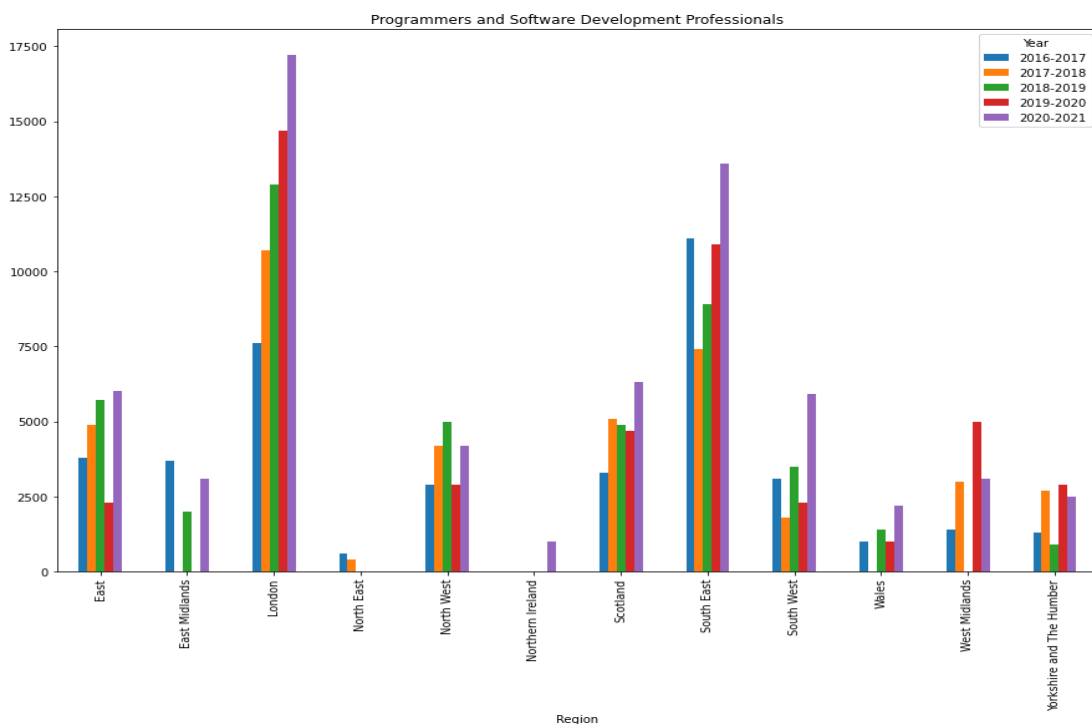
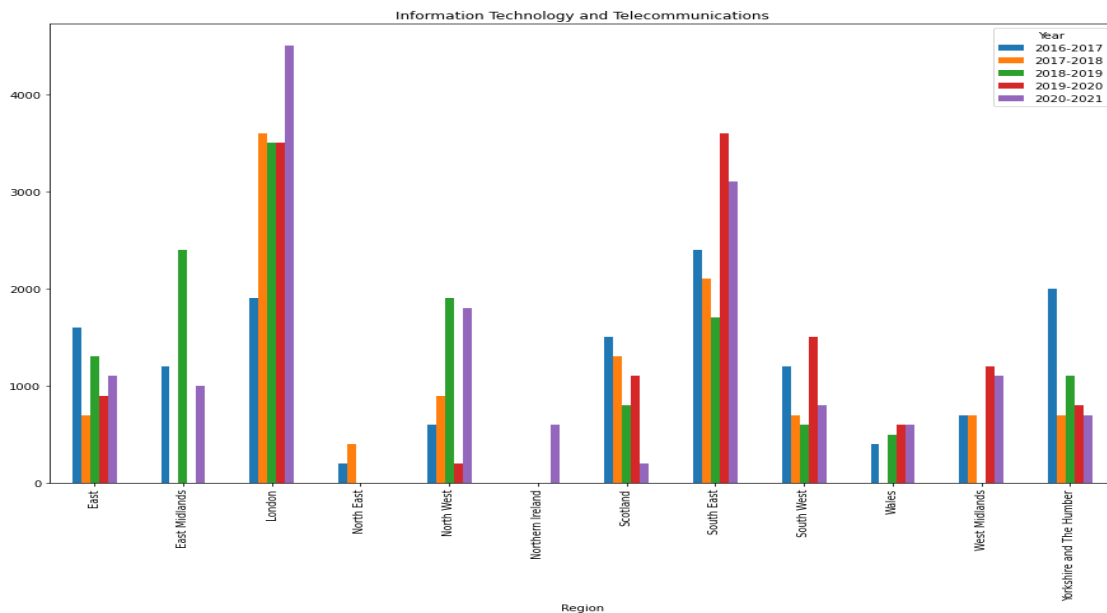
Females

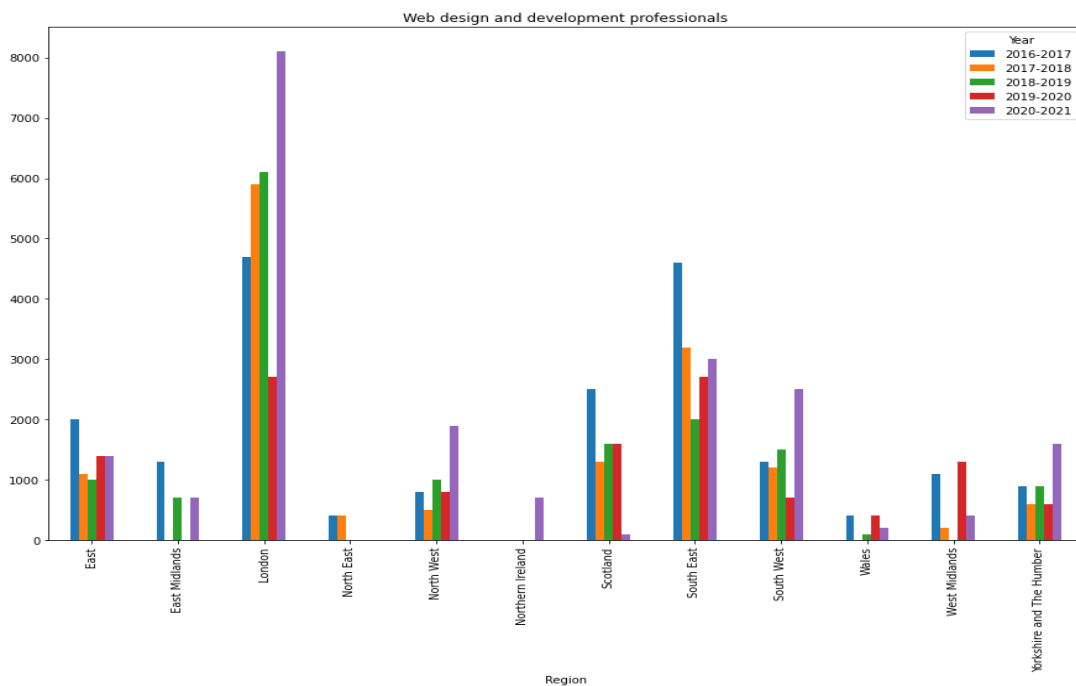
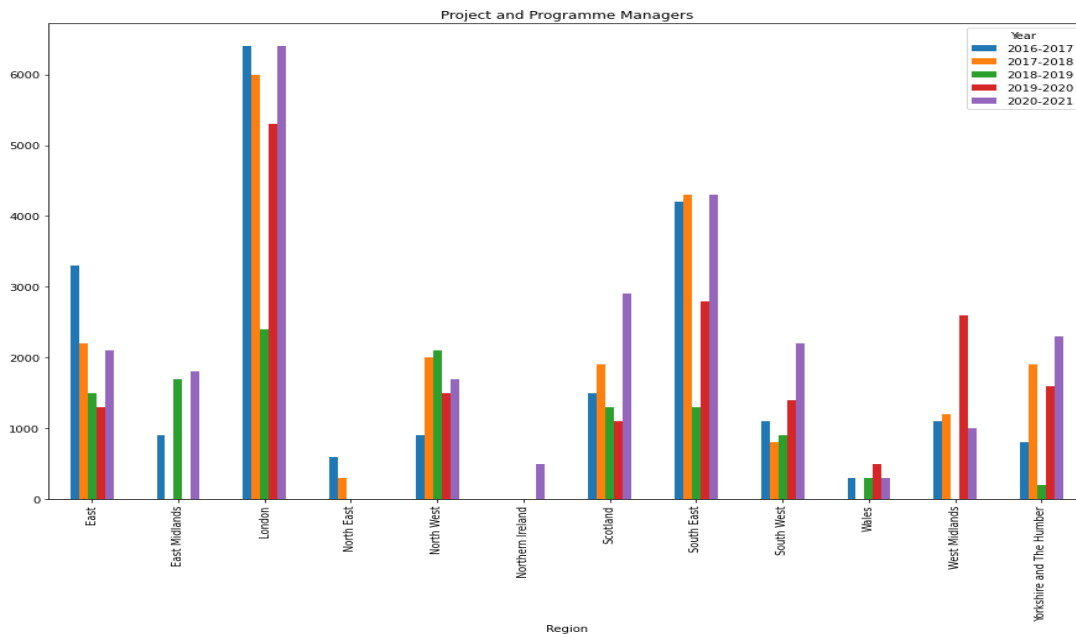


5.2 Occupation and Region Trends

When analysing the occupation gender data, the rate of increase in females differs from region to region and between sectors. In general, London and the South East have the highest number of females working in technology across the majority of the sectors, with the trends seemingly increasing from 2016 to 2021, although that increase is not consistent or entirely linear. However, these regions do have a higher population overall, so it's difficult to say if the rate of increase is actually higher or proportionate to that of other regions.

Below are some example bar charts that visualise the count of females in certain regions and certain occupations (information technology and telecommunications; programmers and software development professionals; IT project and programme managers; web design and development professionals) over 2016–2021. We can immediately see that being the largest regions, London and the South East have the highest number of females working in the four occupations. However, the overall trend is not clear. As mentioned before, the number of females working in various technology roles has *not* been increasing consistently or in a linear fashion as it greatly varies between the years. In addition, there is regional and occupational variation. We can only say that there is little evidence of a certain positive trend in the number of women in technology in the UK over the years.





6 CONCLUSION

Our results indicate that there have been few changes to the difference between the number of males and females in the technology industry in the last five years. However, the change in the difference between males and females in the UK does seem to be decreasing over time.

There have been several initiatives to encourage greater female representation in this industry. Our data suggests that either this will take several years to be reflected in the workforce numbers, or more needs to be done to encourage female participation.

7 REFERENCES

- We are tech women: <https://wearetechwomen.com/campaigns/>
- <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/employmentbyindustryemp13>
- <https://www.hntechsurvey.com/>