MANCHESTER 1824
The University of Manchester

Q-Step
A step-change in quantitative social science skills
Funded by the Nuffield Foundation, ESRC and HEFCE

DEI Data Analyst hosted by FDM

Do people in different demographic groups 'talk' about tech differently?

Alice Eddyshaw | **BA (Hons) Linguistics and German**

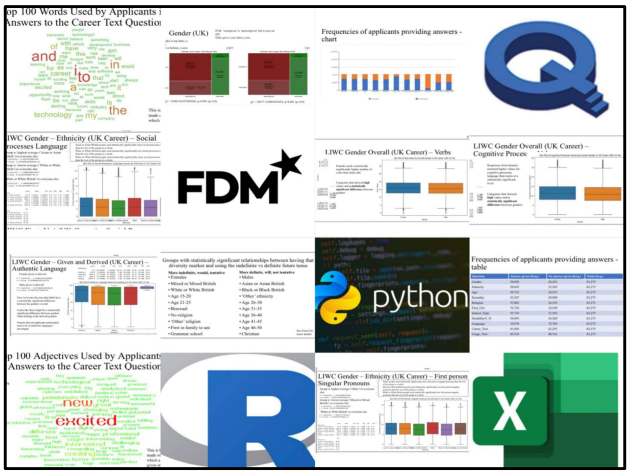

Figure 1: A selection of logos, tables, data visualisations, statistics, slides and more to represent my time as a Q-Step Data Fellow at FDM.

Overview of the Data Fellowship

In my data fellowship, hosted by **FDM, a tech recruitment company**, I analysed data given by applicants to the company's graduate programme My aim was for my analysis to go some way towards answering these questions: **'Do people in different demographic groups 'talk' about starting a career in tech differently?'** and **'What impact could this be having on DEI in tech recruitment?'** The work was based on Jonathan Young's PhD dataset. Once I'd subsetted and cleaned the data set for my analysis, there were four new datasets (2 UK, 2 US), the one I worked on most throughout the fellowship containing **10 million words** written by UK applicants in relation to starting a career in tech.

Data Analysis

From the point of having created the subsets, I produced a series of **plots** (largely mosaic and box plots), fetched **descriptive statistics** and analysed **statistical significance** of certain relationships relating to applicants' **answering/non-answering** of certain questions, **LIWC-22 (Linguistic Inquiry Word Count 22)** data and use of **definite and indefinite future tense**. To do this, I used a mixture of **Excel**, **R** and **Python**.

Findings

- In the case of all **diversity markers** that were asked about **except language**, **more people answered** the question **than did not** in the **UK** dataset. This could be due to applicants' interpretation of the question as asking for a second language, assuming that it's already clear they speak English. Alternatively, applicants could have feared discrimination on the basis of being presumed not to speak English as well as others. In the **US** dataset, **far fewer applicants gave their diversity markers** on the whole. The only two markers that more than half of the applicants gave were gender and disability, yes or no.
- Simply put, **LIWC 22 analysis** confirmed that '**yes, people in different demographic groups do talk about starting a career in tech differently**' in a variety of ways, though more investigation would be needed in order to understand why.
- The **younger** applicants (15-20 and 21-25) used a **higher ratio of indefinite future tense to definite future tense than** the **older** applicants. This could be interpreted as them being more **tentative** in their use of language. Other groups that also used a statistically significantly higher ratio of indefinite future tense to definite future tense in the UK career text dataset were **females**, **white or white British** and **mixed or mixed British** people, **bisexual** people, people with **no religion** or a **religion in the category 'other'**, people who were the **first in their family to go to university**, people who went to **grammar** or **state schools** and people who **have a disability**.

Key Skills Learnt

I learnt, built upon and consolidated **a variety of professional, research and analytical skills**, comprising both soft and hard skills.

In terms of **technical skills**, I built upon my previous knowledge and experience of **Excel** and using **R in RStudio** to **explore**, **analyse** and **create visualisations** of the **data**. I also learnt **Python** from scratch and proceeded to use it for the majority of the analysis and visualisations. I also learnt a lot about **working with large datasets**. I definitely developed my ability to **manipulate, analyse** and **filter data** and gained experience **dealing with** pieces of **data** that are partial, in some way ambiguous or potentially unreliable. Additionally, I furthered a range of **research skills** such as **developing research questions**, **designing tests** and **selecting relevant data**.

The **professional skills** that I worked on most prominently as a Data Fellow are **project management**, **adaptability** and **networking**. The project I worked on was a very self-guided, open-ended one and this provided the opportunity for me to learn to **guide** and **manage** a **complex project** independently. One example of a situation that required me to be adaptable is when RStudio on my laptop stopped being able to handle the large dataset and I decided to solve this problem by learning and using Python instead. Examples of experiences I had as a Data Fellow that enabled me to learn more about networking were: a series of **meetings with** different **professionals** in the company to get to know their roles; a **LinkedIn workshopping** session and, generally, **joining a corporate office**. All in all, I now feel more able to speak the **language of** other **networkers**.

To conclude, I certainly learnt a lot in my data fellowship and I am very grateful for this unique and fantastic opportunity that allowed me to develop myself professionally, personally and academically.

Figure 2: A word cloud of demographic groups according to ratio of indefinite to definite future tense in their written responses (UK Career Text dataset), with larger and redder groups corresponding to higher ratios.

Contact: **alice.eddyshaw@gmail.com**