UNIVERSITY OF LEEDS

Leeds Data Analytics and Society CDT

LEEDS Institute for Data Analytics

Data Analytics & Society

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Overview

UNIVERSITY OF LEEDS

- A bit about us
- Masters overview
- PhD work to date
- Progress and plans

LEEDS Institute for Data Analytics

A bit about us…

UNIVERSITY OF LEEDS

Keiran Suchak
Leeds City Council

Fran Pontin
Active Inspiration

Victoria Jenneson
Leading Supermarket

Jennie Gray
TransUnion, formerly Callcredit

Eugeni Vidal Tortosa

Annabel Whipp
Leeds City Council

Ryan Urquhart
Sainsburys

LEEDS *Institute for Data Analytics*

Masters modules

UNIVERSITY OF LEEDS

Programming for Social Science: Core Skills

Researching Society and Culture

LEEDS Institute for Data Analytics

Masters modules

UNIVERSITY OF LEEDS

Manchester: Understanding Data and their Environment

- Ethical and legal issues when using data in research
- Disclosure control
- Anonymisation
- Data pre-processing
- Data linkage
- Identification of appropriate data sources for projects and the issues associated with using them

Liverpool: Geographic Data Science in Python

- Python as a tool for data science
- Data wrangling
- Working with APIs
- Exploratory data visualisation
- Supervised vs unsupervised learning
- Modelling: prediction vs inference
- Overfitting and cross-validation
- Model predictive performance

LEEDS Institute for Data Analytics

Data Science in Python: Pokemon Battles

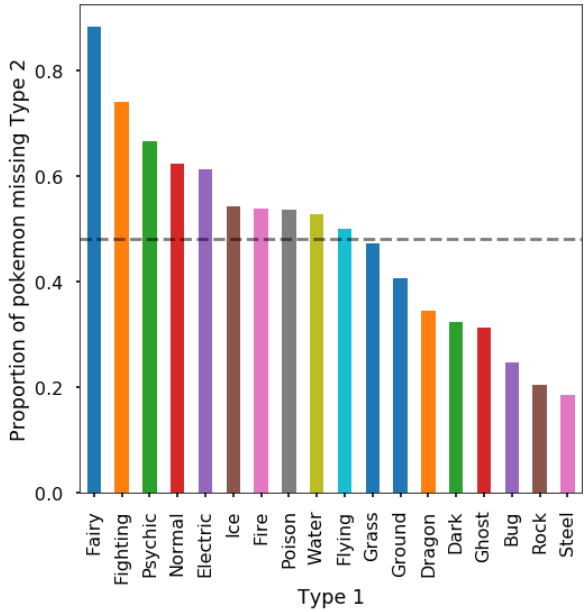
UNIVERSITY OF LEEDS

- Data pre-processing
  - Identifying missing values
  - Creating compound variables
- Exploratory data visualisation
  - Identifying trends in pokemon attributes
- Unsupervised learning
  - Clustering pokemon based on numerical attributes
- Supervised learning
  - Predicting the outcome in 1-on-1 battles between pokemon

```
[23]: def plotMissingType2ProportionBy(column):
          """Function to visualise how the proportion of pokemon that don't have a Type 2 varies."""

          # Make dataframe of proportion of pokemon missing Type 2 by chosen column
          missingColumn = missingType2[column].value_counts().to_frame()
          missingColumn['Total'] = pokemon[column].value_counts()
          # Calculate proportion
          missingColumn['Proportion'] = missingColumn[column] / missingColumn['Total']
          missingColumn = missingColumn.sort_values(by='Proportion', ascending=False)
          print(missingColumn)

          # Plot bar chart
          missingColumn['Proportion'].plot.bar(y=column,
                                               legend=False,
                                               figsize=(8, 8))
          # Plot horizontal line of mean proportion
          plt.axhline(y=missingColumn['Proportion'].mean(), color='black',
                      linestyle='--', alpha=0.5)
          # Add axis labels
          plt.xlabel(column)
          plt.ylabel('Proportion of pokemon missing Type 2')
          plt.show()
```



LEEDS Institute for Data Analytics

Masters modules

UNIVERSITY OF LEEDS

Optional

- Big Data and Consumer Analytics
- Predictive Analytics
- Data Visualisation and Analysis
- Applied Population and Demographic Analysis
- Forecasting and Advanced Business Analytics

Internship

UNIVERSITY OF LEEDS

- 10 day placement with our data partner
- Component of the Masters course
- A project is designed in collaboration
- A report summarising findings is produced and any other outputs are also supplied to the partner

LEEDS Institute for Data Analytics

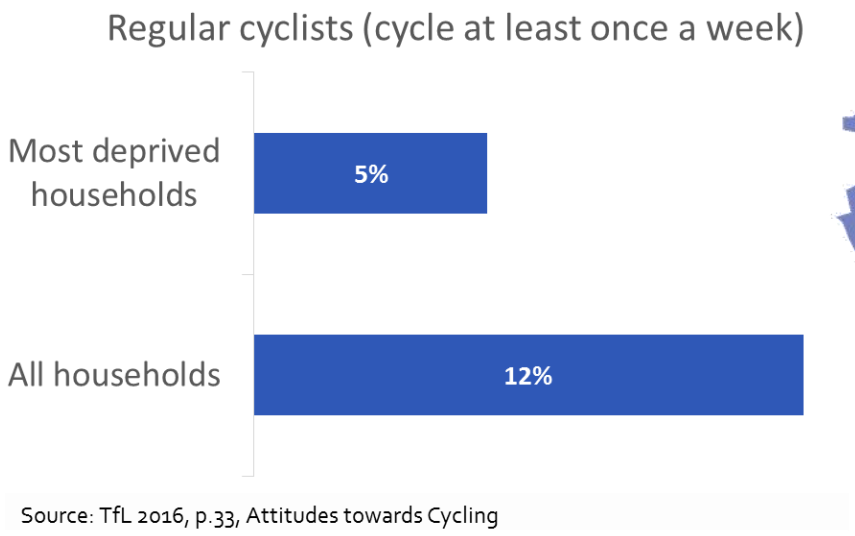
UNIVERSITY OF LEEDS

Cycling rates, cycle-friendly infrastructure, and deprivation in the South London Partnership area

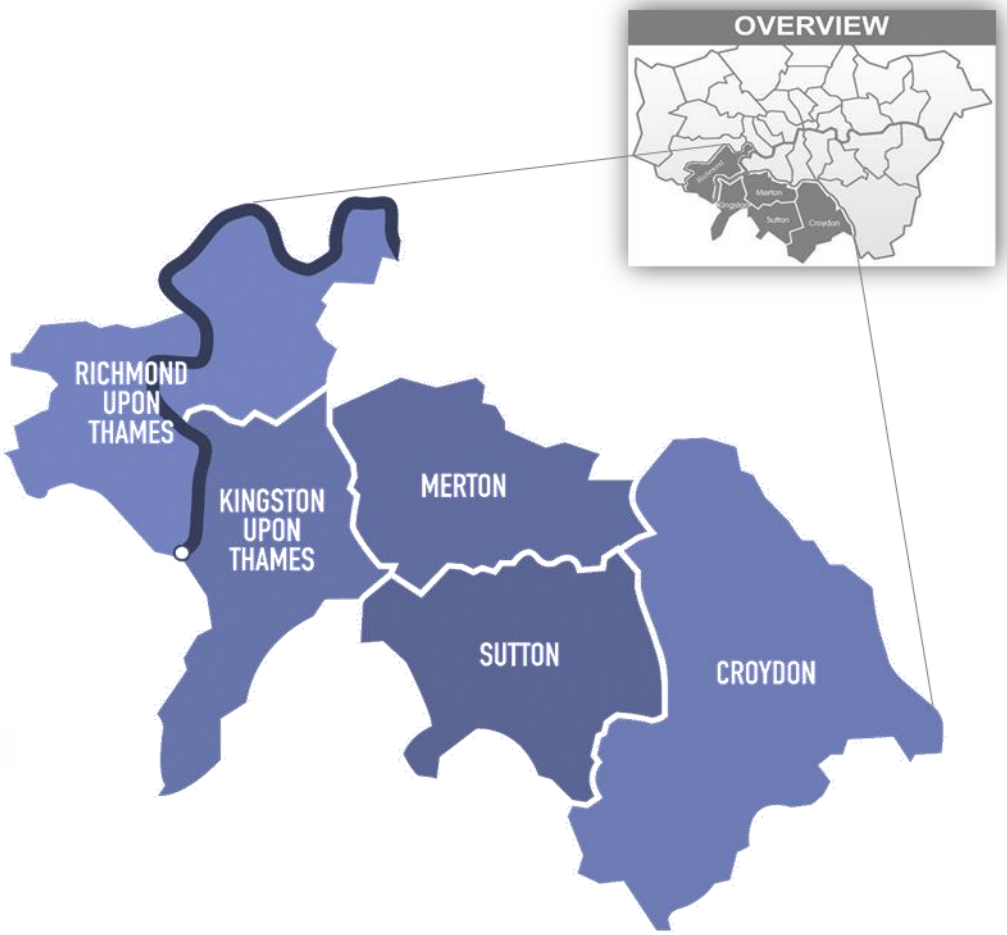
LEEDS Institute for Data Analytics

Internship: *Cycling rates, cycle-friendly infrastructure, and deprivation in the South London Partnership area*

UNIVERSITY OF LEEDS



Source: TfL 2016, p.33, Attitudes towards Cycling



Internship: *Cycling rates, cycle-friendly infrastructure, and deprivation in the South London Partnership area*

UNIVERSITY OF LEEDS

OBJECTIVES

1. Analyse the link between cycling rates, cycle-friendly infrastructure and deprivation.

2. Assess if the cycle-friendly infrastructure is equally distributed.

3. Identify areas in which future investment could increase cycling equity.

Internship: *Cycling rates, cycle-friendly infrastructure, and deprivation in the South London Partnership area*

UNIVERSITY OF LEEDS

METHOD

Units of analysis: LSOAs.

Variables:

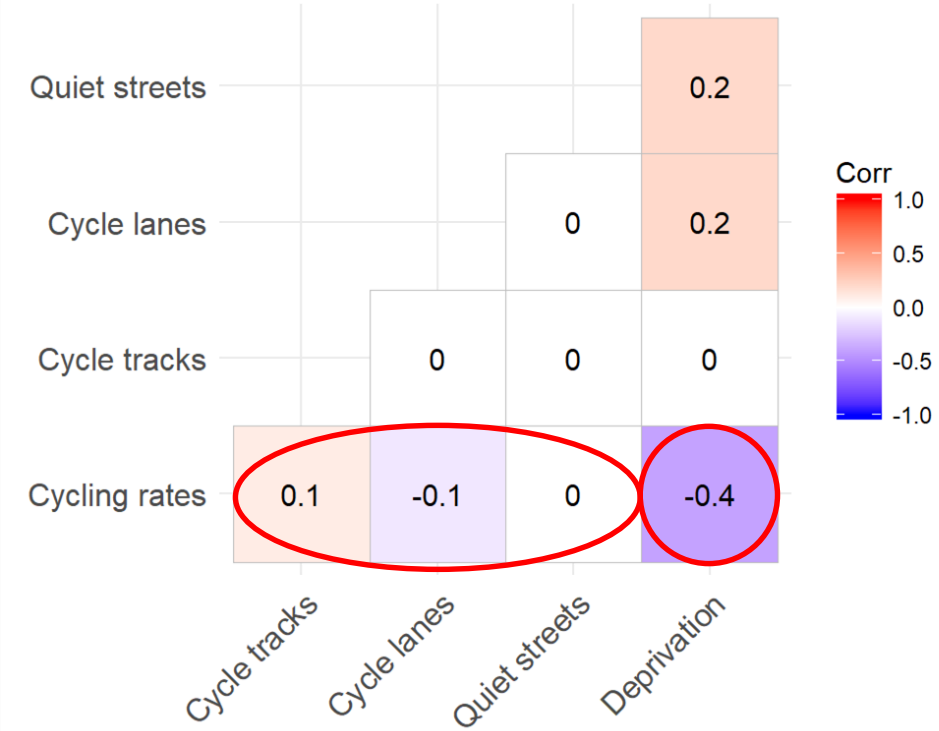
- Cycling rates: % cycling to work
- Cycle tracks: cycle tracks (km)/km^2
- Cycle lanes: cycle lanes (km)/km^2
- Quiet streets: quiet streets (km)/km^2
- Deprivation: Index of Multiple Deprivation score

Analysis: Correlograms, boxplots diagrams, and a bivariate map.

Internship: *Cycling rates, cycle-friendly infrastructure, and deprivation in the South London Partnership area*

UNIVERSITY OF LEEDS

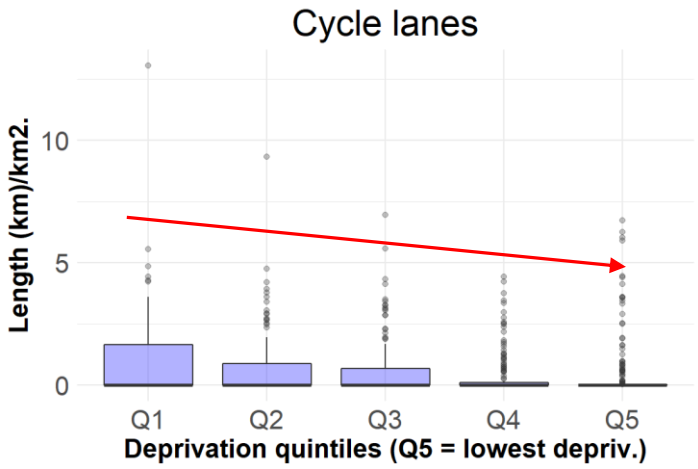
1. Cycling rates, cycle-friendly infrastructure and deprivation link

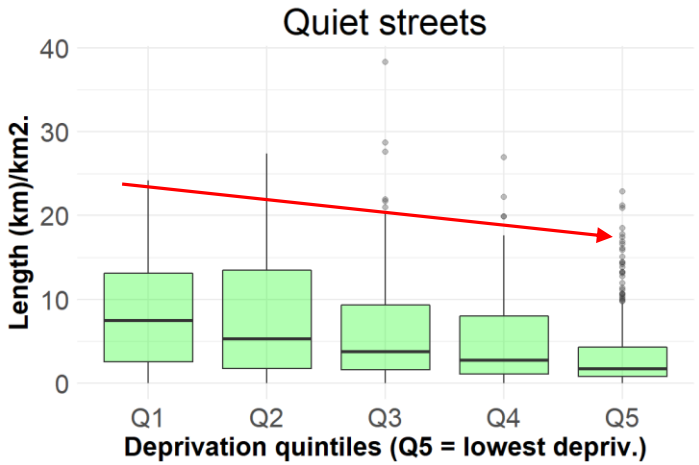


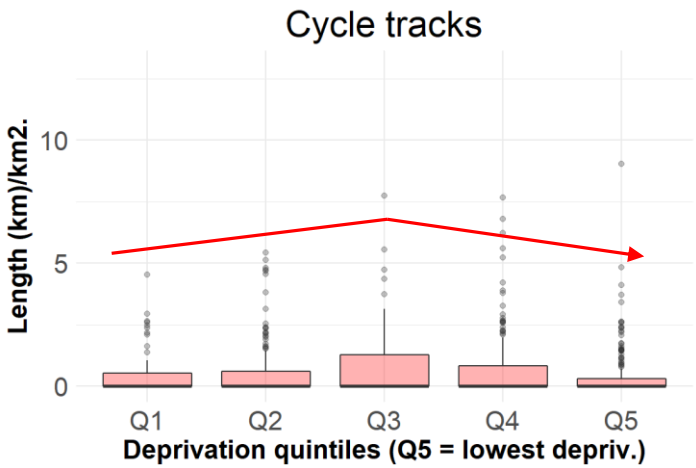
Internship: *Cycling rates, cycle-friendly infrastructure, and deprivation in the South London Partnership area*

UNIVERSITY OF LEEDS

2. Cycle-friendly infrastructure distribution assessment



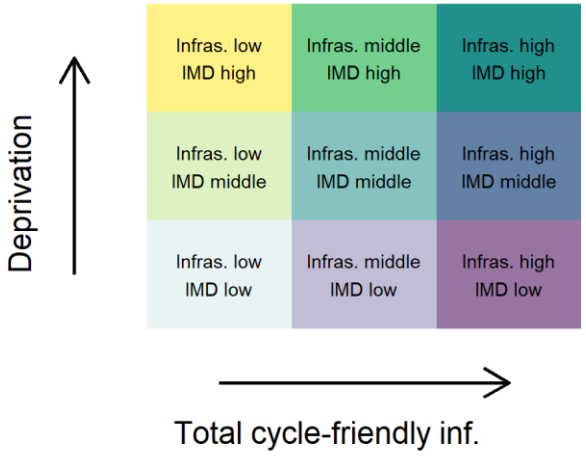


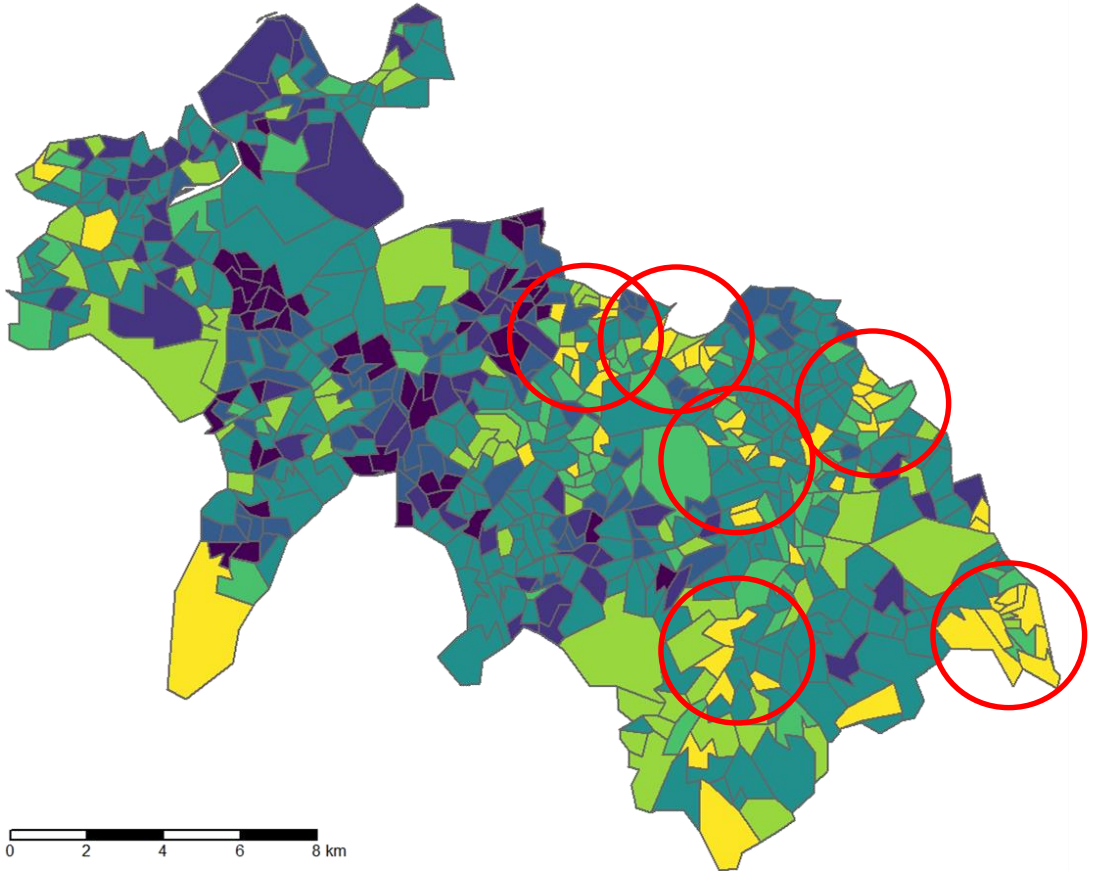


Internship: *Cycling rates, cycle-friendly infrastructure, and deprivation in the South London Partnership area*

UNIVERSITY OF LEEDS

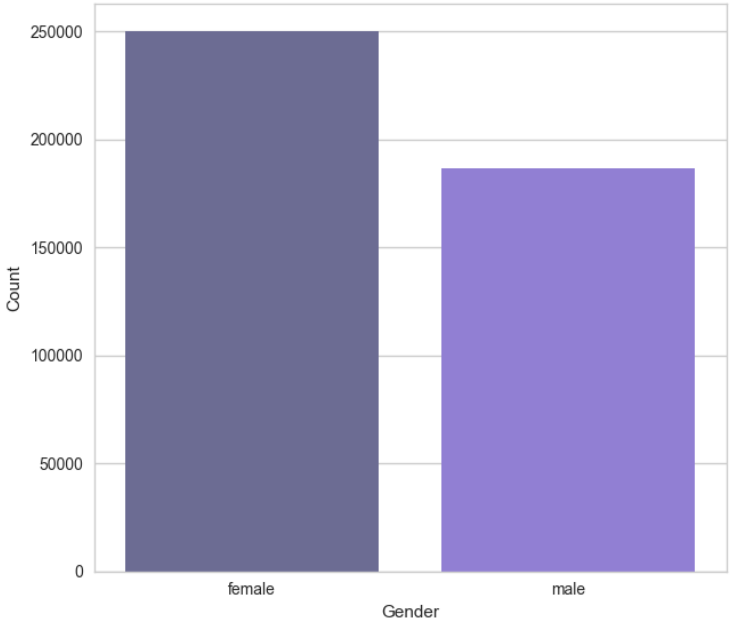
3. Identification of areas to increase cycling equity





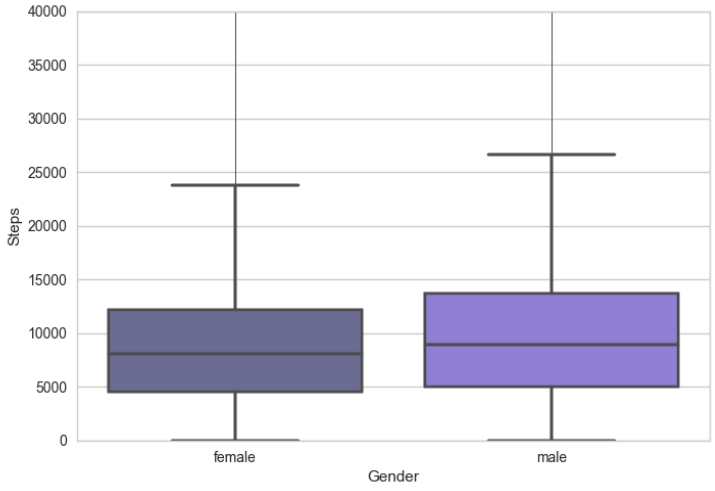
PhD work: *Demographic and Geographic determinants of physical activity*

UNIVERSITY OF LEEDS



340,000 unique users

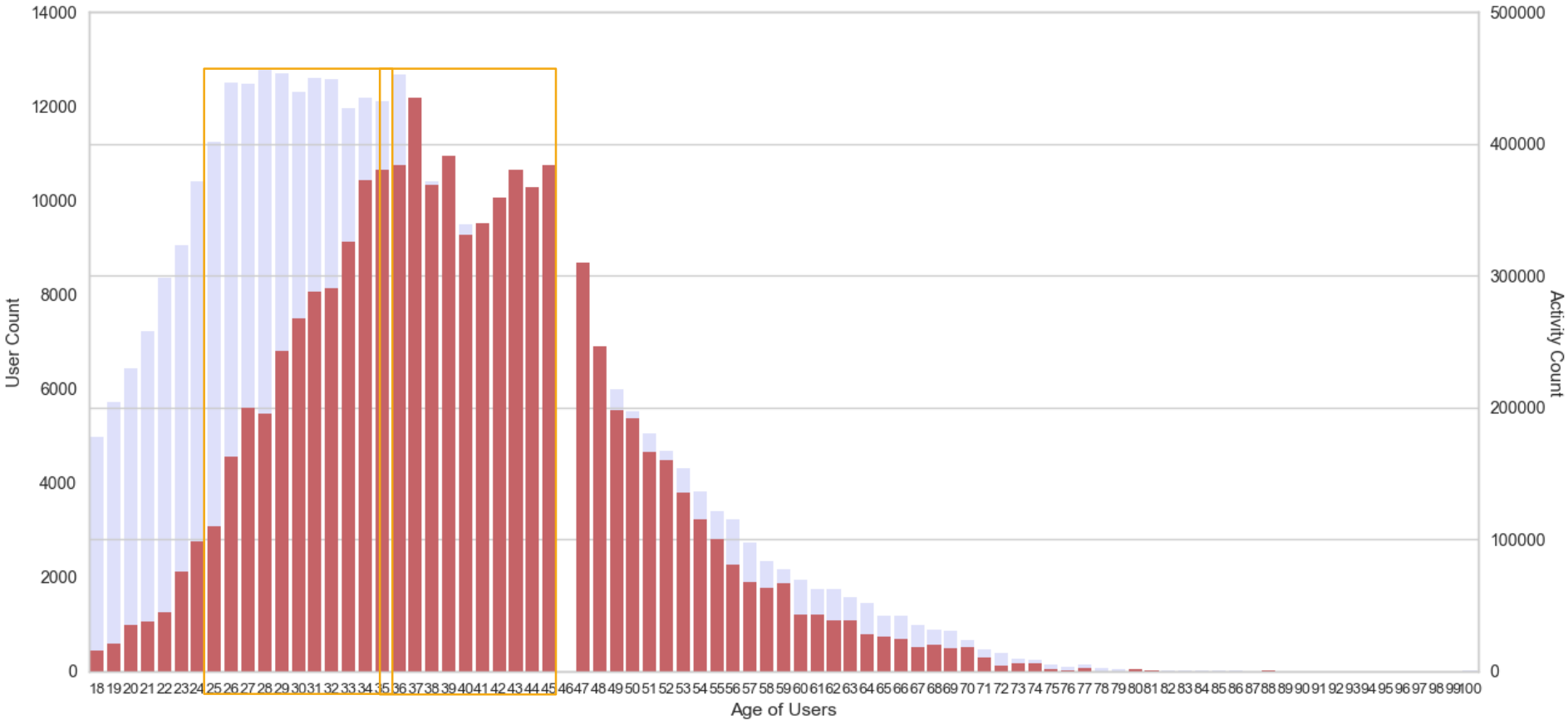
9.6 million activities



LEEDS *Institute for Data Analytics*

PhD work: *Demographic and Geographic determinants of physical activity*

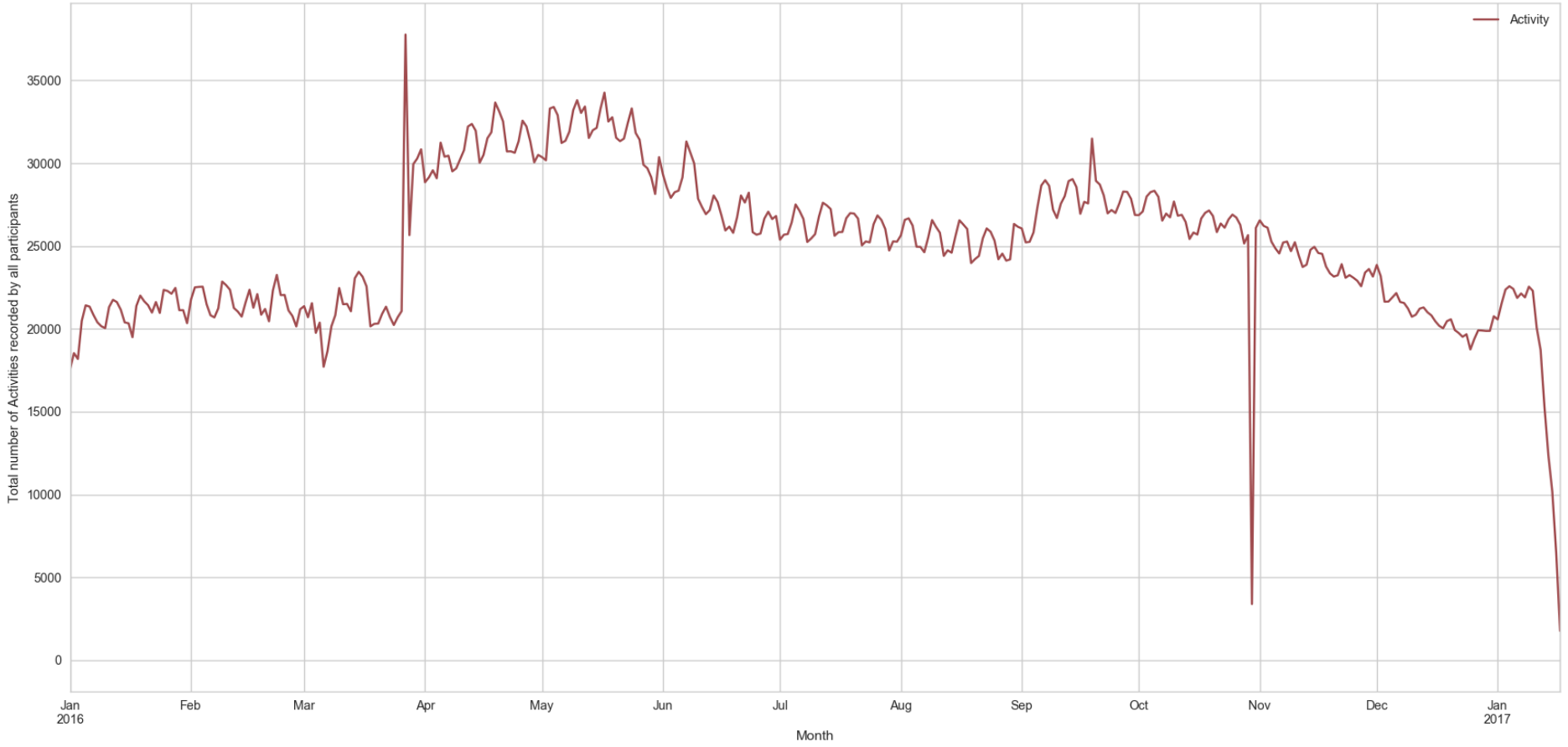
UNIVERSITY OF LEEDS



LEEDS Institute for Data Analytics

PhD work: *Demographic and Geographic determinants of physical activity*

UNIVERSITY OF LEEDS



LEEDS Institute for Data Analytics

PhD work: *Incorporating E-commerce into Grocery Sector Retail Location modelling*

UNIVERSITY OF LEEDS

- UK Grocery E-commerce industry is amongst the most developed Grocery E-commerce markets in the world.
- UK Grocery E-commerce valued at £10.4Billion in 2017 – predicted to be the fastest growing market in the wider grocery industry over the next 5 years – forecasted value of £16 Billion by 2022 (IGD, 2017).
- Online grocery sales contributed 7.3% of total grocery sector spend in 2016 (Rigby, 2017) – A figure expected to grow.
- Fundamental change in the nature of interactions between supply and demand.

LEEDS *Institute for Data Analytics*

PhD work: *Incorporating E-commerce into Grocery Sector Retail Location modelling*

UNIVERSITY OF LEEDS

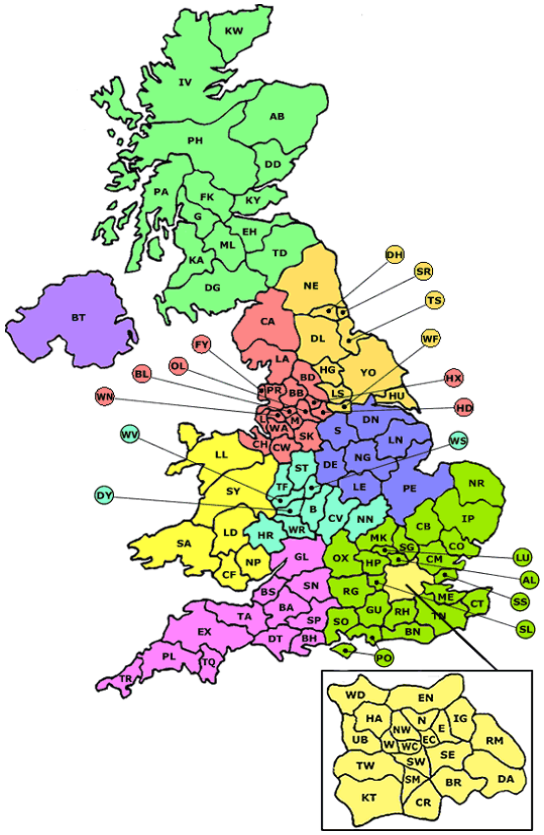
- Incorporating E-commerce interactions (Home Delivery and Click and Collect) into a spatial interaction modelling framework to support location based decision making.
- Omni-channel focus
- Research and understanding of both Consumer demand in the UK and how this driven and the Supply side dynamics of offering these services.
- The real novelty and the real challenge of the project comes in understanding and capturing the consumer interactions in a spatial interaction framework that can be used to support decision making.

LEEDS Institute for Data Analytics

PhD work: *Incorporating E-commerce into Grocery Sector Retail Location modelling*

UNIVERSITY OF LEEDS

Appraisal Areas



- Breaking down the wider estate in smaller areas containing a number of stores
- A method to monitor performance
- Each store has a service area they deliver to. The area which they deliver to is determined by the predicted demand in the area, the capacity of the store and a tradeoff between market share and profitability.
- Designing appraisal areas is very a much an art rather than a science at the moment.

LEEDS *Institute for Data Analytics*

Next Steps?

UNIVERSITY OF LEEDS

LEEDS Institute for Data Analytics

Data Analytics & Society

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL