Use of Data Science to Augment Existing Economic Statistics
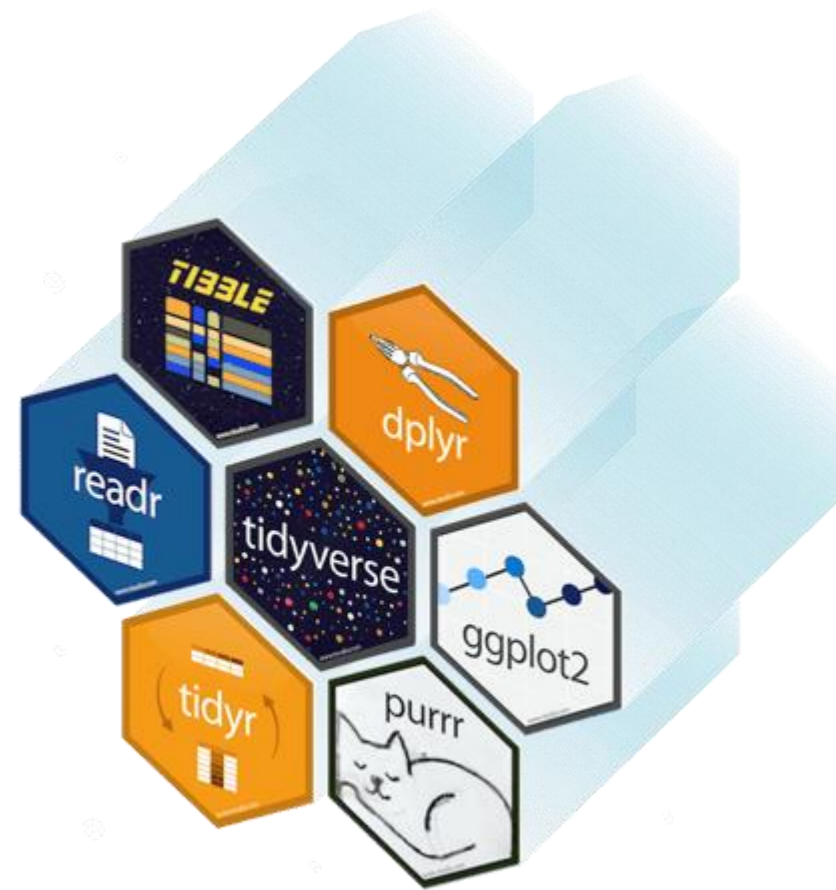
Stephen Nicholson
MSc in Big Data

Overview

- The Scottish Government Directorate for the Chief Economist wished to investigate how data science techniques could be used to augment existing economic statistics and improve the understanding of the Scottish economy.

Objectives

5 broad objectives:

1. Identify data sources to augment current data availability
2. Anything that corroborates published results
3. Modelling with the Sub-Scotland Productivity Model
4. Longer term history
5. Sub-Scotland geographies

Approach

The R programming language is being used for data representation, manipulation, and modelling.

The "Tidyverse" libraries have been central for most of the current work.

Task: Manufacturing
Historic Data Availability

- There is a lot of data from 1998 onwards from the Inter-Departmental Business Register by the Office of National Statistics (ONS)
- Insufficent data from 1946 (post-war) to 1998

The Regional Data System (RDS)

- Predates the ONS IDBR
- 1946 - 1996
- Manufacturing companies with 10 or more employees
- SIC92 classifications
- Addresses and postcodes
- 12,700 records

Current Work

Work focused on RDS data:
- Mapped to new Local authorities
- Mapped to new SIC codes
- Matched to 1998 business register
- Exploratory data analysis
- Modelling