****

***Data Science: Training and Capacity Building***

**Data Wrangling – Organising and Enabling Data**

The Industrial Strategy recognises that a major challenge facing UK businesses and industry is how best to utilise big data to improve economic performance and increase productivity. A substantial barrier to exploiting the potential offered by emerging forms of big data is the lack of a suitably trained workforce with appropriate analytical skills. Many statistical techniques used in the social sciences are also suitable for the analysis of big data in non-academic settings, however social science graduates often lack experience in applying their skills and knowledge in non-academic research domains.

This three-day workshop will provide a fast-track introduction for individuals wishing to learn how to work with data suitable for statistical analysis of business problems. Preparing and enabling data (data wrangling) is an essential aspect of undertaking data intensive statistical research. Data wrangling is highly time consuming and can be complex especially when dealing with messy data, which is often encountered in the business world.

The workshop will provide participants with a practical introduction to the fundamental concepts that characterise data wrangling and which are critical for successful data analyses.

The workshop will comprise both lectures and practical hands-on exercises in the programming languages R and Python. The final day of the workshop will be a group exercise in the form of a ‘hackathon’, where participants will be tasked with addressing an industry challenge using the techniques learned during the first two days.

As a result of attending this workshop, participants will be equipped with a range of skills that are highly desirable in industry, including the use of the R and Python programming languages, working with real-world (i.e. messy) data, and harvesting data from the web and other online sources.

Topics covered will include:

* Organising the data wrangling workflow
* Importing data sets into R
* Preparing and enabling data for statistical analyses
* Collecting data from online sources using Python
* Dealing with messy data (e.g. web data, business data)
* Using loops and functions to make the workflow more efficient
* Dealing with missing data
* Practical hands-on sessions introducing the methods and techniques outlined above.

The workshop is free to attend, and refreshments and lunch will be provided on each day.

The workshop is suitable for postgraduate social science students and early-career researchers with a good general knowledge of statistical methods and data analysis. Some experience of analysing large-scale and complex social surveys (e.g. the birth cohort studies or household panel surveys) or administrative datasets would be an advantage.

The workshop is also suitable for academic staff engaged in teaching data analysis and statistical methods in social science settings who are seeking to incorporate data science and big data analysis skills into their programmes.