EMILY M. TAYLOR PHD

+49 1724618275 ▪ Berlin, Germany
emily.michie.taylor@gmail.com ▪ linkedin.com/in/emilytaylor203 ▪ etay203.github.io

DATA SCIENTIST | DATA ANALYIST | PROJECT MANAGER

Works in Python and R, with skills in statistical analysis, predictive modelling and data wrangling

Data analyst professional who enjoys exploring and visualising data, finding patterns and quantifying insights. Highly numerate with great math and statistical skills. Experienced in communicating complex technical information to a non-technical audience. Continuously loves developing and expanding skills and experience.

PROFESSIONAL EXPERIENCE & CONTINUED SKILL DEVELOPMENT

Lead Researcher, SourcingBot, Berlin, Germany **May 2019 – May 2020**

Sourcingbot.com is an online platform that optimises the sourcing and purchasing of electronic components for manufacturing. It provides a parametric search of around 6.6 million parts combining lab measured data with manufacturer datasheets. A matching algorithm determines similar parts as replacements.

- Took ownership of entire data normalization pipeline - cleaned and restructured data of over 5 million parts
- Adapted and contributed to a mature repository of python code using git and CircleCI.
- Designed and implemented quality control processes of data from raw input to front-end.
- Coded and implemented batch processing of capacitor calculations and their graphical results.
- Wrote and ran SQL queries to match and verify 1000s of equivalent parts in the sourcingbot database on BigQuery. Became the go-to person for extracting data from BigQuery.

Self-Employed/Self-Educated/Freelance Consultant **2017 – May 2019**

John Hopkins University Data Science Certification composed of ten courses including: R Programming, Getting and Cleaning Data, Exploratory Data Analysis, Reproducible Research, Statistical Inference, Regression Modelling and Machine Learning. Project submissions can be viewed on **etay203.github.io/projects**. A few favourites were:

- Used a weight lifting exercise dataset to build three different models (a simple classification tree, a boosted predictor using the boosted trees method and a random forest predictor) in order to predict how well participants did the exercise.
- A two-part statistics project. First, a simulation of the exponential distribution in R to demonstrate the central limit theorem. Second explored the effect of Vitamin C supplements on tooth growth using the ToothGrowth data in R datasets package.
- Applied knowledge of regression to answer “Are Manual or Automatic transmission cars more fuel efficient?”. Explored the relationship between a set of variables and miles per gallon, with least squares and inference using regression of linear and multivariable models.

Engineering Consultant, Galson Sciences Ltd. (GSL), Oakham, UK **2009 –2017**

GSL provides a range of nuclear decommissioning and radioactive waste management services. Independently managed and supported a range of projects for both domestic and international clients. Developed prediction models, and coded programmes to support environmental impact assessments, for example:

- Coded, tested and implemented a bespoke computational tool – enabled the client to quantify and specify radioactive doses.
- Developed and coded a statistical analysis tool - optimized strategies for monitoring and inspection by waste store operators. The tool enabled different sites to plan and tailor their most efficient schedule for monitoring and inspecting waste canisters.
- Designed and built a large relational database from scratch - established data integrity ensuring government compliance by tracking lists of design features, events and processes. Made the process reproducible and saved the client time.
- Presented complex data to non-technical audience as easy-to-interpret graphics and reports. Demonstrated compliance with nuclear safety regulations.

Graduate and Post Graduate Student, Imperial College, London, UK **2000 –2009**

- Utilized numerous statistical techniques, including Markov Chain Monte Carlo for predicting material structures.
- Optimized high volumes of simulations by running batch calculations across different processors in parallel, maximising speed and efficiency.
- Published work in peer reviewed papers
- Presented work publically at several international conferences.

EDUCATION

2005 – 2009 PhD, Department of Materials, Imperial College London. Developed predictive models to better understand the structure and behaviour of ceramic materials for radioactive waste applications. Used high performance computing techniques to run models and calculations. Supervisor Prof. R. W. Grimes.

2000 – 2005 MEng with Upper Second Class Honours, in Materials Science and Engineering, Department of Materials, Imperial College London, with one year placement at CENIM, Madrid 2002-2003.

1992 – 1999 A levels in Maths, Physics, and Chemistry, GCSEs in 9 subjects, South Hampstead High School for Girls, London NW3 5SS.

TECHNOLOGY APPLICATIONS AND PROGRAMMING SKILLS

Markup/Programming: Python (Pandas, NumPy, matplotlib, json, scikit-learn), R (ggplot, knitr), R-Markdown, SQL, MySQL, XPath, XQuery and XSLT, Perl, Visual Basic for Applications (VBA), LaTeX, HTML, CSS

Applications: Microsoft Office, Google Drive, Git, RStudio, BigQuery,

OS/Environments: Windows, Linux, Unix, Mac

LANGUAGES

English (Native), Spanish (Intermediate), French (Basic), German (recently completed Level A2.2).

EXTRA-CURRICULAR

Co-organiser of WiMLDS Berlin since it was founded last year. Supported its rapid growth to 1000+ members, organizing monthly events in Berlin, including Berlin's first scikit-learn open source sprint. Recently hosted and moderated several online zoom meetings for large groups collaborating with a group in Paris as well as a panel discussion.

My other passions include running, cooking and eating great food, and travel