

# STEPHEN RICHER

+44(7513) 065 917 ◊ London, U.K.

[stephen.richer@proton.me](mailto:stephen.richer@proton.me) ◊ [linkedin.com/in/stephenricher](https://www.linkedin.com/in/stephenricher) ◊ [github.com/StephenRicher](https://github.com/StephenRicher)

## OBJECTIVE

---

Python programmer and data scientist currently working with the NHS to build best practise analytical workflows and machine learning models to address health inequality. Seeking a full-time position in software engineering.

## SKILLS

---

<b>Programming Languages</b>	Python (NumPy, Pandas, Scikit-learn), SQL, Git, Bash, C++, Java
<b>Data Science</b>	Supervised Machine Learning (Decision Trees, Regression), Tableau
<b>Cloud Platforms</b>	Microsoft Azure
<b>Soft Skills</b>	Communication, Critical Thinking, Problem Solving, Adaptability

## EXPERIENCE

---

<b>Data Scientist Intern</b> NHS England	Jun 2022 - Present <i>London, U.K</i>
---	--

- Built and deployed an AutoML pipeline, implementing CatBoost and Logistic Regression, for predicting health-care appointment non-attendance at regional NHS trusts. Available at: [github.com/nhsx/dna-risk-predict](https://github.com/nhsx/dna-risk-predict)
- Developed ETL data engineering pipelines for processing and aggregating healthcare and demographic data from multiple public source.
- Collaborated with NHS stakeholders and analysts to design and implement reproducible analytical workflows for studying healthcare inequalities.
- Utilised a variety of statistical approaches and hypothesis tests, including permutation testing and partial correlation, to robustly analyse complex and biased real-world healthcare data with numerous confounding variables.

<b>Doctoral Researcher, Biology &amp; Mathematics</b> University of Bath	Oct 2018 - Jun 2022 <i>Bath, U.K</i>
---	---

- Conducted independent research and developed novel computational tools to understand how the structure of DNA impacts cellular function.
- Developed HiCFlow, a user-friendly analytical workflow for conducting robust and reproducible bioinformatics analysis. Available at: [github.com/StephenRicher/HiCFlow](https://github.com/StephenRicher/HiCFlow)
- Supervised final-year undergraduate students completing projects in computational biology.
- Presented research at national and international conferences.
- Awarded best final-year PhD presentation at the Bath Departmental Research Day.

<b>Research Software Skills Trainer</b> University of Bath, Doctoral College	Apr 2019 - Jun 2022 <i>Bath, U.K</i>
---	---

- Developed and delivered materials to train doctoral students in research software engineering best practises.
- Curriculum: Python programming and style (PEP8), version control, testing and continuous integration. Available at: [github.com/Research-Software-Skills-Bath](https://github.com/Research-Software-Skills-Bath)

<b>Technical Support (IT Services)</b> University of Manchester, Hornet	Oct 2017 - Sep 2018 <i>Manchester, U.K</i>
--	---

- Provided IT support, education and security guidance for residential students.

<b>Study Coordinator</b> MAC Clinical Research	Oct 2016 - Jul 2017 <i>Leeds, U.K</i>
---	--

- Oversaw a phase 3 clinical trial investigating a novel treatment for Alzheimers disease.

## EDUCATION

---

**PhD, Biology & Mathematics**, University of Bath Expected 2022  
Mathematical and bioinformatics-based tools to explore the impact of gene editing on the geometric principles governing the 3D structure of the genome.

**MSc, Bioinformatics & Systems Biology (Distinction)**, University of Manchester 2017 - 2018

**BSc, Biology (First)**, University of Bath 2011 - 2015

## PROJECTS

---

**Real World Data Validation with *validatum*** Lightweight, user-friendly Python module to automatically detect and flag suspicious data inconsistencies and common data entry errors within real world data.

Available at: [pypi.org/project/validatum](https://pypi.org/project/validatum)

**Secure Data Encryption with *datasafe***. Python based command-line utility for encryption of text files and encryption of Pandas DataFrames while preserving datatype.

Available at: [pypi.org/project/datasafe](https://pypi.org/project/datasafe)

**Data Science Best Practises - The Titanic Dataset**. Wrote a popular Kaggle notebook, describing best practise approaches for performing supervised tabular classification using Scikit-learn.

Available at: [Kaggle - Titanic Data Science](https://www.kaggle.com/leedsclinicaltrials/titanic-data-science)