LUKE MCGUINNESS

As PhD candidate and researcher, I have derived insights from large-scale electronic health record datasets, created reproducible reports and visualisations to communicate with stakeholders, and built research tools that are used by thousands of researchers globally each month.

As a freelance data scientist, I have worked with clients across a range of industries to create data pipelines and reporting dashboards, deliver market insights, and inform business decisions.

SKILLS

Skills: Data analysis (*experimental design, hypothesis testing, data ETL, model building, regression, feature selection, bias assessment*), Data visualisation, Pipeline Design, Automation, CI/CD, Web-scraping

Technologies: R, SQL, Git, JavaScript (D3.js), Python, HTML, CSS, Plotly/Tableau, Spark

CONTACT DETAILS

🔗 lukemcguinness.com
🐙 github.com/mcguinlu
in linkedin.com/in/mcguinlu
📞 +44 7732 539 343

🎓 EDUCATION

PhD (Candidate), Epidemiology/Population Health Sciences
Present — 2019
University of Bristol 📍 Bristol, UK
- My thesis examines the relationship between cholesterol and dementia risk, using advanced analytical and causal inference methods to interrogate large datasets of electronic health records and scientific publications.
- *Thesis title:* "Determining the causal effects of lipid levels on risk of dementia: a triangulation of the available evidence and individual participant data meta-analysis".
- Dissertation written as a reproducible RMarkdown document for push-button replicability.

MSc, Epidemiology
2017 — 2016
London School of Hygiene & Tropical Medicine 📍 London, UK
- Assessed the bias inherent to analyses of electronic health record (CPRD) databases by reviewing validations of codelists used to identify dementia cases.

BA (Mod.) in Genetics
2016 — 2012
Trinity College Dublin 📍 Dublin, Ireland

💻 EMPLOYMENT (RESEARCH)

Senior Associate in Research Synthesis
Present — 2021
Bristol Medical School 📍 University of Bristol
- Responsible for the delivery of a government-funded project, with a budget of £300,000, to examine the effect of cholesterol and statin use on dementia risk.
- Working closely with relevant stakeholders, created and led a global collaboration to extract, clean and analyse multiple existing datasets (>3 million participants).
- Involved in a critical review of machine-learning systems for the automatic extraction of data from scientific articles, helping to define best-practice guidelines for the application of ML to my field.
- Supervised two advanced degree students, focusing specifically on developing their data science skills.

Associate in Research Synthesis
2021 — 2016
Bristol Medical School 📍 University of Bristol
- Led a team of three to develop and maintain a data ingestion and classification pipeline to screen scientific literature produced in response of the COVID-19 pandemic, saving our wider team 10 hours of manual work per week.
- Developed expertise in the assessment of bias in analytical studies, recognised by invitations to deliver training on this topic to international audiences.
- Created several research-orientated open-source R packages (and associated Shiny web interfaces), used by >1200 researchers monthly to save hundreds of research-hours and improve the quality/accuracy of published research.
- Automated a machine-learning classification pipeline to identify new scientific publications related to my thesis project, using sciBERT, Python and R.

💼 EMPLOYMENT (INDUSTRY)

Data Scientist
Present — 2019
Freelance 📍 Bristol, UK
- Project: market-scoping exercise for a large pharmaceutical company
  - Contracted to assess a new dissemination method for clinical trial results, resulting in a 20% increase in engagement.
  - Involved both static (`rvest`) and dynamic (`RSelenium`) web-scraping, data munging and analysis, and preparation of a reproducible report and presentation for senior executives.

- **Project: Dashboard/web-app creation for an international charity**
  - Contracted to create an interactive dashboard for an international substance abuse charity, enabling >500 policymakers/month to customise the analysis to their region's circumstances by changing model parameters.
  - Involved extensive communication with client team, including advice on how to prepare the original analysis for deployment into production, and development of a Shiny web application.
- **Project: Application management pipeline for an international consultancy**
  - Contracted to create a pipeline for the processing of grant applications, improving the quality of assessed applications and reducing response-to-applicant time by 50%.
  - Involved database creation and maintenance, extraction of free-text data from application forms, quality assurance checks (e.g. field-completion rate), and creation of a dashboard of summary statistics for the project lead.

2015 ● **Performance Improvement Intern**

EY Financial Services Advisory Team 📍 Dublin, Ireland

- Supported a number of teams throughout my internship, working on both internal and client-facing projects and engagements.
- **Project: Customer retention analysis**
  - Adopted genetic network tools to analyze client relationship data
  - Implemented several relationship management recommendations based on this insight.

⧉ TEACHING EXPERIENCE

Current | 2021 ● **Assessing bias in randomised controlled trials**

Evidence Synthesis Ireland 📍 Online

- Invited to deliver a two-day training course on assessing bias in randomised controlled trials. Subsequently engaged to deliver a "Train-the-trainers" session on this topic.

Present | 2021 ● **Introduction to Data Visualisation and Web Applications in R**

University of Bristol 📍 Bristol, UK

- Course co-lead, responsible for designing and delivering teaching on data science topics to a mix of internal and external attendees.
- Course runs twice per year, and is regularly over-subscribed by 50% or more.

Present | 2020 ● **Statistics Module, MSc in Epidemiology**

University of Bristol 📍 Bristol, UK

- Responsible for delivering teaching to undergraduate students on best practice in the use of statistical software (R / STATA) for research.

Present | 2020 ● **Introduction to R** 🔗

University of Bristol 📍 Bristol, UK

- Developed and delivered lectures on a range of data science topics (data visualisation, literate programming and web applications).

⬈ OTHER WORK & INTERESTS

Data science communication: I enjoy writing about data science on my blog, Hexations. 🔗 Example posts include using promotional blurbs to predict if films will pass the Bechdel test using machine learning, 🔗 and creating a unique website banner for each visitor using *D3.js*. 🔗

Other interests: Running, LEGO and really terrible movies.