

# Nic Crane

Email: [thisisnic@gmail.com](mailto:thisisnic@gmail.com) GitHub: [github.com/thisisnic](https://github.com/thisisnic)

---

## Summary

A motivated and determined individual with experience delivering production-grade data science code in both R and Python. A passionate communicator, skilled in conveying knowledge to both domain experts and non-technical audiences. A quick learner with entrepreneurial spirit, and the ability to apply existing skills across a wide range of contexts.

## Work

*March 2021 - Present*      Voltron Data - Software Engineer

- Contributed to the development and maintenance of Apache Arrow.

*July 2018 – March 2021*      Unai – Senior Consultant/Data Scientist

- Developed a tool for identifying subgroups in Twitter communities using clustering techniques on structured and unstructured data, using Python libraries such as scikit-learn and igraph, which reduced analysis time from several hours to minutes.
- Created a complex multi-step NLP pipeline in Python for conducting aspect-based sentiment analysis on TripAdvisor reviews. Utilised modelling and NLP techniques such as Conditional Random Fields, fuzzy matching, and a pre-trained Word2vec model, improving initial model performance on a small dataset from 40% accuracy to around 90%. Successfully deployed analytics pipeline on AWS using Docker and Pulumi (infrastructure-as-a-service tool) to automate repeated deployment steps.
- Developed a large scale Shiny app containing advanced visualisations. Responsible for writing PostgreSQL functions to supply data for the application. Co-designed a scaled-up version of the app which is currently being successfully rolled out within the NHS.
- Delivered a presentation about this work at RStudio Conference in January 2019 (<https://resources.rstudio.com/rstudio-conf-2019/the-futures-shiny-dashboards-for-genomic-medicine-in-r>). Used acquired knowledge to shape internal best practice guidance on R and Shiny in production environments.

*May 2016 – July 2018*      Mango Solutions – Data Science Consultant

- Consulted on a number of projects involving developing R packages and Shiny applications, developing in-depth knowledge in these areas, and gaining experience of using a wide selection of R packages. Successfully co-authored “Developing R Packages” course, which is now live on DataCamp, and has very positive course ratings.
- Created and delivered workshops and training courses on a range of topics, ranging from introductory to advanced R and machine learning. Delivered private training courses and conference workshops internationally, consistently receiving excellent feedback from attendees.

- Developed an automated static code analysis package and application in R, which has been used to guide part of a large-scale redevelopment drive at the Office for National Statistics. Independently carried out an in-depth needs analysis and scoping exercise, involving use of this software, for a large organisation in the financial services industry which led to agreement of a much larger subsequent SAS to R and Spark conversion project.
- Developed a Shiny application for a client in the pharmaceutical industry to combine data from multiple sources, automate literature searches, and perform statistical analyses, resulting in enthusiastic adoption by the end users and a later project to further extend its functionality.

*July 2015 – May 2016*                      IBM - Smarter Workforce Specialist

- Utilised analytical expertise to modify and create dummy datasets, resulting in sellers delivering compelling product demonstrations.

*Oct 2011 – March 2015*                      Lancaster University – Graduate Teaching Assistant

- Engaged with postgraduate students on a statistics course, explaining complex statistical topics in simple terms, ensuring a 100% course pass rate for all students during my time in this role.

## **Education**

*2011 – 2015*                                      Lancaster University   PhD Applied Social Statistics

Examining whether the cognitive biases which affect reasoning can be reduced or eliminated. I analysed this data with advanced statistical techniques such as multilevel modelling and signal detection theoretical analyses, and translated the results into practical recommendations.

*2010 – 2011*                                      Lancaster University   MRes Applied Social Statistics

*2009 – 2010*                                      Lancaster University   MSc Psychological Research Methods

*2006 – 2009*                                      Lancaster University   BSc Psychology

## **Technical Skills**

- Analytics tools: R, Python, SPSS, SQL, Excel
- Programming languages: Python, Java, PHP
- Technologies - git, subversion
- Platforms – Windows, Unix