

Careers in Data Science

What do data scientists do and how do they do it?

1 July 2022

Nicola Rennie



About me

- BSc(Hons) Mathematics
- MRes Statistics and Operational Research
- PhD Statistics and Operational Research
- Data Scientist @ **Jumping Rivers**



A typical week

- Data science consultancy
- Internal projects
- Training courses
- Outreach work e.g. conferences

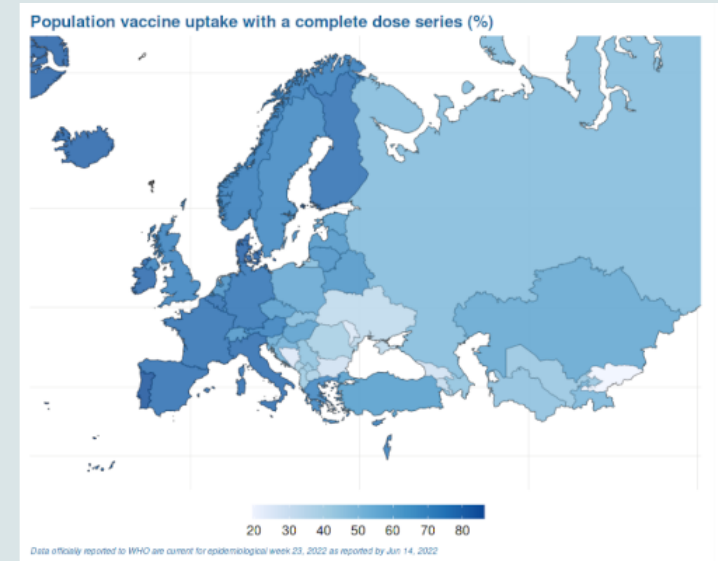
Projects

Nicola Rennie



Dashboard development

- Developing dashboards to display data for monitoring
- Dealing with data updates
- Maintenance of dashboards
- Deploying dashboards



Blog

Statistical modelling

- Statistical advice
- Reviewing potential methods
- Building models
- Scoping projects

The Project

Achieving sustainable development of fisheries is absolutely critical to the overall health of the world's oceans.

The UN's mandated work includes effective, quantitative assessment of fish stocks, which provides data-driven input for monitoring the health of fisheries, as required by Sustainable Development Goals Indicator 14.4.1.

The Food and Agriculture Organization (FAO) plays an important role in fisheries governance, it:

- collects and disseminates data and information;
- acts as a forum for states to develop and adopt treaties and guidelines;
- monitors the development and sustainability of the world's fisheries;
- and, provides policy and technical guidance to member states.

In partnership with Jumping Rivers, NU Solve is supporting the FAO in its endeavours to better monitor the world's fishing stocks.



FAO and NU Solve

Programming and implementation

- Writing code to implement models
- Reviewing code
- Upgrading from e.g. Excel

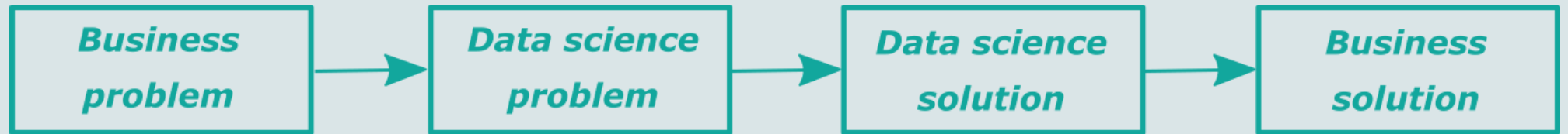


Skills

Nicola Rennie



Problem definition



Data visualisation

Visualise your data at the start...

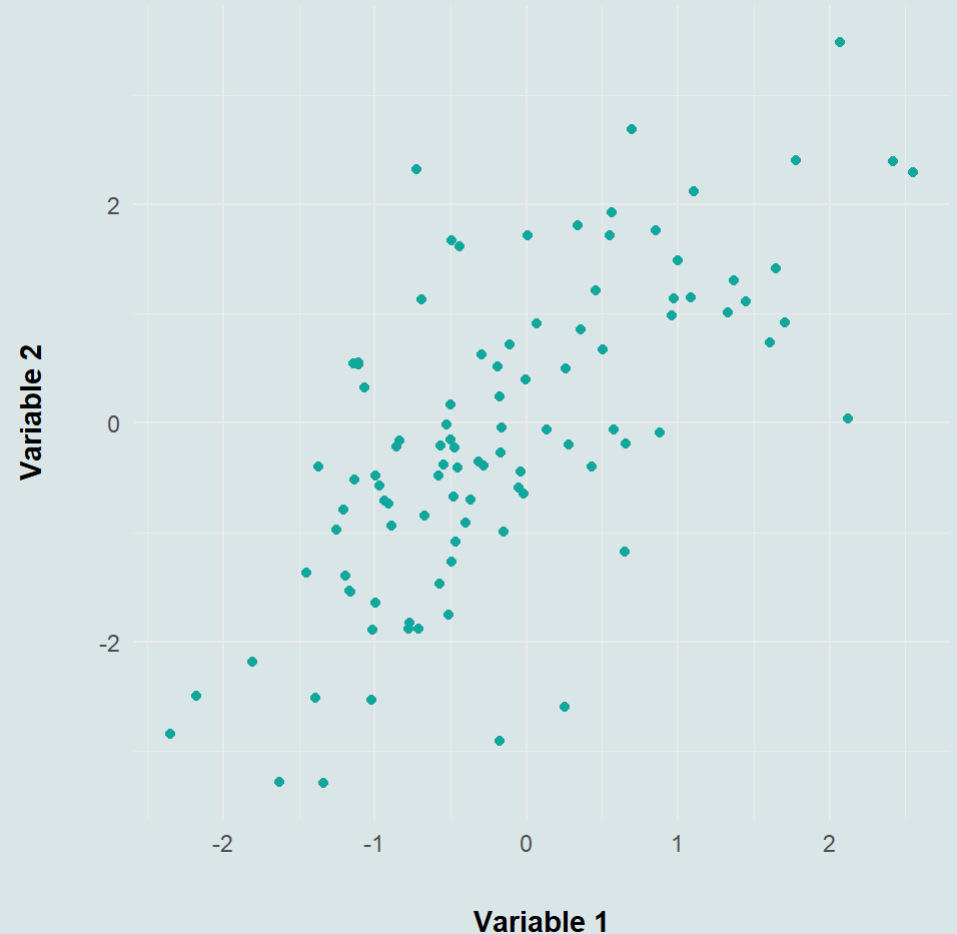
... and at the end.



Data visualisation

Initial data exploration:

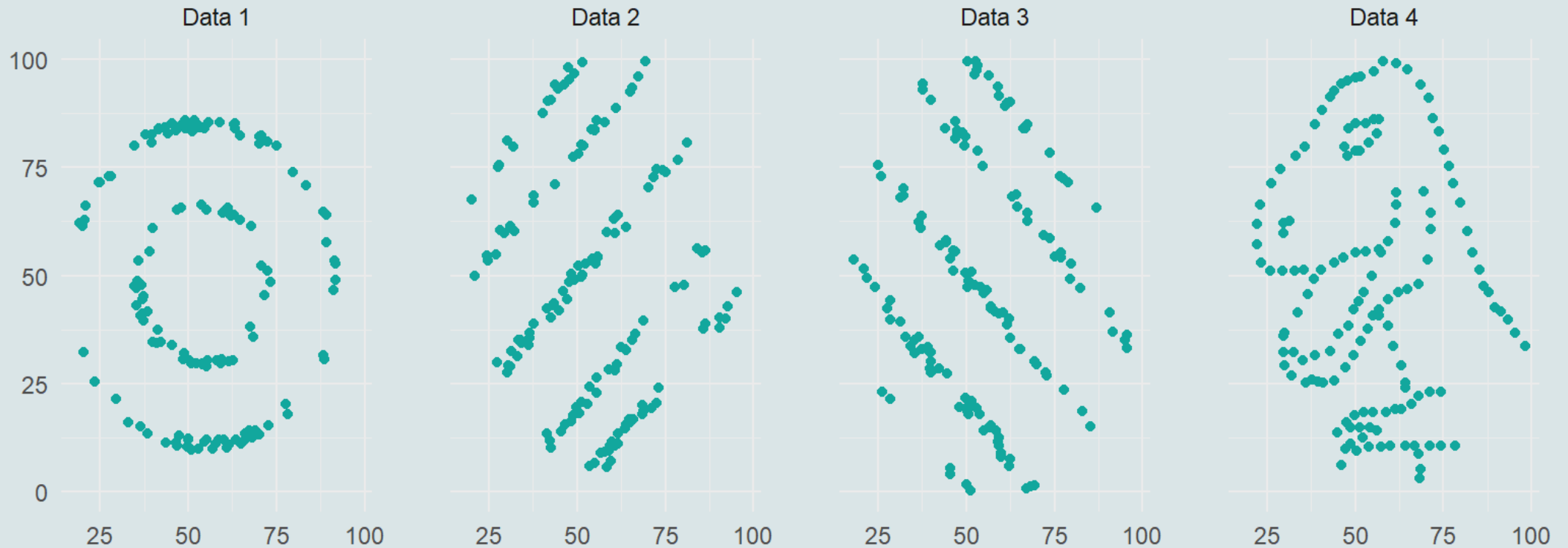
- Usually *standard* plots
- Check model assumptions
- Look at relationships between variables
- Why not summary statistics?



Data visualisation

dataset	mean_x	mean_y	std_dev_x	std_dev_y	corr_x_y
Data 1	54.26873	47.83082	16.76924	26.93573	-0.0685864
Data 2	54.26588	47.83150	16.76885	26.93861	-0.0686092
Data 3	54.26785	47.83590	16.76676	26.93610	-0.0689797
Data 4	54.26327	47.83225	16.76514	26.93540	-0.0644719

Data visualisation



Data: {datasauRus}

Data visualisation

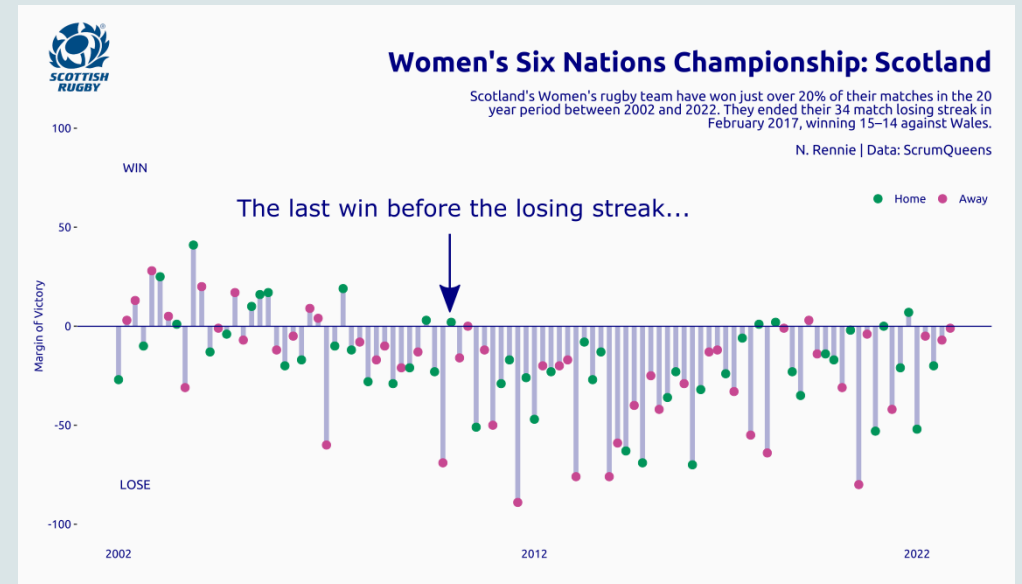
Equations look scary...

... plots not so much.



Data visualisation

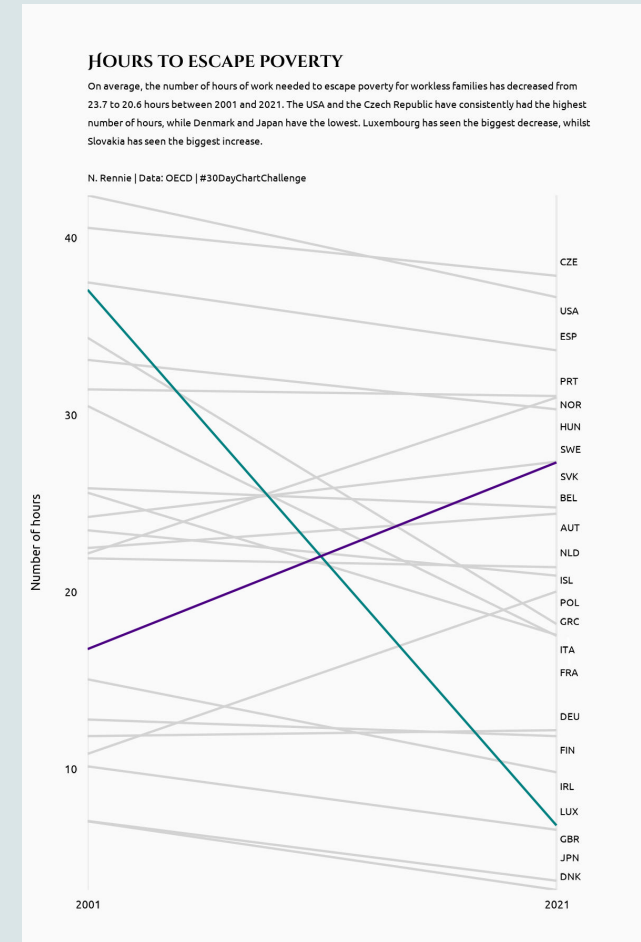
Use annotations



Data visualisation

Use annotations

Colour sparingly

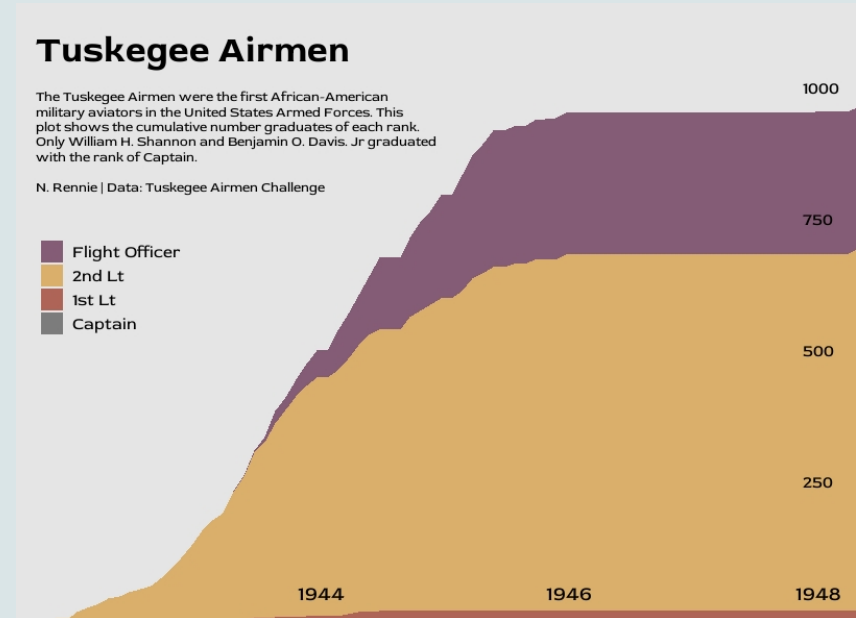


Data visualisation

Use annotations

Colour sparingly

Keep it simple



Statistical methods

- Hypothesis testing
- Time series analysis
- Monitoring
- Modelling and predicting



Statistical methods

- Start simple
- Awareness of methods (not necessarily experience)
- If you know a method, you also need to know it's assumptions

Biggest differences in industry

- General vs tailored solutions
- Explainability sometimes matters as much as statistical performance
- Faster deadlines

Questions?

- Twitter: [@nrennie35](#)
- GitHub: github.com/nrennie
- Website: nrennie.rbind.io