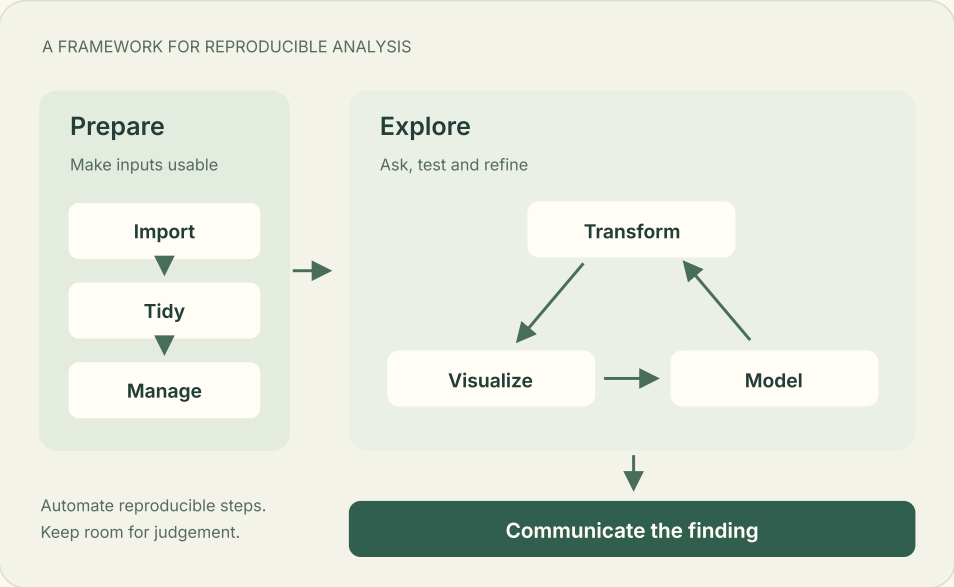
An Improved Data Exploration Pipeline

I argued that reproducible preparation gives analysts more time to explore their data and communicate useful findings.

Edited excerpt from the paper · pages 2–4

Preparation · exploration · communication



Framework redrawn from Figure 1 of the paper: prepare the inputs, explore through a repeating cycle, then communicate the finding. No operational data is shown.

01 / RETRIEVE

Make preparation reproducible

Use programmatic access and tidy structures to make retrieval and transformation reproducible across collection cycles.

02 / MANAGE

Keep the rules with the data

Use structured storage, explicit relationships and verification rules to support consistent calculations and readable outputs.

03 / DEVELOP

Invest in exploration

Choose tools for the problem, assess the process and keep learning so more effort can go into analysis and communication.

> The paper describes programmatic retrieval and structured data management already in use, then argues for continued investment in tools, learning and exploration.

Condensed and edited from *An Improved Data Exploration Pipeline*, sections 2–6. The paper cites R for Data Science for its exploration framework.

Thomas Eapen
nowlabs.net/data-officer

Selected argument and recommendations from the paper.

Contact me to discuss this work.