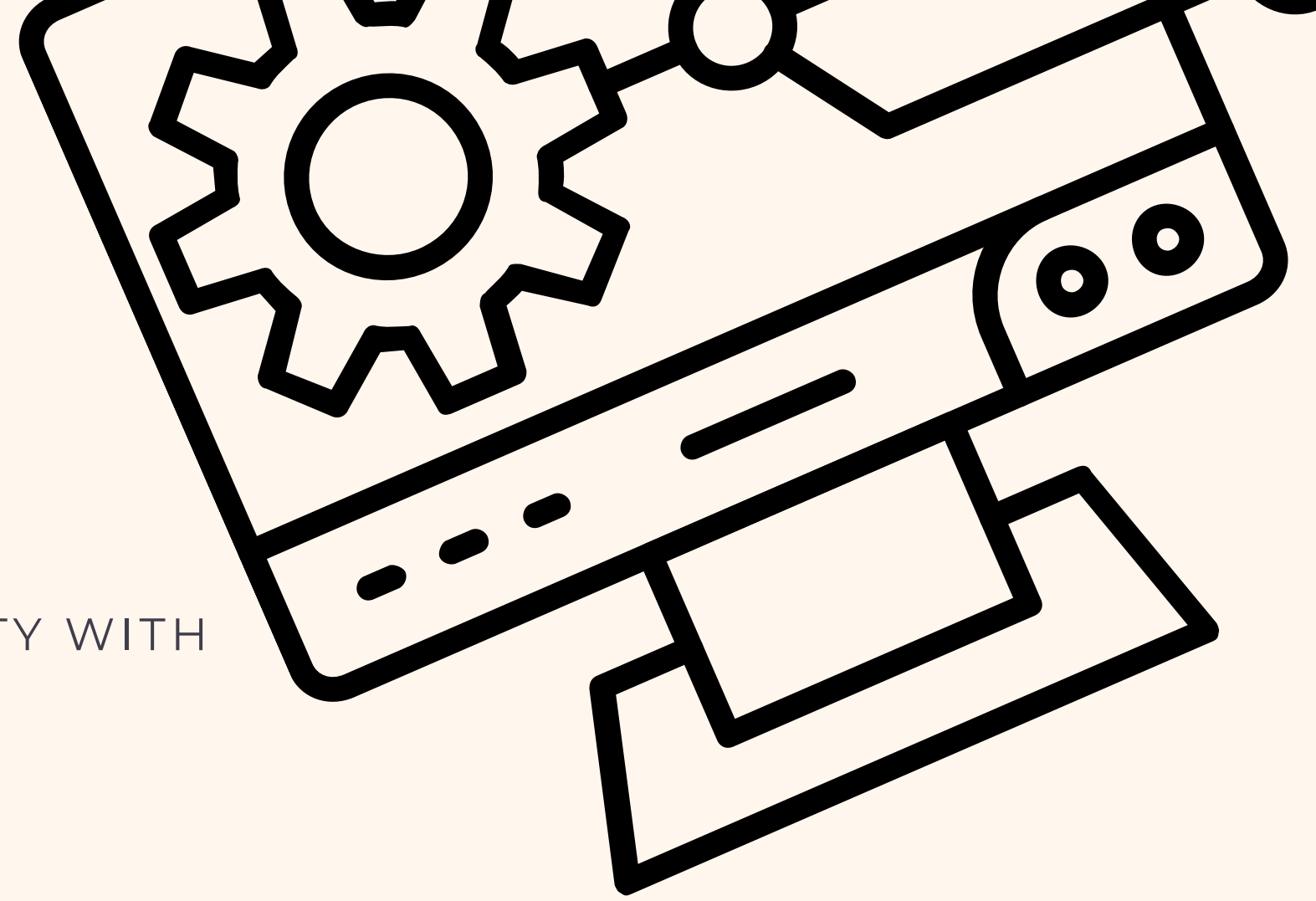


REPRODUCIBLE RESEARCH WITH GIT AND GITHUB

ENSURING SCIENTIFIC REPRODUCIBILITY WITH VERSION CONTROL



WHY REPRODUCIBILITY MATTERS

In research and data science, reproducibility means others can repeat your work and get the same results.

Git + GitHub help track every change — from code to datasets — ensuring transparency, collaboration, and trust.

Without version control, data, code, and results can quickly become inconsistent.

SETTING UP A REPRODUCIBLE WORKFLOW

Step 1: Use a consistent folder structure:

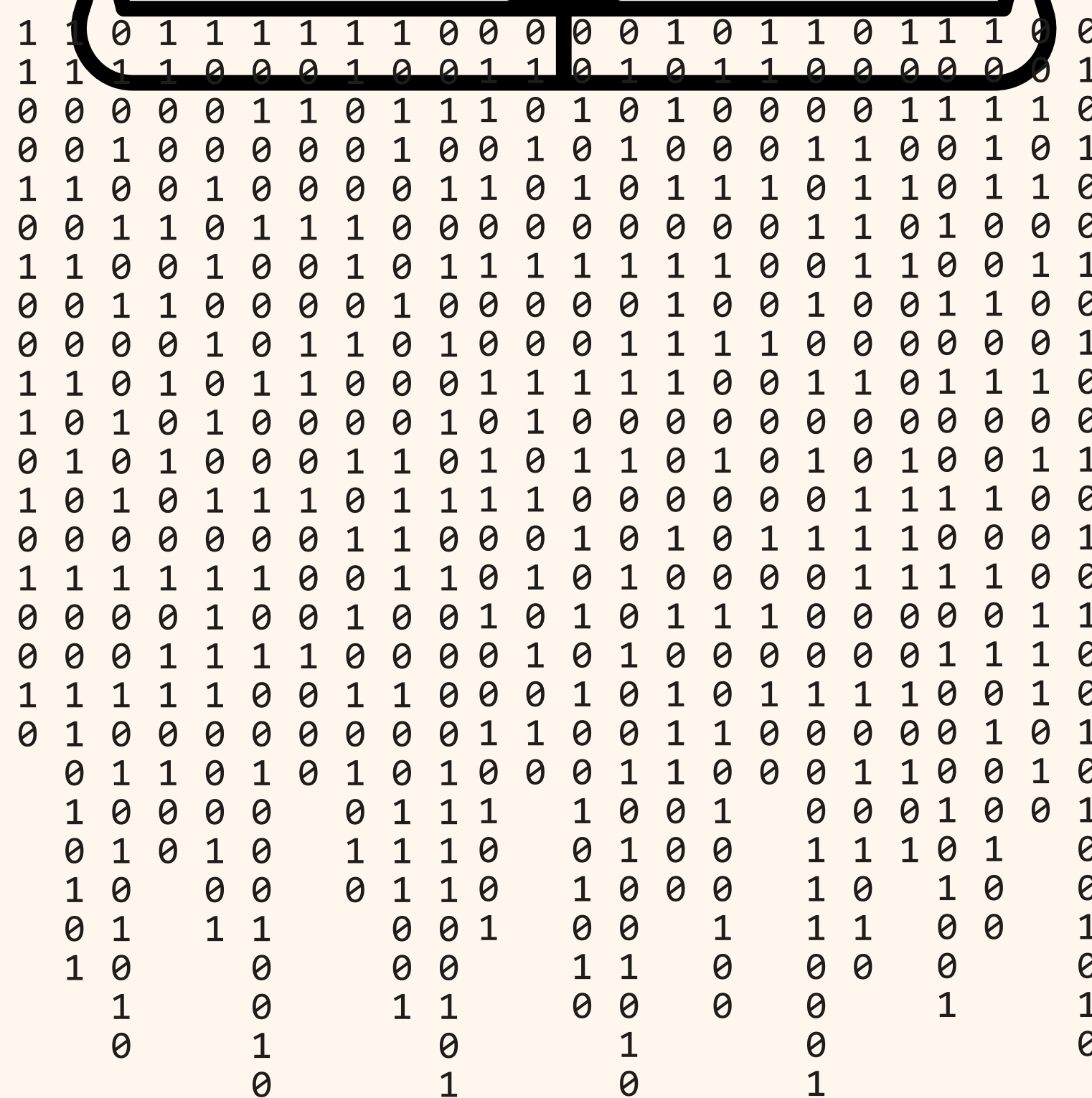
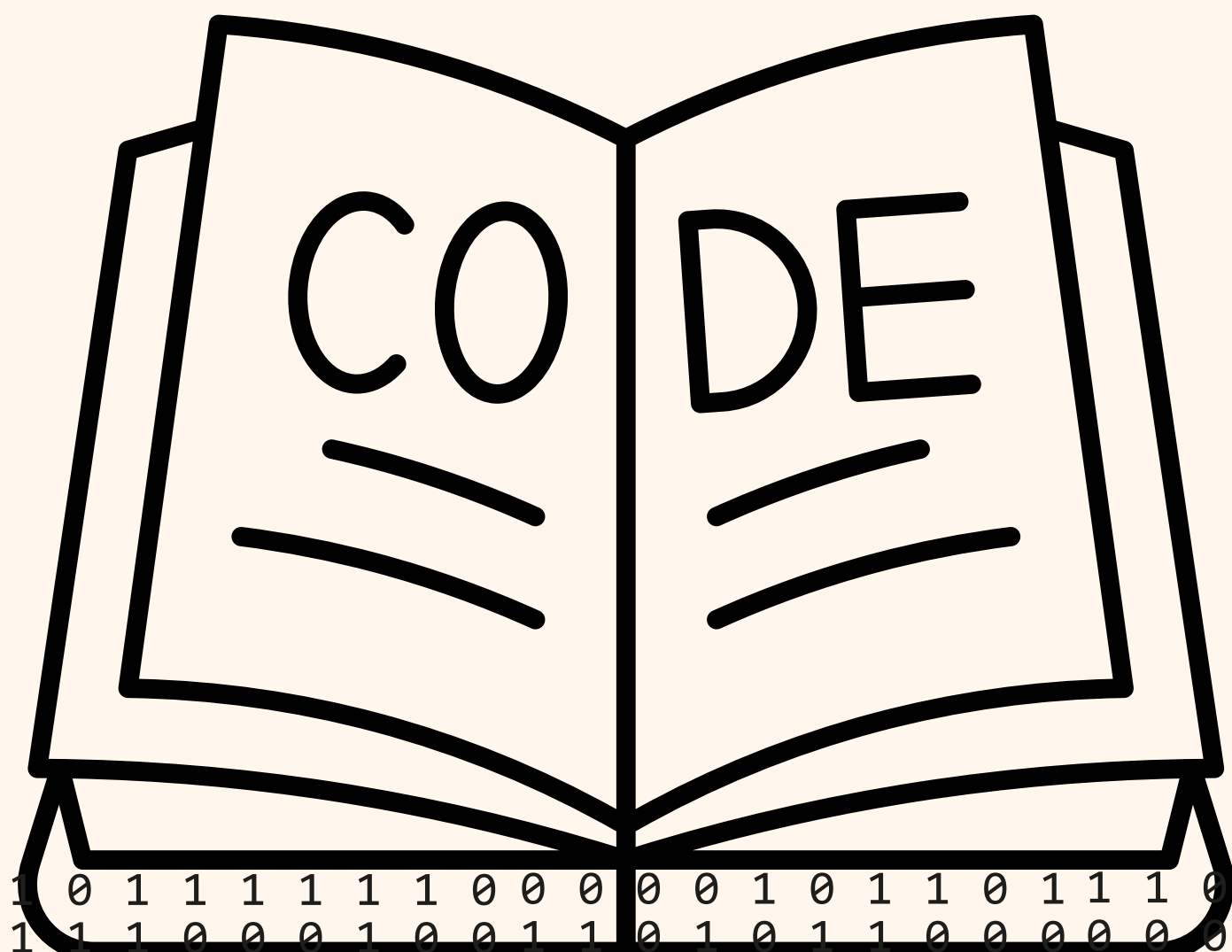
```
project/  
├── data/  
├── src/  
├── results/  
└── docs/
```

Step 2: Add a README.md explaining:

- Purpose of the study
- Dependencies and environment setup
- How to run the analysis

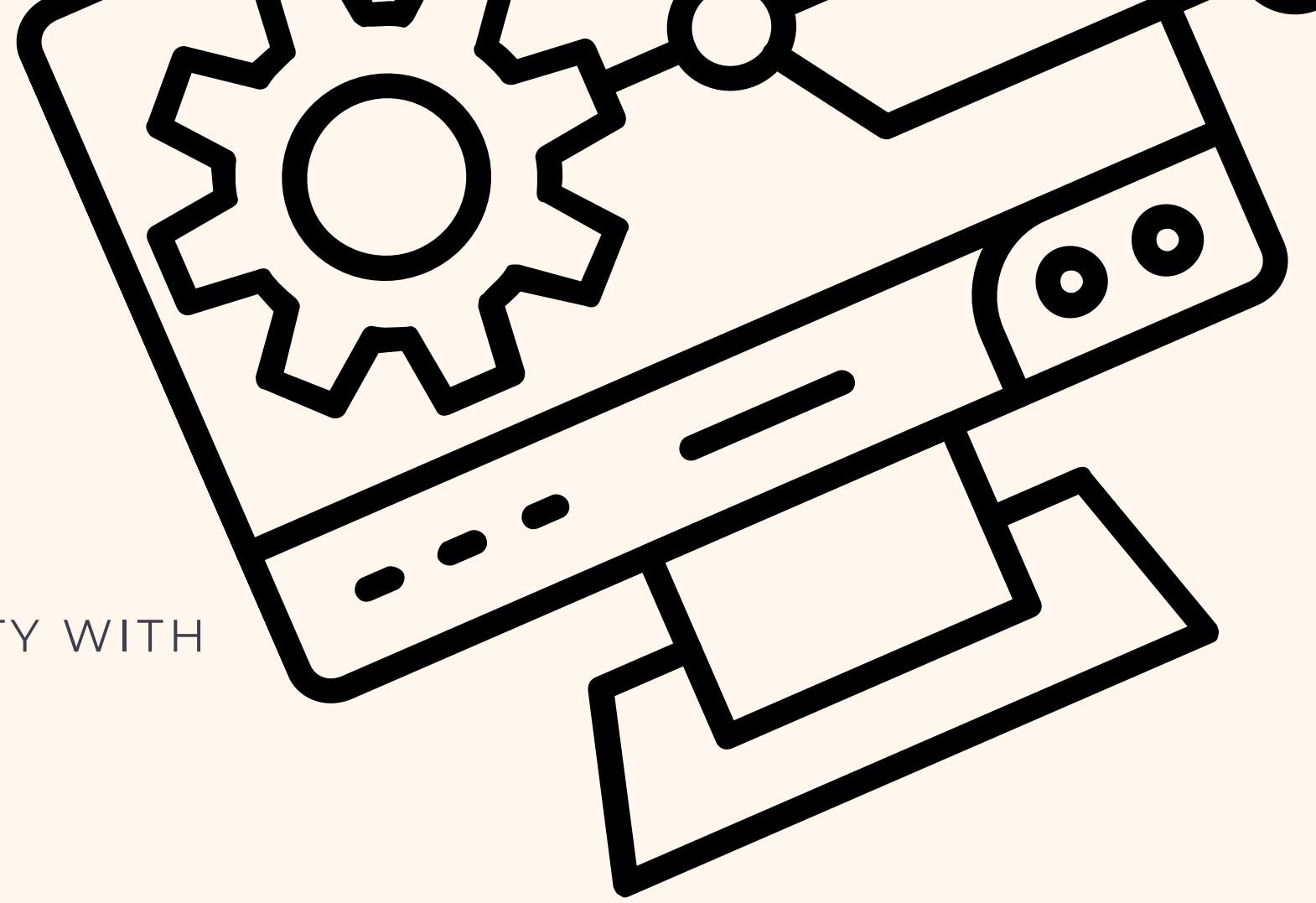
Step 3: Use .gitignore to exclude temporary or large files:

```
*.CSV  
*.log  
__pycache__/
```



REPRODUCIBLE RESEARCH WITH GIT AND GITHUB

ENSURING SCIENTIFIC REPRODUCIBILITY WITH VERSION CONTROL



GIT FOR RESEARCH TRACKING

- Commit frequent lying with clear messages:

```
git commit -m "Added regression model for 2022 data"
```

- Tag important milestone:

```
git tag -a v1.0 -m "Final analysis version"
```

- Document code and results in Markdown or Jupyter notebooks



SHARING AND COLLABORATING ON GITHUB

- Push repositories to Github for peer access.
- Use **Issues** to log bugs or analysis changes.
- Use **Pull Requests** for reviewing updates to analysis scripts.
- Enable **GitHub Actions** for automated testing or data validation.



OUTCOME

Reproducible projects increase credibility and enable others to build on your work and in some case more efficiently



LET'S DO OUR PART

If it can't be reproduced, it does not exist.
--Principle of modern computational science.

