Code Cleaning

MK Lau, Ana Trisovic, Thomas Pasquier

January 24, 2019

- Open-source explosion
- Issues with reproducibility (R, python)
- What's easy to do?
- Needs for automated cleaning
  - Libraries and system info
  - Check paths
  - Formatting (formatR)
  - Results checking
  - Results focused code base reduction (Rclean, CodeDepends, pyCharm, encapsulator)
- Linking to notebooks, capsules and continuous integration
  - Jupyter
  - Rmarkdown
  - Code Ocean
  - Gigantum
  - Globus
  - Whole Tale
  - CORe2
  - DatBlog
- Conclusion: low-bar to significant reproducibility, reuse and benefaction

Notebooks

- `https://www.datacamp.com/community/blog/jupyter-notebook-r#alternatives`
- `http://tkf.github.io/emacs-ipython-notebook/`
- `https://medium.com/@boyanangelov/is-this-the-best-data-science-ide-jupyter-lab-review-`

Journal Targets

- Nature Scientific Data
- Scientific Reports
- Royal Society of Open Science
- Journal of Open Research Software

Possible co-authors:

- Chris Chen
- Margo Seltzer
- Merce Crosas