Collaboration on GitHub

QBS 101.5: Applied Data Science

Simon Stone

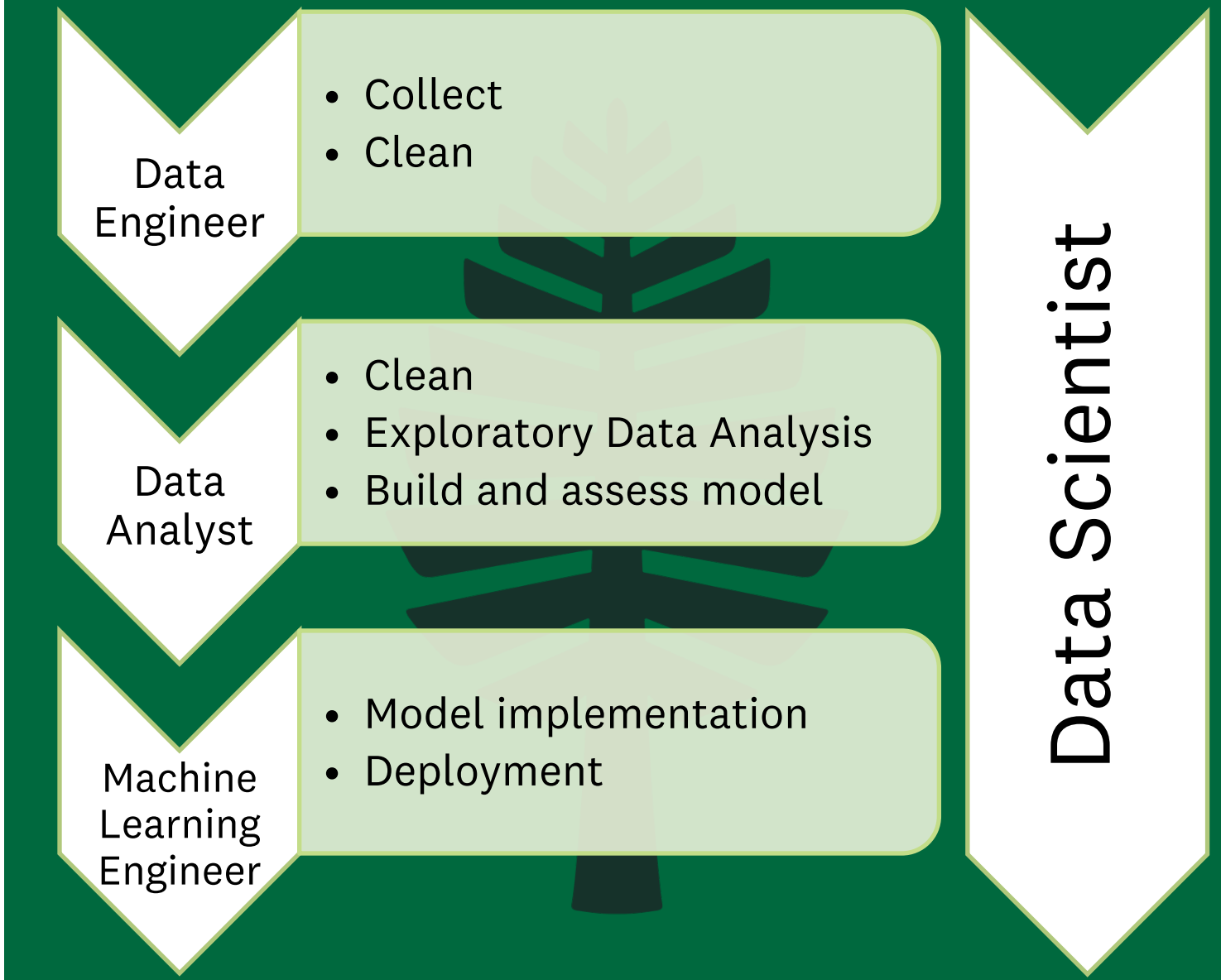
Research Data Services

Dartmouth College Library

Introduction
Data Science

Data Science is OSEMN!*

- 📥 Obtain
- 🧼 Scrub
- 🔎 Explore
- 🤖 Model
- ⁉️ iNterpret



*pronounced “awesome” - /ˈɔ.səm/
https://www.datascience-pm.com/osemn

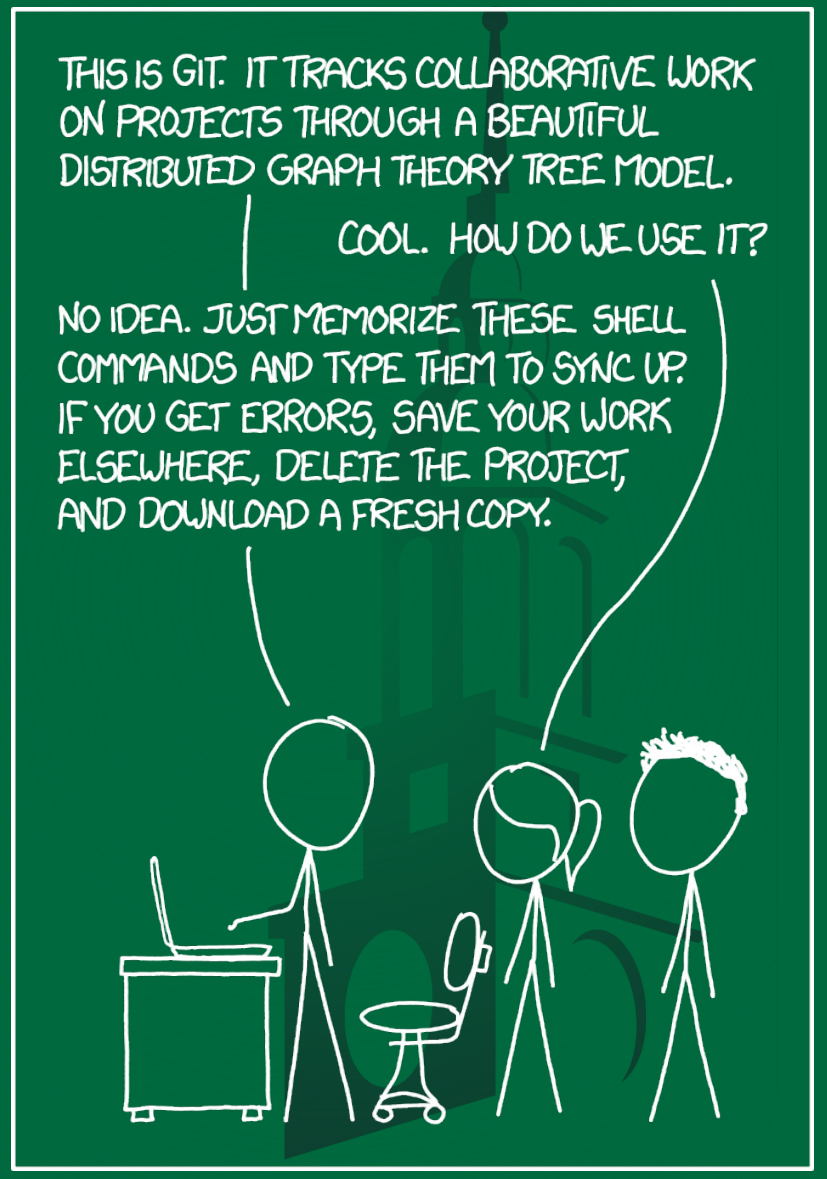
Collaboration on GitHub
Outline

1. Introduction to GitHub
2. Setting Up Collaboration
3. Collaborative Workflows
4. Managing Collaborative
5. Further resources

Caveat Emptor

Most of today's content is from a general software developer's perspective.

Not everything may fit a Data Scientist's needs or use cases all the time.



Munroe, R. Git. *XKCD*. Retrieved August 3, 2023, from https://xkcd.com/1597.

Collaboration on GitHub

1. Introduction to GitHub

Git:

- 🎚️ Distributed version control system
- 💻 Lives and works on your local machine
- ☝️ Responsible for tracking changes and managing branches

GitHub:

- ☁️ Cloud-based hosting platform for Git repositories
- 🌎 Allows remote collaboration
- 📋 Offers tools for project management, collaboration, code sharing, and more

Collaboration on GitHub

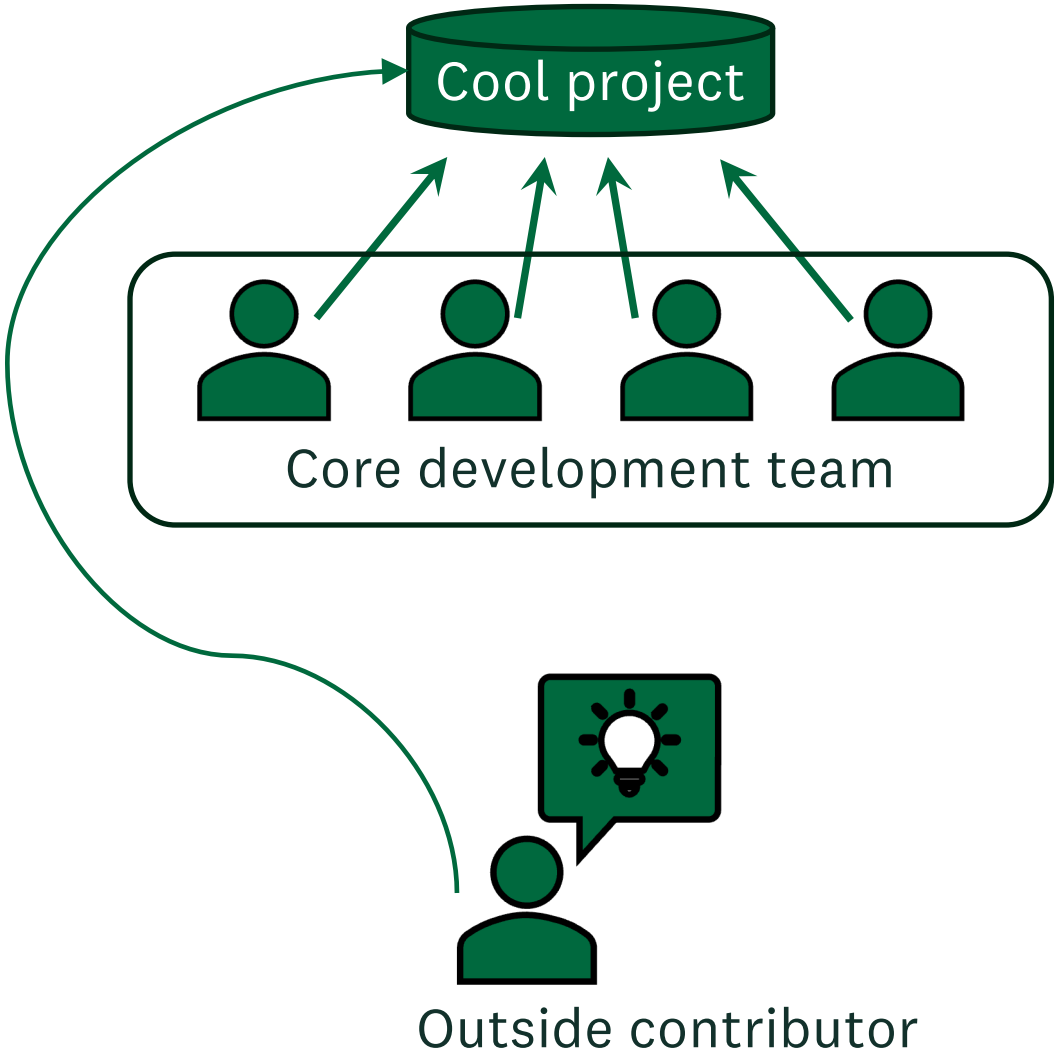
2. Setting up collaboration

Prerequisites:

- Every collaborator has a GitHub account
- Every collaborator has Git installed on their local machine

Two paradigms of collaboration:

1. Being part of a team working on a shared code base
2. Contributing to another team's code base

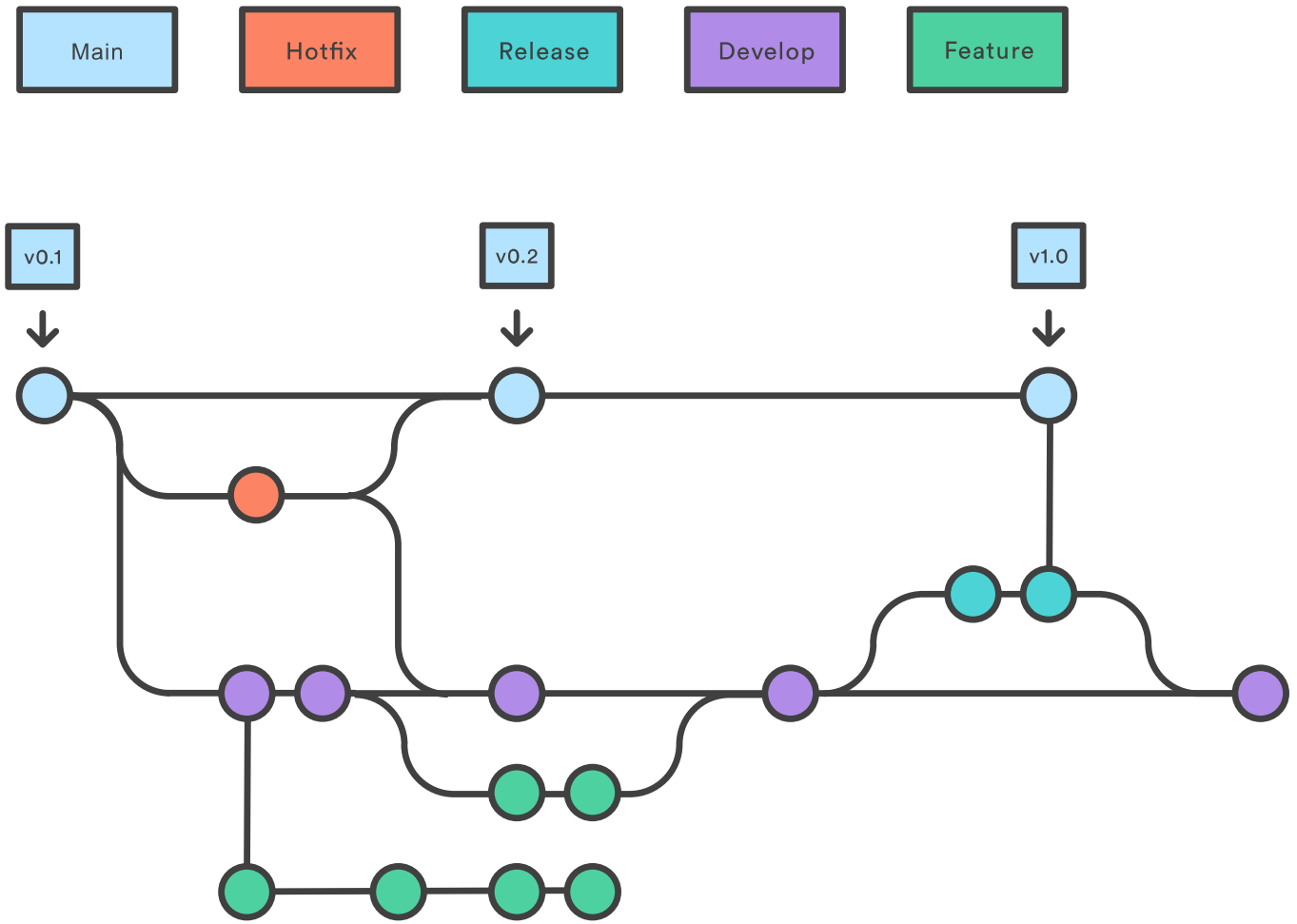


Collaboration on GitHub

3. Collaborative Workflows

The Git Flow:

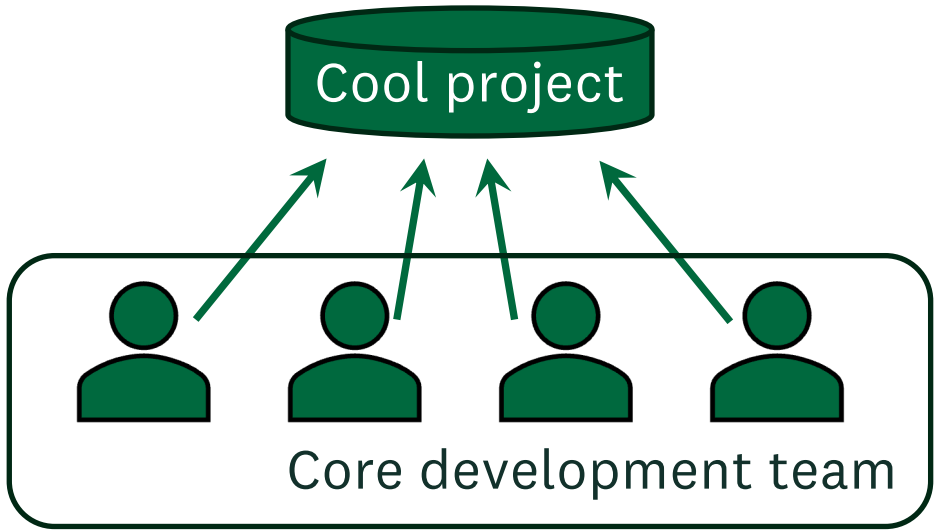
- A branch-based model
- Meaningful branching makes for a clean and informative history at different levels



3. Collaborative Workflows

Working as a team on a shared code base

- The project lives in a single repository hosted on GitHub
- All team members are added as collaborators
- Each team member works on their own branch
- When work is finished, a *pull request* is opened to request inclusion of the changes into the main branch



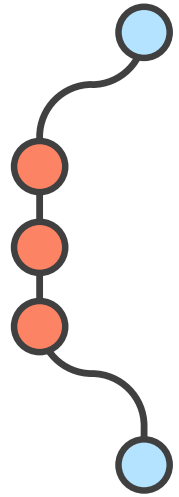
3. Collaborative Workflows

Pull requests

- A pull request is the request to merge a source branch to a target branch

Example:

> You have been working on the branch “`cool_new_feature`”. After a couple of commits to this branch, the feature is now ready to be published. You initiate a pull request from “`cool_new_feature`” to “`main`”

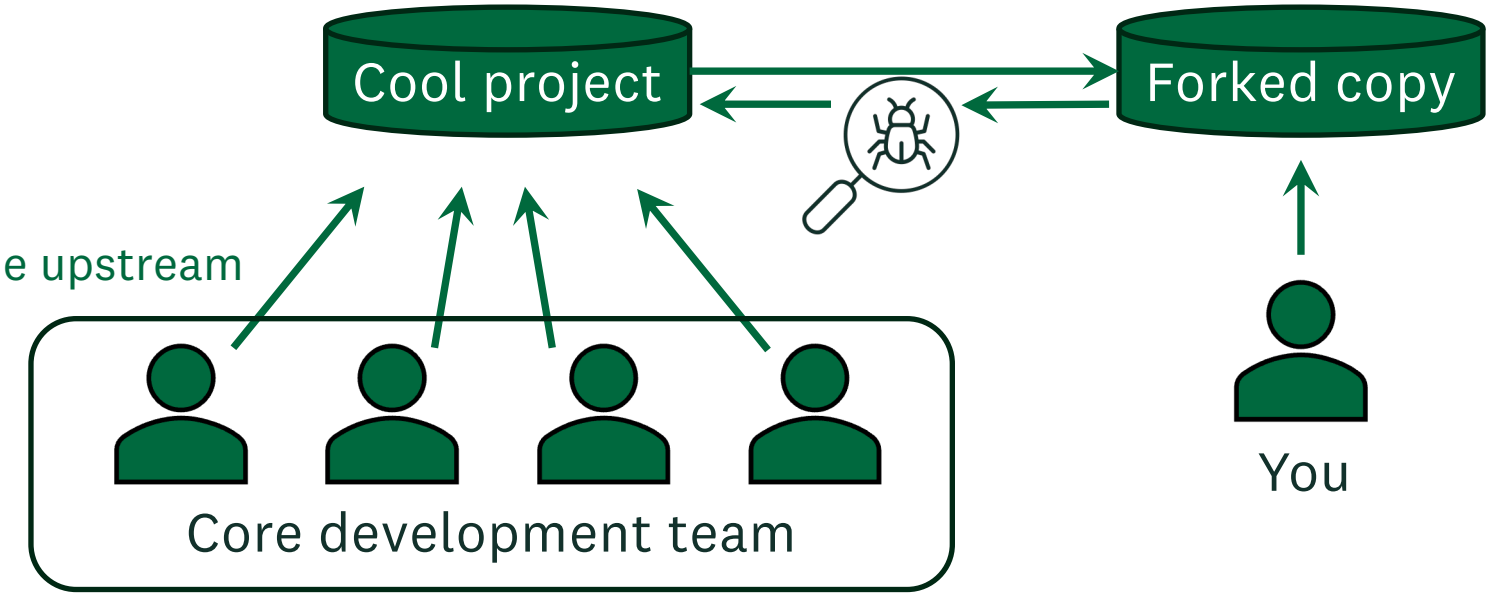


- A pull request helps to adhere to the four-eyes principle
- Pull requests are usually reviewed and often changes are requested before they are accepted (see later in this session)

3. Collaborative Workflows

Contributing to another team's code base

- Create a copy of the original project in your GitHub account (a "fork")
- You can always pull changes from the upstream repo
- The fork is your repo and you have full control over it
- You can ask for changes in your fork to be pulled into the original project ("pull request")



4. Collaborative Workflows
Managing Collaborative Projects

🚩 Issues

- Flag issues
- Plan and track a chunk of work

📋 Project boards

- Plan and track work
- Assign tasks
- Tasks turn into issues

📝 Pull request reviews

- Comment or approve proposed changes in a pull request
- Request further changes before merging

💬 Discussions

- A messageboard attached to your project
- Possible references to issues, pull requests, files, …

5. Further resources

- 📚 GitHub Docs: https://docs.github.com/
- 👌 GitHub integration in VS Code: https://code.visualstudio.com/docs/sourcecontrol/github
- 🤖 Automation (CI/CD): https://github.com/features/actions

Munroe, R. Git Commit. *XKCD*.
Retrieved August 8, 2023, from https://xkcd.com/1296.

Thank you.