Facilitating team-based data science: agile and scrum for undergraduates

Nicholas J. Horton, Randi Garcia, Chelsey Legacy

Outline

1. DSC-WAV
2. Agile & Scrum
3. Kanban Boards
4. Sprint Planning
5. Sprint Retrospective
6. Student Perspectives
7. Lessons Learned

Activities

Activity 1: Populating a backlog for a sprint

Activity 3: Sprint Planning

Activity 5: Stand-up Meeting

Activity 2: Kanban Boards in Github
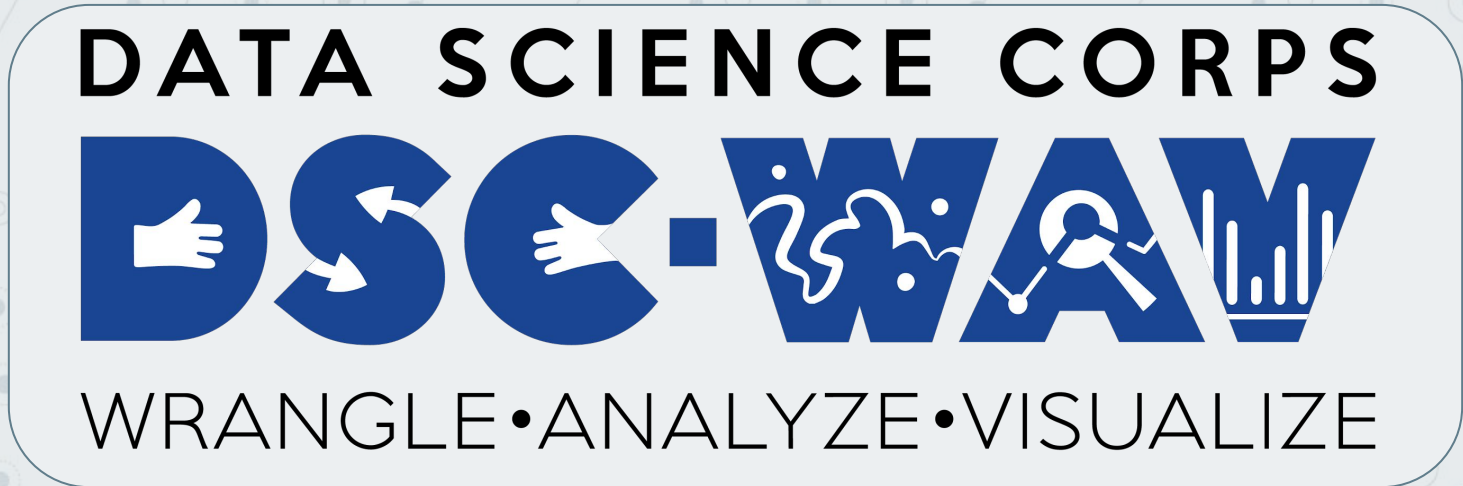
Activity 4: Work on Sprint

Activity 6: Map Plans to Semester

DSC-WAV (Wrangle-Analyze-Visualize)

NSF funded effort from the Harnessing the Data Revolution (HDR) Data Science Corps (DSC) initiative, 1923388, 1923700, 1923934, and 1924017.
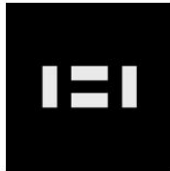
DSC-WAV (Wrangle-Analyze-Visualize)

◎ https://dsc-wav.github.io/www

◎ Collaboration with:

- Five Colleges (Amherst, Smith, Hampshire, Mount Holyoke, and UMass/Amherst)

- Greenfield Community College, Holyoke Community College, Springfield Technical Community College

- University of Minnesota

Goal 1

create opportunities for undergraduate students to work on Data Science for Social Good projects for community organizations

ValleyBike

girls inc.
of the Valley

Western Massachusetts Health Equity Network
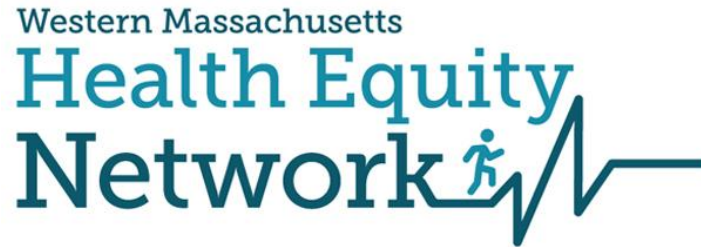
Goal 2

help build data science programs at two-year colleges

GREEN FIELD cc

HOLYOKE COMMUNITY COLLEGE Foundation, Inc.

STCC

Roundtable on Data Science Postsecondary Education

A Compilation of Meeting Highlights

(2020)

| Download Free PDF | Read Free Online | Buy Paperback:$60.00 | Buy Ebook:$48.99 |

Epub, Kindle, MobiPocket
What is an Ebook?

ROUNDTABLE ON DATA SCIENCE POSTSECONDARY EDUCATION
A Compilation of Meeting Highlights

Data Science Course Design Principles

Design Principle	Students will . . .
Active Learning. The course provides regular opportunities for students to actively engage in data explorations using a variety of different instructional strategies	• Be active and engaged participants in discussion, in working on data explorations with classmates, and in making decisions about the direction of instruction based on their work.

Two-Year College Data Science Summit

May 10-11, 2018, Washington, DC metro area

Final Report

March 20, 2019 Webinar, "Data Science for Two-Year Colleges: A report of the Two-Year College Data Science Summit"; Recording, Slides

The University of Texas at Austin
Charles A. Dana Center

DSC-WAV (Wrangle-Analyze-Visualize)

- Building data acumen for undergraduate students
  - *HDSR*, 2021
    https://hdsr.mitpress.mit.edu/pub/nvflcexe/release/1
- Facilitating team-based data science: lessons learned
  - *FoDS*, 2022, https://arxiv.org/abs/2106.11209
- Data Science Transfer Pathways From Associate's to Bachelor's Programs
  - *HDSR*, 2023, https://hdsr.mitpress.mit.edu/pub/k4jt0uu0/release/1

FACILITATING TEAM-BASED DATA SCIENCE: LESSONS LEARNED FROM THE DSC-WAV PROJECT

CHELSEY LEGACY*

Department of Educational Psychology
University of Minnesota
Minneapolis, MN 55455, USA

ANDREW ZIEFFLER

Department of Educational Psychology
University of Minnesota
Minneapolis, MN 55455, USA

BENJAMIN S. BAUMER

Program in Statistical & Data Sciences
Smith College
Northampton, MA 01063, USA

VALERIE BARR

Department of Computer Science
Mount Holyoke College
South Hadley, MA 01075, USA

NICHOLAS J. HORTON

Department of Mathematics and Statistics
Amherst College
Amherst, MA 01002, USA

(Communicated by the associate editor name)

Harvard Data Science Review • Issue 3.1, Winter 2021

The Data Science Corps Wrangle-Analyze-Visualize Program: Building Data Acumen for Undergraduate Students

Nicholas J. Horton[1], Benjamin S. Baumer[2], Andrew Zieffler[3], Valerie Barr[4]

[1]Amherst College, [2]Smith College, [3]University of Minnesota, [4]Mount Holyoke College

Published on: Feb 25, 2021

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

Your Goals for This Workshop

1. Using GitHub and other technology to support collaborative data analysis
2. Leveraging Agile and scrum for fun and profit (in and out of the classroom)
3. Fostering team-based learning approaches
4. Generating project ideas for entry level courses
5. What and how to teach Data Science II

1.

Scrum and Agile

A short introduction

Agile Philosophy

1. Individuals and interactions over processes and tools
2. Working software over comprehensive documentation
3. Customer collaboration over contract negotiation
4. Responding to change over following a plan

Q: How can we port these insights from software development into statistics and data science education?
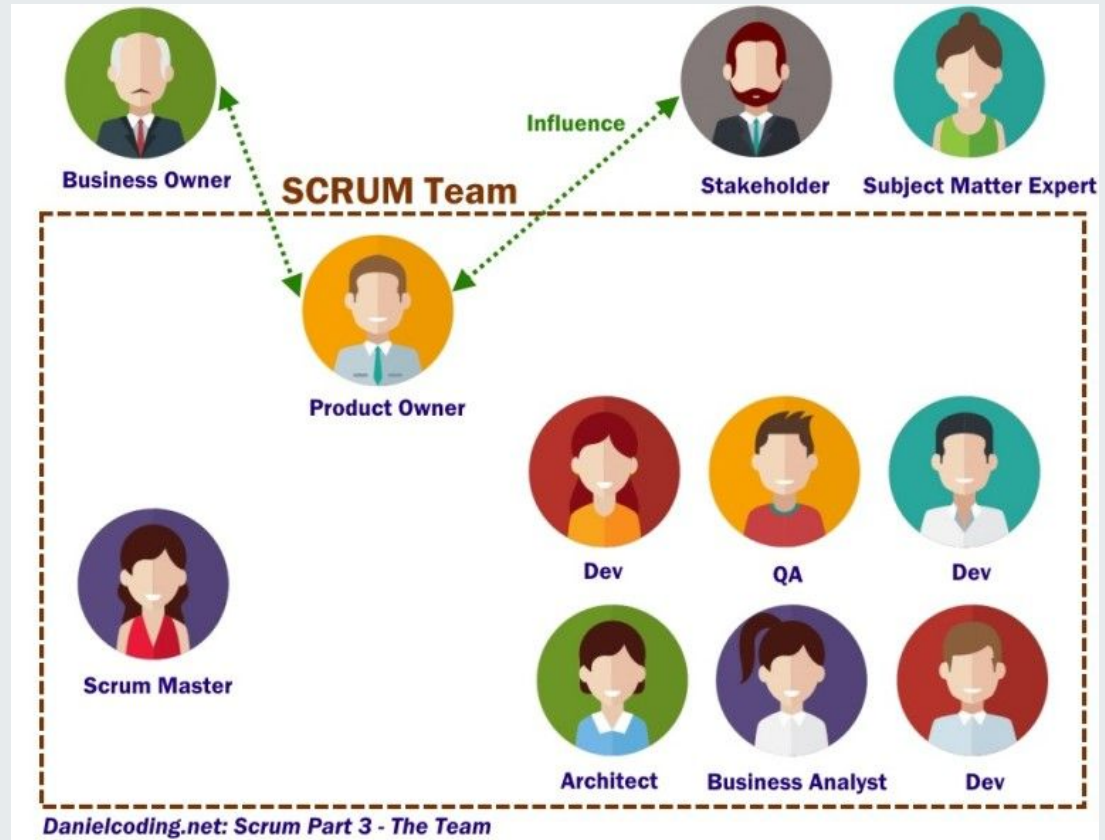
Q: How can we use these approaches to improve data analysis and team-based collaboration?

https://scrumguides.org/scrum-guide.html
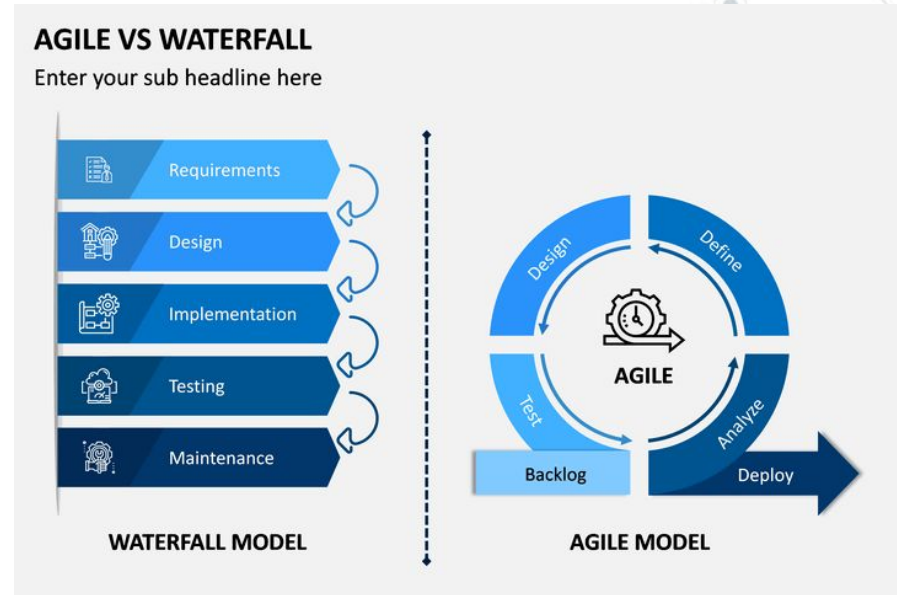
Scrum: a method of implementing the Agile Philosophy

- 3 Key Roles:
  - Product owner
  - Scrum master
  - Development team

Danielcoding.net: Scrum Part 3 - The Team

Key concepts and terms of Agile/Scrum

- A project is broken down into **Sprints**
  - Sprints are typically a 1-4 week period to achieve certain project goals

Some key parts of a Sprint:

- ***Sprint planning:*** goal setting and task defining
- *Sprint demos:* product demonstration meeting
  - with product owner
- ***Sprint retrospectives:*** team reflections on collaboration and project progress

Key concepts and terms of Agile/Scrum

Other concepts we will explore in this workshop:

- ○ **Daily stand-ups***: brief meetings to touch base*
- ○ **Backlogs***: list of tasks defined for that sprint*
- ○ **User stories***: A form of writing tasks to clarify a project objective*
    - ■ Language used for a user story:

        "As a _____, I want to _____ so that I can_____"

        Example: As a web developer, I want to be able to see all sections of the website clearly so the website is easy to navigate.

    - ■ Further clarify by adding "I will know when this task is completed when…"
- ○ **Kanban board***: a visual representation of the backlog and the progress made on the project*
    - ■ More on this later in the workshop!

The Agile Scrum Framework at a glance
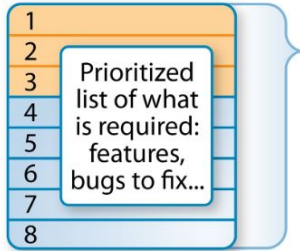
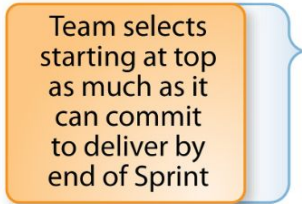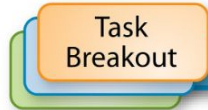**Inputs from Customers, Team, Managers, Execs**

Product Owner

The Team

Scrum Master

Daily Standup Meeting
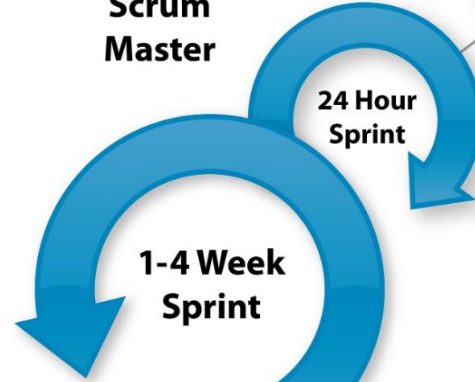
24 Hour Sprint

1-4 Week Sprint

Sprint Review

Product Backlog

1
2
3
4
5
6
7
8

Prioritized list of what is required: features, bugs to fix...

Sprint Planning Meeting

Team selects starting at top as much as it can commit to deliver by end of Sprint

Sprint Backlog

Task Breakout

Sprint end date and team deliverable do not change

Finished Work

Sprint Retrospective

neon rain interactive

AGILE FOR ALL
Making Agile a Reality®

Introductions

Introductions

◎ Introductions (start within pairs)
- ○ Name/Pronouns
- ○ Institution
- ○ A little about yourself
- ○ One thing to share

◎ You will then introduce the person you spoke with to the whole group

Activity 1

Populating the Backlog for Sprints

Activity 1

◎ Learn about your project
◎ Determine research questions
◎ Set goals to answer research questions

Instructors' mock project: Bike counts

An advocacy group in Northampton, Massachusetts has undertaken a set of bike counts at intersections across the city. They are interested in analysis of these data: what intersections are most heavily used? What times are bikes most commonly seen? Are there patterns in usage that relate to on and off-street accommodations?

We will use this mock project to demonstrate aspects of agile and scrum.

Participants' project: modeling movies

A group of students at the University of Minnesota has curated a rich dataset of movies. What insights can be extract?

You will use this mock project to experience aspects of agile and scrum.

Participants' project: Juniata Voices

Kim Roth from Juniata College has extracted data about usage of the "Juniata Voices", a showcase for lectures and creative works (https://www.juniata.edu/offices/juniata-voices/about.php).

Members of team 2 will have the opportunity to use this real project to experience aspects of agile and scrum (or they can also explore the movie data).

Activity 1

◎ Learn about your project
◎ Determine research questions
◎ Set goals to answer research questions

Activity 1

◎ Learn about your project
◎ Determine research questions
◎ Set goals to answer research questions
◎ Use paper/pencil or word/google doc to *populate your backlog*

Example finished DSC-WAV backlog

2.

GitHub & Kanban Boards

A short introduction

GitHub (a subsidiary of Microsoft)

From the GitHub website:

*GitHub is a **code hosting platform** for **version control** and **collaboration**. It lets you and others work together on projects from anywhere.*

More on GitHub & GitHub Classroom

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS   Check for updates

Implementing Version Control With Git and GitHub as a Learning Objective in Statistics and Data Science Courses

Matthew D. Beckman[a], Mine Çetinkaya-Rundel[b,c,d], Nicholas J. Horton[e], Colin W. Rundel[b,d], Adam J. Sullivan[f], and Maria Tackett[d]

[a]Department of Statistics, Penn State University, State College, PA; [b]School of Mathematics, University of Edinburgh, Edinburgh, UK; [c]RStudio, Boston, MA; [d]Department of Statistical Science, Duke University, Durham, NC; [e]Department of Mathematics and Statistics, Amherst College, Amherst, MA; [f]Department of Biostatistics, Brown University, Providence, RI

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS   Check for updates

DATA SCIENCE

Using GitHub Classroom To Teach Statistics

Jacob Fiksel[a], Leah R. Jager[a], Johanna S. Hardin[b], and Margaret A. Taub[a]

[a]Department of Biostatistics, Johns Hopkins University, Baltimore, MD; [b]Department of Mathematics, Pomona College, Claremont, CA
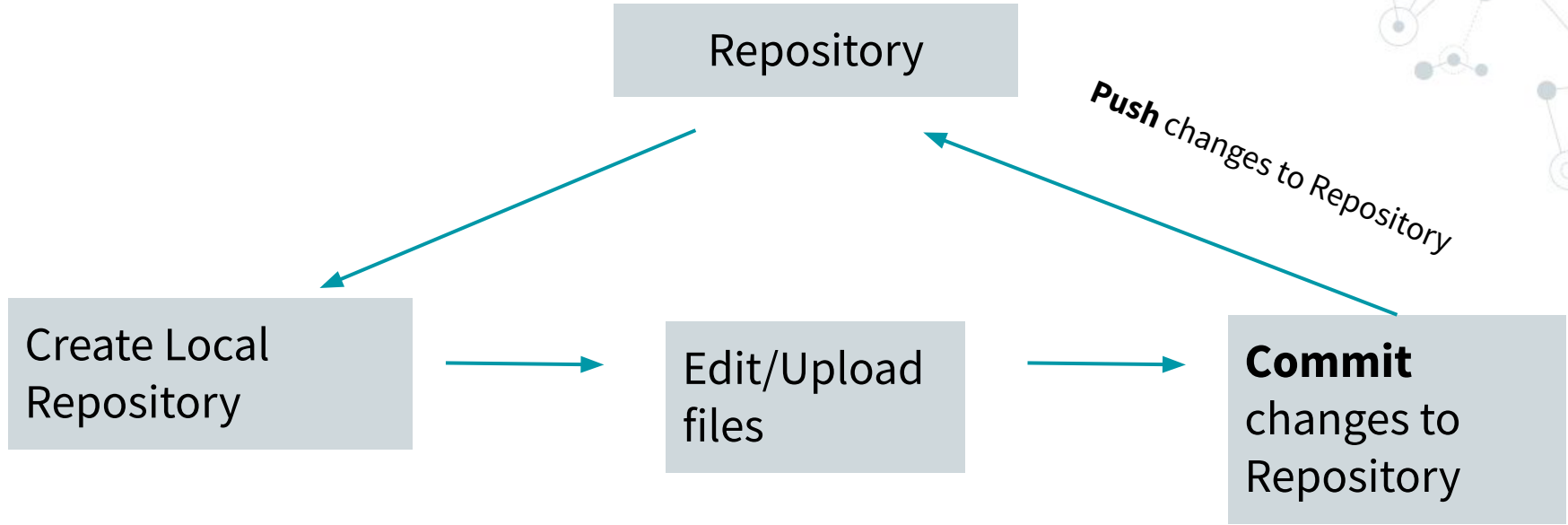
GitHub

◎ Your project materials (code, codebooks, .pdf files etc) are all stored in a **repository (or "repo")**
- ○ Often public, can be made private

◎ Several ways to access and update your GitHub repository
- ○ GitHub website
- ○ GitHub Desktop
- ○ R Studio…..and more!

Github Vocabulary
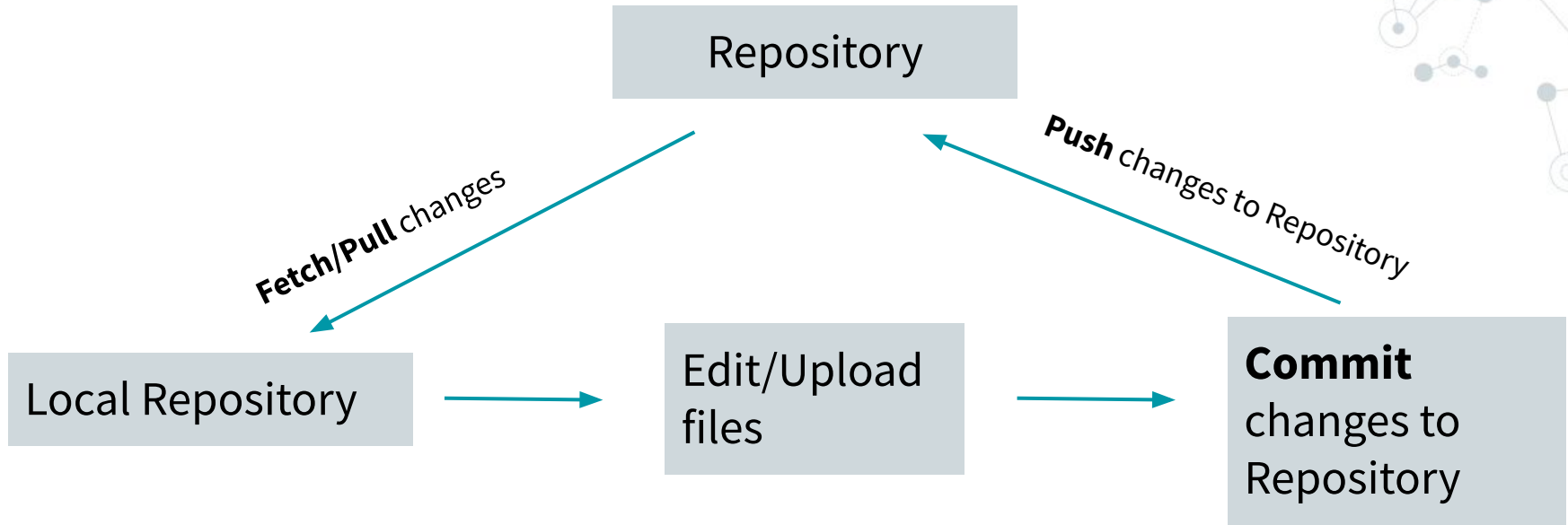
◎ **Clone** a repository to get a version of it on your computer for editing

◎ You can add/edit files on your own computer the update the repository in GitHub by making **commits**

◎ **Pull/Fetch** any commits made by other team members to your local repository before updating documents

◎ **Push** your commits from your local repository to the main repository

Typical GitHub Workflow

Repository

Push changes to Repository

Create Local Repository

Edit/Upload files

Commit changes to Repository

Don't forget to Fetch/Pull often to avoid merge conflicts

Typical GitHub Workflow

Repository

Fetch/**Pull** changes

Push changes to Repository

Local Repository

Edit/Upload files

Commit changes to Repository

GitHub - > Kanban Boards

◎ Github:
- ○ Storing files
- ○ Version Control
- ○ Collaboration
  - ◉ Kanban boards

Kanban Boards

◎ Help organize a list of tasks needed to complete a sprint
◎ During sprint planning tasks are assigned to team members

Ultimate Goal: Keep every team member up to date on the status of tasks during the sprint

Example of a Kanban Board

Software to facilitate this

◎ GitHub
◎ Trello
◎ Google Jam Boards
◎ Many more!

Kanban Boards in GitHub

◎ Project section can facilitate Kanban Boards
◎ Populate the backlog by creating issues
◎ Assign issues
◎ Mark tasks completed
◎ Comments can facilitate discussion within those issues
  ○ Sends notifications to students when they are mentioned

Great for communication!

Kanban Boards in GitHub

🔒 **Sprint 4**

Updated on Jun 23, 2021

| 🔍 Filter cards | + Add ca |

3 To do + ⋯

⊙ **Create a button that saves .png of map** ⋯
#74 opened by ▓▓▓▓
`2 hour`

⊙ **Create dropdown menu of "Data Sources"** ⋯
#76 opened by ▓▓▓▓

⊙ **explore automatic deployment of Shiny app** ⋯
#57 opened by ▓▓▓▓ ▓

Automated as `To do`   Manage

1 In progress + ⋯

⊘ **Organize the UI shiny app object to reflect desired layout** ⋯
#68 opened by ▓▓▓▓
`2 hour`

Automated as `In progress`   Manage

10 Done + ⋯

⊘ **Figure out why the website won't update** ⋯
#82 opened by ▓▓▓▓
`1 hour`

⊘ **Create CSS file for size/design customizations** ⋯
#67 opened by ▓▓▓▓ ▓
`2 hour`

⊘ **Choose a website theme** ⋯
#65 opened by ▓▓▓▓
`1 hour`

Automated as `Done`   Manage

We have populated Kanban boards for your project groups

🔒 **DSC-WAV** / **uscots-team2** Private

generated from DSC-WAV/uscots-bikecounts

<> **Code**   ⊙ Issues 4   ⎇ Pull requests   ▶ Actions   ⊞ Projects 1

⊞ **1 Open**   🗄 0 Closed

⊟ **USCOTS Team 2 Kanban board** Private

#14 updated 10 hours ago

View 1 ▾   + New View

Filter by keyword or by field

🟢 Todo 3
This item hasn't been started

⊙ uscots-team2 #1 •••

review codebook for the movie data

⊙ uscots-team2 #2 •••

identify key issues for the movie data

⊙ uscots-team2 #3

test of Github access

🟤 In Progress 0
This is actively being worked on

🟣 Done 0
This has been completed

DEMO: Add a new Board (this will be the Kanban board style)

A blank board will appear
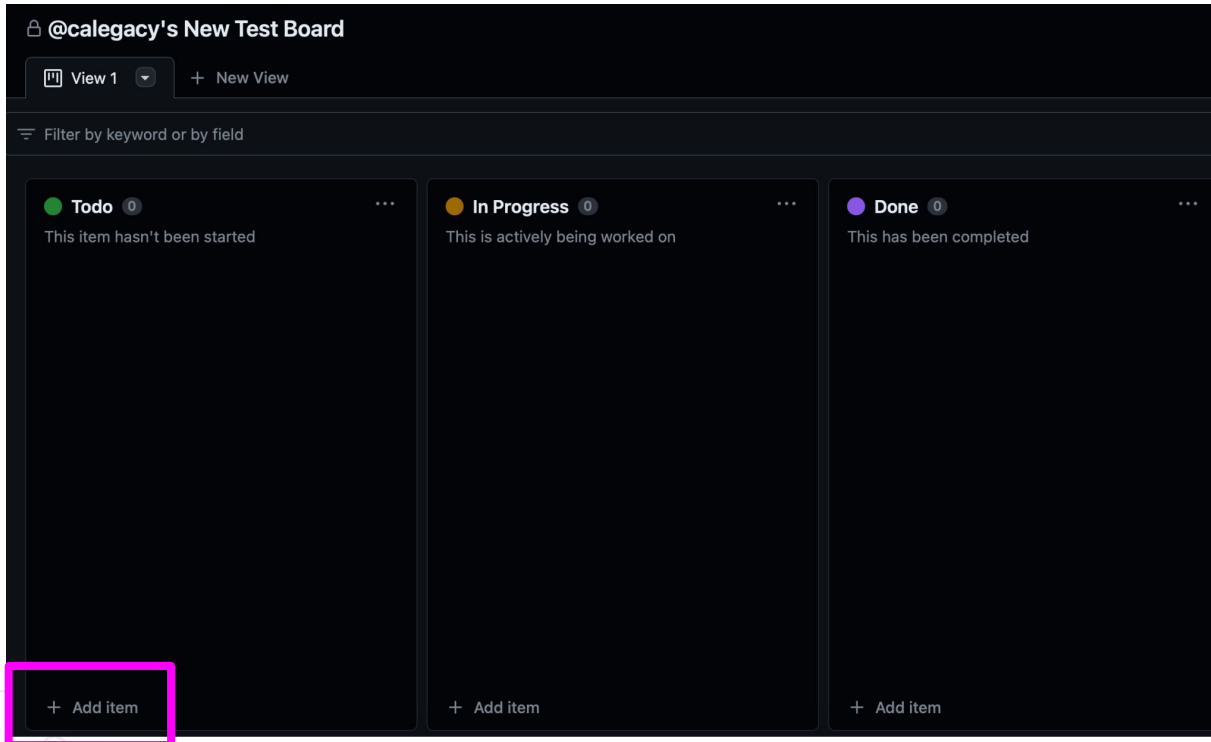
Now you can add things from your backlog to it

As a

Draft

then do this

Draft •••

then do

⊙ Convert to issue

⎘ Copy link in project

🗄 Archive   E

🗑 Delete from project Del
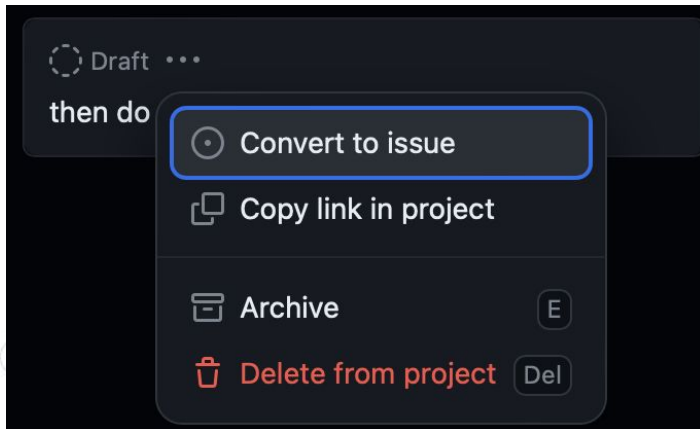
Finish by selecting your project repository

Activity 2

Kanban Board in GitHub

The Agile Scrum Framework at a glance

Inputs from Customers, Team, Managers, Execs

Product Owner

The Team

Scrum Master

Burn Down/Up Chart

Daily Standup Meeting

24 Hour Sprint

1-4 Week Sprint

Sprint Review

Product Backlog

1
2
3
Prioritized list of what is required: features, bugs to fix...
4
5
6
7
8

Sprint Planning Meeting

Team selects starting at top as much as it can commit to deliver by end of Sprint

Task Breakout

Sprint Backlog

Sprint end date and team deliverable do not change

Finished Work

Sprint Retrospective

neon rain interactive

AGILE FOR ALL
Making Agile a Reality®

Activity 2

- ◎ Go to GitHub
- ◎ Populate the kanban board with your tasks
- ◎ Don't worry about assigning them to people - you will do that in the next activity

3.

Sprint Planning Meeting

A short introduction

The Agile Scrum Framework at a glance

Inputs from Customers, Team, Managers, Execs

Product Owner

The Team

Scrum Master

Burn Down/Up Chart

Daily Standup Meeting

24 Hour Sprint

1-4 Week Sprint

Sprint Review

| 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| 6 |
| 7 |
| 8 |

Prioritized list of what is required: features, bugs to fix...

Product Backlog

Team selects starting at top as much as it can commit to deliver by end of Sprint
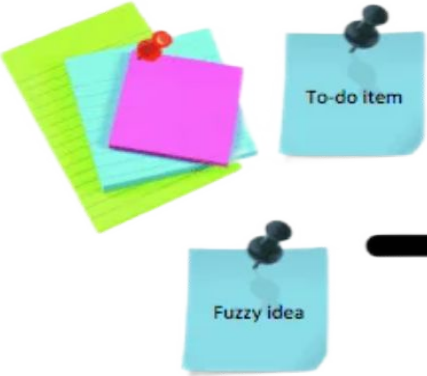
Sprint Planning Meeting

Task Breakout

Sprint Backlog

Sprint end date and team deliverable do not change

Finished Work

Sprint Retrospective

neon rain interactive

AGILEFORALL
Making Agile a Reality®

CC BY ND

Sprint Planning Meeting

Backlog

Sprint Planning Meeting

Kanban Board

Sprint Planning Meeting

During the spring planning meeting we:

1. Greet each other
2. Hear product updates
3. Decide on which issues to work on during this sprint
4. Assign Github issues to team members
5. Write user stories and detailed todo lists in Github

Sprint Planning Meeting

During the spring planning meeting we:

1. Greet each other
   - Sprint #1: ice breakers
   - Sprints #>1: reminders agreed upon procedural changes
2. Hear product updates
3. Decide on which issues to work on during this sprint
4. Assign Github issues to team members
5. Write user stories and detailed todo lists in Github

Sprint Planning Meeting

During the spring planning meeting we:

1. Greet each other
2. Hear product updates
   - Sprint #1: intro to project; review backlog
   - Sprints #>1: updates from client via product manager
3. Decide on which issues to work on during this sprint
4. Assign Github issues to team members
5. Write user stories and detailed todo lists in Github

Sprint Planning Meeting

During the spring planning meeting we:
1. Greet each other
2. Hear product updates
3. Decide on which issues to work on during this sprint
   ○ For students: discuss upcoming assignments/exams
4. Assign Github issues to team members
5. Write user stories and detailed todo lists in Github

Sprint Planning Meeting

During the spring planning meeting we:

1. Greet each other
2. Hear product updates
3. Decide on which issues to work on during this sprint
4. Assign Github issues to team members
   - Sprint #1: maybe light issues; "learn how to use Github"
   - Sprints #>1: scrum master review backlog beforehand
5. Write user stories and detailed todo lists in Github

Sprint Planning Meeting

During the spring planning meeting we:

1. Greet each other
2. Hear product updates
3. Decide on which issues to work on during this sprint
4. Assign Github issues to team members
5. Write user stories and detailed todo lists in Github
   - "As a _____, I want to _____ so that I can_____"
   - "I will know when this task is completed when…"

Assigning Issues and Writing User Stories in Github

Link to bikecount kanban board:

USCOTS bikecount - Sprint 1

Sprint Planning Meeting

◎ Important to set the tone for the sprint
◎ This is your primary full group meeting
  ○ Along with demos and retrospectives
◎ Forward looking
  ○ It is about the coming sprint – resist the urge to do any postmortem for the last sprint
◎ Helpful to have a **pre-planning meeting** for some groups/projects
  ○ Backlog refinement

Activity 3

Sprint Planning

Sprint Planning

In your groups:

1. Greet each other — Briefly check-in: How's this all feeling?
2. Hear product updates — Any new thoughts for the backlog? Need any issues like "update RStudio"?
3. Decide on which issues to work on during this sprint — Are they all in the Kanban board?
4. Assign GitHub issues to team members
5. Write user stories and detailed todo lists in GitHub

Break for Lunch

11:45am-1:00pm

Activity 4

Time to work on Sprint issues (until 1:30pm)
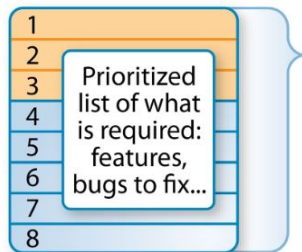
The Agile Scrum Framework at a glance

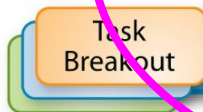**Inputs from Customers, Team, Managers, Execs**

Product Owner

Product Backlog

| 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| 6 |
| 7 |
| 8 |

Prioritized list of what is required: features, bugs to fix...

The Team

Team selects starting at top as much as it can commit to deliver by end of Sprint

Sprint Planning Meeting

Sprint Backlog

Task Breakout

Scrum Master

Burn Down/Up Chart

Daily Standup Meeting
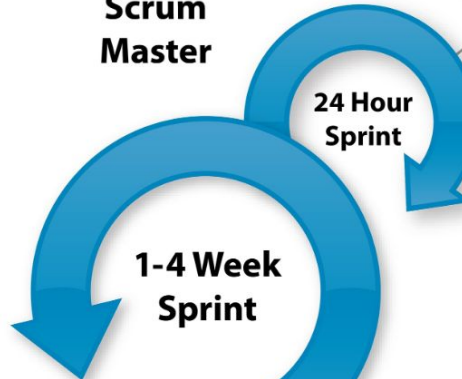
24 Hour Sprint

1-4 Week Sprint

Sprint end date and team deliverable do not change

Sprint Review

Finished Work

Sprint Retrospective

neon rain interactive

AGILEFORALL
Making Agile a Reality®

4.

Stand-ups

A short introduction

The Agile Scrum Framework at a glance

Inputs from Customers, Team, Managers, Execs
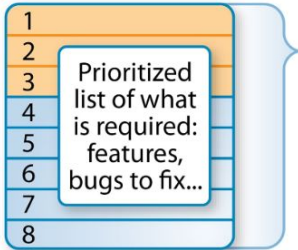
Product Owner

The Team

Scrum Master

Burn Down/Up Chart

Daily Standup Meeting

24 Hour Sprint

1-4 Week Sprint

Sprint Review

1	
2	Prioritized list of what is required: features, bugs to fix...
3	
4	
5	
6	
7	
8	

Product Backlog

Team selects starting at top as much as it can commit to deliver by end of Sprint

Sprint Planning Meeting

Task Breakout

Sprint Backlog

Sprint end date and team deliverable do not change

Finished Work

Sprint Retrospective

neon rain interactive

AGILE FOR ALL
Making Agile a Reality®

Stand-up

◎ During a stand-up these things happen:
- Review progress on sprint goals
- Identify barriers, questions, or impediments
- Adapt the backlog as necessary
- Clarifying and documenting discussions in the GitHub issues ("user stories")
- Produces an actionable plan for the next day

Activity 5

Stand-up

Sprint Planning

◎ We will do a mock stand-up for the instructors' project
◎ Each group will then do a mock-stand-up

5.

Sprint Retrospective

A short introduction

The Agile Scrum Framework at a glance

Sprint Retrospective

One of the scrum ceremonies:

- ◎ Sprint planning meeting
- ◎ Stand-ups
- ◎ Sprint review\product demos
- ◎ Sprint retrospective

The sprint retrospective is what will help your future sprints get better and better!

It's how your team customizes the scrum framework so that it works for YOU.

Five steps for Sprint Retrospective

1. Prepare
2. Set the stage
3. What went well?
   a. Find themes
4. What needs improvement?
   a. Find themes
5. Wrap up and identify action items. Agree on next steps.

Let's try it

Think about the last time you used **a group project** to teach…

What went well?

Retro Jamboard - use blue sticky notes!

(bit.ly/scrumRetro)

Let's try it

Think about the last time you used **a group project** to teach…

What needs improvement?

Retro Jamboard **- use orange sticky notes!**

(bit.ly/scrumRetro)

Activity 6

Map Scrum to Semester

Planning: What would this look like in your course?

◎ Take out your syllabi
◎ Consider the steps in the Agile/Scrum Process:
  ○ Creating a Sprint Backlog
  ○ Sprint Planning Meetings
  ○ Stand up meetings
  ○ Sprint Retrospective
◎ How might these steps could fit into your schedule?
  ○ Take some time to consider this
  ○ Pair up in your groups to discuss

Discussion

◎ What barriers do you foresee in the scheduling?
◎ What concerns/questions do you have based on your:
  ○ Current schedule
  ○ Curriculum
  ○ Other constraints

6.

Student Perspective

Student Panel Discussion

7.

Lessons Learned

Faculty Coordinator Panel Discussion

Lesson Learned: Overview

◎ Initial cohorts completed projects in a "divide and conquer" style
- ○ Later cohorts were able to use the agile/scrum framework with more faculty involvement

◎ This method requires faculty involvement
- ○ At the start, but can lessen over the term

◎ Overall, students learned to be more autonomous
- ○ Learned organization and collaboration skills

Questions?

End Survey

THANK YOU

Nicholas J. Horton

nhorton@amherst.edu

AMHERST COLLEGE

Randi Garcia

rgarcia@smith.edu

SMITH COLLEGE

Chelsey Legacy

legac006@umn.edu

UNIVERSITY OF MINNESOTA