The following work for the GitHub course template was completed under the PIT-UN grant:

Creation of generic course template

The Intro to Data Science course material was packaged as a course template that is public and freely available on GitHub:

https://github.com/DS4PS/intro-data-science-TEMPLATE

All of the ASU/GSU branding was stripped so that anyone can clone the course and customize it for their own school. All of the lab and project instructions and data are included in the repo – solutions are available via invite to a private repository:

https://github.com/DS4PS/intro-data-science-SOLUTIONS

Modularization of Intro Courses

The Intro to Data Science course material was heavily revised and reorganized from content type folders into self-contained units with clear delineation of content to make it easy for new faculty to navigate or adapt. For example, one unit might contain the following:

- topic-02x01-operators rmd/html
- topic-02x02-groups rmd/html
- vid-02x01-operators.mp4
- vid-02x02-vectors.mp4
- practice-02x01-logical-statements.R
- lab-02-instructions rmd/html
- lab-02-data-dictionary rmd/html

Every unit folder includes the data files and helper scripts necessary for the lectures and labs.

Preparation of tutorials on using open-source course shells

The README.md file in the course template repo on GitHub contains detailed instructions on configuring and hosting your own course using the Jekyll template:

https://github.com/DS4PS/intro-data-science-TEMPLATE/blob/main/README.md

We worked directly with two new instructors to test the template during the process of delivering the course to students. Both instructors were able to configure their own classrooms and utilize the material effectively.

Organization and documentation of project files

Course projects were more thoroughly documented, and templates for labs and dashboards were created to facilitate content delivery:

https://lalmada1.github.io/Intro-to-Data-Science-Econ-4970/templates/

Completion of R package to distribute lab datasets

A large number of custom datasets were created for the delivery of program content, including a number of synthetic datasets that allow instructors to replicate the analysis from published studies using simulated data that is free from privacy concerns and licensing constraints. For example:

https://watts-college.github.io/cpp-524-fall-2021/labs/lab-05-diff-in-diff.html

These functions have been made available as a draft package:

https://github.com/DS4PS/pe4ps

And the process of generating synthetic datasets for lectures and labs is documented in several of the course repositories:

https://github.com/DS4PS/pe4ps/blob/master/DataDocumentation.pdf

Creation of demo files to pilot shared assessment pools using R Exams and other tools for automated exercises and exams

Labs are designed as pass or fail assessments to create a productive learning environment that allows students to embrace challenging new material without the pressure of high-stakes assessments. Content is then reinforced in several ways, including regular practice exercises that review topics and explore nuance on specific questions:
https://watts-college.github.io/cpp-527-fall-2021/practice/

And also as low-stakes quizzes that help students assess whether they have grasped the core skills presented on labs. We piloted the use of R Exam questions, which allow instructors to incorporate auto-graded quiz modules into Canvas or other learning platforms:

https://github.com/DS4PS/intro-to-data-science-QUIZZES