

# Homework 1

## Instructions

This assignment is entirely about Git, GitHub, and data management. The goal is to give you a chance to practice wrangling and tidying data. We do this very early in the class because we will start doing some empirical analysis using real data soon. The faster you are comfortable with the datasets, the better. For more detailed instructions on how to submit your homework answers, please see the overview page [here](#).

Since this first homework assignment is a little unique and more about your software setup, homework 1 only consists of two submission stages instead of three. The due date for initial submission is **1/22**, and the final due date is **1/23**.

## Setup

The first half of this assignment is entirely about finalizing your research environment, including version control in Git, setting up your GitHub repository, committing and pushing changes to your GitHub repository, and organizing your project folders. You'll need to implement all of this using our designated high-performance computing resources via Open OnDemand. These steps are worth 9 points at each submission stage. Once you have your software setup complete, move on to the next part of the assignment.

## Building the data

With your setup in place, let's do some very minor data management (import, cleaning, and merging). You should first review this [Medicare Advantage GitHub Repo](#). This repo has a lot more than we need for this first homework assignment, but it provides a general overview of the data. If you're working from the code files in this repo, you'll need to make some simplifications to address the specific questions below.

All we need to do for this first homework assignment is organize the [Monthly Plan Enrollment Data](#), which includes data on plan and contract types, and the [Service Area](#) files, focusing

specifically on **2018**. Note that the raw enrollment data are monthly, so you'll need to combine all of the months and collapse to a single plan-county-year. You'll need to merge the service area data into the monthly enrollment data using an "inner merge" (i.e., take only those rows that match between the datasets).

Once you've created this dataset, answer the following:

1. Provide a table of the count of plans under each plan type. Your table should look something like Table 1. (2 points)

Table 1: Plan Count by Year

	2018
Type 1	47
Type 2	31
Type 3	24

2. Remove all special needs plans (SNP), employer group plans (eghp), and all "800-series" plans. Provide an updated version of Table 1 after making these exclusions. (2 points)
3. Now repeat the same filters from part 2 but also focus only on counties in which plans are approved as per the service area files. With these data, provide a table of the average enrollments for each plan type. (2 points)