

Homework 1

Instructions

This assignment is entirely about 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 **1/29**, and the final due date is **1/31**.

Setup

The first half of this assignment is entirely about version control in Git, setting up your GitHub repository, committing and pushing changes to your GitHub repository, and organizing your project folders. 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

Most of your professional lives will likely involve managing data. It can be tedious but also extremely rewarding when you finally get to find out what's going on in the analysis stage. Anyway, let's get to work! For this homework, you should first take a look at our [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](#) and the [Service Area](#) files, focusing specifically on 2015. Note that the enrollment data

are monthly, so you'll need to either select a given month of data to work with or 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

	2015
Type 1	39
Type 2	21
Type 3	41

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. Provide a table of the average enrollments for each plan type in 2015. (2 points)