

Homework 1

Research Methods, Spring 2024

Leila Mulveny

[Homework 1: Repository](#)

Enrollment Data

Answer the following based on the enrollment data:

1. How many observations exist in your current dataset?

After creating the “full.ma.data” object which represents the enrollment data from 2010-2015, counting the total number of plans in the full data set yields 13,276,162 observations. So, there are 13,276,162 unique combinations of contract/plan/county/year.

2. How many different *plan_types* exist in the data?

Creating a table of the unique plan types and summarizing the counts of each unique value in the “plan_type” column in the “full.ma.data” object shows that we have 18 total plan types.

3. Provide a table of the count of plans under each plan type in each year.

See Table 1.

Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be resized.

Table 1: Plan types by year

Plan Type	2010	2011	2012	2013	2014	2015
Medicare Prescription Drug Plan	893,609	771,694	815,223	826,907	1,122,209	991,457
HMO/HMOPOS	506,802	528,473	507,272	530,909	523,304	479,275
Local PPO	417,551	515,700	636,701	633,884	664,716	704,993
PFFS	385,733	45,781	36,423	31,919	24,905	13,658
Employer/Union Only Direct Contract PDP	28,700	28,697	28,669	25,526	25,528	25,630
Regional PPO	24,442	22,773	21,602	19,970	19,773	17,578
1876 Cost	6,035	6,851	7,633	7,731	7,069	7,157
HCPP - 1833 Cost	3,604	11	11	10	9	9
Employer/Union Only Direct Contract PFFS	3,332	3,329	3,323	0	0	0
National PACE	717	781	858	953	1,118	1,216
Continuing Care Retirement Community	142	0	0	0	0	0
MSA	135	6,421	6,416	6,431	6,449	6,518
PSO (State License)	123	176	171	0	0	0
ESRD I	117	0	0	0	0	0
Pilot	53	3	3	2	2	2
ESRD II	8	0	0	0	0	0
Medicare-Medicaid Plan HMO/HMOPOS	0	0	0	265	1,319	4,130

4. Remove all special needs plans (SNP), employer group plans (eghp), and all “800-series” plans. Provide an updated table after making these exclusions.

To remove all special needs plans (SNP), employer group plans (eghp), and all “800-series” plans I filtered the ‘full.ma.data’ under those restrictions and created a new final.plans object for which the second table of plan types displays. Table 2

Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be resized.

Table 2: Revised plan types by year

Plan Type	2010	2011	2012	2013	2014	2015
Medicare Prescription Drug Plan	391,205	295,458	289,044	278,091	301,082	269,153
PFFS	54,119	22,038	17,449	12,945	6,053	4,232
HMO/HMOPOS	34,460	33,931	37,551	37,179	38,893	36,588
0	29,733	0	0	0	0	0
Local PPO	11,652	13,874	17,030	17,089	17,169	16,728
Regional PPO	10,659	10,995	11,279	9,660	10,420	8,531
1876 Cost	4,923	5,829	6,647	6,759	6,207	6,329
National PACE	717	781	858	953	1,118	1,216
ESRD I	117	0	0	0	0	0
PSO (State License)	97	141	143	0	0	0
MSA	68	131	132	145	163	232
Continuing Care Retirement Community	64	0	0	0	0	0
Medicare-Medicaid Plan HMO/HMOPOS	0	0	0	265	1,319	4,130

5. Merge the the contract service area data to the enrollment data and restrict the data only to contracts that are approved in their respective counties. Limit your dataset only to plans with non-missing enrollment data. Provide a graph showing the average number of Medicare Advantage enrollees per county from 2010 to 2015.

To merge the contract service area data to the enrollment data is to use the inner join operation that combines rows from both datasets based on a related column between them (contractid, fips, year). To limit the dataset to plans with non-missing enrollment data I filtered out the missing enrollment data just using the remove function. The resulting dataset forms the graph of average enrollments per country from 2010-2015. Figure 1.

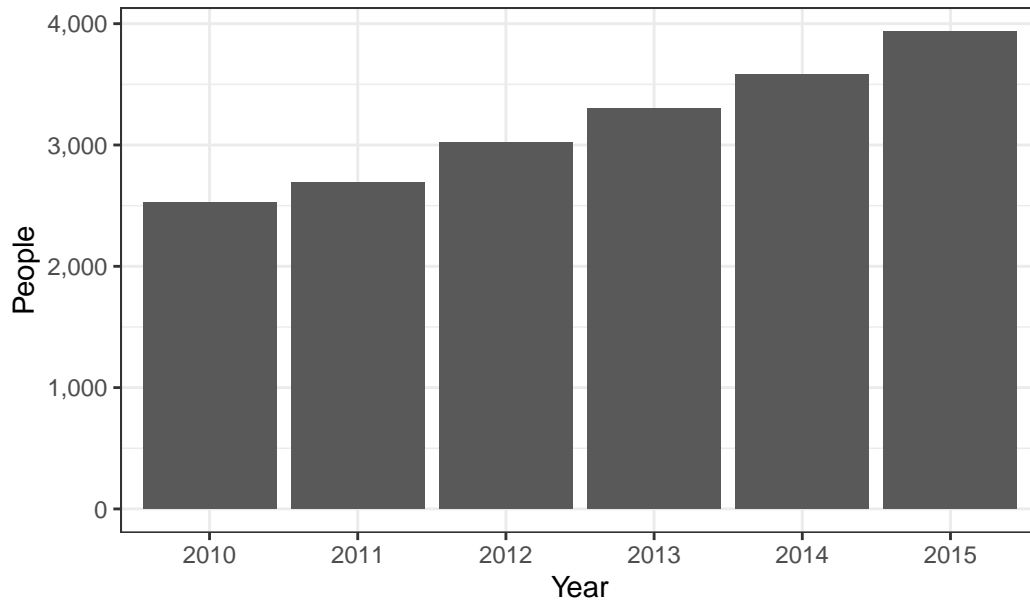


Figure 1: Average Enrollment

Premium Data

6. Merge the plan characteristics data to the dataset you created in Step 5 above. Provide a graph showing the average premium over time.

To merge in the market penetration data I used the left join operation which combines the rows from “final.data” with the rows from the modified “ma.penetration.data” based on the common columns specified by the “fips” and “year” argument. The rows from “final.data” are thus retained in the result and if there are matching rows in “ma.penetration.data”, those values are included. Then we fill in the state information by creating a table of unique state names and then merging it back into the original data. After reading in the premium data, we can merge the premium data into the final data set to form the graph of average premiums over time. The spike in premiums in 2014 occurs because all of the \$0 premiums for 2014 were recorded as “missing” and therefore excluded from the data, inflating the average premium value for that year. [Figure 2](#).

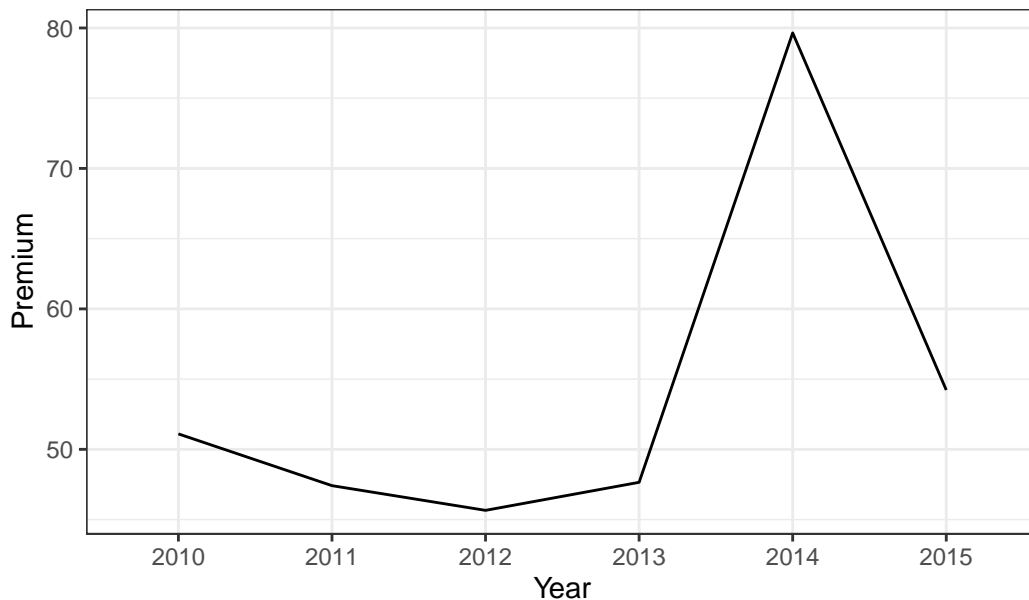


Figure 2: Average Premiums

7. Provide a graph showing the percentage of \$0 premium plans over time.

The graph of the percentage of \$0 premium plans over time, Figure 3, confirms the reasoning behind the spike in premiums in Figure 2.

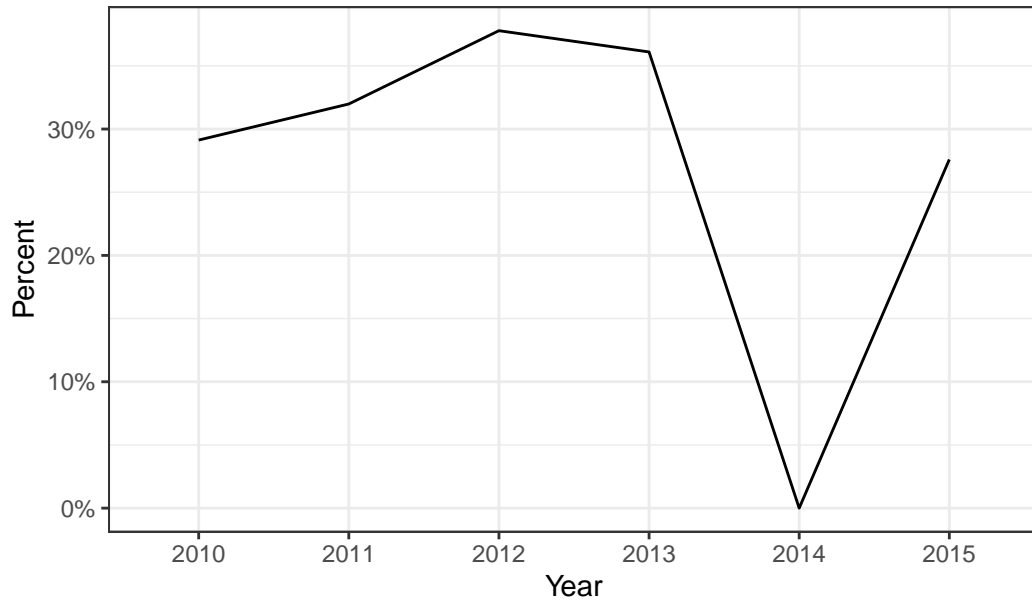


Figure 3: Share of 0 premium plans

Summary Questions

8. Why did we drop the “800-series” plans?

We drop the “800-series” (“Employer Group Waiver Plans”) plans because they are not available to all people. Summaries including these plans are not reflective of the average enrollee’s experience in Medicare Advantage, so we drop them.

9. Why do so many plans charge a \$0 premium? What does that really mean to a beneficiary?

The \$0 premium is really just a reflection of an added premium above Part B of Medicare. Medicare Advantage plans (Part C) provide the same coverage as original Medicare (Part A and Part B), but with additional benefits. Beneficiaries are still paying premiums on Part B, but a \$0 premium just means that they are not paying premiums in excess of the Part B premium.

10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated you.

This data analysis has been on a much larger scale than my previous ECON lab assignments, so there was definitely a learning curve in trying to understand what is going on in the background of my computer when I am creating these datasets. I learned the left join and inner join operations, for which the in-class explanation helped me to intuitively think about/understand what is happening when carrying out the operation in merging datasets. I think where I was getting the most frustrated was in the simple things like making sure my file paths were consistent, and having to frequently check my current working directory and then change it to be able to execute any of the code. However I think that I came to understand what I was actually doing through the trial and error process, and debugging those simple directory-related error messages.