

Analysis

Link to Github

<https://github.com/safiaread/homework-1>

Enrollment Data

1. How many observations exist in your current dataset?

```
[1] 13276162
```

2. How many different plan_types exist in the data?

```
[1] "HCPP - 1833 Cost"
[2] "Employer/Union Only Direct Contract PDP"
[3] "Employer/Union Only Direct Contract PFFS"
[4] NA
[5] "Local PPO"
[6] "National PACE"
[7] "HMO/HMOPOS"
[8] "Continuing Care Retirement Community"
[9] "PFFS"
[10] "1876 Cost"
[11] "Pilot"
[12] "ESRD I"
[13] "PSO (State License)"
[14] "ESRD II"
[15] "MSA"
[16] "Regional PPO"
[17] "Medicare Prescription Drug Plan"
[18] "Medicare-Medicaid Plan HMO/HMOPOS"
```

3. Provide a table of the count of plans under each plan type in each year. Your table should look something like Table 1.

```
# A tibble: 18 × 7
# Groups:   plan_type [18]
  plan_type      `2010` `2011` `2012` `2013` `2014` `2015`
  <chr>      <int> <int> <int> <int> <int> <int>
1 1876 Cost      6035  6851  7633  7731  7069  7157
2 Continuing Care Retirement Commun...    142    NA    NA    NA    NA    NA
3 ESRD I         117    NA    NA    NA    NA    NA
4 ESRD II          8    NA    NA    NA    NA    NA
```

5	Employer/Union Only Direct Contra...	28700	28697	28669	25526	25528	25630
6	Employer/Union Only Direct Contra...	3332	3329	3323	NA	NA	NA
7	HCPP - 1833 Cost	3604	11	11	10	9	9
8	HMO/HMOPOS	506802	528473	507272	530909	523304	479275
9	Local PPO	417551	515700	636701	633884	664716	704993
10	MSA	135	6421	6416	6431	6449	6518
11	Medicare Prescription Drug Plan	893609	771694	815223	826907	1122209	991457
12	Medicare-Medicaid Plan HMO/HMOPOS	NA	NA	NA	265	1319	4130
13	National PACE	717	781	858	953	1118	1216
14	PFFS	385733	45781	36423	31919	24905	13658
15	PSO (State License)	123	176	171	NA	NA	NA
16	Pilot	53	3	3	2	2	2
17	Regional PPO	24442	22773	21602	19970	19773	17578
18	<NA>	277533	NA	NA	NA	NA	NA

4. 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.

```
# A tibble: 15 × 7
# Groups:   plan_type [15]
```

plan_type	`2010`	`2011`	`2012`	`2013`	`2014`	`2015`
<chr>	<int>	<int>	<int>	<int>	<int>	<int>
1 1876 Cost	4952	5838	6650	6761	6207	6329
2 Continuing Care Retirement Communi...	64	NA	NA	NA	NA	NA
3 ESRD I	117	NA	NA	NA	NA	NA
4 HCPP - 1833 Cost	3604	11	11	10	9	9
5 HMO/HMOPOS	34460	33931	37551	37179	38893	36588
6 Local PPO	11652	13874	17030	17089	17169	16728
7 MSA	68	131	132	145	163	232
8 Medicare Prescription Drug Plan	391205	295458	289044	278091	301082	269153
9 Medicare-Medicaid Plan HMO/HMOPOS	NA	NA	NA	265	1319	4130
10 National PACE	717	781	858	953	1118	1216
11 PFFS	54119	23812	17449	12945	6053	4232
12 PSO (State License)	97	141	143	NA	NA	NA
13 Pilot	53	3	3	2	2	2
14 Regional PPO	10659	10995	11279	9660	10420	8531
15 <NA>	277533	NA	NA	NA	NA	NA

5. Merge the contract service area data to the enrollment data, and restrict the data only to contracts that are approved in their respective counties. The R script to create the service area dataset is here: [Contract Service Area](#). And you can follow the [_BuildFinalData.R](#) script to see where/how I join the datasets. Limiting 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. Be sure to format your graph in a meaningful way.

```
# A tibble: 6 × 18
# Groups:   plan_type [1]
  plan_type `2010` `2011` `2012` `2013` `2014` `2015` contractid org_name
  <chr>      <int> <int> <int> <int> <int> <int> <chr>      <chr>
1 1876 Cost  4952  5838  6650  6761  6207  6329 H0502  CONTRA COSTA H...
2 1876 Cost  4952  5838  6650  6761  6207  6329 H0602  ROCKY MOUNTAIN...
3 1876 Cost  4952  5838  6650  6761  6207  6329 H0602  ROCKY MOUNTAIN...
4 1876 Cost  4952  5838  6650  6761  6207  6329 H0602  ROCKY MOUNTAIN...
5 1876 Cost  4952  5838  6650  6761  6207  6329 H0602  ROCKY MOUNTAIN...
6 1876 Cost  4952  5838  6650  6761  6207  6329 H0602  ROCKY MOUNTAIN...
# i 9 more variables: org_type <chr>, partial <lgl>, eghp <chr>, ssa <dbl>,
# fips <dbl>, county <chr>, state <chr>, notes <chr>, year <int>
```

Premium Data

6. I am still working on how to merge this data.

```
#
```

7. I am still working on the graph for this data.

```
#
```

Summary Questions

8. I couldn't find what variable stored the information about 800 series data, so I am unsure about why we would omit them. Perhaps they contained a lot of NAs.
9. Maybe they charged \$0 because they had really high deductibles. They also could have contracts with a small network of providers with high out of pocket costs.
10. With this data, it was definitely an adjustment to manipulate it in VSCode. I have previously used RStudio, and I think you were right that this is a less intuitive interface. I had trouble figuring out how to view the data to get a grasp of how to manipulate it. I also have never worked with a dataset so large before, so it took some time to figure out how to upload it to my computer and organize everything so my computer would run. I'm still having some trouble figuring out the best ways to view the variables and merge them, and the relative paths aren't working for now.