ECON 470 Homework 1

Srijon Sarkar

2026-01-23

Setup: - Created the hwk1 repository with a `.gitignore` file to avoid large data file commits, and synced the SSH key from OpenDemand onto GitHub. - Structured the hwk1 directory into subfolders of the first and second submission, with analysis, data-code, and results. (The second one primarily contains the main work) - Extracted the data only for 2018 as desired from enrollment and service area data of the large MA dataset. (Codes under submission 2 folder, data-code subfolder).

```r
library("rmarkdown")
library("tidyverse")
library("dplyr")
```

```
  Attaching core tidyverse packages              tidyverse 2.0.0
 dplyr     1.1.3      readr     2.1.4
 forcats   1.0.0      stringr   1.5.0
 ggplot2   3.4.4      tibble    3.2.1
 lubridate 1.9.3      tidyr     1.3.0
 purrr     1.0.2
  Conflicts                        tidyverse_conflicts()
 dplyr::filter() masks stats::filter()
 dplyr::lag()    masks stats::lag()
 Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
setwd('/home/ssark38/econ470/a0/work/hwk1')
```

```r
plan_data <- read_csv("data/plan_data.csv")
```

```
Rows: 2475118 Columns: 23
  Column specification
Delimiter: ","
```

```
chr (13): contractid, state, county, org_type, plan_type, partd, snp, eghp, ...
dbl (10): planid, fips, year, n_nonmiss, avg_enrollment, sd_enrollment, min_...

  Use `spec()` to retrieve the full column specification for this data.
  Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
service_data <- read_csv("data/service_data.csv")
```

```
Rows: 331593 Columns: 12
  Column specification
Delimiter: ","
chr (7): contractid, state, county, org_name, org_type, plan_type, notes
dbl (3): fips, year, ssa
lgl (2): partial, eghp

  Use `spec()` to retrieve the full column specification for this data.
  Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Exercise 1

Here we count the number of plans under each plan type:

```
table1 <- plan_data %>%
    distinct(contractid, planid, plan_type) %>%
    count(plan_type, name = "n_plans")

table1
```

A tibble: 11 × 2

plan_type <chr>	n_plans <int>
1876 Cost	101
Employer/Union Only Direct Contract PDP	3
HCPP - 1833 Cost	9
HMO/HMOPOS	2678
Local PPO	966
MSA	5
Medicare Prescription Drug Plan	1011
Medicare-Medicaid Plan HMO/HMOPOS	54
National PACE	258

plan_type <chr>	n_plans <int>
PFFS	50
Regional PPO	109

Exercise 2

Now, we remove all special needs plans (SNP), employer group plans (eghp), and all “800-series” plans as follows:

```r
filtered_plans <- plan_data %>%
    filter(
        snp == "No",
        eghp == "No",
        !(planid >= 800 & planid < 900)
    )
```

```r
updated_table1 <- filtered_plans %>%
    distinct(contractid, planid, plan_type) %>%
    count(plan_type, name = "n_plans")

updated_table1
```

A tibble: 9 × 2

plan_type <chr>	n_plans <int>
1876 Cost	93
HMO/HMOPOS	1569
Local PPO	569
MSA	3
Medicare Prescription Drug Plan	794
Medicare-Medicaid Plan HMO/HMOPOS	54
National PACE	258
PFFS	46
Regional PPO	49

Exercise 3

Finally, before merge we filter the data county-wise in the service area to avoid double-counting in the mean calculation, as a plan spans over several months

```r
filtered_service <- service_data %>%
    filter(!is.na(fips)) %>%
    distinct(contractid, fips, year)
```

As the contract ID uniquely identifies the plan group and FIPS assists in the choice of counties, we use those as primary keys for the merge with year for 2018 as a precautionary measure, though only 2018 data was extracted.

```r
merged_data <- filtered_plans %>%
  inner_join(filtered_service,
             by = c("contractid", "fips", "year"))
```

```r
glimpse(merged_data)
```

```
Rows: 87,672
Columns: 23
$ contractid          <chr> "H0022", "H0022", "H0022", "H0022", "H0022", "H0022…
$ planid              <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 4, 4, 8, 8, …
$ fips                <dbl> 39023, 39035, 39051, 39055, 39057, 39085, 39093, 39…
$ year                <dbl> 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2018, 201…
$ n_nonmiss           <dbl> 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,…
$ avg_enrollment      <dbl> 598.41667, 3653.00000, 115.83333, 77.33333, 571.083…
$ sd_enrollment       <dbl> 23.592982, 92.164872, 6.873312, 4.658001, 24.396193…
$ min_enrollment      <dbl> 558, 3549, 107, 68, 539, 278, 531, 2776, 180, 3115,…
$ max_enrollment      <dbl> 638, 3829, 126, 84, 618, 341, 591, 3006, 217, 3351,…
$ first_enrollment    <dbl> 558, 3596, 107, 80, 539, 278, 559, 2782, 192, 3128,…
$ last_enrollment     <dbl> 622, 3657, 126, 80, 601, 324, 535, 2892, 205, 3278,…
$ state               <chr> "OH", "OH", "OH", "OH", "OH", "OH", "OH", "OH", "OH…
$ county              <chr> "Clark", "Cuyahoga", "Fulton", "Geauga", "Greene", …
$ org_type            <chr> "Demo", "Demo", "Demo", "Demo", "Demo", "Demo", "De…
$ plan_type           <chr> "Medicare-Medicaid Plan HMO/HMOPOS", "Medicare-Medi…
$ partd               <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Y…
$ snp                 <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No…
$ eghp                <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No…
$ org_name            <chr> "BUCKEYE COMMUNITY HEALTH PLAN, INC.", "BUCKEYE COM…
$ org_marketing_name  <chr> "Buckeye Health Plan - MyCare Ohio", "Buckeye Healt…
$ plan_name           <chr> "Buckeye Health Plan - MyCare Ohio (Medicare-Medica…
$ parent_org          <chr> "Centene Corporation", "Centene Corporation", "Cent…
$ contract_date       <chr> "05/01/2014 0:00:00", "05/01/2014 0:00:00", "05/01/…
```

```r
write_csv(merged_data, "data/intermediate_data.csv")
```

Finally, considering counties in which plans are approved as per the service area files, we get a table of the average enrollments for each plan type

```r
table2 <- merged_data %>%
  group_by(plan_type) %>%
  summarise(avg_enrollment = mean(avg_enrollment, na.rm = TRUE),
            .groups="drop") %>%
  arrange(desc(avg_enrollment))

table2
```

A tibble: 8 × 2

plan_type <chr>	avg_enrollment <dbl>
Medicare-Medicaid Plan HMO/HMOPOS	989.16876
HMO/HMOPOS	755.54963
Local PPO	330.62289
1876 Cost	251.56522
Regional PPO	188.78840
National PACE	144.32795
PFFS	93.65923
MSA	58.13192