Homework 1 - Research in Health Economics

Ammarah Ahmed

2023-01-25

```
pacman::p_load(tidyverse, ggplot2, dplyr, lubridate, stringr, readxl, data.table, gdata, readr)
```

#Enrollment Data

```
full.ma.data<- readRDS("data/output/full_ma_data.rds")
```

1. There are 19,126,783 observations in the dataset.
2. There are 26 different plans in the dataset.
3. I was unable to display the table as the code wouldn't run

```
full.ma.data %>%
  group_by(year, plan_type)%>%
  summarize(n_under_plan_type = n())%>%
  spread(year, n_under_plan_type)
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 27 x 10
##    plan_type      `2007` `2008` `2009` `2010` `2011` `2012` `2013` `2014` `2015`
##    <chr>           <int>  <int>  <int>  <int>  <int>  <int>  <int>  <int>  <int>
##  1 1876 Cost        5855   5459   5825   6035   6851   7633   7731   7069   7157
##  2 Continuing Ca~     95    122    158    142     NA     NA     NA     NA     NA
##  3 Employer Dire~   3247     NA     NA     NA     NA     NA     NA     NA     NA
##  4 Employer/Unio~  32358  29113  25860  28700  28697  28669  25526  25528  25630
##  5 Employer/Unio~     NA   3332   3335   3332   3329   3323     NA     NA     NA
##  6 ESRD I             75    122    123    117     NA     NA     NA     NA     NA
##  7 ESRD II            12     12      7      8     NA     NA     NA     NA     NA
##  8 HCPP - 1833 C~     13     13   3938   3604     11     11     10      9      9
##  9 HMO/HMOPOS      60012  70176 479978 506802 528473 507272 530909 523304 479275
## 10 Local PPO       17427  38470 405197 417551 515700 636701 633884 664716 704993
## # ... with 17 more rows
```

```
full.ma.data2<- full.ma.data %>%
  filter(snp == 'No' & eghp == 'No' & !(planid %in% 800:899))
```

```
full.ma.data2 %>%
  group_by(year, plan_type)%>%
  summarize(n_under_plan_type = n())%>%
  spread(year, n_under_plan_type)
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

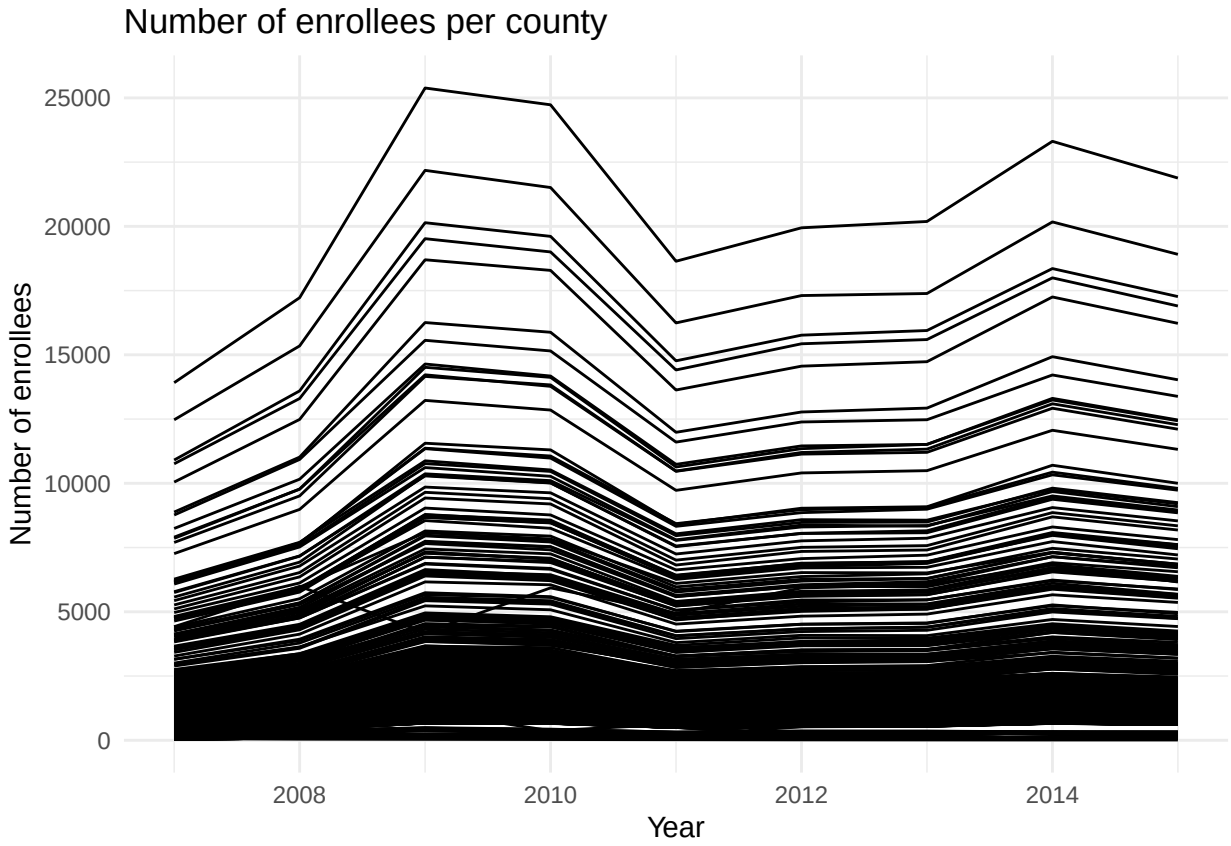
```
## # A tibble: 19 x 10
##    plan_type      `2007` `2008` `2009` `2010` `2011` `2012` `2013` `2014` `2015`
```

```
##    <chr>           <int>  <int>  <int>  <int>  <int>  <int>  <int>  <int>  <int>
##  1 1876 Cost        5074   4595   4845   4952   5838   6650   6761   6207   6329
##  2 Continuing Ca~     68     66     60     64     NA     NA     NA     NA     NA
##  3 ESRD I             75    122    123    117     NA     NA     NA     NA     NA
##  4 HCPP - 1833 C~     13     13   3938   3604     11     11     10      9      9
##  5 HMO/HMOPOS      30670  34545  36166  34460  33931  37551  37179  38893  36588
##  6 Local PPO        6116   7612   9929  11652  13874  17030  17089  17169  16728
##  7 Medicare Pres~ 398167 428936 415027 391205 295458 289044 278091 301082 269153
##  8 Medicare-Medi~     NA     NA     NA     NA     NA     NA    265   1319   4130
##  9 MSA              2177   3303   2459     68    131    132    145    163    232
## 10 MSA Demo          129     NA     NA     NA     NA     NA     NA     NA     NA
## 11 National PACE     405    548    616    717    781    858    953   1118   1216
## 12 PFFS            51987 105859  89586  54119  22038  17449  12945   6053   4232
## 13 Pilot              15     12    201     53      3      3      2      2      2
## 14 PSO (Federal ~    110     NA     NA     NA     NA     NA     NA     NA     NA
## 15 PSO (State Li~    376    394     75     97    141    143     NA     NA     NA
## 16 Regional PPO     7254   7794   8470  10659  10995  11279   9660  10420   8531
## 17 RFB PFFS           NA     NA   3006     NA     NA     NA     NA     NA     NA
## 18 SHMO              458     NA     NA     NA     NA     NA     NA     NA     NA
## 19 <NA>               NA     NA  13619  29733     NA     NA     NA     NA     NA
```

```
final.data1<- readRDS("data/output/full_ma_data.rds")
```

```
final.data1 %>%
  group_by(year,county)%>%
  summarize(count = n())%>%
  ggplot(aes( x = year, y  = count, group = county))+
  geom_line()+
  labs( title = 'Number of enrollees per county', x = 'Year', y = 'Number of enrollees')+
  theme_minimal()
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```



#Premium Data I was unable to figure out how to display the graphs from r script in the r markdown

“‘