```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.
```

```r
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

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

```r
filtered_plans <- plan_data %>%
    filter(
        snp == "No",
        eghp == "No",
        !str_detect(planid, "^8")
    )
```

```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	89
HMO/HMOPOS	1498
Local PPO	529
MSA	3
Medicare Prescription Drug Plan	771
Medicare-Medicaid Plan HMO/HMOPOS	54
National PACE	258
PFFS	41
Regional PPO	48

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

```
Rows: 1364770 Columns: 27
  Column specification
Delimiter: ","
chr (14): contractid, state, county, org_type, plan_type, partd, snp, eghp.x...
dbl (11): planid, fips, year, n_nonmiss, avg_enrollment, sd_enrollment, min_...
lgl  (2): partial, eghp.y

  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.
```

```r
filtered_plans_2 <- merged_data %>%
    filter(
        snp == "No",
        eghp.x == "No",
        !str_detect(planid, "^8")
    )
```

```r
table2 <- filtered_plans_2 %>%
  group_by(plan_type) %>%
  summarise(avg_enrollment = mean(avg_enrollment, na.rm = TRUE),
            .groups = "drop") %>%
  arrange(desc(avg_enrollment))
```

```r
table2
```

A tibble: 8 × 2

plan_type <chr>	avg_enrollment <dbl>
Medicare-Medicaid Plan HMO/HMOPOS	989.16876
HMO/HMOPOS	762.29813
Local PPO	332.34328
1876 Cost	258.61757
Regional PPO	179.17024
National PACE	144.32795
PFFS	96.25695
MSA	58.13192