

Task 1 Question 2

Meredith Gavin

7/14/2023

PREP WORK

Libraries

```
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6          v purrr  0.3.4
## v tibble  3.1.8          v stringr 1.4.0.9000
## v tidyr   1.2.0.9000     v forcats 0.5.1
## v readr   2.1.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':  
##  
##      set_names
```

```
## The following object is masked from 'package:tidyr':  
##  
##      extract
```

```
library(readr)  
library(ggplot2)  
library(lubridate)
```

```
##  
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':  
##  
##      date, intersect, setdiff, union
```

Set Working Directory

```
setwd("~/Desktop/pre_doc_task_1")
```

Load Data

```
ma_enrollment <- read_csv("scp_1205.csv", col_names = FALSE)
```

```
## Rows: 38013 Columns: 10  
## -- Column specification -----  
## Delimiter: ","  
## chr (7): X1, X2, X3, X4, X5, X6, X10  
## dbl (3): X7, X8, X9  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

QUESTION 2: DATA MANIPULATION

Rename Headers

```
ma_enrollment <- ma_enrollment %>%  
  rename(countyname = X1,  
         state = X2,  
         healthplanname = X3,  
         typeofplan = X4,  
         countyssa = X5,
```

```

eligibles = X6,
enrollees = X7,
penetration = X8,
ABrate = X9)

```

Change NAs to 0 for eligibles, enrollees, and penetration variables and remove Puerto Rico and Guam

```

ma_enrollment <- ma_enrollment %>%
  mutate_at(c("eligibles", "enrollees", "penetration"),
            as.double) %>% # make variables numeric
  mutate_at(c("eligibles", "enrollees", "penetration"),
            ~replace_na(., 0)) %>% # change NAs to zeroes
  filter(state != "GU" & state != "PR") # filter out Puerto Rico and Guam

```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```

# create numberofplans1 variable
ma_enrollment_numberofplans1 <- ma_enrollment %>%
  filter(enrollees > 10) %>%
  group_by(state, countyname) %>%
  summarise(numberofplans1 = n()) %>%
  ungroup()

```

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

```

# create numberofplans2 variable

ma_enrollment_numberofplans2 <-
  ma_enrollment %>%
  filter(penetration > 0.5) %>%
  group_by(state, countyname) %>%
  summarise(numberofplans2 = n())

```

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

Merge new variables to master dataframe

```

ma_enrollment <- ma_enrollment %>%
  left_join(ma_enrollment_numberofplans1,
            by = c("state", "countyname")) %>%
  left_join(ma_enrollment_numberofplans2,
            by = c("state", "countyname"))

```

create totalenrollees variable

```
county_enrollment <- ma_enrollment %>%  
  group_by(state, countyname) %>%  
  summarise(totalenrollees = sum(enrollees)) %>%  
  ungroup()
```

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

merge totalenrollees to master df and create totalpenetration

```
ma_enrollment_final <- ma_enrollment %>%  
  left_join(county_enrollment,  
            by = c("state", "countyname")) %>%  
  mutate(totalpenetration = (totalenrollees / eligibles) * 100) %>%  
  select(  
    countyname,  
    state,  
    numberofplans1,  
    numberofplans2,  
    countyssa,  
    eligibles,  
    totalenrollees,  
    totalpenetration  
  ) %>%  
  arrange(state, countyname)
```

save final dataframe as csv

```
write_csv(ma_enrollment_final, "ma_enrollment_final.csv")
```