

558 Project 1

Susan H. and Holly P.

Load in necessary libraries:

```
library(tidyverse)
library(readr)
```

Read in data using read_csv():

```
census_data <- read_csv("https://www4.stat.ncsu.edu/~online/datasets/EDU01a.csv")
```

Question 1

Select only Area_name, STCOU, and any column that ends in “D”

```
census_data1 <- census_data |> select(Area_name, STCOU, ends_with("D"))

#displaying first 5 rows
head(census_data1, 5)
```

```
# A tibble: 5 x 12
  Area_name      STCOU EDU010187D EDU010188D EDU010189D EDU010190D EDU010191D
  <chr>         <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1 UNITED STATES 00000    40024299   39967624   40317775   40737600   41385442
2 ALABAMA       01000     733735    728234     730048     728252     725541
3 Autauga, AL    01001      6829      6900       6920       6847       7008
4 Baldwin, AL   01003     16417     16465     16799     17054     17479
5 Barbour, AL   01005      5071      5098      5068      5156      5173
# i 5 more variables: EDU010192D <dbl>, EDU010193D <dbl>, EDU010194D <dbl>,
#   EDU010195D <dbl>, EDU010196D <dbl>
```

Question 2

converting data into long format, where we want each Area_name entry to have only one Enrollment value with its own unique ID

```
census_long <- census_data1 |> pivot_longer(cols = ends_with("D"),
      names_to = "EnrollmentID",
      values_to = "Enrollment_Total")

#displaying first 5 rows
head(census_long,5)
```

```
# A tibble: 5 x 4
  Area_name      STCOU EnrollmentID Enrollment_Total
  <chr>          <chr> <chr>              <dbl>
1 UNITED STATES 00000 EDU010187D          40024299
2 UNITED STATES 00000 EDU010188D          39967624
3 UNITED STATES 00000 EDU010189D          40317775
4 UNITED STATES 00000 EDU010190D          40737600
5 UNITED STATES 00000 EDU010191D          41385442
```

Question 3

```
long_updated <- census_long |>
  #pull out the year and convert the year into a numeric
  mutate( Year = as.numeric(substr(EnrollmentID, start = 8, stop = 9))) |>
  #no dates above 1996
  mutate(Year = Year + 1900) |>
  #creating new variable for identifying which measurement was grabbed
  mutate(Measurement = substr(EnrollmentID, start = 1, stop = 7) )

#displaying first 5 rows
head(long_updated, 5)
```

```
# A tibble: 5 x 6
  Area_name      STCOU EnrollmentID Enrollment_Total  Year Measurement
  <chr>          <chr> <chr>              <dbl> <dbl> <chr>
1 UNITED STATES 00000 EDU010187D          40024299  1987 EDU0101
2 UNITED STATES 00000 EDU010188D          39967624  1988 EDU0101
3 UNITED STATES 00000 EDU010189D          40317775  1989 EDU0101
```

```
4 UNITED STATES 00000 EDU010190D      40737600  1990 EDU0101
5 UNITED STATES 00000 EDU010191D      41385442  1991 EDU0101
```

Question 4

```
#County Dataset
row_names <- rownames(long_updated)
#using grep to subset the original data for county
county_subset <- row_names %in% grep(pattern = ", \\w\\w", long_updated$Area_name)
county_tibble <- subset(long_updated, county_subset)
class(county_tibble) <- c("county", class(county_tibble)) #changing class
head(county_tibble, 10)
```

```
# A tibble: 10 x 6
  Area_name STCOU EnrollmentID Enrollment_Total Year Measurement
  <chr>      <chr> <chr>          <dbl> <dbl> <chr>
1 Autauga, AL 01001 EDU010187D      6829  1987 EDU0101
2 Autauga, AL 01001 EDU010188D      6900  1988 EDU0101
3 Autauga, AL 01001 EDU010189D      6920  1989 EDU0101
4 Autauga, AL 01001 EDU010190D      6847  1990 EDU0101
5 Autauga, AL 01001 EDU010191D      7008  1991 EDU0101
6 Autauga, AL 01001 EDU010192D      7137  1992 EDU0101
7 Autauga, AL 01001 EDU010193D      7152  1993 EDU0101
8 Autauga, AL 01001 EDU010194D      7381  1994 EDU0101
9 Autauga, AL 01001 EDU010195D      7568  1995 EDU0101
10 Autauga, AL 01001 EDU010196D      7834  1996 EDU0101
```

```
#State Dataset
state_tibble <- subset(long_updated, !(row_names %in% grep(pattern = ", \\w\\w",
long_updated$Area_name)))
# state is what is not included in the grep for county
class(state_tibble) <- c("state", class(state_tibble))
#changing class
head(state_tibble, 10)
```

```
# A tibble: 10 x 6
  Area_name STCOU EnrollmentID Enrollment_Total Year Measurement
  <chr>      <chr> <chr>          <dbl> <dbl> <chr>
1 UNITED STATES 00000 EDU010187D      40024299  1987 EDU0101
2 UNITED STATES 00000 EDU010188D      39967624  1988 EDU0101
```

3	UNITED STATES	00000	EDU010189D	40317775	1989	EDU0101
4	UNITED STATES	00000	EDU010190D	40737600	1990	EDU0101
5	UNITED STATES	00000	EDU010191D	41385442	1991	EDU0101
6	UNITED STATES	00000	EDU010192D	42088151	1992	EDU0101
7	UNITED STATES	00000	EDU010193D	42724710	1993	EDU0101
8	UNITED STATES	00000	EDU010194D	43369917	1994	EDU0101
9	UNITED STATES	00000	EDU010195D	43993459	1995	EDU0101
10	UNITED STATES	00000	EDU010196D	44715737	1996	EDU0101

Question 5

```
#use mutate to create a new variable for state abbreviation
county_tibble1 <- county_tibble |> mutate(State = substr(county_tibble$Area_name,
start = nchar(Area_name)-1, stop = nchar(Area_name)))
#nchar allows for differing name lengths

county_tibble1
```

A tibble: 31,450 x 7

	Area_name	STCOU	EnrollmentID	Enrollment_Total	Year	Measurement	State
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>
1	Autauga, AL	01001	EDU010187D	6829	1987	EDU0101	AL
2	Autauga, AL	01001	EDU010188D	6900	1988	EDU0101	AL
3	Autauga, AL	01001	EDU010189D	6920	1989	EDU0101	AL
4	Autauga, AL	01001	EDU010190D	6847	1990	EDU0101	AL
5	Autauga, AL	01001	EDU010191D	7008	1991	EDU0101	AL
6	Autauga, AL	01001	EDU010192D	7137	1992	EDU0101	AL
7	Autauga, AL	01001	EDU010193D	7152	1993	EDU0101	AL
8	Autauga, AL	01001	EDU010194D	7381	1994	EDU0101	AL
9	Autauga, AL	01001	EDU010195D	7568	1995	EDU0101	AL
10	Autauga, AL	01001	EDU010196D	7834	1996	EDU0101	AL

i 31,440 more rows

Question 6

Use case_when logic to create state tibble

```

state_tibble1 <- state_tibble |> mutate(Division = case_when(

  Area_name %in% c("CONNECTICUT", "MAINE", "MASSACHUSETTS", "NEW HAMPSHIRE",
    "RHODE ISLAND", "VERMONT") ~ "New England",
  Area_name %in% c("NEW JERSEY", "NEW YORK", "PENNSYLVANIA") ~ "Mid-Atlantic",

  Area_name %in% c("ILLINOIS", "INDIANA", "MICHIGAN", "OHIO", "WISCONSIN")
    ~ "East North Central",

  Area_name %in% c("IOWA", "KANSAS", "MINNESOTA", "MISSOURI",
    "NEBRASKA", "NORTH DAKOTA", "SOUTH DAKOTA")
    ~ "West North Central",

  Area_name %in% c("DELAWARE", "DISTRICT OF COLUMBIA", "District of Columbia",
    "FLORIDA", "GEORGIA", "MARYLAND", "NORTH CAROLINA",
    "SOUTH CAROLINA",
    "VIRGINIA", "WEST VIRGINIA")
    ~ "South Atlantic",

  Area_name %in% c("ALABAMA", "KENTUCKY", "MISSISSIPPI", "TENNESSEE")
    ~ "East South Central",

  Area_name %in% c("ARKANSAS", "LOUISIANA", "OKLAHOMA", "TEXAS")
    ~ "West South Central",

  Area_name %in% c("ARIZONA", "COLORADO", "IDAHO", "NEVADA",
    "MONTANA", "NEW MEXICO", "UTAH", "WYOMING")
    ~ "Mountain",

  Area_name %in% c("ALASKA", "CALIFORNIA", "HAWAII", "OREGON",
    "WASHINGTON") ~ "Pacific", .default = "ERROR"

))

state_tibble1

```

A tibble: 530 x 7

	Area_name	STCOU	EnrollmentID	Enrollment_Total	Year	Measurement	Division
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>
1	UNITED STATES	00000	EDU010187D	40024299	1987	EDU0101	ERROR
2	UNITED STATES	00000	EDU010188D	39967624	1988	EDU0101	ERROR

```

3 UNITED STATES 00000 EDU010189D      40317775  1989 EDU0101      ERROR
4 UNITED STATES 00000 EDU010190D      40737600  1990 EDU0101      ERROR
5 UNITED STATES 00000 EDU010191D      41385442  1991 EDU0101      ERROR
6 UNITED STATES 00000 EDU010192D      42088151  1992 EDU0101      ERROR
7 UNITED STATES 00000 EDU010193D      42724710  1993 EDU0101      ERROR
8 UNITED STATES 00000 EDU010194D      43369917  1994 EDU0101      ERROR
9 UNITED STATES 00000 EDU010195D      43993459  1995 EDU0101      ERROR
10 UNITED STATES 00000 EDU010196D      44715737  1996 EDU0101      ERROR
# i 520 more rows

```

Read in second data set

```
census_data2 <- read_csv("https://www4.stat.ncsu.edu/~online/datasets/EDU01b.csv")
```

Write Function for steps 1 and 2

```

step12func <- function(data, values = "Enrollment_Total") {

  long_data <- data |>
    select(Area_name, STCOU, ends_with("D")) |>
    pivot_longer(cols = ends_with("D"), names_to = "EnrollmentID", values_to = values)

  return(long_data)
}

```

```
step12func(census_data2)
```

```

# A tibble: 31,980 x 4
  Area_name      STCOU EnrollmentID Enrollment_Total
  <chr>          <chr> <chr>          <dbl>
1 UNITED STATES 00000 EDU010197D      44534459
2 UNITED STATES 00000 EDU010198D      46245814
3 UNITED STATES 00000 EDU010199D      46368903
4 UNITED STATES 00000 EDU010200D      46818690
5 UNITED STATES 00000 EDU010201D      47127066
6 UNITED STATES 00000 EDU010202D      47606570
7 UNITED STATES 00000 EDU015203D      48506317
8 UNITED STATES 00000 EDU015204D      48693287
9 UNITED STATES 00000 EDU015205D      48978555
10 UNITED STATES 00000 EDU015206D      49140702
# i 31,970 more rows

```

Write Function for step 3

There are now years after 1999 so we have to change our year mutate function

```
step3func <- function(long_data, values = "Enrollment_Total") {  
  
  long_updated <- long_data |>  
    mutate(Year = as.numeric(substr(EnrollmentID, start = 8, stop = 9))) |>  
    mutate(Year = ifelse(Year > 25, Year + 1900, Year + 2000)) |>  
    mutate(Measurement = substr(EnrollmentID, start = 1, stop = 7))  
  return(long_updated)  
}
```

```
step3func(census_long)
```

A tibble: 31,980 x 6

	Area_name	STCOU	EnrollmentID	Enrollment_Total	Year	Measurement
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>
1	UNITED STATES	00000	EDU010187D	40024299	1987	EDU0101
2	UNITED STATES	00000	EDU010188D	39967624	1988	EDU0101
3	UNITED STATES	00000	EDU010189D	40317775	1989	EDU0101
4	UNITED STATES	00000	EDU010190D	40737600	1990	EDU0101
5	UNITED STATES	00000	EDU010191D	41385442	1991	EDU0101
6	UNITED STATES	00000	EDU010192D	42088151	1992	EDU0101
7	UNITED STATES	00000	EDU010193D	42724710	1993	EDU0101
8	UNITED STATES	00000	EDU010194D	43369917	1994	EDU0101
9	UNITED STATES	00000	EDU010195D	43993459	1995	EDU0101
10	UNITED STATES	00000	EDU010196D	44715737	1996	EDU0101

i 31,970 more rows

Write Function for step 5

```
step5func <- function(tibble) {  
  county_tibble_result <- tibble |> mutate(State = substr(tibble$Area_name,  
    start = nchar(Area_name)-1, stop = nchar(Area_name)))  
  
  return(county_tibble_result)  
}
```

```
step5func(county_tibble1)
```

```
# A tibble: 31,450 x 7
```

	Area_name	STCOU	EnrollmentID	Enrollment_Total	Year	Measurement	State
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>
1	Autauga, AL	01001	EDU010187D	6829	1987	EDU0101	AL
2	Autauga, AL	01001	EDU010188D	6900	1988	EDU0101	AL
3	Autauga, AL	01001	EDU010189D	6920	1989	EDU0101	AL
4	Autauga, AL	01001	EDU010190D	6847	1990	EDU0101	AL
5	Autauga, AL	01001	EDU010191D	7008	1991	EDU0101	AL
6	Autauga, AL	01001	EDU010192D	7137	1992	EDU0101	AL
7	Autauga, AL	01001	EDU010193D	7152	1993	EDU0101	AL
8	Autauga, AL	01001	EDU010194D	7381	1994	EDU0101	AL
9	Autauga, AL	01001	EDU010195D	7568	1995	EDU0101	AL
10	Autauga, AL	01001	EDU010196D	7834	1996	EDU0101	AL

```
# i 31,440 more rows
```

Write Function for step 6

```
step6func <- function(tibble) {  
  state_tibble1 <- tibble |> mutate(Division = case_when(  
  
    Area_name %in% c("CONNECTICUT", "MAINE", "MASSACHUSETTS",  
                     "NEW HAMPSHIRE", "RHODE ISLAND", "VERMONT")  
    ~ "New England",  
    Area_name %in% c("NEW JERSEY", "NEW YORK", "PENNSYLVANIA")  
    ~ "Mid-Atlantic",  
  
    Area_name %in% c("ILLINOIS", "INDIANA", "MICHIGAN", "OHIO",  
                     "WISCONSIN") ~ "East North Central",  
  
    Area_name %in% c("IOWA", "KANSAS", "MINNESOTA", "MISSOURI",  
                     "NEBRASKA", "NORTH DAKOTA", "SOUTH DAKOTA")  
    ~ "West North Central",  
  
    Area_name %in% c("DELAWARE", "DISTRICT OF COLUMBIA",  
                     "District of Columbia", "FLORIDA", "GEORGIA",  
                     "MARYLAND", "NORTH CAROLINA", "SOUTH CAROLINA", "VIRGINIA", "WEST VIRGINIA")  
    ~ "South Atlantic",  
  )  
}
```



```

Area_name %in% c("ALABAMA", "KENTUCKY", "MISSISSIPPI",
                 "TENNESSEE") ~ "East South Central",

Area_name %in% c("ARKANSAS", "LOUISIANA", "OKLAHOMA", "TEXAS")
~"West South Central",

Area_name %in% c("ARIZONA", "COLORADO", "IDAHO",
                 "NEVADA", "MONTANA", "NEW MEXICO",
                 "UTAH", "WYOMING") ~ "Mountain",

Area_name %in% c("ALASKA", "CALIFORNIA", "HAWAII",
                 "OREGON", "WASHINGTON")
~ "Pacific", .default = "ERROR"

))

return(state_tibble1)
}

```

```
step6func(state_tibble)
```

```
# A tibble: 530 x 7
```

	Area_name	STCOU	EnrollmentID	Enrollment_Total	Year	Measurement	Division
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>
1	UNITED STATES	00000	EDU010187D	40024299	1987	EDU0101	ERROR
2	UNITED STATES	00000	EDU010188D	39967624	1988	EDU0101	ERROR
3	UNITED STATES	00000	EDU010189D	40317775	1989	EDU0101	ERROR
4	UNITED STATES	00000	EDU010190D	40737600	1990	EDU0101	ERROR
5	UNITED STATES	00000	EDU010191D	41385442	1991	EDU0101	ERROR
6	UNITED STATES	00000	EDU010192D	42088151	1992	EDU0101	ERROR
7	UNITED STATES	00000	EDU010193D	42724710	1993	EDU0101	ERROR
8	UNITED STATES	00000	EDU010194D	43369917	1994	EDU0101	ERROR
9	UNITED STATES	00000	EDU010195D	43993459	1995	EDU0101	ERROR
10	UNITED STATES	00000	EDU010196D	44715737	1996	EDU0101	ERROR

```
# i 520 more rows
```

Write Function for steps 4,5,6

```
step456func <- function(long_data, values = "Enrollment_Total") {
```

```

row_names <- rownames(long_data)
county_subset <- row_names %in% grep(pattern = "\\.\\.w\\.\\.w", long_data$Area_name)
county_tibble <- subset(long_data, county_subset)
class(county_tibble) <- c("county", class(county_tibble))

state_tibble <- subset(long_data, !(row_names %in% grep(pattern = "\\.\\.w\\.\\.w",
                                                       long_data$Area_name)))
class(state_tibble) <- c("state", class(state_tibble))

#return both tibbles in a list
return(list(step5func(county_tibble), step6func(state_tibble)))
}

```

```
step456func(long_updated)
```

[[1]]

A tibble: 31,450 x 7

	Area_name	STCOU	EnrollmentID	Enrollment_Total	Year	Measurement	State
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>
1	Autauga, AL	01001	EDU010187D	6829	1987	EDU0101	AL
2	Autauga, AL	01001	EDU010188D	6900	1988	EDU0101	AL
3	Autauga, AL	01001	EDU010189D	6920	1989	EDU0101	AL
4	Autauga, AL	01001	EDU010190D	6847	1990	EDU0101	AL
5	Autauga, AL	01001	EDU010191D	7008	1991	EDU0101	AL
6	Autauga, AL	01001	EDU010192D	7137	1992	EDU0101	AL
7	Autauga, AL	01001	EDU010193D	7152	1993	EDU0101	AL
8	Autauga, AL	01001	EDU010194D	7381	1994	EDU0101	AL
9	Autauga, AL	01001	EDU010195D	7568	1995	EDU0101	AL
10	Autauga, AL	01001	EDU010196D	7834	1996	EDU0101	AL

i 31,440 more rows

[[2]]

A tibble: 530 x 7

	Area_name	STCOU	EnrollmentID	Enrollment_Total	Year	Measurement	Division
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>
1	UNITED STATES	00000	EDU010187D	40024299	1987	EDU0101	ERROR
2	UNITED STATES	00000	EDU010188D	39967624	1988	EDU0101	ERROR
3	UNITED STATES	00000	EDU010189D	40317775	1989	EDU0101	ERROR
4	UNITED STATES	00000	EDU010190D	40737600	1990	EDU0101	ERROR
5	UNITED STATES	00000	EDU010191D	41385442	1991	EDU0101	ERROR
6	UNITED STATES	00000	EDU010192D	42088151	1992	EDU0101	ERROR
7	UNITED STATES	00000	EDU010193D	42724710	1993	EDU0101	ERROR

```

8 UNITED STATES 00000 EDU010194D      43369917  1994 EDU0101      ERROR
9 UNITED STATES 00000 EDU010195D      43993459  1995 EDU0101      ERROR
10 UNITED STATES 00000 EDU010196D     44715737  1996 EDU0101      ERROR
# i 520 more rows

```

Wrapper function

```

my_wrapper <- function(url, values = "Enrollment_Total") {
  result <- read_csv(url) |>
    step12func() |>
    step3func() |>
    step456func()
  return(result)
}

```

Call It and Combine Your Data

```

CensusA <- my_wrapper(url = "https://www4.stat.ncsu.edu/~online/datasets/EDU01a.csv",
  values = "Enrollment_Total")

```

```

CensusB <- my_wrapper(url = "https://www4.stat.ncsu.edu/~online/datasets/EDU01b.csv",
  values = "Enrollment_Total")

```

```

#Combining results of the two wrapper functions
combine_function <- function(data1, data2, values = "Enrollment_Total") {
  county = dplyr::bind_rows(data1[[1]], data2[[1]])
  state = dplyr::bind_rows(data1[[2]], data2[[2]])

  return(list(county,state))
}

```

```

combined_data <- combine_function(CensusA,CensusB)
combined_data

```

```

[[1]]
# A tibble: 62,900 x 7
  Area_name   STCOU EnrollmentID Enrollment_Total  Year Measurement State
  <chr>       <chr> <chr>           <dbl> <dbl> <chr>      <chr>

```

```

1 Autauga, AL 01001 EDU010187D      6829 1987 EDU0101      AL
2 Autauga, AL 01001 EDU010188D      6900 1988 EDU0101      AL
3 Autauga, AL 01001 EDU010189D      6920 1989 EDU0101      AL
4 Autauga, AL 01001 EDU010190D      6847 1990 EDU0101      AL
5 Autauga, AL 01001 EDU010191D      7008 1991 EDU0101      AL
6 Autauga, AL 01001 EDU010192D      7137 1992 EDU0101      AL
7 Autauga, AL 01001 EDU010193D      7152 1993 EDU0101      AL
8 Autauga, AL 01001 EDU010194D      7381 1994 EDU0101      AL
9 Autauga, AL 01001 EDU010195D      7568 1995 EDU0101      AL
10 Autauga, AL 01001 EDU010196D      7834 1996 EDU0101      AL
# i 62,890 more rows

```

```

[[2]]
# A tibble: 1,060 x 7
  Area_name      STCOU EnrollmentID Enrollment_Total Year Measurement Division
  <chr>          <chr> <chr>          <dbl> <dbl> <chr>      <chr>
1 UNITED STATES 00000 EDU010187D      40024299 1987 EDU0101      ERROR
2 UNITED STATES 00000 EDU010188D      39967624 1988 EDU0101      ERROR
3 UNITED STATES 00000 EDU010189D      40317775 1989 EDU0101      ERROR
4 UNITED STATES 00000 EDU010190D      40737600 1990 EDU0101      ERROR
5 UNITED STATES 00000 EDU010191D      41385442 1991 EDU0101      ERROR
6 UNITED STATES 00000 EDU010192D      42088151 1992 EDU0101      ERROR
7 UNITED STATES 00000 EDU010193D      42724710 1993 EDU0101      ERROR
8 UNITED STATES 00000 EDU010194D      43369917 1994 EDU0101      ERROR
9 UNITED STATES 00000 EDU010195D      43993459 1995 EDU0101      ERROR
10 UNITED STATES 00000 EDU010196D      44715737 1996 EDU0101      ERROR
# i 1,050 more rows

```

Writing a Generic Function for Summarizing

Write a function for state data that plots the mean value of Enrollment over the years for each Division.

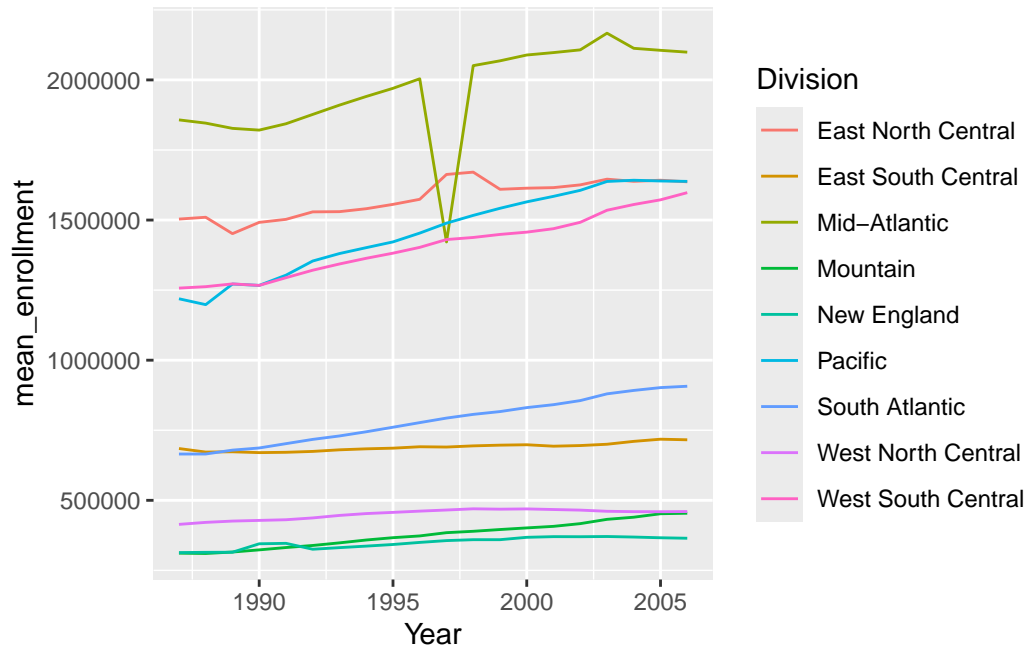
```

plot.state <- function(data, var_name = "Enrollment_Total"){
  data |>
    #group by division and year
    group_by(Division, Year) |>
    #get mean enrollment values for the divisions across years
    summarize(mean_enrollment = mean(get(var_name), na.rm = TRUE), .groups = "drop") |>
    #exclude divisions that are "ERROR"
    filter( Division != "ERROR") |>
    #ggplot where x axis is year and y axis are mean enrollments,

```

```
#distinguish divisions by color
ggplot(aes(x = Year, y = mean_enrollment, color = Division)) +
  #geom_line to show trend
  geom_line() }

#call the plot for the state dataset in the combined data:
plot(combined_data[[2]])
```



Write a function to plot county data, users can: specify state (if not specified default is NJ), specify top or bottom counties (top is default), specify how many top or bottom counties to show (default is 5):

```
#plot function for county data
plot.county <- function(data, state = "NJ", type = "top", n = 5,
                        var_name = "Enrollment_Total") {
  #Filter data for just one state's counties
  data_onestate <- data |>
    filter(State == state) |>
  #Get mean enrollment values grouped by area_names
  group_by(Area_name, Year) |>
    summarize(mean_enrollment = mean(get(var_name), na.rm = TRUE), .groups = "drop")

  #Take filtered data and arrange it ascending and descending, use ifelse to specify n rows:
```

```

select_rows <- if (type == "top") { data_onestate |>
  arrange(desc(mean_enrollment)) |>
  slice_head(n = n) } else { data_onestate |>
  arrange((mean_enrollment)) |>
  slice_head(n = n)}

#filter for where the counties names are the same as the top/bottom selected rows
plot_data <- data_onestate |>
  filter(Area_name %in% select_rows$Area_name)

#now plot the plot_data with geom_line to show trends over years
ggplot(plot_data, aes(x = Year, y = mean_enrollment, color = Area_name)) +
  geom_line()
  }

plot(combined_data[[1]])

```



Put it Together!

Run your data processing function on the two enrollment URLs given previously:

```
CensusA <- my_wrapper(url = "https://www4.stat.ncsu.edu/~online/datasets/EDU01a.csv",
  values = "Enrollment_Total")

CensusB <- my_wrapper(url = "https://www4.stat.ncsu.edu/~online/datasets/EDU01b.csv",
  values = "Enrollment_Total")
```

Run your data combining function to put these into one object:

```
combined_data <- combine_function(CensusA,CensusB)
combined_data
```

```
[[1]]
# A tibble: 62,900 x 7
  Area_name STCOU EnrollmentID Enrollment_Total Year Measurement State
  <chr>      <chr> <chr>          <dbl> <dbl> <chr>      <chr>
1 Autauga, AL 01001 EDU010187D          6829 1987 EDU0101 AL
2 Autauga, AL 01001 EDU010188D          6900 1988 EDU0101 AL
3 Autauga, AL 01001 EDU010189D          6920 1989 EDU0101 AL
4 Autauga, AL 01001 EDU010190D          6847 1990 EDU0101 AL
5 Autauga, AL 01001 EDU010191D          7008 1991 EDU0101 AL
6 Autauga, AL 01001 EDU010192D          7137 1992 EDU0101 AL
7 Autauga, AL 01001 EDU010193D          7152 1993 EDU0101 AL
8 Autauga, AL 01001 EDU010194D          7381 1994 EDU0101 AL
9 Autauga, AL 01001 EDU010195D          7568 1995 EDU0101 AL
10 Autauga, AL 01001 EDU010196D          7834 1996 EDU0101 AL
# i 62,890 more rows
```

```
[[2]]
# A tibble: 1,060 x 7
  Area_name STCOU EnrollmentID Enrollment_Total Year Measurement Division
  <chr>      <chr> <chr>          <dbl> <dbl> <chr>      <chr>
1 UNITED STATES 00000 EDU010187D      40024299 1987 EDU0101 ERROR
2 UNITED STATES 00000 EDU010188D      39967624 1988 EDU0101 ERROR
3 UNITED STATES 00000 EDU010189D      40317775 1989 EDU0101 ERROR
4 UNITED STATES 00000 EDU010190D      40737600 1990 EDU0101 ERROR
5 UNITED STATES 00000 EDU010191D      41385442 1991 EDU0101 ERROR
6 UNITED STATES 00000 EDU010192D      42088151 1992 EDU0101 ERROR
7 UNITED STATES 00000 EDU010193D      42724710 1993 EDU0101 ERROR
8 UNITED STATES 00000 EDU010194D      43369917 1994 EDU0101 ERROR
9 UNITED STATES 00000 EDU010195D      43993459 1995 EDU0101 ERROR
10 UNITED STATES 00000 EDU010196D      44715737 1996 EDU0101 ERROR
# i 1,050 more rows
```

```
#County data:
combined_data[[1]]
```

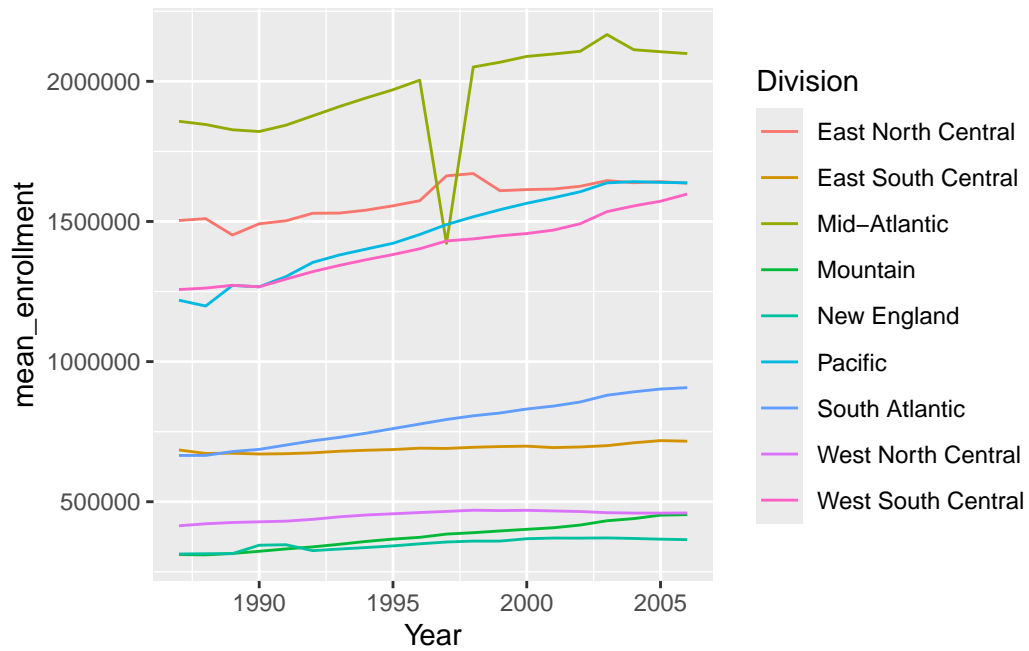
```
# A tibble: 62,900 x 7
  Area_name STCOU EnrollmentID Enrollment_Total Year Measurement State
  <chr>      <chr> <chr>          <dbl> <dbl> <chr>      <chr>
1 Autauga, AL 01001 EDU010187D      6829 1987 EDU0101 AL
2 Autauga, AL 01001 EDU010188D      6900 1988 EDU0101 AL
3 Autauga, AL 01001 EDU010189D      6920 1989 EDU0101 AL
4 Autauga, AL 01001 EDU010190D      6847 1990 EDU0101 AL
5 Autauga, AL 01001 EDU010191D      7008 1991 EDU0101 AL
6 Autauga, AL 01001 EDU010192D      7137 1992 EDU0101 AL
7 Autauga, AL 01001 EDU010193D      7152 1993 EDU0101 AL
8 Autauga, AL 01001 EDU010194D      7381 1994 EDU0101 AL
9 Autauga, AL 01001 EDU010195D      7568 1995 EDU0101 AL
10 Autauga, AL 01001 EDU010196D      7834 1996 EDU0101 AL
# i 62,890 more rows
```

```
#State data
combined_data[[2]]
```

```
# A tibble: 1,060 x 7
  Area_name STCOU EnrollmentID Enrollment_Total Year Measurement Division
  <chr>      <chr> <chr>          <dbl> <dbl> <chr>      <chr>
1 UNITED STATES 00000 EDU010187D      40024299 1987 EDU0101 ERROR
2 UNITED STATES 00000 EDU010188D      39967624 1988 EDU0101 ERROR
3 UNITED STATES 00000 EDU010189D      40317775 1989 EDU0101 ERROR
4 UNITED STATES 00000 EDU010190D      40737600 1990 EDU0101 ERROR
5 UNITED STATES 00000 EDU010191D      41385442 1991 EDU0101 ERROR
6 UNITED STATES 00000 EDU010192D      42088151 1992 EDU0101 ERROR
7 UNITED STATES 00000 EDU010193D      42724710 1993 EDU0101 ERROR
8 UNITED STATES 00000 EDU010194D      43369917 1994 EDU0101 ERROR
9 UNITED STATES 00000 EDU010195D      43993459 1995 EDU0101 ERROR
10 UNITED STATES 00000 EDU010196D      44715737 1996 EDU0101 ERROR
# i 1,050 more rows
```

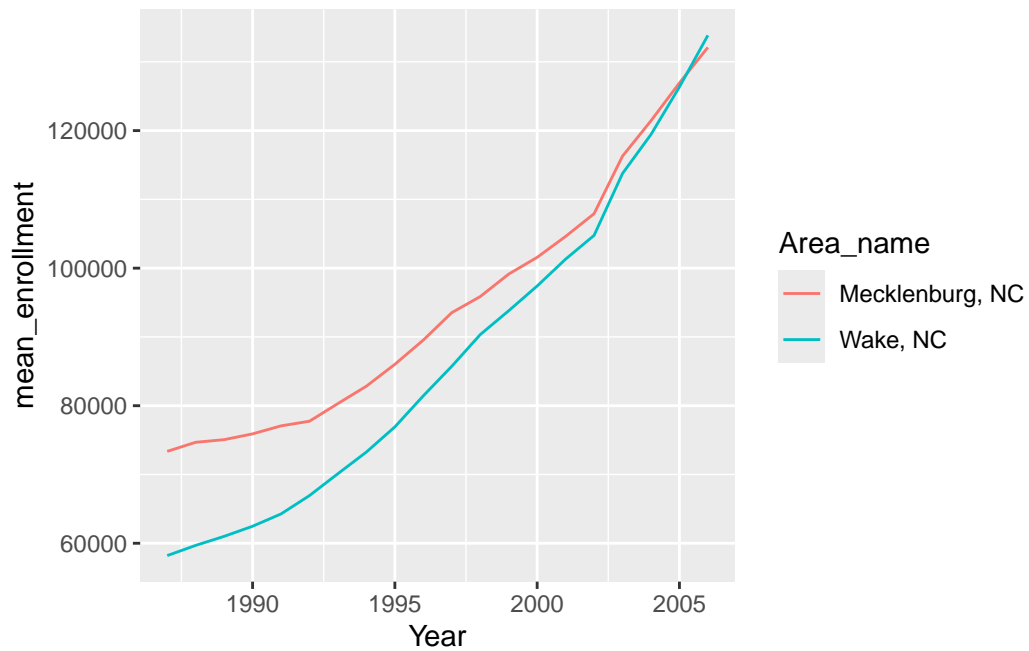
Plot State data frame:

```
plot(combined_data[[2]])
```

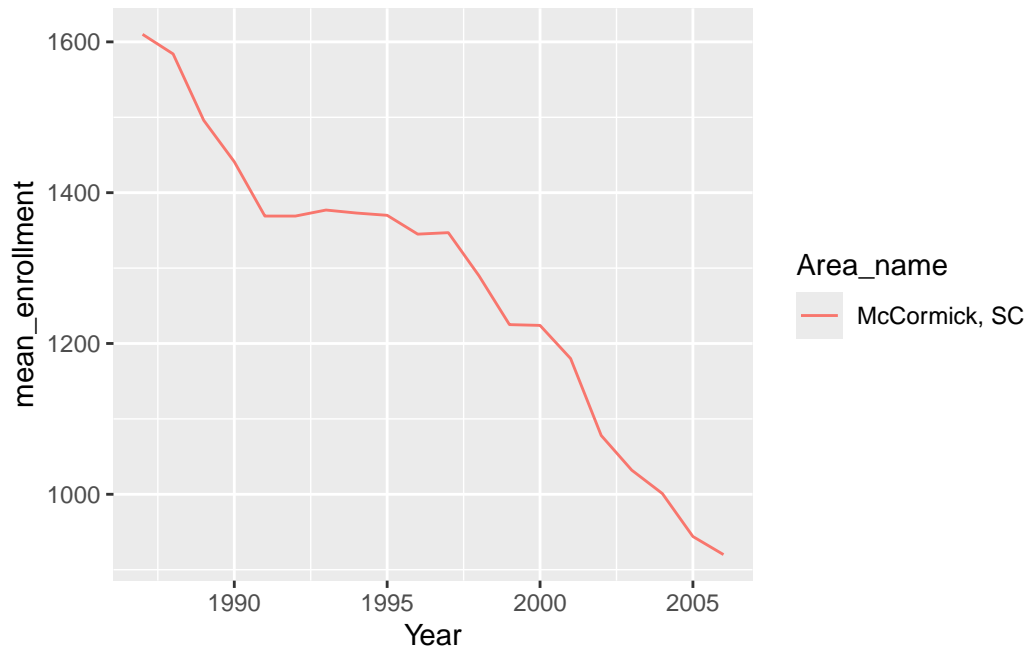
Plot County data: *specifying the state to be "NC", the group being the top, the number looked at being 20*

```
plot(combined_data[[1]], state = "NC", type = "top", n = 20)
```



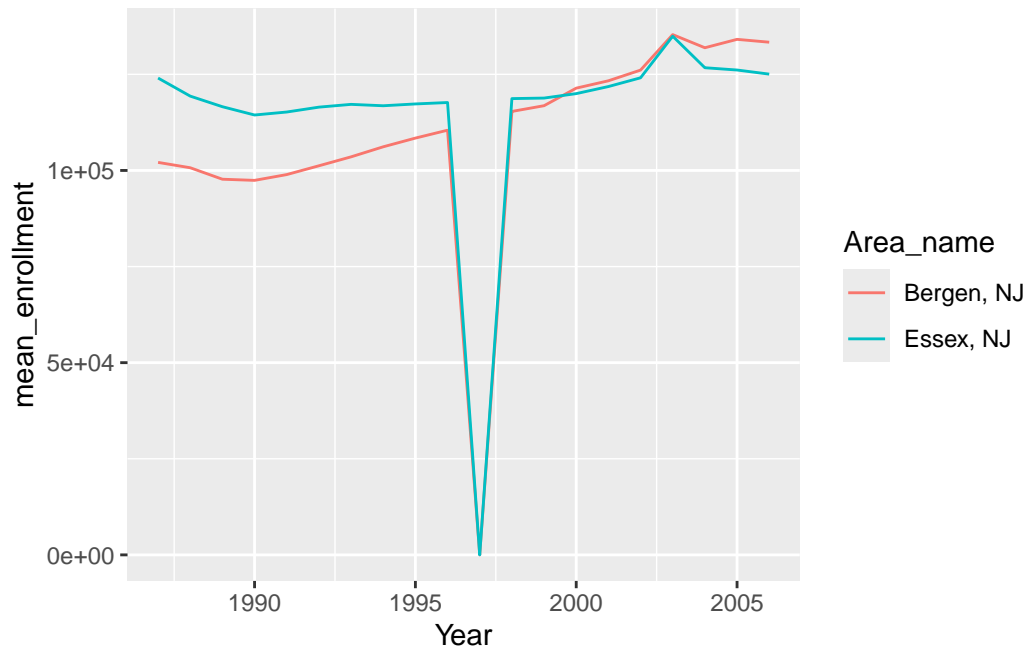
specifying the state to be “SC”, the group being the bottom, the number looked at being 7

```
plot(combined_data[[1]], state = "SC", type = "bottom", n = 7)
```



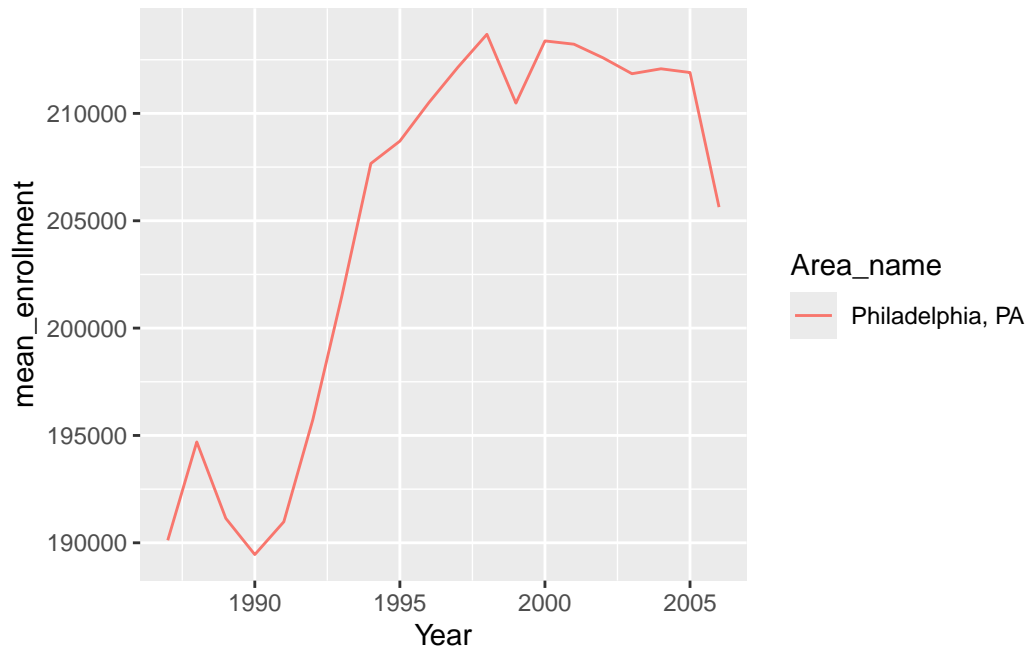
without specifying anything (defaults used)

```
plot(combined_data[[1]])
```



specifying the state to be “PA”, the group being the top, the number looked at being 8

```
plot(combined_data[[1]], state = "PA", type = "top", n = 8)
```



Run your data processing function on the four data sets:

```
URL1 <- my_wrapper(url = "https://www4.stat.ncsu.edu/~online/datasets/PST01a.csv",
  values = "Enrollment_Total")

URL2 <- my_wrapper(url = "https://www4.stat.ncsu.edu/~online/datasets/PST01b.csv",
  values = "Enrollment_Total")

URL3 <- my_wrapper(url = "https://www4.stat.ncsu.edu/~online/datasets/PST01c.csv",
  values = "Enrollment_Total")

URL4 <- my_wrapper(url = "https://www4.stat.ncsu.edu/~online/datasets/PST01d.csv",
  values = "Enrollment_Total")
```

Data combining function to put these into one object:

```
combined_data12 <- combine_function(URL1,URL2)
combined_data34 <- combine_function(URL3,URL4)
combined_data1234 <- combine_function(combined_data12,combined_data34)
combined_data1234
```

```
[[1]]
# A tibble: 125,800 x 7
  Area_name STCOU EnrollmentID Enrollment_Total Year Measurement State
  <chr>      <chr> <chr>                <dbl> <dbl> <chr>      <chr>
1 Autauga, AL 01001 PST015171D          25508  1971 PST0151    AL
2 Autauga, AL 01001 PST015172D          27166  1972 PST0151    AL
3 Autauga, AL 01001 PST015173D          28463  1973 PST0151    AL
4 Autauga, AL 01001 PST015174D          29266  1974 PST0151    AL
5 Autauga, AL 01001 PST015175D          29718  1975 PST0151    AL
6 Autauga, AL 01001 PST015176D          29896  1976 PST0151    AL
7 Autauga, AL 01001 PST015177D          30462  1977 PST0151    AL
8 Autauga, AL 01001 PST015178D          30882  1978 PST0151    AL
9 Autauga, AL 01001 PST015179D          32055  1979 PST0151    AL
10 Autauga, AL 01001 PST025181D          31985  1981 PST0251    AL
# i 125,790 more rows
```

```
[[2]]
# A tibble: 2,120 x 7
  Area_name STCOU EnrollmentID Enrollment_Total Year Measurement Division
  <chr>      <chr> <chr>                <dbl> <dbl> <chr>      <chr>
1 UNITED STATES 00000 PST015171D          206827028  1971 PST0151    ERROR
```

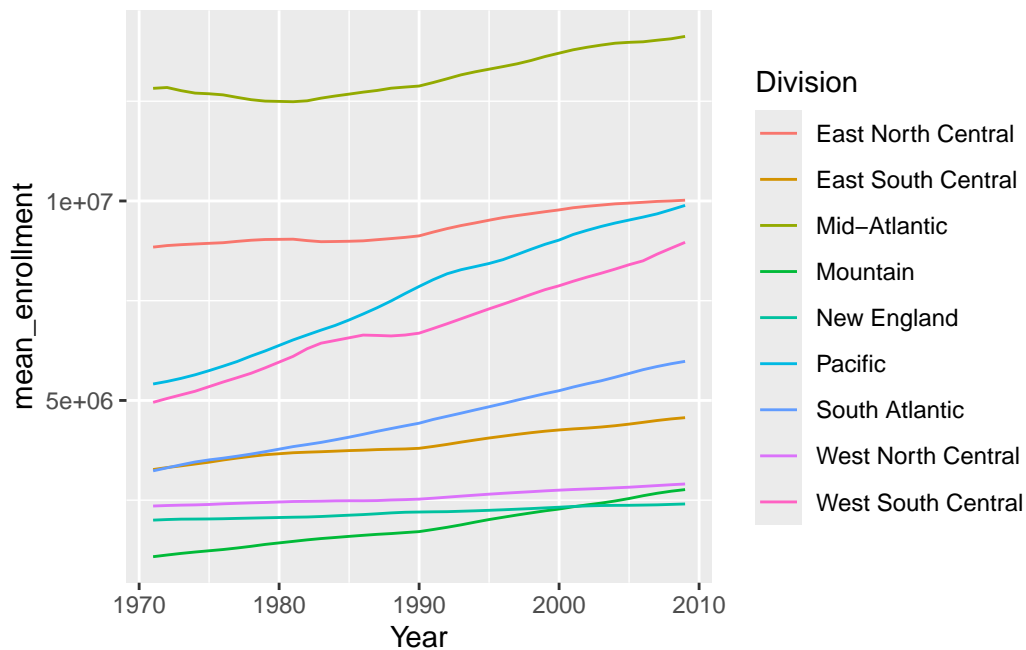
```

2 UNITED STATES 00000 PST015172D      209283904 1972 PST0151  ERROR
3 UNITED STATES 00000 PST015173D      211357490 1973 PST0151  ERROR
4 UNITED STATES 00000 PST015174D      213341552 1974 PST0151  ERROR
5 UNITED STATES 00000 PST015175D      215465246 1975 PST0151  ERROR
6 UNITED STATES 00000 PST015176D      217562728 1976 PST0151  ERROR
7 UNITED STATES 00000 PST015177D      219759860 1977 PST0151  ERROR
8 UNITED STATES 00000 PST015178D      222095080 1978 PST0151  ERROR
9 UNITED STATES 00000 PST015179D      224567234 1979 PST0151  ERROR
10 UNITED STATES 00000 PST025181D      229466391 1981 PST0251  ERROR
# i 2,110 more rows

```

Use the plot function on the state data frame:

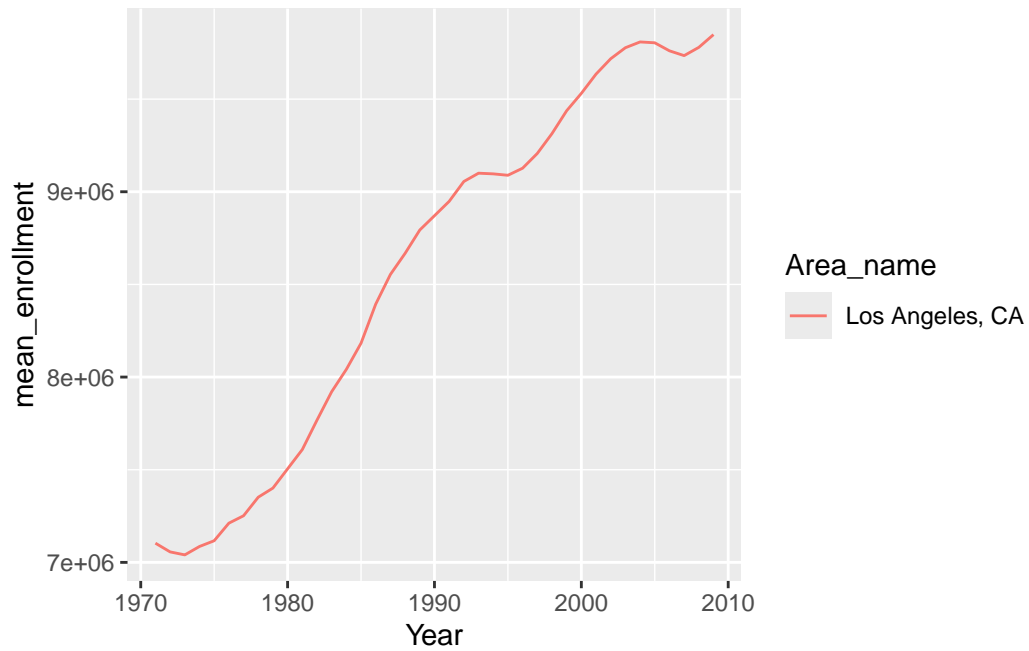
```
plot(combined_data1234[[2]])
```



Use the plot function on the county data frame:

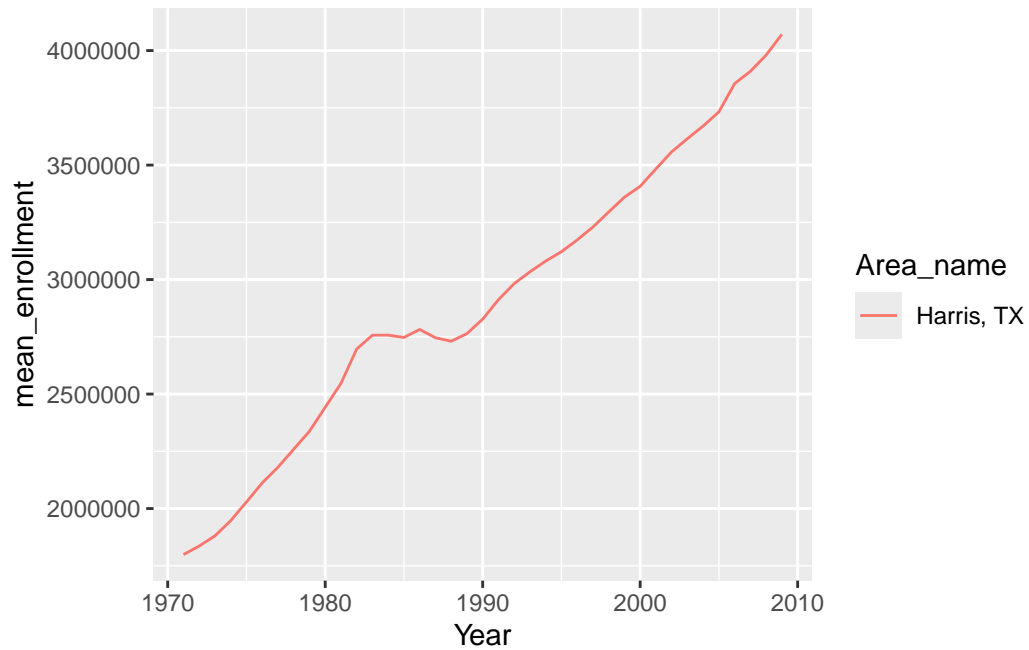
Specifying the state to be "CA", the group being the top, the number looked at being 15:

```
plot(combined_data1234[[1]], state = "CA", type = "top", n = 15)
```



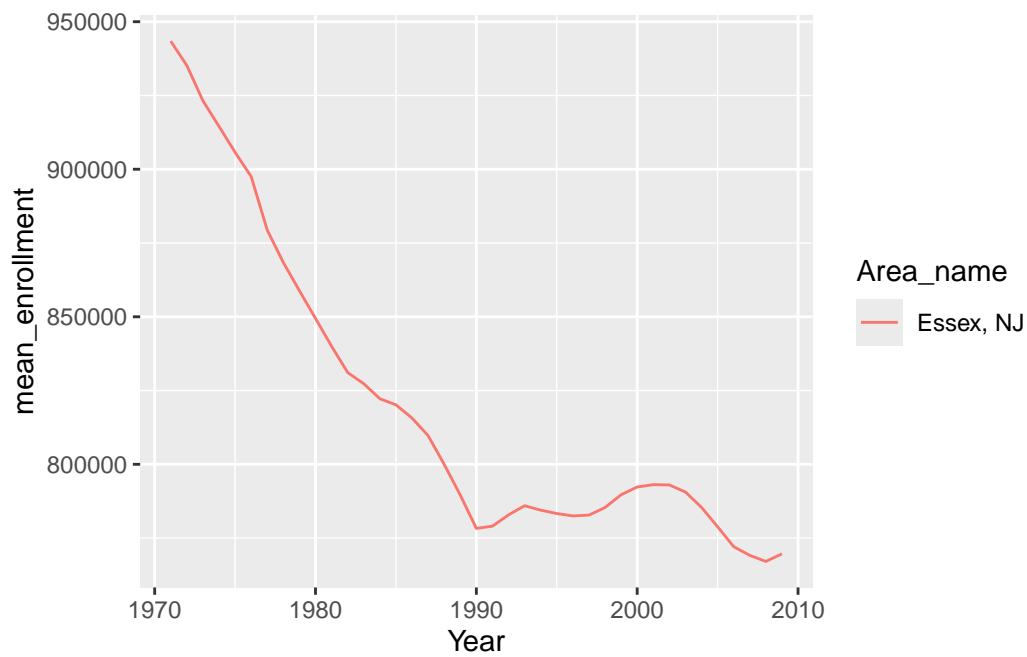
specifying the state to be “TX”, the group being the top, the number looked at being 4

```
plot(combined_data1234[[1]], state = "TX", type = "top", n = 4)
```



without specifying anything (defaults used)

```
plot(combined_data1234[[1]])
```



specifying the state to be “NY”, the group being the top, the number looked at being 10

```
plot(combined_data1234[[1]], state = "NY", type = "top", n = 10)
```

