Project 1

Hongjing Mao Yuhan Hu

2025-06-14

Step 1: Data preprocessing

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ---- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.2     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.4
-- Conflicts ---------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
url <- "https://www4.stat.ncsu.edu/~online/datasets/EDU01a.csv"

edu01a <- read_csv(url, show_col_types = FALSE) %>%
  select(
    area_name = Area_name,
    STCOU,
    ends_with("D")
  )

head(edu01a, 5)
```

```
# A tibble: 5 x 12
  area_name     STCOU EDU010187D EDU010188D EDU010189D
```

```
  <chr>          <chr>      <dbl>      <dbl>      <dbl>
1 UNITED STATES 00000   40024299   39967624   40317775
2 ALABAMA       01000     733735     728234     730048
3 Autauga, AL   01001       6829       6900       6920
4 Baldwin, AL   01003      16417      16465      16799
5 Barbour, AL   01005       5071       5098       5068
# i 7 more variables: EDU010190D <dbl>, EDU010191D <dbl>,
#   EDU010192D <dbl>, EDU010193D <dbl>, EDU010194D <dbl>,
#   EDU010195D <dbl>, EDU010196D <dbl>
```

Step 2: wide string to long string

```
library(tidyr)
edu01a_long <- edu01a %>%
  pivot_longer(
    cols  = -c(area_name, STCOU),
    names_to  = "raw_var",
    values_to = "enrollment",
    values_drop_na = TRUE
  )
head(edu01a_long, 5)
```

```
# A tibble: 5 x 4
  area_name     STCOU raw_var    enrollment
  <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
```

Step 3: Parse variable name string

```
long_updated <- edu01a_long %>%
  mutate(
    # last two digits before the "D"
    yr_two      = substr(raw_var, nchar(raw_var) - 2, nchar(raw_var) - 1),  # e.g. "97" or "
    year        = as.integer(
```

```
                    if_else(as.integer(yr_two) <= 30,
                            paste0("20", yr_two),
                            paste0("19", yr_two))
                  ),
    measurement = substr(raw_var, 1, 7)
  ) %>%
  select(-yr_two)

head(long_updated, 5)
```

```
# A tibble: 5 x 6
  area_name     STCOU raw_var   enrollment  year measurement
  <chr>         <chr> <chr>          <dbl> <int> <chr>
1 UNITED STATES 00000 EDU01018~   40024299  1987 EDU0101
2 UNITED STATES 00000 EDU01018~   39967624  1988 EDU0101
3 UNITED STATES 00000 EDU01018~   40317775  1989 EDU0101
4 UNITED STATES 00000 EDU01019~   40737600  1990 EDU0101
5 UNITED STATES 00000 EDU01019~   41385442  1991 EDU0101
```

Step 4: Separate county vs non-county data & add classes

```
library(dplyr)
library(stringr)

county_idx <- grep(", \\w\\w$", long_updated$area_name)

county_tibble <- long_updated[county_idx, ] %>%
  mutate(area_name = str_trim(area_name))
class(county_tibble) <- c("county", class(county_tibble))

state_tibble  <- long_updated[-county_idx, ]
class(state_tibble)  <- c("state",  class(state_tibble))

# quick check: first 10 rows of each
head(county_tibble, 10)
```

```
# A tibble: 10 x 6
   area_name   STCOU raw_var    enrollment  year measurement
   <chr>       <chr> <chr>           <dbl> <int> <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
```

```
head(state_tibble,  10)
```

```
# A tibble: 10 x 6
   area_name     STCOU raw_var  enrollment  year measurement
   <chr>         <chr> <chr>         <dbl> <int> <chr>
 1 UNITED STATES 00000 EDU0101~   40024299  1987 EDU0101
 2 UNITED STATES 00000 EDU0101~   39967624  1988 EDU0101
 3 UNITED STATES 00000 EDU0101~   40317775  1989 EDU0101
 4 UNITED STATES 00000 EDU0101~   40737600  1990 EDU0101
 5 UNITED STATES 00000 EDU0101~   41385442  1991 EDU0101
 6 UNITED STATES 00000 EDU0101~   42088151  1992 EDU0101
 7 UNITED STATES 00000 EDU0101~   42724710  1993 EDU0101
 8 UNITED STATES 00000 EDU0101~   43369917  1994 EDU0101
 9 UNITED STATES 00000 EDU0101~   43993459  1995 EDU0101
10 UNITED STATES 00000 EDU0101~   44715737  1996 EDU0101
```

Step 5: Create new variable

```
library(stringr)
new_county_tibble <- county_tibble %>%
  mutate(
    state = str_sub(area_name, -2, -1)
  )
head(new_county_tibble, 10)
```

```
# A tibble: 10 x 7
   area_name   STCOU raw_var    enrollment  year measurement
   <chr>       <chr> <chr>           <dbl> <int> <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
# i 1 more variable: state <chr>
```

Step 6: Create new "division" variables

```
state_division <- c(
  "CONNECTICUT" = "New England", "MAINE" = "New England", "MASSACHUSETTS" = "New England",
  "NEW HAMPSHIRE" = "New England", "RHODE ISLAND" = "New England", "VERMONT" = "New England"

  "NEW JERSEY" = "Mid-Atlantic", "NEW YORK" = "Mid-Atlantic", "PENNSYLVANIA" = "Mid-Atlantic"

  "ILLINOIS" = "East North Central", "INDIANA" = "East North Central", "MICHIGAN" = "East No
  "OHIO" = "East North Central", "WISCONSIN" = "East North Central",

  "IOWA" = "West North Central", "KANSAS" = "West North Central", "MINNESOTA" = "West North
  "MISSOURI" = "West North Central", "NEBRASKA" = "West North Central",
  "NORTH DAKOTA" = "West North Central", "SOUTH DAKOTA" = "West North Central",

  "DELAWARE" = "South Atlantic", "DISTRICT OF COLUMBIA" = "South Atlantic", "District of Colu
  "GEORGIA" = "South Atlantic", "MARYLAND" = "South Atlantic", "NORTH CAROLINA" = "South Atla
  "SOUTH CAROLINA" = "South Atlantic", "VIRGINIA" = "South Atlantic", "WEST VIRGINIA" = "Sou

  "ALABAMA" = "East South Central", "KENTUCKY" = "East South Central",
  "MISSISSIPPI" = "East South Central", "TENNESSEE" = "East South Central",

  "ARKANSAS" = "West South Central", "LOUISIANA" = "West South Central",
  "OKLAHOMA" = "West South Central", "TEXAS" = "West South Central",

  "ARIZONA" = "Mountain", "COLORADO" = "Mountain", "IDAHO" = "Mountain",
  "MONTANA" = "Mountain", "NEVADA" = "Mountain", "NEW MEXICO" = "Mountain",
  "UTAH" = "Mountain", "WYOMING" = "Mountain",

  "ALASKA" = "Pacific", "CALIFORNIA" = "Pacific", "HAWAII" = "Pacific",
```

```
  "OREGON" = "Pacific", "WASHINGTON" = "Pacific"
)

division <- state_tibble %>%
  mutate(
    division = if_else(
      area_name %in% names(state_division),
      state_division[area_name],
      "ERROR"
    )
  )
head(division)
```

```
# A tibble: 6 x 7
  area_name     STCOU raw_var   enrollment  year measurement
  <chr>         <chr> <chr>          <dbl> <int> <chr>
1 UNITED STATES 00000 EDU01018~   40024299  1987 EDU0101
2 UNITED STATES 00000 EDU01018~   39967624  1988 EDU0101
3 UNITED STATES 00000 EDU01018~   40317775  1989 EDU0101
4 UNITED STATES 00000 EDU01019~   40737600  1990 EDU0101
5 UNITED STATES 00000 EDU01019~   41385442  1991 EDU0101
6 UNITED STATES 00000 EDU01019~   42088151  1992 EDU0101
# i 1 more variable: division <chr>
```

Functions and Wrapper:

```
library(tidyverse)
library(rvest)
```

```
Attaching package: 'rvest'

The following object is masked from 'package:readr':

    guess_encoding
```

```
library(purrr)
library(tidyr)
library(stringr)
```

Global helper: mapping states to census divisions

```
state_division <- c(
  "CONNECTICUT" = "New England", "MAINE" = "New England", "MASSACHUSETTS" = "New England",
  "NEW HAMPSHIRE" = "New England", "RHODE ISLAND" = "New England", "VERMONT" = "New England"

  "NEW JERSEY" = "Mid-Atlantic", "NEW YORK" = "Mid-Atlantic", "PENNSYLVANIA" = "Mid-Atlantic"

  "ILLINOIS" = "East North Central", "INDIANA" = "East North Central", "MICHIGAN" = "East No
  "OHIO" = "East North Central", "WISCONSIN" = "East North Central",

  "IOWA" = "West North Central", "KANSAS" = "West North Central", "MINNESOTA" = "West North
  "MISSOURI" = "West North Central", "NEBRASKA" = "West North Central",
  "NORTH DAKOTA" = "West North Central", "SOUTH DAKOTA" = "West North Central",

  "DELAWARE" = "South Atlantic", "DISTRICT OF COLUMBIA" = "South Atlantic", "District of Colu
  "GEORGIA" = "South Atlantic", "MARYLAND" = "South Atlantic", "NORTH CAROLINA" = "South Atla
  "SOUTH CAROLINA" = "South Atlantic", "VIRGINIA" = "South Atlantic", "WEST VIRGINIA" = "Sou

  "ALABAMA" = "East South Central", "KENTUCKY" = "East South Central",
  "MISSISSIPPI" = "East South Central", "TENNESSEE" = "East South Central",

  "ARKANSAS" = "West South Central", "LOUISIANA" = "West South Central",
  "OKLAHOMA" = "West South Central", "TEXAS" = "West South Central",

  "ARIZONA" = "Mountain", "COLORADO" = "Mountain", "IDAHO" = "Mountain",
  "MONTANA" = "Mountain", "NEVADA" = "Mountain", "NEW MEXICO" = "Mountain",
  "UTAH" = "Mountain", "WYOMING" = "Mountain",

  "ALASKA" = "Pacific", "CALIFORNIA" = "Pacific", "HAWAII" = "Pacific",
  "OREGON" = "Pacific", "WASHINGTON" = "Pacific"
)
```

Step 1&2: read & pivot to long form:

```
read_and_pivot <- function(url, value_var = "enrollment") {
  read_csv(url, show_col_types = FALSE) %>%
    select(
      area_name = Area_name,
      STCOU,
```

```
      ends_with("D")
    ) %>%
    pivot_longer(
      cols           = -c(area_name, STCOU),
      names_to       = "raw_var",
      values_to      = value_var,
      values_drop_na = TRUE
    )
}

# Test  on edu01a.csv
edu01a_long <- read_and_pivot("https://www4.stat.ncsu.edu/~online/datasets/EDU01a.csv")
head(edu01a_long, 5)
```

```
# A tibble: 5 x 4
  area_name     STCOU raw_var    enrollment
  <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
```

Step 3: parse raw_var into year & measurement

```
parse_raw <- function(df) {
  df %>%
    mutate(
      yr_code     = str_sub(raw_var, -3, -2),
      year        = as.integer(if_else(
                      as.integer(yr_code) <= 30,
                      paste0("20", yr_code),
                      paste0("19", yr_code)
                    )),
      measurement = str_sub(raw_var, 1, 7)
    ) %>%
    select(-yr_code)
}

# Test:
```

```r
edu01a_parsed <- parse_raw(edu01a_long)
head(edu01a_parsed, 5)
```

```
# A tibble: 5 x 6
  area_name     STCOU raw_var   enrollment  year measurement
  <chr>         <chr> <chr>          <dbl> <int> <chr>
1 UNITED STATES 00000 EDU01018~   40024299  1987 EDU0101
2 UNITED STATES 00000 EDU01018~   39967624  1988 EDU0101
3 UNITED STATES 00000 EDU01018~   40317775  1989 EDU0101
4 UNITED STATES 00000 EDU01019~   40737600  1990 EDU0101
5 UNITED STATES 00000 EDU01019~   41385442  1991 EDU0101
```

Step 5:for county data, extract state code and set class

```r
add_state_county <- function(county_df) {
  county_df %>%
    mutate(state = str_sub(area_name, -2, -1))
}

# Test it on the county subset:
county_subset <- edu01a_parsed %>% filter(str_detect(area_name, ", \\w\\w$"))
head(add_state_county(county_subset), 5)
```

```
# A tibble: 5 x 7
  area_name STCOU raw_var enrollment  year measurement state
  <chr>     <chr> <chr>        <dbl> <int> <chr>       <chr>
1 Autauga,~ 01001 EDU010~       6829  1987 EDU0101     AL
2 Autauga,~ 01001 EDU010~       6900  1988 EDU0101     AL
3 Autauga,~ 01001 EDU010~       6920  1989 EDU0101     AL
4 Autauga,~ 01001 EDU010~       6847  1990 EDU0101     AL
5 Autauga,~ 01001 EDU010~       7008  1991 EDU0101     AL
```

Step 6: for state data, map to division and set class

```r
add_division <- function(state_df) {
  state_df %>%
    mutate(division = if_else(
```

```
      area_name %in% names(state_division),
      state_division[area_name],
      "ERROR"
    )) %>%
    structure(class = c("state", class(state_df)))
}
# Test it on the non-county subset:
state_subset <- edu01a_parsed %>% filter(!str_detect(area_name, ", \\w\\w$"))
head(add_division(state_subset), 5)
```

```
# A tibble: 5 x 7
  area_name     STCOU raw_var   enrollment  year measurement
  <chr>         <chr> <chr>          <dbl> <int> <chr>
1 UNITED STATES 00000 EDU01018~   40024299  1987 EDU0101
2 UNITED STATES 00000 EDU01018~   39967624  1988 EDU0101
3 UNITED STATES 00000 EDU01018~   40317775  1989 EDU0101
4 UNITED STATES 00000 EDU01019~   40737600  1990 EDU0101
5 UNITED STATES 00000 EDU01019~   41385442  1991 EDU0101
# i 1 more variable: division <chr>
```

step 7: split and classify the data

```
split_and_classify <- function(df) {
  county_idx  <- grep(", \\w\\w$", df$area_name)
  county_part <- df[county_idx, ]
  state_part  <- df[-county_idx, ]

  list(
    county = add_state_county(county_part),
    state  = add_division(state_part)
  )
}

# example
parts01b <- split_and_classify(edu01a_parsed)
head(parts01b$county, 5)
```

```
# A tibble: 5 x 7
  area_name STCOU raw_var enrollment  year measurement state
```

```
  <chr>     <chr> <chr>        <dbl> <int> <chr>       <chr>
1 Autauga,~ 01001 EDU010~       6829  1987 EDU0101     AL
2 Autauga,~ 01001 EDU010~       6900  1988 EDU0101     AL
3 Autauga,~ 01001 EDU010~       6920  1989 EDU0101     AL
4 Autauga,~ 01001 EDU010~       6847  1990 EDU0101     AL
5 Autauga,~ 01001 EDU010~       7008  1991 EDU0101     AL
```

```r
head(parts01b$state,  5)
```

```
# A tibble: 5 x 7
  area_name     STCOU raw_var   enrollment  year measurement
  <chr>         <chr> <chr>          <dbl> <int> <chr>
1 UNITED STATES 00000 EDU01018~   40024299  1987 EDU0101
2 UNITED STATES 00000 EDU01018~   39967624  1988 EDU0101
3 UNITED STATES 00000 EDU01018~   40317775  1989 EDU0101
4 UNITED STATES 00000 EDU01019~   40737600  1990 EDU0101
5 UNITED STATES 00000 EDU01019~   41385442  1991 EDU0101
# i 1 more variable: division <chr>
```

Wrapper & combine functions

```r
my_wrapper <- function(url, value_var = "enrollment") {
    read_and_pivot(url, value_var) %>%
    parse_raw() %>%
    add_state_county() %>%
    add_division() %>%
    split_and_classify()
}

url_a <- "https://www4.stat.ncsu.edu/~online/datasets/EDU01a.csv"
url_b <- "https://www4.stat.ncsu.edu/~online/datasets/EDU01b.csv"

edu_a <- my_wrapper(url_a, value_var = "enrollment")
edu_b <- my_wrapper(url_b, value_var = "enrollment")

combine_two <- function(w1, w2) {
  list(
    county = bind_rows(w1$county, w2$county),
    state  = bind_rows(w1$state,  w2$state)
  )
```

```
}

edu_combined <- combine_two(edu_a, edu_b)

# Check results
head(edu_combined$county,  5)
```

```
# A tibble: 5 x 8
  area_name STCOU raw_var enrollment  year measurement state
  <chr>     <chr> <chr>        <dbl> <int> <chr>       <chr>
1 Autauga,~ 01001 EDU010~       6829  1987 EDU0101     AL
2 Autauga,~ 01001 EDU010~       6900  1988 EDU0101     AL
3 Autauga,~ 01001 EDU010~       6920  1989 EDU0101     AL
4 Autauga,~ 01001 EDU010~       6847  1990 EDU0101     AL
5 Autauga,~ 01001 EDU010~       7008  1991 EDU0101     AL
# i 1 more variable: division <chr>
```

```
head(edu_combined$state,   5)
```

```
# A tibble: 5 x 8
  area_name STCOU raw_var enrollment  year measurement state
  <chr>     <chr> <chr>        <dbl> <int> <chr>       <chr>
1 UNITED S~ 00000 EDU010~   40024299  1987 EDU0101     ES
2 UNITED S~ 00000 EDU010~   39967624  1988 EDU0101     ES
3 UNITED S~ 00000 EDU010~   40317775  1989 EDU0101     ES
4 UNITED S~ 00000 EDU010~   40737600  1990 EDU0101     ES
5 UNITED S~ 00000 EDU010~   41385442  1991 EDU0101     ES
# i 1 more variable: division <chr>
```

Plot function:

1) State-level plot method

```
plot.state <- function(df, var_name = "enrollment", ...) {
  df %>%
    filter(division != "ERROR") %>%
    group_by(division, year) %>%
    summarize(
```

```
      mean_val = mean(.data[[var_name]], na.rm = TRUE),
      .groups   = "drop"
    ) %>%
    ggplot(aes(x = year, y = mean_val, color = division)) +
    geom_line() +
    labs(
      x     = "Year",
      y     = paste("Mean", var_name),
      title = paste("Mean", var_name, "by Division over Time")
    )
}
```

2) County-level plot method

```
plot.county <- function(df,
                        var_name   = "enrollment",
                        state_name = NULL,
                        top_bottom = "top",
                        n          = 5,
                        ...) {
  # optionally filter to a single state
  if (!is.null(state_name)) {
    df <- df %>% filter(state == state_name)
  }

  # compute overall means by county
  means <- df %>%
    group_by(area_name) %>%
    summarize(
      mean_val = mean(.data[[var_name]], na.rm = TRUE),
      .groups   = "drop"
    ) %>%
    arrange(if (top_bottom == "top") desc(mean_val) else mean_val)

  # select the top or bottom N counties
  selected <- head(means$area_name, n)

  # filter original data to those counties
  plot_df <- df %>% filter(area_name %in% selected)
```

```
  # draw the time series of the raw values
  ggplot(plot_df, aes(x = year, y = .data[[var_name]], color = area_name)) +
    geom_line() +
    labs(
      x     = "Year",
      y     = var_name,
      title = paste(
        str_to_title(top_bottom), n,
        "counties",
        if (!is.null(state_name)) paste("in", state_name) else ""
      )
    )
}
```

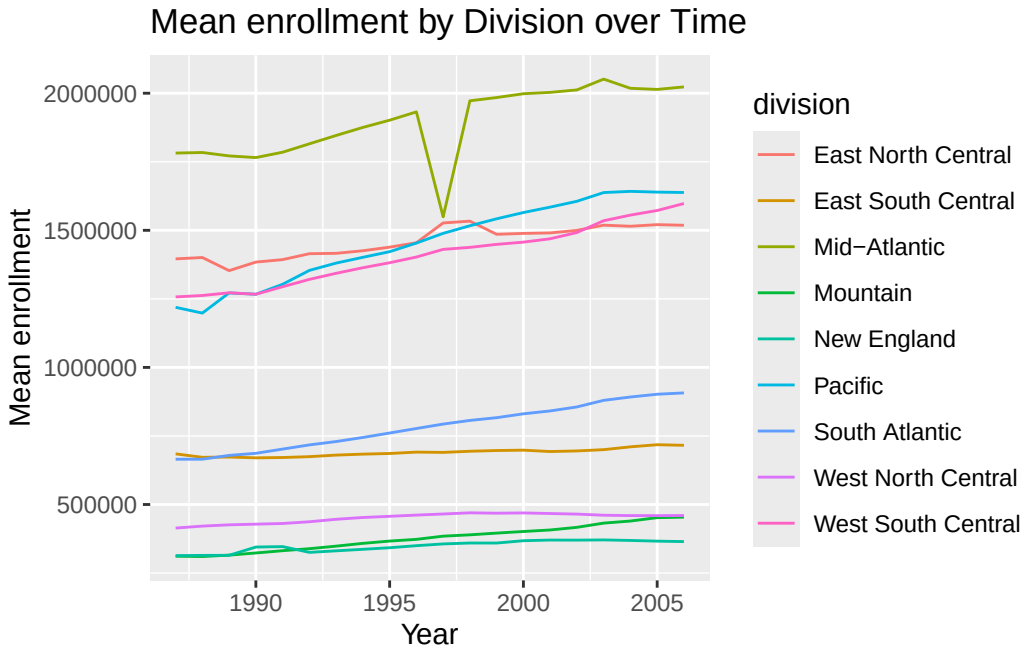
Put it together

```
## 1) Load url
edu_a <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/EDU01a.csv",
                    value_var = "enrollment")
edu_b <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/EDU01b.csv",
                    value_var = "enrollment")

## 2) Combine
edu_combined <- combine_two(edu_a, edu_b)

# Inspect structure: [[1]] = county, [[2]] = state
str(edu_combined)
```

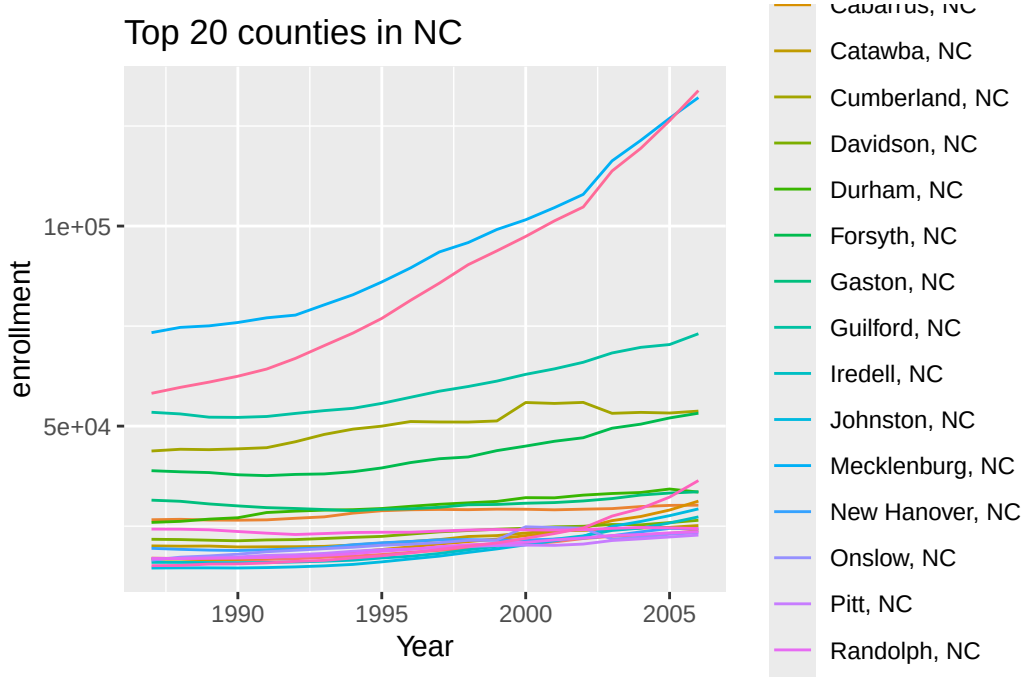
```
List of 2
 $ county: state [62,900 x 8] (S3: state/tbl_df/tbl/data.frame)
  ..$ area_name  : chr [1:62900] "Autauga, AL" "Autauga, AL" "Autauga, AL" "Autauga, AL" ...
  ..$ STCOU      : chr [1:62900] "01001" "01001" "01001" "01001" ...
  ..$ raw_var    : chr [1:62900] "EDU010187D" "EDU010188D" "EDU010189D" "EDU010190D" ...
  ..$ enrollment : num [1:62900] 6829 6900 6920 6847 7008 ...
  ..$ year       : int [1:62900] 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 ...
  ..$ measurement: chr [1:62900] "EDU0101" "EDU0101" "EDU0101" "EDU0101" ...
  ..$ state      : chr [1:62900] "AL" "AL" "AL" "AL" ...
  ..$ division   : Named chr [1:62900] "ERROR" "ERROR" "ERROR" "ERROR" ...
  .. ..- attr(*, "names")= chr [1:62900] "" "" "" "" ...
 $ state : state [1,060 x 8] (S3: state/state/tbl_df/tbl/data.frame)
```

```
  ..$ area_name  : chr [1:1060] "UNITED STATES" "UNITED STATES" "UNITED STATES" "UNITED STATE
  ..$ STCOU      : chr [1:1060] "00000" "00000" "00000" "00000" ...
  ..$ raw_var    : chr [1:1060] "EDU010187D" "EDU010188D" "EDU010189D" "EDU010190D" ...
  ..$ enrollment : num [1:1060] 40024299 39967624 40317775 40737600 41385442 ...
  ..$ year       : int [1:1060] 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 ...
  ..$ measurement: chr [1:1060] "EDU0101" "EDU0101" "EDU0101" "EDU0101" ...
  ..$ state      : chr [1:1060] "ES" "ES" "ES" "ES" ...
  ..$ division   : Named chr [1:1060] "ERROR" "ERROR" "ERROR" "ERROR" ...
  .. ..- attr(*, "names")= chr [1:1060] "" "" "" "" ...
```

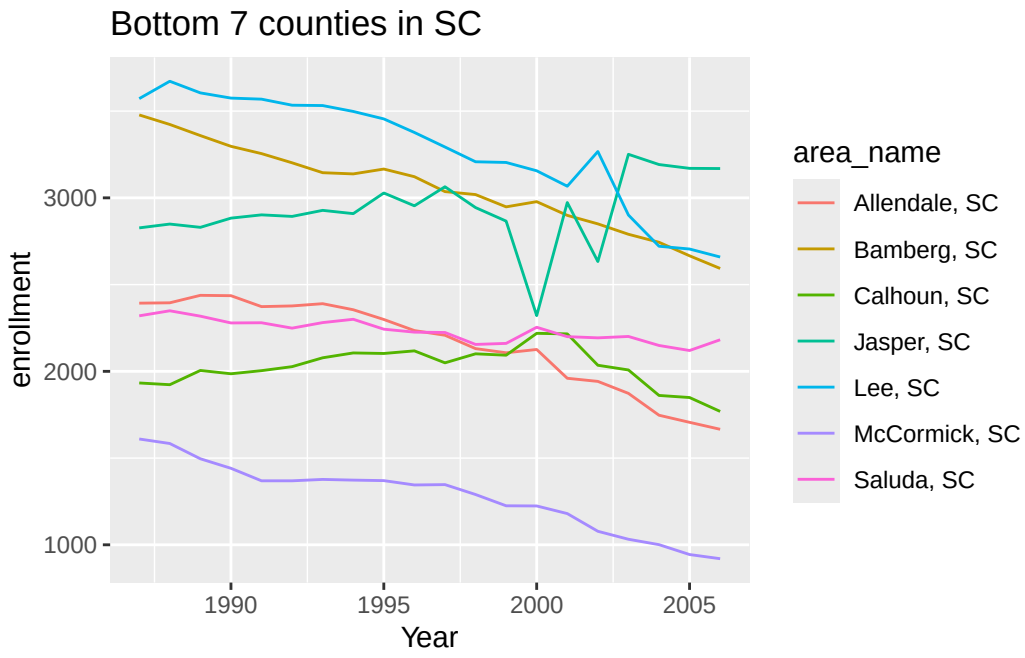
```r
## 3) Extract the state-level tibble via indexing and plot it
state_df  <- edu_combined[[2]]
plot(state_df, var_name = "enrollment")
```



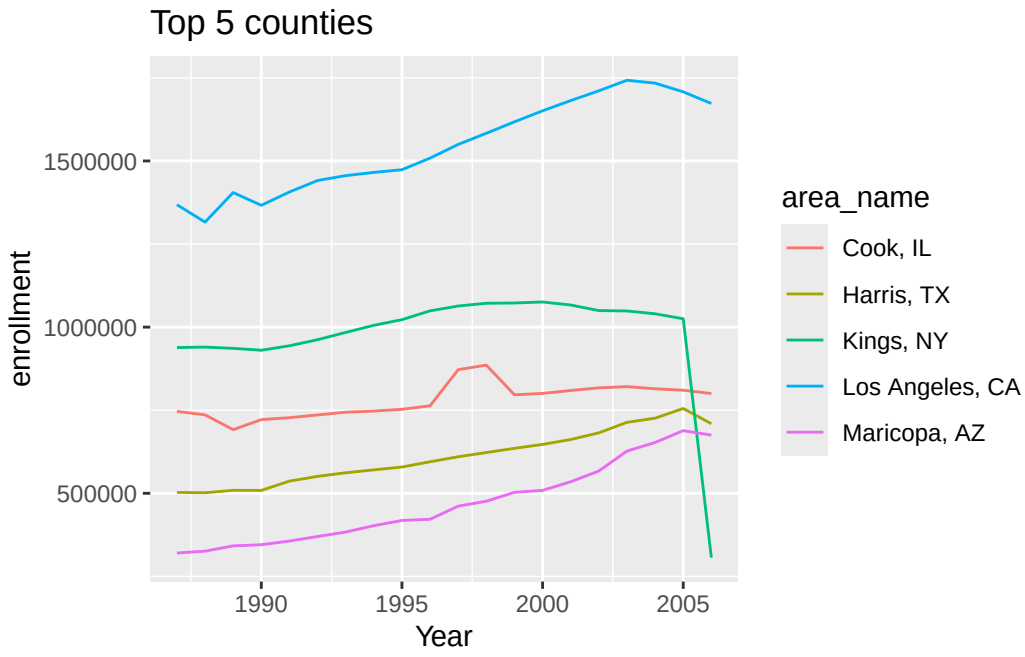

```r
## 4) Extract the county-level tibble and plot with various options
county_df <- edu_combined[[1]]

# 4a) NC, top 20 counties
plot.county(county_df,
     var_name   = "enrollment",
     state_name = "NC",
     top_bottom = "top",
     n          = 20)
```



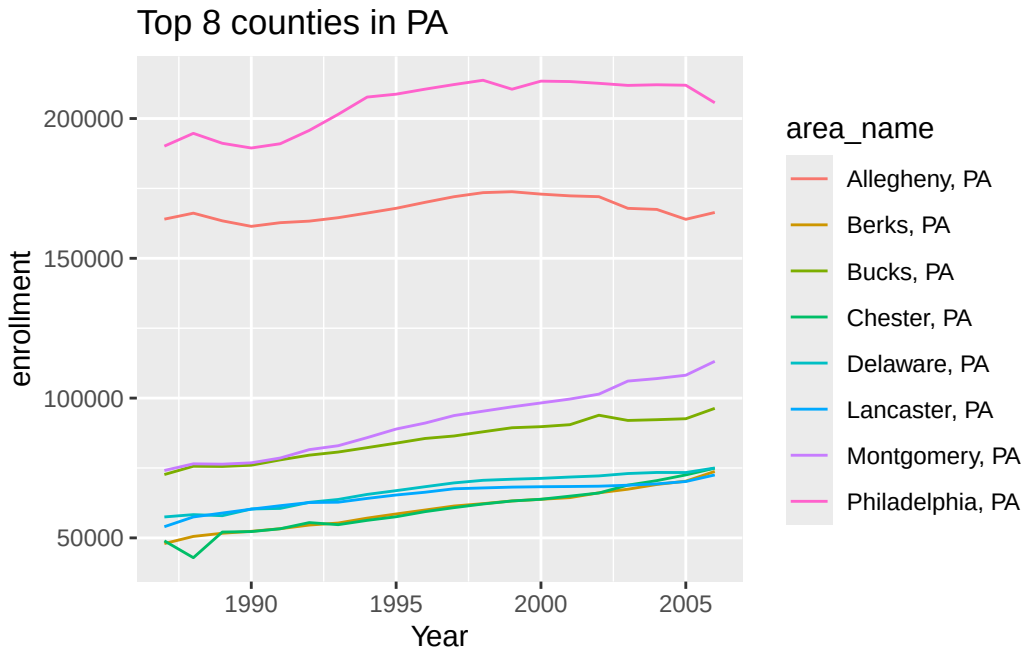

```
# 4b) SC, bottom 7 counties
plot.county(county_df,
     var_name   = "enrollment",
     state_name = "SC",
     top_bottom = "bottom",
     n          = 7)
```




```
# 4c) defaults (all states, top 5)
plot.county(county_df,
     var_name = "enrollment")
```



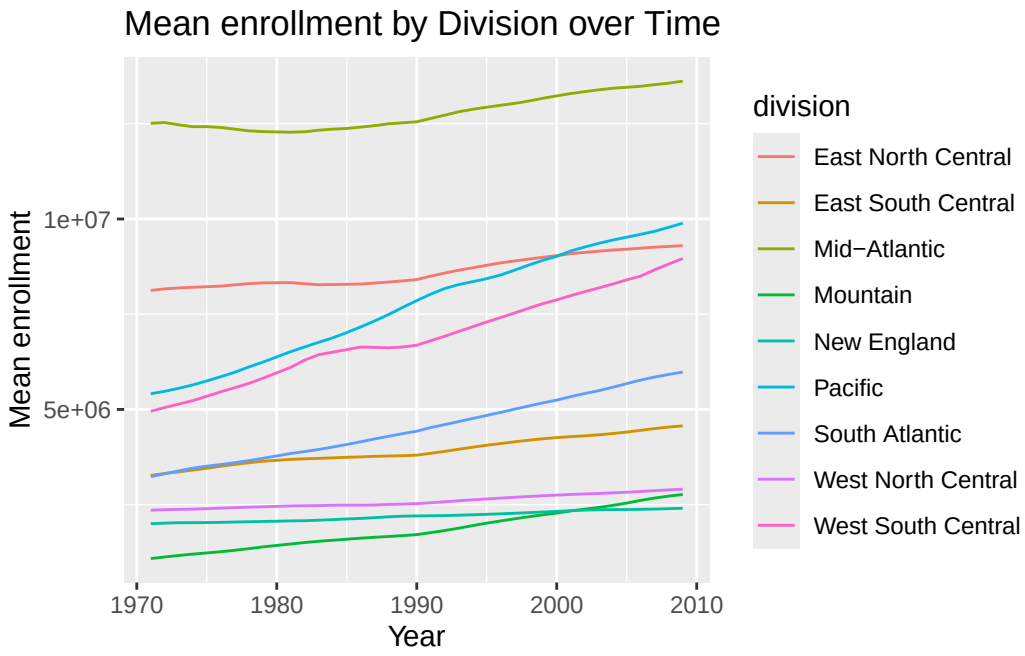
```r
# 4d) PA, top 8 counties
plot.county(county_df,
     var_name   = "enrollment",
     state_name = "PA",
     top_bottom = "top",
     n          = 8)
```



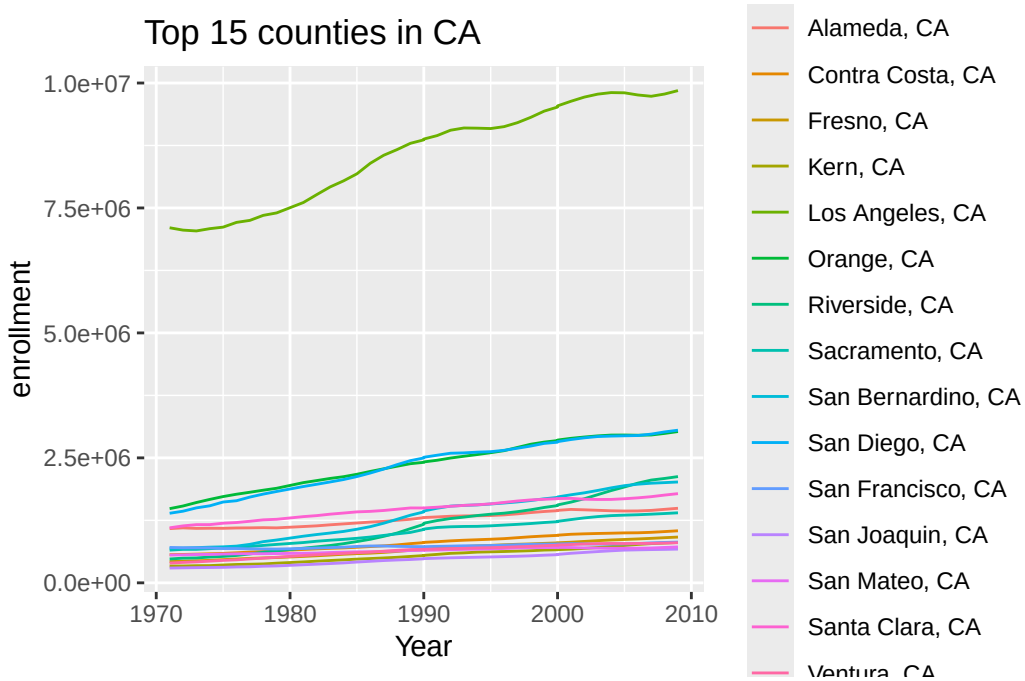
Process the four PST01 datasets

```r
pst_01a <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/PST01a.csv",
                      value_var = "enrollment")
pst_01b <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/PST01b.csv",
                      value_var = "enrollment")
pst_01c <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/PST01c.csv",
                      value_var = "enrollment")
pst_01d <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/PST01d.csv",
                      value_var = "enrollment")

## 2) Combine
pst_combined_1 <- combine_two(pst_01a, pst_01b)
pst_combined_2 <- combine_two(pst_01c, pst_01d)
```

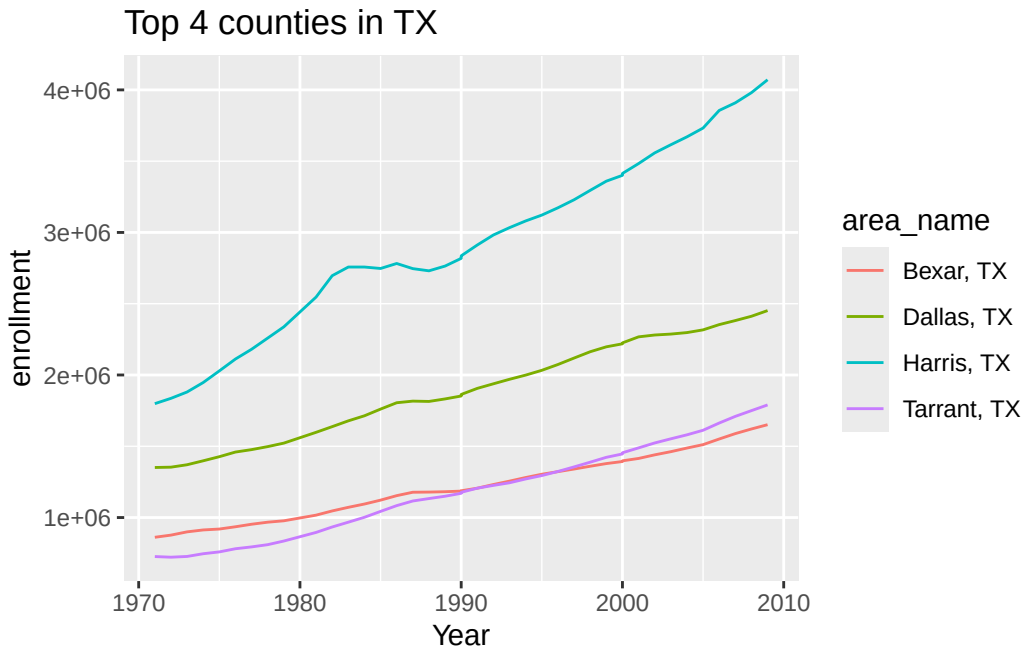
```r
pst_combined <- combine_two(pst_combined_1, pst_combined_2)

## 3) Plot the PST01 combined results on State data
plot.state(pst_combined[[2]],  var_name = "enrollment")
```



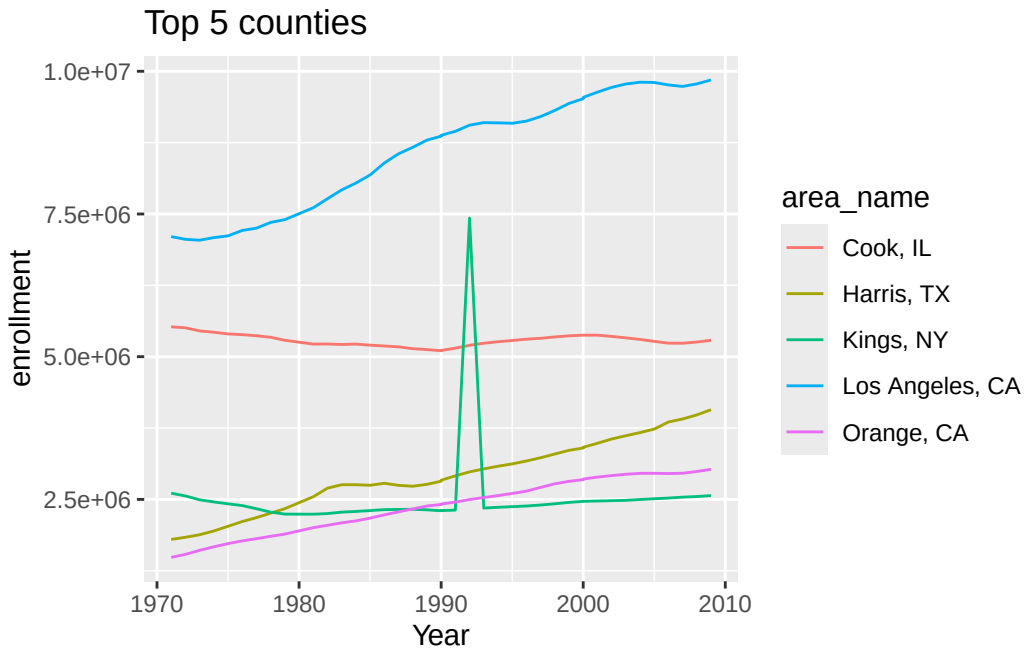

```r
## 3) Plot the PST01 combined results on Specified County data
plot.county(pst_combined[[1]], var_name = "enrollment", state_name = "CA", top_bottom = "top
```




```
plot.county(pst_combined[[1]], var_name = "enrollment", state_name = "TX", top_bottom = "top"
```



```
plot.county(pst_combined[[1]], var_name = "enrollment")  # defaults
```




```
plot.county(pst_combined[[1]], var_name = "enrollment", state_name = "NY", top_bottom = "top"
```