Project 1

Trever Yoder and Koji Takagi

Question 1: Load Packages and Functions

In this section, we load all necessary libraries and our custom functions file.

```
library(tidyverse)
library(readr)
library(ggplot2)

# Load custom functions
source("functions.R")
```

Question 2: Process the EDU Data Sets

We run our wrapper function on the two EDU datasets and inspect the results.

```
edu1 <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/EDU01a.csv", value = "Enro
edu2 <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/EDU01b.csv", value = "Enro

# Inspect to ensure correctness
head(edu1$county)
```

```
# A tibble: 6 x 7
  area_name   STCOU Survey     `Enrollment Value`  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
```

```
head(edu1$noncounty)
```

```
# A tibble: 6 x 7
  area_name     STCOU Survey     `Enrollment Value`  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
```

Question 3: Combine EDU Data Sets

Here we use our combining function to merge the two processed data sets.

```
edu_combined <- combine_wrapper_results(edu1, edu2)
head(edu_combined$county)
```

```
# A tibble: 6 x 7
  area_name   STCOU Survey     `Enrollment Value`  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
```

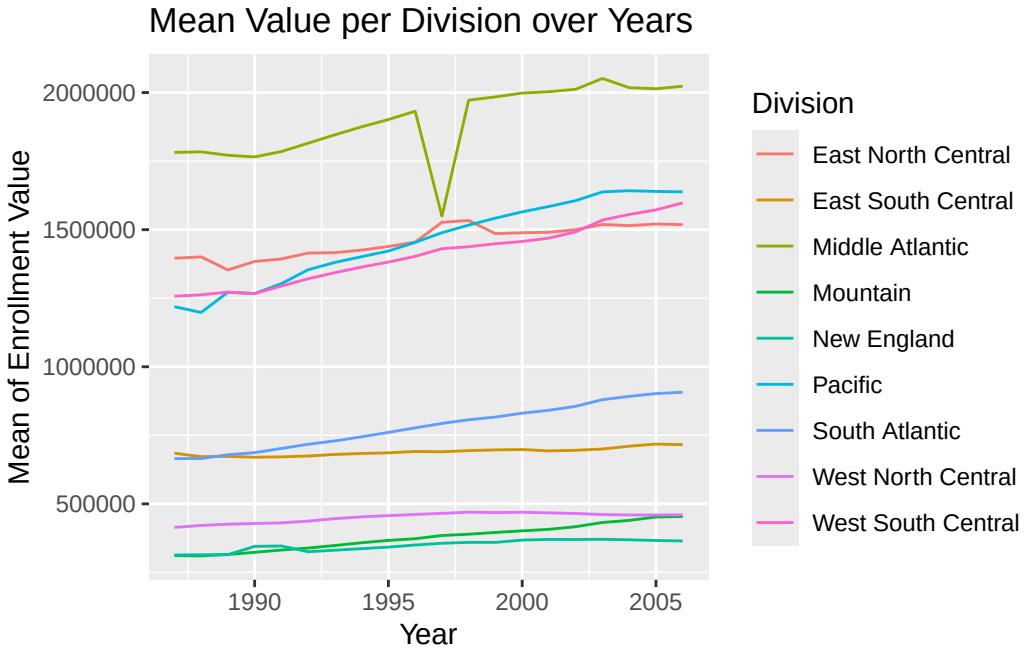
```
head(edu_combined$noncounty)
```

```
# A tibble: 6 x 7
  area_name     STCOU Survey     `Enrollment Value`  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
```

Question 4: State Plot for EDU Data

This plot shows the mean enrollment by Division across years.

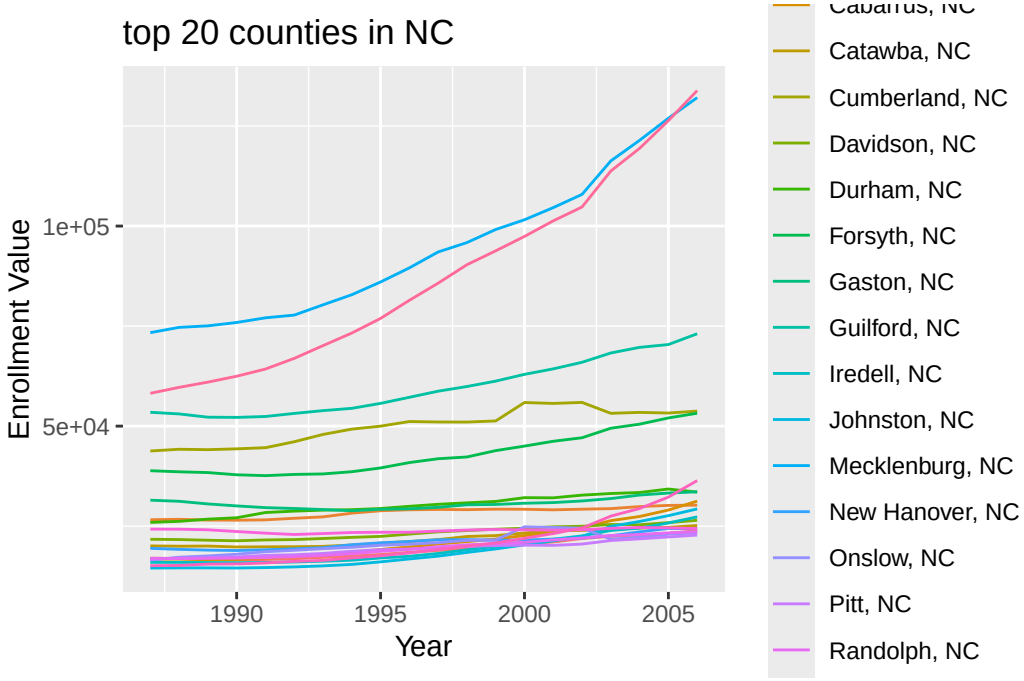
```
plot(edu_combined$noncounty, var_name = "Enrollment Value")
```



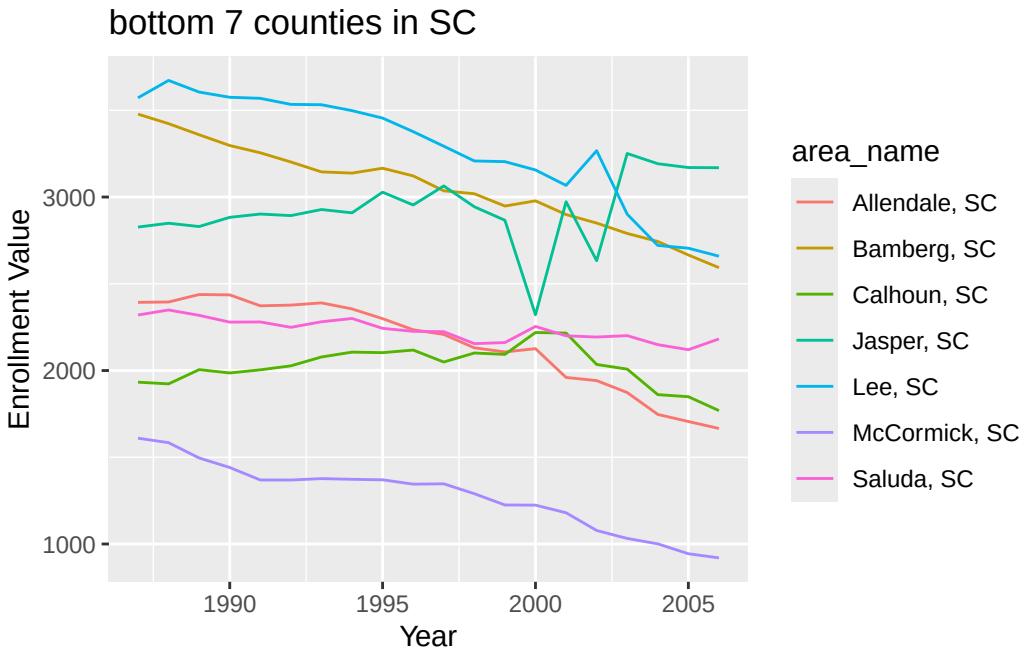
Question 5: County Plots for EDU Data

Below are various plots for county data, demonstrating flexibility in selecting state, top/bottom, and count.

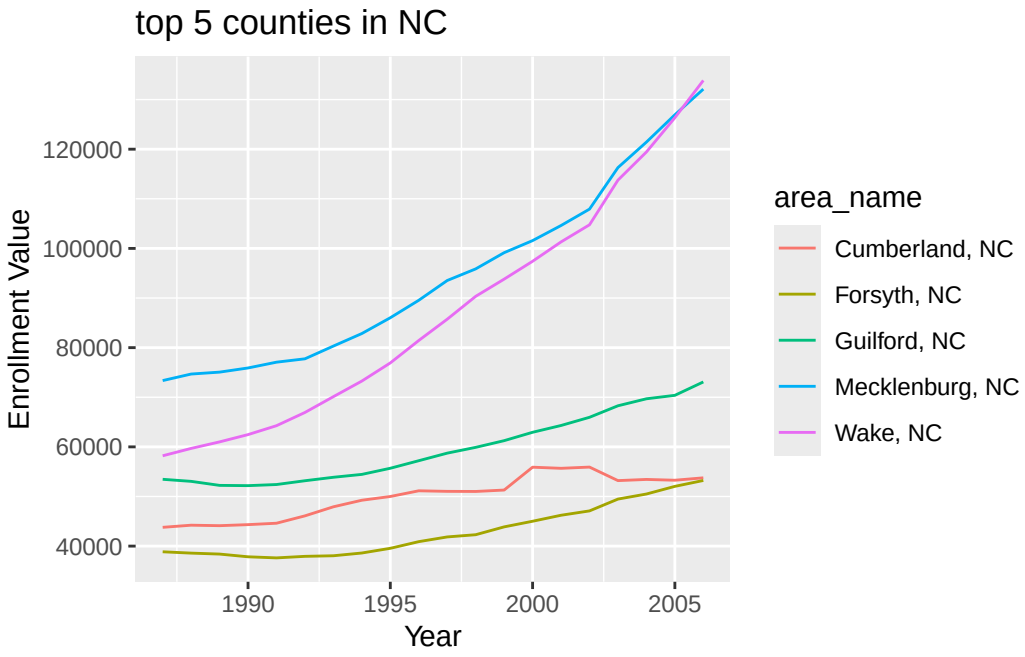
```
# NC, top 20
plot(edu_combined$county, var_name = "Enrollment Value", state = "NC", top_or_bottom = "to
```



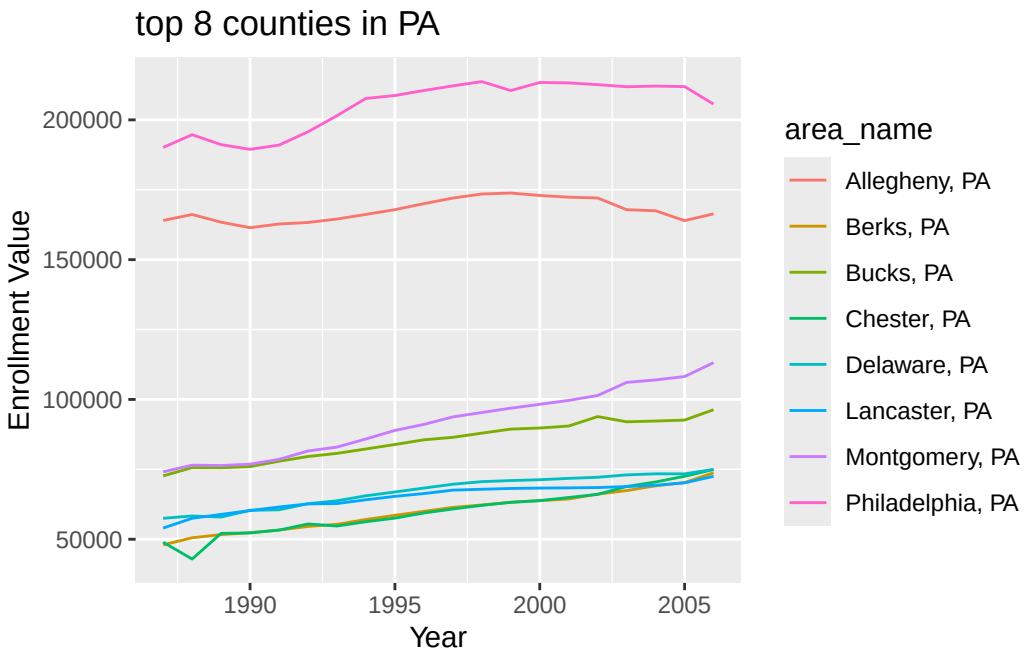

```
# SC, bottom 7
plot(edu_combined$county, var_name = "Enrollment Value", state = "SC", top_or_bottom = "bo
```



```r
# Default (uses NC top 5)
plot(edu_combined$county, var_name = "Enrollment Value")
```




```r
# PA, top 8
plot(edu_combined$county, var_name = "Enrollment Value", state = "PA", top_or_bottom = "to
```



Question 6: Process PST Data Sets

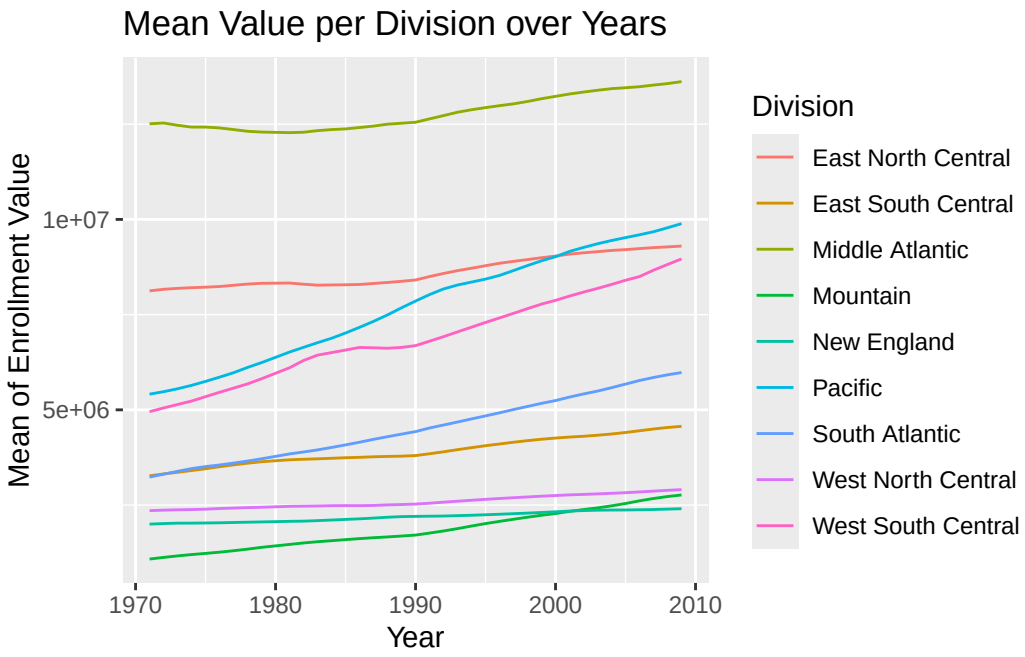
We repeat the same workflow for the four PST datasets.

```
pst1 <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/PST01a.csv", value = "Enro
pst2 <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/PST01b.csv", value = "Enro
pst3 <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/PST01c.csv", value = "Enro
pst4 <- my_wrapper("https://www4.stat.ncsu.edu/~online/datasets/PST01d.csv", value = "Enro

# Combine step by step
pst12 <- combine_wrapper_results(pst1, pst2)
pst34 <- combine_wrapper_results(pst3, pst4)
pst_combined <- combine_wrapper_results(pst12, pst34)
```

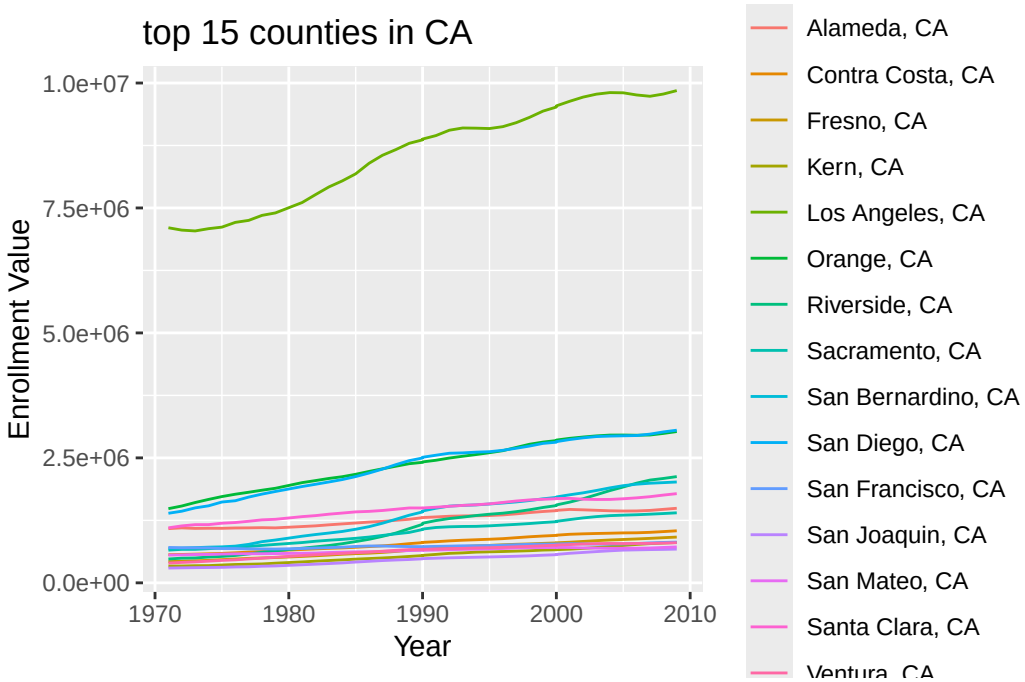
Question 7: State Plot for PST Data

```
plot(pst_combined$noncounty, var_name = "Enrollment Value")
```

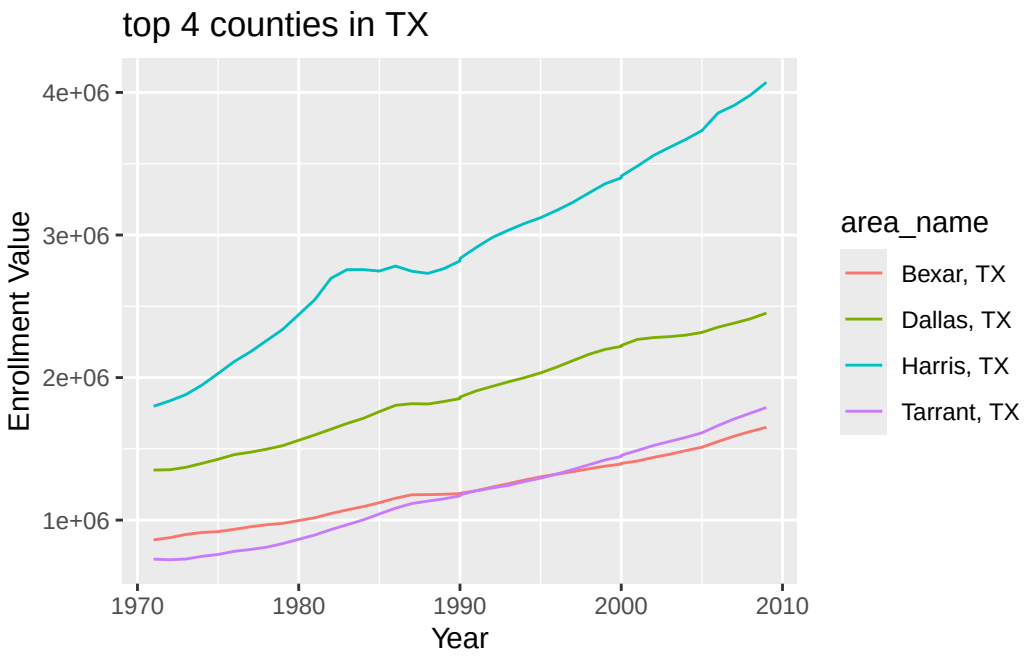


Question 8: County Plots for PST Data

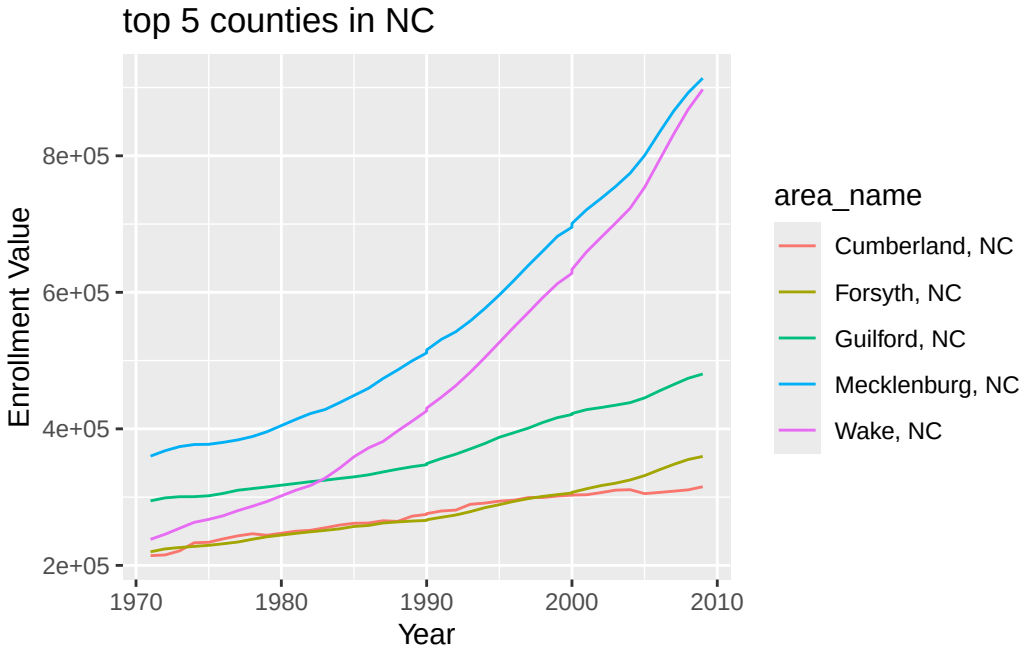
```
# CA, top 15
plot(pst_combined$county, var_name = "Enrollment Value", state = "CA", top_or_bottom = "to
```



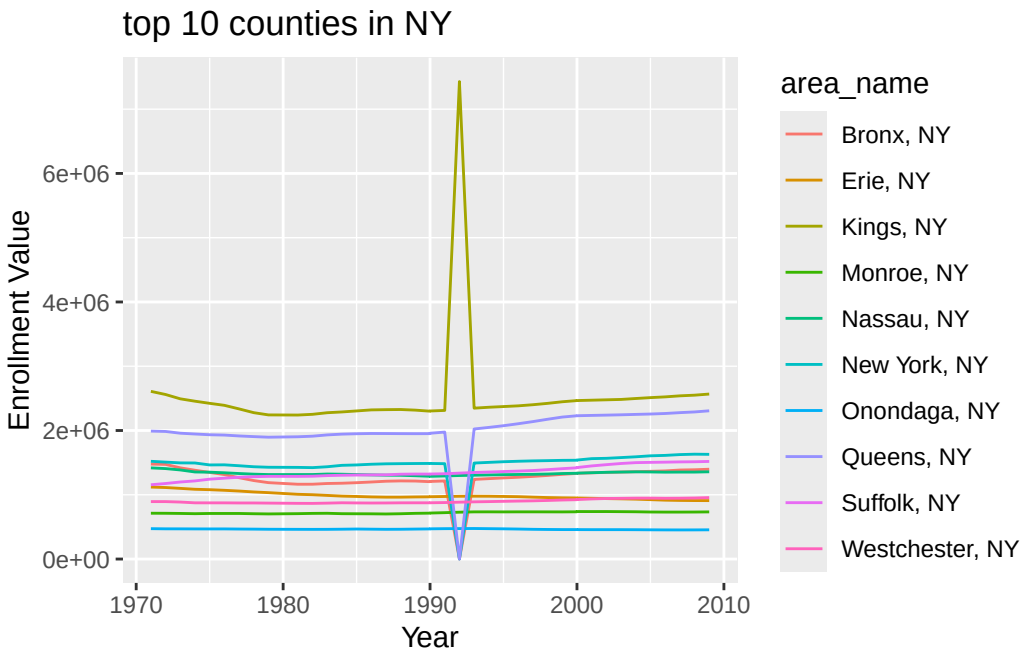

```
# TX, top 4
plot(pst_combined$county, var_name = "Enrollment Value", state = "TX", top_or_bottom = "to
```



```r
# Default
plot(pst_combined$county, var_name = "Enrollment Value")
```




```r
# NY, top 10
plot(pst_combined$county, var_name = "Enrollment Value", state = "NY", top_or_bottom = "to
```


top 10 counties in NY
Enrollment Value
6e+06
4e+06
2e+06
0e+00
1970
1980
1990
2000
2010
Year
area_name
Bronx, NY
Erie, NY
Kings, NY
Monroe, NY
Nassau, NY
New York, NY
Onondaga, NY
Queens, NY
Suffolk, NY
Westchester, NY