

Introduction to Geographic Information Systems

Aditya Ranganath

2022-02-08

Contents

1	Introduction	2
2	Loading data	2
3	Cleaning and filtering data	3
3.1	Create a “Year” field	3
3.2	Filter by year	3
3.3	View the newly created <code>co_traffic_stops_2010</code> object	3
4	Transforming Data	4
4.1	Tabulate county-level count of traffic stops by race	4
4.2	Reshape the data	4
4.3	Calculate total stops for each county in <code>co_county_summary_wide</code>	4
4.4	Clean <code>co_county_summary_wide</code> and assign to new object	5
5	Defining an index of racial bias in traffic stops	5
5.1	Read in and join 2010 census data to <code>co_county_black_stops</code>	5
5.2	Join census data to <code>co_county_black_stops</code>	6
5.3	Define the variables that will be used in the bias index	6
5.4	Calculate the bias index	7
5.5	Compute summary statistics for the bias index	7
5.6	Visualize county-level variation in the bias index using <code>ggplot</code>	8
6	Mapping the bias index	12
6.1	Read in and view the shapefile of CO counties	12
6.2	Join <code>co_counties_census_trafficstops</code> to the shapefile of Colorado counties	15
6.3	Display the bias index on a map of Colorado counties	16

7 Refining and formatting the maps of the Colorado traffic-stop bias index	20
7.1 Refining the map of the continuous bias variable	20
7.2 Refining the map of the categorical bias variable	22
8 Exporting maps	23
9 Summary script	24
A Using tidycensus to extract relevant census data	26
A.1 Step 1: Define your variables	26
A.2 Step 2: Extract the variables using tidycensus	28
A.3 Step 3: Clean the tidycensus dataset	28
A.4 Step 4: Define new variables	28
A.5 Step 5: Finalize and export the dataset	28

1 Introduction

```
library(tidyverse)
library(tmap)
library(sf)
```

2 Loading data

Reading in data

```
# Read in Stanford police data for Colorado and assign to object named "co_traffic_stops"
co_traffic_stops<-read_csv("co_statewide_2020_04_01.csv")
```

```
##
## -- Column specification -----
## cols(
##   .default = col_character(),
##   date = col_date(format = ""),
##   time = col_logical(),
##   subject_age = col_double(),
##   arrest_made = col_logical(),
##   citation_issued = col_logical(),
##   warning_issued = col_logical(),
##   contraband_found = col_logical(),
##   search_conducted = col_logical()
## )
## i Use 'spec()' for the full column specifications.
```

Printing contents of co_traffic_stops object

```
co_traffic_stops
```

```
## # A tibble: 3,112,853 x 20
##   raw_row_number     date       time   location county_name subject_age subject_race subject_sex ...
##   <chr>           <date>     <lgl> <chr>    <chr>          <dbl> <chr>      <chr>    <chr>
## 1 1947986|1947987 2013-06-19 NA    19, I70~ Mesa County        26 hispanic   male    b9...
## 2 1537576          2012-08-24 NA    254, H2~ Jefferson ~        NA <NA>      <NA>    f3...
## 3 1581594          2012-09-23 NA    115, I7~ Logan Coun~        52 white     male    6e...
## 4 1009205          2011-08-25 NA    197, H8~ Douglas Co~        32 white     female   ea...
## 5 1932619          2013-06-08 NA    107, H2~ Kiowa Coun~        33 hispanic   male    d1...
## 6 1179436          2011-12-23 NA    48, 384~ Boulder Co~        NA <NA>      <NA>    b8...
## 7 1326795          2012-04-07 NA    0, R250~ Boulder Co~        39 white     male    4c...
## 8 1786795          2013-03-03 NA    19, E47~ Arapahoe C~        44 white     female   e6...
## 9 1552164          2012-09-02 NA    224, H2~ Park County        NA <NA>      <NA>    43...
## 10 1004281|1004282|10~ 2011-08-21 NA    R2000, ~ Adams Coun~       32 hispanic   male    dd...
## # ... with 3,112,843 more rows, and 11 more variables: officer_sex <chr>, type <chr>, violation <chr>,
## #   arrest_made <lgl>, citation_issued <lgl>, warning_issued <lgl>, outcome <chr>, contraband_found ...
## #   search_conducted <lgl>, search_basis <chr>, raw_Ethnicity <chr>
```

3 Cleaning and filtering data

3.1 Create a “Year” field

```
# Create "year" field
co_traffic_stops<-co_traffic_stops %>% mutate(Year=substr(co_traffic_stops$date, 1,4))
```

3.2 Filter by year

```
# Filter 2010 observations and assign to a new object
co_traffic_stops_2010<-co_traffic_stops %>% filter(Year==2010)
```

3.3 View the newly created co_traffic_stops_2010 object

```
# Print contents of "co_traffic_stops_2010" object
co_traffic_stops_2010
```

```
## # A tibble: 470,284 x 21
##   raw_row_number     date       time   location county_name subject_age subject_race subject_sex ...
##   <chr>           <date>     <lgl> <chr>    <chr>          <dbl> <chr>      <chr>    <chr>
## 1 188721|188722  2010-04-17 NA    2, 989,~ Montezuma ~        16 white     female   b7...
## 2 187958          2010-04-17 NA    991, 32 Montezuma ~        54 white     male    89...
## 3 188451          2010-04-17 NA    9, 280,~ Montezuma ~        49 hispanic   male    73...
## 4 186989|186990|1869~ 2010-04-17 NA    3, 277,~ Montezuma ~        16 white     male    3c...
## 5 186997|186998|1869~ 2010-04-17 NA    3, 277,~ Montezuma ~        37 white     male    3c...
## 6 186993|186994|1869~ 2010-04-17 NA    3, 277,~ Montezuma ~        39 white     male    3c...
```

```

## 7 600865          2010-12-21 NA    164.5, ~ Mineral Co~      110 <NA>      <NA>      0e
## 8 600477          2010-12-21 NA    163, 29~ Mineral Co~      110 <NA>      <NA>      c8
## 9 36625|36626     2010-01-20 NA    312, H5~ Pueblo Cou~      45 hispanic   male      50
## 10 275           2010-01-01 NA    127, H2~ Chaffee Co~      17 white     female     75
## # ... with 470,274 more rows, and 12 more variables: officer_sex <chr>, type <chr>, violation <chr>,
## #   arrest_made <lgl>, citation_issued <lgl>, warning_issued <lgl>, outcome <chr>, contraband_found <
## #   search_conducted <lgl>, search_basis <chr>, raw_Ethnicity <chr>, Year <chr>

```

4 Transforming Data

4.1 Tabulate county-level count of traffic stops by race

```

# Compute county-level count of traffic stops by race
co_county_summary<-co_traffic_stops_2010 %>%
  group_by(county_name) %>%
  count(subject_race)

```

4.2 Reshape the data

```

co_county_summary_wide<-co_county_summary %>%
  pivot_wider(names_from=subject_race, values_from=n)

```

4.3 Calculate total stops for each county in co_county_summary_wide

```

co_county_summary_wide<-co_county_summary_wide %>%
  rowwise() %>%
  mutate(total_stops=sum(c_across(where(is.integer))), na.rm=TRUE))

```

```
co_county_summary_wide
```

```

## # A tibble: 65 x 9
## # Rowwise: county_name
##   county_name      `asian/pacific islander` black hispanic other unknown white  `NA` total_stops
##   <chr>                <int> <int>    <int> <int> <int> <int> <int>      <int>
## 1 Adams County            582  1208     8012    36    462 20225  3825      34350
## 2 Alamosa County           18    43     1537     9    30  2427   414      4478
## 3 Arapahoe County          540  1819     1862    12    300 11089  1898      17520
## 4 Archuleta County          17    28     392     71    41  4125   417      5091
## 5 Baca County              11    61     288     NA     6  971    174      1511
## 6 Bent County               8    46     314     1     6 1155    278      1808
## 7 Boulder County            345   192     1050    10    180 9682   1594      13053
## 8 Broomfield County          32    22     104     3     18  690    226      1095
## 9 Chaffee County             43    37     361     9     71 4806   1194      6521
## 10 Cheyenne County            10    38     147     3     2  821     85      1106
## # ... with 55 more rows

```

4.4 Clean co_county_summary_wide and assign to new object

```
co_county_black_stops<-co_county_summary_wide %>%
  select(county_name, black, total_stops)
```

```
co_county_black_stops<-co_county_black_stops %>%
  rename(black_stops=black)
```

```
co_county_black_stops<-co_county_black_stops %>%
  filter(county_name!="NA")
```

```
co_county_black_stops
```

```
## # A tibble: 64 x 3
## # Rowwise: county_name
##   county_name     black_stops total_stops
##   <chr>           <int>      <int>
## 1 Adams County     1208      34350
## 2 Alamosa County    43       4478
## 3 Arapahoe County   1819      17520
## 4 Archuleta County   28       5091
## 5 Baca County        61       1511
## 6 Bent County         46       1808
## 7 Boulder County     192      13053
## 8 Broomfield County    22       1095
## 9 Chaffee County      37       6521
## 10 Cheyenne County     38       1106
## # ... with 54 more rows
```

5 Defining an index of racial bias in traffic stops

5.1 Read in and join 2010 census data to co_county_black_stops

Read in census data

```
co_counties_census_2010<-read_csv("co_county_decennial_census.csv")
```

```
##
## -- Column specification --
## cols(
##   GEOID = col_character(),
##   County = col_character(),
##   total_pop = col_double(),
##   total_black_pop_over17 = col_double(),
##   total_pop_over17 = col_double()
## )
```

```
# Prints contents of "co_counties_census_2010"
co_counties_census_2010

## # A tibble: 64 x 5
##   GEOID County      total_pop total_black_pop_over17 total_pop_over17
##   <chr> <chr>       <dbl>             <dbl>            <dbl>
## 1 08023 Costilla County     3524              18            2788
## 2 08025 Crowley County    5823              556            5034
## 3 08027 Custer County     4255              37            3525
## 4 08029 Delta County      30952             139            24101
## 5 08031 Denver County     600158            45338           471392
## 6 08035 Douglas County    285465            2447           198453
## 7 08033 Dolores County    2064                4            1602
## 8 08049 Grand County      14843               43            11825
## 9 08039 Elbert County     23086              122            17232
## 10 08041 El Paso County   622263             27280           459587
## # ... with 54 more rows
```

5.2 Join census data to co_county_black_stops

```
co_counties_census_trafficstops<-full_join(co_county_black_stops, co_counties_census_2010,
                                             by=c("county_name"="County"))
```

```
co_counties_census_trafficstops
```

```
## # A tibble: 64 x 7
## # Rowwise: county_name
##   county_name black_stops total_stops GEOID total_pop total_black_pop_over17 total_pop_over17
##   <chr>        <int>      <int> <chr>    <dbl>            <dbl>            <dbl>
## 1 Adams County      1208      34350 08001    441603            9396            315480
## 2 Alamosa County     43        4478 08003    15445             142            11617
## 3 Arapahoe County    1819      17520 08005    572003            40558            424679
## 4 Archuleta County     28        5091 08007    12084              19            9676
## 5 Baca County         61        1511 08009    3788              15            2974
## 6 Bent County          46        1808 08011    6499              486            5403
## 7 Boulder County      192       13053 08013    294567            1961            231813
## 8 Broomfield County     22        1095 08014    55889             415            41237
## 9 Chaffee County        37       6521 08015    17809             264            14821
## 10 Cheyenne County      38       1106 08017    1836              4            1386
## # ... with 54 more rows
```

5.3 Define the variables that will be used in the bias index

```
co_counties_census_trafficstops<-co_counties_census_trafficstops %>%
  mutate(black_stop_pct=((black_stops/total_stops)*100),
        black_pop_pct=((total_black_pop_over17/total_pop_over17)*100)
```

```
co_counties_census_trafficstops
```

```
## # A tibble: 64 x 9
## # Rowwise: county_name
##   county_name    black_stops total_stops GEOID total_pop total_black_pop_ove~ total_pop_over17 b
##   <chr>           <int>      <int> <chr>     <dbl>          <dbl>            <dbl>
## 1 Adams County      1208      34350 08001    441603         9396        315480
## 2 Alamosa County       43      4478 08003    15445          142        11617
## 3 Arapahoe County     1819      17520 08005    572003        40558        424679
## 4 Archuleta County      28      5091 08007    12084          19        9676
## 5 Baca County          61      1511 08009    3788          15        2974
## 6 Bent County           46      1808 08011    6499          486        5403
## 7 Boulder County        92      13053 08013    294567        1961       231813
## 8 Broomfield County      22      1095 08014    55889          415       41237
## 9 Chaffee County         37      6521 08015    17809          264       14821
## 10 Cheyenne County        38      1106 08017    1836             4        1386
## # ... with 54 more rows, and 1 more variable: black_pop_pct <dbl>
```

5.4 Calculate the bias index

```
co_counties_census_trafficstops<-co_counties_census_trafficstops %>%
  mutate(excess_stops_index=black_stop_pct-black_pop_pct)
```

```
co_counties_census_trafficstops
```

```
## # A tibble: 64 x 10
## # Rowwise: county_name
##   county_name    black_stops total_stops GEOID total_pop total_black_pop_ove~ total_pop_over17 b
##   <chr>           <int>      <int> <chr>     <dbl>          <dbl>            <dbl>
## 1 Adams County      1208      34350 08001    441603         9396        315480
## 2 Alamosa County       43      4478 08003    15445          142        11617
## 3 Arapahoe County     1819      17520 08005    572003        40558        424679
## 4 Archuleta County      28      5091 08007    12084          19        9676
## 5 Baca County          61      1511 08009    3788          15        2974
## 6 Bent County           46      1808 08011    6499          486        5403
## 7 Boulder County        92      13053 08013    294567        1961       231813
## 8 Broomfield County      22      1095 08014    55889          415       41237
## 9 Chaffee County         37      6521 08015    17809          264       14821
## 10 Cheyenne County        38      1106 08017    1836             4        1386
## # ... with 54 more rows, and 2 more variables: black_pop_pct <dbl>, excess_stops_index <dbl>
```

5.5 Compute summary statistics for the bias index

```
describe(co_counties_census_trafficstops$excess_stops_index)
```

```
##    vars   n mean    sd median trimmed  mad   min  max range skew kurtosis    se
## X1     1 62 0.53 2.04    0.47    0.65 0.59 -9.84 4.26 14.1 -2.45   10.67 0.26
```

5.6 Visualize county-level variation in the bias index using ggplot

```
co_counties_census_trafficstops<-co_counties_census_trafficstops %>%
  mutate(County=str_remove(county_name, " County"))

bias_graph<-co_counties_census_trafficstops %>%
  drop_na(excess_stops_index) %>%
  ggplot()+
  geom_col(aes(x=reorder(County, excess_stops_index), y=excess_stops_index))+
  coord_flip()+
  labs(title="Racial Discrimination in Police Stops, by County (Colorado)", x="County", y="Bias Index")
  theme(axis.title.x = element_text(size = 9))+  

  scale_y_continuous(breaks=c(-10, -7.5, -5, -2.5, 0, 2.5, 5, 7.5))+  

  expand_limits(y=c(-10,10))+  

  theme(plot.title=element_text(hjust=0.5))

bias_graph

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'Black'## of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs': dot subs##
## <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'Black'## of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs': dot subs##
## <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'Black'## of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs': dot subs##
## <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'Black'## of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs': dot subs##
## <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'Black'## of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs': dot subs##
## <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'Black'## of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs': dot subs##
## <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'Black'## of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs': dot subs##
## <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'Black'## of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs': dot subs##
## <80>
```

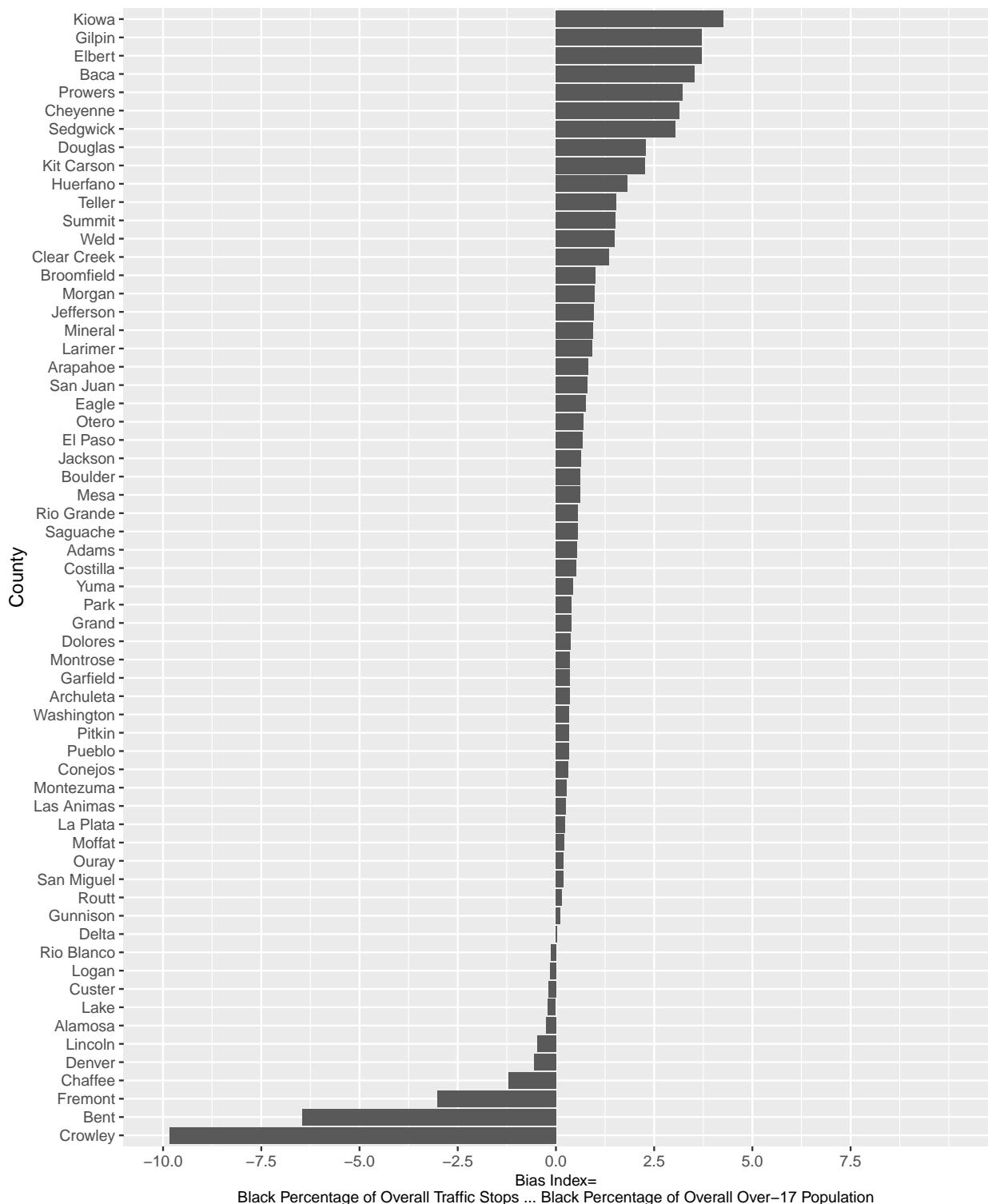


```
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'mbcsToSbcs'
## Percentage of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs'
## substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'mbcsToSbcs'
## Percentage of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs'
## substituted for <80>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, : conversion failure on 'mbcsToSbcs'
## Percentage of Overall Traffic Stops - Black Percentage of Overall Over-17 Population' in 'mbcsToSbcs'
## substituted for <93>
```

Racial Discrimination in Police Stops, by County (Colorado)



6 Mapping the bias index

6.1 Read in and view the shapefile of CO counties

```
# Reads in shapefile and assigns to object named "co_counties_shapefile"
co_counties_shapefile<-st_read("tl_2019_08_county.shp")

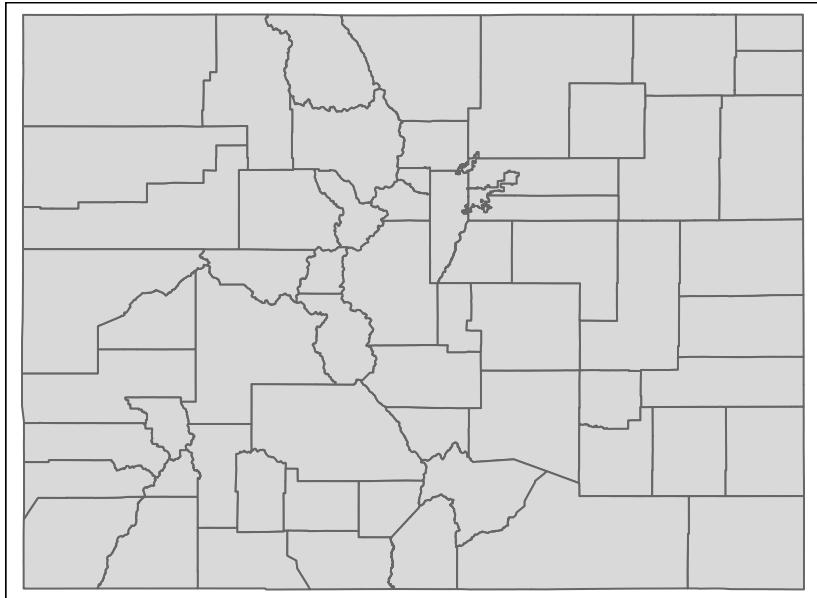
## Reading layer 'tl_2019_08_county' from data source '/Users/adra7980/Documents/CU_workshops/gis/data/'
## Simple feature collection with 64 features and 17 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:            xmin: -109.0602 ymin: 36.99245 xmax: -102.0415 ymax: 41.00344
## geographic CRS: NAD83

co_counties_shapefile

## Simple feature collection with 64 features and 17 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:            xmin: -109.0602 ymin: 36.99245 xmax: -102.0415 ymax: 41.00344
## geographic CRS: NAD83
## First 10 features:
##   STATEFP COUNTYFP COUNTYNS GEOID      NAME      NAMELSAD LSAD CLASSFP MTFCC CSAFP CBSAfp METDIVFP
## 1       08        109 00198170 08109 Saguache Saguache County    06     H1 G4020 <NA> <NA> <NA>
## 2       08        115 00198173 08115 Sedgwick Sedgwick County    06     H1 G4020 <NA> <NA> <NA>
## 3       08        017 00198124 08017 Cheyenne Cheyenne County    06     H1 G4020 <NA> <NA> <NA>
## 4       08        027 00198129 08027 Custer   Custer County    06     H1 G4020 <NA> <NA> <NA>
## 5       08        067 00198148 08067 La Plata La Plata County    06     H1 G4020 <NA> 20420 <NA>
## 6       08        111 00198171 08111 San Juan San Juan County    06     H1 G4020 <NA> <NA> <NA>
## 7       08        097 00198164 08097 Pitkin   Pitkin County    06     H1 G4020 233 24060 <NA>
## 8       08        093 00198162 08093 Park     Park County    06     H1 G4020 216 19740 <NA>
## 9       08        003 00198117 08003 Alamosa Alamosa County    06     H1 G4020 <NA> <NA> <NA>
## 10      08        099 00198165 08099 Prowers Prowers County    06     H1 G4020 <NA> <NA> <NA>
##   ALAND    AWATER   INTPTLAT   INTPTLON
## 1 8206547705 4454510 +38.0316514 -106.2346662 MULTIPOLYGON ((((-106.8714 3...
## 2 1419419016 3530746 +40.8715679 -102.3553579 MULTIPOLYGON ((((-102.6521 4...
## 3 4605713960 8166129 +38.8356456 -102.6017914 MULTIPOLYGON ((((-102.5769 3...
## 4 1913031921 3364150 +38.1019955 -105.3735123 MULTIPOLYGON ((((-105.7969 3...
## 5 4376255148 25642578 +37.2873673 -107.8397178 MULTIPOLYGON ((((-108.2952 3...
## 6 1003660672 2035929 +37.7810492 -107.6702567 MULTIPOLYGON ((((-107.9751 3...
## 7 2514104907 6472577 +39.2175376 -106.9161587 MULTIPOLYGON ((((-106.9154 3...
## 8 5682182508 43519840 +39.1189141 -105.7176479 MULTIPOLYGON ((((-105.9751 3...
## 9 1871465874 1847610 +37.5684423 -105.7880414 MULTIPOLYGON ((((-106.0393 3...
## 10 4243429484 15345176 +37.9581814 -102.3921613 MULTIPOLYGON ((((-102.2111 3...

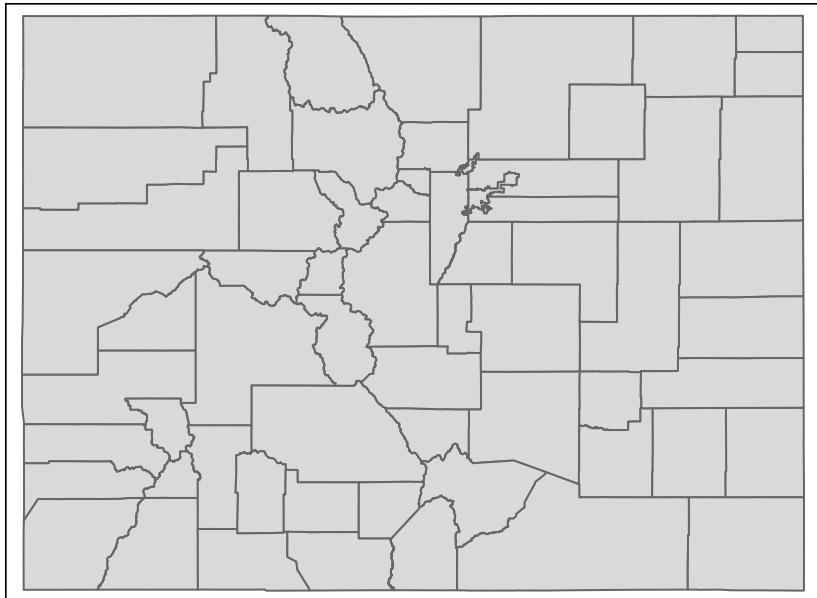
## tmap mode set to plotting

tm_shape(co_counties_shapefile) +
  tm_polygons()
```



```
co_counties_map<-tm_shape(co_counties_shapefile)+  
  tm_polygons()
```

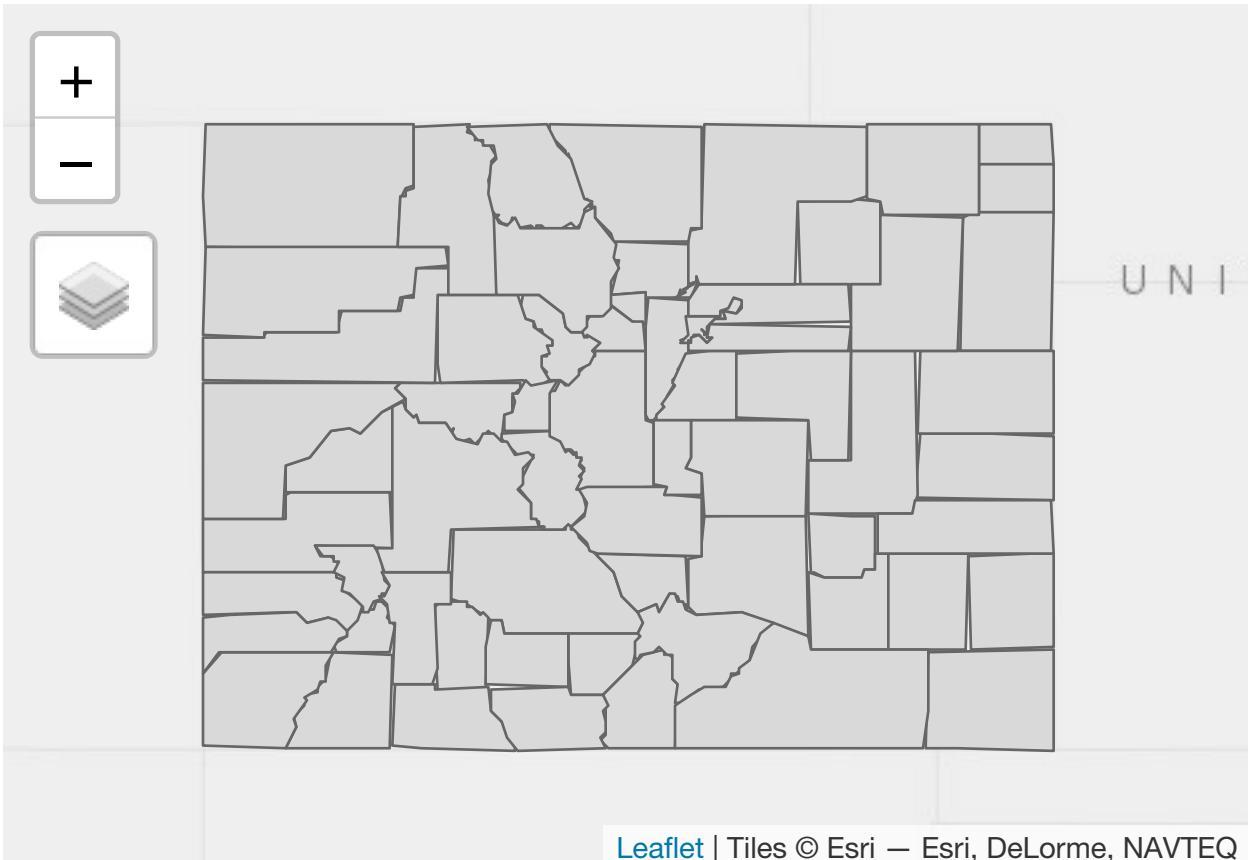
```
co_counties_map
```



```
tmap_mode("view")
```

```
## tmap mode set to interactive viewing
```

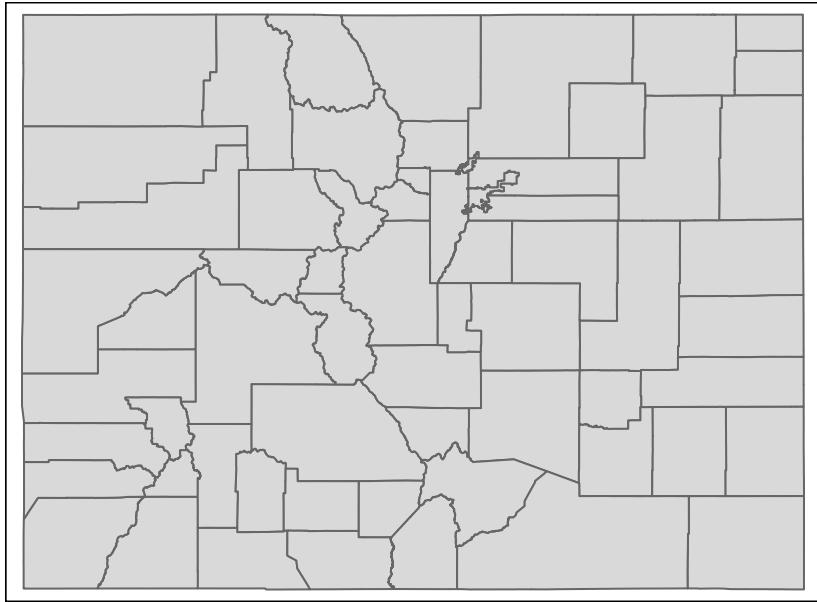
```
co_counties_map
```



```
tmap_mode("plot")
```

```
## tmap mode set to plotting
```

```
co_counties_map
```



6.2 Join co_counties_census_trafficstops to the shapefile of Colorado counties

```

county_shapefile_biasIndex<-full_join(co_counties_shapefile, co_counties_census_trafficstops,
                                         by="GEOID")

county_shapefile_biasIndex

## Simple feature collection with 64 features and 27 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:            xmin: -109.0602 ymin: 36.99245 xmax: -102.0415 ymax: 41.00344
## geographic CRS: NAD83
## First 10 features:
##   STATEFP COUNTYFP COUNTYNS GEOID      NAME      NAMELSAD LSAD CLASSFP MTFCC CSAFP CBSAFP METDIVFP
## 1       08     109 00198170 08109 Saguache Saguache County    06    H1 G4020 <NA> <NA> <NA>
## 2       08     115 00198173 08115 Sedgwick Sedgwick County    06    H1 G4020 <NA> <NA> <NA>
## 3       08     017 00198124 08017 Cheyenne Cheyenne County    06    H1 G4020 <NA> <NA> <NA>
## 4       08     027 00198129 08027 Custer   Custer County    06    H1 G4020 <NA> <NA> <NA>
## 5       08     067 00198148 08067 La Plata La Plata County    06    H1 G4020 <NA> 20420 <NA>
## 6       08     111 00198171 08111 San Juan San Juan County    06    H1 G4020 <NA> <NA> <NA>
## 7       08     097 00198164 08097 Pitkin   Pitkin County    06    H1 G4020 233 24060 <NA>
## 8       08     093 00198162 08093 Park     Park County    06    H1 G4020 216 19740 <NA>
## 9       08     003 00198117 08003 Alamosa Alamosa County    06    H1 G4020 <NA> <NA> <NA>
## 10      08     099 00198165 08099 Prowers Prowers County    06    H1 G4020 <NA> <NA> <NA>
##   ALAND    AWATER    INTPTLAT    INTPTLON      county_name black_stops total_stops total_pop
## 1 8206547705 4454510 +38.0316514 -106.2346662 Saguache County        20     2741    6108
## 2 1419419016 3530746 +40.8715679 -102.3553579 Sedgwick County        26      762    2379
## 3 4605713960 8166129 +38.8356456 -102.6017914 Cheyenne County       38     1106    1836
## 4 1913031921 3364150 +38.1019955 -105.3735123 Custer County         1      118    4255
## 5 4376255148 25642578 +37.2873673 -107.8397178 La Plata County      70    11305    51334
## 6 1003660672 2035929 +37.7810492 -107.6702567 San Juan County        1      125     699

```

```

## 7 2514104907 6472577 +39.2175376 -106.9161587 Pitkin County 4 487 17148
## 8 5682182508 43519840 +39.1189141 -105.7176479 Park County 64 8057 16206
## 9 1871465874 1847610 +37.5684423 -105.7880414 Alamosa County 43 4478 15445
## 10 4243429484 15345176 +37.9581814 -102.3921613 Prowers County 247 6594 12551
## total_black_pop_over17 total_pop_over17 black_stop_pct black_pop_pct excess_stops_index County
## 1 8 4692 0.7296607 0.1705030 0.5591577 Saguache
## 2 7 1919 3.4120735 0.3647733 3.0473002 Sedgwick
## 3 4 1386 3.4358047 0.2886003 3.1472044 Cheyenne
## 4 37 3525 0.8474576 1.0496454 -0.2021878 Custer
## 5 160 40822 0.6191950 0.3919455 0.2272495 La Plata
## 6 0 571 0.8000000 0.0000000 0.8000000 San Juan
## 7 69 14149 0.8213552 0.4876670 0.3336883 Pitkin
## 8 52 13098 0.7943403 0.3970072 0.3973331 Park
## 9 142 11617 0.9602501 1.2223466 -0.2620964 Alamosa
## 10 47 9147 3.7458295 0.5138297 3.2319999 Prowers
## geometry
## 1 MULTIPOLYGON (((-106.8714 3...
## 2 MULTIPOLYGON (((-102.6521 4...
## 3 MULTIPOLYGON (((-102.5769 3...
## 4 MULTIPOLYGON (((-105.7969 3...
## 5 MULTIPOLYGON (((-108.2952 3...
## 6 MULTIPOLYGON (((-107.9751 3...
## 7 MULTIPOLYGON (((-106.9154 3...
## 8 MULTIPOLYGON (((-105.9751 3...
## 9 MULTIPOLYGON (((-106.0393 3...
## 10 MULTIPOLYGON (((-102.2111 3...

```

6.3 Display the bias index on a map of Colorado counties

6.3.1 Make a rough draft of a map

```

traffic_stop_bias_map<-tm_shape(county_shapefile_biasIndex)+  

  tm_polygons(col="excess_stops_index",  

              palette="YlOrRd",  

              textNA="No Data",  

              n=5,  

              style="jenks",  

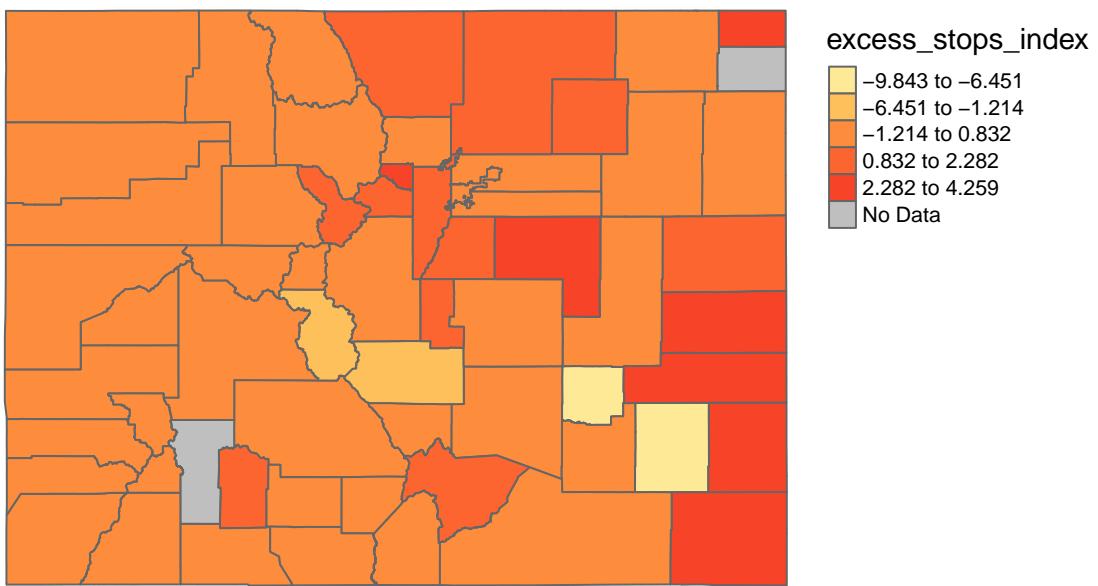
              midpoint=F)+  

  tm_layout(frame=FALSE,  

            legend.outside=TRUE)

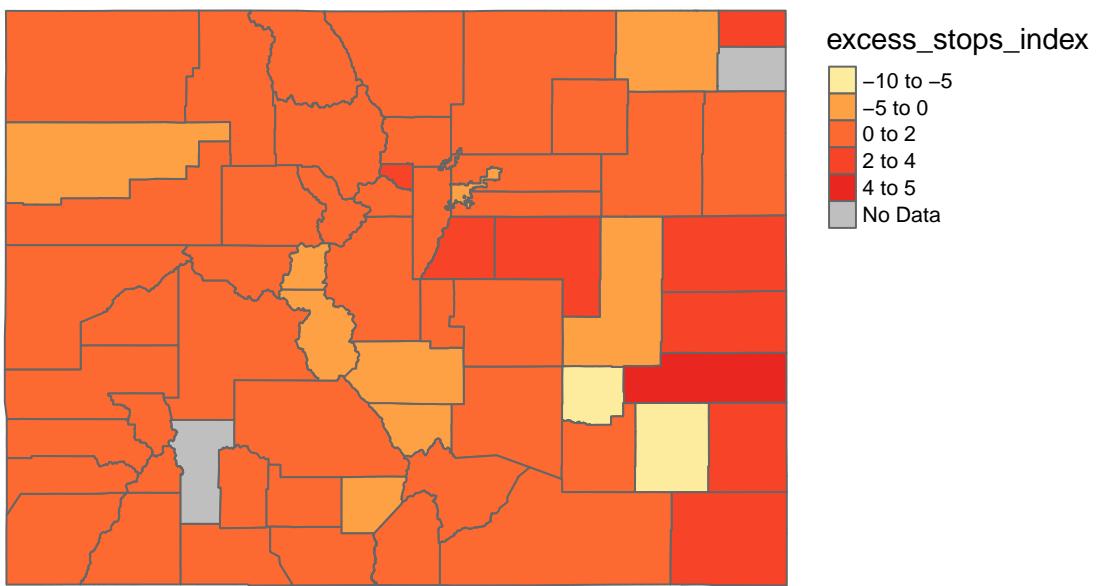
traffic_stop_bias_map

```



6.3.2 Make a map with custom breaks

```
traffic_stop_bias_map<-tm_shape(county_shapefile_biasIndex)+  
  tm_polygons(col="excess_stops_index",  
              palette="YlOrRd",  
              textNA="No Data",  
              n=5,  
              breaks=c(-10,-5, 0, 2, 4, 5),  
              midpoint=F)+  
  tm_layout(frame=FALSE,  
           legend.outside=TRUE)  
traffic_stop_bias_map
```

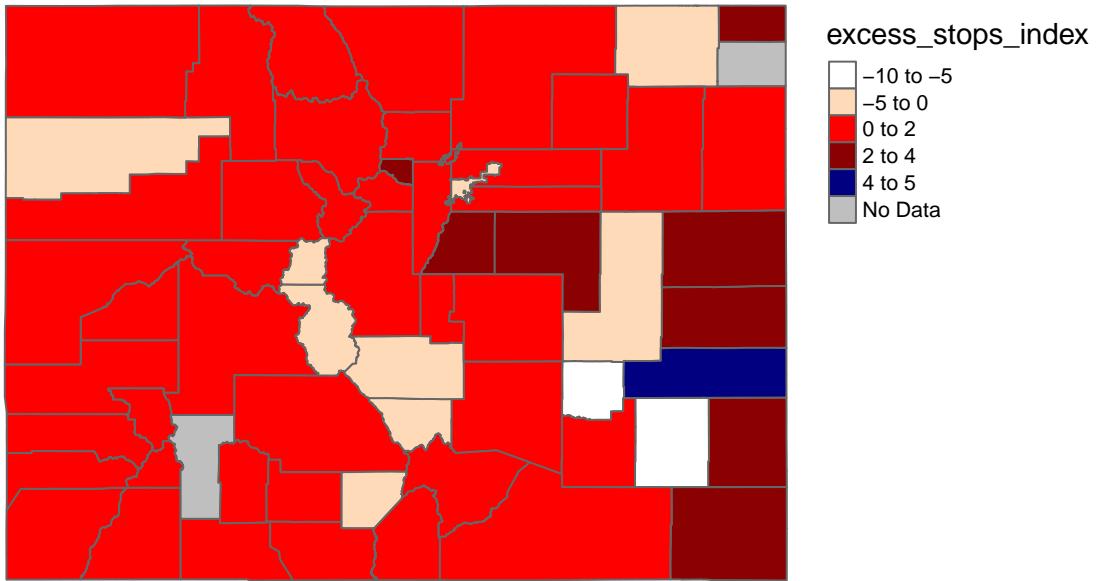


6.3.3 Make a map with custom colors

```
my_colors<-c("white", "peachpuff", "red1", "red4", "navy")
traffic_stop_bias_map<-tm_shape(county_shapefile_biasIndex)+  

  tm_polygons(col="excess_stops_index",
  palette=my_colors,
  textNA="No Data",
  n=5,
  breaks=c(-10,-5, 0, 2, 4, 5))+  

  tm_layout(frame=FALSE,
  legend.outside=TRUE)
traffic_stop_bias_map
```



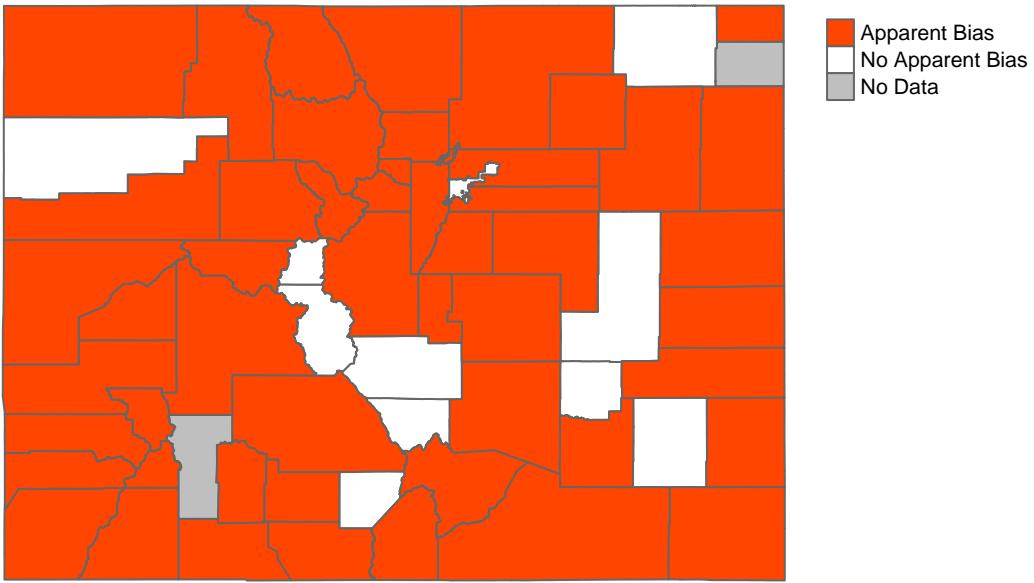
6.3.4 Make a categorical map

Define categorical column

```
county_shapefile_biasIndex<-
  county_shapefile_biasIndex %>%
    mutate(apparent_bias=ifelse(excess_stops_index>0, "Apparent Bias", "No Apparent Bias"))

categorical_map<-tm_shape(county_shapefile_biasIndex)+
  tm_polygons(col="apparent_bias", title="", pal=c("orangered1", "white"),
              textNA="No Data")+
  tm_layout(frame=FALSE,
            legend.outside=TRUE)

categorical_map
```



7 Refining and formatting the maps of the Colorado traffic-stop bias index

7.1 Refining the map of the continuous bias variable

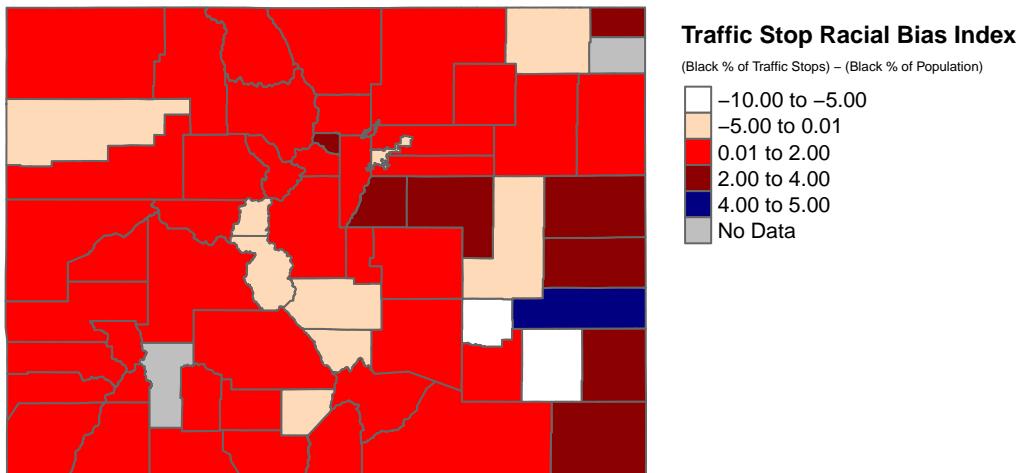
```

my_colors<-c("white", "peachpuff", "red1", "red4", "navy")
traffic_stop_bias_map<-tm_shape(county_shapefile_biasIndex)+ 
  tm_polygons(col="excess_stops_index",
              palette=my_colors,
              title="(Black % of Traffic Stops) - (Black % of Population)",
              textNA="No Data",
              n=5,
              breaks=c(-10,-5, 0.01, 2, 4, 5))+ 
  tm_layout(frame=FALSE,
            legend.outside=TRUE,
            legend.text.size=0.68,
            legend.title.size=0.75,
            title="Traffic Stop Racial Bias Index",
            title.size=0.75,
            title.fontface = 2,
            main.title="Racial Bias in Traffic Stops, by County (Colorado)",
            main.title.position=0.03,
            main.title.size=1,
            attr.outside=TRUE)+ 
  tm_credits("Map Author: NAME\nData Sources: 2010 Decennial Census via tidyverse,\nStanford Open Policy Group\nposition=c(0.02,0.01), # Specifies location of map credits
size=0.38)

traffic_stop_bias_map

```

Racial Bias in Traffic Stops, by County (Colorado)

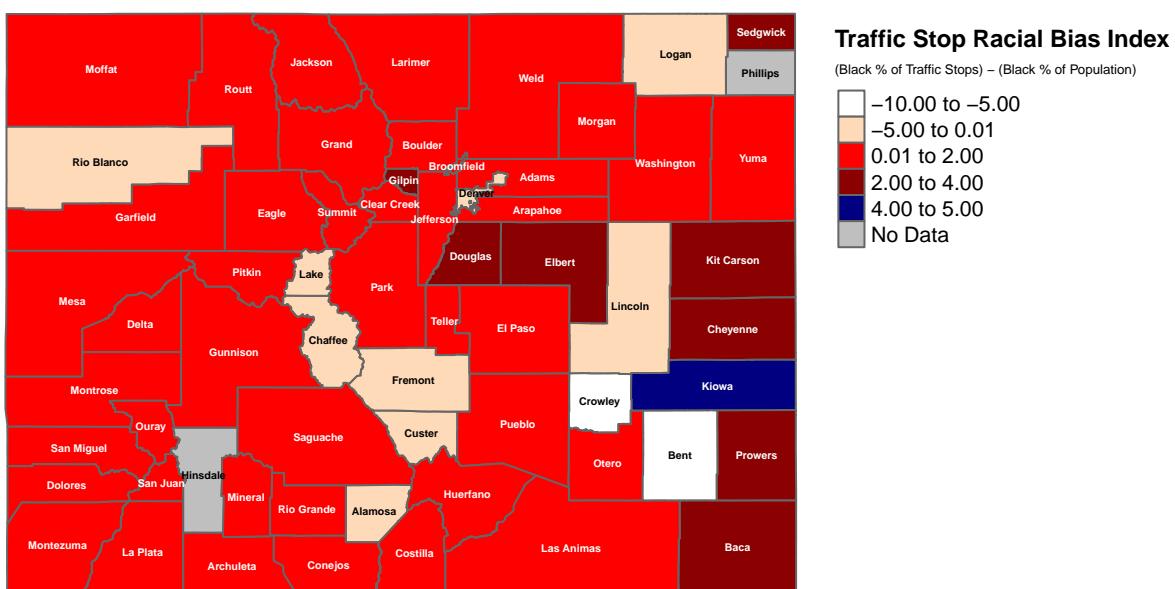


Map Author: NAME
Data Sources: 2010 Decennial Census via tidyverse,
Stanford Open Policing Project,
Colorado GeoLibrary

```
traffic_stop_bias_labeled<-traffic_stop_bias_map+
  tm_text("NAME", size=0.30, fontface=2)

traffic_stop_bias_labeled
```

Racial Bias in Traffic Stops, by County (Colorado)



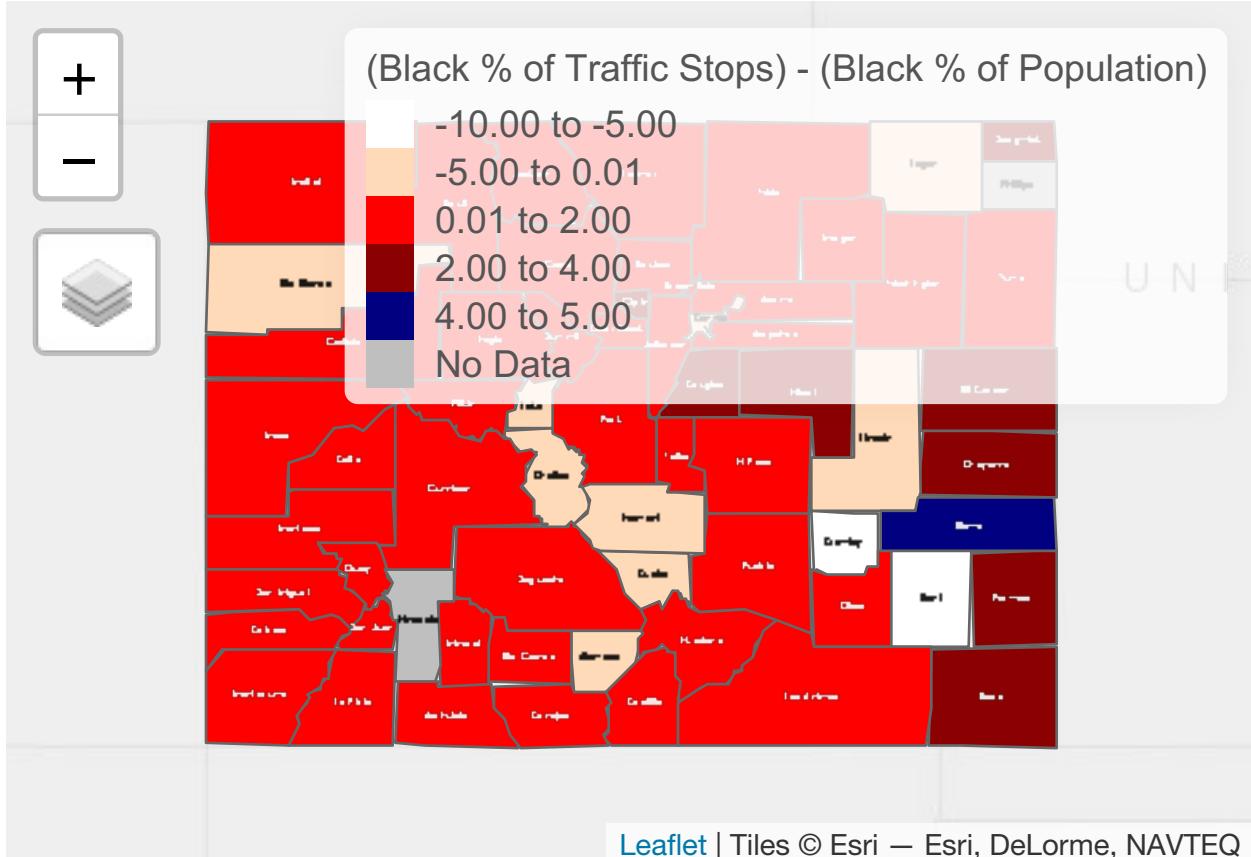
Map Author: NAME
Data Sources: 2010 Decennial Census via tidyverse,
Stanford Open Policing Project,
Colorado GeoLibrary

```
tmap_mode("view")

## tmap mode set to interactive viewing

traffic_stop_bias_labeled

## Credits not supported in view mode.
```



```
tmap_mode("plot")

## tmap mode set to plotting
```

7.2 Refining the map of the categorical bias variable

```
categorical_map<-tm_shape(county_shapefile_biasIndex)+  
  tm_polygons(col="apparent_bias", title="", pal=c("orangered1", "white"),  
  textNA="No Data")+  
  tm_layout(frame=FALSE,  
  legend.outside=TRUE,  
  main.title="Racial Bias in Traffic Stops, by County (Colorado)",  
  main.title.position=0.03,
```

```

    main.title.size=1,
    attr.outside=TRUE)+

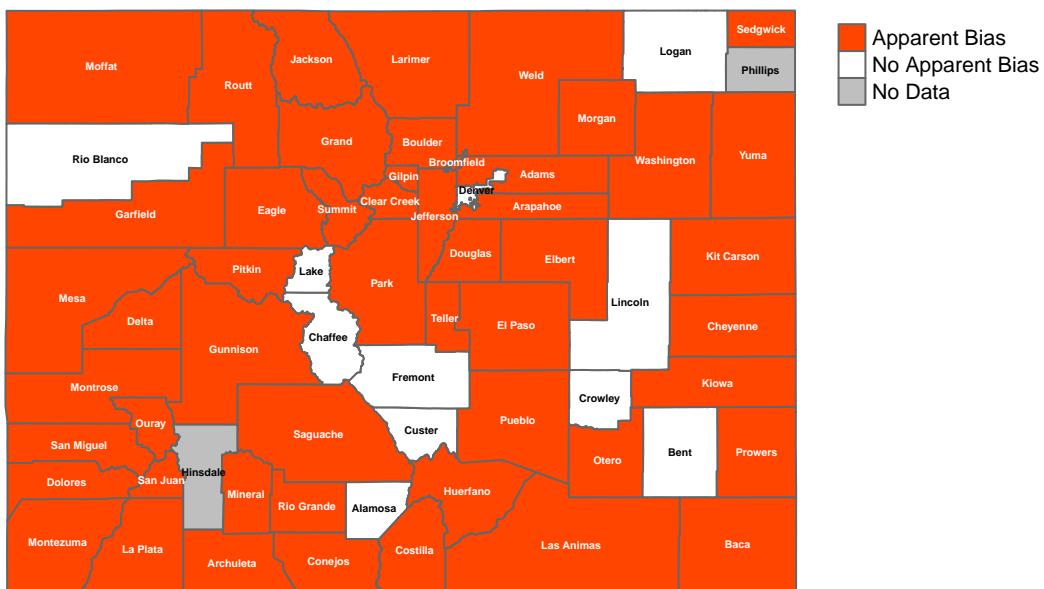
tm_credits("Map Author: NAME\nData Sources: 2010 Decennial Census via tidyverse,\nStanford Open Policing Project,\nColorado Geolibrary", position=c(0.02,0.01), # Specifies location of map credits
           size=0.38)+

tm_text("NAME", size=0.30, fontface=2)

categorical_map

```

Racial Bias in Traffic Stops, by County (Colorado)



Map Author: NAME
 Data Sources: 2010 Decennial Census via tidyverse,
 Stanford Open Policing Project,
 Colorado Geolibrary

8 Exporting maps

```

tmap_save(traffic_stop_bias_labeled, "traffic_stop_bias_labeled.png")

## Map saved to /Users/adra7980/Documents/git_repositories/intro_GIS/traffic_stop_bias_labeled.png

## Resolution: 2448.943 by 1800.777 pixels

## Size: 8.163142 by 6.002591 inches (300 dpi)

```

9 Summary script

```
# Read in Stanford police data for Colorado and assign to object named "co_traffic_stops"
co_traffic_stops<-read_csv("co_statewide_2020_04_01.csv")

# Create "Year" field based on existing "date" field
co_traffic_stops<-co_traffic_stops %>%
  mutate(Year=substr(co_traffic_stops$date, 1,4))# Filter 2010 observations and assign
# Filter 2010 observations and assign to a new object named "co_traffic_stops_2010"
co_traffic_stops_2010<-co_traffic_stops %>% filter(Year==2010)

# Compute county-level count of traffic stops by race and assign to object named "co_county_summary"
co_county_summary<-co_traffic_stops_2010 %>%
  group_by(county_name) %>%
  count(subject_race)

# Reshape the data so that the racial categories are transposed
# from rows into columns and assign the result to an object named
# "co_county_summary_wide"
co_county_summary_wide<-co_county_summary %>%
  pivot_wider(names_from=subject_race, values_from=n)

# Creates a new column named "total_stops" in "co_county_summary_wide" that
# contains information on the total number of stops for each county (across all racial categories)
co_county_summary_wide<-co_county_summary_wide %>%
  rowwise() %>%
  mutate(total_stops=sum(c_across(where(is.integer))), na.rm=TRUE))

# Selects "county_name", "black", and "total_stops" variables from "co_county_summary_wide";
# then renames the "black" variable to "black_stops" for clarity; then removes counties that
# are named "NA" due to an error in the dataset
co_county_black_stops<-co_county_summary_wide %>%
  select(county_name, black, total_stops) %>%
  rename(black_stops=black) %>%
  filter(county_name!="NA")

# Read in the pre-prepared demographic data from the 2010 decennial census and assign
# to an object named "co_counties_census_2010"
co_counties_census_2010<-read_csv("co_county_decennial_census.csv")

# Join "co_counties_census_2010" to "co_county_black_stops" and assign the result
# to an object named "co_counties_census_trafficstops"
co_counties_census_trafficstops<-full_join(co_county_black_stops, co_counties_census_2010,
                                             by=c("county_name"="County"))

# Use the information in "co_counties_census_trafficstops" to define new variables that will be used
# to compute the racial bias index: "black_stop_pct" (the black percentage of overall traffic stops within
# a county) and "black_pop_pct" (the black percentage of the county's over-17 population)

co_counties_census_trafficstops<-
  co_counties_census_trafficstops %>%
  mutate(black_stop_pct=((black_stops/total_stops)*100),
```

```

black_pop_pct=((total_black_pop_over17/total_pop_over17)*100))

# Calculate the bias index and include it as a new variable in "co_counties_census_trafficstops"
co_counties_census_trafficstops<-co_counties_census_trafficstops %>%
  mutate(excess_stops_index=black_stop_pct-black_pop_pct)

# Reads in Colorado county shapefile and assigns the shapefile to a new object named
# "co_counties_shapefile"
co_counties_shapefile<-st_read("tl_2019_08_county.shp")

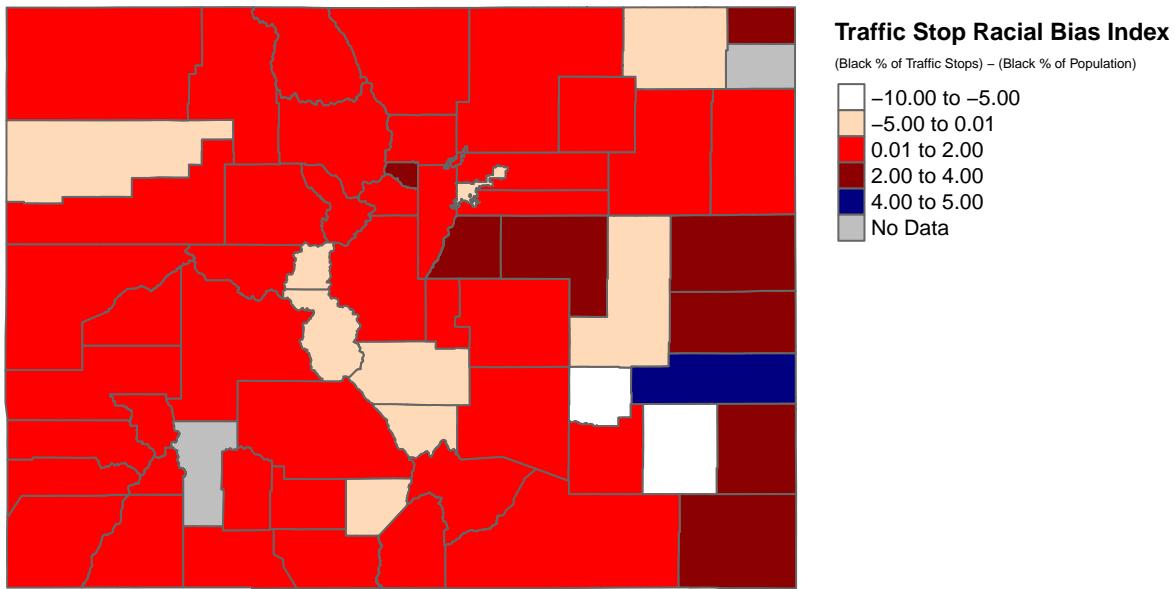
# Join "co_counties_census_trafficstops" to "co_counties_shapefile" using "GEOID" as the join field;
# assign the result to a new object named "county_shapefile_biasIndex"
county_shapefile_biasIndex<-full_join(co_counties_shapefile, co_counties_census_trafficstops,
                                         by="GEOID")

# make a map of the continuous "excess_stops_index"
my_colors<-c("white", "peachpuff", "red1", "red4", "navy") # create color vector
traffic_stop_bias_map<- # object assignment; assigns map to object named "traffic_stop_bias_map"
  tm_shape(county_shapefile_biasIndex)+ # declares the spatial object that is the basis for the map
  tm_polygons(col="excess_stops_index", # declares variable containing data to be mapped
              palette=my_colors, # sets color scheme (based on "my_colors" vector)
              title="(Black % of Traffic Stops) - (Black % of Population)", # sets legend subtitle
              textNA="No Data", # Sets the name of the legend label for "NA" values
              n=5, # defines the number of intervals in the legend
              breaks=c(-10,-5, 0.01, 2, 4, 5))+ # sets custom legend breaks
  tm_layout(frame=FALSE, # removes bounding box
            legend.outside=TRUE, # sets legend outside (invisible) bounding box
            legend.text.size=0.68, # sets size of legend text elements
            legend.title.size=0.75, # sets size of legend main title
            title="Traffic Stop Racial Bias Index", # sets legend's main title
            title.size=0.75, # sets relative size of legend's subtitle
            title.fontface = 2, # Makes legend title bold
            main.title="Racial Bias in Traffic Stops, by County (Colorado)", # specifies main title of map
            main.title.position=0.03, # specifies position of main title
            main.title.size=1, # specifies size of main title
            attr.outside=TRUE)+ # specifies that map credits should be placed outside bounding box
  tm_credits("Map Author: NAME\nData Sources: 2010 Decennial Census via tidyverse,\nStanford Open Policy Group")
  position=c(0.02,0.01), # Specifies location of map credits
  size=0.38) # sets title of credits

# Prints map
traffic_stop_bias_map

```

Racial Bias in Traffic Stops, by County (Colorado)



lower bound on legend is inclusive, upper bound exclusive

<https://geo.colorado.edu/catalog/47540-5e712aeda3d91e0009f59fc7>

A Using tidycensus to extract relevant census data

This section provides a script used to extract the census dataset that was read into R Studio in Section 5.1. To save time during a workshop, it is recommended to prepare the census dataset required to create the relevant index beforehand, and simply provide students with the relevant dataset. However, if you are looking for a way to extract the census dataset within R, the following script can be used as a guide.

Load packages

```
library(tidycensus)
library(tidyverse)
```

Enter your census API key with the following code:

```
census_api_key("INSERT HERE")
```

A.1 Step 1: Define your variables

```

# Variable list for 2010 Decennial
decennial_2010_variables<-load_variables(2010, "sf1")

# Define and name variables for census API call

my_vars<-c(total_pop="P001001",
           totalpop_men_u5="P012003",
           totalpop_men_5to9="P012004",
           totalpop_men_10to14="P012005",
           totalpop_men_15to17="P012006",
           totalpop_women_u5="P012027",
           totalpop_women_5to9="P012028",
           totalpop_women_10to14="P012029",
           totalpop_women_15to17="P012030",
           black_totalpop="PCT012B001",
           black_men_u1="PCT012B003",
           black_men_1="PCT012B004",
           black_men_2="PCT012B005",
           black_men_3="PCT012B006",
           black_men_4="PCT012B007",
           black_men_5="PCT012B008",
           black_men_6="PCT012B009",
           black_men_7="PCT012B010",
           black_men_8="PCT012B011",
           black_men_9="PCT012B012",
           black_men_10="PCT012B013",
           black_men_11="PCT012B014",
           black_men_12="PCT012B015",
           black_men_13="PCT012B016",
           black_men_14="PCT012B017",
           black_men_15="PCT012B018",
           black_men_16="PCT012B019",
           black_men_17="PCT012B020",
           black_women_u1="PCT012B107",
           black_women_1="PCT012B108",
           black_women_2="PCT012B109",
           black_women_3="PCT012B110",
           black_women_4="PCT012B111",
           black_women_5="PCT012B112",
           black_women_6="PCT012B113",
           black_women_7="PCT012B114",
           black_women_8="PCT012B115",
           black_women_9="PCT012B116",
           black_women_10="PCT012B117",
           black_women_11="PCT012B118",
           black_women_12="PCT012B119",
           black_women_13="PCT012B120",
           black_women_14="PCT012B121",
           black_women_15="PCT012B122",
           black_women_16="PCT012B123",
           black_women_17="PCT012B124")

```

A.2 Step 2: Extract the variables using tidycensus

```
# Issue call to Census API
co_counties_race<-get_decennial(
  geography="county",
  variables=my_vars,
  state="CO",
  survey="sf1",
  output="wide",
  year=2010,
  geometry=FALSE)

## Getting data from the 2010 decennial Census

## Using FIPS code '08' for state 'CO'

## Using Census Summary File 1
```

A.3 Step 3: Clean the tidycensus dataset

```
# Remove state name from name field
co_counties_race<-co_counties_race %>% separate(col=NAME, c("County", "x"), sep=",") %>%
  select(-x)
```

A.4 Step 4: Define new variables

```
# Create variable for total over-17 population
co_counties_race<-co_counties_race %>%
  mutate(total_pop_over17=total_pop-totalpop_men_u5-totalpop_men_5to9-
    totalpop_men_10to14-totalpop_men_15to17-totalpop_women_u5-
    totalpop_women_5to9-totalpop_women_10to14-totalpop_women_15to17)

# Create variable for total over-17 black population
co_counties_race<-co_counties_race %>%
  mutate(total_black_pop_over17=black_totalpop-black_men_u1-black_men_1-
    black_men_2-black_men_3-black_men_4-black_men_5-black_men_6-
    black_men_7-black_men_8-black_men_9-black_men_10-black_men_11-
    black_men_12-black_men_13-black_men_14-black_men_15-black_men_16-
    black_men_17-black_women_u1-black_women_1-black_women_2-black_women_3-
    black_women_4-black_women_5-black_women_6-black_women_7-black_women_8-
    black_women_9-black_women_10-black_women_11-black_women_12-
    black_women_13-black_women_14-black_women_15-black_women_16-
    black_women_17)
```

A.5 Step 5: Finalize and export the dataset

```

#Select relevant variables
co_counties_census_2010<-co_counties_race %>%
  select(GEOID, County, total_pop, total_black_pop_over17, total_pop_over17)

co_counties_census_2010

## # A tibble: 64 x 5
##   GEOID County      total_pop total_black_pop_over17 total_pop_over17
##   <chr> <chr>       <dbl>             <dbl>            <dbl>
## 1 08023 Costilla County     3524                18        2788
## 2 08025 Crowley County    5823               556        5034
## 3 08027 Custer County     4255                37        3525
## 4 08029 Delta County      30952               139       24101
## 5 08031 Denver County     600158              45338       471392
## 6 08035 Douglas County    285465              2447       198453
## 7 08033 Dolores County    2064                 4        1602
## 8 08049 Grand County      14843               43        11825
## 9 08039 Elbert County     23086               122       17232
## 10 08041 El Paso County    622263              27280       459587
## # ... with 54 more rows

# Export the data
write_csv(co_counties_census_2010, "co_counties_census_2010.csv")

```