Introduction to Geographic Information Systems

2022-02-09

# 1 Introduction

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

# 2 Loading data

Let’s begin by reading in the Stanford Policing Project State Patrol data for the state of Colorado; the data can be downloaded from this [website](https://openpolicing.stanford.edu/data/), but has also been made available to you as part of the workshop materials as a CSV file. Once, the traffic patrol data has been downloaded into your working directory, pass the name of the file to the read\_csv function, and assign it to an object. Here, we’ll assign the traffic patrol data to an object named co\_traffic\_stops. Note that object names are arbitrary, but ideally should meaningfully describe their contents.

# 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.

Once the traffic patrol data has been read into R studio and assigned to the object, we can print the contents of the dataset to the console by typing the name of the object into the R Studio console (note that only the first few records will be printed)

# Print the contents of "co\_traffic\_stops" (i.e. the CO traffic patrol data) to the console; the first few records of the dataset will print  
co\_traffic\_stops

## # A tibble: 3,112,853 × 20  
## raw\_row\_number date time location county\_name subject\_age subject\_race subject\_sex officer\_id\_hash  
## <chr> <date> <lgl> <chr> <chr> <dbl> <chr> <chr> <chr>   
## 1 1947986|1947987 2013-06-19 NA 19, I70… Mesa County 26 hispanic male b942632983   
## 2 1537576 2012-08-24 NA 254, H2… Jefferson … NA <NA> <NA> f3d4f46927   
## 3 1581594 2012-09-23 NA 115, I7… Logan Coun… 52 white male 6e49e2fbc8   
## 4 1009205 2011-08-25 NA 197, H8… Douglas Co… 32 white female eaea851669   
## 5 1932619 2013-06-08 NA 107, H2… Kiowa Coun… 33 hispanic male d18e34d749   
## 6 1179436 2011-12-23 NA 48, 384… Boulder Co… NA <NA> <NA> b84c696aed   
## 7 1326795 2012-04-07 NA 0, R250… Boulder Co… 39 white male 4c0279748e   
## 8 1786795 2013-03-03 NA 19, E47… Arapahoe C… 44 white female e6b5b9bb98   
## 9 1552164 2012-09-02 NA 224, H2… Park County NA <NA> <NA> 43f1f150d3   
## 10 1004281|1004282|10… 2011-08-21 NA R2000, … Adams Coun… 32 hispanic male dd2f10b6f8   
## # … 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 <lgl>,  
## # search\_conducted <lgl>, search\_basis <chr>, raw\_Ethnicity <chr>

We can also view co\_traffic\_stops object (or, for that matter, any dataset in R Studio) within the R Studio data viewer by passing the name of the object to the View function:

# Inspect ```co\_traffic\_stops``` in the R Studio data viewer  
View(co\_traffic\_stops)

# 3 Cleaning and filtering data

To make things tractable,

## 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 × 21  
## raw\_row\_number date time location county\_name subject\_age subject\_race subject\_sex officer\_id\_hash  
## <chr> <date> <lgl> <chr> <chr> <dbl> <chr> <chr> <chr>   
## 1 188721|188722 2010-04-17 NA 2, 989,… Montezuma … 16 white female b737c978e4   
## 2 187958 2010-04-17 NA 991, 32 Montezuma … 54 white male 89066a8ce8   
## 3 188451 2010-04-17 NA 9, 280,… Montezuma … 49 hispanic male 7346f3aeb5   
## 4 186989|186990|1869… 2010-04-17 NA 3, 277,… Montezuma … 16 white male 3c174e7733   
## 5 186997|186998|1869… 2010-04-17 NA 3, 277,… Montezuma … 37 white male 3c174e7733   
## 6 186993|186994|1869… 2010-04-17 NA 3, 277,… Montezuma … 39 white male 3c174e7733   
## 7 600865 2010-12-21 NA 164.5, … Mineral Co… 110 <NA> <NA> 0e48f16a6a   
## 8 600477 2010-12-21 NA 163, 29… Mineral Co… 110 <NA> <NA> c85c943fb9   
## 9 36625|36626 2010-01-20 NA 312, H5… Pueblo Cou… 45 hispanic male 5054b33e13   
## 10 275 2010-01-01 NA 127, H2… Chaffee Co… 17 white female 750fd84f12   
## # … 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 <lgl>,  
## # 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 × 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 × 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 × 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 × 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 × 9  
## # Rowwise: county\_name  
## county\_name black\_stops total\_stops GEOID total\_pop total\_black\_pop\_ove… total\_pop\_over17 black\_stop\_pct  
## <chr> <int> <int> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Adams County 1208 34350 08001 441603 9396 315480 3.52   
## 2 Alamosa County 43 4478 08003 15445 142 11617 0.960  
## 3 Arapahoe County 1819 17520 08005 572003 40558 424679 10.4   
## 4 Archuleta County 28 5091 08007 12084 19 9676 0.550  
## 5 Baca County 61 1511 08009 3788 15 2974 4.04   
## 6 Bent County 46 1808 08011 6499 486 5403 2.54   
## 7 Boulder County 192 13053 08013 294567 1961 231813 1.47   
## 8 Broomfield County 22 1095 08014 55889 415 41237 2.01   
## 9 Chaffee County 37 6521 08015 17809 264 14821 0.567  
## 10 Cheyenne County 38 1106 08017 1836 4 1386 3.44   
## # … 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 × 10  
## # Rowwise: county\_name  
## county\_name black\_stops total\_stops GEOID total\_pop total\_black\_pop\_ove… total\_pop\_over17 black\_stop\_pct  
## <chr> <int> <int> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Adams County 1208 34350 08001 441603 9396 315480 3.52   
## 2 Alamosa County 43 4478 08003 15445 142 11617 0.960  
## 3 Arapahoe County 1819 17520 08005 572003 40558 424679 10.4   
## 4 Archuleta County 28 5091 08007 12084 19 9676 0.550  
## 5 Baca County 61 1511 08009 3788 15 2974 4.04   
## 6 Bent County 46 1808 08011 6499 486 5403 2.54   
## 7 Boulder County 192 13053 08013 294567 1961 231813 1.47   
## 8 Broomfield County 22 1095 08014 55889 415 41237 2.01   
## 9 Chaffee County 37 6521 08015 17809 264 14821 0.567  
## 10 Cheyenne County 38 1106 08017 1836 4 1386 3.44   
## # … 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)

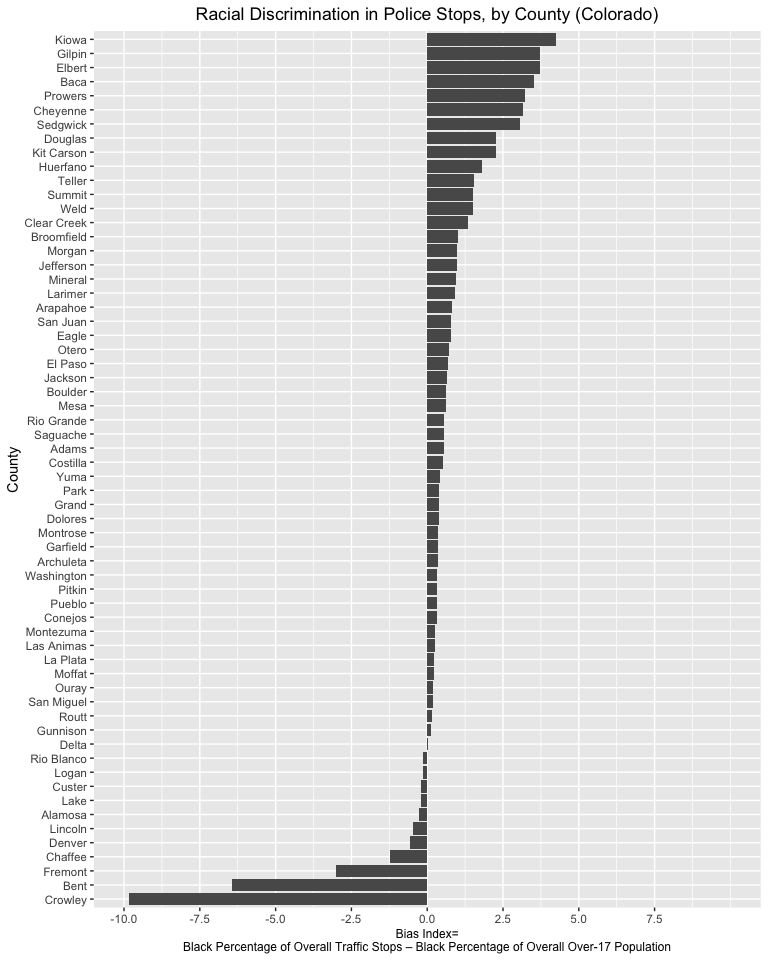
## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 62 0.53 2.04 0.47 0.65 0.59 -9.84 4.26 14.1 -2.45 10.67 0.26

## 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=\nBlack Percentage of Overall Traffic Stops – Black Percentage of Overall Over-17 Population")+  
 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



# 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")

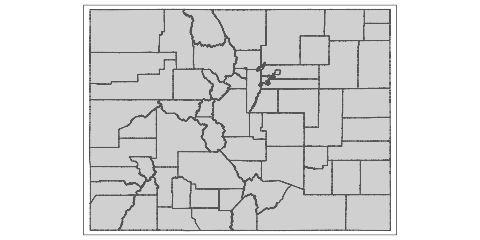
## Reading layer `tl\_2019\_08\_county' from data source `/Users/adra7980/Documents/CU\_workshops/gis/data/tl\_2019\_08\_county/tl\_2019\_08\_county.shp' using driver `ESRI Shapefile'  
## Simple feature collection with 64 features and 17 fields  
## geometry type: MULTIPOLYGON  
## dimension: XY  
## bbox: xmin: -109.0602 ymin: 36.99245 xmax: -102.0415 ymax: 41.00344  
## geographic CRS: NAD83

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 FUNCSTAT  
## 1 08 109 00198170 08109 Saguache Saguache County 06 H1 G4020 <NA> <NA> <NA> A  
## 2 08 115 00198173 08115 Sedgwick Sedgwick County 06 H1 G4020 <NA> <NA> <NA> A  
## 3 08 017 00198124 08017 Cheyenne Cheyenne County 06 H1 G4020 <NA> <NA> <NA> A  
## 4 08 027 00198129 08027 Custer Custer County 06 H1 G4020 <NA> <NA> <NA> A  
## 5 08 067 00198148 08067 La Plata La Plata County 06 H1 G4020 <NA> 20420 <NA> A  
## 6 08 111 00198171 08111 San Juan San Juan County 06 H1 G4020 <NA> <NA> <NA> A  
## 7 08 097 00198164 08097 Pitkin Pitkin County 06 H1 G4020 233 24060 <NA> A  
## 8 08 093 00198162 08093 Park Park County 06 H1 G4020 216 19740 <NA> A  
## 9 08 003 00198117 08003 Alamosa Alamosa County 06 H1 G4020 <NA> <NA> <NA> A  
## 10 08 099 00198165 08099 Prowers Prowers County 06 H1 G4020 <NA> <NA> <NA> A  
## ALAND AWATER INTPTLAT INTPTLON geometry  
## 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...

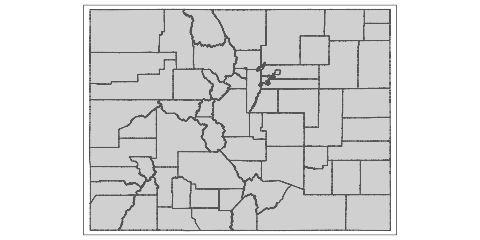
## tmap mode set to plotting

tm\_shape(co\_counties\_shapefile)+  
 tm\_polygons()



co\_counties\_map<-tm\_shape(co\_counties\_shapefile)+  
 tm\_polygons()

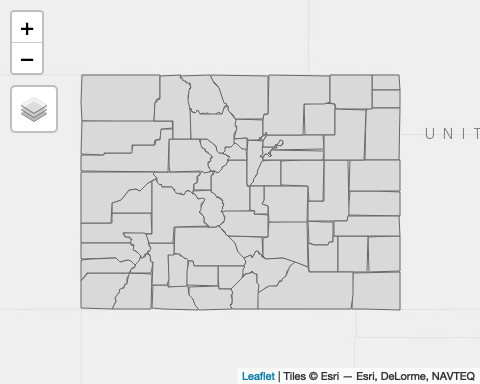
co\_counties\_map



tmap\_mode("view")

## tmap mode set to interactive viewing

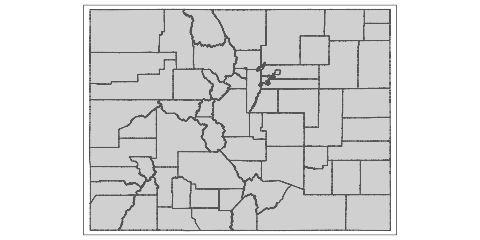
co\_counties\_map



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")

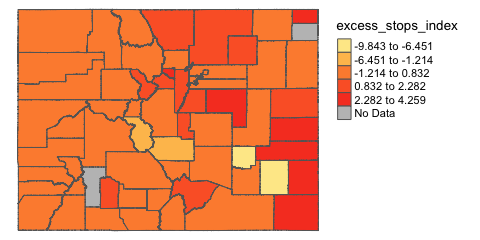
county\_shapefile\_biasIndex

## 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 FUNCSTAT  
## 1 08 109 00198170 08109 Saguache Saguache County 06 H1 G4020 <NA> <NA> <NA> A  
## 2 08 115 00198173 08115 Sedgwick Sedgwick County 06 H1 G4020 <NA> <NA> <NA> A  
## 3 08 017 00198124 08017 Cheyenne Cheyenne County 06 H1 G4020 <NA> <NA> <NA> A  
## 4 08 027 00198129 08027 Custer Custer County 06 H1 G4020 <NA> <NA> <NA> A  
## 5 08 067 00198148 08067 La Plata La Plata County 06 H1 G4020 <NA> 20420 <NA> A  
## 6 08 111 00198171 08111 San Juan San Juan County 06 H1 G4020 <NA> <NA> <NA> A  
## 7 08 097 00198164 08097 Pitkin Pitkin County 06 H1 G4020 233 24060 <NA> A  
## 8 08 093 00198162 08093 Park Park County 06 H1 G4020 216 19740 <NA> A  
## 9 08 003 00198117 08003 Alamosa Alamosa County 06 H1 G4020 <NA> <NA> <NA> A  
## 10 08 099 00198165 08099 Prowers Prowers County 06 H1 G4020 <NA> <NA> <NA> A  
## 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

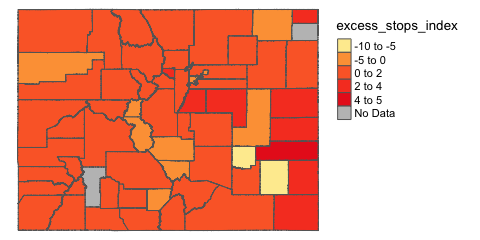
### 6.3.1 Make a rough draft of a map

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



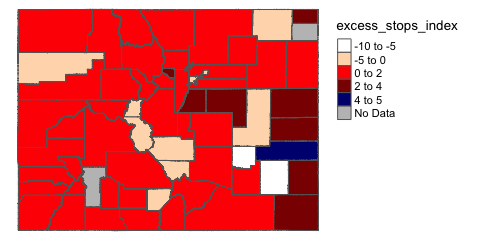
### 6.3.2 Make a map with custom breaks

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

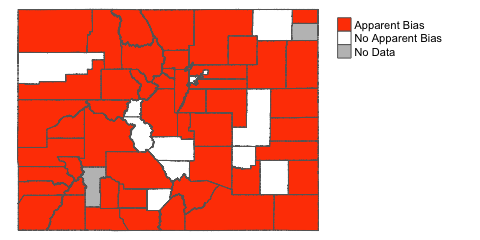


### 6.3.4 Make a categorical map

Define categorical column

county\_shapefile\_biasIndex<-  
 county\_shapefile\_biasIndex %>%   
 mutate(apparent\_bias=ifelse(excess\_stops\_index>0, "Apparent Bias", "No Apparent Bias"))

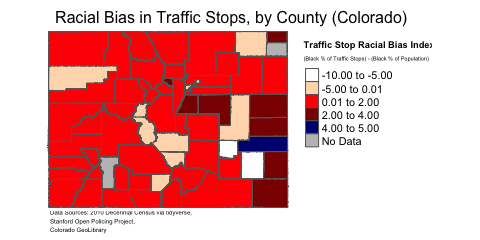
categorical\_map<-tm\_shape(county\_shapefile\_biasIndex)+  
 tm\_polygons(col="apparent\_bias", title="", pal=c("orangered1", "white"),   
 textNA="No Data")+  
 tm\_layout(frame=FALSE,   
 legend.outside=TRUE)  
   
categorical\_map



# 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 Policing Project,\nColorado GeoLibrary ", # Sets text for map credits  
 position=c(0.02,0.01), # Specifies location of map credits  
 size=0.38)  
  
traffic\_stop\_bias\_map



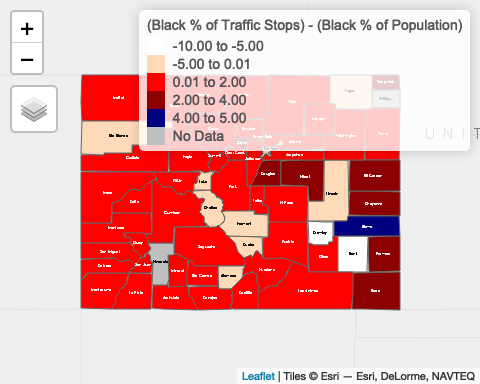
traffic\_stop\_bias\_labeled<-traffic\_stop\_bias\_map+  
 tm\_text("NAME", size=0.30, fontface=2)  
  
traffic\_stop\_bias\_labeled

tmap\_mode("view")

## tmap mode set to interactive viewing

traffic\_stop\_bias\_labeled

## Credits not supported in view mode.

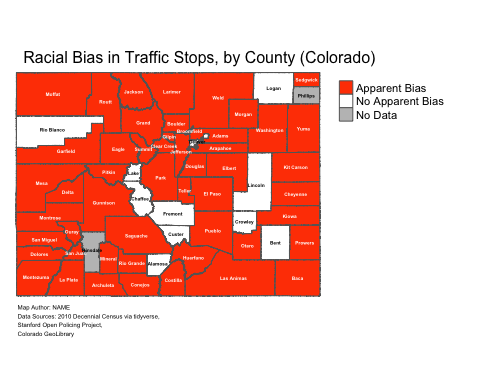


tmap\_mode("plot")

## 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 ", # Sets text for map credits  
 position=c(0.02,0.01), # Specifies location of map credits  
 size=0.38)+  
 tm\_text("NAME", size=0.30, fontface=2)  
   
   
categorical\_map



# 8 Exporting maps

tmap\_save(traffic\_stop\_bias\_labeled, "traffic\_stop\_bias\_labeled.png")

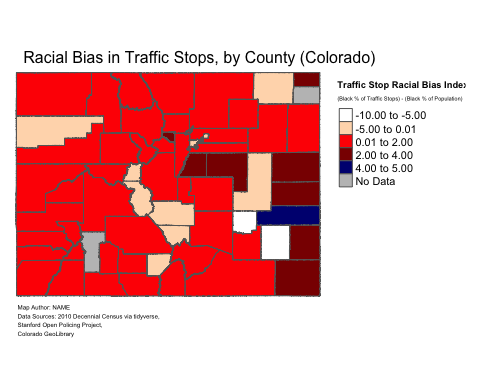
## Map saved to /Users/adra7980/Documents/git\_repositories/intro\_GIS/traffic\_stop\_bias\_labeled.png

## 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 to a #   
# 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 Policing Project,\nColorado GeoLibrary ", # Sets text for map credits  
 position=c(0.02,0.01), # Specifies location of map credits  
 size=0.38) # sets title of credits  
  
# Prints map  
traffic\_stop\_bias\_map



lower bound on legend is inclusive, upper bound exclusive

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

# (APPENDIX) Appendix

# 10 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")

## 10.1 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")

## 10.2 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

## 10.3 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)

## 10.4 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)

## 10.5 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 × 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")