Homework 5

Alison Barbee

2022-11-30

library(readr)  
  
homicide\_data <- read\_csv("https://raw.githubusercontent.com/washingtonpost/data-homicides/master/homicide-data.csv")

## Rows: 52179 Columns: 12  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): uid, victim\_last, victim\_first, victim\_race, victim\_age, victim\_sex...  
## dbl (3): reported\_date, lat, lon  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#Picked Los Angeles as my primary city

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ dplyr 1.0.10  
## ✔ tibble 3.1.8 ✔ stringr 1.4.1   
## ✔ tidyr 1.2.1 ✔ forcats 0.5.2   
## ✔ purrr 0.3.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(forcats)  
  
losAngeles <- homicide\_data %>%  
 mutate(city\_name = str\_c(homicide\_data$city, homicide\_data$state, sep = ", ")) %>%  
 filter(city\_name == "Los Angeles, CA") %>%  
 mutate(Status = case\_when(  
 grepl("Closed by arrest", disposition) ~ "solved",   
 grepl("Closed without arrest", disposition) ~ "unsolved",  
 grepl("Open/No arrest", disposition) ~ "unsolved",  
 TRUE ~ "NA")) %>%  
 mutate(homicide\_race = fct\_lump\_min(victim\_race, min = 100))

Use different colors to show the three race groups with the highest number of homicides for that city (you may find the fct\_lump function from forcats useful for this).

sum(losAngeles$victim\_race == "Hispanic")

## [1] 1088

sum(losAngeles$victim\_race == "Black")

## [1] 886

sum(losAngeles$victim\_race == "White")

## [1] 192

sum(losAngeles$victim\_race == "Other")

## [1] 59

sum(losAngeles$victim\_race == "Asian")

## [1] 29

sum(losAngeles$victim\_race == "Unknown")

## [1] 3

#Mapping it out

library(sf)

## Warning: package 'sf' was built under R version 4.2.2

## Linking to GEOS 3.9.3, GDAL 3.5.2, PROJ 8.2.1; sf\_use\_s2() is TRUE

library(tigris)

## Warning: package 'tigris' was built under R version 4.2.2

## To enable caching of data, set `options(tigris\_use\_cache = TRUE)`  
## in your R script or .Rprofile.

library(ggplot2)  
library(viridis)

## Warning: package 'viridis' was built under R version 4.2.2

## Loading required package: viridisLite

library(MAP)

## Warning: package 'MAP' was built under R version 4.2.2

## Loading required package: flexmix

## Warning: package 'flexmix' was built under R version 4.2.2

## Loading required package: lattice

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

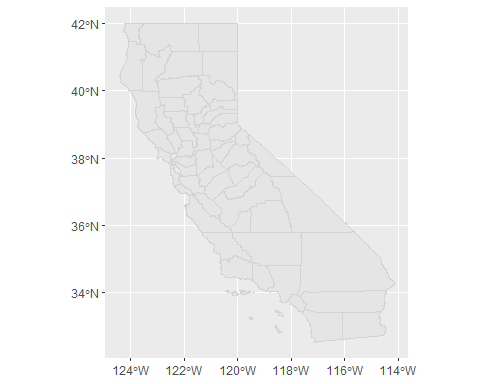
## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

ca\_counties <- counties(state = "CA", cb = TRUE, class = "sf")

## Retrieving data for the year 2020

## | | | 0% | | | 1% | |= | 1% | |= | 2% | |== | 2% | |== | 3% | |=== | 4% | |=== | 5% | |==== | 5% | |==== | 6% | |===== | 7% | |===== | 8% | |====== | 8% | |====== | 9% | |======= | 9% | |======= | 10% | |======= | 11% | |======== | 11% | |======== | 12% | |========= | 12% | |========= | 13% | |========= | 14% | |========== | 14% | |========== | 15% | |=========== | 15% | |=========== | 16% | |============ | 16% | |============ | 17% | |============ | 18% | |============= | 18% | |============= | 19% | |============== | 19% | |============== | 20% | |============== | 21% | |=============== | 21% | |=============== | 22% | |================ | 22% | |================ | 23% | |================= | 24% | |================= | 25% | |================== | 25% | |================== | 26% | |=================== | 27% | |=================== | 28% | |==================== | 28% | |==================== | 29% | |===================== | 29% | |===================== | 30% | |===================== | 31% | |====================== | 31% | |====================== | 32% | |======================= | 32% | |======================= | 33% | |======================= | 34% | |======================== | 34% | |======================== | 35% | |========================= | 35% | |========================= | 36% | |========================== | 37% | |========================== | 38% | |=========================== | 38% | |=========================== | 39% | |============================ | 39% | |============================ | 40% | |============================ | 41% | |============================= | 41% | |============================= | 42% | |============================== | 42% | |============================== | 43% | |=============================== | 44% | |=============================== | 45% | |================================ | 45% | |================================ | 46% | |================================= | 46% | |================================= | 47% | |================================= | 48% | |================================== | 48% | |================================== | 49% | |=================================== | 49% | |=================================== | 50% | |=================================== | 51% | |==================================== | 51% | |==================================== | 52% | |===================================== | 52% | |===================================== | 53% | |===================================== | 54% | |====================================== | 54% | |====================================== | 55% | |======================================= | 55% | |======================================= | 56% | |======================================== | 57% | |======================================== | 58% | |========================================= | 58% | |========================================= | 59% | |========================================== | 59% | |========================================== | 60% | |========================================== | 61% | |=========================================== | 61% | |=========================================== | 62% | |============================================ | 62% | |============================================ | 63% | |============================================= | 64% | |============================================= | 65% | |============================================== | 65% | |============================================== | 66% | |=============================================== | 67% | |=============================================== | 68% | |================================================ | 68% | |================================================ | 69% | |================================================= | 69% | |================================================= | 70% | |================================================== | 71% | |================================================== | 72% | |=================================================== | 72% | |=================================================== | 73% | |==================================================== | 74% | |==================================================== | 75% | |===================================================== | 75% | |===================================================== | 76% | |====================================================== | 77% | |====================================================== | 78% | |======================================================= | 78% | |======================================================= | 79% | |======================================================== | 79% | |======================================================== | 80% | |======================================================== | 81% | |========================================================= | 81% | |========================================================= | 82% | |========================================================== | 82% | |========================================================== | 83% | |========================================================== | 84% | |=========================================================== | 84% | |=========================================================== | 85% | |============================================================ | 85% | |============================================================ | 86% | |============================================================= | 86% | |============================================================= | 87% | |============================================================= | 88% | |============================================================== | 88% | |============================================================== | 89% | |=============================================================== | 89% | |=============================================================== | 90% | |=============================================================== | 91% | |================================================================ | 91% | |================================================================ | 92% | |================================================================= | 92% | |================================================================= | 93% | |================================================================== | 94% | |================================================================== | 95% | |=================================================================== | 95% | |=================================================================== | 96% | |==================================================================== | 97% | |==================================================================== | 98% | |===================================================================== | 98% | |===================================================================== | 99% | |======================================================================| 100%

ggplot() +   
 geom\_sf(data = ca\_counties, color = "lightgray")



la\_homicides <- losAngeles %>%   
 select(uid, reported\_date, homicide\_race, lat, lon, Status)  
  
la\_homicides <- st\_as\_sf(la\_homicides, coords = c("lon", "lat")) %>%  
 st\_set\_crs(4269)

ggplot() +  
 geom\_sf(data = ca\_counties, color = "lightgray", fill = "gray") +  
 xlim(c(-118.8, -118)) + ylim(c(33.7, 34.35)) +  
 geom\_sf(data = la\_homicides, aes(color = homicide\_race,  
 fill = homicide\_race,  
 shape = Status)) +  
 scale\_fill\_discrete(name = "Race") +  
 scale\_color\_discrete(name = "Race") +  
 scale\_shape\_manual(values = c(18, 8)) +  
 ggtitle("Location of homicides in Los Angeles, CA")

