HW5

Amy Gill

2025-11-30

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```r
packages <- c("tidyverse", "sf", "tigris", "forcats", "lubridate", "knitr")
install.packages(setdiff(packages, rownames(installed.packages())))

library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.5.2
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(sf)
```

```
## Warning: package 'sf' was built under R version 4.5.2
```

```
## Linking to GEOS 3.13.1, GDAL 3.11.4, PROJ 9.7.0; sf_use_s2() is TRUE
```

```r
library(tigris)
```

```
## Warning: package 'tigris' was built under R version 4.5.2
```

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

```
library(forcats)
library(lubridate)

options(tigris_use_cache = TRUE)
```

```
## 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.
```

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

#Choosing city

```
city_name <- "Baltimore"
state_abbrev <- "MD"

city_data <- homicides %>%
  filter(city == city_name, state == state_abbrev)
```

#Convert homicide points to sf

```
city_sf <- city_data %>%
  st_as_sf(coords = c("lon", "lat"), crs = 4326, remove = FALSE)
```

#Download Census boundaries with tigris and use tracts.

```
tracts <- tracts(
  state = state_abbrev,
  county = "510",
  year = 2020) %>%
  st_transform(4326)
```

#5

```
city_sf <- city_sf %>%
  mutate(
    race_clean = fct_lump(f = as.factor(victim_race), n = 3))
```
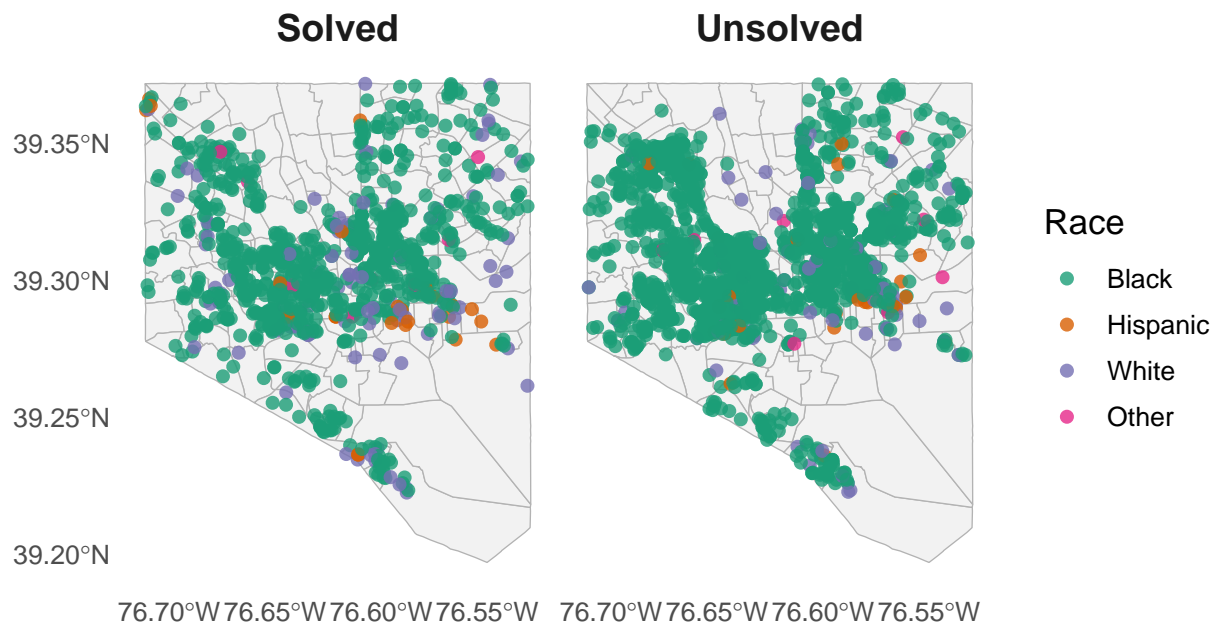
#6

```
city_sf <- city_sf %>%
  mutate(
    solved = case_when(
      disposition %in% c("Closed without arrest", "Open/No arrest") ~ "Unsolved",
      TRUE ~ "Solved"
    ))
```

#7 Plot ~Homicides with county

```r
ggplot() +
  geom_sf(data = tracts, fill = "gray95", color = "gray70") +
  geom_sf(data = city_sf, aes(color = race_clean), alpha = 0.8, size = 1.8) +
  facet_wrap(~solved) +
  scale_color_brewer(palette = "Dark2", name = "Race") +
  labs(
    title = paste("Homicides in", city_name, state_abbrev),
    subtitle = "Faceted by case status (Solved vs Unsolved)",
    caption = "Source: Washington Post homicide dataset"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    panel.grid = element_blank(),
    strip.text = element_text(size = 14, face = "bold")
  )
```

Homicides in Baltimore MD
Faceted by case status (Solved vs Unsolved)

Source: Washington Post homicide dataset