

HWK #5 – Homicides

Melanie Luu

1. Introduction

For Homework #5, I will be working with the Washington Post homicide dataset. This document will eventually contain all analysis, plots, and interpretation, but for the first commit it only includes the structure and setup.

For this assignment, I chose **Choice 2** from Homework #5: recreating the monthly homicide figure for Baltimore with a reference line for the arrest of Freddie Gray and colors to show cold versus warm months.

The data come from the Washington Post's `homicide-data.csv` file, which I downloaded from the Washington Post GitHub repository and saved in my local `data/` folder inside this R project.

Source: <https://github.com/washingtonpost/data-homicides>

2. Project Setup Summary

- Created an R Project for HWK #5
- Added required folders:
 - `data/`
 - `figures/`
 - `writing/`
- Created this RMarkdown document inside the `writing/` directory
- Initialized git for the project
- Ready to make the first commit

3. Load Libraries (placeholder)

4. Placeholder: Data Import

```
# Read data from the `data` subdirectory (relative to project root)
homicides_raw <- readr::read_csv("data/homicide-data.csv")

# Glimpse at the structure
dplyr::glimpse(homicides_raw)
```

```
## Rows: 52,179
## Columns: 12
## $ uid      <chr> "A1b-000001", "A1b-000002", "A1b-000003", "A1b-000004", ~
## $ reported_date <dbl> 20100504, 20100216, 20100601, 20100101, 20100102, 201001~
## $ victim_last  <chr> "GARCIA", "MONTAYA", "SATTERFIELD", "MENDIOLA", "MULA", ~
```

```
## $ victim_first <chr> "JUAN", "CAMERON", "VIVIANA", "CARLOS", "VIVIAN", "GERAL~
## $ victim_race <chr> "Hispanic", "Hispanic", "White", "Hispanic", "White", "W~
## $ victim_age <chr> "78", "17", "15", "32", "72", "91", "52", "52", "56", "4~
## $ victim_sex <chr> "Male", "Male", "Female", "Male", "Female", "Female", "M~
## $ city <chr> "Albuquerque", "Albuquerque", "Albuquerque", "Albuquerque~
## $ state <chr> "NM", "NM", "NM", "NM", "NM", "NM", "NM", "NM", "NM", "N~
## $ lat <dbl> 35.09579, 35.05681, 35.08609, 35.07849, 35.13036, 35.151~
## $ lon <dbl> -106.5386, -106.7153, -106.6956, -106.5561, -106.5810, --
## $ disposition <chr> "Closed without arrest", "Closed by arrest", "Closed wit~
```

5. Create monthly homicide counts for Baltimore

```
baltimore_monthly <-
  homicides_raw |>
  dplyr::filter(city == "Baltimore") |>
  # reported_date is an integer like 20150412; convert to Date
  dplyr::mutate(
    date = lubridate::ymd(as.character(reported_date)),
    month = lubridate::floor_date(date, unit = "month"),
    season = dplyr::if_else(
      lubridate::month(month) %in% c(11, 12, 1, 2, 3, 4),
      "Cold months (Nov-Apr)",
      "Warm months (May-Oct)"
    )
  ) |>
  dplyr::count(month, season, name = "homicides")

head(baltimore_monthly)
```

```
## # A tibble: 6 x 3
##   month      season      homicides
##   <date>    <chr>          <int>
## 1 2007-01-01 Cold months (Nov-Apr)      28
## 2 2007-02-01 Cold months (Nov-Apr)      17
## 3 2007-03-01 Cold months (Nov-Apr)      26
## 4 2007-04-01 Cold months (Nov-Apr)      19
## 5 2007-05-01 Warm months (May-Oct)      32
## 6 2007-06-01 Warm months (May-Oct)      31
```

6. Plot monthly homicides in Baltimore

```
freddie_arrest <- as.Date("2015-04-12")

ggplot2::ggplot(baltimore_monthly,
  ggplot2::aes(x = month, y = homicides, color = season)) +
  ggplot2::geom_line() +
  ggplot2::geom_point(size = 1.5) +
  # smooth line for seasonal / long-term trends
  ggplot2::geom_smooth(se = FALSE, size = 1, linetype = "solid") +
  # Vertical reference line: arrest of Freddie Gray on April 12, 2015
  ggplot2::geom_vline(xintercept = freddie_arrest, linetype = "dashed") +
  ggplot2::annotate(
    "text",
    x = freddie_arrest,
```

```

y = max(baltimore_monthly$homicides, na.rm = TRUE) + 2,
label = "Freddie Gray arrest\nApr 12, 2015",
hjust = -0.05,
vjust = 1,
size = 3
) +
ggplot2::scale_color_manual(
  values = c(
    "Cold months (Nov-Apr)" = "#1f78b4",
    "Warm months (May-Oct)" = "#e31a1c"
  )
) +
ggplot2::scale_x_date(
  date_breaks = "1 year",
  date_labels = "%Y"
) +
ggplot2::labs(
  title = "Monthly Homicides in Baltimore",
  subtitle = "Cold vs warm months, with reference line for the arrest of Freddie Gray",
  x = "Month",
  y = "Number of homicides",
  color = "Season"
) +
ggplot2::theme_minimal(base_size = 11) +
ggplot2::theme(
  legend.position = "top",
  panel.grid.minor = ggplot2::element_blank()
)

```

Monthly Homicides in Baltimore

Cold vs warm months, with reference line for the arrest of Freddie Gray

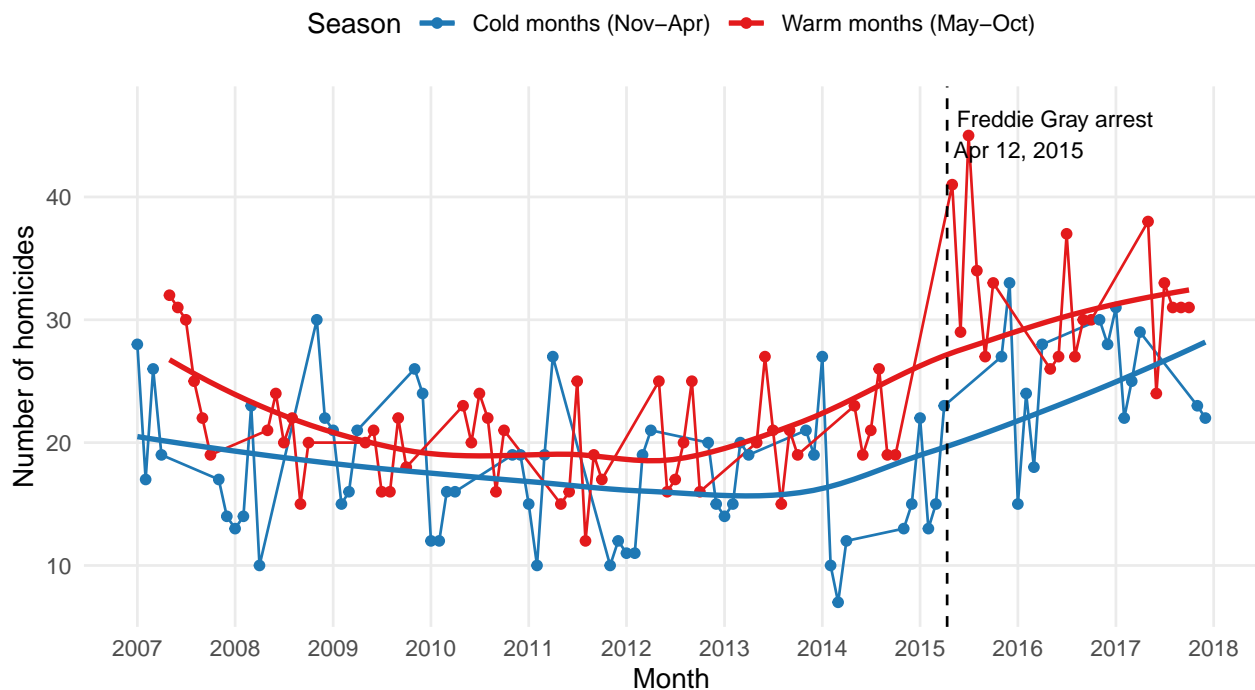


Figure 1. Monthly homicides in Baltimore, 2007–2017.

Cold months were defined as November–April, while warm months were defined as May–October. This follows the pattern used in the original Washington Post figure.

7. Methods summary

I used `ggplot2::geom_smooth()` with `method = "loess"`, which is appropriate for seasonal trend visualization.

Steps included: - Filtering Baltimore cases

- Converting `reported_date` to YYYY-MM-DD

- Aggregating monthly

- Creating season variable

- Plotting with loess trend + reference line

8. Interpretation

Warm months (May–October) consistently show higher homicide counts than cold months.

A noticeable spike occurs shortly after Freddie Gray’s arrest in April 2015.

The smooth trend indicates homicides rose from 2014–2016 before leveling.

9. Reproducibility notes

The project root is enforced using `knitr::opts_knit$set(root.dir = '.')` so knitting works regardless of the working directory. This analysis uses only CRAN packages and reads data from `data/homicide-data.csv`. Anyone cloning the repository with the same structure can knit the document.

10. Session Information

```
sessionInfo()
```

```
## R version 4.5.2 (2025-10-31)
## Platform: aarch64-apple-darwin20
## Running under: macOS Sequoia 15.6.1
##
## Matrix products: default
## BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/Libraries/libBLAS.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRlapack.dylib; LAPACK version 3.12.0
##
## locale:
##  [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/Denver
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
##  [1] scales_1.4.0    lubridate_1.9.4 forcats_1.0.1  stringr_1.6.0
##  [5] dplyr_1.1.4     purrr_1.2.0    readr_2.1.6    tidyr_1.3.1
##  [9] tibble_3.3.0    ggplot2_4.0.0  tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] utf8_1.2.6      generics_0.1.4  stringi_1.8.7  lattice_0.22-7
##  [5] hms_1.1.4       digest_0.6.38   magrittr_2.0.4  evaluate_1.0.5
```

## [9]	grid_4.5.2	timechange_0.3.0	RColorBrewer_1.1-3	fastmap_1.2.0
## [13]	Matrix_1.7-4	tinytex_0.58	mgcv_1.9-3	textshaping_1.0.4
## [17]	cli_3.6.5	rlang_1.1.6	crayon_1.5.3	bit64_4.6.0-1
## [21]	splines_4.5.2	withr_3.0.2	yaml_2.3.10	tools_4.5.2
## [25]	parallel_4.5.2	tzdb_0.5.0	vctrs_0.6.5	R6_2.6.1
## [29]	lifecycle_1.0.4	bit_4.6.0	vroom_1.6.7	ragg_1.5.0
## [33]	pkgconfig_2.0.3	pillar_1.11.1	gtable_0.3.6	glue_1.8.0
## [37]	systemfonts_1.3.1	xfun_0.54	tidyselect_1.2.1	rstudioapi_0.17.1
## [41]	knitr_1.50	farver_2.1.2	htmltools_0.5.8.1	nlme_3.1-168
## [45]	rmarkdown_2.30	labeling_0.4.3	compiler_4.5.2	S7_0.2.0