Homework #5

Rachel Brady

2023-11-28

Packages

```
library(readr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v purrr     1.0.2
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## -- 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
```

```
library(purrr)
library(broom)
library(ggplot2)
library(ggthemes)
library(lubridate)
library(stringr)
```

Get data, filter to one city

```
get_homicide <- paste0("https://raw.githubusercontent.com/washingtonpost/",
                       "data-homicides/master/homicide-data.csv")

homicide <- read_csv(get_homicide) %>%
  mutate(city_name = (paste(city, state, sep=", ")))
```

```
## 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.
```

```
balt <- homicide %>%
  filter(city_name == "Baltimore, MD" )
```

Get data ready for plotting

```
balt2 <- balt %>%
  mutate(reported_date = ymd(reported_date)) %>%
  mutate(year = substr(reported_date, 1,4)) %>%
  mutate(mo = substr(reported_date, 6,7)) %>%
  select(reported_date, year, mo) %>%
  group_by(year, mo) %>%
  summarize(monthly_count = n()) %>%
  mutate(date = paste0(year, mo, sep = "01")) %>%
  mutate(date = ymd(date)) %>%
  mutate(season = case_when(mo %in% c("05", "06", "07", "08", "09", "10") ~ "Summer",
                            mo %in% c("11", "12", "01", "02", "03", "04") ~ "Winter"))
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

Plotting

```
color <- c("#CCCCCC","#99CCFF", "white") #making color scheme

vline.data <- data.frame(z=c("2015-04-01"))
vline.data <- vline.data %>%
  mutate(z = ymd(z)) #getting date set for vertical line

balt2 %>%
  ggplot() +
  geom_col(aes(x = date, y = monthly_count, fill = season)) +
  geom_smooth(aes(x = date, y = monthly_count), color = "#6666FF", se = FALSE,
              span = 0.1, linewidth = 1.5) +
  theme_dark() +
  scale_fill_manual(values= color) +
  labs(title = "Homicides in Baltimore, MD",
       x = "Date",
       y = "Monthly homicides") +
  theme(legend.position = "bottom") +
  theme(legend.title=element_blank()) +
  geom_vline(aes(xintercept = z), color = "#CC0000", linewidth = 2,
             linetype = "dashed", vline.data) +
  annotate("text", x = ymd("2014-08-01"), y = 40,
           label = "        Arrest of \n Freddie Grey",
           color = "white")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```