Homework_5

Nikki Johnson

11/22/2019

```
#load required packages
library(readr)
library(magrittr)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:magrittr':
##
##     extract
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(forcats)
library(ggplot2)

#read in data
homicides <- read_csv (file = "https://raw.githubusercontent.com/washingtonpost/data-homicides/master/h
```

```
## Parsed with column specification:
## cols(
##   uid = col_character(),
##   reported_date = col_double(),
##   victim_last = col_character(),
##   victim_first = col_character(),
##   victim_race = col_character(),
##   victim_age = col_character(),
##   victim_sex = col_character(),
##   city = col_character(),
##   state = col_character(),
```

```
##   lat = col_double(),
##   lon = col_double(),
##   disposition = col_character()
## )
```

```
homicides
```

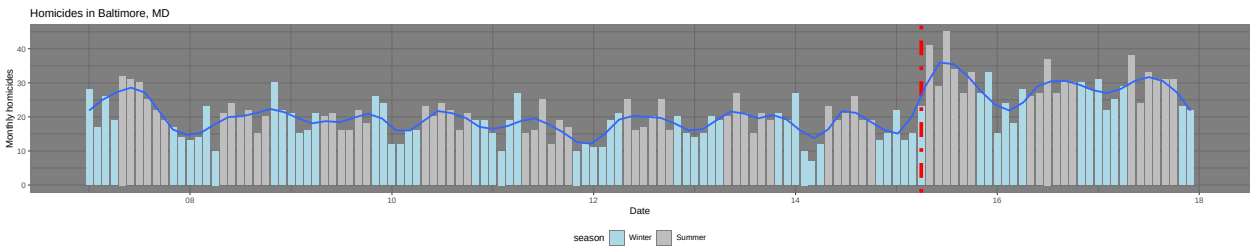
```
## # A tibble: 52,179 x 12
##    uid   reported_date victim_last victim_first victim_race victim_age
##    <chr>         <dbl> <chr>       <chr>        <chr>       <chr>
##  1 Alb-~      20100504 GARCIA      JUAN         Hispanic    78
##  2 Alb-~      20100216 MONTOYA     CAMERON      Hispanic    17
##  3 Alb-~      20100601 SATTERFIELD VIVIANA      White       15
##  4 Alb-~      20100101 MENDIOLA    CARLOS       Hispanic    32
##  5 Alb-~      20100102 MULA        VIVIAN       White       72
##  6 Alb-~      20100126 BOOK        GERALDINE    White       91
##  7 Alb-~      20100127 MALDONADO   DAVID        Hispanic    52
##  8 Alb-~      20100127 MALDONADO   CONNIE       Hispanic    52
##  9 Alb-~      20100130 MARTIN-LEY~ GUSTAVO      White       56
## 10 Alb-~      20100210 HERRERA     ISRAEL       Hispanic    43
## # ... with 52,169 more rows, and 6 more variables: victim_sex <chr>,
## #   city <chr>, state <chr>, lat <dbl>, lon <dbl>, disposition <chr>
```

```
#filter for baltimore homicides
baltimore <- homicides %>%
  filter(city == "Baltimore")

#pull out month and year
month <- baltimore %>%
  mutate(reported_date = ymd(reported_date)) %>%
  mutate(month = format(reported_date, "%m")) %>%
  mutate(year = format(reported_date, "%y"))

#making month a factor variable
month$month <- as.factor(month$month)

#setting up dates for graph
month$month_year <- format(as.Date(month$reported_date), "%y-%m")

#create season column
season <- month %>%
  mutate(season = fct_recode(month, 'Winter' = "01",
                             'Winter' = "02",
                             'Winter' = "03",
                             'Winter' = "04",
                             'Summer' = "05",
                             'Summer' = "06",
                             'Summer' = "07",
                             'Summer' = "08",
                             'Summer' = "09",
                             'Summer' = "10",
                             'Winter' = "11",
                             'Winter' = "12"))

#aggregate data and find total by month
```

```
season_2 <- season %>%
  select(year, month, season, month_year) %>%
  mutate(month_year = parse_date_time(month_year, "ym")) %>%
  mutate(month_year = ymd(month_year)) %>%
  count(month_year, season, name = 'total')
```

```
#data for arrest
arrest <- data.frame(month_year = as.Date('2015-04-12'), total = 45)

#plot data
plot_data <- season_2 %>%
  ggplot(aes(x = month_year, y = total)) +
  geom_col(aes(fill = season)) +
  geom_vline(xintercept = as.numeric(as.Date("2015-04-01")),
             linetype = 4, size = 2, color = "red") +
  geom_smooth(se = FALSE, span = 0.10) +
  scale_fill_manual(values = c("Winter" = "lightblue", "Summer" = "gray")) +
  scale_x_date(name = "Date", date_labels = "%y") +
  labs(y = "Monthly homicides") +
  ggtitle("Homicides in Baltimore, MD") +
  theme_dark() +
  theme(legend.position = "bottom")
plot_data
```

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




```
#add highlight
plot_highlight <- plot_data +
        geom_text(data = arrest,
                  label = "Arrest of Freddie Gray",
                  size = 7,
                  color = "white", vjust = 2, hjust = 1.15)
plot_highlight
```

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