BIOST 509: Homework 4

Instructor: Amy Willis, Biostatistics, UW

Due date: 1:00pm on November 1, 2019 via Canvas

Content advisory: This homework involves analysing of use of force data by Seattle Police. Two optional questions invite you to analyse this data from a racial equity lens. If you anticipate finding this homework challenging or upsetting to complete and would prefer to analyze different data instead, I welcome you to e-mail me (adwillis@uw.edu) and I will write a different homework for you. Please reach out to me if you have any questions.

Instructions

Last week, as part of the Health Sciences School's Common Book program, Ijeoma Oluo, author of *So you want to talk about race*, visited the School of Public Health. In keeping with the theme of the Common Book and School of Public Health's commitment to health equity, this homework looks at data on Use of Force incidents by police. There are two optional questions that invite you to analyse this data through the lens of racial equity.

The dataset is available from Canvas in `Pages/Module 4 materials` or `Files/datasets`. It was downloaded on October 24, 2019 from https://data.seattle.gov/Public-Safety/Use-Of-Force/ppi5-g2bj. The website data is updated daily (to my knowledge) and so the latest dataset may not match the dataset available on Canvas.

Each observation documents a incident of "Use of Force (UOF) by sworn law enforcement officers of the Seattle Police Department". This is data collected by the Seattle Police Department (i.e., not by civilians, bystanders, or those that experiences the UOF incident).

The variables included in the dataset are as follows:

- `ID`: Composite key for the identification of use of force.
- `Incident_Num`: Key identifying a force incident.
- `Incident_Type`: Use of force classification.
- `Occured_date_time`: Date and time that force occurred.
- `Precinct`: Precinct where the force occurred.
- `Sector`: Sector where the force occurred.
- `Beat`: Beat where the force occurred.
- `Officer_ID`: Key identifying unique officers.
- `Subject_ID`: Key identifying unique subjects.
- `Subject_Race`: Race of the subject of the use of force
- `Subject_Gender`: Gender of the subject of the use of force

Submit your answers to the below questions in a R Script (.R), Word (.doc or .docx) or pdf file to Canvas. Don't forget to include both your code and answers in your response.

Question 1

Before analyzing this dataset, we will need to modify the dates (`Occured_date_time`) into a useable format. This question will guide you through this process. A template with the code to preprocess the dates is available on Canvas under `Pages/Module 4 materials`.

First, if you do not already have the `lubridate` package installed, you can install it with the following command:

```
install.packages("lubridate")
```

You can then load the packages and the data as follows, modifying the file path:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1      v purrr   0.3.3
## v tibble  2.1.3      v dplyr   0.8.3
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
```

```
## -- Conflicts ----------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

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

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

```
use_of_force <- read_csv("datasets/Use_Of_Force.csv")
```

```
## Parsed with column specification:
## cols(
##   ID = col_character(),
##   Incident_Num = col_double(),
##   Incident_Type = col_character(),
##   Occured_date_time = col_character(),
##   Precinct = col_character(),
##   Sector = col_character(),
##   Beat = col_character(),
##   Officer_ID = col_double(),
##   Subject_ID = col_double(),
##   Subject_Race = col_character(),
##   Subject_Gender = col_character()
## )
```

You can then create a new tibble called `uof_dates` that contains the year of each incident in the column `Year` as follows:

```
uof_dates <- use_of_force %>%
  separate(Occured_date_time, into = c("Day", "Time", "AM-PM"), sep=" ") %>%
  mutate(Date = lubridate::mdy(Day, quiet = TRUE)) %>%
  mutate(Year = lubridate::year(Date))
```

Date manipulation is not a critical component of this course, which is why I provided the above code for you. If you're interested in understanding more about **separate** *and functions from the* **lubridate** *package, please come to office hours and I'll be happy to talk with you about them further.*

What years do we have data on?

Question 2

Make a scatterplot or line plot of the number of UOF incidents documented in each year. That is, make a plot with Year on the x-axis and number of incidents on the y-axis. You may use `plot` or `ggplot`.

Question 3

Save the plot from Question 2 as a `pdf` file. You do not need to upload the plot with your submission, but provide the code that you used to save the file.

Question 4 (optional)

Use `ggplot` to make a line plot of the number of UOF incidents documented in each year, with one line for each Racial group. In which year did UOF incidents spike for every race?

Optional: Add a title to your plot, begin the y-axis at zero, and format the axis labels nicely.

Question 5 (optional)

Race and Ethnicity statistics on the Seattle population are available at https://www.seattle.gov/opcd/population-and-demographics/about-seattle#raceethnicity. What fraction of UOF incidents were committed against each racial group in 2017? Contrast these numbers with the race and ethnicity demographic data on Seattle (linked above). Which groups comprise a higher proportion of UOF incidents relative to their proportion of the population?