Hw_5

Amy Fox Randy Xun Ana Velasquez

November 5, 2018

Load necessary packages

```
library(readr)
library(dplyr)
library(sf)
library(tigris)
library(ggplot2)
library(forcats)
library(ggthemes)
library(lubridate)
library(tidyr)
library(zoo)
library(knitr)
library(rgeos)
library(broom)
library(purrr)
```

Part 3

Pick one city in the data. Create a map showing the locations of the homicides in that city, using the sf framework discussed in class. Use tigris to download boundaries for some sub-city geography (e.g., tracts, block groups, county subdivisions) to show as a layer underneath the points showing homicides. Use different facets for solved versus unsolved homicides and different colors to show the three race groups with the highest number of homicides for that city (you may find the fct_lump function from forcats useful for this).

Read in homicide data

```
homicides <- read_csv("./data/homicide-data.csv")
homicides_4 <- homicides
homicides3 <- homicides
head(homicides)
```

```
## # A tibble: 6 x 12
##   uid        reported_date victim_last victim_first victim_race victim_age
##   <chr>              <int> <chr>       <chr>        <chr>       <chr>
## 1 Alb-000001      20100504 GARCIA      JUAN         Hispanic    78
## 2 Alb-000002      20100216 MONTOYA     CAMERON      Hispanic    17
## 3 Alb-000003      20100601 SATTERFIELD VIVIANA      White       15
## 4 Alb-000004      20100101 MENDIOLA    CARLOS       Hispanic    32
## 5 Alb-000005      20100102 MULA        VIVIAN       White       72
## 6 Alb-000006      20100126 BOOK        GERALDINE    White       91
## # ... with 6 more variables: victim_sex <chr>, city <chr>, state <chr>,
## #   lat <dbl>, lon <dbl>, disposition <chr>
```

Filter homicide data to only homicides in houston. Remove unnecessary columns. Create an unsolved logical column for the homicides that are unsolved. Remove rows that do not contain latitidue or longitude data

Add another column that takes the 3 races with the highest homicide numbers

```r
homicides <- homicides %>%
  filter(city == "Houston") %>%
  select(victim_race, lat, lon, disposition) %>%
  mutate(unsolved = disposition != "Closed by arrest") %>%
  filter(!is.na(lon), !is.na(lat))

Houston_homicides <- homicides %>%
  mutate(victim_race_factor = fct_lump(as.factor(victim_race), n = 3))
```

Convert Houston homicides to an sf object. Get the tracts (subdivisions) and bodies of water for the different counties in Houston and convert the sp objects to sf objects. Create pretty names for the facet wrapping headers. Finally, plot the homicide data facet wrapping by the unsolved vs solved cases overlayed onto the tracts and bodies of water.

```r
homicides_sf <- st_as_sf(Houston_homicides,
                         coords = c("lon", "lat")) %>%
  st_set_crs(4326)

TX_tracts <- tracts("Texas", county = c("Harris County", "Brazoria", "Fort Bend", "Montgomery",
                                        "Liberty", "Waller"),
                    cb = TRUE)
```

```
##
  |
  |                                                                  |   0%
  |
  |                                                                  |   1%
  |
  |=                                                                 |   1%
  |
  |=                                                                 |   2%
  |
  |==                                                                |   3%
  |
  |==                                                                |   4%
  |
  |===                                                               |   4%
  |
  |===                                                               |   5%
  |
  |====                                                              |   6%
  |
  |====                                                              |   7%
  |
  |=====                                                             |   7%
  |
  |=====                                                             |   8%
  |
  |======                                                            |   9%
  |
  |======                                                            |  10%
  |
  |=======                                                           |  10%
  |
|=======                                                          |  11%
|
|========                                                         |  12%
|
|========                                                         |  13%
|
|=========                                                        |  13%
|
|=========                                                        |  14%
|
|==========                                                       |  15%
|
|==========                                                       |  16%
|
|===========                                                      |  16%
|
|===========                                                      |  17%
|
|============                                                     |  18%
|
|============                                                     |  19%
|
|=============                                                    |  19%
|
|=============                                                    |  20%
|
|==============                                                   |  21%
|
|==============                                                   |  22%
|
|===============                                                  |  22%
|
|===============                                                  |  23%
|
|===============                                                  |  24%
|
|================                                                 |  24%
|
|================                                                 |  25%
|
|=================                                                |  26%
|
|=================                                                |  27%
|
|==================                                               |  27%
|
|==================                                               |  28%
|
|===================                                              |  29%
|
|===================                                              |  30%
|
|====================                                             |  30%
|
|====================                                             |  31%
|
|====================                                             |  32%
|
|=====================                                            |  32%
|
|=====================                                            |  33%
|
|======================                                           |  34%
|
|======================                                           |  35%
|
|=======================                                          |  35%
|
|=======================                                          |  36%
|
|========================                                         |  37%
|
|========================                                         |  38%
|
|=========================                                        |  38%
|
|=========================                                        |  39%
|
|==========================                                       |  40%
|
|==========================                                       |  41%
|
|===========================                                      |  41%
|
|===========================                                      |  42%
|
|============================                                     |  43%
|
|============================                                     |  44%
|
|=============================                                    |  44%
|
|=============================                                    |  45%
|
|==============================                                   |  45%
|
|==============================                                   |  46%
|
|==============================                                   |  47%
|
|===============================                                  |  47%
|
|===============================                                  |  48%
|
|================================                                 |  49%
|
|================================                                 |  50%
|
|=================================                                |  50%
|
|=================================                                |  51%
|
|==================================                               |  52%
|
|==================================                               |  53%
|
|===================================                              |  53%
|
|===================================                              |  54%
|
|====================================                             |  55%
|
|====================================                             |  56%
|
|=====================================                            |  56%
|
|=====================================                            |  57%
|
|=====================================                            |  58%
|
|======================================                           |  58%
|
|======================================                           |  59%
|
|=======================================                          |  59%
|
|=======================================                          |  60%
|
|========================================                         |  61%
|
|========================================                         |  62%
|
|=========================================                        |  62%
|
|=========================================                        |  63%
|
|==========================================                       |  64%
|
|==========================================                       |  65%
|
|===========================================                      |  66%
|
|===========================================                      |  67%
|
|============================================                     |  67%
|
|============================================                     |  68%
|
|=============================================                    |  69%
|
|=============================================                    |  70%
|
|==============================================                   |  71%
|
|===============================================                  |  72%
|
|===============================================                  |  73%
|
|================================================                 |  73%
|
|================================================                 |  74%
|
|=================================================                |  75%
|
|=================================================                |  76%
|
|==================================================               |  77%
|
|===================================================              |  78%
|
|===================================================              |  79%
|
|====================================================             |  80%
|
|=====================================================            |  81%
|
|=====================================================            |  82%
|
|======================================================           |  83%
|
|=======================================================          |  84%
|
|=======================================================          |  85%
|
|========================================================         |  86%
|
|=========================================================        |  87%
|
|=========================================================        |  88%
|
|==========================================================       |  89%
|
|===========================================================      |  90%
|
|===========================================================      |  91%
|
|============================================================     |  92%
|
|=============================================================    |  93%
|
|=============================================================    |  94%
|
|==============================================================   |  95%
|
|===============================================================  |  96%
|
  |================================================================  |  97%
  |
  |================================================================  |  98%
  |
  |================================================================= |  98%
  |
  |================================================================= |  99%
  |
  |==================================================================| 100%
```

```
TX_tracts <- st_as_sf(TX_tracts)
Houston_water <- area_water("Texas", county = c("Harris County", "Brazoria", "Fort Bend",
                                                "Montgomery", "Liberty", "Waller"))
```

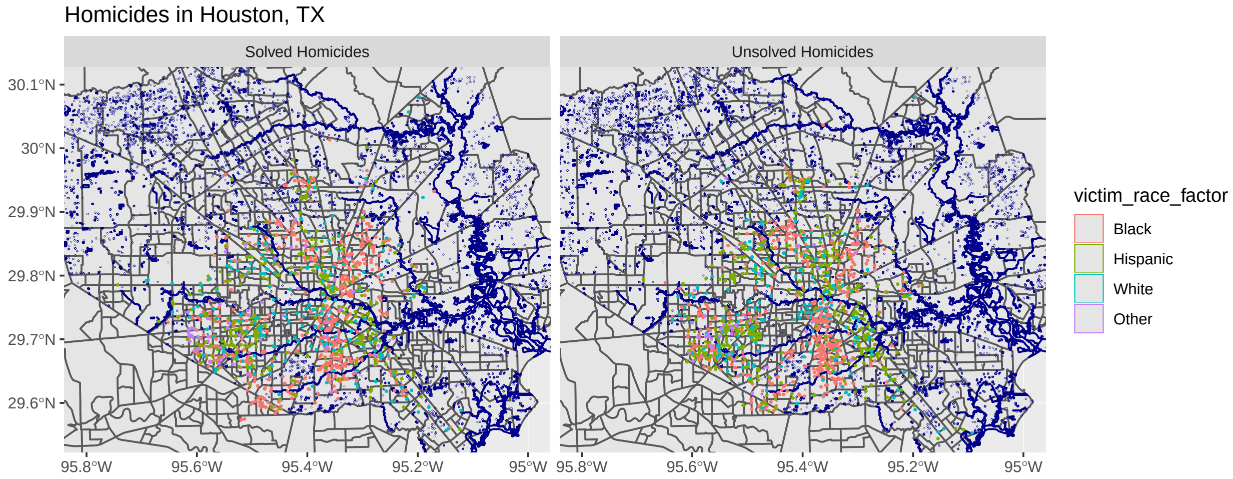
```
##
  |
  |                                                                  |   0%
  |
  |                                                                  |   1%
  |
  |=                                                                 |   1%
  |
  |==                                                                |   2%
  |
  |==                                                                |   4%
  |
  |===                                                               |   5%
  |
  |====                                                              |   6%
  |
  |=====                                                             |   8%
  |
  |======                                                            |   9%
  |
  |=======                                                           |  10%
  |
  |=======                                                           |  11%
  |
  |========                                                          |  13%
  |
  |=========                                                         |  14%
  |
  |==========                                                        |  15%
  |
  |===========                                                       |  17%
  |
  |============                                                      |  18%
  |
  |============                                                      |  19%
  |
  |=============                                                     |  20%
  |
  |==============                                                    |  22%
  |
  |===============                                                   |  23%
```

```
|
|================                                                 |  24%
|
|=================                                                |  26%
|
|=================                                                |  27%
|
|==================                                               |  28%
|
|===================                                              |  29%
|
|====================                                             |  30%
|
|====================                                             |  31%
|
|=====================                                            |  32%
|
|======================                                           |  33%
|
|======================                                           |  35%
|
|=======================                                          |  36%
|
|========================                                         |  37%
|
|=========================                                        |  38%
|
|==========================                                       |  40%
|
|===========================                                      |  41%
|
|===========================                                      |  42%
|
|============================                                     |  44%
|
|=============================                                    |  45%
|
|==============================                                   |  46%
|
|==============================                                   |  47%
|
|===============================                                  |  47%
|
|================================                                 |  49%
|
|=================================                                |  50%
|
|=================================                                |  51%
|
|==================================                               |  52%
|
|==================================                               |  53%
|
|===================================                              |  54%
```

```
|
|===================================                              | 55%
|
|====================================                             | 55%
|
|=====================================                            | 56%
|
|======================================                           | 58%
|
|======================================                           | 59%
|
|=======================================                          | 60%
|
|========================================                         | 62%
|
|=========================================                        | 63%
|
|==========================================                       | 64%
|
|==========================================                       | 65%
|
|===========================================                      | 65%
|
|===========================================                      | 67%
|
|============================================                     | 68%
|
|=============================================                    | 69%
|
|==============================================                   | 71%
|
|===============================================                  | 72%
|
|================================================                 | 73%
|
|================================================                 | 74%
|
|=================================================                | 76%
|
|==================================================               | 77%
|
|===================================================              | 78%
|
|====================================================             | 80%
|
|=====================================================            | 81%
|
|=====================================================            | 82%
|
|======================================================           | 83%
|
|=======================================================          | 85%
|
|========================================================         | 86%
```

```
  |
  |=========================================================        |  87%
  |
  |==========================================================       |  89%
  |
  |==========================================================       |  90%
  |
  |===========================================================      |  91%
  |
  |============================================================     |  92%
  |
  |=============================================================    |  94%
  |
  |==============================================================   |  95%
  |
  |===============================================================  |  96%
  |
  |===============================================================  |  98%
  |
  |================================================================ |  99%
  |
  |=================================================================| 100%
```

```r
Houston_water <- st_as_sf(Houston_water)

facet_names <- c("TRUE" = "Unsolved Homicides",
                 "FALSE" = "Solved Homicides")

ggplot(homicides_sf) +
  facet_wrap(~unsolved, ncol = 2,
             labeller = as_labeller(facet_names)) +
  geom_sf(data = TX_tracts) +
  geom_sf(data = Houston_water, color = "blue4" ) +
  geom_sf(data = homicides_sf, aes(color = victim_race_factor), size = .3) +
  coord_sf(xlim = c(-95.8, -95), ylim = c(29.55, 30.1)) +
  labs(fill = "Victim Race") +
  ggtitle("Homicides in Houston, TX")
```



Part 4

Recreate the graph shown below. It shows monthly homicides in Baltimore, with a reference added for the date of the arrest of Freddie Gray and color used to show colder months (November through April) versus warmer months (May through October). There is a smooth line added to help show seasonal and long-term trends in this data

```
#Selecting Baltimore data and adding season (Winter or Summer).

baltimore <- homicides_4 %>%
  unite(city_names, city, state, sep = ",") %>%
  unite(victim_name, victim_first, victim_last, sep = " ") %>%
  select(reported_date, city_names) %>%
  filter(city_names == "Baltimore,MD") %>%
  mutate(reported_date = ymd(reported_date)) %>%
  mutate(dates = as.yearmon(reported_date))
```

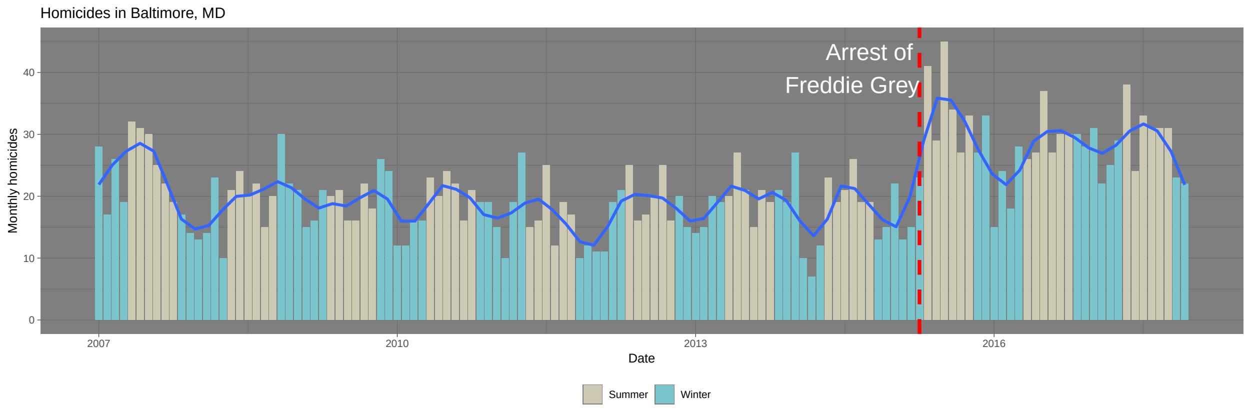
```r
seasons = function(x) {
  if(x %in% 5:10) return ("Summer")
  if(x %in% c(11,12,1,2,3,4)) return ("Winter")
}
baltimore$season = sapply(month(baltimore$reported_date), seasons)

#Counting Baltimorehomicides cases by month/year.

counts_homicides <- baltimore %>%
  group_by(dates) %>%
  count() %>%
  ungroup()

#Joining Baltomore data with homicides cases.

baltimore<-baltimore %>%
  right_join(counts_homicides, by = "dates")

#Selecting unique values per month

baltimore <- unique(data.frame(event_date = baltimore$dates,
                               event_count = baltimore$n,
                               season = baltimore$season))

head(baltimore)
```

```
##     event_date event_count season
## 1     Jan 2007          28 Winter
## 29    Feb 2007          17 Winter
## 46    Mar 2007          26 Winter
## 72    Apr 2007          19 Winter
## 91    May 2007          32 Summer
## 123   Jun 2007          31 Summer
```

```r
#Creating an object called "Freddie" to add to the final plot.

freddie <- homicides_4 %>%
  unite(victim_name, victim_first, victim_last, sep = " ") %>%
  select(reported_date, victim_name) %>%
  mutate(reported_date = ymd(reported_date)) %>%
  mutate(dates = as.yearmon(reported_date))

freddie <- freddie %>%
  right_join(counts_homicides, by = "dates") %>%
  filter(victim_name == "FREDDIE CARLOS GREY")
head(freddie)
```

```
## # A tibble: 1 x 4
##   reported_date victim_name         dates              n
##   <date>        <chr>               <S3: yearmon> <int>
## 1 2015-04-30    FREDDIE CARLOS GREY Apr 2015         23
```

```r
baltimore %>%
  ggplot() +
  geom_bar(aes(x = event_date, y = event_count, fill = season),
           stat = "identity") +
  scale_fill_manual(values = c("cornsilk3", "cadetblue3")) +
  geom_smooth( aes(x = event_date, y = event_count),
               method = "loess", se = FALSE, span = 0.1, size = 1.2) +
  labs(x = "Date", y = "Monthly homicides", fill = " ") +
  theme_dark() +
  theme(legend.position = "bottom") +
  ggtitle("Homicides in Baltimore, MD") +
  geom_vline(data = freddie, xintercept = (freddie$dates),
             color = "red", linetype = "dashed", size = 1.5) +
  geom_text(data = freddie, aes(x = dates, y = n, label = "Arrest of
                                Freddie Grey"), size = 7, hjust = 1,
                                vjust = -1.7, colour = "white")
```



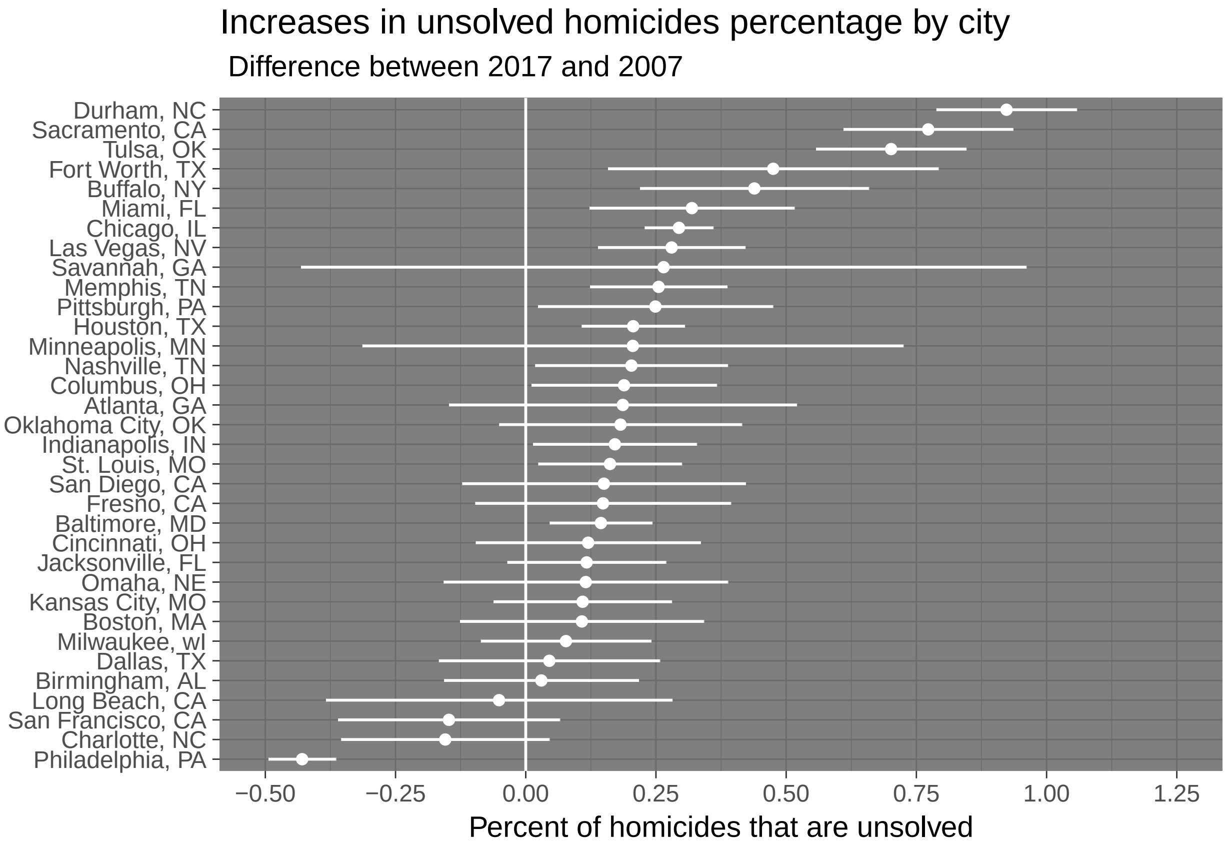
Part 5

Create one more plot using this data. Work with your group to create a plot that follows the principles of good plotting and that you think illustrates something interesting in the data. Write a paragraph explaining what the plot is showing and why you find it interesting.

```
homicides3 <- homicides3 %>%
  unite(col = city_name, city, state, sep = ", ") %>%
  mutate(reported_date = ymd(reported_date)) %>%
  arrange(reported_date)

homi2017 <- filter(homicides3, reported_date >= ymd(20170101) &
                     reported_date <= ymd(20171231))

homi2007 <- filter(homicides3, reported_date >= ymd(20070101) &
                     reported_date <= ymd(20071231))

unsolved2017 <- homi2017 %>%
  select(city_name, disposition) %>%
  group_by(city_name) %>%
  mutate(unsolved = disposition != 'Closed by arrest') %>%
  summarize(total_homicides17 = n(),
            total_unsolved17 = sum(unsolved))

unsolved2007 <- homi2007 %>%
```

```r
  select(city_name, disposition) %>%
  group_by(city_name) %>%
  mutate(unsolved = disposition != 'Closed by arrest') %>%
  summarize(total_homicides07 = n(),
            total_unsolved07 = sum(unsolved))

proportion_unsolved2017 <- unsolved2017 %>%
  mutate(test = map2(total_unsolved17, total_homicides17,
                     ~ prop.test(.x, n = .y))) %>%
  mutate(test = map(test,
                    ~ tidy(.x))) %>%
  unnest(.drop = TRUE) %>%
  select(city_name, estimate, total_homicides17)

proportion_unsolved2007 <- unsolved2007 %>%
  mutate(test = map2(total_unsolved07, total_homicides07,
                     ~ prop.test(.x, n = .y))) %>%
  mutate(test = map(test,
                    ~ tidy(.x))) %>%
  unnest(.drop = TRUE) %>%
  select(city_name, estimate, total_homicides07)

combo_proportion <- inner_join(proportion_unsolved2017, proportion_unsolved2007,
                               by = "city_name")

diff_proportion <- combo_proportion %>%
  mutate(diff_percent = (estimate.x - estimate.y))

diff_proportion <- diff_proportion %>%
  mutate(x_min = (diff_percent -
          2.575 * sqrt((estimate.x * (1-estimate.x)/total_homicides17)
          + (estimate.y*(1-estimate.y)/total_homicides07)))) %>%
  mutate(x_max = (diff_percent +
          2.575 * sqrt((estimate.x * (1-estimate.x)/total_homicides17)
          + (estimate.y*(1-estimate.y)/total_homicides07))))
```

```r
diff_proportion %>%
  mutate(city_name = fct_reorder(city_name, diff_percent)) %>%
  ggplot(aes(x = diff_percent, y = city_name)) +
  geom_point(color= "white") +
  theme_dark() +
  labs(title = "Increases in unsolved homicides percentage by city",
     subtitle = " Difference between 2017 and 2007",
     y = NULL, x = "Percent of homicides that are unsolved") +
  scale_x_continuous(limits = c(-0.5, 1.25),
                 breaks = c(-0.5, -0.25, 0, 0.25, 0.5, 0.75, 1.0, 1.25)) +
  geom_errorbarh(color = "white",
                 aes(y = city_name, xmin = x_min,
                     xmax = x_max, height = 0)) +
  geom_vline(xintercept = 0, color = "white")
```



For this part of the assignment, our group looked at how the percentage of unsolved murders in the major cities changed from the first year of the study (2007) to the last year of the study (2017). This finding is interesting because it could potentially show which police forces are getting better at solving murders. At the same time, this graph can help identify cities which are declining over time. As you can see by this graph, with the help of the error bars and the white vertical line indicating no change, only Philadelphia, PA has significantly improved, statistically, its ability to solved homicides in 10 years time.