

# ERHS 535: Homework 4

Ana Velasquez Munoz

November 4, 2018

## Part 1.

Please check my github. README file is under the folder **Writing**

[github ana] ([https://github.com/anakvm/hw\\_4](https://github.com/anakvm/hw_4))

## Part 2: Estimating the proportion of unsolved homicides in each city

1. and 2. Reading the data as an R object named `homicides` and creating a new column called `city_name` that combines the city and state

```
homicides <- read_csv("data-homicide/homicide-data.csv") %>%
  unite(city_names, city, state, sep = ",")
head(homicides)
```

```
## # A tibble: 6 x 11
##   uid   reported_date victim_last victim_first victim_race victim_age
##   <chr>         <int> <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
## # ... with 5 more variables: victim_sex <chr>, city_names <chr>,
## #   lat <dbl>, lon <dbl>, disposition <chr>
```

3. Creating a dataframe called *unsolved* with one row per city that gives the total number of homicides for the city and the number of unsolved homicides.

```
## # A tibble: 6 x 3
##   city_names      n_homicides unsolved
##   <chr>          <int>      <int>
## 1 Albuquerque,NM      378      146
## 2 Atlanta,GA          973      373
## 3 Baltimore,MD       2827     1825
## 4 Baton Rouge,LA       424      196
## 5 Birmingham,AL        800      347
## 6 Boston,MA           614      310
```

4. Baltimore prop.test and confidence interval.

estimate	statistic	p.value	parameter	conf.low	conf.high	method
0.6455607	239.011	0	1	0.6275625	0.6631599	1-sample proportions test with continuity correction

## 5. Prop.test all cities and confidence interval using purrr::map2.

```
unsolved <- homicides %>%
  select(city_names, disposition)%>%
  mutate(unsolved = disposition %in% c("Closed without arrest",
                                       "Open/No arrest")) %>%

  group_by(city_names) %>%
  summarize(n_homicides = n(),
            unsolved = sum(unsolved)) %>%

  #applying prop test to all cities, and making a statement to
  #call for the estimates and confidence intervals
  #in the data frame.
  mutate(newcol = purrr::map2(unsolved, n_homicides,
                              ~ prop.test(.x, n = .y) %>%
                                {data.frame(estimate = .[["estimate"]],
                                              ci_lower = .[["conf.int"]][[1]],
                                              ci_upper = .[["conf.int"]][[2]]})) %>%

  unnest()

kable(head(unsolved))
```

city_names	n_homicides	unsolved	estimate	ci_lower	ci_upper
Albuquerque,NM	378	146	0.3862434	0.3372604	0.4375766
Atlanta,GA	973	373	0.3833505	0.3528119	0.4148219
Baltimore,MD	2827	1825	0.6455607	0.6275625	0.6631599
Baton Rouge,LA	424	196	0.4622642	0.4141987	0.5110240
Birmingham,AL	800	347	0.4337500	0.3991889	0.4689557
Boston,MA	614	310	0.5048860	0.4646219	0.5450881

## 6. Plot of unsolved homicides for all cities using geom\_errorbarh.

```
unsolved %>%
  slice(-49) %>% #Removing Tulsa,AL
  mutate(city_names = fct_reorder(city_names, estimate)) %>%
  ggplot() +
  geom_errorbarh(aes(y = city_names, x = estimate, xmin = ci_lower,
                    xmax = ci_upper), height = 0, color = "white")+
  geom_point(mapping = aes(y = city_names, x = estimate),
            size = 1, shape = 21, color = "white", fill = "white") +
  theme_dark()+
  scale_x_continuous(labels = percent)+
  labs(x = "Percent of homicides that are unsolved", y = " ") +
  ggtitle("Unsolved homicides by city",
          subtitle = "Bars show 95% confidence interval")
```

## Unsolved homicides by city

Bars show 95% confidence interval

