Homework\_4

## import the homecides data

url <- paste0("https://raw.githubusercontent.com/washingtonpost/", "data-homicides/master/homicide-data.csv")  
homicides <- read.csv(url, header = T)

## create new columns including city\_name and unsolved, filter out Tulsa, AL because it doens’t exist

homicides <- homicides %>%   
 mutate(city\_name = paste(city, state, sep = ", ")) %>%   
 filter(city\_name != "Tulsa, AL")

## create total cases and unsolved cases for each city

total\_homicides <- homicides %>%   
 group\_by(city\_name) %>%   
 summarize(n()) %>%   
 rename(cases\_total = `n()`)  
unsolved\_homicides <- homicides %>%   
 group\_by(city\_name) %>%   
 select(city\_name, disposition) %>%   
 filter(str\_detect(disposition,   
 "Closed without arrest|Open/No arrest")) %>%   
 summarize(n()) %>%   
 rename(unsolved\_cases = `n()`)  
unsolved <- full\_join(total\_homicides, unsolved\_homicides, by = "city\_name")

## estimate the proportion of unsolved homicide cases in Baltimore, MD

baltimore <- unsolved[unsolved$city\_name == "Baltimore, MD",]  
baltimore\_prop <- prop.test(x = baltimore$unsolved\_cases,   
 n = baltimore$cases\_total,  
 conf.level = 0.95)  
kable(tidy(baltimore\_prop) %>%   
 select(estimate,  
 conf.low,  
 conf.high))

estimate

conf.low

conf.high

0.6455607

0.6275625

0.6631599

## perform same test on all other cities

all\_unsolved <- unsolved %>%   
 mutate(prop\_test = map2(unsolved\_cases, cases\_total, prop.test),   
 tidy\_prop\_test = map(prop\_test, tidy)) %>%   
 unnest(tidy\_prop\_test, .drop = TRUE)

## recreate the given plot

all\_unsolved %>%   
 select(city\_name, estimate, conf.low, conf.high) %>%   
 mutate(city\_name = reorder(city\_name, estimate)) %>%   
 ggplot(aes(x = estimate, y = city\_name)) +  
 geom\_point(color = "white") +  
 geom\_errorbarh(xmin = all\_unsolved$conf.low,  
 xmax = all\_unsolved$conf.high,  
 height = 0,  
 color = "white") +  
 scale\_x\_continuous(breaks = c(0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8),  
 label = c("20.0%", "30.0%", "40.0%",   
 "50.0%", "60.0%", "70.0%", "80.0%")) +  
 labs(x = "Percent of homicides that are unsolved",  
 y = "") +  
 ggtitle("Unsolved homicides by city",  
 subtitle = "Bars show 95% confidence interval") +  
 theme\_dark()

