Homework 4, Burton Karger

Load required libraries.

library(tidyverse)  
library(tidyr)  
library(broom)  
library(purrr)  
library(scales)

Read data in, combine city/state columns.

homicides <- read\_csv("homicide-data.csv", col\_names = TRUE) %>%   
 unite(col = "city\_state", city, state, sep = ",") %>%   
 print()

## # A tibble: 52,179 x 11  
## 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 5 more variables: victim\_sex <chr>,  
## # city\_state <chr>, lat <dbl>, lon <dbl>, disposition <chr>

Summary of dispostion of cases (closed by arrest, or not).

unsolved <- homicides %>%   
 mutate(unsolved = disposition != "Closed by arrest") %>%   
 group\_by(city\_state) %>%   
 summarize(total = n(), total\_unsolved = sum(unsolved)) %>%  
 select(city\_state, total\_unsolved, total) %>%   
 print()

## # A tibble: 51 x 3  
## city\_state total\_unsolved total  
## <chr> <int> <int>  
## 1 Albuquerque,NM 146 378  
## 2 Atlanta,GA 373 973  
## 3 Baltimore,MD 1825 2827  
## 4 Baton Rouge,LA 196 424  
## 5 Birmingham,AL 347 800  
## 6 Boston,MA 310 614  
## 7 Buffalo,NY 319 521  
## 8 Charlotte,NC 206 687  
## 9 Chicago,IL 4073 5535  
## 10 Cincinnati,OH 309 694  
## # ... with 41 more rows

Making frame work for map2 call in which prop.test is applied to each row (city\_state).

baltimore <- unsolved[3,]  
balt\_prop <- prop.test(baltimore$total\_unsolved, baltimore$total) %>%   
 tidy()  
  
balt\_prop %>%   
 select(estimate, conf.high, conf.low)

## # A tibble: 1 x 3  
## estimate conf.high conf.low  
## <dbl> <dbl> <dbl>  
## 1 0.646 0.663 0.628

Applying prop.test and tidying output of statistical analysis.

estimates\_unsolved <- unsolved %>%   
 mutate(results = map2(total\_unsolved, total, ~prop.test(.x, .y))) %>%   
 mutate(results = map(results, ~tidy(.x))) %>%   
 unnest(.drop = TRUE) %>%   
 select(city\_state, estimate, conf.low, conf.high) %>%   
 print()

## # A tibble: 51 x 4  
## city\_state estimate conf.low conf.high  
## <chr> <dbl> <dbl> <dbl>  
## 1 Albuquerque,NM 0.386 0.337 0.438  
## 2 Atlanta,GA 0.383 0.353 0.415  
## 3 Baltimore,MD 0.646 0.628 0.663  
## 4 Baton Rouge,LA 0.462 0.414 0.511  
## 5 Birmingham,AL 0.434 0.399 0.469  
## 6 Boston,MA 0.505 0.465 0.545  
## 7 Buffalo,NY 0.612 0.569 0.654  
## 8 Charlotte,NC 0.300 0.266 0.336  
## 9 Chicago,IL 0.736 0.724 0.747  
## 10 Cincinnati,OH 0.445 0.408 0.483  
## # ... with 41 more rows

Mimicking graph.

estimates\_unsolved %>%  
 slice(-49) %>%   
 ggplot(aes(x = estimate, y = factor(city\_state, levels = city\_state[order(estimate)]))) +  
 geom\_point(color = "white") +  
 geom\_errorbarh(aes(xmax = conf.high, xmin = conf.low, height = 0), color = "white") +  
 ggtitle("Unsolved homicides by city", subtitle = "Bars show 95% confidence interval") +  
 labs(x = "Percent of homicides that are unsolved", y = "") +  
 theme(plot.subtitle = element\_text(face = "bold")) +  
 theme\_dark() +  
 scale\_x\_continuous(labels = percent)

