Homework 4

Camron Pearce

November 8, 2018

Tidying the Data

```
total_homicides <- homicides %>%
  group_by(city_name) %>%
  count()

unsolved <- homicides %>%
  select(city_name, disposition) %>%
  mutate(not_solved = disposition %in% c("Closed without arrest",
                                         "Open/No arrest")) %>%
  filter(not_solved == T) %>%
  group_by(city_name) %>%
  count() %>%
  rename(total_unsolved = n) %>%
  ungroup()

total_unsolved <- unsolved %>%
  left_join(total_homicides, by = "city_name") %>%
  rename(unsolved =  "total_unsolved",
         total_homicides = "n")

balt_homicides <- total_unsolved %>%
  filter(city_name == "Baltimore, MD")
balt_homicides
```

```
## # A tibble: 1 x 3
##   city_name     unsolved total_homicides
##   <chr>            <int>           <int>
## 1 Baltimore, MD     1825            2827
```

Statistical Analysis of Baltimore

```
##
##  1-sample proportions test with continuity correction
##
## data:  balt_homicides$unsolved out of balt_homicides$total_homicides, null probability 0.5
## X-squared = 239.01, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.6275625 0.6631599
## sample estimates:
##         p
## 0.6455607
```

```
## # A tibble: 1 x 8
##   estimate statistic  p.value parameter conf.low conf.high method
##      <dbl>     <dbl>    <dbl>     <int>    <dbl>     <dbl> <chr>
## 1    0.646      239. 6.46e-54         1    0.628     0.663 1-sam~
## # ... with 1 more variable: alternative <chr>
```
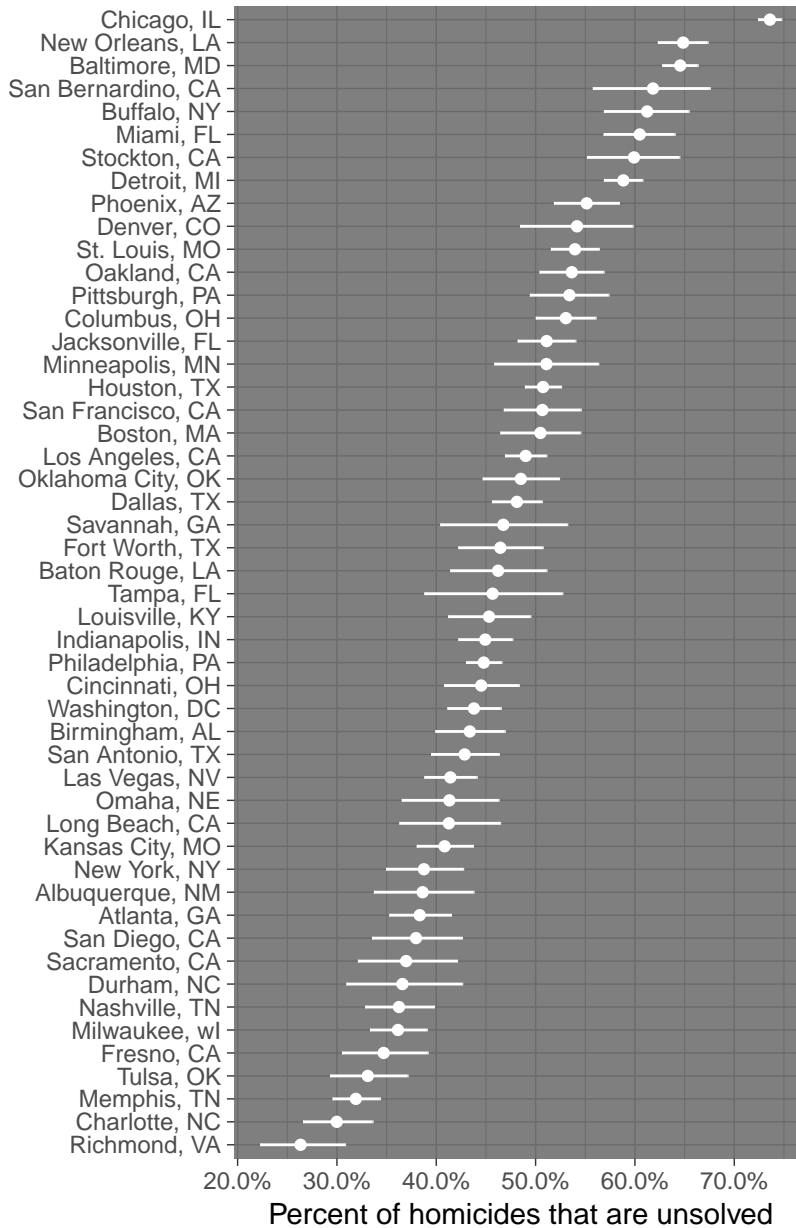
Statistical Testing for Unsolved Homicides by City

```r
unsolved_df <- total_unsolved %>%
  mutate(newcol = map2(unsolved, total_homicides,
                       ~ prop.test(.x, n = .y) %>%
                {tibble(estimate = .[["estimate"]],
                        CI_lower = .[["conf.int"]][[1]],
                        CI_upper = .[["conf.int"]][[2]])})) %>%
  unnest() %>%
  mutate(city_name = factor(city_name, levels = city_name[order(estimate)]))


unsolved_df %>%
  ggplot(aes(estimate, city_name)) +
  geom_point(color = "white")+
  geom_errorbarh(aes(xmin = CI_lower,
                     xmax = CI_upper,
                      y = city_name,
                      height = 0), color = "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% confience interval")
```

Unsolved homicides by city
Bars show 95% confience interval

Percent of homicides that are unsolved

```
#library(rmarkdown)
#render("homework_4.rmd", output_format = "pdf_document")
```