homework4

Amanda Walz

October 27, 2018

Homicide arrest rates in different major cities across the United States.

```
homicide_url <- paste0("https://raw.githubusercontent.com/washingtonpost/data-homicides/master/homicide-

#Load in data and tidy data
homi_data <- read_csv(homicide_url)
```

```
## Parsed with column specification:
## cols(
##   uid = col_character(),
##   reported_date = col_integer(),
##   victim_last = col_character(),
##   victim_first = col_character(),
##   victim_race = col_character(),
##   victim_age = col_character(),
##   victim_sex = col_character(),
##   city = col_character(),
##   state = col_character(),
##   lat = col_double(),
##   lon = col_double(),
##   disposition = col_character()
## )
```

```
homi_table <- homi_data %>%
  unite(col = city_name, city, state, sep = ", ")

#Creating table with counts by city
pre_unsolved <- homi_table %>%
  select(city_name, disposition) %>%
  mutate(not_solved = disposition %in% c("Closed without arrest", "Open/No arrest")) %>%
  filter(not_solved == TRUE) %>%
  group_by(city_name) %>%
  count() %>%
  rename(total_unsolved = n) %>%
  ungroup()

pre_unsolved2 <- homi_table %>%
  select(city_name) %>%
  group_by(city_name) %>%
  count() %>%
  rename(total_homicide = n) %>%
  ungroup()

unsolved <- left_join(pre_unsolved, pre_unsolved2, by = "city_name")
```

```r
#Run prop test on baltimore data
baltimore_summary <- unsolved %>%
  slice(3)
baltimore_prob <- prop.test(
  x = baltimore_summary$total_unsolved,
  n = baltimore_summary$total_homicide)
baltimore_prob
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  baltimore_summary$total_unsolved out of baltimore_summary$total_homicide, 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
```

```r
tidy(baltimore_prob)
```

```
## # 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>
```

```r
#Tidying data for final graph
unsolved_prop <- map2(unsolved$total_unsolved, unsolved$total_homicide, prop.test) %>%
  map_df(tidy) %>%
  select(estimate, conf.low, conf.high) %>%
  mutate(city_name = unsolved$city_name)
kable(unsolved_prop,
      caption = "Unsolved Homicides by City")
```

Table 1: Unsolved Homicides by City

estimate	conf.low	conf.high	city_name
0.3862434	0.3372604	0.4375766	Albuquerque, NM
0.3833505	0.3528119	0.4148219	Atlanta, GA
0.6455607	0.6275625	0.6631599	Baltimore, MD
0.4622642	0.4141987	0.5110240	Baton Rouge, LA
0.4337500	0.3991889	0.4689557	Birmingham, AL
0.5048860	0.4646219	0.5450881	Boston, MA
0.6122841	0.5687990	0.6540879	Buffalo, NY
0.2998544	0.2660820	0.3358999	Charlotte, NC
0.7358627	0.7239959	0.7473998	Chicago, IL
0.4452450	0.4079606	0.4831439	Cincinnati, OH
0.5304428	0.5002167	0.5604506	Columbus, OH
0.4811742	0.4561942	0.5062475	Dallas, TX
0.5416667	0.4846098	0.5976807	Denver, CO
0.5883287	0.5687903	0.6075953	Detroit, MI
0.3659420	0.3095874	0.4260936	Durham, NC
0.4644809	0.4222542	0.5072119	Fort Worth, TX

estimate	conf.low	conf.high	city_name
0.3470226	0.3051013	0.3913963	Fresno, CA
0.5074779	0.4892447	0.5256914	Houston, TX
0.4493192	0.4223156	0.4766207	Indianapolis, IN
0.5111301	0.4820460	0.5401402	Jacksonville, FL
0.4084034	0.3803996	0.4370054	Kansas City, MO
0.4141926	0.3881284	0.4407395	Las Vegas, NV
0.4126984	0.3629026	0.4642973	Long Beach, CA
0.4900310	0.4692208	0.5108754	Los Angeles, CA
0.4531250	0.4120609	0.4948235	Louisville, KY
0.3190225	0.2957047	0.3432691	Memphis, TN
0.6048387	0.5685783	0.6400015	Miami, FL
0.3614350	0.3333172	0.3905194	Milwaukee, wI
0.5109290	0.4585150	0.5631099	Minneapolis, MN
0.3624511	0.3285592	0.3977401	Nashville, TN
0.6485356	0.6231048	0.6731615	New Orleans, LA
0.3875598	0.3494421	0.4270755	New York, NY
0.5364308	0.5040588	0.5685037	Oakland, CA
0.4851190	0.4467861	0.5236245	Oklahoma City, OK
0.4132029	0.3653146	0.4627477	Omaha, NE
0.4478103	0.4300380	0.4657157	Philadelphia, PA
0.5514223	0.5184825	0.5839244	Phoenix, AZ
0.5340729	0.4942706	0.5734545	Pittsburgh, PA
0.2634033	0.2228571	0.3082658	Richmond, VA
0.3696809	0.3211559	0.4209131	Sacramento, CA
0.4285714	0.3947772	0.4630331	San Antonio, TX
0.6181818	0.5576628	0.6753422	San Bernardino, CA
0.3796095	0.3354259	0.4258315	San Diego, CA
0.5067873	0.4680516	0.5454433	San Francisco, CA
0.4674797	0.4041252	0.5318665	Savannah, GA
0.5396541	0.5154369	0.5636879	St. Louis, MO
0.5990991	0.5517145	0.6447418	Stockton, CA
0.4567308	0.3881009	0.5269851	Tampa, FL
0.3310463	0.2932349	0.3711192	Tulsa, OK
0.4379182	0.4112495	0.4649455	Washington, DC

```
plot_data <- full_join(unsolved, unsolved_prop) %>%
  select(-total_homicide, -total_unsolved) %>%
  arrange(desc(estimate))
```

```
## Joining, by = "city_name"
```

```
homicide_plot <- plot_data %>%
  mutate(city_name = reorder(city_name, estimate)) %>%
  ggplot(mapping = aes(x = estimate, y = city_name)) +
  geom_point(color = "White") +
  geom_errorbarh(aes(xmin = plot_data$conf.low,
                     xmax = plot_data$conf.high),
                     height = 0,
                     color = "white") +
  scale_x_continuous(name = "Percent of homicides that are unsolved", labels = percent) +
  theme_dark() +
  ggtitle("Unsolved homicides by city",
```

```
          "Bars show 95% confidence inverval") +
  ylab(NULL)
homicide_plot
```