Homework4

Devin Nelson

10/31/2018

Background The Washington Post collected data on more than 52,000 criminal homicides over the past decade in 50 of the largest American cities.The data included the location of the killing, whether an arrest was made and, in most cases, basic demographic information about each victim. This analysis will focus on the proportion of homicides that went unsolved.

Analysis

First, you need to read in the homicide dataset, which is in a CSV format.

```r
homicide <- read.csv("/Users/devinnelson/Desktop/Homework4/Data/homicide-data.csv")
```

First you will have to combine the city and the state into a single variable using the `unite` function.

```r
homicide <- homicide %>%
  unite(cityname, city, state, sep = ", ")
```

Next, you will create a new dataset titled 'unsolved' that consists of three variables: city name, total number of homicides, and total number of unsolved homicides with one row for each city.To do so, you need to create a new variable titled "totalunsolved" that is a true/false variable for the disposition finding. You can now use the sum function to count the number of "Trues".

```r
unsolved <- homicide %>%
  select(cityname, disposition) %>%
  mutate(disposition = factor(disposition)) %>%
  group_by(cityname) %>%
  mutate(totalhomicides = n()) %>%
  mutate(totalunsolved = disposition == 'Closed without arrest'| disposition == 'Open/No arrest') %>%
  mutate(totalunsolved = sum(totalunsolved)) %>%
  group_by(cityname, totalunsolved) %>%
  count() %>%
  ungroup() %>%
  rename(totalhomicides = n)
```

In order to understand how to find the proportion of unsolved on all observations, it is helpful to practice on one. Therefore, 'Baltimore, MD' was pulled out to find the proportion using the `prop.test` and `tidy` functions.

```r
baltimore <- unsolved %>%
  filter(cityname == 'Baltimore, MD')

baltimore_prop <- prop.test(x = baltimore$totalunsolved,
                            n = baltimore$totalhomicides)

baltimore_prop <- tidy(baltimore_prop)

baltimore_prop
```

```
## # 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>
```

Now that we have a better understanding of what the output of one observation looks like, it can be mapped to all of the observations using the `map` and `map2` functions, as shown below. Additionally, it can be placed into a ggplot to demonstrate all observations' proportions alongside the error bars.

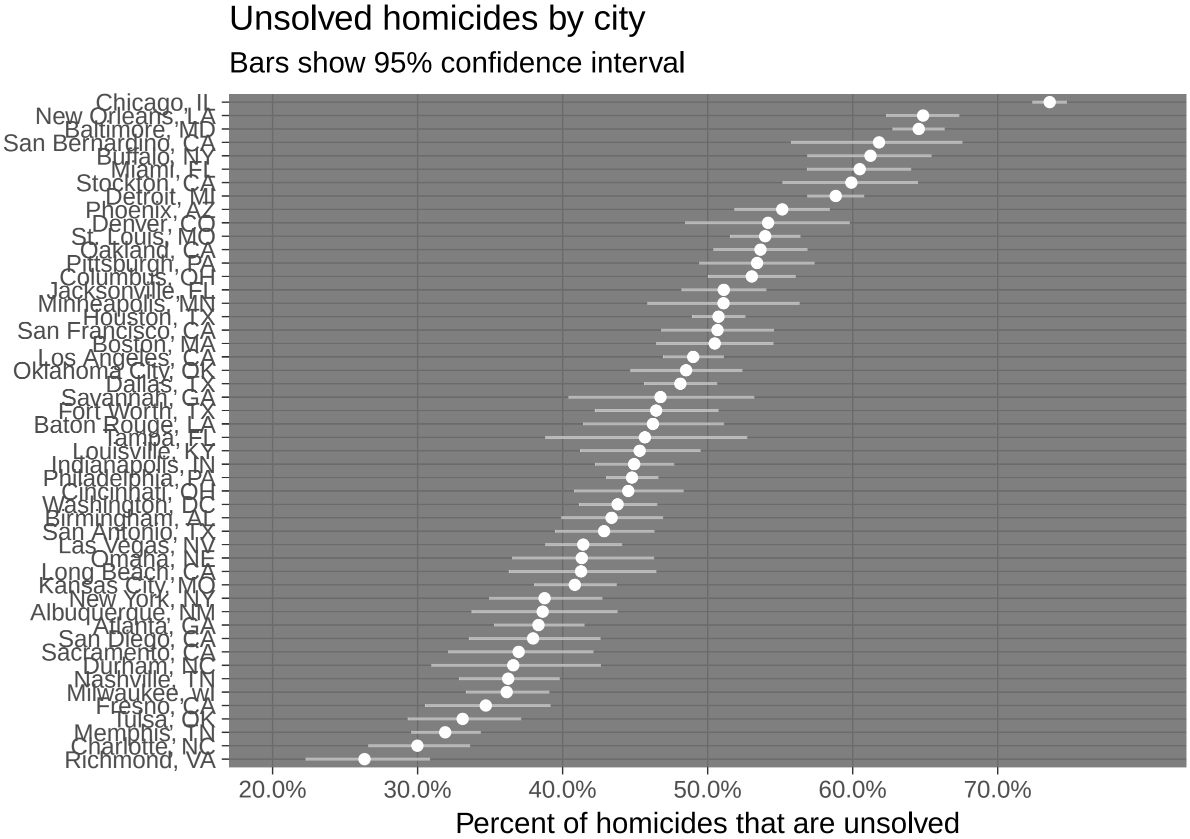
```
unsolvedprop <- unsolved %>%
  mutate(result = map2(totalunsolved, totalhomicides,  ~ prop.test(.x, n = .y))) %>%
  mutate(result = map(result, tidy)) %>%
  unnest(.drop = TRUE) %>%
  select(cityname, estimate, conf.low, conf.high) %>%
  filter(cityname != 'Tulsa, AL') %>%
  ggplot() +
  geom_point(mapping = aes(x = estimate, y = reorder(cityname, estimate)),
           color = 'white') +
  geom_errorbarh(mapping = aes(y = cityname, x = estimate, xmin = conf.low, xmax = conf.high),
               color = 'white',
               height = 0,
               alpha = .5) +
  scale_x_continuous(labels = percent,
                   breaks = c(.2,.3,.4,.5,.6,.7),
                   limits = c(0.2, 0.8),
                   minor_breaks = c(.2,.3,.4,.5,.6,.7)) +
  labs(x = 'Percent of homicides that are unsolved', y = '') +
  ggtitle("Unsolved homicides by city",
        subtitle = "Bars show 95% confidence interval") +
  theme_dark()
```

```
## Warning in prop.test(.x, n = .y): Chi-squared approximation may be
## incorrect
```

```
## Warning: Ignoring unknown aesthetics: x
```

Final Plot

The final graph will look as follows:


Unsolved homicides by city
Bars show 95% confidence interval
Chicago, IL
New Orleans, LA
Baltimore, MD
San Bernardino, CA
Buffalo, NY
Miami, FL
Stockton, CA
Detroit, MI
Phoenix, AZ
Denver, CO
St. Louis, MO
Oakland, CA
Pittsburgh, PA
Columbus, OH
Jacksonville, FL
Minneapolis, MN
Houston, TX
San Francisco, CA
Boston, MA
Los Angeles, CA
Oklahoma City, OK
Dallas, TX
Savannah, GA
Fort Worth, TX
Baton Rouge, LA
Tampa, FL
Louisville, KY
Indianapolis, IN
Philadelphia, PA
Cincinnati, OH
Washington, DC
Birmingham, AL
San Antonio, TX
Las Vegas, NV
Omaha, NE
Long Beach, CA
Kansas City, MO
New York, NY
Albuquerque, NM
Atlanta, GA
San Diego, CA
Sacramento, CA
Durham, NC
Nashville, TN
Milwaukee, WI
Fresno, CA
Tulsa, OK
Memphis, TN
Charlotte, NC
Richmond, VA
20.0%
30.0%
40.0%
50.0%
60.0%
70.0%
Percent of homicides that are unsolved