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))
```

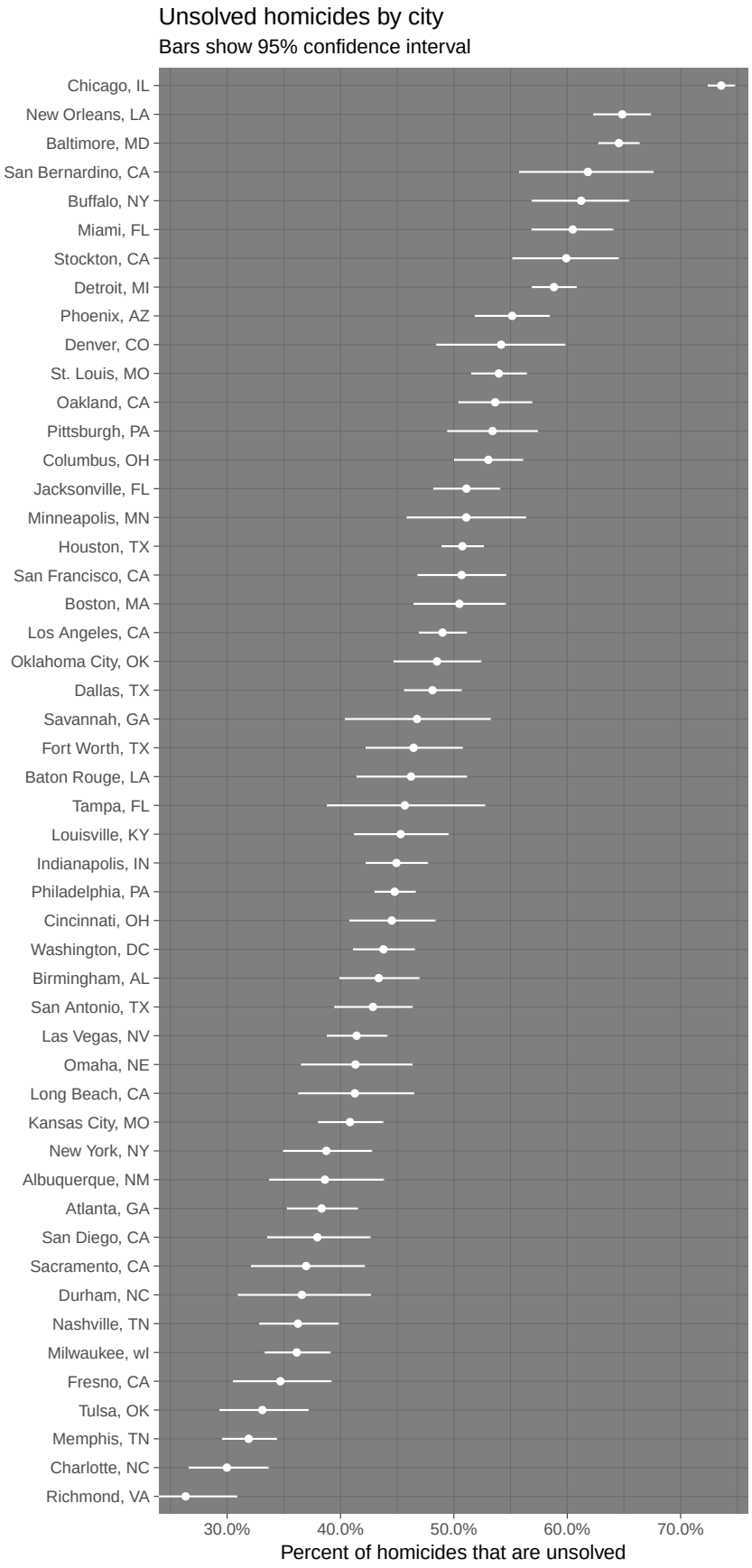
estimate	conf.low	conf.high
0.6455607	0.6275625	0.6631599

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()
```


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
30.0%
40.0%
50.0%
60.0%
70.0%
Percent of homicides that are unsolved