

# ECON 3255 Data Assignment 1

Christine Shi

2022-09-22

## Setup

```
# load data
ucr_1960_2020 <- read.csv("ucr_1960_2020.csv")

# load packages
library(tidyverse)
library(knitr)
library(kableExtra)
```

1)

How many police agencies report data to the UCR in each decadal year (e.g., 1960, 1970, etc.)?

```
# create table of number of reporting agencies by decade
agencies_by_decade <- ucr_1960_2020 %>%
  # split data into decades based on year
  mutate(decade = case_when(year %>% between(1960, 1969) ~ 1960,
                             year %>% between(1970, 1979) ~ 1970,
                             year %>% between(1980, 1989) ~ 1980,
                             year %>% between(1990, 1999) ~ 1990,
                             year %>% between(2000, 2009) ~ 2000,
                             year %>% between(2010, 2020) ~ 2010)) %>%

  # calculate number of distinct agency names by decade
  group_by(decade) %>%
  summarize(num_agencies = n_distinct(ori)) # ORI = unique identifier

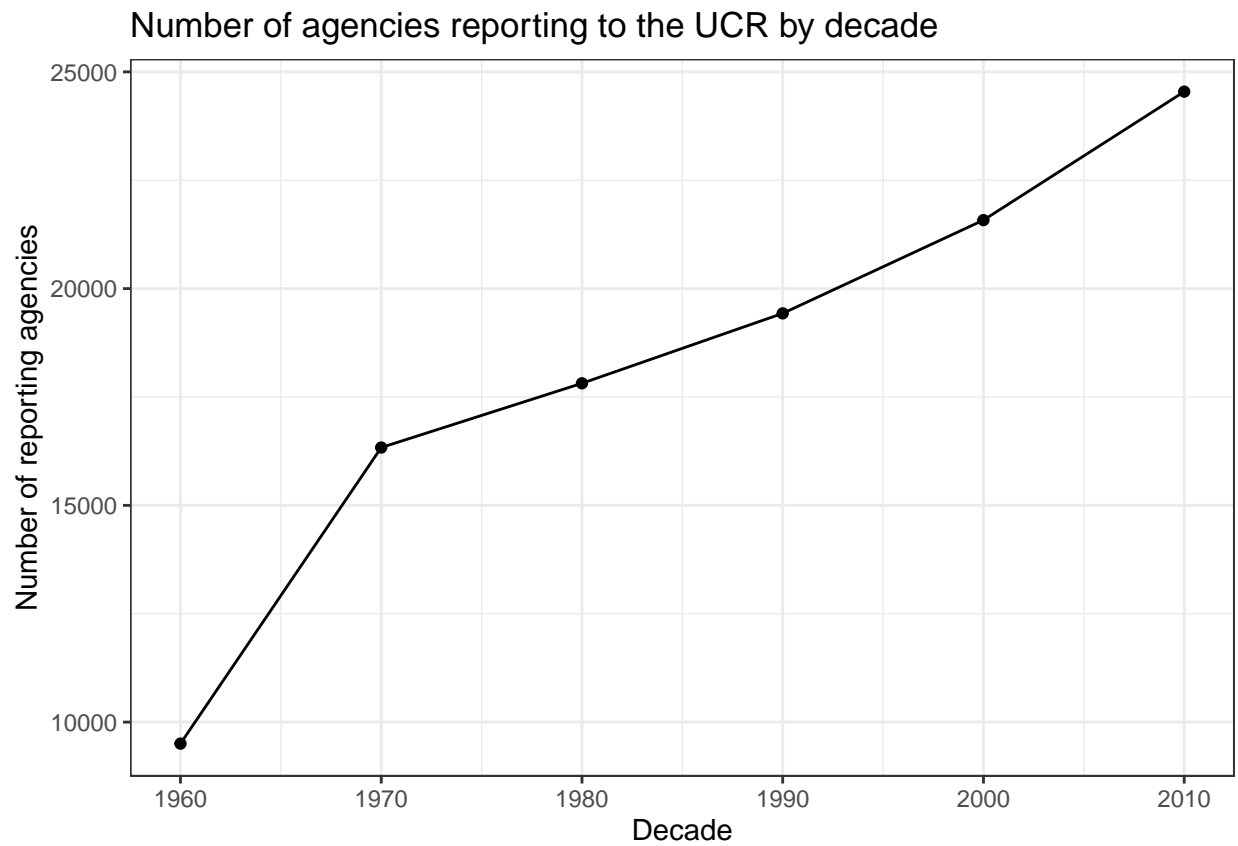
# display table with descriptive title and column names
agencies_by_decade %>%
  rename("Number of reporting agencies" = num_agencies,
         "Decade" = decade) %>%
  kable(caption = "Number of agencies reporting to the UCR by decade")

# line plot of data
agencies_by_decade %>% ggplot(aes(decade, num_agencies)) +
  geom_line() +
```

Table 1: Number of agencies reporting to the UCR by decade

Decade	Number of reporting agencies
1960	9504
1970	16334
1980	17816
1990	19429
2000	21580
2010	24544

```
geom_point() +
labs(title = "Number of agencies reporting to the UCR by decade",
      x = "Decade",
      y = "Number of reporting agencies") +
theme_bw()
```



The number of agencies reporting to the UCR has increased decade by decade since 1960, from around 10,000 to around 25,000.

Table 2: Number of local police departments reporting by year

Year	Number of local police departments
1960	5335
1961	5340
1962	5070
1963	5587
1964	5622
1965	5642

2)

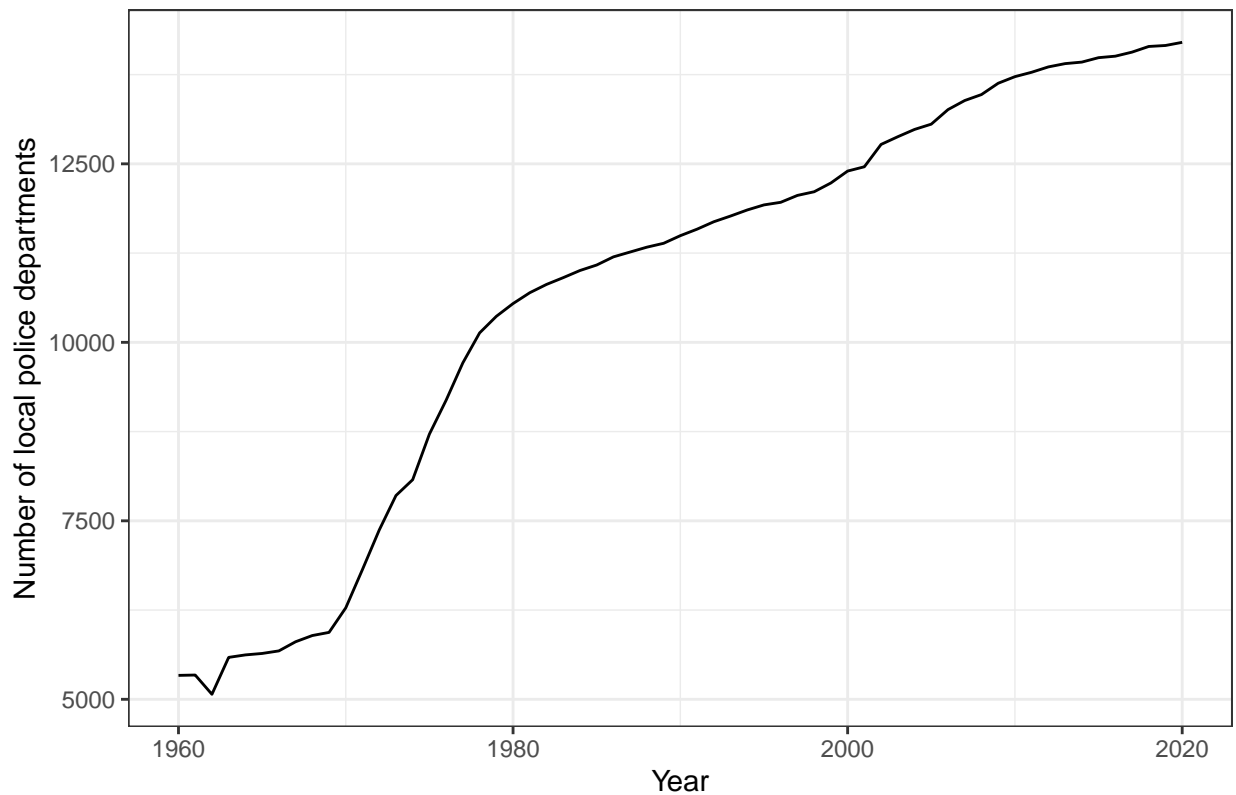
Across all years of the data, how many of the agencies are local police departments? (Note: this is not the number of agency-year observations for local police departments.)

```
# create table of local police departments by year
local_by_year <- ucr_1960_2020 %>%
  # group data by year
  group_by(year) %>%
  # filter for local police departments
  filter(agency_type == "local police department") %>%
  # count number of departments
  summarize(num_local = n_distinct(ori))

# display the table with descriptive title and column names
local_by_year %>%
  rename("Number of local police departments" = num_local,
         "Year" = year) %>%
  head() %>%
  kable(caption = "Number of local police departments reporting by year")

# plot the data
local_by_year %>% ggplot(aes(year, num_local)) +
  geom_line() +
  labs(title = "Number of local police departments reporting by year",
       x = "Year",
       y = "Number of local police departments") +
  theme_bw()
```

Number of local police departments reporting by year



The number of agencies reporting to the UCR that are local police departments has been increasing since 1960, from around 5,000 to around 14,000.

3)

For each police agency and year, calculate the total number of reported violent and property index crimes. Then, for each police agency and year, calculate the total rate of reported violent and property index crimes per 100,000 residents.

```
total_by_year_and_agency <- ucr_1960_2020 %>%
  #group by agency and year
  group_by(crosswalk_agency_name, year) %>%
  rowwise() %>%
  # sum number of violent crimes
  mutate(total_violent_crimes = sum(c(actual_murder,
                                     actual_manslaughter,
                                     actual_rape_total,
                                     actual_robbery_total,
                                     actual_assault_aggravated)),

         # rate per 100,000 residents
         total_violent_crime_rate = total_violent_crimes/(population_1/100000),
         # repeat for property crimes
         total_property_crimes = sum(c(actual_burg_total,
                                     actual_theft_total,
                                     actual_mtr_veh_theft_total)),
```

Table 3: Total number and rate of violent and property crimes by police agency and year from 1960 to 2020

Year	Agency	Population	Total violent crimes	Total violent crime rate (per 100,000 residents)	Total property crimes	Total property crime rate (per 100,000 residents)
2020	anchorage police department	286388	3472	1212.341	9872	3447.072
2019	anchorage police department	287731	3584	1245.608	12261	4261.272
2018	anchorage police department	291992	3826	1310.310	14389	4927.875
2017	anchorage police department	296188	3567	1204.303	16041	5415.817
2016	anchorage police department	299097	3427	1145.782	14649	4897.742
2015	anchorage police department	301239	3226	1070.910	11801	3917.487

```

    total_property_crime_rate = total_property_crimes/(population_1/100000)) %>%
# calculate these variables for each agency and year
summarize(ori,
  agency_type,
  population_1,
  total_violent_crimes,
  total_violent_crime_rate,
  total_property_crimes,
  total_property_crime_rate)

# display table with descriptive column names
total_by_year_and_agency %>%
  select(-c(ori, agency_type)) %>%
  rename("Year" = year,
    "Agency" = crosswalk_agency_name,
    "Population" = population_1,
    "Total violent crimes" = total_violent_crimes,
    "Total violent crime rate (per 100,000 residents)" = total_violent_crime_rate,
    "Total property crimes" = total_property_crimes,
    "Total property crime rate (per 100,000 residents)" = total_property_crime_rate) %>%
  head() %>%
  kable(caption = "Total number and rate of violent and property crimes by police agency and year from 1960 to 2020",
    kable_styling() %>% column_spec(4:7, width = "2cm"))

```

The `total_by_year_and_agency` table contains data on the total number and rate (per 100,000 residents) of reported violent and property index crimes for each police agency and year. Violent index crimes include murder (and manslaughter), rape, robbery, and aggravated assault. Property index crimes include burglary, theft, and motor vehicle theft.

Which police agency had the highest total number of reported violent index crimes in 1980?

```

total_by_year_and_agency %>%
  #filter to the year 1980
  filter(year == 1980) %>%
  #arrange from highest to lowest number of reported violent index crimes
  arrange(desc(total_violent_crimes)) %>%

```

Table 4: NYC had the highest number of reported violent crimes in 1980

Year	Agency	Total number of reported violent index crimes in 1980	Total rate of reported violent index crimes in 1980
1980	new york city police department	149555	2125.765

Table 5: Vernon had the highest total rate of reported violent index crimes in 1980

Year	Agency	Total reported violent index crimes in 1980	Total rate of reported violent index crimes in 1980
1980	vernon police department	70	78651.69

```
# keep the agency with the highest number of reported violent index crime
head(1) %>%
select(c(year, crosswalk_agency_name, total_violent_crimes, total_violent_crime_rate)) %>%
# rename columns for readability
rename("Year" = year,
       "Agency" = crosswalk_agency_name,
       "Total number of reported violent index crimes in 1980" =
         total_violent_crimes,
       "Total rate of reported violent index crimes in 1980" =
         total_violent_crime_rate) %>%
kable(caption = "NYC had the highest number of reported violent crimes in 1980") %>%
kable_styling() %>% column_spec(2:4, width = "5cm")
```

The New York City Police Department had the highest total number of reported violent index crimes (149,555) in 1980.

Which police agency had the highest total rate of reported violent index crimes in 1980, and what was the population of its jurisdiction that year?

```
# repeat for violent crime rate
total_by_year_and_agency %>%
# data quality issue: in some years, some agencies reported 0 population
# this results in some crime rates being calculated as `Inf`
# filter these zero-population observations out
filter(year == 1980, population_1 > 0) %>%
group_by(crosswalk_agency_name) %>%
rowwise() %>%
arrange(desc(total_violent_crime_rate)) %>%
head(1) %>%
select(c(year, crosswalk_agency_name, total_violent_crimes, total_violent_crime_rate)) %>%
rename("Year" = year,
       "Agency" = crosswalk_agency_name,
       "Total reported violent index crimes in 1980" = total_violent_crimes,
       "Total rate of reported violent index crimes in 1980" = total_violent_crime_rate) %>%
kable(caption = "Vernon had the highest total rate of reported violent index crimes in 1980") %>%
kable_styling() %>% column_spec(2:4, width = "5cm")
```

The Vernon Police Department had the highest total rate of reported violent index crimes in 1980 (78,652 per 100,000 residents) with a jurisdiction population of 89.

Table 6: Total number and rate of violent and property crimes across all police agencies by year from 1960 to 2020

Year	Total Population	Total violent crimes	Total violent crime rate (per 100,000 residents)	Total property crimes	Total property crime rate (per 100,000 residents)
1960	181302747	211440	116.6226	2663661	1469.179
1961	180897952	243449	134.5781	2973838	1643.931
1962	171201833	240757	140.6276	2880316	1682.410
1963	191357191	265085	138.5289	3455339	1805.701
1964	194237698	324013	166.8126	3855916	1985.153
1965	196730164	342057	173.8712	3985370	2025.805

4)

For each year, calculate the total number of reported violent and property index crimes, along with the total population, across all police agencies. Then, for each year, calculate the total rate of reported violent and property index crimes per 100,000 residents.

```
total_by_year <- total_by_year_and_agency %>%
  group_by(year) %>%
  summarize(total_population = sum(population_1),
            total_violent_crimes = sum(total_violent_crimes),
            total_violent_crime_rate = total_violent_crimes/(total_population/100000),
            total_property_crimes = sum(total_property_crimes),
            total_property_crime_rate = total_property_crimes/(total_population/100000))

total_by_year %>%
  rename("Year" = year,
         "Total Population" = total_population,
         "Total violent crimes" = total_violent_crimes,
         "Total violent crime rate (per 100,000 residents)" = total_violent_crime_rate,
         "Total property crimes" = total_property_crimes,
         "Total property crime rate (per 100,000 residents)" = total_property_crime_rate) %>%
  head() %>%
  kable(caption = "Total number and rate of violent and property crimes across all police agencies by year",
        kable_styling() %>% column_spec(-(1), width = "2.5cm"))
```

Plot the total rates of both reported violent and property index crimes from 1960 to 2020 on one figure. To make the figure readable, plot violent and property crime rates on different Y-axes (labeled accordingly) and include a legend to distinguish the two crime types from each other.

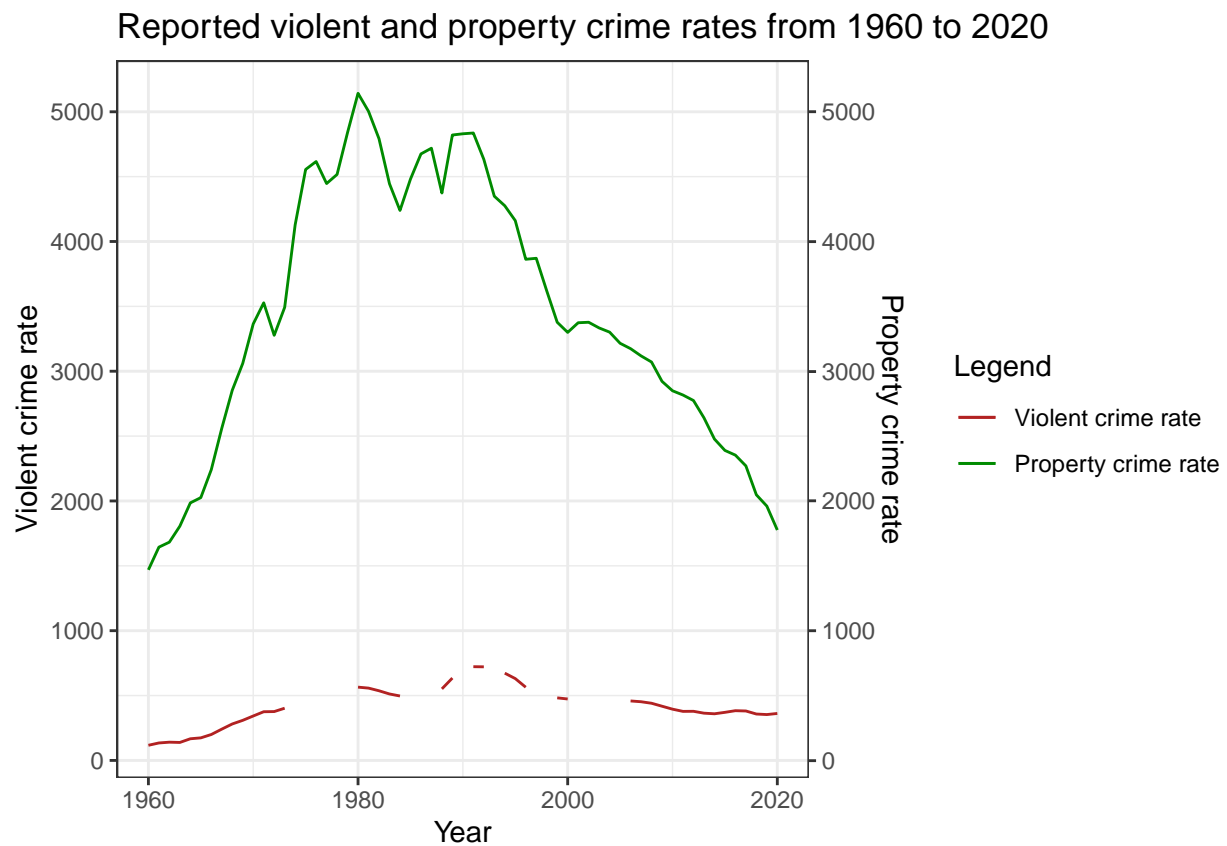
```
# assigning colors to different variables
colors <- c("Violent crime rate" = "firebrick", "Property crime rate" = "green4")

total_by_year %>% ggplot(aes(x = year)) +
  geom_line(aes(y = total_violent_crime_rate, # line for violent crime
```

```

    color = "Violent crime rate")) +
  geom_line(aes(y = total_property_crime_rate, # line for property crime
    color = "Property crime rate")) +
  scale_y_continuous(
    name = "Violent crime rate",
    sec.axis = sec_axis(trans = ~., name = "Property crime rate")
  ) + # labeling the second axis for property crime
  labs(title = "Reported violent and property crime rates from 1960 to 2020",
    x = "Year",
    color = "Legend") +
  scale_color_manual(values = colors) +
  theme_bw()

```



5)

In what year did the rate of reported violent index crimes reach its peak? Reported property index crimes?

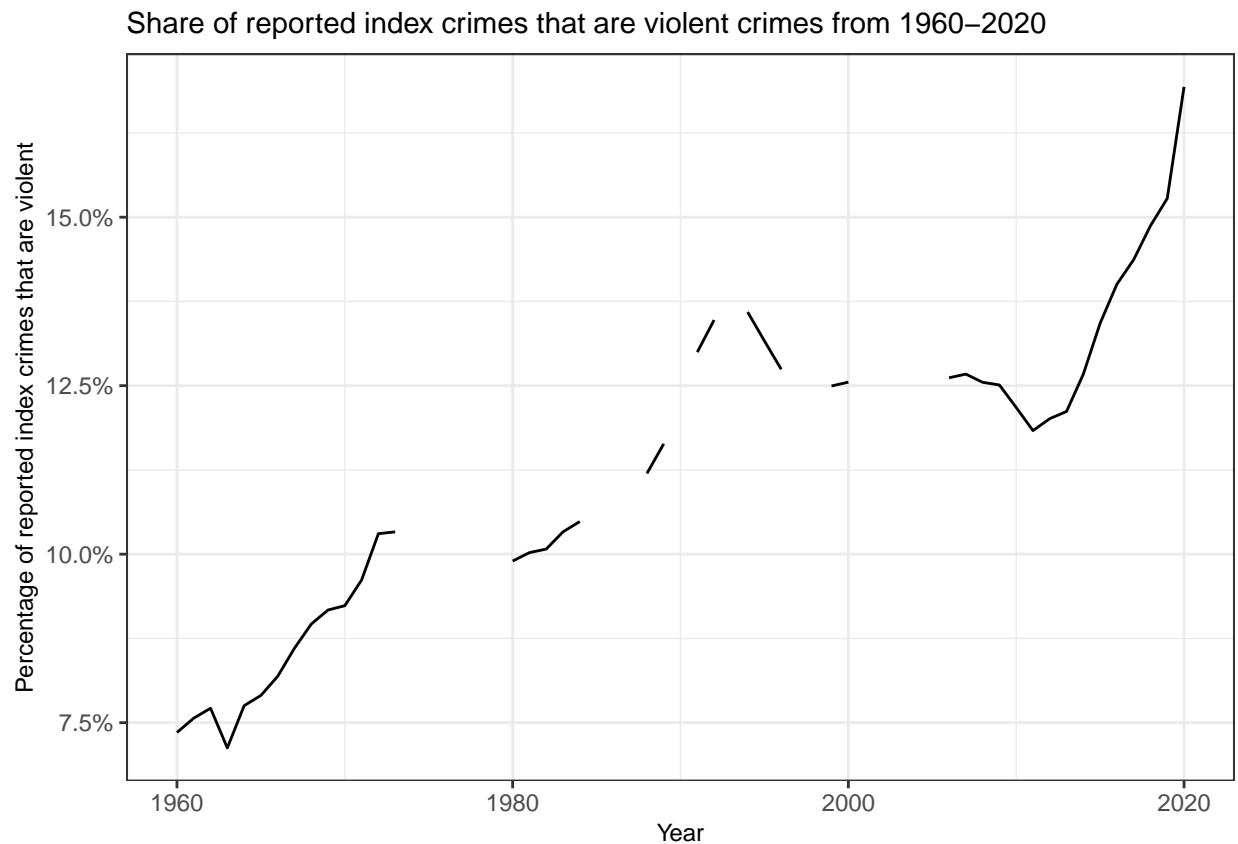
From the graph, it looks like reported violent index crimes reached their peak in around 1992. Reported property index crimes reached their peak in around 1980.



6)

Plot the share of all reported index crimes that are violent crimes from 1960 to 2020, labeling the Y-axis accordingly.

```
total_by_year %>%
  mutate(total_crimes = total_violent_crimes + total_property_crimes,
         share_violent = total_violent_crimes/total_crimes) %>%
  ggplot(aes(year, share_violent)) +
  geom_line() +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Share of reported index crimes that are violent crimes from 1960-2020",
       x = "Year",
       y = "Percentage of reported index crimes that are violent") +
  theme_bw() +
  theme(title = element_text(size = rel(0.85)),
        axis.title = element_text(size = rel(0.95)))
```



7)

For the remaining questions, please limit the analysis sample to the years 1973 to 2020, and limit police agencies to:

- those that are local police departments,
- with populations equal to or greater than 250,000 in 1990, and
- with non-zero population in all the years 1973 to 2020.

```
# filter the dataset with total number/rate of violent/property crime for year/agency from 1960-2020
ucr_1973_2020 <- total_by_year_and_agency %>%
  filter(year %>% between(1973, 2020), # 1973-2020
         agency_type == "local police department") %>% #local police depts
  # deselect columns we don't need
  select(-c(total_violent_crimes, total_property_crimes, total_property_crime_rate))

# get the ORIs of the agencies with >= 250,000 population in 1990
check_1990 <- ucr_1973_2020 %>%
  filter(year == 1990) %>%
  filter(population_1 >= 250000)

# of those agencies, get the agencies reporting 0 population in any year
check_zero_pop <- ucr_1973_2020 %>%
  filter(ori %in% check_1990$ori) %>%
  filter(population_1 == 0)

# apply filters
ucr_1973_2020_filtered <- ucr_1973_2020 %>%
  filter(ori %in% check_1990$ori) %>% #1990 filter
  filter(!(ori %in% check_zero_pop$ori)) # zero pop filter

ucr_1973_2020_filtered %>%
  head() %>%
  select(c(year, crosswalk_agency_name, ori, population_1, total_violent_crime_rate)) %>%
  rename("Year" = year,
         "Agency" = crosswalk_agency_name,
         "ORI" = ori,
         "Population" = population_1,
         "Rate of violent index crimes" = total_violent_crime_rate) %>%
  kable(caption = "UCR 1973-2020 data post-filter") %>%
  kable_styling %>% column_spec(5, width = "2.5cm")
```

You should be left with 77 police departments each observed over 48 years, or 3,696 observations in total.

```
# check that the filtering was done correctly
length(unique(ucr_1973_2020_filtered$ori)) # should be 77
```

Table 7: UCR 1973-2020 data post-filter

Year	Agency	ORI	Population	Rate of violent index crimes
2020	birmingham police department	AL00102	207443	417.9461
2019	birmingham police department	AL00102	208025	0.0000
2018	birmingham police department	AL00102	208933	1926.4549
2017	birmingham police department	AL00102	210383	1147.9064
2016	birmingham police department	AL00102	210758	1622.7142
2015	birmingham police department	AL00102	210515	1760.9196

```
## [1] 77
```

```
length(unique(ucr_1973_2020_filtered$year)) # should be 48
```

```
## [1] 48
```

```
nrow(ucr_1973_2020_filtered) # should be 3696
```

```
## [1] 3696
```

7)

Which three police departments saw the largest drops (in percentage terms) in their rates of reported violent index crimes from 1990 to 2010, and what were those drops?

```
# create table with differences in violent crime rate from 1990 to 2010
ucr_1990_2010 <- ucr_1973_2020_filtered %>%
  select(ori, crosswalk_agency_name, total_violent_crime_rate) %>%
  filter(year %in% c(1990, 2010)) %>%
  pivot_wider(names_from = year,
              names_prefix = "crime_",
              values_from = total_violent_crime_rate) %>%
  # calculate percent difference between total violent crime rate in 2010 and 1990
  summarize(ori,
            crosswalk_agency_name,
            crime_1990_to_2010 = 100*(crime_2010 - crime_1990)/crime_1990)

ucr_1990_2010 %>%
  arrange(crime_1990_to_2010) %>%
  head(3) %>%
  rename("Agency" = crosswalk_agency_name,
         "ORI" = ori,
         "Percent change in reported violent index crime rate from 1990 to 2010" = crime_1990_to_2010) %>%
  kable(caption = "Cities with largest drops in rates of reported violent index crimes from 1973 to 2020",
        kable_styling() %>% column_spec(3, width = "5cm"))
```

the Tampa, LA, and NYC police departments saw the largest (percent) drops in their rates of reported violent index crimes from 1990 to 2010, of 81%, 76%, and 75%, respectively.

Table 8: Cities with largest drops in rates of reported violent index crimes from 1973 to 2020

Agency	ORI	Percent change in reported violent index crime rate from 1990 to 2010
tampa police department	FL02902	-80.52214
los angeles police department	CA01942	-76.43879
new york city police department	NY03030	-75.11735

Table 9: City with largest increase in rate of reported violent index crimes from 1973 to 2020

Agency	ORI	Percent change in reported violent index crime rate from 1990 to 2010
minneapolis police department	MN02711	1096.701

8)

Which police department saw the largest increase (in percentage terms) in its rate of reported violent index crimes from 1990 to 2010? Do you suspect this increase is real or the result of a data quality issue (and if so, why)?

```
ucr_1990_2010 %>%
  arrange(desc(crime_1990_to_2010)) %>%
  head(1) %>%
  rename("Agency" = crosswalk_agency_name,
         "ORI" = ori,
         "Percent change in reported violent index crime rate from 1990 to 2010" = crime_1990_to_2010) %>%
  kable(caption = "City with largest increase in rate of reported violent index crimes from 1973 to 2020") %>%
  kable_styling() %>% column_spec(3, width = "4cm")
```

The Minneapolis police department saw the largest percent increase in its rate of reported violent index crimes from 1990 to 2010, with an increase of 1097%.

Taking a closer look at the Minneapolis police department's data:

```
# isolating just minneapolis
ucr_1973_2020_filtered %>%
  filter(ori == "MN02711", year %>% between(1985, 1995)) %>%
  select(c(year, population_1, total_violent_crime_rate)) %>%
  rename("Year" = year,
         "Population" = population_1,
         "Total rate of violent index crimes" = total_violent_crime_rate) %>%
  kable(caption = "Minneapolis crime data 1985-1995") %>%
  kable_styling() %>% column_spec(4, width = "4cm")
```

It looks like on average, the Minneapolis reports rates of around 1,000 violent index crimes per 100,000 residents, but in 1990 they reported 89. The fact that their reported violent crime rate in 1990 starkly differs from any other year suggests that this is the result of a data quality issue.

Table 10: Minneapolis crime data 1985-1995

crosswalk_agency_name	Year	Population	Total rate of violent index crimes
minneapolis police department	1995	357709	1978.14425
minneapolis police department	1994	370834	1907.59208
minneapolis police department	1993	366642	1767.66437
minneapolis police department	1992	377345	1642.79373
minneapolis police department	1991	373303	1577.53889
minneapolis police department	1990	368383	88.76631
minneapolis police department	1989	348384	1541.11555
minneapolis police department	1988	364745	1507.90278
minneapolis police department	1987	359371	1541.30411
minneapolis police department	1986	362845	1642.29905
minneapolis police department	1985	361021	1422.35493

Table 11: Minneapolis crime data from 1990 and 2010 and surrounding 5-year ranges

year	population_1	actual_murder	actual_rape_total	actual_robbery_total	actual_assault_aggravated
2012	390240	39	403	1719	1711
2011	385531	32	386	1589	1715
2010	382578	37	438	1596	1993
2009	382618	18	413	1663	2148
2008	376753	37	367	2005	2370
1992	377345	60	600	3075	2464
1991	373303	64	744	2610	2471
1990	368383	6	44	163	114
1989	348384	46	462	2524	2335
1988	364745	55	460	2565	2420

```

# looking at just 1990 and 2010
ucr_1960_2020 %>%
  filter(ori == "MN02711",
         year %in% c(1988:1992, 2008:2012)) %>%
  # can only fit so many columns on one page
  select(c(year, population_1, actual_murder, actual_rape_total, actual_robbery_total, actual_assault_aggravated))
  kable(caption = "Minneapolis crime data from 1990 and 2010 and surrounding 5-year ranges") %>%
  kable_styling() %>% column_spec(-(1), width = "2.5cm") %>% row_spec(0, angle = 40)

```

However, when investigating more closely by looking at the original UCR 1960-2020 dataset for the years surrounding 1990 and 2010, it's not automatically obvious that there is a data quality error. There seems to be no error with the reported population in 1990, so the data quality issue, if it exists, must be with the crime reporting. It seems that the numbers of index crimes reported in 1990 are around an order of 10 lower than the numbers for other years from 1998 to 1992 (and this is not the case for 2010). It's possible that one digit is left off of the reported numbers of each crime (for example, the true value of actual\_murder in

1990 could have been 60, but was incorrectly entered as 6). However, there are some fields for which that also does not seem as likely—for example, robbery. It could be that the number of reported robberies was meant to be entered as, say, 1,630, but that number also seems low compared to the other values between 1988 and 1992, which all lie between 2,500 and 3,100. The same can be said for the burglary, theft, motor vehicle theft, and assault fields.

Regardless, for one reason or another, the data in 1990 is very different from the surrounding years, and should be viewed with skepticism.

9)

Plot the rates of reported violent index crimes from 1973 to 2020 for New York and Los Angeles, including a legend to distinguish the two cities from each other.

```
ucr_1973_2020_filtered %>%
  filter(crosswalk_agency_name %in% c("new york city police department", "los angeles police department")) +
  ggplot(aes(year, total_violent_crime_rate, color = crosswalk_agency_name)) +
  geom_line() +
  scale_color_manual(values = c("new york city police department" = "cornflowerblue", "los angeles police department" = "firebrick"),
                    labels = c("New York", "Los Angeles")) +
  labs(title = "Rates of reported violent index crimes in New York and Los Angeles from 1973 to 2020 ",
       x = "Year",
       y = "Violent crime rate (per 100,000 residents)",
       color = "City") +
  theme_bw()
```

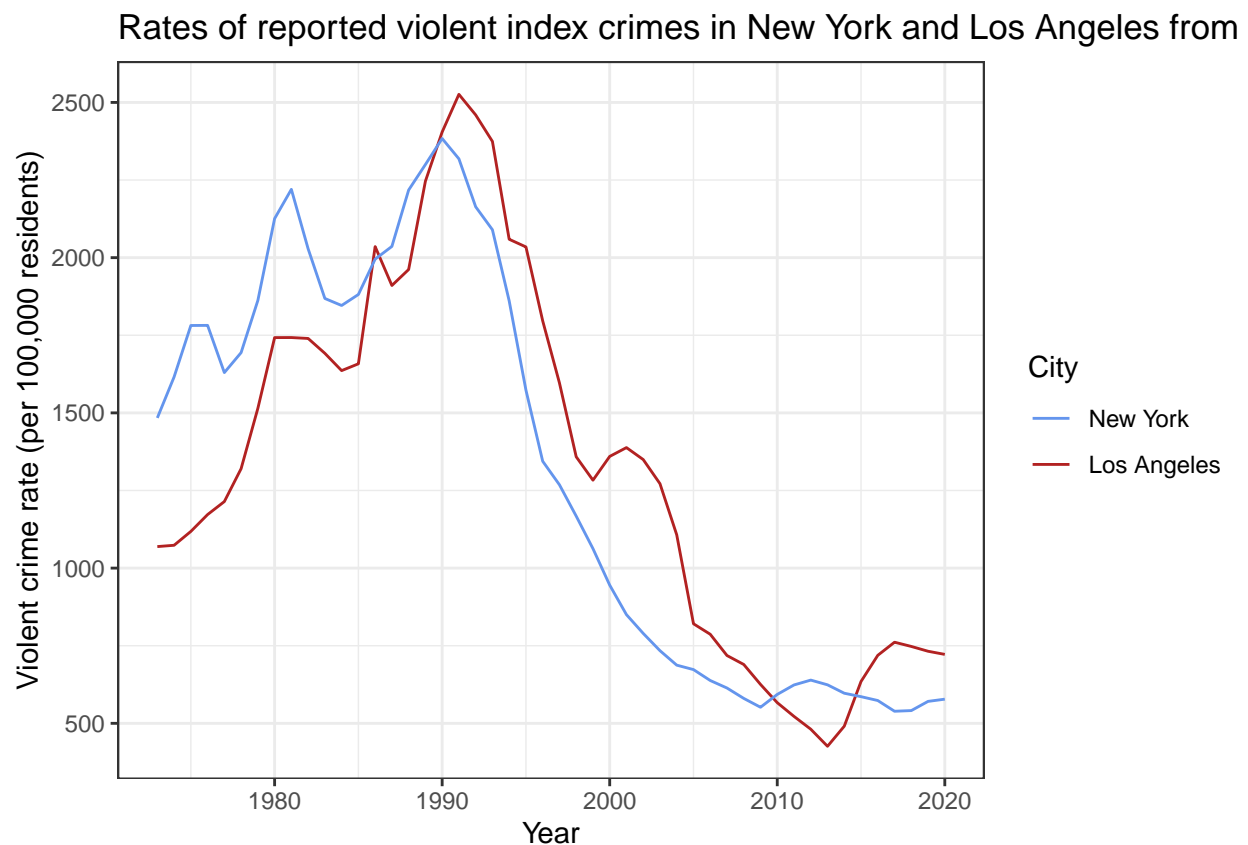


Table 12: Police departments with highest rates of reported violent index crimes over 1973-2020 in 2020

ORI	Agency	Rate of reported violent index crimes in 2020	Highest rate of violent crime from 1973-2020 in 2020
KS08703	wichita police department	1246.846	TRUE
TNMPD00	memphis police department	2356.449	TRUE
WIMPD00	milwaukee police department	1601.922	TRUE

10)

Which three police departments saw their highest rates of reported violent index crimes over this period occur in 2020? (Note: exclude the two departments that do not report any violent index crimes over this period, as this is likely a data quality issue.)

```
ucr_highest_2020 <- ucr_1973_2020_filtered %>%
  select(crosswalk_agency_name, ori, year, total_violent_crime_rate) %>%
  # create a new column for rate of violent index crime for every year
  pivot_wider(names_from = year,
              names_prefix = "crime_",
              values_from = total_violent_crime_rate) %>%
  group_by(ori) %>%
  summarize(crosswalk_agency_name,
            ori,
            crime_2020,
            # this column will be TRUE if the rate of violent index crimes in 2020 > all other years
            highest_2020 = if_all(c(crime_2019:crime_1973), ~ . < crime_2020)) %>%
  # find depts that saw highest rates of reported violent index crimes over this period in 2020
  filter(highest_2020 == TRUE)

ucr_highest_2020 %>%
  rename("ORI" = ori,
         "Agency" = crosswalk_agency_name,
         "Rate of reported violent index crimes in 2020" = crime_2020,
         "Highest rate of violent crime from 1973-2020 in 2020" = highest_2020) %>%
  kable(caption = "Police departments with highest rates of reported violent index crimes over 1973-2020",
        kable_styling() %>% column_spec(3:4, "4cm"))
```

The Wichita, Memphis, and Milwaukee police departments saw their highest rates of reported violent index crimes over the period 1973-2020 occur in 2020.

Plot their rates of reported violent index crimes from 1973 to 2020, including a legend to distinguish the three cities from each other.

```
# setting colors and labels for the three cities
colors <- c("wichita police department" = "green4",
           "memphis police department" = "firebrick",
           "milwaukee police department" = "cornflowerblue")

labels <- c("wichita police department" = "Wichita",
```

```

    "memphis police department" = "Memphis",
    "milwaukee police department" = "Milwaukee")

ucr_1973_2020_filtered %>%
  filter(ori %in% ucr_highest_2020$ori) %>% # filter for desired cities
  ggplot(aes(year, total_violent_crime_rate, color = crosswalk_agency_name)) +
  geom_line() +
  scale_color_manual(values = colors, labels = labels) +
  labs(title = "Rates of reported violent index crimes from 1973-2020 in three cities",
       x = "Year",
       y = "Rate of reported violent index crimes (per 100,000 residents)",
       color = "City") +
  theme_bw()

```

