

Info Final Project

Eric Lan
Quinton Baughman
Mohamed Camara
Gabrielle Morris
2024-05-22

Introduction

This topic interested us because we knew that the population in California was much bigger than the population in Washington and we wanted to observe the difference in crime rates between the states and counties with a considerable population discrepancy. Also as UW students a majority of the student body is made up of these two states so doing research on this topic would be relevant to us and most people in the school. We already knew the difference in population size and we also knew that California had more counties than the state of Washington. In this research assignment we will be looking at four main things which are; comparing the crime rates between California and Washington, looking at the 5 most populated counties of each state to compare in population size, the crime rates by county, and what the most prevalent violent crimes were overall. This data is important because understanding where the most violent crimes happen can help lead to solutions that can effectively address these crimes. This analysis can also identify environmental factors contributing to higher crime rates and increase the public's awareness of violent crimes in their communities, equipping residents with knowledge on how to stay safe.

Explaining our Data

We used these data sets because they came from a highly reliable source and they aligned with the topic that we wanted to conduct our research on. The data source is from DATA.GOV which is an official website of the United States government that other entities can directly report to. For the dataset we used to find violent crimes in Washington the group that reported this data was the Washington State Statistical Analysis Center. This group conducts policy-relevant research and analysis on justice issues, provides technical assistance, and maintains a clearinghouse of state justice-related data. The California dataset was collected by the California Department of Public Health. Both of these datasets are free and available to the public to use and view. In our Washington dataset it includes the year and total population of each county in Washington including both the female and male population from 1990 to 2022. Columns are represented by the type of crime committed and totals of those crimes. Unlike the California dataset this data includes violent crimes as well as non-violent crimes which needed to be filtered out. In the California data set columns are denoted by identifiers like the county and region, the name of the crime, crime rates, and population totals in the years 2000 to 2013. In this dataset the type of violent crime is in one column with each row representing a different crime committed. The sample used when merging the two data sets were from the years 2000 to 2013 where we focused on only four violent crimes those crimes being; rape, assault, murder, and robbery.

Our datasets

Washington State Crime Data: 1419 rows 228 columns

```
## [1] 1419 228
```

California State Crime Data: 49228 rows and 27 columns

```
## [1] 49227 27
```

In both Washington and California datasets, each row corresponds to the number of reported crimes of various types in a specific county during a particular year.

In Washington State, the data is clean, devoid of any missing or implausible values. For California State, we employ the pivot_wider function to organize each type of crime into columns with the respective number of reported crimes as values. We drop columns labeled 'Jurisdiction does not report' and 'NA' since they solely contain "NA" values. Rows with missing values in the 'Violent crime total' column are excluded. While some rows contain NA values in other columns(one or two columns), removing all of them would lead to significant data loss. Thus, we choose to skip them, utilizing "na.rm = TRUE" during calculations. Additionally, as the dataset includes state information, we need to avoid using these rows when computing averages to prevent skewing the data due to the inclusion of state-specific information. For Washington State, the crime categories include SRS_AG_ASSLT, SRS Rape, SRS_MURDER, SRS_ROBBERY, POP_TOTAL, while for California State, they comprise Aggravated Assault, Forcible Rape, Murder and non-negligent Manslaughter, Robbery, and Total Population.

```
## # A tibble: 6 × 9
## # Groups:   year, county [6]
##   year county POP_TOTAL SRS_AG_ASSLT SRS Rape SRS_MURDER SRS_ROBBERY
##   <dbl> <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 1990 adams      13603         80         14         3         4
## 2 1991 adams      13798         52         6         0         2
## 3 1992 adams      14143         28        13         2         2
## 4 1993 adams      14535         45         3         0         3
## 5 1994 adams      15006         57         5         0         5
## 6 1995 adams      15366         34         5         4         3
## # i 2 more variables: TOT_violent <dbl>, wa_crime_rate <dbl>
```

```
## # A tibble: 6 × 10
##   year geoname    ca_county TOT_POP `Aggravated assault` `Forcible rape`
##   <dbl> <chr>      <chr>      <dbl>          <dbl>          <dbl>
## 1  2000 California <NA>      33847694      138325          9784
## 2  2000 Alameda   Alameda   1443741       5457            568
## 3  2000 Alpine     Alpine    1208           9              1
## 4  2000 Amador     Amador    34120         153             22
## 5  2000 Butte      Butte     203171        475             77
## 6  2000 Calaveras Calaveras  40554         101             8
## # i 4 more variables: `Murder and non-negligent manslaughter` <dbl>,
## #   Robbery <dbl>, `Violent crime total` <dbl>, rate <dbl>
```

Methods

Our first step was to filter the data sets to include the information we would use for our analysis. From the Washington crime set, the data had the ideal layout with the year, county, population, and types of crimes listed in the columns, using only integer values. We selected from this set the types of crime to analyze, total population for each county, and year. We wanted to choose only the crimes that were reported from the California data, which generalized more than the Washington set. As a result we chose to keep only the Assault, Rape, Murder, and Robbery columns. We computed the rate as a ratio of the total crime value for each county divided by its population. On the other hand, the California crime set reported the values in a long format, as opposed to wide. This required us to reshape the data to match the layout of the Washington set. We reported the types of crime within the columns and selected only the columns necessary for our analysis, which were the same as for the Washington set. The data originally reported from 1990 to 2022, but we had to select the years from 2000 - 2011 due to the overlap and inconsistency in criminal reporting systems. Some of the data after 2011 was missing or would be inconsistent with what we had collected for California. Due to the large amount of data, we chose to remove columns that had missing values. We believe our data and analysis remains accurate because we removed only a fraction of values and our data remained large enough to be representative. We chose to join the two data sets by year, allowing us to create representative graphs and analyze the crime trends for both Washington and California.

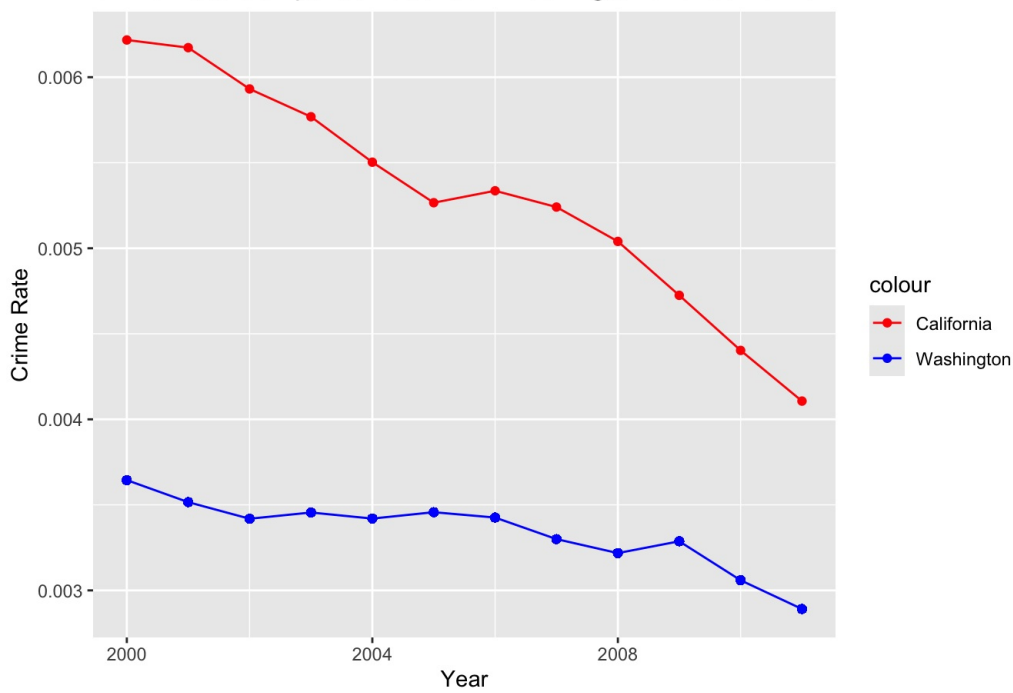
```
## # A tibble: 6 × 18
## # Groups:   year, county [1]
##   year county POP_TOTAL SRS_AG_ASSLT SRS Rape SRS_MURDER SRS_ROBBERY
##   <dbl> <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1  2000 adams    16428        21        14          0          5
## 2  2000 adams    16428        21        14          0          5
## 3  2000 adams    16428        21        14          0          5
## 4  2000 adams    16428        21        14          0          5
## 5  2000 adams    16428        21        14          0          5
## 6  2000 adams    16428        21        14          0          5
## # i 11 more variables: TOT_violent <dbl>, wa_crime_rate <dbl>, geoname <chr>,
## #   ca_county <chr>, TOT_POP <dbl>, `Aggravated assault` <dbl>,
## #   `Forcible rape` <dbl>, `Murder and non-negligent manslaughter` <dbl>,
## #   Robbery <dbl>, `Violent crime total` <dbl>, rate <dbl>
```

Findings

After filtering and creating our graphs, we found the following answers to our research questions:

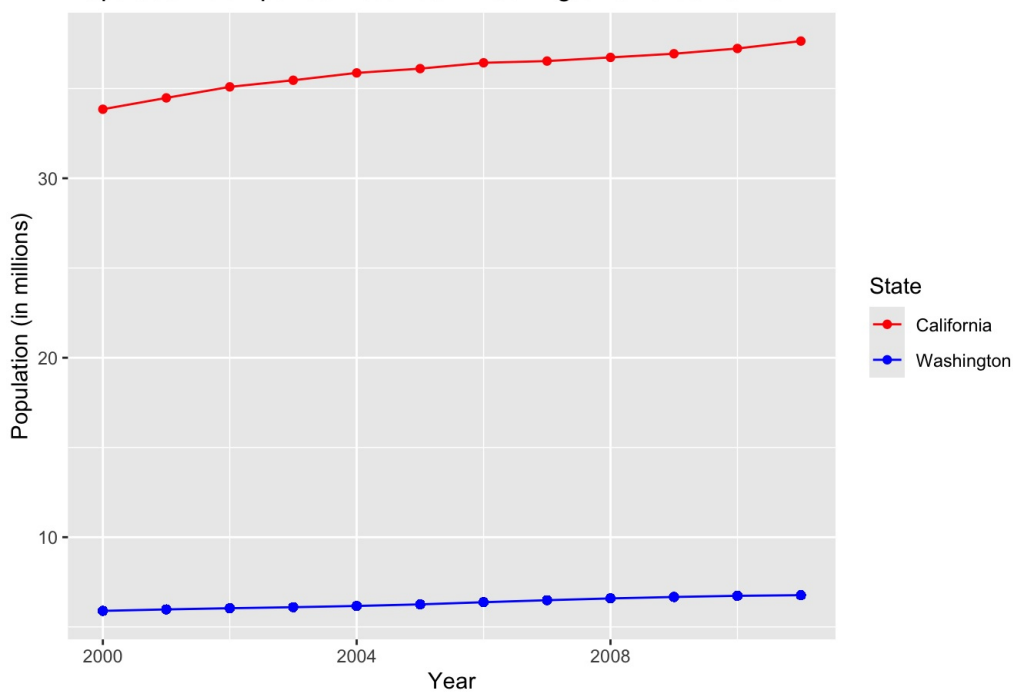
After comparing the crime rates between Washington and California, California has an overall higher crime rate

Crime Rate Comparison between Washington and California



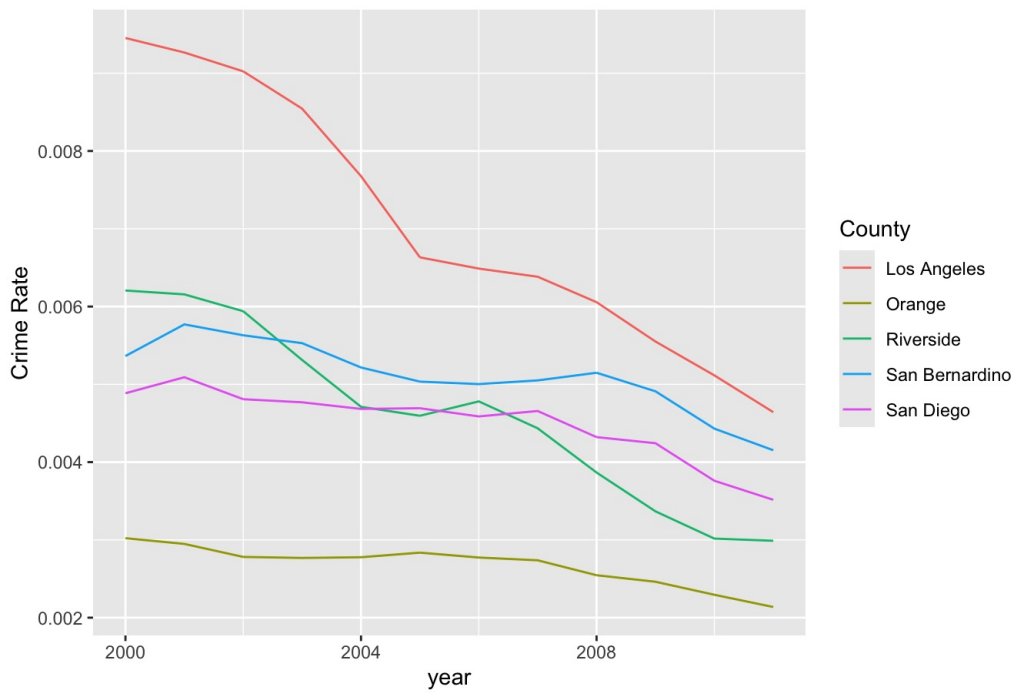
Washington and California have experienced population growth over the time from 2000 - 2013

Population Comparison between Washington and California

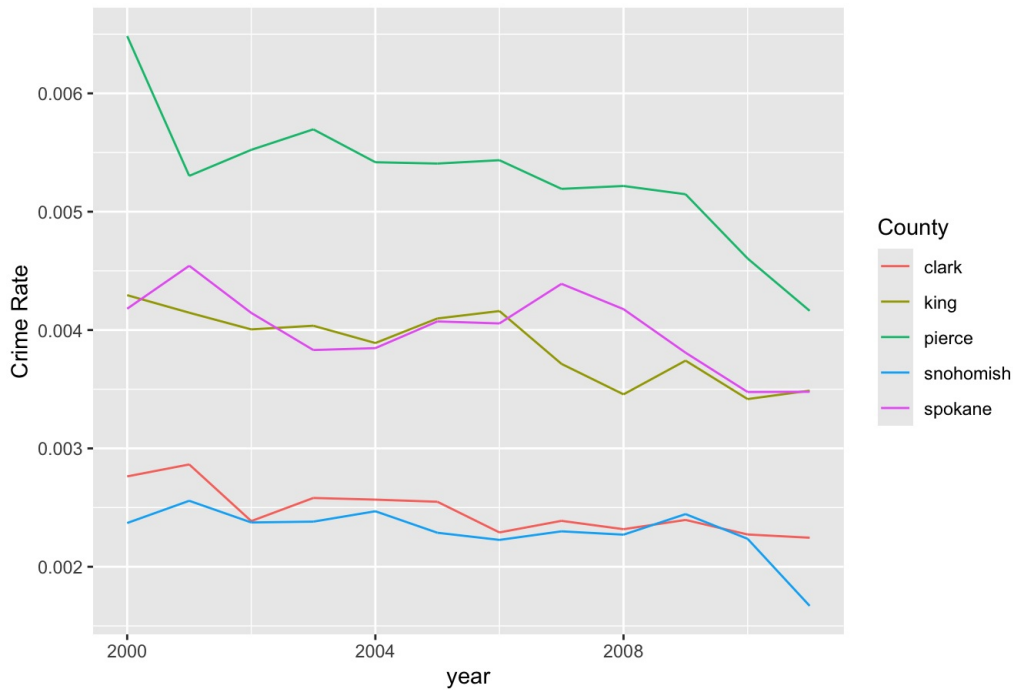


Based on the 5 most populous counties, California's counties had higher violent crime rates compared to Washington, except for Rape

Crime Rates of Top 5 Population Counties over Time in California



Crime Rates of Top 5 Population Counties over Time in Washington

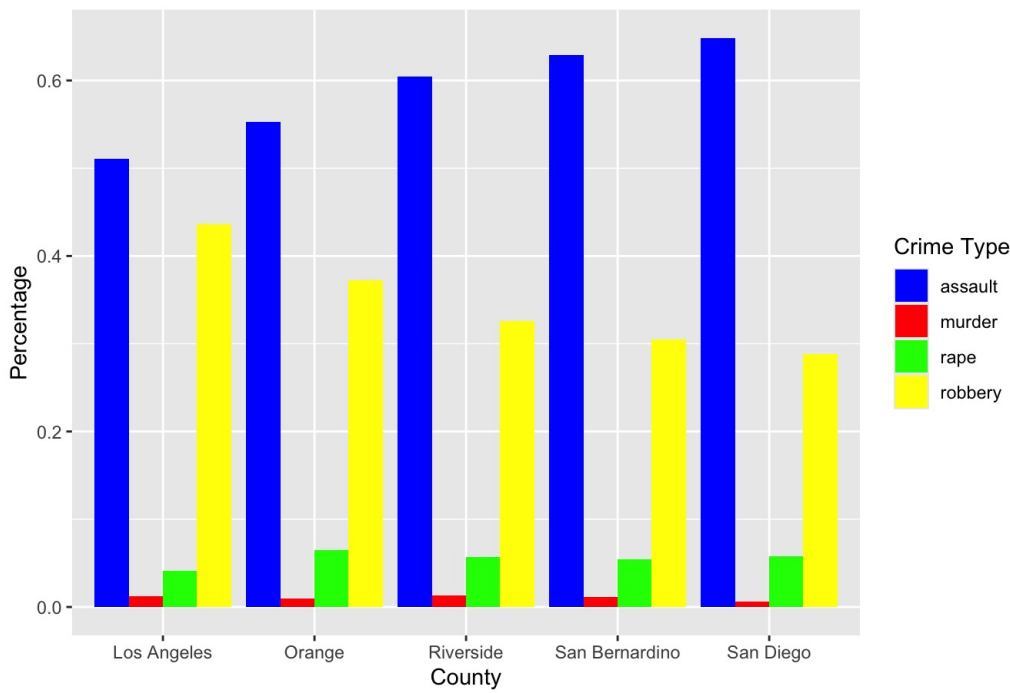


```
## # A tibble: 5 × 3
##   county      population crime_rate
##   <chr>      <dbl>      <dbl>
## 1 king      1834578.    0.00387
## 2 pierce    764539.     0.00530
## 3 snohomish 664962.     0.00230
## 4 spokane   443809.     0.00400
## 5 clark     395611.     0.00247
```

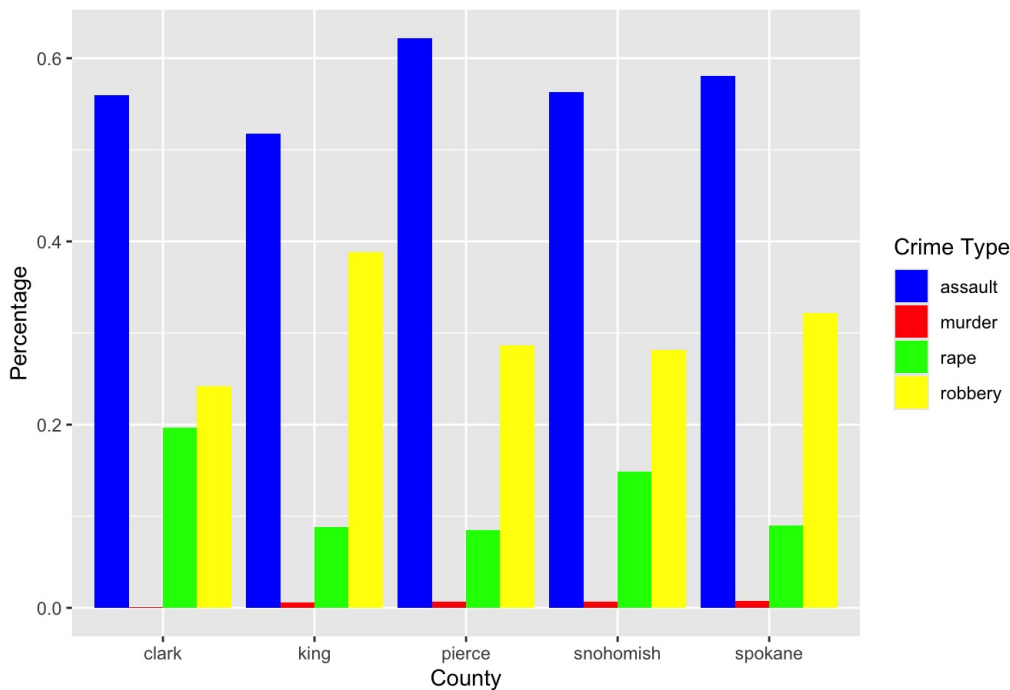
```
## # A tibble: 5 × 3
##   ca_county      population crime_rate
##   <chr>          <dbl>      <dbl>
## 1 Los Angeles   9868890.    0.00707
## 2 Orange        2979494.    0.00267
## 3 San Diego     2963590.    0.00450
## 4 San Bernardino 1921449.    0.00510
## 5 Riverside     1907132.    0.00461
```

The most prevalent crime overall is Assault

Crime Rates Across Top 5 Population Counties in California



Crime Rates Across Top 5 Population Counties in Washington



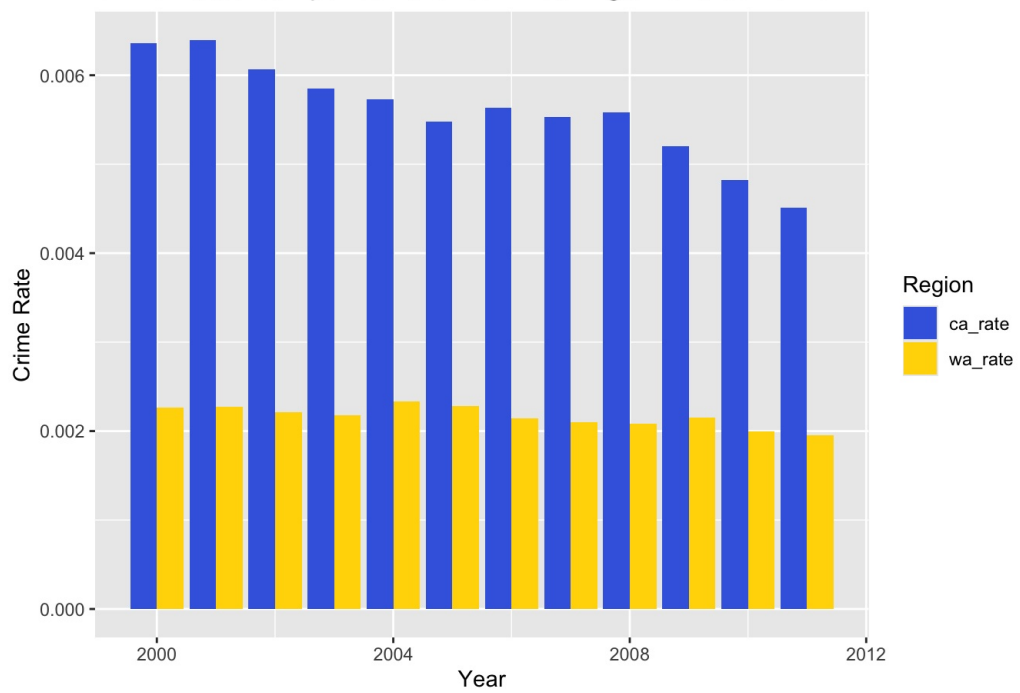
There were plenty of stories, but we found that the overall message is that California has higher violent crime rates than Washington, which could be a factor in California being a more dangerous state than Washington. One way that we could extend our story is to include all different types of crimes. Sure, violent crimes are a large factor in debating about the safety of an area, however, there are other crimes to factor in when it comes to safety. To factor in other crimes, we would need more data sets that include non-violent crimes. That way, we can measure both violent and non-violent crime rates, and compare safety on a larger scale of crime.

Summary

From a violent crime standpoint, California is a more dangerous state than Washington. California has a higher population and population growth rate, yet has higher crime rates than Washington. As UW students, we can gain a little bit of an understanding as to why a large population of California students attend universities in other states to avoid higher violent crime rates, and how safety is factored into deciding on which college to attend.

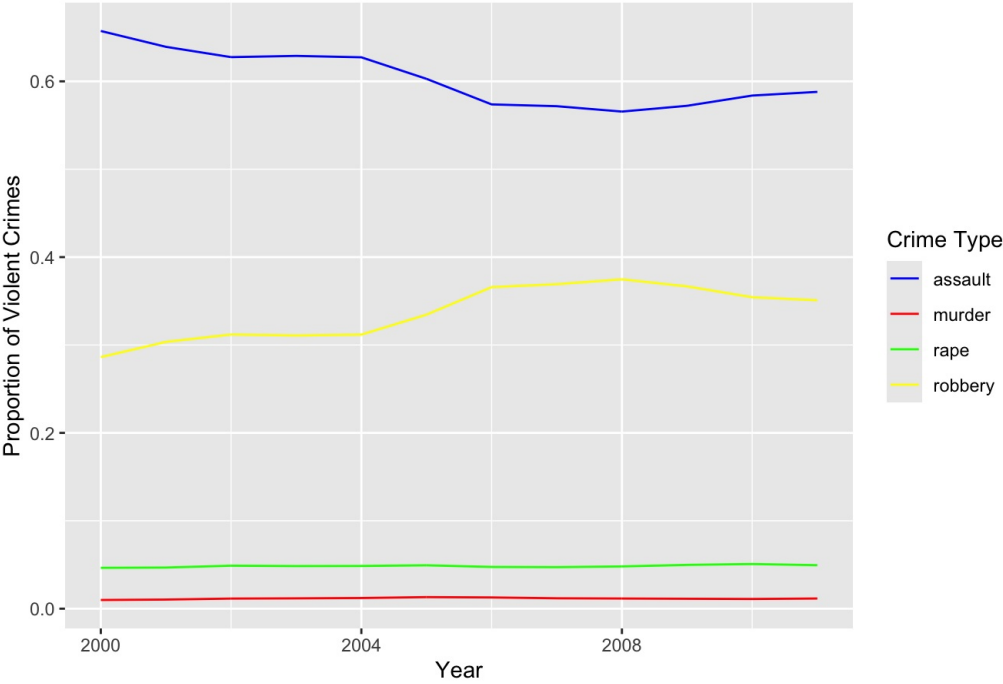
More Grpahs

Crime Rates Comparison between Washington and California



```
## # A tibble: 5 × 5
##   year assault  rape murder robbery
##   <dbl>   <dbl> <dbl>   <dbl>   <dbl>
## 1  2000    60.0  12.4  0.898    26.7
## 2  2001    58.9  12.3  0.843    27.9
## 3  2002    58.5  12.9  0.886    27.7
## 4  2003    58.9  13.2  0.864    27.1
## 5  2004    58.0  13.4  0.896    27.7
```

Proportion of Violent Crimes Over Time in California



Proportion of Violent Crimes Over Time in Washington

