Crime Rate and Unemployment Rate

Matthew E. Heino

2025-03-19

Load the required modules for the project

```
library(tidyverse)
library(raster)
library(sf)
library(ggspatial)
library(ggnewscale)
library(ggsn)
library(plotly)
```

Set the working directory

```
setwd(dirname(rstudioapi::getSourceEditorContext()$path))
```

Crime Rate and Unemployment Rate

1 Data

The data used was for this analysis is from three files. Each File contained data that needed to be joined in to a single coherent data frame that contained all the relevant information. The three files are the following:

- 1. **crime_and_incarceration_by_state.csv** - holds information about crime statistics for each state and other data. It is composed of 18 different variables. Only a few will be used in the analysis. They are listed below. Variables, violent_crime_total and state_population, will be used to calculate the crime rate. The calculation can be found in **Section**: 3.0.5. “jurisdiction” and “year” column names will be changed in **Section**: 3.0.4. Additional information about the processing can be found in subsequent sections.

  The features that will be used from the file are:

  - Variable 1: **violent_crime_total** - the total amount of violent crime in the state
  - Variable 2: **state_population** - the total population of the state
  - Variable 3: **jurisdiction** - holds the state name.
  - Variable 4: **year** - holds the year the data was recorded

- 2. **unemployment_county.csv** - This data will be processed in **Section**: 3.0.3. Additional information will be found in this section.

  The features that will be used from the file are:

  - Variable 1: **Employed** - the number of people employed in each of the counties
  - Variable 2: **Unemployed** - the number of people unemployed in each of the counties
  - Variable 3: **Unemployment Rate** - the unemployment rate for each of the counties

- 3. **tl_2019_us_state.shp** - this is the shape file that will be used in the mapping the data in **Sections**: 3.2.1 and 3.2.2

Note: All other columns will not be used and will discarded during further processing.

2 Project Objectives

The objective of this project is to determine if there is a correlation between the rate of crime and unemployment. This project seeks to see if there is an observable correlation that be seen either graphically or seen in a mathematical supported way. There will be a discussion of the findings. The data topic that will covered in this project will be the temporal-spatial changes of unemployment rate in the contiguous USA.

3 Data Processing and Data Visualization

Data preprocessing:

Steps:

- Read the data from the CSV files into individual data frames. **Section**: 3.0.1 .
- Remove the parts of the United States that are not contiguous. **Section**: 3.0.2 .
- Process the unemployment rate data. **Section**: 3.0.3
- Process the crime rate data. **Section**: 3.0.4
- Join relational tables. **Section**: 3.0.6
- Save the final combined and cleaned data. **Section**: 3.0.7

3.0.1 Read in the data from the data files.

This code will read in data from the data files that were discussed in **Section** 1

```
`File Name` <- c('crime_and_incarceration_by_state.csv',
                 'Murder Rates, States By Region_Full Data_data',
                 'tl_2019_us_state.shp')

file_df <- data.frame(`File Name`)

knitr::kable(file_df,col.names = c("File Name"), caption = "Files Used",)
```

Table 3.1: Files Used

File Name
crime_and_incarceration_by_state.csv
Murder Rates, States By Region_Full Data_data
tl_2019_us_state.shp

```
# Read in the unemployment rate from the CSV file
Unemployrate <-  read_csv("data/unemployment_county.csv")

# Read in the Crime rate from the CSV file
Crimerate <- read_csv ("data/crime_and_incarceration_by_state.csv")

# Read the states shape file
States <- st_read("data/tl_2019_us_state/tl_2019_us_state.shp")
```

```
knitr::kable(head(Unemployrate, 10), caption = "Unemployment Rate")
```

Table 3.2: Unemployment Rate

County	State	Labor Force	Employed	Unemployed	Unemployment Rate	Year
Autauga County	AL	24383	23577	806	3.3	2007
Baldwin County	AL	82659	80099	2560	3.1	2007
Barbour County	AL	10334	9684	650	6.3	2007
Bibb County	AL	8791	8432	359	4.1	2007
Blount County	AL	26629	25780	849	3.2	2007
Bullock County	AL	3653	3308	345	9.4	2007
Butler County	AL	9099	8539	560	6.2	2007
Calhoun County	AL	54861	52709	2152	3.9	2007
Chambers County	AL	15474	14469	1005	6.5	2007
Cherokee County	AL	11984	11484	500	4.2	2007

```
knitr::kable(head(Crimerate[ , 1:9], 10),
             caption = "Crime and Incarceration by State Part 1")
```

Table 3.2: Crime and Incarceration by State Part 1

jurisdiction	includes_jails	year	prisoner_count	crime_reporting_change	crimes_estimated	state_population	violent_crime_total	murder_m
FEDERAL	FALSE	2001	149852	NA	NA	NA	NA	
ALABAMA	FALSE	2001	24741	FALSE	FALSE	4468912	19582	
ALASKA	TRUE	2001	4570	FALSE	FALSE	633630	3735	
ARIZONA	FALSE	2001	27710	FALSE	FALSE	5306966	28675	
ARKANSAS	FALSE	2001	11489	FALSE	FALSE	2694698	12190	
CALIFORNIA	FALSE	2001	157142	FALSE	FALSE	34600463	212867	
COLORADO	FALSE	2001	17278	FALSE	FALSE	4430989	15492	
CONNECTICUT	TRUE	2001	17507	FALSE	FALSE	3434602	11492	
DELAWARE	TRUE	2001	6841	FALSE	FALSE	796599	4868	
FLORIDA	FALSE	2001	72404	FALSE	FALSE	16373330	130713	

```
knitr::kable(head(Crimerate[ , 10: 17], 10),
             caption = "Crime and Incarceration by State Part 2")
```

Table 3.2: Crime and Incarceration by State Part 2

rape_legacy	rape_revised	robbery	agg_assault	property_crime_total	burglary	larceny	vehicle_theft
NA	NA	NA	NA	NA	NA	NA	NA
1369	NA	5584	12250	173253	40642	119992	12619
501	NA	514	2681	23160	3847	16695	2618
1518	NA	8868	17889	293874	54821	186850	52203

rape_legacy	rape_revised	robbery	agg_assault	property_crime_total	burglary	larceny	vehicle_theft
892	NA	2181	8969	99106	22196	69590	7320
9960	NA	64614	136087	1134189	232273	697739	204177
1930	NA	3555	9849	170887	28533	121360	20994
639	NA	4183	6565	95299	17159	65762	12378
420	NA	1156	3269	27399	5144	19476	2779
6641	NA	32867	90331	782517	176052	516548	89917

```
knitr::kable(head(States, 10), caption = "States")
```

Table 3.2: States

REGION	DIVISION	STATEFP	STATENS	GEOID	STUSPS	NAME	LSAD	MTFCC	FUNCSTAT	ALAND	AWATER	INTPTLAT	INTPT
3	5	54	01779805	54	WV	West Virginia	00	G4000	A	62266231560	489271086	+38.6472854	-080.6
3	5	12	00294478	12	FL	Florida	00	G4000	A	138947364717	31362872853	+28.4574302	-082.4
2	3	17	01779784	17	IL	Illinois	00	G4000	A	143779863817	6215723896	+40.1028754	-089.1
2	4	27	00662849	27	MN	Minnesota	00	G4000	A	206230065476	18942261495	+46.3159573	-094.1
3	5	24	01714934	24	MD	Maryland	00	G4000	A	25151726296	6979340970	+38.9466584	-076.6
1	1	44	01219835	44	RI	Rhode Island	00	G4000	A	2677787140	1323663210	+41.5974187	-071.5
4	8	16	01779783	16	ID	Idaho	00	G4000	A	214049897859	2391604238	+44.3484222	-114.5
1	1	33	01779794	33	NH	New Hampshire	00	G4000	A	23189198255	1026903434	+43.6726907	-071.5
3	5	37	01027616	37	NC	North Carolina	00	G4000	A	125925929633	13463401534	+35.5397100	-079.1
1	1	50	01779802	50	VT	Vermont	00	G4000	A	23874197924	1030383955	+44.0685773	-072.6

3.0.2 Remove the parts of the United States that are not contiguous.

The states of Alaska, American Samoa, Northern Mariana Islands, Puerto Rico, US Virgin Islands, Hawaii, and Guam. The projects analysis will only focus on the contiguous United States or the mainland United States. Analysis will focus on the lower 48 states.

```
Contiguous_state <- States %>% filter(STUSPS != "AK" & STUSPS != "AS" &
                                      STUSPS != "MP" & STUSPS != "PR" &
                                      STUSPS != "VI" & STUSPS != "HI" &
                                      STUSPS != "GU")
```

3.0.3 Process the unemployment rate data

The data will be filtered to remove Alaska and Hawaii from the data set. This analysis will only focus on the contiguous United States. It is not needed so it will be removed from the data. The data will be grouped by state and then by the Year in which the data was collected. Three variables will created. These variables are the following:

- **TotalForce**: This variable will hold the total number of workers. This includes all workers both employed and unemployed.
- **Totalemployed**: This variable will hold the total number of employed workers.
- **Totalunemployed**: This variable will hold the total number of unemployed workers.
- **Meanrate**: This variable will hold the mean rate of unemployment

```
Unemployrate <- Unemployrate %>% filter(State != 'AK' & State != "HI") %>%
  group_by(State, Year) %>%
  summarise(Totalforce = sum(`Labor Force`), Totalemployed=sum(Employed),
            Totalunemployed=sum(Unemployed), Meanrate = mean(`Unemployment Rate`,
                                                             rm.na=TRUE))
```

The column in this data frame will need to have a column name changed from “State” to “STUSPS”. The years that will required will be also filtered from the data set. The years that are required for this project were from 2007 to 2014.

```
Unemployrate <- Unemployrate %>% rename("STUSPS" = "State") %>%
  filter(Year %in% c(2007:2014))
```

3.0.4 Process the Crime rate

In this step the crime rate will need to have two columns renamed using the **rename()** function. The two columns are jurisdiction and the year columns. The “jurisdiction” column will be changed to “STUSPS”. This will aid joining the frames in a later step. Changing “year” to “Year” will help keep the naming convention consistent among the data frames that are to be used in the final

project.

```
Crimerate <- Crimerate %>%
  rename("STUSPS" = "jurisdiction") %>%
  rename("Year" = "year") %>%
  filter(STUSPS != "FEDERAL" & STUSPS != "ALASKA" & STUSPS != "HAWAII") %>%
  filter(Year %in% c(2007:2014))
```

There will be a need to change the state names in the STUSPS column.

```
Crimerate$STUSPS <- state.abb[match(str_to_title(Crimerate$STUSPS), state.name)]
```

3.0.5 Calculate the crime rate

The crime rate was calculated using two columns from the Crimerate data frame. The columns were:

- **violent_crime_total**: the total number of violent crime in the state
- **state_population**: the population of the state

```
Crimerate <- Crimerate %>%
  mutate(Crimerate=(violent_crime_total/state_population) * 100) %>%
  dplyr::mutate_if(is.numeric, round, 1)
```

3.0.6 Join relational tables

The data frames will be joined so all the data will be contained in one frame. Only unique columns will be included within the final data frame. From the joined data frames select columns that are relevant for final use in the creation of the final project.

```
CS_Erate <- right_join(Contiguous_state, Unemployrate, by= c("STUSPS"))

CS_Erate_Crate <- right_join(CS_Erate, Crimerate, by= c("STUSPS", "Year"))

CS_Erate_Crate1 <- CS_Erate_Crate %>%
  select(REGION, STUSPS, NAME, Year, Meanrate,Crimerate) %>%
  rename("Unemplyrate"="Meanrate")
```

```
knitr::kable(head(CS_Erate_Crate1, 10), caption = "Combined Data")
```

Table 3.3: Combined Data

REGION	STUSPS	NAME	Year	Unemplyrate	Crimerate	geometry
3	WV	West Virginia	2007	5.138182	0.3	MULTIPOLYGON (((-81.74725 3…
3	WV	West Virginia	2008	4.914546	0.3	MULTIPOLYGON (((-81.74725 3…
3	WV	West Virginia	2009	8.801818	0.3	MULTIPOLYGON (((-81.74725 3…
3	WV	West Virginia	2010	9.740000	0.3	MULTIPOLYGON (((-81.74725 3…
3	WV	West Virginia	2011	8.985454	0.3	MULTIPOLYGON (((-81.74725 3…
3	WV	West Virginia	2012	8.443636	0.3	MULTIPOLYGON (((-81.74725 3…
3	WV	West Virginia	2013	7.716364	0.3	MULTIPOLYGON (((-81.74725 3…
3	WV	West Virginia	2014	7.532727	0.3	MULTIPOLYGON (((-81.74725 3…
3	FL	Florida	2007	4.186567	0.7	MULTIPOLYGON (((-86.38865 3…
3	FL	Florida	2008	6.473134	0.7	MULTIPOLYGON (((-86.38865 3…

3.0.7 Save the final combined and cleaned data.

```
saveRDS(CS_Erate_Crate1, file = "CS_Erate_CrateCombined1.Rds")
```

3.1 EDA analysis

```
# Create a copy of the data frame
stats_df <- data.frame(CS_Erate_Crate1) %>% select(-geometry)

region_unemploy <- stats_df %>%
group_by(REGION) %>%
  summarise(
    `Region Mean` = mean(Unemplyrate),
    `Maximum Unemployment Rate` = max(Unemplyrate),
    `Minimum Unemployment Rate` = min(Unemplyrate),
    `Quantiles Unemployment` = list(round(quantile(Unemplyrate, type=1), 2)),
    `Standard Deviation` = sd(Unemplyrate),
  )

knitr::kable(region_unemploy, caption = "Regional Unemployment Statistics.", align = "cccc", digits = 2)
```

Table 3.4: Regional Unemployment Statistics.

REGION	Region Mean	Maximum Unemployment Rate	Minimum Unemployment Rate	Quantiles Unemployment	Standard Deviation
1	7.02	10.54	3.50	3.50, 5.40, 7.19, 8.61, 10.54	1.83
2	6.52	14.12	2.87	2.87, 4.39, 5.97, 8.18, 14.12	2.53

REGION	Region Mean	Maximum Unemployment Rate	Minimum Unemployment Rate	Quantiles Unemployment	Standard Deviation
3	8.04	13.27	3.43	3.43, 6.46, 7.75, 9.31, 13.27	2.33
4	7.69	13.81	2.92	2.92, 5.43, 7.58, 9.51, 13.81	2.71

```
region_unemploy_box <- stats_df %>%
group_by(REGION) %>% ggplot(mapping=aes(x=REGION, y=Unemplyrate, fill=REGION))+
  geom_boxplot()+
  labs(colour="Year", y="Unemployment Rate", x="Region",
       title="Unemployment Rate by Region") +
  theme(panel.background = element_blank(), text=element_text(size=16),
        plot.title=element_text(hjust=0.5, size=20))

ggplotly(region_unemploy_box)
```

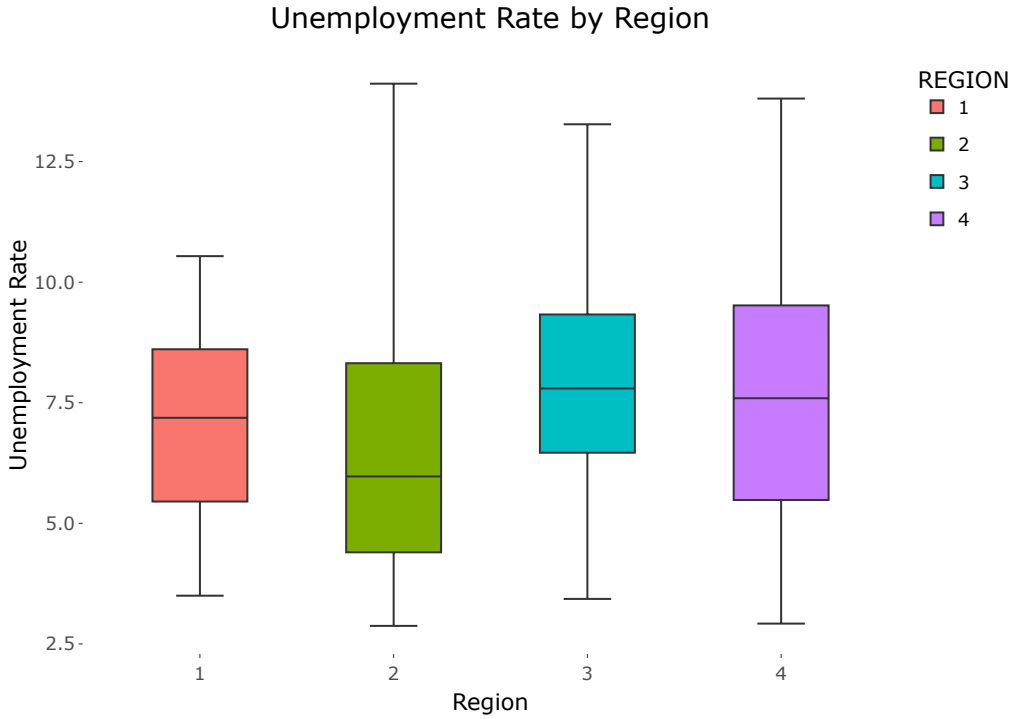


Figure 3.1: Unemployment Rate by Region.

```
region_year_unemploy <-stats_df %>%
group_by(Year, REGION) %>%
  summarise(
    `Region Mean` = mean(Unemplyrate),
    `Maximum Unemployment Rate` = max(Unemplyrate),
    `Minimum Unemployment Rate` = min(Unemplyrate),
    `Quantiles: 0%  25%  50%  75%  100%` =
      list(round(quantile(Unemplyrate, type=1), 2)),
    `Standard Deviation` = sd(Unemplyrate),
  )

knitr::kable(region_year_unemploy, caption = "Regional Unemployment Statistics by Year and Region.",
             align = "cccc", digits = 2)
```

Table 3.5: Regional Unemployment Statistics by Year and Region.

Year	REGION	Region Mean	Maximum Unemployment Rate	Minimum Unemployment Rate	Quantiles: 0% 25% 50% 75% 100%	Standard Deviation
2007	1	4.52	5.34	3.50	3.50, 4.39, 4.46, 4.72, 5.34	0.49
2007	2	4.79	8.05	2.87	2.87, 3.73, 4.71, 5.33, 8.05	1.39
2007	3	5.03	7.15	3.43	3.43, 4.12, 5.00, 5.49, 7.15	1.13
2007	4	4.48	6.78	2.92	2.92, 3.43, 4.05, 5.69, 6.78	1.32
2008	1	5.58	7.20	3.84	3.84, 5.38, 5.54, 5.77, 7.20	0.89
2008	2	5.43	8.90	3.21	3.21, 3.60, 5.41, 6.26, 8.90	1.71
2008	3	6.10	8.38	3.69	3.69, 4.79, 6.21, 6.98, 8.38	1.38
2008	4	5.81	8.64	3.15	3.15, 4.60, 5.37, 7.22, 8.64	1.73
2009	1	8.27	10.22	6.17	6.17, 7.74, 8.22, 8.97, 10.22	1.21
2009	2	8.29	14.12	4.26	4.26, 5.17, 8.03, 9.95, 14.12	3.12
2009	3	9.80	13.27	6.48	6.48, 8.00, 8.80, 11.25, 13.27	2.18

Year	REGION	Region Mean	Maximum Unemployment Rate	Minimum Unemployment Rate	Quantiles: 0% 25% 50% 75% 100%	Standard Deviation
2009	4	9.08	12.93	6.31	6.31, 6.49, 8.79, 11.67, 12.93	2.43
2010	1	8.55	10.54	5.82	5.82, 8.60, 8.78, 8.93, 10.54	1.45
2010	2	8.19	13.33	3.96	3.96, 5.25, 7.69, 10.16, 13.33	3.05
2010	3	10.20	13.15	7.16	7.16, 8.50, 9.74, 11.63, 13.15	1.82
2010	4	9.92	13.81	6.16	6.16, 8.48, 9.51, 12.09, 13.81	2.51
2011	1	8.13	10.40	5.38	5.38, 7.73, 8.42, 8.61, 10.40	1.59
2011	2	7.35	11.37	3.76	3.76, 5.21, 6.86, 9.09, 11.37	2.48
2011	3	9.60	12.58	6.20	6.20, 7.75, 9.31, 11.22, 12.58	1.81
2011	4	9.39	13.43	5.62	5.62, 7.53, 8.97, 11.59, 13.43	2.46
2012	1	7.85	9.79	5.40	5.40, 7.16, 8.18, 8.64, 9.79	1.58
2012	2	6.54	10.07	3.53	3.53, 4.85, 5.96, 8.07, 10.07	2.12
2012	3	8.62	11.12	5.52	5.52, 7.37, 8.53, 9.36, 11.12	1.55
2012	4	8.52	12.14	5.19	5.19, 6.29, 7.86, 10.23, 12.14	2.29
2013	1	7.19	8.70	4.90	4.90, 7.19, 7.68, 7.70, 8.70	1.40
2013	2	6.30	9.93	3.48	3.48, 4.41, 5.48, 7.75, 9.93	2.16
2013	3	8.03	10.13	5.65	5.65, 6.87, 7.72, 9.13, 10.13	1.29
2013	4	7.72	10.73	4.67	4.67, 5.53, 7.61, 9.06, 10.73	2.05
2014	1	6.03	7.22	4.19	4.19, 6.10, 6.19, 6.47, 7.22	1.07
2014	2	5.27	8.18	3.12	3.12, 3.99, 4.76, 6.21, 8.18	1.59
2014	3	6.95	8.91	4.81	4.81, 6.02, 7.07, 7.64, 8.91	1.11
2014	4	6.59	9.45	4.10	4.10, 4.74, 7.28, 7.81, 9.45	1.83

```
region_year_unemploy_box <-
 ggplot(stats_df) + geom_boxplot(aes(x=REGION, y=Unemplyrate, fill=REGION)) +
  facet_wrap(~Year, ncol=2) +
  labs(colour="Year", y="Unemployment Rate", x="Region",
       title="Unemployment Rate by Year and Region") +
  theme_classic() +
  theme(plot.title = element_text(hjust = 0.5, size=20),
        text=element_text(size=16))

ggplotly(region_year_unemploy_box)
```

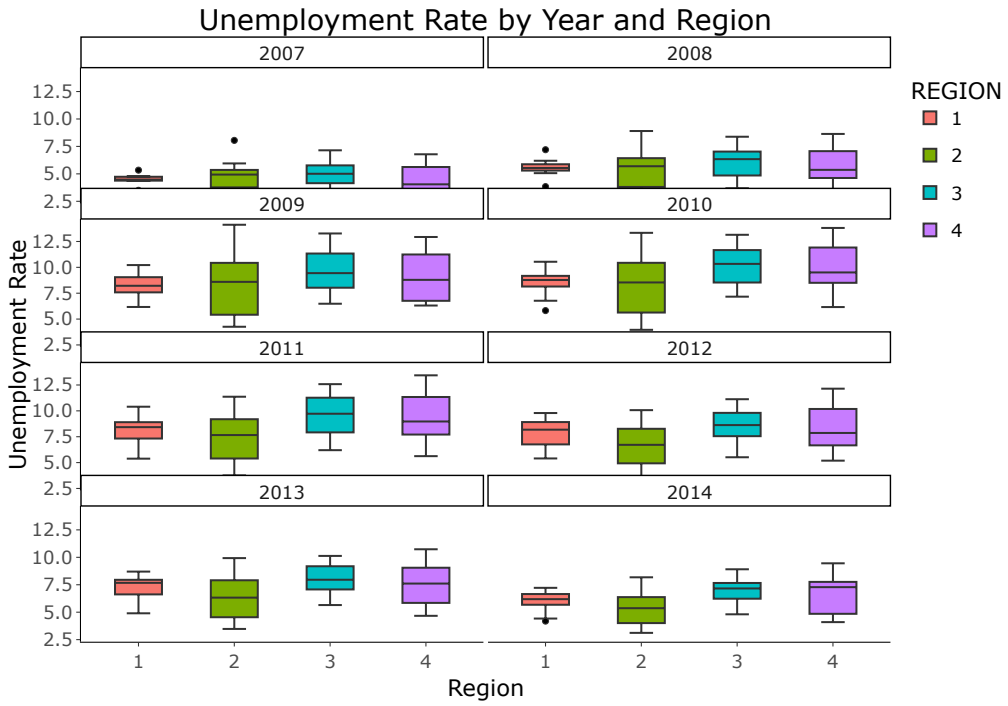


Figure 3.2: Unemployment Rate by Year and Region.

```
region_crime <- stats_df %>%
group_by(REGION) %>%
  summarise(
    `Region Mean` = mean(Crimerate),
    `Maximum Crime Rate` = max(Crimerate),
    `Minimum Crime Rate` = min(Crimerate),
    `Quantiles: 0%  25%  50%  75%  100%` = list(quantile(Crimerate, type=1)),
    `Standard Deviation` = sd(Crimerate)
  )

knitr::kable(region_crime, caption = "Regional Crime Statistics",
             align = "cccc", digits = 2)
```

Table 3.6: Regional Crime Statistics

REGION	Region Mean	Maximum Crime Rate	Minimum Crime Rate	Quantiles: 0% 25% 50% 75% 100%	Standard Deviation
1	0.27	0.5	0.1	0.1, 0.2, 0.3, 0.4, 0.5	0.12
2	0.34	0.6	0.2	0.2, 0.3, 0.3, 0.4, 0.6	0.10
3	0.45	0.8	0.2	0.2, 0.3, 0.5, 0.5, 0.8	0.15
4	0.36	0.8	0.2	0.2, 0.2, 0.3, 0.4, 0.8	0.16

```
region_crime_box <- stats_df %>%
group_by(REGION) %>% ggplot(mapping=aes(x=REGION, y=Crimerate, fill=REGION))+
  geom_boxplot() +
   labs(colour="Year", y="Crime Rate", x="Region",
       title="Crime Rate by Region") +
  theme(panel.background = element_blank(),
        plot.title=element_text(hjust=0.5, size=20), text=element_text(size=16))

ggplotly(region_crime_box)
```

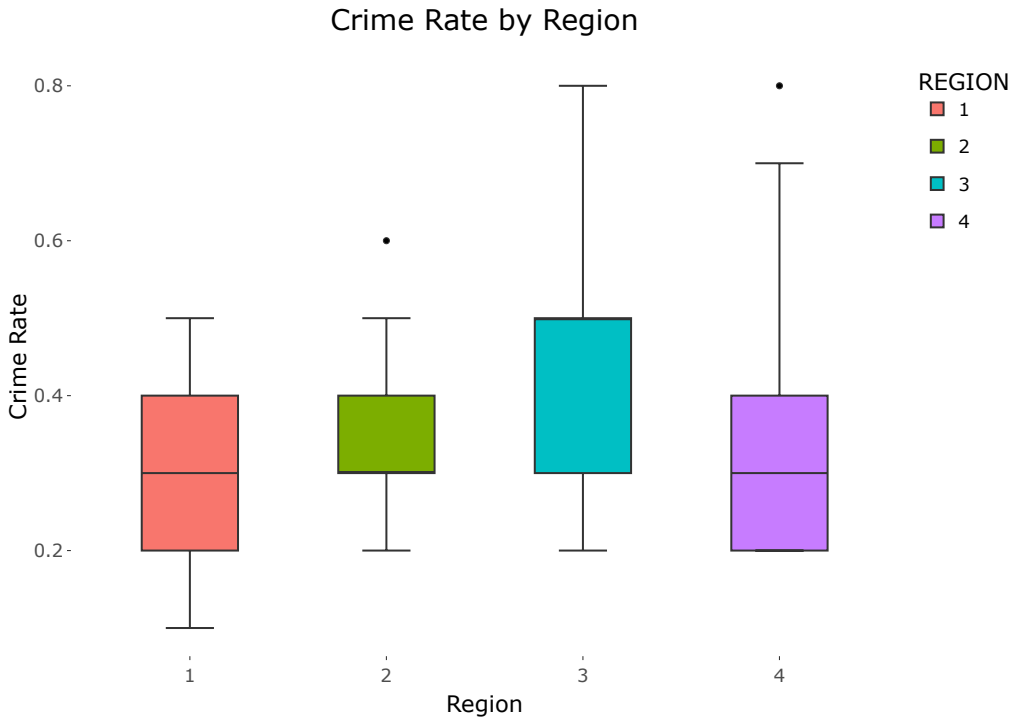


Figure 3.3: Crime Rate by Region.

```
region_year_crime <-stats_df %>%
group_by(Year, REGION) %>%
  summarise(
    `Region Mean` = mean(Crimerate),
    `Maximum Crime Rate` = max(Crimerate),
    `Minimum Crime Rate` = min(Crimerate),
    `Quantiles: 0%  25%  50%  75%  100%` = list(quantile(Crimerate, type=1)),
    `Standard Deviation` = sd(Crimerate),
  )

knitr::kable(region_year_crime, caption = "Regional Crime Statistics by Year and Region.",
             align = "cccc", digits = 2)
```

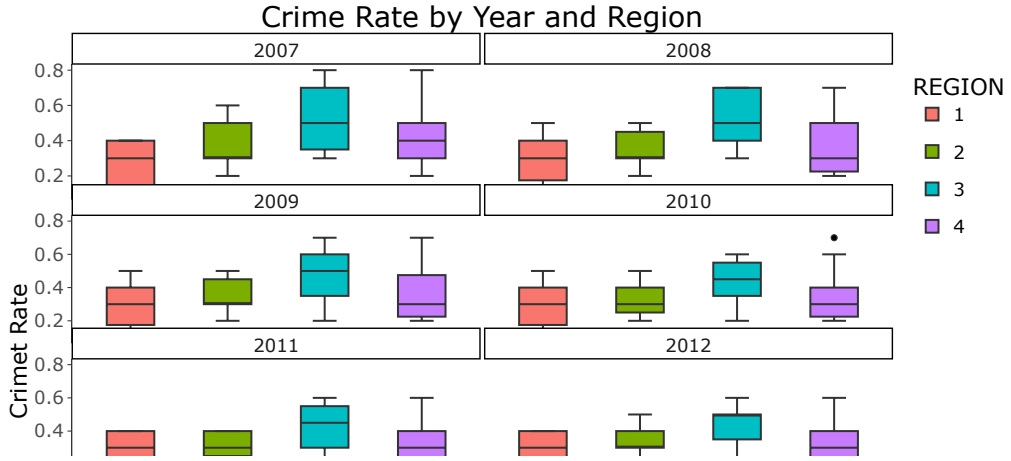
Table 3.7: Regional Crime Statistics by Year and Region.

Year	REGION	Region Mean	Maximum Crime Rate	Minimum Crime Rate	Quantiles: 0% 25% 50% 75% 100%	Standard Deviation

Year	REGION	Region Mean	Maximum Crime Rate	Minimum Crime Rate	Quantiles: 0% 25% 50% 75% 100%	Standard Deviation
2007	1	0.26	0.4	0.1	0.1, 0.1, 0.3, 0.4, 0.4	0.13
2007	2	0.37	0.6	0.2	0.2, 0.3, 0.3, 0.5, 0.6	0.13
2007	3	0.52	0.8	0.3	0.3, 0.3, 0.5, 0.7, 0.8	0.18
2007	4	0.43	0.8	0.2	0.2, 0.3, 0.4, 0.5, 0.8	0.18
2008	1	0.29	0.5	0.1	0.1, 0.2, 0.3, 0.4, 0.5	0.14
2008	2	0.36	0.5	0.2	0.2, 0.3, 0.3, 0.4, 0.5	0.10
2008	3	0.52	0.7	0.3	0.3, 0.3, 0.5, 0.7, 0.7	0.16
2008	4	0.39	0.7	0.2	0.2, 0.2, 0.3, 0.5, 0.7	0.19
2009	1	0.29	0.5	0.1	0.1, 0.2, 0.3, 0.4, 0.5	0.14
2009	2	0.34	0.5	0.2	0.2, 0.3, 0.3, 0.4, 0.5	0.11
2009	3	0.48	0.7	0.2	0.2, 0.3, 0.5, 0.6, 0.7	0.15
2009	4	0.36	0.7	0.2	0.2, 0.2, 0.3, 0.5, 0.7	0.17
2010	1	0.29	0.5	0.1	0.1, 0.2, 0.3, 0.4, 0.5	0.14
2010	2	0.32	0.5	0.2	0.2, 0.2, 0.3, 0.4, 0.5	0.11
2010	3	0.44	0.6	0.2	0.2, 0.3, 0.4, 0.5, 0.6	0.14
2010	4	0.35	0.7	0.2	0.2, 0.2, 0.3, 0.4, 0.7	0.16
2011	1	0.27	0.4	0.1	0.1, 0.2, 0.3, 0.4, 0.4	0.12
2011	2	0.31	0.4	0.2	0.2, 0.2, 0.3, 0.4, 0.4	0.08
2011	3	0.43	0.6	0.2	0.2, 0.3, 0.4, 0.5, 0.6	0.14
2011	4	0.34	0.6	0.2	0.2, 0.2, 0.3, 0.4, 0.6	0.15
2012	1	0.28	0.4	0.1	0.1, 0.2, 0.3, 0.4, 0.4	0.12
2012	2	0.33	0.5	0.2	0.2, 0.3, 0.3, 0.4, 0.5	0.10
2012	3	0.44	0.6	0.2	0.2, 0.3, 0.5, 0.5, 0.6	0.13
2012	4	0.34	0.6	0.2	0.2, 0.2, 0.3, 0.4, 0.6	0.15
2013	1	0.27	0.4	0.1	0.1, 0.2, 0.3, 0.3, 0.4	0.11
2013	2	0.33	0.5	0.2	0.2, 0.3, 0.3, 0.4, 0.5	0.08
2013	3	0.41	0.6	0.2	0.2, 0.3, 0.4, 0.5, 0.6	0.12
2013	4	0.34	0.6	0.2	0.2, 0.2, 0.3, 0.4, 0.6	0.15
2014	1	0.24	0.4	0.1	0.1, 0.2, 0.2, 0.3, 0.4	0.11
2014	2	0.32	0.4	0.2	0.2, 0.3, 0.3, 0.4, 0.4	0.06
2014	3	0.40	0.6	0.2	0.2, 0.3, 0.4, 0.5, 0.6	0.12
2014	4	0.34	0.6	0.2	0.2, 0.2, 0.3, 0.4, 0.6	0.15

```
region_year_crime_box <-
 ggplot(stats_df) + geom_boxplot(aes(x=REGION, y=Crimerate, fill=REGION)) +
  facet_wrap(~Year, ncol=2) +
  labs(colour="Year", y="Crimet Rate", x="Region",
       title="Crime Rate by Year and Region") +
  theme_classic() +
  theme(plot.title = element_text(hjust = 0.5, size=20),
        text=element_text(size=16))

ggplotly(region_year_crime_box)
```



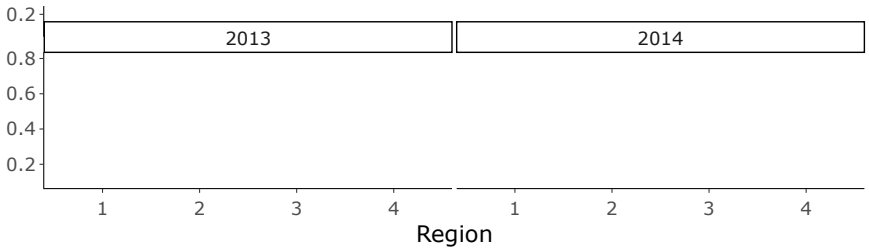


Figure 3.4: Crime Rate by Year and Region.

3.2 Data analytics method

The data visualizations that were produced for the project were the following:

- A spatial map over the contiguous USA for the unemployment rate for the specific year 2014.
- A spatial map over the contiguous USA for the crime rate for the specific year 2014.
- Scatter plot for the data relationship between the unemployment rate and crime rate.
- Time series plot for the four states for the unemployment rate
- Time series plot for the four states for the crime rate

Data for the creation of the graphs is loaded from the RDS file that was created in a previous section of the project. The file is a ".Rds" the name of the file is:

- **CS_Erate_CrateCombined1.Rds**

This file will read in using the **readRDS()**. The data found in this will then be used to create the plots that are found in this section of the project.

Read the cleaned data from the ".Rds" file.

```
all_info_from_RDS <- readRDS("CS_Erate_CrateCombined1.Rds")
```

3.2.1 A spatial map over the contiguous USA for the unemployment rate for the specific year 2014.

This is a map of the unemployment rate for the year 2014. This will be an interactive plot using the **plot_ly** function to create it.

The only year that will plotted on this time series plot will be for the year 2014. This data will be filtered from the **all_info_from_RDS**.

Note: This step could have been done using a pipe, but this makes it easier to see what is going on.

```
info_for_year_2014 <- all_info_from_RDS %>% filter(all_info_from_RDS$Year == 2014)
```

Using the **info_for_year_2014** data frame a graph of the contiguous United States will be created showing unemployment rate as a layer on the graph.

```
sp1 <- ggplot(data=info_for_year_2014) +
  geom_sf(data= info_for_year_2014$geometry,
          aes(fill=info_for_year_2014$Unemplyrate,
              text=paste("State: ",info_for_year_2014$NAME ,
                         "\nUnemployment Rate: ",
                         round(info_for_year_2014$Unemplyrate, 2 )))) +
  xlab("Longitude") +
  ylab("Latitude") +
  guides(fill=guide_legend(title= "Unemployment Rate for 2014")) +
  labs(title = "Unemployment Rate Over Contiguous USA ",
       subtitle = "Unemployment Color Coded by State",
       caption = "Data source: Unknown") +
  scalebar(data= info_for_year_2014, location="bottomleft", dist= 500, st.size=2,
           dist_unit = "km", transform= TRUE, model= "WGS84", st.dist=0.04) +
  annotation_north_arrow(location = "br", which_north = "true",
                         style = north_arrow_fancy_orienteering) +
  theme(panel.background = element_blank(), legend.position = "right",
        plot.title = element_text(hjust = 0.5, size=20),
        plot.subtitle = element_text(hjust = 0.5, size=16),
        text=element_text(size=16))

sp1
```

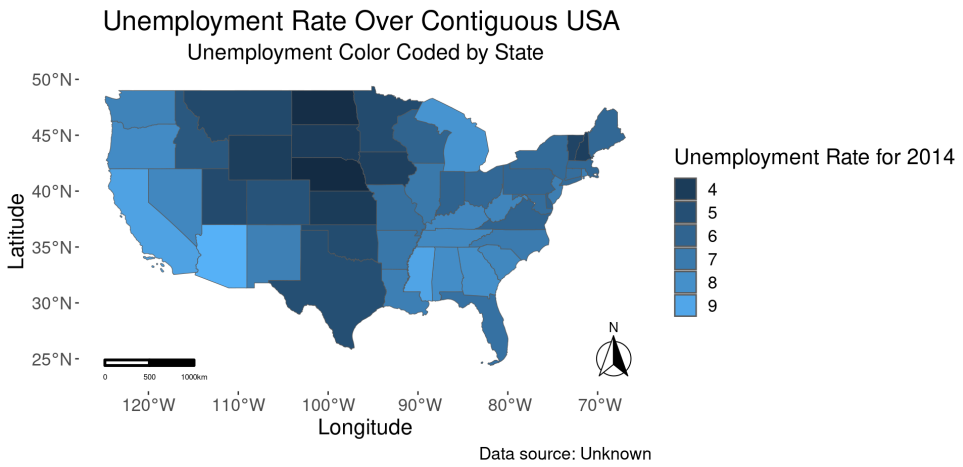


Figure 3.5: A spatial map over the contiguous USA for the unemployment rate for the year 2014

3.2.2 A spatial map over the contiguous USA for the crime rate for the specific year 2014.

Using the **info_for_year_2014** data frame a graph of the contiguous United States will be created showing crime rate as a layer on the graph.

```
ggplot(data=info_for_year_2014) +
  geom_sf(data= info_for_year_2014$geometry,
          aes(fill=info_for_year_2014$Crimerate)) +
  xlab("Longitude") +
  ylab("Latitude") +
  guides(fill=guide_legend(title= "Crime Rate for 2014")) +
  labs(title = "Crime Rate Over Contiguous USA ",
       subtitle = "Crime Rate Color Coded by State",
       caption = "Data source: Unknown") +
  scalebar(data= info_for_year_2014, location="bottomleft", dist= 500, st.size=2,
           dist_unit = "km", transform= TRUE, model= "WGS84", st.dist=0.04) +
  annotation_north_arrow(location = "br", which_north = "true",
                         style = north_arrow_fancy_orienteering) +
  theme(panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5, size=20),
        plot.subtitle = element_text(hjust = 0.5, size=16),
        text=element_text(size=16))
```

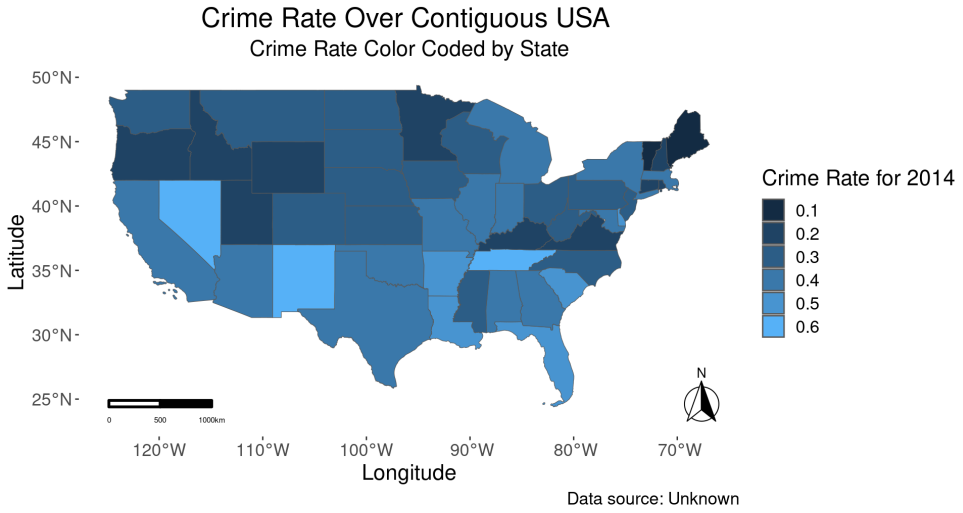


Figure 3.6: Spatial map over the contiguous USA for the crime rate for the year 2014

Scatter plot for the data relationship between the unemployment rate and crime rate.

Creates a scatter plot using crime rate (x-axis) and unemployment rate (y-axis).

```
fig <- plot_ly(data= info_for_year_2014, x= ~Crimerate, y= ~Unemplyrate,
               color= ~REGION) %>%
  add_markers() %>%
  layout(title="<b>Unemployment Rate and Crime Rate for 2014 </b>",
         margin=list(b = 10, l= 10)) %>%
  layout(xaxis=list(title= "<b>Crime Rate Per 100,000 People</b>"),
         yaxis=list(title="<b>Unemployment Rate Per 100 People </b>"),
         legend=list(title=list(text='<b> Region </b>'),
                     showlegend=TRUE)) %>%
  layout(xaxis=list(titlefont= list(size= 14)),
         yaxis=list(titlefont= list(size= 14)))


fig
```

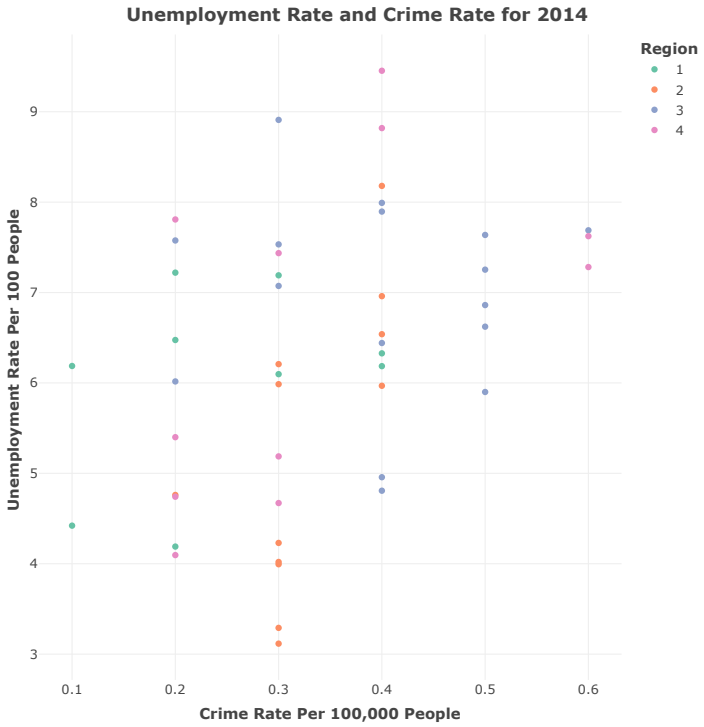


Figure 3.7: Scatter plot for the data relationship between the unemployment rate and crime rate

4) Time series plot for the four states for the unemployment rate.

This will be an interactive plot of the unemployment rate for four states:

- California
- Idaho
- Illinois
- Indiana

Steps to create the time series plot:

- Data will be filtered from the **all_info_from_RDS** data frame and a new data frame will be created. **Section**: 3.3
- The new data frame created is four_states_year_2014. **Section**: 3.3
- Create the unemployment rate time series plot. **Section**: 3.4
- Create the crime rate time series plot.**Section**: 3.5

Section 3.3 data filtered from the **all_info_from_RDS** data frame and a new data frame will be created. A vector of states was created to form the list of states that were to plotted on the graph. These states will be used for this time series plot and the one that follows.

3.3 Data will be filtered from the all_info_from_RDS data frame and a new data frame will be created.

```
states <- c("California", "Idaho", "Illinois", "Indiana")
four_states_year_2014 <- all_info_from_RDS %>% filter(NAME %in% states)

stats_df <-  as.data.frame(four_states_year_2014)
```

3.4 Create the unemployment rate time series plot.

```
une <- plot_ly(data=stats_df, x= ~as.factor(Year), y= ~Unemplyrate,color= ~NAME) %>%
  filter(NAME %in% states) %>%
  group_by(NAME) %>%
  add_lines() %>%
  layout(title="<b>Unemployment Rate Changes by Year</b>",
         xaxis=list(title= "<b>Year</b>"),
         yaxis=list(title="<b>Unemployment Rate</b>"),
         legend=list(title=list(text='<b> State </b>'), showlegend=TRUE))

une
```

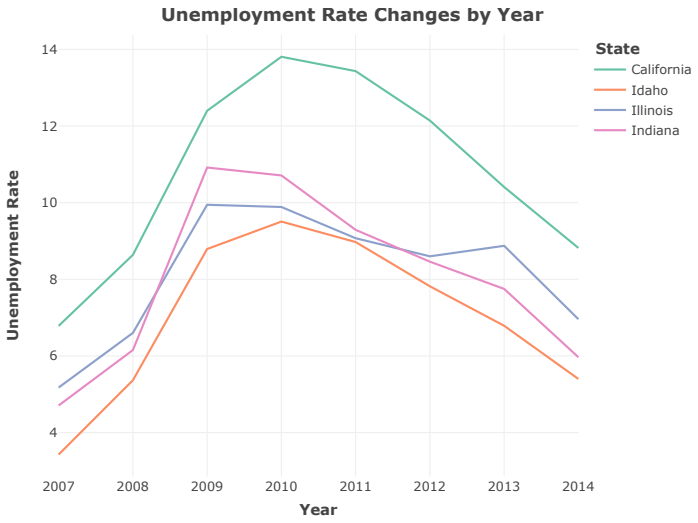


Figure 3.8: Unemployment rate time series plot.

3.5 Create the crime rate time series plot.

Note: To better see the crime rate for California select it from the legend on the right of the plot.

```
cr <- plot_ly(data=stats_df, x= ~as.factor(Year), y= ~Crimerate, color= ~NAME) %>%
  filter(NAME %in% states) %>%
  group_by(NAME) %>%
  add_lines() %>%
  layout(title="<b>Crime Rate Changes by Year</b>",
         xaxis=list(title= "<b>Year</b>"),
         yaxis=list(title="<b>Crime Rate</b>"), yaxis=list(range(c(0, .7))),
         legend=list(title=list(text='<b> State </b>'), showlegend=TRUE))


cr
```

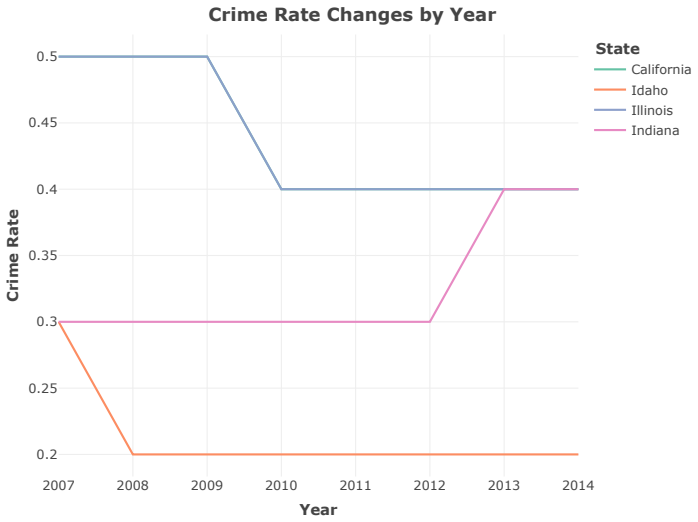


Figure 3.9: Crime rate time series plot.

4 Discussion and Conclusion

This project proposed to investigate if there was a relationship between changes in the unemployment rate and the crime rate. This project resulted in numeric and graphical data that can be used to answer the following questions:

1. Does the employment rate change over time and in a specific region of the country?
2. Can we discern a pattern or a correlation between the changes in the unemployment rate and the geography in which the data has been taken?

While it might be self-evident that an increase in unemployment would increase the crime rate, we can use the analysis that resulted from the graphs and the numerical data to see if there is a correlation. Many methods were employed to see if there was a correlation between the two.

Examining the statistics found in Section 3.1 EDA Analysis, we can see in tables that show some statistical measurements for the data. The tables and graphs show the quantiles, the mean, maximum, minimum, and any outliers for the available data. Table 3.4 Regional Unemployment Statistics. Shows the grouped data for all the years within the dataset. In Figure 3.1 Unemployment Rate by Region. Based on the boxplots regions 3 and 4 seem to have an extreme range in the unemployment rate for the period.

If we examine the data for each year as given in Table 3.5 Regional Unemployment Statistics by Year and Region, regions 3 and 4 usually have the highest mean unemployment rate over the period. To better illustrate the measurements that were given in the table, it may be helpful to look at a boxplot of the data over time. Using a faceted boxplot we can visualize the data for the regions for each year that there is data. Studying Figure 3.2 Unemployment Rate by Year and Region we can see that regions 2 and 4 do show the extremes in the unemployment rate.

The next item to look at is the crime rate in aggregate. Looking at the data over time we can see that the regions with the highest mean change in crime rate are regions 3 and 4. In Table 3.6 Regional Crime Statistics we can see the statistics for aggregated crime rate. These statistics seem to correspond with the high unemployment rates in regions 3 and 4, as discussed previously. To visualize this data we can refer to Figure 3.3 Crime Rate by Region. In this boxplot, we can see the range in the rate of change in the crime rate for regions. Regions 3 and 4 have extreme changes in crime for the data.

Referring to Table 3.7 Regional Crime Statistics by Year and Region shows the crime rate statistics by year and the regions. Over the period regions 3 and 4 continue to show the highest change in crime. To make it easier to understand it may be beneficial to look at the boxplots of the crime rate change over the year 2007 to 2014.

This can be seen in Figure 3.4 Crime Rate by Year and Region the regions show consistently high changes in the crime rate. If we look at a spatial map for a year, is there a suggestion of any correlation between the crime rate and the unemployment rate? The first spatial map to be examined is for the unemployment this map can be found in Figure 3.5. The highest rate of unemployment for 2014 is about 9% for the states of Mississippi and Arizona. The state of Mississippi is in Region 3 and Arizona is in Region 4.

What does this mean for the crime rate for the states looking at Figure 3.6? The rates for these states in 2014 were in the range of 0.1 for Mississippi and 0.4 for the state of Arizona. While there does seem to be some discrepancy in terms of the crime rate change, this may be due to the time frame that was chosen to examine for the spatial map.

Sometimes it may be beneficial to look at the data points as they relate to each other. This can be accomplished using the scatterplot. In Figure 3.7, there is a scatterplot of the data for the year 2014, based on this plot there does not seem to be a linear relation between the unemployment rate and the crime rate. There does not seem to be a linear correlation between the data points as indicated by the graph (Question Video: Identifying the Linear Correlation from the Scattergraph, Nagwa, n.d.).

When looking at a few examples from the selection of states we can further look for a trend in in crime and the unemployment rate. Using the states of California, Idaho, Illinois, and Indiana we can see the unemployment rates from 2007 to 2014 (Please refer to Figure 3.8 Unemployment Rate Changes by Year and Figure Crime Rate Changes by Year 3.9. (Please note it may be beneficial to toggle the different states on and off by selecting them from the legend located on the right side of the graph.)

We can see that California had the highest unemployment in the years 2010 through 2014. California had a change in the crime rate that was greater than the other states that were listed in the graph. One thing to note is Illinois had a spike in the unemployment rate in 2013 and did not see a commensurate change in the crime rate as indicated by no change in the crime rate shown in Figure 3.9.

To look at this relationship in a numerical manner we can look at the correlation coefficient. If we look at the correlation coefficient 0.17. The value of 0.17 indicates that there is almost no correlation between the variables to indicate that there is a meaningful correlation between crime and unemployment rates. This value is very close to the value zero. For there to be a correlation between the variables the value needs to fall closer to either -1 or to 1. Values closer to -1 indicate a negative correlation indicating that the variables change in a negative relation to each other. If one of the variables increases the other will decrease and vice versa (Soetewey, 2020).

While looking at data utilizing different methods. It seems that we can determine that there may be a possible correlation based on the graphs, but using a scatterplot there does not seem to be a correlation between the unemployment rate and the crime rate. There is a natural tendency to think that as unemployment increases the crime rate will also increase and it seems to be true.

Upon further research into the subject of crime rate and unemployment, there is a correlation between the two. This correlation depends on the type of crime that is being examined. When there is high unemployment the type of crime will vary. Crimes that involve possible quality of living may be only marginally affected. For instance, larceny, burglary, and robbery have a positive relationship. Crimes that may not affect the quality of living like car theft do not show a positive correlation but illustrate a negative relationship. This is not to say that car theft does not go up if given the right circumstances (Fallahi & Rodríguez, 2014). For instance, car theft will go up if there is a period of expansion within the economy. This is logical since there will be a market for the cars and the parts that they contain.

For further investigative purposes, it might be beneficial to look into the regions that were covered in the data to see what the prevalent form of employment is and see if there are any similarities in the type of work that may cause such high rates of unemployment.

5 References

Fallahi, F., & Rodríguez, G. (2014). Link between unemployment and crime in the US: A Markov-Switching approach. *ScienceDirect*.

Question Video: Identifying the Linear Correlation from the Scattergraph, Nagwa. (n.d.). *Identifying the Linear Correlation from the Scattergraph [Video].* Nagwa. https://www.nagwa.com/en/videos/909167139353/ (https://www.nagwa.com/en/videos/909167139353/)

Soetewey, A. (2020, May 28). *Correlation coefficient and correlation test in R. Stats and R.* Retrieved February 13, 2025, from https://statsandr.com/blog/correlation-coefficient-and-correlation-test-in-r/ (https://statsandr.com/blog/correlation-coefficient-and-correlation-test-in-r/)