

# Project\_Stat184

Catherine Galant

## Contents

<b>Project Introduction</b>	<b>1</b>
Primary Dataset . . . . .	1
Secondary Dataset . . . . .	2
Data Wrangling . . . . .	2
<b>Data Visualization</b>	<b>5</b>
Code Appendix . . . . .	8

## Project Introduction

My goal with this project is to look at how crime in Montgomery County has changed over time.

To answer my guiding questions I plan to look at how crime changes over time for a location. My first guiding question is really to help narrow down my scope, by picking out 5-10 of the most prominent offenses in Montgomery County I can take a closer look at they change through the past 9 years.

I plan to facet out the data by crime and/or city and look at how it changes by month/year by plotting the total number of crimes.

## Primary Dataset

My primary datasets is from catalog.Data.gov and is public crime data from Montgomery County, MD that catalogs every crime reported in the county from July 1 2016 to Aug 2 2025. It has 38 variables detailing the specifics of the crime like the exact time, location, charge, victims, location, etc.

This dataset is useful for my guiding questions because it gives specific location, crime, and time data. It will tell me exactly what crime was committed, divided into categories 3 times so to specify the exact incident. It also includes multiple different location categories, like the exact address, city, and police district.

## **Secondary Dataset**

My secondary dataset is from [opendata.Maryland.gov](https://opendata.maryland.gov/), it includes a tally of violent crimes committed every year from 1975 to 2022 broken up by counties in Maryland. It also gives the percent change of each crime by year, the population of the county, and the crime rate per 100,000 people for each violent crime.

This dataset is important for my guiding questions because it will help compare how Montgomery County compares to its surrounding counties in terms of crime rates.

## **Data Wrangling**

To get my data into a usable format, I will select only the variables that I need from and filter it so that only cases relevant to my data are observed. For my primary dataset 'Crime.csv' from [catalog.Data.gov](https://catalog.data.gov/) that has every crime reported from July 1, 2016 to August 2, 2025.

I first separate the data into 2 different sets that I would like to look at, the first is how all crime rates change over the time period, the second is how the most common crimes change over time. To identify the most common crimes I group by the common crime name in `Crime.Name2` and summarise the count, arrange in descending order and get the top 9 crimes.

## **These next steps I do to both dataframes.**

I only want the start date and time, the specific crime, and the Police District it happened in. I have to convert the start date and time into a POSIXct date time so that I can use it later and I mutate the column to do that. Then I create new columns that store only the year and the month and get rid of the original `Start_Date_Time` column. Next I filter out all the minor offenses underneath "All Other Offenses" and all the unfinished Police District information, usually written as "OTHER" or left blank.

For the first dataframe that looks at all crimes over the months and years, named `mont_crime_year_district`. To get the summary data I need for graphing, I group it by Year, Month, and Police.District, and then summarize it by count. This gets me my tidied dataframe needed for graphing. For the second dataframe, named `mont_year_bycrime`, that looks at 9 of the top crimes by year, I first filter the data to only have those top 9 crimes, and then I group by year and crime, and summarize again by count. Now this dataframe is also ready to be plotted.

Table 1: Montgomery County Crime by Police District/Year

Year	Month	Police District	Count
2017	10	SILVER SPRING	909
2017	09	SILVER SPRING	889
2019	05	SILVER SPRING	885
2017	08	SILVER SPRING	871
2019	10	SILVER SPRING	868

Table 2: Montgomery County Top Crimes by Year

Year	Crime	Count
2018	Drug/Narcotic Violations	6045
2017	Drug/Narcotic Violations	5889
2024	Shoplifting	5455
2020	Theft From Motor Vehicle	4993
2017	Theft From Motor Vehicle	4956

Below are the first 5 lines of each dataframe.

For my secondary dataset, it has the total number of the crimes it looks at as individual columns, so I need to pivot the dataframe to be longer and make a new column that has what crime was committed.

After exploratory data analysis I discovered that the number of crimes for murder and rape were much lower than the others and observing the trend for them would be difficult if plotted on the same graph, so I separated them into 2 dataframes for plotting. The first dataframe is `md_crime_high` for high volume crimes. These crimes are Robbery, Aggravated Assault, Breaking & Entering, and Motor Vehicle Theft. The second dataframe is `md_crime_low` for low volume crimes, Murder and Rape.

After using `filter` to get only the counties needed and `select` to separate the data, I pivoted each dataframe so that instead of multiple columns for each crime, crime was one column with a value and each year and county had multiple rows, making each crime a case rather than each year and county. Now after some cleaning up of the column names the 2 dataframes are complete and ready for plotting.

Below are the first 5 lines of the new dataframes.

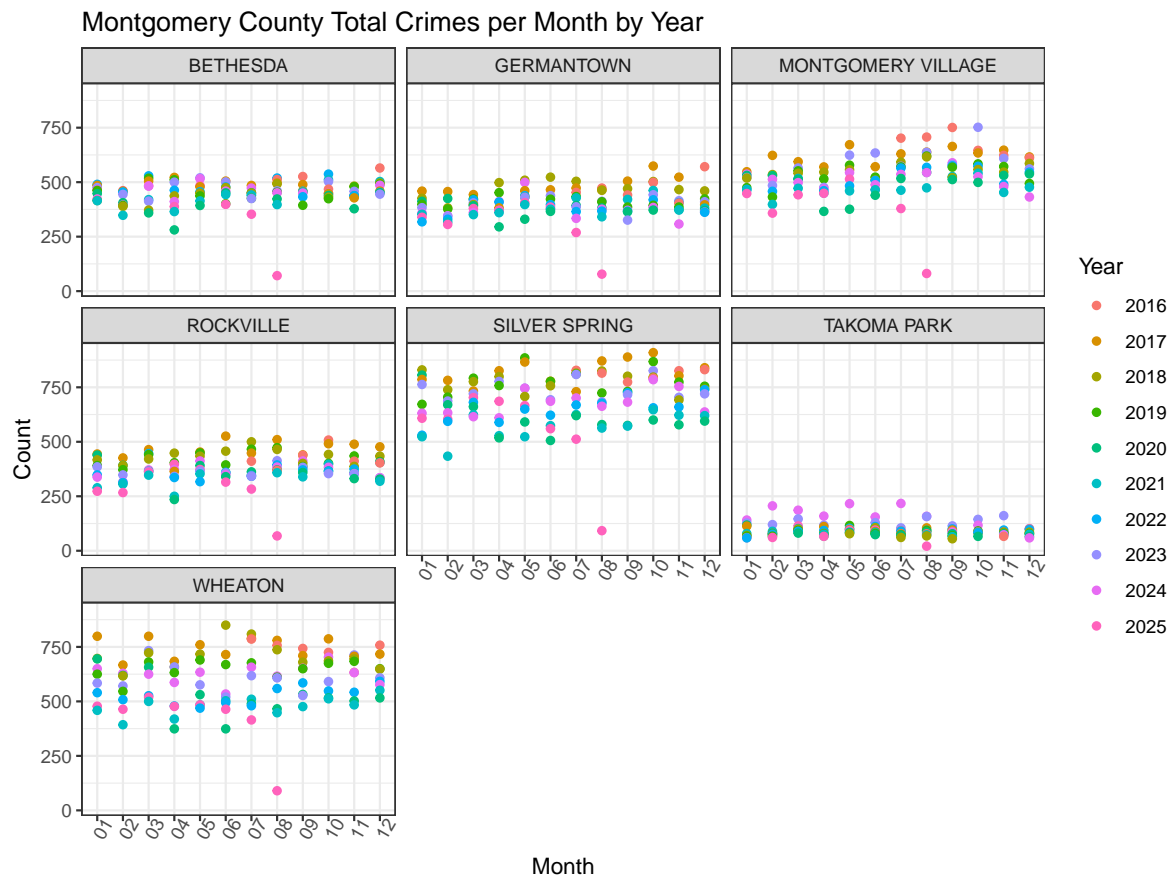
Table 3: Maryland Crime Numbers by Year for Popular Counties: High Number Crimes

JURISDICTION	YEAR	CRIME	NUMBER
Anne Arundel County	1975	ROBBERY	413
Anne Arundel County	1975	AGGRAVATED ASSAULT	1514
Anne Arundel County	1975	BREAKING & ENTERING	5662
Anne Arundel County	1975	MOTOR VEHICLE THEFT	1986
Anne Arundel County	1976	ROBBERY	330

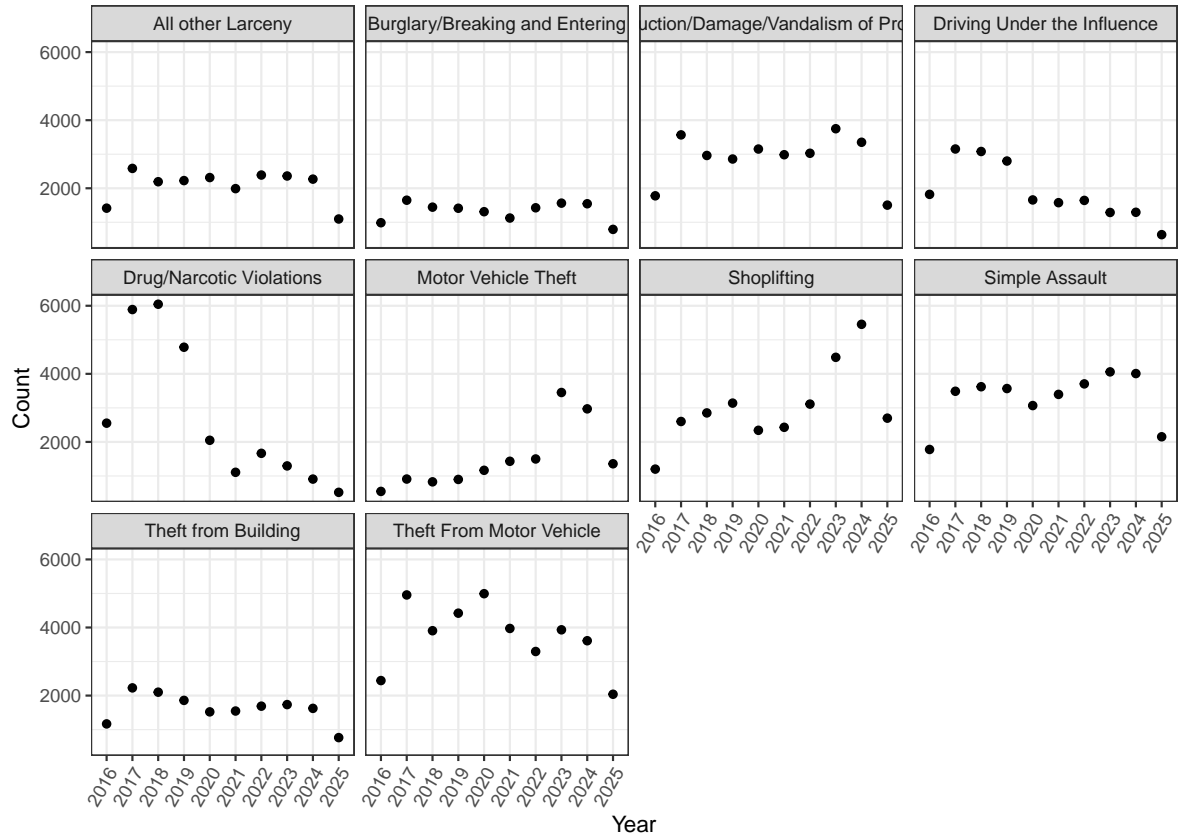
Table 4: Maryland Crime Numbers by Year for Popular Counties: Low Number Crimes

JURISDICTION	YEAR	CRIME	NUMBER
Anne Arundel County	1975	MURDER	16
Anne Arundel County	1975	RAPE	74
Anne Arundel County	1976	MURDER	10
Anne Arundel County	1976	RAPE	77
Anne Arundel County	1977	MURDER	16

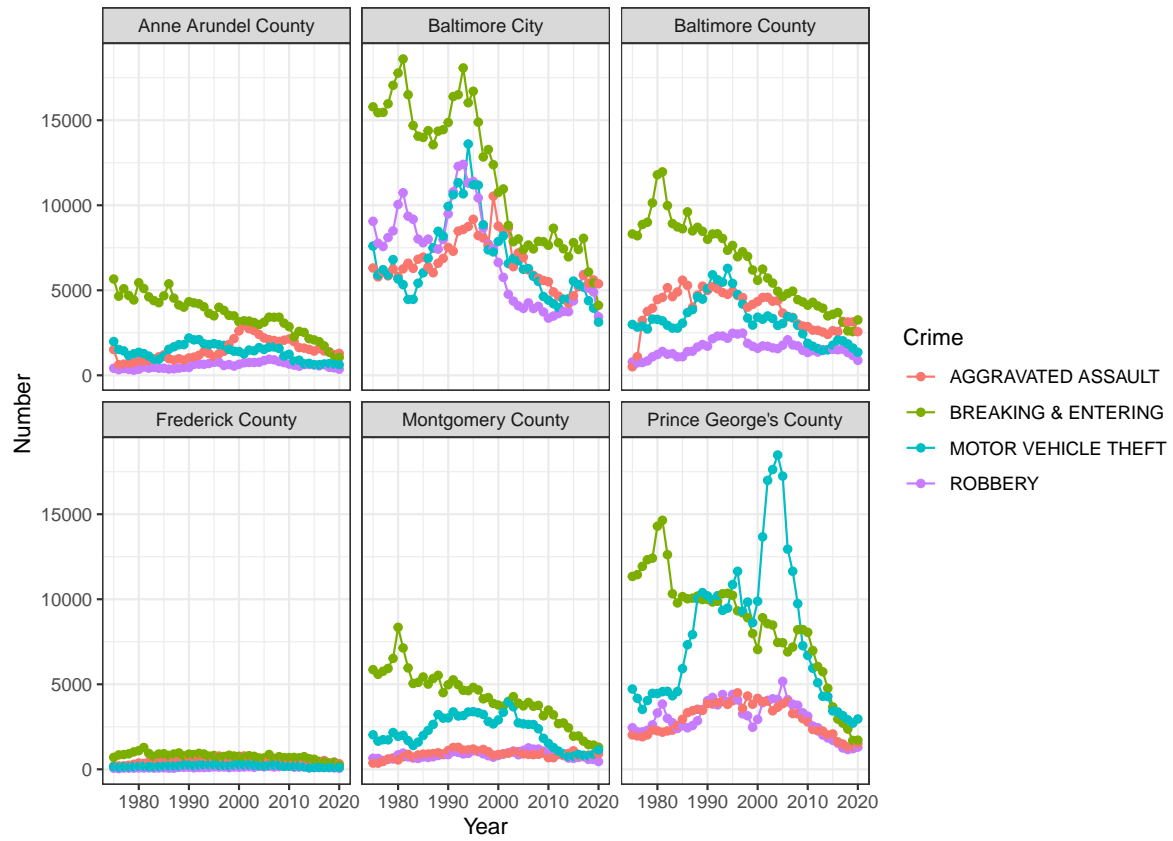
## Data Visualization



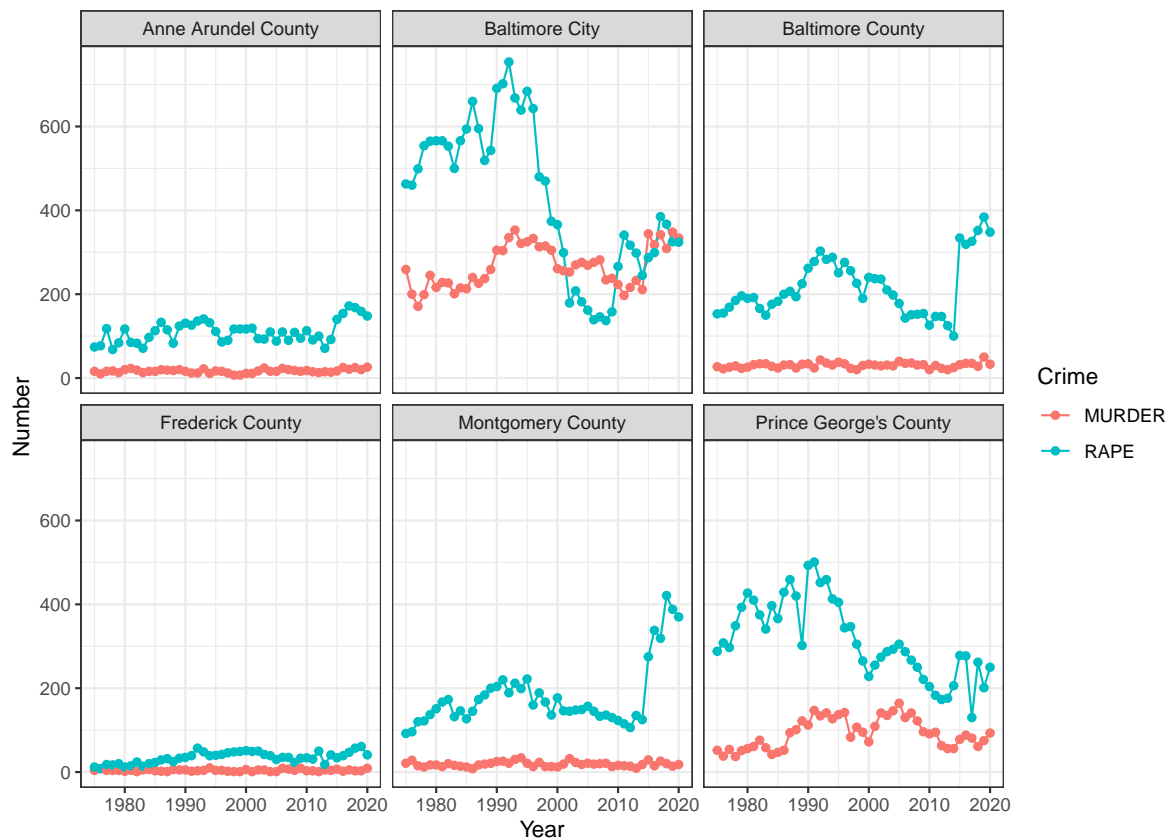
Montgomery County Top Crimes by Year



Number of other Crime Incidents in Prominent Maryland Counties 1975–2022



Number of Murder and Rape Incidents in Prominent Maryland Counties 1975–2022



## Code Appendix

```
#My chosen style guide is the tidyverse style guide
```

```
library(tidyverse)
library(ggplot2)
library(kableExtra)
```

```
mont_crime <- read.csv("C:\\Users\\catie\\Downloads\\Crime.csv")
```

```
md_crime <- read.csv("C:\\Users\\catie\\Downloads\\Violent_Crime___Property_Crime_by_County_")
```

```
mont_crime <- mont_crime%>%
```

```
  select(Start_Date_Time, Crime.Name2, Police.District.Name)
```

```
mont_crime_year_district <- mont_crime %>%
```



```

mutate(Start_Date_Time = as.POSIXct(Start_Date_Time, format = "%m/%d/%Y %H:%M:%S %p")) %>%
mutate(Year = format(Start_Date_Time, format = '%Y')) %>%
mutate(Month = format(Start_Date_Time, format='%m')) %>%
select(-Start_Date_Time) %>%
filter(Crime.Name2 != 'All Other Offenses') %>%
filter(Police.District.Name != 'OTHER') %>%
filter(Police.District.Name != '') %>%
group_by(Year, Month, Police.District.Name) %>%
summarise(num=n()) %>%
arrange(desc(num))

colnames(mont_crime_year_district) <- c('Year', 'Month', 'Police District','Count')
mont_crime_year_district <- na.omit(mont_crime_year_district)

top9_crimes <- (mont_crime %>%
  group_by(Crime.Name2) %>%
  summarise(count = n()) %>%
  filter(Crime.Name2 != 'All Other Offenses') %>%
  arrange(desc(count)))$Crime.Name2[1:10]
mont_year_bycrime <- mont_crime %>%
  mutate(Start_Date_Time = as.POSIXct(Start_Date_Time,format = "%m/%d/%Y %H:%M:%S %p")) %>%
  mutate(Year = format(Start_Date_Time, format = '%Y')) %>%
  mutate(Month = format(Start_Date_Time, format = '%m')) %>%
  select(-Start_Date_Time) %>%
  filter(Crime.Name2 %in% top9_crimes) %>%
  filter(Police.District.Name != 'OTHER') %>%
  filter(Police.District.Name != '') %>%
  group_by(Year, Crime.Name2) %>%
  summarise(num = n()) %>%
  arrange(desc(num))
mont_year_bycrime <- na.omit(mont_year_bycrime)
colnames(mont_year_bycrime) <- c('Year', 'Crime', 'Count')

output_mont_cyd <- mont_crime_year_district[1:5,]
output_mont_cyd %>%
  kbl(caption = "Montgomery County Crime by Police District/Year") %>%
  kable_classic(full_width = F, html_font = "Cambria")

output_mont_ybc <- mont_year_bycrime[1:5,]
output_mont_ybc %>%

```

```

kbl(caption = "Montgomery County Top Crimes by Year") %>%
kable_classic(full_width = F, html_font = "Cambria")

md_crime <- md_crime %>%
  filter(JURISDICTION %in% c("Montgomery County", "Baltimore City", "Anne Arundel County", "I

md_crime_high <- md_crime %>%
  select(JURISDICTION, YEAR, ROBBERY, AGG..ASSAULT, B...E, M.V.THEFT)
md_crime_low <- md_crime %>%
  select(JURISDICTION, YEAR, RAPE, MURDER)
md_crime_high <- md_crime_high %>%
  pivot_longer(cols = c(ROBBERY, AGG..ASSAULT, B...E, M.V.THEFT), names_to = 'CRIME')

md_crime_high$CRIME[md_crime_high$CRIME == 'AGG..ASSAULT'] = 'AGGRAVATED ASSAULT'
md_crime_high$CRIME[md_crime_high$CRIME == 'B...E'] = 'BREAKING & ENTERING'
md_crime_high$CRIME[md_crime_high$CRIME == 'M.V.THEFT'] = 'MOTOR VEHICLE THEFT'

md_crime_low <- md_crime_low %>%
  pivot_longer(cols = c(MURDER, RAPE), names_to = 'CRIME')
colnames(md_crime_high)<-c('JURISDICTION', 'YEAR', 'CRIME', 'NUMBER')
colnames(md_crime_low)<-c('JURISDICTION', 'YEAR', 'CRIME', 'NUMBER')

md_crime_high[1:5,] %>%
  kbl(caption = "Maryland Crime Numbers by Year for Popular Counties: High Number Crimes") %>%
  kable_classic(full_width = F, html_font = "Cambria")

md_crime_low[1:5,] %>%
  kbl(caption = "Maryland Crime Numbers by Year for Popular Counties: Low Number Crimes") %>%
  kable_classic(full_width = F, html_font = "Cambria")
p_year_district <- ggplot(data = mont_crime_year_district, mapping = aes(x = Month, y = Count)) +
  geom_point() +
  facet_wrap(vars(`Police District`)) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 60, hjust=0.5)) +
  labs(title='Montgomery County Total Crimes per Month by Year')

p_year_district

p_year_crime <- ggplot(data = mont_year_bycrime, mapping = aes(x = Year, y = Count)) +
  geom_point(mapping = aes(x = Year, y = Count)) +
  facet_wrap(vars(Crime)) +
  theme_bw() +

```

```

    theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
    labs(title = 'Montgomery County Top Crimes by Year')

p_year_crime

high_crime <- md_crime_high %>%
  ggplot(mapping = aes(x = YEAR, y = NUMBER, colour = CRIME)) +
  facet_wrap(vars(JURISDICTION)) +
  geom_line() +
  geom_point() +
  theme_bw() +
  ylab('Number') +
  xlab("Year") +
  labs(title = "Number of other Crime Incidents in Prominent Maryland Counties 1975-2022", c

low_crime <- md_crime_low %>%
  ggplot(mapping = aes(x = YEAR, y = NUMBER, colour = CRIME)) +
  facet_wrap(vars(JURISDICTION)) +
  geom_line() +
  geom_point() +
  theme_bw() +
  ylab('Number') +
  xlab("Year") +
  labs(title = "Number of Murder and Rape Incidents in Prominent Maryland Counties 1975-2022", c

high_crime
low_crime

```