In [31]:
```
#Import Libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
import seaborn as sns
import scipy.stats as stats
import plotly.figure_factory as ff
```

In [32]:
```
#Read data from CSV
df_crime =pd.read_csv("crimeratesbystate-formatted.csv")
```

In [33]:
```
df_crime.head()
```

Out[33]:

	state	murder	forcible_rape	robbery	aggravated_assault	burglary	larceny_theft	motor_vehicle_theft
0	United States	5.6	31.7	140.7	291.1	726.7	2286.3	416.7
1	Alabama	8.2	34.3	141.4	247.8	953.8	2650.0	288.3
2	Alaska	4.8	81.1	80.9	465.1	622.5	2599.1	391.0
3	Arizona	7.5	33.8	144.4	327.4	948.4	2965.2	924.4
4	Arkansas	6.7	42.9	91.1	386.8	1084.6	2711.2	262.1

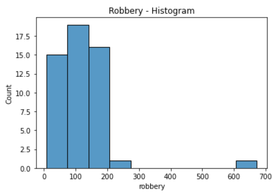
In [34]:
```
# Set up the dataset.
df_education = pd.read_csv('education.csv')
df_education.head()
```

Out[34]:

	state	reading	math	writing	percent_graduates_sat	pupil_staff_ratio	dropout_rate
0	United States	501	515	493	46	7.9	4.4
1	Alabama	557	552	549	7	6.7	2.3
2	Alaska	520	516	492	46	7.9	7.3
3	Arizona	516	521	497	26	10.4	7.6
4	Arkansas	572	572	556	5	6.8	4.6

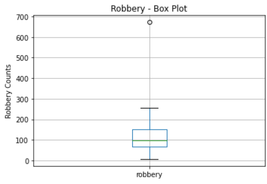
Histogram

In [35]:
```
sns.histplot(data=df_crime, x="robbery", bins=10)
plt.title("Robbery - Histogram")
plt.show()
```



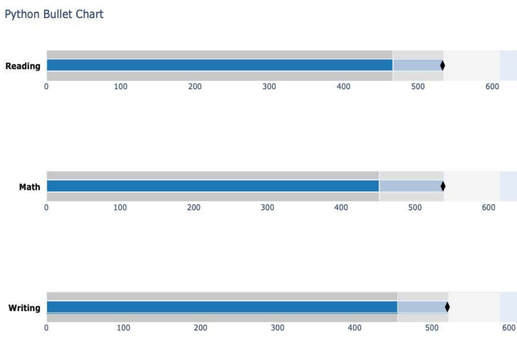
Box Plot

In [36]:
```
fig, ax = plt.subplots()
boxplot = df_crime.boxplot(column=['robbery'], return_type='axes')
plt.ylabel("Robbery Counts")
plt.title("Robbery - Box Plot")
plt.show()
```



Bullet Graph

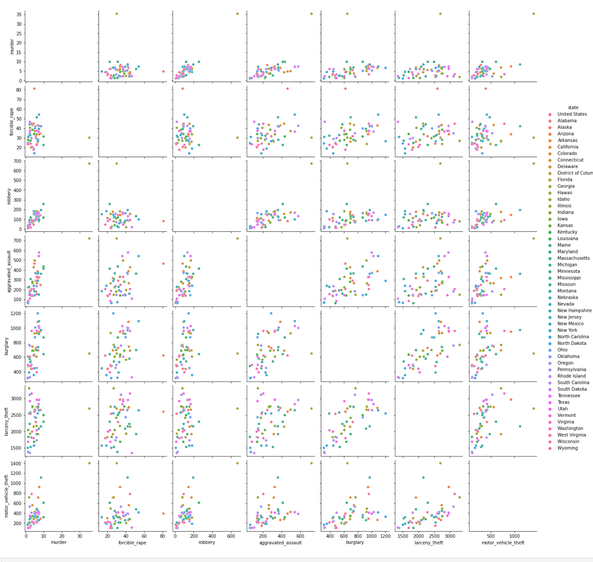
In [37]:
```
# Create a data object to hold the values to display.
data = [
    {"title": "Reading",
     "ranges": [df_education["reading"].min(), df_education["reading"].mean(), df_education["reading"].max()],
     "measures":[466, 533],
     "markers": [int(df_education["reading"].mean())]},

    {"title": "Math",
     "ranges": [df_education["math"].min(), df_education["math"].mean(), df_education["math"].max()],
     "measures":[451, 538],
     "markers": [int(df_education["math"].mean())]},

    {"title": "Writing",
     "ranges": [df_education["writing"].min(), df_education["writing"].mean(), df_education["writing"].max()],
     "measures":[455,520],
     "markers": [int(df_education["writing"].mean())]}
]

#Create the bullet chart.
fig = ff.create_bullet(
    data, titles='title',
    markers='markers',
    measures='measures',
    ranges='ranges',
    orientation='h',
    title='Python Bullet Chart'
)

fig.show()
```



Pair Plot

In [38]:
```
sns.pairplot(df_crime,hue="state")
```

Out[38]:
```
<seaborn.axisgrid.PairGrid at 0x7faf7b3f08e0>
```



In [ ]:
```
```

Assignment_6.2_Vayuvegula_Soma_Shekar_R

Soma Shekar Vayuvegula

02/25/2023

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.1.7     v purrr   0.3.4
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
##
## Attaching package: 'reshape2'
##
##
## The following object is masked from 'package:tidyr':
##
##     smiths
##
##
##
## Attaching package: 'data.table'
##
##
## The following objects are masked from 'package:reshape2':
##
##     dcast, melt
##
##
## The following object is masked from 'package:purrr':
##
##     transpose
##
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
##
##
##
## Attaching package: 'plotly'
##
##
## The following object is masked from 'package:ggplot2':
##
##     last_plot
##
##
## The following object is masked from 'package:stats':
##
##     filter
##
##
## The following object is masked from 'package:graphics':
##
##     layout
##
##
##
## Attaching package: 'reshape'
##
##
## The following object is masked from 'package:plotly':
##
##     rename
##
##
## The following object is masked from 'package:data.table':
##
##     melt
##
##
## The following objects are masked from 'package:reshape2':
##
##     colsplit, melt, recast
##
##
## The following objects are masked from 'package:tidyr':
##
##     expand, smiths
##
##
## The following object is masked from 'package:dplyr':
##
##     rename
##
##
## ------------------------------------------------------------------------------
##
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
```

```
## library(plyr); library(dplyr)
##
## ------------------------------------------------------------------------------
##
##
## Attaching package: 'plyr'
##
##
## The following objects are masked from 'package:reshape':
##
##     rename, round_any
##
##
## The following objects are masked from 'package:plotly':
##
##     arrange, mutate, rename, summarise
##
##
## The following object is masked from 'package:purrr':
##
##     compact
##
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
##
##
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
df_crime<-read.csv("crimeratesbystate-formatted.csv")
head(df_crime,5)
```

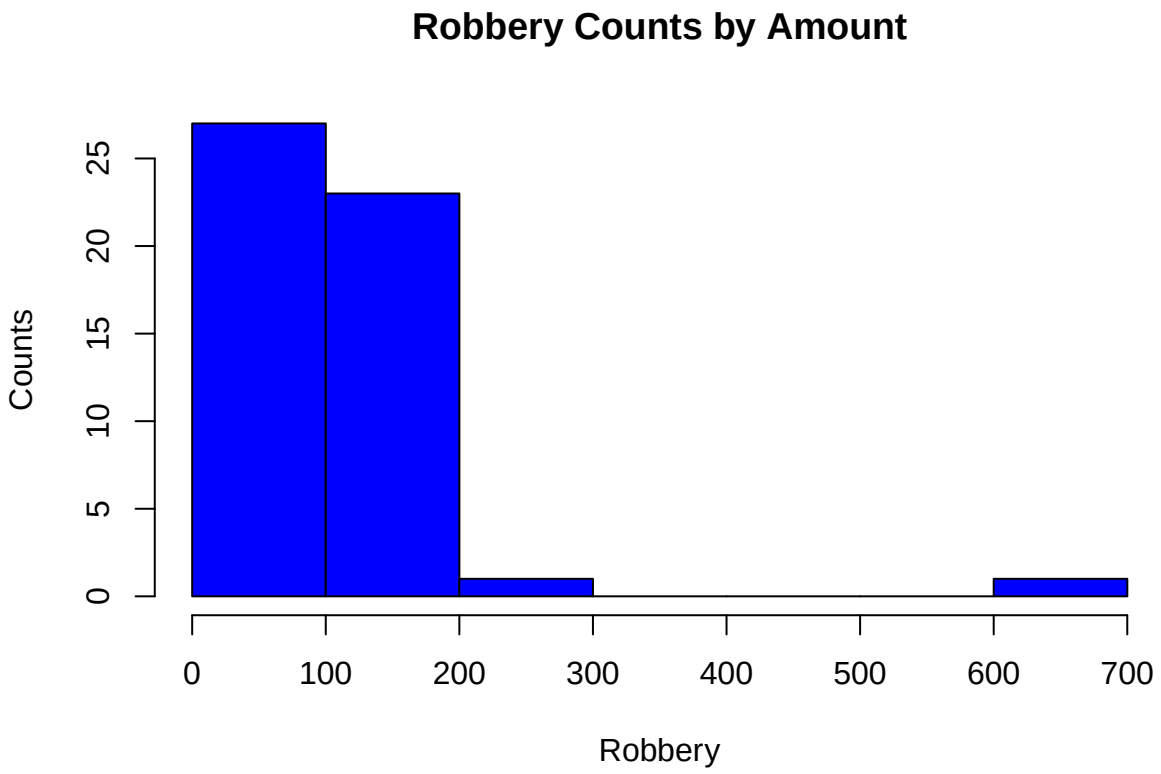
```
##            state murder forcible_rape robbery aggravated_assault burglary
## 1 United States     5.6          31.7   140.7              291.1    726.7
## 2       Alabama     8.2          34.3   141.4              247.8    953.8
## 3        Alaska     4.8          81.1    80.9              465.1    622.5
## 4       Arizona     7.5          33.8   144.4              327.4    948.4
## 5       Arkansas    6.7          42.9    91.1              386.8   1084.6
##   larceny_theft motor_vehicle_theft
## 1        2286.3               416.7
## 2        2650.0               288.3
## 3        2599.1               391.0
## 4        2965.2               924.4
## 5        2711.2               262.1
```

```
df_education<-read.csv("education.csv")
head(df_education,5)
```

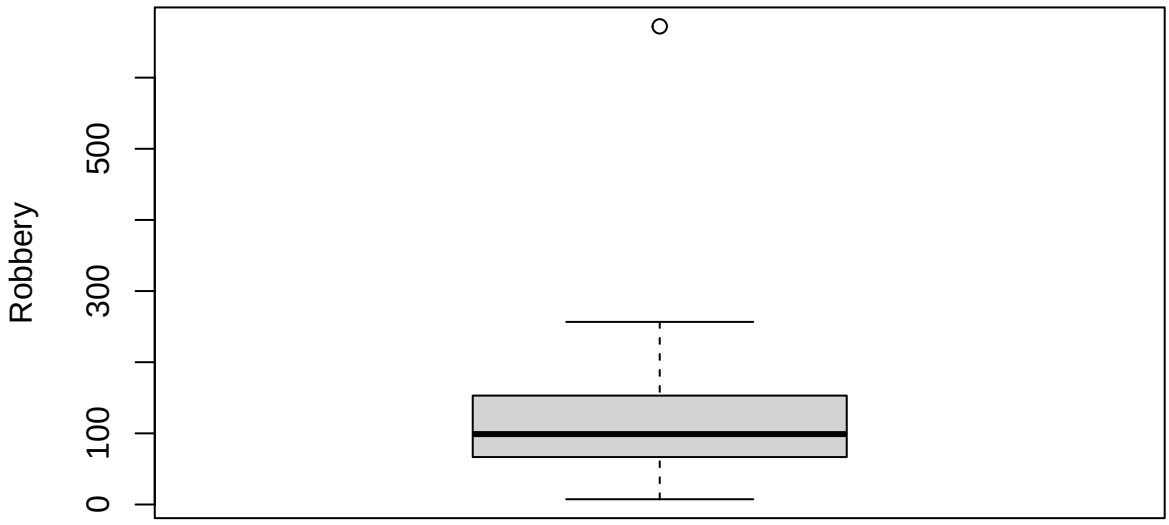
```
##           state reading math writing percent_graduates_sat pupil_staff_ratio
## 1 United States     501  515     493                    46              7.9
## 2       Alabama     557  552     549                     7              6.7
```

```
## 3        Alaska     520  516     492                  46              7.9
## 4       Arizona     516  521     497                  26             10.4
## 5      Arkansas     572  572     556                   5              6.8
##   dropout_rate
## 1          4.4
## 2          2.3
## 3          7.3
## 4          7.6
## 5          4.6
```

```
hist(df_crime$robbery,main="Robbery Counts by Amount",xlab="Robbery",
      ylab="Counts",
      col="blue",
      freq=TRUE)
```




```
options(warn=-1)
crime <- subset(df_crime, select = -c(state))
boxplot(x=crime$robbery,data=crime,xlab="Frequency",ylab="Robbery",main="Box Plot")
```

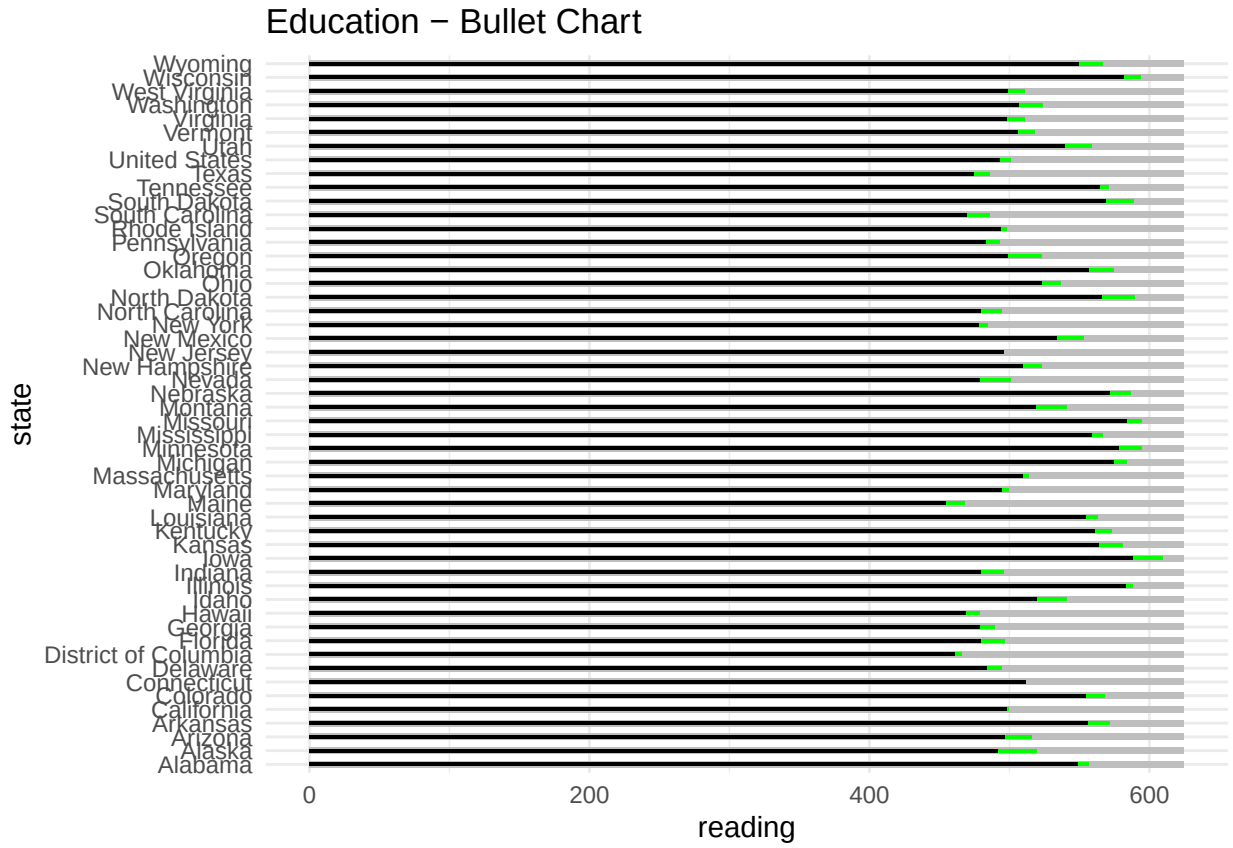


Frequency

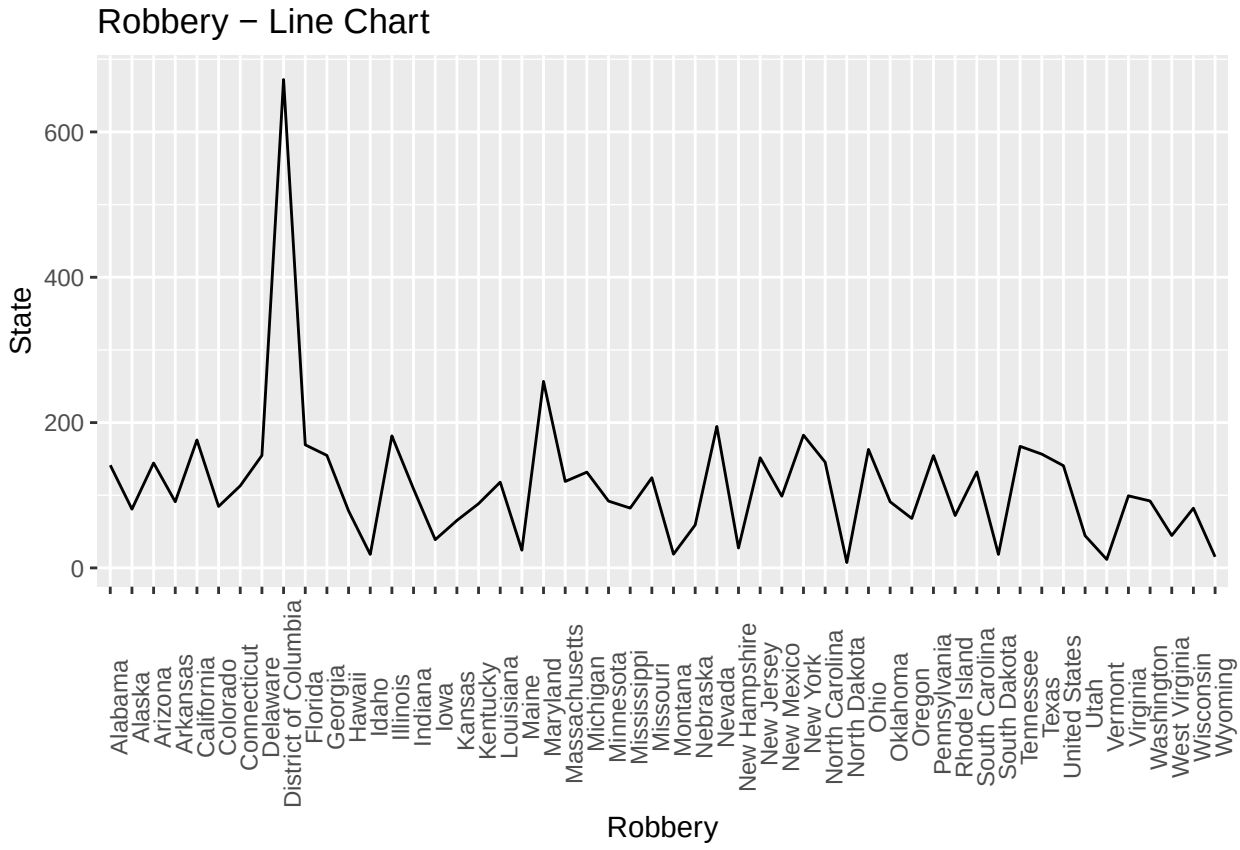
```
tibble(
  name = "R Bullet Chart",
  quant_value = 550,
  qualitative = 600
)
```

```
## # A tibble: 1 x 3
##   name           quant_value qualitative
##   <chr>                <dbl>       <dbl>
## 1 R Bullet Chart         550         600
```

```
df_education %>%
  ggplot(aes(x = reading, y =  state)) +
  geom_col(width = 0.5, aes(x = 625), fill = "grey") +
  geom_col(width = 0.25,aes(x = reading), fill = "green") +
  geom_col(aes(x = writing),fill = "black",color = NA,width = 0.25) +
  theme_minimal() +
  labs(title = "Education - Bullet Chart")
```



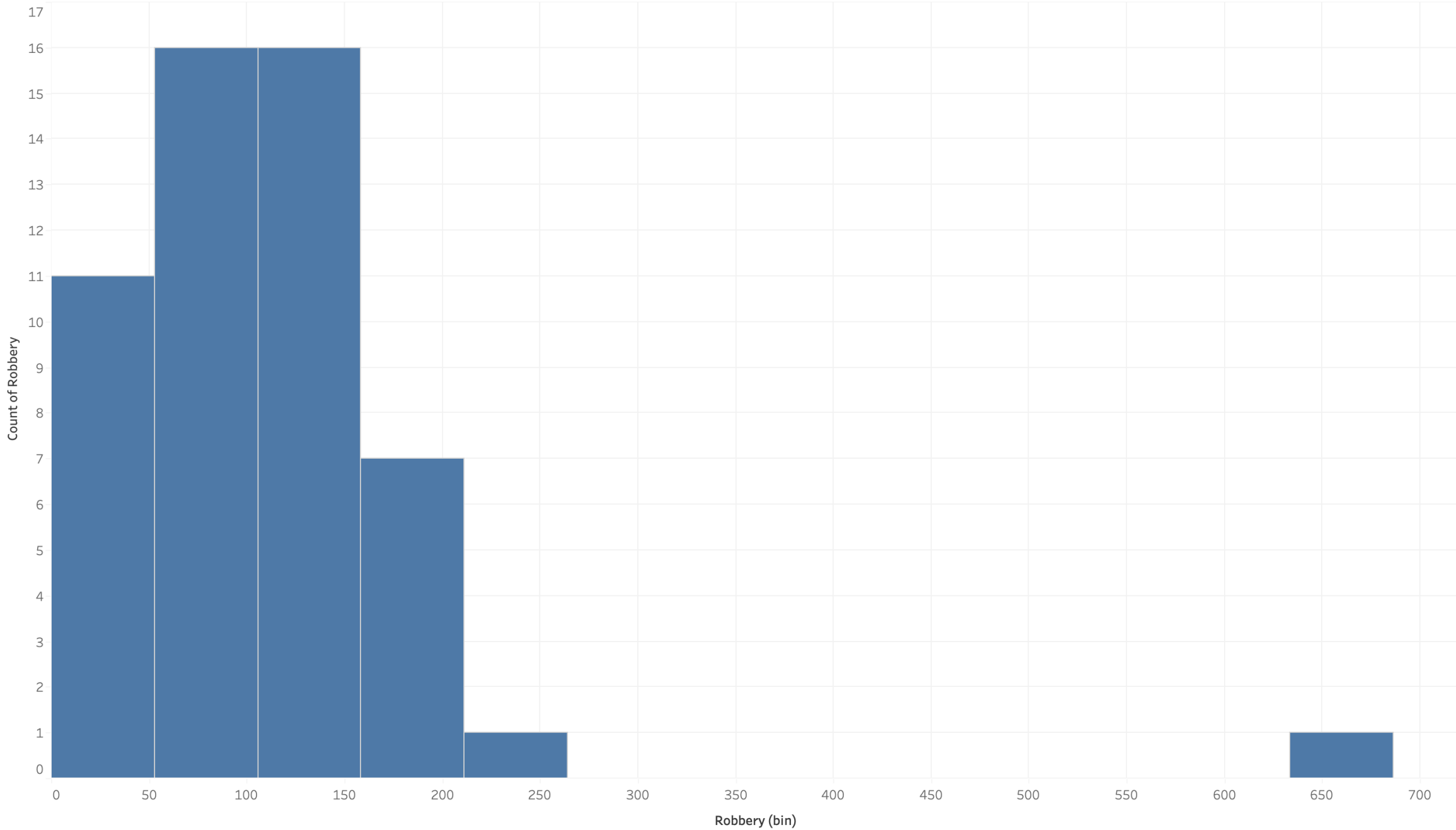

```
crime <- df_crime[,c("state","robbery")]
crime <- crime[!(crime$state=='United States'),]
p<-ggplot(crime,aes(x=state,y=robbery,group=1))+geom_line()+ggtitle("Robbery - Line Chart") +
  xlab("Robbery") +
  ylab("State")+theme(axis.text.x = element_text(angle = 90))
p
```


Robbery − Line Chart
State
600
400
200
0
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
United States
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
Robbery

Assignement_6.2_Vayuvegula_Soma_Shekar_Tableu

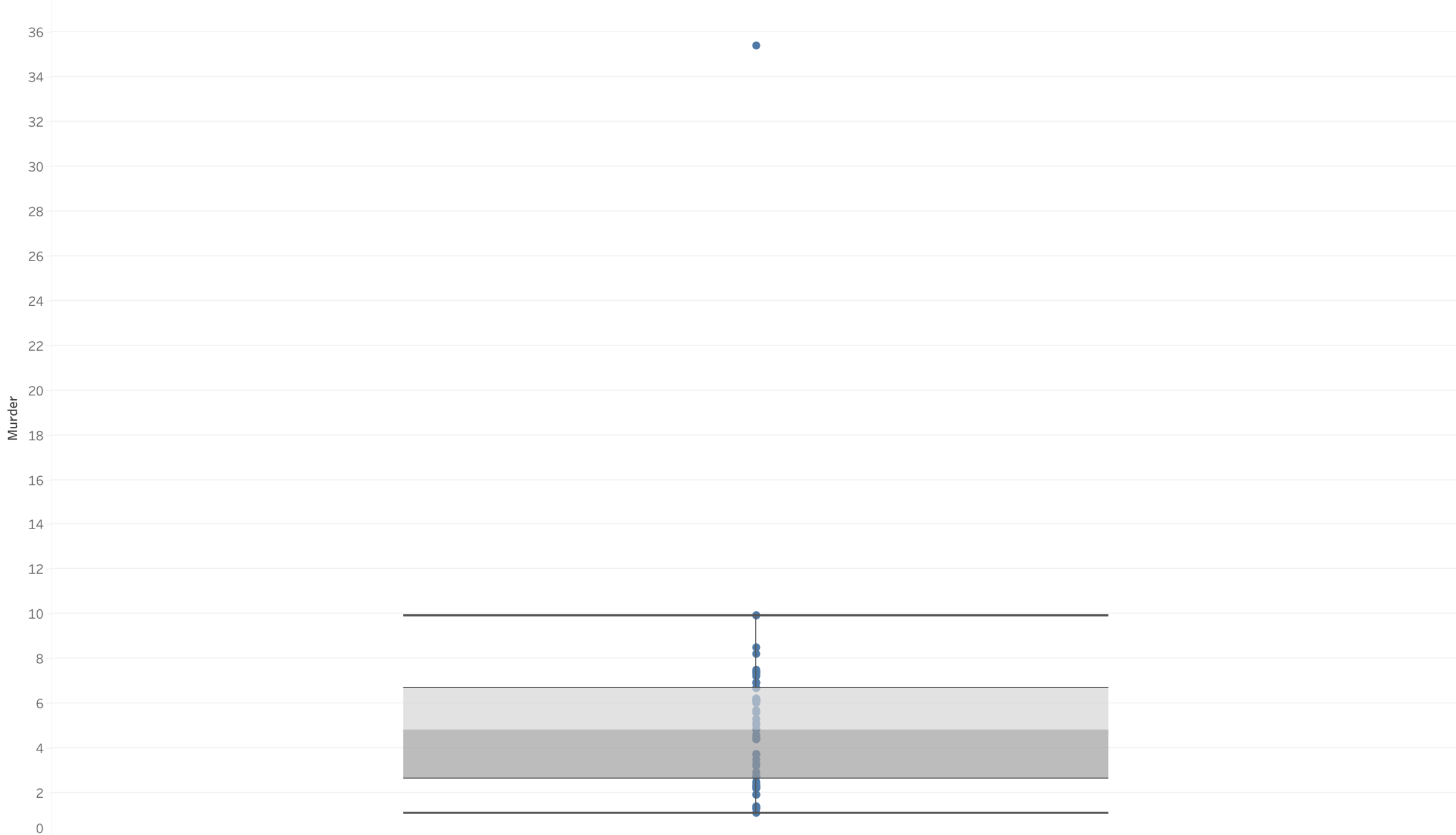
File created on: 2/26/23 12:46:28 AM CST

Histogram



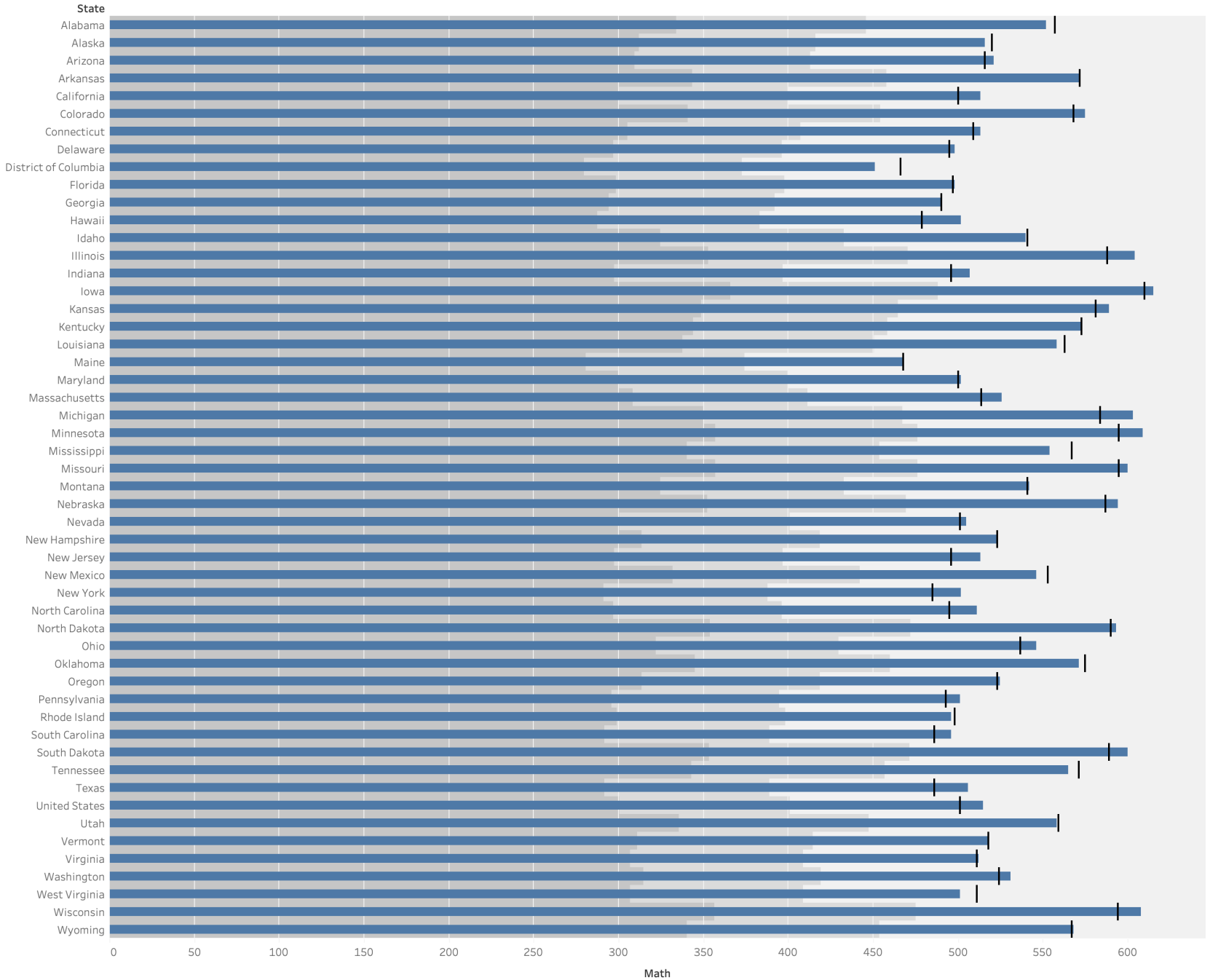
The trend of count of Robbery for Robbery (bin).

Box Plot



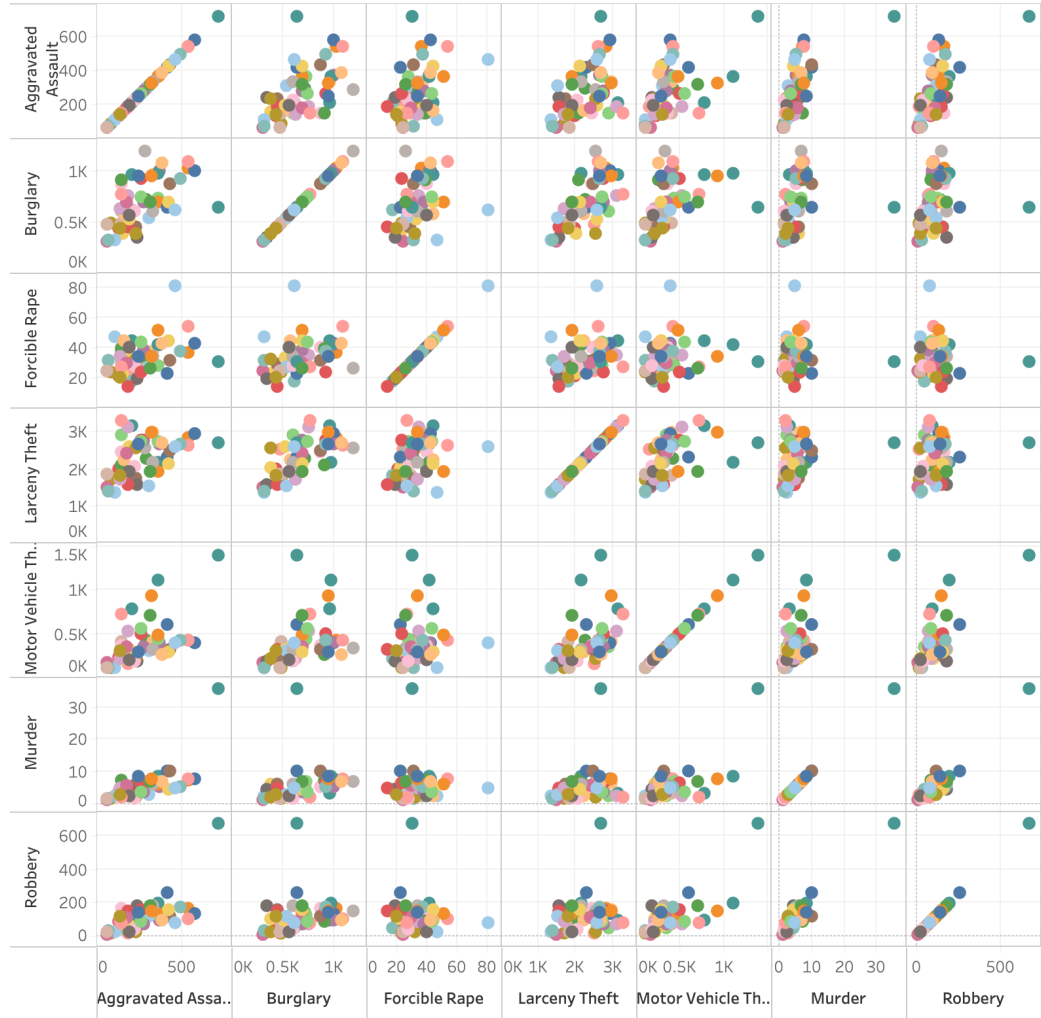
Sum of Murder. Details are shown for State.

Bullet Chart



Sum of Math for each State.

Pair Scatter plot



Sum of Aggravated Assault, sum of Burglary, sum of Forcible Rape, sum of Larceny Theft, sum of Motor Vehicle Theft, sum of Murder and sum of Robbery vs. sum of Aggravated Assault, sum of Burglary, sum of Forcible Rape, sum of Larceny Theft, sum of Motor Vehicle Theft, sum of Murder and sum of Robbery. Color shows details about State. Details are shown for State.

State
- Alabama
- Alaska
- Arizona
- Arkansas
- California
- Colorado
- Connecticut
- Delaware
- District of Columbia
- Florida
- Georgia
- Hawaii
- Idaho
- Illinois
- Indiana
- Iowa
- Kansas
- Kentucky
- Louisiana
- Maine
- Maryland
- Massachusetts
- Michigan
- Minnesota
- Mississippi
- Missouri
- Montana
- Nebraska
- Nevada
- New Hampshire
- New Jersey
- New Mexico
- New York
- North Carolina
- North Dakota
- Ohio
- Oklahoma
- Oregon
- Pennsylvania
- Rhode Island
- South Carolina
- South Dakota
- Tennessee
- Texas
- United States
- Utah
- Vermont
- Virginia
- Washington
- West Virginia
- Wisconsin
- Wyoming