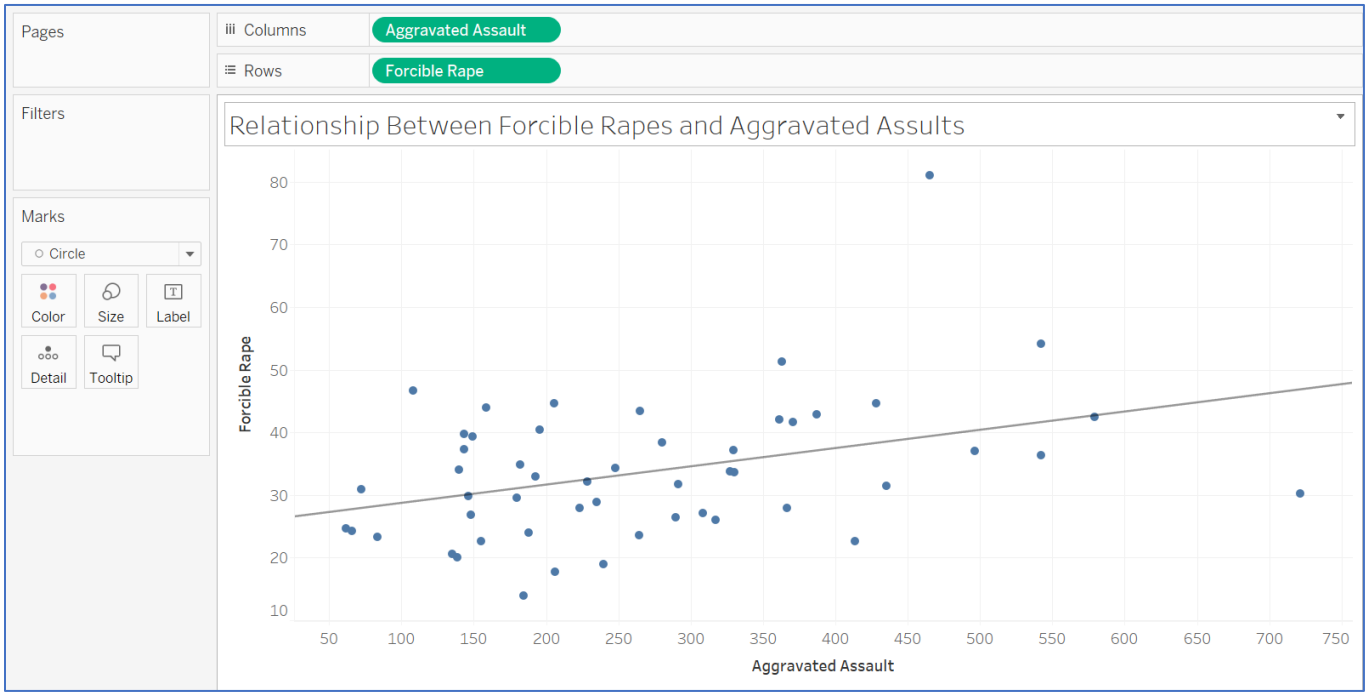
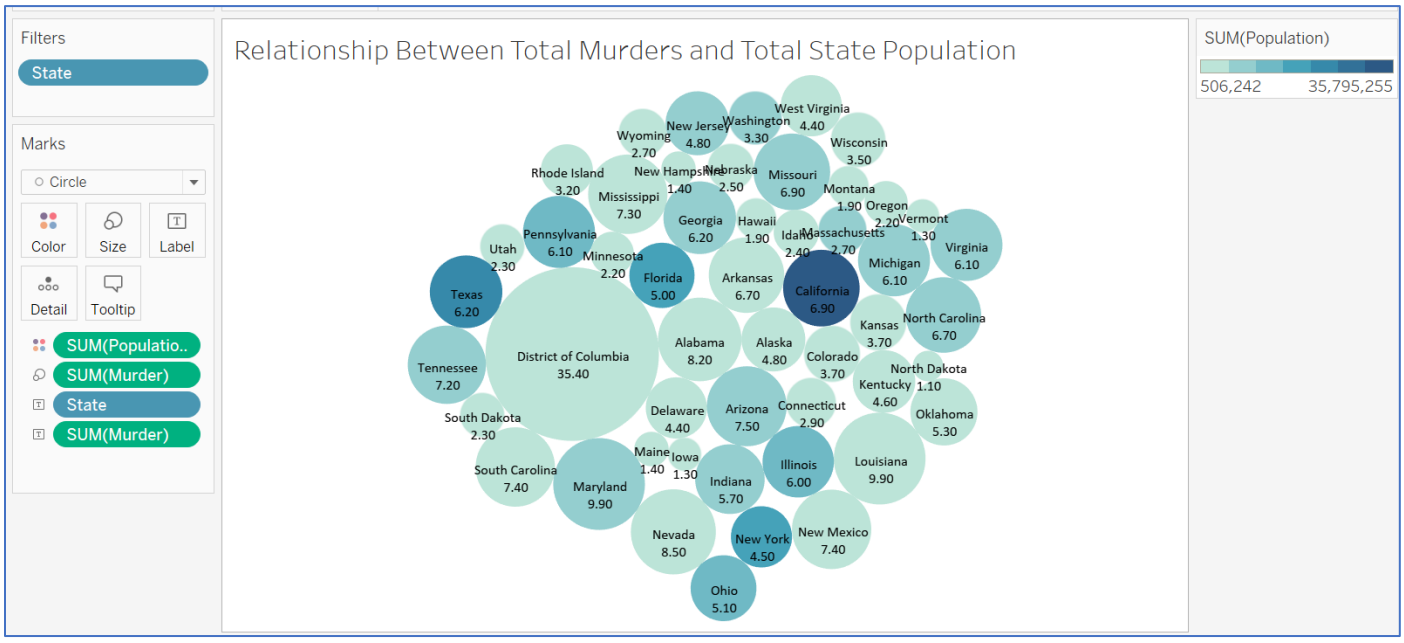
4.2 Exercises: Scatterplots, Bubble Charts & Density Plots/Maps

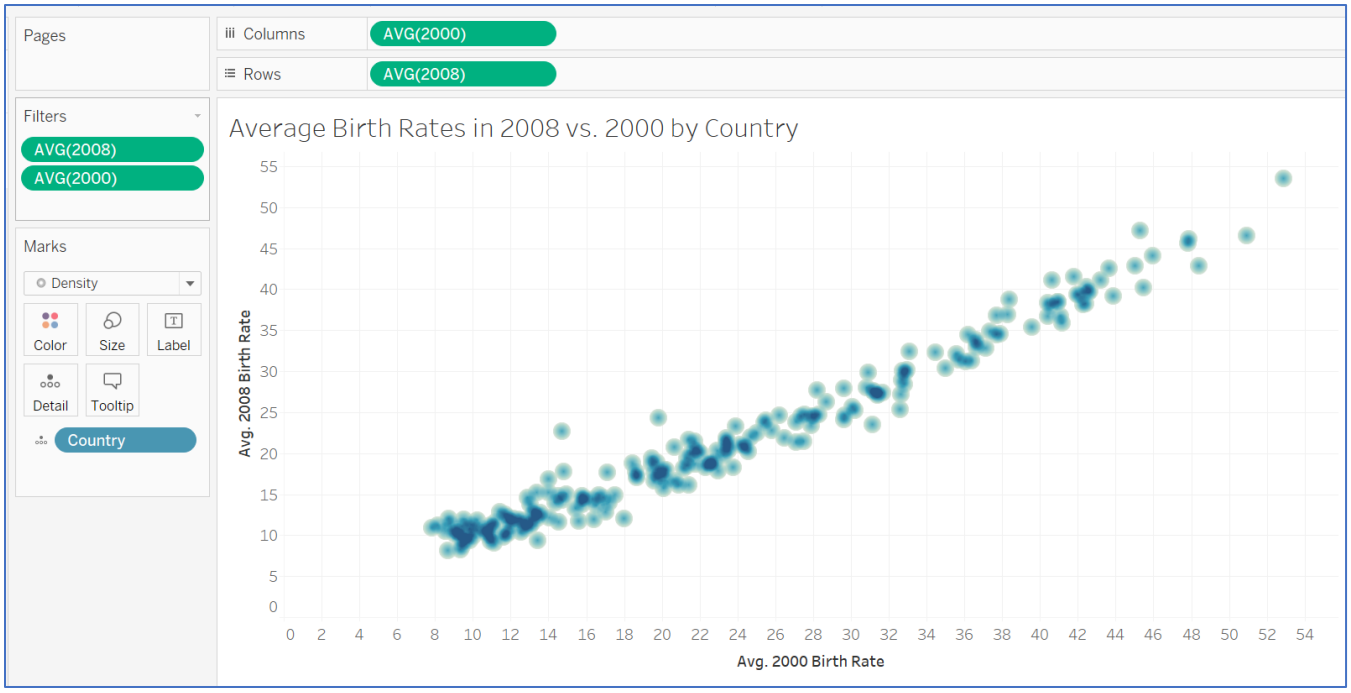
A. Scatterplot – Tableau

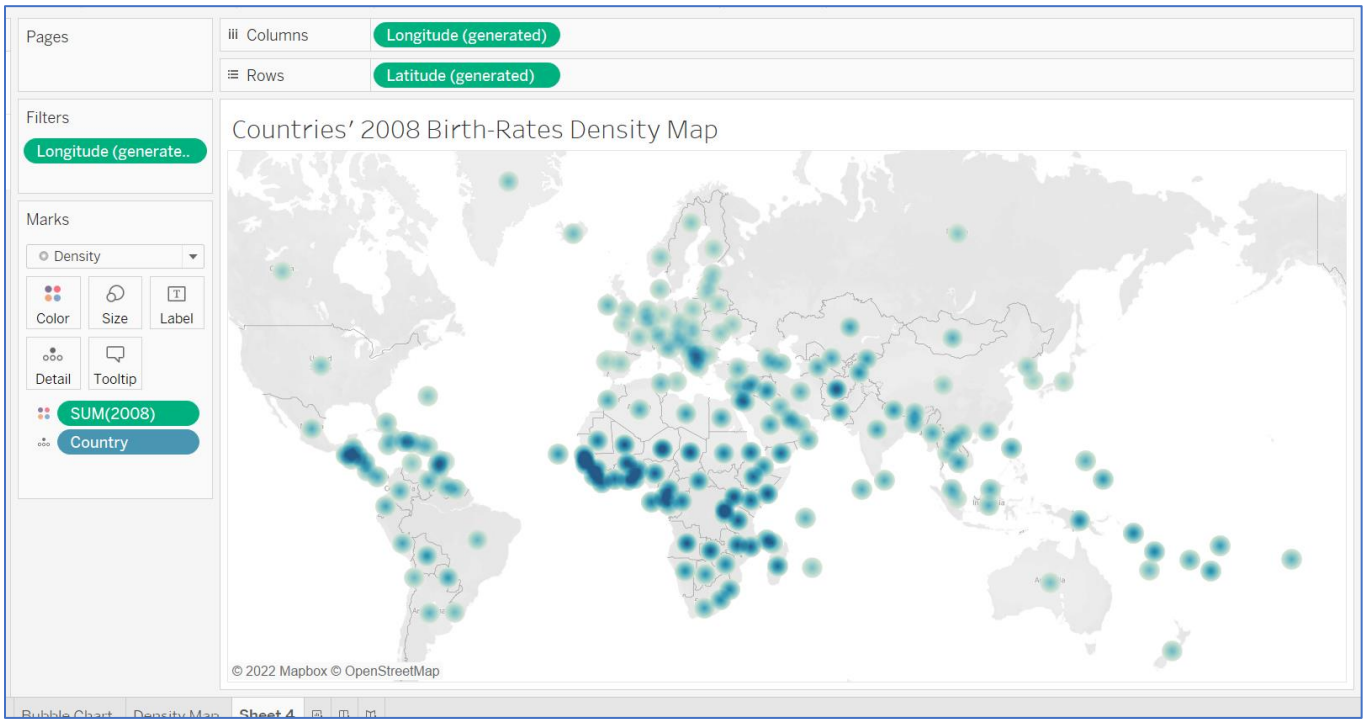


B. Bubble Chart – Tableau



C. Density Plot/Map– Tableau

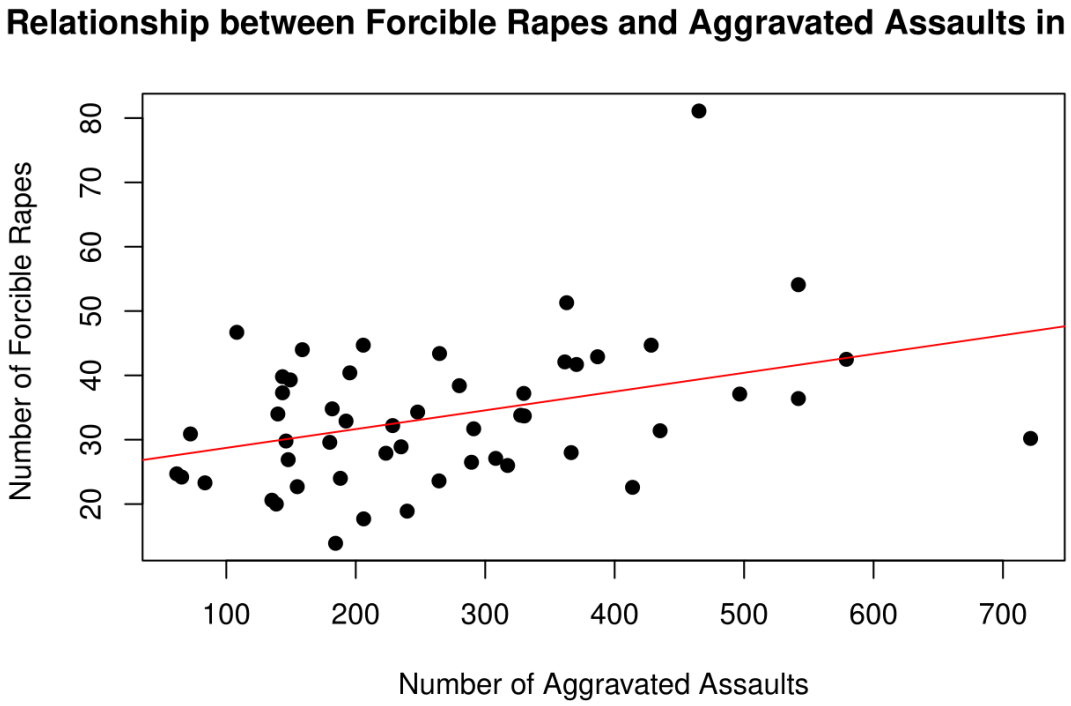




D. Scatter Plot – R

```
#Scatter Plot
```{r scatter}
plot(crime$aggravated_assault, crime$forcible_rape, main="Relationship between Forcible Rapes and Aggravated Assaults in the U.S.",
   xlab="Number of Aggravated Assaults", ylab="Number of Forcible Rapes ", pch=19)

# Add fit lines
abline(lm(crime$forcible_rape~crime$aggravated_assault), col="red") # regression line (y~x)
```
```

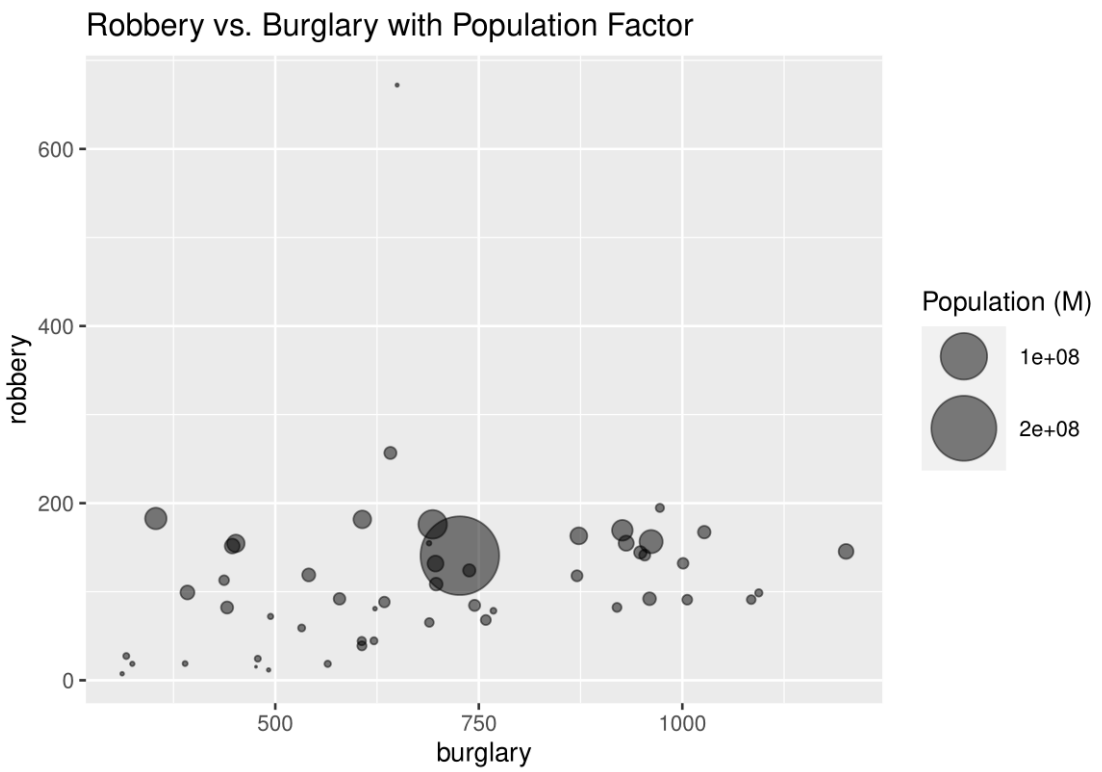


E. Bubble Chart – R

```
library(ggplot2)

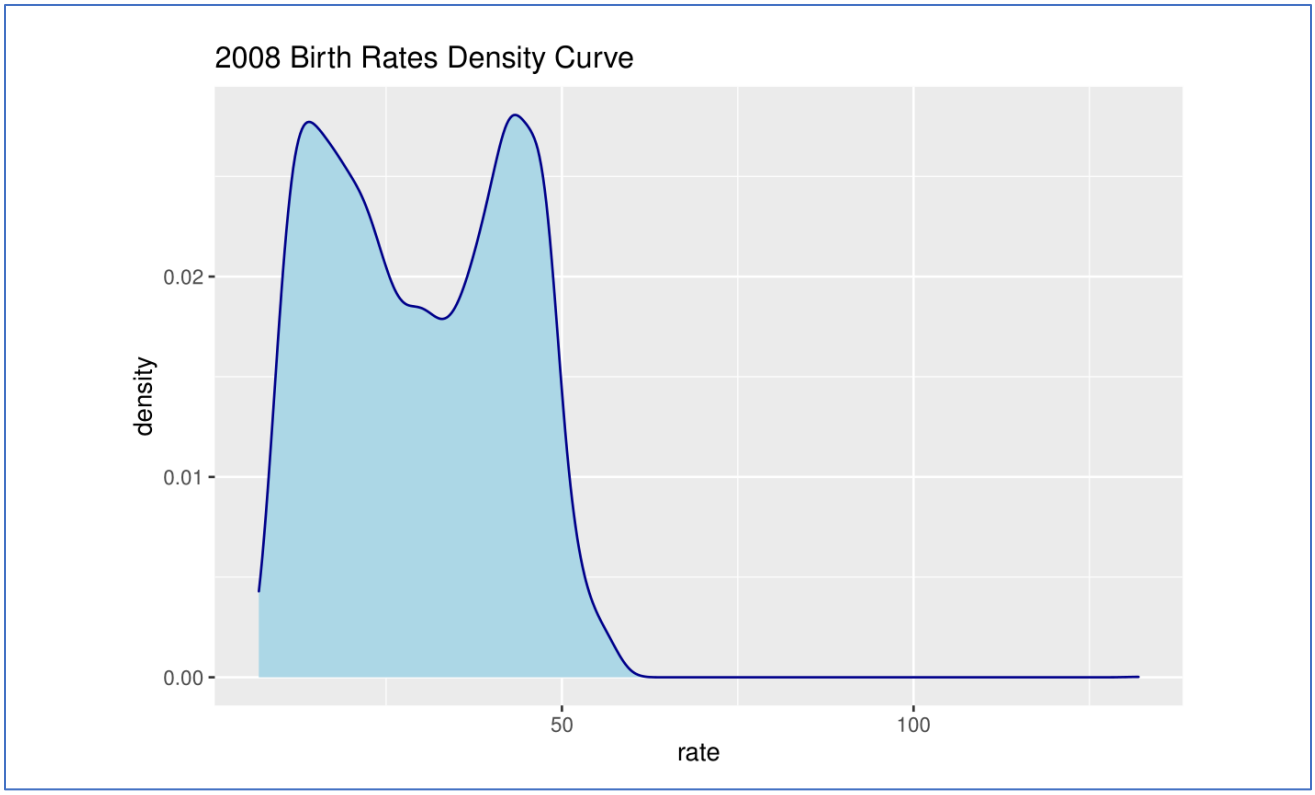
# Most basic bubble plot
#ggplot(crime, aes(x=robbery, y=burglary, size = population)) +
    #geom_point(alpha=0.7)

# Most basic bubble plot
crime %>%
  arrange(desc(population)) %>%
  mutate(state = factor(state, state)) %>%
  ggplot(aes(x=burglary, y=robbery, size = population)) +
    geom_point(alpha=0.5) +
    scale_size(range = c(.1, 15), name="Population (M)") +
  ggtitle("Robbery vs. Burglary with Population Factor")
```



F. Density Plot – R

```
subset_year <- subset(birth_year, year=2008,
select=c(year, rate))
# Basic density plot in ggplot2
ggplot(subset_year, aes(x = rate, colour = year)) +
  geom_density(color="darkblue", fill="lightblue")+ ggtitle("2008 Birth Rates Density Curve")
```

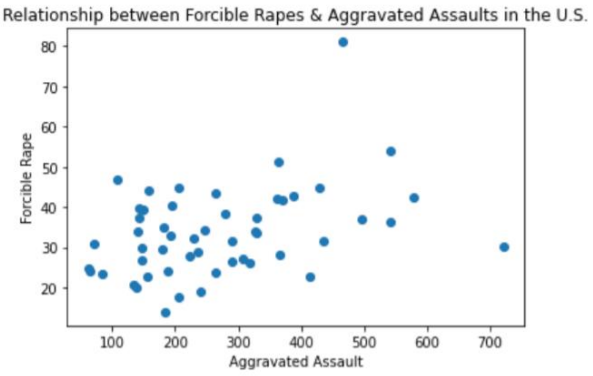


G. Scatter Plot – Python

Scatter Plot

In [32]:

```
#forcible rape values on the y-axis
#aggravated assult values on the x-axis
plt.scatter(crime['aggravated_assault'], crime['forcible_rape'])
plt.xlabel("Aggravated Assault")
plt.ylabel("Forcible Rape")
plt.title("Relationship between Forcible Rapes & Aggravated Assaults in the U.S.")
plt.show()
```

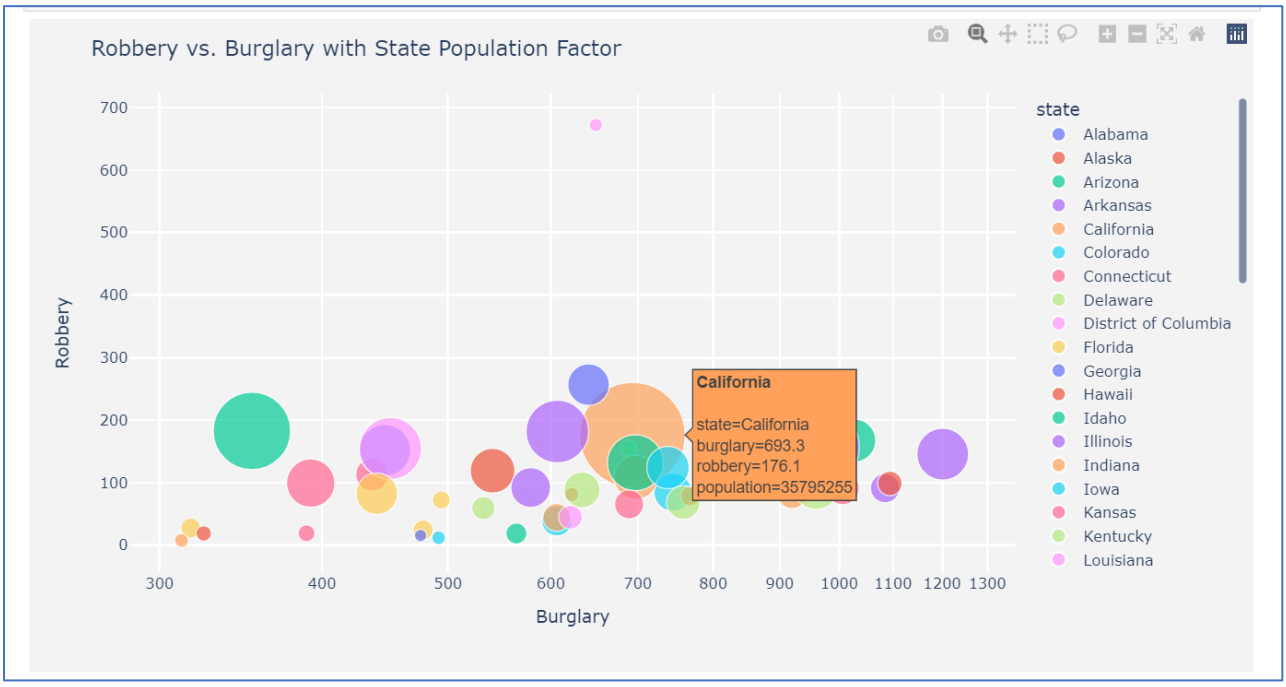


H. Bubble Chart – Python

Bubble Chart

In [12]:
```
import plotly.express as px
```

In [16]:
```
fig = px.scatter(crime_df.query("state!='United States'"), x="burglary", y="robbery",
         size="population", color="state",
                  hover_name="state", log_x=True, size_max=60)
fig.update_layout(
    title='Robbery vs. Burglary with State Population Factor',
    xaxis=dict(
        title='Burglary',
        gridcolor='white',
        type='log',
        gridwidth=2,
    ),
    yaxis=dict(
        title='Robbery',
        gridcolor='white',
        gridwidth=2,
    ),
    paper_bgcolor='rgb(243, 243, 243)',
    plot_bgcolor='rgb(243, 243, 243)',
)
fig.show()
```



I. Density Plot – Python

Density Plot

In [17]:
```
import seaborn as sns
```

In [24]:
```
birth_year_df['year'].unique()
```

Out[24]:
```
array([1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970,
       1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981,
       1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992,
       1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003,
       2004, 2005, 2006, 2007, 2008], dtype=int64)
```

In [30]:
```
#looking into start of different decades
years = [1960,1970,1980,1990,2000]

# Iterate through the designated years
for year in years:
    # Subset to the year
    subset = birth_year_df[birth_year_df['year'] == year]

    # Draw the density plot
    sns.distplot(subset['rate'], hist = False, kde = True,
                 kde_kws = {'linewidth': 3},
                 label = year)

# Plot formatting
plt.legend(prop={'size': 10}, title = 'Year')
plt.title('Density Plot for Birth Rates at Decade Starts')
plt.xlabel('Birth Rate')
plt.ylabel('Density')
```

Out[30]:
```
Text(0, 0.5, 'Density')
```