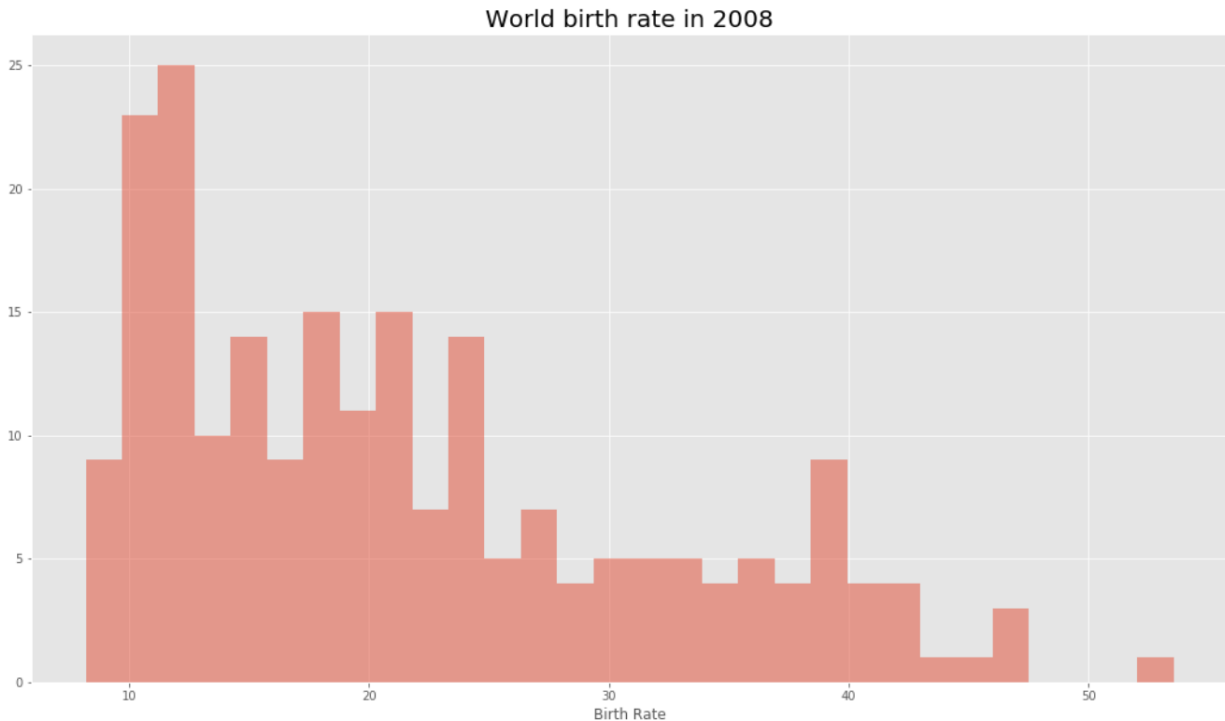


## Histogram – Python

```
# plot a histogram of world birth rate in 2008
x = df['2008']
num_bins = 30

plt.figure(figsize=(18,10))
plt.hist(x, num_bins, alpha=0.5)
plt.title("World birth rate in 2008", fontsize=20)
plt.xlabel("Birth Rate", fontsize=12)
plt.show()
```



### Background Information:

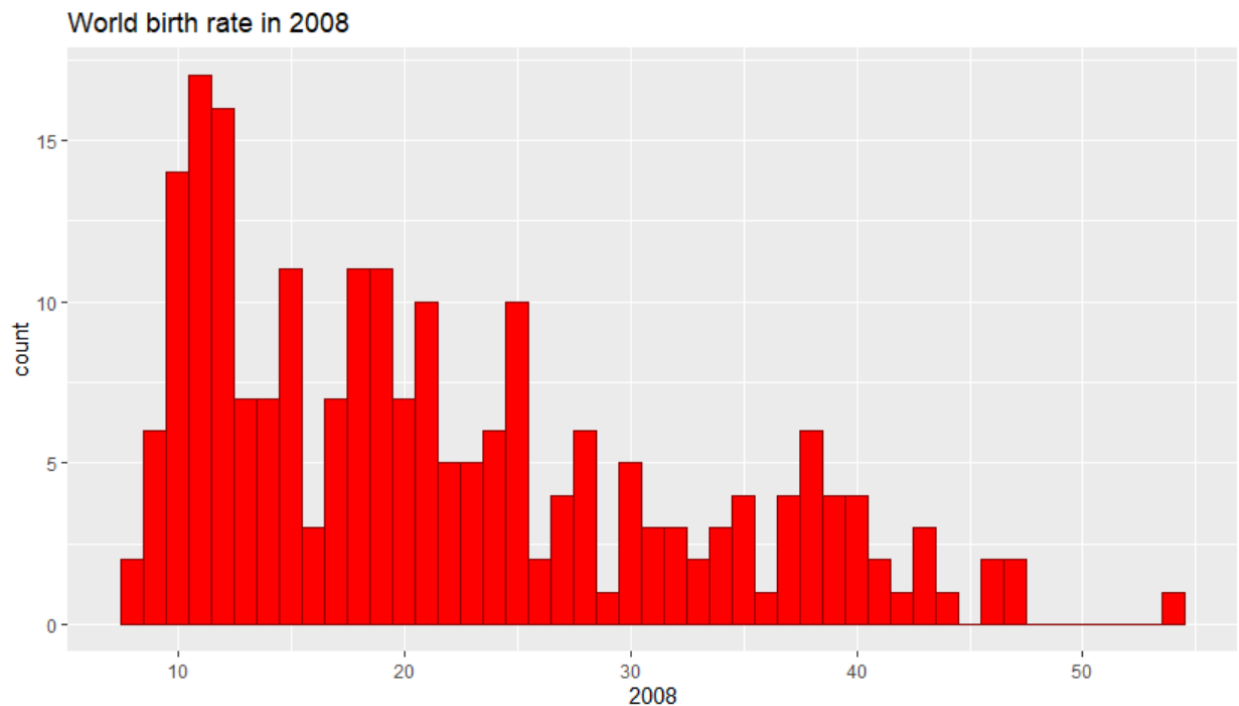
A histogram is created by using Python – matplotlib. The data set is birth-rate. In this data set, we can find the data of the birth rate for all countries in the world. The data is plotted for the year of 2008. The peak of the birth rate occurs at around 12%.

### Method:

- Load the libraries and the dataset into a data frame
- Select the data from a column of the data frame
- Use Matplotlib to plot the histogram with the `plt.hist()` function

## Histogram – R

```
1 # Libraries
2 library(tidyverse)
3 library(hrbrthemes)
4 library(ggplot2)
5
6 # load data
7 df <- read_csv('birth-rate.csv')
8 df1 <- read_csv('crimeratesbystate-formatted.csv')
9 df2 <- read_csv('education.csv')
10
11
12 # plot histogram
13 ggplot(df, aes(x=`2008`)) +
14   geom_histogram(binwidth=1, color='darkred', fill='red') +
15   ggtitle('World birth rate in 2008')
16   theme_ipsum() +
17   theme(
18     plot.title = element_text(size=15)
19   )
```



### Background Information:

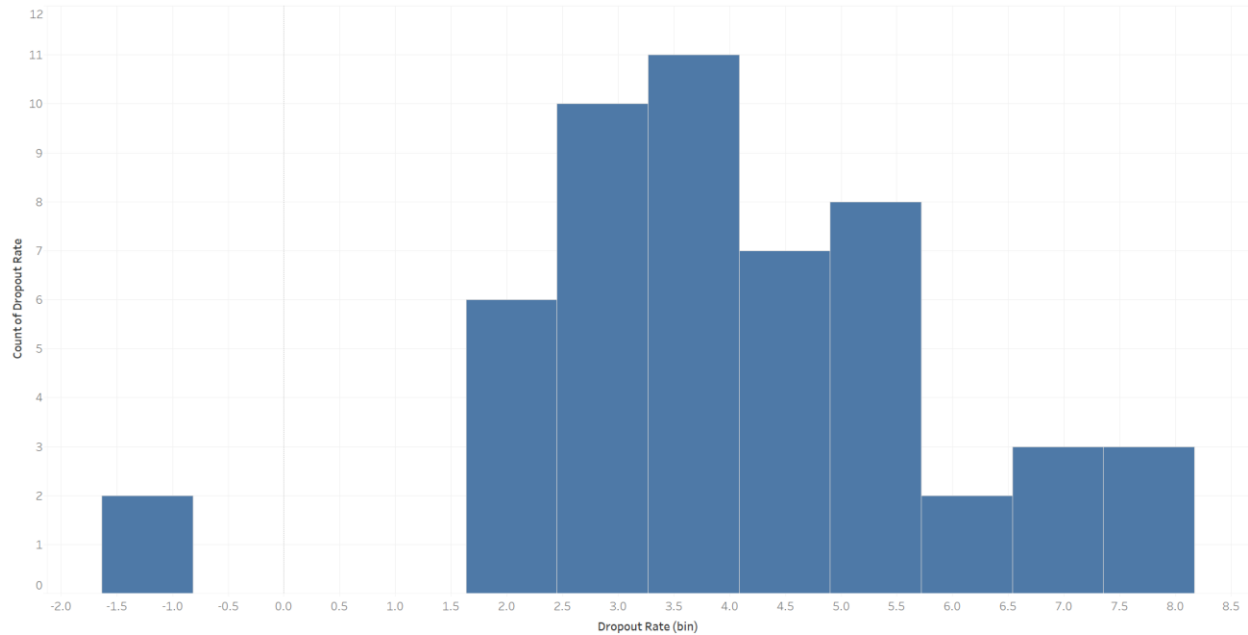
A histogram is created by using R – ggplot2. The data set is birth-rate. In this data set, we can find the data of the birth rate for all countries in the world. The data is plotted for the year of 2008. The peak of the birth rate occurs at around 12%.

### Method:

- Load the libraries and the dataset into a data frame
- Select the data from a column of the data frame
- Use ggplot() and geom\_histogram() to plot the data

## Histogram – Tableau

Dropout Rate in the US



### Background Information:

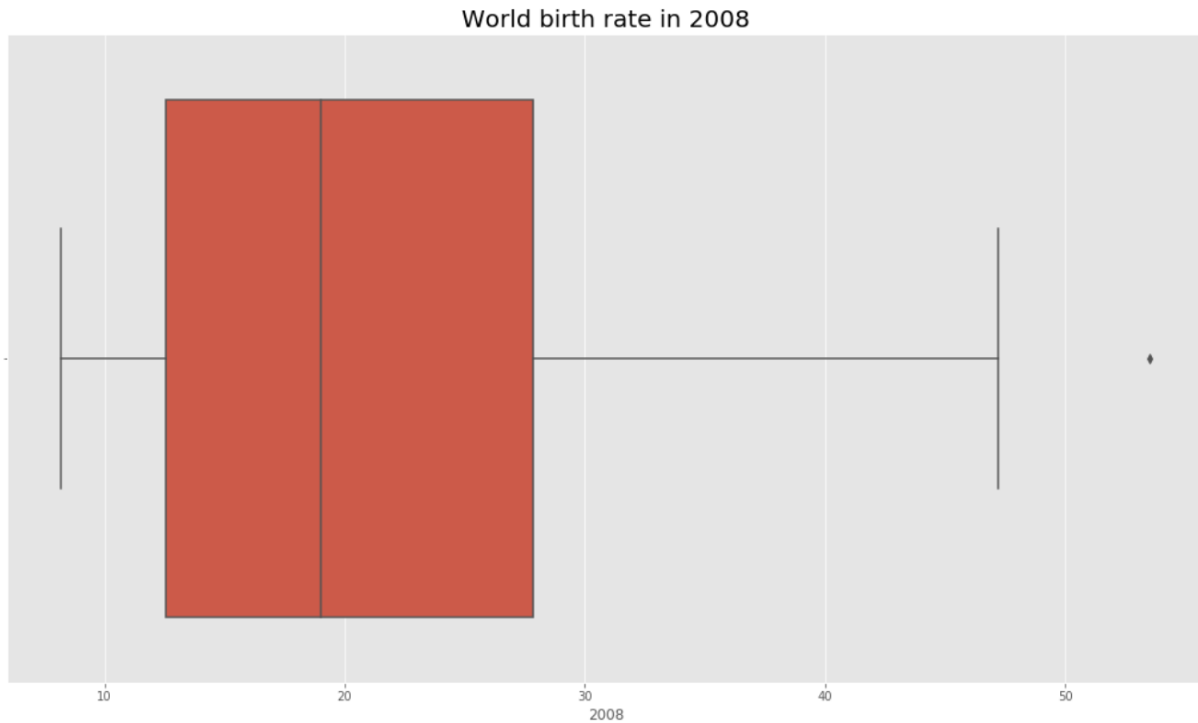
A histogram is created by Tableau. The data set is education. In this data set, we can find the data of many education statistics for different states in the US. The histogram of the dropout rate is plotted. Among different states, the peak of the dropout rate is about 3.5%

### Method:

- Load the data into the datasource
- Drag the DropoutRate to the column
- Select the histogram on the Show Me tab

## Box Plot – Python

```
# plot a box plot
plt.figure(figsize=(18,10))
sns.boxplot(df['2008'])
plt.title('World birth rate in 2008', fontsize=20)
plt.xlabel('2008', fontsize=12)
plt.show()
```



### Background Information:

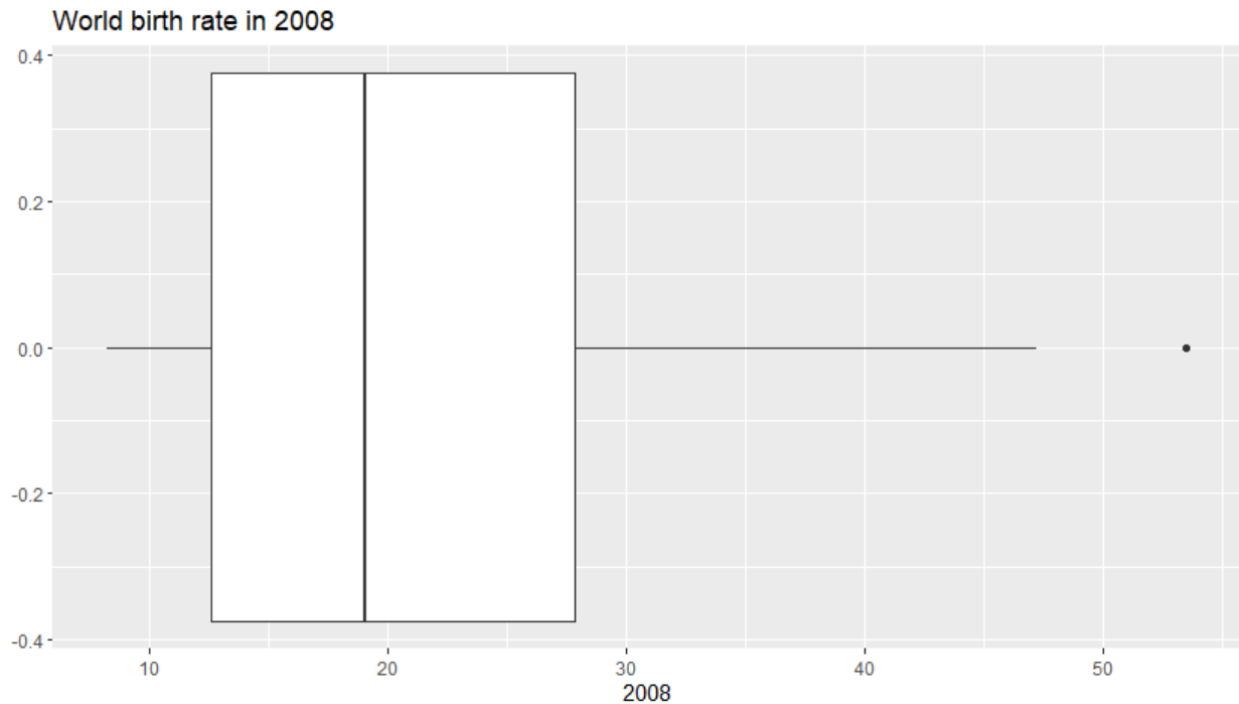
A box plot is created by using Python – seaborn. The data set is birth-rate. In this data set, we can find the data of the birth rate for all countries in the world. The minimum is about 8%, the maximum is about 46%. The Q1 is about 12%, Q3 is about 27%, and the median is about 19%. There is an outlier at 53%.

### Method:

- Load the libraries and the dataset into a data frame
- Select the data from a column of the data frame
- Use Seaborn and Matplotlib with the `sns.boxplot()` function to plot the box plot

## Box Plot – R

```
22 # box plot
23 ggplot(df, aes(x=`2008`)) +
24   geom_boxplot() +
25   xlab("2008") +
26   ggtitle("World birth rate in 2008")
27
```



### Background Information:

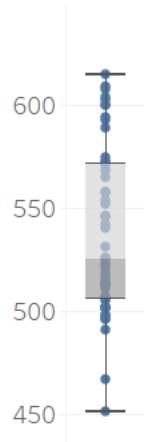
A box plot is created by using R – ggplot2. The data set is birth-rate. In this data set, we can find the data of the birth rate for all countries in the world. The minimum is about 8%, the maximum is about 46%. The Q1 is about 12%, Q3 is about 27%, and the median is about 19%. There is an outlier at 53%.

### Method:

- Load the libraries and the dataset into a data frame
- Select the data from a column of the data frame
- Use ggplot() and geom\_boxplot() to plot the box plot

## Box Plot – Tableau

### Math Score in the US



#### Background Information:

A box plot is created by using Tableau. The data set is education. In this data set, we can find the data of many education statistics. In this box plot, the data of math score is shown. The median across different states is about 530. The Q1 is about 510 and the Q3 is about 570.

#### Method:

- Load the data into the data source
- Drag the math to the Column
- Drag the states to the Detail
- Select the box plot on the Show Me tab

## Bullet Chart – Python

```
limits = [1, 7, 36]
data_to_plot = ('Murder in US', df1['murder'])
palette = sns.color_palette("Reds_r", len(limits))

fig, ax = plt.subplots()
ax.set_aspect('equal')
ax.set_yticks([1])
ax.set_yticklabels([data_to_plot[0]])

prev_limit = 0
for idx, lim in enumerate(limits):
    ax.barh([1], lim-prev_limit, left=prev_limit, height=15, color=palette[idx])
    prev_limit = lim
```

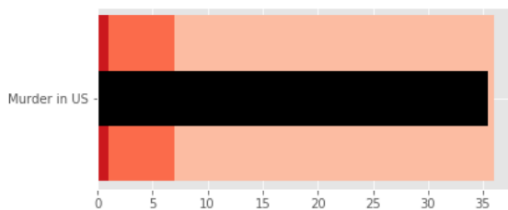
...

```
fig, ax = plt.subplots()
ax.set_aspect('equal')
ax.set_yticks([1])
ax.set_yticklabels([data_to_plot[0]])

prev_limit = 0
for idx, lim in enumerate(limits):
    ax.barh([1], lim-prev_limit, left=prev_limit, height=15, color=palette[idx])
    prev_limit = lim

# Draw the value we're measuring
ax.barh([1], data_to_plot[1], color='black', height=5)

<BarContainer object of 52 artists>
```



### Background Information:

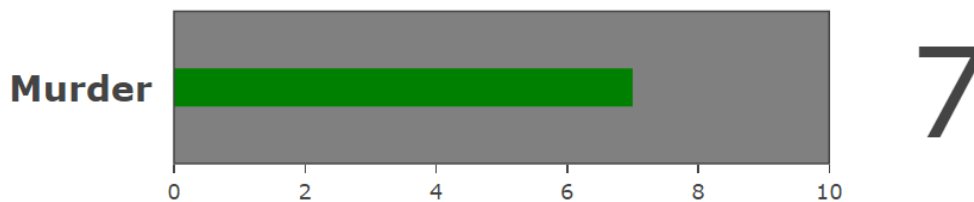
A bullet chart is created by using Python – plotly. The data set is `crimeratesbystate`-formatted. In this data set, we can find the data of crimes in different states in US. The minimum number of murder is 1.1 as indicated by the red bar. The maximum number is 35.4 as indicated by the black bar. The average number of murder is 7 as indicated by the orange bar.

### Method:

- Load the libraries and the dataset into a data frame
- Select the data from a column of the data frame
- Use plotly and the `barh()` function to plot the bullet chart

## Bullet Chart – R

```
29 library(plotly)
30
31 fig <- plot_ly(
32   type = "indicator",
33   mode = "number+gauge",
34   value = 7,
35   domain = df1$murder,
36   title = list(text = "<b>Murder</b>"),
37   gauge = list(
38     shape = "bullet",
39     axis = list(range = list(NULL, 10)),
40     threshold = list(
41       line = list(color = "red", width = 2),
42       thickness = 0.75,
43       value = 15),
44     steps = list(
45       list(range = c(0, 10), color = "lightgray"),
46       list(range = c(0, 10), color = "gray"))),
47   height = 150, width = 600)
48 fig <- fig %>%
49   layout(margin = list(l= 100, r= 10))
50
51 fig
```



### Background Information:

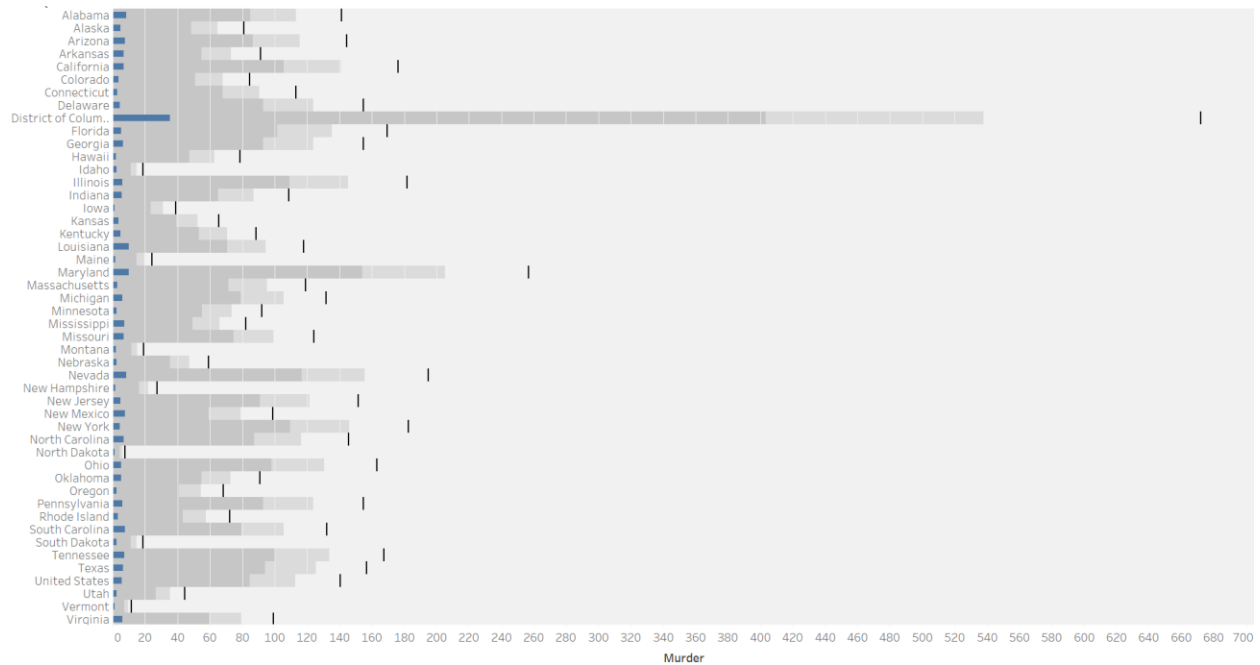
A bullet chart is created by using R – plotly. The data set is `crimeratesbystate`-formatted. In this data set, we can find the data of crimes in different states in US. The average number of murder across the states is 7.

### Method:

- Load the libraries and the dataset into a data frame
- Select the data from a column of the data frame
- Use `plot_ly()` function to plot the bullet chart



## Bullet Chart – Tableau



### Background Information:

A bullet chart is created by using Tableau. The data set is `crimeratesbystate`-formatted. In this data set, we can find the data of crimes in different states in US. In this bullet chart, the data for murder is plotted with the data for robbery.

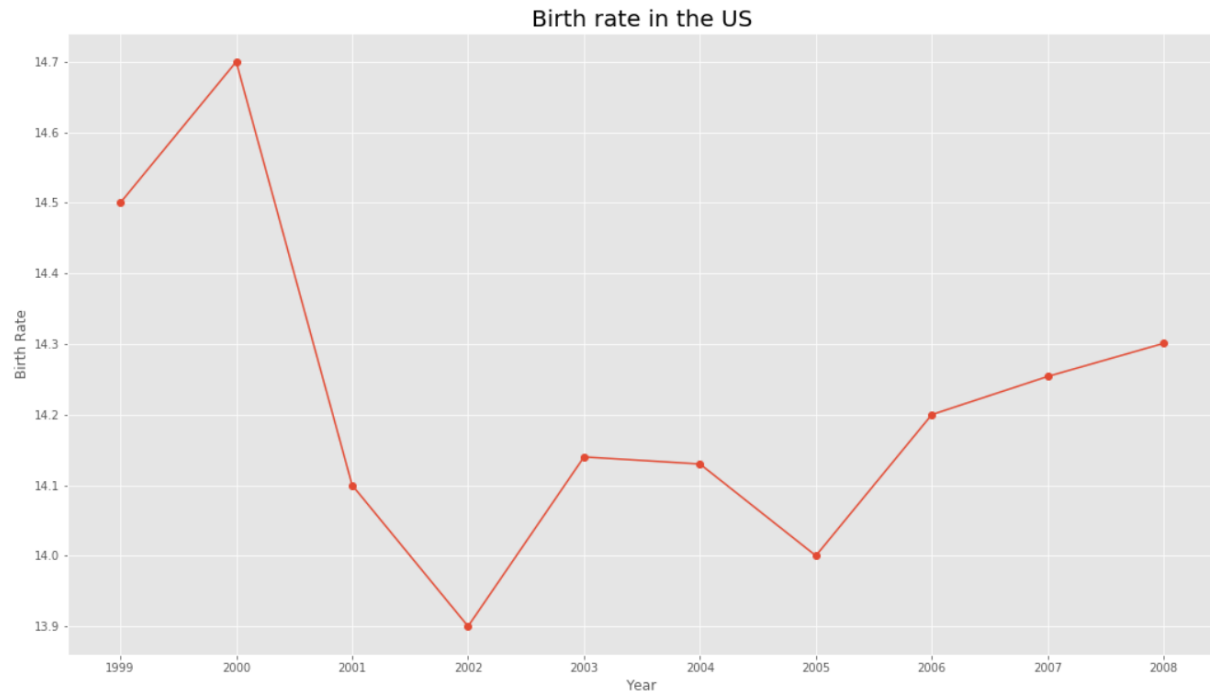
### Method:

- Load the data into the data source
- Drag the Murder into the Columns
- Drag the Robbery into the Columns
- Drag the State into the Rows
- Select Bullet Chart on the Show Me tab

## Scatter Plot with connected lines – Python (Option)

```
x = df.iloc[221].drop(['Country'])  
x = x[39:]
```

```
plt.figure(figsize=(18,10))  
plt.plot(x, marker='o')  
plt.title("Birth rate in the US", fontsize=20)  
plt.xlabel("Year", fontsize=12)  
plt.ylabel("Birth Rate", fontsize=12)  
plt.show()
```



### Background Information:

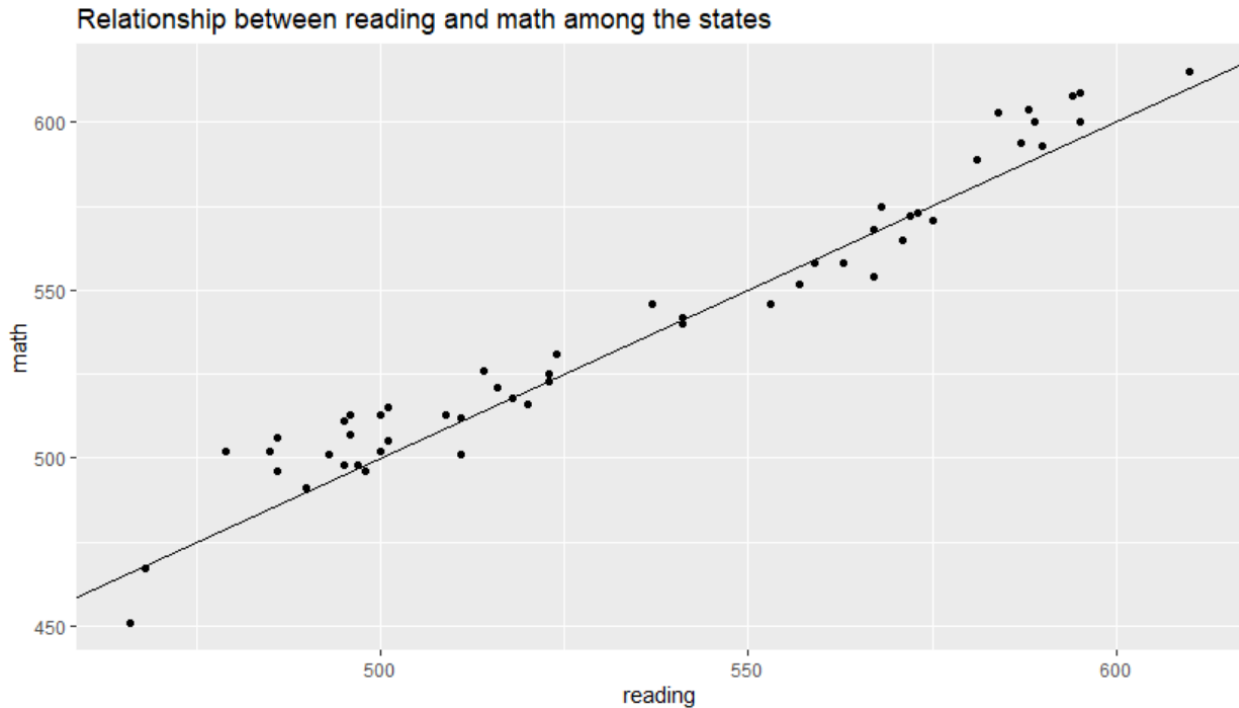
A scatter plot is created by using Python – matplotlib. The data set is birth-rate. In this data set, we can find the data of birth rates for different countries in the world. In this plot, the data for the birth rate in the US from 1999 to 2008 are shown. It appears the birth rate drops the most in 2002 after a peak in 2000.

### Method:

- Load the libraries and the dataset into a data frame
- Select the data from a column of the data frame
- Use matplotlib and the plot() function to plot the scatter plot with a connected line

### Scatter Plot – R (Option)

```
54 # plot reading and math in us
55 ggplot(df2, aes(x=reading, y=math))+
56   geom_point() +
57   geom_abline() +
58   ggtitle("Relationship between reading and math among the states")
59
```



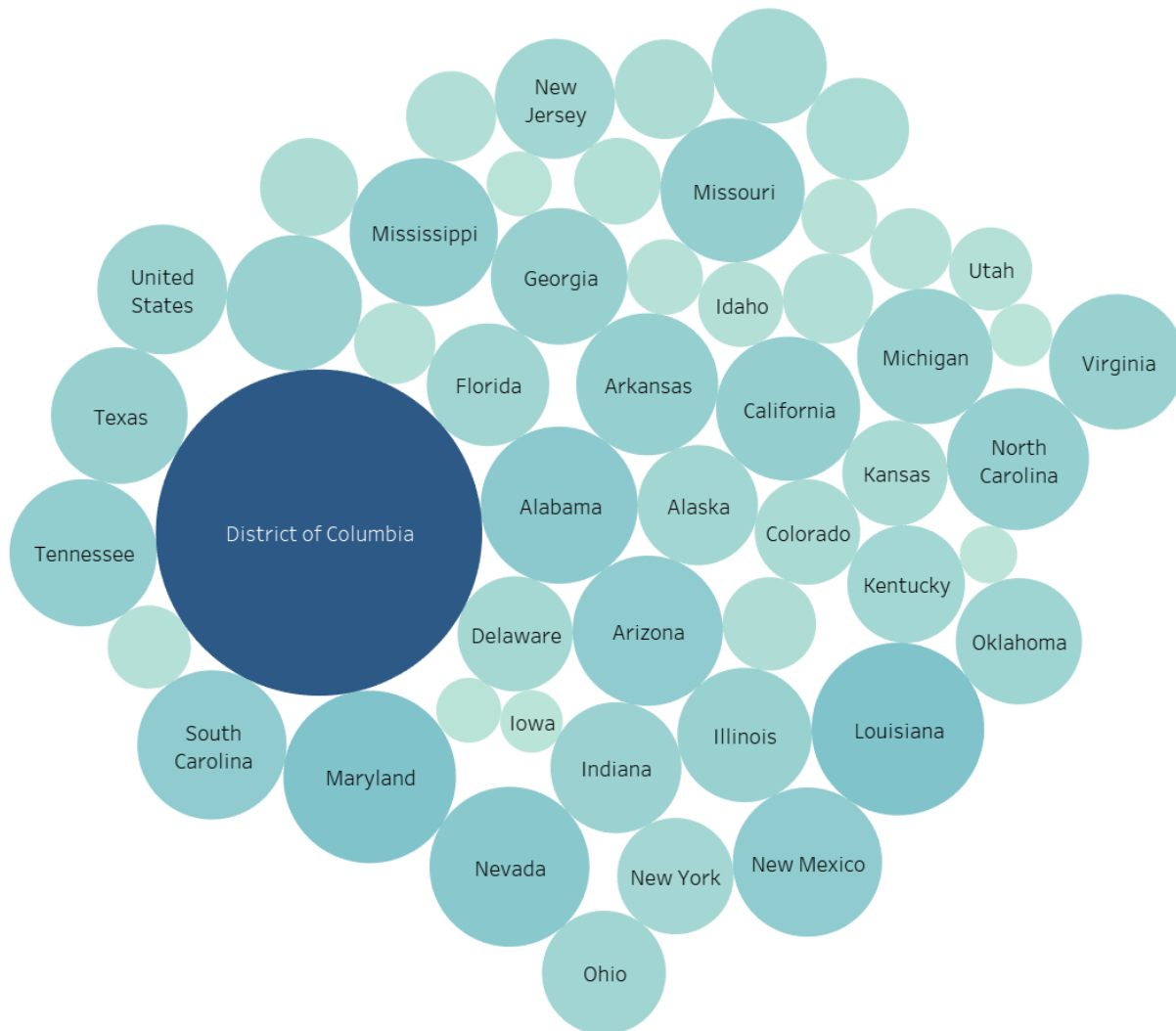
#### Background Information:

A scatter plot is created by using R – ggplot. The data set is education. In this data set, we can find the data of many education statistics for different states in the US. In this plot, the data of reading and math are plotted to show how their relationship looks like. As shown in the plot, the reading and math seems to have a linear relationship.

#### Method:

- Load the libraries and the dataset into a data frame
- Select the data from a column of the data frame
- Use ggplot(), geom\_point() and geom\_line() function to plot the scatter plot

## Bubble Chart – Tableau (Option)



### Background Information:

A bubble chart is created by using Tableau. The data set is `crimeratesbystate-formatted`. In this data set, we can find the data of crimes in different states in US. In this bubble chart, the data for murder is plotted. The states with the bigger bubble size have higher murder count.

### Method:

- Load the data into the data source
- Drag the Murder into the Color
- Drag the State into the Label
- Drag the Murder into the Size