MIS 0855 – Data Science (Section 006) – Fall 2017
In-Class Exercise (Day 25-26) – Advanced Data Visualization with Tableau

Objective: Learn how to use Tableau's advanced data visualization tools

Learning Outcomes:

- Use advanced visualization toolkits in Tableau for more effective presentation of data analyses

- Understand the purpose and effective use of the following tools – dual line chart, moving average, box plot, bubble chart, bullet graph, histogram, and pareto chart.


Part 0. Understand the Dataset

In this exercise, we will use the crime statistics in large U.S. cities for 2009-2014. This dataset is obtained from the Uniform Crime Reports (UCR) published by the FBI (https://ucr.fbi.gov/crime-in-the-u.s/2014/crime-in-the-u.s.-2014). This file includes the number of crime occurrences per population (crime rates) and the number of police officers killed or assaulted in the line of duty.

1) Download "Crimes 1999-2014.xlsx" and save it to your computer. Remember where you saved it!

2) Open the file, and browse the data.

3) Take a look at Data Dictionary tab, and understand what each crime data is for.
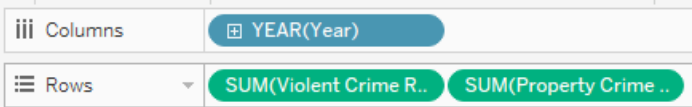

Part 1. Dual Line Chart

1) Start Tableau.

2) Click on "Microsoft Excel" under "In a file."

3) Navigate to the location where your data file ("Crimes 1999-2014.xlsx") is stored and select it.

4) Drag the "Crimes 1999-2014" sheet to the workspace.

5) Click "Sheet 1" to Go to Worksheet.

6) For us to draw a line chart, there must be at least one date-type dimension. If you look at Dimensions panel, Year dimension is designated as a text field. Right-click Year and select "Change Data Type" > "Date."
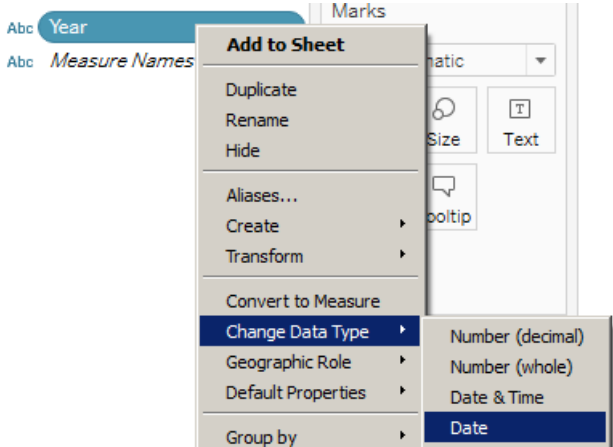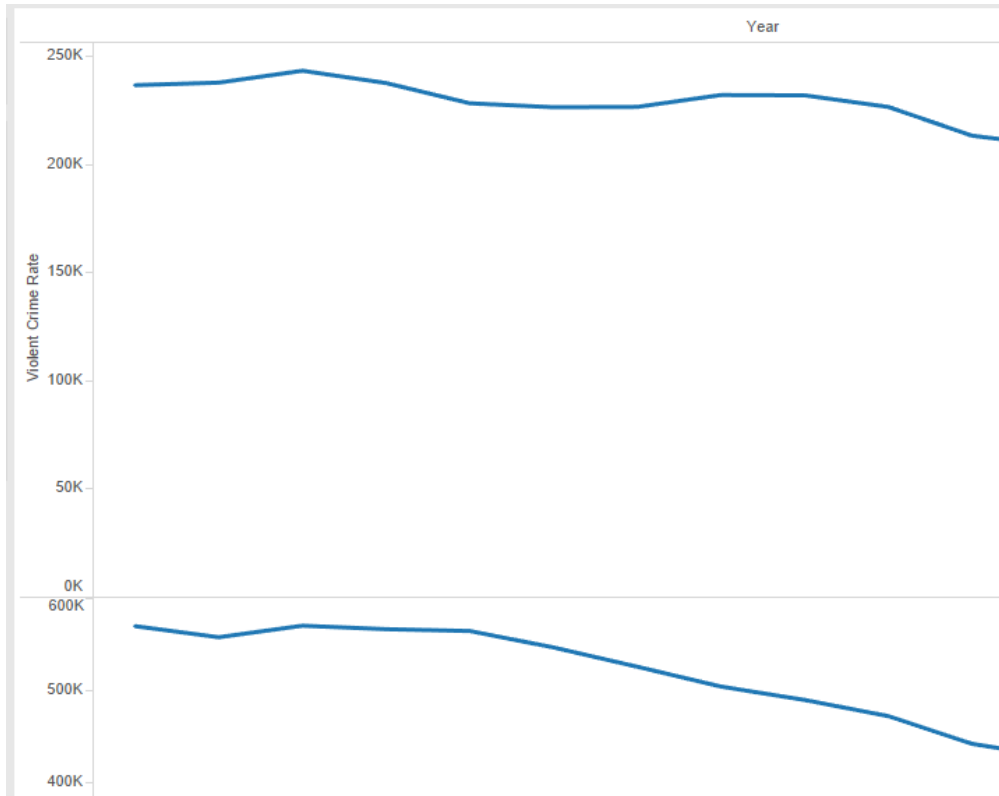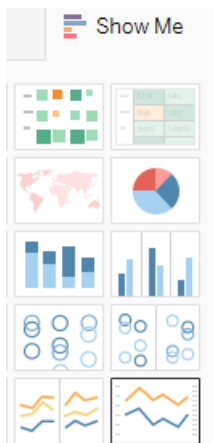

You will see a date ▦ icon next to Year.

7) Drag Year from Dimensions to Columns, and drag "Violent Crime Rate" and "Property Crime Rate" from Measures to Rows.
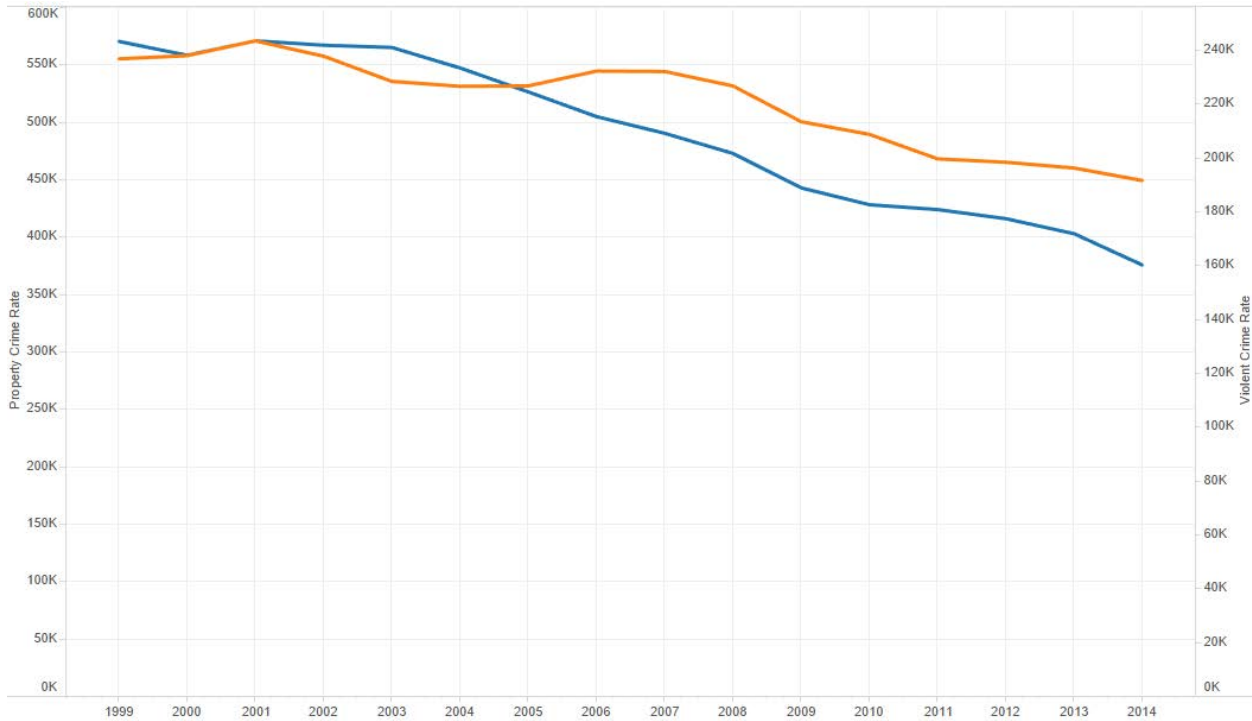
You will see two separate line charts – the top one for the violent crime rate and the bottom one for the property rate.
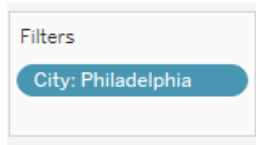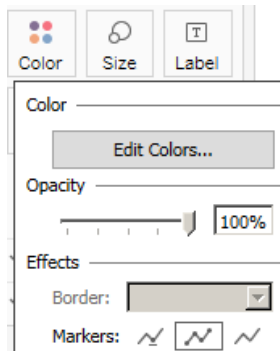

8) Click "Dual Line Chart" on Show Me tab.

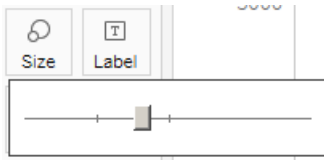You will see that the two charts are combined into one.

9) We'd like to know the crime trends in the city where we live – Philadelphia. Drag City from Dimensions to Filters and check Philadelphia.
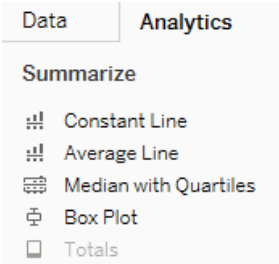

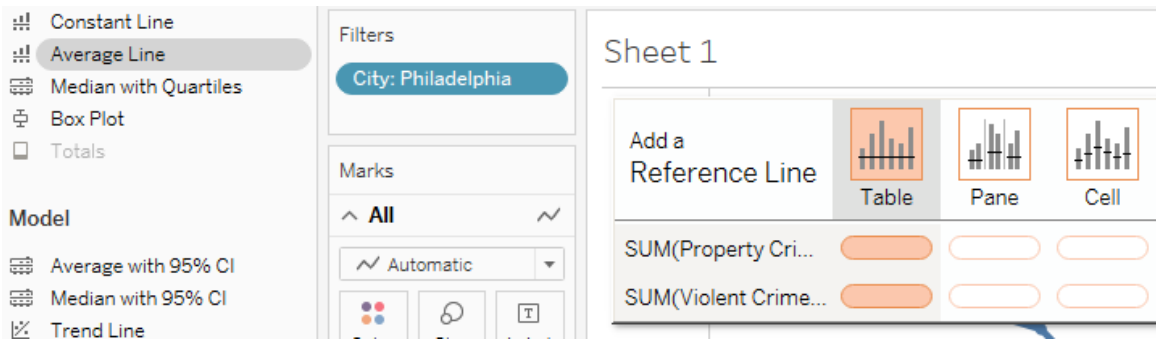10) On Marks panel, click Color and select a dotted line.

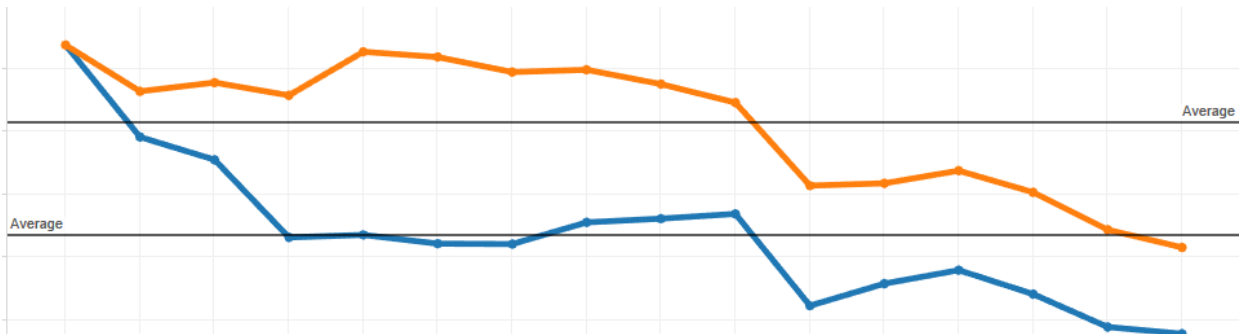11) Right next to Color, click Size, and you can make the lines thicker.

12) Click Analytics tab next to Data.

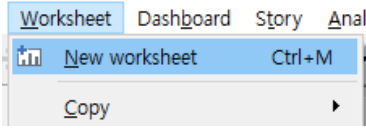13) Drag Average Line from Analytics to "Table."

Now you'll see that the chart now displays the average crime rates in 1999-2014, which demonstrates that Philadelphia has become safer to live in recent years.

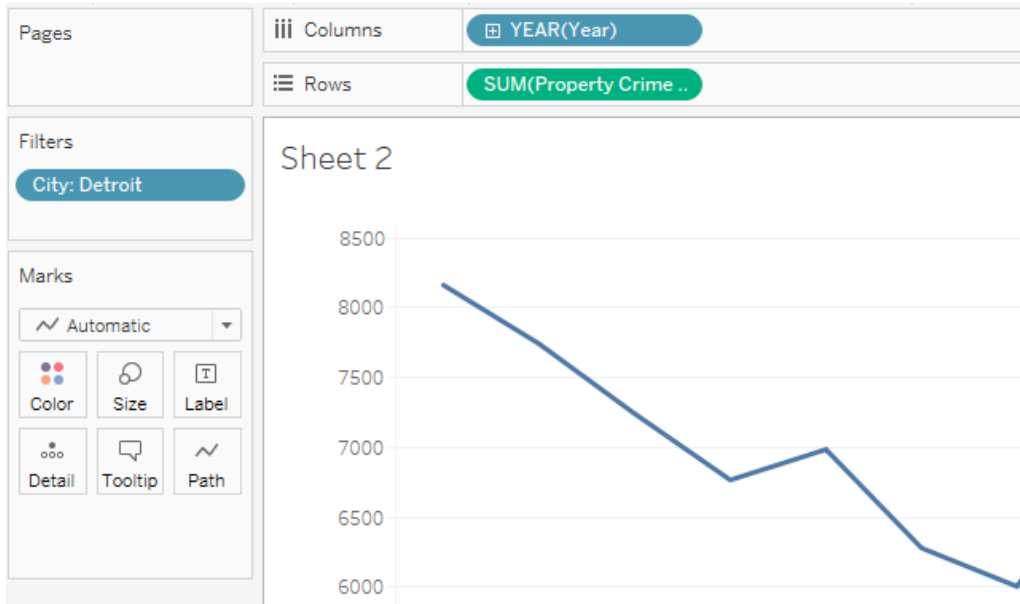14) Rename "Sheet 1" to "Dual Line Charts" and save Tableau.

Part 2. Moving Average

1) Create a new worksheet.

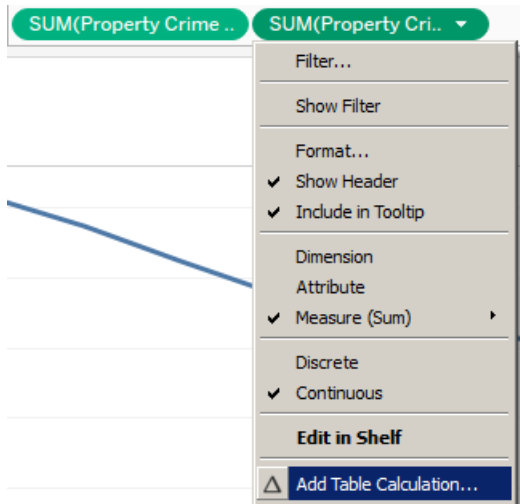2) Drag Year to Columns and Property Crime Rates to Rows.

3) Drag City to Filters and Select "Detroit." The chart will show the trend in the property crime in Detroit in 1999-2014.
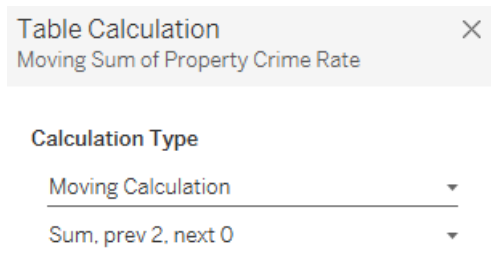

4) Drag Property Crime Rates one more time to Rows. You will see the same two line charts.
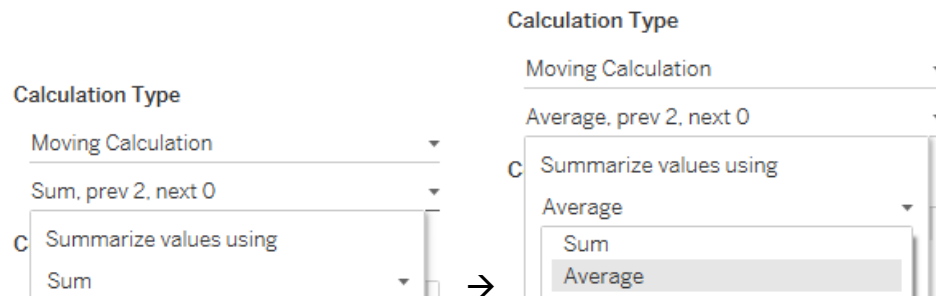
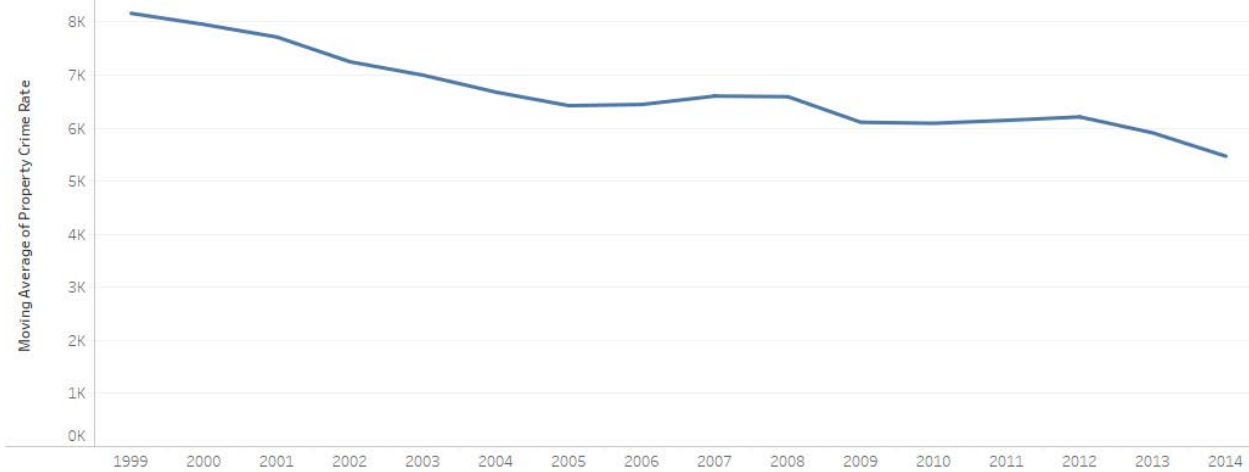5) Select the second SUM(Property Crime Rates) on Rows and click "Add Table Calculation."

6) Select "Moving Calculation" in Calculation Type.

7) Click "Sum, prev 2, next 0," click "Sum," and select "Average."

8) You will see a smoother line chart on the bottom, which shows the average of crime rates in a moving three-year window. For example, for 2007, the chart shows the average of crime rates in 2005, 2006, and 2007.
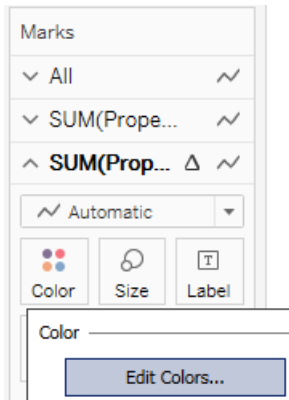

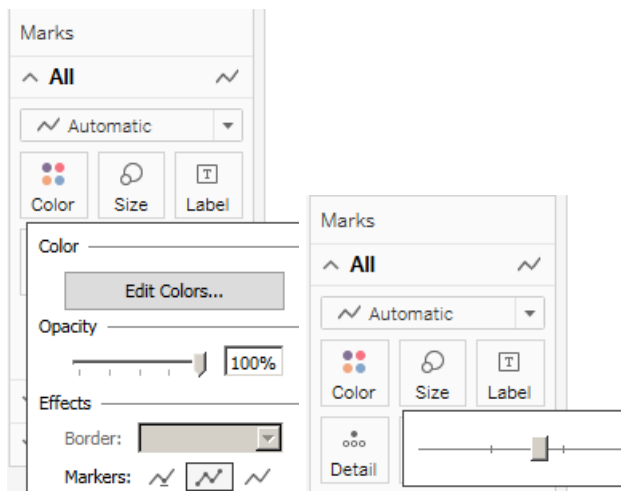9) Click the second "Property Crime Rates" on Rows and select "Dual Axis."
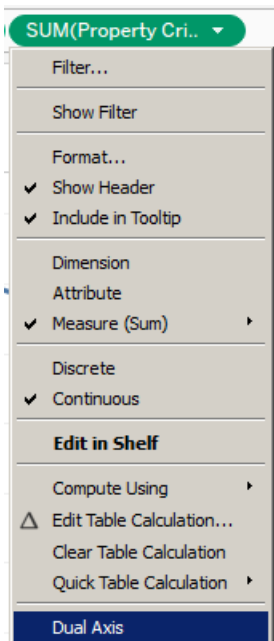
10) Over Marks, click the second SUM(Property Crime Rates), select Color, and click "Edit Color."
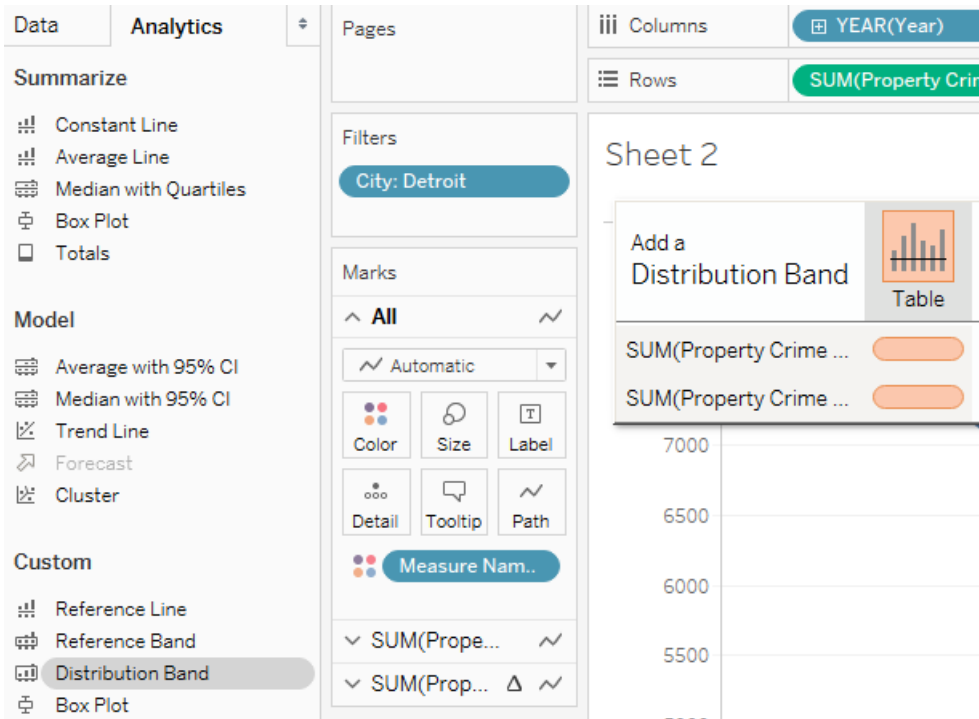

11) Click "Moving Average of Property Crime Rates" and click the green box.
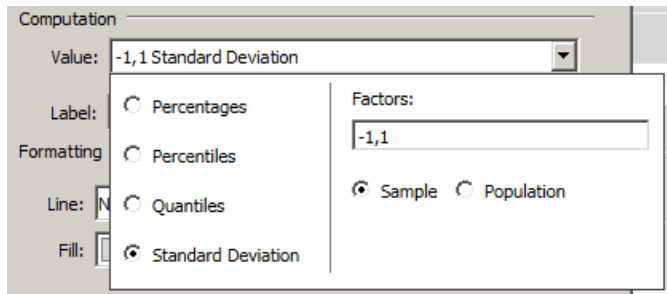

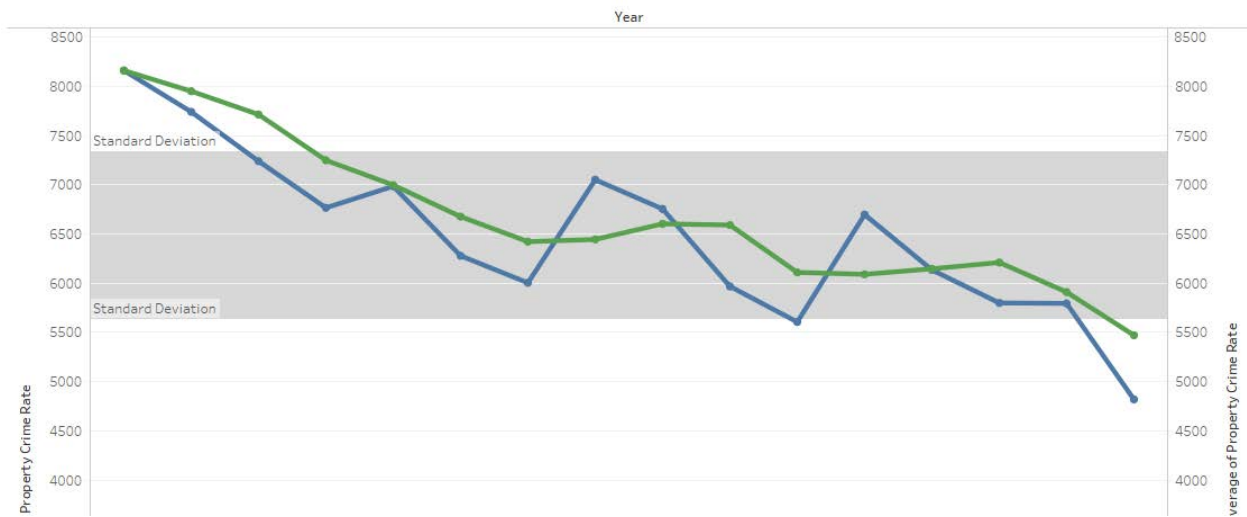12) Under Marks, click "All" and change the market and its size.

13) Finally, click Analytics next to Data and drag "Distribution Band" over the chart.

14) Under "Computation," click "Value" and select "Standard Deviation."

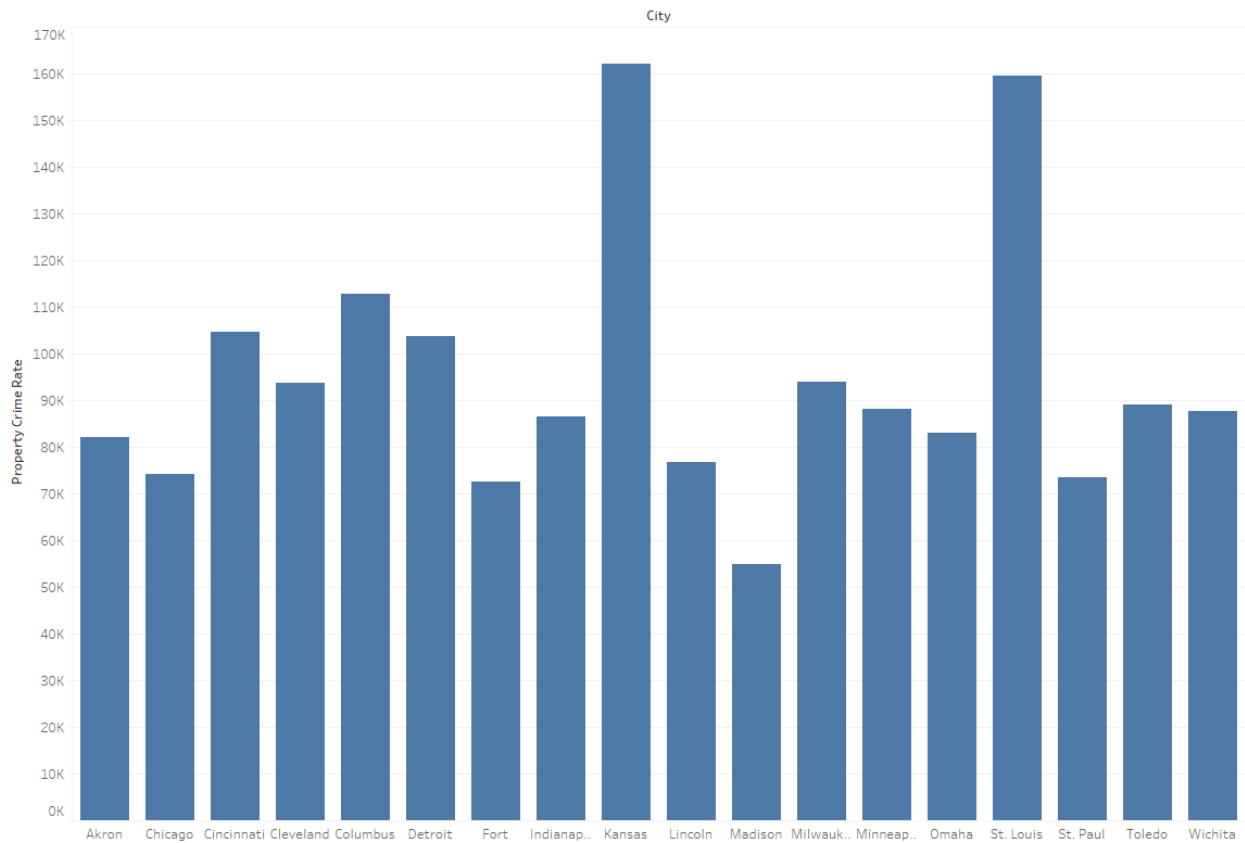15) Now the chart shows the range where most values of crime rates falls into.

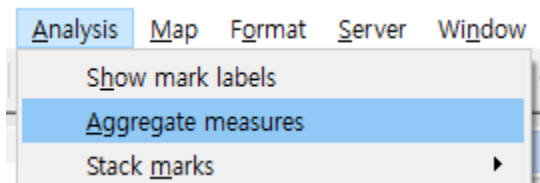
16) Rename "Sheet 2" to "Moving Average" and save.

Part 3. Box Plot

1) Create a new worksheet.

2) Drag City from Dimensions to Columns, and drag Property Crime Rate from Measures to Rows.

3) Let's focus on the cities in Midwest. Drag Region from Dimensions to Filter. Select East Midwest and West Midwest.
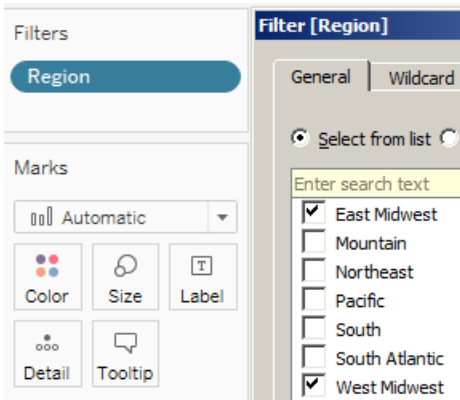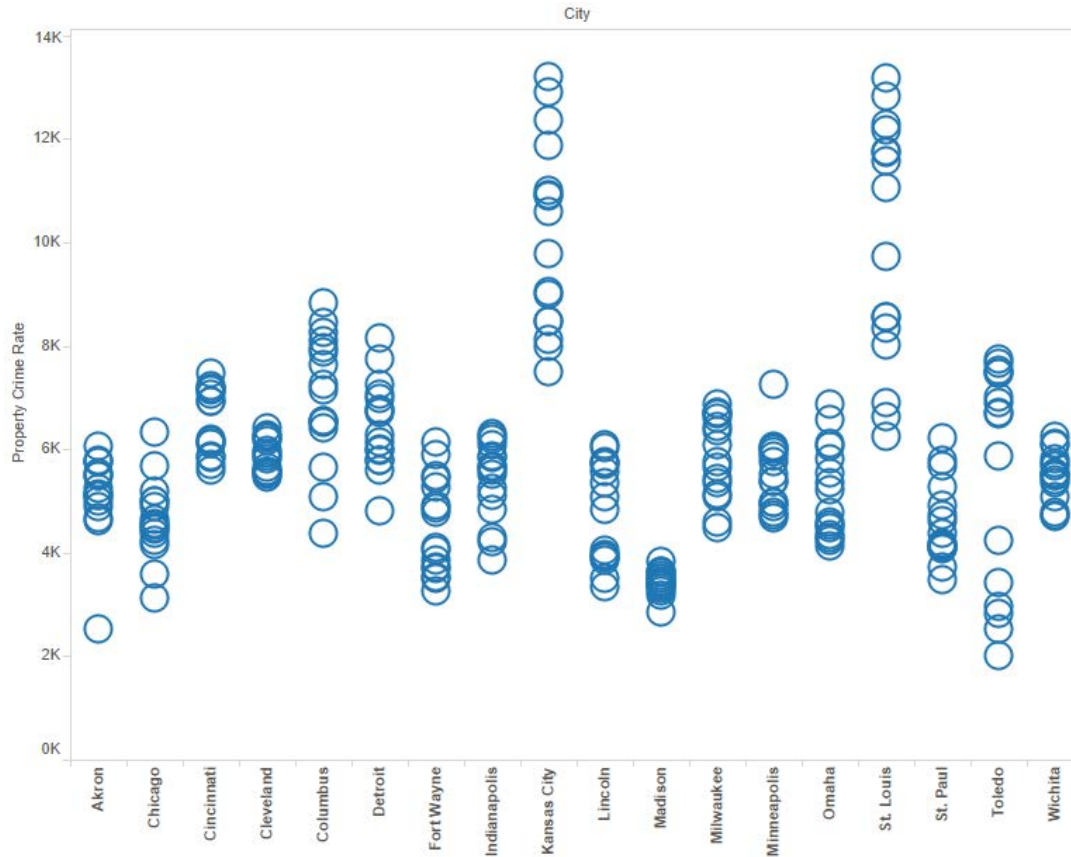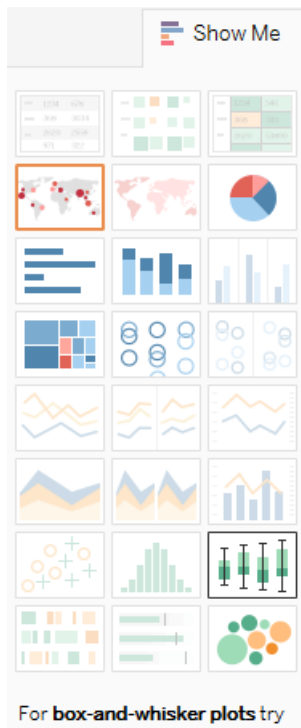
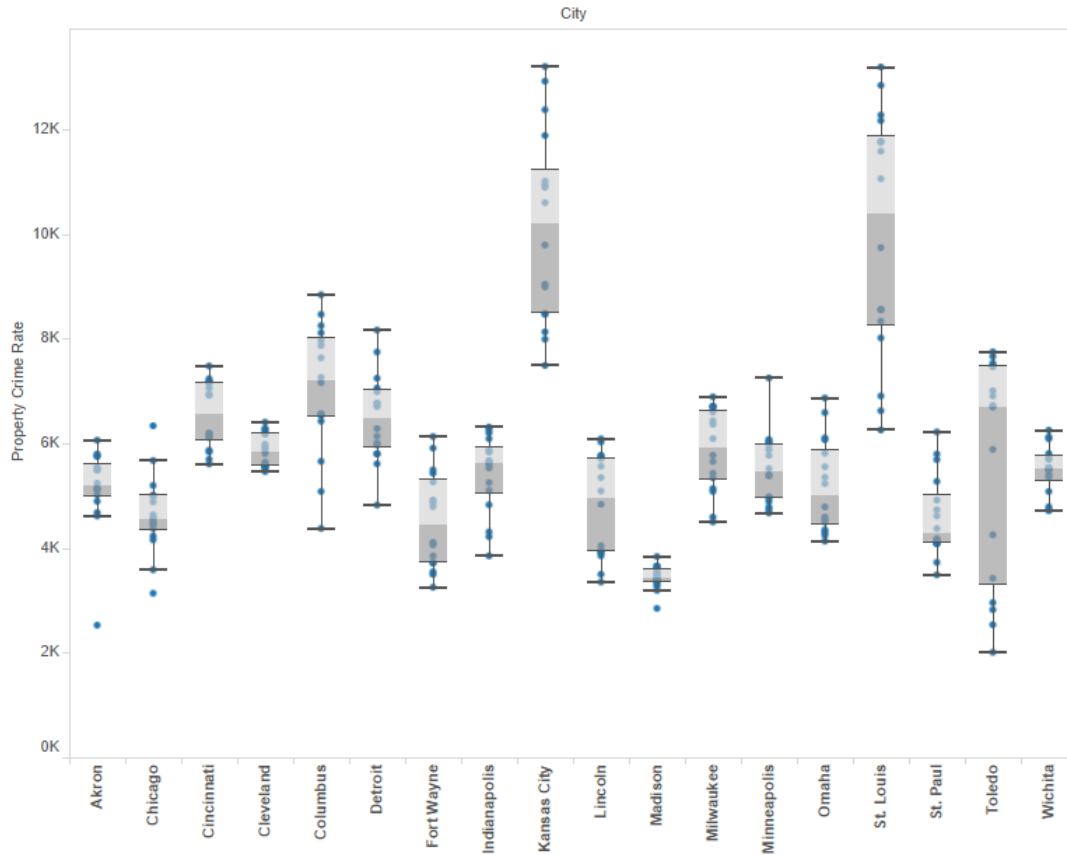4) On the top menu, click Analysis and uncheck Aggregate Measures.

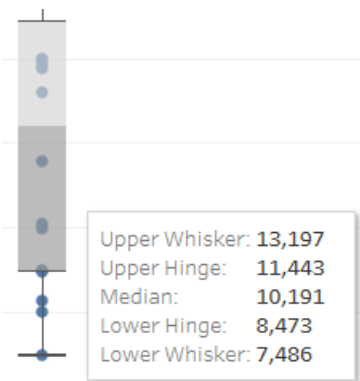You'll see the property crime rate in each year of 1999-2014 appears as a circle.

5) On Show Me tab, click "box-and-whisker plot."

For **box-and-whisker plots** try

6) Move the mouse over a box, which shows the range of property crime rates in each city in 1999-2014 (minimum, maximum, 25%, 50% and 75% quartiles).
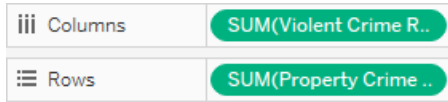

| Upper Whisker: | 13,197 |
Upper Hinge:	11,443
Median:	10,191
Lower Hinge:	8,473
Lower Whisker:	7,486

7) Rename "Sheet 3" to "Box Plot" and save.

Part 4. Bubble Chart

1) Create a new worksheet.

2) Drag Violent Crime Rate from Measures to Columns, and drag Property Crime Rate to Rows.

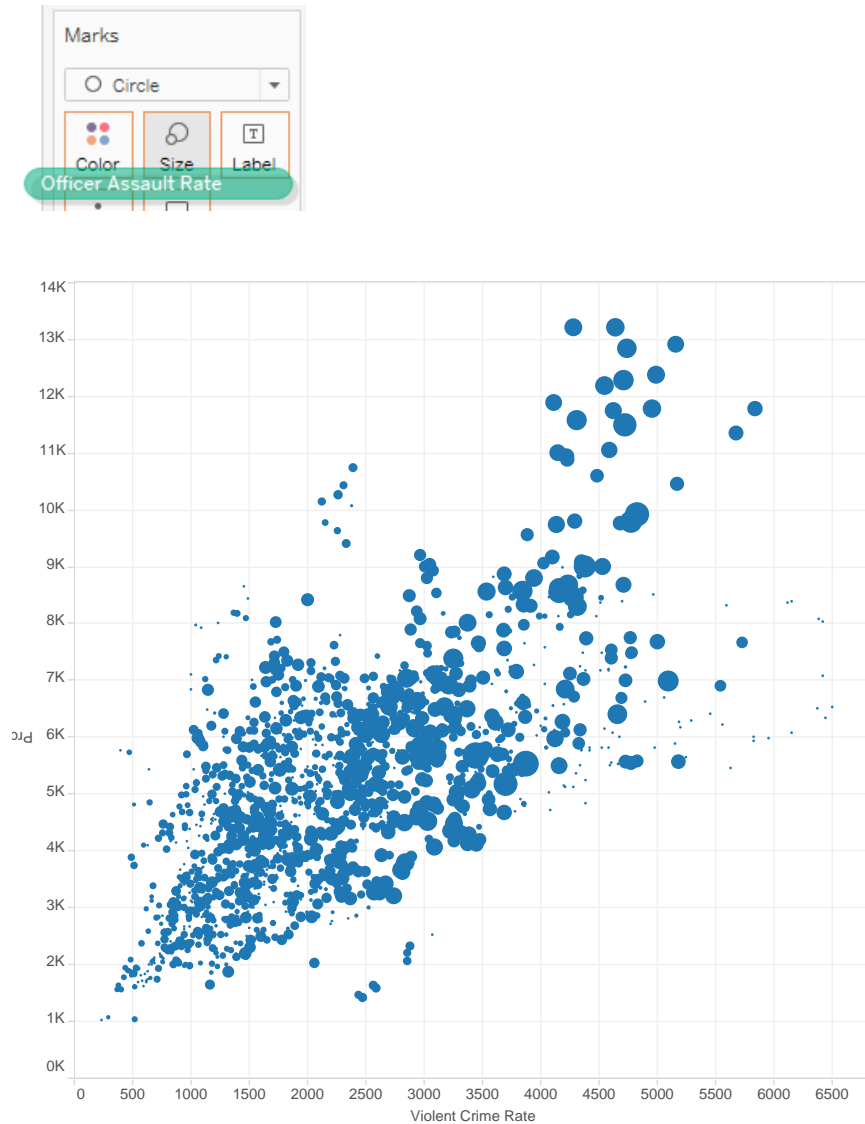| iii Columns	SUM(Violent Crime R..
☰ Rows	SUM(Property Crime ..

3) On the top menu, click Analysis and uncheck Aggregate Measures, which shows a scatter plot of violent and property crime rates. As expected, the two types of crimes are highly correlated.
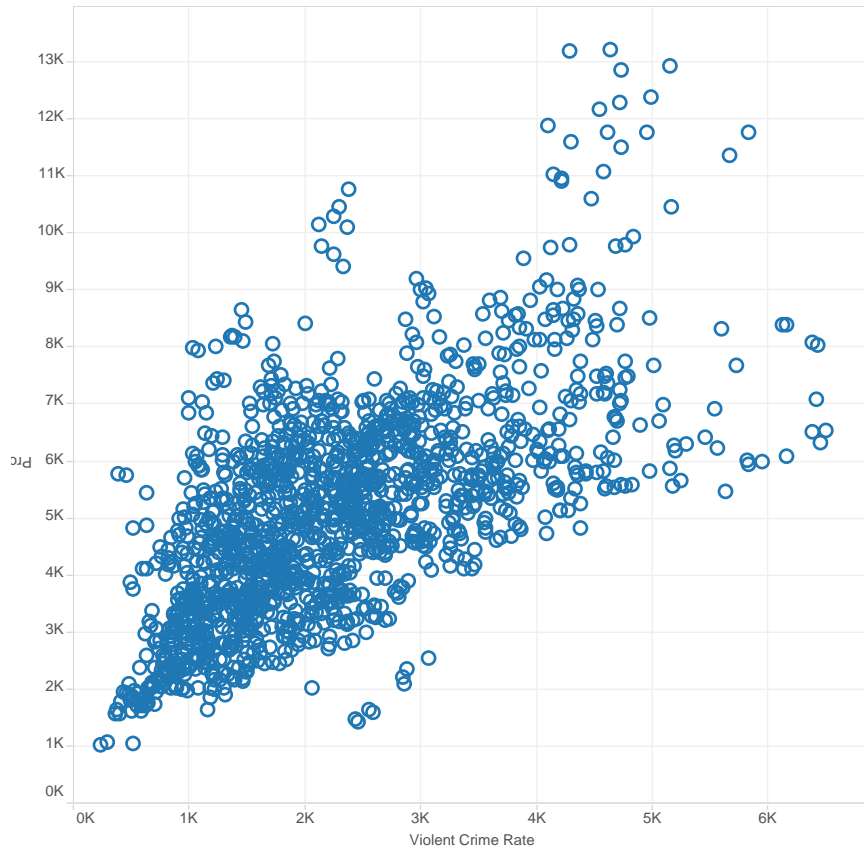
4) On Marks panel, change Automatic to Circle.
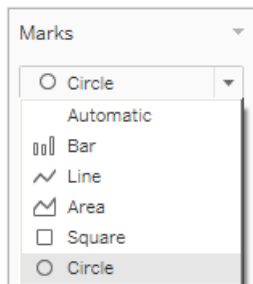
5) Drag Officer Assault Rate to Size on Marks panel.

Note that Officer Assault Rate is the number of police officers killed or assaulted in the line of duty. This bubble chart now shows that police officers are in more danger (as shown in bigger circles) in cities with higher crime rates.
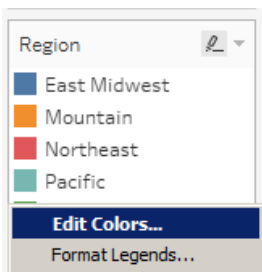
6) Drag Region from Dimensions from Color on Marks panel.
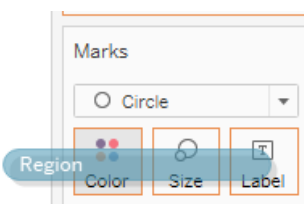
7) While the plot might look pretty, we can make this more informative and visually appealing by using fewer colors. (It is less distracting to viewers if you use as few different colors as possible.) Right-click Region panel, and select Edit colors.

8) Let's assign colors to Regions as follows.

|	
🟩	East Midwest
🟥	Mountain
🟦	Northeast
🟥	Pacific
🟦	South
🟦	South Atlantic
🟩	West Midwest

Now this one looks easier to understand It is now clearer that cities in Midwest (in green) and South (in blue) are more crime-prone.

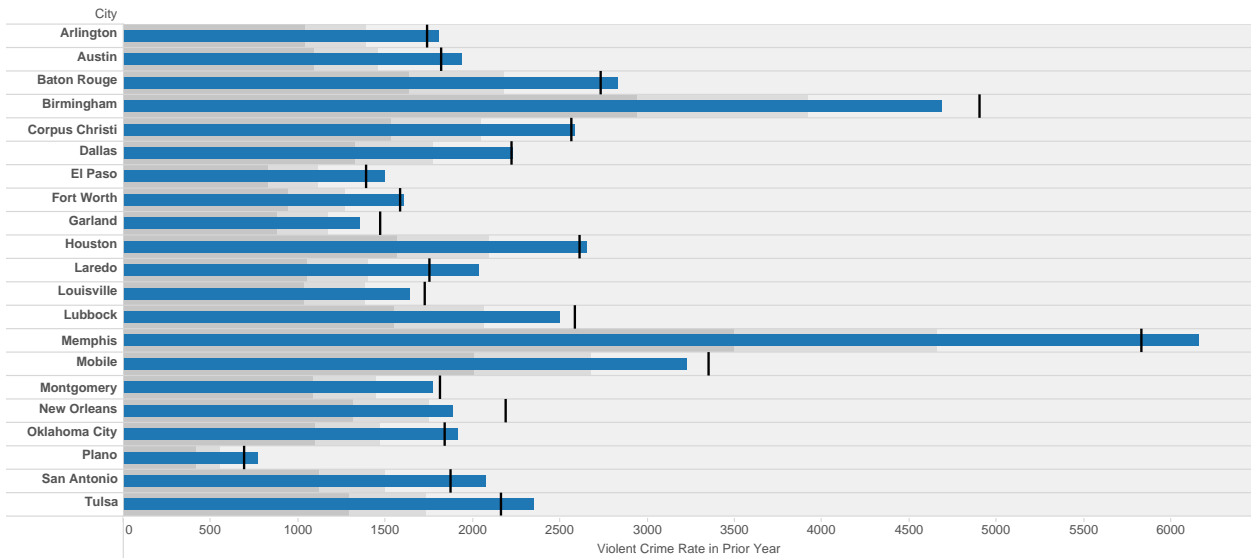7) Rename "Sheet 4" to "Bubble Chart" and save.

Part 5. Bullet Graph

1) Create a new worksheet.

2) Drag Region from Dimensions to Filters. Select South.

3) Drag Year from Dimensions to Filters. Double-click "Years" and select 2014.
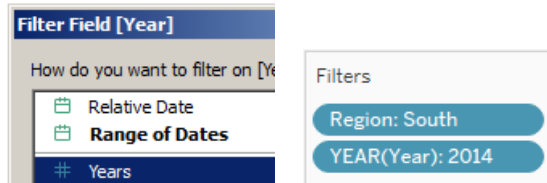
4) With pressing Ctrl key, click City, Violent Crime Rate, and Violet Crime Rate in Prior Year.
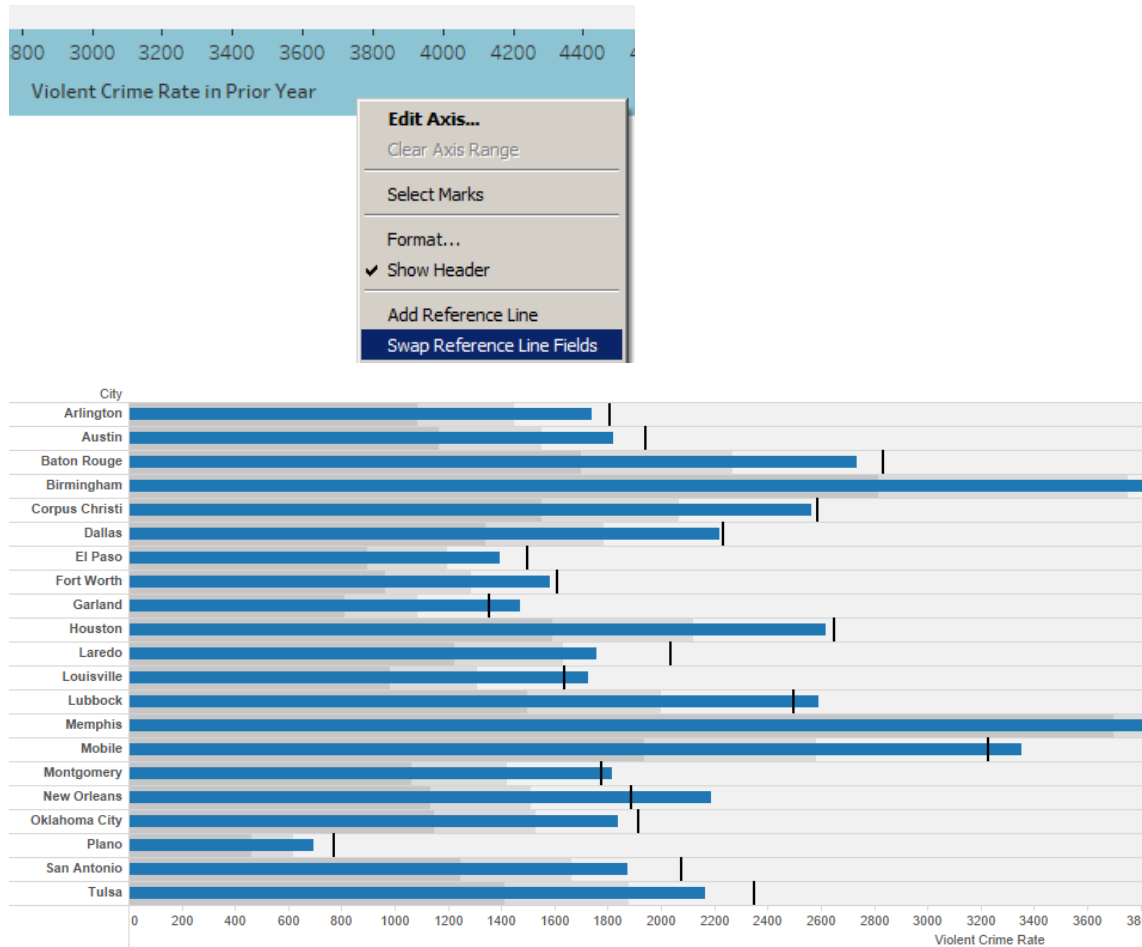
5) On Show me tab, click Bullet Graphs.

This graph shows the violent crime rates in a prior year (2013) as bars and the crime rate in a current year (2014) as horizontal line as a reference. It is not informative. We'd rather like a graph that shows the 2014 crime rate as a bar and the 2013 crime rate as a reference.

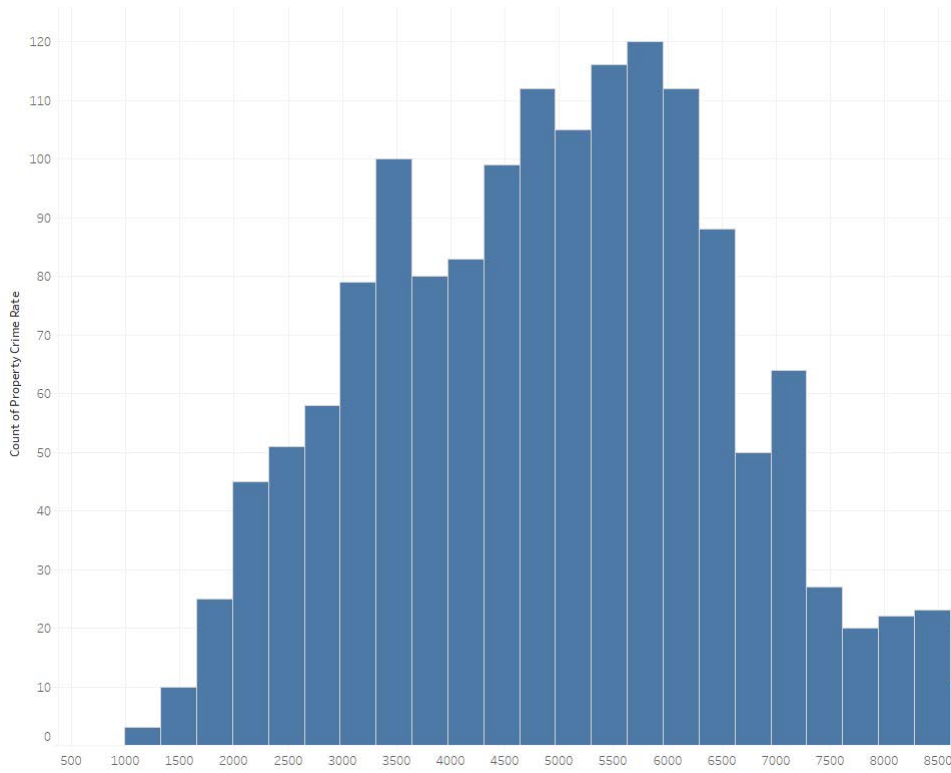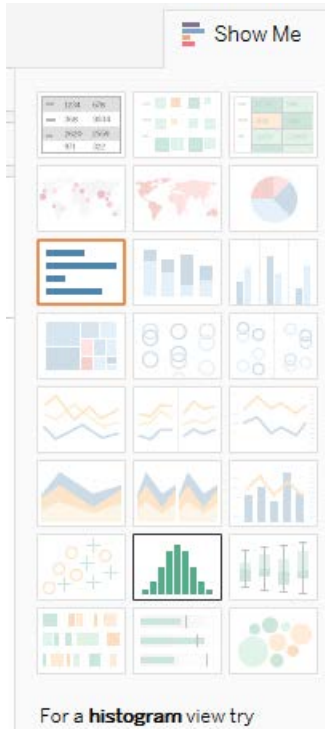6) Right-click the horizontal axis and select "Swap reference line fields."

Now we can see how the crime rate has changed from 2013 (line) to 2014 (bar) in each southern city.
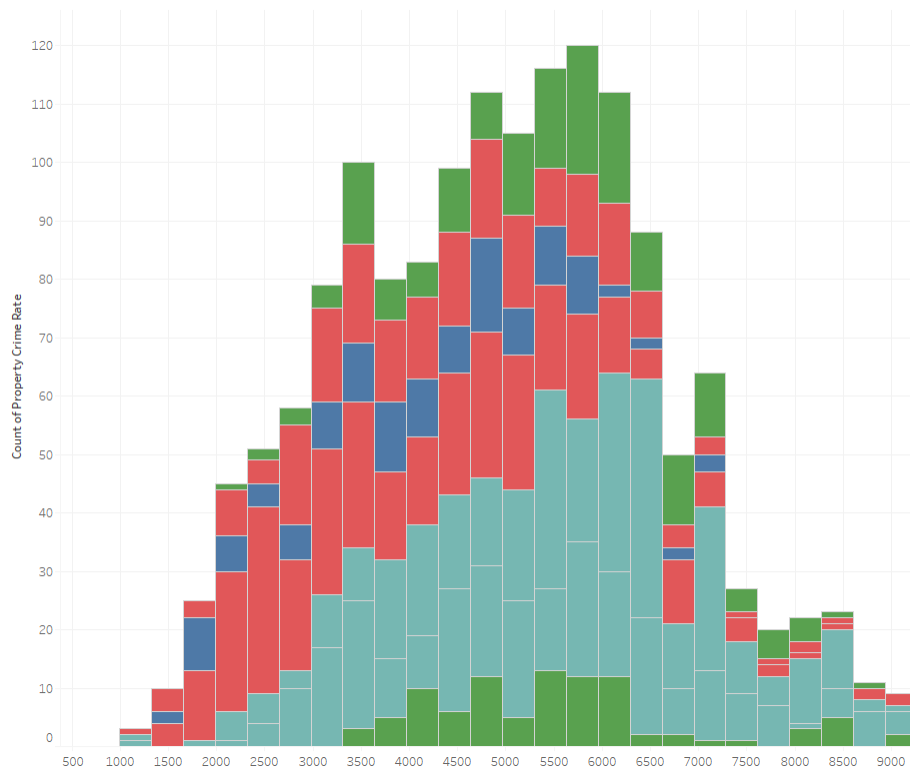
7) Rename "Sheet 5" to "Bullet Graph" and save.

Part 6. Histogram

1) Create a new worksheet.

2) Click Property Crime Rate on Measures.

3) On Show Me tab, click Histogram.
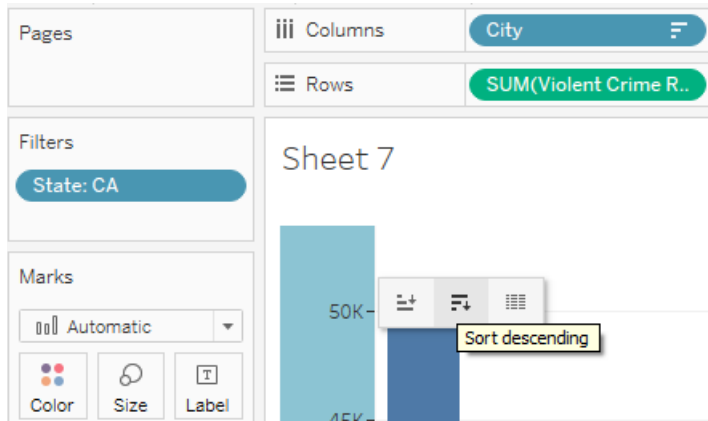
4) Drag Region from Dimensions to Color.

5) Rename "Sheet 6" to "Histogram" and save.

Part 7. Pareto Chart

1) Open a new worksheet.

2) Drag City to Columns and Violent Crime Rates to Rows.

3) Drag State to Filters and select California (CA).

4) Click the vertical axis and select Descending Sort.

5) Drag Violent Crime Rates again to Rows.

6) Click the second Violent Crime Rates and select Add Table Calculation.

7) Under Calculation Type, select "Running Total."

8) Click "Add secondary calculation" at the bottom of Table Calculation window.

Table Calculation   ✕
Difference in Running Sum of Violent Crime Rate

Primary Calculation Type	Secondary Calculation Type
Running Total ▾	Difference From ▾
Sum ▾	**Compute Using**

Compute Using

Table (across)
Cell
Specific Dimensions

☑ City

Restarting every ▾

Table (across)
Cell
Specific Dimensions

☑ City

At the level ▾

Relative to Previous ▾

☑ Add secondary calculation
☑ Show calculation assistance

9) Under Secondary Calculation Type, select Percent of Total.

Secondary Calculation Type
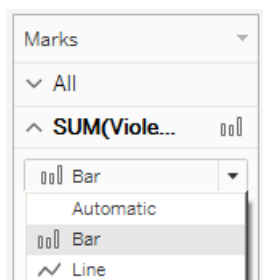
Percent of Total ▾

Difference From
Percent Difference From
C Percent From
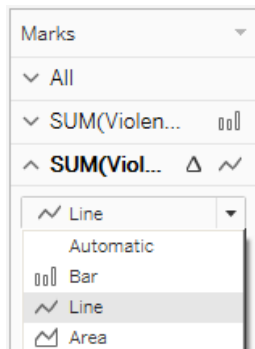Percent of Total
Rank
Percentile
Table

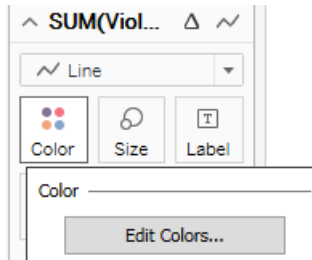8) Click the second Violent Crime Rates again and select Dual Axis.
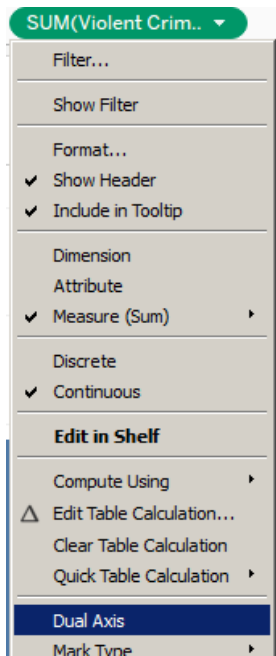
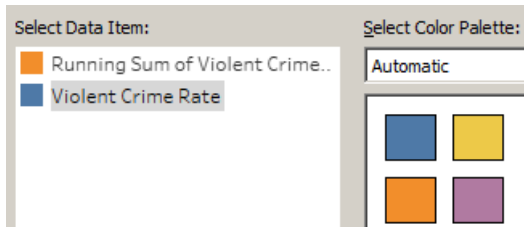9) Under Marks, click the first SUM(Violent Crime Rates) and change Automatic to Bar.

10) Under Marks, click the second SUM(Violent Crime Rates) and change Automatic to Line.
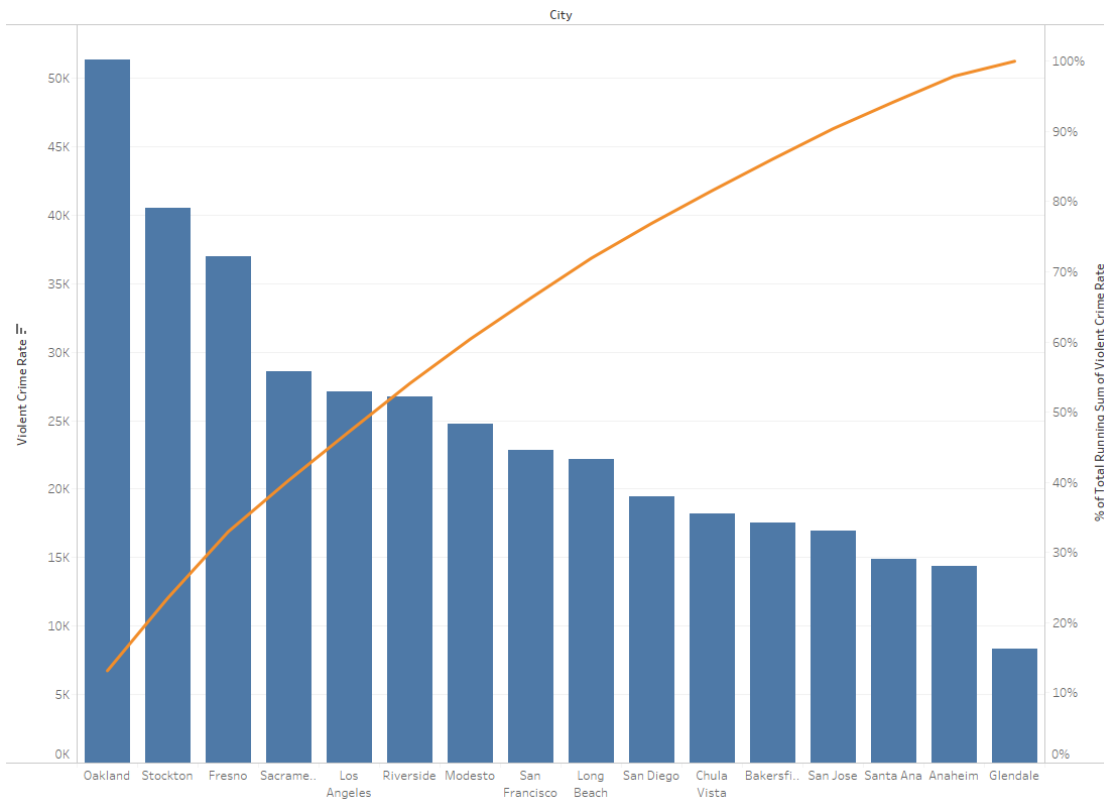
11) Click Color -> Edit Color.

12) Click Violent Crime Rate and click Blue.

And you will see your Pareto Chart.

With this chart, you can find out, for example, how much percentage of crime rates the top five cities share among the all California cities.

13) Rename Sheet 7 to Pareto Chart.