HOMEWORK 4
Submitted by: SHARANG BIRADAR N11407631

Activity 1: Creating data visualizations in Tableau

Data: sales.xls

This lab activity will allow you to do the following:

- Practice importing data into Tableau;
- Understand the difference between dimensions and measures;
- Create several different display types and select the most appropriate display;
- Add a chart title and caption for the your audience; and
- Use the grouping feature and quick filter tool to show only the data you want to display.

1) Go to Tableau and import the sales.xls dataset that you used in assignment 1

- Open Tableau. Select connect to data > In a file > Microsoft Excel
- Navigate to and select the **sales.xls** file from your computer
- Select and drag the sales.xls worksheet to the Go To Worksheet window
- Click on the Go to Worksheet button in orange

2) Save your Tableau workbook as a Tableau Packaged Workbook.

- Go to File > Save as >
- Enter a name for your file.
- Select file type as Tableau Packaged Workbook (.twbx)
- What's the difference between a Tableau Workbook and a Tableau Packaged Workbook?

3) Dimensions and measures

How has your data been categorized by Tableau? You should see your data organized by dimensions and measures.

- What do you remember about the difference between measures and dimensions?[1]

- Write down the names of the fields that are measures:

- Write down the names of the fields that are dimensions:

- Did Tableau encode your data correctly? Are the right variables encoded as dimensions and measures?

---

[1] http://onlinehelp.tableausoftware.com/v6.1/public/online/en-us/Id112A8A00YEX.html

4) Create a chart that displays the number of purchases by age

- What variable do you want to put on the x-axis, or on the columns?

- What variable do you want to put on the y-axis or the rows?

- What is the default output? What type of chart is this?

5) Describe what is happening with the data? Are you seeing what you want to see?

6) Select another display? This can be done through the Marks Card or Show Me.

- What are the available options? Which one did you select?

---STOP---

7) Change Age variable from a measure to a dimension.

- A measure can be changed to a dimension by right clicking on the variable name in the measures window > Convert to dimension.
- Repeat step 4, using the age variable as a dimension. Note the options available for chart types in the Show Me window. Select the bar chart.
- Note the differences among the first bar chart you graphed (in step 4), the histogram or table (in step 6) and your current bar chart?

8) In a new tab, create a stacked bar of the purchases by age and gender.

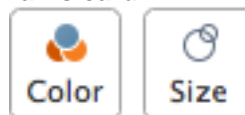**9) Swap the x-axis and the y-axis by mouse over the variable on the y-axis and select the**

icon from the Tableau menu   .

Which arrangement works best?

10) Next, order the data using the sort ascending or descending icons   .

Which order works best?

11) Change the size and color of the bars using the marks card.

On the marks card you will see icons for color and size.

- Select Color > Edit Colors > Select the data item (i.e. female, male) and then select the color for that data item.
- Select Size > Drag the slider to the desired size.

12) Create a third worksheet in your workbook. Group your data by age by right clicking on the Age dimension. Then, create your groupings.

Chart your group.

13) Exclude all men and women under 30 years old from your graph.

This can be done using the quick filter on Age and unchecking the ages you would like to exclude from the display.
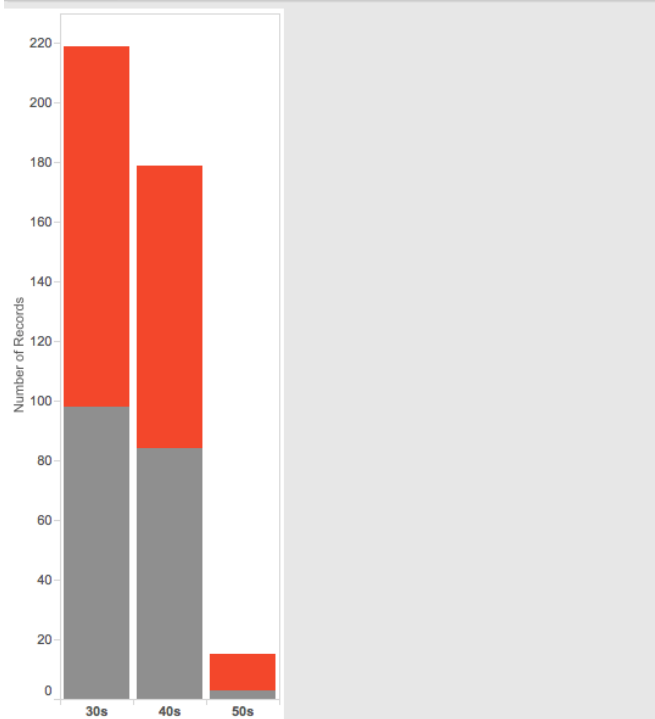
14) Add a title and caption for each worksheet

- Go to Worksheet > Show Title. Revise the title in the title box
- Go to Worksheet > Show Caption. Revise the caption in the caption box.

15) Save your work
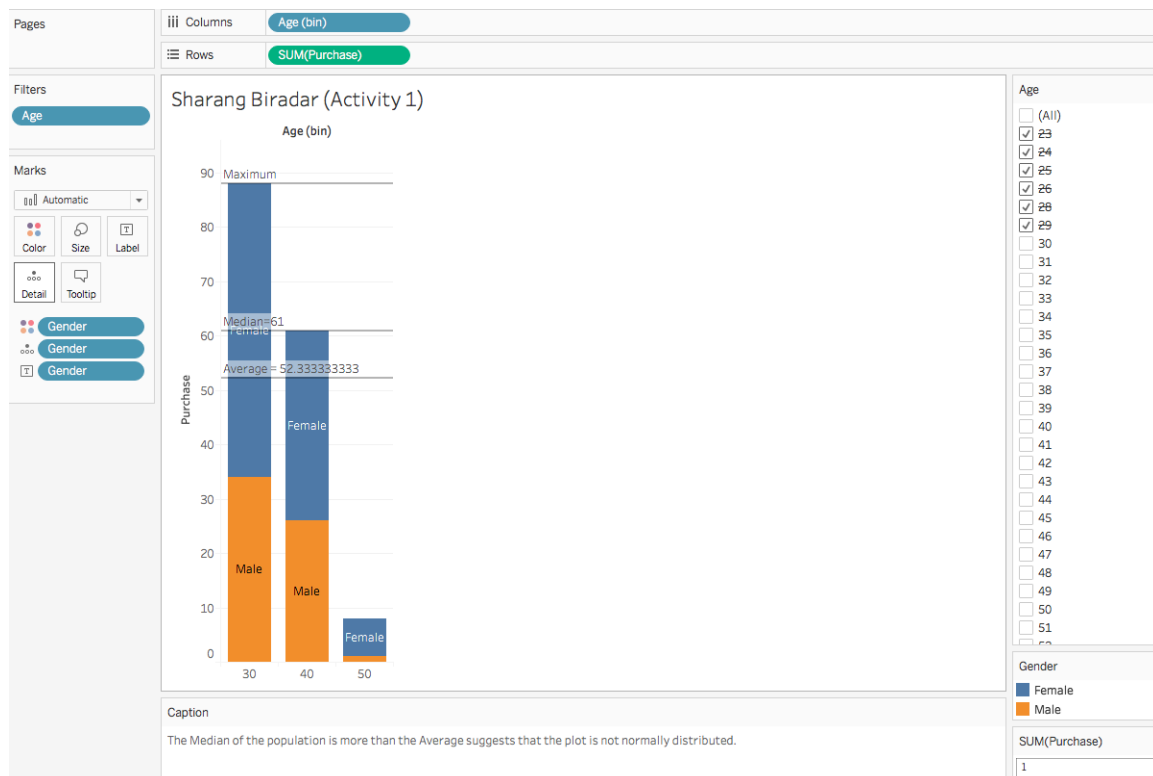
Title   Purchases by Age and Gender for Customers 30 and Above

Caption
Number of purchases by customers above the age of 30, displayed by gender

--STOP--

Bins

Earlier, we grouped our data using the *Create Group* function. We can also group our data using bins. Try grouping the Age variable using bins.

- Create a new sheet.
- Right click on Age (age should be a dimension, if it is not change it from a measure to a dimension) and select *Create Bins*
- Select the number of bins you would like and click Ok. Tableau tends to treat bins different than other programs. The bins number refers to multiples. For example, a bin size of 10 would yield an x-axis with the following values = 20, 30, 40, 50. A bin size of 8 would yield an x-axis with the following values = 16, 24, 32, 40, 48, 56.

Note the difference between Bins and Groups.

Activity 2: The process of data visualization

What are the steps involved in producing a data visualization? Write them down and define them. Did you follow the process for the Tableau exercise (with the casino data)?

## Activity 3: Adding pre-attentive attributes and reference lines in Tableau

A. Reference Lines

Adding reference lines can focus the audience's attention and situate your data in context of a larger trend or pattern.

For this lab, add a reference line to the sales chart you created in activity 6.
P

1. Navigate to the bar chart that shows *Purchases by age group and gender.* This should look similar to the figure below.
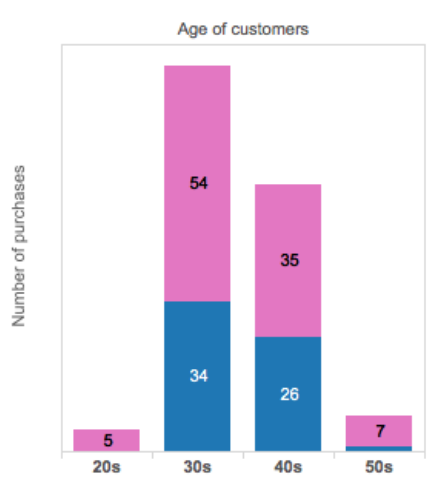
2. Next, right click on the y-axis and select > *Add reference line, band or box.*
3. In the dialog window select:
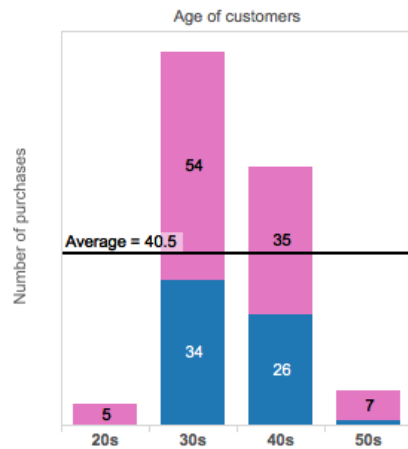
   Scope: Entire table
   Line: Value > Sum (Purchase) > Average
   Formatting: Line with no fills
   select Ok.

4. Then, modify the line and label for the average reference line. Right click on the average line > Edit

5. In the dialog box select the Custom Label and enter
    <Computation> = <Value>.
6. Change the line color to black and select Ok.
7. Your bar chart should look similar to the one below.


9. Try applying other types of reference lines:

    - Constant- places a line at the specified value on the axis.
    - Maximum- places a line at the maximum value.
    - Median- places a line at the median value.
    - Minimum- places a line at the minimum value.
    - Sum -places a line at the SUM of all the values in either the cell, pane, or entire view.
    - Total -places a line at the aggregate of all the values in either the cell, pane, or the entire view. This option is useful when computing a weighted average rather than an average of averages. It is also useful when working with a calculation with a custom aggregation. The total is computed using the underlying data and behaves the same as selecting one of the Totals options on the Analysis menu. If Total All Using has been assigned, the function applied matches the one that has been selected.

--STOP--

B. Reference bands

Reference bands can highlight the data that falls within a certain window value. The reference bands are shown as shaded areas behind the marks in the view between two constant or computed values on the axis. The options are the same as for Reference Lines.

Add reference band to your chart that looks similar to the chart below.
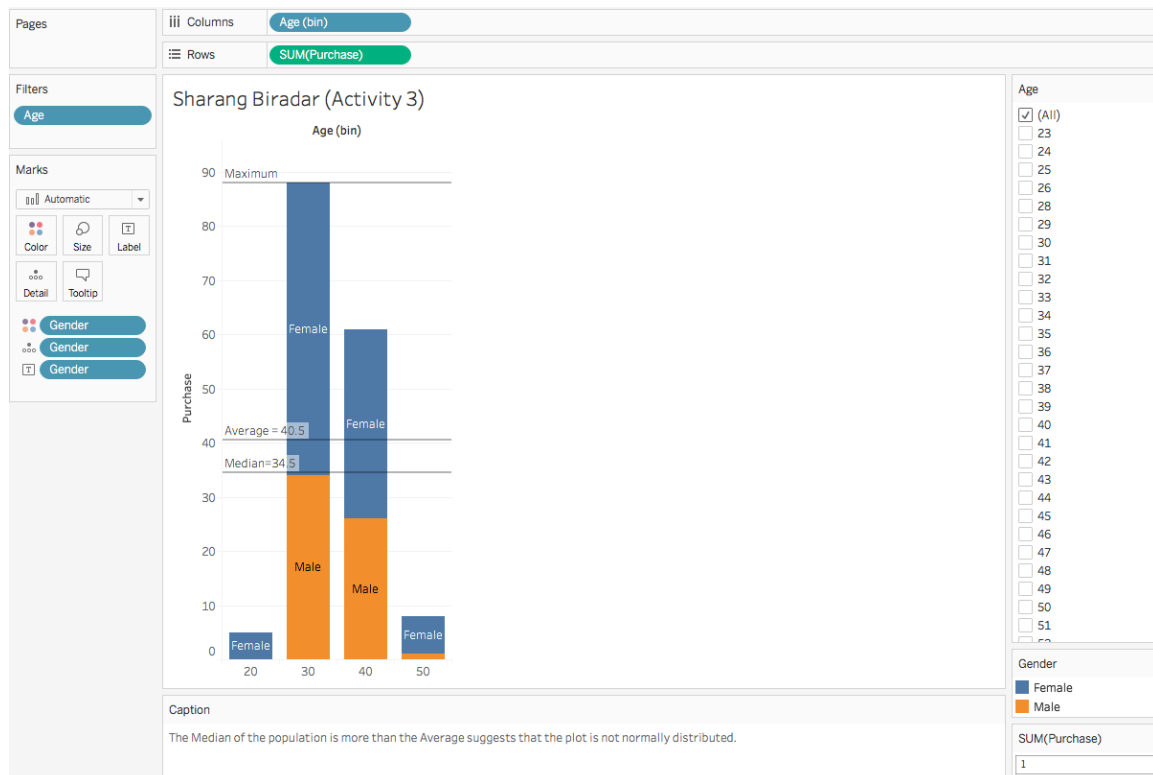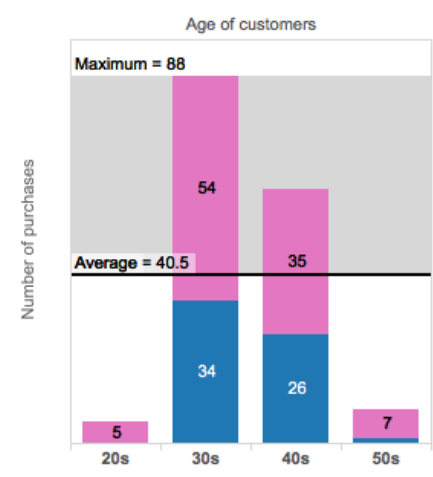
Activity 4: Recycling bin data

This is an open-ended lab. Your general task is to create at least 3 static explanatory visualizations from the **recycle.csv** dataset. This dataset contains a list of all of the recycling bins in NYC.

- Try to create a bar, pie, treemap, stacked bar, and bubble chart. **No maps!**
- Apply effective design principles. Use the Format menu to format your visualization, add a title, caption, and annotations.
- Add pre-attentive attributes
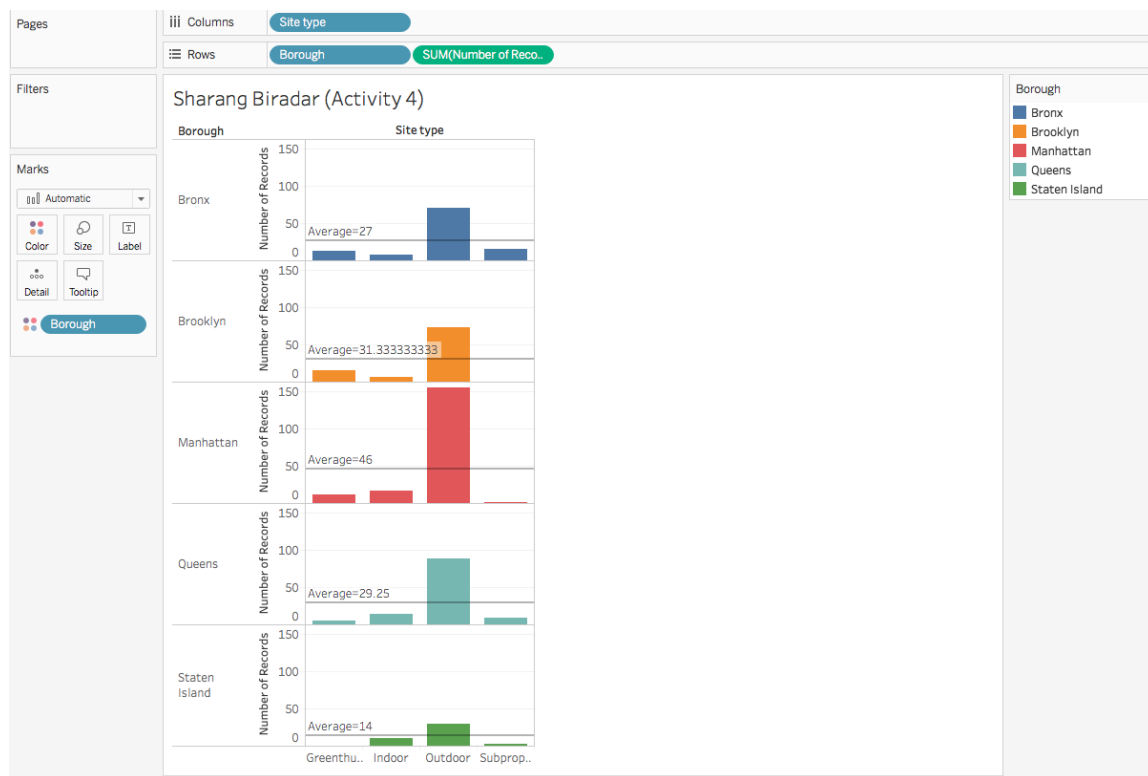
Activity 5: Slicing your data by date. Multiple measures on the same axis, and panel

A. Dates

Years can be thought of as discrete or continuous variables. Let's begin working with a data set you are familiar with: The China internet usage world economic indicator (**chinainternet.csv**).

- Import data
- Drag year to columns and Internet Usage to rows.

This will display a line chart by default. Note that *Years* is coded as a continuous variable. In Tableau time elements such at years, days, hours, minutes will need to be coded as continuous variables to create a time-series line graph. In the example below, each year is connected by a continuous line.
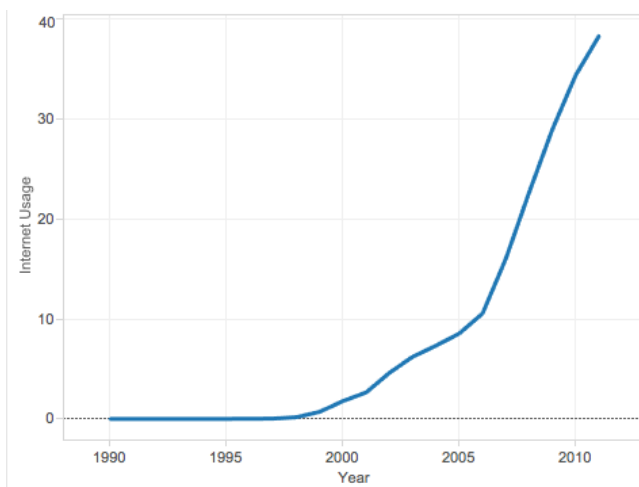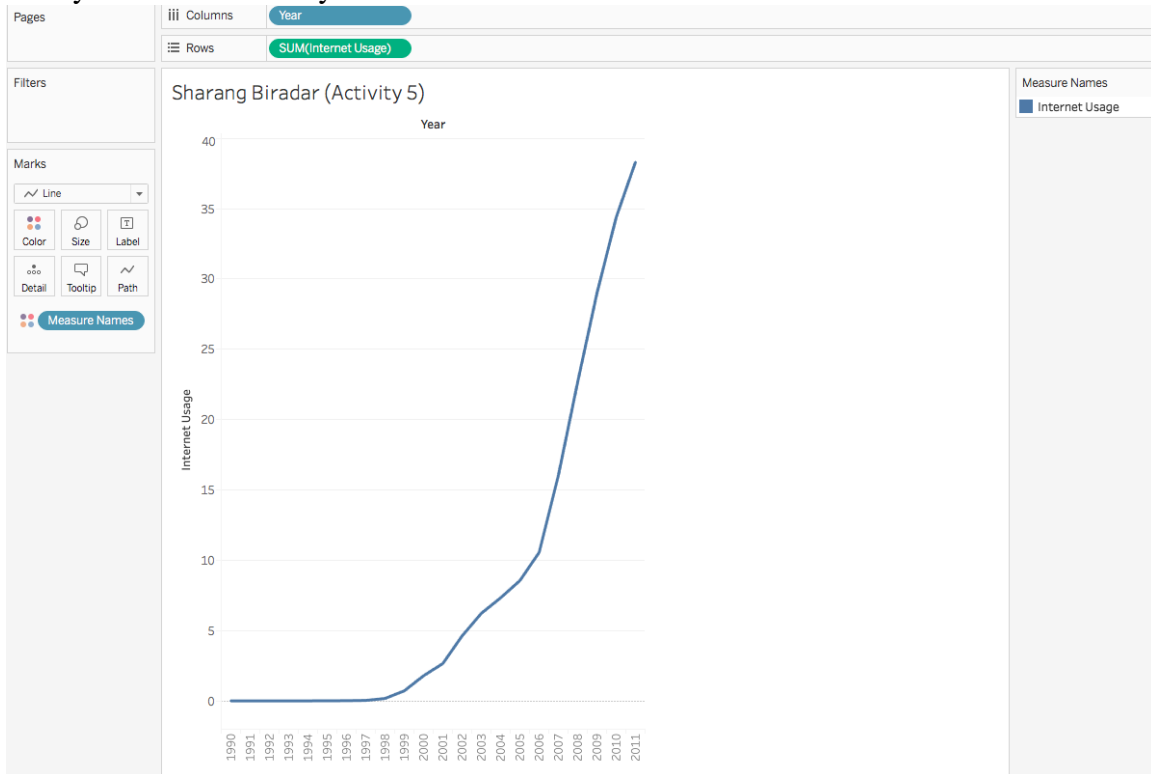
B. Dates as a categorical variable

What if you wanted to graph years as a categorical variable?

You would need to change *Years* to a discrete variable. To do this, right click on *Years* > Discrete. Now you should see your line chart transform into a bar chart (see the figure below)
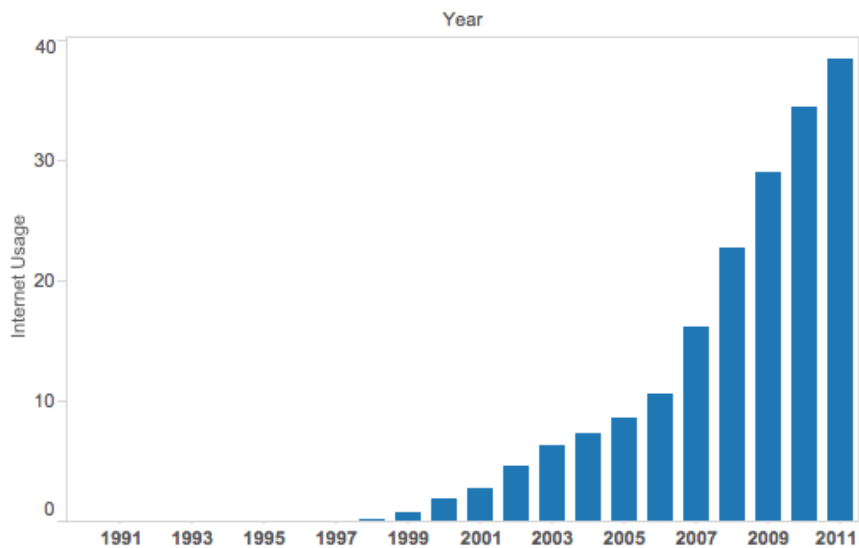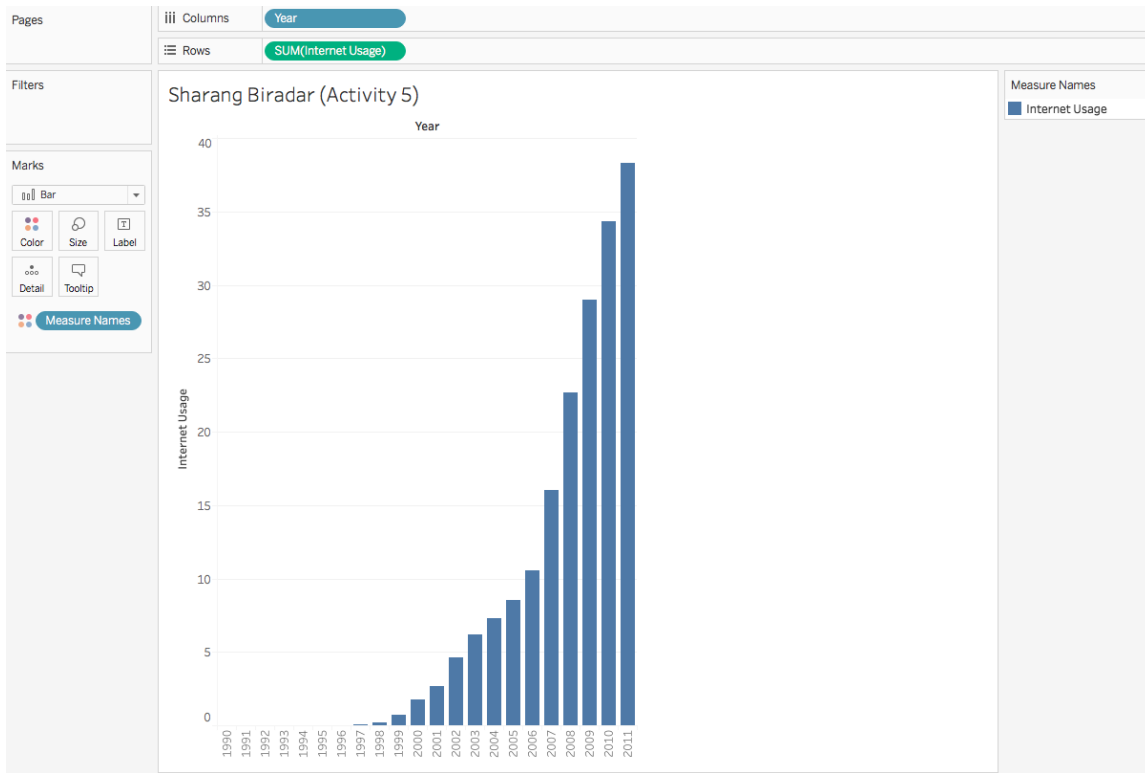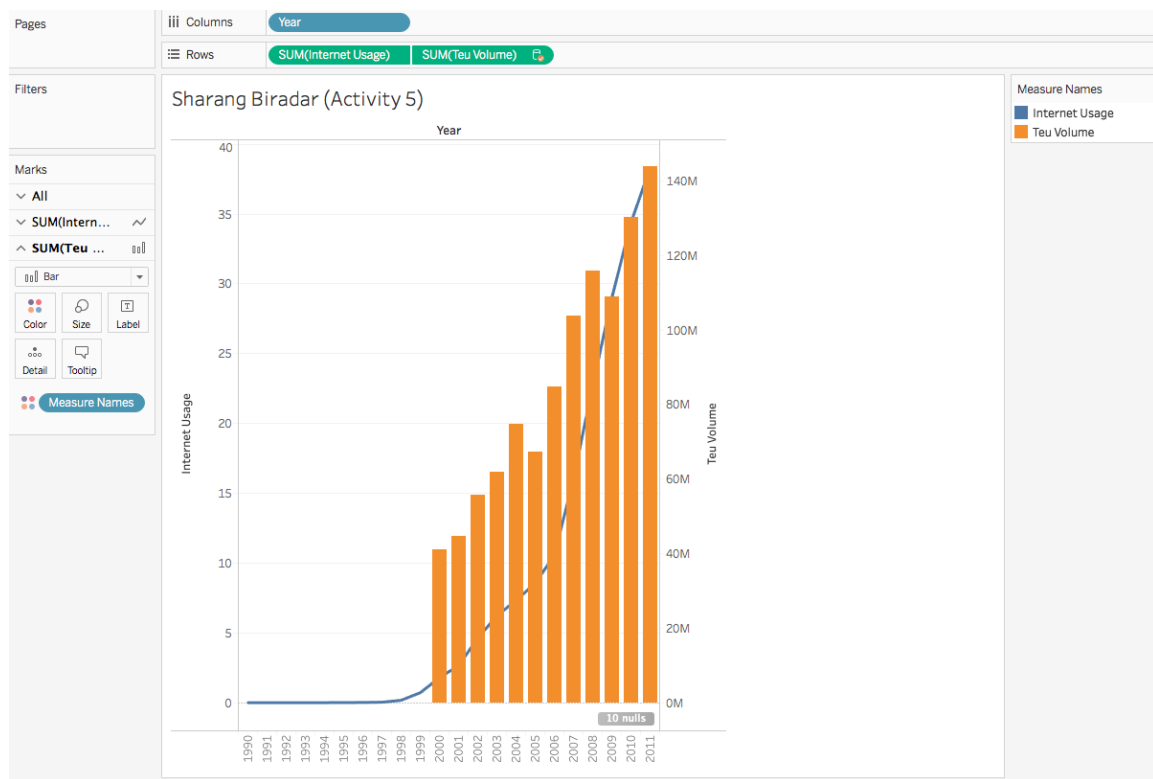
---STOP---

Dual axis charts

What if you wanted to create a second axis to map another variable, such as China's TEU volume? First import the data (**china_teu.xlsx**). Next, drag the TEU volume to the right of the chart on the y-axis, see the figure below. Use the Marks card to change the display of each variable on the y-axis.

Reference: http://downloads.tableausoftware.com/quickstart/feature-guides/secondary_axes.pdf
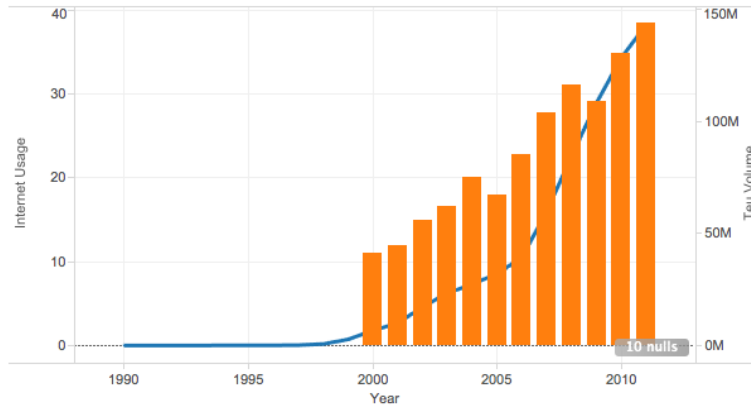
--STOP--

# Activity 6: Creating basic statistical visualizations in Tableau

PART 1

1. Import the following data into tableau.

dataset_a
dataset_b
dataset_c
dataset_d

2. Create a simple box plot for each x value in each data set. Create a single boxplot per sheet.

Steps:

1. Create a new worksheet and label it BoxPlot X A
2. Select sheet 1 (data set A) from the data window.
3. Drag the measure X A to rows.
4. Go to Analysis > disaggregate measures
5. Select the box-and-whiskers plot from the Show Me window. You should see something similar to this boxplot. Repeat the steps for X B, X C, and X D using the other data sets.
6. You should have 4 tabs, each with a boxplot

Sharang Biradar (Activity 6a) — X A, X B, X C, X D box plots
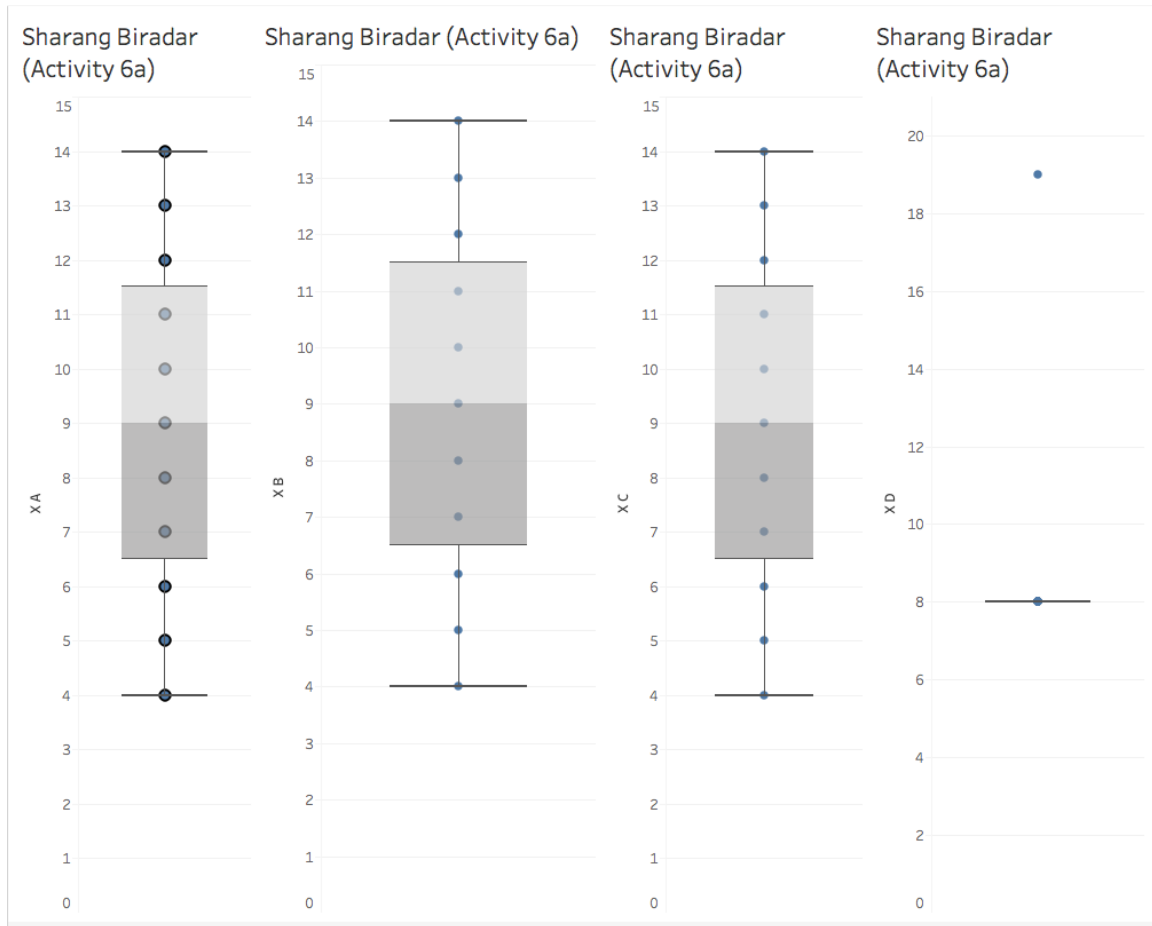
PART 2

Create a histogram for each x value in each data set. Create a single histogram per sheet.

Steps:

1. Create a new worksheet and label it Histogram X A
2. Select sheet 1 (data set A) from the data window.
3. Drag the measure X A to columns.
4. Go to Analysis > Aggregate measures (in case they are not aggregated)
5. Select the histogram plot from the Show Me window. You should see something similar to the histogram below.
6. Drag number of records from measures to **Label** on the Marks card
7. Repeat the steps for X B, X C, and X D using the other data sets.

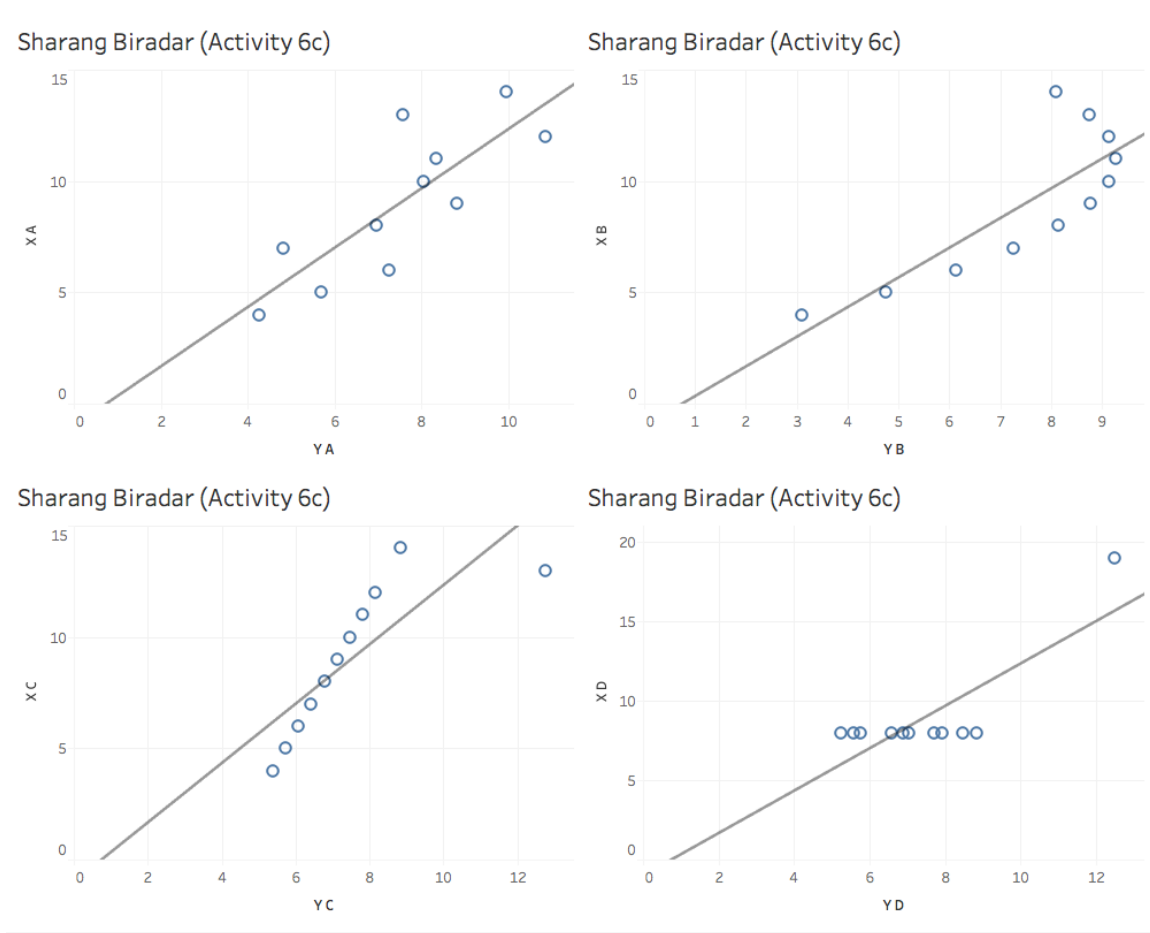8. You should have 4 tabs, each with a histogram
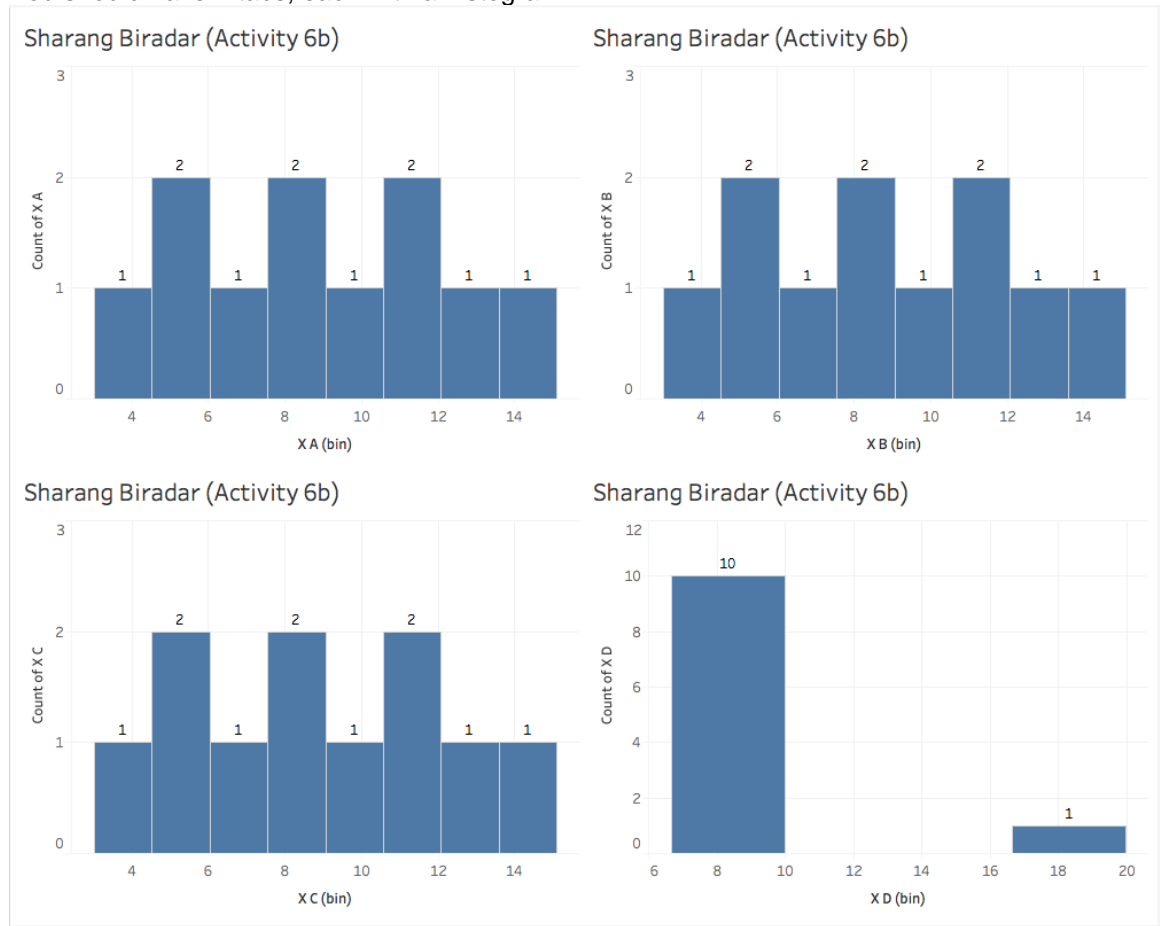
PART 3

Create a scatter plot for each x-y value in each data set with a trend line. Create a single scatter plot per sheet.

Steps:

1. Create a new worksheet and label it X Y - A
2. Select sheet 1 (data set A) from the data window.
3. Drag the measure X A to rows and Y A to columns.
4. Go to Analysis > uncheck Aggregate measures (in case they are aggregated)
5. Select the scatter plot from the Show Me window. You should see something similar to the scatter plot below.
6. Add a trend line.
   a. Go to analysis > Show Trend lines
   b. To learn more about the model used, go to analysis > Describe Trend Model
   c. To change the model used, go to analysis > Edit Trend Line (For purposes of this exercise, keep the linear model).
7. Repeat the steps for X Y - B, X Y- C, and X Y- D using the other data sets.
8. You should have 4 tabs, each with a scatter plot with a trend line visible.
9. Next, merge the 4 charts on a single worksheet, using the dashboard feature.
   a. Create a new dashboard by going to Dashboard > New Dashboard

b. Drag your 4 sheets X Y - A, X Y - B, X Y- C, and X Y- D to the dashboard white space.
c. Note the differences between the 4 data sets.

Sharang Biradar (Activity 6c) — X A vs Y A
Sharang Biradar (Activity 6c) — X B vs Y B
Sharang Biradar (Activity 6c) — X C vs Y C
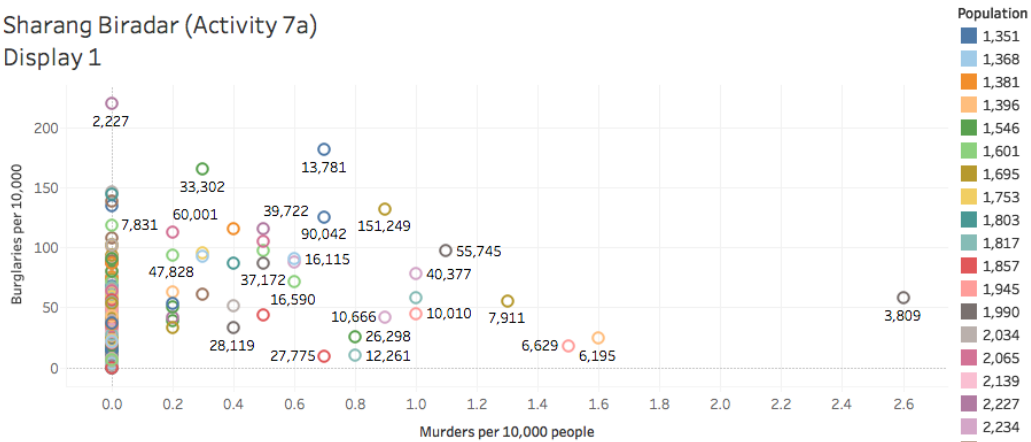Sharang Biradar (Activity 6c) — X D vs Y D

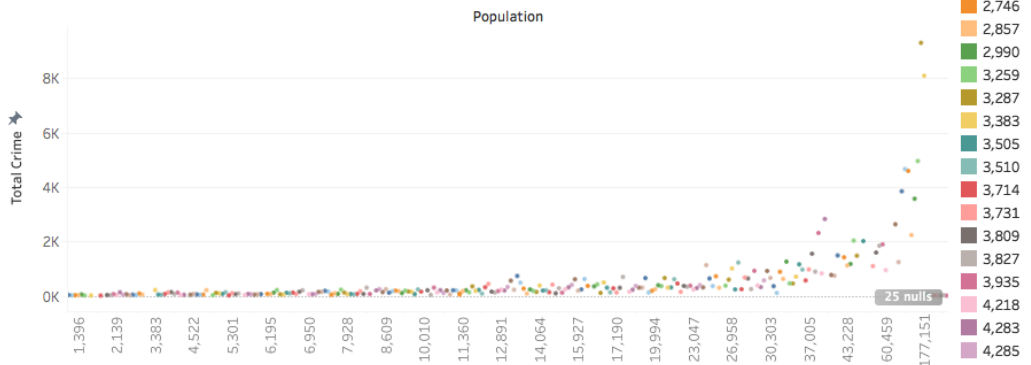## Activity 7: Visualizing multivariate data in Tableau

Using the **crime.xls** data set create 2 displays. In addition, create a calculated field called totalcrime.

- Display 1: Create a scatter plot to show the relationship between murders per 10,000 people vs. burglaries per 10,000 people by population. Encode population by size or color. Determine the encoding that best displays the relationship.
- Display 2: Create a calculated field called TotalCrime by right clicking in the measures field > create calculated field. This field should sum all the crimes. See the figure below.
- Create a scatter plot that shows the relationship between total crimes and population.

Sharang Biradar (Activity 7a)
Display 1

Population
1,351
1,368
1,381
1,396
1,546
1,601
1,695
1,753
1,803
1,817
1,857
1,945
1,990
2,034
2,065
2,139
2,227
2,234
2,442
2,646
2,653
2,740
2,746
2,857
2,990
3,259
3,287
3,383
3,505
3,510
3,714
3,731
3,809
3,827
3,935
4,218
4,283
4,285

Burglaries per 10,000

Murders per 10,000 people

2,227
13,781
33,302
60,001
39,722
7,831
90,042
151,249
47,828
16,115
55,745
37,172
40,377
16,590
10,666
10,010
7,911
28,119
26,298
6,629
27,775 12,261
6,195

Sharang Biradar (Activity 7b)
Display 2

Population

Total Crime

25 nulls

Calculated Field [TotalCrime]

Name: TotalCrime

Formula:

[Arsons]+[Aggravated assaults]+[Burglaries]+[Forcible rapes]+[Larceny-thefts]+[Motor vehicle thefts]+[Murders and manslaugh

✔ The calculation is valid.   Clear

Fields: Parameters: Create   Functions: Help
All   All   All

Enter Text to Search   Enter Text to Search   Enter Text to Search   Arsons
Aggravated assaults   ABS   Data type: Integer
Arsons   ACOS
Arsons per 10,000   ASCII
Assaults per 10,000   ASIN
Burglaries   ATAN
Burglaries per 10,000   ATAN2
⊕ City   ATTR
Forcible rapes   AVG
Larceny-thefts   CASE
Larceny-thefts per 1...   CHAR
Motor vehicle thefts   CONTAINS
Motor vehicle thefts ...   COS
Murders and mansla...   COT
Murders per 10,000 ...   COUNT   Describe...
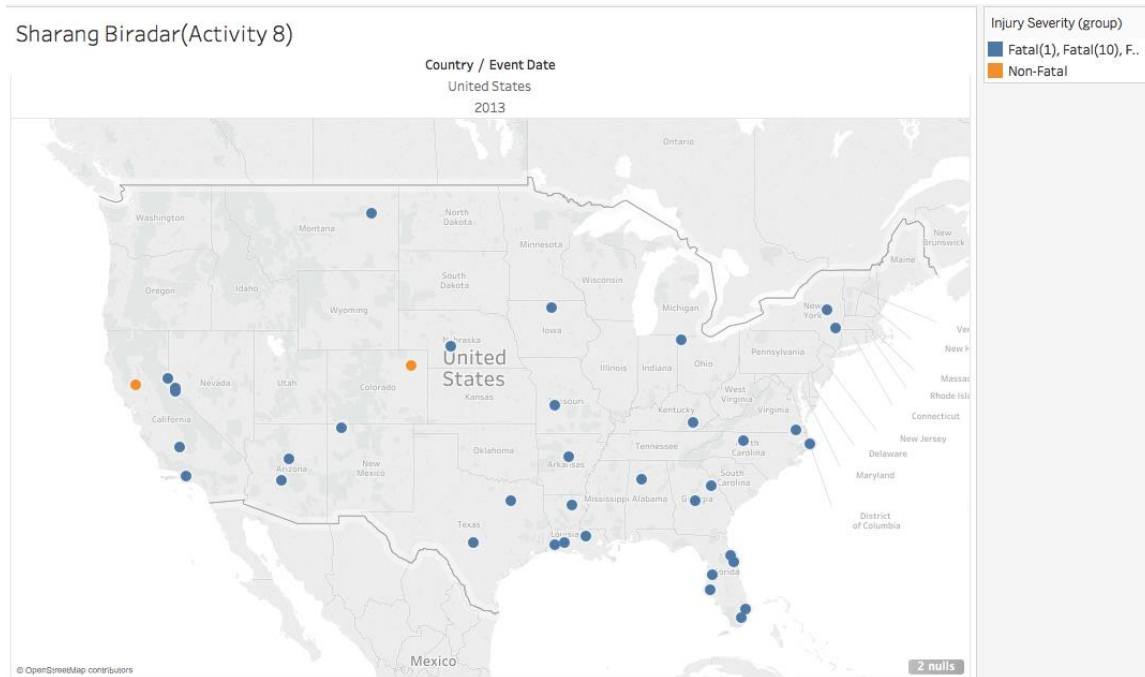⊕# Number of Records   COUNTD

⚠ Changes to this calculation might change the following:
  Worksheet "TotalCrimes by city's population"   Apply   Cancel   OK

16

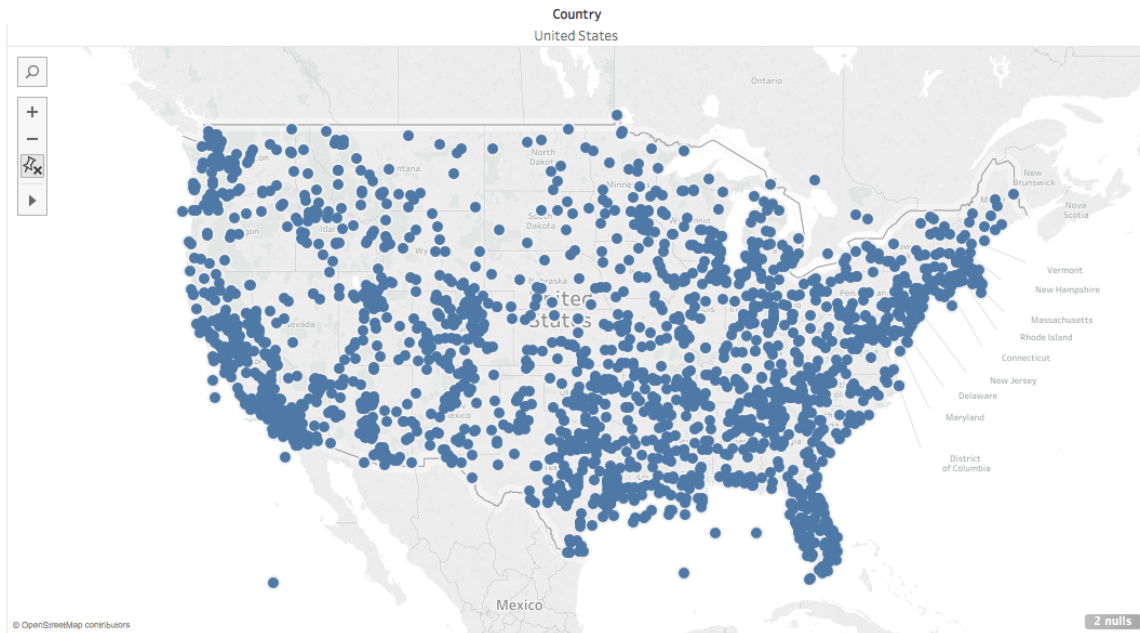## Activity 8: Visualizing geospatial data in Tableau

Use the **aviation** data set. Show the plane crashes in the US for 2013.
Create a group for fatal vs. non-fatal crashes. Filter out unknowns.

Activity 9: Showing History in Tableau

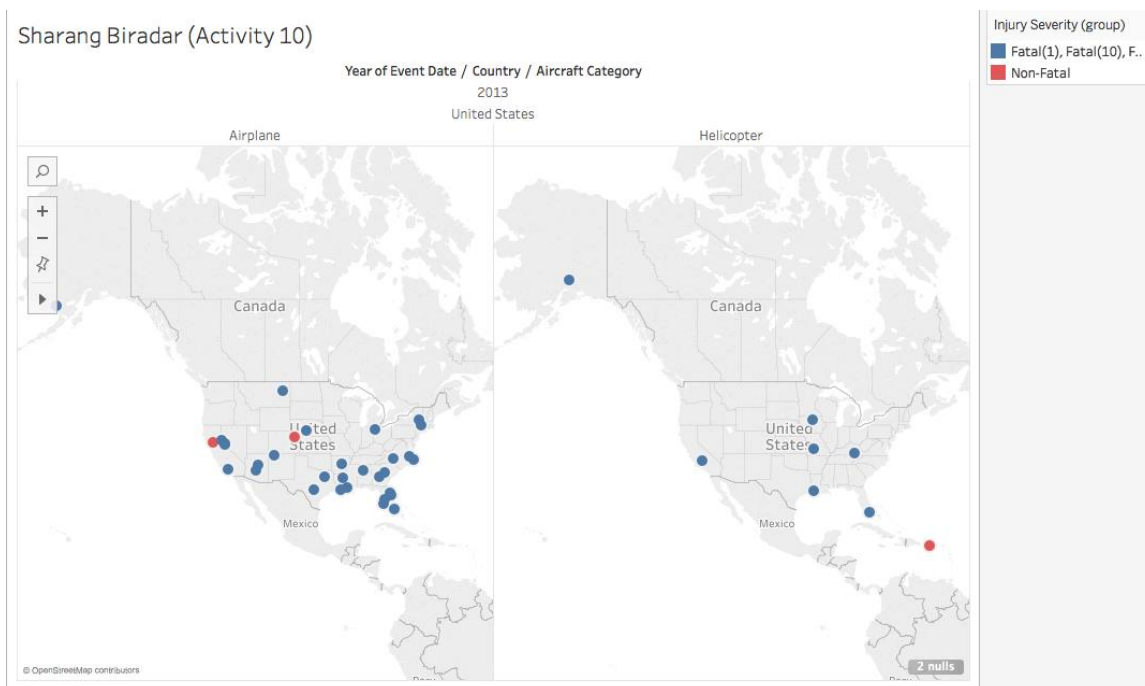Show the plane crashes in the US from the 1940s through 2013.

Activity 10: Highlight and filter

Create additional displays of your aviation data. Link those displays together on a dashboard. Then, use one display as a filter.

Activity 11: Cheese factory dashboard

Use the Cheesemakers.csv data.

Audience: VP of Sales at cheese maker that sells to the public and gourmet retailers

Task: Updates for monthly review by the VP of Sales
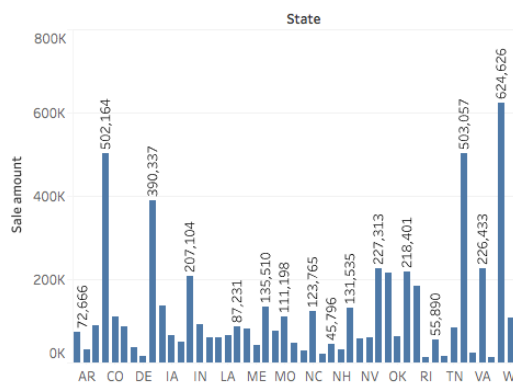What are the sales by state?
What were the sales by customer contact method in 2013 compared to 2012?
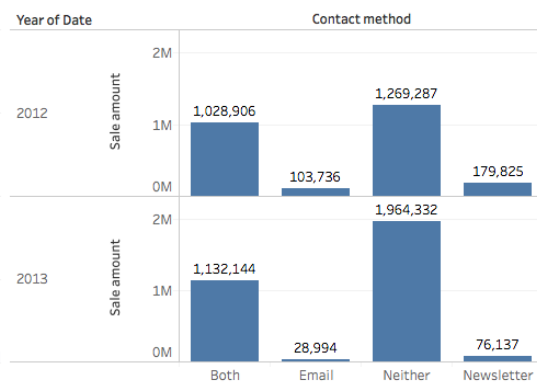What are the annual sales by item versus target sales?
What are the actual sales by customer contact method versus the target sales?

Reference: http://kb.tableausoftware.com/articles/knowledgebase/interactive-dashboard
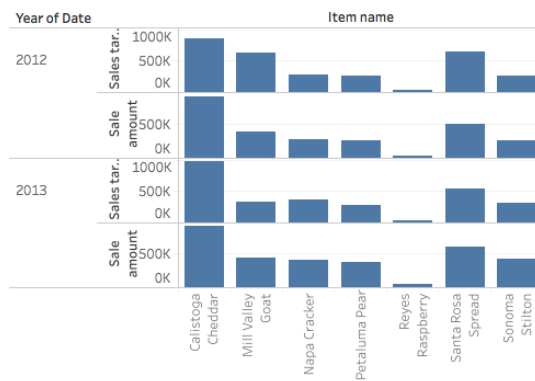

Sharang Biradar (Activity 11a)

Sharang Biradar (Activity 11b)

Sharang Biradar (Activity 11c)

Sharang Biradar (Activity 11d)