Tableau 9 Overview

Dr. Philip E Cannata

Oracle Data Scientist, Oracle Certified Professional, and Adjunct Professor at the University of Texas Computer Science Department in Austin
Objective

This presentation will communicate at a high level my methodology for systematically analyzing data using Tableau’s visual techniques. Fundamental and Advanced Tableau concepts will be presented and you will be introduced to many very useful visualizations along the way.

This presentation contains detailed, step-by-step instructions for reproducing each visualization. During the presentation, you can do these steps along with me or just watch.

At the end of each section, I’ll ask “How might you use what you’ve just seen on your project?”
Elements of Data Visualization

Dr. Philip Cannata, phil.cannata@oracle.com,
office hours: TR 4:00 - 4:45 in GDC 3.402.

TAs: Xun Li, xun.bhsfer@utexas.edu, office hours: GDC 3.802A, W 3:00 - 5:00 pm

Canvas link.

We will be using Piazza.

Make sure you are registered with piazza for this course and be sure to check and read your piazza email several times a day because this will be the primary means of communication outside of class time. Dr. Cannata will not alter his procedure of always sending all of his messages to the entire class.

Students with disabilities link.

Students with disabilities may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471-6259, and the link above.

Draft Syllabus for Elements of Programming Languages link.

---

Project 1

Interesting article “Letter written by Titanic’s coward goes up for auction” here. And another one here.

1. Open the Titanic SQL file at this link. Copy the contents into a SQL Developer worksheet for your account and click the Run script button. This will create...

Project 2

Project 3 - Due Friday October 16 at 11:59 pm
Project 4 - Due Monday November 2 at 11:59 pm
Project 5 - Due Friday November 13 at 11:59 pm
Project 6 - Due Monday November 23 at 11:59 pm
Final Project - December 7, and 8, 2015

---

My Elements of Data Visualization UT Class Website
• http://www.cs.utexas.edu/~cannata/dataVis/
Two Useful Books

1. Tableau 9: The Official Guide by George Peck
2. Key Performance Indicators: The 75 Measures Every Manager Needs to Know by Bernard Marr
Tableau Terminology
1. Open Super Store Orders Extract.tde
2. Select Sheet 1
1. Click on Department under Dimensions and select Describe
2. Click Load if necessary
Tableau Terminology – Measure and Green Things

1. Click on Discount under Measures and select Describe
2. Click Load if necessary
Methodology

To analyze a Data Source
1. Start with a green thing (usually a Measure) in non-aggregated mode and make a Boxplot
2. Start with a green thing in aggregated mode and make a Histogram
3. Start with 2 green things and make a Scatter plot
4. Start with two blue things (usually Dimensions) plus a green thing and make a Crosstab
5. Start with a blue thing and a green thing and make a Barchart

Non-Aggregated Measures Analysis (start with a green thing) - also demonstrates Boxplots, Dates, and Pages

   Boxplot Show Me.twbx

Aggregated Measures Analysis (start with a green thing) - also demonstrates Histograms, Dual-axis Plots and Show Me

   Histogram Show Me.twbx

Scatter Plots (start with 2 green things) -also demonstrates Maps, Actions, Dashboards, and the Analytics Tab - Models and Forecasting

   Scatter Plot.twbx

Crosstabs (start with two blue things and a green thing) - also demonstrates Key Performance Indicators (KPIs), Calculated Fields, Sets, Parameters, and Hierarchies

   Crosstab with KPI.twbx

Barcharts (start with a blue thing and a green thing) - also demonstrate Table Calculations, the Analytics Tab - Reference Lines, Formatting, and ID Sets

   Barchart.twbx

Joining Tables - this section demonstrates how to join two tables.

Data Blending - this section demonstrates how to blend two tables, which is not the same as joining tables.
Boxplot Story (start with a green thing)

1. Open Super Store Orders Extract.tde
2. Uncheck Aggregate Measures under the Analysis tab and Rename the Current Sheet to Sales Boxplot.
3. Click on Sales Under Measures and then click on the boxplot icon on "Show Me" i.e., the right icon on row 7.
4. Drag Category onto Columns.
5. Drag Region onto the Color Shelf.
6. Drag Region onto the Filters Shelf and from this Pill's menu, select Show Quick Filter.
7. Drag Sales onto the Filters Shelf and from this Pill's menu, select Show Quick Filter.
8. Drag Customer Name onto Detail
9. Explore the outliers
10. Drag Region onto Pages
11. Set the minimum Sales to $20,000
12. Click through the Regions on Pages and notice “office Supplies” is common to all
1. Duplicate the Sales Boxplot Sheet and rename it to Sales Boxplot over Time.
2. On the Order Date item under Dimensions, select Change Data Type -> Date
3. Drag Order Date on top of Regions on the Pages Shelf and on this Pill's menu, change it to Quarter (make sure you select the second Quarter option)
4. On the Analytics Tab. Click and hold on the Average Line pill and drag to Pane
5. Select the forward play button.
Boxplot Story (start with a green thing) + Profit over Quarter Time using Pages

1. Duplicate the Sales Boxplot over Time Sheet to a new sheet named Profit Boxplot over Time.
2. Drag Profit on top of Sales on Row to replace Sales with Profit
3. Select a region of interest and play the movie.
How might you use what you’ve just seen on your project?
Histogram Story (start with a green thing)

1. Open Super Store Orders Extract.tde
2. Rename Sheet 1 to Shipping Cost Histogram
3. Click on Shipping Costs under Measures and then click on the Histogram Icon on Show Me (i.e., the middle icon on row 7).
4. Notice a Shipping Costs (bin) item was created under Dimensions. Open the menu on this item and select Edit. Change the bin size to 10.
5. Drag Container onto the Pages Shelf.
6. On the Analytics Tab. Click and hold on the Average Line pill and drag to Pane
7. Step through each Container type.
8. Drag Category on top of Container on Pages
Histogram Story (start with a green thing) + Profit on a Dual-axis

1. Duplicate the current Sheet to a new Sheet named Shipping Cost + Profit Histogram
2. Drag Profit to the right axis to create a dual axis plot.
3. Click on the CNT(Shipping Costs) Pill on the Marks Card and change its Mark to Bar.
4. On the Analytics Tab. Click and hold on the Average Line pill and drag to Sum(Profit)/Pane
5. Step through each Container type.
Histogram Story (start with a green thing) + Packed Bubbles Plot

1. Duplicate the Shipping Cost + Profit Histogram Sheet to a new Sheet named Packed Bubbles
2. Go to Show me and pick Packed Bubbles
3. Change the Type of Order Date to Date if necessary
4. Drag Order Date to Pages
5. Click through the Quarters
Histogram Story (start with a green thing) + Treemap Plot

1. Duplicate the Shipping Cost + Profit Histogram Sheet to a new Sheet named TreemapDrag
2. Go to Show me and pick Treemap
3. Change the Type of Order Date to Date if necessary
4. Drag Order Date to Pages
5. Click through the Quarters
How might you use what you’ve just seen on your project?
Scatter Plot Story (start with 2 green things)
1. Open Super Store Orders Extract.tde
2. Uncheck Aggregate Measures under the Analysis tab
3. Rename the Current Sheet to Scatter Plot
4. Drag Sales onto the Columns Shelf
5. Drag Profit onto the Rows Shelf
6. Drag Category onto the Pages Shelf
7. Click on the Analytics Tab and double click on Trend Line
8. Click on the Data Tab
9. Drag State onto Color
10. Remove Country Region from Color
Scatter Plot Story (start with 2 green things) +
Maps and Actions
1. Create a new Sheet named Map
2. Double click on State
3. Click on the Worksheet Tab and select Actions
4. Select Add Action -> Filter
5. Select Run Action on Select
6. Select Map for the Source
7. Select Scatter Plot for Target
8. Click on Ok and Ok
9. Click on Texas on the map
Scatter Plot Story (start with 2 green things) + Dashboards

1. Create a new Dashboard
2. Drag the Scatter Plot and Map Sheets onto the Dashboard
3. Zoom into the United States
4. Drag select Washington, Oregon, and California
5. Click on the white “Use as Filter” icon just to the left of the X in the Scatter Plot Sheet
6. Step through the Categories and compare the States
Scatter Plot Story (start with 2 green things) + Trend Model

1. Right click on the blue trend line and select Describe Trend Model
2. For more on Trend Lines, see the following video: https://www.tableau.com/learn/tutorials/on-demand/trend-lines-residuals-and-forecasting
Scatter Plot Story (start with 2 green things) + Another Action Plot
1. Create a new sheet named “High Profit States”.
2. Drag State onto Columns and Profit onto Rows.
3. Click on the Worksheet Tab and select Actions
4. Select Add Action -> Filter.
5. Select Run Action on Select.
7. Select Scatter Plot for Target.
8. Click on Ok and Ok.
9. Click on Illinois on the “High Profit States.
10. Add this sheet to the dashboard also and select the four highest profit states.
Scatter Plot Story (start with 2 green things) + Forecasting

1. Create a new Sheet named “Forecasting”
2. Change the type of Order Date to Date
3. Drag Order Date onto Columns
4. Change Order Date to Quarters
5. Drag Sales onto Rows
6. On the Analytics Tab select Forecast and drag it to Add a Forecast
7. Drag Category to Pages
8. Step through Category and see how many are forecast to be flat
How might you use what you’ve just seen on your project?
Crosstabs + KPI Story (start with two blue things and a green thing)

1. Open Super Store Orders Extract.tde
2. Rename Sheet 1 to Crosstab + KPI
3. Drag Category to the Columns Shelf
4. Drag State to the Rows Shelf
5. Drag Country Region to the Filters Shelf and select United States of America
6. Drag Sales to Text on the Marks Card
7. View the data by clicking on the View Data icon on the Dimensions Tab, notice there is no column named “KPI – Profit Ratio”
8. On the Dimensions Tab, click on the menu icon and select Create Calculated Field
9. Name the Calculated Field “KPI - Profit Ratio”
10. Enter \texttt{\textit{Sum(Profit)/Sum([Sales])}} as the calculation
11. On the Dimensions Tab, click the View Data icon and see that a new field named “KPI - Profit Ratio” was added to the data.
12. View the data again, now notice there is a column named “KPI – Profit Ratio”
13. Drag KPI - Profit Ratio onto Color
14. Drag KPI - Profit Ratio onto Size
15. Change the Color Palette to Red-White-Green Diverge
Crosstabs + KPI Story (start with two blue things and a green thing) + Sets

1. Create a new Sheet named “Four Highest”
2. Drag Category to the Columns Shelf
3. Drag Sales to the Rows Shelf
4. Drag Country Region onto Filters and select United States of America
5. Select the four highest Sales Categories and hover over one of them
6. Click on the Set icon and select Create Set
7. Name the Set “Four Highest Sales Categories”
Crosstabs + KPI Story (start with two blue things and a green thing) +
Sets
1. Duplicate the “Crosstab + KPI” Sheet a new Sheet named “Crosstab + KPI 2”
2. Drag the Four Highest Sales Category Set to Filters and change it to Show In/Out of Set
3. Show Quick Filter on this Set
4. Set this Filter to In
Crosstabs + KPI Story (start with two blue things and a green thing) + Sets
1. Create a new Sheet named “Region Sales Map”
2. Drag Country Region to Filters and select United States of America
3. Double click on State
4. Drag KPI – Profit Ratio to Color
5. Drag the Four Highest Sales Categories to the Filters Shelf
6. Drag Region to Label
Crosstabs + KPI Story (start with two blue things and a green thing) + Parameters

1. Duplicate Sheet “Region Sales Map” to a new Sheet named “Region Sales Map 2”
2. On the Dimensions Tab click on the menu and select “Create Parameter” and name it “KPI Low”
3. Click on Range
4. Set Minimum to 0, Maximum to 0.1, and Step size to 0.01
5. Duplicate this Parameter, name it “KPI Medium” and set Minimum to 0.1, Maximum to 0.2, and Step size to 0.01
6. Duplicate KPI – Profit Ratio to KPR – Profit Ratio (copy) and Edit it
7. Rename it to KPI – Profit Ratio Low to High and enter the following calculation,
   
   IF sum([Profit])/sum([Sales]) <= [KPI Low] THEN "Low"
   ELSEIF sum([Profit])/sum([Sales]) <= [KPI Medium] THEN "Medium"
   ELSE "High"
   END

8. Remove KPI – Profit Ratio from Color and replace it with KPI – Profit Ratio Low to High
9. Click on the menu for each Parameter and select Show Parameter Control
10. Adjust Parameter settings as desired.
Crosstabs + KPI Story (start with two blue things and a green thing) + Hierarchies

1. On the Marks Card, click on the + sign on the left side of the State Pill
How might you use what you’ve just seen on your project?
Barcharts Story (start with a blue thing and a green thing)

1. Open Super Store Orders Extract.tde
2. Rename Sheet 1 Barchart
3. Drag Sales to Columns
4. Drag Category and Region to Rows
5. Drag Region onto Filters and select East, West, and South
6. Click on the Analytics Tab
7. Click on Average Line, drag it a bit and then select Pane
8. Click on one of the Average lines and Edit it – change the Label to Value
9. Click on one of the Averages lines and select Format
10. A Format Pane shows up on the left.
11. On this Pane select Numbers -> Currency(Custom)
12. Change Decimal Places to 0
13. Click on the bottom Axis and select Format
14. A Format Pane shows up on the left.
15. On this Pane select Numbers -> Currency(Custom)
16. Change Decimal Places to 0
Barcharts Story (start with a blue thing and a green thing) + Table Calculations

1. Duplicate the Barchart Sheet to a Table Calculations Sheet.
2. Right click on the View and select View Data – notice there is no column named “Calculated Difference from Average”.
3. Create a Calculated Field named "Calculated Difference from Average" as
   \[ \text{SUM}([\text{Sales}]) - \text{WINDOW_AVG} \left( \text{sum}([\text{Sales}]) \right) \]
4. Drag "Calculated Difference from Average" onto Label.
5. Right click on the View and select View Data – notice there is not a column named “Calculated Difference from Average” as there would be for a simple Calculated Field.
6. Hold command (on Mac) down and drag \text{SUM}([\text{Sales}]) from Columns to Label. This will copy \text{SUM}([\text{Sales}]) to Label.
7. Drag "Calculated Difference from Average" onto Color.
Barcharts Story (start with a blue thing and a green thing) + ID Sets

1. Create a new Sheet named ID Sets
2. Drag Order Id onto Columns
3. Drag Discount onto Rows
4. Set the View to be Entire View
5. Highlight all lines between 0.40 and 0.30 by dragging a box over them
6. Hover over a highlighted line, select the Set icon and click on Create Set
7. Name the Set High Discount Orders
8. Drag the Set onto Pages and click the play button so that the status shows Out
9. Highlight all lines between 0.30 and 0.20 by dragging a box over them
10. Hover over a highlighted line, select the Set icon and click on Create Set
11. Name the Set Medium Discount Orders
Barcharts Story (start with a blue thing and a green thing) + ID Sets
1. Create a new Sheet named ID Sets on a Map
2. Double click State
3. Highlight the United States with the Zoom Area tool
4. Drag the 2 Sets to Pages
5. Drag Customer Name to Color
6. Drag Customer Names to Detail
7. Drag City to Detail
8. Change the Marks to Circle
9. Click the Pages slider and hover over a Circle to see the Details
Barcharts Story (start with a blue thing and a green thing) + ID Sets Part 2

1. Create a new Sheet named High Sales Customers
2. Drag Customer Id onto Columns
3. Drag Sales onto Rows
4. Set the View to be Entire View
5. Highlight select Customers with Sales > $100k
6. Drag High Discount Orders onto Filters
7. Drag Profit on top of Sales on Rows
How might you use what you’ve just seen on your project?
Joining Tables:
1. Open Super Store Orders Extract.tde
2. Open US Super Store Quota.tde
3. You should see the following on the connections page:

![Diagram of Extract and Extract1 connections]

4. Click on the “Join” icon, i.e., the icon with the blue and white circles.
5. Fill in the Join as follows, i.e., select “Inner” and set State = US_STATE

![Join window with Inner selected]

6. Inspect the data and notice only the US States appear.

![Table showing only US States]

7. This shows that in an Inner Join, only rows that meet the join criteria are retained in the resulting table.
Joining Tables continued:

1. Change the Join to be Left

2. Inspect the data and notice that US States appear with a QUOTA and non-US States appear with null for their QUOTA.

3. This shows that in a Left Join, all of the rows in the Left table appear at least once in the resulting table even if their value needs to be set to null.
Joining Tables continued:

1. Open Sheet 1 and notice there are Dimensions and Measures for both tables.
How might you use what you’ve just seen on your project?
Blending Data:
1. Open Super Store Orders Extract.tde.
2. Drag Sales onto Columns and Region onto Rows.
3. Add Coffee Chain extract as a new data source by choosing:

![New Data Source](image)

4. Try to add coffee sales to Columns, you should get the following warning:

![Warning](image)

5. Under Data go to Edit Relationships.
6. Do a Custom edit and add Region from Super Store and Market from Coffee Chain:

![Relationships](image)

7. Now you should be able to drag Coffee Sale to Columns.
How might you use what you’ve just seen on your project?