Tableau Take-Home Assignment

Getting started:
Download a free copy of Tableau (Tableau Desktop, not Tableau Prep) for personal use at: http://www.tableau.com/academic/students. When installing Tableau you will likely be entering your information into a form twice - while odd, this is how Tableau does it…

Download the assignment data located in the Tableau folder on the course website and start working on the *Data Loading and Data Transformation* requirement (see below). Note that I have created tutorials that guide you through these steps (use ctrl+click to open the links). All the tutorials are also available directly from the Panopto folder.

Then work on the *Visual Analytics* tutorials. These tutorials will introduce you to many basic and some more advanced visual analytics functions and presentation options in Tableau. While you should perform all the steps in these tutorials, you do not need to study them and it is OK if you do not understand everything. *Some of these tutorials are related directly to questions in a fraud case. If you are assigned the fraud case (this is not part of the regular takehome assignment), then you might want to refer back to the Visual Analytics tutorials when working on the case.*

After you have completed the visual analytics tutorials, then start working on the *Problems*. These questions build on the content in the earlier tutorials but also introduce some new concepts. After you have completed the Problems you are done with the Tableau Introduction (the take-home assignment). The skills test is based on the work you did in the Problems section.

If you are assigned the Tableau case (if so then this will be shown as a separate assignment after the Tableau skills test), then continue with the instructions after the problem solutions. Note that the case uses the same data that you use in the take-home assignment and it is as such not necessary to repeat the data loading and data transformation steps.

Data Loading and Data Transformation

1) Follow the importing and validating data tutorial and connect to the data from Tableau and validate the data using the check figures in Exhibit 1 (the table notes contain instructions that may be helpful) of the case (way down towards the end of this document). After reviewing the tutorial, only validate the transactions table, all other tables should important fine if the transactions table has imported correctly. *When validating the number of records note that the tutorial uses a version of Tableau that included a pill named Number of records. Your version of Tableau will likely instead name this pill Transactions.txt (count).*
2) Follow the creating table joins tutorial and create table joins in Tableau using the ERD diagram in Exhibit 2 (the table notes contain instructions that may be helpful).
3) Follow the creating calculated fields tutorial and calculate fields as indicated in the notes to the Transactions table in Exhibit 5 (shown below for ease of access):

(1) *Sales* (Gross Sales subtotal) defined as IF TransType = "S" and Subtotal > 0 THEN Subtotal, (2) *Returns* (Returns subtotal) defined as IF Subtotal < 0 THEN -Subtotal, (3) *Net sales* (sales subtotal net of returns subtotal) defined as IF TransType = "S" THEN Subtotal, and (4) *Purchases* (Purchases subtotal) defined as IF TransType = "O" THEN Subtotal).

Note that the tutorial only shows how to create one calculated field. You need to follow similar procedures and use the logic above to also create returns, net sales, and purchases.

4) Follow the converting between Measures and Dimensions tutorial and convert fields from measures to dimensions or from dimensions to measures as needed. Note that dimensions are fields that contain discrete data that either cannot be aggregated (e.g., it is not possible to sum all names in a table) or that have no practical meaning if aggregated (e.g., the sum of all zip codes in a table is useless information). Measures are fields that contain numerical data that when aggregated have practical meaning (e.g., the sum of TR Quantity for all sales transactions represents the total number of units sold). *The tutorial uses a version of Tableau that contains two header labels with the words Measures and Dimensions, respectively. Your version of Tableau will likely not have these labels (you still have fields grouped by dimensions at the top and measures at the bottom, separate by a light grey line).*

Visual Analytics

Important - just go through these tutorials (perform all steps) and get a general feel for Tableau, do not study this and don't worry if you do not understand it all; instead focus on studying the problems further down and then refer back to these tutorials when you go through the case.

5) Dimensions (grouping) and measures (aggregating), rows and columns, and dates
6) Marks area
7) Axis
   a) Chart types, dual axis, labels, axis titles, and formatting
   b) Independent axis ranges
8) Quick table calculations
   a) Quick table calculations
   b) Quick table calculations – compute using
9) Trend and reference lines
   a) Trend lines and forecasting
   b) Reference lines
10) Level of detail expressions (LOD)
11) Ranking and filtering Top N within categories

Problems 1)

See section appendix for graph solutions; tutorial links are provided in the question numbers.

Q1 - What is the percentage of total purchases for each material in 2015? Use Purchases for purchases, Mat MaterialDescription for materials, and TR TransDate for dates. Create a bar graph with percentage of total annual purchases on the y-axis and each material on the x-axis and filter out null values and only include transaction from 2015.

Formatting:
x-axis

- show x-axis labels vertically (material names)
- make sure that the entire material names are shown
- remove the Column Field Label (the text that reads Mat MaterialDescription)

y-axis

- change the y-axis label to "Percent of Total Purchases"
- include no decimals on the y-axis scale percentages

general chart area

- change the Chart Title to "Percent of Total Purchases in 2015 - Company Wide"
- add labels on top of each bar indicating the percentage of total purchases (by for example using the marks area) and display these labels horizontally and formatted to one decimal (right click purchases on the Rows shelf → Format → Pane → Default Numbers → Percentage with 1 decimal)

Q2 - Add to Q1 and answer the following question: what is the percentage of total purchases for each material in 2015 for each site? Use TR SiteCode for site and make sure that the percentages for the different materials add up to 100% for each site? Display the materials on the x-axis with bars for the different sites within each material type.

Formatting:

- add boarders around each bar and apply consistent colors for different sites; use blue for all sites but use different color intensity for each site
- display bar labels vertically
- change the Chart Title to "Percent of Total Purchases in 2015 per Site"

Q3 - Add to the graph above and show 2012 data in addition to 2015 data. To make this work you need to create two calculated fields, one that shows purchases for 2012 and one the shows purchases for 2015. Use the YEAR() function to determine the year of TR TransDate. When calculating the first variable, if the year is 2012 then return the purchase amount for that row otherwise return null. Then add these two calculated fields to the graph (and remove purchases and the 2015 filter), perform a quick table calculation for each calculated field to get the percentage of total (for each calculated field, again make sure that the percentages adds up to 100% for each site), and use a dual axis chart (right click on the y-axis in the bottom graph and select Dual Axis) to combine the two measures into one graph.

Formatting:

- Remove the bar labels
- Change the chart title to: "Percent of Total Purchases in 2015 (bars) and 2012 (circles) per Site"
- Dual Axis

    - Synchronize the axis
    - Remove the axis on the right side of the graph
    - Change the y-axis title to “Percent of Total Purchases per Site”
- In the marks area
    - change the type of graph to show bars for 2015 data and circles for 2012 data (make sure that the circles are in front of the bars)
    - Use the same colors for different sites used in Q2 (for both the bars and the circles) – this may happen automatically
    - Have boarders around the circles and bars
    - Change the size of the circles to the width of the bars
- Annotations
    - Add marks annotations to one circle and one bar
    - Enter the text “2012” and “2015” in respective text box
    - Remove the text box boarders

Q4 - Show purchases per state on a map that shows color intensity for purchases (have the entire state with this color). Only show purchases for 2015 and exclude Alaska and Hawaii. Show Marks Labels for the min and max values.

Formatting: None

Q5 - Use a bar chart to show for each site total Sales (TR Subtotal), average sales per transaction (TR Subtotal), number of sales transactions (TR Subtotal or any field that does not have null values), total quantity sold (TR Quantity), average unit price per transaction (TR UnitPrice), average unit price per item sold (TR Subtotal and TR Quantity), number of customers (Ent EntityID or any other field that uniquely identifies a customer), and average sales per customers (TR Subtotal and Ent EntityID) in 2015.
Formatting: None

Q6 - Show the average transaction sales amount for all employees in NY for each quarter. Add a reference line showing average transaction sales amount per quarter for all employees in NY. Also add a trend line (a single trend line) representing the average trend for all NY employees (i.e., do not show one trend line per employee).

Formatting: Change the y-axis scale to have a fixed start of 14K and change the y-axis and x-axis label to “Average Transaction Sales Amount” and “Transaction Quarter”, respectively.

Q7 - Duplicate the graph in Q6 and add a line showing the average transaction sales amount for the entire company (not just NY). To accomplish this create an LOD statements (and use a dual axis chart to add this calculated field to the graph).

For the skills test, focus on FIXED LOD statements. Note that FIXED LOD statements follow this syntax: {FIXED grouping declaration : aggregate expression}, where grouping declaration

is a list of fields to use to group the data, e.g., [TR SiteCode], [TR Manager], and aggregate expression is a single aggregate function and field to perform the calculation on, e.g., SUM([TR Subtotal]). For example, {FIXED [TR SiteCode], [TR Manager] : SUM([TR Subtotal])} groups by TR SiteCode and TR Manager and sums TR Subtotal. This produces a separate result set (that can be added to an existing graph) created by the following SQL statement: SELECT SUM([TR Subtotal]) FROM… GROUP BY [TR SiteCode], [TR Manager]. When added to an existing graph, this result set is only filtered and aggregated further if the graph contains filters and aggregation dimensions that are at the same or higher level of detail as the data in the LOD result set (i.e., the result set is generated first and cannot be disaggregated by the graph).

Formatting: Change the y-axis scale to have a fixed start of 13K and remove the trend-line for the newly added company average. If added automatically, remove the measure name pill from the Color marks area. Add an annotation stating "Company-wide average transaction sales amount" to the newly added company average, remove the arrow, and the box boarder.

Q8 - Duplicate the graph in Q7 and add a line showing the transaction average for each quarter for the entire company (not just NY). To accomplish this create an LOD statements and use DATETRUNC('quarter', [TR TransDate]) to group by truncated quarter (quarter and year). Use a dual axis chart to add this calculated field to the graph.

Formatting: Change the y-axis scale to have a fixed start of 12K and if added automatically, remove the measure name pill from the Color marks area. Add an annotation stating "Company-wide average transaction sales amount per quarter" to the newly added company average, remove the arrow, and the box boarder.

Q9 - Show total quarterly sales (for all years) for each employee in NY. Display TR TransDate Years and Quarters on the x-axis and total quarterly sales on the y-axis. Add a reference line showing the average quarterly sales for all employees in NY. Also add a trend line showing the overall trend for the employees in NY (show a single trend line, not one trend line per employee).

Formatting: Change the y-axis scale to have a fixed start of 6,5 million, and change the y-axis and x-axis labels to "Quartely Sales per Employee" and "Transaction Quarter", respectively.

Q10 - Add two company quarterly sales averages to Q9 (duplicate the graph in Q9 and add to this graph). The first average should show average quarterly sales for all employees (not just NY), but only one average across all quarters. The second average should show average quarterly sales for each quarter for all employees. To accomplish this create two LOD statements (and use them in two separate graphs). Use DATETRUNC('quarter', [TR TransDate]) to group by truncated quarter (quarter and year) in the second LOD statement. Note that employees are parts of sites and that if you aggregate on employees the results from the LOD statement will be filtered based on the view filter that excludes all sites except NY. Therefore you need to somehow calculate average sales per employee within the LOD statement

without grouping by employee, i.e., calculate average sales as sum of sales divided by the number of unique employees or use two LOD statements, a first statement that calculates quarterly sales per employee and a second statement that calculates the average of the output from the first statement.

Formatting: Add annotations to the two lines stating: “Company-wide average quarterly employee sales across all periods” and “Company-wide average quarterly employee sales each quarter”.

Q11 - Create a scatter plot with sales quantity on the y-axis, sales on the x-axis, and product number as bubble color. Do not aggregate the measures (each bubble should represent a single data point).

Formatting: None

Q12 - Create a scatter plot with purchases on the y-axis, sales on the x-axis, returns on bubble size, and site code (Site SiteCode) as bubble color. Aggregate based on site code and truncated week number.

Formatting: Use filled bubbles with boarders, set the x-axis and y-axis tick marks to 1M increments, remove the gridlines, and add on one of the larger bubbles an annotation stating “Size indicates amount of returns”

Q13 - Create a histogram of sales (dollar value).

Formatting: Change the bin size to 250.

Q14 - Create a box-and-whisker plot of returns per customer (Ent EntityID) for each truncated month (TR TransDate). Show returns on the y-axis, truncated month on the x-axis, and Ent EntityID in detail. Also show the site code as color.

Formatting: None

Q15 - Create a horizontal bar chart showing the top three employees per site in terms of total sales in 2015. Show sales on the x-axis and site code and employee last name on the y-axis.

Formatting: Remove the field labels for rows. Rank order the result (sort) within each site.

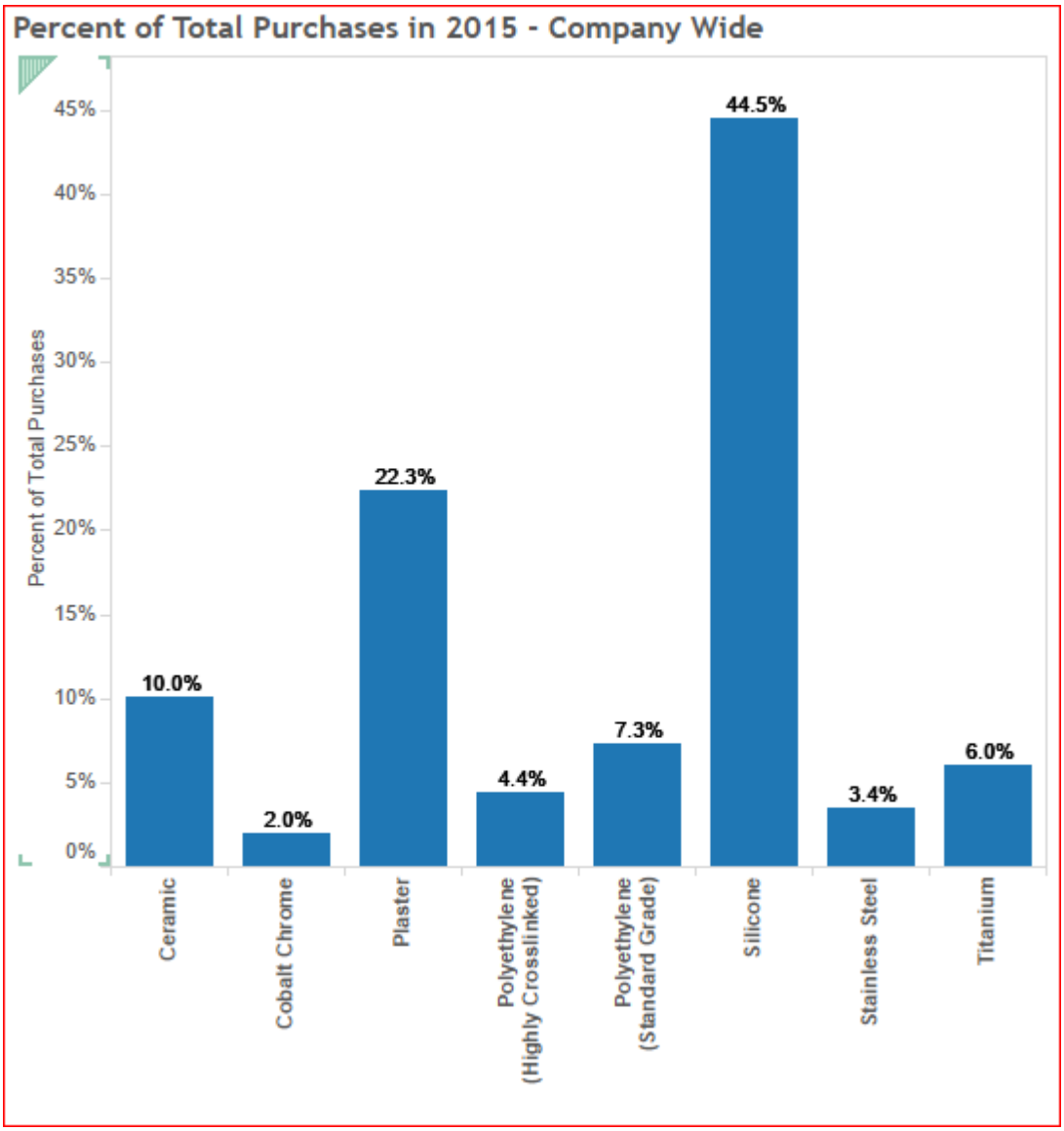
Q16 - Duplicate Q1 and Q2 and name them Q16b and Q16c, respectively. Change the Q16b and Q16c titles to “Percent of Total Purchases – Company Wide” and “Percent of Total Purchases per Site”, respectively. Remove the date filter in both graphs and then create a dashboard with Q16b on top and Q16c at the bottom. Create a new highlight action that selects all bars related to a specific material when the material is selected in one of the dashboard graphs. Also create a

filter based on TR TransDate that allows the selection of a range of dates using a slider. Place this filter at the bottom across the entire dashboard and remove the filter title.

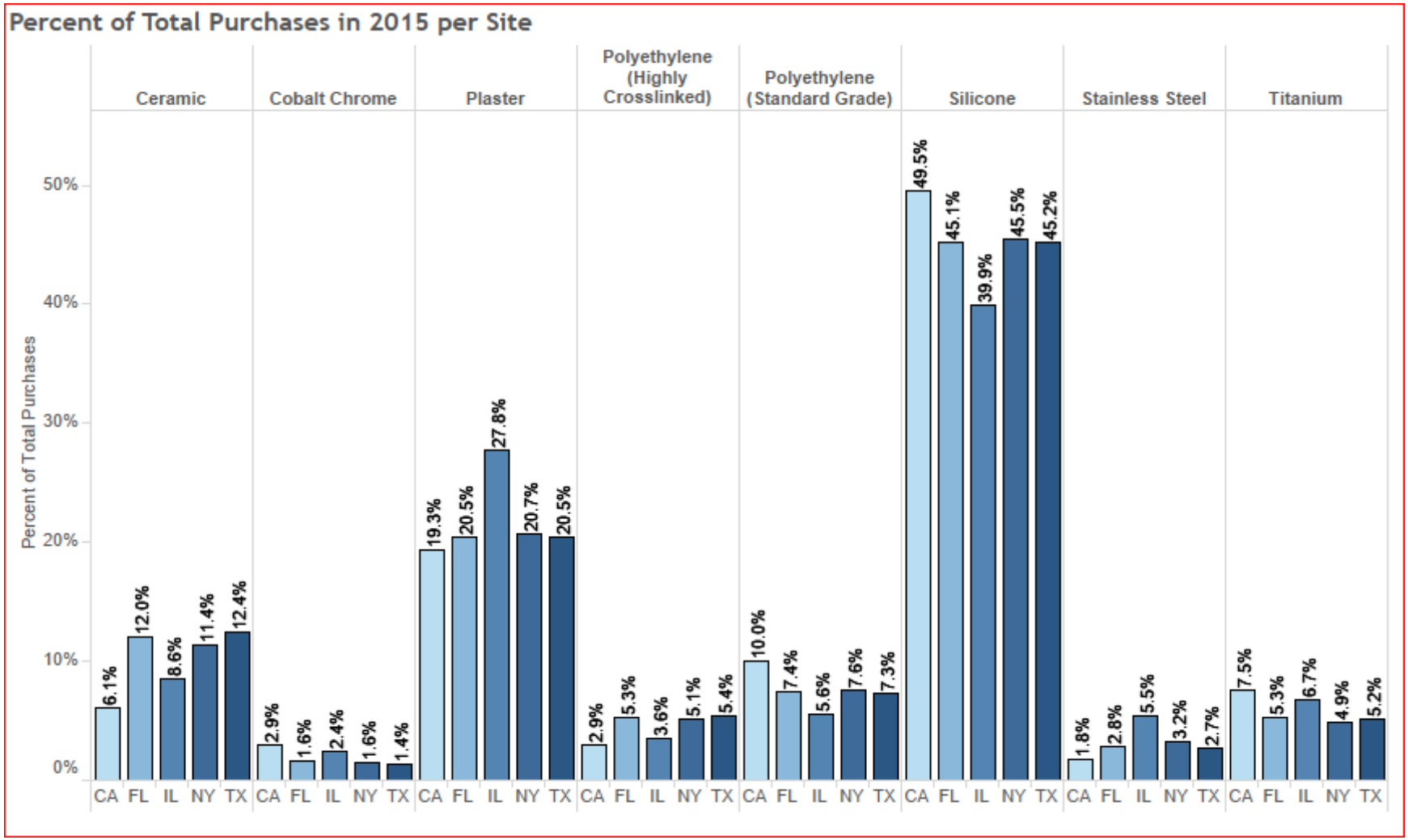
Formatting: No additional formatting.

TAKE-HOME ASSIGNMENT EXHIBIT
Solution Graphs

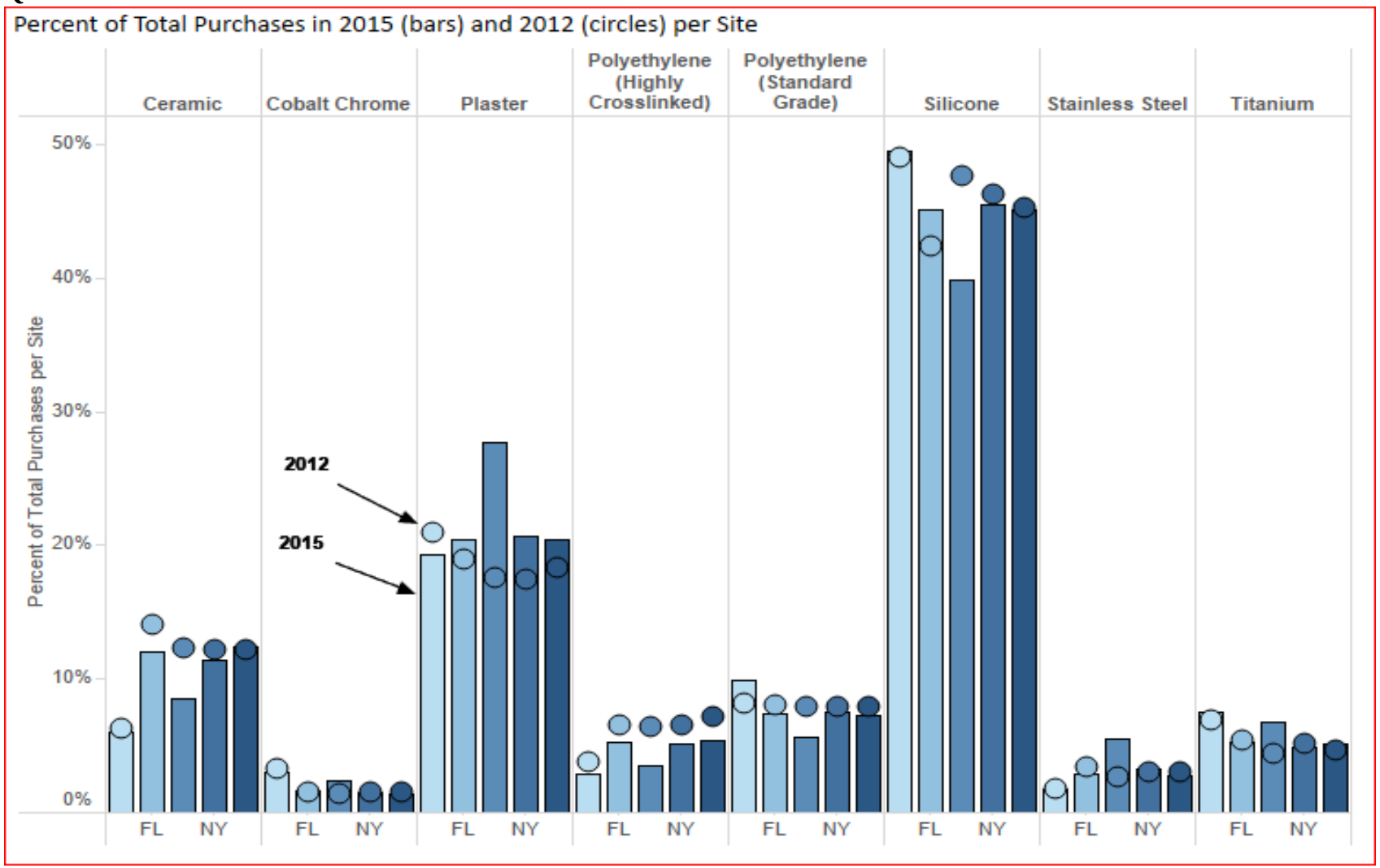
Q1



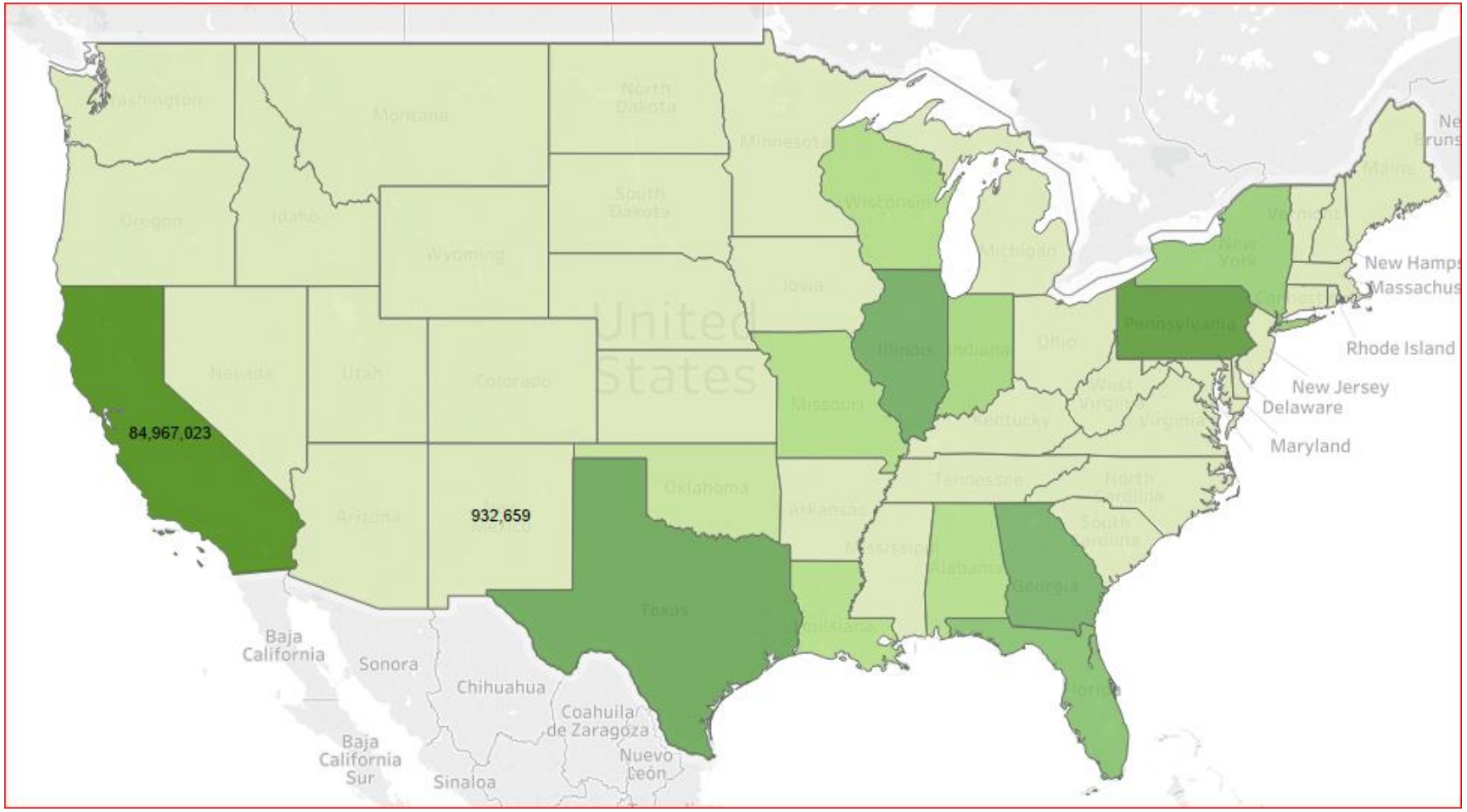
Q2


Percent of Total Purchases in 2015 per Site
Ceramic
Cobalt Chrome
Plaster
Polyethylene (Highly Crosslinked)
Polyethylene (Standard Grade)
Silicone
Stainless Steel
Titanium
Percent of Total Purchases
50%
40%
30%
20%
10%
0%
CA FL IL NY TX
6.1% 12.0% 8.6% 11.4% 12.4%
2.9% 1.6% 2.4% 1.6% 1.4%
19.3% 20.5% 27.8% 20.7% 20.5%
2.9% 5.3% 3.6% 5.1% 5.4%
10.0% 7.4% 5.6% 7.6% 7.3%
49.5% 45.1% 39.9% 45.5% 45.2%
1.8% 2.8% 5.5% 3.2% 2.7%
7.5% 5.3% 6.7% 4.9% 5.2%

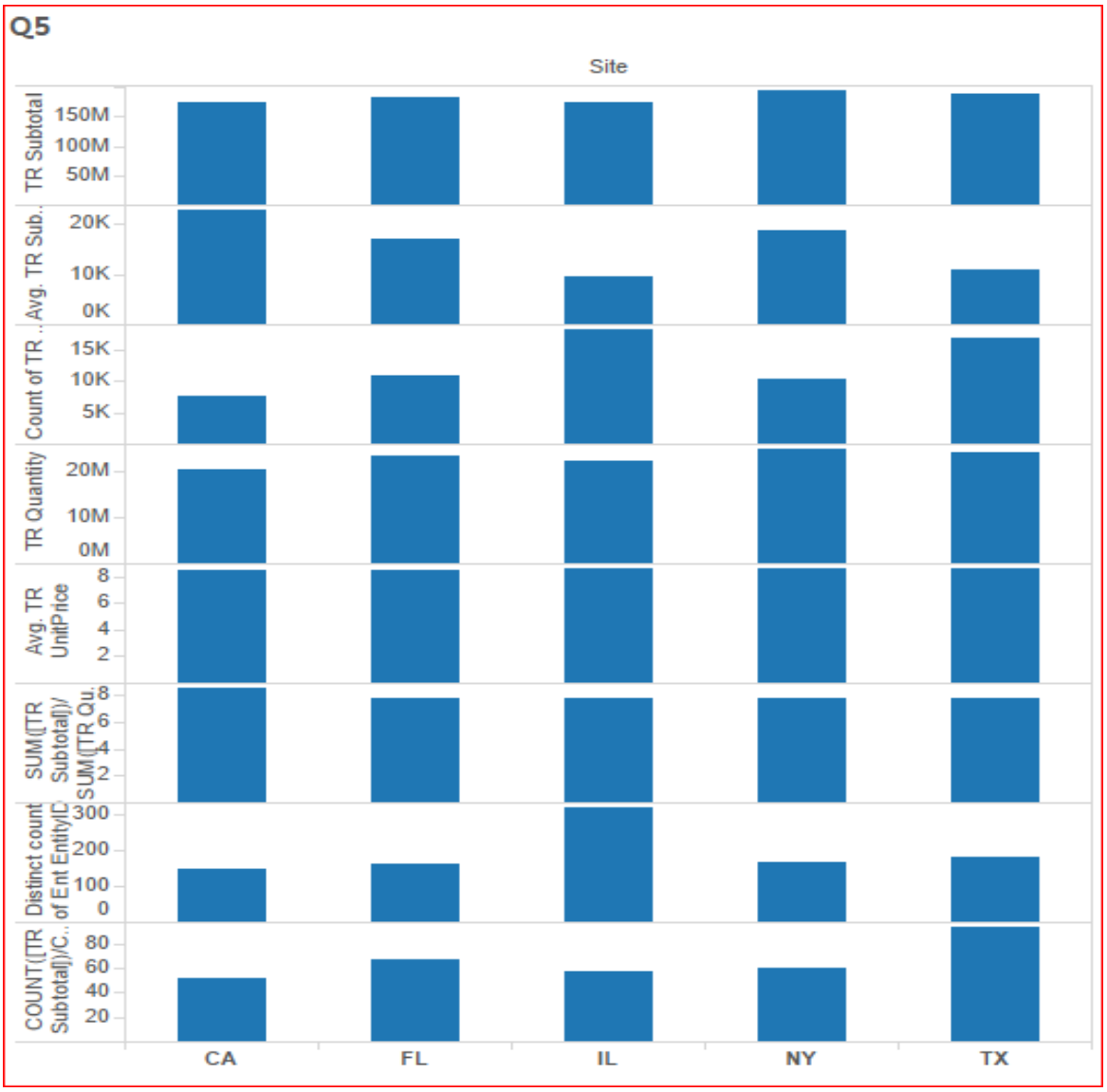
Q3


Percent of Total Purchases in 2015 (bars) and 2012 (circles) per Site
Ceramic
Cobalt Chrome
Plaster
Polyethylene (Highly Crosslinked)
Polyethylene (Standard Grade)
Silicone
Stainless Steel
Titanium
Percent of Total Purchases per Site
50%
40%
30%
20%
10%
0%
2012
2015
FL
NY

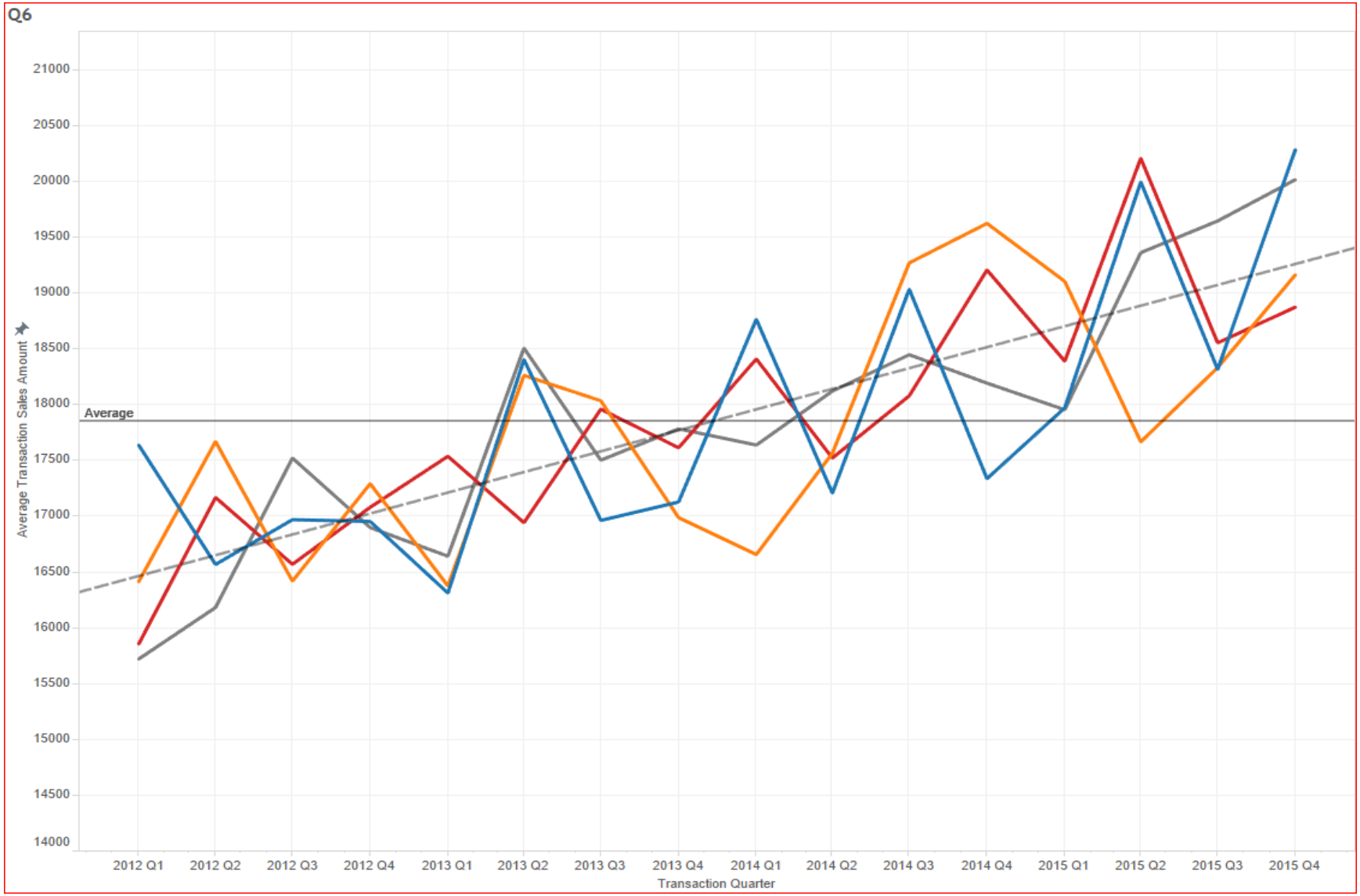
Q4


84,967,023
932,659
Pennsylvania
Rhode Island
New Jersey
Delaware
Maryland
Texas
Georgia
Baja California
Sonora
Chihuahua
Coahuila de Zaragoza
Nuevo León
Baja California Sur
Sinaloa

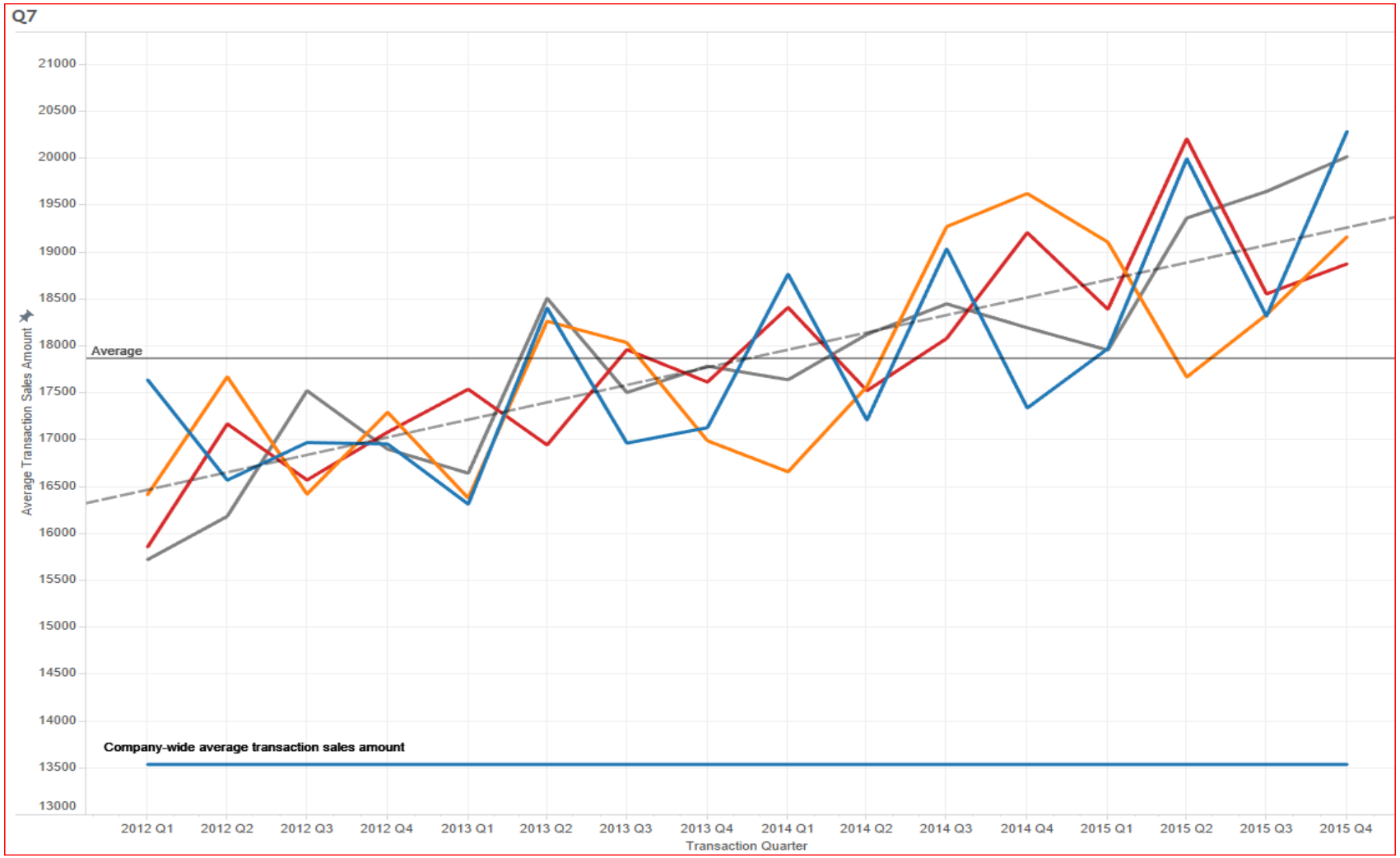
Q5


Q5
Site
TR Subtotal
150M
100M
50M
Avg. TR Sub..
20K
10K
0K
Count of TR ..
15K
10K
5K
TR Quantity
20M
10M
0M
Avg. TR UnitPrice
8
6
4
2
SUM([TR Subtotal])/SUM([TR Qu..
8
6
4
2
Distinct count of Ent EntityID
300
200
100
0
COUNT([TR Subtotal])/C..
80
60
40
20
CA
FL
IL
NY
TX

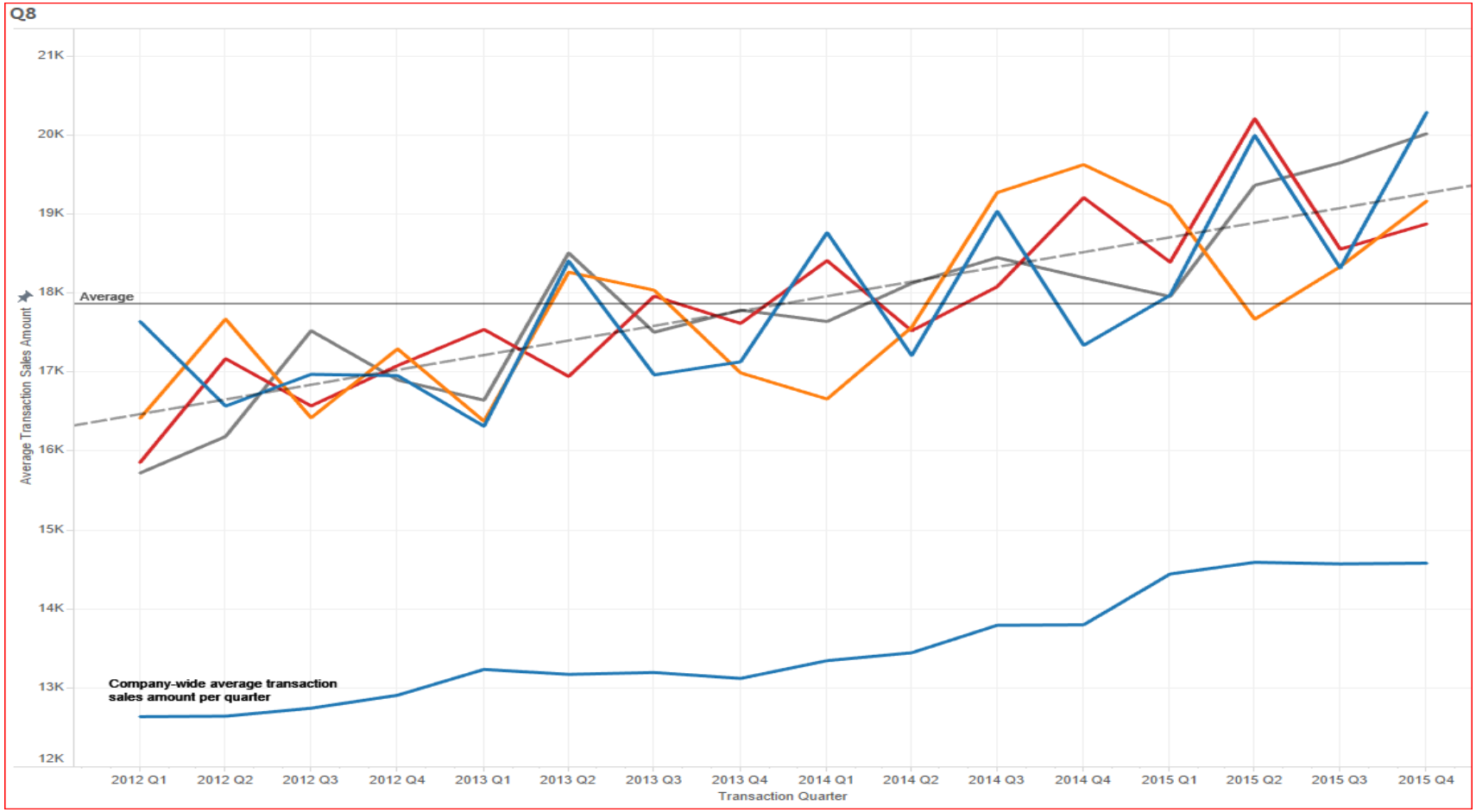
Q6


Q6
21000
20500
20000
19500
19000
18500
18000
17500
17000
16500
16000
15500
15000
14500
14000
Average Transaction Sales Amount
Average
2012 Q1
2012 Q2
2012 Q3
2012 Q4
2013 Q1
2013 Q2
2013 Q3
2013 Q4
2014 Q1
2014 Q2
2014 Q3
2014 Q4
2015 Q1
2015 Q2
2015 Q3
2015 Q4
Transaction Quarter

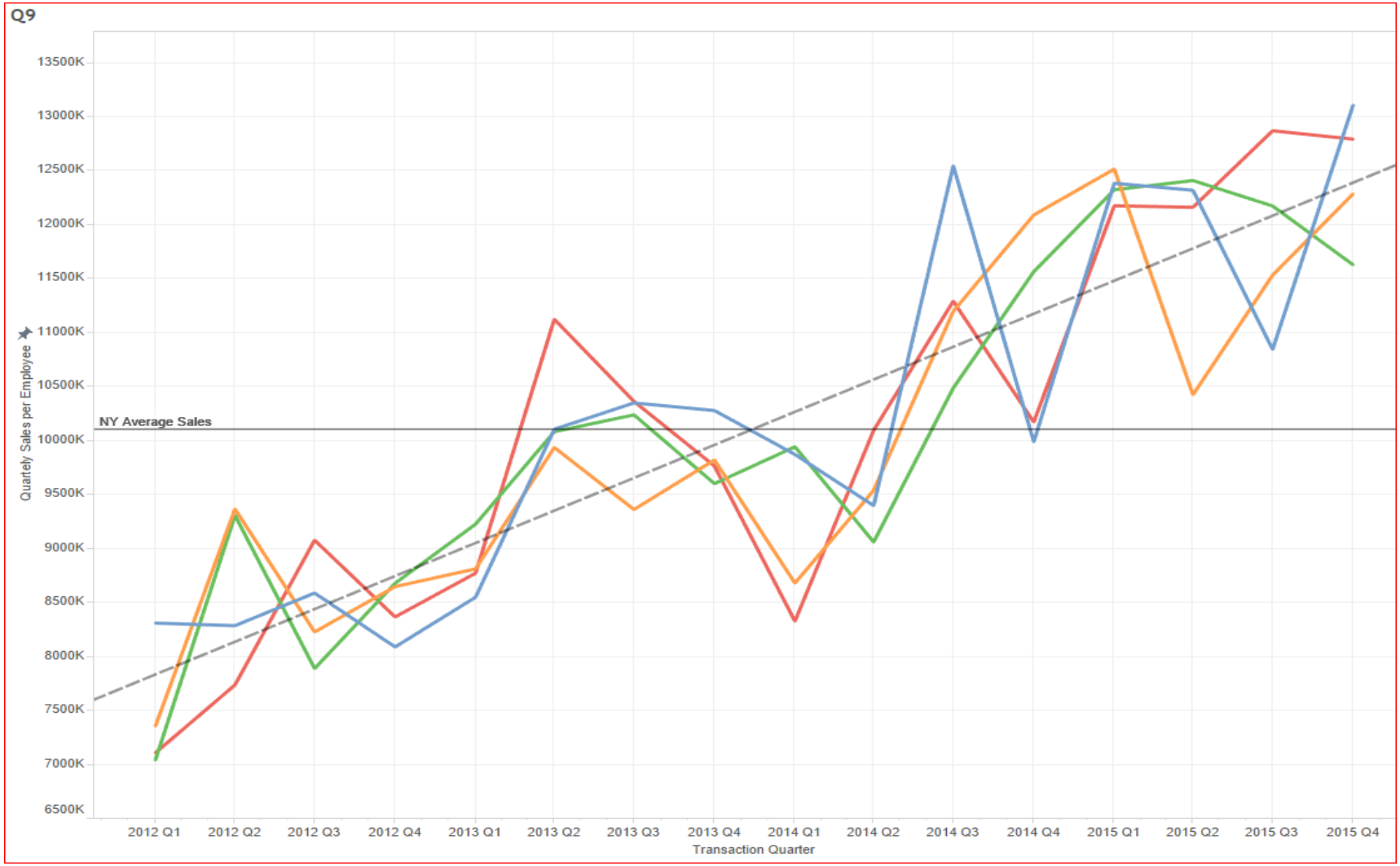
Q7


Q7
21000
20500
20000
19500
19000
18500
18000
17500
17000
16500
16000
15500
15000
14500
14000
13500
13000
Average Transaction Sales Amount
Average
Company-wide average transaction sales amount
2012 Q1
2012 Q2
2012 Q3
2012 Q4
2013 Q1
2013 Q2
2013 Q3
2013 Q4
2014 Q1
2014 Q2
2014 Q3
2014 Q4
2015 Q1
2015 Q2
2015 Q3
2015 Q4
Transaction Quarter

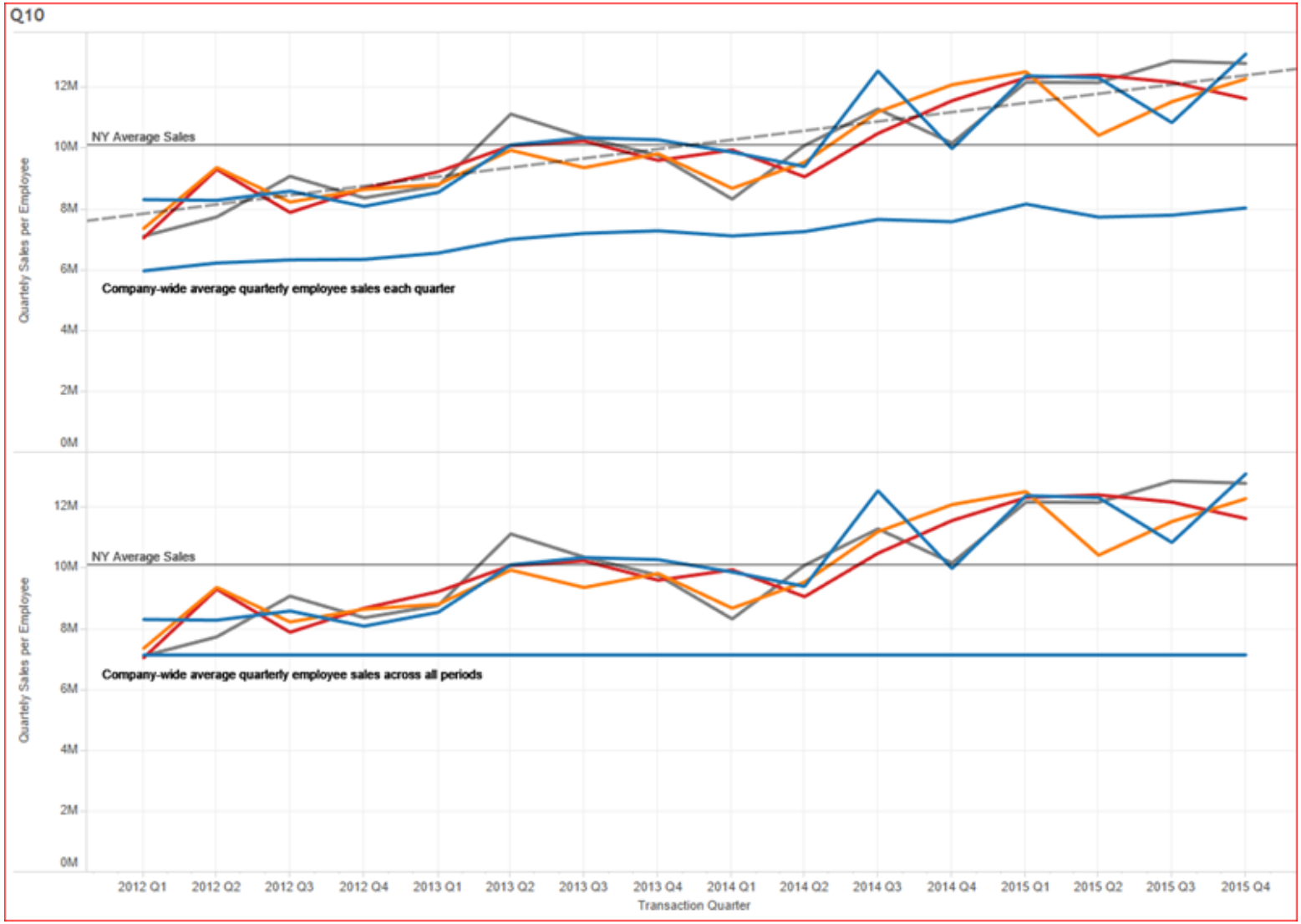
Q8


Q8
21K
20K
19K
18K
17K
16K
15K
14K
13K
12K
Average
Average Transaction Sales Amount
Company-wide average transaction sales amount per quarter
2012 Q1
2012 Q2
2012 Q3
2012 Q4
2013 Q1
2013 Q2
2013 Q3
2013 Q4
2014 Q1
2014 Q2
2014 Q3
2014 Q4
2015 Q1
2015 Q2
2015 Q3
2015 Q4
Transaction Quarter

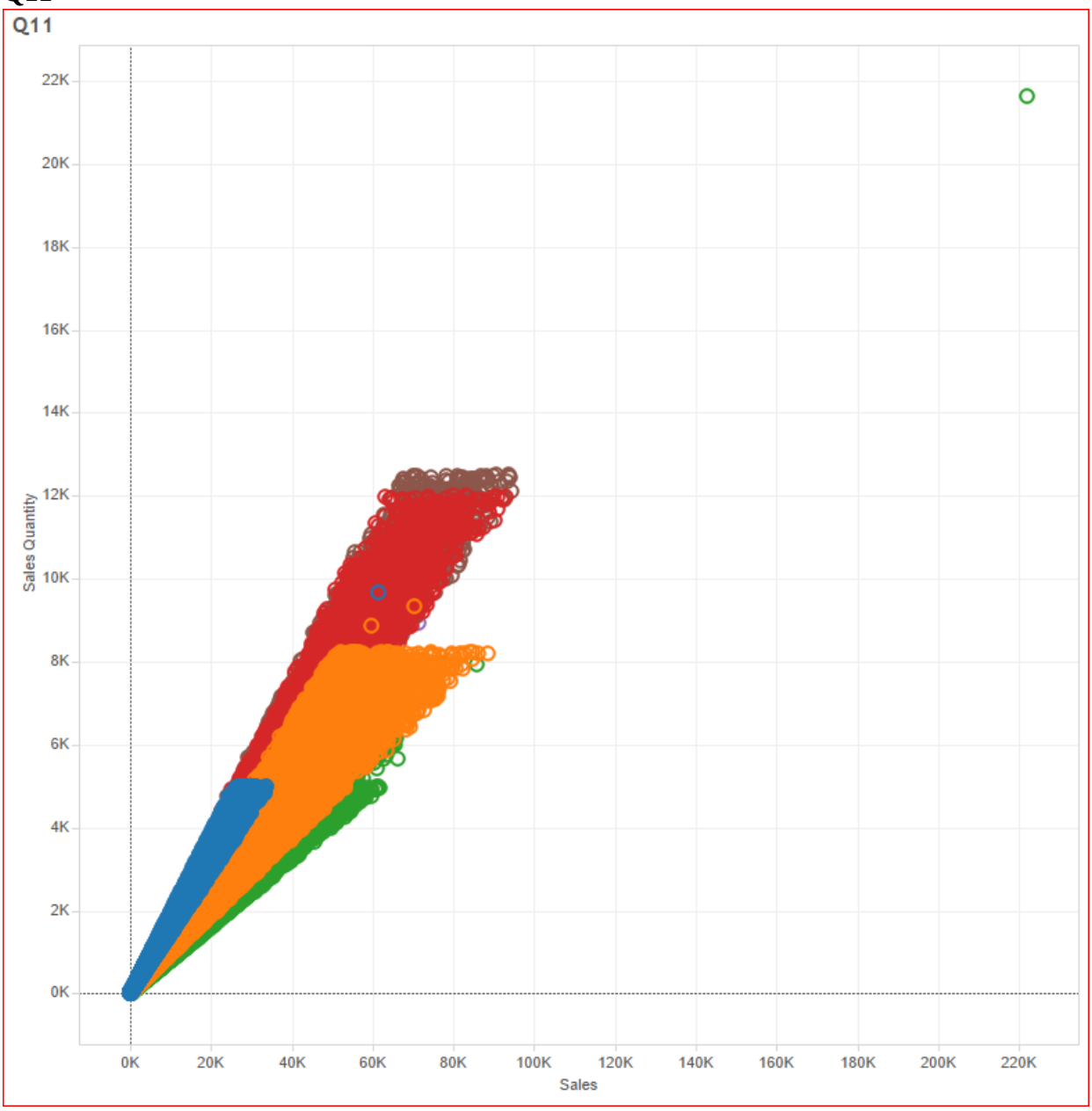
Q9


Q9
13500K
13000K
12500K
12000K
11500K
11000K
10500K
10000K
9500K
9000K
8500K
8000K
7500K
7000K
6500K
Quartely Sales per Employee
NY Average Sales
2012 Q1
2012 Q2
2012 Q3
2012 Q4
2013 Q1
2013 Q2
2013 Q3
2013 Q4
2014 Q1
2014 Q2
2014 Q3
2014 Q4
2015 Q1
2015 Q2
2015 Q3
2015 Q4
Transaction Quarter

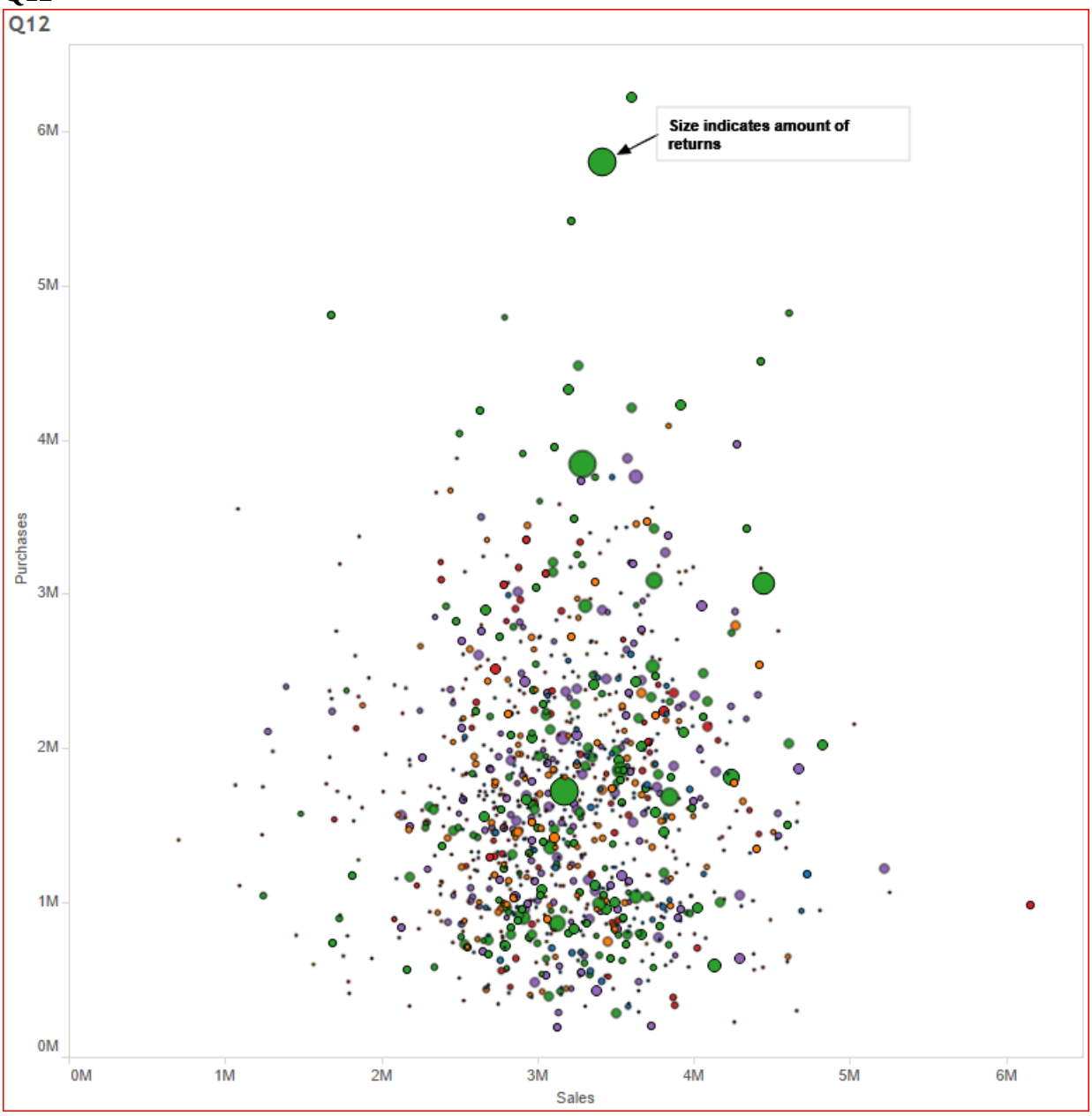
Q10


Q10
12M
10M
8M
6M
4M
2M
0M
Quartely Sales per Employee
NY Average Sales
Company-wide average quarterly employee sales each quarter
Company-wide average quarterly employee sales across all periods
2012 Q1
2012 Q2
2012 Q3
2012 Q4
2013 Q1
2013 Q2
2013 Q3
2013 Q4
2014 Q1
2014 Q2
2014 Q3
2014 Q4
2015 Q1
2015 Q2
2015 Q3
2015 Q4
Transaction Quarter

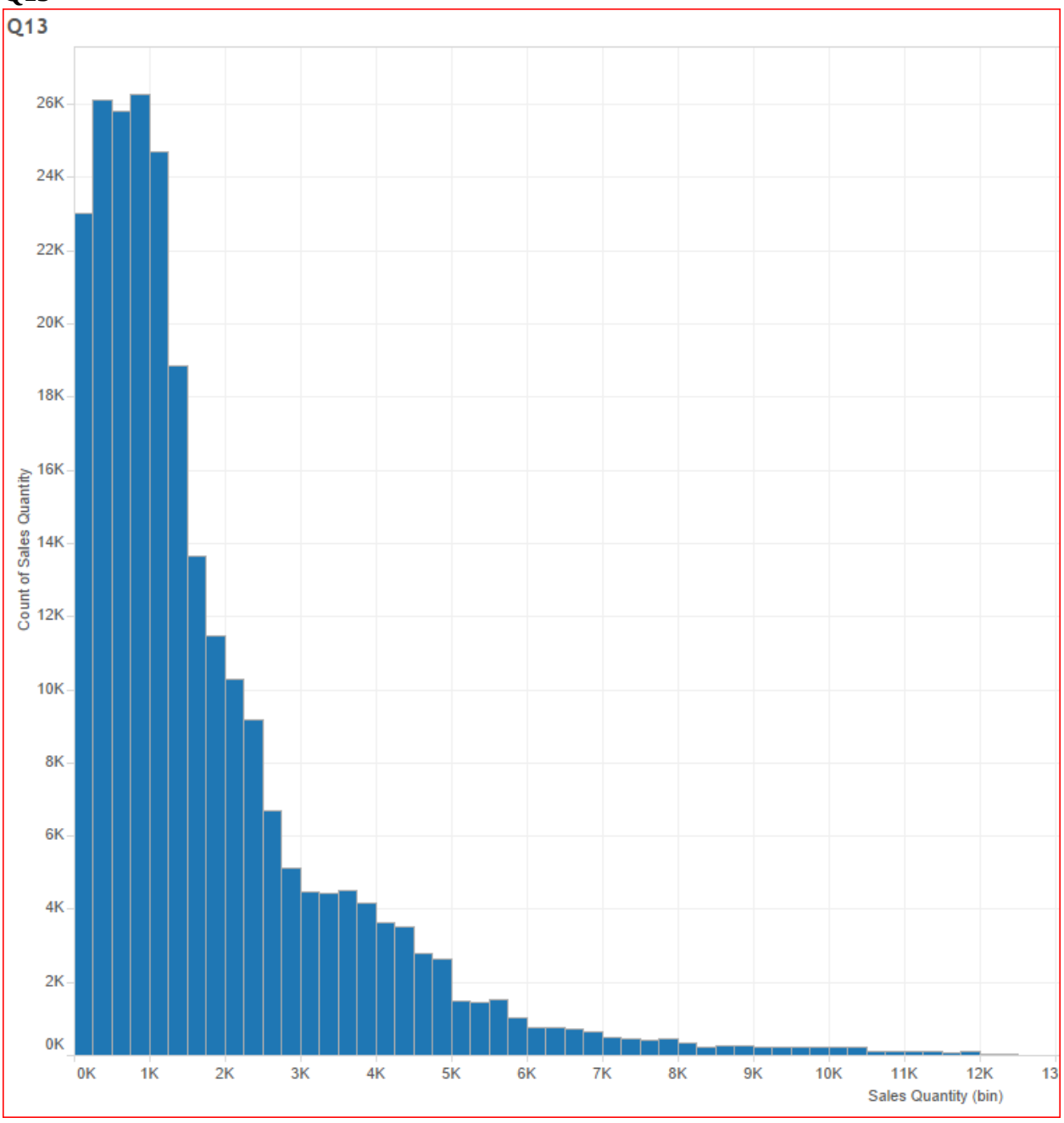
Q11


Q11
22K
20K
18K
16K
14K
12K
10K
8K
6K
4K
2K
0K
Sales Quantity
0K
20K
40K
60K
80K
100K
120K
140K
160K
180K
200K
220K
Sales

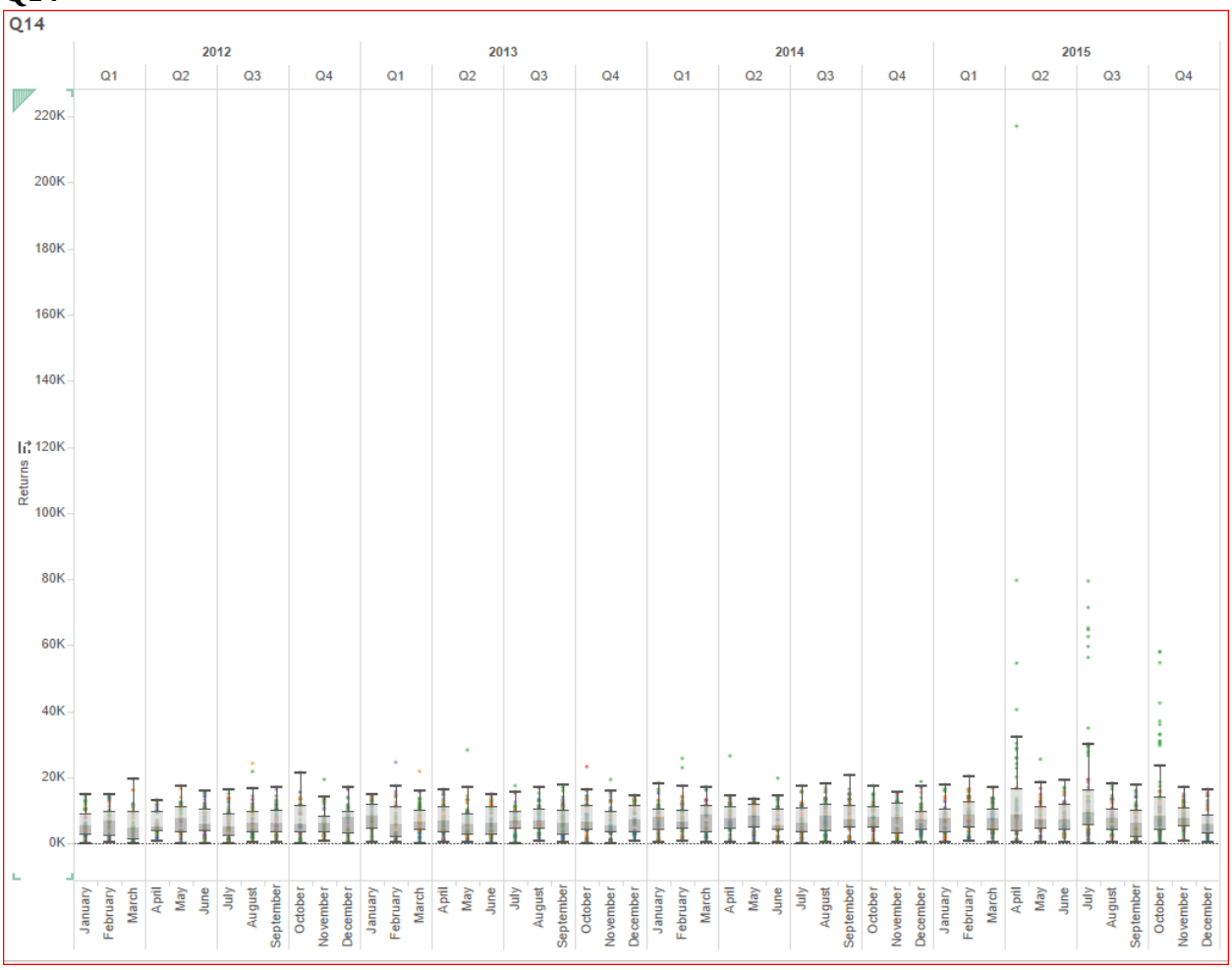
Q12


Q12
Size indicates amount of returns
6M
5M
4M
3M
2M
1M
0M
Purchases
0M
1M
2M
3M
4M
5M
6M
Sales

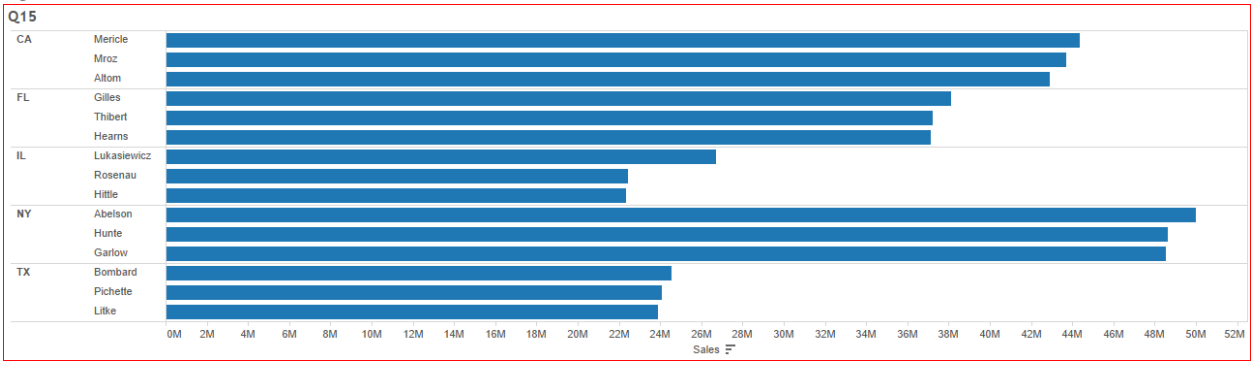
Q13


Q13
26K
24K
22K
20K
18K
16K
14K
12K
10K
8K
6K
4K
2K
0K
Count of Sales Quantity
0K
1K
2K
3K
4K
5K
6K
7K
8K
9K
10K
11K
12K
13
Sales Quantity (bin)

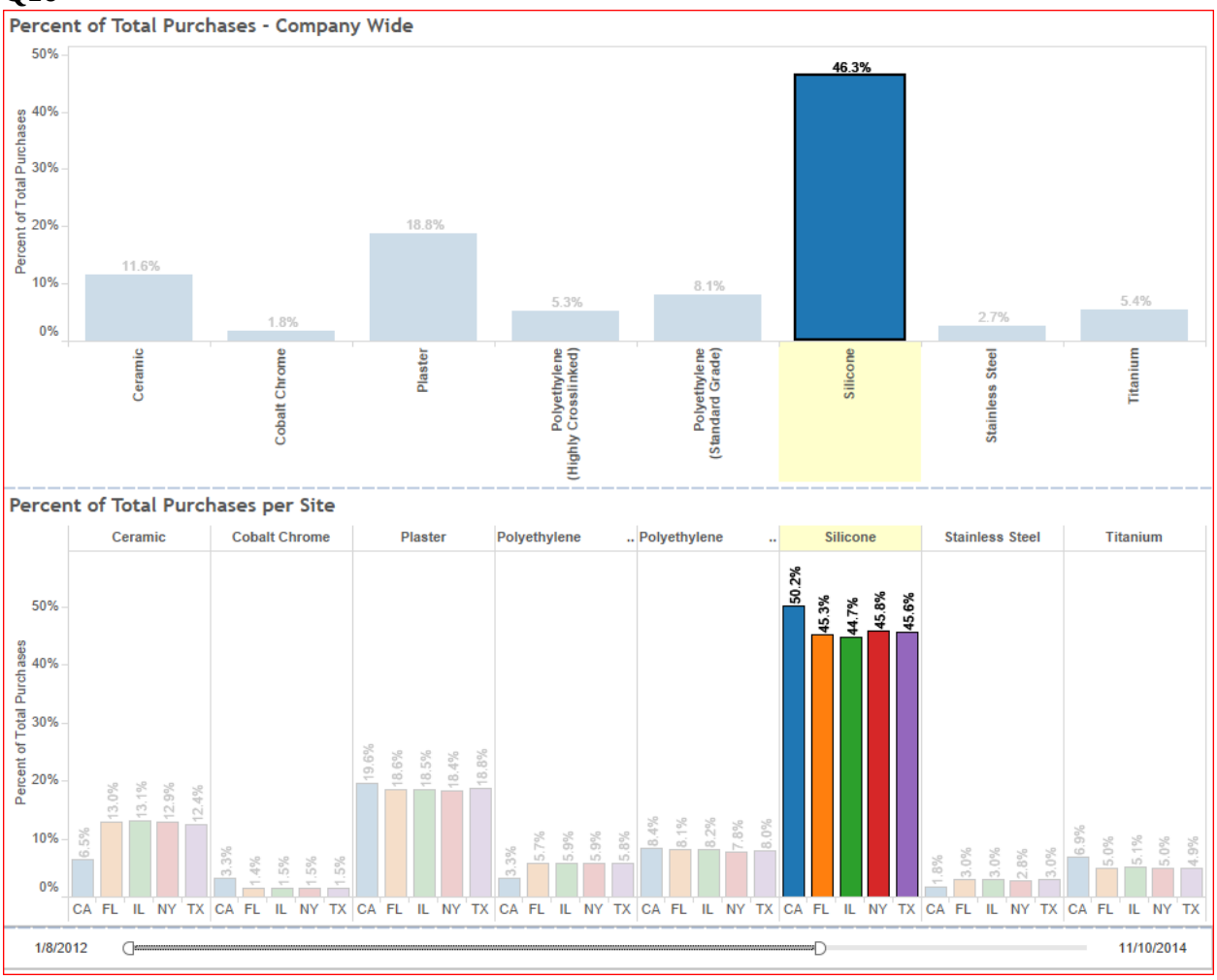
Q14


Q14
2012
2013
2014
2015
Q1
Q2
Q3
Q4
Returns
220K
200K
180K
160K
140K
120K
100K
80K
60K
40K
20K
0K
January
February
March
April
May
June
July
August
September
October
November
December

Q15


Q15
CA
Mericle
Mroz
Altom
FL
Gilles
Thibert
Hearns
IL
Lukasiewicz
Rosenau
Hittle
NY
Abelson
Hunte
Garlow
TX
Bombard
Pichette
Litke
0M 2M 4M 6M 8M 10M 12M 14M 16M 18M 20M 22M 24M 26M 28M 30M 32M 34M 36M 38M 40M 42M 44M 46M 48M 50M 52M
Sales

Q16


Percent of Total Purchases - Company Wide
Percent of Total Purchases
50%
40%
30%
20%
10%
0%
11.6%
1.8%
18.8%
5.3%
8.1%
46.3%
2.7%
5.4%
Ceramic
Cobalt Chrome
Plaster
Polyethylene (Highly Crosslinked)
Polyethylene (Standard Grade)
Silicone
Stainless Steel
Titanium
Percent of Total Purchases per Site
Ceramic
Cobalt Chrome
Plaster
Polyethylene ..
Polyethylene ..
Silicone
Stainless Steel
Titanium
6.5% 13.0% 13.1% 12.9% 12.4%
3.3% 1.4% 1.5% 1.5% 1.5%
19.6% 18.6% 18.5% 18.4% 18.8%
3.3% 5.7% 5.9% 5.9% 5.8%
8.4% 8.1% 8.2% 7.8% 8.0%
50.2% 45.3% 44.7% 45.8% 45.6%
1.8% 3.0% 3.0% 2.8% 3.0%
6.9% 5.0% 5.1% 5.0% 4.9%
CA FL IL NY TX
1/8/2012
11/10/2014

CASE

A Picture is Worth a Thousand Words: Using Interactive Data Visualization to Assess Fraud Risk

"What now?" Bryan thought to himself. Bryan had been assigned to the Acme Company for the third year in a row and he was now the audit senior on the project. Acme had previously been a fairly low-risk audit client, but the company's stock price had plummeted in the last year, and management was now aggressively pushing sales to increase Acme's market share and eventually profitability. While not a specific fraud risk, this increasingly aggressive environment at Acme was a concern that had come up during the audit engagement team's fraud brainstorming session. Adding to this concern, the team had also been informed by Acme's internal audit department that a recent caller to Acme's anonymous tip line had claimed that at least one person in the company was involved in fraud of some kind. No further information was provided. During the fraud brainstorming session, however, the audit engagement team identified a number of more specific fraud risks. Bryan had now been tasked with using "data visualization" to perform an initial evaluation of the risks identified during the fraud brainstorming session.[1] However, he was not sure where to start. He didn't want to admit it, but he didn't really know much about data visualization.

The Client

The Acme Company is a medical device manufacturer specializing in hip implants. Acme's founder, Dr. Ralph Anderson, started producing custom-made hip implants in 1922, and three years later Acme was formed. Since its inception until recent times, Acme has been on the forefront of innovation and has been the US market leader for hip implants for the past three decades. However, competition has intensified during the past five years, putting enormous pressure on Acme's leaders and employees.

In 2015, Acme generated total sales of almost a billion dollars (see Exhibit 1). Although net sales had increased by 7.0 percent over the previous year, this increase was four percent below the industry average. Also, Acme's gross profit margin decreased from 37 percent in 2014 to 36 percent in 2015 and the total cost of sales was $579 million in 2015 compared to $542 million in 2014. Despite these factors, Acme's profitability had improved in the past few years due to sales growth and, more importantly, as a result of cutting funding for research and development along with selling, general, and administrative expenses. Going forward, however, Acme's leadership is determined to improve sales by increasing its research and development

[1] "Data visualization is the presentation of data in a pictorial or graphical format. It enables decision makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns. With interactive [data] visualization, you can take the concept a step further by using technology to drill down into charts and graphs for more detail, interactively changing what data you see and how it's processed." (SAS, 2017).

efforts while maintaining its operating profit margins by making more efficient use of selling, general, and administrative resources.

Acme is a decentralized organization, with five manufacturing and sales sites across the US employing a total of 28 product managers and 29 sales managers. Each of Acme's sites operates independently and is responsible for procurement of raw materials, production, and sales of finished products in its own geographic region. Sales managers maintain their own customer relationships, while product managers monitor inventory levels and order raw materials as needed. The production team leader explained that silicone and plaster are used to create the molds for all the company's hip replacements. The materials poured into molds vary by product: some are metal, some are plastic, and some are ceramic. Each final product is built from a specific amount of raw materials as outlined in Acme's Manufacturing Specifications (see Exhibit 2).

Each site also has flexibility in its compensation arrangements, but Acme generally pays sales managers using a combination of a relatively small salary, a commission calculated as a percentage of sales, and a quarterly bonus dependent on attaining certain sales targets. In contrast, a much larger portion of the product managers' compensation is a fixed salary with smaller quarterly performance-based bonuses. To cut costs, Acme has not authorized any salary increases in the last three years (making performance-based bonuses more important to the employees) and has also made the quarterly performance targets more difficult to attain. However, to balance cost cutting with Acme's strategic objective to increase sales, the company substantially increased the size of its potential bonuses in the last year.

The Audit Engagement

Earlier in the week the audit engagement team conducted a fraud brainstorming session in which it identified a wide range of potential fraud risks. By the end of the meeting, the team had identified one specific type of fraud as high risk (see Exhibit 3). The team also identified some typical red flags associated with this fraud. After the meeting, Sarah, the team manager, and Bryan, the audit senior, had the following conversation:

Sarah: Bryan, why don't you take a look at the fraud risk that we identified during the fraud brainstorming meeting?

Bryan: Okay, but I'm not sure where to start.

Sarah: As you know, the national office is encouraging us to use analytics and data visualization throughout the various stages of the audit.

Bryan: That's true. We went through a short session on Tableau during new senior training.

Sarah: I've also been hearing some rumors about more and more engagement teams using Tableau. It would be great if we could use it too. I don't want us to fall behind the curve.

Bryan: Absolutely, although I must admit that I don't remember much from the training–it was more of a demo than an actual hands-on exercise.

Sarah: Okay, so why don't we both spend some time this weekend familiarizing ourselves with Tableau? You can then start the analysis first thing Monday morning.

Bryan: Sure, but how do we get the data to analyze?

Sarah: I've already obtained the data that we need from Acme. They were helpful and made the table structure easier for us to work with. The data were sent in csv format together with file descriptions [see Exhibits 4-6]. The check figures in these file descriptions are important to use to validate that all data files were imported correctly. We should be able to replicate the numbers in the check figures before we start analyzing the data. The ERD diagram will tell us which tables to join, which fields to use in each join, and what types of joins to use. Finally, the data dictionary can be useful for understanding what data each field contains. This might be of particular value when we create calculated fields. Also, I noticed something unusual when reviewing the data dictionary: all sales and purchase order transactions are in one large table.

Bryan: All right, I will make sure to look at these documents before I start playing around with the data.

CASE REQUIREMENTS

1) Understanding Sales

   a) Use Tableau to gain an overall understanding of *sales* at Acme. In your analyses try to answer the following questions:

      i) How has total sales changed for Acme overall and for different sites? What are the drivers of differences among the sites? Potential drivers to analyze include:

         - changes in average unit prices across different products due to
            - changes in unit prices for specific products, or
            - changes in sales mix of products with different average unit prices (e.g., unit prices for specific products have not changed, but a site may be selling more (or less) of expensive products relative to cheaper products
         - changes in quantity sold due to changes in
            - the number of customers, or
            - quantity sold per customer

      ii) What is the sales mix (i.e., what is the contribution of different products to total sales)? Has the sales mix changed over time? What was the sales mix for different sites in the most recent year? Does the sales mix differ for different sites? Is there any relation between the sales mix and performance differences detected earlier?

         Rather than using a pie chart, show percentage of total in a bar graph by using a quick table calculation (do not use a stacked bar chart).

      iii) Are there any dominant customers overall? Are there any dominant customers at different locations?

         When performing this analysis at the site level, create a visualization that shows the top 5 customers for each site. While not required, you may want to consider also using a quick table calculation to show percent of total site sales for each of the top 5 customers.

   b) Create a Tableau Story to present your analyses.[2] In your story, start at a high (i.e., general) level and then drill down into details. Make sure that you include analyses that

[2] Stories are increasingly used in data visualization to communicate about data analyses. Stories contain a sequence of story points that each present a narrative visualization pulled from analyses performed in various worksheets and dashboards. The story points, taken as a whole, show how different visualizations are connected and are useful for explaining the analytical process and revealing how work was performed.

answer all questions in requirement 1(a). One effective approach to keep the reader's interest is to show a figure and ask a question related to this figure (e.g., for a figure that shows sales over time for different sites, you could ask, "Which site has experienced the lowest/highest sales growth?") The next story point should answer the question and show the same figure highlighting the area of the figure that shows sites with the lowest/highest sales growth.

To include multiple graphs in a single story point, first create a dashboard that includes all the graphs you want to show in one display and then include the dashboard in the story.

2) Fraud Risks

a) Evaluate the ***kickback fraud risk***[3] identified in the fraud brainstorming session (see Exhibit 3). Focus your analysis on examining the red flags associated with the kickback fraud risk, but note that you may not have the required data to evaluate all red flags. For each red flag, perform your analysis at various levels. For example, begin by examining trends over time at the aggregate level, then perform similar analyses at the site level and at the employee/supplier/raw material level. If you identify any anomalies in one of these analyses, be sure to explore this finding further. For example, if total purchase order amount appear to be abnormal for a specific site in a specific year, then analyze this subset of the data further (e.g., examine whether the abnormal purchase order amounts is associated with specific employees, raw materials, suppliers, etc.). If you identify something suspicious make sure to perform a more detailed examination of these data points for the other red flags (but still also perform the high-level to more detailed analyses when analyzing the other red flags).

Your analyses of the red flags should include (but not be limited to) the following specific analyses:

i) When analyzing the red flag related to abnormal changes in total purchases, analyze changes in total purchase over time and comparing these changes in purchases to changes in sales for the same time period. Use a dual combination graph to show purchases and sales on the same graph.

ii) When comparing unit prices for different items purchased at the product manager level (or even product manager - supplier level) use a LOD statement and include a comparison to the company mean for the same item and period.

iii) When analyzing the red flag related to abnormal increase in inventory levels, consider analyzing changes in inventory levels over time by comparing raw material purchases

[3] The fraud brainstorming meeting specifically identified kickback fraud associated with purchases (rather than sales) as high fraud risk.

to raw materials sales. To perform this analysis you need to calculate how many units of each raw material is sold when different products are sold – see Exhibit 2. However, note this analysis will generate some results that do not make sense. More specifically, you will see that some changes in inventory levels are consistently negative or positive (it would generally make sense that changes in inventory levels should be close to zero, which would indicate that inventory levels are not increasing or decreasing). Ignore this and instead see if there are any deviations from the consistent negative or positive changes, e.g., Cobalt consistently decreases by 1.4 to 1.8 million units per month but there is no significant deviations to this trend so there does not appear to be anything unusual with Cobalt.

iv) Displaying the inventory levels changes analysis using a control chart.

b) Document your fraud risk evaluation by creating a Tableau Story. In your story, start at a high (i.e., general) level and then drill down into details to provide the accounting department with insights that they can use to analyze the fraud risk further. Make sure that you include analyses of all red flags you examine. As explained earlier, one effective approach to keep the reader's interest is to show a figure and ask a question related to this figure (e.g., for a figure that shows sales over time for different sites, you could ask, "Do you see anything unusual in sales over time for each site?") The next story point should answer the question and show the same figure, highlighting the area of the figure that looks unusual. Next, drill down and look at the highlighted area in more detail (e.g., if a specific site looked abnormal, then the analysis could consider different employees at that site). This drill-down can be presented either in a new figure below the main highlighted figure or in a subsequent story point. To keep the readers interest, you will likely have to exclude some of your analyses, especially analyses that did not indicate any problems. To make the reader feel more comfortable with the completeness of your analyses, you can mention that you performed these analyses.

EXHIBIT 1
Consolidated Operations
(in millions)

	2015	2014	2013	2012
Net Sales	**$ 913**	**$ 854**	**$ 808**	**$ 717**
Cost of Sales	579	542	480	415
Gross Profit	**$ 334**	**$ 311**	**$ 329**	**$ 302**
Research and development	57	59	66	65
Selling, general, and administrative	191	196	234	243
Recall charges, net of insurance proceeds	5	5	4	3
Intangible asset amortization	3	2	2	1
Total operating expenses	**$ 256**	**$ 262**	**$ 306**	**$ 312**
Operating income	78	49	23	(10)
Other income (expense), net	1	2	(5)	11
Earnings before income taxes	**$ 79**	**$ 51**	**$ 18**	**$ 1**
Income taxes	11	8	2	-
Net earnings	**$ 68**	**$ 44**	**$ 16**	**$ 1**

EXHIBIT 2
Acme Company Manufacturing Specifications

Acme Co.

Producing medical devices since 1925

Manufacturing Specifications

Prod #	Product Name	Mat#	Material Name	Qnt Used
70041	Metal and Plastic Hip	300328	Stainless steel	0.5
		300922	Polyethylene (standard grade)	1
		400221	Silicone	0.2
		400222	Plaster	0.2
70042	Metal and Crosslinked Polyethylene Hip	300321	Titanium	0.4
		300932	Polyethylene (highly crosslinked)	1
		400221	Silicone	0.3
		400222	Plaster	0.25
70051	Metal-on-Metal Hip (Titanium)	300321	Titanium	0.6
		400221	Silicone	0.1
		400222	Plaster	0.2
70052	Metal-on-Metal Hip (Stainless steel)	300328	Stainless steel	0.5
		400221	Silicone	0.1
		400222	Plaster	0.1
70053	Metal-on-Metal Hip (Cobalt chrome)	300339	Cobalt chrome	1
		400221	Silicone	0.1
		400222	Plaster	0.15
70061	Ceramic-on-Ceramic Hip	300773	Ceramic	1
		400221	Silicone	0.2
		400222	Plaster	0.1

Notes: Grey text indicates materials that are common across all products.

EXHIBIT 3
Fraud Risks Identified in Fraud Brainstorming Session

Fraud Description	Red Flags[a]
Kickback: A percentage of the transaction value paid "off-book" by a seller to a buyer as payment for having made the sale possible.	On the buyer's[b] books (when Acme purchases raw materials): • abnormal increase in purchases subtotals • abnormal increase in inventory levels • abnormal increase in raw materials purchases unit costs • abnormal reduction in payment period

[a] When looking for abnormal increases/reductions, look for changes over time (e.g., relatively large monthly increases), cross-sectional differences (e.g., one product manager with large levels or changes in purchase quantities relative to other product mangers), deviations from patterns of other measures that should follow similar patterns (e.g., purchases should increase when sales increases), etc.

[b] The seller's books can contain similar red flags, including: (1) an unusual increase in sales (volume or dollar value) to a particular customer (or for a particular sales manager-customer pair), (2) higher than normal selling prices to a particular customer (or pair), and (3) abnormally quick cash receipts. However, the fraud brainstorming meeting specifically identified kickback associated with purchases (rather than sales) as high fraud risk.

EXHIBIT 4
Check Figures

Material Master Table

Number of records: 8
No null values in any columns
Sum of Material: 2,604,058

Product Master Table

Number of records: 6
No null values in any columns
Sum of Product: 420,300

Transactions Table

Number of records: 258,602
No null values in any columns
Sum of Quantity: 1,168,069,908
Sum of UnitPrice: 2,047,028
Sum of Subtotal: 5,058,319,250
Sum of Tax: 354,113,001
First of TR_TransDate: 1/1/2012
Last of TR_TransDate: 12/31/2015
First of CompleteDate: 1/3/2012
Last of CompleteDate: 1/20/2016

Sites Table

Number of records: 5
No null values in any columns

Employee Master Table

Number of records: 57
No null values in any columns
Sum of Order Approval Authority: 81,250,000

Entity Master Table

Number of records: 1,072
Sum of Relationship Manager: 87,627,132
Number of null values - Relationship Manager: 979
Number of null values - Zip: 66

Notes: The check figures were calculated by Acme before the data files were exported. Check figures are used to validate that data are exported, transferred, and imported correctly. Before joining the tables, the check figures should be validated one table at a time. For example, first import the Transactions table and validate this table. To import the Transactions table, press Ctrl + D, select Text, browse to the case data files, and open Transactions. Please see below for instructions on how to perform the validations. After the Transactions table has been validated, remove this table from the data source canvas and then import the Material Master table and validate this table, etc.

To validate: (1) the *Number of Records* in a table, drag the Number of Records[4] measure (this measure is automatically included and is located at the bottom of the Measures on the left side of the screen) to Text in the marks area. By adding this field to a worksheet and summing (or counting) it, the number of records is returned; (2) the *Sum* of a column, drag the column to Text in the marks area. If the column is classified as a *Dimension*[5] then use the dropdown menu to change the aggregation to a sum (use Measure → Sum); (3) *First and Last Dates* of a date column, drag the date column to Text in the marks area and change the aggregation performed on the column to Measure → Min (or Max) to obtain the earliest (most recent) date; and (4) the *Number of Null Values* in a column, drag the column to be counted to Text in the marks area and change the aggregation performed on the column to Measure → Count. This returns the number of non-null rows in the column. To manually obtain the number of null values in the columns, take the difference between the Number of Records in the table and the number of non-null values in the column.[6]

[4] *Number of Records* is a calculated field with the number 1 that simply adds a new field with 1 on each row.
[5] *Dimensions* are fields that contain discrete data that either cannot be aggregated (e.g., it is not possible to sum all names in a table) or that have no practical meaning if aggregated (e.g., the sum of all zip codes in a table is useless information). *Measures* are fields that contain numerical data that when aggregated have practical meaning (e.g., the sum of TR Quantity for all sales transactions represents the total number of units sold).
[6] Alternatively, change the pill with the count to subtract the count from the total number of records; e.g., replace COUNT([Ent RelationshipManager]) with SUM([Number of Records])-COUNT([Ent RelationshipManager]).

EXHIBIT 5
Data Dictionary[7]

Material Master Table

Physical Name	Data Type	Req'd	PK	Notes
MaterialCategory	CHAR(10)	☑	☐	Constant stating raw material
Material	CHAR(10)	☑	☑	Material item identifier
MaterialDescription	CHAR(10)	☐	☐	Description of material

Notes: The Material Master table contains information about different raw materials purchased by Acme. All fields in this table should be dimensions.

Product Master Table

Physical Name	Data Type	Req'd	PK	Notes
ProductCategory	CHAR(10)	☑	☐	Constant stating hip implant
Product	CHAR(10)	☑	☑	Product item identifier
ProductDescription	CHAR(10)	☐	☐	Description of product

Notes: The Product Master table contains information about different products sold by Acme. All fields in this table should be dimensions.

Transactions Table

Physical Name	Data Type	Req'd	PK	Notes
TransType	CHAR(10)	☑	☑	Indicates the type of transaction (O for orders and S for sales)
TransNumber	CHAR(10)	☑	☑	Identifier for orders and sales
SiteCode	CHAR(10)	☑	☐	Transaction site (one of five locations)
Manager	DOUBLE	☑	☐	Employee associated with the transaction
EntityID	CHAR(10)	☑	☐	Supplier/Customer associated with the transaction
Quantity	DOUBLE	☐	☐	Transaction quantity
Unit	CHAR(10)	☐	☐	Unit of transaction (lb or kg for orders, ea for sales)
UnitPrice	CURRENCY	☐	☐	Price per unit
TransDate	DATETIME	☐	☐	Sales/purchase order date
CompleteDate	DATETIME	☐	☐	Shipping/receiving date (sales/orders)
Item	CHAR(10)	☐	☐	The material purchased or product sold
CreatedBy	CHAR(10)	☐	☐	The source of the data (e.g., daily batch)
Subtotal	CURRENCY	☐	☐	Transaction total before tax (UnitPrice * Quantity)
Tax	CURRENCY	☐	☐	Tax amount

Notes: The Transactions table contains a combination of purchase order, sales order, and sales return transactions (each row is either a purchase, a sale, or a return). Quantity, UnitPrice, Subtotal, and Tax should be measures; all other fields should be dimensions. To correctly aggregate sales, orders, and returns data, it is important keep in mind whether a transaction is an order, a sale, or a return. To avoid having to constantly think about these distinctions, it is helpful to create calculated columns in Tableau for sales, sales net of returns, returns, and purchases subtotals and quantities (and perhaps unit prices and taxes). To calculate these fields, *TransType* can be used to distinguish between sales (TransType = "S") and purchase order (TransType = "O") transactions. Sales returns are identified as transactions with negative values in Quantity, Subtotal, and Tax. For example, four columns can be used to

[7] The data dictionary contains metadata (data about the data) that are useful for understanding the content of each table. For example, the data dictionary shows that the Material Master table contains three columns (MaterialCategory, Material, and MaterialDescription) and that the column Material contains an identifier. Understanding the content of the data is not only important when analyzing the data but also during the data transformation stage (e.g., to create calculated fields and to determine which fields should be dimensions vs. measures).

reference the subtotal column: (1) *Sales* (Gross Sales subtotal) defined as IF TransType = "S" and Subtotal > 0 THEN Subtotal, (2) *Returns* (Returns subtotal) defined as IF Subtotal < 0 THEN Subtotal, (3) *Net sales* (sales subtotal net of returns subtotal) defined as IF TransType = "S" THEN Subtotal, and (4) *Purchases* (Purchases subtotal) defined as IF TransType = "O" THEN Subtotal). To create a new calculated measure, right click the TR Subtotal measure, select Create → Calculated field… (or select Create calculated field when in the data source view), and enter the calculation. For example, to calculate the field *Sales Subtotal* (which is all sales, including sales that were later returned), add the following formula (excluding the period): IF [TR Subtotal]>0 and [TR TransType]= "S" THEN [TR Subtotal] END.

Additional fields descriptions:
UnitPrice is the sales price of product for sales transactions and the purchase price of raw materials for purchase transactions. For a given sales transaction, UnitPrice is the price per unit paid by the customer for the particular product sold. For a given purchase transaction, UnitPrice is the price per unit of weight (one lb or one kg) of the purchased raw material that Acme paid the supplier.

TransDate indicates the date the transaction was recorded. For sales order transactions, TransDate represents the date the sales order transaction was recorded. For purchase order transactions, TransDate represents the date the purchase order transaction was recorded.

CompleteDate indicates the date the transaction was completed. For sales order transactions, CompleteDate represents the shipping date. For purchase order transactions, CompleteDate represents the receiving date.

Item is an identifier that indicates which product was sold (for sales transactions) or which raw material was purchased (for purchase transactions).

Sites Table

Physical Name	Data Type	Req'd	PK	Notes
SiteCode	CHAR(10)	☑	☑	Identified for company sites (five state abbreviations)
SiteName	CHAR(10)	☐	☐	Name of site (states)

Notes: The Sites table contains the name of each site, which is the name of the state where the site is located. This table provides a mapping of state acronyms to state names. All fields in this table should be dimensions.

Employee Master Table

Physical Name	Data Type	Req'd	PK	Notes
EmployeeID	INTEGER	☑	☑	Employee identifier
FirstName	CHAR(10)	☐	☐	Employee first name
LastName	CHAR(10)	☐	☐	Employee last name
SiteCode	CHAR(10)	☐	☐	Company site identified (state)
Position	CHAR(10)	☐	☐	Employee position
Order Approval Au...	CURRENCY	☐	☐	Employee order approval authority

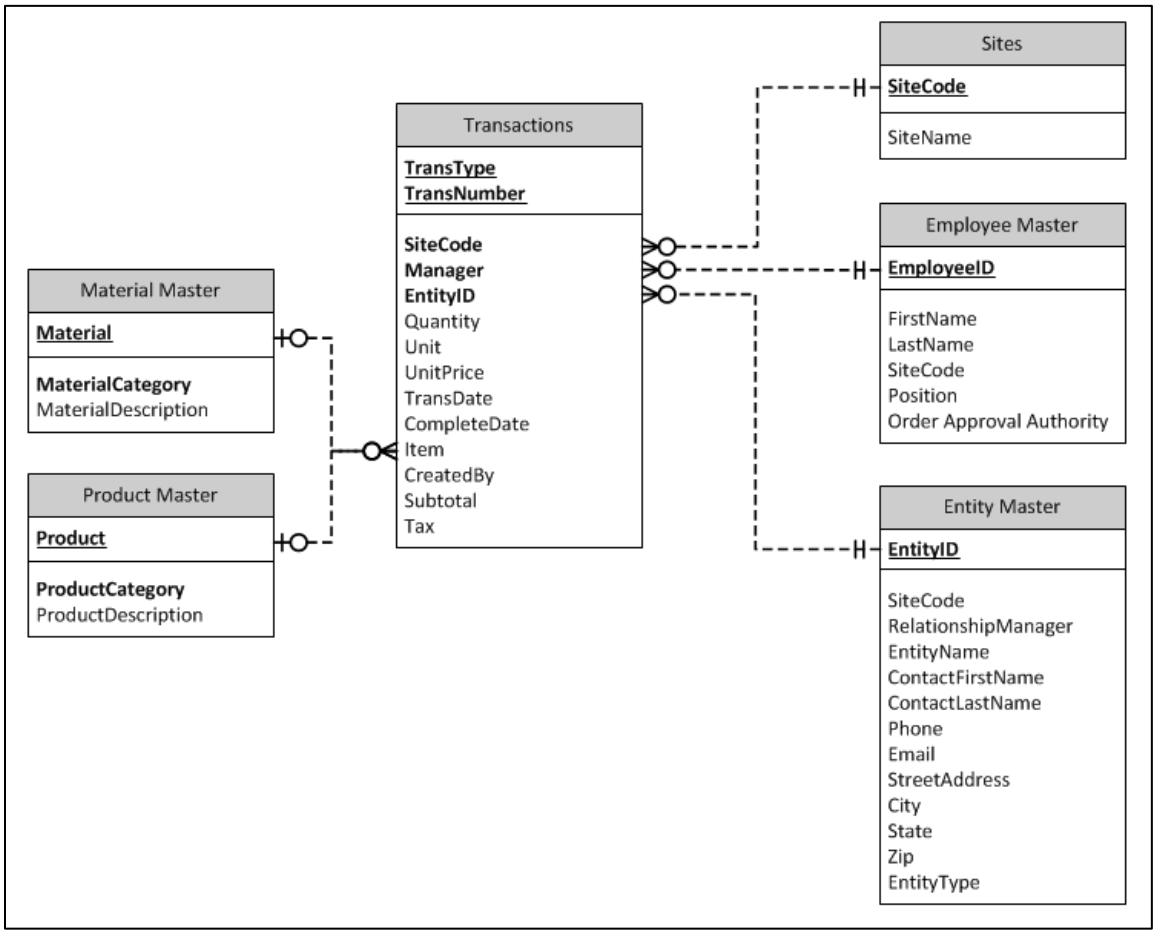
Notes: The Employee Master table contains information about employees associated with the sales order and purchase order transactions, e.g., sales and product managers.

Entity Master Table

Physical Name	Data Type	Req'd	PK	Notes
EntityID	CHAR(10)	☑	☑	Entity (supplier/customer/Acme Co.) identifier
SiteCode	CHAR(10)	☐	☐	Indicates which ACME site the entity is associated with
RelationshipManager	CHAR(10)	☐	☐	Identifies which purchasing manager manages the relationship with this supplier (null if customer)
EntityName	CHAR(10)	☐	☐	Entity company name
ContactFirstName	CHAR(10)	☐	☐	Entity contact first name
ContactLastName	CHAR(10)	☐	☐	Entity contact last name
Phone	CHAR(10)	☐	☐	Entity phone
Email	CHAR(10)	☐	☐	Entity email
StreetAddress	CHAR(10)	☐	☐	Entity street address
City	CHAR(10)	☐	☐	Entity city
State	CHAR(10)	☐	☐	Entity state
Zip	CHAR(10)	☐	☐	Entity zip code
EntityType	CHAR(10)	☐	☐	Entity type ("S" for customer and "O" for supplier)

Notes: The Entity Master table contains information about customers and suppliers, i.e., entities associated with the sales order and purchase order transactions, respectively. The *RelationshipManager* field identifies the Acme product manager associated with each supplier. Customer accounts do not have assigned sales managers and this field is blank for all customer entities (EntityType "S"). All fields in this table should be dimensions.

EXHIBIT 6
Entity-Relationship-Diagram (ERD)



Notes: The ERD diagram indicates which tables to join and which columns to use in the joins. For example, the Transaction table should be joined with the Employee table and the join should use the Manager field from the Transactions table and the EmployeeID field from the Employee Master table. The ERD diagram also indicates what types of joins to use. The ERD diagram shows that all Transactions must have at least one corresponding record in the Sites, Employee Master, and Entity Master tables. As such, *inner joins* should be used between Transactions and Sites, Transactions and Employee Master, and Transactions and Entity Master.[8] The ERD diagram also indicates that some records in Transactions may not have a corresponding record in the Material Master or Product Master tables.[9] To ensure that all transactions are available for analysis in Tableau, use *outer joins* between Transactions and Material Master and between Transactions and Product Master to include all transactions even if there is no matching record in the Material Master or Product Master tables.

[8] Given that the focus is on the Transactions table, the joined data only need to include rows from the other tables when matching records in the Transactions table are available. Since the ERD diagram additionally specifies that each transaction is associated with one and only one site, one and only one employee, and one and only one entity, these tables can be joined with the Transactions table using *inner joins*.

[9] Purchase order transactions have corresponding records in the Material Master table but not in the Product Master table. Sales order transactions have corresponding records in the Product Master table but not in the Material Master table.