In the professional world, turning massive amounts of data into something that can be seen and understood is vitally important. This is where Tableau steps in. It has emerged as a clear leader in data visualization because it translates your actions into a database query and expresses the response graphically.

It also has the unique ability to do ad hoc analysis of millions of rows of data in just a matter of seconds with Tableau’s Data Engine. Tableau is a rare software platform that is intuitive and even fun to use, which also enables you to dive deep into answering complex questions about your data.

Starting with creating your first dashboard in Tableau 9.0, this book will let you in on some useful tips and tricks, teach you to tell data stories using dashboards, and teach you how to share these data stories. Practical examples along with detailed explanations of how and why various techniques work will help you learn and master Tableau quickly.

Who this book is written for
If you want to understand your data using data visualization and don’t know where to start, then this is the book for you. Whether you are a beginner or have years of experience, this book will help you to quickly acquire the skills and techniques used to discover, analyze, and communicate data visually. Some familiarity with databases and data structures is helpful, but not required.

What you will learn from this book
- Explore and analyze your data by creating basic and advanced data visualizations
- Fix data problems, enhance your analysis, and create rich interactivity using custom calculations
- Perform effective analysis by joining and blending data from different sources
- Enhance your visualizations with custom formatting, labels, and annotations
- Explore advanced topics such as sheet swapping, custom maps, and LOD calculations
- Create meaningful dashboards in Tableau
- Extend the value and functionality of your data
- Share your data story using story points and fully interactive dashboards
In this package, you will find:

- The author biography
- A preview chapter from the book, Chapter 8 'Adding Value to Analysis Trends, Distributions, and Forecasting'
- A synopsis of the book’s content
- More information on Learning Tableau

About the Author

Joshua N. Milligan has been a consultant with Teknion Data Solutions since 2004, where he currently serves as a team leader and project manager. With a strong background in software development and custom .NET solutions, he uses a blend of analytical and creative thinking in BI solutions, data visualization, and data storytelling. His years of consulting have given him hands-on experience in all aspects of the BI development cycle, including data modeling, ETL, enterprise deployment, data visualization, and dashboard design. He has worked with clients in numerous industries, including financial, healthcare, marketing, and government.

In 2014, Joshua was named a Tableau Zen Master, the highest recognition of excellence from Tableau Software. As a Tableau-accredited trainer, mentor, and leader in the online Tableau community, he is passionate about helping others gain insights into their data. He was a technical reviewer of Tableau Data Visualization Cookbook, Packt Publishing, and is currently reviewing Creating Data Stories with Tableau Public, Packt Publishing. His work has appeared multiple times on Tableau Public’s Viz of the Day and Tableau’s website. Joshua also shares frequent Tableau tips, tricks, and advice on his blog at www.VizPainter.com
Learning Tableau

The Tableau community is full of individuals passionate about the software. We use software every day—web browsers, word processors, e-mail applications, instant messaging, and numerous other apps. What is it about Tableau that inspires people to write books and blogs and spend hours volunteering to help others visualize their data?

Tableau is unique in several ways. It is easy and transparent. You can immediately connect to nearly any data source and start asking and answering questions about your data in a visual way. It's also intuitive. Its interface allows hands-on interaction with data, it's easy to get into a flow, and every action uncovers new insights. It's fun! It allows creativity and gives freedom. You're not locked into chart types and wizards that give only one path to a solution. Tableau designers feel like artists, with data as paint and Tableau as a blank canvas.

At the same time, Tableau introduces a paradigm vastly different from traditional BI tools. This book presents the fundamentals for understanding and working within that paradigm. It will equip you with the foundational concepts that will help you use Tableau to explore, analyze, visualize, and share the stories contained in your data.

What This Book Covers

Chapter 1, Creating Your First Visualizations and Dashboard, introduces the basic concepts of data visualization and multiple examples of individual visualizations, which are ultimately put together in an interactive dashboard.

Chapter 2, Working with Data in Tableau, shows that Tableau has a very distinctive paradigm for working with data. This chapter explores that paradigm and gives examples of connecting to and working with various data sources.

Chapter 3, Moving from Foundational to Advanced Visualizations, expands upon the basic concepts of data visualization to show you how standard visualization types can be extended.

Chapter 4, Using Row-level and Aggregate Calculations, introduces the concepts of calculated fields and the practical use of calculations, and walks through the foundational concepts for creating row-level and aggregate calculations.

Chapter 5, Table Calculations, proves that table calculations are one of the most complex and powerful features in Tableau. This chapter breaks down the basics of scope, direction, partitioning, and addressing to help you understand and use these to solve practical problems.
Chapter 6, *Formatting a Visualization to Look Great and Work Well*, shows how formatting can make a standard visualization look great, have appeal, and communicate well. This chapter introduces and explains the concept of formatting in Tableau.

Chapter 7, *Telling a Data Story with Dashboards*, dives into the details of building dashboards and telling stories with data. It covers the types of dashboards, objectives of dashboards, and concepts such as actions and filters. All of this is done in the context of practical examples.

Chapter 8, *Adding Value to Analysis – Trends, Distributions, and Forecasting*, explores the analytical capabilities of Tableau and demonstrates how to use trend lines, distributions, and forecasting to dive deeper into the analysis of your data.

Chapter 9, *Making Data Work for You*, explains that data in the real world isn't always structured well. This chapter examines the structures that work best and the techniques that can be used to address data that can't be fixed.

Chapter 10, *Advanced Techniques, Tips, and Tricks*, builds upon the concepts covered in the previous chapters. This chapter expands your horizons by introducing numerous advanced techniques while giving practical advice and tips.

Chapter 11, *Sharing Your Data Story*, throws light on the fact that once you've built your visualizations and dashboards, you'll want to share them. This chapter explores numerous ways of sharing your stories with others.
Adding Value to Analysis – Trends, Distributions, and Forecasting

Sometimes, quick data visualization needs a little deeper analysis. For example, a simple scatterplot can reveal outliers and clusters of values. However, often, you want to understand the distribution. A simple time series helps you see the rise and fall of a measure over time. But many times, you want to see the trend or make predictions of future values.

Tableau enables you to quickly enhance your data visualizations with statistical analysis. Built-in features, such as trend models, distributions, and forecasting, allow you to quickly add value to your visual analysis. Additionally, Tableau integrates with R, an extensive statistical platform that opens up endless options for the statistical analysis of your data. This chapter will cover built-in statistical models and analysis.

This chapter will cover the following topics:

- Trending
- Forecasting
- Distributions

We'll take a look at these concepts in the context of a few examples using some sample datasets. You can follow and reproduce these examples using this chapter's workbook.
Trends

Let’s say you are analyzing populations of various countries using the World Population dataset in the provided workbook. This dataset produces one record containing the population for each country for each year from 1960 to 2013. Create a view similar to the one shown in the following screenshot, which shows you the change in population over time for Afghanistan and Australia. You’ll notice that Country Name has been filtered and added to the Color and Label shelves.

From this visualization alone, you can make several interesting observations. The growth of the two countries’ populations was fairly similar up to 1980. At that point, the population of Afghanistan went into decline until 1988 when the population of Afghanistan started to recover. At some point around 1996, the population of Afghanistan exceeded that of Australia. The gap has grown wider ever since.
While we have a sense of the two trends, they become even more obvious when we see them. Tableau offers several ways to add trend lines:

- From the menu, navigate to Analysis | Trend Lines | Show Trend Lines
- Right-click on an empty area in the pane of the view and select Show Trend Lines
- Switch to the Analytics tab in the left-hand side pane and drag and drop Trend Line on the trend model of your choice (we’ll use Linear for now and discuss the others later in this chapter)

Once you have added the trend line, your view should look like this:
Trends are calculated by Tableau after querying the data source. Trend lines are drawn based on various elements in the view:

- **The two fields that define x and y coordinates**: The last (right-most) field on Rows and Columns will define the axes that give Tableau x and y coordinates to calculate various trend models. In order to show trend lines, you must use a continuous (green) field or discrete (blue) date fields and have one such field on both Rows and Columns. If you use a discrete (blue) date field to define headers, the other field must be continuous (green).

- **Additional fields that create multiple, distinct trend lines**: Discrete (blue) fields on the Rows, Columns, or Color shelves can be used as factors to split a single trend line into multiple, distinct trend lines.

- **The trend model selected**: We'll examine the differences in models in the next section.

Observe in the view that there are two trend lines. As Country Name is a discrete (blue) field on Color, it defines a trend line per color by default.

Earlier, we observed that the population of Afghanistan increased and decreased within various historical periods. Notice that the trend lines are calculated along the entire date range. What if we want to see different trend lines for these time periods? We can force Tableau to draw distinct trend lines using a discrete field on Rows, Columns, or Color.

Go ahead and create a calculated field called Period that defines discrete values for the different historical periods and using code like this:

```plaintext
IF Year([Year]) <= 1979
    THEN "1960 to 1979"
ELSEIF Year([Year]) <= 1988
    THEN "1980 to 1988"
ELSE "1988 to 2013"
END
```
When you place **Period** on columns, you'll get a header for each time period, which breaks the lines and causes separate trends to be shown for each time period. You'll also observe that Tableau keeps the full date range in the axis for each period. You can set an independent range by right-clicking on one of the date axes, selecting **Edit Axis**, and then checking the option for **Independent axis range for each row or column**.

In this view, transparency has been applied to **Color** and the trend lines have been formatted to make them stand out. Additionally, the axis for **Year** was hidden (by unchecking the **Show Header** option on the field). Now, you can clearly see the difference in trends for different periods of time. Australia's trends only change slightly in each period. Afghanistan's trends were quite different.
Customizing trend lines

Let's take a look at another example that will allow us to consider various options for trend lines. Create a new sheet and use the Real Estate data source connection to create a view similar to this one:

Here, we created a scatterplot with the sum of Area on Columns to define the x axis and the sum of Price on Rows to define the y axis. Address has been added to the level of detail on the Marks card to define the level of aggregation. So, each mark on the scatterplot is a distinct address at a location defined by the area and price. Type has been added to Color. We've also shown the trend lines and are getting one trend line per color by default. Assuming a good model, the trend lines demonstrate how much and how quickly Price is expected to rise with an increase in Area.
In this dataset, we have two fields, Address and ID, either of which define a unique record. Adding one of these fields to the level of detail effectively disaggregates the data and allows us to plot a mark for each address. Sometimes, you may not have a field in the data that defines uniqueness. In these cases, you can disaggregate the data by unchecking Aggregate Measures from the Analysis menu.

Alternately, you can use the drop-down menu on each of the measure fields on Rows and Columns to change them from measures to dimensions while keeping them continuous. As dimensions, each individual value will define a mark. Keeping them continuous will retain the axes required for trend lines.

Let's consider some of the options available for trend lines. You can edit trend lines by using the menu and navigating to Analysis | Trend Lines | Edit Trend Lines or clicking/right-clicking on a trend line and then selecting Edit. When you do this, you'll see a dialog box similar to this:

Here, you have options to select a model type, including applicable fields as factors in the model, allowing discrete colors to define distinct trend lines, showing confidence bands, and forcing the $y$ intercept to zero. Experiment with the options for a bit. Notice how either removing the Type field as a factor or unchecking the Allow a trend line per color option results in a single trend line.
You can also see the result of excluding a field as a factor in the following view, where **Type** has been added to **Rows**:

As represented in the left portion of the preceding screenshot, **Type** is included as a factor. This results in a distinct trend line for each type of sale. When **Type** is excluded as a factor of the same trend line, which is the overall trend for all types, a trend line is drawn three times. This technique can be quite useful to compare subsets of data with the overall trend.
Trend models
We'll return to the original view and stick with a single trend line as we consider the trend models available. The following models can be selected from the Trend Line Options window:

- **Linear**: We'll use this model if we assume that as Area increases, Price will increase at a constant rate. No matter how high Area increases, we'll expect Price to increase such that new data points fall close to the straight line.

![Linear Trend Model](image1)

- **Logarithmic**: We'll use this model if we believe that there is a "law of diminishing returns" in effect. That is, area can only increase to a certain extent before buyers stop paying much more:

![Logarithmic Trend Model](image2)
• **Exponential**: We’ll use this model to test the idea that each additional increase in area results in a dramatic (exponential) increase in price:

![Exponential Model Diagram]

• **Polynomial**: We’ll use this model if we feel the relationship between Area and Price is complex and follows more of an S-shaped curve, where, though initially increasing the area dramatically increases the price, at some point the price levels. You can set the degree of the polynomial model anywhere from 2 to 8. The trend line shown here is a third-degree polynomial:

![Polynomial Model Diagram]
Analyzing trend models

It can be useful to observe trend lines, but often, we'll want to understand whether the trend model we've selected is statistically meaningful. Fortunately, Tableau gives us some visibility into trend models and calculations.

Simply hovering over a single trend line will reveal the calculation as well as P-value for that trend line.

A p-value is a statistical concept that describes the probability that the results of assuming no relationship between values (random chance) are at least as close as results predicted by the trend model. A p-value of 5 percent (.05) will indicate a 5 percent random chance describing the relationship between values as well as the trend model. This is why p-values of 5 percent or less are considered to indicate a significant trend model. If your p-value is higher than 5 percent, then you should not consider that trend to significantly describe any correlation.
Additionally, you can see a much more detailed description of the trend model by navigating to **Analysis | Trend Lines | Describe Trend Model...** from the menu or using the similar menu from a right-click on the view's pane. When you view the trend model, you will see the **Describe Trend Model** window:

![Describe Trend Model window](image)

You can also get a trend model description in the worksheet description, which is available from the **Worksheet** menu or by pressing **Ctrl + E**. The worksheet description includes quite a bit of other useful summary information about the current view.
The wealth of statistical information shown in the window includes a description of the trend model, the formula, the number of observations, and the p-value for the model as a whole as well as for each trend line. Note that in the window shown in the preceding screenshot, the **Type** field was included as a factor that defined three trend lines. At times, you may observe that the model as a whole is statistically significant even though one or more trend lines may not be.

Additional summary statistical information can be displayed in Tableau Desktop for a given view by showing the summary. From the menu, navigate to **Worksheet | Show Summary**. The information displayed in the summary can be expanded using the drop-down menu on the **Summary** card.

Tableau also gives you the ability to export data, including data related to trend models. This allows you to more deeply, and even visually, analyze the trend model itself. Let's analyze the third-degree polynomial trend line of the real estate price and area scatterplot without any factors. To export data related to the current view, use the menu and navigate to **Worksheet | Export | Data**. The data will be exported as a Microsoft Access Database (\*.mdb) and you will be prompted as to where to save the file.
Adding Value to Analysis – Trends, Distributions, and Forecasting

On the Export Data to Access screen, specify an access table name and select whether you wish to export data from the entire view or the current selection. You may also specify that Tableau should connect to the data. This will generate the data connection and make it available with the specified name in the current workbook.

The new data source connection will contain all the fields that were present in the original view as well as additional fields related to the trend model. This allows us to build a view such as the following using residuals and predictions:
A scatterplot of predictions and residuals allows you visually see how far each mark was from the location predicted by the trend line. It also allows you to see whether residuals are distributed evenly on either side of a zero. An uneven distribution would indicate problems with the trend model.

You can include this new view along with the original one in a dashboard to explore the trend model visually. Use the highlight button on the toolbar to highlight the **Address** field.

With the highlight action defined, selecting marks in one view will allow you to see them in the other. You could extend this technique to export multiple trend models and dashboards to evaluate several trend models at the same time, as shown in the following screenshot:

You can achieve even more sophisticated statistical analysis, leveraging Tableau’s ability to integrate with R. R is an open source statistical analysis platform and a programming language with which you can define advanced statistical models. R functions can be called from Tableau using special table calculations (all of which start with `SCRIPT_`). These functions allow you to pass expressions and values to a running R Server that will evaluate the expressions using built-in libraries or custom-written R scripts and return results to Tableau.

You can learn more about Tableau and R integration from this whitepaper (you will need to register a free account first): [http://www.tableausoftware.com/learn/whitepapers/using-r-and-tableau](http://www.tableausoftware.com/learn/whitepapers/using-r-and-tableau)
Distributions

Analyzing distributions can be quite useful. We've already seen that certain table calculations are available to determine statistical information such as averages, percentiles, and standard deviations. Tableau also makes it easy to quickly visualize various distributions including confidence intervals, percentages, percentiles, quantiles, and standard deviations.

You may add any of these visual analytic features using the Analytics tab (alternately, you can right-click on an axis and select Add Reference Line). Just like reference lines and bands, distribution analytics can be applied within the scope of a table, pane, or cell. When you drag and drop the desired visual analytic, you'll have options to select the scope and the axis. In the following example, we've dragged and dropped Distribution Band from the Analytics tab onto the scope of Pane for the axis defined by Sum(Price):

Once you have selected the scope and axis, you will be given options to change settings. You may also edit lines, bands, distributions, and box plots by right-clicking on the analytic feature in the view or by right-clicking on the axis. Here, we'll define settings for one and two standard deviations above and below the mean:
Each specific **Distribution** option specified in the **Value** dropdown under **Computation** has unique settings. **Confidence Interval**, for example, allows you to specify a percent value for the interval. **Standard Deviation** allows you to enter a comma-delimited list of values that describe how many standard deviations are used, and at what intervals. This, for example, is the result of specifying standard deviations of -2, -1, 1, 2:
Adding Value to Analysis – Trends, Distributions, and Forecasting

Each axis can support multiple distributions, reference lines, and bands. Here, first and second standard deviations on both sides of the average (the solid line) are shown. You'll notice that the Type field defines three panes and the standard deviations have been set to be calculated per pane.

On a scatterplot, using a distribution for each axis can yield a very useful way to analyze outliers. Showing a single standard deviation for both Area and Price allows you to easily see properties that fall within norms for both, one, or neither.

Forecasting

As we've seen, trend models make predictions. Given a good model, you expect additional data to follow the trend. When the trend is over time, you can get an idea about where future values may fall. However, predicting future values often requires a different type of model. Factors such as seasonality can make a difference not predicted by a trend alone. Starting with version 8.0, Tableau includes built-in forecasting models that can be used to predict and visualize future values.

To use forecasting, you'll need a view that includes a date field or enough date parts for Tableau to reconstruct a date (for example, a Year and a Month field). You may drag and drop Forecast from the Analytics tab, navigate to Analysis | Forecast | Show Forecast from the menu, or right-click on the view’s pane and select the option from the context menu.
Here, for example, is the view of the population growth of Afghanistan and Australia with forecasts shown over time:

Note that when you show the forecast, Tableau adds a forecast icon to the \texttt{SUM(Population)} field on \texttt{Rows} to indicate that the measure is being forecast. Additionally, Tableau adds a new special \texttt{Forecast Indicator} field to \texttt{Color} so that forecast values are differentiated from actual values in the view.

You can move the \texttt{Forecast Indicator} field or even copy it (hold \texttt{Ctrl} while dragging and dropping) to other shelves to further customize your view.
When you edit the forecast by navigating to **Analysis > Forecast > Forecast Options...** from the menu or when you use the right-click context menu on the view, you will be presented with various options to customize the trend model, like this:

Here, you have options to set the length of the forecast, determine aggregations, customize the model, and set whether you wish to show prediction intervals. The forecast length is set to **Auto** by default, but you can extend the forecast by a custom value.

The options under **Source Data** allow you to optionally specify a different grain of data for the model. For example, your view might show a measure by year but you could allow Tableau to query the source data to retrieve values by month and use a finer grain to potentially achieve better results.
Tableau’s ability to separately query the data source to obtain data at a finer grain for more precise results works well with relational data sources. However, OLAP data sources are not compatible with this approach, which is one reason forecasting is not available when working with cubes.

By default, the last value is excluded from the model. This is useful when you are working with data where the most recent time period is incomplete. For example, when records are added daily, the last (current) month is not complete until the final records are added on the last day of the month. Prior to this last day, the incomplete time period might skew the model unless it is ignored.

The model itself can be set to **Automatic** with or without seasonality or can be customized to set options for seasonality and trend. To understand the options, consider the following view of **Sales by MONTH** from the Superstore sample data:
The data displays a distinct cyclical or seasonal pattern. This is very typical for retail sales. The following are the results of selecting various custom options:

<table>
<thead>
<tr>
<th>No Trend</th>
<th>Additive Seasonality</th>
<th>Multiplicative Seasonality</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
<td><img src="image3.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additive Trend</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.png" alt="Graph" /></td>
<td><img src="image5.png" alt="Graph" /></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiplicative Trend</th>
<th>Not Allowed</th>
<th>Numerically Unstable</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image6.png" alt="Graph" /></td>
<td><img src="image7.png" alt="Graph" /></td>
<td></td>
</tr>
</tbody>
</table>

Much like trends, forecast models and summary information can be accessed using the menu. Navigating to **Analysis | Forecast | Describe Forecast** will display a window with tabs for both the summary and details concerning the model.
Clicking on the link at the bottom of the window will give you much more information on the forecast models used in Tableau.

Forecast models are only enabled with a certain set of conditions. If the option is disabled, ensure that you are connected to a relational database and not OLAP, are not using table calculations, and have at least five data points.

Summary
Tableau provides an extensive set of features to add value to your analysis. Trend lines allow you to more precisely identify outliers, determine which values fall within the predictions of certain models, and even make predictions of where measurements are expected. Tableau gives you extensive visibility into the trend models and even allows you to export data containing trend model predictions and residuals. Distributions are useful to understand how measurements are distributed. Forecasting allows a complex model of trends and seasonality to predict future results. Having a good understanding of these tools will give you the ability to clarify and validate your initial visual analyses.

Next, we'll turn our attention back to the data. We considered very early on how to connect to data, and we've been working with data ever since. However, we've spent most of our time working with clean, well-structured data. In the next chapter, we'll consider how to deal with messy data.
Where to buy this book
You can buy Learning Tableau from the Packt Publishing website.
Alternatively, you can buy the book from Amazon, BN.com, Computer Manuals and most internet book retailers.
Click here for ordering and shipping details.