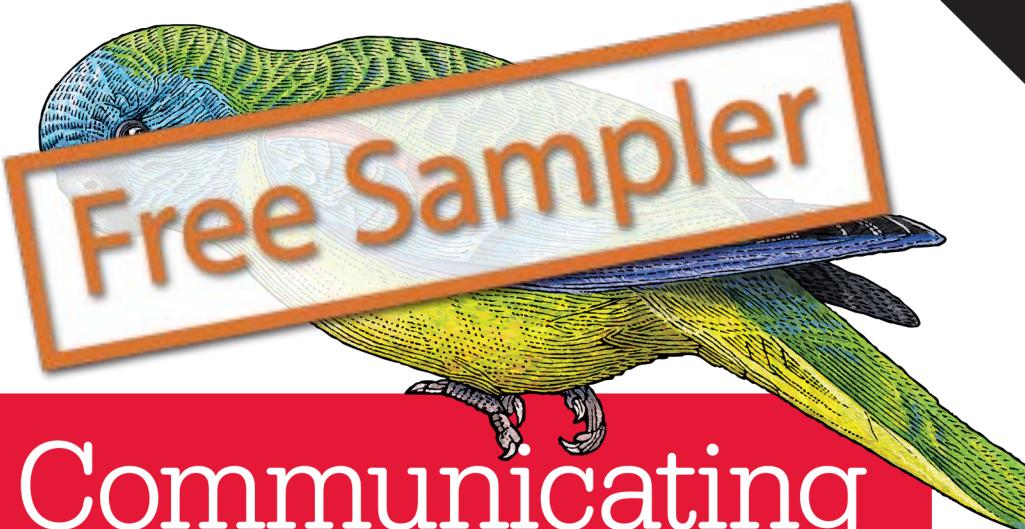


O'REILLY®

Covers  
Tableau version 8.1



Free Sampler

# Communicating Data with Tableau

---

DESIGNING, DEVELOPING, AND DELIVERING  
DATA VISUALIZATIONS

Ben Jones

# Communicating Data with Tableau

Go beyond spreadsheets and tables and design a data presentation that really makes an impact. This practical guide shows you how to use Tableau Software to convert raw data into compelling data visualizations that provide insight or allow viewers to explore the data for themselves.

Ideal for analysts, engineers, marketers, journalists, and researchers, this book describes the principles of communicating data and takes you on an in-depth tour of common visualization methods. You'll learn how to craft articulate and creative data visualizations with Tableau Desktop 8.1 and Tableau Public 8.1.

- Present comparisons of how much and how many
- Use blended data sources to create ratios and rates
- Create charts to depict proportions and percentages
- Visualize measures of mean, median, and mode
- Learn how to deal with variation and uncertainty
- Communicate multiple quantities in the same view
- Show how quantities and events change over time
- Use maps to communicate positional data
- Build dashboards to combine several visualizations

“Ben Jones has written that unusual thing: a reference book that entertains as well as educates. This book is a pleasure.”

—Andy Cotgreave  
Social Content Manager  
at Tableau Software

---

**Ben Jones** is the Tableau Public product marketing manager at Tableau Software in Seattle, and an award-winning Tableau Public author. Previously, he worked on data-driven projects for every department, from facilities to finance, as an engineer and business analytics manager for a Fortune 500 company.

---

DATA

US \$39.99

CAN \$41.99

ISBN: 978-1-449-37202-6



9 781449 372026



Twitter: @oreillymedia  
facebook.com/oreilly

# Want to read more?

You can [buy this book](#) at [oreilly.com](#)  
in print and ebook format.

**Buy 2 books, get the 3rd FREE!**

Use discount code: OPC10

All orders over \$29.95 qualify for **free shipping** within the US.

---

It's also available at your favorite book retailer,  
including the iBookstore, the [Android Marketplace](#),  
and [Amazon.com](#).

O'REILLY®

Spreading the knowledge of innovators



oreilly.com

## **Communicating Data with Tableau**

by Ben Jones

Copyright © 2014 Ben Jones. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://my.safaribooksonline.com>). For more information, contact our corporate/institutional sales department: 800-998-9938 or [corporate@oreilly.com](mailto:corporate@oreilly.com).

**Editor:** Julie Steele

**Indexer:** Lucie Haskins

**Production Editor:** Kristen Brown

**Cover Designer:** Karen Montgomery

**Copieditor:** Jasmine Kwityn

**Interior Designer:** David Futato

**Proofreader:** Eliahua Sussman

**Illustrator:** Rebecca Demarest

June 2014: First Edition

### **Revision History for the First Edition:**

2014-06-12: First release

See <http://oreilly.com/catalog/errata.csp?isbn=9781449372026> for release details.

Nutshell Handbook, the Nutshell Handbook logo, and the O'Reilly logo are registered trademarks of O'Reilly Media, Inc. *Communicating Data with Tableau*, the image of a turquoise parrot, and related trade dress are trademarks of O'Reilly Media, Inc.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O'Reilly Media, Inc. was aware of a trademark claim, the designations have been printed in caps or initial caps.

While every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

ISBN: 978-1-449-37202-6

[LSI]

---

# Table of Contents

Preface.....	ix
<b>1. Communicating Data.....</b>	<b>1</b>
A Step in the Process	2
A Model of Communication	3
Three Types of Communication Problems	4
Six Principles of Communicating Data	5
Principle #1: Know Your Goal	6
Principle #2: Use the Right Data	7
Principle #3: Select Suitable Visualizations	8
Principle #4: Design for Aesthetics	9
Principle #5: Choose an Effective Medium and Channel	11
Principle #6: Check the Results	13
Summary	13
<b>2. Introduction to Tableau.....</b>	<b>15</b>
Using Tableau	15
My Tableau Story	16
Tableau Products	16
Connecting to Data	18
The Tableau User Interface	18
Summary	29
<b>3. How Much and How Many.....</b>	<b>31</b>
Communicating “How Much”	32
An Example of How Much	33
Comparing Comparisons	35

Fine-Tuning the Default	36
Sorting	37
The Dot Chart	39
Communicating “How Many”	42
A Tale of Two Formats	43
Counting Dimensions	43
Histograms: How Many of How Much?	46
Summary	50
<b>4. Ratios and Rates.....</b>	<b>51</b>
Ratios	52
Two Ways of Adding Rank	60
Rates	63
Blending Data Sources	64
Visualizing Rates	65
Summary	67
<b>5. Proportions and Percentages.....</b>	<b>69</b>
Part-to-Whole	69
Introducing Filters and Quick Filters	71
Introducing Table Calculations	74
Proportions as Waterfall Charts Using Gantt	79
Current-to-Historical	82
The Bullet Graph	82
Reference Lines	84
Actual-to-Target	85
Summary	86
<b>6. Mean and Median.....</b>	<b>87</b>
The Normal Distribution	88
An Example of “Normal” Data	89
Box Plots	90
An Example of “Non-Normal” Data	96
Sensitivity to Outliers	97
Visualizing Typical Values of Non-Normal Distributions	98
Summary	99
<b>7. Variation and Uncertainty.....</b>	<b>101</b>
Respecting Variation	101
Visualizing Variation	104

Variation Over Time: Control Charts	106
Anatomy of a Control Chart	107
How to Create a Control Chart in Tableau	107
Understanding Uncertainty	115
Summary	123
<b>8. Multiple Quantities.....</b>	<b>125</b>
Scatterplots	126
Who Is Who?	130
Making it Exploratory	134
Adding Background Images	136
Stacked Bars	137
Regression and Trend Lines	141
The Quadrant Chart	146
Summary	148
<b>9. Changes Over Time.....</b>	<b>149</b>
The Origin of Time Charts	150
The Line Chart	151
The Dual-Axis Line Chart	154
The Connected Scatterplot	158
The Date Field Type and Seasonality	162
The Timeline	166
The Slopegraph	171
Step 1: Get the Data	171
Step 2: Connect Tableau	172
Step 3: Create a Parameter and Matching Calculated Field	172
Step 4: Create the Basic Slopegraph	174
Step 5: Add Line Coloring and Thickness	175
Step 6: Design the Dashboard	178
Summary	179
<b>10. Maps and Location.....</b>	<b>181</b>
One Special Map	182
Circle Maps	183
Adding a Second Encoding	185
When Marks Multiply	186
Filled Maps	190
Dual-Encoded Maps	196
A Dual-Axis Map	197

A Dual-Encoded Circle Map	199
Summary	201
<b>11. Advanced Maps.....</b>	<b>203</b>
Maps with Shapes	204
Maps Showing Paths	211
Plotting Map Shapes Using Axes	216
Summary	223
<b>12. The Joy of Dashboards.....</b>	<b>225</b>
Dashboards in Tableau	226
A Word of Caution	228
“Begin with the End in Mind”	229
Types of Dashboards	230
Context Is King	235
Summary	238
<b>13. Building Dashboards.....</b>	<b>239</b>
Building an Exploratory Dashboard	243
Step 1: Design	243
Step 2: Sheets	243
Moving Things Around	248
Step 3: Annotations	250
Step 4: Objects	253
Step 5: Actions	258
Step 6: Formatting	270
Steps 7 and 8: Delivery and Results	271
Building an Explanatory Dashboard	272
A Key Point to Explain: Nordic Countries in the Lead	272
Another Key Point to Explain: The Emergence of China	274
Summary	275
<b>14. Advanced Dashboard Features.....</b>	<b>277</b>
Animating Dashboards	278
Showing Multiple Tabs	282
Adding Navigation with Filters	285
Adding Custom Header Images	290
Adding Google Maps to Dashboards	292
Create the URLs	294

Adding Dynamic Google Maps Satellite Images to Our Dashboard	295
Adding YouTube Videos to Dashboards	297
Summary	302
<b>A. Resources.....</b>	<b>303</b>
<b>Index.....</b>	<b>305</b>

## CHAPTER 4

# Ratios and Rates

*“We are the safest large city in America,  
but any crime rate is too high.”*

—Michael Bloomberg, former mayor  
of New York City

Until now, we've considered absolute amounts of countable and measurable nouns, like population, tons of garbage collected, and number of rat sightings by borough. Pleasant, I know.

But not all boroughs are created equal. As we saw in [Chapter 1](#), they have different surface areas and different numbers of people living in them. The fact that they produce different amounts of garbage isn't surprising, and probably doesn't make for a very interesting message to communicate to an audience.

These absolute comparisons, while helpful for showing the big picture, aren't generally considered “apples-to-apples” comparisons. There are apples, and, well, there are Big Apples.

A helpful way to “normalize” comparisons is using ratios, rates, proportions, and percentages. What's the difference between these four?

- A *ratio* is a comparison of two terms expressed as a quotient. For example, Manhattan produced 0.264 tons of recycle for every ton of refuse. Ratios can be expressed as “x to y,” “x:y,” “x/y,” or as a decimal.
- A *rate* is a ratio in which the two terms have different units. For example, the population density of Brooklyn is 36,136 residents per square mile. Rates are often predictive because time can be

used as the denominator (crime rates, population growth rates, etc.).

- A *proportion* is a ratio in which the numerator is a partial amount and the denominator is the total amount (expressed as a number between 0 and 1). For example, the proportion of the NYC population living in the Bronx is 0.169. A proportion is expressed as a number between 0 and 1.
- A *percentage* is a ratio comparing a number to 100. For example, 16.9% of NYC residents live in the Bronx. A percentage is generally a number between 0 and 100, but can be larger than 100 (e.g., “sales have increased by 150% year-over-year”).

These types of normalized comparisons can make for much more interesting messages to communicate. Luckily, they are quite easy to create in Tableau. We’ll explore the first two in this chapter, and the last two in the next.

## Ratios

The data source itself may include a field that is a ratio, but most often the raw data includes terms that an analyst needs to combine to create a ratio. Tableau can handle these types of computations with Calculated Fields, in which new fields can be created by combining existing fields using a variety of operations. We introduced the concept in [Chapter 2](#), and now we’ll build on it.

Let’s return to the DSNY data set to illustrate the technique, and we’ll see if the ratio comparison is any more interesting than the absolute figures.

The data reported by DSNY includes three continuous Measures (quantitative fields) for each community district:

`RefuseTonsCollected`

Amount of nonrecyclable garbage

`PaperTonsCollected`

Amount of recyclable paper

`MGPTonsCollected`

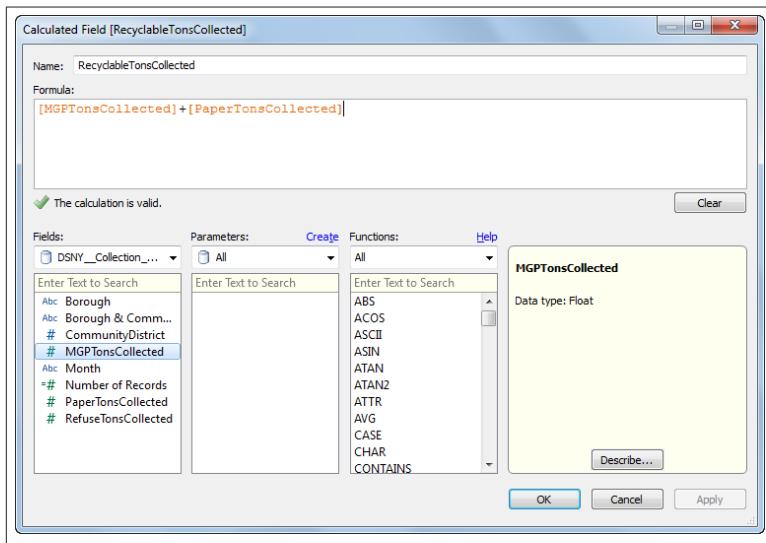
Amount of recyclable metal, glass, and plastic

The second two data types (paper and MGP) are both recyclable, and so can be summed to create a new variable for each community district

that we will call `RecyclableTonsCollected` to stay consistent with the provided nomenclature.

A comparison of the ratio of recyclable material to refuse would be interesting to consider, as it would indicate which communities in New York City are doing a better job recycling.

First, we'll need to create a new calculated field that sums the two recyclable fields. Right-click anywhere in the lefthand Dimensions or Measures panels, and select *Create Calculated Field*. Then, fill out the resulting dialog box, as shown in [Figure 4-1](#).



*Figure 4-1. Summing two fields with a Calculated Field*

Next, let's create a second calculated field that creates the ratio of recyclable material to refuse. Right-click in the Dimensions or Measures area again and fill out the dialog box as shown in [Figure 4-2](#).

Notice that we put the numerator and denominator in brackets, preceded by SUM. This is critical, because as we move up a level of from community district to borough, we don't want to *add* the individual ratios of each community district; we want to create a new ratio that represents the quotient of the aggregate amounts. The first approach would create a larger ratio that is misleading.

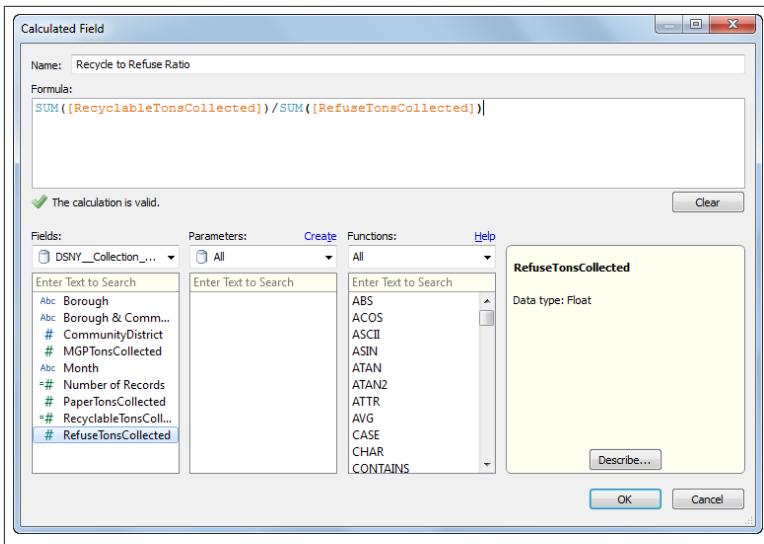


Figure 4-2. Creating a ratio with a Calculated Field

Now that we've created the ratios, how can we best visualize them? Let's start with a simple bar chart of the aggregate ratios of each of the boroughs. To create one, we'll drag the Borough field from the Dimensions area on the left to the Rows shelf, and we'll drag the Recycle to Refuse Ratio calculated field to the Columns shelf. Color wouldn't strictly be needed, but we'll add it anyway for aesthetics by dragging Borough to the Color shelf. Lastly, we'll sort the bars in descending order to put the borough with the highest recycle ratio at the top. The screen shot in [Figure 4-3](#) shows the resulting view.

We can easily see from this simple bar chart that Manhattan has the highest ratio of recycle to refuse, but that it still only recycles about one ton of material for every four tons of trash. The ratio for Manhattan is about twice as great as the ratio for the Bronx.

Does this mean each community district in Manhattan had a higher recycle ratio than each community district in the Bronx? Not necessarily, as we're dealing with the terms in the aggregate.

How would we communicate the results at a community district level? If we just replace Borough with CommunityDistrict in the Rows shelf, then we'll get a bar chart where all the community district #1s will be aggregated, and all the community district #2s will be aggregated, and so on. Aggregating community districts in this way isn't very

meaningful. Instead, we'd like to see community district ratios for each borough.

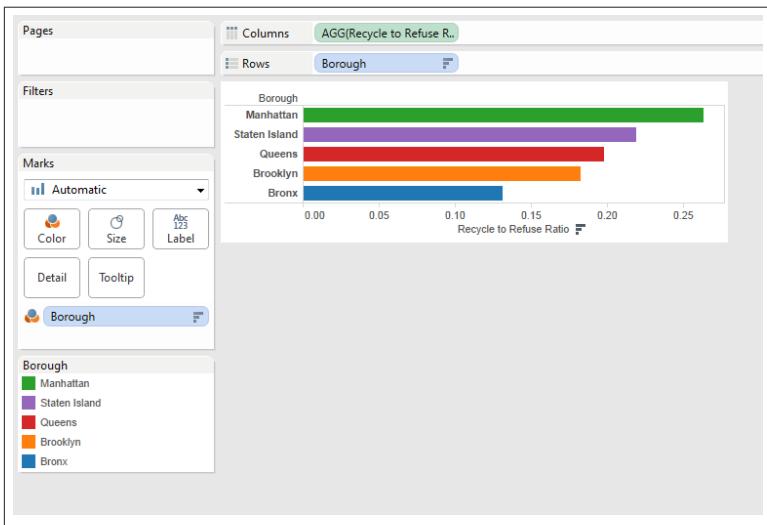


Figure 4-3. Visualizing the ratio of recycle to refuse for each borough

To make this comparison, let's click and drag the Borough pill from the Rows shelf up to the Columns shelf, and add the CommunityDistrict field from the Measures panel to the Rows shelf to create a grid of bar charts, as shown in [Figure 4-4](#).



Figure 4-4. Ratios by community district and borough

From this view, we can now see that even within boroughs, not all community districts are created equal. Manhattan community district

#1 has the highest ratio overall, but it's clear that some community districts in the Bronx have higher ratios than some in Manhattan.

But what if we wanted to know the exact ratios? The bar charts in [Figure 4-4](#) make it fairly easy to get a general sense of the relative ratios for the community districts, but precision isn't easily obtained, is it? Can you tell the exact ratio for Manhattan community district #1? I can't. If knowing the precise amounts is a task our audience will want to perform, we'll need to think about how to make it easy for them to accomplish that task.

We could increase precision by dragging the Recycle to Refuse Ratio from the Measures area to the Label shelf (or Ctrl-selecting AGG(Recycle to Refuse Ratio) and dragging it from the Columns shelf to the Label shelf to duplicate). The resulting bar chart grid with labels is shown in [Figure 4-5](#).



*Figure 4-5. Adding labels to the bar chart grid*

That's a lot of decimal places, isn't it? Our audience probably doesn't need to know the ratio to the fourth decimal place (ten-thousandths). We can change the number of digits shown by right-clicking the Recycle to Refuse Ratio in the Measures panel, selecting *Default Properties*, then *Number Format*, and choosing *Number (Custom)*. In the resulting dialog box, we'll change the decimal places to three, as shown in [Figure 4-6](#).

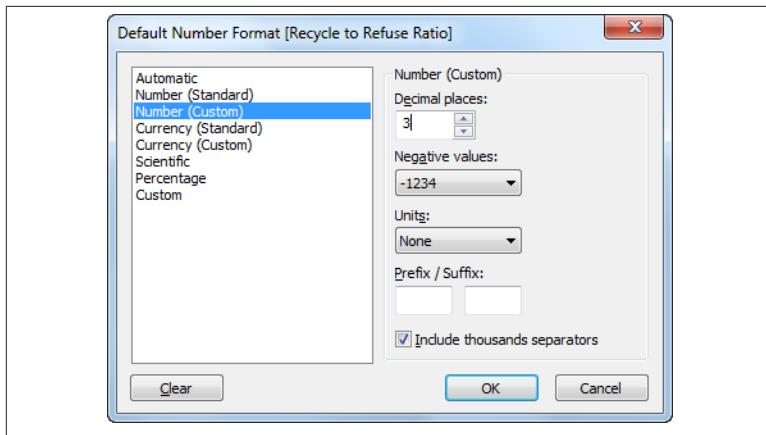


Figure 4-6. Specifying the number of decimal places in the Measures panel

Notice that the labels in the bar chart grid change accordingly.

What we really have here is a table that has been augmented by adding bars with lengths proportional to the values in each cell. Another way to show this precise comparison is by getting rid of the bars and instead coloring each cell by the ratio—a highlight table.

Let's make a highlight table by starting with a new sheet. Ctrl-click Borough, CommunityDistrict, and Recycle to Refuse Ratio, and then open the *Show Me* panel and select *highlight table*. Tableau creates the view shown in [Figure 4-7](#).

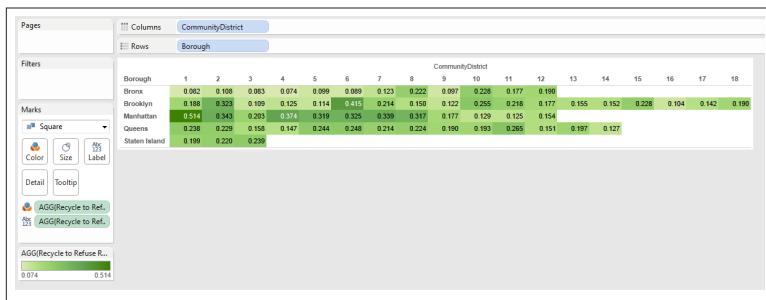
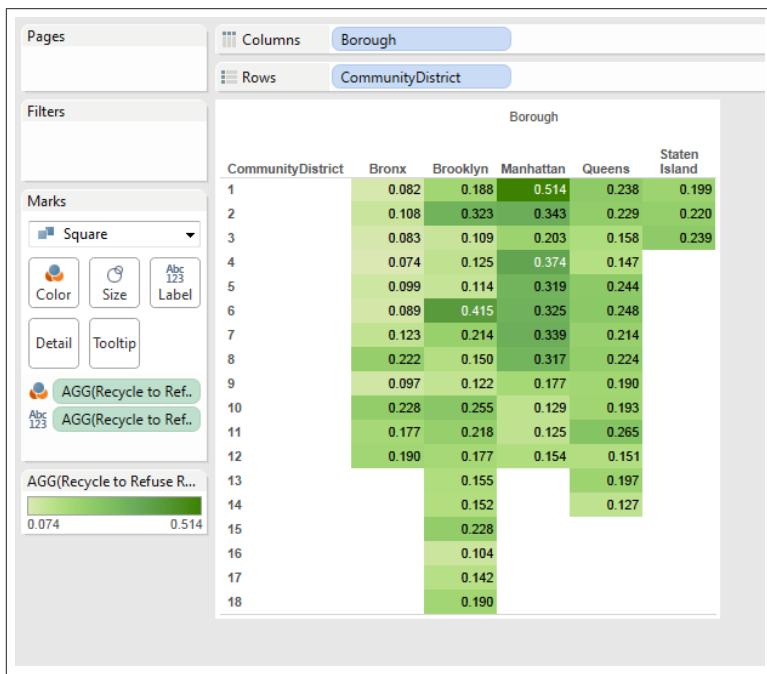


Figure 4-7. The default highlight table for Recycle to Refuse Ratio

Notice that Tableau created a highlight table that's “flipped” compared to our bar chart grid shown in [Figure 4-5](#) (Borough is in Rows instead

of Columns, and `CommunityDistrict` is in Columns instead of Rows). To flip it to match, we'll just click the *Swap* button in the toolbar, , resulting in the taller table shown in [Figure 4-8](#).



*Figure 4-8. The highlight table with Rows and Columns swapped*

Tableau has defaulted to a sequential green color palette, meaning the gradient increases proportionately with increasing `Recycle to Refuse Ratio`. Our eyes can easily pick out the darkest cell, and we also find the lightest rather easily. Most people can distinguish between about six different levels of intensity. If the goal were to communicate the precise values, we could have created a simple table without the highlights. Adding the highlights makes it easier to quickly compare cells and find the highest and lowest values.

It's still not immediately obvious which community district has the second-highest ratio, though. Our eyes have to dart around to all the dark cells and read and memorize the values. Eventually we settle on Brooklyn community district #6 as having the second-highest ratio, but the task gets harder as we move along. Which has the third highest? The fourth? Is there an easier way to facilitate this assessment of rank?

What we'd like is a simple list of community districts in descending order of the ratio. To make this list, let's first create a new field that combines the Borough and CommunityDistrict fields into a single Dimension.

Ctrl-click Borough and CommunityDistrict so they are both selected, then right-click on one of the two selected fields (either will do fine) and select *Combine Fields*. A new string field will appear in the Dimensions area called Borough & CommunityDistrict (Combined).

Create the list shown in **Figure 4-9** by dragging Borough & CommunityDistrict to the Rows shelf, Recycle to Refuse Ratio to the Columns shelf, and Borough to the Color shelf, sorting in descending order.

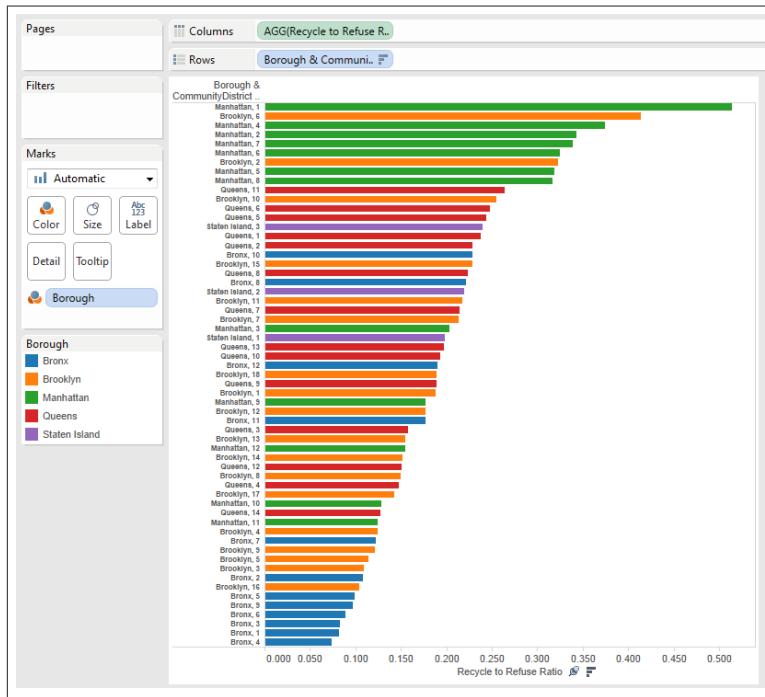


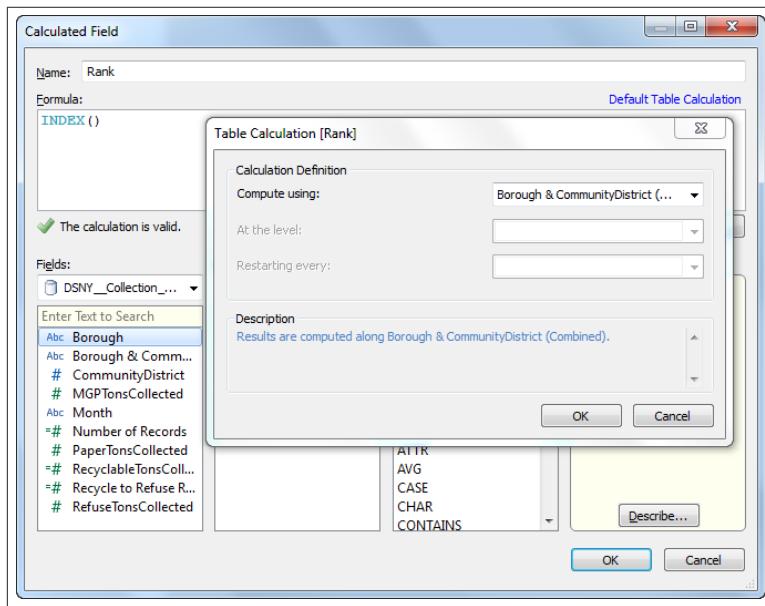
Figure 4-9. Bar chart of community districts

Now we can quickly tell which are the top four or five districts, and which are the bottom four or five. Using the colored bars, we also can see some patterns, like that the districts with the lowest ratio are all in the Bronx (blue), and most of the top ten are in Manhattan (green).

But if I asked you to pick out the tenth or the fifteenth districts, you'd have to count down from the top, right? Not the easiest task to perform. How can we add rank to this list to easily pick out the order from 1 to 59?

## Two Ways of Adding Rank

Let's create a Rank field. To do so, right-click anywhere in the Dimensions or Measures panel on the left and select *Create a Calculated Field* once again. This time, name the field Rank and enter `INDEX()`. As soon as the message appears in the bottom left indicating "the calculation is valid," blue text will also appear in the top right that says "Default Table Calculation." Click on this blue text and change the *Compute using* drop-down to *Borough & CommunityDistrict (Combined)*, as shown in [Figure 4-10](#).



*Figure 4-10. Creating a rank for each of the 59 community districts*

This tells Tableau that the rank should apply to the combination of *Borough* and *CommunityDistrict*.

After clicking *OK* twice, we now see a new field in the Measures panel called Rank. Notice that this field is continuous (the # symbol is green), but we can consider the rankings to be discrete integers, because there is no rank between any two successive ranks. To change the Rank data field type, right-click on it in the Measures panel and select *Convert to Discrete*. The # symbol next to Rank will turn blue, indicating that it's discrete. We're now ready to add Rank to the chart.

To add the Rank field to the bar chart, simply drag it to the left of Borough & CommunityDistrict (Combined) on the Rows shelf. After removing the row dividers, we have the colored, sorted, and now ranked bar chart shown in [Figure 4-11](#).

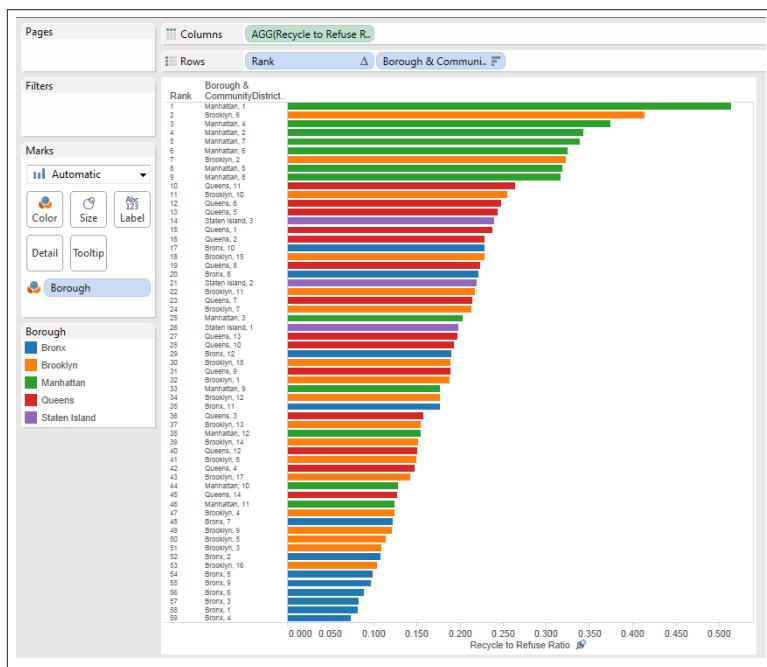


Figure 4-11. Adding the Rank field to the bar chart

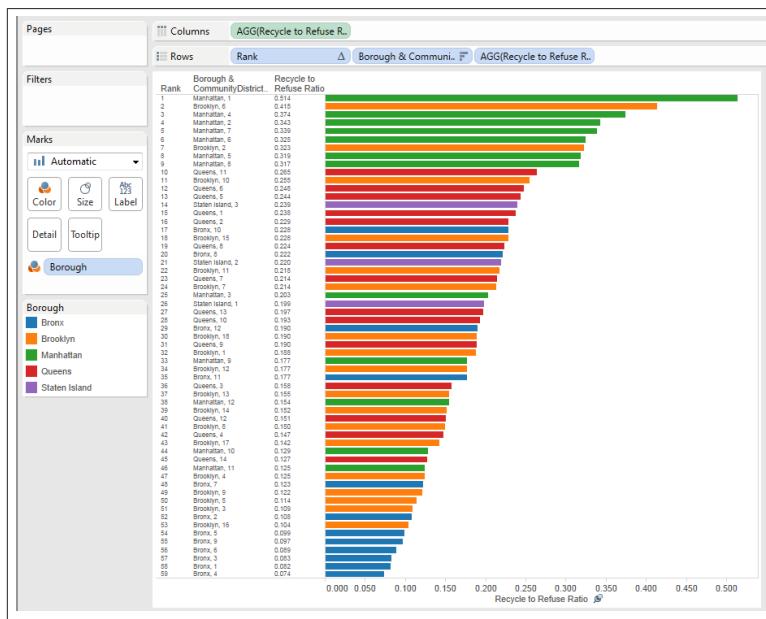
Alternatively, since the launch of Version 8.1, users can create a similar Rank field by making use of the new Rank Table Calculation. To do so, drag another instance of Recycle to Refuse Ratio to the columns shelf, click in the down arrow of the new pill, and change it to *Discrete*, and then click the down arrow again and select *Quick Table Calculation* and then *Rank*. Finally, click the down arrow once more

and select *Compute using*, then choose *Borough CommunityDistrict (combined)*. (This calculated field also lets you handle ties in different ways.)

Now the only task this chart doesn't facilitate is gleaning precise values. We could add the labels to the right of the bars as we did in [Figure 4-5](#) by simply dragging the *Ratio* field to the *Label* shelf, but perhaps a better approach would be to add the labels to the left of the bars so as not to interfere with the perception of the bars' lengths.

Because *Recycle to Refuse Ratio* is a continuous data field (the # symbol is green), if we just drag it out onto the *Rows* shelf to the right of *Borough & CommunityDistrict*, Tableau will create 59 y-axes for each of the bars. Try it and see what I mean.

We just want the ratio value itself to appear, so we'll first convert *Recycle to Refuse Ratio* to a discrete field the same way we converted *Rank*, and then we'll drag the ratio onto the *Rows* shelf to the far right, leaving a blue pill and creating the updated view shown in [Figure 4-12](#).



*Figure 4-12. Precise values added to the left of the bars*

We've now visualized the ratio in a way that makes it easy for our audience to perform a number of different tasks very quickly:

- Get a general sense of how the 59 community districts compare relative to one another
- Determine which districts have the highest ratios, and which have the lowest
- Get a general sense of which boroughs have districts at the top of the list, and which have districts at the bottom
- Know the precise ratio for any district
- Pick out any district in rank order (e.g., the 12<sup>th</sup> or the 28<sup>th</sup>)

This view is a very versatile and useful way to communicate the ratio of recycle to refuse. But it doesn't tell us which community districts produced more trash overall *per person*. For all we know, the districts at the top may be recycling a lot, but they may also be producing a much higher amount per person. That's where rates come into play.

## Rates

Recall that a rate is just a special type of ratio in which the numerator and the denominator have different units. In the previous example, the ratio of recycle to refuse involved terms with the same units—namely, tons. When we consider the trash production rate per person, our numerator is still tons, but now our denominator is the number of people.

But the data set we have been working with so far doesn't include community district population. In order to determine trash production per person, we'll need to find a way to include population data in our analysis.

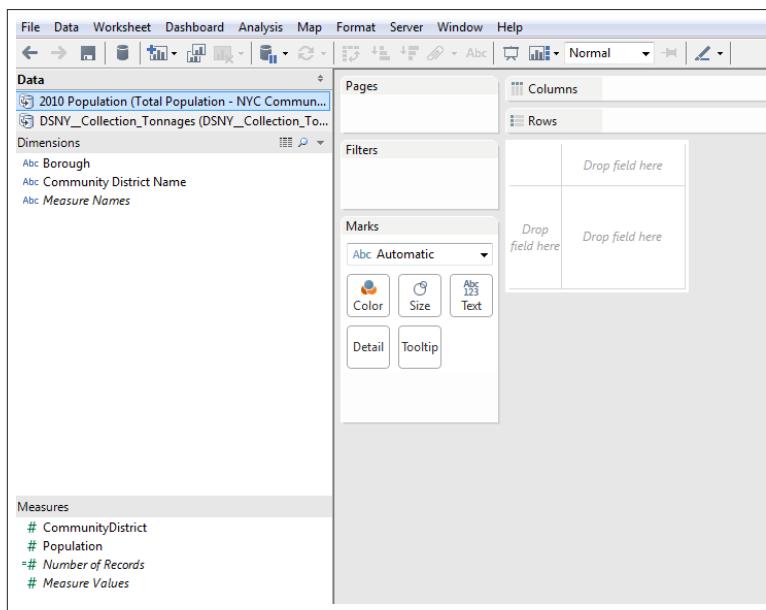
Going back to the data discovery process in [Figure 1-1](#), we've gone all the way around the horse track, and now we have a brand-new question that involves gathering new data.

Ideally, we would find population by community district for September 2011, which is when the DSNY collection data was taken. We can't find that, but we came close: [2010 census data](#) puts us within a year's time of the trash collection data, and is probably suitable to get a reasonable approximation of the trash production rate.

Now that we found population data, how can we create rates with it? Luckily, Tableau allows us to connect our current workbook to this new data set and “blend” it with the existing data in the workbook.

## Blending Data Sources

From the *Data* menu of our current workbook, select *Connect to Data* and then *Microsoft Excel*, and browse to the population data file, connecting to the correct sheet in the file and importing the data as an extract. We now see the new data source appear in the Data area along with the DSNY data, as shown in [Figure 4-13](#).



*Figure 4-13. Adding a second data source to a workbook*

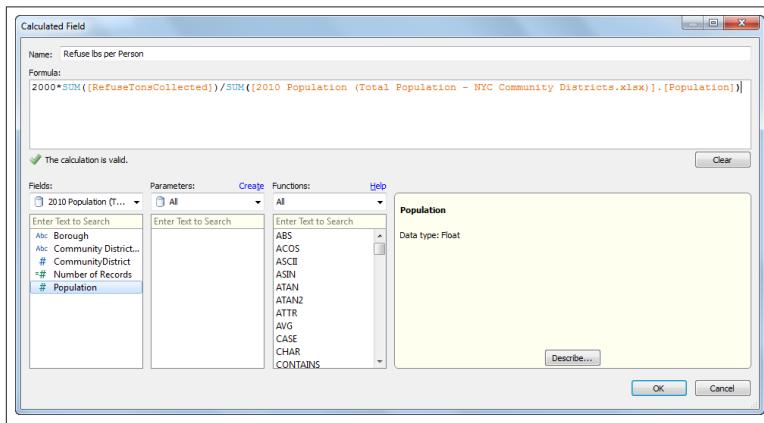
The newly imported data set is highlighted in the Data area, and the fields for that data set are shown in the Dimensions and Measures panels below. If we click on the DSNY data set, the data fields for that data set will show below.

Notice that `CommunityDistrict` is in the Dimensions panel of the DSNY set, but it's in the Measures area of the new population data source. Let's match them up by dragging `CommunityDistrict` from the Measures panel of the population table into the Dimensions area.

Because both data sources now have a Dimension called `Borough` and a Dimension called `CommunityDistrict`, Tableau will link them together for us. If the fields to link had different names, we would have to manually link them using `Data → Edit Relationships`.

## Visualizing Rates

Now that we have linked these two data sources, our updated Tableau workbook contains both terms of the refuse per person ratio: the measurement of the amount of refuse in tons (the numerator) and the count of the number of people in each community district (the denominator). We'll create the rate the same way we created the ratio in the previous section, but now we need to create a calculated field that includes fields from two different sources, as shown in [Figure 4-14](#).



*Figure 4-14. Creating a rate using fields from two data sources*

Using the `SUM` aggregation type for both terms in the rate means that we can get meaningful rates for boroughs as well as community districts. Also, notice we are converting from short tons (US) to pounds (lbs.) by multiplying the numerator by 2,000—which is how many pounds are in each ton. On a per person basis, it's much easier for us to think in terms of pounds, so this is a better unit to use for this comparison.

Now that we have created the rate, we can visualize it. We'll follow the same steps that we used to create the ratio bar chart in [Figure 4-12](#) to create the rate comparison bar chart shown in [Figure 4-15](#).

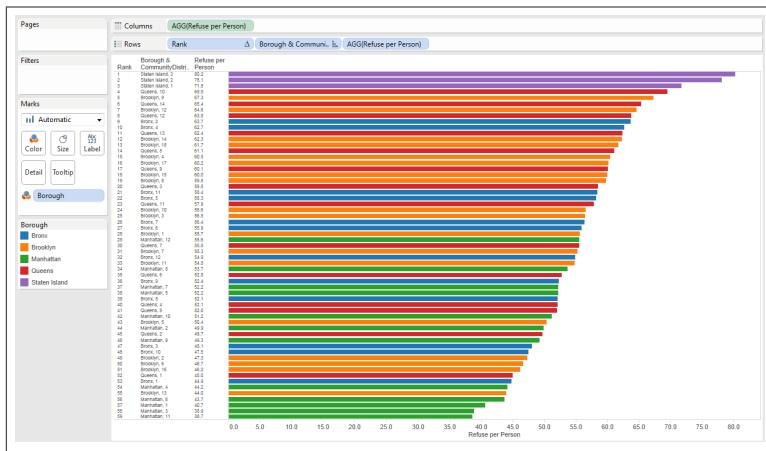


Figure 4-15. Comparing rates in a bar chart

It's clear from this view that the three community districts in Staten Island produced the most refuse per person in September 2011, at a rate of more than 70 pounds per person.



A word of caution about these results: often, data visualization and data analysis are best used to propose new questions to ask. Data is great at helping us make comparisons, but it doesn't always answer "why?" and "how?" It can be easy for us to jump to conclusions based on what the data shows, but some further qualitative investigation is frequently required.

Why are the Staten Island community districts at the top of this chart? The chart itself doesn't tell us. It helps us formulate the next question. Does this mean every person who lives on Staten Island produced 70 to 80 pounds of trash that month? No, that's not what the data shows. The data just shows how much trash was collected from their communities. DSNY provided *overall* refuse collected, not just for residences, so perhaps there are industries or businesses on Staten Island that produced a lot of trash.

The point is to be careful with what we communicate, and not to communicate more than the data gives us license to say.

# Summary

In this chapter, we learned how to use Calculated Fields and blended data sources to create ratios (quotients) and rates (quotients with mixed units). We also learned the Index function as well as the Rank table calculation, and created highlight tables and bar charts to compare different dimensions—at both the borough and community district levels. This enabled us to go beyond overall amounts (“how much”) and to consider relative amounts (“how much per”).

In the next chapter, we’ll consider another type of normalized comparison: proportions and percentages.

## O'Reilly Ebooks—Your bookshelf on your devices!



PDF



ePub



Mobi



APK



DAISY

When you buy an ebook through [oreilly.com](http://oreilly.com) you get lifetime access to the book, and whenever possible we provide it to you in five, DRM-free file formats—PDF, .epub, Kindle-compatible .mobi, Android .apk, and DAISY—that you can use on the devices of your choice. Our ebook files are fully searchable, and you can cut-and-paste and print them. We also alert you when we've updated the files with corrections and additions.

Learn more at [ebooks.oreilly.com](http://ebooks.oreilly.com)

You can also purchase O'Reilly ebooks through the iBookstore, the [Android Marketplace](#), and [Amazon.com](#).

**O'REILLY®**

Spreading the knowledge of innovators

[oreilly.com](http://oreilly.com)