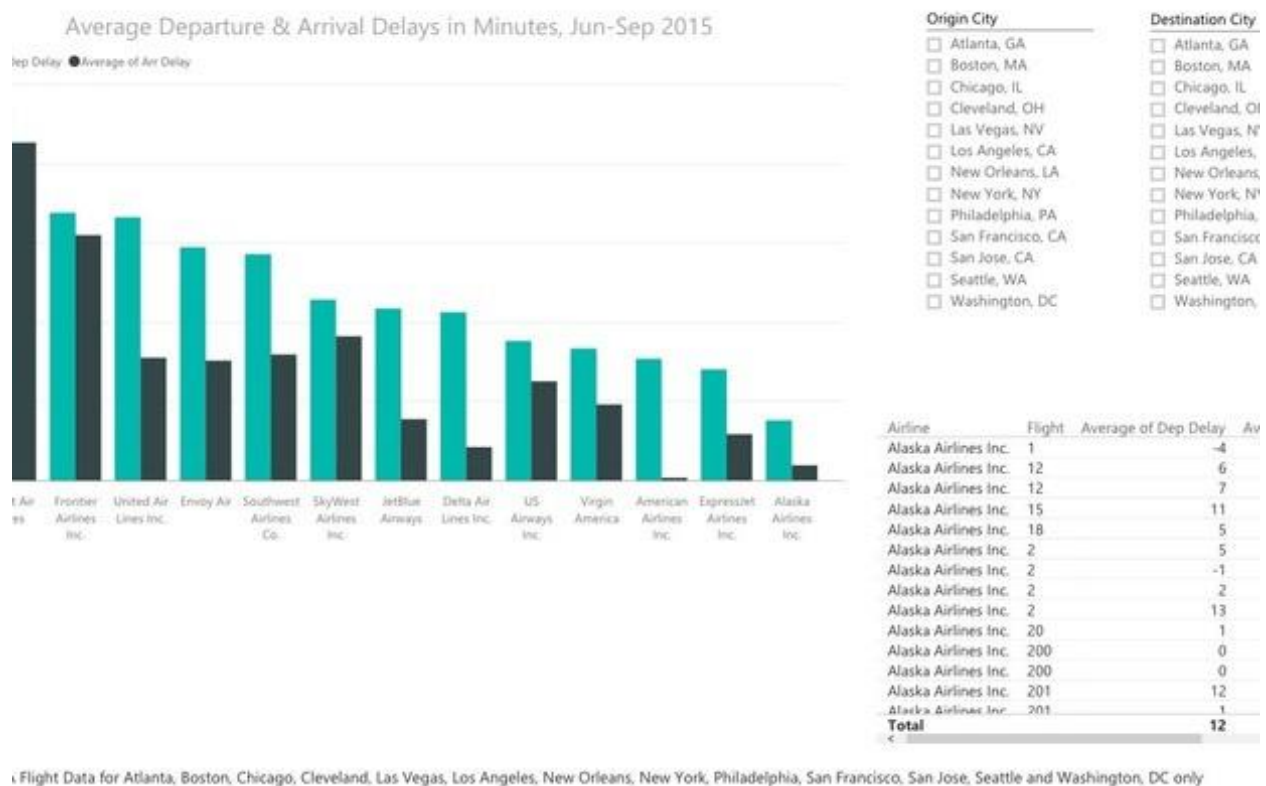


## Free data visualization with Microsoft Power BI: Your step-by-step guide

We'll show you how to analyze a file with more than two million records of U.S. airline flight delays in this hands-on tutorial.

Sharon Machlis | July 11, 2016



Microsoft has jumped into the free, self-service data analysis space with [Power BI](#).

Power BI offers basic data wrangling capabilities similar to Excel's Power Query. It also lets you create interactive visualizations, reports and dashboards with a few clicks or drag-and-drops;

type natural-language questions about your data on a dashboard; and handle files that are too large for Excel.

It can work with dozens of data types -- not only Excel, Access and CSV files, but also Salesforce, Google Analytics, MailChimp, GitHub, QuickBooks Online and dozens of others. And, it will run R scripts -- meaning that any data you can pull in and massage via R you can import into Power BI.

## A guide to Power BI

In this article, I've put together a step-by-step guide to starting with Power BI, along with a number of other resources to help you along:

### What is Power BI exactly?

Power BI includes both a downloadable desktop program and a cloud service, each of which offers different but overlapping capabilities. Data wrangling is desktop-only; visualizations and reports can be created in either; dashboards and report sharing are cloud-only. In addition, there are mobile apps for iOS, Android and Windows that let you view your Power BI or [SQL Server Reporting Services \(SSRS\)](#) reports and dashboards.

At least for now, you can take advantage of most Power BI capabilities without paying -- although Microsoft is clearly betting that you'll like the basic cloud service enough to spring for a \$9.99/month paid account. Chief benefits of the paid account are increased data storage (10GB vs. 1GB), more timely automated data refreshes, the ability to create enterprise "content packs" and higher streaming capacity.

Be advised, though, that Microsoft wants a business email address when you sign up for Power BI cloud service -- while it can't screen out all non-commercial addresses, it won't accept known free consumer addresses like Gmail.com. Accounts from .gov and .mil addresses aren't supported for direct sign-up at powerbi.com either, although addresses at .edu and .org are.

And if you'd like to use any of Power BI's free mobile apps, you'll need a Power BI cloud account or access to your organization's SQL Server.

On the other hand, Power BI Desktop (at least for now) is not only free but doesn't require an account, an email address or a credit card -- just a Windows PC.

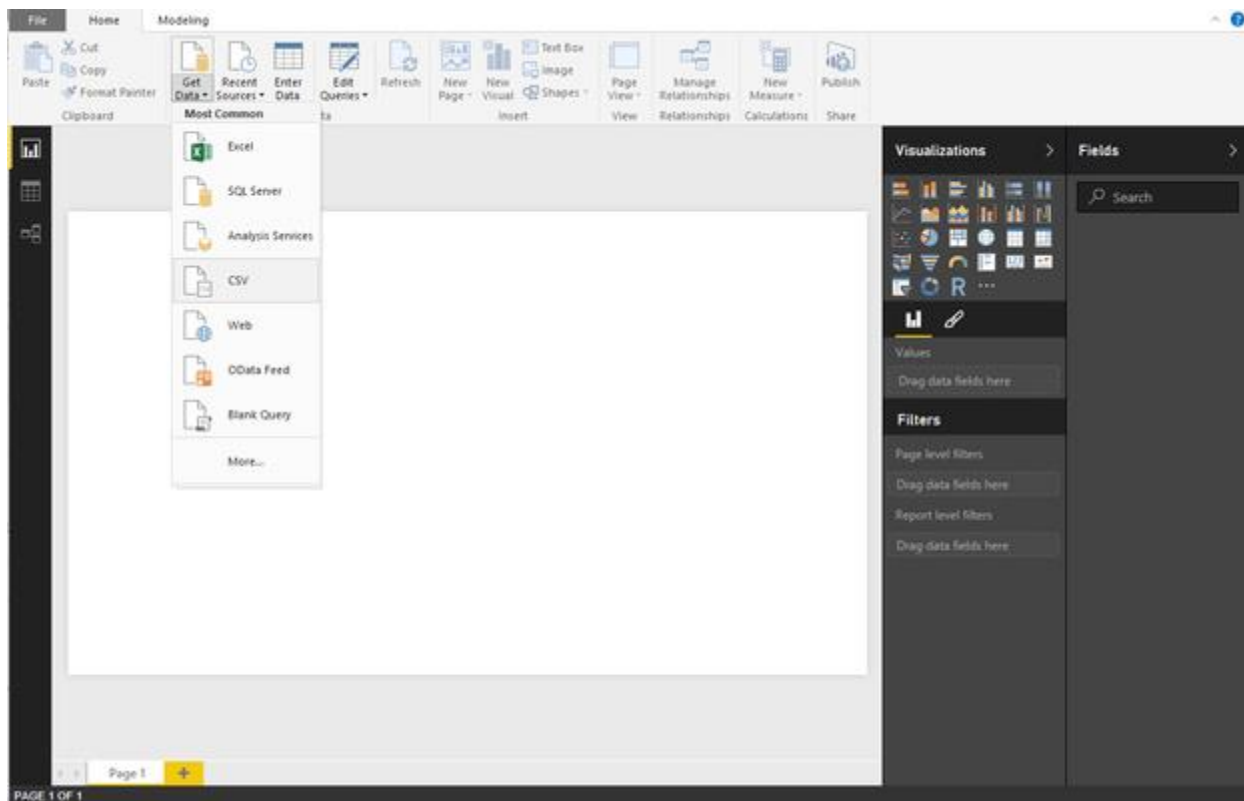
If you'd like to learn how to use this new, still-evolving tool to create reports and dashboards, read on.

## Importing data into Power BI

Power BI Desktop is the better place to begin, unless you're sure that your data is already in the format you need for visualization. (Which may be the case if, like me, you prefer to do your data wrangling with a scripting language like R or Python.)

If you're used to Excel, you might think that selecting File > Open is the way to start analyzing your data in Power BI. But you'd be wrong -- File > Open is only for an *already existing* Power BI project.

Instead, to import new data, click the Get Data button on the Home tab, choose your data source type and click Connect.



To load a file, go to Get Data and select your data type.

This will bring up a familiar Windows file-selection dialog. Choose your file and you'll see a preview of your data. If it looks okay and there's nothing more you want to do to the data before

starting to graph and chart, hit Load. Otherwise, click Edit, which brings up the Power BI Query Editor.

In this article, I'm going to use monthly files of airline flight-delay information from last summer that I downloaded from the [Federal Aviation Administration \(FAA\)](#) website. I know -- especially where airlines are concerned, past performance is no guarantee of future results. But if you're going to book a flight this summer, it might be fun (if not necessarily predictive) to answer questions such as: Which airlines had the best and worst delays last summer? Are there any specific flights that do especially poorly or well? [These Power BI charts](#) can help you easily answer these questions.

If you want to follow along, you can [download your own data files from the Department of Transportation website](#). Or if you want, you can download the same files I'm using here -- the [file download](#) is available to all members of the [Computerworld Insider](#) program; registration is free, so if you're not already an Insider, it's easy enough to sign up. Files include data for domestic flights in the U.S. by month (so if you want to check flights to Paris, this won't help). There are separate files for June, July, August and September.

Start by loading in the June file (*2015\_06\_ONTIME.csv*): Go to Get Data > CSV in Power BI. Select and open your file, and you'll see a preview of your data. Then click Edit (*not* Load) to bring up the Query Editor. Now we can do some data wrangling.

One thing that can be useful to check at this point is whether number columns are loading in as numeric (aligned to the right) or text (aligned to the left). In other words, if you see numbers that are flush left in your data preview, they're not importing correctly -- which is one reason to choose Edit and bring up the Query Editor window even if you don't think you need to make changes in your data's structure.

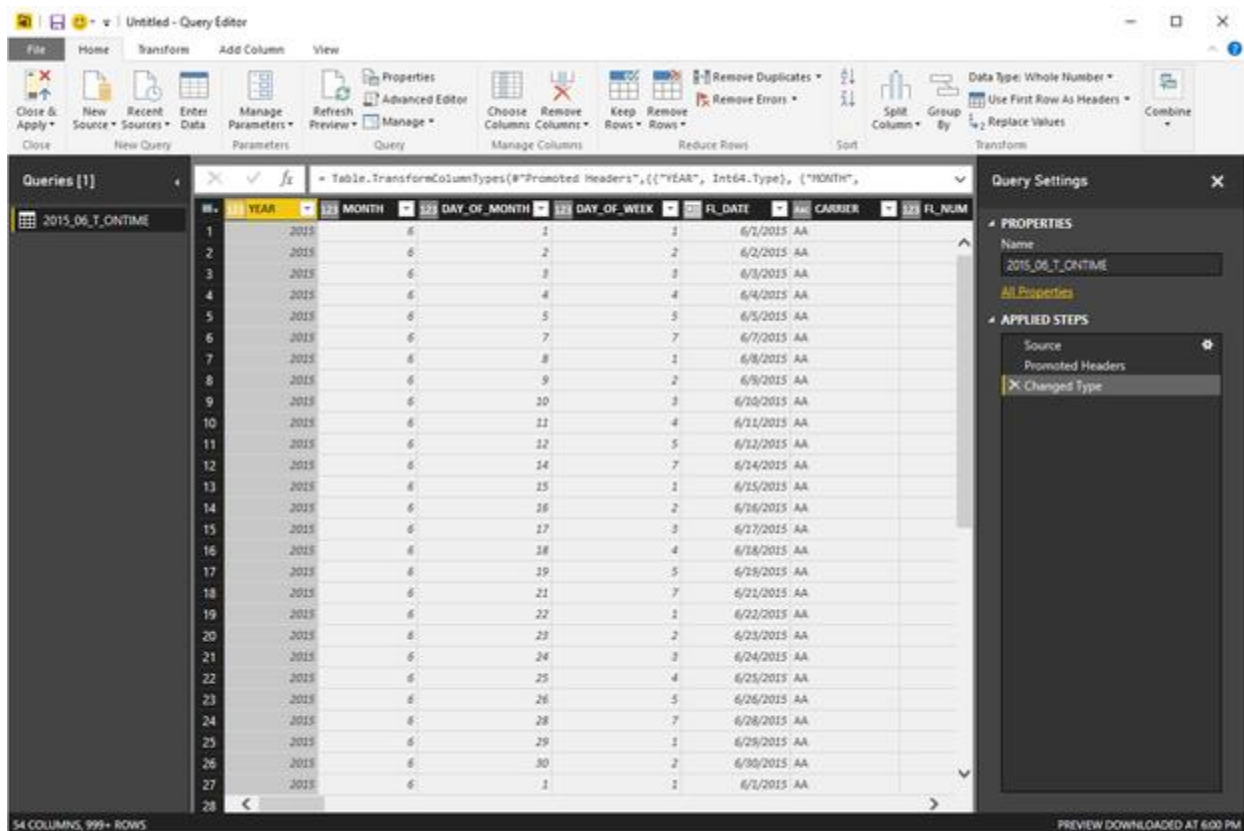
2015\_06\_T\_ONTIME.csv

YEAR	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	FL_DATE	CARRIER	FL_NUM	ORIGIN	ORIGIN_CI
2015	6	1	1	6/1/2015	AA	1	JFK	New York,
2015	6	2	2	6/2/2015	AA	1	JFK	New York,
2015	6	3	3	6/3/2015	AA	1	JFK	New York,
2015	6	4	4	6/4/2015	AA	1	JFK	New York,
2015	6	5	5	6/5/2015	AA	1	JFK	New York,
2015	6	7	7	6/7/2015	AA	1	JFK	New York,
2015	6	8	1	6/8/2015	AA	1	JFK	New York,
2015	6	9	2	6/9/2015	AA	1	JFK	New York,
2015	6	10	3	6/10/2015	AA	1	JFK	New York,
2015	6	11	4	6/11/2015	AA	1	JFK	New York,
2015	6	12	5	6/12/2015	AA	1	JFK	New York,
2015	6	14	7	6/14/2015	AA	1	JFK	New York,
2015	6	15	1	6/15/2015	AA	1	JFK	New York,
2015	6	16	2	6/16/2015	AA	1	JFK	New York,

Load Edit Cancel

Make sure that the number columns are loading in as numeric -- in other words, aligned to the right.

Once in the Query Editor, you can right-click on a column header and select "Change Type" in order to manually select a data type such as whole number, decimal number, date, date/time, etc. But there's plenty more we can do with this data besides checking column types.



Once in the Query Editor, you can start tweaking your data.

*Note:* If you're not interested in data wrangling and want to get started with charts and graphs, load the summer15delays.csv file instead and skip ahead to the Easy visualizations section -- but do make sure that the flight number is changed from numeric to text when you import the file.

## Power BI Desktop vs cloud service

*It can be confusing for beginners to understand when to use Power BI's desktop software and when to use the Power BI cloud service, since they do some but not all of the same things.*

*To begin with, the desktop application runs only on Windows. The cloud service is multi-platform, running on Microsoft Edge, Internet Explorer, Chrome, Safari for Mac and Firefox. In addition, mobile apps are available for iOS, Android and Windows 10 for viewing your reports and dashboards.*

*The desktop app is designed for **authoring**. It's where you can do data modeling, merging different data sources and so on as well as visualization.*

The Power BI cloud service is for **sharing and creating dashboards**. It's also where you can access features like **natural-language questions** ("Hey Cortana...."). Cortana natural-language technology is getting baked into Power BI cloud. You can build visualizations and reports in the service, too, but you can wrangle data only in the desktop app.

Here's a bit more detail:

If you need to **massage your data for analysis** -- combine data sources, add or delete columns, or reshape tables -- you'll want to use the desktop software, since it's the only option for data modeling.

Both the desktop software and the cloud service do basic data visualizations and reports. I asked Microsoft about the difference; a rep told me via email in May: "For **simple data visualization and exploration Power BI Desktop and the Power BI Service offer similar capabilities**, but you'll encounter some differences. For example, R visuals aren't available in the service yet, but are planned for later this summer." (Update: R visuals were added to the cloud service in early July; a paid Pro account is needed to use them.)

To **create dashboards** -- visual summaries of data from one or more different sources -- and ask natural-language questions about data on those dashboards, you need the Power BI cloud service.

To get some **rudimentary automated analysis** (Quick Insights) from your data, you need the cloud service.

To share your analysis with others, you need the cloud service.

Some online data sources, such as Adobe Analytics, require using the cloud service.

You can start off working in Power BI desktop if you need its data-wrangling capabilities and then publish your data or report to your cloud account to create dashboards and share your work.

---

## Data wrangling

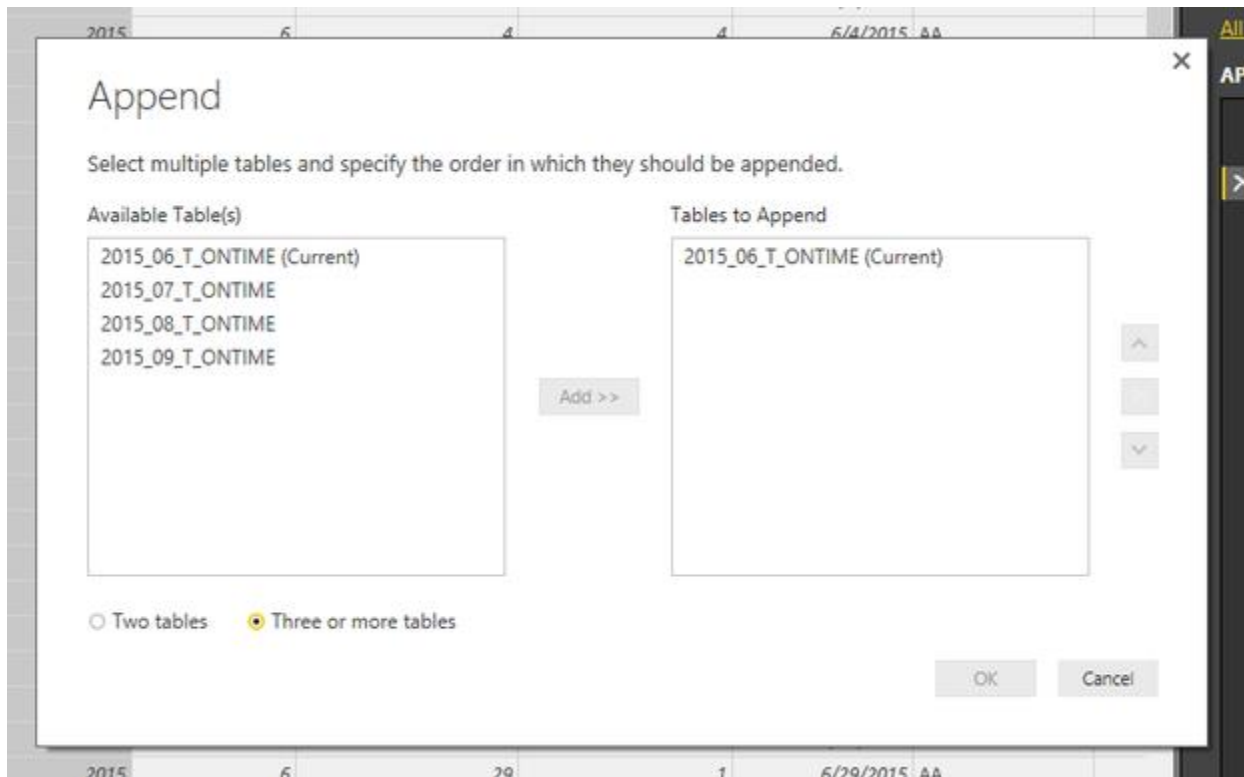
Query Editor lets you reshape and transform data by, for example, merging multiple data sources and pivoting or unpivoting data. It can also handle common data drudgery like adding, deleting, renaming and moving columns.

For the flight-delay data, you want to combine multiple months into a single table, so don't start changing data within June until you've added July, August and September. Otherwise, you'd have to repeat the same transformations on each file before combining them.

To import more CSV files into this active Query Editor window, click on New Source > CSV and choose a file (you can select only one at a time), give the data preview a quick look and click OK.

Each data source should now be listed under Queries in the left panel. Next, in the Queries panel, click on the table that you want to add data to (for this example, you can choose 2015\_06\_ONTIME and add other months to that).

Click Append Queries (if your browser window is narrow, Append Queries may be under a Combine button), click on Three or more tables (the default is two tables) and follow the instructions in the dialog box.



Append Queries lets you combine data sources.

Your 2015\_06\_ONTIME file should now have four months of data, not just June, so it's probably worth renaming it to something like 2015\_SUMMER\_ONTIME. You can right-click on it in the left-side Queries panel to rename it; or, with that query active (clicked on from the Queries panel), look at the Query Settings on the right panel and change text in the Name box under PROPERTIES.

Next, let's try merging two tables using a common column -- what's often called a join in SQL or using a lookup table in Excel.

The CARRIER column in 2015\_06\_ONTIME uses airline codes (such as AA) instead of airline names. It would be nice to have the complete airline names available since not all codes are



intuitive -- AA may be obvious, but VX and B6 less so. I created a file of airline codes and carrier names, *airlinecodes.csv*, that's also available in the *Computerworld Insider* [data download](#). You can import the file into the current Query Editor window using the New Source button.

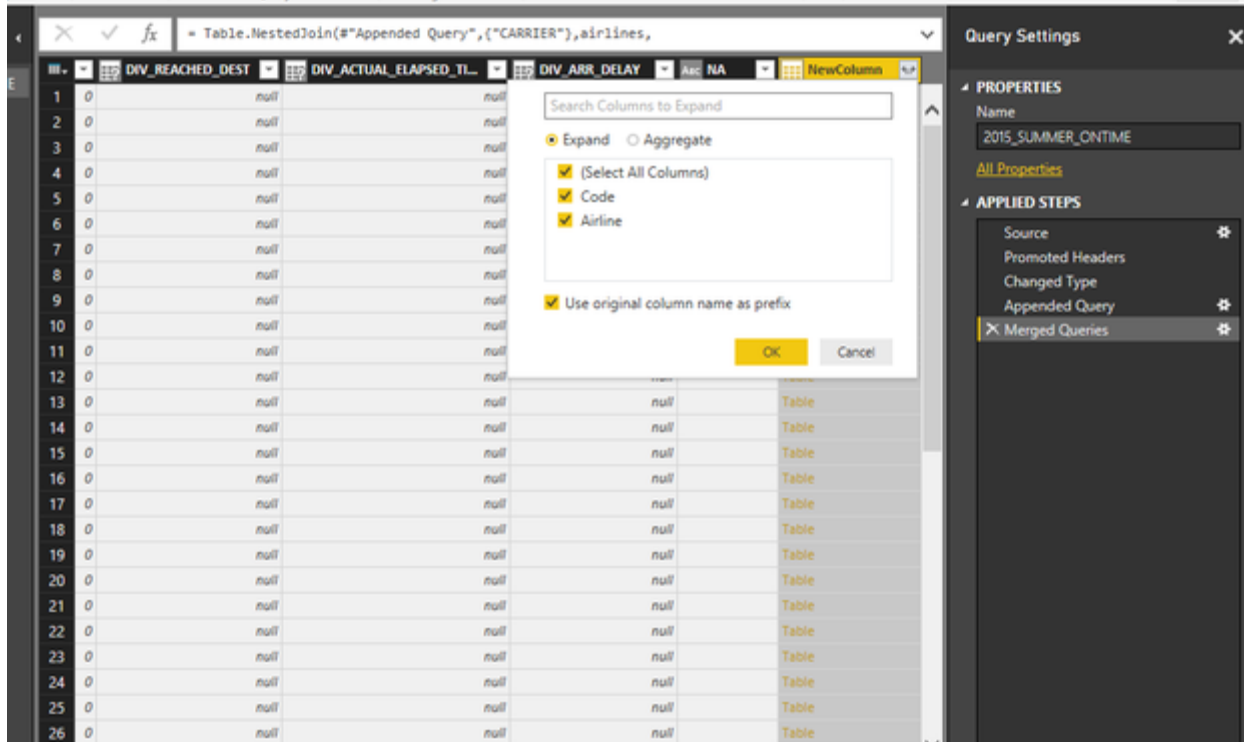
On my system, *airlinecodes.csv* imported into Power BI without the correct column names. Instead, the columns were named "Column1" and "Column2" (obviously the system defaults) while the real column names, "Code" and "Airline," showed up as the first row of data. If that happens to you as well, look to the right on the Home ribbon and you should see an option to Use First Row As Headers. Click on that to fix it.

Next, let's add airline name information to the flight-delay data. With your main data source 2015\_SUMMER\_ONTIME active, select Merge Queries. In the dialog box that pops up, choose the query table with lookup information, in this case *airlinecodes*, and then click on the column header for the common column in each table: CARRIER in 2015\_SUMMER\_ONTIME and Code in the airlinecodes table.

Finally, choose what kind of merge/join you want -- they're described in the drop-down list. For example, the one we want, "Left Outer," means "Keep all rows from the first table whether or not there's a match in the second table." Power BI will now attempt to estimate how many matching rows there are, which is helpful if you've made a mistake and end up with zero matches and thus need to re-examine which columns you're using for the join and whether the data is formatted the same in both.

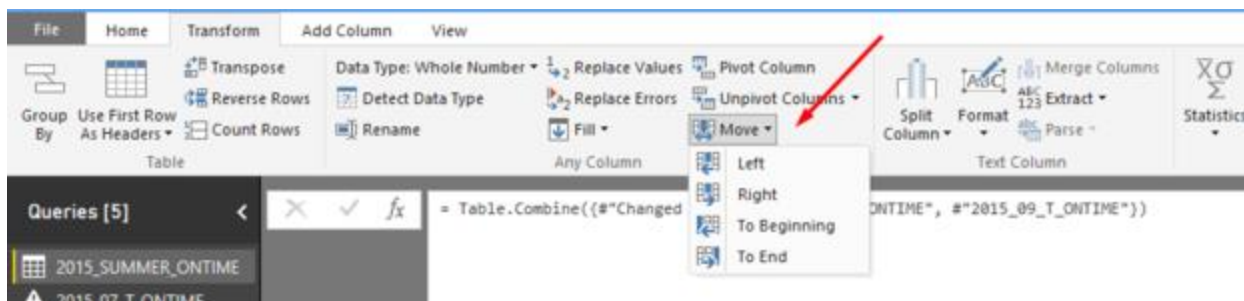
Everything good? Click OK and the join is done.

After the merge, you'll see a column on the far right of your data table labeled NewColumn, with a table icon to the left of the column name and arrows to the right. Click the arrows to expand that table column to view the data -- you can choose the columns you want to add to the original table.



Expand the table column to view the data.

If you want to move that Airline column closer to Carrier, head to the Transform tab. There you'll have the choice to move the new (or any other) column. In this case, using the Power BI interface, you can select the new column and then use the Move > To Beginning and then Move > Right four times in order to get the Airline column next to the Carrier column. But there's a less cumbersome way.



The Transform ribbon has a number of options including moving a column.

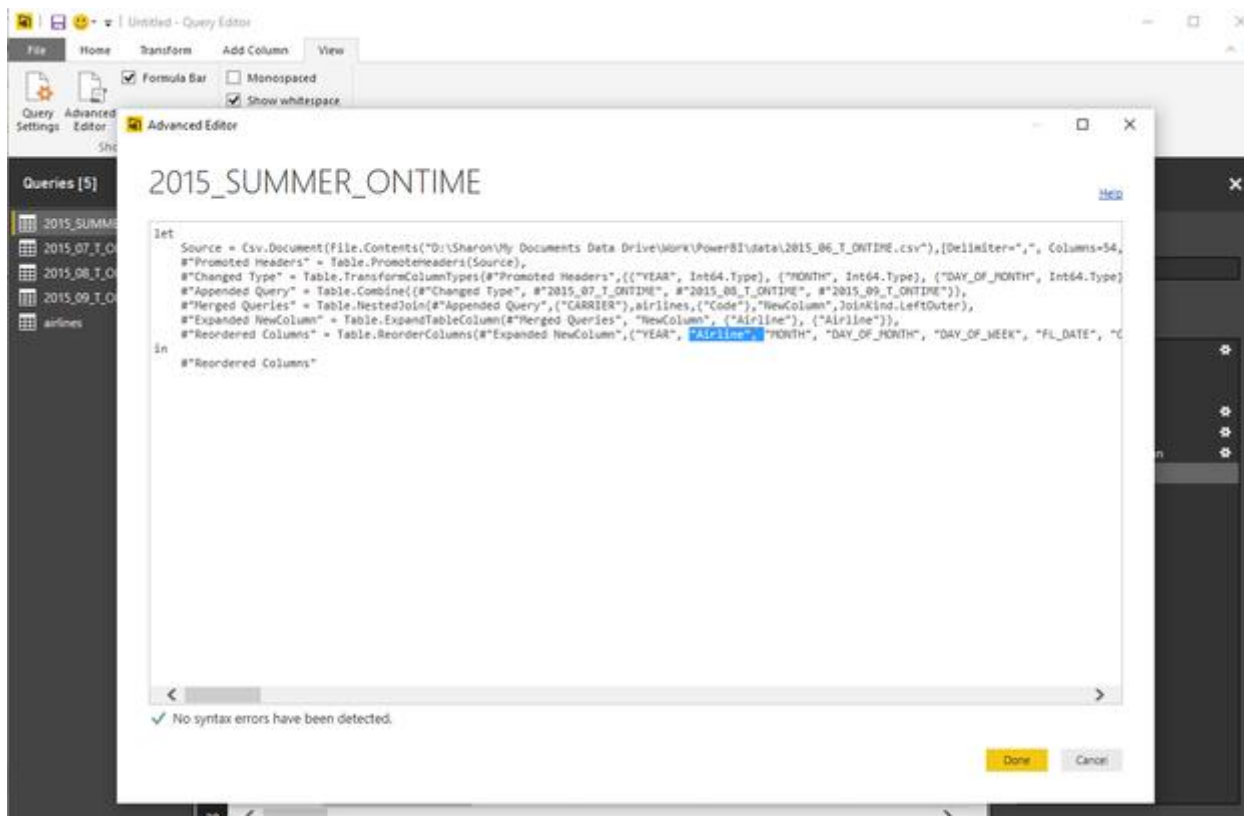
First, if you don't see a formula bar above your table, go to the View ribbon and check Formula Bar so it's visible. Then head to the Transform ribbon, make sure the Airline column is selected, and choose Move > To Beginning. You'll now see a formula in the M language that reorders the

columns. Click Move > Right and you'll see how the formula changes. It looks like you could just cut and paste the Airline column where you want it to go -- but you can't make changes here.



Check the Formula bar under the View ribbon in order to see underlying M language scripts.

Instead, go to View > Advanced Editor and you'll see all your steps as a script. Edit the Table.ReorderColumns line to put Airline when you want it, and click Done. Your Airline column should have moved.



Edit the Table.ReorderColumns line to put Airline when you want it.

Note: As of the July 2016 Power BI Desktop update, you can run R scripts directly in the Power BI Query Editor in order to wrangle your data (or do other tasks) -- but there are a couple of things to know. One is obvious: You need R installed locally on your machine. In addition, the first time you try to run an R script in a project, you'll be asked to submit some information about data privacy; hit Continue and then set both R and your data source to Public. There are also two crucial variable names to keep in mind. The currently active data set in the Power BI Query Editor is stored in the R variable `dataset`; whatever data frame you want returned from your R script should be saved in an R variable called `output`. After running your script, you should get a table similar to the one you saw when doing the merge query command above; that table will need to be expanded. Shortly after this was unveiled I ran into problems when running R scripts on complex data sets after merging or appending queries, but I'm hoping this fledgling functionality will improve.

There are a number of other operations available on the Transform menu, such as grouping values by one or more columns, splitting columns based on a separator or specific number of characters, and pivoting (similar to creating Excel pivot tables) or unpivoting (moving multiple columns of data into two columns: variable data, such as "Month" and value data, such as "June").

You can also do some basic statistical exploration using options on the Transform ribbon, such as counting rows or creating summary stats like count, average, median and standard deviation. When you perform a data summary, a table with the new summarized values will appear in place of your original data table. To get the view of your original data back, go to the Applied Steps column to the right of your data table and delete the summarizing function.

In fact, you can use Applied Steps to delete any step you've taken here, not just the most recent, which is rather handy. Unfortunately, there's not also an Undo after deleting a step in the Query Editor, so you can't bring something back that you've erased.

Select the Count Rows menu option on the Transform ribbon to see the number of rows; if you're using flight-delay files downloaded from *Computerworld*, you'll see there are slightly over two million rows -- more than you could pull into Excel and its [limit of 1,048,576 rows](#). Delete the Count Rows step to return to the full data.

Right now there are over 50 columns in the summer flight-delay table. To make things more manageable, select just a few that you might want to use in some data visualizations: MONTH, FL\_DATE, Airline, FL\_NUM, ORIGIN, ORIGIN\_CITY\_NAME, DEST, DEST\_CITY\_NAME, CRS\_DEP\_TIME, DEP\_DELAY, ARR\_DELAY, CANCELLED, CARRIER\_DELAY, WEATHER\_DELAY, NAS\_DELAY, SECURITY\_DELAY, LATE\_AIRCRAFT\_DELAY.

Click the Choose Columns button on the Home ribbon, deselect Select All Columns and then check the boxes next to the columns you want to keep and click OK.

If you later decide you want to add back any column, you can return to the Query Editor by clicking Edit Queries and then click the settings gear next to Removed Other Columns in the Applied Steps column. That brings the dialog box back up with your prior selections, and you can add more columns to your project.

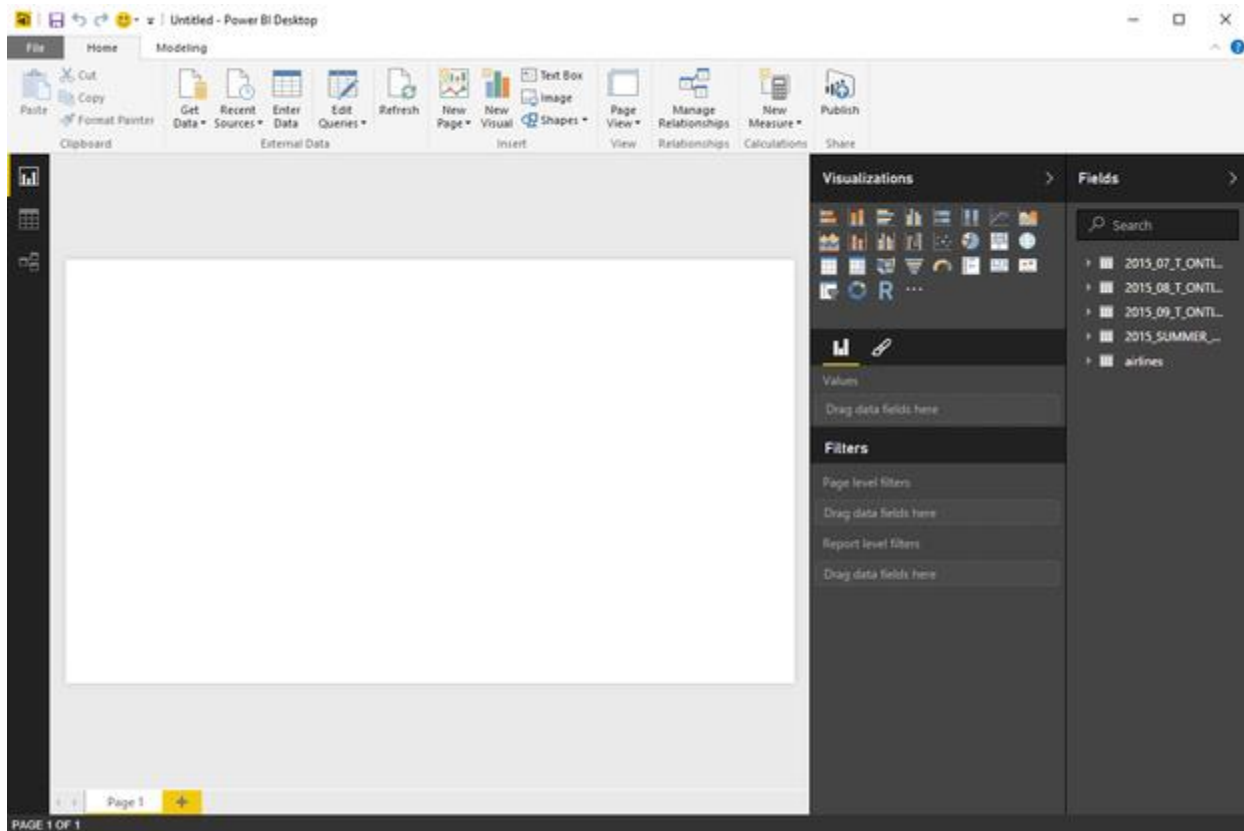
Finally, make sure to change columns that don't make sense as numbers into text. For example, unless you want Power BI to calculate sums of your flight numbers -- and you don't -- turn them into text. Otherwise, attempting to make graphs with these fields as categories won't work, and you'll end up wasting a fair amount of time (don't ask me how I know). You can convert FL\_NUM and MONTH into text by right-clicking the column header and choosing Change Type > Text. Another tweak: The CANCELLED column shows 0 for false and 1 for true -- if you want, you can change that type to True/False.

Now comes the fun part.

## **Easy visualizations**

Close out of the Query Editor with Close & Apply to get back to the main Power BI application. (Because these files are so large, this may take a little while.) At this point you might also want to do a File > Save to save your work as a Power BI project so that if something happens, you won't need to re-create your steps.

Your Power BI blank canvas should look something like this:



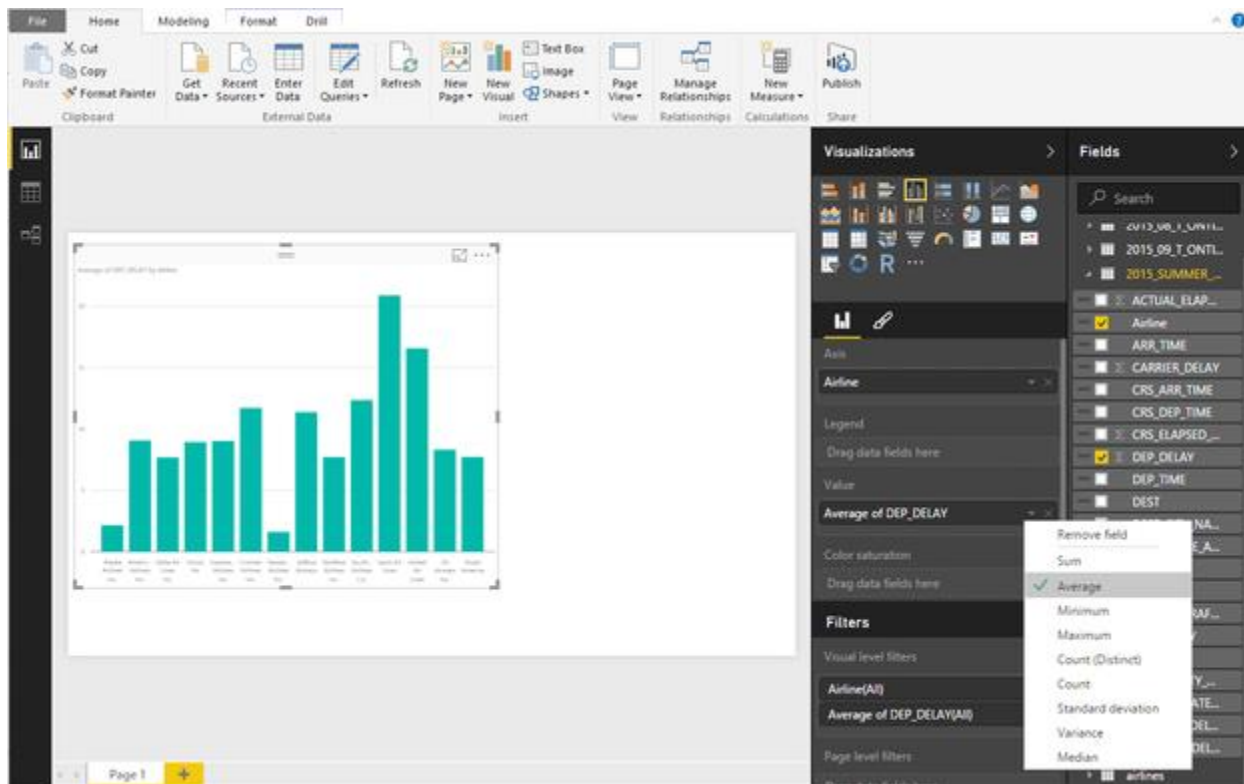
A Power BI blank canvas.

It would be nice to get rid of all the other query files that we merged into 2015\_SUMMER\_ONTIME, since we won't be using them again. If you go to the Fields panel on the right and try to right-click and choose Delete, you'll get a warning that you can't delete those queries because they're referenced by another query. However, you can right-click and then *Hide* them to get them out of the way.

Do that for all the queries except 2015\_SUMMER\_ONTIME. Now that there's only one query table left, the 2015\_SUMMER\_ONTIME table will automatically expand to show its column names. (When there are multiple queries in the menu, the column names don't appear by default; click the triangle next to one to expand it; right-click and choose Expand all to see all the queries' fields.)

Want to graph departure delays? Click the check box next to DEP\_DELAY. You'll get a bar graph with one bar summing all the delay time in your data. (Power BI tries to guess what visualization will work best with your data if you don't specify.) Click Airline, and the graph will morph into a bar graph of each airline's total flight delays.

A sum of total delays by airline isn't that useful, though; the more flights an airline has, the more minutes delayed it's likely to have. But if you look under the Visualizations panel and find DEP\_DELAY under Value, you can click the triangle drop-down menu and change the Sum of DEP\_DELAY minutes to Average (or Median, if you prefer).



Clicking the triangle next to a Value field brings up visualization options, such as changing Sum to Average.

(If the graph is too small, just click and drag a corner of the box to resize it.)

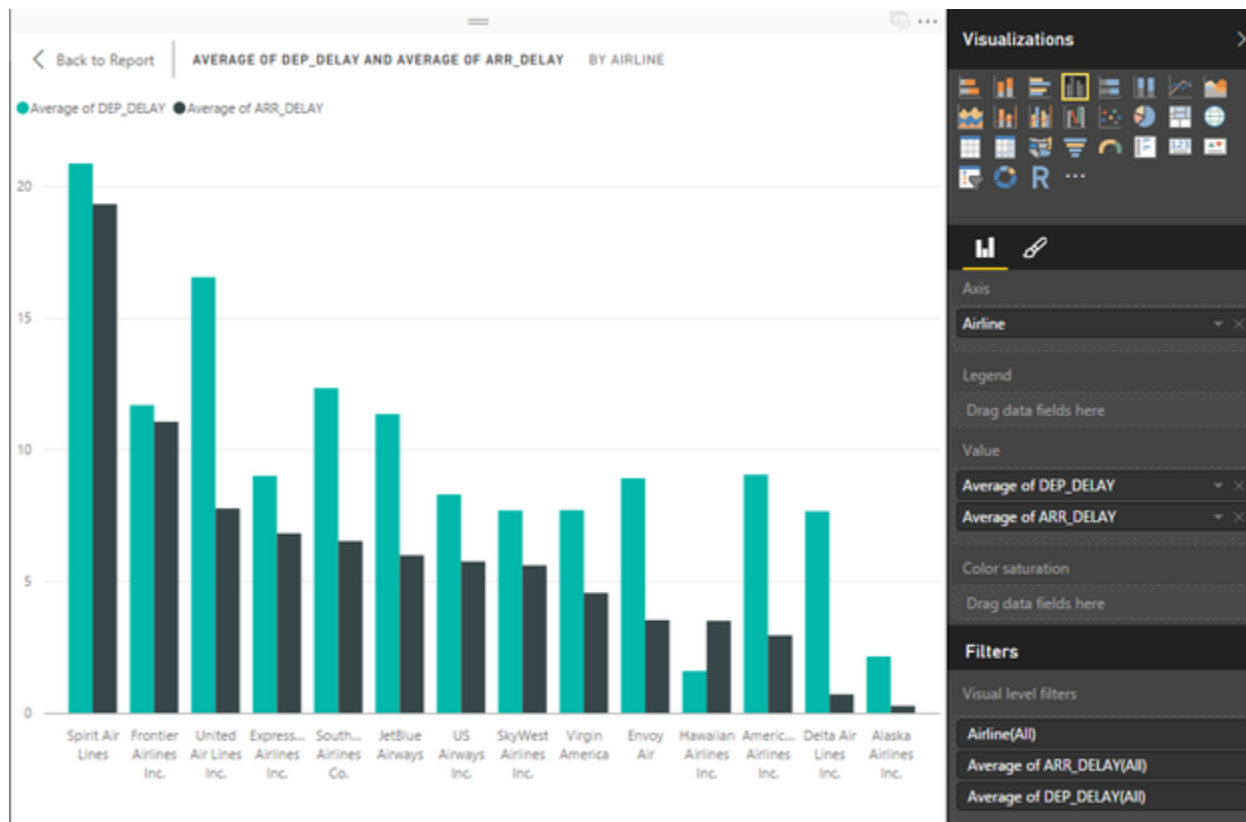
The graph is currently sorted by airline alphabetically. If you want it sorted from highest to lowest, click the ellipsis at the top right and choose Sort By > Average of DEP\_DELAY.



Clicking the ellipsis at the top right of a graph brings up sorting options.

Arrival delays would be interesting, too. Drag `ARR_DELAY` onto the Value area of the panel and change it from Sum to Average. Click the ellipsis again at top right to sort by `ARR_DELAY` descending (if it defaults to ascending, just click again). You should now have a graph that looks like this:





A graph showing departure and arrival delays.

(You can see a version of your graph that takes up the whole canvas by clicking on the "Focus mode" icon next to the ellipsis at the top right of the graph's borders.)

The default title is a little lame. You can change the title and a lot of design defaults by clicking the brush icon just below the Visualizations options. To edit a specific graph, make sure you click it so it's active (the bounding box is visible). Click on Title and you'll see options to change the title text, font color, size, alignment and so on.

There are also options to add data labels to the bars, change the X- and Y-axes, and more.

For example, you can change the title to Average Departure & Arrival Delays in Minutes, Jun - Sep 2015, bump up the text size and center-align. If you want to rename fields, you can do that on the Fields panel -- right-click a field and choose rename. Try changing DEP\_DELAY to Dep Delay and ARR\_DELAY to Arr Delay.

Well, it sure looks like Spirit Airlines wasn't the best choice last summer -- but maybe the airline has big delays in some airports but not others? We can revise the visualization to include just a few airports.

There are a couple of ways to narrow in on a subset of data. Filters can apply to one graph, the entire page or all pages in a multi-page report. You can change filters while editing a report; but if you subsequently share your work as a Power BI dashboard, read-only users will see only what you filtered and won't be able to add data back in.

Another choice is to add an interactive slicer. With slicers, if you later share your work with others, they can interactively filter data and add it back in.

Let's try both.

---

## **Personal vs. Enterprise gateways**

*Power BI has two different "gateways" that allow you to automatically refresh data from local data sources: [personal](#) and [enterprise](#).*

*If you'd like to automatically refresh local files residing on your system, you can use the personal gateway. It requires a paid Power BI Pro account and 64-bit Windows on your PC. For scheduled data refreshes, your system needs to be on and not asleep for the refresh to take place.*

*As you might imagine, the enterprise gateway is designed to be installed on a server (although it can also be installed locally) and, like personal, requires a Power BI Pro account. With the enterprise version, IT professionals can manage access for multiple accounts.*

*To automatically update data that's elsewhere in your organization, such as an in-house database server, you can use either the personal gateway or the enterprise gateway.*

*If you want to refresh data from an R source, you currently need the personal gateway, although adding that capability to the enterprise gateway is on Microsoft's roadmap.*

*Note that if you're connecting to an external online data source such as Salesforce, you don't need a gateway.*

---

## **Filtering your data**

When I started working with Power BI in March, a major drawback was that you couldn't add a text search box to a report, table or slicer. If you were analyzing information with a lot different categories, such as U.S. flight data, it was pretty annoying to have to scroll through hundreds of cities on a list in order, say, to find St. Louis.

As of the June 30 Power BI Desktop software update, you can add a text-searchable slicer to your reports, making it easier to hone in on one item amidst hundreds (or thousands). More on that in

a bit. But it's also possible that you know there are only a few items of interest among the hundreds in your list, and you'd like to create a report with just a subset of the data.

One way to do this is to filter a report down to a few key categories -- in this case, perhaps showing only some cities that are of known interest, such as where your company has offices.

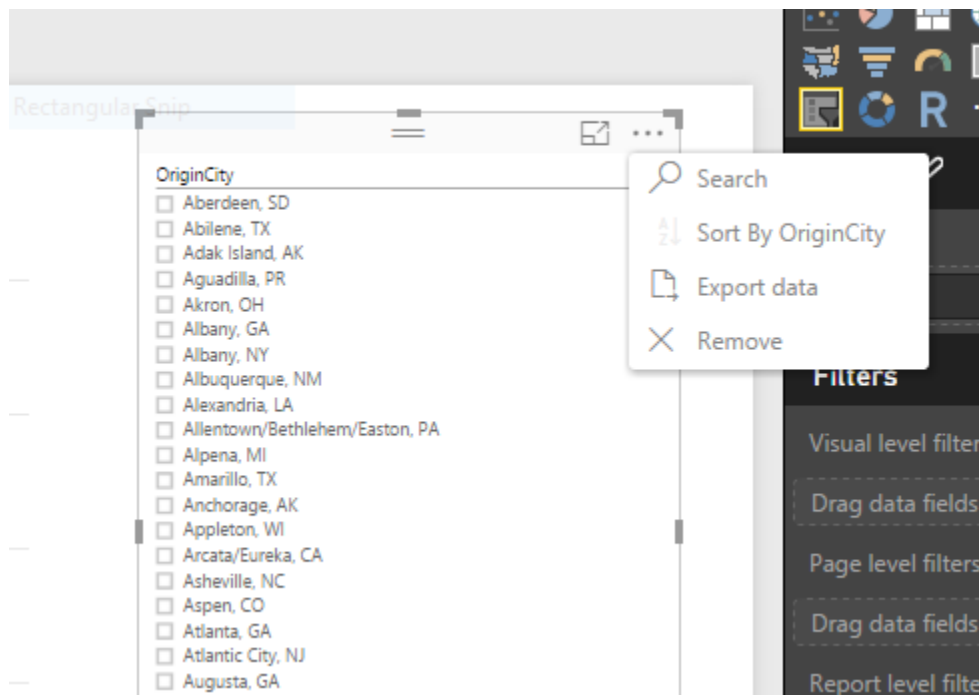
To do this, click on an empty area of the canvas and then drag `DEST_CITY_NAME` onto the Report level filters (where you see the "Drag data fields here" area). Pick a few cities. If you're following along, I chose Atlanta, Boston, Chicago, Cleveland, Las Vegas, Los Angeles, New Orleans, New York, Philadelphia, San Francisco, San Jose, Seattle and Washington, D.C.

Click the `DEST_CITY_NAME` header on the filter to close it. Then do the same for `ORIGIN_CITY_NAME` -- drag it on top of the `DEST_CITY_NAME` filter and select the cities you want -- and you'll just have info for flights between your key cities.

At this point, it may be worth noting on the report itself that the data is now only for a few cities. You can add text to the page by clicking on the Text Box button on the Home ribbon. Move and shape it the way you want on the canvas and then write some text explaining which cities the report covers.

We can now make it easy for users to pick origin and destination cities by adding a couple of slicers. Click onto an empty area of the canvas, then click on the slicer visualization icon (it looks like a little filter/funnel on a table icon under Visualizations -- in the May 2016 version of Power BI, it's the third from last icon under Visualizations). Check `ORIGIN_CITY_NAME`. Now click on an empty area of the canvas again, click on the slicer icon a second time, and then click on `DEST_CITY_NAME`. Size and move slicers around the canvas as you like.

If you still have enough cities in your slicer that adding a search box is worthwhile, click the ellipsis in the top right of the slicer and select Search. That will add a text search box to the slicer.



Adding a search box to a slicer.

If the text is a little small and hard to read, click on each slicer, then click the brush icon and choose a new text size under Items. Just as with the graph, you can change the title and click on the fields to rename them (from, say, ORIGIN\_CITY\_NAME to Origin City and DEST\_CITY\_NAME to Destination City) and increase the Header font size.

You can probably now see the benefit of filtering the data first: Without that page-level filter, there would be more than 300 cities to scroll through on each slicer.

Finally, it might be interesting to see the actual flights, not just the airline. Drag Airline to an empty spot on the canvas and then add FL\_NUM. You'll get a table. Add Dep Delay and Arr Delay, and then once again make sure to change both from Sum to Average (under Values). Rename FL\_NUM to Flight. You can add the scheduled departure time by clicking on CRS\_DEP\_TIME and adding that to the table, too.

Now when you click on an origin and destination city in the slicers, you'll see all available flights and their average arrival and departure delays. If you click on one airline's bar in the graph, the table will show just that airline's flights.

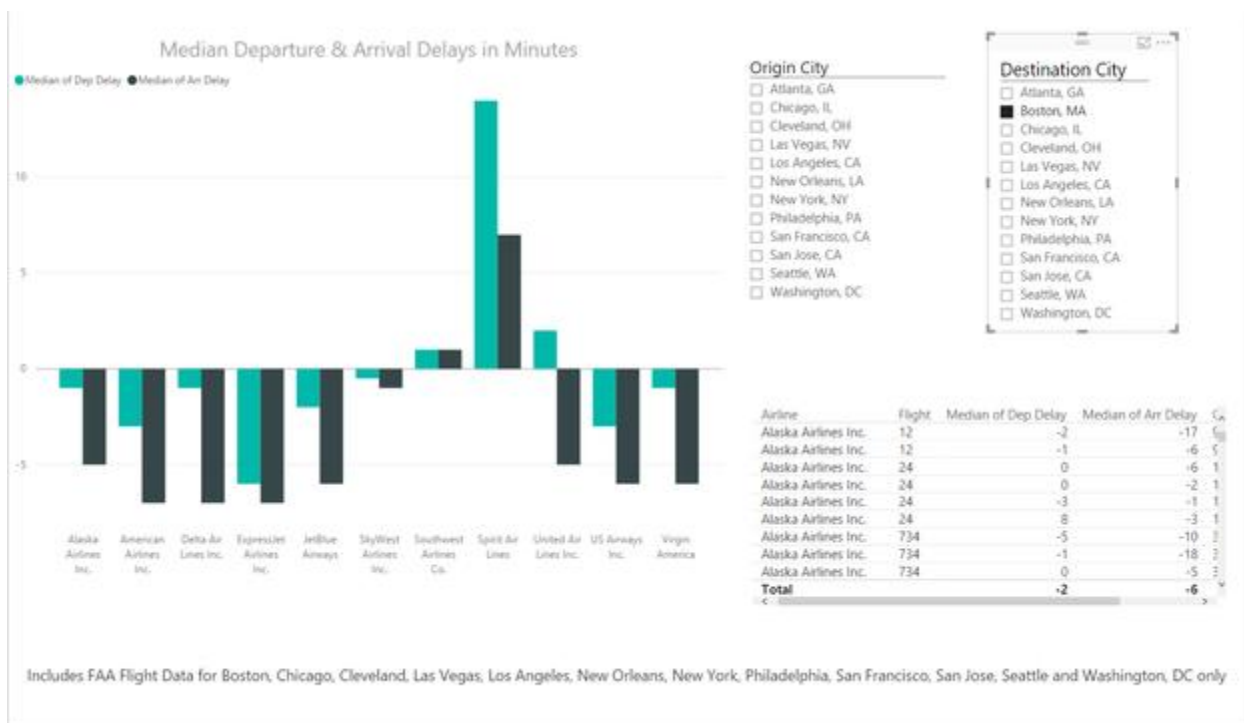
(Note: It's not very easy to find, but you can customize how the graphics on your page interact with each other. Click on one graphic to activate it; then on the Format ribbon, choose Edit

Interactions. The other graphics on the page will all have some additional icons: A filter and a circle with slash through it. Clicking on the filter means that graphic will change based on what happens in the active graphic; clicking the circle with slash means it will not.)

It's also easy to take the same graphics and decide to look at medians instead of averages, since a few unlucky *very* late flights could have an outsized effect on overall averages. As in Excel, you can add a page to your Power BI report by clicking the plus sign next to the tab with the page name (default should be Page 1).

Even handier, since we've got slicers and a graph all set up: Right-click on the page and duplicate it. It's now pretty easy to click on the graph; under the Value section, click on Average of Dep Delay and Average of Arr Delay and change each to Median. If you're following along, you'll also want to change the title of the graph and the table with flight data from Average to Median.

Every airline had a 0 or below median arrival delay for all these cities combined -- except for Spirit. When I just look at flights arriving in Boston, Spirit's delays look even more pronounced -- although to be fair, they might have just had a bad summer in 2015 and improved since then.



A graph showing flights arriving in Boston.

## Interactive drilldowns

Interested in how average delays break down by month? Power BI has automatic drilldown by date fields, which we can see by creating a new visualization on a new page.

Again, right-click on Page 1 and duplicate it, click on the graph to activate it, then unclick Airlines and click on FL\_DATE. You'll only see two bars on the chart, one each for arrival and departure.

That's because Power BI defaults to graphing by year, and we've only got one year's worth of data. Under Axis, you can click the x next to Year to delete that so the graph will stop aggregating annually (which is somewhat useless for this data). It now defaults to Quarter. That, too, isn't much use for this particular four-month data set, but let's pretend it is.

To enable Power BI's date drilldown, click the down arrow at the top right of the graph. Now, if you click one of the third-quarter bars, it will drill down to show you months. Click a month's bar, and it will zoom in on days for that month.

To go back up to larger time groups, click the up arrow at the top left of the graph. Note that while you're drilling up and down, you'll no longer be affecting other visuals on the page, so data on the table won't change.

The date drilldowns are automated for date fields, but you can set up drilldowns for any hierarchy. Activate the graph on the first page, then drag FL\_NUM onto the Axis field, making sure it ends up below Airline. Nothing will appear to change on the graph except that drilldown icons appear.

Click on the down arrow at the top right to activate drilldown, click on an airline's bar, and you'll see all the data for that airline's individual flights. Again, because drilldown is active, you won't see any changes on the table. If you want to be able to manually filter the table for a specific airline while this is going on, you can either temporarily add Airline as a page-level filter or add a third slicer for Airline.

Click the up arrow at the top left to get back to the original graph, click the down arrow again to deactivate drilldown capabilities if it's still selected, and the graphic will work as it did before.

If you want to change the headline for the graph on this page, make the graph active once again, click the brush icon on the Visualizations panel and then click on Title.

For a final step, you may want to rename the page tabs from "Page 1" and "Duplicate of Page 1" to something more meaningful. This doesn't currently work the same way as in Excel -- instead of right-clicking a tab, you need to double-click on the tab name.

There are a lot more visualizations you can generate within Power BI. In addition to all the icons in the Visualizations panel, including tree maps and actual geographic maps, there are other graphics available for import from the [Custom Visuals Gallery](#). If you find one you like, download it from the gallery and then import it using the ellipsis next to the last icon in the Visualizations panel. You've got to import it separately into any report where you want to use it. (You can find an example of one of the more recent custom visuals created by Microsoft Research [here](#).)

There are a lot of other ways to visualize this data, such as looking at the columns with reasons for delays, but for now I'll move on.

---

## 12 important Power BI visualization concepts

*If you'd like to give Microsoft's Power BI platform a try for analyzing and visualizing your data, there are a few key ideas that will help you master the software.*

*1. One way to make a new visualization in Power BI is to click on the visualization type you want under Visualizations on the right-hand panel. You'll see icons for various types of bar charts plus options like scatter plot, tree maps, geographical maps, line charts and a few more. That puts a blank visualization on your canvas.*

*Next, you can drag fields from your data tables onto Axis, Value, Legend and other choices in the right-side panel; or you can just check boxes next to the fields and Power BI will try to determine how to use them.*

*Another way to make a new visualization is to simply click the check boxes next to fields without first picking a visualization type, and Power BI will try to guess the type of data visualization that will work best. If it's not what you want, you can then click a Visualizations icon to change the type.*

*2. If you want to work on an existing visualization, make sure you've clicked on it so it's active -- you can tell that it's active if there's a bounding box with draggable corners visible. If you want to create a new visualization, make sure none of your existing visualizations on the page show that bounding box before checking boxes next to your fields (clicking on an empty part of your canvas will deactivate any existing visualizations). Or you can just start dragging fields onto an empty area of your canvas.*

3. To change a field's calculation from, say, Sum to Average or Count, click the triangle next to that field under Value in the Visualizations panel and pick what you want. As of this writing there were nine options for numerical fields.

4. To change design options such as title, fonts and colors, click on the pen icon under the visualization choices. The three-bar bar-chart icon lets you go back to working with your data.

5. You can narrow down the amount of data that shows up in a single visualization, on a page or on an entire report by adding filters. Drag a field onto the "Visual level filters," "Page level filters" or "Report level filters" area. The basic filtering options will let you select one, several or all options in a category; advanced filtering gives you options such as contains or starts with. (Alas, still no regular expression filtering here and none planned.)

Any fields you add to a particular graph will also automatically show up as visual-level filters. So if you only want a subset of your data to show on one specific visualization, click that visualization to make it active. Then click on it under "Visual level filters" under the Filters heading in the right-hand Visualizations panel to get your filtering options.

6. If a chart or graph isn't sorted the way you want, click the ellipsis at the top right while the visualization is active and then choose Sort By. You'll get choices to sort by fields, ascending or descending.

7. Allow your users (and yourself) to interact with data in a graph by adding a slicer visualization for data categories.

8. Don't overlook the desktop software's Quick Access toolbar -- the icons above the ribbon listings that include a smiley face (which lets you send feedback to Microsoft). Just as in Excel, you can add your own most-used icons up there by right-clicking any menu item. The Query Editor has its own, separate Quick Access toolbar.

9. If you think you'll need to tweak your data in any way -- even just rename columns -- work in Power BI Desktop until you're ready to share. (And remember that to share, you need to first publish to Power BI cloud.) You can't do any data wrangling using the cloud service, but you can publish multiple project changes from your desktop to the cloud as you're working. Publish currently only works one way, though; you can't send changes back from the cloud to Desktop.

10. To share your work privately among other Power BI users, create a dashboard in the cloud service and share. To share your work publicly, create a report in the cloud service and share to the web. If you want to share a report page with colleagues privately, pin the entire report page to a dashboard and then share the dashboard.

11. Microsoft has fairly robust [Power BI documentation online](#).

12. If there's a way you think Power BI should be improved -- new feature, change how something works -- head over to the [Power BI Ideas forum](#). Someone else may have already posted the idea and you can vote for it, or you can add a new one. Microsoft officials say they



*pay close attention to this user community when deciding what to include in Power BI updates.*

---

## Sharing your work

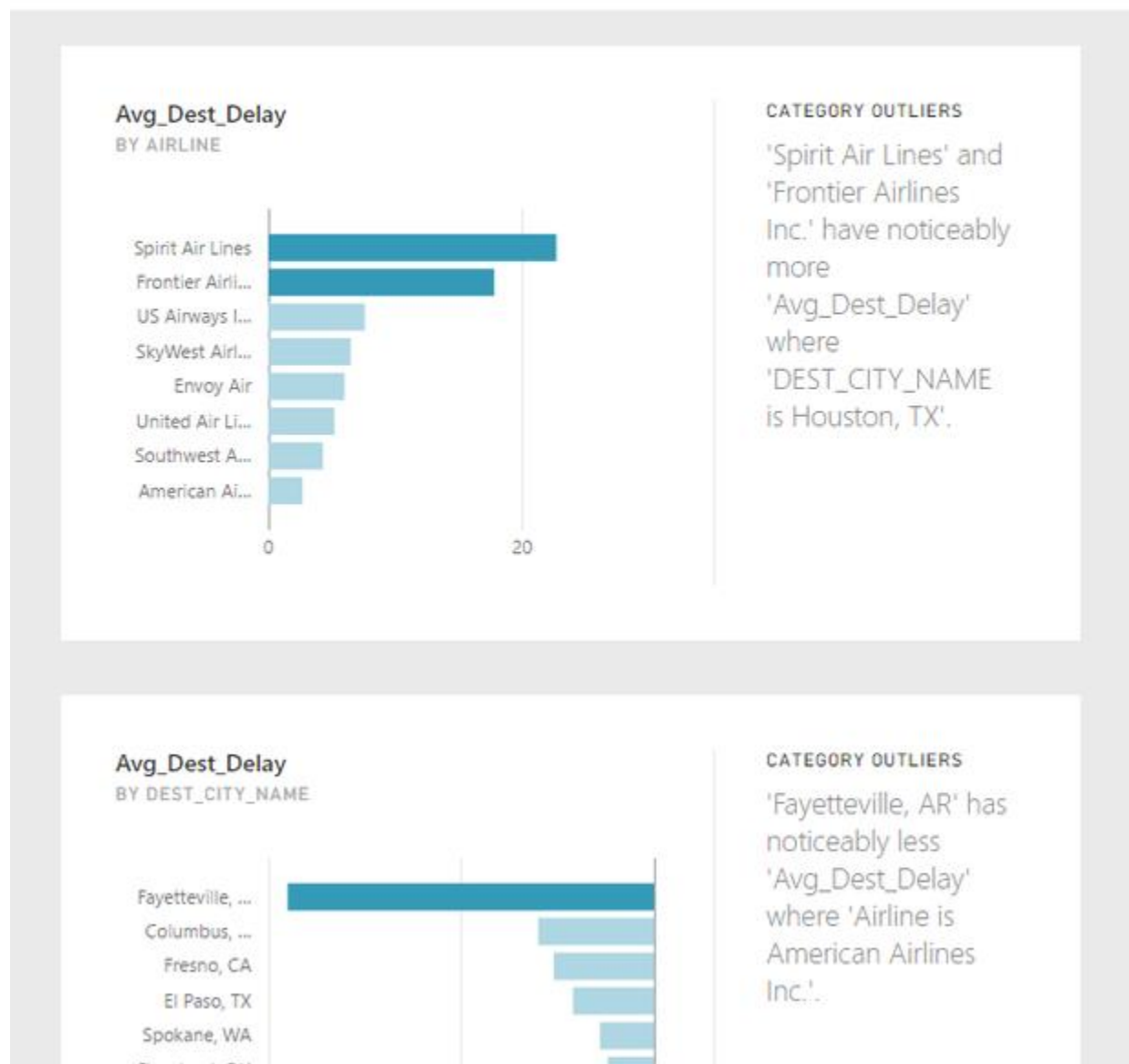
After you've done some exploratory visualizations, you might want to share some of your analyses with co-workers. To do this, you first need to publish your data and report to the Power BI cloud service. Save your work, then go to Home > Publish. You'll need to sign into your Power BI account if you're not currently logged in (or create one if you don't yet have one).

After data is published to the Power BI cloud service, you'll be offered the chance to "Get Quick Insights" from your data. This is Microsoft's automated look through various data points in order to [highlight things like outliers and correlations](#). Unfortunately, Quick Insights doesn't know that it shouldn't add together all the delay times -- it needs to average them to be meaningful. (I expect the default would work a lot better for files of data like sales and profits, where sums by region or store over time could make sense.)

I tried out Quick Insights by duplicating my project (I did a Save As to another name), and then grouping data by destination city, airline and flight. (If you want to follow along, go to Edit Queries > Transform > Group By and follow the dialog box to create a new column of average arrival delays, select Close & Apply, re-do your visualizations, then publish).

## Quick Insights for summer15grouped

A subset of your data was analyzed and the following insights were found. [Learn more](#)



Quick Insights attempts to call out interesting aspects of your data. In order for this to work properly, however, your data needs to be in a format that it can understand.

To get Quick Insights on a data set within the Power BI cloud service, click on the ellipsis next to its name in the left-hand menu (click the three-line menu icon at the top left to show the menu if it's not visible) and choose Quick Insights. (Can't find Quick Insights? Make sure you're doing this under Datasets and not under Reports.)

After looking at Quick Insights, go the main Power BI cloud page, sign in if you need to, click the three-horizontal-line icon at the top left to view the menu (if necessary), and you should see all your Datasets, Reports and Dashboards. If you already created visualization report pages in the desktop software before publishing to the service, they'll show up under Reports (your data, not surprisingly, shows up under Datasets).

You can share a report directly to the web for public viewing, which then also allows you to embed it somewhere on your own website: Go to File > Publish to Web. Be advised that although this was free while in beta, I haven't seen Microsoft guarantee that it will stay free since it moved out of preview in early July.

If this is private data you want to share with just a few other people, you'll need to share your work as a dashboard instead. If you'd like to share an entire report page as a dashboard, click the pin (thumbtack) icon just above the page and choose to add it to a new dashboard. You can also mix and match visualizations from different pages -- or even different data sets -- onto a dashboard by clicking on an individual visualization and then clicking on the pin icon at the top right of that visualization and pinning it to an existing or new dashboard. You can then click the share icon at the top right and share with others by entering their email addresses.

The dialog box is pretty self-explanatory, letting you decide whether recipients can re-share or just view. (As of this writing, row-level security is available for some data sets, allowing only certain users to see more sensitive information in a data set, but some of the features are in preview and are likely to change.)

You can embed web content, images, text boxes and videos into a dashboard from the dashboard itself; you can also create additional visualizations in a report and pin them to a dashboard, including adding "cards" that call out one or a couple of key metrics. You can edit your report in Power BI service by clicking the Edit Report link above the page and then pinning a new visualization to any dashboard.

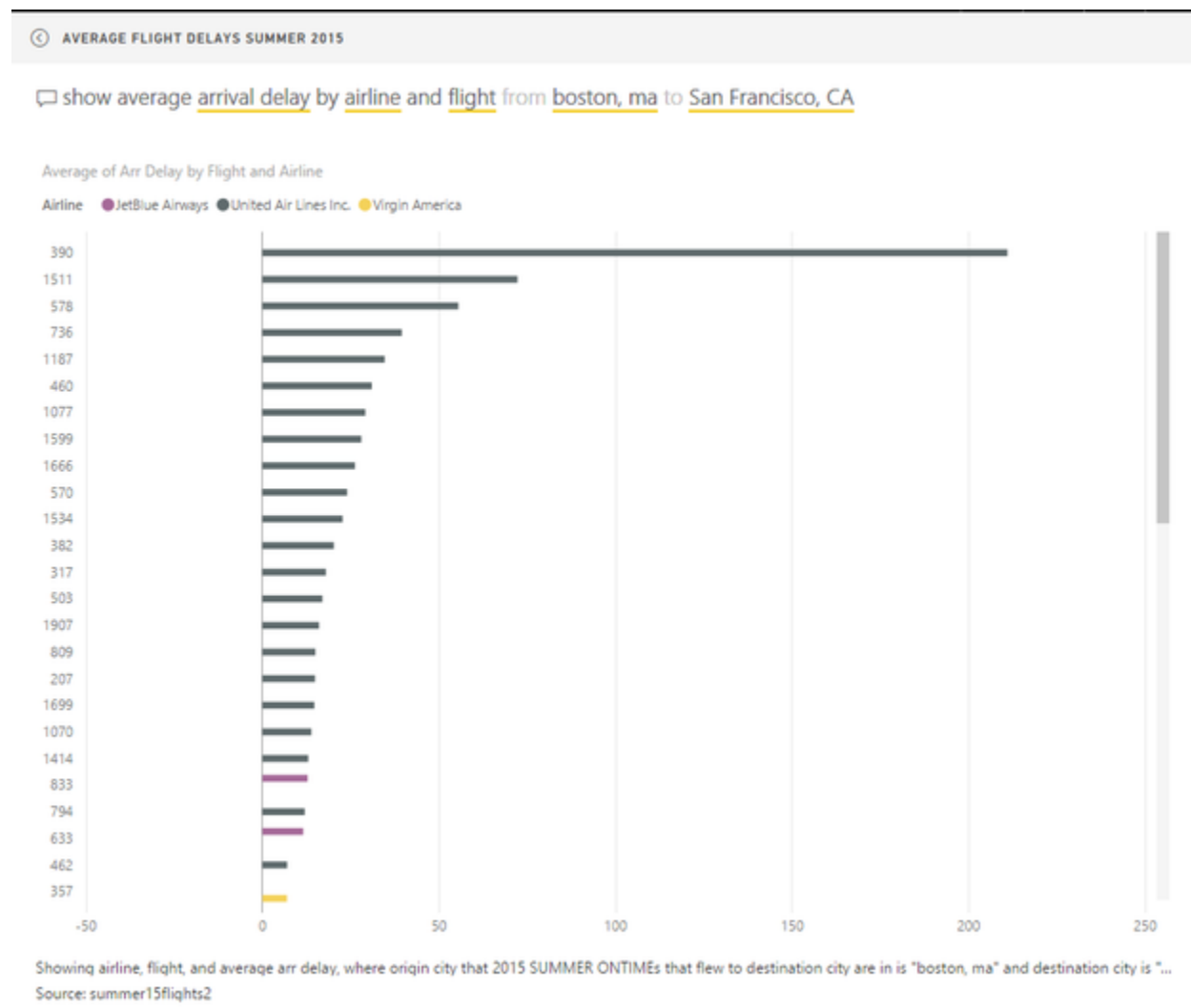
You can allow your users to ask natural-language questions by clicking the setting gear icon on the top (black) nav bar, clicking Dashboards and then the specific dashboard you want to edit. Click the checkbox next to "Show the Q&A search box on this dashboard."

This feature lets you (or your dashboard viewers) generate new visualizations from a data set by typing in questions. Once I figured out appropriate formats for my queries, I found this a pretty compelling interface. For example, I imported data for monthly search visits to IDG's U.S. websites and then was able to view various slices of the data by typing in things like "Show me search visits by month for *Computerworld*" or "What month has the largest search visits

for *CIO*'. And while I could have created the same interactivity with slicers on a report page that I pinned to a dashboard, the value-add here is that users can ask questions that I didn't necessarily think to model for them.

For the flight data, I queried my dashboard with questions such as "What origin city has the largest average security delay?" (Adak Island Alaska. Who knew?) and "What destination city has the largest average weather delay from Boston?"

As you type in something that Power BI recognizes as possibly coming from your data set, the word is underlined and you see choices for what to select. Do you want Boston in Origin City or Destination City? Click and pick. If visualization type -- such as column or bar -- isn't the one you want, you can specify one in your query.



Power BI dashboards offer natural-language queries, underlining column names in your data sets.

One quibble: Trying to pull data for a city like San Francisco or New York didn't work for me because the system didn't start recognizing the city name until I typed the fourth letter; three letters and a space didn't work so well. Fortunately, I could also use SFO (for San Francisco) or JFK and LGA (for New York) from the Origin airport code -- but that's something to keep in mind if you want to use this feature and have cities in your data.

If your data doesn't have intuitive column names, you can add a "synonym" to a data model -- in Desktop only -- for alternate ways to refer to a column. Click on the Relationships icon in the left panel (it's the third icon showing three boxes), select the Modeling ribbon, and you should see a Synonyms icon and list of fields to the right. If you click in the text box under Origin City, for example, you can enter additional substitute names such as Leaving.

## Conclusion

In just a few weeks of working with Power BI, I found it to be a fairly intuitive platform for loading in data and quickly doing basic data exploration and visualization.

As of the Microsoft Data Insights Summit in March, Power BI was missing some functionality I consider basic -- text searching for slicers, conditional formatting for tables, calendar data-pickers as slicers -- but Microsoft has made progress since then on improving the platform. Conditional formatting for tables was added in the [May 31 desktop update](#), searchable slicers arrived in the [June 30 update](#) and Microsoft says [better date-range pickers are in the works](#).

The platform is being updated fairly frequently -- once a month for the Desktop, more often for the cloud service -- offering the promise of continued new features and functionality. And as an R user, I'm intrigued by the integration of R scripts within the software and Microsoft incorporating [R graphics in the Power BI service](#) (Pro account needed) -- although you definitely don't need to know or use R for Power BI.

If you already use Excel -- especially on Windows -- or are otherwise a Microsoft shop with corporate data stored in Microsoft infrastructure, I'd recommend downloading Power BI, signing up for a free account and taking Power BI for a test drive. If you don't use other Microsoft products, it may still be worthwhile to try the platform if you want to do some quick data exploration and analysis and don't currently have a favorite platform, since Power BI will pull in a lot of other data formats besides spreadsheets, SQL Server and Microsoft Azure.

While I was somewhat underwhelmed with the two online Content Pack data connections I tried (Adobe Analytics and Google Analytics), largely because they didn't cover long-term site content trends, that's not to say some of the others like Salesforce or MailChimp wouldn't work

for you. And Microsoft is developing what it calls enterprise-class ["solution templates"](#) -- there are a couple of offerings for Salesforce and Dynamics CRM, for example -- that are aimed to make it easier to gear up corporate analytics reporting. (Microsoft says that Content Pack publishers are typically responsible for what scenarios to focus on.)

Bottom line? Microsoft designed Power BI for business users to do self-service analysis, in some cases with the help of IT to set up gateways or internal "Content Packs" connecting to internal data. It's not as robust a platform as, say, [Tableau](#) -- but it also doesn't carry Tableau's robust price tag for private data analysis. (Tableau Public is free but has less functionality than the paid software and isn't designed for sharing private data, which is usually critical for enterprise users.)

If you are a full-time data analyst or commercial designer creating publication-quality graphics, this may not be for you. But if you don't need high-end customization and functionality while analyzing data and sharing your work, this could be an attractive alternative.

---

## **Data cleaning via scripts**

*If you program, you might be more comfortable preparing your data for analysis with a scripting language like Python or R; I know that I am. Obviously, you can prep some data with any scripting language, save a new file and then pull that resulting file into Power BI. But with R, you can run R scripts right from within Power BI, including refreshing data from those scripts right inside Power BI (the Desktop version).*

*To run a script, go to Get Data > Other > R Script and you'll get a dialog box to execute an R script. You don't actually have to copy and paste your entire script into the dialog box; just use the R source ("path/to/filename.R") function, making sure you use the full path to the script. Any data frames that are creating by your script will then appear for possible import into Power BI.*

*Power BI also supports a couple of Microsoft scripting languages: M (the Power Query Formula Language) and DAX (Data Analysis Expressions).*

*M is the language that underlies querying and mashing up data. What you do in the Query Editor has M underneath. If you see automatically generated scripts in the formula bar within the Query Editor, that's M.*

*Once you've loaded your data from the Query Editor into the Power BI report view for analysis and visualization, the scripting language is DAX. It's designed so Excel power users will be comfortable, and has similar syntax and functions to writing Excel formulas. If you create a new column in your report view, the formula bar there is DAX. If you want to do a lot of complex data work within Power BI, at some point you may be interested in learning DAX.*

*Like Excel formulas, though, DAX is not as robust as a scripting language like R. For example, its SUBSTITUTE function will just replace one exact text string with another and does not use regular expressions that can match a pattern such as "find the letter A followed by two or more numbers, with or without a space in between."*

*(When asked about regular expression support, Microsoft responded that Power BI was designed for business users, most of whom are unfamiliar with regular expressions. The company believes that a fair amount text manipulation can be accomplished using Power BI queries.)*

## **Scripting resources**

### **DAX**

[Introduction to DAX](#) -- a Microsoft Guided Learning tutorial with both text and videos.

[DAX 101](#) -- a video recording of a two-hour intro session presented at Microsoft Data Insights in March 2016.

### **M**

[Introduction to Power Query \(informally known as "M"\) Formula Language](#) -- a quick overview and list of functions for using M to build data queries.

### **R**

[Learn to use R: Your hands-on guide](#) -- Computerworld's free PDF download for getting started with R, including some simple data visualization techniques. Already know R basics? Download our [Advanced Beginner's Guide to R](#).

[R Language](#) -- The place to go for all Computerworld's coverage of R.

---

*Sharon Machlis is Executive Editor, Online & Data Analytics at Computerworld, where she works on data journalism projects and in-house editor tools in addition to writing and editing.*