Serials Review

ISSN: 0098-7913 (Print) 1879-095X (Online) Journal homepage: https://www.tandfonline.com/loi/usrv20

Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data

Louis T. Becker(Contributor) & Elyssa M. Gould (Column Editor)

To cite this article: Louis T. Becker(Contributor) & Elyssa M. Gould (Column Editor) (2019): Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data, Serials Review, DOI: 10.1080/00987913.2019.1644891

To link to this article: https://doi.org/10.1080/00987913.2019.1644891

Published online: 29 Jul 2019.

SERIALS REVIEW
https://doi.org/10.1080/00987913.2019.1644891

Routledge
Taylor & Francis Group

SHARPEST TOOL IN THE SHED: ELYSSA M. GOULD, COLUMN EDITOR

Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data

Louis T. Becker, Contributor

University of Tennessee Libraries, Knoxville, Tennessee, USA

ABSTRACT
This segment of the Sharpest Tool in the Shed column introduces Microsoft's Power BI software and associated functionality built into recent (2013 and newer) versions of Microsoft's Excel. Librarians in technical services and other areas can use Power BI to combine, analyze, visualize, and share data from the wide variety of data sources encountered in library operations. As the benefits of visualization in the era of "big data" are widely discussed in the library literature, and Excel is nearly ubiquitous in office environments of all types, Power BI offers a user-friendly—but possibly underutilized—way for librarians to expand their grasp of the data around them.

KEYWORDS
data manipulation; data visualization; Excel; Power BI; software

Introduction

In recent years, data visualization has been a staple topic of discussion in libraries and in the broader world of business and journalism. Phil Simon's book, *The Visual Organization*, laid out case studies on the value of data made accessible and visual to organizational end users and the trend toward "self-service" in business intelligence and data analysis (Simon, 2014, p. 154). Ann Emery compiled a list of nine separate major conferences focused entirely on data visualization this year (Emery, 2019). Within our field, EBSCO's *Library and Information Science Source* indexes 163 articles with the subject "Information Visualization" published since January 2015.

Libraries are ideal consumers of this trend. They have information-savvy, technically competent employees and a lot of data. Modern integrated library systems (ILS) have included tools for business intelligence-style analysis and visualization of collection and acquisitions data for several years (Breeding, 2013). Despite these all-in-one solutions, many activities are still tracked in disparate systems; at this author's library, these systems include ILLiad (https://www.oclc.org/illiad.html), SpringShare LibApps (https://www.springshare.com/), Google Analytics (https://analytics.google.com/analytics), and a few more manual tabulations alongside the ILS. Even if each of these platforms provides useful, modern reporting solutions, the data remain siloed. Vendors and librarians end up distributing information in spreadsheets, and even if solutions existed to permanently mesh all of these systems, it is doubtful that any library could afford the disruption necessary to implement them. In a world of disparate data sources and widely differing reporting needs, simplicity, flexibility, and wide distribution are key features. Power BI, and the related Power Query tools within Excel, provide advanced analysis capabilities along with these key features.

Much of the library literature on visualization has focused on Tableau (https://www.tableau.com) and a few other visualization-specific software packages. In 2015, Sarah Anne Murphy introduced Tableau in *College and Research Libraries News* by asking "Why visualize library data? Why invest in purchasing, installing, and learning data visualization software and tools? Why can't we just use Excel?" (2015, p. 482). Since then, while the library literature has seen numerous articles on the use of Tableau in various projects, Microsoft has added considerably to Excel's capabilities in this regard and developed Power BI from a set of Excel add-ons to a completely new software/service package. There's always more to learn, and Tableau remains a formidable solution for advanced visualizations, but Town and Thabtah

CONTACT Elyssa M. Gould egould1@utk.edu Electronic Resources Librarian, University of Tennessee Libraries, 1015 Volunteer Boulevard, Knoxville, TN 37996.

(2019) found Power BI to be broadly competitive with Tableau. Comparing the two programs for architects and designers, Carlisle (2018) emphasized the “flexible and forgiving” nature of Power BI’s data-cleaning tools (p. 258). With the addition of Power BI, it is possible that librarians can now accept Murphy’s recommendations regarding visual analysis and broadly “just use Excel.”

The Power BI model

Microsoft developed Power BI out of several “Power” add-ons for Excel beginning with the 2013 edition (Kline, 2014). The foundational layer is a tool called Power Query, which has been fully integrated with Excel as “Get and Transform” beginning in 2016 (Microsoft, n.d.a). Power Query provides functions for accessing tabular data in a wide variety of formats, ranging from large-scale databases to simple text files. These tables can then be reformatted and combined before being loaded to Excel as tables or as part of a “data model” that can be manipulated with pivot table and graphing tools. Other add-ons (Power Pivot and Power View) added capacity for mathematical calculations on tables without the spreadsheet grid reference system and for interactive visualizations based on those calculations. The pivot table system, together with basic charting functions, remains a part of recent Excel releases. The Power View system has become the visualization engine of Power BI. Power Query fully replaces the text import functions of Excel in 2016 and newer versions and allows significantly more control over data types and delimiters along with more advanced transformations.

Outside of these Excel tools, Power BI consists of three offerings. The first is a freely available standalone Windows application, Power BI Desktop (https://powerbi.microsoft.com/en-us/desktop/), which combines Power Query with a visualization and analysis layer. The second is the Power BI cloud service (https://app.powerbi.com). The cloud service can be used without the desktop software for basic analysis and visualization. Custom calculations beyond standard descriptive statistics require the desktop version. Reports created with the desktop version can be uploaded to the cloud platform for sharing. Visualizations can be published to the open web (for embedding on a public web page) for free. If reports on the cloud service need to be shared within an organization (for collaboration or because the data involved is sensitive), both the creator and any collaborators need to be “Pro” users, at a cost of $9.99 monthly, reduced to $3 for nonprofits (Microsoft, n.d.b). This compares favorably with Tableau’s current pricing of $70 per user per month for report designers and $12–$15 per user per month for viewers of content that is not shared publicly (Tableau Software, 2019).

At the author’s institution, the per-user cost has been avoided through the third offering, an on-premises Report Server that allows sharing of desktop-created reports through an intranet interface with locally managed permissions. The Report Server does not provide all the features of the cloud service but does allow for the export of the data behind individual visualizations, which is not permitted from the open-web publications. The underlying desktop files can also be distributed from the server for individual work. Report Server licenses are sold as part of the “Premium” cloud service or with certain institutional licenses of Microsoft SQL Server (Sparkman et al., 2018). Purchase of a Report Server is likely to be an institution-wide decision rather than a library initiative.

Power Query and the M language: Data retrieval, cleaning, and shaping

Power Query, also known as Get and Transform, is presented as a graphical user interface (GUI) window that opens when summoned from Excel or Power BI Desktop but is actually a front end for a scripting language referred to as “M.” Whether by pointing and clicking within the GUI or by writing a text file, users compile a series of steps—called a query—that are interpreted to assemble a table of data to be loaded to Excel or to Power BI’s visualization tool. Each step is labeled with a name and consists of a function done to another named step or to a previously created query. Functions exist to open a wide range of data files, including databases, Excel, HTML web tables, XML API results, or delimited text files. If there’s a way to make it look like a table, M can find a way to read it. Adjusting for nonstandard delimiters or sectioning text into columns is easy with GUI prompts. The resulting tables can then be filtered and sectioned to focus on the data needed for the specific project. Because the data import process is scripted, users can step through the various transformations to troubleshoot errors or retrieve earlier phases of the transformation. Users who have established data-cleaning protocols using the R statistical computing language (https://www.r-project.org/) can integrate those scripts into M steps. The M interpreter will simply call the user’s R instance, and data will flow seamlessly through both systems. The data transformation process is completely nondestructive; all of the original data remain in the referenced files or databases, and only the extracts are actually imported to the resulting Excel or Power BI file. Well-organized documentation of the M language is provided by Duncan and Kinsman (2018).

A few features of the M system have particularly increased the speed and consistency of data analysis projects at the University of Tennessee. In addition to pulling information from external files, the Excel implementation of Power Query can access data from the workbook in which the query is running. These data can then be fed into further queries. To compare serial renewal lists sent by vendors with data from the ILS, a query was written to pull data from ILS exports and from a vendor return form and to return the lines that differed in certain key respects. The resulting table displayed the key columns from each file as a single table. Sections of the queries that referred to individual file names were then changed to variables; a subquery defined those variables as cells in a worksheet within the comparator workbook. To repeat the analysis, staff can simply enter the file paths to the two files to be compared in the labeled cells and trigger a refresh of the data table. Updated results load automatically, and the files can be compared without opening either of the source tables individually.

In addition to easily accepting parameters, M queries can be transformed into custom functions. This allows the same transformations to be made to all files in a list, which can easily be retrieved from a given Windows folder or SharePoint list. A single function written to transform an Excel COUNTER report from an individual platform's wide-format pivot report to a long-format table can be applied to a directory storing a year's worth of archived reports. As well as quickly creating a summary of e-resources use for annual statistics, the combined table allows comparison of use across platforms and a clearer picture of the demand for resources that exist in several different formats. Once created, functions like this can be easily copied to other Excel files or to Power BI Desktop for visualization; M code is not limited to a specific Excel file but can be modularized and reused as needed for different projects and reports.

Visualization and analysis

Once data have been shaped and configured by Power Query, tables can be loaded to the Power BI visualization layer. Multiple tables can be used in a given set of visualizations and can be linked together in a "data model" similar to those found in traditional relational database systems. It is not necessary to put all data into a single table before beginning exploratory analysis and visualization. A Power BI report contains one or more pages on which one or more visuals can be grouped. Assembling a visualization is a drag-and-drop process; the tables and fields of available data are listed in a side pane, and a second pane displays a selector for the form of the visualization and blanks for the data. Fields of data can be moved easily between axis labels, legends, and area values without redefining the underlying table.

The visualizations provided by Power BI are not terribly groundbreaking in form; bar, line, and area charts are provided in a number of forms, and scatter plots, treemaps, and map-based displays are also available. Dashboard-type presentations are enabled by tables, pivot matrices, and a variety of indicator cards to spotlight a single figure. Fonts, colors, and other presentation settings can be easily customized, although the defaults come closer to contemporary design practice than Excel's standard graphs. For projects with advanced statistical or graphic design needs, Power BI provides "R visuals" that export selected fields to run any R script the user might specify. Users can extend the basic graphic and statistical capabilities of Power BI while drawing on their cleaned data store and presenting results in a unified report (Iseminger, Blythe, Casey, & Hu, 2018). The real value in Power BI visualizations comes less from visual artistry than from their ease of construction and interactivity. A table can be transformed into a graph or a graph transmuted from bar to line or another form at the click of a button. When exploring data, it can be revelatory to quickly extract a table from the larger data store and then test different ways to visualize it.

Once established, the visualizations on any report page can interact with each other. In the visual version of the combined COUNTER report, a bar graph showed the total number of downloads in each month of the time period covered. Below this simple time line, a table showed the individual journal titles with their total number of downloads. Clicking on a single month in the bar graph forced the table to show only the journals used in the selected month, and the number of downloads was automatically adjusted to show only the downloads in that month. Selecting a single journal title in the table, on the other hand, modified the time line graph to show the use of that journal over time. Additional graphs could be added to such a report to section usage by vendor, subject classification, and so on. A major challenge of report design is determining which aspects a user will want to compare and interact with without running out of space or making each graph too small.

In addition to filtering based on side-by-side visualizations, any particular graph can be filtered by the particular data in the graph or by other data fields brought into a "Filters" pane to one side of the report. Similar filters can also be made available to adjust the content of an entire page or a multipage report. On the page level, drop-down and checkbox filters can also be added to the visual area; these can improve efficiency for users who want a "quick check" tool and easy awareness of how the data are being filtered.

The measures available for visualizations are standard descriptive statistics—sum totals, counts, measures of central tendency, minimums, and maximums. These measures can be applied to an entire column of data or to the portion of the data specified by the user's configuration of visuals and filters. For more advanced calculations, PowerBI provides Data Analysis Expressions (DAX) (Microsoft, n.d.c). The DAX system is similar to Excel's formulas, adapted for tabular data and dynamic filtering. Rather than specifying a range of cells on which to perform a calculation, DAX formulas specify a table or column to be manipulated. This allows for tables to grow in length through updates. If another month of data is added to the files underlying our example COUNTER report, the totals and other comparative measures will take that into account when the report is updated, and the formulas do not have to be rewritten. Along with the ability to handle growing data sets, DAX's use of column names rather than grid coordinates improves readability. Row-level specificity is added through filter statements in the formula. For example, our COUNTER report can explicitly count the titles with no usage or compare the average usage of a specific vendor's titles to the average overall, without necessarily including all vendor names as an axis on the graph. DAX measures allow Power BI to go beyond exploratory visualizations to extensive analysis.

Conclusion

Power BI offers a set of tools that bridge gaps—between databases and spreadsheets, between spreadsheets and statistical tools, between knowledge workers and the piles of data they would all like to find stories in to share with colleagues, administrators, and the public. There are numerous tools that are purpose built for data cleaning, for statistics, and for interactive graphics. Some of them may be better at their precise tasks. Power BI brings these functions together in a user-friendly package with Microsoft's industrial-sized support infrastructure and widespread user base. Beyond the documentation already referenced here, there are numerous blogs, tutorials, and books to help users learn the interfaces and languages involved. Feldmann (2019) provides a useful list of these resources. The different components of Power BI may seem complex in the abstract, but the graphical interfaces, built on widely used tools, allow for easy visualization of data and of the processes needed to connect our different data stores together. As users gain an initial understanding of their data, the combination of interactive and scripted tools will let them extend their analysis without abandoning what has already been built. Librarians have been building on existing knowledge, plans, and data stores for generations. Power BI is a tool that follows that pattern and works with it.

ORCID

Louis T. Becker http://orcid.org/0000-0002-0027-7953

References

Breeding, M. (2013). Mining data for library decision support. *Computers in Libraries*, *33*(5), 23–25.

Carlisle, S. (2018). Software: Tableau and Microsoft Power BI. *Technology|Architecture + Design*, *2*(2), 256–259. https://doi.org/10.1080/24751448.2018.1497381

Duncan, O., & Kinsman, M. (2018, October 17). *Power Query M reference [Web site]*. Retrieved from https://docs.microsoft.com/en-us/powerquery-m/power-query-m-reference

Emery, A. (2019, January 22). *The complete listing of data visualization conferences [blog post]*. Retrieved from https://depictdatastudio.com/complete-listing-data-visualization-conferences/

Feldmann, I. (2019). *The BIccountant: Learning resources [web page]*. Retrieved from https://www.thebiccountant.com/learning-resources/

Iseminger, D., Blythe, M., Casey, L., & Hu, J. (2018, November 27). *Create Power BI visuals using R [web page]*. Retrieved from https://docs.microsoft.com/en-us/power-bi/desktop-r-visuals

Kline, K. (2014, August-September). PowerBI: The self-service BI app that can change your life. *Database Trends & Applications*, *28*(4), 39.

Microsoft. (n.d.a). *Get & transform in Excel [web page]*. Retrieved from https://support.office.com/en-us/article/get-transform-in-excel-881c63c6-37c5-4ca2-b616-59e18d75b4de

Microsoft. (n.d.b). *Power BI pricing: Analytics for every organization [web page]*. Retrieved from https://powerbi.microsoft.com/en-us/pricing/

Microsoft. (n.d.c). *Data Analysis Expressions (DAX) Reference [web site]*. Retrieved from https://docs.microsoft.com/en-us/dax/data-analysis-expressions-dax-reference

Murphy, S. A. (2015). How data visualization supports academic library assessment. *College & Research Libraries News*, *76*(9), 482–486. doi:10.5860/crln.76.9.9379

Simon, P. (2014). *The visual organization data visualization, big data, and the quest for better decisions (Wiley and SAS business series)*. Hoboken, NJ: John Wiley & Sons.

Sparkman, M., Saxton, A., Blythe, M., Hamilton, B., Clem, E., Duncan, O., & Ghanayem, M. (2018, November 19). *What is Power BI Report Server? [web page]*. Retrieved from https://docs.microsoft.com/en-us/power-bi/report-server/get-started

Tableau Software. (2019). *Pricing for data people [web page]*. Retrieved from https://www.tableau.com/pricing/teams-orgs

Town, P., & Thabtah, F. (2019). Data analytics tools: A user perspective. *Journal of Information & Knowledge Management*, *18*(1), 1950002. doi:10.1142/S0219649219500023