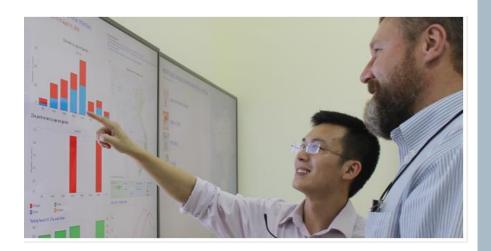
June 19, 2017



Health Data Visualization Selection Assistant

Helping Data Users Make Better Choices

Prepared for PATH by the Global Health Security Partnership.

This report and the activities described therein were supported by Cooperative Agreement Number 1U2GGH001812-01, funded by the Centers for Disease Control and Prevention.

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

MAILING ADDRESS PO Box 900922 Seattle, WA 98109 USA

ADDRESS 2201 Westlake Avenue Suite 200 Seattle, WA, USA

TEL: 206.285.3500 **FAX:** 206.285.6619

www.path.org



Seeing Your Health Data Differently

In addition to becoming an integral part of business intelligence solutions, data visualization has become an important way for public health specialists to interpret and share health information quickly and efficiently. Presenting and visualizing health information can often change the way public health officials make decisions, and the way they think about their most pressing public health issues.

But how do countries and data users sort through the growing number of software platforms and tools available for data visualization and workflows? What tools are affordable and fit within users' budgets? And how do countries ensure users have the knowledge and capacity to use the various tools?

Many experts have some prior exposure to data visualization technologies, but as software options grow, staying up-to-speed can be a challenge. It has become a crowded world out there, and often the tool used to store/parse health data might not be the right tool to visualize it.

In order to help country stakeholders, donors, and public health experts choose the best data visualization platform that is suited for their needs, the PATH Global Health Security Partnership team has put together this report to help ministries of health, NGOs, and other users of health data. The report is divided into two separate sections: the first outlines and adapts some of the most important questions stakeholders ought to be asking as they decide how to make these choices for their data visualization needs. This section helps to better illustrate and formalize the data platform selection *process*. Next, we compile and compare some features of the most popular self-service automation and visual data discovery tools side-by-side-focusing on *products*. By addressing both *processes* and *products*, we hope users are able to deliberate and assess their available options with more information at hand. Additional resources, including useful websites and related reports, are appended as a conclusion.

1. Process

General advice for data users

It is difficult to provide any "generic" or universal guidelines for selecting a data visualization solution. The bottom line is that almost any software package will present both strengths and weaknesses, and should only be chosen based on the context in which the software solution will be utilized. With that said, it is important to note a few common lessons passed down from experts with years of assisting in software solutions.¹

• "Don't start with the tool"

Users and groups often jump straight to a software solution and then build a case to justify why they chose it. In many circumstances, the tool chosen might not be the right choice. Perhaps an extra tool for visualization is not even needed for the project. The process

ⁱ Credit goes to the experts and solutions **engineers at Novel-T (Philippe** Veltsos and Patrick Briand, http://www.novel-t.ch/) for their insights and input into this consultation process.

undertaken in support of selecting the appropriate software package is as important as the final product selected.

• Data quality comes from data use

Data that gets used, in most cases, is of better quality than data that does not. Let data use guide your thinking about your needs and requirements for business intelligence/data visualization solutions.

• Think about the "total cost of ownership"

The initial purchase price of a software package for your organization or team is almost certainly not the only cost to consider in the scope of your budget for data visualization. Beyond licenses, you can expect to incur expenses based on setup; maintenance; training; plug-ins; updates; and/or cloud hosting.

Recognize that you are buying into and building a relationship with a new suite of software, often with unexpected licensing restrictions or upkeep fees.

Does each machine need a license to use the software? Does the software require special licensing for cloud access? Unexpected costs can add up quickly.

• Open- $Source \neq Free$

Although open-source solutions can be much more cost-effective, it is important not to conflate "open-source" with "free" when considering data visualization tool options.

Open source options allow users and staff to "peek under the hood" and recalibrate features as needed in the source code. This means if you have skilled users, they can modify the software to the needs of your projects, rather than waiting for a company to integrate your solutions into their proprietary source code.

Be cognizant of your project needs, and the complexity involved in tinkering with your data platforms. Open-source solutions can take more resources and time to set up and get started, but can often be more tailored solutions fit-to-purpose once integrated into your workflow.

Business intelligence/data visualization tools sit on top of data warehouses.

Data visualization tools like Tableau and PowerBI are meant to be plugged into functioning and sustainable data warehouses. If you do not have a good data analysis plan to support your business intelligence or data visualization needs, you may run into problems down the road. *Good visualization outputs only come from high quality data inputs.*

• What questions will your software solution answer?

What kinds of questions will your data visualization packages need to answer? Are these simple questions, or very complicated ones?

Often, simple questions can be addressed with very simple tools. Sometimes Excel is enough to visualize and help answer simple, but important questions such as, "How are we doing over time?"

Think about resource constraints. Do you need to buy a full package with all the "bells and whistles," potentially at great cost to your budget? Or will simpler, targeted solutions suffice?

Approaching Your Choices: General Considerations for Selecting A Data Visualization Platform

Based on research conducted by Engine Room's Data Tool Selection Assistant Projectⁱⁱ, there are generally three important steps end-users should address before selecting a data visualization platform:

- 1. The **goal** or **problem** you want the tool to address.
- 2. The **interests** and **needs** of the people you want to use the tool.
- 3. The **tools or platforms** that are available within your constraints.

The following is a list of questions to guide end-users (individuals, groups, or organizations) further as they articulate their goals, needs, and requirements in a data visualization package. The answers to these questions should be developed collaboratively with the individual(s) and team(s) who will be using both the data visualization package, *as well as its outputs*.

- Project Objective:
 - What is your project's objective? What do you want your data visualizations to accomplish?
 - How could a data visualization tool help you achieve these goals and objectives?
 - What kind of decisions will data visualization support?
- Needs and Requirements:
 - What things must the tool be able to do? Are there specific platforms that must integrate with the tool?
 - What kinds of data must the tool be able to interpret? How will this data be sourced/compiled?
 - How will the visualizations be used, and who will need to use/see them?
- Description of Users:
 - Who do you expect to use the tool? What do you know about them?
 - Are these users mainly internal or external to your organization, or both?
 - Will users need remote access to the platform and/or the visualizations that it produces?

4

ii https://toolselect.theengineroom.org/

- What skills and experience do your users already have? Are they comfortable with programming?
- How familiar are your users with data visualization tools? How much experience do they have with data management or data analysis (# of months/years)?
- What level of access will users have or need to produce strong visualizations? Will users be able to regularly access the data needed for visualization work?
- What tools/products do your users already use?
- How much time/training will your users/staff be able to dedicate to learning the platform?

Potential Obstacles/Challenges

- What budget is available for purchasing/licensing a tool? How much money will be available to maintain and update the tool? How long can that funding be sustained?
- What might prevent users from using the tool? (e.g., cost, hardware/software requirements, connectivity, maintenance integration issues, pre-existing workflows/solutions, lack of skills, etc.)

2. Products

Please note: The following table is a partial list of commonly-used enterprise and open-source data visualization and management platforms. In no way should this list be considered exhaustive. For more examples, please refer to the following section on "additional resources."

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
VENDOR	Tableau Foundation	Microsoft	Microsoft	University of Oslo (HISP)	R and R-Studio	Mathworks
WEBSITE	www.tableau.co m	https://powerbi.micro soft.com/	https://products .office.com/en- us/excel	http://www.dhis 2.org	https://cran.r- project.org/	https://www.mathw orks.com/products/ matlab.html
YEAR LAUNCHED and # of USERS	2003; Over 35,000 customer accounts (organizations and individuals)	2014; Over 5 million subscribers	1987; Over 750 million users worldwide	1996; used in over 47 countries	mid-1980s; about 1– 2 million users	Mathworks founded in 1984; about 1 million users worldwide
WEB-BASED (Sharable Online) vs. RICH CLIENT SYSTEM (Requiring Local Software Install)	Both web-based (Tableau Server) and client-based systems (Tableau Desktop) available and well integrated.	Both web-based (PowerBI Embedded) and client-based systems (PowerBI Desktop). Sharing also available over mobile client.	Client-side only (requires desktop installation)	Both web/mobile based and client-based systems available and integrated.	Client-side by default, can be installed as a server but requires extensive programming	Client-side (requiring desktop installation) by default, but options available to export calculations and operations to a "Distributed Computing Server" without extra licensing fees (for computationally- intensive jobs).

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
PRICE/ LICENSING	Licensed product, requires purchase: personal or professional editions, ranging from \$1,000–2,000 USD/license. Additional cost for Tableau Server licenses. Low-cost options negotiable with LMICs. Free desktop license available for students/educatio nal institutions.	Basic package is free for users and organizations with Office365 license. "Power User" option requires subscription and is \$9.99 USD per user per month. MS notes, "Academic, Government and Non-profit pricing options are available." Subscribers can have a mix of free and power users based on needs.	Included as part of the standard MS Office365 licensing suite. Various pricing options for businesses, non-profits, governments, or educational institutions.	Open-source platform (free- to-install).	Open-source platform (free-to- install).	Enterprise solution, with various licensing/pricing options available for users and groups (Business/Home/Stu dent/Education). Individual license starts at \$2,150 USD. Quotes can be requested from Mathworks to option multiple "flex user" networked accounts. Reports indicate that each node in a network requires its own licensing agreement, so MatLab can often become very expensive for large- group userbases.
QUALITY	Tableau is a well-tested, reliable, and mature software package. It is used across businesses worldwide and does not require	PowerBI is newer to the market than Tableau, but is updated frequently with features requested by its userbase in order to become competitive. It is a secure	Provides basic analytics and graphics. Strong and standardized option to build data models, but very limited options	Provides basics for analytics and BI in single application (pivot tables, charts/graphs, data prep and integration), but not intended to	Generally reliable but packages are developed by the community and all packages need to be vetted/verified. "GGPlot2" is a widely used and liked visualization	Integrated "MatPlot" system is well-liked, but fewer community- packages available than in similar R- Studio. Output graphs and plots are similar in
	coding or	platform, with	to conduct data	replace/compete	package for R, with	complexity and

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
	programming skills to use. It is a secure platform with user permissions, credentials, and publication controls.	features resembling Tableau.	analyses within single interface.	with Tableau or PowerBI for advanced reporting or analyses.	lots of supportive documentation (http://ggplot2.org/)	quality to those from R-Studio.
COMPLEXITY (Pros)	Easy to use, end- user focused, highly interactive. Drag & drop tools and pop-up dialogues facilitate quick learning and powerful options. Does not require specific programming language or skills to execute complex functions.	Interface will be immediately familiar with Office365 users. Easy to use features laid out intuitively. Strong integration with Excel features and options. Product is evolving with new features added continuously with customer feedback and demand. Dragand-drop options, and simple outputs based on contextual needs are available.	Fairly easy-to- use for basic functions. Capability to produce complex analysis and visualization with supplemental tools and third- party applications. Used and understood widely.	Basic visualization functionality within- application. Health-specific platform that links warehouse with simple visualization options.	Very flexible; does not require any reliance on prefabricated or preset functions. Highly customizable.	Matlab, like R- Studio, requires specific knowledge of its data-language to operate. It has a robust programming environment, well- suited to developing new mathematical algorithms to parse large data-sets. This makes it very popular for machine-learning projects, and engineering applications.

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
COMPLEX (Cons)	Some advanced functions require specific training or familiarity with tools/programming language. DHIS2 integration requires third-party application.	Data vizualizations limited to 3,500 data points. Some basic operations require DAX programming/formul a skills. Limited trend lines, no forecasting capabilities.	Limited graphical user interface (GUI). Advanced visualizations require programming and skilled use of external tools and suites not integrated into the primary interface (PowerPivot, Power Query, PowerMap, Silverlight, HTML5, etc.). Complex analyses can limit options for sharing outputs to other users or platforms.	Advanced visualization functions require coding (e.g., in R), or Google Viz APIs.	Very steep learning curve; around a year of training and experience needed until user will be proficient. Workflow is based almost entirely in programming language, no 'drag & drop' functions like Tableau/PowerBI.	Like R and R- Studio, MatLab has a very steep learning curve, and requires significant experience and expertise to learn and understand its programming language. This makes exploratory data visualization more difficult than in other applications, such as PowerBI, Tableau, or even Excel. R- Studio is widely the preferred programming environment for data visualizations, but MatLab can be more powerful if the user is sufficiently advanced in its language/syntax.

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
DATA INTEGRATION/ CONNECTIONS	Tableau supports many options for data connectivity, and can import almost any file-type from stat programs (R, SPSS, SAS, Excel, etc.). A personal license allows for 7 data file sources, from Excel to SAS (see below). A professional license includes "hundreds" of database and file source options for importing. There is a third-party option to link DHIS2 data with Tableau directly. Tableau operates seamlessly in the cloud with Tableau Online, Server, and Mobile.	Well-integrated with MS Office suite, particularly Excel (e.g. PowerPivot tables). Statistical file importing options limited compared to Tableau. Other database and cloud connections available. PowerBI is integrated across Android mobile devices and the cloud.	Excel is the standard for producing, organizing, and manipulating data sets of all sizes and shapes. Almost all programs accept its file formats. Compatible with OpenOffice (license-free) and other dataediting suites. Extremely limited connectivity to external data sources, databases, or import capabilities compared to PowerBI/Table au. Does not allow integration of other non-Office products.	Standard importing options from Excel and statistical programs. Limited options for cloud or other database connectivity. Third-party applications available to connect datasets with Tableau for further analysis and visualization options. Data is self-hosted, with limited/no cloud-based options available.	Can integrate with almost any other software language or program environment. Database connections will require specific command-line programming or SQL knowledge to create relational connections to databases. Optional third-party plug-ins are available to minimize coding/programming required, but these may cost additional licensing or purchasing fees.	Matlab has multiple options for importing and database connection, including SQL queries, command line functions, and a Database Explorer application. The data source setup will differ depending on the database drivers that you are using for connection. Once data sources are available, users can connect to their databases, modify and delete database connections, and work with multiple databases at once.

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
DATA BLENDING/ MANIPULATION	Tableau intelligently determines the relationships between multiple imported data sources. Data links can be modified and created manually based on user's preferences.	PowerBI automatically determines the relationships between multiple imported data sources. Like Tableau, data links can be modified based on user needs.	Heavily restricted and limited options for importing, blending, and combining data from multiple sources/databa ses. Often requires third- party solutions or manual re- entry.	$(N\!/\!A)$	Data blending is possible with other programs, but this requires additionally extensive programming knowledge and experience. R can export to almost any other data visualization package, including PowerBI and Excel.	Data blending and manipulation are features embedded at the core of Matlab, but designed for large scientific datasets. Matlab is not as intuitive and straightforward as its competitors are when it comes to data manipulation.
DATA SOURCES	- Excel spreadsheets - Cloud Services -Microsoft Access - Multiple text file formats - Statistical files - Database connectors	- Excel spreadsheets - Cloud services (Salesforce, Google Docs and others) - Streaming data - On-premises databases - Power BI Desktop files For more info: https://powerbi.micro soft.com/en- us/documentation/po wer-bi-desktop-data- sources/	Text files	- XML - CSV - PDF - DHIS2 database: pre- built data entry form collected through SMS, plain HTML and Java for feature phones, and a high-end Web-based solution with offline support for smartphones	Almost any kind of file or list type available.	Extensive import/export of data files supported.

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
DATA VISUALIZATION / OUTPUTS	Tableau's suite of tools produce dynamic charts, graphs, and reports that are highly customizable. Standard and expanded options for reports are available, including map views, as long as geographic fields exist in the data source.	Evolving options and features, but more limited capabilities to tweak or fine-tune outputs. Strong default options, but fewer possibilities for advanced or highly complex outputs. MS is focusing on expanding customizable options for visualization and reporting. Map views are available as long as geographic fields exist in the data source.	Excel can produce strong, highly customizable charts, graphs, and analyses, but requires very high level of expertise, skill and training. External tools, applications, and systems are needed to produce robust data visualizations and dashboards similar to Tableau/Power BI.	Basic outputs available, including: - GIS (web- based GIS features) - Charts (column, line, pie, stacked column and area) - Pivot tables - Dashboards	R Programming offers a satisfactory set of inbuilt function and libraries (such as ggplot2, leaflet, lattice) to build visualizations and present data. Visualization options are very broad and customizable, including web-hosted animations, but this requires extensive time-coding and additional training. R's "visualizations are great for exploratory analyses, but do not produce very high-quality graphs," according to some.	Mathworks has developed a number of modules for data visualization and outputs for Matlab. Even if they can be highly customized and provide real time updates, they are not designed to become a dashboard.

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
PRODUCT SUPPORT	The software is commonly used, and plenty of online and inperson training options are available (webinars, YouTube videos, hosted instruction, etc.). There is an extensive network of Alliance Partners with expertise available. Tableau offers product and customer support online and over the telephone for its users.	As PowerBI is fairly new to market, the userbase and expert communities are smaller and less prominent than with Tableau. Limited documentation/educa tional opportunities online or in-person. User community growing rapidly with new adopters. Guided learning options now available on Microsoft website: http://community.powerbi.com/https://powerbi.microsoft.com/en-us/guided-learning/	Excel is supported as part of the Office365 Suite, through Microsoft's standard customer service provisions. The userbase is almost unlimited, and online training is ubiquitous.	The University of Oslo's Health Information Systems Program (HISP), with a well-established community of practice and over 10 centers in various countries support DHIS2 globally. The University holds trainings often, and on request. Strong support and community.	No formal product support teams, but the developers are available to provide packages and preset formulas. Online communities for training, sharing, and support discussions available widely. See e.g., https://support.rstudio.com/hc/en-us	MatLab has extensive online learning options, including free tutorials, MOOCs, and other community guides. Although it takes time to learn the programming language, the extensive resources for user-support can assist newcomers to become proficient in the basics quickly.

	Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
SUMMARY & NOTES	Tableau is the "gold standard" for BI software. Its advantage is that it is powerful, robust, well integrated, and commonly used by companies and organizations worldwide. It can be expensive to license without special agreements with Tableau Foundation. It is available for both Mac and Windows. It has mobile clients available for Android and iOS. Tableau allows for a deep and advanced user-experience, with a comprehensive feature-set.	PowerBI offers a powerful, affordable alternative to Tableau that will be familiar and easy-to-learn with those comfortable in Office products. It is intended to simply, streamline, and replace Excel data analytics into a single package. It is tightly integrated with Office365, but its advanced features and automated functions are limited compared to Tableau. It has a growing user-base and recognition among business and institutional communities. It is only available as a desktop client on Windows PCs. It has a mobile client available for Android. Where Tableau allows users to "work with the data like clay,"	Excel is the generally accepted standard for data collection, entry, manipulation, or modeling. Most users agree that it is necessary but not sufficient to complete advanced data visualizations. PowerBI has been designed to take over most/many of Excel's visualization and analysis functions. Excel is available both for Windows and Mac. It has mobile clients available for Android and iOS, but these have very limited	DHIS2 is a widely-used health information system, most often for data warehousing. Its functionality for advanced analytics and visualization is limited, requiring export to third-party programs for expanded analysis. There are more established options to link DHIS2 with Tableau than with PowerBI at the time of writing.	R-Studio programming environment offers a comprehensive package of visualization options, ranging from basic to extremely advanced, with fully customizable outputs. However, it requires sufficient training in R-specific programming, external third-party packages, and longer time commitments than other "drag & drop" visualizers like Tableau/PowerBI. R- Studio offers a very powerful suite of options for savvy users, but requires long-term skill development and design focus to generate advanced outputs.	Matlab has been around for quite some time as a powerful data analytics suite, with more recent efforts to integrate visualization. It is primarily targeted toward engineering applications, including Big Datasets. As the website says, it is "optimized for solving engineering and scientific problems", not necessarily visualizing data. However, its strength is in its ability to combine a data manipulation/analys is back-end with a flexible data visualization front-end.

Tableau	MS PowerBI	MS Excel	DHIS2	R-Studio	Matlab
	PowerBI functions more "like a vending machine"—with presets available to organize and visualize your data.	visualization options.			

3. Additional Resources

The following is a list of additional online resources to help users explore more comparisons and discussions over the relative advantages of available data visualization platforms. They are listed in no particular order. **Please note:** all of the following websites and pages listed are external resources, and are not maintained, sponsored, or endorsed by PATH or any of its employees or staff. Use at your own risk.

Dataviz.Tools

http://dataviz.tools

(Highly recommended – selection of open-source/free data visualization options.)

A curated guide to the best tools, resources, and technologies for data visualization. A comprehensive and impressive collection of information on data visualization tools, with an emphasis on free/open-source options. As the author states, it is meant to bridge the gap between programmers/statisticians and the general public by only highlighting free/freemium, responsive and relatively simple-to-learn technologies for displaying both basic and complex, multivariate datasets. It leans heavily toward open-source software and plugins, rather than "enterprise, expensive B.I. solutions."

• Data Visualization for Development

Knowledge Gateway & Community https://knowledge-gateway.org/dataviz

(Great resources and a helpful listsery to sign up and receive data visualization updates and suggestions from experts.)

According to the website, "The Data Viz for Development Community was created as a space to share examples of great visualizations, resources, tools, and ideas for making information accessible. [Their] goal is to highlight simple visual principles to think about that can improve basic visualizations, basic tools for viz creation, and more advanced tools for visualizing information, in order to share appropriate resources for the data-curious to the data viz ninjas."

• Andrei Pandre's Data Visualization Comparison

Blogger's Data Viz Package Comparison https://apandre.wordpress.com/tools/comparison/

A similar document to this one, with a comprehensive comparison table looking at Tableau, PowerBI, Qlickview, and Spotfire. Andrei's comparisons are muddled and sometimes unclear, and the layout is less than ideal. But it is helpful to view his commentary as a supplement to some of the considerations listed in this product guide. Please see his "additional factors to consider" at the bottom of his blog post, in case you are looking for more distinguishing criteria in your software package decision.

• **Tool Selection Assistant** (Highly recommended, cited in this report.)

From 'the Engine Lab' at University of the Witwatersrand https://toolselect.theengineroom.org/

A generic blueprint to help organizations in the development world choose their data processing packages. Does not cover individual options or package pros/cons, but provides robust "six rules of thumb" to help guide decision-making process. Most questions from the "general considerations" section in this document are pulled from this resource. The website also features reports on data visualization integration in selected projects across the developing world, a useful insight into how to manage visualization solutions in various contexts.

• K4Health Data Visualization Crash Course

https://www.globalhealthlearning.org/course/data-visualization-brave-new-world

A great crash course for beginners who want to understand data viz basics.

PCMag.com 2016 Best Data Visualization Tool Roundup

http://www.pcmag.com/roundup/346417/the-best-data-visualization-tools

An excellent overview of data visualization requirements and options, including an overview for how to conduct a needs assessment and links to individual software package reviews. A helpful comparison between some of the platforms examined in this document.

• Qlik Data Suite

http://www.qlik.com/us/products

Qlik offers a suite of tools and platforms that combine strong data management with elegant data visualization options. Qlikview and Qliksense are strong competitors in the data management marketplace, but due to space constraints, these software packages were not analyzed in this report.

GIS Lounge Website

https://www.gislounge.com/

A great resource for all things GIS (geographic information systems), including overviews of available software platforms, learning resources, newsletters, maps, and upcoming conferences.

Ovum Decision Matrix: Selecting a Business Intelligence Solution, 2014–15

Ovum Consultancy Business Intelligence Analysis (*Of limited use, for reference purposes.*) https://www.sas.com/content/dam/SAS/en_us/doc/analystreport/ovum-decision-matrix-bi-105875.pdf (online PDF report).

An extremely comprehensive and detailed overview of BI options and data analytics packages. Although an excellent overview of the software field (as of 2014), the only specific mention of "visual discovery and self-service automation" is on page 12, with a minor chart that plots Tableau against IBM and Tibco (PowerBI is not mentioned in the document at all). Its application and usefulness for the public health field is questionable.