Anti-Corruption Report
An Acuris company

anti-corruption.com

December 2, 2020

DATA ANALYTICS

How Visual Analytics Can Fuel a Compliance Program

By Andy Miller, *Lextegrity*

Company leaders, whose experience traditionally has not included data-focused roles, have become conscious of the value that their company's data holds and, in turn, relentlessly curious about how their organization is capturing and extracting value from their data.

While profit centers have forged ahead with data analytics, functions that were historically less data-focused, such as compliance, legal and internal audit, have lagged even further behind. Adapting data analytics was more difficult for these functions because they have less control of their applications and data, and the breadth and quality of software that addressed their needs was limited. With the use of data to inform risk decisions now a firm expectation, these risk functions are racing to modernize their processes to utilize data to drive greater efficiency and effectiveness in their work.

Regardless of a company's path to greater data efficiency, there is still confusion around what to do once data is ready to be analyzed. It is often not clear what type of analysis will provide a compliance department with the most value. In my experience, a focus on providing data in a user-friendly, eye-catching way is the most efficient route to gaining insights from compliance data. Here we discuss how data visualizations can add value to a compliance program, and give three examples of the contexts in which a compliance program could effectively make use of visual analytics.

See "Continuous Spend Monitoring for End-to-End Third-Party Risk Management" (Dec. 11, 2019).

The Value of Visuals

A key goal of compliance departments is to monitor and mitigate risk to the organization. Monitoring requires constantly assessing key metrics and the data underlying those metrics, looking for anything that falls outside of expectations or is trending in the wrong direction. Then, the compliance team can gather appropriate context around the data trends and investigate to identify the root cause of an issue.

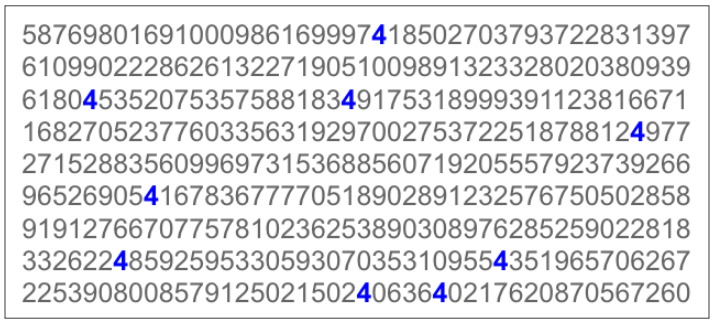
Effective monitoring requires being able to see and understand data in a variety of ways using visual analytics.

Visual Communication Is Primal

First, take a look at this data below for 5 seconds. How many 4's can you spot?

```
5876980169100098616999741850270379372283139 7
6109902228626132271905100989132332802038093 9
6180453520753575881834917531899939112381667 1
1682705237760335631929700275372251878812497 7
2715288356099697315368856071920555792373926 6
9652690541678367777051890289123257675050285 8
9191276670775781023625389030897628525902281 8
3326224859259533059307035310955435196570626 7
2253908008579125021502406364021762087056726 0
```

Now look at this image. How many 4's do you see now?



The answer is 9, and I am certain your brain spotted the 4's in the second image immediately. When we translate data into a visual element for efficient comprehension, what is called visual encoding, our brain is able to process the information far more efficiently and effectively.

According to David Williams, the director of the University of Rochester's Center for Visual Science, "[m]ore than 50 percent of the cortex, the surface of the brain, is devoted to processing visual information." What we perceive through our eyes surpasses our other senses, both in terms of pure volume of information and speed of input, which makes humans especially talented and agile visual pattern detectors. Indeed, it is not accidental that we utilize the expression "I see" as an acceptable stand-in for "I understand." Human vision is powerful and something we can harness to extract insights from data.

Descriptive Statistics Alone Are Not Enough

Another reason for visual analytics comes from Anscombe's quartet, which illustrates the ways in which summary statistics may not capture the ways in which data sets differ. First, he presents four datasets:

Each of these data sets have similar summary statistics, for example:

- the average x value is 9 for each dataset;
- the average y value is 7.50 for each dataset;
- the variance for x is 11 and the variance for y is 4.12;
- the correlation between x and y is 0.816 for each dataset; and
- a linear regression (line of best fit) for each dataset follows the equation y = 0.5x + 3

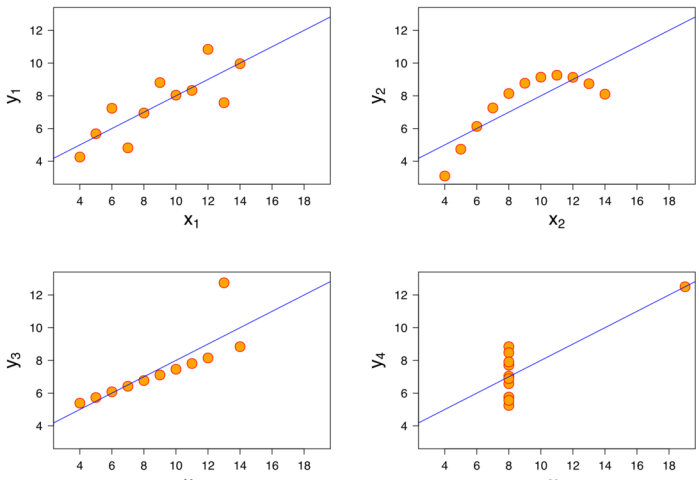
Thus, in terms of pure statistical measures, these four datasets appear to be similar. However, when the datasets are graphed, the ways in which they differ become obvious.

Simply computing and comparing summary statistics or staring at a table of data would not have told the same stories as these visuals.

See "Using Data Analytics to Boost Compliance Program Effectiveness" (Jun. 27, 2018).

The Four Datasets: Raw Data & Graph

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



Each dataset has eleven (11) x & y points.

Employing Visual Analytics

Compliance uses the data it gathers from disparate sources in multiple ways, including for periodic reporting to compliance committees, sharing with the business, aiding internal investigations or generally monitoring for risk in the organization.

The following three "contexts" or visual views of information can serve as a visual-analytics framework. This foundation can fuel a company's ability to drive more real-time and higher quality information for the compliance program and stakeholders.

For each context, we look at the key characteristics, some examples of these views and make suggestions as to what compliance data it might best fit.

Monitoring Context

Compliance is periodically monitoring the effectiveness of their various compliance programs. This is the classic "compliance scorecard" that is generated monthly or quarterly to provide information to the compliance organization, business partners or leadership. Visualizing data in this context should be the top priority for any compliance program due to the sheer importance of managing efforts efficiently.

A compliance scorecard will cover a company's proactive and reactive programs, including compliance related pre-approvals, hotline calls, due diligence reviews, training completion rates, spend monitoring and investigation matters. While voluminous in terms of the number of areas from which to gather data, it is the easiest data to gather, wrangle and visualize.

Data visualizations in this context should be:

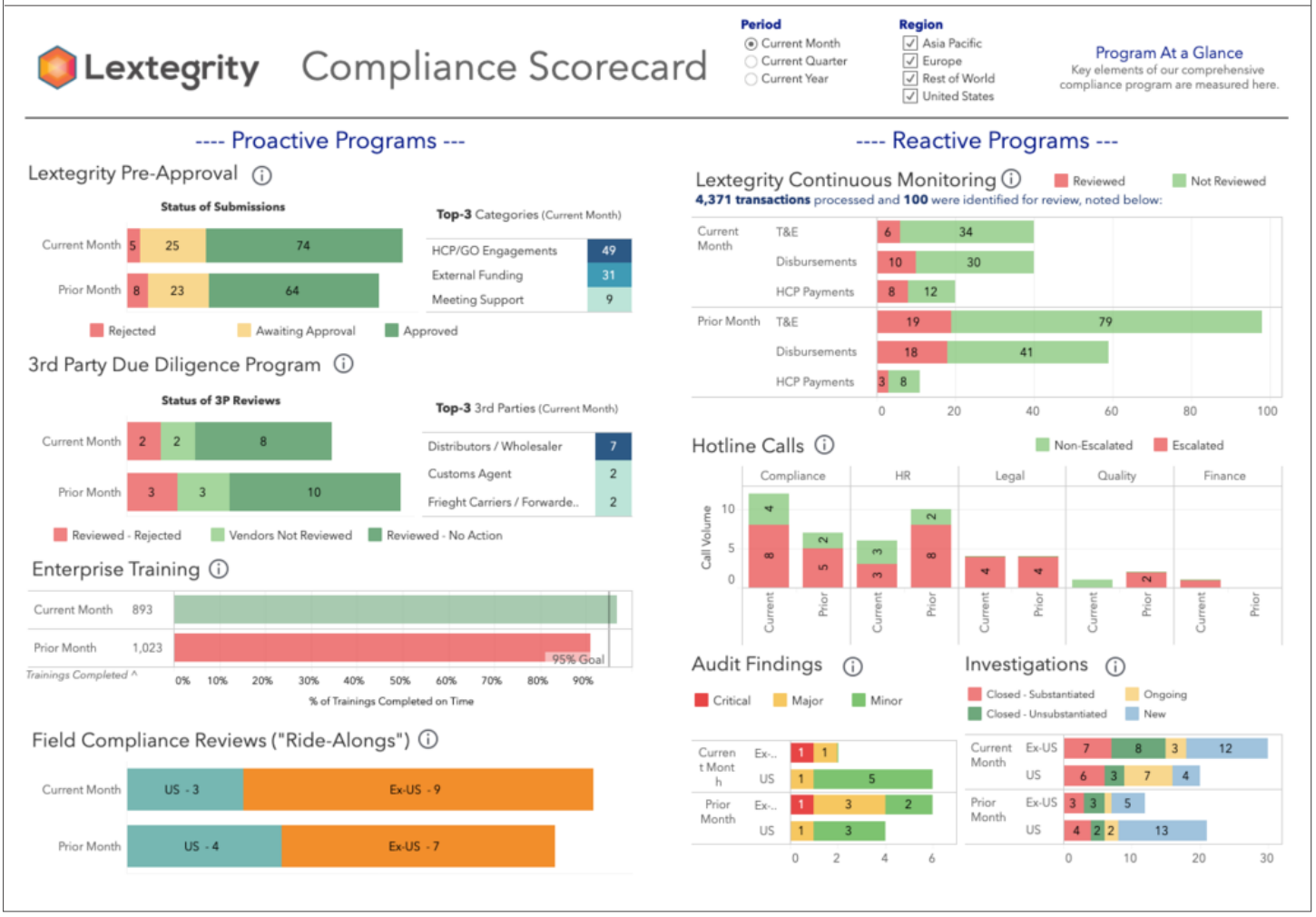
- periodic in nature, in that the data should be aligned with key intervals of time, such as weeks, months or quarters;
- based on highly aggregated data, typically aligned with a few key dimensions such as time, region, business unit, category or nature of values;
- inclusive of simple aggregate measures, such as the counts of hotline calls or investigations;
- easy to focus via filters of key dimensions such as region or business unit;
- comparative in nature to either the prior period or an expectation, benchmark, target or goal, typically by way of simple charts like bar charts; and
- delivered in a dashboard format.

Data visualization expert Stephen Few defines a dashboard as "a visual display of the most important *information needed to achieve one or more objectives*; consolidated and *arranged on a single screen* so the information can be *monitored at a glance*." Information visualized within the monitoring context should conform to that definition and not devolve into a "kitchen sink" of metrics and graphs, lest the audience struggle to efficiently consume the information and formulate appropriate conclusions.

Dashboards are the most common way in which executives and other decisionmakers are presented with data, so the information must be concise and understood quickly with little or no training provided.

See the Anti-Corruption Report's four-part series on measuring compliance: "Getting Started" (Aug. 2, 2017); "Seven Areas of Compliance to Measure" (Aug. 16, 2017); "How to Measure Quality" (Sep. 6, 2017); and "Gathering and Analyzing Data" (Sep. 20, 2017).

Figure 1. An example "Compliance Scorecard" for a robust compliance program.



Investigations Context

The investigations context of data visualizations allows the user to make specific requests for narrow data sets. For example, if a hotline call comes in about an employee's spend practices, a visualization of that employee's previous spends would be helpful or if a vendor is up for reapproval, a visualization of the value and frequency of the spend with that vendor may be necessary.

Data visualizations in the context of an investigation should:

- offer easy lookup of information, typically in a tabular format (see **Figure 3**);
- have filters for users to quickly limit what is within their scope of interest; and
- have views which provide a "profile" of a specific subject of interest (see **Figure 2**).

Figure 2. A "profile" view that enables detailed information focused on a sole subject of interest, including information such as spend over time, nature of spend, comparisons with peers and finally transactional details.

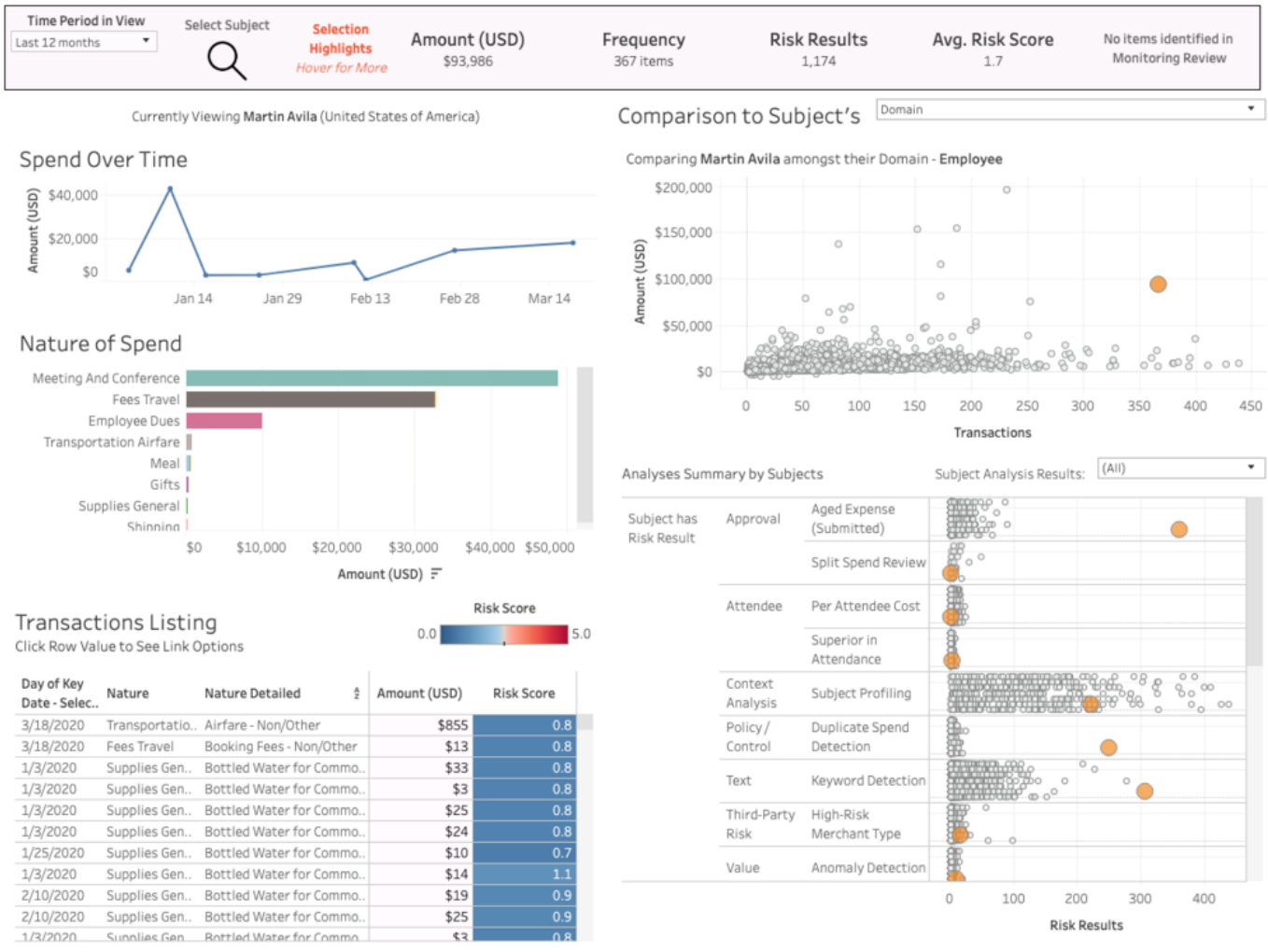
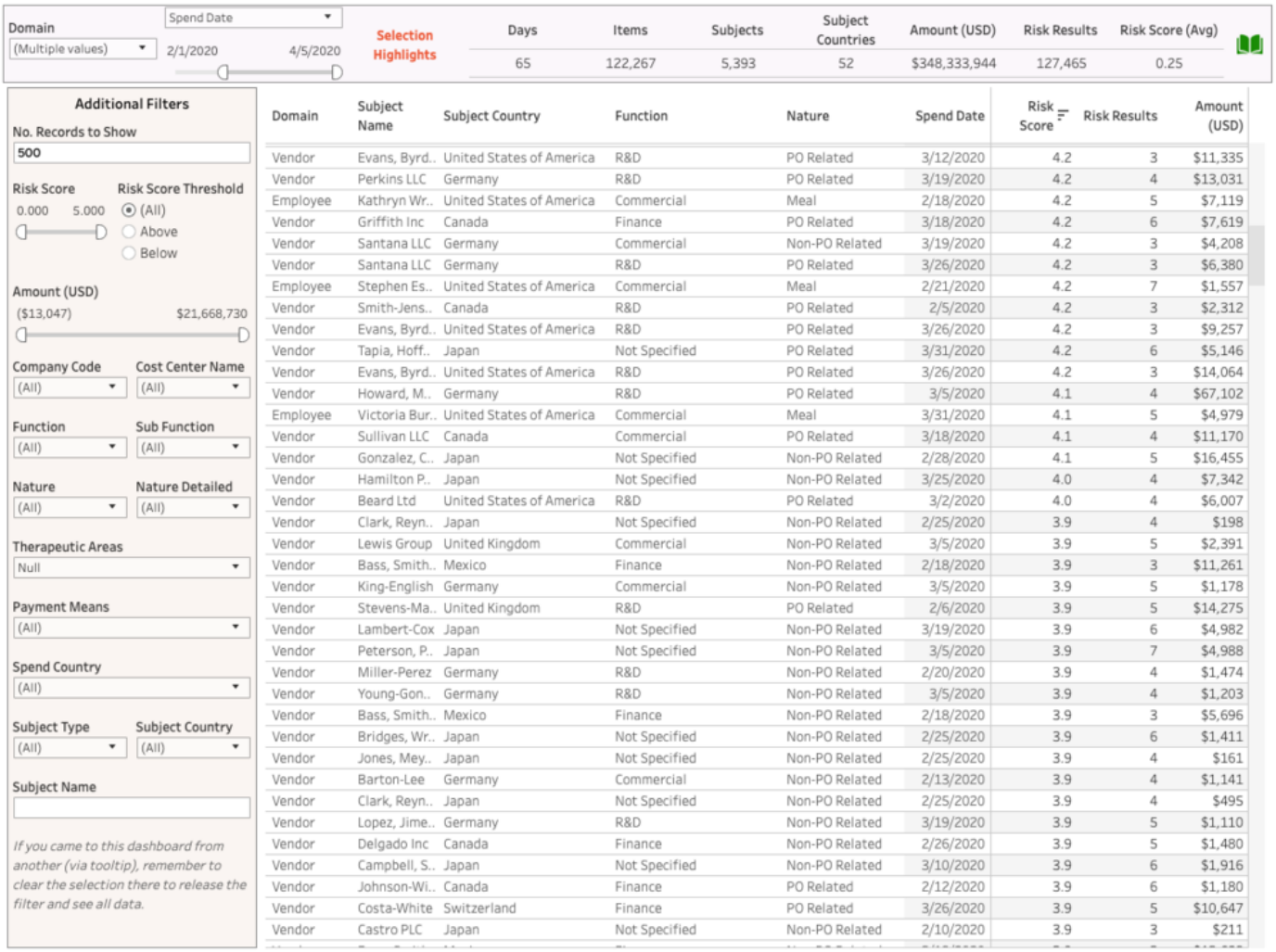


Figure 3. An example of a lookup view where the user can easily filter information in a tabular format. Further, this view can offer drill-downs from other views.



Domain: (Multiple values) | Spend Date: 2/1/2020 – 4/5/2020 | Selection Highlights

Days	Items	Subjects	Subject Countries	Amount (USD)	Risk Results	Risk Score (Avg)
65	122,267	5,393	52	$348,333,944	127,465	0.25

Additional Filters

No. Records to Show: 500

Risk Score: 0.000 – 5.000 | Risk Score Threshold: (All) / Above / Below

Amount (USD): ($13,047) – $21,668,730

Company Code: (All) | Cost Center Name: (All)

Function: (All) | Sub Function: (All)

Nature: (All) | Nature Detailed: (All)

Therapeutic Areas: Null

Payment Means: (All)

Spend Country: (All)

Subject Type: (All) | Subject Country: (All)

Subject Name:

If you came to this dashboard from another (via tooltip), remember to clear the selection there to release the filter and see all data.

Domain	Subject Name	Subject Country	Function	Nature	Spend Date	Risk Score	Risk Results	Amount (USD)
Vendor	Evans, Byrd..	United States of America	R&D	PO Related	3/12/2020	4.2	3	$11,335
Vendor	Perkins LLC	Germany	R&D	PO Related	3/19/2020	4.2	4	$13,031
Employee	Kathryn Wr..	United States of America	Commercial	Meal	2/18/2020	4.2	5	$7,119
Vendor	Griffith Inc	Canada	Finance	PO Related	3/18/2020	4.2	6	$7,619
Vendor	Santana LLC	Germany	Commercial	Non-PO Related	3/19/2020	4.2	3	$4,208
Vendor	Santana LLC	Germany	R&D	PO Related	3/26/2020	4.2	3	$6,380
Employee	Stephen Es..	United States of America	Commercial	Meal	2/21/2020	4.2	7	$1,557
Vendor	Smith-Jens..	Canada	R&D	PO Related	2/5/2020	4.2	3	$2,312
Vendor	Evans, Byrd..	United States of America	R&D	PO Related	3/26/2020	4.2	3	$9,257
Vendor	Tapia, Hoff..	Japan	Not Specified	PO Related	3/31/2020	4.2	6	$5,146
Vendor	Evans, Byrd..	United States of America	R&D	PO Related	3/26/2020	4.2	3	$14,064
Vendor	Howard, M..	Germany	R&D	PO Related	3/5/2020	4.1	4	$67,102
Employee	Victoria Bur..	United States of America	Commercial	Meal	3/31/2020	4.1	5	$4,979
Vendor	Sullivan LLC	Canada	Commercial	PO Related	3/18/2020	4.1	4	$11,170
Vendor	Gonzalez, C..	Japan	Not Specified	Non-PO Related	2/28/2020	4.1	5	$16,455
Vendor	Hamilton P..	Japan	Not Specified	Non-PO Related	3/25/2020	4.0	4	$7,342
Vendor	Beard Ltd	United States of America	R&D	PO Related	3/2/2020	4.0	4	$6,007
Vendor	Clark, Reyn..	Japan	Not Specified	Non-PO Related	2/25/2020	3.9	4	$198
Vendor	Lewis Group	United Kingdom	Commercial	Non-PO Related	3/5/2020	3.9	5	$2,391
Vendor	Bass, Smith..	Mexico	Finance	Non-PO Related	2/18/2020	3.9	3	$11,261
Vendor	King-English	Germany	Commercial	Non-PO Related	3/5/2020	3.9	5	$1,178
Vendor	Stevens-Ma..	United Kingdom	R&D	PO Related	2/6/2020	3.9	5	$14,275
Vendor	Lambert-Cox	Japan	Not Specified	Non-PO Related	3/19/2020	3.9	6	$4,982
Vendor	Peterson, P..	Japan	Not Specified	Non-PO Related	3/5/2020	3.9	7	$4,988
Vendor	Miller-Perez	Germany	R&D	Non-PO Related	2/20/2020	3.9	4	$1,474
Vendor	Young-Gon..	Germany	R&D	Non-PO Related	3/5/2020	3.9	4	$1,203
Vendor	Bass, Smith..	Mexico	Finance	Non-PO Related	2/18/2020	3.9	3	$5,696
Vendor	Bridges, Wr..	Japan	Not Specified	Non-PO Related	2/25/2020	3.9	6	$1,411
Vendor	Jones, Mey..	Japan	Not Specified	Non-PO Related	2/25/2020	3.9	4	$161
Vendor	Barton-Lee	Germany	Commercial	Non-PO Related	2/13/2020	3.9	4	$1,141
Vendor	Clark, Reyn..	Japan	Not Specified	Non-PO Related	2/25/2020	3.9	4	$495
Vendor	Lopez, Jime..	Germany	R&D	Non-PO Related	3/19/2020	3.9	5	$1,110
Vendor	Delgado Inc	Canada	Finance	Non-PO Related	2/26/2020	3.9	5	$1,480
Vendor	Campbell, S..	Japan	Not Specified	Non-PO Related	3/10/2020	3.9	6	$1,916
Vendor	Johnson-Wi..	Canada	Finance	PO Related	2/12/2020	3.9	6	$1,180
Vendor	Costa-White	Switzerland	Finance	PO Related	3/26/2020	3.9	5	$10,647
Vendor	Castro PLC	Japan	Not Specified	Non-PO Related	2/10/2020	3.9	3	$211

These views provide users quick ways to answer the questions they have in real time. Further, they can alleviate significant burden on those IT or business professionals that have historically supported "requests for information."

However, it is critically important to ensure these views incorporate the appropriate risk context alongside the raw data.

See "Treating Like Cases Alike – A Tool for Quantifying Compliance Issue Severity" (Jul. 13, 2016).

Insights Context

The final context is using visualizations to get a big picture view of a compliance program in order to glean insights, spot issues and make improvements. Data often includes significant noise and can yield false positives so spotting real insights from data requires careful design consideration and inquisitive users that can "connect the dots" as they explore the information interactively.

This context can be particularly useful in reviewing transactional data, such as employee travel and entertainment spend or vendor disbursements. These data sets are usually voluminous and rich in available attributes. However, they can be difficult to wrangle, cleanse and visualize in a useful way.

When visualizing data in this context, designers may take more liberty in their creations but doing so may require providing the audience with additional training and time to absorb and extract insights from the information.

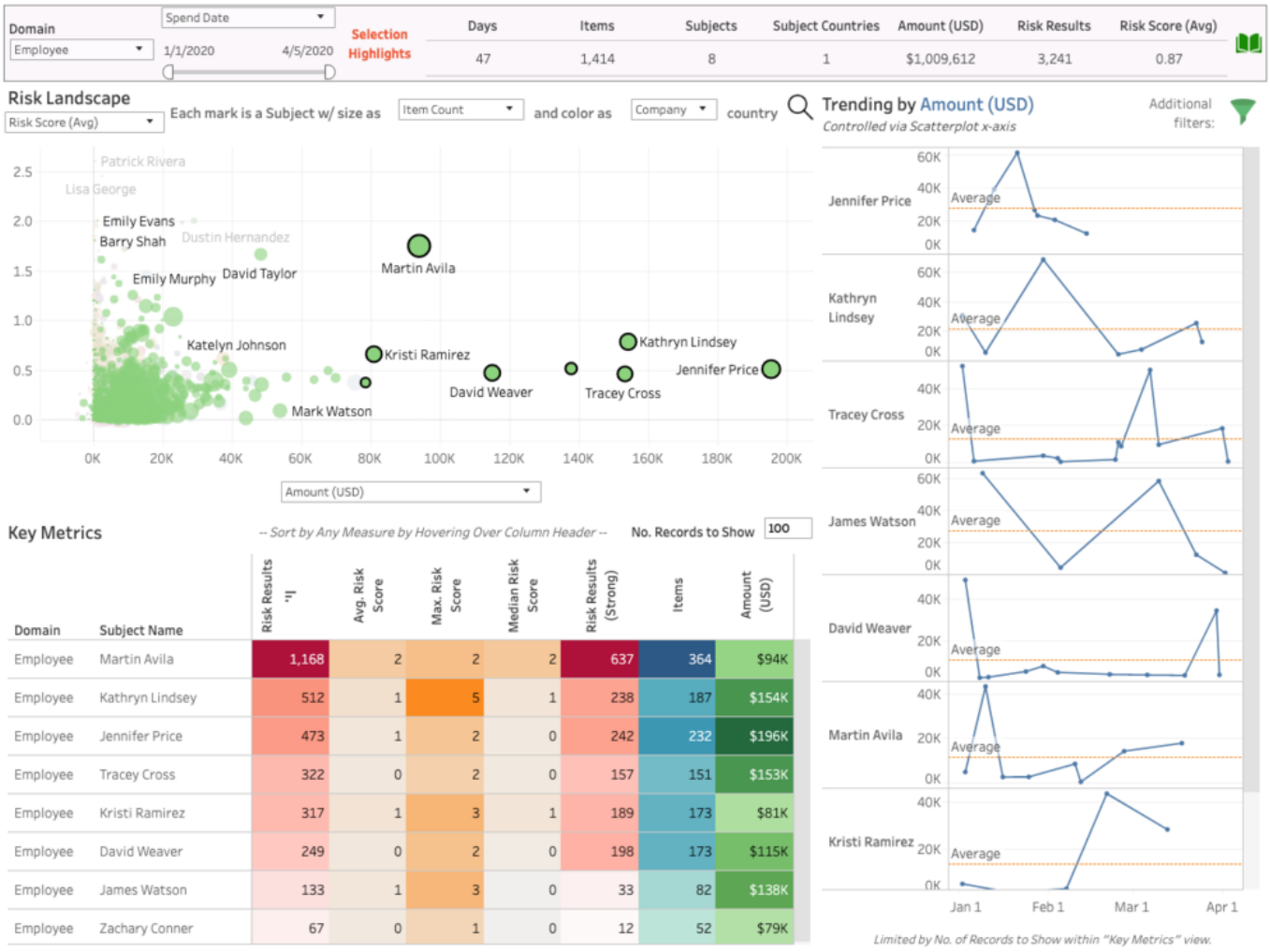
Data visualizations for insights should:

- provide visuals that quickly enable users to identify trends or outliers (see **Figure 4**) across specific dimensions of data;
- include a variety of visualization forms of similar information, such as pairing a scatter plot of subjects aligned with axes of two critical measures alongside a line chart for each subject to view spend trends (see **Figure 5**);
- be interactive such that complimentary visuals are dynamically adjusted when the user interacts with another visual;
- presenting multiple facets of information which dynamically adjust as users make selections within a visual; and
- have the ability to drill-down into finer grain "levels" of information, such as starting from a geographic view, drilling into the vendors for a specific geography, then selecting a specific vendor to see their transactions.

See "Compliance Records Are a Strategic Gold Mine" (Feb. 5, 2020).

Figure 4. An example of visualizations in an insights context enabling risk score outlier identification at the transactional level (note that an impressive 150K transactions are presented in this single view).

Figure 5. An example of visualizations in an insights context enabling exploration of subject (*e.g.*, vendors, employees) level information. The user has selected a cohort of subjects, which has filtered the other visuals for closer inspection.



Domain	Subject Name	Risk Results	Avg. Risk Score	Max. Risk Score	Median Risk Score	Risk Results (Strong)	Items	Amount (USD)
Employee	Martin Avila	1,168	2	2	2	637	364	$94K
Employee	Kathryn Lindsey	512	1	5	1	238	187	$154K
Employee	Jennifer Price	473	1	2	0	242	232	$196K
Employee	Tracey Cross	322	0	2	0	157	151	$153K
Employee	Kristi Ramirez	317	1	3	1	189	173	$81K
Employee	David Weaver	249	0	2	0	198	173	$115K
Employee	James Watson	133	1	3	0	33	82	$138K
Employee	Zachary Conner	67	0	1	0	12	52	$79K

Conclusion

Visual analytics are a critical part of any company's data journey, as they provide the most straightforward way to monitor, investigate and generate insights from data efficiently and effectively. While there are some challenges to overcome, such as wrangling and automating data, there are tools available to assist employees in meeting their information needs, such as Tableau and Power BI.

There is immense value to be gained from working visually with compliance data. All compliance programs should consider taking at least small steps towards employing robust visual analytics and then iteratively improving over time.

Andy Miller is chief analytics officer at Lextegrity, a compliance software company focused on helping organizations mitigate risk across the enterprise, specifically through employing advanced analytics to expose fraud, bribery, corruption, sanctions violations and conflicts of interest within enterprise data.