SHRM-SIOP Science of HR White Paper Series

Data Visualization: Get Visual to Drive HR's Impact and Influence

Evan F. Sinar
DDI

Evan F. Sinar
DDI
Evan.Sinar@ddiworld.com

Evan Sinar is Chief Scientist and a Vice President at DDI. He designs and executes analytics linking leadership programs to business outcomes and produces actionable insights about talent trends and people practices, including leading DDI's Global Leadership Forecast and High-Resolution Leadership research programs. Evan authored data visualization chapters in the book *Big Data at Work* and the *SAGE Encyclopedia of Industrial and Organizational Psychology*. He is an active social media author and curator and has been recognized as a top worldwide influencer on data visualization as well as other leading-edge topics such as HR analytics, big data and the Internet of things. He has authored over 70 professional presentations and articles for major publications and professional conferences and serves on the editorial board of several journals. Evan holds a Ph.D. in industrial and organizational psychology and is a Fellow of the Society for Industrial and Organizational Psychology.

ABSTRACT

The goal of this paper is to establish a rationale and roadmap for adopting visualization techniques to represent HR data. By using these approaches, HR professionals can unlock a level of data engagement and influence difficult to achieve using tables and numbers alone. Many HR datasets are well-suited to visualization as a component of storytelling and messaging for nontechnical, professional and executive audiences. Data visualization is also a more approachable skill than ever, with many easy-to-use and quick-to-learn tools available. This paper provides an overview of foundational concepts for effective visualization design and a recommended set of practical resources and principles for rapidly incorporating data visualization into your own work.

Data Visualization: Get Visual to Drive HR's Impact and Influence

Data visualization—defined as representing quantitative variables through visual properties such as position, length, area and color—is an immensely powerful technique for exploring and communicating information. To a degree far beyond traditional text or number-based formats, well-designed visualizations make data accessible, clear, enticing and engaging to business audiences. Yet, despite the recent surge in visualization's use, many HR professionals aren't proficient in these compelling approaches. Thus, they miss golden opportunities to expand their influence and audience reach.

The purpose of this paper is to:

- Provide an overview of key concepts for effective data visualization.
- Review the evidence base supporting its use.
- Recommend a set of practical resources for adopting these techniques in your own practices.

Why Visualization Now?

The historical roots of data visualization extend back centuries: from the early 1700s-era work of pioneers such as William Playfair (1759-1823, considered a pioneer in graphically displaying statistical information) and Charles Minard (1781-1870, best known for creating data-rich geographic representations, most notably his stunning map of Napoleon's 1812 Russian Campaign) to the foundations built by (still active) late 20th-century practitioners William Cleveland and Edward Tufte. However, the past several years have seen an explosion of prominence for data visualization within the business community.

I see three main trends driving this upswell: big data availability, data-driven decision-making and information transparency.

1. **Big Data Availability.** The ever-increasing proliferation of information—in scale, speed and variety—has created business environments where analysis expertise to sort through and extract insights from these large-scale data is increasingly valuable. Data visualization taps into our formidable visual processing skills to counteract and inoculate against big data overload and confusion.

   Visualization techniques also pair well with other facets of data science. These techniques serve as a "face" for the advanced machine learning and artificial intelligence methods in expanded use in many organizations with the data gathered from and about their employees.

2. **Data-Driven Decision-Making.** Many organizations have shifted away from "gut instinct"-based decision making, to instead adopt an expectation—and often a

mandate—for data to drive strategic and tactical choices. This has hugely increased the proportion and number of business professionals being asked to make decisions based on the ever-growing amount of data. Through data visualization techniques, organizations can involve a broader range of experts to explore and extract value from their datasets.

3. **Information Transparency and Ethics.** The initial wave of unbridled optimism about big data's impact has faded, replaced—particularly for employee data—by an orientation toward data-gathering that guards against bias and unethical data use. "Black box" analytical models resulting from complex data science procedures are also facing increased scrutiny because they are largely impenetrable to lay audiences. These forces place a premium on data transparency. Visualization is inherently well-suited to information transparency, providing a platform for data openness, engagement and interactivity. It also provides a more direct view of the raw data, counteracting the potential for biased interpretations on the part of the researchers and giving the audience an opportunity to see the underlying data for themselves.

The HR Skills Gap—and Opportunity

For HR professionals—serving as the conduit between employee data and talent strategy—visualization is a vital area of proficiency to fuel success. Yet, it's not typically a strength for those in the HR function.

In data gathered in 2017 (DDI, The Conference Board & EY, 2018), we found that only 16% of organizations were successfully implementing data visualization and

storytelling techniques. This is despite the utility these methods have in linking an organization's bench strength (supply of leaders ready to immediately step in to fill critical roles) and its financial success.

These skills gaps within HR have further consequences at the organizational level—for example, where HR professionals are responsible for growing others in these areas despite often lacking these skills themselves.

Business Applications and Research

Data visualization has three main objectives: exploring data, explaining it and building others' engagement in resulting actions.

1. **Exploring Data**: Data visualization allows more efficient detection of relationships, patterns and differences within people datasets than information presented in tabular or textual form.
2. **Explaining Data**: Data visualizations convey complex concepts in a format that surfaces and clarifies data trends. Visualized forms of data also guide audiences to contribute their observations and ideas about why and how patterns exist.
3. **Building Engagement**: Data visualizations elicit a powerful and unique type of audience connection to quantitative information that is unlikely using more common business presentation formats of endless bulleted text lists or inscrutable tables of numbers.

Data visualization research conducted in a business environment is in early states but rapidly gaining momentum. Three recent research studies provide a

promising foundation for visualization's advantages over more traditional forms of business communication.

A 2014 study by Pandey and colleagues found the visually-presented information was more persuasive and resulted in greater attitude change to identical information presented using tables. Al-Kassab and his research team (2014) found that visualization improved managers' information synthesis and processing abilities while making decisions. And a 2015 study by Kernbach, Eppler and Bresciani demonstrated higher attention to, agreement with and recall of business strategies when presented visually.

Foundational Issues

Objectives and Business Questions. Data visualization is most effective when well-aligned with a business question and sought-after action. That is, what is the goal of the analysis and resulting visualization? Five common visualization objectives and example associated business questions for HR/people data are to:

1. **Compare values across groups**—for example, employee engagement scores by region, generation or function to target wellness investment to remedy deficiencies.
2. **Display connections or relationships between variables**—for example, between employee level and turnover risk score category to guide retention efforts.
3. **Show hierarchical or part-to-whole structures** – for example, number of yearly hires by role within a department within a division to allocate onboarding budgets.

4. **Illustrate change over time**—for example, training completions by course and by quarter to design learning paths and phase out underused courses.
5. **Display data patterns within maps or geographies**—for example, average assessment score by country to inform recruitment strategies.

What Data are Best to Visualize? While there is no one "right" kind of data to visualize, certain data characteristics are better for taking advantage of visualization's strengths.

Generally, higher-volume data (for example, from large-scale employee populations) are more suited to visualization than low-volume data. Over-time data—for example, multi-year survey or test administrations—are particularly well-matched to visualization techniques, as there are entire classes of visualizations designed to be used with data spanning time periods.

Visualization is also especially powerful for multivariate data—that is, described by several distinct dimensions such as competency scores, as well as for data segmented into hierarchical, nested categories such as region, department and job title. Geographical, map-based data can also be represented visually using techniques designed for that purpose.

Of course, baseline standards for data quality and veracity must be met for visualization as with other analytics. For a thorough discussion of this topic, I recommend the SHRM-SIOP Science of HR White Paper "Trends and Practices in Talent Analytics" (Kaur & Fink, 2017).

Audience and Format. Before creating any data visualization, consider your audience and format(s) for presentation. What's the audience's familiarity with visualized data? What's worked well (and what hasn't) for this audience in the past? Ultimately, what insight, decision and actions are you looking to drive with your audience?

Answers to these questions will guide what types of visualizations you use, as well as the level of detail you share. A group of senior executives may expect a more straightforward, recommendation-focused visualization whereas a technical audience may seek additional detail causing you to include more such information in the visualization. If possible, create separate presentations for audiences varying widely in their orientation toward data visualization to avoid creating a single view of the data that doesn't fully meet the needs of either group.

In terms of presentation format, visualizations presented on slides will typically need to be higher-resolution, with larger text and visual elements, and less detailed than visualizations shared in a printed report. Slide presentations can have their own challenges—factors such as room lighting, screen size and projector resolution can have a major impact on which visualizations you'll use.

If presenting using a projector, I strongly recommend a trial run of your visualizations in the same room and with the same A/V setup as the final presentation. Be sure to factor in enough time to make adjustments (e.g., enhancing contrast, enlarging graphics, adding call-out boxes to highlight key data trends) to ensure that visualizations show up in high legibility and interpretability.

Visualization Techniques

I recommend three overarching techniques to increase the accuracy, interpretability and persuasiveness for the data visualizations you create: contrast, annotate and sequence.

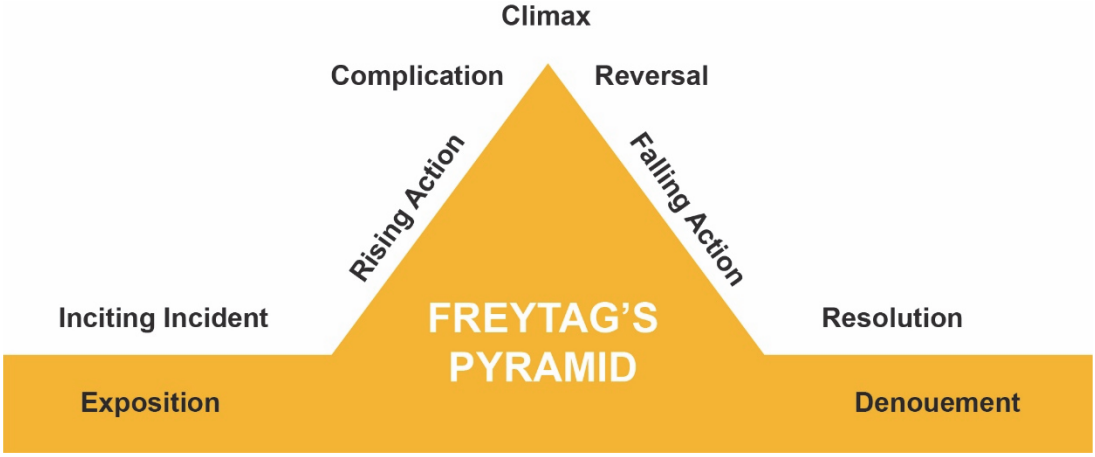
Contrast. Use contrast within the visualization to draw attention to the most notable segments of or patterns within the data. This can be done by using focal colors (see color recommendations below) to accentuate critical visualization features. It's important to note that creating contrast also involves deemphasizing less important data elements. This may be because they aren't statistically significant, are too small to be meaningful or simply aren't the focus of a specific part of the presentation (but may be given focus at a later point). This is best done by using gray to fade nonfocus elements out of the foreground.

Annotate. Data visualizations are rarely complete immediately after emerging from the program used to create them. Post-creation annotations are often essential for clarifying your message for the audience. Annotations include labeling the visualization not merely with a descriptive title (e.g., "Applicant volume over time by region"). Create a title that drives action (e.g., "Recruitment staff needed in the Southeast to support a surge in candidates"). Also, a thorough analyst will already know the background for notable visualization features such as a spike in the trend line in a particular quarter. "Callout" annotations to layer in this context will help the audience understand the reasons for key patterns.

Annotations can also be used to note data provenance such as the source and date for the data. This pre-empts basic clarification questions from the audience and helps build confidence in data quality. When used properly, annotations can improve memorability of data visualizations, reduce working memory demands the audience faces, and help establish a common frame of reference and understanding for the visualization. Another valuable outcome is that informative visualizations allow discussion time to focus less on basic questions and more on sophisticated discussions about causes, actions and implications.

Sequence. Think of visualizations not as stand-alone data views but as components in a larger narrative that you're building and sharing with the audience. Well-sequenced visualizations can match and take advantage of established storytelling principles. For example, consider Freytag's Pyramid (see Figure 1) —from early "exposition" use within a presentation to introduce and set up the topic, to next highlight "rising action" as supporting facts are revealed, to the "climax" revealing the primary insight, to "falling action" showing how the conflict is addressed, and finally to the "conclusion" as a visual summary of the entire story.

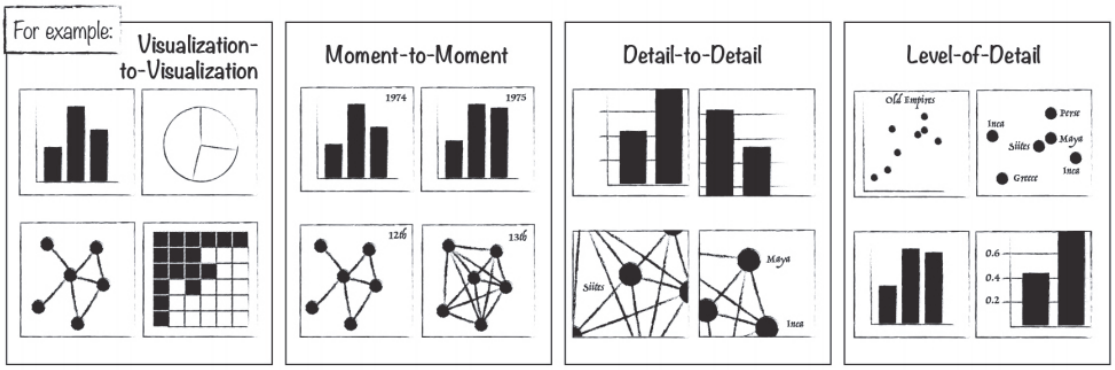
Figure 1: Freytag's Pyramid



Source: https://www.clearvoice.com/blog/freytags-pyramid-using-classic-storytelling-techniques-successful-marketing/

Sequencing visualizations can also involve adopting concepts of a "data comic" (Bach, Riche, Carpendale, & Pfister, 2017): progressive panel-by-panel (or slide-by-slide) views shifting from one to an alternate view of the same data, changing the time scale (one year to the next) or moving from a higher to a lower level of detail. See Figure 2 for an example of a data comic-style sequence of visualizations.

Figure 2: Data Comic Panel-by-Panel Example



Source: http://aviz.fr/~bbach/datacomics/Bach2017datacomics.pdf

Sequencing that gradually steps up from a basic to a complex data visualization (orienting the audience to each step before moving on) is also a counterweight to

pressures based on the overstated assumption that "the audience needs to be able to understand a visualization in five seconds!" Taken literally, this approach will lead to simplistic data views (and weaker impressions about the presenter). Rather than deploying visualization to simplify the complex (often futile or misleading given the true sophistication of the modern business environment), instead use well-designed (high-contrast, annotated and sequenced) visualizations to convey and clarify the complex.

Avoiding Common Design Missteps. With visualization, minimalist design often produces maximum clarity—extraneous elements such as backgrounds, 3D effects, gridlines, borders and shadows draw focus from the data, can reduce accuracy of interpretation and should be stripped back as much as possible.

Responsible data visualization also includes awareness and avoidance of misleading (and, if done intentionally, unethical) design choices—for example, truncating the Y-axis for column charts (which can make even small changes appear massive); using dual Y-axes (which can show a false correlation between two variables).

Make visualizations easy for the audience by direct-labeling data points whenever possible. This is preferable to relying on axes and captions focusing them to oscillate back and forth across the visualization. Also, provide common baselines—for example, aligning bar segments for different groups by their left edge—to aid comparisons when presenting multiple groups or categories in the same visualization.

Data Visualization Versus Infographics. Infographics are related to data visualization. Yet, the two forms of information presentation are distinct in several key aspects.

Infographics place a stronger emphasis on design and aesthetic features, sometimes at the expense of data accuracy. Data visualizations are typically defined as being automatically generated (e.g., using software), in contrast to infographics, which are often illustrated with the involvement of a graphic designer.

Infographics tend toward subjectivity whereas data visualizations are more objective in their message. Infographics are often made up of several individual data visualizations, arranged in a story sequence or around a common theme.

Though this white paper is about data visualization, well-designed infographics share many of the same effective design principles (e.g., use of color; graphical properties; annotations). I recommend the work of the Accurat Studio (https://www.accurat.it/) and David McCandless (https://informationisbeautiful.net/).

Visualization Principles and Techniques

Accuracy of Visual Features. With an eye toward accurate audience interpretation of the patterns illustrated by a data visualization, the various graphical features used to denote data are far from equal and should be prioritized accordingly.

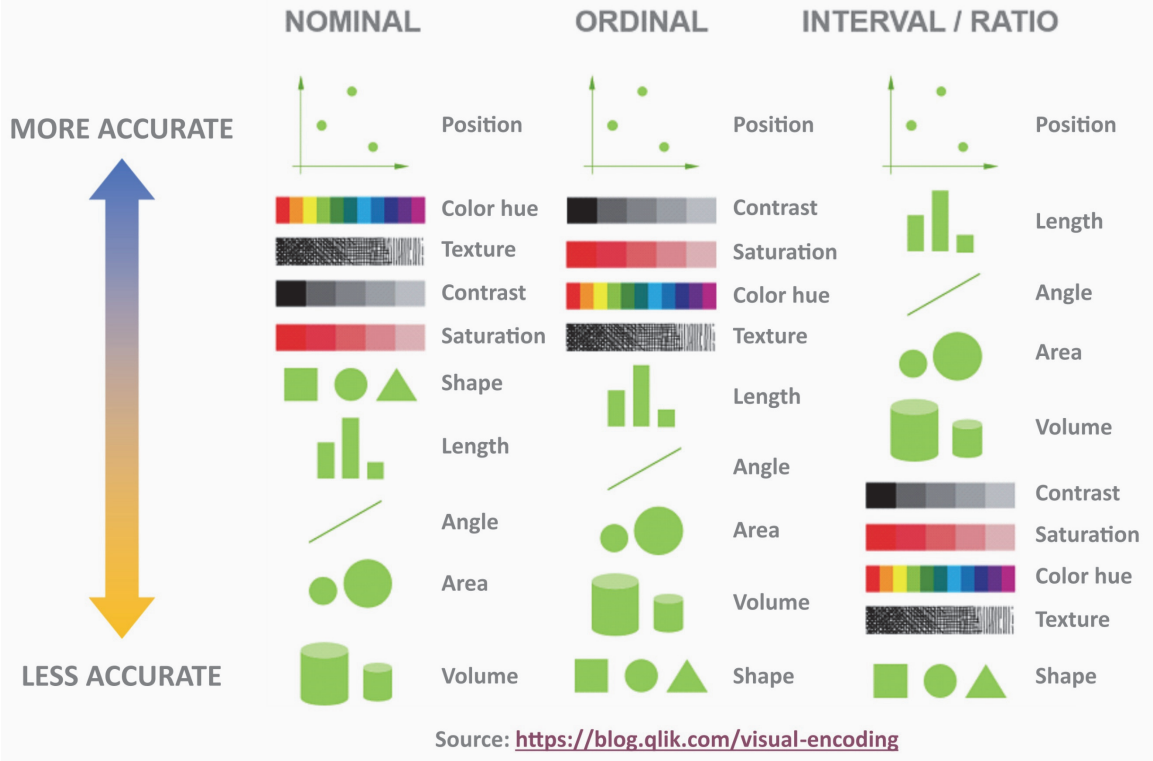
Research summarized by Lunbald (2015) rank-ordered visual features by accuracy (Figure 3). Features higher on the list will be better-suited for visual design features, leading to higher interpretative accuracy than those lower on the list.

For example, for quantitative interval/ratio data, varying data element positions will be more accurately interpreted than varying their length, which, in turn, will be more accurate than varying their volume.

It's important to understand these distinctions when making design choices for a visualization. Use features higher in these lists to illustrate more important distinctions within the data. Avoid those lower in the list unless necessary (Figure 3).

Variation in relative accuracy also explains why certain visualizations are recommended over others. For example, because volume differences are interpreted less accurately than those based on area or length, 3D graphics are rarely appropriate.

Figure 3: Visual Features by Data Type, Ranked by Interpretative Accuracy



Source: https://blog.qlik.com/visual-encoding

Using Color. Color is both a powerful and problematic graphical property for representing data. Many visualizations used in business settings use color excessively or inappropriately as a graphical property.

One particular risk area is the use of "rainbow" color scales. These fail to recognize that visual distinctions are not consistent across the color spectrum, leading to inaccuracy and misinterpretation about the magnitude of score differences (Kosara, 2013).

A more accurate approach is to consider color variation using saturation rather than hue for continuous scales. A second color-related risk area is use of palettes that don't account for colorblindness. Online resources exist for checking graphics for colorblindness suitability and for identifying color schemes that accurately represent quantitative distinctions (see, for example, http://colorbrewer2.org).

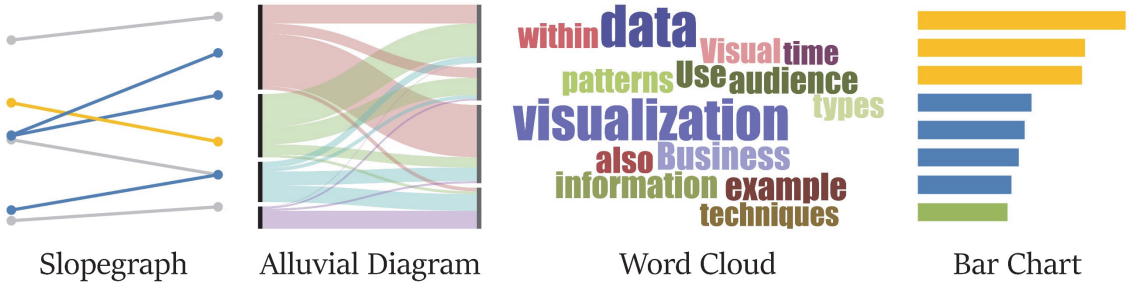
Types and Techniques

Data visualization types are guided by the research question being asked—or the data story to be told. I use Kirk's (2012) five-category taxonomy to classify visualization types, which also build on the five main objectives for visualization as listed above.

For each category, I provide a brief description of the key question addressed by the visualization, identify several representative types and show simplified versions (without labels or annotations for this purpose). For further "in the wild" examples of each type—very useful for getting ideas and gauging feasibility for your own data—I highly recommend searching Google Images for the name of each type.

- **Comparing Values Across Groups**—data visualizations that aid comparisons among multiple groups based on their quantitative properties. These visualizations are designed to illustrate data patterns (e.g., which groups are larger than others; which groups differ most between two time periods; which groups are over-represented in how their members are distributed across categories). Examples are slopegraphs, Sankey/Alluvial diagrams, word clouds, bar charts (Figure 4).

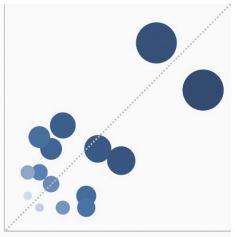
Figure 4: Representative Visualizations for Comparing Values



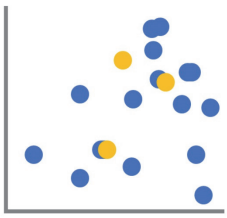
Slopegraph | Alluvial Diagram | Word Cloud | Bar Chart

- **Displaying Connections or Relationships Between Variables**—data visualizations designed to show how two or more variables relate to one another. These visualizations display patterns such as strength of relationship between variables, what values most often co-occur with others, and the multivariate profiles of particular cases. Four examples are bubble plots, scatterplots, parallel coordinates and word trees (Figure 5).

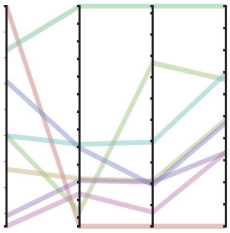
Figure 5: Representative Visualizations for Displaying Connections or Relationships Between Variables



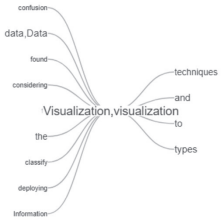
Bubble Plot



Scatterplot



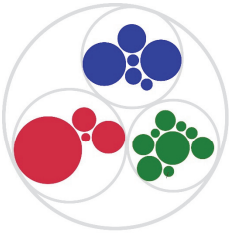
Parallel Coordinates



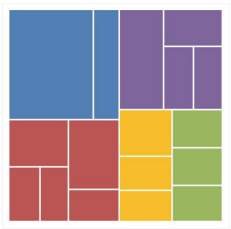
Word Tree

- **Showing Hierarchical or Part-to-Whole Structures**—data visualizations designed to depict patterns such how a single data category relates within a broader grouping (e.g., the scope and depth of an ordered data structure; how a group breaks down into its proportionate members). Four examples are circle packing diagrams, tree maps, sunburst diagrams and pie/donut charts (Figure 6).

Figure 6: Representative Visualizations for Showing Hierarchical or Part-to-Whole Structures



Circle Packing



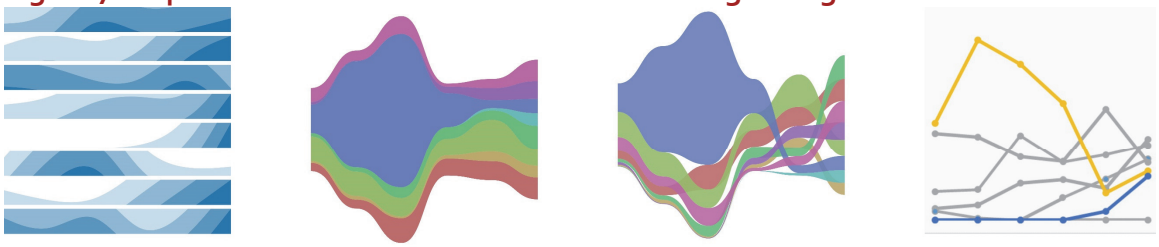
Tree Map

Sunburst

Donut Chart

- **Illustrating Change over Time**—data visualizations designed to show variation across a time span. These visualizations show data patterns such as how certain groups become larger in an absolute or relative sense compared with others across years and the trajectory and magnitude of a category's growth or decline over time. Four examples are horizon charts, stream graphs, bump charts and line charts (Figure 7).

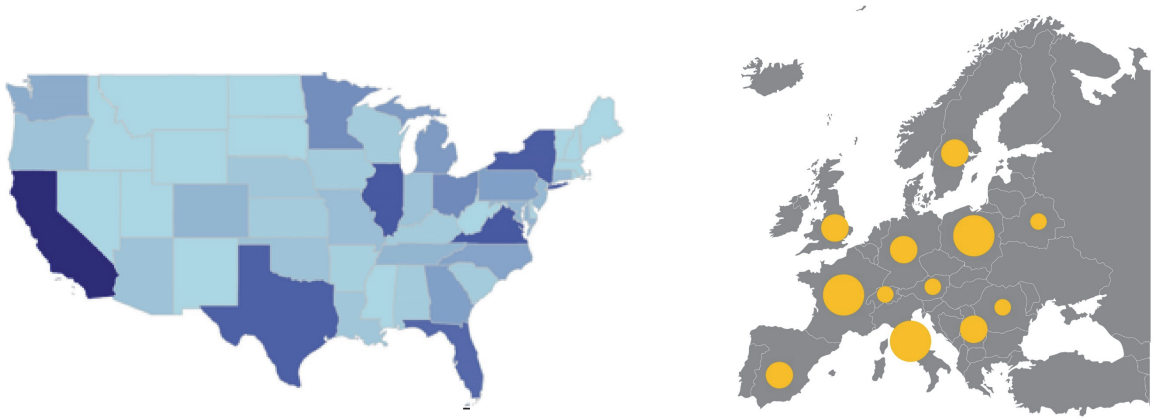
Figure 7: Representative Visualizations for Illustrating Change over Time



Horizon Chart | Stream Graph | Bump Chart | Line Chart

- **Displaying Data Patterns Within Maps or Geographies**—data visualizations designed to display data patterns overlaid on maps, with geographical entities the foundation onto which other visual properties (for example, value-sized shapes or color saturation) are placed. Two example types are Choropleth maps and dot maps (Figure 8).

Figure 8: Representative Visualizations for Displaying Data Patterns Within Maps or Geographies



Choropleth Map | Dot Map

For an expanded, searchable view of over 150 data visualization types, including descriptions, definitions, data structure guidance, and examples, peruse The Data Viz Project (http://datavizproject.com/).

Tools and Continuous Learning

Tools. Approachability and ability to create high-quality data visualizations may be easier than you think. Although it's certainly possible to invest time and expense to either learn a visualization-focused programming language or to invest heavily in commercial visualization software, it's no longer necessary to do so for the vast majority of HR professionals seeking to adopt and apply these techniques.

With a wide range of data visualization tools available—and growing every day—I focus in this paper on three tools that are open-source or widely available, involve a shallow learning curve and are capable of producing presentation-quality graphics.

1. **RawGraphs (**http://rawgraphs.io/**)** is an open-source online tool that allows users to enter their own data and to generate many of the visualization types listed above, and that can be paired with a vector graphics editor to rapidly produce free, high-quality visualizations.
2. **Voyant** (http://voyant-tools.org/) is an open-source online tool for processing text-based data to produce word trees, bubble lines and a dozen other visualizations for graphically displaying patterns in a large-scale text corpus.
3. **Microsoft Excel**. Excel is surprisingly capable for creating visualizations, particularly after overruling its default settings, which violate many of the foundational principles above. Several prominent data visualization practitioners also release templates designed specifically for Excel, and the 2016 version of the program further extends its visualization abilities.

A Small-Scale Example. The data and steps below walk through a small-scale example of data visualization using the RawGraphs tool described above; this example can be scaled up to match your own data specifications.

Table 1. Sample Data for Bump Chart

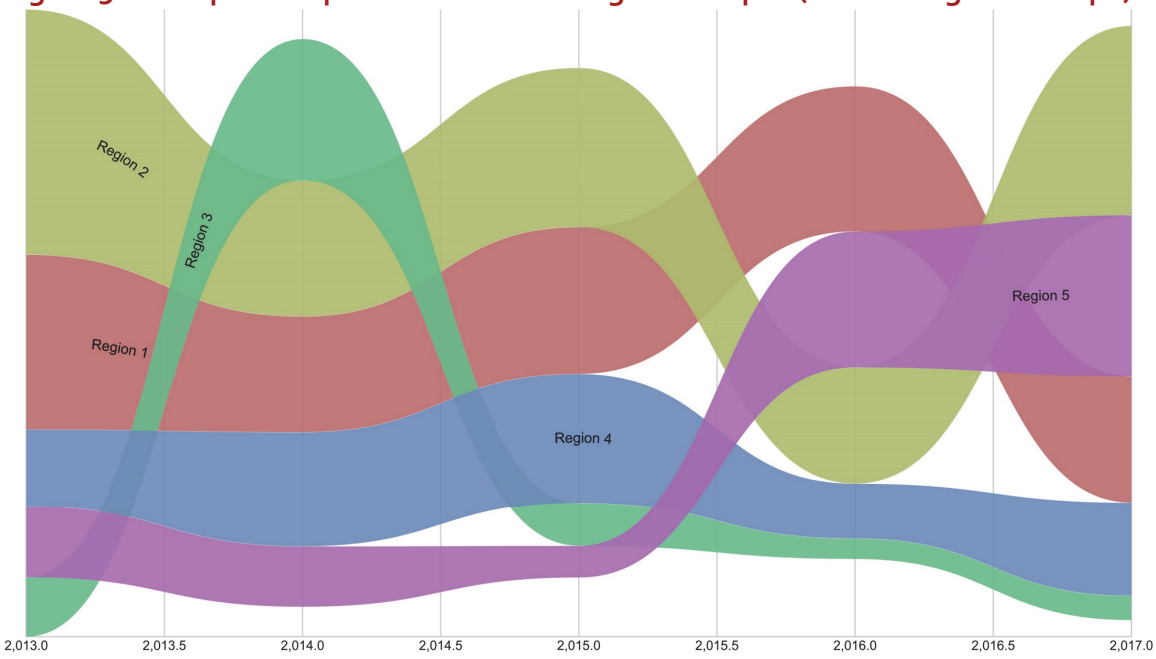
Region	2013	2014	2015	2016	2017
Region 1	800	529	672	664	578
Region 2	1122	622	728	531	866
Region 3	268	647	196	92	109
Region 4	353	524	591	253	425
Region 5	322	273	143	622	737

To create a visualization using this dataset, start by opening the website http://rawgraphs.io/.

- Click on the "Use It Now!" button to progress to the "Load your data" screen.
- Copy and paste the data above into the entry box.
- Click on the "Click here to stack it" button in the lower right, then select "Region" as the dimension to stack on (this reorients the data into a vertical format).
- Scroll down to view the available visualization types and select "Bump Chart." (Note: Select "Try our samples" in the "Load your data" section to view and experiment with additional visualization types.)
- Scroll further to configure the visualization: Drag and drop "Region" into the Group box, "column" into the Date box and "value" into the Size box.

- The visualization below will appear. The "Download" section can be exported directly into an image (.png) format (picture file with transparent background), or to a vector graphics (.svg) file for further editing in a graphics program to adjust font sizes, labels and so on. (I recommend Inkscape, https://inkscape.org/en/, for this as a free and richly featured image processing tool.)
- The resulting example graphic (Figure 9) shows year-over-year trends by region (X-axis is year; wider streams indicate larger values; higher placement on the Y-axis indicates rank-ordering of regions by year).

Figure 9: Example Bump Chart Created Using RawGraphs (Pre-editing in Inkscape)



Sources for Continuous Learning and Inspiration. Befitting a rapidly emerging and energetic field, many of the leading data visualization practitioners are prolific curators and authors on the topic. I recommend the work of the four experts below for staying current on new data visualization tips, templates and examples to emulate:

- Andy Kirk (http://www.visualisingdata.com/blog/).
- Stephanie Evergreen (http://stephanieevergreen.com/category/blog/).
- Cole Nussbaumer Knaflic (http://www.storytellingwithdata.com/).
- Robert Kosara (https://eagereyes.org/).

Beyond the work of these experts, I advocate tracking the Twitter hashtag #dataviz to efficiently monitor a steady stream of visualization examples and guidance. Seek further visualization inspiration through compilations such as the Information is Beautiful awards (https://www.informationisbeautifulawards.com/showcase) and Harvard Business Review's Visual Library (https://hbr.org/offers/visual-library) and year-end summaries (e.g., https://hbr.org/2016/12/what-we-learned-about-management-in-2016-in-19-charts).

Conclusion

HR professionals armed with skills in information visualization will be more effective in exploring the data they draw on to make workplace decisions, in explaining quantitative trends to nontechnical audiences, and in engaging and persuading others about their messages.

Visual storytelling skills are highly developable, and they are also exceedingly valuable. As HR professionals become proficient in data visualization (currently not a skill strength for most), they can make their messages more compelling, memorable and engaging to the full range of employees and business stakeholders they must influence.

References

Al-Kassab, J., Ouertani, Z. M., Schiuma, G., & Neely, A. (2014). Information visualization to support management decisions. *International Journal of Information Technology & Decision Making, 13,* 407.

Bach, B., Riche, N. H., Carpendale, S., & Pfister, H. (2017). The Emerging Genre of Data Comics. *IEEE Computer Graphics and Applications, 38*(3), 6-13.

DDI, The Conference Board, & EY. (2018). *Global Leadership Forecast 2018: 25 research insights to fuel your people strategy.* Pittsburgh, PA: DDI.

Kaur, J., & Fink, A., (2017). *Trends and Practices in Talent Analytics.* Society for Human Resource Management (SHRM)-Society for Industrial-Organizational Psychology (SIOP) Science of HR White Paper Series.

Kernbach, S., Eppler, M. J., & Bresciani, S. (2015). The Use of Visualization in the Communication of Business Strategies: An Experimental Evaluation. *International Journal of Business Communication, 52*(2), 164-187.

Kirk, A. (2012). *Data visualization: A successful design process.* Birmingham, UK: Packt Pub.

Kosara, R. (2013). *How the rainbow color map misleads.* Retrieved January 7, 2018, from https://eagereyes.org/basics/rainbow-color-map.

Lunbald, P. (2015) *Second Pillar of Mapping Data to Visualizations: Visual Encoding.* Retrieved December 14, 2017, from https://blog.qlik.com/visual-encoding.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG), 5*(2), 110-141.

Pandey, A. V., Manivannan, A., Nov, O., Satterthwaite, M. L., & Bertini, E. (2014). The persuasive power of data visualization. *New York University Public Law and Legal Theory Working Papers*, Paper 474.

Sinar, E. F. (2015). Data visualization. In S. Tonidandel, E. King, & J. Cortina (Eds.), *Big data at work: The data science revolution and organizational psychology* (pp. 115-157). New York, NY: Taylor & Francis.

How to cite this paper:

Sinar, E. F. (2018). *Data Visualization: Get Visual to Drive HR's Impact and Influence.* Society for Human Resource Management (SHRM)-Society for Industrial-Organizational Psychology (SIOP) Science of HR White Paper Series.