

THE EYES HAVE IT

Why your career needs better data-visualization skills

By Nathan Jaye, CFA

We live in a visual world. With the increasing availability of big datasets, visual data is becoming more and more essential for providing investors with powerful, actionable insights. Mark C. Hoogendijk, CFA, founder and managing director of E8 Consulting Asia, specializes in data analytics and data visualization and has presented to such local member societies as the Hong Kong Society of Financial Analysts and CFA Society Singapore. In this interview, Hoogendijk discusses the advantages and efficiencies of using visual data, the available software packages and training sessions, and the ways in which big data is changing hiring practices in the profession.



Mark C. Hoogendijk, CFA

How is big data changing the investment profession?

In the finance industry now, there's an increased emphasis being placed not only on being a quant but also on being a data scientist. If you look at investments analytics, more and more is being requested of the analyst in terms of how much data they analyze.

Historically, analysts look at the balance sheet. They interview the board. They compare similar companies in the industry. These days, research analysts, portfolio managers, and hedge fund managers have the ability to access many other datasets. Given that datasets are becoming bigger and bigger, many tools that analysts use (such as Excel) are coming up short. People are slowly understanding that they need more powerful programs. So, they're starting to hire programming analysts.

How fast is data analytics growing?

It's definitely picking up momentum. We're actually advancing very fast. Bloomberg recently

posted an article about the huge number of hedge funds that are now deploying machine learning—deep-learning AI—in their investment tactics.

Firms are hiring data scientists more than ever. These are people who have a quantitative background and also a statistical background—who are also able to do data mining, to work with huge datasets. One of the best examples of this is Renaissance Technologies, given how well they've done. This is a trend that will definitely continue and at a much faster pace than we're currently experiencing.

Why is visual data so helpful?

It's directly related to data analytics. Data visualization illuminates relationships within the data. Most people are visual. You can see that on risk dashboards—it's all visual these days. Quickly, with one glance, you can see where your risk is in the portfolio. You can direct your attention to those areas that are increasing in risk or try to understand why certain stocks are not performing as you expect.

How can visual data help investors make decisions?

If you're a portfolio manager, you're there to create alpha. You're there to beat the benchmark. With data visualization, you can quickly see those stocks providing the alpha in your portfolio.

Another case is showing the unexpected volatility of a stock. You can search different volatility measures and compare them across industries. This quickly highlights those stocks that are showing too much volatility. Also, if you look at a large portfolio, you can quickly see the drawdowns. At a glance, you'll see the depth of the drawdown and when it started. If required, you can also capture when the index went into

drawdown and how each stock in the index started its drawdown in perspective.

These kinds of visualizations quickly show the fund manager what's happening in the portfolio. The same can be said from a risk-analytic perspective.

What's the goal of visualization?

The ultimate goal is to gain actionable insights. Insights in themselves are great, but the most important thing is being able to act on it. In a visual risk dashboard, right away, the risk manager can see the issues in their portfolio. Is the risk building up? Where is it coming from? Which geographical locations, sectors, or industries are contributing to the increased risk? Is a particular asset class showing signs of distress?

If a particular sector, industry, country, or region in your portfolio is showing stress signals, then, as a risk manager, you want to take action. You might want to reduce your exposure. You might want to put a hedge on. The goal of visualization is to directly alert the users to potential areas where they can take action right away.

Is it easy to make visualizations?

Data visualization is almost a science in itself. It's a whole field these days. Edward Tufte is one of the leading pioneers. Learning data visualization definitely requires some training. It's not easy to come up with a graph where the viewer immediately sees the relationship. You have to choose the right datasets, the right graph types, and the right aesthetics (color, shape, etc.).

Illustration by Kelly Alder



People are now spending more time on learning proper data visualization. It's not just at the end of the course anymore—the last half afternoon on data visualization. Now, there are dedicated courses purely on capturing visual information.

How do we move from data analytics to visualization?

Do you recommend hiring a full-time person?

I would take most of the people who are spending time on data analytics and train them on data visualization. If you simply hired one or two specialists in data visualization, they might not directly understand what the risk manager

or portfolio manager is looking for. Visualization is useful for the people working with the data. But, it's also important for the portfolio managers and risk managers to understand what is possible. So, the quant or the data analyst should definitely be spending more time on training in data visualization.

What are some helpful tools?

First, you have commercial packages that are quite useful. Tableau, Qlik, Microsoft Power BI, Highcharts, and TIBCO Spotfire are a few such software packages. They all offer powerful data-visualization tools.

Alternatively, one can choose among open-source software. One of my favorites is the ggplot2 package, within the R programming language. Plotly is another I would also highly recommend.

Where can we go for training?

Those interested can spend one or two days doing a data-visualization course. There are many companies that provide these trainings. There are online courses, too. You could do a course on Coursera or edX. They all have courses related to data visualization.

You can do this effectively with a bit of training?

Absolutely. The thing is, most people who work with Excel or PowerPoint are already doing data visualizations. They are already presenting information in pie charts or bar graphs. It's just becoming more of a dedicated science, because we're using more and more data.

Data visualization has become more difficult, because the amount of data is increasing. If you've got 100,000 data points, it's easy to over-plot, and it's going to be one big blur. You need to see through the density to the main story and then show that to your audience. But, it's something that anybody can learn.

You recommend using R, an open-source programming language. Why?

R originated in New Zealand and was built by two professors there. It originally was meant to be a statistical programming language and was often used by students in statistics and hard sciences. Over time, however, many people from other industries have implemented R into their daily workflow. I use R specifically for investment analytics on big datasets. R is also known to be used for credit analytics.

The true value of R comes from a large community of experts creating their own packages within R. This means that you can leverage what specialists in your industry have already built. When someone creates a number of plotting functions, they can assemble that into what they call a package and then share with the larger community. It goes into an archive, where anyone then can access the packages. This is what makes R so powerful.

By now, R contains close to 10,000 packages on CRAN (Comprehensive R Archive Network). Every package is dedicated to a specific area. For example, a package can

be dedicated to portfolio analytics or performance analytics. In regard to data visualization, there are at least 40–50 dedicated plotting packages, from highly specific to highly flexible.

What's the ggplot2 package?

ggplot2 is a very famous R package based on *The Grammar of Graphics*, a book by Leland Wilkinson. It gives you the ability to capture multidimensional data very, very easily. Every time I teach it, participants are amazed and baffled by how easy it is to interpret high-dimensional data with this package.

Does R work with other languages?

Some of the R packages give you the ability to connect to other languages. You could connect to Python or to C++. There's a package called Rcpp, which gives you the ability to integrate C++ code and use that in the R work environment.

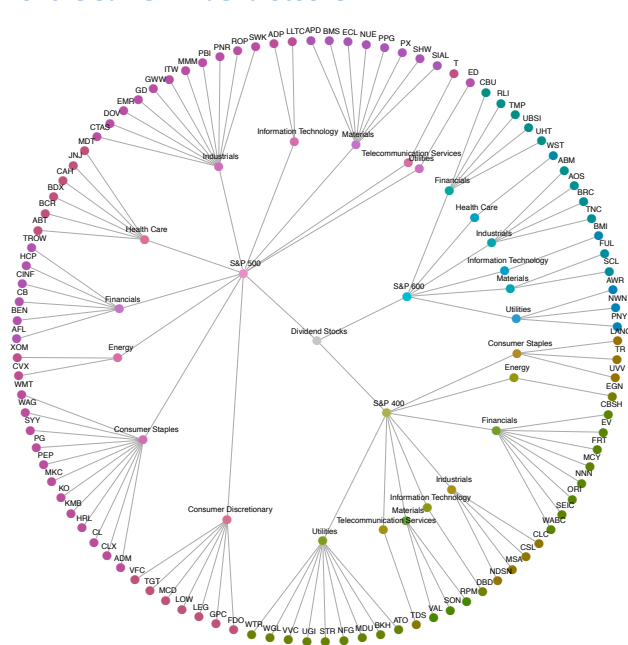
If you look at R together with RStudio (a free, integrated development environment for R), it's an extremely powerful tool set. It takes you from importing and cleaning your data all the way to creating interactive market reports. It also lets you create powerful web applications.

That's one of the reasons I have a preference for R at the moment. If I want people to have the ability to interact with the data and do their own analysis, I can build a web application with R without having to learn HTML, CSS, or JavaScript.

Why is open source powerful for investment firms?

Open-source software lets you gain huge efficiencies with daily workflow. Open-source software is software in which

Treemap Visualization of Index & Sector Split for the same Dividend Stocks.



Graphics: Martijn Tennekes (2017). treemap: Treemap Visualization. R package version 2.4-2. <https://CRAN.R-project.org/package=treemap>

you can see all the code. It's not a black box. You can access the code; you can change the code; you can do almost anything you would like (although it's important to read the various licenses). Most importantly, you do not have to reinvent the wheel.

For example, I spent about 13 years in investment banking, working within the structured derivatives space. I was frequently asking my team to build efficient frontiers and analyze what happens when we incorporate various derivative solutions. Somehow, it felt we were always starting from the ground up. Given the silos within banks, this was probably happening a lot within many different departments. If you use open-source software, most of the building blocks are already there. Many applications for the financial industry have already been built.

Leveraging on open-source software is a great way to reduce the time spent on the data analytics pipeline. And, the software typically will be free. You download directly to your computer and start working with it right away. A lot of people are not aware that this is available, and that's the biggest problem. Maybe they've heard of it. Maybe they've briefly worked with it at university. But, they've never really looked into it and the depth it covers.

In my talks to CFA [local member] societies, I show members how powerful it is to use open-source software like R, Python, or Julia. Members will say, "Wow, that was very easy." With one line of code, they can do something they've always wanted to do through Excel but which might typically take multiple hours or even days. That's why I developed these courses: to train CFA Institute members in using open-source software.

What are the basic steps in visualization?

One of my first steps in understanding data is to make a very quick plot. Just quickly plot a graph—what does the data look like? It can be a histogram, box plot, or scatter plot. You're just trying to get a sense of whether there is any structure in the data. Are there any outliers? From there, you can do more advanced algorithms.

The fact that you have more and more data isn't a problem. Anything can be done automatically these days. Every morning, for example, I have code running that downloads all the data that I require and starts going through the statistics, starts running the numbers that I programmed, and quickly does the visualizations. All those visualizations will be then saved at a specific location. I have direct access to all the different graphics. Given the use of web applications, it's very easy to capture 30–40 graphics on your screen and click through it very quickly. I visualize almost all my datasets, as long as I know that I can get actual insights from them.

Where can we access databases?

It's crucial for investment managers and risk managers to have absolutely clean data. Depending on the type of analysis you're doing, you might link to Bloomberg or FactSet. There's an R package that lets you connect your Bloomberg terminal and download data from Bloomberg (the Rblpapi

CLIENTS WANT TO HAVE A QUICK OVERVIEW OF THEIR PORTFOLIO, AND IT NEEDS TO BE VISUAL. ... IN TERMS OF A PORTFOLIO MANAGER OR HEDGE FUND MANAGER, MOST OF THEM ALREADY USE DATA VISUALIZATION IN MANY FORMS, BUT THERE IS STILL A LARGE DEMAND TO IMPROVE.

package). There are also packages that let you link to Thomson Reuters datasets.

You can also link to publicly available datasets. FRED (Federal Reserve Economic Data) is a well-known database. It's basically the Federal Reserve Bank of St. Louis. R can directly link to FRED and download data from there. If you're looking for economic variables or need information on unemployment or consumer spending in the US, you can link R to that database. FRED has more than 400,000 sets of data.

You could also link to the ECB. This is another reason why R is so powerful. You can connect to both commercial and publicly available datasets.

How quickly can one gain proficiency?

That really depends on each individual. Obviously, if they have previous experience with coding—perhaps in C, C++, or C#—that will significantly help. If they have no coding experience, they can still pick up the basics of R within a sufficiently short time.

The main ingredient is using it on a daily basis. This is what happens when people learn Excel from the first day. If you use something every day, you pick it up quickly. It's the same with everything—learning a language or playing sports. Ultimately, you just have to practice.

I definitely recommend DataCamp, an online learning platform for R and Python. They've hired the top programmers and specialists in that area. It's very interactive, meaning that you spend a lot of time actually coding on the presented data.

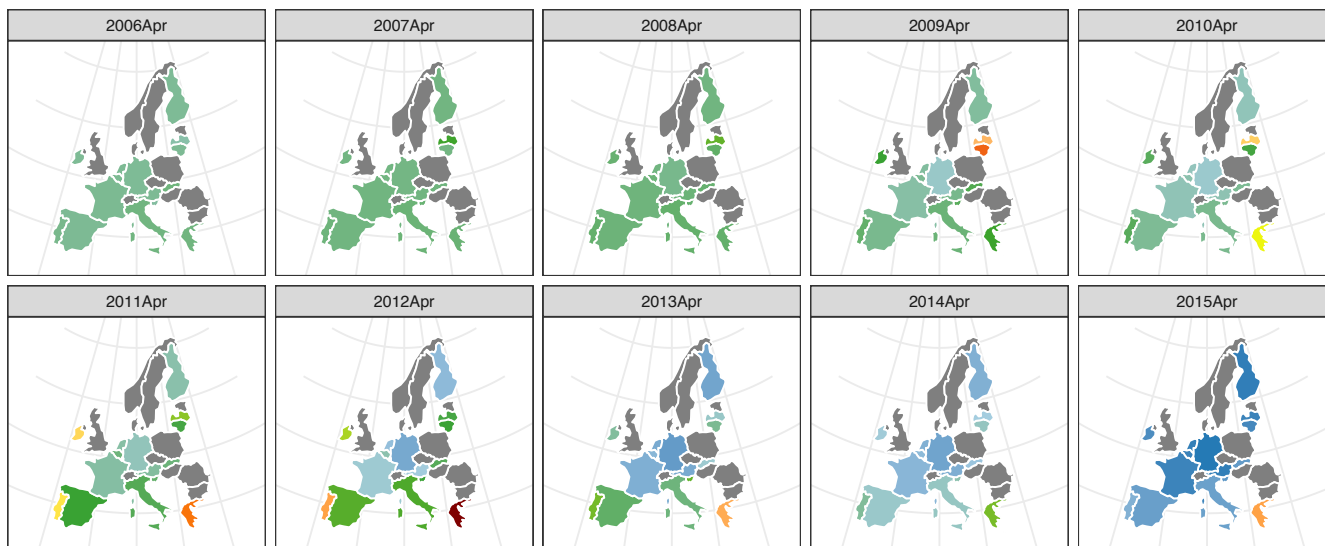
One of the easiest ways to learn is to give yourself a project. You say, "Okay, this what I want to do" and just start working on it.

It's definitely much faster than people expect. Many people are kind of driven away when they think of programming languages. But, when I run my courses, most people are up and running within two days. They're able to input their own data, analyze it, plot it, and create an efficient frontier and perform back-testing on their portfolios, for example.

How important are data skills in an investment career?

It's going to be very important. Creating value for your clients is becoming ever more difficult. It's important to utilize

ggplot2 Visualization of European Debt Crisis Changes in Interest Rates



Graphics: H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

10Y Rates
5% 10% 15% 20%

all the data available. Not only standard economic and fundamental data and technical analysis but also consensus data, sentiment analysis, and linking to other public datasets that potentially could relate to the underlying investment. Given that there's so much data out there, it's increasingly important to learn these skill sets.

How can visual data help with clients?

In the retail industry, there's a fair bit of data visualization already taking place. Log in to trading systems; right away, when you're looking at your iPhone, you'll see a visualization of your portfolio and your positions. Perhaps you'll see a visualization of various recommended portfolios based on your peers, or it might recommend a change in the portfolio if you believe certain events could take place. Visuals keep it simple. It gives a quick overview, one summary in a nice dashboard.

It also depends on how you define client. If you're a life insurance company that has 10 different asset managers, you can quickly benchmark them by looking at the data. You can plot their performance to see if they're generating alpha or not and in which areas they are underperforming or where the manager may not be in line with the mandate.

With visual data, you can easily track your fund managers on a daily basis: their positions, their change in positions, how that affects your capital position, and how that potentially matches with your liabilities. For the client, there are many benefits.

Has presenting data visually become an industry standard?

In short, yes. Clients want to have a quick overview of their portfolio, and it needs to be visual. For retail, that may be more directly related to the investment company or maybe

online trading platforms or private banks. In terms of a portfolio manager or hedge fund manager, most of them already use data visualization in many forms, but there is still a large demand to improve.

Where is this field going in the future?

The field keeps on advancing, especially when using high-dimensional data for global portfolios. There's already been a fair amount of research on using machine learning algorithms to find different ways of creating a robust efficient frontier, trading algorithms, or creating unique stock screeners. People are going to be using machine learning more and more as it becomes more accessible to the general investment industry.

Machine learning is already doing part of the job of financial advisers. I think that can go very far, especially in the area of robo-advisers. For example, through the use of cameras and other tools, we may be able to analyze the body language of the clients. We'll better understand how they react to certain questions and also to judge whether they're risk averse or not. How will they react if they are down 10% or 20% or 30% on their portfolio? (For more on accurately predicting the emotional reactions of clients, see "Hero Types, Roller Coasters, and Dark Zones" on page 38 of this issue.)

Algorithms will become smarter and smarter, quicker and quicker. We've seen it already in high-frequency trading. Five years down the road, a lot of people will be wanting software that is related to data analytics. Where exactly will it bring us? That's a difficult question to answer.

Nathan Jaye, CFA, is a keynote speaker and member of CFA Society San Francisco.