# Why Two Y-Axes (Y2Y): A Case Study for Visual Correlation with Dual Axes

line 1: 1st Given Name Surname
line 2: *dept. name of organization*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 2nd Given Name Surname
line 2: *dept. name of organization*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 3rd Given Name Surname
line 2: *dept. name of organization*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

***Abstract*—Many researchers and authors recommend against dual axis charts. We provide a case study of timeseries correlation analysis in financial services where dual axis charts superimpose series to facilitate observation of local patterns. Fine-grain patterns cannot be otherwise perceived using a single axis charts, normalization charts, scatterplots nor statistical correlation analysis.**

***Keywords—visualization, timeseries, correlation, leading indicators, normalization.***

## I. Introduction

Within the visualization community, some researchers are dubious of the benefits of multiple axes displayed on a single chart, such as a line chart with two y-axes. This position has been well reasoned with examples of misleading charts, misleading patterns and problems with usability. This has led to visualization implementations that deliberately do not allow dual axes.

We contend that earlier research has over-generalized the recommendations against the use of dual y-axis charts. We illustrate this with a case study from professional financial services, where multiple y-axis timeseries charts are common. Further, we argue multiple y-axis charts may be valuable beyond finance and more broadly applicable for deeper data analysis.

## II. Why Dual Axis Charts are Bad

In *A Study on Dual-Scale Data Charts*, Isenberg et al points out tasks such as comparing distances and slopes are difficult when two different scales are used [1]. Position, length and slope are tested in five chart variants where a *single* timeseries is split into two ranges; where each range has very different data characteristics. These are then visualized as a regular chart (single x and single y axis); two charts side-by-side (broken); a single plot area with multiple x and y axes (superimposed); a single plot wherein different scales apply to different regions (lens and bifocal); and a chart of the full range with a chart with a subset of the range (cut-out). They conclude that superimposed charts – i.e. with dual axes – should be avoided.

Stephen Few also argues against dual scaled axes in charts [2]. Few provides a well-reasoned argument against dual-axes bar charts. With regards to line charts, Few argues against dual axes (e.g. lines crossing are visually salient but are not semantically meaningful), and instead normalizes lines as a percentages and plots them with a single common scale.

Kaiser Fung, also discusses the misuse of dual axis charts, with many examples at the weblog junkcharts. In addition to the comments from other researchers, Fung points out that each axis can be manipulated independently to accentuate trends and crossings to potentially create a misleading message [3].

Hadley Wickham, the influential author of the visualization library *ggplot2*, was asked about adding support for two y-axes at a public seminar regarding visualization in 2016 [4]. Hadley responded to the effect that there should never be two y-axes on a line chart or scatterplot and that he would never add it to ggplot2 although the open source community might be inclined to add it. This is further elaborated in a public posting by Wickham [5] citing many issues with dual axes charts, including references to Isenberg's study, Few's analysis, Fung's examples, and additional aspects, such as the lack of being able to map any given pixel unambiguously to a single value in data space. Other authors have also written and illustrated potentially confusing dual-axis charts [6,7].

Given the strength of evidence and the reasoning of researchers and professionals, one may conclude, such as Wickham, that multi-axes should never be used. However, dual axis charts are easily accessible in some software such as Microsoft Excel, and multi-axis charts are available in some professional software, such as financial charting software which may allow up to eight y-axes on a chart (e.g. ChartIQ, Bloomberg, Eikon, FactSet, etc.), such as the line chart with 3 y-axes in Fig. 1. If multiple axes charts are bad, then how are financial professionals using charts to facilitate their analysis and trading strategies, using multi-axes charts?

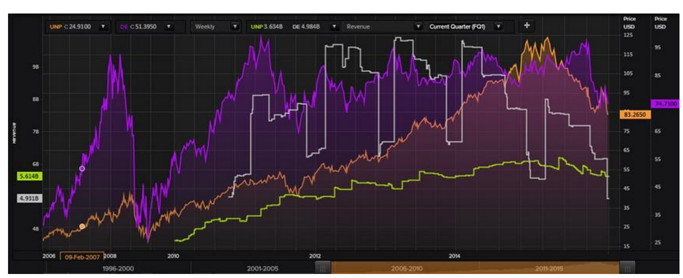

Fig. 1. Sample Refinitiv Eikon chart with four lines and three y-axes.

We contend that prior research has over-generalized findings. Isenberg et al's study used charts with multiple *x*-axes in their study – not multiple y-axes. They used a singular timeseries split into two date ranges, not two timeseries with the same date range. Furthermore, dual-axes may be effective for some types of tasks: for example, the analyst may be more interested in correlation than slope or distance.

Instead, we provide a case study from financial services, with a dual-axes timeseries chart used for correlation analysis. Users need to rapidly assess and visually confirm ongoing interdependence among multiple timeseries, which is greatly facilitated by the use of multiple y-axes on a singular chart.

## III. FINANCIAL TIMESERIES ANALYSIS

In financial services, high-frequency timeseries charts are heavily used by professionals. These include *specialized chart types* which have evolved over *centuries*, such as candlestick charts [8], kagi charts and renko charts [9], point-and-figure charts [10], market profile charts [11], equivolume charts [12], and so on. In addition to timeseries charts, scatterplots, bar charts, pie charts, heat maps, 3D plots and so on, are also used for analysis of particular aspects of timeseries analysis.

These charts and visualizations are used by financial professionals every day in their work. Some professionals explicitly study and earn the designation of a Chartered Market Technician (cmtassociation.org).

### *A. Timeseries analysis, spreads and correlations*

One common task in financial timeseries analysis is the comparison of two or more related timeseries. For example, a spread trade focuses on the relative movement of two financial instruments: the wager is on the relative movement of the two instruments, not simply the price change of the individual instruments. A common business objective of the financial professional is to identify and understand the supply/demand imbalances across related markets. Similarly, an economist may analyze the interrelationship between commodities produced within a country and its currency to assess exports and economic growth.

## IV. OIL VS. CANADIAN DOLLAR CASE STUDY

For this paper, we consider one case study that illustrates the benefit of multiple y-axes for a single chart superimposing two independent timeseries for analysis, in addition to scatterplots, and statistical analysis.

For this example, we use the price of oil and the Canadian dollar from the period January 1, 2009 to December 30, 2016 for 2087 observations for each timeseries. Economists, commodity and currency traders regularly compare the price of oil and the Canadian dollar to gain insight into the strength of the Canadian economy. Energy products represent 9.1% of Canadian GDP, 21.6% of its total exports and Canada is the fourth largest global exporter of oil.

### *A. Correlation via line charts, scatterplot and statistics*

During this time period, oil had a low price of $26.21 and a high of $113.93, more than four times higher than the low. The Canadian dollar ranged from a low of $0.69 to a high of $1.06, only 1.5 times the low. As shown in Fig. 2, a single y-axis renders the Canadian dollar as a flat orange line, given the difference in the magnitude of the nominal values.

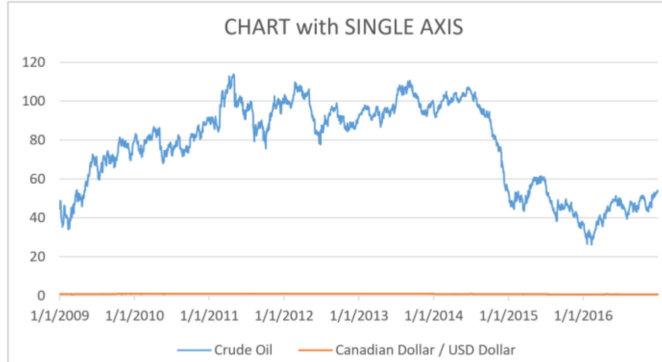


Fig. 2. Single axis. Given the difference in magnitude, one series is flat.

Alternatively, both series could be normalized to a value of 100 on the start date, as shown in Fig. 3. This aligns the starting point of the two timeseries, however, the magnitude of the change in oil price differs greatly from the magnitude of the change in the Canadian dollar, again rendering the Canadian dollar to a nearly flat orange line. Furthermore, the normalized data adds burden on a trader who needs to convert back to market prices prior to executing an order.

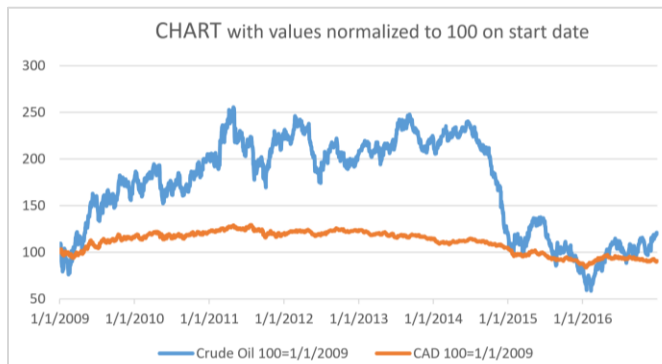


Fig. 3. Single normalized axis. The magnitude of the change renders one series nearly flat.

Small multiples are an array of charts, each chart in a separate panel [13]. Splitting the above into small multiples are shown in Fig. 4. Clearly, both series demonstrate similar patterns as they start low, trend upward, drop to new lows, then somewhat recover, but it is difficult to see if these movements are coordinated: do local highs and lows occur at the same date? Do uptrends and downtrends occur at the same time?

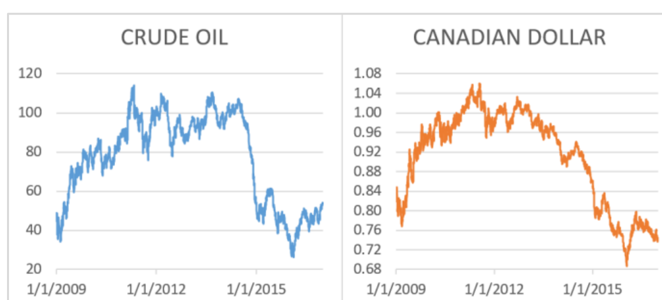


Fig. 4. Small multiples. The two charrts are side-by-side.

Horizon charts also split charts into separate panels [14]. These are stacked vertically with a shared horizontal axis, and a consistent vertical axis, stacking intervals for large movements. However, the large different in magnitude makes it difficult to see time related patterns such as the normalized horizon charts shown in Fig. 5. The most salient visual aspects the additional dark bands in crude oil, and the longer negative range in red for the Canadian Dollar vs the shorter but deeper negative for oil. Neither of these observations are relevant to the tasks for spread trades and related movement.

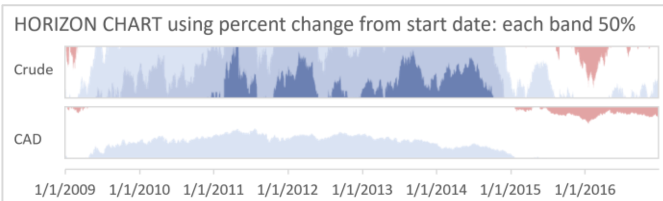


Fig. 5. Horizon chart for the same data. Bands and colors are most salient.

Another potential solution is to vertically stack the series using separate Y-axes and a shared X-axis as shown in Fig 6. This more clearly shows the temporal alignment of the activity in the two series. However, the challenge of this layout is the eye must move up and down continually trying to see which of the prices are moving together or diverging or if one security is leading and the other is lagging. Fine comparisons across a few points is difficult.

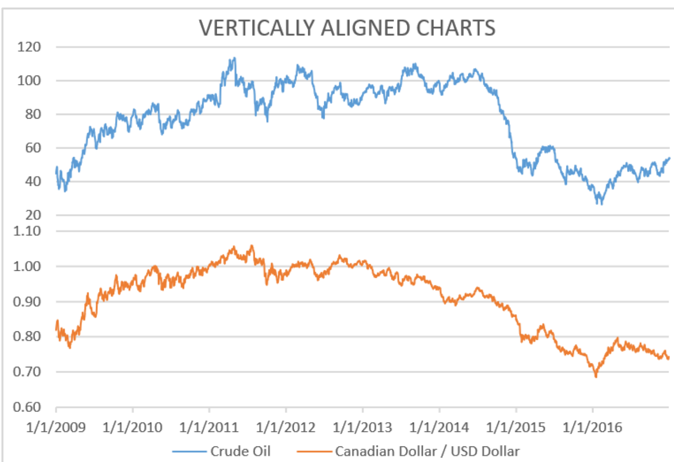


Fig. 6. Vertically aligned charts facilitate visual comparision over time.

Instead, a scatterplot can be used to visually indicate the relationship between the two series, plotting the Canadian dollar on the x-axis and Crude oil on the y-axis, as shown in Fig. 7. A strong linear correlation exists as indicated by points mostly lining up along a diagonal (shown with a regression line in orange). The exception is a cluster of points that deviate from the regression near the top right corner of the plot.

This strong correlation and deviation are of interest to the financial analyst. Financial securities may move together for a period of time but then change: this may be referred to as a *correlation breakdown* or as a *regime change*. A spread trading strategy is susceptible to risk and significant financial losses in the event of a correlation breakdown.

Experienced traders are attuned to track various metrics and adjust their pricing relative to these shifts. For example, an interview with an oil trader said, “On the trading floor, we’d all be closely tracking inventory data for a few months, and then it would change, we’d all switch to tracking storage, or refinery outages, or some other metric as a leading indicator driving our pricing.” [15]. A recent example is the decoupling of West Texas oil and Brent oil, typically highly correlated but recently decoupled when West Texas oil became negatively priced due to lack of U. S. storage capacity [16]. The underlying drivers of a change may not be explicit or obvious, so detection of these changes via analysis of relative price movement is important.

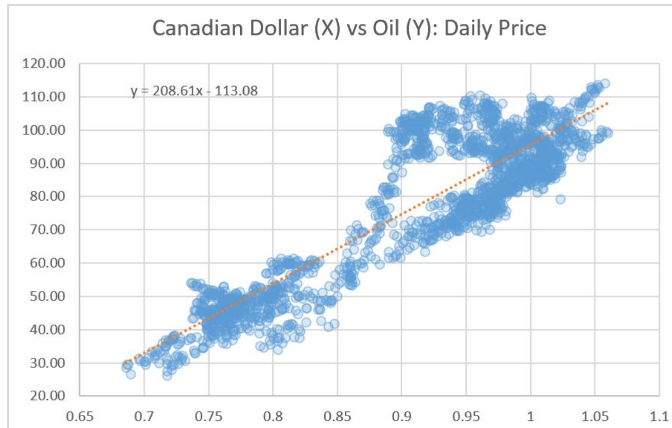


Fig. 7. Scatterplot. Most points tend to align near the regression.

The time sequence can be indicated within a scatterplot, for example, by joining successive points with a line (i.e. a worm plot) or by color-coding different time periods, as shown in Fig. 8. In this example, each year of data is assigned a different color, starting in yellow (2009), moving through orange, brown, purple and blue (2016). Regression lines are also shown, and the deviation upper right largely occurs in 2013 (brownish points with a blue regression line), when the correlation between oil and Canadian dollar are negative.

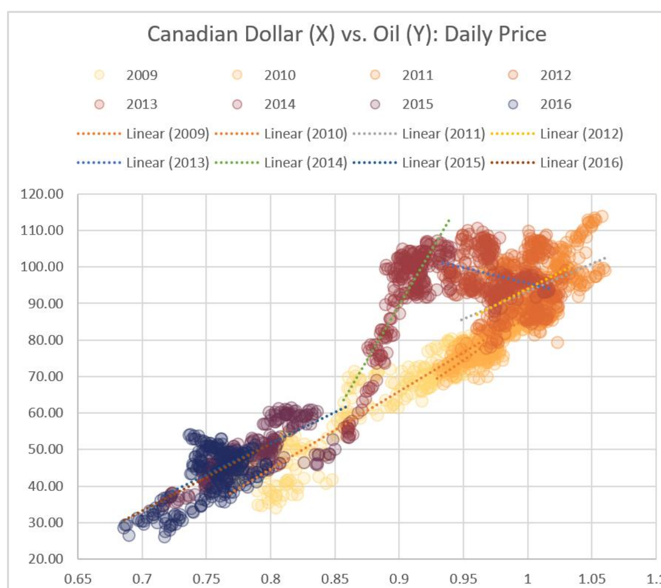


Fig. 8. Same scatterplot, with time encoded via color.

While both the above scatterplots show the long-term trending of the timeseries together, the daily pattern is not visible: do these relationships exist at the weekly or daily level (or even intraday)? Instead, the daily percent change can be computed for each series and plotted as a scatterplot, as shown in Fig. 9. This scatterplot indicates that when the price of oil increases, typically the Canadian dollar increases and vice versa. But the relationship is not strict: sometimes the price of one increases while the other decreases or other variation. The linear regression (orange dashed line), indicates a tendency for the for the prices to move in the same direction.

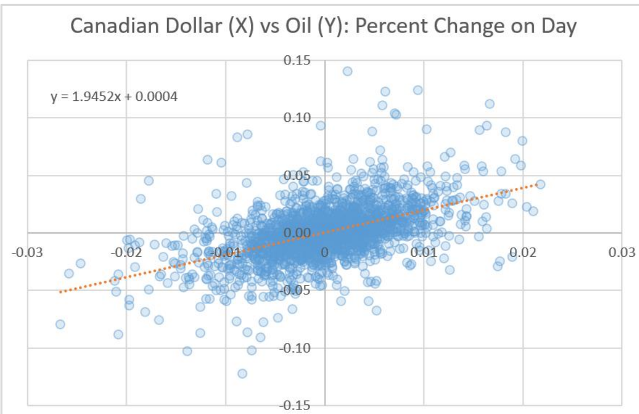


Fig. 9. Scatterplot using daily percent change in value.

The scatterplots confirm that there is a correlation between the securities both long term and short term, with occasional negative correlations – however, these analyses for insufficient as a trading tool. Finer grain detection of when the correlation changes start and subsequently revert cannot be discerned.

Instead, a statistical correlation based on the daily percent change can be computed. A new timeseries can be created using a short period for the correlation, such as a correlation based on the prior 20 days. A value of one indicates perfect correlation (all points would align to a perfect diagonal line in the scatterplot), a value of zero indicates no correlation (both prices move completely independently), and a negative value indicates an inverse correlation (the price of one security tends to move in the opposite direction to the other security). As can be seen in Fig. 10, the 20 day correlation typically fluctuates between 0.4 and 0.8, with some periods of inverse correlation, such as the sharp flip for a sort duration in early 2011, and the gradual shift to a negative correlation in late 2013-early 2014.

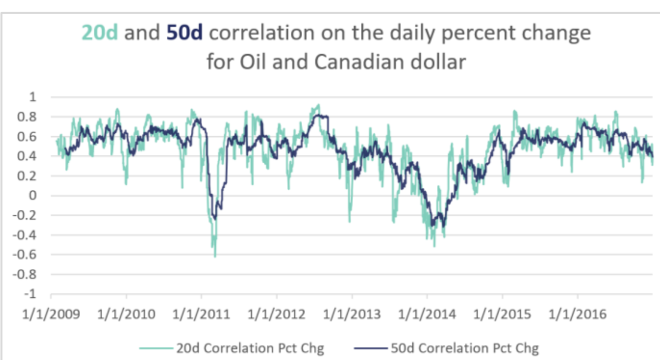


Fig. 10. Rolling statistical correlation of oil and Canadian dollar.

While scatterplots and correlation timeseries are useful, they are still insufficient. A correlation calculation is based on prior observations: a 20 day correlation requires 20 days of observations to reveal the correlation change. This is a *lagging* indicator, confirming the change, rather than a *leading* indicator which helps with early assessment of the change.

Instead, consider a line chart with dual axes as shown in Fig. 11. Starting at 2009, it can be seen how both securities move sharply down together, then bound back up, in many cases peaks and troughs are mirrored in each series. There is clearly a period of divergence starting in mid-2012 that becomes more accentuated by mid-2013 when the series sometimes are moving in opposite directions. During this period, oil is relatively stable showing some short upward and downward trends while the Canadian dollar is trending lower. Then in the summer of 2014, both securities quickly fall in unison, with oil losing more than ½ of its value, dropping from $100 to $44 in 5 months, with a corresponding drop (but smaller magnitude drop in percentage terms) in the currency from $0.92 to $0.79.

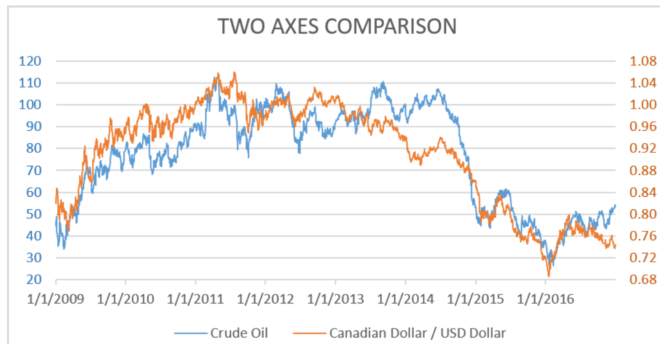


Fig. 11. Dual-axis timeries. See similar movement patterns.

By superimposing the two charts, with similar movement, the visual proximity of related points facilitates comparisons of local trends. If two separate charts were used (such as the small multiples in Fig. 4 or the horizon charts in Fig. 5), the proximity would be lost, making it more difficult to visually compare the movement in the same time period. The axiom "time is money" is truer in finance than most other cases and introduction of any complexity or delay in the evaluation will result in lost opportunity and money.

Another advantage of this method is it clearly communicates when the behavior in each security confirms/supports the action in the other security. That is, the financial analyst is interested in knowing both securities are making new highs or lows. Two critical pieces of information are derived from this analysis. First, it highlights which security is fundamentally driving the trend and which is responding. Second, if only one security demonstrates the strength to sustain the trend then this provides an early indication of a change in the supply/demand balance and the potential for the trend to stall or reverse. The core business objective in this use case is not to understand slope or magnitude, but rather to identify alignment in the two markets.

### *B. Alignment and Interaction*

Note that the price chart and the correlation chart can be compared to aid visual analysis. Fig. 12 shows a chart zoomed in to December 2010 – March 2011, with the price chart above and the correlation chart, aligned to a common x-axis below. From mid-December 2010 through to February 2011, there is a decoupling where prices diverge. In mid-December prices sharply diverge (short arrows), then oil stays flat and drops for a few months but the Canadian dollar continues on a fairly consistent upward trend during the same period (longer arrows).

The divergence in the correlation is visible in the 20 day and 50 day correlations on the lower panel. Prior to mid-December, the 20 day and 50 day correlations are stable and overlapping, both around 0.75, then begin to move down slowly. Note the sharp divergence in behavior in the timeseries in the first few days (upper panel) and the time lag for these divergences to become noticeable in the correlation timeseries (lower panel). Identifying corresponding points between panels can be facilitated by vertical grid lines in printed charts. Most financial charts will provide an interactive vertical tracking line to facilitate visually aligning times of interest.

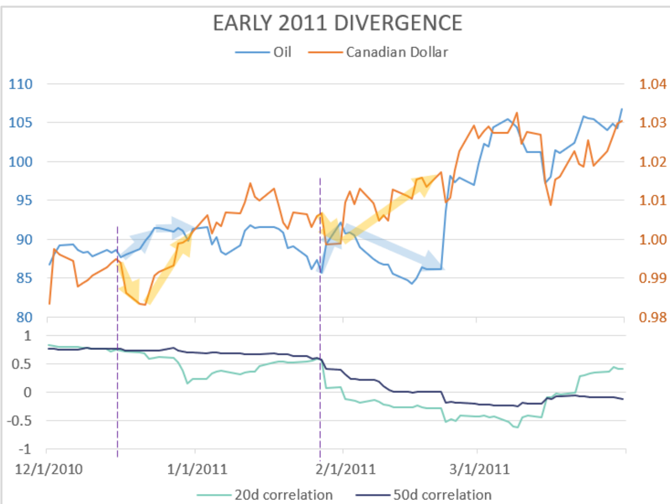


Fig. 12. Zoom-in of prior chart with annotations indicating divergence.

Figure 12 also shows annotations that a trader may add to a chart. In this example, large arrows have been added. Other annotations may be used to quantify the changes, such as measuring the magnitude of the change.

Note that the vertical axes scaling is typically automatic, as financial analysts may be viewing tens to hundreds of such charts per day. With interaction such as pan and zoom, the vertical axes automatically rescale based on the minimum and maximum values in the plot area, plus some padding, so there is no fiddling required to maximize the respective series.

Further interactions include linked selection between the scatterplot and the timeseries. For example, one can lasso a subset of points in the scatterplot to highlight corresponding points in the line chart. This can facilitate identification of anomalous movements that might not be visible, divergences of short durations, or confirmation that an apparent visual pattern in the dual-axis chart can be verified by the scatterplot.

### C. *Inverted Axis for Inverse Correlation*

Local proximity is also important for inverse correlation. When two securities are expected to be inversely correlated, one of the axes may be inverted, thereby more closely plotting the points for visual comparison. Credit default swaps (CDS) typically move inversely to related securities: in Fig. 13 the CDS is shown on the left axis (green), while the price of oil is shown *inverted* on the left axis (22 at the top, 88 at the bottom). As can be seen, the green and blue lines are plotted close to each other.

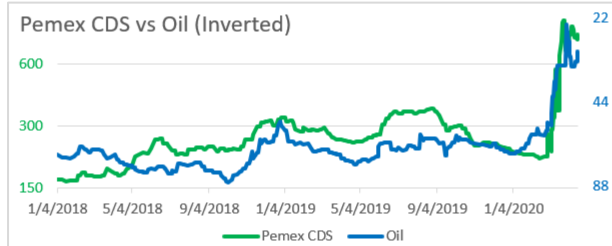


Fig. 13. Inversely correlated securities may be plotted with one axis inverted.

### D. *Multiple Y-Axes*

The prior example in Fig. 13 compares the price of oil to the CDS for an oil company. A multi-axis visual correlation may involve some combinations of commodities (e.g. West Texas Intermediate crude oil, Brent crude oil), petrocurrencies (e.g. Canadian dollar, Norwegian krone), oil companies (e.g. Exxon, BP, Shell) and derivatives such as puts, calls, or CDS associated with currencies or companies. A number of these may have completely different ranges and magnitudes, and thus a viewer will create a chart with many y-axes as shown in Fig. 1 to understand the interrelationship of all these markets.

## V. Supplemental Non-Timseries Example

The prior compelling examples for dual-axes are high-frequency timeseries charts. Are there other chart types where dual axes may be useful, in other domains?

### A. *Pareto charts*

Pareto charts are an essential chart type used in quality control. They are always dual axis, the left axis showing a unit of measure corresponding to bars on the plot; and a right axis showing cumulative total typically as a percentage (Fig. 14). Both axes share a common baseline (zero) and are scaled to align. The combination of axes allows the viewer to quickly answer questions of magnitude and proportion.

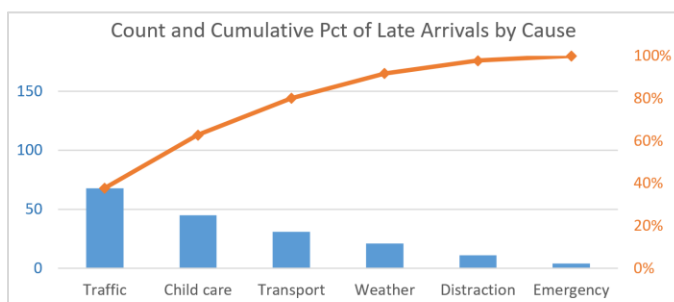


Fig. 14. Sample Pareto chart, which always have dual-axes.

### B. *Histograms*

Histograms are often overlaid and may vary in range. Dual-axes allows heights (or area under curve) to be normalized to facilitate comparison. For example, Fig. 15 shows website traffic analysis comparing ages of anonymous viewers (purple) to registered users (cyan). The right image, with dual axis, readily shows the relative similarity in heights for ages 20+ but a large discrepancy for age 1-20.

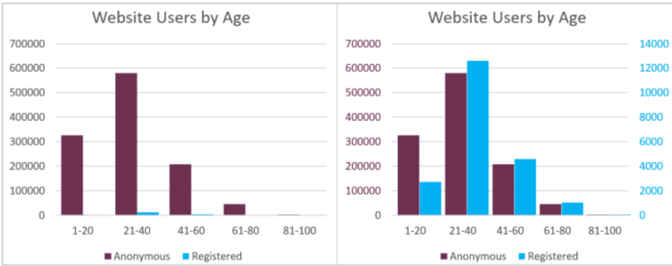


Fig. 15. Histograms, left single axis, right dual axis.

Figure 16 shows the same distributions with smaller bins (yearly). Left shows a single axis with counts: heights are not readable for the registered users. Middle shows data normalized by area under curve and a single axis indicating the percent of the given population. The difference in shape at young ages is immediately visible: there are no registered users under 16 years of age, the minimum age for registering.

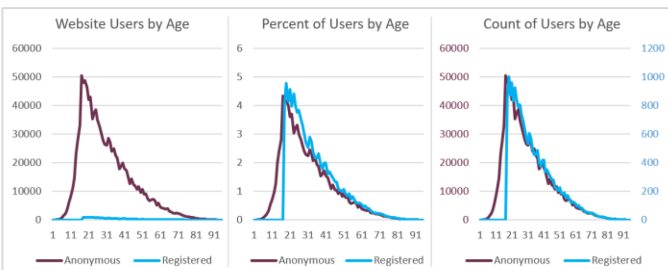


Fig. 16. Superimposed histograms, left single axis, middle normalized, right dual axis with counts.

The right chart shows counts on dual axes: while shape and patterns are similar to the normalized chart, there is additional information conveyed by the axes, i.e. the counts. In this example, it shows approximately 50,000 under-aged viewers one year below the legal age, whereas there are 1,000 registered viewers who are at the legal age. The combination of comparison and magnitude may be critically important, for example, for analysis of marketing controlled goods, such as alcohol, tobacco, gambling, vehicles and so on.

While normalization with a single axis can reveal the pattern, the dual-axis reveals both the pattern and the magnitude of the value without interaction. Reading without interaction is useful for applications that may need to export visualizations such as print or data science notebooks (e.g. Jupyter); for an audience to interpret the results in a seminar; and provides faster access to approximate values than moving the mouse for a tooltip – which is important in real-time systems requiring fast assessment.

## VI. Discussion

The financial case study shows dual-axes timeseries line charts facilitate the analysis of local relative movement between timeseries. The Pareto and histogram charts show the benefit of dual axes for reading values. While the Pareto and histograms could still be useful with only normalized data and interaction, the financial analysis example *cannot* be completed without the use of the dual-axis chart.

In both examples a singular value could be modeled and computed to capture the convergence/divergence of these series, however, this model would lose the underlying raw data: for example, in the financial case, it will be unknown which security caused the divergence. The visualizations exist in the context of the real-world knowledge of the professional analyst. For example, the financial professional may recognize a spike in one series as being a result by a one-off external event (such as news, government statistics data, commentary by an industry expert, and so on), and thus include this real-world knowledge in their consideration. Their mental model of the world supplements the data and visualization.

Also to note, with increased interest in data science, it will be more important to visually represent complex statistical topics in ways that aid perception and understanding. Correlations can be shown with scatterplots, computed as regressions and plotted as a series, as shown in Figs 6-9. However, the dual-axis chart can also intuitively aid this understanding, assuming the viewer is attending to the relative movement of the lines (not crossings).

The examples shown all assign the color of the line to the color of the axis labels. This is a recommended design pattern to assist the user when displaying data with multiple y-axes.

## VII. Conclusion

Data exploration is a central element of data science. Using charts with multiple Y-axes is a quick way to see if there is a data relationship which should be statistically analyzed.

The examples shown are all created and consumed by professionals: financial analysts, quality control specialists, and data scientists. They are for understanding correlations and patterns, to learn and gain insights, rather than to manipulate or persuade. They are not casual readers of a chart in a news story. As such, given the concerns of prior researchers, it may be that multi-axis charts should not be used for casual charts, but, are appropriate when used by professionals or for deeper analysis. Further user studies are required to determine where the boundary lies for use/non-use of multiple axes.

At a time when there is a greater need for increasing data literacy, the visualization community should be actively identifying familiar, usable, accessible techniques, to aid understanding of statistical concepts that underpin data science. Re-examining existing familiar visualization techniques – such as multi-axis charts – may be highly relevant to specific analytical tasks.

## References

[1] Isenberg, Petra, Anastasia Bezerianos, Pierre Dragicevic, and Jean-Daniel Fekete. "A study on dual-scale data charts." *IEEE Transactions on Visualization and Computer Graphics* 17, no. 12 (2011): 2469-2478.

[2] Few, Stephen. "Dual-Scaled Axes in Graphs: Are They Ever the Best Solution", *Perceptual Edge Newsletter*, perceptualedge.com/articles/visual_business_intelligence/dual-scaled_axes.pdf

[3] Fung, Kaiser. "The hidden bad assumption behind most dual-axis time-series charts." *Junk Charts: Recycling chart junk as junk art.* junkcharts.typepad.com/junk_charts/axis/

[4] Wickham, Hadley. "Nine visualisation challenges." *Data Visualization New York* at The New School, NYC, Nov. 17, 2016.

[5] " "ggplot with 2 y axes on each side and different scales", *Stack Overflow,* Nov 5, 2013. stackoverflow.com/a/3101876

[6] Rost, Lisa Charlotte. "Why not to use two axes, and what to use instead." Datawrapper, 2018. blog.datawrapper.de/dualaxis/

[7] Qui, Linda. "Chart shown at Planned Parenthood hearing is misleading and 'ethically wrong'". Politifact. Oct. 1, 2015. politifact.com/factchecks/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading-/

[8] Nison, Steve. *Japanese Candlestick Charting Techniques A Contemporary Guide to the Ancient Investment Techniques of the Far East.* New York Institute of Finance. 1991.

[9] Nison, Steve. *Beyond candlesticks: new Japanese charting techniques revealed*. John Wiley & Sons, 1994.

[10] De Villiers, Victor. "The point and figure method of anticipating stock price movements: complete theory and practice." *Stock market publications*, 1933.

[11] Chicago Board of Trade. *A Six-Part Study Guide to Market Profile.* 1996. web.archive.org/web/20170829185705/https://www.cmegroup.com/education/interactive/marketprofile/handbook.pdf.

[12] Arms, Richard W. *Volume Cycles in the Stock Market.* Dow Jones – Irwin, 1983.

[13] Tufte, Edward R. *The visual display of quantitative information*. Vol. 2. Cheshire, CT: Graphics press, 2001.

[14] Heer, Jeffrey, Nicholas Kong, and Maneesh Agrawala. "Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations." *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1303-1312. 2009.

[15] Sindel, Bill. *Navigating the Crypto Markets.* TradersExpo. New York. March 11, 2019

[16] Nawaz, Sehr. "Making History: Coronavirus and Negative Oil Prices" Global Risk Insights. May 29, 2020. globalriskinsights.com/2020/05/making-history-coronavirus-and-negative-oil-prices/.