



<b>Title</b>	Playing fair with time series data
<b>Authors(s)</b>	Cuffe, Paul
<b>Publication date</b>	2020-11
<b>Publication information</b>	Cuffe, Paul. "Playing Fair with Time Series Data." IEEE, November 2020. <a href="https://doi.org/10.1109/MPOT.2018.2868000">https://doi.org/10.1109/MPOT.2018.2868000</a> .
<b>Publisher</b>	IEEE
<b>Item record/more information</b>	<a href="http://hdl.handle.net/10197/10284">http://hdl.handle.net/10197/10284</a>
<b>Publisher's statement</b>	© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
<b>Publisher's version (DOI)</b>	10.1109/MPOT.2018.2868000

Downloaded 2026-05-01 23:44:16

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd\_oa)



© Some rights reserved. For more information

# *Figured Out*: Playing fair with time series data

Paul Cuffe,<sup>1</sup>

<sup>1</sup>University College Dublin

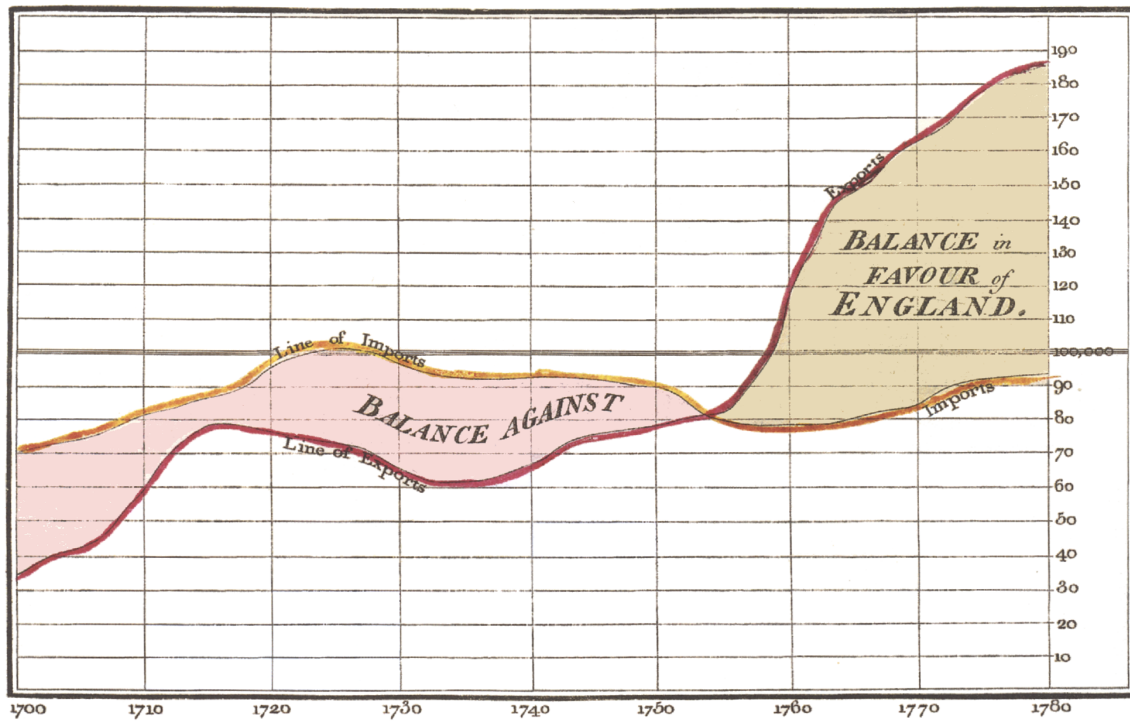
*Figured Out* is a recurring column about how to produce clear and effective data graphics. This installment considers a basic question: what is the clearest way to portray the fluctuation of a quantity over time?

## **Data graphics: the early history**

Data graphics had to be invented: the idea of using diagrams to represent numerical data only emerged towards the end of the 18<sup>th</sup> century. While graphs of mathematical functions had existed before this, it fell to a Scottish scoundrel, William Playfair, to popularise the idea of graphically representing *empirical* data in all its messiness.

Playfair led a colourful life: he was born in Scotland in 1759, participated in the storming of the Bastille in 1789, and over his career participated in a wide variety of business ventures across several countries, many of dubious probity, before dying in London in 1823. His career résumé testifies to his restlessness and pluck: he tried his hand in such disparate professions as draftsman, silversmith, pamphleteer, banker and journalist. Perhaps his greatest legacy is a book he published as a young man in 1786, the *Commercial and Political Atlas and Statistical Breviary*. Within this volume, he invented such foundational statistical graphics as pie charts, bar charts and line graphs. At last, people had graphics to stab at as they argued the topics of the day!

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



*The Bottom line is divided into Years, the Right hand line into L10,000 each.*  
*Published as the Act direct, 14<sup>th</sup> May 1786. by W<sup>m</sup> Playfair* *Neale sculpt 352, Strand, London.*

Figure 1: An early line chart visualisation of time series data by William Playfair: then, as now, the economics of *balance-of-payments* could be a contentious issue!

An example of Playfair’s work is given in figure 1: before diagrams like this, such trade dynamics could only be inferred by labourious squinting at long columns of tabulated numbers. In the book from which this graphics is taken, published in 1798, Playfair lays out his motivations for using visual representation, in succinct words that are still relevant today:

*“As the eye is the best judge of proportion, being able to estimate it with more quickness and accuracy than any other of our organs, it follows that wherever relative quantities are in question [...] this mode of representing is peculiarly applicable; it gives a simple, accurate and permanent idea, by giving form and shape to a number of separate ideas, which are otherwise abstract and unconnected.”*

This remains the core goal of data visualisation: to transmute opaque and lifeless numbers into a clear, concrete and meaningful display. A line graph, as in figure 1, is a tangible and direct way to show the variation of a quantity over time. All the supporting elements in a graphic – the axes, the annotations, the colour scheme – should be carefully chosen to foreground the quantity fluctuating over time, and to remove confusing layers of abstractions.

In the next section, we will consider some practical advice on how to achieve such clarity.

## **Principles for plotting time series data**

**Labelling the time axis:** Defining suitable axes is often the most important step in creating a legible data graphic. The axes should serve as a contextualising frame for the data, designed to facilitate interpretation. Time should nearly always be placed on the horizontal axis, increasing from left to right. To make this axis easy to parse, it is essential to label it in a suitable time format. So, if you are showing a value fluctuating over the course of the day, be sure to label the horizontal tickmarks in the 24 hour clock format (HH:MM) and at logical intervals (perhaps every eight hours?). If working with monthly data, tick off the start day of each month: these spacings will be somewhat irregular, but that’s okay! The internal format of your raw data is

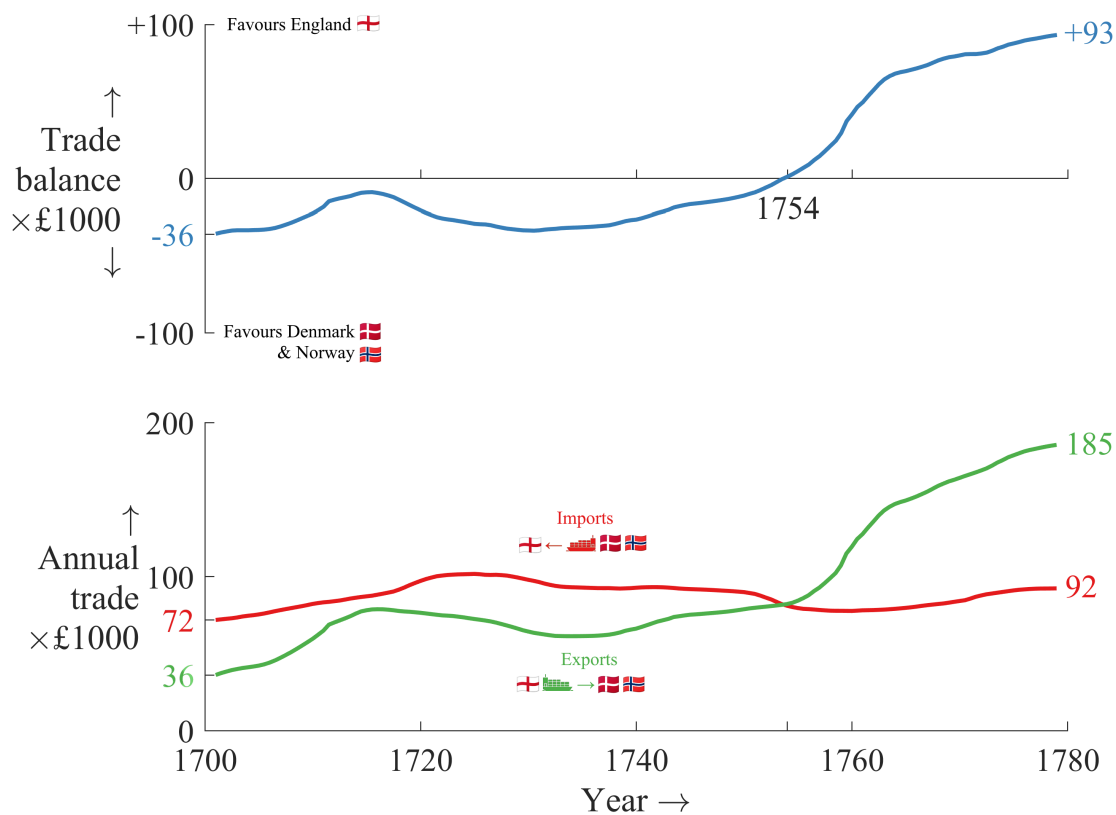


Figure 2: The author’s modern reworking of Playfair’s early data graphic portraying trade between England and Denmark & Norway

irrelevant when plotting: even if your data comes encoded with an awkward scheme, such as number of hours elapsed since the start of the year, it is essential to transform it to meaningful units for presentation.

**Showing multiple traces:** Very often, you will want to compare how different quantities vary over time. One obvious way to do this is to plot a distinct trace for each quantity. Generally, it is best to use a colour coding to differentiate each trace. These colours should be chosen carefully: ideally, each colour will be distinctive even when printed in black and white, and will have a mnemonic relationship to the quantity it depicts. Exotic line styles (e.g. dashings,

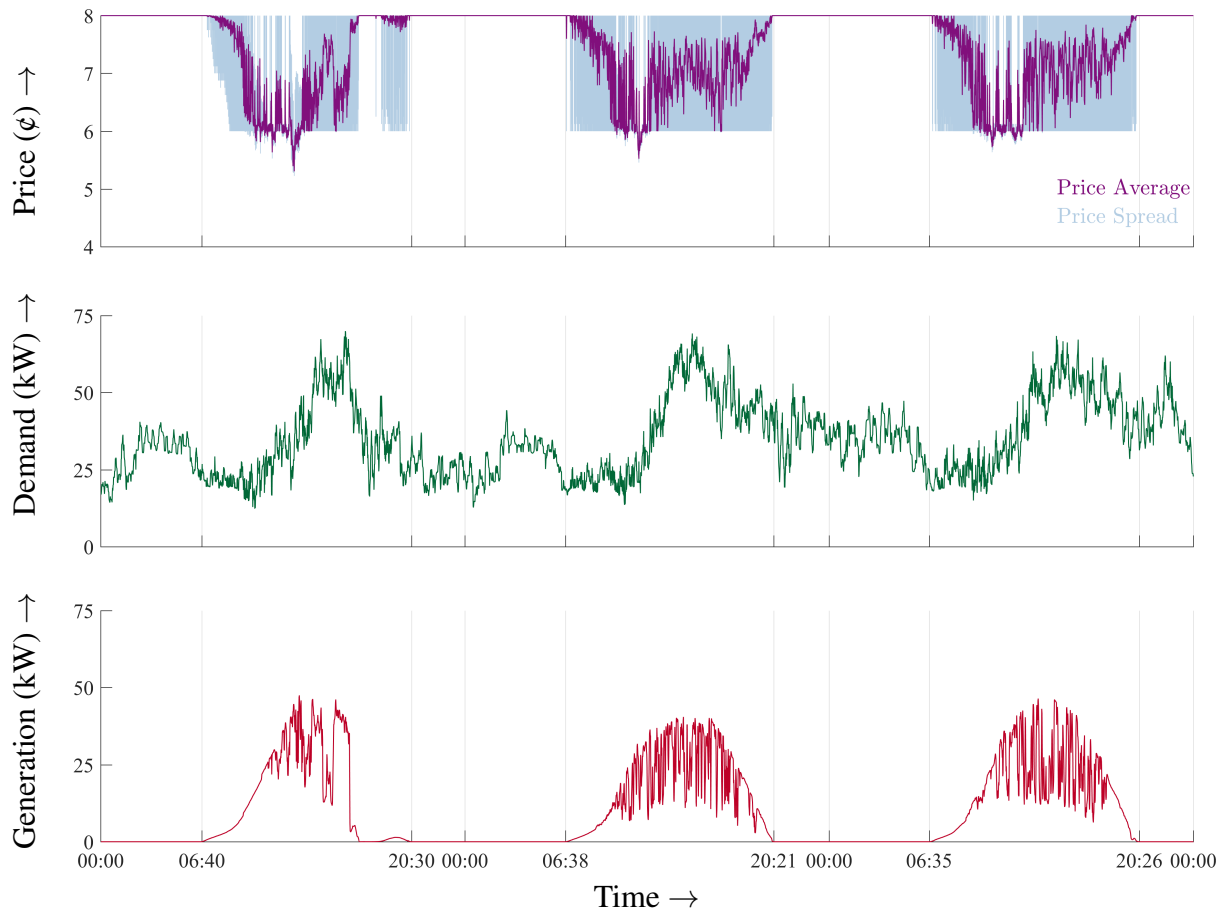


Figure 3: An ensemble graphic showing the operation of the local electricity marketplace over the three test days

diverse thicknesses, fiddly point markers) should be avoided, as they tend to make a graphic look too busy and cluttered. Direct annotation of each trace is usually sufficient to maintain clarity: legends should be avoided if possible.

Some of these design principles are showcased in figure 2, which is a modern reworking of figure 1:

**Minimalism:** Figure 1 is very *busy*: notice all the gridlines, the wordy axis labels, and the copious vertical axis tickmarks. A more modern, minimalist sensibility is embodied in figure 2,

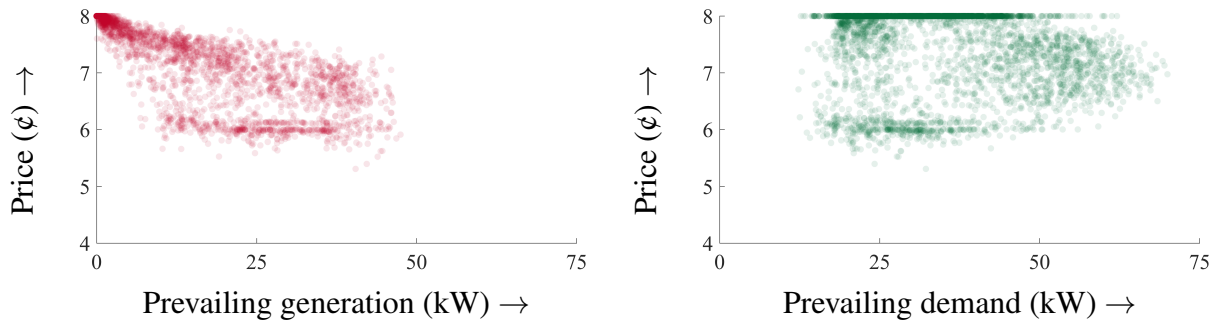


Figure 4: The relationship between average energy prices and the prevailing network generation and demand

which leaves plenty of whitespace and room for the data to breathe.

**Explicit representation of trade balance:** In figure 1, the trade balance is represented as a shaded area between the import and export traces. However, as both the top and bottom of this area is varying over time, it is hard to properly perceive its fluctuations. To overcome this, figure 2 explicitly portrays the trade balance in the upper pane.

**Deliberately matched axes:** The axes in both panes of figure 2 are carefully matched. Crucially, the horizontal axis is identical in both cases, so the year labels on the bottom axis are accurate across both panes. This equivalence allows the crossover year to be explicitly tagged on the upper pane: this annotation is matched with a corresponding bare tickmark on the bottom axis (trying to mark the year ‘1754’ on the bottom axis would create a clash with the ‘1760’ label) The vertical axes are also matched, each having the same height and both spanning 200 units. Note that this leaves some whitespace in the upper panel: software might have automatically select a range of [-40 → +100] here. However, axes that are symmetrical about zero are important here, to give an at-a-glance summary of how much the balance favours England. Zooming in too close on data often fails to tell the big-picture story!

**Direct labelling:** The data displayed in figure 2 is a bit confusing: it would be easy to lose track of the ‘polarity’ of imports and exports. The traces in the bottom pane are directly annotated with colour-matched labels, and ‘Imports’ and ‘Exports’ are clarified using a miniature pictogram of a ship sailing between the relevant flags. Humans are visual creatures: don’t be shy to use imagery to clarify a point whenever you can! Likewise, flags and corresponding labels are used to contextualise the vertical axis in the upper pane.

**Values of interest:** On each trace, the opening and closing values are directly annotated in the appropriate colour. This emphasises how much the trade figures have varied over the eighty year period depicted.

## **Working with more technical data**

When visualising complicated technical data, it is often useful to present multiple complementary panes. An example of this is shown in figures 3 and 4. The data here concerns the simulated operation of a local electricity marketplace over the course of three days. Figure 3 presents the variation in electricity consumption, local electricity generation, and traded energy prices over this period:

**Time axis:** The bottom axis of figure 3 is formatted like a 24 hour clock, and includes orientating tickmarks at midnight each night. In addition to this, the onset and completion times of the daily generation cycles are marked: this is solar generation, so this corresponds to daily sunlight levels. These markings are projected upwards to the other panes using a faint gridline, so that the time-matched relationships between generation, demand and resulting prices can be explored.

**Price spreads:** The top pane of figure 3 shows the average spot price of energy transactions: it can be seen that when the solar generation ramps up, this surplus causes energy prices to decline. As the average strike price isn't the only quantity of interest here, the *spread* of prices arising in each period is also depicted in a paler shade, as a backdrop. These price spreads recall the famous *candlestick charts* used to show financial data, which depict the opening and closing prices of a security over a time period. Famously, green candles depict rising prices, with red reserved for bearish periods!

**Temporal trends:** The data in figure 3 has a granularity of just sixty seconds, and so the noisy minute-to-minute variance is quite evident. In many cases like this, it is wise to overlay a *smoothing line*, to cleanly portray the temporal trend. However, the purpose of figure 3 is actually to show how short-term fluctuations in solar power affect the local electricity price. So, in this specific case de-emphasising these noisy fluctuations would risk detracting from the graph's intended message. Generally, though, overlaid trendlines are a useful tool for showing time series data.

To complement the time series perspective, figure 4 explicitly explores the relationship between price and prevailing demand and generations:

**Scatterplots for comparison:** Time series data need not always be plotted versus time. If the relationship between two particular variables is of interest, this can be shown directly and explicitly using a scatterplot. For instance, the left pane of figure 4 shows quite clearly that increasing power generation exerts downward pressure on energy prices. Conversely, the rightmost pane shows that price and demand levels appear largely independent.

**Opacity:** It is always a good idea to use *opacity* when plotting scatterplots of large datasets. This allows the natural clusters in the data to reveal themselves, as areas of greater density. For

instance, the dark green band in the right pane of figure 4 indicates that 8¢ is the most typical energy price.

**Colour coding:** Note that the colour coding remains consistent between figures 3 and 4: red representing generation and green representing demand. This makes it easier for the reader to make comparison back-and-forth between the figures. Consistency always helps graphics to work together and to complement each other, the whole being greater than the parts.

## Parting comments

The time series plot is an ubiquitous and useful data graphic. As with other data graphics, it works best when it has sensible axes, minimum clutter, and consistent styling with other complementary graphics. While many software packages allow you to create a time series graph, not all them make it easy to embrace these design principles. Popular software choices amongst serious data visualisation practitioners would include the *R* language combined with *ggplot2* library, or *Python* with *matplotlib*. Fortunately, these are free and open source solutions! Another useful tool is the lightweight raster editor *Paint.net*, which can be used for manual fine-tuning of graphics (such as adding the flag annotations in figure 2).

## About the author

**Paul Cuffe** (paul.cuffe@ucd.ie) is an Assistant Professor in the School of Electrical & Electronic Engineering at University College Dublin. With his regular *Figured Out* column in IEEE Potentials, he seeks to equip early career engineers with the graphical skills they need to understand complex data and to make an impact in their organisations. Paul would like to acknowledge Hugo Gil for his help with earlier versions of some of the graphics presented in this article. Ship icon by *Symbolon*.

## References and Notes

1. W. Playfair, *Lineal Arithmetic; applied to shew the progress of the commerce and revenue of England during the present century, which is represented ... by thirty-three copperplate charts, etc.* 1798.
2. C. Giaimo, “The Scottish scoundrel who changed how we see data,” *Atlas Obscura*, 2016.
3. H. Wainer and I. Spence, “Visual revelations,” *CHANCE*, vol. 10, no. 1, pp. 35–37, 1997.
4. W. Playfair, H. Wainer, and I. Spence, *Playfair’s Commercial and Political Atlas and Statistical Breviary*. Cambridge University Press, 2005.