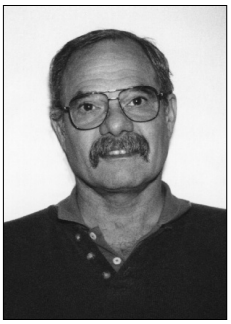
VISUAL REVELATIONS



Howard Wainer, Column Editor

Nobody's perfect

Michael Friendly, York University and Howard Wainer, National Board of Medical Examiners

One of the great advantages of the graphical depiction of information is the vast flexibility of graphical formats. The same format can be used for many different purposes. Nevertheless, line charts, bar charts and pie charts seem to be most frequently used for presentation of quantitative information in the popular media, whereas the scatterplot's principal home is on the pages of more technical outlets. But although its fame may be a bit circumscribed, the scatterplot's value within science is fully appreciated. Friendly & Denis (1984) described it as the most versatile, polymorphic, and generally useful invention in the entire history of statistical graphics.

In view of the division between typical audiences of these two classes of displays (scatterplots vs. everything else) it should not be surprising that there were two different inventors. The iconoclastic Scot William Playfair (1759-1823), is generally credited with the invention of the pie chart, bar chart and line chart (Funkhouser & Walker, 1935; Funkhouser, 1937), whereas the British astronomer John Frederick William Herschel (1792-1871) is the most likely candidate as the inventor of the scatterplot (Friendly & Denis, 2004). Herschel's use of the scatter plot occurs in his "investigation of the orbits of revolving double stars," which he read to the Royal Astronomical Society on January 13, 1832 and published a year later. The term "scatterplot" appeared somewhat later; Moore (1911) credits it to Karl Pearson.

But why didn't Playfair invent the scatter plot? He did such a marvelous job inventing so many graphical tools, how could he have missed this one? Could it be because the data of greatest interest to him would not have yielded deeper insights as a scatterplot? Playfair's initial plots (in his 1786 Atlas) were almost all line graphs. He was interested in showing trends in the commerce between England and its various trading partners. Hence line graphs of imports and exports over time were the logical design of choice. He would also shade the area between the two resulting curves to indicate the balance of trade, and label this space as balance in favor or against England as the case might be. Such time series data are not a fertile field from which might grow a scatterplot.

Playfair did not have Scotland's trade data for more than a single year (1780-1781). Thus the time-line plot was not suitable. So instead, he depicted Scotland's trade with its eighteen partners as a bar chart (Figure 1). He complained about the insufficiency of the data, but note how well he used the form: he paired imports and exports together for each of Scotland's trading partners for easy comparison (he could have made two separate charts), he placed the bars horizontally to make reading of the labels easier, and he ordered the countries by the total value of each country's trade with Scotland. Let us emphasize the insight that this last aspect represents. It is easy to imagine someone of lesser insight succumbing to "America first" and ordering the bars alphabetically. The resulting figure would be much the worse (see Figure 2).

Playfair invented so many graphical forms, and because his taste was almost always impeccable, it is shocking to find one that has serious flaws. In his Letter on our agricultural distresses, their causes and remedies; accompanied with tables and copperplate charts shewing and comparing the prices of wheat, bread and labour, from 1565 to 1821 , he produced an apparently breathtaking figure (Figure 3) showing three parallel time-series: the price of a quarter of wheat, the weekly wages of a 'good mechanic', and who was the reigning monarch during each of the time periods shown.

His use of the line graph for such time series data was natural given how successful he had been in the past with similar data. But the inferences he wished to draw from this data set were subtly different than the ones he had made before. Previously the questions posed were "How have England's exports to X changed over the past 200 years?" Or "what has been the character of England's debt over the past 200 years?" Sometimes, when a second parallel data set was also included on the plot a new, but parallel, question pertaining to it was posed "How have England's imports from X changed over the past 200 years?" And then natural comparisons between the two data series were made – When were exports greater than imports?

Column Editor: Howard Wainer, Distinguished Research Scientist, National Board of Medical Examiners, 3750 Market Street, Philadelphia, Pennsylvania 19104; *hwainer@nbme.org.*

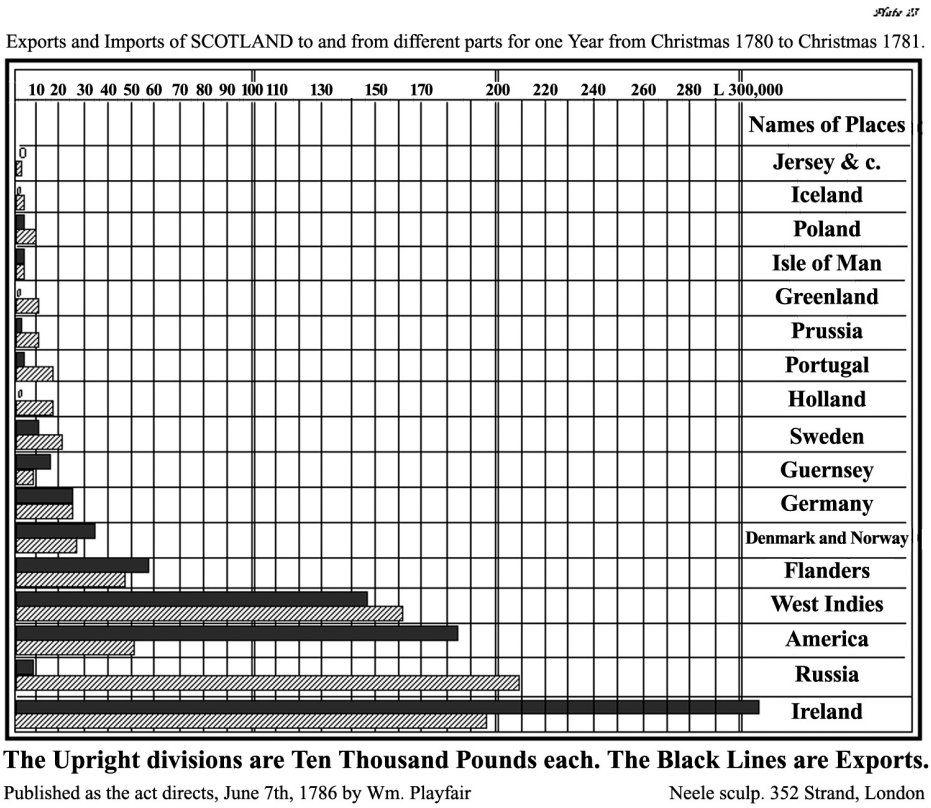


Figure 1. The imports (cross-hatched lines) and exports (solid lines) to and from Scotland in 1781 for 18 countries, ordered by the total volume of trade.

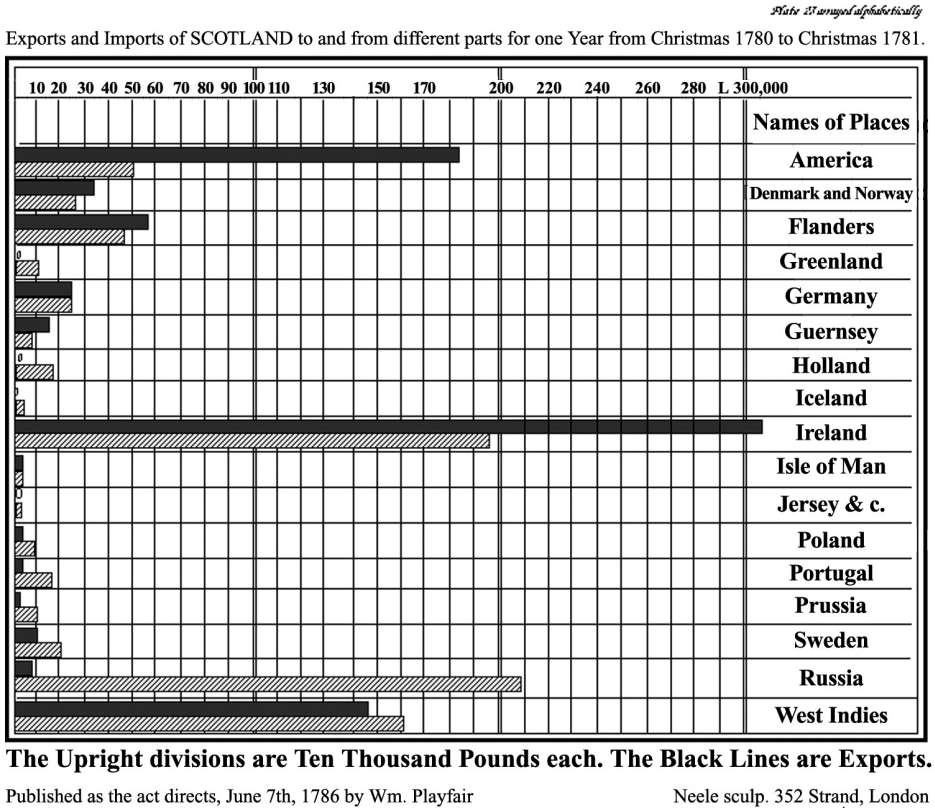


Figure 2. The imports (cross-hatched lines) and exports (solid lines) to and from Scotland in 1781 for 18 countries, ordered alphabetically.

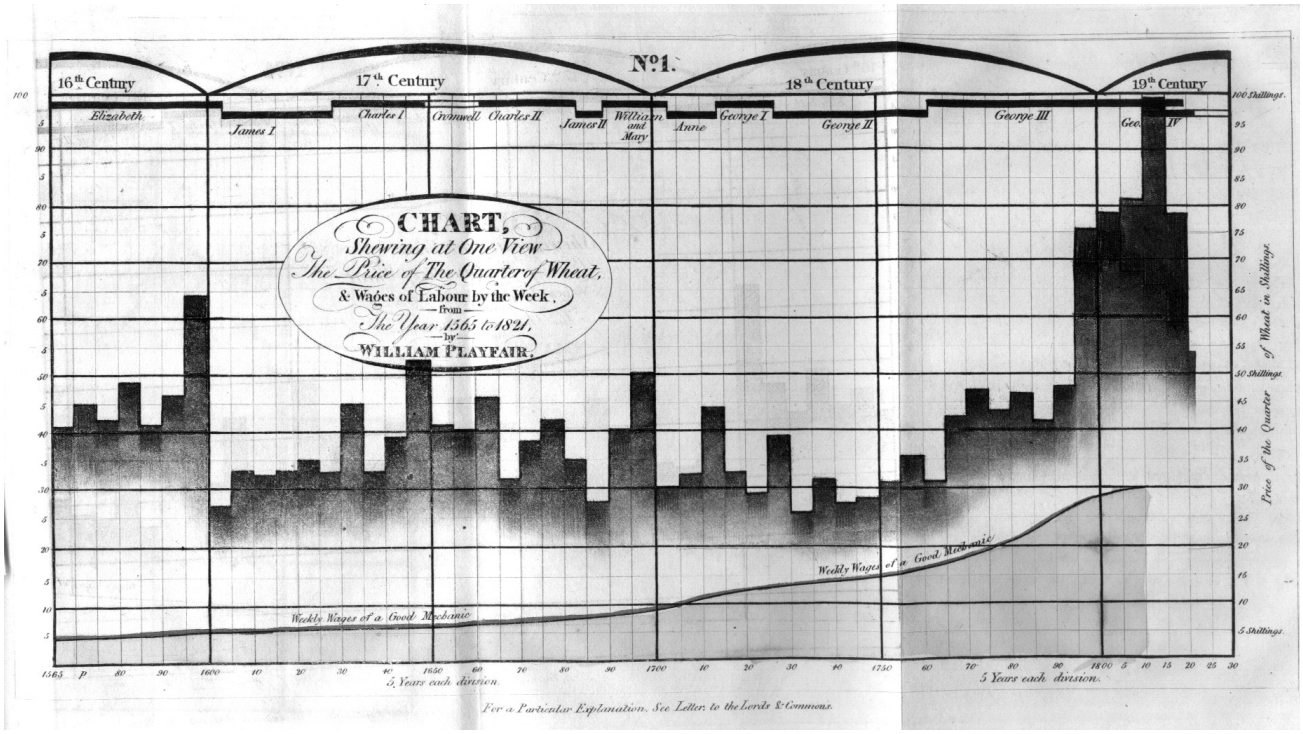


Figure 3. **A time series display showing three parallel time-series: prices of a quarter of wheat (the histogram bars), wages of a good mechanic (the line beneath it) and the reigns of English monarchs from Elizabeth I to George IV (1565 through 1820). Our gratitude for this copy of Playfair's figure to Stephen Ferguson and the Department of Rare Books and Special Collections. Princeton University Library.**

All of these questions could be well answered by the format Playfair chose. But the data in Figure 3 were meant to illustrate a deeper question. Playfair wrote (page 29-30)

"You have before you, my Lords and Gentlemen, a chart of the prices of wheat for 250 years, made from official returns; on the same plate I have traced a line, representing, as nearly as I can, the wages of good mechanics, such as smiths, masons, and carpenters, in order to compare the proportion between them and the price of wheat at every different period …. the main fact deserving of consideration is, that never at any former period was wheat so cheap, in proportion to mechanical labour as it is at the present time…" (emphasis ours).

Is this conclusion true? It is not easy to see in Playfair's figure as he produced it. Apparently Playfair was not fully acquainted with the benefits of combining variables into new variables to examine specific issues. In this instance if we make a line plot in

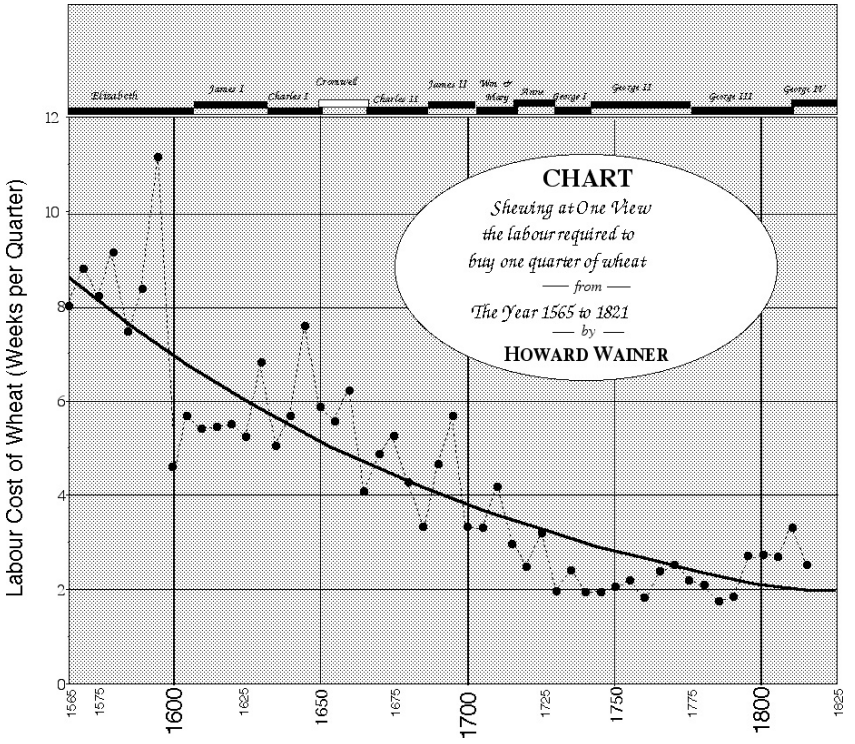


Figure 4. **The times series from Figure 3 recalculated to show the number of weeks of work required to buy one quarter of wheat. The individual data points are connected by a dotted line, a fitted quadratic is drawn through them (solid line).**

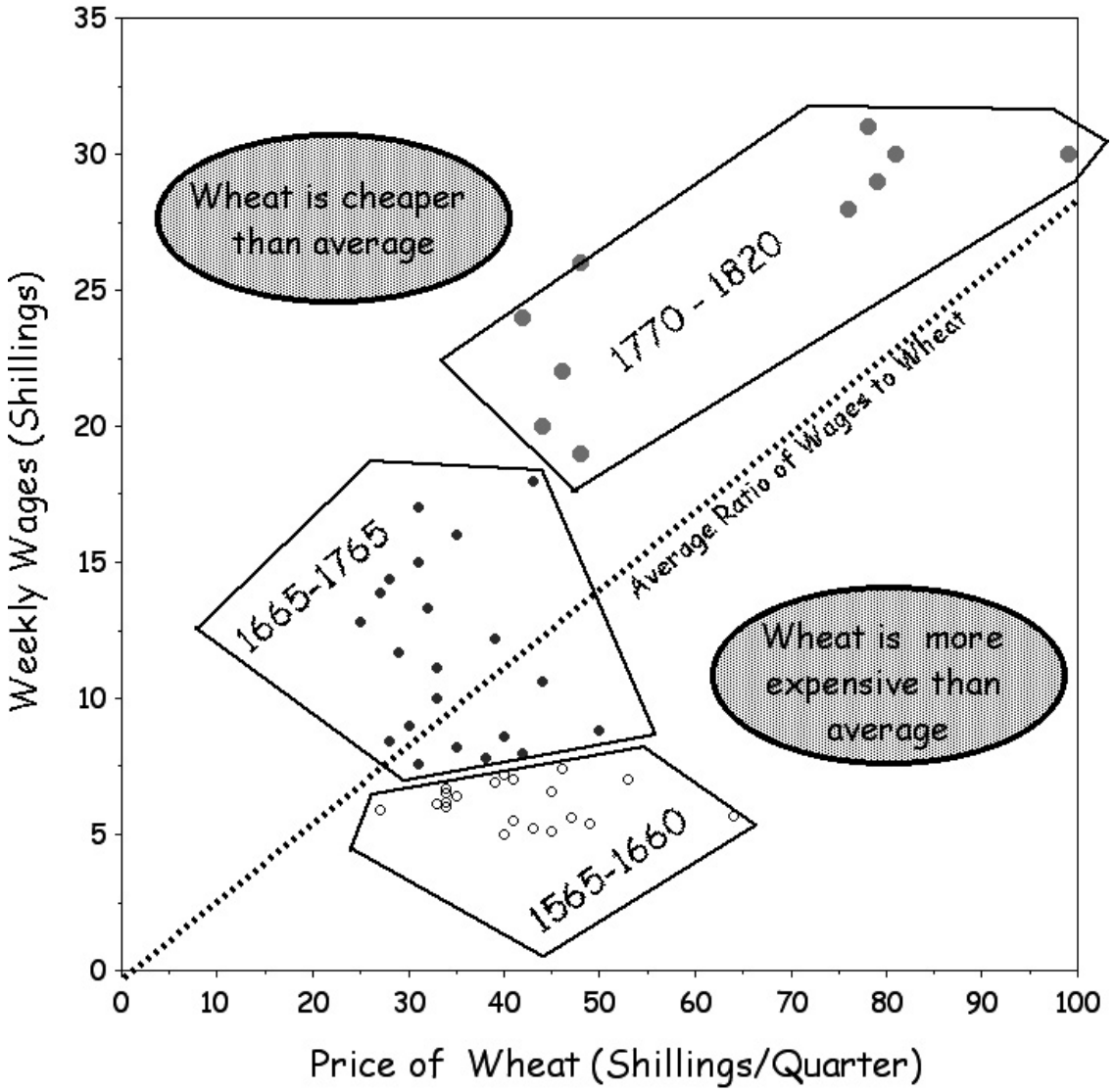


Figure 5. **A scatterplot of Wheat price (on the horizontal axis) vs. Wages (on the vertical axis). The diagonal line represents the average ratio between the two over the 255 years depicted. Data points for three different eras are indicated.**

which the variable being displayed is the ratio "Labor cost of wheat (weeks/quarter)" the truth of Playfair's conclusion is evident.

Another way to look at the relationship between the cost of wheat and the amount of wages directly might be a scatter plot. But to construct a scatter plot that provides such an insight (see Figure 5) takes a lot of work, and some experience in peering at them. Playfair might have considered a scatter plot, and rejected it as worthless – or at least as worthless for the sorts of time series data that he was primarily concerned with. Of course, had Playfair been interested in the relationship between parent's and children's heights he might have scooped Galton. But he wasn't and he didn't. Nobody's perfect.

References

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, Experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79, 531-554.

Friendly, M. & Denis, D. (2004). Who invented the scatterplot? Manuscript submitted for publication..

Funkhouser, H. G. & Walker, H. M. (1935). Playfair and his charts, *Economic History*, 3, 103-109.

Funkhouser, H. G. (1937). Historical development of the graphic representation of statistical data. *Osiris*, 3, 269-404.

Herschel, J. F. W. (1833). On the investigation of the orbits of revolving double stars. *Memoirs of the Royal Astronomical Society,* 5, 171-222.

Moore, H. L. (1911). *Laws of Wages: An Essay in Statistical Economics*, New York: The Macmillan Co. Reprinted in 1967 by Augustus M. Kelley: New York.