

Visualizing Cereal World

Edward J. Wegman and Alan Dorfman
George Mason University and the Bureau of Labor Statistics

The Consumer Price Index is arguably one of the most important statistics produced in the US. Price indices determine a wide range of economic benefits and generally indicate the economic health of the nation. Traditionally these indices are based on survey data from household surveys (of consumer purchases) and retail establishment surveys (to determine selling prices). The linkage between amount spent by households on items bought and their prices is therefore indirect. Scanner data based on UPC codes directly links price, quantity, and total expenditures and therefore offers the possibility of more accurate prices indices. In addition, because expenditure data is collected directly, use of scanner data offers the possibility of exploiting alternate price index formulae. Because price indices are typically averages, the possibility exists that unforeseen outliers and other anomalous structures in the data can bias or skew this important index. The Bureau of Labor Statistics has launched an experimental study of the possibility of revising the CPI using scanner data. The experiment involves using scanner data on breakfast cereals. This paper will briefly review price indices and then discuss some findings discovered through a data visualization process.