Data mining uncovers 19th century Britain's fat habit

April 3 2014, by Uta Hinrichs, Beatrice Alex, Jim Clifford

Full speed ahead in search of globalisation.

A collaboration between historians, text mining, and information visualisation researchers has thrown up new insight into the hunger for sugar, coffee and rubber in the 19th century, as well as how fat became a worldwide commodity.

Text mining, the process of searching large volumes of data to dig out nuggets of information, in combination with information visualisation,

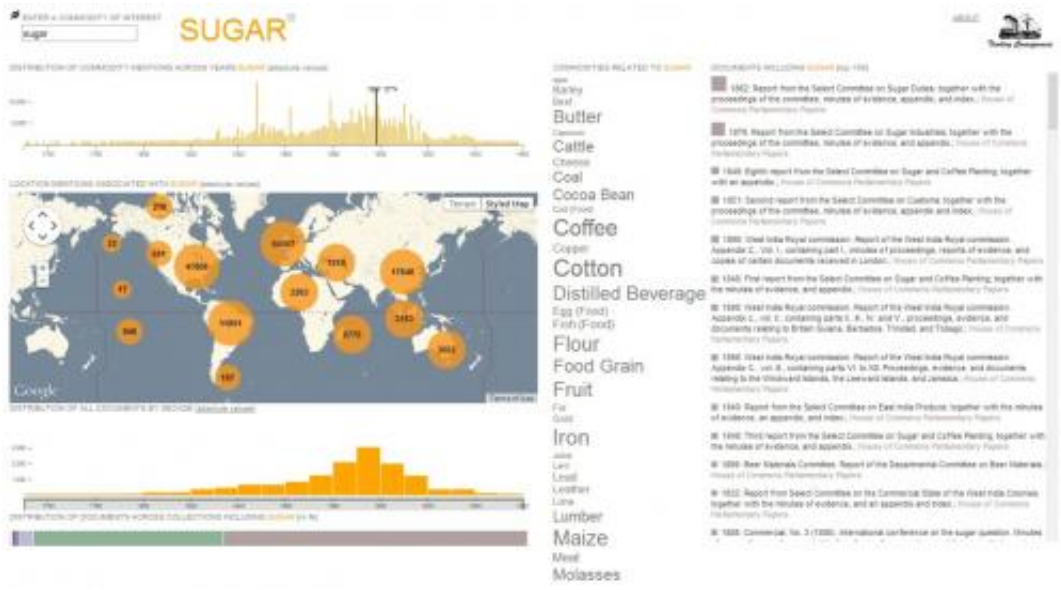
which can help to reveal and explore trends in large data sets, has opened up more than 200,000 historic documents to investigate the trading of goods since the 19th Century.

Historians found that not only did the British import cotton, tea, and rubber from the United States, China and Brazil, but they also relied on many hundreds more obscure commodities such as gutta-percha, a natural latex, cinchona, which is used to produce quinine, and gum arabic for the printing trade. British industrialists sourced these commodities from Peru, West Africa, Sri Lanka and the Malaya Peninsula.

The global supply chains that developed as a result enabled industrial development in Britain. But they also had significant environmental consequences, such as deforestation and resource depletion, all around the world.

This is particularly the case in the global supply of oils and fats that developed over the course of the 19th century.

At the end of the 18th century, the British imported small amounts of tallow, fat rendered from beef or mutton, from Russia and Ireland. But over the next hundred years, the global fats supply chain grew to include palm oil from West Africa, cottonseed from Egypt and the United States, whale blubber from the South Pacific along with new sources of tallow from Australia and Argentina. This was all to supply the burgeoning soap, candle, margarine and explosives industries.



Most frequently mentioned commodities and locations, 1750-1960.

The environmental consequences of this hunger for fat varied from the dramatic over-harvesting of whales and significant deforestation associated with coconut plantations, through to more benign results such as the increase in palm oil harvested in the Niger Delta region.

Light work of big data

The Trading Consequences project, that was part of the second round of the Digging into Data challenge, has produced a database that allows users to search for new detail of this kind using text mining techniques.

That information can then be displayed using information visualisations to give a rich picture of how Britain sourced common commodities such as cotton and more obscure ones such as gum arabic at a large scale and where in the world trade went on.

Traditionally, historians have to read through vast numbers of documents

to research a topic. But the amount of text one person can take in is limited to a couple of hundred documents a year, at most. That makes them ill-equipped to make the most of the millions of pages of historic documents available in archives worldwide, even though many of them are now easily accessible online.

Vast amounts of digitised historical records are available to inform us about the extent to which Britain relied on overseas commodities in the 19th century but despite this huge resource, and maybe even because of its size, the story is impossible to accurately track manually. It spreads across documents discussing locations all over the globe and spans decades.

But by identifying hundreds of terms associated with commodity trading, such as "sugar" "coffee" or "cattle" and geographic locations such as London or Canada, which have been computationally extracted from the documents, the mountain of information becomes infinitely easier to search.

Text mining also makes it possible to identify links between commodities, locations and dates to uncover how commodities were discussed in space and time; information visualisation can make this information accessible and explorable. Both techniques in combination can build a picture of how trade developed over time, based on dates or places mentioned in the records of a trading company operating several hundred years ago.

The visualisation provides an overview of the most frequently mentioned commodities and locations between 1800 and 1910. Searching for iconic commodities such as "sugar" refines these overviews and also shows related commodities that were mentioned in the same context. So we can see how British planters experimented with a range of commodities in Sri Lanka during the course of the 19th century, starting with coffee,

then cinchona before settling on tea and rubber at the century's end. Zooming in on certain time frames or geographic locations provides more details.

The location cloud visualisation provides an overview of the most frequent locations that were mentioned in relation to a selected commodity, such as coffee, which helps build an picture of how the coffee trade spread. Location labels are scaled according to their relative importance in that decade, so you can see who ruled the roost for a particular commodity at a particular time.

These new techniques can help produce a better understanding of how Britain came to occupy the place it did in global trade right up until the early 20th century. That, in turn, can help us understand the continued growth in global trade and the environmental consequences of that trade that are still being felt today.

This story is published courtesy of The Conversation *(under Creative Commons-Attribution/No derivatives).*

Provided by The Conversation