WORCESTER POLYTECHNIC INSTITUTE

FEED GRAIN DATA VISUALIZATION PROCESS BOOK

December 13, 2016

Xitong Zhang (XitongZhang1994)
Bolun Lin (Aghost527)
Meng Li (li545)

Contents

1 Overview **4**

2 Motivations **4**

3 Related Work **5**

4 Questions **5**

5 Data **5**

6 Exploratory Data Analysis **6**

7 Design Evolution **7**
7.1 Overview Charts 7
7.2 The World Map 7
7.3 Heatmaps 8
7.4 The US Map 8
7.5 History Layout 9
7.6 Assembly For The Final Deliverable 9

8 Implementation **9**
8.1 Overview Charts 9
8.2 Supply and Disappearance 11
8.3 History 12
8.4 Exports and Imports (Heatmaps) 14
8.5 The World Map 15
8.6 The Domestic Price 18

9 Evaluation **19**
9.1 Discussion Of Results 19
9.1.1 Corn 20
9.1.2 Sorghum 20
9.1.3 Barley 20
9.1.4 Oats 21
9.2 Future Endeavors 21

10 References **21**

11 Appendices **22**

1 OVERVIEW

The project aims to create a visualization system for the US feed grains data. The system enables users to visualize the US historical feed grain data and gain insights into supply and demand of feed grains in domestic market and international trade of the US with import and export partners. The system even can help users briefly review history after the American Civil War. Feed grains are grains grown for feeding livestock, such as corn, sorghum, and barley. The data consists of four major parts, supplies, disappearance, i.e., consumption, domestic prices, and imports/exports data. The four major goals are,

- Demonstrate inbound and outbound flows of feed grains.
- Discover the US exports/imports trends.
- Visualize evolution of both consumption patterns on the demand side and sources of supplies.
- Finally, since our data covers a long time period, from 1866, right after the American civil war, to 2016, it's also quite interesting to us that whether the data matches history.

2 MOTIVATIONS

Feed grains, as a component of food chain, their sufficient and reliable supplies help stabilize food supplies and ensure food security, and even contribute to national security. On the other hand, as a major source of animal feeds, feed grains production has significant impact on feed industry which directly determines production of animal protein and therefore is closely related to public health and welfare. Additionally, feed grain is also an important agricultural commodity trading in global market which plays an essential role in determining ranking and influence of a nation across the world. Creating a visualization system for feed grains enables us not only to perceive supply and demand patterns, but also to acquire in-depth understanding on structure of feed grain consumption and detect trends of domestic prices of feed grains. Besides, we are motivated as well by challenges of designing visualizations raised by complexity of the data.

3 Related Work

Picking this topic actually is no more than an accident. We in the very beginning, were interested in creating visualizations for food grain data, since food shortage has been a popular topic across the globe. However, somehow we mixed definition of the feed grain with that of the food grain. Luckily, we are quickly fascinated by the feed grain data, because there are tremendous stories inside. Plus, we are inspired a lot by deep reasons and stories behind data patterns. The data appears more and more fascinating to us.

4 Questions

The questions we are curious at and would like our visualization system to answer are listed below,

- How does feed grain flow in (produced) and flow out (consumed)?
- How does consumption patterns of feed grain evolve?
- How does sources of feed grains supplies transform structurally?
- Does the trends of the US imports/exports change over time? If yes, how?
- Can the data narrate the US or even the world history?

5 Data

The data is collected from Data.gov, which is managed and hosted by the US General Services Administration, Office of Citizen Services and Innovative Technologies. The link to the data is here.
The data incorporates four feed grains, corn, sorghum, barley, oats and foreign coarse grains. Important tables and attributes are listed below,

- Supply: beginning stock, production, imports
- Disappearance: food, alcohol use and industry use, seed use, feed and residual use, exports and ending stock
- Domestic and International Prices: average prices received by farmers and cash prices at principal markets

- Exports and Imports

Our data is in excel which is converted into csv for d3 to process. The data processing is implemented with js functions and python.

Data cleaning is necessary, since the dataset contains missing data and in the meantime, data tables have different time frames. For an instance, supply and demand data covers a long time span from 1975 to 2016, while import and export data is in a range from 1988 to 2016. Also, the data needs discretization for some graphics, since it is not a good idea to use color to represent continuous variables. At last, our data is in excel which has to be converted into csv for d3 to process. The data processing will be implemented with js functions.

The whole dataset of the grain is so large that it has hundreds thousand pieces of data. If we load the whole dataset before loading the visualization, it would take a lot of time and make the users annoyed. Moreover, keeping too big a dataset in memory is a burden of computer and makes the query very slow, so we have to divide the whole dataset into pieces and for each page only load the part of the data it plots.

Thus, we sometimes need to load more than one CSV or JSON files in one page, especially for a map, whose implementation details are usually saved in a JSON file. Then, we need to avoid the situation that the website runs without having all the files prepared, so we use the 'queue'function in D3 to make sure that the visualization is present after all the files it needs are ready.

6 EXPLORATORY DATA ANALYSIS

We initially visualized the data simply with line charts, which show clear trends in production, planted acreage, yield per acre, price, feed use, and food use. By visualizing the first four statistics, we realized that those trends somehow were coincident with the US and world history. By examining trends of other statistics altogether, we were surprised that feed grains consumption patterns and sources of supplies transformed evidently over time. We therefore, decided to conduct further investigation by creating more visualizations and building interactions with them. Also we developed great interests in the US exports and imports data in the very beginning.

7 DESIGN EVOLUTION

We initially planned to visualize the statistics with slope charts (Figure 25), because we focus on comparisons across grains and trend detection. However, after creating a couple of alternative visualization types, including streamgraph (Figure 23) and line charts (Figure 24), it's quite obvious that streamgraph and line charts are more effective than slope charts. For example, the streamgraph (Figure 26) showing harvested acreages of all four grains not only demonstrates trends for individual grains and makes comparisons across grains possible, but also presents the trends of total harvested acreages, since the height of the outer curve at a given time point is the sum of harvested acreages of four grains.

In addition, as we are lectured in classes that, slope charts are able to compare only two variables that are next to each other directly and in general do not perform well when there are a fair amount of variables as what we have, we abandoned the original design.

7.1 Overview Charts

To make an overview more informative, we modified the original design (Figure 22) by populating it with pie charts that present percents of components in total supplies, total disappearance, and domestic use, as well as small multiple graphs for all important statistics (Figure 27). Also, to enable users visualize data in a specific year, we built a time control widget.

7.2 The World Map

For annual imports and exports data, we initially planned to visualize it in a map with arrows pointing from the US to countries that are trading with the US and thickness of each arrow indicating trading volume as Figure 31. However, we realized that it's almost impossible to represent trading volume by arrow thickness, because too many arrows overshadowed each other, causing thickness of arrows indiscernible. We modified the original design by representing trading volumes with colors filled in countries. The Figure 29 is half-finished, since the world map has not been padded with colors. However, after finishing it, we found the map cluttered with different colors and arrows. We therefore decided to build a control widget for users to choose whether hide or display arrows. In addition, we added a mouseover

to the world map so that users can view trading volume between the US and a selected country in a selected year. To make the layout concise, a filter was built for users to choose either display exports or imports. The final version is given by Figure 37. Since the data is highly concentrated, we divided the range unevenly into five intervals for better renderings. The original design of visualizing export and import volumes with slope charts (Figure 30) is discarded.

7.3 Heatmaps

Besides, we noticed that the world map does not provide users any information on the US aggregate exports and imports over all trading partners of the US. Hence, we designed three heatmaps for monthly aggregate exports (Figure 32), imports (Figure 33), and net exports (Figure 34) data. To map continuous data with five colors, monthly volumes are discretized into five intervals.

7.4 The US Map

The US feed grain prices are visualized with the US map and sparklines. In the very first design version, we plan to pad the US map with colors that indicate prices of grains in the market located in a state (Figure 28). However, after closely checking the data, we had to abandon the design, because not every state has a trading market for grains. Some of states even have more than one market. In this case, it 's not feasible to color a state with two different hues. Then we switched to another design which presents domestic markets and prices separately in the US map and line charts respectively (Figure 35). When a user selects a particular year on the time selection widget and hover a mouse over a specific market, monthly prices will appear in the line chart. However, the design suffers a serious drawback: when the year is fixed, prices in other years will be hidden in the line chart, which contradicts the purpose of presenting price trend over time. It does not make sense to display only one segment of the entire trend, because users are not able to perceive anything without inspecting the segment over the entire trend. Moreover, it could be misleading, because trivial patterns may be amplified. Finally, we came up with a design which replaces the line charts with sparklines showing price data covering the entire time span and removes time selection widget. For users to make comparisons across grains effortlessly, sparklines for prices of four grains will be vertical-aligned (Figure 38).

7.5 History Layout

Since we realized that our data reflected history, a new design for linking the data and history is appended. The key of the design is the time line of historical moments, in which individual moments are highlighted by colors and users can check them by clicking colored segments. Figure 36 only shows partial layout, because the whole layout is too large to be captured in a screenshot.

7.6 Assembly For The Final Deliverable

Finally, the visual system is assembled by organizing all charts into five sections: overview, supply and demand, imports and exports, price, and history. For users to easily navigate to a particular section, a menu bar is built.

8 IMPLEMENTATION

8.1 Overview Charts

At first, we intend to know where grains are from and how they disappear. There are 3 parts of supply and 6 parts of disappearance. We want to show the connections between different variables. And then we design this chart.The image in the Ending Stocks rectangle indicates which grain's data shown on the website.

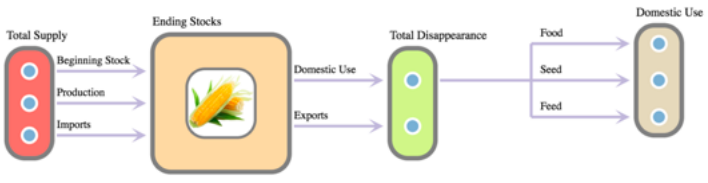

Figure 1: Inbound and Outbound Of Feed Grain Flowchart

We choose bar chart to show the raw data of each variable, we can get the trend directly.

Figure 2: Bar Chart For Beginning Stock

Because there are multiple variables we want to show, so we design the matrix plot in the end. And the 12th plot is the unit of all charts. At first, we want to use one scale for all bar charts to show the amount relation between different variables. But length of some bars will be really short. For example, as the image shown below, data of imports is 1.5 and data of production is 5841. If we use one certain scale for all bar charts, the trend of some charts will be hard to be recognized.

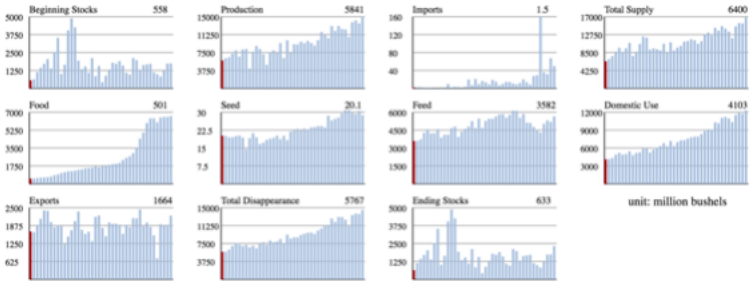

Figure 3: Small Multiples

And then we have a problem. We have about 50 years' data, it's difficult to select one certain year for users by using mouse. At the same time, what users want is to get all data of one certain year. Interaction using mouseover is not enough.

To solve the problem, we design a slider to select a certain year. When user choose one year, chosen data will be highlighted.

At first the length of the slider is short when showing prototype, it's difficult to choose one year too, so we make it longer.

Figure 4: Slider For Time

And then we want to show more stories of data because there are 3 parts of supply, 2 parts of disappearance and 2 parts of domestic use. We consider that it'll be better if we can show the proportion situation. So we design 3 pie charts then.

When users select one year by slider, the pie charts will change with matrix plot we design above.

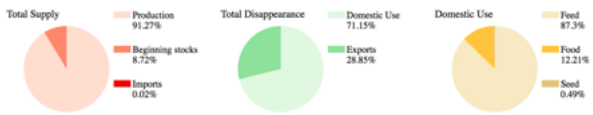

Figure 5: Pie Charts

We don't design many interactions in this part, because we consider what we design should be effective, we want to show all information in data. Too much interactions need more manipulations from users. We believe it's inefficient.

The completed overview is given in Figure 32.

8.2 Supply and Disappearance

And then, we think it's not enough to show raw data of supply and disappearance. Because matrix bar plots can't show amount relations between different variables because the scale of every bar chart is different. We think stream chart or stacked bar chart is a good choice to gather all data together into one chart. Because we have already designed matrix bar charts. So we choose stream chart to implement.

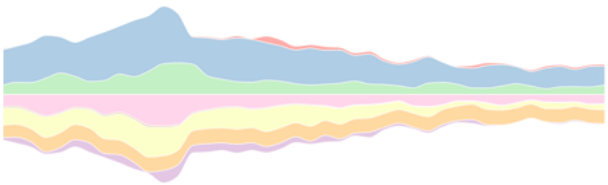

Figure 6: Streamgraph For Supply and Demand

And then we add legends and a slider for users choosing years. The data of legend will be

updated when users select a different year using the slider.

8.3 History

We think harvested acre is the most important variable to show the history of agriculture. The harvested acre will changed rapidly when some import event happened. For example, the harvested acre will decrease when there is a drought. And we have 4 types of grains, we think it's not appropriate to jump into the conclusion with showing only one type of grain. We have data of 4 grains from 1866 to 2016, so stream chart is a good choice for us because the width will be too short if we choose stacked bar charts to implement.

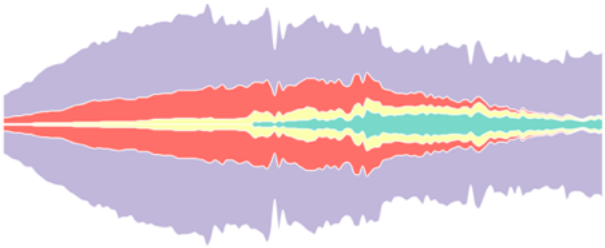

Figure 7: Streamgraph For Harvested Acreages

And then, it's not enough to show raw data. Users should get more information from the data. Users need to know more about the amount relations of 4 grains. So we design a bar chart and a pie chart to show the relations. Bar chart shows the value of every grain. And the pie chart shows the proportion of 4 grain harvest acres.

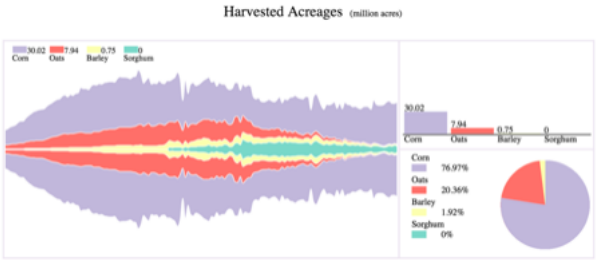

Figure 8: Streamgraph v2 For Harvested Acreages

And then we design a slider to select a certain year. The bar chart and pie chart will be

updated.

And then we think some other variables may reflect the agriculture history. And we select price, production, yield per acre and planted acres. And we design line chart to show trend because we have already designed bar charts and pie charts. We add interaction to show data selected by slider too, a black circle will indicate which data is selected.

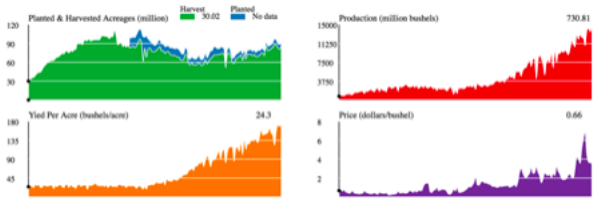

Figure 9: Small Multiples For History

And then to save page space, we design a selector to select which grain data will show.

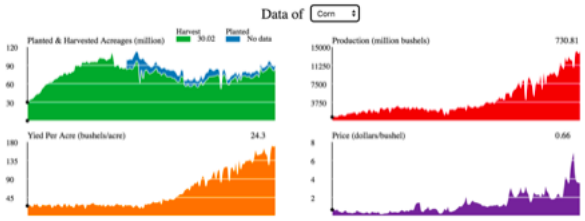

Figure 10: Small Multiples v2 For History

Next, after prototype feedback, we consider that it's not enough to show history event with variables we picked. It's a better choice if we can provide a 'time axis'to show all important events in the U.S. agriculture history. So we design this part finally.

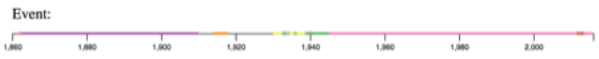

Figure 11: Time Line For Historical Events

It's a 'time axis'which can be zoomed in/out and dragged. Different color represents different event. Users can get what happened by clicking. Chosen part's color will changed by

clicking.

Figure 12: Time Line v2 For Historical Events

8.4 Exports and Imports (Heatmaps)

Imports and exports are two essential variables we considered although we didn't write this part in the proposal. Because when some events happened, the amount of imports and exports would change. We want to show the detailed imports and exports situation. And we have detailed data of imports and exports from 1989 2015.

Then we should choose one chart to show the data. We have already designed bar charts, pie charts, stream charts, line charts. And we believe that heat map is a really good choice because heat map can show some patterns of data directly. And then we designed two heat map for imports and exports.

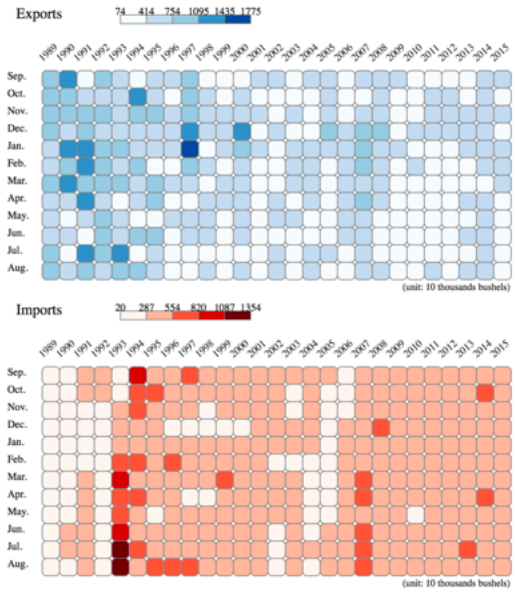

Figure 13: Heatmaps For Exports and Imports

However, heat map can't show the relations between imports and exports. So we designed a third chart by calculation. We can get the result easily that the grain was imported more since 1993.

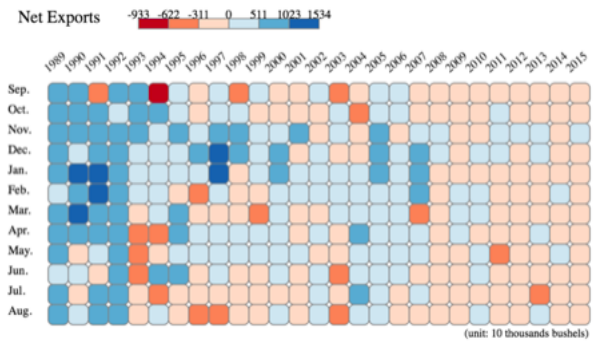

Figure 14: Heatmaps For Net Exports

8.5 The World Map

This is a world map where mark the trading lines between the US and the other country. The color maps to the trading volume for a particular product in one certain year.

Users can query the data with the menu above where can filter the data by time, grain type and export or import. We also provide the trigger to turn the trading lines on or off, in case the lines block users'eyesight when they want to see the detail information of a small area country.

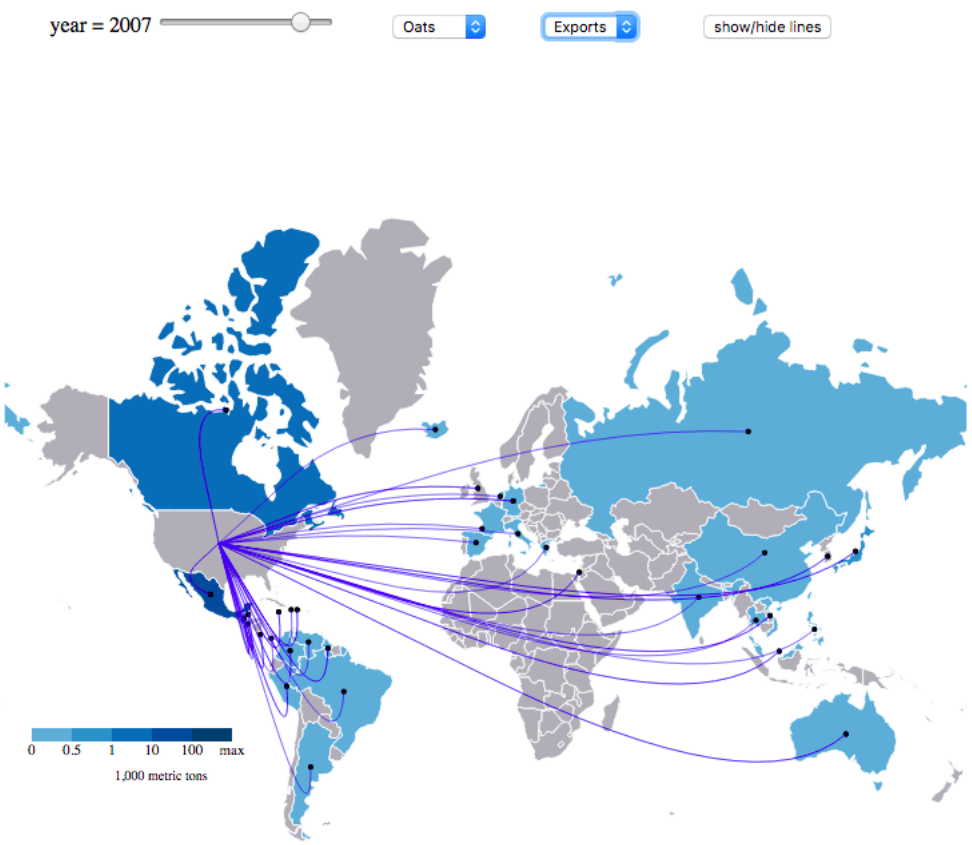

Figure 15: The World Map For Oats Exports

The trading lines are drawn in the form of Bezier curve. We formulate the curves automatically by its midpoint and the ratio. It would make the curves look more elegant than the straight lines and have less chance to mess up with each other. They can be shown or hidden with a click on the 'show/hide lines'bottom and the change would be a smooth transition, without abrupt feeling to users.

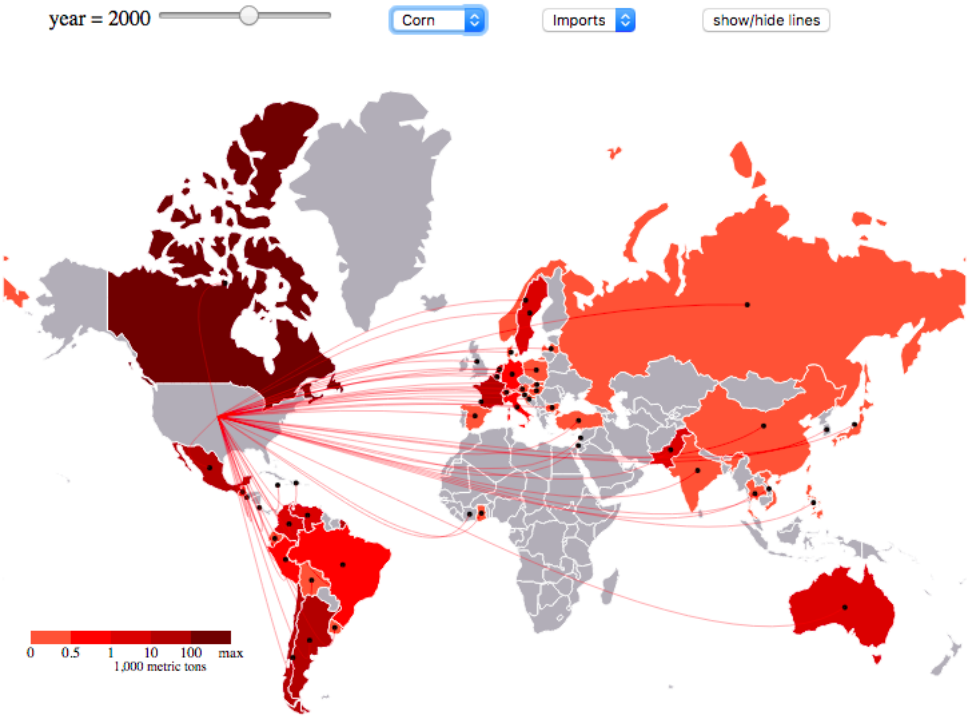

Figure 16: The World Map For Corn Imports

In order to help the users differ the data of import and export, we use different color and re-color all the figure depending on the data involved whenever the user change the setting in the menu above.

Figure 17: The Color Scale Of The World Map

You might have notice that the scales do not equal proportion to each other, because we find that the trading amount with different country varies too much. The amount could be less than 1 thousand metric tons with many countries but more than 300 thousand metric tons with Canada. We have to form the scale in this way. In fact, the color maps to the orders of magnitude of the trading amount.

Figure 18: The World Map For Corn Imports

Since the color only maps to the magnitude and the color isn't good at indicating precise number anyway, we have to add mouse over function in every country to show the precise trading amount. We also have to apply the style 'pointer-events:none'to many figures to prevent them from blocking the interaction. In the picture, the USA imports 2.29 thousand metric tons of Corn form Australia in 2000. The other information is shown in the menu above. The tip would show up when mouse is over a country on the map. In fact, it's a div whose opacity is always 0 except it's triggered by the interaction.

8.6 The Domestic Price

This visualization shows the price trend of a main grain market in the USA for around 40 years. The map on the left marks the primary grain market's geography information in the US. When the mouse hovers on the market, the visualization would show the price trend of several primary grains on the right side.

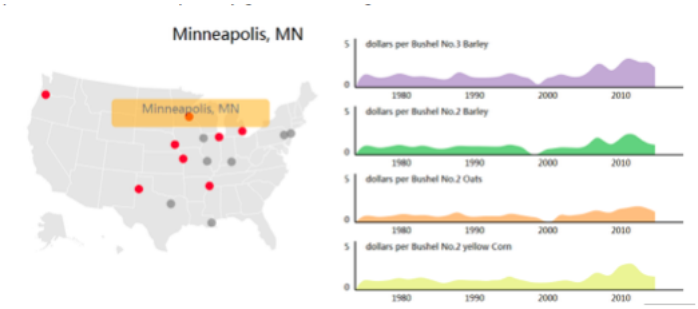

Figure 19: Grain Domestic Price at Minneapolis

Even though a market could sell many kinds of corn or corn product, we only present the unprocessed grain's price trend, so there are some markets without the information we want for the time being. Thus we fill these kinds of markets grey but keep them remaining in case we need to query other product in the future.

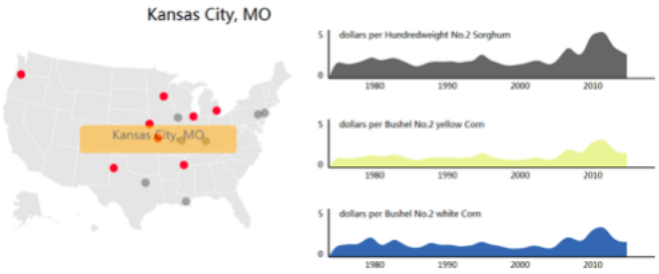

Figure 20: Grain Domestic Price at Kansas City

Different markets keep trading different kinds of commodity, so we may don't know how many trends we should visualize. Thus, we have to implement the flow chart fit the container automatically according to their number. Different grain map to different color, but most of the grains are either green or yellow, so we canâĂŹt color all the flow chart with the grains color in origin.

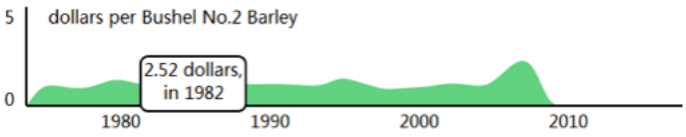

Figure 21: Line Chart For No.2 Barley Price

In order to help the user query the exact price of a certain grain in a year, we implement hover interaction on the flow chart. Whenever the mouse is on the chart, it would show the price in that year with a tip, which would disappear as soon as the mouse out of the chart.

9 Evaluation

9.1 Discussion Of Results

The visualization of the data presents plenty of interesting facts. Primarily, it shows how the US agriculture develops over time as technologies advance and history develops. Secondly, evolution of consumption patterns of grains even reflects changes in diet structure. For an instance, the vast major of barley was consumed as feed grains before the 20th century. However, in recent years, most of barley has been consumed as food grains, indicating that diet structure of the US people has been changed. The structural transformation in sources of supplies illustrates changes in agricultural policies and adjustment of agriculture industrial

structure. For example, we observed that the US oats supply relies on domestic production in the 19th century. However, in the 20th century, the major source of supply has shifted to imports. One reason is that the total consumption declined. Another important reason is that the US farmers plant soybean and other crops that are more profitable than oats instead.

9.1.1 Corn

The US is the world's largest producer and exporter of corn, and yet domestic use consumes a majority of total supplies. Production has increased steadily over time since 1975. Imports also has risen, especially over recent years, to meet the ever increasing domestic demand. Additionally, the US has maintained high export volumes over ~ 40 years with occasional slight fluctuations, indicating importance of corn to the US economy as it contributes significantly to the US agricultural trade balance. Another prominent observation on corn consumption pattern reveals that corn has been consumed primarily by food use, rather than by feed use since 2008, which implies increments in intake of corn by the US people, in line with their pursuit of health diets.

9.1.2 Sorghum

The US sorghum production has experienced decreases roughly over ~ 40 years and slightly increased over recent a couple of years. It was demanded mainly as a feed ingredient before 2011. However, lately sorghum has been consumed more by food use similar to corn for the same reason. In general, total demand and total supplies have dropped since the end of 20th century, as demand of feed use has decreased sharply. Import volume surged in 2012 to respond to a sudden decrease of sorghum production and increase in food use demand. Besides, seed use also has decreased over nearly 30 years, suggesting a shrinkage in the growth of sorghum and explaining the constant reduction of production. Consequently, sorghum's share of the US feed grain production probably drops.

9.1.3 Barley

Similar to sorghum, barley production, total supplies, total demand, and feed use has reduced over time. However, over the recent several years, it seems all statistics has reached a plateau. Different from corn and sorghum, food use demand of barley is invariant over ~ 40 years. The most plausible reason is that beer production guarantees food use demand of barley, since it is the essential ingredient of beers.

9.1.4 Oats

Compared to sorghum and barley, the US oats production has reduced even more quickly. Since 1998, imports have accounted for almost one third of total supplies. Domestic use consumes a vast majority of total supplies and therefore, import volumes has to be maintained high to meet the demand. With increasing demand by food use and decreasing production, the US has shifted from oats exporter to oats importer since 1982. Evidently, as a health food providing numerous nutritional benefits, oats has attained popularity, which results in increasing demand by food use.

9.2 Future Endeavors

Our visual system provides users a set of functions, which enhance user experience and also polish the layout, such as navigation, selection and filtering. However, there still exists room for continued development. A zoom in/out function can be added to the world map so that users are able to zoom in on the map and view any area of their interests. Also, timeline of historical events and time selection widget in the history page can be merged in a way that a description of a historical moment pops up when a user selects the corresponding year.

10 REFERENCES

http://bl.ocks.org/d3noob/10633421
http://bl.ocks.org/d3noob/5987480
https://bl.ocks.org/mbostock/db6b4335bf1662b413e7968910104f0f
http://bl.ocks.org/mbostock/4254963

11 APPENDICES

Figure 22: Overview Chart

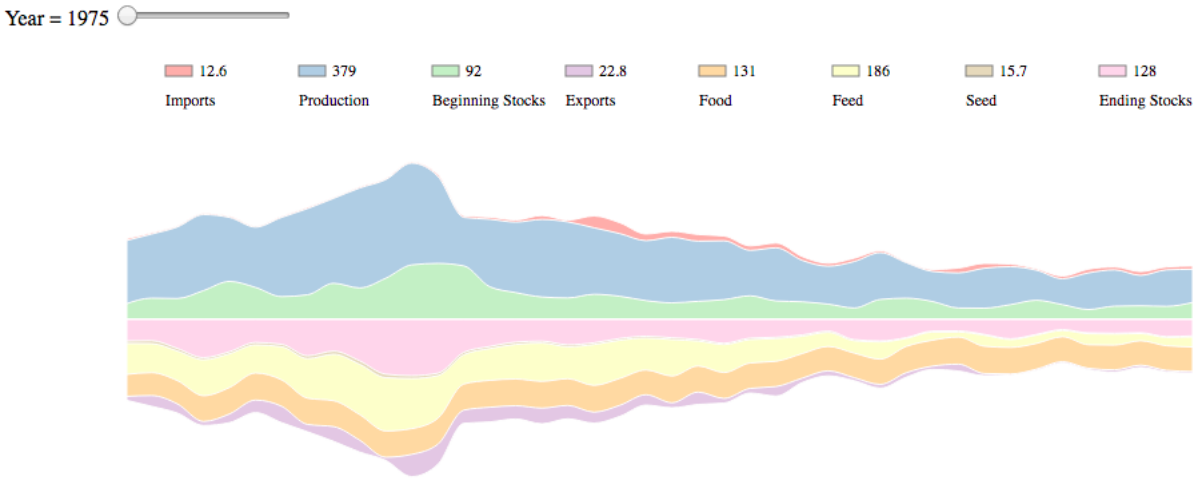

Figure 23: Streamgraph for Barley

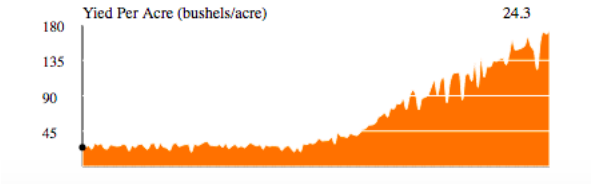

Figure 24: Line Chart For Corn Yield Per Acre

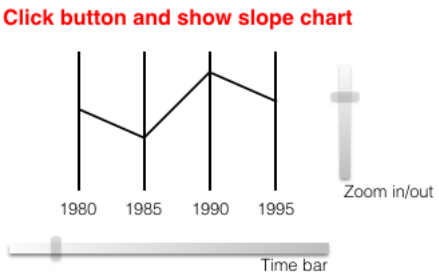

Figure 25: Slope Chart for Supplies and Demands

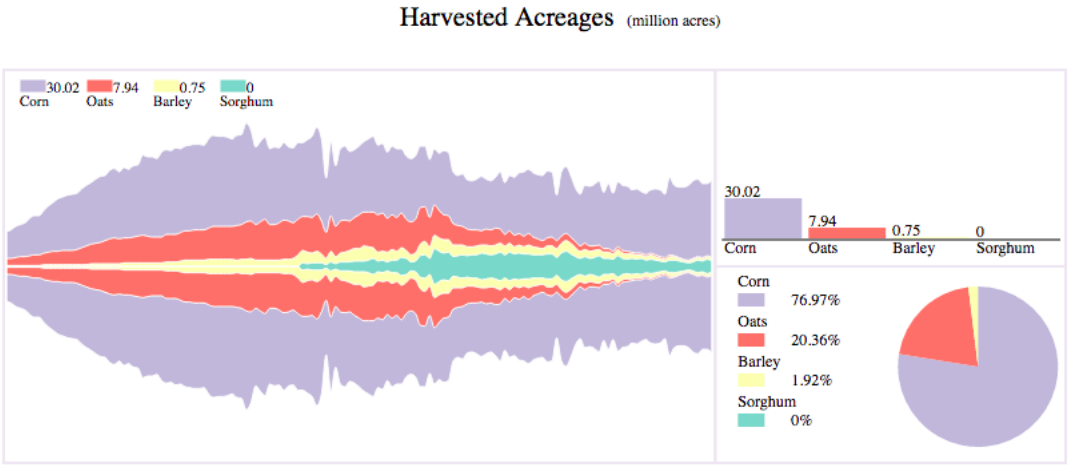

Figure 26: Streamgraph For Feed Grain Harvested Acreages

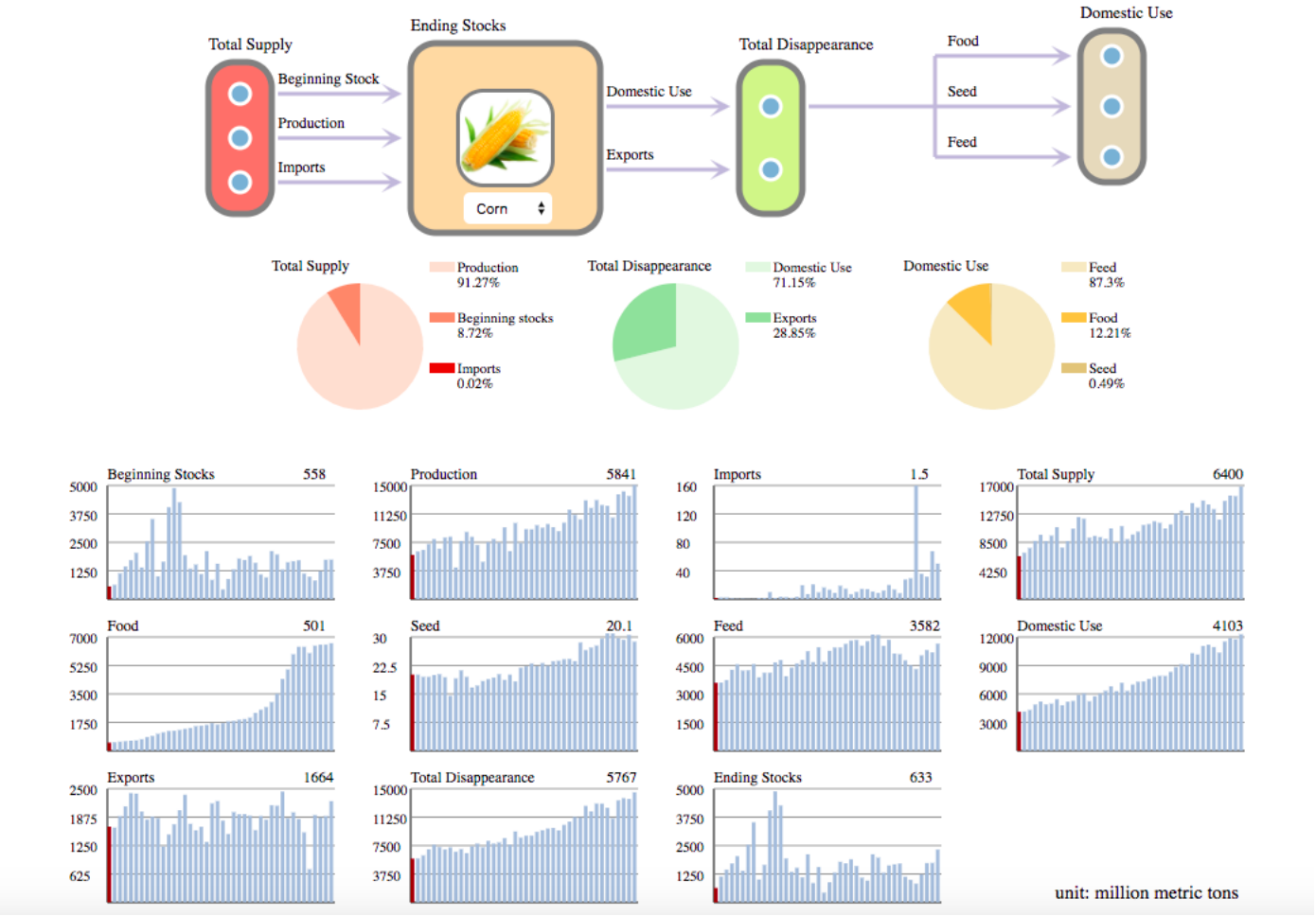

Figure 27: New Design of Overview

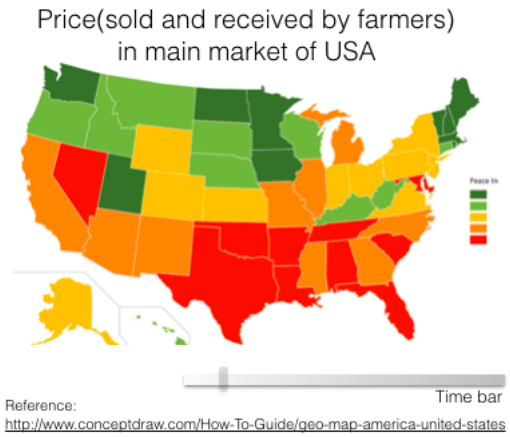

Figure 28: The US Map for Prices at Domestic Markets

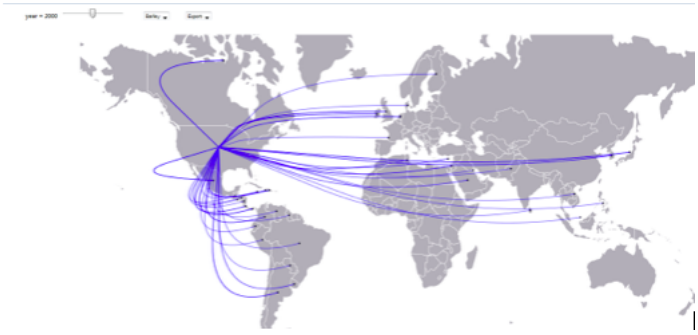

Figure 29: World Map for Exports and Imports

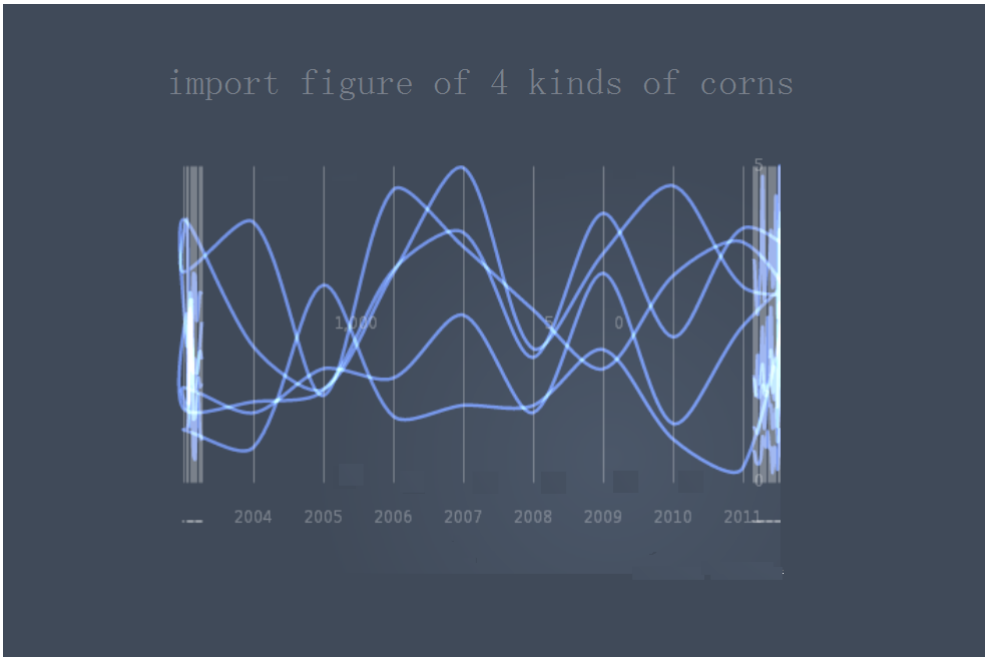

Figure 30: Slope Chart for Exports/Imports

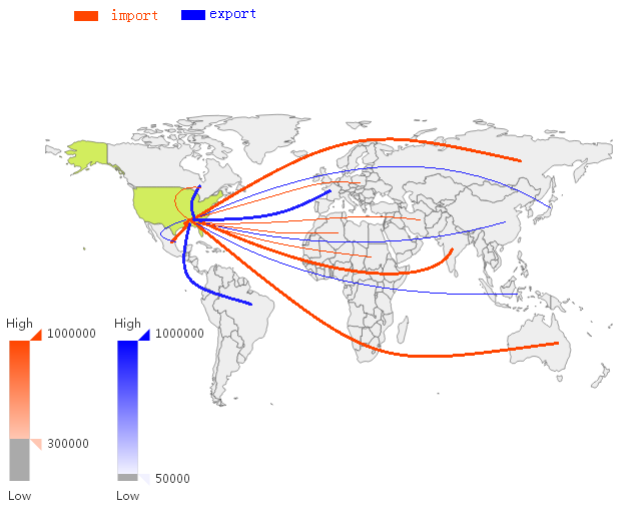

Figure 31: World Map for Exports/Imports

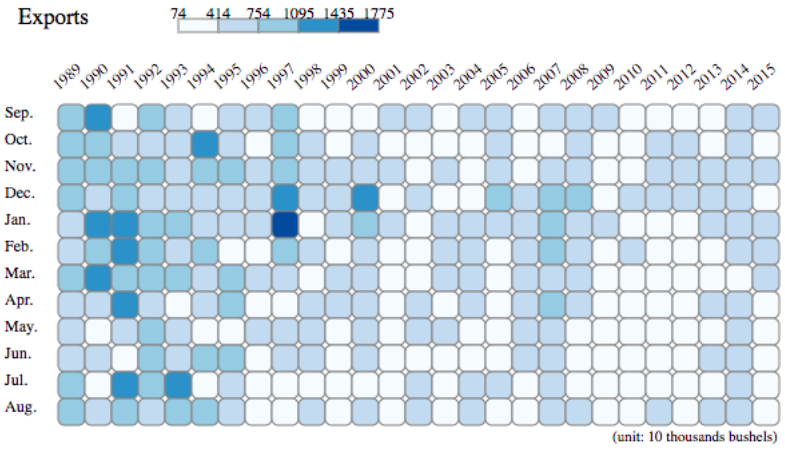

Figure 32: the US Total Monthly Barley Exports

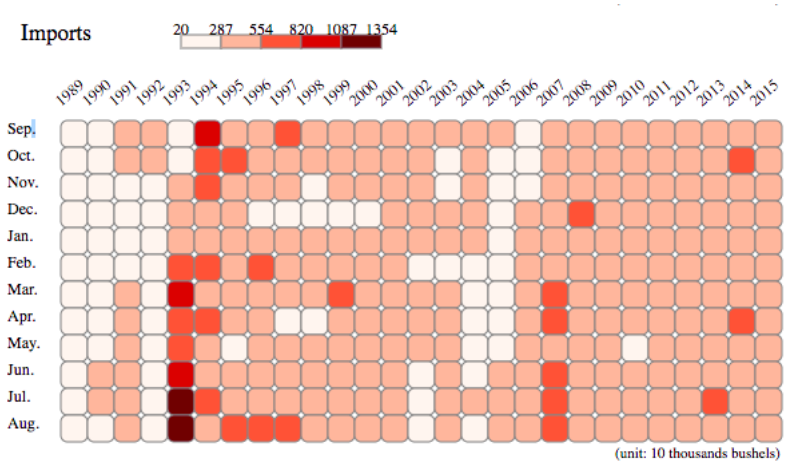

Figure 33: the US Total Monthly Barley Imports

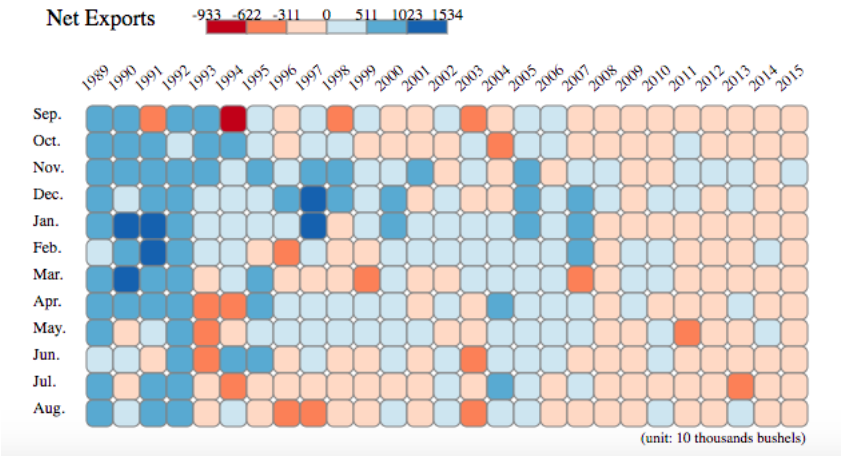

Figure 34: the US Total Monthly Barley Net Exports

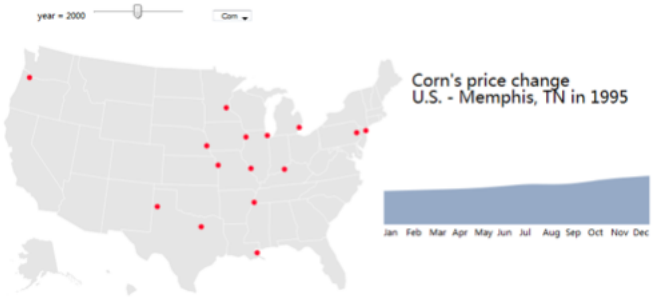

Figure 35: the US Feed Grain Prices

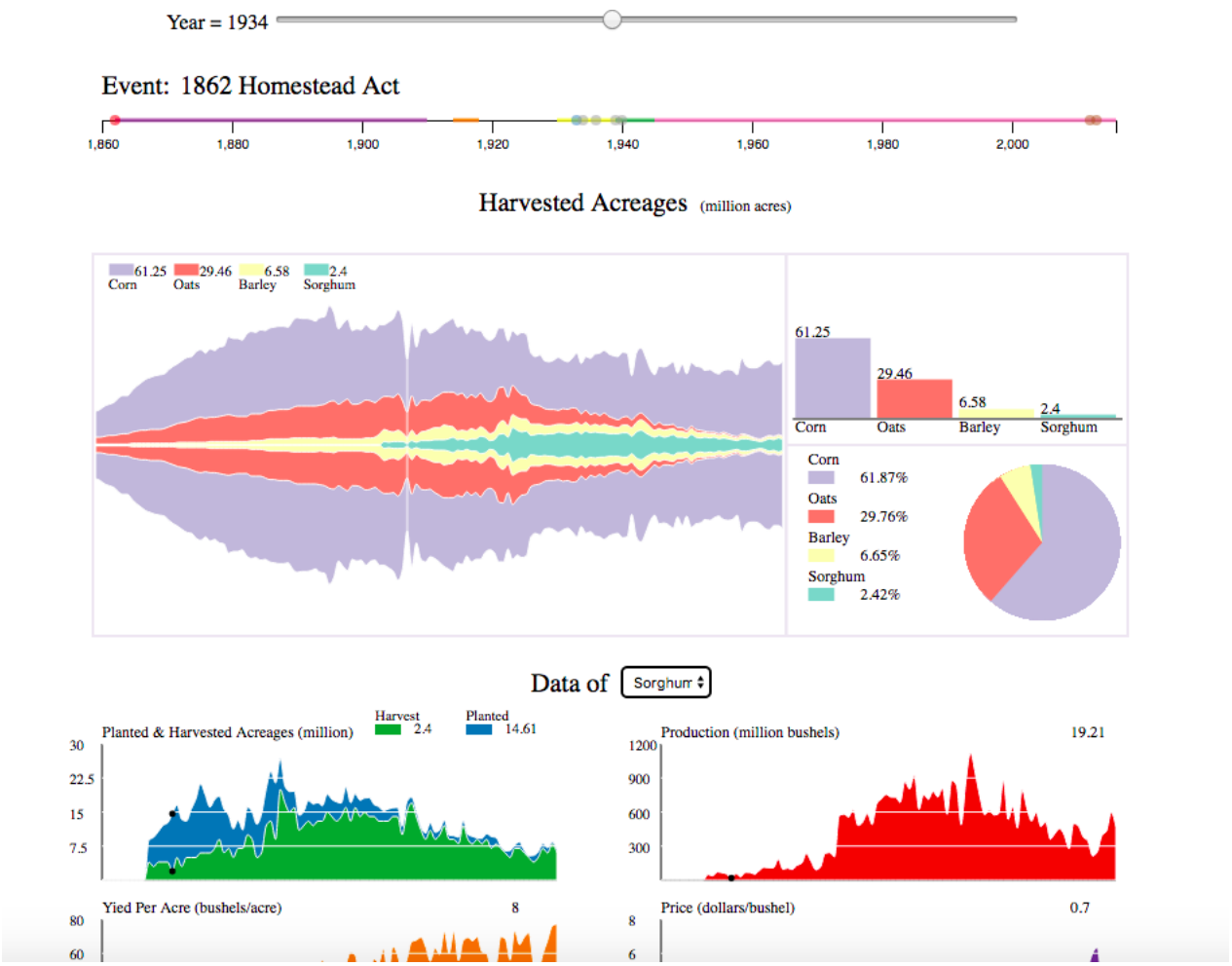

Figure 36: Visualization for Linking Data and History

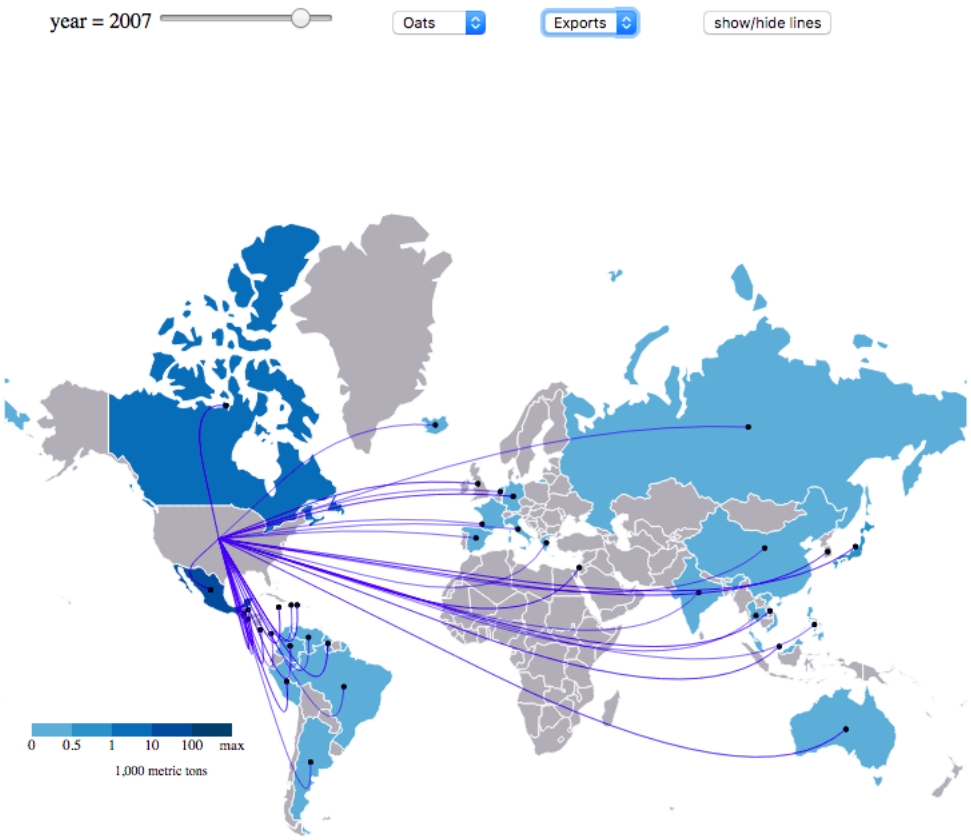

Figure 37: World Map For Exports And Imports

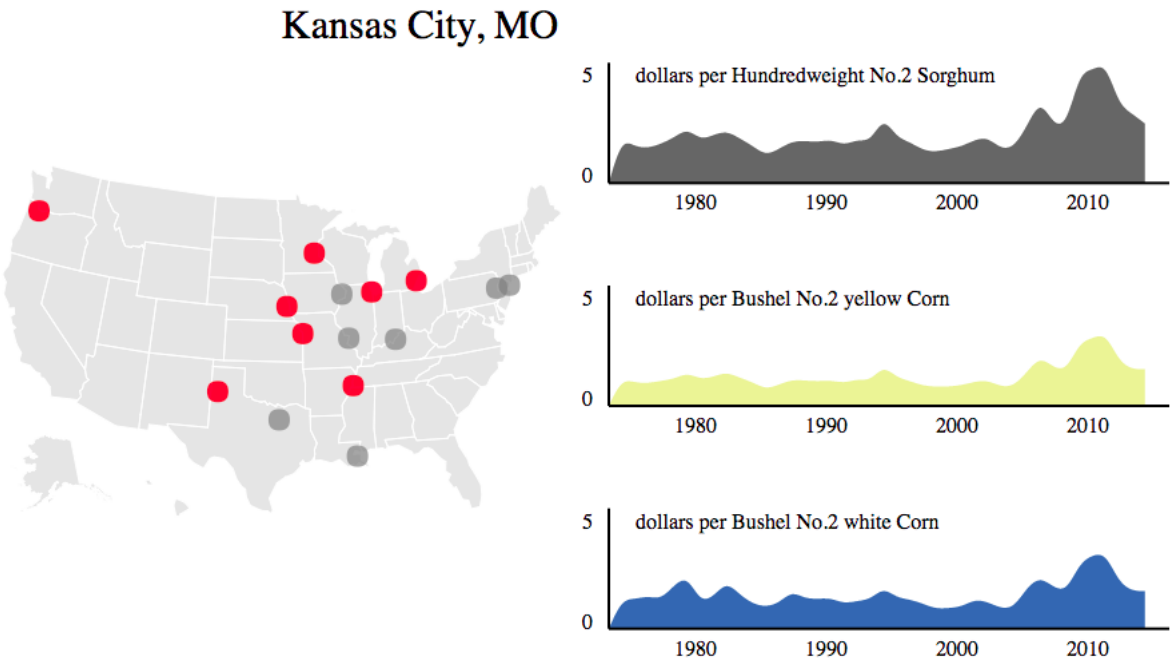

Figure 38: The US Map For Feed Grain Domestic Prices