

Visualizing World Crop Production

Process Book

Christian Felt u0809693@utah.edu

CS6630 November 2020

repository: <https://github.com/christianfelt/dataviscourse-pr-visualizing-world-crop-production>

Background and Motivation

I chose this project simply by answering the question “What data visualization would I be most interested in looking at?” The answer was a map with linked views showing the agricultural production of the world. I’m an amateur gardener and cook and find everything to do with food and farming very interesting. I also love geography and wanted to create a visualization that makes good use of a map.

Goals

The main thing I would like to accomplish is to consolidate my learning this semester by using D3, JavaScript, HTML, and CSS to create a sophisticated visualization website from scratch. The benefits of this are that I will be able to create good visualizations on other topics in the future. The second thing I would like to accomplish is to learn more precisely what is grown in the world, how much, and where. The benefits are that I will have a better understanding of the world’s agriculture.

Related Work

Homework 4 for this class gave me the idea of supplementing a world map with linked views and chronological data. The National Geographic magazines I read as a child gave me many attractive examples of static visualizations that compare countries, while this project gives me the opportunity to make an interactive visualization that allows the user to explore much more data, make more accurate comparisons, and see changes over time.

Questions

This project will help to answer such questions as:

Which countries grow the most or least of a given crop?

What crops are grown in the world and how much in each country?

Where are crops of particular interest grown?

Are there interesting patterns in the geographical distribution of certain crops?

How has the production of a given crop in different countries changed over time?

What are the exact rankings of countries in terms of production of a given crop?

Data

My data is from the “Crops” data set from the Food and Agriculture Organization of the United Nations (<http://www.fao.org/faostat/en/#data>). I am using all the countries, all the items (crops), and all the years that were available when I downloaded the data (1961-2014), but only the “Production Quantity” element, which corresponds to the tonnes of dry crop produced. I downloaded the data from here: <https://data.world/agriculture/crop-production>.

The screenshot shows the FAO FAOSTAT website interface. At the top is the FAO logo and navigation links. The main content area is titled 'Crops' and includes a search bar and a 'Back to domains' button. Below this, there are four panels for configuring the data set: 'COUNTRIES', 'ELEMENTS', 'ITEMS', and 'YEARS'. Each panel has a search bar and a list of options with radio buttons. The 'COUNTRIES' panel shows a list of countries including Afghanistan, Albania, Algeria, American Samoa, Andorra, and Angola. The 'ELEMENTS' panel shows 'Area harvested', 'Yield', and 'Production Quantity'. The 'ITEMS' panel shows 'Agave fibres nes', 'Almonds, with shell', 'Anise, badian, fennel, coriander', 'Apples', and 'Apricots'. The 'YEARS' panel shows a list of years from 2014 to 2018. On the right side, there is a 'Bulk Downloads' section with a table of download options and their sizes.

Download Option	Size
All Data	12.78 MB
All Data Normalized	18.16 MB
All Area Groups	4.02 MB
Africa	2.07 MB
Americas	1.92 MB
Asia	2.33 MB
Europe	1.9 MB
Oceania	434 KB

Data Cleanup

The main cleanup I had to do was to shorten some countries' names, e.g. change “Bolivia (Plurinational State of)” to Bolivia, “Venezuela (Bolivarian Republic of)” to Venezuela, and “People’s Democratic Republic of Korea” to North Korea, and many others. This was necessary partly in order to display the names more compactly, but mainly in order to make the country names match the names in the file (cmu5.csv) that maps between full country names and the shortened ISO country code used in the world.json file (e.g. AFG for Afghanistan).

There were many missing crop production values in the data set, and I left them blank rather than manually interpolating them. As a result, “N/A” will show up for such

country/crop/element/year combinations in my table, white space will show up in the choropleth map and bar chart, and the line chart D3 will interpolate between missing values.

There were some odd strings in the crop names, such as “nes” in “Agave fibres nes,” which would probably be confusing to the user even if I provided a legend that explained what all of them mean. According to the FAOSTAT documentation, NES means “not elsewhere specified,” so I simply replaced “nes” with this full string.

Data Wrangling

The data wrangling was quite extensive. First, I read every row of the .csv data files and inserted all their fields into a single dictionary with the following structure: {countries : crops : elements (especially production) : attributes (especially years) : values}. At the same time, I precomputed the maximum production values for each crop for each year, since I would need these max values for setting the bounds in my scales later, and created sets of the country and crop names, since I would need these for populating my list views.

Then I had to filter and reshape the master data dictionary into different formats, suitable for binding in each of the views. For instance, in the line chart view, I had to create an array of the production values for every year between 1961 and 2014 for the selected crop for each of the selected countries. I used my precomputed max values of crop production for the selected crop for each year to quickly compute the max value over all years, which I needed to set the upper limit of the domain of my y axis scale for this view.

Using the cmu5.csv file from Homework 4, which had a mapping between full country names and the ISO codes in the world.json file, I created a map between the ISO country codes and the full country names in my data files, but I had to change by hand many of these country names to match the names given in the cmu5.csv file (e.g. replace “Democratic Republic of the Congo” with “Congo”).

I had to format the production values into scientific notation, using just the first two digits to label the axes in my line and bar chart views in order to save space, and displaying the units (e.g. “1e+7 tonnes”) dynamically as they changed.

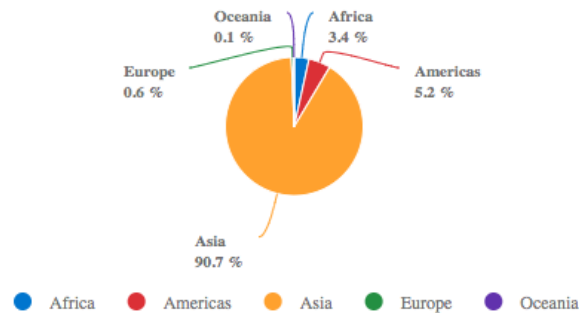
Exploratory Data Analysis

I used Excel to view the .csv data files in tabular form. It seemed obvious that the production values for each year would make a good line chart, and that it would be convenient to compare the production values of different countries in the same year using a bar chart.

Later, I looked at the “visualize data” tools on the FAOSTAT website where I got my data <http://www.fao.org/faostat/en/#data/QC/visualize>. I found that, like me, they had decided to use a choropleth world map to give an overview of production and had used a line chart to show production over time. Unlike me, they had used a pie chart and had aggregated countries into different regions:

Production share of Rice, paddy by region

Average 1994 - 2018



I found the pie chart to be less helpful than the other charts, though, since it is hard to read the tiny slices and to make accurate comparisons between slice areas. The FAOSTAT bar charts for rice production are below:

Seed of Rice, paddy by region

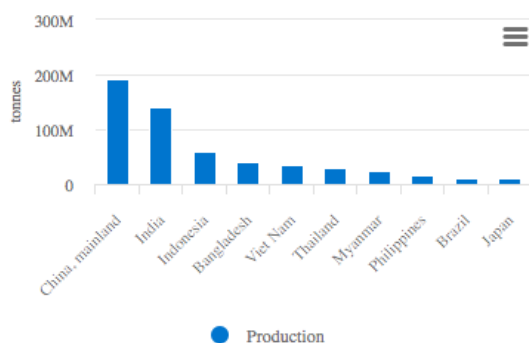
Average 1994 - 2018



No data available.

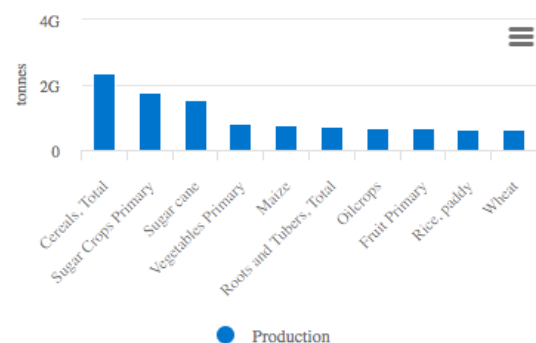
Production of Rice, paddy: top 10 producers

Average 1994 - 2018



Most produced commodities, World + (Total)

Average 1994 - 2018



Seed of Rice, paddy: top 10 producers

Average 1994 - 2018



No data available.

Most produced Seed, World + (Total)

Average 1994 - 2018



No data available.

The FAOSTAT bar charts are easy to read, but I hope that my design improves on them by making the colors of the bars correspond to the colors of the borders of the selected countries.

Thus, it is easier to remember which bar is which. Another way I hope my design improves on FAOSTAT's is that in my choropleth it is possible to select countries by clicking on them on the map rather than having to select them from a scroll list. I chose to leave out the "Seed" statistics altogether since for most country/crop/year combinations, there is no data, so the boxes announcing "No data available" just clutter up the page. Overall, compared to FAOSTAT's, my visualization focuses less on providing access to every detail and more on making the main trends in production easy to see and attractive to explore.

Design Evolution

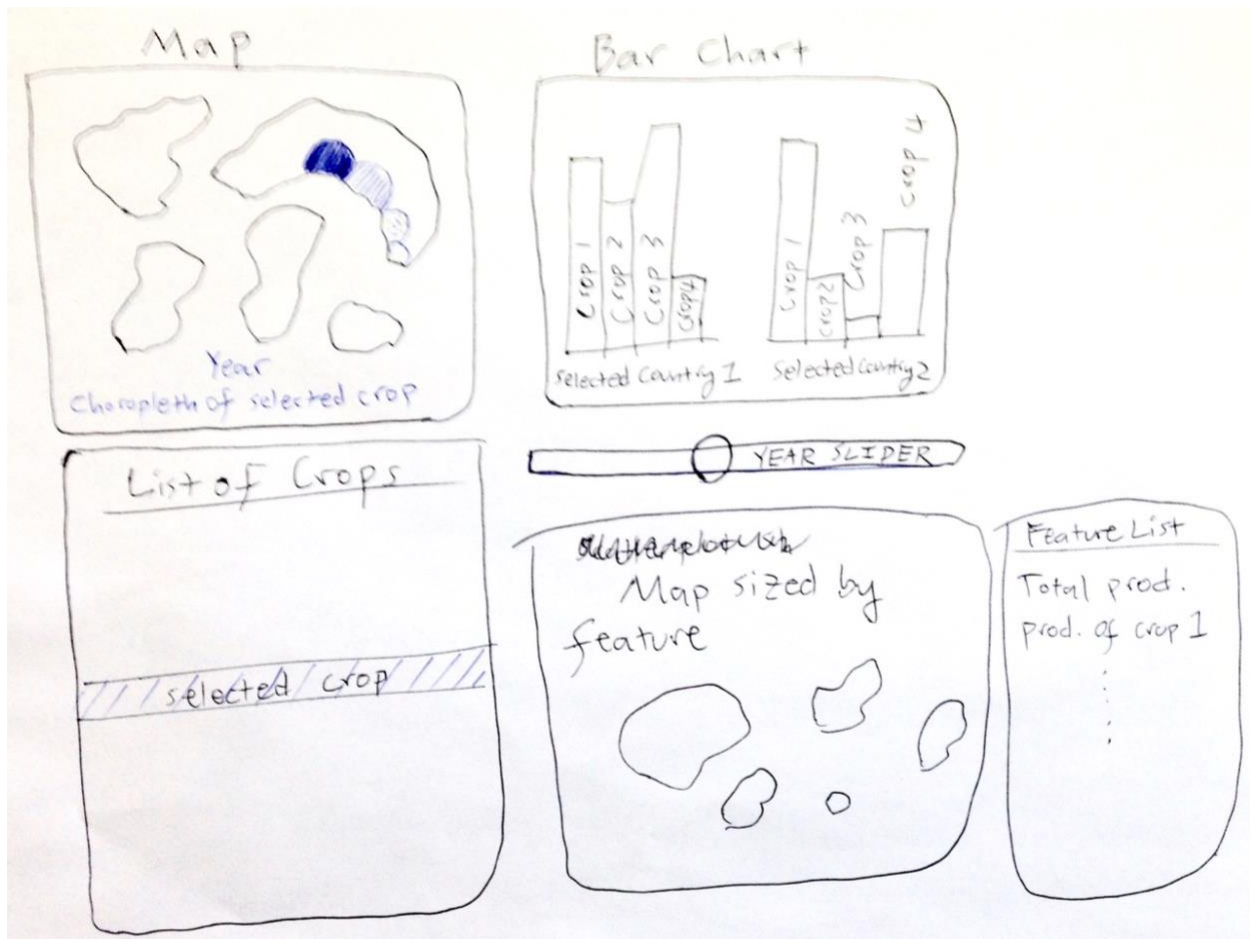
My general goals when starting were: I wanted to have at least 3 linked views that show different aspects of the data (where, when, and how much). It should be easy to find each crop or country. It should be possible to select and accurately compare multiple countries and see how their production has changed over time. There should be some storytelling, guiding the user to the most interesting parts of the data as revealed in these particular views. It should be possible to sort countries by production for each crop and easily find the top and bottom producers (in the table view).

Prototype Design 1:

I began with the idea of a choropleth map and linked bar chart and crop list. Initially, as in the sketch below, I wanted to have another map that would encode production amount using size, i.e. a cartogram. This would have been interesting from a coding and aesthetic point of view, but the information presented would have been redundant with the information in the bar chart and choropleth, so I decided against it.

I chose a choropleth because it looks pretty and will provide an easy way for the user to select a country quickly and get an overview of the relative production levels for that crop throughout the world. For precise comparisons between countries, the bar chart will be used. This way, it doesn't matter that from the map alone it might seem like, say, China produces much more tomatoes than, say, the Netherlands if both were equally dark blue.

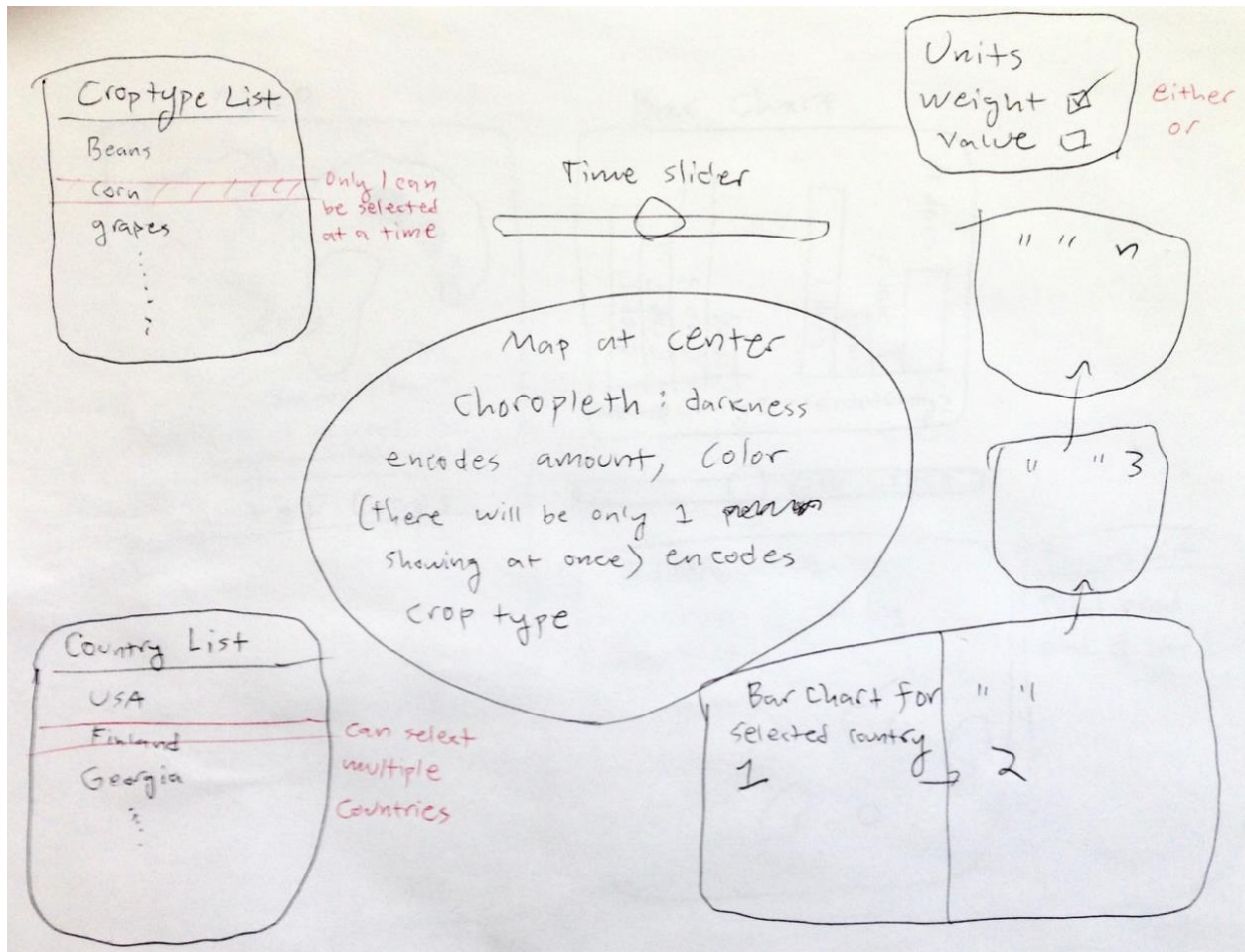
The feature list contains the Cartesian product of the set of all crops and the set of all units, namely weight and monetary value.



Prototype Design 2:

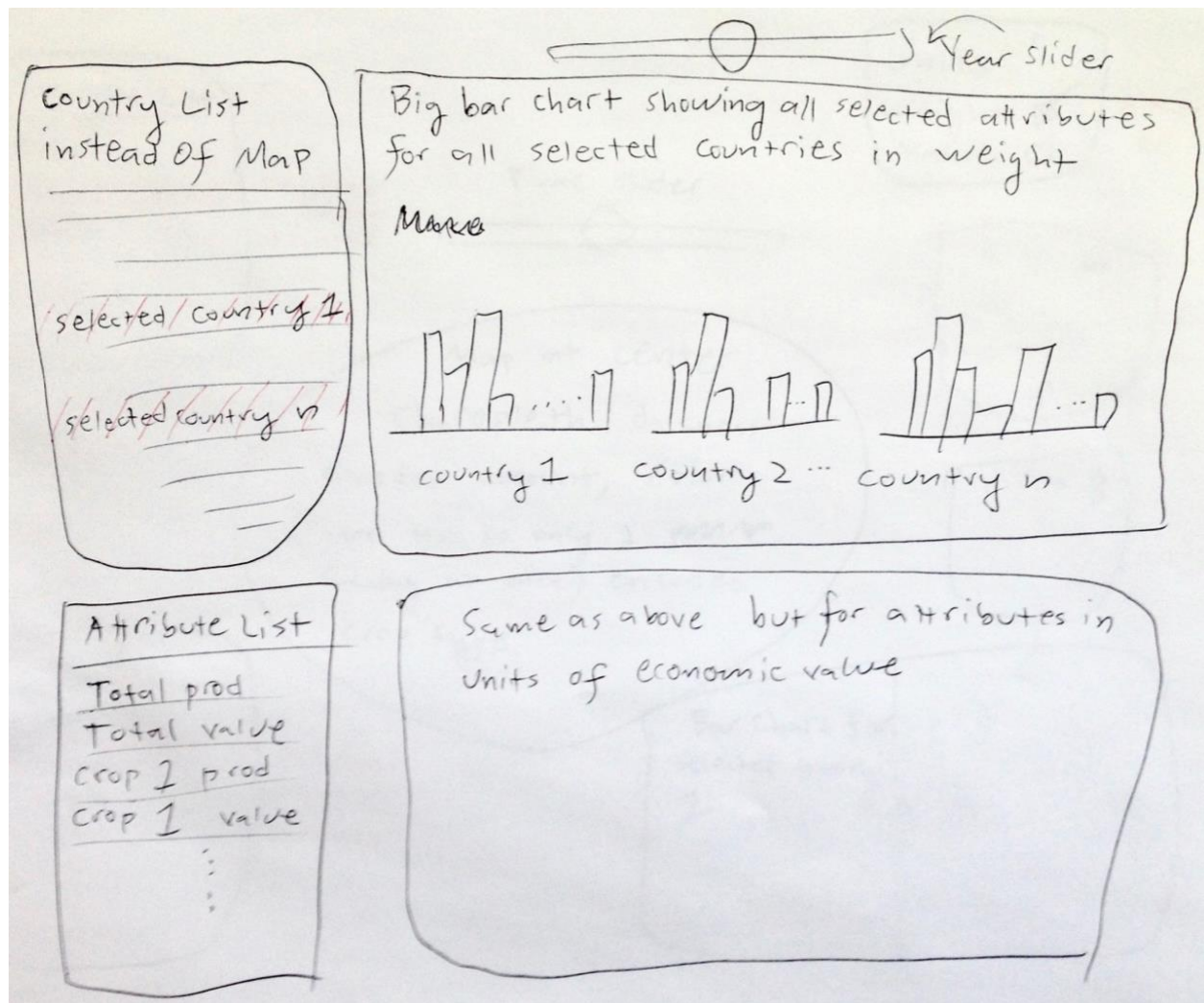
Next, I considered putting the map at the center of the visualization and creating an arbitrarily large number of separate bar charts as needed all around it to show the production of the selected crop in the selected countries. The downside of this is that it makes it harder to compare countries' output than if they are all aligned against the same axes. Here, countries can be selected from a list, as well, in case it is easier for the user to find a country by name than by clicking on it on the map. This would be especially helpful for very small countries.

I also introduce a switch that allows one to choose either weight or value so that the attributes list in the previous prototype can be simplified to a list of crops.



Prototype Design 3:

Next, I considered doing without the map altogether. The user selects countries from a list, and the data for all selected countries and all selected attributes shows up in the bar charts. The top bar chart is in units of weight and the bottom bar chart is in units of monetary value, so now you can see these units at the same time rather than having to toggle them with a switch.



Prototype Design 4:

I won't say too much about this one since it's the same as my final design on the next page, except that for this prototype I considered using little pictures of the most important crops to allow people to select them more easily and pleasantly than by hunting for them in the full list.

Map

Clear Selection

Can select multiple countries
Border changes when selected
Choropleth showing amount of selected attribute
Only one attribute can be selected at a time.

Bar Chart shows the values of the selected attribute for all the selected countries side by side

Attribute List (Crops)

Peas

Wheat

grapes

⋮
⋮
⋮

value ☒
weight ☐

only one

only one
can be
selected
at a time.

Graphic
selection



Wheel for
most important
crops?



Year slider: updates

Both Map and Bar Chart.

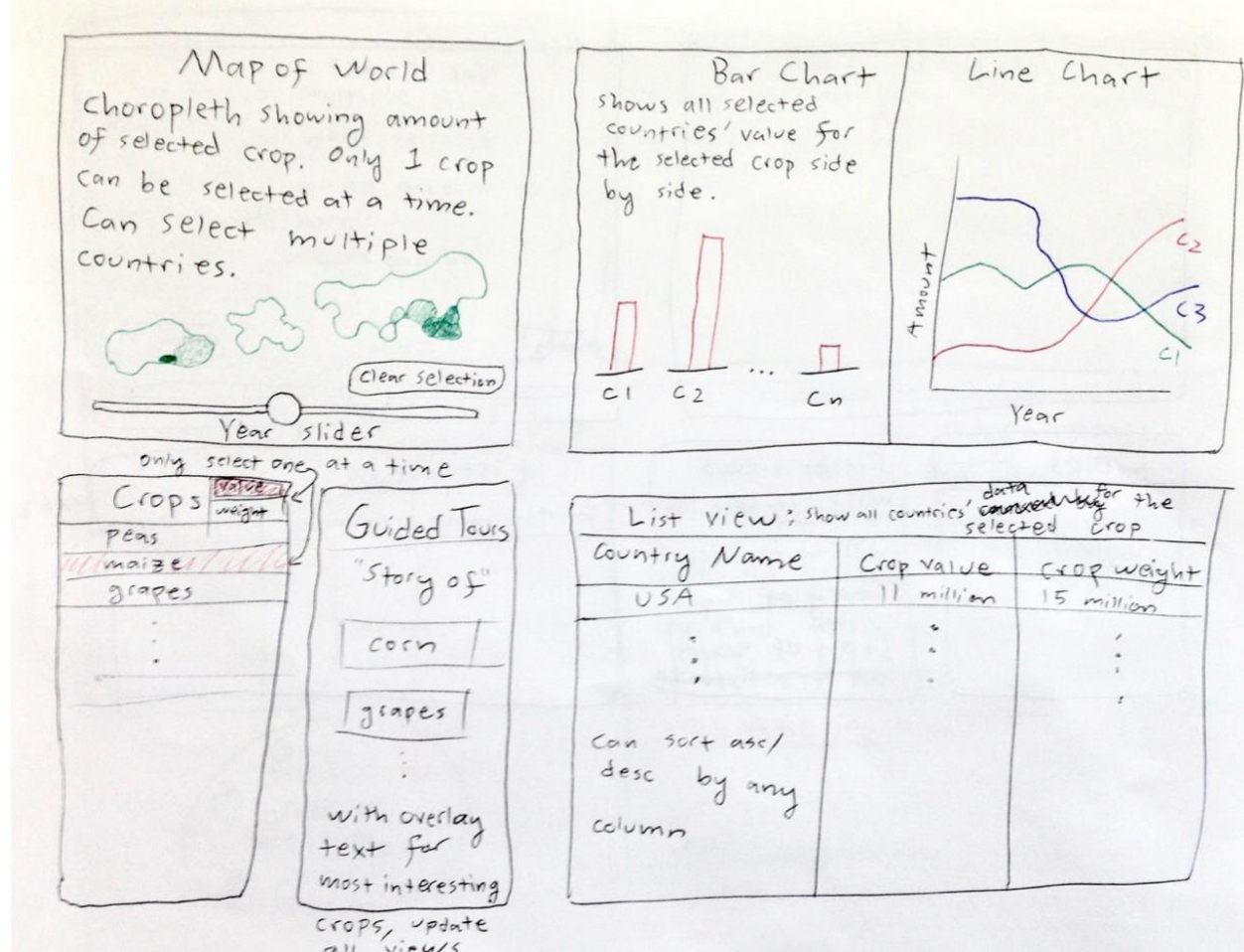
List view: For given attribute,
shows all countries in table ranked
by that attribute ~~and with other~~
~~attributes if helpful in other columns.~~

country Name	Crop Value	crop weight	

can sort ascending or descending by any column

Final Prototype (described in detail, since I used this text and image to help me write the code):

The choropleth allows the user to select countries easily and get a quick overview of the production of the selected crop throughout the world. Only one crop can be selected at a time, so the choropleth will have only one hue at a time. The crop is selected using the "Crops" table in the bottom left. The map view has a year slider and "clear selection" button. Selections made in one view will propagate to all the others. Multiple countries can be selected; their borders will be thickened and darkened. The bar chart shows the amount of the selected crop grown in all of the selected countries side by side, making accurate comparisons easy. The line chart shows the production of the selected crop in the selected countries over time, making it easy to see how crop production has changed over time. The table view in the bottom right shows the data for the selected crop for every country. It is linked only to the "Crops" panel, not to the other views. The columns can be sorted ascending or descending by any attribute by clicking on the column header, which will then be highlighted. The "Guided Tours" panel allows you to choose from 3-5 crops for which I will create an overlay with text, explaining what is especially interesting about the given crop in each of the views. Tooltips with country name, crop type, unit type, and exact value will appear whenever you mouse over a country, bar, or line.



My final edits to this design, as I implemented it, were to color the borders of each selected country to match the color of the corresponding line and bar and to place the line

chart under the bar chart rather than by its side. I also removed the value/weight switch and the crop value column from the list view, since I used only production weight, not value.

Below are some selected screenshots of my visualization in chronological order.

The first important thing I learned was how to manage HTML div formatting using flex and grid layouts, and I laid out the divs for every element I wanted in my final visualization even before I had written any of the content for them. Then I drew a blank map and read in the crop data:

World Crop Visualization

Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693



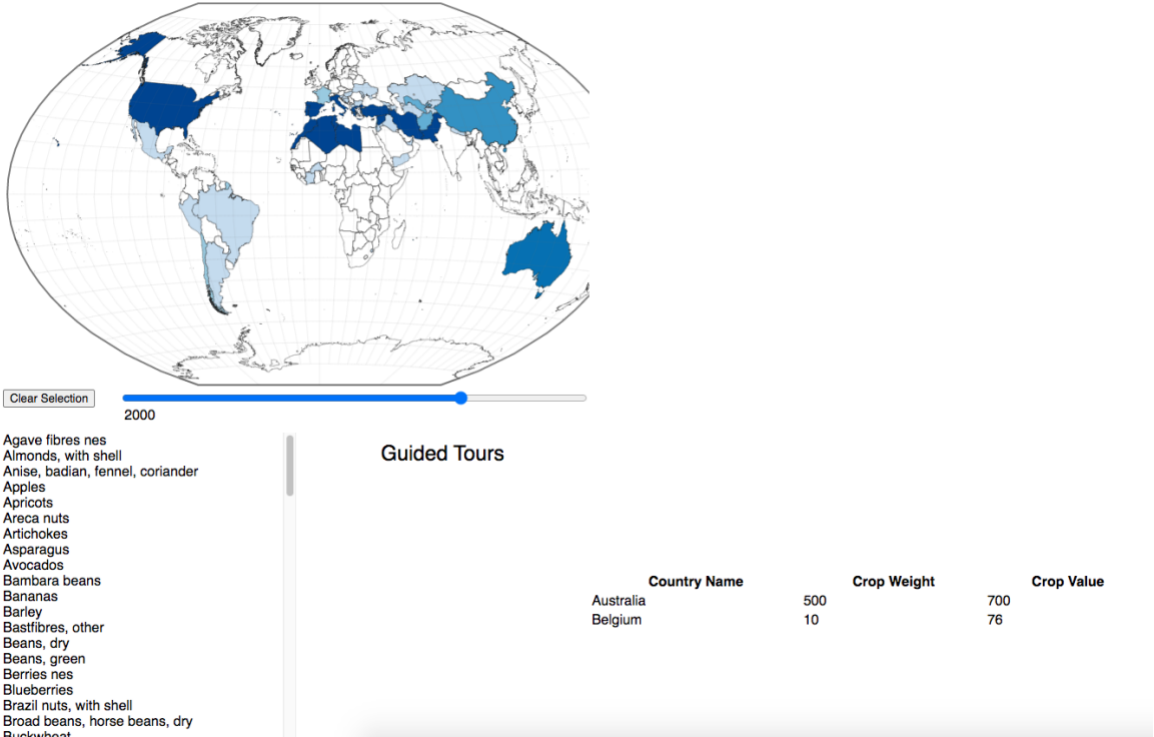
Almonds, with shell
Almonds, with shell
Almonds, with shell
Apples
Apples
Apples
Apricots
Apricots
Apricots
Artichokes
Artichokes
Artichokes
Bananas
Bananas
Bananas
Barley

Guided Tours

Next, I wrangled the crop data into the form I needed to implement the choropleth:

World Crop Visualization

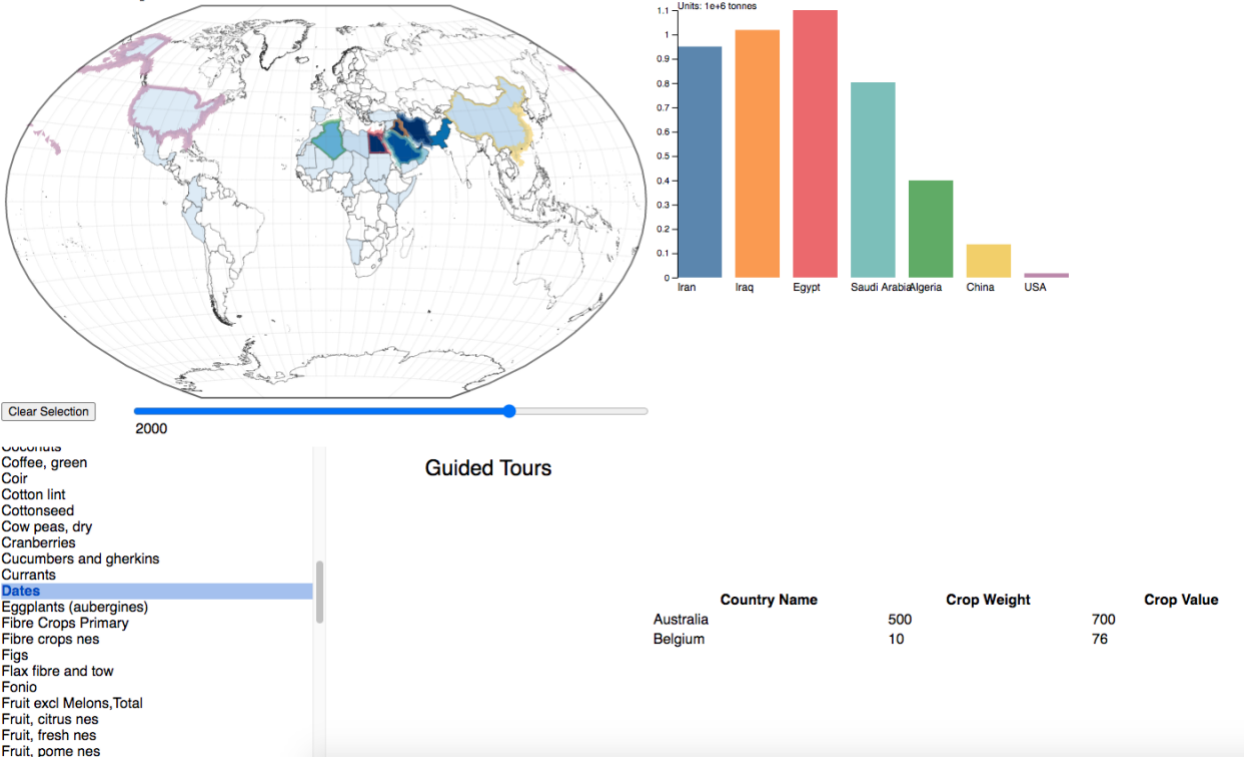
Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693



Next, I implemented the bar chart:

World Crop Visualization

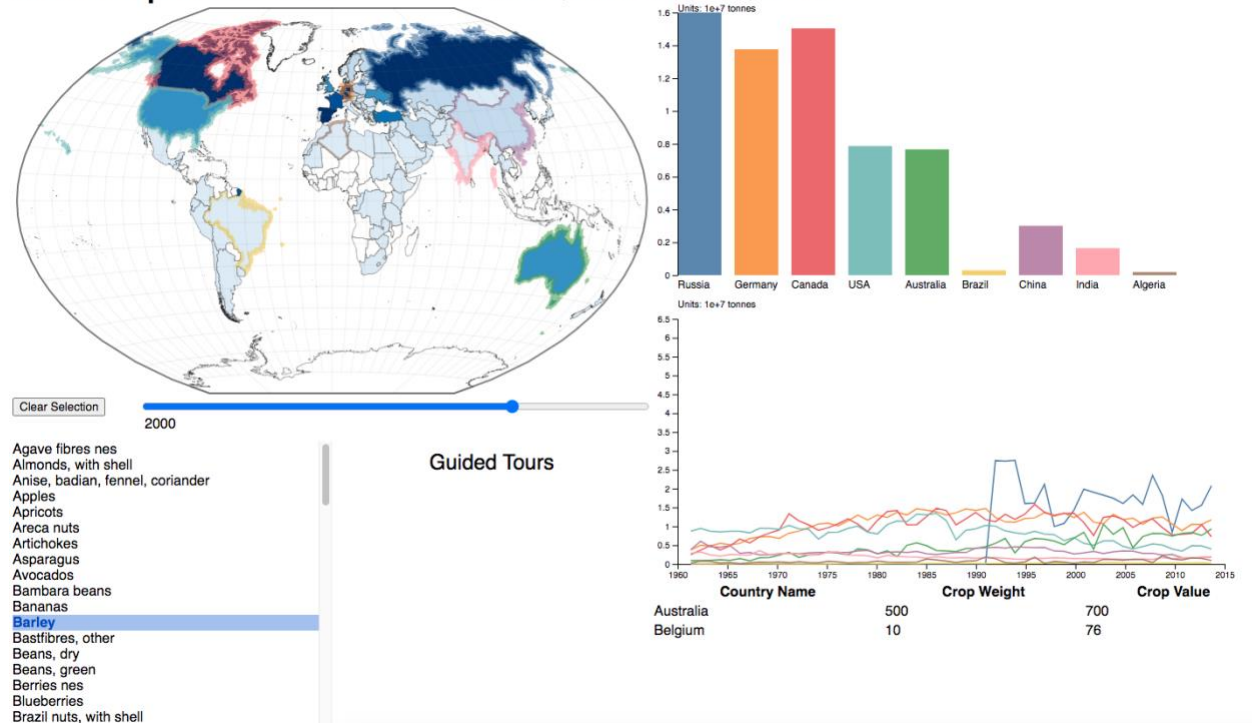
Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693



Next, I implemented the line chart:

World Crop Visualization

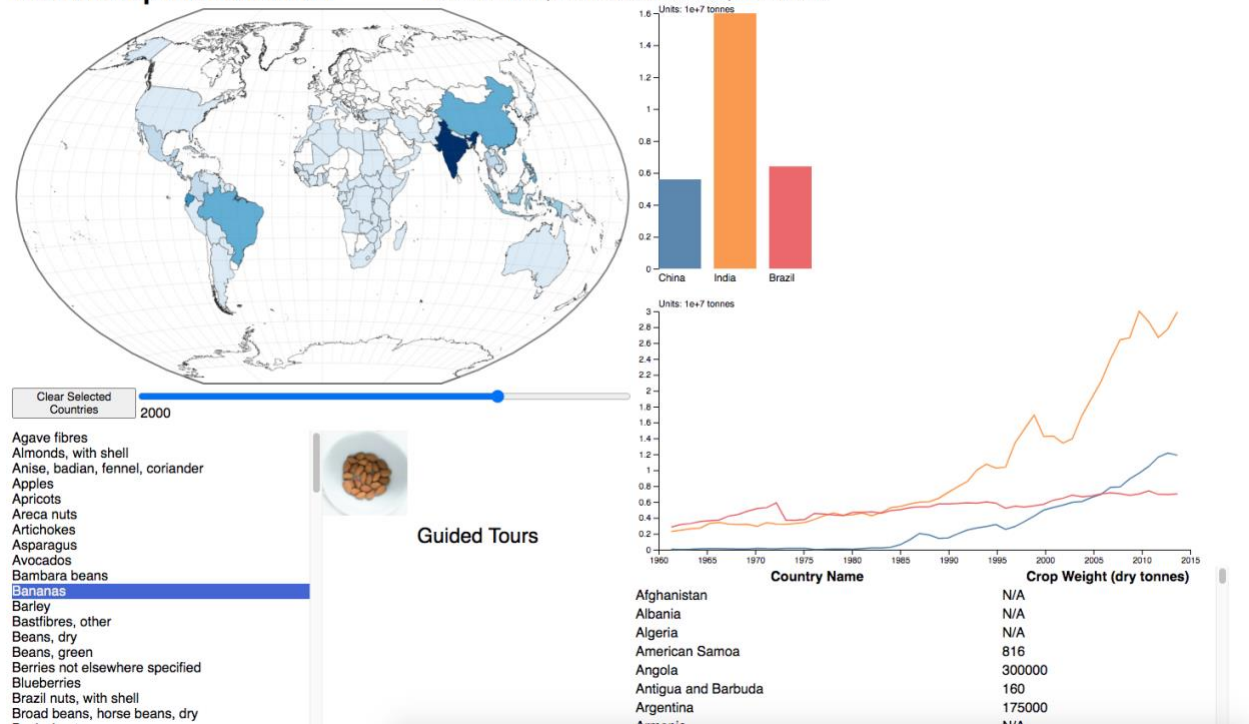
Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693



Next, I implemented the table:

World Crop Visualization

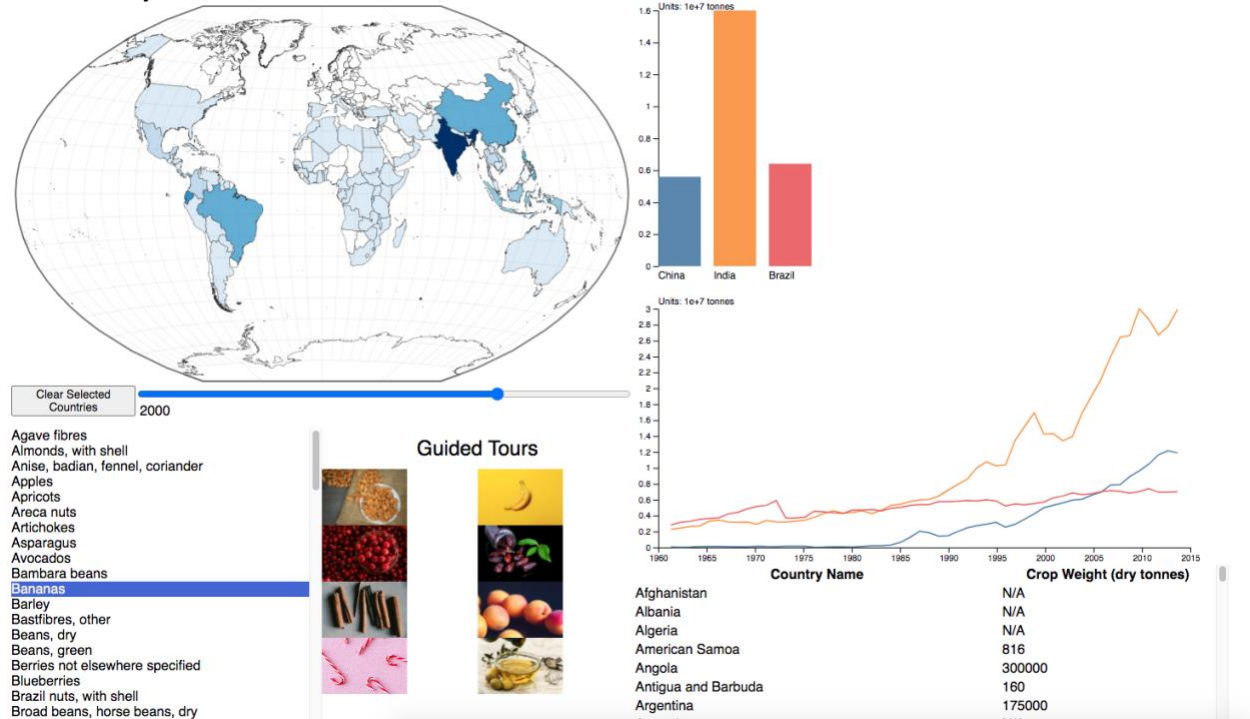
Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693



Next, I implemented the guided tours:

World Crop Visualization

Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693

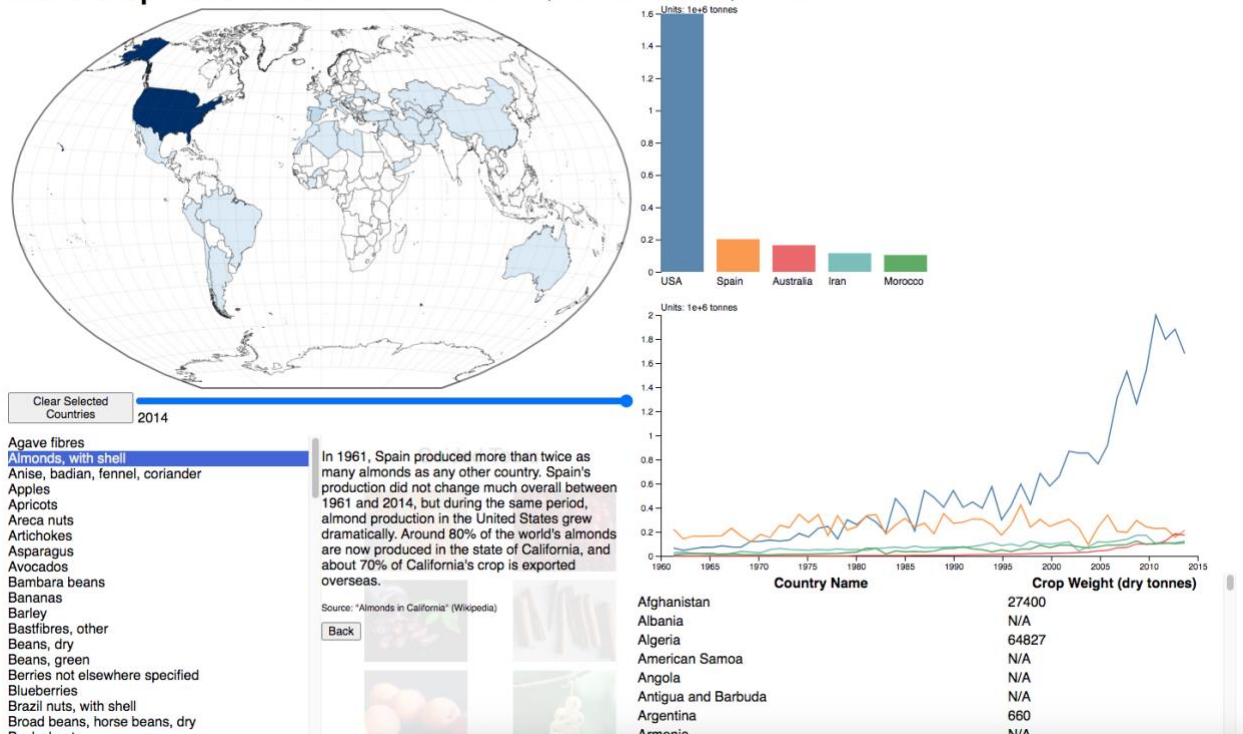


Next, I tried scaling the bar chart y axis the same as line chart, that is, according to the max production value over all years, not changing it according to the max of the given year. This looked better and was easier to read, especially if you were using the year slider to animate the bar chart.

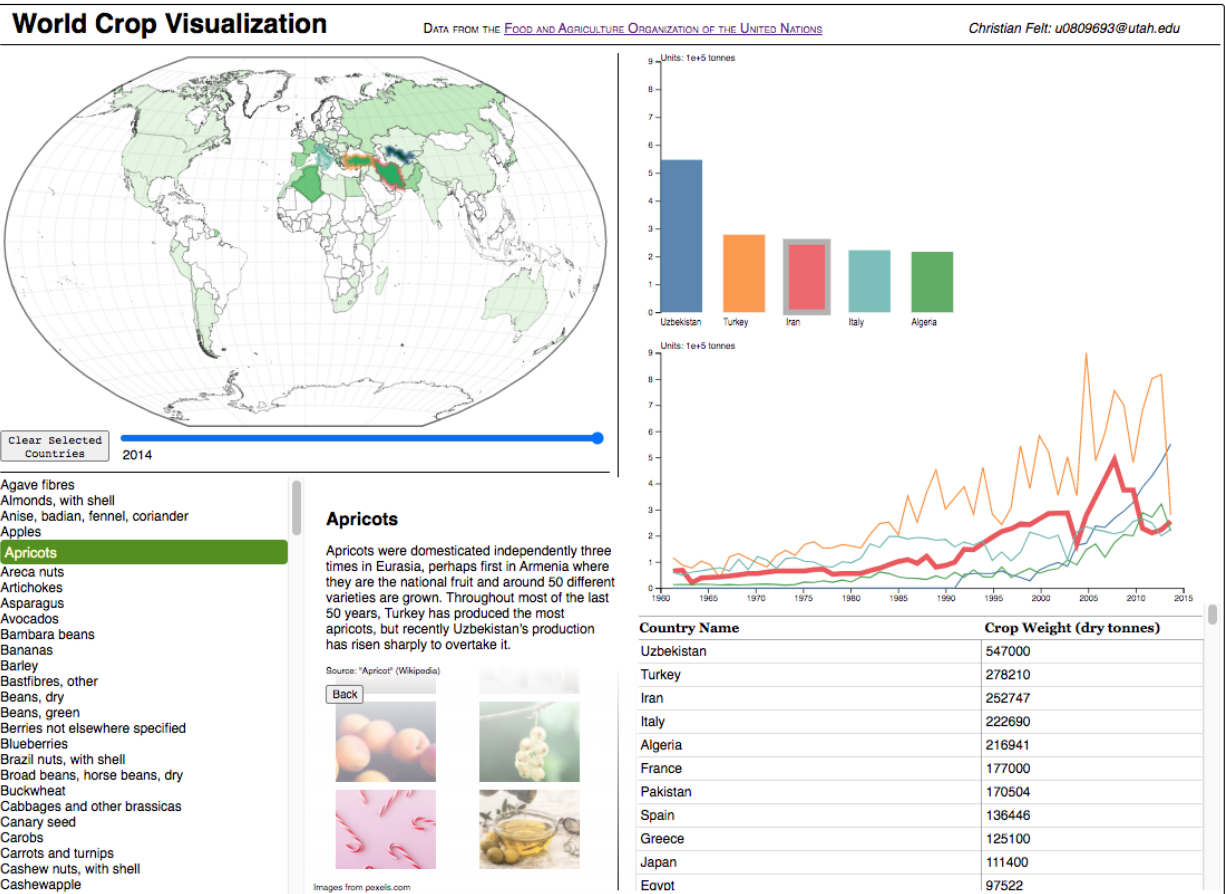
I made small fixes/improvements such as highlighting country borders on the map when you select the country from the table, loading the table with it sorted in descending order by weight, rather than ascending alphabetical order by name. I fixed a bug with countries that have spaces in their names not getting selected correctly by id for mouseover events, and a similar bug with commas in crop names. I fixed a bug where I was calling nice() on the axis scale but not the bar scale so they were slightly out of alignment. Also fixed tooltips so that their data was updated when year or crop were changed, and many other little things.

World Crop Visualization

Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693



Finally, I changed the choropleth from blue to green, since green seems more thematically appropriate for plants, and it doesn't really matter what the choropleth color is. Unfortunately, there is no easy way to change the color of the slider for Chrome, or I would have made that green, too. I also framed the visualization, chose different fonts, drew some lines separating the divs, and generally did a lot of little tweaks to make it look nicer.



Implementation

My final visualization is shown in the image above. The world map is a choropleth that gives you a general overview of the amount of the selected crop grown in each country. The bar chart allows for more precise comparisons between selected countries. The line chart shows the change in production of the selected crop in the selected countries over time. The table shows all countries and their production of the selected crop for the selected year. Countries can be selected by clicking on their shapes on the world map or by clicking on their names on the table. The year can be changed using the slider below the world map. The crop is selected in the list of crops at the bottom left. The lines are animated to be drawn from left to right. Selected countries are given a colored border to match the color of the corresponding bar and line for that country. The lines are enlarged when you mouse over them, and the corresponding bars are given a grey outline at the same. The bars are outlined when you mouse over them and the corresponding line is enlarged. Tooltips on the country shapes on the map and the bars in the bar chart show the country name and production value for the selected crop and year, and tooltips in the line chart show the name of the country that line corresponds to. The guided tours consist of eight images of particularly interesting crops. When you click on them, the main countries that grow them are selected in all the views and a box with explanatory text and a "back" button appears in the Guided Tours pane. The columns of the table can be sorted in ascending or descending order by clicking on the column headers. The default sort is

descending by crop weight, since this was the sorting I as a user most frequently wanted to select. The selected crop and countries persist when you change the year so you can use the slider to watch the change over time in the bar chart and choropleth. The axis of the bar chart is scaled the same as the axis of the line chart to make it easier to compare production changes over time (rather than if the bar chart y axis were rescaled according to the max crop for the given year only). There is a button below the map to clear the selected countries.

Evaluation

This visualization is effective for getting a quick and accurate idea of where a given crop is grown, how much, and how this has changed over time (from 1961 to 2014). It is fun to search it for “fun facts,” e.g. that Italy is the world’s largest grower of artichokes, and India is the largest producer of coir and jute (used to make gunny sacks). Germany and Eastern Europe are the largest growers of gooseberries. The USSR was the largest producer of barley by far and since its collapse no other country has come close to producing even half so much. In general, there were surprisingly many crops that were grown very predominantly in just one country. China was by far the largest producer of Tallowtree seed and many other crops. Mexico was by far the largest producer of avocados. Bambara beans, a fatty legume which ripens underground like a peanut, are grown mainly in West Africa. The United States is the largest producer of almonds, followed by Spain. Turkey was by far the largest producer of apricots (which is the most delicious flavor of Turkish delight) until it was recently surpassed by dramatic growth in Uzbekistan. The United States, Canada, and Chile together grow almost all of the world’s cranberries. In general, China’s agricultural production of many crops has grown dramatically since the 1990s.

One weakness of this visualization is that it only shows crop production in terms of weight, not value, area harvested or other units. Still, I wasn’t terribly interested in seeing these other units, and it would have added complexity both to the implementation and the user interface. As a user, I wish there was an informative blurb about every crop, not just the Guided Tours, but still one can easily right-click and look up a crop on Wikipedia.

It can be a little hard to tell the countries apart on the line chart when there are lots of lines, and it is hard to distinguish nuance or trends in the lines that lie close to the bottom the whole time. Only the main producers show up well on the line chart. Still, the bar chart and table view make precise comparisons easy.

Further improvements I could make would be to include other attributes besides crop weight, e.g. monetary value or percentage exported, and to have another visualization to show all the crops grown in a particular country (as opposed to showing all the countries that grow a particular crop, as I do now). I could also add a choropleth legend, and find and add the data for 2014 to present, and pre 1961 years, too, if possible.