# Visualizing World Crop Production

Process Book

Christian Felt u0809693@utah.edu

## 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 a dynamic 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?

//(maybe)What are the exact rankings of countries in terms of production of a given crop?

//(maybe) What is the economic value of various crops in various countries?

//(maybe) How has the economic value of a given crop in different countries changed over time?

# Data

My data is 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, but only the "Production Quantity" element, which corresponds to the tonnes of dry crop produced.



# 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 my file that maps between full country names and the shortened ISO country code used in my 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, nothing will show up for such country/crop/element/year combinations in my choropleth map and bar chart, but in the line chart D3 will automatically interpolate between them.

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." This would seem to imply that agave fibres appear in other items on the list, yet I could not find them anywhere, so there seemed to be no harm in just deleting the "nes." But in the case of "berries nes," since there were also e.g. "blueberries" on the list, I replaced the term with "berries not elsewhere specified."

## 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.

I created a map between the ISO country codes in my geojson file and the full country names in my data files.

I formatted 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.
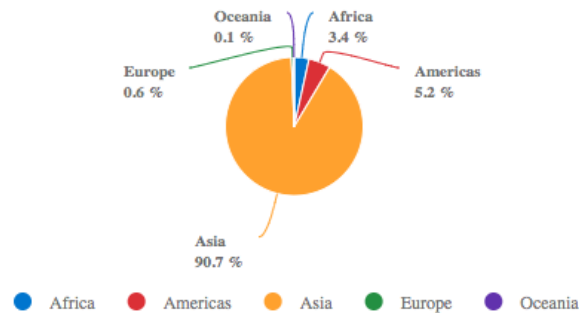
## Exploratory Data Analysis

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

Oceania 0.1 %    Africa 3.4 %

Europe 0.6 %    Americas 5.2 %

Asia 90.7 %

● Africa  ● Americas  ● Asia  ● Europe  ● Oceania

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:
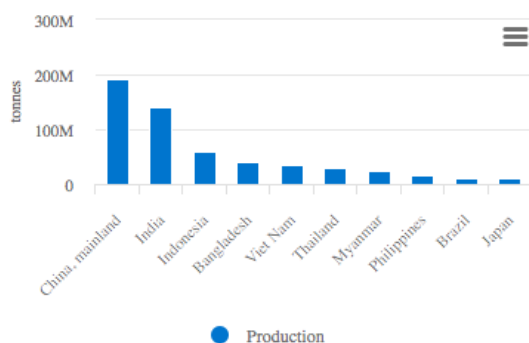
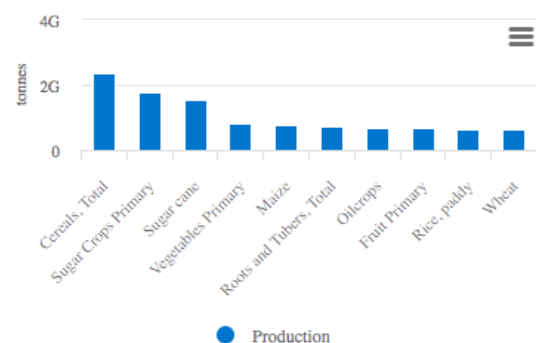Seed of Rice, paddy by region

Average 1994 - 2018

No data available.

Production of Rice, paddy: top 10 producers

Average 1994 - 2018

300M

200M

100M

0

tonnes

China, mainland · India · Indonesia · Bangladesh · Viet Nam · Thailand · Myanmar · Philippines · Brazil · Japan

● Production

Most produced commodities, World + (Total)

Average 1994 - 2018

4G

2G

0

tonnes

Cereals, Total · Sugar Crops Primary · Sugar cane · Vegetables Primary · Maize · Roots and Tubers, Total · Oilcrops · Fruit Primary · Rice, paddy · Wheat

● Production

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

General goals: I want 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. //(maybe)(There should be a convenient way to show both weight and monetary value of crops.) 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.
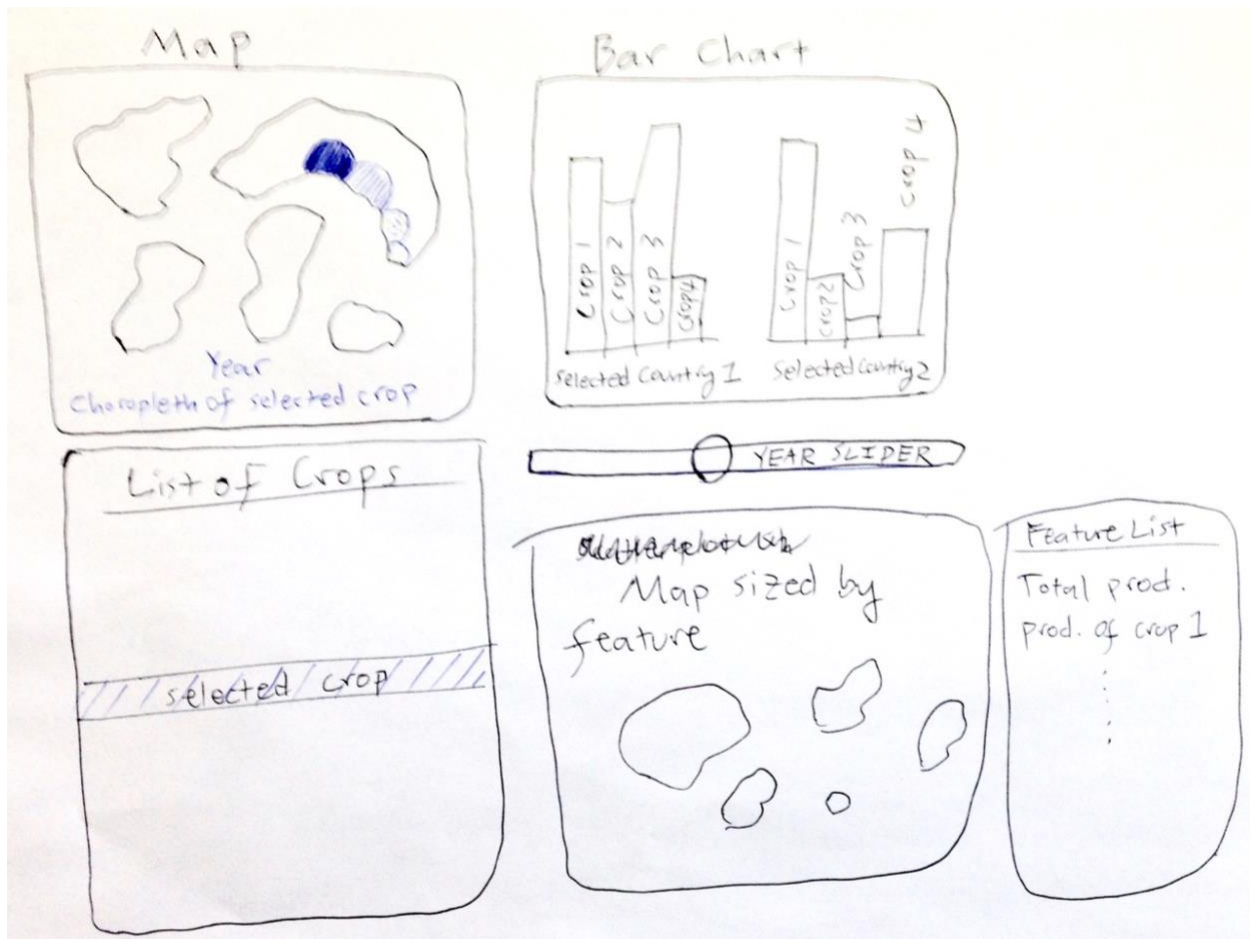
Prototype Design 2:

Here, 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.
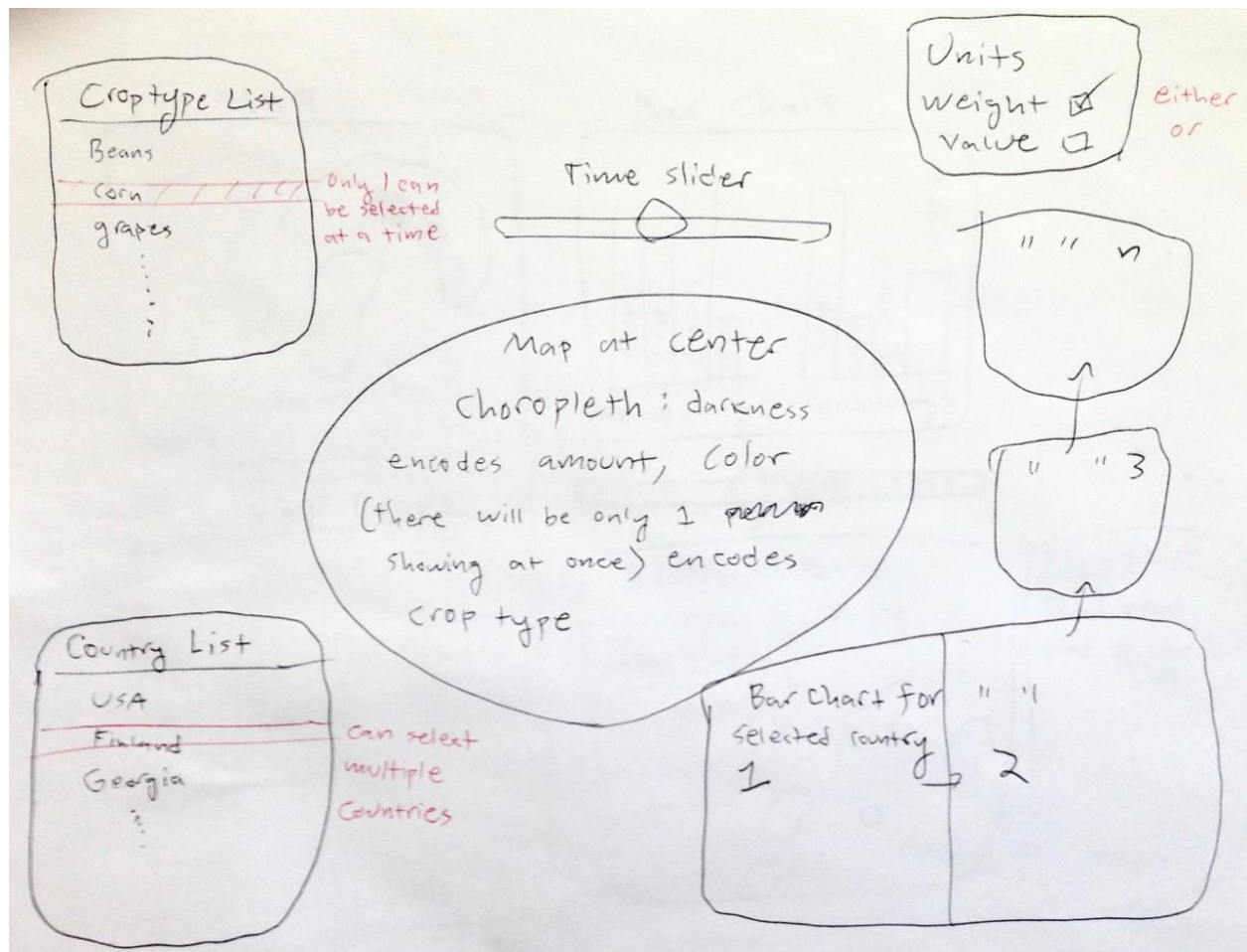
Prototype Design 3:

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.
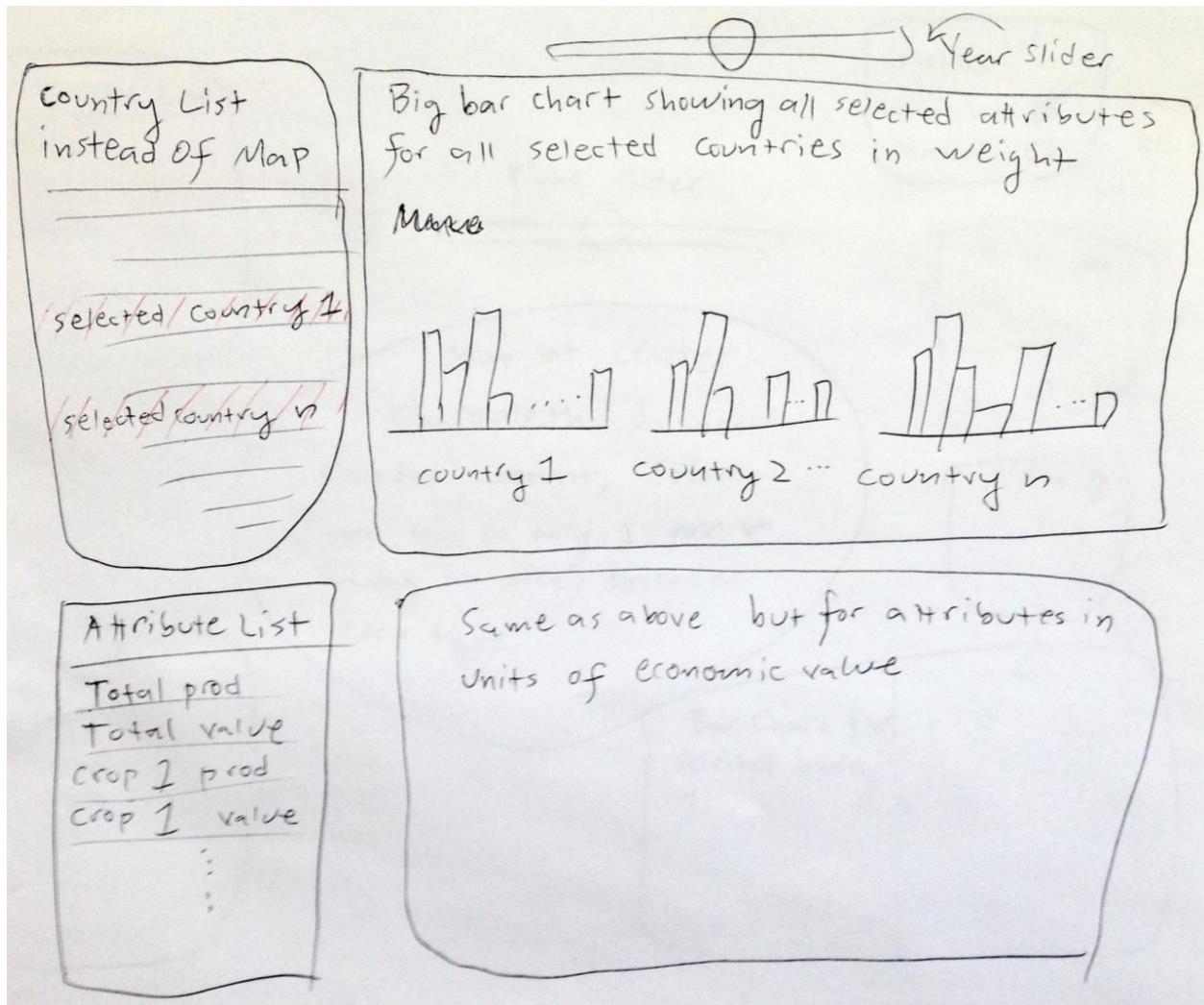
Prototype Design 4:

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)**

| value ☑ | only |
| weight ☐ | one |

Peas
Wheat
grapes
.
.
.

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 Design (described in detail, since I'll be using 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. //(maybe)(There is also a switch in this table that allows the user to display either the monetary value or the weight of crops.) 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.
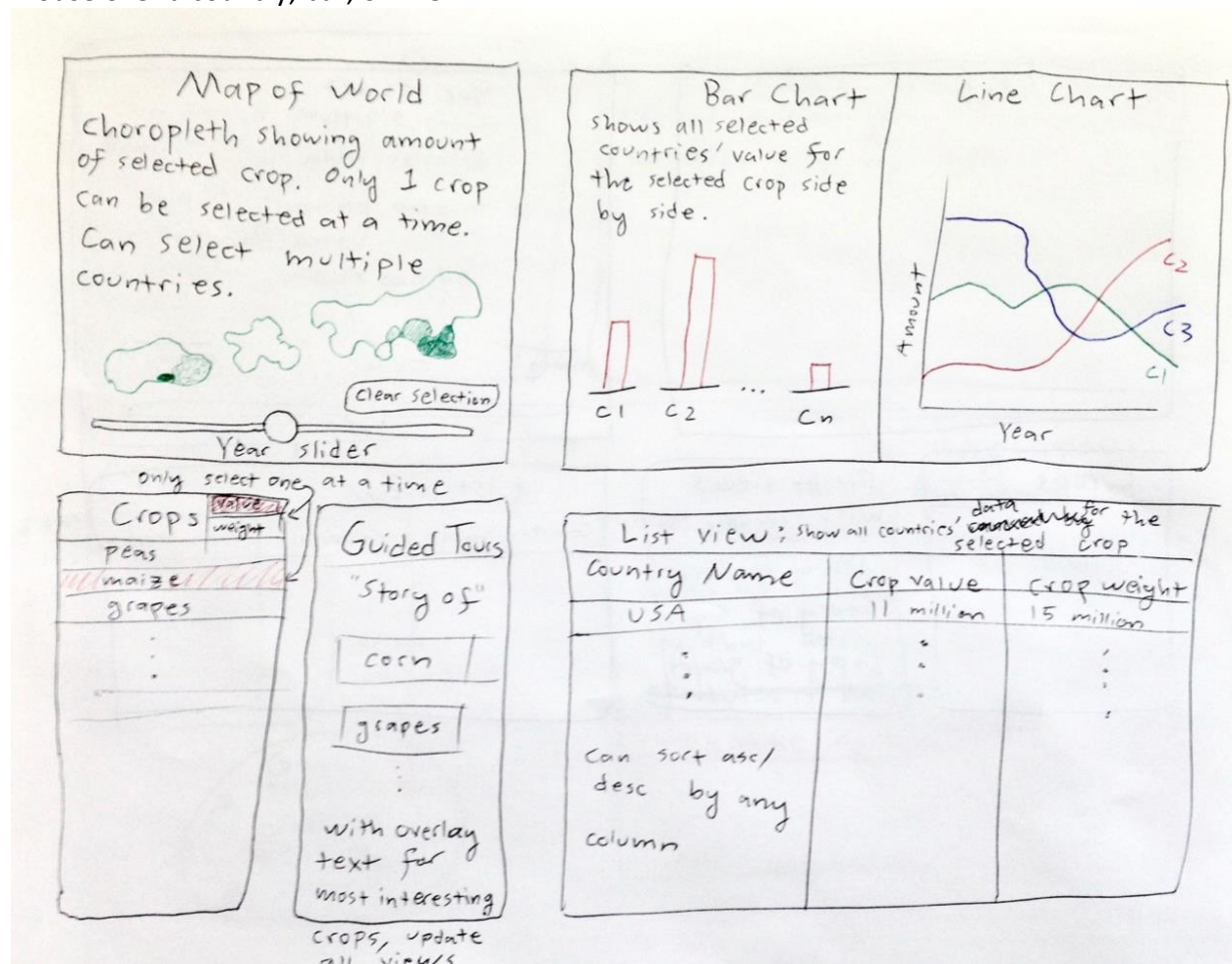
My final edits to this design, as I implemented it, were to color the borders of each selected country to match the color of the corresponding line and bar and to place the line chart under the bar chart rather than by its side. //(maybe)(I also removed the value/weight switch and the crop value column from the list view, since I used only production weight, not value.)

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

# World Crop Visualization

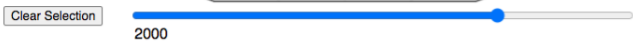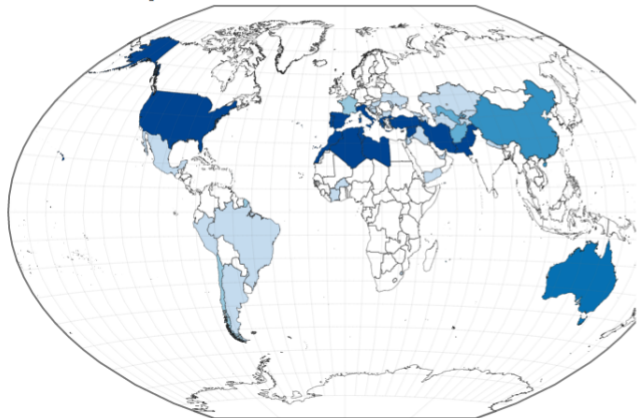Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693

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

## Guided Tours

# World Crop Visualization

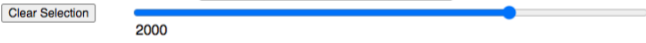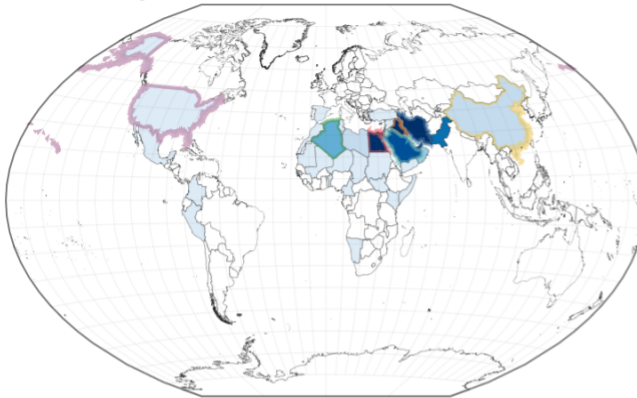Name: Christian Felt; E-Mail: u0809693@utah.edu; UID: u0809693

Clear Selection

2000

Agave fibres nes
Almonds, with shell
Anise, badian, fennel, coriander
Apples
Apricots
Areca nuts
Artichokes
Asparagus
Avocados
Bambara beans
Bananas
Barley
Bastfibres, other
Beans, dry
Beans, green
Berries nes
Blueberries
Brazil nuts, with shell
Broad beans, horse beans, dry
Buckwheat

## Guided Tours

| Country Name | Crop Weight | Crop Value |
|---|---|---|
| Australia | 500 | 700 |
| Belgium | 10 | 76 |

---

# World Crop Visualization

Units: 1e+6 tonnes

1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Iran   Iraq   Egypt   Saudi Arabia   Algeria   China   USA

Clear Selection

2000

Coconuts
Coffee, green
Coir
Cotton lint
Cottonseed
Cow peas, dry
Cranberries
Cucumbers and gherkins
Currants
Dates
Eggplants (aubergines)
Fibre Crops Primary
Fibre crops nes
Figs
Flax fibre and tow
Fonio
Fruit excl Melons,Total
Fruit, citrus nes
Fruit, fresh nes
Fruit, pome nes

## Guided Tours
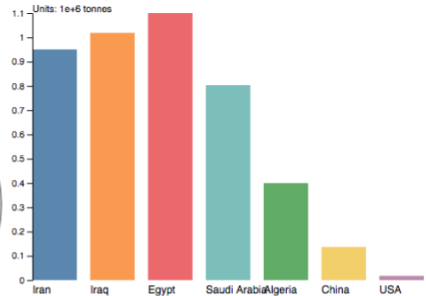
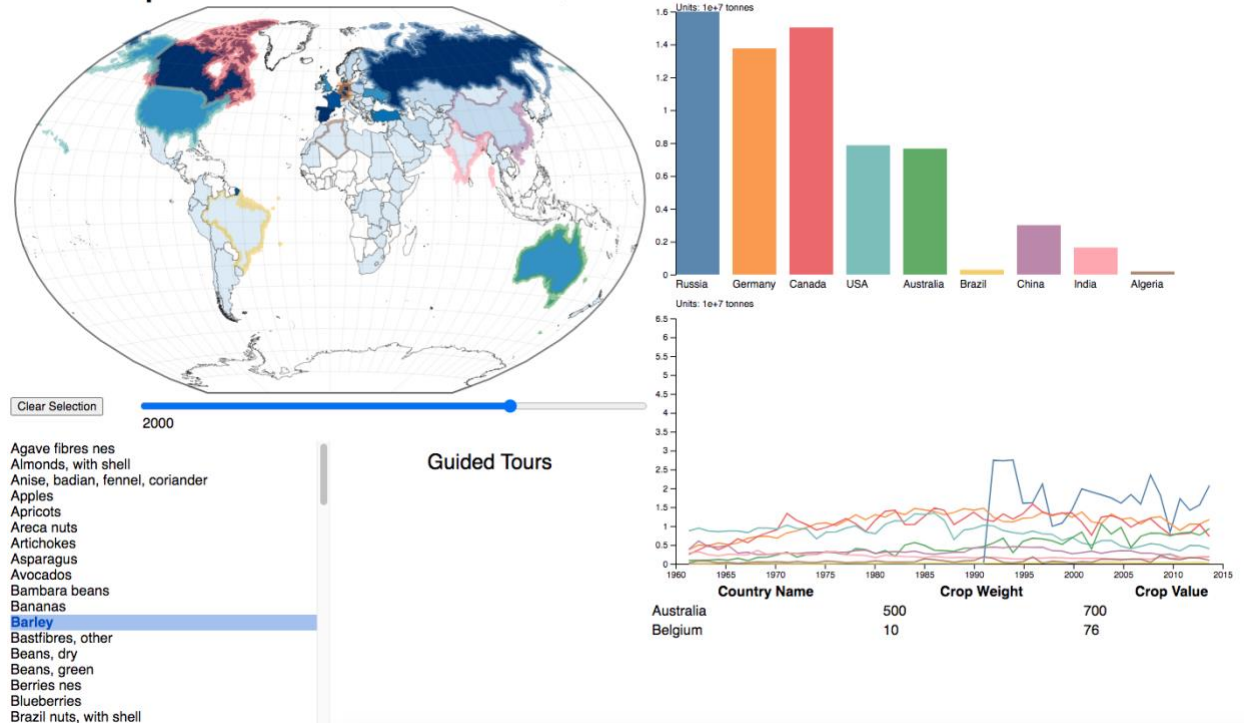| Country Name | Crop Weight | Crop Value |
|---|---|---|
| Australia | 500 | 700 |
| Belgium | 10 | 76 |

**World Crop Visualization**

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

Units: 1e+7 tonnes

Russia  Germany  Canada  USA  Australia  Brazil  China  India  Algeria

Units: 1e+7 tonnes

1960  1965  1970  1975  1980  1985  1990  1995  2000  2005  2010  2015

Clear Selection

2000

Agave fibres nes
Almonds, with shell
Anise, badian, fennel, coriander
Apples
Apricots
Areca nuts
Artichokes
Asparagus
Avocados
Bambara beans
Bananas
Barley
Bastfibres, other
Beans, dry
Beans, green
Berries nes
Blueberries
Brazil nuts, with shell

Guided Tours

| Country Name | Crop Weight | Crop Value |
|---|---|---|
| Australia | 500 | 700 |
| Belgium | 10 | 76 |

# Evaluation

This visualization is very 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). I learned some interesting facts, e.g. that Italy is the world's largest grower of artichokes, and India is the largest producer of coir and jute. 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. Cold-hardy grains like barley were grown mainly in countries with large chilly plains like Canada, Germany, USA, and Russia (but also Australia). 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 is by far the largest producer of apricots, which is the most delicious flavor of Turkish delight. Germany grows the most gooseberries and the United States, Canada, and Chile together grow almost all of the world's cranberries.

One weakness of this visualization is that it only shows crop production in terms of weight, not value or area harvested or other metrics. Still, I wasn't terribly interested in seeing these other metrics, and it would have added complexity both to the implementation and the user interface. The main weakness from my perspective as a user is that there is no little blurb and picture about each crop; this would have been cool, but one can always take the trouble of right-clicking on a word and looking it up 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.

# Milestone Stats

I put a ★ next to each item that has been completed or mostly completed for the milestone.

**Must-Have Features:**

★ Line chart showing multiple countries' production of a given crop over time.

★ Bar chart showing multiple countries' production of a given crop side by side.

★ Choropleth map showing all countries' production of a given crop at a given time and allowing the user to select multiple countries.

★ A year slider bar to change the year in bar chart and map views.

★ A button to clear the map selection.

★ A list of crops, allowing the user to change the selected crop.
A table, showing the country name and crop weight for the selected crop.
The columns of this table should be sortable in ascending or descending order.

**Optional Features:**

Tooltips for all lines, bars, and countries.
"Guided Tour" panel with some overlays explaining particularly interesting crops.
Buttons to automatically select top n producers of selected crop for display in line chart, bar chart, and choropleth.
Allow user to select country from the table view.
Use monetary value as well as weight of crops and have a switch, allowing the user to toggle between units of weight and monetary value.
Add other interesting features besides crops, e.g. import/export ratio or live animals that could show up in the line chart, bar chart, and choropleth in the same ways, and provide new columns in the table. Put these other features in a panel separate from the crops for conceptual clarity. Still only one feature can be selected at a time.
In addition to or instead of choropleth, try a grey map that shows crop production by size of overlaid circles.

**Project Schedule:**

★ Nov. 1: Have repo and workspace set up, gather data, write code to read and bind data.

★ Nov. 8: Implement choropleth map, crop list and selection.

★ Nov. 15: Implement bar chart, line chart, and year slider. Milestone due.
Nov. 22: Implement table view and guided tours.
Nov. 29: Implement tooltips and selection from table. Finish process book. Make screencast and website. Host website on GitHub and screencast on YouTube or Vimeo. Write README and finish/polish comments in code. Release project on GitHub.
Dec. 2: Wrap up submission. Project due.