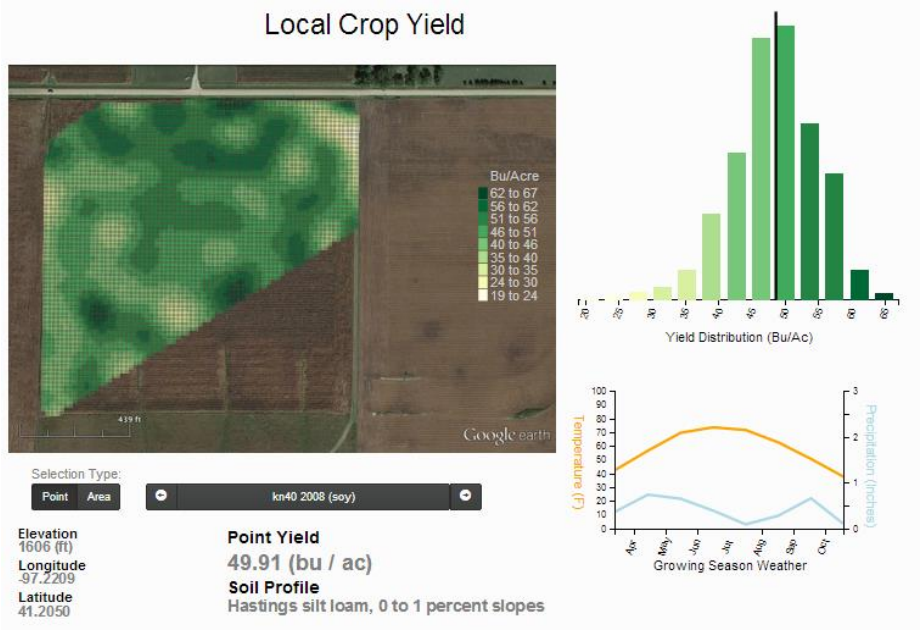
A Century of Corn

Harvesting U.S. Crop Yield Data

J. Benjamin Cook, Ryan King, Chad Hornbaker, Conor Myhrvold

Local Visualization Dashboard, Corn Field Yield, 2008 Growing Year



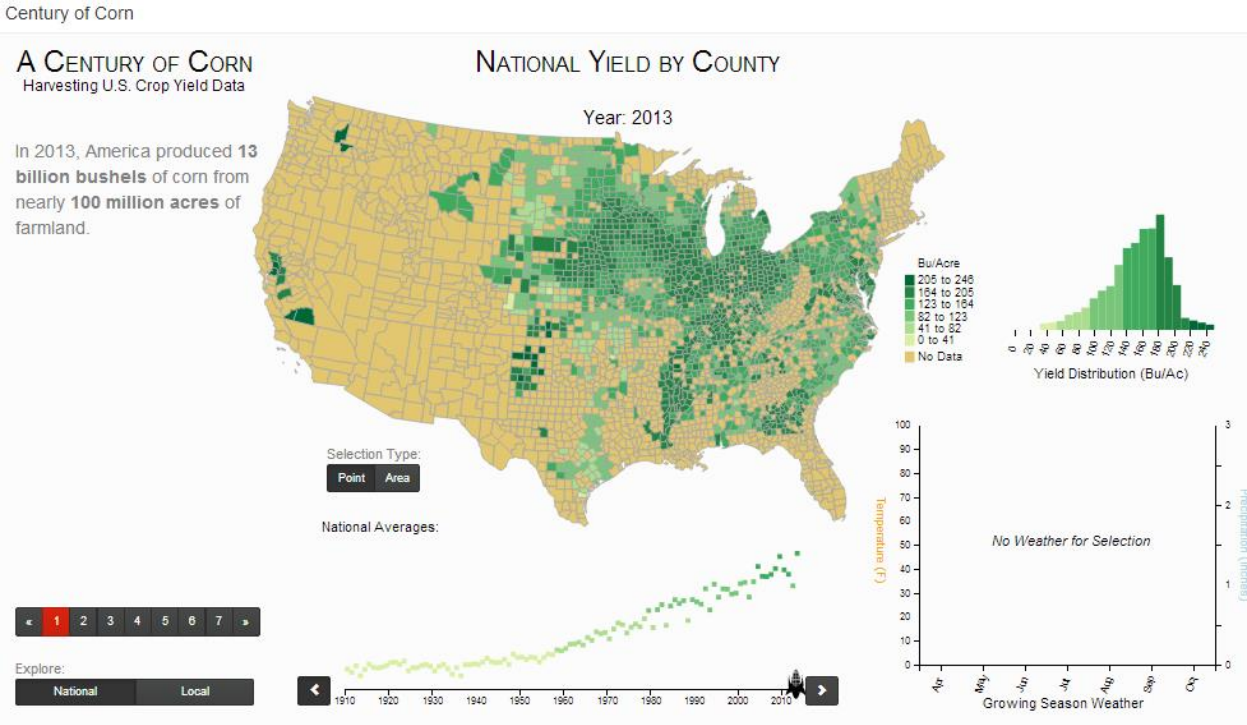
Index

Click on page number of section to go to it.

Overview and Motivation

In American, corn is king. We produced over 13 billion bushels of corn in 2013, across almost 100 million acres of growing area. Corn growing is now an everyone-affair; taxpayers helped contribute to this record-breaking haul, to the tune of several billion dollars a year[1]. How did we get here?

The setup to this story begins over a century ago and our project, BitFarm, shows how. BitFarm is an interactive d3.js data visualization tool that merges information from a local farm with county, state, and national corn yield trends from 1910-2013. With access to over a century of crop yield data, we are able to visualize the changing face of farming in America. What does this face look like?



[1] $80b+ from 1995-2012.

We're familiar with Big Pharma and Biotech, immortalized in franchises such as Resident Evil's Raccoon City, but a lesser known face is Big Farm. Technology, climate, and politics have a significant effect on farming efficiency.

	Harvest Area	*Sales*
2011 US Corn Harvest	84 million acres	$63.9 billion
2013 US Corn Harvest	97.3 million acres	

Source: http://www.epa.gov/oecaagct/ag101/cropmajor.html , http://www.nass.usda.gov/Newsroom/2013/03_28_2013.asp

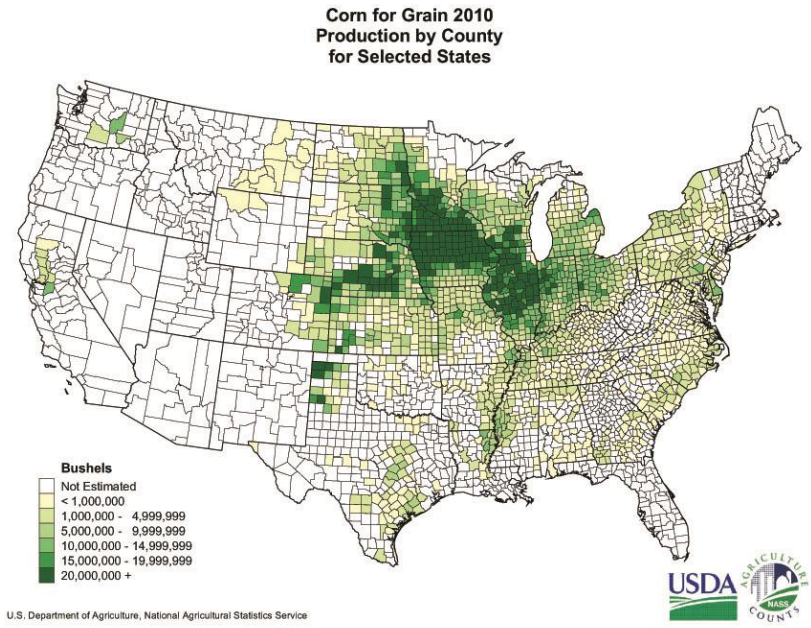
Corn is our unofficial national crop. If you had to pick one cash crop, corn would be it:

> *The United States is, by far, the largest producer of corn in the world, producing 32 percent of the world's corn crop in the early 2010s. Corn is grown on over 400,000 U.S. farms. The U.S. exports about 20 percent of the U.S. farmer's corn production.*
>
> *Additionally, corn farming has become exponentially more efficient. If U.S. farmers in 1931 wanted to equivalently yield the same amount of corn as farmers in 2008, the 1931 farmers would need an additional 490 million acres! – United States Environmental Protection Agency*

While the meteorological climate has been variable throughout yield history, propping up the corn kingdom is the reliable political climate. The US government doesn't just say it with words -- they say it with dollars. Corn is easily the most subsidized crop in the United States, and consumers support it: corn is found in almost every aisle of the supermarket. Therefore, the story of farming the United States is in many respects the story of corn.

Related Work

For all of the talk about the changing face of agriculture and US land use, there have been few attempts to show a concomitant picture of what this change looked like in the 20th century, and now into the 21st. There are other factors at hand which are relevant to the farm. Droughts and bad weather can take the cash out of cash crops. Yet favorable farm policies, fertilizers and new technology farmers can increase yield. What has been the end result of this ongoing challenge in the 20th century?



National corn yield by county. We used the same data and color scheme, but took this static visualization several steps further by making it selectively interactive. We also combined this map with more detailed information about what the yield from a recent individual farm in Nebraska at the height and heat of the Corn Belt, looks like.

We dug deeper and gathered many disparate, but related datasets, related to corn yield, America's biggest cash crop for many decades. Our displays of corn production use two main visualizations to show the big picture, the local picture, and everything in between (through interactive selection).

Questions

We were motivated by the general question: what affects a farm over space and time in the United States? More specifically, we aimed to answer, how do different farm influences fit together to show the big picture? As we made our visualizations, we were able to specify our questions. Our visualization now directly answers:

1) What does the 20th century US corn crop yield change look like?
2) What do key climate indicators affecting the yield of a local farm look like over that same time period?
3) How does the above compare to the context of one county or a larger region of counties (up to national)?

Data

Here we list the datasets used in our visualizations, where they can be obtained, and how they are used:

1 – National Weather Data, United States Historical Climatology Network

Daily precipitation, and min/max temperatures affect yield. We show the relationship between climate (aggregated weather trends) and yield in the County Level visualization. After pulling down weather data from the last century we cleaned and transformed the data with a Map/Reduce job to obtain a manageable clean sample that we can visualize.

2 – County Yield Data, USDA National Agricultural Statistics Service (NASS)

Shows yield by county from 1910-2013. Together, it's a national picture. NASS maintains relatively updated and clean survey data from farmers around the country. To use this data we simply downloaded CSV files and removed irrelevant columns with a simple Python script.

3 – Local Yield Data came from correspondence with a Nebraska farmer.

From combines (geo-located data, sampled 6 times/sec). Proprietary to others, but we can use publically. An example of a farmer's individual yield. The county and national data place this small plot into context. Local yield data is summarized by fitting a Gaussian Process to the data so we can show the spatial variability in local yield.

4 – Soil Type Data, UC Davis Soil Resource Laboratory

Affects yield and type of crop suitable to be grown (although we focus on corn). Soil data is downloaded through the UC Davis soil lab API, which allows us to access a high dimensional overview of soil by longitude and latitude.

5 – Grain Price Data, USDA Economic Research Service (ERS)

Affects the crop value; how much a farmer can get from their land. We summarize this with monthly averages of price in dollars per bushel. We use this data to show how valuable a farmer's field is at any given time.

IPython Notebooks used to clean the data:

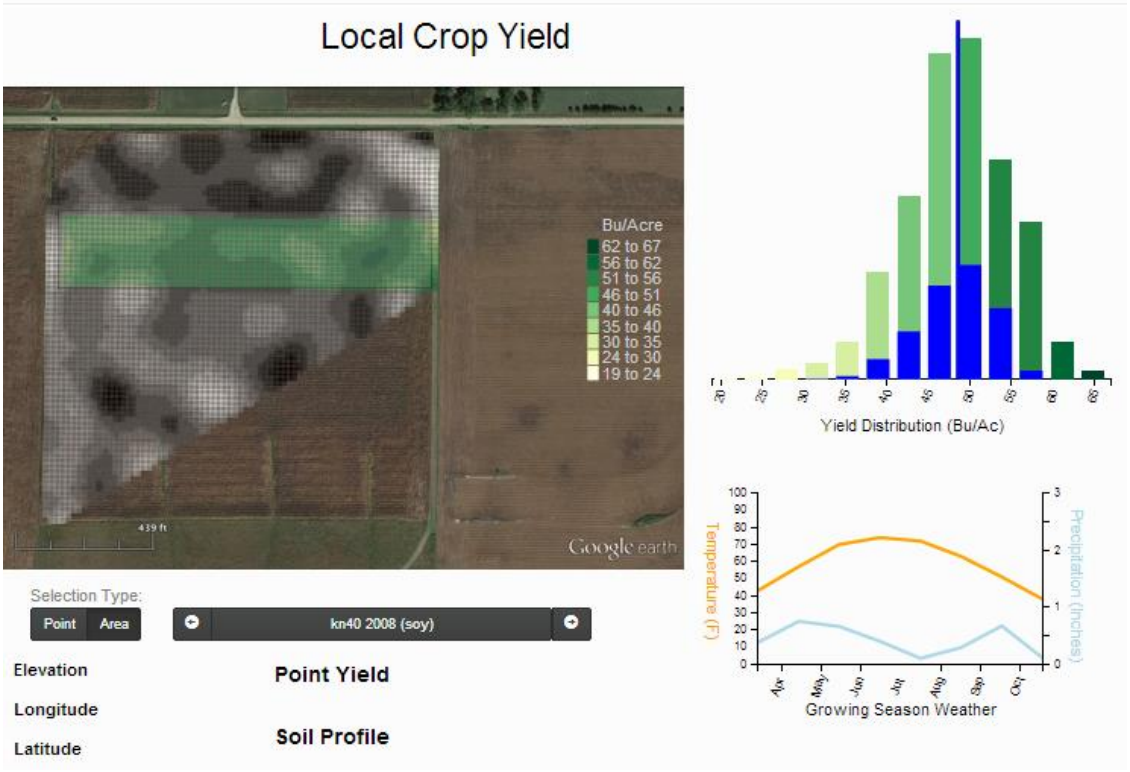
Weather Data IPython Notebook

County Data IPython Notebook

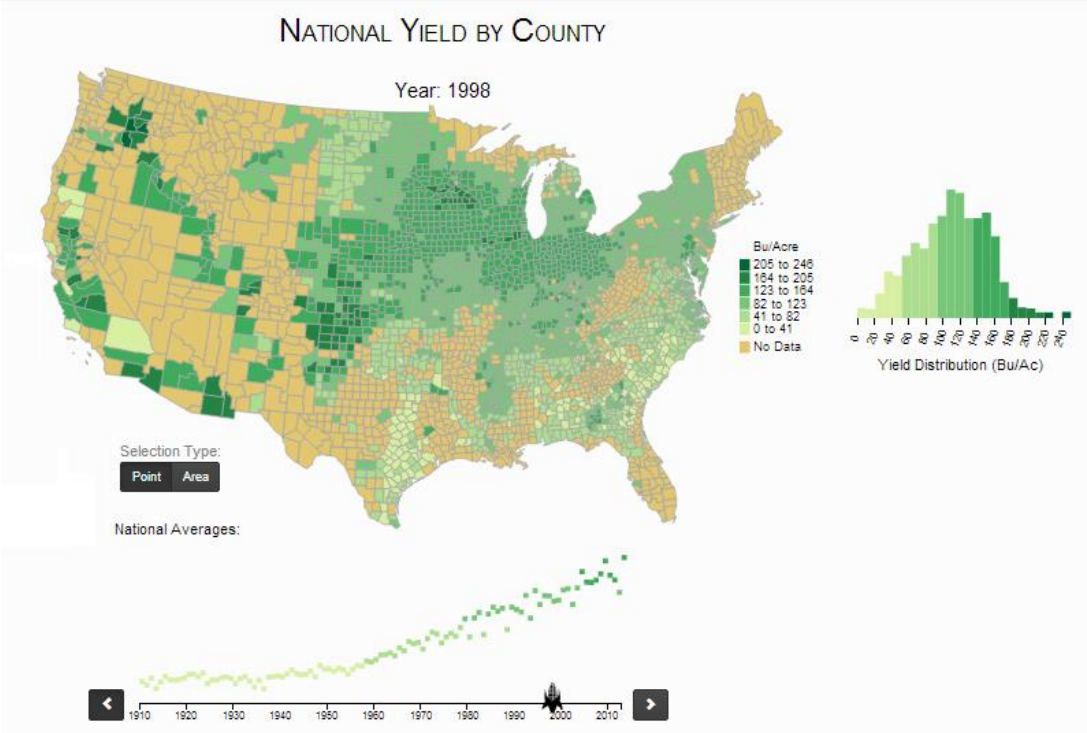
Exploratory Data Analysis:

We started with separate visualizations of local yield, county-level yield and climate data. We also did some basic statistics to aggregate and summarize the data. For example, we trained a Gaussian Process on the local-level yield to show the patterns of high yield in a spatially smooth way. Because the climate data are several gigabytes in raw form, we also aggregated temperature by plotting average monthly values. We learned several important things from these exploratory efforts that informed our final visualization.

There is a major technology trend, and the corn yield has increased over time overall, with some variance for any individual area (i.e., some have decreased, but the concomitant picture is one of effusive efficiency). There is a high degree of spatial variation in corn yield. Yield tends to be highest in the Midwest United States; however, corn is still throughout the US.



An example of local variation of corn yield on a field (on the histogram blue is the selected area; the remainder of the regular histogram colors on each bin represents the grayed out area in aggregate.)

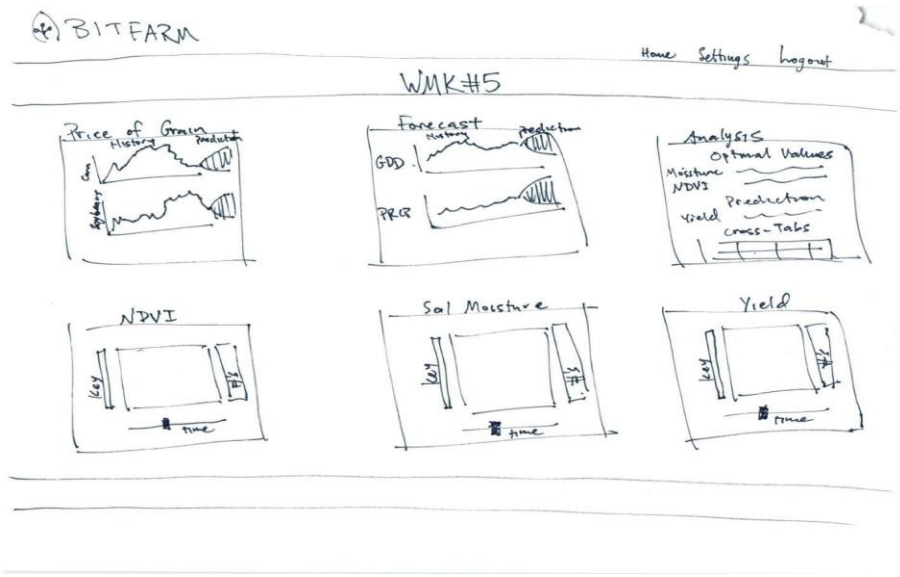


An example of county variation of corn yield on a field in 1998, when a big cumulative disparity between the lowest yields (< 40 bu/acre) and highest yields (> 180 bu/acre) developed across the country.

Design Evolution

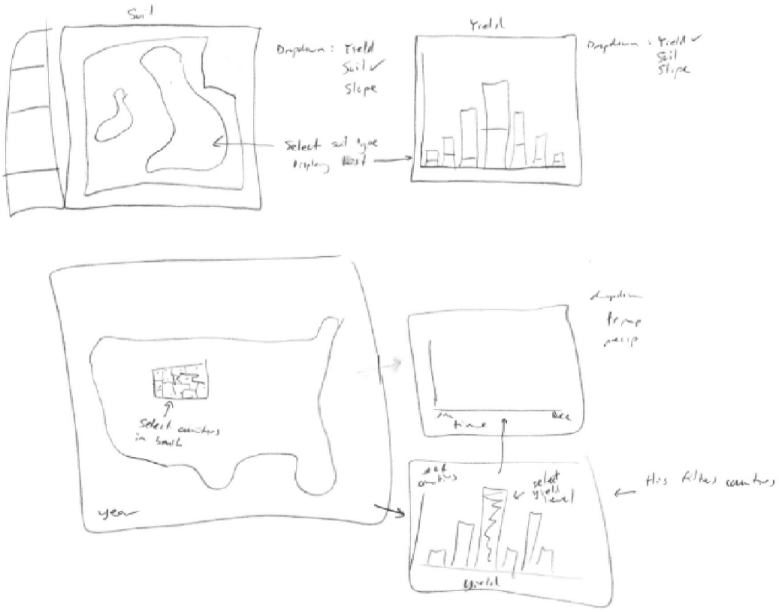
Our initial design concept was for a dashboard with several floatable visualizations, each showing a metric affecting farming success, the crop yield. The idea was that the layout could be customized; each visualization would be expandable, draggable and snapped into place, much like an app interface on a Windows Phone or an app dashboard for a PC.

First Concept Sketches



As we began to better understand the data and explore possible implementations, we gradually began to move away from the dashboard concept and focus on the visual components that communicated the best story for our data. The dashboard concept may be useful as a future implementation in an app, but we did not see it as the best way to communicate our intended story and allow a user to intuitively discover and explore the data.

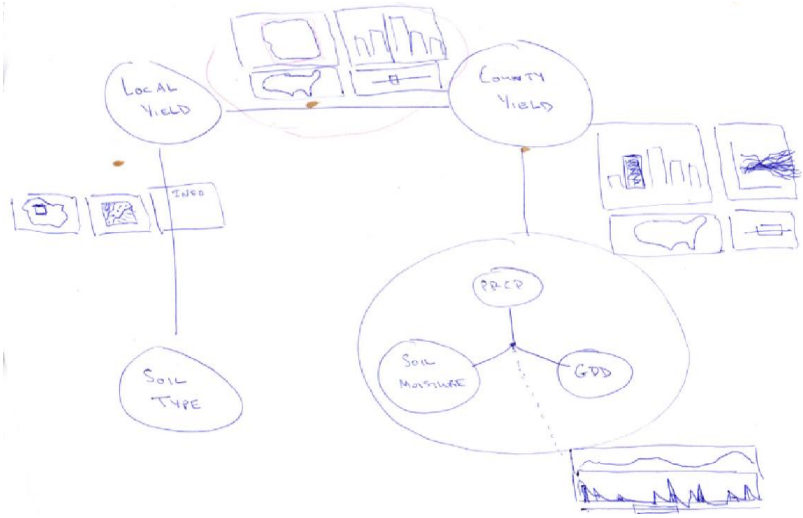
More Early Stage Sketches



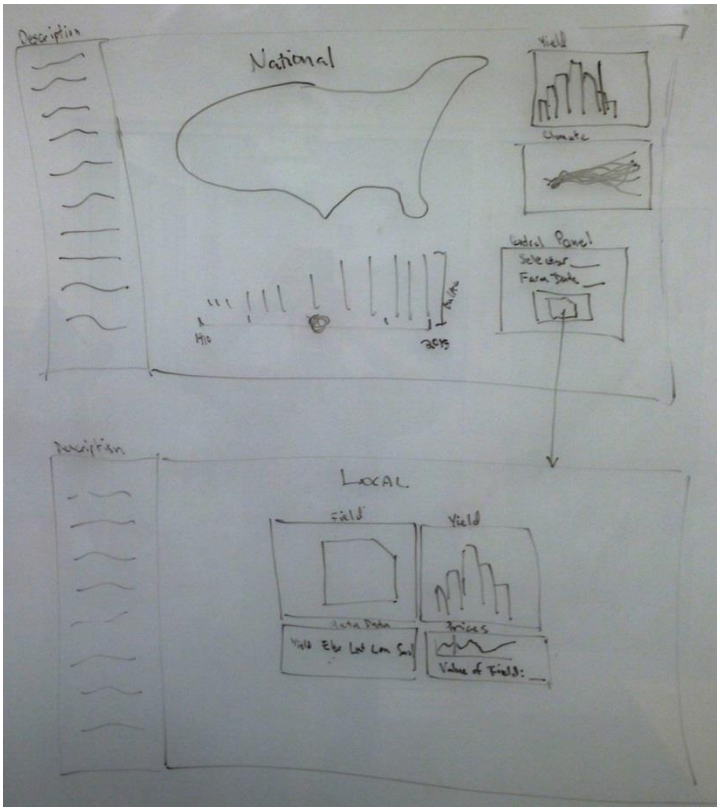
The sketch above shows our initial concept for the simplified interface. Rather than multiple disparate views, we focus on a single view at the national level *(bottom)*, and a view of a single farm *(top)*. Nearly all of our available data can be displayed in these two simple views. A user can look at the yield data at specific points in a field and see where the field is producing the most, as well as compare the yield to the current and future market price. At the national level, the user can view other counties around the country with similar output, as well as seeing the historical trends in yield in the context of weather and location.

Linking the data between the micro and macro level was a difficult conceptual task. The sketch below depicts the information flow we believed would work well for addressing our research questions. At the local level, we can see which parts of the field have a certain soil type. We thought this may be an interesting link to the point yield data. At the field level, we want to not only see the yield patterns, but have a way to compare it to how well other counties across the country are doing. This broadens our view to the county and national level, where we can see trends in weather. The addition of historical data provided a more compelling view of how well modern American farms are doing in comparison to only a few decades ago.

Second Stage Sketches



Milestone Design Sketches

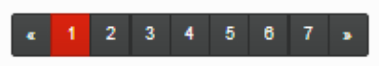


We narrowed down our ideas further to detail the specific information we want to see in the two primary views. The National view (above) shows the historical trends and has the option to view

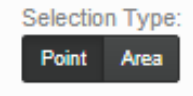
any of the individual farms in the dataset (still very limited at this time). The Local view implements our desire for a way to explore a farm's data at a very specific level.

Final Design Product

The transition and cross-information between these views was further refined in our final product, incorporated in the playback of key transitions in corn yield throughout the 20th century:

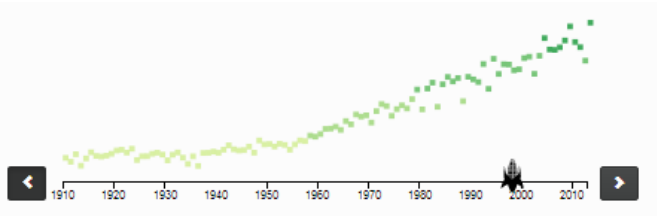


The selection by point (county) or by area (defining a rectangle or group of counties):



The selection by local or national:

The selection by year:

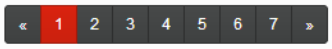


Double click on the arrows to either end of the years, and the corn slider creates an animation that transitions through all years.

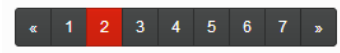
Obviously, the final implementation is free for perusal so we have kept this section shorter, since its existence is self-evident as the finished product of our project.

Story Line

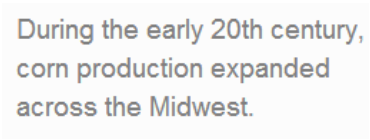
As high fructose corn syrup, corn products are in almost every aisle of the supermarket. We show how this happened using an animated story line, revealing the massive extent of corn yield in the mid-20th century which made the US the #1 producers of corn in the world.

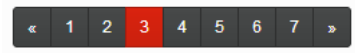


2013 – start at present day (the last year for which growing season information is available – note that weather data is not available yet)

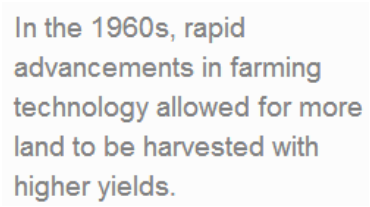


1910 – 1959 Midwest expansion (and expansion of collected data, which is also some inference to where corn was popular to grow)





1960 – 1979 The Green Revolution ("Get Big or Get Out" US farm policy encourages expansion of farming, along with technology to increase efficiency: more corn over more space)



1980 – 1999 Agribusiness Aggregation (Consolidation of corn, and leveling off and decline of total area farmed; yet production continues to grow thanks to efficiency)

In the 1990s, the number of corn producing counties decreases while from an efficiency standpoint, corn yields reached higher still.

This is the rise of smaller numbers of bigger farms: the aggregation of agribusiness.

2000-2013 Re-regionalization to Midwest (Farming returns at the start of the 21st Century to more or less where it was toward the start of the 20th Century – the Midwest region, albeit with more land farmed in that area, and much higher yields)

« 1 2 3 4 5 6 7 »

Since 2000, the Southwest has experienced several periods of long-term droughts. Corn became less economically viable to grow in regions where it was marginally successful before.

However, farms in the Corn Belt states continue to produce higher yields.

Present-day: Local View (How farmers can see their land, and local yield variation on a farm)

« 1 2 3 4 5 6 7 »

Today's growers have access to unprecedented amounts of data related to their fields.

Yield at the individual field level is driven by factors such as soil type, climate, crop rotation, and irrigation.

Return to National View 2013 (with 2013 EPA quotation from 2013 USDA data)

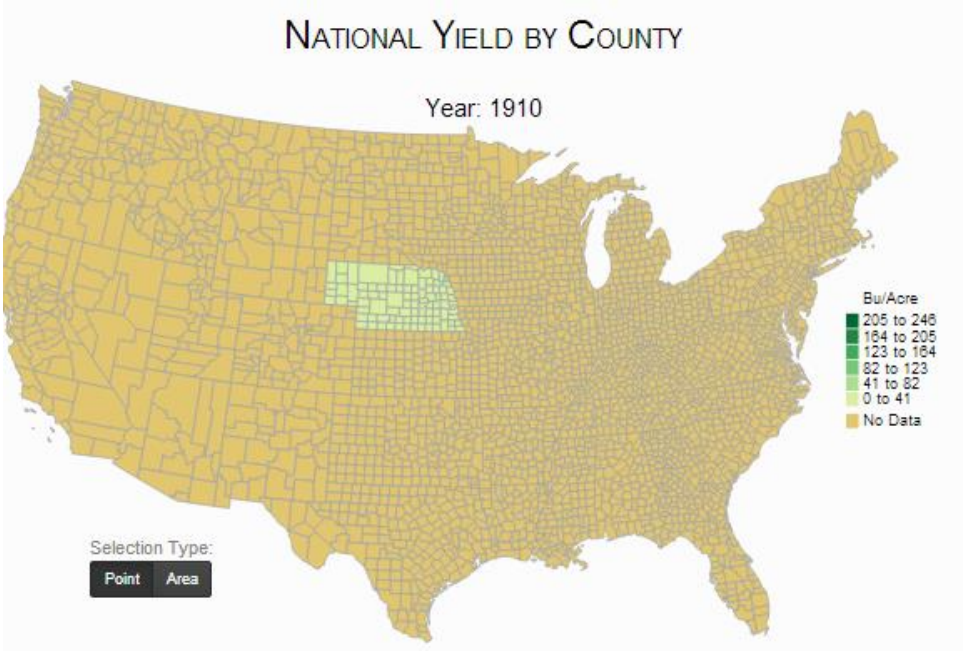
« 1 2 3 4 5 6 7 »

Implementation

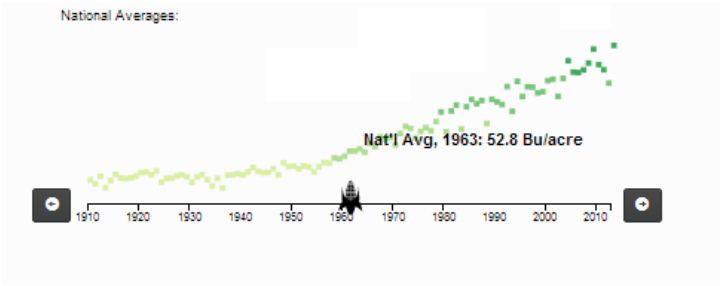
Our layout consists of a primary visualization of national crop yield of corn for the past 100+ years (since 1910, although in early years data unavailability is high) and a secondary visualization of local field level data (for several fields, for several single years of corn along with soybean examples to illustrate the cross-crop potential of our visualization approach.)

National Yield View

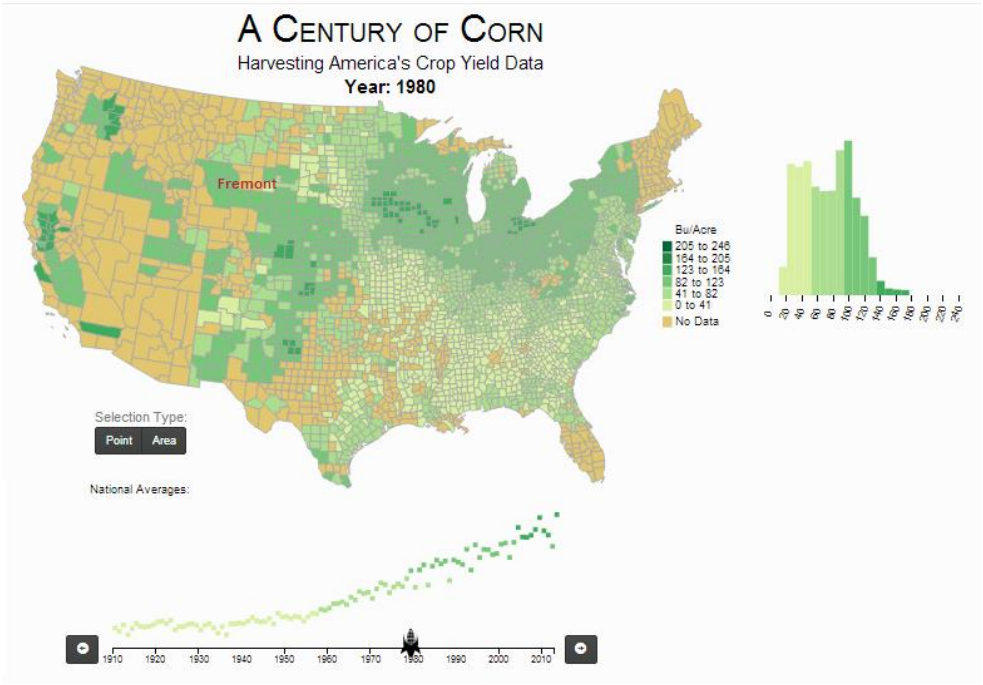
The National Yield View starts by loading corn yield data for the earliest year in the dataset (currently 1910). We then generate a choropleth map using the county yield data. For comparative power, its color range is scaled to the minimum and maximum yield across the entire dataset. Darker green shades correspond to higher yields (see later years than the 1910s for this on the map). We chose green because of the intuitive association with plant growth. We used a light brown color to shade in any county without yield data, analogous to an un-farmed field's soil.



We also generate a scatter plot with the average yield of all counties per year:



This axis of the scatter plot also serves as range slider that allows for interactive selection of which year's data is displayed on the choropleth. For aesthetic reasons, the axis ticks were placed at ten year intervals. To keep track of what year has been selected, a tooltip below the title logs the currently selected year. Highlight your mouse over a county, and the county name is displayed in red:



In addition to the aforementioned aspects of the visualization, there are several additional features that are designed to enhance the user's ability to glean trends about crop yields.

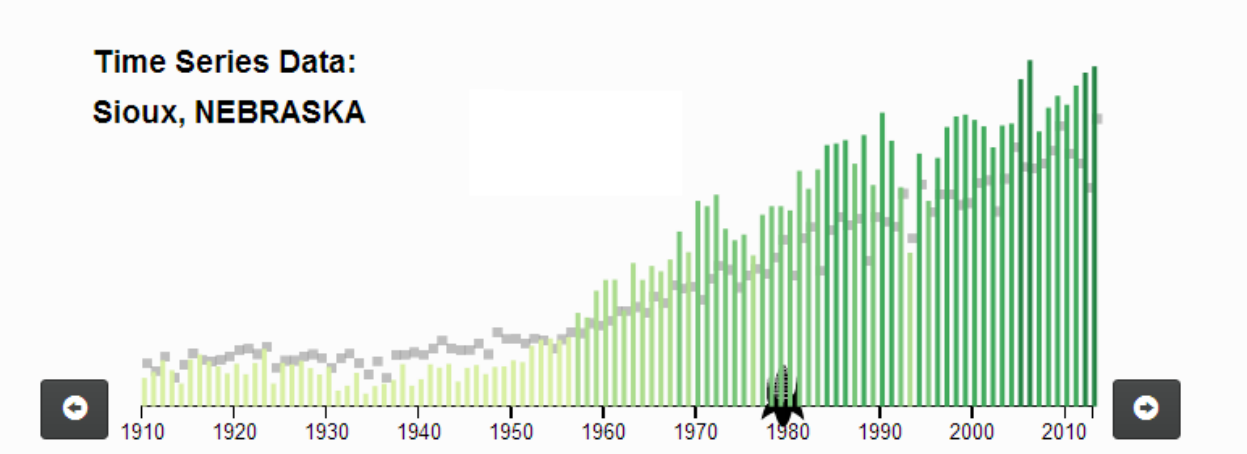
Toggle Selection Mode

There are two selection methods: by point (pixel) and by area (brushing by creating a rectangular area over which to examine).

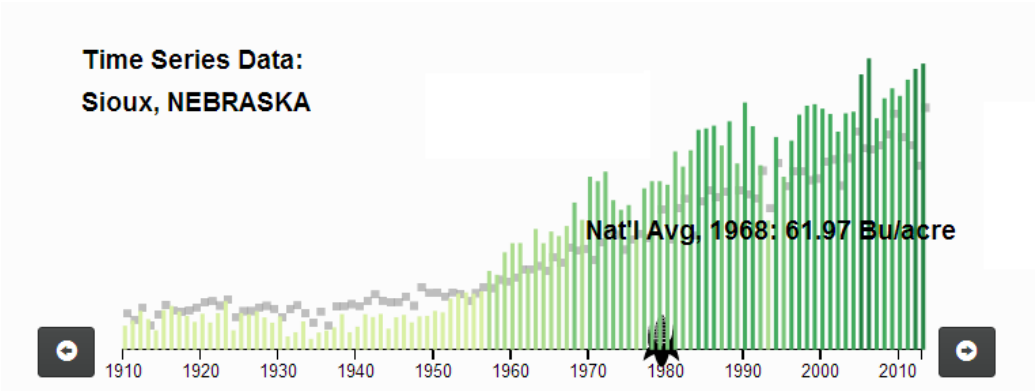
1) Point Selection:

When the user mouse hovers over any county, the name is displayed (see above image).

Additionally, the user can click on any county to display historic yield data of that county. This county data is plotted against the national average data for comparative purposes. To simplify the graphic, the national averages are greyed:

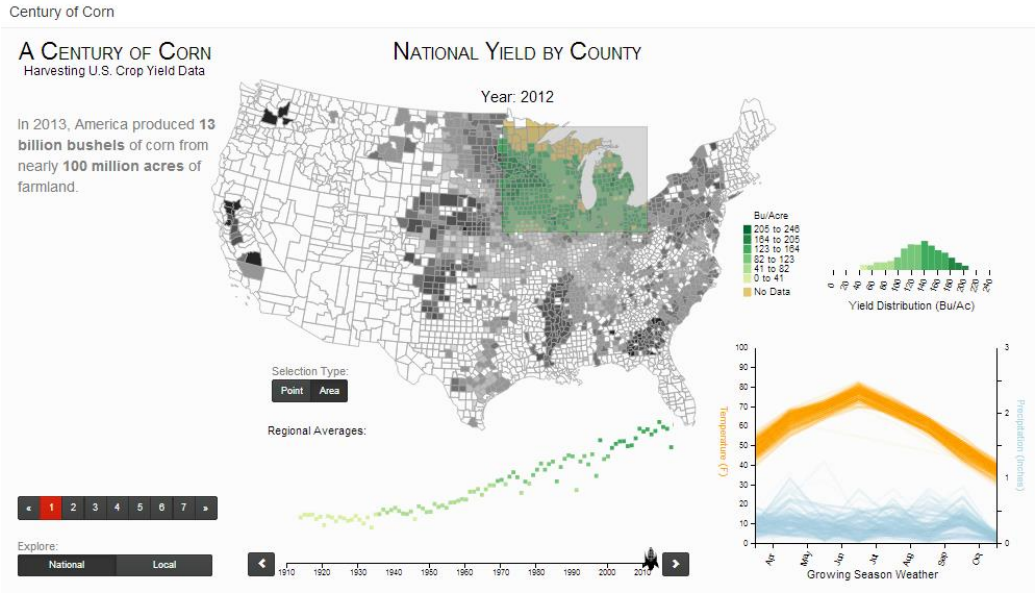


To see the actual value at any year, the user can simply hover over the appropriate bar. Additionally, the user can hover over national average scatter plot value to generate a tooltip that will display the numerical average. This feature allows the user to quantify exactly how well any county is doing compared to the national average:



2) Selection by Area (Two Dimensional Brushing):

This second selection method allows for the user to choose a group of counties to focus on. When selected, the brushed counties maintain their original color scheme while the rest of map is gray scaled.

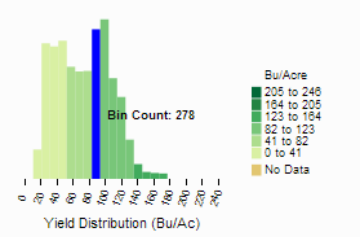


When a region is selected, the data displayed is truncated to only include the select counties. This allows a user to obtain better insight to regional trends as opposed to national averages.

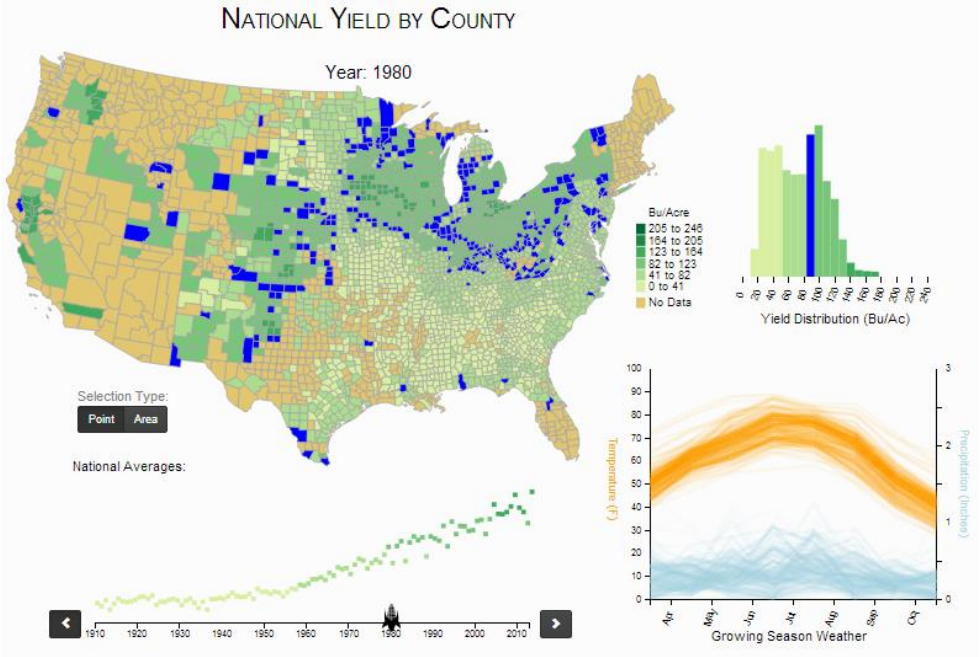
You can select by area, and then run the story line visualization to get a dynamic change for just that area, instead of nationally.

Side Panel Histogram

Yields for select counties are used to create a histogram of the distribution. A user can click on the histogram bars to select all counties with that yield range:



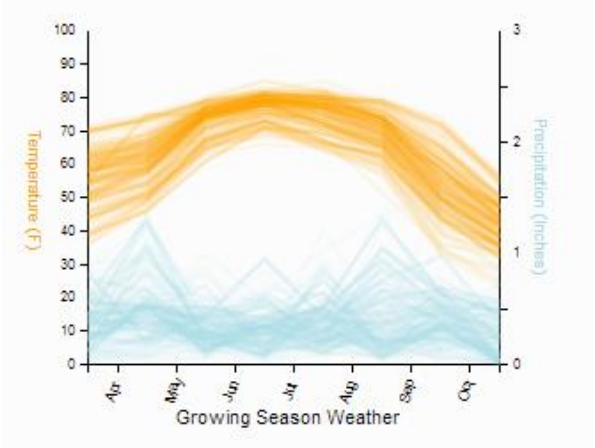
Bin there, done that! Selecting a bin from the 1980 corn yields is simple.



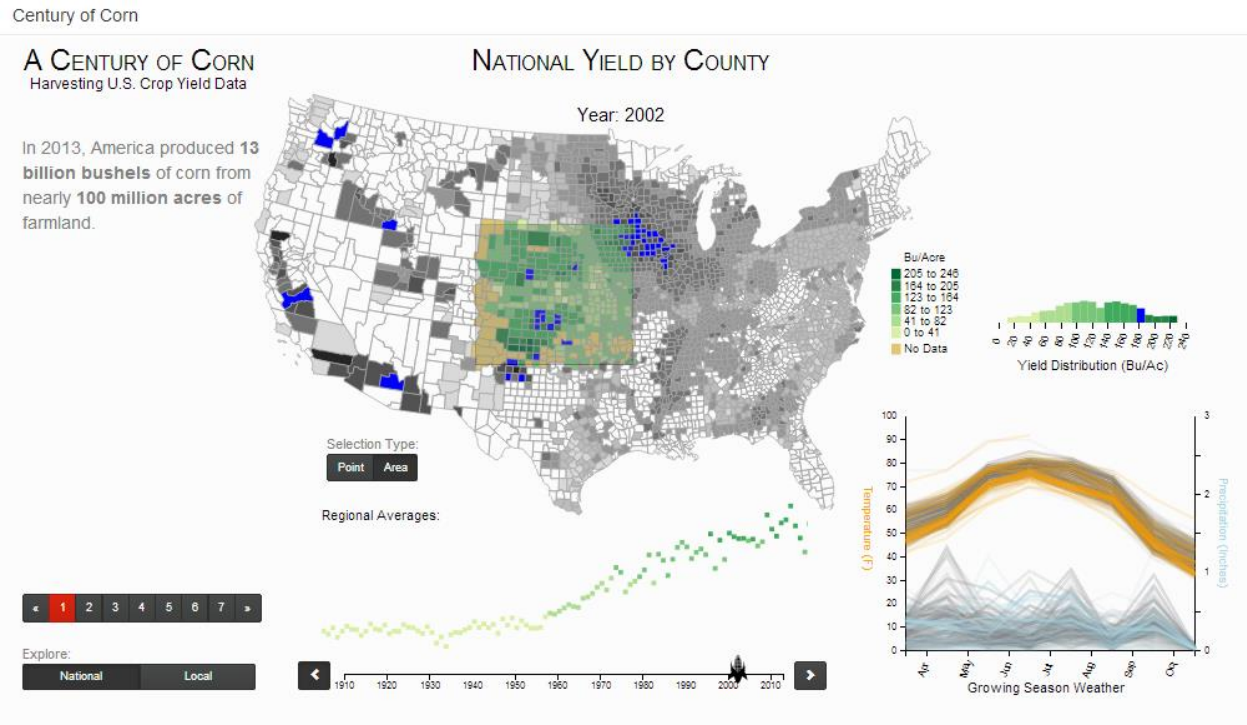
The idea of this bin selection feature is identify regions which had similar yields to hopefully discover trends about what generated those similar yields.

Side Panel Weather Data

Shows temperature and precipitation data for national or selected regions. Each path element represents a specific county:



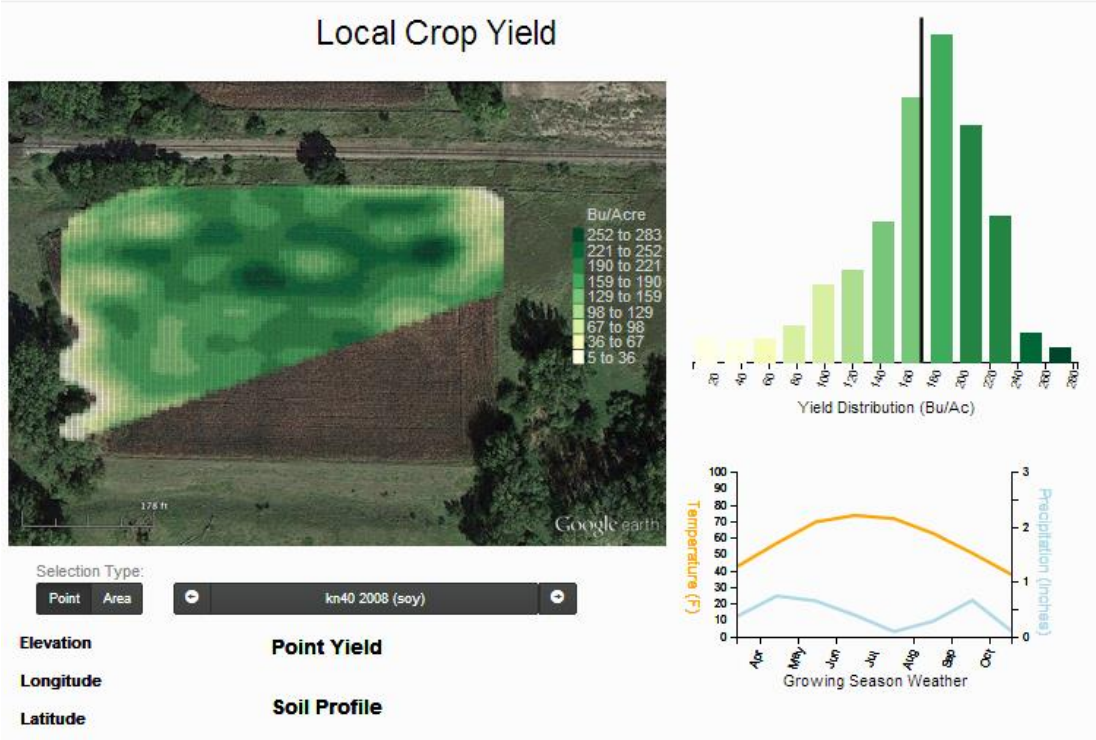
Whether or not the weather cooperated can make a good or bad corn crop year.



After selecting a subset region (by Area), you can subselect a bin in the histogram of that area, and highlights those specific counties' weather. (The counties that fall under the selected area.)

Local Field View

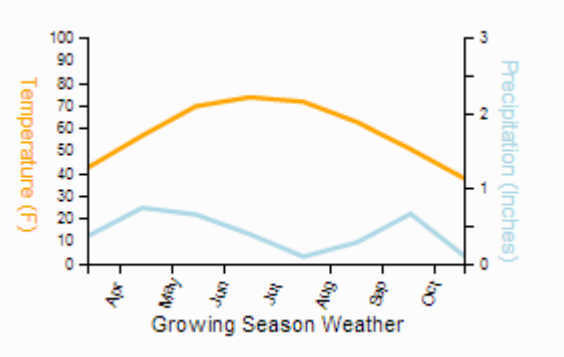
The goal of the Local Field View is to show an in-depth look at the (a)maize(ing) yield on a farm in Nebraska. The Local Field view starts by loading a heat map style view of the field (inferred via a Gaussian Process), a histogram of the yield distribution in the field for that crop (corn, in our case) over a time range (currently, in months).



Field View: The field is shaded by bushels per acre. The user can brush over any rectangular portion of the field to view the distribution of yield for that portion in relation to the overall field distribution. The selected area on the field relates to the orange bars on the histogram at the right.

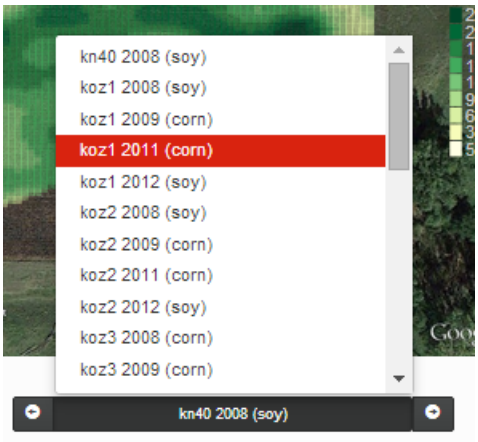
Yield Histogram: This contains both a static view of the crop distribution and the relative distribution of any selected areas of the field. The user can also brush the histogram to select a range of yield values, which will color the field map accordingly.

Local Weather Panel: Details the temperature and precipitation for the growing season in that county, for that year:

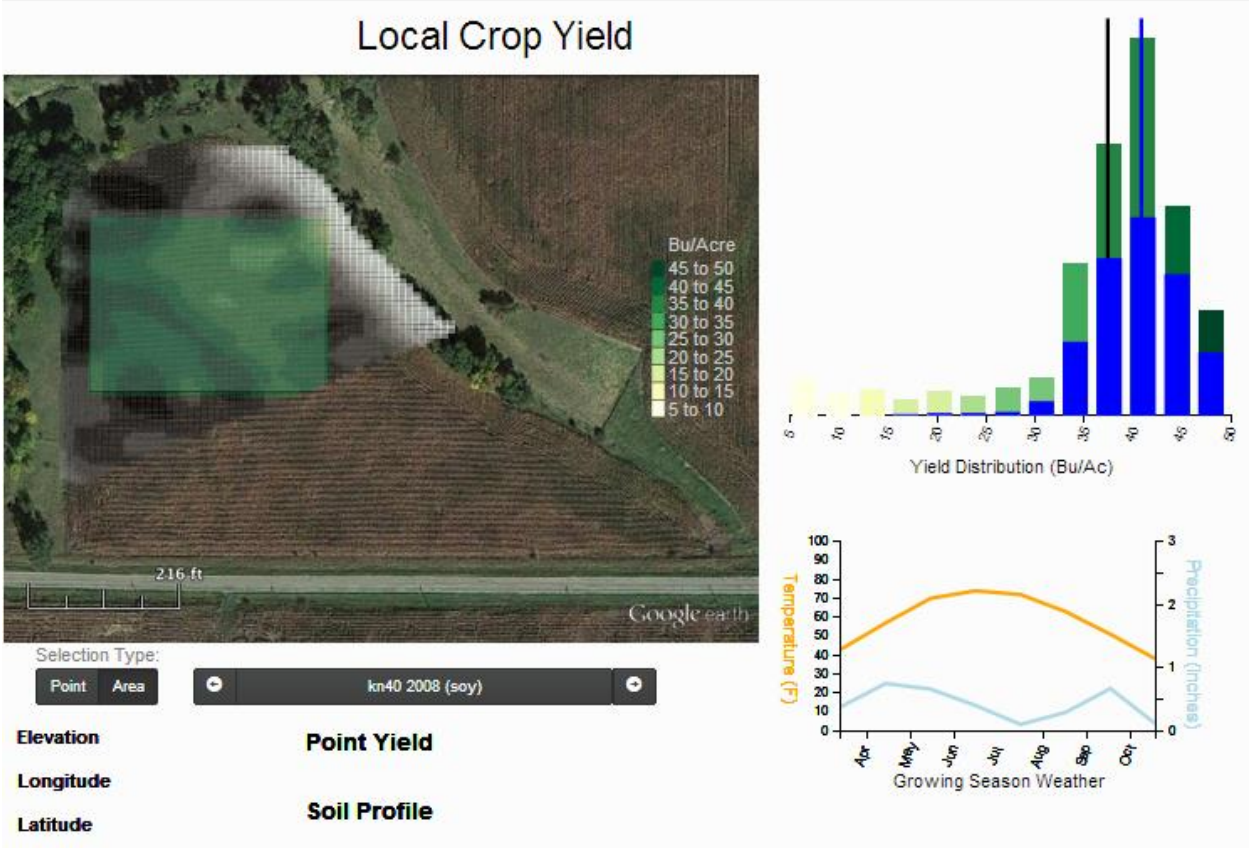


Information pane: This contains basic metadata for a point on the field as the user hovers over it with the mouse.

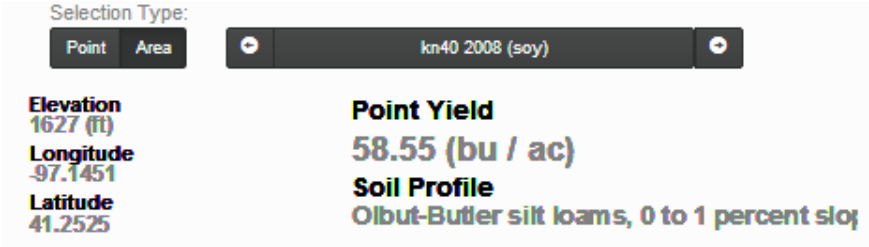
Additional fields: There are many fields available in several recent growing seasons for corn and soybean:



As previously mentioned, the area selection shows the portion of the yield distribution of the given field a user highlights:



In the Point selection mode, as a user hovers their mouse over the field pixels, information about that pixel (soil type, yield, coordinates, and elevation) appears below:



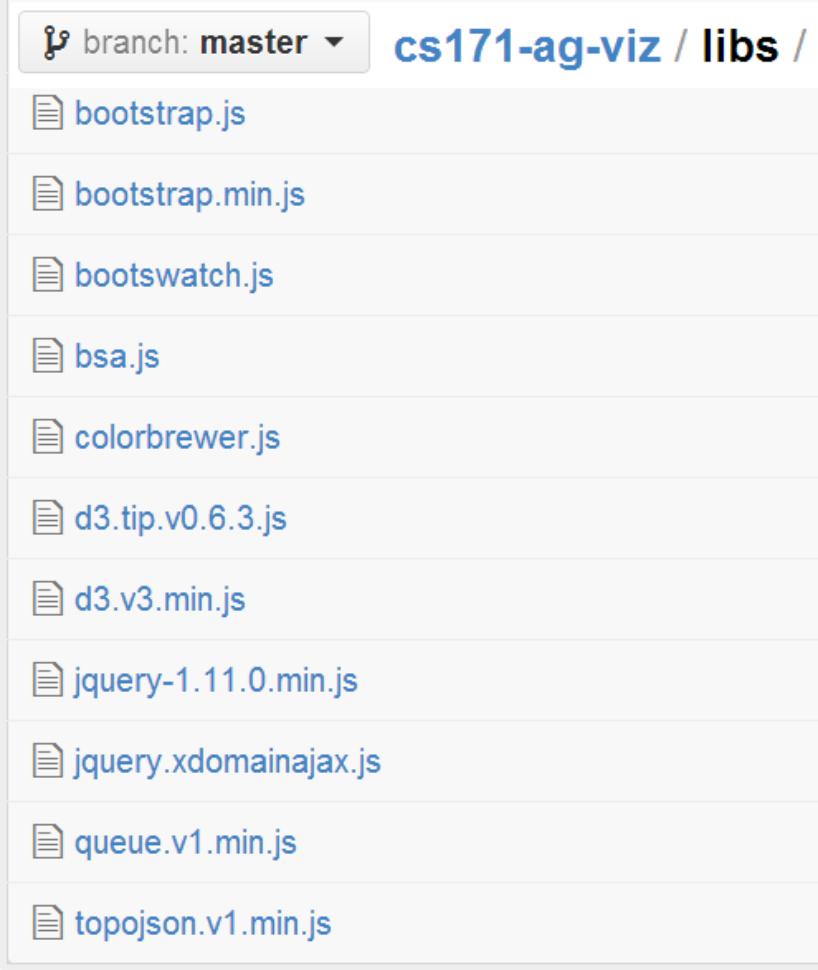
Evaluation

There were several key trends we noticed as part of the changing story of corn crop yields over time. First, if you look at the corn crop yield over time, you'll see the geographic area dramatically change throughout the 20th century.

- The time period this occurred in, from the 1950s to the 1970s, corresponds with the rise of big agriculture and the Green Revolution, with more intensive farming techniques.
- The 1970s policy of "Get Big or Get Out".
- In the 1990s on, the contraction of growing area to support an ever growing population with ever growing yield. Note the remaining corn crop areas in 2013 all have high yields relative to the typical yield just several decades prior.
- From 2000 to 2013, the Southwest of the US experienced record droughts. From our visualization, we see a noticeable decline in crop yield in the Southwest over this same time period, which warrants further scrutiny, and is exactly the type of observation we have in mind to detect with this visualization.

Libraries

We used the following JavaScript libraries in constructing our visualizations:



Additional Resources

EWG Farm Subsidies

Historical Track Record – Crop Production

King Corn (documentary), 2007

Major Crops Grown in the United States

The Insanity of Our Food Policy

USDA Forecasts Record-High Corn Production in 2013

US Farm Subsidies in 2005, Wikipedia

US Farms: Numbers, Size, and Ownership

Visualize GPS Agronomic Data with Maps to Make Data Easier to Understand

Visualizing Historical Agricultural Data: The Current State of the Art

Visualization of a Crop Season: The Integration of Remotely Sensed Data and Survey Data

Weather Anomalies, Crop Yields, and Migration in the US Corn Belt

Weather Effects on Expected Corn and Soybean Yields