



# CS 171: Final Project Process Book

Eric Slater, Luciano Arango, Monarch Wadia



# Overview and Scope

- Design and implement a web-based interactive visualization that allows you to answer questions about a topic
  - Choose the topic to answer questions about
  - Acquire data
  - Design visualization
  - Implement using modern frameworks
  - Evaluate the results



# Project Team

- Eric Slater
  - Harvard Class of 2016
  - A.B. Degree Candidate in Statistics
  - Mather House
  - First Year Programmer (has taken CS50)
- Luciano Arango
  - Harvard Class of 2016
  - A.B. Degree Candidate in Computer Science
  - Adams House
  - Second Year Programmer (has taken CS50, CS61)



## Project Team (cont.)

- Monarch Wadia
  - Harvard Extension School Student
  - Lives in Toronto



# Project Milestones

- Thursday, March 13: Project proposal due (part of Homework 3)
- Tuesday, April 8th, 11:59 pm: **Announce your project**
- Thursday, April 10, 11:59 pm: **Functional project prototype due**
- Week of April 14: Project review with the TFs
- Thursday, May 1, 11:59pm: **Projects due (including screencast)**
- Thursday, May 8: Best project presentations and prizes



# Motivation

- There was much debate in our group about what the main theme of our project should be—we thought about everything from sports statistics to consumer products like cell phones and automobiles to music listening preferences as illustrated in the visualizations discussed in class. All of the above seemed to be niche categories with respect to our group, with only one or two members really excited about them. We finally came around to the Financial Crisis, a topic that is perhaps a bit overplayed. However, one thing that we unanimously noted was the difference in impact that the financial crisis had on different areas around the country. For example, St. Paul, MN seemed to be less much affected than areas like Phoenix, AZ. The entire group is interested in economics and we believe a topic that everyone is attached to would produce a much more effective team.



# Primary Objectives

- The primary questions we are trying to answer with this visualization are:
  - 1) Is there a significant difference in impact that the financial crisis had on different states in America?
  - 2) Did states see a differentiated response time to the financial crisis in terms of economic indicators?
  - 3) What data best supports the argument that there was a significant difference in impact on states?
  - 4) Can we show unbiased data clearly and concisely, and allow for the audience to both playfully interact with and make concrete conclusions from it?
  - 5) How big of a “wow” factor can we introduce to visitors?
- With the accomplishment of all of the aforementioned goals, we would gain significant specific field-knowledge about best-practice visualization portrayal, the usage of D3 and its incorporation into Javascript, and how to efficiently scrape all of the related data from the web.



# Original Considerations for Data

- **FDIC Failed Bank List**
  - <http://catalog.data.gov/dataset/fdic-failed-bank-list>
- **Quarterly House Price Indexes for Metropolitan Areas**
  - <http://catalog.data.gov/dataset/quarterly-house-price-indexes-for-metropolitan-areas/resource/9fb1e508-f257-46c0-b23f-4c3eac0e9a9a>
- **Local Area Unemployment Statistics**
  - <http://data.bls.gov/pdq/querytool.jsp?survey=la>
- **Terminated Multifamily Mortgages**
  - [http://portal.hud.gov/hudportal/HUD?src=/program\\_offices/housing/comp/rpts/mfh/mf\\_f47t](http://portal.hud.gov/hudportal/HUD?src=/program_offices/housing/comp/rpts/mfh/mf_f47t)
- **Employment Hours/other data by area**
  - <http://catalog.data.gov/dataset/employment-hours-and-earnings-from-the-current-employment-statistics-survey-stat>
- **Low-income Housing problems**
  - <http://catalog.data.gov/dataset/housing-problems-of-low-income-households/resource/799c2406-28b9-453d-b736-da6fdb617a00>
- **County Business Patterns**
  - <http://catalog.data.gov/dataset/county-business-patterns/resource/2bfcc388-170b-4c60-8b41-addba3bef1d4>



# What's next?

- Target
  - Choose domain and research questions
  - Find and clean data
- Translate
  - Formulate data analysis tasks
  - Transform and summarize data
- Design
  - Design visual encodings and interactions
  - Include many sketches
- Implement
  - Use sketches to define data structures
- Validate
  - Post-Project Assessment



# Finding and Cleaning Data

- We queried data from the aforementioned sources
- We anticipated substantial data cleanup and that is what we got
- We found the following obstacles in all of our data
  - Redundant values
  - Junk data points
  - Excessive data points for the same thing
  - Different date-ranged data (ie most were per year while some were every month)
  - Different collection areas (ie some were state-wide, some county-wide, some city-wide)
- Parsing the data was also much more difficult than we expected



# Parsing Data

- Our initial scope included a dataset with many different economic indicators for each state per year
- We decided on a JSON reduction of data size as seen in Problem Set 4
- The initial structure was based on the following sketch:
  - Object for each state
  - Array of years in each state
  - Each year would be a key for another object
    - Economic indicators would be here

Data - JSON

State  
→ [2000:{}, ..., 2011:{}]

  ↳ crime:  
    Unemp:  
    GDP:  
    ~~:  
    ~~:

State  
State  
:



# The Next Data Structure

- We decided on a more object oriented process after running initial JSON directors
  - Object for State
  - Object for Year
    - Key: Values of every indicator

This is my thinking for the structure of the data (haven't narrowed data down yet)

- State Identifier
- Year
  - State Name
  - GSP
  - Median Income
  - Population
  - Number of Patents
  - Unemployment
  - Venture Capital
  - Union Data
  - Bankruptcy

# Final JSON object format

- Objects of indicators inside
  - Crime
  - GSP
  - Homeowner Rates
  - Unemployment
  - Venture Capital

```
▼ Object {US: Object, AL: Object, AK: Object, AZ: Object, AR: Object...}
  ► AK: Object
  ► AL: Object
  ▼ AR: Object
    ► 2000: Object
    ▼ 2001: Object
      ▼ Crime: Object
        Aggravated_assault: "17889"
        Aggravated_assault_rate: "337.1"
        Burglary: "54821"
        Burglary_rate: "1033"
        Forcible_rape: "1518"
        Forcible_rape_rate: "28.6"
        Motor_vehicle_theft: "52203"
        Motor_vehicle_theft_rate: "52203"
        Murder_and_nonnegligent_Manslaughter: "400"
        Murder_and_nonnegligent_manslaughter_rate: "7.5"
        Population: "5306966"
        Property_crime_rate: "5537.5"
        Property_crime_total: "293874"
        Robbery: "8868"
        Robbery_rate: "167.1"
        Violent_Crime_rate: "540.3"
        Violent_crime_total: "28675"
      ► __proto__: Object
    ▼ GSP: Object
      RealGSP: "78447"
      TotalGSP: "71082"
      ► __proto__: Object
    ▼ HomeownerRates: Object
      ownerRate: 71.2
      vacancyRate: 2.5
      ► __proto__: Object
    ▼ Unemployment: Object
      LaborForce: 1357069.5
      TotalUnemp: 63543.67
      UnempRate: 4.75
      ► __proto__: Object
    ▼ VC: Object
      dollarsInvested: "10400000"
      numberOfDeals: "3"
      ► __proto__: Object
    ► __proto__: Object
  ► 2002: Object
  ► 2003: Object
  ► 2004: Object
  ► 2005: Object
  ► 2006: Object
  ► 2007: Object
  ► 2008: Object
  ► 2009: Object
  ► 2010: Object
```



# Normalizing Data

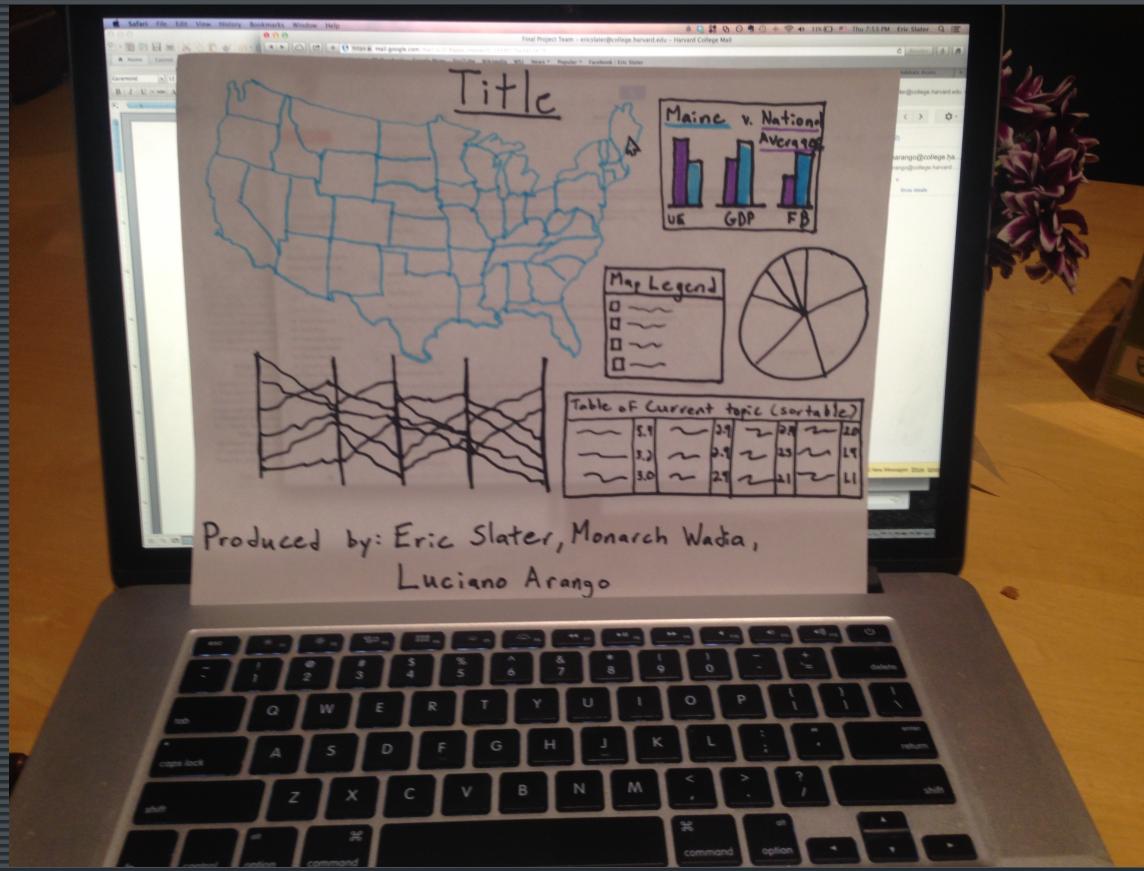
- Crime
  - Aggregate data
  - And rates per 100,000 people
- Gross State Product
  - Divided by population to get per capita
  - In Real and Total terms
- Homeowner Rates
  - Rate out of 100 for owning a home
  - Vacant homes per 100
- Unemployment
  - Unemployment Rate
  - Total number of people Unemployed
  - Total number of people in the Labor Force
- Venture Capital
  - Number of deals and dollars

# Some Mandatory Procrastination Before Design



# Translate

- Our original outline for displaying the United States in indicators during the Great recession:





# Original Scope (before HW3)

- **Must Have Features**

- A central, interactive Chloropleth map of the United States
- An interactive visualization of some shape/form that portrays individual state data (preferably something that loads as you scroll over a state)

- **Optional Features**

- An interactive Parallel Coordinate visualization to compare states in different categories.
- An interactive Hierarchical Edge Bundling that shows connections between different states, times, and/or specific economic indicators.
- Cool Javascript tour of our website and how to do certain interactions with it

# Unfortunately, we did not know that we would be implementing something similar on Homework 4...

- Feedback on our initial Proposal:
  - “You do seem to be a bit ambitious / disorganized here -- you seem to put forth a lot of nonrelated goals, but if want to accomplish all of them that's great”
  - Our Assessment: Our original picture did seem a bit cluttered. Initially we want to remove the pie graph from scope. Additionally the table of current indicators would be a good thing to scale back on.
  - Further evidence “It might be worth considering a better way to 'organize' the different visualizations that you would like to use. The line between a cool 'dashboard' of information and a disorganized agglomeration of different visualization types is a thin one. I would recommend discussing internally a way to organize the visualization so it comes across coherently”



# Further Feedback

- “Your must-have features seem a bit too limited. A chloropleth map, especially given the last homework, will be very straightforward (as will a simple hover tooltip). I'd recommend you move one or more of your optional features into your must-have list.”
  - Assessment: We were not aware of the chloropleth portion of Homework 4 as it was not yet released when we turned this in. So off the bat we are thinking about ways to spice our visualization beyond a map.
  - Further Evidence: “I do want to caution you from doing too little in this cause because of the popularity of your topic and the scope of the homework 4”



# Considering Examples

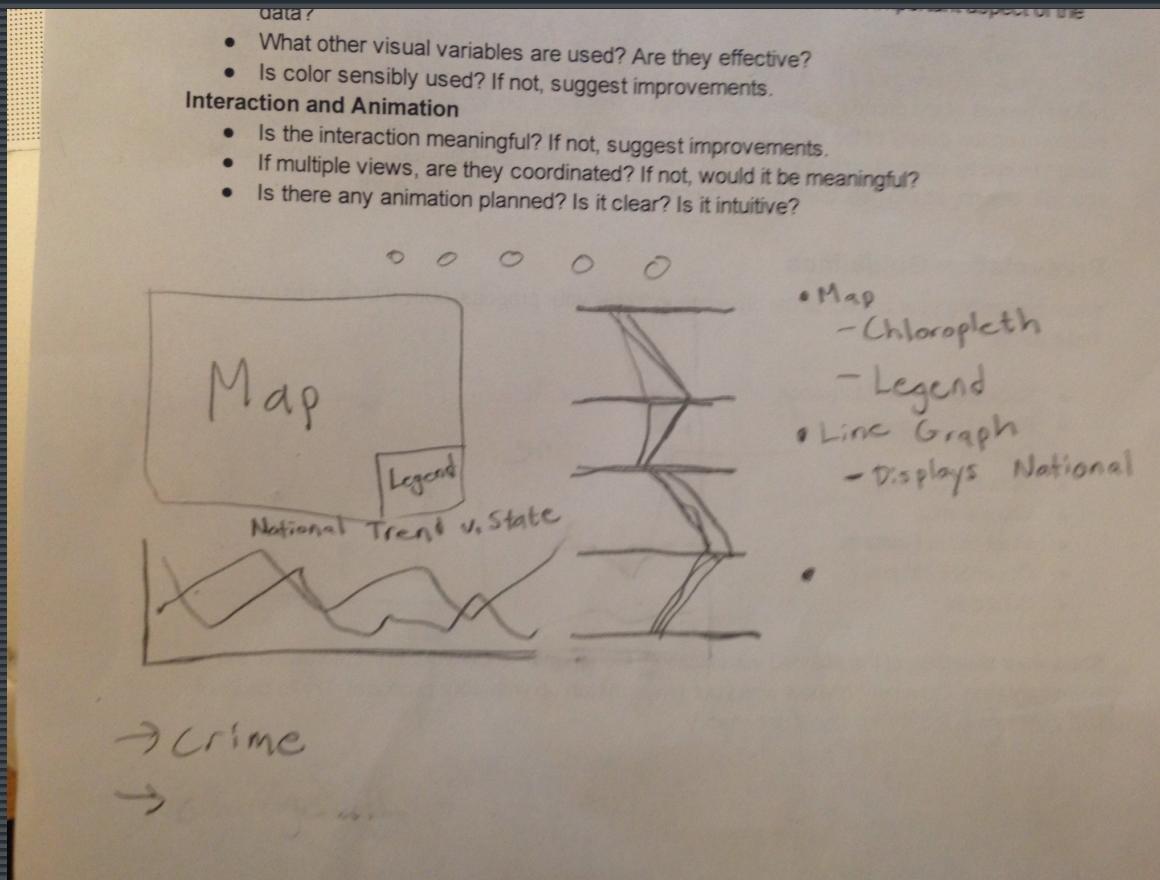
- With our scope under construction we reached out to the internet for further examples of good visualization technique and creativity
  - The first visualization we considered was the winner of last year's projects
  - <https://googledrive.com/host/0B956l87SY0KpRUdrX19TUDhXb1U/index.html>
  - Having talked about Parallel Coordinates and already in our scope, we thought it would broaden our horizons to add a brushable version that would remove color from unselected states on our graph.
  - Turns out our TF was of the same opinion “I think having a linked parallel coordinates brushing graph would be a great idea.”
- We then turned our attention to a less interactive version of mapping state data
  - <http://www.pewstates.org/research/data-visualizations/financial-crisis-5-years-later-and-still-not-recovered-85899502641>
  - Another graph dealing with the financial crisis, but breaking up the states into regions and having the simple user interface of adding and removing regions



# Further Scope Considerations

- “I still am a bit concerned about the scope of your choloropleth (given that is resembles somewhat the bonus of the last homework). While I understand that the data is not given *a priori*; it might make more sense to narrow your project to detailed information for a particular dataset and then be able to ‘dive down’ into a particular state”
- This was definitely a tough thing to hear from our TF
- But understanding his concerns we started to think about new directions to take our project
- After discussing with many other TFs during Office Hours and after lecture we realized that we had quite a challenge on our hands
  - Display 5 different sets of data, for 50 states, over a period of 10 years
  - One might be thinking 3D at this point! But we knew better

# A Solution Sketch 1



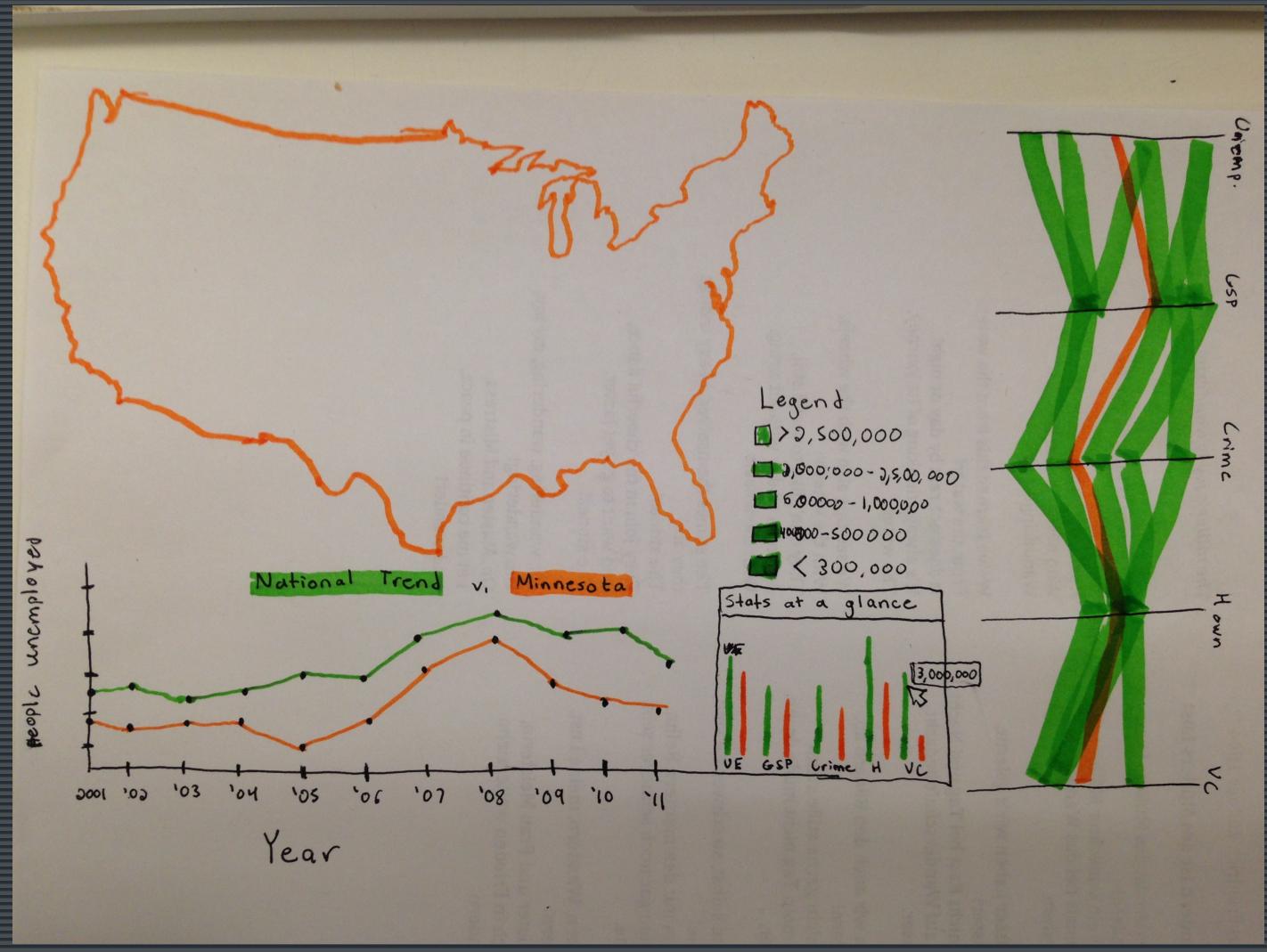


# A Written Solution

- Map
  - connected to parallel coordinates
  - connected to line graph
  - shows one data point at a time that is selected by user
  - Chloropleth - hooked up to a selector on the x-axis of line graph
- Map Legend
  - dynamically changes as data changes
- Line Graph
  - x-axis -- year
  - y-axis -- indicator scale
  - always displays national average
  - click on a state to display it as well
- Indicator Selector
  - radio buttons or something similar to select one of five data points
- Parallel Coordinates
  - Brushable
  - axis -- each indicator
  - displays states with chloropleth color
  - national average also displayed, in a unique differentiating color
  - all in all its main purpose will be as a brushing tool
- Stats at a glance
  - allow user to see all indicators for a selected state in relation to the national average
  - completes the goal of displaying a readable version of all indicators
  - This will be for a given year, which will be picked with a selector hooked up to the x-axis of the line graph

# Solution Sketch 2

- “I do appreciate the increase in scope and think it is more appropriate for the final project”



Daydreaming about possible solutions I  
took the following noteworthy photo





# Considerations

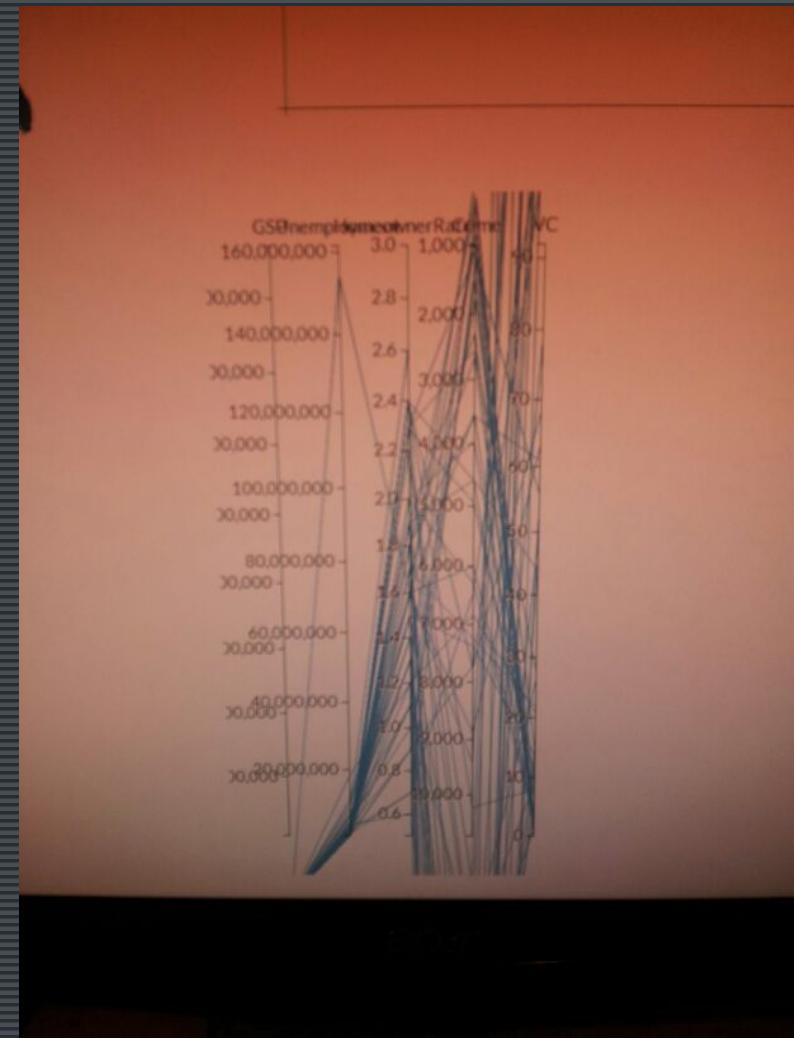
- We are scanning visualizations in reading order and are attracted to titles, text, and labels
  - Put titles at the top left and put labels and textual explanations close to the visualization
- Our visual system sees differences, not absolute values, and is attracted to high-contrast edges
  - Maximize contrast between visual elements to make them stand out and avoid busy textures
- Draw the user's eye to the most important part of the visualization
  - Provide a visual hierarchy of information that will help make it clear to the user how they should interact with the information
- We can easily see objects that are different in color or that are in motion
  - Use color and motion sparingly to make the important information pop out

# Implementation

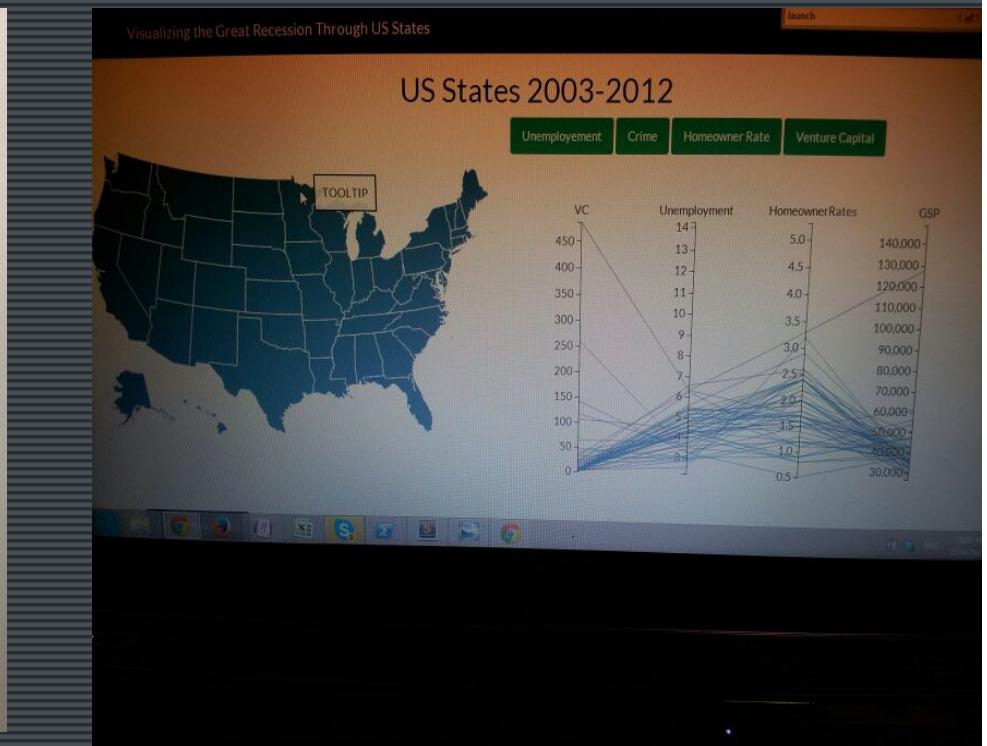
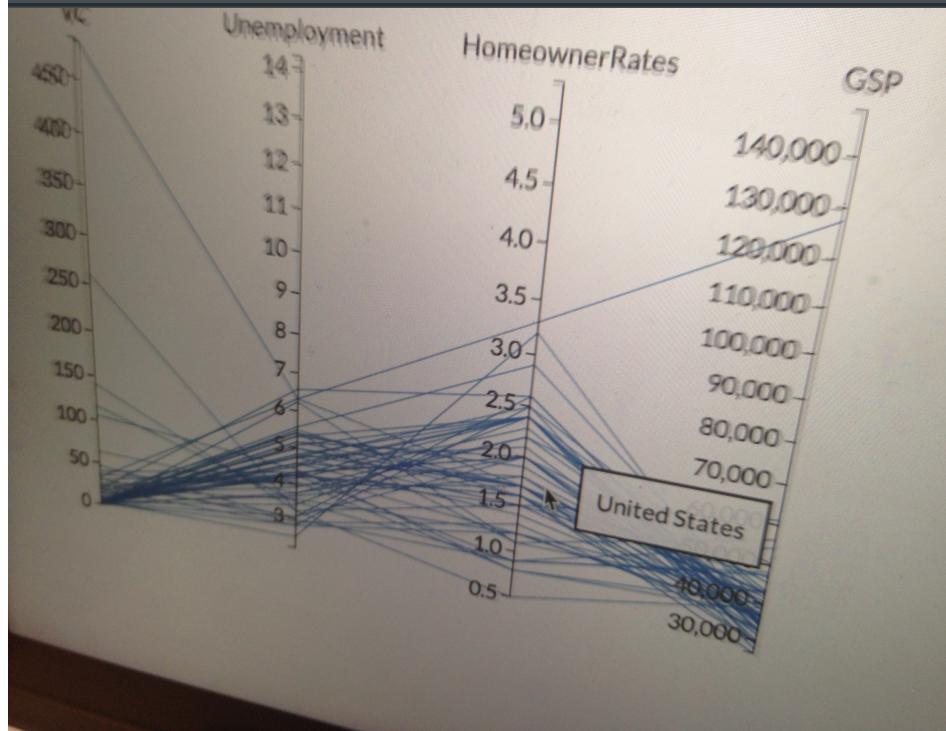
CS171 Project				Documentation			
Assorted Data				Documentation			
GSP_2001-2012	Today, 9:47 PM	--	Folder	Financial Data Mining.docx	Today, 10:03 PM	--	Folder
HomeownerRates.2000_2013	Yesterday, 3:35 PM	--	Folder	IMG_3764.JPG	Mar 7, 2014, 11:10 AM	225 KB	Micros...ument
MedianIncome_2000.2012	Mar 31, 2014, 4:53 PM	95 KB	Micros...kbook	Log.docx	Today, 10:03 PM	1.8 MB	JPEG image
Population_2000.2012	Apr 5, 2014, 2:22 PM	98 KB	Micros...kbook	Monarch Wadia - In progress.docx	Apr 3, 2014, 8:28 PM	94 KB	Micros...ument
Population_2000.2013	Mar 31, 2014, 5:05 PM	78 KB	Micros...kbook	Monarch Wadia - Mus...nd Nice to Haves.docx	Mar 13, 2014, 7:42 PM	12 KB	Micros...ument
ResidentialBuildingPermits.2000_2013	Mar 31, 2014, 5:08 PM	71 KB	Micros...kbook	proposal_Slater_Arango_Wadia.docx	Mar 13, 2014, 7:40 PM	11 KB	Micros...ument
Bugs.docx	Mar 31, 2014, 5:08 PM	71 KB	Micros...kbook		Mar 13, 2014, 7:53 PM	100 KB	Micros...ument
Code	Apr 4, 2014, 8:19 AM	133 KB	Micros...kbook		Yesterday, 9:18 PM	--	Folder
CrimeRateToJSON	Today, 2:13 PM	27 KB	Micros...ument		Today, 9:35 PM	--	Folder
CrimeRateToJSON.html	Yesterday, 10:37 PM	--	Folder	FinalJSONwithCrime.json	Yesterday, 7:43 PM	491 KB	JSON
DataToJson.html	Apr 4, 2014, 8:58 AM	2 KB	Document	GSP_Ue_Homeowner.json	Apr 4, 2014, 10:05 AM	138 KB	JSON
HomeOwnerRates.html	Yesterday, 3:39 PM	1 KB	HTML	GSP_Ue_HOWN_CRIME_VC_POP.json	Yesterday, 3:45 PM	527 KB	JSON
Monarch_formatData_array.html	Apr 4, 2014, 10:04 AM	3 KB	HTML	GSP_Ue_HOWN_CRIME_VC.json	Apr 4, 2014, 10:09 AM	510 KB	JSON
Monarch_formatData.html	Apr 4, 2014, 10:05 AM	2 KB	HTML	GSP_Ue_HOWN_CRIME.json	Apr 4, 2014, 10:08 AM	474 KB	JSON
newArray.html	Apr 7, 2014, 4:21 PM	3 KB	HTML	GSP_UnemploymentByYear.json	Apr 4, 2014, 10:04 AM	99 KB	JSON
newArray2.html	Apr 7, 2014, 3:46 PM	2 KB	HTML	GSP_UnemploymentByYear2.json	Apr 4, 2014, 8:35 AM	89 KB	JSON
newFinalToJson.html	Yesterday, 10:37 PM	2 KB	HTML	ihopethisisit.json	Yesterday, 11:33 PM	44 KB	JSON
OneMoreGo.html	Today, 9:34 PM	3 KB	HTML	lucArry.json	Yesterday, 4:14 PM	527 KB	JSON
OneMoreGoDiff.html	Yesterday, 7:42 PM	4 KB	HTML	lucArry2.json	Yesterday, 4:16 PM	40 KB	JSON
PopulationToJson.html	Yesterday, 7:42 PM	4 KB	HTML	lucArry3.json	Yesterday, 4:24 PM	39 KB	JSON
UnemploymentToJson.html	Yesterday, 9:28 PM	3 KB	HTML	LucianoArrayWithCrime.json	Yesterday, 10:42 PM	491 KB	JSON
VentureCapitalToJson.html	Today, 6:29 PM	--	Folder	LucianoPlusID.json	Yesterday, 11:12 PM	44 KB	JSON
cs-171-final-project copy	Today, 2:21 PM	--	Folder	LucianoPlusD2.json	Yesterday, 11:16 PM	44 KB	JSON
DataCSV	Today, 2:21 PM	--	Folder	LucianoPlusD3.json	Today, 7:11 PM	37 KB	JSON
Bankruptcy_2000.2012	Apr 1, 2014, 2:57 PM	1.7 MB	comm...values	LucianoPlusD4.json	Today, 9:29 PM	37 KB	JSON
CrimeStatebyState.csv	Apr 4, 2014, 10:06 AM	100 KB	comm...values	LucianoPlusD5.json	Today, 9:35 PM	36 KB	JSON
CrimeStatebyState2.csv	Apr 5, 2014, 2:22 PM	100 KB	comm...values	LucianoplusD.json	Yesterday, 11:07 PM	491 KB	JSON
crimeUS	Yesterday, 6:29 PM	34 KB	Micros...kbook	MonarchJSONpp.json	Yesterday, 7:45 PM	491 KB	JSON
GSP_2001.2012	Mar 31, 2014, 8:40 PM	49 KB	comm...values	MonarchStyle.json	Yesterday, 8:59 PM	63 KB	JSON
GSP_2001.2012.2.csv	Apr 1, 2014, 2:57 PM	49 KB	comm...values	MonarchStyle2.json	Yesterday, 9:26 PM	75 KB	JSON
HomeownerRates.2000_2013	Apr 4, 2014, 8:18 AM	73 KB	comm...values	MonarchStyleDiffs.json	Yesterday, 10:04 PM	107 KB	JSON
MedianIncome_2000.2012	Apr 1, 2014, 2:57 PM	22 KB	comm...values	newMonarchJSON.json	Yesterday, 5:59 PM	578 KB	JSON
Nativity_2003-2011.csv	Apr 2, 2014, 4:53 PM	24 KB	comm...values	plusUE.json	Apr 1, 2014, 3:21 PM	187 KB	JSON
PatentsPerState_2000.2012	Apr 1, 2014, 2:57 PM	24 KB	comm...values	plusUEByMonth.json	Apr 1, 2014, 3:23 PM	635 KB	JSON
Population_2000.2013	Apr 1, 2014, 2:57 PM	21 KB	comm...values	reducedStateData.json	Apr 4, 2014, 9:54 AM	36 KB	JSON
ResidentialBuildingPermits.2000_2013	Apr 4, 2014, 8:19 AM	67 KB	comm...values	MONARCH_CLEANED_...-IndicatorsObject.json	Apr 7, 2014, 4:21 PM	97 KB	JSON
salesofmealawayfromhome.csv	Apr 2, 2014, 4:53 PM	4 KB	comm...values	MONARCH_CLEANED_DATA.json	Apr 7, 2014, 2:56 PM	53 KB	JSON
Taxes on Production and Imports Less Subsidies	Apr 3, 2014, 10:39 PM	--	Folder	MonarchsPretty.json	Yesterday, 9:18 PM	117 KB	JSON
UnemploymentByMonth_2000.2014	Apr 1, 2014, 3:13 PM	468 KB	comm...values	plusUE_pretty.json	Apr 3, 2014, 9:28 AM	399 KB	JSON
Unions_2001.2013	Apr 1, 2014, 2:57 PM	68 KB	comm...values	plusUEByMonth_pretty.json	Apr 3, 2014, 9:29 AM	1.2 MB	JSON
VentureCapital_2000.2012	Apr 4, 2014, 9:40 AM	20 KB	comm...values	reducedStateData_pretty.json	Apr 3, 2014, 9:29 AM	48 KB	JSON
VentureCapital_2000.2012.2.csv	Apr 5, 2014, 2:22 PM	20 KB	comm...values	ideas for linked coordinates.js	Apr 6, 2014, 6:36 PM	26 KB	JavaScript
VentureCapital2000.2012.csv	Apr 1, 2014, 2:57 PM	20 KB	comm...values	LectureNotes.docx	Today, 5:32 PM	364 KB	Micros...ument
Documentation	Today, 10:03 PM	--	Folder	libs	Apr 4, 2014, 8:10 AM	--	Folder
Financial Data Mining.docx	Mar 7, 2014, 11:10 AM	225 KB	Micros...ument	colorbrewer.js	Mar 18, 2014, 11:10 AM	19 KB	JavaScript
IMG_3764.JPG	Today, 10:03 PM	1.8 MB	JPEG image	d3.v3.min.js	Mar 9, 2014, 2:38 PM	147 KB	JavaScript
Log.docx	Apr 3, 2014, 8:28 PM	94 KB	Micros...ument	FileSaver.js	Mar 18, 2014, 11:10 AM	8 KB	JavaScript
Monarch Wadia - In progress.docx	Mar 13, 2014, 7:42 PM	12 KB	Micros...ument	jquery-1.11.0.min.js	Mar 18, 2014, 11:10 AM	96 KB	JavaScript
Monarch Wadia - Mus...nd Nice to Haves.docx	Mar 13, 2014, 7:40 PM	11 KB	Micros...ument	queue.v1.min.js	Apr 4, 2014, 8:10 AM	692 bytes	JavaScript
proposal_Slater_Arango_Wadia.docx	Mar 13, 2014, 7:53 PM	100 KB	Micros...ument	topojson.v1.min.js	Apr 4, 2014, 8:09 AM	6 KB	JavaScript
				mongoose-free-5.2.exe	Mar 8, 2014, 5:25 PM	82 KB	Windo...rchive
				mongoose.conf	Apr 7, 2014, 4:23 PM	415 bytes	Document
				Process Book	Today, 10:30 PM	10.8 MB	Micros...tation
				Screen Shot 2014-04-09 at 8.11.01 PM.png	Yesterday, 8:11 PM	15 KB	PNG image

# Parallel Coordinates

- Shows primarily relationships between adjacent axis
  - Clustering & aggregating records
  - Transparency of lines
  - Interaction is crucial
  - Axis reordering
  - Brushing
  - Filtering
- 
- Our first attempt was somewhat fruitful!



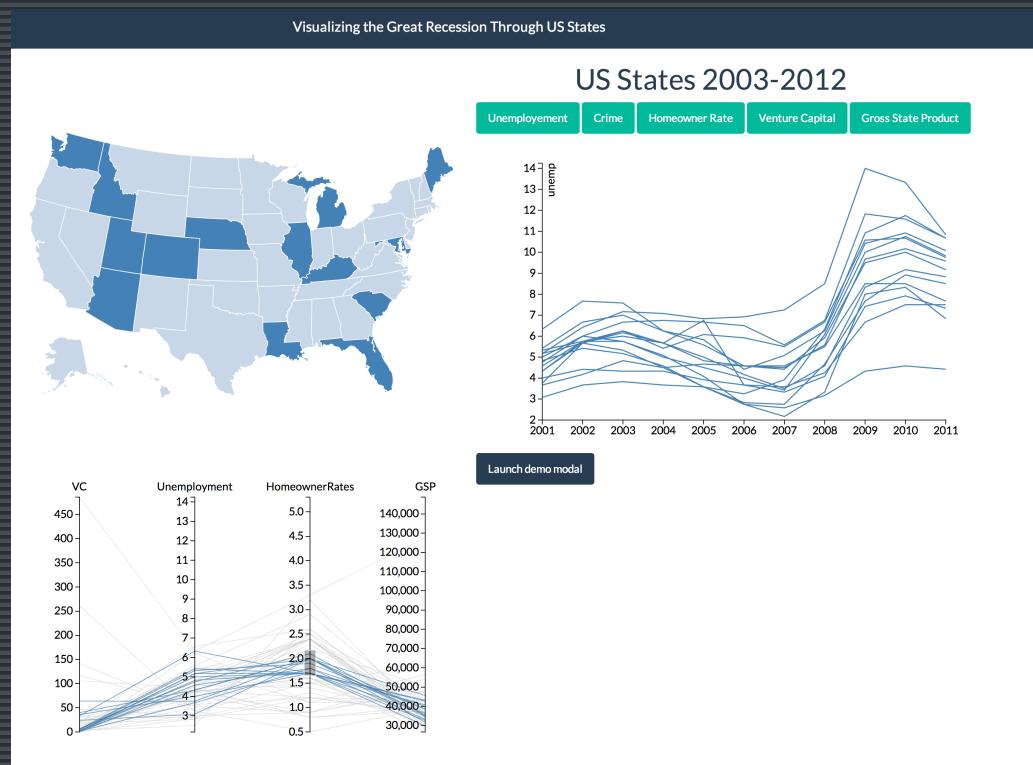
# Parallel Coordinates Take 2 (&3)



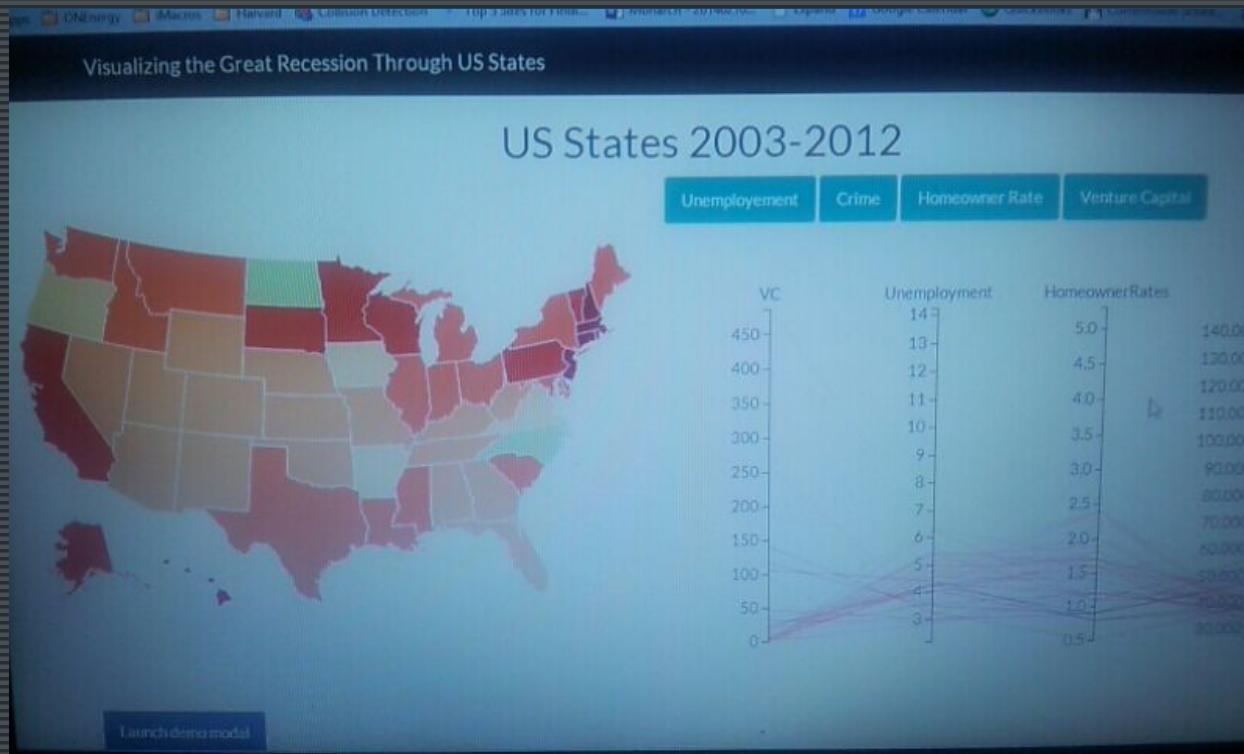


# A Line Graph

- But not just any line graph
  - We want interactivity with both the map and the Parallel Coordinates
  - Usually displays 50 lines (one for each state)
  - Gives us a time range from 2001 to 2011



# Enter Chloro



- Still working on the color scheme...

# A list of other details to include in next process book

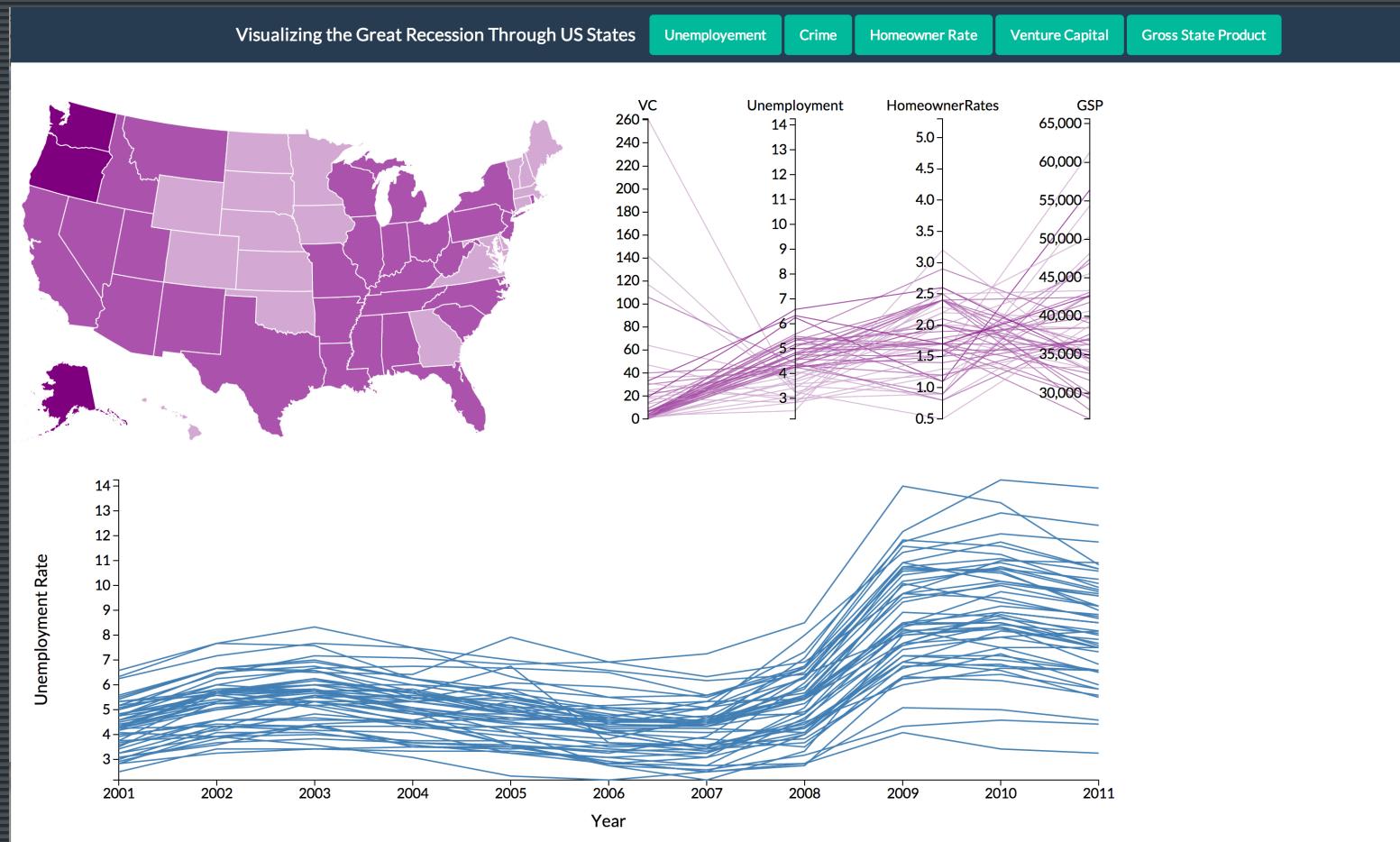
- Data
  - There was a constant re-arrangement of data into different arrays and objects to make all graphs work in sync
  - We are thinking about changing some of the data indicators, as it seems that it is harder to see a correlation than originally thought for some (data was parsed in STATA before entering it)
  - Literally spent more time on data manipulation than the coding so far
- Design Decisions
  - Many design decisions were made along the line that are not discussed in depth here. When we near completion, I will better be able to elaborate on how these impacted our final product
  - Scope was a major setback for us given that Problem Set 4 dealt with our chloropleth ideals. However, we have arrived at a solution that goes above and beyond such simple ideas as projecting a map and coloring it.



# Where are we now?

- US Chloropleth Map hooked up to five different sets of data controlled by the parallel coordinates.
- Parallel Coordinates that are brushable and that affect our line graph and map when brushed.
- A Line Graph for a specific indicator and any selected states over the time period we are observing.
  - Highlighting functionality within the Line Graph
- Buttons that change the indicator of interest
- Tooltip that displays the name of the state selected

# Here





# Where are we going?

- A Map with dynamic transitions, five different color schemes, a tooltip that shows “stats at a glance,” and hover over capability.
- Parallel Coordinates that take all indicators into account and that will highlight with the highlighting of the line graph
- A selector for year, preferably on the axis of the line graph
- A more synced color scheme and typeface
- Responsiveness to browser size