

# Process Book

[Google doc link:](#)

[https://docs.google.com/document/d/1tT17falrFtT0OJriwHM398GTD0MGfH807ySTJ4-4es0/edit  
?usp=sharing](https://docs.google.com/document/d/1tT17falrFtT0OJriwHM398GTD0MGfH807ySTJ4-4es0/edit?usp=sharing)

## Project Proposal: **Visualizing Mass Incarceration**

### [Visualizing Mass Incarceration In the United States - Presentation](#)

The United States has the highest incarceration rate in the world. Those incarcerated are disproportionately economically and socially disadvantaged. Often, imprisonment begins in school in what is known as the “School to Prison” pipeline, whereby students are arrested by resource officers and sent to juvenile detention centers or Disciplinary Alternative Schools as opposed to being disciplined by their teachers within the context of their schools. Some potential sources of data for this project are:

- <https://www.sentencingproject.org/the-facts/#rankings?dataset-option=SIR>
- <https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system>
- [https://www.kaggle.com/christophercorrea/prisoners-and-crime-in-united-states#crime\\_and\\_incarceration\\_by\\_state.csv](https://www.kaggle.com/christophercorrea/prisoners-and-crime-in-united-states#crime_and_incarceration_by_state.csv)
- <https://dataverse.harvard.edu/api/access/datafile/:persistentId?persistentId=doi:10.7910/DVN/24827/HNKFPV>
- <https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/charter-schools-civil-rights-and-school-discipline-a-comprehensive-review>
- <http://trends.vera.org/incarceration-rates?data=pretrial>

This data lends itself well to a narrative visualization style. Our goal is to tell a story regarding incarceration in the United States, beginning with visualizing national and statewide data demographic trends over time, then delving into the more specific issue of juvenile incarceration. Of particular interest is the effect of the charter school boom on juvenile incarceration. We would like to analyze data from a variety of dimensions, including socioeconomic status, race, location, and other aspects of people’s backgrounds (depending on datasets we continue to find and choose to use) in order to show which of those dimensions have a greater correlation with incarceration, and how those influences have changed over time. We hope to draw from visualization styles such as choropleth maps and scatterplots, but also to create project-specific.

# TABLE OF CONTENTS

TEAM AGREEMENT	3
MAP	4
SKETCH	8
DECIDE	13
PROTOTYPE	14
TESTING	26
FINAL CHAPTER	34

# TEAM AGREEMENT

## Communication:

We will use Slack for our primary communications.

## Meeting time and Location:

Sundays at noon at Smith.

## Underperforming team members:

We will try to mitigate this by setting clear deliverables for each team member in our weekly meeting. We will also try to communicate directly instead of subtle/passive clues about expectations.

## Collaboration on Implementation:

GitHub repository (created & shared)

### Collaboration Process:

1. Create a branch
2. Make changes
3. Make pull request
4. Notify team members about changes
5. Merge into the master branch
6. Resolve conflicts

Visualizations will be modular in their own individual JS files.

## Team Roles:

We all decide by vote

Shruthi: video narration, writing, creative consultant

Chris: cs, video narration, design/prototype consultant

Nikhil: cs, data wrangler

Simon: cs, Project Manager

## Signatures

  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_

# MAP

## Data:

<b>Charter School Data</b> Enrollment % total enrollment Out of school suspension Out of school suspension rates State School District  <i>Each of the above for races:</i> Asian Black Latino Native White Other  <i>Each of the above for gender:</i> Male Female  <i>via The Civil Rights Project</i>	<b>Crime Data:</b> State Year County Population (Gender/Race) Total Violent Crime Murders & Manslaughters Rape Robbery Assault Burglary Property Crime Vehicular Theft Larceny Arson Jail population (Gender/Race) Prison population (Gender/Race)  <i>via Vera institute of Justice</i>
---	--

## **Data Sets:**

- <https://www.sentencingproject.org/the-facts/#rankings?dataset-option=SIR>
- <https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system>
- [https://www.kaggle.com/christophercorrea/prisoners-and-crime-in-united-states#crime\\_and\\_incarceration\\_by\\_state.csv](https://www.kaggle.com/christophercorrea/prisoners-and-crime-in-united-states#crime_and_incarceration_by_state.csv)
- <https://dataverse.harvard.edu/api/access/datafile/:persistentId?persistentId=doi:10.7910/DVN/24827/HNKFPV>

- <https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/charter-schools-civil-rights-and-school-discipline-a-comprehensive-review>
- <http://trends.vera.org/incarceration-rates?data=pretrial>

**Note:** We may need to do some data manipulation in order to standardize things by county and by state

### **Data Questions:**

1. What is the effect of the charter school boom on juvenile incarceration?
    - Scatter Plot with trend lines
  2. What states have the highest rates of incarceration/various types of crimes?
    - Choropleth (Nation/State/County Data)
    - Animation over time
  3. What are the correlations of background dimensions (socioeconomic status, race, location) with incarceration?
    - Correlation Matrix
  4. How have those influences have changed over time?  
 Animation - Overtime  
 The slider moves with animation, and once it's done the user can manipulate it as they want: guided and then exploratory.
- Line Chart  
 Scatterplot with trendlines  
 Bubble Chart - Hans Rosling  
 Linked views
5. How do incarceration rates vary between urban and rural areas?
  6. Which states have the worst prison conditions?

### **Notes:**

- Powerful comparisons(States and external countries) to show gravity
- Choropleth transitions to cartogram.

# Goals and Tasks

## **Background:**

Our goal is to tell a story regarding incarceration in the United States, beginning with visualizing national and statewide data demographic trends over time, then delving into the more specific issue of juvenile incarceration. Of particular interest is the effect of the charter school boom on juvenile incarceration. We would like to analyze data from a variety of dimensions, including socioeconomic status, race, location, and other aspects of people's backgrounds (depending on datasets we continue to find and choose to use) in order to show which of those dimensions have a greater correlation with incarceration, and how those influences have changed over time. In order to do so we will offer the following 3 visualizations.

## **Visualization #1 (Interactions)**

We will begin by visualizing the prison population in the United States. We will use a dot matrix to visualize this data (see sketches for reference). Once we've shown this data, we want to dive into the specific demographics of this imprisoned group. The visualization will thus zoom into the square that represents the incarcerated population of the United States, at which points the dots will rearrange themselves into bars for a bar chart, allowing the viewer to see the racial, gender, socio-economic, age, etc. breakdown of those imprisoned in the united states.

## **Visualization #2**

Choropleth of the United States, which breaks down the incarceration rate by state and by county. We will allow filtered based on the demographic data in the previous visualization. We want this to be zoomable so that the user can interact with it.

## **Visualization #3**

Dynamic bubble chart representing school suspensions vs incarcerations over years on a district-level. We are also looking at the connection to charter school populations in the context of different district location types.

## **Visualization #4**

Matrix with custom icons showing the average facilities available in prisons across the country.

## Tasks

1. **Consolidate Data:** Our datasets differ slightly in that some are in terms of state, city, county, etc. We want to consolidate this, which may require some scripting.
2. **Implement Each Visualization:** Visualization #1 will transition between multiple views, so first, we'll implement them in isolation before working on the transitions
3. **Research** mass-incarceration and write up good background paragraphs for each section of our webpage. We'll also want to grab some quotes and key stats that we'll accentuate on the webpage. Since each view

## Features List

Must-haves:

1. Choropleth map
2. An innovative visualization
  - a. Visualization #1 (interactive) - ask the user to guess the incarcerated population, and dots become bar chart
  - b. Viz 2: bubble chart
  - c. Visualization 3: Choropleth of US incarceration rates, by state and by county
3. Quotes from incarcerated people, prison activists, and/or policy-makers
4. A couple of key outlier data points that we'll highlight as part of our page
4. Images and case studies

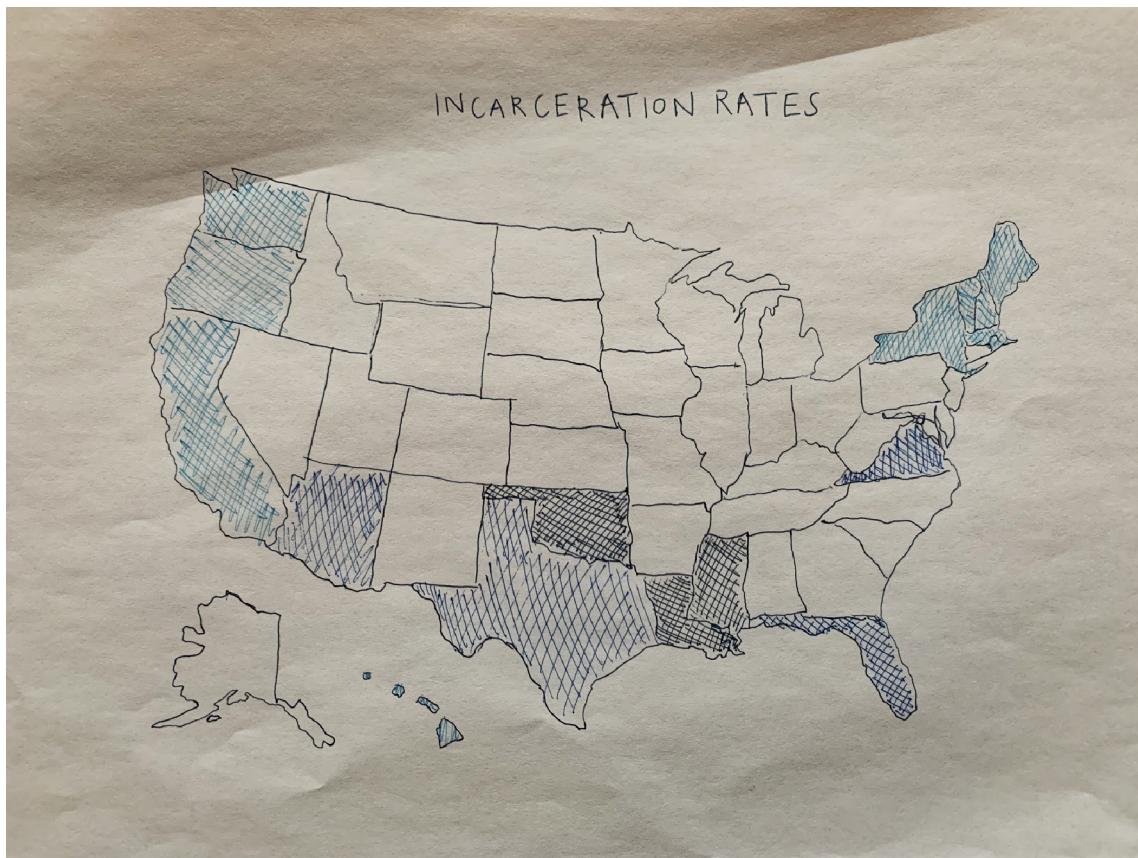
## Timeline

- **November 17th** - Consolidate Datasets into a Master Data Set and augment the data as necessary
- **November 24th** - Begin work on each of the visualizations, meet up before thanksgiving. 3 People will work on individual visualizations and one person will work on the overall webpage, the next, quotes, notable stats, etc.
- **December 1st** - Reconvene after Thanksgiving, review code, and assess needs before the final deadline. Spend the next weeks consolidating code and working on bringing it together for the actual webpage.
- **December 3rd** - Finish up the first version of the webpage. Get some feedback from peers and make these changes before the due date.

# SKETCH

## Visualization Sketches

### Sketch 1:

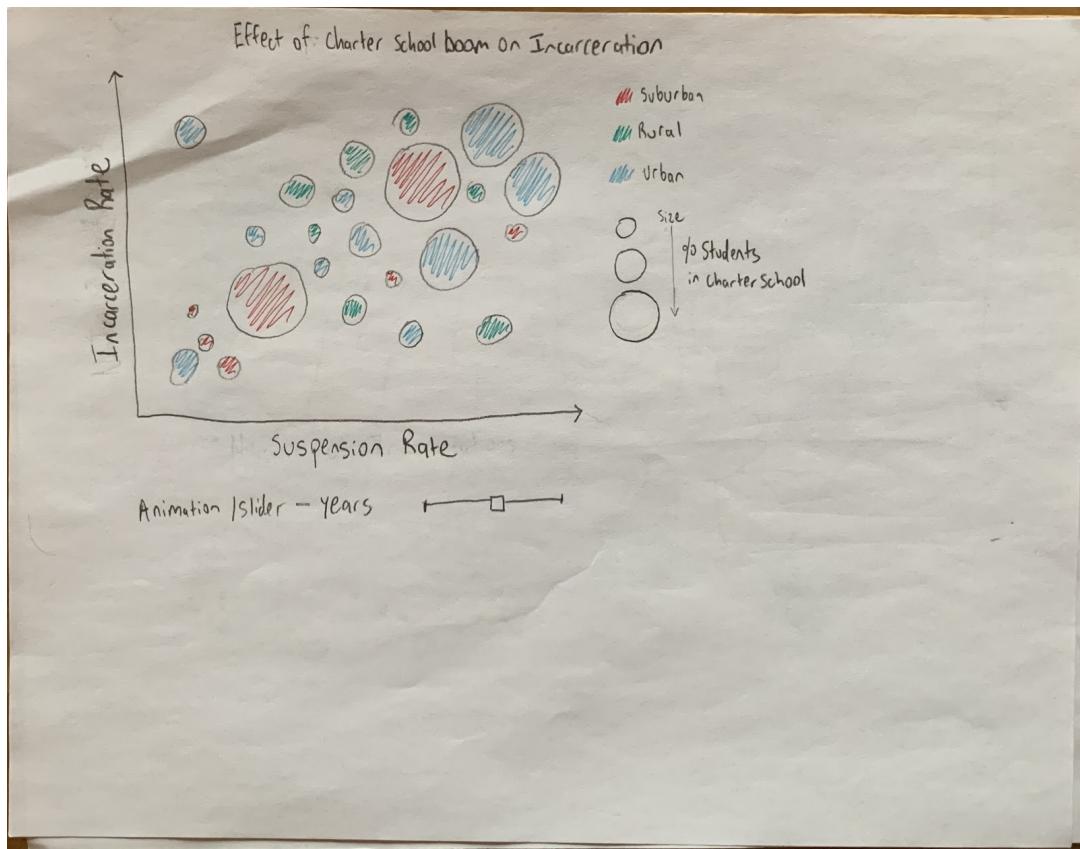


Choropleth of states - incarceration rates

Filter by crimes, jail conditions, prison populations (race, gender)

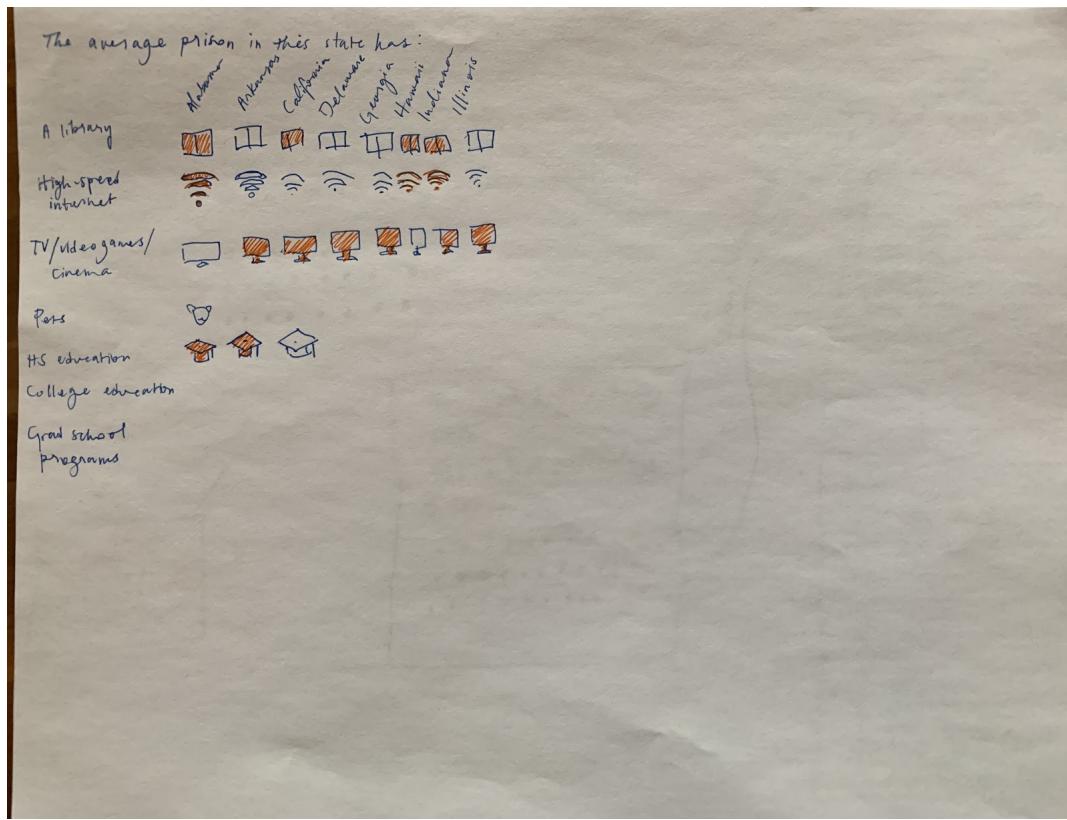
Include timeline filtering/animation

## Sketch 2:



Bubble chart of suspensions vs incarcerations over the years. View by region type and see how popular charter schools are.

### Sketch 3:



Matrix of amenities available in prisons in a specific state.

## Interaction Storyboard

Interactions

Us population in M

Each of these dots rep 1000 people.  
How many of these do you think are incarcerated? (Drag to make selection)

Drag

Zoom/Transition

Dot of the 35%  
55% are male  
45% are female  
(dots become bars)

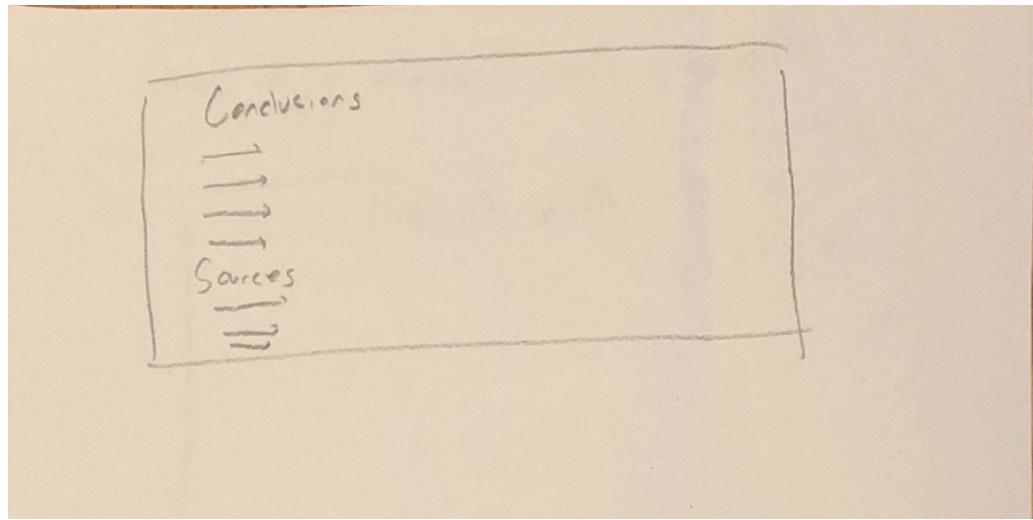
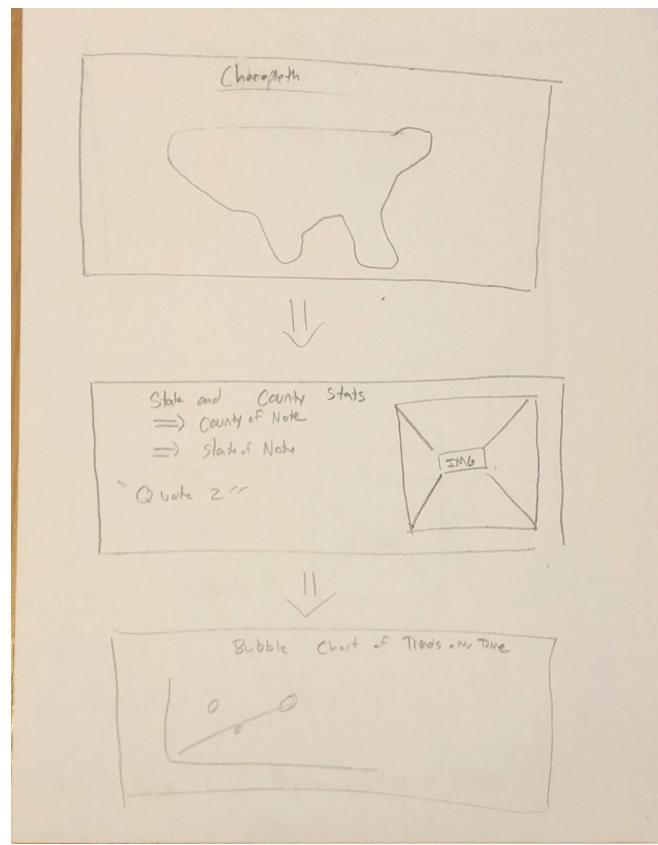
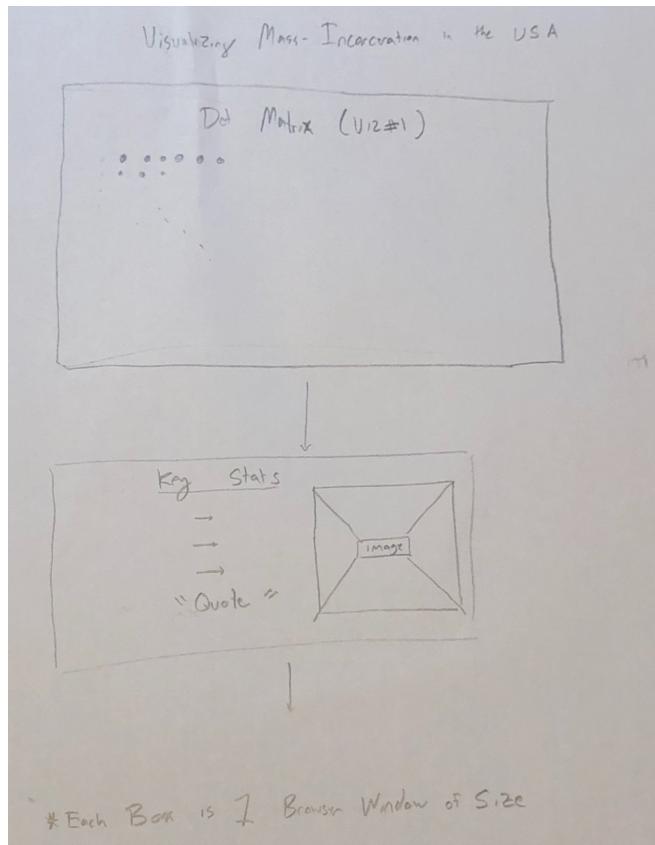
Female Male

You selected 30% ! Very close.  
The actual percentage is 35%

25% are Black  
15% are Latino  
10% are Native  
2% are White

Black Lat Net White

## Webpage Layout



# DECIDE

Honed in on Civil Rights Data Collection from Ed.gov (96,360 schools, 1,836 attributes), pulled this data and joined school districts on county

<https://ocrdata.ed.gov/DownloadDataFile>

School District to County Mappings

<https://catalog.data.gov/dataset/school-district-geographic-relationship-files>

Vera Incarceration Trends

<http://trends.vera.org/incarceration-rates>

# PROTOTYPE

Students who worked on V1:

Shruthi Venkata

---

Chris Kinyua

---

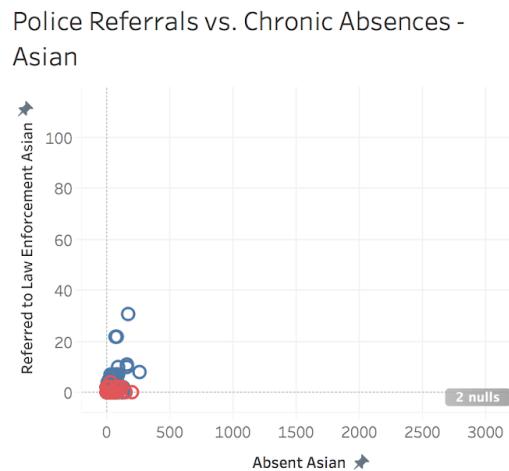
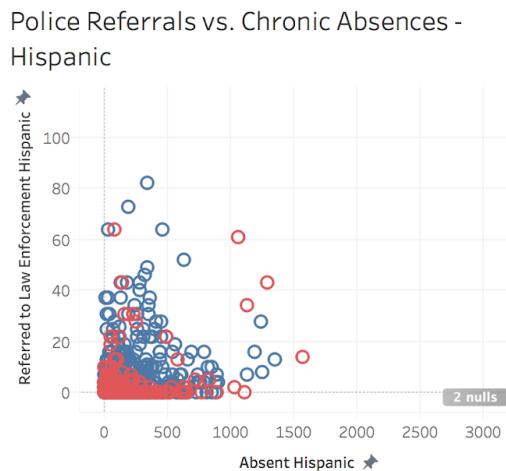
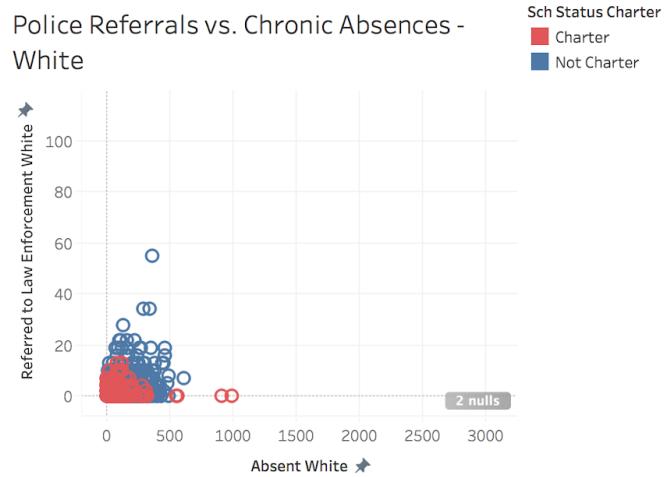
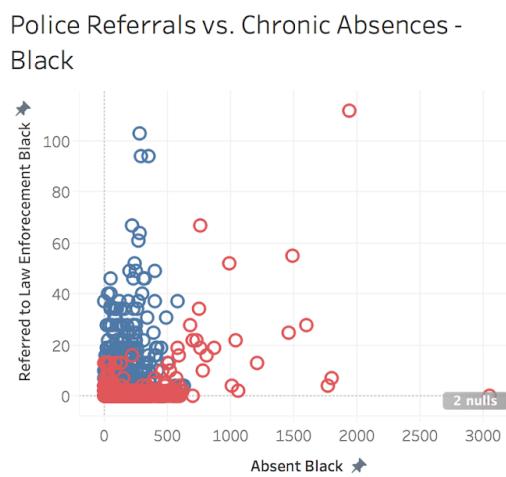
Nikhil Vanderklaauw

---

Simon Warchol

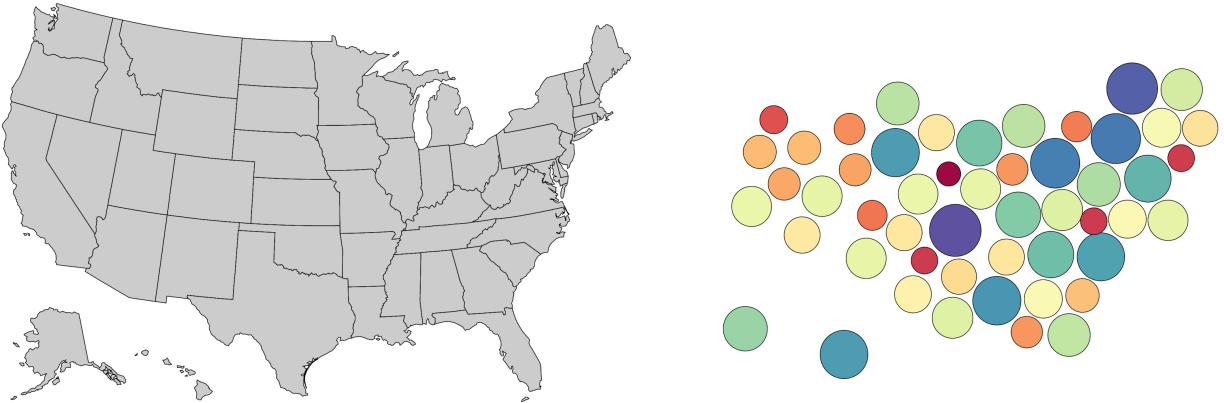
<https://github.com/Cs171-Final-Project/Final-Project>

## Scatter Plot Prototypes



## Interactions

This will be presented as one graph, with toggle to switch between different race groups. Showcasing difference between charter and non-charter schools, expressed as a percentage for each school.



## Choropleth/Cartogram

Color: Crime rate of state (normalized population)

Size:

- % of students suspended: in-school + out-of-school
- % of students expelled / expelled w zero tolerance
- % of students referred to authorities
- % of students receiving school related arrest

Scatter

## Stacked Bar Chart

For each school type on X axis

Y is the total number of offenses, broken into categories:

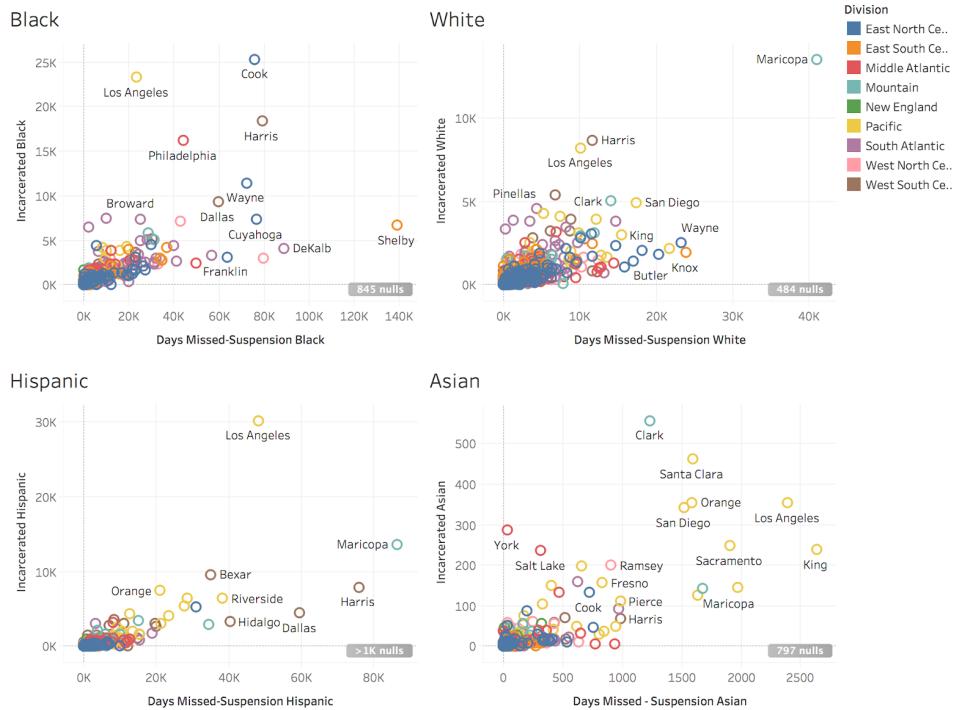
- Incidents of rape
- Incidents of sexual assault
- Incidents of robbery
- Incidents of physical attack
- Incidents of threats of physical attack
- Incidents of possession of firearm

For race on X, y is the counts of suspension categories by gender

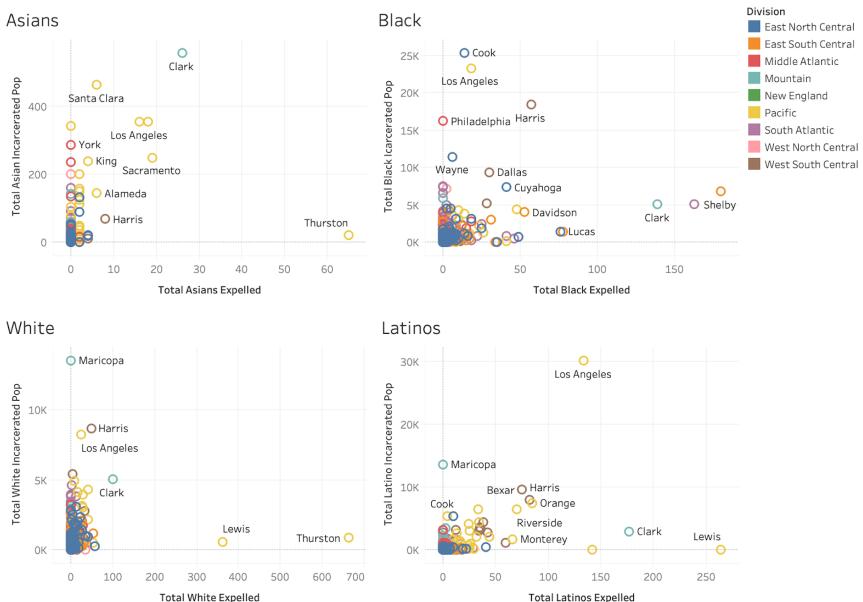
School Employees:

- Number of school counsellors, psychologists, social workers
- Vs: Number of security guards, law enforcement officers

### Incarcerations vs. Days of School Missed Due to Suspensions



### Total Incarceration vs Total Expulsions UZT by Race



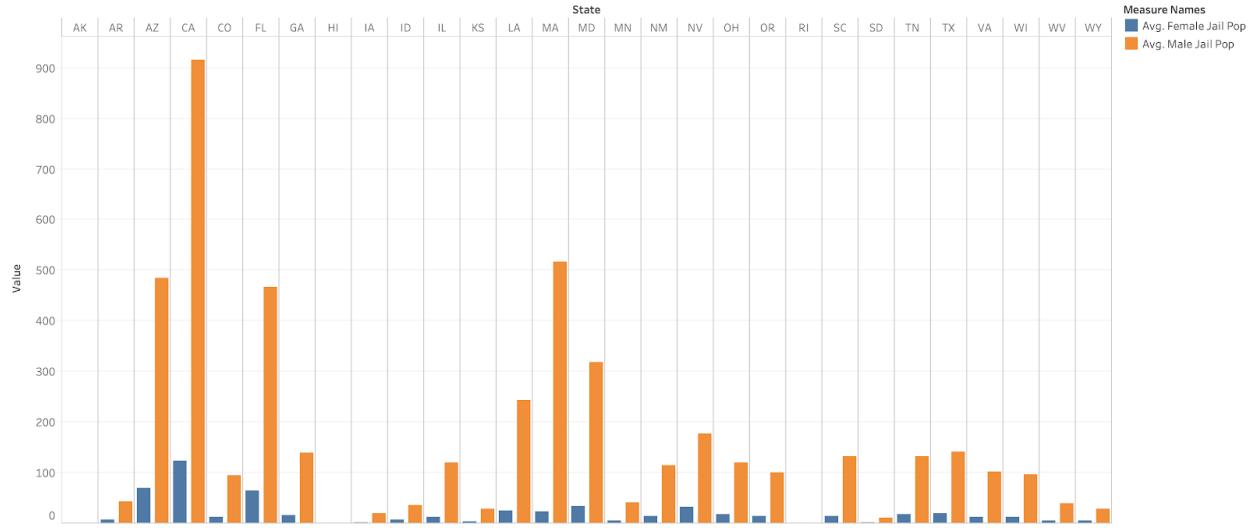
### Interactions:

Filtering - The user can click the legend to highlight related points in the graph and grey out unrelated areas.

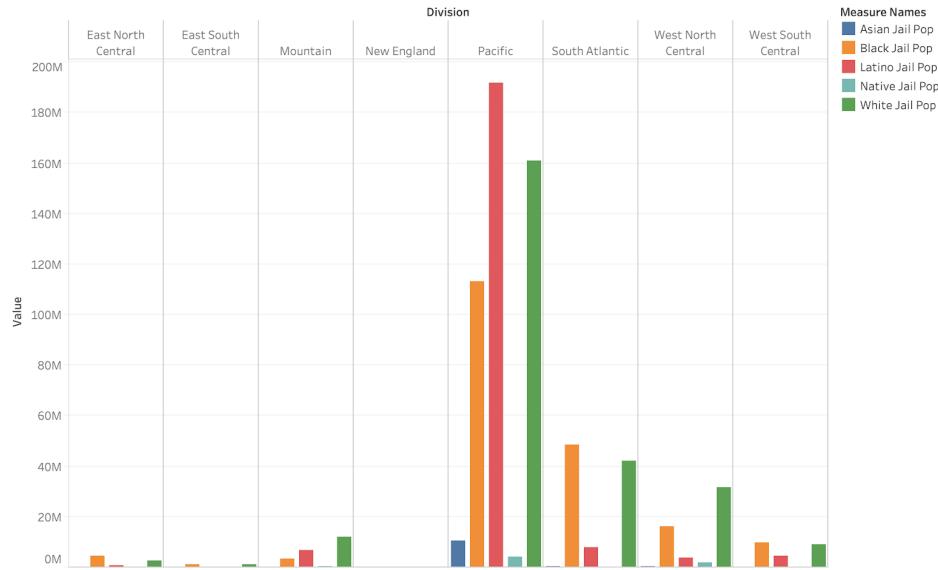
**Hover** - A tooltip with additional information about a datapoint is shown when the user hovers above the point.

### Bar Chart Prototypes

Average State Jail Population by Gender



Average Region Jail Population By Race

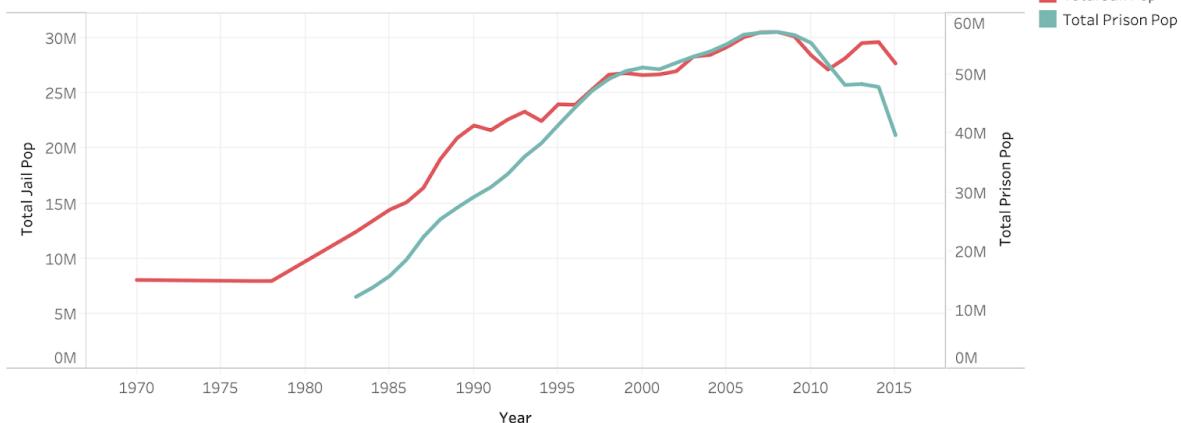


### Interactions:

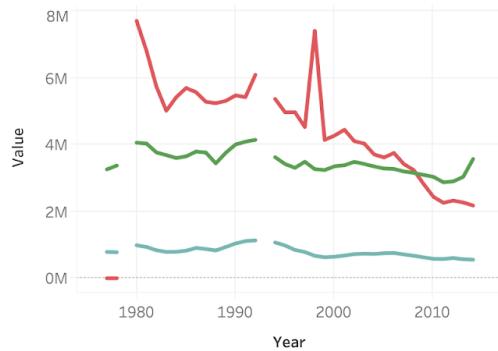
**Hover** - A tooltip with additional information about a datapoint is shown when the user hovers above the point.

## Line Chart (Trend) Prototypes

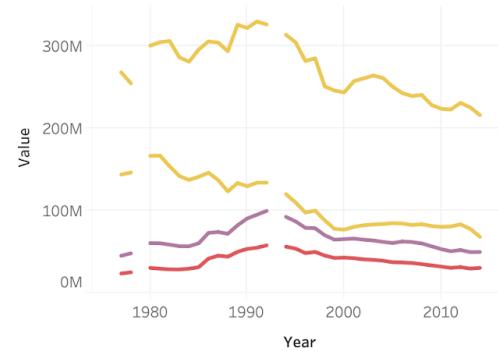
Trend in US total jail and prison population between 1980 and 2016.



Trends in reported cases of Murder, Rape and Arson crimes between 1970 and 2015.



Trends in Agr Assault, Burglary, Larceny and Violent Crimes between 1970 and 2015

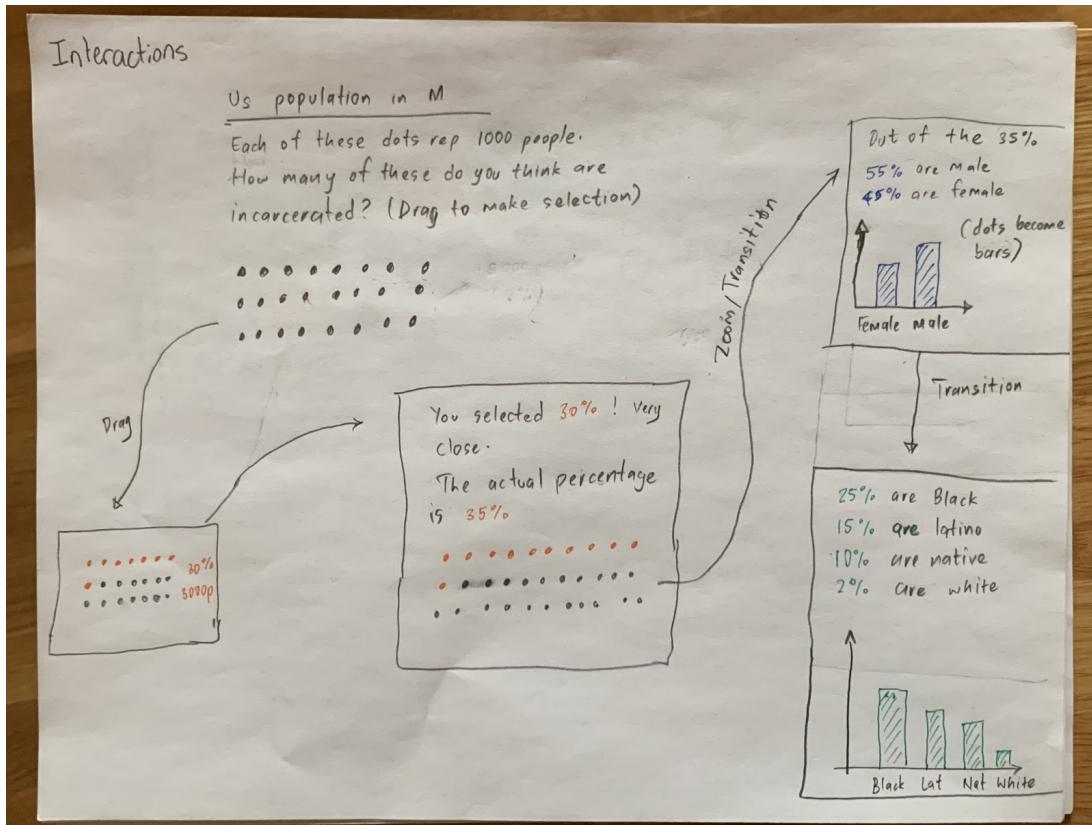


### Interactions:

Hover - A tooltip with additional information about a data point is shown when the user hovers above the point.

Filtering - The user can click the legend to highlight related points in the graph and grey out unrelated areas.

## Interaction Storyboard



This shall be the first interaction a user encounters on our website. Its primary role is to make the user aware of their own assumptions about the incarceration rates in the US.

The user is prompted to make a selection of how many individuals they think are currently incarcerated in the US. They can do this by clicking and dragging over their preferred number of dots. The selected dots shall change color to yellow. We shall then compare their selection with the actual data and show it to them in a new dot matrix. Any erroneous selections(points that they should or should not have selected) shall be colored red, to show how off they were from the mark. The dot-matrix shall then transition to bar charts showing the proportion of different groups in the incarcerated lot i.e genders and races. This data shall be presented alongside US population data to show any uneven distribution in the incarceration rates and population for each of these groups.

This will also be tested with racial breakdown of schools in boston- showcasing how the groups that are in the minority (black etc.) make up the majority of the population of students who are suspended, expelled etc.

## STORY CONTENT

Mass incarceration is one of the most jarring realities of the United States--the US holds one quarter of the world's incarcerated population, while it only constitutes 5% of the global population. The practice of imprisoning socioeconomic and racial minorities, as well as other specific groups, stems from a harsh and complex history of racial discrimination unique to the United States, whose infrastructure and economy originally emerged from slavery, and in which under the 13th Amendment, involuntary servitude is only permissible in the context of jail. This history has transformed into an equally complex modern reality, in which demographics are jailed in varying patterns, for varying reasons, across different regions of the US. These visualizations aim to expose trends in incarceration via a number of those factors, and then more specifically address the role of educational institutions and their punitive systems in feeding the prison-industrial complex.

“[Prison] relieves us of the responsibility of seriously engaging with the problems of our society, especially those produced by racism and, increasingly, global capitalism.”

— Angela Y. Davis, *Are Prisons Obsolete?*

“When children attend schools that place a greater value on discipline and security than on knowledge and intellectual development, they are attending prep schools for prison”

— Angela Y. Davis, *Are Prisons Obsolete?*

“How are prisons supposed to produce stability through controlling what counts as crime? Four theories condense two and a quarter centuries of experience into conflicting and generally overlapping explanations for why societies decide they should lock people out by locking them in. Each theory, which has its intellectuals, practitioners, and critics, turns on one of four key concepts: retribution, deterrence, rehabilitation, or incapacitation. Let’s take them in turn. The shock of retribution—loss of liberty—supposedly keeps convicted persons from doing again, upon release, what sent them to prison. Retribution’s specter, deterrence, allegedly dissuades people who can project themselves into a convicted person’s jumpsuit from doing what might result in lost liberty.”

— Ruth Wilson Gilmore, *Golden Gulag: Prisons, Surplus, Crisis, and Opposition in Globalizing California*

In my view, prisons are partial geographical solutions to political economic crises, organized by the state, which is itself in crisis. Crisis means instability that can be fixed only through radical

measures, which include developing new relationships and new or renovated institutions out of what already exists.”

— Ruth Wilson Gilmore, *Golden Gulag: Prisons, Surplus, Crisis, and Opposition in Globalizing California*

“I don’t know how I survived it, but I did.”

— Korey Wise

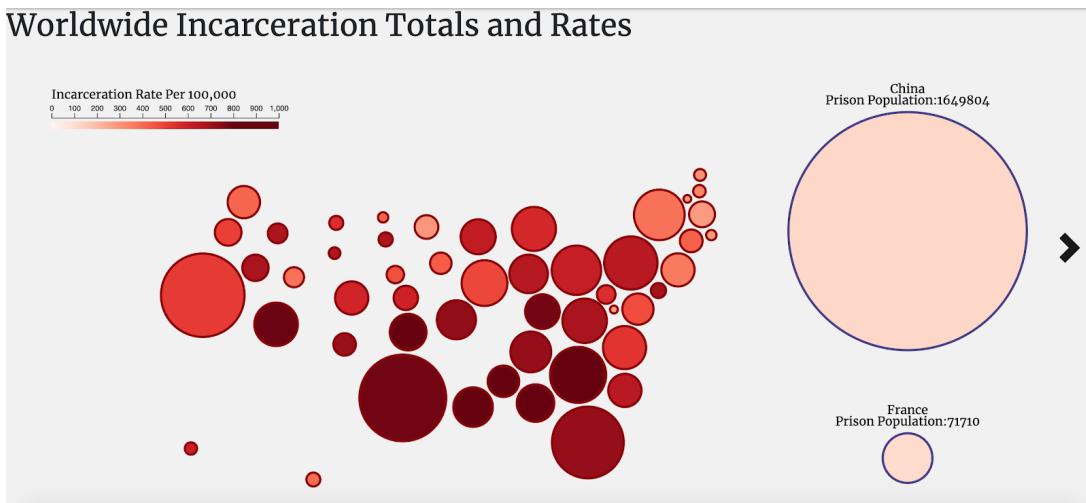
The ACLU describes the “school-to-prison pipeline” as a national trend in which youth are funneled out of schools into juvenile and criminal justice systems. Especially affected are students with learning disabilities, and children from minority and disadvantaged groups, including racial minorities and children who grew up in poverty. These scatterplots demonstrate the correlations between county-wide incarceration rates and disability, and incarceration rates and punitive or disciplinary methods used in county schools, including suspension, expulsion, and corporal punishment. These graphs additionally expose variation in strength of correlation between races [WILL HAVE DIFFERENT/ADDITIONAL TEXT AS THESE GRAPHS ARE COMPLETE AND WE KNOW WHAT THE DATA SHOWS]

V2 Prototype:

Cartogram:

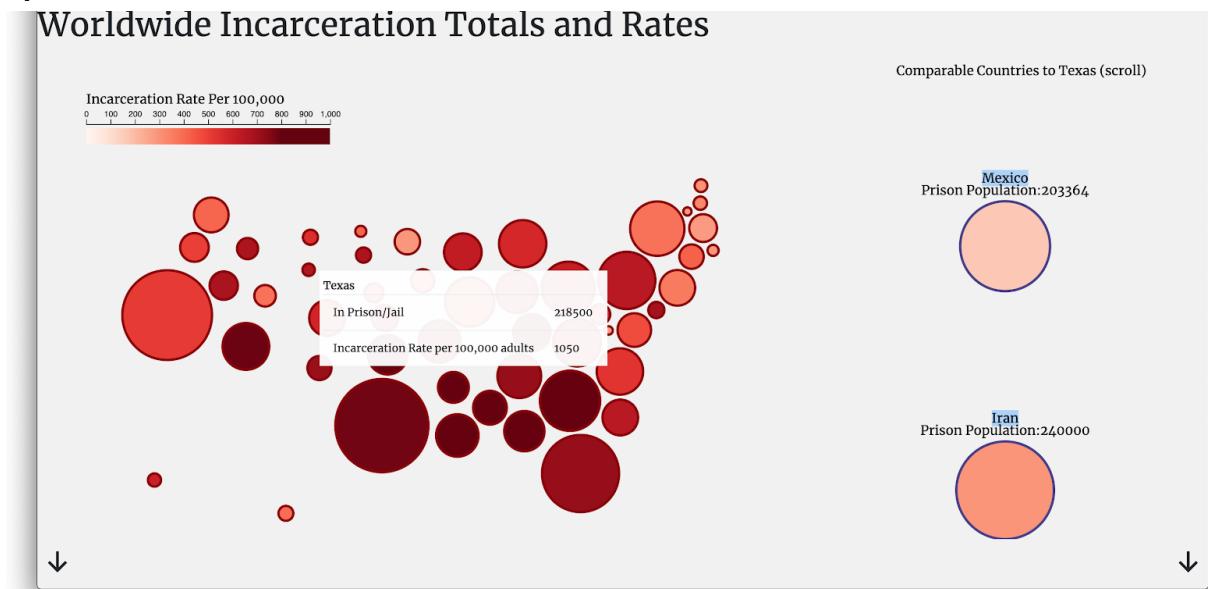
Implemented with help from <https://bl.ocks.org/veltman/938ea2d0ef98c02633bec15d6fb3a177>

Size corresponds to the number of incarcerated,  
Worldwide Incarceration Totals and Rates



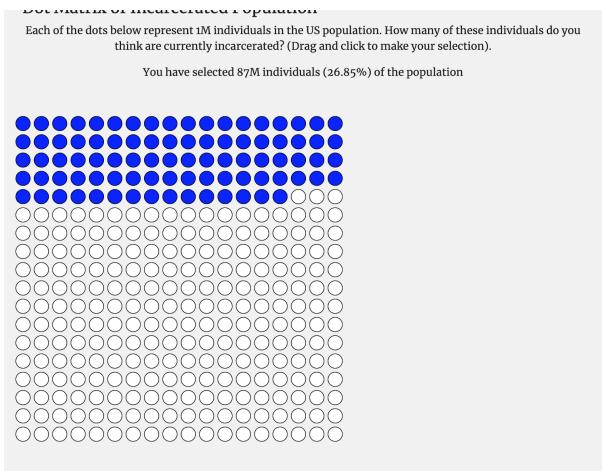
Working on making the left list scrollable and updated depending on your hover

Update:



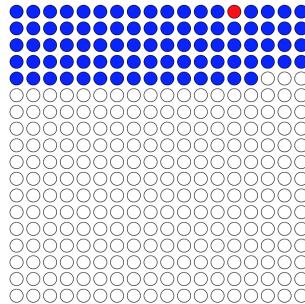
## Dot Matrix:

The first matrix is interactive and allows the user to make a selection. The second matrix is read only and appears after a selection is made to compare the users guess to the actual values.



Thanks for selecting. Here's how you did:

- You have selected 26.85% of the population(87M individuals)
- You overestimated the real value by 22.45%. In reality, about 4.4% of the population(14M) in the US are currently incarcerated.(The actual number is shown as a red dot)



# TESTING

## Dot Matrix

First page (dots) shouldn't have as much vis  
Title - need some kinda title, explanation

I think you could have more of a legend instead of writing it down as text

I like the percentage - it's good that it changes

Not quite sure what the red means - color scale...

Oh this is the real value, interesting (probably would help for all the dots up to the red to be red)

Having some kinda pointer or text above the circle could be interesting

Bridge gap between color, text and label

Have this bigger font at top

Not sure why there are so many dots when the real value is so close to top

It would be interesting to have different inputs

Not really sure why you chose red and blue, could have better color choices in that one

## Cartogram

Incarceration totals and rates:

It would be great to have a map

It would be great to have map or drop down menu / selector -- see one in comparison to another

Whereas it's not just text above the dot

Or some sort of barchart accompanying it

Circles make it hard to compare proportions

Colors are hard to compare if they're so far apart

What's the alignment of circles

Want to see the actual line

Found the right arrow

Maybe having the bubbles overlayed over an actual map - she's not sure it's possible

Why is it changing

Scroll where? Kinda confusing

This is really cool - took a while to notice what the other comparable countries were

You have some room in corner, could have areas to compare other countries together

### **Bar charts**

These are a lot of dropdowns

Why did we chose these specific 7

Maybe say something about why u chose those

Potentially have another thing

Suspensions and expulsion are separate ideas

2 side by side comparison graphs

I like taht ur able to sort

Transition with labels?

Make sure you label the axes

### **Scatter**

Seems like the axes cutoff the datapoints

Kinda hard to click on this

Wouldn't expect to have to click

Need interactive feature -- tooltip or sorting would be beneficial

Scatterplot

Shows correlation but its also not very clear

Potentially think about doing something other than a scatter plot

Trendline potentially

Its not super clear how steep correlation is

Have a lot of text there too -- take out some making more visualization per text on the page

The blue and red colors aren't the best in matix

I like the grey

Need more images

Have one really big image at top and never again

Need another image

## **Shruthi's notes**

PAGE 2

Have a barchart accompanying

Circles make the comparison between country and state

Colors are hard to compare as well

Having the circles representing states overlaid on an actual map

Unsure where to scroll on the side

Show scroll bar (Jess couldn't figure out how to scroll)

This part is cool though!

There's some room--so including another vis where you can compare the countries side by side?

PAGE 3

There are LOTS in the dropdown menu

Say something to explain why we chose these specific 7 measures

Maybe another slide, because suspension and expulsion are very different types of punishment

Sorting is great

Label axes + make sure scaling is clean

PAGE 4

Make sure none of the data points are cut off,

Fix this whole vis

Interactive tooltip (maybe click/hover for county) or sorting

Correlation unclear

Maybe do something other than a scatterplot

How steep is the correlation? Everything is bunched toward the lower x and y

Too much text on this page

More vis per text

**+ images**

Blue and red color scheme doesn't make much sense + color cohesion

Grey is good

## PRESENTATION 2 FEEDBACK

- Scroll bar
  - overflow
- Grey out dots + add trendline to scatterplot
- Big fan of the cartogram
- Alaska and Hawaii are awkwardly placed so add a label for them
- Adding more context to the tooltip per state / more facts

## Takeaways and To-Dos

Make the circle initially as a bar graphs

Make scroll bar visible

Map should come after scatter

Scatter should have bullet points next to it of quick facts, no paragraphs

Make it clearer on dots why there are so many dots

Need next arrows

Make map bubble thing click focused - cant hover on DC and then scroll on the side

## Overall Site Design

- Fix the scrolling technique--it would move through pages unintentionally in its current design, and the scrolling happened chunkily
- Make the color scheme consistent
- Include dots on L/R side as a kind of table of contents to navigate through
- Make sure that it snaps into same place each time and that each one doesn't overflow - sometimes it would get cut-off
- Scroll fix on bubble chart next to cartogram
- Add images, and intersperse visualizations and text better (overall format fixes)

## Matrix Interaction Vis

- Redo the first view
- Reduce the number of dots (because the actual answer is so low)
- Make it all one page (instead of snapping to a new vis of the same kind, updating the current vis)
- Change color scheme
- Have 3 separate step-throughs -- one overall, one race based, one school based
  - Do all in sequence and then show all results at once
- Callout what the red dot is ("Your guess here") with an arrow or some kind of highlighting mechanism
  - Make all dots leading up to the red dot also red
- Center + make prettier
- Add text to side

## Cartogram

- Make it click rather than mouseover (can't scroll with a middle state selected by hover)
- Make scroll bar visible on right
- Start with bar chart instead of the big circle representing USA--or just guide them through on how to get the USA bubble to expand
  - Or start with the world as a bubble
- Make button on right that transitions more obvious - or make it so you can click on usa bubble
- Bubble legend -- make it clear that this is the absolute magnitude of incarcerations, not rate (which is the color)
- Make it clear that the right side is comparing by magnitude of # ppl incarcerated
- Make the numbers with commas
- Start with world
- On next do USA with stuff on right
- Have scroll greyed out
- Then have this current view

## Scatterplot

- Connection between race incarcerated and race suspended?
- Scatterplot should show more than just suspension -- maybe a dropdown similar to that in the barchart
- Popping statistics to show how race disparities between punishment in school match up to incarceration rate race disparities (transition to barchart)
- Show trend lines and grey out dots

## Bar + Chloropleth : SHOWS DEEPER DIVE INTO SCHOOL STATS

- Add chloropleth under the bar that's connected to the filter dropdown
- Showing rates overall in each state

# CS 171 Project Presentations

(Give the completed form to the team you gave feedback on. They will have to scan it in and attach it to their final submission.)

**Your Names:** John, Michael, Thomas, Sid

**Your E-mail:** beaumaster@hbs.edu

Name of group you evaluated:

School to prison pipeline

What is good about the group's visualization?

The map visualization is really neat and a good representation with the choice of using per capita.

What could be improved?

Adding contextual notes to the map could be helpful

Is the message clear? What is the message?

The message is overall clear but could use some transitions.

# CS 171 Project Presentations

(Give the completed form to the team you gave feedback on. They will have to scan it in and attach it to their final submission.)

**Your Names:** Gunnar Plunkett

**Your E-mail:** Gplunkett@college.harvard.edu

---

**Name of group you evaluated:**

"Visualizing Mass Incarceration"

**What is good about the group's visualization?**

The message and the "so what" of the visualization are very clear.

The visualizations are engaging and interactive, especially where you guess the overall incarceration rate.

**What could be improved?**

It would be good to have:

- a more consistent color scheme

- more easily readable/interpretable shapes (although it's pretty fine already)

**Is the message clear? What is the message?**

The message is clear:

~~The~~ Incarceration and the school-prison pipeline have a major especially on people of color.

# FINAL CHAPTER

**Overview and Motivation:** Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

**Related Work:** Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

The ACLU had a number of pages that provided high-level overviews of both mass incarceration in the United States and the school to prison pipeline which were particularly influential. Of particular note was the following infographic

:<https://www.aclu.org/issues/juvenile-justice/school-prison-pipeline/school-prison-pipeline-infographic>. We were motivated to create a more detailed and nuanced visualizations than those in this infographic. Our dot matrix was influenced by the Syria visualized project, while the various cartographs that were displayed in class following the trump “Impeach This” were inspiration for the Dorling cartogram.

**Questions:** What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

1. How many people are in jails and prisons in the United States?
2. How does this number compare to other nations?
3. How does race affect this number?
4. How are schools related to this incarceration total?
5. How does race impact the way students are disciplined in schools?

We initially thought it would be easy to correlate schooling and incarceration, but this ended up being a lot harder to quantify than anticipated. We were particularly taken by the disparity in incarceration rate between the US and other comparable nations, thus spurring the state cartogram and comparable nations visualization.

**Data: Source, scraping method, cleanup, etc.**

We used two primary data sources:

- Civil Rights Data Collection (CRDC) comprehensive school dataset
- Vera Institute of Justice Incarceration dataset

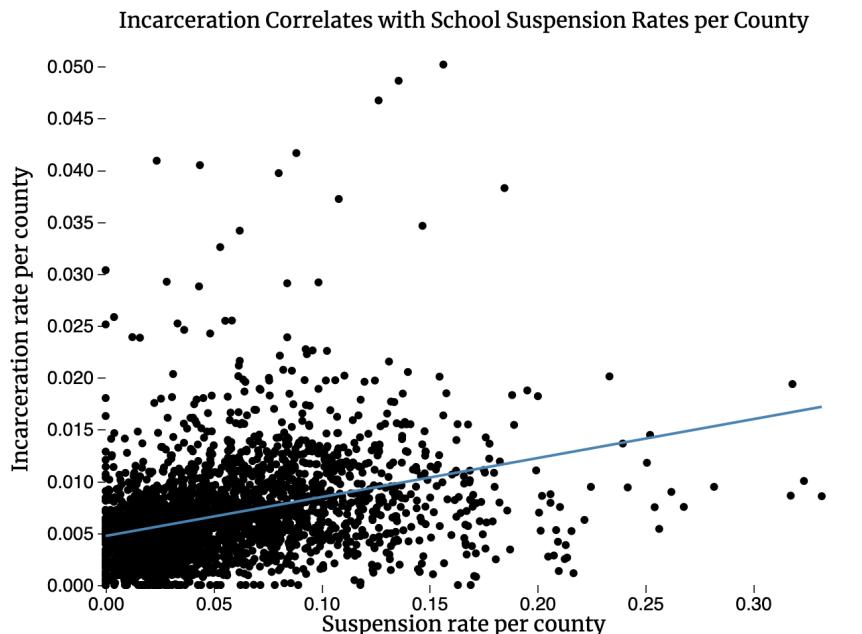
Pandas was used to clean these datasets: removing null data, filtering relevant columns to be used in our visualizations and performing merge/joins.

We used various levels of aggregation for state and county level. In order to match these two datasets we used ED.gov’s school district dataset which allowed for grouping of school districts into counties which was the granularity used in the Vera set. This allowed for county-level data with both school statistics and incarceration. We converted this aggregated data into json format to be more easily read into D3 for processing.

We then selected for certain statistics (columns of the Vera or CRDC dataset) and included them in separate spreadsheets to facilitate loading in data in our js files.

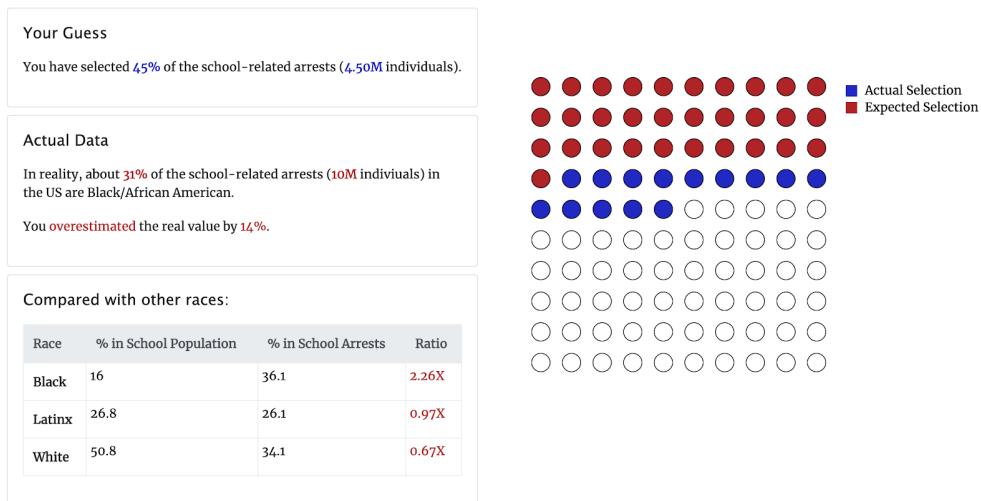
**Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?**

We initially used Tableau to look at our datasets. We implemented simple visualizations like bar charts, line charts, and scatterplots to identify trends and correlations in the data such as the direct correlation between school-related arrests and incarceration rates in various counties in the US. This led to us using a scatterplot as one of our visualizations to show this correlation.



We also noted a mismatch between the percentage of black/african american adults in the US population(12%) and the percentage of the same group among the incarcerated population(33%). The same mismatch was also present when we compared the percentage of black kids in school() vs the percentage of school arrests involving black kids(36%). This inspired us to create the matrix visualizations where we'd ask users to guess these percentages and then later find out what the true values were. This would allow them to identify these discrepancies in the distribution of the races among the incarcerated population and also among school-related arrests as well.

Thanks for selecting. Here's how you did:



We also noted that the US had a much higher incarceration rate than all other countries. In fact, some of the states in the US have a larger incarcerated population than entire countries that have much bigger size and population. We highlighted this fact via the cartogram and world map. These two visualizations allow users to compare incarceration rates for many countries and even at first glance it is clear that the United States is an outlier, and not in a good way.

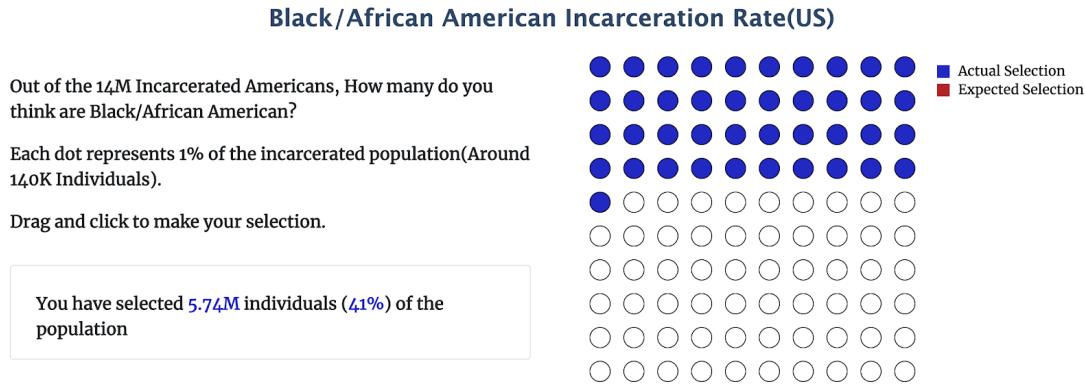
**Design Evolution: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?**

We initially considered implementing a dot matrix which transitioned to bar charts, similar to what the Syria group did. However, we soon realized that we had a lot of different data sets, and instead of delving deep into one individual set, we would rather move from data set to data set, visualizing the most important stats from each. Initially, our cartogram only displayed the United states, and other nations were displayed as bubbles to the right of the visualization. However, after student and TF feedback, we decided to make two distinct cartograms, one of the US and one of the world, with a bar chart on the right to compare states to nations. World and US maps were added to the cartogram in the background to make it more readable, and the force simulation which determines the dots' locations was updated to keep dots as close as possible to the state boundaries.

We established several visual motifs which are consistent throughout the project. The repeated use of circles is of particular note, though we made sure to not use pie charts. After the feedback of the TF, we decided that while these circles are important, it was more important to use bars to compare values as opposed to having a scrollable list of circles.

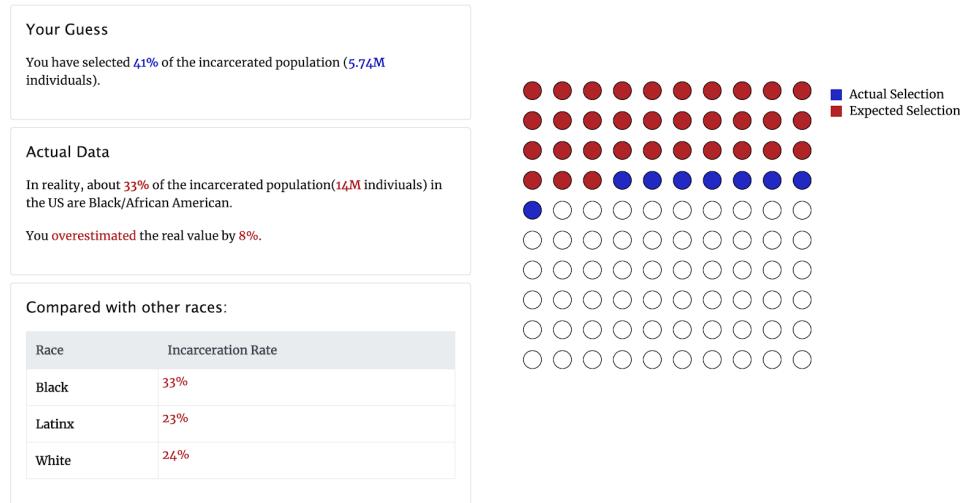
**Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.**

The first interactive visualizations are the dot matrices: the user is asked a question relating to one of the main datasets and is allowed to guess what the correct answer may be by interacting with the matrix .

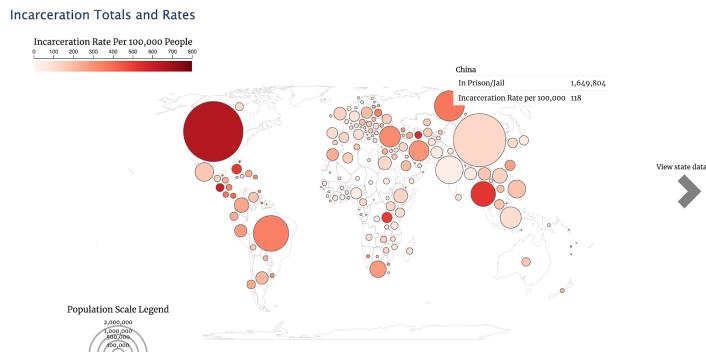


After selection, the user is shown the real value and how far off their guess was. This is also accompanied by a table of supporting data related to the question. The purpose of this interaction is to make the user aware of their own misconceptions before they delve into the dataset in later visualizations and also to add a shock factor since often times their guess is expected to be very far from the actual values.

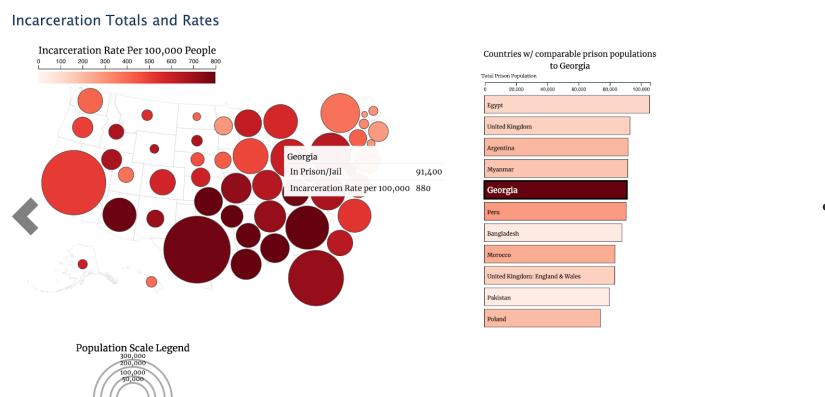
Thanks for selecting. Here's how you did:



The cartogram offerers several key points on interactivity. The initial world view has tooltips which show the total incarcerated and incarceration rate.



By pressing next, this view transitions into a view of the United States, which offers similar tooltips. When one clicks on one of the states, a bar chart of comparable nations appears on the right, which then changes depending on the selected state.



At any point, the user can transition back to the previous view.

The interactive barchart and choropleth visualizations (on the same page) have linked views. Clicking on any one of the bars in the barchart corresponding to a race allows the user to see a

The scatterplot has interactive hover tooltips for each point for users to see the actual incarcerated and suspended population percentages, which may be hard to discern with the axis labels, and that further show county names for each point.

**Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?**

The most notable takeaway from this data is that the United States is an exception with regards to incarceration rate and schooling. Those who have grown up within American public schools in the last two decades may think that police presence in schools is the status quo, and it is not, compared to both other nations, and the US pre-Columbine and 9-11. Thus, in comparing US states to other nations, we hoped to make a point as to how outrageous the US incarceration rate is. In general, our visualizations focused on the most marked stats and comparisons we came across. Our project could be improved by more effectively conveying the connection between overall incarceration and the school to prison pipeline. In retrospect, our topic was

perhaps a bit too ambitious, and would have been a bit more effective if we either focused on just the school to prison pipeline or simply overall incarceration. We were in part limited by the datasets that we found and the correlations we could show between them (some of the visualizations we generated did not support our points, and so we chose not to include them.) Nonetheless, in first showing the gravity of this topic, and then narrowing in on the school-to-prison pipeline as a facet of this overarching topic, we constructed a story that shows why we should care about these topics as interconnected and correlational.

## Script

Our website visualizes mass incarceration globally and in the United States, and then more specifically addresses how public schools are related to mass incarceration, given the phenomenon of the “school-to-prison pipeline,” where students are funnelled into jails. In this first visualization, the user guesses the incarceration rate of the US