

What Does Sex Look Like?

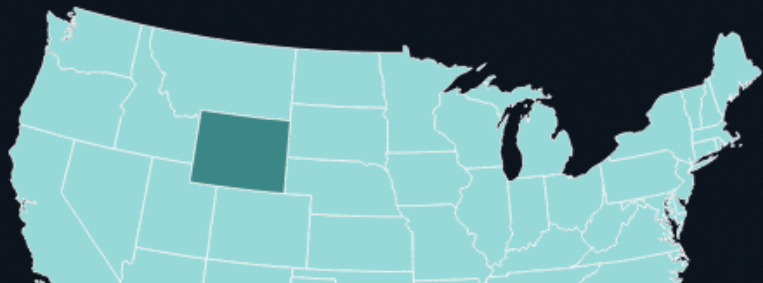
cs171 final project Process Book

Vivian Leung, Lucas Freitas, and Charlie Lovett

Goals and Motivations

Through “What Does Sex Look Like,” we seek to understand and make clear a current portrait of the United States and its position relative to sex, specifically STI- contraction rates, teen-pregnancy rates, and trending porn interests. While not attempting to conflate many of these statistics, “What Does Sex Look Like” strives to provide for the user an understanding of their own state’s and its constituents’ current risk and preferences.

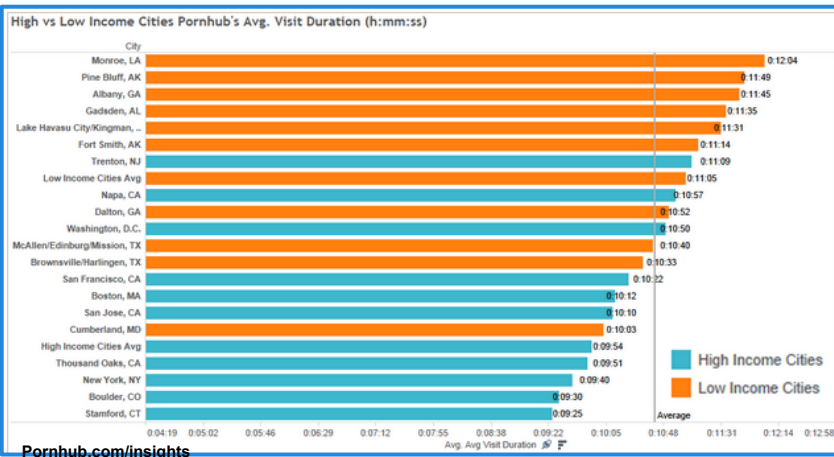
Because sex and sexual health is so multifaceted, this data visualization strives to consolidate these important sex-related statistics that are available, but difficult accessed, in a clean and user-friendly way to better inform healthy and smart choices regarding the user’s own sexual practices.



The Project Proposal

Original Direction

We were incredibly interested by the static graphs and charts from the website pornhub.com/insights, whose representations were stunning albeit non-interactive. We hoped to gain access to some of their data to better represent it. Do trends in porn consumption have parallels to a state's sexual health practice and statistics?



Problems

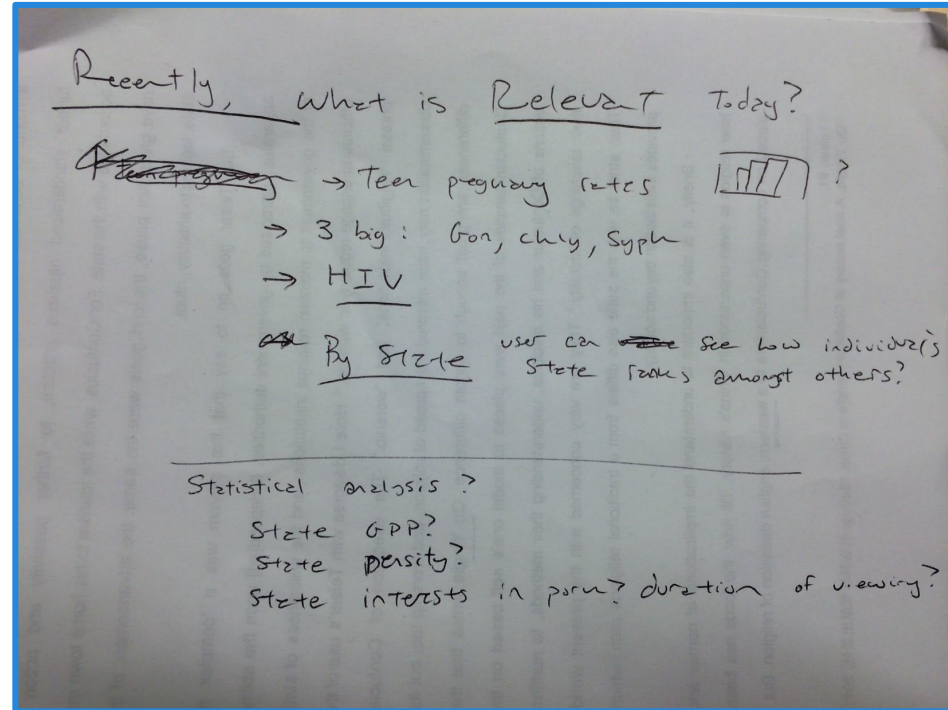
Pornhub was reluctant to share some of their data with us. Without this, we turned our focus away from a porn-centric visualization, and looked instead to how we could visualize sexual health, rather than sexual urges.



http://3.bp.blogspot.com/-x0-xU52WWq0/TmdcrjCnEhI/AAAAAAAAAd0/R6_aCK7V0ss/s1600/437046r6s71ehus.jpg

A New Direction

As our initial, potentially hormonal, excitement over working with pornhub faded, a newer passion arose: **a desire to better consolidate relevant, and yet under-advertised/ accessed, sexual health statistics.** Inspired by the Center for Disease Controls massive database on sexually transmitted infections and sexual health trends, we set our eyes on re-imagining sexual health for a regular internet citizen. With all of these studies, statistics, insights, feelings and misconceptions about sexual health available, how and what would we communicate to this user?



Questions We Sought to Address

What STI's are still prevalent to teens today?

Pregnancy rates?

Do these factors have anything to do with state GDP? Population density?

Sex isn't just disease, and "normal" sex is a myth? How can we supplement this data with positive sexual facts?

Obtaining our Data

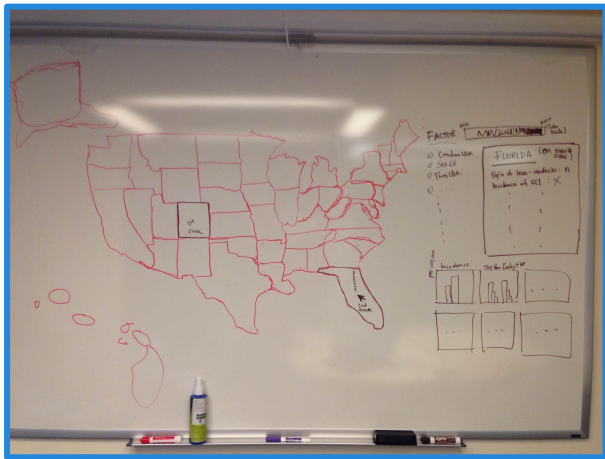
The last US census taken by the CDC released readily for consumption was in 2010. Because of this **we knew our data set would be unaltered**; we simply downloaded the relevant surveys and findings from the CDC related to teen pregnancy and rates of infection for Gonorrhea, Syphilis, Chlamydia, and HIV and converted them to .csv files in excel and google spreadsheet, were they not .csv to begin with.

We sorted these .csv files by “state,” “cases” (number of reports), and “rate_per_pop” (or rate per population of 100,000).

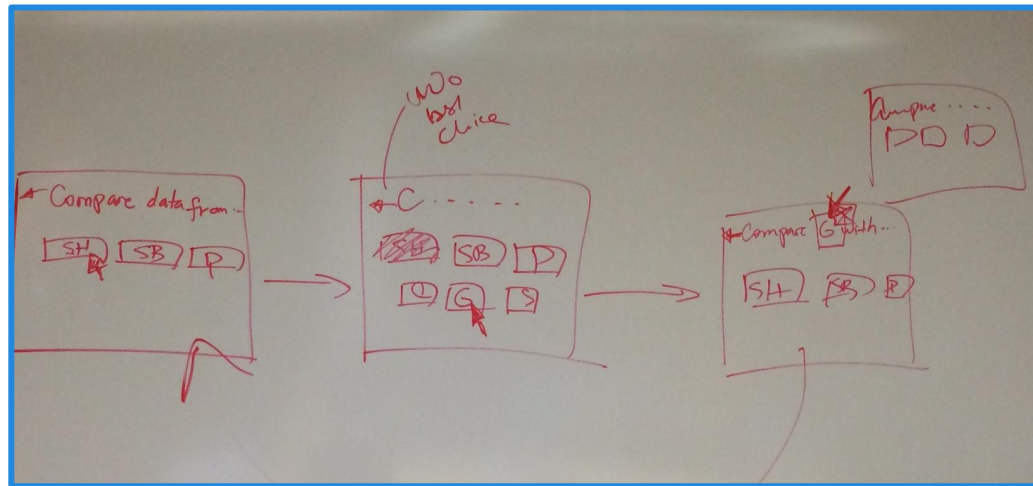
Pornhub did have one csv file they were willing to share with us. **It contained the top three searched terms and averaged duration of use on the website per state.** We opted to include this data as well. A state that may be very pro-life still searches “creampie” just as regularly as most blue states. While not desiring to draw comparisons between a particular search term and rates of contraction, the pornhub data served **to balance the, while informative, dark nature of STI statistics, with funny and leveling sex facts.**

We sorted this file by genre (“Teen” “Ebony” etc.), followed by the three states who searched the most. This data is scaled simply, top searched state 1,2,3. (Further data manipulation was done, explained in full later).

Visualizing the Data

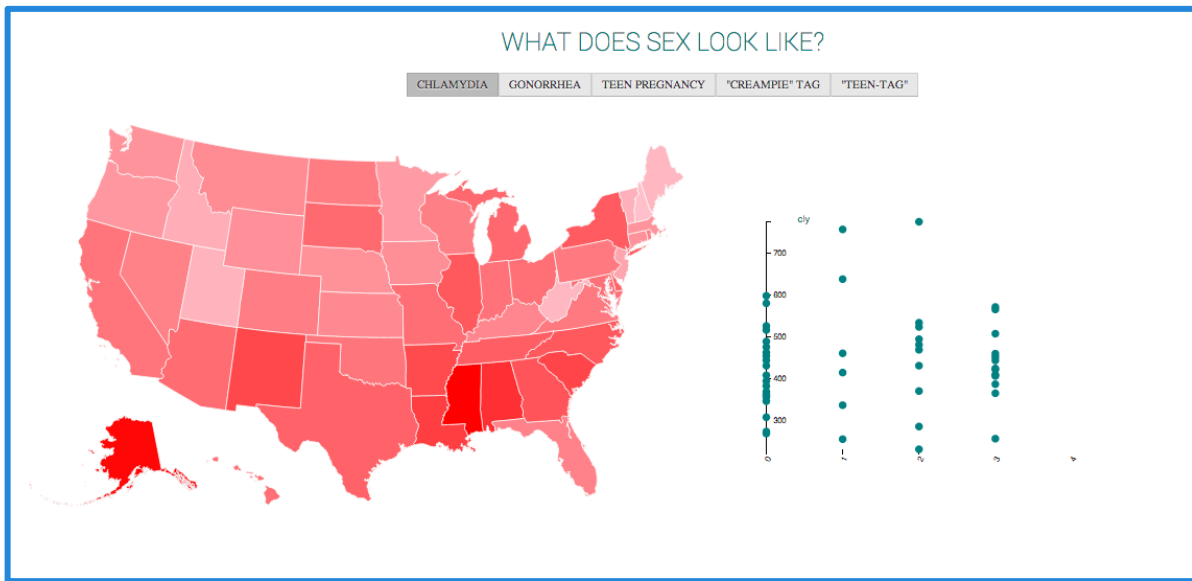


Statemap was an obvious choice as an introductory viz. By **selecting a state**, the user would be given access to **state-relative information across all factors**. We initially envisioned this would be portrayed across multiple graphs of some sort (explained below)



An alternative method of search would be nationwide. **Selecting a variable would display dynamically the variation across all states** (shown by color gradients). Multiple factors could be chosen simultaneously and layered on the map.

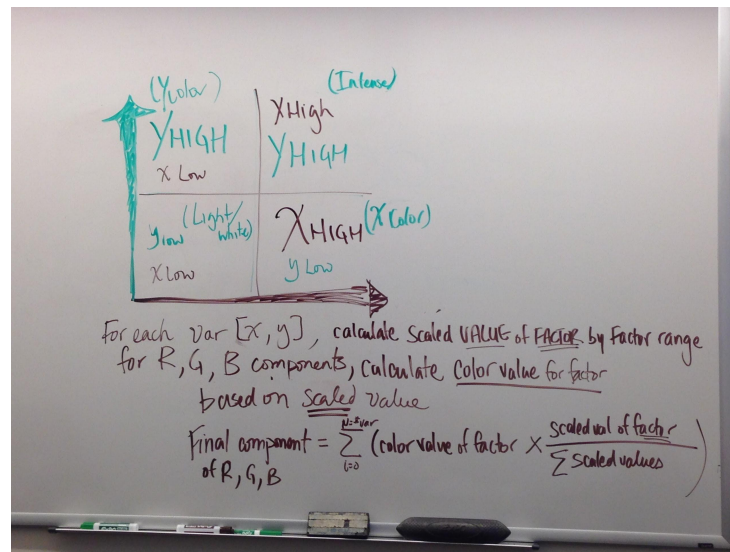
Visualizing the Data: Prototype



By the first milestone, we had a draft of this initial plan. Using sample tags from the pornhub csv file, two STIs, and the teen pregnancy statistics, we created buttons that allowed the user to dynamically visualize his or her selection on the state map, as well as graph their nationwide prevalence alongside multiple selected options.

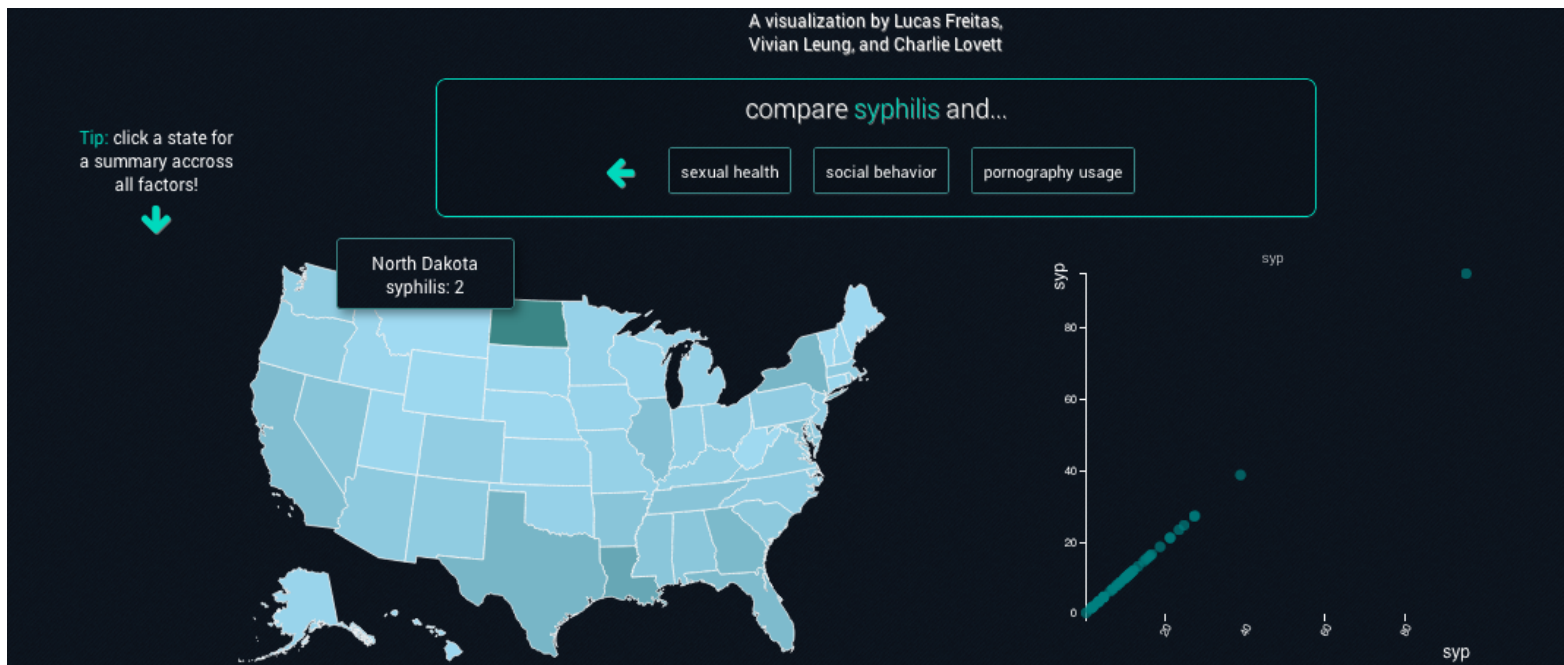
Visualizing the Data: Layering Map

Originally, we were hoping layer up to three data sets on mapviz on top of one another when more than one factor was chosen. However, we realized that three factors layered would imply too much potential causation amongst data sets that were, while at times certainly correlative or followed a similar pattern, potentially harmful to conflate (while gonorrhea and chlamydia, two commonly simultaneously contracted STIs shared similar states, we would hate for the user to think any combination of three porn searches; social behaviors; or disease prevalence rates would have great relation. So we stuck with two.



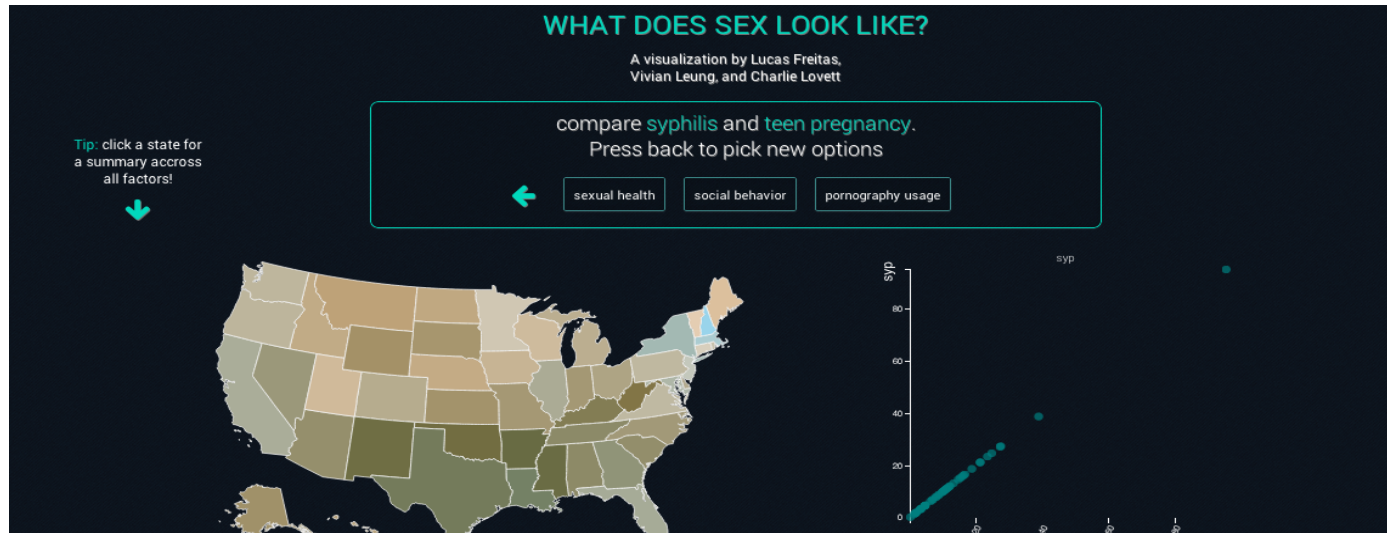
The image on the left displays the formula we used to **calculate the hue and color of combining two datasets with different ranges on top of one another**. This specific scale ensured that states with greater X than Y was more xcolor, the reverse was more ycolor, and a combination was a mix, factoring in hue to denote strong presences of both (dark) or not (light).

Visualizing the Data: Layering Map



Selecting one, syphilis, shows a dynamic blue graph, where shades of blue shows higher prevalence of syphilis.

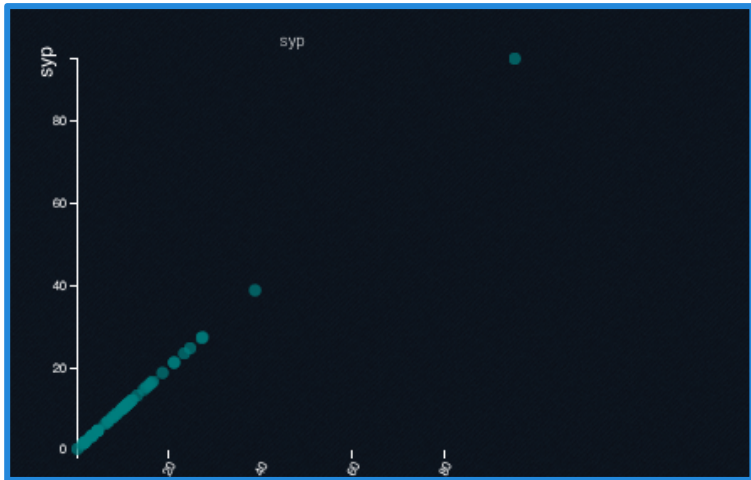
Visualizing the Data: Layering Map



Clicking now on teen pregnancy, we see how the graph has changed, blue layered with a red orange. States that are bluer, have higher rates of syphilis per 100,000 people, while the pregnancy rates of their state per 1000 teens is lower. Conversely, more red means more pregnant teens.

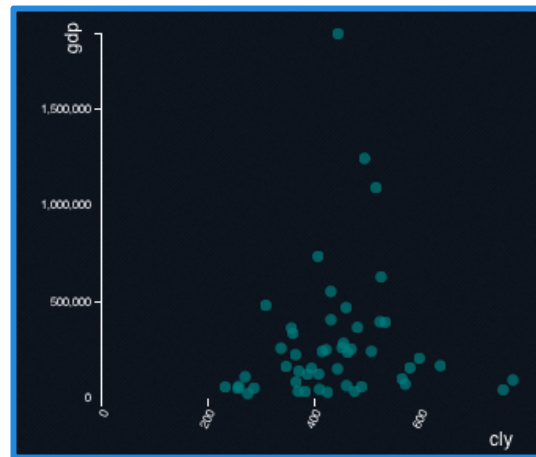
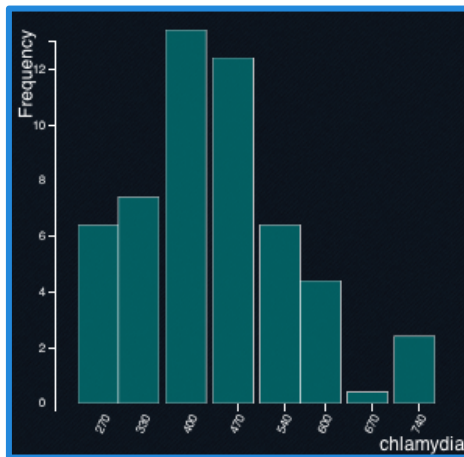
Green denotes crossover. Lighter states means that they have overall less rates of both, while darker states have higher rates of both.

Visualizing the Data: Supplemental

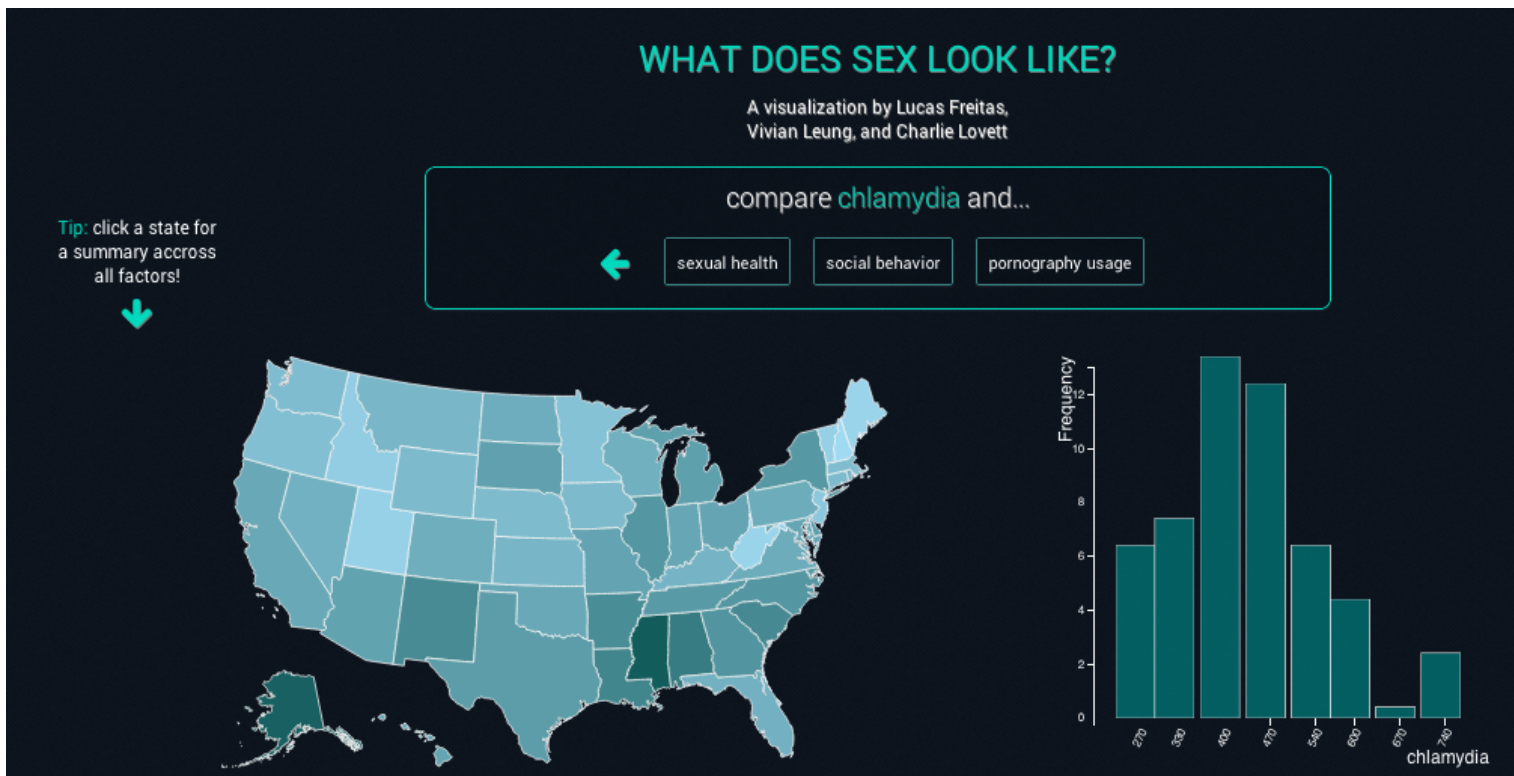


This graph as shown a slide above was a prototype. We originally planned to plot points as each new factor was added by the user.

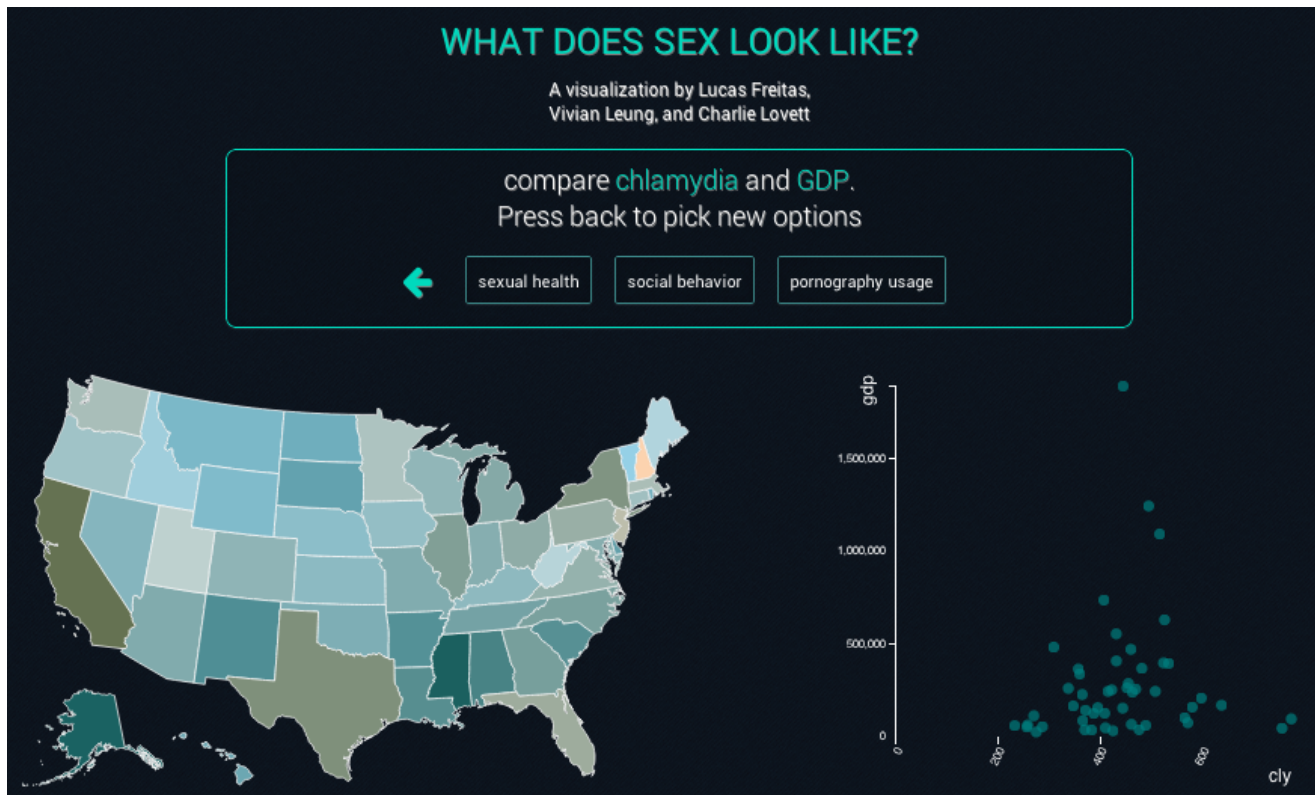
However, as different factors change the way we view static data, we too thought our graphs should reflex this alteration. When only **one factor is selected**, the **graph to the side is displayed as histogram**. It changes when a **second factor is added to an x-y plot**, so the viewer can see, like the mapviz discussed previously, whether there is a clearer trend.



Examples of the Front Page, One Factor



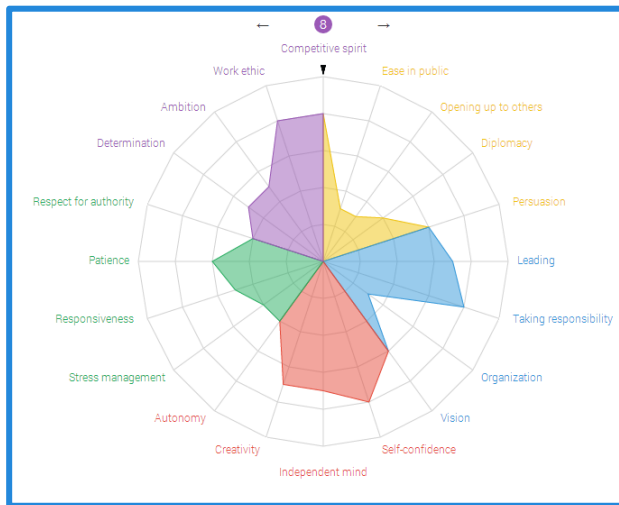
Examples of the Front Page: Two Factors



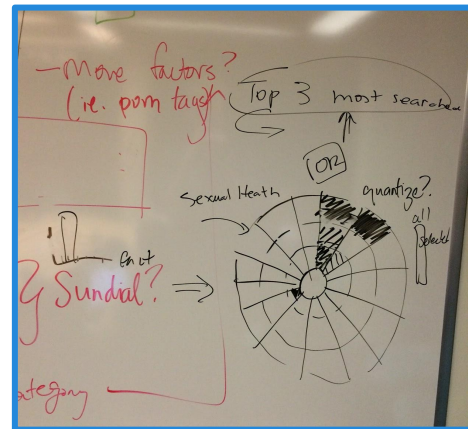
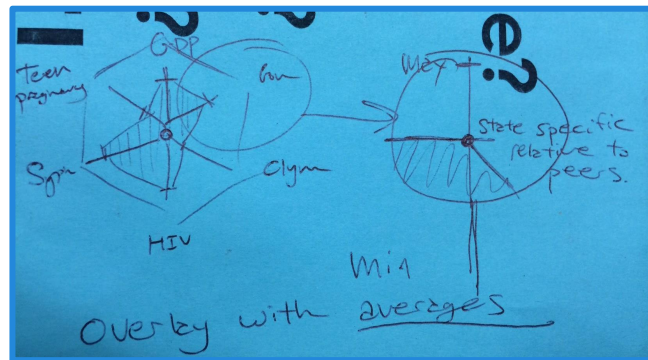
Visualizing the Data: Radar Chart

However, we were still conflicted about how best to communicate, across all factors, a state-specific visualization. Rather than creating multiple, individual visualizations on state-click, **we opted instead to do a radar chart**, inspired by the recent visualizations from talentoday.com. We chose to plot each factor on a ray.

Because each factor has its own scale, **they are weighted differently** to show the most compelling picture: rather than making all STIs values graphed as a percentage (which, if 8000 cases per 100,000 people, would be meaningless visually), we plotted them between their min and max values nationwide.

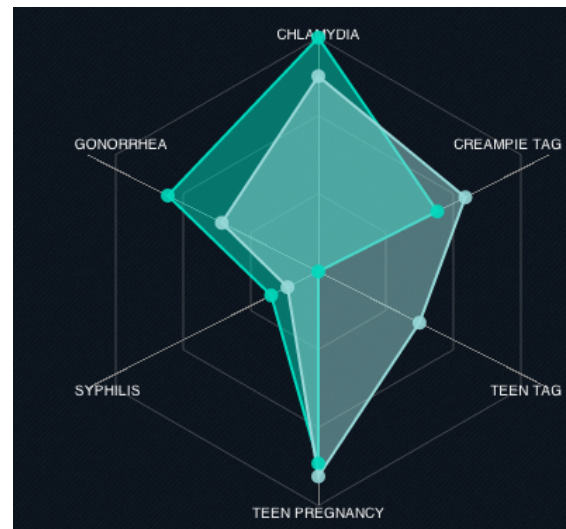
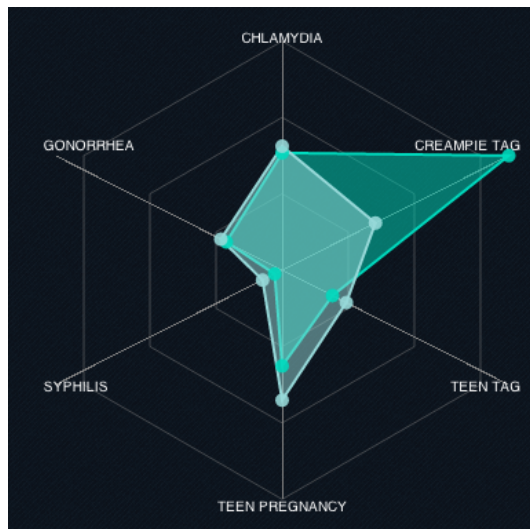


<http://foxyreign.com/2014/04/25/personality-test-by-talentoday-com/>



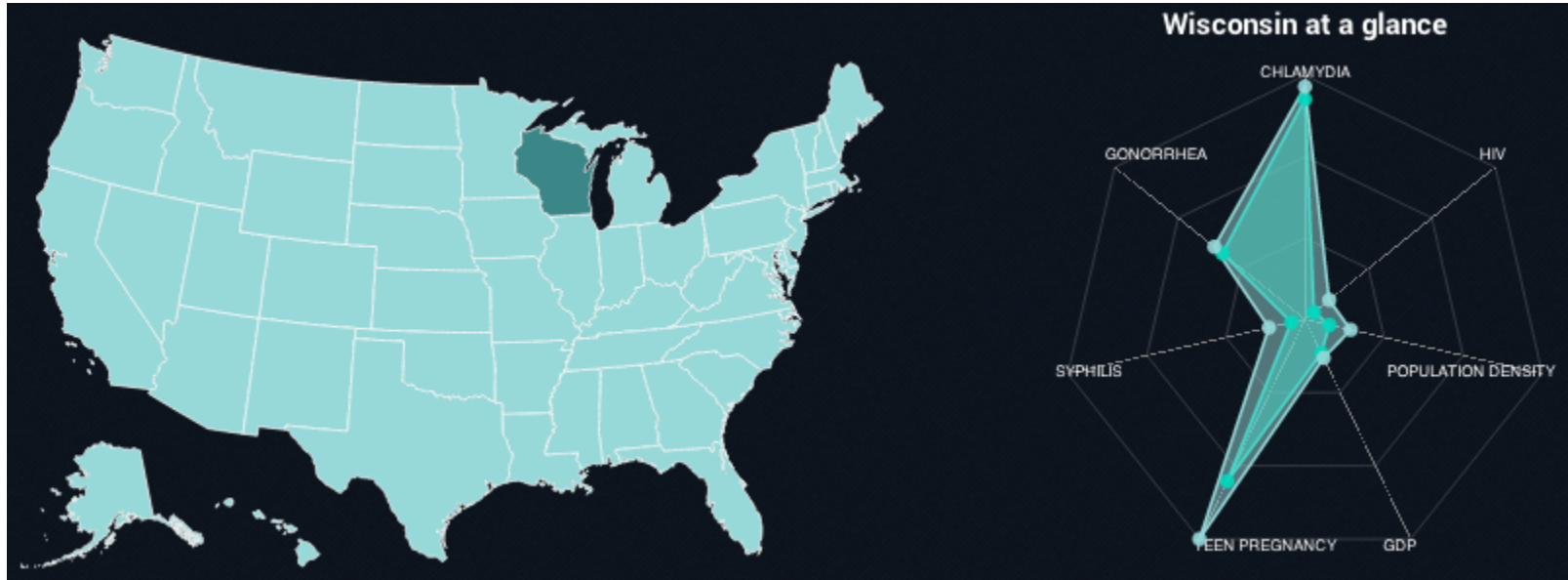
Visualizing the Data: Radar Chart

Borrowing from the radar-chart-d3 (alangrafu) repository, we began to plot our radar charts. Though not shown here, the lighter, white polygon is the national average graph, while the darker teal is state-specific. Scaling the rays was beneficial in creating a poignant nation-average radar chart to supplement the state-specific one. Additional **hover interactivity** was included for nodes to list their values.



In this earlier prototype version, the leftmost chart (Wisconsin), has most of its data points in line with the national average. Conversely, the rightmost (Illinois) has a larger prevalence of all STI's, though teenage pregnancy rates are below the national average. We scaled the chart such that larger discrepancies between national average and state are most apparent to the viewer.

Example of the State Specific Viz Option



Visualizing the Data: Porn Revelation

Working with the raw CDC data and its neatness made us revisit our raw pornhub dataset, pictured here as it was received, unchanged. Could altered to be more useful?

state	avg_time	1	2	3
Arizona	0:10:59	compilation	creampie	teen
Florida	0:10:55	creampie	teen	milf
Nevada	0:11:15	anita queen	asian	teen
Hawaii	0:11:53	asian	hawaii	creampie
California	0:11:18	asian	teen	massage
Washington	0:10:15	asian	teen	pov
Iowa	0:10:46	college	creampie	pov
New York	0:10:18	college	massage	milf
Connecticut	0:10:12	college	milf	pov
Minnesota	0:10:40	college	teen	creampie
Pennsylvania	0:10:26	college	teen	milf
Montana	0:10:34	compilation	anal	wife

The current layout, while interesting, was not useful enough for what we wanted to do with it. Rather, we envisioned an array of objects by genre containing each state and corresponding “rank” .

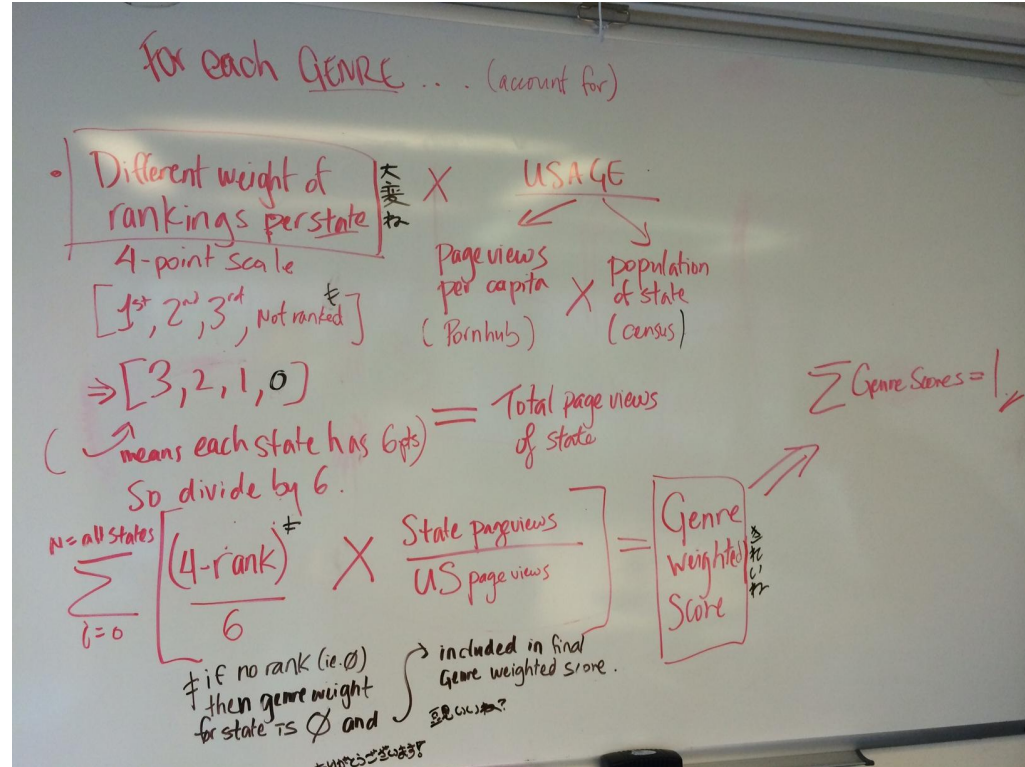
However, we wanted to accurately represent the popular genres of the United States given this data. And that’s when Vivian had a revelation: (next page)

Visualizing the Data: Porn Revelation

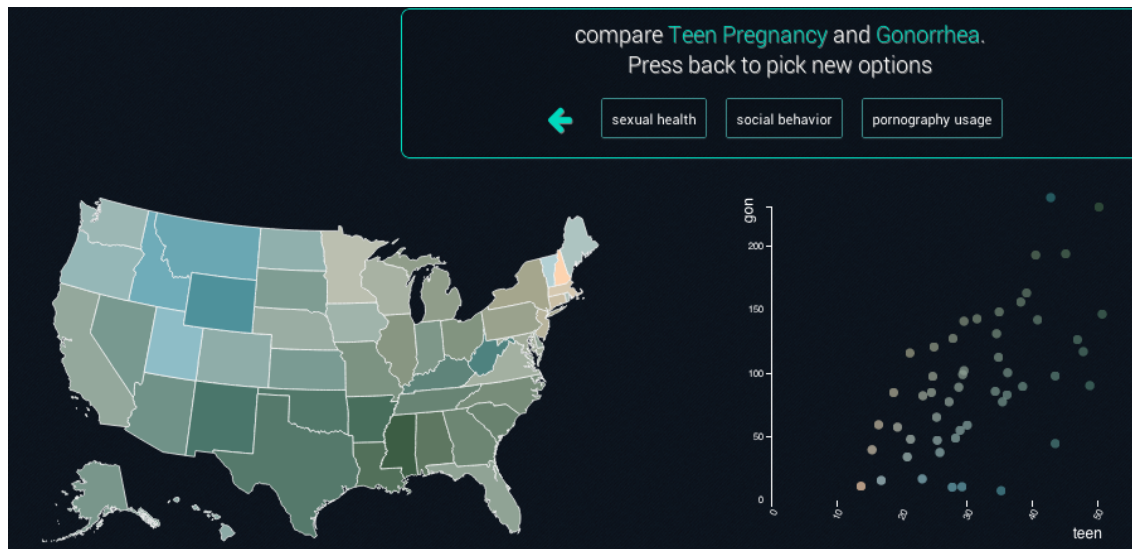
We wanted an **overall figure** that considered the different weight of rankings as well as the usage of the state. To calculate this, we utilized page views per capita for each state* and combining that from the US census data on population totals, we could normalize and accurately weigh these state preferences according to their own consumption. The formula devised on the right details how this figure was calculated and eventually entered into our revised dataset.

Per one year, pornhub is visited 63.2 billion times. Of this, 41.4 billion total pageviews from the United States

*(from pornhub.com/insights static visualizations)



Neato

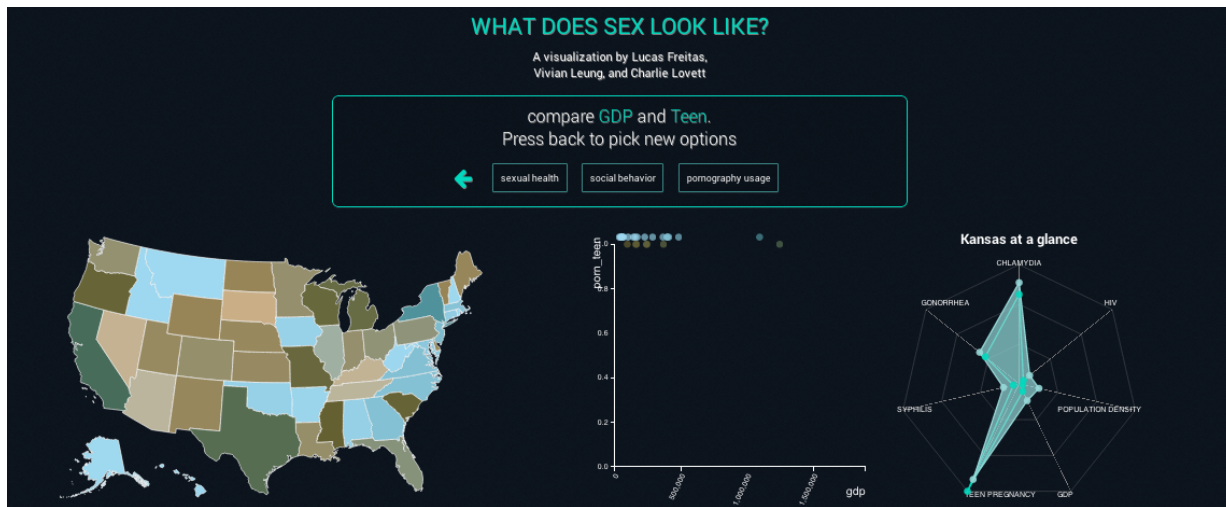


It was neat to see how STI rates by state were so highly correlated, as would be expected. (left)

Of course some more “exciting” trends could be seen, with rising teen pregnancy rates and Gonorrhea... (above)

Conclusions and Evaluations

We strove to provide small facets of the multifaceted body that is sexual practice and health in America. What was fascinating about our end product was not necessarily the trends that STIs shared, or that Teen Pregnancy and GDP. It was playing around with the different toggles and discovering outliers: a state we were prejudiced against that had almost no reported cases of Chlamydia, and whose favorite porn search term is “Teen”. By consolidating these many different factors we were able to see not only their intersections, but their divergence as well.



Sources

STI Data

Center of Disease Control, 2012, 2012 Sexually Transmitted Diseases Surveillance, last accessed Thursday, May 1st. <http://www.cdc.gov/std/stats12/default.htm>,

Pornhub Data

Pornhub, 2014, Pornhub Insights, last accessed Monday, April 28th.

<http://www.pornhub.com/insights/>

And a csv from pornhub, included in our data file. This csv file is also floating around on the web, so they didn't really send us much...

Teen Pregnancy

Office of Adolescent Health, 2012, Trends in Teen Pregnancy and Childbearing, last accessed Wednesday, April 30th.

<http://www.hhs.gov/ash/oah/adolescent-health-topics/reproductive-health/teen-pregnancy/trends.html>

Sources

Population

Wikipedia, 2014, List of U.S. States by Population Density, last accessed Monday, April 28th. http://en.wikipedia.org/wiki/List_of_U.S._states_by_population_density

Wikipedia, 2014, List of U.S. States by Population, last accessed Thursday, May 1st.
http://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_population

HIV

Henry J. Kaiser Family Foundation, 2010, Estimated Rates per 100000 of Adults and Adolescents Living with an HIV Diagnosis, last accessed Monday, April 28th. <http://kff.org/hivaids/state-indicator/estimated-rates-per-100000-of-adults-and-adolescents-living-with-an-hiv-diagnosis/#>

Henry J. Kaiser Family Foundation, 2010, Estimated Numbers of Adults and Adolescents Living with an HIV Diagnosis, last accessed Monday, April 28th. <http://kff.org/hivaids/state-indicator/estimated-numbers-of-adults-and-adolescents-living-with-an-hiv-diagnosis/>

Sources

GDP

The World Bank, 2014, Indicators, last accessed Monday, April 28th. <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

Wikipedia, 2014, List of U.S. States by GDP, last accessed Monday, April 28th. http://en.wikipedia.org/wiki/List_of_U.S._states_by_GDP

Libraries

Mike Bostock, 2014, D3, last accessed Thursday, May 1st. <http://d3js.org/d3.v3.min.js>

Mike Bostock, 2013, TOPOjson, last accessed Thursday, May 1st. <https://github.com/mbostock/topojson>

John Resig, 2014, jQuery, last accessed Thursday, May 1st. [js/jquery-1.11.0.min.js](http://jquery.com/jquery-1.11.0.min.js)

Alan Grafu, 2013, Radar Chart, last accessed Thursday, May 1st. <https://github.com/alandgrafu/radar-chart-d3> [alandgrafu](#)

Fonts

Dave Gandy, 2014, Font Awesome, last accessed Thursday, May 1st. <http://fontawesome.github.io/Font-Awesome/>

netdna.bootstrapcdn.com/font-awesome/4.0.3/css/font-awesome.css

<http://fonts.googleapis.com/css?family=Roboto:400,300,100,700> This google font. unclear how to cite?