CS/INFO 3300; INFO 5100
Project 2
In-class Demo Day on Friday, November 10
Final Report due 11:59pm Tuesday, November 14 (no slipdays)

In this project, you will create an **interactive data visualization**. Your visualization interface should use **dynamic elements** to show more information than can be presented in a single view. One common mantra for adding interactions is to follow the pattern: "overview first, zoom and/or filter, then details on demand." You are **required to integrate multiple data sources** for this project. Your project must be **responsive and performant** in our web browser, with no lengthy loading times or distracting latencies during interactions. Finally, you will write a **final report** on your design and implementation. You will turn in your code via CMS along with your final report. Groups are encouraged to create a **Cornell GitHub repository** for their project to make collaborating on code easier and add accountability. **All projects must obey a 50mB maximum file size limit for their final ZIP submission** (this is a mechanism to ensure that you scope datasets down to a level that will enable responsive interactions in a web browser). **Project visualizations can take many forms, including but not limited to: data dashboards, components of a data-driven user interface, interactive data journalism, quizzes/personalized pages that incorporate visualizations of complex data, and interactive research articles.**

This project is intentionally very open-ended. It is up to your team what data you will visualize, how users will interact with data, and how you will go about using D3 to realize your design. When thinking about your project, you ought to think in terms of the **insights you want to convey through your visualization** rather than displaying the most points, the most attributes possible, or all kinds of interactions. What points or trends do you want viewers to notice or what argument do you want to make? How will viewers navigate your visualization? How will interactions help users to understand your message more deeply?

All groups are encouraged throughout the project to seek feedback from fellow classmates and course staff. The best way to improve a visualization is to have others critique it. You will be required to submit **1 milestone response** to CMS which will be graded on a completion credit basis. This milestone will be a chance to get some feedback on your progress. We will cap off the project with an in-class demo day where we will critique each other's work.

Examples and sources: You are encouraged to find data and inspiration from other sites. Make sure you acknowledge these in comments and in your written description. If you choose to scrape data, please scrape respectfully and ethically. Any code that you did not write yourself (such as d3) must go in a separate .js file. It is okay if your project uses a dataset that has already been visualized or that it *resembles* an existing visualization on a shallow level. Just be sure that your final submission represents your own *design and coding*. **Unacknowledged code or concept reuse will be handled with standard academic integrity procedures.**

Regarding grading: This is an **open-ended assignment**. With homework we have a specific idea of what we want and we "take off" points when your work deviates from that. The reason project experience is the single most valuable asset you can bring to a new job is that we do *not* have specific ideas about what projects should look like: **it's up to you and your teammates**. As a result, think of the criteria below as an opportunity to "earn" points, not "lose" points. Our principle with projects is that better work should get better grades. That does not mean that we curve: there's no reason we wouldn't in theory give everyone 100s, but in practice "perfect" grades are extremely rare.

Regarding teamwork and conflicts: If you have concerns about how your team is working contact a graduate TA individually as soon as possible. In rare circumstances we will differentiate grades with a group, but we are much happier to help a group succeed.

Turning your project in

You will upload a final zip file containing your project code and final report to CMS. Graders will then run a web server from the project root directory in order to view your submission. **It is absolutely critical for you to turn in correct, working files. Please re-download and test your project file prior to the deadline to verify you have submitted what you intended. We suggest that you elect one group member to turn in the file via CMS and have the other group member(s) download and check the contents. Do this well before the deadline.**

Milestones & Timeline

Submit a written status report, one per group, to CMS by 11:59pm on the following date. On-time completion of these reports will be worth **3 points in total**. No slipdays.

- **Friday, November 3**:
  - Begin by summarizing your final project idea in 1-2 sentences. Make sure to describe the interactions you envision for this idea (storyboarding interactions in sketches can help to flesh out your ideas before you code them up).
  - For each team member, write 1-2 sentences describing how they contributed to the project during the week.
  - For each member, list assigned tasks for the next week prior to the deadline.
  - If you have any questions / requests for feedback, feel free to include them and provide sketches or drawings

Finally, submit your project via CMS and attend demo day.

- **Friday, November 10**:
  - Have at least one team member attend class virtually to demo your project via Zoom screen share. **Make sure that you have close-to-final code to demo.** Test ahead of time to make sure that your project is ready to share.

- **Tuesday, November 15**:
  - Turn in your final report to CMS.

Grading criteria

Your final submission has two parts, a **d3-based interactive data visualization (55 points)** and a **written final report (30 points).** The remaining 10 points come from completing an external **team evaluation (3 points; details later during the term), meeting milestones (3 points each)** and having at least one team member **attend demo day (4 points)**.

1. **The root directory of your project ZIP file must include** a**n HTML page called index.html containing your visualization.** Include any additional script files and any additional data files in your repository, preferably in JSON form. Make certain that you have included all data files necessary and that your paths work properly if a web server is run from your project root directory. We will not be able to grade projects that do not run. You may import the d3 and topoJSON libraries from d3js.org, but all other libraries and data files must be stored in the repository. Your visualization will be graded on the following elements:

   a) **Complexity of the data**. Find a dataset that is manageable, however you ought to avoid trivial data. There should be more than two variables, for example. An advanced project combines multiple datasets to provide a unique, novel perspective. Be sure to curate and preprocess your data! Beginning projects often have too little data or too much. Don't overwhelm us. You must integrate at least two data sources in this project. A topoJSON file can count as one of your datasets, though we encourage you to think bigger. **It can be challenging to make geographic visualizations both compelling and interactive.** Advanced projects will deliver a seamless presentation of data, while beginning projects will show obvious splits in the interface between data sources.

   b) **Interactivity**. Advanced projects will provide clear, intuitive tools for exploring a complex data set. Each view should have an appropriate amount of information -- not too much, not too little. Projects that use motion or transitions to highlight contrasts and similarities are encouraged (though be careful with overuse of animation). Beginning projects might only add tooltips or similar descriptive elements to a fundamentally static interface. Do not rely solely on mouseover interactions – they are very hard to users to discover and often go unused. If you add an interactive affordance, signal to the user that it is there.

   c) **Technical correctness**. The code must actually do what you intend it to do. We also prefer good style in coding: use informative variable names, consistent indenting and whitespace, and informative comments.

   d) **Creativity**. Beginning projects often look like online examples or things we've seen before. Advanced projects will make us think "how did they do that?" or use something familiar in an unfamiliar way. Don't be boring.

e) **Proper use of visual channels**. Use scales such as position, shape, color, and text appropriately for variables. Advanced projects give us accurate impressions of the underlying data values, allow us to make comparisons between relevant data points, and balance between focus and context. Beginning projects are often hard to interpret and make comparisons difficult.

f) **Usability**. Someone viewing your work should be able to understand the data values represented in the visualization easily and accurately. Advanced projects make choices that are clear and intuitive and may walk us through specific examples. Beginning projects often leave us wondering what we're looking at or make us read long descriptive paragraphs to figure out what's going on.

g) **Overall polish**. Beginning projects will look like a collection of parts, with default styles. Advanced projects will have a sense of unity, even if they have multiple sections.

h) **Motivation**. What's the point? What are you trying to say? Beginning projects will present information. More advanced projects will have a clear argument and use carefully chosen combinations of marks and channels to guide our attention to the evidence that supports that argument. Advanced projects deliver insights.

2. **Submit to CMS a PDF file containing a written description of your project**. There are no specific page or word limits. This document should contain:

   a) **A description of the data**. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. **Editing is important! You are not required to use every part of the dataset.** Selectively choosing a subset can improve usability. Describe any criteria you used for data selection.

   b) **An overview of your visual design rationale.** A good rule of thumb to follow is "**every pixel must be justified."** Instead of a 100,000-element breakdown, give us an overview of the design decisions you made and the trade-offs inherent in how you displayed the data. **This part ought to include a description of the mapping from data to visual elements**. Describe marks and channels you employ such as position, color, or shape. Mention any transformations you performed, such as log scales.

   c) **An overview of your interactive elements and their design rationale**. Give us an outline of the design decisions that went into the interaction affordances you added to your visualization. What process did you use to choose the interactions you developed? How did you make them discoverable, usable, and interesting?

   d) **The story**. What does your visualization tell us? What was surprising about it? What insights do you want to convey to the viewer of your visualization?

3. At the end of your PDF file, include an **outline of team contributions to the project**. Identify how work was broken down in the group and **explain each group member's contributions to the project**. Give a rough breakdown of how much time you spent developing and which parts of the project took the most time.

External team evaluation

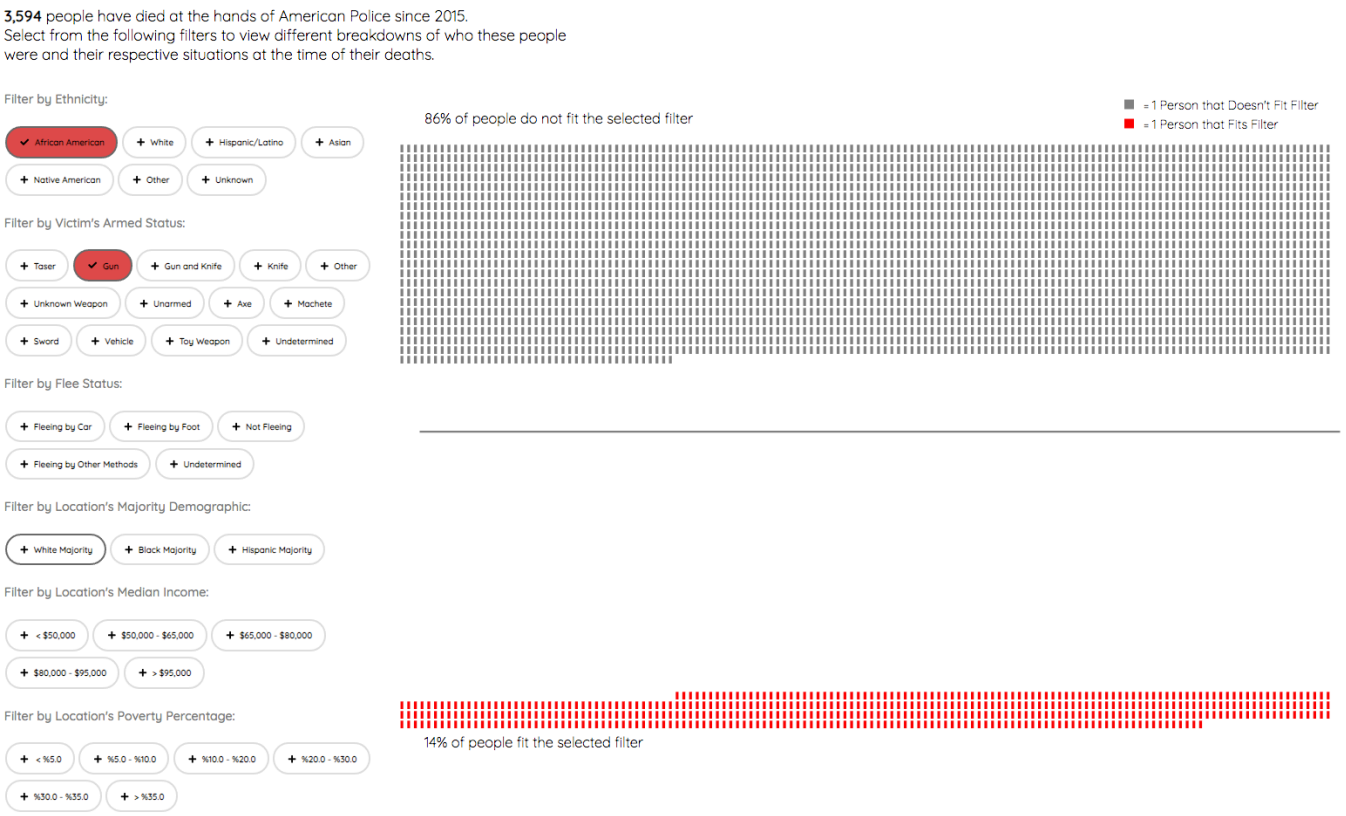
You will complete an external evaluation of your team members prior to final submission. This process will be outlined in class on a later date. Plan on completing an independent evaluation of your teammates.

Best practices

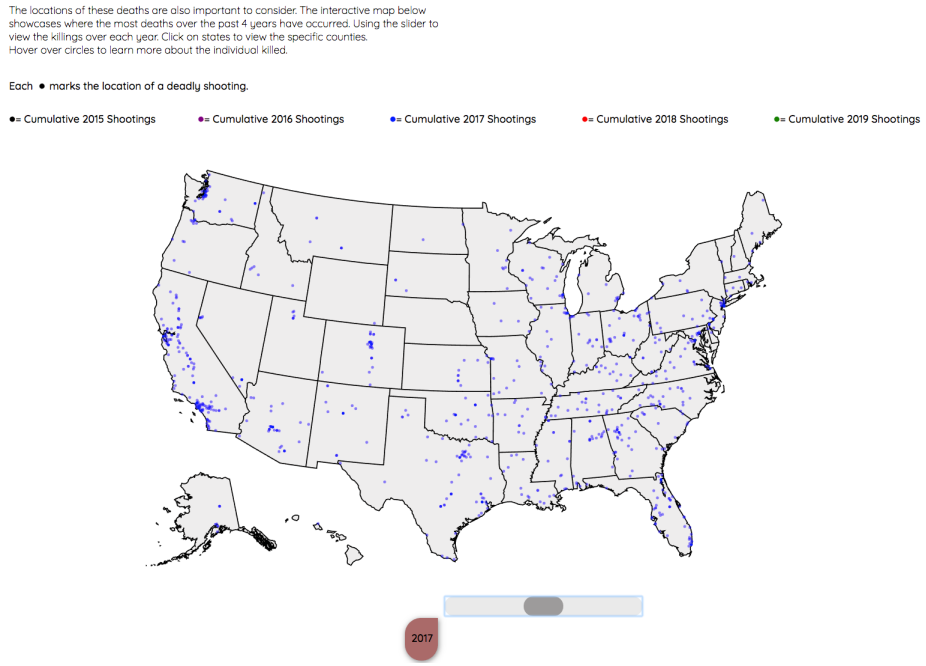
- Start now.
- Talk to each other. Listen and value each other's different perspectives.
- If a group member will be unavailable for any period during the project, figure out in advance how you will work around that absence.
- Discuss how you plan to collaborate on the project early on and set up a schedule to meet and a list of group expectations.
- Use relative paths for data, images, and other resources: do not start URLs with "/". Your project will be one directory among many, not the document root.
- Set up a code repository, like Github or Bitbucket. This is always a good idea, but it can also provide insurance if something goes wrong with your CMS submission.
- No, seriously, start now.

Example projects:

Visualizing police violence as it relates to race and socioeconomic status in the US



This view allowed users to select different criteria. An animation moved individual markers from the top to the bottom. The design doc explains why they chose individual markers instead of bars or some other chart.



The second chart features a time slider (insufficient on its own due to issues with spotting differences as you slide) that helps to show trends over time to tell another part of the story.

At first it seems like there is not much in this project, however it draws your eyes to two interactive elements which invite you to personalize the visualization.

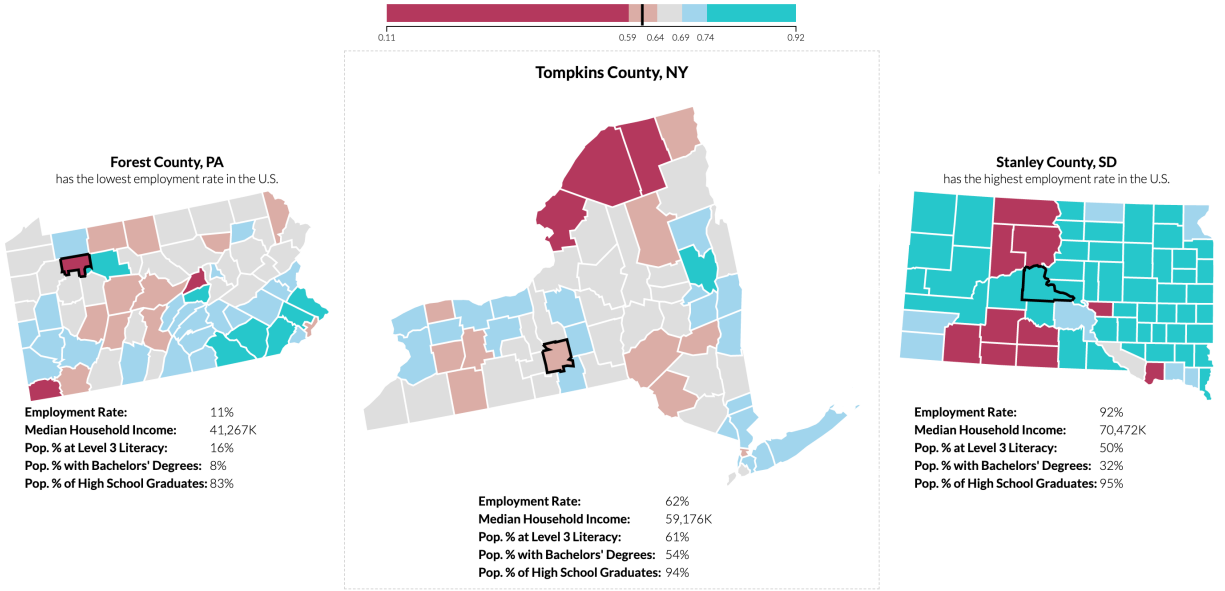
Employment & Education in the United States

Part I: How well-employed is your county?

Select your home county to see its employment rate ranking as compared to other counties.

Home State: Select your state Home County: Select your county

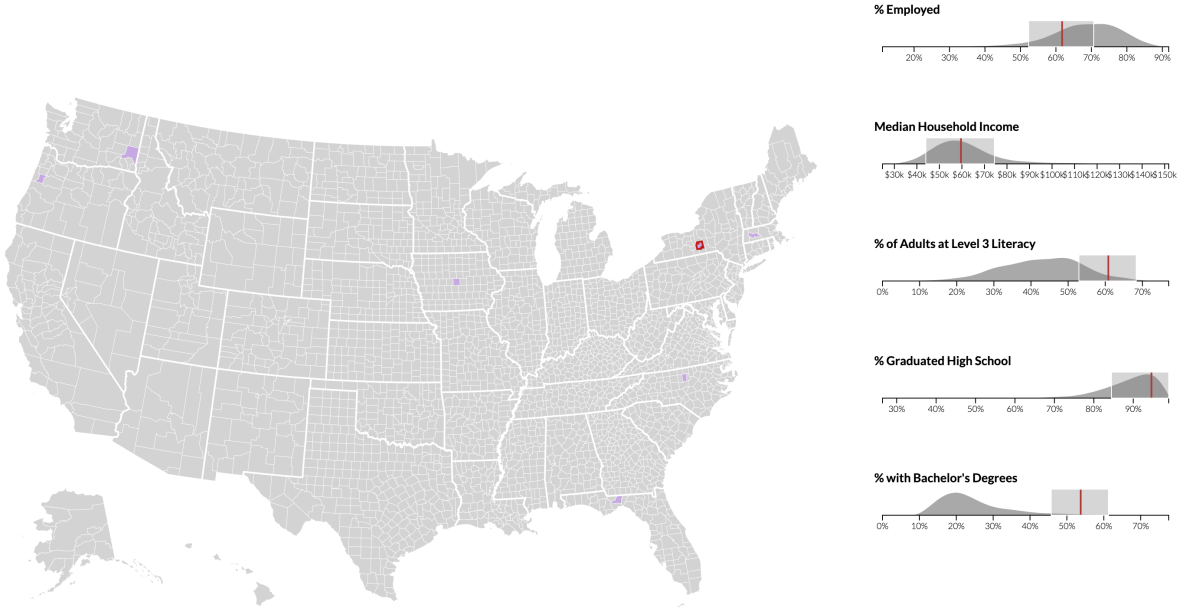
Afterwards, a variety of visualizations and select comparison points help to illustrate how the selected county compares to other places in the USA.



…and a similarity metric invites users to explore further.

Part II: Discover other counties

Starting with the counties most similar to yours, explore other counties in the US by filtering for metrics such as income, literacy, and education. Click on any county to compare it with yours!



It then continues to offer additional, useful views of the data in several more panels

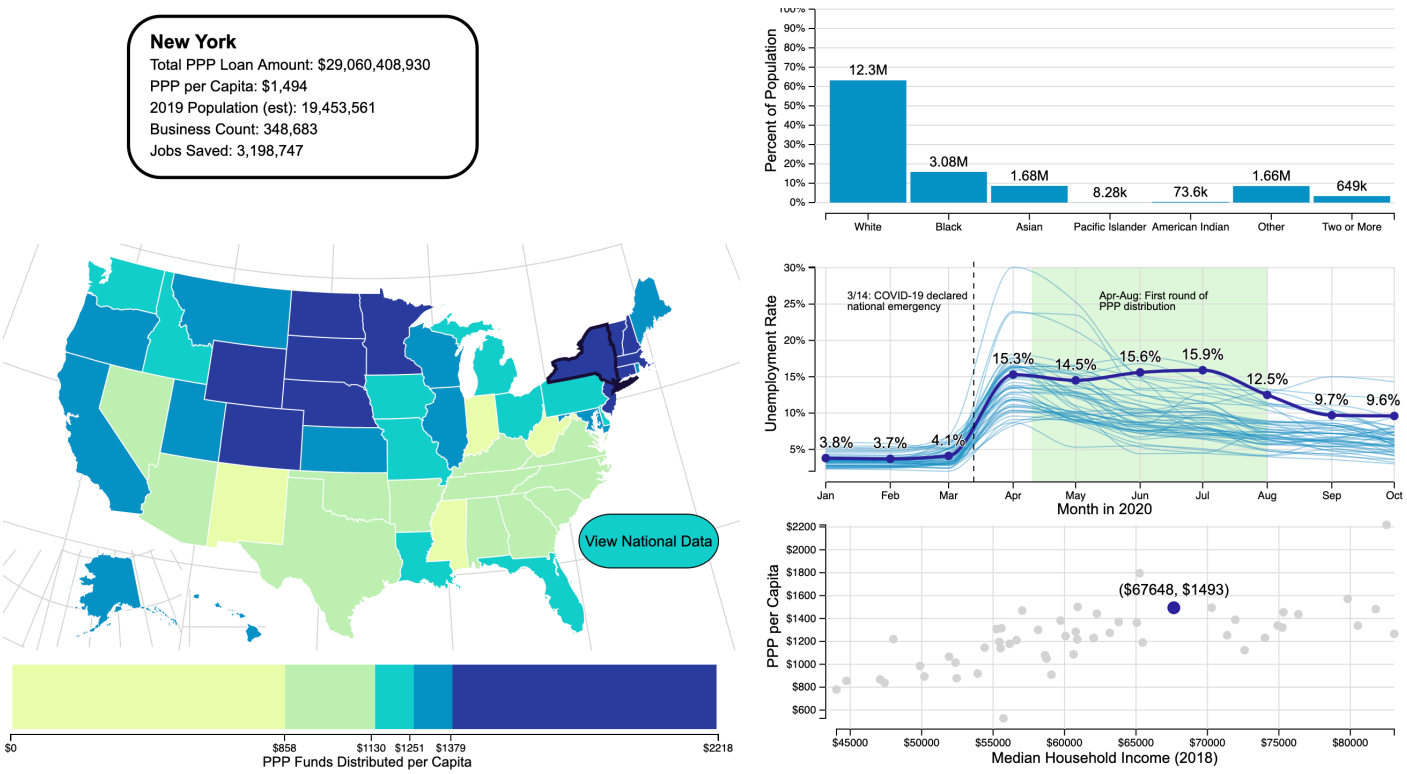
This project visualized data from the Paycheck Protection Program using an interactive dashboard. Visual features were carefully aligned by color scheme and interactive tools were linked across the dashboard to provide a consistent experience. Drill-down was provided via map interactions for individual states

Paycheck Protection Program

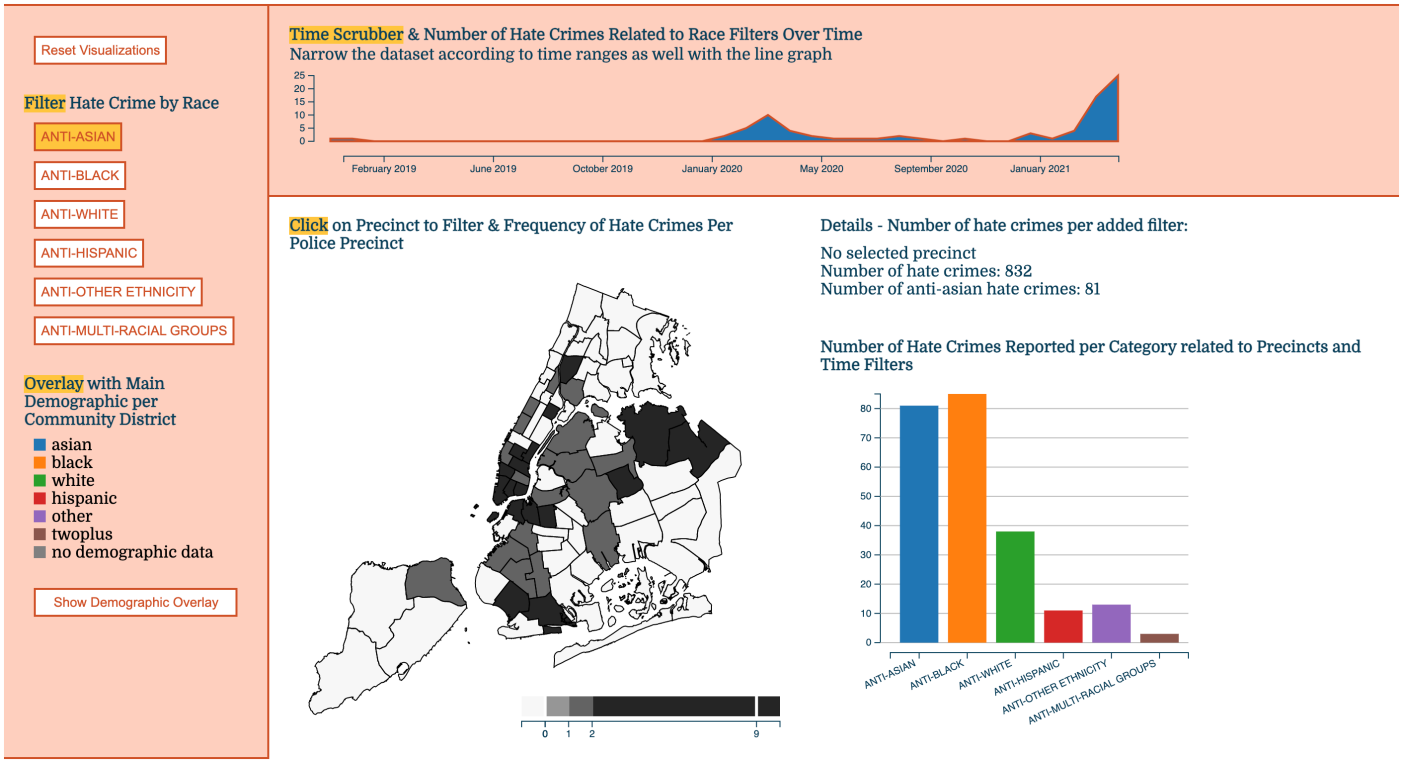
This visualization tells a story about the trajectory of the Paycheck Protection Program created by the CARES Act and managed by the U.S. Small Business Administration (SBA) throughout the pandemic. This program, referred to by the acronym PPP, is meant to help businesses (primarily small businesses) keep its workers employed and paid throughout the COVID-19 pandemic to avoid the consequences of unemployment on both the individuals and the economy. These loans were also meant to be easy to forgive and have a low-interest rate.

Our visualization tries to answer the questions: Who were the winners and losers of the PPP? Was this money given where it was needed the most? And was racial bias involved in its distribution of funds? While investigative journalism and more careful examination of the data can tell us smaller stories of misuse of funds or a heartwarming story of success, we wanted to give the big picture of its distribution.

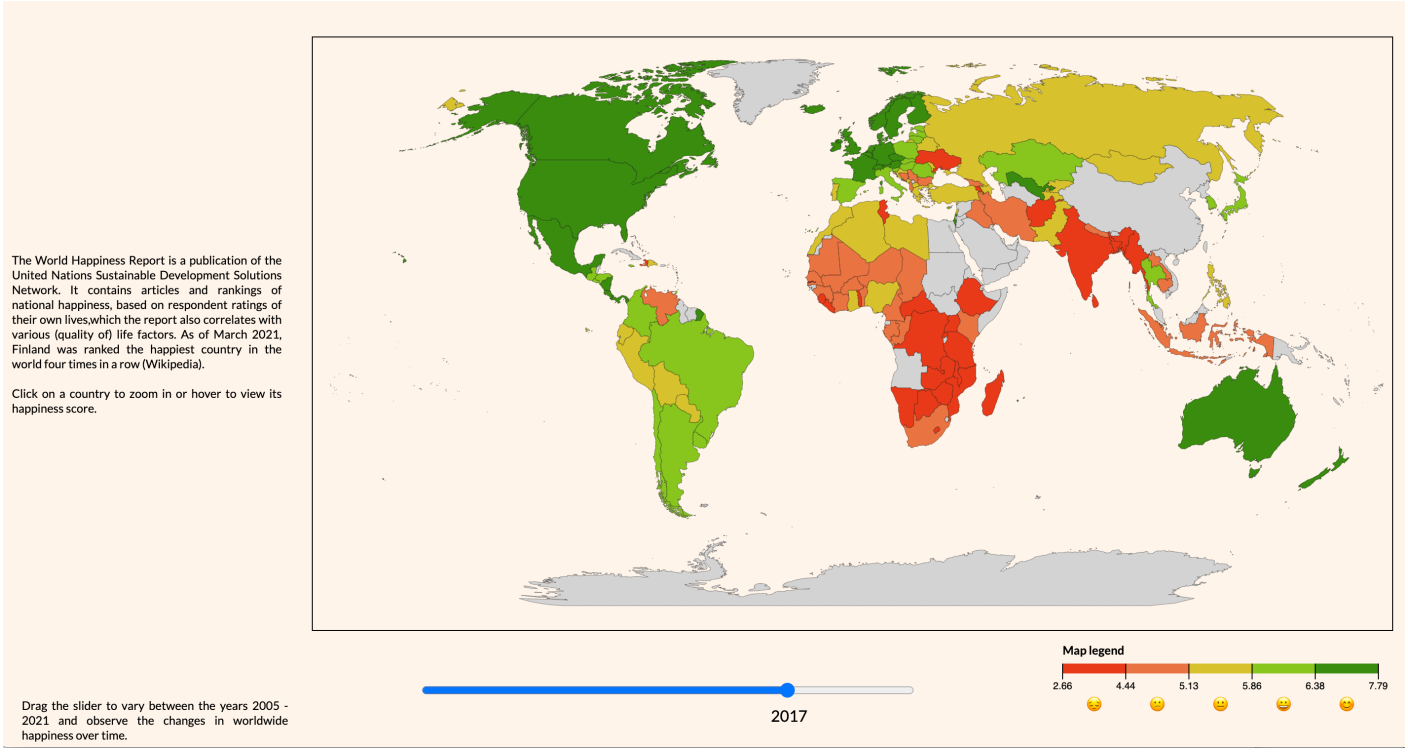
Click on a state to see a more detailed breakdown of its statistics.



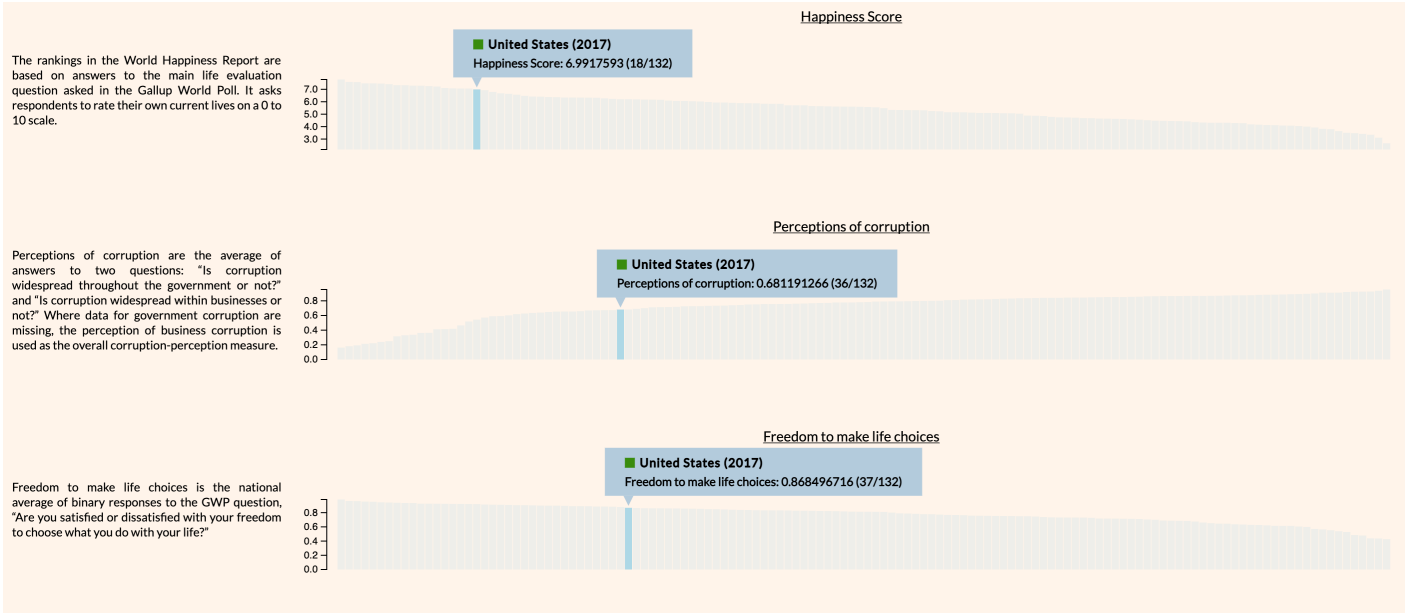
One group of students examined a dataset of hate crimes committed in New York City. Through a mixture of filtering and detail focus tools, the project helps users triangulate on different demographic factors underpinning hate crimes. Consistent use of color schemes, clear signaling of interactive elements, and ease of interaction all effective show aspects of the data.



Each year a few projects consider datasets related to world happiness scores. This one combines an interactive map view with supporting charts. The design doc explains how small elements such as emoji help to make the project usable.

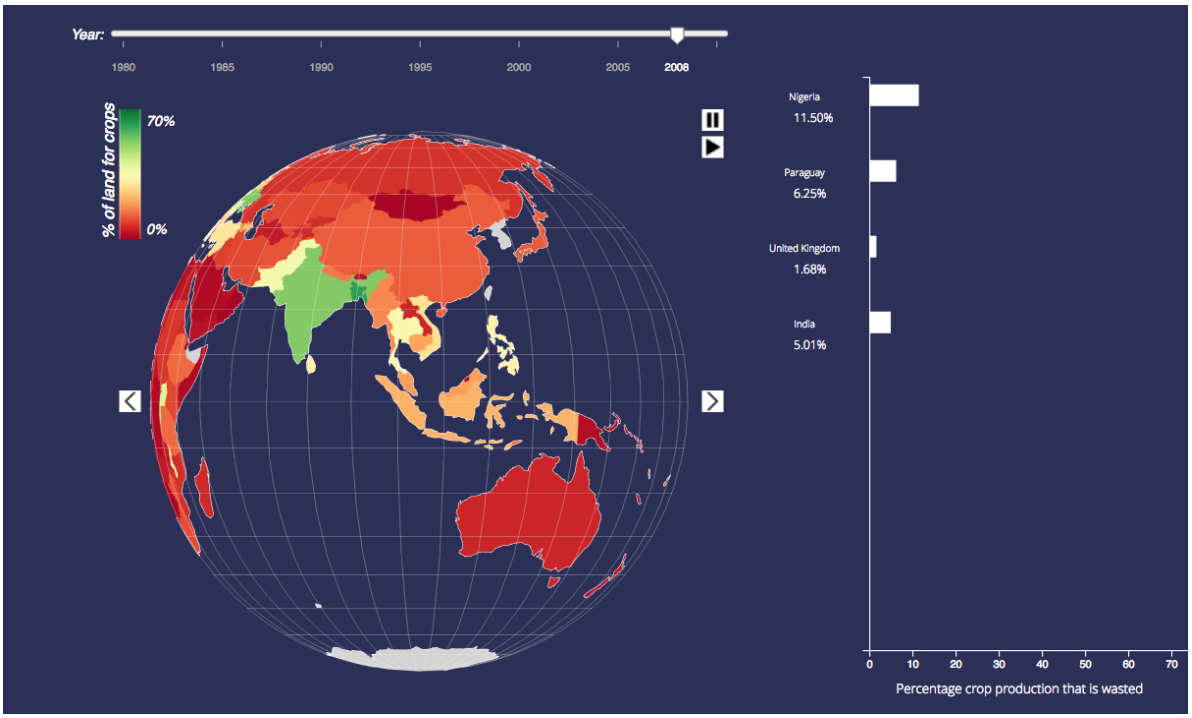


While a 'time slider' as seen at the bottom of the first chart doesn't add a huge degree of richness on its own, combined with supporting charts and interactive details views it helps to add depth.

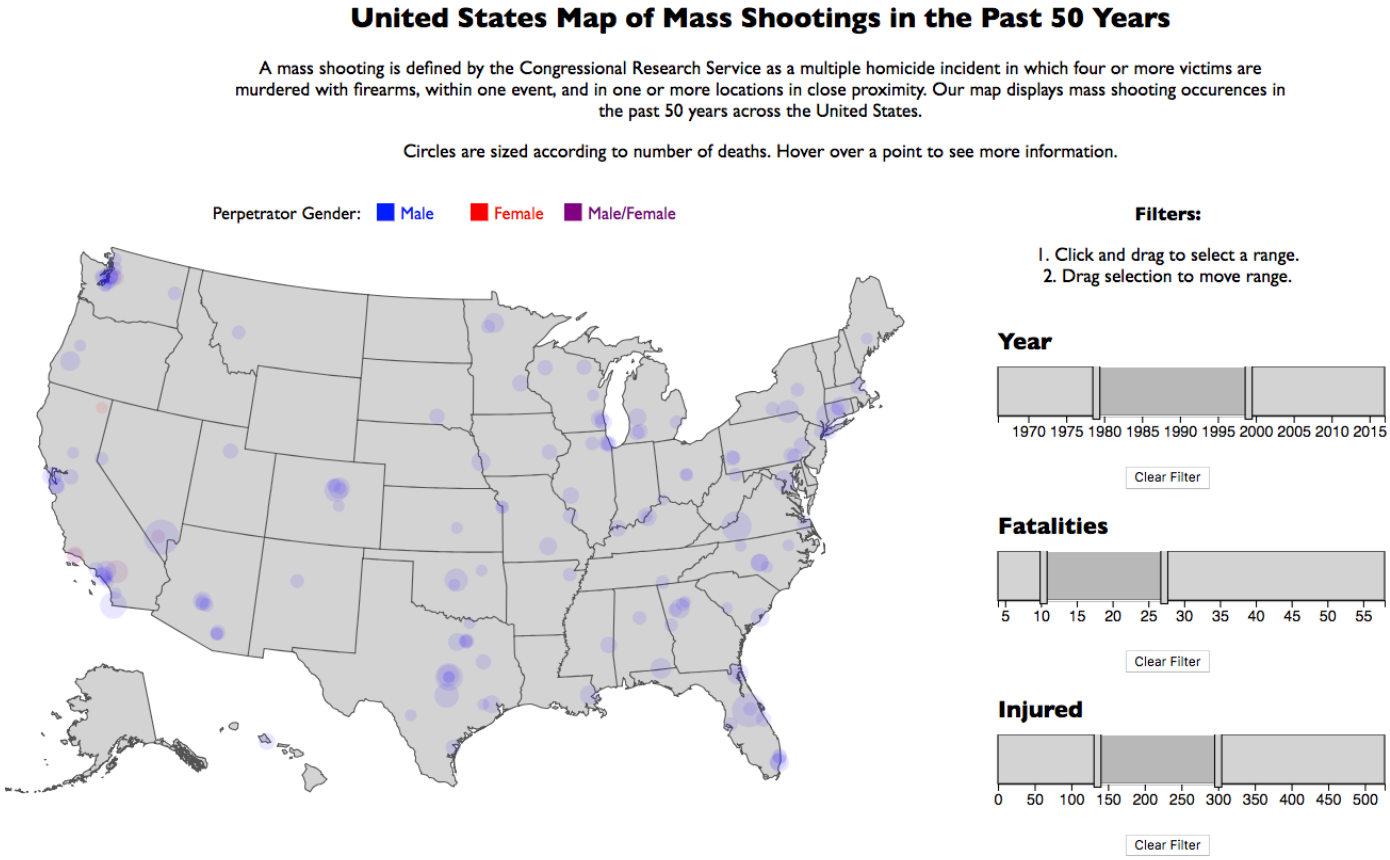


As a way to bring users into the story, this project on the efficiency of crop production started with a short quiz to customize the following views

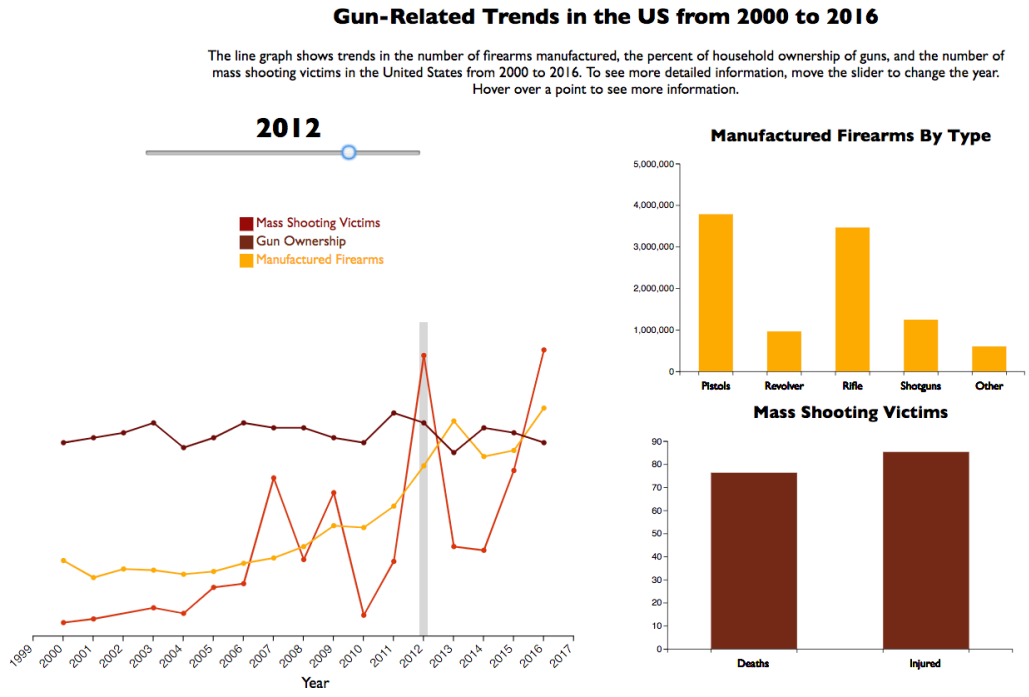
The rotating globe they then provide showing different values has "wow factor", but a static flat map might have been more effective here. A time slider and additional charts showing comparative information based on user selections add context.



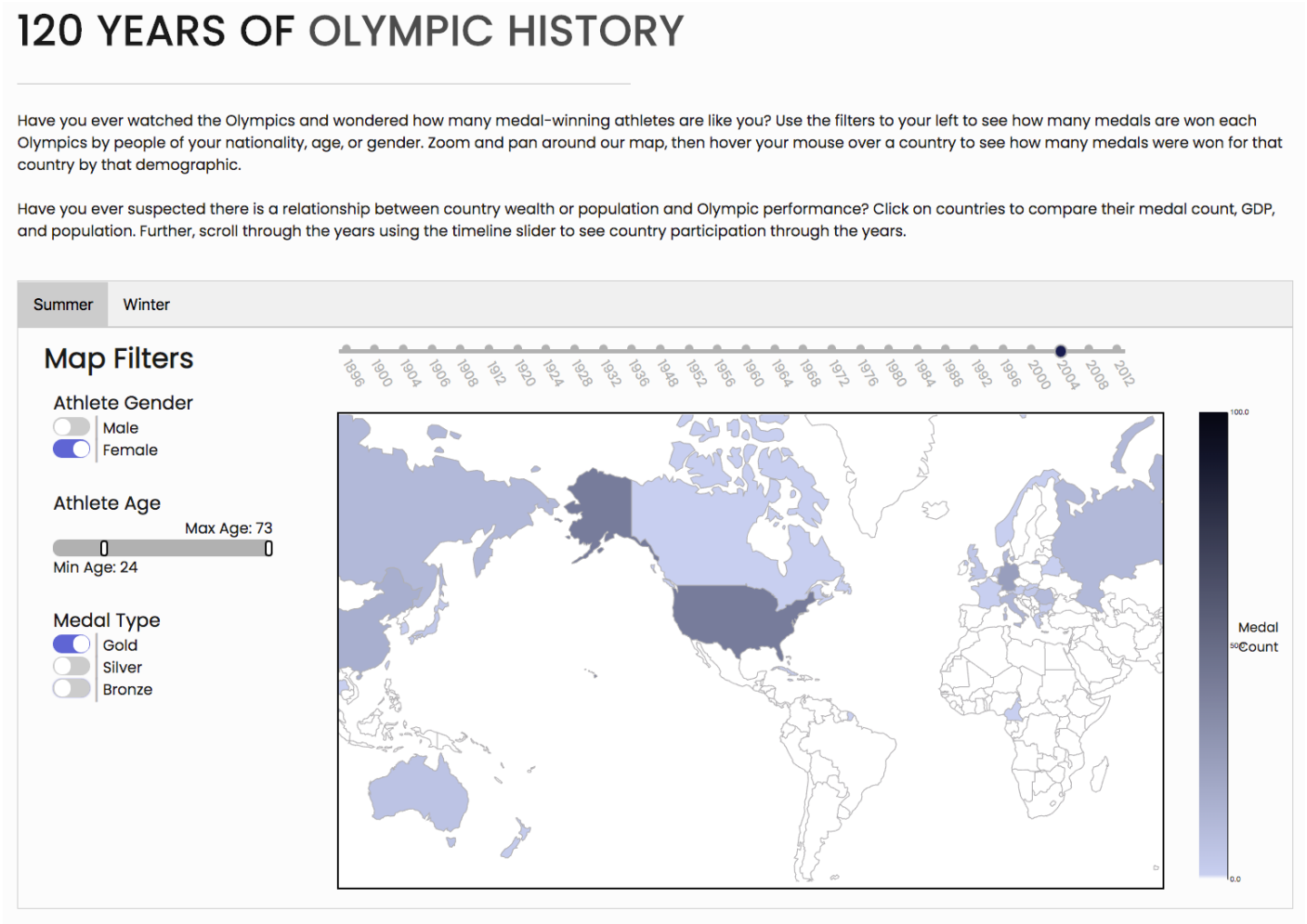
This project visualized violent data (notice how both projects dealing with crime and violence treat the subject seriously and orient their typography and color choices to reflect this). This time the authors allow users to drag dynamic query sliders to filter to explore the data.



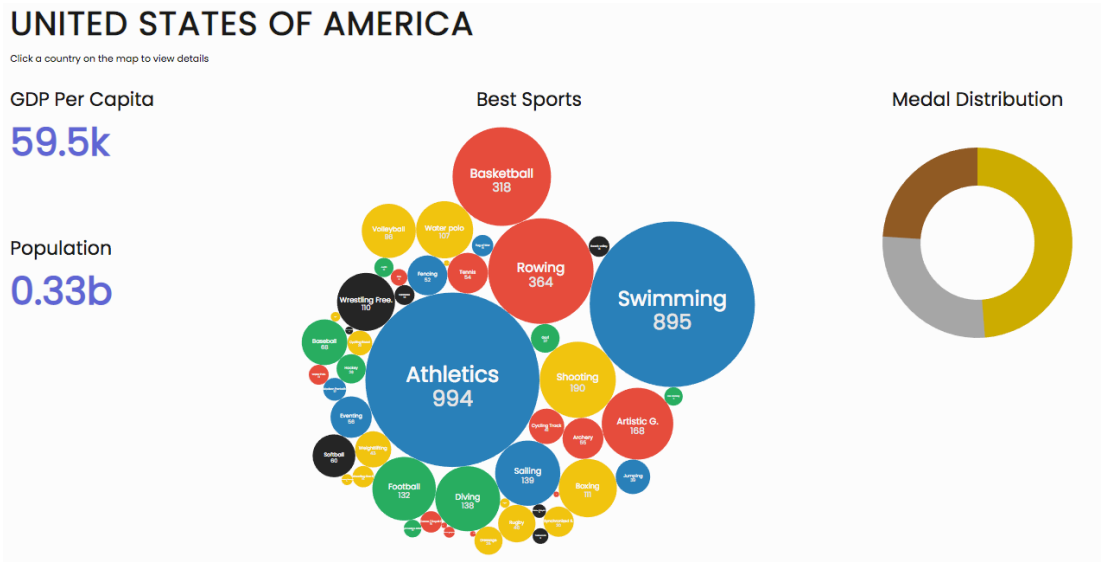
They then augment this view with a time slider showing multiple linked views of contextual data



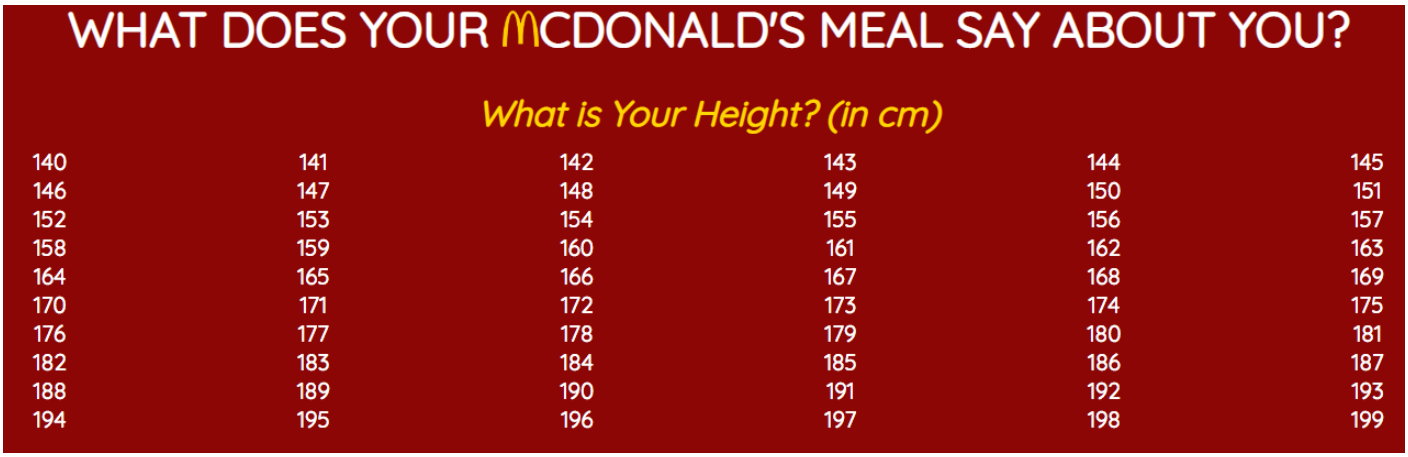
This team provided an interactive tool for filtering and exploring world Olympic medal data. Multiple coordinated filters provide immediate filter responses, allowing the user to explore different patterns. The time slider is not the only interactive tool in this linked chart.



The authors also provided additional views of other aspects of the data based on the user's filters



This team created an interactive quiz that allows a user to input personal data and then select a meal at McDonalds (fast food restaurant). Though some interactive choices for entering data might not be optimal, the color scheme and typography invite interest.



After the quiz, many different data factors and comparisons are provided to the user so they can understand how their choices compare to others. The interaction here comes from the quiz. If this team were larger, they would have been expected to add more "details on demand" tools to this final view to add depth.