CS/INFO 3300; INFO 5100
Project 2
Due 11:59pm Thursday, April 30th (no slipdays)

You are creating an **interactive data visualization**. Your 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).

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?

Successful projects tend to focus on one core message and align all design elements (marks, visual channels, annotations, interactions) towards one goal. Strong projects will go beyond hover-for-details popups and integrate interactions that help to answer questions like "what if," "what about this segment of the data," and "tell me more about this outlier in context to peers".

On the day that the project is due, April 30th, there will be an **in-class demo session for your group projects**. We will split into breakout groups to critique each other's work. **Make sure that you have a prototype of your project prepared for the in-class demo, and that at least one team member attends to demo and critique**. If your team cannot attend due to time zone or holiday observance, please let a graduate TA know ASAP via the staff email address.

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. The milestone will occur earlier in the project period to give you a chance to float ideas and receive feedback on them.

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. **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: Please be respectful of your teammates time, effort, autonomy, and backgrounds. 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.**

Milestones

Submit a written status report, one per group, to CMS by 11:59pm on **Monday, April 5**. Your milestone report should:

- List five ideas you have for your visualization project (1-2 sentences each).
- For each of those five ideas, make sure you describe the interactions you envision.
- Each team member should list their assigned tasks for the following week.
- If you have any questions / requests for feedback, feel free to include them and provide sketches or drawings.

On **April 30th** we will have an in-class synchronous **demo and critique** session. Make sure that you have a prototype of your project prepared for the **in-class demo via screen sharing**, and that at least one team member **attends to demo and critique**. If your team cannot attend due to time zone or holiday observance, **let a graduate TA know ASAP** via the staff email.

Grading criteria:

Your final submission has two parts, a **d3-based static data visualization (55 pts),** a **written description of your visualization (35 pts),** an **outline of team-member contributions to the project (4 pts), your milestone grade (3 pts), and demo day attendance (3 pts)**. Your submission must meet the following requirements in order to earn a high grade:

1. **The root directory of your project ZIP file must include an 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 absolutely 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. 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.

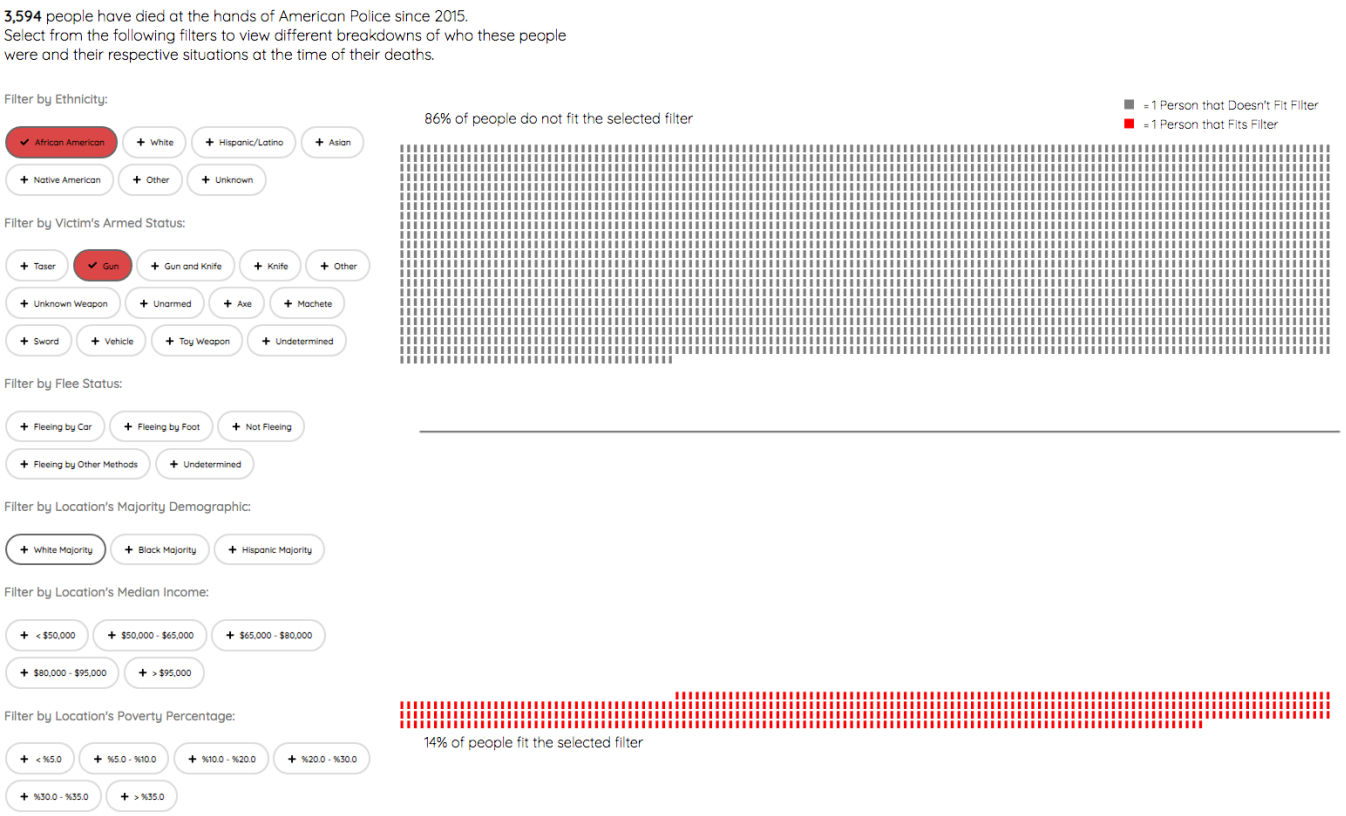
   c) **An overview of your interactive elements and their design rationale**. Give us an outline of the design decisions that went into the interaction affordances you added to your visualization. What process did you use to choose the interactions you developed? How did you make them discoverable, usable,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

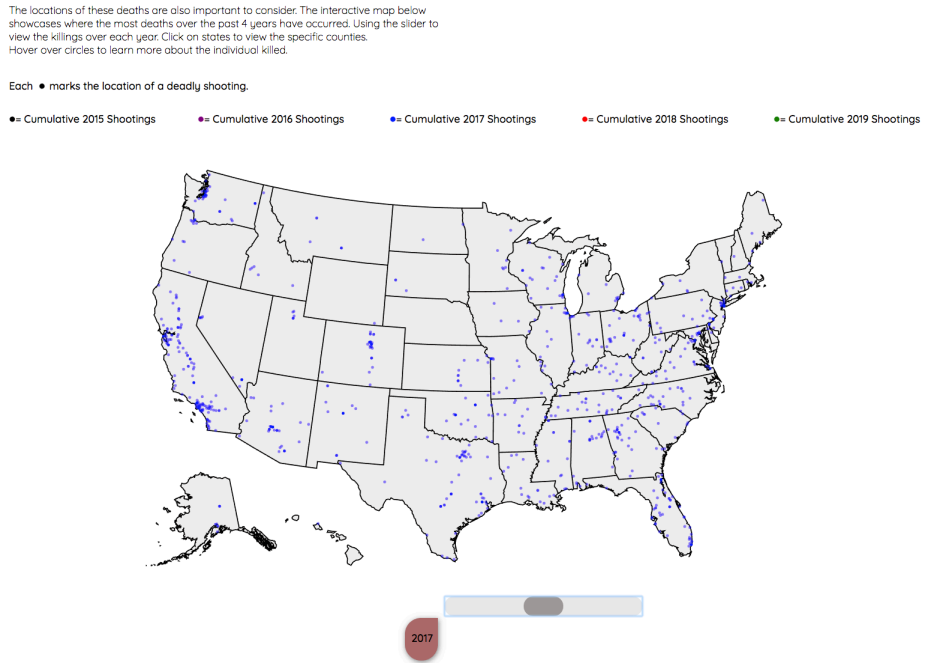
3. At the end of your PDF file, include an **outline of team contributions to the project**. Identify how work was broken down in the group and **explain each group member's contributions to the project**. Give a rough breakdown of how much time you spent developing and which parts of the project took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

Example projects:

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



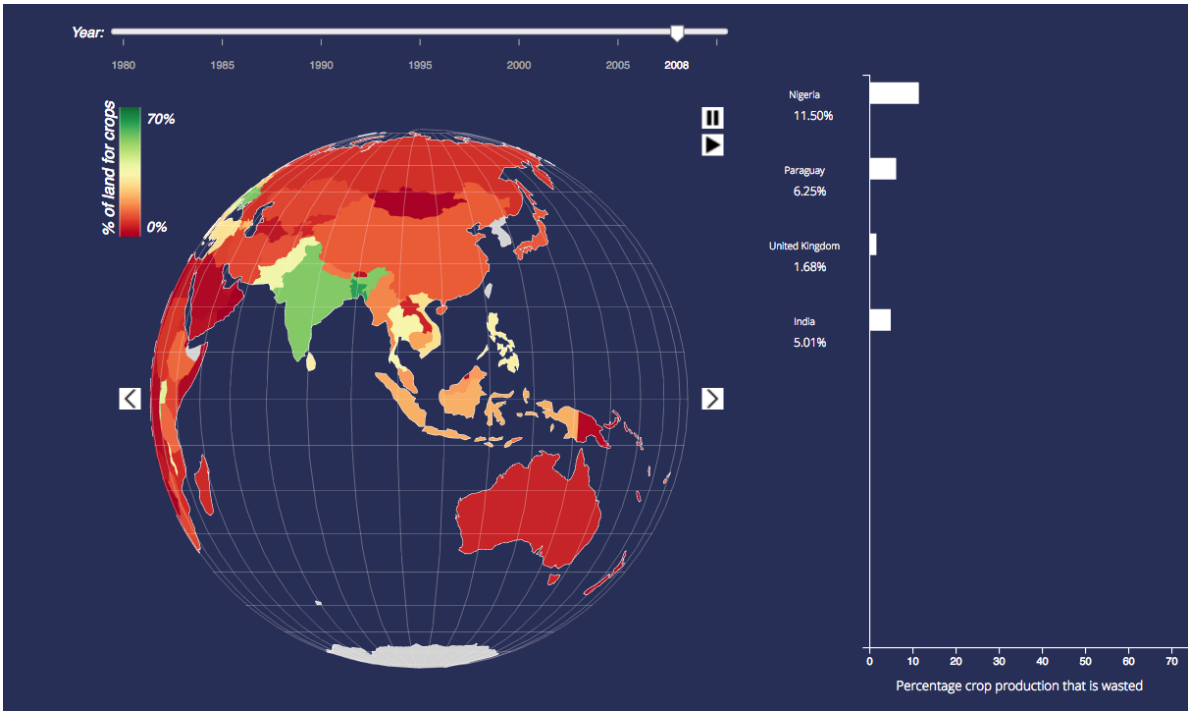
This view allowed users to select different criteria. An animation moved individual markers from the top to the bottom. The design doc explains why they chose individual markers instead of bars or some other chart.



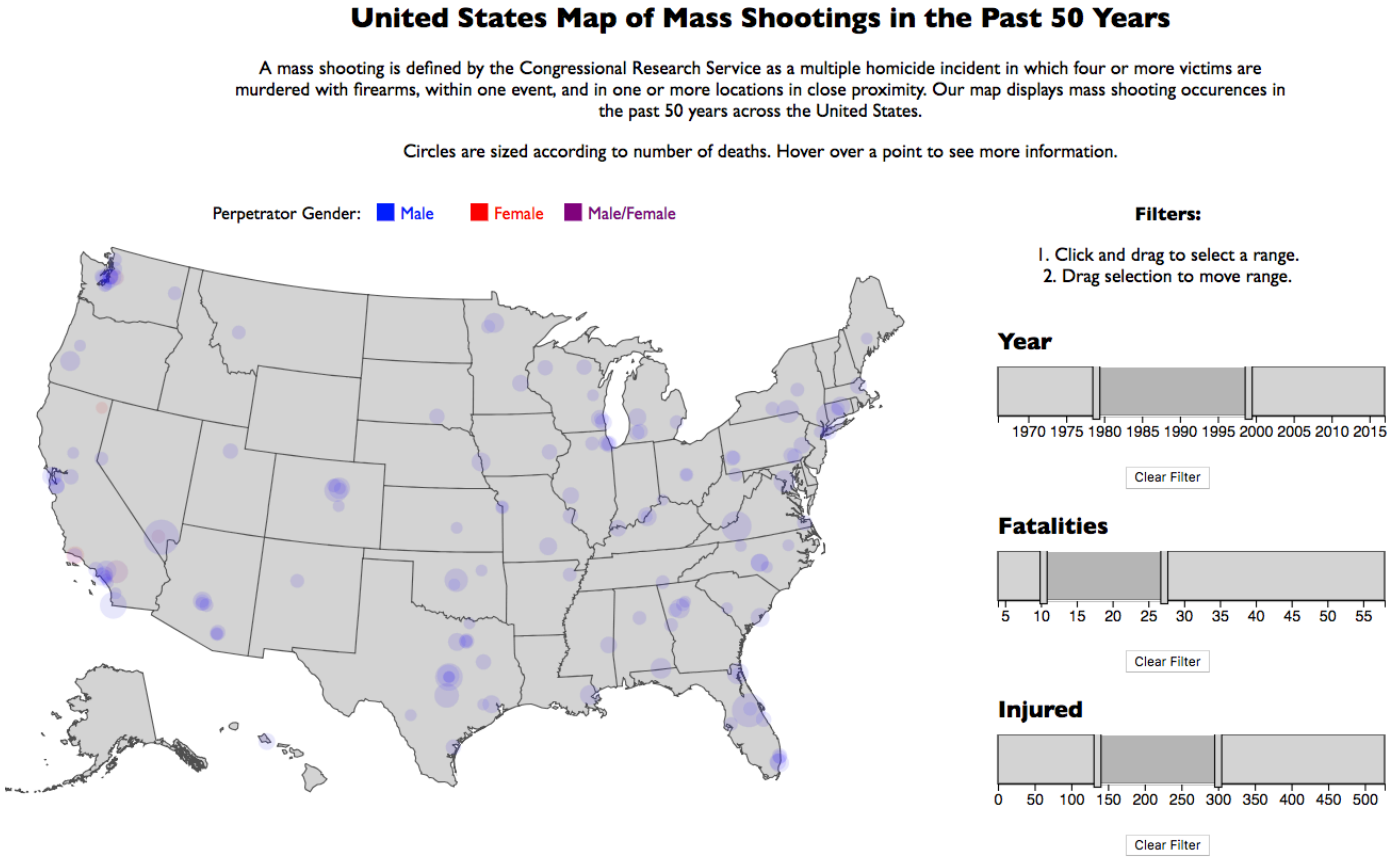
The second chart features a time slider (insufficient on its own due to issues with spotting differences as you slide) that helps to show trends over time to tell another part of the story.

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

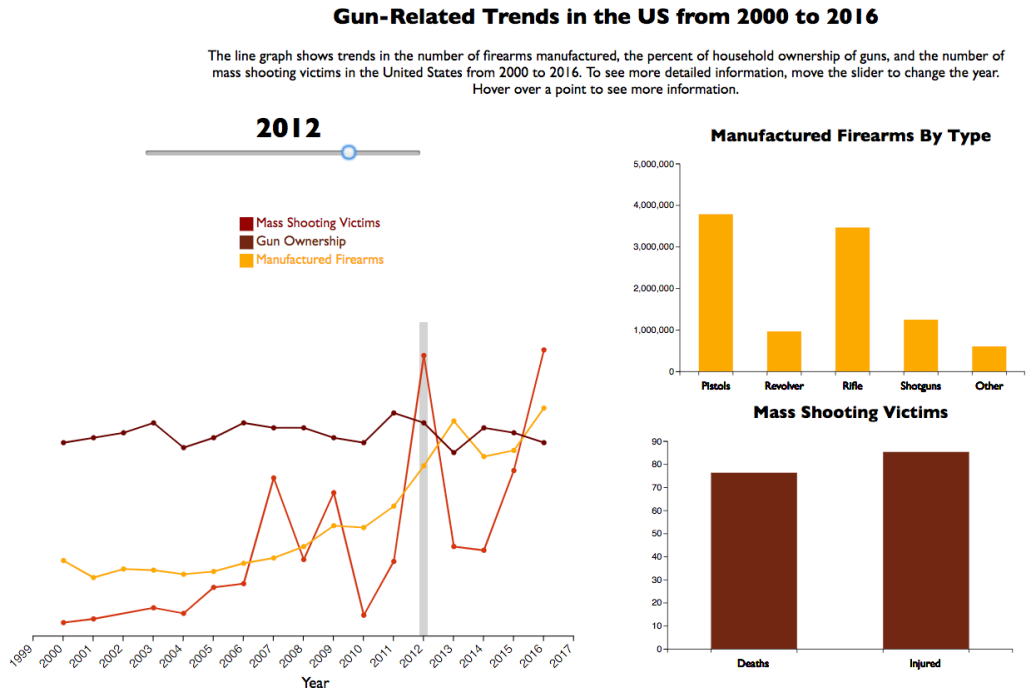
The rotating globe they then provide showing different values has "wow factor", but a static flat map might have been more effective here. A time slider and additional charts showing comparative information based on user selections add context.



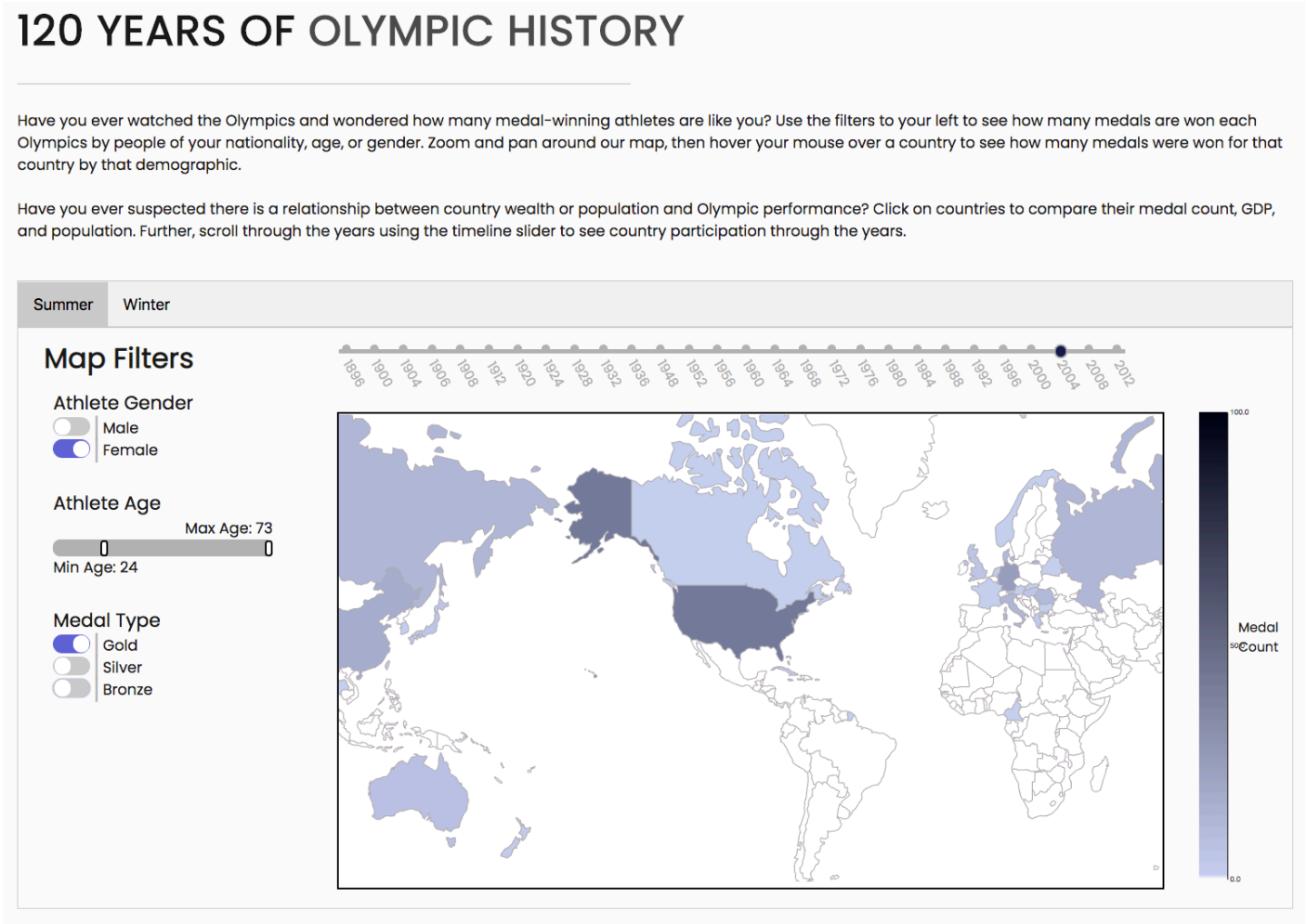
This project visualized violent data (notice how both projects dealing with crime and violence treat the subject seriously and orient their typography and color choices to reflect this). This time the authors allow users to drag dynamic query sliders to filter to explore the data.



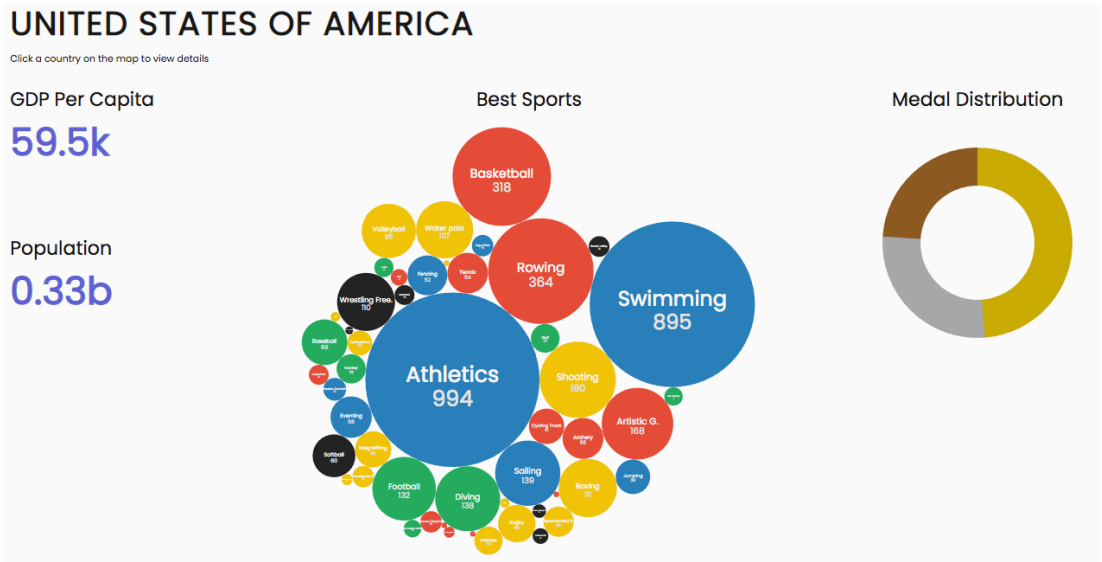
They then augment this view with a time slider showing multiple linked views of contextual data



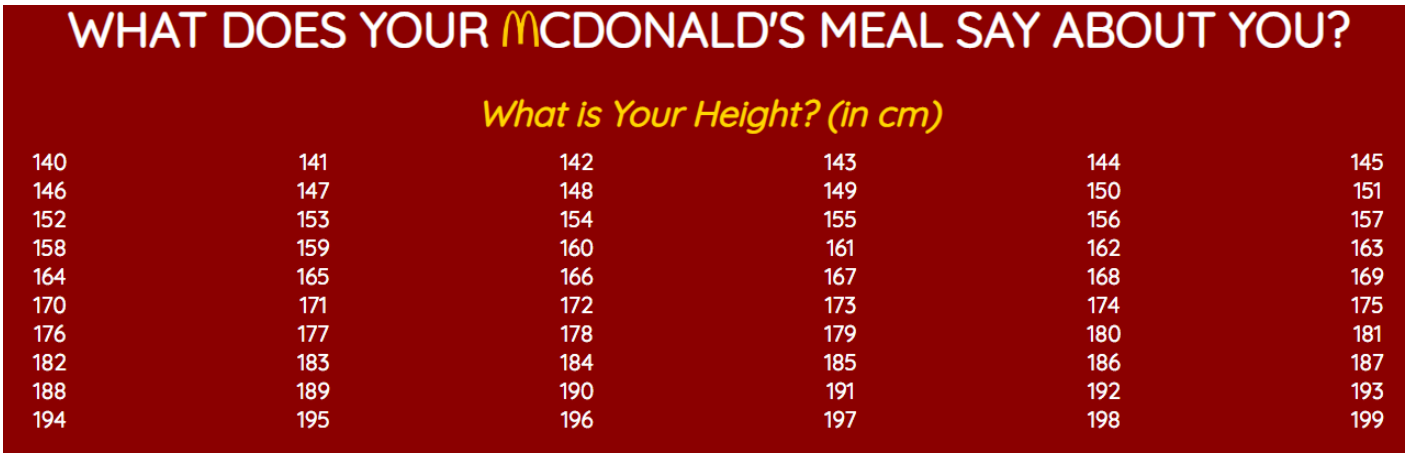
This team provided an interactive tool for filtering and exploring world Olympic medal data. Multiple coordinated filters provide immediate filter responses, allowing the user to explore different patterns. The time slider is not the only interactive tool in this linked chart.



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.