The GA Demographics / Constituent / Resource Map

# Overall:

This is a series of data projects that will have a visual component. There are two parts:

1. Georgia Science Constituent Map
2. State of Science in Georgia (SoS Ga)

They are supported by the Backend Data Project.

**Purpose**: Understand the make-up of GA so we can identify audiences and messaging that resonates with those audiences. Understand political history. Understand sentiment of the GA Science and Technology community. Identify where politics may be at odds with constituent demographics.

This map is of large interest to our current (and potential future) funders! It adds pizazz as well as information.

* A map that people can use to understand the demographics of their area
* A map that people can use to contextualize events / policies to their local area
* A map that people can use to write contextual narratives around
  + How the S&T Community thinks about various issues
  + How a specific Issue or Topic effects Georgians (or a GA county / region)
* A map that Sci4Ga can use to understand
  + People and places to engage – and how best to engage them (what topics and arguments might resonate based on their needs?)

Purpose of specific parts

1. Georgia Science Constituent Map – add context with numbers around race, age, income level, access to services (healthcare, fresh foods), employment sector

Gauges public sentiment on science-adjacent social issues and shows legislative activity against demographic, geographic, and other contexts to inform Science for Georgia and its partner network of advocacy opportunities. For Sci4Ga and partners and anyone who wants to advocate for a specific policy. This can help to identify areas, populations, and policymakers to effectively target and who to ignore.

* 1. How does a legislator’s voting record compare to his/her constituent base?
  2. What was voter turnout?
  3. What is local budget spend on education vs law enforcement vs infrastructure vs R&D?
  4. Which issues are people least informed on?
  5. Which issues might be easiest to connect with people on?
  6. Where does it make sense to advocate for a given issue?
  7. Who are those with the greatest interest on a given policy?

1. State of Science Survey, SOS Ga – for policy makers, business leaders, and advocates to contextualize the Science and Technology professional population in Georgia, answering key questions that policy makers want answered.
   1. How does R&D spending relate to local economies?
   2. Where are R&D dollars spent in Georgia?
   3. What is the demographics of Science and Technology professionals (vs the overall GA population)?
   4. What things do Science & Technology professionals consider when contemplating whether to move to or stay in Georgia?
   5. Where do Science & Technology professionals stand on policy issues?
2. Backend Data Project – the above two rely heavily on data – the data needs to be input in a transparent and uniform manner, easily updatable, and available for people to verify / add to / perform research on.

Bounding Conditions / Requirements

* The demographics and visuals will be listed in a publicly accessible server
  + Two User Types and access
    - One: General Public – can sort and filter the visuals but cannot change the data – they can download the data if they want to do their own research.
    - Two: Data Volunteers – can create outputs, can manipulate the data – these are people vetted and known by Sci4Ga.
* Sources must be clearly identified and labeled.
* Clicking on a county or filtering by a demographic will update all the other “panels” in the visual.
* Stats must be shown relative to the other counties and/or national averages (to add perspective).

### Cool Examples

<https://ensia.com/features/6-ways-environmental-advocates-can-change-climate-concern-into-action/>

<https://model.earth/community/tools/#partners>

<https://neighborhood.org/localsite/info/>

# User Stories

## Georgia Lawmaker (Staff)

Are a county’s residents who have heart disease better served by employing a cardiologist or by adding another health clinic (i.e. do they not have access to the right doctor, or do they not have access to any doctor?)

Questions to ask: Who needs to go to clinic? Average distance to clinic? Number of clinics in a county? Chronic disease burden? County demographics (what services do people need?)

Go to the demographics map. Plug in your county.

See the demographics, heath stats, access to clinics and access to healthy food in your county (vs. state as a whole).

Decide if people need another clinic or if they need the existing clinic to better meet their needs.

## Concerned Citizen who cares about water quality

Wants to advocate for cleaner rivers in their county.

Demographics in county, including disease burden and air and water quality.

(Science for GA should have links to “GA Water Project” so you could see the cleanliness of the rivers in the area vs the general demographics.)

Go to the demographics map. Plug in your county. See demographics across your county vs other counties in GA.

Go to the SOS GA map and see the sentiment of the population and if they care about water quality. Understand ways to make an argument.

## Town in GA wants to open a VC Incubator

Show R&D drives economic growth.

Show what S&T employees want to see the local town care about.

Show where your local town is vs. the rest of GA.

Show education rate vs job sector (educational pipeline).

Go to the demographic map: Look at racial and class divide.

Go to the Policy Map: Budget spend on research, infrastructure, growth, law & order, education

Higher education institutes.

Go to the SOS Map and see the sentiment of the current S&T residents in your city.

## Potential projects

* Tie R&D funding to economic activity
  + Data should already exist somewhere
* Environmental Effects of Poverty Meta-analysis
  + What are the contributing factors?
    - Put all the data together and see what shakes out
  + Contributing factors and where they occur the most in Georgia
    - Where should we spend policy dollars and on what interventions?
  + Data side – fill in missing data on environmental quality factors and lifestyle factors and budget spending
* Fill in the missing data gaps
* Host the data in a way that
  + Interns can use nicely
  + People can download and use themselves
* Nice visual of GA demographics (that tie to various things: EEP, Economic Activity)

# MVP

Demographics Map / Chart - Show county by county and contextualize vs the rest of the state. Contextually what the heck is going on in Georgia.

* We have a starter of this Map in Tableau.

SoS GA - Show visually and contextualize the S&T populations. A series of interconnected maps and charts that show demographics and priorities.

Policy Map – overlay the demographics map with Twitter sentiment analysis on topics of interest.

* Pick one topic and do 2 weeks of sentiment analysis
* Show contextualized sentiment on a topic
* Suggest topics for us to pick. (correlation with vaccine hesitance)
* Identify polling orgs that already do this already? Steal it.

Be able to tell a story around a policy related question.

# Data Issues & Ideas

* Some of stats needed (such as budget spending percentage) requires a county by county record search – potential to automate?
* About every other year update the base demographics from US Census
  + Missing employment sector from our demographics right now
* Missing Data –
  + State of Science GA – being conducted this fall
  + Budget-spend per county
  + Links to other GIS-type science data (<https://adoptastream.georgia.gov/>)
  + Cool GA stats site - <https://datausa.io/profile/geo/georgia>

# Progress so far

Two Tableau files have been created on top of a comprehensive sheet of statistics.

The Tableau files and source data are located in Programs > GA Policy Map > Tableau Files

There are “read only packaged workbooks” – enables sorting and filtering of the data – but not changing the data.

[Policy Map (focused on legislators)](https://sci4ga.sharepoint.com/sites/programming_committee/Shared%20Documents/Data%20Focused%20Programs/GA%20Constituent%20Map%20and%20Science/Tableau%20Files/SCI4GA%20Policy%20Map%20(1).twbx)

[Demographics Dashboard](https://sci4ga.sharepoint.com/sites/programming_committee/Shared%20Documents/Data%20Focused%20Programs/GA%20Constituent%20Map%20and%20Science/Tableau%20Files/Demographics%20Dashboard%20CONNECTION%20EDIT.twbx)

To read the Packaged Workbooks without Tableau on the free Tableau Reader

1. Download Tableau Reader - <https://www.tableau.com/products/reader>
2. Download the Packaged Workbook
3. Open Tableau Reader
4. Open the workbook

There are a lot of source Excel Workbooks in there as well.

Older Demographics files (from a 2018 project) are located in “2018 – Demographics and Tableau from Amy” – these can be helpful references if needed. There is a readme in there on how the files were created.

# Notes of Interest

Liam pointed out that you can put legislative boundaries in Tableau using ArcGIS - <https://www.tableau.com/about/blog/2020/4/more-spatial-data-power-tableau-connect-esri-and-oracle#:~:text=Connect%20directly%20to%20your%20Esri,URL%20or%20GeoService%20API%20URL.>

<https://opendata.atlantaregional.com/datasets/0a318fbe1cec4040abbc440b501f902e_75>

Also Story Maps - <https://storymaps-classic.arcgis.com/en/>

# Display Details

Details around what should be listed and how displayed.

**On host website – we need to show sources of data**

* When you click on a county/region all the charts are linked and all filter to that county/region
  + If you filter down to a county – make it really obvious that all stats are reflecting that county
* For each chart we need to show things relative to the rest of the state or the rest of the country
  + What are we going to display (i.e. number of people per square mile or percentage minority)
  + How are we going to display it (map, barchart)
* Remember that we can have supplemental information on the “host pages”
* We can collapse the different sections so as to not overwhelm

State Map shows

* Show things by county
* Relative to entire state
* Visual representation of regional trends

Bar Charts shows

* Show things by non-county groupings

## Demographics Map

|  |  |  |
| --- | --- | --- |
| Policy Areas | Communities | GA Regions |
| * Education * Environment * Tax reform * Criminal Justice * Healthcare * Health access * Food access * Housing * National Security * Immigration * Weapons of Mass Destruction * Voting * Jobs | * Low-income * Underserved * Vulnerable * Food insecure * Housing insecure * Youth * Elderly * Veterans * Special needs * People with disabilities * Professionals * Business * Universities / Higher Education * University Students * University Educators * K-12 Educators * K-12 Students * Policymakers * Press * Incarcerated * First responders * Mental Health * LGBTQ+ * Minorities | * + <https://www.dca.ga.gov/local-government-assistance/planning/regional-planning>   + Northwest   + Mountains   + Atlanta   + Northeast   + Three Rivers   + Middle   + Central Savannah   + River Valley   + Heart of Georgia   + Coastal   + Southwest   + Southern |

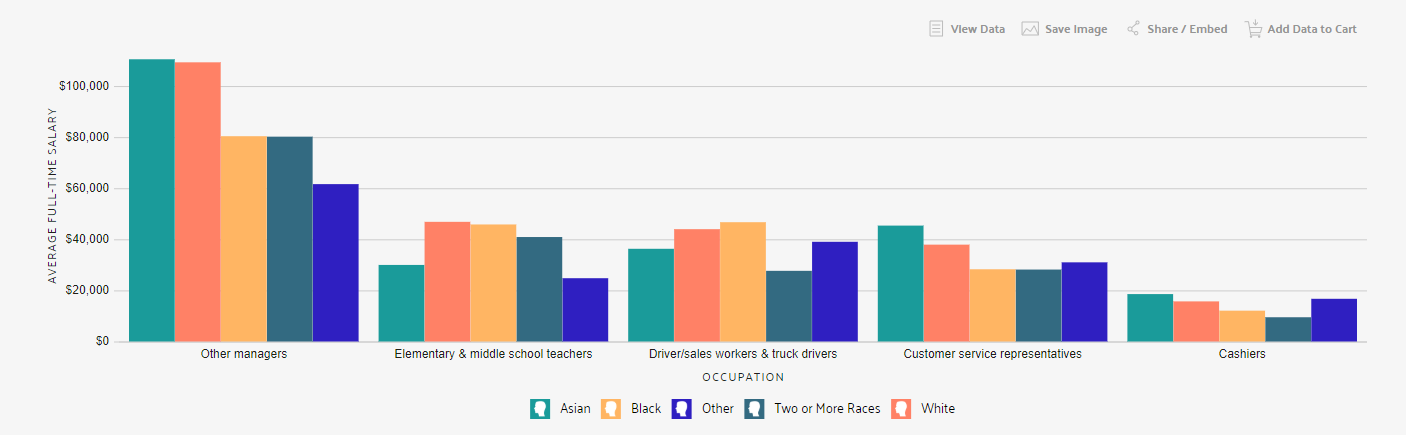
### General Demographics

Purpose: understand where the people are and how we can serve them.

We will need to maintain a visual database of:

* Population statistics - by county – race, sex, age, income, education, occupation, who takes care of the kids, single parent household, venture capital investment, religion - <https://www.census.gov/data.html>
* How people voted in the last few years – in this [excel file](https://teams.microsoft.com/l/file/7B0F67A6-6CDC-49FF-A3E3-0D933085EE2B?tenantId=38a3f439-99bc-428a-b228-cae5307c68f9&fileType=xlsx&objectUrl=https%3A%2F%2Fsci4ga.sharepoint.com%2Fsites%2Fprogramming_committee%2FShared%20Documents%2FState%20of%20GA%20Science%2FDemographics%20information%2Fincome%20race%20voting%20history%20BASE%20plus%20Polling%20AVGS.xlsx&baseUrl=https%3A%2F%2Fsci4ga.sharepoint.com%2Fsites%2Fprogramming_committee&serviceName=teams&threadId=19:dfd05d822ce043aabf5856d5f99d397c@thread.skype&groupId=42d2ec74-d1ea-4f9e-b8f0-a4dd84579d20) (income race voting history BASE plus Polling)
* Location of power plants, universities, waste disposal
* Location of natural resources
* Are there public health and environmental databases with percent with asthma, or air quality, or water quality or water source or tree cover or type of land use?

One map in the middle with 6 on the sides.

1. Population Density
   1. What: # of people per county
   2. Color Scale – peg the color scale with min and max values (i.e. 1000 to 800K)
   3. How: Yellow Map
   4. Selector – just click the county to zoom in
2. Income Level
   1. Bar chart – compare to national average
3. Race
   1. State map per county
   2. Shaded map with the filters
4. Age
   1. Keep as a bar graph – need to show relative to GA averages or national averages
5. Religious Adherence – Iqra researching more
   1. “Percentage of Religious Believers”
   2. Color scale – peg it
6. Education Level
   1. Grouped Bar chart –
   2. Kind of look like this:
   3. 

### Health

1. Insurance Coverage
   1. Like the pie charts – make “no coverage” red or yellow
      1. Public / private can be dark and light purple
2. Chronic health issues
   1. Make sure you can either see one at a time in a county map (toggle between) or as a grouped bar chart by region
   2. Obesity, heart disease, diabetes, hypertension, high cholesterol
3. Smokers
   1. Group by region similar to other charts

### Quality of life

1. Medical facilities
2. Farmer Markers
3. Libraries

## Policy Map

Starting in 2016 - From each session identify

* Bills that are “science related” - try and categorize them into themes (ie. There were a ton of coal ash bills. Gun bills. Etc)
* Who voted which way on important things?
* To score them: 1 point for pro science bill, 0 points for meh, -1 for anti-science bill
* Place this on the map
* Good Orgs to Look at:
  + <https://gabio.org/advocacy/legislative-watch/>

Do we put power plants here?

### Composition

1. House
2. Senate

### Demographics

1. Income
2. Race
3. # of polling places
4. Last presidential election

### Presidential Margin of Victory

* No change

### Percentage of Turnout

### Number of Polling Places per County

* Make it look like the library chart

### Budget Priorities

1. Education Spend
2. Law Enforcement Spend
3. Social Services Spend

### Bills – this is on pause until Quorum access

1. 2017-2018
2. 2019-2020

How do we find a way to provide more information?

Need to fix the headers and titles

Have links to the sources of this – Grouped by Committees - website navigation is tricky

Quorum has a way better search feature – use this in Quorum as well -

# Infrastructure of Data Map

## Hosting

## Data Structure

### Data Dictionary

# Questions from Rob

KEY to Highlights:

Demographics Map & SoS GA

Policy Map

All

What online hosting options are available to you / preferable to you?

              AWS/Azure – preference for Azure since we have more available credits there.

What hosting experience would you ideally like for an enthusiastic newcomer to your project?

              Unclear what this is asking

Would you be interested in organizing a lightweight project management tool around projects involving the dataset? E.g. trello or github's own

              We currently use Microsoft 365’s Planner for project management, but we do already have a [GitHub org](https://github.com/sci4ga) and can see how it would be useful to use a GitHub project for this purpose.

Is it static data at a point in time (or over some period) or will it be refreshed and updated in the future?

Our dataset is static. We would update it at most yearly.

Ideally there would be components that would automagically update at least monthly.

Is the data structured, semi-structured, unstructured?

              Structured

              Unstructured and structured

What format(s) is it currently in?

              .csv

              various

Do you have a detailed schema and/or data dictionary?

              Our source data all have data dictionaries with the exception of data yet to be collected.

              Our structured data will have data dictionaries.

What information do you have about the source(s), reliability / accuracy of the data?

              The source data we currently are all collected by government or recognized institutions. We have data yet to be collected for which the quality will be our responsibility.

              Determining the reliability of given data sets will be part of the project.

Is it at all clear whether some data fields are likely more valuable than others?

              Yes, with some exceptions

              Not immediately clear

What kinds of past qualitative or quantitative knowledge is there about this kind of data or its subject matter? E.g. is there past work on deriving insights from this kind of data? What does a "feature" in this data look like?

              Things like economic and census data are pretty widely understood and some of the questions we intend to answer have existing whitepapers (not by us) that answer in whole or in part, or data that provides an obvious answer. The data that we intend to collect and add to the existing data may provide novel insights.

              Using sentiment as a feature will bring with it all the issues that naturally go along with it. Lots of unknowns.

Do you have a reading list for preparing a volunteer with domain expertise?

              Not at present

Has anyone assessed the nature and extent of any missing or non-compliant data entries?

              Yes, the data quality is high and requires minimal cleaning.

              Data munging will be part of the project.

If applicable, can you imagine having a validation process that provides a robust, reproducible way to know whether entries in the data are "valid" according to some specifications?

              Given it will be done approximately once a year, we expect this will be a process that is documented and iteratively improved upon.

              This will be a fun challenge depending on the sources.

Do you have any already-existent use cases / applications that use the data? E.g. by a past volunteer

              We’re in the process of building our minimum viable product. We know where we’re going with this, but it will serve as the point that we build off of each subsequent year. – the interns did some great work this summer in Tableau.

              We’re starting completely from scratch here.

Do you have any potential project descriptions that go into some detail about how you can imagine the data could be used? You mentioned sentiment analysis in a loose way, so that could be a good example to flesh out with a bit of imagination.

              Our initial interest would be to see if the data available for sentiment analysis would be useful at all, comparing sentiment on social media around, for example, vaccination to what is reflected in polling data. If useful, we would like to determine if it is useful enough to inform advocacy campaigns and make note of whether highly public events that sway sentiment are followed by a reversion to the mean or if they have lasting effects.