JHU 315 Final Project

Natalie Vallejo - 315
nvallej1@jhu.edu
Jonathan Young - 315
jyoun127@jhu.edu

Exploration of COVID-19 and its effects on patient outcomes and economic factors on a national, regional, and state level.

Link to project: https://jonathanjymd.grafana.net/goto/0ceI0C5Vk?orgId=1
Username: jyoun127, Password: grafanajhu315

1. Phase I Information

Our project description and database design has not changed from Phase I. To view the Phase I information and representative list of questions addressed by our system, please click on the following link: https://docs.google.com/document/d/1Rul5aImlCH5llU2GmvObuldIbBv7ifLCnVZgdCZgOcc/edit?usp=sharing

2. Loading Data and Sources

Data Loading

Data was loaded one of two ways.

First, our team created a python script that generated `INSERT` lines to populate a `sql` script. The script was subsequently copied and pasted into the terminal connecting to the database and ran to insert the values. The file can be found in the file directory under the file called `python_scripts/generator.py`. This method was used as a backup and prior to discovery of DBeaver.

The second method was using DBeaver, an application that allows one to insert csv files, modify keys, attributes, and other aspects of the database via GUI. This was the primary method of loading data.

Sources

COVID

Description	URL
CDC Weekly Cases	https://data.cdc.gov/Case-Surveillance/Weekly-United-States-COVID-19-Cases-and-Deaths-by-/pwn4-m3yp
CDC Deaths by Sex and Age	https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-by-Week-Sex-and-Age/vsak-wrfu
CDC Racial COVID data	https://covid.cdc.gov/covid-data-tracker/#demographicsovertime
CDC Racial COVID Deaths data	https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-Distribution-of-Deaths/pj7m-y5uh
COVID Gender and Sex data	https://www.genderscilab.org/gender-and-sex-in-covid19/#DataTable
CDC PCR testing data	https://healthdata.gov/dataset/COVID-19-Diagnostic-Laboratory-Testing-PCR-Testing/j8mb-icvb
CDC Vaccination Data	https://data.cdc.gov/Public-Health-Surveillance/Rates-of-COVID-19-Cases-or-Deaths-by-Age-Group-and/3rge-nu2a

Census

Description	URL
US Census Data Per State	https://www2.census.gov/programs-surveys/popest/tables/2010-2020/state/totals/nst-est2020.xlsx
US Census Racial Composition Per State	https://data.census.gov/table?tid=ACSDT5Y2020.B03002&g=0100000US
Rural vs. Urban Population Per State	https://www.nationalpopularvote.com/rural-states-are-almost-entirely-ignored-under-current-state-state-system

Insurance

Description	URL
Percentage of insured individuals by program by state	https://www.kff.org/other/state-indicator/total-population/?currentTimeframe=0&sortModel={"colId":"Location","sort":"asc"}

Political

Description	URL
Racial composition of Democratic Party	https://www.pewresearch.org/religion/religious-landscape-study/compare/racial-and-ethnic-composition/by/state/among/party-affiliation/democrat-lean-dem/
Overall affiliation by state	https://www.pewresearch.org/religion/religious-landscape-study/compare/party-affiliation/by/state/

Unemployment Factors

Description	URL
Key economic indicators in the time of COVID-19	https://data.bls.gov/apps/covid-dashboard/home.htm
Employment Levels for Different Age Groups	https://www.bls.gov/webapps/legacy/cpsatab9.htm

Characteristics of the Unemployed/Employed

Description	URL
Educational Attainment By Gender, Race	https://www.census.gov/data/tables/2021/demo/educational-attainment/cps-detailed-tables.html
Occupation By Gender, Race	https://www.bls.gov/cps/cps_aa2020.htm#charunem
Multiple Job Holders By Gender, Race	https://www.bls.gov/cps/cps_aa2020.htm#multjob
Remote Work During the Pandemic	https://www.bls.gov/cps/effects-of-the-coronavirus-covid-19-pandemic.htm
Full/PartTime Work Status	https://www.bls.gov/cps/lfcharacteristics.htm#fullpart
Marital Status	https://www.bls.gov/cps/cps_aa2020.htm#charemp

GDP

Description	URL
Monthly National GDP Data	https://www.spglobal.com/marketintelligence/en/mi/products/us-monthly-gdp-index.html
GDP By Industry	https://apps.bea.gov/iTable/?reqid=150&step=2&isuri=1&categories=gdpxind
GDP By State	https://apps.bea.gov/itable/?ReqID=70&step=1
GDP Factors/Components	https://www.bea.gov/data/gdp/gdp-state

State Characteristics

Description	URL
Personal Income and Its Outlays	https://apps.bea.gov/iTable/?reqid=19&step=2&isuri=1&categories=survey#eyJhcHBpZCI6MTksInN0ZXBzIjpbMSwyXSwiZGF0YSI
Personal Consumption	https://apps.bea.gov/itable/?ReqID=70&step=1#eyJhcHBpZCI6NzAsInN0ZXBzIjpbMSwyNCwyOSwyNSwzMV0sImRhdGEiOltbIlRhYmxlSWQiLCI1MjQiXSxbIk
Top Industries in Each State	https://ssti.org/blog/useful-stats-top-industries-state-net-establishment-and-job-creation-2005-2019

Description	URL
Poverty Information and Median Household Income	https://www.census.gov/data/datasets/2020/demo/saipe/2020-state-and-county.html

3. MySQL Server

Our team utilized an AWS RDS (Amazon Web Services Relational Database Service) utilizing a MySQL Community engine. Hardware specifications goes as follows:

Compute	Value
AWS Instance Type	db.t3.micro
vCPU's	2
Memory (GiB)	1
Processor	Intel Skylake E5 2686 v5 (2.5 GHz)
CPU Architecture	64-bit

4. Utilizing Our Project

Our team decided to present our data utilizing Grafana, an open source analytics and interactive visualization program.

For the purposes of demonstration, to access our dashboards, utilize the following link to access the landing page.

> 💡 NOTE: The credentials below will grant READ ONLY access.

Landing Page Link

https://jonathanjymd.grafana.net/goto/0ceI0C5Vk?orgId=1

Credentials to Login

Username: jyoun127

Password: grafanajhu315

> 💡 In the event that the above credentials do not work, please refer to the below for snapshots of our project.
>
> ⚠️ Snapshots generally provide the same level of functionality as normal dashboards, except that the drop down menus cannot be selected. Please see the next section in regards to how to utilize the functionality of the dashboards.
>
> Snapshots can be accessed to **anyone.**

National COVID Age/Gender Trends

Grafana

If you're seeing this Grafana has failed to load its application files 1. This could be caused by your reverse proxy settings. 2. If you host grafana under subpath make sure your grafana.ini root_url setting includes subpath. If not using a reverse proxy make sure to set serve_from_sub_path to true.

https://snapshots.raintank.io/dashboard/snapshot/2TosRfjQFGirOjF1rdduDNKgO85kdUR9

Deeper Dive Into Private Sector GDP

Grafana

If you're seeing this Grafana has failed to load its application files 1. This could be caused by your reverse proxy settings. 2. If you host grafana under subpath make sure your grafana.ini root_url setting includes subpath. If not using a reverse proxy make sure to set serve_from_sub_path to true.

https://snapshots.raintank.io/dashboard/snapshot/g7evdgCMgjjc01RogEUpBJ1VHEN5LFiq

National COVID Statistics and Patient Outcomes

Grafana

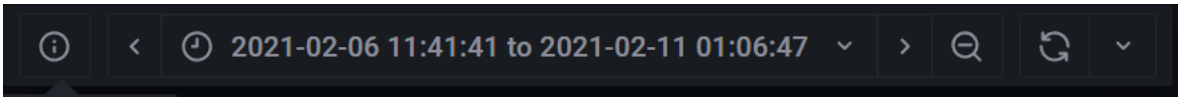
If you're seeing this Grafana has failed to load its application files 1. This could be caused by your reverse proxy settings. 2. If you host grafana under subpath make sure your grafana.ini root_url setting includes subpath. If not using a reverse proxy make sure to set serve_from_sub_path to true.

https://snapshots.raintank.io/dashboard/snapshot/0Paa6NwlRwU3Bd12VdcGCxo0PLZK5uC5

3. Some Shortcuts

Changing Viewing Period

To view a specific time period, navigate to the three dot menu to the top right. The following menu should appear



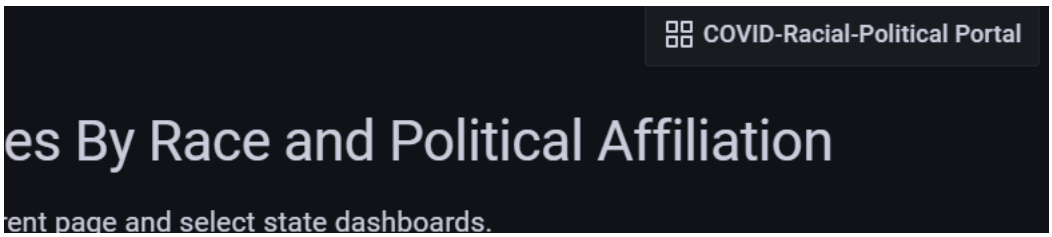
By the time listed above, click the drop down arrow to select another time frame to view. You can also click the minus button to select a wider field of time to view. The arrows on either side of the time can be utilized to navigate to future or past time periods.

These time periods are only valid for dashboard components that utilize `time series` (most likely in the form of line graphs). Not all graphs may have data presented for the time period selected.

You can also manually zoom into the data set by highlighting a specific region that that you would like to examine. This can be done by clicking and dragging a specific window on the time graph.

Navigation

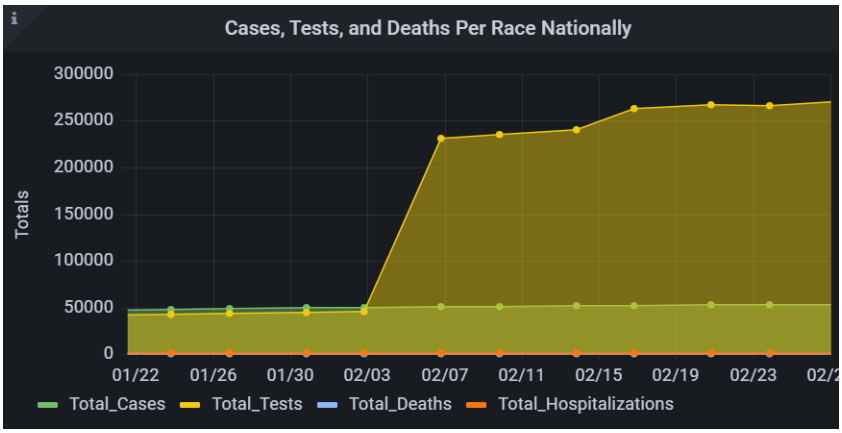
To navigate between the dashboards, either click on the button in the top right to navigate to other dashboards or click on `JHU-315-Final-Project` on the top left. The former will redirect the user to the parent page. The latter links to the parent folder containing all dashboards. Below is a screenshot of the former method, in which `COVID-Racial-Political-Portal` will direct you to the parent page.



Focusing on Select Datasets

In line graphs, there's often a myriad of data points that appear. To focus on one specific dataset, simply click on the dataset desired. An example is presented below:

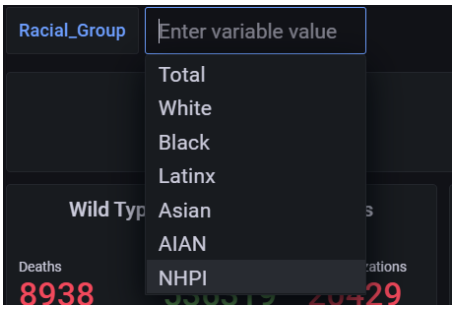
In this example, `Total_Cases`, `Total_Tests` , `Total_Deaths` , and `Total Hospitalizations` are the data sets presented. To view, for example, only `Total_Cases` , simply click on



`Total_Cases` . The same applies for the other data sets.

To view select datasets in conjunction with each other, hold the `shift` button and click on the datasets you wish to present.

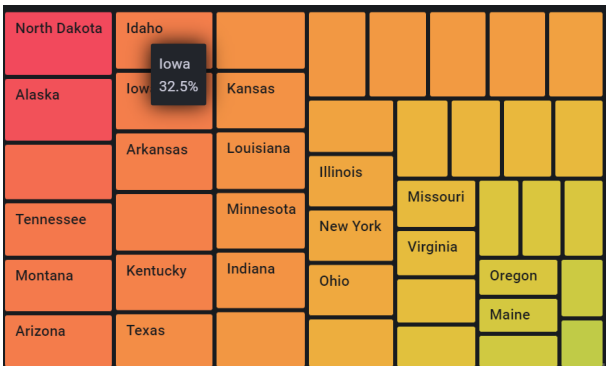
Toggling Filters



On some dashboards, there are filters on the top left section that enable a user to select a specific group. This can include states, COVID variants, and other metrics. To toggle them, navigate to the top left and click the drop down menu. An example is on the left.

Due to current limitations of MySQL, multiple values cannot be selected at once per drop down menu. However, multiple drop down options can be configured at once.

Map Features



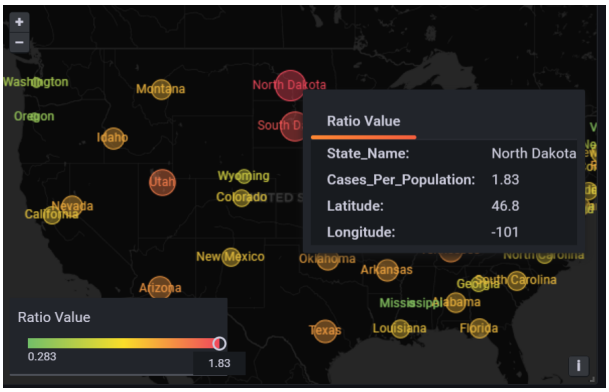
Data Hover

Other dashboard components usually have additional information to present itself with if hovered over. For instance, the map tile feature, in the dashboard `National Trends By Race` , can be hovered over to yield numerical results.

Layer Data Retrieval

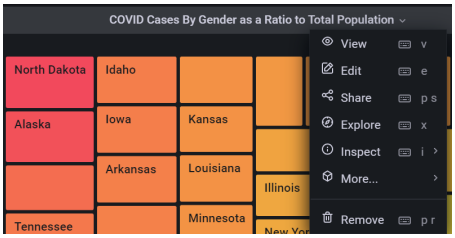
To view more data regarding on a geo map, click on the corresponding data point. A pop up will occur similar to the figure on the right.

> 💡 Note: Due to Grafana limitations, the latitude and longitude cannot be removed and are extraneous to present.



In general on all maps, there is a zoom in and out button for to navigate the map. A user can also click and drag to navigate as well.

Misc.



Panel Drop Down

To examine a dashboard component more closely, click the drop down at the top of the dashboard component. A menu appears.

Below are the list options that are available for the user.

`View` button enables the user to examine the panel on a larger screen.

`Inspect` button provides the user the option to…

- View the the raw data from the SQL Query unformatted by clicking the `Data` button
- Examine the SQL Query utilized to fetch the data by clicking the `Query` button

Additional Information

Further information on a panel can be retrieved by clicking the `i` button.

> Note: for almost ALL Grafana dashboard features, there's an option to hover over the panel for more information.

4. Focus/Specialization

Our team chose to focus in the following areas:

Visualization and Data Interaction

The primary focus on the project was creating a GUI.

In our end product, the queries and resulting visualizations provide an highly in-depth analysis on COVID-19's impact on key economic indicators and patient outcomes, performing analysis at the national, regional, and state levels.

Using Grafana enabled our team to explore creating interfaces to present meaningful data and in a unique way that best addresses questions regarding COVID, economics, and patient outcomes. For some dashboard components, there was a challenge of making the data presented highly interactive and intuitive.

The dashboards also had to be designed for versatility as well. There are many data points and sets to present, and the dashboard's functionality had to adapt to that.

Cloud Computing - Performance and Security

Using Amazon Web Services (AWS) relational database service (RDS) enabled our team to explore acquiring machines that balanced performance and cost. There were a variety of instances (machines) to choose from, and our team needed to ensure that the data processing required by Grafana dashboards could be effectively handled by the instance. There were a few instances in which our instance had to be rebooted due to the complexity of certain queries overwhelming the system.

On the security side of things, AWS RDS has many features to protect and and maintain the access of the database. There were two sides of security that had to be configured:

Inbound Rules - Our team had two layers of security that prevented unauthorized use. First, users were created with certain permissions to access the different databases on the instance. Passwords and usernames were also created. Second, only users from a certain IP could access the database.

Outbound Rules - Our team had one layer of security on this end. To secure who is reading the data, our team had the option of filtering who can retrieve information from the database.

MySQL Integration

Linking Grafana to our MySQL database was not a trivial task, as there are many limitations to MySQL that hindered the presentation of data. Therefore, our team had to process the data uniquely via tools such as Google Sheets to optimize the data and explore using unconventional MySQL methods to adapt to Grafana.

5. Strengths Of the Project

Data Advantages

Highly Extensive and Comprehensive Data

Our project explores a variety of datasets from a variety of angles. As such, our MySQL database contains over 100 tables of uploaded data as well as interesting views and conversions of the data. Integration of these datasets was also challenging as we aimed to cover both economics and COVID information.

Adaptability

The dashboard and system as a whole is designed in such a way that users can toggle what information they would like to see. This is thoroughly exemplified in the above sections, where numerous tools and options gives the user an option to adjust the perspective of the data.

Technological Advantages

Versatility

Choosing Grafana enabled our dashboards to be hosted on a variety of mediums. For instance, if one desired to deploy the dashboard on a local instance or their own server, it is incredibly easy to do so. Our dashboards are currently hosted on a Grafana itself. However, one can easily make a copy of our dashboards, modify it for their use, and host the dashboards on their own systems.

This is practical for institutions and entities who desire to present our dashboard on a variety of platforms.

AWS RDS Scalability

AWS RDS enables a scalability if handling large volumes of transactions. Our database instance can be scaled if necessary.

Adaptability

Grafana provides a variety of libraries that contain numerous ways to present data. Our team chose to utilize a few of them to show different angles of the same datasets purely from a visualization standpoint. In summary, our dashboards can be adapted for a variety of datasets, and easily manipulated to show different angles of the dataset.

Ease of Use

Grafana combined with our dashboard/system design is easy to use, making it efficient for the user to access the data they require or compare data that the user desires.

6. Limitations

MySQL/Grafana Integration

Integrating MySQL and Grafana was not easy, as there were many limitations from MySQL that hindered production of certain dashboards. Some functionality that was not doable with MySQL included having multiple parameters for a filter.

MySQL datatypes are not as easy to work with with in Grafana, requiring certain hacks in order Grafana to recognize a date for instance in time series.

The above limitations significantly diminished the capability to present data in a certain type of way.

Improvements to this issue would to use database engines that enable increased versatility.

AWS RDS Limitations

Due to cost limitations, our team chose not to utilize a more powerful instance or scale the instance. There were a few instances where the instance maxed out at 100% utilization, which required a reboot to reset the system.

Rolling Data

Our data isn't rolling. Meaning that there is no automated system to retrieve up to date information about the pandemic or economic data. Therefore, the information is static.

Improvements could improve developing a system that references data sources' API to fetch data on a daily basis.

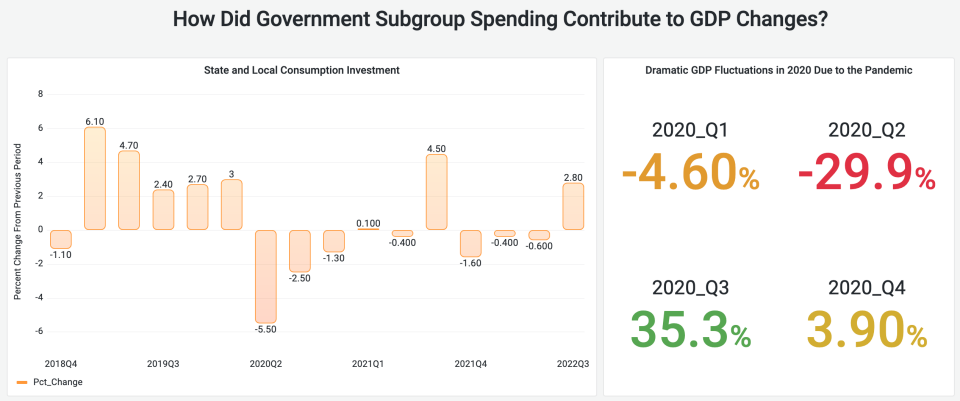
Grafana Limitations

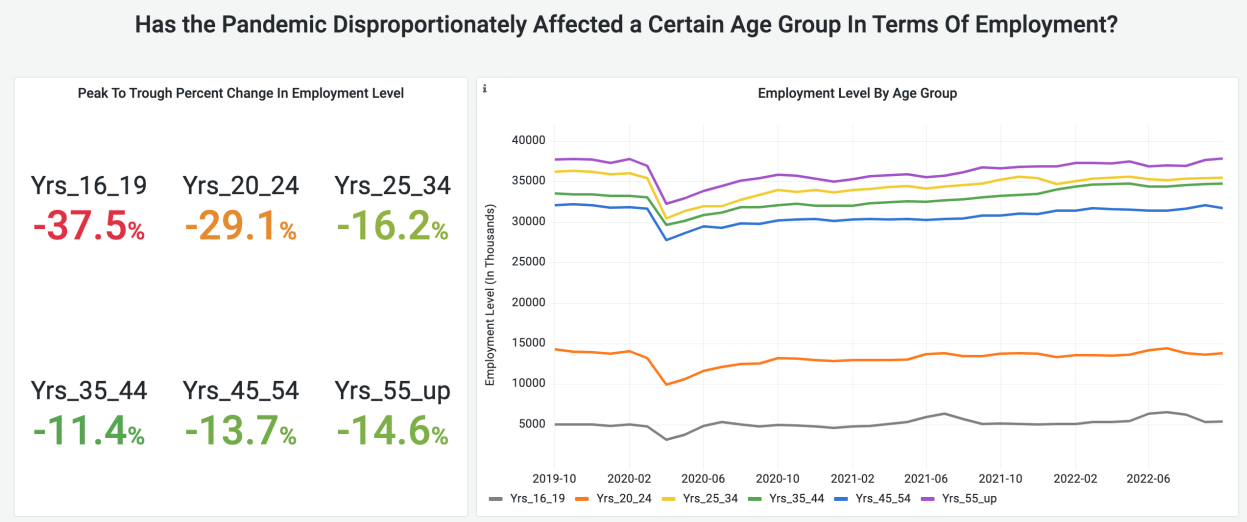
There are certain actions on Grafana that are not doable, preventing our data from being adequately presented in a manner that is most appropriate for a particular application. These limitations are often attributed to the limited features found in the library of panels.

Improvements are not doable from our end and our team believe these issues will be resolved as Grafana matures.

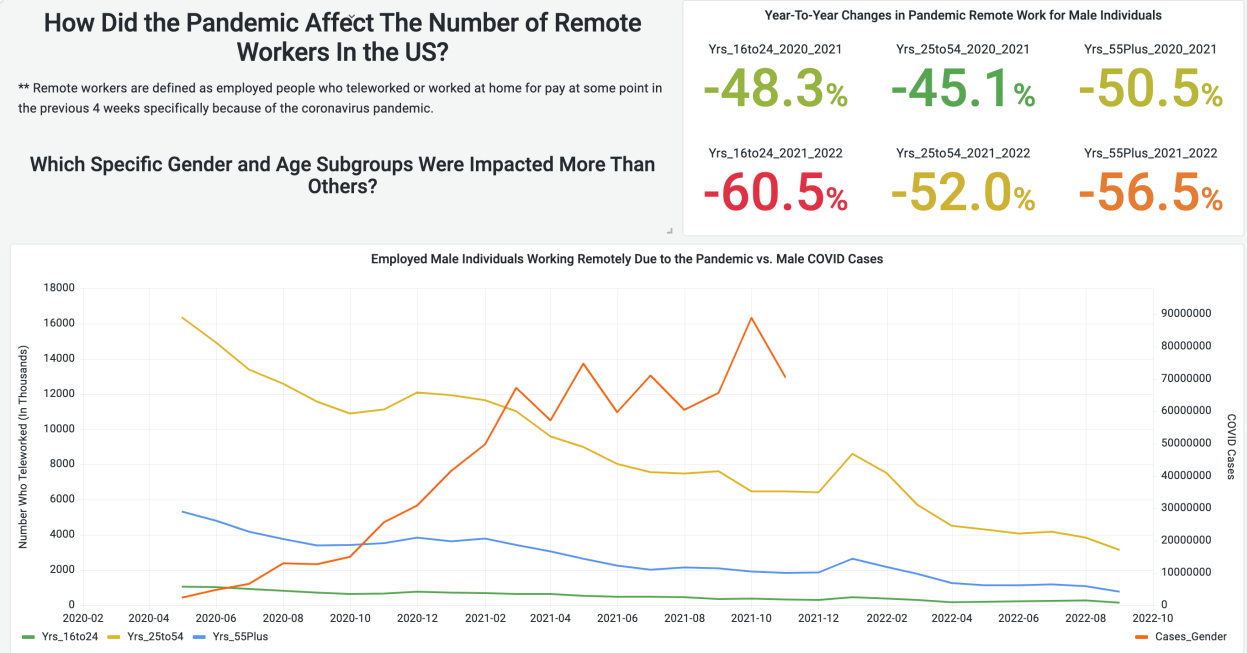
7. Highlighting Dashboard Aspects

Our dashboard integrates graphical data with numerical statistics computed from queries in an understandable and clear way. This combination of outputs helps to emphasize the trends that can be seen in the graphs.

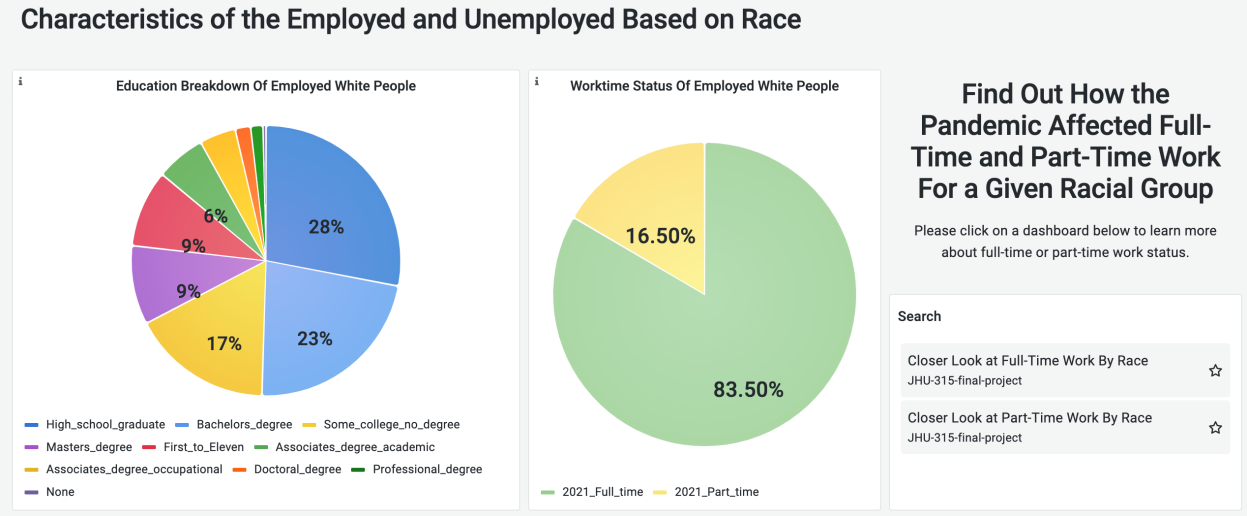




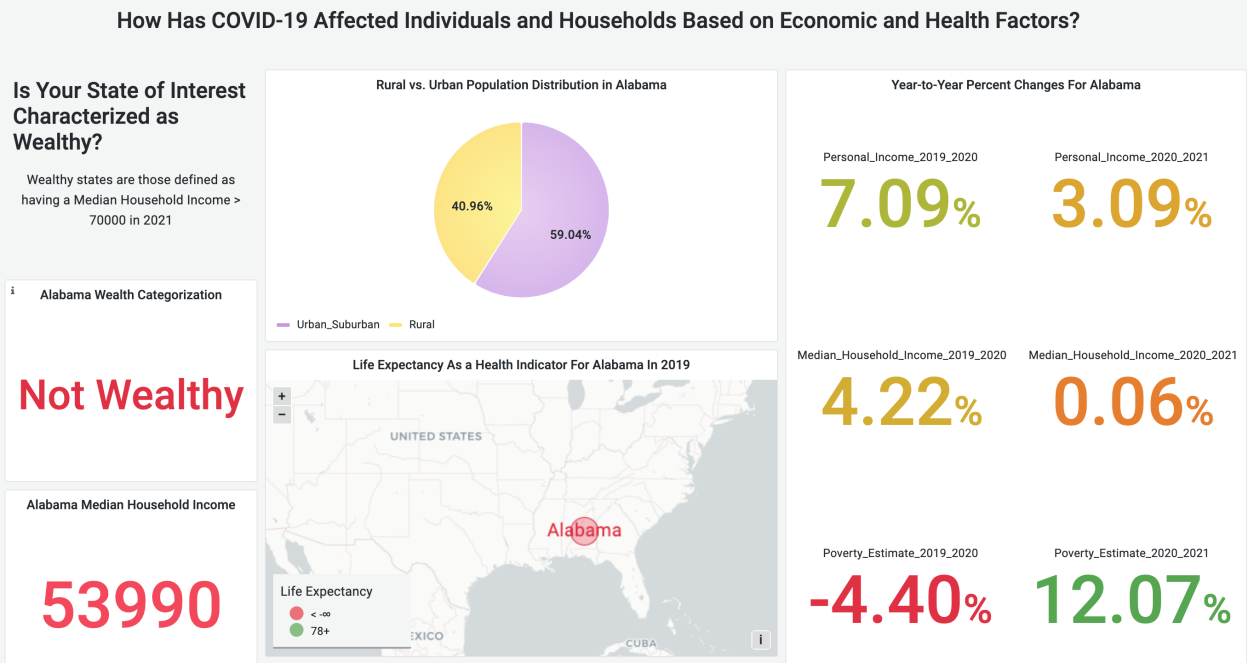
Our dashboard also includes lots of user friendly explanations on what questions are being highlighted in a given section as well as directions on which drop-downs are available to use.



Our dashboard also provides means for users who want to know more to navigate to different sub-dashboards to gain more interesting insights.



The dashboard provides a mix of visual displays to make the data more interactive, approachable, and appealing.



8. Misc.

Link to just the user guide can be found with the link below

User Guide

Description of the files in this directory can be found in the `README.md` file.