

Nate Everett
CPSC408
May 22, 2021

Covid-19 Database and Analytics Dashboard

Introduction

The Covid-19 pandemic took the whole world by storm in the beginning of 2020. Since then, data analysts, psychologists, medical professions, and business owners have tracked billions of points of data expanding over broad topics relating to the disease. As per usual in this world of big data, having large amounts of data can lead to key insights to drive success in any of the aforementioned fields. A business owner may want to track data to see how Covid has impacted sales in heavily restricted zones. A pharmaceutical company may want to analyze the biological protein structures to create a vaccine. A psychologist may want to know the impact that shelter-in-place has on teenagers with pre-existing mental health conditions. The possibilities for insight during this pandemic are infinite, but we need to be able to actually access and utilize the data correctly in order to arrive at these insights. Unless you have a well-versed data scientist or analyst on your team, tracking and analyzing important data may seem like a daunting task.

The Problem: Inaccessibility to Data

Data is the new currency. Covid has generated hundreds of millions of dollars worth of data and a large portion of it is sitting waiting for someone to use it. As was previously mentioned, unless any team looking to use Covid data for any reason has a professional data scientist on their team, it is probably unlikely that any valuable results are going to arise. This will be, and has proven to be, a common struggle among fields that may have a hard time

managing large sums of data. For instance, a group of psychologists may be familiar with gathering data from 30 or even 100 participants and running analysis of variance tests to see how results change with different parameters. But what happens when this same team is given millions of participants with hundreds of parameters. The methods fall apart.

Without having access or even working knowledge of big data tools, the entire world suffers the consequences of the valuable results that could come from actionable analysis in these data. This is why businesses with infinite resources to pour into data like Amazon, Facebook, and Google have seen business operations and profits soar, while smaller brick and mortar shop owners are filing for unemployment. The issue is inaccessibility and failure to optimize and utilize readily available data.

Solving the Problem: Accessible Interfaces

The obvious problem to this solution is to provide an easy-to-use user interface that allows individuals without a serious technical background to access this data. While the data engineering side of operations might require some outside sourcing, accessibility and analysis of the data an individual is interested in should be virtually automated by this interface.

Many companies and even skilled individuals have already come up with data dashboards that can provide valuable insights at the click of a drop down menu. For instance, the UX Collective¹ team has provided a solution that shows aggregated data and visual analysis just by clicking a few drop down menus and buttons. Many of the existing dashboards and analytics

¹ <https://uxdesignn.cc/covid-19-coronavirus-data-visualisation-dashboard-3a15ed76eb7e>

tools are currently only for internal use, especially any of the ones worth mentioning. It is for this reason that there is a need for a more accessible interface.

One issue to be weary of in collecting data in a space like this is data privacy, HIPAA, and GDPR. Covid has affected the entire world, and usually each country has its own laws and regulations regarding data security and privatization. Should the user interface choose to show information about Covid patients from Germany, the interface most definitely needs to take into account the incredibly strict and rigorous data regulations of the EU. Should the user interface show information about patients in the United States, the information must be anonymized by law according to the Health Insurance Portability and Accounting Act. Should the user interface show information of many or even all countries, these sorts of data protection measures must not be understated.

Using R's Shiny to Provide Easy Visual Access to Data

The scripting language R provides a dashboard library called Shiny. Users can access the methods of the Shiny API to construct dashboards relatively simply while still being able to call upon all of the other functionalities of R, which happens to be an incredibly powerful tool for analysis. This solution seeks to provide a seamless interface that can connect to a SQL database and generate instant results in the form of graphical and tabular outputs.

While not fully developed at its current stage, the Covid Dashboard connects to a SQL database in the cloud hosted by Google Cloud Platform. We can query this database through the

use of R's DBI package which is a library that connects R and SQL together. The database in current use consists of five tables to provide the most general Covid data: *Country*, *Country Statistics*, *Person*, *Individual Statistics*, and *Vaccine*. From here we created an easy to use user interface with plenty of drop down menus and buttons to serve the functionality needs of the user. The user can dynamically select the country or vaccine type data that they are interested in and we will, utilizing DBI, send a query to that database depending on the input of the user and return the results. If the user finds something particularly interesting that they would like to do further analysis with, the application can generate a report of the data they queried and they will find a comma-separated value file containing the results of that query.

In addition to visualizing the data, this dashboard also serves as a traditional CRUD application. Say the user has collected new data or would like to make any updates to existing data, or maybe even the European Union is chasing them down to remove data violating GDPR. There is an interface that, with only small working knowledge of SQL and mostly a working knowledge of a drop down menu, can solve all of these needs at the click of a create, update, or delete button.

Conclusion

To recap, data is the new currency. While the Covid pandemic has been chaotic and destructive to almost all aspect of the operations of society, if anyone could be an investor in "data," they would surely have reached El Dorado with the presence of Covid. The problem everyone except the large tech moguls face is how to deal with all of the newly-generated data. Smaller businesses and researchers with minimal funding are starved of the possibilities for

cutting edge research. Our solution seeks to alleviate some of this inaccessibility by providing an easy to use interface for all Covid data related needs. Thanks to the dashboarding tools provided by Shiny and the SQL integration tools provided by DBI, small businesses and underfunded research facilities no longer need to painfully watch tech giants steal all of the wealth.