Data Visualization Tools in Industry

Gaoyi Shi (gs3012), Xiaonan Guo (xg2308)

Background

We will Introduce several data visualization tools currently used in industry. Companies use different data visualization tools based on their business purposes and requirements. Based on our daily work and some research on our friends' companies, we have picked the 4 most popular data visualization tools to demonstrate how the industries are using different tools for graphical data analysis. We strongly suggest that our classmates take some time to make themselves familiar with those tools which will be beneficial for seeking data scientist related jobs.

Content

Part1: R/Python Plotly

Part2: R Shiny App

Part3: AWS QuickSight

Part4: Tableau

Appendix: Python code

Part 1: R/Python Plotly

Plotly is a free and open-source graphing library for making interactive graphs with R, Python, MATLAB, and Excel. The plotly library is an interactive, open-source plotting library that supports various chart types covering statistical, financial, geographic, scientific, and 3-dimensional use-cases.

Built on top of the Plotly JavaScriptlibrary (plotly.js), plotly enables both R and Python users to create beautiful interactive web-based visualizations. Here, we will show the examples in both R and Python languages, which are very much similar to each other based on user's preference for which language to choose.

Plotly Package Installation

Python: `# conda install -c plotly plotly=4.12.0`

R: `devtools::install_github("ropensci/plotly")`

Examples in Python

In the attached Jupyter Notebook, we showed plots for some commonly used charts, including histogram, box plot, parallel coordinate plot, heatmap and some interesting graphs, like map plot and 3-D plots.

Sources:

https://plotly.com/r/getting-started/

https://plotly.com/python/getting-started/#overview

Part 2: R Shiny App

What is an R shiny app?

- An app created using R code

- Shiny has a lot of opportunities, but the basic use of it is to allow the user to specify specific inputs such as filters or other selections, which results in an interactive output, such as a graph, table, or text that is based on these inputs

- These are some examples of inputs for PI's shiny app, described in the next slide

Self Insured vs. Fully Insured

Both	▲

Both

Self Insured

Fully Insured

Service Location State

All States
Alabama
Alaska
Arizona
Arkansas
California
Colorado

Provider PINs separated by commas

Range of Paid Amount

257   52,038   100,000

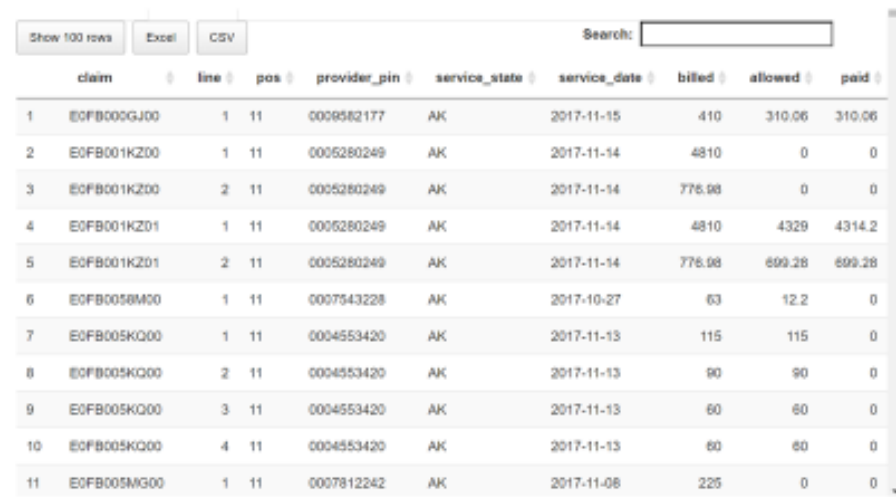-7,000 3,700 14,400 25,100 35,800 46,500 57,200 67,900 78,600 89,300 100,000

Proprietary
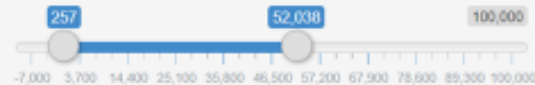
Used it as Database front end for Non-tecnical users

- This specific app has a variety of data options for the user to specify, such as filters, list of fields to view, and fields to order the data by

- Once the resulting table loads, the user has the option to export the table to Excel

- This allows users who are not familiar with SQL to pull their own data and work with it in excel

Show 100 rows	Excel	CSV				Search:			
	claim	line	pos	provider_pin	service_state	service_date	billed	allowed	paid
1	E0FB000GJ00	1	11	0009582177	AK	2017-11-15	410	310.06	310.06
2	E0FB001KZ00	1	11	0005280249	AK	2017-11-14	4810	0	0
3	E0FB001KZ00	2	11	0005280249	AK	2017-11-14	776.98	0	0
4	E0FB001KZ01	1	11	0005280249	AK	2017-11-14	4810	4329	4314.2
5	E0FB001KZ01	2	11	0005280249	AK	2017-11-14	776.98	699.28	699.28
6	E0FB0058M00	1	11	0007543228	AK	2017-10-27	63	12.2	0
7	E0FB005KQ00	1	11	0004553420	AK	2017-11-13	115	115	0
8	E0FB005KQ00	2	11	0004553420	AK	2017-11-13	90	90	0
9	E0FB005KQ00	3	11	0004553420	AK	2017-11-13	60	60	0
10	E0FB005KQ00	4	11	0004553420	AK	2017-11-13	60	60	0
11	E0FB005MG00	1	11	0007812242	AK	2017-11-08	225	0	0

For Visualization

R Shiny app also can be used as graphical analysis, please refer to the following website for Demonstration, they have the dashboard, Github and R code which you can learn by yourself.

Examples

1. COVID-19 tracker by Edward Parker

In December 2019, cases of severe respiratory illness began to be reported across the city of Wuhan in China. These were caused by a new type of coronavirus, and the disease is now commonly referred to as COVID-19. The number of COVID-19 cases started to escalate more quickly in mid-January and the virus soon spread beyond China's borders. This story has been rapidly evolving ever since, and each day we are faced by worrying headlines regarding the current state of the outbreak.

In isolation, these headlines can be hard to interpret. How fast is the virus spreading? Are efforts to control the disease working? How does the situation compare with previous epidemics? This site is updated daily based on data published by Johns Hopkins University. By looking beyond the headlines, we hope it is possible to get a deeper understanding of this unfolding pandemic.

An article discussing this app was published in The Conversation. The map was also featured on the BBC World Service program Science in Action.

This app is updated once daily. For more regular updates, see the JHU COVID-19 dashboard.

View app

View code

2. New Zealand Trade Intelligence Dashboard by [Wei Zhang](#)

The New Zealand Trade Intelligence Dashboard provides up-to-date annual information on trade by commodities, services and trading partners. It presents a full picture of New Zealand?s trading profile through intuitive and interactive graphs and tables. It also provides powerful functionality enabling users to generate their own reports for different commodities, and markets or groups of markets that interest them. All figures, tables and data are downloadable. It works on both PCs and mobile devices.

The philosophy of user-centered design is executed throughout the entire UI and functionality design of the shiny app. There will be less than three clicks away for almost any of user's needs.

The shiny app has the potential to be easily adapted to generate trade intelligence for other countries.

View app

View code

Try it on RStudio Cloud

Sources:

https://shiny.rstudio.com/gallery/

https://shiny.rstudio.com/gallery/covid19-tracker.html

https://shiny.rstudio.com/gallery/nz-trade-dash.html

Part 3: AWS QuickSight

Amazon QuickSight is a business intelligence tool based on the Amazon Web Services (AWS) platform. QuickSight features data visualizations, interactive dashboards, and machine learning insights. According to Amazon, it powers the QuickSight platform with its Super-Fast, Parallel, In-Memory Calculation Engine, or SPICE. The QuickSight SPICE engine interacts with data from multiple AWS sources such as Redshift, Athena, S3, RDS, and Aurora; and other cloud data sources, including SaaS applications such as Salesforce, Square, ServiceNow, Twitter, Github and JIRA; third-party databases such as MySQL, Postgres, and SQL Server; Users may also upload their own excel, csv and JSON files.

Companies using AWS cloud service may also use this data visualization tool to quickly import data and do data visualization. For personal use, we will need register an AWS, which is paid for use.
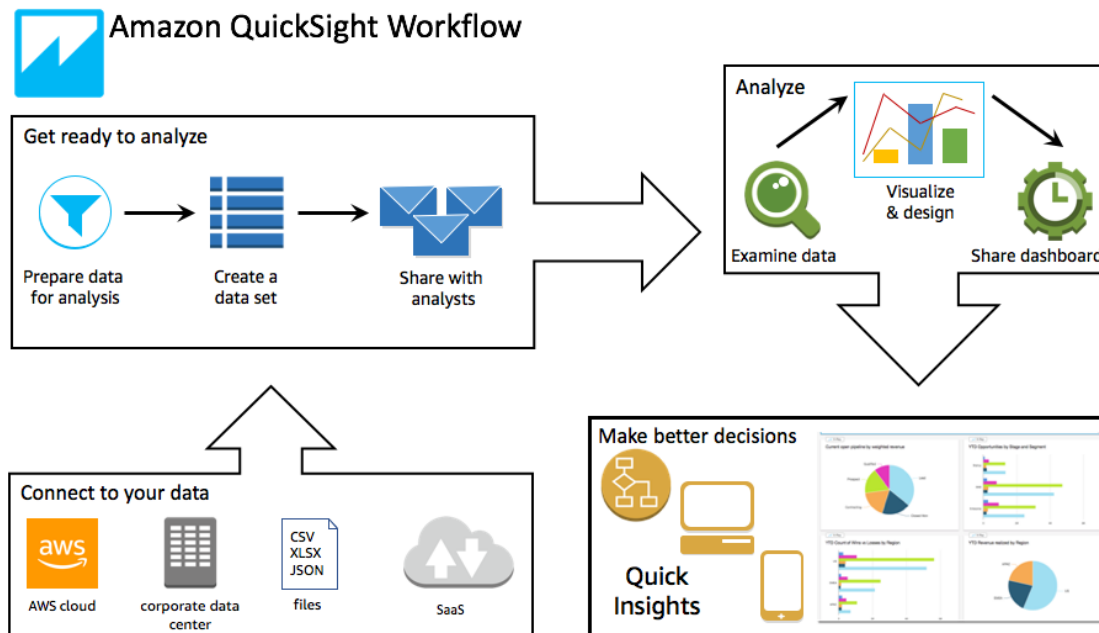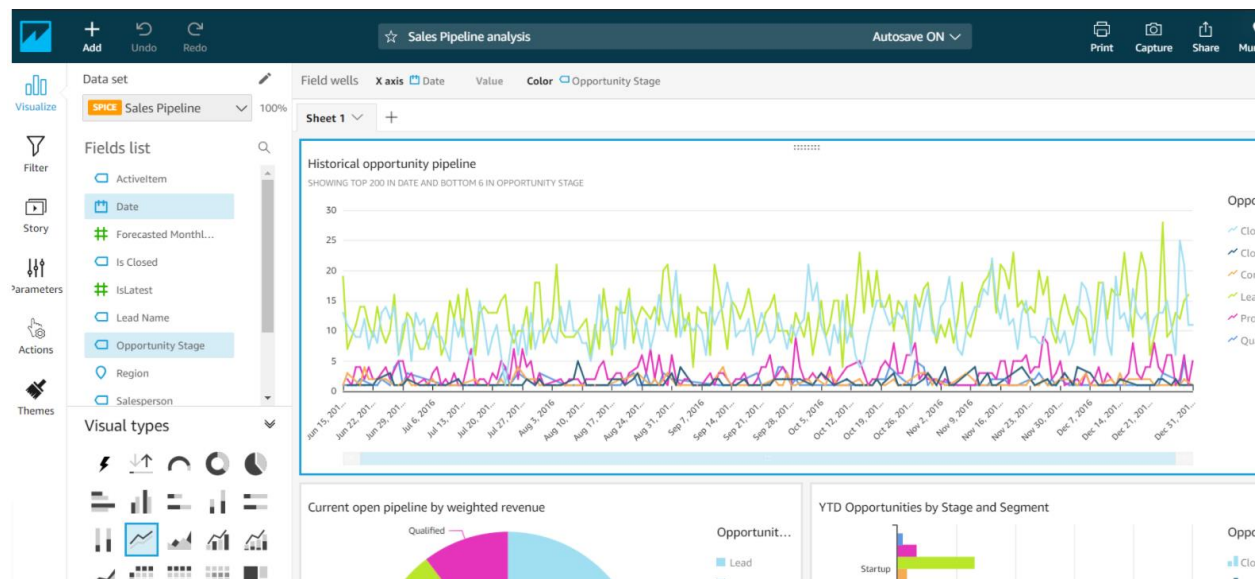
If we have AWS account, we can log into console and find QuickSight. Go to the main dashboard, we will see several templates ready for use.
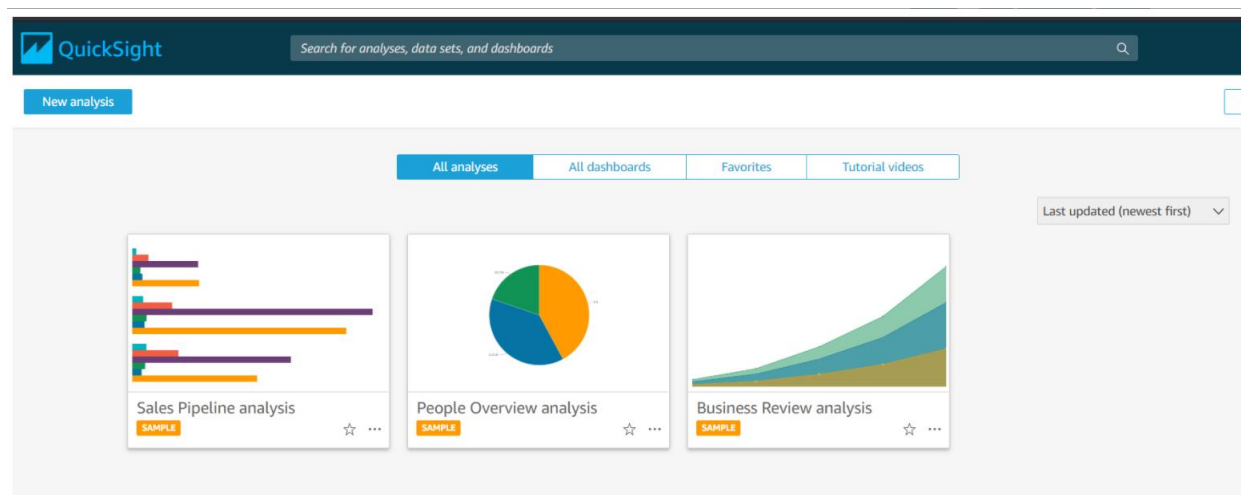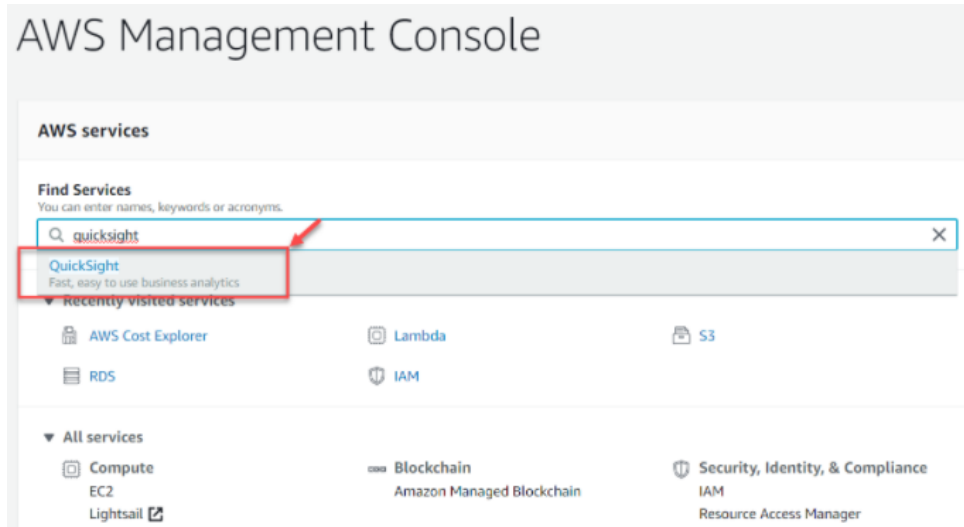
For example, choose Sales Pipeline, we will see the dashboard as below, which is the layout of an Amazon QuickSight analysis. We can see many icons that are ready for use to visualize the report. It gives the option to define custom themes and color schemes so that the reports can be blended into the same theme as web application or the branded look of the organization. The Dataset pane displays a list of all the available fields that are categorized according to the data types. You can use these fields to create the various analyses as and when required.

Visuals will be created on the right-hand pane where you can adjust the size of the visuals by just dragging the right handles of those visuals. You can also make two or three visuals appear side by side by just adjusting the width of each of those visuals.

Sources:

https://www.sqlshack.com/getting-started-with-amazon-quicksight/

https://docs.aws.amazon.com/quicksight/latest/user/how-quicksight-works.html

Part 4: Tableau

Tableau helps people and organizations be more data-driven

As the market-leading choice for modern business intelligence, our analytics platform makes it easier for people to explore and manage data, and faster to discover and share insights that can change businesses and the world.

Everything we do is driven by our mission to help people see and understand data, which is why our products are designed to put the user first—whether they're an analyst, data scientist, student, teacher, executive, or business user. From connection through collaboration, Tableau is the most powerful, secure, and flexible end-to-end analytics platform.

In my team, we use the tableau to track all the medical fraud, waste abuse and overpayment machine learning models and queries' performance. It is an interactive dashboard which you can use it to check different items.

The bar charts show the trend
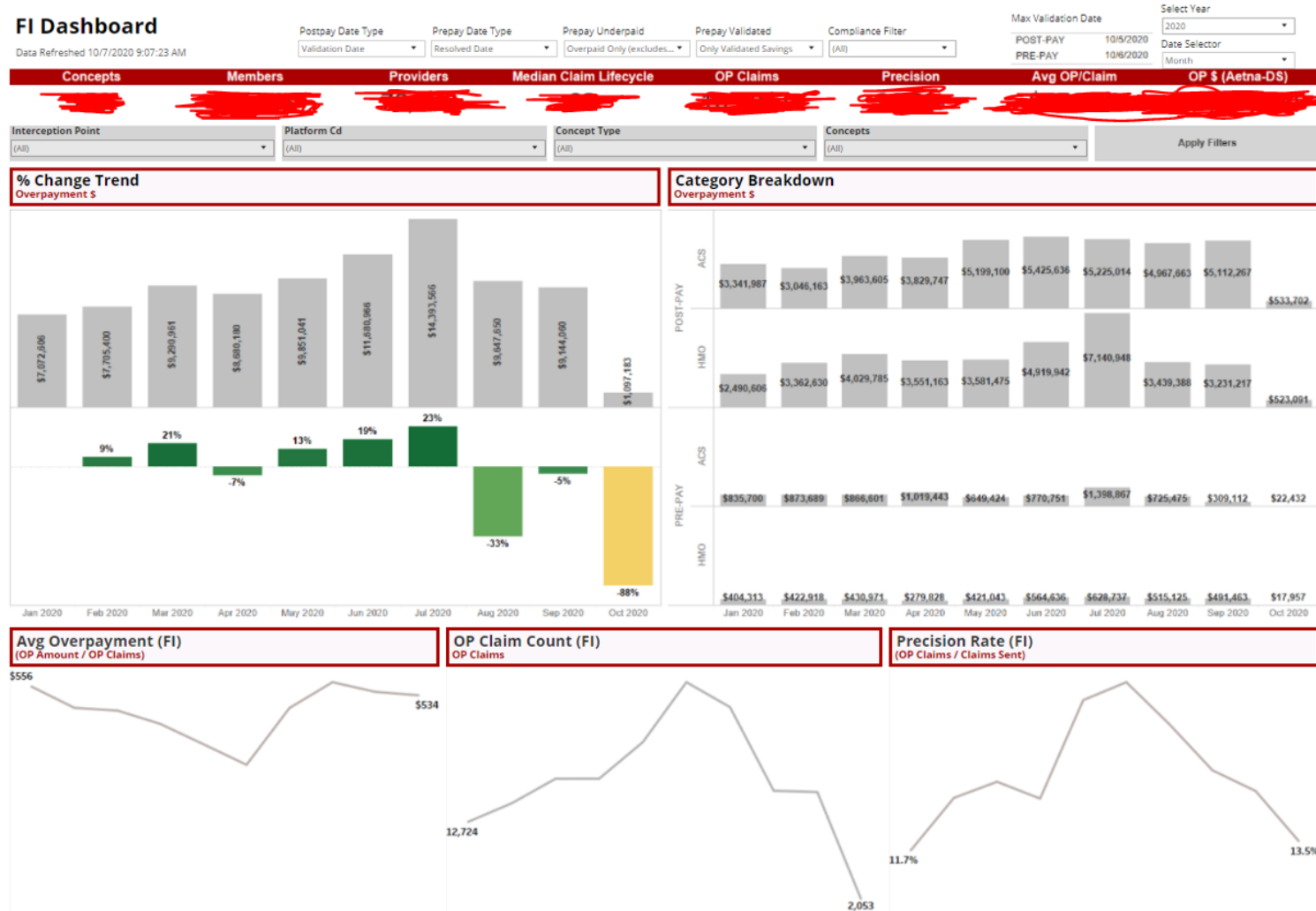
Models and Queries performance by time (Time series)
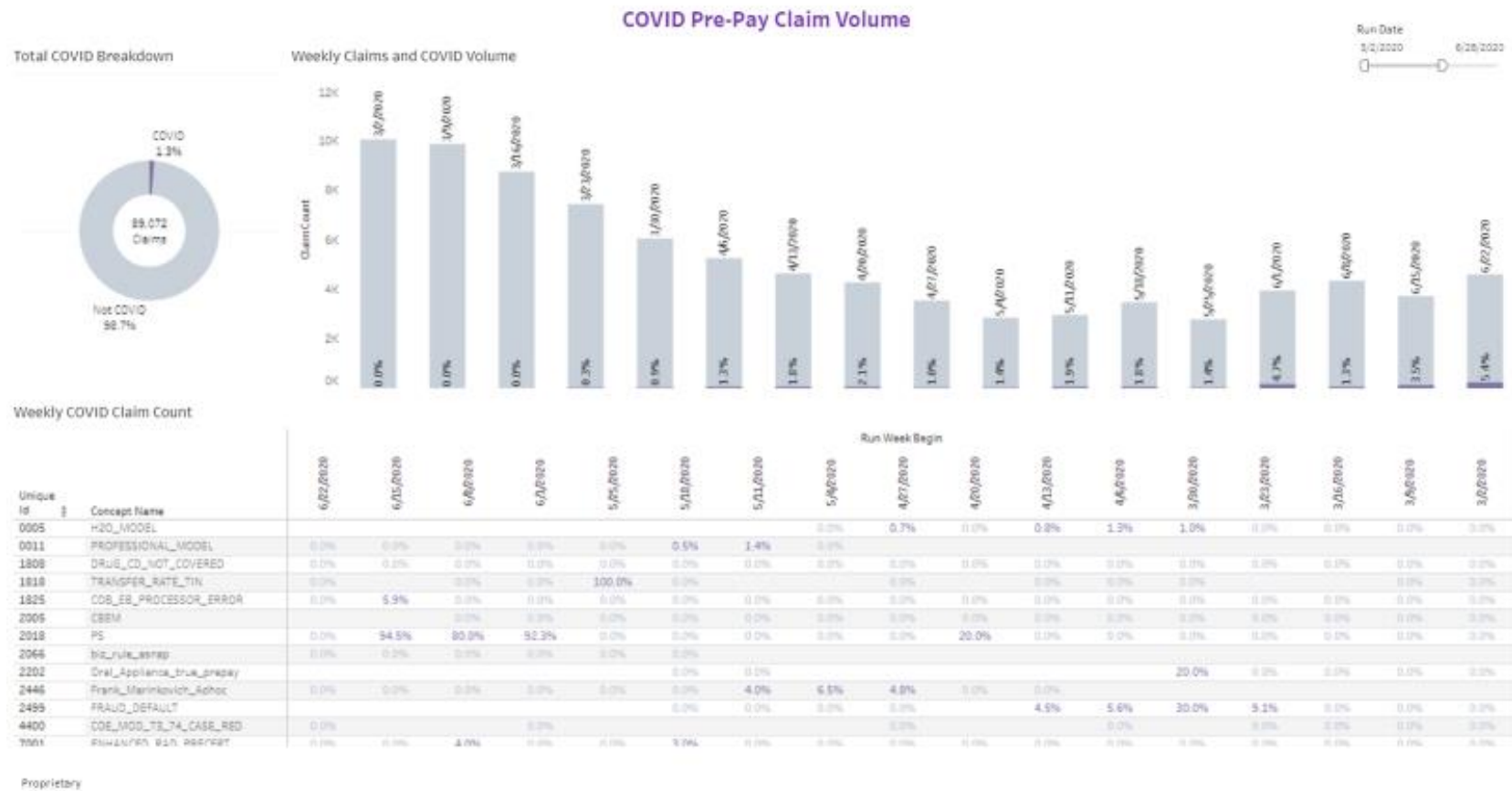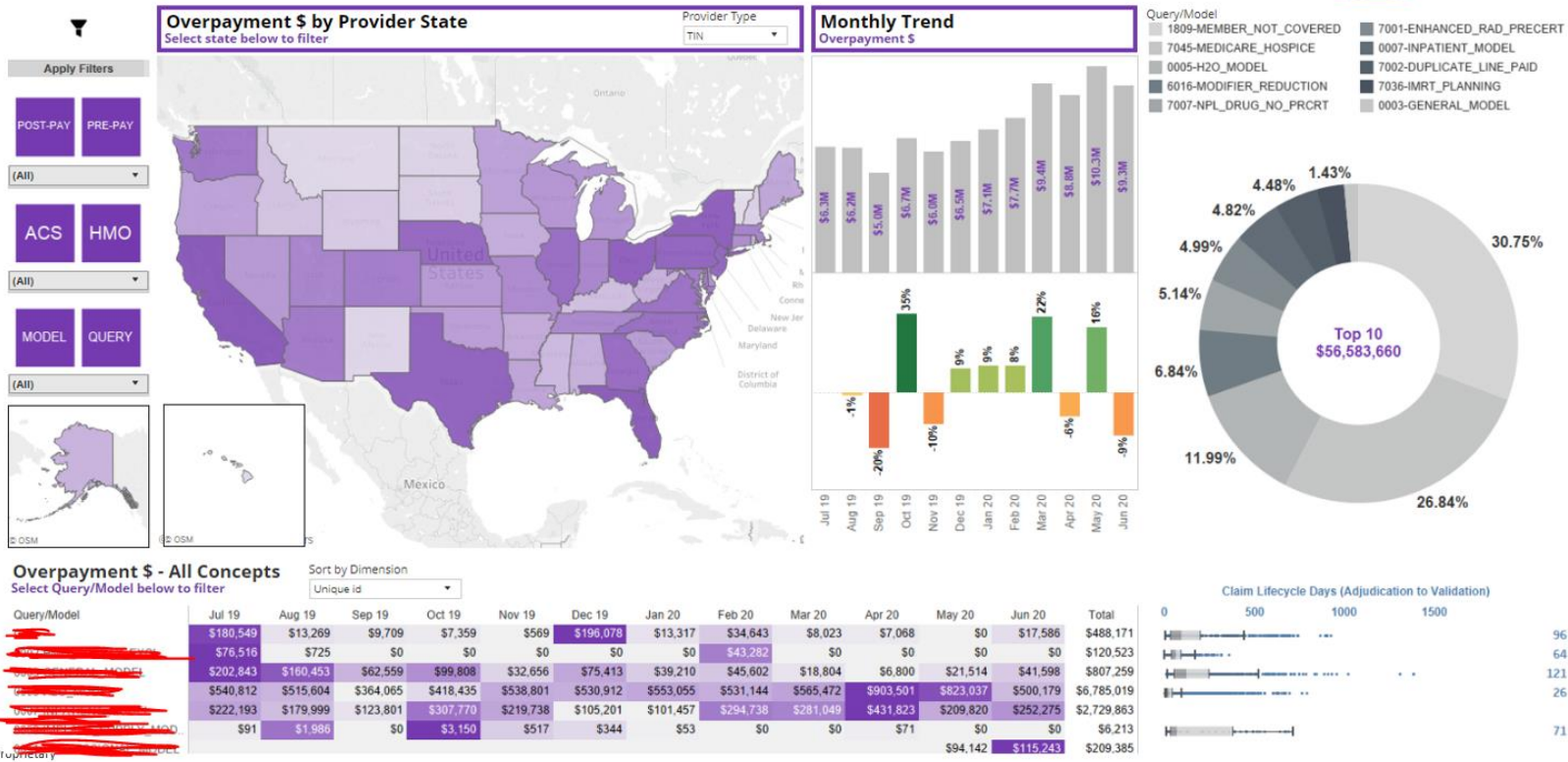
Mapping and Pie charts

Another time series example:

Sources:

https://www.tableau.com/