

# Process Notebook

*Visualizing the consumer complaints from the Bureau of Consumer Financial Protection (CFPB)*



**Group E:** Cunhonghu Ding, Weiran Yu, Qianyun Wang, Hongcen Tu

04.19.2021

QMSS GR5063 - Columbia University

Spring 2021

## INTRODUCTION

The purpose of this project is to develop a web application that allows users to quickly explore the Consumer Complaint Database through visualizations. The tool would serve as a starting point for exploratory analysis into this dataset. We have made the raw data available to download for everyone who comes across this tool. Some of the questions that can be answered using this web application include:

1. Which states receive the highest number of consumer complaints?
2. What are the most frequent terms associated with complaints?
3. Do the sentiments expressed differ by financial product?
4. Have the trends in consumer complaints changed following COVID?
5. Which companies receive the most amount of complaints?

## DATA

### Description:

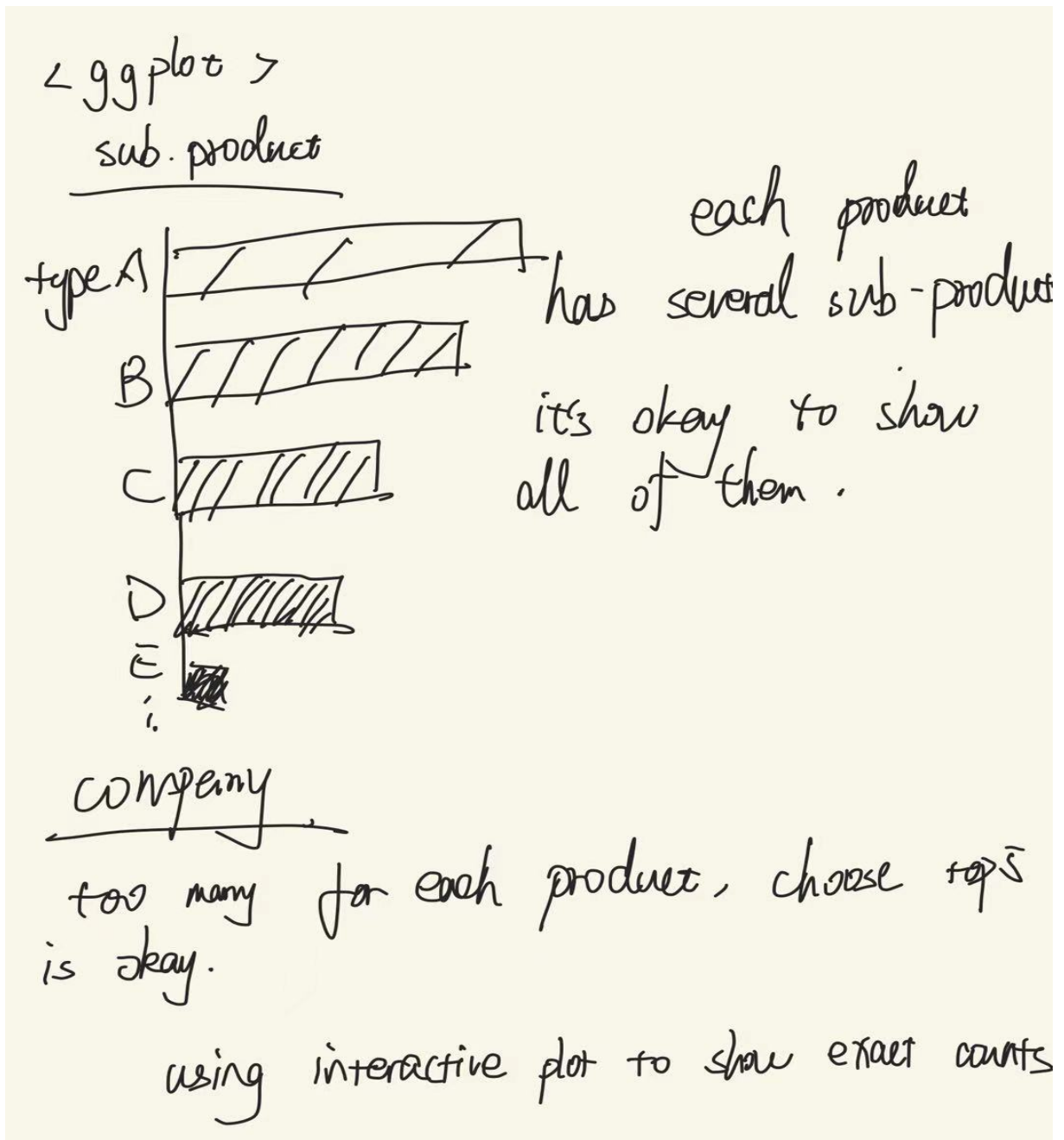
The Consumer Complaint Database is a collection of complaints about consumer financial products and services that we sent to companies for response.

Complaints are published after the company responds, confirming a commercial relationship with the consumer, or after 15 days, whichever comes first.

Complaints referred to other regulators, such as complaints about depository institutions with less than \$10 billion in assets, are not published in the Consumer Complaint Database. The database generally updates daily. Each row in this data set corresponds to a complaint that was given by a customer and the related attributes such as a narrative, issue and sub-issue type, related financial product, etc.

Source: <https://www.consumerfinance.gov/data-research/consumer-complaints/>

## INITIAL PROJECT PLANS (hand drawing)

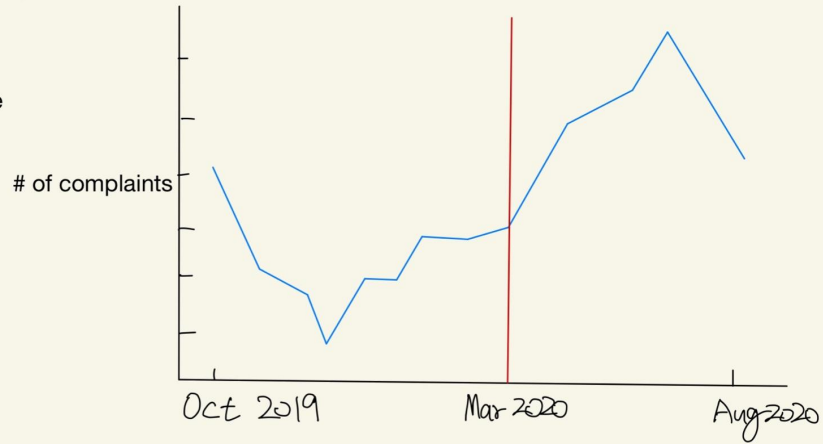


## Consumer complaints

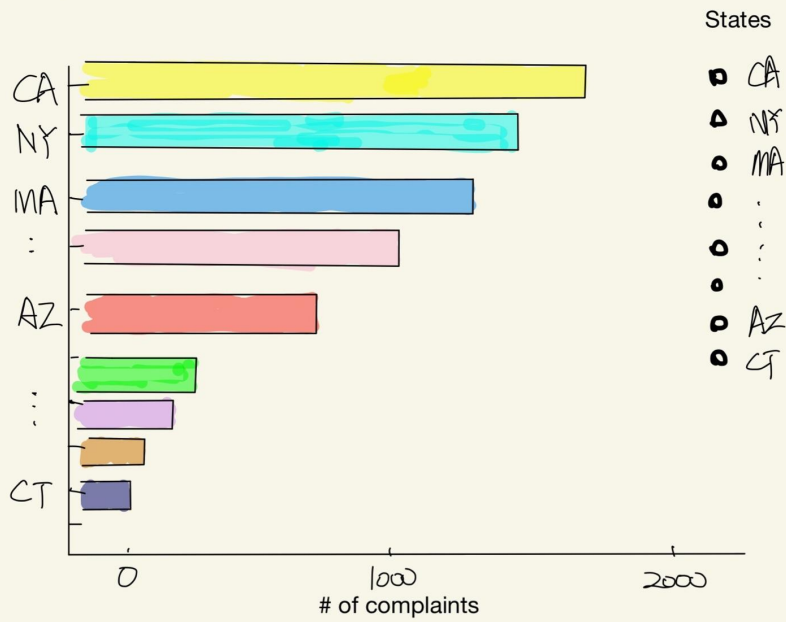
Products:  
Debt collection  
Credit reporting, credit  
repair services, or other  
personal consumer  
reports  
Mortgage  
Vehicle loan or lease  
Checking or saving  
account  
...

Timely response  
☐ Yes  
☐ No

Complaint by date



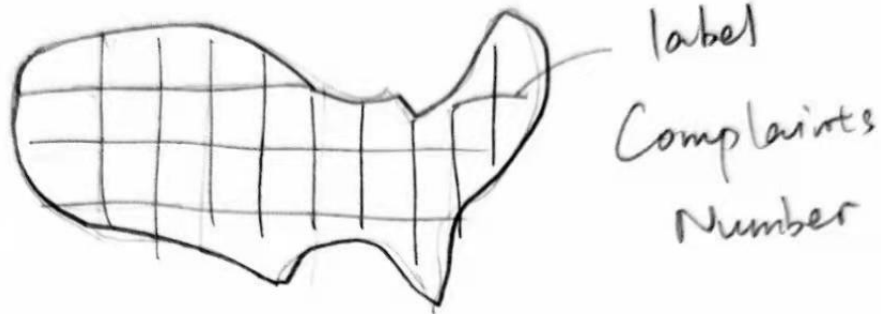
Complaints by states



<Map>

State level

Option 1:



Option 2: Statebins



Analysis } by Product

Top 1: Credit

by Issue

Top 1: Report

# 1. Wordcloud

In nice colours,  
↓ of course!

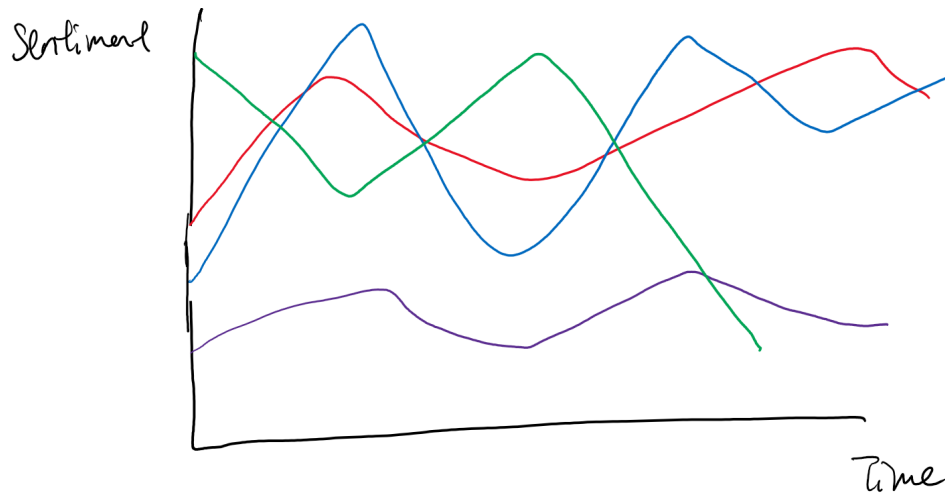
SCAM CREDIT  
LOAN MONEY  
DEBT BAD CA  
NOPE HSA THEFT CARD

# 2. Comparison Cloud

CREDIT CARD MORTGAGE  
BANK CREDIT SCORES

Good to know what's different among  
the products!

3. How have sentiment changed?



## VISUALIZATION

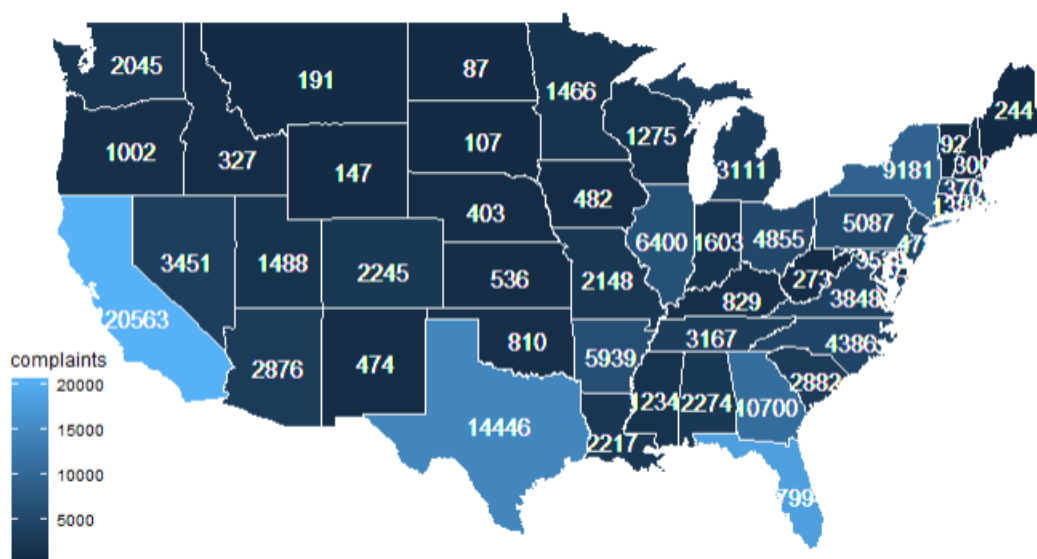
This is an interactive tool to help you explore the financial customer complaints data from 2019 to 2020. Please feel free to navigate to any of the tabs above to explore visualizations broken down by geographical location, text content, and complaint type.

The Consumer Financial Protection Bureau is a U.S. government agency that makes sure banks, lenders, and other financial companies treat you fairly. Consumer Complaint Database is a collection of complaints about consumer financial products and services that we sent to companies for response.

### 1. Maps

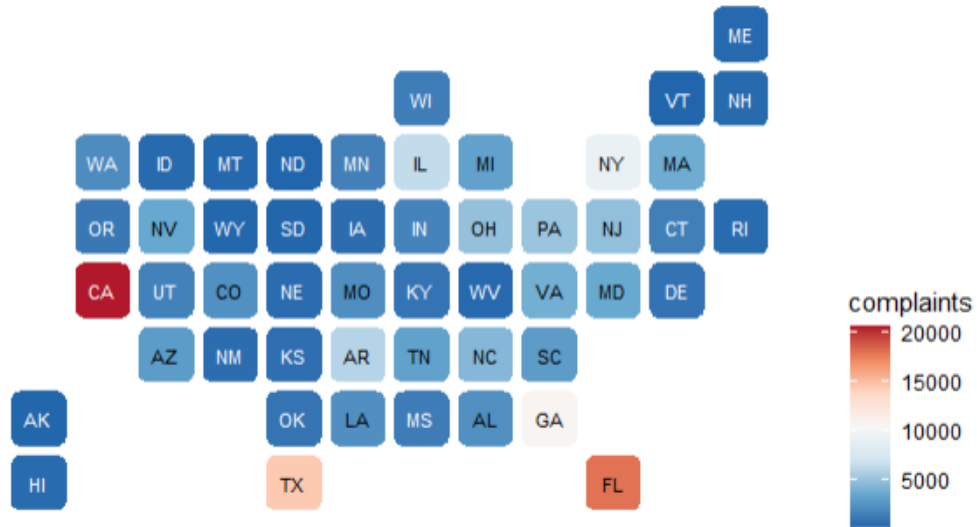
#### Geographical Overview

Consumers from all 50 states and the District of Columbia submitted complaints to the Bureau. To understand state and regional trends, we analyze the geographic distribution of complaints. Map 1 shows that, from 2019 to 2020, the Bureau received more complaints from consumers in California than anywhere else in the United States, followed by consumers in Florida, Texas, and Georgia. Consumers in North Dakota submitted the fewest complaints of any state.

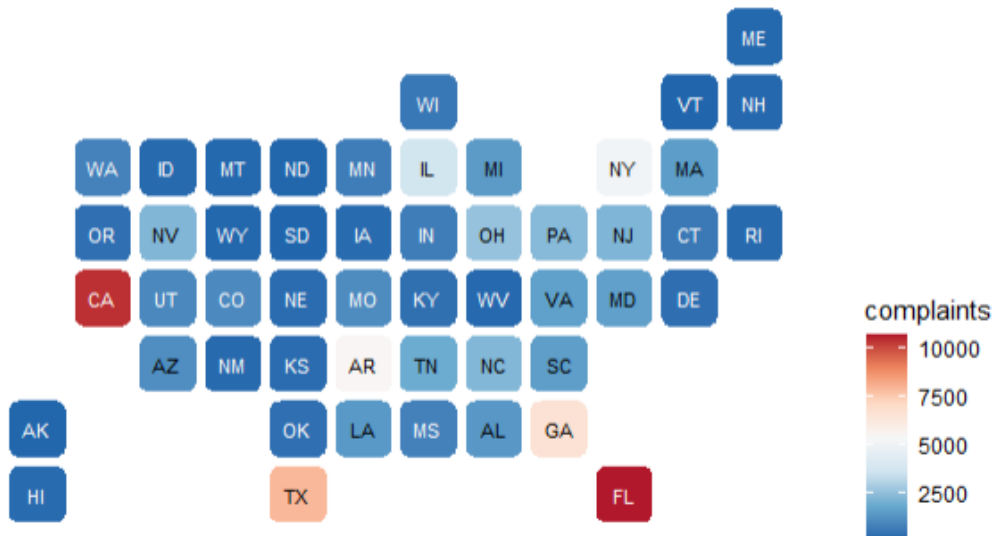




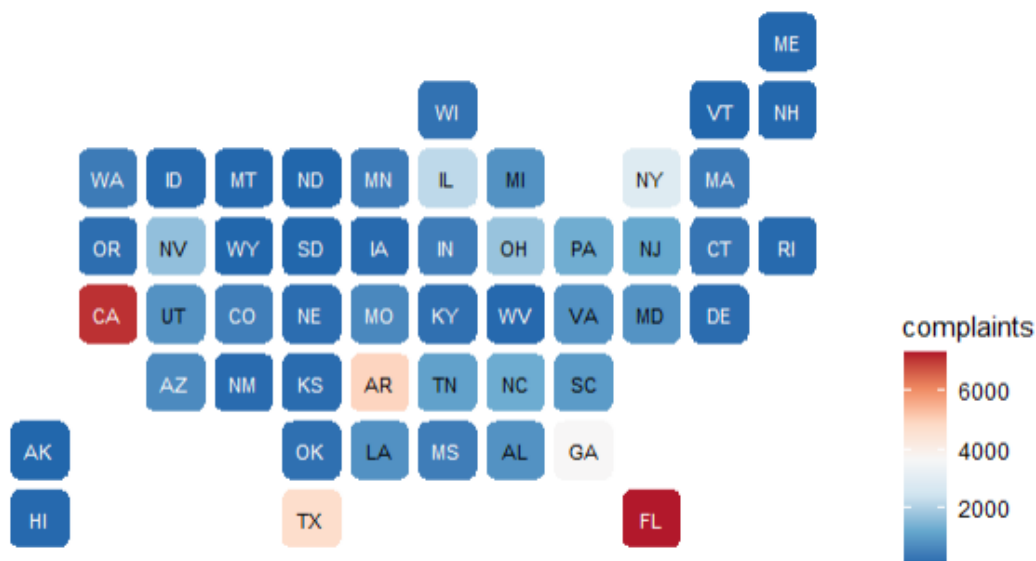
## Total Customer Complaints by State



Product: Credit reporting, credit repair services, or other personal co



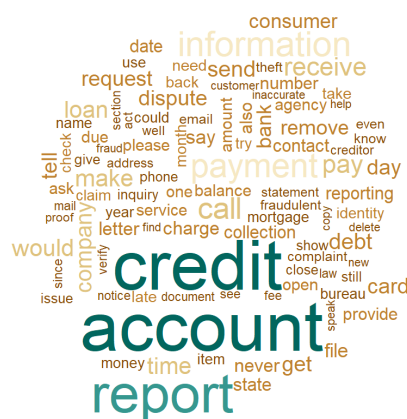
Issue: Incorrect information on your report



## 2. Visualizations of text analyses

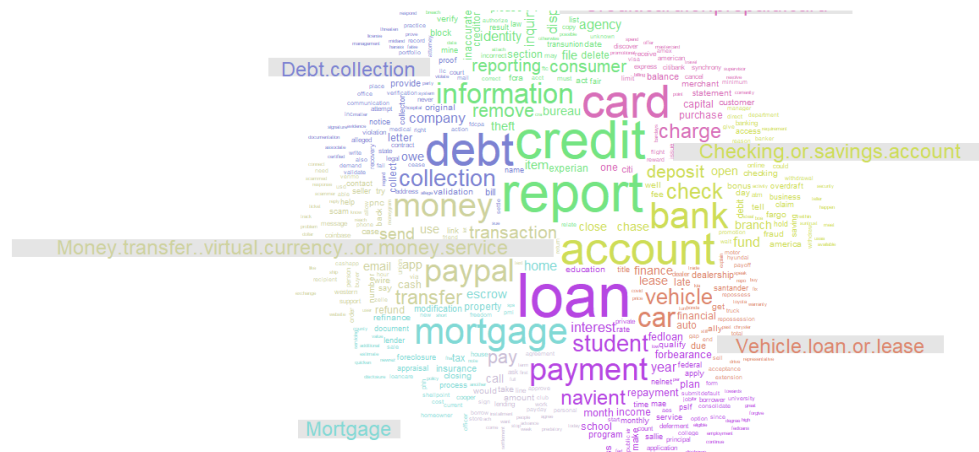
## 1. Word Cloud

We toyed with the idea of a word cloud identifying the most frequent phrases used among complaints in general, but ultimately realized that the value of information overlaps with the comparison cloud, so we chose not to proceed further with this visualization on the final product. Regardless, here's what we came up with anyway.

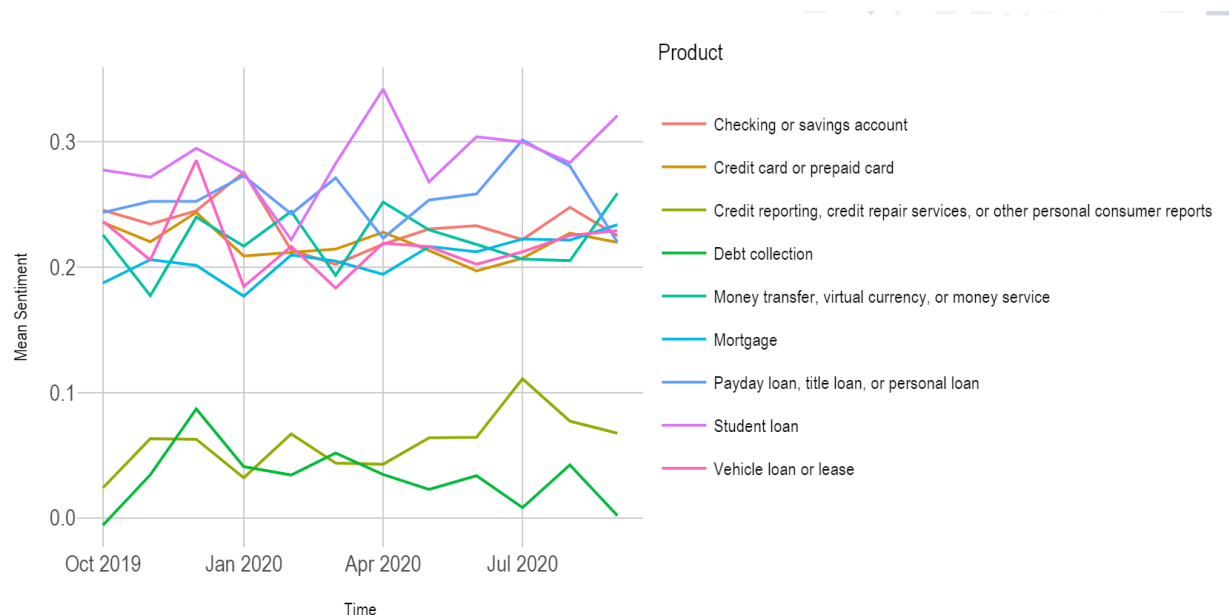


## 2. Comparison Cloud

We generated a comparison cloud to compare the different terms that are being used across the different types of product complaints. For the most part, the terms are what we expect (i.e. loan, student standing out in the student loan category, and mortgage standing out in the mortgage category). It's a nice visualization and a nice overview of what words are used, but there's really not much value to be extracted out of it.

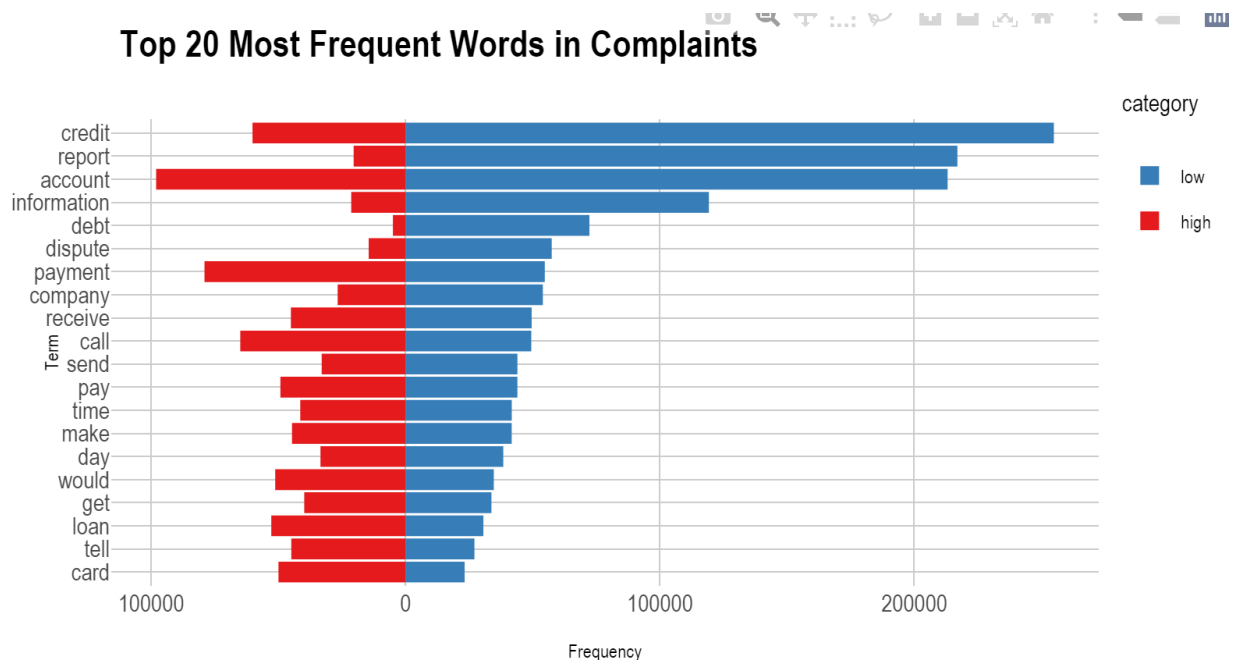


### 3. Sentiment over Time



We generated this plot in ggplotly, and immediately realized that there are really 2 clusters of products. Debt collection and credit reporting related complaints have vastly lower sentiment scores than the other categories. I suppose perhaps these could be more personal events than the other ones? There is really not a significant variation over time for these sentiment scores. As well, all of these sentiment scores are positive, even though the dataset is for consumer complaints. Perhaps the formal language used in data entry has all to do with it.

#### 4. Pyramid Plot



We decided to add this plot in the wake of the findings of our previous plot, and compare the term frequencies of the most common words found in these complaints, grouped by sentiment category. Much of the terms are self-explanatory, but there are some notable items in here. Most significantly, the term “debt” is rarely seen in high sentiment categories. Terms such as “report”, “credit” and “information” are seen far more in the low sentiment categories as well. Perhaps people get more angry when they are asking for information? Hasty conclusion drawn there.

### 3.Static images based on ggplot2

a)complaints by date:

We made several plots to show the trends of consumer complaints changing during the 12 months for different products and whether getting timely response. The products include: debt collection; credit reporting, credit repair services, or other personal consumer reports; mortgage; vehicle loan or lease; checking or saving account; payday loan, title loan, or personal loan; credit card or prepaid card; student loan; money transfer, virtual currency, or money service. A read line, considering March, 2020 as the covid start month, was made to show the comparisons of the trend before and after the start of the pandemic in the U.S.A. All the plots show apparent increases of consumer complaints right after March, 2020.

Products:

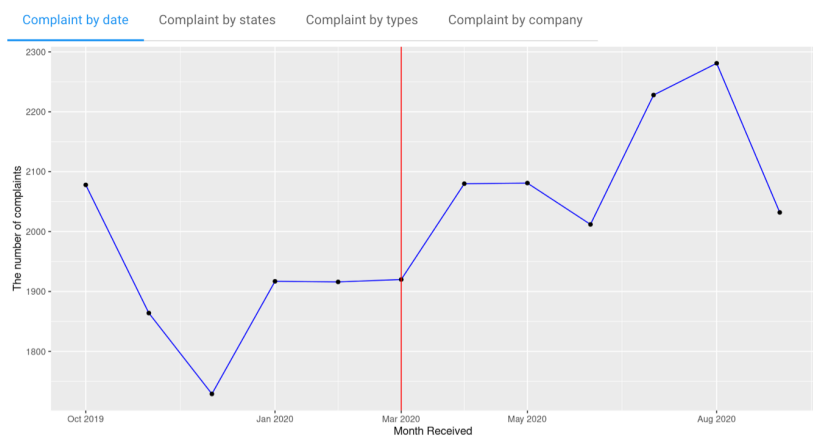
Debt collection

Timely response

☒ Yes

☐ No

DOWNLOAD DATA



Products:

Money transfer, virtual currency, or money service

personal consumer reports

Mortgage

Vehicle loan or lease

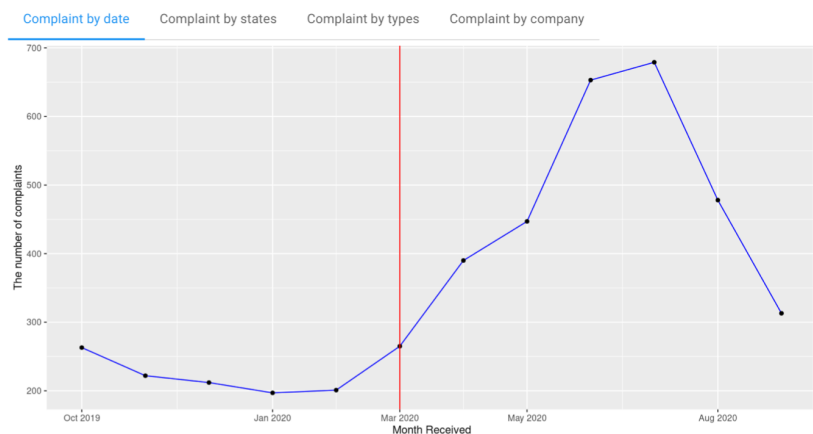
Checking or savings account

Payday loan, title loan, or personal loan

Credit card or prepaid card

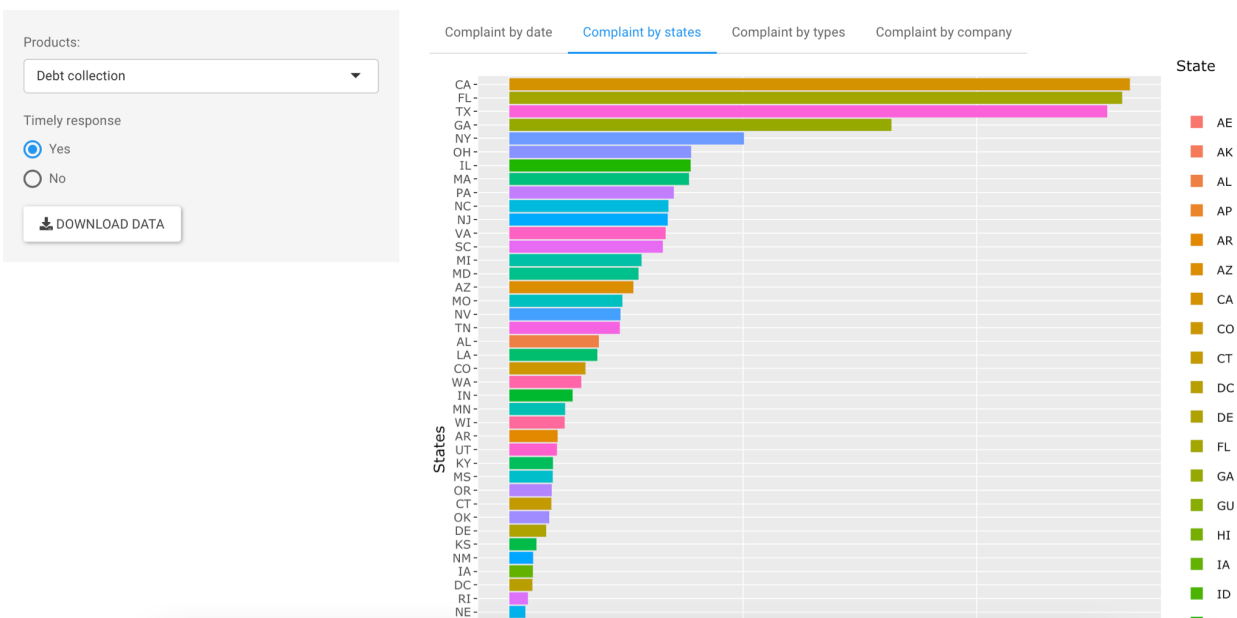
Student loan

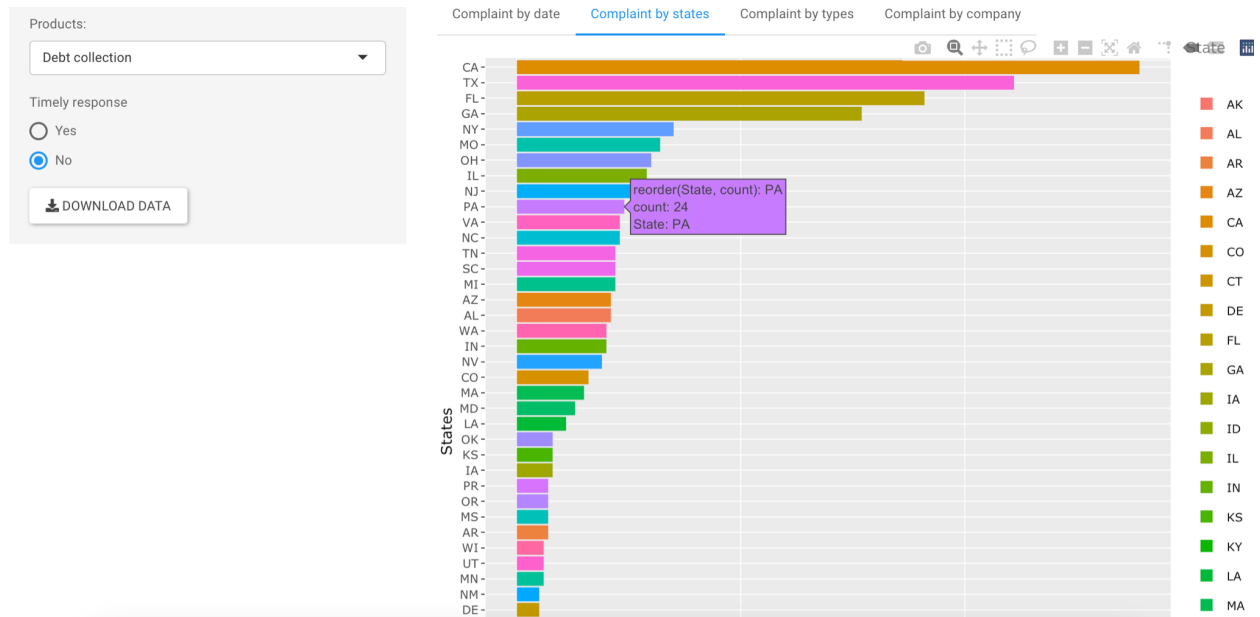
Money transfer, virtual currency, or money service



## b)complaints by states:

We were also curious about how the number of consumer complaints differ in states and regions. We made bar plots interactive to show the count of complaints based on different products and whether getting timely response. Using debt collection as an example, the top five states were CA, FL, TX, GA, NY, no matter getting timely responses or not. More details can be seen on the project.

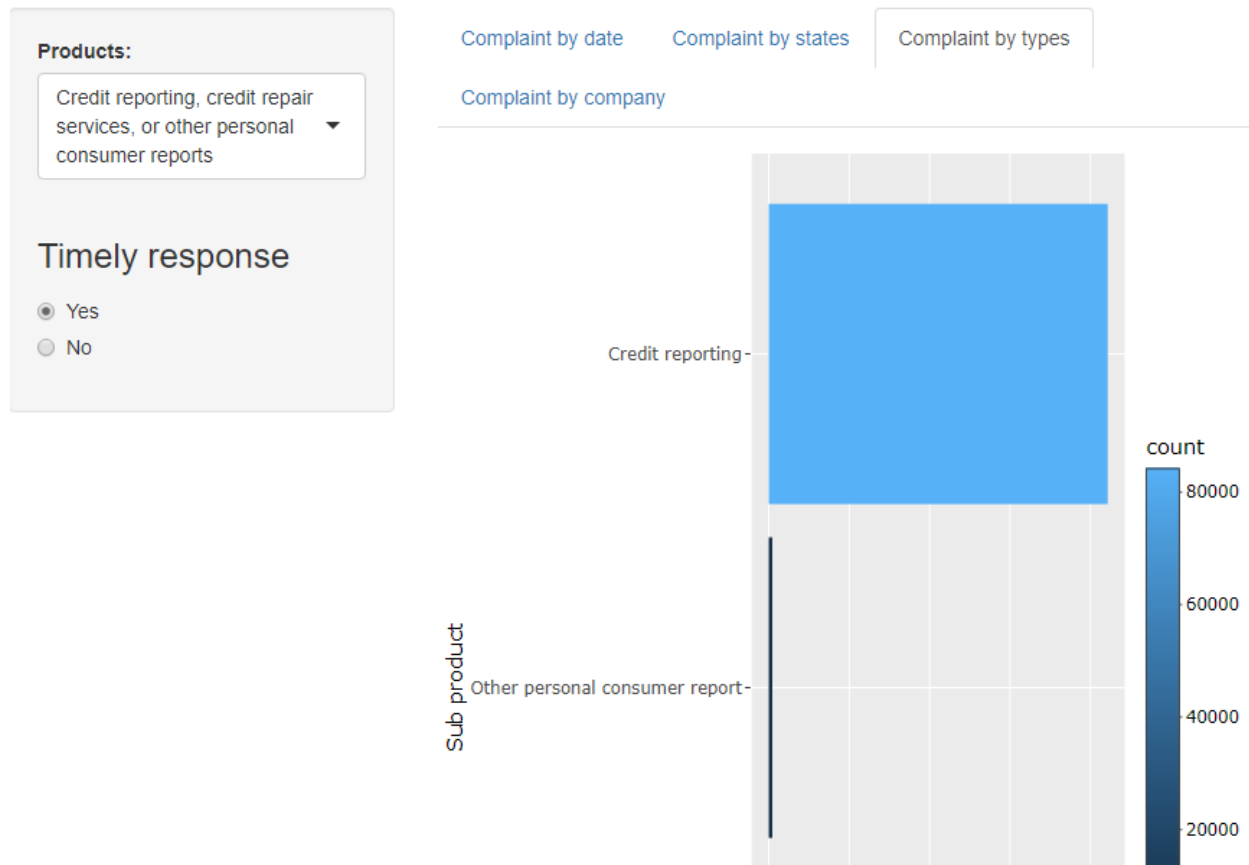




### c) Complaints by type:

Since each product has several sub-products, in order to make the complaint types more specific, we made a bar graph to show all sub-products of each product. We also reordered them from most frequent to least frequent for the audience to have a clear look at which complaint type was received the most. For example, under *credit reporting*, *credit repair services*, or *other personal consumer reports* product, credit reporting was complained about the most, which makes sense since credit reporting took a large part in people's daily life.

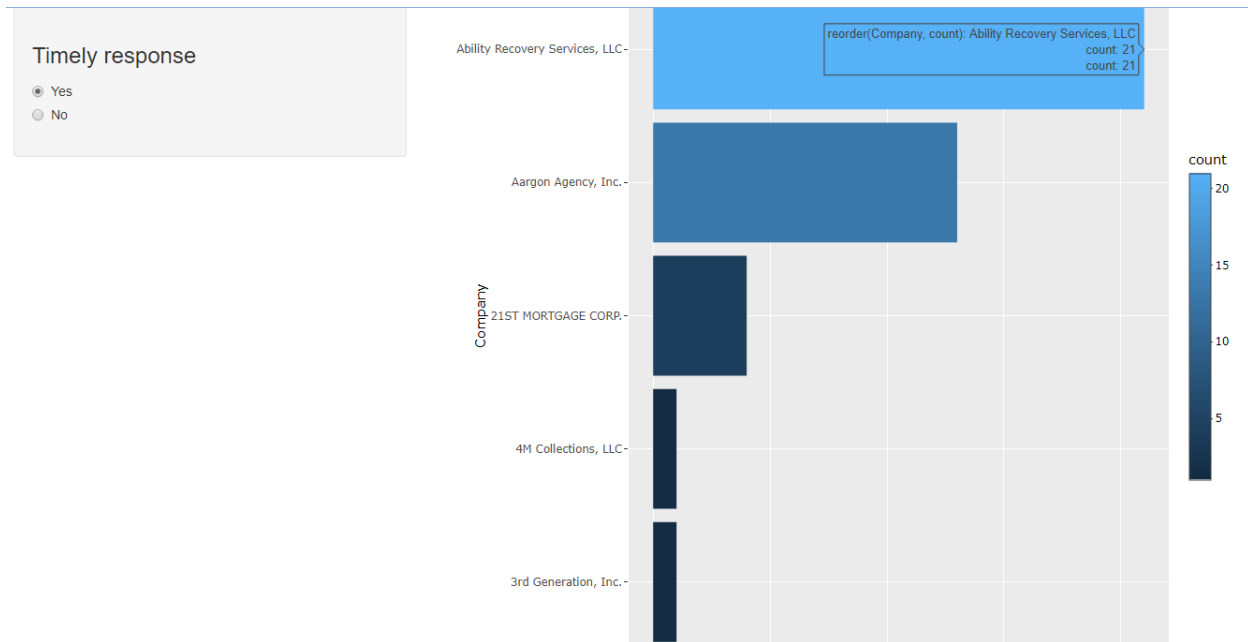
## Consumer Complaints



### d) Complaints by company

We also found that we should visualize the complaints made by each company. Firstly we included all companies, however, that ended up with a crowded and unclear graph since there were too many companies for each type of product. Then we concluded only the top 5 companies to make the graph understandable. For example, if we look at *credit reporting, credit repair services, or other personal consumer reports* product, we could see that Ability Recovery Services, LLC made 21 complaints on this product. We also reordered the bar plot from most frequent to least frequent so we could keep the same theme the entire time.





## CONCLUSION

We can make the following conclusions using the visualizations presented:

1. There's no time trend of sentiment scores across the different financial products.
2. There's 2 distinct groups of financial products based on sentiment scores, with debt collection and credit reporting averaging significantly lower sentiment scores than the other product categories.
3. "Debt" and "information" are most notably associated with low sentiment categories, among other things.
4. There are apparent increases of consumer complaints right after March, 2020, regarding different products.
5. Credit reporting is complained about the most, and General-purpose credit card or charge card follows behind during Oct 2019 to Aug 2020.
6. The Bureau received complaints from consumers in California the most compared to other states in the United States. Consumers in North Dakota submitted the fewest complaints of any state.