**STEPS SPECIFICATIONS: DIAGRAMS/IMAGES WHERE POSSIBLE:**

* + data visualisation- Joey

The questions we are trying to answer for crime v tansactions are:

1. Do areas with higher crime rate commit more CC frauds?
2. What are crime rates in comparison to the CC Frauds rate?

The question we are trying to answer for covid v transaction is:

1. Did the effects of COVID restrictions result in more Credit Card Fraud than different periods of time throughout the year?

Datasets at our disposition:

* Simulated transaction dataset: time coverage 01/01/2019-06/21/2021 -daily records
* COVID API dataset: 02/02/2020-present day -daily records
* Crime API dataset: 1991-2020 -yearly records
* US population dataset: time coverage 1991-2020 -yearly records

How we want to answer this:

For Covid v transactions:

* We will first **analyse the trend of stringency actual values of COVID in USA over 2020 & 2021**: **Figure 1**
  + **The distribution of stringency actual values of COVID in USA in 2020 & 2021**: **Figure 2**
* Then we will compare the percentage change of fraudulent transactions in the different states, to the percentage change of covid stringency values over the time period of 2020. (**Figures 3 & 4**)

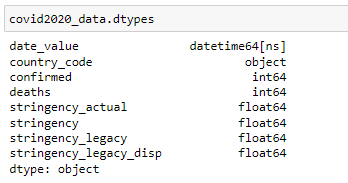
For Crime v transactions:

* + We will first **analyse the trend of crime rate related to CC and ATM Frauds** compared to the population in the US: we will show the **trend of crime rate and population**:
    - **since 1991** (when our data begin): **Graph\_1**,
    - **in the last 10 years**: **Graph\_2**
    - **in the last 3 years**: **Graph\_3**
  + We will then show how our transaction database is composed of fraudulent/non-fraudulent transactions in 2019 and 2020 (**Graph\_4**) (we unfortunately only have data for these 2 years, it would have been interesting to show the trend of the total number of fraudulent transactions in the same time periods - since 1991 or in the last 10 years).
  + Finally, we will compare the rate of fraudulent transactions in the different states to the national crime rate related to CC and ATM Frauds to see if it reflects the national data (**Graph\_5**)

This was done in 2 stages:

* Data cleansing & preparation
* Data visualisations

Data cleansing & preparation:

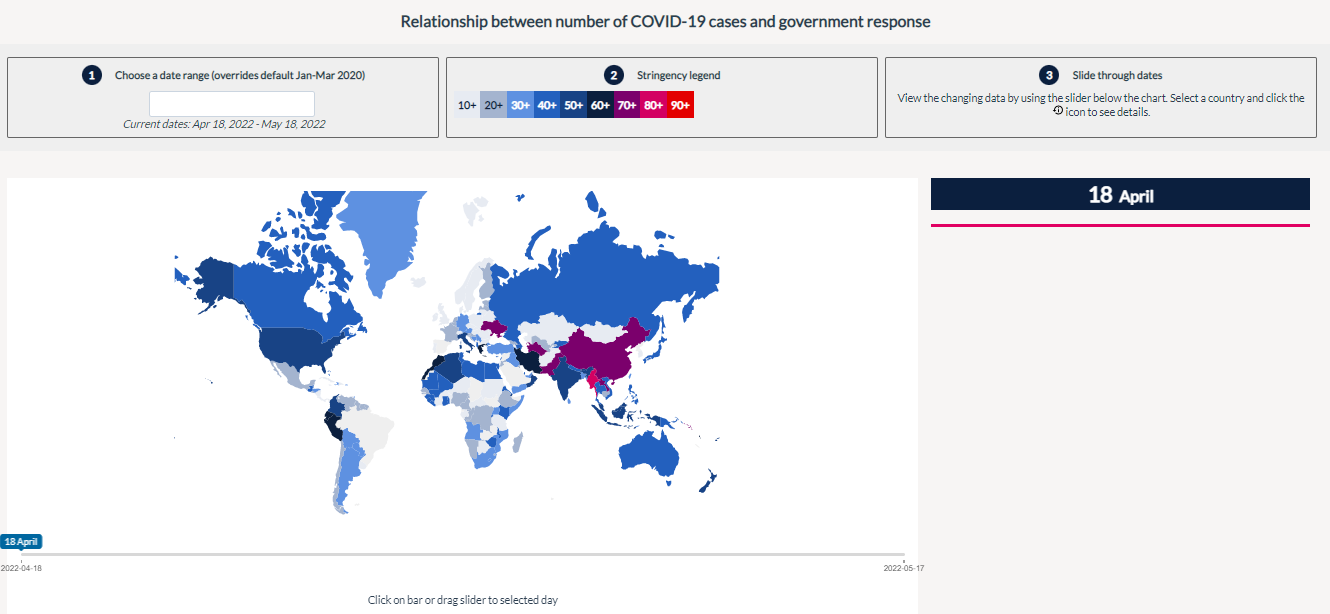
* Importation of datasets
* Figuring out what we wanted to show on the graphs & therefore what calculations were required to be made
* Creating DF which were more concise
* Exploring the DF and acquiring more information
* 
* Ridding/replacing null values
* Filtering out the information we need from datasets into DF
* Combining DF which would be of use for comparison
* Converting data types which may cause issues for the visualisations

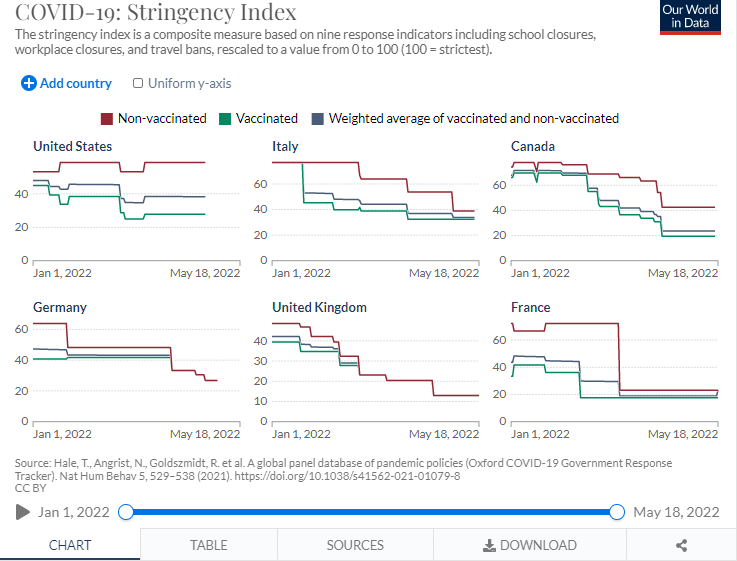
Data visualisations:

For Crime v transactions:

* Before getting into the comparison between the number of the frauds and the crime rate to try to answer our questions, we consider it useful to show the trend of the crime rate limited to CC and ATM frauds in the time period 1991-2020.
* The no. of offences related to 2020 must be considered as the total number of offences reported by end of 2020.
* Since our Crime dataset is the result of the offenses reported to the FBI by the Agencies, but only 63% Agencies sends data, we must consider our dataset as a sample of the real number of offenses occurred. Also, we cannot be sure that the offenses occurred in the competence area of these agencies correspond to the overall number of offenses.
* Hence, for our considerations, we assumed that their data reflects the behaviour of the overall population.

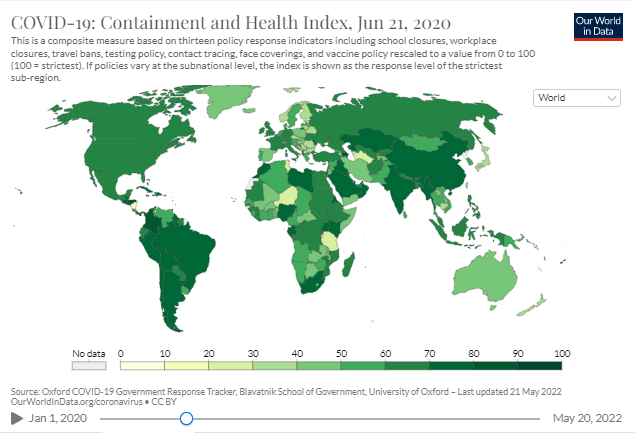
For Covid v transactions:

* Before we start the comparison between stringency percentage change for USA during COVID v. fraudulent transactions percentage change in 2019 & 2020. We found it important to have an understanding of how the stringency values did change over the course of 2020 and 2021 in the USA. These 2 figures endeavour to provide some insight from the data we have to give us an understanding of what the USA government policies were like. We have used the stringency\_actual values as opposed to the other calculated ones as we do not know all the initial values for which they were used for the calculations.
* Some background on the COVID API → There is a lot of interactive visualisations available from the source and the associated links to where they gained their data from:
* 
* We aren’t looking at the entire world but only US so it’s not as valid for us but a good indicator.



As COVID19 went on, their data collection values started to be altered too. Inclusion of vaccination status.

There’s also some pretty cool glider indicators to depict how covid19 stringency played into effect across the world. Nice heatmap + world map conversion.



* We have to remember that we can draw conclusions to the best of our abilities with the data provided and collected for us. The source is a trusted one and we believe this dataset has minimialised as much bias as possible.
* Therefore we have decided this is the best method for us to have a clearer idea of what the data is trying to tell us.

**Tools and Libraries Used:**

pandas

numpy

matplotlib.pyplot

matplotlib

seaborn

pprint

Datetime (and datetime from datetime)

**Implementation Process:** Achievements **- DIAGRAMS/IMAGES WHERE POSSIBLE**

WHAT WENT WELL FOR YOU

Working on this project has given me the confidence in coding to retrieve APIs and create visualisations to a standard that I am more comfortable with. I now better understand the steps and processes required and how to better clean and prepare my data. I decided to swap APIs for our visualisations to gain more insight into our data and have a better understanding. Working so closely with Nicol, it was great teamwork and I believe it provided us with gaining towards our common goal for this project.

**Implementation Process:** Challenges **- DIAGRAMS/IMAGES WHERE POSSIBLE**

WHAT YOU FOUND DIFFICULT

A lot more research is required for obtaining the correct API than I initially expected. This is to ensure it provides you with the information you seek from a reliable source.

The selection of visualisations was also vital in how we wanted to depict the analysis we gained from the datasets. It was difficult to select what type to not cause bias.

We were limited by the datasets due to the simulated aspect of the transactions and drawing conclusions from them will be based purely on what we have. It may not represent society as a whole as ultimately it was simulated.

**Implementation Process:** Decision to Change Something  **- DIAGRAMS/IMAGES WHERE POSSIBLE**

Scatter Graph Change - JOEY AND NIKI

As noted in the challenges, the graph type choices were difficult to make at the beginning without cleansing and preparing the data first. The decision to make bar plots and line graphs as opposed to a scatter graph for the comparison of stringency v fraudulent transactions became more apparent as the data was leading us in that direction. This may also be due to us trying to wrap our heads around what we wanted the visualisations to depict and making sure we answer the core question of whether the effects of COVID restrictions result in more Credit Card Fraud than different periods of time throughout the year? To avoid creating bias from visualisation depictions we changed the graph type.

**RESULT REPORTING:**

**EXPLAIN YOUR FINDINGS HERE - Data Visualisation Results/Modelling Results**

**INCLUDE DIAGRAMS/IMAGES WHERE POSSIBLE**

**DATA VISUALISATION:**

Before we start the comparison between stringency percentage change for USA during COVID v. fraudulent transactions percentage change in 2019 & 2020. We found it important to have an understanding of how the stringency values did change over the course of 2020 and 2021 in the USA.

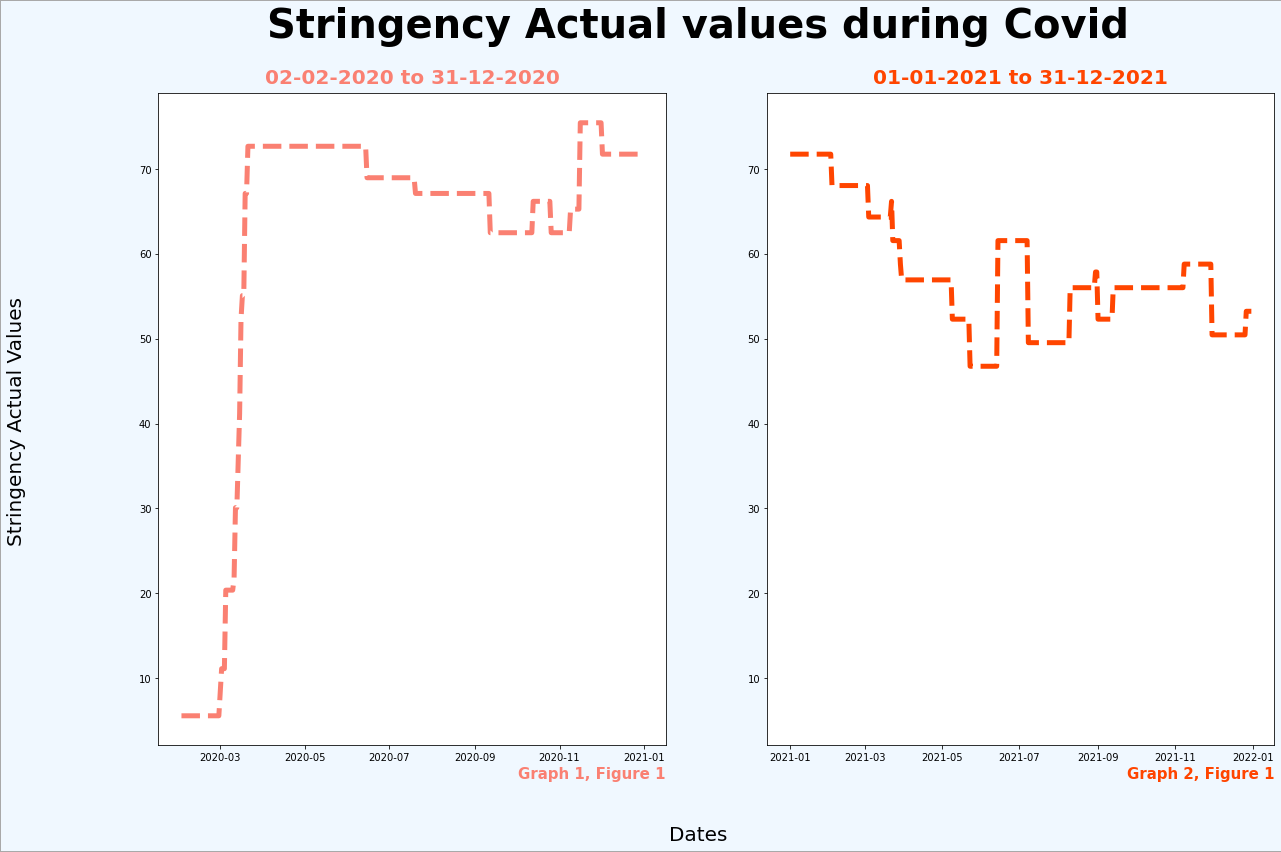
These 2 figures endeavour to provide some insight from the data we have to give us an understanding of what the USA government policies were like.

We have used the stringency\_actual values as opposed to the other calculated ones as we do not know all the initial values for which they were used for the calculations.

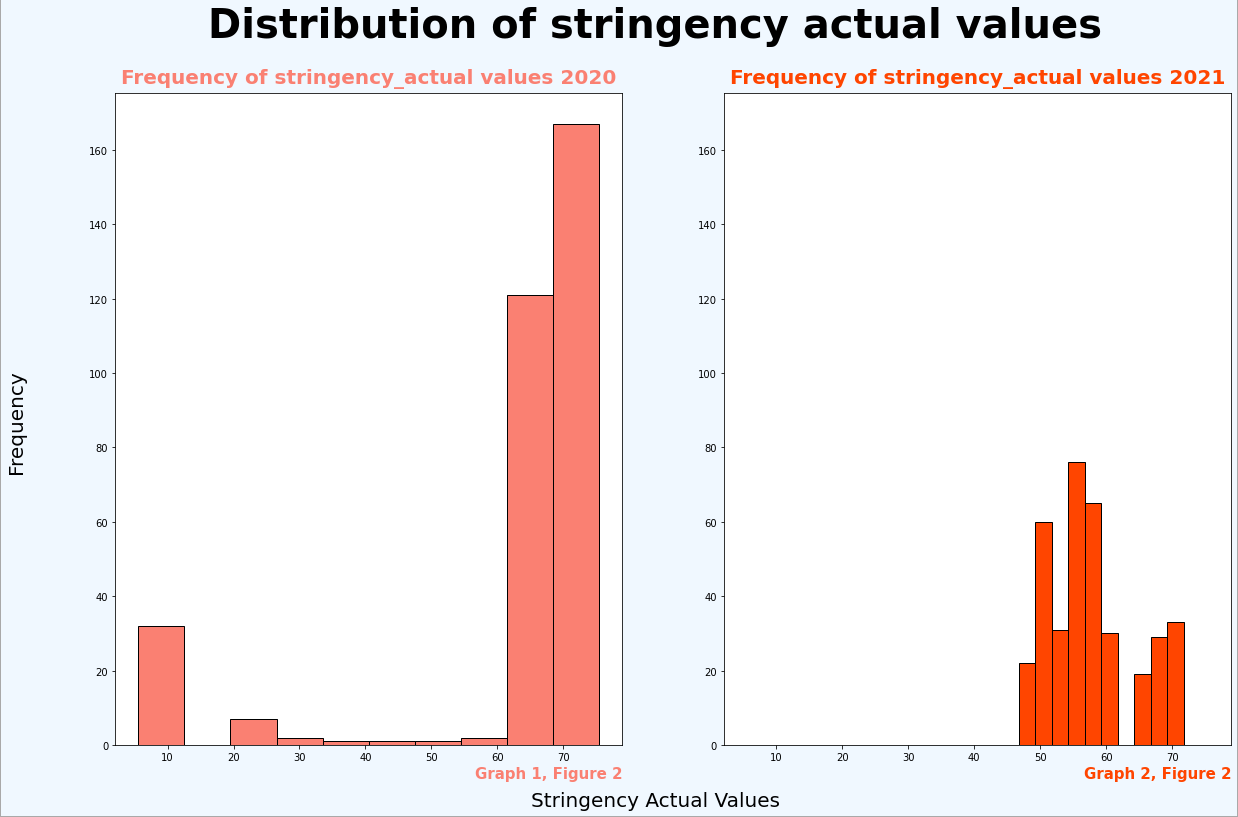
We have to remember that we can draw conclusions to the best of our abilities with the data provided and collected for us.

The source is a trusted one and we believe this dataset has minimialised as much bias as possible.

Therefore we have decided this is the best method for us to have a clearer idea of what the data is trying to tell us.



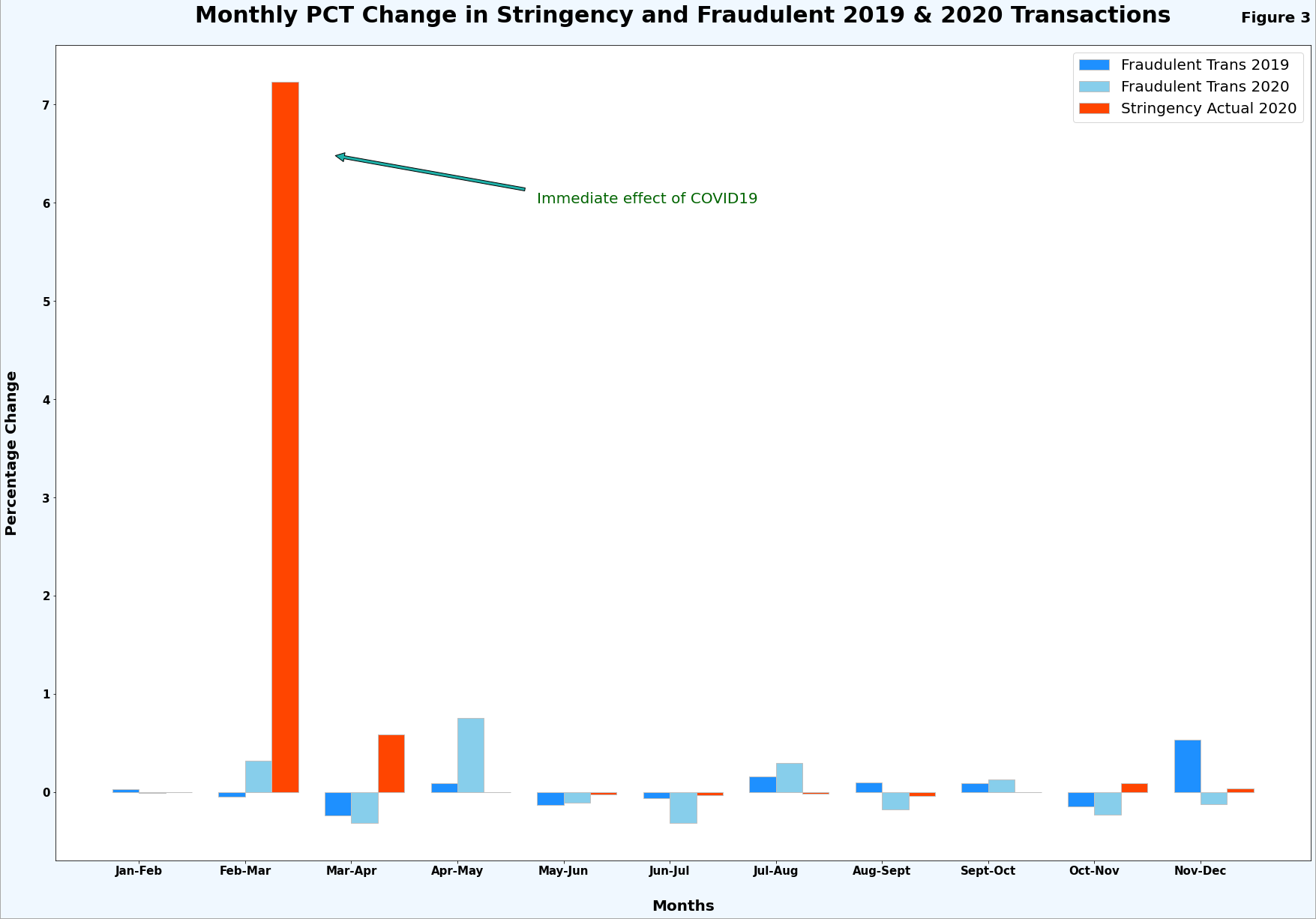
In Figure 1, the sharp rise in stringency values in 2020-03 indicates when COVID19 first impacted the USA government's policies. From these we are able to see the dips and rises according to what you can infer as seasonal changes. There is a slight increase of stringency in 2020-11 at the beginning of winter, then as winter draws to an end the stringency values slowly decrease. This is not to say they are gone completely. With COVID19 there was a lot of uncertainty and even though Summer came around 2021-05/06 with stringency being at the lowest levels for the past year, it sharply increased back to 63 possibly due to the increase in number of COVID19 cases.



In Figure 2 we compare the distribution of stringency actual values between 2020 and 2021. This is to see how the USA changed their stringency as we got to understand COVID19 better. They share the same x and y axes so you can see the difference in values immediately.

With these graphs it is clear to see that consistently higher levels of stringency were present for the year of 2021. There is a spike for the year 2020 but there was more uncertainty and so that is why there are values towards the lower end of the spectrum.

QUESTION: Did the effects of COVID restrictions result in more Credit Card Fraud than different periods of time throughout the year?



In Figure 3 we are able to see a major spike in stringency percentage change for the months of February to March. We do not see this same effect for fraudulent transactions in 2020 as one would expect with everyone being indoors. This could be due to the dataset being simulated and not real-world data.

From what we can infer from the datasets, there is no correlation at all between fraudulent transactions and the impact of stringency on the USA.

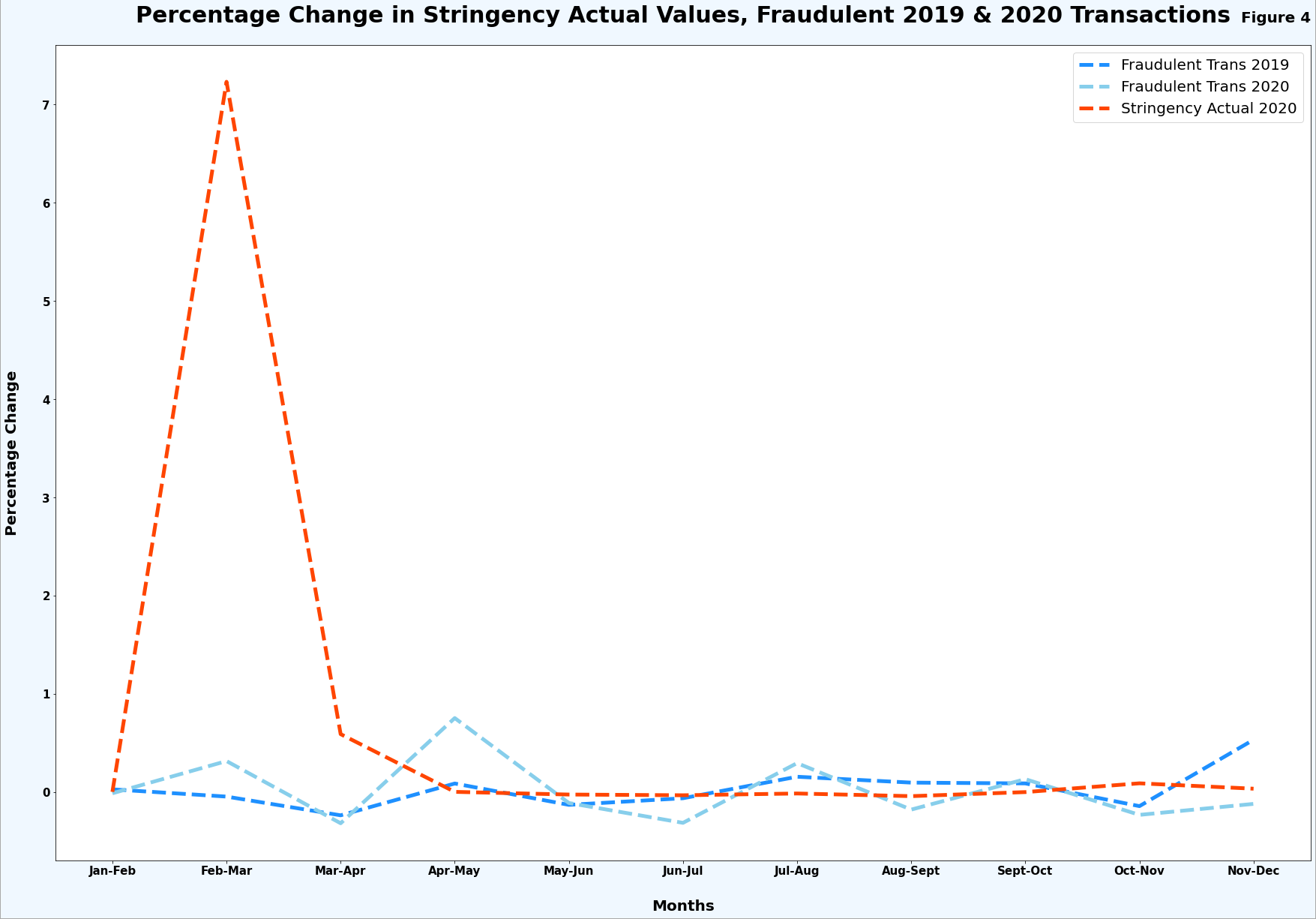


Figure 4 depicts the same lack of correlation between transactions and stringency. This visualisation was created to highlight this in a more effective way. It is from creating multiple graphs which we are able to determine which would best depict the information the data is trying to show us.