Visualization of the Trade History between China and U.S. States

ABSTRACT

This project aims to use various visualization techniques and machine learning methods to study the history of trade between China and U.S. states. To achieve this goal, we analyzed the data from different sources, performed manifold learning-based clustering analysis, and created the best visualizations to illustrate our results. We hope that these visualizations and insights gained in our analyses can help us understand the trend of the U.S.-China trade and provide us some clues on the possible impact of the trade war on individual U.S. states.

INTRODUCTION

China and the United States (U.S.) are the two largest economies in the world, and their economic and trade relations play an important role in the global economic system. In 2017, China and the United States, as the two leading economies of the WTO members, accounted for more than 20% of world merchandise trade[1] (**Fig. 1**). In history, the trade between China and the United States can be dated back two centuries ago. When Sino-U.S. relations began to recover in 1972, the bilateral trade was just as much as 12.28 million U.S. dollars[2]. Today, China is the U.S.'s largest trading partner, the third largest export market and the largest source of imports, while the United States is China's second largest trading partner, the largest export market and the fifth largest source of imports[3]. Behind the increasingly frequent U.S.-China economic and trade exchanges is the emergence of an economic relationship that is deeply integrated and mutually beneficial. U.S. companies have established more than 20,000 joint ventures or sole proprietorships in China. Meanwhile, some of China's largest companies (e.g. Alibaba, Baidu, New Oriental, etc.) have been successfully approved for NASDAQ listings, becoming multinational companies. According to the statistics of General Administration of Customs of the People's Republic of China, the total trade volume between China and the United States has grown from 33 billion U.S. dollars in 1992 to 633.5 billion U.S. dollars in 2018. At the same time, however, the U.S. trade deficit with China has also increased from $28.1 billion in 2001 to $419.2 billion in 2018, an increase of 11.6% ($43.6 billion) from 2017[4,5].

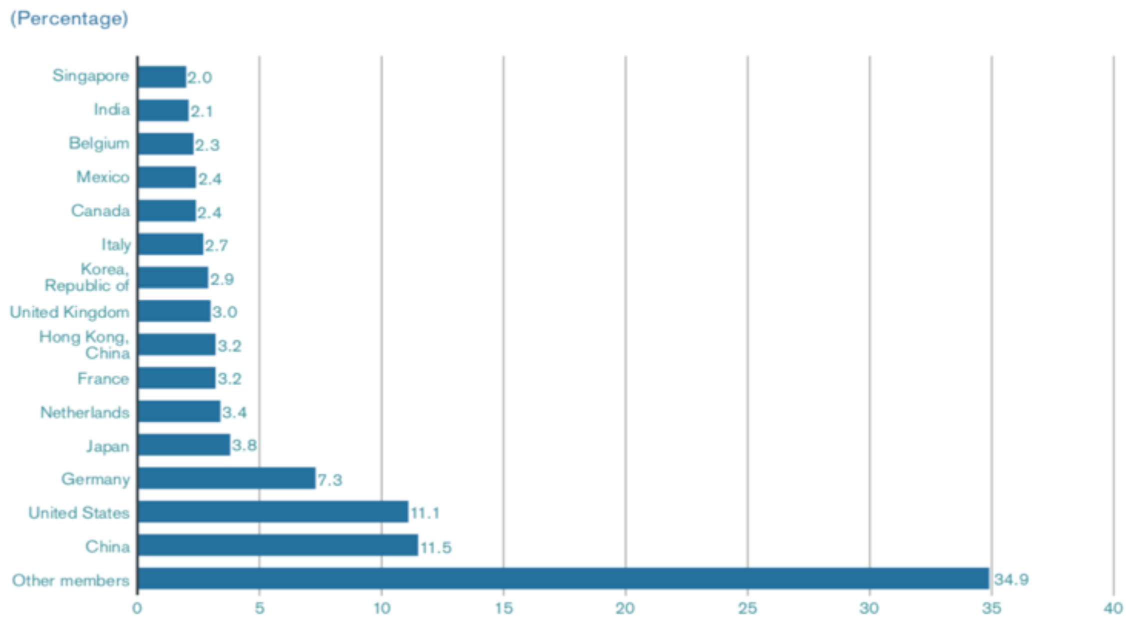


Figure 1:Share in merchandise trade of WTO members, 2017.
(Source: WORLD TRADE STATISTICAL REVIEW 2018, World Trade Organization)

In March 2019, the United States announced that it would increase tariffs on 50 billion worthy of goods imported from China (including clothing, shoes, electronics, steel, and aluminum). In response, China announced in April that it would retaliate against a range of U.S. products, signifying the start of the trade war between the two world's largest economies. Until recently, after multiple rounds of negotiations and repetitions, the trade war was still in the process of fermentation. The tension between the two countries has been further escalated after President Trump administration decided to raise the tariffs from 10% to 25% on $200 billion worthy of Chinese imports on the 10th of May. Economists and business leaders generally expect that the trade war between the two largest economies in the world will have a profound impact on both sides and the global economy.

There have been several existing visualizations depicting the evolution history of China-U.S. trade and mutual influence, presented in the format of choropleth map, time series, treemap, piechart, *etc*..

Choropleth map: The ITA International Trade Administration website collects historical data on U.S. exports to countries around the world since 1999. The picture below shows the regional differences in U.S. exports to China in different states in a choropleth map[6] (**Fig. 2**). Readers can quickly understand state-by-state exports to China and identify specific visually striking patterns. These preliminary visual displays provide a general overview of the regional differences in merchandise export, and meaningful insights for further downstream analysis. One downside is that the traditional choropleth map does not take land area into account potentially leading to visual illusion.

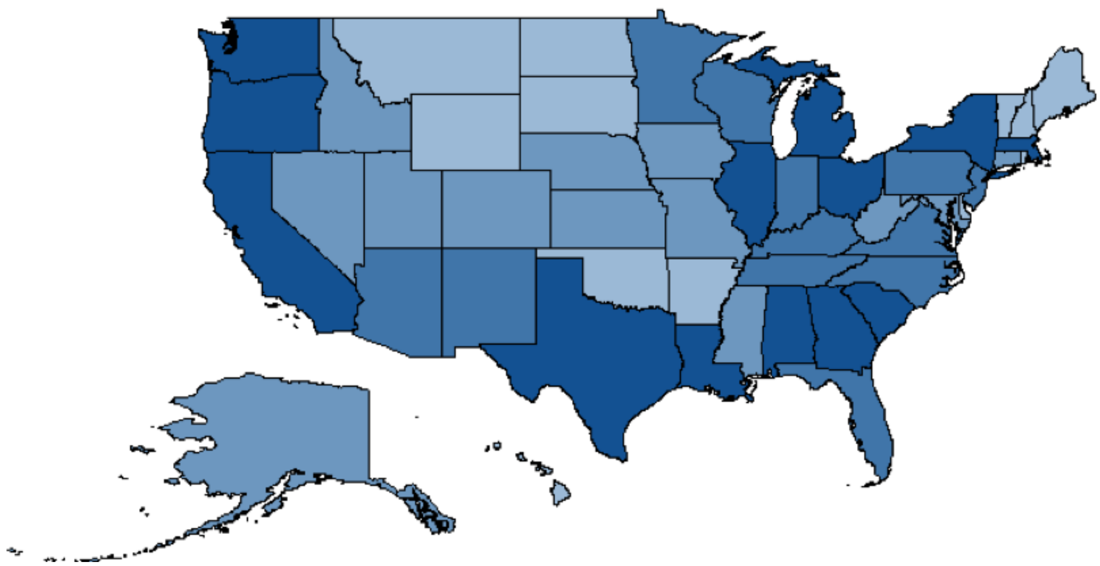


Figure 2: 2018 NAICS Total All Merchandise exports to China
(Source: International Trade Administration)

Treemap: Data visualization from the Observatory of Economic complexity website showcases the top export destinations of the United States using a metro-style treemap[7] (**Fig. 3**). The treemap correlates the fractions of the export with the area size of rectangles and labels the name and the corresponding export fraction for each destination country. The metro style embodies the principles of cleanliness, readability, and objectivity. It can be intuitively learned that Mexico (15%), Canada (12%) and China (11%) are among the top three export destinations. Unfortunately, because too much information needs to be presented, the visualization seems a bit busy especially for countries with small export fractions.

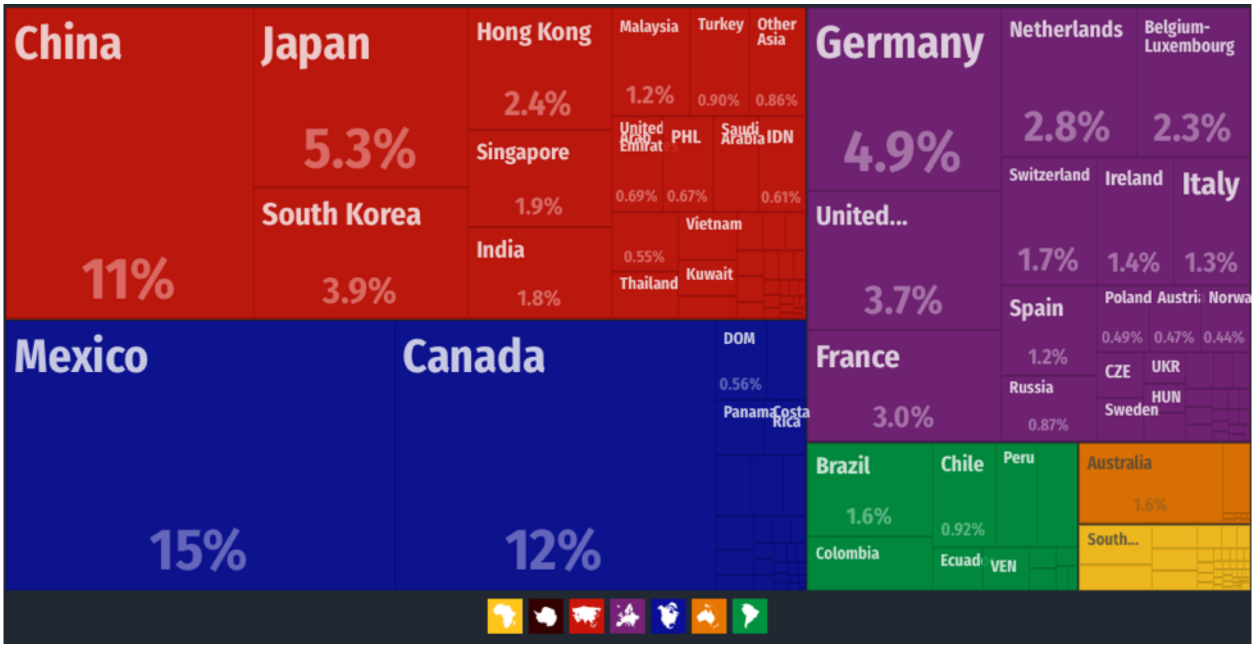


Figure 3: Export destinations of the U.S. foreign trade

(Source: the Observatory of Economic complexity)

Time series plot: This is another very common way of presenting economic data. It can be used to determine the trend of data in a given time period (e.g. some seasonal patterns), as well as analyze the impact of an economic or political event during a particular time period. For example, the time series plot of the US trade deficit with China clearly demonstrates the trend of an ever-expanding trade deficit that reached $419 billions in 2018[4] (**Fig. 4A**). Likewise, the time series chart of stock indexes since January 2018 also shows that the trade war has caused significant impacts on stock markets in both countries[4] (**Fig. 4B**).

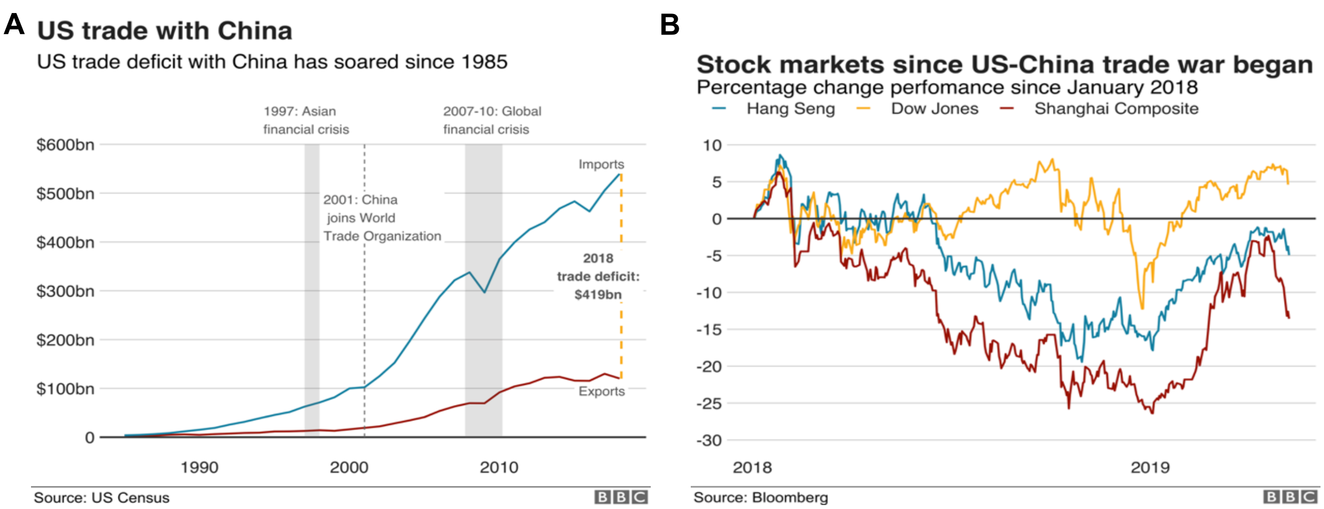
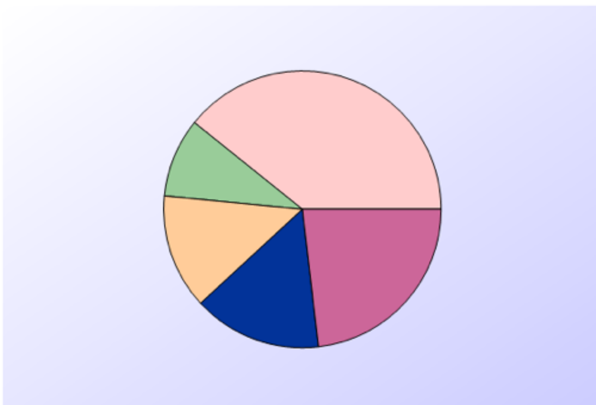


Figure 4: Time series plots of U.S trade with China and stock markets

Piechart: The following piechart displays the type fractions of products exported to China[5] (**Fig. 5**). It presents data in fractions and saves the reader's time to read the statistical figures directly. However, since our visual perception is not good at evaluating and comparing the size of the area, piechart is not optimal for presenting large amount of data or data with similar fractions.



Product	Value ($)	Percent
336--TRANSPORTATION EQUIPMENT	$27,813,712,334	23.1 %
334--COMPUTER AND ELECTRONIC PRODUCTS	$17,991,522,227	15 %
325--CHEMICALS	$16,132,263,343	13.4 %
333--MACHINERY, EXCEPT ELECTRICAL	$11,085,548,071	9.2 %
All Others	$47,125,095,154	39.2 %
Grand Total	$120,148,141,129	100 %

Figure 5: 2018 export to China of NAICS Total All Merchandise
(Source: International Trade Administration)

Overall, these existing visualizations provide us with a general overview of U.S.-China trade history from different perspectives. However, these visualizations focusing on simple two-dimensional, static display, often lack impressive interactivity and in-depth analysis of future trends and potential causal relationships, greatly reducing the visual experience of potential audiences. In this report, we will combine many of these data visualization techniques and sought to gain insights into the future direction of China-U.S. trade with a special focus on trade relations between different U.S. states and China, and further analyze the potential impact of political preference, regional culture, economy scale, and geographic location on the bilateral trade between the U.S. and China.

DATA & METHODS

After doing extensive research, we have identified the following datasets for use in our project:

- International Trade Administration's official website for export and import data. Their database has both the export (1999-2008) and import data (2009-2008) between U.S. states and individual foreign countries and economic entities[8, 9] (China, EU, etc.);
- U.S. Census Bureau's official website for export and import data between U.S. and China[10] (1985-2018);
- St Louis Federal Reserve's official website for producer price index data to adjust trade numbers[11];

- U.S. Federal Election Commission's official website for 2016 presidential election result data[12];
- U.S. Department of Commerce Bureau of Economic Analysis' official website for GDP by state data[13].

We have two main objectives for this study: the first is to use different visualization methods to illustrate the trade data (over time) between the individual U.S. states and China; and the second is to choose the best visualization methods to perform analysis on the trade data based on the work on the first objective.

Various visualization techniques to illustrate the trade history between China and individual U.S. states

In order to compare the import and export of different states in one graph, we have two basic options for visualization: one is to compare all states at each point of time, while the other is to compare the change in trade over time for each individual state.

As the first option, a state pyramid of bars can be created to show the distribution of trade dollars of different states and a slider widget that is bound to the year to visualize the trade dollar distribution over time. The advantage of using a bar chart is that we can compare both the difference between different states and the difference between import and export for each state at the same time. However, the bar chart also suffers the disadvantage of becoming too long with fifty states. A way to address this problem is to only show a few selected states by economy scale, import or export. Another option is to use interactive choropleth map of U.S. states to illustrate the comparison of export, import and surplus/deficit. Compared to a bar chart, a heatmap can be visually intuitive by using a map to illustrate all fifty states together as well as using different colors to show the numerical values.

If we choose the second visualization option, we can use interactive time-series plots of export, import and surplus/deficit over time for individual states and then add a drop-menu that can be used to choose the state of interest. A time-series plot will allow us to easily see the trend over time for each individual state but does not do a good job comparing different states.

The export of goods to foreign countries is important for a state because it creates jobs. In addition to (absolute) export, we are also interested in understanding the relative importance of export to China as compared to total export for each state, as it gives us critical insight about how each state would be affected differently if their export to China gets reduced. Using the same method in the previous section, we have plotted the time-series of the share of the export to China in total export for each individual state. As we have explained before, a time-series illustrates trend very well but is sometimes not very effective in performing multiple comparisons.

Visualizations for manifold learning

In the second part of the article, we tried to build a manifold learning model that clustered U.S. states according to foreign trade statistics, geographic location, export orientation, and party preferences of each state. The following visualization techniques were considered to display the results and insights gained from the analysis.

- A scatterplot to show the UMAP clustering of U.S. states based on the manifold learning.
- A choropleth map to show the overall geographic distribution of different clusters of U.S. states.
- A normalized Bar chart to compare the fractions of the democratic states and republican states based on the classification results. This comparison is particularly interesting, because it will tell us how the political dynamics possibly interacted with economies of U.S. states;
- A boxplot chart to compare the economic and trade characteristics of different clusters of U.S. states. The result of this plot will shed some light on which states would be more prone be impacted by the trade war;
- A waffle chart to display the export orientation of classified U.S. states. As we know, although one of its main targets is China, President Trump's foreign trade policy is also targeting many other regions and countries. It is highly likely that the geopolitical dynamics will determine how the policy targets different economies differently. There are multiple visualization choices for this kind of comparison: pie chart, tree map and waffle chart are the most common among them. Pie chart's main problem is that human eyes do not usually does not read angles very effectively, while tree map is more useful for comparing a large number of objects. Therefore, we have utilized waffle chart to compare the export destinations of different clusters of U.S. states to China and other major trading partners (EU, Japan, and NAFTA) in 2018.

RESULTS

Part I: Visualizations of U.S. trade with China and other major trading partners.

Visualize the Import and Export Over Time for Individual States

The three types of visualization - pyramid bar chart, choropleth map and time-series - for the import and export data between U.S. states and China are shown in **Fig. 6, 7** and **8**, respectively. Only static plots are shown here, while interactive plots will be demonstrated in the presentation. As explained previously, three types of comparisons can be easily obtained from the pyramid bar chart: import vs. export for an individual state; the import between different states; and the export between different states. Because the plot would get too long if all fifty states are included, only the top 10 states by 2018 GDP are shown (ordered by export). It should be noted that the export and import is in different scales ($16B vs. $160B), as there would be no visible difference in export between different states if the same scale were used.

A lot of interesting observations can be made from the pyramid bar chart. First, the import is often a magnitude more than the export, indicating that many states have a deficit in trade with China. Secondly, a comparison between different years indicates that the export and import ranks less monotonically over time. For example, the export and import in general ranks order in 2008, while the import does not rank monotonically as the export in 2018. This observation may suggest a different growth rate in economy for different states after the economic crisis in 2007. Last but not the least, it is expected that the states that both import from China and export to China significantly will be impacted the most by the trade war. For example, Carlifornia falls into this category and may be hit hard by the trade war.

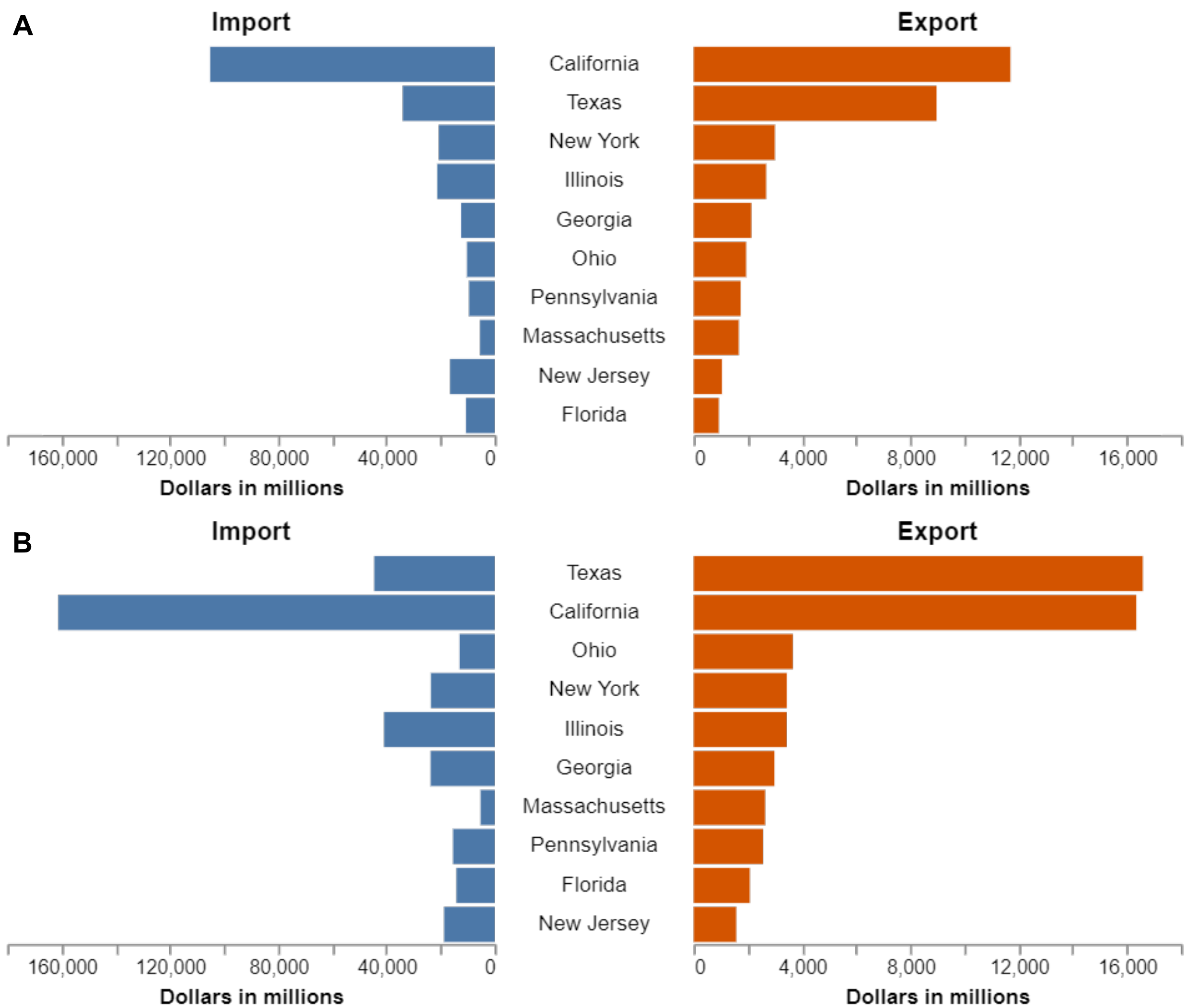


Figure 6. Pyramid Bar Graph of Trade between China and Top 10 States by GDP: **A** 2008; **B** 2018.

Compared to the pyramid bar chart, the cholorepth map has the advantage of effectively showing 50 states together, as demonstrated from **Fig. 7**. It can be seen from the plot that, export and import did not change equally over time, although both have grown over the 10-year time span. While California was clearly the No.1 in export in 2008, Texas and Washington exported about the same as California in 2018. In addition, South Carolina seems to have increased its export very significantly from 2008 to 2018. On the other hand, as we have seen from the pyramid bar chart, all states seem to increase more or less the same in import during the same time period. It is worth noting that it is more challenging for human eyes to accurately recognize the quantitative difference in color scale than in length scale. Therefore, it may make more sense for us to use the cholorepth maps in Figure 8 for semi-quantitative purpose, e.g. determining if the export is similar other than trying to accurately rank similar export numbers.

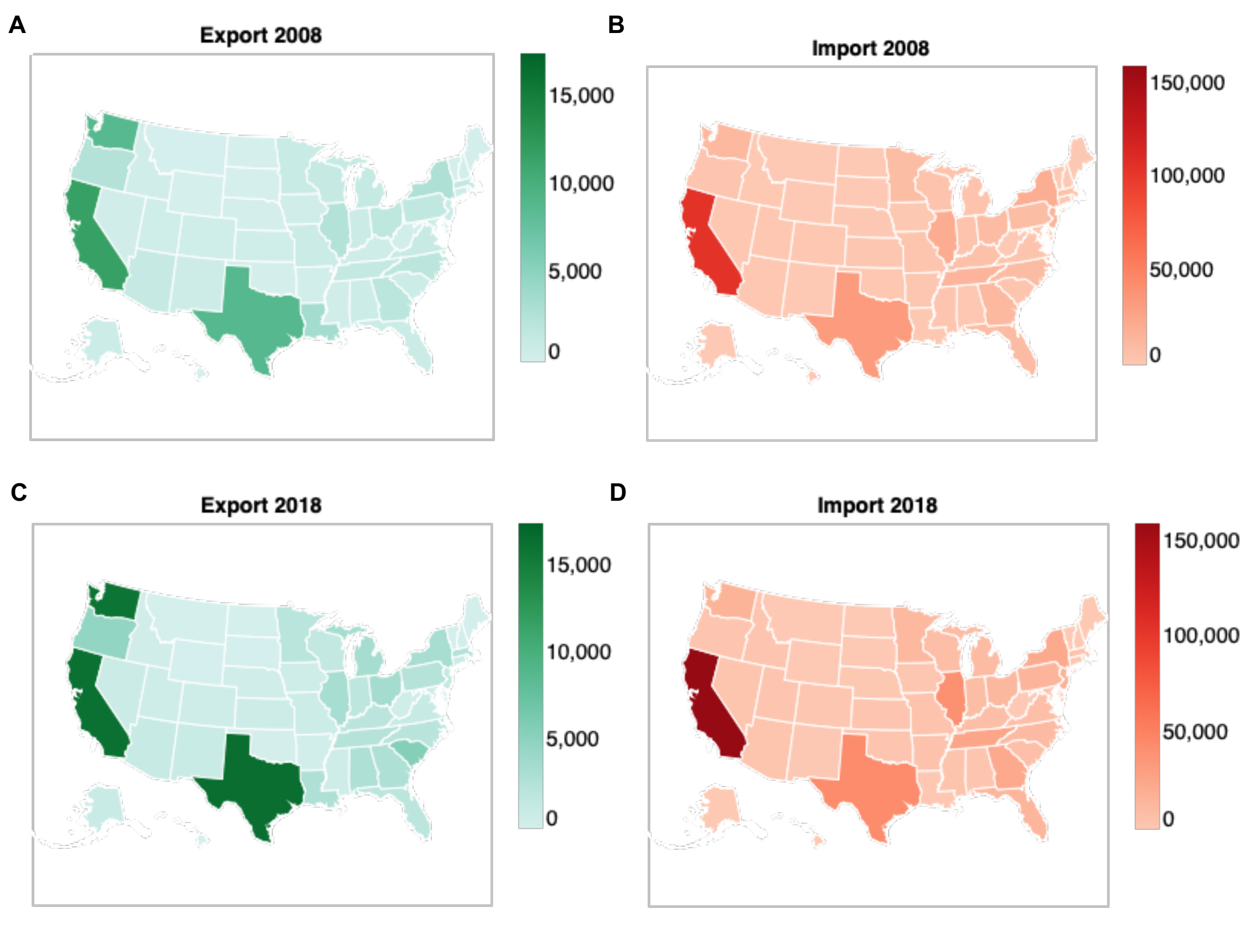


Figure 7. Cholorepth Map: **A** Export 2008; **B** Import 2008; **C** Export 2018; and **D** Export 2018.

While **Fig. 6** and **7** essentially show the comparison of different states at the same time point, Figure 8 uses time-series to illustrate the trend of the export, import and surplus/deficit over time for each individual state. Therefore, the time-series plots allow us to study in more details the change of trade between a certain state and China and potentially dig deeper into the underlying reasons. For example, four different patterns are shown in **Fig**. **9**: **A** export stayed the same, while import increased; **B** both export and import increased; **C** both export and import stayed the same; and **D** other (irregular) patterns.

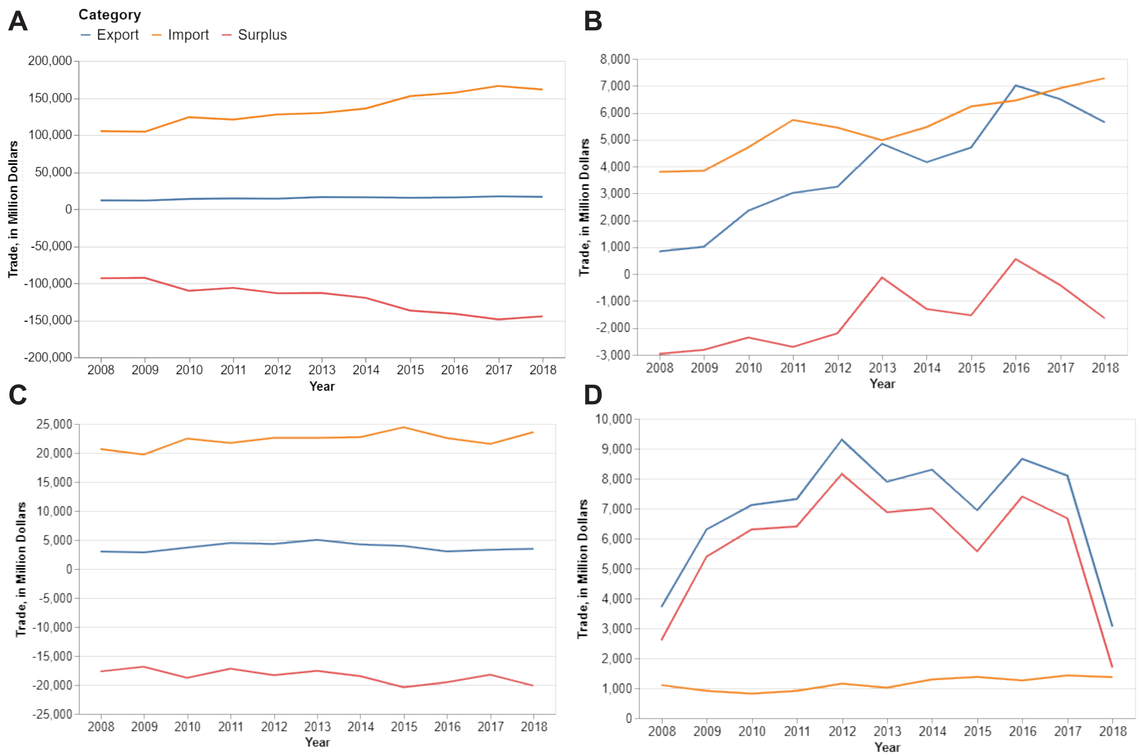


Figure 8. Time Series of Export, Import and Surplus/Deficit: **A** California; **B** South Carolina; **C** New York; and **D** Lousianna

Similar to **Fig. 8** and **9** uses time-series to visualize both the absolute export to China and the percent of export of China in total export for two individual states. In general, the two lines both increased over time, as shown in **Fig. 9A**. However, we have also observed a few cases where the change was irregular. One such example is the state of Vermont, as shown in **Fig. 9B**. The export to China from Vermont grew rapidly after the recession in 2007, peaked at 2011 and then rapidly dropped. In fact, the decline in exports was largely due to the drop in electronic parts and components, probably because the intermediate electronic goods from other countries became much competitive.

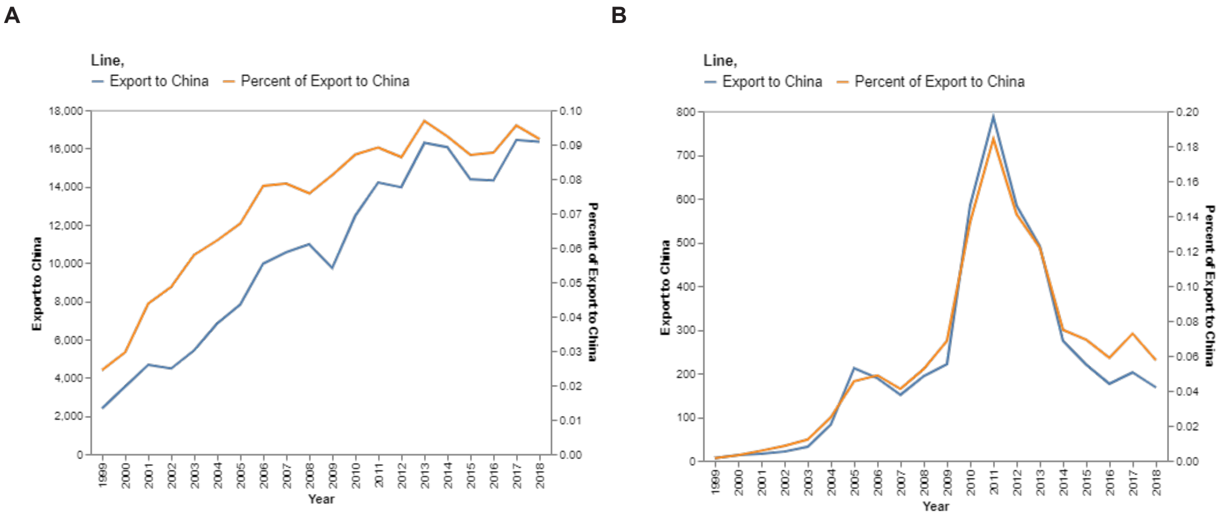


Figure 9. Time Series of Export to China and Percent of Export to China in Total Export: **A** California; **B** Vermont.

U.S. exports to major world's trading partners

We also investigated the fractions of U.S. exports to the major world's trade partners (China, Japan, EU, NAFTA, others) in the top 10 U.S. States (by GDP) (**Fig. 10**). In general, NAFTA is the single largest export destination for these 10 states. Ohio (51.4%), Texas (43.5%), and Illinois (42.5%) are among the top 3 states whose exports to NAFTA accounted for more than 40% of total merchandise trade; For the EU countries, the top 3 export states are Massachusetts (28.9%), New Jersey (26.7%), and New York (24.6%), which is not surprising considering that there resides a vast population of European descendants and immigrants. Intriguingly, the fractions of exports to China are relatively small compared to other trading partners (the export to Japan is similar). No state has an export fraction more than 10%. The top three export-leading states are Massachusetts (9.7%), California (9.2%), and Georgia (7.4%). However, due to the huge economy of these states (e.g. California's GDP could rank fifth in the world if it were regarded as a country), a nearly 10% of export fractions may still have a considerable impact on the overall U.S. exports.

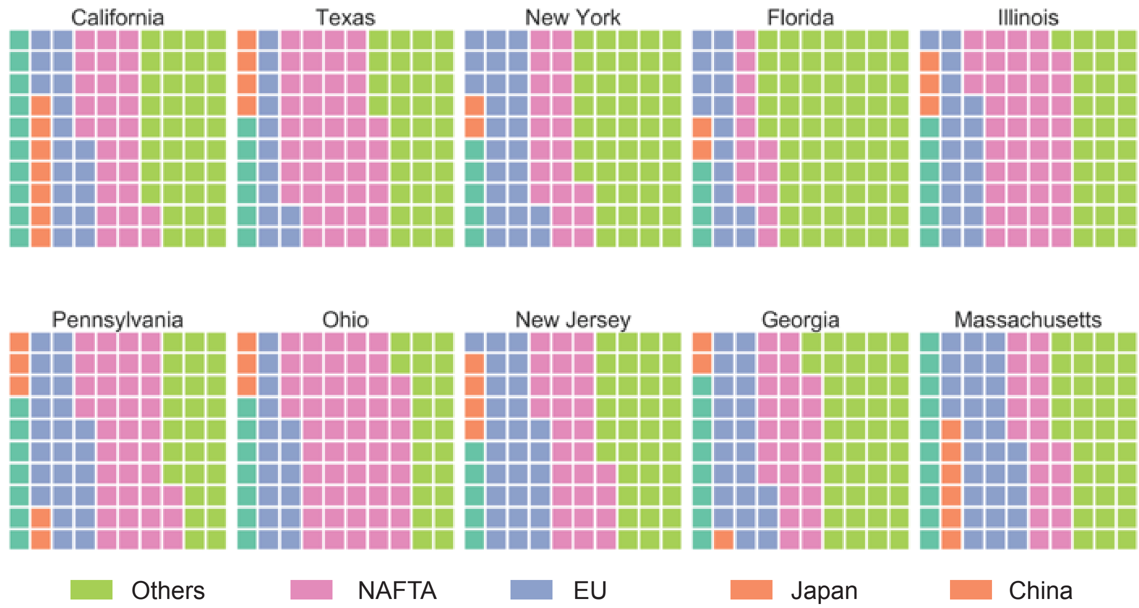


Figure 10: Fractions of U.S. exports to major world's trading partners

PART II: Classification of U.S. states based on manifold learning

Different from the traditional, distance-based dimension reduction techniques (such as PCA or MDS), the Uniform Manifold Approximation and Projection (UMAP) is the latest, weighted k-neighbor graph-based learning algorithm for general purpose dimension reduction[14]. Similar to t-SNE, UMAP can better preserve fine detailed local manifold structure in the low dimensional embedding space; meanwhile, it also captures more global and topological structure than the former. Due to its ability to effectively reveal local data structures, it has been successfully applied in clustering task on datasets with homogenous data types[15].

Based on our previous visualizations, it is not difficult to see that different U.S. states have distinct characteristics influenced by multiculturalism, geographic location, political preference and export orientation. In order to gain insights into potential intrinsic links between these states, we further built a manifold learning model (UMAP) that attempts to classify the U.S. states based on these characteristics. Specifically, we first constructed a dataset that included a number of export-related economic data (e.g. economic aggregates (GDP), exports to major trading partners and their fractions, exports to China and their fractions), as well as geographical location and political preference as features for manifold learning. We then conducted grid search to explore the optimal hyperparameter space of UMAP for values of n_neighbor (number of neighbors) and min_dist (desired separation of the neighboring points in the embedding space) (**Fig. 11**). As shown below, the optimal clustering was achieved with n_neighbor = 3 and min-dist = 0.001, which successfully yielded 4 meaningful structures of U.S. state clusters (**Fig. 12**).

- ***Category I*****:** Washington, South Carolina, Oregon, Alabama, Kentucky, Puerto Rico, New Mexico, Alaska, Connecticut, Utah, New Hampshire
- ***Category II*****:** Texas, California, New York, Illinois, Georgia, Massachusetts, Pennsylvania, Tennessee, North Carolina, Minnesota, Florida, New Jersey, Virginia, Maryland, Colorado
- ***Category III*****:** Louisiana, Nevada, Mississippi, West Virginia, Idaho, Delaware, Arkansas, Maine, Vermont, Rhode Island, Montana, Wyoming, Hawaii, District of Columbia
- ***Category IV*****:** Ohio, Michigan, Indiana, Wisconsin, Arizona, Missouri, Kansas, Iowa, Nebraska, Oklahoma, South Dakota, North Dakota

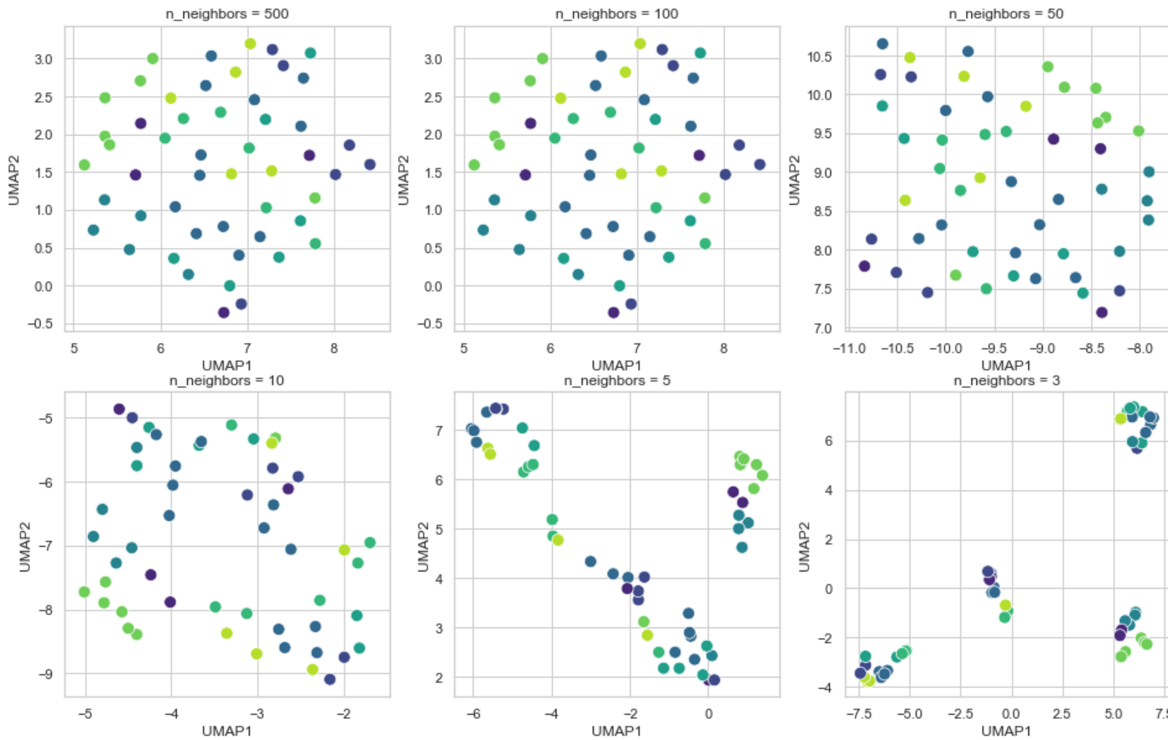


Figure 11: Hyperparameter optimization of UMAP manifold learning

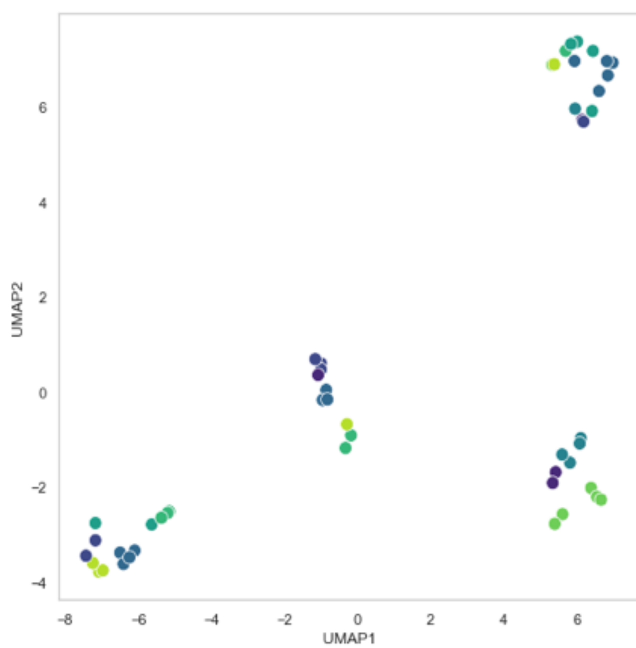


Figure 12: UMAP embeddings of U.S. state clusters

The states of these 4 categories have the following characteristics in terms of geographical locations (**Fig. 13**):

- Category I has a wide geographical span and presents a sporadic distribution. It includes Washington and Oregon in the northwestern U.S., New Mexico on the southern border of the U.S., Tennessee in the middle, Alabama and South Carolina in the southwest, and New Hampshire and Connecticut in the northeastern U.S.
- Category II is heavily concentrated on the East and West coasts of the U.S. (California, Massachusetts, Pennsylvania, New York), the Great Lakes region (Illinois), and some of the economic centers in the southern U.S. (Texas, Georgia, Florida).
- States of categories III and IV are predominantly located in the Midwestern U.S. that includes some traditional agricultural states such as Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio and Wisconsin.

From the perspective of political preference, the states of category I and category III have balanced Democratic and Republican parties; the Democratic Party is slightly dominant in the category II states with no obvious difference. Of note, the Republican party dominates 100% of the states in category IV (**Fig. 14**). In terms of export orientation, the four categories of U.S. states present distinct patterns as shown by the waffle chart (**Fig. 15**). It is clear that the type I states rank 1st in terms of fraction of exports to China (~15%); the category I states also take the lead in exporting to the EU. The export destinations of category IV states are dominated by NAFTA, which accounts for more than 50% of total merchandise trade. In general, all four categories of states have relatively small fraction of export to Japan, compared to other major trading partners.

Economically, category II states have the largest GDP and imports from China among all four categories. In terms of absolute volume, the states of category I and II dominated exports to China. However, intriguingly, only exports of category I states accounted for more than 15% of the U.S.'s total exports, with the states of the other three categories all significantly left behind (**Fig. 16**). Taken together, we expect that category II states would have a strong impact on U.S. exports due to their huge import & export volume and overall GDP. Additionally, although the category I states have relatively small GDP and imports from China, it would also be highly possible that their economies would be more vulnerable to the effect of trade war given their large-scale exports to China and the leading export fractions.

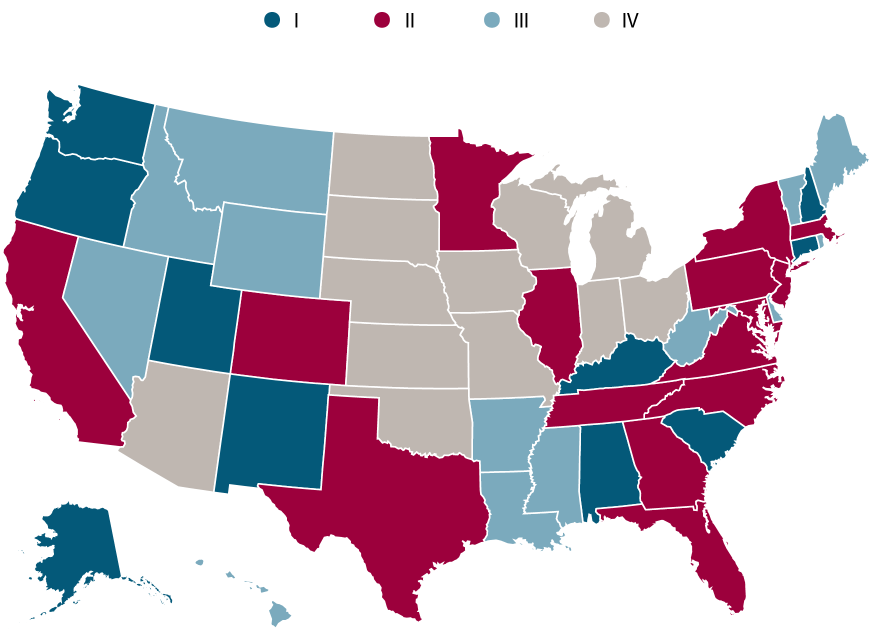


Figure 13: Geographical distribution of the 4 categories of U.S states

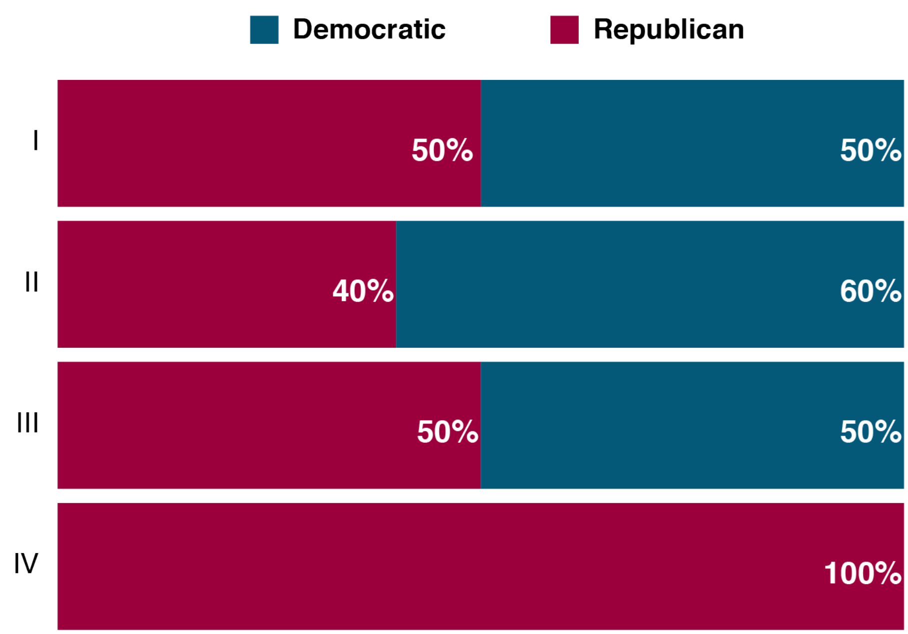


Figure 14: Political preferences among 4 categories of U.S states

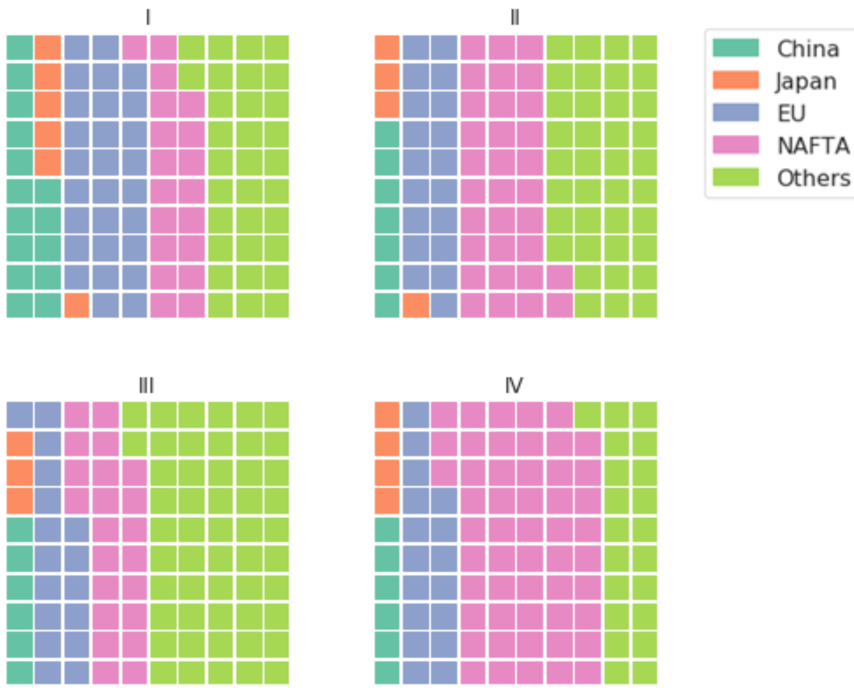


Figure 15: Export orientation of the 4 categories of U.S states

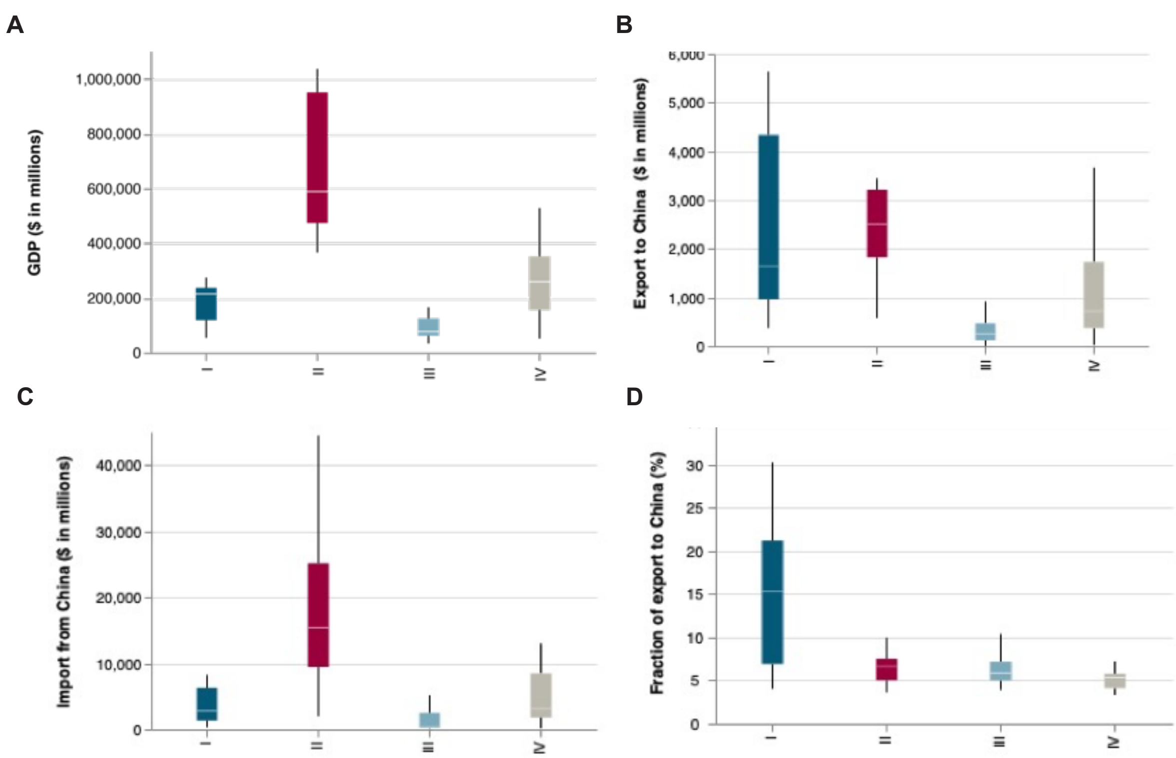


Figure 16: Economic and trade characteristics of the 4 categories of U.S states

DISCUSSION & CONCLUSION

In this study, we have used different visualization techniques (bar chart, choropleth and times-series) to illustrate the trade data between China and individual U.S. states. As we have discussed in the previous section, each of the techniques has its strength and weakness: bar chart is able to compare different trade categories (export and import) and different states simultaneously, while it is not the most appropriate if we want to show a lot of states in one graph; choropleth allows us to compare all fifty states all together intuitively, while it does not convey quantitative information very accurately ; time-series show the trend of trade of an individual state over time, but it can only show one state at a time.

To the best of our knowledge, this is the first study to build a machine learning model for U.S. state classification using data from U.S. foreign trade, geography, and party preferences. In this study, we explored the application of manifold learning in solving classification problems in economic data. We used the latest universal dimension reduction technique - UMAP, that is significantly faster than t-SNE and exhibits better classification results (the classification result of t-SNE is not shown). This study has several limitations. First, in this study, we did not attempt to distinguish types of exports of different U.S. states for visualization. Second, the dataset used for building the manifold learning model only includes the latest historical data for 2018, even though the data coincides with the starting time of the U.S.-China trade war. Lastly, because the trade war is still an ongoing event, its economic consequence could take months or even years to show up. This makes it impossible for us to fully assess the impact of the trade war on the U.S. economy and its states.

Due to time limits, we haven't explored many of the areas that could expand the findings obtained in this study. We would propose the following 4 directions for future work: 1) Investigate the underlying reasons for the different trends observed for different states; 2) Study the details of export and import portfolios by different states (e.g. what kinds of goods each state imports and how they change over time); 3) Build models of the trade data (e.g. ARIMA models) and perform forecasting based on different scenarios; and 4) Compare the results from the modeling work with manifold learning results.

REFERENCES

1. WORLD TRADE STATISTICAL REVIEW 2018. (2019). Retrieved from https://www.wto.org/english/res_e/statis_e/wts2018_e/wts2018_e.pdf
2. US Census Bureau's official website. Retrieved from https://www.census.gov
3. The People's Republic of China. (2019). Retrieved from https://ustr.gov/countries-regions/china-mongolia-taiwan/peoples-republic-china#
4. The US-China trade war in charts. (2019). Retrieved from https://www.bbc.com/news/business-48196495
5. International Trade Administration. Retrieved from http://tse.export.gov/TSE/ChartDisplay.aspx
6. Kafchinski, J. (2019). Foreign Trade: U.S. Top Trading Partners. Retrieved from https://www.census.gov/foreign-trade/statistics/highlights/top/index.html
7. OEC - United States (USA) Exports, Imports, and Trade Partners. (2019). Retrieved from https://oec.world/en/profile/country/usa/
8. International Trade Administration,, State-by-State Export from A Selected Market, http://tse.export.gov/TSE/TSEOptions.aspx?ReportID=101&Referrer=TSEReports.aspx&DataSource=SED
9. International Trade Administration, State-by-State Import from A Selected Market, http://tse.export.gov/stateimports/TSIOptions.aspx?ReportID=101&Referrer=TSIReports.aspx&DataSource=SED
10. US Census Bureau, Trade in Goods with China, https://www.census.gov/foreign-trade/balance/c5700.html
11. St Louis Federal Reserve, Producer Price for All Commodities, https://fred.stlouisfed.org/series/PPIACO
12. US Federal Election Commision, Official 2016 Presidential General Election Result, https://transition.fec.gov/pubrec/fe2016/2016presgeresults.pdf
13. Bureau of Economic Analysis, US Department of Commerce, Gross Domestic Product by State, https://www.bea.gov/system/files/2019-04/qgdpstate0519_4.pdf
14. Becht, E., McInnes, L., Healy, J., Dutertre, C., Kwok, I., & Ng, L. et al. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nature Biotechnology, 37(1), 38-44. doi: 10.1038/nbt.4314
15. McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. Journal Of Open Source Software, 3(29), 861. doi: 10.21105/joss.00861