COVID-19 Pandemic Impact on the US Refrigerated Truck Transportation and Fruits & Product Prices

Authors

Quoc-Huy Nguyen (quochuyn), Yifei Dong (yifeid), Ning Xu (nixu)

Motivation

The trucking industry remains one of the most important means of transportation of goods in the US. This is especially the case with perishable items, which require quick and efficient transportation. The Agricultural Marketing Service of the USDA regularly issues reports to provide a consistent overview of the state of the agricultural market. One particular report tracks the daily refrigerated truck volume of fruits and vegetables from US domestic origins. Most movements are reported, but not all commodities nor origins are included.

Food and product prices have been constantly increasing ever since COVID-19 hit the US in early 2020. Holding the hypothesis that the COVID-19 pandemic has impacted much of the food supply chain, thus increasing food prices, we want to conduct a study that can prove the relation between these three factors: COVID-19 cases, refrigerated truck volume, and refrigerated product prices.

From a previous study by the U.S. Bureau Of Labor Statistics[3], the CPI for fruit and vegetables in Q1 2020 was negative 1.9 indicating a big drop in prices. In this study, we will reference data from various sources and conduct analysis to find out whether these hypotheses are true and look at the long-term impacts of the COVID-19 pandemic.

Data Sources

Name	Description	Main Variables	Details	Access
Refrigerated Truck Volumes	The primary data set on refrigerated truck volumes on various staple fruits and vegetables by region. Some of the data collection methods used include Federal marketing orders, telephone interviews, faxes, emails, and access to other data sources. The time frame of interest will be from Jan. 1, 2015 to Dec. 31, 2022.	Date, region of origin, commodity, volume	**Size**: 108 MB **Shape**: 1M rows & 14 columns **Format**: CSV	USDA
COVID-19 Cases	The COVID-19 data set is a preprocessed clean data set with weekly updated COVID-19 total cases and death counts aggregated on the state level of 60 different states. The time frame covers Jan. 23, 2020 to Jan. 12, 2023.	Date, State, Total_cases	**Size**: 550 KB **Shape**: 9K rows & 10 columns **Format**: CSV	CDC
Fruit Prices	The fruit prices data set comes from the Economic Research Service department of the USDA. The data allows for observations of the weekly economic performance of various fruits over time. The time frame covers May 2, 2020 to Apr. 23, 2022.	Week, commodity, retailed price	**Size**: 976 KB **Shape**: 2K rows & 3 columns **Format**: XLSX	USDA

Data Manipulation Methods

Truck Volume Data Set

The steps for importing and cleaning the refrigerated truck volume data set are as follows:

1. Aggregating some rows due to unnecessary columns. Some rows for truck volume may be assigned to different commodity marketing seasons (e.g., Jan. 1, 2015 assigned to 2015 and 2014) indicated by the `Season` column, but otherwise are complete duplicates. For our analysis, we disregard this fact because we are more interested in the date of the movement. The method for aggregation is a summation.
2. Keeping only important columns. There are extra columns relating to the date such as which week out of the year it is.
3. Removing rows with no/useless information. Particularly, rows that have a value of 0 for the `10,000 LBS` column (the column for the truck volume) are removed. This is because the column is encoded as an integer value. Therefore, truck volumes below the 10,000 lbs threshold will just have a value of 0. As for negative values, we decide to keep them as the metadata does not provide a reason for their existence. For the few rows that do have negative values, they may be for readjusting previous movements.
4. Casting the data type of the `date` column to be a DateTime type. This allows us to take advantage of the powerful time series manipulation methods within pandas.
5. Resampling/aggregating rows based on a DateTime frequency conversion (e.g., every month or every week) similar to the `resample()` method of a `pandas.DataFrame`. This is due to the fact that we have very large data where the truck data set covers from Jan. 1, 2015 to Dec. 31, 2022 on a daily basis. When grouping the data based on its columns, additional handling is necessary with the use of the `pandas.Group` object. This object allows us to resample the `date` column while also grouping by the other columns.
6. Filtering to only rows for California and Florida for downstream analysis.

COVID-19 Data Set

Since the covid-19 data set is a preprocessed data set without missing values and obvious outliers. All we did for this dataset is to read it and filter out the target region of California and Florida which is abbreviated for “CA” and “FL” in the column “state” in the weekly covid-19 cases dataset.

Fruit Price Data Set

The time range of the Fruit-weekly-price data set is from 2020-05-02 to 2022-04-23, the whole range of which is within the time range covered in covid-19 dataset, therefore we kept all the records from the time range perspective. The data set was in an xlsx format, and several manipulation steps were done after importing the file:

1. The sheets had a lot of wasted areas that are for formatting, and annotating the data. Only specific columns and rows were taken from each sheet.

2. The file was organized in a way that each week's data was stored in a separate sheet. Therefore, modulation needed to be built to combine all weeks' data as columns into one data frame.
3. Then, for the convenience of generating the visualization plots and the coherence of the format with the other two datasets, we also converted this data frame to a long format shape using the variable "date" as a new column and representing each commodity's price in each week with the `pandas.melt` function using column "Commodity" as id_vars and value_vars as sorted dates by ascending order.
4. After comparing the unique value of "Commodity" with the ones in the *Truck Volume* dataset, we found some dispatches between the names and needed to correct them in this dataset so it can be correctly joined with the *Truck Volume* dataset.
5. This dataset uses a different day of the week as the week start than what's used in the *COVID* dataset. We used date between ranges to join this dataset with the *COVID* one.
6. There were a lot of missing values for different commodities in different weeks. To make the missing prices closer to the real prices, we decide to impute the missing prices' value with the commodity's nearest week's / record history price. Therefore, we apply "groupby" function on the "Commodity" column of the converted long format data frame generated previously and fill null values with the method "bfill". The "bfill" method stands for back-fill which takes the previous value after sorting by the date.

Analysis

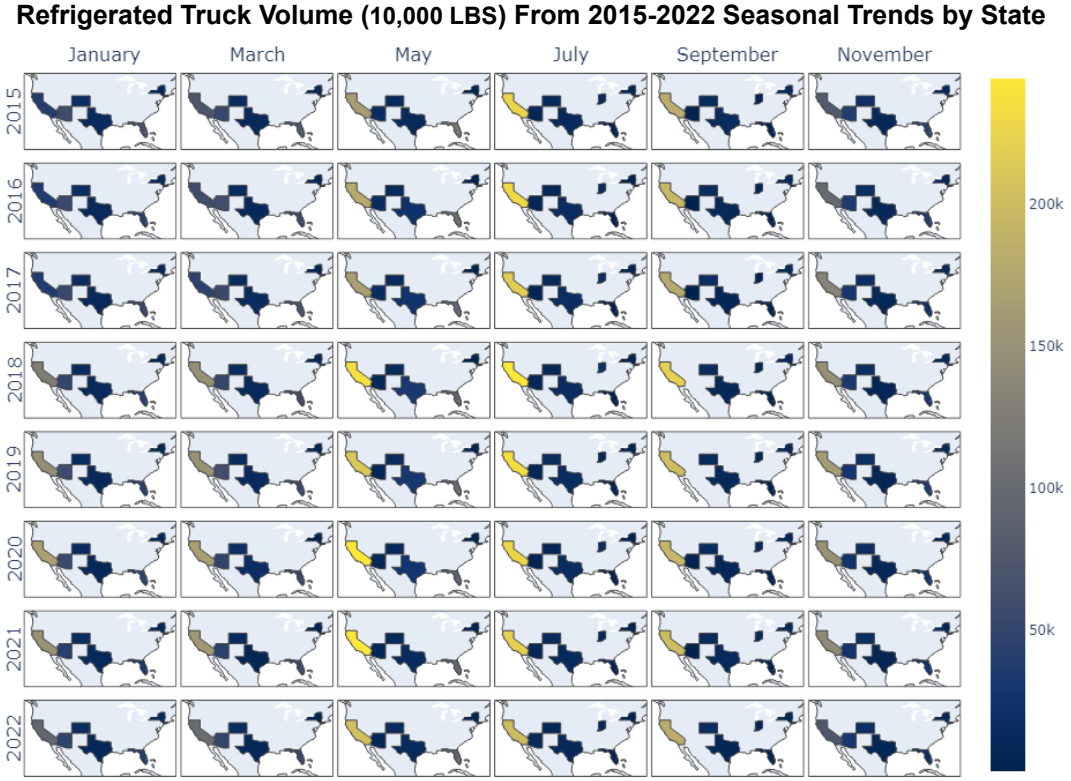
Our goal is to prove that COVID-19 has impacts on refrigerated truck volume and thus changes fruit prices. However, our hypothesis was proved wrong because we didn't find strong correlations between either COVID-19 cases and refrigerated truck volume, or COVID-19 cases and fruit prices. Therefore, we conducted some more analysis to find out what can be some of the other factors that impact refrigerated truck volume and fruit price. We found obvious seasonal patterns in both data sets. The following visualizations will better show our findings.

Visualizations

Choropleth Map

The below is an attempt at using a small multiples visualization for exploratory data analysis. Each visualization is a choropleth map for each month of the year where the total refrigerated truck volume is encoded as the color and the level of aggregation is on the state level. As a caveat, not all states are included in the data set. Although there are defined regions such as the Pacific Northwest that group multiple states, due to the difficulty of redefining geometric boundaries those regions will not be included in this visualization.

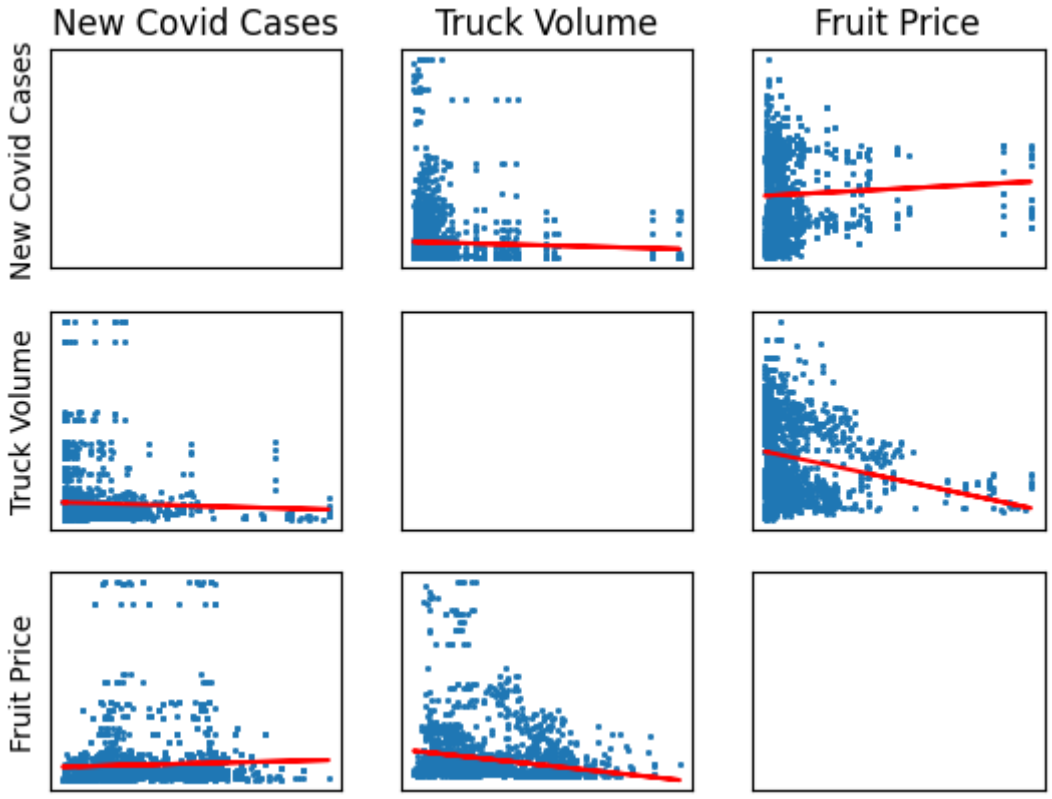
First, notice the accented yellow hues that indicate a high value of total refrigerated truck volume. Immediately, we notice two trends: (1) California is consistently the overwhelming leader in total refrigerated truck volume compared to the other states and (2) California backs down during the non-summer months indicating a seasonal/periodical trend.



SPLOM

Below is a SPLOM chart showing scatterplots between each pair of the three variables (new Covid case numbers, truck volume, and fruit price) and their regression trend line. We can see that the most obvious correlation is the negative relationship between fruit price and truck volume. Next is the slightly positive correlation between new Covid cases and fruit prices. We barely see any correlation between Covid case increases and truck volume changes.

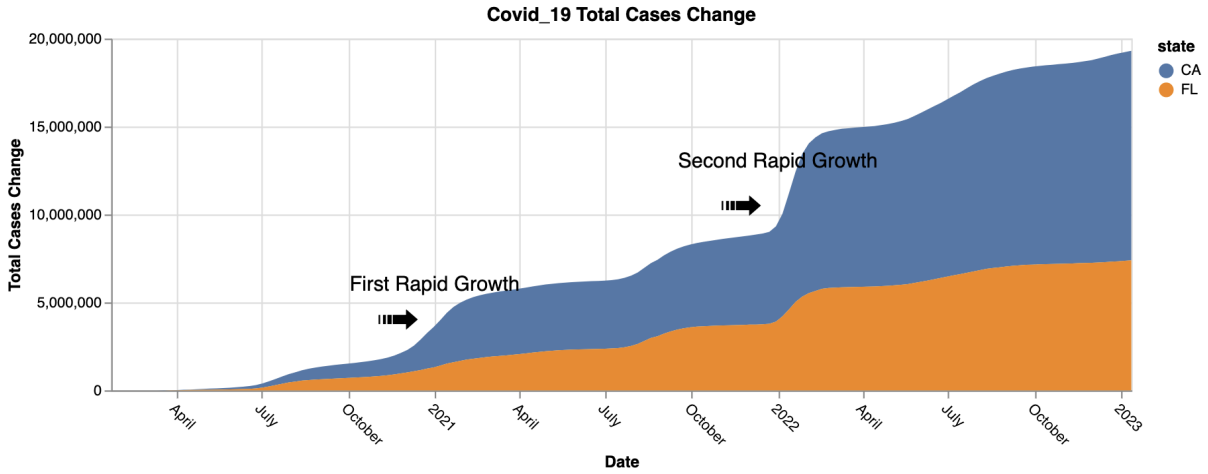
Scatterplot Matrix Between Covid Cases, Truck Volume, Fruit Price, with Regression Lines



(Diagonal graphs were intentionally left blank to not show the correlation between a variable with itself.)
Since we didn't find much correlation between COVID-19 cases and the other two variables, we decided to look deeper into each variable for their trend.

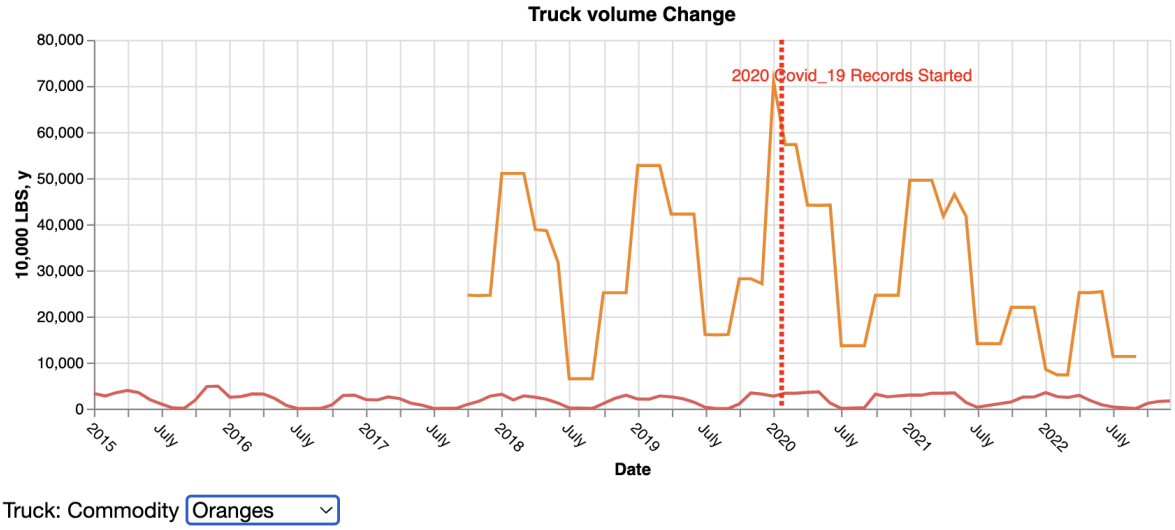
Concatenated Interactive Line Plots

Below is a line chart showing cumulative COVID-19 cases in both CA and FL.



We can see that the total number of cases for California is much higher than in Florida about 4 times up to the beginning of 2023. During the case expansion period, we found that there was a significant increase in the total cases for both target states in the first week of 2021 and 2022. Besides, after the first week of 2022, the total Covid-19 cases in California has increased about three times faster than in the state of Florida reaching about 20,000,000 cases. By computation and sorting by the weekly new_cases increase and visualization, there are 2 rapid growth periods, one is from 2020-12-17 to 2021-01-14 and the other was from 2022-01-06 to 2022-02-03. Both of these periods are concentrated from the end of last year to the beginning of the next year, which is winter for both states.

The following chart shows truck volume with a red dash stroke dividing the period before and after Covid-19 began (Using Oranges as an example).



In the notebook, by choosing different commodities from the dropdown list of “Truck: Commodity” we can compare horizontally the parts to the left and right of the red divide line. It can be found that:

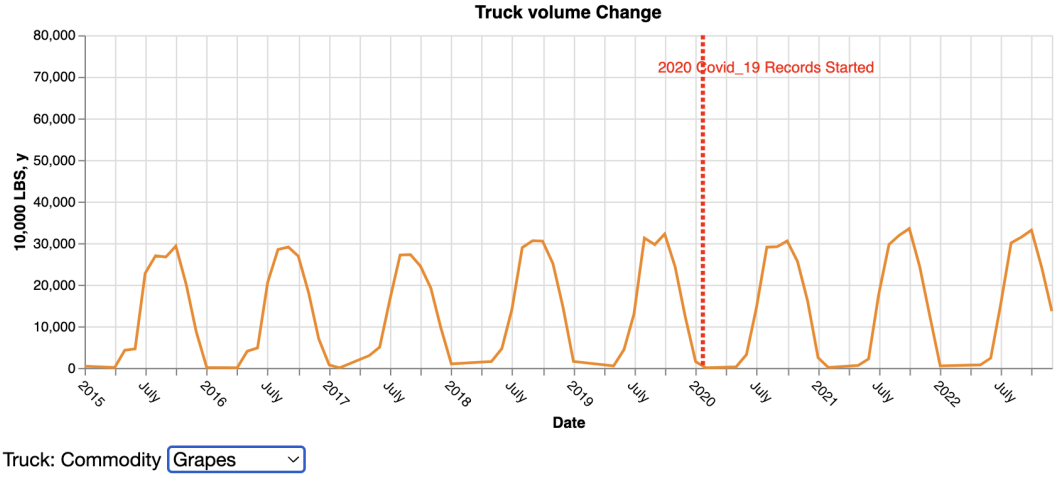
1. The monthly total truck transportation volume for each fruit keeps stable at the same level before and after Covid-19 for both states. If we look at the date exactly around the red line, we can see there is a significant decrease after the peak around the time that

Covid-19 began, but it still followed the same tendency of the same time period before Covid-19 started.

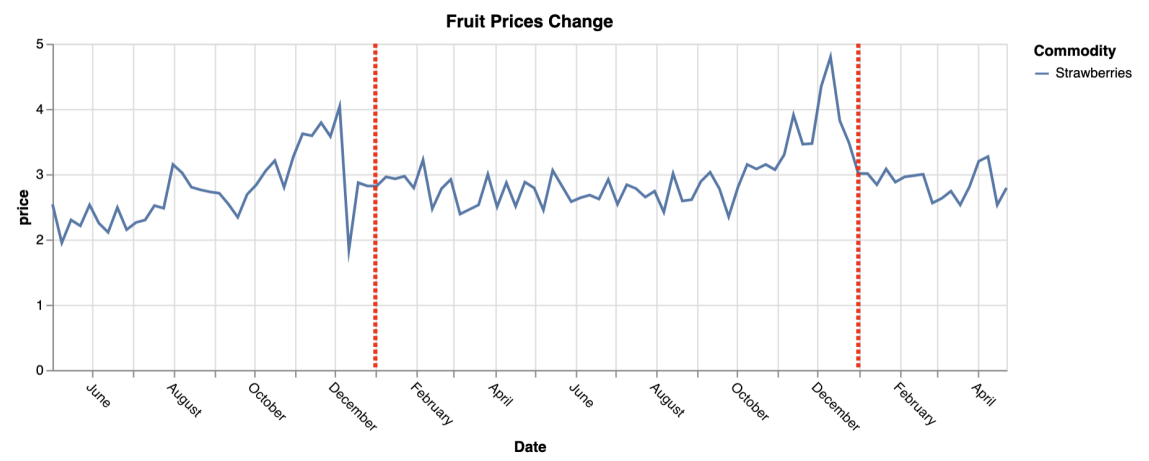
2. Under the prerequisite of the previous point, we find that there are only truck transportation volumes for apples, cherries, grapes, lemons, peaches, pears, and raspberries in California but no truck volume records in the state of Florida. The difference in the quantity and changing direction of truck volumes for the same fruit is caused by the geographical condition of the origin of the specific products, not due to the impact of covid cases increase. For the commodities that have truck transportation for both states, the total truck volume for the same fruit type is shown as follows:

Commodity	**California vs. Florida**
Strawberries	5 times higher
Grapefruit	Almost the same but opposite tendency during the same time period.
Blueberries	Almost the same, with the same trend during the same time period.

3. For the fruit with obvious seasonal production like cherries and strawberries, the truck volume is fluctuating completely with the seasonal production and supply change presenting an obvious periodical characteristic as the following (we use grapes as an example). We can see that the truck volume is relatively very high up to 30,000 (10,000lbs) during the summer and very low to almost 0 in the state of California (referenced below using Grapes as an example).



4. As for fruit prices, we marked 2 red divided lines for the first and second rapid growth periods of Covid-19 cases. We observed that the fruit price for all the fruits decreased around the first growth period (beginning of 2021) and the second fast growth (beginning of 2022). In addition, this is consistent with the study report done by the U.S. Bureau of Labor Statistics[3] that the 12-month (March 2019 - March 2020) percentage change for fruits and vegetables was -1.9 which means that on average the urban household expenditure on fruit decreased by 1.9% compared to the period before Covid-19 pandemic. We can see the exact figure in the following table.



Fruit Price : Commodity Strawberries

Conclusion

In conclusion, our initial hypothesis that COVID-19 has a strong impact on the refrigerated truck volume and fruit prices was not proven by our analysis as we didn't find strong correlations between the number of new COVID-19 cases with either of the other two variables. There was indeed a sudden and obvious fruit price drop when COVID-19 first hit the US, but that seemed like a one-time shock. Rather, we found that the truck volume and fruit prices mainly have seasonal trends with winter times truck volume lower and fruit prices higher, and summer times the opposite.

Statement of Work

All: Initial research on the theme, data source, and hypothesis. Meet weekly to review each other's work and provide feedback. Edit and modify the report.
Quoc-Huy Nguyen: Import, manipulate, and clean, the refrigerated truck volume data set. Set up a GitHub repository. Create choropleth map small multiples visualization. Work on the sections for the motivation, data sources and manipulation, and statement of work of the report. Aggregate and clean up the Jupyter notebook to be cohesive and understandable.
Ning Xu: Manipulate and clean the COVID-19 cases data set. Create line plot visualizations to show patterns/trends of truck volume, COVID-19 cases, and product prices over time. Work on the sections for the data sources, data manipulation methods, visualization, and analysis of the report.
Yifei Dong: Manipulate and clean the fruit product prices data set. Create SPLOM visualization between truck volume, COVID-19 cases, and product prices while fitting the best lines of fit to determine correlation. Work on the sections for the data manipulation, analysis, visualizations, and conclusion of the report. Polish language of the report of its professionalism and insightfulness.

References

1. Billiau, Seth. "Weekly United States COVID-19 Cases and Deaths by State." *Centers for Disease Control and Prevention*, 16 Feb. 2023, https://data.cdc.gov/Case-Surveillance/Weekly-United-States-COVID-19-Cases-and-Deaths-by-/pwn4-m3yp.
2. Gastelle, Jesse. "Refrigerated Truck Volumes: Open AG Transport Data." *United States Department of Agriculture, Agricultural Marketing Service*, 9 Feb. 2023, https://agtransport.usda.gov/Truck/Refrigerated-Truck-Volumes/rfpn-7etz.
3. Mead, Dave, et al. "The Impact of the COVID-19 Pandemic on Food Price Indexes and Data Collection : Monthly Labor Review." *U.S. Bureau of Labor Statistics*, Aug. 2020, https://www.bls.gov/opub/mlr/2020/article/the-impact-of-the-covid-19-pandemic-on-food-price-indexes-and-data-collection.htm.
4. Weber, Catharine. "Selected Weekly Fruit Movement and Price." *United States Department of Agriculture, Economic Research Service*, 17 June 2022, https://www.ers.usda.gov/data-products/fruit-and-tree-nuts-data/selected-weekly-fruit-movement-and-price/.