Strawberries Production in the US

An Exploratory Data Analysis

Nicole Kingdon

October 18, 2023

Table of contents

Introduction **2**
- Strawberries & Positive Health . . . 2
- Strawberries & Pesticides . . . 2
- Organic Strawberries . . . 3
- Analyzing Strawberry Production . . . 3

Data Acquisition & Assessment **5**
- USDA-NASS Data . . . 5
- Census Data . . . 5
- Survey Data . . . 6
- States . . . 7
- Years . . . 7
- Assumptions & Motivations . . . 7
  - Census Data . . . 7
  - Survey Data . . . 8

Data Cleaning & Organizing **9**
- R Packages . . . 9
- Initial Data Cleaning & Organization . . . 9
- Separating Data Frames . . . 9
  - Census . . . 10
  - Survey . . . 11

Exploratory Data Analysis **15**

References **15**

Introduction

Strawberries & Positive Health

Strawberries are a fruit that holds several minerals, vitamins, and nutrients (CRAIG 1997), which have positive implications on human health (Afrin et al. 2016). Specifically, strawberries have been found to help reduce likelihood of cancer, diabetes, obesity, neurodegeneration, cardiovascular disease, and metabolic syndrome (see Figure 1) (Afrin et al. 2016). Although strawberries as a healthy food is the norm, pesticides appear to be harming the beneficial factors of this fruit.

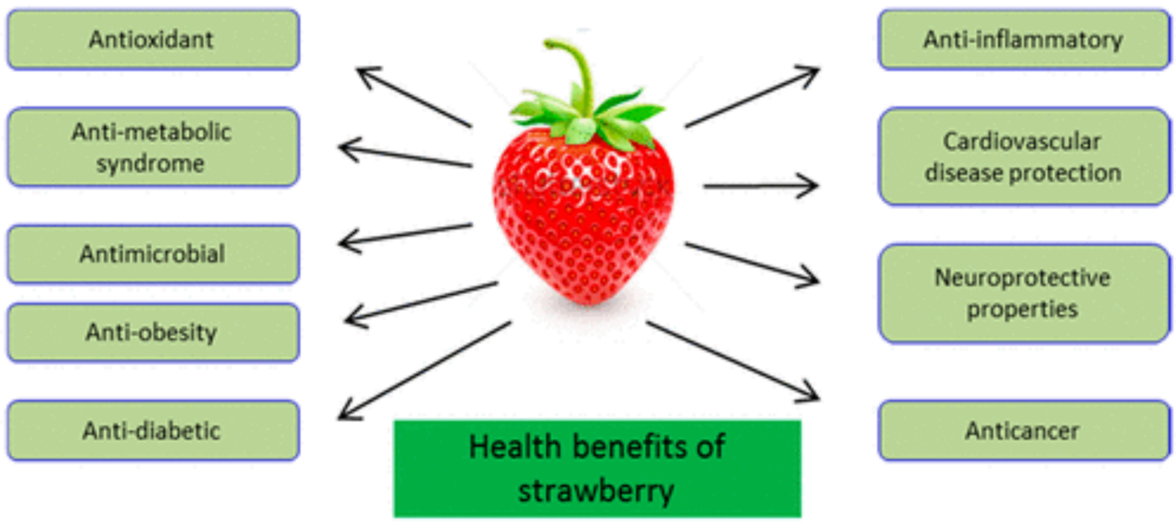


Figure 1: Health benefits of strawberries (Afrin et al., 2016)

Strawberries & Pesticides

Pesticides are used on fruit and vegetable crops, including strawberries, with hopes to increase the quantity (Fenik, Tankiewicz, and Biziuk 2011). Pesticides are made of chemical compounds to reduce or completely eliminate pests from impacting crops (Afrin et al. 2016). These chemical compounds may increase the yield of the crop, but may have a large risk on human health. Additionally, they may contaminate bodies of water and soil with the chemicals, help pests develop resistance to the chemicals, and impact helpful organisms from persisting in areas where pesticides are used. Overall, there are positive and negative impacts of using pesticides (see Figure 2), but it is important to further examine these impacts, specifically on strawberries, to understand the implications of using such.

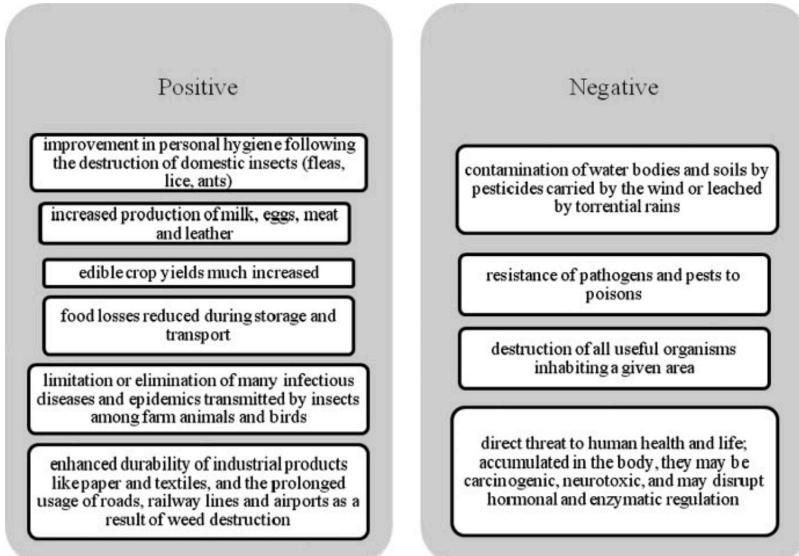
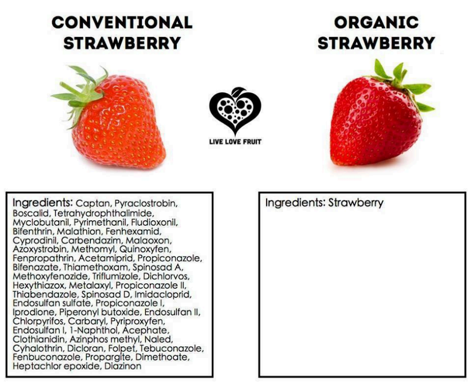


Figure 2: Positive and negative implications of pesticides on fruit and vegetable crops (Fenik et al., 2011)

Organic Strawberries

Strawberries that are organic are simply strawberries. They use alternative and safer methods to production, and those fruits and vegetables, in this case strawberries, are considered organic (Verteramo Chiu and Gomez 2023). Along with these alternative methods comes a higher cost of production and care, which increases the price to the consumer.



Analyzing Strawberry Production

The literature varies on if organic or non-organic is environmentally better (Afrin et al. 2016). This exploratory data analysis on production of strawberries, which encompasses both processed (non-organic) and fresh (organic) market data, will help us better understand production of strawberries in the United States. Additionally, it dives deeper into the pesticide usage in states, which will provide more information as to the level of toxic ingredients that go into the production of non-organic (or conventional) strawberries .

Data Acquisition & Assessment

USDA-NASS Data

The data was acquired from U.S. Department of Agriculture (USDA) and the National Agricultural Statistics Service (NASS). The data was uploaded for data cleaning and organizing and exploratory data analysis by Professor Haviland Wright, who chose the following data: USDA-NASS.

The data frame loaded into the environment by Professor Wright is titled `strawberry`.

```
Rows: 4,314
Columns: 21
$ Program             <chr> "CENSUS", "CENSUS", "CENSUS", "CENSUS", "CENSUS", "~
$ Year                <dbl> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 202~
$ Period              <chr> "YEAR", "YEAR", "YEAR", "YEAR", "YEAR", "YEAR", "YE~
$ `Week Ending`       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `Geo Level`         <chr> "STATE", "STATE", "STATE", "STATE", "STATE", "STATE~
$ State               <chr> "ALASKA", "ALASKA", "ALASKA", "ALASKA", "ALASKA", "~
$ `State ANSI`        <chr> "02", "02", "02", "02", "02", "02", "02", "06", "06~
$ `Ag District`       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `Ag District Code`  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ County              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `County ANSI`       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `Zip Code`          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ Region              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ watershed_code      <chr> "00000000", "00000000", "00000000", "00000000", "00~
$ Watershed           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ Commodity           <chr> "STRAWBERRIES", "STRAWBERRIES", "STRAWBERRIES", "ST~
$ `Data Item`         <chr> "STRAWBERRIES, ORGANIC - OPERATIONS WITH SALES", "S~
$ Domain              <chr> "ORGANIC STATUS", "ORGANIC STATUS", "ORGANIC STATUS~
$ `Domain Category`   <chr> "ORGANIC STATUS: (NOP USDA CERTIFIED)", "ORGANIC ST~
$ Value               <chr> "2", "(D)", "(D)", "(D)", "2", "(D)", "(D)", "142",~
$ `CV (%)`            <chr> "(H)", "(D)", "(D)", "(D)", "(H)", "(D)", "(D)", "1~
```

Census Data

The data offers census data based on state that represents fresh market (organic) and process market (non-organic) sales.

```
Rows: 864
```

```
Columns: 21
$ Program            <chr> "CENSUS", "CENSUS", "CENSUS", "CENSUS", "CENSUS", "~
$ Year               <dbl> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 202~
$ Period             <chr> "YEAR", "YEAR", "YEAR", "YEAR", "YEAR", "YEAR", "YE~
$ `Week Ending`      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `Geo Level`        <chr> "STATE", "STATE", "STATE", "STATE", "STATE", "STATE~
$ State              <chr> "ALASKA", "ALASKA", "ALASKA", "ALASKA", "ALASKA", "~
$ `State ANSI`       <chr> "02", "02", "02", "02", "02", "02", "02", "06", "06~
$ `Ag District`      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `Ag District Code` <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ County             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `County ANSI`      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `Zip Code`         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ Region             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ watershed_code     <chr> "00000000", "00000000", "00000000", "00000000", "00~
$ Watershed          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ Commodity          <chr> "STRAWBERRIES", "STRAWBERRIES", "STRAWBERRIES", "ST~
$ `Data Item`        <chr> "STRAWBERRIES, ORGANIC - OPERATIONS WITH SALES", "S~
$ Domain             <chr> "ORGANIC STATUS", "ORGANIC STATUS", "ORGANIC STATUS~
$ `Domain Category`  <chr> "ORGANIC STATUS: (NOP USDA CERTIFIED)", "ORGANIC ST~
$ Value              <chr> "2", "(D)", "(D)", "(D)", "2", "(D)", "(D)", "142",~
$ `CV (%)`           <chr> "(H)", "(D)", "(D)", "(D)", "(H)", "(D)", "(D)", "1~
```

Survey Data

Additionally, it holds survey information for each state, specifically indicating pesticides and bacterium used to preserve strawberry crop yield. In addition, it offers fresh and process market data.

```
Rows: 3,450
Columns: 21
$ Program            <chr> "SURVEY", "SURVEY", "SURVEY", "SURVEY", "SURVEY", "~
$ Year               <dbl> 2022, 2022, 2022, 2022, 2022, 2022, 2022, 2022, 202~
$ Period             <chr> "MARKETING YEAR", "MARKETING YEAR", "MARKETING YEAR~
$ `Week Ending`      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `Geo Level`        <chr> "STATE", "STATE", "STATE", "STATE", "STATE", "STATE~
$ State              <chr> "CALIFORNIA", "CALIFORNIA", "CALIFORNIA", "FLORIDA"~
$ `State ANSI`       <chr> "06", "06", "06", "12", "12", "12", NA, NA, NA, "06~
$ `Ag District`      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `Ag District Code` <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ County             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ `County ANSI`      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```
$ `Zip Code`          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ Region              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ watershed_code      <chr> "00000000", "00000000", "00000000", "00000000", "00~
$ Watershed           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ Commodity           <chr> "STRAWBERRIES", "STRAWBERRIES", "STRAWBERRIES", "ST~
$ `Data Item`         <chr> "STRAWBERRIES - PRICE RECEIVED, MEASURED IN $ / CWT~
$ Domain              <chr> "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL~
$ `Domain Category`   <chr> "NOT SPECIFIED", "NOT SPECIFIED", "NOT SPECIFIED", ~
$ Value               <chr> "108", "(D)", "(D)", "169", "(D)", "(D)", "0", "135~
$ `CV (%)`            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

States

There were 47 states (c("ALASKA", "CALIFORNIA", "CONNECTICUT", "FLORIDA", "GEORGIA", "IDAHO", "ILLINOIS", "INDIANA", "IOWA", "KENTUCKY", "LOUISIANA", "MAINE", "MARYLAND", "MASSACHUSETTS", "MICHIGAN", "MINNESOTA", "MONTANA", "NEBRASKA", "NEW HAMPSHIRE", "NEW JERSEY", "NEW YORK", "NORTH CAROLINA", "OHIO", "OKLAHOMA", "OREGON", "PENNSYLVANIA", "RHODE ISLAND", "SOUTH CAROLINA", "SOUTH DAKOTA", "TENNESSEE", "VERMONT", "WASHINGTON", "WEST VIRGINIA", "WISCONSIN", "ALABAMA", "ARIZONA", "COLORADO", "KANSAS", "MISSOURI", "NEVADA", "NEW MEXICO", "VIRGINIA", "ARKANSAS", "NORTH DAKOTA", "TEXAS", "UTAH", "OTHER STATES")) with two states considered as "other states".

Years

The data was from the years c(2021, 2019, 2016, 2022, 2020, 2018, 2017).

Assumptions & Motivations

Census Data

The `census` data was a nation-wide collection of data about the fresh and process markets related to strawberries. This data has values that are indicated as `(D)`, which are data that was withheld upon request by the strawberry market in that particular state. This could leave out important information in the data.

Survey Data

The `survey` data was collected via a survey sent out to each state in the United States. There were only 11 out of 47 states who returned the survey (c("CALIFORNIA", "FLORIDA", "OTHER STATES", "NEW YORK", "NORTH CAROLINA", "OREGON", "WASHINGTON", "MICHIGAN", "OHIO", "PENNSYLVANIA", "WISCONSIN")), which includes the "other states". (The "other states" did not have any data relating to pesticides and bacterium.) This is only a 23% response rate, which is not comprehensive of all the states and the entire United States process market. The states that did return the survey will still be able to show a report of pesticide and bacterium usage on their processed strawberry crops.

Data Cleaning & Organizing

R Packages

The following R packages were used to clean and organize the data:

```
library(knitr)
library(kableExtra)
library(tidyverse)
library(stringr)
library(dplyr)
```

The data was organized into two main categories: `census` and `survey`. The `census` data was cleaned and organized to show fresh and process market sales, and the `survey` data consisted of pesticide and fertilizer usage.

Initial Data Cleaning & Organization

The following initial data cleaning derived from Professor Wright.

Removed columns with a single value in all columns using a function. This function produced all columns with one value from the `strawberry` data frame. From there, all but the single value columns were added back to the data frame.

Next, Professor Wright checked if every row in the `state` column held a value, which it did.

After that, Professor Wright checked to see what state had the most rows.

The state with the most rows is CALIFORNIA.

Separating Data Frames

The following separation of data frames derived from Professor Wright.

By using `dplyr`, the `strawberry` data frame was split by `Program`. The rows that contained `CENSUS` were moved to `strwb_census` , and the rows containing `SURVEY` were moved to `strwb_survey`.

Census

After splitting `CENSUS` and `SURVEY` rows into two data frames, Professor Wright has first organized the `CENSUS` data.

First, he separated composite columns and cleaned the Value column.

The composite columns in the `strwb_census` are as follows: `Data Item` and `Domain category`.

The column separators in `CENSUS` are ",", "-", ":".

Separated `Data Item` into the new columns `Fruit`, `temp1`, `temp2`, and `temp3` by ",".

Separated `temp1` into the new columns `crop_type` and `prop_acct` by "-".

To finish this first section of cleaning and organizing the `strwb_census` data frame, Professor Wright string trimmed both sides of the following three columns: `crop_type`, `temp2`, and `temp3`.

Now, Professor Haviland created a "Fresh Market" column. To do this, he duplicated the `temp2` column with the new column name as `Fresh Market`. Next, he removed all the cells in the new column that began with `MEASURED`. Same for the cells that begin with `PROCESSING`. He substituted `NA` values for empty strings. Finally, the `FRESH MARKET` parts of the strings were removed, as they were unneccesary anymore with the new column creation of `Fresh Market`.

Now, to manipulate the `temp2` column, Professor Wright removed all the cells that began with `FRESH`, which would now set up the cleaning and organizing to create the `Process Market` column.

Professor Wright created a "Process Market" column.

To do this, he followed the same method as `Fresh Market`, where he first duplicated `temp2`, then removed cells beginning with `MEASURED`, and removed `PROCESSING` from the beginning of the strings. and remove

Finally, removed the cells starting with `PROCESS MARKET` from `temp2`.

Next, Professor Wright removed NA's from `prop_acct`, `temp2`, and `temp3` by substituting a space.

From here, he combined `temp2` with `temp3` to create a `Metric` column. He also removed parts of string that did not matter, beginning with `MEASURED IN`. To finish, he moved the `Metric` column to the end.

The Value column was transformed. To do this, Professor Wright first pulled the `Value` column from `strwb_census` and put them into `vals`. From there, he string replaced all `vals` to remove the commas in the strings. From there, he converted the strings to numeric values. During this, `NA` values were automatically implemented.

After this, he aimed to find the location and value of the footnotes in the `Value` column. He implemented a string detection on this column, which helped discover that all the cells with `NA` values were also where the footnotes were located. The footnotes indicated that the `Value` was instructed to be left out during the data collection.

I finished cleaning and organizing the `strwb_census` data frame, which is detailed below.

First, I selected particular columns that had necessary data.

Next, I removed the "," from the `Value` column and transformed them into numeric values. This introduced rows with `NA` values.

After that, I cleaned up the `CV (%)` column by changing the values to numbers, instead of strings. This also introduced rows with `NA` values.

Furthermore, I am going to omit all rows in the `Value` and `CV (%)` columns with `NA` values. These are being omitted because they hold no meaning. Only the `Value` column had to be adjusted, as the `CV (%)` column did not have any values if the `Value` column also did not.

Finally, to complete the `strwb_census` cleaning and organizing, I am going to arrange the `State` column to be in ascending order.

Survey

Professor Wright had organized the SURVEY data frame splitting the marketing, and production data from the chemical application data. In the strawberry data frame, The `CENSUS` rows contain marketing, sales, and production data. The `SURVEY` rows contain rows which may be redundant with the CENSUS rows and chemical application rows. These rows contain fresh and process market sales data, which have been removed.

Began cleaning and organizing `strwb_survey` by discovering what columns in this data frame that need to be split.

First, Professor Wright separated the `Data Item` column to `temp1`, `temp2`, `temp3`, and `temp4` by ",". Additionally, he separated `temp1` into `temp1a` and `temp1b`.

Next, he separated the `Domain` column into `temp22` and `temp23` by ",".

Also, he separated `Domain Category` into `temp42` and `temp43` by ",".

To finish the `strwb_survey` cleaning and organizing, this data frame was split into two new data frames, where `strwb_survey_chem` holds pesticide data and `strwb_survey_mkt` contains all the surveyed market and fertilizer usage data.

Survey: Market

Now, I further cleaned up both strwb_survey_mkt and strwb_survey_chem. First, I worked with the strwb_survey_mkt.

Quickly, to begin, Professor Wright Dropped one-value columns in strwb_survey_mkt.

To begin, I made the Value column into numeric values, which introduced NA values. Then, I changed the format of the numeric values, so they did not appear in scientific notation.

To reduce the duplicity of strwb_survey_mkt with strwb_census, I separated the fresh and process market data in strwb_survey_mkt from the data on the fertilizer usage.

Next, I separated the temp42 column to create the columns type and fertilizer_type by the delimiter ":".

I cleaned up the fertilizer_type strings by removing the parentheses on both sides.

I also removed the empty spaces in these strings.

Finally, I selected the rows that are relevant and renamed temp3 to measurement and temp4 to avg.

Survey: Chemical

Finally, I finished cleaning and organizing strwb_survey_chem.

First, Professor Haviland dropped one-value columns in strwb_survey_chem.

Then, I selected the relevant columns.

From here, I split up temp43 into chemical and PC# by "=".

Furthermore, I cleaned up both of the new chemical and the PC# columns by removing the unnecessary parentheses.

I also want to separate the chemical column by chemical type and chemical name.

Trimmed off the spaces on both sides of the chemical column.

Trimmed the same to chem_type.

Finally, trimmed the same with PC#.

Now to remove the rows with NA values in the PC# column.

Addition to Strwb_Survey_Chem: WHO Chemical Toxicity

Now, to better understand the `strwb_survey_chem` data, I added two columns of data. We are first going to begin with "chemical toxicity" (`toxicity`).

To gather this information, I used the World Health Organization (WHO)'s classification of pesticides by hazard [WHO, 2019]. Table 1 shows the WHO's toxicity classification for pesticides.

Class	Description	LD_{50} for the Rat (mg/kg Body Weight)			
		Oral		Dermal	
		Solids	Liquids	Solids	Liquids
Ia	Extremely hazardous	≤5	≤20	≤10	≤40
Ib	Highly hazardous	5–50	20–200	10–100	40–400
II	Moderately hazardous	50–500	200–2,000	100–1,000	400–4,000
III	Slightly hazardous	>500	>2,000	>1,000	>4,000

doi:10.1371/journal.pmed.1000357.t001

Figure 3: Table 1. World Health Organization Pesticide Toxicity Classification

Additionally, some chemicals are presented as fatal or toxic if inhaled, as they are gaseous or volatile fumigants [WHO, 2019]. Others are classified as "unlikely to present acute hazard" by WHO [WHO, 2019], which means that they will not present any hazard if used properly. Furthermore, they are also classified as "no significant acute toxicity" when they are not in the WHO classification and are found to be non-toxic, mostly discovered through the Environmental Protection Agency (EPA) [EPA].

I classified each chemical toxicity by the string `Highly hazardous`, `Moderately hazardous`, `Unlikely to present acute hazard`, `Slightly hazardous`, `Fatal if inhaled`, `Toxic if inhaled`, `No significant acute toxicity` and `Not specified`, based on the WHO chemical toxicity rating system [WHO, 2019].

I searched through each classification table to find each pesticide. Not all pesticides were on the table. To find the missing chemical toxicities, I used the large language model, Chat GPT [Chat GPT], found information through EPA Pesticide Fact Sheets [EPA-Acibenzolar], [EPA-Ammonium.], [EPA-Aureobasidium.], [EPA-Canola], [EPA-Caprylic], [EPA-Capasaicin], [EPA-Clethodim], [EPA-Cyfluefenamid], [EPA-Cytokinin], [EPA-Indole.], [EPA-Iron-Phos.], [EPA-Mefenoxam], [EPA-Metam-sodium], [EPA-Polyoxin], [EPA-Potassium], [EPA-Potassium-salts], [EPA-Potassium-silicate], [EPA-Spiromedifen], [EPA-STREPTOMYCES-LYDICUS], [EPA-SULFENTRAZONE], [EPA-SULFUR], and in other sources (Kilani-Morakchi, Morakchi-Goudjil, and Sifi 2021), [PubChem-Capric], [Carfentrazone-ethyl], [Pub-Chem-Copper.], [Copper-Octanoate], [ACS-Copper-Oxide], [Cyprodinil], [DECYLDIMETHYLOCTYL], [Didecyl.], [Dodine], [Sodium-Ferric-Ethyl.], [Garlic-Oil], [Glyphosate], [Hydrogen-Peroxide], [Isofetamind], [Methoxyfenozide], [Mineral-oil], [Mono-Potassium-Salt], [Mustard-oil], [Peroxyacetic-acid], [Quinoline], [CDC-PYRACLOSTROBIN], [SOYBEAN-OIL].

Furthermore, through this search, I discovered through Chat GPT [Chat GPT], that many of the "chemicals" were actually bacterium. I found the toxicity of the bacterium via web search from various sources [BACILLUS AMYLOLIQUEFAC F727.], [BACILLUS AMYLOLIQUEFACIENS MBI 600], [BACILLUS AMYLOLIQUEFACIENS STRAIN D747], [BACILLUS PUMILUS], [BACILLUS SUBTILIS], [ALKYL. DIM. BENZ. AM], [AUREOBASIDIUM PULLULANS DSM 14940], [BACILLUS SUBT. GB03], [BEAUVERIA BASSIANA], [BLAD], [BT KURSTAK ABTS-1857], [BT], [BURKHOLDERIA A396 CELLS & MEDIA], [CAPSICUM OLEORESIN EXTRACT], [CHROMOBAC SUBTSUGAE PRAA4-1 CELLS AND SPENT MEDIA], [GLIOCLADIUM VIRENS], [HARPIN A B PROTEIN], [HELICOVERPA ZEA NPV], [NEEM OIL], [PAECILOMYCES FUMOSOR], [PETROLEUM DISTILLATE], [PSEUDOMONAS CHLORORAPHIS STRAIN AFS009], [REYNOUTRIA SACHALINE], [TRICHODERMA HARZ.], [TRICHODERMA VIRENS STRAIN G-41].

There were three `Unknown` values for `COPPER ETHANOLAMINE`, `KANTOR`, and `HALOSULFURON-METHYL`, which did not have any clear information on toxicity.

Addition to Strwb_Survey_Chem: CAS Registry Number

Now, the second added column represents each pesticide's Chemical Abstract Service (CAS) Registry Number (`CAS#`). A CAS Registry Number allows each chemical compound, including molecular formulas, chemical structures, generic, systematic, common, and trade names, to have a clear identification number [CAS].

To find the `CAS#`s of each pesticide, I used [WHO, 2019], [CHAT GPT], and, mainly, the United States Environmental Protection Agency's Pesticide Chemical Search [EPA-search]. After discovering that some of the rows in `chemical` are actually bacterium, all values that do not have `CAS#`s are indicated as `Bacteria`. All the other chemicals were matched with their appropriate `CAS#`.

Quickly, I reorganized `strwb_survey_chem` to have the columns in a different order.

I also changed the column name from `chemical` to `strwb_treatment`, as bacterium are not chemicals.

Finally, to finish the data cleaning and organization of `strwb_survey_chem`, I changed the name of the data frame to `strwb_survey_treat` (treat = treatment) to avoid confusion, since it is known that all strawberry pesticide treatments are not only chemicals but also bacterium.

Exploratory Data Analysis

Now, to

References

Afrin, Sadia, Massimiliano Gasparrini, Tamara Y. Forbes-Hernandez, Patricia Reboredo-Rodriguez, Bruno Mezzetti, Alfonso Varela-López, Francesca Giampieri, and Maurizio Battino. 2016. "Promising Health Benefits of the Strawberry: A Focus on Clinical Studies." *Journal of Agricultural and Food Chemistry* 64 (22): 4435–49. https://doi.org/10.1021/acs.jafc.6b00857.

CRAIG, WINSTON J. 1997. "Phytochemicals." *Journal of the American Dietetic Association* 97 (10): S199–204. https://doi.org/10.1016/s0002-8223(97)00765-7.

Fenik, Jolanta, Maciej Tankiewicz, and Marek Biziuk. 2011. "Properties and Determination of Pesticides in Fruits and Vegetables." *TrAC Trends in Analytical Chemistry* 30 (6): 814–26. https://doi.org/10.1016/j.trac.2011.02.008.

Kilani-Morakchi, Samira, Houda Morakchi-Goudjil, and Karima Sifi. 2021. "Azadirachtin-Based Insecticide: Overview, Risk Assessments, and Future Directions." *Frontiers in Agronomy* 3 (July). https://doi.org/10.3389/fagro.2021.676208.

Verteramo Chiu, Leslie J., and Miguel I. Gomez. 2023. "A Tale of Two Strawberries: Conventional and Organic Open-Field Production in California." *Sustainability* 15 (19): 14363. https://doi.org/10.3390/su151914363.