

Herbicide Usage in Strawberry Production

AUTHOR

Shuxin Qian, Rebecca Vithayathil,
Bihong Wang

Introduction

This project aims to examine the use of herbicides across different U.S. states and their regional distribution patterns. The goal is to identify regional patterns and differences in herbicide application practices, and to explore the relationship between farming type and agricultural land use.

Using data collected from sources such as the USDA, we identified several directions to explore correlations. In terms of regional variation, the use of herbicides differs in application methods and intensity across different areas. These comparisons allow us to investigate underlying factors such as environmental conditions and the degree of impact on strawberry cultivation.

In addition, there are distinct effects between organic and conventional pesticide options. By combining information on land use and farm types, we can further analyze the potential for reducing dependence on herbicides.

Through a systematic analysis of these dimensions, this project not only enhances our understanding of herbicide usage in U.S. agricultural production but also provides valuable insights for future environmental regulation, pesticide management, and sustainable farming practices.



Data Overview

This dataset records herbicide usage in strawberry cultivation across different U.S. states and years. Each row represents the application of a specific herbicide in a given state and year. The data include the total amount of herbicide applied and the proportion of treated area. By matching herbicide IDs, we also obtained the corresponding toxicity indicator data.

The toxicity indicators cover various potential hazard categories, such as environmental toxicity, health hazards, and irritancy. These variables allow for an integrated analysis of both the intensity of chemical use and the potential risk levels associated with strawberry cultivation.

Raw data

Program	Year	Period	Week.Ending	Geo.Level	State	State.ANSI	Ag.Dis
CENSUS	2022	YEAR	NA	STATE	ALABAMA	1	NA
CENSUS	2022	YEAR	NA	STATE	ALABAMA	1	NA
CENSUS	2022	YEAR	NA	STATE	ALABAMA	1	NA
CENSUS	2022	YEAR	NA	STATE	ALABAMA	1	NA
CENSUS	2022	YEAR	NA	STATE	ALABAMA	1	NA

Table 1 presents the raw dataset directly obtained from the USDA database. As shown, the data contain over 2,000 observations, but a large number of missing values (NA) exist across several columns. In addition, certain variables are irrelevant to our analysis and need to be removed. Therefore, data cleaning and integration are necessary before conducting

any further statistical modeling or visualization.

We first organized the data set downloaded directly from the USDA website by removing columns that contained only a single unique value. This step eliminates meaningless constants and keeps only variables that provide useful distinctions. Next, we cleaned and standardized several data names, and created new variables to facilitate later comparisons. We then grouped the large and complex dataset so that for each state, each year, and each type of herbicide, only one representative value was retained. Finally, we transformed the cleaned data into a structured table format, making it easier to review and to use in subsequent coding and analysis.

Data Cleaning

Table 2. Final data

Year	State	Domain.Category	APPLICATIONS, MEASURED IN		Acute		Environ
			LB		Toxic	Corrosive	
1990	FLORIDA	GLYPHOSATE = 417300	300		0	1	
1990	FLORIDA	NAPROPAMIDE = 103001	3,100		0	0	
1990	FLORIDA	PARAQUAT = 61601	4,100		1	1	
1990	MICHIGAN	2,4-D = 30001	500		0	1	
1990	MICHIGAN	CHLOROXURON = 25501	400		0	0	

We first organized the data set downloaded directly from the website by removing columns that contained only a single unique value. This step eliminates meaningless constants and keeps only variables that provide useful distinctions.

Next, we cleaned and standardized several data names, and created new variables to facilitate later comparisons. We then grouped the large and complex dataset so that for each state, each year, and each type of herbicide, only one representative value was retained. We transformed the cleaned data into a structured table format, making it easier to review and to use in subsequent coding and analysis.

Due to the large number of missing values in the USDA dataset, we

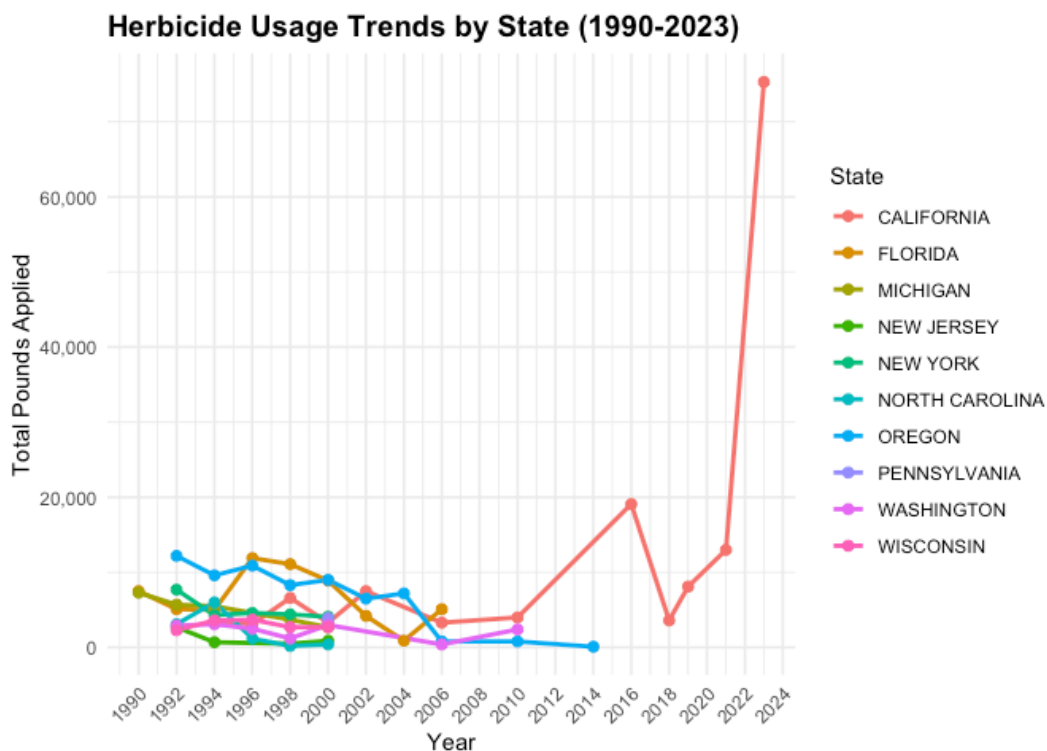
determined that additional data were needed. We found that PubChem, an open chemical information database, provides detailed compound-level records, including toxicity classifications.

First we attempted to scrape the PubChem website search function so that we could import toxicity information that way. After we weren't able to achieve this we decided to search the website ourselves and copy the information into a shared sheet since we only had 38 chemicals to search. This was a more accurate approach as well as there were inconsistencies between naming and IDs in the USDA website and the PubChem website.

After completing the sheet we saw that there were 6 unique Toxicity categories for our herbicides. We then reformatted the toxicity data so that there were 6 binary columns (one for each toxicity category). You can check our Data branch to get code and detail information as well as our .

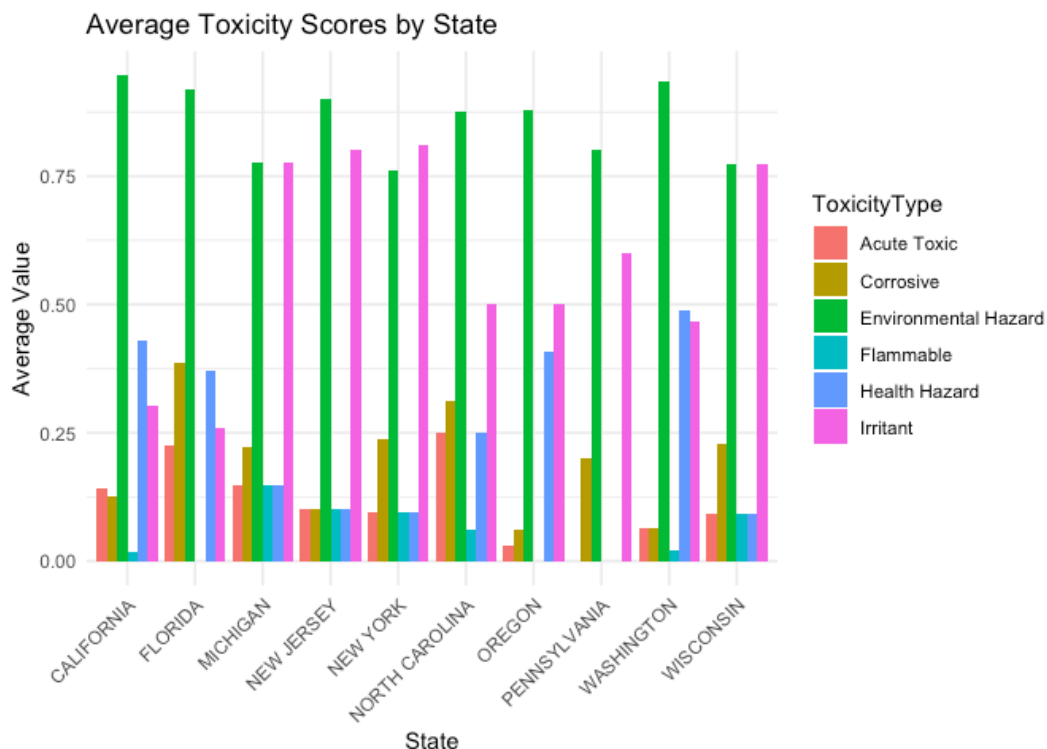
Exploratory Data Analysis

1) Herbicide Coverage Over Time



Our variables include year, state, and herbicide coverage rate. Each color in the figure represents a different state, and the upward or downward slope of the lines indicates the change in herbicide usage in pounds over time. Oregon shows a clear downward trend, suggesting a decrease in herbicide usage over time. California shows clearly upward trend in usage.

2) Average Toxicity by state

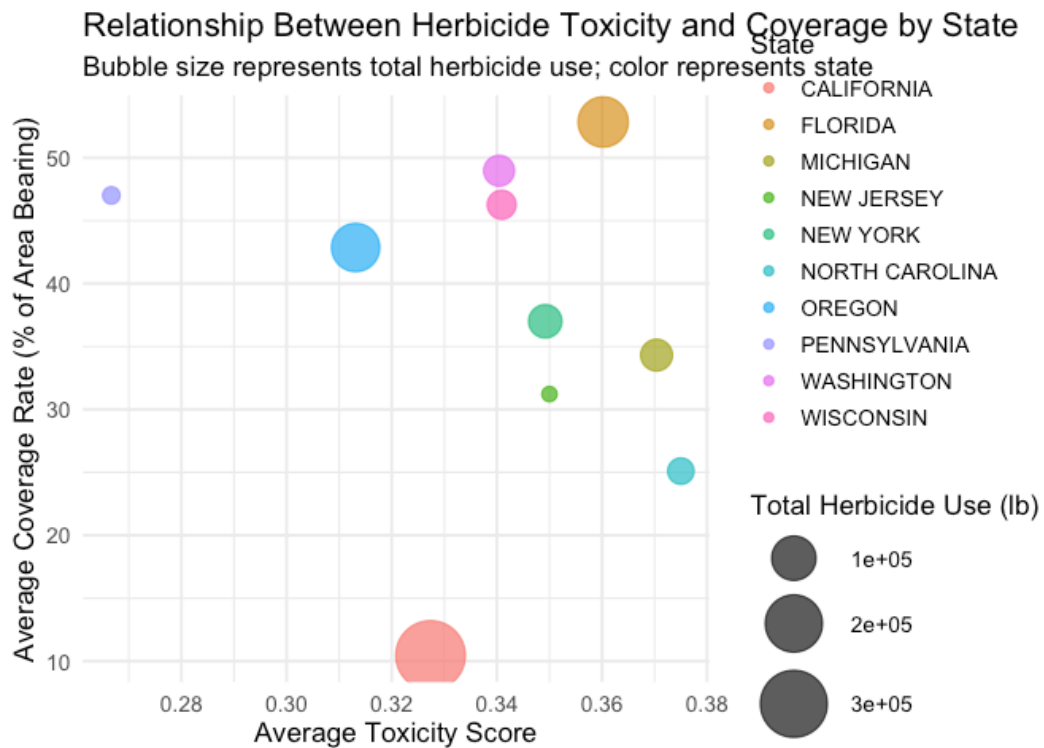


This figure illustrates the average toxicity levels of herbicides across states, based on six toxicity indicators: Acute Toxicity, Corrosive, Environmental Hazard, Flammable, Health Hazard, and Irritant. Each color represents a toxicity category, and the height of each bar indicates the mean score for that category within each state.

Overall, most states exhibit low to moderate levels of toxicity. However, there are marked differences among the six toxicity types: Environmental Hazard and Corrosive scores are higher in several states, suggesting potential ecological and material risks, whereas Flammable and Health Hazard remain consistently low across all regions.

These findings indicate that, although the intensity of herbicide use varies by state, the distribution of toxicity profiles is not uniform, reflecting differences in herbicide selection and agricultural practices across regions.

3) Average herbicide toxicity vs Coverage Rate



This bubble chart illustrates the relationship between average herbicide toxicity and coverage rate across states. Each bubble represents a state:

The x-axis shows the average toxicity score;

The y-axis shows the average coverage rate (% of area bearing);

The bubble size indicates the total amount of herbicide used;

The color distinguishes different states.

Overall, most states cluster around the region of moderate toxicity and moderate coverage, suggesting that herbicide usage is relatively balanced nationwide. However, several clear differences can be observed:

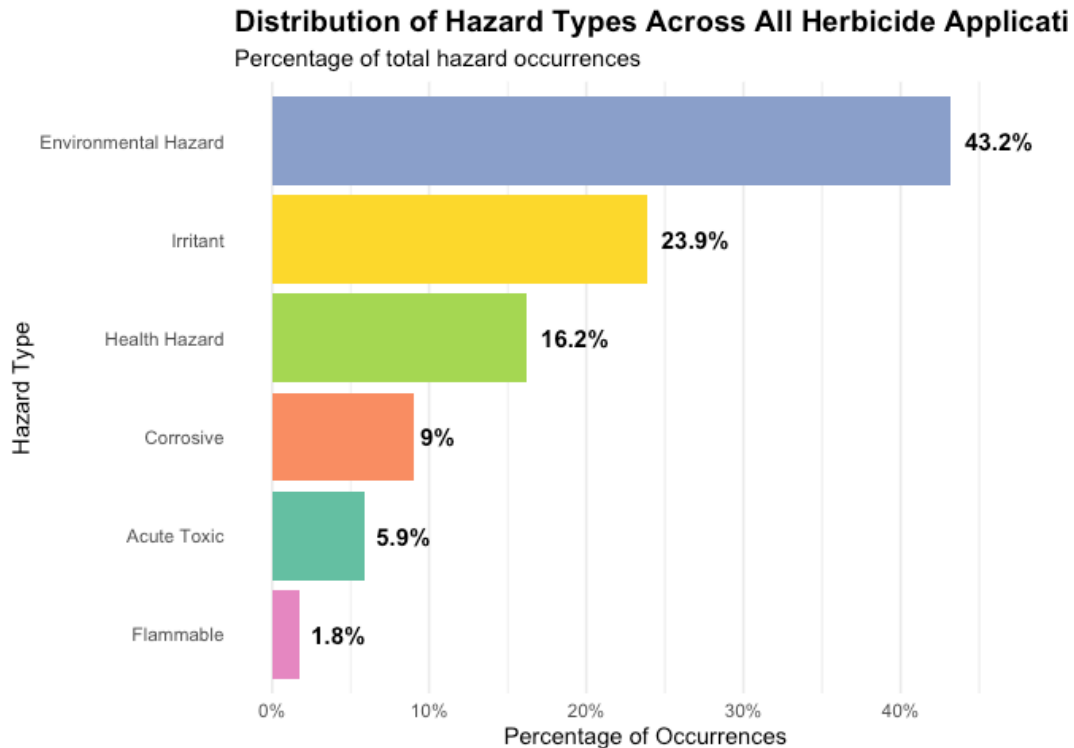
States with high toxicity and high coverage (upper-right quadrant) indicate intensive chemical application and higher potential environmental risks. These regions often have larger strawberry production scales and may rely less on low-toxicity alternatives.

States with low toxicity but high coverage likely use safer chemical compositions, reflecting more advanced pesticide management practices or stronger regulatory oversight.

States with low toxicity and low coverage (lower-left quadrant) typically have smaller strawberry cultivation areas or a higher proportion of organic farming.

The bubble sizes further reveal differences in total herbicide usage across states: even when toxicity levels are similar, the total amount of herbicide applied varies substantially among states.

4) Toxicity



The bar chart illustrates the distribution of different hazard types associated with all herbicides, expressed as the percentage of total hazard occurrences. It is evident that Environmental Hazard is the most prevalent type, accounting for 43.2% of all occurrences. This indicates that herbicide use poses significant risks to ecosystems, potentially contaminating soil, water, and affecting non-target species.

The second most common category is Irritant, representing 23.9%, suggesting frequent risks of skin or eye irritation for individuals handling these substances. Health Hazards (16.2%) also represent a considerable portion, implying possible long-term or systemic health effects from exposure.

Summary

This project integrated herbicide application data from the USDA and incorporated toxicity information from PubChem to systematically analyze the patterns of herbicide use, and potential risks in U.S. strawberry production. The results show that although the overall coverage rate

remains moderate, there are clear differences across states in both usage intensity and toxicity composition. For example, California demonstrates a rising trend in herbicide coverage and higher toxicity levels, while Florida shows a decreasing trend and relatively lower potential risks. Despite variations in state-level practices, herbicides continue to exert the greatest environmental impact among agricultural chemicals.

Beyond data analysis, this project also provided valuable experience in collaborative work using GitHub, data collection and cleaning, and creative data generation when facing incomplete datasets. It further strengthened skills in visualization, interpretation, and reproducible research, all of which are essential for effective communication in data-driven environmental studies.

Citation

<https://quickstats.nass.usda.gov/> - Farming and usage information by state and year

<https://pubchem.ncbi.nlm.nih.gov/#query=108800> - Toxicity information by herbicide

<https://www.epa.gov/newsreleases/epa-issues-emergency-order-stop-use-pesticide-dacthal-address-serious-health-risk-4> - Herbicide DPCA was no longer in the PubChem database likely because it has been banned. Use this source for toxicity information.