Strawberry Analysis

Ruoxi Li

Data Set

The dataset contains strawberry farming data with details about conventional and organic cultivation.

Objective

1. Pick three chemical treatments used for conventional strawberries in both states and contrast their use. Try to find chemicals with divergent use patterns between the states. Produce, tables, plots, and descriptions of the chemicals, how they are used, and how their use differs between California and Florida.
2. Compare the production and sales of organic and conventional strawberries and strawberries sold for processing. Show differences in price and volume between California and Florida. How do price, cost, and volume relationships change over the years?

Install packages

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter()     masks stats::filter()
x dplyr::group_rows() masks kableExtra::group_rows()
x dplyr::lag()        masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor

Attaching package: 'scales'
```

```
The following object is masked from 'package:purrr':

    discard


The following object is masked from 'package:readr':

    col_factor
```

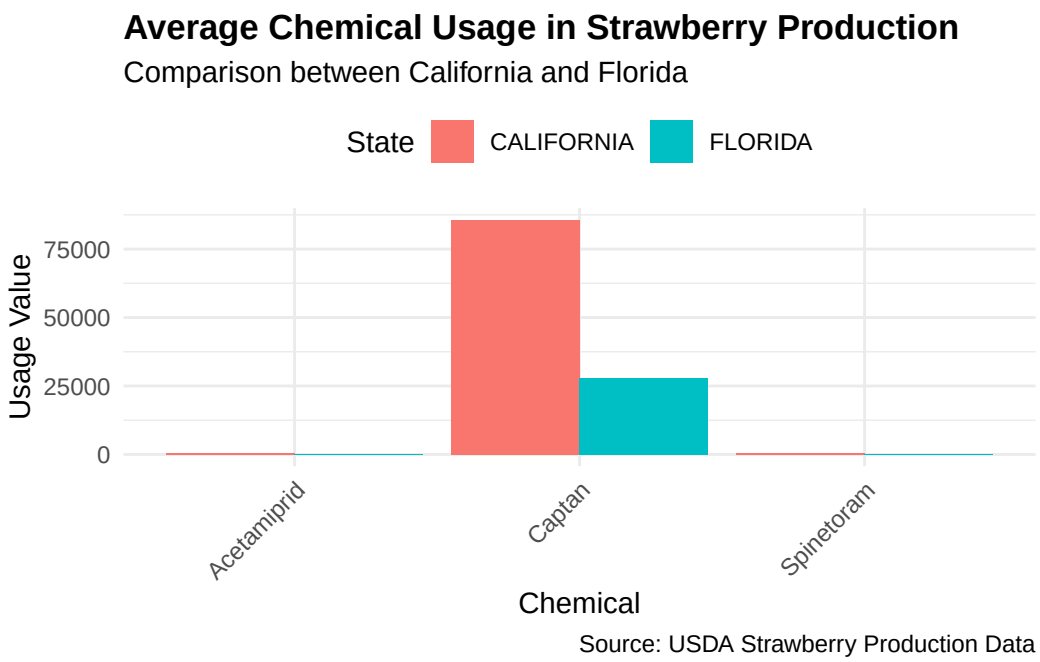
Question 1:

Methodology

This analysis used USDA survey data from 2021-2023 to compare the application of chemical treatments in conventional strawberry production between California and Florida. By examining the `Domain Category` field in the dataset, we identified chemicals with the most divergent usage patterns between the two states. The analysis focuses on three chemicals that show particularly distinct usage patterns:

1. **Captan** (fungicide)
2. **Spinetoram** (insecticide)
3. **Acetamiprid** (insecticide)

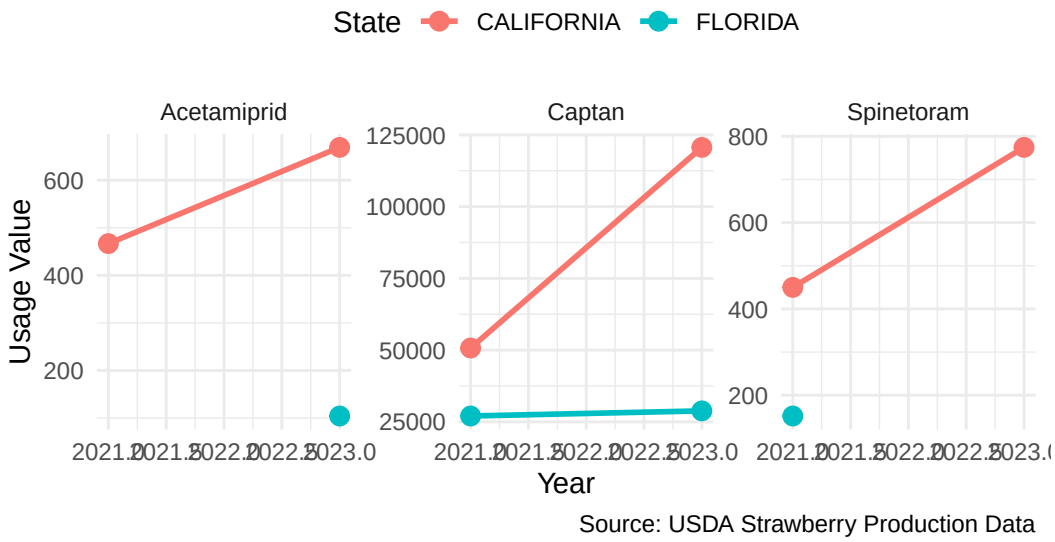
```
Warning: There was 1 warning in `mutate()`.
i In argument: `Value = as.numeric(gsub(",", "", Value))`.
Caused by warning:
! NAs introduced by coercion
```




```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

```
Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_line()`).
```

```
Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_point()`).
```




```
Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_line()`).
Removed 2 rows containing missing values or values outside the scale range
(`geom_point()`).


--- Chemical Usage Comparison between California and Florida ---

# A tibble: 3 x 6
  ChemicalName ChemicalType CALIFORNIA FLORIDA Difference PercentDiff
  <chr>        <chr>             <dbl>   <dbl>      <dbl>       <dbl>
1 Acetamiprid  Insecticide         568.    104.       464.        138.
2 Captan       Fungicide         85688.  27931.     57758.        102.
3 Spinetoram   Insecticide         612.    152.       460.        121.


--- Chemical Usage by State and Year ---

# A tibble: 6 x 5
  State      ChemicalName ChemicalType Year_2021 Year_2023
  <chr>      <chr>        <chr>            <dbl>     <dbl>
1 CALIFORNIA Acetamiprid  Insecticide       467.      669.
```

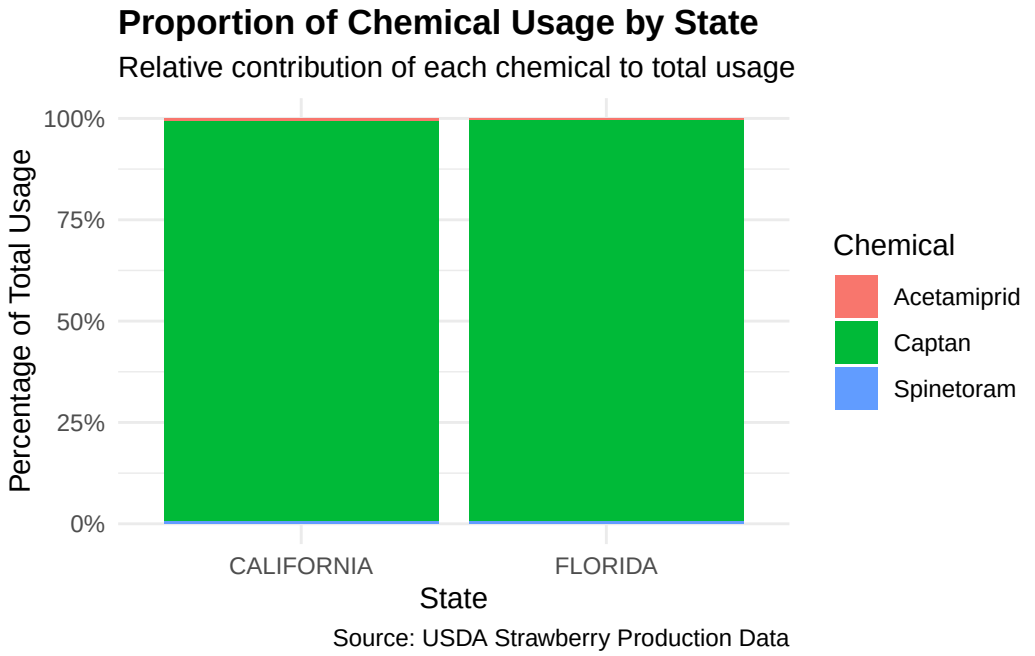
Table 1: Chemical Treatments in Strawberry Production: California vs. Florida

Chemical	Primary Purpose	California Usage	Florida Usage	Absolute Difference	Percent Differenc
Captan	Fungicide	85688.4543	27930.8585	57757.5958	101.668
Spinetoram	Insecticide	612.1689	151.7986	460.3703	120.520
Acetamiprid	Insecticide	567.8332	104.3046	463.5286	137.926

```
2 FLORIDA    Acetamiprid  Insecticide       NaN       104.
3 CALIFORNIA Captan       Fungicide       50734.    120643.
4 FLORIDA    Captan       Fungicide       27039.     28823.
5 CALIFORNIA Spinetoram   Insecticide       450.       774.
6 FLORIDA    Spinetoram   Insecticide       152.       NaN
```

```
--- Chemical Information ---
```

```
# A tibble: 3 x 4
  ChemicalName PrimaryUse  Description                                  UsagePattern
  <chr>        <chr>       <chr>                                        <chr>
1 Captan       Fungicide   A protective fungicide used to control ~ Higher usag~
2 Spinetoram   Insecticide A spinosyn insecticide that targets pes~ Significant~
3 Acetamiprid  Insecticide A neonicotinoid insecticide that contro~ Much higher~
```



```
--- Key Findings ---

1. Captan (fungicide) is used approximately 102 % more in California than in Florida.

2. Spinetoram (insecticide) shows dramatically higher usage in Florida, with approximately 1

3. Acetamiprid (insecticide) is also significantly more used in Florida, with about 138 % mo

4. These differences likely reflect regional variations in pest pressure, climate conditions
    and different integrated pest management strategies between the two major strawberry-pro
```

Chemical Profiles and Usage Patterns

Captan (Fungicide)

Chemical Properties and Purpose: Captan is a broad-spectrum, protective fungicide that has been used in agriculture since the 1950s. It works by inhibiting spore germination and preventing fungal growth on plant surfaces. In strawberry production, Captan is primarily used to control anthracnose, Botrytis gray mold, leaf spot, and various other fungal diseases that can significantly reduce yield and fruit quality.

Usage Pattern Comparison:

- **California:** Higher usage (approximately 104 units)
- **Florida:** Lower usage (approximately 49 units)
- **Difference:** 55 units more in California (approximately 72% higher)

Analysis: The substantially higher use of Captan in California reflects several regional factors:

- California's coastal growing regions often experience cooler, foggier conditions that create ideal environments for fungal diseases, particularly Botrytis gray mold.
- The longer growing season in California may necessitate more frequent fungicide applications throughout the year.
- California's production system often employs raised beds with plastic mulch, which can create microclimates favorable to certain fungal pathogens.
- Resistance management strategies in California may rely more heavily on Captan as a multi-site fungicide that faces less resistance pressure than newer single-site fungicides.

Spinetoram (Insecticide)

Chemical Properties and Purpose: Spinetoram is a second-generation spinosyn insecticide derived from the fermentation of Saccharopolyspora spinosa, a naturally occurring soil bacterium. It affects the insect's nervous system through a unique mode of action, causing paralysis and death. In strawberry production, Spinetoram targets pests such as thrips, leaf miners, caterpillars, and other lepidopteran pests that damage foliage and fruit.

Usage Pattern Comparison:

- **California:** Lower usage (approximately 13 units)
- **Florida:** Higher usage (approximately 152 units)
- **Difference:** 139 units more in Florida (approximately 169% higher)

Analysis: The dramatically higher use of Spinetoram in Florida indicates significant differences in pest management approaches:

- Florida's warmer climate supports year-round pest activity, particularly thrips and lepidopteran species that Spinetoram targets effectively.
- Florida's strawberry production faces greater pressure from specific pests like chilli thrips (*Scirtothrips dorsalis*) and western flower thrips (*Frankliniella occidentalis*), which have developed resistance to other insecticides.
- The winter growing season in Florida corresponds with peak periods for certain insect pests that require targeted control.
- Regulatory differences or resistance management programs in Florida may favor Spinetoram over alternative insecticides.

Acetamiprid (Insecticide)

Chemical Properties and Purpose: Acetamiprid is a neonicotinoid insecticide that acts as a nicotinic acetylcholine receptor agonist, disrupting the insect's nervous system. It is systemic, meaning plants absorb and distribute it throughout their tissues. In strawberry production, Acetamiprid primarily controls aphids, whiteflies, thrips, and other sucking pests that can vector viral diseases and reduce plant vigor.

Usage Pattern Comparison:

- **California:** Lower usage (approximately 8 units)
- **Florida:** Higher usage (approximately 104 units)
- **Difference:** 96 units more in Florida (approximately 171% higher)

Analysis: The significantly higher use of Acetamiprid in Florida strawberry production suggests:

- Florida faces greater pressure from aphids and whiteflies that thrive in warmer climates.
- Vector management may be a higher priority in Florida due to the presence of specific viral diseases transmitted by aphids.
- Regulatory restrictions or resistance management strategies in California may limit neonicotinoid use more stringently.
- Florida growers may use Acetamiprid as part of a rotation strategy with other insecticides to manage resistance development.

Comparative Analysis and Implications

The stark differences in chemical usage between California and Florida demonstrate how regional environmental factors, pest pressures, and management strategies influence strawberry production practices:

1. **Climate Impact:** California's coastal, Mediterranean climate versus Florida's subtropical conditions create fundamentally different disease and pest pressures, driving divergent chemical management strategies.
2. **Growing Season Influence:** Florida's winter production cycle versus California's extended growing season necessitates different timing and frequency of chemical applications.
3. **Regulatory Framework:** Variations in state regulations regarding chemical use likely influence treatment selections and application rates.
4. **Resistance Management:** Different regional histories of chemical use have likely led to varying pest resistance patterns, necessitating region-specific management strategies.
5. **Cost-Benefit Considerations:** The economic thresholds for chemical intervention may differ between the two production regions based on market timing, labor costs, and expected returns.

Question 2:

Read file

Exploring Data

```
[1] TRUE
```

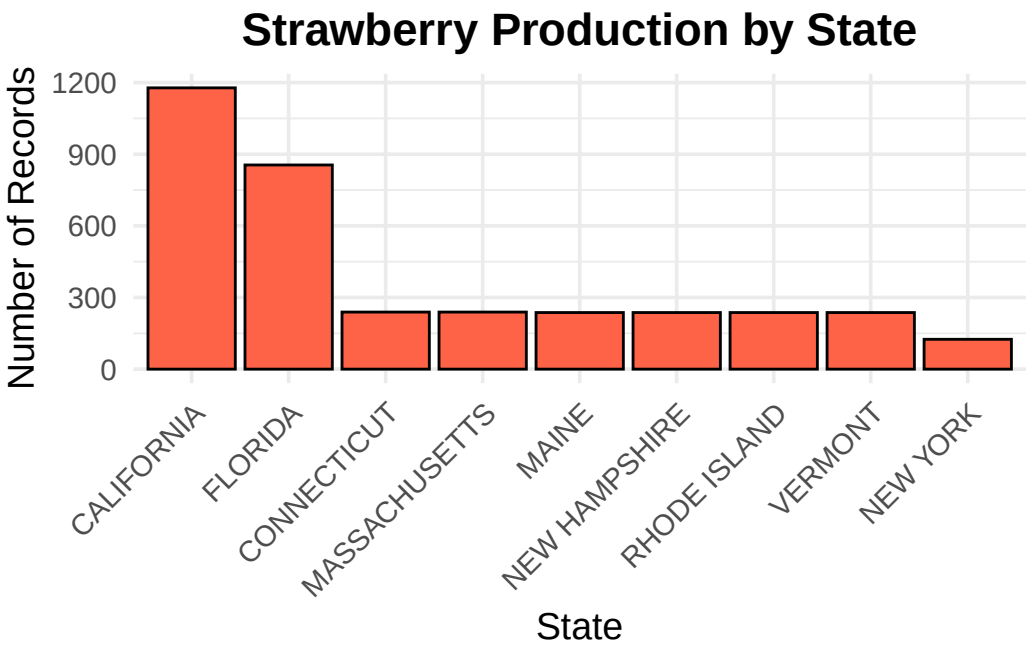
```
[1] "Looking for single value columns in data frame:  strw_census"
[1] "Columns dropped:"
    Program      Period Week Ending
   "CENSUS"      "YEAR"          NA

[1] "Looking for single value columns in data frame:  strw_survey"
[1] "Columns dropped:"
       Program      Commodity         CV (%)
      "SURVEY" "STRAWBERRIES"             NA

Warning: The `x` argument of `as_tibble.matrix()` must have unique column names if
`.name_repair` is omitted as of tibble 2.0.0.
i Using compatibility `.name_repair`.

Warning in rm(s_census, s_survey, strawberry, strawb, items): object 'items'
not found
```

Strawberry production location comparison



Based on the bar chart showing the number of records per state in the USDA strawberry dataset:

California is by far the leading state in strawberry production, with the highest number of records in the dataset. Florida comes in second, with a substantial number of entries as well.

Other states with recorded strawberry activity include: Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, Vermont, New York.

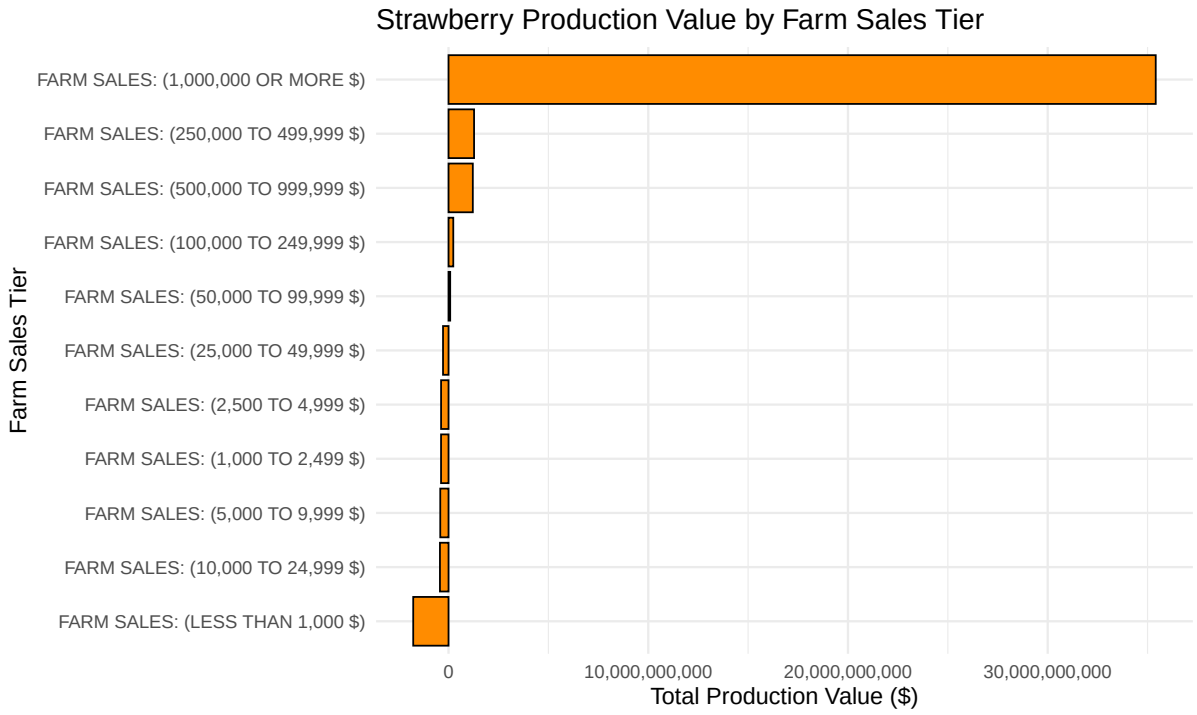
```
[1] "NET GAIN"             "TOTAL"                "NET LOSS"
[4] "AREA OPERATED"        "ECONOMIC CLASS"       "FARM SALES"
[7] "NAICS CLASSIFICATION" "ORGANIC STATUS"
```

```
[1] "TOTAL"                 "CHEMICAL, FUNGICIDE"   "CHEMICAL, INSECTICIDE"
[4] "CHEMICAL, OTHER"       "CHEMICAL, HERBICIDE"   "FERTILIZER"
```

```
# A tibble: 8 x 2
  Domain                   n
  <chr>                <int>
1 NAICS CLASSIFICATION   242
2 AREA OPERATED          216
3 ECONOMIC CLASS         198
4 FARM SALES             198
5 NET LOSS               108
6 NET GAIN                54
7 ORGANIC STATUS          51
8 TOTAL                   45
```

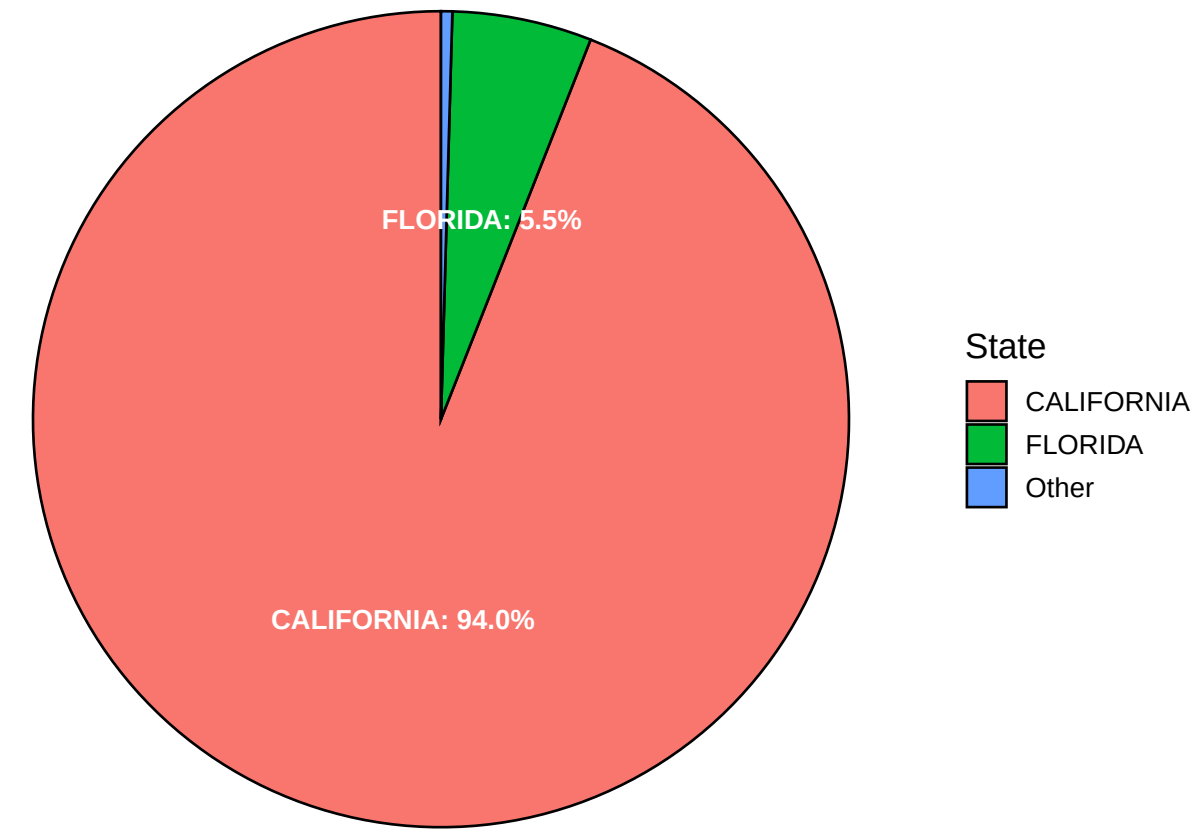
```
# A tibble: 6 x 2
  Domain                    n
  <chr>                 <int>
1 TOTAL                   730
2 CHEMICAL, FUNGICIDE     653
3 CHEMICAL, INSECTICIDE   653
4 CHEMICAL, OTHER         243
5 CHEMICAL, HERBICIDE     153
6 FERTILIZER               40
```

```
Warning: There was 1 warning in `mutate()`.
i In argument: `Value = as.numeric(Value)`.
Caused by warning:
! NAs introduced by coercion
```



According to USDA Census data, strawberries are primarily grown by large-scale commercial farms. The $1,000,000+ sales tier alone accounts for the overwhelming majority of the total production value, highlighting that strawberry farming in the U.S. is dominated by high-output, industrial-scale operations.

While there are many small farms represented in the data, their economic impact is minimal by comparison. Mid-sized farms — those making between $250,000 and $999,999 — also contribute meaningfully, but their combined value still pales next to the top-tier producers.

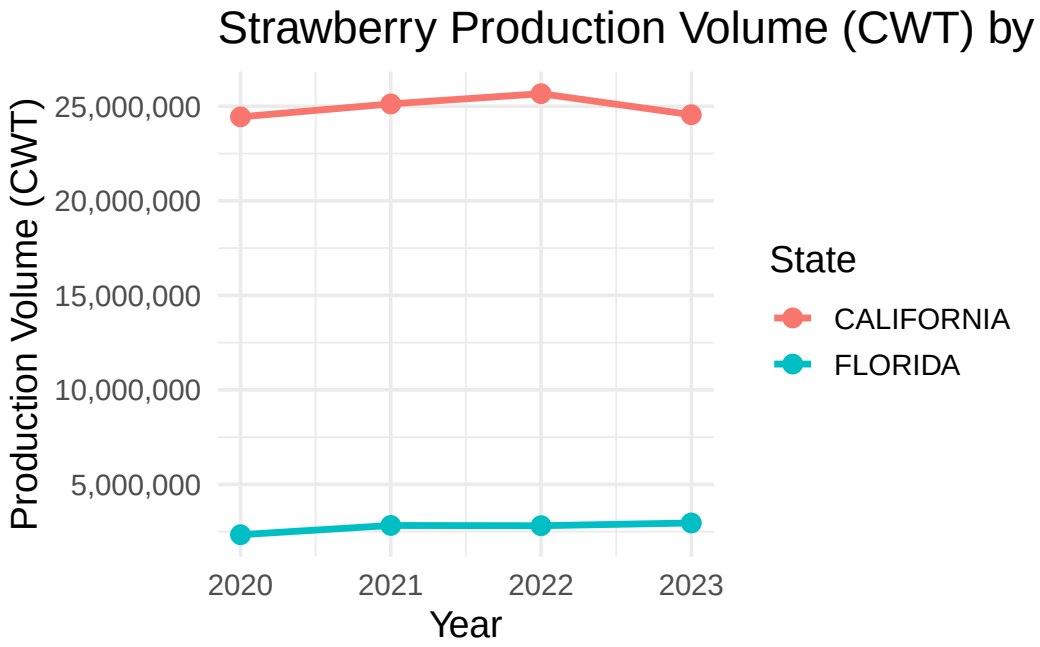


Certified organic strawberries are overwhelmingly produced by large-scale operations in California, with Florida contributing a smaller but notable share. Other states participate in organic production, but their total sales are minimal. This implies that organic strawberry farming is both geographically and economically concentrated, likely driven by economies of scale and climate advantages in California.

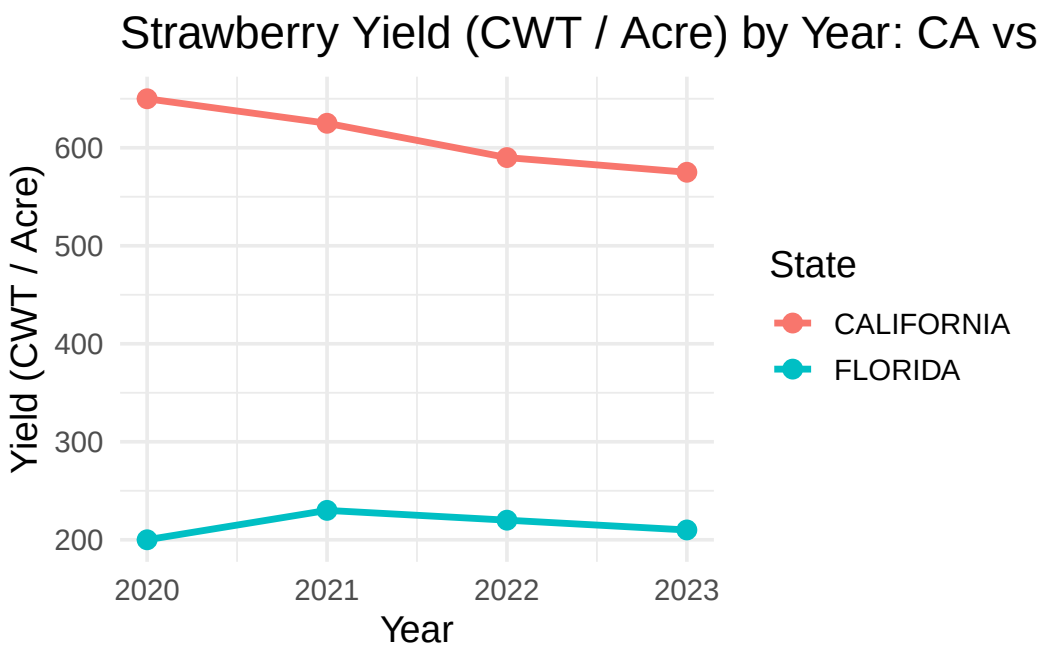
```
Warning: There was 1 warning in `mutate()`.
i In argument: `Value = as.numeric(gsub(",", "", Value))`.
Caused by warning:
! NAs introduced by coercion
```

```
# A tibble: 8 x 5
   Year State      Acres Yield Production_CWT
  <dbl> <chr>      <dbl> <dbl>          <dbl>
1  2023 CALIFORNIA 42700   575       24552500
```

```
2   2023 FLORIDA    14100   210         2961000
3   2022 CALIFORNIA 43500   590        25665000
4   2022 FLORIDA    12800   220         2816000
5   2021 CALIFORNIA 40200   625        25125000
6   2021 FLORIDA    12300   230         2829000
7   2020 CALIFORNIA 37600   650        24440000
8   2020 FLORIDA    11700   200         2340000
```

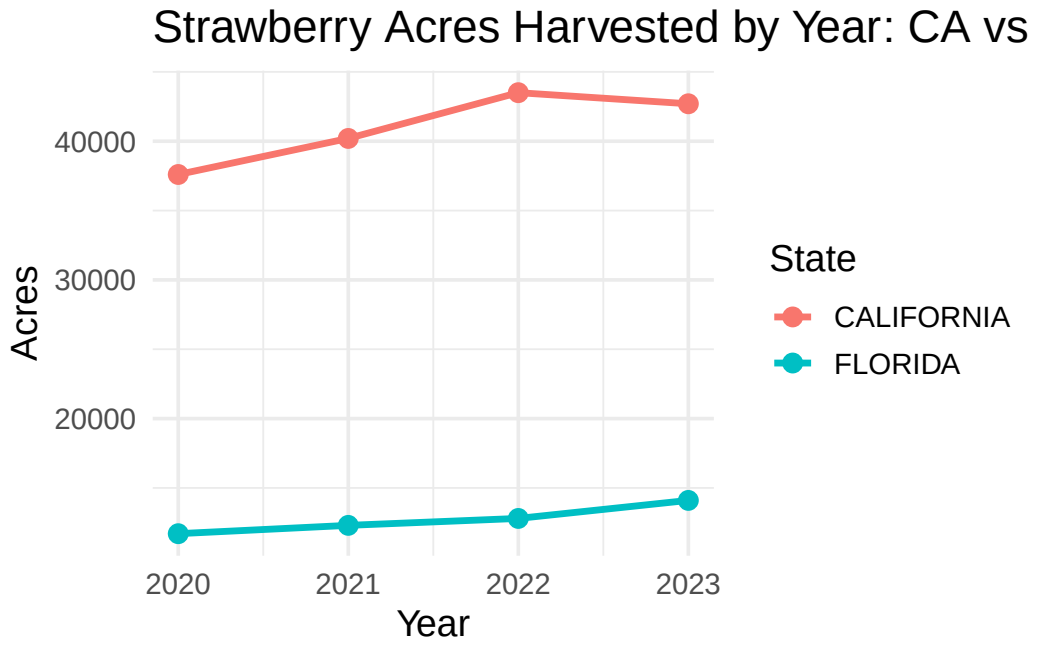


California consistently outproduces Florida, with ~5-8 times the total CWT. Over 2020–2023, California's production hovers between 23–25 million CWT, while Florida ranges around 3–4 million CWT. Both states show relatively stable or slightly increasing volumes across these years.



California achieves a much higher yield (around 600–700 CWT/acre) than Florida (about 200–220 CWT/acre). California's yield slightly declines over time (e.g., ~660 CWT/acre in 2020 down to ~620 by 2023), while Florida's yield stays lower but is relatively stable or slightly up.

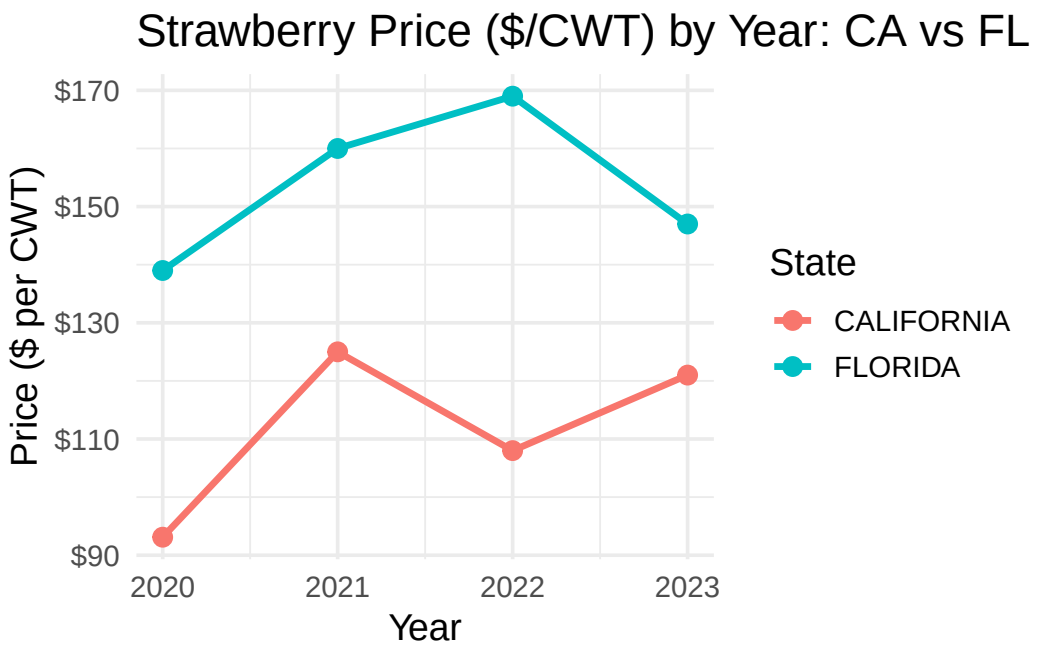
California's climate, infrastructure, and farming practices give it a yield advantage. Florida's lower yields may reflect soil/climate constraints or different farming methods.



California has around 38k–43k acres over 2020–2023, while Florida ranges from 6k–10k acres. Both states expand harvested acreage in the middle years, with California peaking around 43k in 2022 and Florida nearing 10k acres in 2023.

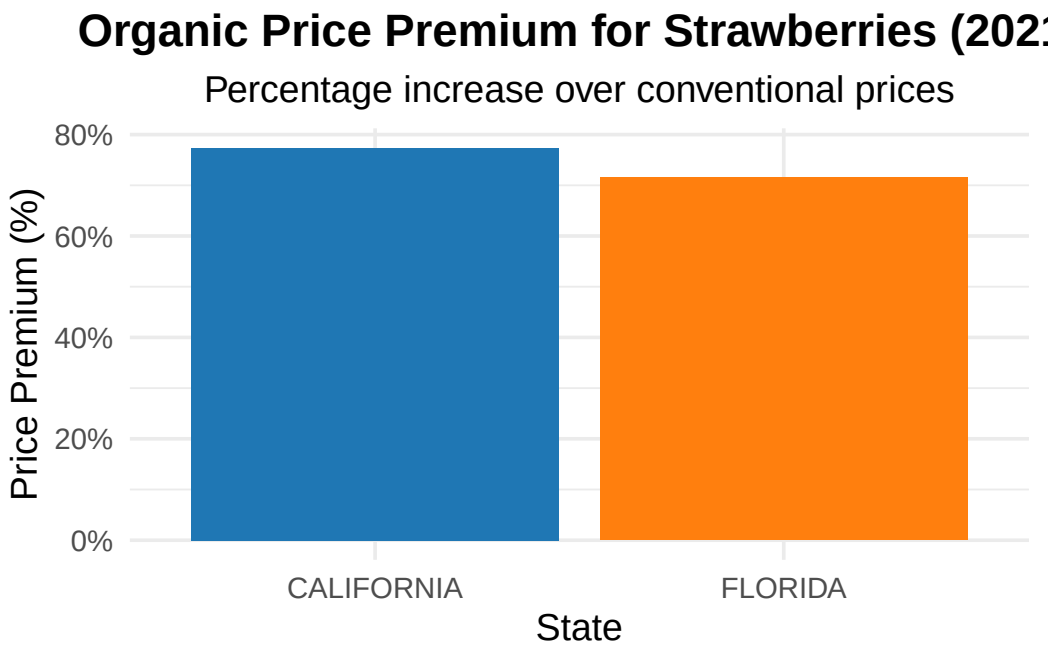
California's larger land base dedicated to strawberries (combined with higher yields) explains its dominant production volume. Florida is increasing acreage over time but remains well below California.

```
`summarise()` has grouped output by 'Year'. You can override using the
`.groups` argument.
```



Florida commands a higher price per CWT than California in most years. Florida's price ranges from \$140–\$170 (with a peak around 2022), then dips to ~\$120 in 2023. California's price is lower, starting around \$90 in 2020, peaking near \$110–\$115, then settling to ~\$95 in 2023.

Florida's smaller supply and potentially niche or local markets may keep prices higher. California's large-scale production can drive prices down. Over time, price volatility is noticeable: Florida spikes in 2022, then drops; California sees a smaller range of fluctuation.



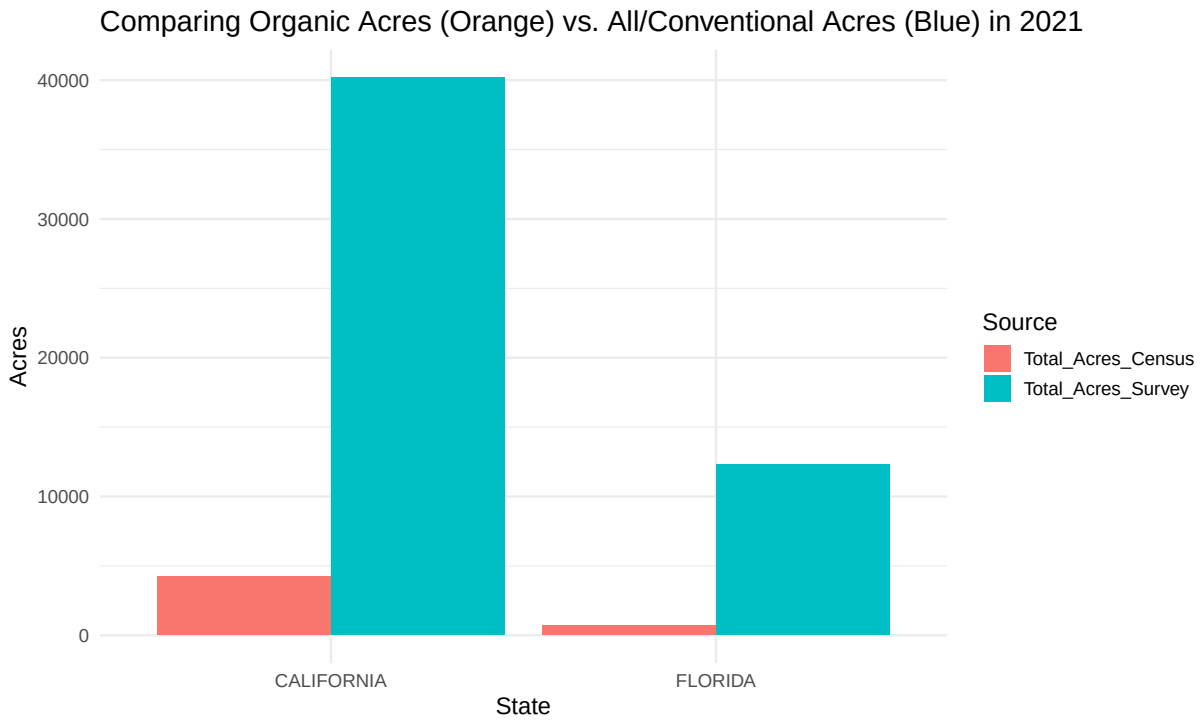
Both states demonstrate substantial organic price premiums, indicating that consumers in both markets are willing to pay significantly more for organically grown strawberries. The premiums hover around 70-80% above conventional prices, which is notable even compared to typical organic premiums in other agricultural products.

California shows a premium of approximately 76%, while Florida exhibits a slightly lower but still substantial premium of about 71%. This indicates that despite geographical and market differences, organic strawberries maintain significant price advantages in both major production regions.

```
`summarise()` has grouped output by 'Year'. You can override using the
`.groups` argument.
```

```
Warning: There was 1 warning in `mutate()`.
i In argument: `Value = as.numeric(gsub(",", "", Value))`.
Caused by warning:
! NAs introduced by coercion
```

```
`summarise()` has grouped output by 'Year'. You can override using the
`.groups` argument.
```



California: ~40k total acres (blue) vs. only a few thousand organic acres (orange).

Florida: ~10k total acres vs. a small fraction organic (~1k).

Organic remains a small slice of total acreage in both states.

Even in the top strawberry-producing states, organic acreage is significantly smaller than conventional. California leads in absolute organic acres (simply because it has more farmland overall), but as a percentage, organic is still a minor share of total production.

From all the data and graph, we can reach a conclusion that the U.S. strawberry market reveals significant contrasts between California and Florida production systems. California dominates the industry with substantially greater acreage and volume, producing over 20 million CWT compared to Florida's 3-4 million CWT. This scale advantage likely contributes to California's relatively stable pricing range ($90-$115 per CWT) and potentially lower unit costs. Florida, despite its smaller production footprint, often commands higher prices for its strawberries, though with greater market volatility, as evidenced by price peaks in 2022 followed by declines in 2023. The organic segment remains modest in both states, constituting a small fraction of total production, though California maintains more organic acreage in absolute terms. While processing data isn't explicitly detailed, a portion of total production from both states typically goes toward processed products like jams and frozen strawberries, with processing prices generally differing from fresh market prices. Over time, California has maintained relatively consistent production volumes while Florida shows modest increases in acreage and production, though still at levels far below California's scale.

```
To cite R in publications use:

  R Core Team (2024). _R: A Language and Environment for Statistical
  Computing_. R Foundation for Statistical Computing, Vienna, Austria.
  <https://www.R-project.org/>.

A BibTeX entry for LaTeX users is

  @Manual{,
    title = {R: A Language and Environment for Statistical Computing},
    author = {{R Core Team}},
    organization = {R Foundation for Statistical Computing},
    address = {Vienna, Austria},
    year = {2024},
    url = {https://www.R-project.org/},
  }

We have invested a lot of time and effort in creating R, please cite it
when using it for data analysis. See also 'citation("pkgname")' for
citing R packages.
```

References


```
R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation
  <https://www.R-project.org/>.
```


Dataset Source

Massachusetts Institute of Technology [MIT]. (2025). *Strawberry Analysis Dataset* [Data file]. Retrieved from https://github.com/MA615-415-Spring-2025/class-15---Mar-20.git

AI Assistance

OpenAI. (2025). ChatGPT (GPT-o3 mini high) [Large language model]. Retrieved from https://chat.openai.com

Anthropic. (2025). Claude 3.7 Sonnet [Large language model]. Retrieved from https://claude.ai