Strawberries in the United States: Exploratory Data Analysis

GRS 615: Data Science in R

Nicole Kingdon

October 18, 2023

Introduction

Strawberries & Positive Health

Strawberries are a fruit that holds several minerals, vitamins, and phytonutrients (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.

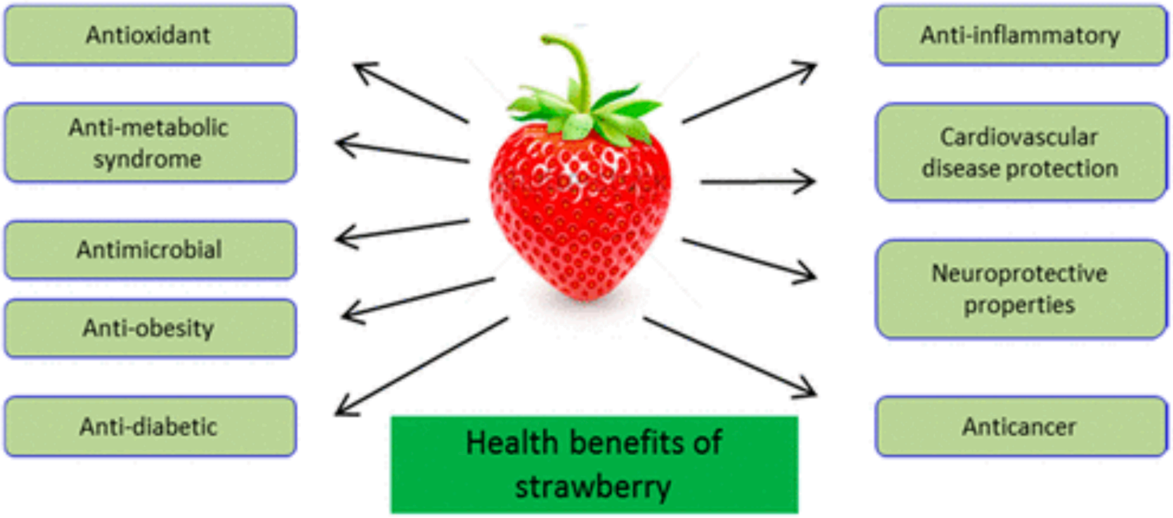

Figure 1: Health benefits of strawberries (Afrin et al., 2016)

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.

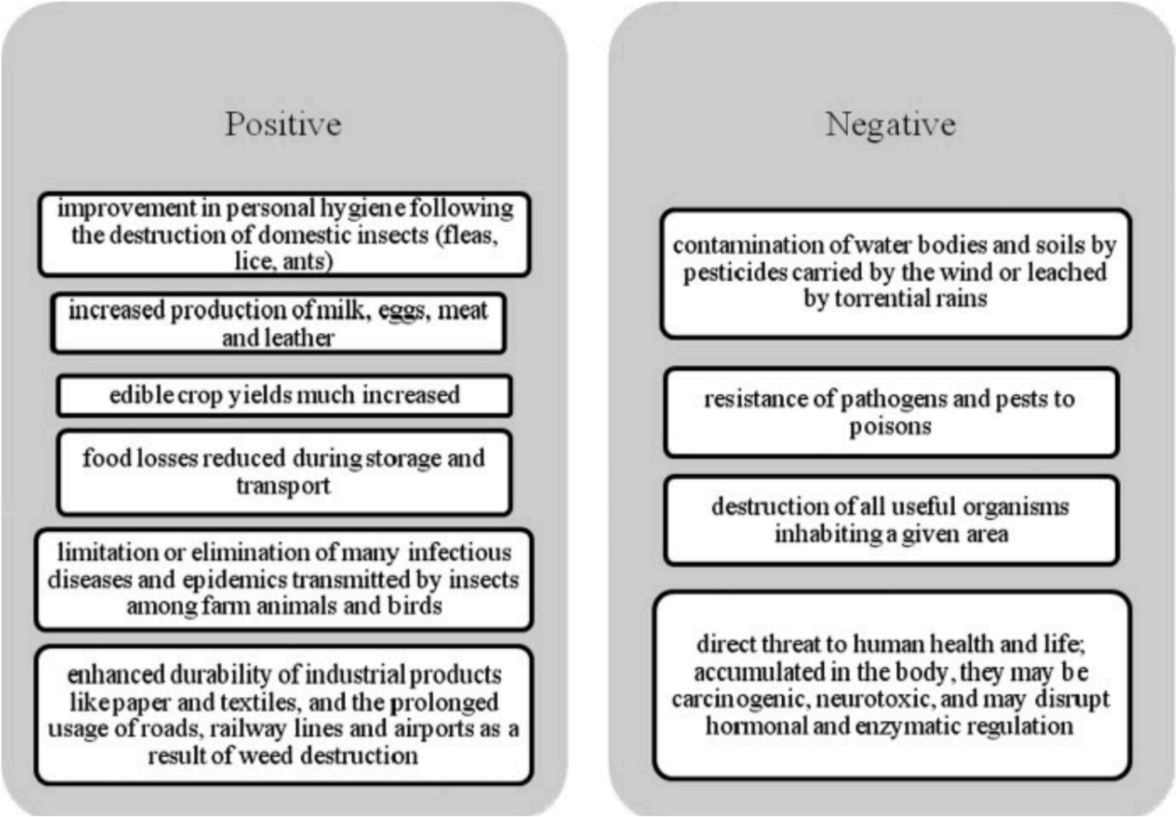
Positive
improvement in personal hygiene following the destruction of domestic insects (fleas, lice, ants)
increased production of milk, eggs, meat and leather
edible crop yields much increased
food losses reduced during storage and transport
limitation or elimination of many infectious diseases and epidemics transmitted by insects among farm animals and birds
enhanced durability of industrial products like paper and textiles, and the prolonged usage of roads, railway lines and airports as a result of weed destruction
Negative
contamination of water bodies and soils by pesticides carried by the wind or leached by torrential rains
resistance of pathogens and pests to poisons
destruction of all useful organisms inhabiting a given area
direct threat to human health and life; accumulated in the body, they may be carcinogenic, neurotoxic, and may disrupt hormonal and enzymatic regulation

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

Organic Strawberries

Strawberries that use alternative and safer methods to production 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.

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 uploaded to R is titled `strawberry` (see below).

```r
strawberry <- read_csv("strawberry.csv", col_names = TRUE)
```

```
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
library(knitr)
library(kableExtra)
```

```

Attaching package: 'kableExtra'


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

    group_rows
```

```r
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.3     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.2
```

```
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter()          masks stats::filter()
x kableExtra::group_rows() masks dplyr::group_rows()
x dplyr::lag()             masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```
library(stringr)
```

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.

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.