Sberry_cleaning

mary liu

10/18/2020

Acquire and read the data

These data were collected from the USDA database selector: https://quickstats.nass.usda.gov (https://quickstats.nass.usda.gov)

The data were stored online (https://quickstats.nass.usda.gov/results/D416E96E-3D5C-324C-9334-1D38DF88FFF1) and then downloaded as a CSV file.

Data selected from the NASS database often has columns without any data or with a single repeated Values. The berries data had only 8 out of 21 columns containing meaningful data.

Year	Period	State	Commodity	Data Item	Domain	Domain Category	Value
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME - PRICE RECEIVED, MEASURED IN $ / LB	TOTAL	NOT SPECIFIED	2.85
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME, FRESH MARKET - PRICE RECEIVED, MEASURED IN $ / LB	TOTAL	NOT SPECIFIED	3.56
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME, PROCESSING - PRICE RECEIVED, MEASURED IN $ / LB	TOTAL	NOT SPECIFIED	0.29
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES - PRICE RECEIVED, MEASURED IN $ / LB	TOTAL	NOT SPECIFIED	2.69
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES, FRESH MARKET - PRICE RECEIVED, MEASURED IN $ / LB	TOTAL	NOT SPECIFIED	D.
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES, PROCESSING - PRICE RECEIVED, MEASURED IN $ / LB	TOTAL	NOT SPECIFIED	D.

This table contains informaton about berries: blueberries, raspberries, and strawberries.

When the data have been cleaned and organized, the three kinds of berries will be separted into tables with the same stucture so that they can be compared. So, working with Blueberries along demonstrates how the data will be cleaned and organized for all three kinds of berries. Only the “YEAR” time periond will be considered.

Data Strawberries

Cleaning Data Item colume

```
## [1] FALSE
```

Cleaning Domain and Domain Category colume

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

Year	State	Type	Measure	Material	Value	Chemical
2019	CALIFORNIA	ACRES HARVESTED			35,400	
2019	CALIFORNIA	ACRES PLANTED			36,000	
2019	CALIFORNIA	PRODUCTION	$		2,221,320,000	
2019	CALIFORNIA	PRODUCTION	CWT		20,500,000	
2019	CALIFORNIA	YIELD	CWT / ACRE		580	
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(AZOXYSTROBIN = 128810)	5,500	FUNGICIDE
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(BACILLUS AMYLOLIQUEFACIENS MBI 600 = 129082)	(NA)	FUNGICIDE

Year	State	Type	Measure	Material	Value	Chemical
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(BACILLUS AMYLOLIQUEFACIENS STRAIN D747 = 16482)	(NA)	FUNGICIDE
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(BACILLUS PUMILUS = 6485)	(NA)	FUNGICIDE
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(BACILLUS SUBT. GB03 = 129068)	(NA)	FUNGICIDE

EDA(Exploratory Data Analysis)

Variables

There are total 10 variables after cleaning. The ten variables in data Strawberries is Year, State, Type, Measure, Domain 1, Domain 2, Domain Category 1, Domain Category 2, Domain Category Detail.

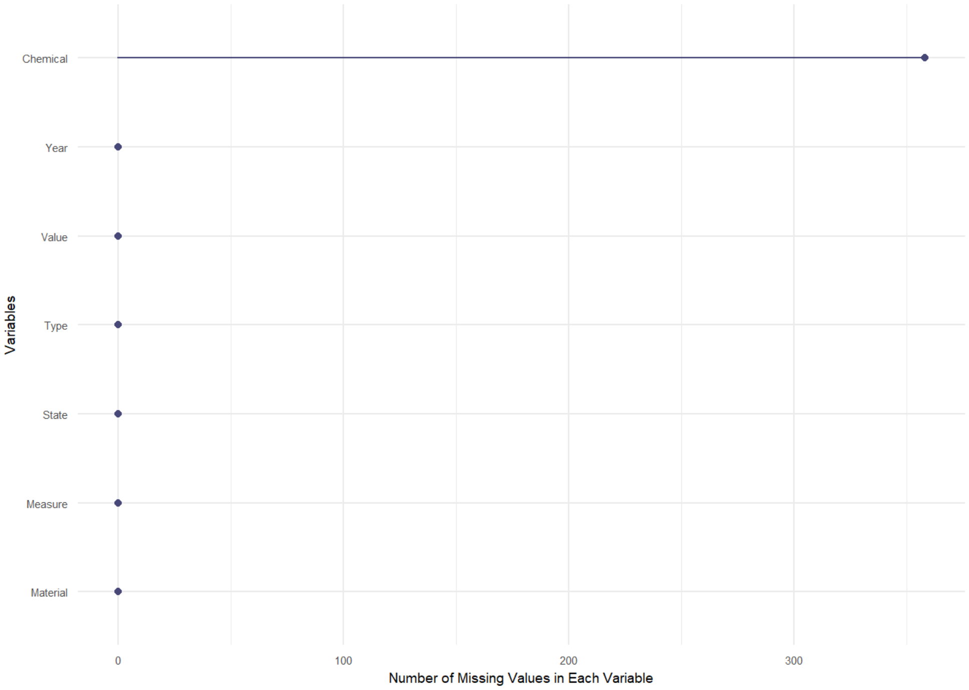
Observations

There are total 3220 observations. The head eight observations are listed below:

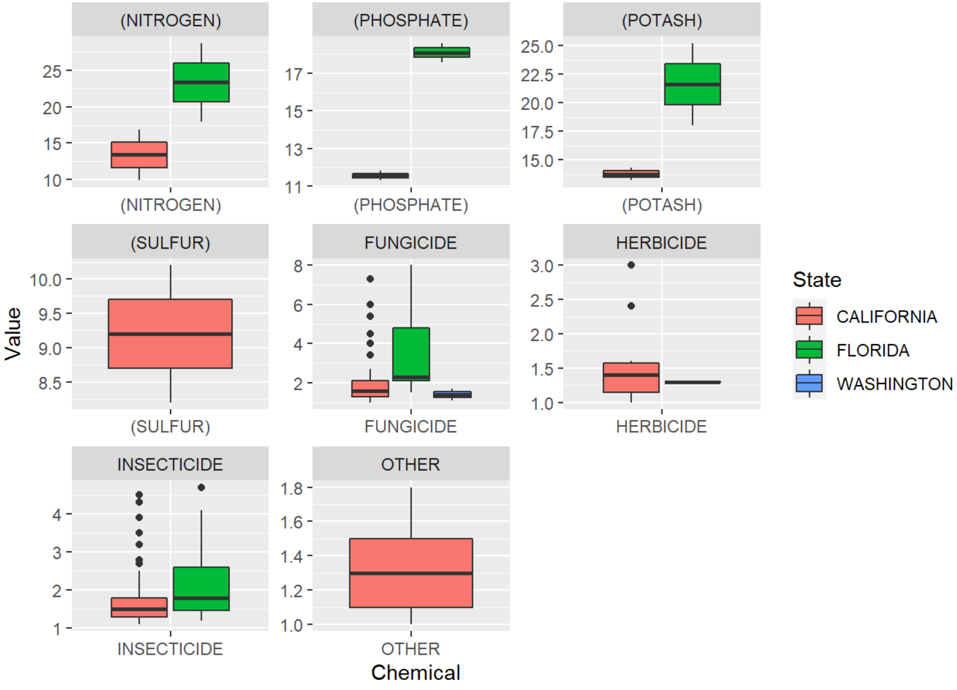
Year	State	Type	Measure	Material	Value	Chemical
2019	CALIFORNIA	ACRES HARVESTED			35,400	
2019	CALIFORNIA	ACRES PLANTED			36,000	
2019	CALIFORNIA	PRODUCTION	$		2,221,320,000	
2019	CALIFORNIA	PRODUCTION	CWT		20,500,000	
2019	CALIFORNIA	YIELD	CWT / ACRE		580	
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(AZOXYSTROBIN = 128810)	5,500	FUNGICIDE
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(BACILLUS AMYLOLIQUEFACIENS MBI 600 = 129082)	(NA)	FUNGICIDE
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(BACILLUS AMYLOLIQUEFACIENS STRAIN D747 = 16482)	(NA)	FUNGICIDE
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(BACILLUS PUMILUS = 6485)	(NA)	FUNGICIDE
2019	CALIFORNIA	BEARING - APPLICATIONS	LB	(BACILLUS SUBT. GB03 = 129068)	(NA)	FUNGICIDE

Missing Values

First I conduct basic data reprocessing. Missing values for sberry dataset are shown in the histogram below. The plot below shows that Material, Chemical and Measure variable has missing value.



### Compared variables I compared the chemical values by different states.



## Shiny app

please see file app.R

slide

please see file berry.PPT

reference

Class recording 11, 14-18, MA615

http://rstudio.github.io/shiny/tutorial/ (http://rstudio.github.io/shiny/tutorial/)

https://shiny.rstudio.com/tutorial/ (https://shiny.rstudio.com/tutorial/)

Garrett Grolemund, Hadley Wickham, R for Data Science, https://r4ds.had.co.nz/ (https://r4ds.had.co.nz/)