Report of Berry Project

Haoqi Wang

10/20/2020

1.Overview

In this assignment, I use the data that were collected from the USDA data base, and I choose strawberry as my data set to do the data cleaning, data organization, exploratory data analysis. Finally, I find some correlations between variables.

2.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.

3. Choose Strawberries-Data clean and Data organize

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.

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 Strawberries along demonstrates how the data will be cleaned and organized.

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

Year	Period	State	Commodity	Data.Item	Domain	Domain.Category	Value
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.

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

```
## [1] TRUE
```

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

```
##  [1] ""                                  "MEASURED IN $"
##  [3] "MEASURED IN CWT"                   "MEASURED IN CWT / ACRE"
##  [5] "MEASURED IN LB"                    "MEASURED IN LB / ACRE / APPLICATION"
##  [7] "MEASURED IN LB / ACRE / YEAR"      "MEASURED IN NUMBER"
##  [9] "MEASURED IN PCT OF AREA BEARING"   "MEASURED IN TONS"
```

```
##  [1] "Year"        "State"       "Type"        "Avg"         "Marketing"
##  [6] "Harvest"     "Chem_family" "Materials"   "Chemical"    "Value"
## [11] "m_in_1"      "m_in_2"      "Measures"
```

```
## NULL
```

```
## NULL
```

```
## [1] " AVG"
```

```
## [1] "MEASURED IN LB / ACRE / APPLICATION"
```

```
##   [1] "(AZOXYSTROBIN = 128810)"
##   [2] "(BACILLUS AMYLOLIQUEFACIENS MBI 600 = 129082)"
##   [3] "(BACILLUS AMYLOLIQUEFACIENS STRAIN D747 = 16482)"
##   [4] "(BACILLUS PUMILUS = 6485)"
##   [5] "(BACILLUS SUBT. GB03 = 129068)"
##   [6] "(BACILLUS SUBTILIS = 6479)"
##   [7] "(BLAD = 30006)"
##   [8] "(BORAX DECAHYDRATE = 11102)"
##   [9] "(BOSCALID = 128008)"
##  [10] "(BT SUBSP KURSTAKI EVB-113-19 = 6544)"
##  [11] "(CAPTAN = 81301)"
##  [12] "(COPPER HYDROXIDE = 23401)"
##  [13] "(COPPER OCTANOATE = 23306)"
##  [14] "(CYFLUFENAMID = 555550)"
##  [15] "(CYPRODINIL = 288202)"
##  [16] "(DIFENOCONAZOLE = 128847)"
##  [17] "(FENHEXAMID = 90209)"
##  [18] "(FLUDIOXONIL = 71503)"
##  [19] "(FLUOPYRAM = 80302)"
##  [20] "(FLUXAPYROXAD = 138009)"
##  [21] "(FOSETYL-AL = 123301)"
##  [22] "(ISOFETAMID = 270000)"
##  [23] "(MEFENOXAM = 113502)"
##  [24] "(MONO-POTASSIUM SALT = 76416)"
##  [25] "(MYCLOBUTANIL = 128857)"
##  [26] "(PENTHIOPYRAD = 90112)"
##  [27] "(POLYOXIN D ZINC SALT = 230000)"
##  [28] "(POTASSIUM BICARBON. = 73508)"
##  [29] "(PROPICONAZOLE = 122101)"
##  [30] "(PYRACLOSTROBIN = 99100)"
##  [31] "(PYRIMETHANIL = 288201)"
##  [32] "(QUINOLINE = 55459)"
##  [33] "(STREPTOMYCES LYDICUS = 6327)"
##  [34] "(SULFUR = 77501)"
##  [35] "(TETRACONAZOLE = 120603)"
##  [36] "(THIOPHANATE-METHYL = 102001)"
##  [37] "(THIRAM = 79801)"
##  [38] "(TRICHODERMA HARZ. = 119202)"
##  [39] "(TRIFLOXYSTROBIN = 129112)"
##  [40] "(TRIFLUMIZOLE = 128879)"
##  [41] "(CARFENTRAZONE-ETHYL = 128712)"
##  [42] "(FLUMIOXAZIN = 129034)"
##  [43] "(GLUFOSINATE-AMMONIUM = 128850)"
##  [44] "(GLYPHOSATE ISO. SALT = 103601)"
##  [45] "(GLYPHOSATE POT. SALT = 103613)"
##  [46] "(NAPROPAMIDE = 103001)"
##  [47] "(OXYFLUORFEN = 111601)"
##  [48] "(PARAQUAT = 61601)"
##  [49] "(PENDIMETHALIN = 108501)"
##  [50] "(SULFENTRAZONE = 129081)"
##  [51] "(ABAMECTIN = 122804)"
##  [52] "(ACEQUINOCYL = 6329)"
##  [53] "(ACETAMIPRID = 99050)"
```

```
##  [54] "(AZADIRACHTIN = 121701)"
##  [55] "(BEAUVERIA BASSIANA = 128924)"
##  [56] "(BIFENAZATE = 586)"
##  [57] "(BIFENTHRIN = 128825)"
##  [58] "(BT KURSTAK ABTS-1857 = 6523)"
##  [59] "(BT KURSTAKI ABTS-351 = 6522)"
##  [60] "(BT KURSTAKI SA-11 = 6519)"
##  [61] "(BT SUB AIZAWAI GC-91 = 6426)"
##  [62] "(BUPROFEZIN = 275100)"
##  [63] "(BURKHOLDERIA A396 CELLS & MEDIA = 6534)"
##  [64] "(CARBARYL = 56801)"
##  [65] "(CHLORANTRANILIPROLE = 90100)"
##  [66] "(CHLORPYRIFOS = 59101)"
##  [67] "(CHROMOBAC SUBTSUGAE PRAA4-1 CELLS AND SPENT MEDIA = 16329)"
##  [68] "(CYANTRANILIPROLE = 90098)"
##  [69] "(CYFLUMETOFEN = 138831)"
##  [70] "(ETOXAZOLE = 107091)"
##  [71] "(FENBUTATIN-OXIDE = 104601)"
##  [72] "(FENPROPATHRIN = 127901)"
##  [73] "(FENPYROXIMATE = 129131)"
##  [74] "(FLONICAMID = 128016)"
##  [75] "(FLUPYRADIFURONE = 122304)"
##  [76] "(HELICOVERPA ZEA NPV = 107300)"
##  [77] "(HEXYTHIAZOX = 128849)"
##  [78] "(IMIDACLOPRID = 129099)"
##  [79] "(MALATHION = 57701)"
##  [80] "(METHOXYFENOZIDE = 121027)"
##  [81] "(NALED = 34401)"
##  [82] "(NEEM OIL = 25006)"
##  [83] "(NEEM OIL"
##  [84] "(NOVALURON = 124002)"
##  [85] "(PIPERONYL BUTOXIDE = 67501)"
##  [86] "(POTASSIUM SALTS = 79021)"
##  [87] "(PYRETHRINS = 69001)"
##  [88] "(PYRIDABEN = 129105)"
##  [89] "(PYRIPROXYFEN = 129032)"
##  [90] "(SOYBEAN OIL = 31605)"
##  [91] "(SPINETORAM = 110007)"
##  [92] "(SPINOSAD = 110003)"
##  [93] "(SPIROMESIFEN = 24875)"
##  [94] "(SULFOXAFLOR = 5210)"
##  [95] "(THIAMETHOXAM = 60109)"
##  [96] "(ZETA-CYPERMETHRIN = 129064)"
##  [97] "(ACIBENZOLAR-S-METHYL = 61402)"
##  [98] "(AUREOBASIDIUM PULLULANS DSM 14940 = 46010)"
##  [99] "(AUREOBASIDIUM PULLULANS DSM 14941 = 36010)"
## [100] "(BT KURSTAKI SA-12 = 6518)"
## [101] "(CAPRIC ACID = 128955)"
## [102] "(CAPRYLIC ACID = 128919)"
## [103] "(CHLOROPICRIN = 81501)"
## [104] "(DICHLOROPROPENE = 29001)"
## [105] "(FLUTRIAFOL = 128940)"
## [106] "(GLIOCLADIUM VIRENS = 129000)"
## [107] "(HYDROGEN PEROXIDE = 595)"
```

```
## [108] "(IRON PHOSPHATE = 34903)"
## [109] "(METALDEHYDE = 53001)"
## [110] "(METAM-POTASSIUM = 39002)"
## [111] "(METAM-SODIUM = 39003)"
## [112] "(PAECILOMYCES FUMOSOR = 115002)"
## [113] "(PEROXYACETIC ACID = 63201)"
## [114] "(POTASSIUM SILICATE = 72606)"
## [115] "(REYNOUTRIA SACHALINE = 55809)"
## [116] "(TRICHODERMA VIRENS STRAIN G-41 = 176604)"
## [117] " "
## [118] "(COPPER CHLORIDE HYD. = 23501)"
## [119] "(DODINE = 44301)"
## [120] "(FLUTOLANIL = 128975)"
## [121] "(2,4-D"
## [122] "(CYPERMETHRIN = 109702)"
## [123] "(ALKYL. DIM. BENZ. AM = 69105)"
## [124] "(DECYLDIMETHYLOCTYL = 69165)"
## [125] "(DIDECYL DIM. AMMON. = 69166)"
## [126] "(DIMETHYLDIOCTYL = 69149)"
## [127] "(IPRODIONE = 109801)"
## [128] "(CYFLUMETOFEN= 138831)"
## [129] "(DIAZINON = 57801)"
## [130] "(EMAMECTIN BENZOATE = 122806)"
## [131] "(LAMBDA-CYHALOTHRIN = 128897)"
## [132] "(SPIROTETRAMAT = 392201)"
## [133] "(MUSTARD OIL = 4901)"
## [134] "(CYTOKININS = 116801)"
## [135] "(DIMETHYL DISULFIDE (DMDS) = 29088)"
## [136] "(COPPER OXIDE = 25601)"
## [137] "(AMMONIUM PELARGONATE = 31802)"
## [138] "(FLUBENDIAMIDE = 27602)"
## [139] "(METHYL BROMIDE = 53201)"
## [140] "(MINERAL OIL = 63502)"
## [141] "(CHLOROTHALONIL = 81901)"
## [142] "(CYAZOFAMID = 85651)"
## [143] "(MANCOZEB = 14504)"
## [144] "(HALOSULFURON-METHYL = 128721)"
## [145] "(ENDOSULFAN = 79401)"
## [146] "(CAPSAICIN = 70701)"
## [147] "(HARPIN A B PROTEIN = 6506)"
## [148] "(CLETHODIM = 121011)"
## [149] "(CLOPYRALID MONO SALT = 117401)"
## [150] "(SIMAZINE = 80807)"
## [151] "(TERBACIL = 12701)"
## [152] "(FERRIC SODIUM EDTA = 139114)"
## [153] "(CLOMAZONE = 125401)"
```

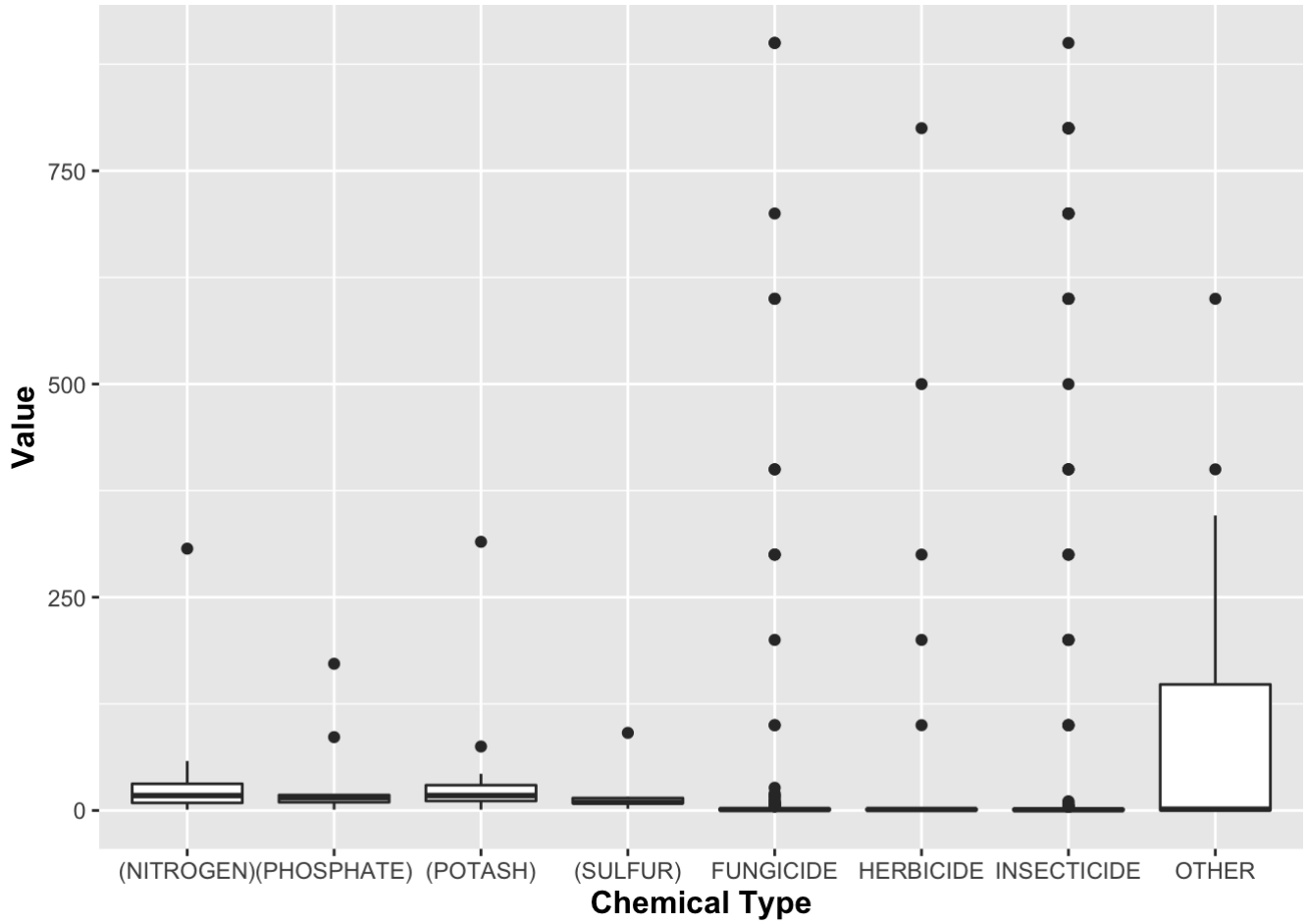
```
## [1] "FUNGICIDE"   "HERBICIDE"   "INSECTICIDE" "OTHER"       "(NITROGEN)"
## [6] "(PHOSPHATE)" "(POTASH)"    "(SULFUR)"
```

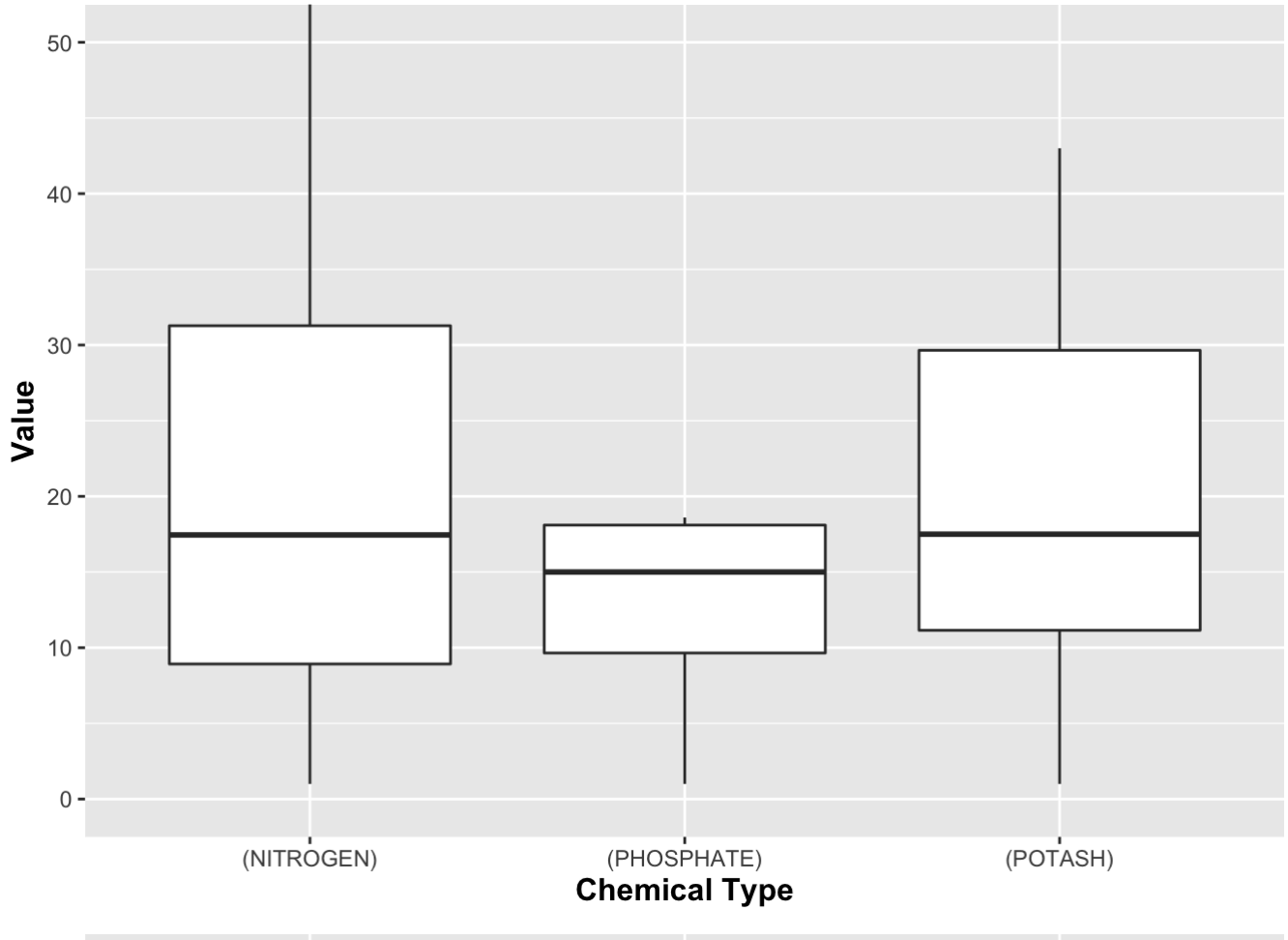
Year	State	Type	Production	Avg	Measures	Materials	Chemical	Value

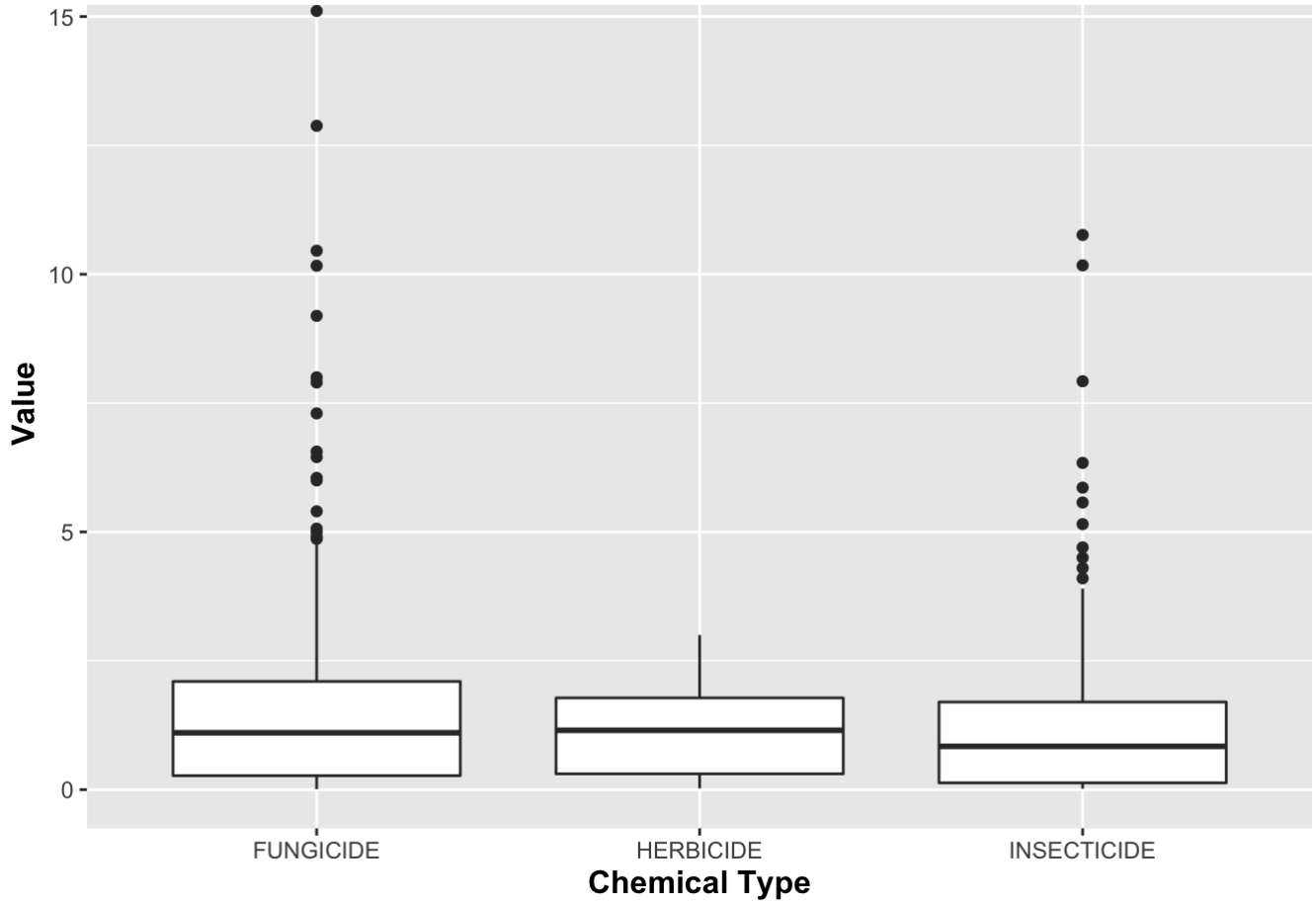
Year	State	Type	Production	Avg	Measures	Materials	Chemical	Value
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(AZOXYSTROBIN = 128810)	FUNGICIDE	5,500
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BACILLUS AMYLOLIQUEFACIENS MBI 600 = 129082)	FUNGICIDE	(NA)
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BACILLUS AMYLOLIQUEFACIENS STRAIN D747 = 16482)	FUNGICIDE	(NA)
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BACILLUS PUMILUS = 6485)	FUNGICIDE	(NA)
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BACILLUS SUBT. GB03 = 129068)	FUNGICIDE	(NA)
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BACILLUS SUBTILIS = 6479)	FUNGICIDE	(NA)
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BLAD = 30006)	FUNGICIDE	1,200
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BORAX DECAHYDRATE = 11102)	FUNGICIDE	300
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BOSCALID = 128008)	FUNGICIDE	6,200
2019	CALIFORNIA	BEARING	APPLICATIONS		MEASURED IN LB	(BT SUBSP KURSTAKI EVB-113-19 = 6544)	FUNGICIDE	(NA)

4.Exploratory data analysis

(1)chemical and value

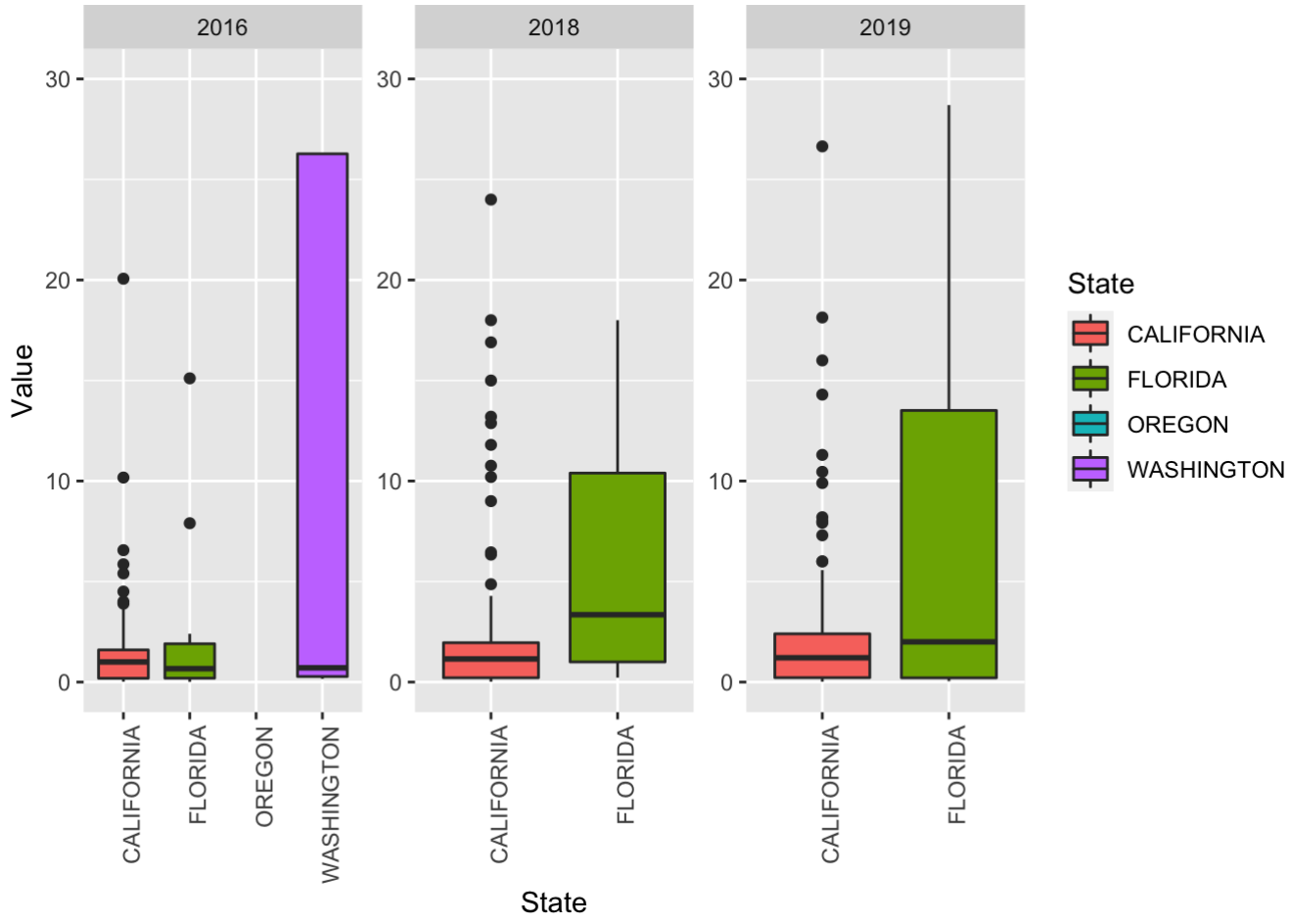

750
500
250
0
Value
(NITROGEN)
(PHOSPHATE)
(POTASH)
(SULFUR)
FUNGICIDE
HERBICIDE
INSECTICIDE
OTHER
Chemical Type


50
40
30
20
10
0
Value
(NITROGEN)
(PHOSPHATE)
(POTASH)
Chemical Type

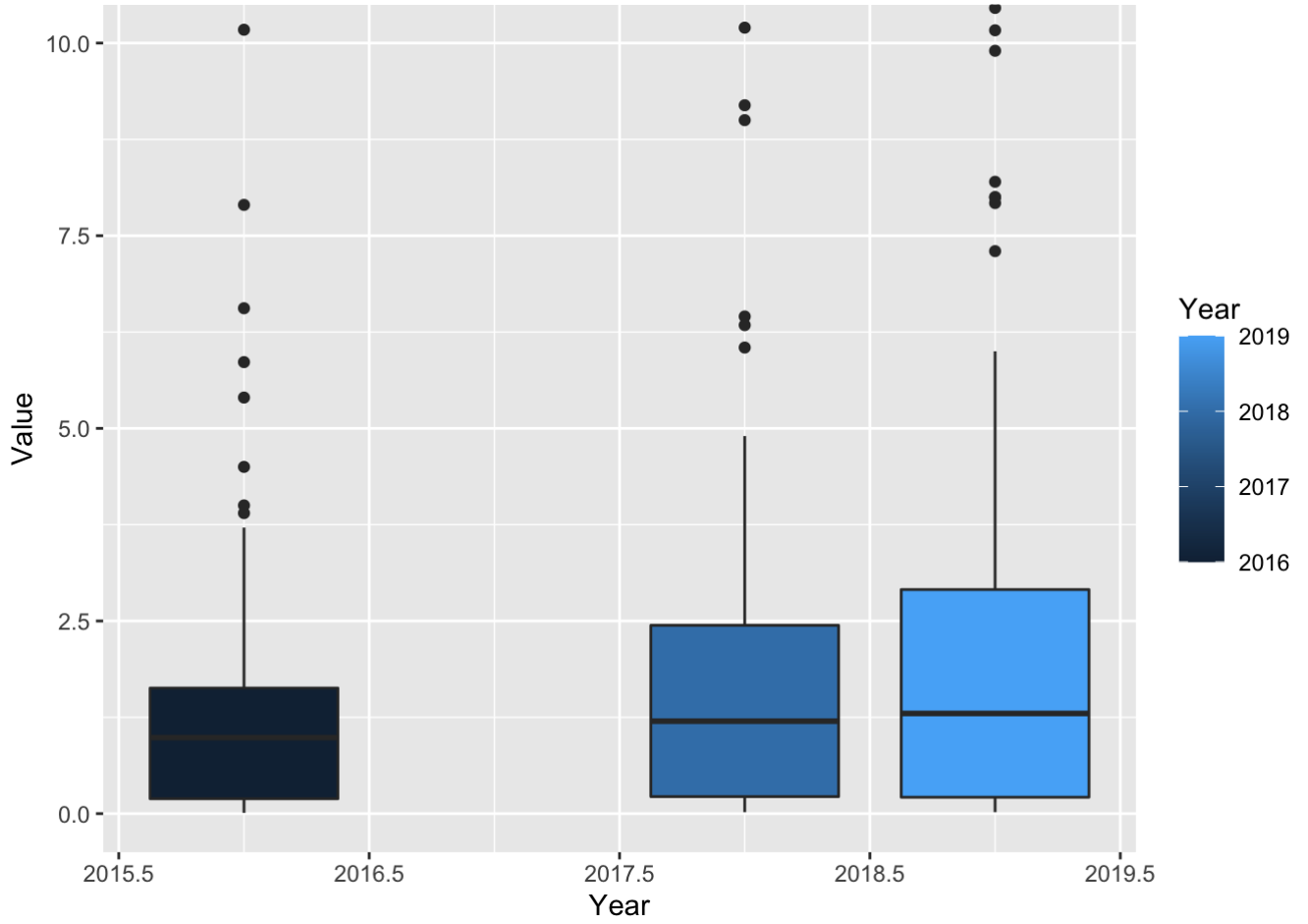


##

(2)State and value



(3)Year and value




```
## List of 1
##  $ axis.text.x:List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 1
##   ..$ vjust        : NULL
##   ..$ angle        : num 90
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

5.Conclusion

According the dataset and EDA distribution, I can conclude that the correlations of chemical type, state and years are different. But the simple boxplot does not show more right now. I may need to more time to research deeper.

During this time, I learned how to clean data, and how to organize data. When I did EDA part, I really want to use pca to predict , but there are some errors that need me to fix later. I will review the class recording to fix them.

6.Reference

```
##
## To cite ggplot2 in publications, please use:
##
##   H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
##   Springer-Verlag New York, 2016.
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     author = {Hadley Wickham},
##     title = {ggplot2: Elegant Graphics for Data Analysis},
##     publisher = {Springer-Verlag New York},
##     year = {2016},
##     isbn = {978-3-319-24277-4},
##     url = {https://ggplot2.tidyverse.org},
##   }
```

```
##
## To cite the 'knitr' package in publications use:
##
##   Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report
##   Generation in R. R package version 1.29.
##
##   Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition.
##   Chapman and Hall/CRC. ISBN 978-1498716963
##
##   Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible
##   Research in R. In Victoria Stodden, Friedrich Leisch and Roger D.
##   Peng, editors, Implementing Reproducible Computational Research.
##   Chapman and Hall/CRC. ISBN 978-1466561595
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

```
##
## To cite package 'magrittr' in publications use:
##
##   Stefan Milton Bache and Hadley Wickham (2014). magrittr: A
##   Forward-Pipe Operator for R. R package version 1.5.
##   https://CRAN.R-project.org/package=magrittr
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {magrittr: A Forward-Pipe Operator for R},
##     author = {Stefan Milton Bache and Hadley Wickham},
##     year = {2014},
##     note = {R package version 1.5},
##     url = {https://CRAN.R-project.org/package=magrittr},
##   }
##
## ATTENTION: This citation information has been auto-generated from the
## package DESCRIPTION file and may need manual editing, see
## 'help("citation")'.
```

```
##
## To cite package 'kableExtra' in publications use:
##
##   Hao Zhu (2020). kableExtra: Construct Complex Table with 'kable' and
##   Pipe Syntax. R package version 1.2.1.
##   https://CRAN.R-project.org/package=kableExtra
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {kableExtra: Construct Complex Table with 'kable' and Pipe Syntax},
##     author = {Hao Zhu},
##     year = {2020},
##     note = {R package version 1.2.1},
##     url = {https://CRAN.R-project.org/package=kableExtra},
##   }
```

```
##
## Dietrich J (2020). _citation: Software Citation Tools_. R package
## version 0.4.1.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {citation: Software Citation Tools},
##     author = {Jan Philipp Dietrich},
##     year = {2020},
##     note = {R package version 0.4.1},
##   }
```

Class recording 14-18, MA615

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