Strawberries

Haoran Su

10/19/2020

#Strawberries ##Background The data were collected from the USDA database selector: https://quickstats.nass.usda.gov

The data were stored online and then downloaded as a CSV file.

In this assignment, after cleaning and reorganizing data, we select the strawberries from three kinds of berries and saperate columns to leave the required data. Then we make boxplots for strawberries of years and states. It can also be visualized on the shiny app.

Data cleaning and reorganizing

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
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME - PRICE R
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME, FRESH M
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME, PROCESS
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES - PRICE RECEIVE
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES, FRESH MARKET
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES, PROCESSING - P

Year	Period	State	Commodity	Data Item
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME - PRICE R
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME, FRESH M
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME, PROCESS
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES - PRICE RECEIVE
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES, FRESH MARKET
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES, PROCESSING - P

This table contains information about berries: blueberries, raspberries, and strawberries. We only choose Strawberries for analyzing and data need further cleaning and organizing. Only YEAR" time period will be considered. Also, we separate the `Data Item`, `Domain`, `Domain Category` columns for further cleaning process and remove the redundant columns.

Within the “what” and “meas” columns, there are data begins with “MEASURED IN”.We want to put them all in the same column.

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

```
## [7] "MEASURED IN $"                       "MEASURED IN CWT"
## [9] "MEASURED IN TONS"
```

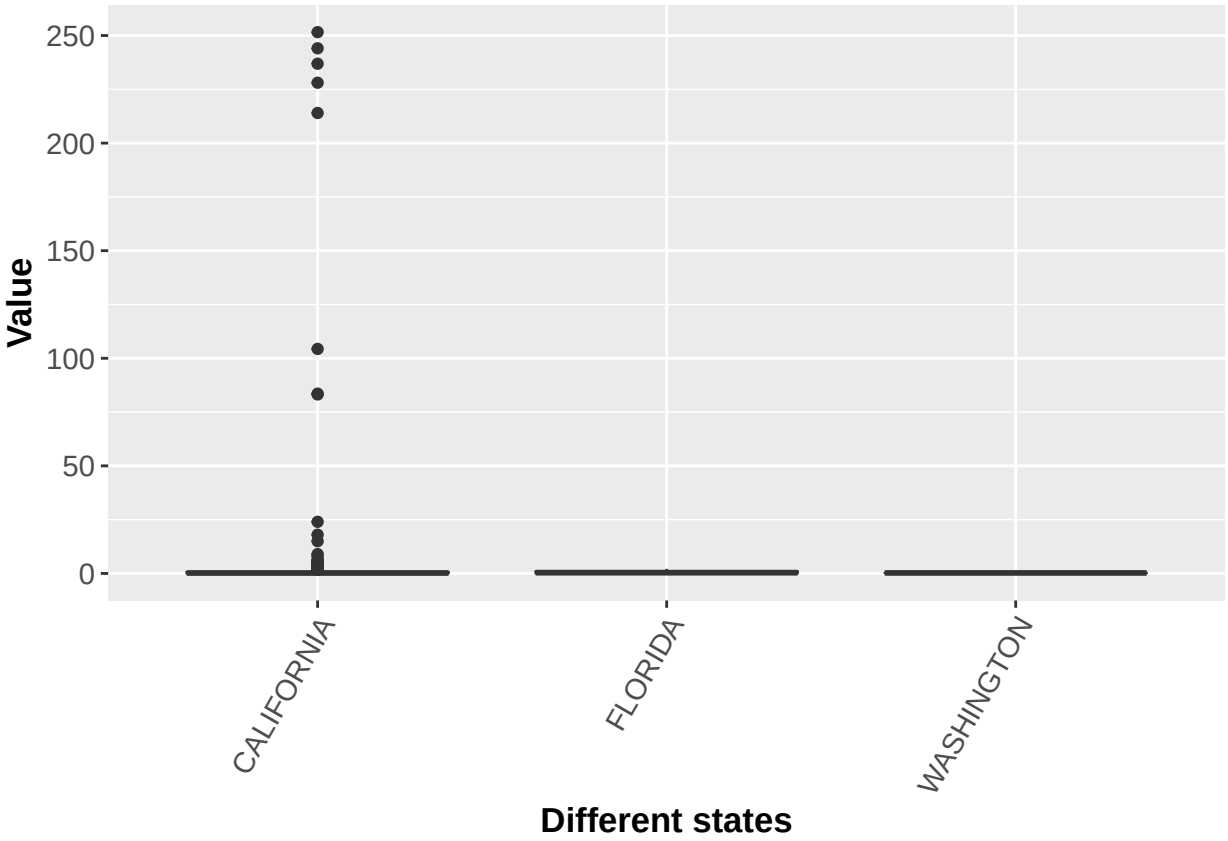
Finally, we rename the columns of the dataset and select those that are required in this analysis.

```
## [1] ""                      "$"                      "CWT"
## [4] "CWT /"                 "APPLICATIONS"           "TREATED"
## [7] "- PRODUCTION"          "UTILIZED - PRODUCTION" "PRODUCTION"
```

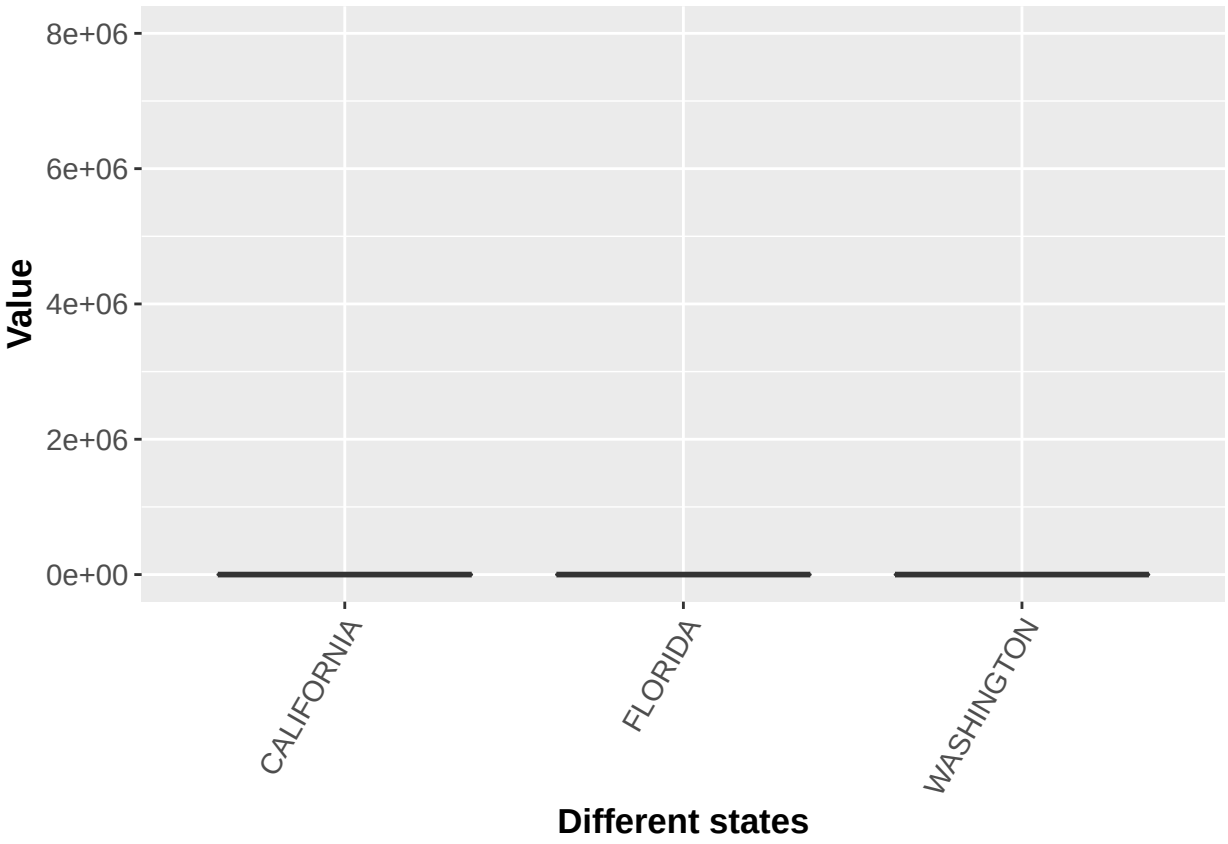
```
## [1] "Year"       "State"      "type"       "production" "Avg"
## [6] "Measures"   "Materials"  "Chemical"   "Value"
```

After cleaning and reorganizing the columns, current headings of data "sberry" is as above. As the chemicals of strawberries is the factor we want to analysize, we do further filtering for the data.

1. We make a box plot of chemical values being applied to strawberries according to different states.

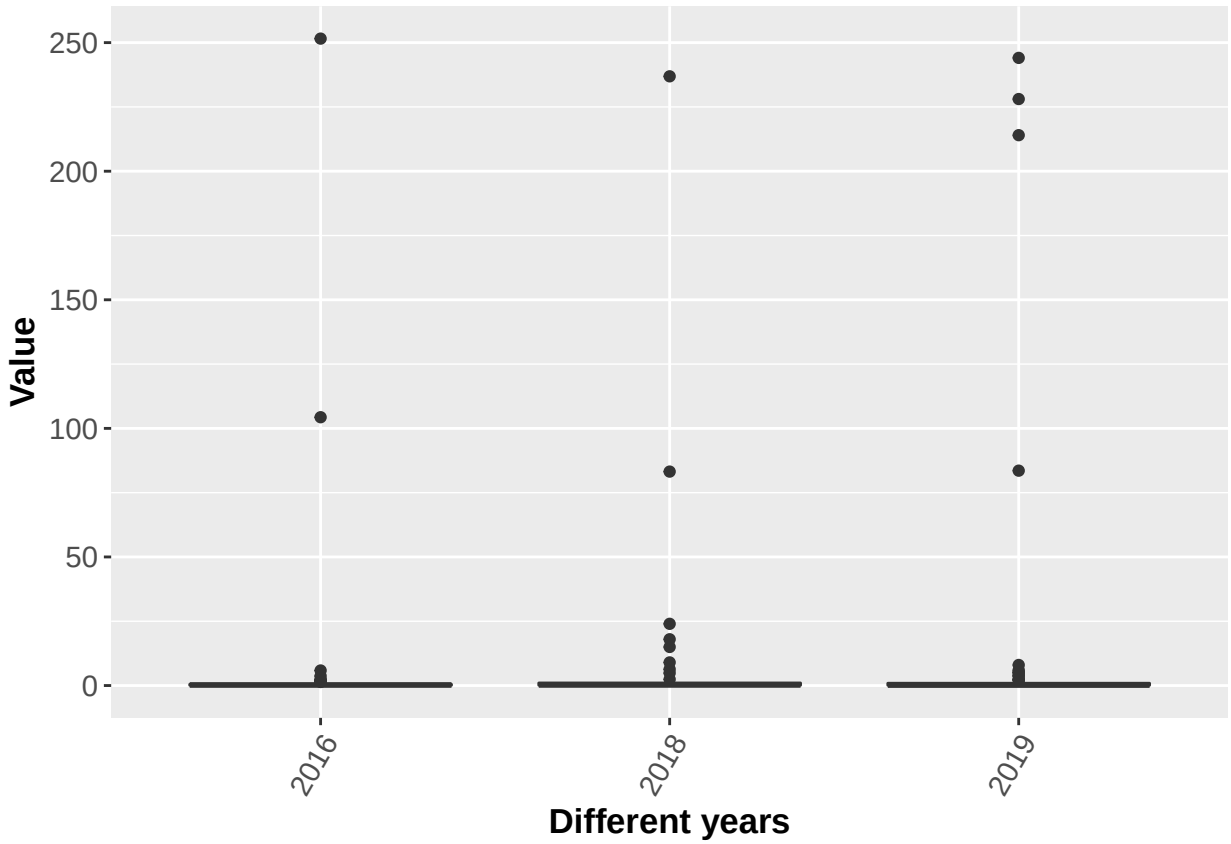


From the plot, we find many outliers influencing the presentation so we remove the outliners and make another box plot.

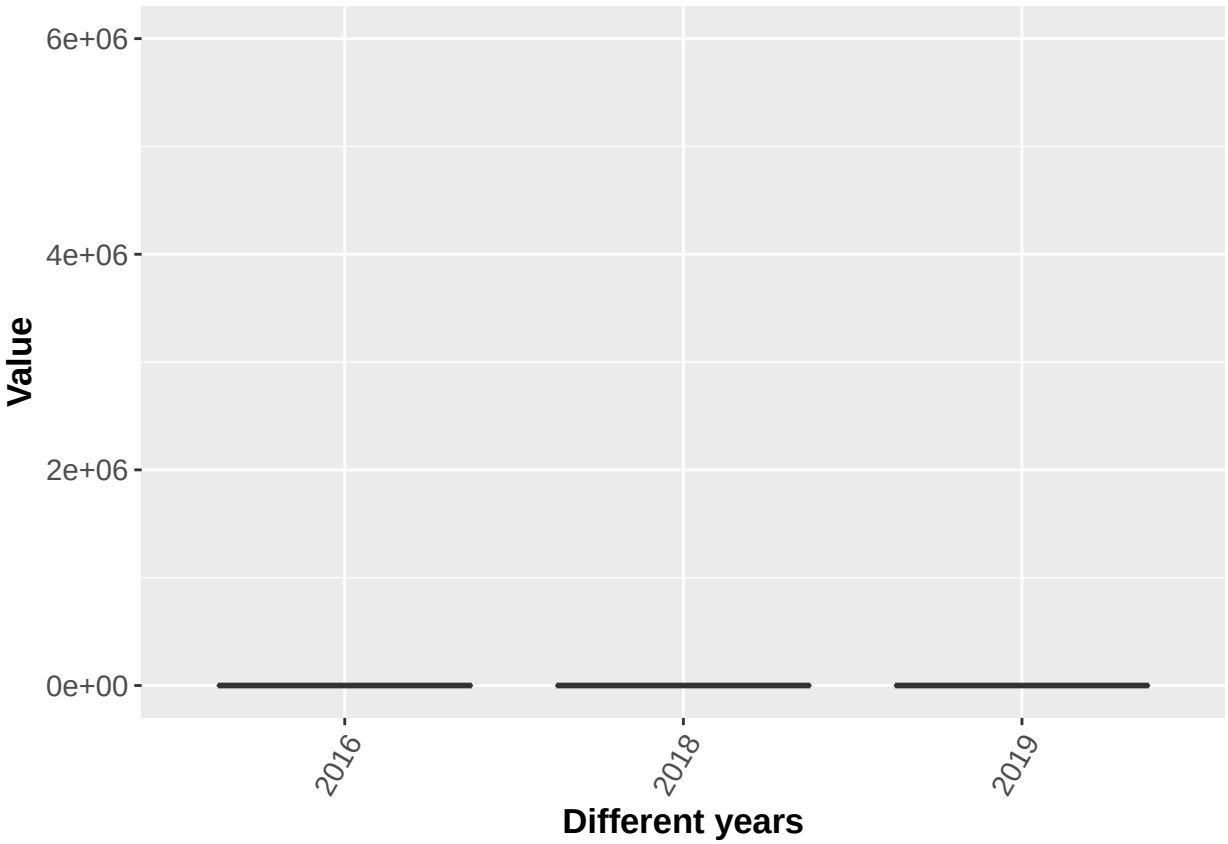


From the plot, we can see that in most states, the chemical values of straberries are quite small. Only in Ohio, Pennsylvania, Wisconsin and other states, the chemical value is high.

2. Then we make a box plot of chemical values being applied to strawberries according to years.



Similarly, the box plot of years also have many outliers so we remove them and get another bow plot.



Within the years 2015-2019, chemical values of strawberries is quite low in 2016, 2018 and 2019. It has the highest values in 2015 and year 2017 gets the second rank.

Citation

- David Morison (2017) Exploratory data analysis into the relationship between different types of crime in London. Available at: https://towardsdatascience.com/exploratory-data-analysis-into-the-relationship-between-different-types-of-crime-in-london-20c328e193ff (Accessed: 15 Oct 2020).
- United States Department of Agriculture National Agricultural Statistics Service (n.p.) Quick Stats. Available at: https://quickstats.nass.usda.gov/results/D416E96E-3D5C-324C-9334-1D38DF88FFF1 (Accessed: 15 Oct 2020).


```
## {
##   "title": "knitr: A General-Purpose Package for Dynamic Report Generation in R",
##   "version": "1.29",
##   "description": "<p>Provides a general-purpose tool for dynamic report generation in R\n    using L
##   "creators": [
##     {
##       "name": "Xie, Yihui"
##     },
##     {
##       "name": "Vogt, Adam"
##     },
##     {
##       "name": "Andrew, Alastair"
##     },
##     {
##       "name": "Zvoleff, Alex"
```

```
##     },
##     {
##       "name": "Simon, Andre"
##     },
##     {
##       "name": "Atkins, Aron"
##     },
##     {
##       "name": "Wolen, Aaron"
##     },
##     {
##       "name": "Manton, Ashley"
##     },
##     {
##       "name": "Yasumoto, Atsushi"
##     },
##     {
##       "name": "Baumer, Ben"
##     },
##     {
##       "name": "Diggs, Brian"
##     },
##     {
##       "name": "Zhang, Brian"
##     },
##     {
##       "name": "Pereira, Cassio"
##     },
##     {
##       "name": "Dervieux, Christophe"
##     },
##     {
##       "name": "Hugh-Jones, David"
##     },
##     {
##       "name": "Robinson, David"
##     },
##     {
##       "name": "Hemken, Doug"
##     },
##     {
##       "name": "Murdoch, Duncan"
##     },
##     {
##       "name": "Campitelli, Elio"
##     },
##     {
##       "name": "Hughes, Ellis"
##     },
##     {
##       "name": "Riederer, Emily"
##     },
##     {
##       "name": "Hirschmann, Fabian"
```

```
##     },
##     {
##       "name": "Simeon, Fitch"
##     },
##     {
##       "name": "Fang, Forest"
##     },
##     {
##       "name": "Frank E Harrell Jr"
##     },
##     {
##       "name": "Aden-Buie, Garrick"
##     },
##     {
##       "name": "Detrez, Gregoire"
##     },
##     {
##       "name": "Wickham, Hadley"
##     },
##     {
##       "name": "Zhu, Hao"
##     },
##     {
##       "name": "Jeon, Heewon"
##     },
##     {
##       "name": "Bengtsson, Henrik"
##     },
##     {
##       "name": "Yutani, Hiroaki"
##     },
##     {
##       "name": "Lyttle, Ian"
##     },
##     {
##       "name": "Daniel, Hodges"
##     },
##     {
##       "name": "Burkhead, Jake"
##     },
##     {
##       "name": "Manton, James"
##     },
##     {
##       "name": "Lander, Jared"
##     },
##     {
##       "name": "Punyon, Jason"
##     },
##     {
##       "name": "Luraschi, Javier"
##     },
##     {
##       "name": "Arnold, Jeff"
```

```
##     },
##     {
##       "name": "Bryan, Jenny"
##     },
##     {
##       "name": "Ashkenas, Jeremy"
##     },
##     {
##       "name": "Stephens, Jeremy"
##     },
##     {
##       "name": "Hester, Jim"
##     },
##     {
##       "name": "Cheng, Joe"
##     },
##     {
##       "name": "Ranke, Johannes"
##     },
##     {
##       "name": "Honaker, John"
##     },
##     {
##       "name": "Muschelli, John"
##     },
##     {
##       "name": "Keane, Jonathan"
##     },
##     {
##       "name": "Allaire, JJ"
##     },
##     {
##       "name": "Toloe, Johan"
##     },
##     {
##       "name": "Sidi, Jonathan"
##     },
##     {
##       "name": "Larmarange, Joseph"
##     },
##     {
##       "name": "Barnier, Julien"
##     },
##     {
##       "name": "Zhong, Kaiyin"
##     },
##     {
##       "name": "Slowikowski, Kamil"
##     },
##     {
##       "name": "Forner, Karl"
##     },
##     {
##       "name": "Smith, Kevin K."
```

```
##     },
##     {
##       "name": "Mueller, Kirill"
##     },
##     {
##       "name": "Takahashi, Kohske"
##     },
##     {
##       "name": "Walthert, Lorenz"
##     },
##     {
##       "name": "Gallindo, Lucas"
##     },
##     {
##       "name": "Hofert, Marius"
##     },
##     {
##       "name": "Modrák, Martin"
##     },
##     {
##       "name": "Chirico, Michael"
##     },
##     {
##       "name": "Friendly, Michael"
##     },
##     {
##       "name": "Bojanowski, Michal"
##     },
##     {
##       "name": "Kuhlmann, Michel"
##     },
##     {
##       "name": "Patrick, Miller"
##     },
##     {
##       "name": "Caballero, Nacho"
##     },
##     {
##       "name": "Salkowski, Nick"
##     },
##     {
##       "name": "Hansen, Niels Richard"
##     },
##     {
##       "name": "Ross, Noam"
##     },
##     {
##       "name": "Mahdi, Obada"
##     },
##     {
##       "name": "Li, Qiang"
##     },
##     {
##       "name": "Vaidyanathan, Ramnath"
```

```
##     },
##     {
##       "name": "Cotton, Richard"
##     },
##     {
##       "name": "Krzyzanowski, Robert"
##     },
##     {
##       "name": "Francois, Romain"
##     },
##     {
##       "name": "Williamson, Ruaridh"
##     },
##     {
##       "name": "Kostyshak, Scott"
##     },
##     {
##       "name": "Meyer, Sebastian"
##     },
##     {
##       "name": "Brouwer, Sietse"
##     },
##     {
##       "name": "Bernard, Simon de"
##     },
##     {
##       "name": "Rousseau, Sylvain"
##     },
##     {
##       "name": "Wei, Taiyun"
##     },
##     {
##       "name": "Assus, Thibaut"
##     },
##     {
##       "name": "Lamadon, Thibaut"
##     },
##     {
##       "name": "Leeper, Thomas"
##     },
##     {
##       "name": "Mastny, Tim"
##     },
##     {
##       "name": "Torsney-Weir, Tom"
##     },
##     {
##       "name": "Davis, Trevor"
##     },
##     {
##       "name": "Veitas, Viktoras"
##     },
##     {
##       "name": "Zhu, Weicheng"
```

```
##     },
##     {
##       "name": "Wu, Wush"
##     },
##     {
##       "name": "Foster, Zachary"
##     }
##   ]
## }
```

```
## {
##   "title": "tidyverse: Easily Install and Load the 'Tidyverse'",
##   "version": "1.3.0",
##   "description": "<p>The 'tidyverse' is a set of packages that work in\n    harmony because they sha
##   "creators": [
##     {
##       "name": "Wickham, Hadley"
##     },
##     {
##       "name": "RStudio"
##     }
##   ]
## }
```

```
## {
##   "title": "kableExtra: Construct Complex Table with 'kable' and Pipe Syntax",
##   "version": "1.2.1",
##   "description": "<p>Build complex HTML or 'LaTeX' tables using 'kable()' from 'knitr' \n    and the
##   "creators": [
##     {
##       "name": "Zhu, Hao"
##     },
##     {
##       "name": "Travison, Thomas"
##     },
##     {
##       "name": "Tsai, Timothy"
##     },
##     {
##       "name": "Beasley, Will"
##     },
##     {
##       "name": "Xie, Yihui"
##     },
##     {
##       "name": "Yu, GuangChuang"
##     },
##     {
##       "name": "Laurent, Stéphane"
##     },
##     {
##       "name": "Shepherd, Rob"
##     },
##     {
##       "name": "Sidi, Yoni"
```

```
##     },
##     {
##       "name": "Salzer, Brian"
##     },
##     {
##       "name": "Gui, George"
##     },
##     {
##       "name": "Fan, Yeliang"
##     },
##     {
##       "name": "Murdoch, Duncan"
##     }
##   ]
## }
```

```
## {
##   "title": "citation: Software Citation Tools",
##   "version": "0.4.1",
##   "description": "<p>A collection of functions to extract citation information from 'R' packages and
##   "creators": [
##     {
##       "name": "Dietrich, Jan Philipp"
##     }
##   ]
## }
```