Strawberries

bcm

10/13/2020

(1) Read and examine

These 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.

The data has 21 columns.

```
##  [1] "Program"          "Year"             "Period"           "Week Ending"
##  [5] "Geo Level"        "State"            "State ANSI"       "Ag District"
##  [9] "Ag District Code" "County"           "County ANSI"      "Zip Code"
## [13] "Region"           "watershed_code"   "Watershed"        "Commodity"
## [17] "Data Item"        "Domain"           "Domain Category"  "Value"
## [21] "CV (%)"
```

(2) Remove the columns that only had NAs

This leaves 12 columns.

```
##  [1] "Program"         "Year"            "Period"          "Geo Level"
##  [5] "State"           "State ANSI"      "watershed_code"  "Commodity"
##  [9] "Data Item"       "Domain"          "Domain Category" "Value"
```

(3) Remove the columns that provide no new information

"Program" and "Geo Level" have only 1 value. And "State ANSI" is a duplicate of "State"

This leaves 8 columns.

```
## [1] "Year"            "Period"          "State"           "Commodity"
## [5] "Data Item"       "Domain"          "Domain Category" "Value"
```

(4) Select the rows that contain "Strawberry" & only the rows where the period = "YEAR" and eliminate the Period and Commodity column.

NOTE: The Period column has three values: "MARKETING YEAR", "YEAR", and "YEAR - AUG FORECAST." We are only keeping the records where Period = "YEAR" so that we have a consistent comparison.

This leaves 6 columns.

```
## [1] "Year"            "State"           "Data Item"       "Domain"
## [5] "Domain Category" "Value"
```

(5)