615HW5

Chang Lu

2024-10-05

```
# Load the dataset
strawberry_data <- read.csv("strawberries25_v3.csv")

# Splitting the 'Domain' column into separate components
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)

# Use separate to split the Domain column into "Chemical_Type", "Chemical_Name", and "Chemical_Code"
strawberry_data_cleaned <- strawberry_data %>%
  separate(Domain, into = c("Chemical_Type", "Chemical_Name"), sep = ", ", extra = "merge", fill = "rig
  separate(Chemical_Name, into = c("Chemical_Name", "Chemical_Code"), sep = " = ", fill = "right")

# If necessary, you can filter rows to remove non-chemical entries if they are not relevant:
strawberry_data_cleaned <- strawberry_data_cleaned %>%
  filter(!is.na(Chemical_Type)) # Retain only rows where Chemical_Type is not NA

# Check the cleaned dataset
head(strawberry_data_cleaned)
```

```
##   Program Year Period Week.Ending Geo.Level   State State.ANSI Ag.District
## 1  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 2  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 3  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 4  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 5  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 6  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
##   Ag.District.Code  County County.ANSI Zip.Code Region watershed_code Watershed
## 1               40 BULLOCK          11       NA     NA              0         NA
```

```
## 2               40 BULLOCK          11       NA     NA              0        NA
## 3               40 BULLOCK          11       NA     NA              0        NA
## 4               40 BULLOCK          11       NA     NA              0        NA
## 5               40 BULLOCK          11       NA     NA              0        NA
## 6               40 BULLOCK          11       NA     NA              0        NA
##      Commodity                                    Data.Item Chemical_Type
## 1 STRAWBERRIES                   STRAWBERRIES - ACRES BEARING         TOTAL
## 2 STRAWBERRIES                     STRAWBERRIES - ACRES GROWN         TOTAL
## 3 STRAWBERRIES                 STRAWBERRIES - ACRES NON-BEARING         TOTAL
## 4 STRAWBERRIES     STRAWBERRIES - OPERATIONS WITH AREA BEARING         TOTAL
## 5 STRAWBERRIES       STRAWBERRIES - OPERATIONS WITH AREA GROWN         TOTAL
## 6 STRAWBERRIES STRAWBERRIES - OPERATIONS WITH AREA NON-BEARING         TOTAL
##   Chemical_Name Chemical_Code Domain.Category Value CV....
## 1          <NA>          <NA>   NOT SPECIFIED   (D)    (D)
## 2          <NA>          <NA>   NOT SPECIFIED     3   15.7
## 3          <NA>          <NA>   NOT SPECIFIED   (D)    (D)
## 4          <NA>          <NA>   NOT SPECIFIED     1    (L)
## 5          <NA>          <NA>   NOT SPECIFIED     6   52.7
## 6          <NA>          <NA>   NOT SPECIFIED     5   47.6
```

```r
# Save the cleaned dataset
write.csv(strawberry_data_cleaned, "strawberries_cleaned.csv", row.names = FALSE)
```