MA615 strawberry

2024-10-21

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(knitr)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(stringr)
strawberry<-read.csv("cleaned_strawberry_data.csv")
na_summary <- colSums(is.na(strawberry))
strawberry_clean <- strawberry %>% drop_na(Value)
summary(strawberry_clean$Value)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       2       4    4526      18  895054
```

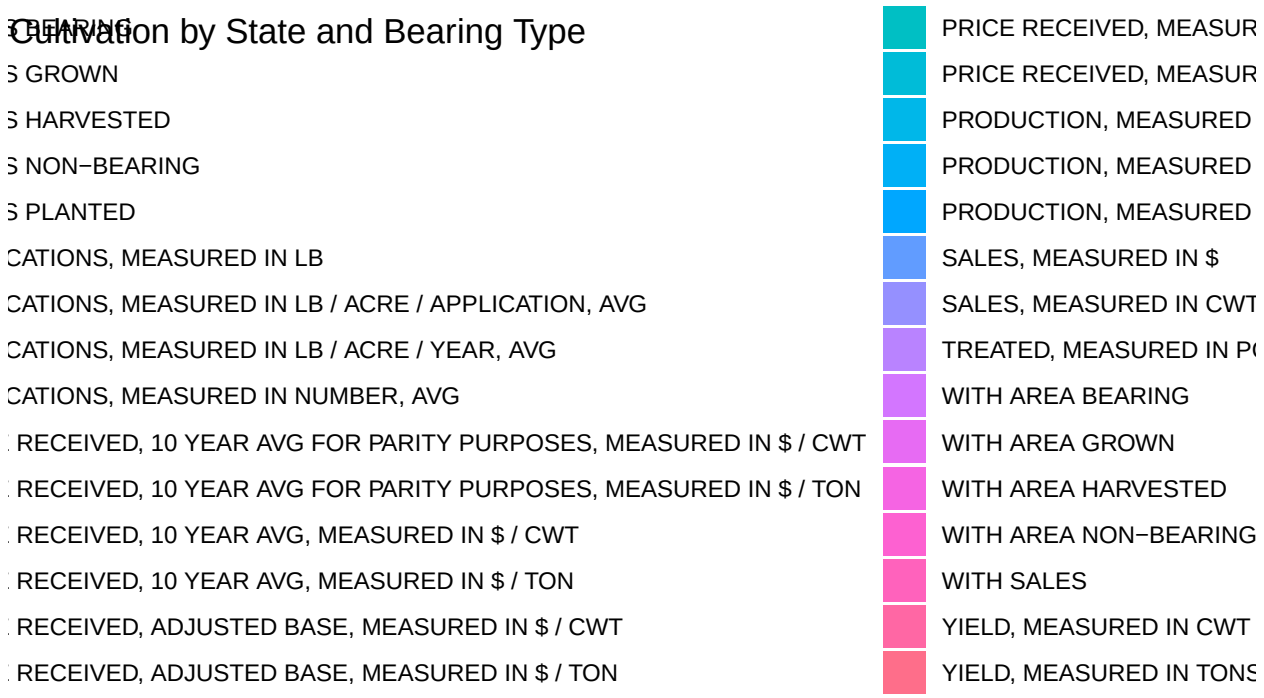
```
state_measure_summary <- strawberry_clean %>%
  group_by(State, Measure, Bearing_type) %>%
  summarise(Total_Value = sum(Value, na.rm = TRUE)) %>%
  arrange(desc(Total_Value))
```

```
## `summarise()` has grouped output by 'State', 'Measure'. You can override using
## the `.groups` argument.
```

```r
head(state_measure_summary)
```

```
## # A tibble: 6 x 4
## # Groups:   State, Measure [6]
##   State      Measure Bearing_type                 Total_Value
##   <chr>      <chr>   <chr>                              <dbl>
## 1 CALIFORNIA <NA>    APPLICATIONS, MEASURED IN LB    10433500
## 2 FLORIDA    <NA>    APPLICATIONS, MEASURED IN LB     4231300
## 3 WASHINGTON <NA>    SALES, MEASURED IN $             2485043
## 4 OREGON     <NA>    SALES, MEASURED IN $             2295766
## 5 VERMONT    <NA>    SALES, MEASURED IN $             1934348
## 6 NEW YORK   <NA>    SALES, MEASURED IN $             1277266
```

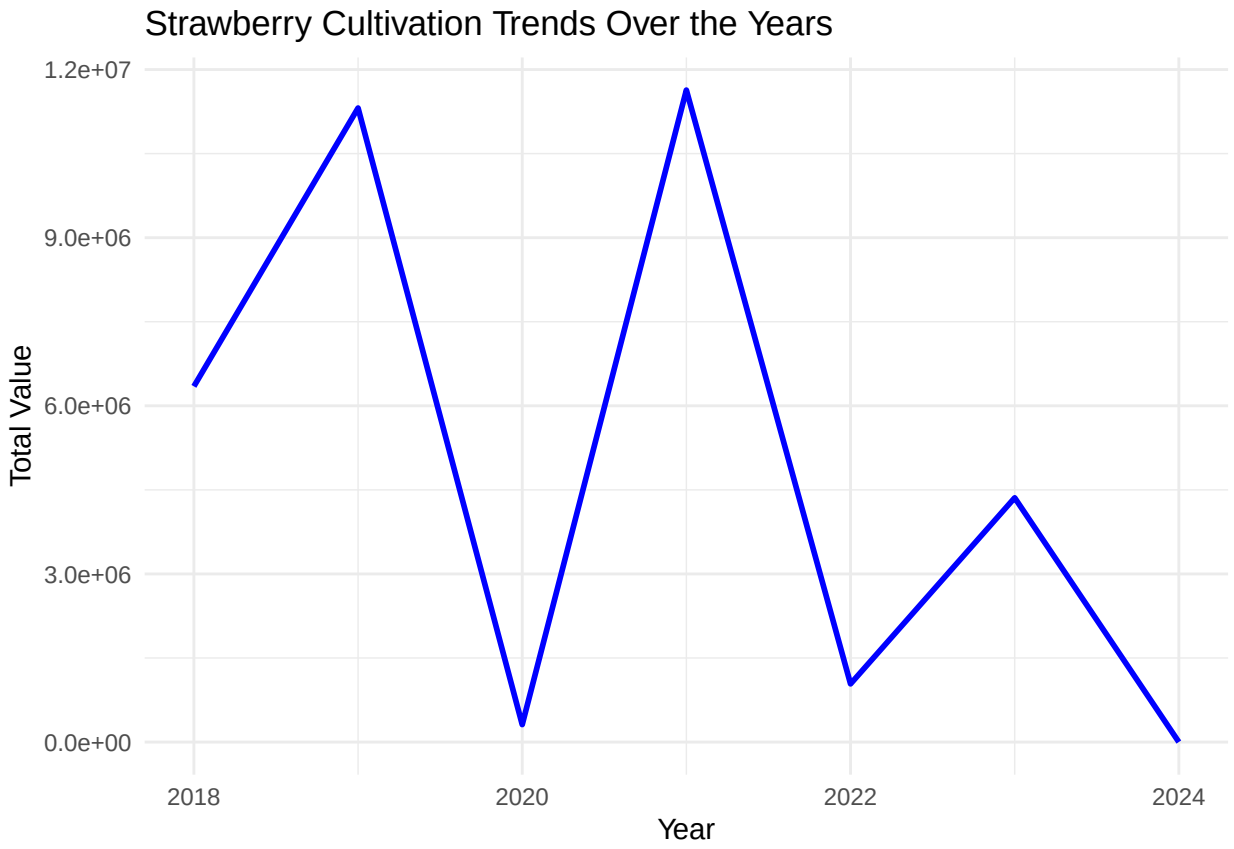
```r
library(ggplot2)
ggplot(state_measure_summary, aes(x = reorder(State, -Total_Value), y = Total_Value, fill = Bearing_type
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Strawberry Cultivation by State and Bearing Type", x = "State", y = "Total Value") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```




```r
yearly_summary <- strawberry_clean %>%
  group_by(Year) %>%
  summarise(Total_Value = sum(Value, na.rm = TRUE))
ggplot(yearly_summary, aes(x = Year, y = Total_Value)) +
  geom_line(color = "blue", size = 1) +
  theme_minimal() +
  labs(title = "Strawberry Cultivation Trends Over the Years", x = "Year", y = "Total Value")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

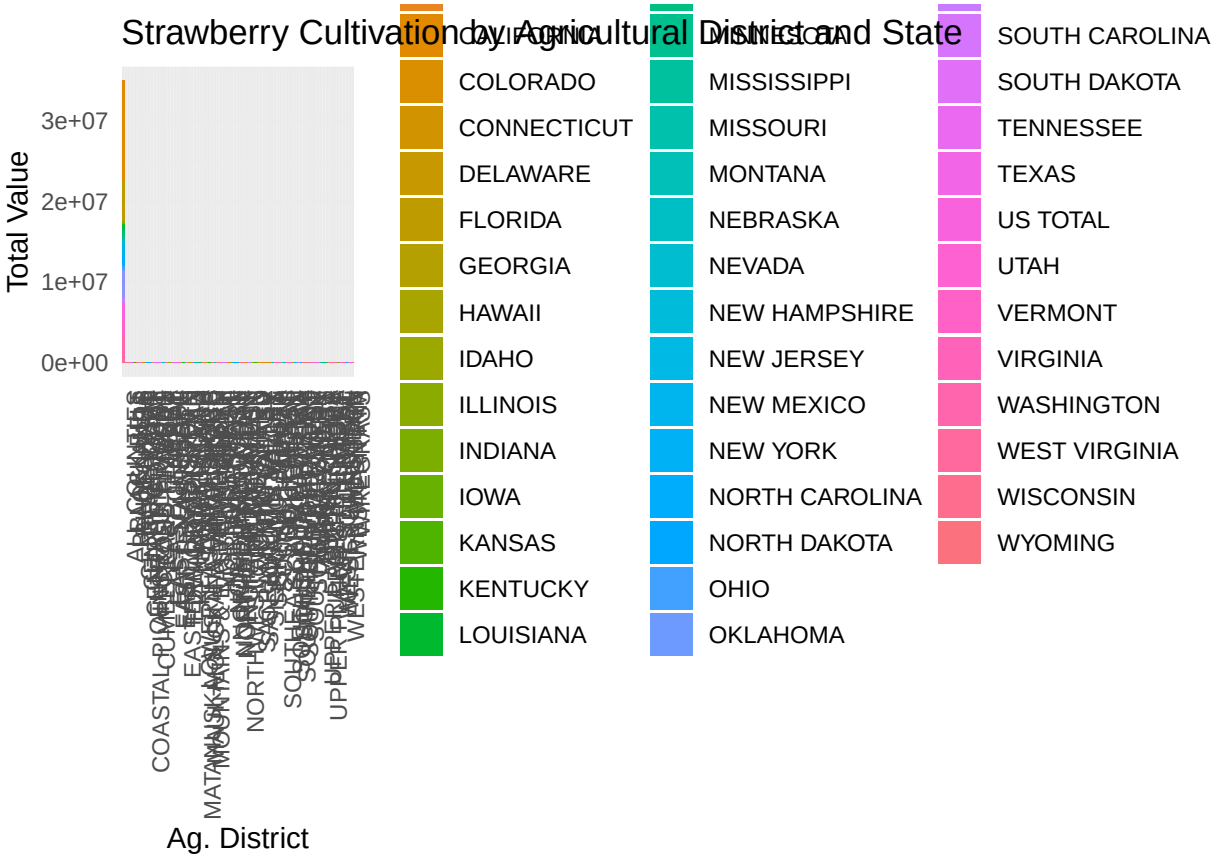
```
## generated.
```




```
district_summary <- strawberry_clean %>%
  group_by(State, Ag.District) %>%
  summarise(Total_Value = sum(Value, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'State'. You can override using the
## `.groups` argument.
```

```
ggplot(district_summary, aes(x = Ag.District, y = Total_Value, fill = State)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Strawberry Cultivation by Agricultural District and State", x = "Ag. District", y = "To
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

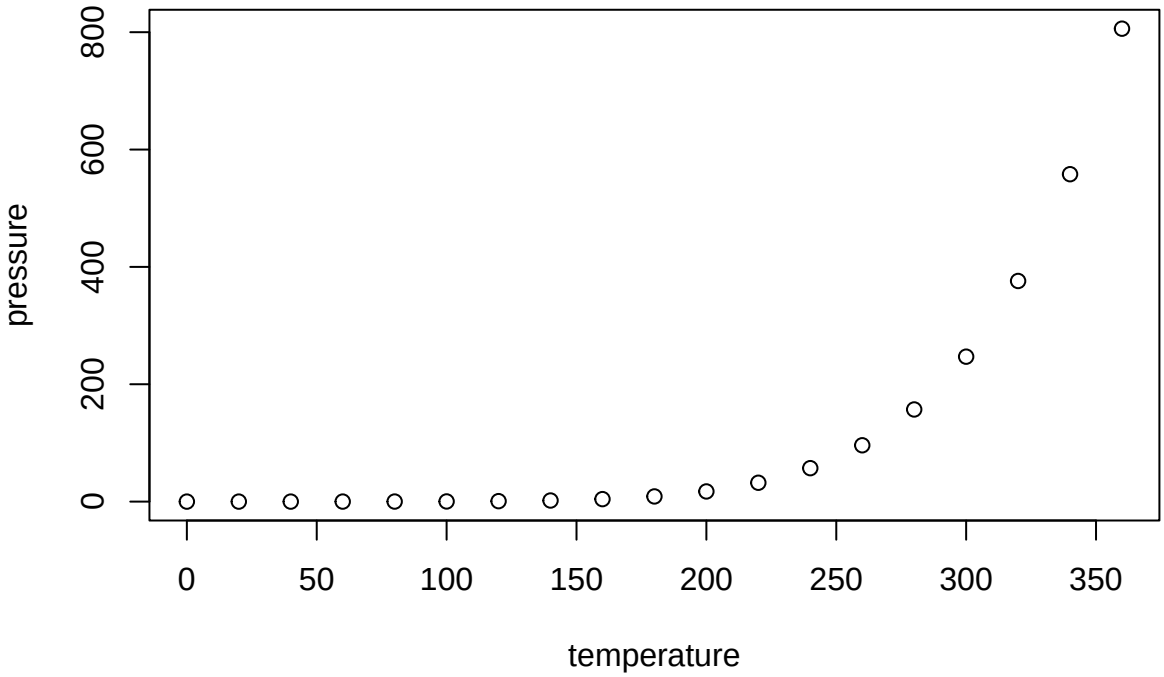


#Conclusion #1. Regional distribution of strawberry planting: As can be seen from the bar chart, there are obvious differences in strawberry planting among different states. Some states have particularly large strawberry planting areas, and the planting characteristics and policy support of these states can be further studied in the future. #2. Changes in planting trends: Strawberry planting area has fluctuated over the past few years. Using the time series graph, we can identify whether there is a cyclical change and further analyze the possible causes, such as climate, market demand, etc. #3. The use of chemical substances: For the use of toxic chemicals, we can see whether the carcinogens listed by WHO are frequently used in strawberry cultivation, which has an important impact on health and the environment.

#New question #Is the trend of strawberry planting area related to climate and policy changes? #Are the differences between different agricultural areas due to natural conditions or differences in growing techniques? #Can climate data or economic data be combined to further analyze factors affecting strawberry cultivation in the future?

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.