Assignment 1

Jinran Li

2024-10-02

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

```
strawberry <- read_csv("strawberries25_v3.csv", col_names = TRUE)
```

```
## Rows: 12669 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (15): Program, Period, Geo Level, State, State ANSI, Ag District, County...
## dbl  (2): Year, Ag District Code
## lgl  (4): Week Ending, Zip Code, Region, Watershed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(strawberry)
```

```
## Rows: 12,669
## Columns: 21
## $ Program            <chr> "CENSUS", "CENSUS", "CENSUS", "CENSUS", "CENSUS", "~
## $ Year               <dbl> 2022, 2022, 2022, 2022, 2022, 2022, 2022, 2022, 202~
## $ Period             <chr> "YEAR", "YEAR", "YEAR", "YEAR", "YEAR", "YEAR", "YE~
## $ `Week Ending`      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ `Geo Level`        <chr> "COUNTY", "COUNTY", "COUNTY", "COUNTY", "COUNTY", "~
## $ State              <chr> "ALABAMA", "ALABAMA", "ALABAMA", "ALABAMA", "ALABAM~
## $ `State ANSI`       <chr> "01", "01", "01", "01", "01", "01", "01", "01", "01~
## $ `Ag District`      <chr> "BLACK BELT", "BLACK BELT", "BLACK BELT", "BLACK BE~
## $ `Ag District Code` <dbl> 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40,~
## $ County             <chr> "BULLOCK", "BULLOCK", "BULLOCK", "BULLOCK", "BULLOC~
## $ `County ANSI`      <chr> "011", "011", "011", "011", "011", "011", "101", "1~
## $ `Zip Code`         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Region             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ watershed_code     <chr> "00000000", "00000000", "00000000", "00000000", "00~
## $ Watershed          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Commodity          <chr> "STRAWBERRIES", "STRAWBERRIES", "STRAWBERRIES", "ST~
## $ `Data Item`        <chr> "STRAWBERRIES - ACRES BEARING", "STRAWBERRIES - ACR~
## $ Domain             <chr> "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL~
## $ `Domain Category`  <chr> "NOT SPECIFIED", "NOT SPECIFIED", "NOT SPECIFIED", ~
## $ Value              <chr> "(D)", "3", "(D)", "1", "6", "5", "(D)", "(D)", "2"~
## $ `CV (%)`           <chr> "(D)", "15.7", "(D)", "(L)", "52.7", "47.6", "(D)",~
```

```r
strawberry <- strawberry %>%
mutate(Value = as.numeric(str_replace_all(Value, "[^0-9\\.]", "")))

str(strawberry$Value)
```

```
##  num [1:12669] NA 3 NA 1 6 5 NA NA 2 2 ...
```

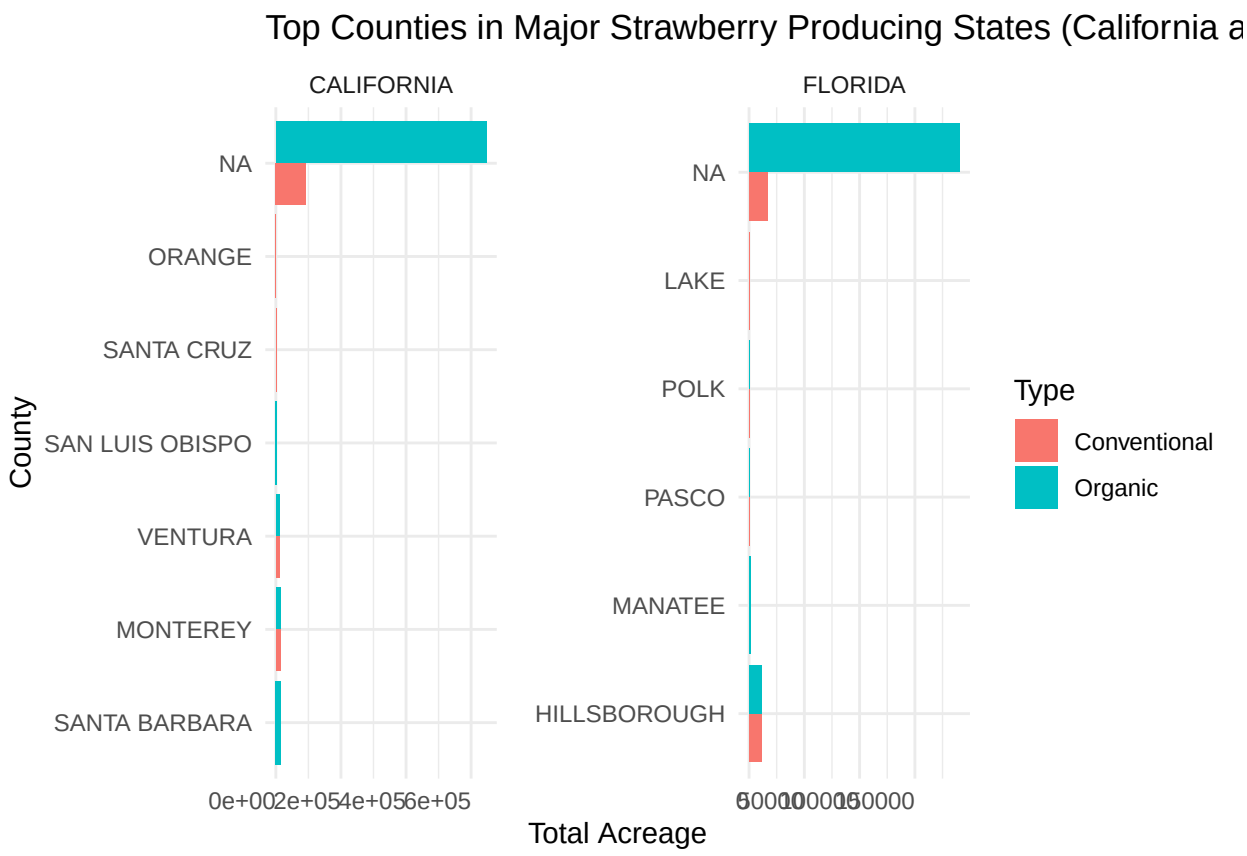
```r
head(strawberry$Value)
```

```
## [1] NA  3 NA  1  6  5
```

```r
state_county_acres <- strawberry %>%
  filter(State %in% c("CALIFORNIA", "FLORIDA")) %>%
  filter(str_detect(`Data Item`, "ACRES"))

state_county_acres <- state_county_acres %>%
  mutate(Type = ifelse(str_detect(`Data Item`, "BEARING"), "Conventional", "Organic"))

state_county_acres <- state_county_acres %>%
  select(State, County, Type, Value) %>%
  mutate(Value = as.numeric(Value))

state_county_acres <- state_county_acres %>%
  group_by(State, County, Type) %>%
  summarize(Total_Acres = sum(Value, na.rm = TRUE)) %>%
  ungroup()
```

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

```r
top2_states_all_types <- state_county_acres %>%
  filter(State %in% c("CALIFORNIA", "FLORIDA")) %>%
  group_by(State, Type) %>%
  top_n(5, wt = Total_Acres) %>%
  ungroup()

# plot
ggplot(top2_states_all_types, aes(x = reorder(County, -Total_Acres), y = Total_Acres, fill = Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  coord_flip() +  # flip
  facet_wrap(~ State, scales = "free") +
  labs(title = "Top Counties in Major Strawberry Producing States (California and Florida)",
       x = "County",
       y = "Total Acreage",
       fill = "Type") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 1),
        panel.spacing = unit(1, "lines"))
```




```r
county_acreage_proportion <- state_county_acres %>%
  group_by(State) %>%
  mutate(State_Total_Acres = sum(Total_Acres, na.rm = TRUE)) %>%
  ungroup() %>%
  mutate(Percentage = (Total_Acres / State_Total_Acres) * 100) %>%
  select(State, County, Type, Total_Acres, State_Total_Acres, Percentage)
```

```r
# acreage proportion
county_acreage_proportion <- county_acreage_proportion %>%
  filter(County %in% c("SANTA BARBARA", "VENTURA", "MIAMI-DADE", "HILLSBOROUGH", "PASCO", "POLK", "ORAN

#
ggplot(county_acreage_proportion, aes(x = reorder(County, -Percentage), y = Percentage, fill = County))
  geom_bar(stat = "identity") +
  coord_flip() +  # flip
  facet_wrap(~ State, scales = "free") +
  labs(title = "Proportion of Strawberry Acreage by Major Counties in California and Florida",
       x = "County",
       y = "Proportion (%)",
       fill = "County") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5))
```

Proportion of Strawberry Acreage by Major Counties in California and Florida