Berries

Data analysis for data "berries"

Ruiyi Lian

Data Cleaning

1. Data Clean

Load data.

```
dtRaw <- fread("C:/Users/user/Desktop/1900_20201018/berries(3)(1).csv")
dt <- dtRaw
```

Convert value from character to numeric.

```
dt$Value <- str_remove_all(dt$Value,",")
dt$Value <- as.numeric(dt$Value)
```

Distinguish different method of measure.
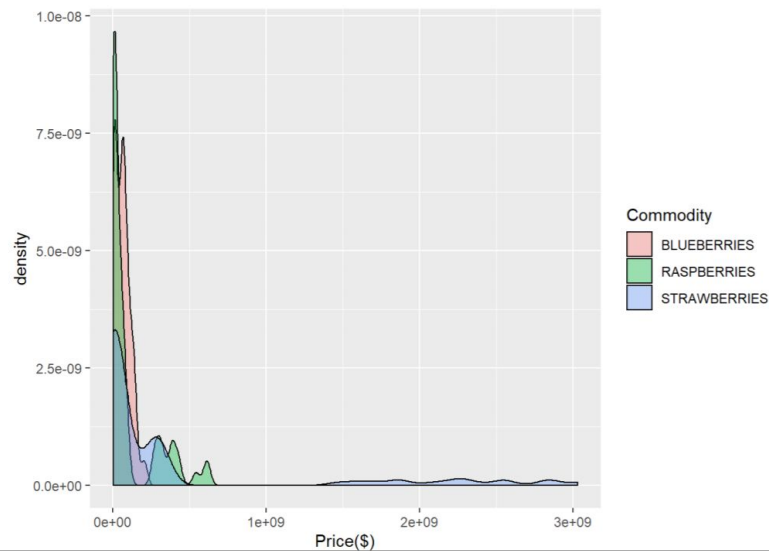
```
dt$measure <- sub('.*MEASURED IN ', '', dt$`Data Item`)
unique(dt$measure)
```

EDA

- Price of Production
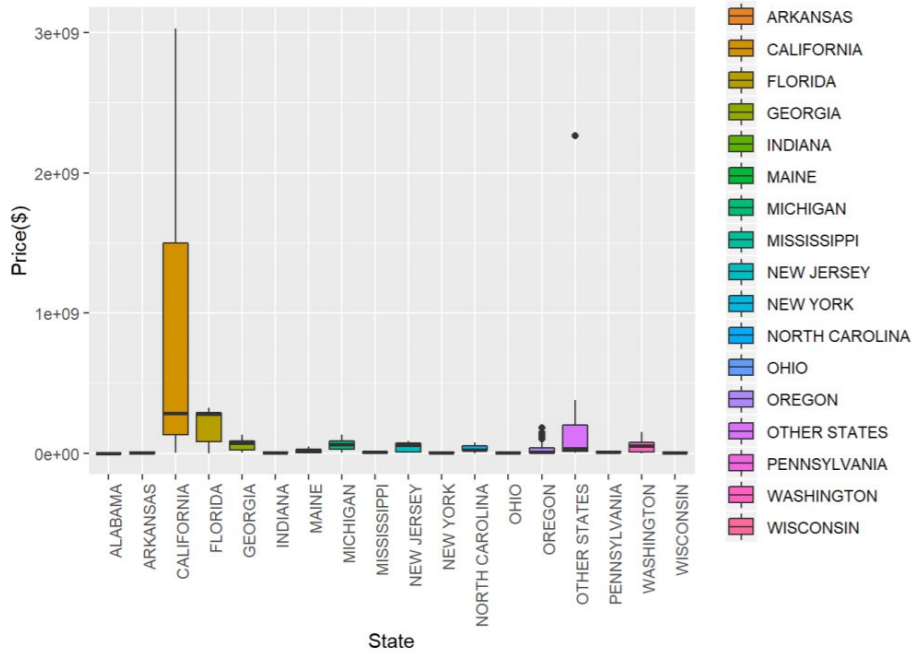
- Price of procession

- Gross Product of Production

Price of Production
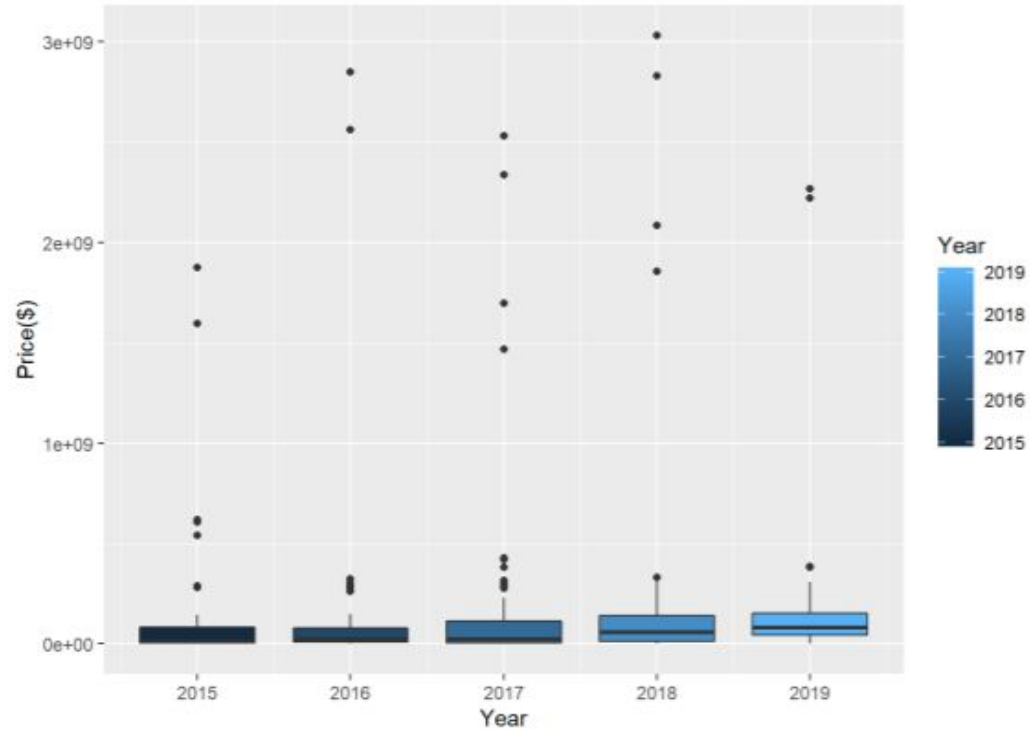
- Affect of Commodity

Price of Production

- Affect of State

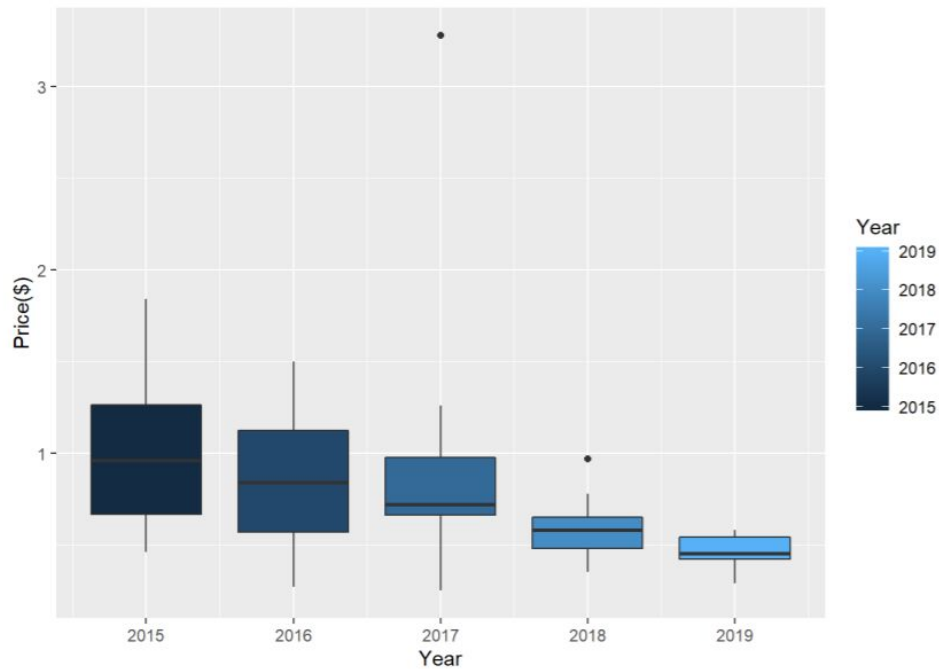# Price of Production
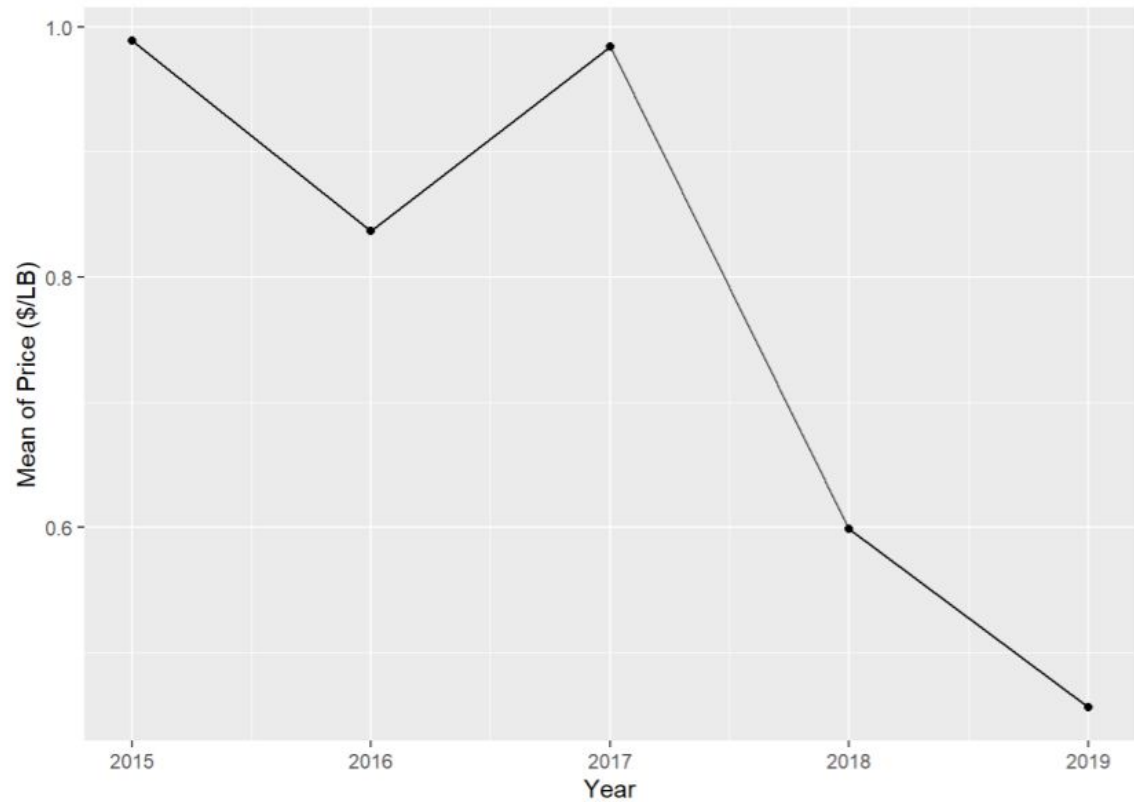
- Affect of Year

Price of procession

- Affect of Year

```
ggplot(procession, aes(x=Year, y=Value, group=Year, fill=Year))+
   geom_boxplot()+
   labs(y="Price($)")
```
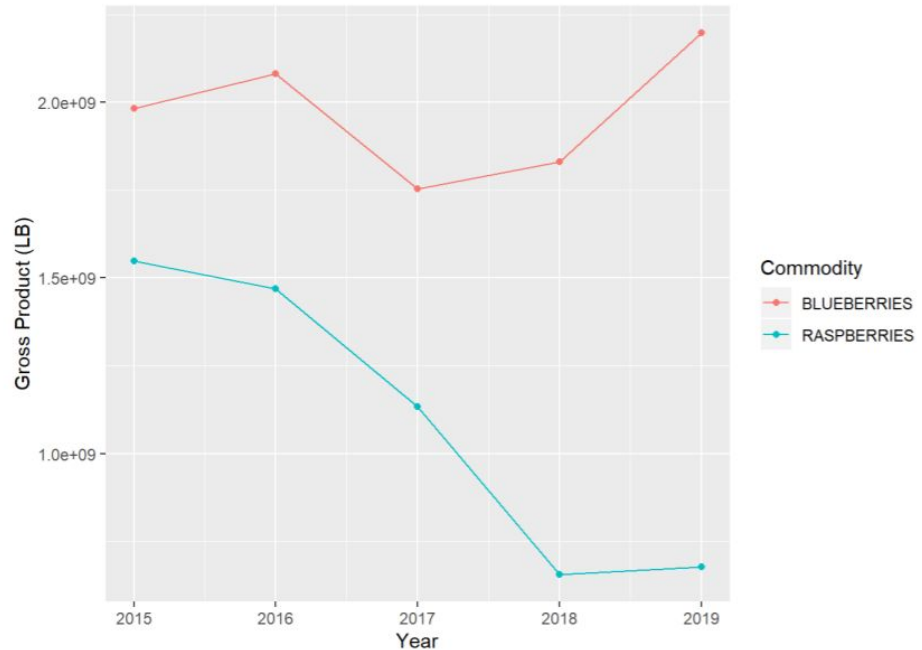
Price of procession

- Trend

Gross Product of Production

Change of Gross Product of Procession

Gross Product of Production

Change of Raspberries in Different States

```
rasp <- gpprd[gpprd$Commodity=="RASPBERRIES"][, (sum(Value, na.rm = T)),
                                                by=c("Year", "State")]
rasp <- dcast(rasp, Year~State, value.var = "V1")
rasp
```

```
##    Year CALIFORNIA    OREGON OTHER STATES WASHINGTON
## 1: 2015 1122680000 56330000           NA  369975000
## 2: 2016 1017680000 51240000           NA  400250000
## 3: 2017  715950000 26720000           NA  389940000
## 4: 2018  286000000       NA    217320000  151600000
## 5: 2019  287000000       NA    225840000  165000000
```

Shiny

```r
# Define UI ----
ui <- fluidP  verticalLayout(..., fluid = TRUE)
  titlePanel("Analysis of Berries Dataset"),
  verticalLayout(
    mainPanel(
      h3("1. Price of Production"),
      h4("1.2 Affect of Commodity"),
      h5("Confirm whether price of production were affect by commodit
      plotOutput("p1"),
      plotOutput("p2"),
```

```r
# Define server logic ----
server <- function(input, output) {
  output$p1 <- renderPlot({
    ggplot(production,aes(x=Commodity,y=Value,fill=Commodity))+
      geom_boxplot()+
      labs(y="Price($)"
  })
  output$p2 <- renderPlot({
    ggplot(production,aes(x=Value,group=Commodity,fill=Commodity))+
```

renderPlot(expr, width = "auto", height = "auto", res = 72, ..., env = parent.fra
quoted = FALSE, execOnResize = FALSE, outputArgs = list())