615_Assignment3_Honey

Dave Anderson

October 1, 2018

```
Honey <- read_csv("Honey.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   Year = col_integer(),
##   `State ANSI` = col_integer(),
##   watershed_code = col_integer(),
##   Value = col_number()
## )
```

```
## See spec(...) for full column specifications.
```

```
Deadout <- read_csv("Deadout.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   Year = col_integer(),
##   `State ANSI` = col_integer(),
##   watershed_code = col_integer(),
##   Value = col_number()
## )
## See spec(...) for full column specifications.
```

```
Price_per_lb <- read_csv("Price per lb.csv")
```

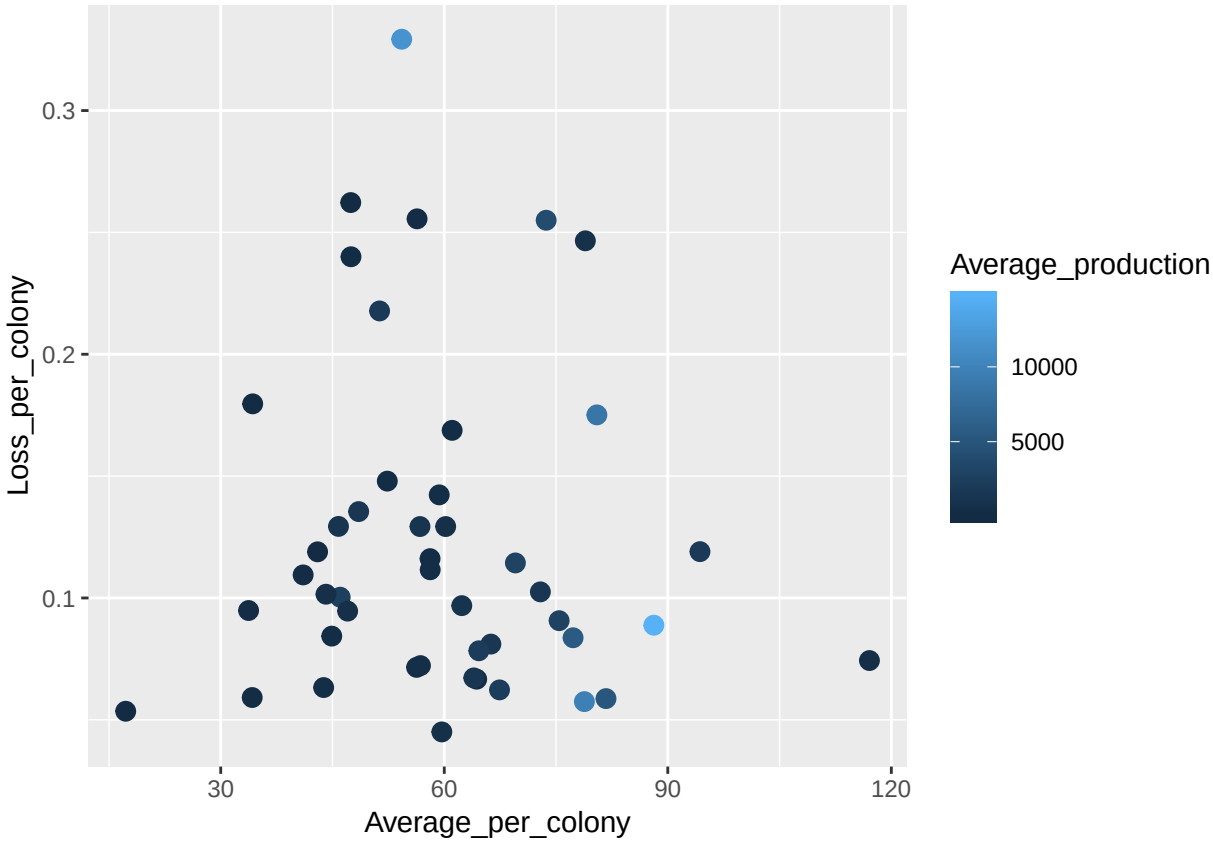
```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   Year = col_integer(),
##   `State ANSI` = col_integer(),
##   watershed_code = col_integer(),
##   Value = col_double()
## )
## See spec(...) for full column specifications.
```

```
Production_per_Colony <- read_csv("Production per Colony.csv")
```

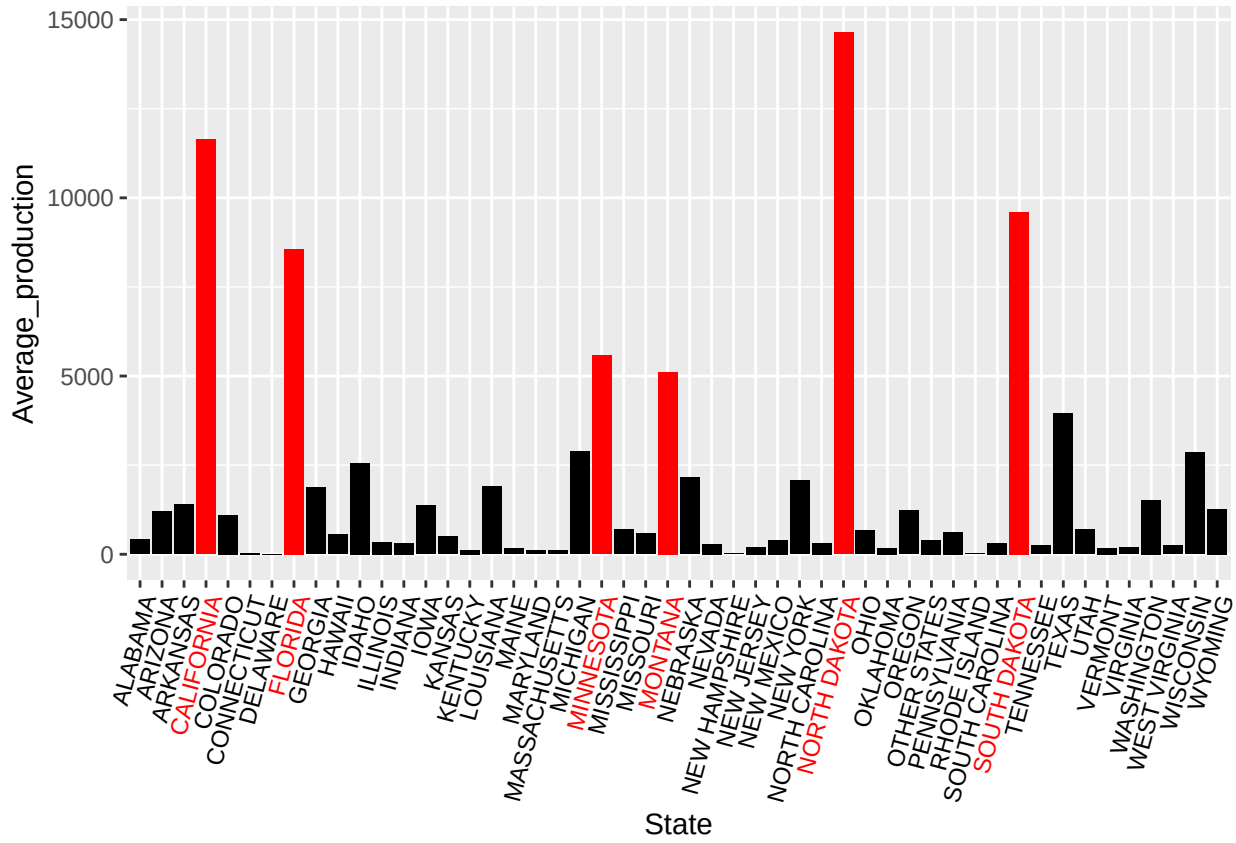
```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   Year = col_integer(),
##   `State ANSI` = col_integer(),
##   watershed_code = col_integer(),
##   Value = col_double()
## )
## See spec(...) for full column specifications.
```

```r
#sort each data set to variables we want.
Honey <- dplyr::select(Honey,Year,State,Value)
Deadout <- dplyr::select(Deadout,Year,Period,State,Value)
Price_per_lb <- dplyr::select(Price_per_lb,Year,State,Value)
Production_per_Colony <- dplyr::select(Production_per_Colony,Year,State,Value)



#Filter out totals, group each variable by state, average values from each year
Production <- filter(Honey, State != "US TOTAL") %>% group_by(State) %>% summarise(Average_production =
Loss <- filter(Deadout, State != "US TOTAL") %>% group_by(State) %>% summarise(Average_loss = mean(Valu
Price <- filter(Price_per_lb, State != "US TOTAL") %>% group_by(State) %>% summarise(Average_price = mea
Colony_production <- filter(Production_per_Colony, State != "US TOTAL") %>% group_by(State) %>% summaris

#Combine into one set. Add new variables to show number of colonies and loss/colony
Honey_by_State <- full_join(Production, Loss, by = "State")
Honey_by_State <- full_join(Honey_by_State, Price, by = "State")
Honey_by_State <- full_join(Honey_by_State, Colony_production, by = "State") %>%
  mutate(Colonies = Average_production*2000/Average_per_colony) %>%
  mutate(Loss_per_colony = Average_loss/Colonies)


#Scatterplot to show relationship between productivity and loss by state.
#Colored to show overall high-producing states.
ggplot(data = Honey_by_State, mapping = aes(Average_per_colony,Loss_per_colony))+
  geom_point(aes(color = Average_production),size = 3)
```

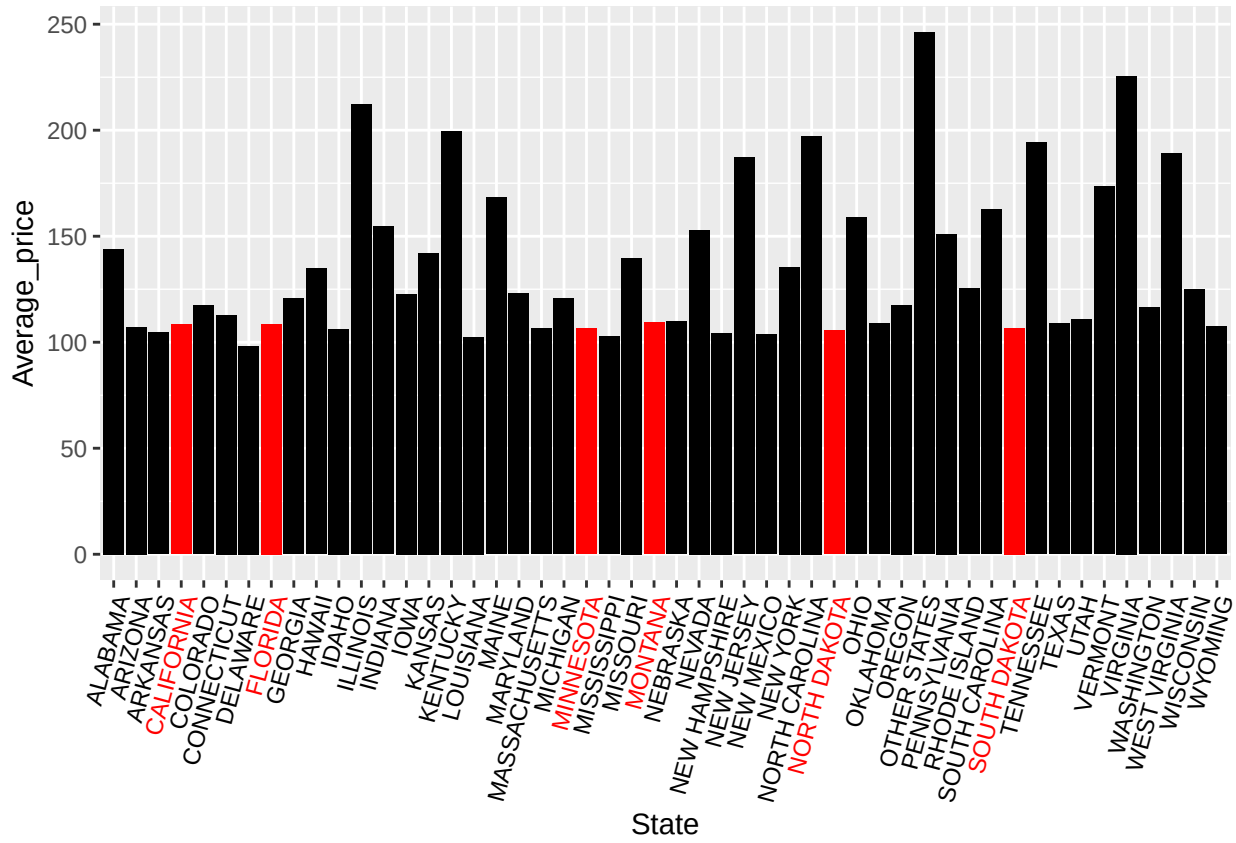
```
## Warning: Removed 4 rows containing missing values (geom_point).
```




```
#Total Production by state, largest six states indicated in red.
ggplot(data = Honey_by_State, mapping = aes(State, Average_production, fill=ifelse(Average_production >
  geom_col()+
  scale_fill_manual(guide=FALSE, values=c("red", "black"))+
  theme(axis.text.x = element_text(color = ifelse(Honey_by_State$Average_production > 5000, "red", "bla
```



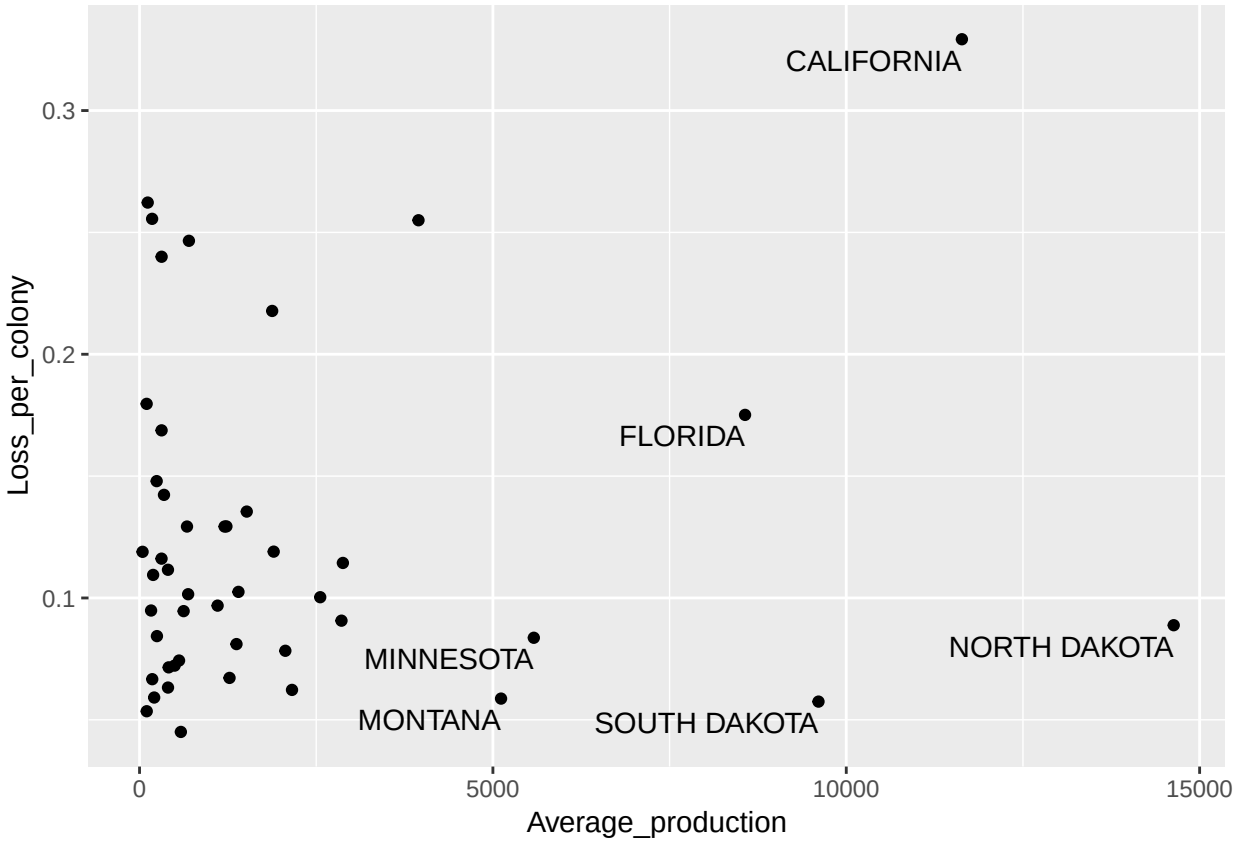

```
#Display of prices per lb. by state. Top six producing states still in red to show their low prices.
ggplot(data = Honey_by_State, mapping = aes(State ,Average_price, fill = ifelse(Average_production > 500
  geom_col()+
  scale_fill_manual(guide=FALSE, values=c("red", "black"))+
  theme(axis.text.x = element_text(color = ifelse(Honey_by_State$Average_production > 5000, "red", "bla
```



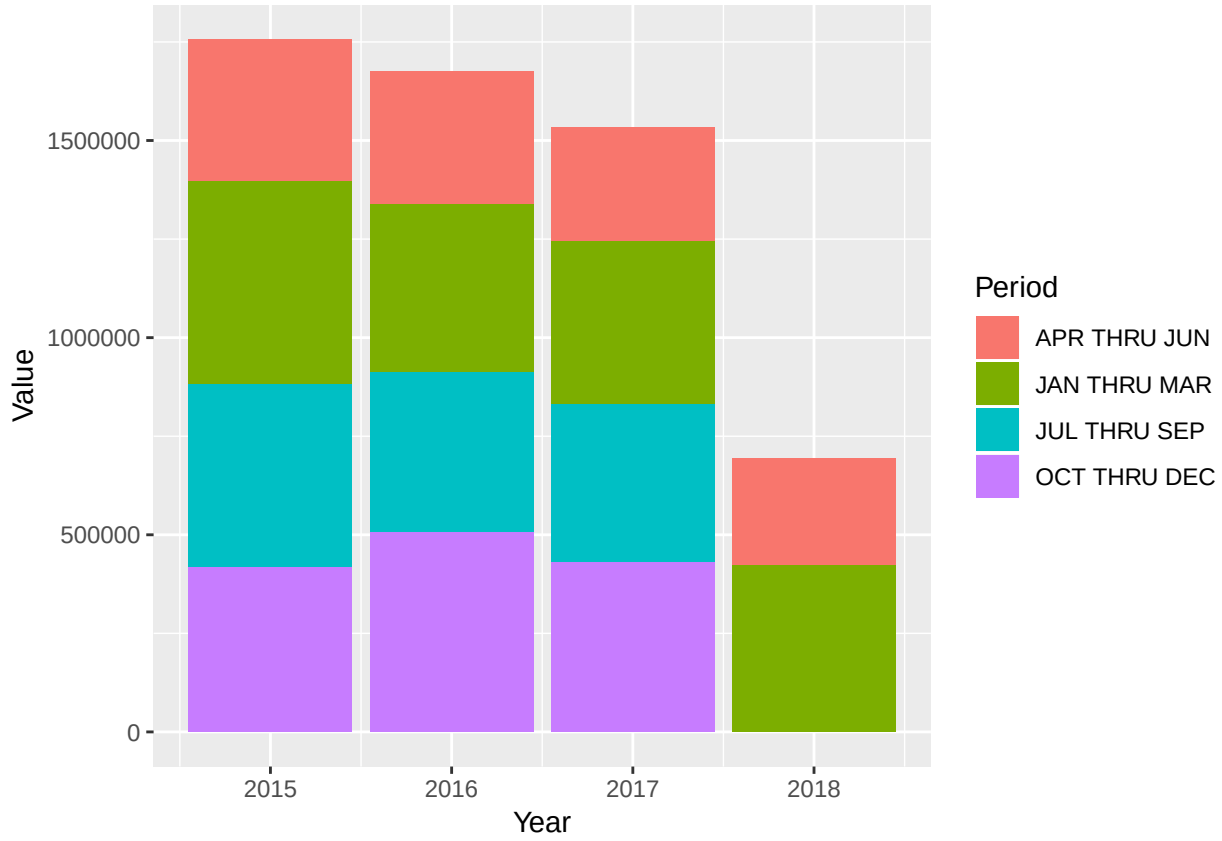

```
#Display of production vs. loss per colony with out top states labeled.
ggplot(data = Honey_by_State, mapping = aes(Average_production,Loss_per_colony, label = State))+
  geom_point() + geom_text(aes(label = ifelse(Average_production > 5000,as.character(State),'')),vjust =
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

```
## Warning: Removed 4 rows containing missing values (geom_text).
```




```
#Display of colonies lost by year and peiod.
loss_by_year <- Deadout %>% filter(State == "US TOTAL") %>% group_by(Year)
ggplot(data = loss_by_year)+
  geom_col(mapping = aes(x = Year, y = Value, fill = Period))
```



It is interesting to see two of our top producers from, big, southern,

costal states while the other 4 are from the midwest.

Looking at prices, we see that the big producers are also among cheapest states.

The two large states have high rates of deadout colonies. The 4 big producers from the midwest have low loss rates.