EDLD652_Lab_PS3

Michelle Cui

2025-02-12

Data

```
honey <- read.csv(here("data","honeyproduction.csv"))
summary(honey)
```

```
##     state              numcol        yieldpercol       totalprod
##  Length:626         Min.   :  2000   Min.   : 19.00   Min.   :   84000
##  Class :character   1st Qu.:  9000   1st Qu.: 48.00   1st Qu.:  475000
##  Mode  :character   Median : 26000   Median : 60.00   Median : 1533000
##                     Mean   : 60284   Mean   : 62.01   Mean   : 4169086
##                     3rd Qu.: 63750   3rd Qu.: 74.00   3rd Qu.: 4175250
##                     Max.   :510000   Max.   :136.00   Max.   :46410000
##      stocks           priceperlb       prodvalue            year
##  Min.   :    8000   Min.   :0.4900   Min.   :  162000   Min.   :1998
##  1st Qu.:  143000   1st Qu.:0.9325   1st Qu.:  759250   1st Qu.:2001
##  Median :  439500   Median :1.3600   Median : 1841500   Median :2005
##  Mean   : 1318859   Mean   :1.4096   Mean   : 4715741   Mean   :2005
##  3rd Qu.: 1489500   3rd Qu.:1.6800   3rd Qu.: 4703250   3rd Qu.:2009
##  Max.   :13800000   Max.   :4.1500   Max.   :69615000   Max.   :2012
```

```
str(honey)
```

```
## 'data.frame':    626 obs. of  8 variables:
##  $ state      : chr  "AL" "AZ" "AR" "CA" ...
##  $ numcol     : num  16000 55000 53000 450000 27000 230000 75000 8000 120000 9000 ...
##  $ yieldpercol: int  71 60 65 83 72 98 56 118 50 71 ...
##  $ totalprod  : num  1136000 3300000 3445000 37350000 1944000 ...
##  $ stocks     : num  159000 1485000 1688000 12326000 1594000 ...
##  $ priceperlb : num  0.72 0.64 0.59 0.62 0.7 0.64 0.69 0.77 0.65 1.19 ...
##  $ prodvalue  : num  818000 2112000 2033000 23157000 1361000 ...
##  $ year       : int  1998 1998 1998 1998 1998 1998 1998 1998 1998 1998 ...
```

Lab Tasks

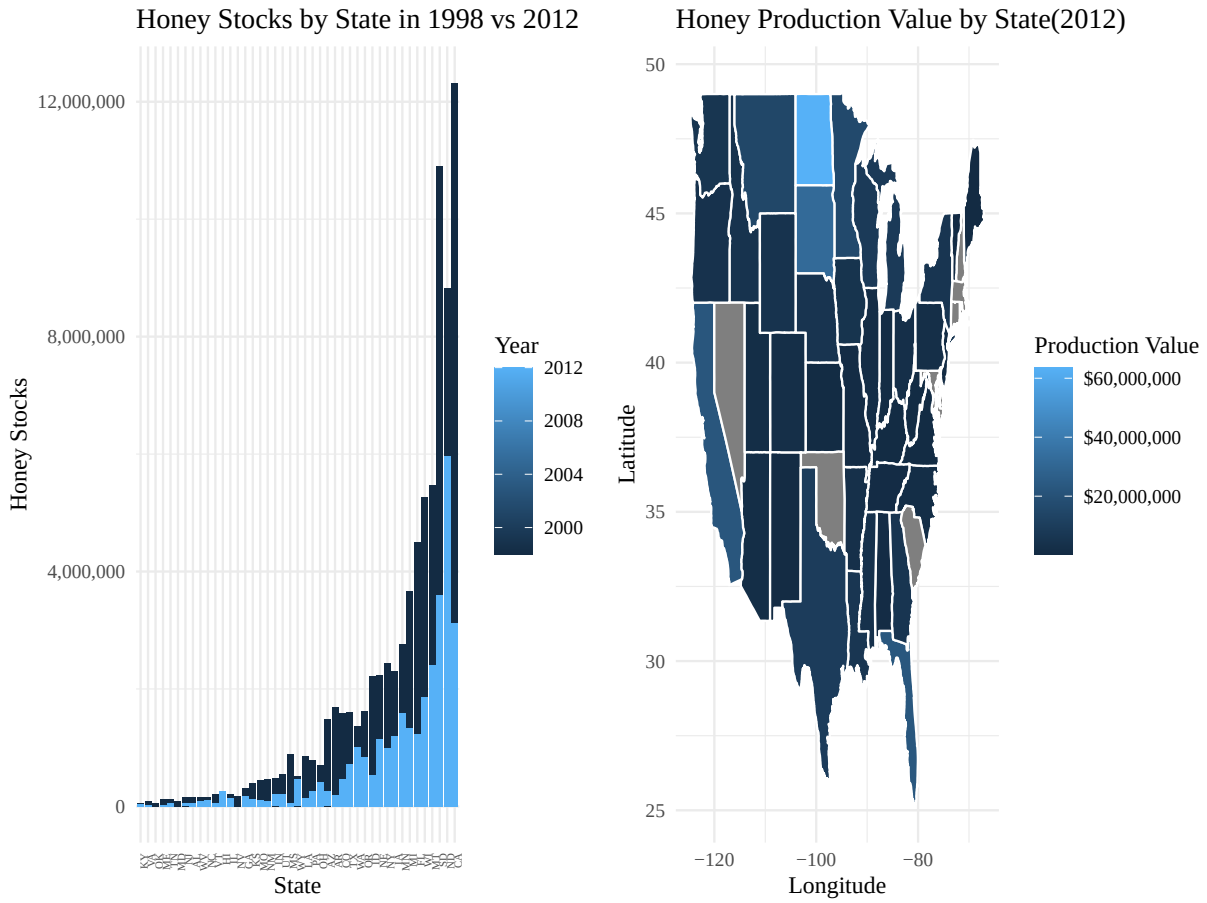
1. Create two plots from the same data. Choose wisely - so that they are two different types of plots (also note: plots with facet_wrap doesn't count here!)

```
honey_98_12 <- honey %>%
  filter(year == 1998 | year == 2012) %>%
  select(state, year, stocks) %>%
  mutate(state = fct_reorder(state, stocks))

p1 <- ggplot(honey_98_12, aes(x = state, y = stocks, fill = year)) +
  geom_col(position = "dodge") +
  labs(title = "Honey Stocks by State in 1998 vs 2012",
```

```
       x = "State",
       y = "Honey Stocks",
       fill = "Year") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 5))+
  scale_y_continuous(labels = label_comma())


honey_2012 <- honey %>%
  filter(year == 2012)

honey_2012 <- honey_2012 %>%
  mutate(state = tolower(state.name[match(state, state.abb)]))

us_states <- map_data("state")

map_data_merged <- us_states %>%
  left_join(honey_2012, by = c("region" = "state"))

p2 <- ggplot(map_data_merged, aes(x = long, y = lat, group = group, fill = prodvalue)) +
  geom_polygon(color = "white") +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90") +
  scale_fill_continuous(name = "Production Value", labels = label_dollar()) +
  labs(title = "Honey Production Value by State(2012)",
       x = "Longitude",
       y = "Latitude",
       fill = "Production Value ($)")
```

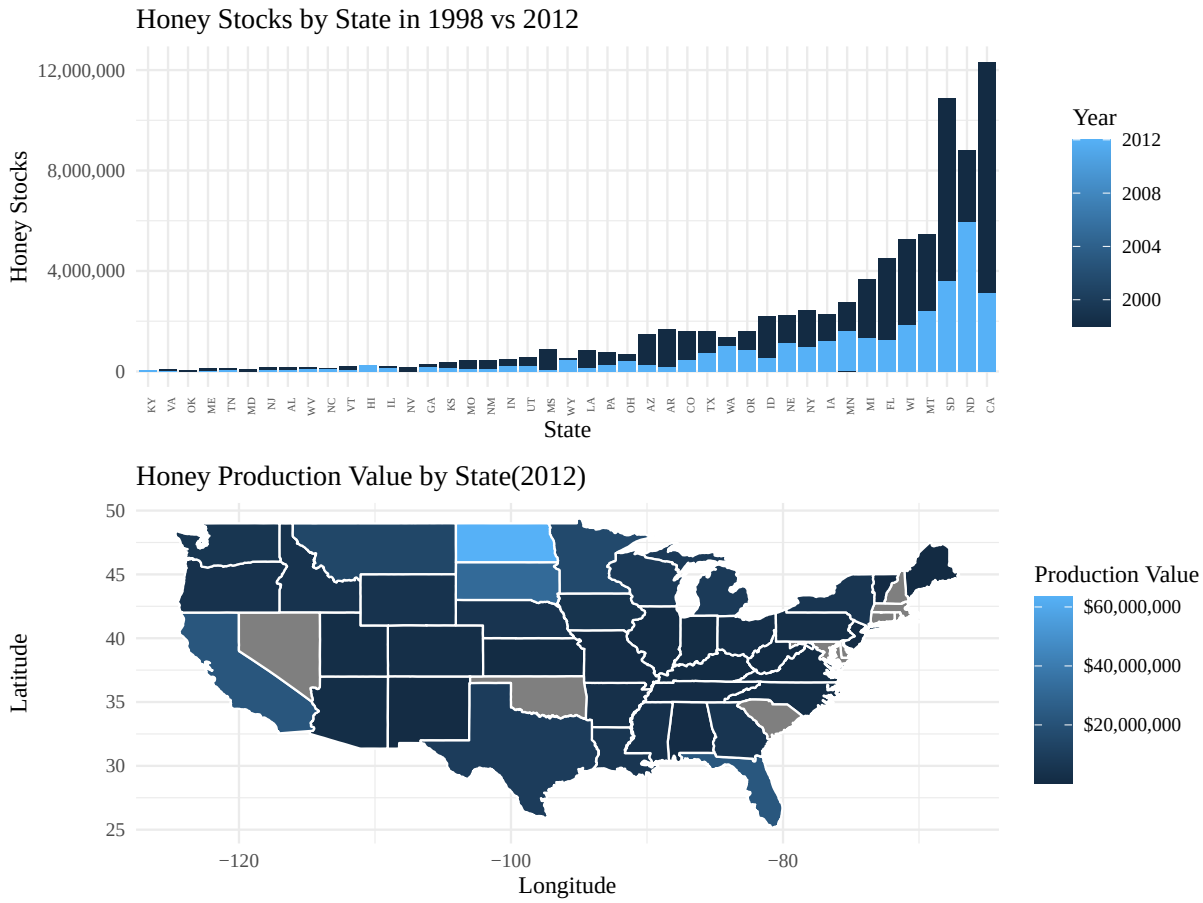
2. Now combine those plots side-by-side using library(patchwork) we learned in week 5

```
p1 + p2
```



3. Now combine those plots by stacking them vertically (also using library(patchwork) we learned in week 5

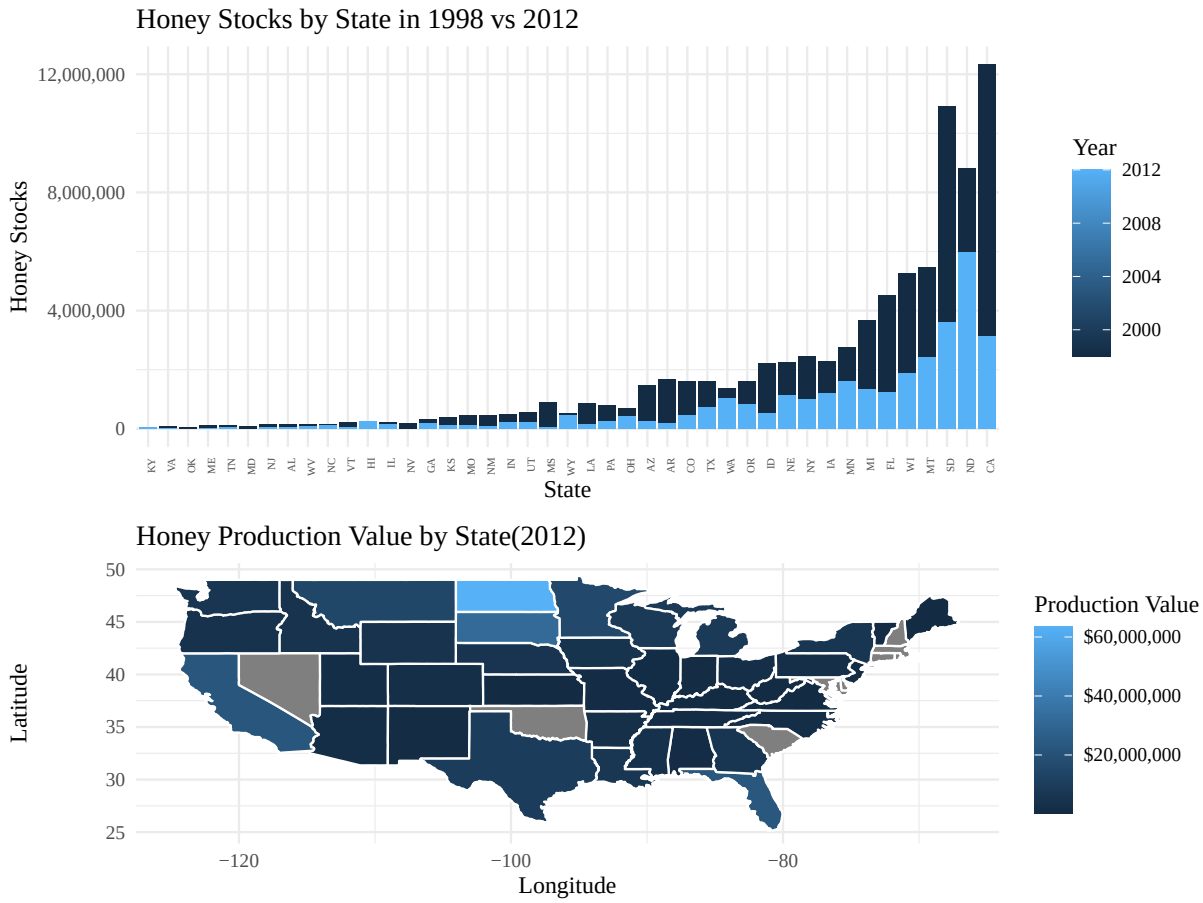
```
p1/p2
```



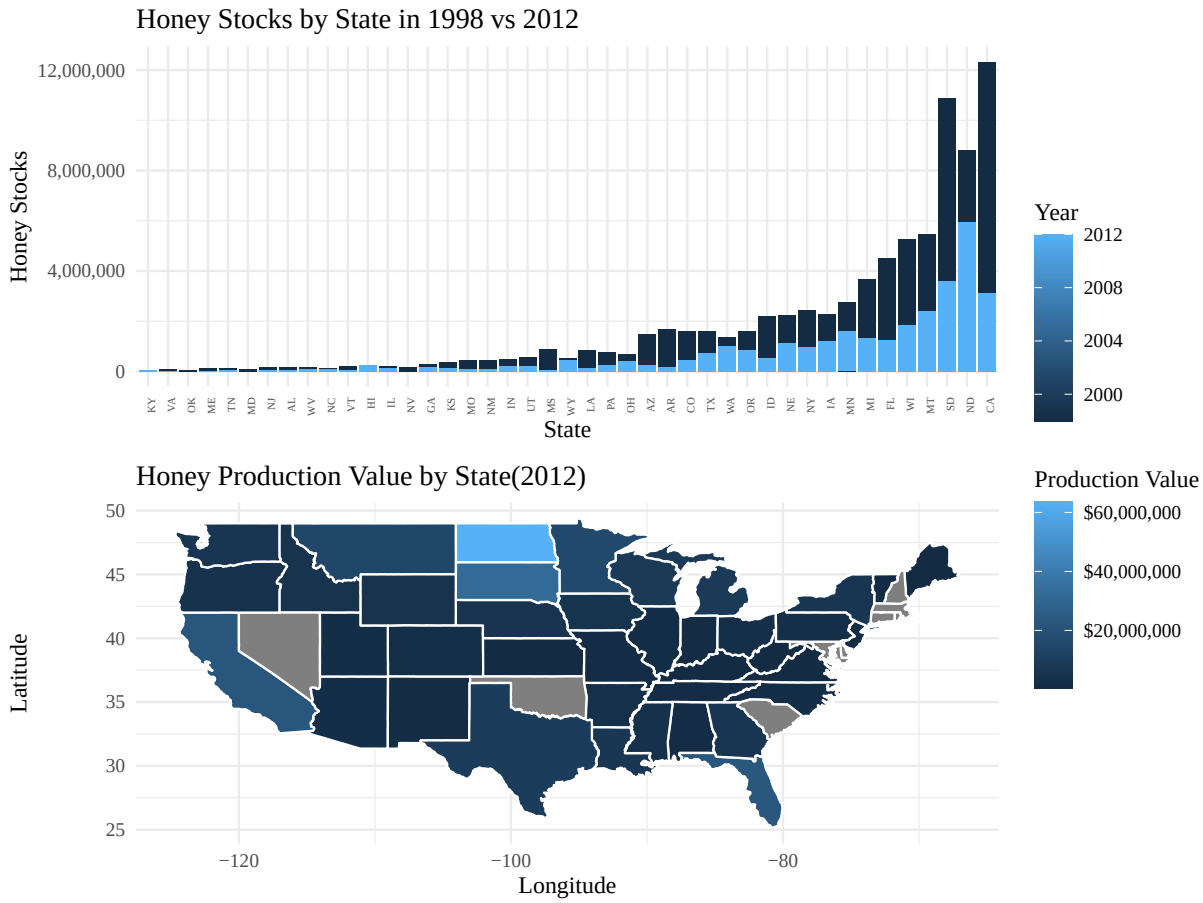
4. Which one do you prefer when comparing #1 vs. #2 vs. #3? Reflect briefly.

Answer: I like the #3 most because the side by side stacking makes the plots look clustered. However, the #3 is not satisfying because the width of the p2 is affected by p1. I wonder how can I fix it. So I tried a few things.

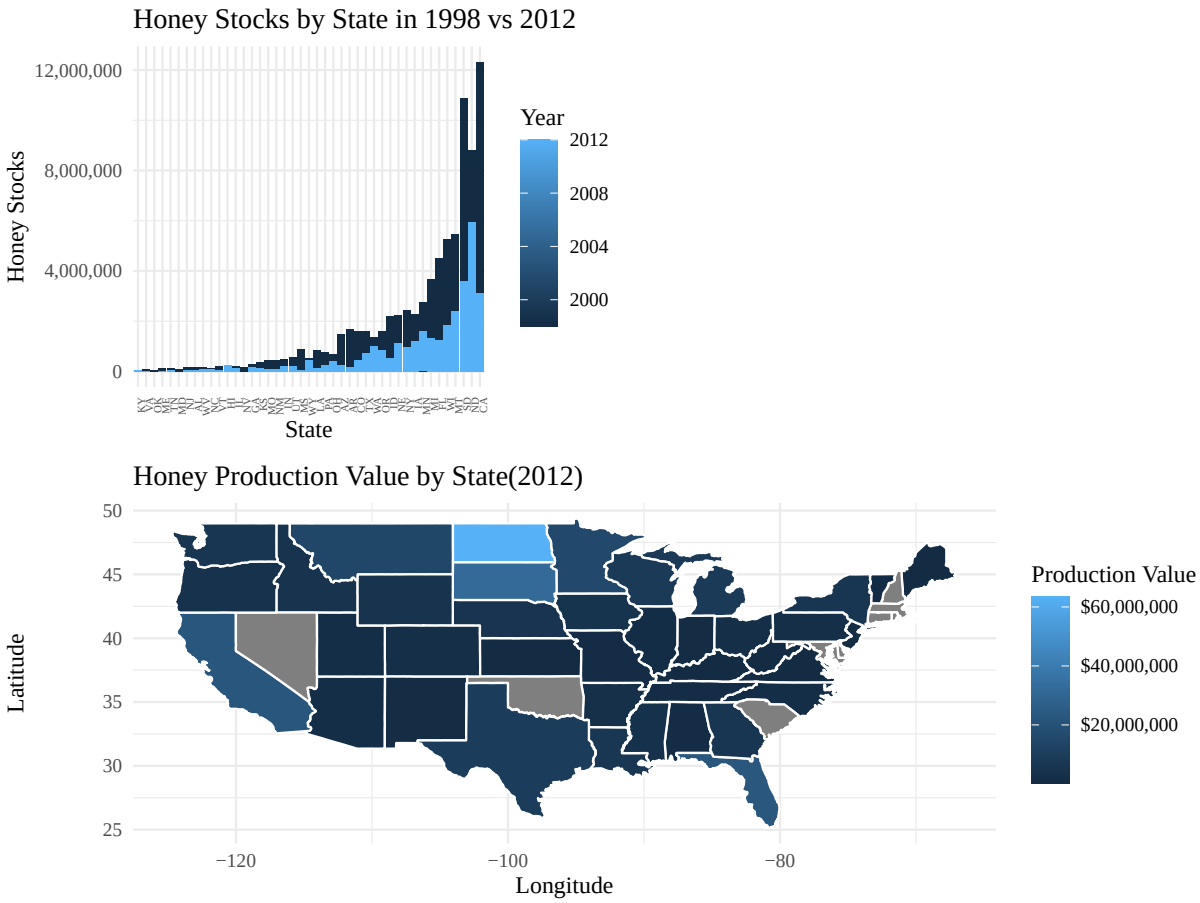
```
p1/p2 + plot_layout(heights = c(1,0.7)) # Equal heights. This doesn't work well
```




```
p1/p2 + plot_layout(guides = "collect", heights = c(1,1))
```




```
# This one looks slightly better

(p1 + plot_spacer()) / p2 + plot_layout(heights = c(1,1))
```


Honey Stocks by State in 1998 vs 2012
12,000,000
8,000,000
4,000,000
0
Honey Stocks
State
Year
2012
2008
2004
2000
Honey Production Value by State(2012)
50
45
40
35
30
25
Latitude
−120
−100
−80
Longitude
Production Value
$60,000,000
$40,000,000
$20,000,000