# Lab_2

Emmanuel Maduneme

2023-02-01

The purpose of this lab is to use color to your advantage. You will be asked to use a variety of color palettes, and use color for its three main purposes: (a) distinguish groups from each other, (b) represent data values, and (c) highlight particular data points.

## Data

We'll be working with the honey production data from #tidytuesday. The #tidytuesday repo contains the full data, but we'll work with just the cleaned up version, using the **honeyproduction.csv** file, which is posted on the website.

The data is in under Dataset tab of Week 4 module on Canvas.

You can import the dataset using the code below.

```
here()
```

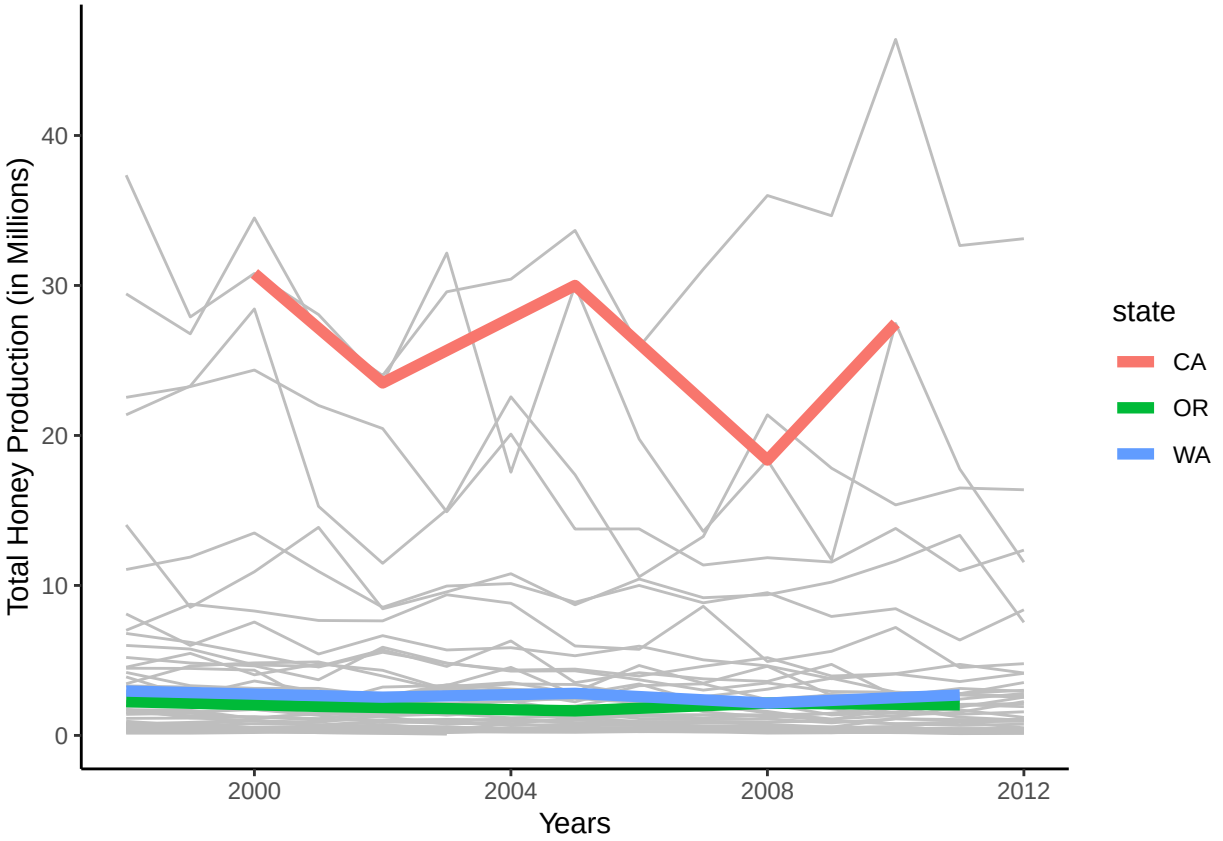
```
## [1] "/Users/emmanuelmaduneme/Desktop/UO SOJC PhD classes/Winter 23 Materials/Data Viz/Data Viz Proje
```

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

1. Visualize the total production of honey (**totalprod**) across years (**year**) by state (**state**). Use color to highlight the west coast (Washington, Oregon, and California) with a different color used for each west coast state.

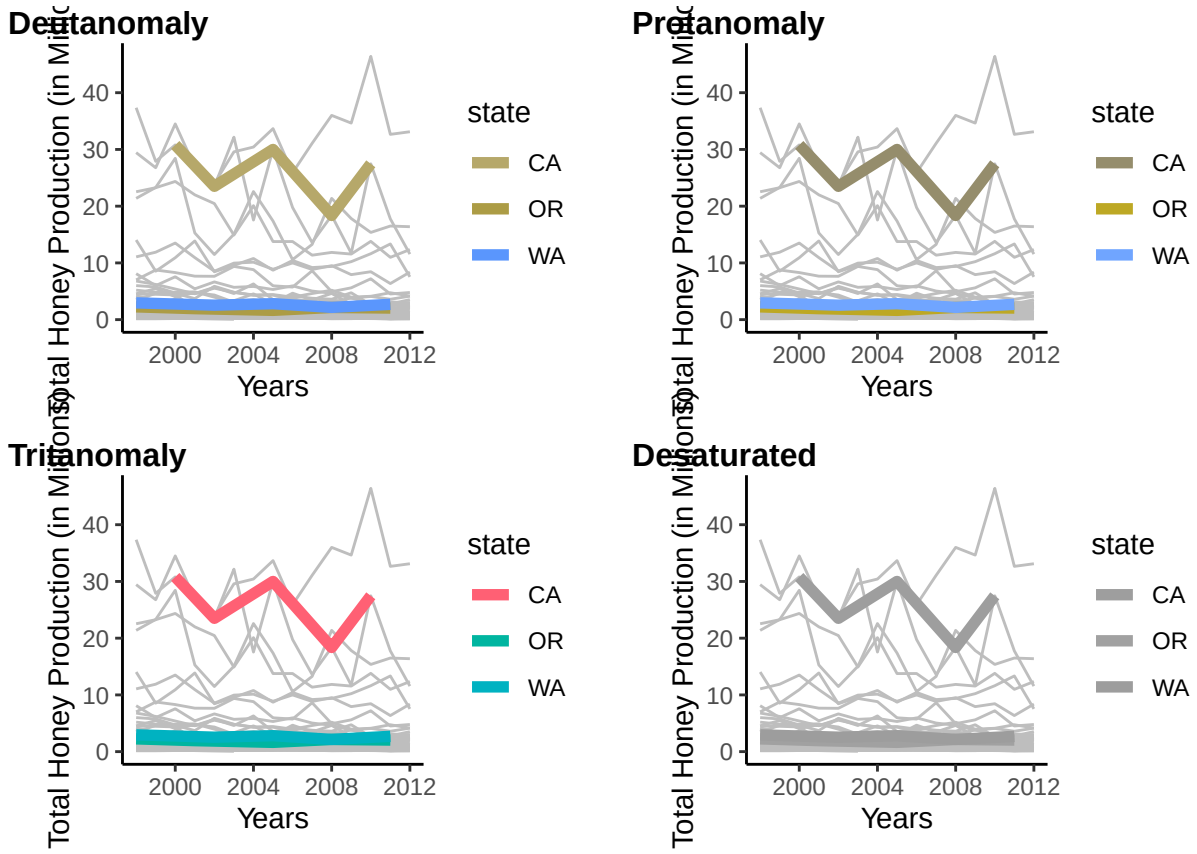
- **Hint 1**: I'm not asking for a specific kind of plot, just one that does the preceding. But if you're trying to visualize change over time, a bar chart is likely not going to be the best choice.
- **Hint 2**: To get each state to be a different color you should either map state to color (for your layer that adds the west coast colors) or use the gghighlight package.

```
bee_plot <- d %>%
  ggplot(aes(year, totalprod/1e6, group = state)) +
  geom_line(col = "grey") +
  geom_line(aes(col = state),
            data = filter(d, state == c("WA", "OR", "CA")), size = 2)+
  theme_classic() +
  labs(x = "Years",
       y = "Total Honey Production (in Millions)")
bee_plot
```



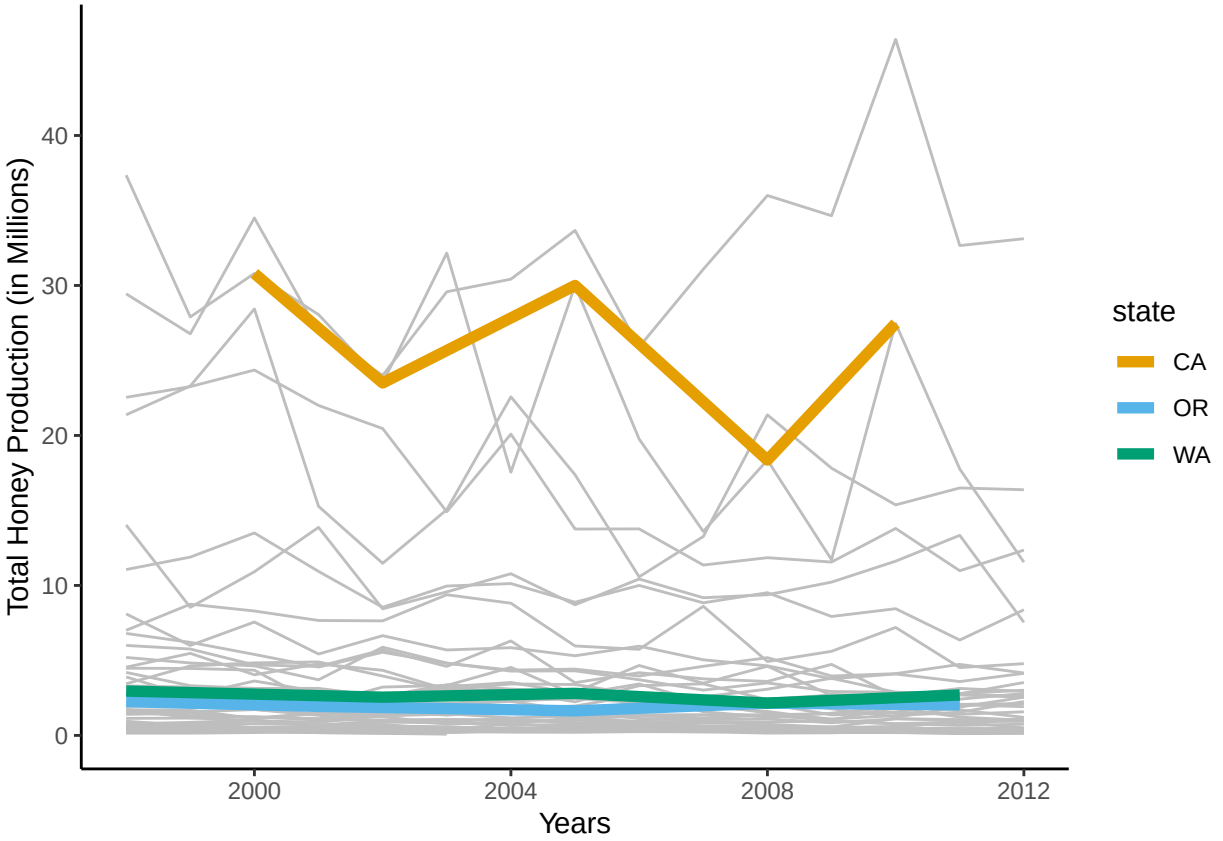

2. Reproduce the plot according three different kinds of color blindness using the `cvd_grid` package from the `colorblindr` package.

```
p_load(colorblindr)
cvd_grid(bee_plot)
```




3. Reproduce the plot using a color blind safe palette of your choice.

```
### Using a CVD inclusive Palettes
bee_plot +
  scale_color_OkabeIto()
```




4. Download the file **'us census bureau regions and divisions.csv'** from the course website denoting the region and division of each state.

- Join the file with your honey file.
- Produce a bar plot displaying the average honey for each state (collapsing across years).
- Use color to highlight the region of the country the state is from.
- Note patterns you notice.

I noticed that Northeast states do not produce great amounts of honey, compared to Northwestern and Western states.

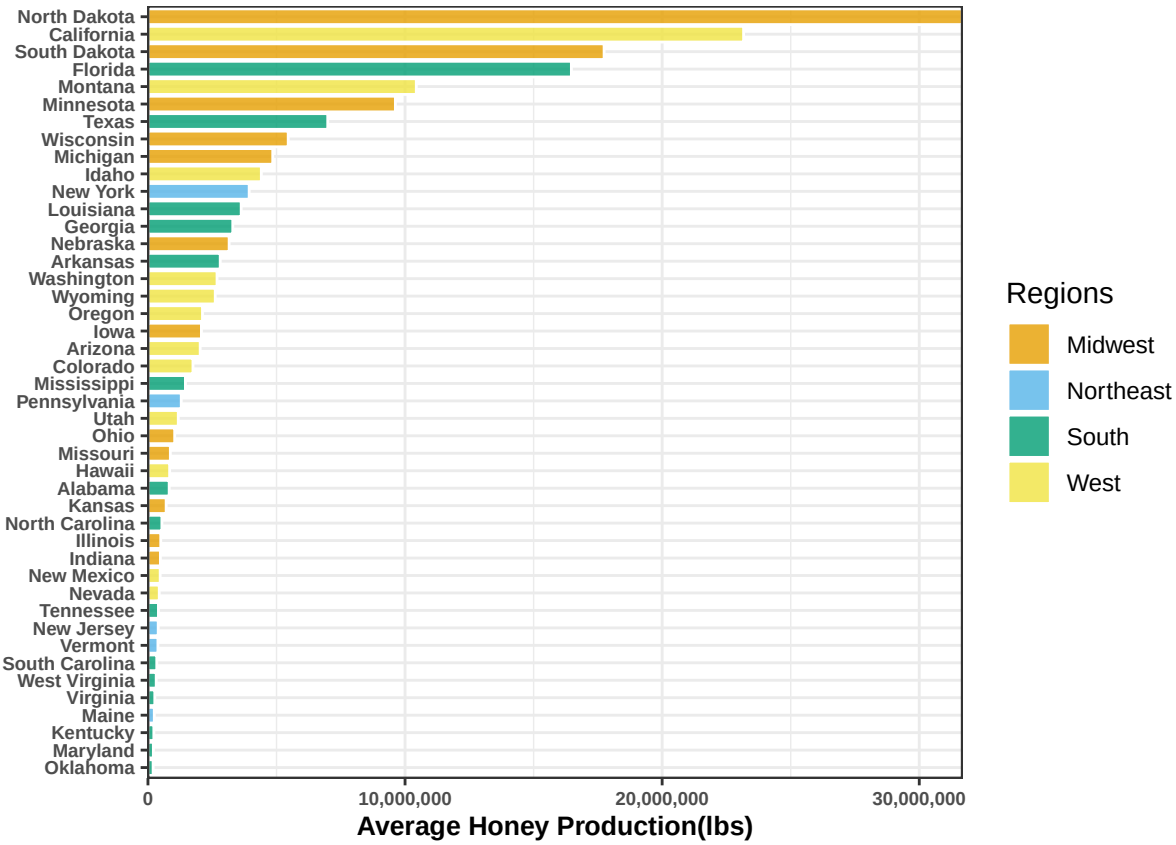
```
us_regions <- read.csv(here("data", "us census bureau regions and divisions.csv")) %>%
  clean_names()


us_regions <- us_regions %>%
   rename(state_name = state,
          state = state_code)

honey_data <- left_join(d, us_regions)


honey_data %>%
  group_by(state_name, region) %>%
    summarise(avg_honey_prod = mean(totalprod, na.rm = T)) %>%
```

```
  ggplot(aes(avg_honey_prod, fct_reorder(state_name, avg_honey_prod))) +
   geom_col(aes(fill = region),
            col = "white",
            alpha = 0.8) +
  ## Changing x axis into nominal numbers with comma
  scale_x_continuous(expand = c(0,0),
                     labels = scales::comma) +
  scale_fill_OkabeIto(name = "Regions") +
  theme_bw() +
  labs(x = "Average Honey Production(lbs)",
       y = "") +
   theme(axis.text = element_text(face = "bold", size = 7),
     axis.title.x = element_text(face = "bold", size = 10))
```

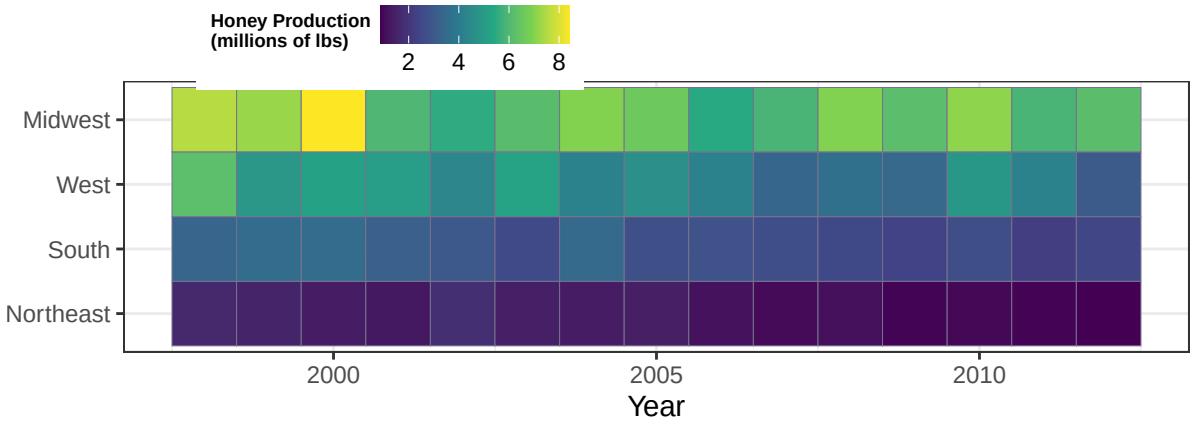



The plot should look like similar to the following plot (see the pdf).

5. Create a heatmap displaying the average honey production across years by *region* (averaging across states within region). The plot should look like similar to the following plot (see the pdf).

```
honey_data %>%
  group_by(year, region) %>%
  summarise(avg_prodctn = mean(totalprod, na.rm = T)/1e6) %>%
  ggplot(aes(year, fct_reorder(region, avg_prodctn)))  +
  geom_tile(aes(fill = avg_prodctn), col = "#76728d") +
  scale_fill_viridis_c("Honey Production\n(millions of lbs)\n") +
```

```
    coord_fixed() +
  labs(x = "Year",
       y = "") +
  theme_bw() +
  theme(legend.position = c(0.25, 1.15),
                            legend.direction = "horizontal",
        legend.title = element_text(size = 7, face = "bold"),
        legend.key.size = unit(.5, 'cm'),
        legend.key.height = unit(.5, "cm"))
```




6. Create at least one more plot of your choosing using color to distinguish, represent data values, or highlight. If you are interested in producing maps, I suggest grabbing a simple features data frame of the US using the Albers projection by doing the following:

```
us <- usa_sf()

honey_sub <- honey_data %>%
  mutate(iso_3166_2 = state) %>%
  select(iso_3166_2, totalprod, year)


full_set <- expand_grid(iso_3166_2 = unique(us$iso_3166_2),
                        year = 1998:2012)
honey_sub <- left_join(full_set, honey_sub)
```

```
honey_geoset <- left_join(us, honey_sub)

honey_geoset %>%
  ggplot() +
  geom_sf(aes(fill = totalprod/1e6)) +
  facet_wrap(~year) +
  scale_fill_continuous_sequential( name =  "Honey Production\n(millions of lbs)\n",
                                    na.value = "white",
                                    palette = "YlOrBr") +
   theme_void() +
    theme(legend.position = "bottom",
          legend.justification = "center",
          legend.direction = "horizontal",
          legend.title = element_text(size = 7, face = "bold"),
          legend.key.size = unit(1, 'cm'),
          legend.key.height = unit(.5, "cm"))
```