

# Class Activity 5

Your name here

2024-04-02

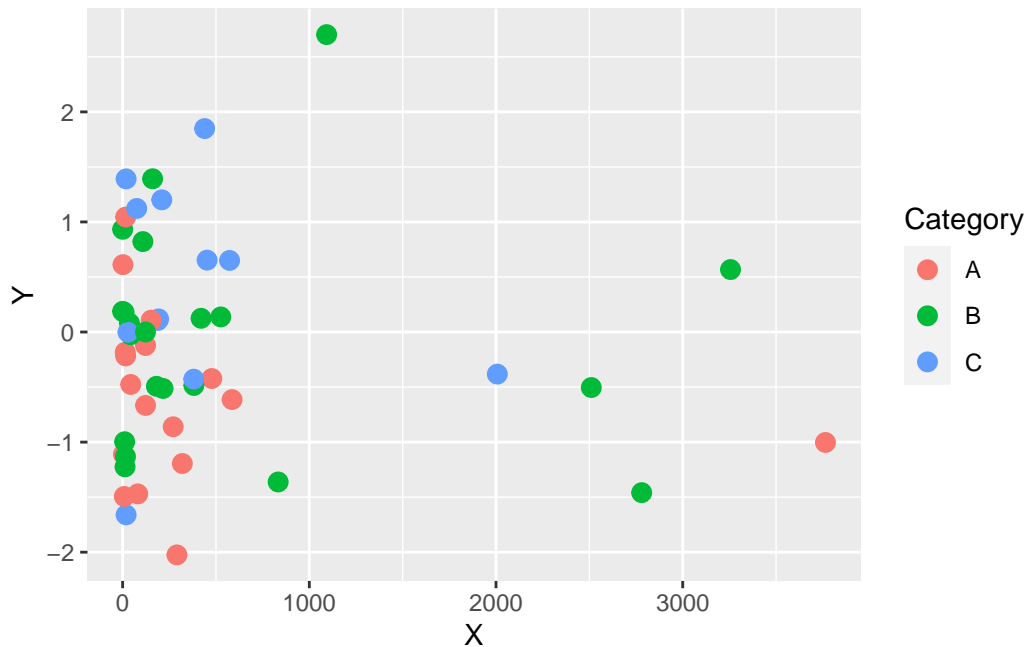
## Problem 1: Changing color and shape scales

In this problem, you will learn about the effects of changing colors, scales, and shapes in `ggplot2` for both gradient and discrete color choices. You will be given a series of questions and examples to enhance your understanding. Consider the following scatter plot

```
# Generate sample data
set.seed(42)
data <- data.frame(
  Category = factor(sample(1:3, 50, replace = TRUE), labels = c("A", "B", "C")),
  X = 10 ^ rnorm(50, mean = 2, sd = 1),
  Y = rnorm(50, mean = 0, sd = 1)
)

p <- ggplot(data, aes(x = X, y = Y, color = Category)) +
  geom_point(size = 3)

p
```



- a. Modify the scatter plot to use custom colors for each category using `scale_color_manual()`. What is the effect of changing the colors on the plot's readability?

*Answer:*

```
p <- ggplot(data, aes(x = , y = , color = )) +
  geom_point(size = 3)

p
```

```
Error in `geom_point()`:
! Problem while setting up geom.
! Error occurred in the 1st layer.
Caused by error in `compute_geom_1()`:
! `geom_point()` requires the following missing aesthetics: x and y
```

- b. Modify the scatter plot to use custom shapes for each category using `scale_shape_manual()`. What is the effect of changing the shapes on the plot's readability?

*Answer:* Changing the shapes using `scale_shape_manual()` helps to distinguish between categories and improves the plot's readability

```
p <- ggplot(data, aes(x = , y = , shape = , color = )) +
  geom_point(size = 3)
```

p

Error in `geom\_point()`:

! Problem while setting up geom.

i Error occurred in the 1st layer.

Caused by error in `compute\_geom\_1()`:

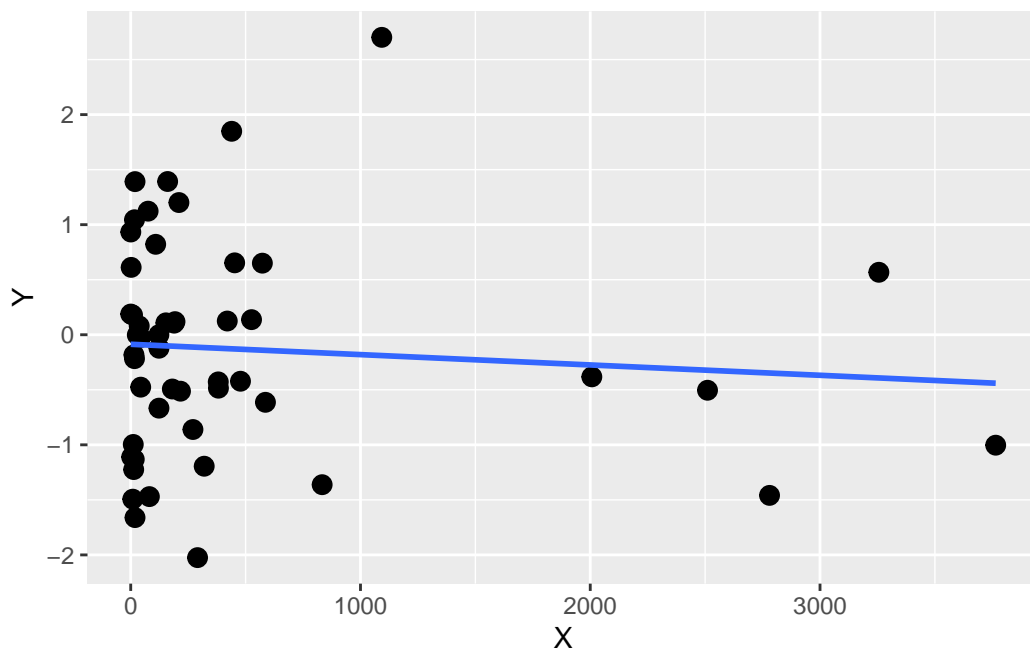
! `geom\_point()` requires the following missing aesthetics: x and y

- c. Try modifying the plot by combining color, shape, and theme customizations. Additionally, try using `geom_smooth()` to add trend lines for each category. Pay attention to how each element affects the overall readability and interpretability of the plot.

*Answer:*

```
p <- ggplot(data, aes(x = X, y = Y)) +
  geom_point(aes(color = , shape = ), size = 3) +
  geom_smooth(aes(group = , color = ), method = "lm", se = FALSE)
```

p



## Problem 2: US maps

Now, let's learn about the effect of changing various coordinate systems in `ggplot2` using a map example from the `usmap` package. We will explore the different types of coordinate systems available in `ggplot2` and how they can be applied to the map visualization.

```
#install.packages("usmap")    #uncomment to install
library(usmap)
```

**a. Plot a simple map of the United States using `ggplot2` and the `usmap` package.**

*Answer:*

```
us <-  
us
```

Error in eval(expr, envir, enclos): object 'us' not found

**b. Apply the `coord_flip()` function to the map to flip the x and y axes.**

*Answer:*

```
us_flipped <-  
us_flipped
```

Error in eval(expr, envir, enclos): object 'us\_flipped' not found

**c. Apply the `coord_polar()` function to the map to transform the plot to a polar coordinate system**

*Answer:*

```
us_polar <-  
us_polar
```

Error in eval(expr, envir, enclos): object 'us\_polar' not found

d. Apply the `coord_quickmap()` function to the map to provide an approximation for a map projection.

*Answer:*

```
us_quickmap <-  
us_quickmap
```

Error in eval(expr, envir, enclos): object 'us\_quickmap' not found

## Problem 3: Choropleth map

In today's class we created choropleth maps of states in the US based on ACS data.

```
states <- map_data("state")  
ACS <- read_csv("https://raw.githubusercontent.com/deepbas/statdatasets/main/ACS.csv")  
ACS <- dplyr::filter(ACS, !(region %in% c("Alaska", "Hawaii"))) # only 48+D.C.  
ACS$region <- tolower(ACS$region) # lower case (match states regions)
```

### (a) Mapping median income

Create a choropleth plot that uses color to create a MedianIncome map of the US.

*Answer:*

```
# map median income  
ggplot(data= ) + coord_map() +  
  geom_map(aes(map_id = , fill = ), map = ) +  
  expand_limits(x= , y= )
```

Error in `geom\_map()`:  
! `map` must be a data frame, not absent.

### (b) Mapping deviations from national median income

The median income in the US in 2016 was estimated to be \$27,000. Redraw your map in (a) to visualize each state's deviation from national median income.

Answer:

```
# compare state income to national income
ggplot(data= ) + coord_map() +
  geom_map(aes(map_id = , fill = ), map = ) +
  expand_limits(x= , y= )
```

```
Error in `geom_map()` :
! `map` must be a data frame, not absent.
```

### (c) Changing numerically scaled color

You should use a *diverging* color for (b) to highlight larger deviations from the national median. Add `scale_fill_distiller` to the map from (b) and select a diverging palette.

Answer:

```
# change to a diverging color
ggplot(data= ) + coord_map() +
  geom_map(aes(map_id = , fill = ), map = ) +
  expand_limits(x= , y= )
```

```
Error in `geom_map()` :
! `map` must be a data frame, not absent.
```

### (d) Fixing a midpoint on a diverging scale

Use `scale_fill_gradient2` to fix a midpoint scale value at a white color, with diverging colors for larger positive and negative values. Apply this color to your map in (b) and fix the midpoint at an appropriate value.

Answer:

```
# change to a gradient fill color
ggplot(data=) + coord_map() +
  geom_map(aes(map_id = , fill = ), map = ) +
  expand_limits(x=, y=) + ggtitle("")
```

```
Error in `geom_map()``:  
! `map` must be a data frame, not absent.
```

### (e) Polygon map

```
# Merge income data with geographic information  
income_data <- left_join(states, ACS, by = c("region" = "region"))
```

For this task, you will create a polygon map to visualize the `MedianIncome` across different states. Pay attention to the shapes and sizes of states as depicted on the map.

```
library(sf)  
  
ggplot(data = income_data) +  
  geom_polygon() +  
  coord_sf() +  
  labs(fill = "Median Income", title = "Median Income by State") +  
  theme_minimal()
```

```
Error in `geom_polygon()``:  
! Problem while setting up geom.  
i Error occurred in the 1st layer.  
Caused by error in `compute_geom_1()``:  
! `geom_polygon()` requires the following missing aesthetics: x and y
```

### (f) Visualizing Relative Income Deviation

```
national_median <- 27000  
  
ACS$IncomeDeviationPercent <- ((ACS$MedianIncome - national_median) / national_median) * 100  
income_data <- left_join(states, ACS, by = c("region" = "region"))  
  
ggplot(data = income_data) +  
  geom_polygon() +  
  coord_sf() +  
  labs(fill = "Income Deviation (%)", title = "Income Deviation from National Median by State") +  
  theme_minimal()
```

```
Error in `geom_polygon()`:  
! Problem while setting up geom.  
i Error occurred in the 1st layer.  
Caused by error in `compute_geom_1()`:  
! `geom_polygon()` requires the following missing aesthetics: x and y
```