Class Activity 5

Arthur Viegas Eguia

April 03 2024

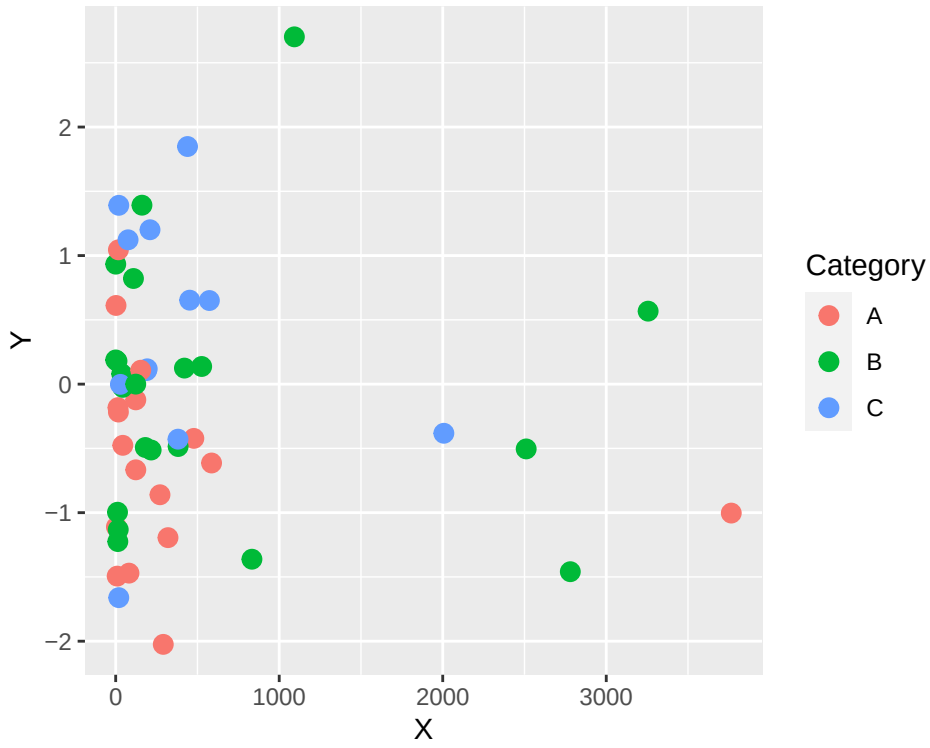
Problem 1: Changing color and shape scales

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
```



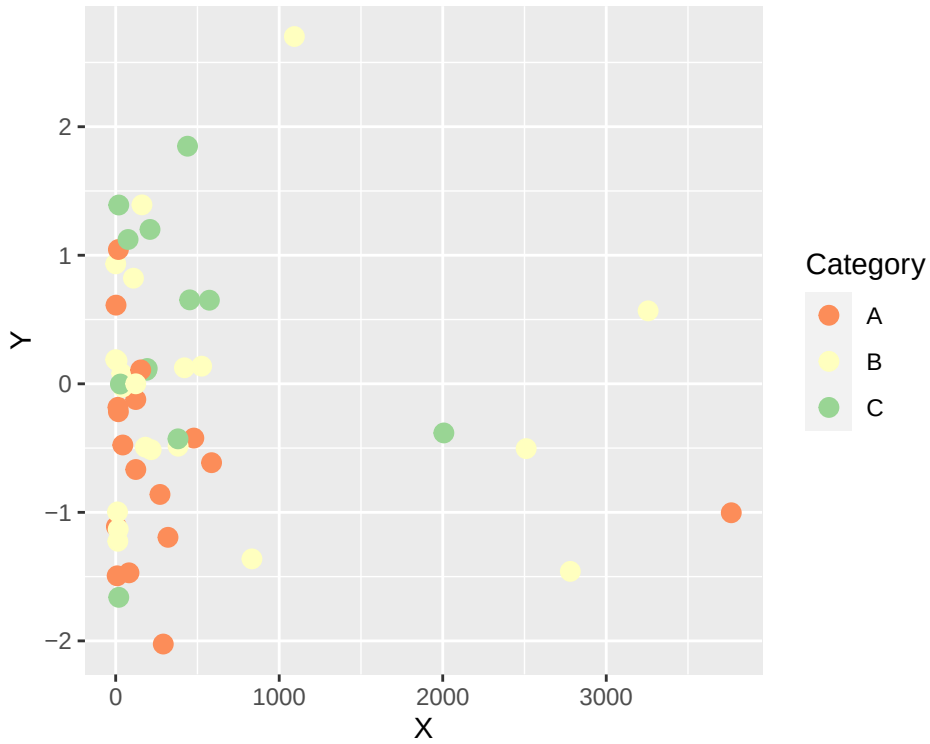
a. Modify the scatter plot to use custom colors for each category using `scale_color_manual()`. What is

the effect of changing the colors on the plot's readability?

Answer: It makes the graph more readable. it is easier to tell the points apart

```
p <- ggplot(data, aes(x = X, y = Y, color = Category)) +
  geom_point(size = 3) + scale_color_manual(values = c("#fc8d59", "#ffffbf", "#99d594"))

p
```

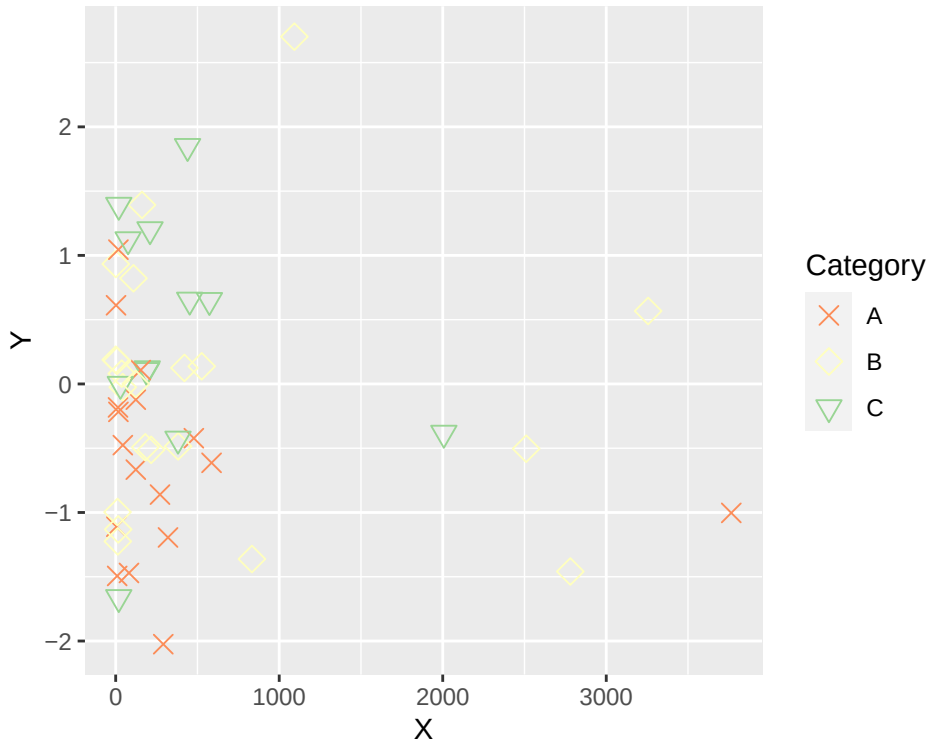


b. Modify the scatter plot to use custom shapes for each category using `scale_shape_manual()`. What is the effect of changing the shapes on the plot's readability?

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

```
p <- ggplot(data, aes(x = X, y = Y, shape = Category,  color = Category)) +
  geom_point(size = 3) +scale_shape_manual(values= c("A" = 4, "B" = 5, "C" = 6)) + scale_color_manual(va

p
```

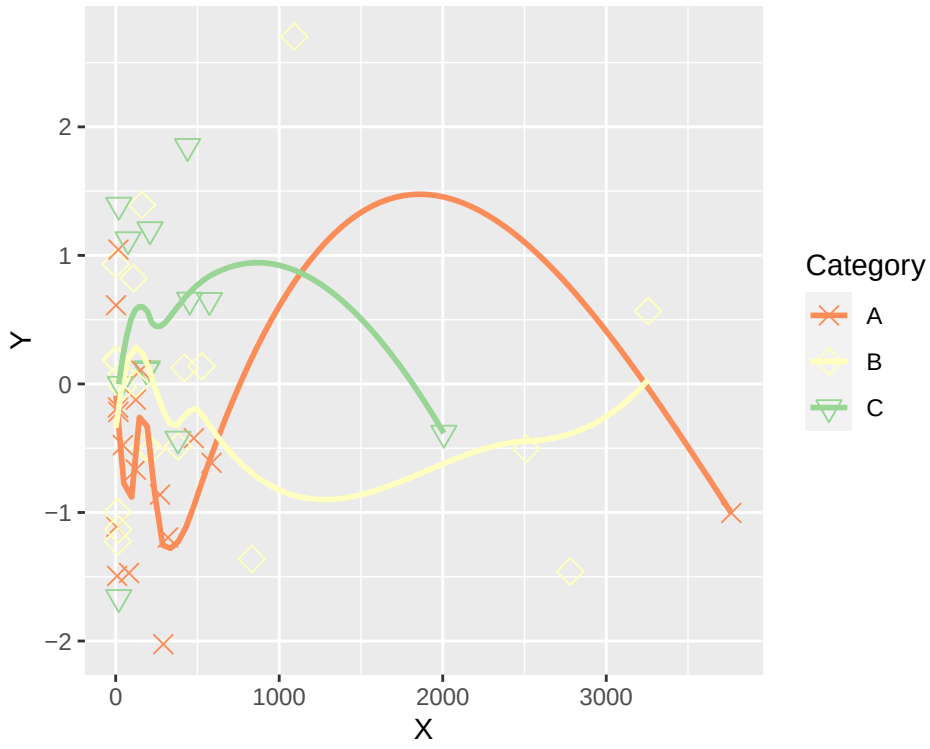


c. Try modifying the plot by combining color, shape, and theme customizations. Additionally, try using `geom_smooth()` to add trend lines for each category. Pay attention to how each element affects the overall readability and interpretability of the plot.

Answer: They all make the plot more readable. Each helps distinguish a category in a different way.

```
p <- ggplot(data, aes(x = X, y = Y)) +
  geom_point(aes(color = Category, shape = Category), size = 3) +
  geom_smooth(aes(color = Category), method = "loess", se = FALSE) + scale_color_brewer(palette="Dark")


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 <- plot_usmap(regions = "states")
#us
```

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

Answer:

```
#us_flipped <- us + coord_flipped()
#us_flipped
```

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

Answer:

```
#us_polar <- us + coord_sf()
#us_polar
```

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

Answer:

```
#us_quickmap <- us + coord_quickmap()
#us_quickmap
```

Problem 3: Chloropeth map

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

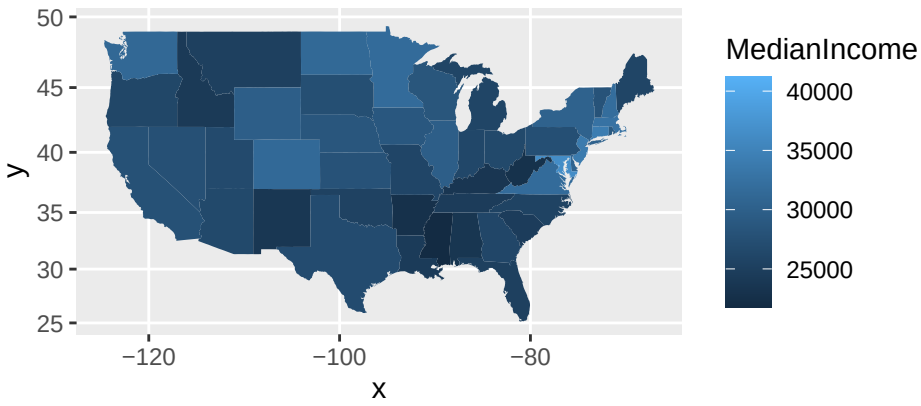
```
states <- map_data("state")
ACS <- read.csv("https://raw.githubusercontent.com/deepbas/statdatasets/main/ACS.csv")
ACS <- dplyr::filter(ACS, !(region  %in% c("Alaska", "Hawaii"))) # only 48+D.C.
ACS$region <- tolower(ACS$region)  # lower case (match states regions)
```

(a) Mapping median income

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

Answer:

```
# map median income
ggplot(data= ACS) + coord_map() +
  geom_map(aes(map_id = region, fill = MedianIncome), map = states) +
  expand_limits(x= states$long, y= states$lat)
```

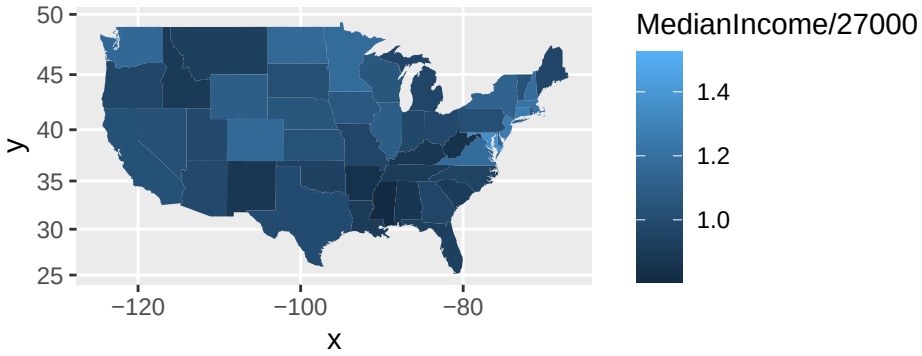


(b) Mapping deviations from national median income

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= ACS) + coord_map() +
  geom_map(aes(map_id = region, fill = MedianIncome/27000), map = states) +
  expand_limits(x= states$long, y= states$lat)
```

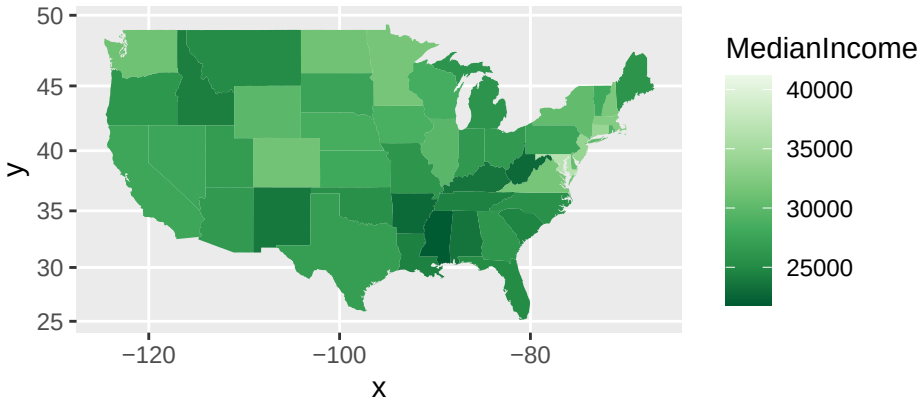


(c) Changing numerically scaled color

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= ACS) + coord_map() +
  geom_map(aes(map_id = region, fill = MedianIncome), map = states) +
  expand_limits(x=states$long, y= states$lat) + scale_fill_distiller(palette="diverging")
```

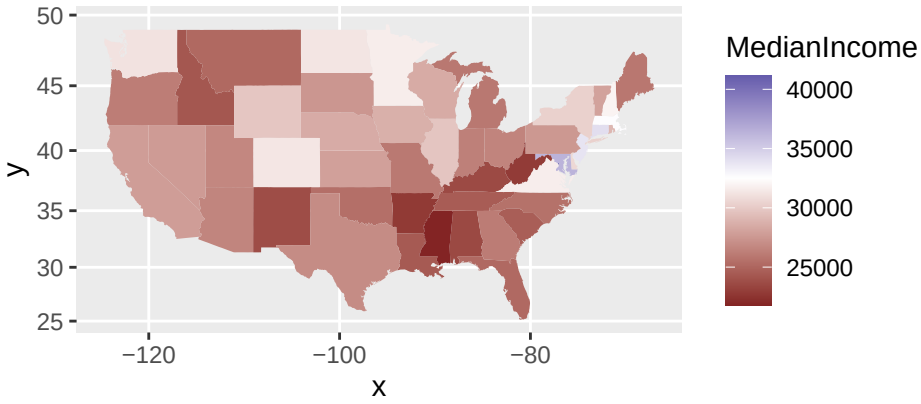


(d) Fixing a midpoint on a diverging scale

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=ACS) + coord_map() +
  geom_map(aes(map_id = region, fill = MedianIncome), map = states) +
  expand_limits(x=states$long, y= states$lat) + ggtitle("") + scale_fill_gradient2(midpoint = 32500)
```



(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()
```

(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()
```