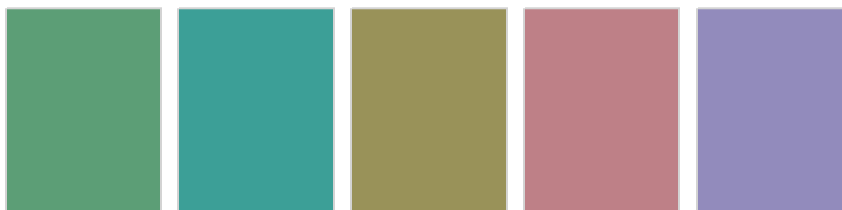


# Homework 5

This homework is due on the deadline posted on edX. Please submit a .pdf file of your output and upload a .zip file with your .Rmd file.

**Problem 1:** Use the color picker app from the **colorspace** package ( `colorspace::choose_color()` ) to create a qualitative color scale containing five colors. One of the five colors should be `#5C9E76` , so you need to find four additional colors that go with this one.

```
# replace "#FFFFFF" with your own colors
colors <- c("#5C9E76", "#3C9F97", "#999259", "#BE8087", "#928BBC")
swatchplot(colors)
```



For the rest of this homework, we will be working with the `midwest_clean` dataset, which is a cleaned up version of the **ggplot2** `midwest` dataset.

```
midwest_clean <- midwest %>%
  select(
    state, county, area, popdensity, percbelowpoverty, inmetro
  ) %>%      # keep only a subset of data
  na.omit()  # remove any rows with missing data
```

**Problem 2:** Perform a PCA of the `midwest_clean` dataset and make a rotation plot of components 1 and 2.

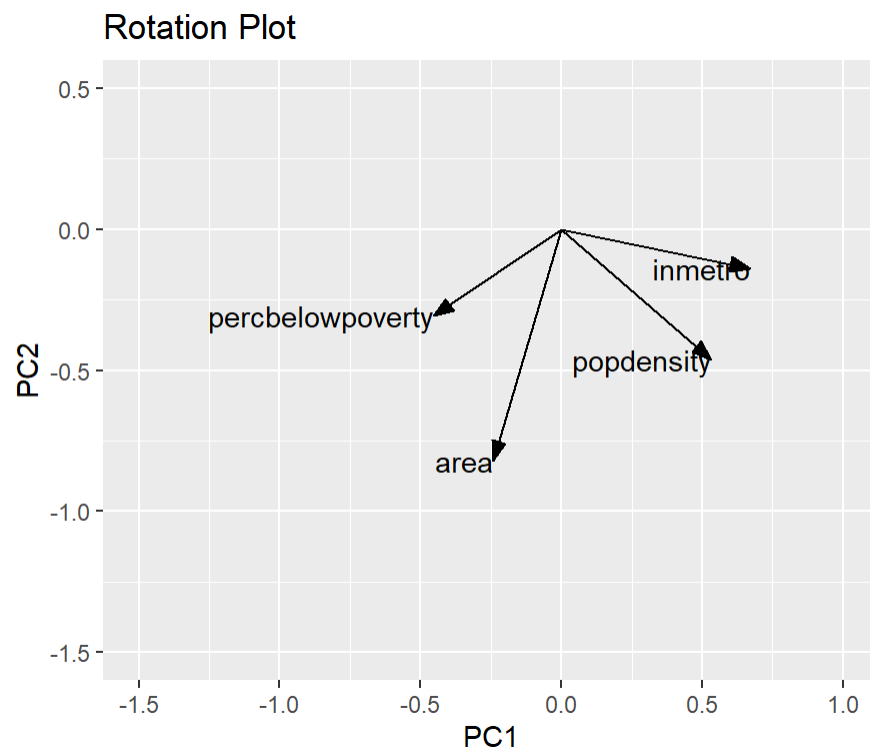
```

pca_fit <- midwest_clean %>%
  select(where(is.numeric)) %>%
  scale() %>%
  prcomp()

arrow_style <- arrow(
  angle = 20, length = grid::unit(8, "pt"),
  ends = "first", type = "closed"
)

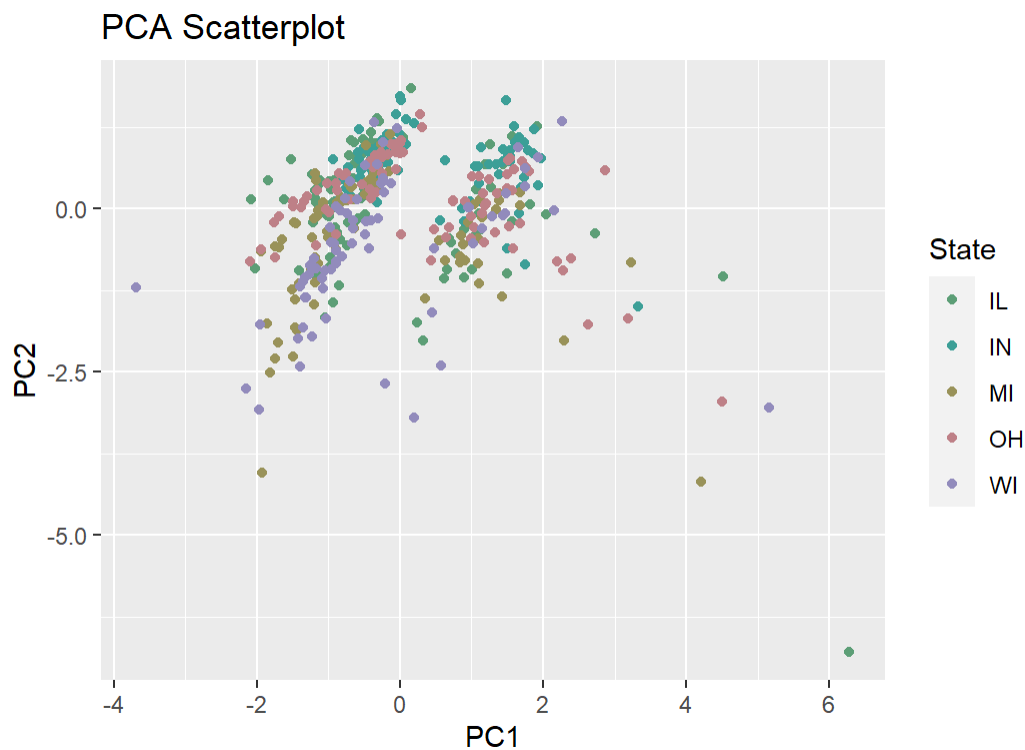
pca_fit %>%
  tidy(matrix = "rotation") %>%
  pivot_wider(
    names_from = "PC",
    values_from = "value",
    names_prefix = "PC"
  ) %>%
  ggplot(aes(PC1, PC2)) +
  geom_segment(
    xend = 0,
    yend = 0,
    arrow = arrow_style
  ) +
  geom_text(aes(label = column), hjust = 1) +
  xlim(-1.5, 1) +
  ylim(-1.5, 0.5) +
  coord_fixed() +
  labs(title = "Rotation Plot")

```



**Problem 3:** Make a scatter plot of PC 2 versus PC 1 and color by state. You should use the custom colorscale you created in Problem 1. Then use the rotation plot from Problem 2 to describe where Chicago, Illinois can be found on the scatter plot. Provide any additional evidence used to support your answer.

```
pca_fit %>%
  augment(midwest_clean) %>%
  ggplot(aes(.fittedPC1, .fittedPC2)) +
  geom_point(aes(color = state)) +
  scale_color_manual(
    name = "State",
    values = colors
  ) +
  labs(
    title = "PCA Scatterplot",
    x = "PC1",
    y = "PC2"
  )
)
```



Based on the rotation plot in part 2, we can see that data points with higher population density will be pushed down and to the right. Given that by far the farthest point on the scatter plot to the bottom right is isolated by itself, it is colored to match Illinois, and knowing that Chicago is a major metropolitan, we can deduce that the data point in the far bottom right is Chicago.