

Homework 9

Sookja Kang, sk26949

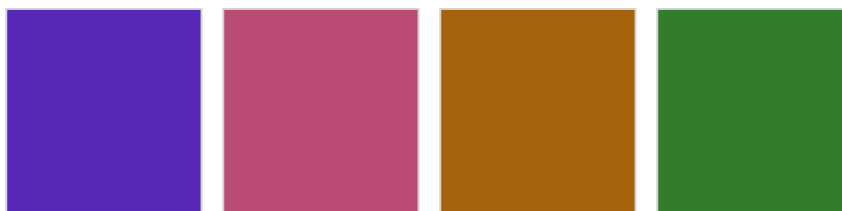
This homework is due on April 26, 2021 at 11:00pm. Please submit as a pdf file on Canvas.

Problem 1: (2 pts)

Use the color picker app from the **colorspace** package (`colorspace::choose_color()`) to create a qualitative color scale containing four colors. One of the four colors should be #5626B4, so you need to find three additional colors that go with this one.

```
# replace "#FFFFFF" with your own colors
colors <- c("#5626B4", "#BA4B75", "#A6620A", "#337C2C")

swatchplot(colors)
```



Problem 2: (4 pts) Take the following scatter plot of the penguins dataset and make three modifications:

1. Use the colors you chose in Problem 1.
2. Improve the visual appearance by choosing a theme and cleaning up axis labels.
3. Remove the need for a legend by direct-labeling the points.

```
#ggplot(penguins, aes(bill_length_mm, body_mass_g, color = species)) +
# geom_point(size = 2, na.rm = TRUE)
penguins_labels <- tibble(
  species = c("Adelie", "Chinstrap", "Gentoo"),
  body_mass_g = c(2500, 3500, 6000),
  bill_length_mm = c(35, 55, 55),
  label = c("Adelie", "Chinstrap", "Gentoo"),
  hjust = c(0, 0.5, 0),
  vjust = c(0, 0.5, 0)
```

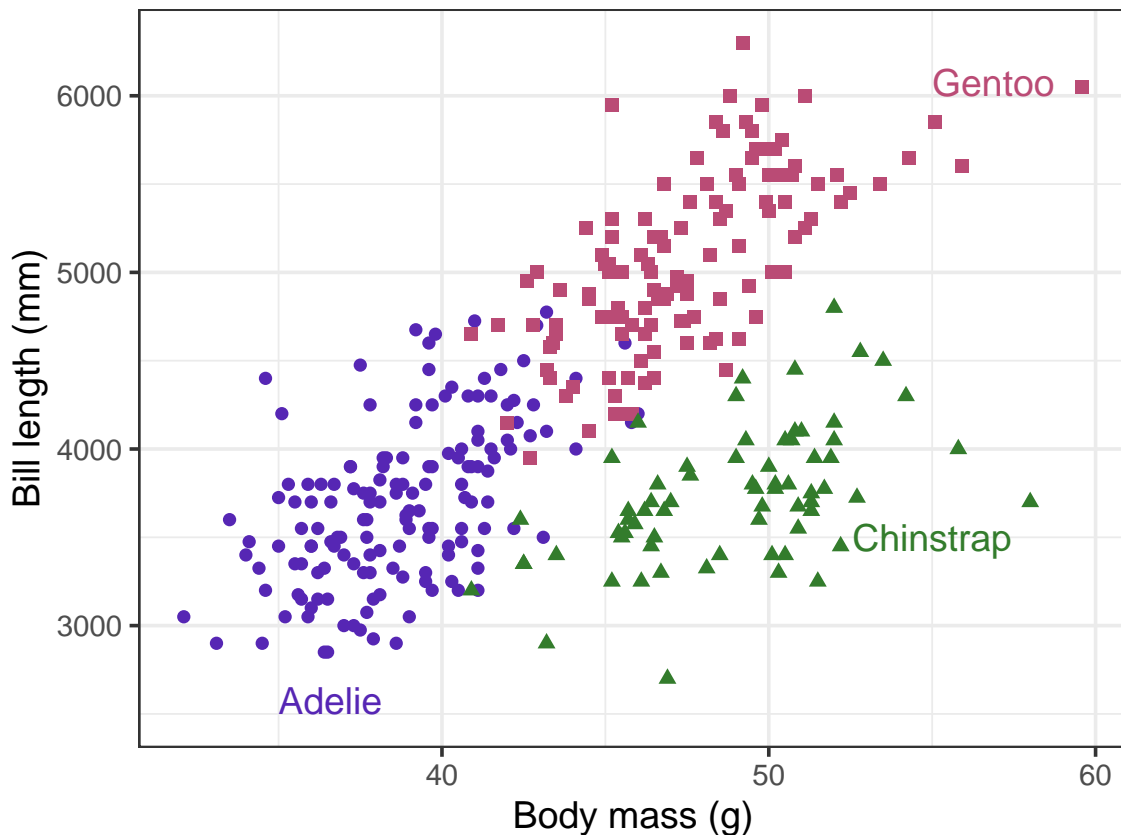
```

)
penguins_labels

## # A tibble: 3 x 6
##   species  body_mass_g bill_length_mm label      hjust vjust
##   <chr>      <dbl>      <dbl> <chr>      <dbl> <dbl>
## 1 Adelie      2500         35 Adelie        0     0
## 2 Chinstrap   3500         55 Chinstrap    0.5   0.5
## 3 Gentoo      6000         55 Gentoo        0     0

ggplot(penguins, aes(bill_length_mm, body_mass_g, color = species)) +
  geom_point(aes(shape = species), size = 2, na.rm = TRUE) +
  scale_x_continuous(name = "Body mass (g)") +
  scale_y_continuous(name = "Bill length (mm)") +
  geom_text(data = penguins_labels,
    aes(label = label,
      hjust = hjust,
      vjust = vjust),
    size = 14/.pt) +
  guides(color = "none", shape = "none") +
  scale_color_manual(
    values = c(Adelie = "#5626B4", Chinstrap = "#337C2C", "Gentoo" = "#BA4B75")
  ) +
  theme_bw(14)

```



Problem 3: (4 pts) The following scatter plot shows per-capita income versus number of inhabitants in all Texas counties in 2010. Use `geom_text_repel()` to label a subset of the counties by name. You can choose

the counties to subset as you wish. Also, choose a theme and clean up the axis labeling, and make any other improvements to the plot design you consider appropriate.

Hint: If you're not sure how to select a subset of counties to label, check out the examples on the **ggrepel** website for some inspiration: <https://ggrepel.slowkow.com/articles/examples.html#examples-1>

```
tx_census <- read_csv("https://wilkelab.org/SDS375/datasets/US_census.csv") %>%
  filter(state == "Texas") %>%
  select(county = name, pop2010, per_capita_income)
tx_census_f <- tx_census %>%
  filter(per_capita_income > 35000)
tx_census_f

## # A tibble: 5 x 3
##   county      pop2010 per_capita_income
##   <chr>      <dbl>      <dbl>
## 1 Borden County      641      40916
## 2 Collin County  782341      37362
## 3 Kendall County  33410      36418
## 4 King County      286      39511
## 5 Loving County     82      42220

tx_census %>%
  mutate(
    county = ifelse(county %in% c("Borden County", "Collin County", "Kendall County",
                                "King County", "Loving County"), county, "")
  ) %>%
  ggplot(aes(pop2010, per_capita_income)) +
  geom_point(size = 1.5) +
  scale_x_log10(name = "Population (2010)") +
  scale_y_continuous(name = "Per capita income") +
  geom_text_repel(aes(label = county),
                  max.overlaps = Inf,
                  box.padding = 0.6) +
  theme_bw(14)
```

