Homework 9

Wenxuan Zhou wz4388

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", "#D986D1", "#8BC99B", "#387FC9")

swatchplot(colors)
```

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

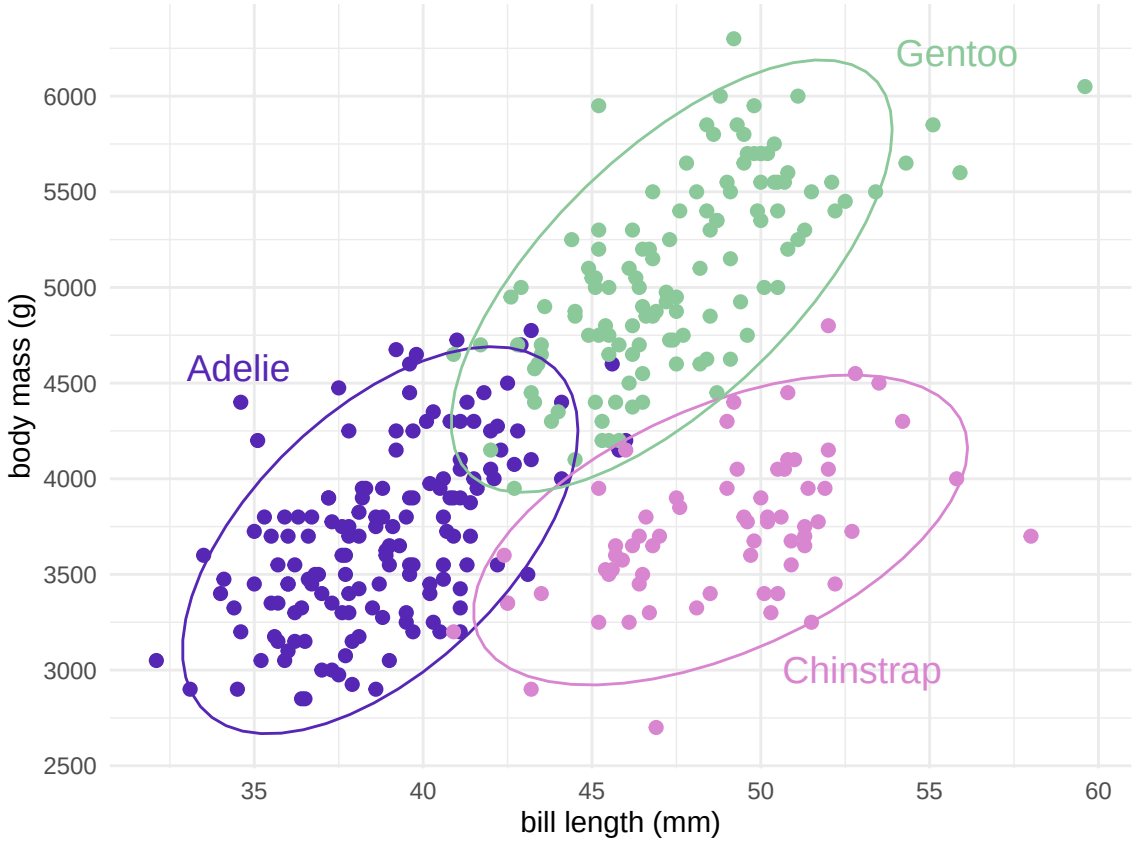
1. Use the colors you chose in Problem 1.
2. Improve the visual appearance by choosing a theme and cleaning up axis labels.
3. Remove the need for a legend by direct-labeling the points.

```
penguins_labels <- tibble(
  species = c("Adelie", "Chinstrap", "Gentoo"),
  bill_length_mm = c(33, 53, 54),
  body_mass_g = c(4510, 3000, 6300),
  hjust = c(0, 0.5, 0),
  vjust = c(0, 0.5, 1)
)

ggplot(penguins, aes(bill_length_mm, body_mass_g, color = species)) +
  geom_point(size = 2, na.rm = TRUE) + scale_color_manual(values = colors) +
```

```
  scale_x_continuous(name = 'bill length (mm)', breaks = seq(30, 60, by = 5)) +
  scale_y_continuous(name = 'body mass (g)',
                     breaks = seq(2000, 7000, by = 500)) +
  theme_minimal() +
  geom_text(data = penguins_labels, aes(label = species, hjust = hjust,
                                        vjust = vjust), size = 14/.pt) +
  stat_ellipse(size = 0.5) +
  guides(color = "none", shape = "none")
```

```
## Warning: Removed 2 rows containing non-finite values (stat_ellipse).
```



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$label <- ""
ix_label <- c(3, 43, 101, 118, 135, 173, 227)
tx_census$label[ix_label] <- tx_census$county[ix_label]
```

```
tx_census %>%
  ggplot(aes(pop2010, per_capita_income)) +
  geom_point(size = 1.5,
             color = ifelse(tx_census$label == "", "grey50", "red")) +
  scale_x_log10() +
  geom_text_repel(aes(label = label), max.overlaps = Inf, box.padding = 0.5,
                  size = 10/.pt) +
  labs(x = 'population', y = 'per capita income') +
  theme_minimal()
```