

Homework 9

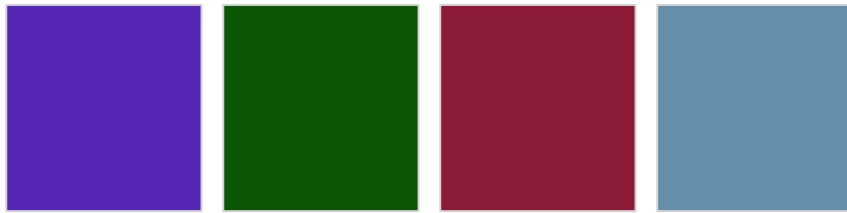
Jeremy Ulfohn // jau392

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', '#0B5505', '#8C1B37', '#658EAB')
# using all colors of the same brightness was unreadable for the scatterplot, so I changed to a lighter
swatchplot(colors)
```



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

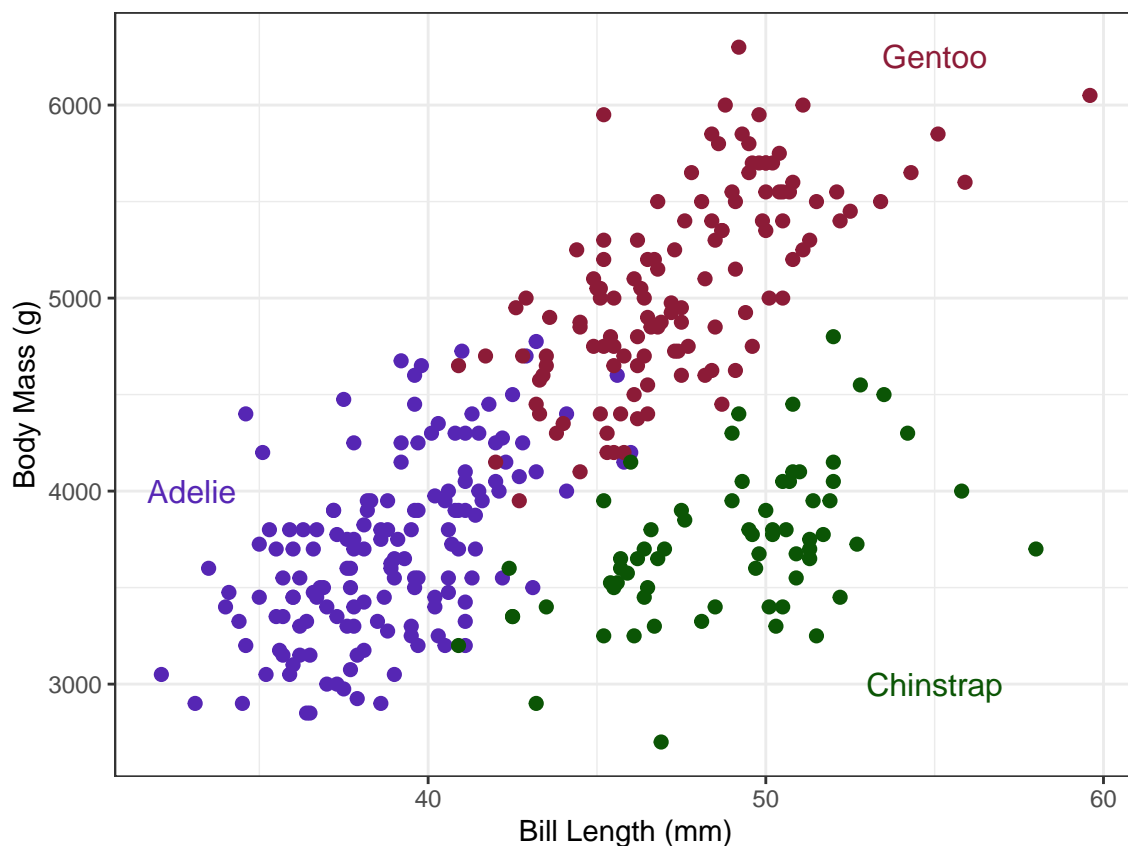
1. Use the colors you chose in Problem 1.
2. Improve the visual appearance by choosing a theme and cleaning up axis labels.
3. Remove the need for a legend by direct-labeling the points.

```
# create table with label positions: peng_labels
peng_labels <- tibble(
  species = c("Adelie", "Chinstrap", "Gentoo"),
  bill_length_mm = c(33, 55, 55),
  body_mass_g = c(4000, 3000, 6250),
  label = c("Adelie", "Chinstrap", "Gentoo"),
  hjust = c(0.5, 0.5, 0.5),
  vjust = c(0.5, 0.5, 0.5)
)
```

```

# remove legend, rename axis labels, choose theme
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)") +
  scale_y_continuous(name = "Body Mass (g)") +
  theme_bw() + # keep the grid in the background; it's unreadable without
  theme(legend.position = "none") + # use + theme(legend.position...) to remove legend from theme
  geom_text(
    data = peng_labels,
    aes(label = label, hjust = hjust, vjust = vjust),
    size = 12/.pt
  )

```



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)

# (1) subset, as in Automatic Labeling lecture (2) theme, rename axes

```

```

set.seed(6644)
tx_census %>% # randomly exclude 50% of the labels to avoid surfeit
  mutate(
    county = ifelse(runif(n()) < 0.5, '', county)
  ) %>%
  ggplot(aes(pop2010, per_capita_income)) +
  geom_point(size = 1.5) +
  geom_text_repel(
    aes(label = county),
    max.overlaps = 20,
    size = 6/.pt,
    box.padding = 0.3
  ) +
  scale_x_log10(name = "2010 Population (Log Scale)") +
  scale_y_continuous(name = "Income Per Capita") +
  theme_bw()

```

```

## Warning: ggrepel: 100 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```

