98 lines (78 sloc) | 2.74 KB

```
# Uncomment those and run them if you haven't downloaded those packages

# install.packages("tidyverse")
# install.packages("palmerpenguins")
# install.packages("gt")

# Go here for more directions on palmer penguins: https://allisonhorst.github.io/palmerpenguins/

# Load in the packages
library(tidyverse)
library(palmerpenguins)
library(gt)

# Put the dataset in your enviroment
penguins <- penguins

# Let's look at the first five rows using head()
```

```r
head(penguins)

# We can do the same thing but with a couple columns
penguins %>% select(species, island, bill_length_mm) %>% head()

# Check how many rows in the dataset
nrow(penguins)

# Check the column nanes
names(penguins)

# Clean the dataset
penguins <- penguins %>%
  filter(!is.na(sex), !is.na(bill_length_mm))

colSums(is.na(penguins))

# check the count and average bill length by island
island_stats <- penguins %>%
  group_by(island) %>%
  summarize(count = n(),
            avg_bill_len = mean(bill_length_mm))

island_stats

# make a boxplot based on island
penguins %>%
  ggplot(aes(x = island, y = bill_length_mm)) +
  geom_boxplot(aes(fill = island)) +
  theme_bw() +
  labs(x = "Island",
       y = "Bill Length (MM)",
       title = "Bill Length by Island in Palmer Penguins Dataset")

# Saving the plot
ggsave('penguin-boxplot', width = 15, height = 10, dpi = "retina")

```

```r
# make a scatter plot of bill length and flipper length
penguins %>%
  ggplot(aes(x = bill_length_mm, y = flipper_length_mm), group = species) +
  geom_point(aes(color = species, shape = species), size = 3) +
  scale_color_brewer(palette = "Set1") +
  theme_bw() +
  labs(x = "Bill Length (MM)",
       y = "Flipper Length (MM)",
       title = "Relationship Between Flipper Length and Bill Length by Species") +
  facet_wrap(~island) # taking the graph one step further

# Using mutate to make a new column
penguins_bill_flipper <- penguins %>%
  mutate(bill_to_flipper = bill_length_mm / flipper_length_mm)

# Using group_by() with two variables to set up and make a table
island_gender_stats <- penguins_bill_flipper %>%
  group_by(species, sex) %>%
  summarize(count = n(),
            avg_b_to_f = round(mean(bill_to_flipper), 2)) %>%
  ungroup() %>%
  mutate(sex = case_when(
    sex == "female" ~ "Female",
    sex == "male" ~ "Male"
  )) %>%
  arrange(-avg_b_to_f)

# Making the gt table
island_gender_stats %>%
  gt() %>%
  cols_label(species = "Species",
             sex = "Gender",
             count = "Count",
             avg_b_to_f = "Average Bill to Flipper") %>%
  cols_align(align = "center") %>%
  tab_header(title = "Average Bill to Flipper Ratio by Species and Gender") %>%
  data_color(
```

```
92      columns = vars(avg_b_to_f),
93      colors = scales::col_numeric(
94        palette = c(
95          "darkorange", "darkblue"),
96        domain = NULL))
97
98
```