 BONBUTTERY 

Exploratory Data Analysis Coding Exam

Codes Prior to Plotting:

 # Set Working Directory

 # Open Packages (after installation)

> Library(ggplot2)

> Library(RColorBrewer)

 # Read csv file

> penguin <- read.csv("penguins_size-1.csv")

 # Removed All Rows with NA values

penguin1 <- na.omit(penguin)

 # Removed Row No. 337 (because of "." in sex column)

penguin2 <- penguin1[-337,]

1.) Body Mass (g) vs Flipper Length (mm) of Penguins in Antarctica

The scatterplot below (Figure 1) shows a clear linear relationship between the body mass (in g) and flipper length (in mm) of the three species of penguins in Antarctica, but the species are not specified. It signifies that as the body mass of the penguins increases, their flipper length increases as well. Another plot below, Figure 2, shows the same relationship but this time specifying the species.

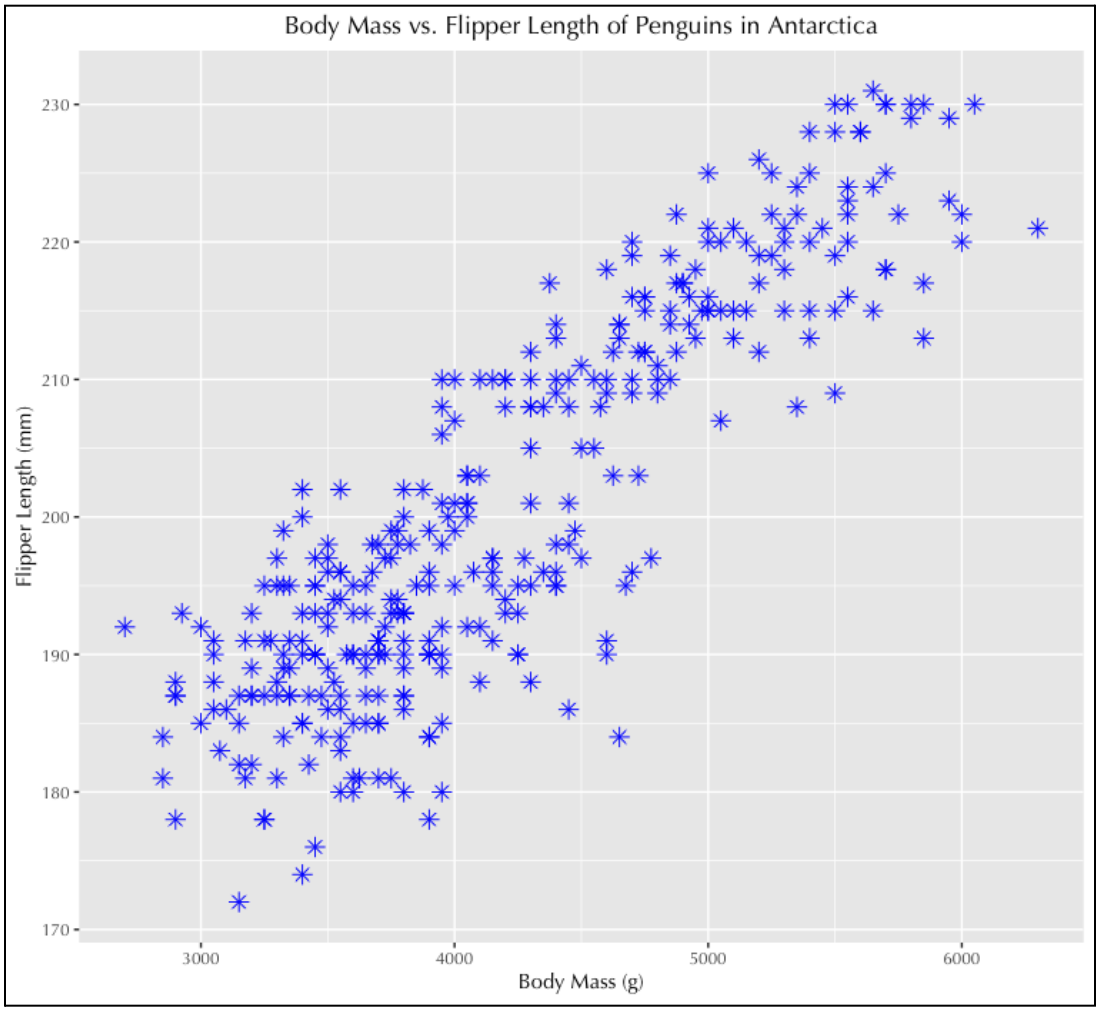

Figure 1: Scatter Plot of Body Mass vs. Flipper Length of the Penguins in Antarctica

Code: > ggplot(data = penguin1, aes(x = body_mass_g, y = flipper_length_mm, color="Blue")) + labs(x = "Body Mass (g)", y = "Flipper Length (mm)") + theme(legend.position = "none") + ggtitle("Body Mass vs. Flipper Length of Penguins in Antarctica") + theme(plot.title = element_text(hjust = 0.5)) + theme(text = element_text(family = "Optima")) + geom_point(size = 3, shape = 8, color="Blue")

2.) Body Mass (g) vs Flipper Length (mm) of Penguins in Antarctica

The scatterplot, this time, specifies the penguin species that tend to have the highest increase in flipper length when their body mass also increases, which is the Gentoo species. This also suggests that this species is the heaviest among the three species. Furthermore, the remaining two species, Adelie and Chinstrap, are more or less on the same level.

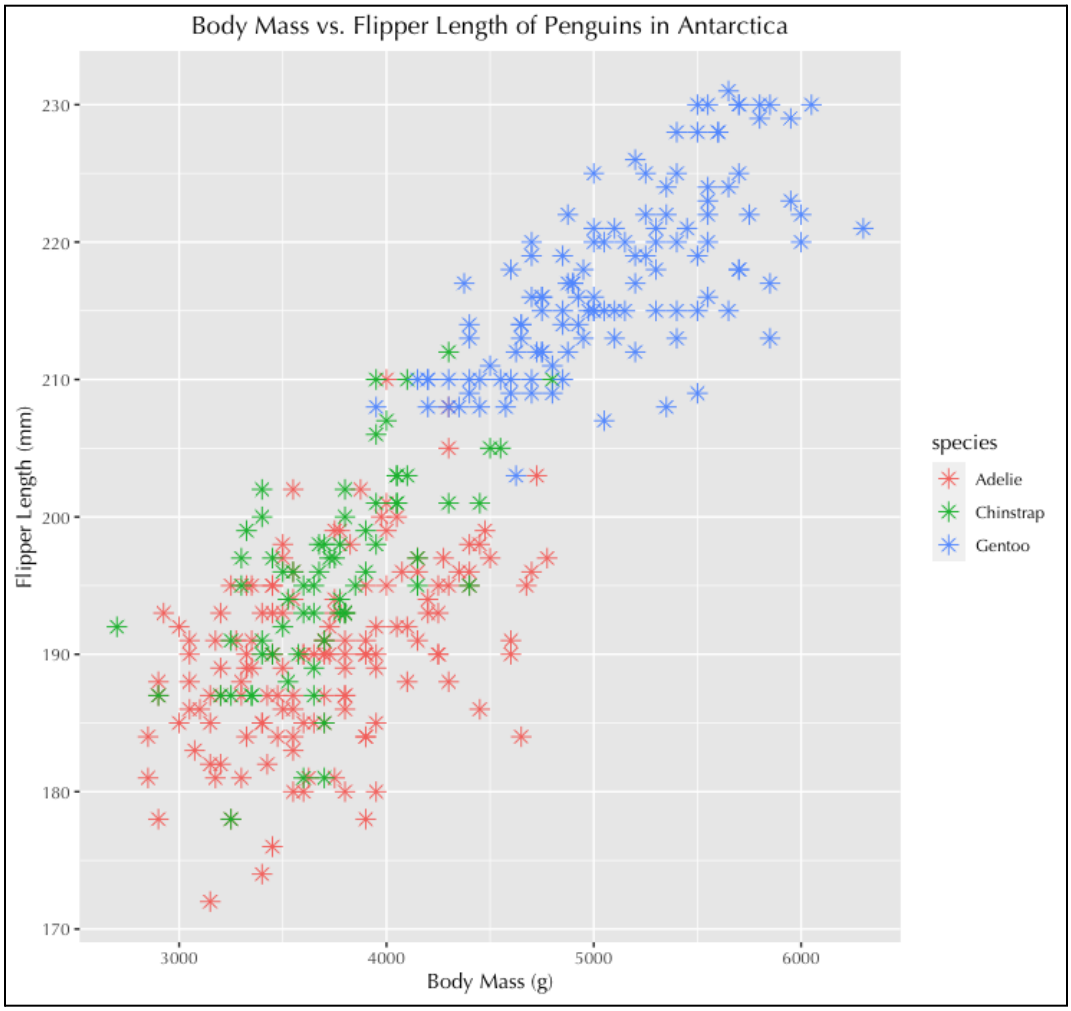

Figure 2: Scatter Plot of Body Mass vs. Flipper Length of the three Specified Penguin Species in Antarctica

Code: > ggplot(data = penguin1, aes(x = body_mass_g, y = flipper_length_mm, color=species)) + labs(x = "Body Mass (g)", y = "Flipper Length (mm)") + ggtitle("Body Mass vs. Flipper Length of Penguins in Antarctica") + theme(plot.title = element_text(hjust = 0.5)) + theme(text = element_text(family = "Optima")) + geom_point(size = 3, shape = 8)

3.) Body Mass (g) of Penguin Species in Different Islands

Figure 3 below shows the comparison of the Body Mass (in g) of the different penguin species in Antarctica. From the three species, it is Gentoo that has the highest body mass, which is in line with the previous figure's output, wherein it is also Gentoo that has the highest body mass. Furthermore, in this same plot, it is noticeable that in some islands there is an absence of certain penguin species, such as Chinstrap in the islands of Torgersen and Biscoe, and Gentoo in Dream and Torgersen islands. It is only the Adelie species that can be found on all three islands. Another figure below, Figure 4, confirms our notion as it also shows similar results in terms of the presence and absence of certain species on a particular island.

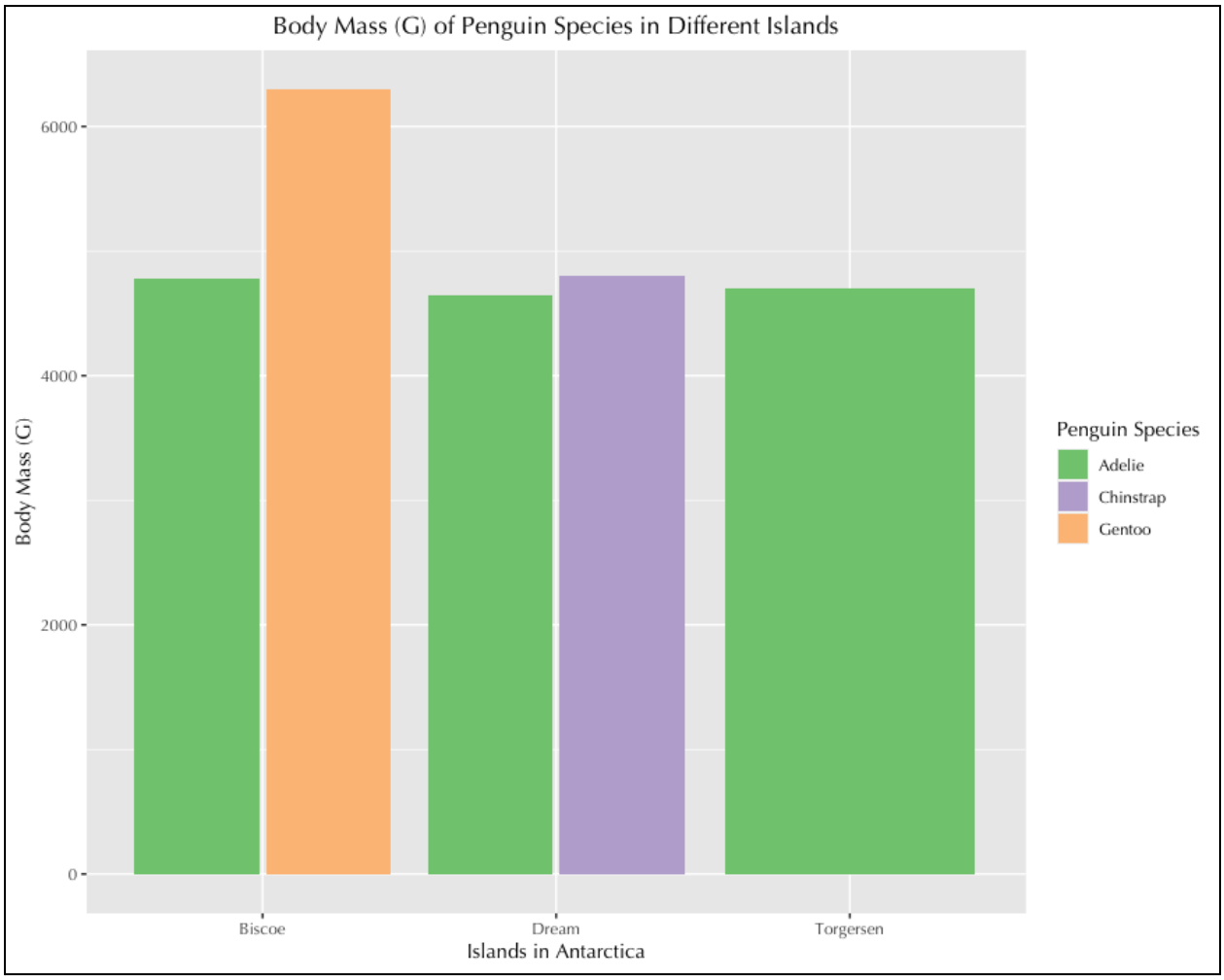

Figure 3: Grouped Bar Plot of Body Masses of the 3 different Penguin Species in Different Antarctic Islands

Code: > ggplot(data = penguin2, aes(x = island, y = body_mass_g, fill = species))+ geom_bar(position=position_dodge(0.9), width=0.85, stat="identity") + theme(text = element_text(family = "Optima")) + ggtitle("Body Mass (G) of Penguin Species in Different Islands") + theme(plot.title = element_text(hjust = 0.5)) + scale_fill_brewer(palette="Accent") + labs(x="Islands in Antarctica", y="Body Mass (G)")

4.) Frequency of Penguin Species in Different Antarctic Islands

As previously introduced in Figure 3, the bar plot below shows the counts of sampled penguins in Antarctica, and it confirms the previous idea that other penguin species are not present on other islands. Additionally, on the island of Biscoe, where there were no recorded Chinstrap species, Gentoo species dominate the island with the highest individual count. On Dream Island, it is the other way around, with no Gentoo species, and their number is also higher than the Adelie.

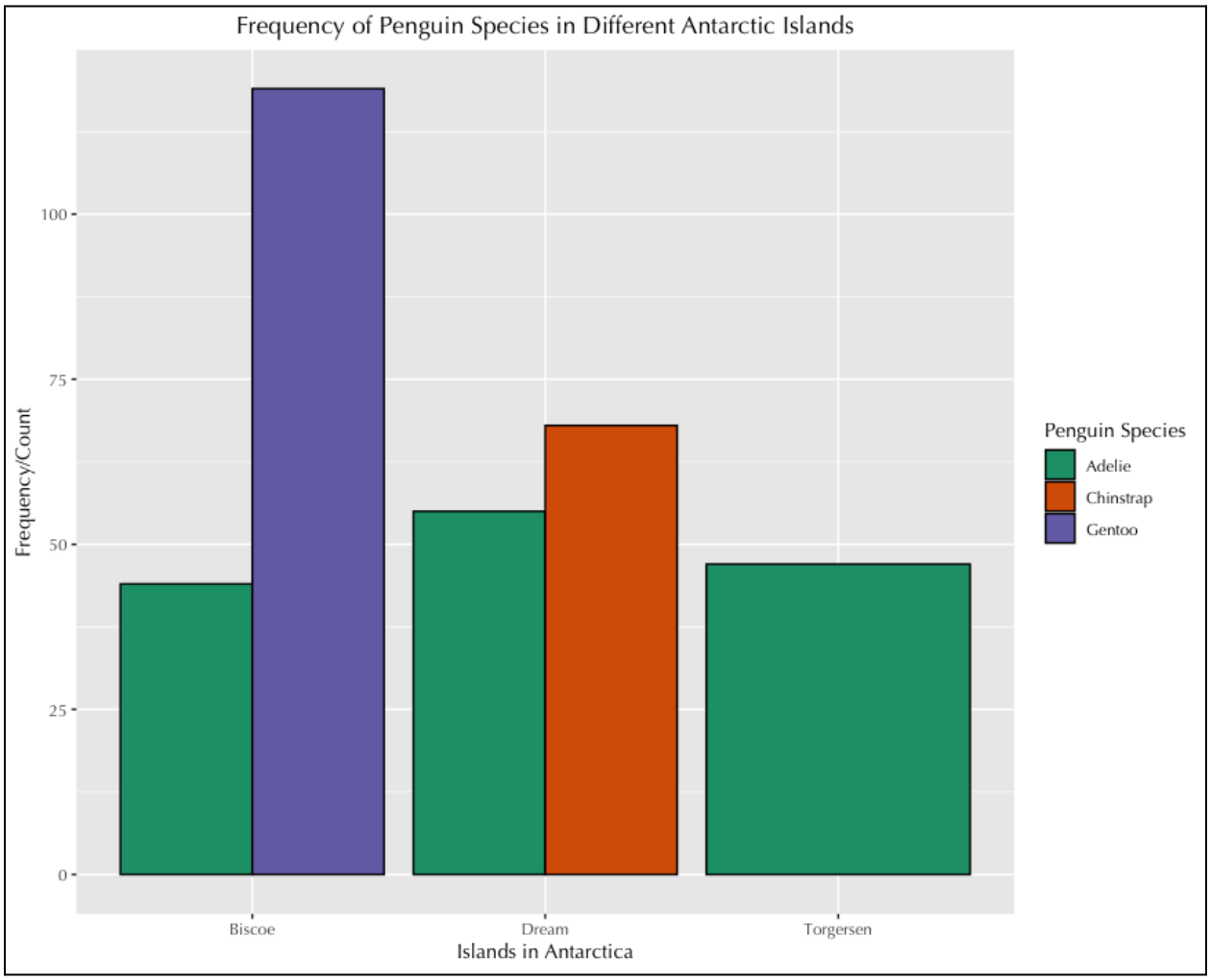

Figure 4: Grouped Bar Plot of the Frequency of Penguin Species in Different Antarctic Islands

Code: > ggplot(data = penguin2, aes(x=island, fill=species)) + geom_bar(position = "dodge", color="black") + ggtitle("Frequency of Penguin Species in Different Antarctic Islands") + theme(plot.title = element_text(hjust = 0.5)) + scale_fill_brewer(palette="Dark2", name = "Penguin Species") + labs(x="Islands in Antarctica", y="Frequency/Count") + theme(text = element_text(family = "Optima"))

5.) Comparison of Bill Length of Penguin Species in Antarctica

In this figure, Figure 5, the bill lengths of the three penguin species are compared using a boxplot. It is found that the Chinstrap species possesses the highest mean bill length among the three species, Adelie with the lowest mean, and Gentoo just closely behind Chinstrap. However, to confirm our visual assumptions, more specific statistical analysis is necessary.

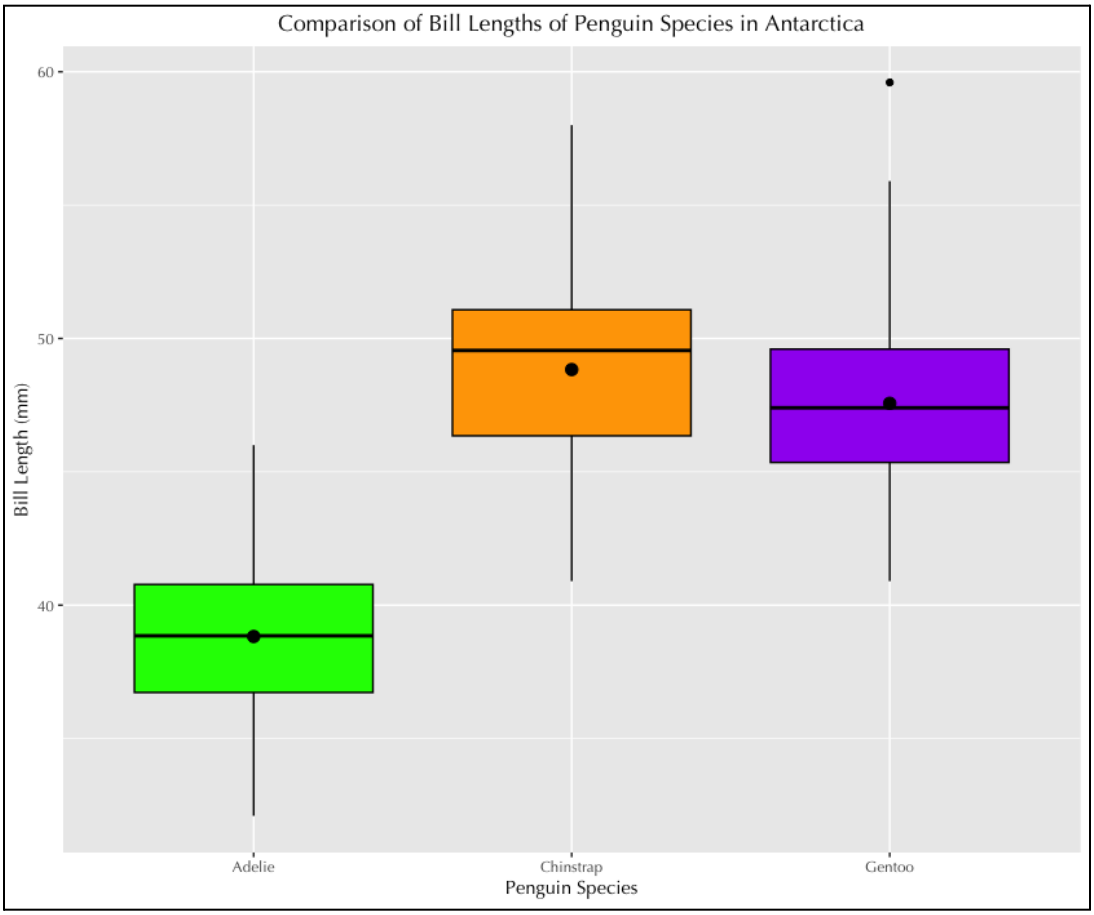

Figure 5: Boxplot Comparing the Bill Lengths of the Three Penguin Species in Antarctic

Code: > ggplot(data = penguin2, aes(x = species, y=bill_length_mm)) + labs(x = "Penguin Species", y = "Bill Length (mm)") + theme(legend.position = "none") + ggtitle("Comparison of Bill Lengths of Penguin Species in Antarctica") + theme(plot.title = element_text(hjust = 0.5)) + theme(text = element_text(family = "Optima")) + geom_boxplot(col="Black", fill=c("Green", "Orange", "Purple")) + stat_summary(fun=mean, geom='point', shape=19, size=3)

For the purpose of improving and doing a more thorough research, this data analysis of the Penguin dataset is available for revisions and comments. You are welcome to modify, analyze, and utilize these visual analyses however you like. Simply remember to give credit :)))