# Midterm 1 W26

**Vaishnavi Maganti**

2026-01-29

# Instructions

Answer the following questions and complete the exercises in RMarkdown. Please embed all of your code and push your final work to your repository. Your code must be organized, clean, and run free from errors. Remember, you must remove the `#` for any included code chunks to run. Be sure to add your name to the author header above.

Your code must knit in order to be considered. If you are stuck and cannot answer a question, then comment out your code and knit the document. You may use your notes, labs, and homework to help you complete this exam. Do not use any other resources- including AI assistance or other students' work.

Don't forget to answer any questions that are asked in the prompt! Each question must be coded; it cannot be answered by a sort in a spreadsheet or a written response only.

For all plots you create, a title and clearly labeled axes must be provided. We also expect pipes `%>%` to be used wherever possible.

Be sure to push your completed midterm to your repository and upload the document to Gradescope. This exam is worth 50 points.

Please load the following libraries.

```
library(tidyverse)
library(janitor)
```

# Part 1: Repository

**Question 1. (3 points) Before you start analyzing data, please put a link to your GitHub repository below. Your repository should have a clear README and be well-organized. Add `jmledford3115` and `bryshalm` as collaborators to your repository if you haven't already done so.**

Link to repository:

https://github.com/vmagiant-coder/BIS15L_W26_vmaganti (https://github.com/vmagiant-coder/BIS15L_W26_vmaganti)

# Part 2: Data and Analysis

In the midterm 1 folder there is a second folder called `data`. Inside the `data` folder, there is a .csv file called `anolis_dat.csv`. These data came from D. Luke Mahler, Liam J. Revell, Richard E. Glor, Jonathan B. Losos, ECOLOGICAL OPPORTUNITY AND THE RATE OF MORPHOLOGICAL EVOLUTION IN THE DIVERSIFICATION OF GREATER ANTILLEAN ANOLES, Evolution, Volume 64, Issue 9, 1 September 2010, Pages 2731–2745 (https://academic.oup.com/evolut/article/64/9/2731/6854302?login=true). The original research article is included in the data folder.

*Anolis* is a genus of lizards commonly known as anoles. Anoles are found throughout the Americas, but are especially diverse in the Caribbean. The data include morphological measurements for *Anolis* lizards from the islands of the Greater Antilles. These data can be used to study patterns of morphological evolution and adaptation in *Anolis* lizards.

The variables include:
- `species` : Species name of the anole lizard.
- `habitat` : Habitat type where the lizard was found.
- `hindlimb_length_mm` : Length of the lizard's hindlimbs (in millimeters).
- `tail_length_mm` : Length of the lizard's tail (in millimeters).
- `body_length_mm` : Length of the lizard's body (in millimeters).
- `toepad_lamellae_count` : Count of lamellae on the lizard's toepads.
- `island` : Island where the lizard was found.

**Question 2. (2 points) Load the data and store it as an object called `anolis`.**

```
anolis <- read_csv("data/anolis_dat.csv")
```

```
## Rows: 52 Columns: 7
## ── Column specification ──────────────────────────────────────────────
## Delimiter: ","
## chr (3): Species, Habitat, Island
## dbl (4): Hindlimb length (mm), Tail length (mm), Body length (mm), Toepad la...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

**Question 3. (2 points) Use a summary function of your choice to get an idea of the structure of the data.**

```
glimpse(anolis)
```

```
## Rows: 52
## Columns: 7
## $ Species                <chr> "A. ahli", "A. alayoni", "A. alfaroi", "A. a…
## $ Habitat                <chr> "Trunk-ground", "Twig", "Grass-bush", "Trunk…
## $ `Hindlimb length (mm)`   <dbl> 50.46, 25.50, 26.17, 36.80, 50.39, 49.37, 29…
## $ `Tail length (mm)`       <dbl> 81.99, 54.75, 79.00, 84.88, 154.45, 91.01, 1…
## $ `Body length (mm)`       <dbl> 51.67, 41.32, 30.95, 51.53, 72.32, 51.72, 32…
## $ `Toepad lamellae (count)` <dbl> 27, 31, 24, 36, 41, 28, 29, 28, 28, 31, 32, …
## $ Island                 <chr> "Cuba", "Cuba", "Cuba", "Hispaniola", "Cuba"…
```

**Question 4. (2 points) Clean the variable names so they are all lowercase and without special characters or spaces. Be sure to use the cleaned data for all subsequent analyses.**

```
anolis <- anolis %>%
  clean_names()
```

```
names(anolis)
```

```
## [1] "species"              "habitat"             "hindlimb_length_mm"
## [4] "tail_length_mm"       "body_length_mm"      "toepad_lamellae_count"
## [7] "island"
```

**Question 5. (4 points) Convert the `habitat` and `island` variables to factors.**

```
anolis$habitat <- as.factor(anolis$habitat)
```

```
anolis$island <- as.factor(anolis$island)
```

**Question 6. (2 points) Anole species were sampled from multiple islands. Which islands are represented in the data? Display the island names.**

```
anolis %>%
  select(island) %>%
distinct(island)
```
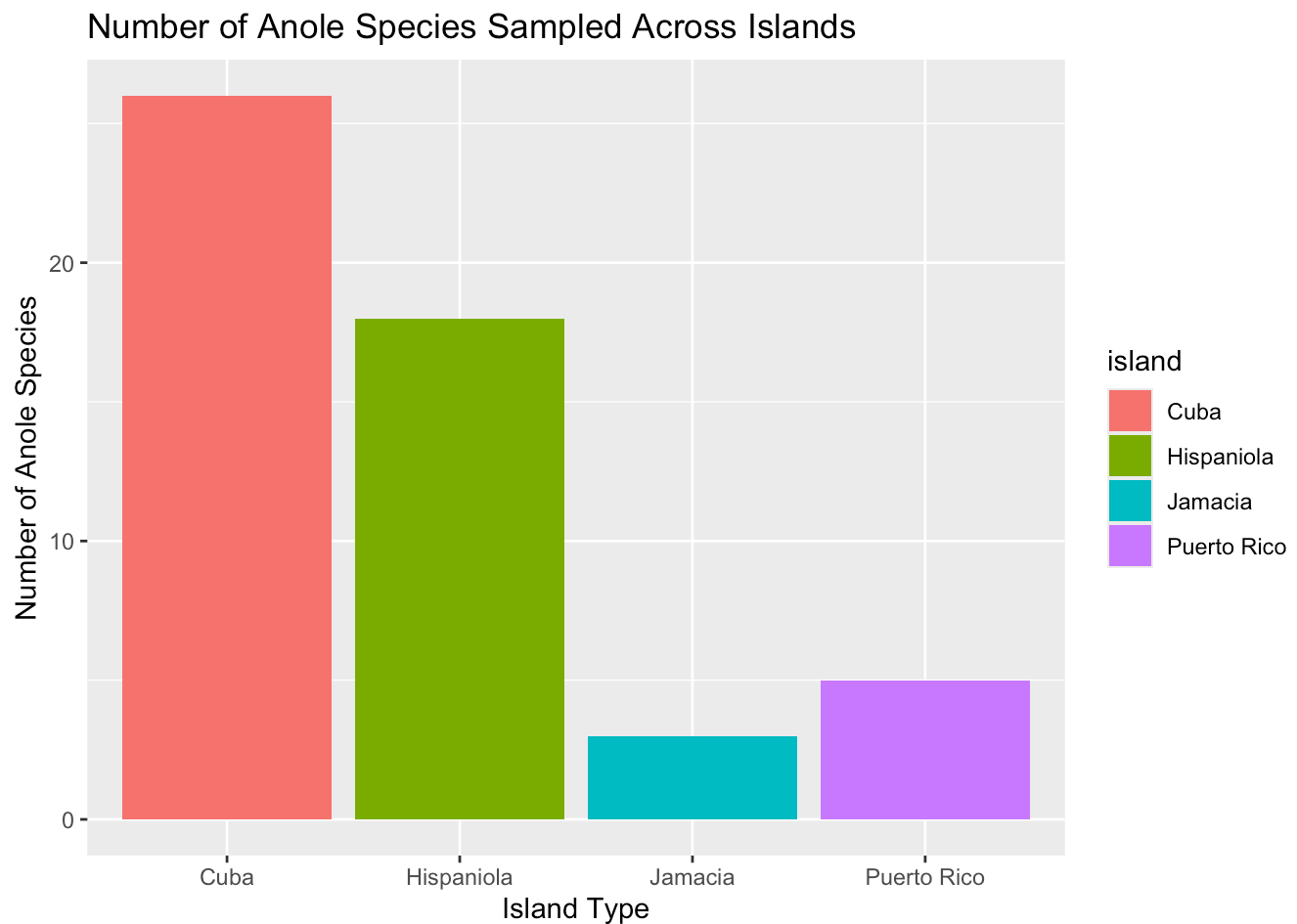
```
## # A tibble: 4 × 1
##   island
##   <fct>
## 1 Cuba
## 2 Hispaniola
## 3 Puerto Rico
## 4 Jamacia
```

**Question 7. (4 points) Is sampling equal across islands? Create a plot to visualize the number of anole species sampled from each island. Be sure to label your axes and add a title.**

```
names(anolis)
```

```
## [1] "species"              "habitat"             "hindlimb_length_mm"
## [4] "tail_length_mm"       "body_length_mm"      "toepad_lamellae_count"
## [7] "island"
```

```
anolis %>%
  select(island, species) %>%
  ggplot(aes(x=island)) +
  geom_bar(aes(fill=island)) +
  labs(title="Number of Anole Species Sampled Across Islands", x="Island Type", y="Numbe
r of Anole Species")
```

## Number of Anole Species Sampled Across Islands



Based on the graph in Question 8, sampling is not equal across islands with Cuba having a larger majority of sampling as opposed to Jamaica which is in minority. Hispaniola follows Cuba as a close second, with Puerto Rico being 3rd and Jamaica being 4th.

**Question 8. (2 points) Which habitat types are represented in the data? Display the names of the habitat types.**

```
anolis %>%
  select(habitat) %>%
  distinct(habitat)
```
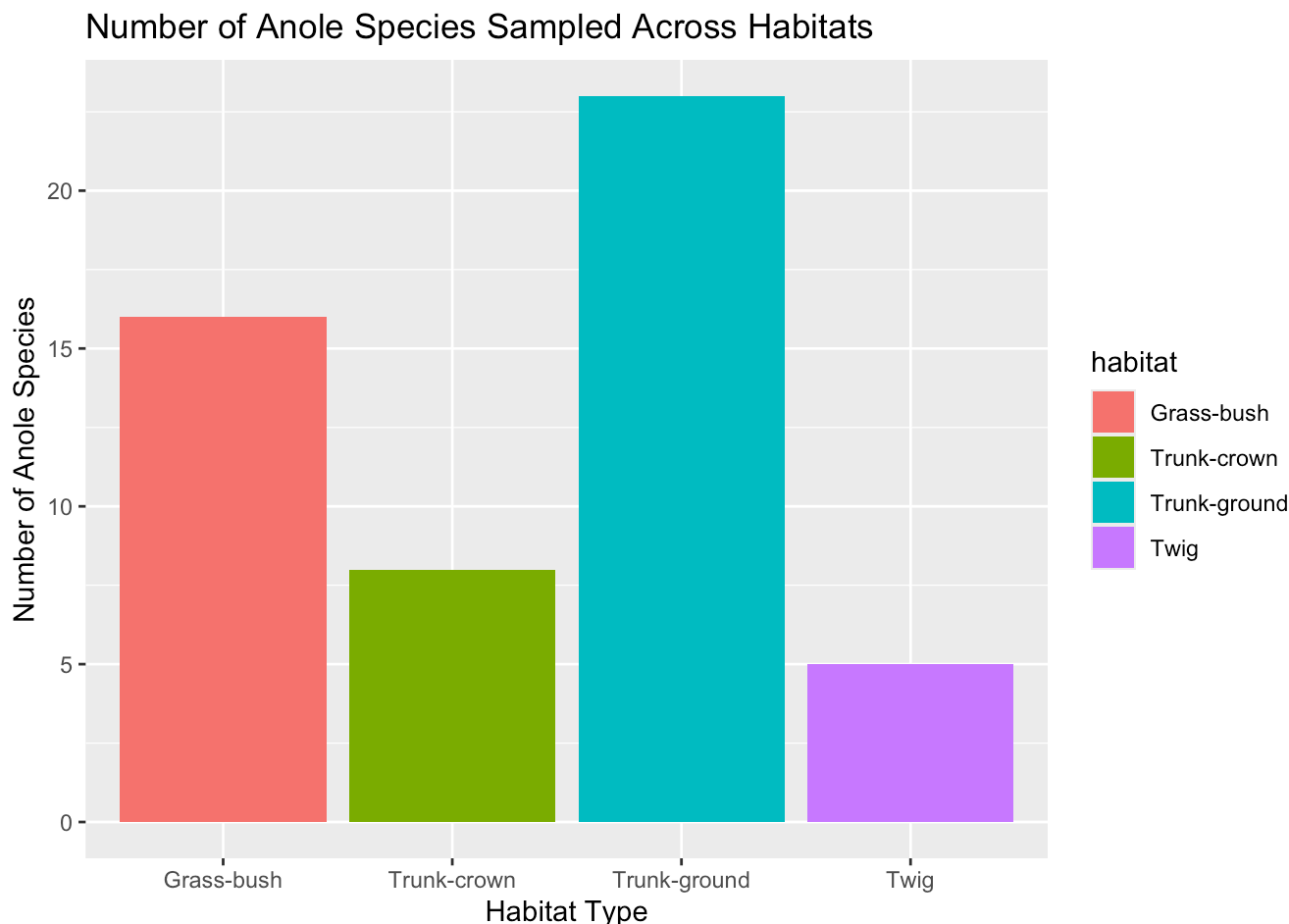
```
## # A tibble: 4 × 1
##   habitat
##   <fct>
## 1 Trunk-ground
## 2 Twig
## 3 Grass-bush
## 4 Trunk-crown
```

**Question 9. (4 points) Is sampling equal across habitat types? Create a plot to visualize the number of anole species sampled from each habitat type. Be sure to label your axes and add a title.**

```
anolis %>%
  select(habitat, species) %>%
  ggplot(aes(x=habitat)) +
  geom_bar(aes(fill=habitat)) +
  labs(title="Number of Anole Species Sampled Across Habitats", x="Habitat Type", y="Num
ber of Anole Species")
```

## Number of Anole Species Sampled Across Habitats



Based on the graph in Question 9, sampling across habitats is once again not equal as the trunk-ground habitat once again takes prevalence in regards to the highest number of anole species sampled. Grass-bush is a close second, followed by trunk-crown and then the twig habitat.

**Question 10. (4 points) The morphology of anoles varies based on their habitat. How does the range of hindlimb length compare among different habitats? Create a plot to visualize the distribution of hindlimb lengths across habitat types. Be sure to label your axes and add a title.**

```
names(anolis)
```
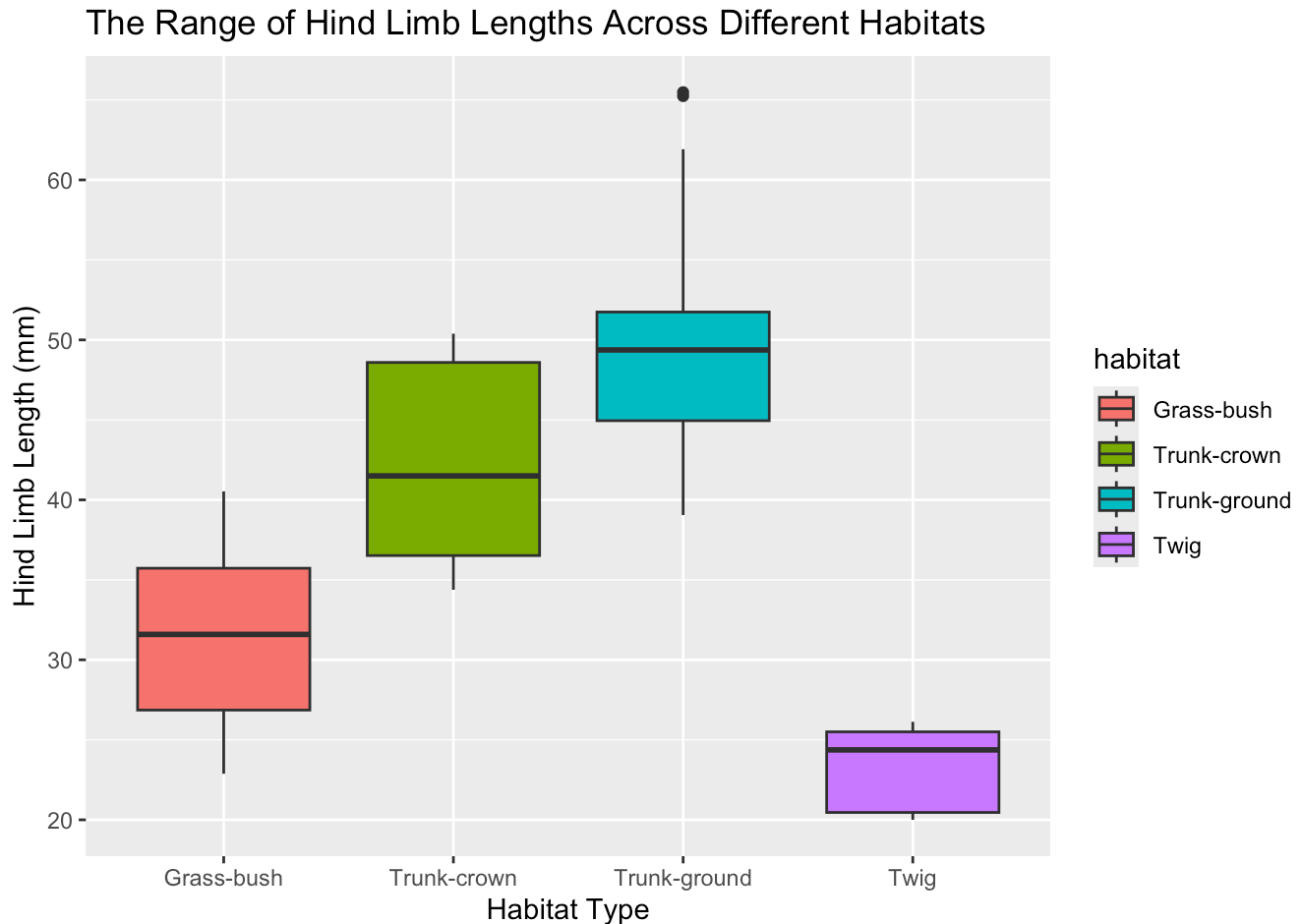
```
## [1] "species"          "habitat"           "hindlimb_length_mm"
## [4] "tail_length_mm"   "body_length_mm"    "toepad_lamellae_count"
## [7] "island"
```

```
anolis %>%
  select(habitat, hindlimb_length_mm) %>%
  ggplot(aes(x=habitat, y=hindlimb_length_mm)) +
  geom_boxplot(aes(fill=habitat)) +
  labs(title="The Range of Hind Limb Lengths Across Different Habitats", x="Habitat Typ
e", y="Hind Limb Length (mm)")
```

## The Range of Hind Limb Lengths Across Different Habitats



Based on the graph for Question 10, the ranges of hind limb length across different habitats distinctly varies. The trunk-crown habitat has a larger range in comparison to the other habitats, with the twig habitat having the smallest range along with average.

**Question 11. (4 points) The plot above is compelling, but don't we expect larger lizards to have longer limbs? What about tail length? Shouldn't longer lizards have longer tails? To correct for this, make two new columns: 1. `ratio_of_hindlimb_to_body`, and 2. `ratio_of_tail_to_body`. Don't forget to add these columns to the anolis data frame.**

```
anolis <- anolis %>%
  mutate(ratio_of_hindlimb_to_body=hindlimb_length_mm/body_length_mm)
anolis
```

```
## # A tibble: 52 × 8
##    species       habitat       hindlimb_length_mm tail_length_mm body_length_mm
##    <chr>         <fct>                      <dbl>          <dbl>          <dbl>
##  1 A. ahli       Trunk-ground               50.5           82.0           51.7
##  2 A. alayoni    Twig                       25.5           54.8           41.3
##  3 A. alfaroi    Grass-bush                 26.2           79             31.0
##  4 A. aliniger   Trunk-crown                36.8           84.9           51.5
##  5 A. allisoni   Trunk-crown                50.4          154.            72.3
##  6 A. allogus    Trunk-ground               49.4           91.0           51.7
##  7 A. alumina    Grass-bush                 30.0          106.            32.9
##  8 A. alutaceus  Grass-bush                 27.4           94.6           31.8
##  9 A. angusticeps Twig                      24.4           65.1           40.2
## 10 A. armouri    Trunk-ground               51.8          101.            56.1
## # ℹ 42 more rows
## # ℹ 3 more variables: toepad_lamellae_count <dbl>, island <fct>,
## #   ratio_of_hindlimb_to_body <dbl>
```
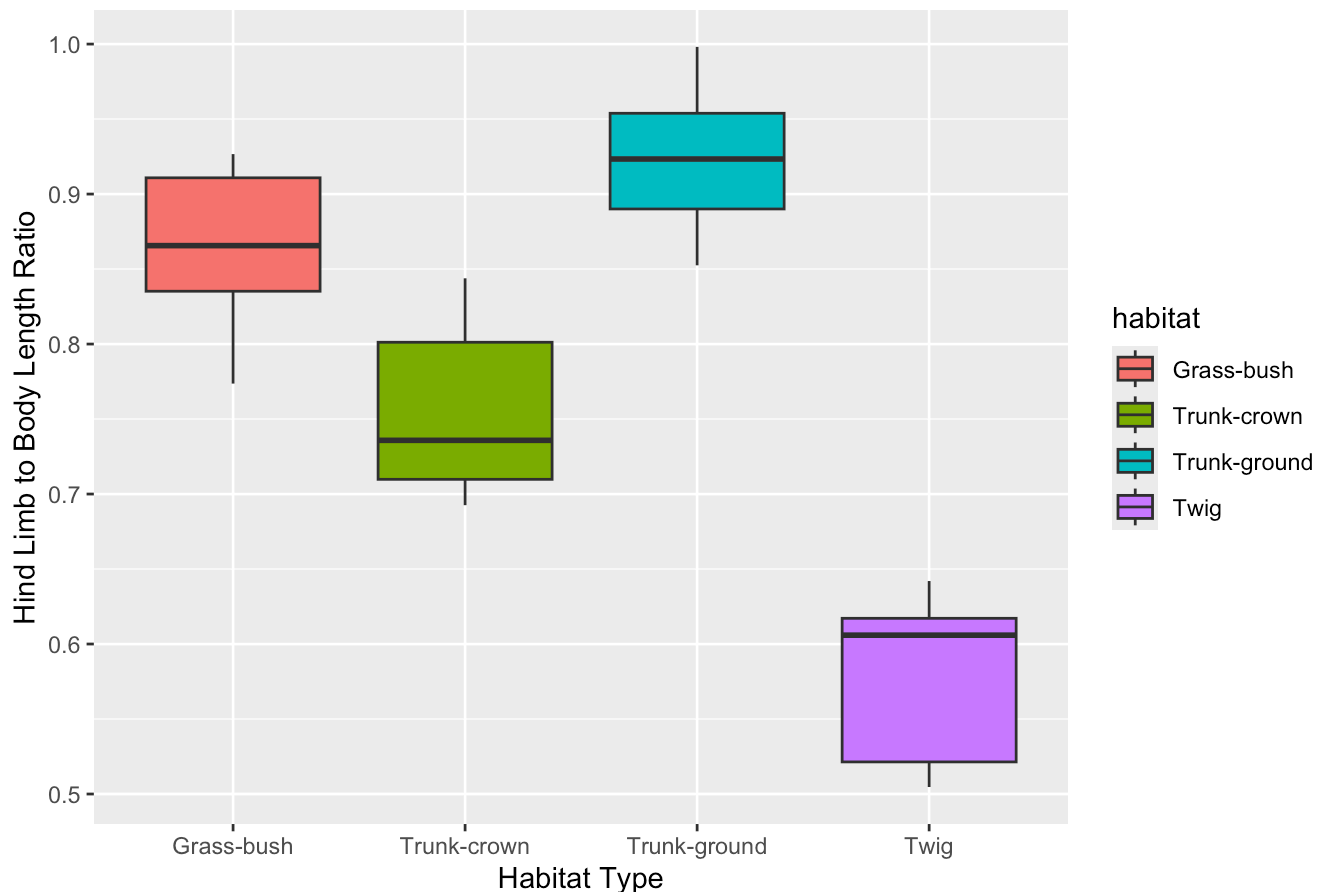
```
anolis <- anolis %>%
  mutate(ratio_of_tail_to_body=tail_length_mm/body_length_mm)
anolis
```

```
## # A tibble: 52 × 9
##    species       habitat       hindlimb_length_mm tail_length_mm body_length_mm
##    <chr>         <fct>                      <dbl>          <dbl>          <dbl>
##  1 A. ahli       Trunk-ground               50.5           82.0           51.7
##  2 A. alayoni    Twig                       25.5           54.8           41.3
##  3 A. alfaroi    Grass-bush                 26.2           79             31.0
##  4 A. aliniger   Trunk-crown                36.8           84.9           51.5
##  5 A. allisoni   Trunk-crown                50.4          154.            72.3
##  6 A. allogus    Trunk-ground               49.4           91.0           51.7
##  7 A. alumina    Grass-bush                 30.0          106.            32.9
##  8 A. alutaceus  Grass-bush                 27.4           94.6           31.8
##  9 A. angusticeps Twig                      24.4           65.1           40.2
## 10 A. armouri    Trunk-ground               51.8          101.            56.1
## # ℹ 42 more rows
## # ℹ 4 more variables: toepad_lamellae_count <dbl>, island <fct>,
## #   ratio_of_hindlimb_to_body <dbl>, ratio_of_tail_to_body <dbl>
```

**Question 12. (4 points) Create a new plot that examines the distribution of `ratio_of_hindlimb_to_body` across habitat types. How does this plot differ from the one you made in Problem 10? Be sure to label your axes and add a title.**

```
anolis %>%
  select(habitat, ratio_of_hindlimb_to_body) %>%
  ggplot(aes(x=habitat, y=ratio_of_hindlimb_to_body)) +
  geom_boxplot(aes(fill=habitat)) +
  labs(title="The Ratio of Hind Limb to Body Length Across Different Habitats", x="Habit
at Type", y="Hind Limb to Body Length Ratio")
```

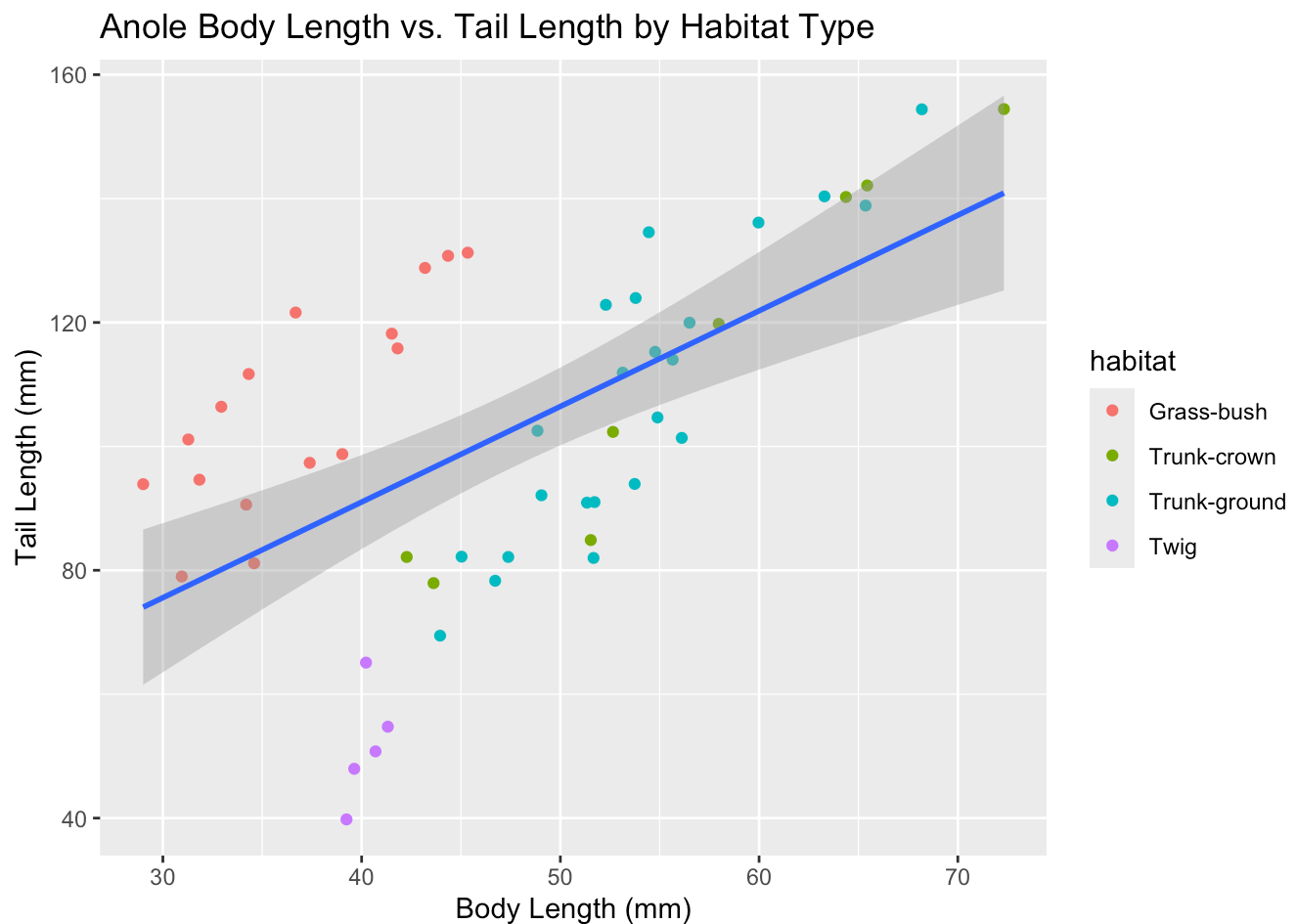## The Ratio of Hind Limb to Body Length Across Different Habitats



The graph made in Question 12 differs from the graph made in Question 10 because the grass-bush habitat range now surpasses the trunk-crown habitat range in regards to hind limb to body length ratio. There are also two different dependent variables being used, with Question 10 using just hind limb length in mm and Question 12 using the ratio between hind limb length and body length.

**Problem 13. (4 points) A longer tail provides better balance and agility. Create a plot that examines the relationship between body length and tail length. Color the points by habitat type and add a line of best fit. What does this plot suggest about the relationship between body length and tail length? What do you notice about lizards in the `Grass-bush` habitat? Be sure to label your axes and add a title.**

```
anolis %>%
  select(habitat, body_length_mm, tail_length_mm) %>%
  ggplot(aes(x=body_length_mm, y=tail_length_mm)) +
  geom_point(aes(color=habitat)) +
  geom_smooth(method=lm, se=T) +
  labs(title="Anole Body Length vs. Tail Length by Habitat Type", x="Body Length (mm)",
y="Tail Length (mm)")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

## Anole Body Length vs. Tail Length by Habitat Type



The plot in Question 13 suggests that there is a moderate (not tightly correlated) relationship between body length and tail length, but overall their relationship does follow a positive trend. The lizards in the grass-bush habitat tend to have a stronger positive relationship between body length and tail length than the other habitats because they follow a tighter positive trend that is steadily increasing.

**Problem 14. (4 points) Toepad lamellae are transverse, plate-like structures found on the ventral surface of the digits. They are a key adaptation that allows anoles to cling to and move efficiently on smooth and vertical surfaces. What is the mean number of toepad lamellae for each habitat type?**

```
names(anolis)
```

```
## [1] "species"                "habitat"
## [3] "hindlimb_length_mm"     "tail_length_mm"
## [5] "body_length_mm"         "toepad_lamellae_count"
## [7] "island"                 "ratio_of_hindlimb_to_body"
## [9] "ratio_of_tail_to_body"
```
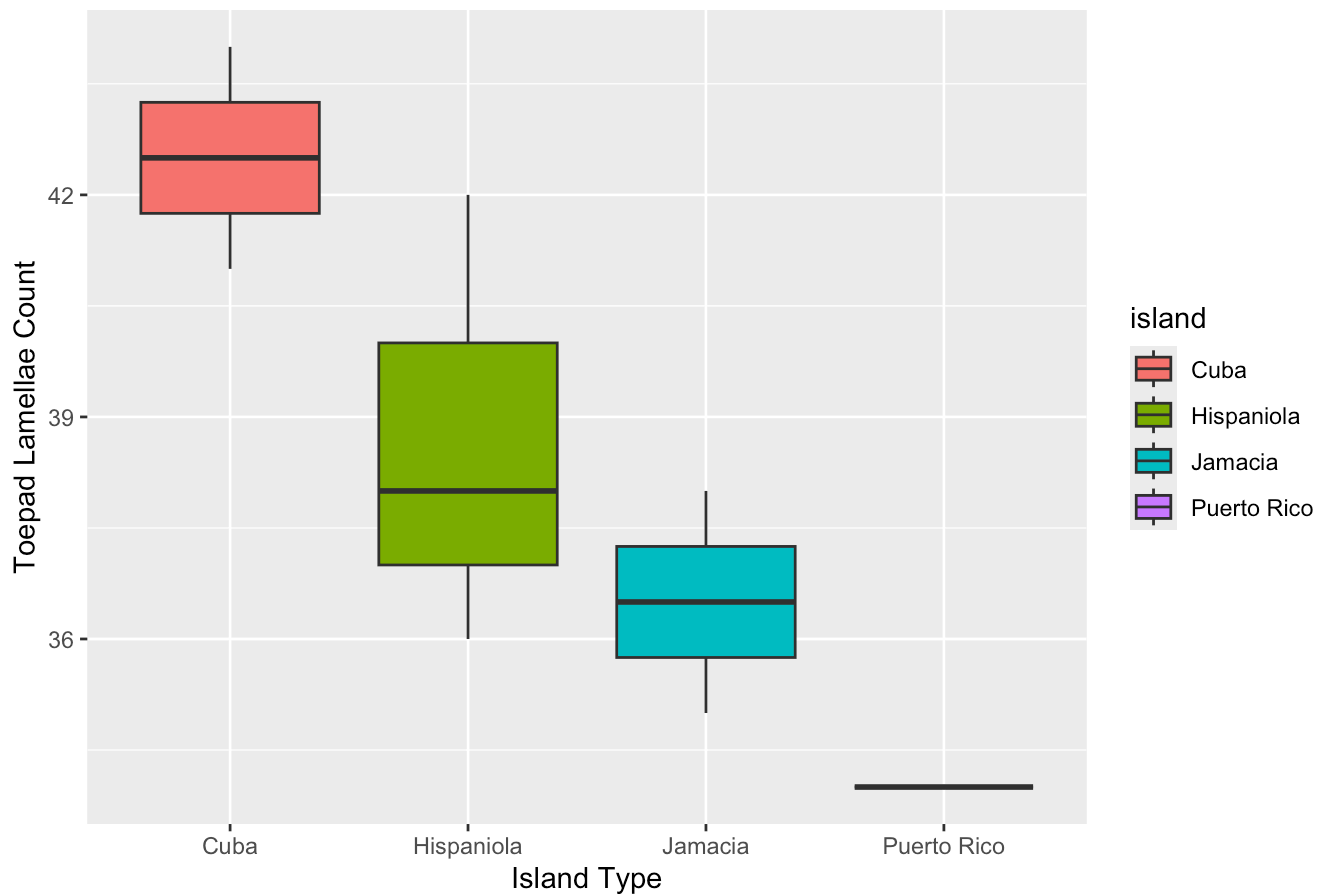
```
anolis %>%
  group_by(habitat) %>%
  summarize(mean_toepad_lamellae=mean(toepad_lamellae_count))
```

```
## # A tibble: 4 × 2
##   habitat      mean_toepad_lamellae
##   <fct>                      <dbl>
## 1 Grass-bush                  28.3
## 2 Trunk-crown                 38.5
## 3 Trunk-ground                30.0
## 4 Twig                        27.6
```

**Problem 15. (5 points) The number of toepad lamellae is significantly different for trunk-crown species. But, is this consistent across all islands? Make a plot that shows the range in number of toepad lamellae by island for trunk-crown species only. Be sure to label your axes and add a title.**

```
anolis %>%
  select(toepad_lamellae_count, island, habitat) %>%
  filter(habitat=="Trunk-crown") %>%
  ggplot(aes(x=island, y=toepad_lamellae_count)) +
  geom_boxplot(aes(fill=island)) +
  labs(title="The Distribution of Toepad Lamellae by Island for Trunk-Crown Anoles", x
="Island Type", y="Toepad Lamellae Count")
```



The Distribution of Toepad Lamellae by Island for Trunk-Crown Anoles

Based on the graph in Question 15, the distribution of toepad lamellae for trunk-crown lizards varies across each island type and is not consistent.

# Submit the Midterm

1. Save your work and knit the .rmd file.
2. Open the .html file and "print" it to a .pdf file in Google Chrome (not Safari).
3. Go to the class Canvas page and open Gradescope.
4. Submit your .pdf file to the midterm assignment- be sure to assign the pages to the correct questions.
5. Commit and push your work to your repository.