

# Midterm 1 W26

Weifan Liu

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 `bryshal` as collaborators to your repository if you haven't already done so.**

Link to repository: [https://github.com/eva2005099/BIS15W2026\\_wliu.git](https://github.com/eva2005099/BIS15W2026_wliu.git)  
([https://github.com/eva2005099/BIS15W2026\\_wliu.git](https://github.com/eva2005099/BIS15W2026_wliu.git))

## 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.**

```
summary(anolis)
```

```
##   Species           Habitat      Hindlimb length (mm) Tail length (mm)
## Length:52      Length:52      Min.   :20.00      Min.    : 39.79
## Class :character Class :character 1st Qu.:31.70      1st Qu.: 82.19
## Mode  :character Mode  :character Median :40.01      Median :101.86
##                                     Mean  :40.51      Mean   :102.81
##                                     3rd Qu.:49.27      3rd Qu.:121.92
##                                     Max.   :65.48      Max.    :154.45
## Body length (mm) Toepad lamellae (count) Island
## Min.   :29.01    Min.   :23.00      Length:52
## 1st Qu.:40.07    1st Qu.:28.00      Class :character
## Median :47.05    Median :29.00      Mode  :character
## Mean   :47.65    Mean   :30.54
## 3rd Qu.:54.53    3rd Qu.:33.00
## Max.   :72.32    Max.   :44.00
```

```
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 <- clean_names(anolis)
```

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

```
names(anolis)
```

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

```
anolis <-
  anolis %>%
  mutate(across(c(habitat, island), as.factor))
```

```
levels(anolis$habitat)
```

```
## [1] "Grass-bush" "Trunk-crown" "Trunk-ground" "Twig"
```

```
levels(anolis$island)
```

```
## [1] "Cuba"      "Hispaniola" "Jamacia"    "Puerto Rico"
```

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

```
anolis %>%
  distinct(island)
```

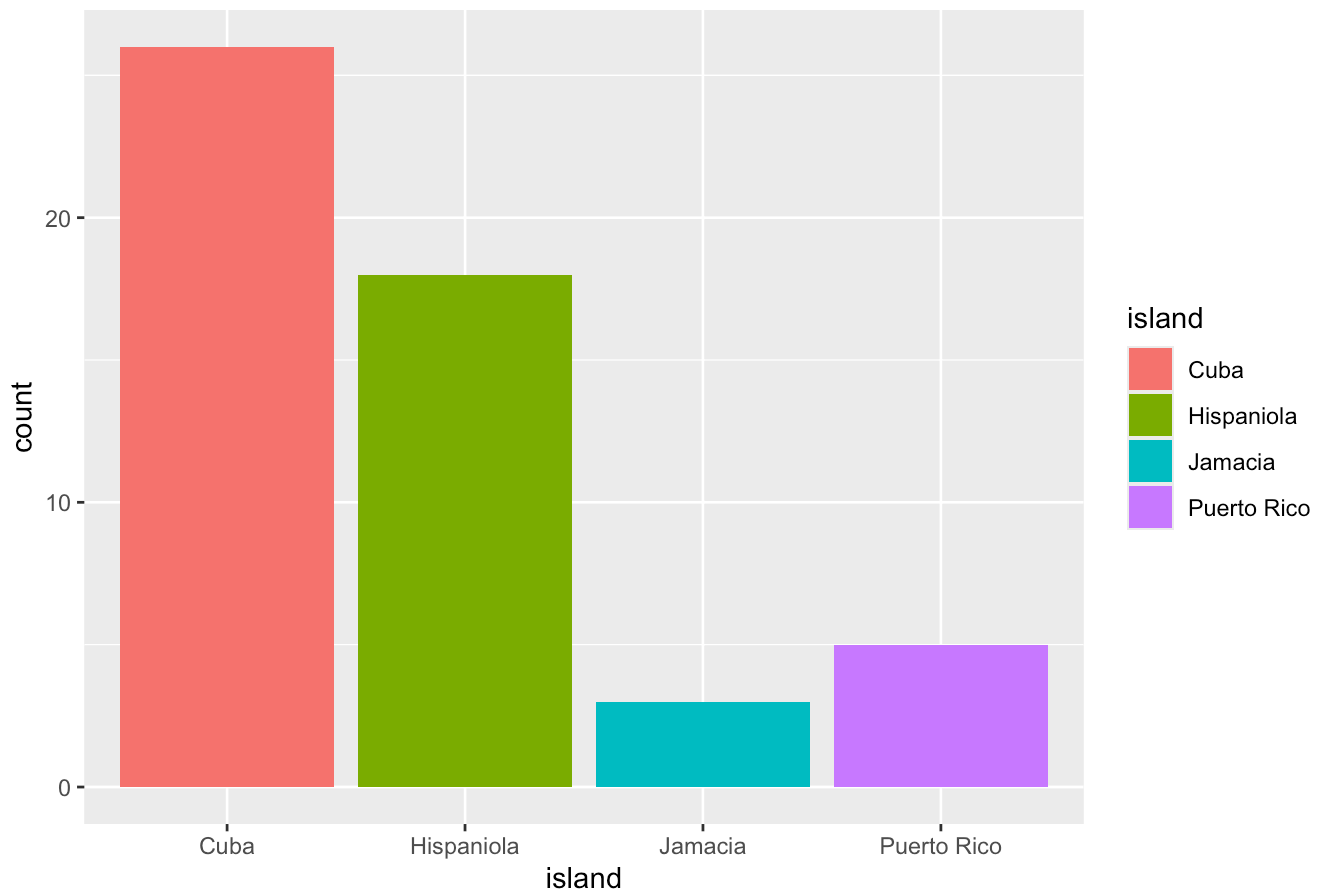
```
## # A tibble: 4 × 1
##   island
##   <fct>
## 1 Cuba
## 2 Hispaniola
## 3 Puerto Rico
## 4 Jamacia
```

Island Cuba, Hispaniola, Puerto Rico, Jamacia are represent in the data.

**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.**

```
anolis %>%
  count(island) %>%
  ggplot(mapping=aes(x=island,y=n,fill=island))+
  geom_col()+
  labs(title="Number of Anole Species Sampled from Each Island",x="island",y="count")
```

Number of Anole Species Sampled from Each Island



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

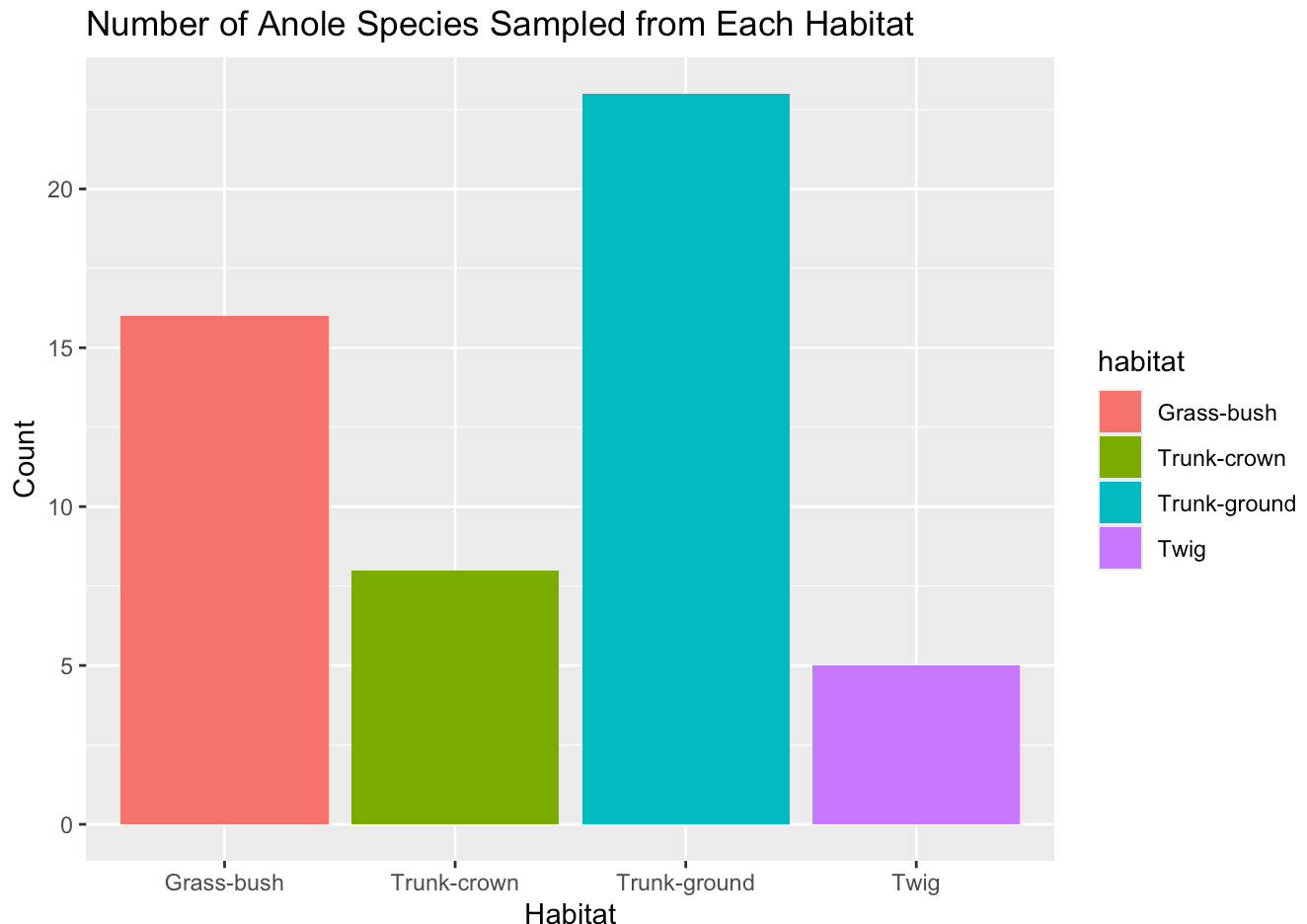
```
anolis %>%
  distinct(habitat)
```

```
## # A tibble: 4 × 1
##   habitat
##   <fct>
## 1 Trunk-ground
## 2 Twig
## 3 Grass-bush
## 4 Trunk-crown
```

Habitat Trunk-ground, Twig, Grass-bush, Trunk-crown are represented in the data.

**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 %>%
  count(habitat) %>%
  ggplot(mapping=aes(x=habitat,y=n,fill=habitat))+
  geom_col()+
  labs(title="Number of Anole Species Sampled from Each Habitat",x="Habitat",y="Count")
```



**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)
```

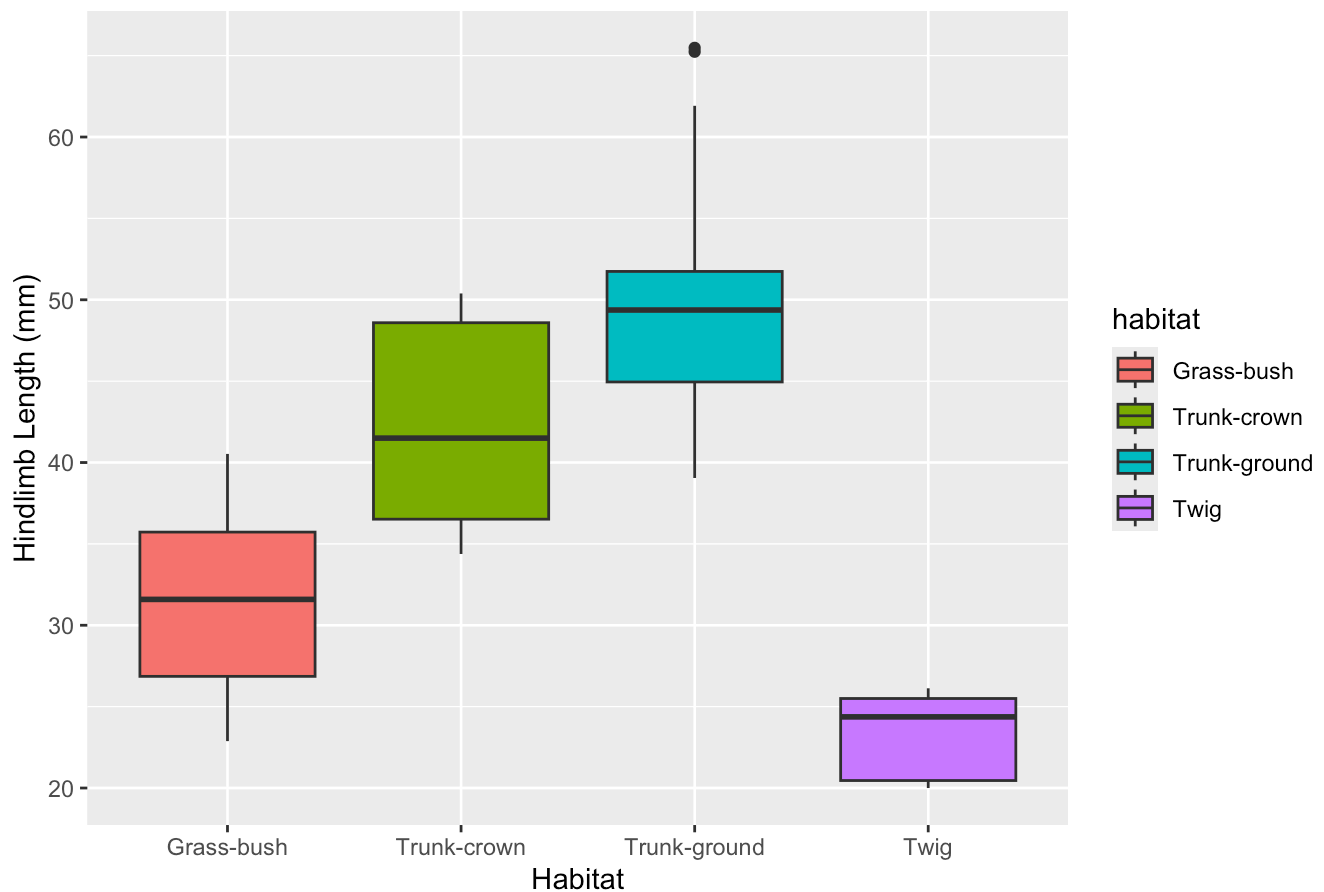
```
## [1] "species"          "habitat"           "hindlimb_length_mm"
## [4] "tail_length_mm"    "body_length_mm"    "toepad_lamellae_count"
## [7] "island"
```

```
is.na(anolis$hindlimb_length_mm)
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE
```

```
anolis %>%
  select(hindlimb_length_mm,habitat) %>%
  ggplot(mapping=aes(x=habitat,y=hindlimb_length_mm,fill=habitat))+
  geom_boxplot()+
  labs(title="How does the Range of Hindlimb Length Compare among Different Habitats",x
    ="Habitat",y="Hindlimb Length (mm)")
```

How does the Range of Hindlimb Length Compare among Different Habitats



Among these habitats, lizards living in the twigs have shorter hindlimb, while lizards living in Trunk-ground have longer hindlimb.

**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.

```
names(anolis)
```

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

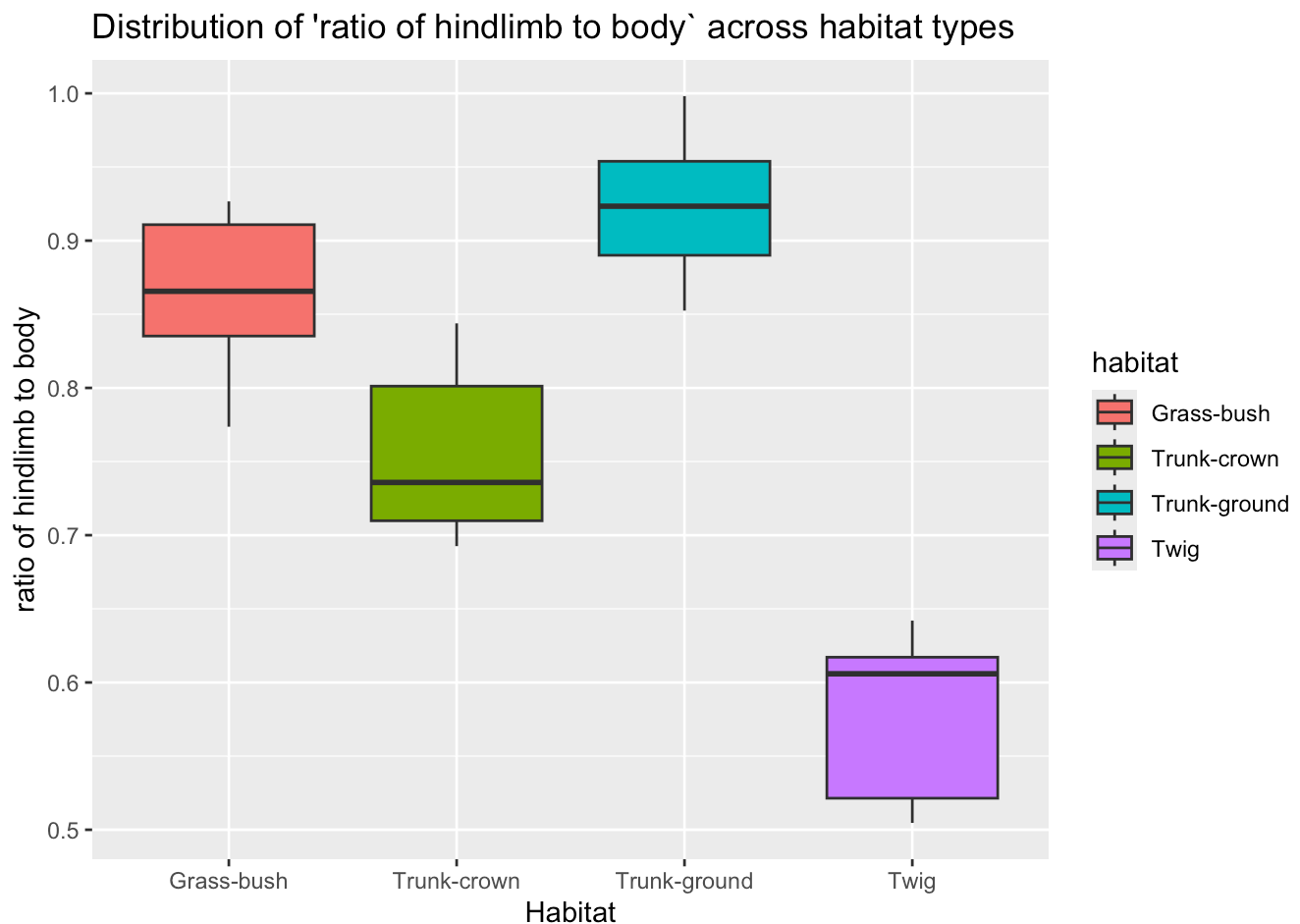
```
anolis <-
  anolis %>%
  mutate(ratio_of_hindlimb_to_body=hindlimb_length_mm/body_length_mm,ratio_of_tail_to_body=tail_length_mm/body_length_mm)
```

**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.

```
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 %>%
  select(ratio_of_hindlimb_to_body,habitat) %>%
  ggplot(mapping=aes(x=habitat,y=ratio_of_hindlimb_to_body,fill=habitat))+
  geom_boxplot()+
  labs(title="Distribution of 'ratio of hindlimb to body` across habitat types",x="Habitat",y="ratio of hindlimb to body")
```



Those two plots are different, especially for the habitat of Grass-bush, if we compare the ratio, which of Grass bush is larger than which of Trunk-ground and Twig, the possible reason can be that the lizards from Grass-bush are smaller, so their hindlimbs are longer compared to their 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.**

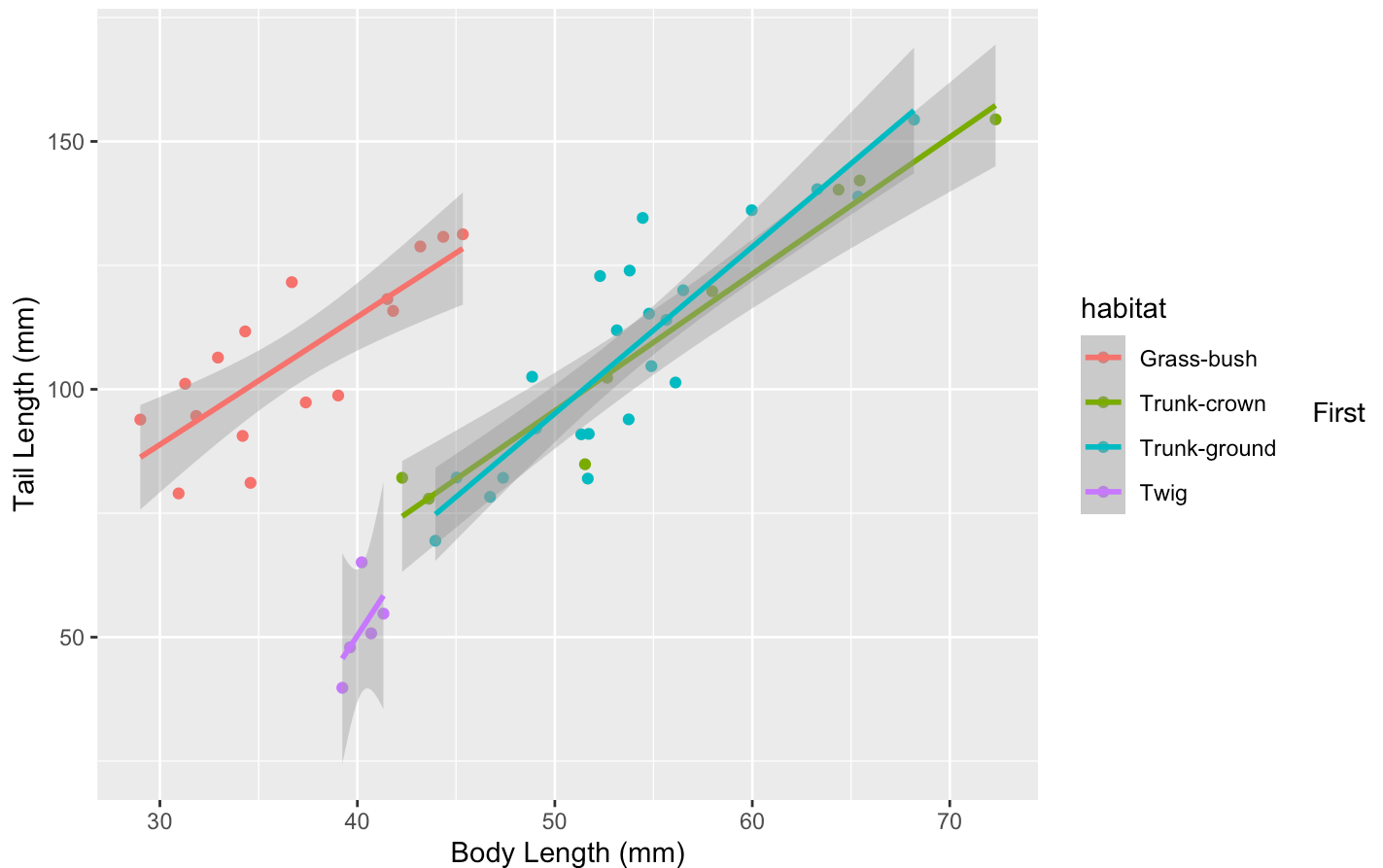
```
names(anolis)
```

```
## [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"
```

```
analis %>%
  select(body_length_mm,tail_length_mm,habitat) %>%
  ggplot(mapping=aes(x=body_length_mm,y=tail_length_mm,colour=habitat))+
  geom_point()+
  geom_smooth(method="lm",na.rm=T)+
  labs(title="The Relationship Between Body Length and Tail Length In Different Habita
t",x="Body Length (mm)",y="Tail Length (mm)")
```

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

### The Relationship Between Body Length and Tail Length In Different Habitat



of all, this plot indicates that with longer body length, the lizards tends to have longer tail, Because for longer body length, those lizards need longer tail to balance themselves. I also noticed that the lizards from Grass-bush have have both shorter body length but their tail lengths are similar to those of lizards from Trunk-crown and Trunk-ground, even those, lizards from those two habitat are much larger than them.

**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 %>%
  distinct(habitat)
```

```
## # A tibble: 4 × 1
##   habitat
##   <fct>
## 1 Trunk-ground
## 2 Twig
## 3 Grass-bush
## 4 Trunk-crown
```

```
anolis %>%
  filter(habitat=="Trunk-ground") %>%
  select(habitat,toepad_lamellae_count) %>%
  summarize(mean_toepad_lamellae_count=mean(toepad_lamellae_count))
```

```
## # A tibble: 1 × 1
##   mean_toepad_lamellae_count
##   <dbl>
## 1                30.0
```

```
anolis %>%
  filter(habitat=="Twig") %>%
  select(habitat,toepad_lamellae_count) %>%
  summarize(mean_toepad_lamellae_count=mean(toepad_lamellae_count))
```

```
## # A tibble: 1 × 1
##   mean_toepad_lamellae_count
##   <dbl>
## 1                27.6
```

```
anolis %>%
  filter(habitat=="Grass-bush") %>%
  select(habitat,toepad_lamellae_count) %>%
  summarize(mean_toepad_lamellae_count=mean(toepad_lamellae_count))
```

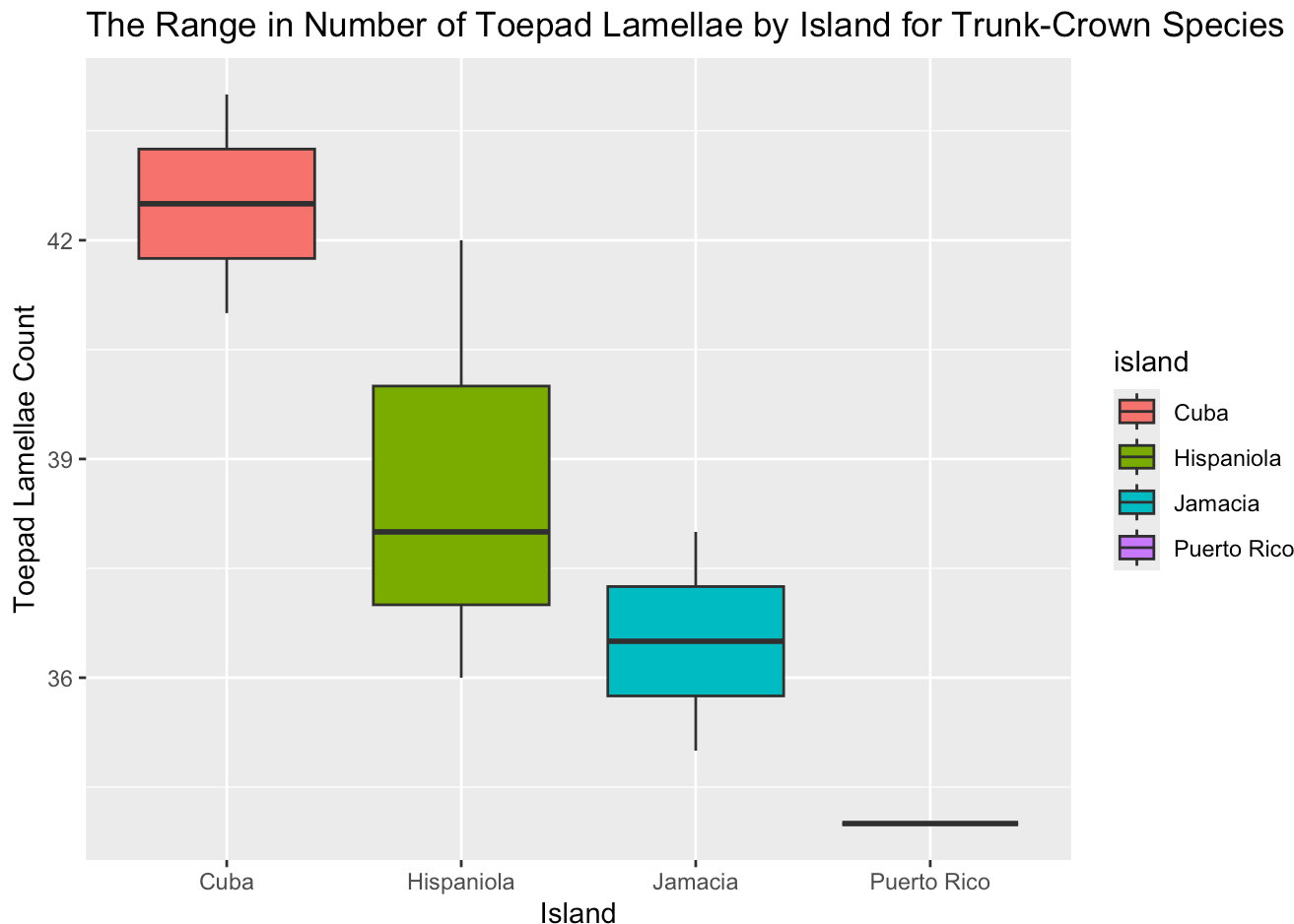
```
## # A tibble: 1 × 1
##   mean_toepad_lamellae_count
##   <dbl>
## 1                28.3
```

```
anolis %>%
  filter(habitat=="Trunk-crown") %>%
  select(habitat,toepad_lamellae_count) %>%
  summarize(mean_toepad_lamellae_count=mean(toepad_lamellae_count))
```

```
## # A tibble: 1 × 1
##   mean_toepad_lamellae_count
##                               <dbl>
## 1                               38.5
```

**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 %>%
  filter(habitat=="Trunk-crown") %>%
  select(habitat,toepad_lamellae_count,island) %>%
  ggplot(mapping=aes(x=island,y=toepad_lamellae_count,fill=island))+
  geom_boxplot()+
  labs(title="The Range in Number of Toepad Lamellae by Island for Trunk-Crown Species",
x="Island",
y="Toepad Lamellae Count")
```



No, it is not same for islands even though these lizards have similar habitat, the toepad lamellae count is much less than Puerto Rico, and significantly more in Cuba island. ## 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.