

Midterm 1 W26

Chengyi Wang

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/chengyi-w/BIS15W2026_ChengyiWang (https://github.com/chengyi-w/BIS15W2026_ChengyiWang)

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. alfaroii", "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.

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

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

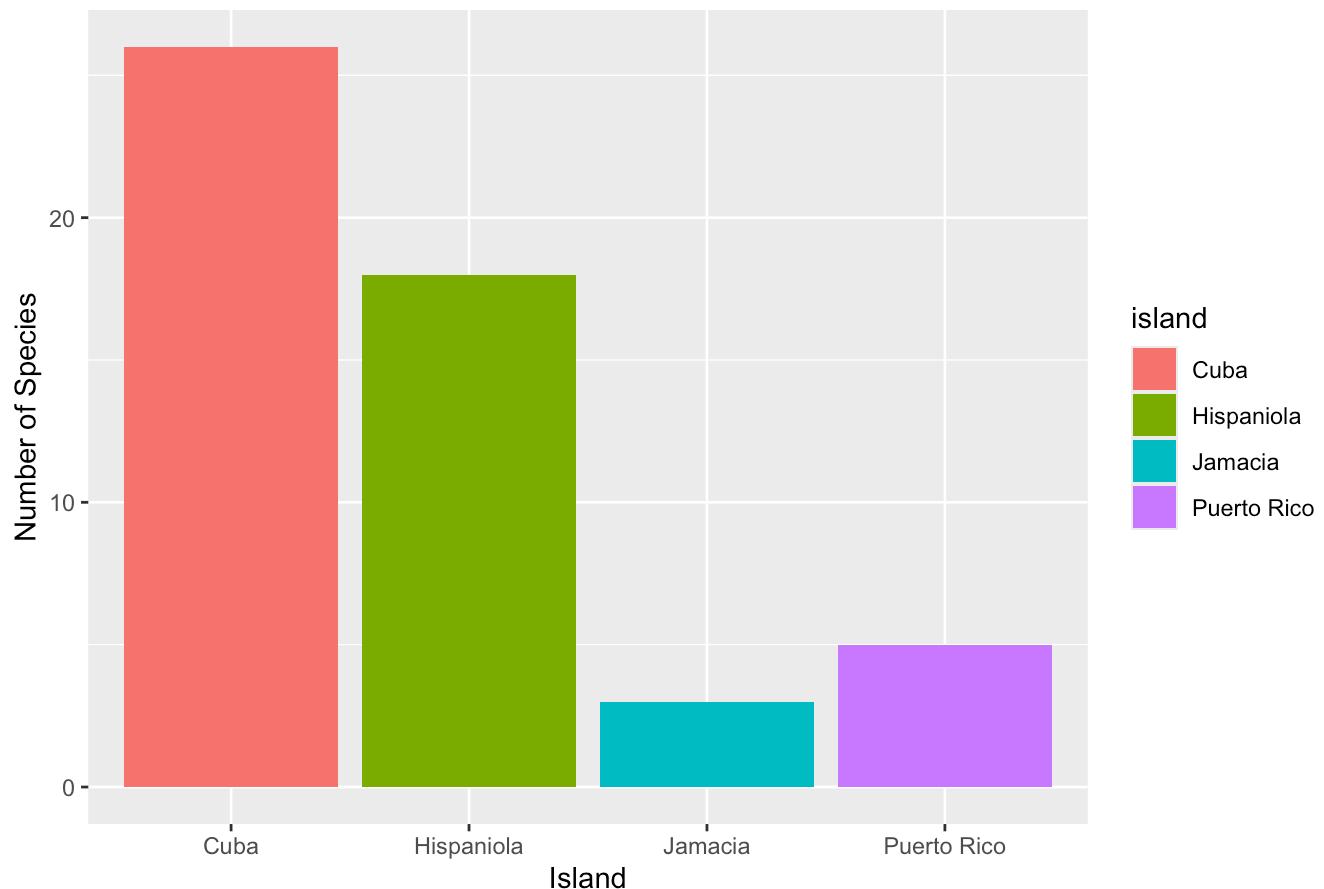
```
anolis %>%
  select(island) %>%
  distinct()
```

```
## # 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.

```
anolis %>%
  ggplot(mapping=aes(x=island,fill=island)) +
  geom_bar()+
  labs(x="Island",y="Number of Species",title="Number of anole species sampled from each island ")
```

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.

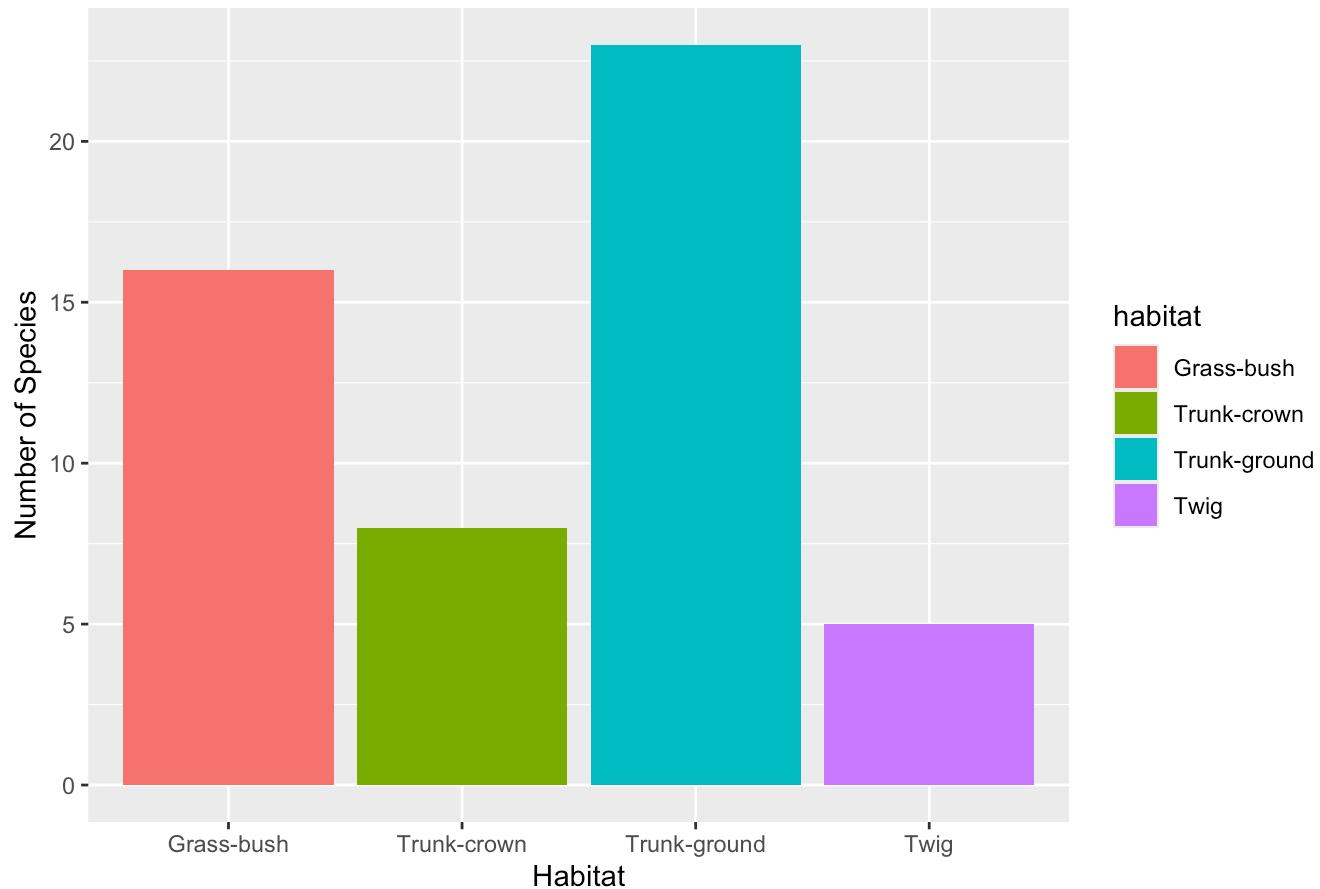
```
anolis %>%
  select(habitat) %>%
  distinct()
```

```
## # 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 %>%
  ggplot(mapping=aes(x=habitat, fill=habitat)) +
  geom_bar()+
  labs(x="Habitat",y="Number of Species",title="Number of anole species sampled from each habitat type")
```

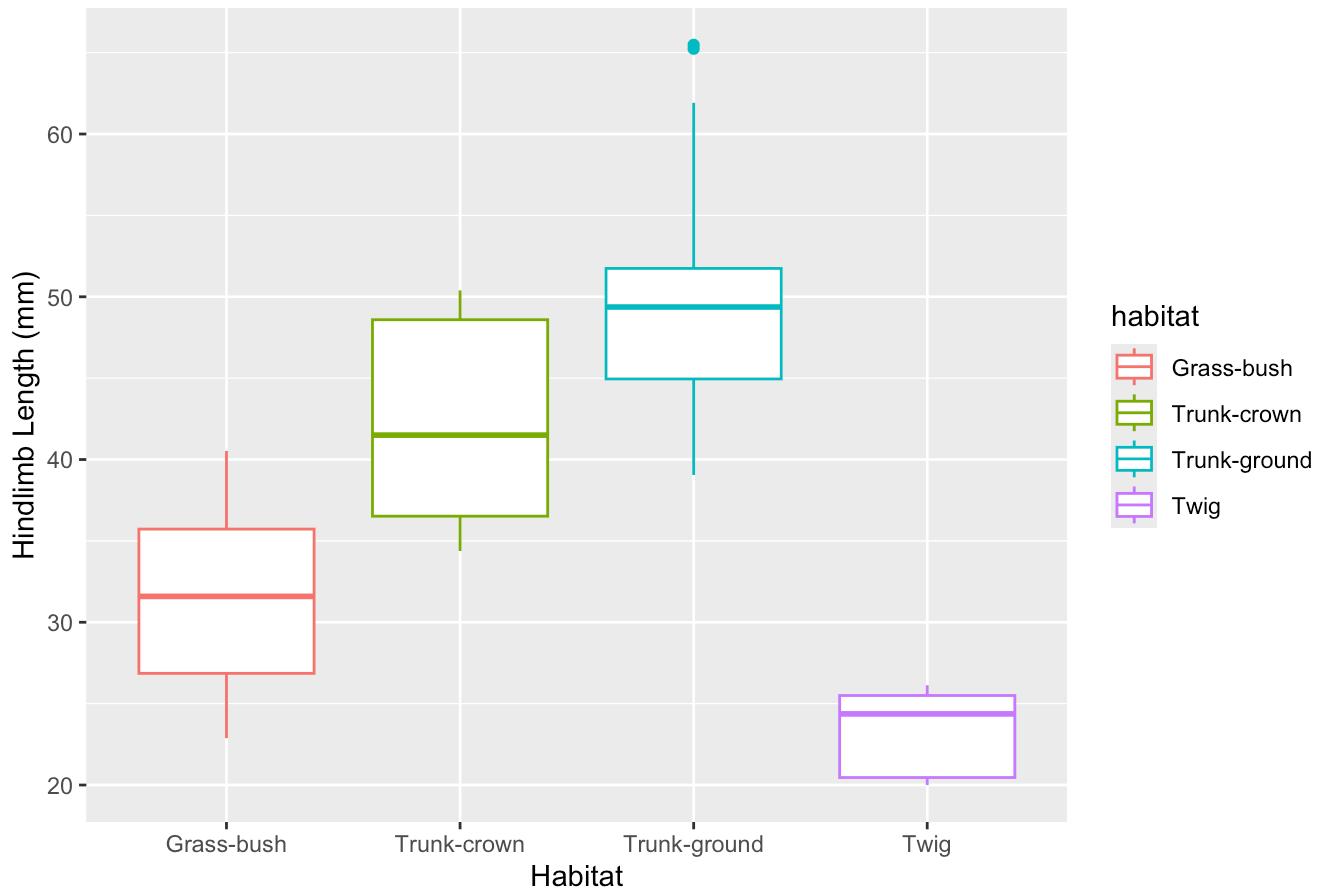
Number of anole species sampled from each habitat type



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.

```
anolis %>%
  ggplot(mapping=aes(x=habitat,y=hindlimb_length_mm,colour=habitat)) +
  geom_boxplot()+
  labs(x="Habitat",y="Hindlimb Length (mm)",title="Hindlimb Length among different habitat types")
```

Hindlimb Length among different habitat type



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) %>%
  mutate(ratio_of_tail_to_body = tail_length_mm/body_length_mm)
head(anolis)
```

```
## # A tibble: 6 × 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
## # i 4 more variables: toepad_lamellae_count <dbl>, island <fct>,
## #   ratio_of_hindlimb_to_body <dbl>, ratio_of_tail_to_body <dbl>
```

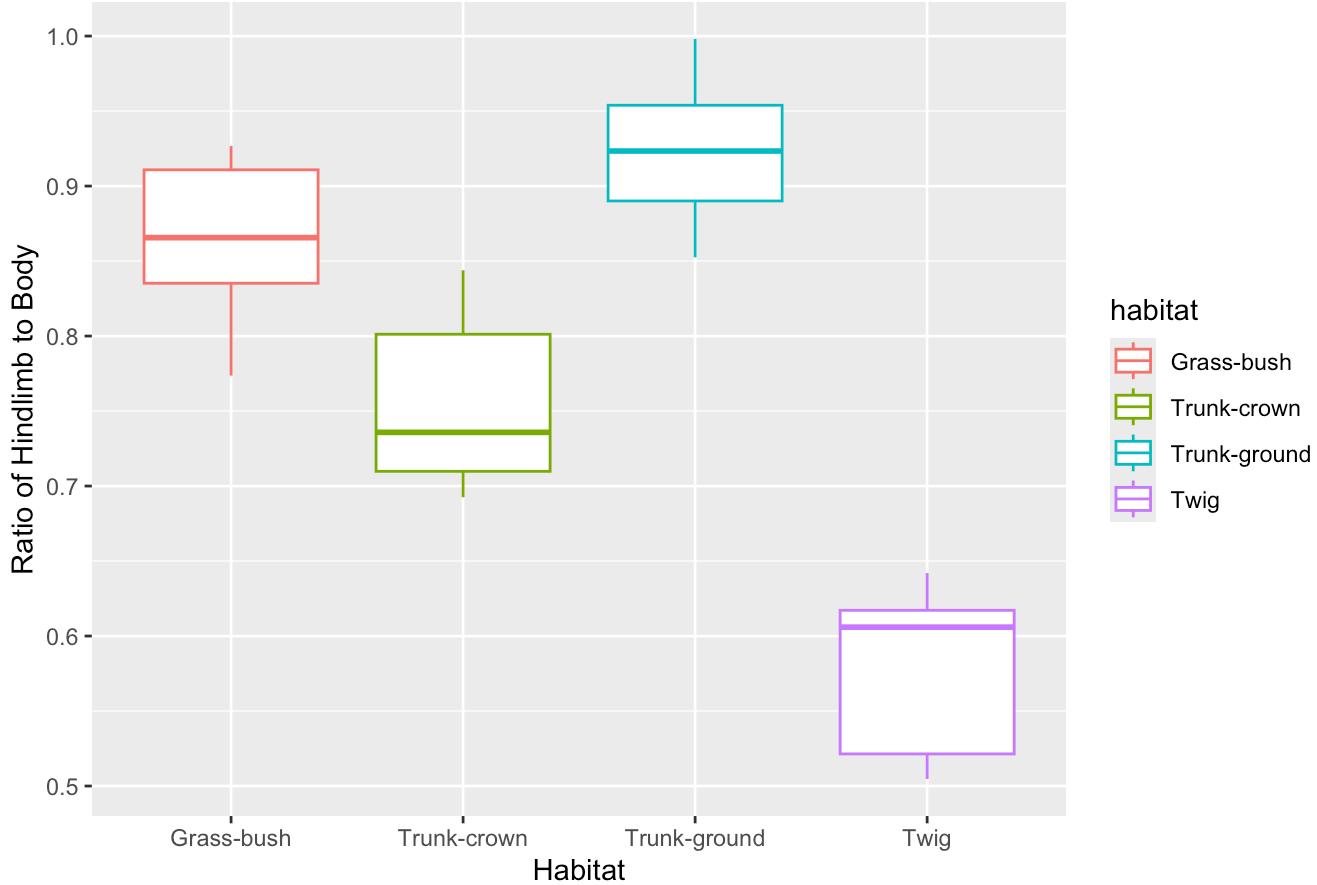
Question 12. (4 points) Create a new plot that examines the distribution of `ratio_of_hindlimb_to_body` across habitat types. How does this plot differ from the one you made in Problem 10? Be sure to label your axes and add a title.

```

anolis %>%
  ggplot(mapping=aes(x=habitat,y=ratio_of_hindlimb_to_body,colour=habitat)) +
  geom_boxplot()+
  labs(x="Habitat",y="Ratio of Hindlimb to Body",title="Ratio of Hindlimb to Body among different habitat type")

```

Ratio of Hindlimb to Body among different habitat type



It is a bit different from the previous one. The differences between species from each habitats are relatively smaller now. Also, in the previous plot the species from grass-bush habitats have shorter hindlimb length compared to species from Trunk-crown, but in this new plot, species from grass-bush habitats have larger hindlimb to body ratio compared to species from Trunk-crown.

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 %>%
  ggplot(mapping=aes(x=body_length_mm,y=tail_length_mm,colour = habitat))+ 
  geom_point()+
  geom_smooth(method = lm, na.rm = T)+
  labs(x="Body Length (mm)",y="Tail Length (mm)",title = "Body Length vs Tail Length of each habitats")

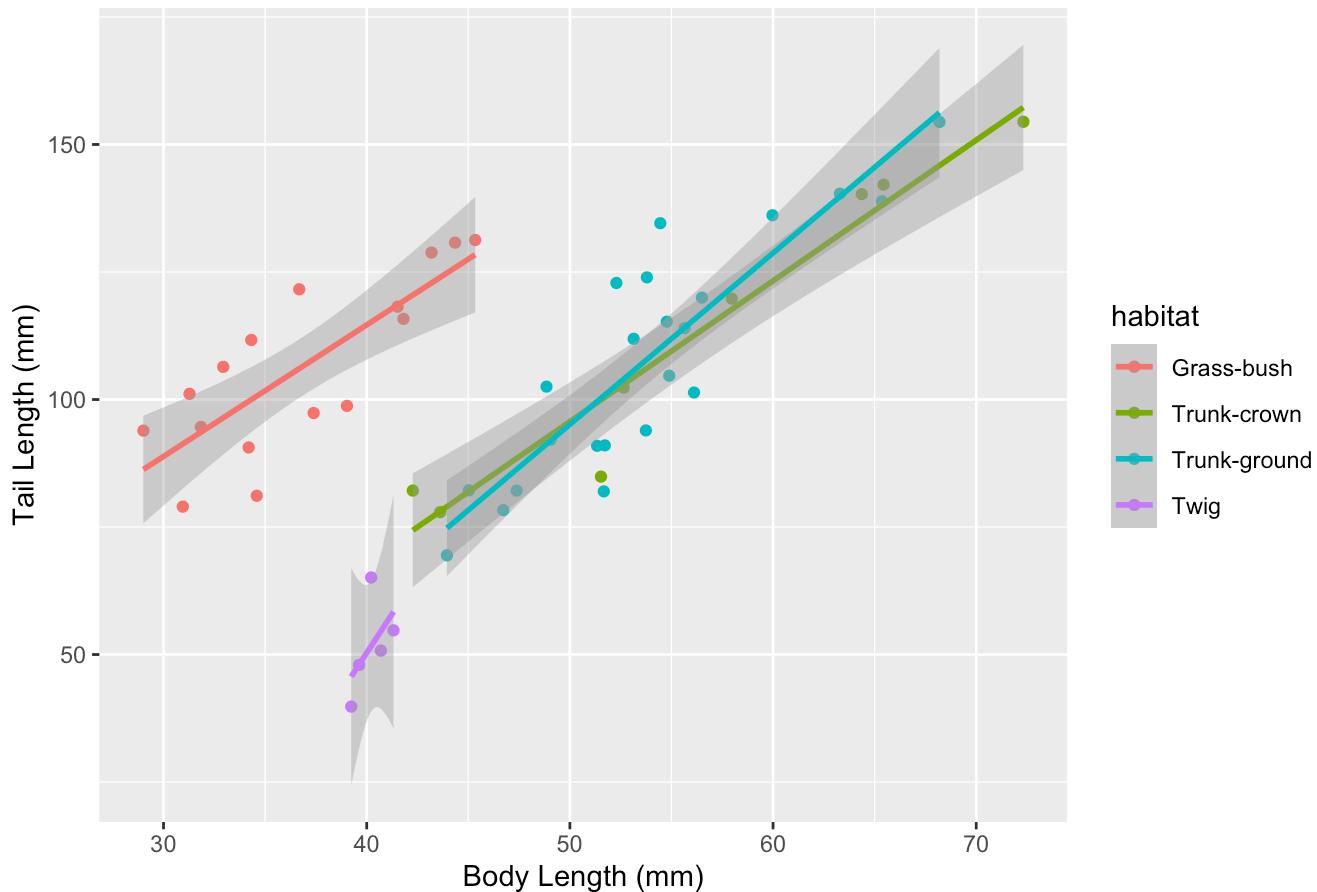
```

```

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

```

Body Length vs Tail Length of each habitats



For all 4 habitats, there is a positive correlations between body length and tail length. Species from Grass-bush habitat have tails of similar length to Trunk-crown and Trunk-ground habitat, while they have smaller body length, and have longer tails compared to species from Twig habitat which have around the same body length.

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?

```
anolis %>%
  select(habitat,toepad_lamellae_count) %>%
  group_by(habitat) %>%
  summarise("mean number of toepad lamellae"=mean(toepad_lamellae_count))
```

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

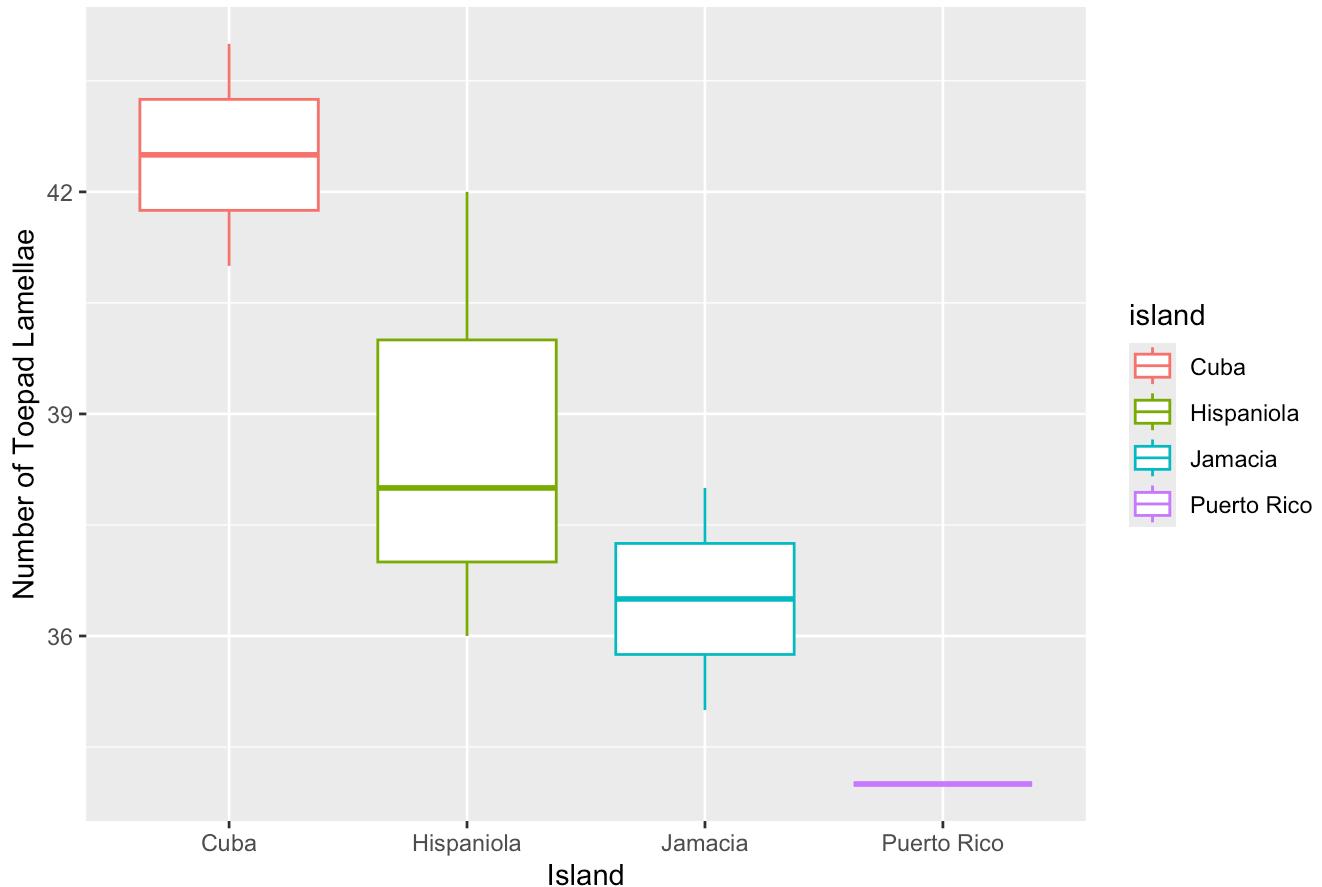
Problem 15. (5 points) The number of toepad lamellae is significantly different for trunk-crown species. But, is this consistent across all islands? Make a plot that shows the range in number of toepad lamellae by island for trunk-crown species only. Be sure to label your axes and add a title.

```

anolis %>%
  filter(habitat=="Trunk-crown") %>%
  ggplot(mapping=aes(x=island,y=toepad_lamellae_count,colour = island))+
  geom_boxplot()+
  labs(x="Island",y="Number of Toepad Lamellae", title="Toepad Lamellae by Island for Trunk-crown Species")

```

Toepad Lamellae by Island for Trunk-crown Species



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.