Final Exam (Modules 1-4)

Keaton Wilson

5/6/2020

Exam 4 (Final)

This exam covers material from Module 4, and also earlier in the semester. It is worth the same number of points (10) as a normal exam.

Question 1

We've spent a lot of time this semester learning the fundamentals of a programming language (R) - at the beginning of this course, we talked a bit about some of the assumptions surrounding data science, programming and statistics. At this point, you have a foundation to continue to build your proficiency in R (or another language, honestly). Thinking back to the beginning of the semester - what are two things that surprised you about learning to program and work with data over the course of the semester, and what is one thing that wasn't surprising? (4-7 sentences).

Question 2

One important part of the Data Science workflow is frequently looking at your data to make sure it's taken the form you except, and to do some quality control to ensure that columns are of the right type, the data is shaped correctly, and a myriad of other important features. Which answer below best outlines the main tools we covered for this piece of the workflow?

a. View()

b. glimpse()

c. head()

d. calling the dataframe from the console (if it's a tibble)

e. all of the above

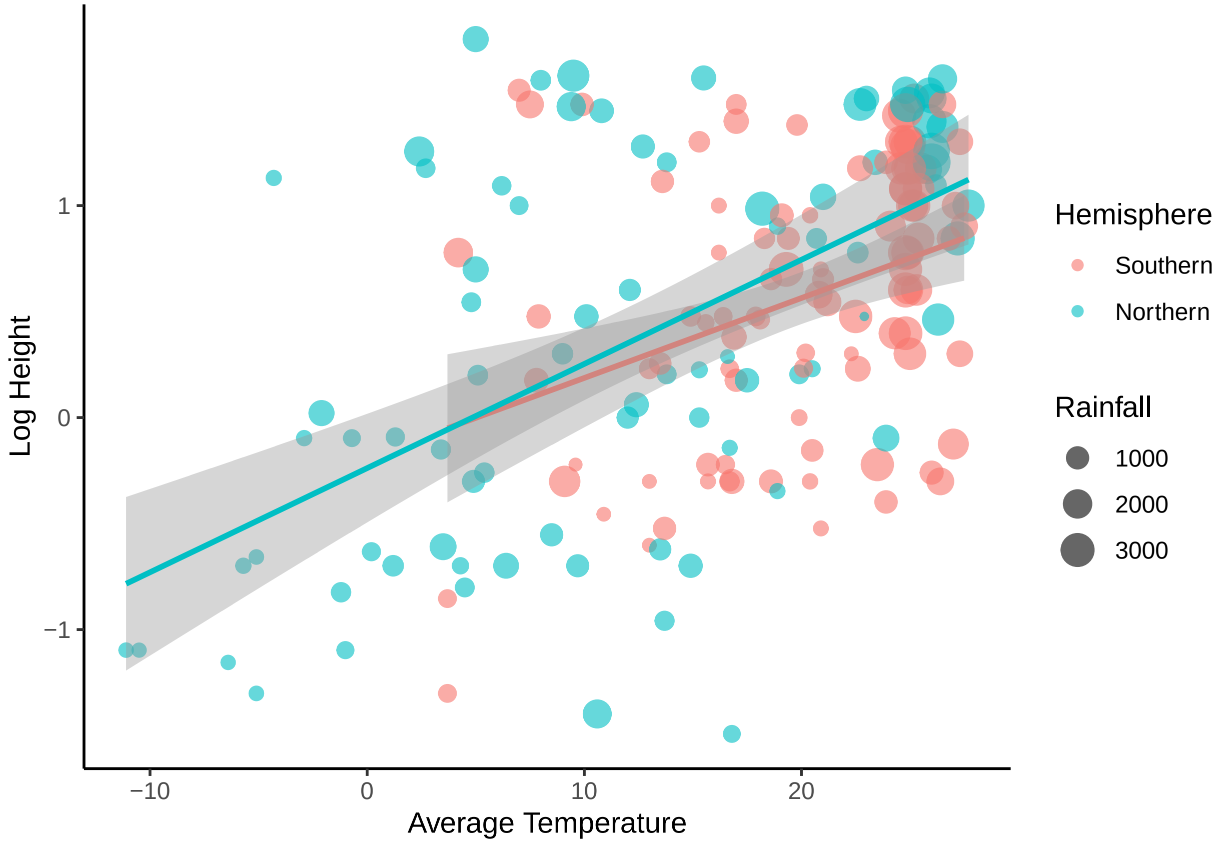
f. b and c

Question 3

Examine the plot below. The lines represent the trendlines from a multiple regression the examines the height (log-transformed) of a bunch of plant species across the world as a function of temperature and what hemisphere (southern or northern) they're in.

E.g. `height_mod = lm(log_height ~ temperature*hemisphere)`

Interpret the figure in every-day language that a colleague with no background in statistics or data science could understand, and describe an additional variable that isn't present in the model above that might be useful.



Question 4

True or false: the data structure we worked mostly frequently with this semester was a list.

Question 5

When starting to write a new script, what do the best practices we learned in class dictate that we do first? Why do we do this? (3-5 sentences).

Question 6

We worked through a few projects throughout the semester that involved resampling our data many, many, many times (permutation tests, bootstrapping, and simulations). What is the main reason that these methods exist and we can use them now, when compared to 50 years ago? (2-4 sentences).

Question 7

The goal of the code below is to create a summarized data frame that calculates the mean and standard deviation of the lifespan of a number of male and female penguins across multiple sites, gets rid of missing lifespan data, and then plots these values. Describe three problems with the code below.

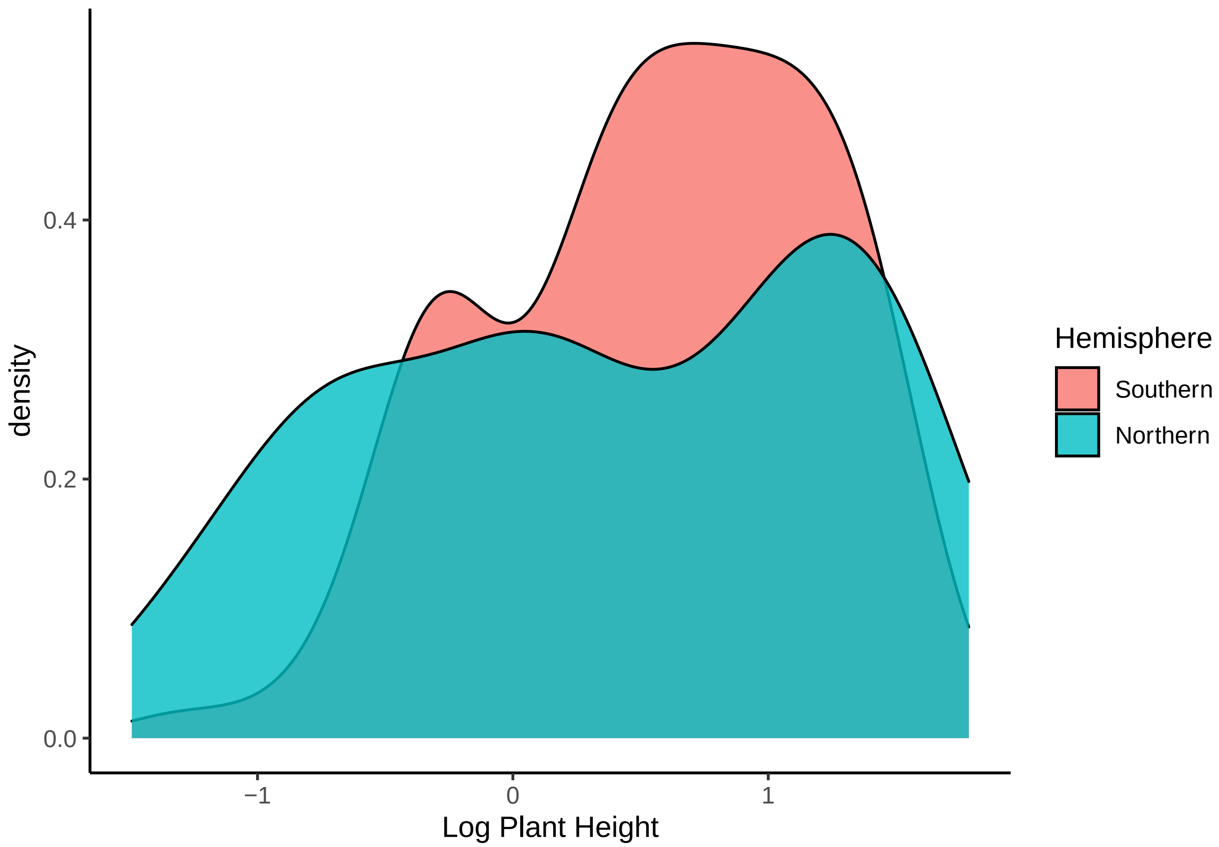
```
penguins %>%
  select(lifespan, sex) %>%
  filter(!is.na(lifespan)) %>%
  group_by(site, sex) %>%
  summarize(mean_lifespan = mean(lifespan),
            sd_lifespan = sd(lifespan))

ggplot(penguins_summary, aes(x = site, y = mean_lifespan, color = sex)) +
  geom_point() +
  geom_errorbar(aes(ymin = mean_lifespan*-2*sd_life_span,
                    ymax = mean_lifespan*2*sd_life_span)) +
  theme_classic()
```

Question 8

Write some psuedo-code (it doesn't have to be syntaxically perfect, but it should convey your thought process and logic) that performs the following task. A for-loop that creates a vector where each item is the log-transformed version of the item in the following vector: c(2, 5, 9, 87, 212, 44). 1 bonus point if you can print the log-transformed value at each step in the for loop.

Question 9

Examine the following figure that shows the log-transformed distribution of plant heights for both the southern and northern hemisphere. What kind of statistical tool would you use to determine whether or not there was a difference in the average height of plants between the two hemispheres? What kind of outcomes would you expect in the output of this test given the plot below?



Question 10

Describe succinctly what the pipe (%>%) operator does and the contexts in which we've used it over the course of the class? (3-7 sentences).