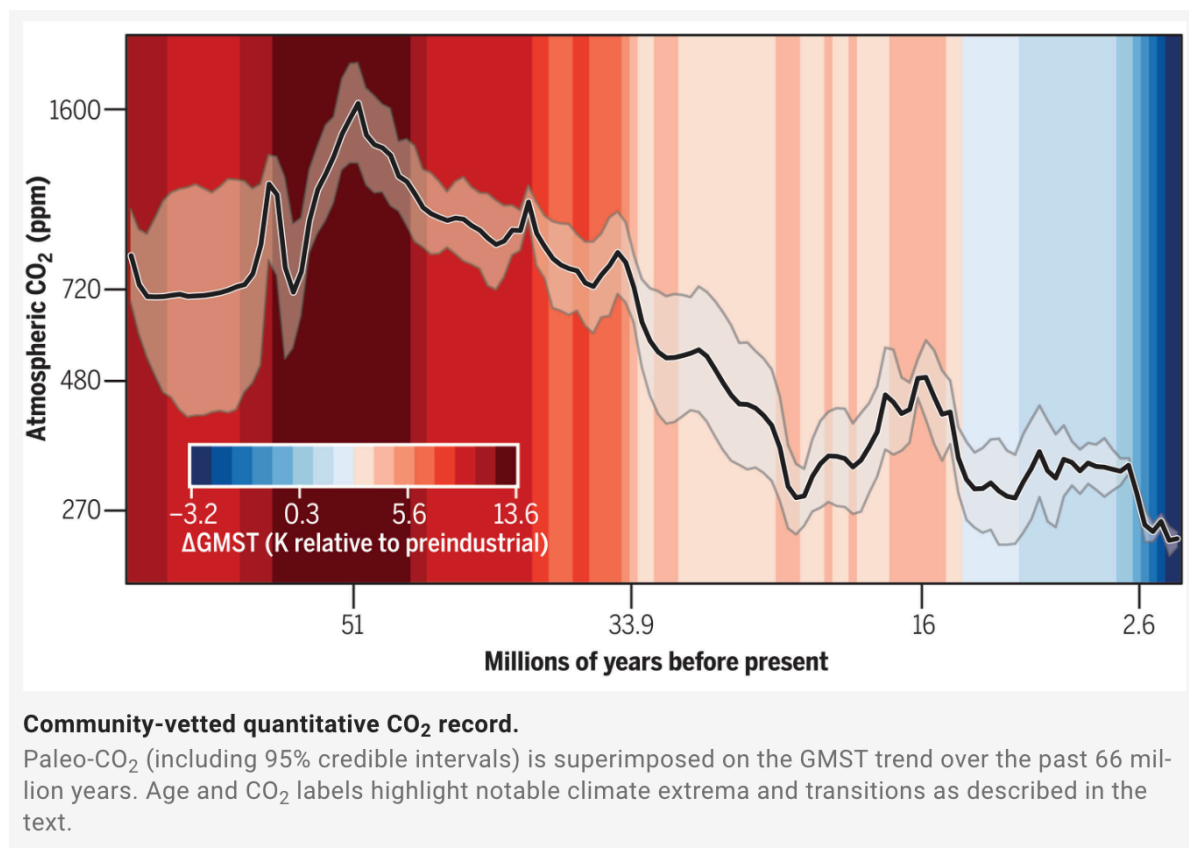


BIOL 220 Problem Set 02: Graphics

Answer Key

Fix this graph!

Consider the graphic below from a recent [article in *Science* by the Cenozoic CO₂ Proxy Integration Project Consortium](#). The data visualization attempts to show complex data about the change in atmospheric CO₂ and average global temperature over the past approximately 65 million years.



💡 Use the figure above to answer questions 1 & 2.

Hints: on the legend for the color scale, where is 0? Look closely at the tick marks of the axes. What *messages* do the authors want to communicate about these data; what are the patterns they're trying to show? Do they achieve their goals?

1. Identify two instances in which this graphic violates the data visualization rules from Whitlock and Schluter's textbook. There are multiple possible answers. [2 points]

i Possible answers

1. The color gradient implies that a 0 difference is at the most pale colors, but 0 is actually located at a darker blue color, making interpreting the figure difficult
2. Overplotting: there's too much information in this one plot
3. The use of acronyms makes it hard to understand what information is being communicated
4. The axis labels appear arbitrary
5. We can't see the actual data
6. Not clear what the message is
7. The recent dramatic increase in temperature toward the present cannot be seen because of the scale
8. Color choice is not ideal, but not the most terrible thing about this visualization

2. For each of those two violations, suggest a remedy to improve it. [2 points]


i Possible answers

1. Center the color gradient at 0
2. Make different plots for the different visualizations
3. Spell out the acronyms
4. Label the axis consistently, then add additional marks if you want the readers to take note of specific times
5. Show the data
6. The message is that CO₂ and temperature have been correlated for a long time, show a scatterplot of that correlation
7. The message is also that there has been enormous recent increase in CO₂ and temperature (with temperature lagging behind but inevitably catching up), show this with an additional plot focused specifically on recent times
8. Breaking the visualization into multiple plots could eliminate the need for a color gradient that may not be well suited to all viewers

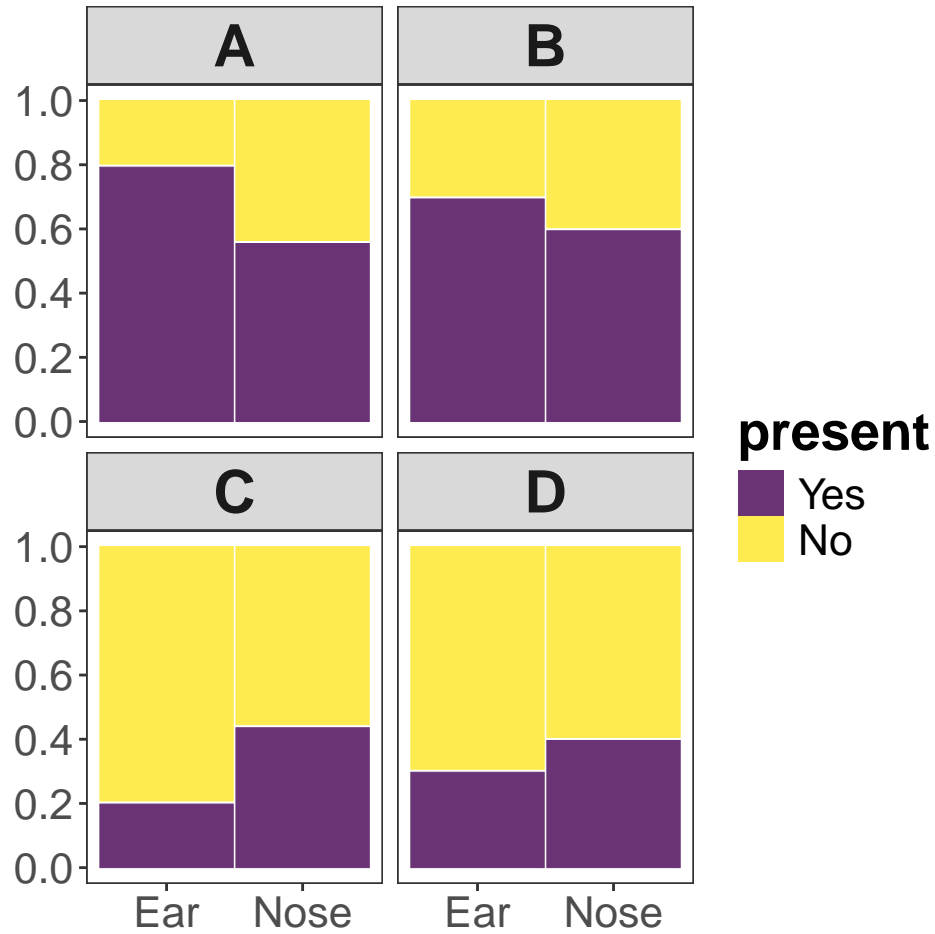
Interpreting figures

Consider a situation in which bacterial swabs were taken from the ears and noses of 50 study subjects, and the number of swabs that showed the presence of *Staphylococcus* were measured. The table below shows the results of the measurements.

	Yes	No
Ear	40	10
Nose	28	22

 Use the information in the above table to answer question 3.

3. Which of the plots shown below (**A**, **B**, **C**, or **D**) correctly depicts the *Staphylococcus* swab data? [1 point]



i Answer

A

Making graphics

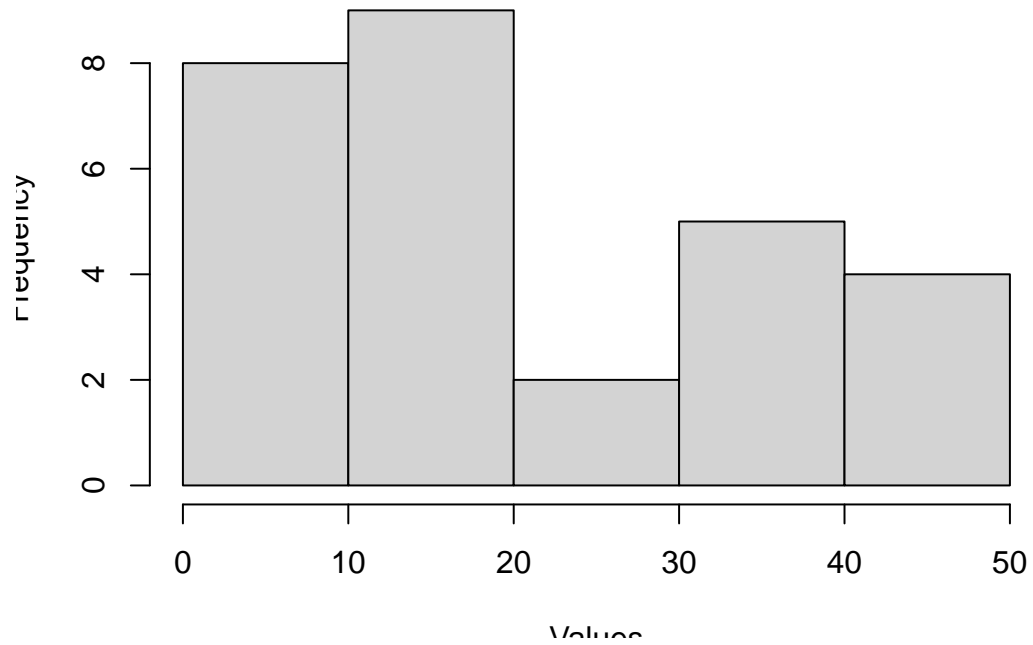
💡 Draw a graph by hand!

- Sketch out a histogram (with 5 bins of equal size as appropriate) showing the distribution of the following values:

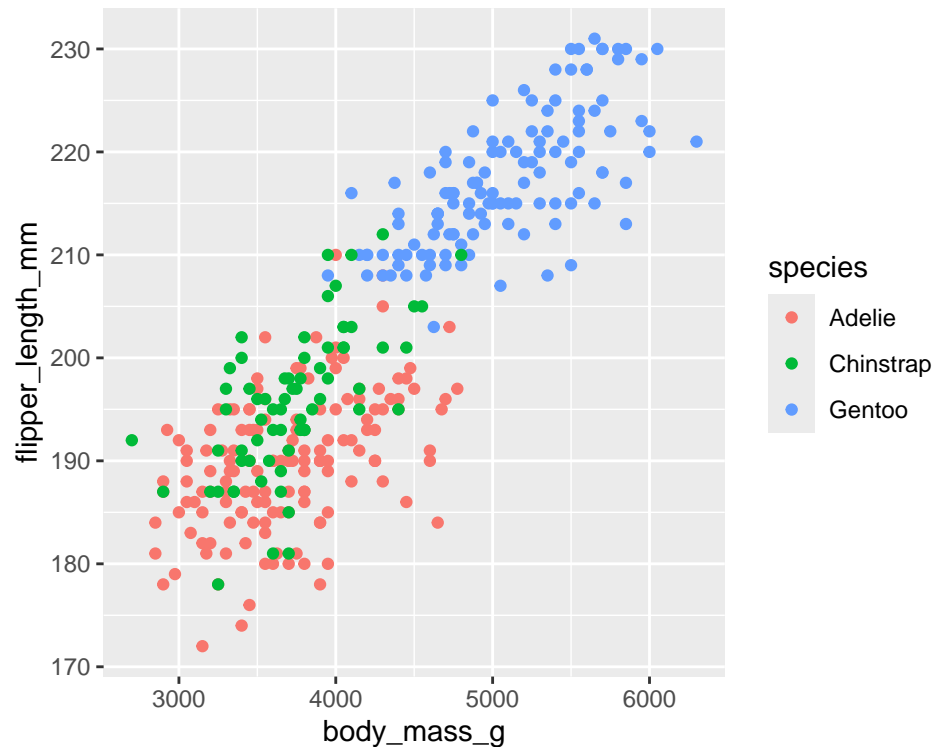
2, 3, 3, 4, 5, 7, 8, 8, 11, 11, 12, 13, 15, 15, 17, 18, 18, 23, 24, 32, 33, 34, 35, 38, 41, 42, 43, 48

You should draw a good, honest, clear histogram BY HAND either on paper or digitally, then submit your graphic by uploading it via the Google Form [2 points].

i Answer



Now you'll fix some *ggplot2* code to improve the quality of the below figure:



This is a *similar* figure to one we have thought about in lecture. In lecture we looked at bird body mass and bird wing length for non-native species in Hawai i. This figure shows body mass and flipper (aka wing) length for three species of penguins.

💡 Improve this figure!

Here is the code needed to produce the above visualization of the penguin data:

```
library(ggplot2)

# note: the `penguin` dataset we need is made available to us via the
# `palmerpenguins` package
library(palmerpenguins)

ggplot(penguins, aes(x = body_mass_g, y = flipper_length_mm,
                     color = species)) +
  geom_point()
```

5. [3 points] Choose two aspects of the iris data plot that you think should be improved. Then starting with the above code, add to it with additional commands to realize the improvements you'd like to make.

To turn in your answer for this question simply paste your final code into the google form. Include comments in your code explaining (briefly) what you chose to improve and why, e.g. (these are *not* good changes to make):

```
# 1. Change the colors to rainbow because go bows!  
# 2. Change the points to diamonds because why not!  
  
ggplot(iris, .....  
  .....  
  .....  
  .....
```

i Possible answers

Warning: Removed 2 rows containing missing values or values outside the scale range (`geom_point()`).

