

---

## 0.1 Question 1a

Let's start by visualizing the distribution of the islands from which the penguins in our dataset come. Your friend Sunya says that using a **histogram** is the best way to do this. However, Josh disagrees and says that using a **horizontal bar chart** is a better move. Who is right? Explain your answer.

*Hint:* What kind of variable is contained within the "**island**" column?

*Type your answer here, replacing this text.*



```
In [ ]: ...  
  
    # Do not change any code below this comment  
    plt.xlabel("Number of Penguins")  
    plt.title("Number of Penguins on Each Island");
```



---

## 0.2 Question 2b

Sandra takes a quick glance at the histogram from **Question 2a** and claims the following: > *“I can tell you how many penguins have body masses between 5000 and 5200 grams, **only** using the histogram above.”*

Leanne disagrees with Sandra, stating: > *“That’s not true. You’d need more than just the histogram above to know that information.”*

Who is correct in this situation: Sandra or Leanne? Explain why the *other person is incorrect*, providing justification based on what you’ve learned about histograms.

*Type your answer here, replacing this text.*



**Task:** In the cell below, write code that produces a histogram that visualizes the distribution of our penguins' flipper lengths. We've provided the bins for you to use as the `my_bins` variable. More specifically, you'll want to re-create the following histogram exactly:

```
In [ ]: my_bins = np.arange(170,240, 5)
        ...

        # Do not change anything below this line
        plt.title("Distribution of Flipper Lengths (mm)");
```





**Task:** In the cell below, compare the two histograms that result from setting the bin width to **100** and **750**. What are the pros and cons of each size? (Remember that these histograms are displaying the same data, just with different bin sizes.)

*Type your answer here, replacing this text.*



---

### 0.3 Question 2e

In the previous question, you plotted the distribution of penguins' body mass measurements for the **entire dataset**. While doing so is important to visualize our data as a whole, we can uncover underlying patterns within our data by grouping. This process is known as **disaggregation**: separating aggregated data into its component parts; we saw this process in Lab 2 with the BRFSS data set.

**Task:** Using the `penguins` table and the `hist` table method and its various optional arguments, re-create the following set of histograms:

*Note:* Use the `my_bins` variable we've defined as the optional `bins` argument of the `hist` method.

```
In [ ]: my_bins = np.arange(2500, 6500, 500)
        ...
```



### 0.3.1 Question 3a (Discussion)

Looking at `cal_data` first glance, what do you observe? Try exploring the dataset by what you have learned so far in class! Reflect in 2-3 sentences on what you have noticed about the data. Feel free to add/ delete more cells beyond our given cell.

*Type your answer here, replacing this text.*



### 0.3.2 Question 3d

Using the adjusted `admissions`, create a bar graph comparing the acceptance rate between female and male applicants.

```
In [ ]: ...
```





### 0.3.3 Question 3f

Using `admission_major` and `num_applicants`, create two separate overlaid bar charts. The first one will plot the number of applicants of males and females across the six majors, and the second will plot their acceptance rates in the majors.

```
In [ ]: ...  
      ...
```



### 0.3.4 Question 3h (Discussion)

Consider the takeaways that you had after looking at the different visualizations you created in **Questions 3d and 3f**. Based on what we've discussed regarding the various sociological paradigms, how do you think these conclusions best align with the paradigms?

*Hint:* The conclusions might not line up perfectly with any one of the paradigms! Just feel free to reflect on which ones they might line up best or worst with.

*Type your answer here, replacing this text.*



### 0.3.5 Question 3i (Discussion)

A media company wants to make a report on UC Berkeley admissions. What would happen if they ONLY report with the visualization you created in **Question 3d**? How would this report potentially impact the society we see today?

*Type your answer here, replacing this text.*

