Session 10: EDA — Variation — Pen-and-Paper Pair Exercise

PSY 410 | Data Science for Psychology

Name: _____ **Date:** _____

No laptop today? No problem. This handout lets you practice the same skills on paper. Work with a partner who has a laptop and compare your work at the end.

The data: `penguins`

This dataset (from `palmerpenguins`) has measurements of 344 penguins from three species. Here's a summary of the variable you'll focus on:

`flipper_length_mm` (flipper length in millimeters):

Statistic	Value
Min	172
1st Qu.	190
Median	197
Mean	200.9
3rd Qu.	213
Max	231
NAs	2

Species breakdown: Adelie (152), Chinstrap (68), Gentoo (124)

Here are 10 representative rows:

species	island	flipper_length_mm	body_mass_g	sex
Adelie	Torgersen	181	3750	male
Adelie	Torgersen	186	3800	female
Adelie	Torgersen	195	3250	female
Adelie	Torgersen	NA	NA	NA
Adelie	Torgersen	193	3450	female
Chinstrap	Dream	192	3500	female
Chinstrap	Dream	196	3900	male
Gentoo	Biscoe	217	4500	female
Gentoo	Biscoe	230	5200	male

species	island	flipper_length_mm	body_mass_g	sex
Gentoo	Biscoe	221	5150	male

The task (same as the slide exercise)

1. Plot the distribution of **flipper_length_mm** — what shape is it?
2. Check for **outliers** in flipper length. Are any values suspicious?

Your pen-and-paper version

Step 1: Sketch a histogram. Using the summary statistics above, draw a rough histogram on the grid below. Think about where most values fall (between Q1 and Q3) and the overall range.

```
Count
  |
  |
  |
  |
  |
  |_________________________________________
  170  180  190  200  210  220  230
             flipper_length_mm
```

Step 2: What shape do you expect? Circle one:

- Symmetric / bell-shaped
- Right-skewed (tail to the right)
- Left-skewed (tail to the left)
- Bimodal (two humps)

Why? (Hint: look at the species breakdown and the sample rows. Do all species have similar flipper lengths?)

Your answer: _____

Step 3: Sketch a boxplot. Using the summary statistics, draw a boxplot:

- Draw the box from Q1 (***) to Q3 (***)

- Draw the median line at _____
- Draw whiskers extending to the min and max (or 1.5 x IQR, whichever is closer)

```
|__|___|_____|______|___|__|
170        190      210       230
```

Step 4: Outlier check. Calculate the IQR and outlier fences:

- IQR = Q3 - Q1 = _____ - _____ = _____
- Lower fence = Q1 - 1.5 x IQR = _____ - _____ = _____
- Upper fence = Q3 + 1.5 x IQR = _____ + _____ = _____

Are any values in the summary statistics beyond these fences? _____

Step 5: Write the code. What code would produce these plots?

```
# Histogram
penguins |> ggplot(aes(x = _________________)) +
  geom_____________(fill = "steelblue", color = "white") +
  theme_minimal()

# Boxplot
penguins |> ggplot(aes(x = _________________)) +
  geom_____________() +
  theme_minimal()
```

Check your work

Compare your sketches, outlier calculations, and code with your partner's screen. Do your answers match?