The Quantitative Center: Location and Average

In Chapter 3, we visualized the "shape" of our data using histograms. However, science requires more than just a picture; it requires a number that describes the "typical" member of a population. We call this the measure of **Central Tendency**.

1. Manual Centers: Feeling the Math

Below is a curated log of **15 platypus weights (kg)** recorded during a single week on the Snowy River.

`1.2, 0.8, 1.5, 2.1, 1.3, 1.1, 1.6, 0.9, 1.4, 1.2, 1.8, 1.0, 2.4, 1.1, 1.3`

Task A: Sort the Data
Before finding the center, you must arrange the values from lightest to heaviest.

Task B: Identify the Median
The Median is the "Middle Seat." Since $n = 15$, the median is the value in the **8th position**.

Task C: Calculate the Mean
The Mean ($\bar{x}$) is the arithmetic average. Sum all 15 values ($\sum x$) and divide by $n = 15$.

Reflection: Superstar Power

Your **Mean** should be slightly higher than your **Median**. Look back at your sorted list. Why did that "superstar" 2.4kg animal pull the mean upward while the median stayed fixed in the middle?

Statypus Insight: Resistance Preview

It appears that the mean is more affected by extreme values than the median. We will make a definitive statement about this in Section 3.4.

CHAPTER 4

2. Scaling Up with R: Scavenger Hunt

Manual calculations help us feel the math, but real science involves the full sample found in the platypus research studies we first saw in the worksheet from Section 3.1.

Bill the Statypus says: G'day, team! We've got a fresh batch of data for our new chapter. Before you start your analysis, make sure you pull the newest archive file from the server! Run the following command in your console to get everything ready:

```
load(url("https://statypus.org/files/StatypusCh4.RData"))
```

It's also in the book like always, but once that's loaded, you're ready to hunt for the truth!

Coding Corner: The Hunt for Centrality

Open your digital textbook at **r.statypus.org** and navigate to **Section 4.1**.

- Find the two R functions used to calculate the "centers" of a quantitative variable.
- In your console, run these functions using the `WeightF` variable from the `PlatypusData2` dataframe.
- What is the result of each calculation?

Sally the Statypus says: Don't worry! We will consult the book to find a fix for this unexpected result.

Reflection: The Missing Link

Why did R return `NA` instead of a number? Go back to the textbook and find the specific argument (the "switch") you must add inside the parentheses to tell R to ignore missing values.

What is the corrected code for the Mean?

Bill the Statypus says: See, we use software that is robust enough to handle nearly any obstacle. You just have to tell it that it can overlook missing data. By default it is simply telling you that the average of 3 and "I don't know" is "I don't know" which it calls `NA`.

Reflection: The Regional Results

While our field logs are local, we don't know if these results represent the broad population of the entire population. Once you have applied the "na.rm" fix, record the true regional centers:

- **Regional Mean:**
- **Regional Median:**

Coding Corner: Using AI to Make Better Graphics

Ask AI to help you add vertical lines showing the mean and median to a histogram of the `WeightF` variable. Make a sketch below.

3. Final Research Conclusion

Replication Report: Central Location

1. **The Resistant Metric:** If we discovered one "Megastatypus" that weighed 10.0kg, which center do you think would change the most: the Mean or the Median? Why?

2. **The Missing Variable:** Our histograms appear to have two peaks rather than the one we normally need to discuss skew. What variable might we be overlooking that would account for 2 different subpopulations within the data.