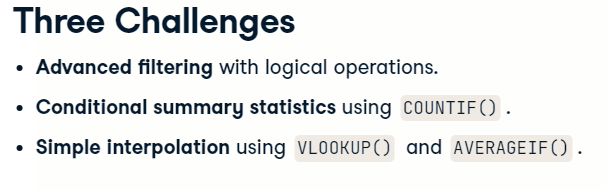
# Intermediate Spreadsheets

**Got It!**

## 1. Bringing it all together

Hurrah! You're almost finished! The last few exercises consist of three data analysis challenges.

## 2. Three Challenges



The first challenge is apply your skills with logical operations from Chapter 3 to filtering datasets. The second challenge is to calculate conditional summary statistics using the COUNTIF() function. Finally, you'll learn how to do simple imputation using VLOOKUP() and AVERAGEIF(). This means calculating estimates for missing values.

## 3. Let's practice!

Have fun with the final exercises.

#### (1) Advanced filtering

Time to practice the logical operations and filtering techniques you learned about in Chapter 3.

##### Instructions

* In column H, define a logical condition where the count of each butterfly type is greater than 50.
* In cell I2, define a filter on the whole dataset (columns A to G), using the logical condition in column H.

#### (2) Conditional summary statistics

In [**Data Analysis with Spreadsheets**](https://spreadsheets.datacamp.com/courses/data-analysis-with-spreadsheets), you saw how to use [**COUNTIF()**](https://support.google.com/docs/answer/3093480) to calculate summary statistics. Here's you'll take it one step further using the related [**COUNTIFS()**](https://support.google.com/docs/answer/3256550) function, which lets you pass multiple conditions to it.

Arguments to COUNTIFS() come in pairs: a range of values to filter on, best given as absolute addresses, and a condition. The condition is text consisting of

1. a value to match, or
2. a comparison operator (=, <, >=, etc.) and a number.

For example, COUNTIFS(A2:A100, "DataCamp", B2:B100, ">10") counts the number of values where column A matches "DataCamp" **and** column B is greater than 10.

##### Instructions

* In cell H2, get the unique Area values using UNIQUE().
* In column I, use COUNTIF() to get the count of each Area. Pass the data range from A2 to A45 as absolute addresses, and the filter criteria from H2 to H9.
* In column J, use COUNTIFS() to get the count of each area with more than 20 Swallow-tails. the first two arguments are the same as the previous step, then add a condition for column D to be greater than 20.

#### (3) Simple imputation

If you have missing data, you can use a set of techniques known as imputation to substitute guesses for those values. A very simple form of imputation is to substitute the average of the non-missing values for that group.

This technique involves 3 steps: use [**AVERAGEIF()**](https://support.google.com/docs/answer/3256529) to get the group averages, [**VLOOKUP()**](https://support.google.com/docs/answer/3093318) to join these back to the original dataset, and [**ISBLANK()**](https://support.google.com/docs/answer/3093290) to locate missing values.

The dataset has been edited so some Brush-footed counts are missing.

##### Instructions

* In cell J2, get the unique Areas.
* In column K, rows 2 to 9, use AVERAGEIF() to calculate the average count of Brush-footeds by Area. Pass it the area data as absolute addresses, the unique areas, and the Brush-footed count data as absolute addresses.
* In column H, join the average counts back to the original dataset. VLOOKUP() takes 4 arguments. It needs the Area from column A, the data range of the table you just created (J2 to K9) as absolute addresses, the column in the that table that contains the averages (2), and whether or not that table is sorted (FALSE).
* In column I, IF() column G is blank, take the value from column H, else take the value from column G.