Getting Started with Dataland
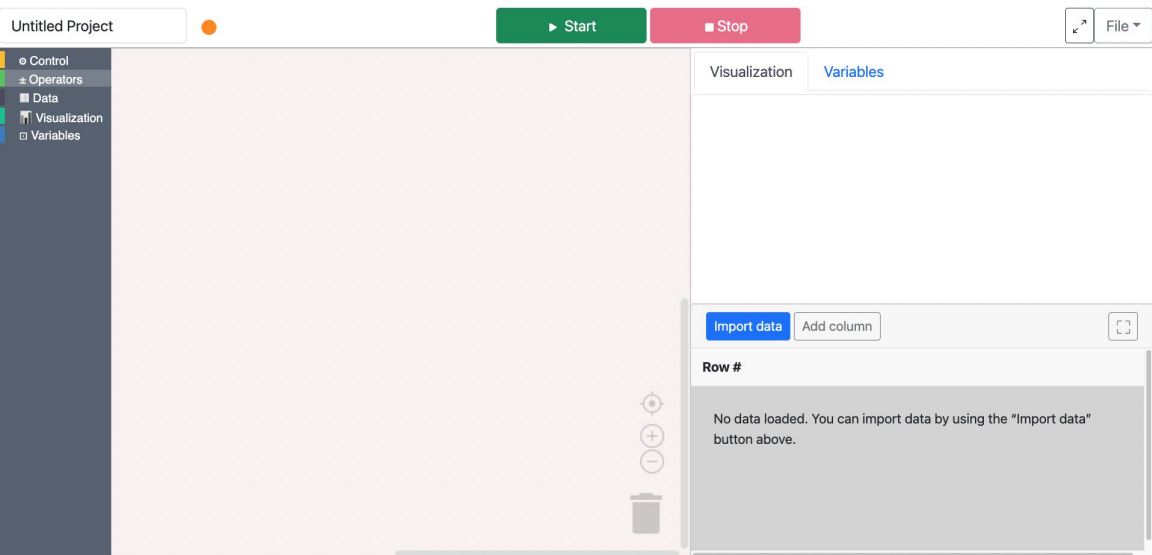
By the Dataland team

Dataland is a programming system. It helps you analyze and plot data by dragging and dropping programming blocks.

This guide will get you started with Dataland!

Getting Started

This is the what you will see when opening up Dataland.
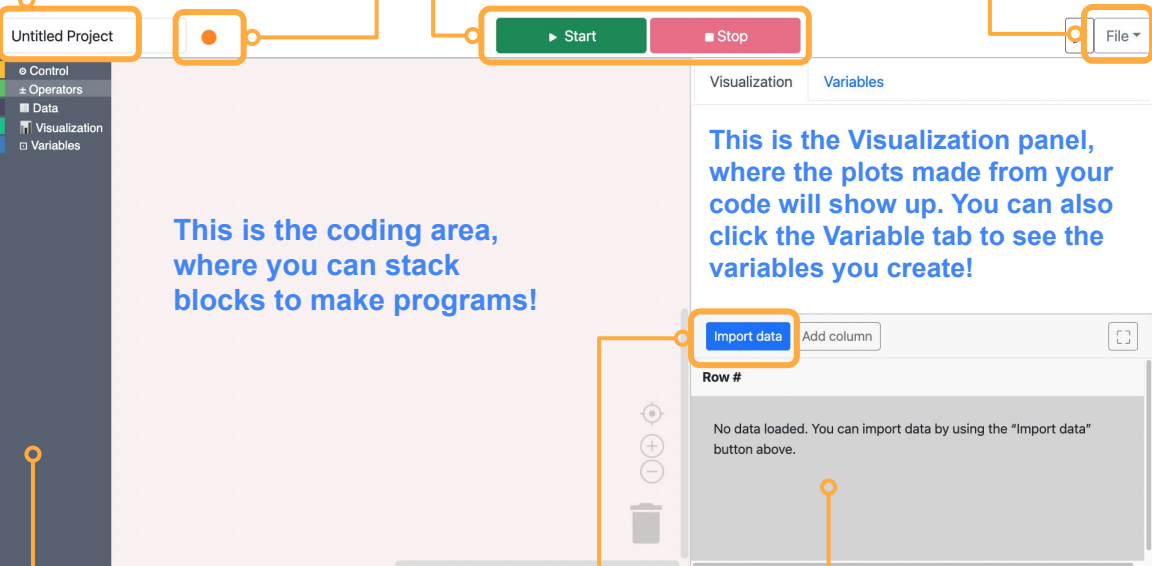Let's take a tour and learn what each part does!

A Tour of Dataland

Dataland automatically saves your project every few seconds. Hover on this doc to see the status of your project - whether it is saved or not!

You can load projects from your computer or download project to your computer here!

You can type the name of your project in this text box!

To run or stop your code, click these buttons!

Untitled Project

● Start ■ Stop

File ▾

Visualization | Variables

⊙ Control
± Operators
▦ Data
📊 Visualization
▢ Variables

This is the coding area, where you can stack blocks to make programs!

This is the Visualization panel, where the plots made from your code will show up. You can also click the Variable tab to see the variables you create!

Import data | Add column

Row #

No data loaded. You can import data by using the "Import data" button above.

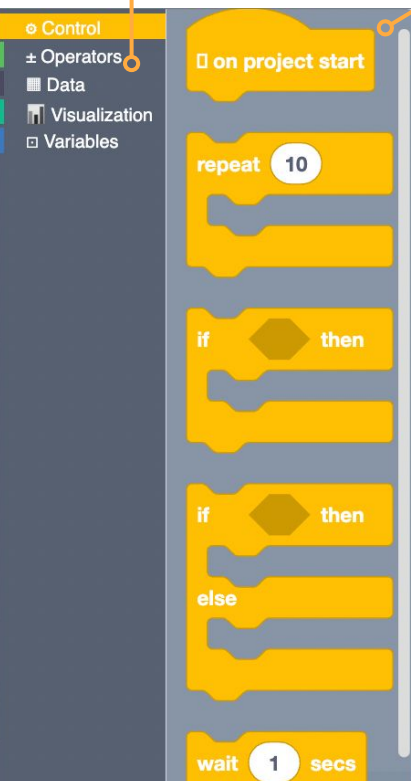This is the toolbox, you can select programming blocks from the different categories above!

You can import data from your computer by clicking this button and select from your files!

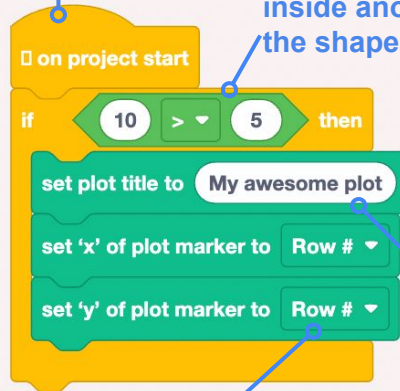The data that you import from your computer will show up here as a table.

A Closer Look at the Coding Area

In the toolbox, you can click on each category and see a panel, from which you can select different programming blocks.

You can drag and drop blocks from the panel to the coding area. Stack blocks together to make a program!

Control
± Operators
Data
Visualization
Variables

on project start

repeat 10

if   then

if   then

else

wait 1 secs

To make a program runnable, you always need to start with this "on project start" block.

You can put some blocks inside another. Look at the shapes for hints!

on project start

if   10 > 5   then

set plot title to   My awesome plot
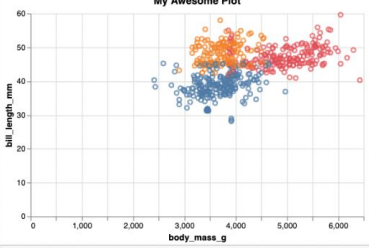
set 'x' of plot marker to   Row # ▼

set 'y' of plot marker to   Row # ▼

You can enter text or numbers in some blocks.

You can select from dropdown menus in some blocks.

This is the trash bin. You can drag the blocks that you don't need here. You can also find reuse the blocks here later!

A Closer Look at the Data Table

You can import a dataset from your own computer to Dataland!
Your dataset must be in the format of a CSV (comma-separated
values) file, with a file name that ends with ".csv".
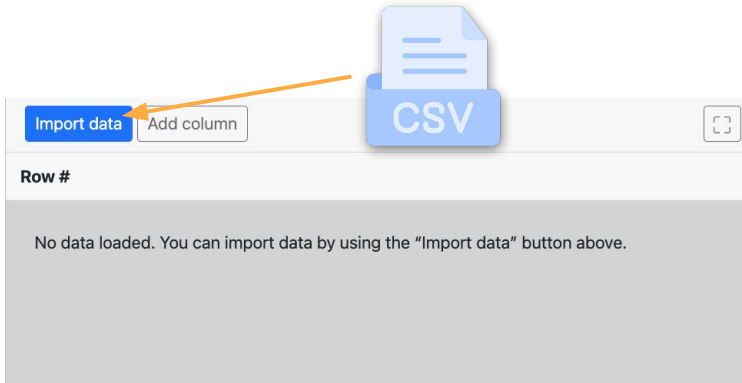
Once successfully imported, the dataset will show up as a table
in the lower right corner of the Dataland screen.

Click here
to expand
the table to
full screen!

A Tour of Dataland

This is what Dataland looks like when you have code and data added! The code in the coding area will talk to the data in the table, and make a plot as a result.

Now let's see some example projects that you can make with data using Dataland!

All the examples in this guide uses a dataset about the characteristics of different species of penguin.
The dataset can be downloaded from: [INSERT URL]

Making a Scatter Plot

A scatter plot shows data as dots. It presents the relationship between a pair of numerical data.

For example, let's make a scatter plot that shows the flipper length and bill depth of three penguin species.

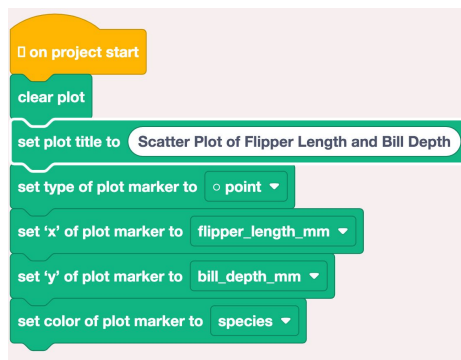Scatter Plot of Flipper Length and Bill Depth
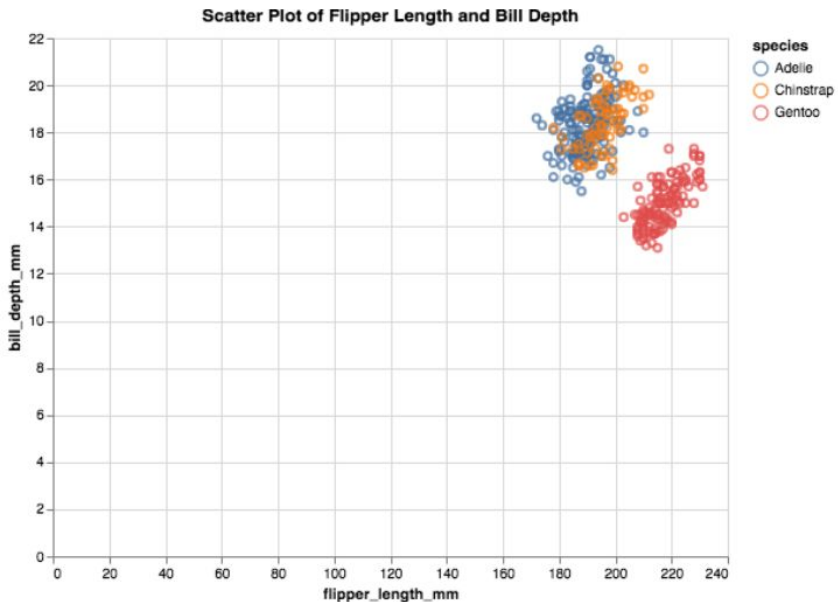
How to make a scatter plot?

To make scatter plot, select blocks from the Visualization category in the toolbox.
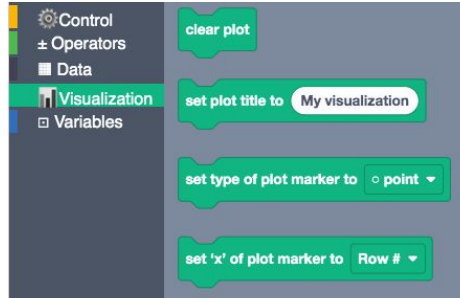
Set the plot marker to *point*. Use the blocks that set "x" axis and "y" axis. If you want to see the groups in your data, set them in different colors.

This block stack generates the example plot on the other side.

Your plot will show up at the Visualization tab on the upper right corner of the screen!

Filtering Data

Filtering data means getting rows in the data table that meet certain conditions. It can be useful when you only need a certain part of the dataset.

For example, let's specifically plot the bill length and depth of the *Adelie* penguins from a dataset full of different penguin species using the technique of filtering data.

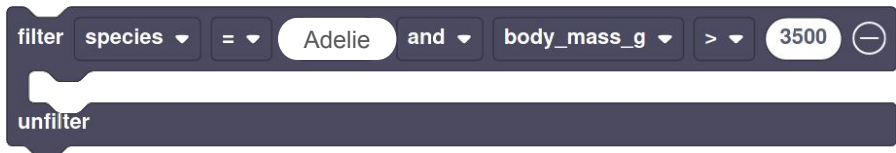Adelie penguins' bill length and depth
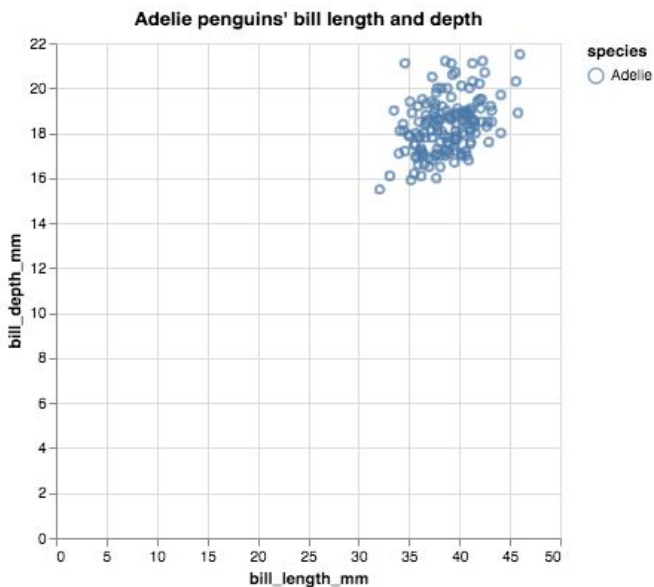
How to filter data?

To create a filter, go to the Data category in the toolbox and select the Filter block. Set the filter condition using the dropdown menus and the text box. In this example, the filter condition is when the value in the "species" column equals "Adelie".
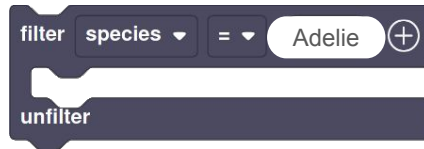
You can also add more conditions to the filter by clicking the "+" sign.

The filter block can be useful when paired with other blocks. For example, this block stack generates the plot shown on the other side. It filters all the rows where the species is *Adelie* and plots the bill length and depth of these rows.
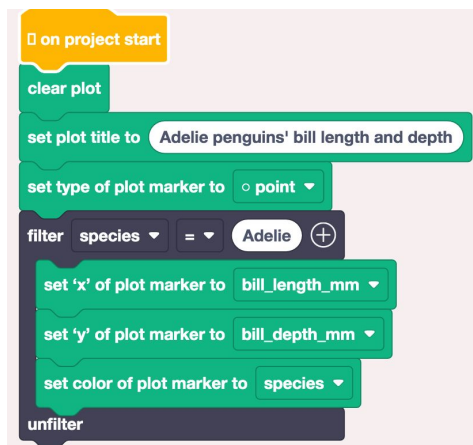
Creating a Variable

A variable is a placeholder for a value that may change. In our case, the value can be a number or text. Variables can be useful when used with other blocks.

Let's create a variable!

How to create a variable?

To create a new variable, go to the Variable category in the toolbox. Click "Create variable". A window will pop up and you can name the new variable in the popup window.
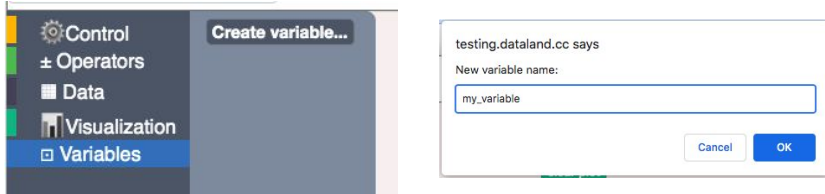
The blocks of the new variable can be found in the toolbox. The oval block is the variable itself. It can be used with the control, operation, and data blocks.

The Set block gives a value to the variable. The value can be a number or text. For example, in the following code we set the variable "my_variable" to 5.
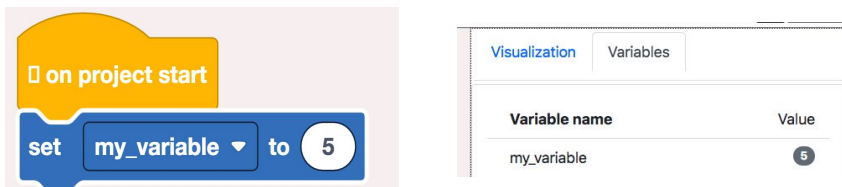
The value of the variables can be found in the Variable tab on the upper right corner of the screen.
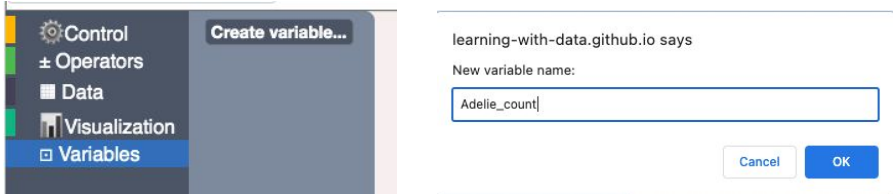
Counting Data

Sometimes, you may want to count certain things in your data and save it for later use. Variables can be handy in this case!

For example, let's count the number of *Adelie* penguins (a particular species of penguin) in a dataset full of different species of penguins.

| Variable name	Value
Adelie_count	217

How to count data?

First, create a new variable called "Adelie_count" to store the count of Adelie penguins.


Since we are only counting the number of *Adelie* penguins, let's prepare a filter that selects data of the *Adelie* species.

The "number of rows" block from the Data category in the toolbox can tell us the number of rows selected in the current condition.
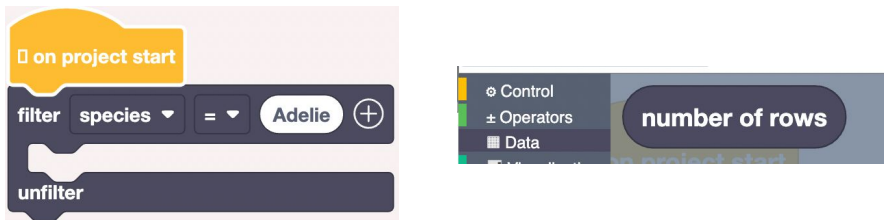
Let's set our variable Adelie_count to be "number of rows" and put it inside the filter. This code will assign the number of rows of Adelie penguin to our variable Adelie_count! You can view the value of the variable in the Variable panel.
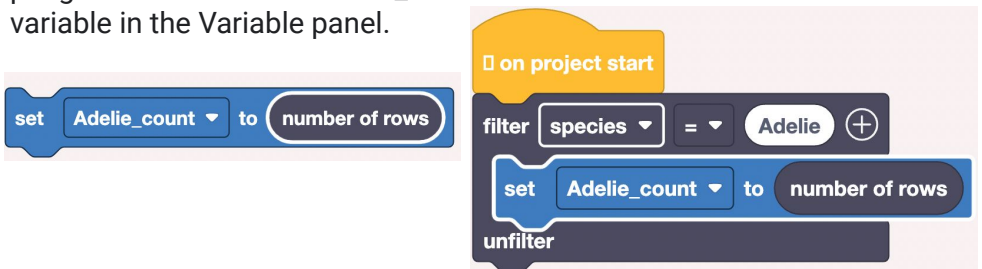
Adding New Data to the Table

Sometimes, you may want to add a new column to the data table to help with your analysis.

For example, let's create a new column in our table. The new column is named "new_column" and is filled with all "1"s.

new_column
1
1
1
1
1

How to add new data to the table?

To add a new column, click the "Add column" button above the table. A window will pop up, and you can name the new column in the popup window. For example, here we name it as "new_column".
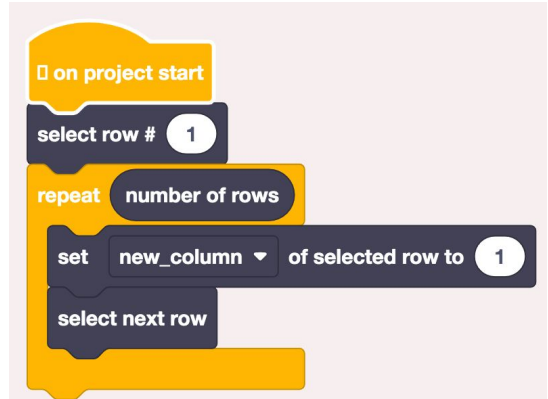

The new column will show up as an empty column at the rightmost of the table.

To fill the new column in with data, select each row of the column and set it with a value. For example, the block stack above generates the column of all "1"s shown on the other side. It starts from row #1, sets it with value "1", and repeats this step for all the rows in the column (i.e., repeat "number of row" times).

Analyzing Data by Groups

Sometimes, a dataset contains different groups. It may be useful to analyze data within each groups.

For example, in our data, we have different species of penguins. Let's find out the maximum bill length of each species.

**Maximum bill length in each species**
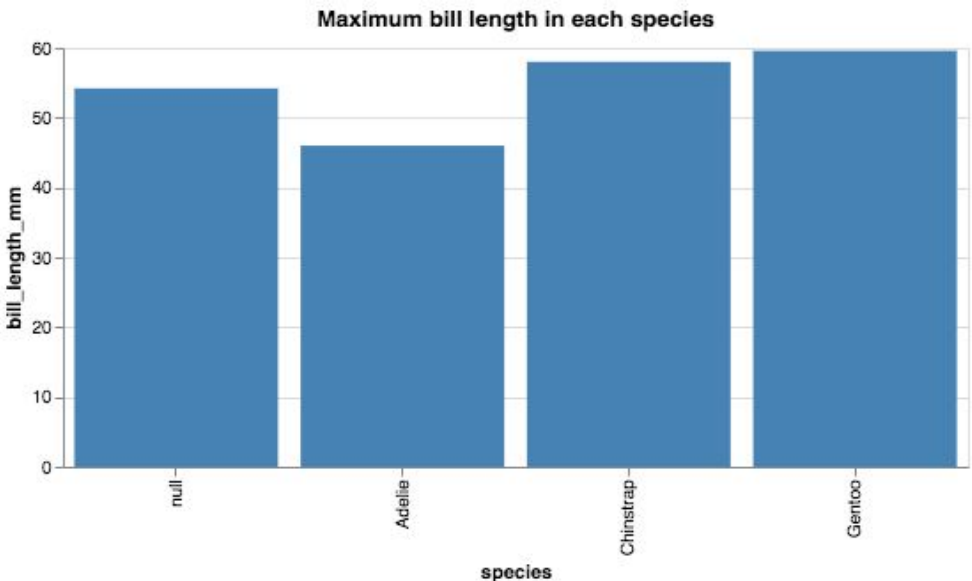
How to analyze data by groups?

Go to the Data category in the toolbox and select the Group By block. The block allows you to say how you'd like to group the data.

In the leftmost dropdown menu, select the column of the table that contains the groups. In the middle dropdown menu, select the column that you'd like to analyze. In the rightmost dropdown menu, set the how you want to analyze it (e.g., mean, maximum, etc.).
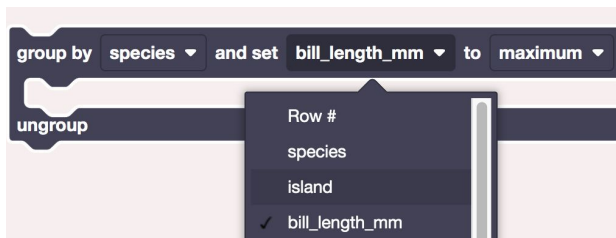

The Group By block can be useful when paired with other blocks. For example, this block stack generates the plot on the front side. It groups the data based on species and plots the maximum bill length within each species.
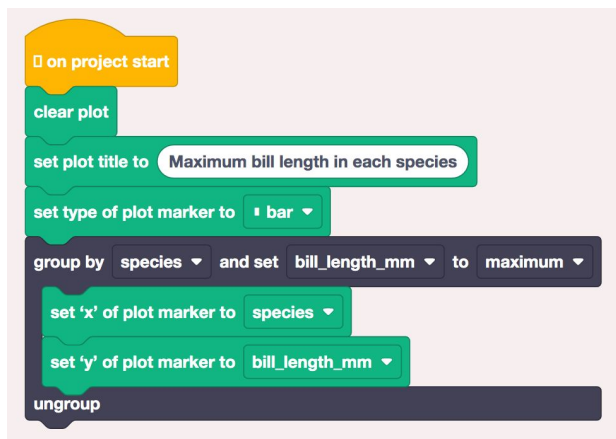
Plotting Data on a Map

Sometimes, we can be working with data that contains geographical information. We may want to plot where each data point is on a map.

For example, let's plot the locations of the penguins in our dataset on a map.
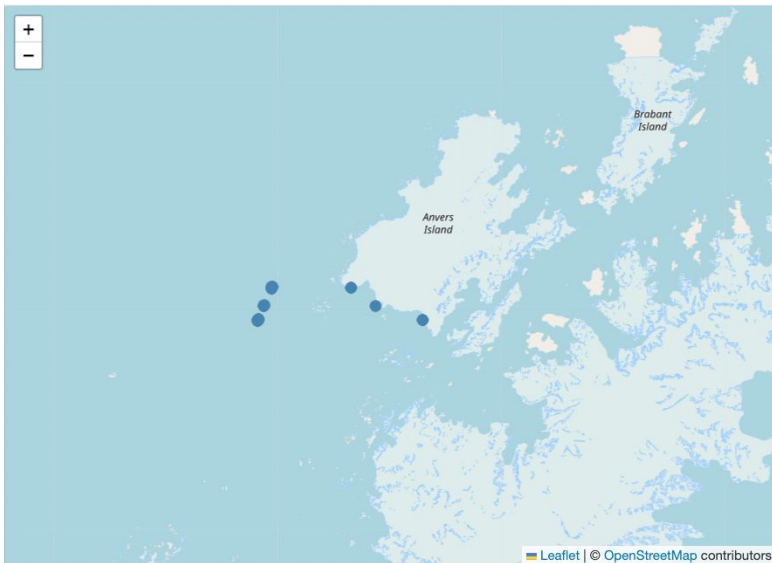
How to plot data on a map?

First, make sure you are in the Map microworld. In this microworld, you can see a map in the visualization panel and a category of Maps blocks in the toolbox. You can select blocks from the Maps category in the toolbox to plot on the map.
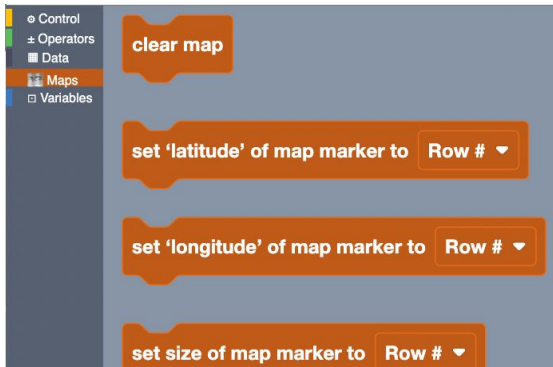
Make sure your data has geographical information, that is, a column called "latitude" and a column called "longitude". The following block stack will plot the latitude and longitude of all the data points in your dataset on the map. You can also combine these blocks with other blocks (e.g., filter) to make a more complicated project!

Now it is your turn to make exciting projects with data using Dataland!