Learning data science with Dataland

By the Dataland team

This handout aims to help you design your own data science curriculum with Dataland.

The following contains a sample list of learning goals with Dataland.

For each learning goal, we offer a sample template of tasks for students to engage with using Dataland. We also provide an example task for each learning goal using the penguins dataset from https://allisonhorst.github.io/palmerpenguins/.

Sample Learning Goals

☐ Reading tables

☐ Filtering

☐ Variables & counting

☐ Grouping & aggregation

☐ Scatter plot, clusters, & outliers

☐ Operating and adding columns

☐ Bar plot

☐ Geographical data

Reading tables

Sample task template:

Read the data table loaded in Dataland. Think about and discuss the following questions:

- What do you think each row represents?

- What do you think each column represents?

- How do you think data were collected for each column? For each row?

- What are the names and scales of measurements? What does it mean to categorize and measure things this way?

Filtering

Sample task template:

Create a scatter plot of data from *[column A]* and *[column B]* that *[fits a certain condition]*.

- **Hint:** use the Filter block.

On top of the current scatter plot, further narrow down the plotted data that *[fits another condition]*.

- **Hint:** use more than one condition in the filter block by clicking on the plus icon at the end of the block. Or, use more than one filter block.

Filtering

Example task:

Create a scatter plot of body mass and flipper length of the Adelie species of penguins.

- **Hint:** use the Filter block.

On top of the current scatter plot, further narrow down the penguins that are under 4,500 grams in body mass.

- **Hint:** use more than one condition in the filter block by clicking on the plus icon at the end of the block. Or, use more than one filter block.

Variables & counting

Sample task template:

Count the number of data points that fits *[a certain condition]*.

- **Hint**: use the Variable block to help with the counting.

- **Hint:** after running the program, view the value of variable in the panel on the right.

Variables & counting

Example task:

Count the number of Adelie penguins that are under 4,500 grams in body mass.

- **Hint:** use the Variable block to help with the counting.

- **Hint:** after running the program, view the value of variable in the panel on the right.

Grouping & aggregation

Sample task template:

Compare the minimum, maximum, mean, or number of unique values in *[a column]* for each *[categories in another column]*.

- **Hint:** use the Groupby block and group on the categories.

Grouping & aggregation

Example task:

Compare the mean values of bill length among Adelie penguins on each island.

- **Hint:** use the Groupby block and group on the categories.

Scatter plots, clusters, & outliers

Sample task template:

Make a scatter plot of *[column A]* and *[column B]*.
Look at the shape of the data, describe what you
discovered in the plot.

- **Hint:** pay attention to the clusters of data
  where a lot of points seem to be close to each
  other.

- **Hint:** pay attention to the outliers which seem
  to be far away from the majority of the data.

What does the plot indicate about the different
[categories]?

- **Hint:** use different colors for the different
  [categories] in the plot.

Scatter plots, clusters, & outliers

Example task:

Make a scatter plot of flipper length and body mass. Look at the shape of the data, describe what you discovered in the plot.

- **Hint:** pay attention to the clusters of data where a lot of points seem to be close to each other.

- **Hint:** pay attention to the outliers which seem to be far away from the majority of the data.

What does the plot indicate about the different species of penguins?

- **Hint:** use different colors for the different species in the plot.

Operating and adding columns

Sample task template:

Operate the data in each row of *[column A]*. Put this new data into a new column.

- **Hint:** add a new column using the "Add column" button on top of the data table.

- **Hint:** use the Operation blocks for calculation.

Operating and adding columns

Example task:

Calculate the ratio of the data in each row of bill length and bill depth. Put this new data into a new column named "Bill Ratios".

- **Hint:** add a new column using the "Add column" button on top of the data table.

- **Hint:** use the Operation blocks for calculation.

Bar plot

Sample task template:

Compare the values of *[certain column(s)]* across different *[categories]*.

- **Hint:** create bar plots of *[values]* over the *[categories]*.

- **Hint:** you may use different colors to indicate the different columns.

Bar plot

Example task:

Compare the count of the different species of penguins across the years.

- **Hint:** create bar plots of count of rows over the species and years.

- **Hint:** you may use different colors to indicate the different columns.

Geographical data

Sample task template:

Find the locations of the data points that fit *[certain conditions]*.

- **Hint:** Plot the latitude and longitude on the map.

- **Hint:** Filter on *[the certain conditions]*.

Geographical data

Example task:

Find the locations of the Adelie penguins on the Dream island.

- **Hint:** Plot the latitude and longitude on the map.

- **Hint:** Filter on the Adelie penguins on the Dream island.