

Practical Exercise 2: Graphing Data

Pathways to Data Analytics: Essentials

August 29, 2018

1 Introduction

There's a reason why data scientists spend so much time exploring data using graphics. Relying only on data summaries like means, variances, and correlations can be dangerous, because wildly different data sets can give similar results.

2 The task

Using the dinosaur data set from the last exercise, this exercise requires you to complete four tasks:

1. Create a valid bar chart.
2. Create a valid pie chart.
3. Create a valid scatterplot.
4. Work out which country has provided us with the most amount of discovered species out of the species in our sample. Find a valid way to convey this information.

Remember, graphs have to be meaningful!

3 Source

If you are confident that you corrected all of the issues in the data set from exercise 1, you can use that data set for this exercise. Alternatively, you can download a cleaned version of that data set from https://github.com/LewBrace/LewBrace-PtDA_essentials_Exercise_2. By default, the file will be downloaded in a zipped folder, along with a Read Me file, which can be ignored as there is no information in there, and a copy of this pdf.

You will therefore need to 'unzip' the folder, which is a fairly simple operation. On Windows machines, just right-click on the file and select 'Extract all'.

4 Support

If you need a refresher on the content covered in the lecture, the lecture slides can be downloaded from https://github.com/LewBrace/PtDA_essentials_lecture_slides.

As always, if you have any questions, I can be contacted via `l.brace@exeter.ac.uk`.