Data Science Assignment 1

Nike Marie Pulow – Henri Paul Heyden
stu239549 – stu240825

1 Data Acquisition

1. We selected „Plum, Christoph; Gerhard, Miriam; Smykala, Mike (2024): Growth rate, carrying capacity and stoichiometry of Antarctic phytoplankton in response to temperature and nitrogen:phosphorus supply interactions [dataset bundled publication]. PANGAEA, https://doi.org/10.1594/PANGAEA.971329"

2.

   3.1 **Who funded the dataset?**
   The dataset was funded by the German Research Foundation (DFG) with grant-number: 442927665: „Altered stoichiometry and species interactions in coastal Antarctic plankton communities under global climate change (https://gepris.dfg.de/gepris/projekt/442927665)"

   3.2 **What data does each instance consist of?**
   The data consists of raw measurements of how different antarctic phytoplankton species react to different conditions. The different dimensions are represented in tables where each dimensional value is either a string to, for example define the specific species used, or a numeric value to, for example display the temperatures or the concentration of phytoplankton.

   3.3 **How was the data associated with each instance acquired?**
   In the four datasets we also define the dimension „Study type" which describes on how the specific values were acquired. Going through them you can see that in each tuple the values is „Laboratory experiment", so all instances were measurements done in laboratory conditions.

   3.4 **Was any preprocessing/cleaning/labeling of the data done?**
   No there seems to have been no change involved, as the measured values under the specific conditions are all raw measurements which they conducted with the specific equipment referenced.

   3.5 **Has the dataset been used for any tasks already?**
   As the datasets have been funded, this seems likely, but they also don't seem to be that used since our download was the first of it on the URL specified, nor are there any references to uses.

   3.6 **How will the dataset be distributed?**
   The datasets have DOIs as well as the collection of datasets that is referenced does, but other than the copies stored on PANGAEA, there seem to be no other (re-)distributors linked or referenced.

   3.7 **How can the owner/curator/manager of the dataset be contacted?**
   The requester, curator as well as the main researcher Dr. Christoph Plum can be contacted via mail as specified on the website of the DFG.

3. The question whether the datasets have been used already is relevant since because the data seems to be mainly distributed by the PANGAEA, we can get a broad estimate on how relevant it is to the general state of research. One must though value that the datasets are quite new as they have been published this year.
   Given the context of research field, it seems likely that the topic is not that popular, generally spoken, but may have a niche for further study. For example as the data provided is raw data, one might look at calculated variances between values, as well as expected values, and graph this processed data to determine which cultures survive under conditions caused by climate change best and which will likely die, while also looking at how definite answers to these questions are.
   Looking at the nature of these datasets, as they contain raw data, it makes it likely, that it could be well used for further, more specific studies that contextualize the data, like in the just given example.

2 Pandas tutorial

`merge this somehow`

3 Fairness

`merge this somehow 2`