A red and white hat with a blue letter

Description automatically generated

Winter Activity

|  |
| --- |
|  |
|  |
|  |

Data Analyst Activity Guide

# Palmer Penguins

## About the data

Resource: penguins.csv

These happy looking penguins’ data were collected and made available by Dr. Kristen Gorman and the Palmer Station Antarctica LTER.

The penguins data was released through a creative commons license with the aim of providing real data that can be used as an alternative to the iris dataset. This makes it great for practicing your Data Analyst skills and getting into the winter season spirit!

Artwork by @allison_horst

An image of Chinstrap, Gentoo, and Adelie penguins- Artwork by @allison\_horst

There are data for 344 penguins. Within the data set there are 3 different species collected from 3 islands in the Palmer Archipelago Antarctica. The data set records (where available) their bill length and depth, flipper length, body mass, and sex, as well as the year the observation was made.

The inclusion of the year column means that the data set can potentially be used to practice techniques and answer question types that were not possible with the iris data set – increasing the range of investigations that can be practiced.

The data was gathered as part of wider research into the western side of the Antarctic Peninsula to gain understanding into the mechanisms of change and the ability to predict changes to the ecosystem.

## Task

Goals:

* To gain insight and understanding into the Palmer Penguins using the penguins.csv data provided.
* To practice and keep your skills *warm*!

You have the freedom to tackle this with whichever tools and methods you choose, but we have provided some ideas below to get you started.

**Online research**

What do people already know about Chinstrap, Gentoo, and Adelie Penguins?

When you analyse the data, look out for anything expected or unexpected!

**Analytical Inspiration**

Choose a tool to use:

* Python with Pandas (or other packages)
* Power BI
* Excel
* Orange 3
* And more!

**Questions**

You may choose to ask and tackle your own investigations, but here are some ideas:

* What data cleaning and preparation is needed?
* What have you practiced with the iris data set before that you could also practice with penguins?
* Is there a difference in bill or flipper measurements for each group of penguins?
* What visualisations could you create and what do they tell you?
* How are things changing over time?
* Can you predict what type of penguin it is from one or more of the other measurements?

# References

Horst AM, Hill AP, Gorman KB (2020). Palmerpenguins: Palmer Archipelago (Antarctica) penguin data. R package version 0.1.0. <https://allisonhorst.github.io/palmerpenguins/>. Doi: 10.5281/zenodo. 3960218

## 