

Data 5100 – Andrex Ibiza

Course Project Checkpoint 2

2024-01-23

The ``penguins_raw`` dataset from the ``palmerpenguins`` package contains data on 344 penguin observations with 17 variables. The original variables in the dataset were “studyName”, “Sample Number”, “Species”, “Region”, “Island”, “Stage”, “Individual ID”, “Clutch Completion”, “Date Egg”, “Culmen Length (mm)”, “Culmen Depth (mm)”, “Flipper Length (mm)”, “Body Mass (g)”, “Sex”, “Delta 15 N (o/oo)”, “Delta 13 C (o/oo)” and “Comments”.

I made the following changes to the dataset:

- “studyName”: changed to “study\_name”
- “Sample Number”: changed to “sample\_number” and converted to integer
- “Species”: changed to “species” and converted to factor
- “Region”: changed to “region” and converted to factor
- “Island”: changed to “island” and converted to factor
- “Stage”: changed to “stage” and converted to factor
- “Individual ID”: changed to “individual\_id”
- “Clutch Completion”: changed to “clutch\_completion” and converted to factor
- “Date Egg”: changed to “date\_egg”
- “Culmen Length (mm)”: changed to “culmen\_length\_mm”
- “Culmen Depth (mm)”: changed to “culmen\_depth\_mm”
- “Flipper Length (mm)”: changed to “flipper\_length\_mm”
- “Body Mass (g)”: changed to “body\_mass\_g”
- “Sex”: changed to “sex” and converted to factor
- “Delta 15 N (o/oo)”: changed to “delta\_15\_n”
- “Delta 13 C (o/oo)” : changed to “delta\_13\_c”
- “Comments”: changed to “comments”

The cleansed dataset contains the following variables:

- “study\_name” (character) – Sampling expedition from which data were collected, generated, etc.
- “sample\_number” (integer) – an integer denoting the continuous numbering sequence for each sample
- “species” (factor) – 3 levels: “Adelie Penguin (*Pygoscelis adeliae*)”, “Chinstrap penguin (*Pygoscelis antarctica*)”, and “Gentoo penguin (*Pygoscelis papua*)”.
- “region” (factor) – only 1 level, “Anvers”, denoting the region of Palmer LTER sampling grid
- “island” (factor) – 3 levels – “Biscoe”, “Dream”, and “Torgersen” denoting the island near Palmer Station where samples were collected
- “stage” (factor) – 1 level – “Adult, 1 Egg Stage” – denotes reproductive stage at sampling
- “individual\_id” (character) – alphanumeric unique identifier for each individual penguin represented in the data

- “clutch\_completion” (factor) – 2 factors – “Yes” or “No” – denoting whether the study nest was observed with a full clutch, i.e. 2 eggs
- “date\_egg” (date) denotes the date study nest observed with one egg (sampled)
- “culmen\_length\_mm” (numeric) – a number in millimeters denoting the length of the dorsal ridge of a bird’s bill
- “culmen\_depth\_mm” (numeric) – a number in millimeters denoting the depth of the dorsal ridge of a bird’s bill
- “flipper\_length\_mm” (numeric) – a number in millimeters denoting the length of the penguin’s flipper
- “body\_mass\_g” (numeric) – a number denoting the body mass of the penguin in grams
- “sex” (factor) – denotes penguin’s sex
- “delta\_15\_n” (numeric) - a number denoting the measure of the ratio of stable isotopes  $^{15}\text{N}:$  $^{14}\text{N}$
- “delta\_13\_c” (numeric) - a number denoting the measure of the ratio of stable isotopes  $^{13}\text{C}:$  $^{12}\text{C}$
- “comments” (character) – text providing additional relevant information for the data

The dimensions of the cleansed dataset are 344 x 17.