

Week 07 Tools for meta data

Open and reproducible science: dependable computations and statistics

Homework solutions

In this exercise we will explore an example dataset which is part of the palmerpenguins R package. More precisely we will explore the associated metadata of the Adelie data set on which the package is partly based on.

You can find the Adelie data set and its XML meta data starting from the section Data citations on <https://allisonhorst.github.io/palmerpenguins/#references>

Or you can download the associated meta data directly using this link: <https://portal.edirepository.org/nis/metadataviewer?packageid=knbnlterpal.219.5&contentType=application/xml>

Download is also possible directly within R with

```
download.file("https://portal.edirepository.org/nis/metadataviewer?packageid=knbnlterpal.219.5&contentType=application/xml",
             here::here("data", "adelie.xml"))
```

Exercise

Read the data into R using the EML package and answer the following questions:

```
library("EML")
eml <- read_eml(here::here("data", "adelie.xml"))
```

1

How many attributes does the data table have? Use the function `get_attributes` of the EMLpackage.

```
attributes<-get_attributes(eml$dataset$dataTable$attributeList)
dim(attributes[[1]])[2]
```

```
## [1] 12
```

There are 12 attributes in the data table.

2

Inspect the attribute list. Which attributes correspond to the bill length and depth provided by the palmerpenguins package?

```
attributes[[1]][10:11,2:3]
```

```
##      attributeName      attributeDefinition
## 10 Culmen Length length of the dorsal ridge of a bird's bill
## 11 Culmen Depth  depth of the dorsal ridge of a bird's bill
```

Culmen Length corresponds to bill length and Culmen Depth corresponds to bill depth.

3

Find the information on the corresponding measurement unit of bill length and depth.

```
attributes[[1]][10:11,8]
```

```
## [1] "millimeter" "millimeter"
```

The unit of measurement of Culmen Length and Culmen Depth are in millimeter.

4

Which numerical variables have been measured but are not contained in the `palmerpenguins` package?

```
library(palmerpenguins)
names(penguins)
```

```
## [1] "species"      "island"      "bill_length_mm"
## [4] "bill_depth_mm" "flipper_length_mm" "body_mass_g"
## [7] "sex"          "year"
```

```
attributes[[1]][15:16,c(2,3,7)]
```

```
##      attributeName      attributeDefinition
## 15      Delta 15 N a measure of the ratio of stable isotopes 15N:14N
## 16      Delta 13 C a measure of the ratio of stable isotopes 13C:12C
##      domain
## 15 numericDomain
## 16 numericDomain
```

Attributes Delta 15 N, Delta 13 C are missing in the `palmerpenguins` package.

5

Find the url of the file containing the data in the `physical` information of the meta data.

```
eml$dataset$dataTable$physical$distribution
```

```
## $online
## $online$url
## [1] "https://pasta.lternet.edu/package/data/eml/knb-lter-pal/219/5/002f3893385f710df69eebe893144ff"
```

The url of the file containing the data in the `physical` information is <https://pasta.lternet.edu/package/data/eml/knb-lter-pal/219/5/002f3893385f710df69eebe893144ff>.

6 (2 points)

Does the meta data contain information on the precise geographic region or the species?

```
coverage <- eml_get(eml, "coverage")
names(coverage)
```

```
## [1] "@context"          "temporalCoverage"
```

No, it only contains temporal information.

7 (2 points)

Find the methods description and describe how was the sex of the penguins was determined?

```
cat(eml_get(eml, "methods")$methodStep$description$section)
```

General Methods

Each season, study nests, where pairs of adults were present, were individually marked and chosen before the onset of egg-laying, and consistently monitored. When study nests were found at the one-egg stage, both adults were captured to obtain blood samples used for molecular sexing and stable isotope analyses, and measurements of structural size and body mass. At the time of capture, each adult penguin was quickly blood sampled (~1 ml) from the brachial vein using a sterile 3 ml syringe and heparinized infusion needle. Collected blood was stored in 1.5 ml micro-centrifuge tubes that were kept cool. In the field, a small amount of whole blood was smeared on clean filter paper stored in a 1.5 ml micro-centrifuge tube for molecular sexing. Measurements of culmen length and depth (using dial calipers ± 0.1 mm), right flipper (using a ruler ± 1 mm), and body mass (using 5 kg ± 25 g or 10 kg ± 50 g Pesola spring scales and a weigh bag) were obtained to quantify body size variation. After handling, individuals at study nests were further monitored to ensure the pair reached clutch completion, i.e., two eggs. Molecular analyses were conducted at Simon Fraser University following standard PCR protocols, and stable isotope analyses were conducted at the Stable Isotope Facility at the University of California, Davis using an elemental analyzer interfaced with an isotope ratio mass spectrometer

8

What is the publication date of the data?

```
eml$dataset$pubDate
```

```
## [1] "2017-03-29"
```

The publication date is 2017-03-29

9

What is the email address of the creator of the dataset?

```
cr <- eml_get(eml, "creator")
cr[[2]]$electronicMailAddress
```

```
## [1] "kgorman@sfu.ca"
```

The email address of the creator is kgorman@sfu.ca

10

How many keywords are listed?

```
ky <- eml_get(eml, "keywordSet")
length(ky$keyword)
```

```
## [1] 3
```

There are 3 keywords listed.

Submission

Homework submission has two parts:

1. Push your Rmd file to Gitlab.
2. Enter the corresponding Gitlab pages URL into the Open edX text box in the next unit.

Note: the URL will be of the form

https://opensciencecourse.pages.uzh.ch/hs22_dcas/hs22_dcas/USERNAME_Homework_06.html,
where USERNAME is your Gitlab user name. E.g for user jdoe:

https://opensciencecourse.pages.uzh.ch/hs22_dcas/hs22_dcas/jdoe_Homework_06.html

Your submission on Open edX will be peer reviewed. Staff will perform random checks on the peer review.

Your submission on Gitlab will be automatically checked, points from the Open edX assessment may be decreased if your submission is not correct.