Appendix 1

Below we show the utility of the functions from the ***arakno*** package in three examples that are mimicking common situations. The package can be loaded with the following command:

```
> library(arakno)
```

<u>Example 1</u>

This example shows how to find valid species names, authorities, LSID codes and prepare this information in a table, which is a common requirement of papers on community ecology, for example.

You may store the list of names that you want to check in an Excel spreadsheet. Open the spreadsheet, copy the column with names to R via clipboard and run the following command (or import the file by specifying the full path in the quotation marks):

```
> tax <- scan("clipboard", what="char", sep="\t")
> tax
[1] "Mangora acalypha" "Clibiona compta" "Pardossa amendata" "Gyrakancium eraticum"
[5] "Theridion impressum" "Zodarum rubidum" "Agelena similis" "Arraneus"
```

The list contains eight names, the first is a valid name, the second is a valid name with a single typo, the third is a valid name with two typos, the fourth has multiple typos, the fifth is a species that changed genus, the sixth is a junior synonym with a typo, the seventh is a junior synonym of a species that changed genus, and the eight is a genus name with a typo. To check the validity of names use function `checknames` and place them into an object named e.g. `tax1`, as it will be used later:

```
> tax1 <- checknames(tax); tax1

Species                  Best match
[1,] "Clibiona compta"       "Clubiona comta"
[2,] "Pardossa amendata"     "Pardosa amentata"
[3,] "Gyrakancium eraticum"  "Cheiracanthium erraticum"
[4,] "Theridion impressum"   "Phylloneta impressa"
[5,] "Zodarum rubidum"       "Zodarion rubidum"
[6,] "Agelena similis"       "Allagelena gracilens"
[7,] "Arraneus"              "Araneus"
     Note                    Alternative match(es)
[1,] "Misspelling"           NA
[2,] "Misspelling"           NA
[3,] "Misspelling"           NA
[4,] "Nomenclature change"   NA
[5,] "Misspelling"           NA
[6,] "Synonym"               NA
[7,] "Misspelling"           NA
```

The function returned four columns, the first containing the list of original but incorrect species names (the first one was omitted as it is correct), the second one lists possible best matches, the third the reason for mismatch, and the fourth alternative match(es). In all cases a correct name was found.

We select first six names from the second column of `tax1`, by specifying this selection in brackets, and call `authors` to find their authorities:

```
> authors(tax1[1:6, 2])
                   Species          Authors
1          Clubiona comta (C. L. Koch, 1841)
2         Pardosa amentata   C. L. Koch, 1839
3 Cheiracanthium erraticum      (Clerck, 1757)
4     Phylloneta impressa (Walckenaer, 1802)
5         Zodarion rubidum    (L. Koch, 1881)
6     Allagelena gracilens        Simon, 1914
```

Authorities and years for all species in the list are given in the second column. Now we find the taxonomic classification for all these species using function `taxonomy`, without specifying any other arguments:

```
> taxonomy(tax1[1:6, 2])
  Sub/Infraorder          Family           Genus                   Species
2 Araneomorphae         Araneidae         Mangora         Mangora acalypha
3 Araneomorphae       Clubionidae        Clubiona           Clubiona comta
4 Araneomorphae         Lycosidae         Pardosa         Pardosa amentata
5 Araneomorphae Cheiracanthiidae Cheiracanthium Cheiracanthium erraticum
6 Araneomorphae       Theridiidae      Phylloneta      Phylloneta impressa
7 Araneomorphae        Zodariidae        Zodarion          Zodarion rubidum
1 Araneomorphae        Agelenidae       Allagelena     Allagelena gracilens
```

For each species name, the genus, family, and infraorder are shown. For the sake of completeness, we ask for LSID from WSC for all the species by calling `lsid`:

```
> lsid(tax1[1:6, 2])
                   Species                                     LSID
1         Mangora acalypha urn:lsid:nmbe.ch:spidersp:016308
2           Clubiona comta urn:lsid:nmbe.ch:spidersp:024502
3         Pardosa amentata urn:lsid:nmbe.ch:spidersp:018522
4 Cheiracanthium erraticum urn:lsid:nmbe.ch:spidersp:023304
5      Phylloneta impressa urn:lsid:nmbe.ch:spidersp:008433
6         Zodarion rubidum urn:lsid:nmbe.ch:spidersp:026117
7     Allagelena gracilens urn:lsid:nmbe.ch:spidersp:020847
```

Eventually, we can get valid names, authorities, and classification in a single command by calling `taxonomy` with two arguments, `aut=T` and `check=T`. We will use the first seven species names from the original vector, `tax`. Results will be stored in an object, named e.g. `tax2`:

```
> tax2 <- taxonomy(tax[1:7], aut=T, check=T); tax2

Sub/Infraorder          Family           Genus
2 Araneomorphae         Araneidae         Mangora
3 Araneomorphae       Clubionidae        Clubiona
4 Araneomorphae         Lycosidae         Pardosa
5 Araneomorphae Cheiracanthiidae Cheiracanthium
6 Araneomorphae       Theridiidae      Phylloneta
7 Araneomorphae        Zodariidae        Zodarion
1 Araneomorphae        Agelenidae       Allagelena
                   Species          Authors
2         Mangora acalypha (C. L. Koch, 1841)
3           Clubiona comta (Walckenaer, 1802)
4         Pardosa amentata   C. L. Koch, 1839
5 Cheiracanthium erraticum      (Clerck, 1757)
```

```
6        Phylloneta impressa (Walckenaer, 1802)
7          Zodarion rubidum    (L. Koch, 1881)
1     Allagelena gracilens       Simon, 1914
```

This output can be either marked and copied via clipboard into other software or exported to a file named e.g. table.txt on a disk (D) using the function `write.table`:

```
> write.table(tax2, file="d:\\table.txt", sep="\t")
```

Example 2

In this example we show how to extract information on species and genus richness, how to gather data on a species distribution, and how to map them, which is often required for a taxonomy paper, for example.

If we need to know how many species there are in the family Palpimanidae and in the genus *Ero*, for example, we can use the function `length` and apply it on the function `species` with specified names:

```
> length(species("Palpimanidae"))
[1] 158
> length(species("Ero"))
[1] 40
```

There are 158 species in the family Palpimanidae, and 40 species in the genus *Ero*.
To obtain verbal description on the distribution of six species from the `tax1` we just type:

```
> distribution(tax1[1:6, 2])
                 Species
   321        Clubiona comta
2        Pardosa amentata
3 Cheiracanthium erraticum
4      Phylloneta impressa
5          Zodarion rubidum
6      Allagelena gracilens

Distribution
   321                                                 Europe, North Africa, Turkey,
                                                       Caucasus
2           Europe, Turkey, Caucasus, Russia (Europe to South Siberia), Kazakhstan?, Iran?
3 Europe, Turkey, Caucasus, Russia (Europe to Far East), Iran, Central Asia, China, Korea,
Japan
4 North America, Europe, Turkey, Caucasus, Russia (Europe to Far East), Kazakhstan, Iran,
Central Asia, China, India
5                    Europe, Turkey, Russia (Europe, Caucasus). Introduced to USA, Canada
6                                                            Europe to Central Asia
```

The verbal distribution information of several species is very long, thus it is returned on separate lines. Moreover, to obtain records for *Cheiracanthium erraticum*, which is listed among species in `tax1` at the 3rd position in the second column, from WST and GBIF we call:

```
> records(tax1[3,2])
Retrieving data for Cheiracanthium erraticum
                  species      long      lat database      reference
1   Cheiracanthium erraticum  16.050972 58.82240    GBIF 10.15468/kllkyl
2   Cheiracanthium erraticum  12.612355 55.71835    GBIF 10.15468/ab3s5x
```

```
...
321 Cheiracanthium erraticum -27.1972 38.73206     WST Macías-Hernández et al. 2020
```

The list is very long, thus only the first two and the last row are displayed. Eventually, if we need to plot the species distribution, we use the function `map`. At first, we use a global world map of a species with a wide distribution, e.g. *Zodarion   ubidium*, with countries of occurrence shaded, without records plotted as points (`records=F`).

```
> map("Zodarion  ubidium", records = F)
```


Zodarion rubidum
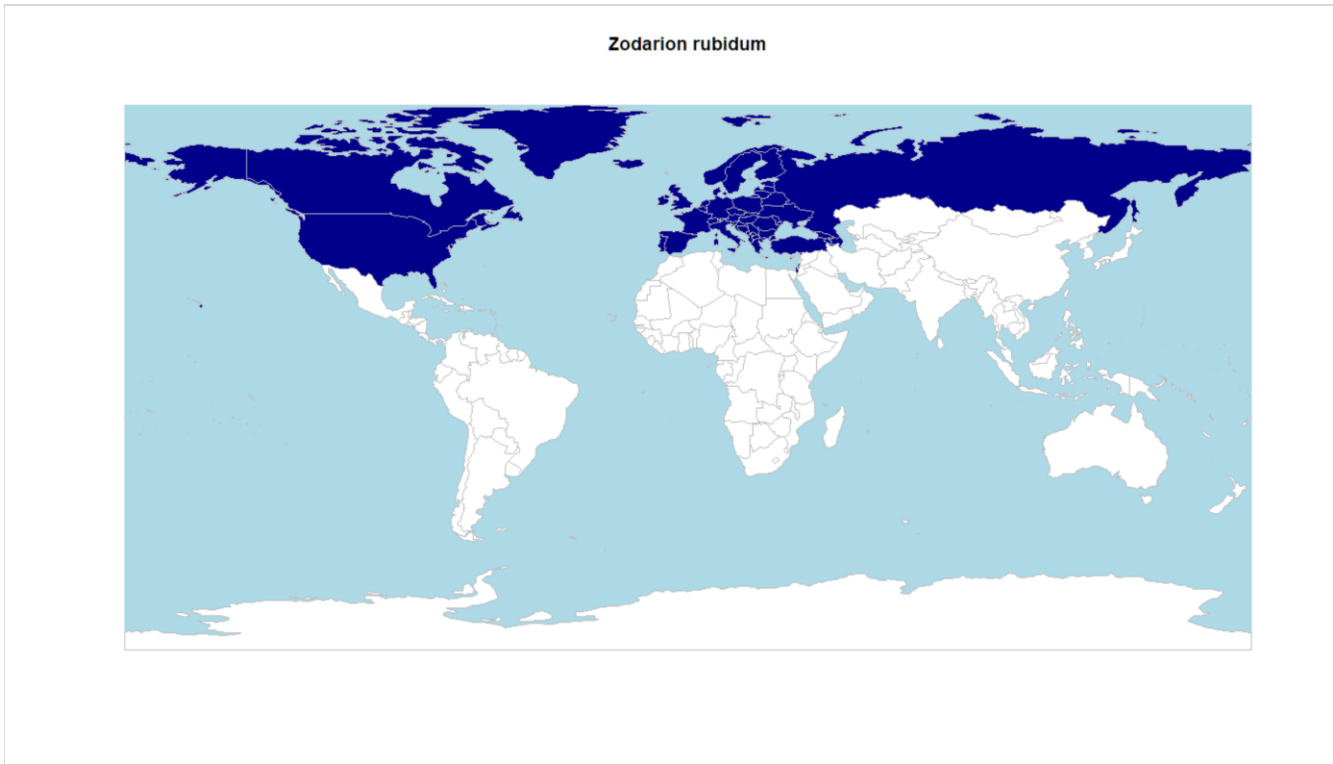
The user should be aware that the maps are only estimates of a species known distribution based on the WSC, with all the caveats associated with a rough textual description.

Secondly, we produce a map for a species with a small distribution, e.g., *Zodarion jozefienae*. For this purpose, we use the argument `zoom=T` and allow the actual records to be plotted as points. In order to get smooth country borders we use argument `hires=T`. Plotting might take longer.

```
> map("Zodarion jozefienae", hires = T, zoom = T)
Retrieving data for Zodarion jozefienae
```

Zodarion jozefienae

The function returned a map of the Iberian Peninsula with individual records.

Example 3

Here we show how to prepare data on traits obtained from the WST for further statistical analysis. For example, we want to compare body length (abbreviation for this trait is "bole") between males and females of *Zodarion jozefienae*. We download the trait data into an object named `tr1` by using the function `traits` with the species name as the first argument, and `trait = "bole"` as the second argument. Then we look at the content by calling `names`.

```
> tr1 <- traits("Zodarion jozefienae", trait = "bole")
> names(tr1)
 [1] "id"                   "originalName"         "value"
 [4] "frequency"            "sampleSize"           "treatment"
 [7] "altitude"             "locality"             "habitat"
[10] "microhabitat"         "note"                 "rowLink"
[13] "eventDate.text"       "eventDate.start"      "eventDate.end"
[16] "taxonomy.lsid"        "taxonomy.order"       "taxonomy.family"
[19] "taxonomy.genus"       "taxonomy.species"     "taxonomy.subspecies"
[22] "trait.id"             "trait.abbrev"         "trait.name"
[25] "trait.category.id"    "trait.category.name"  "measure.id"
[28] "measure.name"         "sex.id"               "sex.name"
[31] "lifeStage.id"         "lifeStage.name"       "method.id"
[34] "method.abbrev"        "method.name"          "location.id"
[37] "location.abbrev"      "location.coords.lat"  "location.coords.lon"
[40] "country.id"           "country.name"         "country.code"
[43] "dataset.id"           "dataset.name"         "dataset.authors"
[46] "dataset.restricted"   "reference.id"         "reference.abbrev"
```

The data frame `tr1` has 48 columns (The explanation of the columns can be found at https://github.com/oookoook/spider-trait-database/blob/master/docs/api.md). We are interested only in two of them, namely value (which contains the measurements) and sex.name (which contains the sex categories). For further analysis it is useful to extract those two variables from the `tr1` and it is essential to make the vector value numeric and the vector sex.name a factor. At last, we can produce a boxplot of the comparison.

```
> body <- as.numeric(tr1$value)
```

```
> sex <- factor(tr1$sex.name)
> plot(body~sex)
```
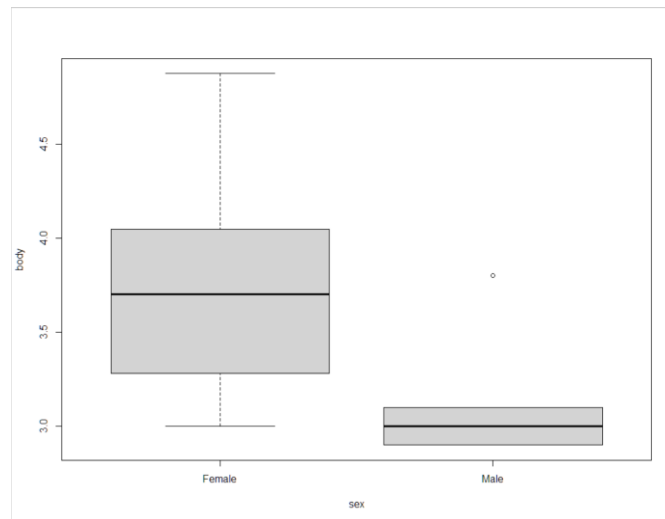

The boxplot shows larger body size in females than in males.

The two variables are ready to be used in statistical analyses, such as an ANOVA implemented in GLM (e.g., Pekár & Brabec 2016).

References

Pekár S. & Brabec M. 2016. Modern Analysis of Biological Data. Generalised Linear Models in R. Masaryk University Press, Brno.