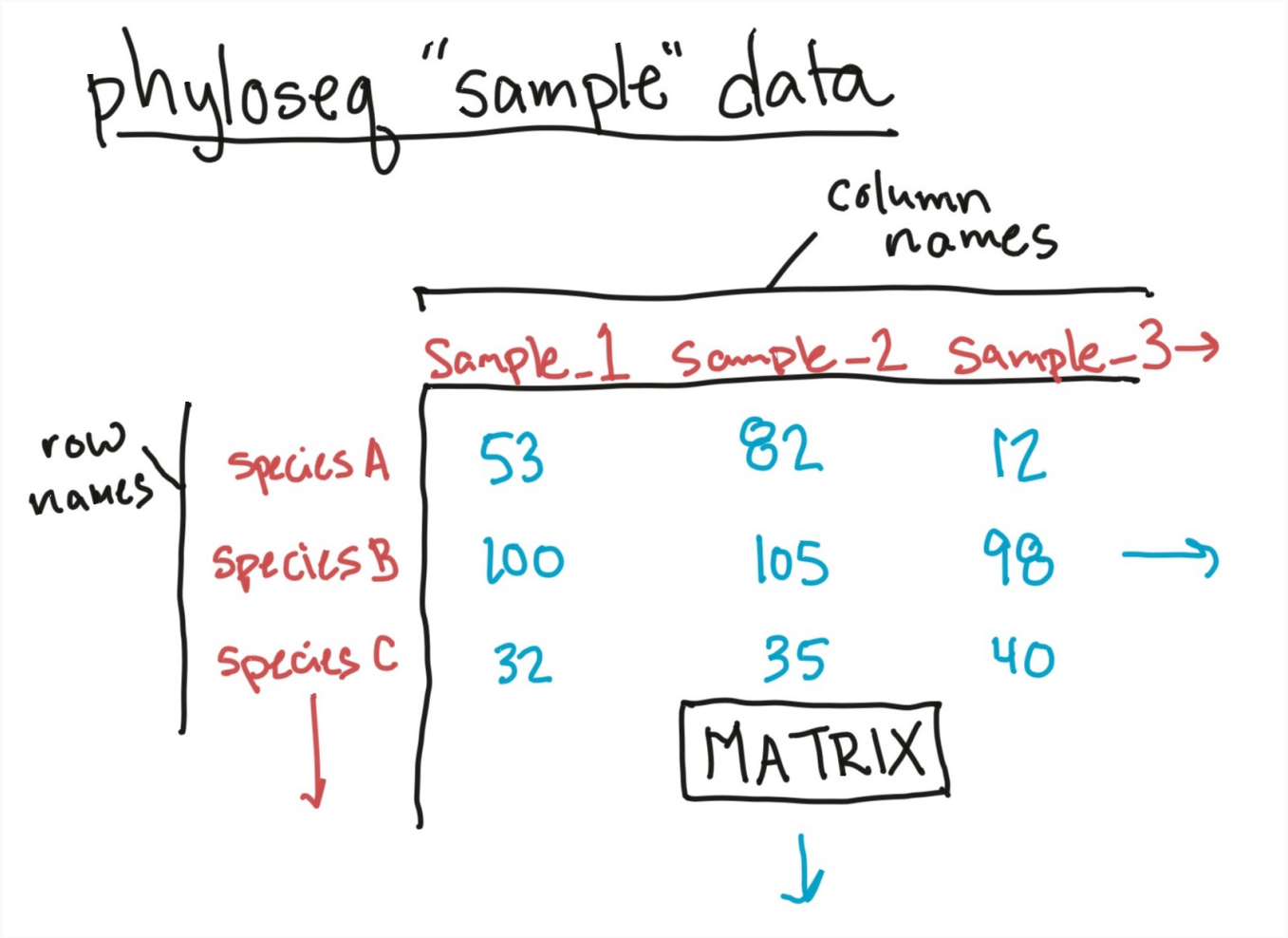
Exploring data 2

Example data

phyloseq sample data



The "phyloseq" object class has a "sample" slot, with a matrix with prevalence for each bacteria in each sample.

phyloseq sample data

You can use the `get_sample` accessor function to extract this data:

```
library("microbiome")
data("atlas1006")
atlas_sample_data <- atlas1006 %>%
  get_sample()
```

phyloseq sample data

You can check that this is a matrix, with column names giving sample number and rownames giving bacteria species:

```
atlas_sample_data %>% is.matrix()
```

```
## [1] TRUE
```

```
atlas_sample_data %>% colnames() %>% head(n = 3)
```

```
## [1] "Sample-1" "Sample-2" "Sample-3"
```

```
atlas_sample_data %>% row.names() %>% head()
```

```
## [1] "Actinomycetaceae"              "Aerococcus"
## [3] "Aeromonas"                     "Akkermansia"
## [5] "Alcaligenes faecalis et rel." "Allistipes et rel."
```

phyloseq sample data

You can use square bracket indexing to check the top left corner of the sample data:

```
atlas_sample_data[1:6, 1:3]
```

```
##                                 Sample-1 Sample-2 Sample-3
## Actinomycetaceae                       0        0        0
## Aerococcus                             0        0        0
## Aeromonas                              0        0        0
## Akkermansia                           21       36      475
## Alcaligenes faecalis et rel.           1        1        1
## Allistipes et rel.                    72      127       34
```

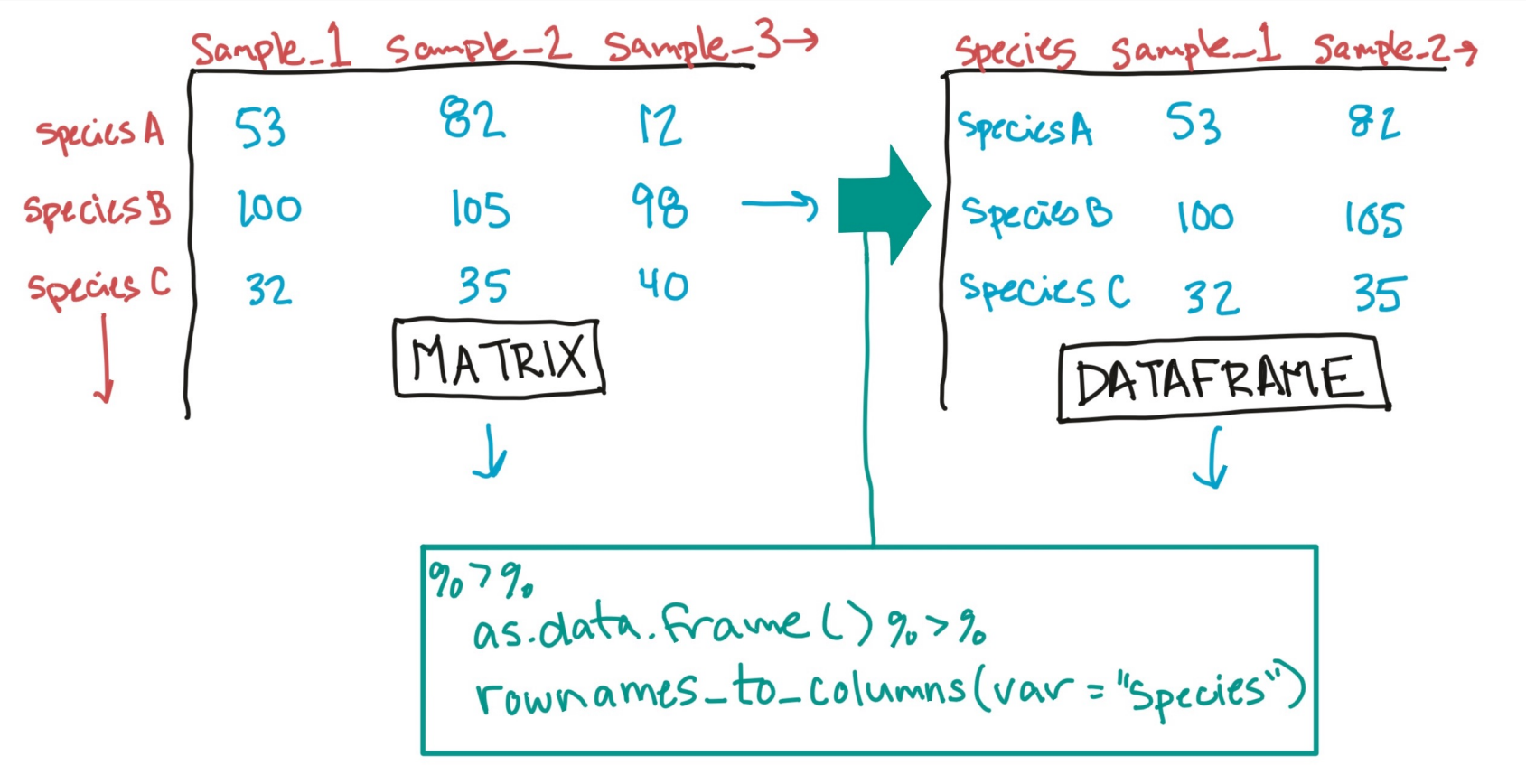
Tidying phyloseq sample data

To tidy this data, we need to:

1. Change to a data frame
2. Move row names into a column
3. Pivot longer so that column names are in their own column as values

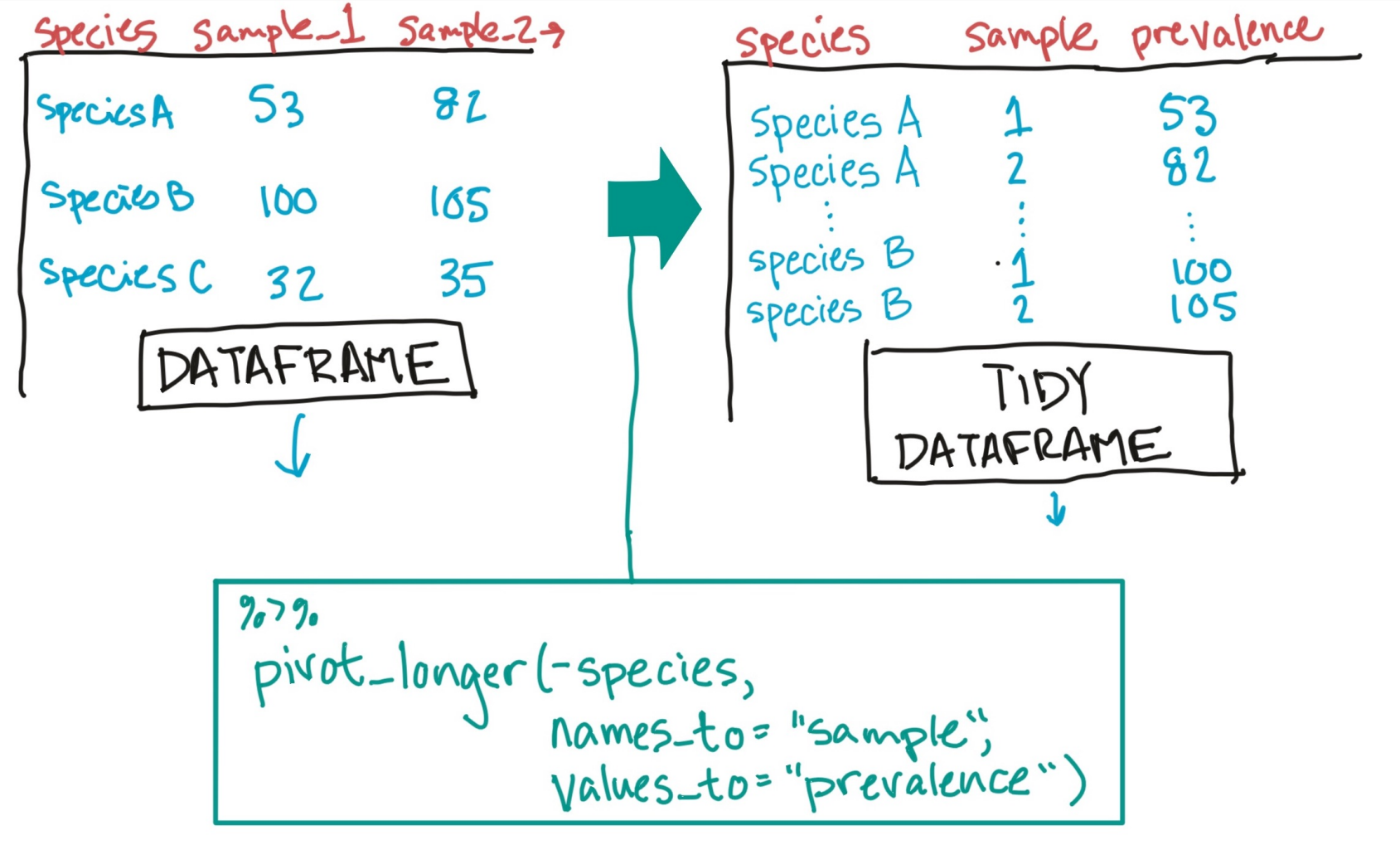
Tidying phyloseq sample data

Change to a data frame and move row names into a column:



Tidying phyloseq sample data

Pivot longer so that column names are in their own column as values:



Tidying phyloseq sample data

Here is everything in code:

```
library(tibble)
library(tidyr)

tidy_samples <- atlas1006 %>%
  get_sample() %>%
  as.data.frame() %>%
  rownames_to_column(var = "species") %>%
  pivot_longer(-species,
               names_to = "sample",
               values_to = "prevalence")
```

Tidying phyloseq sample data

Here's what the beginning of the tidy data looks like:

```
tidy_samples %>%
  slice(1:5)
```

```
## # A tibble: 5 x 3
##   species          sample   prevalence
##   <chr>            <chr>         <dbl>
## 1 Actinomycetaceae Sample-1          0
## 2 Actinomycetaceae Sample-2          0
## 3 Actinomycetaceae Sample-3          0
## 4 Actinomycetaceae Sample-4          0
## 5 Actinomycetaceae Sample-5          0
```