

# **Demo - use case 2: Querying data on species and sample level**

Moritz Mainka

2025-04-09

## **Introduction**

The goal of this script is to show how to query data from sample and species level at the same time. Further, we will have a look at how to aggregate, visualize, and store this query.

## Load plot-info table

The first step is to load the plot-level information table. This table is central as it functions as linking table between sample and species level. As a good practice, we start by assessing the table.

```
# load plot-level information table
# this table is central as it connects site data down to the sample or species level
plot_info <- read.csv("../data/plot_level_info.csv")

head(plot_info)
```

|   | plot_id   | lat      | long     | site | environment | core     | geology_code |
|---|-----------|----------|----------|------|-------------|----------|--------------|
| 1 | AD_T_BP01 | 78.17369 | 15.88830 | AD   | T           | core     | G24          |
| 2 | AD_T_BP02 | 78.17376 | 15.88909 | AD   | T           | core     | G24          |
| 3 | AD_T_BP03 | 78.17356 | 15.88917 | AD   | T           | core     | G24          |
| 4 | AD_T_BP04 | 78.17364 | 15.89029 | AD   | T           | core     | G24          |
| 5 | AD_T_BP05 | 78.17198 | 15.89659 | AD   | T           | gradient | G24          |
| 6 | AD_T_BP06 | 78.17214 | 15.89606 | AD   | T           | gradient | G24          |

|   | geology         | plot_type |
|---|-----------------|-----------|
| 1 | Shale+sandstone | BP        |
| 2 | Shale+sandstone | BP        |
| 3 | Shale+sandstone | BP        |
| 4 | Shale+sandstone | BP        |
| 5 | Shale+sandstone | BP        |
| 6 | Shale+sandstone | BP        |

```
# str(plot_info)
# summary(plot_info)
```

## Aggregate species-level data on plot-level

We will now aggregate species-level data on plot-level. For this we will use the SLA (specific leaf area) data for *Saxifraga cespitosa* and calculate the mean SLA. Then we will do the same for soil chemistry (pH and CEC) and combine both datasets.

```
# load data with plant cover per plant functional type
veg <- read.csv("../data/veg_plant_traits.csv")

# aggregate mean sla for Saxifraga cespitosa
```

```
spec_sla <- aggregate(sla ~ plot_id, data = veg[veg$species == "Saxifraga cespitosa", ], FUN = mean)

# merge plot-level information with plot-level data using the plot_id column
plot_data <- merge(plot_info, spec_sla, by = "plot_id", all.y = T)

# repeat process for sample information
soil <- read.csv("../data/soil_chem.csv")

# aggregate TC
soil_agg <- aggregate(cbind(pH_CaCl2, CECeff_BaCl2) ~ plot_id, data = soil, FUN = mean)

# merge plot-level information with plot-level data using the plot_id column
soil_veg <- merge(plot_data, soil_agg, by = "plot_id", all.x = T)
```

The resulting dataframe has 12 rows.

## Create plot of aggregated data

Let's create a simple scatterplot to visualize our query:

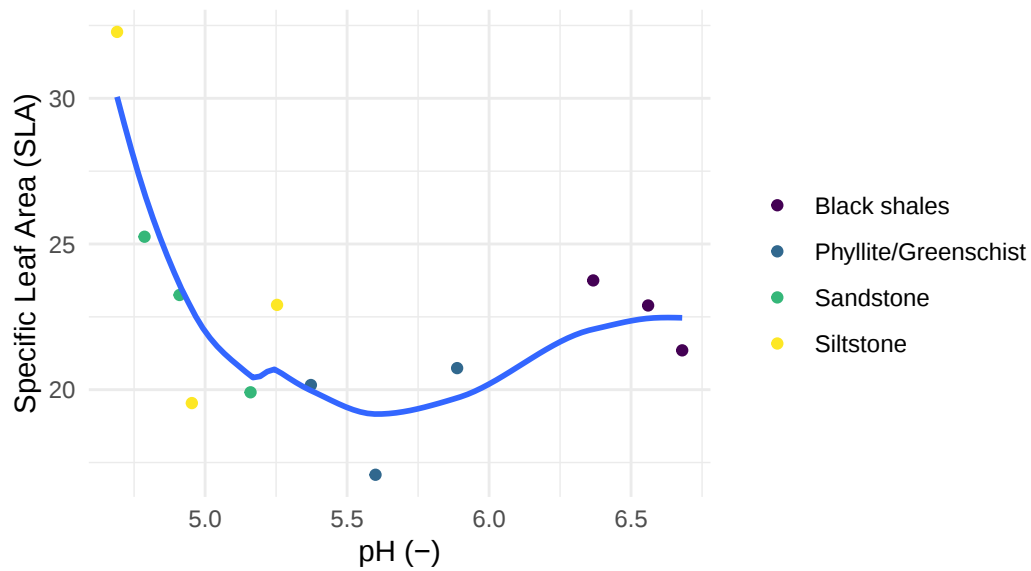
```
# plot of SLA with pH
library(ggplot2)
library(viridis)
```

Loading required package: viridisLite

```
ggplot(soil_veg) +
  geom_point(aes(x = pH_CaCl2, y = sla, color = as.factor(geology))) +
  geom_smooth(aes(x = pH_CaCl2, y = sla), method = "loess", se = FALSE) +
  labs(title = "Relation of specific leaf area (SLA) of Saxifraga cespitosa\nand soil pH",
       x = "pH (-)",
       y = "Specific Leaf Area (SLA)") +
  scale_color_manual(name = "", values = viridis(length(levels(as.factor(soil_veg$geology))))) +
  theme_minimal()
```

`geom\_smooth()` using formula = 'y ~ x'

Relation of specific leaf area (SLA) of *Saxifraga cespitosa* and soil pH



### Save table as csv

Finally, let's save our query as a csv file.

```
# save table as csv
write.csv(soil_veg, file = "../demo/query_demo2.csv", row.names = F)
```