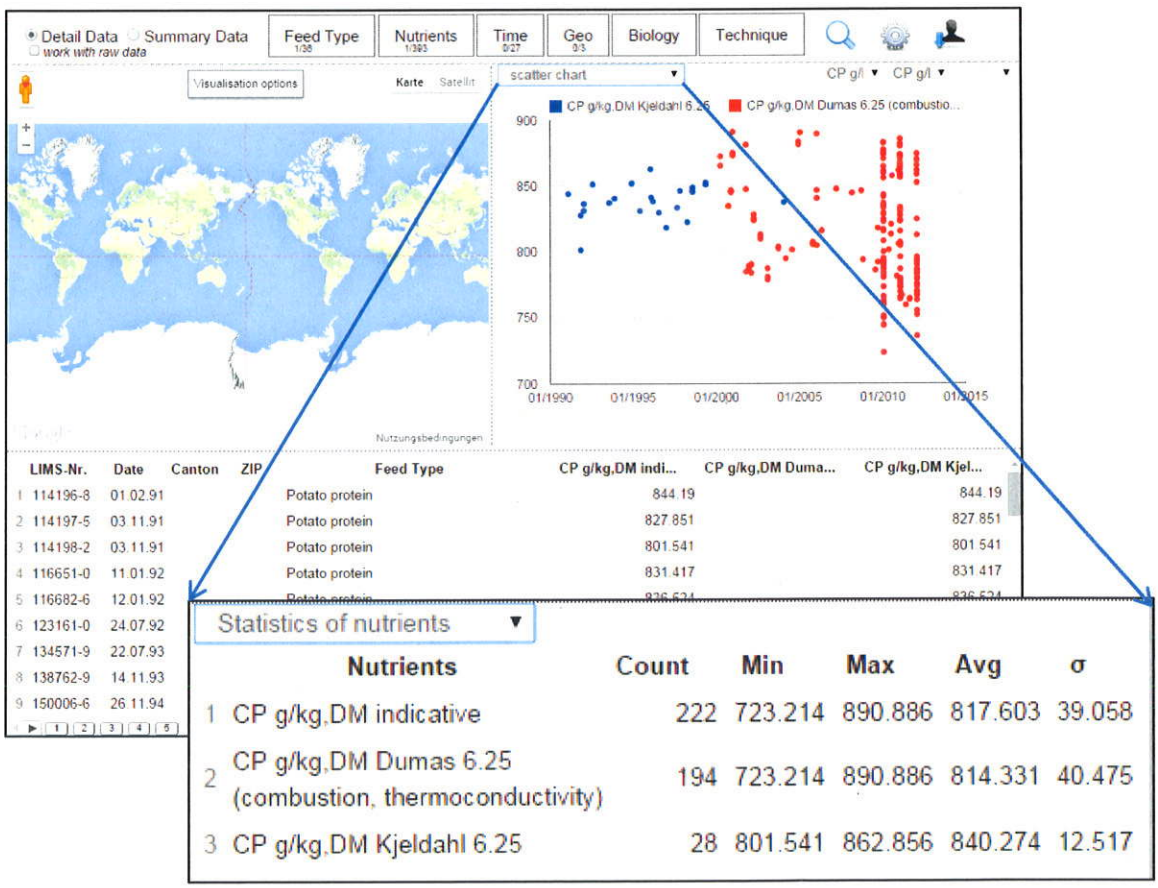
Topic: Implementing grouping factors in nutrient statistics

Problem description: The Swiss feed database contains data about nutrient contents of animal feed. Queries can be executed on summary data or detail data comprising records of individual feed samples. The result output of a query on the detail data level is divided into three views (figure 1, 2):

1. Google map locating the origin of individual feed samples and coloring the spatial nutrient density
2. Scatter chart and statistics of nutrients
3. List containig the selected nutrients of individual feed samples

The view *Statistics of nutrients* displays for each nutrient *count, min, max avg* and σ. In case of more than one analytical method for a defined nutrient, the statistics is given for each method separately. The overall mean corresponds to the *indicative* value. The current result output does for instance not give a statistical answer to the yearly evolution of count, min, max, avg and σ. Although the scatter plot gives a visual indication, the introduction of additional grouping factors in the statistical part would enrich the possibilities of data interpretation. In a simple case, grouping by year is most obvious (figure 1).

Figure 1. Simple case: one feed type, one nutrient and time selection



In more complex queries, feed type, year, canton and altitude are potential grouping factors (figure 2a, b) that make sense.

Figure 2a. Complex case: selection of several feed types, several nutrients, and time and geo

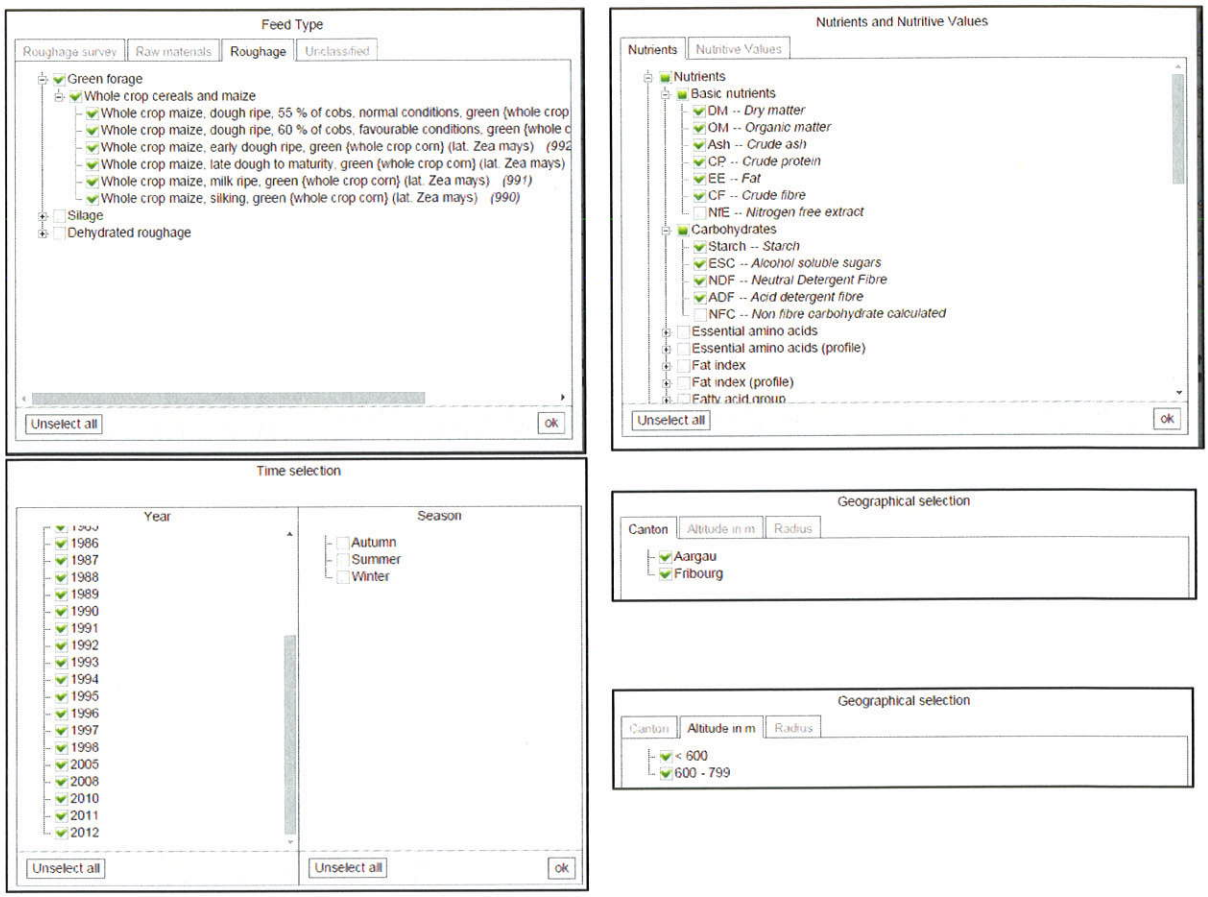
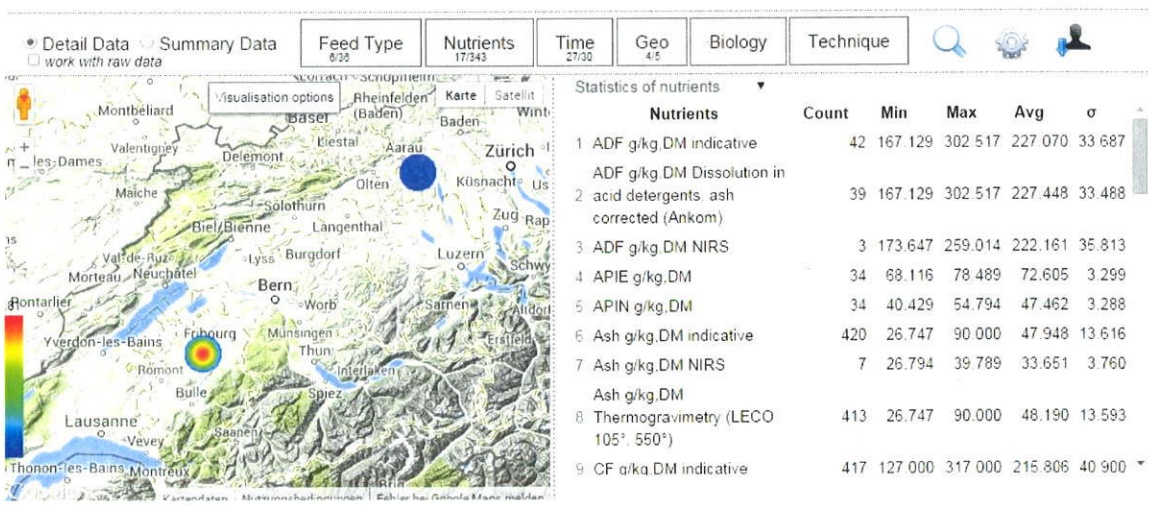


Figure 2b. Complex case: result output of nutrient statistics



The goal of the project is to implement *grouping factors* in the statistics of nutrient that contains the following elements:

- For the selected feed type(s), nutrient(s) and time/geo parameters, a user can specify grouping factors according to which the statistics of nutrients should be subdivided. In a simple case (one feed, one nutrient, time selection, no geo info), additional grouping by year is the most obvious procedure. In a generalized and more complex case, possible grouping candidates are: feed type, year, canton and altitude.
- Each additional grouping factor increases the complexity of the result output. This will limit the number of grouping factors that can be active at a time. A solution for a user friendly display is an important aspect of the implementation. In the case of more than one selected nutrient, most probably the grouping by analytical method must be dropped if additional grouping factors are wished which then can only be applied on the indicative value of the selected nutrients.
- By default, the result view is restricted to 150 samples which the user can manually modify under the *advanced option* button. The implemented functionality must be optimzed for fast response time. Particularly it should be tested whether the default setting must be maintained or not. The default setting could lead to misinterpretations if not the full sample number is considered. In case of a restricted result view, users should be informed by a comment line or window pop-up.