

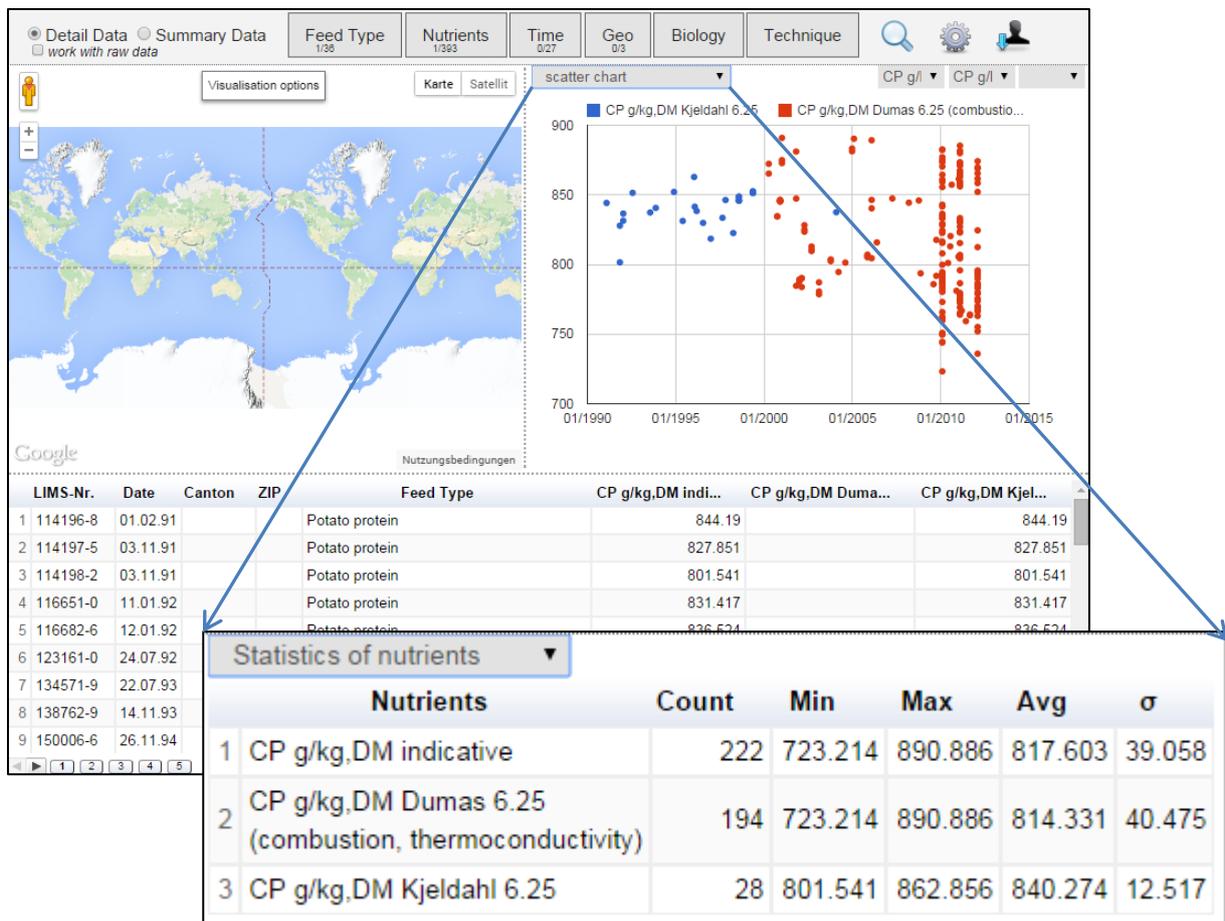
Topic: Implementing grouping factors in nutrient statistics

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 containing 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

The figure consists of four screenshots from a software application:

- Feed Type:** A tree view showing 'Green forage' selected, with sub-items like 'Whole crop cereals and maize' and 'Whole crop maize, dough ripe, 55 % of cobs, normal conditions, green (whole crop)'. Other options include 'Silage' and 'Dehydrated roughage'.
- Nutrients and Nutritive Values:** A tree view showing 'Basic nutrients' selected, with sub-items like 'DM -- Dry matter', 'OM -- Organic matter', 'Ash -- Crude ash', 'CP -- Crude protein', 'EE -- Fat', 'CF -- Crude fibre', and 'NFE -- Nitrogen free extract'. Other categories include 'Carbohydrates', 'Essential amino acids', 'Fat index', and 'Fatty acid group'.
- Time selection:** A list of years from 1986 to 2012, with '1992' selected. A 'Season' dropdown menu is set to 'Summer'.
- Geographical selection:** Two panels. The first shows 'Canton' with 'Aargau' and 'Fribourg' selected. The second shows 'Altitude in m' with '< 600' and '600 - 799' selected.

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

The screenshot shows a software interface with a map of Switzerland on the left and a table of nutrient statistics on the right. The table is titled 'Statistics of nutrients' and has columns for 'Nutrients', 'Count', 'Min', 'Max', 'Avg', and σ .

Nutrients	Count	Min	Max	Avg	σ
1 ADF g/kg,DM indicative	42	167.129	302.517	227.070	33.687
2 ADF g/kg,DM Dissolution in acid detergents, ash corrected (Ankom)	39	167.129	302.517	227.448	33.488
3 ADF g/kg,DM NIRS	3	173.647	259.014	222.161	35.813
4 APIE g/kg,DM	34	68.116	78.489	72.605	3.299
5 APIN g/kg,DM	34	40.429	54.794	47.462	3.288
6 Ash g/kg,DM indicative	420	26.747	90.000	47.948	13.616
7 Ash g/kg,DM NIRS	7	26.794	39.789	33.651	3.760
8 Thermogravimetry (LECO 105°, 550°)	413	26.747	90.000	48.190	13.593
9 CF a/kg,DM indicative	417	127.000	317.000	215.806	40.900

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 optimized 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.