BIOL495\_lab2\_ exercises Thomas Sanchez

* Exercise 1:

Assume you are collecting insects at three different locations, three times a year, over the span of three years. For each day/location of sampling, you record:

§Site data: site of the sampling, its geographic coordinates, description of the site

§Sampling data: day, month, and year, weather conditions, temperature, humidity, etc.;

§Species data: classification according to species and stage of development of all the insects collected, their total, and their measurements

Write out the plan of your design of all the tables you are making, and how you are going to relate them together.

|  |  |
| --- | --- |
| Site | location |
| A | x,y |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Site | Month | year | weather | temp | humidity | etc. |
| A | Jan | 2022 |  |  |  |  |
| A | Jan | 2023 |  |  |  |  |
| A | Jan | 2024 |  |  |  |  |
| A | Apr | 2022 |  |  |  |  |
| A | Apr | 2023 |  |  |  |  |
| A | Apr | 2024 |  |  |  |  |
| Etc. |  |  |  |  |  |  |
| Site | Month | year | weather | temp | humidity | etc. |
| B | Jan | 2022 |  |  |  |  |
| B | Jan | 2023 |  |  |  |  |
| B | Jan | 2024 |  |  |  |  |
| B | Apr | 2022 |  |  |  |  |
| B | Apr | 2023 |  |  |  |  |
| B | Apr | 2024 |  |  |  |  |
| Etc. |  |  |  |  |  |  |
| Site | Month | year | weather | temp | humidity | etc. |
| A | Jan | 2022 |  |  |  |  |
| B | Jan | 2022 |  |  |  |  |
| C | Jan | 2022 |  |  |  |  |
| A | Apr | 2022 |  |  |  |  |
| B | Apr | 2022 |  |  |  |  |
| C | Apr | 2022 |  |  |  |  |
| Etc. |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Site | Month | year | Classification/development | total count | measurements | description |
| A | Jan | 2022 |  |  |  |  |
| B | Jan | etc. |  |  |  |  |
| Etc. |  |  |  |  |  |  |

I would display the date in at least 3 sets of tables. Firstly is just the coordinates of each site, then simply reference the site from then on. The second set would firstly compare the same site over years, and a second set would contrast the different sites during the same timeframe. The third table would be a similar to the last and contrast different sites over the same month and year. If there was little to no change and I needed to show their average differences over time I would only then use the first layout from the second set of tables (same site, same month, dif. Year). Ultimately the exact layout depends on what is considered the most valuable piece of derived data.

The tables are mostly related by simply the site location, but also by the formatting of month and year depending on desired set as stated above.

* Exercise 2:

1. get the first two lines of all the files ending with .csv within *data/* directory, and save them to a single file called “headers.csv” under *sandbox/*

blimko@RANGERIII MINGW64 ~/Documents/BIOL495/DS4Bio\_labs/lab1+2/lab1\_unix\_DSB/data (master)

$ head -n 2 \*.csv > ../sandbox/headers.csv