Witteveen + Bos

Sample Scientist

Bob Brederveld
Luke Moth

Introduction

- system analyses: datasets
- often performing same analyses again
- until now: often re-script or make same graphs in excel

More efficient using tools: *Sample Scientist*

Platform for performing standard analytics

EcoConverter – Dataprofeet – Sample Scientist

Explanation tool

- works on all “n samples vs m observables”

dummy data

some_index	some_location	some_timestamp	some_variable1	some_variable2	some_
0	my_location1	2021-03-21	5.0	99.9	0.000
1	my_location2	2021-04-21	-5.0	44.9	0.000

- supports (for now) the following analyses:
  - PCA
  - hierarchical clustering
  - linear regression
  - stepwise regression
  - correlation matrix
  - boxplots (including ANOVA test when subgroups)
  - Stepwise with PCs
  - IR EGV plot
- available as python package or as stand-alone executable

Explanation tool - input

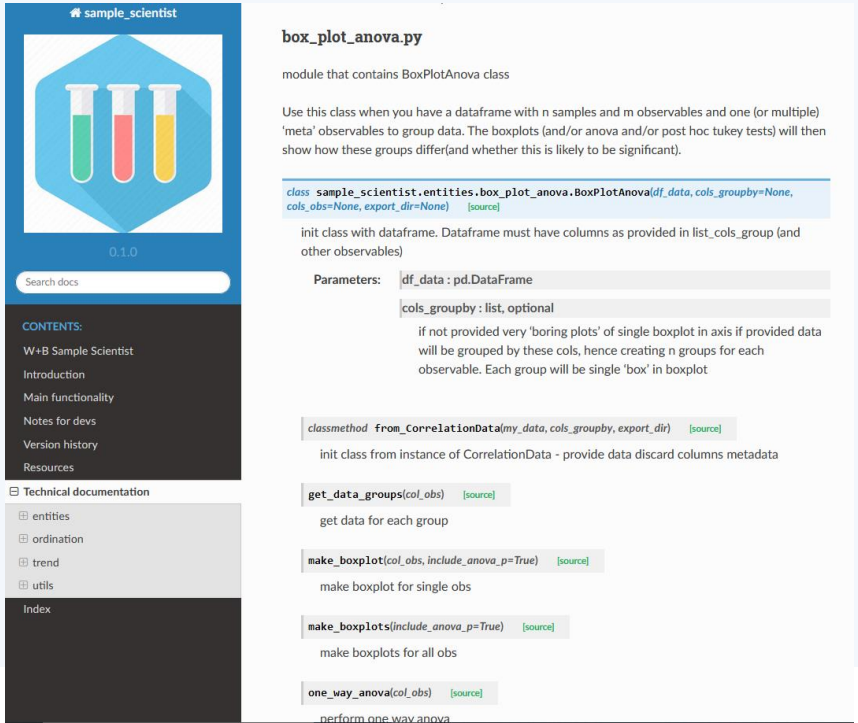
- When using the exe version:

  a *_config.yaml

example1_config.yaml | example_6_all_config.yaml

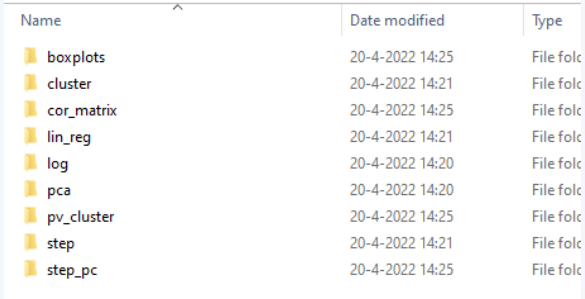
```yaml
---
data_input:
  # REQUIRED
  data_file: "D:/repos/trendanalyse/tests/data/demo_data_bluecan.csv"
  data_file_type: "csv"     # may be csv, xlsx or xlsx-multitab
  # optional, defaults which user can update
  csv_seperator: ";"
  decimal_seperator: ","
  # IF metadata cols and/or respons cols in output, please specify
  # otherwise these will be interpreted like numeric (and set to nan)
  cols:
    cols_metadata:
      - "datum"
      - "locatie"
      - "type_weer"
    cols_respons:
      - "total ghg"
    col_datetime: "datum"
    cols_color_def: ["locatie", "type_weer"]
    cols_ignore: ["data_mist"]
  # OPTIONAL (delete if need be)
  output_dir: "D:/other_work/tests_eco_analyser/testexe6"
pca: #delete entire PCA block if no PCA is required
  # OPTIONAL
  include_varimax: true
  varimax_min_axes: 2
  varimax_max_axes: 8
  biplot_max_axes: 10
linear_regression:  #delete entire linear_regression block if no linear_re
  # OPTIONAL
```

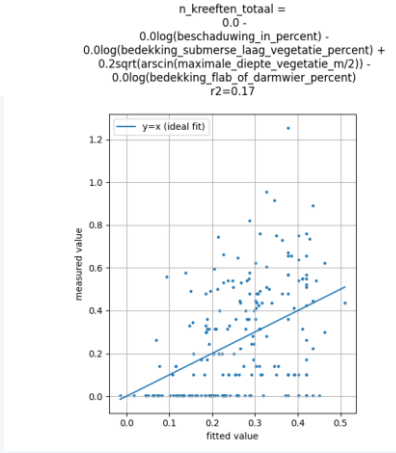
Explanation tool – use in python

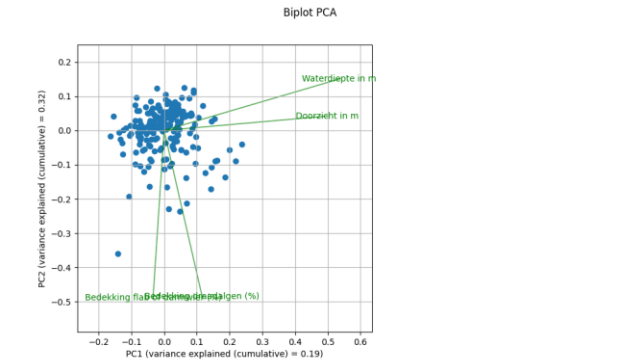


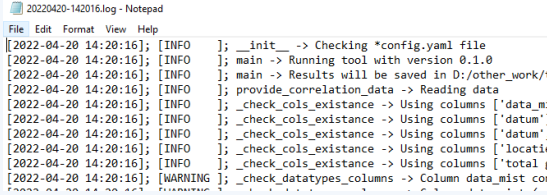
Explanation tool - output

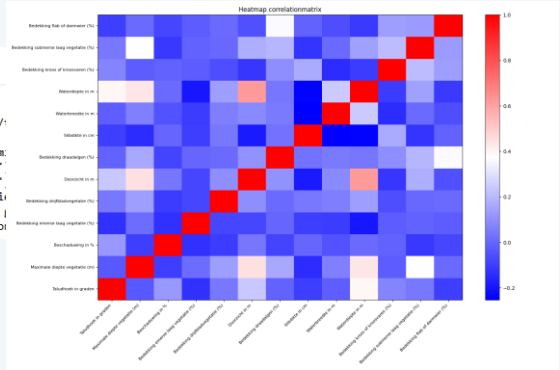
- Folder structure
- Plots (color coded)
- Excel files
- logging












www.witteveenbos.com