BONARES
Centre for Soil Research

DQ-Kit web app: Evaluating and Improving Data Quality for Soil and Agricultural Data in the BonaRes Repository

Susanne Lachmuth[1], Carsten Hoffmann[1], Viet Hoang Nguyen[1], Igo Silva de Almeida[1], Torsten Thalheim[2], Bethia Kynn[2], Stephan Lesch[3], Sebastian Rick[3], Thomas Kühnert[1], Xenia Specka[1], Nikolai Svoboda[1]

[1]Leibniz Center for Agricultural Research (ZALF), susanne.lachmuth@zalf.de
[2]Deutsches Biomasseforschungszentrum (DBFZ), [3]Senckenberg Museum of Natural History (SGN)

DQ-Kit v.1 — Data Quality & Plausibility Tool

Home | Demonstration | FAQ | About

Attention data authors and re-users: Discover DQ-Kit to enhance your data quality assurance effort!

Simply upload your csv table and let DQ-Kit provide comprehensive statistical insight into your scientific data. But that's not all! DQ-Kit takes it a step further by alerting you to any anomalies that could potentially affect the accuracy and reliability of your data.

DQ-Kit aids in refining a dataset before publication, instilling confidence, and extending impact for data authors. Additionally, it empowers data re-users to efficiently compare various datasets and select the ones that best align with their objectives.
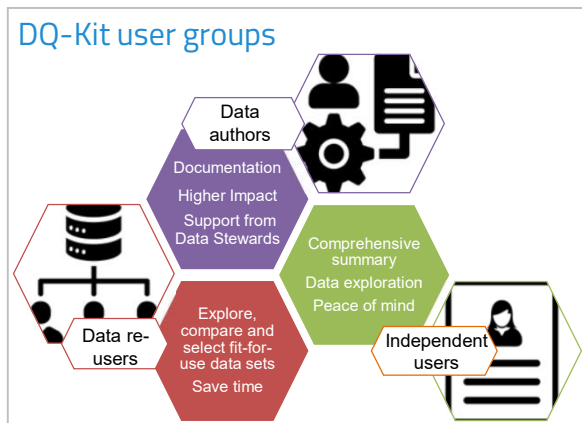
Read more

Drag and Drop or **click here to browse**

Analyse now

A collaboration of: BONARES, DBFZ

dqkit.bonares.de

BONARES REPOSITORY

BonaRes Repository: Download or publish soil and agricultural research data

DQ-Kit user groups

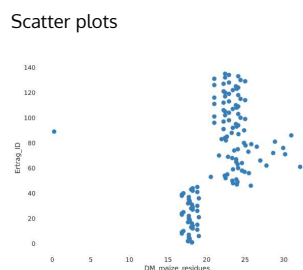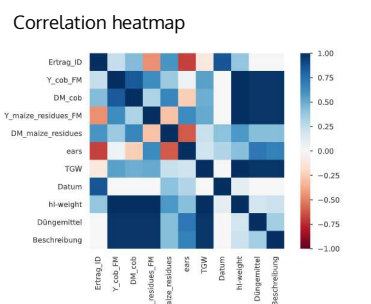Data authors — Documentation, Higher Impact, Support from Data Stewards

Data re-users — Explore, compare and select fit-for-use data sets, Save time

Comprehensive summary, Data exploration, Peace of mind

Independent users

Summary Statistics

Dataset statistics

Number of variables	11
Number of observations	137
Missing cells	92
Missing cells (%)	6.1%
Duplicate rows	2
Duplicate rows (%)	1.5%
Total size in memory	11.9 KiB
Average record size in memory	88.9 B

Variable types

Numeric	7
Categorical	4

DM_maize_residues — Real number (ℝ)

Distinct	65	Minimum	0.25
Distinct (%)	48.1%	Maximum	32.1
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Skewness	-1.01	Memory size	1.2 KiB
Mean	21.7		

hl-weight — Categorical

HIGH CORRELATION MISSING UNIFORM

Distinct	33
Distinct (%)	68.8%
Missing	89
Missing (%)	65.0%
Memory size	1.2 KiB

68,40000...	3
67,30000...	3
66	3
68,59999...	3
66,80000...	3
Other val...	33
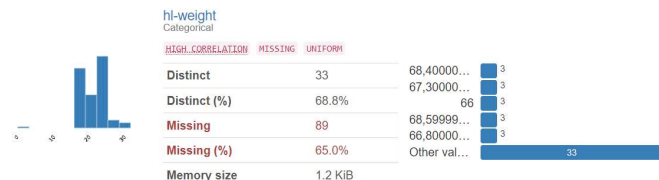
Relationships between Variables

Correlation heatmap

Scatter plots

Quality Alerts

Y_cob_FM is highly overall correlated with hl-weight and 2 other fields	High correlation
DM_cob is highly overall correlated with hl-weight and 2 other fields	High correlation
Y_maize_residues_FM is highly overall correlated with hl-weight and 2 other fields	High correlation
dry-weight is highly overall correlated with hl-weight and 2 other fields	High correlation
hl-weight is highly overall correlated with Y_cob_FM and 3 other fields	High correlation
Fertilizer is highly overall correlated with Y_cob_FM and 3 other fields	High correlation
Comment is highly overall correlated with Y_cob_FM and 3 other fields	High correlation

hl-weight has 89 (65.9%) missing values	Missing
Fertilizer has 2 (1.5%) missing values	Missing
Y_cob_FM is highly skewed (γ1 = 11.60771622)	Skewed
DM_cob is highly skewed (γ1 = 11.61544909)	Skewed
Y_maize_residues_FM is highly skewed (γ1 = 11.61842108)	Skewed
DM_maize_residues is highly skewed (γ1 = -1.008110506)	Skewed
ears is highly skewed (γ1 = 9.914844922)	Skewed
dry-weight is highly skewed (γ1 = 10.86911262)	Skewed
Yield_ID is uniformly distributed	Uniform
hl-weight is uniformly distributed	Uniform
Yield_ID has unique values	Unique

DQ-Kit Open-Source Initiative

Everybody is invited to contribute to DQ-Kit ...

Save the Date: Workshop 24-25 Sept. 2024 at ZALF

DQ-Kit v.1 — Data Quality & Plausibility Tool
dqkit.bonares.de

Your idea! · Metadata export · Time series analysis · Geo-spatial data · Extended graphics · Plausibility checks · Open Source

UFZ HELMHOLTZ Zentrum für Umweltforschung · zalf · SENCKENBERG world of biodiversity · BGR Bundesanstalt für Geowissenschaften und Rohstoffe · TUM Technische Universität München

www.bonares.de

Referenzen: DQ-Kit report - https://docs.profiling.ydata.ai/latest/, Icons - https://thenounproject.com/, QR codes - https://me-qr.com/