

Data Science Style Guide

Lukas E. G. Fischer

Methods Center and Institute of Sports Science • University of Tübingen

lukas.fischer@uni-tuebingen.de

What	How	Example
Variable names in dataframes (columns)	CamelCase	InterestRate
Dataframe defined within the code	Lowercase letters and underscore using nouns	interest_rate
Variables defined at the beginning of the code	- Capital letters and underscore or - Single capital letter (when implementing formulas)	- INTEREST_RATE - T
Variables defined within the code	Lowercase letters and underscore using nouns	interest_rate
Functions	- Identifier and lowercase letters with underscore using verbs or - Lowercase letters with underscore using verbs if the function is used for several projects	- et_calculate_interest_rate - calculate_interest_rate
Function inputs	- Lowercase letters and underscore using nouns or - Single lowercase letter (when implementing formulas)	- interest_rate - t
Objects (Lists, Dictionaries, ...)	Identifier and lowercase letters with specifier	et_interest_rate_list
Folders	- Capital letters and underscore with headline rule (small "and", "with", ...) or - CamelCase	- Economics_and_Finance - JabRef
Documents	- Lowercase letters with underscore. The general structure should be identifier_content_specifier or - content_specifier for not necessarily project-specific files (data, references, certain functions, ...)	- et_data_shaping_1 - anscombe_1

Of course, the variable, function, or file names can be abbreviated in a sensible manner if the names become too long otherwise (e.g., IntRate, int_rate, et_calc_int_rate, ...).