

# Exploratory Data Analysis (EDA)

Statsomat.com

24 June 2021

## Basic Information

Automatic statistics for the file:

File
Finance.csv

Your selection for the encoding: UTF-8

Your selection for the decimal character: ,

Observations (rows with at least one non-missing value): 1662

Variables (columns with at least one non-missing value): 3

Variables considered continuous: 3

Variables considered continuous
return
size
volatility

Variables considered categorical: 0

## Results for Numerical Variables

### Descriptive Statistics

Variables are sorted alphabetically. Missings are omitted in stats. CV only for positive variables.

	N Obs	N Missing	N Valid	% Complete	N Unique	Mean	SD	Median	MAD	Min	Max	Skewness	Kurtosis	CV
return	1662	0	1662	100	1662	22.18	6.25	22.82	6.38	0.31	36.14	-0.55	0.01	0.28
size	1662	0	1662	100	57	6.62	0.37	6.62	0.43	6.21	8.01	0.73	-0.2	0.06
volatility	1662	0	1662	100	1662	0.27	0.28	0.18	0.2	0	1.32	1.59	2.31	

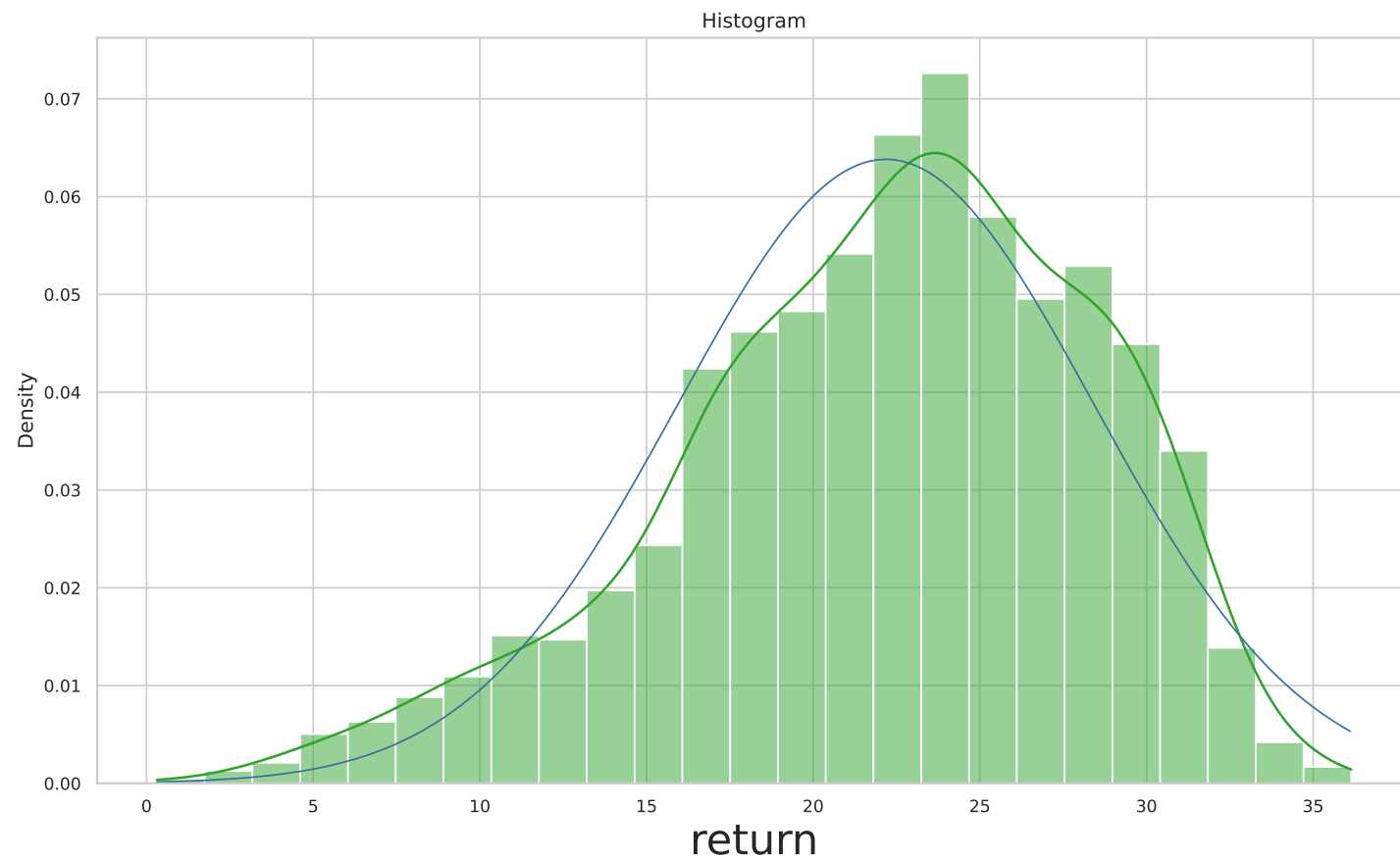
## Graphics

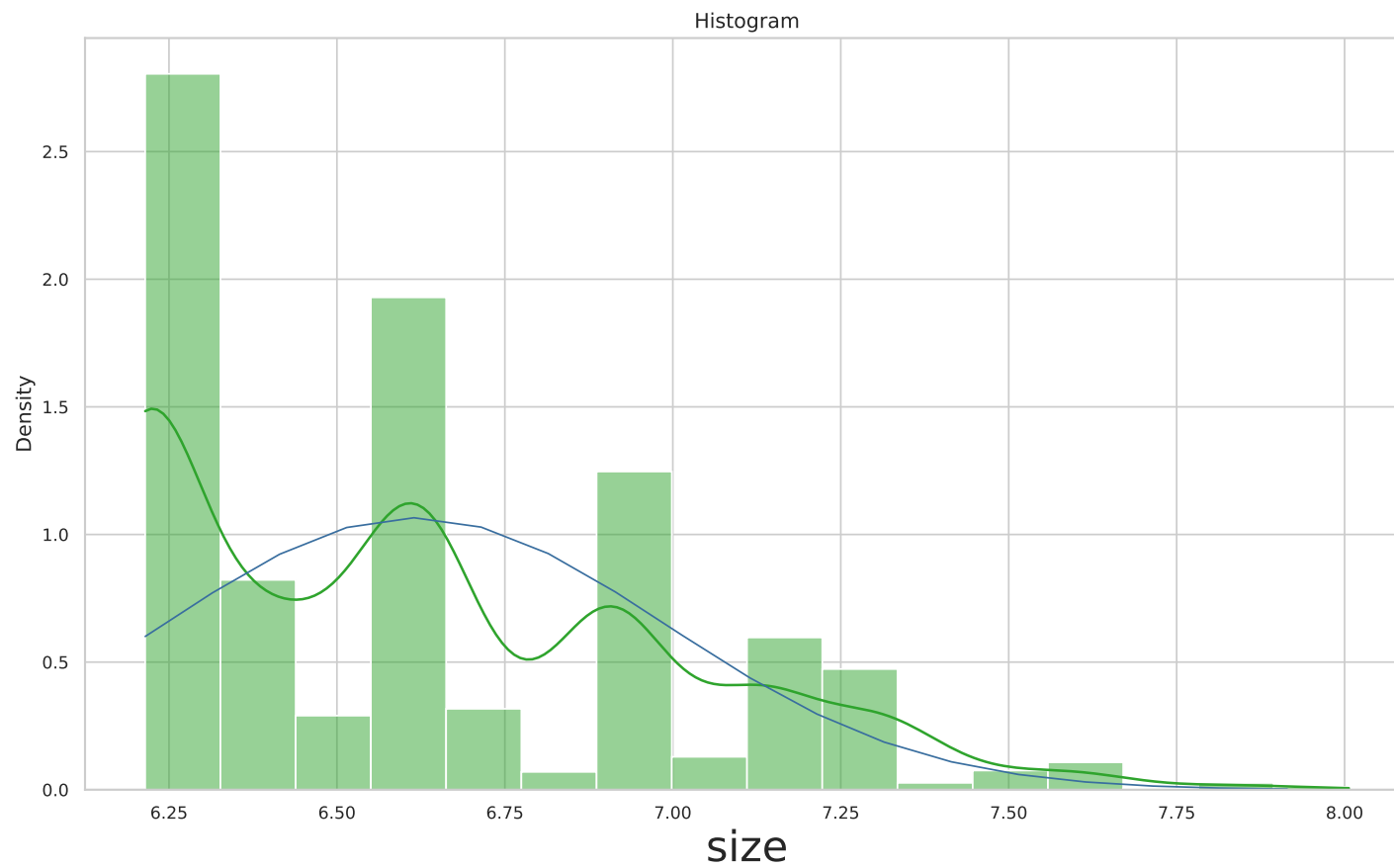
### Histograms

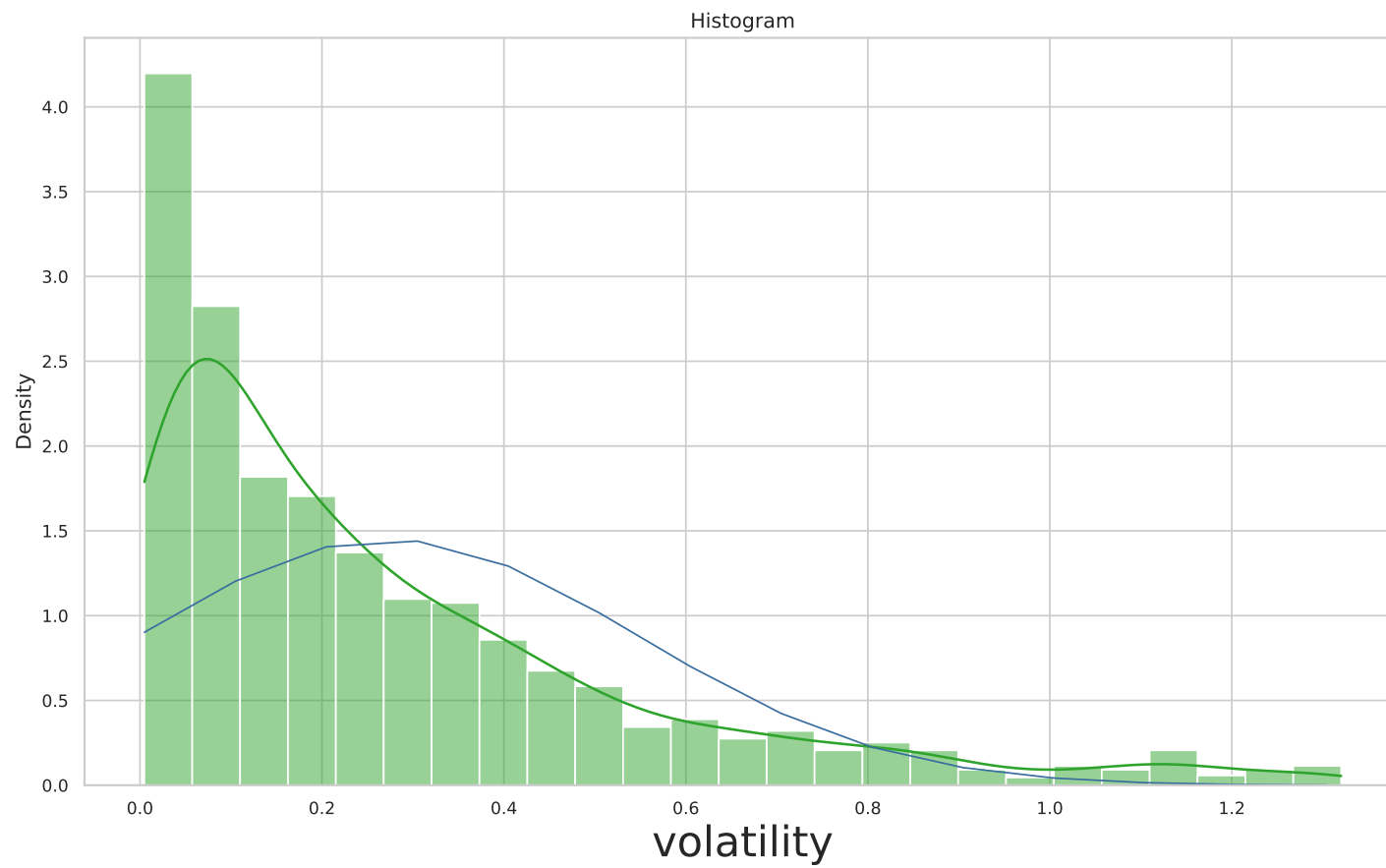
Details: Density Histograms. One large figure per page for each variable, sorted alphabetically. The blue line represents the normal density approximation. The green line represents a special kernel density approximation.

See figures on next page.

□

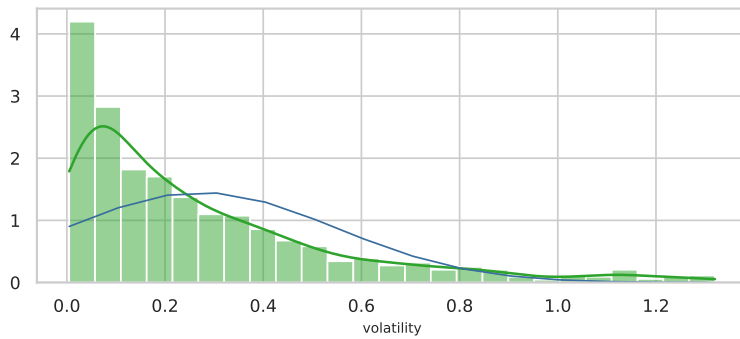
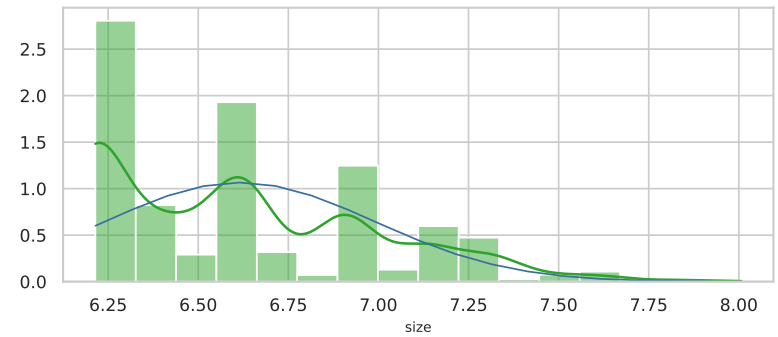
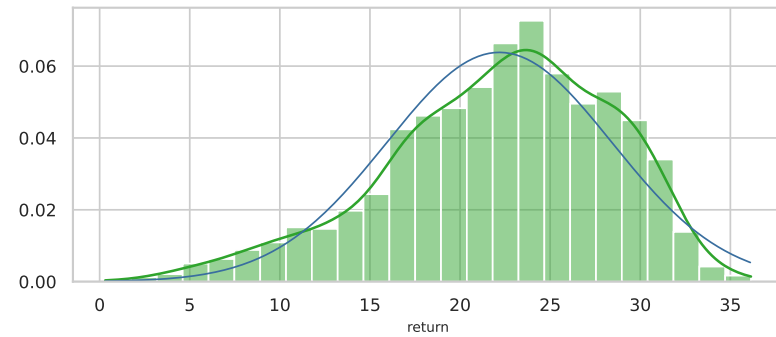






## Histograms Summary

Multiple Relative Frequency Histogram in one figure. Variables are sorted alphabetically. The blue line represents the normal density approximation. The green line represents a special kernel density approximation.

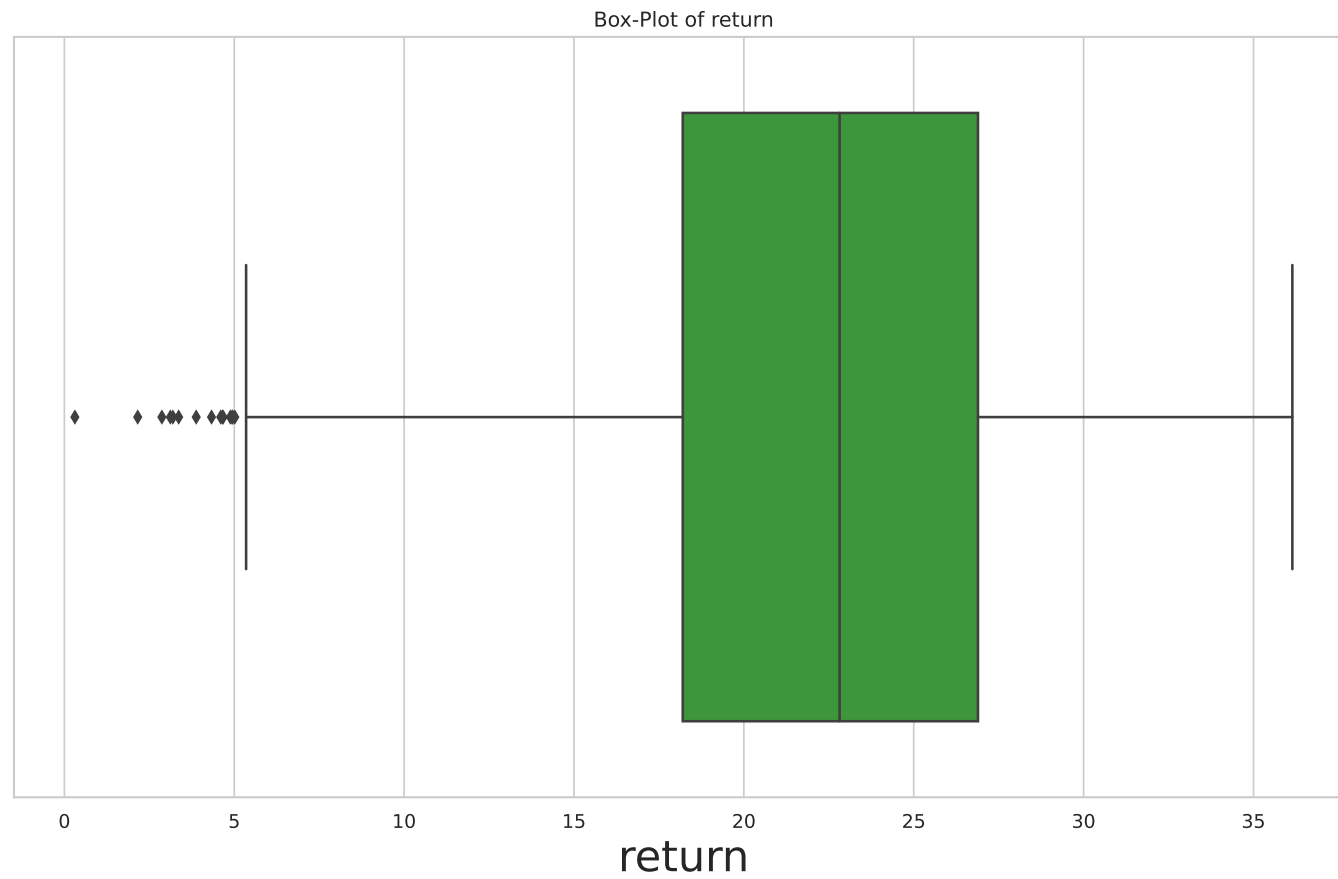


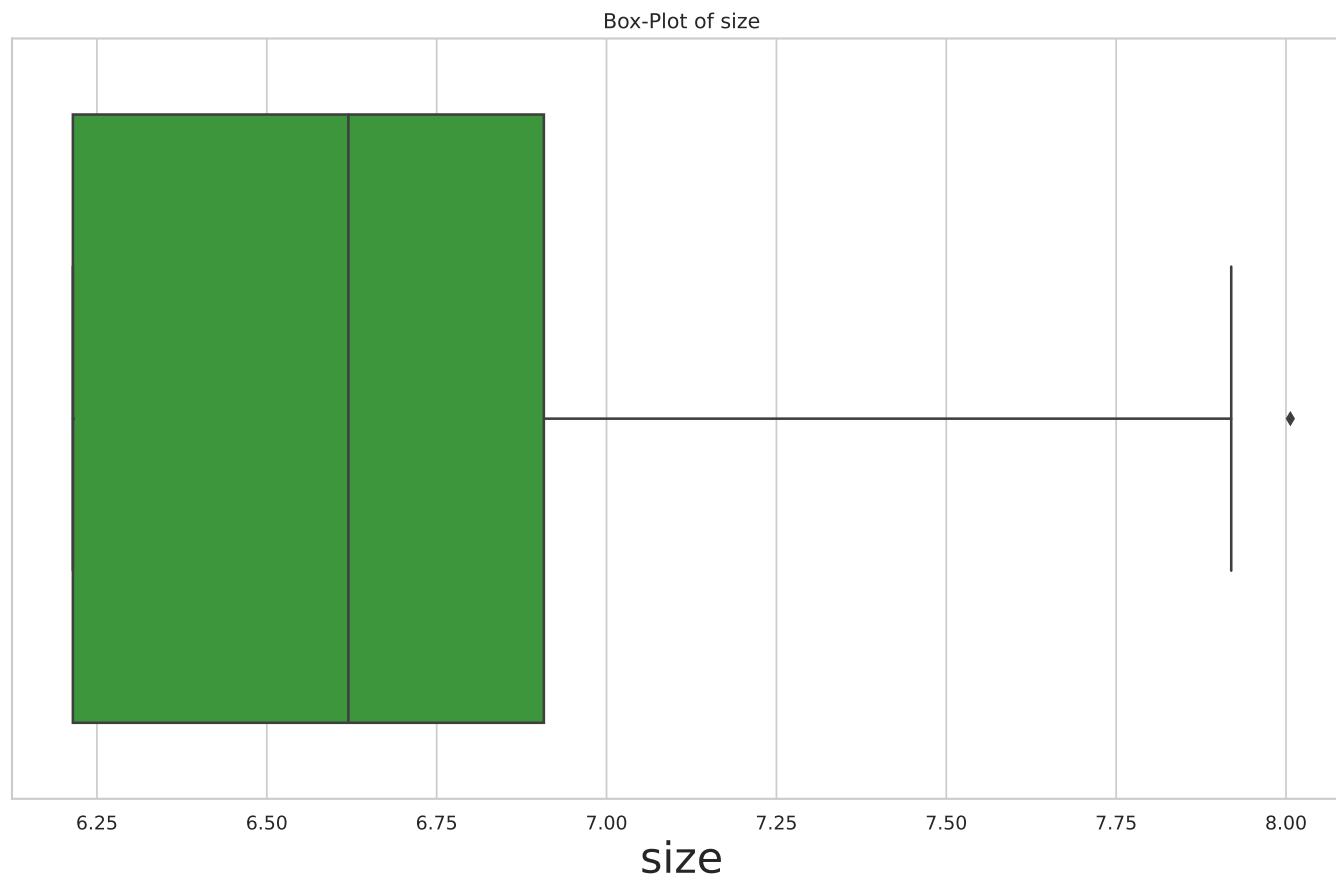


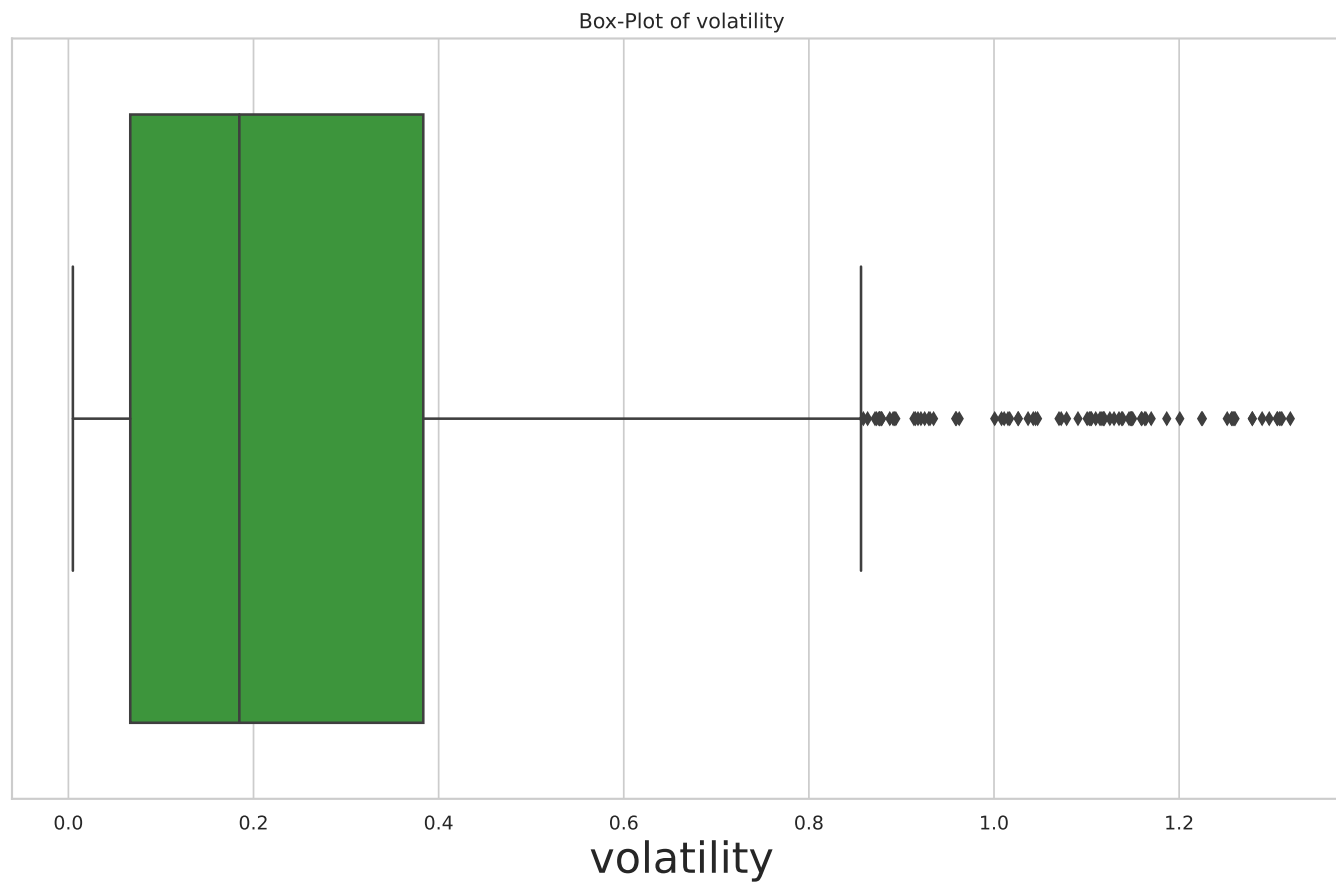
## Box-Plots

One Box-Plot per page for each variable. Variables are sorted alphabetically.

□

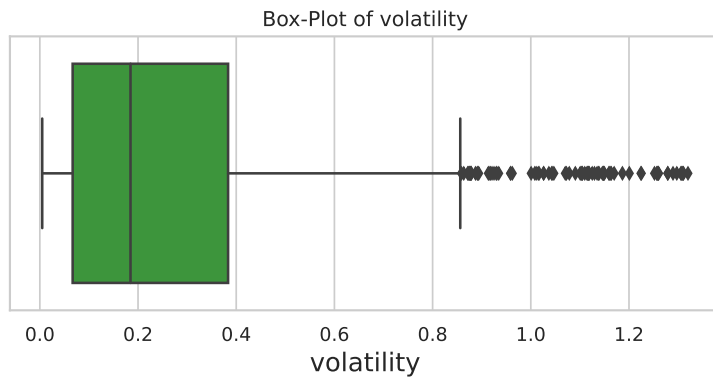
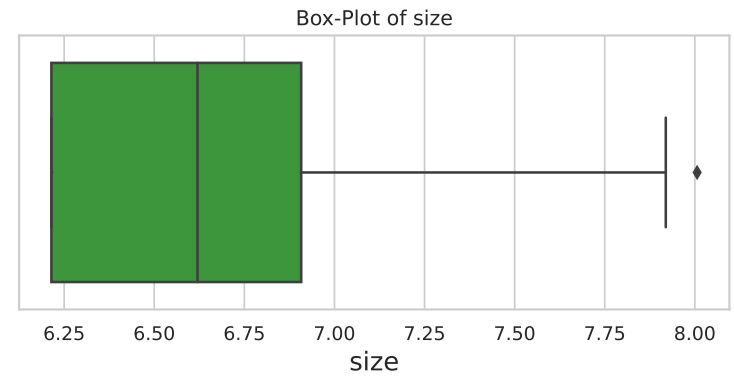
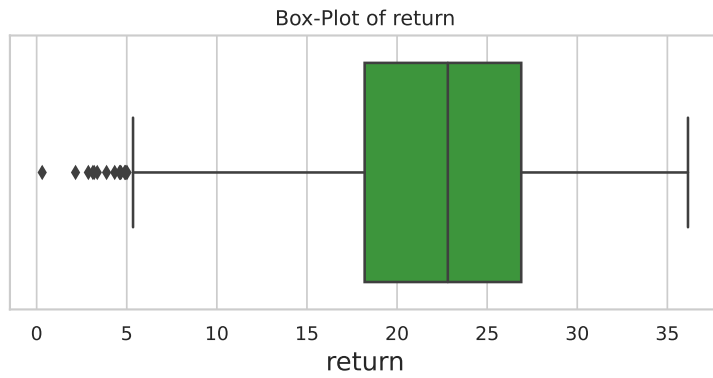






## Box-Plots Summary

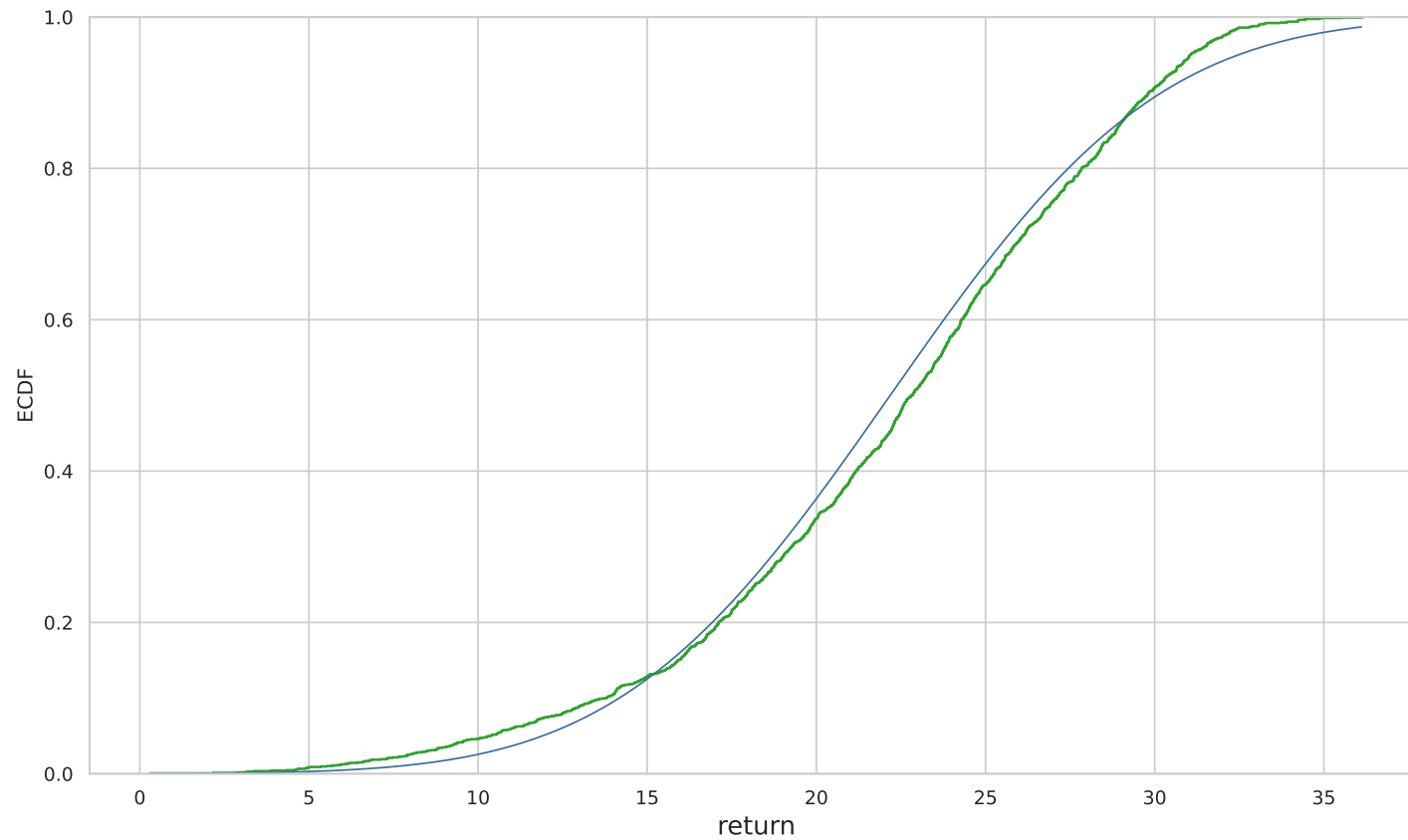
Multiple Box-Plots of variables in one figure. Variables are sorted alphabetically.

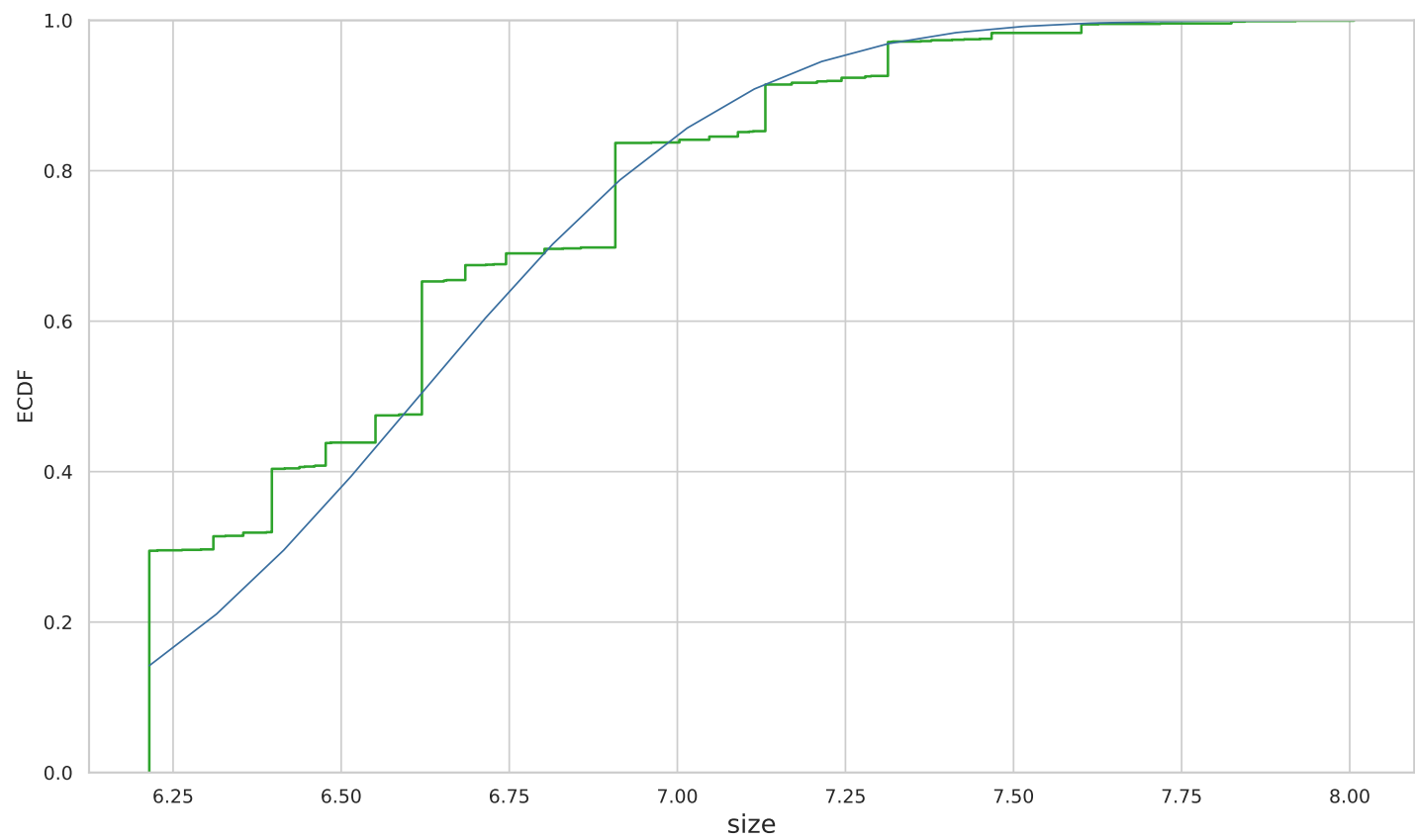


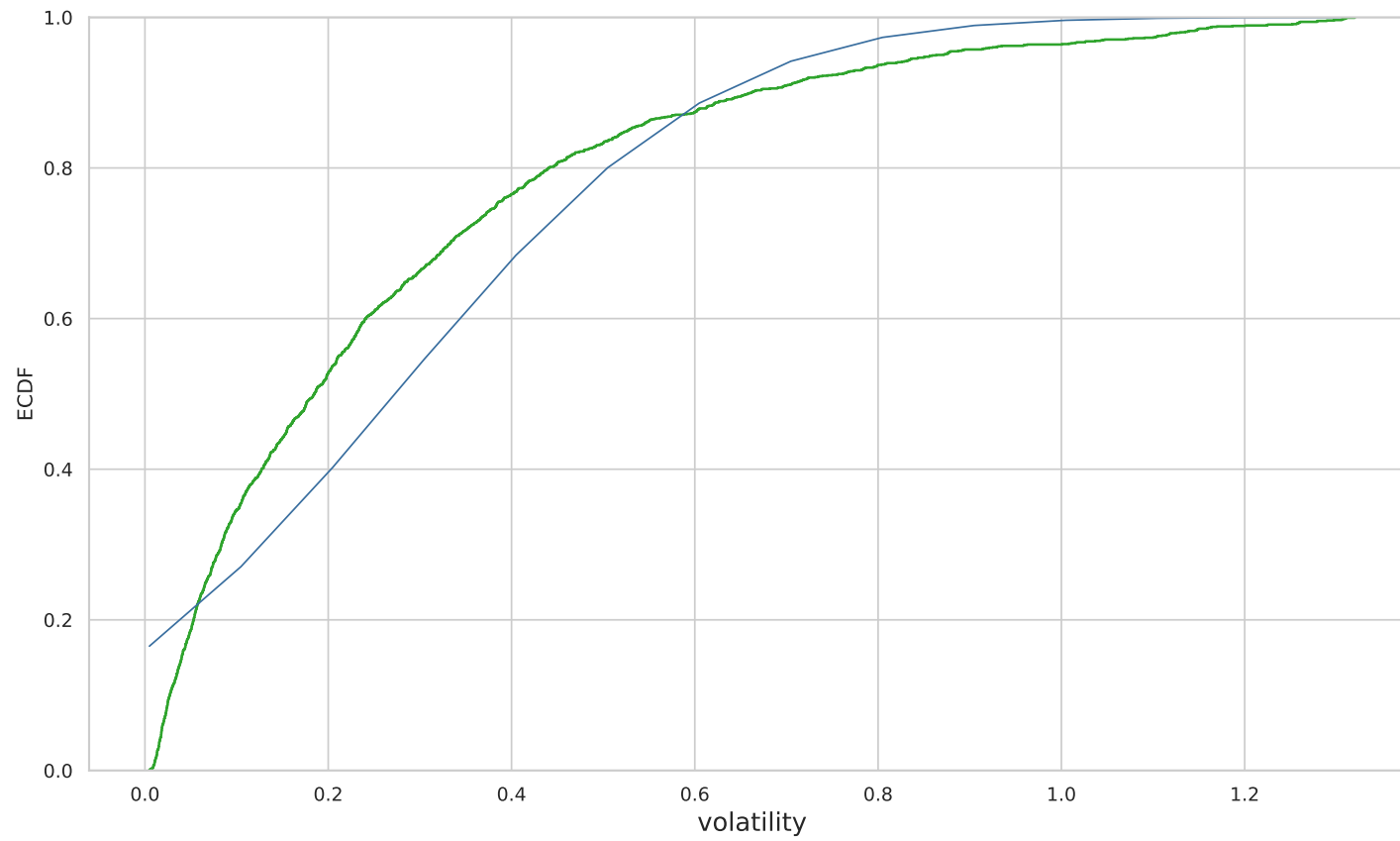
## ECDF Plots

One ECDF (Empirical Cumulative Distribution Function) Plot per page for each variable. Variables are sorted alphabetically. The blue line represents the CDF of a normal distribution. If the variable is normally distributed, the blue line approximates well the ECDF.

□

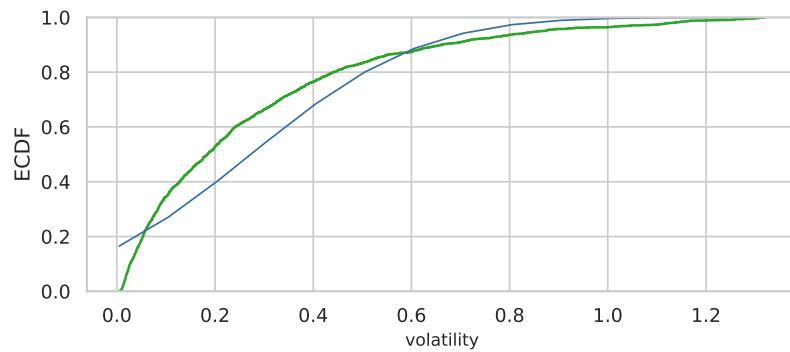
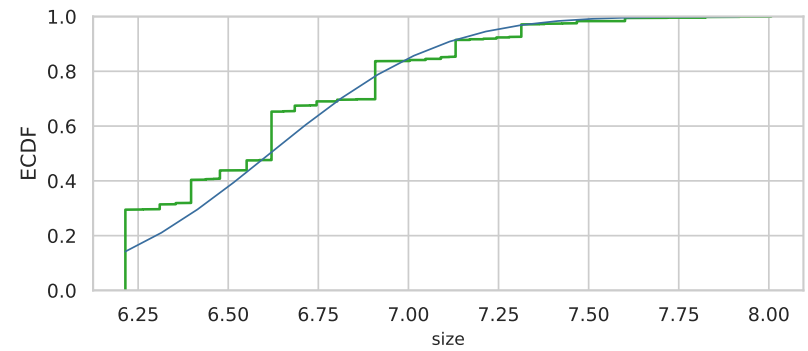
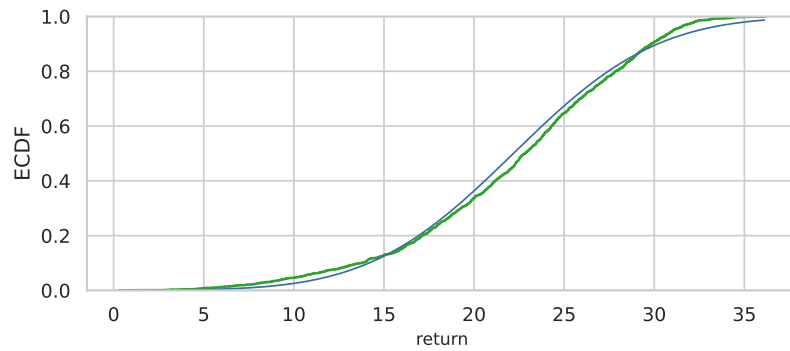






## ECDF Plots Summary

Multiple ECDF Plots of variables in one figure. Variables are sorted alphabetically.

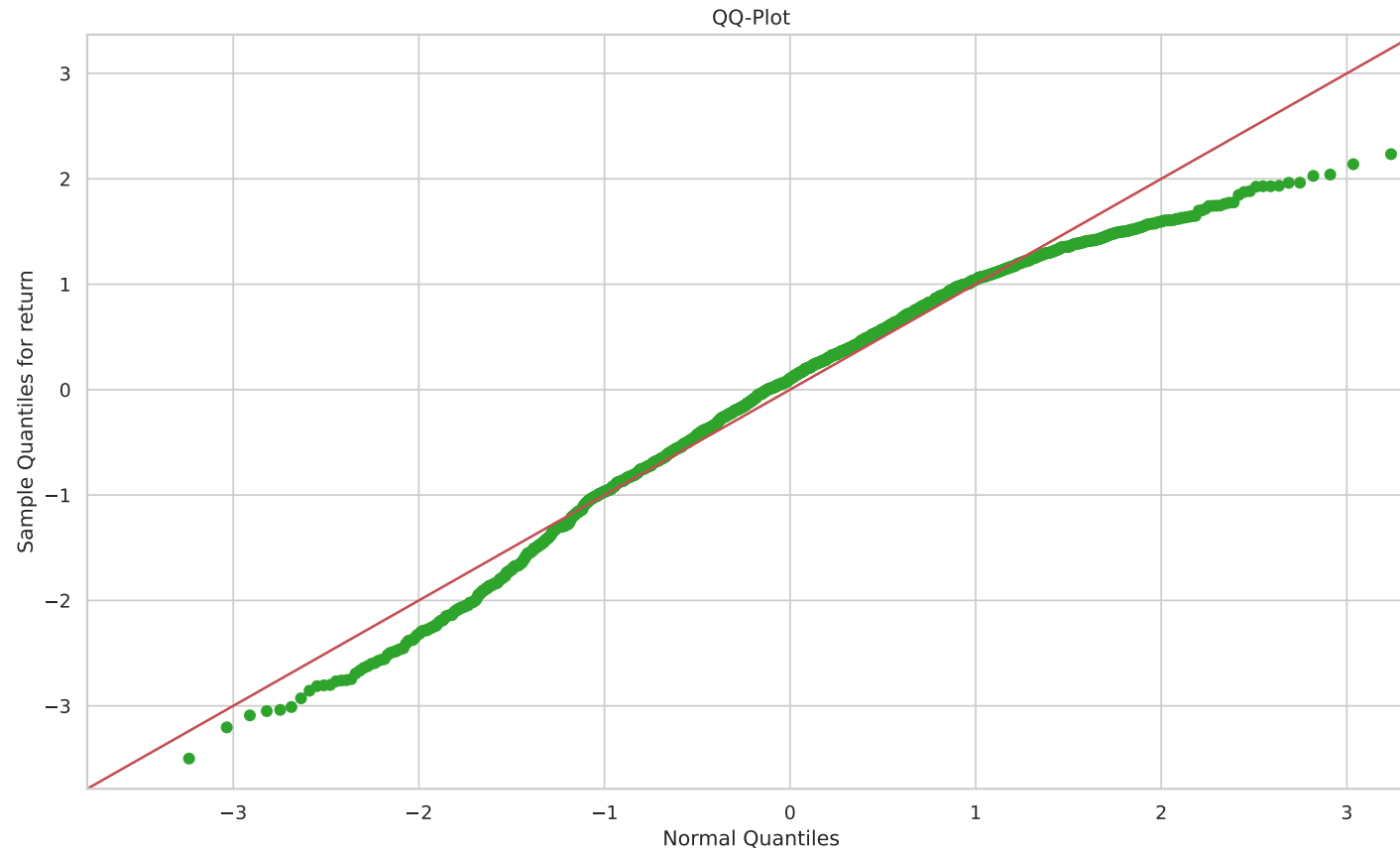


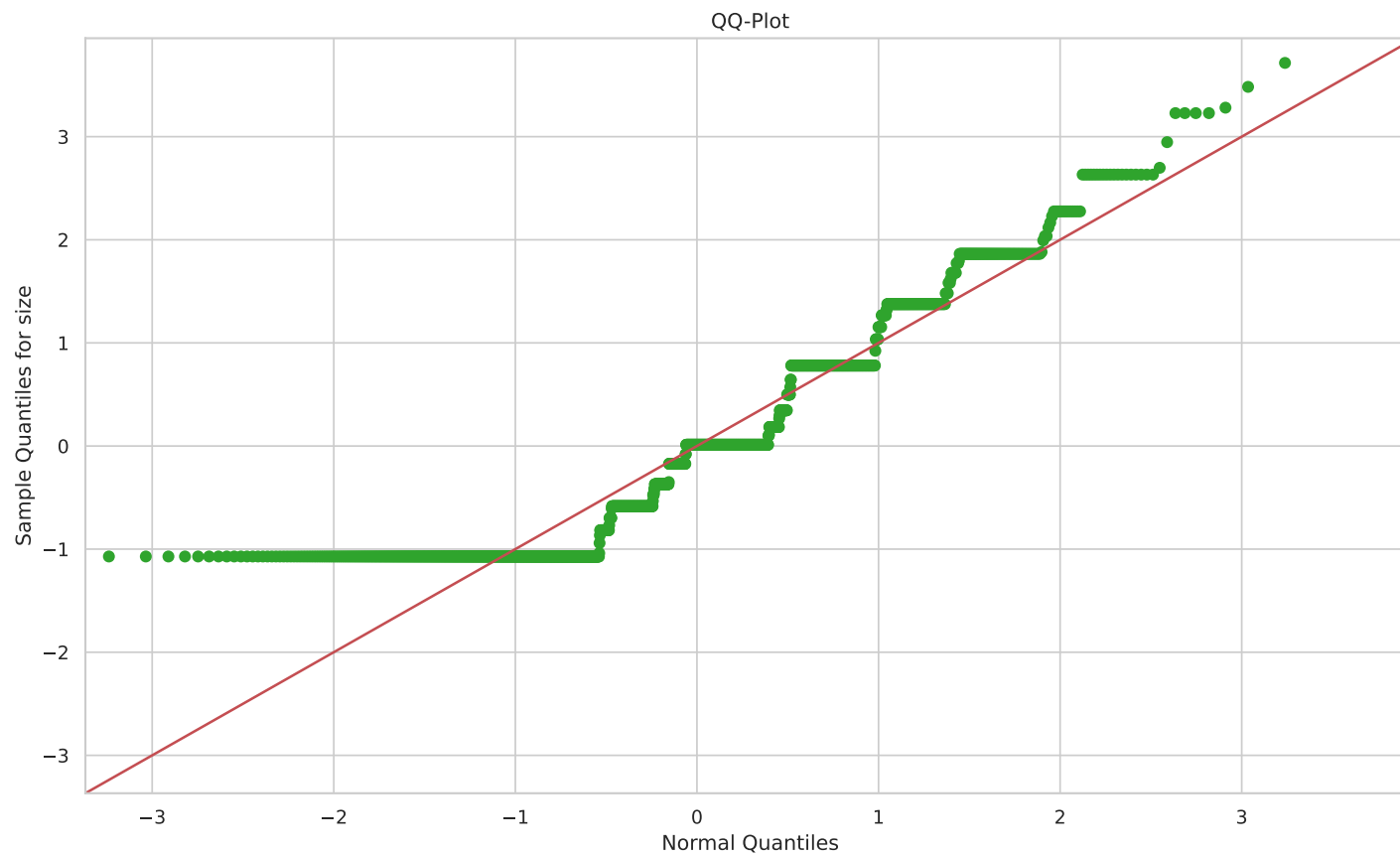


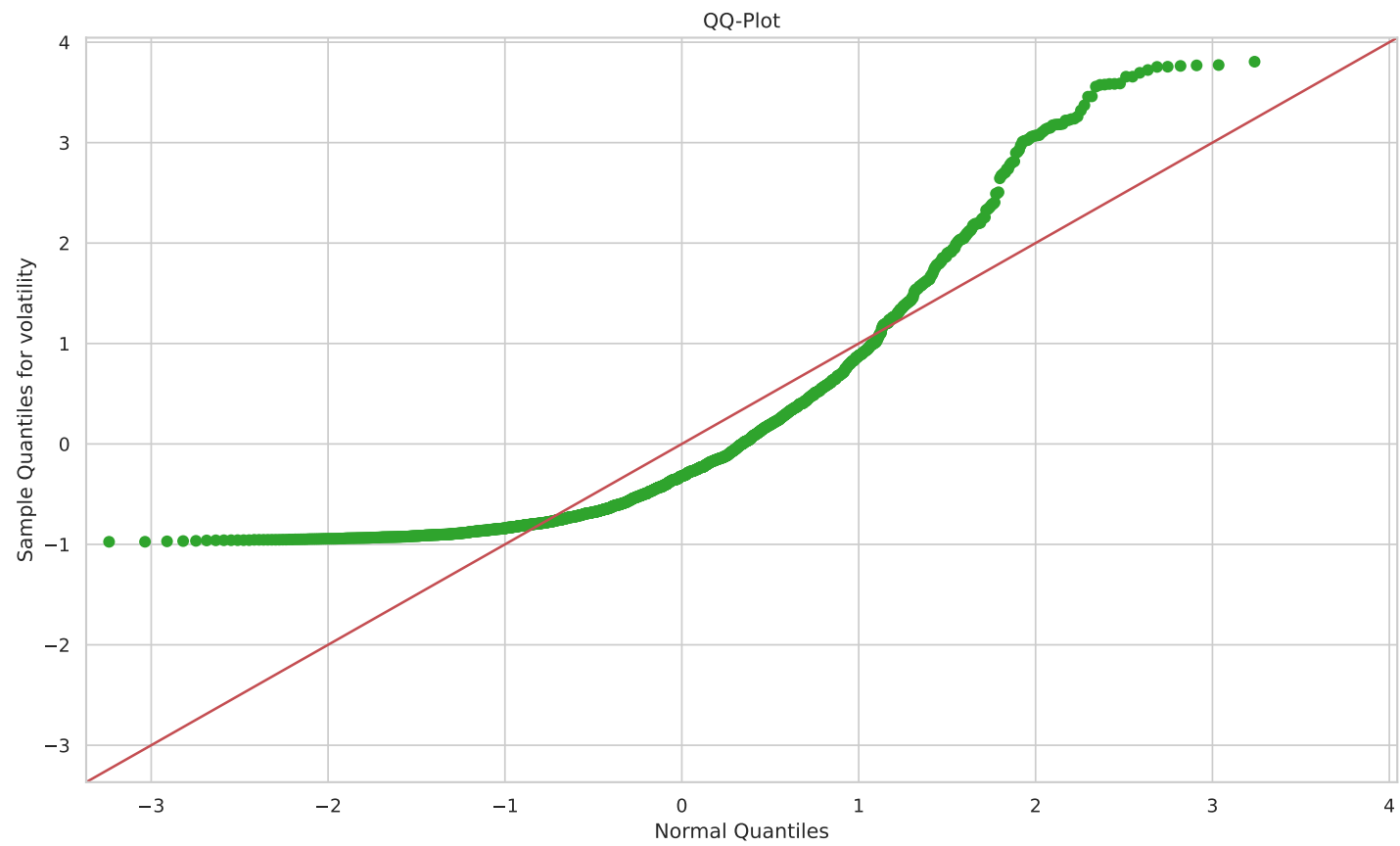
## QQ-Plots

One QQ-Plot per page for each variable. Variables are sorted alphabetically.

[]







## QQ-Plots Summary

Multiple QQ-Plots of variables in one figure. Variables are sorted alphabetically.

