

Exploratory Data Analysis (EDA)

Statsomat.com

19 April 2021

Basic Information

Automatic statistics for the file:

File
Baitingdata.csv

Your selection for the encoding: UTF-8

Your selection for the decimal character: Auto

Observations (rows with at least one non-missing value): 160

Variables (columns with at least one non-missing value): 25

Variables considered continuous: 7

Variables considered continuous
1st attack
1st attack stop
1st locate
2nd attack
CBH (cm)
DBH (cm)
height (m)

Variables considered categorical: 18

Variables considered categorical
2nd attack stop
ant sample
Attacked
Baiting tree no.
date
Detected
elevation (m)
field notes
H: 0
H: 1-5%
H: 33+%
H: 5-33%
lab notes
Recruited
species
Termite/C
transect
Tree number

Warning: More than 90% of the values of these columns could be treated as numeric. Nevertheless, because of some values or the selected decimal character, the columns must be treated as discrete. Are all the values plausible? Please check the data once more before uploading! Column(s): H: 0 H: 1-5% H: 5-33% H: 33+%

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
1st attack	160	44	116	72.5	94	157.53	149.91	103.5	122.31	1	595	1.1	0.4	0.95
1st attack stop	160	44	116	72.5	34	504.39	169.32	600	0	58	600	-1.56	1.01	0.34
1st locate	160	36	124	77.5	100	131.59	132.23	100.5	108.97	1	595	1.56	2.39	1.0
2nd attack	160	137	23	14.37	21	332.61	152.19	319	167.53	73	595	0.3	-0.82	0.46
CBH (cm)	160	0	160	100	52	6.23	6.52	5.3	5.34	0	29.6	1.73	3.44	
DBH (cm)	160	0	160	100	52	1.98	2.08	1.69	1.7	0	9.42	1.73	3.44	
height (m)	160	0	160	100	46	2.9	1.92	2.5	1.56	0.4	9	1.43	2.15	0.66

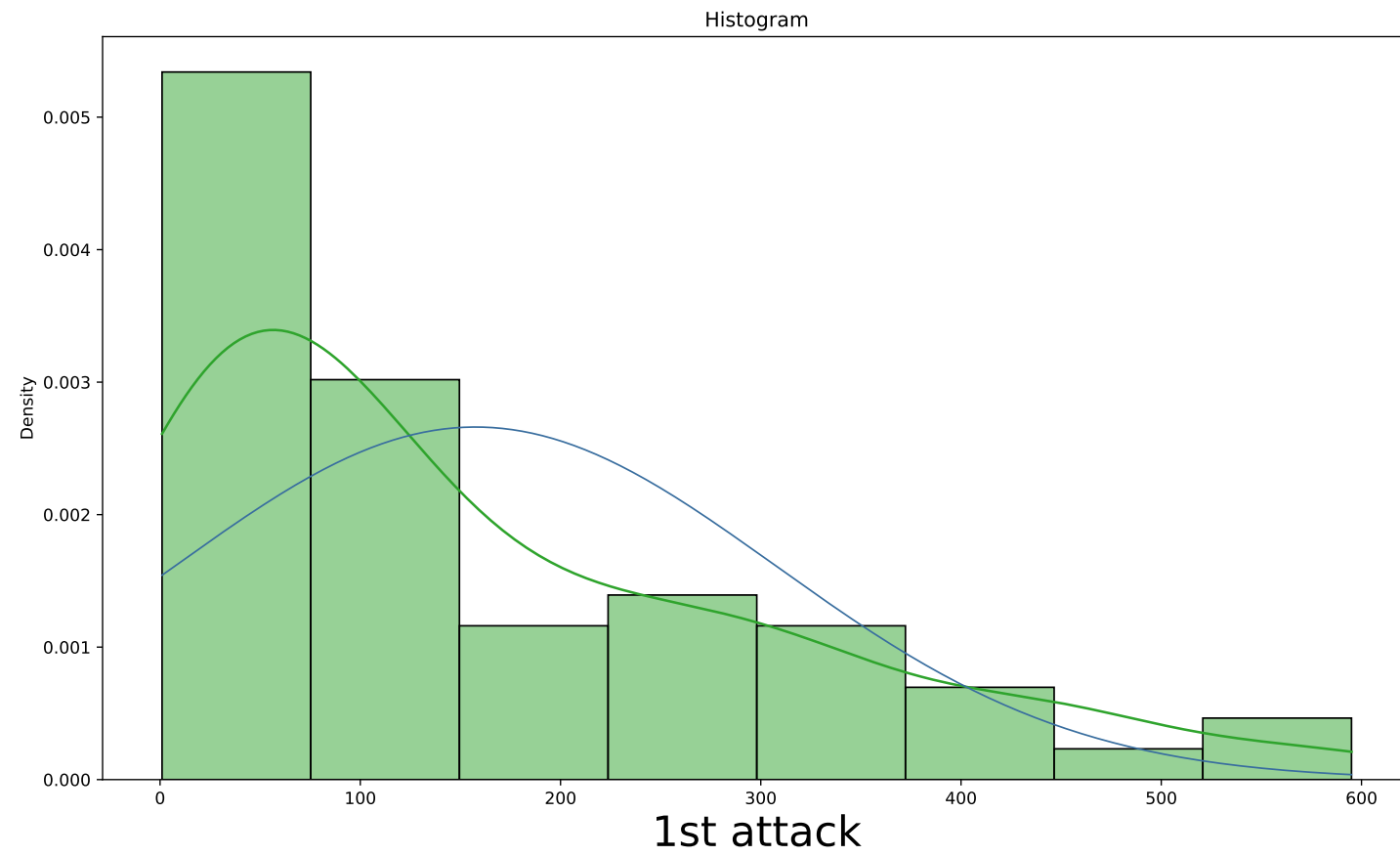
Graphics

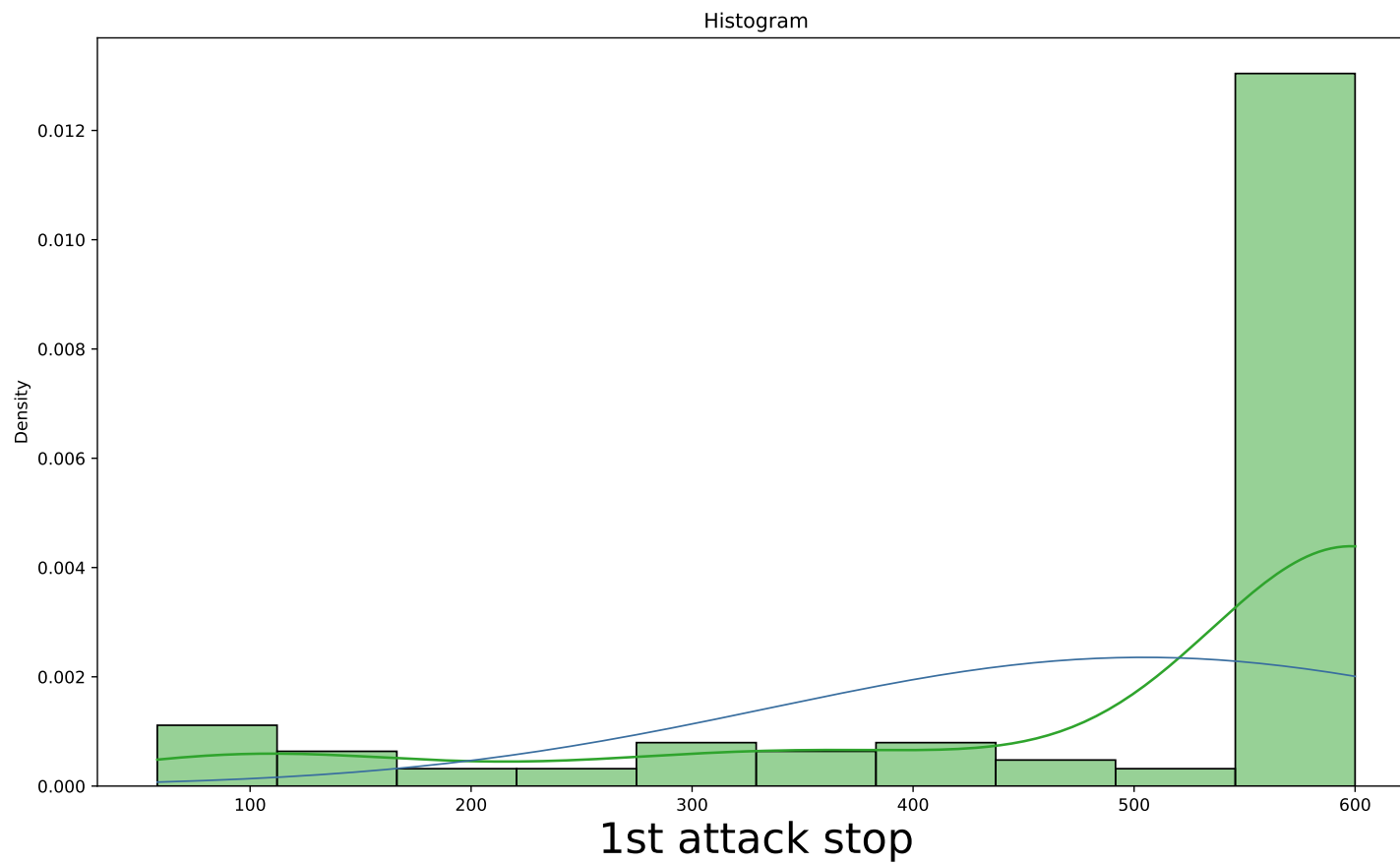
Histograms

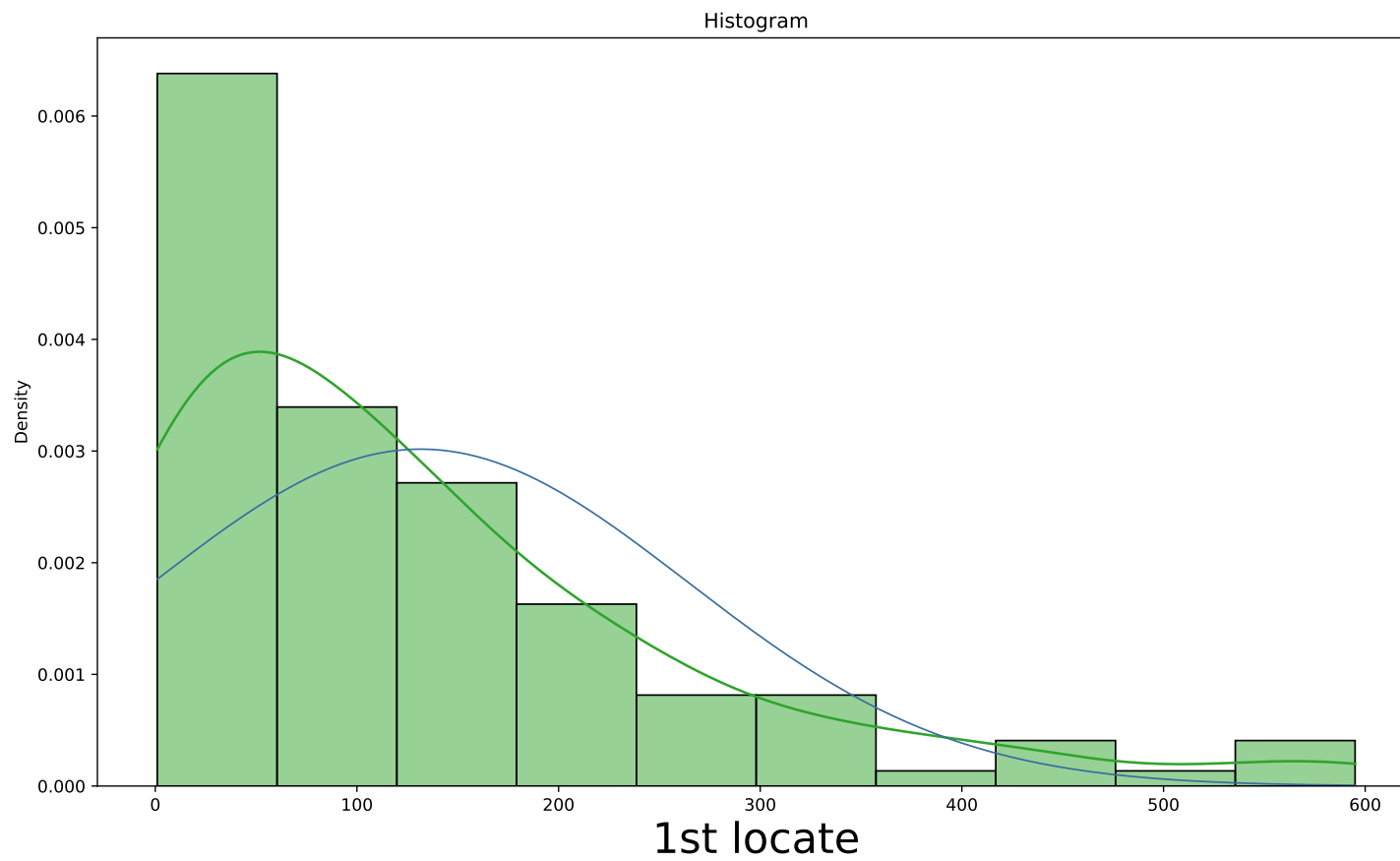
Details: Density Histograms. One large figure per page for each variable, sorted alphabetically. The blue line represents the normal density approximation. The green line represents a special kernel density approximation.

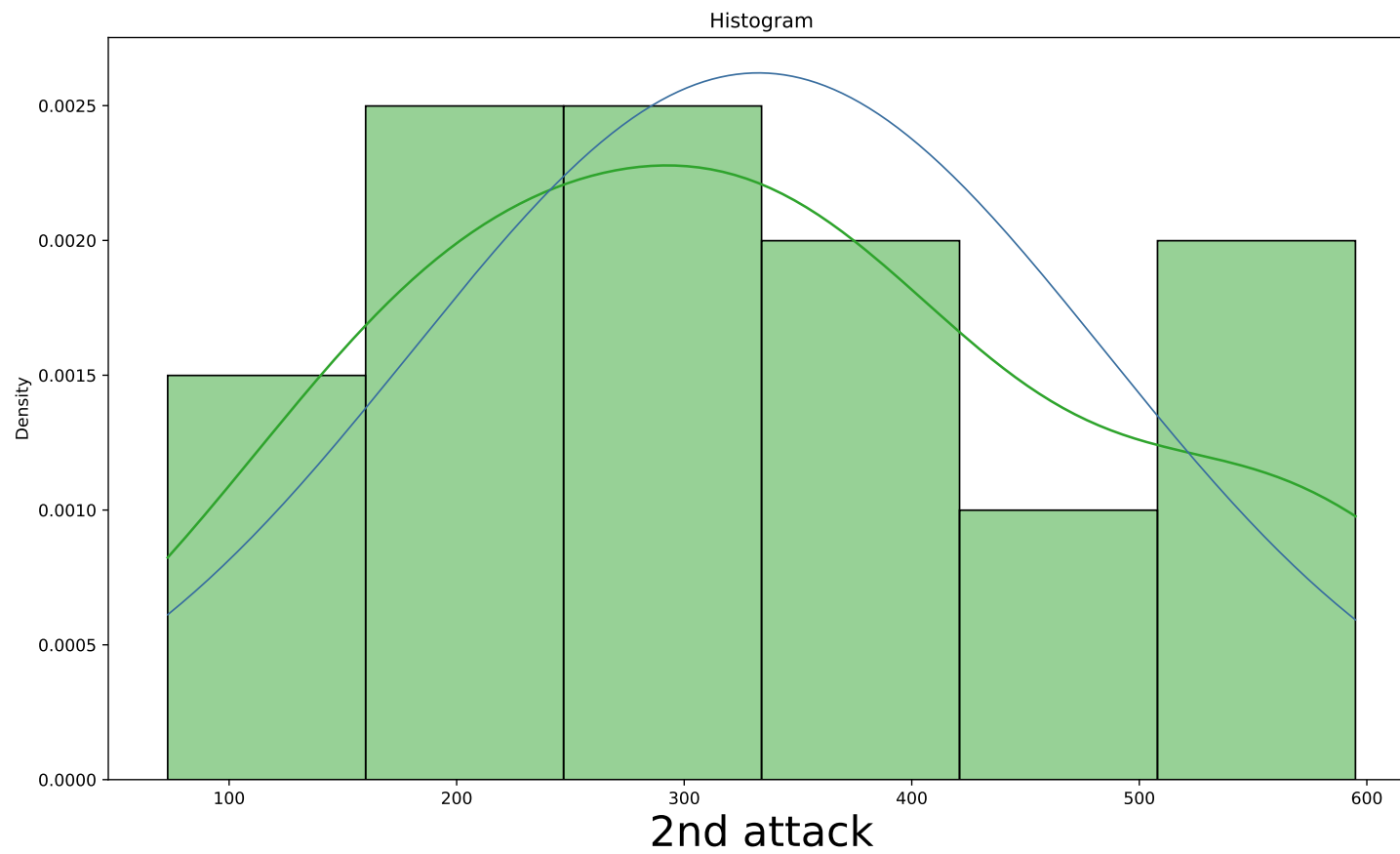
See figures on next page.

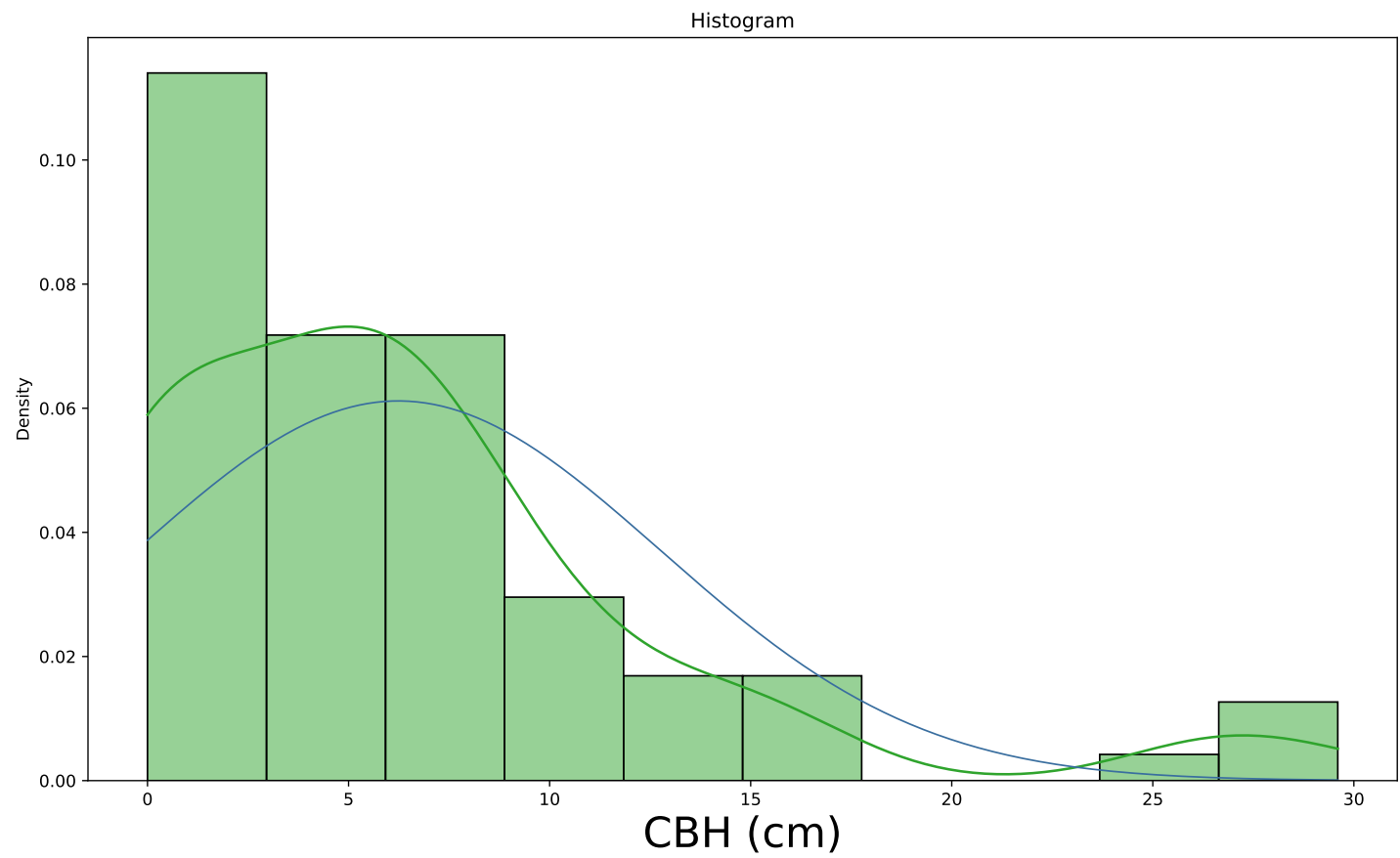
□

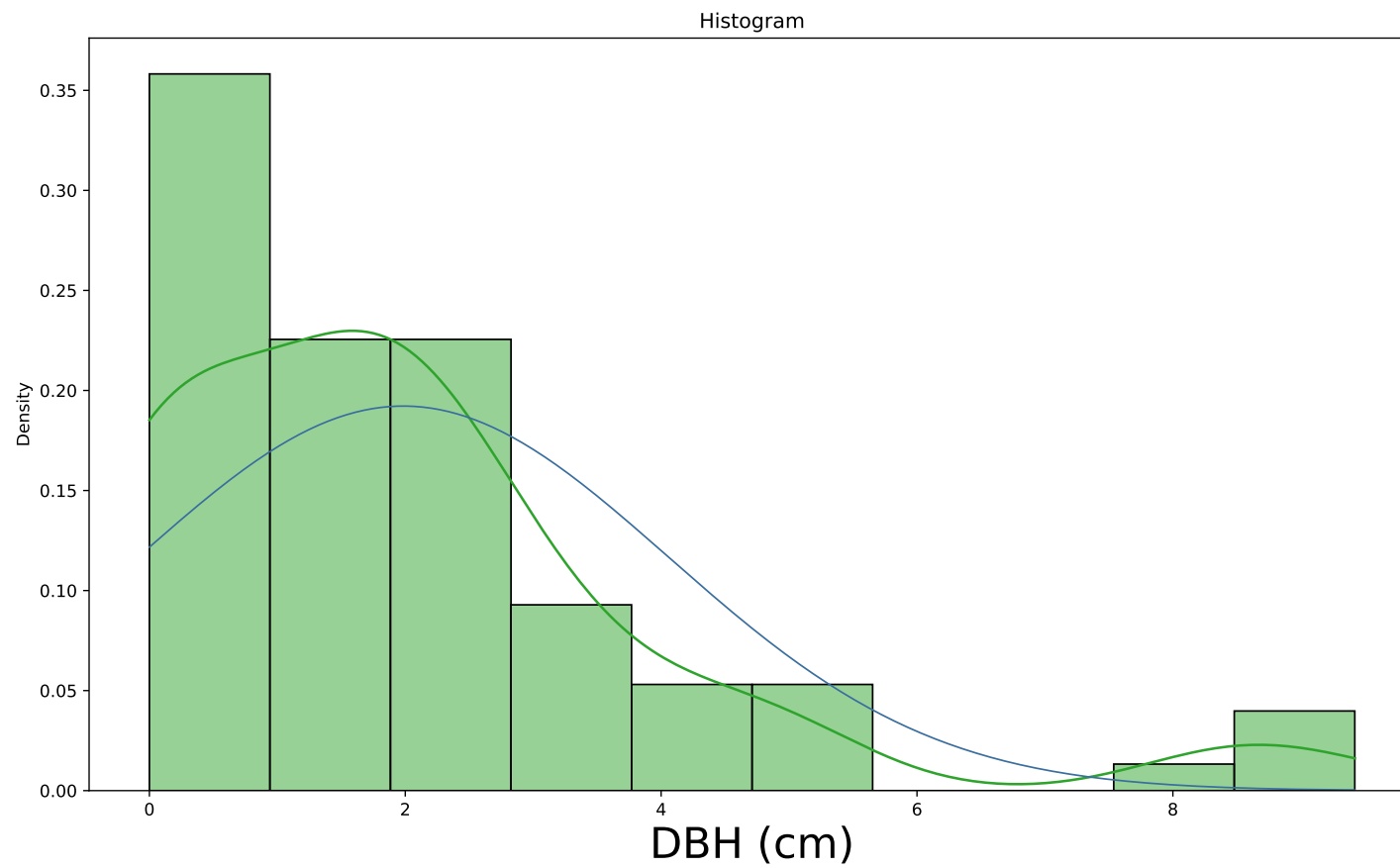


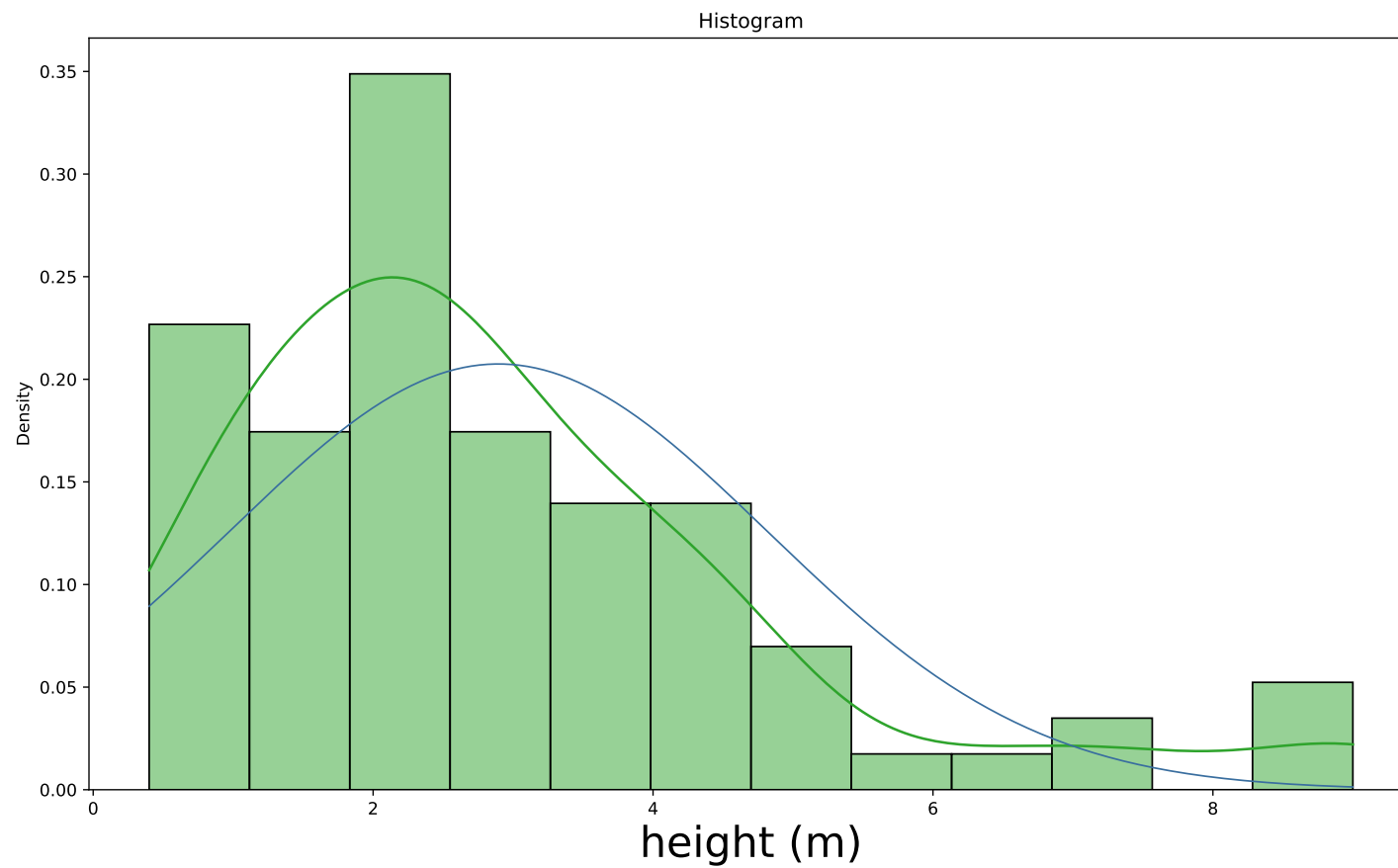






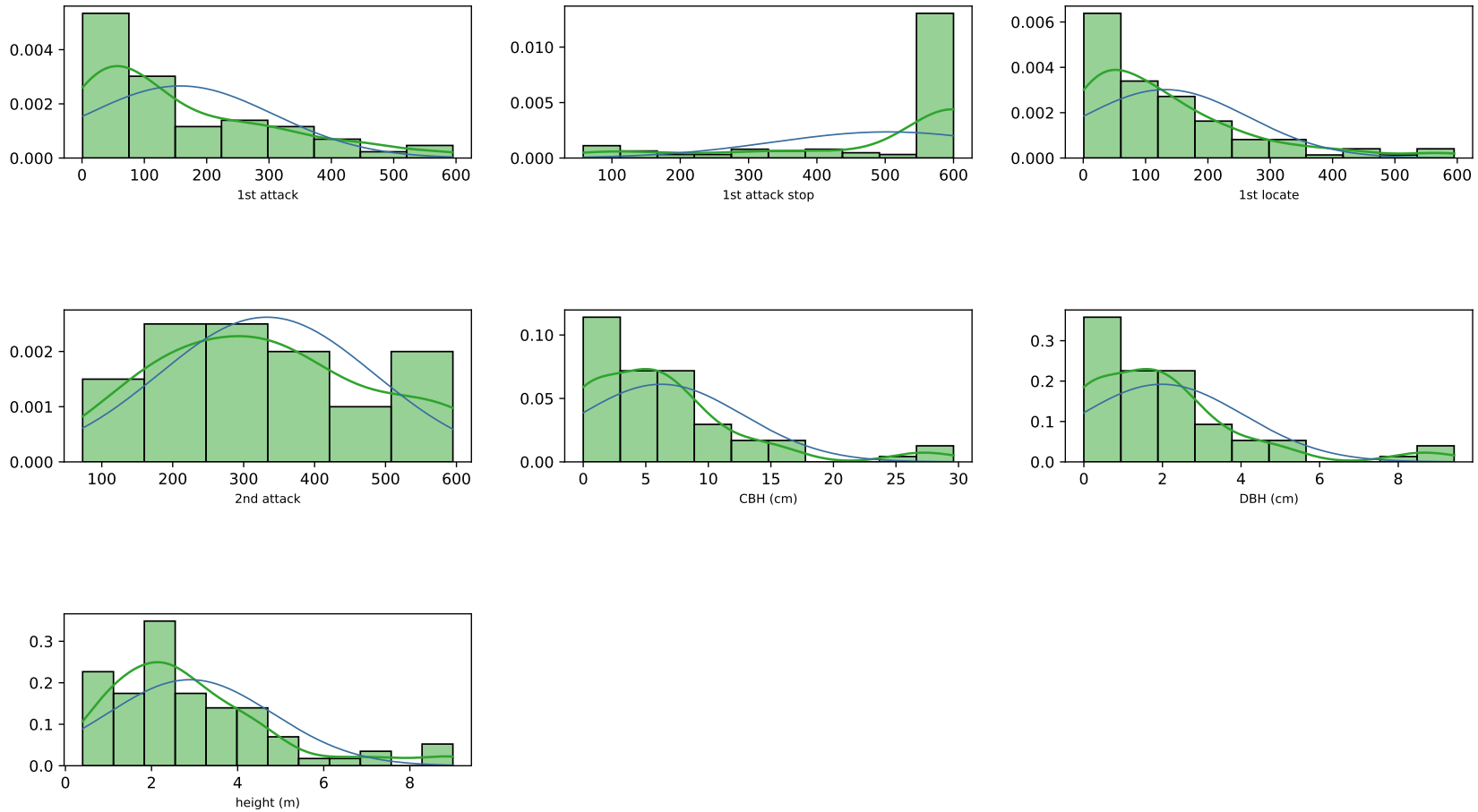






Histograms Summary

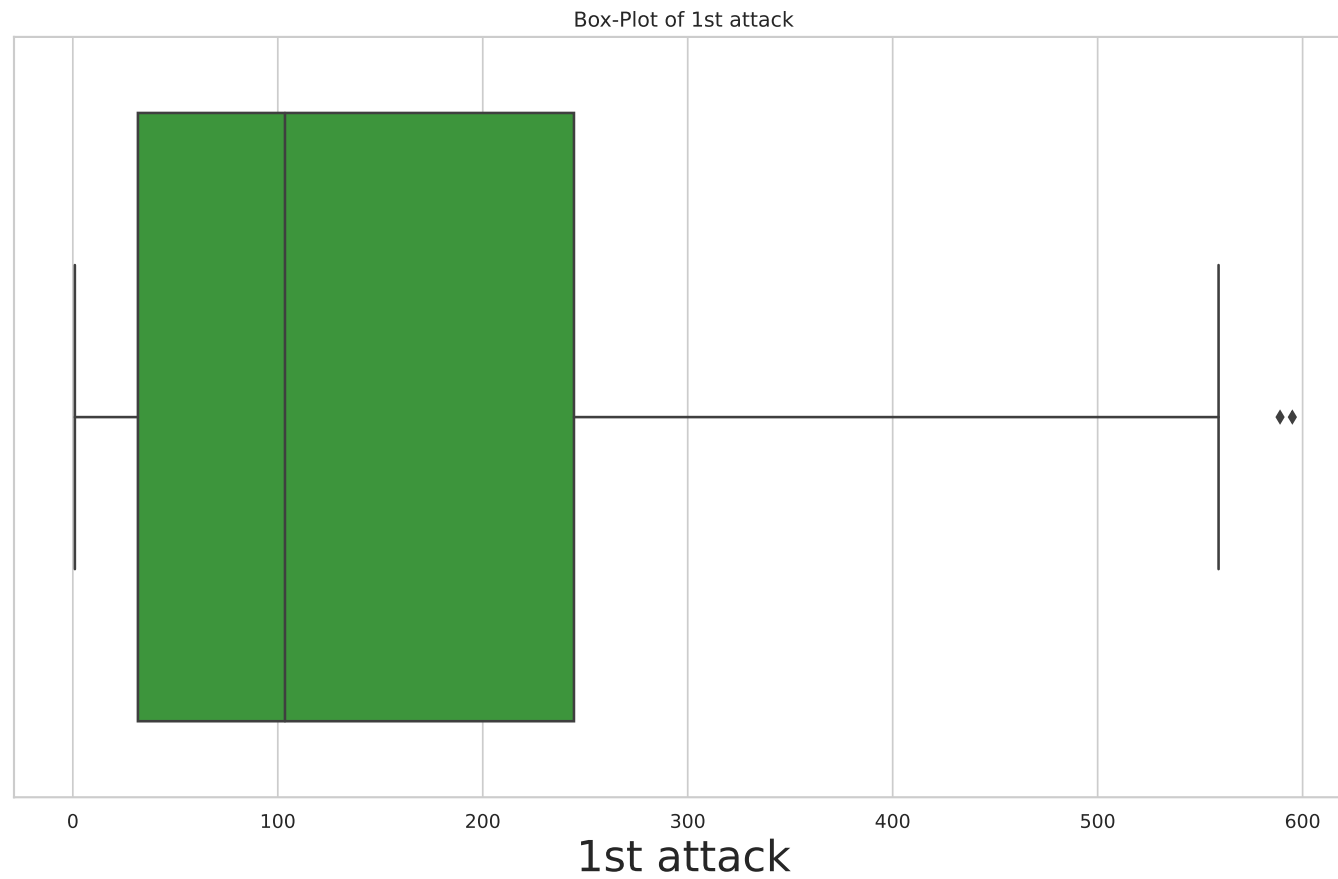
Multiple Relative Frequency Histogram in one figure. Variables are sorted alphabetically. The blue line represents the normal density approximation. The green line represents a special kernel density approximation.

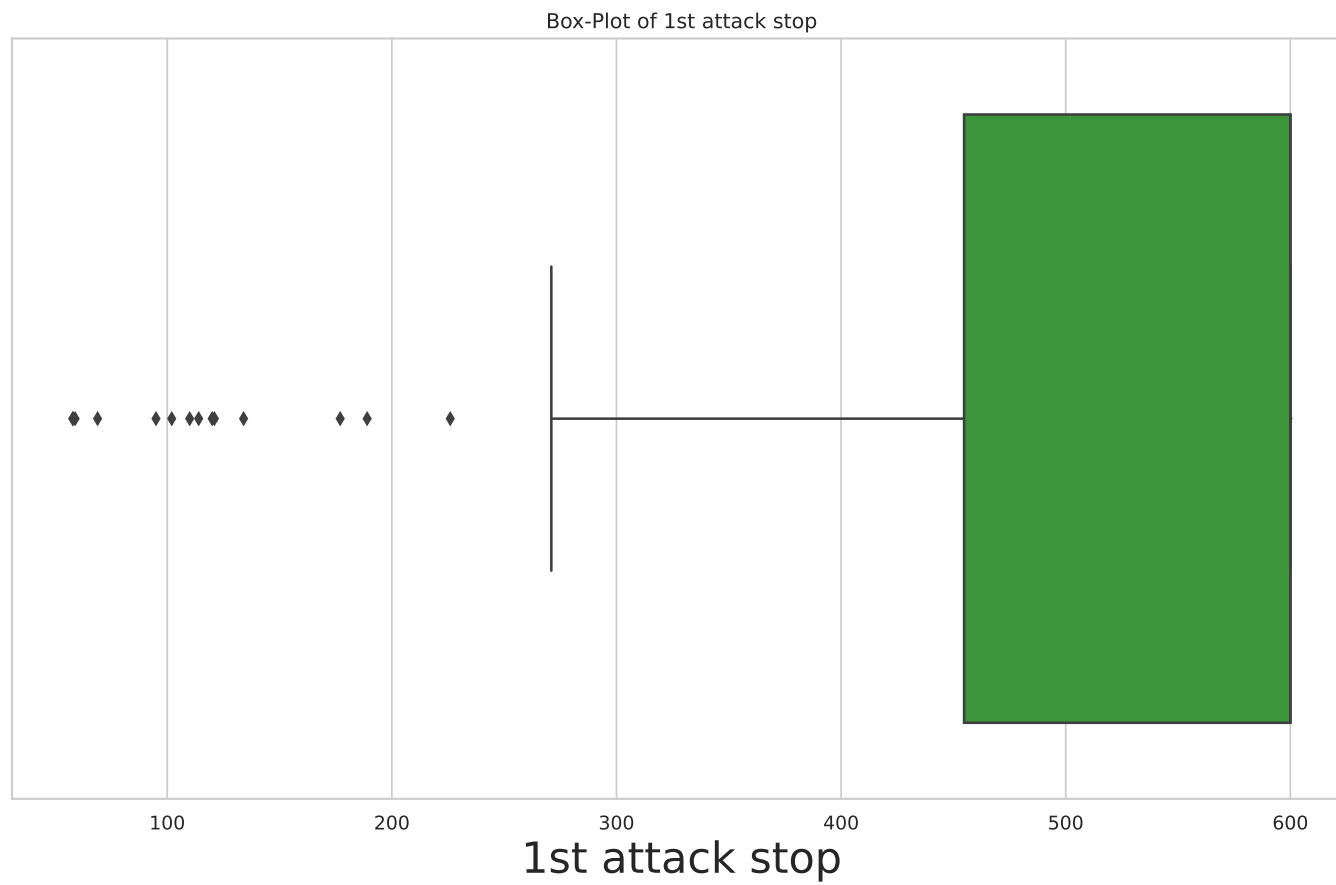


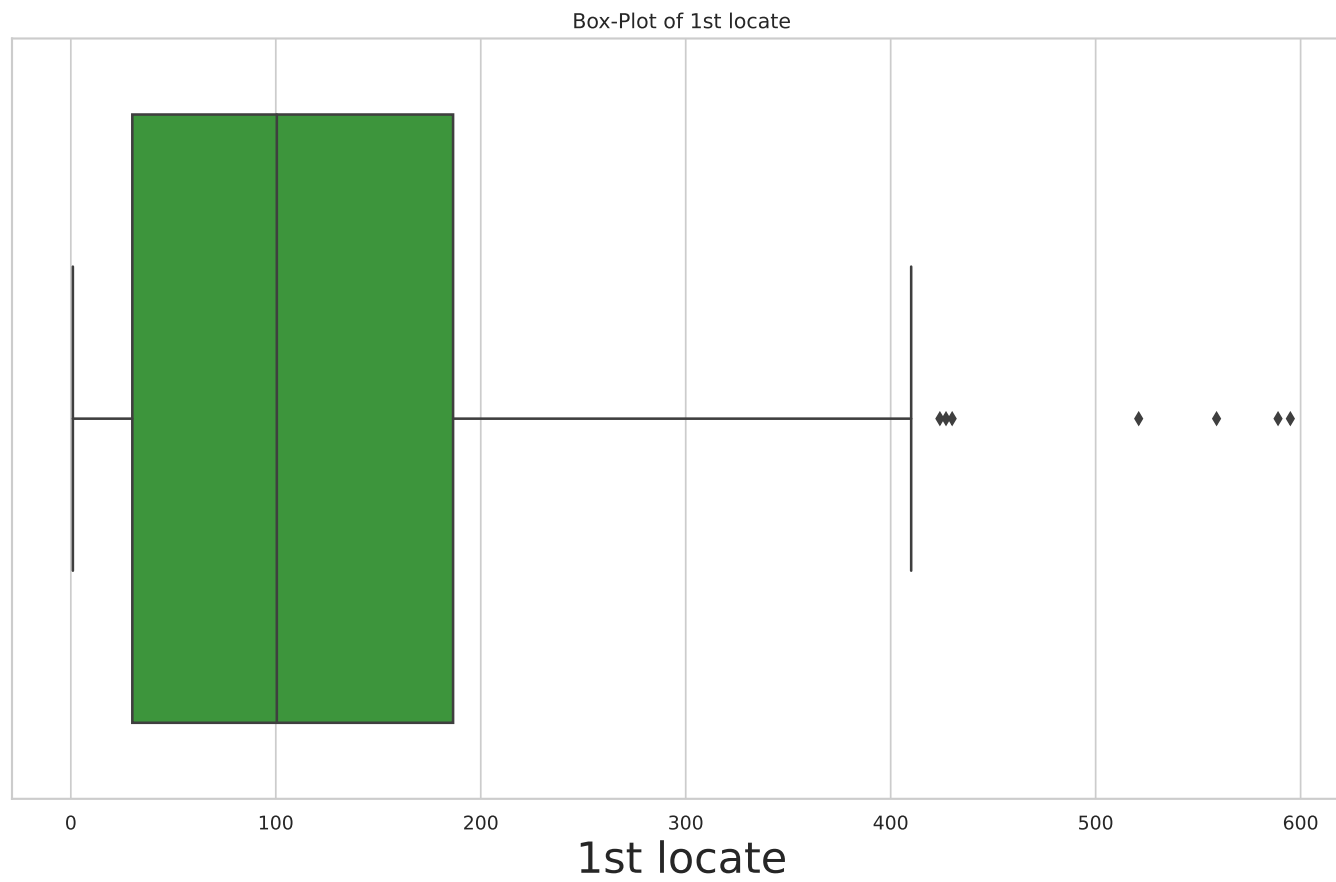
Box-Plots

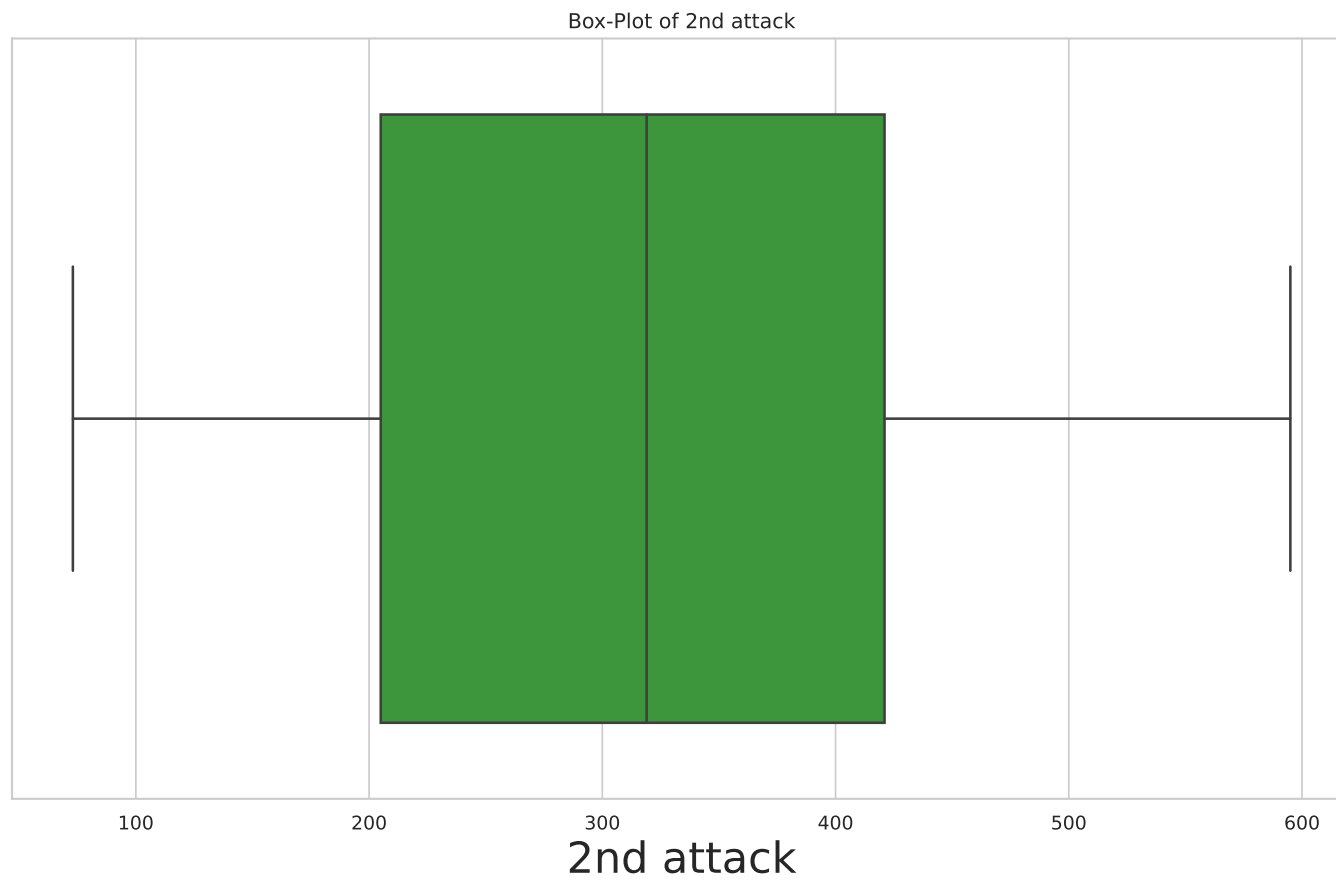
One Box-Plot per page for each variable. Variables are sorted alphabetically.

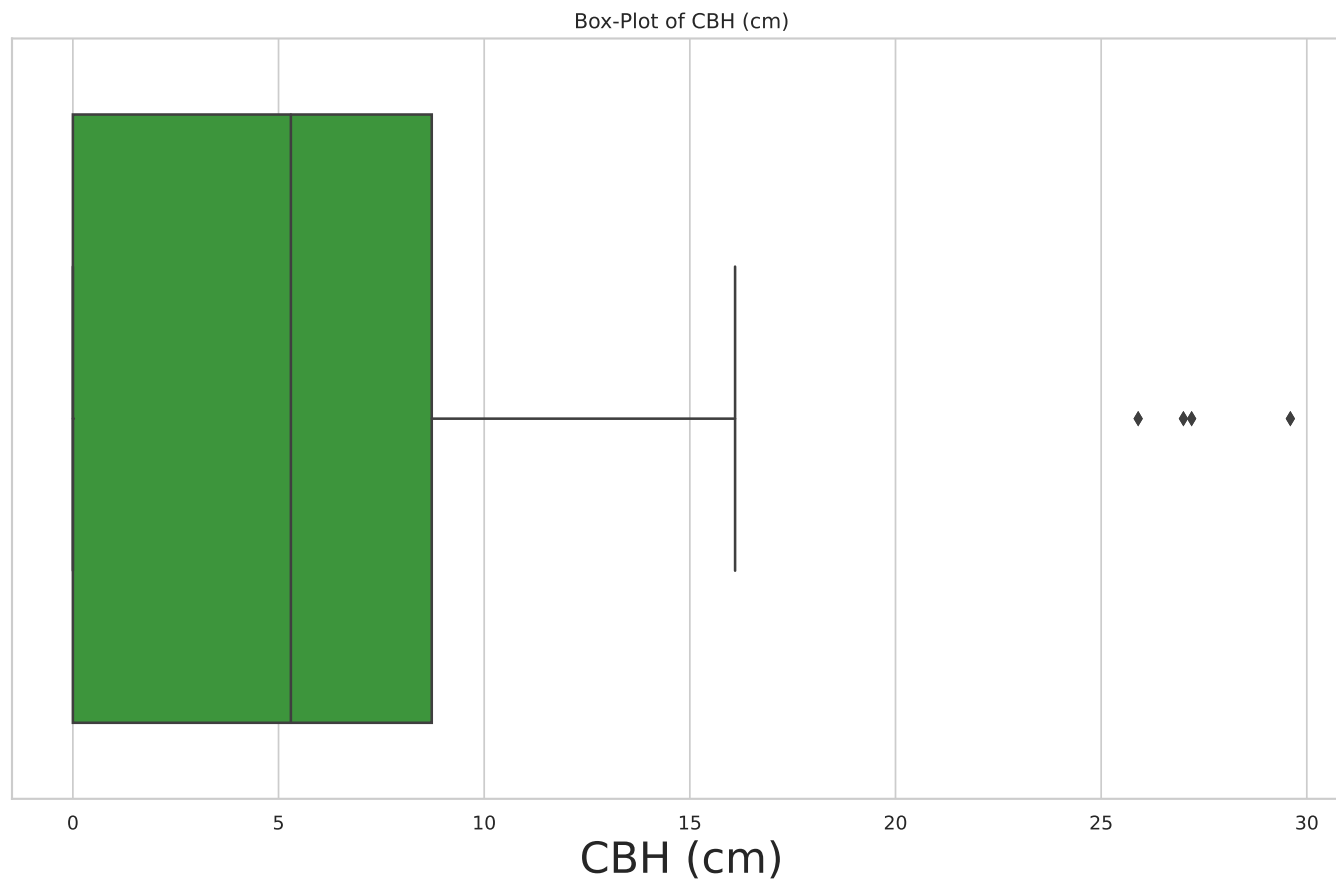
□

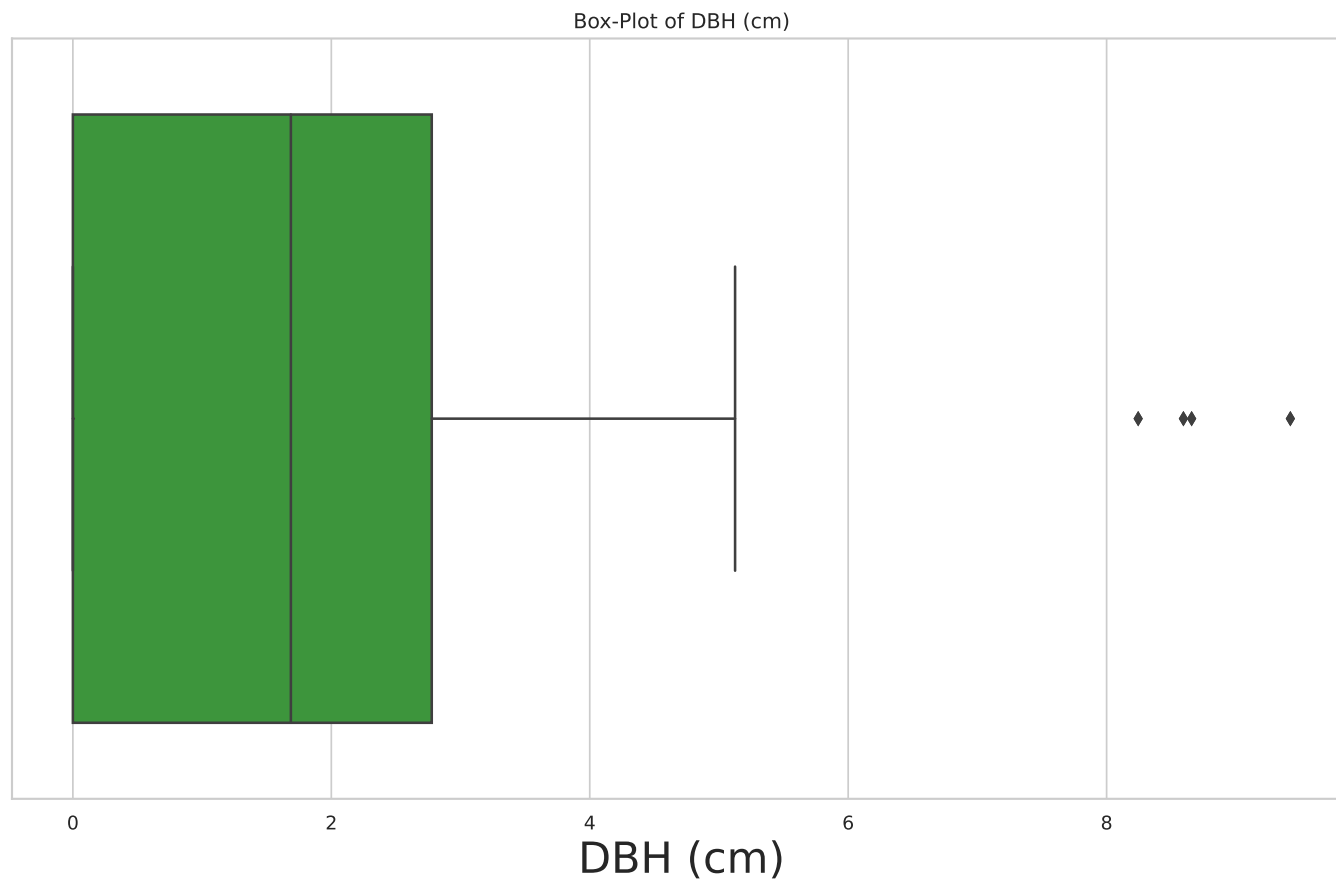


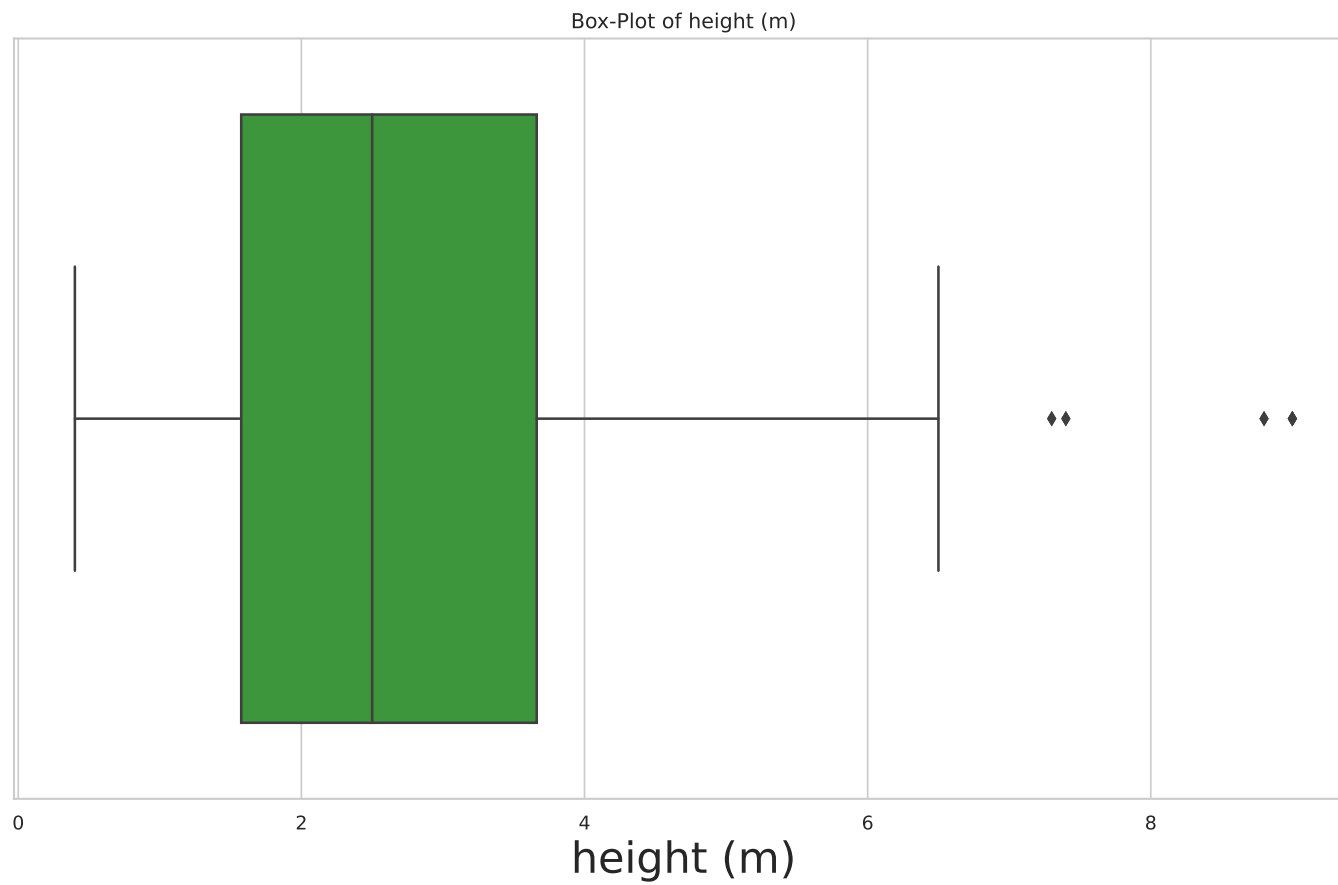






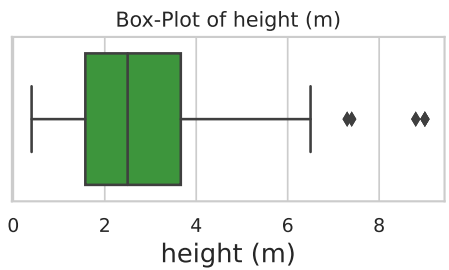
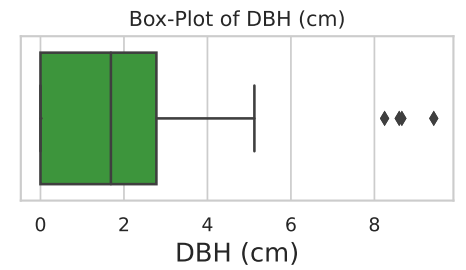
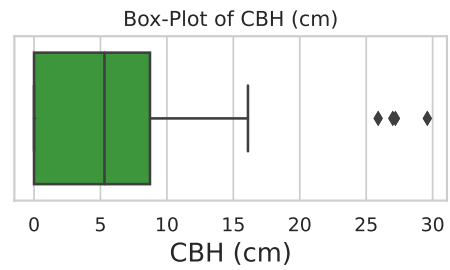
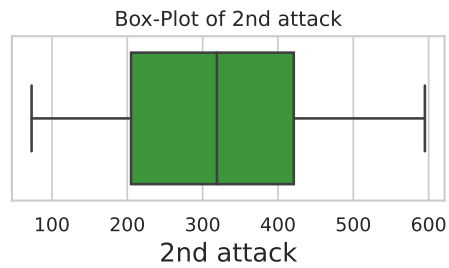
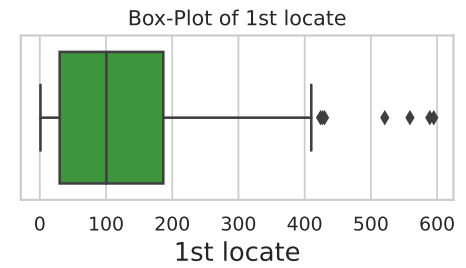
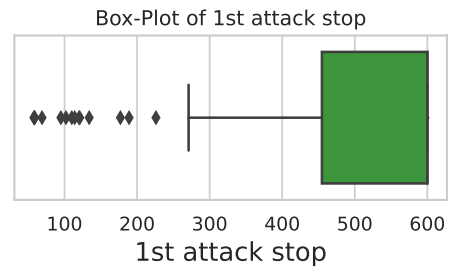
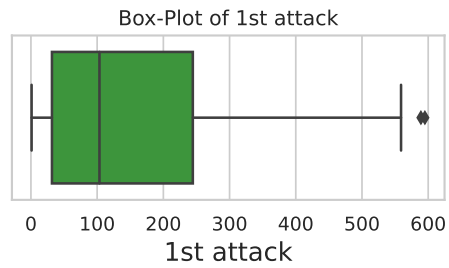






Box-Plots Summary

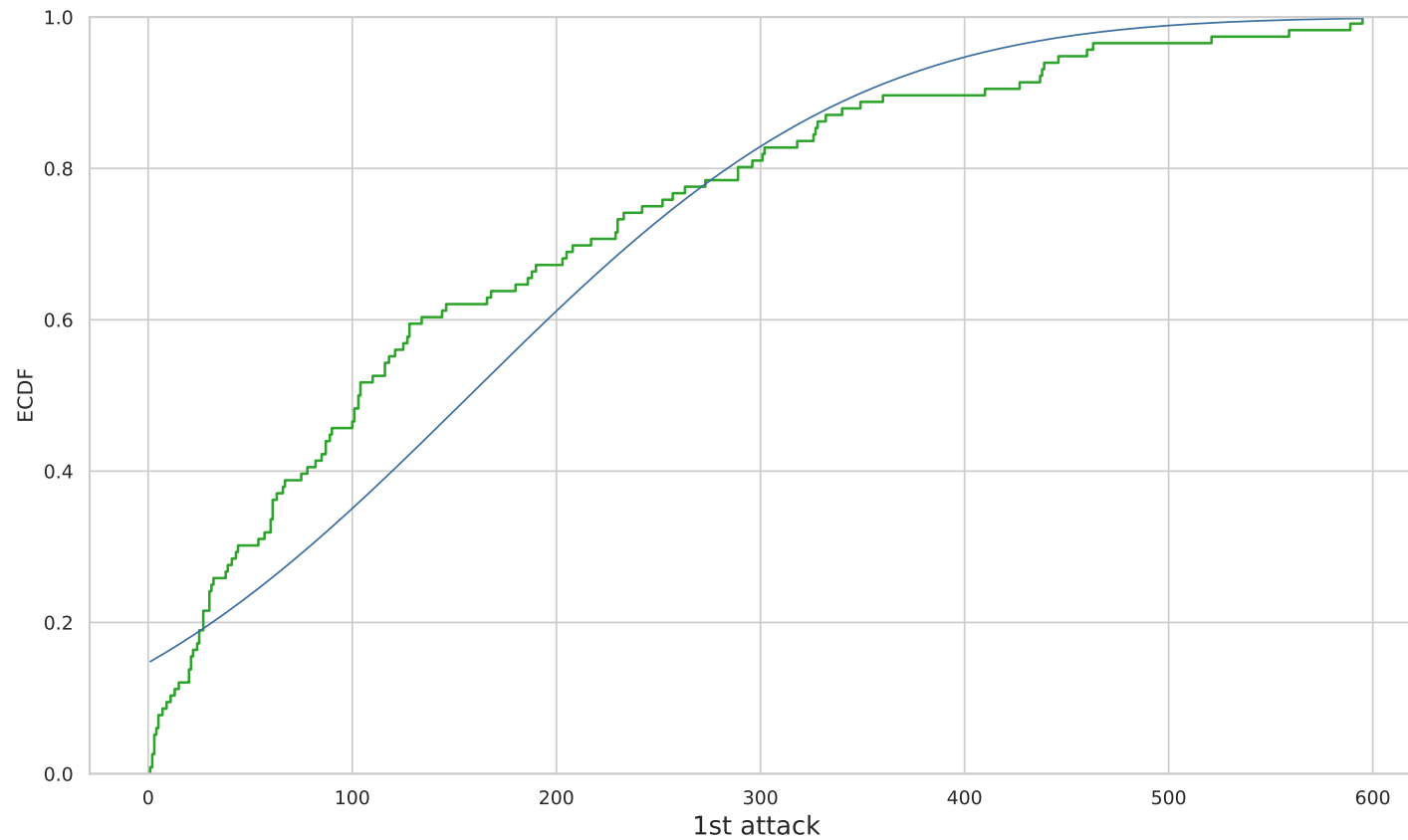
Multiple Box-Plots of variables in one figure. Variables are sorted alphabetically.

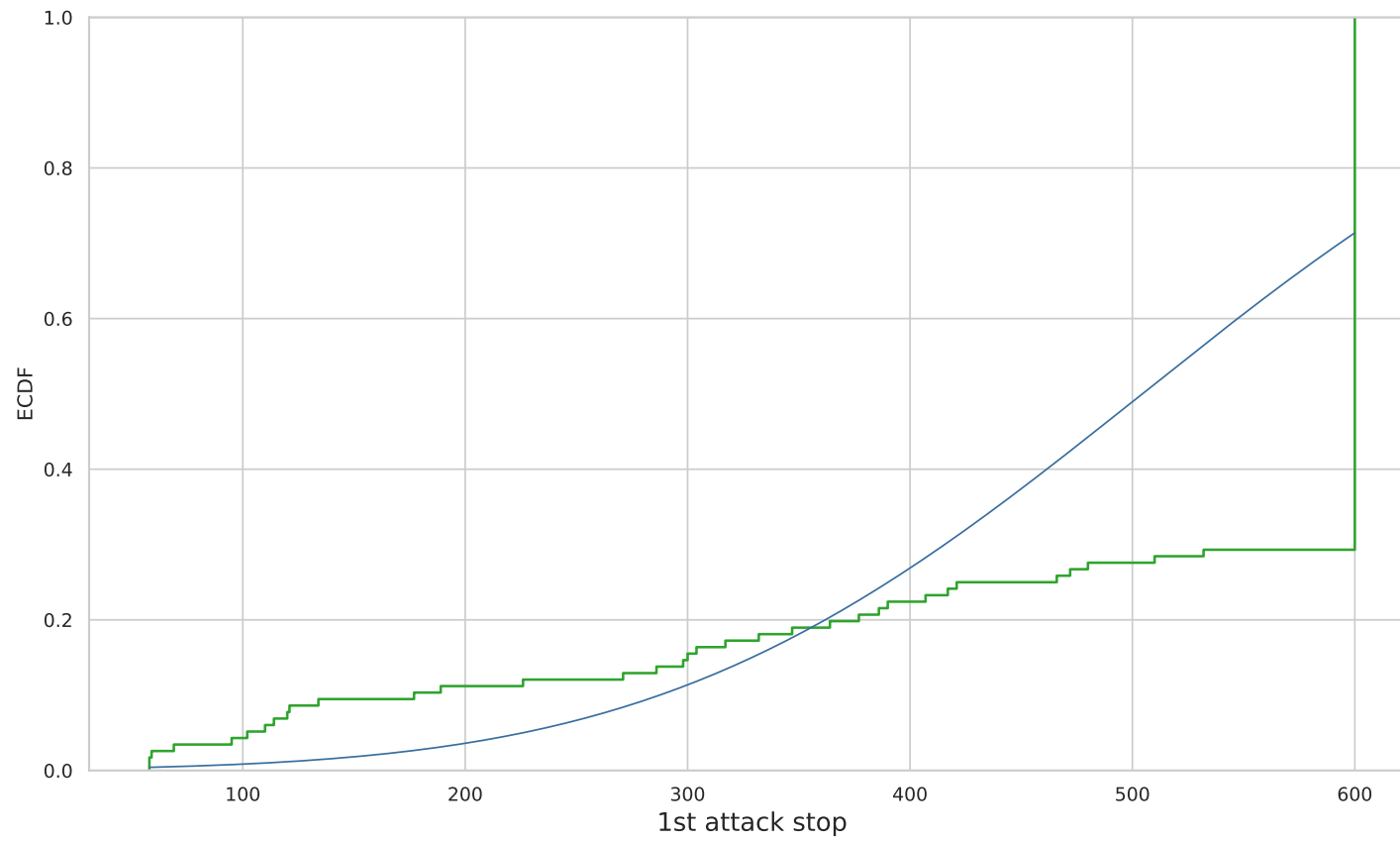


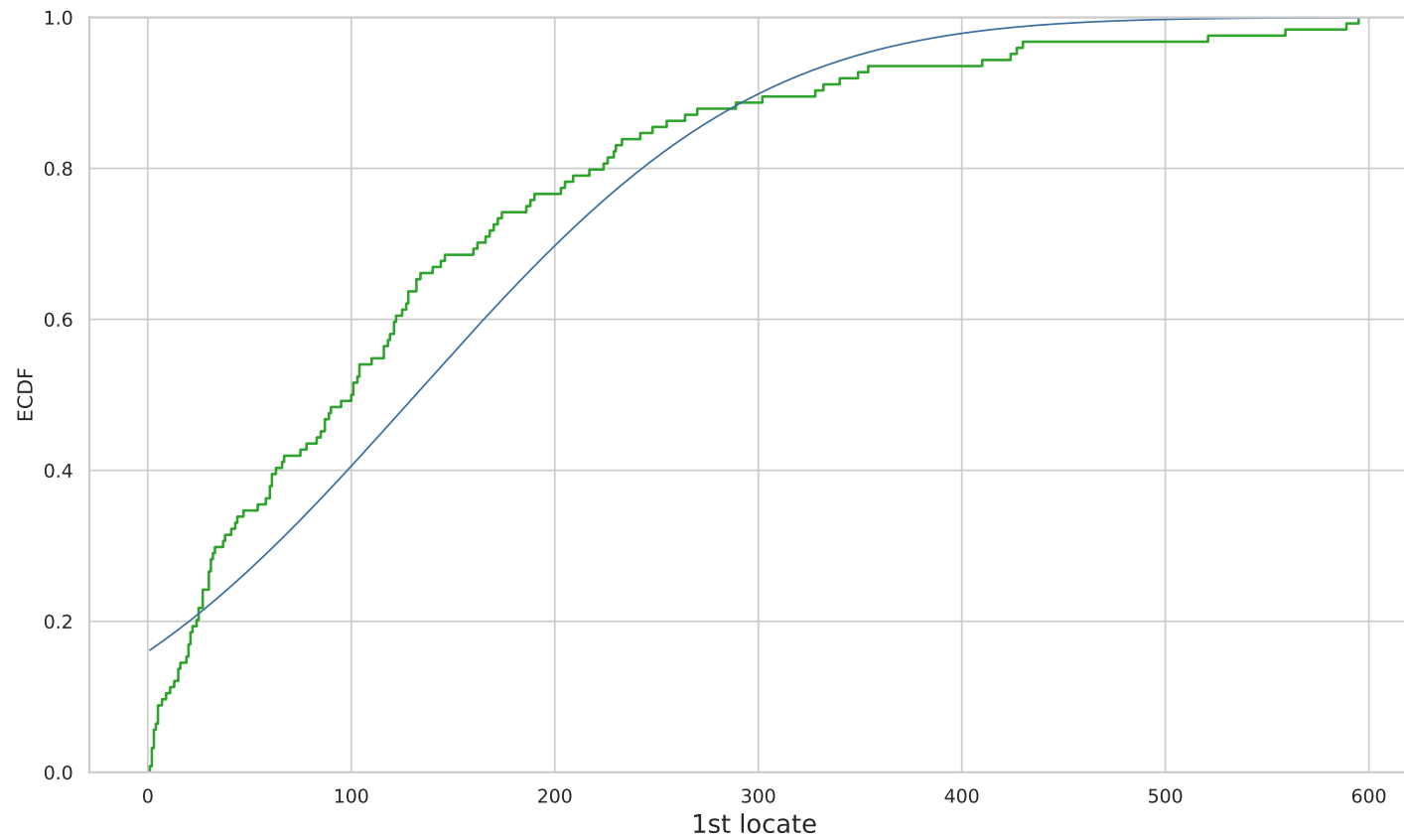
ECDF Plots

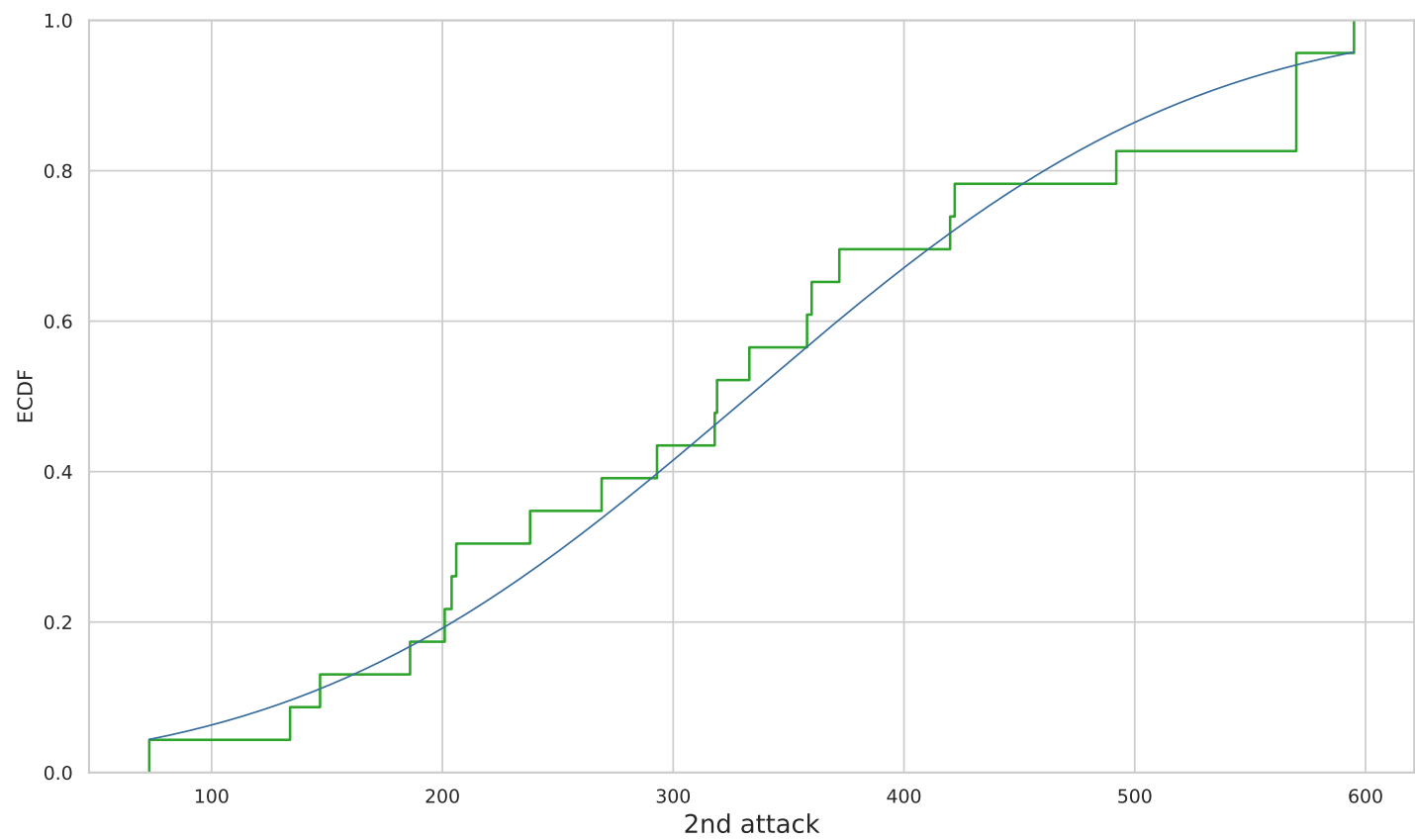
One ECDF (Empirical Cumulative Distribution Function) Plot per page for each variable. Variables are sorted alphabetically. The blue line represents the CDF of a normal distribution. If the variable is normally distributed, the blue line approximates well the ECDF.

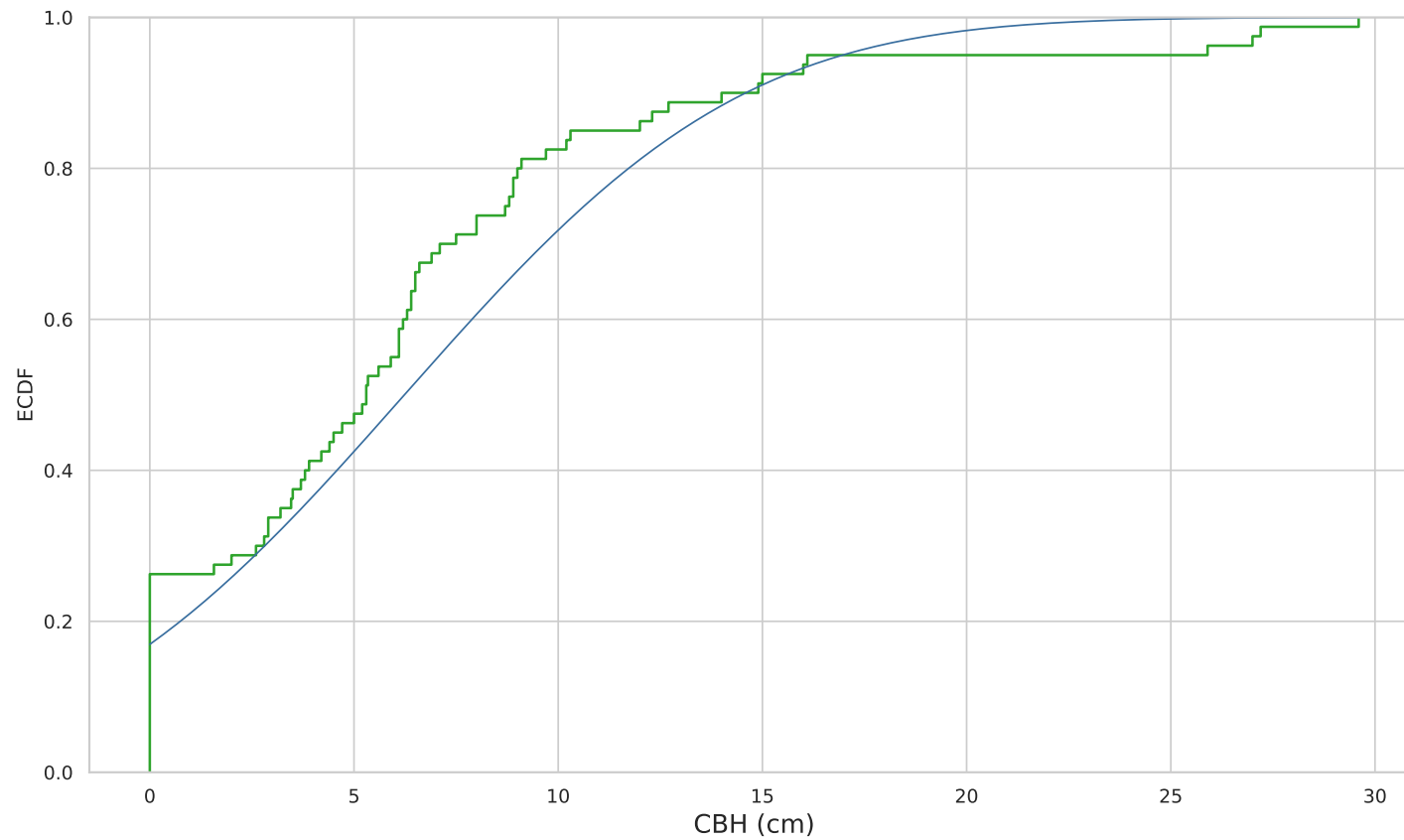
□

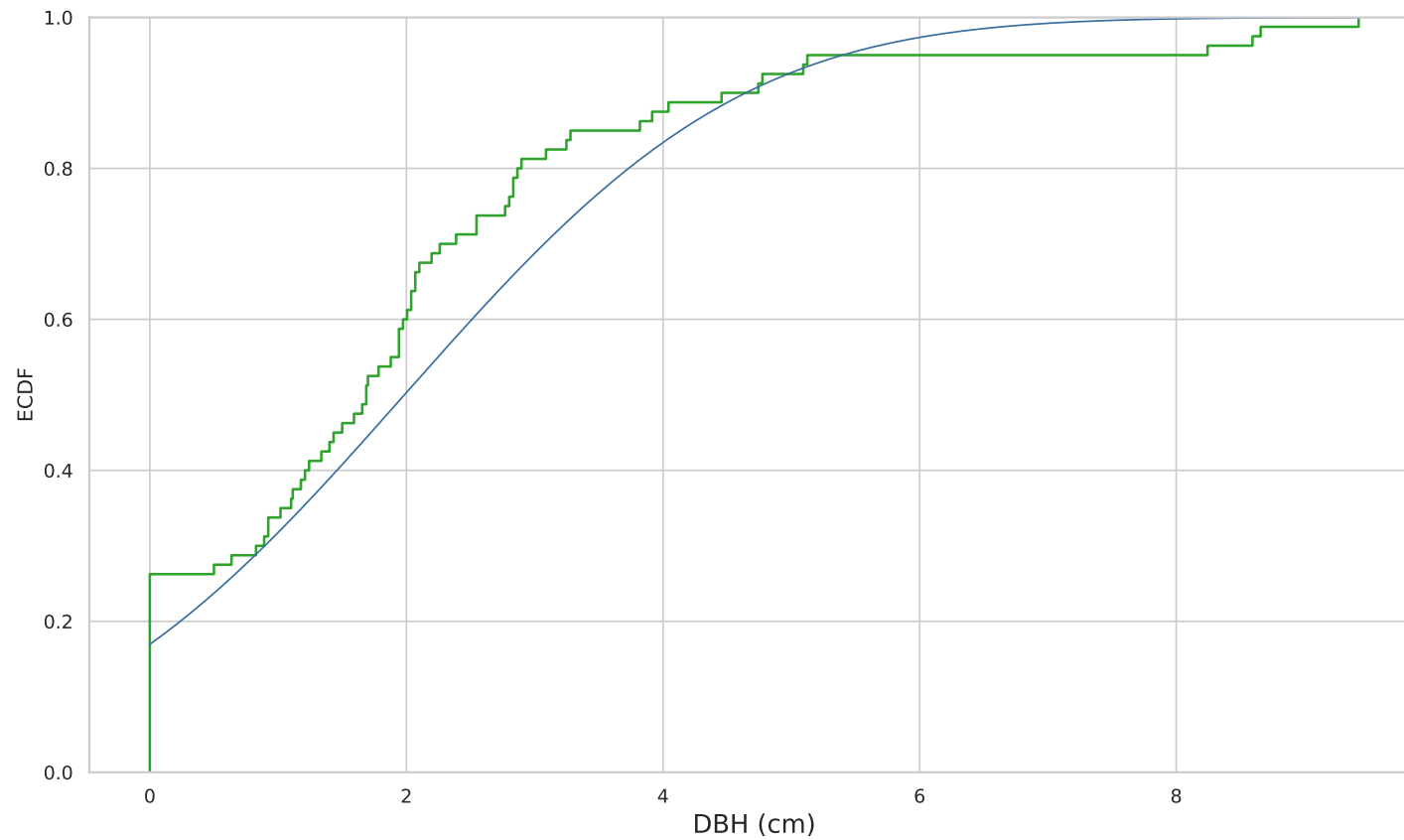


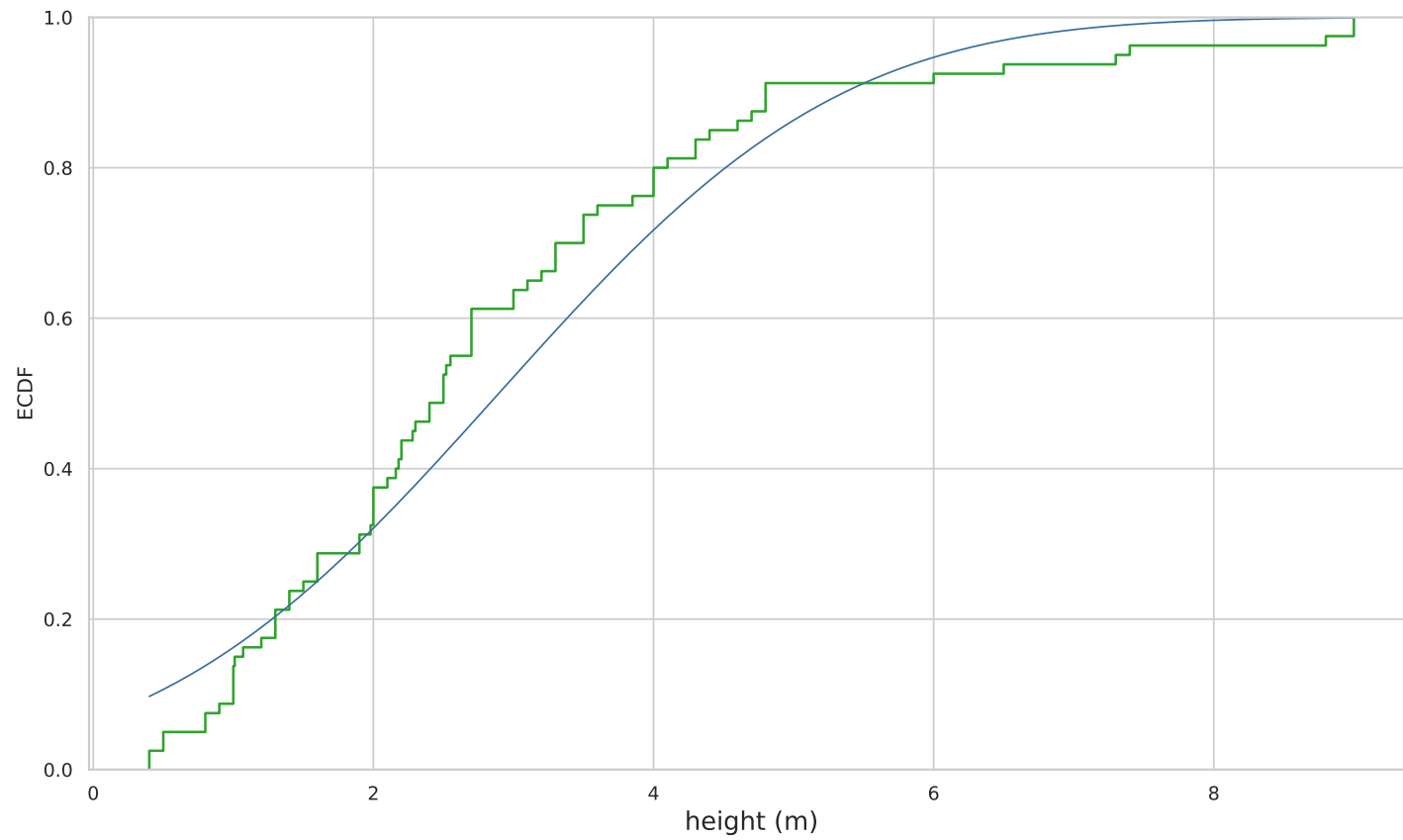






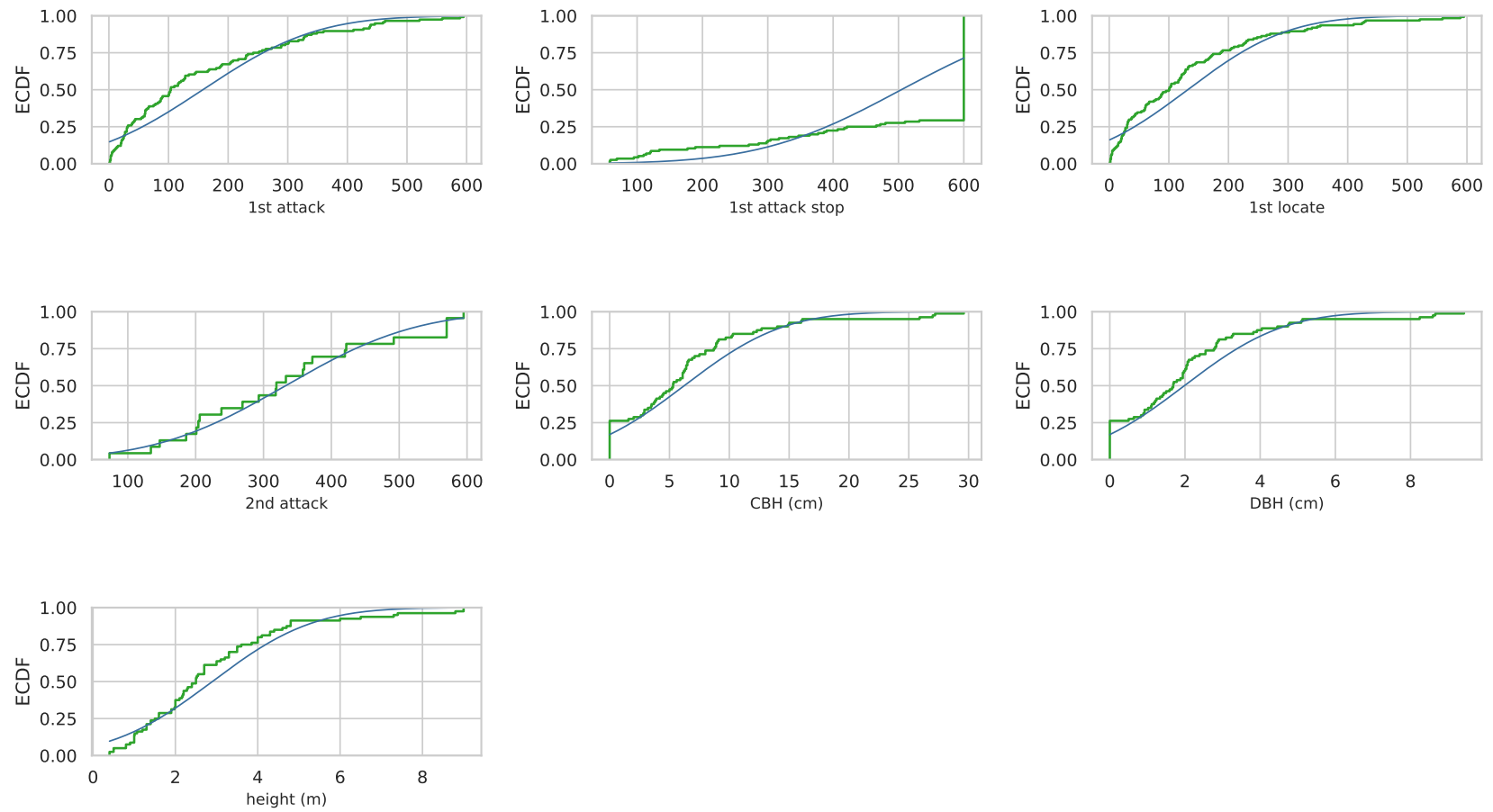






ECDF Plots Summary

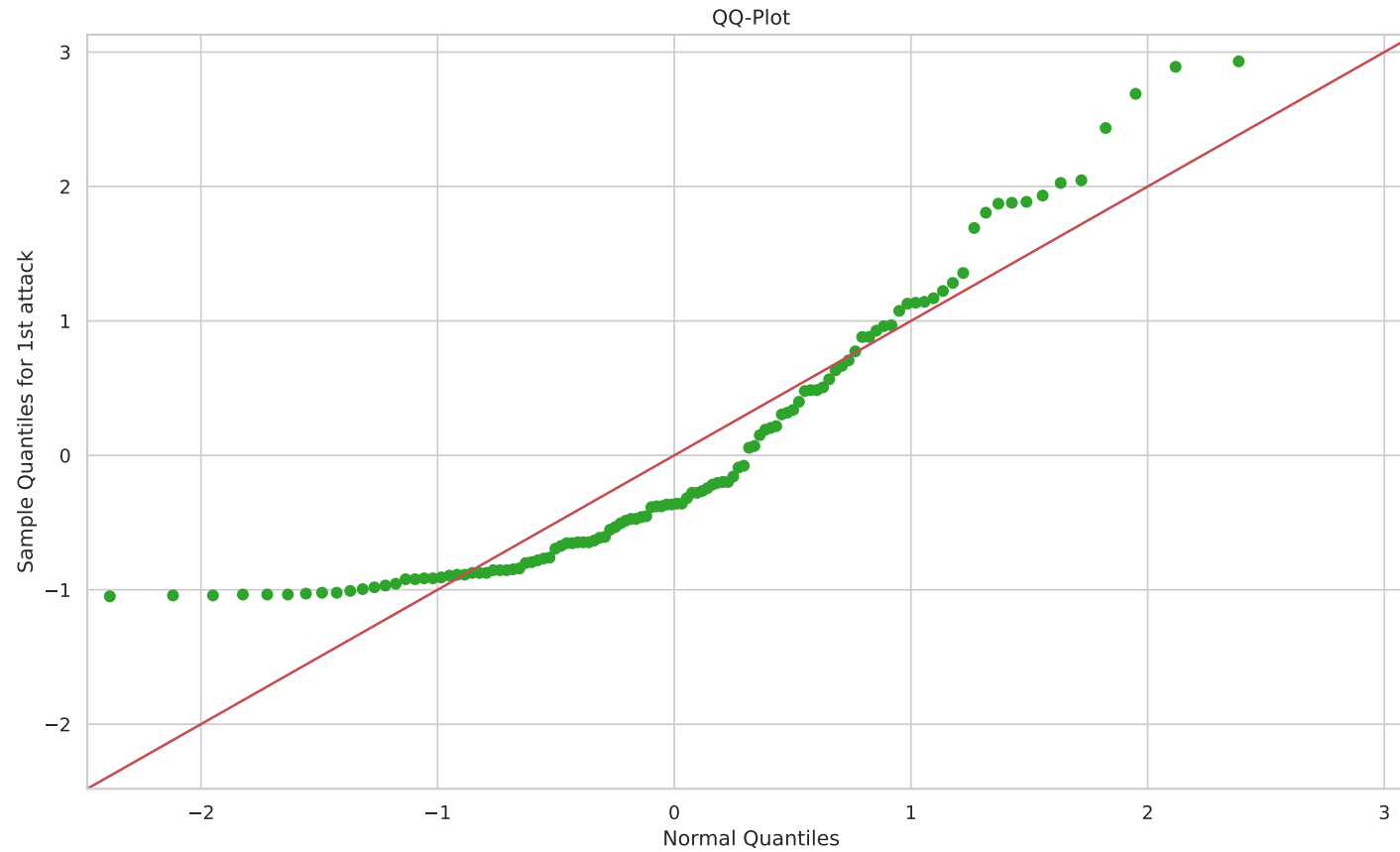
Multiple ECDF Plots of variables in one figure. Variables are sorted alphabetically.

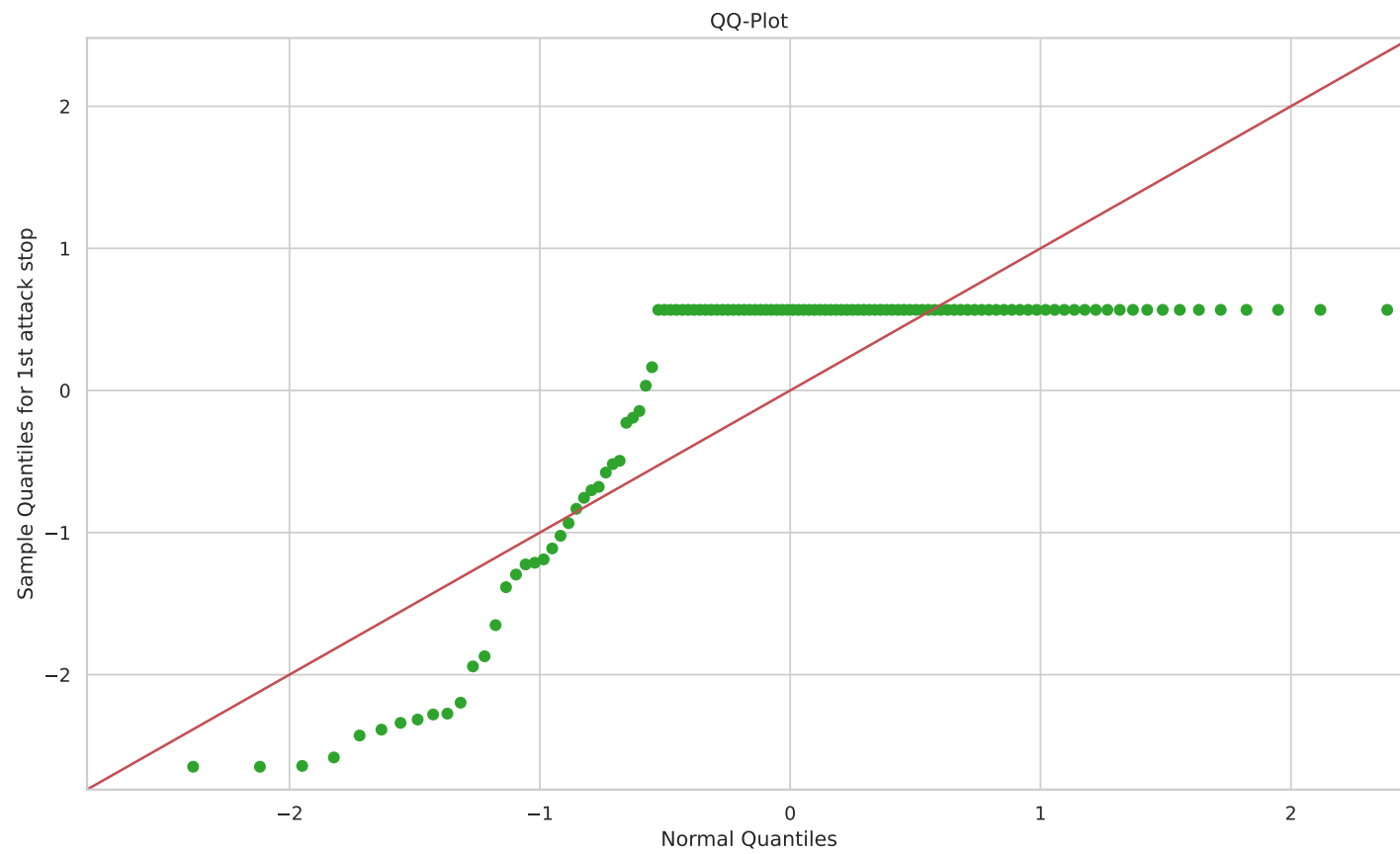


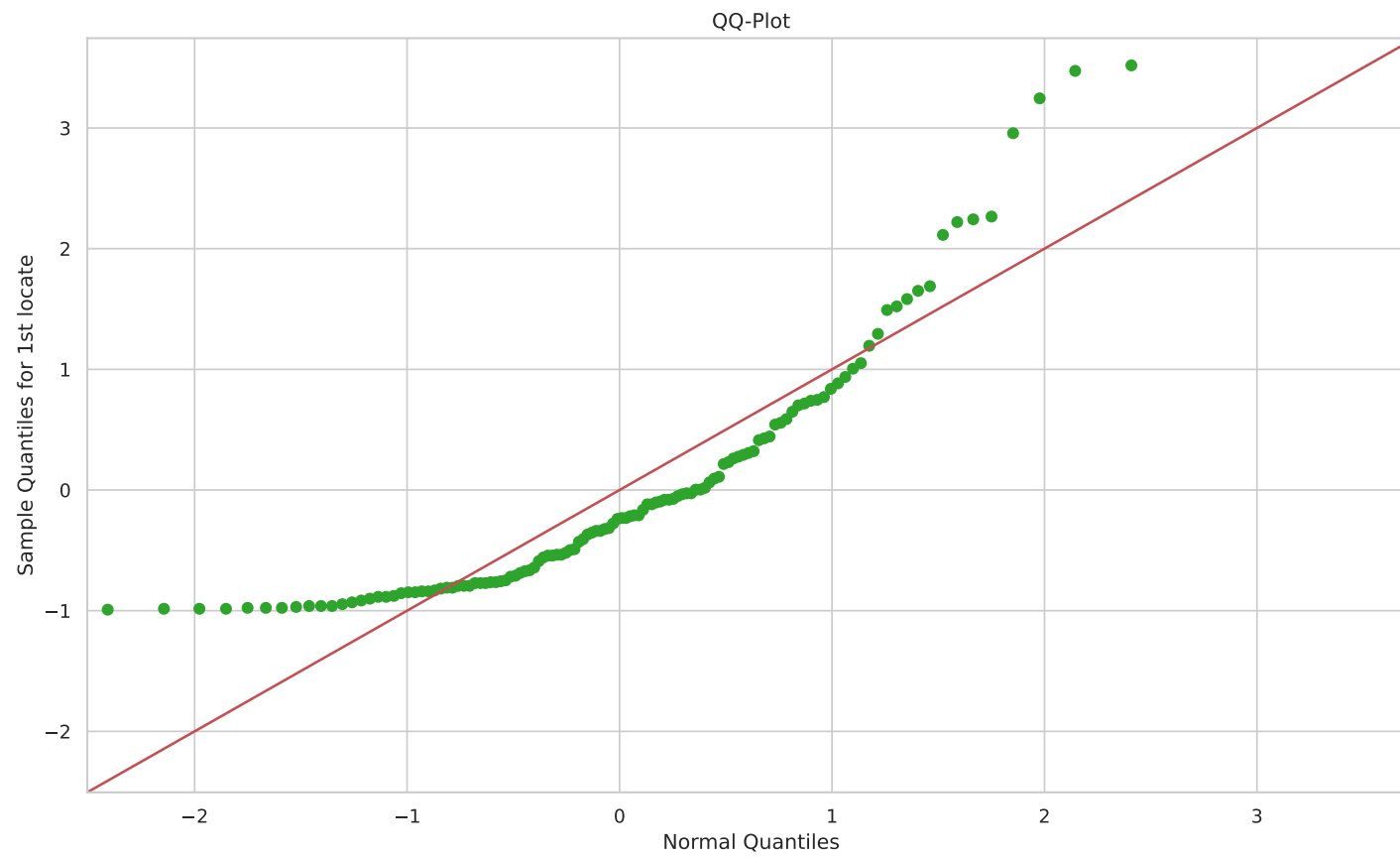
QQ-Plots

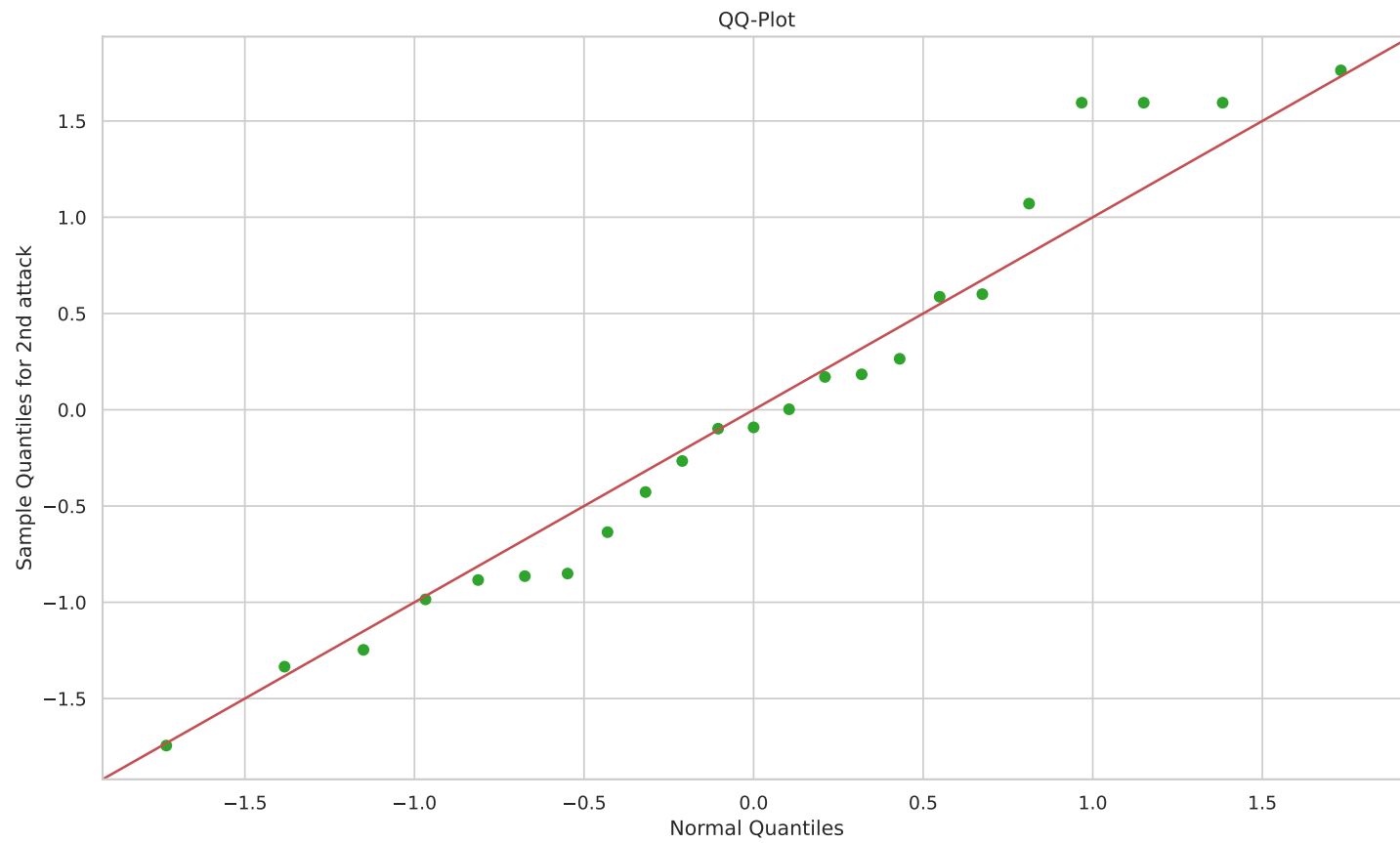
One QQ-Plot per page for each variable. Variables are sorted alphabetically.

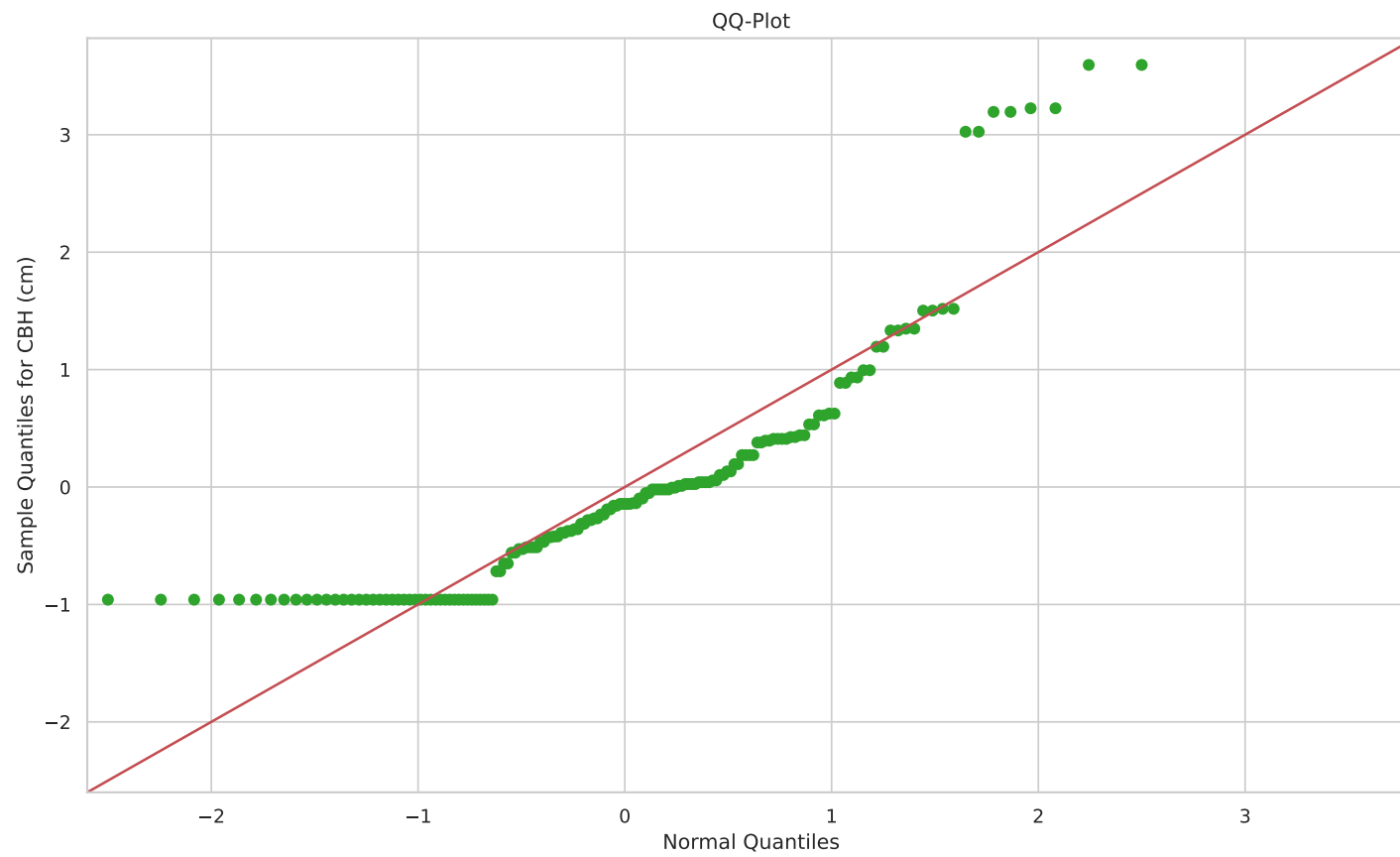
□

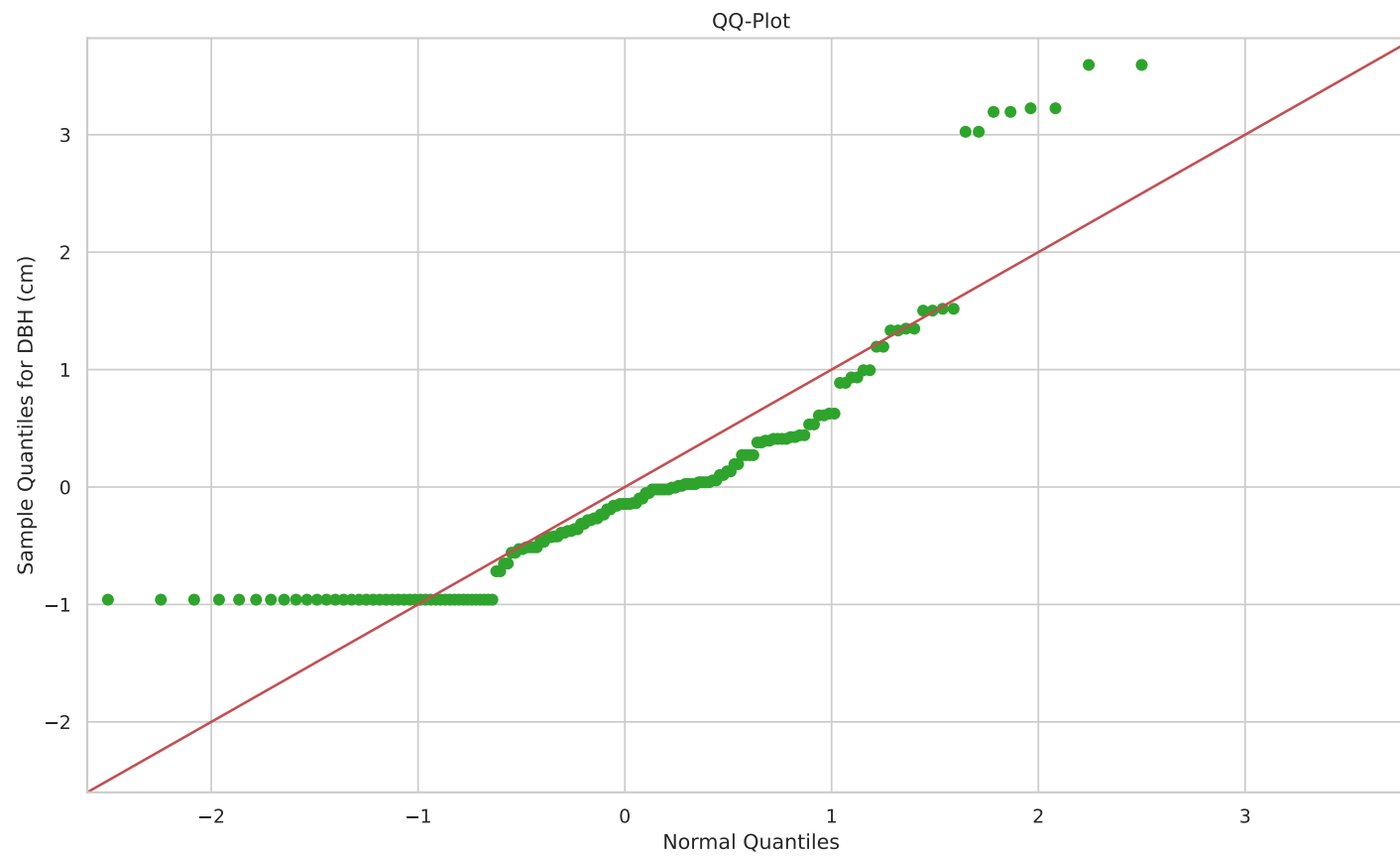


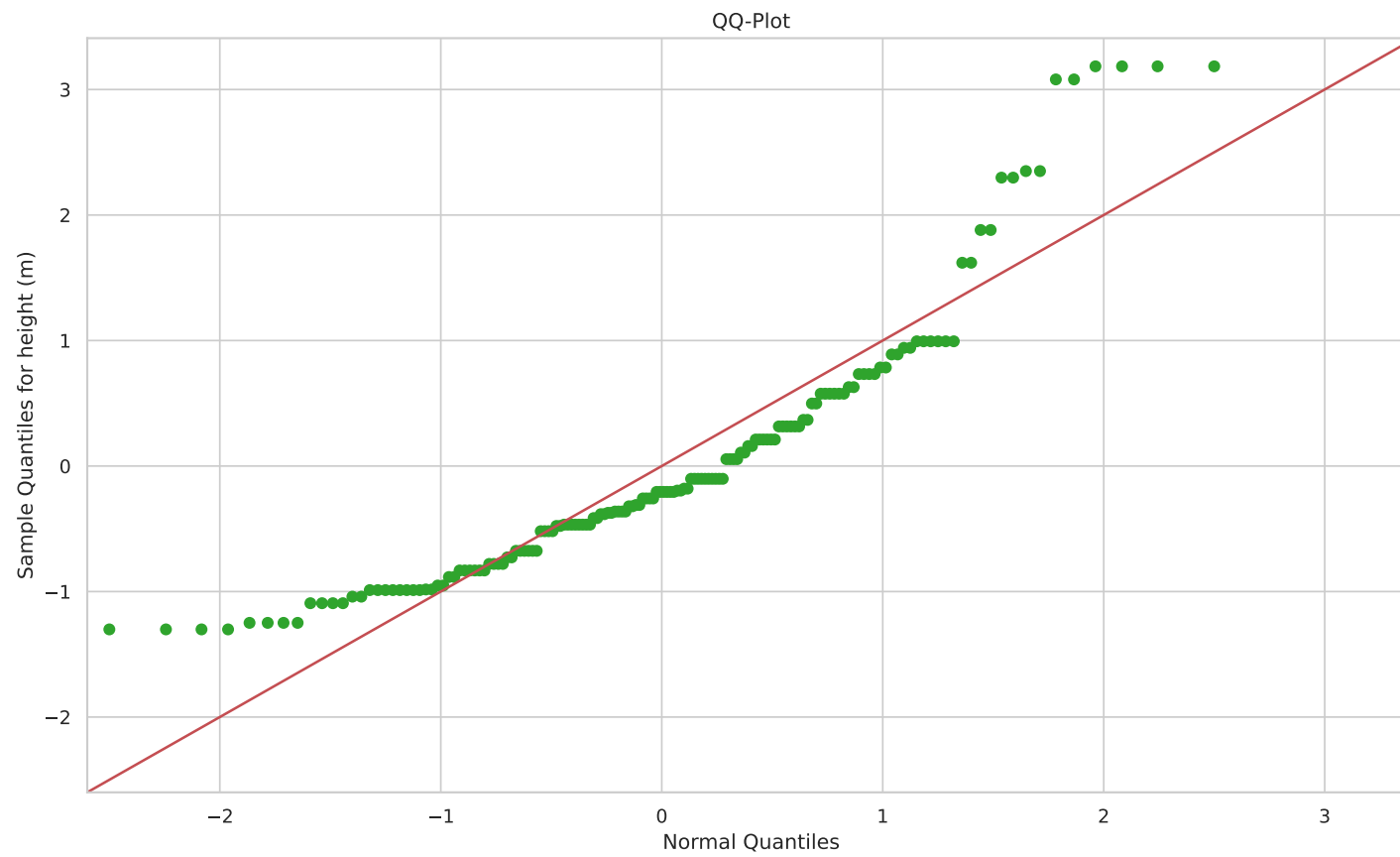






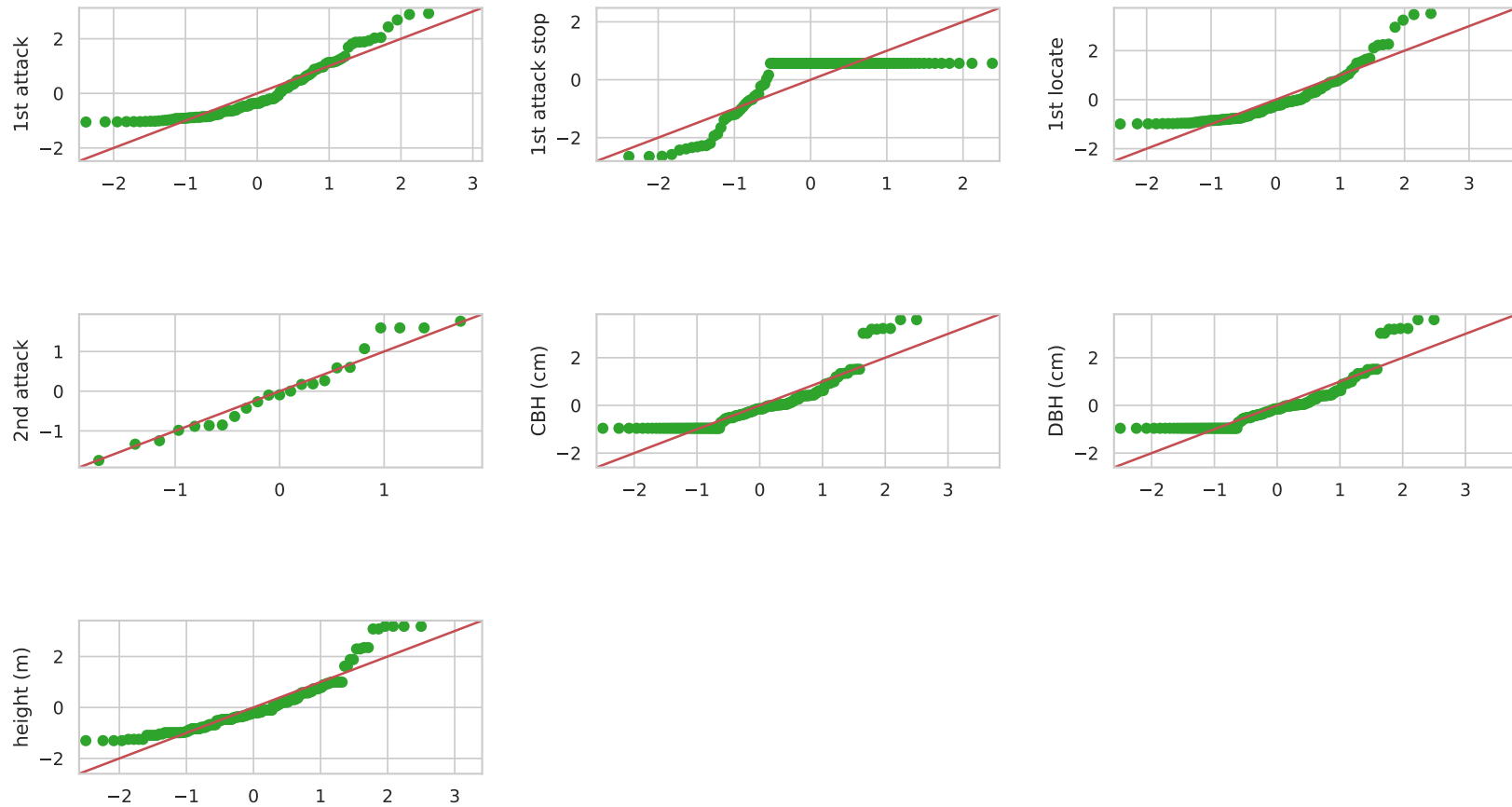






QQ-Plots Summary

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



Results for Discrete Variables

Descriptive Statistics

Totals

The table is sorted by the variable name. If any, N Unique contains the missing category.

	N Obs	N Missing	N Valid	% Complete	N Unique
2nd attack stop	160	137	23	14.37	11
ant sample	160	100	60	37.5	2
Attacked	160	0	160	100	2
Baiting tree no.	160	0	160	100	80
date	160	0	160	100	29
Detected	160	0	160	100	2
elevation (m)	160	0	160	100	8
field notes	160	152	8	5	6
H: 0	160	0	160	100	10
H: 1-5%	160	0	160	100	32
H: 33+%	160	0	160	100	39
H: 5-33%	160	0	160	100	41
lab notes	160	145	15	9.38	9
Recruited	160	0	160	100	2
species	160	0	160	100	5
Termite/C	160	0	160	100	2
transect	160	0	160	100	3
Tree number	160	60	100	62.5	51

Frequencies

The table is sorted by the variable name. For each variable, a maximum of 20 unique values are considered, sorted in decreasing order of their frequency. If any, missings are counted as a category.

Variable	Category	Frequency	Percent
2nd attack stop	Missing	137	0.85625
2nd attack stop	600.0	14	0.0875
2nd attack stop	242.0	1	0.00625
2nd attack stop	501.0	1	0.00625
2nd attack stop	307.0	1	0.00625
2nd attack stop	401.0	1	0.00625
2nd attack stop	492.0	1	0.00625
2nd attack stop	263.0	1	0.00625
2nd attack stop	512.0	1	0.00625
2nd attack stop	370.0	1	0.00625
2nd attack stop	474.0	1	0.00625
ant sample	Missing	100	0.625
ant sample	Y	60	0.375
Attacked	1	116	0.725
Attacked	0	44	0.275
Baiting tree no.	13	2	0.0125
Baiting tree no.	B1004	2	0.0125
Baiting tree no.	B1001	2	0.0125
Baiting tree no.	79	2	0.0125
Baiting tree no.	863	2	0.0125
Baiting tree no.	946	2	0.0125
Baiting tree no.	858	2	0.0125
Baiting tree no.	251	2	0.0125
Baiting tree no.	949	2	0.0125
Baiting tree no.	B0703	2	0.0125
Baiting tree no.	B0801	2	0.0125
Baiting tree no.	B0704	2	0.0125
Baiting tree no.	B01000	2	0.0125
Baiting tree no.	918	2	0.0125
Baiting tree no.	157	2	0.0125
Baiting tree no.	32	2	0.0125
Baiting tree no.	628	2	0.0125
Baiting tree no.	905	2	0.0125
Baiting tree no.	B0803	2	0.0125
Baiting tree no.	B0901	2	0.0125
date	21.08.13	20	0.125
date	14.08.13	20	0.125

Variable	Category	Frequency	Percent
date	17.08.13	20	0.125
date	19.08.13	20	0.125
date	20.08.13	20	0.125
date	16.08.13	18	0.1125
date	15.08.13	18	0.1125
date	26.08.13	2	0.0125
date	24.08.13	2	0.0125
date	22.08.26	1	0.00625
date	22.08.18	1	0.00625
date	22.08.24	1	0.00625
date	22.08.29	1	0.00625
date	22.08.23	1	0.00625
date	22.08.22	1	0.00625
date	22.08.13	1	0.00625
date	22.08.32	1	0.00625
date	22.08.28	1	0.00625
date	22.08.25	1	0.00625
date	22.08.17	1	0.00625
Detected	1	124	0.775
Detected	0	36	0.225
elevation (m)	900	20	0.125
elevation (m)	1400	20	0.125
elevation (m)	1000	20	0.125
elevation (m)	1100	20	0.125
elevation (m)	800	20	0.125
elevation (m)	700	20	0.125
elevation (m)	1200	20	0.125
elevation (m)	1300	20	0.125
field notes	Missing	152	0.95
field notes	(jimmy data)	3	0.01875
field notes	*broken	2	0.0125
field notes	*broken, 5 @ respro	1	0.00625
field notes	few ants	1	0.00625
field notes	few ants on lower b	1	0.00625
H: 0	0	84	0.525
H: 0	1	22	0.1375
H: 0	3	12	0.075
H: 0	2	12	0.075
H: 0	NA	10	0.0625
H: 0	5	8	0.05
H: 0	8	6	0.0375

Variable	Category	Frequency	Percent
H: 0	4	2	0.0125
H: 0	7	2	0.0125
H: 0	12	2	0.0125
H: 1-5%	0	22	0.1375
H: 1-5%	1	18	0.1125
H: 1-5%	3	14	0.0875
H: 1-5%	4	12	0.075
H: 1-5%	10	10	0.0625
H: 1-5%	NA	10	0.0625
H: 1-5%	2	8	0.05
H: 1-5%	19	4	0.025
H: 1-5%	20	4	0.025
H: 1-5%	14	4	0.025
H: 1-5%	22	4	0.025
H: 1-5%	5	4	0.025
H: 1-5%	33	4	0.025
H: 1-5%	23	4	0.025
H: 1-5%	26	4	0.025
H: 1-5%	41	2	0.0125
H: 1-5%	21	2	0.0125
H: 1-5%	28	2	0.0125
H: 1-5%	12	2	0.0125
H: 1-5%	65	2	0.0125
H: 33+%	8	10	0.0625
H: 33+%	NA	10	0.0625
H: 33+%	13	8	0.05
H: 33+%	17	8	0.05
H: 33+%	9	6	0.0375
H: 33+%	24	6	0.0375
H: 33+%	10	6	0.0375
H: 33+%	5	6	0.0375
H: 33+%	4	6	0.0375
H: 33+%	12	6	0.0375
H: 33+%	20	6	0.0375
H: 33+%	7	6	0.0375
H: 33+%	16	6	0.0375
H: 33+%	11	6	0.0375
H: 33+%	35	4	0.025
H: 33+%	26	4	0.025
H: 33+%	36	4	0.025
H: 33+%	1	4	0.025

Variable	Category	Frequency	Percent
H: 33+%	19	4	0.025
H: 33+%	14	4	0.025
H: 5-33%	12	10	0.0625
H: 5-33%	NA	10	0.0625
H: 5-33%	9	10	0.0625
H: 5-33%	21	8	0.05
H: 5-33%	15	8	0.05
H: 5-33%	8	8	0.05
H: 5-33%	7	8	0.05
H: 5-33%	13	6	0.0375
H: 5-33%	14	6	0.0375
H: 5-33%	24	6	0.0375
H: 5-33%	6	4	0.025
H: 5-33%	19	4	0.025
H: 5-33%	0	4	0.025
H: 5-33%	3	4	0.025
H: 5-33%	39	4	0.025
H: 5-33%	27	4	0.025
H: 5-33%	23	4	0.025
H: 5-33%	52	4	0.025
H: 5-33%	16	4	0.025
H: 5-33%	41	2	0.0125
lab notes	Missing	145	0.90625
lab notes	ant ID from main da	6	0.0375
lab notes	Much bigger than B0	2	0.0125
lab notes	Vial transferred fr	2	0.0125
lab notes	250 misread as 230	2	0.0125
lab notes	Small compared to 8	1	0.00625
lab notes	Bigger than 802	1	0.00625
lab notes	Taller petiole than	1	0.00625
Recruited	1	94	0.5875
Recruited	0	66	0.4125
species	ANON012	68	0.425
species	ANON013	42	0.2625
species	ANON009	40	0.25
species	ANON002	6	0.0375
species	ANON001	4	0.025
Termite/C	T	80	0.5
Termite/C	C	80	0.5
transect	Y	100	0.625
transect	N	58	0.3625

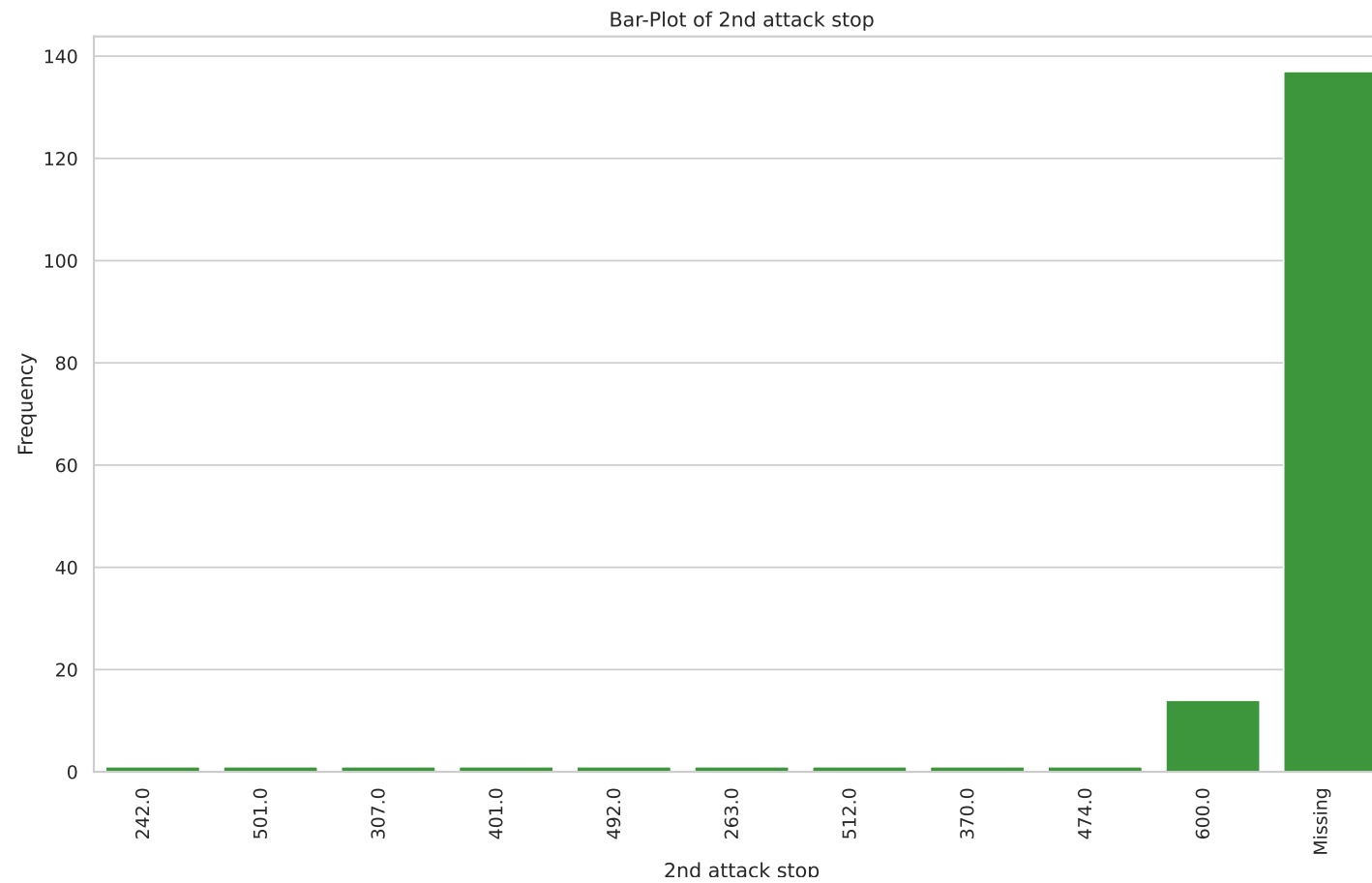
Variable	Category	Frequency	Percent
transect	Y-not m	2	0.0125
Tree number	Missing	60	0.375
Tree number	1300 - 0963	2	0.0125
Tree number	1200 - 0077	2	0.0125
Tree number	1200 - 0898	2	0.0125
Tree number	1400 - 1002	2	0.0125
Tree number	1400 - 0995	2	0.0125
Tree number	1400 - 0013	2	0.0125
Tree number	800 - 0329	2	0.0125
Tree number	1200 - 0918	2	0.0125
Tree number	1100 - 0108	2	0.0125
Tree number	1200 - 0904	2	0.0125
Tree number	1200 - 0902	2	0.0125
Tree number	1400 - 0014	2	0.0125
Tree number	700 - 0250	2	0.0125
Tree number	1400 - 0994	2	0.0125
Tree number	1200 - 0907	2	0.0125
Tree number	1400 - 1001	2	0.0125
Tree number	1400 - 0990	2	0.0125
Tree number	1000 - 0863	2	0.0125
Tree number	1200 - 0943	2	0.0125

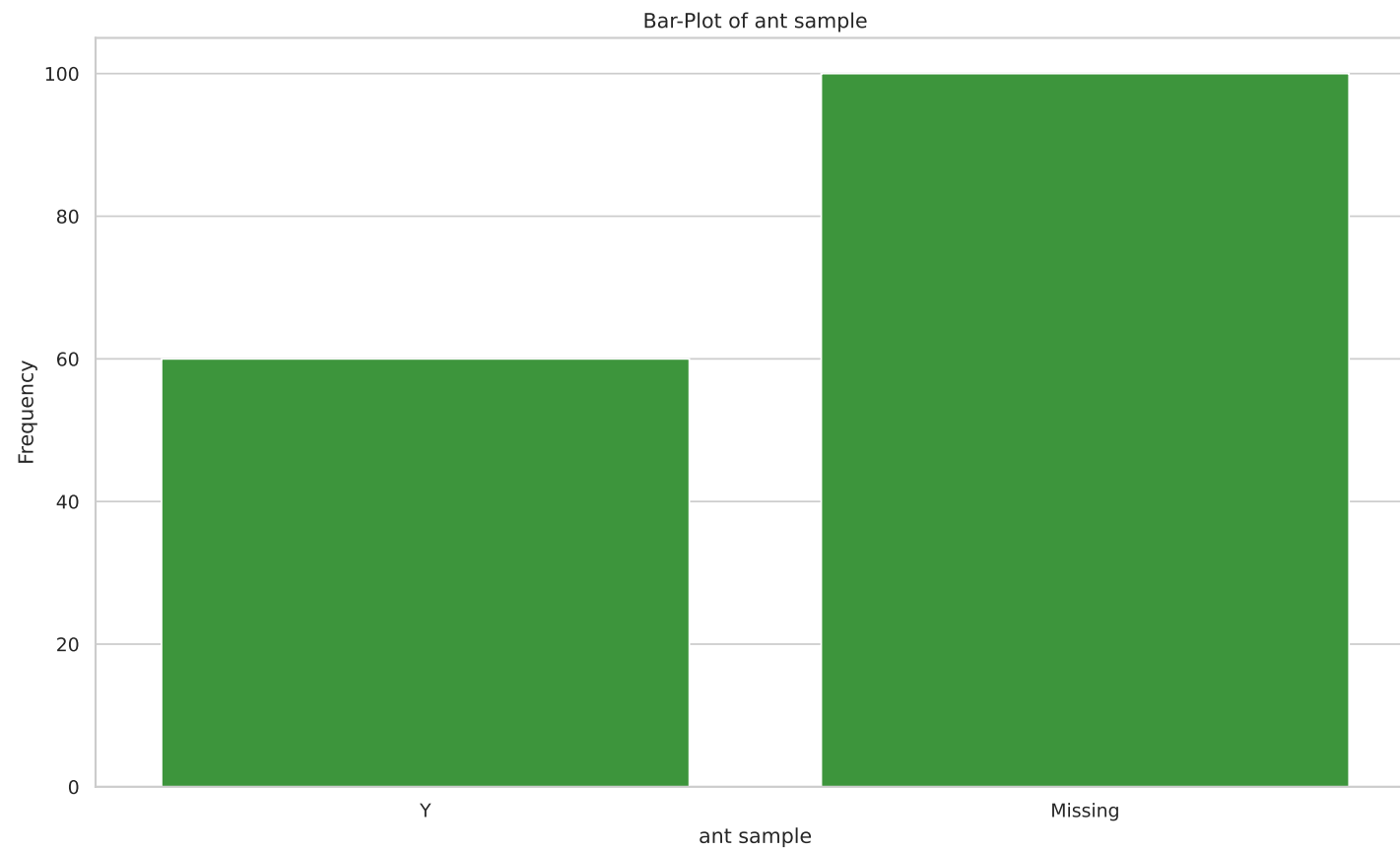
Graphics

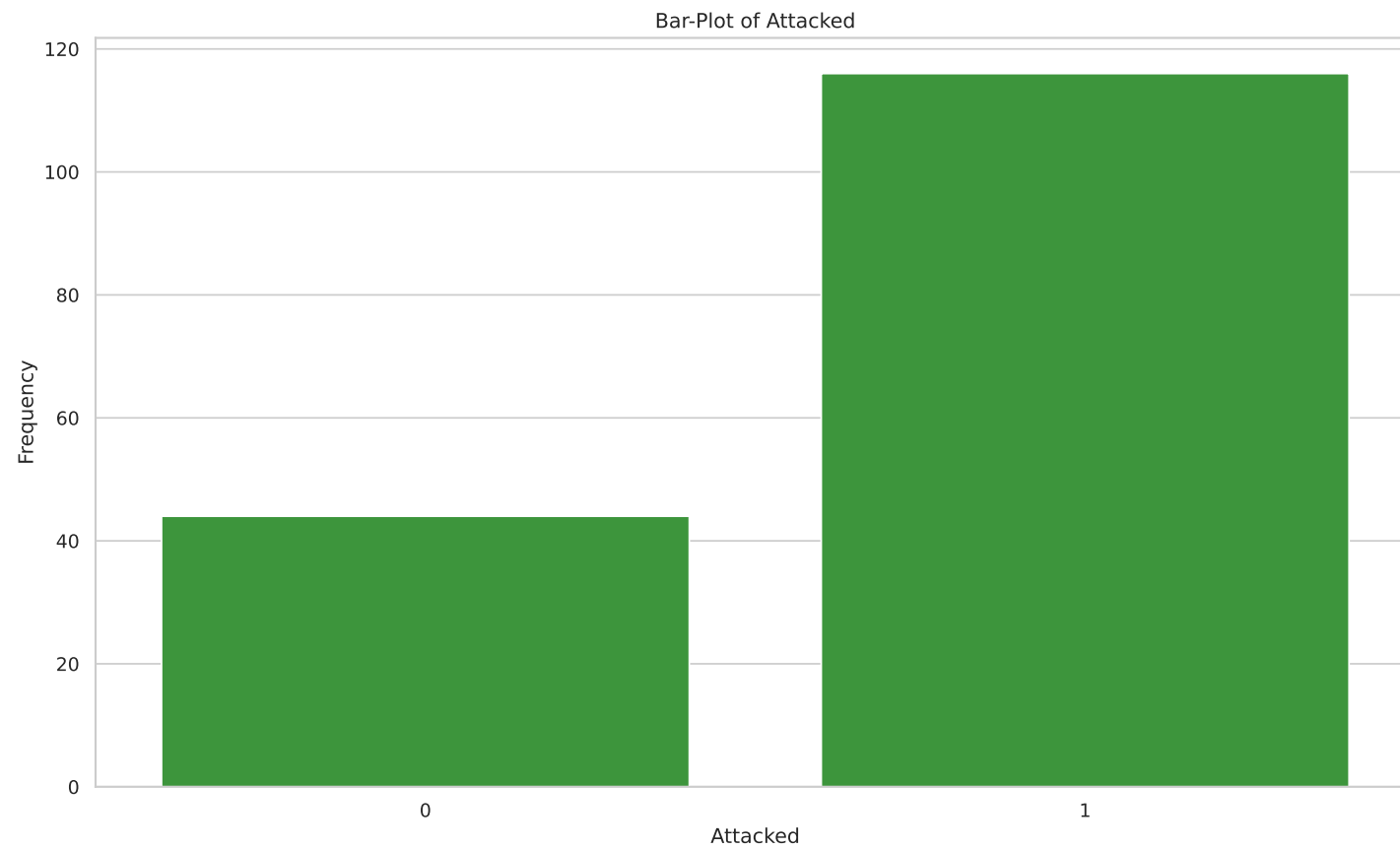
Bar-Plots

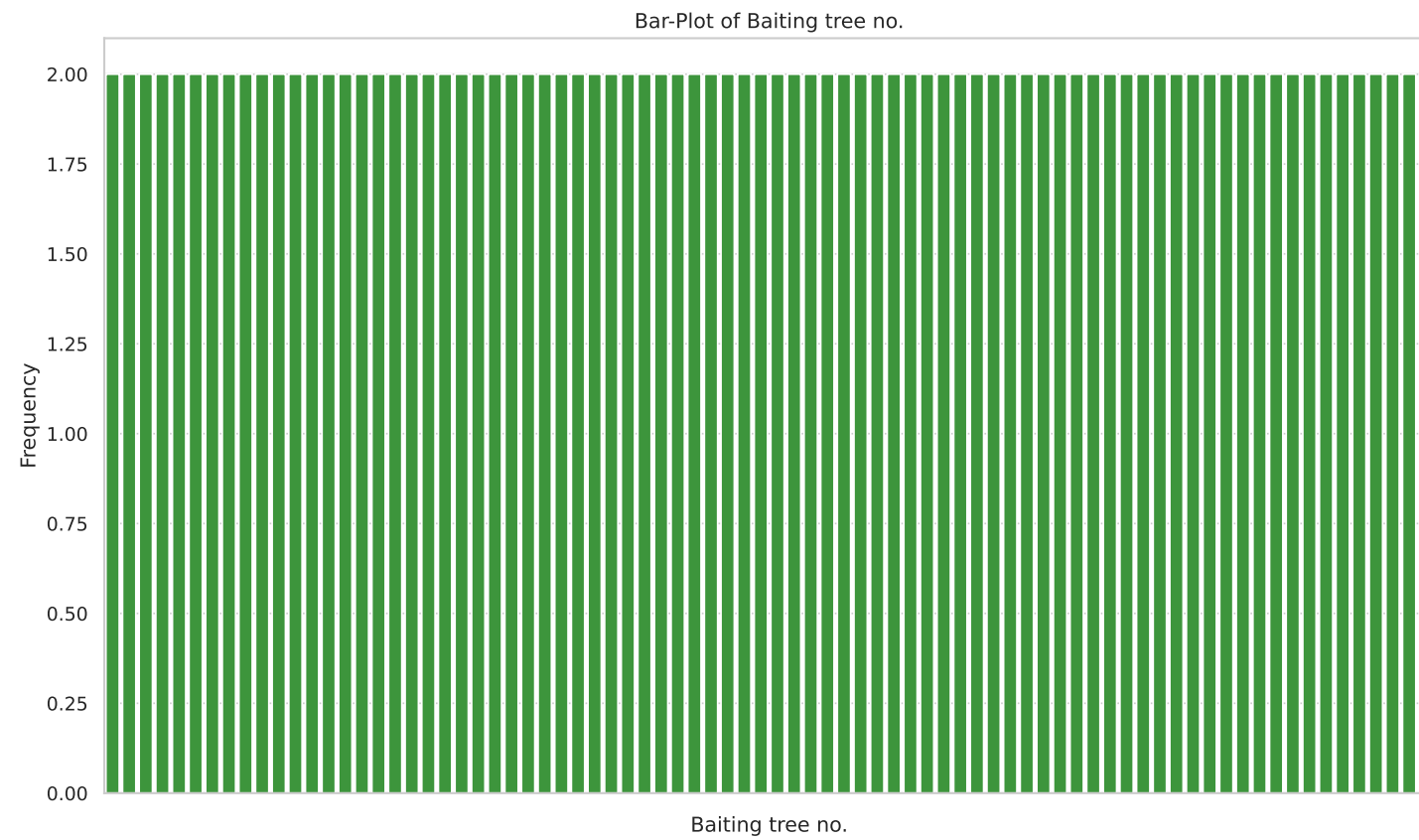
One Bar-Plot per page for each variable. Variables are sorted alphabetically. No labels for variables with more than 40 categories.

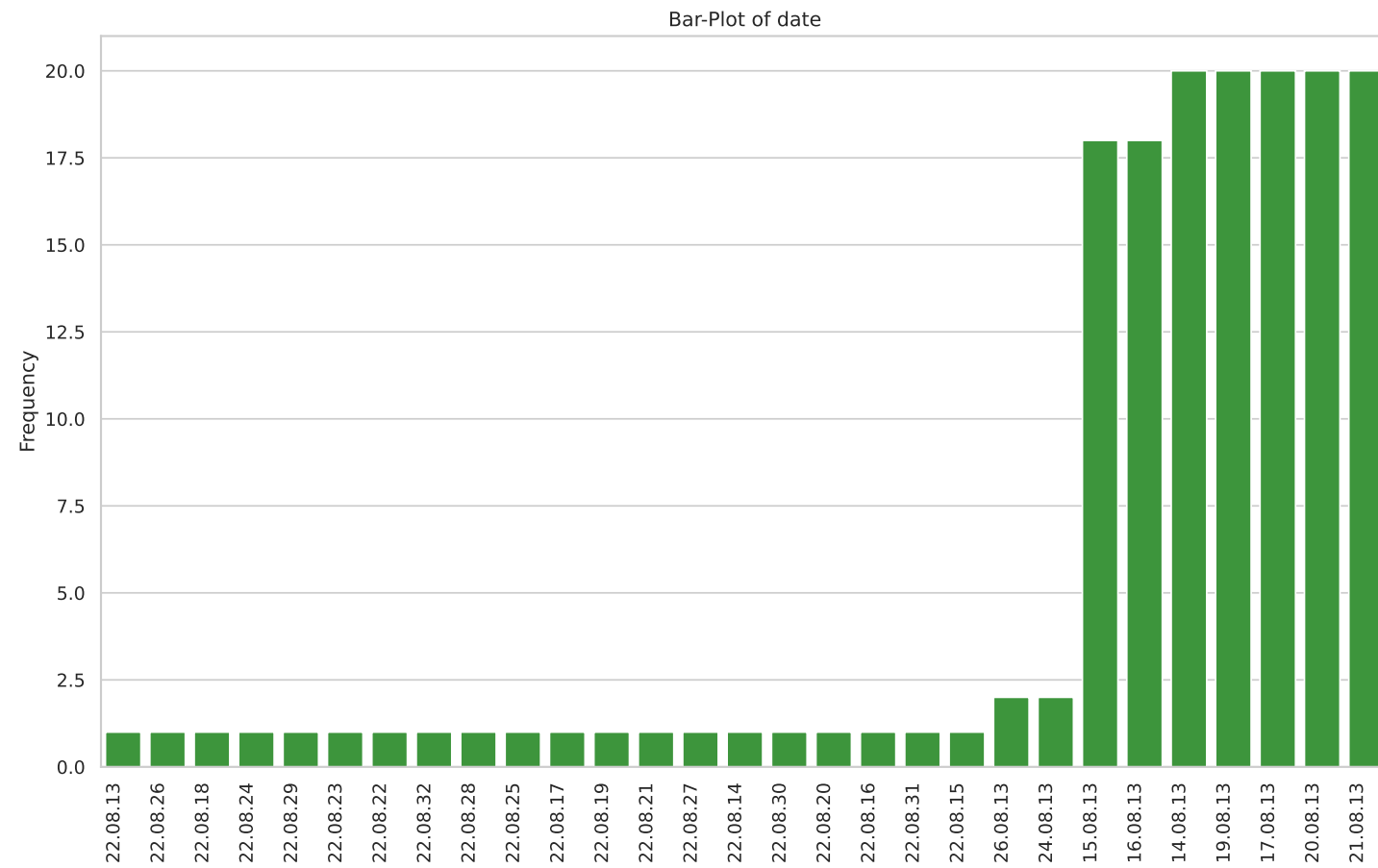
□

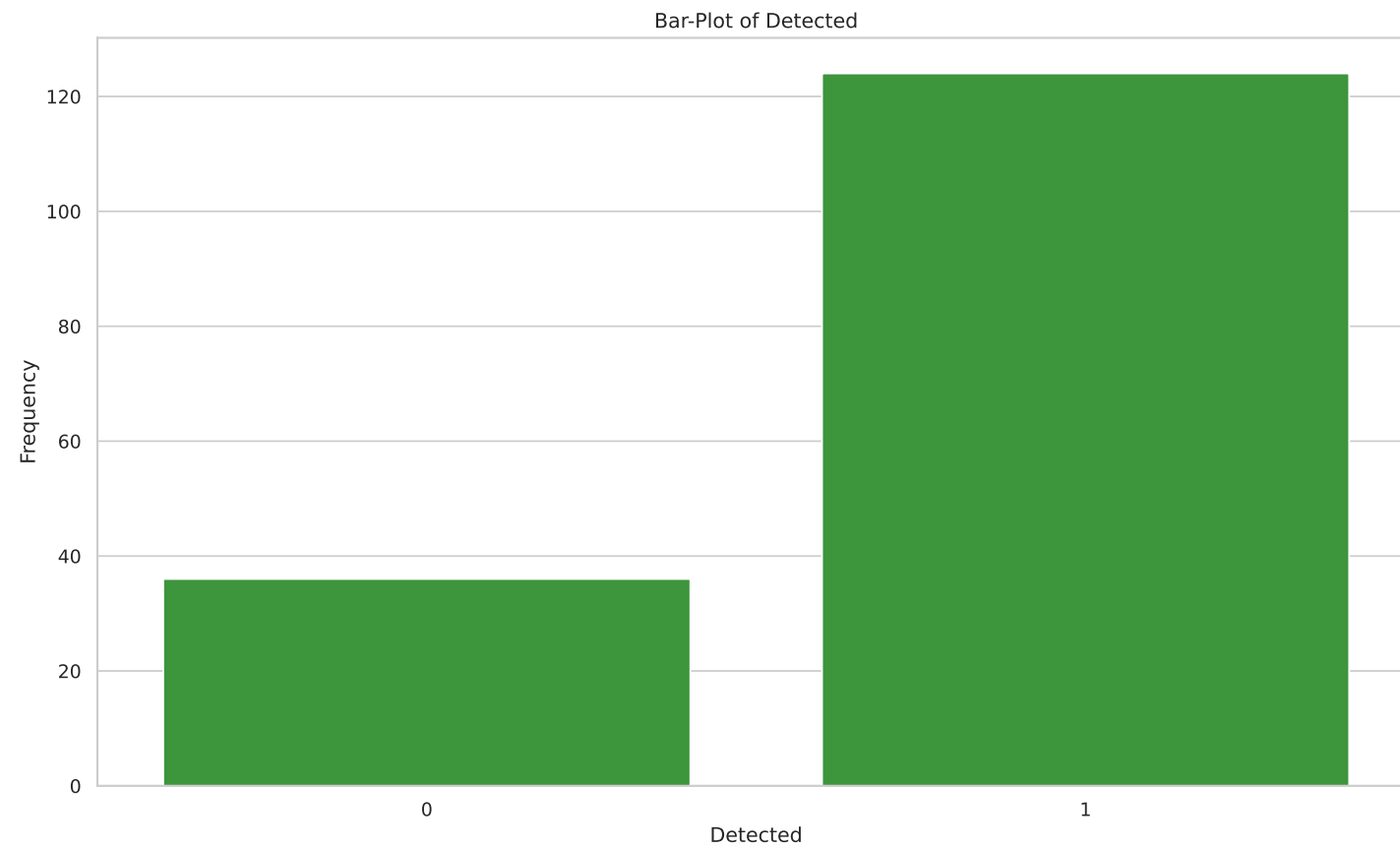


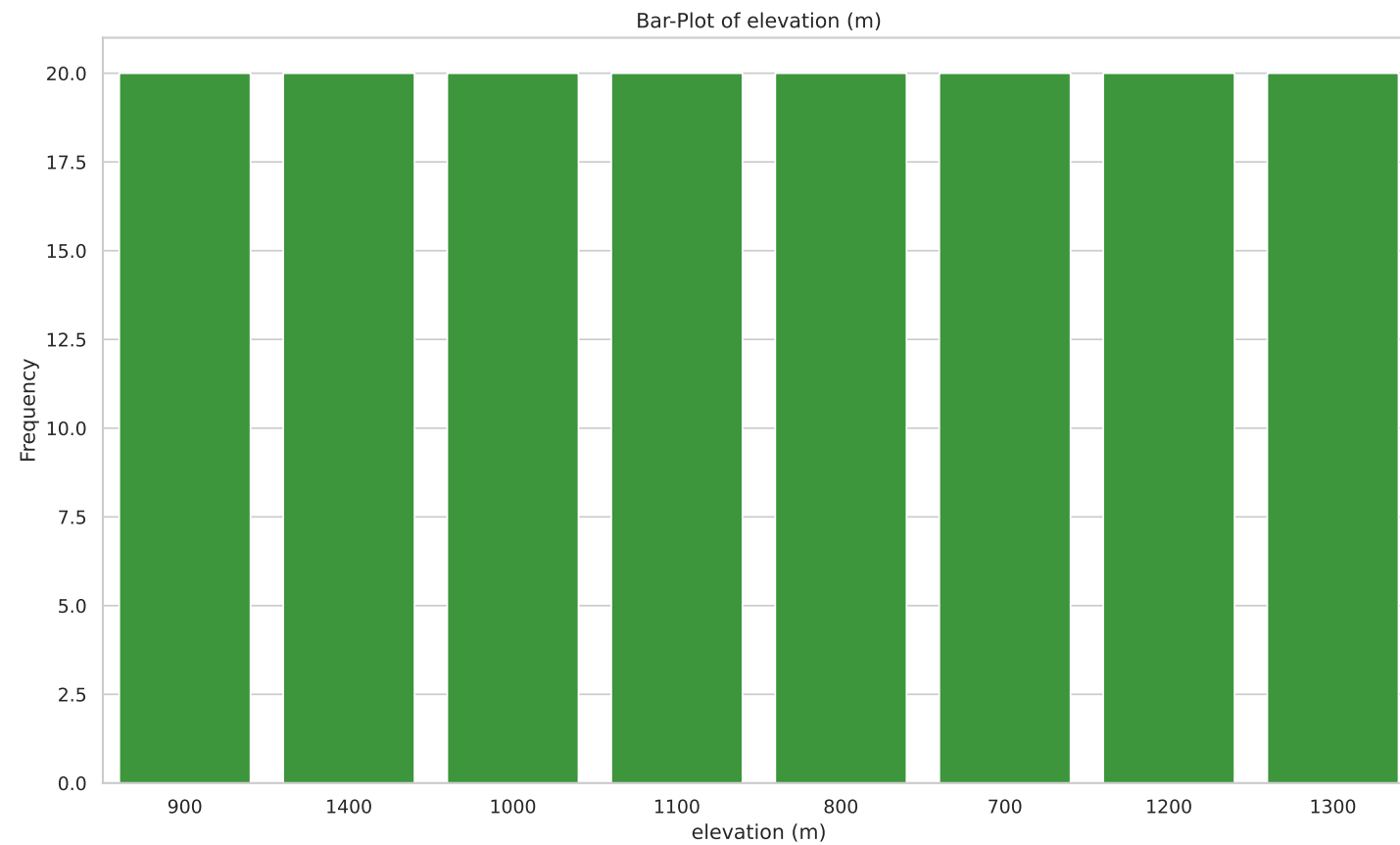


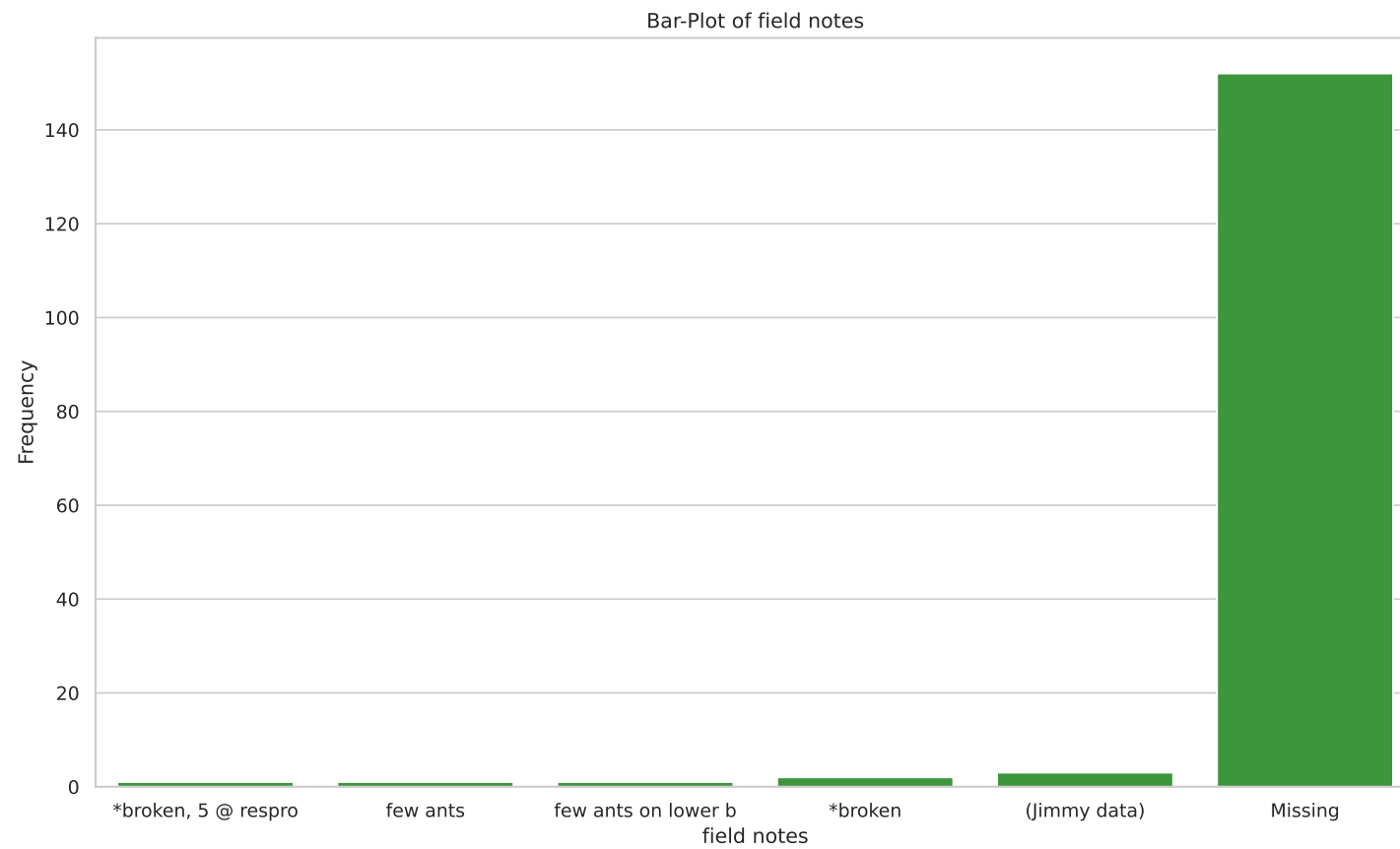


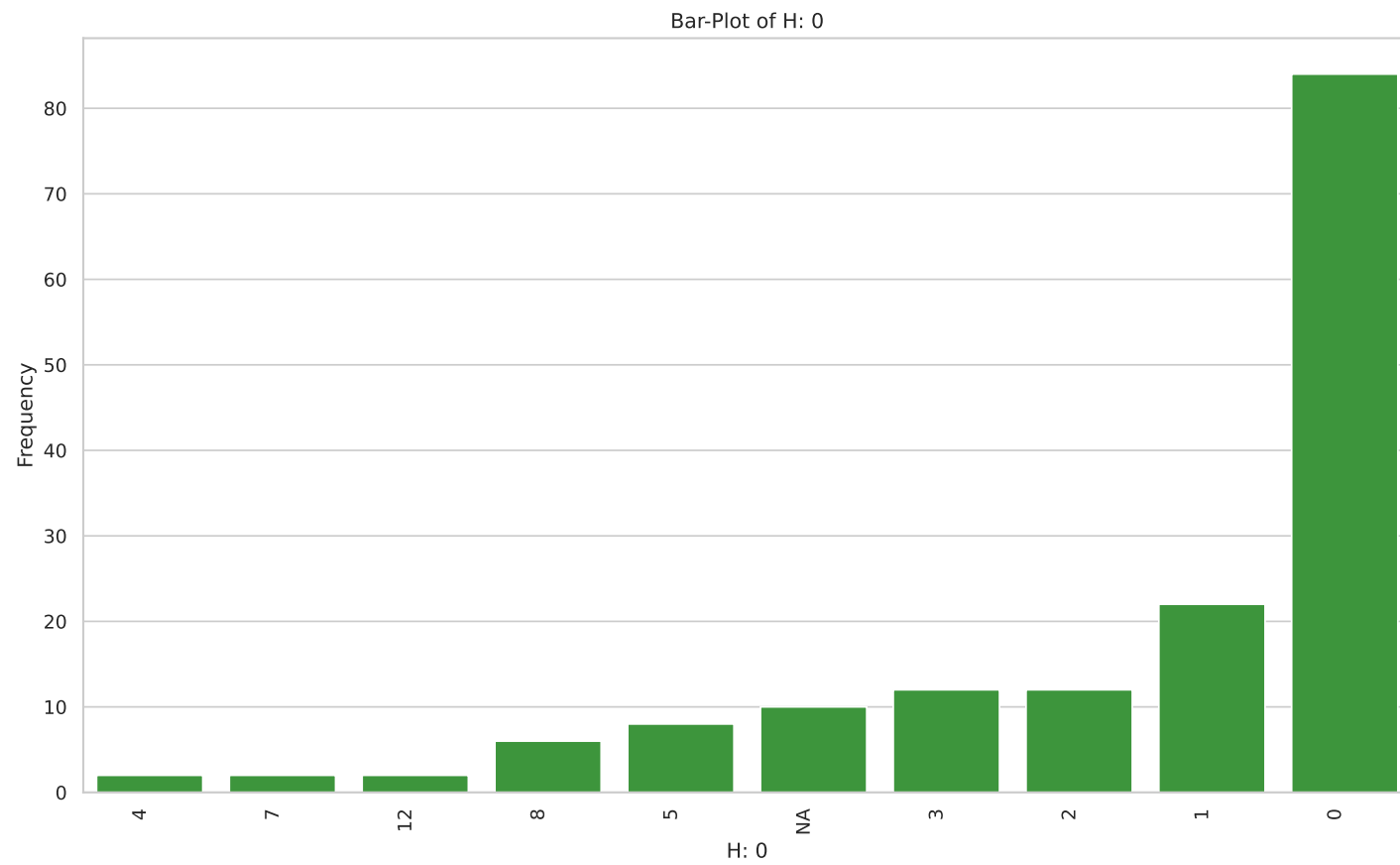


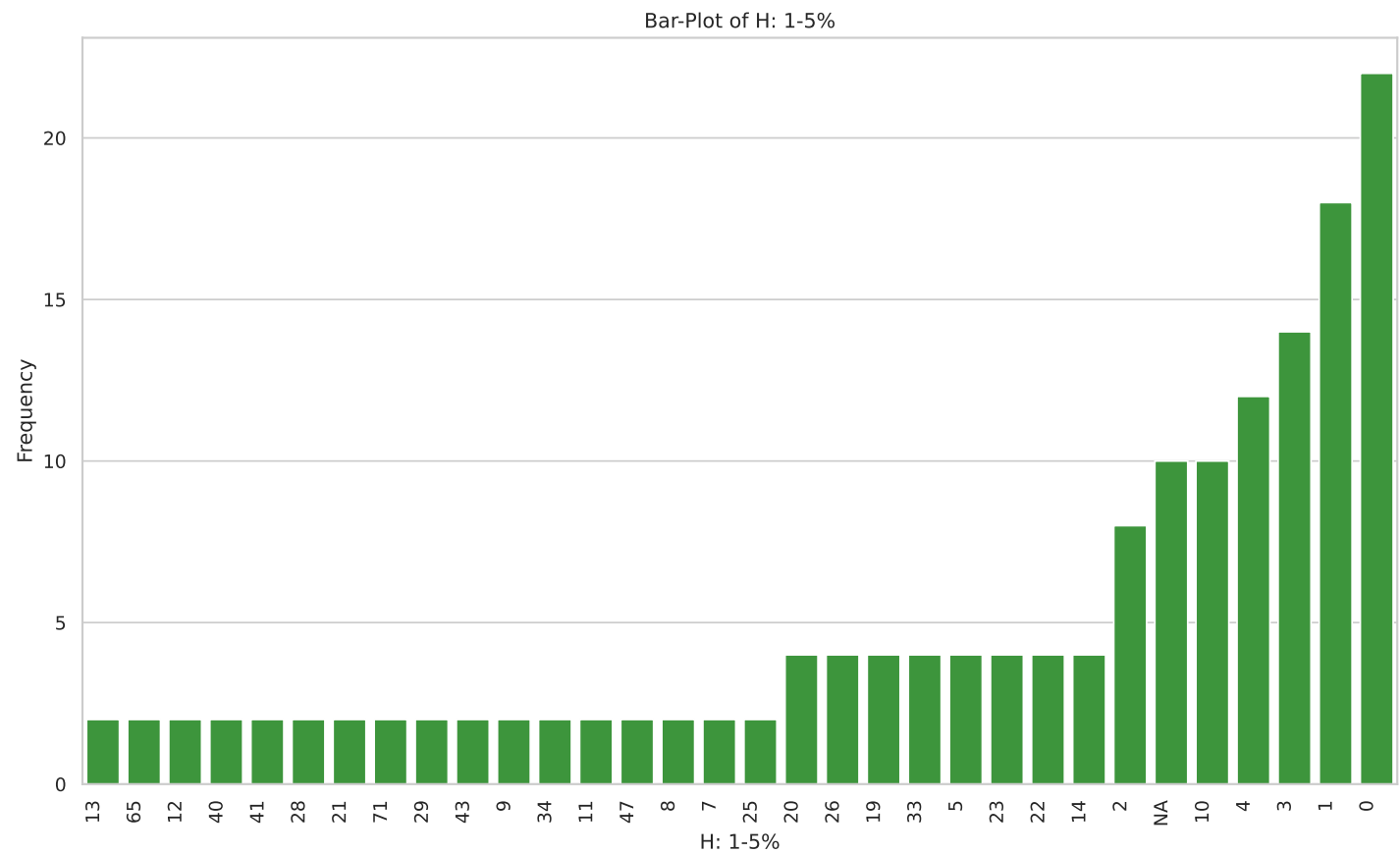


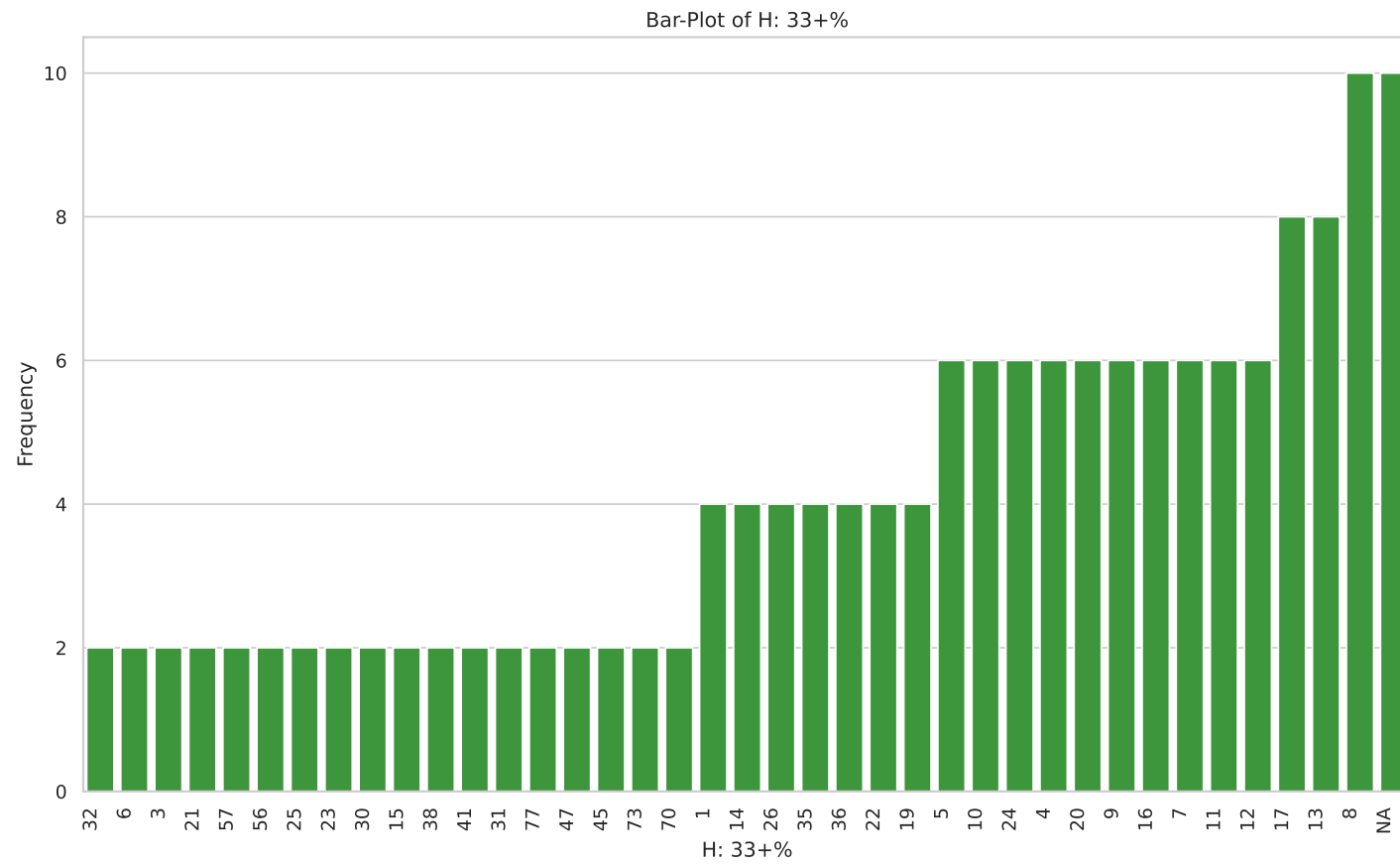


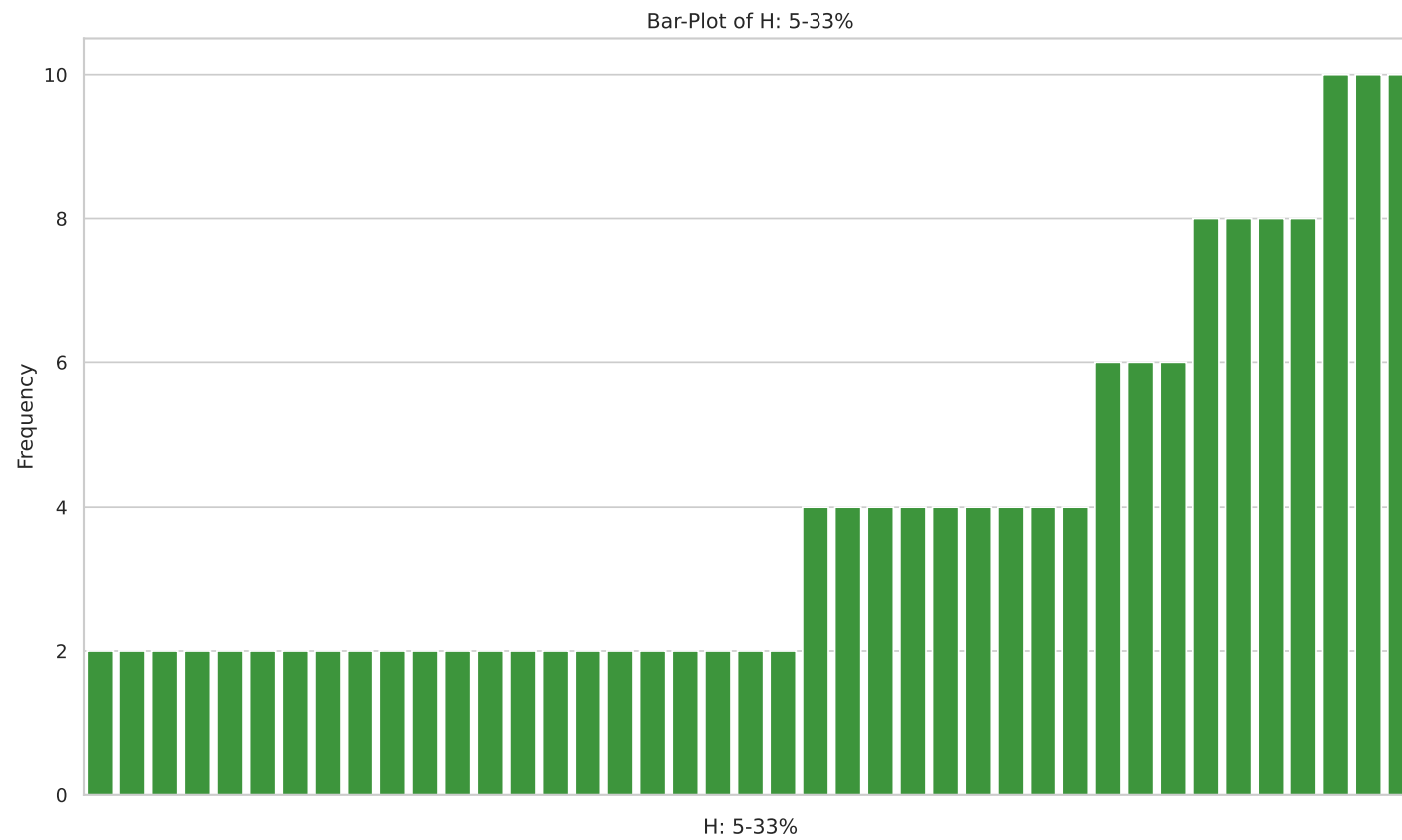


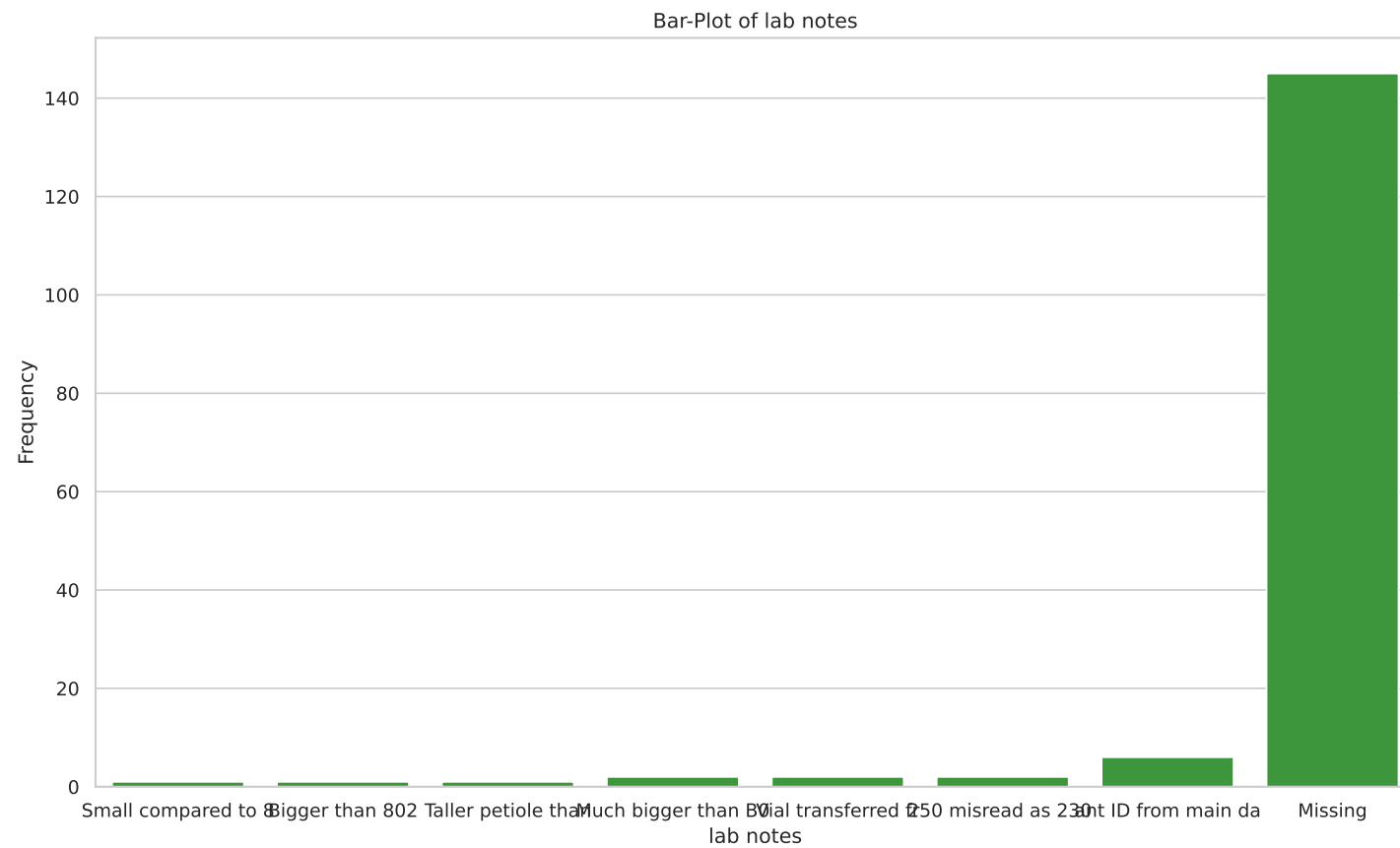


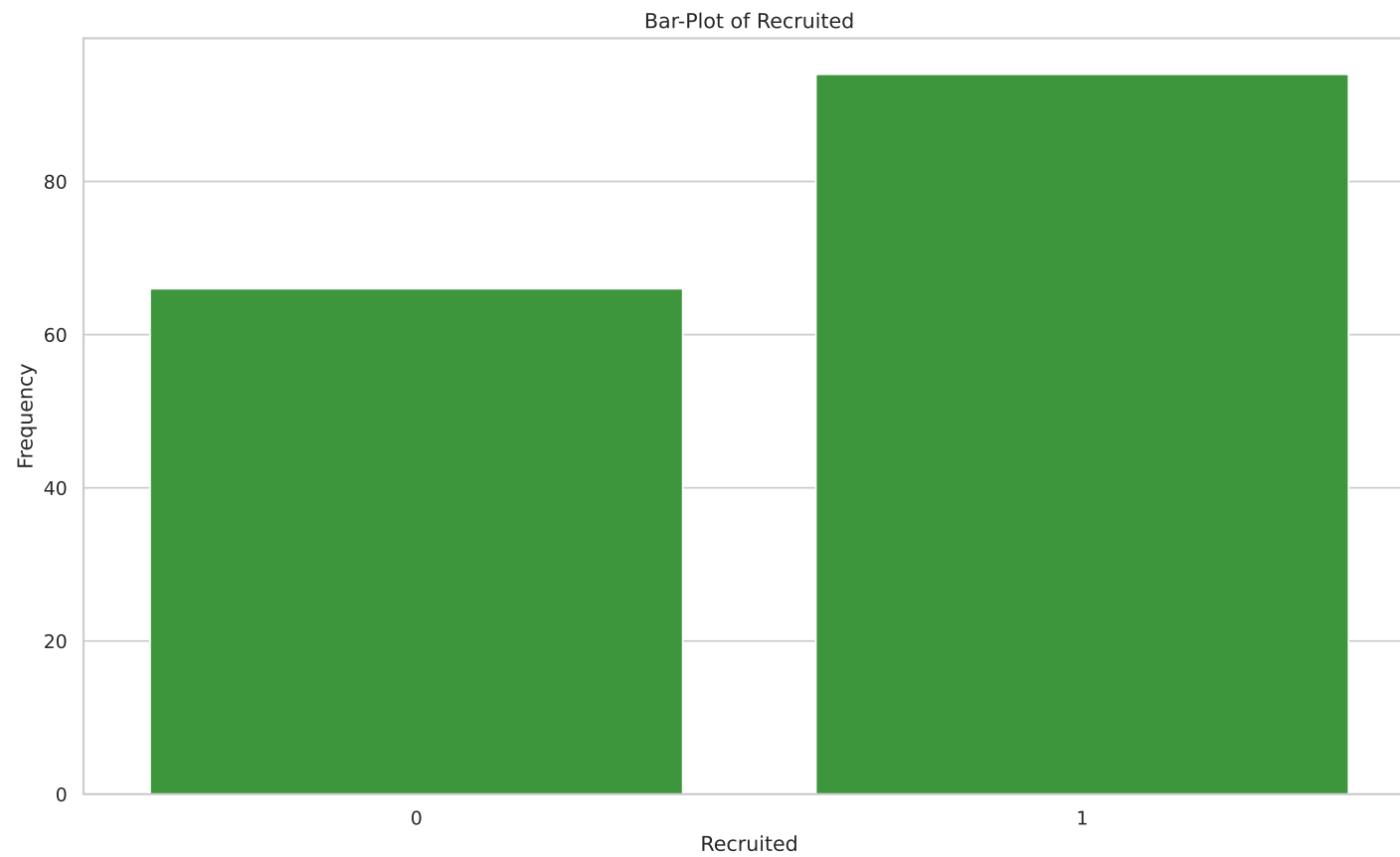


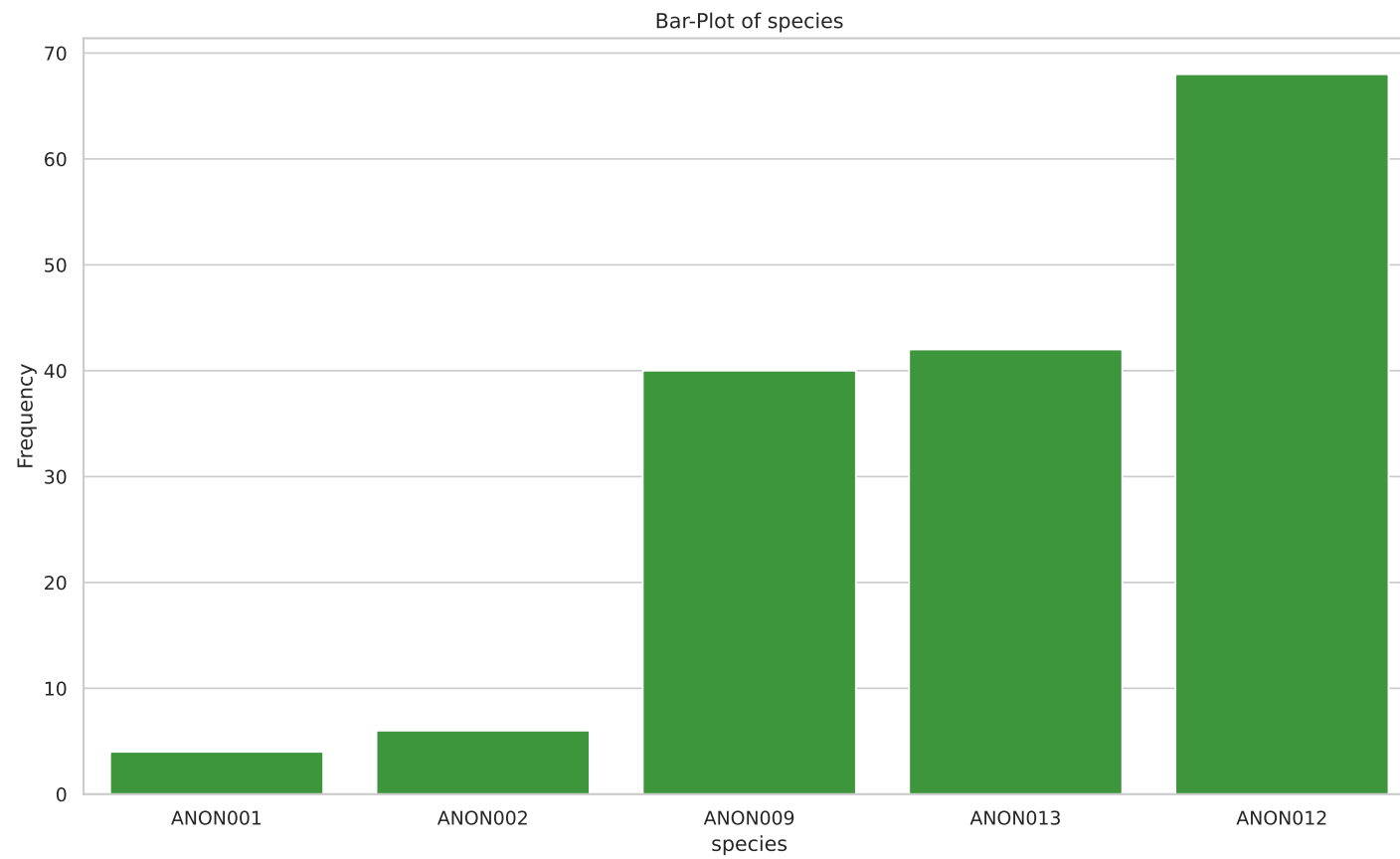


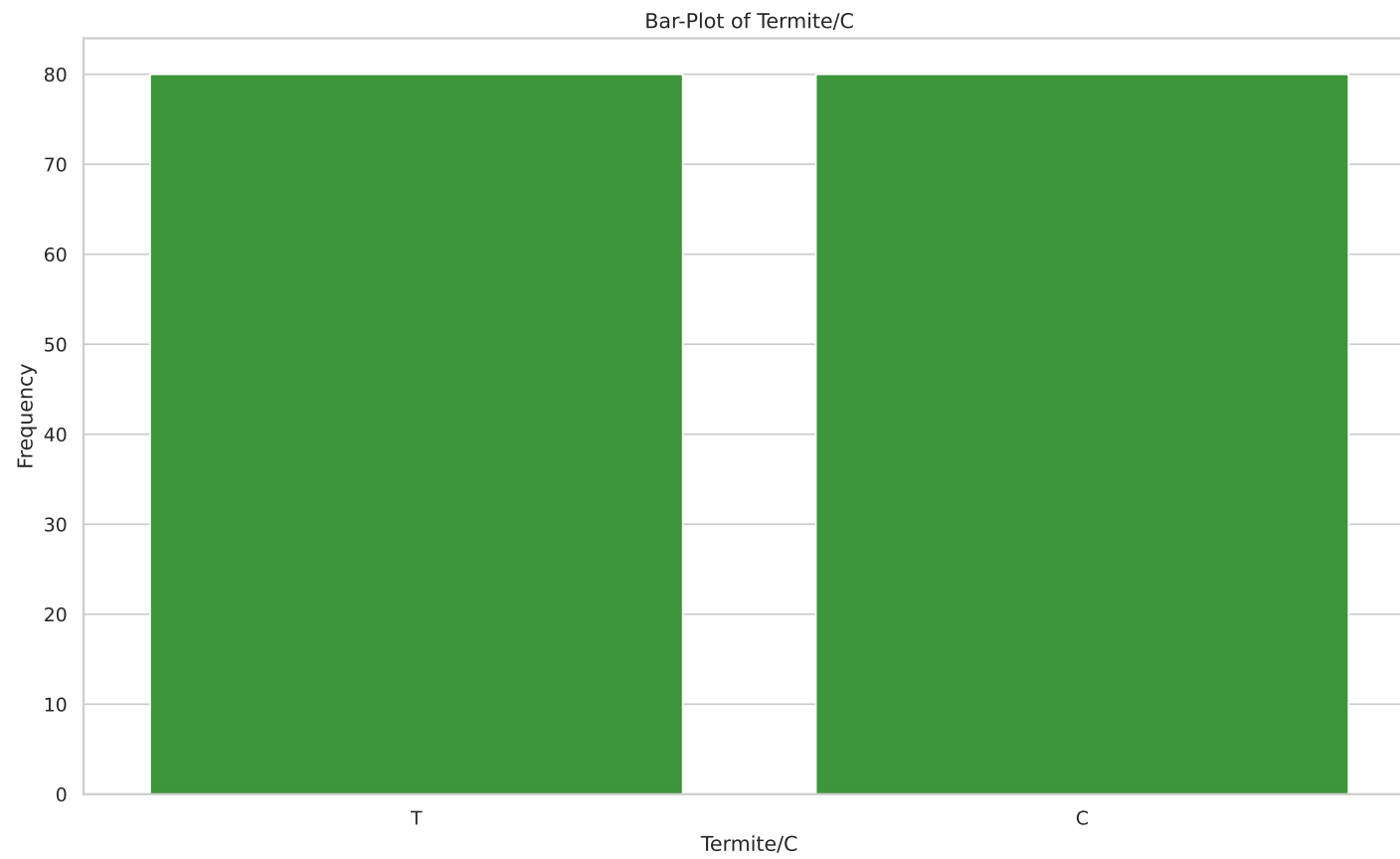


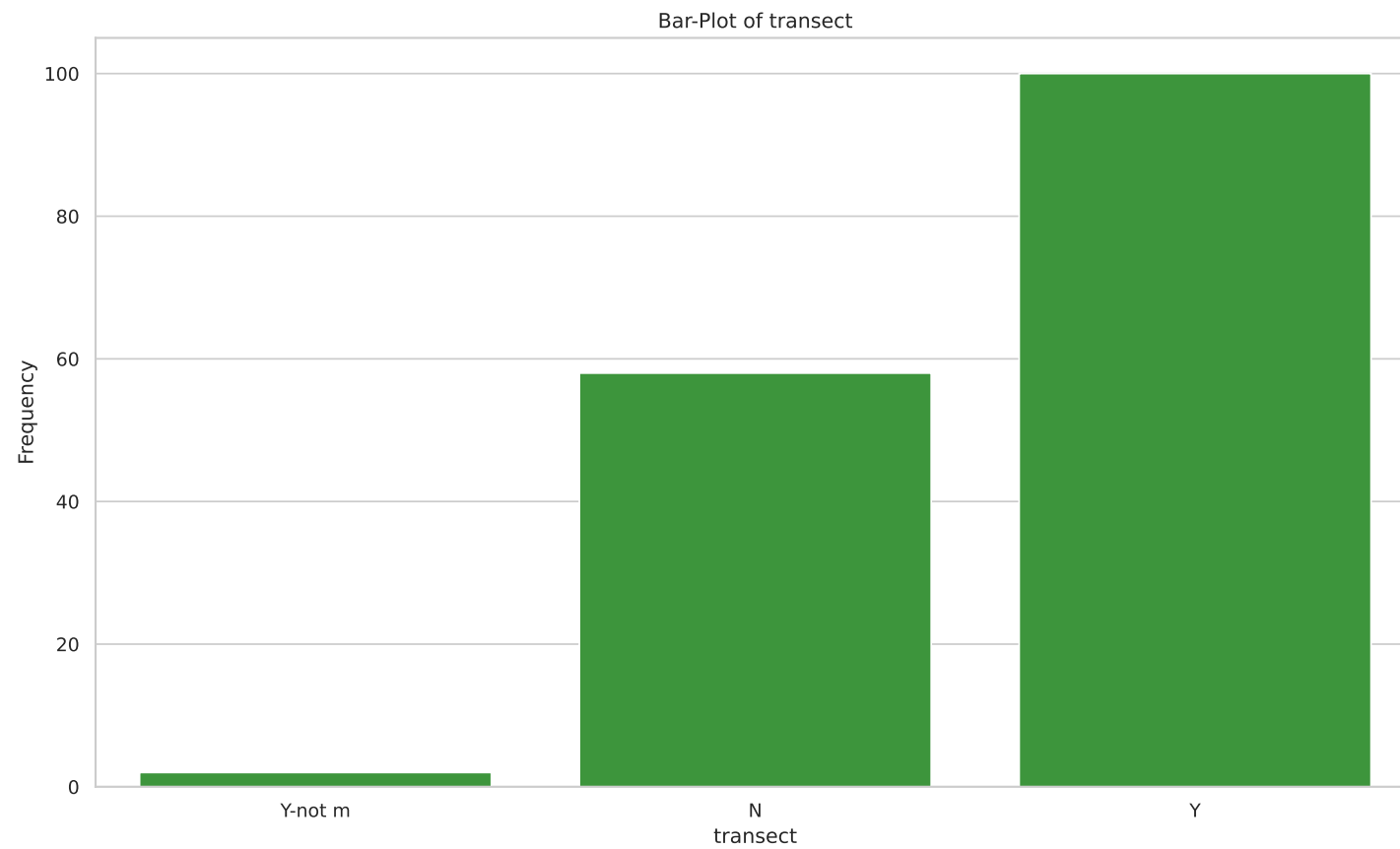


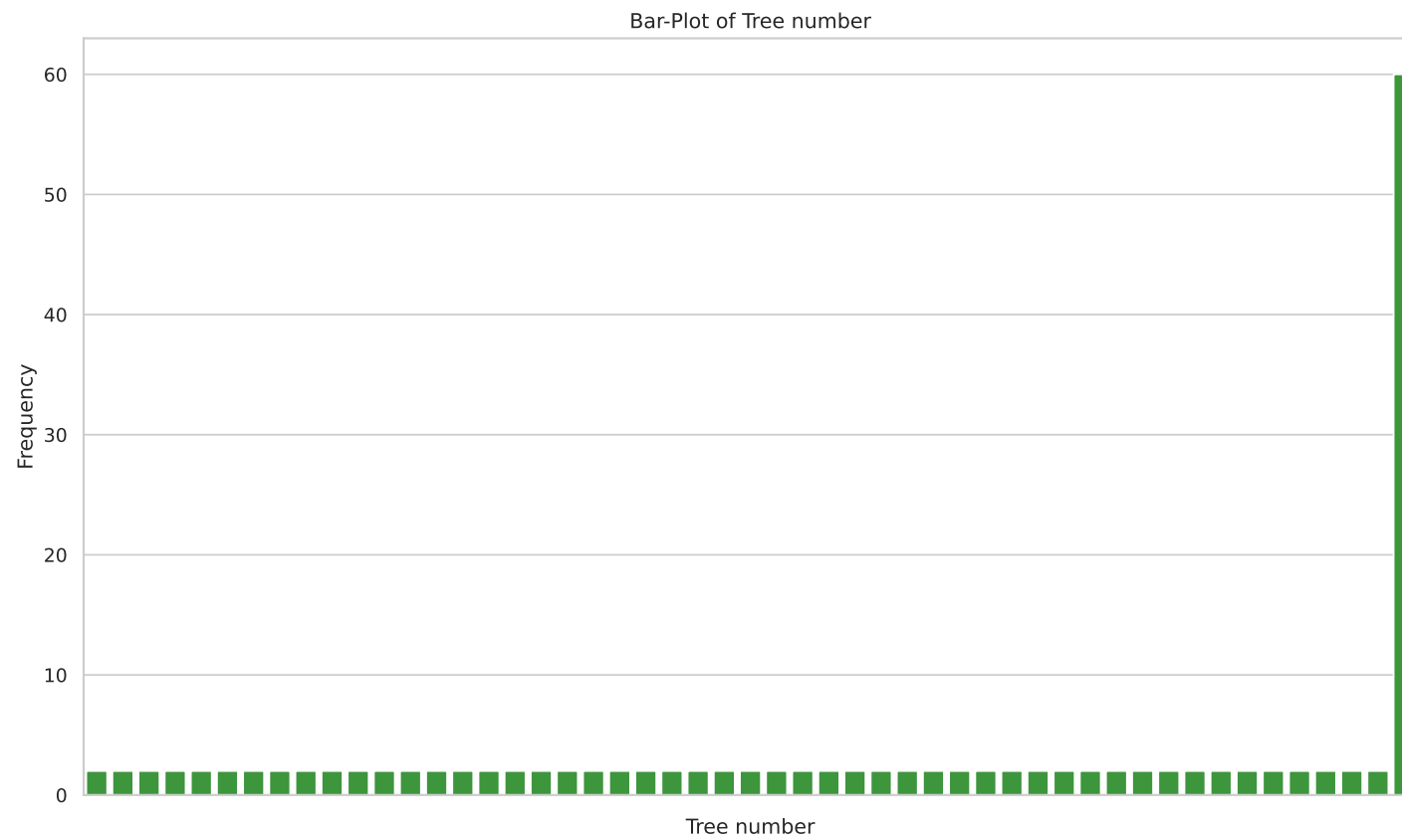






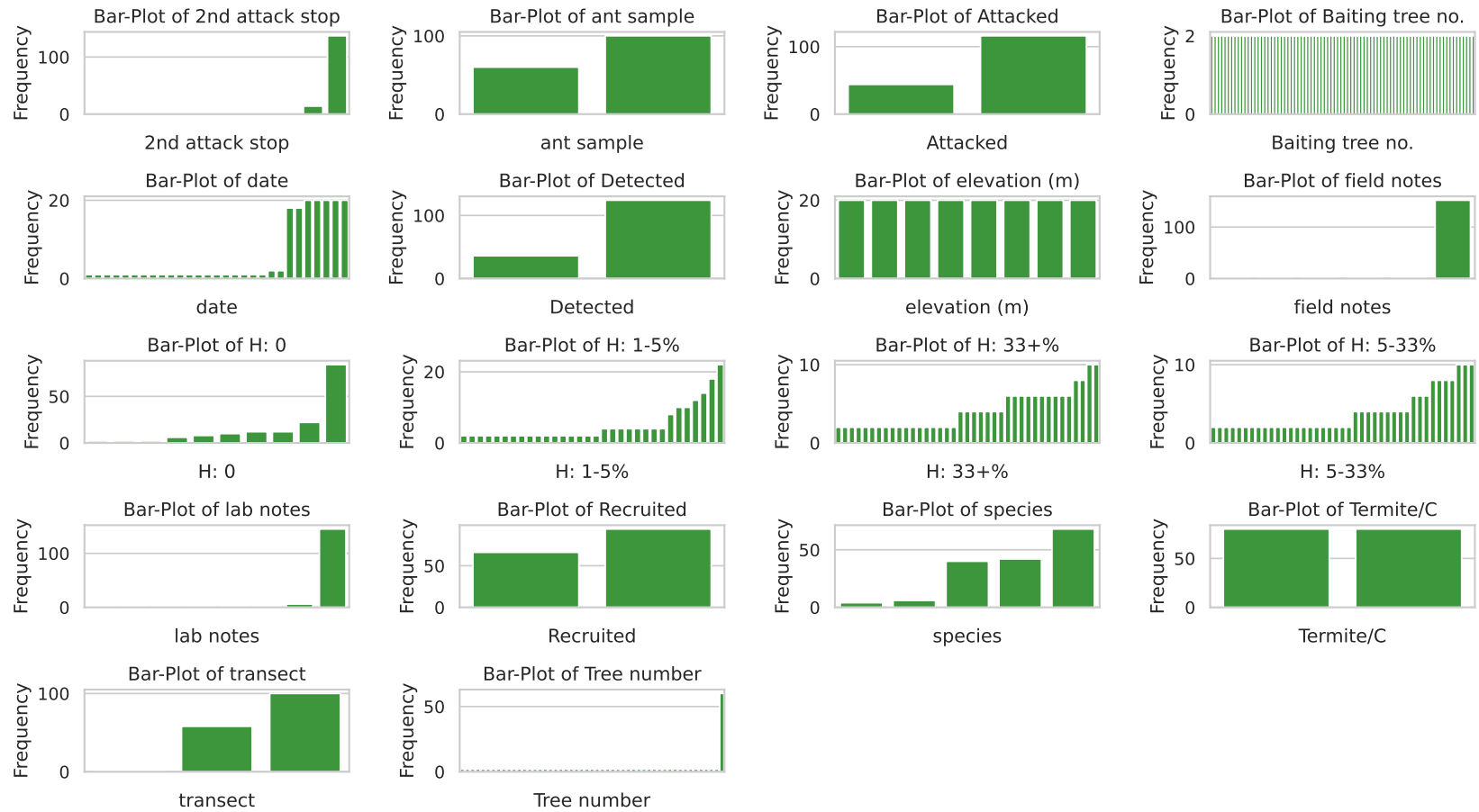






Bar-Plots Summary

Multiple Bar-Plots of variables in one figure. Variables are sorted alphabetically. No labels displayed.

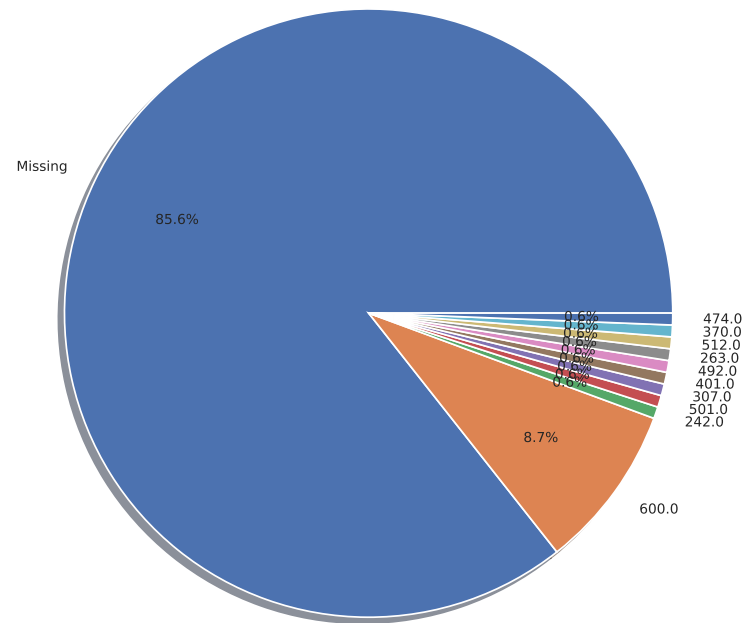


Pie Plots

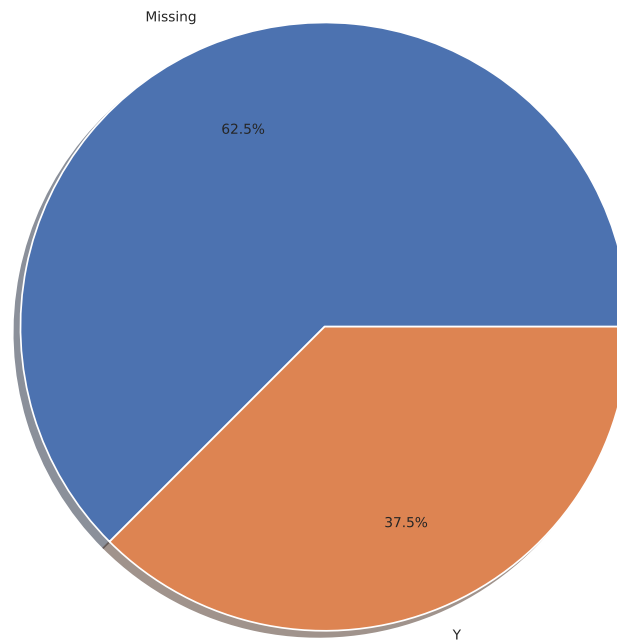
One Pie Plot per page for each variable. Variables are sorted alphabetically.

□

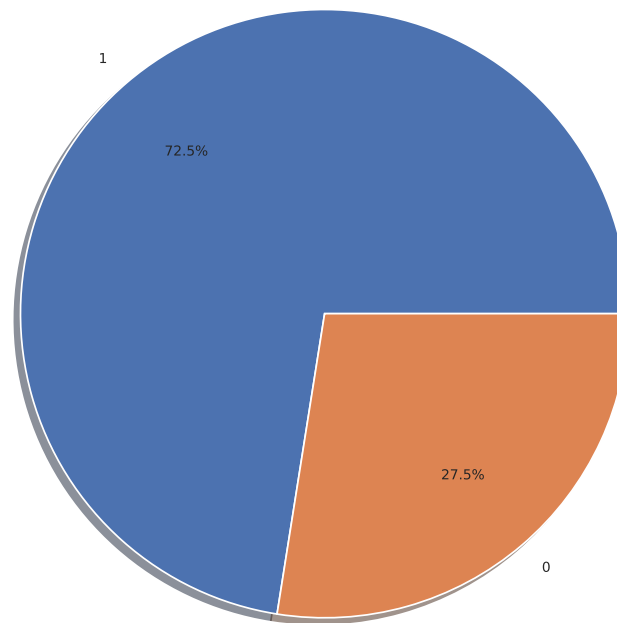
Pie-Plot of 2nd attack stop



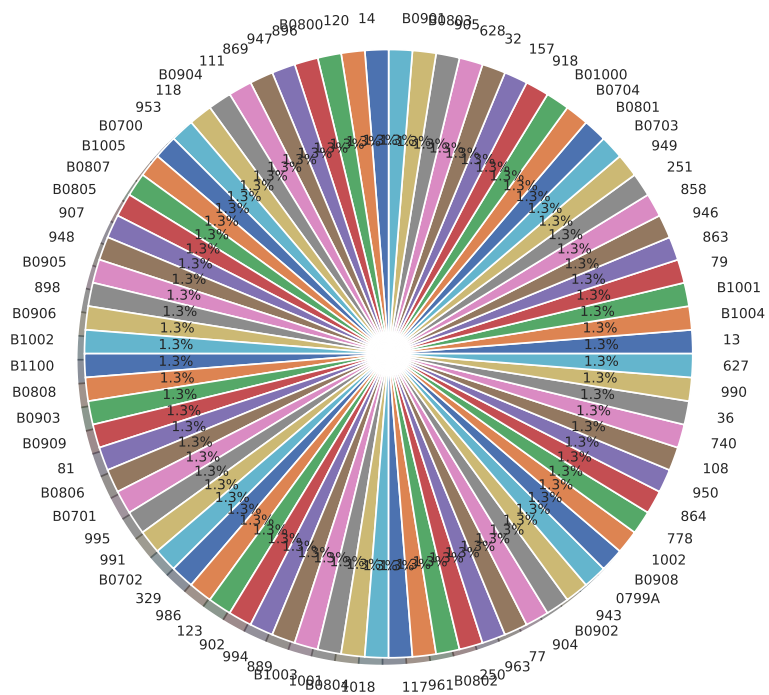
Pie-Plot of ant sample



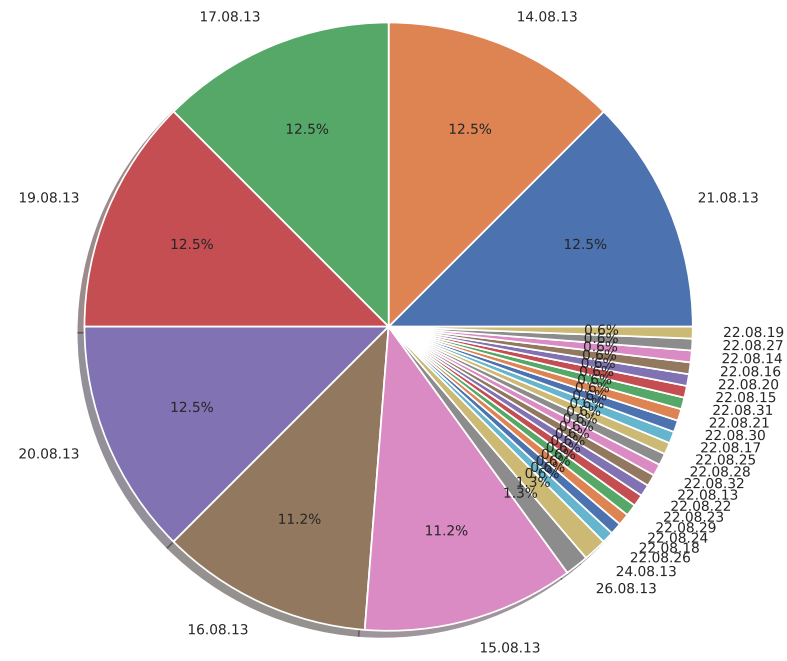
Pie-Plot of Attacked



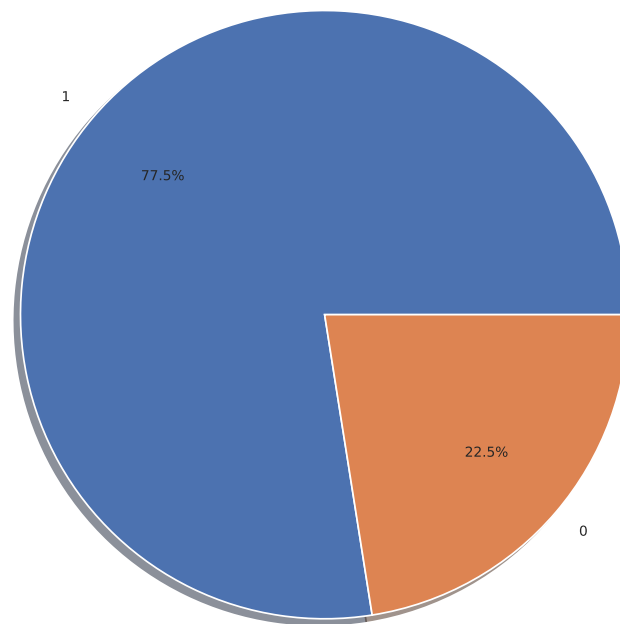
Pie-Plot of Baiting tree no.



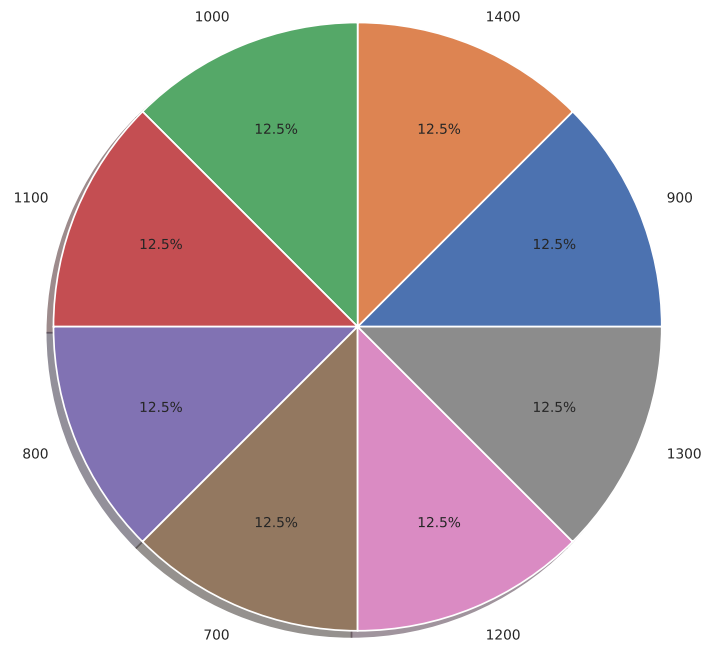
Pie-Plot of date



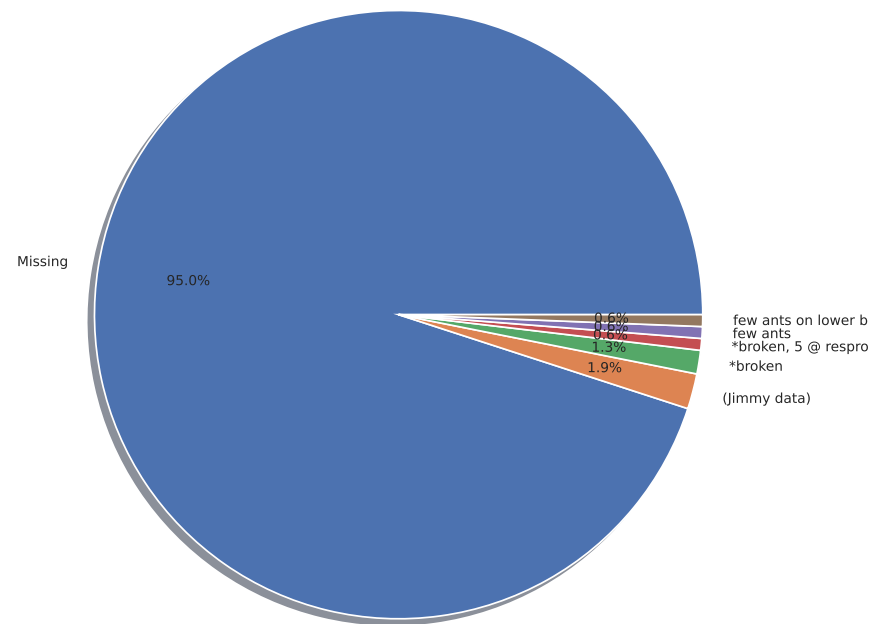
Pie-Plot of Detected



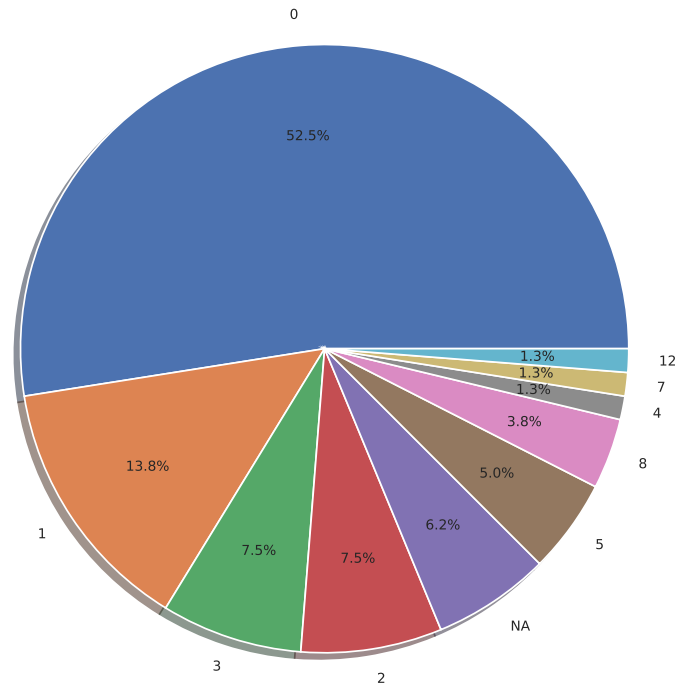
Pie-Plot of elevation (m)



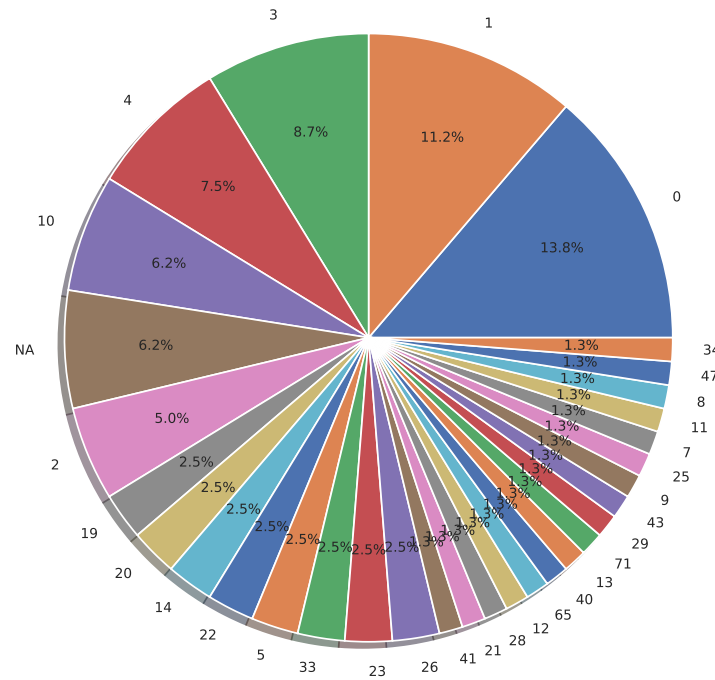
Pie-Plot of field notes



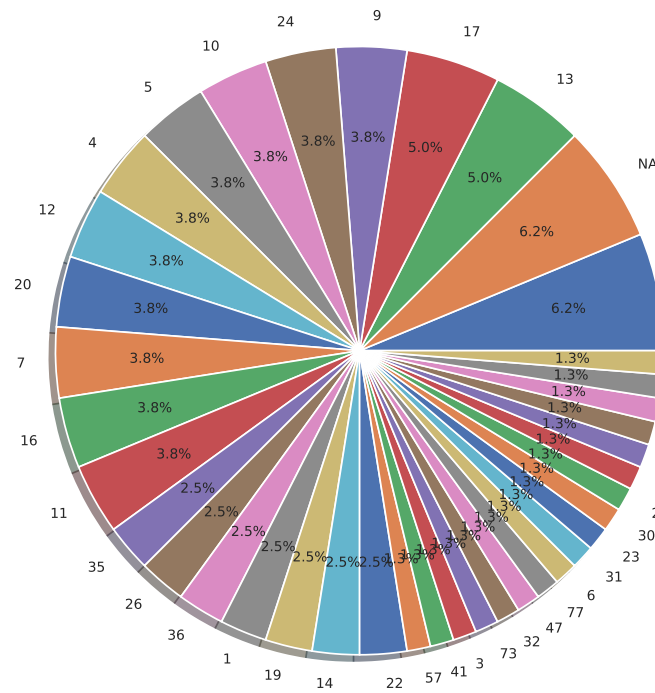
Pie-Plot of H: 0



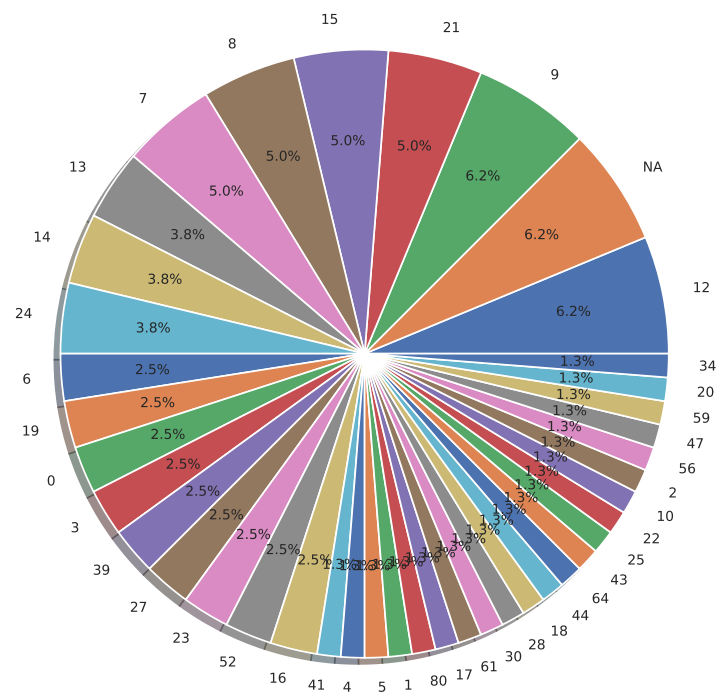
Pie-Plot of H: 1-5%



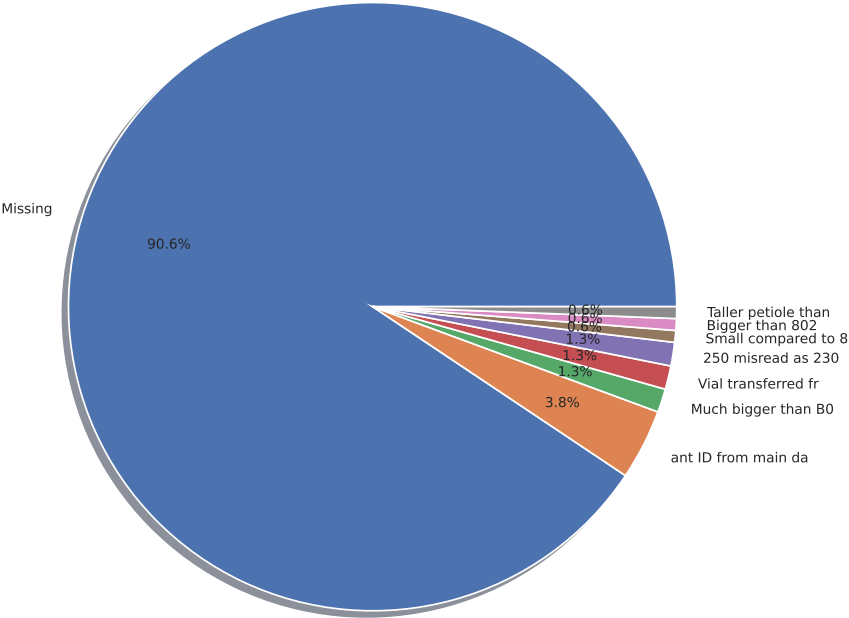
Pie-Plot of H: 33+%



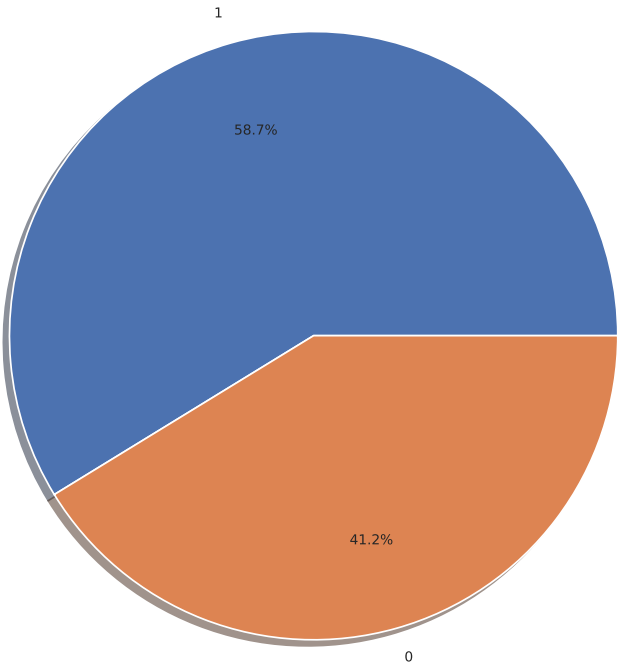
Pie-Plot of H: 5-33%



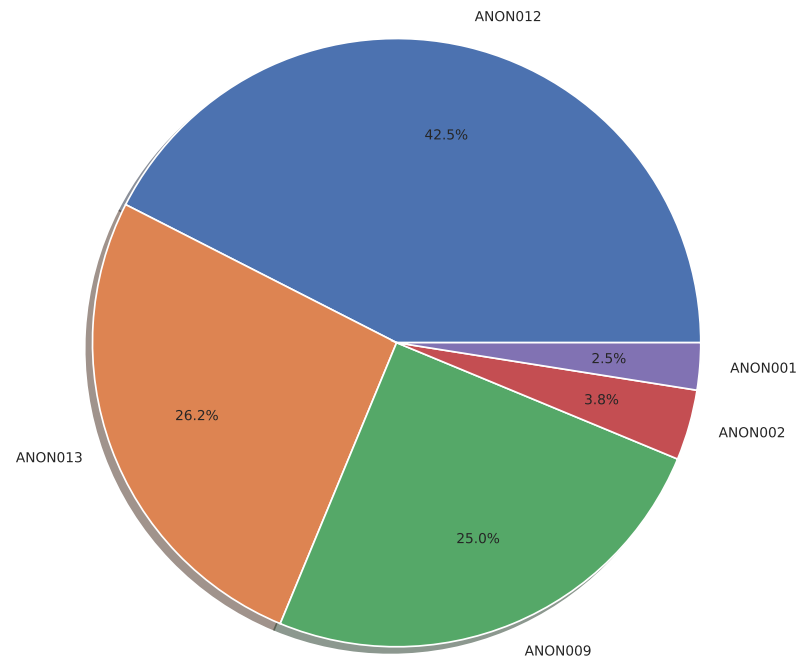
Pie-Plot of lab notes



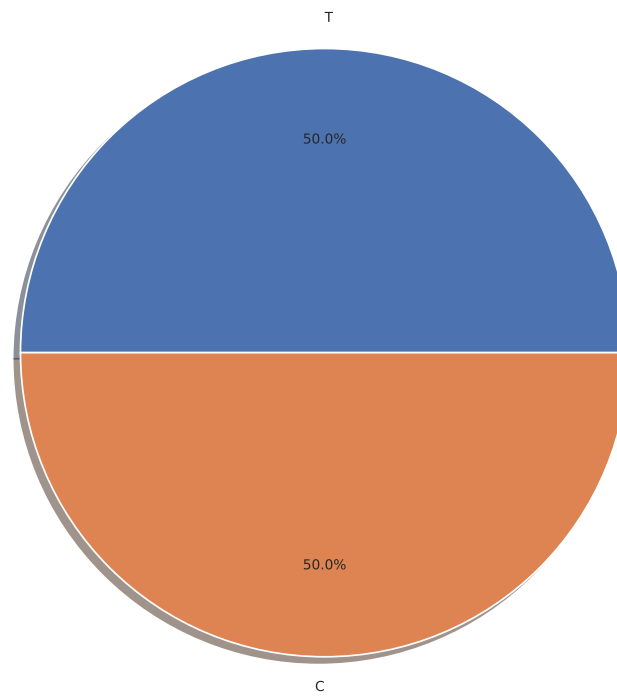
Pie-Plot of Recruited



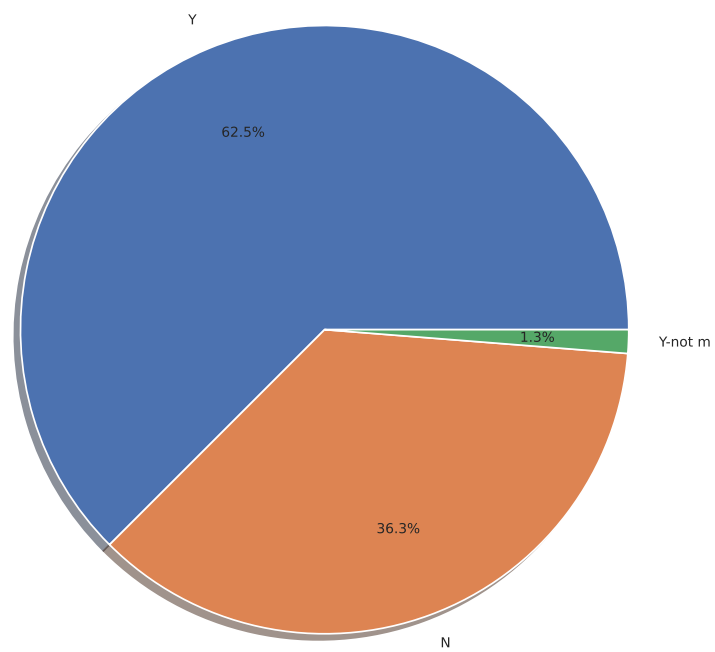
Pie-Plot of species



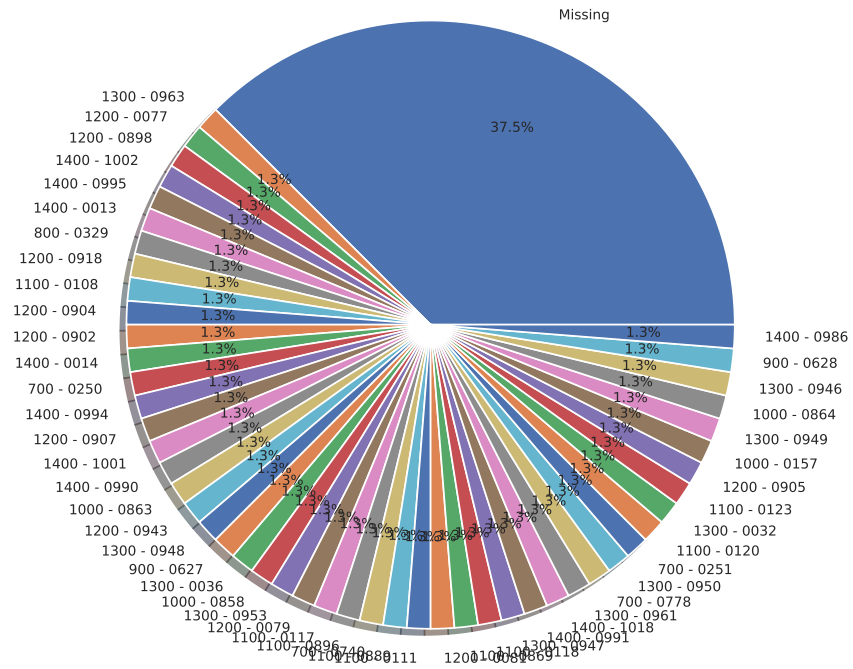
Pie-Plot of Termite/C



Pie-Plot of transect



Pie-Plot of Tree number



Pie Plots Summary

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

See figures on next page.

