Exploratory Data Analysis (EDA)

Statsomat.com

17 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): 24
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: 17

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

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

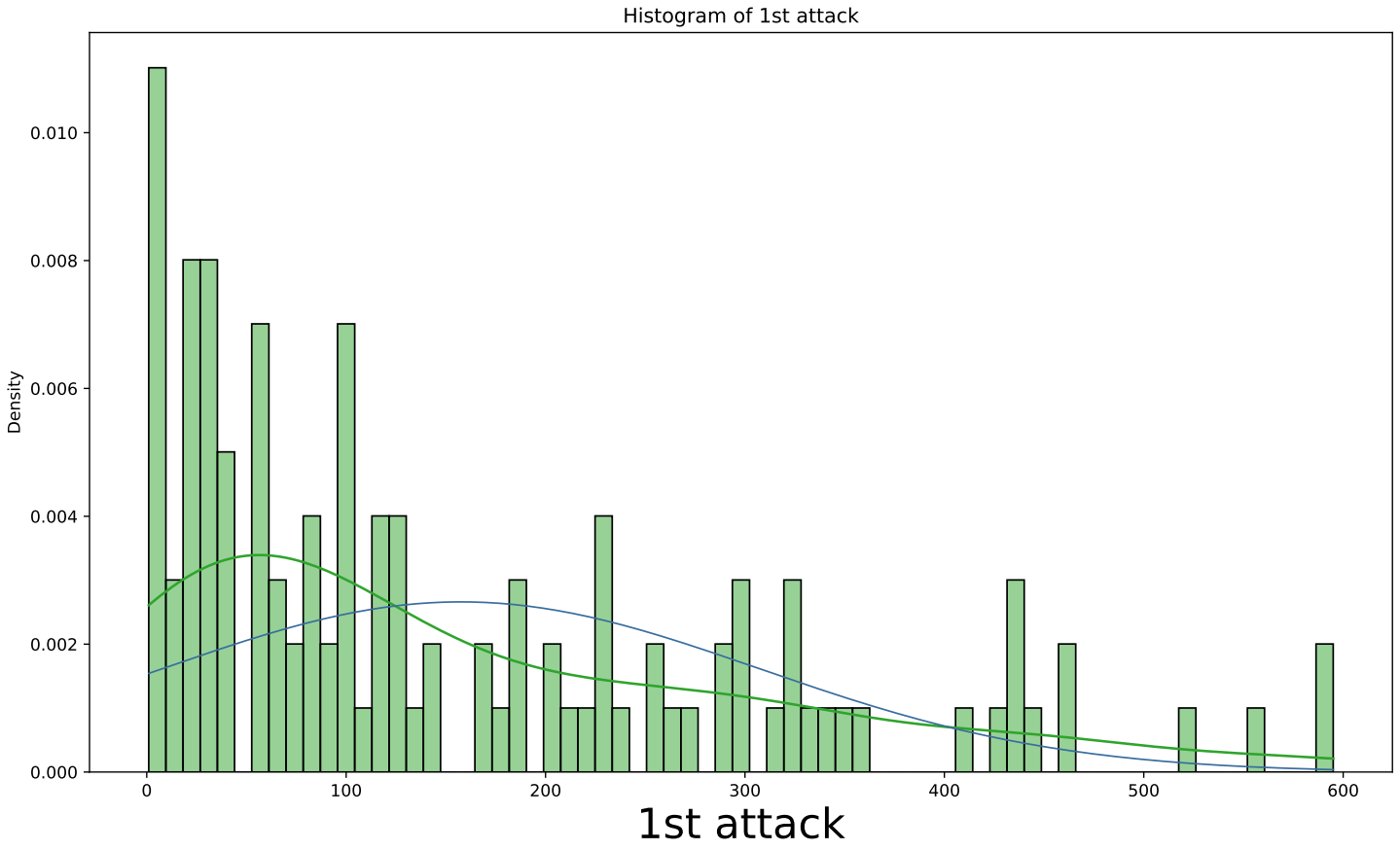
Graphics

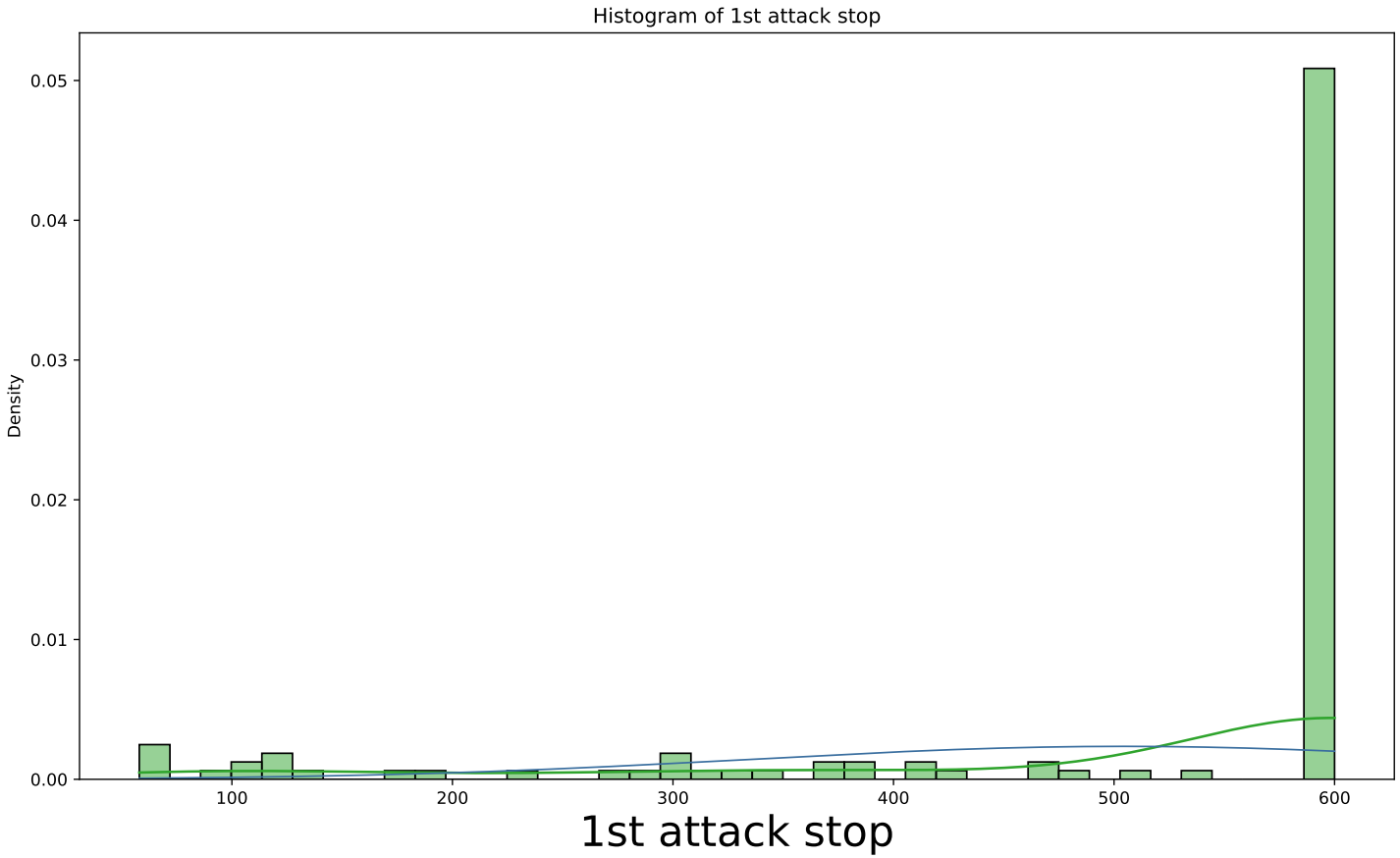
Histograms

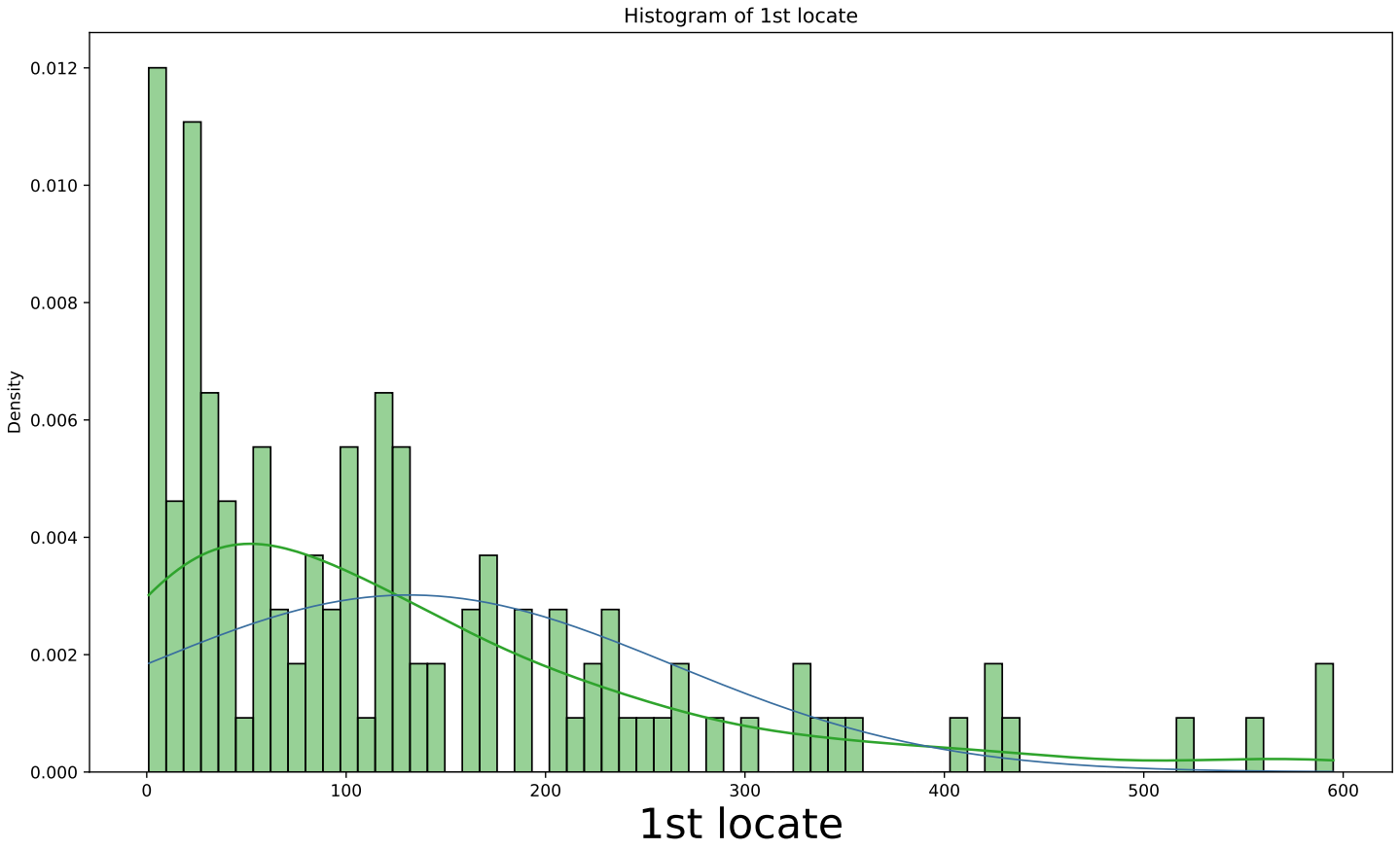
Details: Density Histograms. One large figure per page for each variable, sorted alphabetically. The blue line represents the normal density approximation. The green line represents a special kernel density approximation.

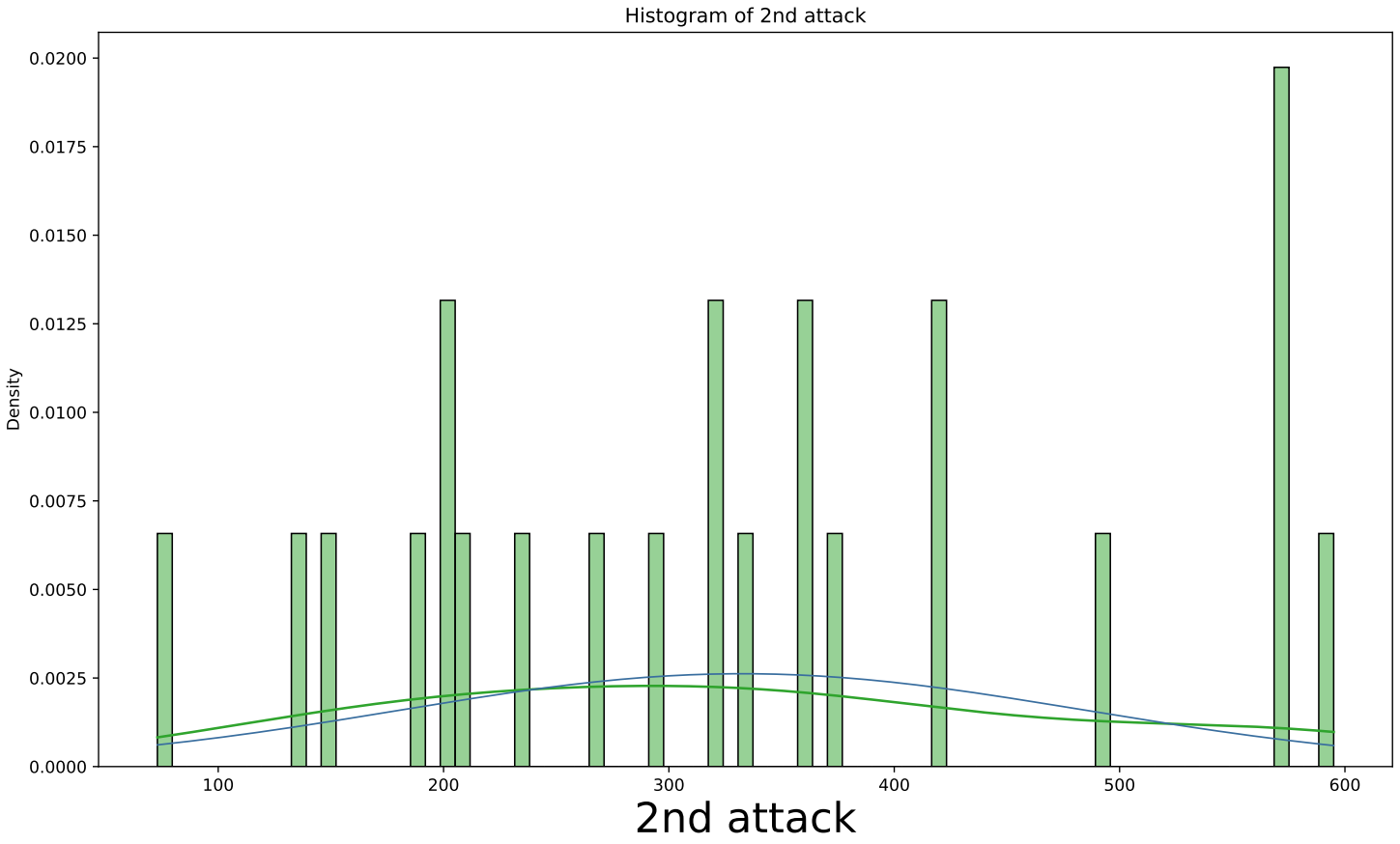
See figures on next page.

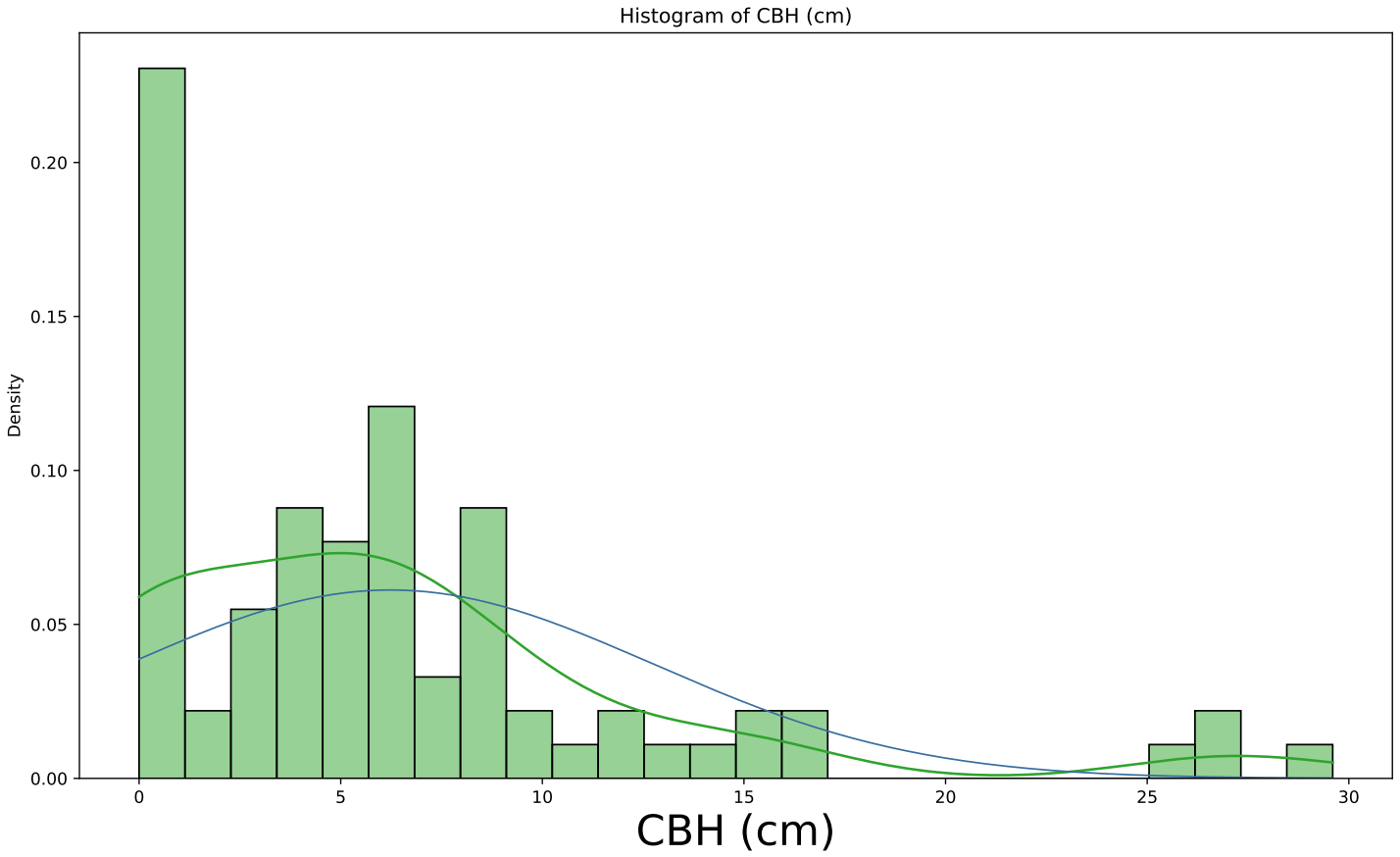
[]

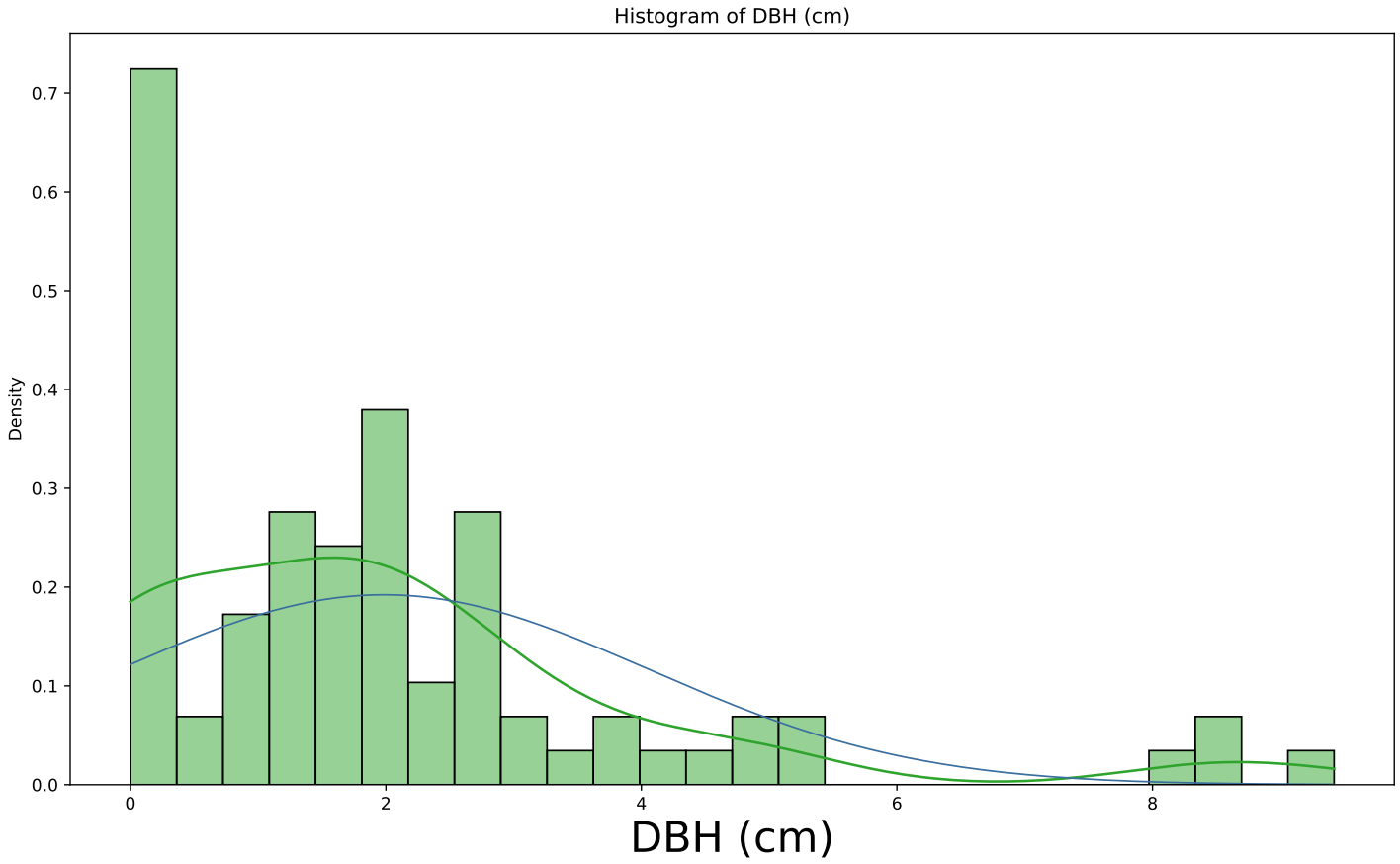

Histogram of 1st attack
Density
0.010
0.008
0.006
0.004
0.002
0.000
0
100
200
300
400
500
600
1st attack

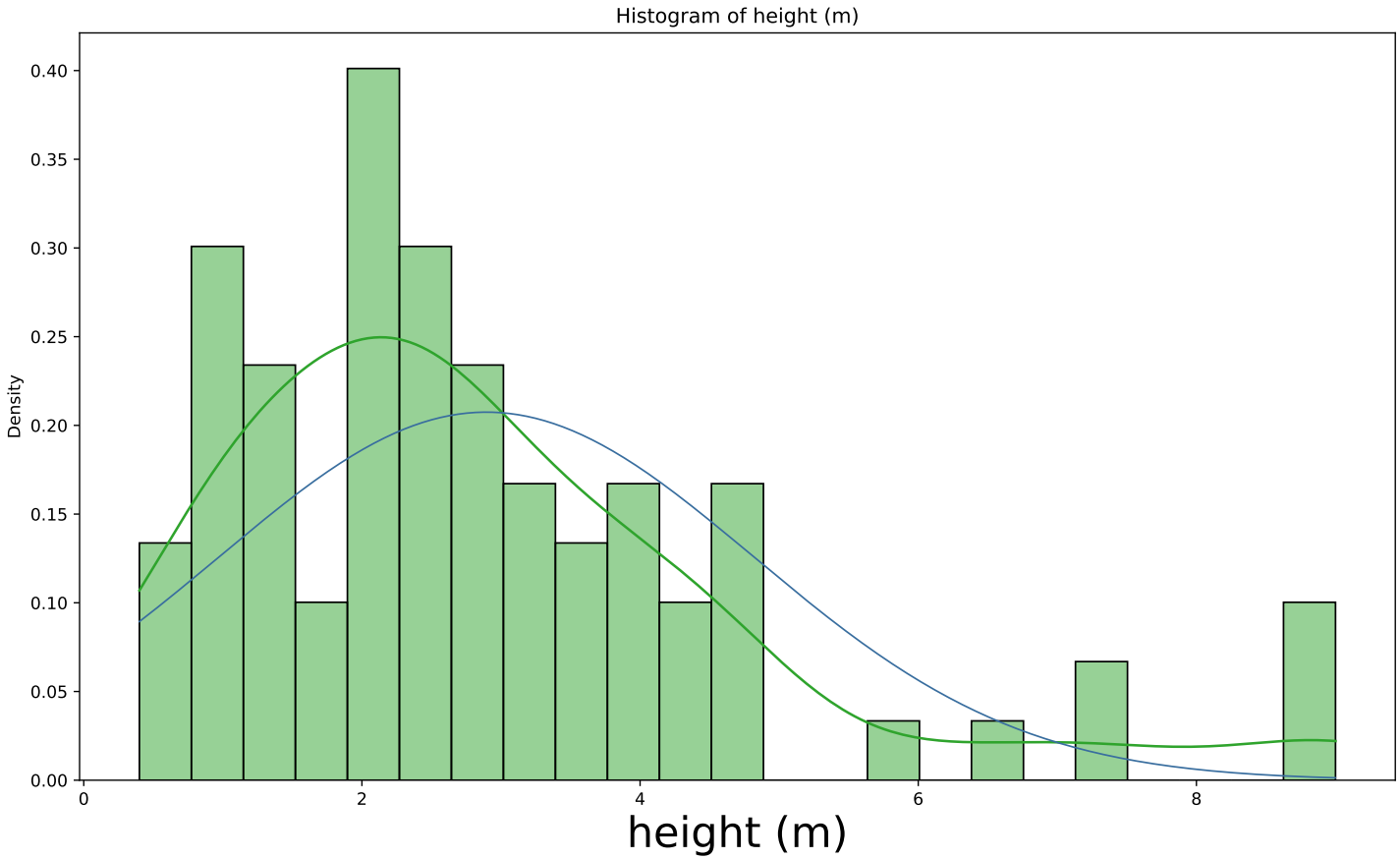

Histogram of 1st attack stop
Density
0.05
0.04
0.03
0.02
0.01
0.00
100
200
300
400
500
600
1st attack stop


Histogram of 1st locate
Density
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
100
200
300
400
500
600
1st locate


Histogram of 2nd attack
Density
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
100
200
300
400
500
600
2nd attack

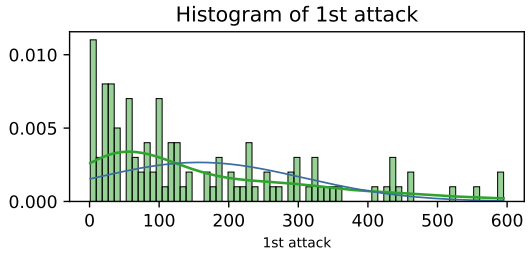

Histogram of CBH (cm)
Density
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
25
30
CBH (cm)

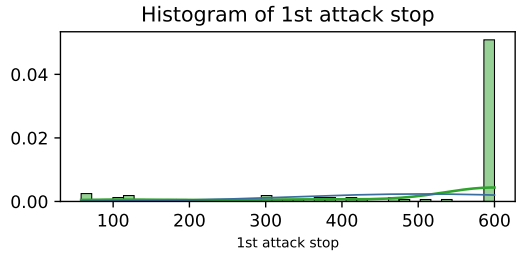

Histogram of DBH (cm)
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
2
4
6
8
DBH (cm)

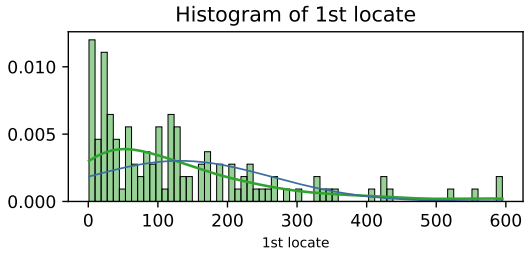

Histogram of height (m)
Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
height (m)

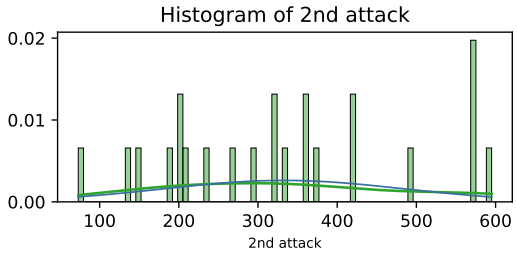
Histograms Summary

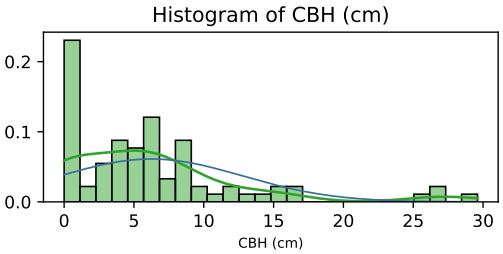
Multiple Relative Frequency Histogram in one figure. Variables are sorted alphabetically. The blue line represents the normal density approximation. The green line represents a special kernel density approximation.

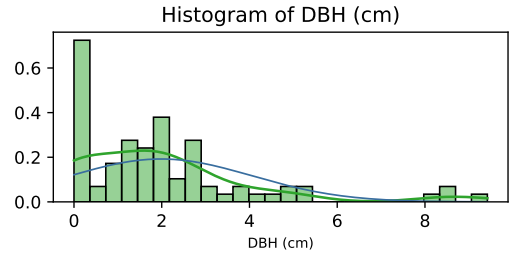


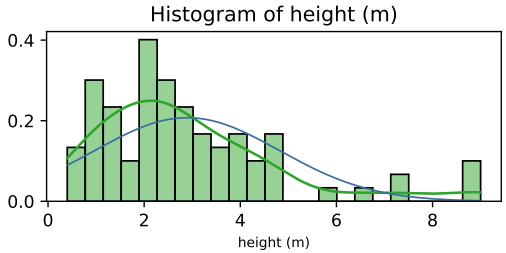








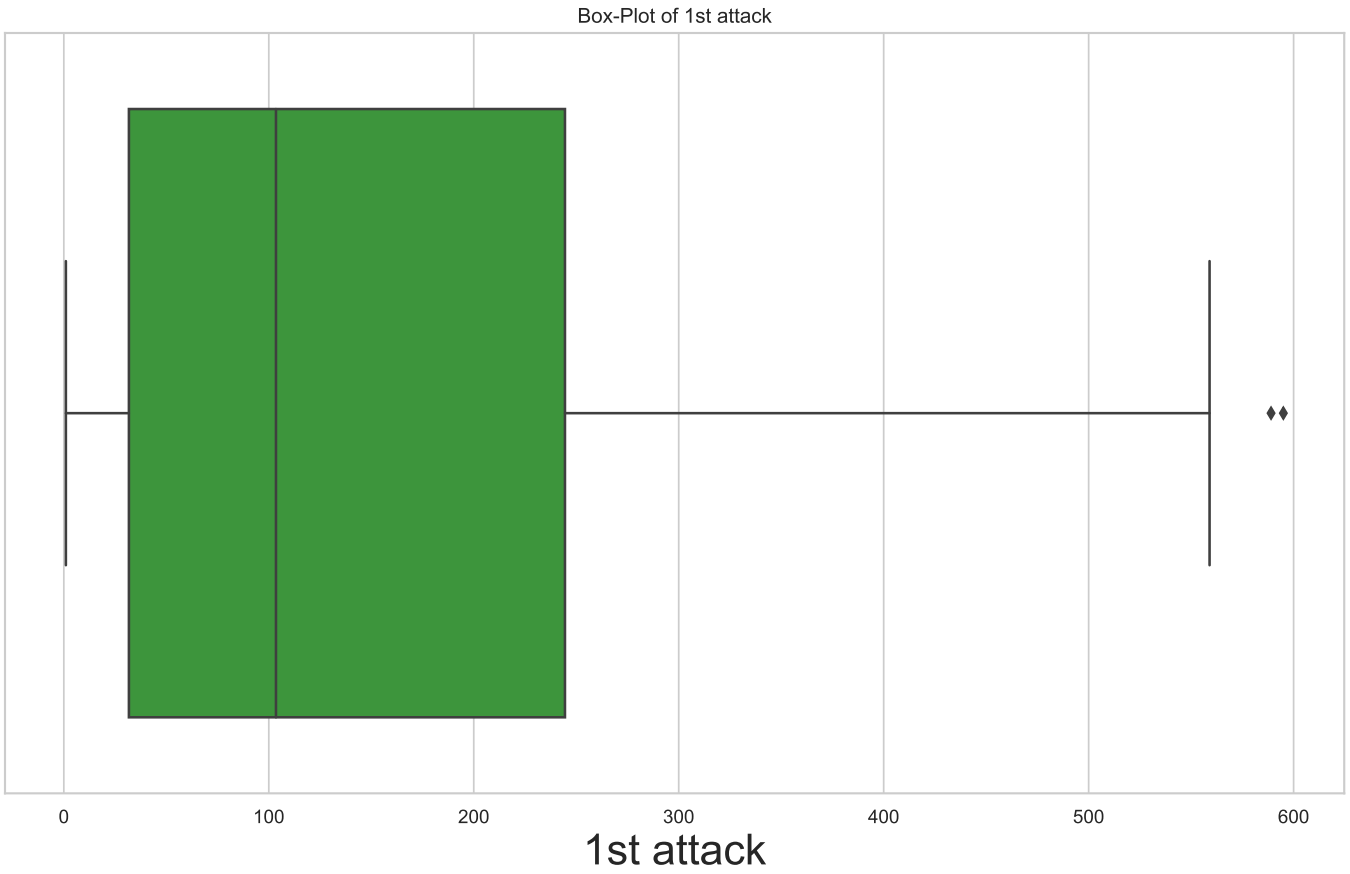


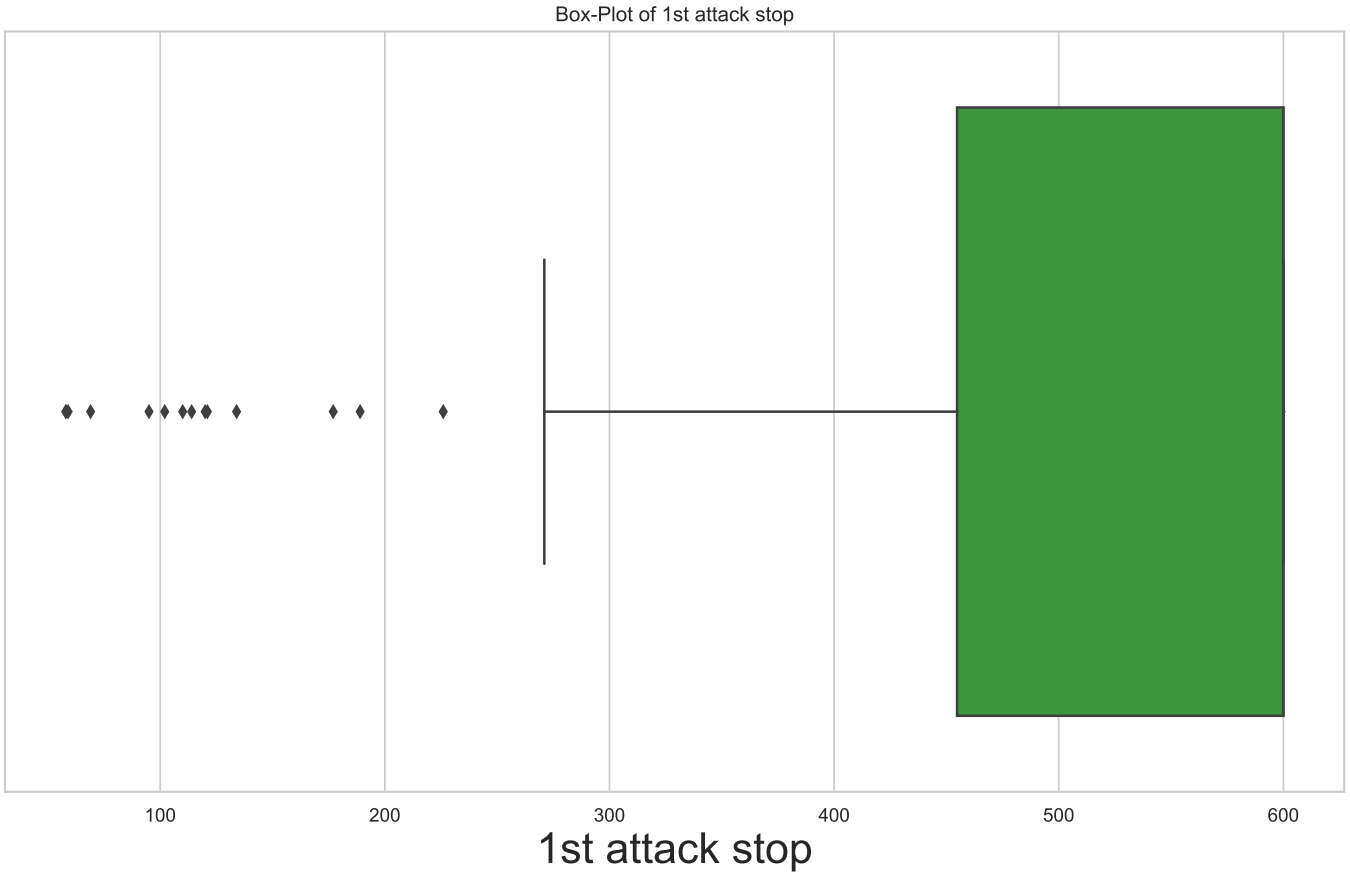


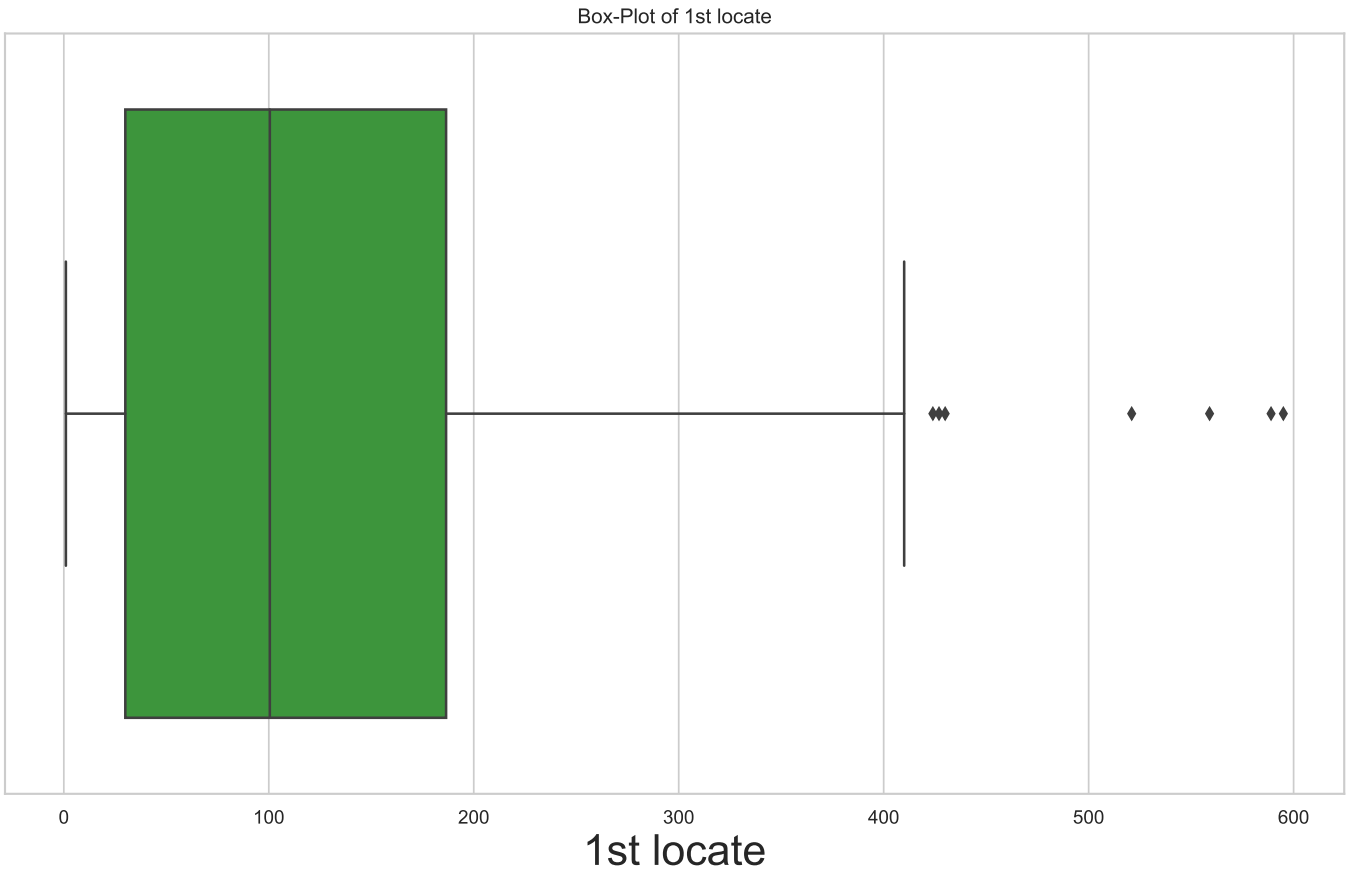
Box-Plots

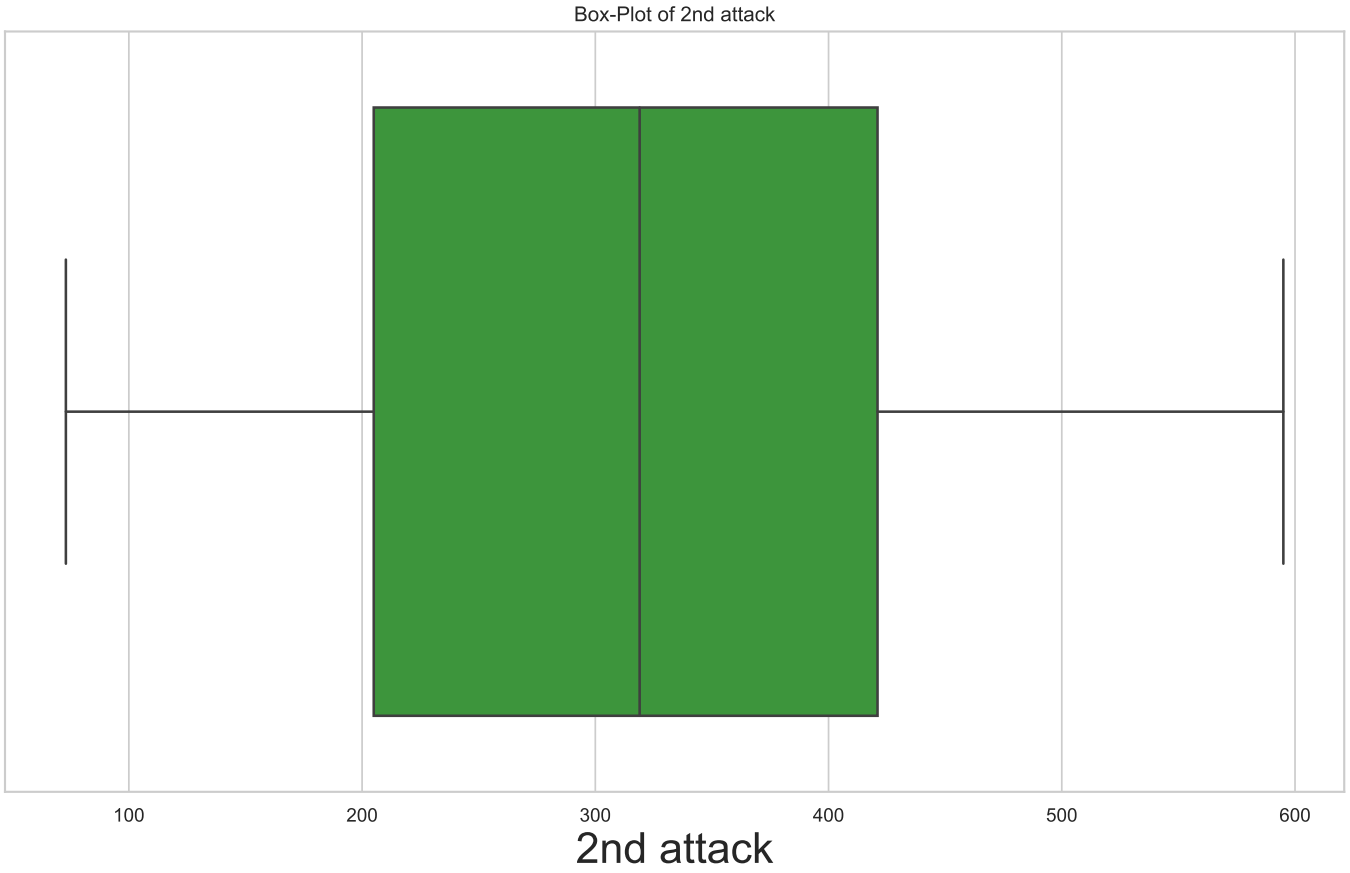
One Box-Plot per page for each variable. Variables are sorted alphabetically.

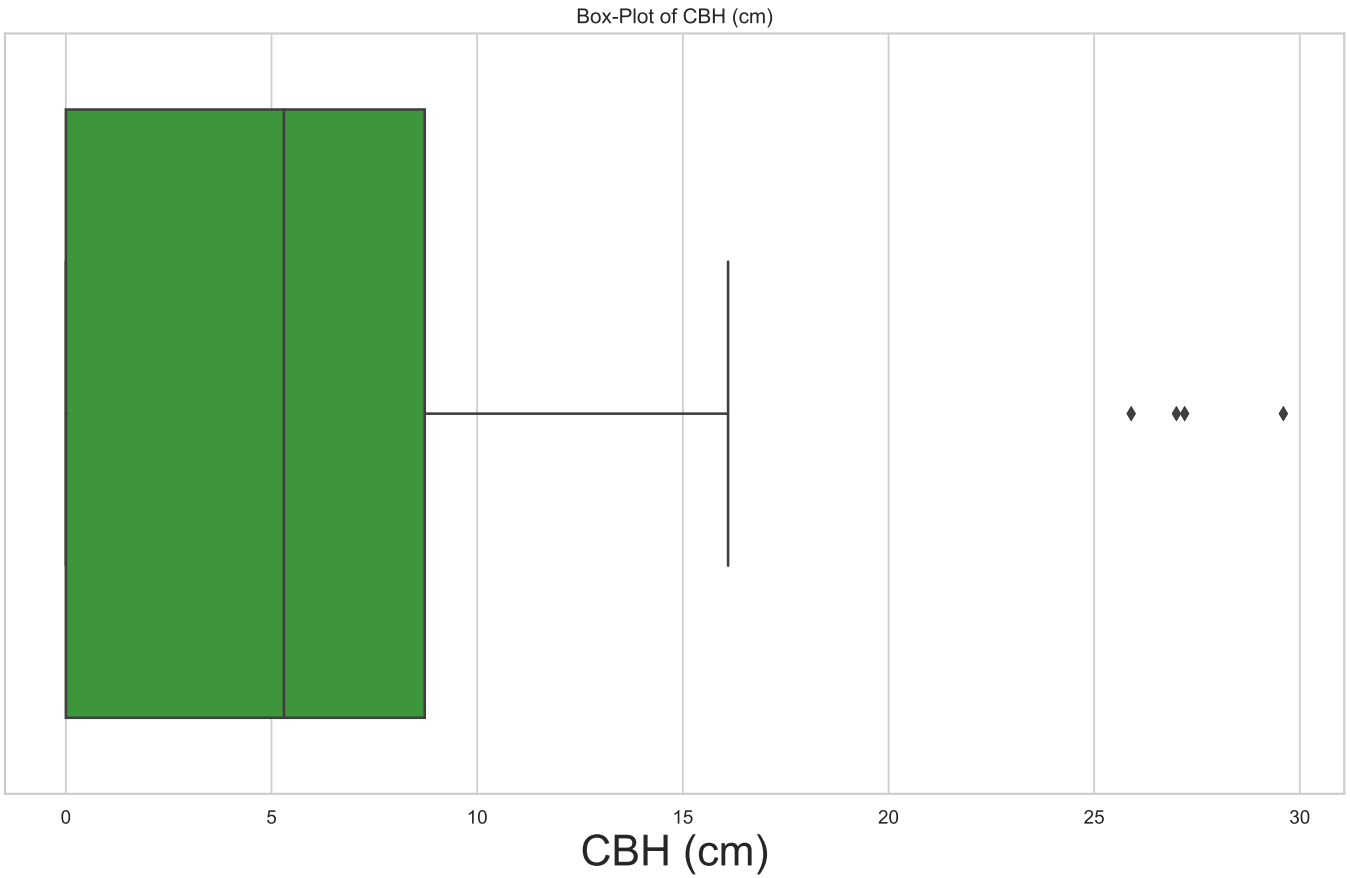
[]

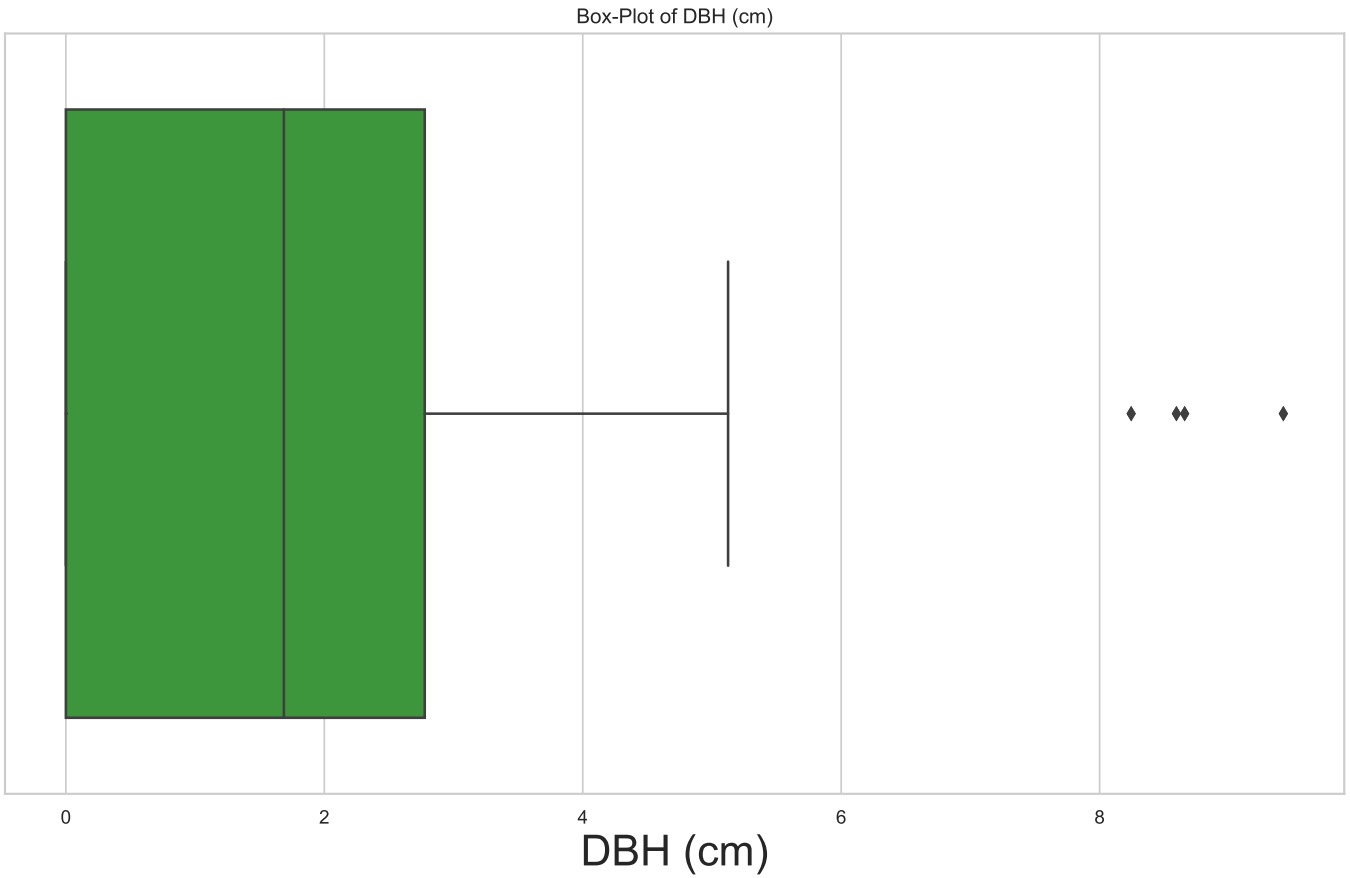


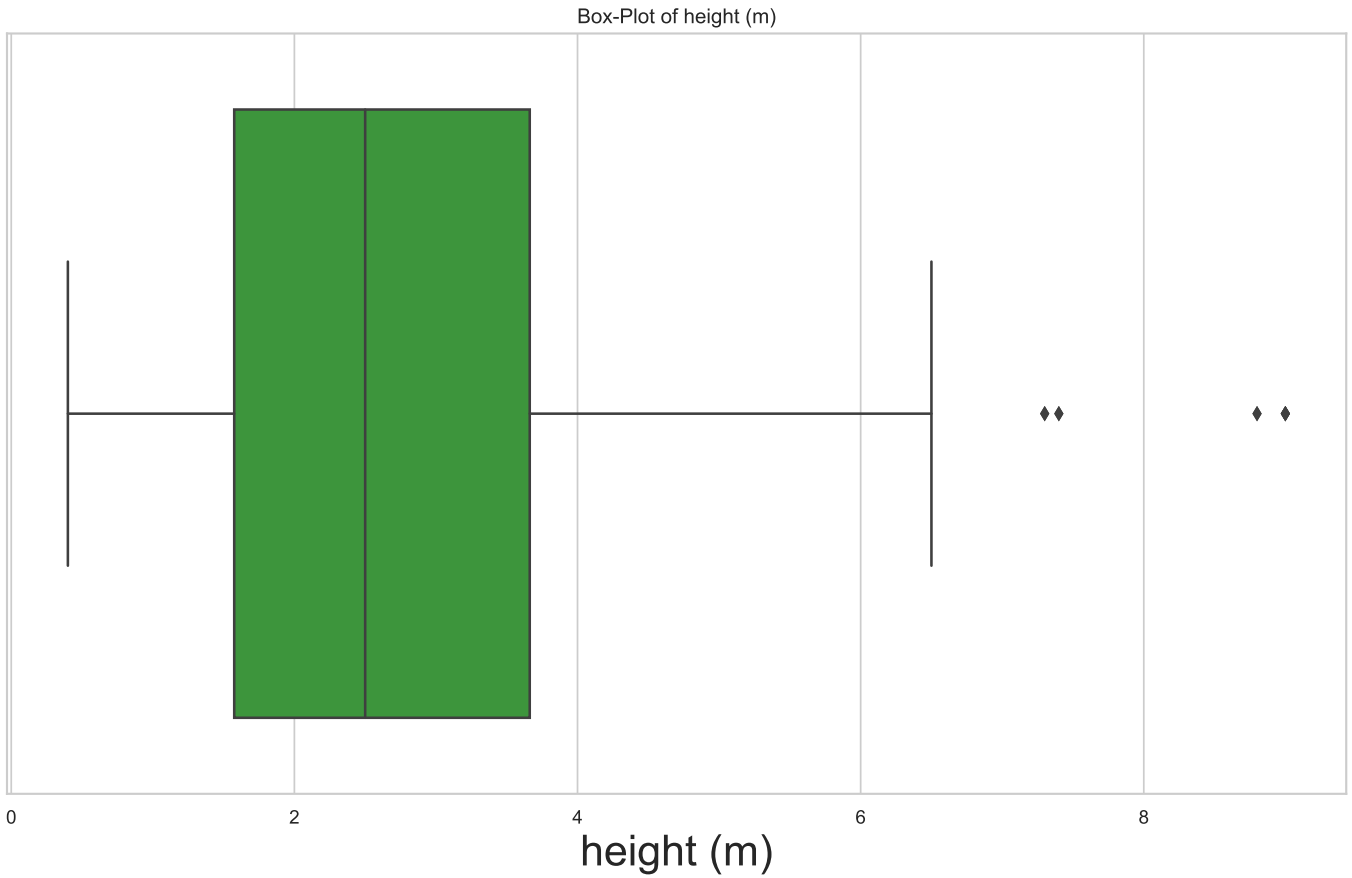

Box-Plot of 1st attack stop
100
200
300
400
500
600
1st attack stop


Box-Plot of 1st locate
0
100
200
300
400
500
600
1st locate


Box-Plot of 2nd attack
100
200
300
400
500
600
2nd attack

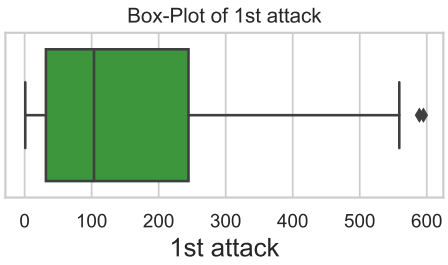

Box-Plot of CBH (cm)
0
5
10
15
20
25
30
CBH (cm)

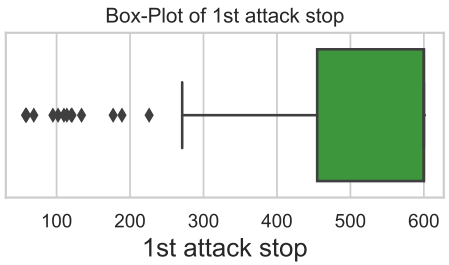

Box-Plot of DBH (cm)
0
2
4
6
8
DBH (cm)

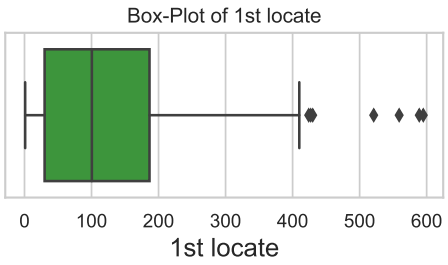

Box-Plot of height (m)
0
2
4
6
8
height (m)

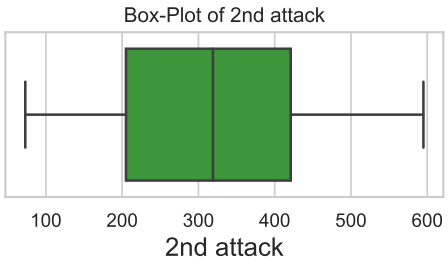
Box-Plots Summary

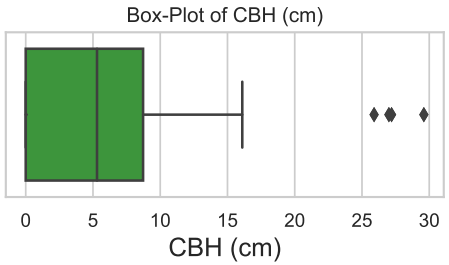
Multiple Box-Plots of variables in one figure. Variables are sorted alphabetically.

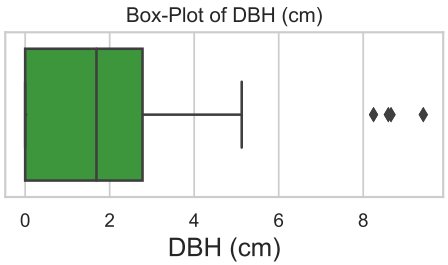


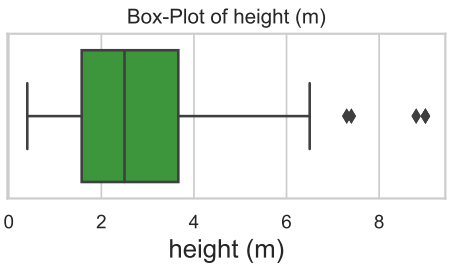








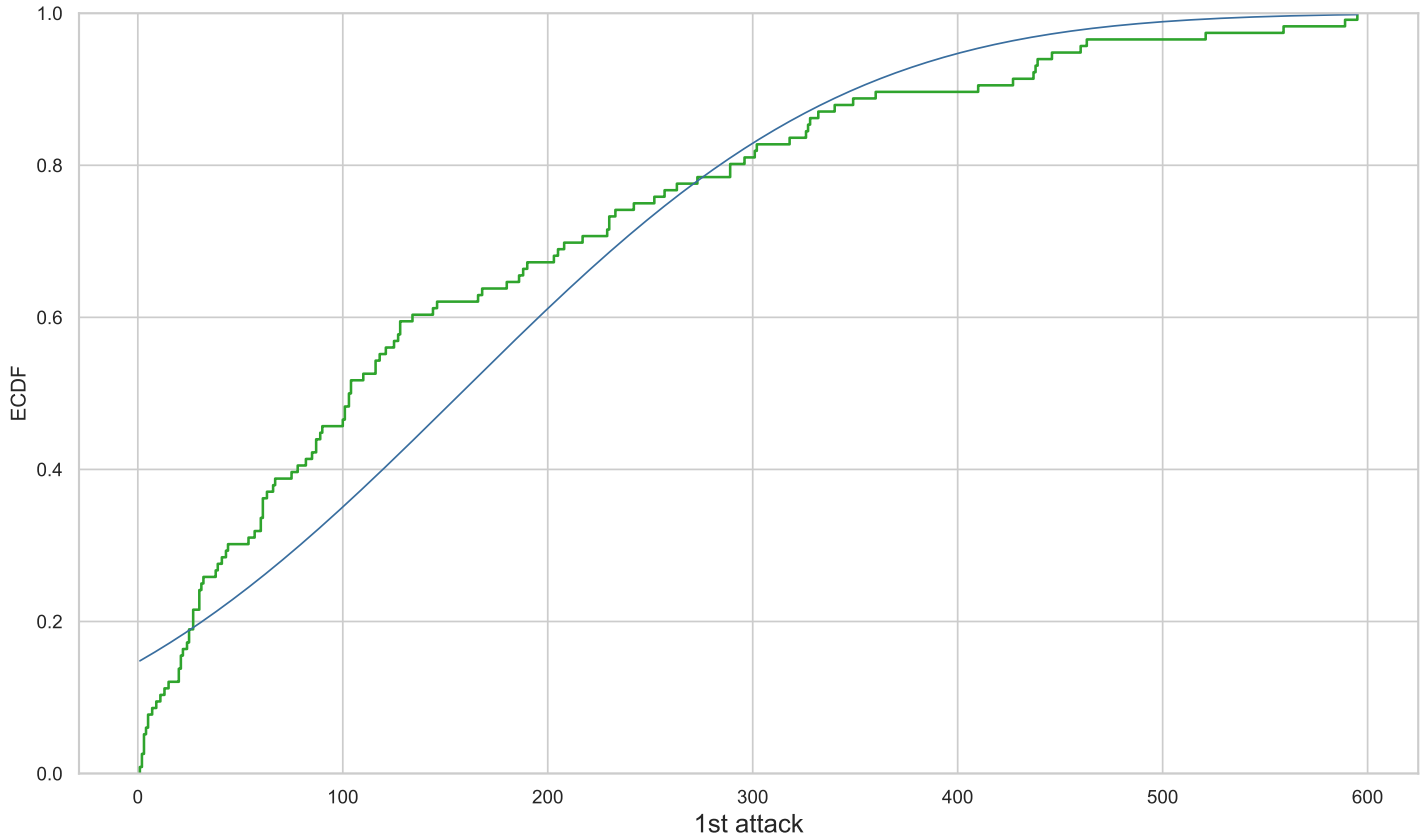


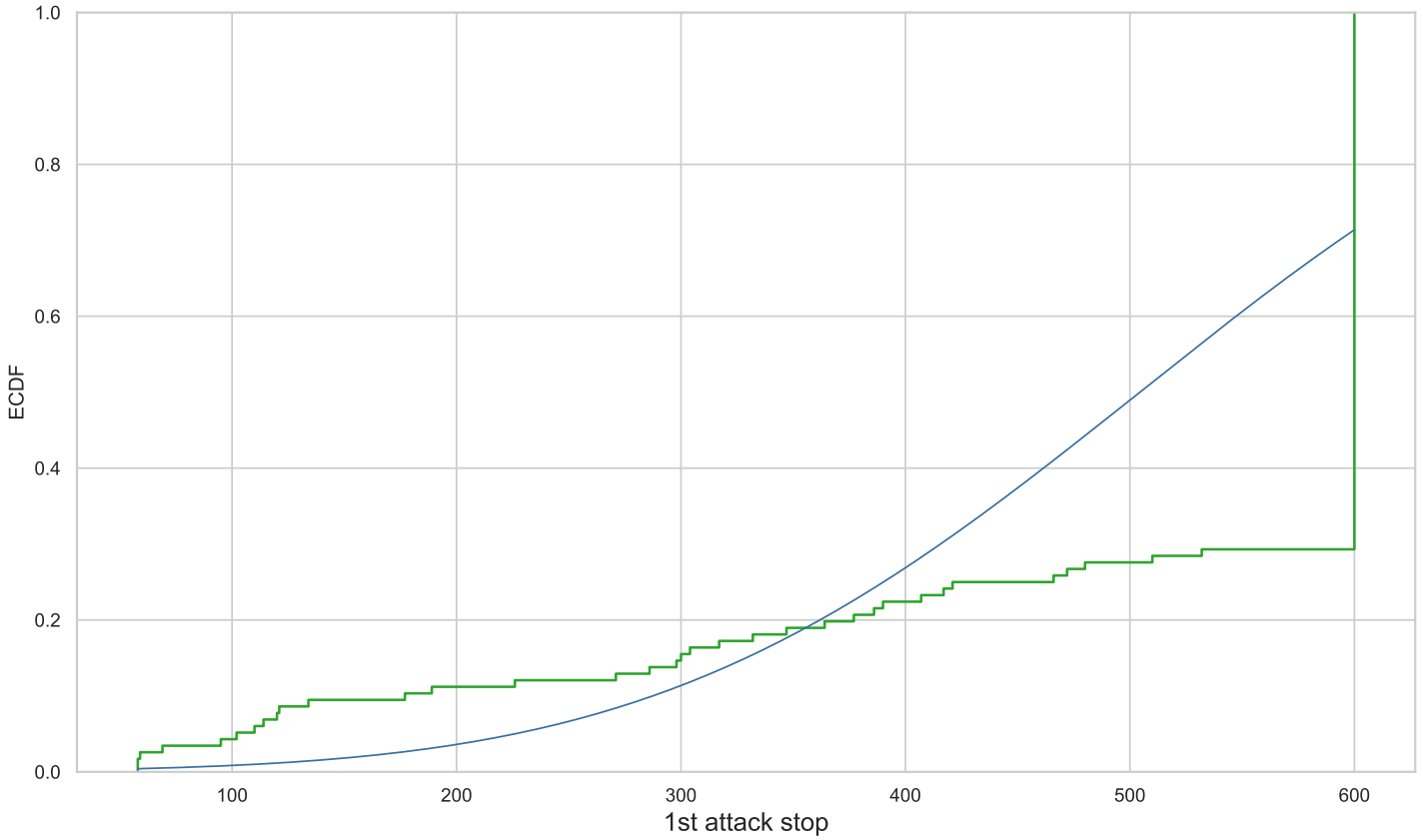


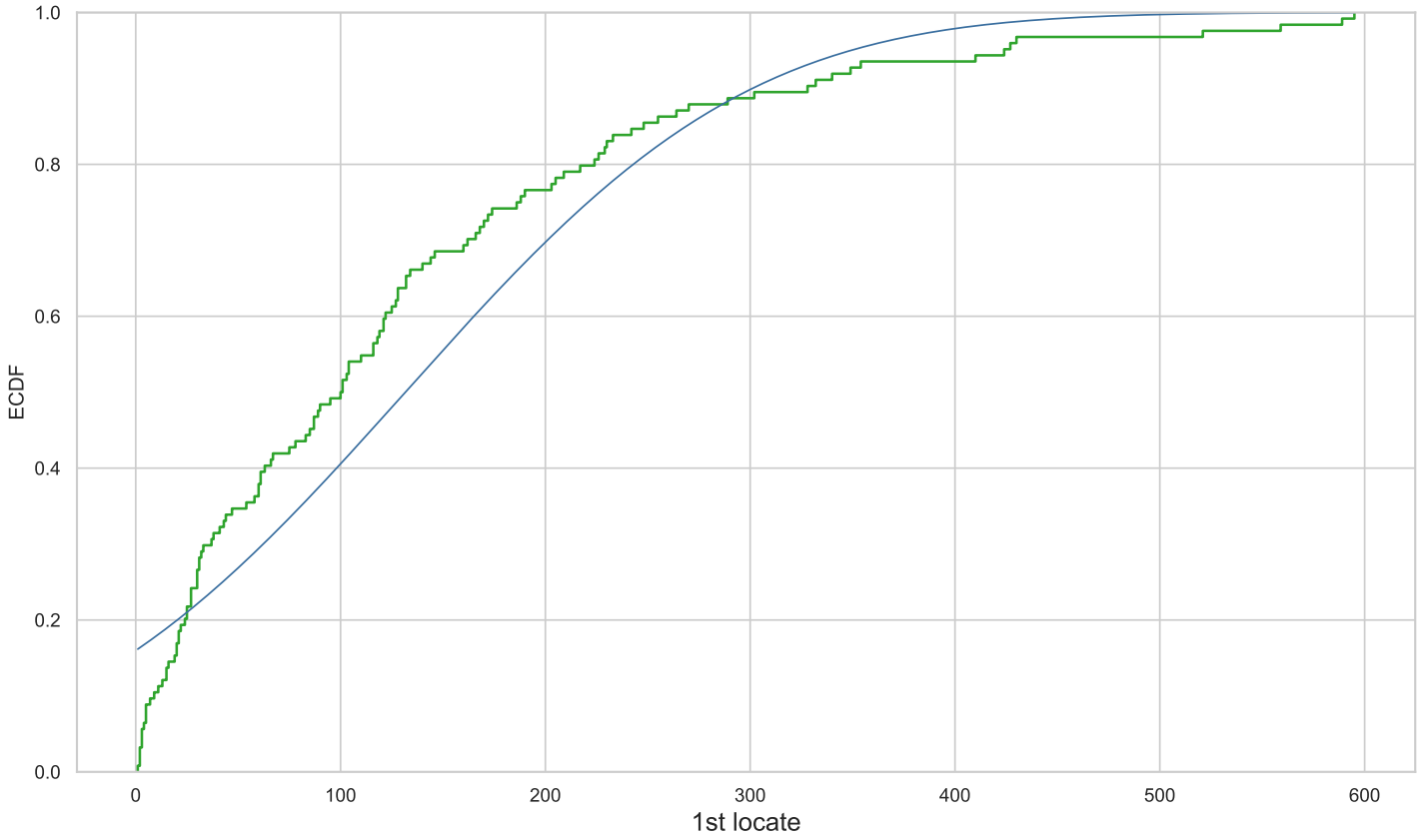
ECDF Plots

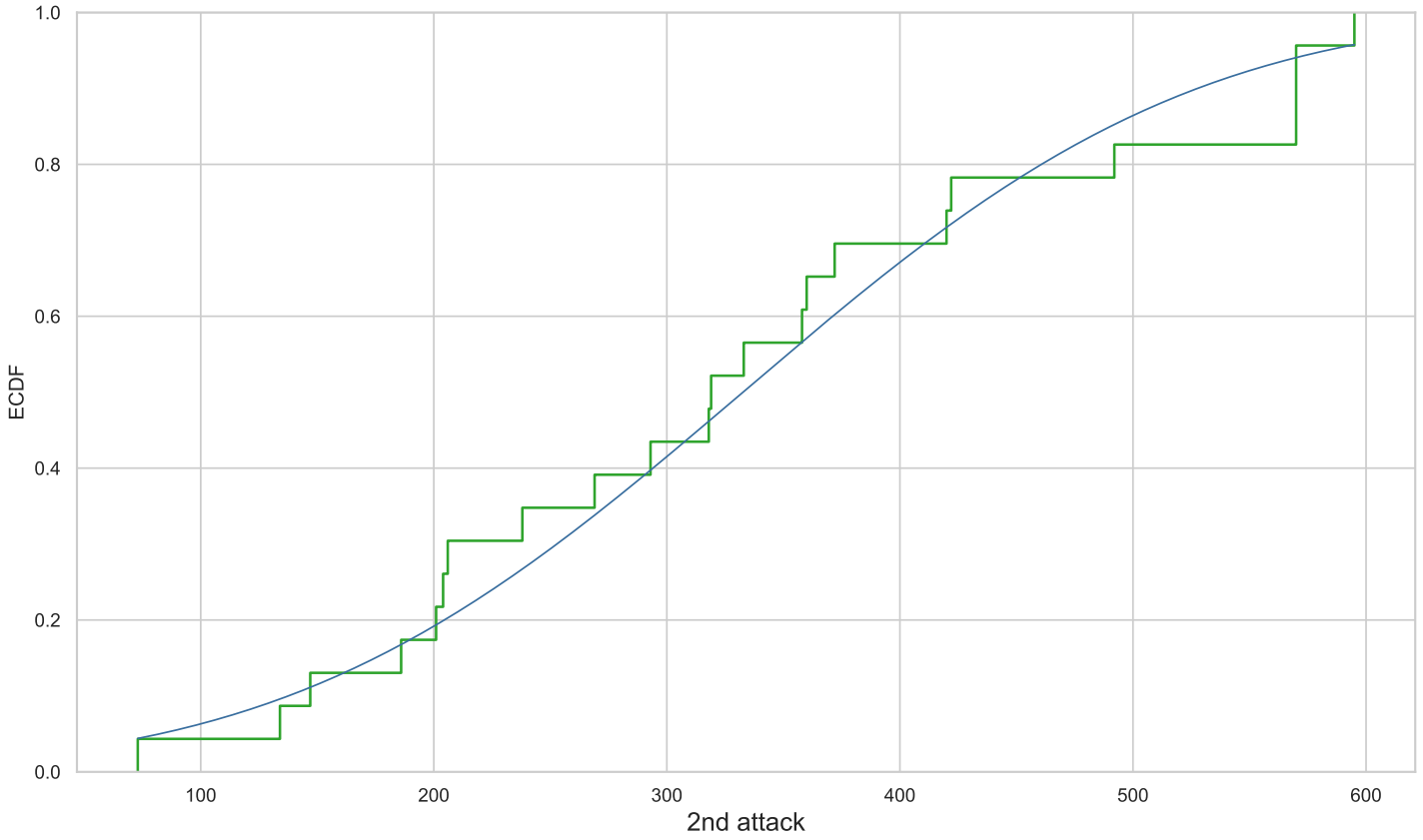
One ECDF (Empirical Cumulative Distribution Function) Plot per page for each variable. Variables are sorted alphabetically. The blue line represents the CDF of a normal distribution. If the variable is normally distributed, the blue line approximates well the ECDF.

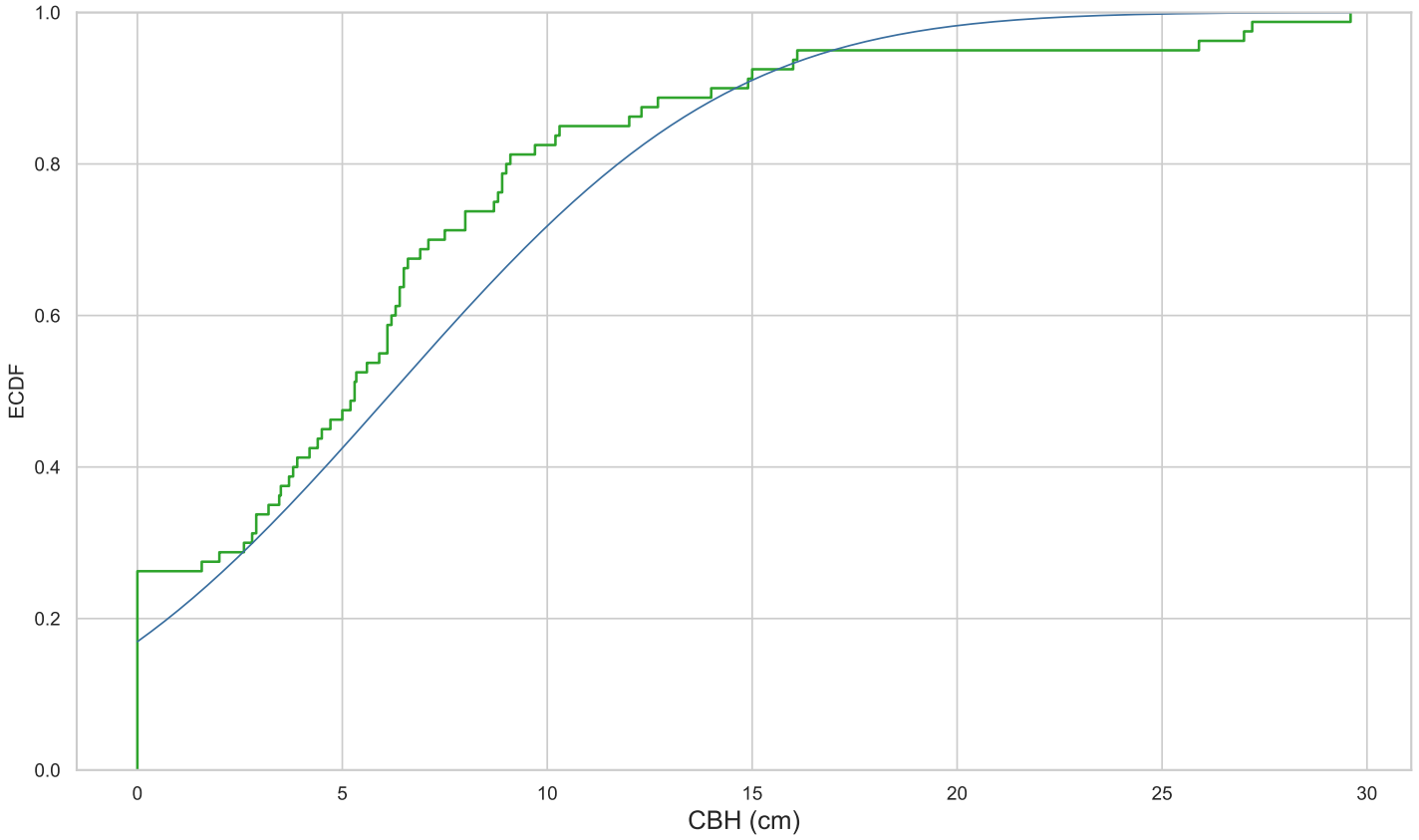
[]

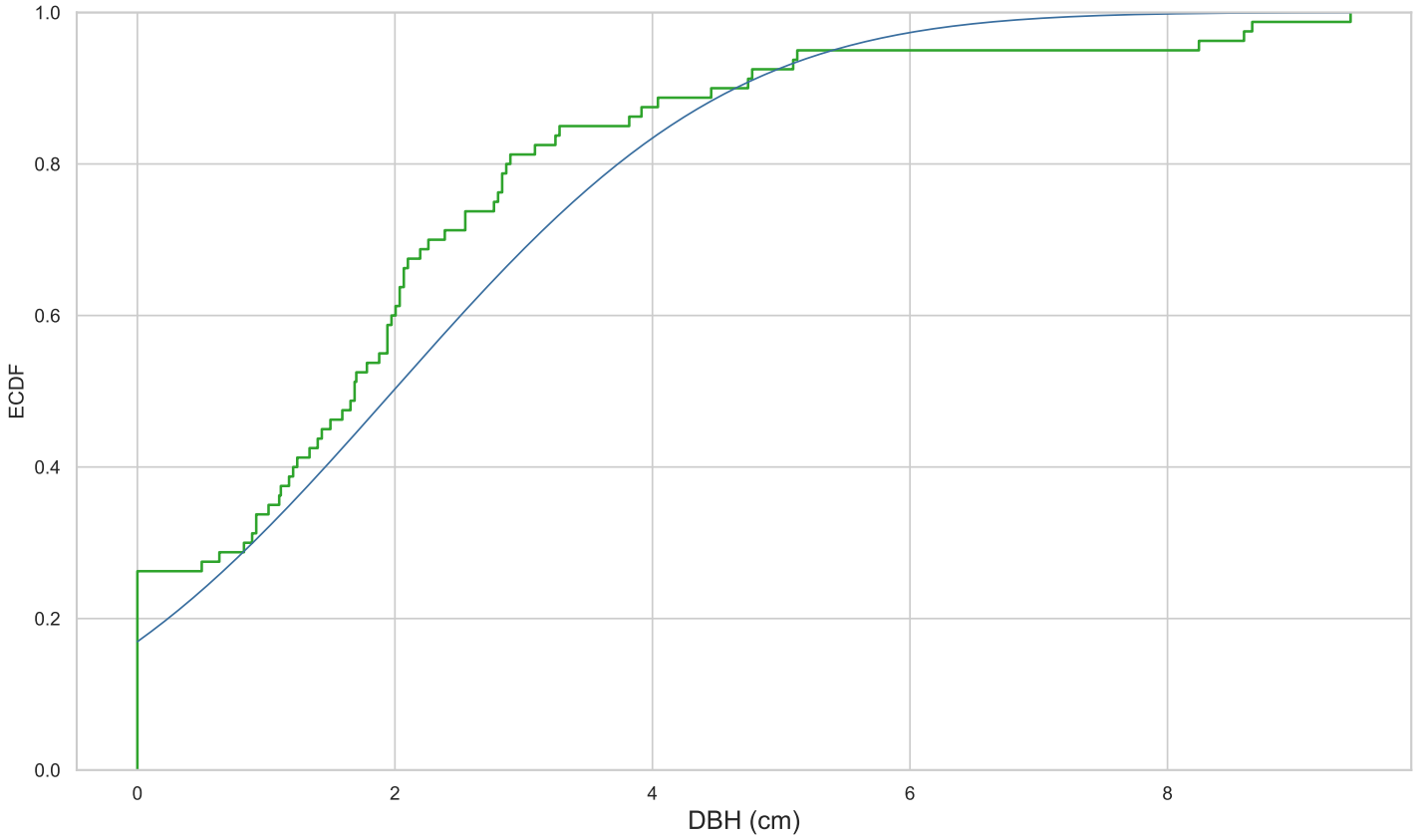


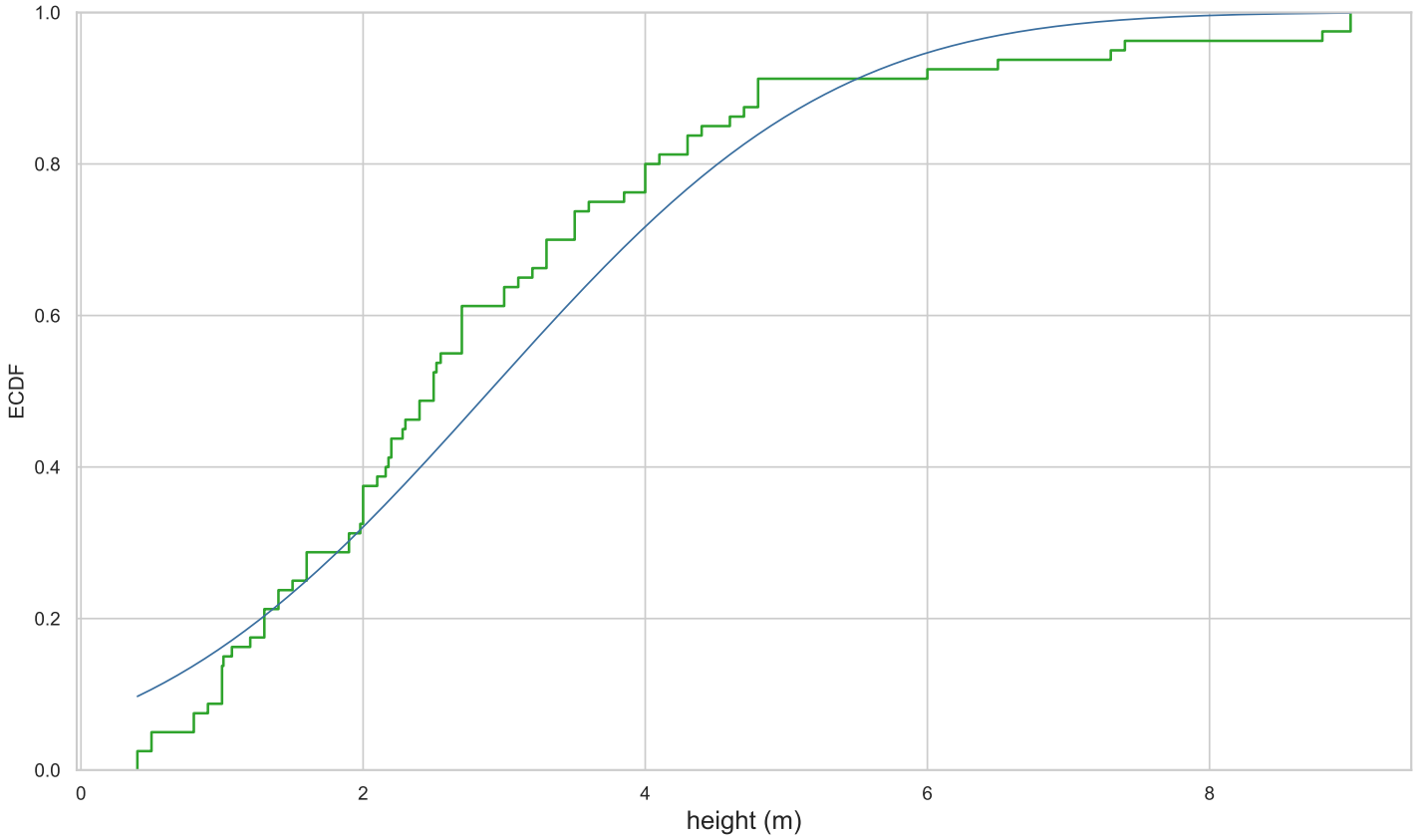

1.0
0.8
0.6
0.4
0.2
0.0
ECDF
100
200
300
400
500
600
1st attack stop


1.0
0.8
0.6
0.4
0.2
0.0
ECDF
0
100
200
300
400
500
600
1st locate


1.0
0.8
0.6
0.4
0.2
0.0
ECDF
100
200
300
400
500
600
2nd attack

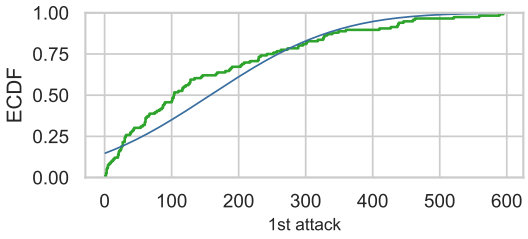

1.0
0.8
0.6
0.4
0.2
0.0
ECDF
0
5
10
15
20
25
30
CBH (cm)

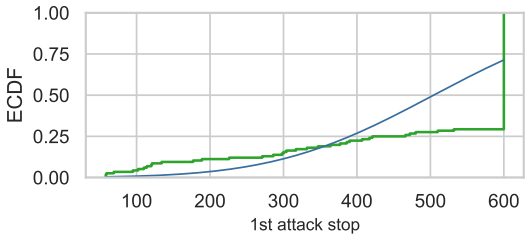

1.0
0.8
0.6
0.4
0.2
0.0
ECDF
0
2
4
6
8
DBH (cm)

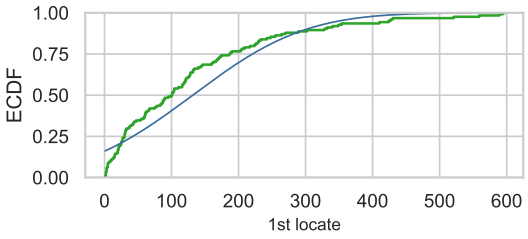

1.0
0.8
0.6
0.4
0.2
0.0
ECDF
0
2
4
6
8
height (m)

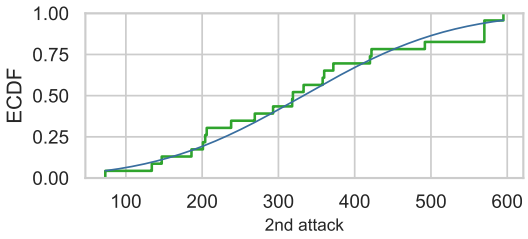
ECDF Plots Summary

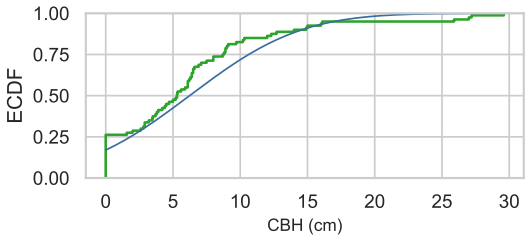
Multiple ECDF Plots of variables in one figure. Variables are sorted alphabetically.

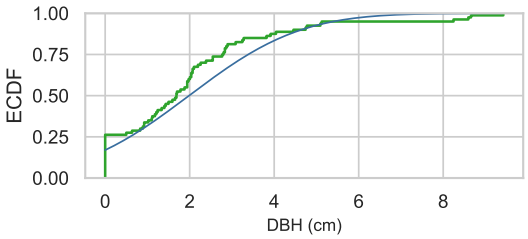


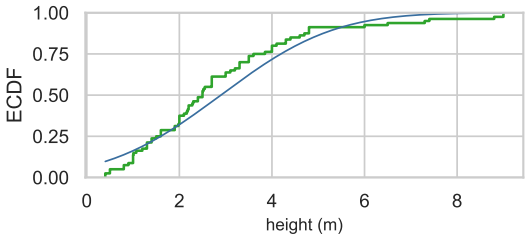








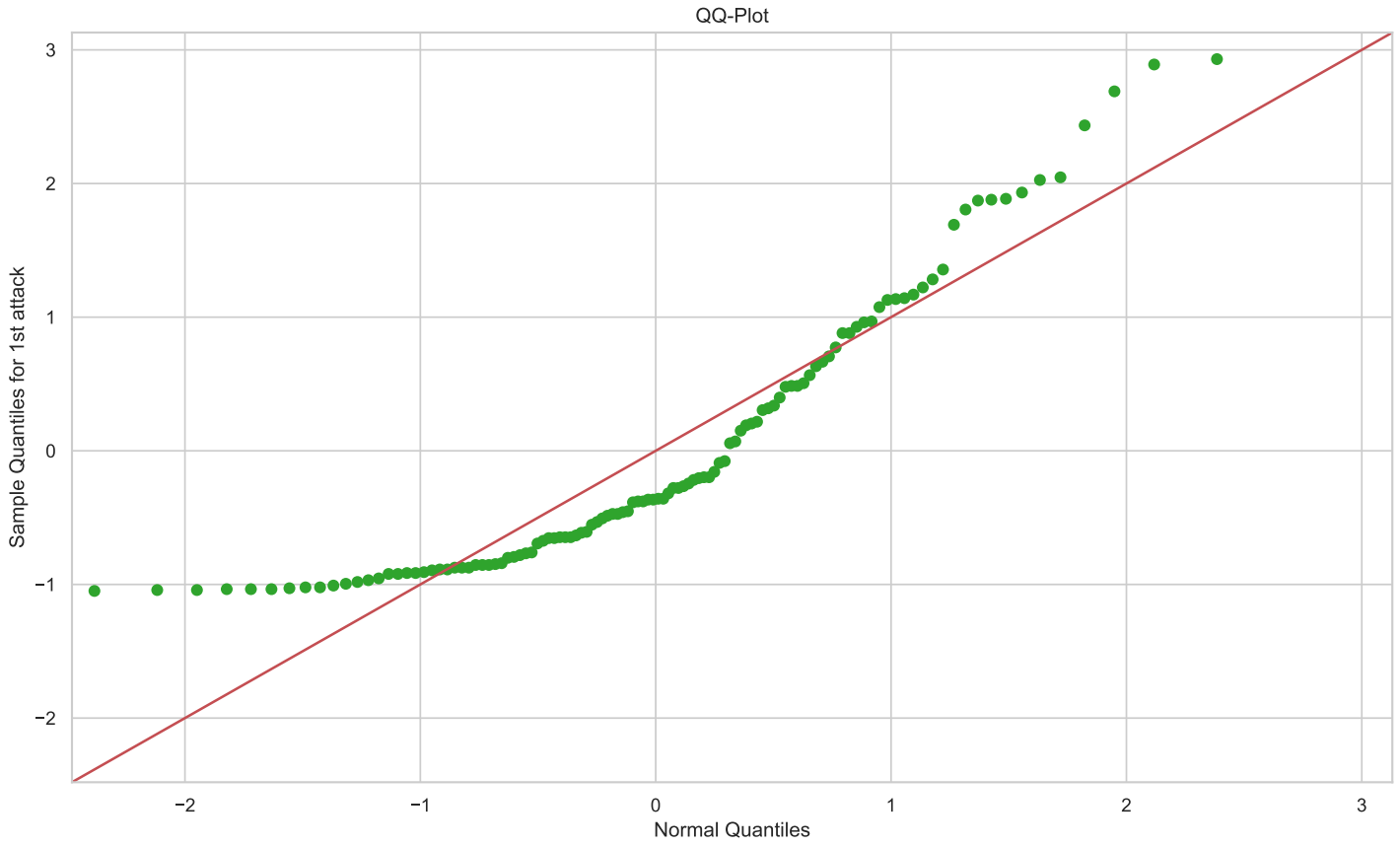


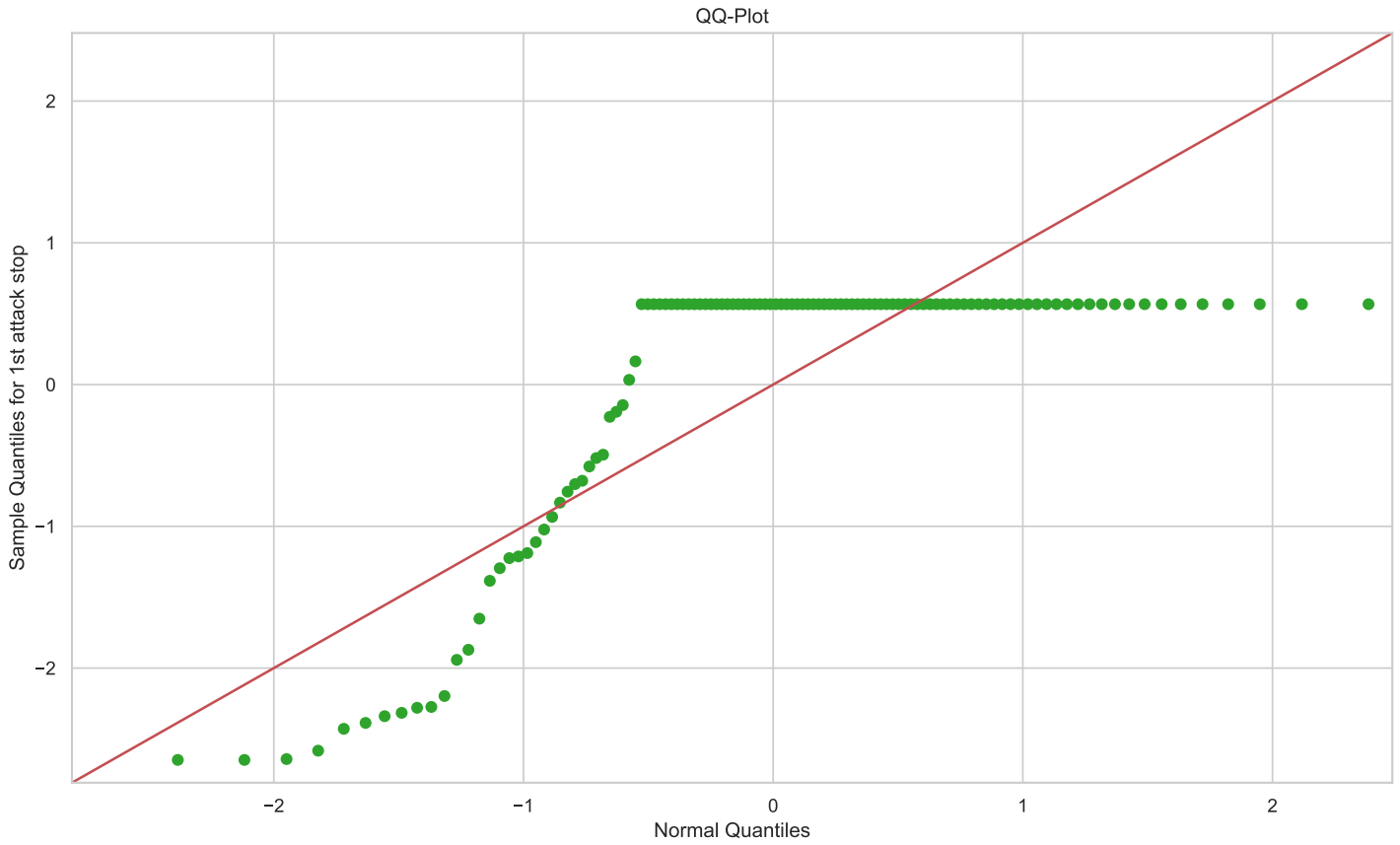


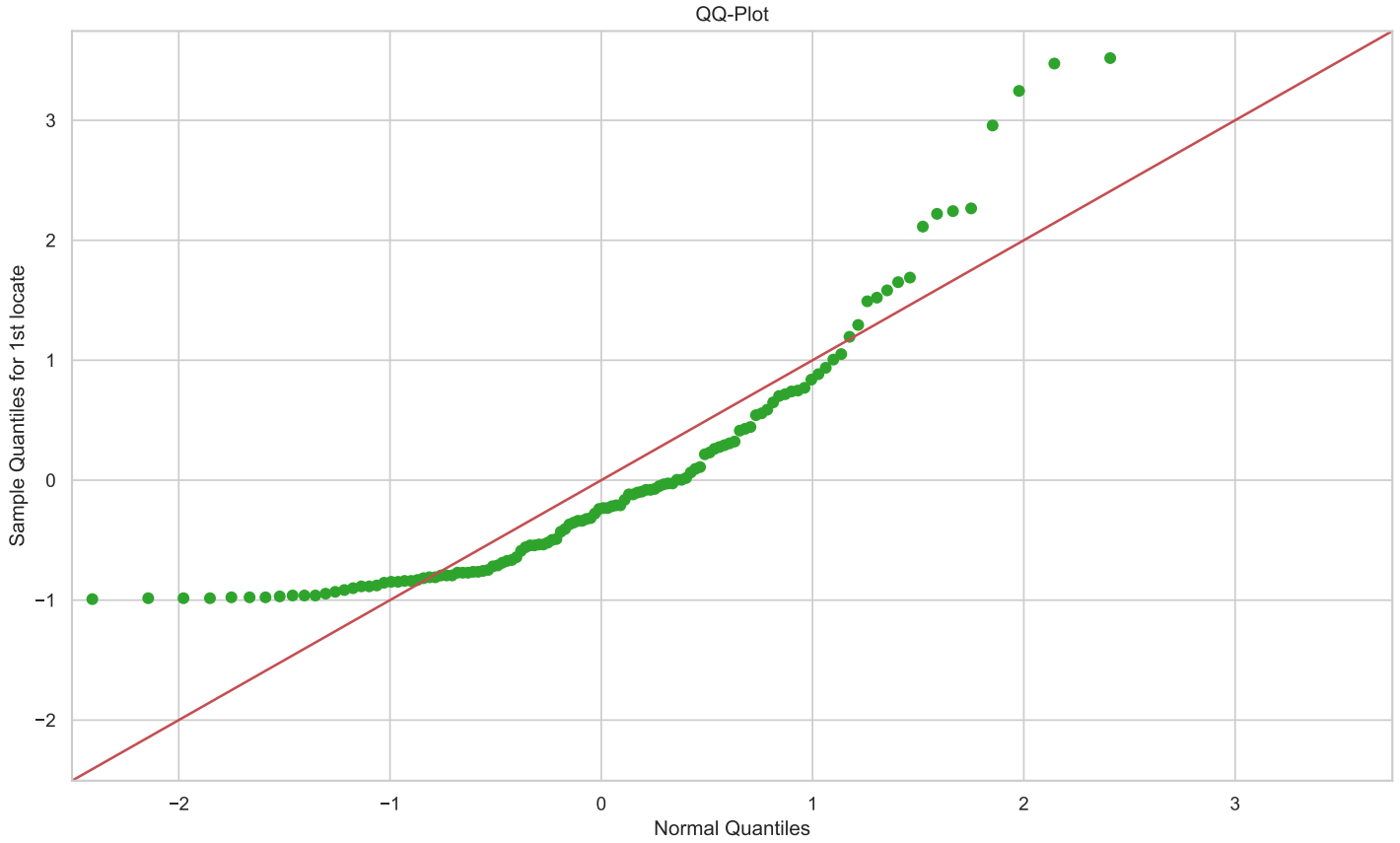
QQ-Plots

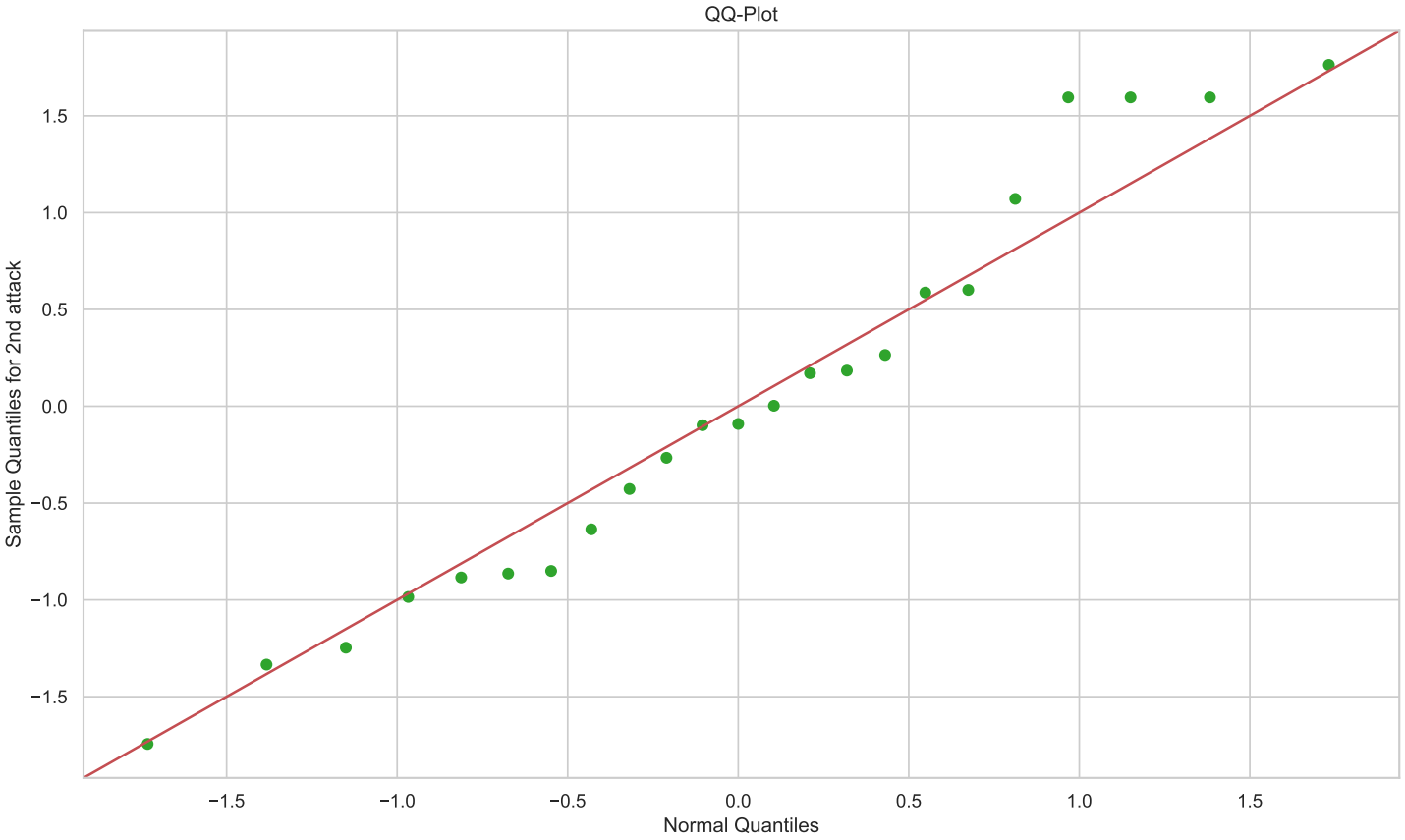
One QQ-Plot per page for each variable. Variables are sorted alphabetically.

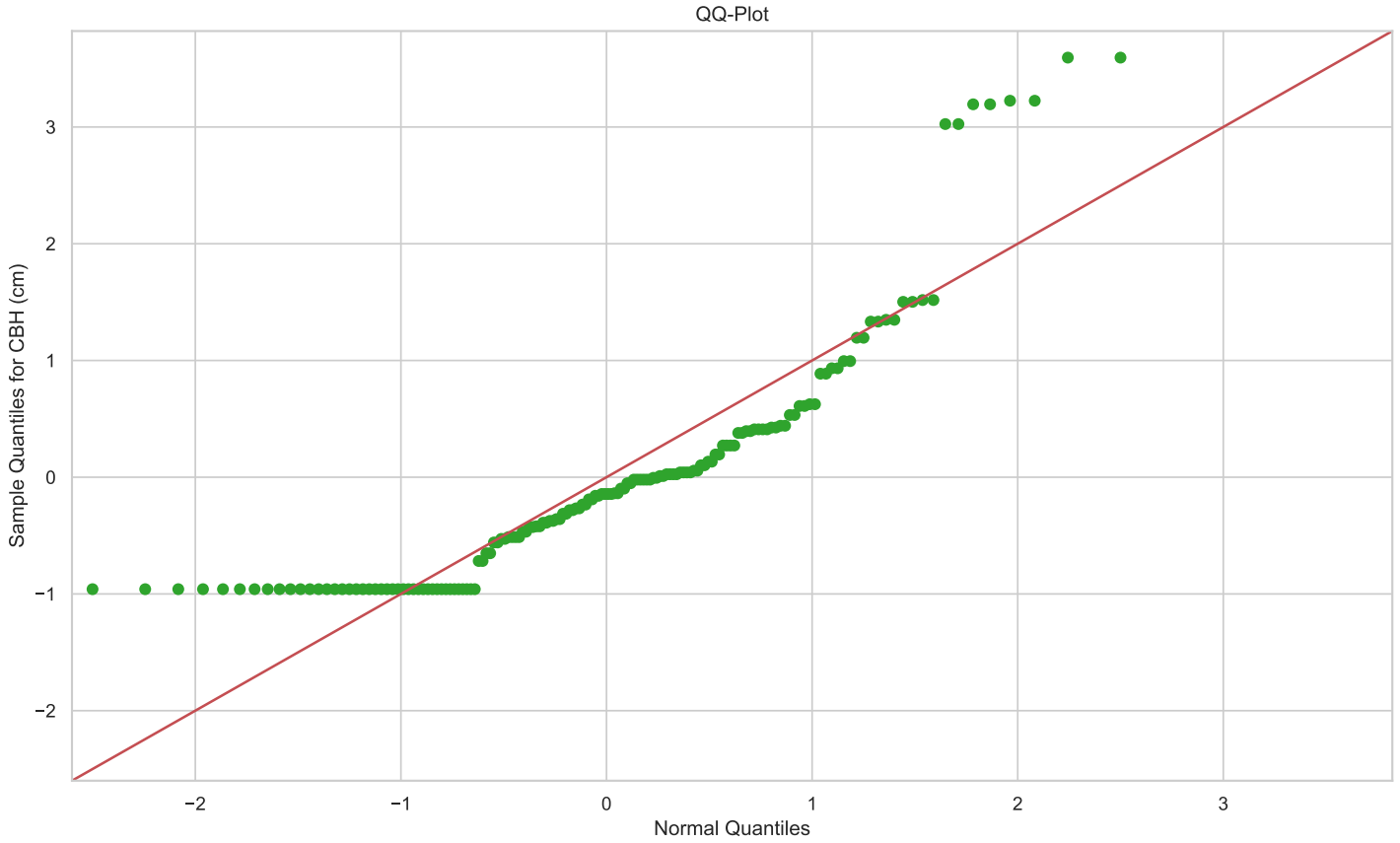
[]

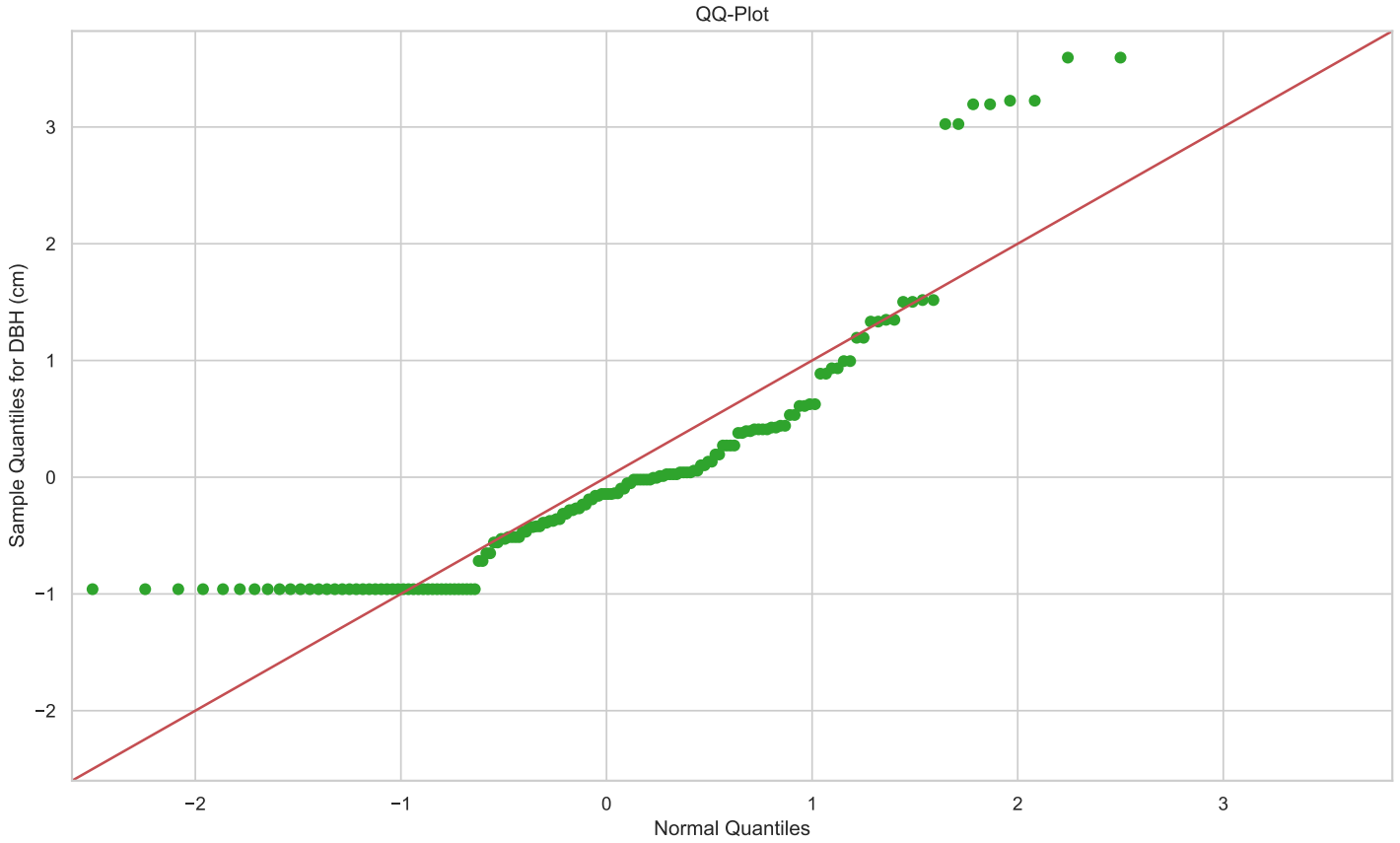


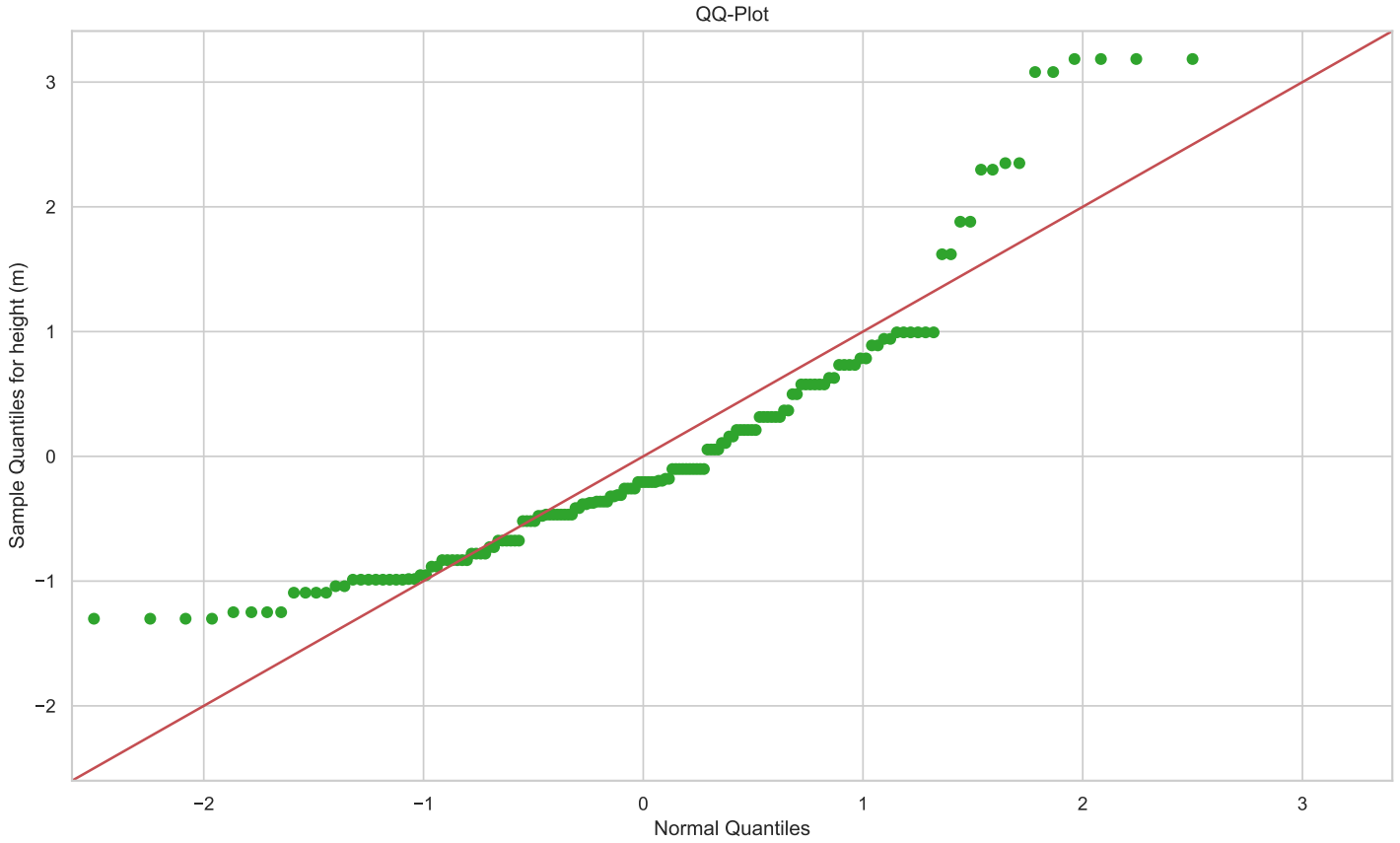

QQ-Plot
Sample Quantiles for 1st attack stop
Normal Quantiles
2
1
0
−1
−2
−2
−1
0
1
2


QQ-Plot
Sample Quantiles for 1st locate
Normal Quantiles
3
2
1
0
−1
−2
−2
−1
0
1
2
3


QQ-Plot
Sample Quantiles for 2nd attack
Normal Quantiles
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

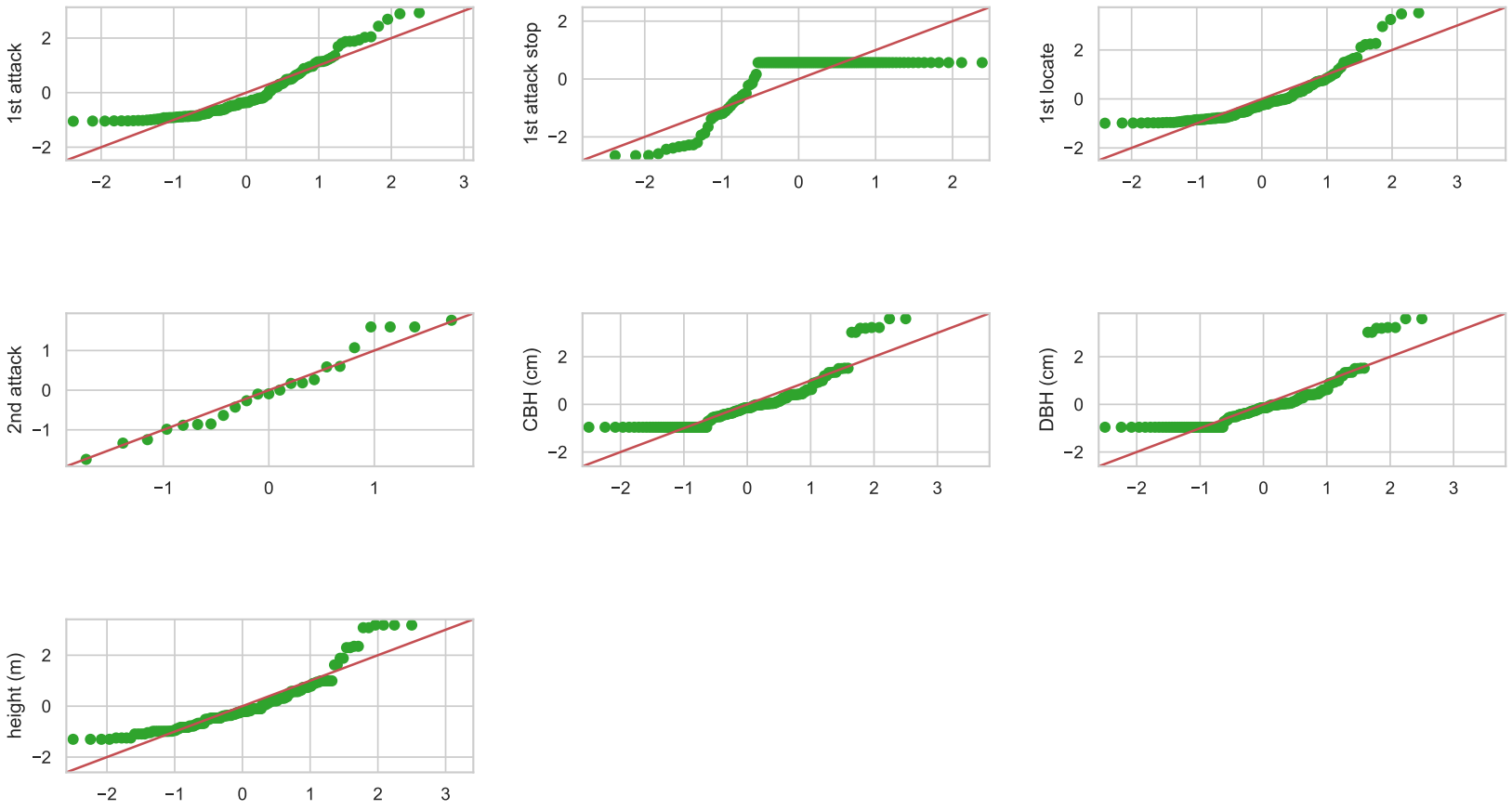

QQ-Plot
Sample Quantiles for CBH (cm)
Normal Quantiles
3
2
1
0
−1
−2
−2
−1
0
1
2
3


QQ-Plot
Sample Quantiles for DBH (cm)
Normal Quantiles
3
2
1
0
−1
−2
−2
−1
0
1
2
3


QQ-Plot
Sample Quantiles for height (m)
Normal Quantiles
3
2
1
0
−1
−2
−2
−1
0
1
2
3

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
Attacked	160	0	160	100	2
Baiting tree no.	160	0	160	100	80
Detected	160	0	160	100	2
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
Recruited	160	0	160	100	2
Termite/C	160	0	160	100	2
Tree number	160	60	160	100	51
ant sample	160	100	160	100	2
date	160	0	160	100	29
elevation (m)	160	0	160	100	8
field notes	160	152	160	100	6
species	160	0	160	100	5
transect	160	0	160	100	3

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.

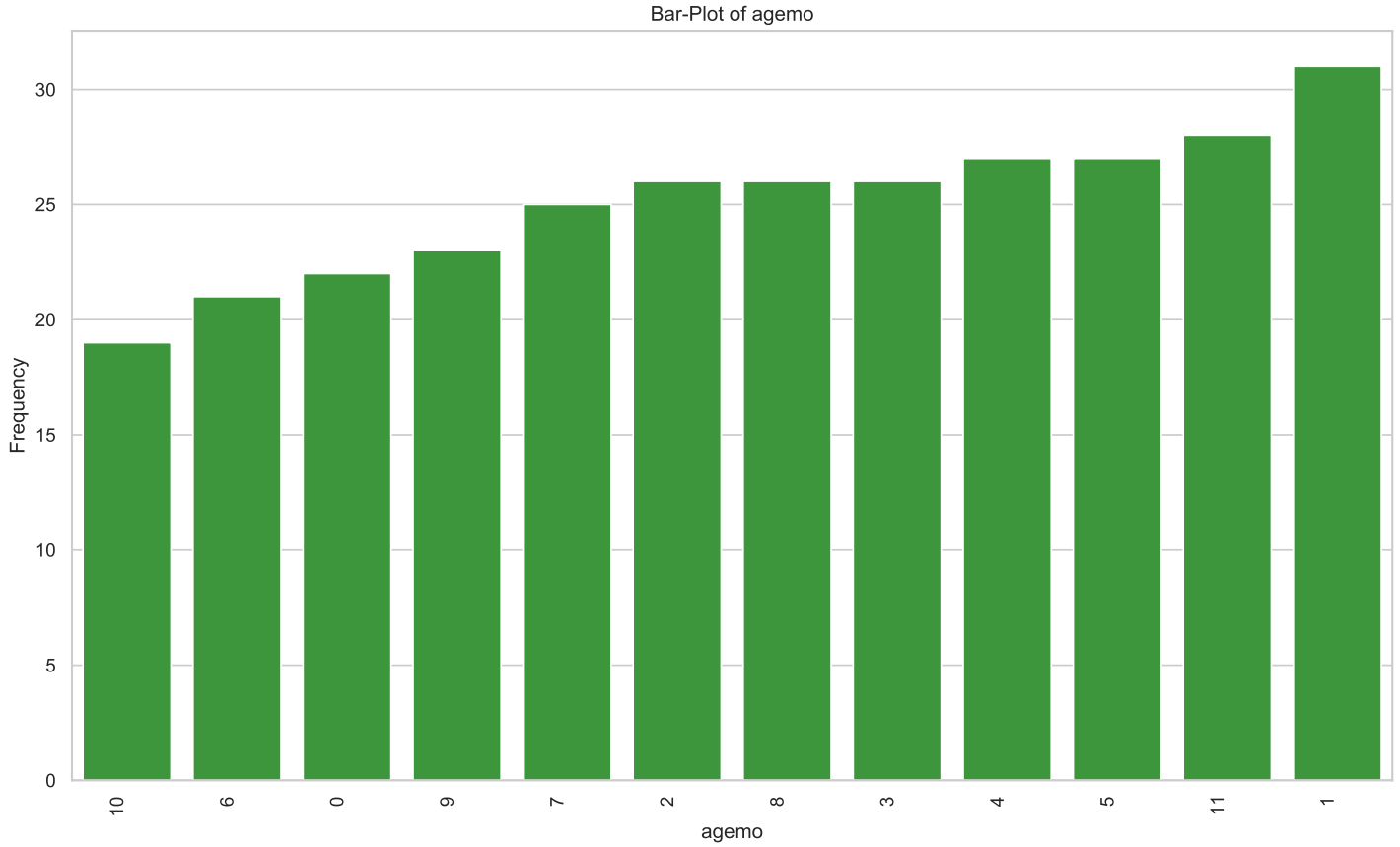
Variable	Category	Frequency	Percent
agemo	1	31	0.10299
agemo	11	28	0.0930233
agemo	5	27	0.089701
agemo	4	27	0.089701
agemo	3	26	0.0863787
agemo	8	26	0.0863787
agemo	2	26	0.0863787
agemo	7	25	0.0830565
agemo	9	23	0.076412
agemo	0	22	0.0730897
agemo	6	21	0.0697674
agemo	10	19	0.0631229
ageyr	13	110	0.365449
ageyr	12	101	0.335548
ageyr	14	55	0.182724
ageyr	15	20	0.0664452
ageyr	11	8	0.0265781
ageyr	16	7	0.0232558
grade	7	157	0.521595
grade	8	143	0.475083
grade	Missin	1	0.00332226
school	Pasteu	156	0.518272
school	Grant-	145	0.481728
sex	2	155	0.51495
sex	1	146	0.48505

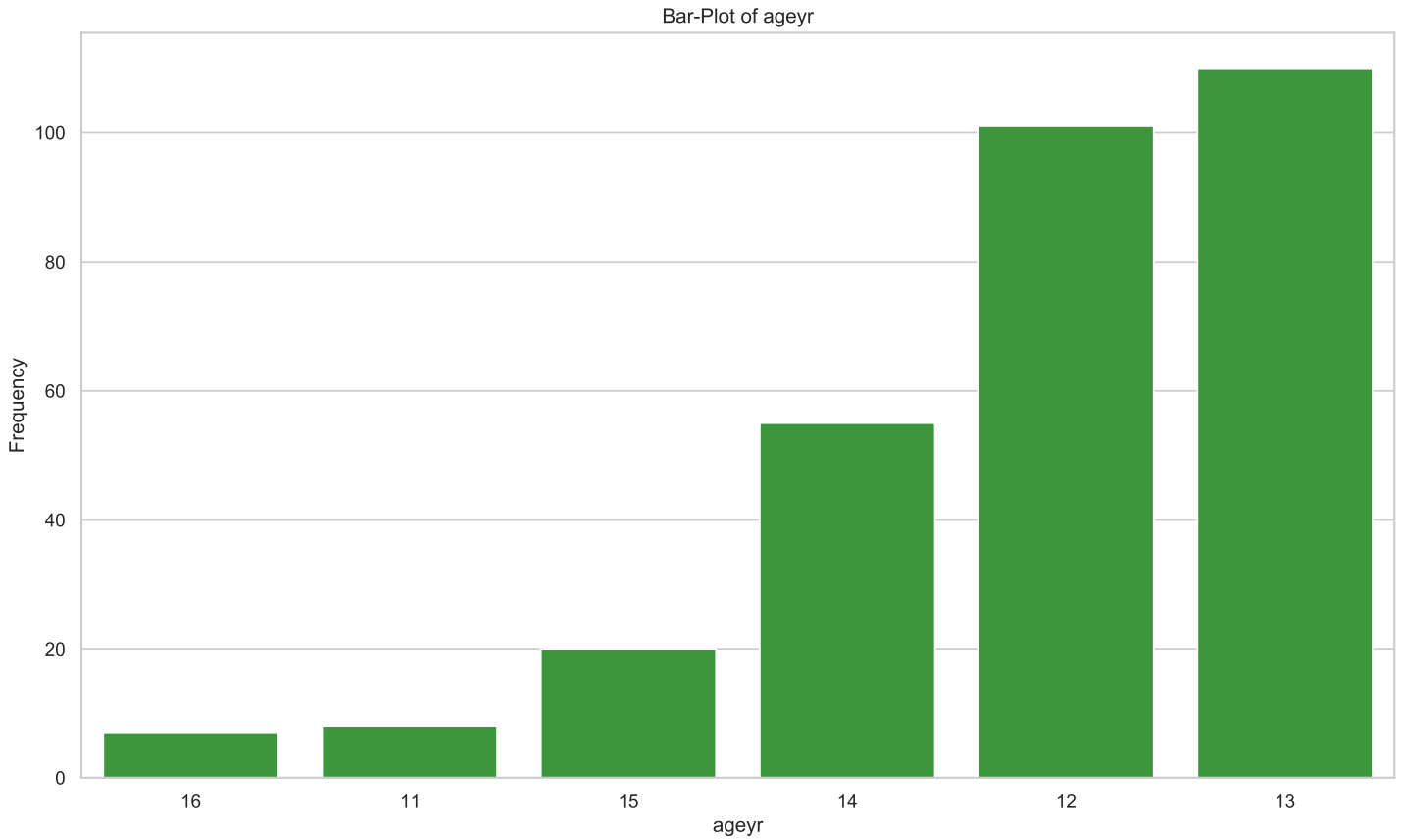
Graphics

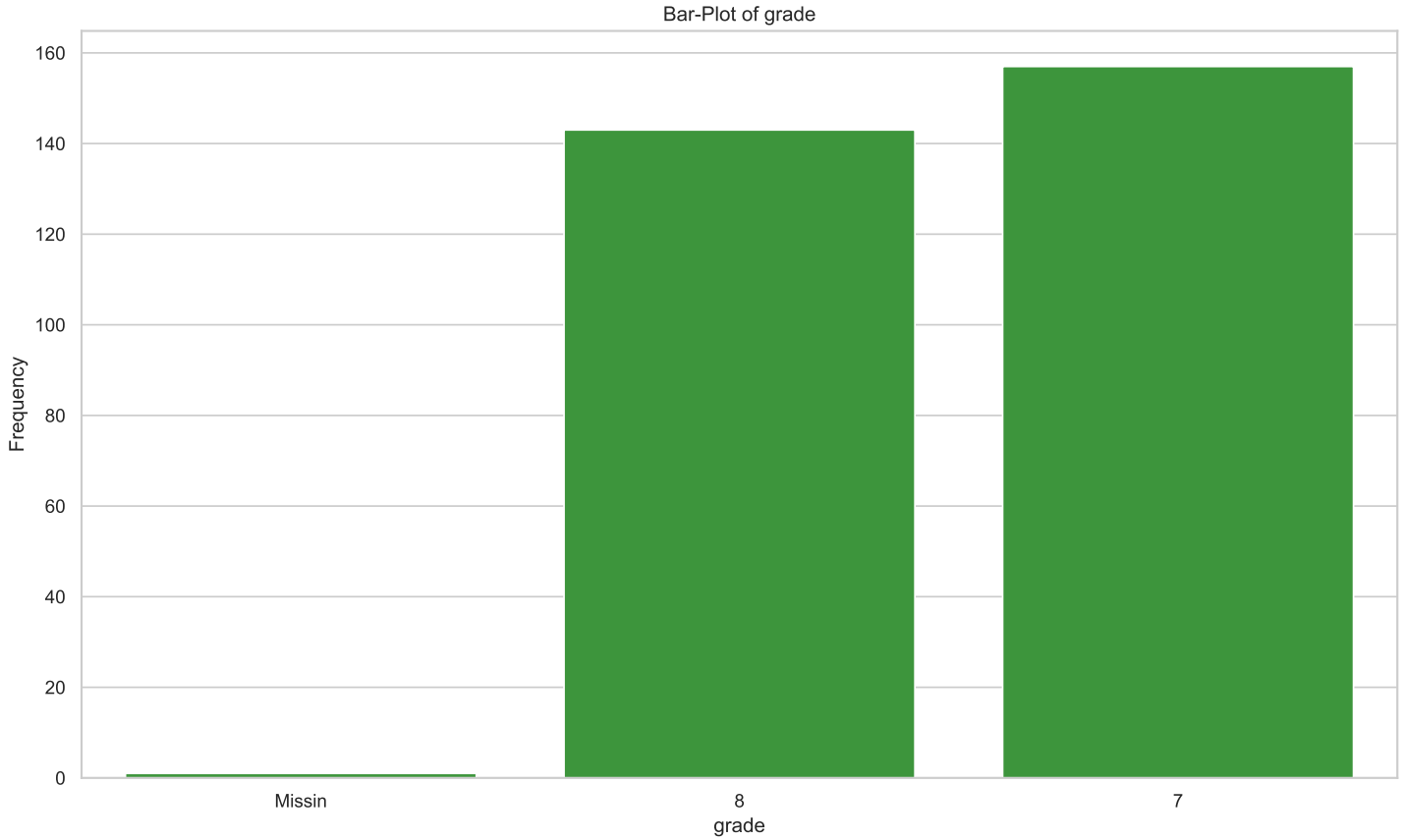
Bar-Plots

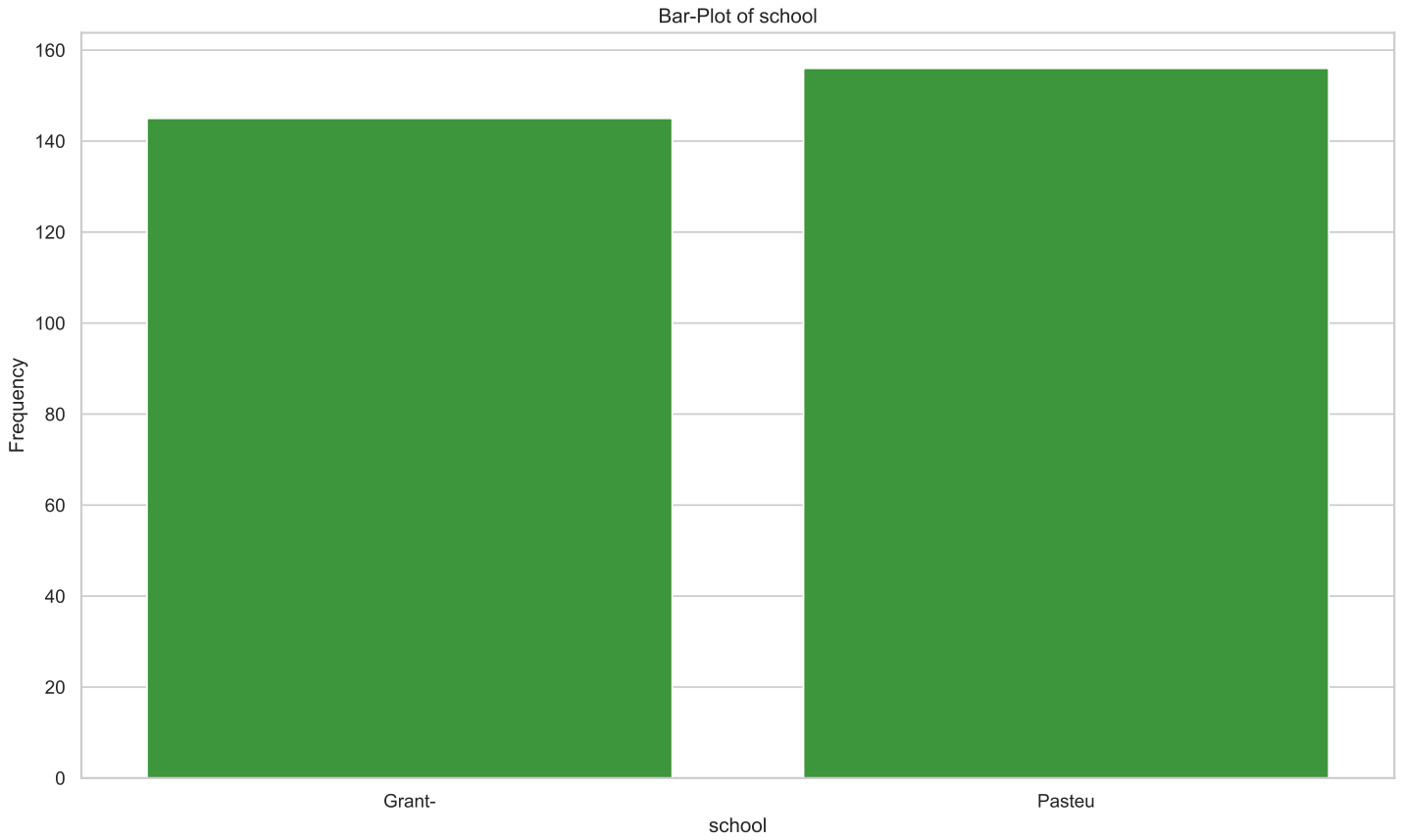
One Bar-Plot per page for each variable. Variables are sorted alphabetically. No labels for variables with more than 40 categories.

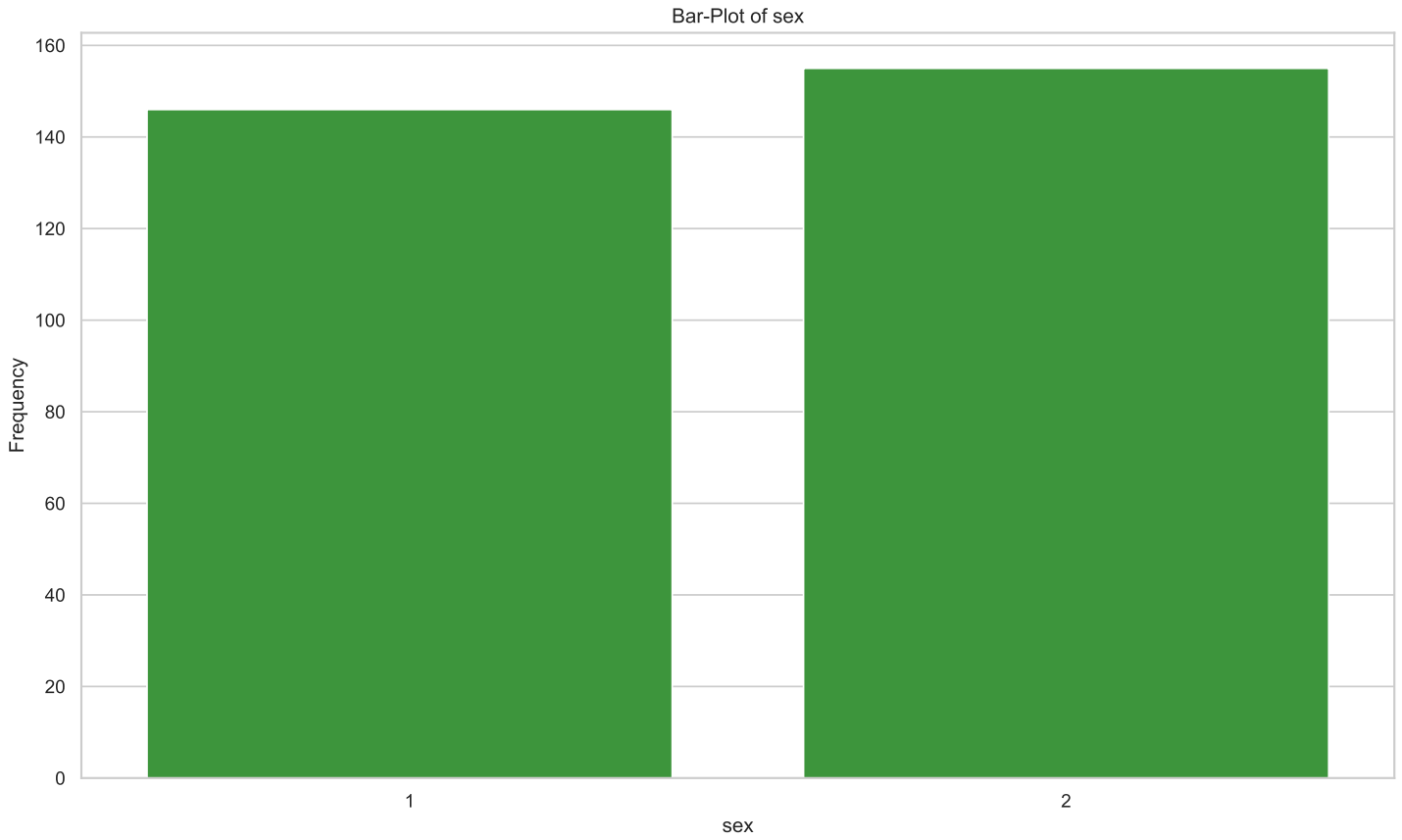
[]




Bar-Plot of ageyr
Frequency
0
20
40
60
80
100
16
11
15
14
12
13
ageyr


Bar-Plot of grade
160
140
120
100
80
60
40
20
0
Frequency
Missin
8
7
grade

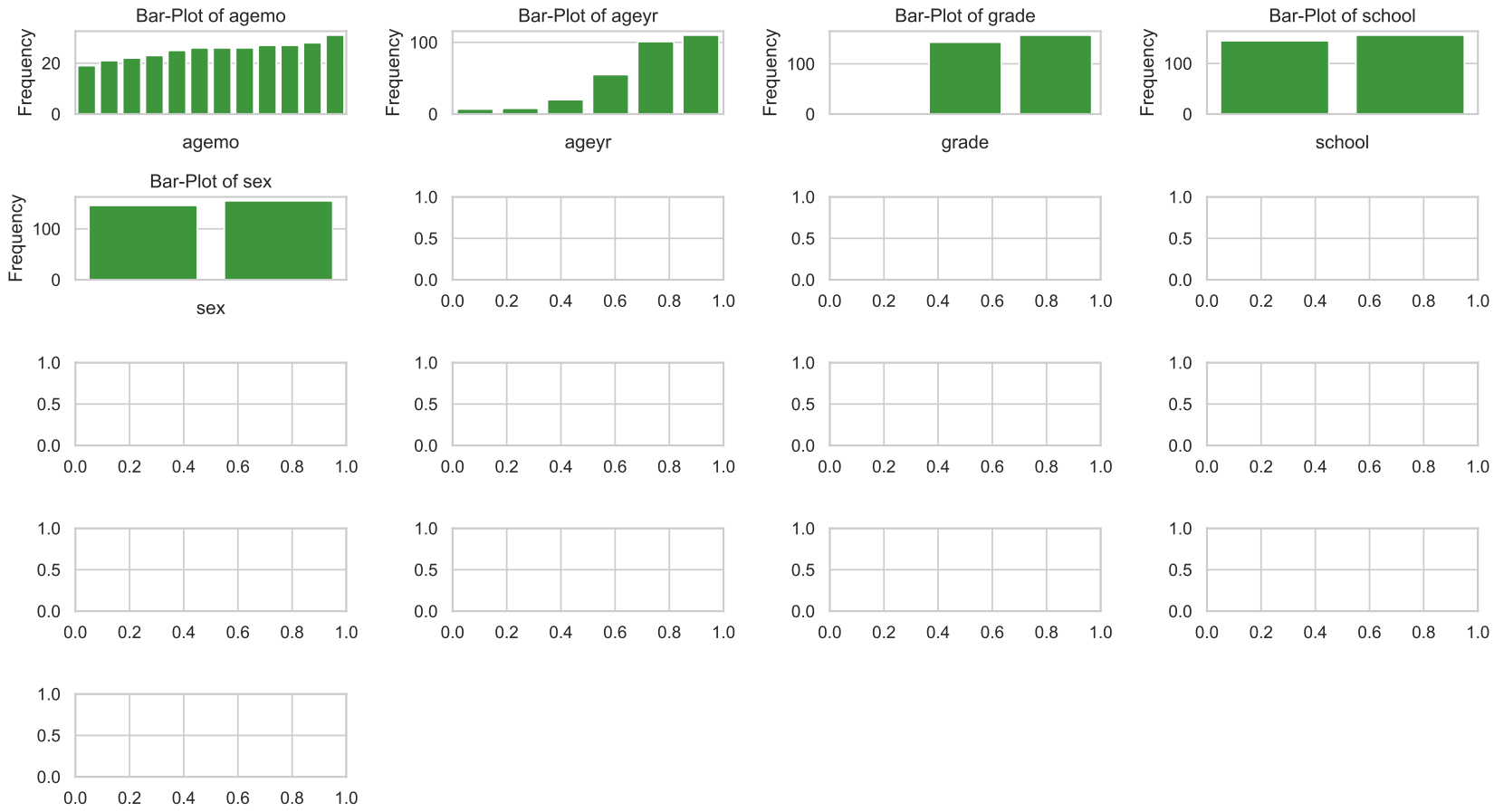

Bar-Plot of school
160
140
120
100
80
60
40
20
0
Frequency
Grant-
Pasteu
school


Bar-Plot of sex
160
140
120
100
80
60
40
20
0
Frequency
1
2
sex

Bar-Plots Summary

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

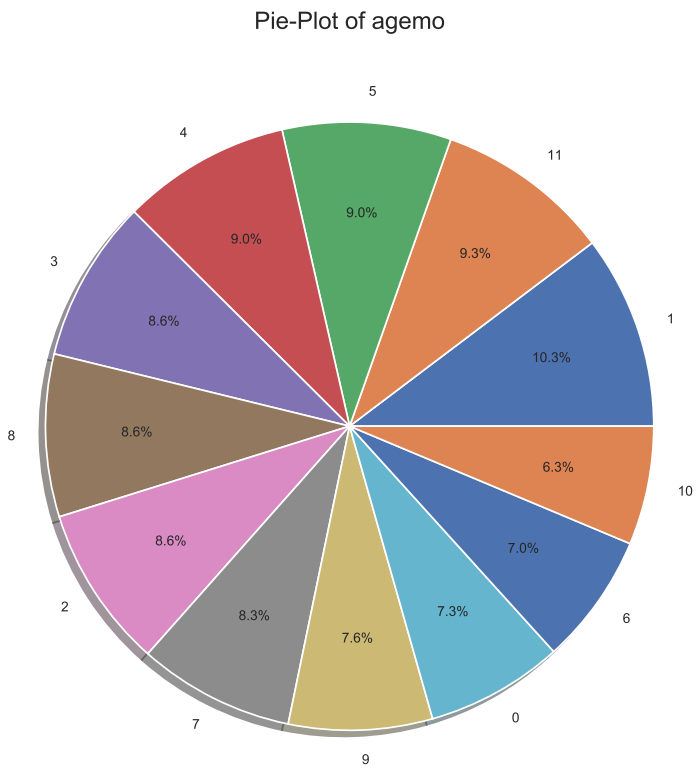
```
Error in py_call_impl(callable, dots$args, dots$keywords): IndexError: index 5 is out of bounds for axis 0 with size 5

Detailed traceback:
  File "<string>", line 2, in <module>
  File "<string>", line 2, in barplot
  File "C:\Users\Denise Welsch\DOCUME~1\.virtualenvs\shiny-app-env\lib\site-packages\pandas\core\indexes\base.py", line 4104, in __getitem__
    return getitem(key)
```

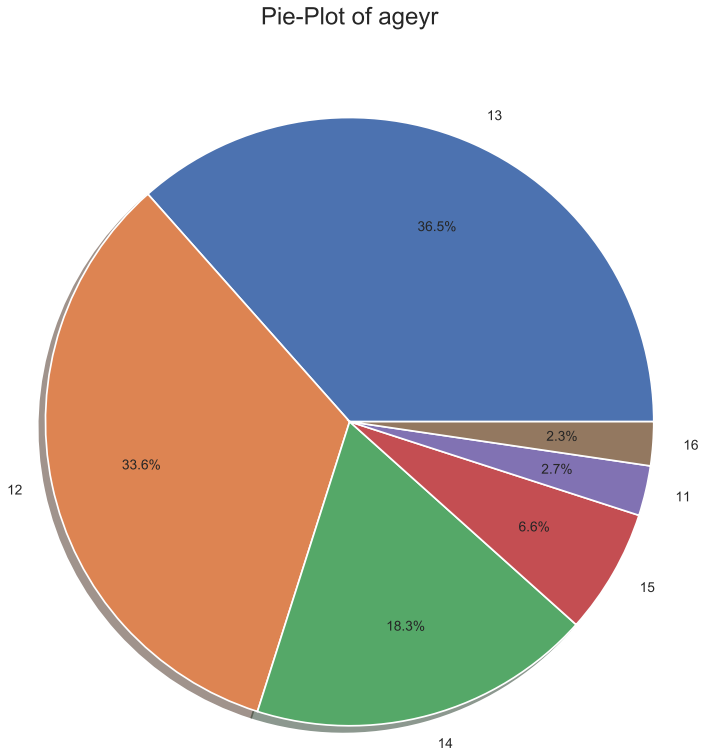

Bar-Plot of agemo
Frequency
0
20
agemo
Bar-Plot of ageyr
Frequency
0
100
ageyr
Bar-Plot of grade
Frequency
0
100
grade
Bar-Plot of school
Frequency
0
100
school
Bar-Plot of sex
Frequency
0
100
sex
0.0
0.5
1.0
0.0
0.2
0.4
0.6
0.8
1.0

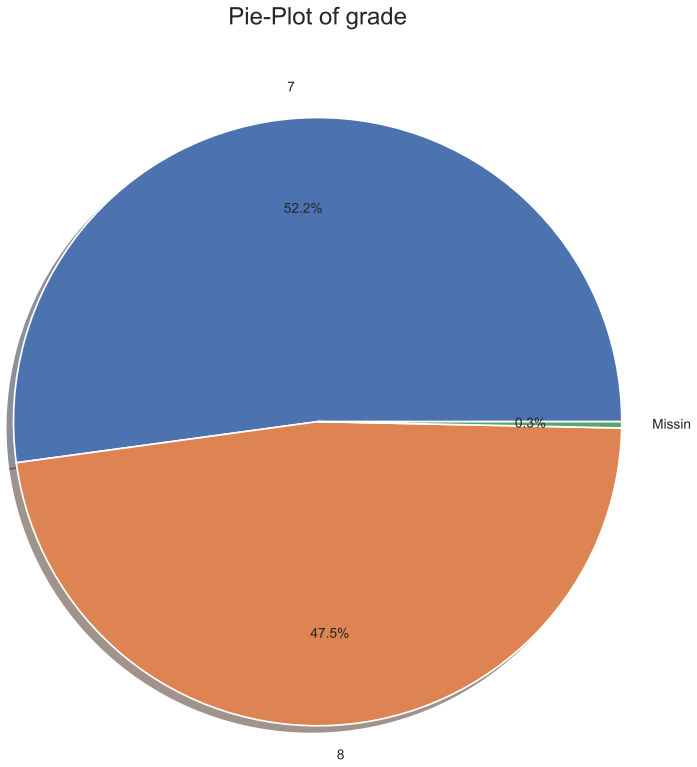
Pie Plots

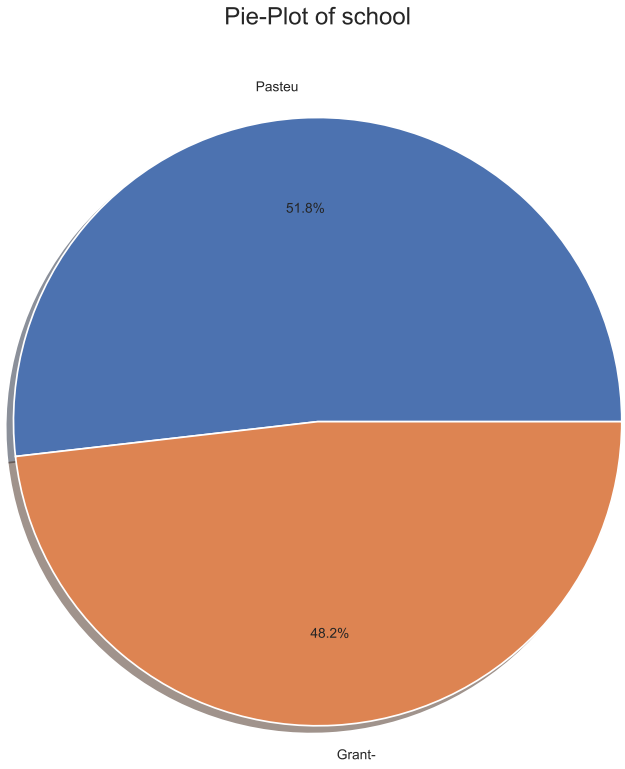
One Pie Plot per page for each variable. Variables are sorted alphabetically.

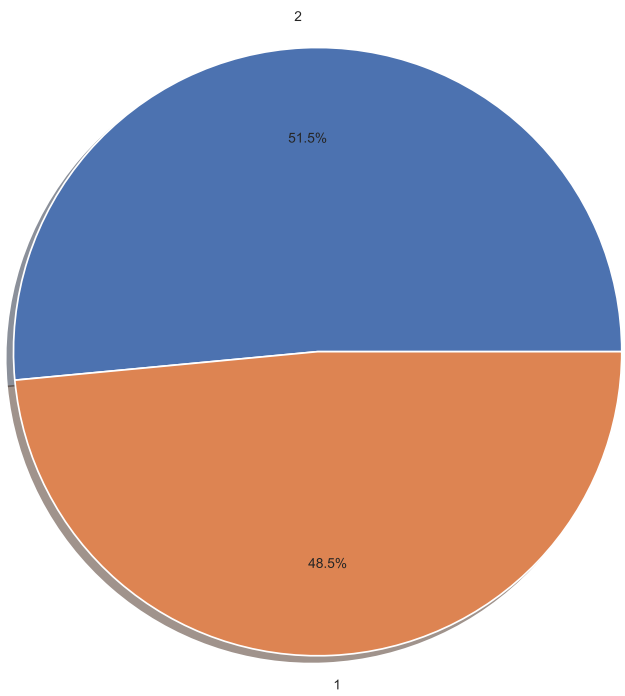
[]




Pie-Plot of ageyr
13
36.5%
12
33.6%
14
18.3%
15
6.6%
11
2.7%
16
2.3%


Pie-Plot of grade
7
52.2%
0.3%
Missin
47.5%
8


Pie-Plot of school
Pasteu
51.8%
48.2%
Grant-

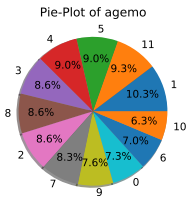

Pie-Plot of sex
2
51.5%
48.5%
1

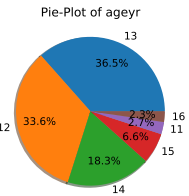
Pie Plots Summary

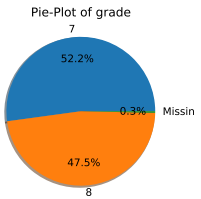
Multiple Pie Plots of variables in one figure. Variables are sorted alphabetically.

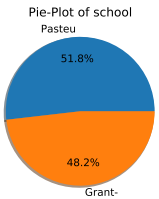
See figures on next page.

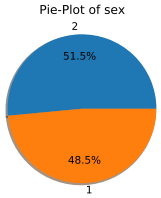
```
Error in py_call_impl(callable, dots$args, dots$keywords): IndexError: single positional indexer is out-of-bounds

Detailed traceback:
  File "<string>", line 2, in <module>
  File "<string>", line 2, in pieplot
  File "C:\Users\Denise Welsch\DOCUME~1\.virtualenvs\shiny-app-env\lib\site-packages\pandas\core\indexing.py", line 873, in __getitem__
    return self._getitem_tuple(key)
  File "C:\Users\Denise Welsch\DOCUME~1\.virtualenvs\shiny-app-env\lib\site-packages\pandas\core\indexing.py", line 1443, in _getitem_tuple
    self._has_valid_tuple(tup)
  File "C:\Users\Denise Welsch\DOCUME~1\.virtualenvs\shiny-app-env\lib\site-packages\pandas\core\indexing.py", line 702, in _has_valid_tuple
    self._validate_key(k, i)
  File "C:\Users\Denise Welsch\DOCUME~1\.virtualenvs\shiny-app-env\lib\site-packages\pandas\core\indexing.py", line 1352, in _validate_key
    self._validate_integer(key, axis)
  File "C:\Users\Denise Welsch\DOCUME~1\.virtualenvs\shiny-app-env\lib\site-packages\pandas\core\indexing.py", line 1437, in _validate_integer
    raise IndexError("single positional indexer is out-of-bounds")
```


Pie-Plot of agemo
5
11
1
10
6
0
7
9
7
2
8
3
4
9.0%
9.3%
10.3%
6.3%
7.0%
7.3%
7.6%
8.3%
8.6%
8.6%
8.6%
9.0%


Pie-Plot of ageyr
13
16
11
15
14
12
36.5%
2.3%
2.7%
6.6%
18.3%
33.6%


Pie-Plot of grade
7
Missin
8
52.2%
0.3%
47.5%


Pie-Plot of school
Pasteu
Grant-
51.8%
48.2%


Pie-Plot of sex
2
1
51.5%
48.5%