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): 25
Variables considered continuous: 7

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

Variables considered categorical: 18

Variables considered categorical
date
transect
Tree number
Baiting tree no.

(continued)

Variables considered categorical
Termite/C
Detected
Attacked
Recruited
H: 0
H: 1-5%
H: 5-33%
H: 33+%
ant sample
field notes
species
lab notes
elevation (m)
2nd attack stop

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

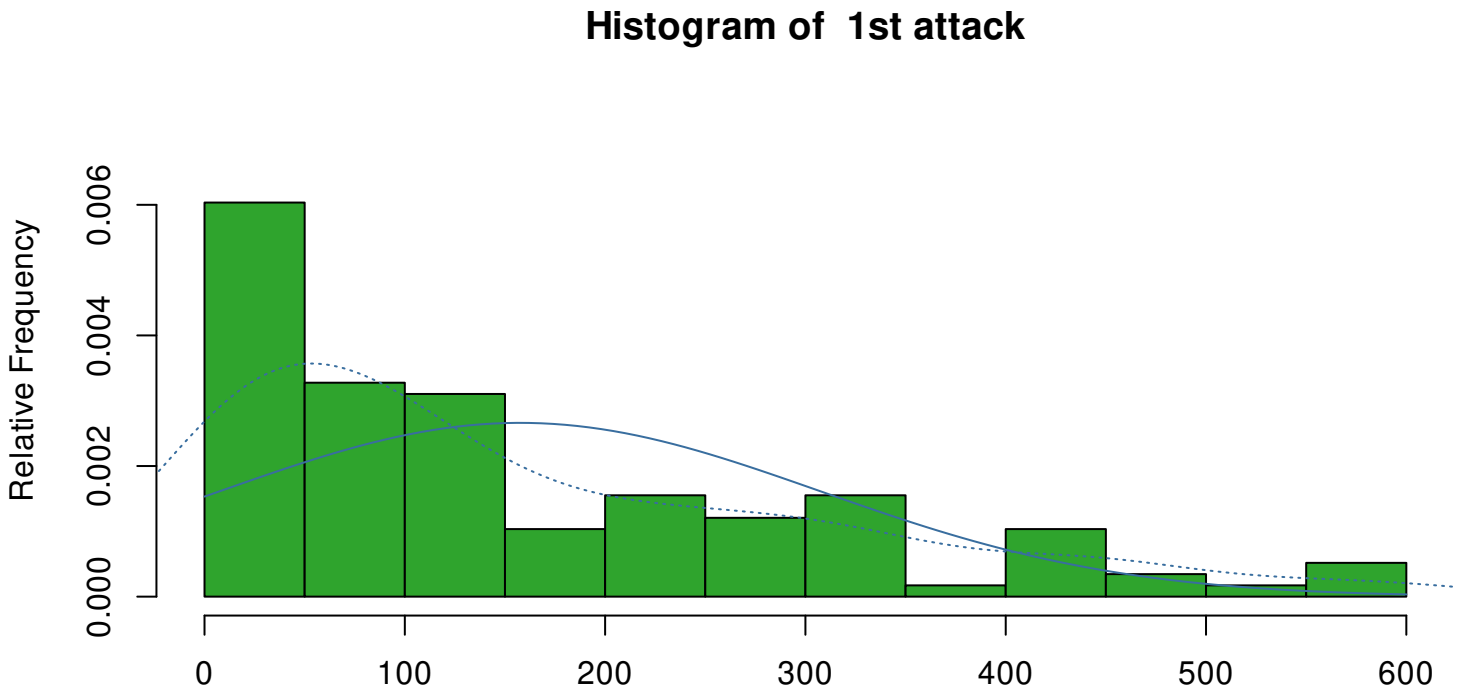
Variables are sorted alphabetically. Missings are omitted in the stats. CV only for positive variables.

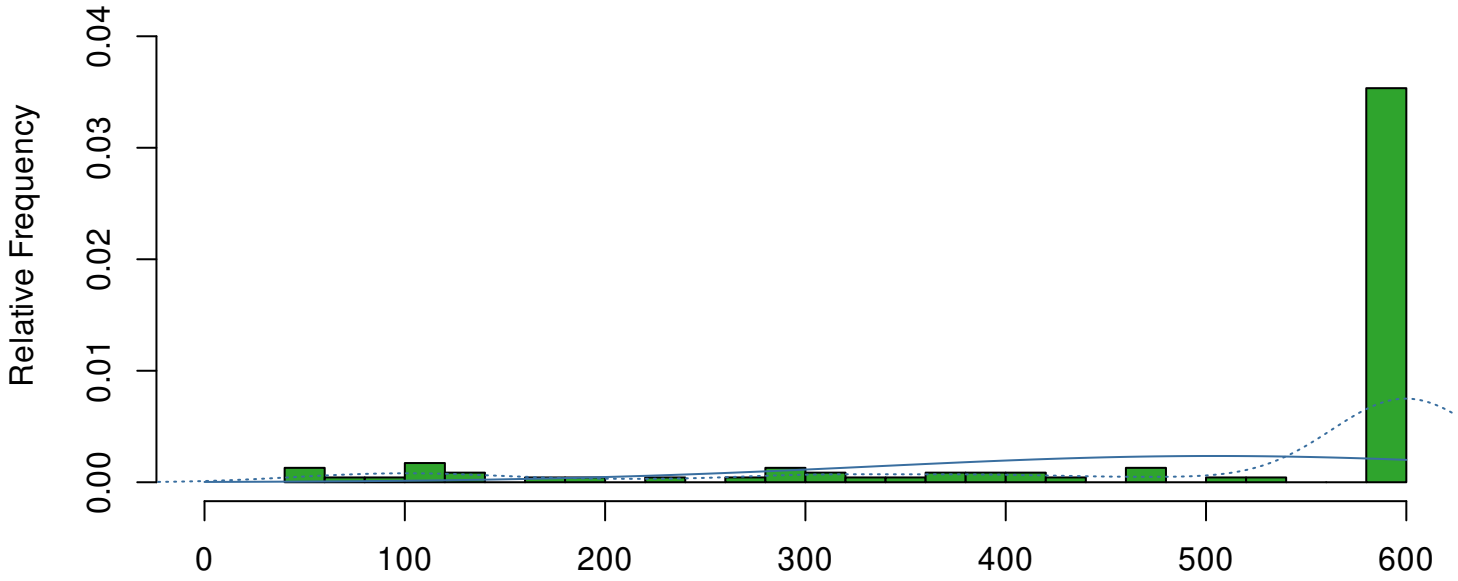
Variable	N Obs	N Missing	N Valid	% Complete	N Unique	Mean	SD	Median	MAD	MIN	MAX	Skewness	Kurtosis	CV
1st attack	160	44	116	72.50	94	157.53	149.91	103.50	122.31	1.0	595.00	1.07	0.27	0.95
1st attack stop	160	44	116	72.50	34	504.39	169.32	600.00	0.00	58.0	600.00	-1.52	0.85	0.34
1st locate	160	36	124	77.50	100	131.59	132.23	100.50	108.97	1.0	595.00	1.52	2.16	1
2nd attack	160	137	23	14.38	21	332.61	152.19	319.00	167.53	73.0	595.00	0.26	-1.08	0.46
CBH (cm)	160	0	160	100.00	52	6.23	6.52	5.30	5.34	0.0	29.60	1.70	3.22	-
DBH (cm)	160	0	160	100.00	52	1.98	2.08	1.69	1.70	0.0	9.42	1.70	3.22	-
height (m)	160	0	160	100.00	46	2.90	1.92	2.50	1.56	0.4	9.00	1.40	1.99	0.66

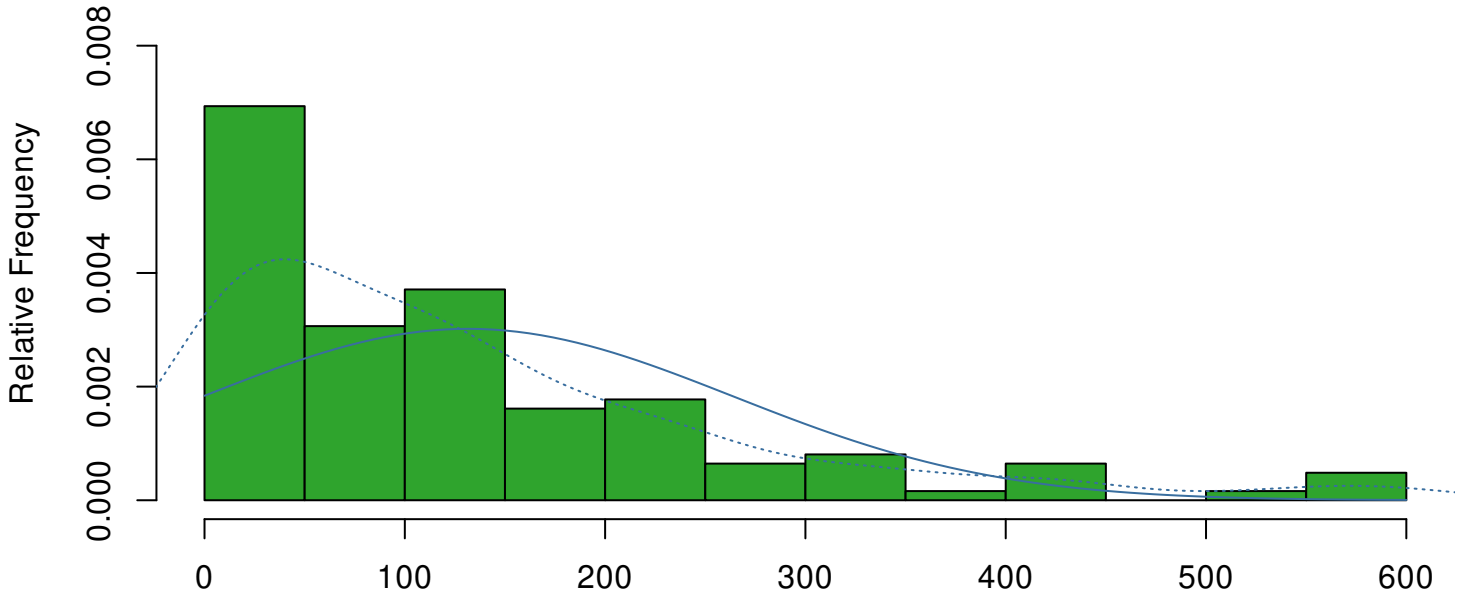
Graphics

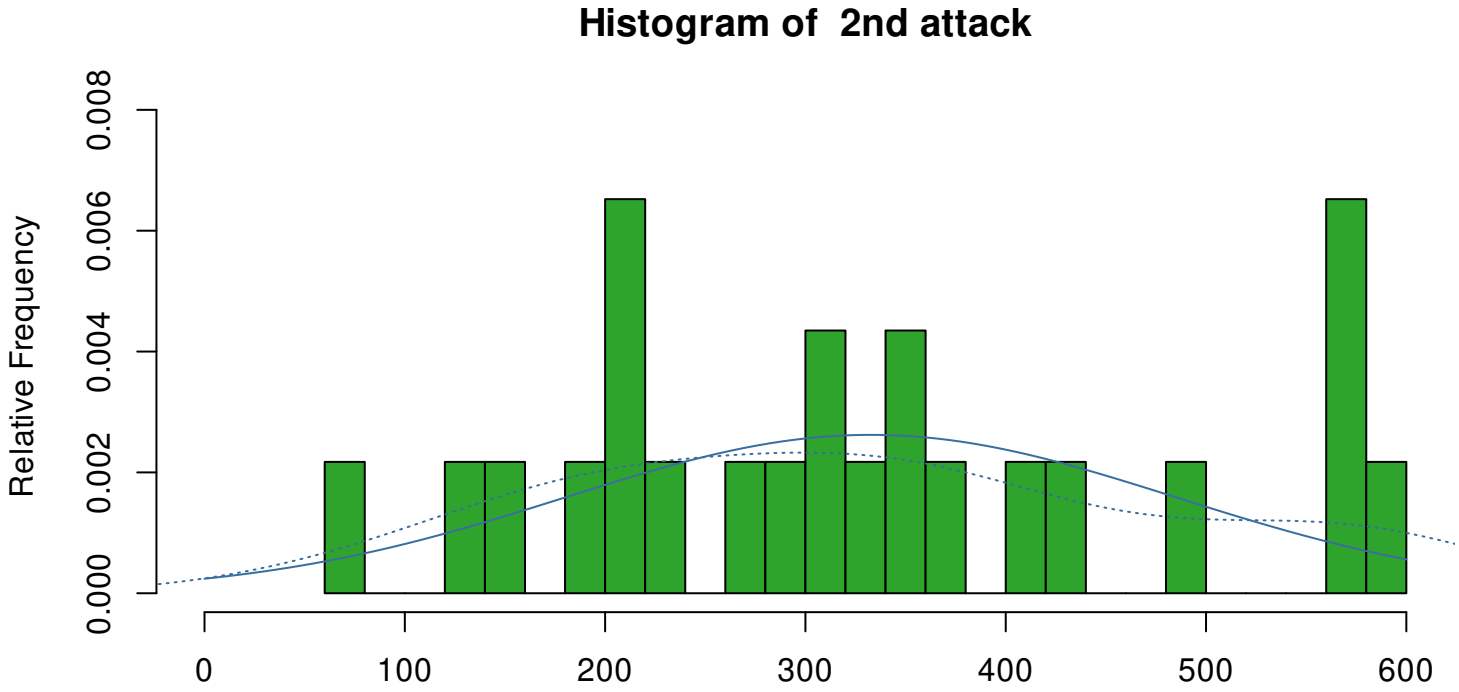
Histograms

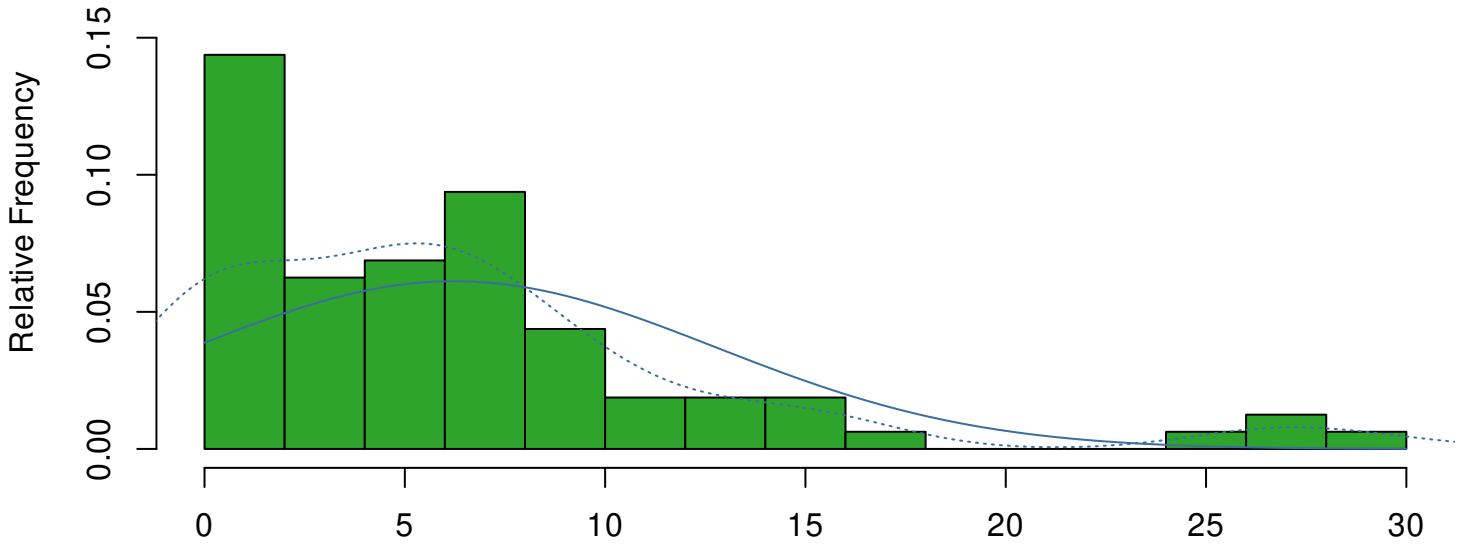
One Relative Frequency Histogram per page for each variable. Variables are sorted alphabetically. The blue line represents the normal density approximation. The blue dotted line represents a special kernel density approximation.

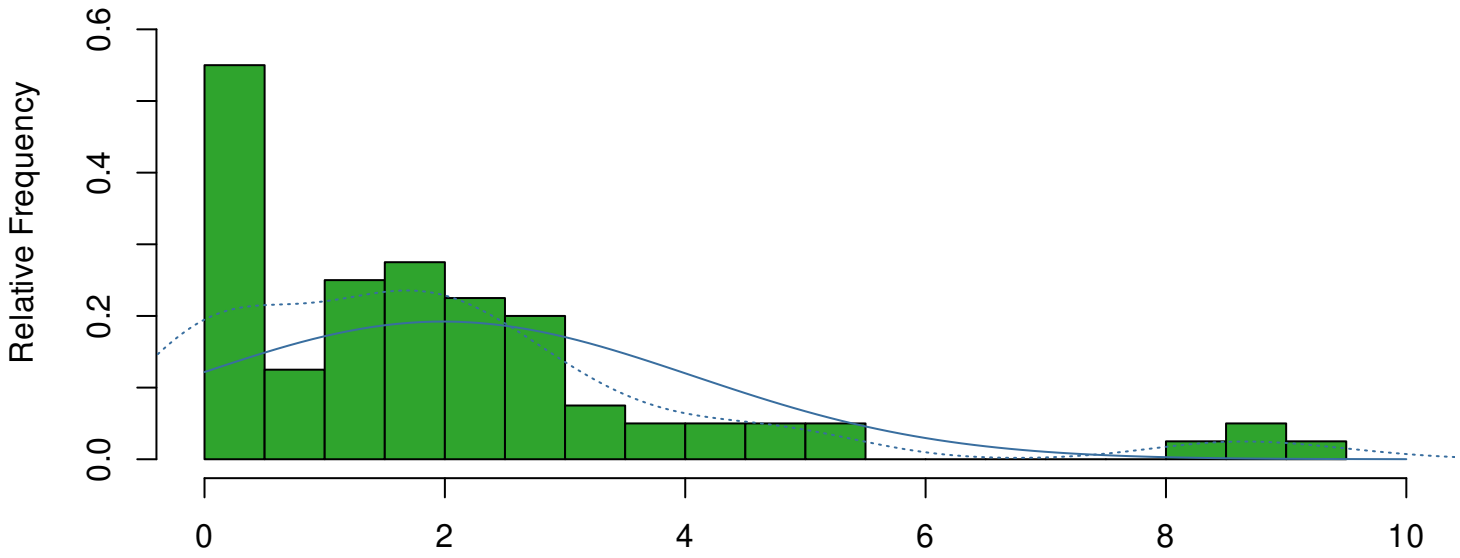


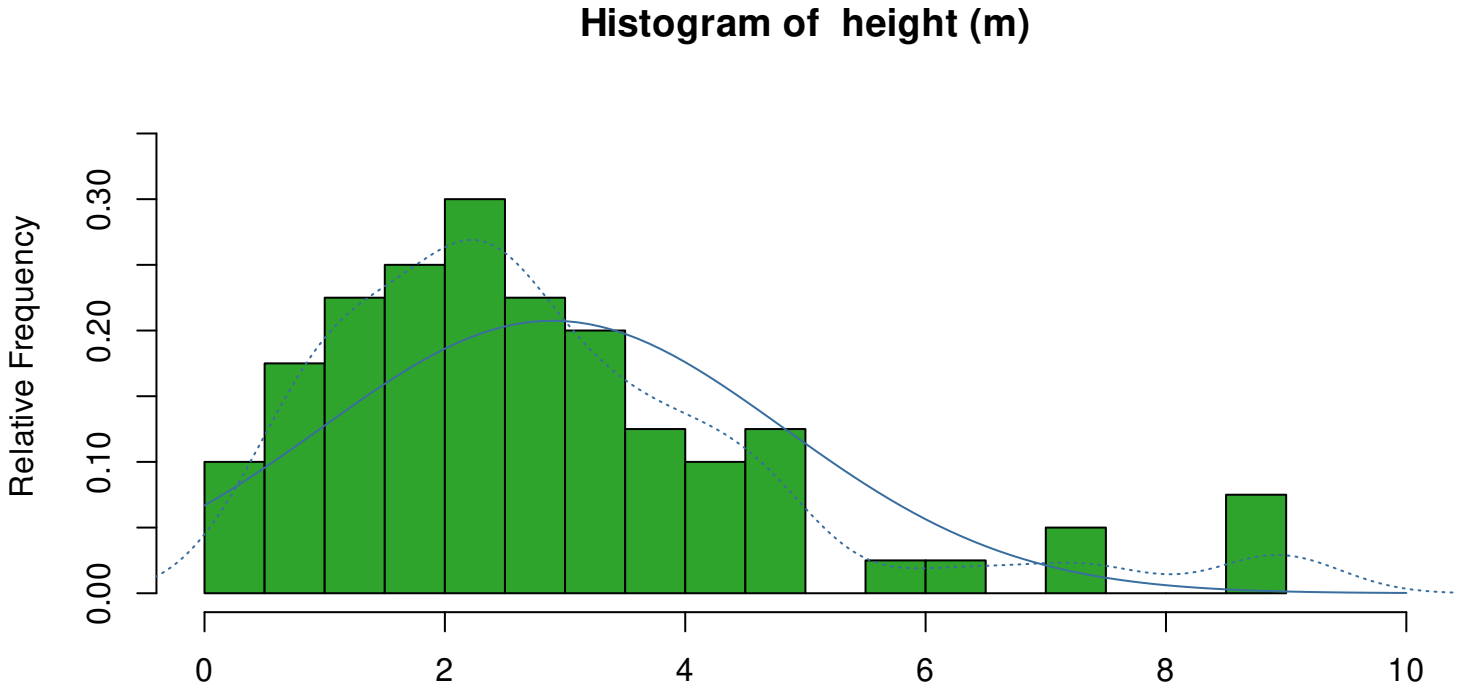

Histogram of 1st attack stop
Relative Frequency
0.00
0.01
0.02
0.03
0.04
0
100
200
300
400
500
600


Histogram of 1st locate
Relative Frequency
0.000
0.002
0.004
0.006
0.008
0
100
200
300
400
500
600


Histogram of 2nd attack
Relative Frequency
0.000
0.002
0.004
0.006
0.008
0
100
200
300
400
500
600

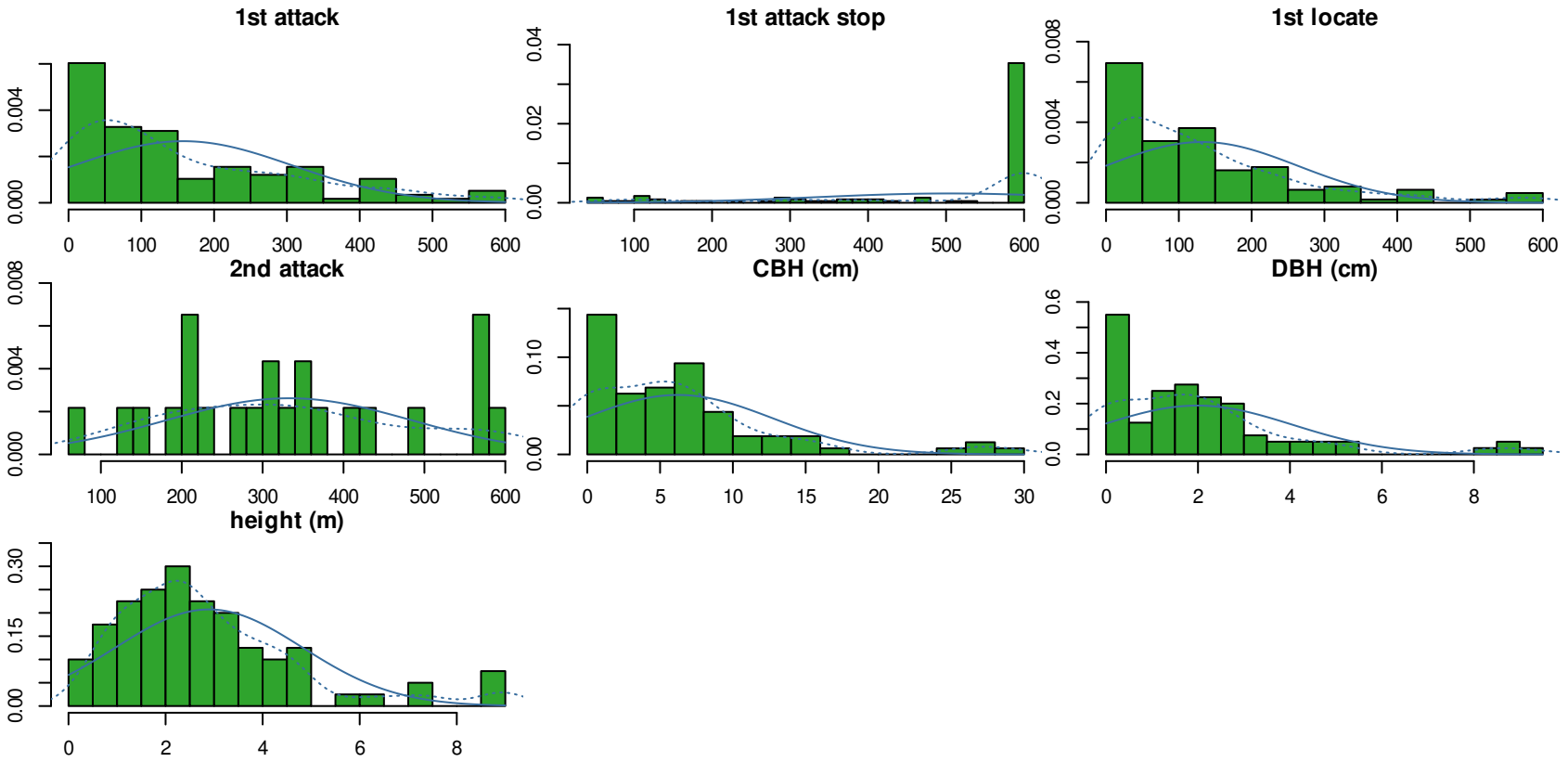

Histogram of CBH (cm)
Relative Frequency
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30


Histogram of DBH (cm)
Relative Frequency
0.0
0.2
0.4
0.6
0
2
4
6
8
10


Histogram of height (m)
Relative Frequency
0.00
0.10
0.20
0.30
0
2
4
6
8
10

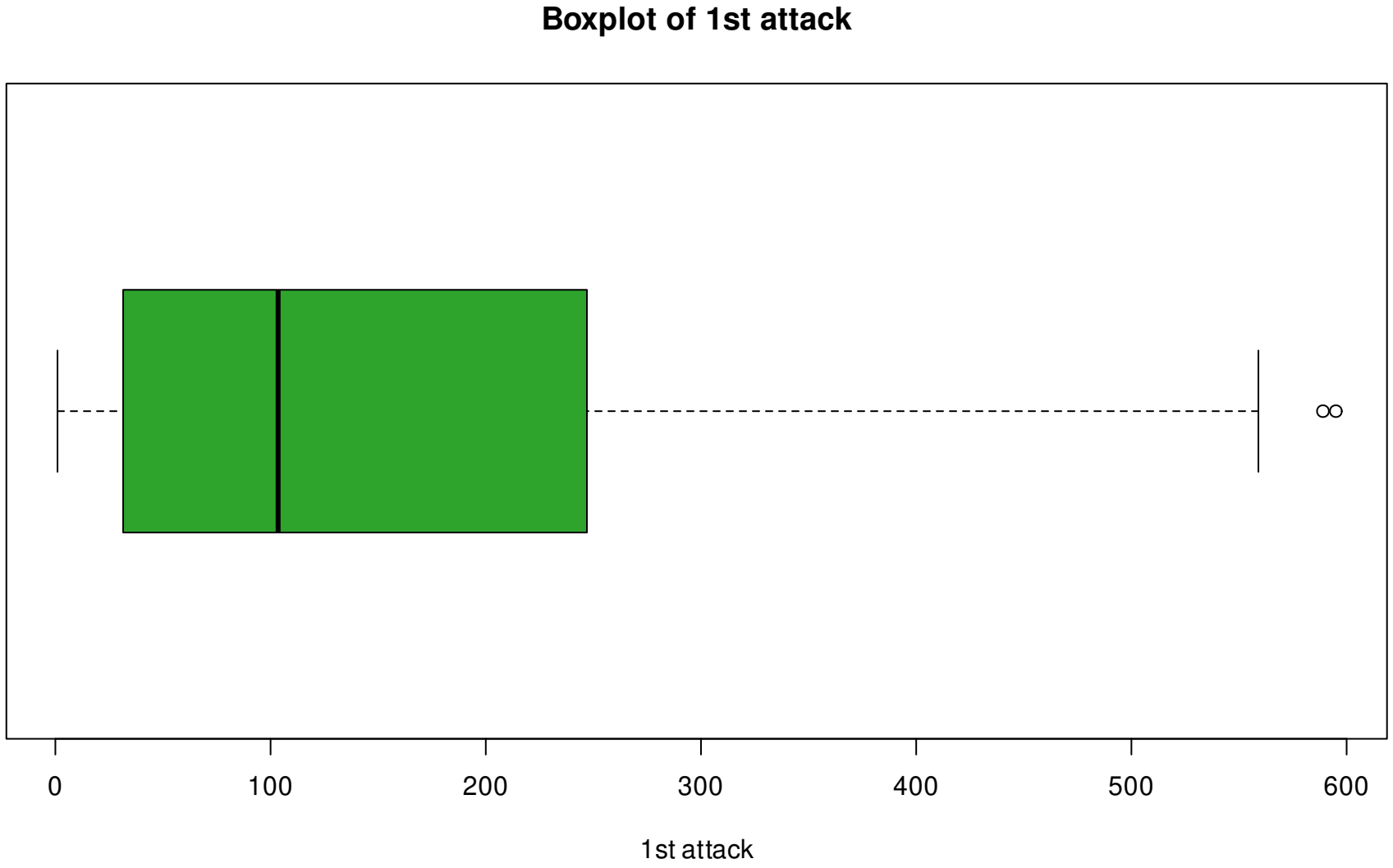
Histograms Summary

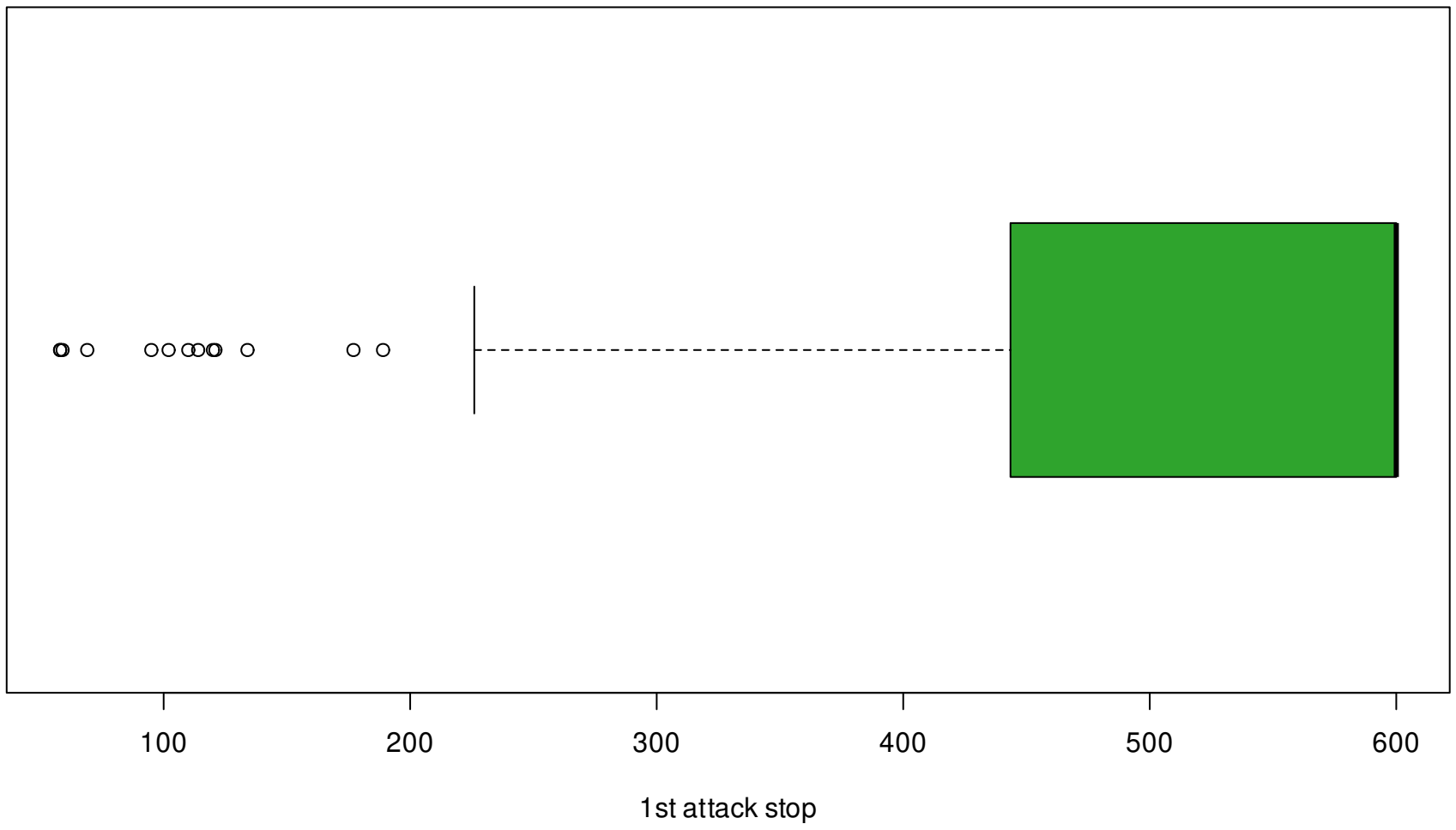
Multiple Relative Frequency Histogram in one figure. Variables are sorted alphabetically. The blue line represents the normal density approximation. The blue dotted line represents a special kernel density approximation.

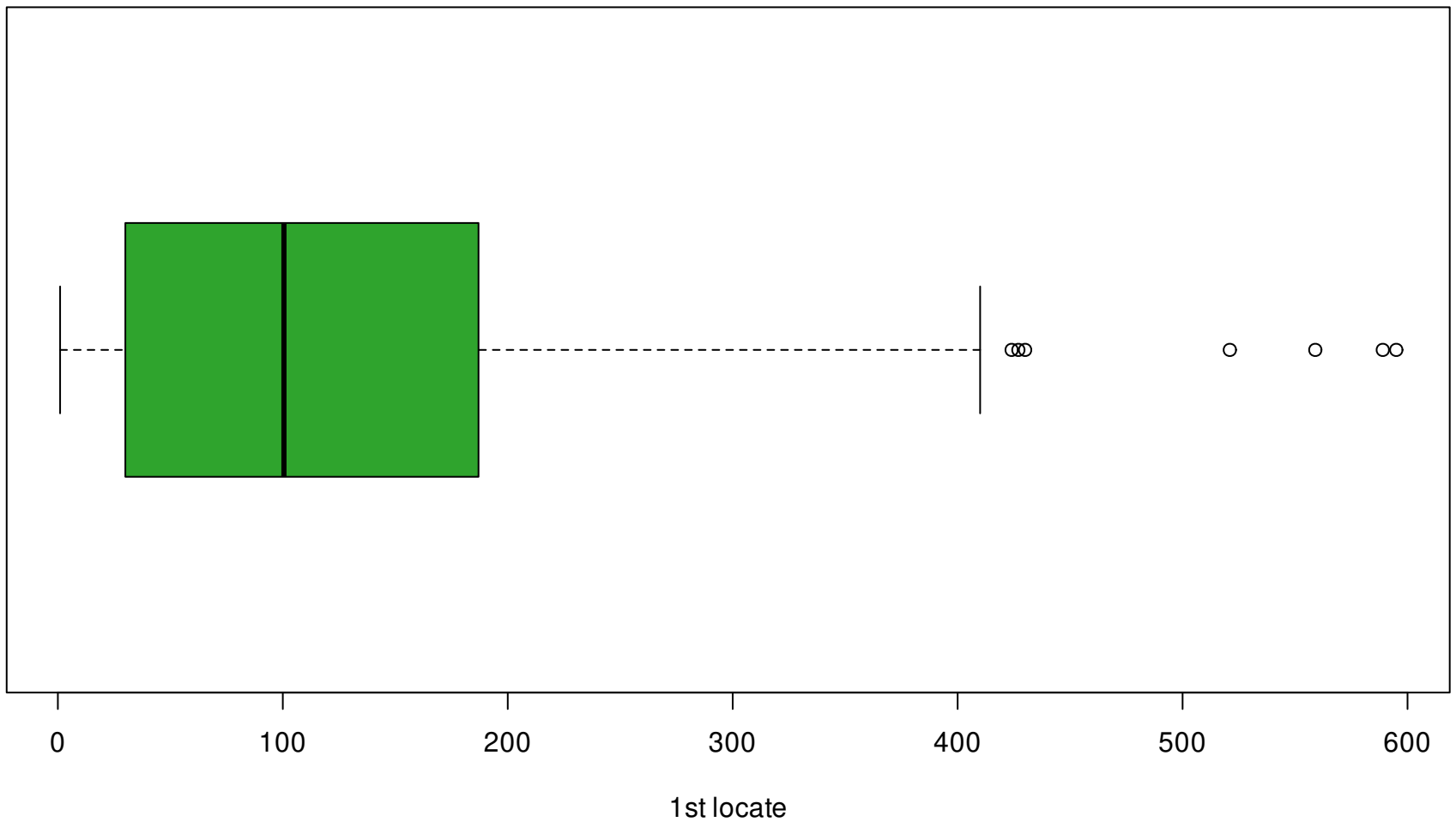


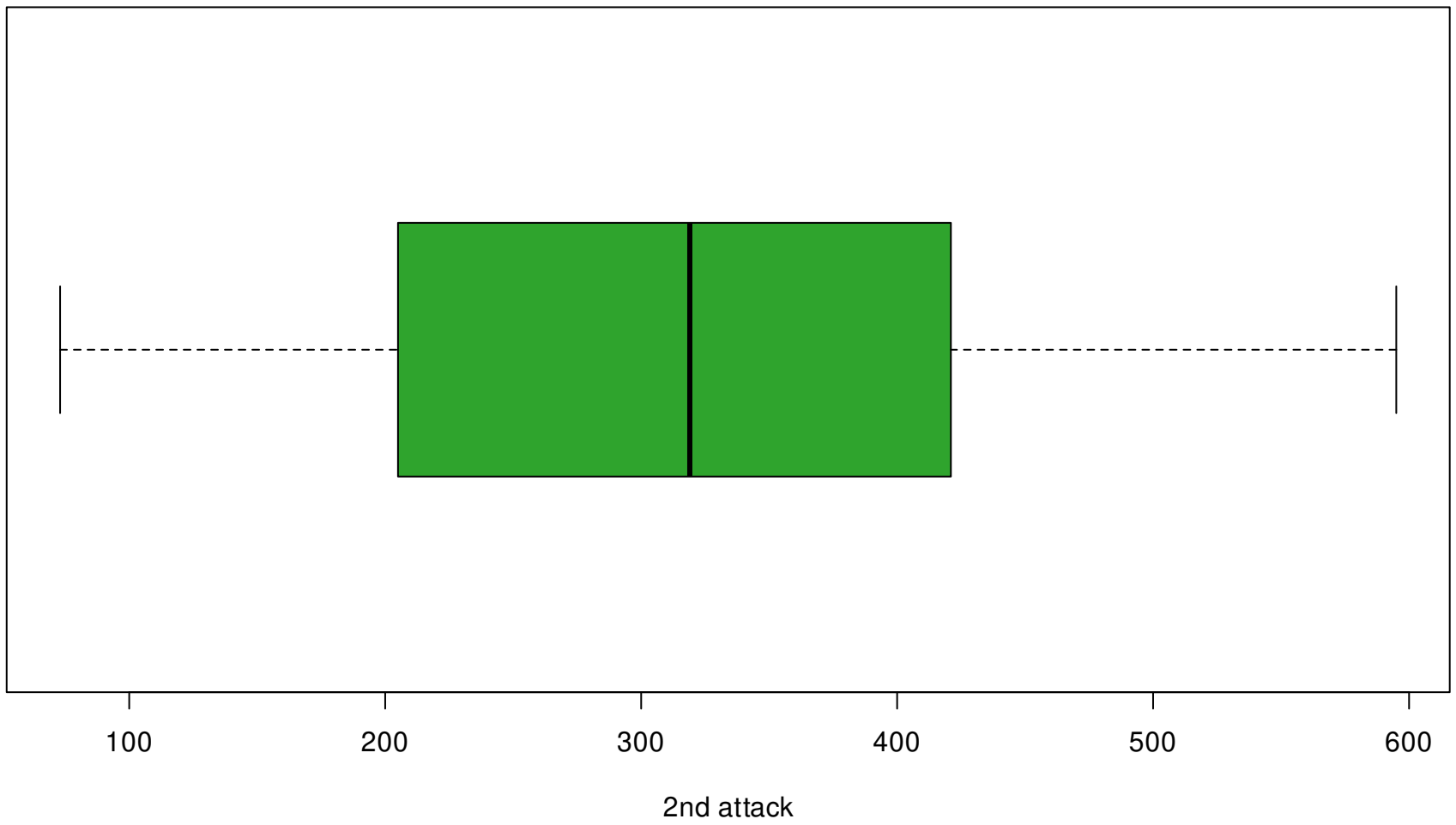
Box-Plots

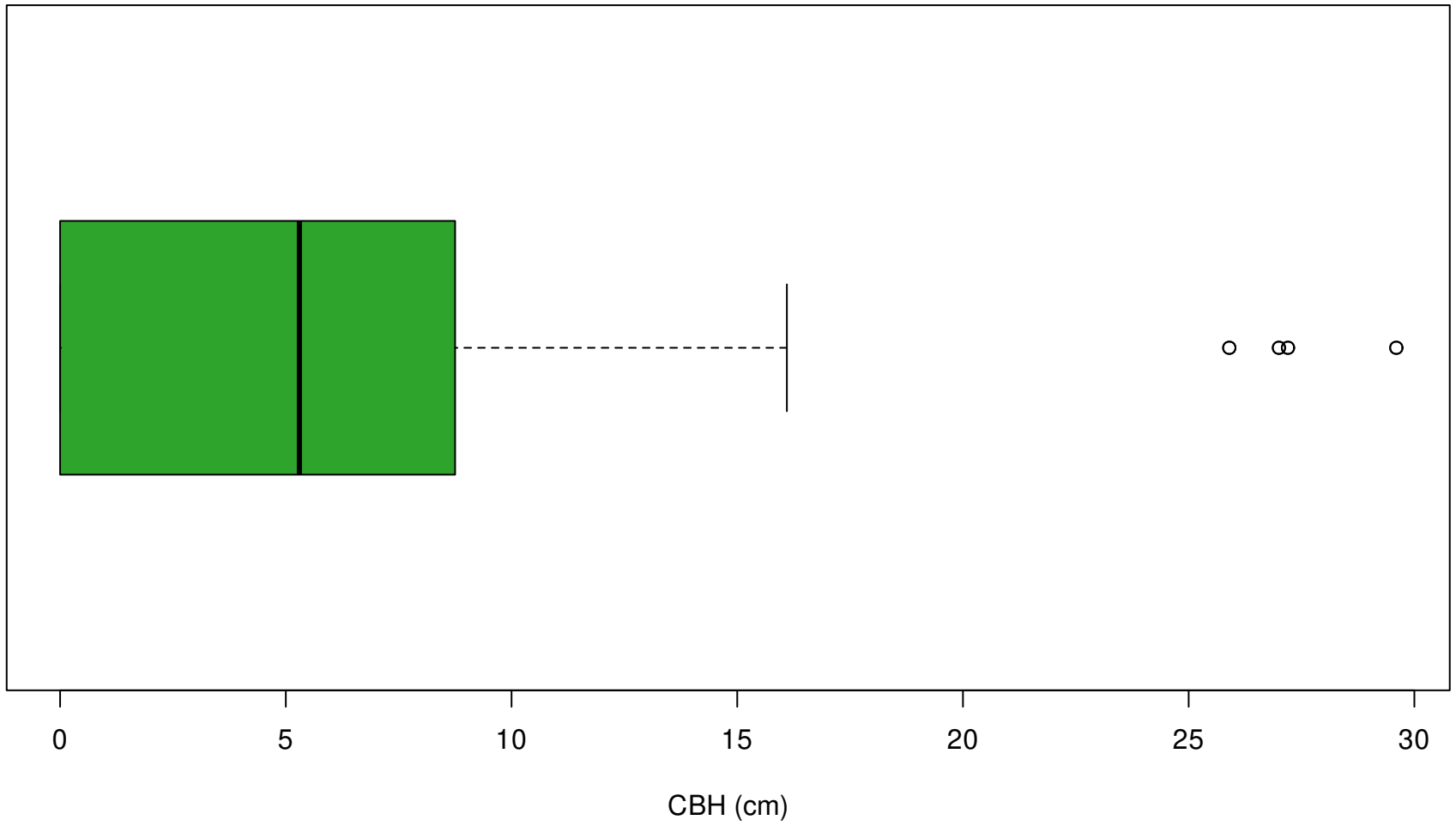
One Box-Plot per page for each variable. Variables are sorted alphabetically.

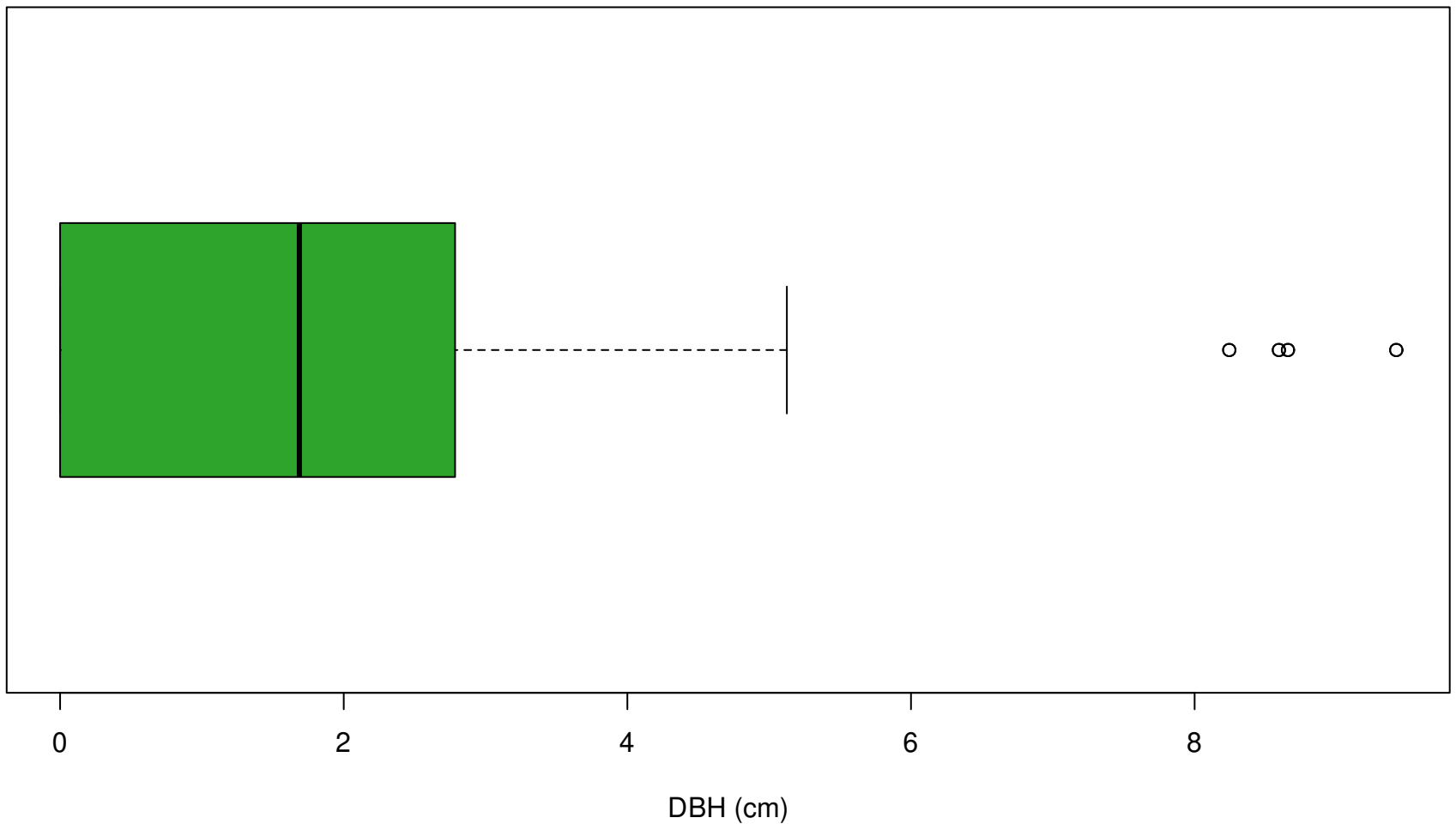


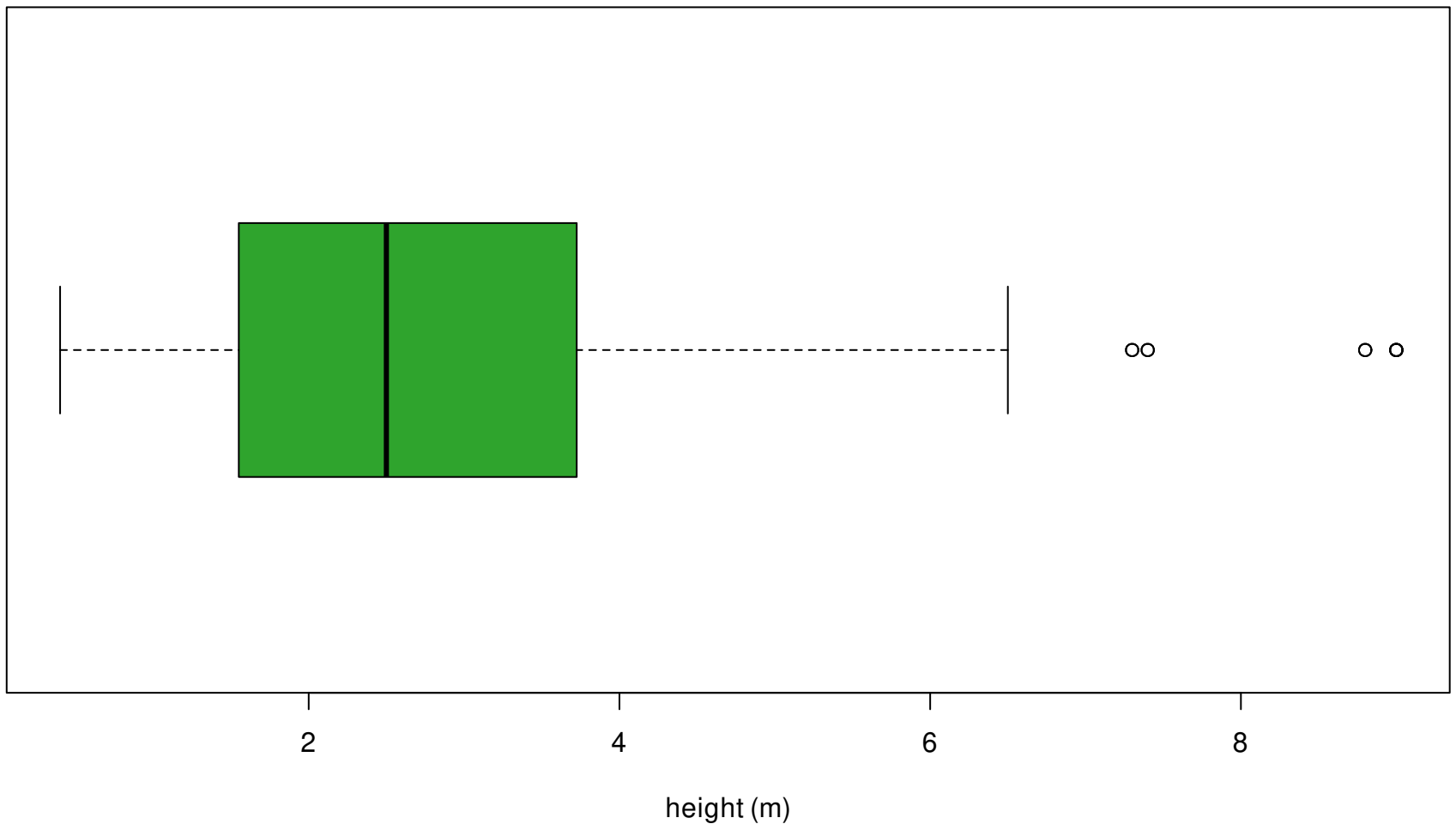

Boxplot of 1st attack stop
100
200
300
400
500
600
1st attack stop


Boxplot of 1st locate
0
100
200
300
400
500
600
1st locate


Boxplot of 2nd attack
100
200
300
400
500
600
2nd attack

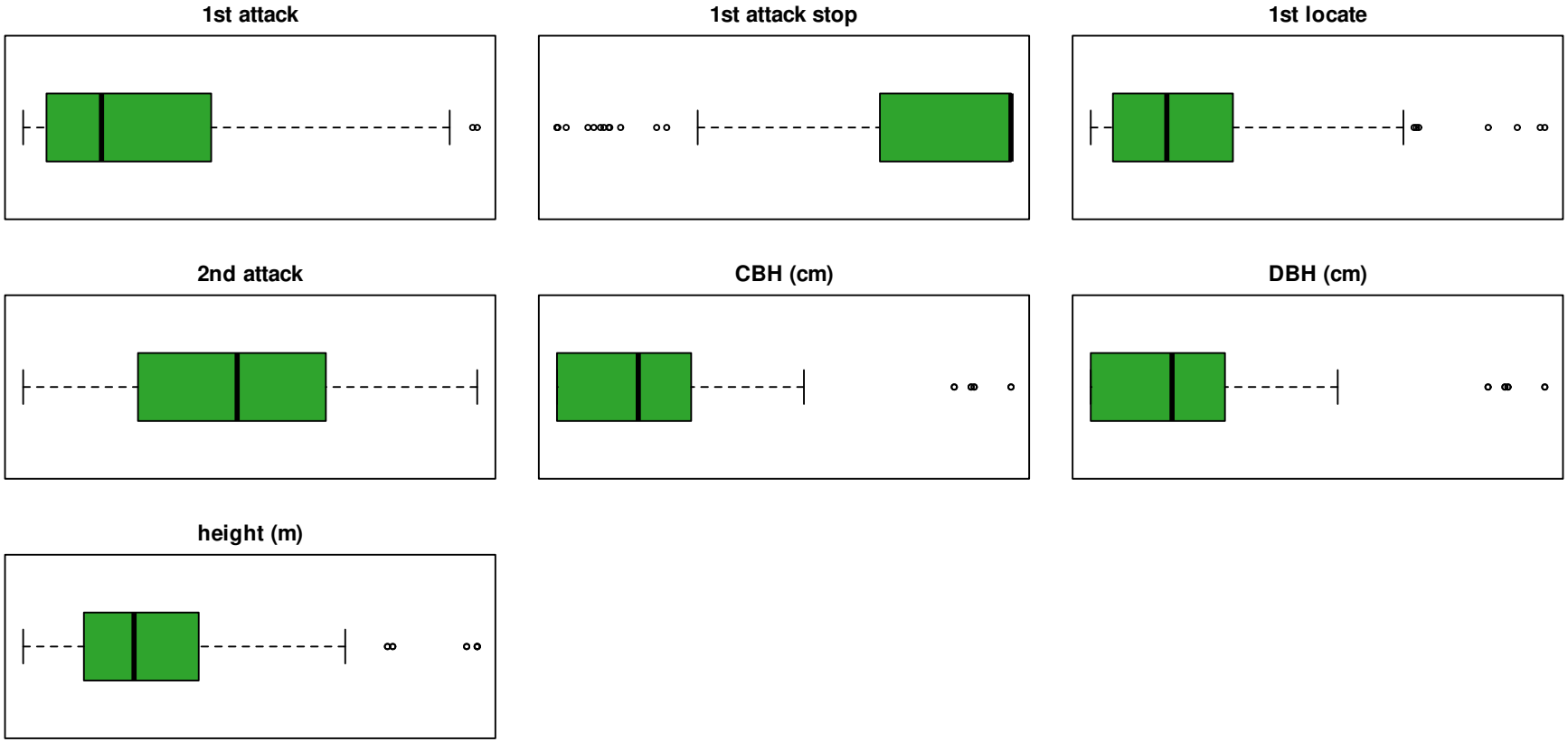

Boxplot of CBH (cm)
0
5
10
15
20
25
30
CBH (cm)


Boxplot of DBH (cm)
0
2
4
6
8
DBH (cm)


Boxplot of height (m)
2
4
6
8
height (m)

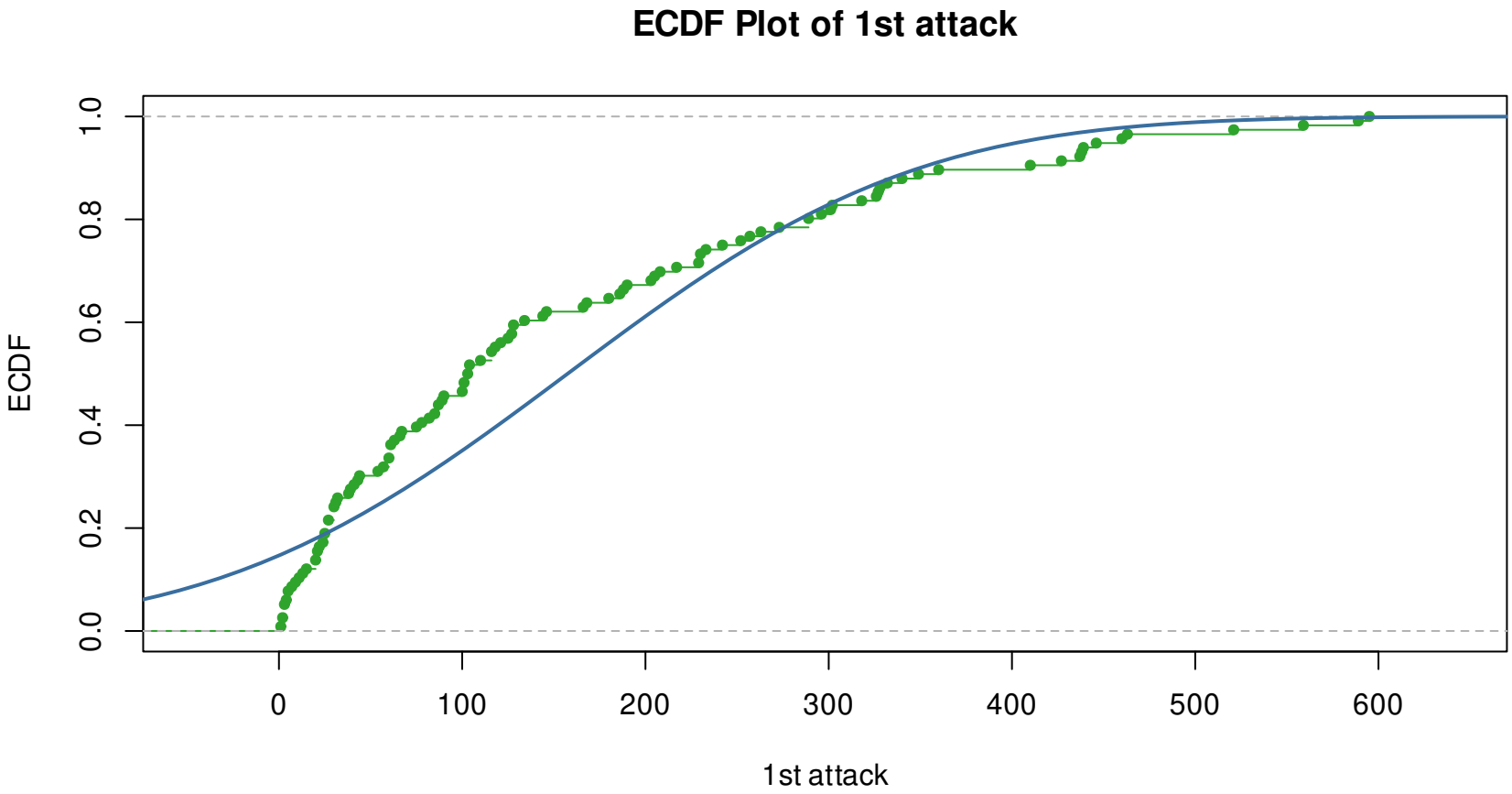
Box-Plots Summary

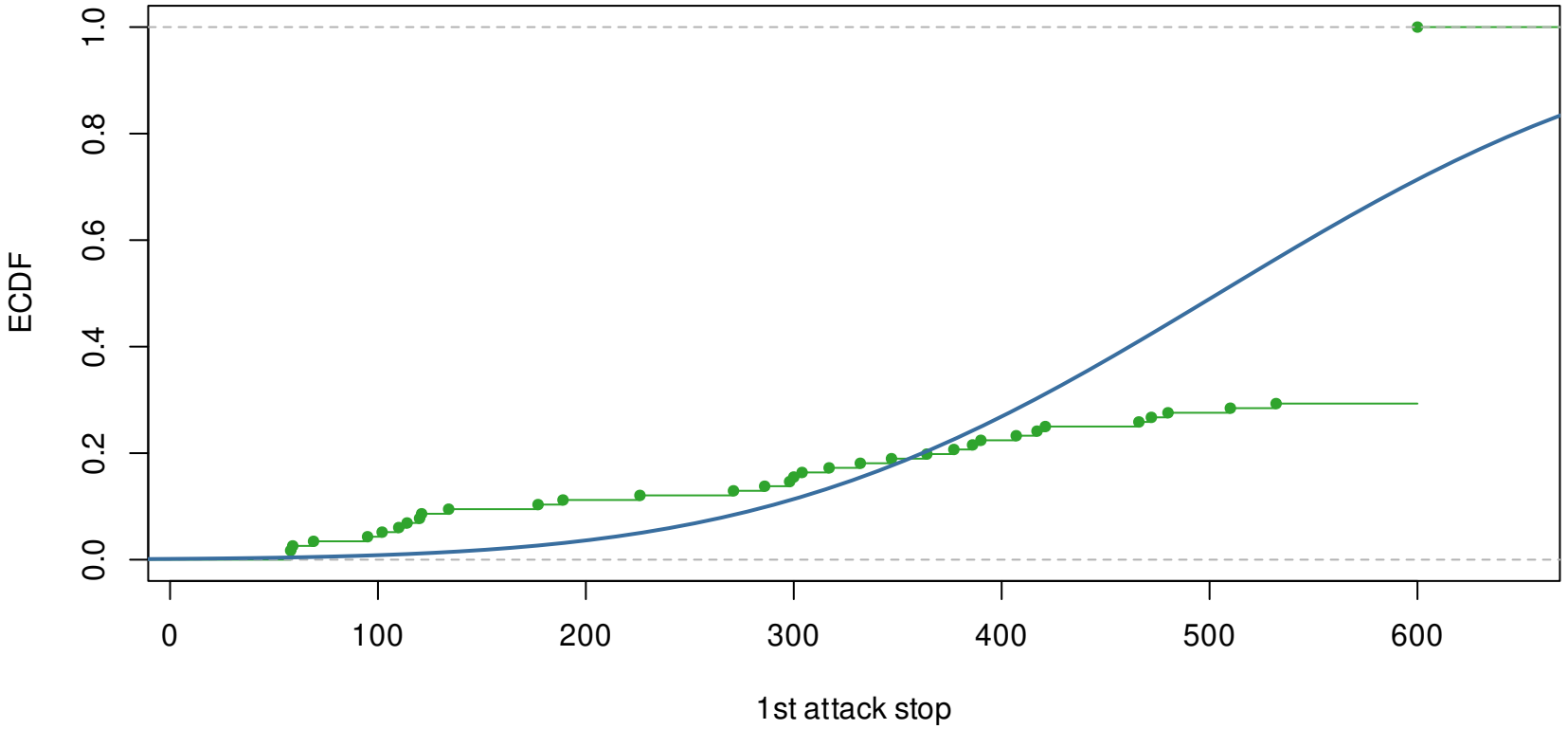
Multiple Box-Plots of variables in one figure. Variables are sorted alphabetically.

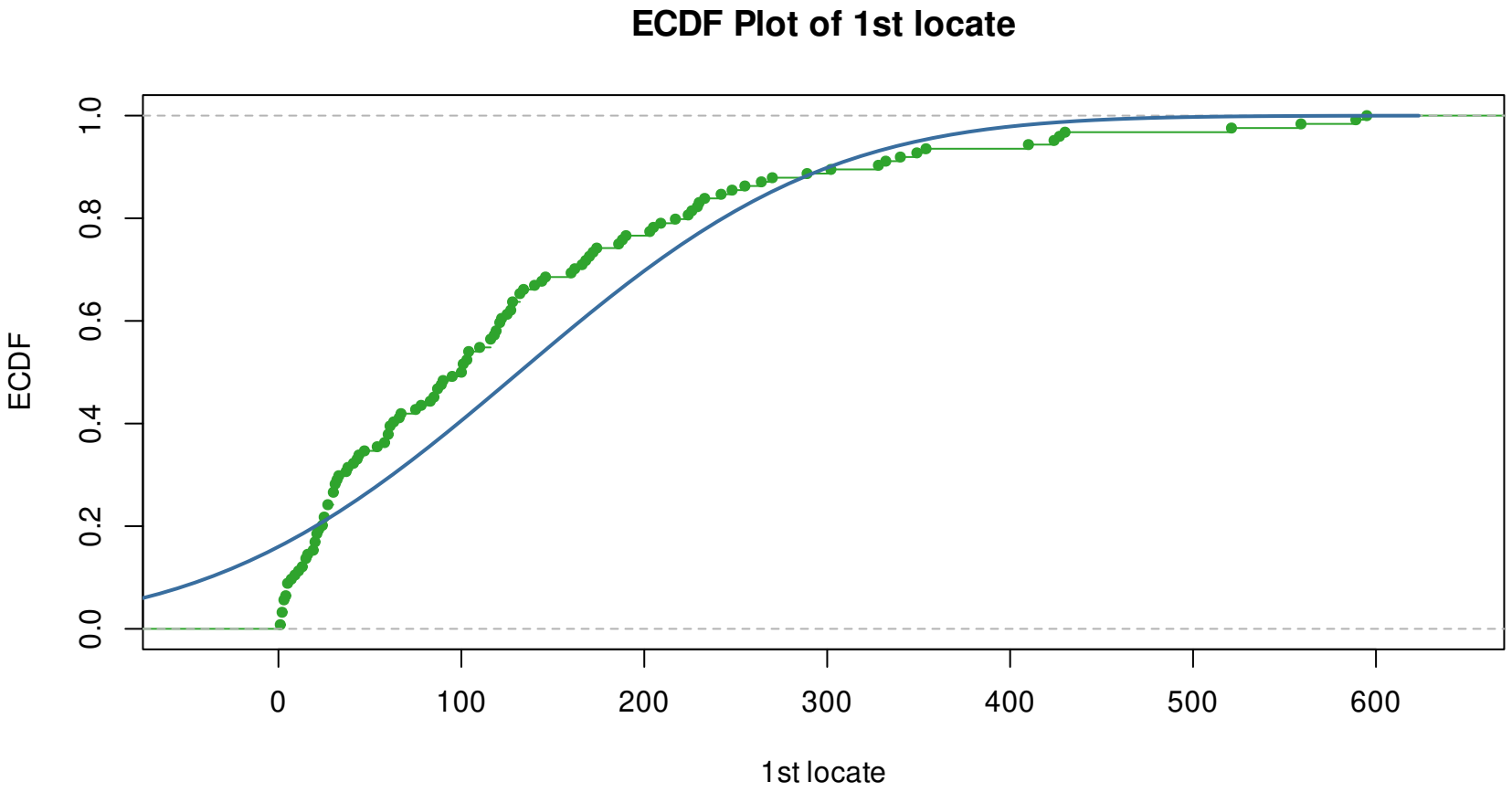


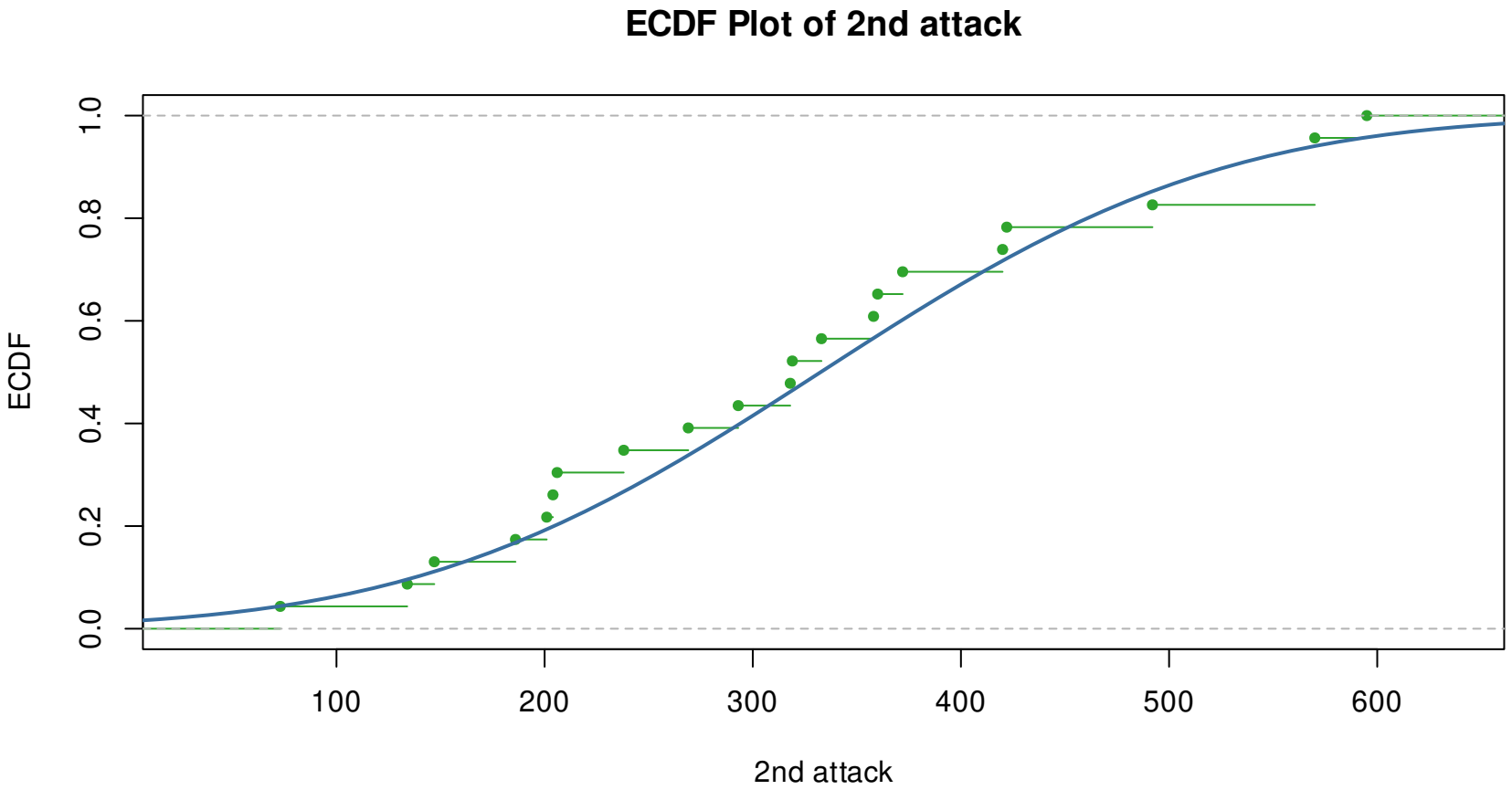
ECDF Plots

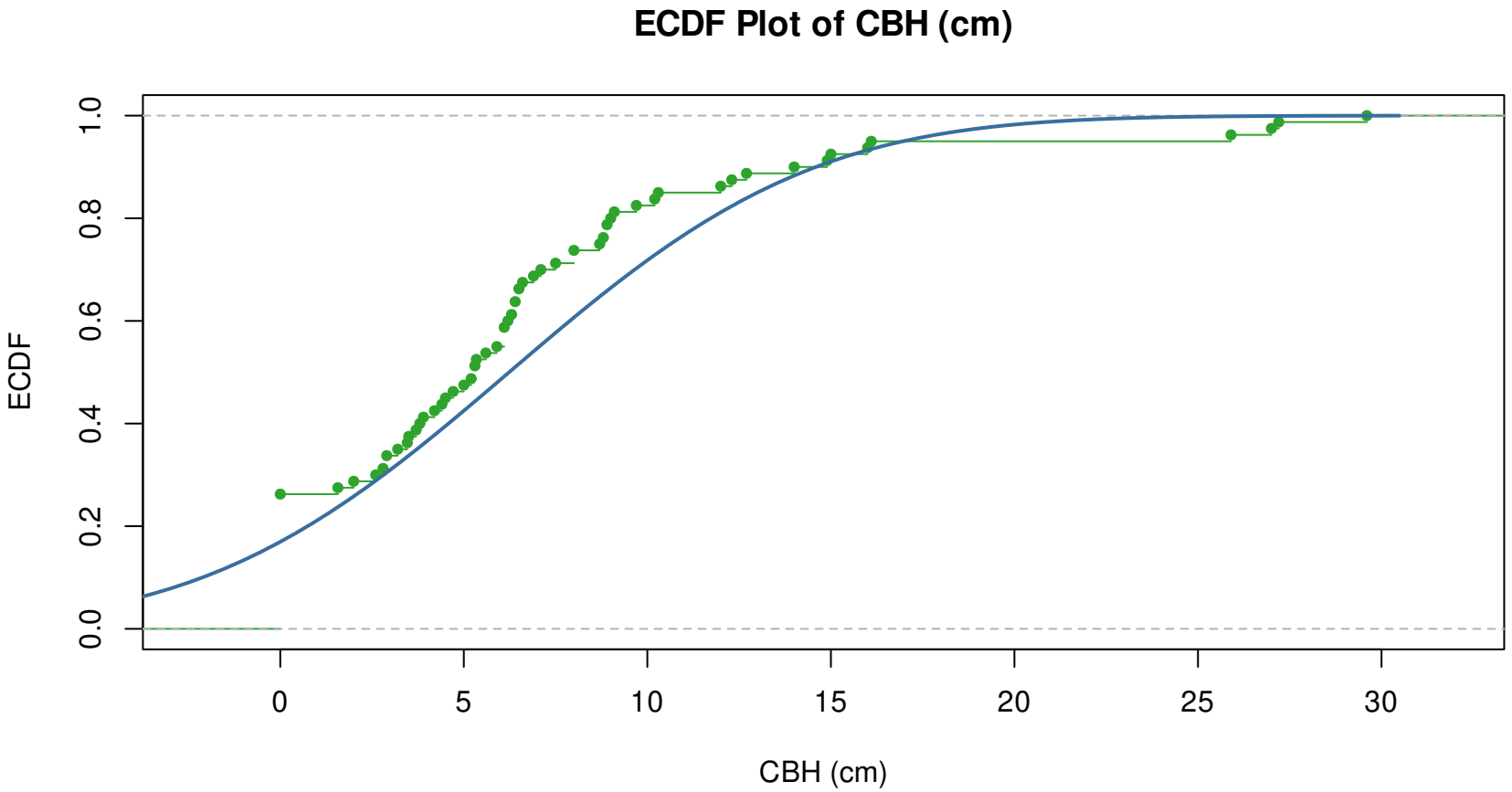
One ECDF (Empirical Cumulative Distribution Function) Plot per page for each variable. Variables are sorted alphabetically. The blue line represents the CDF of a normal distribution. If the variable is normally distributed, the blue line approximates well the ECDF.

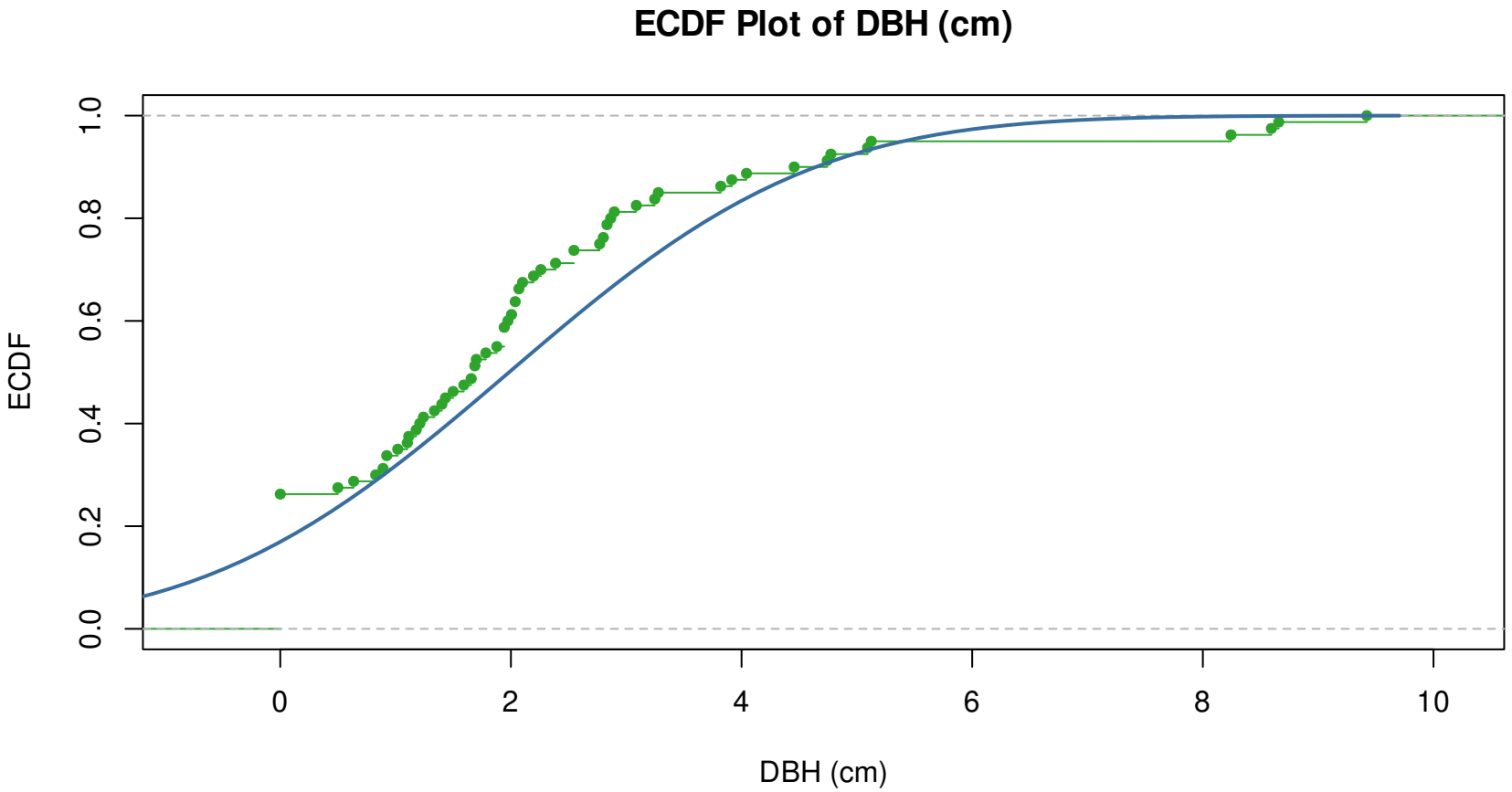


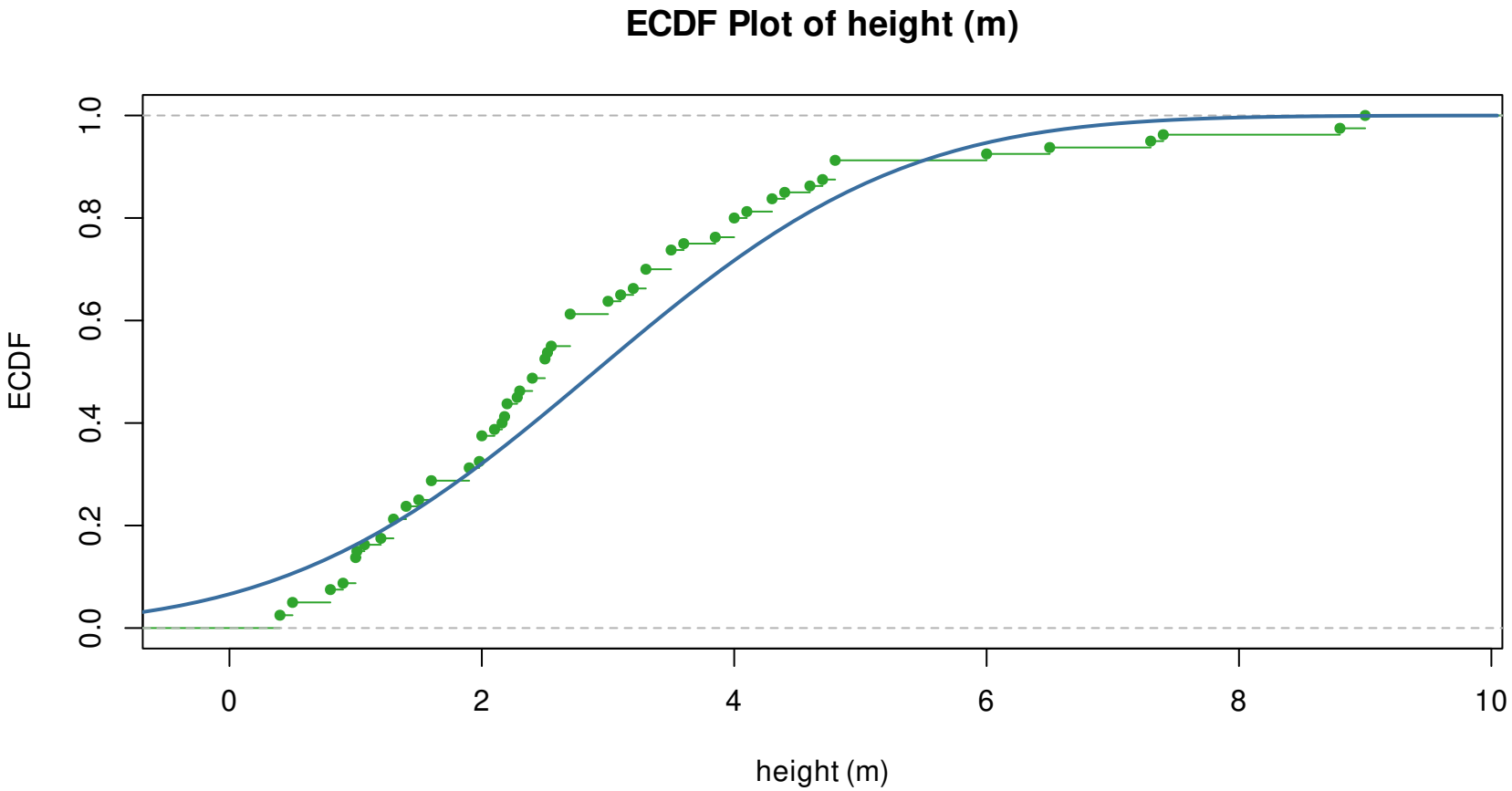

ECDF Plot of 1st attack stop
ECDF
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
600
1st attack stop


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


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

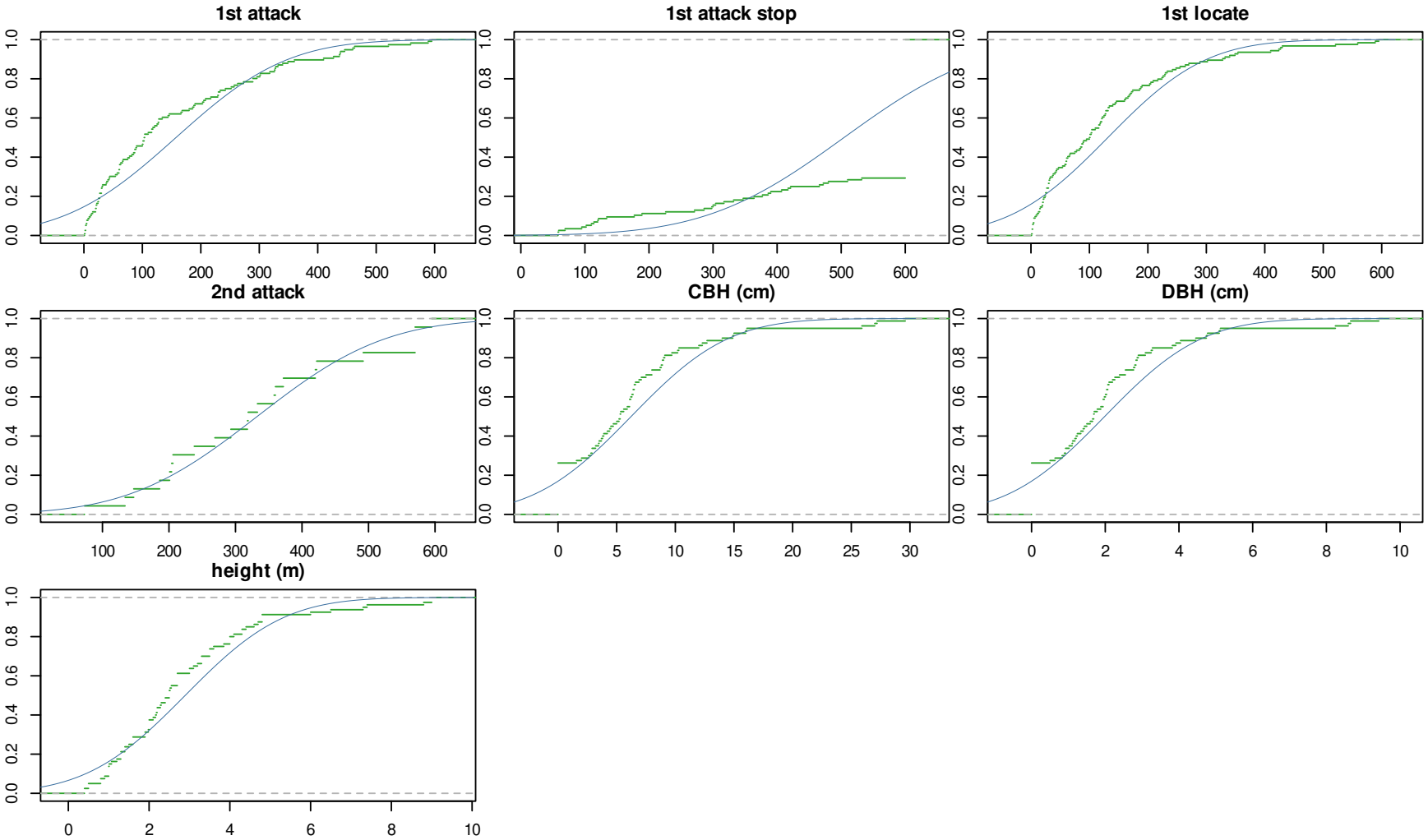

ECDF Plot of CBH (cm)
ECDF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
CBH (cm)


ECDF Plot of DBH (cm)
ECDF
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
DBH (cm)


ECDF Plot of height (m)
ECDF
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
height (m)

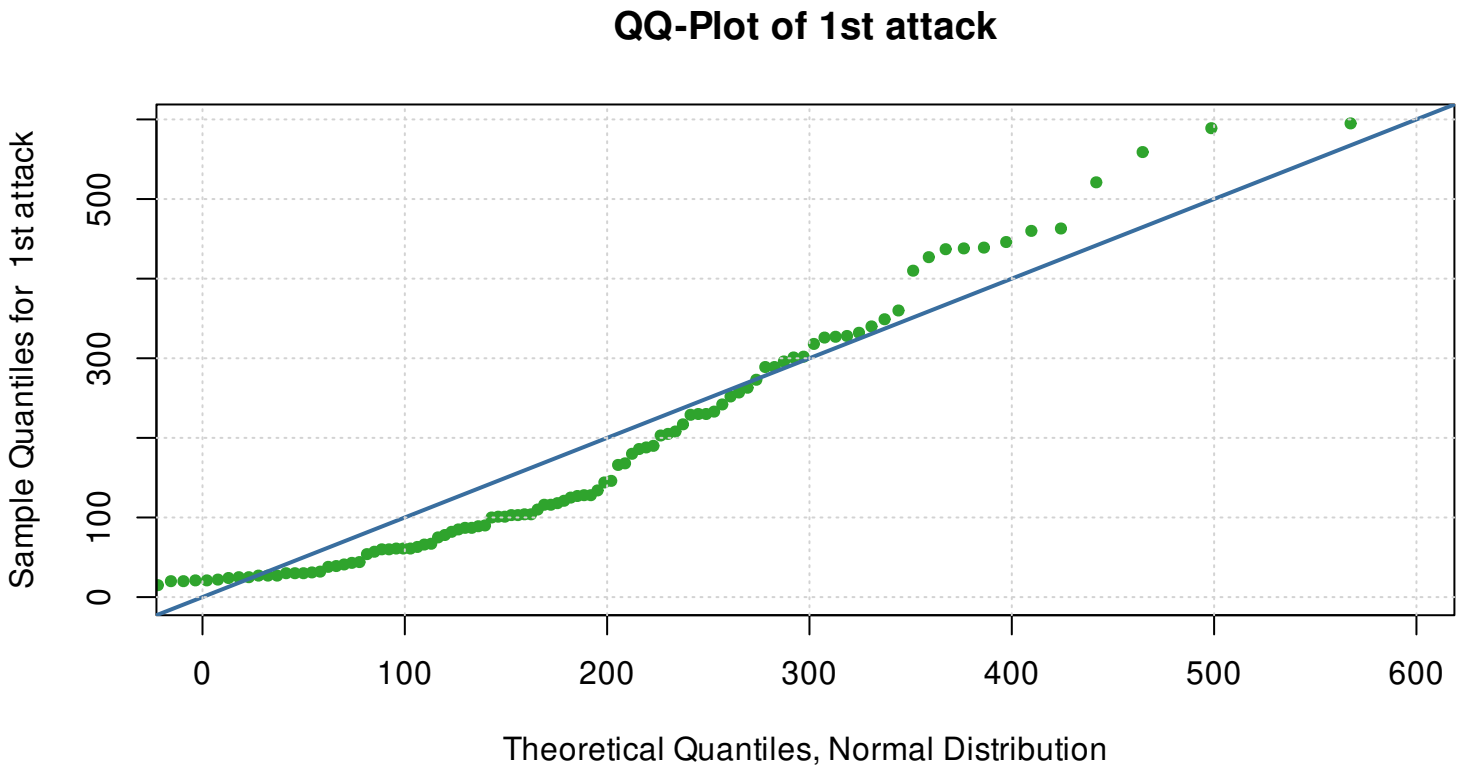
ECDF Plots Summary

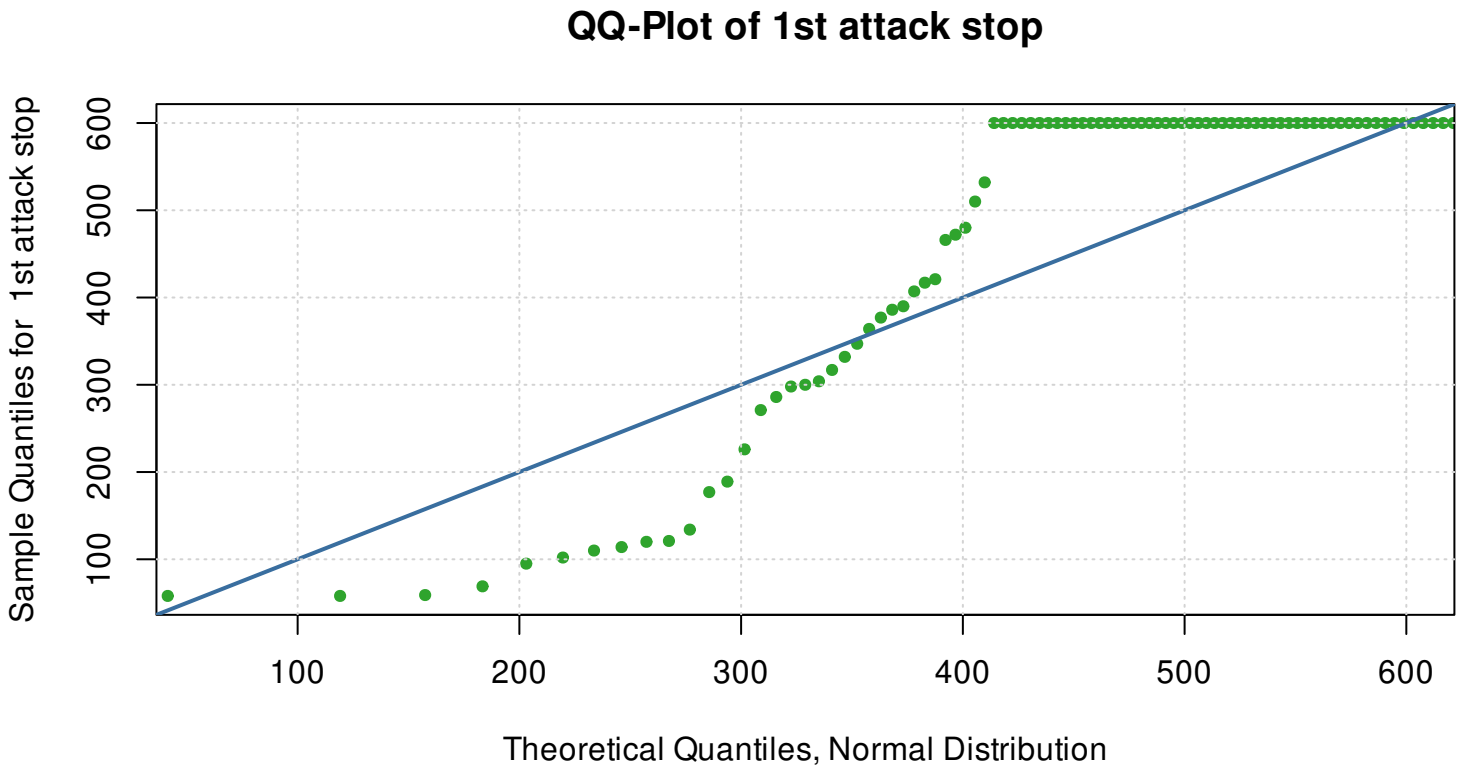
Multiple ECDF Plots of variables in one figure. Variables are sorted alphabetically.

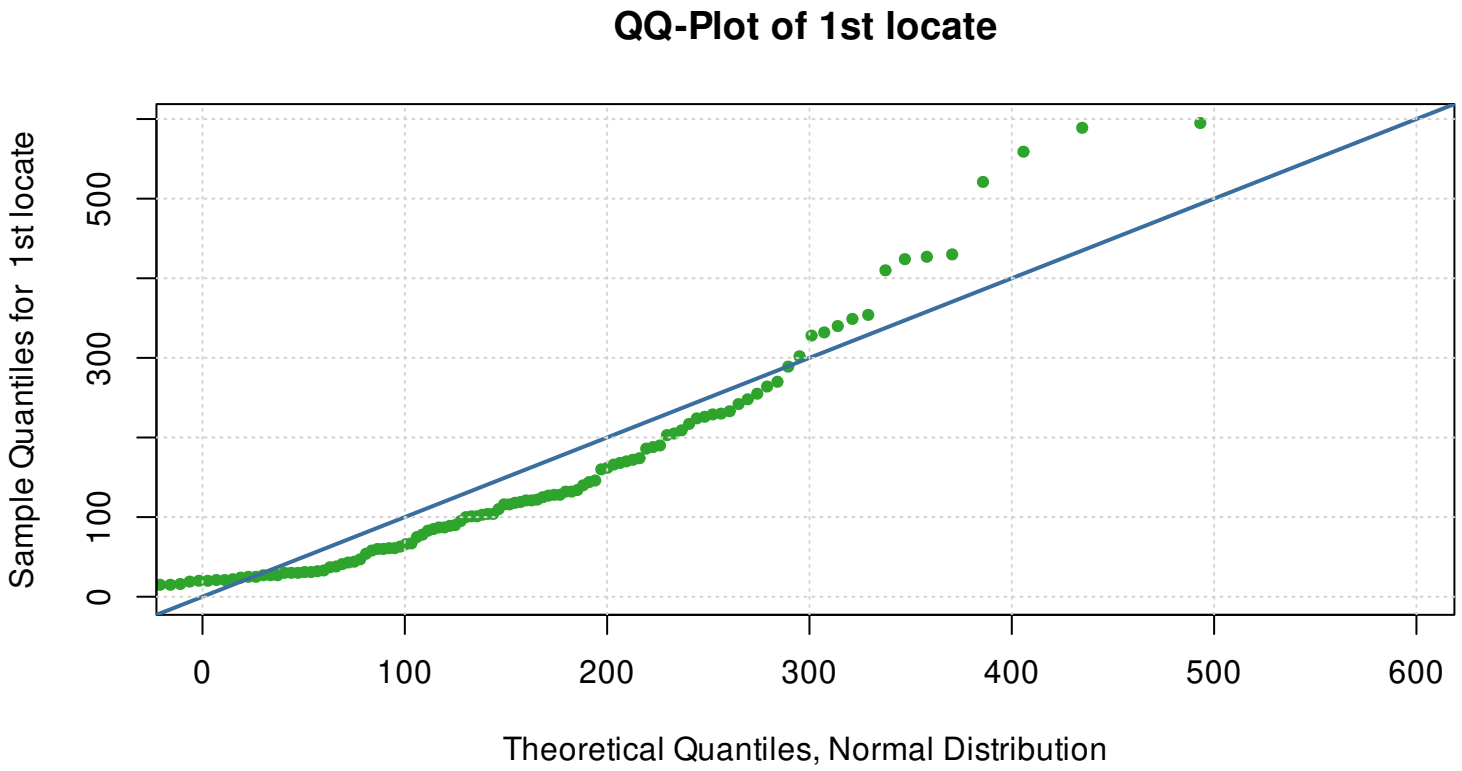


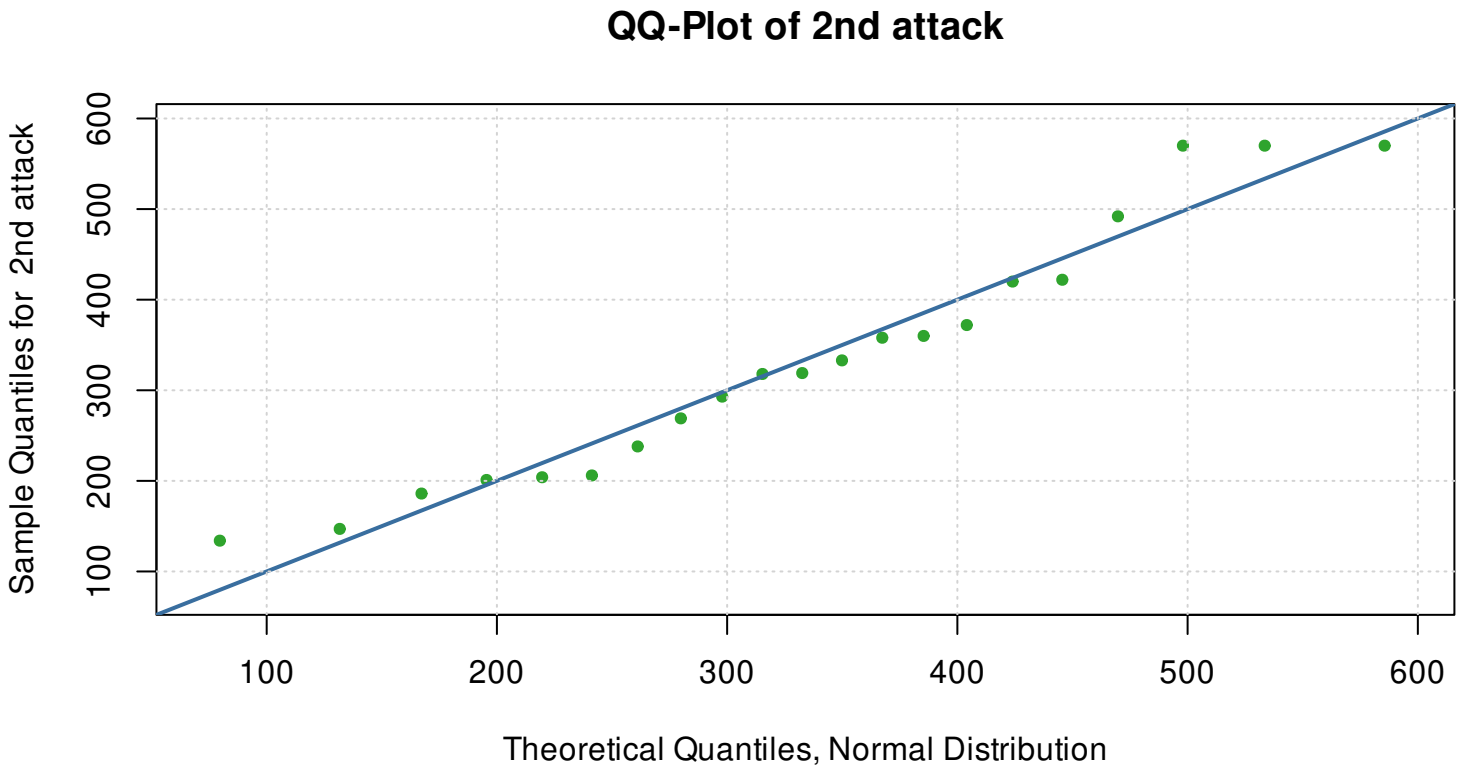
QQ-Plots

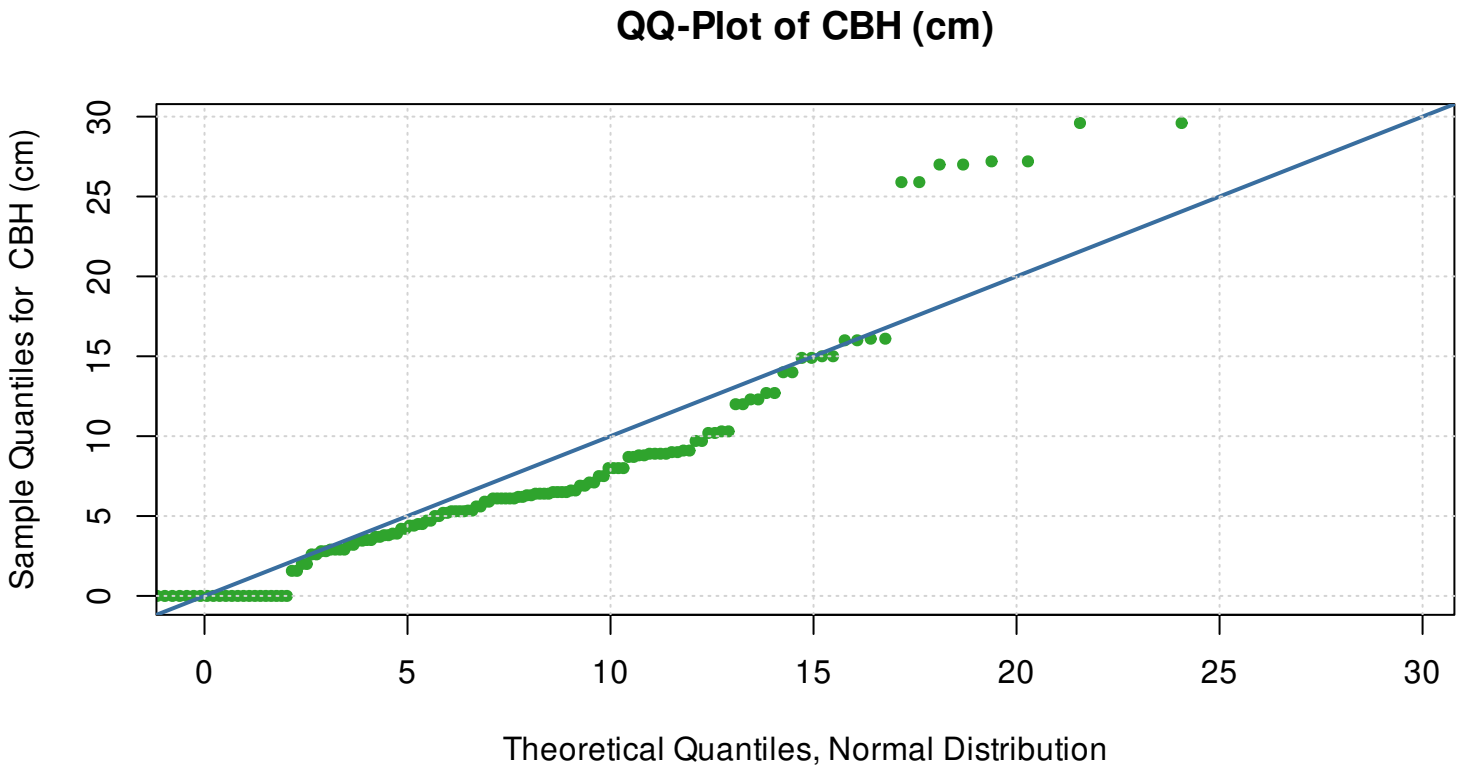
One QQ-Plot per page for each variable. Variables are sorted alphabetically.

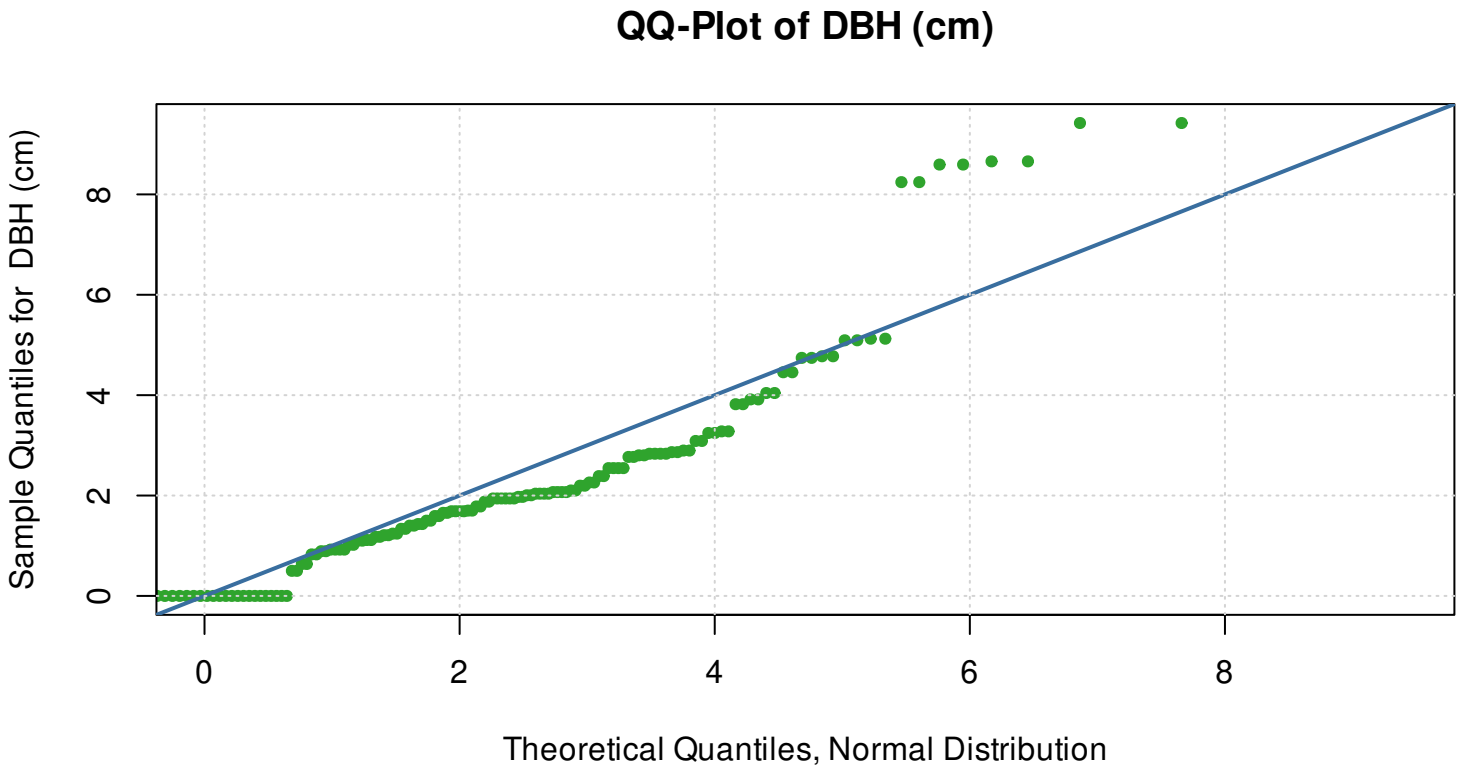


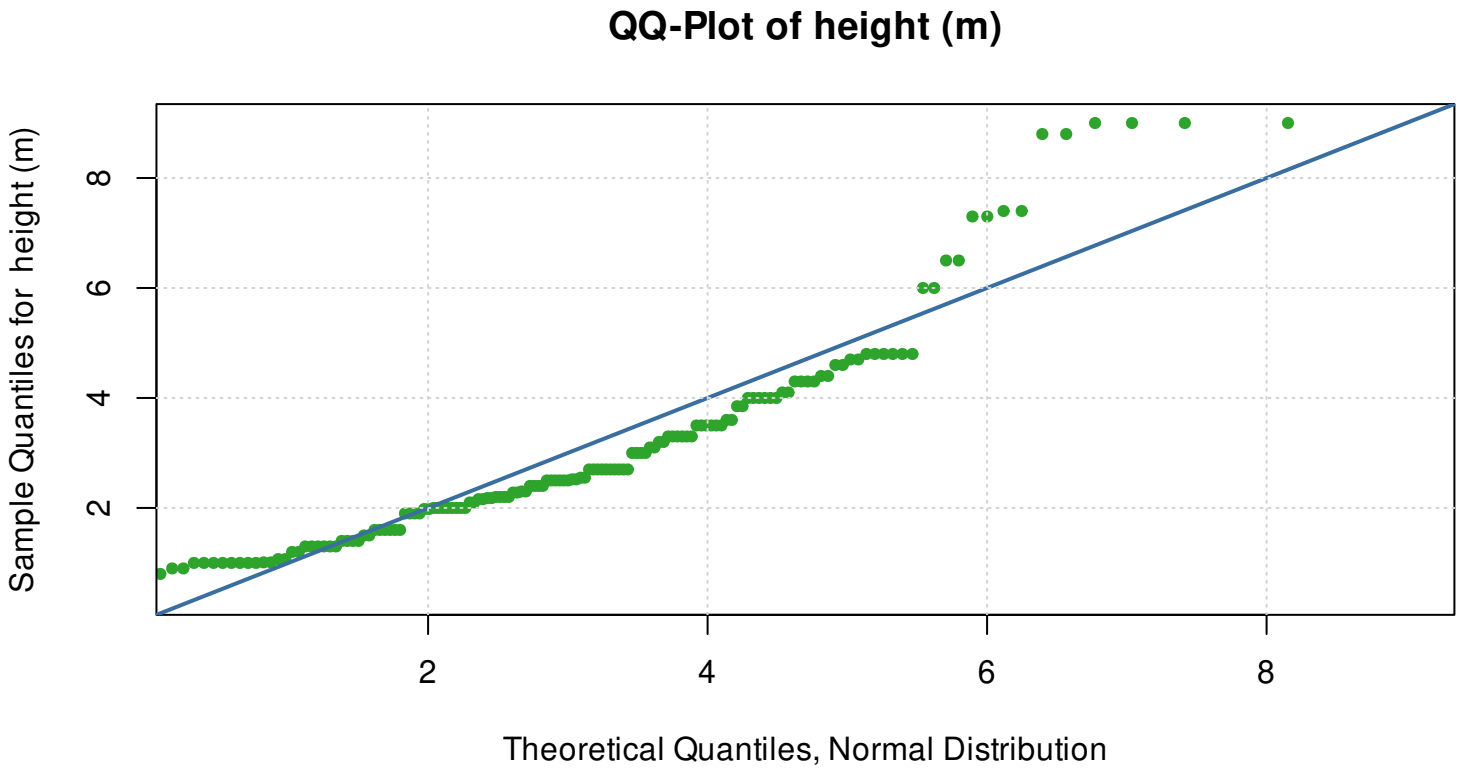

QQ-Plot of 1st attack stop
Sample Quantiles for 1st attack stop
100
200
300
400
500
600
Theoretical Quantiles, Normal Distribution


QQ-Plot of 1st locate
Sample Quantiles for 1st locate
0
100
300
500
0
100
200
300
400
500
600
Theoretical Quantiles, Normal Distribution


QQ-Plot of 2nd attack
Sample Quantiles for 2nd attack
100
200
300
400
500
600
Theoretical Quantiles, Normal Distribution

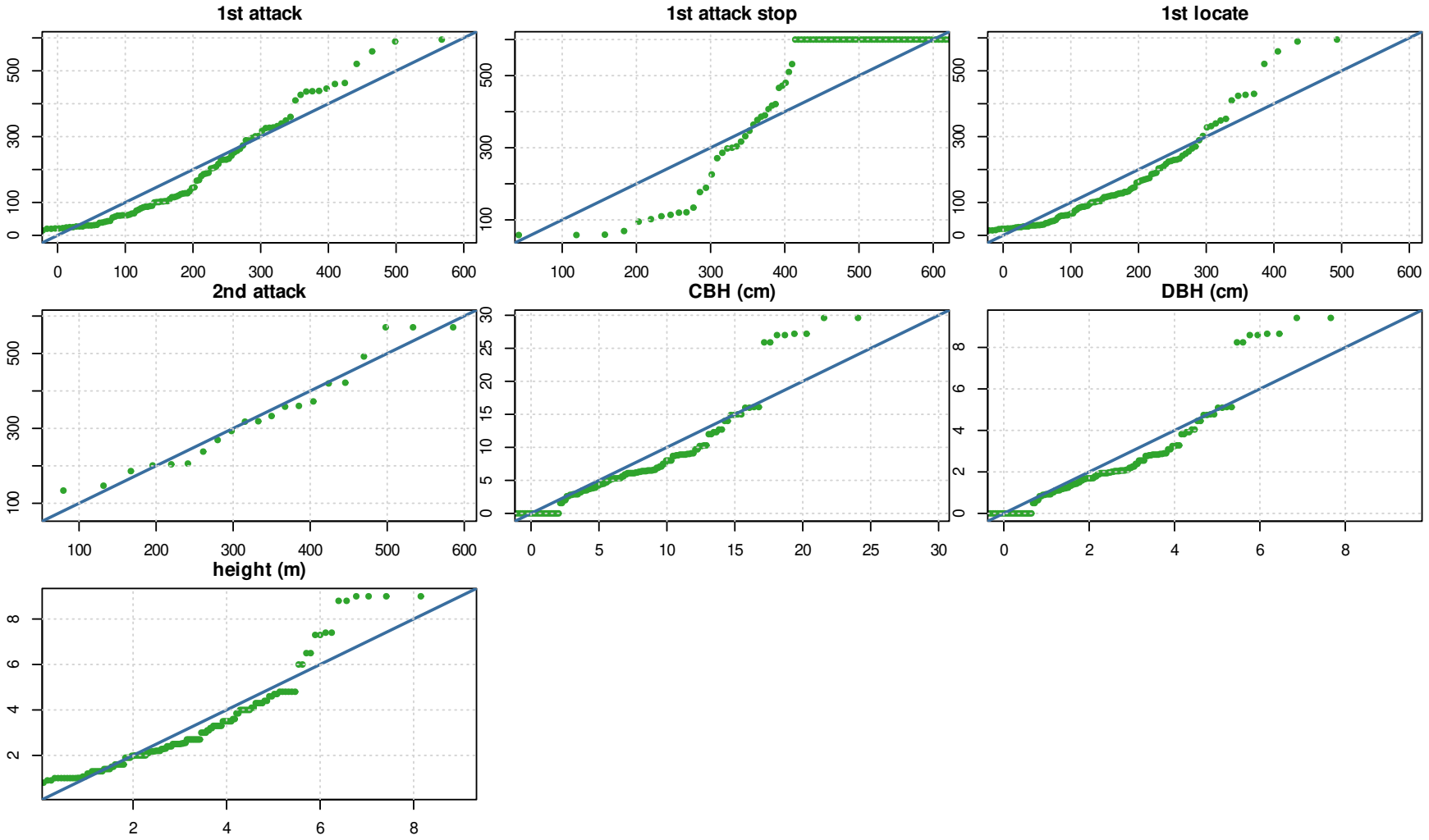

QQ-Plot of CBH (cm)
Sample Quantiles for CBH (cm)
0
5
10
15
20
25
30
Theoretical Quantiles, Normal Distribution


QQ-Plot of DBH (cm)
Sample Quantiles for DBH (cm)
0
2
4
6
8
0
2
4
6
8
Theoretical Quantiles, Normal Distribution


QQ-Plot of height (m)
Sample Quantiles for height (m)
2
4
6
8
2
4
6
8
Theoretical Quantiles, Normal Distribution

QQ-Plots Summary

QQ-Plots of variables in one figure. Theoretical Quantiles of the Normal Distribution.



Results for Discrete Variables

Descriptive Statistics

Totals

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

Variable	N Obs	N Missing	N Valid	% Complete	N Unique
2nd attack stop	160	137	23	14.38	11
ant sample	160	100	60	37.50	2
Attacked	160	0	160	100.00	2
Baiting tree no.	160	0	160	100.00	80
date	160	0	160	100.00	29
Detected	160	0	160	100.00	2
elevation (m)	160	0	160	100.00	8
field notes	160	152	8	5.00	6
H: 0	160	0	160	100.00	10
H: 1-5%	160	0	160	100.00	32
H: 33+%	160	0	160	100.00	39
H: 5-33%	160	0	160	100.00	41
lab notes	160	145	15	9.38	9
Recruited	160	0	160	100.00	2
species	160	0	160	100.00	5
Termite/C	160	0	160	100.00	2
transect	160	0	160	100.00	3
Tree number	160	60	100	62.50	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	85.62
2nd attack stop	600	14	8.75
2nd attack stop	242	1	0.62
2nd attack stop	263	1	0.62
2nd attack stop	307	1	0.62
2nd attack stop	370	1	0.62
2nd attack stop	401	1	0.62
2nd attack stop	474	1	0.62
2nd attack stop	492	1	0.62
2nd attack stop	501	1	0.62
2nd attack stop	512	1	0.62
ant sample	Missing	100	62.50
ant sample	Y	60	37.50
Attacked	1	116	72.50
Attacked	0	44	27.50
Baiting tree no.	0799A	2	1.25
Baiting tree no.	1001	2	1.25
Baiting tree no.	1002	2	1.25
Baiting tree no.	1018	2	1.25
Baiting tree no.	108	2	1.25
Baiting tree no.	111	2	1.25
Baiting tree no.	117	2	1.25
Baiting tree no.	118	2	1.25
Baiting tree no.	120	2	1.25
Baiting tree no.	123	2	1.25
Baiting tree no.	13	2	1.25
Baiting tree no.	14	2	1.25
Baiting tree no.	157	2	1.25
Baiting tree no.	250	2	1.25
Baiting tree no.	251	2	1.25
Baiting tree no.	32	2	1.25
Baiting tree no.	329	2	1.25
Baiting tree no.	36	2	1.25
Baiting tree no.	627	2	1.25
Baiting tree no.	628	2	1.25
Baiting tree no.	****All Other Values****	120	75.00
date	14.08.13	20	12.50
date	17.08.13	20	12.50
date	19.08.13	20	12.50
date	20.08.13	20	12.50
date	21.08.13	20	12.50
date	15.08.13	18	11.25
date	16.08.13	18	11.25
date	24.08.13	2	1.25

(continued)

Variable	Category	Frequency	Percent
date	26.08.13	2	1.25
date	22.08.13	1	0.62
date	22.08.14	1	0.62
date	22.08.15	1	0.62
date	22.08.16	1	0.62
date	22.08.17	1	0.62
date	22.08.18	1	0.62
date	22.08.19	1	0.62
date	22.08.20	1	0.62
date	22.08.21	1	0.62
date	22.08.22	1	0.62
date	22.08.23	1	0.62
date	****All Other Values****	9	5.62
Detected	1	124	77.50
Detected	0	36	22.50
elevation (m)	700	20	12.50
elevation (m)	800	20	12.50
elevation (m)	900	20	12.50
elevation (m)	1000	20	12.50
elevation (m)	1100	20	12.50
elevation (m)	1200	20	12.50
elevation (m)	1300	20	12.50
elevation (m)	1400	20	12.50
field notes	Missing	152	95.00
field notes	(Jimmy data)	3	1.88
field notes	*broken	2	1.25
field notes	*broken, 5 @ resprout	1	0.62
field notes	few ants	1	0.62
field notes	few ants on lower branches but inhabited above	1	0.62
H: 0	0	84	52.50
H: 0	1	22	13.75
H: 0	2	12	7.50
H: 0	3	12	7.50
H: 0	NA	10	6.25
H: 0	5	8	5.00
H: 0	8	6	3.75
H: 0	12	2	1.25
H: 0	4	2	1.25
H: 0	7	2	1.25
H: 1-5%	0	22	13.75
H: 1-5%	1	18	11.25
H: 1-5%	3	14	8.75
H: 1-5%	4	12	7.50
H: 1-5%	10	10	6.25
H: 1-5%	NA	10	6.25
H: 1-5%	2	8	5.00
H: 1-5%	14	4	2.50
H: 1-5%	19	4	2.50
H: 1-5%	20	4	2.50

(continued)

Variable	Category	Frequency	Percent
H: 1-5%	22	4	2.50
H: 1-5%	23	4	2.50
H: 1-5%	26	4	2.50
H: 1-5%	33	4	2.50
H: 1-5%	5	4	2.50
H: 1-5%	11	2	1.25
H: 1-5%	12	2	1.25
H: 1-5%	13	2	1.25
H: 1-5%	21	2	1.25
H: 1-5%	25	2	1.25
H: 1-5%	****All Other Values****	24	15.00
H: 33+%	8	10	6.25
H: 33+%	NA	10	6.25
H: 33+%	13	8	5.00
H: 33+%	17	8	5.00
H: 33+%	10	6	3.75
H: 33+%	11	6	3.75
H: 33+%	12	6	3.75
H: 33+%	16	6	3.75
H: 33+%	20	6	3.75
H: 33+%	24	6	3.75
H: 33+%	4	6	3.75
H: 33+%	5	6	3.75
H: 33+%	7	6	3.75
H: 33+%	9	6	3.75
H: 33+%	1	4	2.50
H: 33+%	14	4	2.50
H: 33+%	19	4	2.50
H: 33+%	22	4	2.50
H: 33+%	26	4	2.50
H: 33+%	35	4	2.50
H: 33+%	****All Other Values****	40	25.00
H: 5-33%	12	10	6.25
H: 5-33%	9	10	6.25
H: 5-33%	NA	10	6.25
H: 5-33%	15	8	5.00
H: 5-33%	21	8	5.00
H: 5-33%	7	8	5.00
H: 5-33%	8	8	5.00
H: 5-33%	13	6	3.75
H: 5-33%	14	6	3.75
H: 5-33%	24	6	3.75
H: 5-33%	0	4	2.50
H: 5-33%	16	4	2.50
H: 5-33%	19	4	2.50
H: 5-33%	23	4	2.50
H: 5-33%	27	4	2.50
H: 5-33%	3	4	2.50
H: 5-33%	39	4	2.50

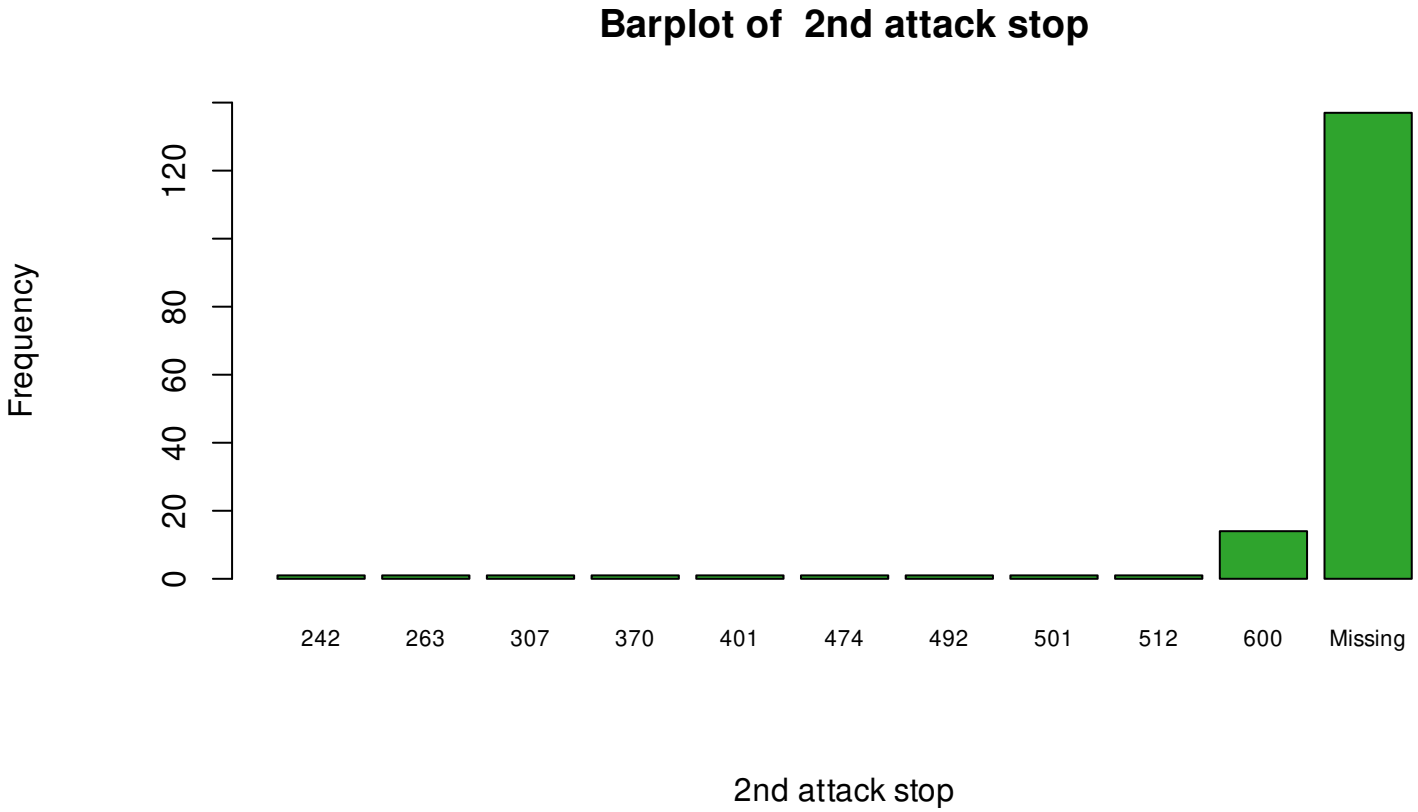
(continued)

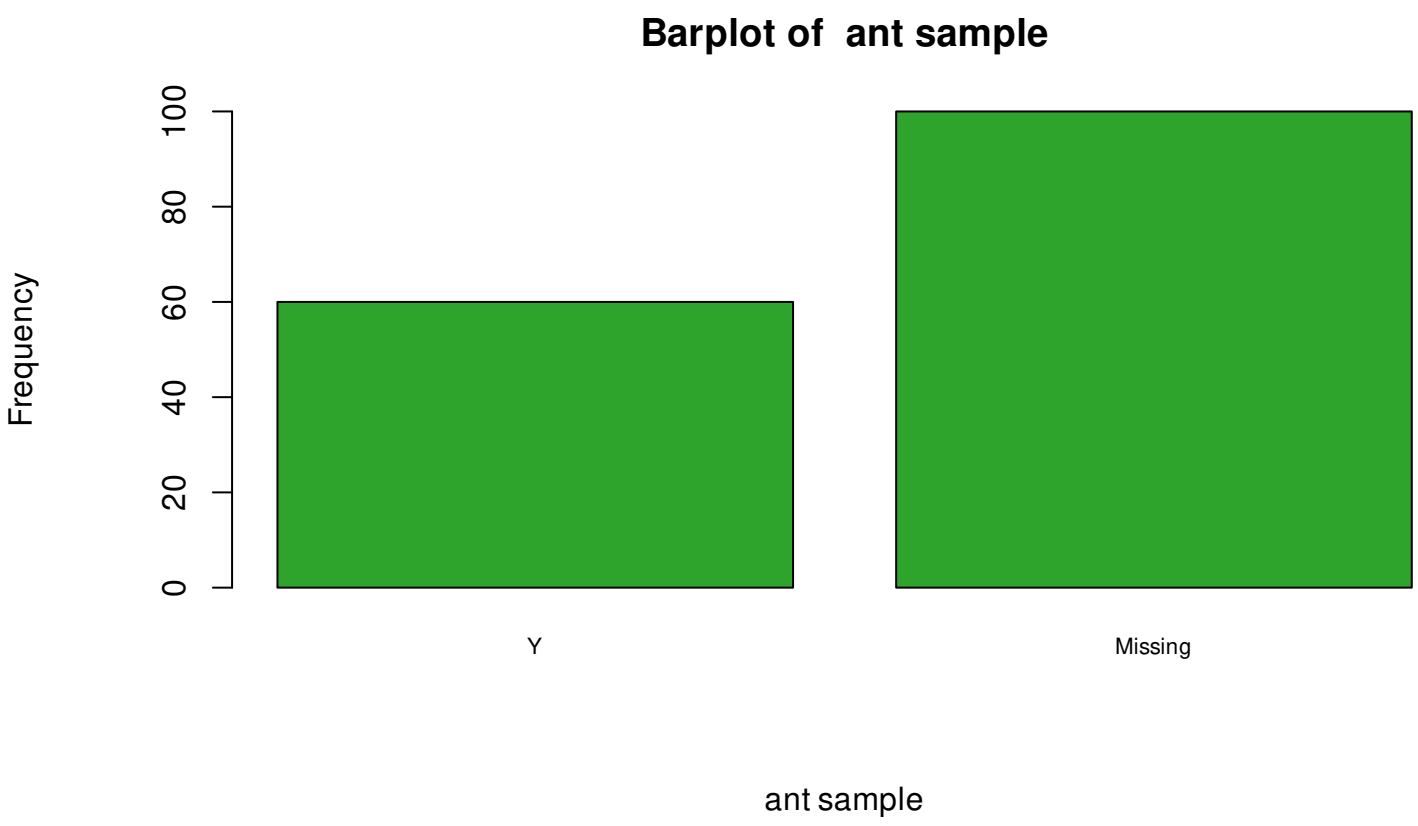
Variable	Category	Frequency	Percent
H: 5-33%	52	4	2.50
H: 5-33%	6	4	2.50
H: 5-33%	1	2	1.25
H: 5-33%	****All Other Values****	42	26.25
lab notes	Missing	145	90.62
lab notes	ant ID from main data	6	3.75
lab notes	250 misread as 230 so changed	2	1.25
lab notes	Vial transferred from Jimmy samples	2	1.25
lab notes	Bigger than 802	1	0.62
lab notes	Much bigger than B0700	1	0.62
lab notes	Much bigger than B0805	1	0.62
lab notes	Small compared to 801	1	0.62
lab notes	Taller petiole than 801	1	0.62
Recruited	1	94	58.75
Recruited	0	66	41.25
species	ANON012	68	42.50
species	ANON013	42	26.25
species	ANON009	40	25.00
species	ANON002	6	3.75
species	ANON001	4	2.50
Termite/C	C	80	50.00
Termite/C	T	80	50.00
transect	Y	100	62.50
transect	N	58	36.25
transect	Y-not m	2	1.25
Tree number	Missing	60	37.50
Tree number	1000 - 0157	2	1.25
Tree number	1000 - 0858	2	1.25
Tree number	1000 - 0863	2	1.25
Tree number	1000 - 0864	2	1.25
Tree number	1100 - 0108	2	1.25
Tree number	1100 - 0111	2	1.25
Tree number	1100 - 0117	2	1.25
Tree number	1100 - 0118	2	1.25
Tree number	1100 - 0120	2	1.25
Tree number	1100 - 0123	2	1.25
Tree number	1100 - 0869	2	1.25
Tree number	1100 - 0889	2	1.25
Tree number	1100 - 0896	2	1.25
Tree number	1200 - 0077	2	1.25
Tree number	1200 - 0079	2	1.25
Tree number	1200 - 0081	2	1.25
Tree number	1200 - 0898	2	1.25
Tree number	1200 - 0902	2	1.25
Tree number	1200 - 0904	2	1.25
Tree number	****All Other Values****	62	38.75

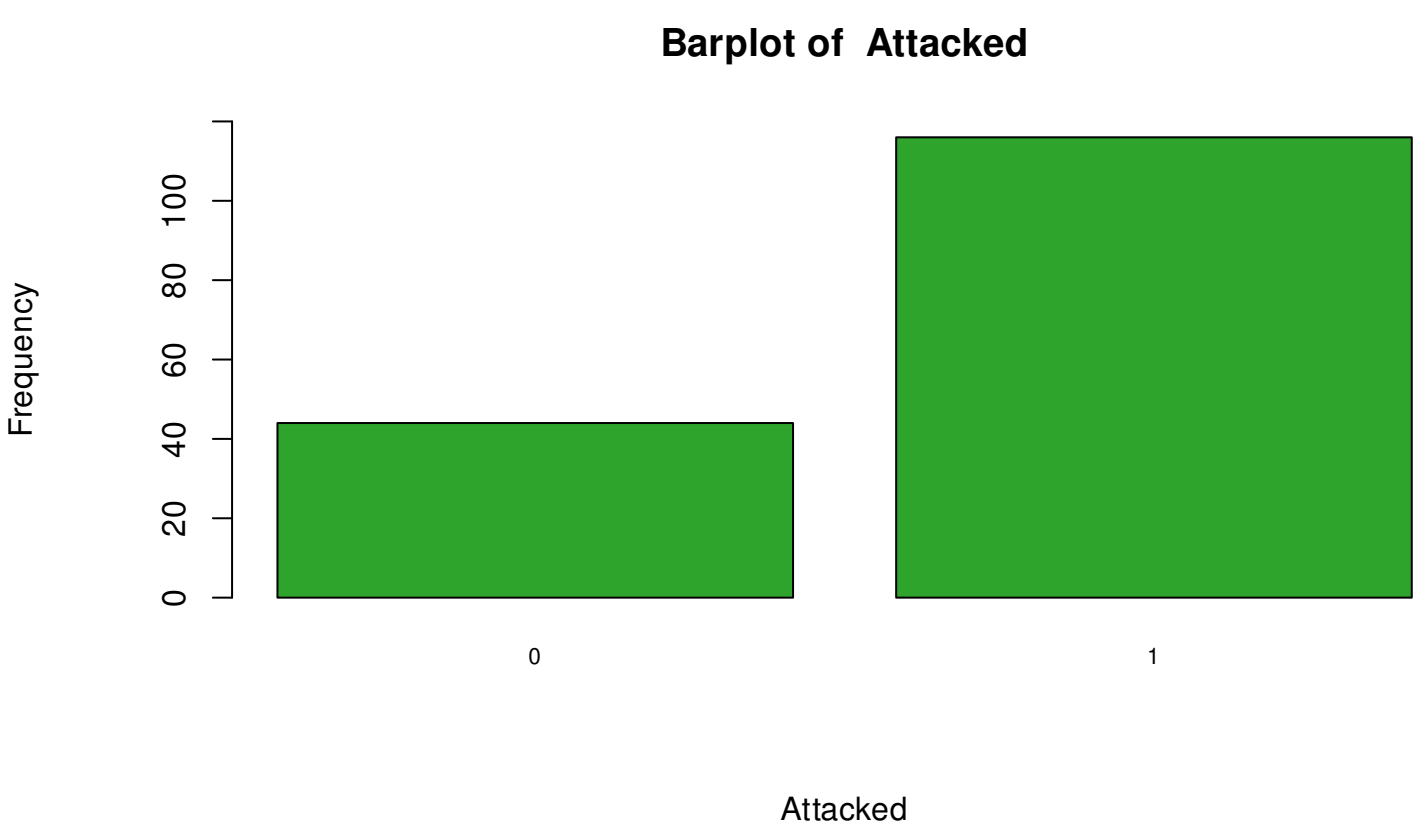
Graphics

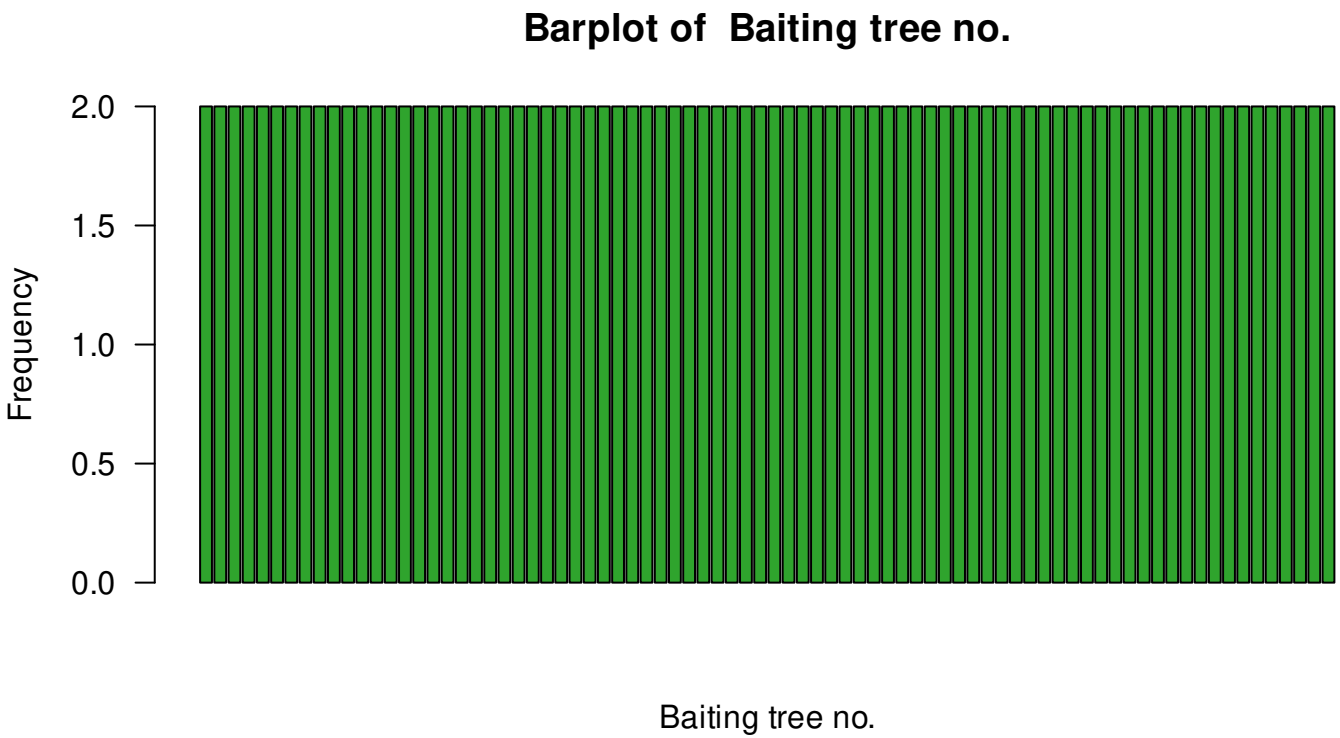
Bar-Plots

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

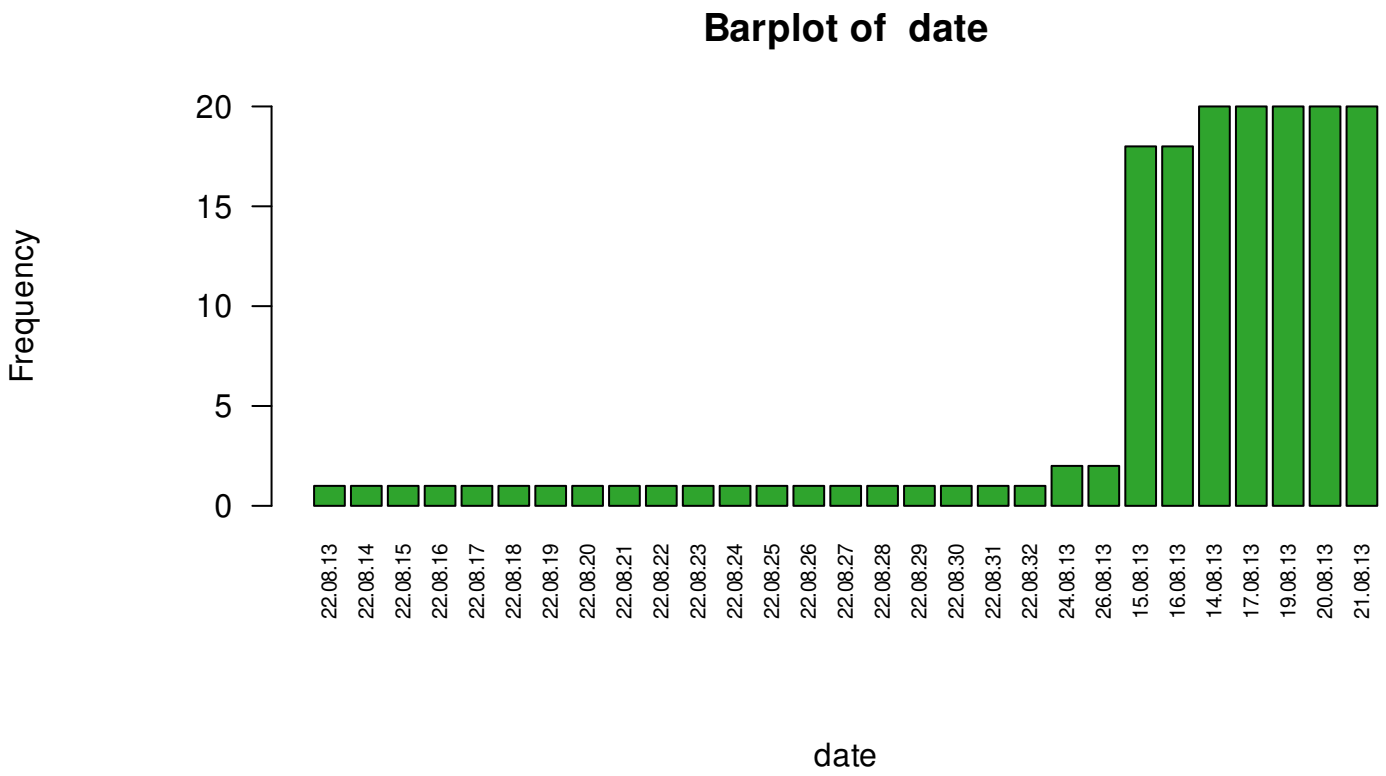


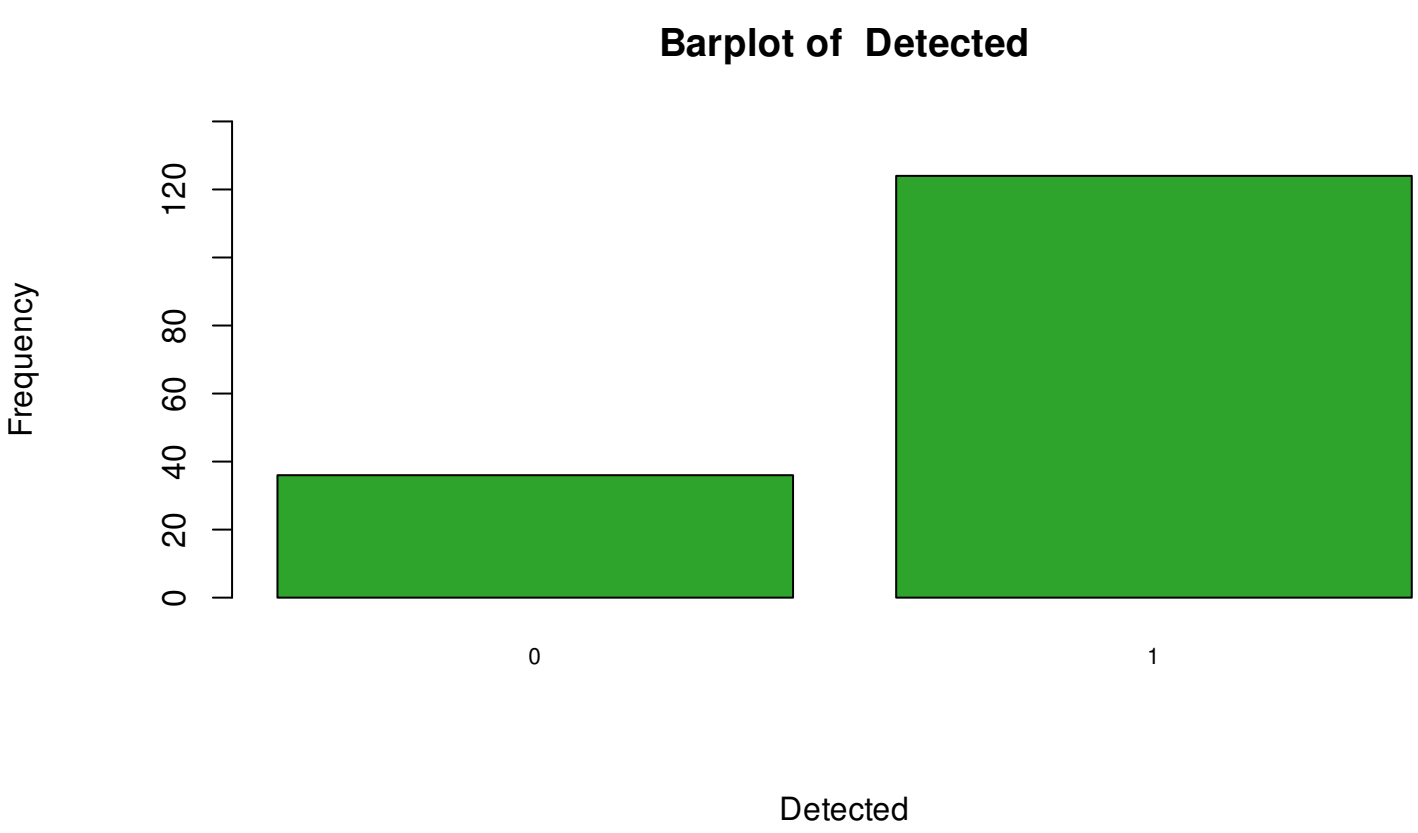

Barplot of ant sample
Frequency
0
20
40
60
80
100
Y
Missing
ant sample

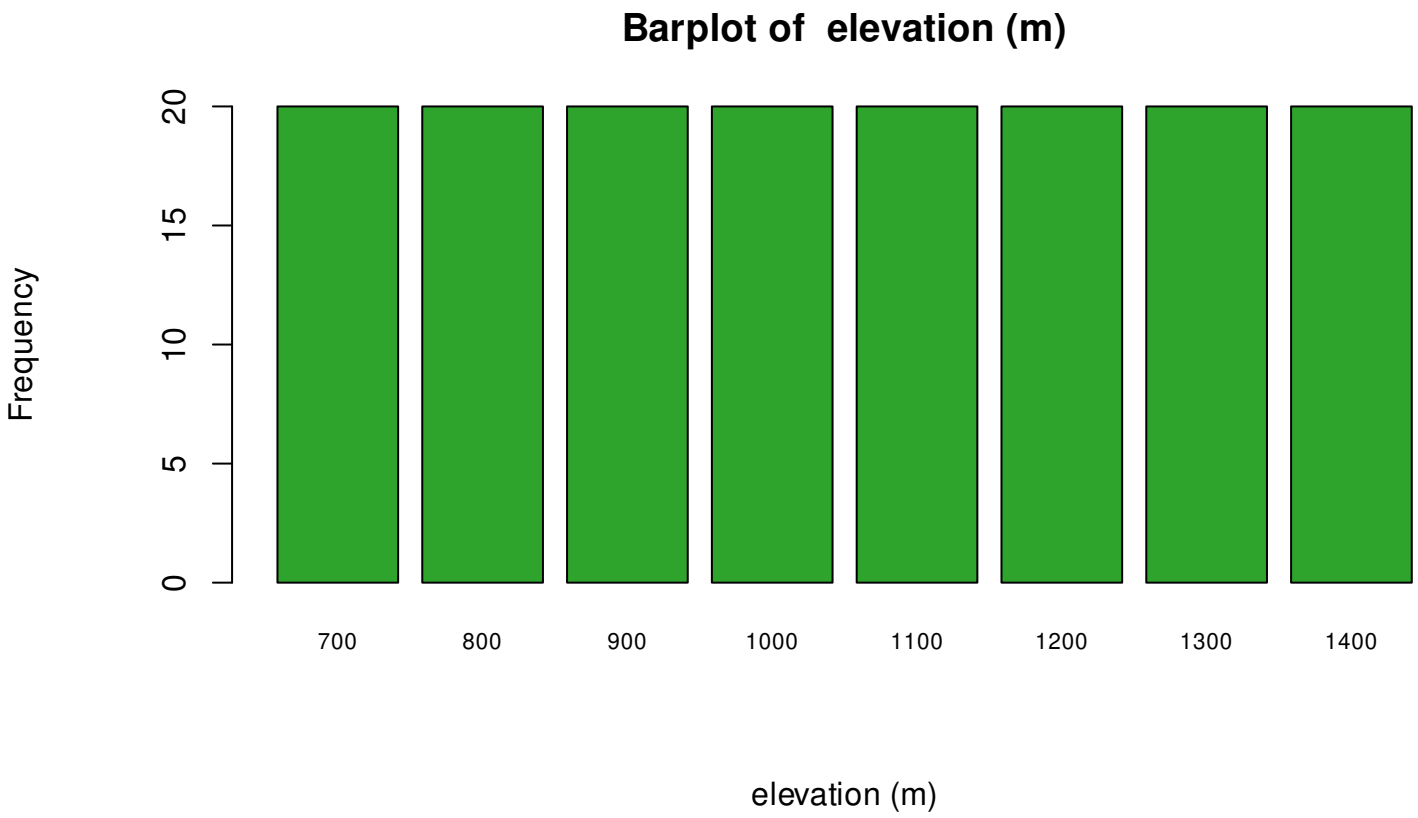

Barplot of Attacked
Frequency
0
20
40
60
80
100
0
1
Attacked

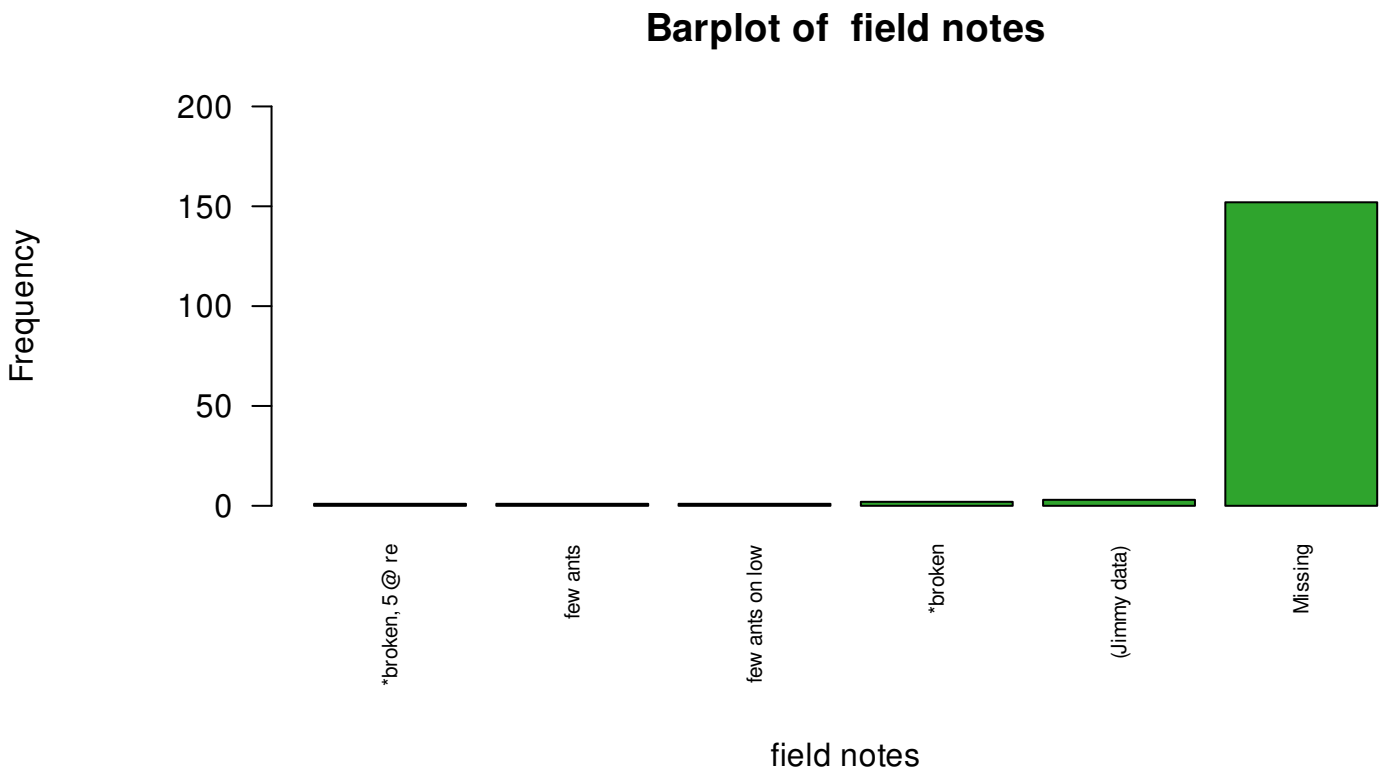


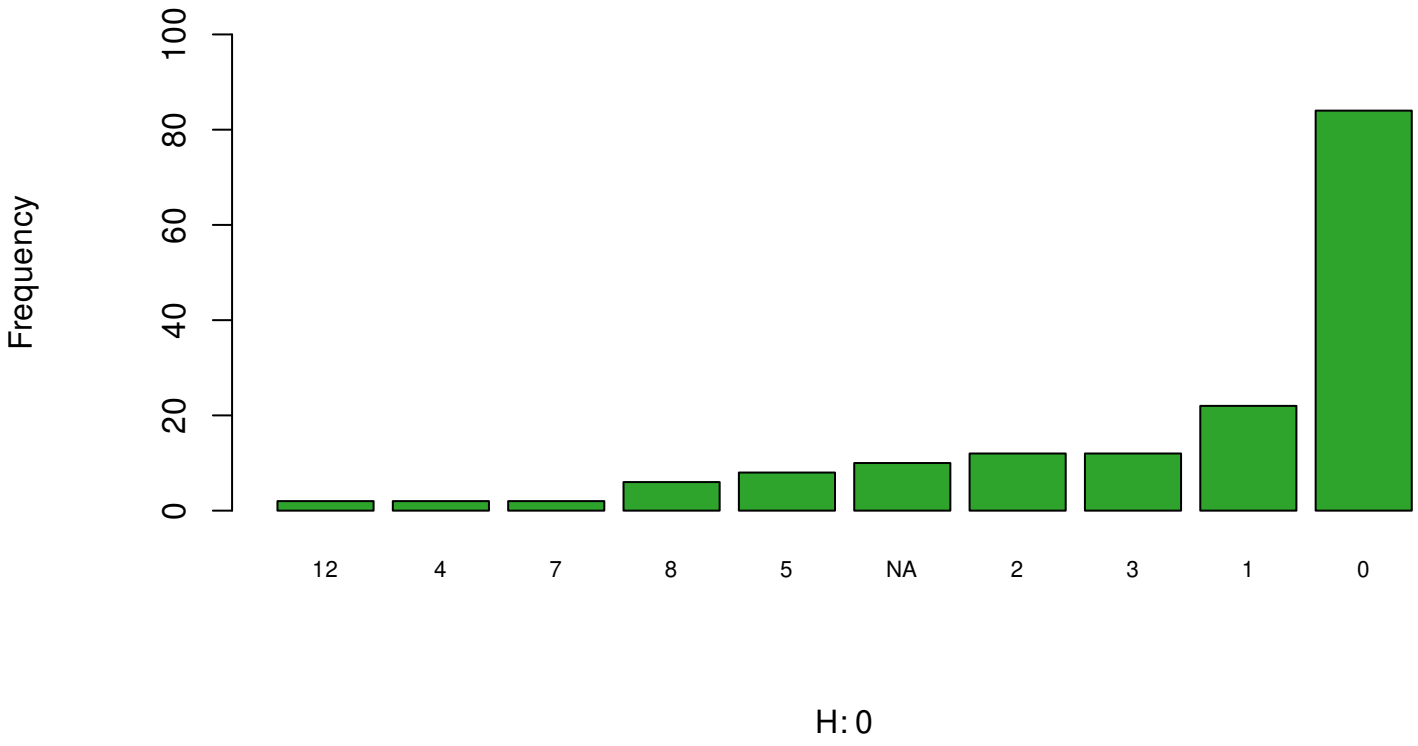
More than 40 categories. Category labels are not displayed.

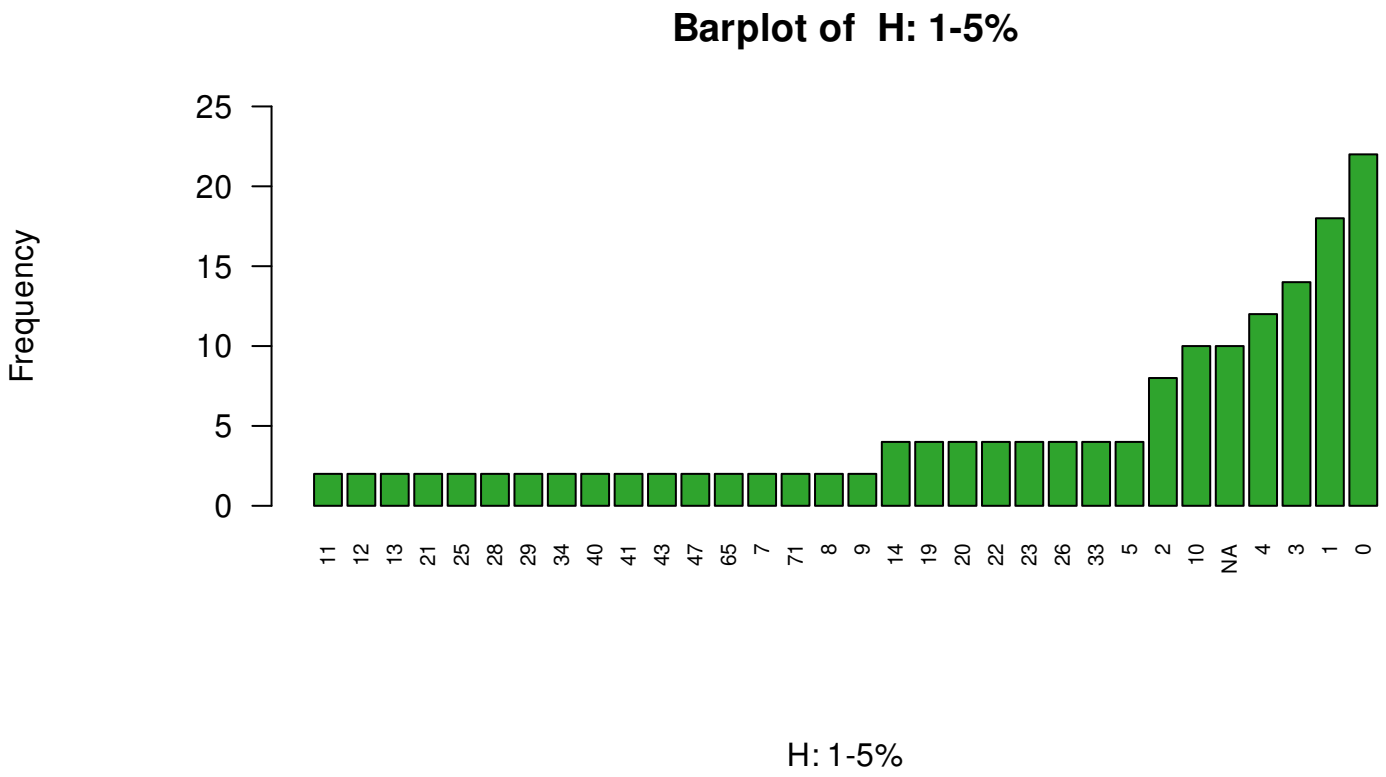

Barplot of date
Frequency
20
15
10
5
0
22.08.13
22.08.14
22.08.15
22.08.16
22.08.17
22.08.18
22.08.19
22.08.20
22.08.21
22.08.22
22.08.23
22.08.24
22.08.25
22.08.26
22.08.27
22.08.28
22.08.29
22.08.30
22.08.31
22.08.32
24.08.13
26.08.13
15.08.13
16.08.13
14.08.13
17.08.13
19.08.13
20.08.13
21.08.13
date

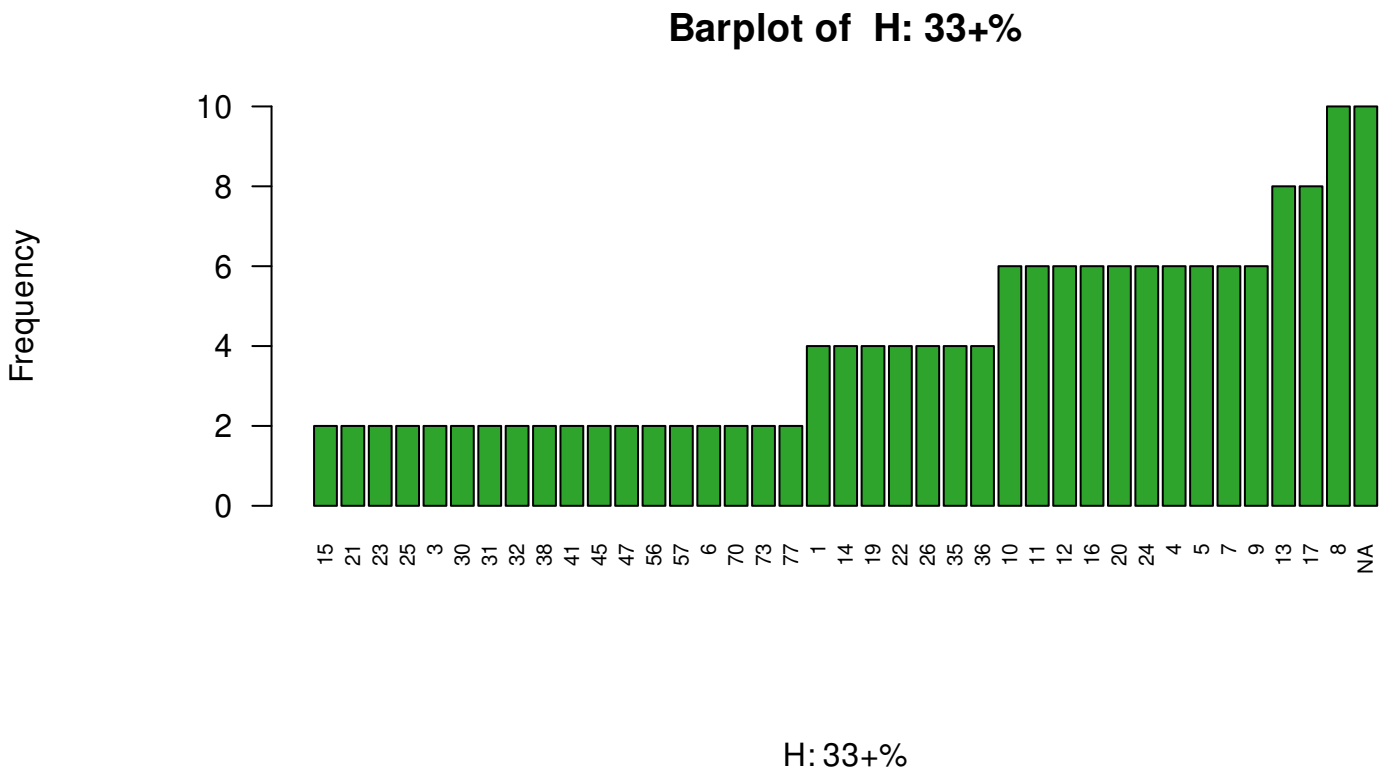

Barplot of Detected
Frequency
0
20
40
60
80
120
0
1
Detected

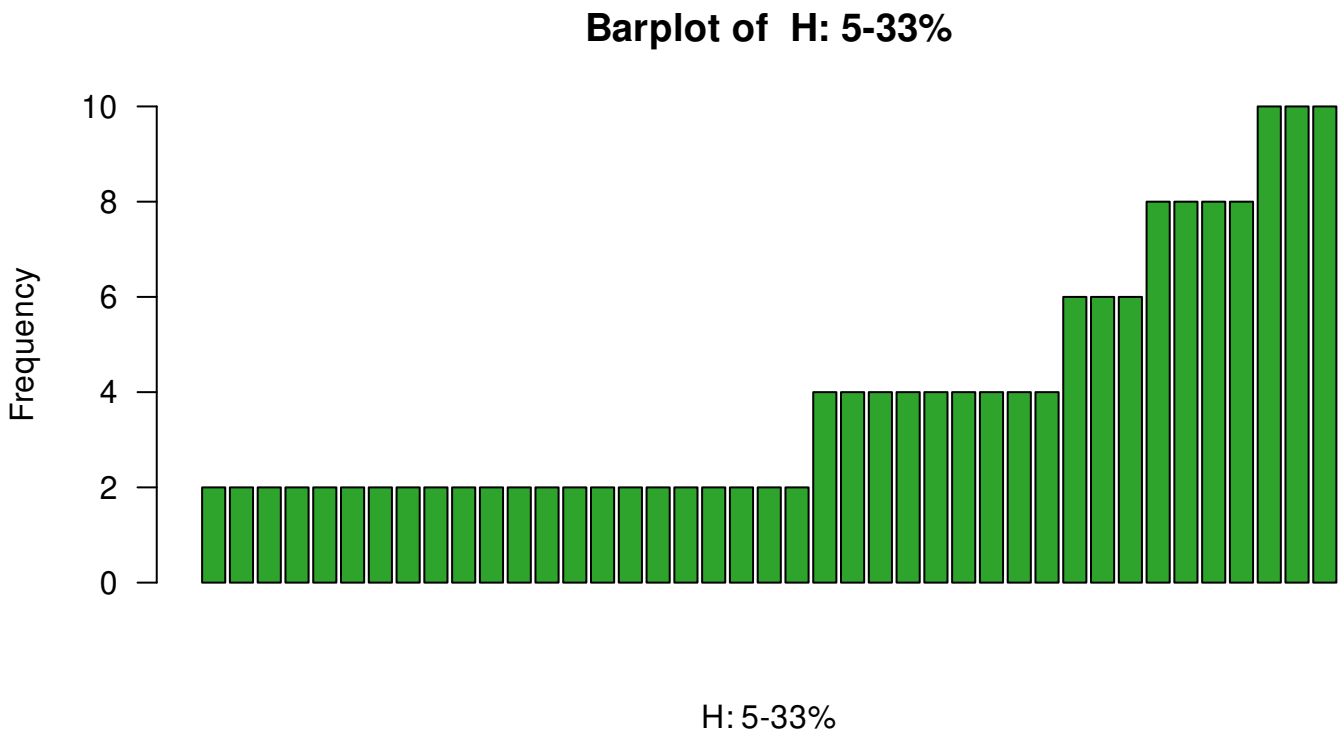

Barplot of elevation (m)
Frequency
0
5
10
15
20
700
800
900
1000
1100
1200
1300
1400
elevation (m)


Barplot of field notes
Frequency
200
150
100
50
0
*broken, 5 @ re
few ants
few ants on low
*broken
(Jimmy data)
Missing
field notes

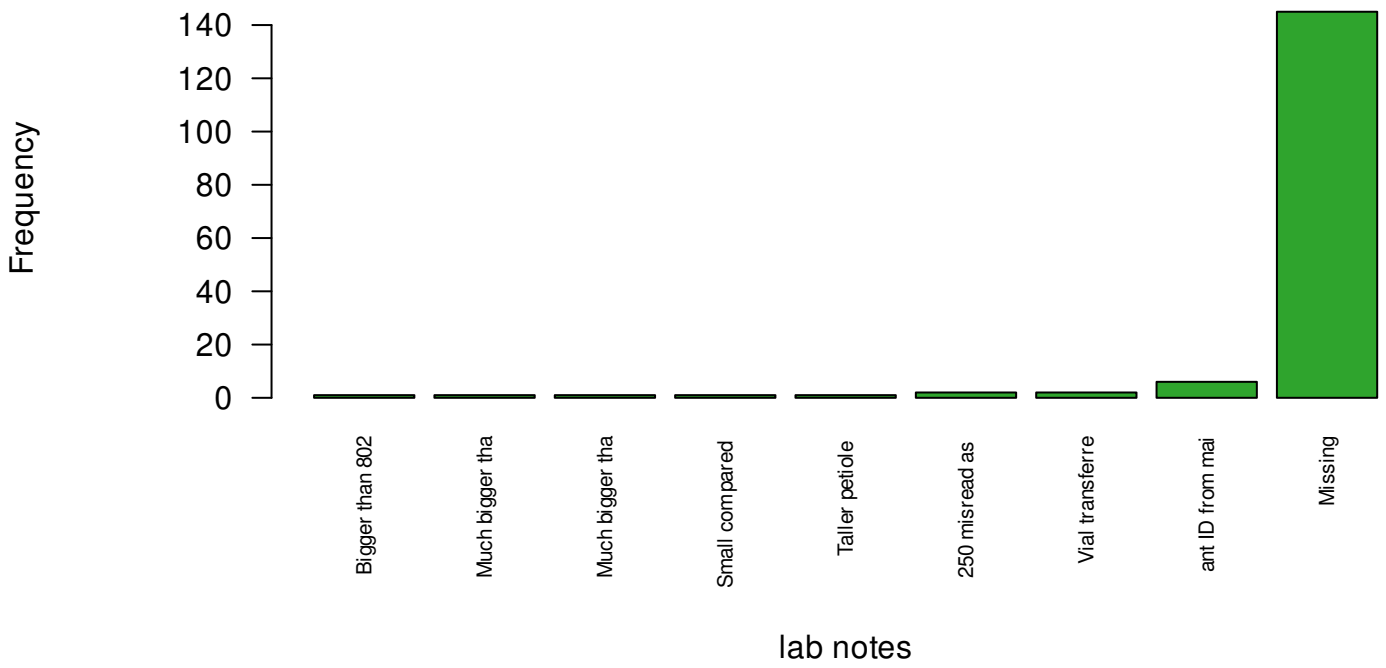

Barplot of H: 0
Frequency
0
20
40
60
80
100
12
4
7
8
5
NA
2
3
1
0
H: 0

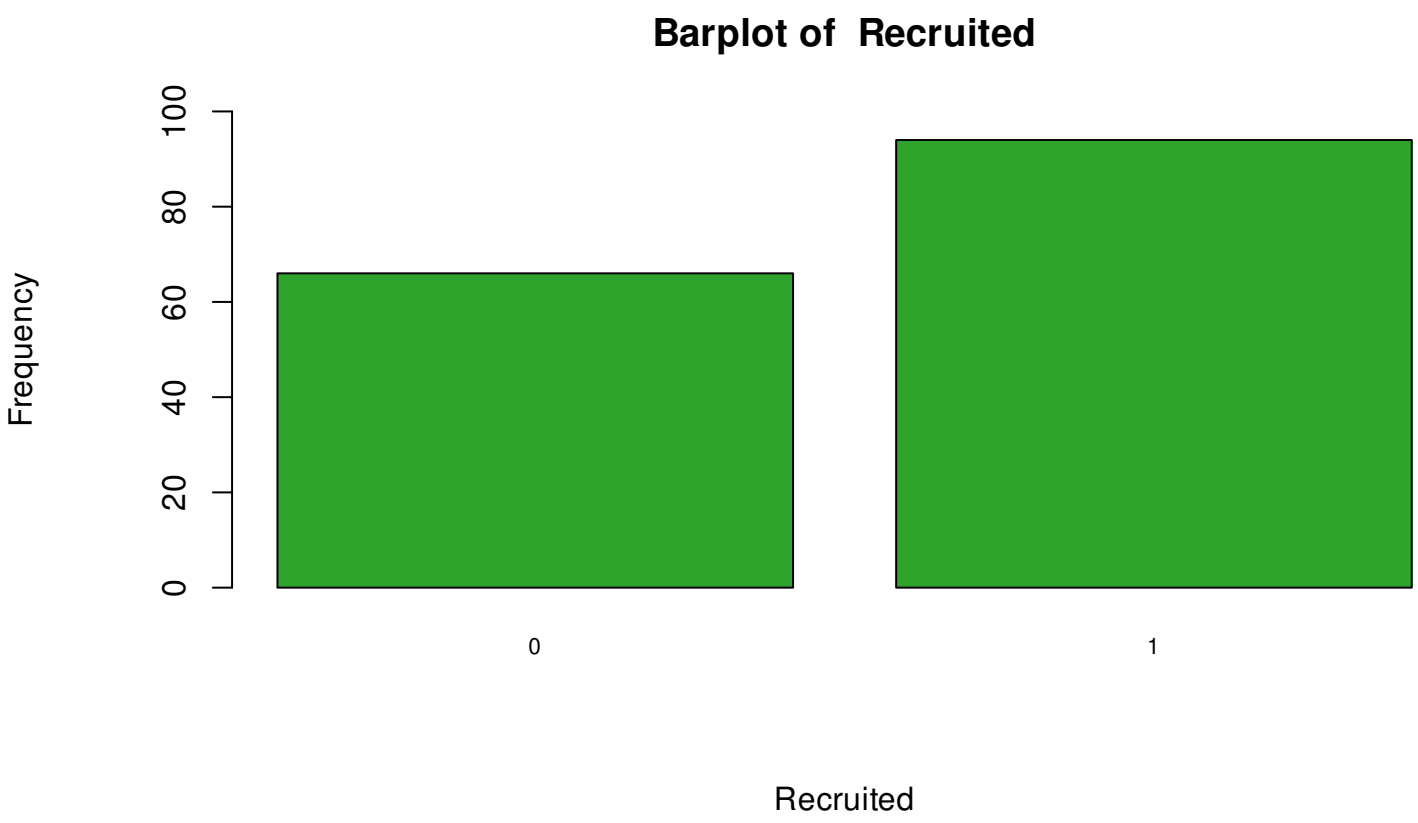

Barplot of H: 1-5%
Frequency
25
20
15
10
5
0
11
12
13
21
25
28
29
34
40
41
43
47
65
7
71
8
9
14
19
20
22
23
26
33
5
2
10
NA
4
3
1
0
H: 1-5%

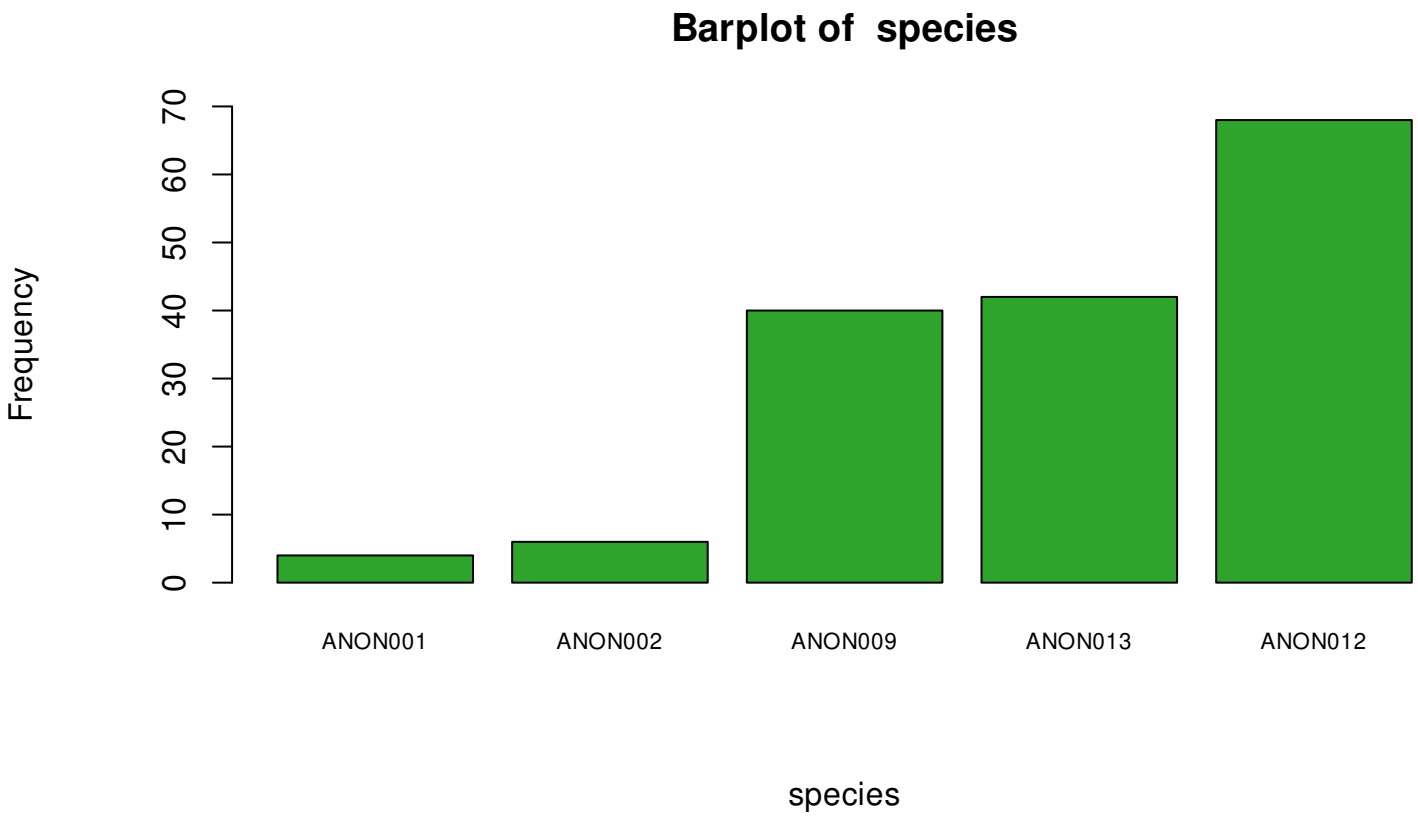

Barplot of H: 33+%
Frequency
10
8
6
4
2
0
15
21
23
25
3
30
31
32
38
41
45
47
56
57
6
70
73
77
1
14
19
22
26
35
36
10
11
12
16
20
24
4
5
7
9
13
17
8
NA
H: 33+%

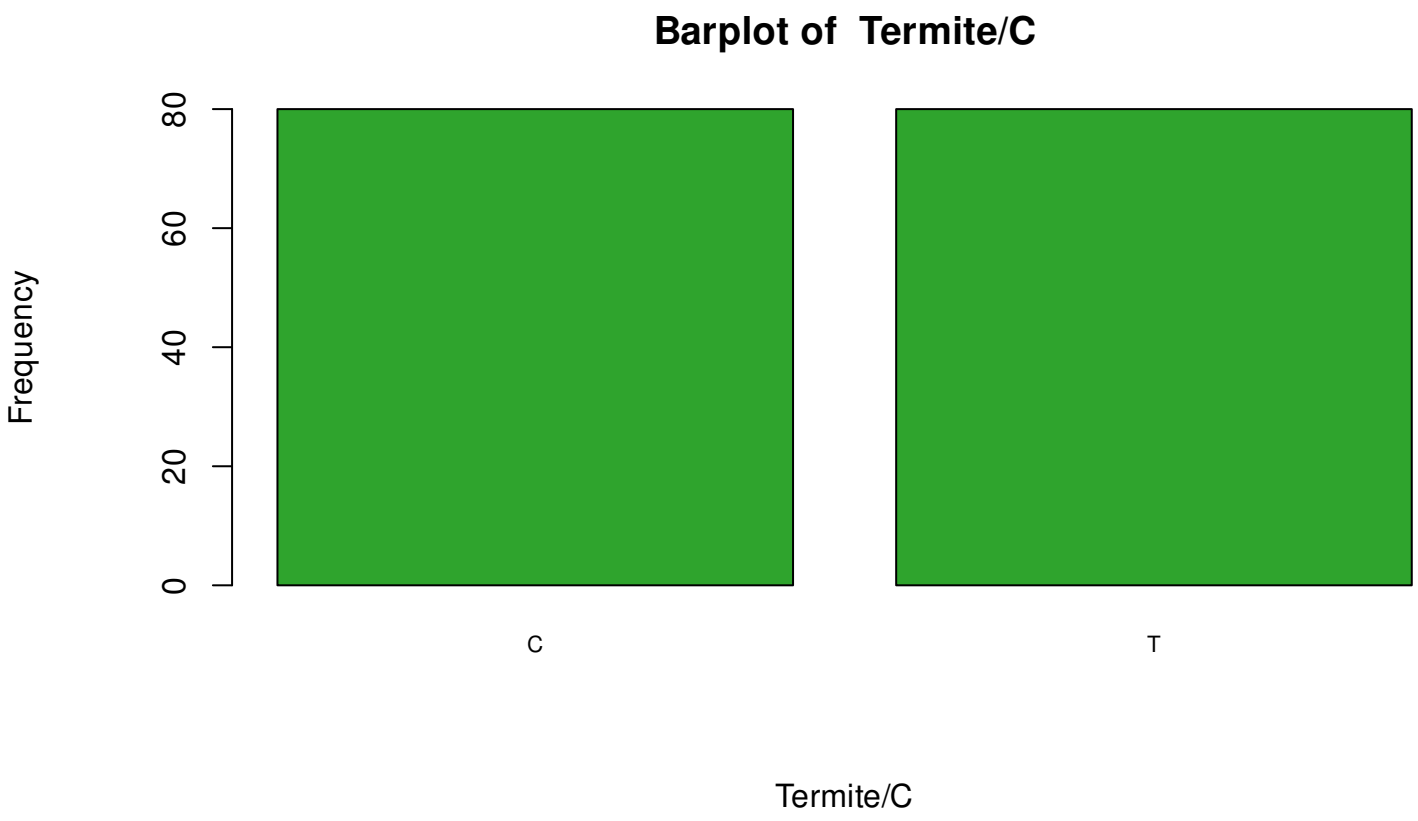


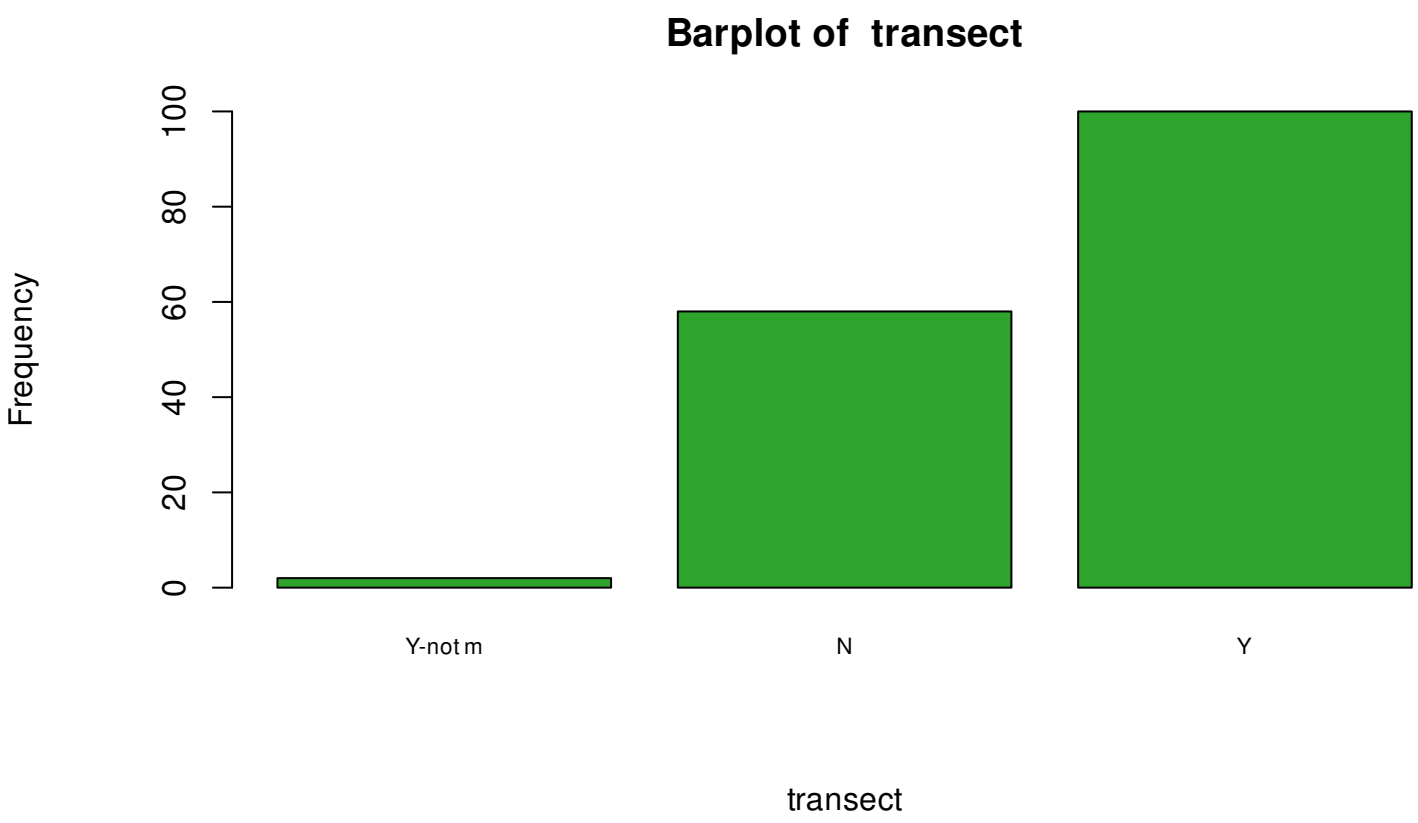
More than 40 categories. Category labels are not displayed.

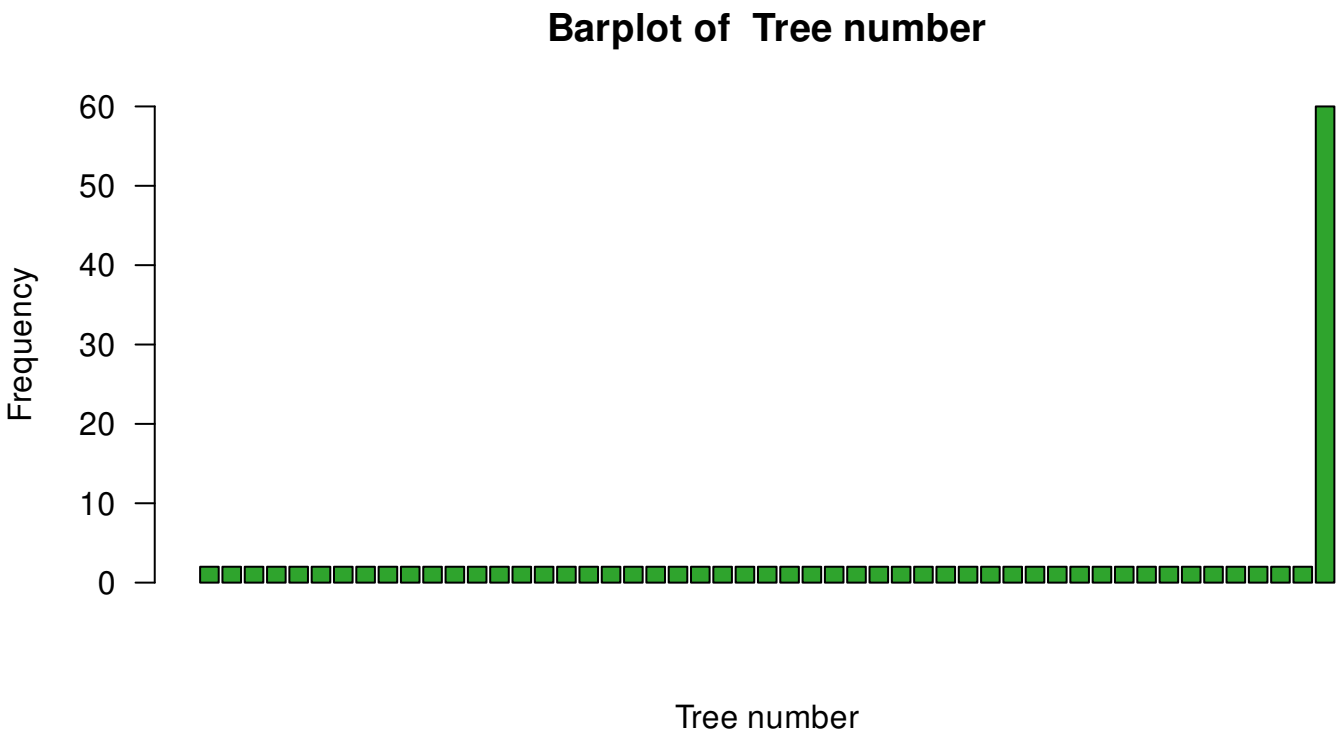

Barplot of lab notes
Frequency
140
120
100
80
60
40
20
0
Bigger than 802
Much bigger tha
Much bigger tha
Small compared
Taller petiole
250 misread as
Vial transferre
ant ID from mai
Missing
lab notes


Barplot of Recruited
Frequency
0
20
40
60
80
100
0
1
Recruited


Barplot of species
Frequency
0
10
20
30
40
50
60
70
ANON001
ANON002
ANON009
ANON013
ANON012
species


Barplot of Termite/C
Frequency
0
20
40
60
80
C
T
Termite/C

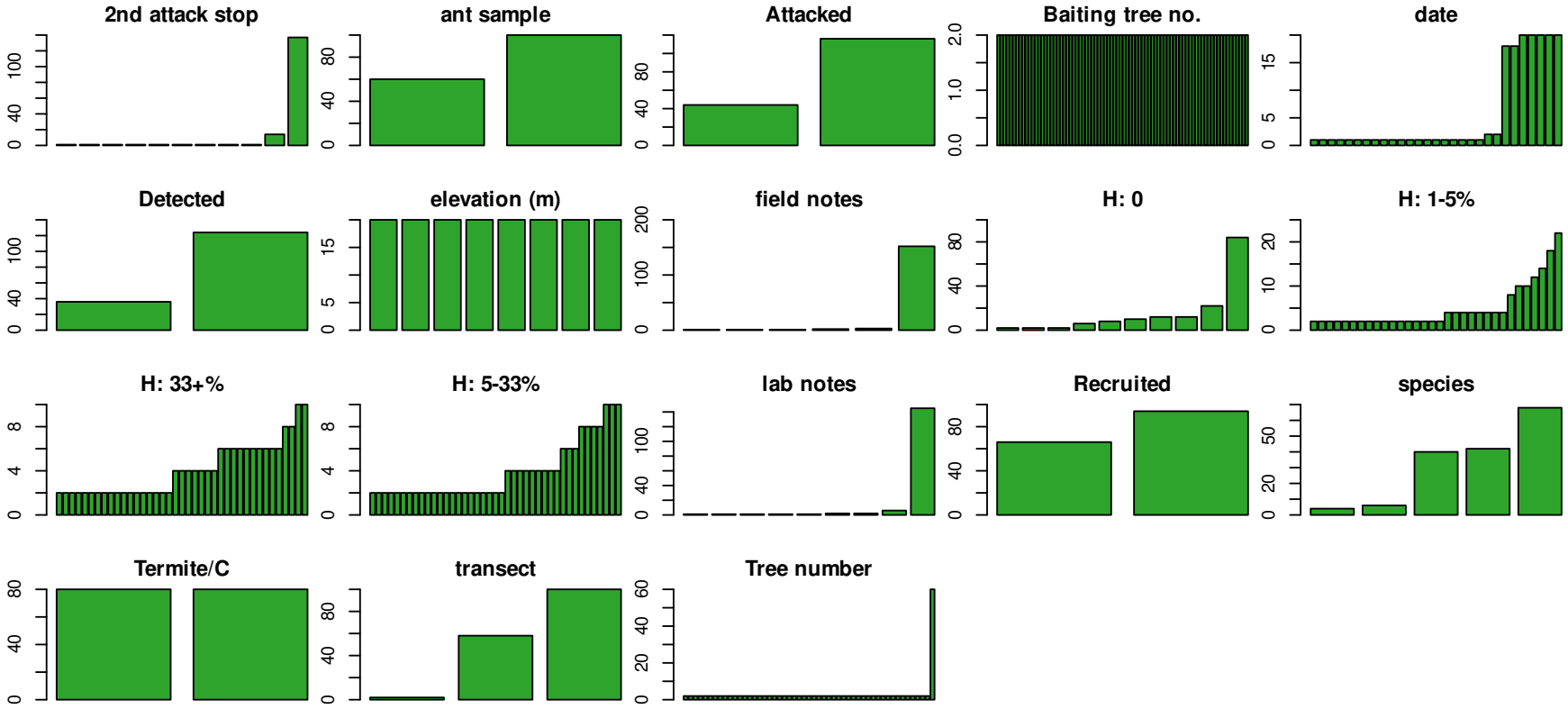

Barplot of transect
Frequency
0
20
40
60
80
100
Y-not m
N
Y
transect



More than 40 categories. Category labels are not displayed.

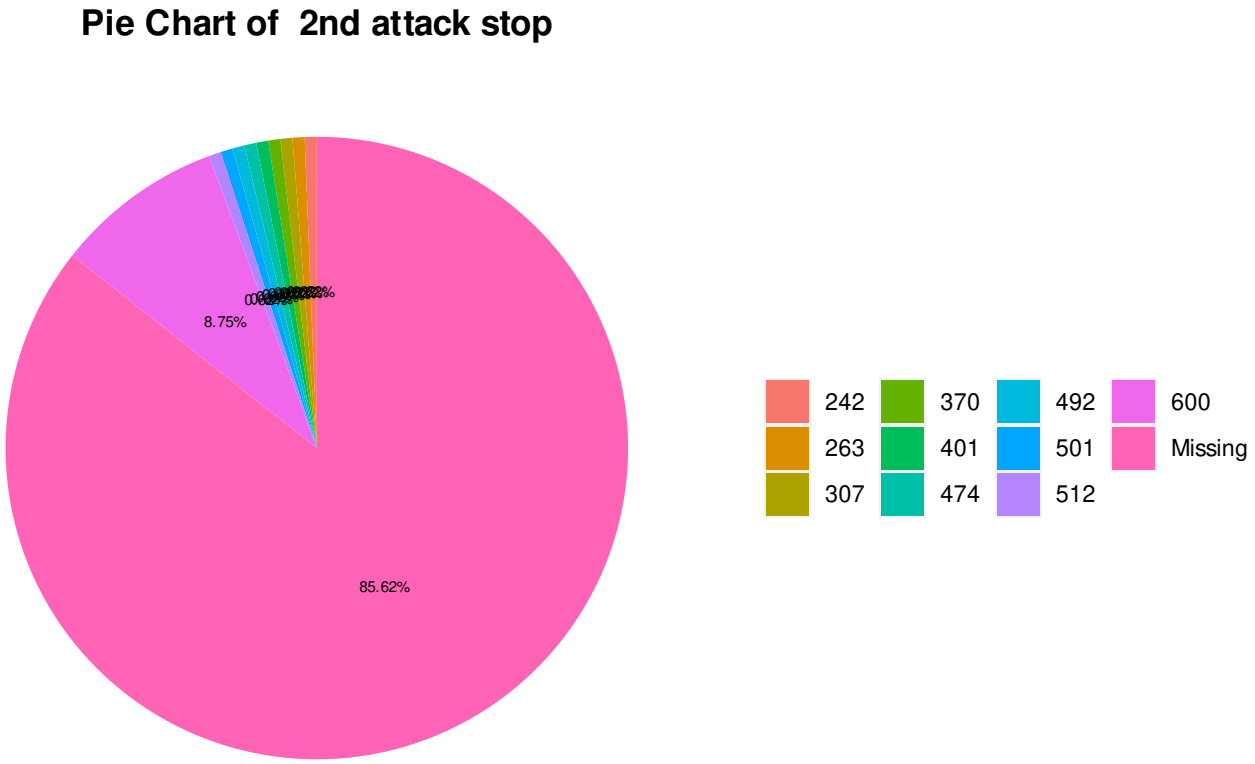
Bar-Plots Summary

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

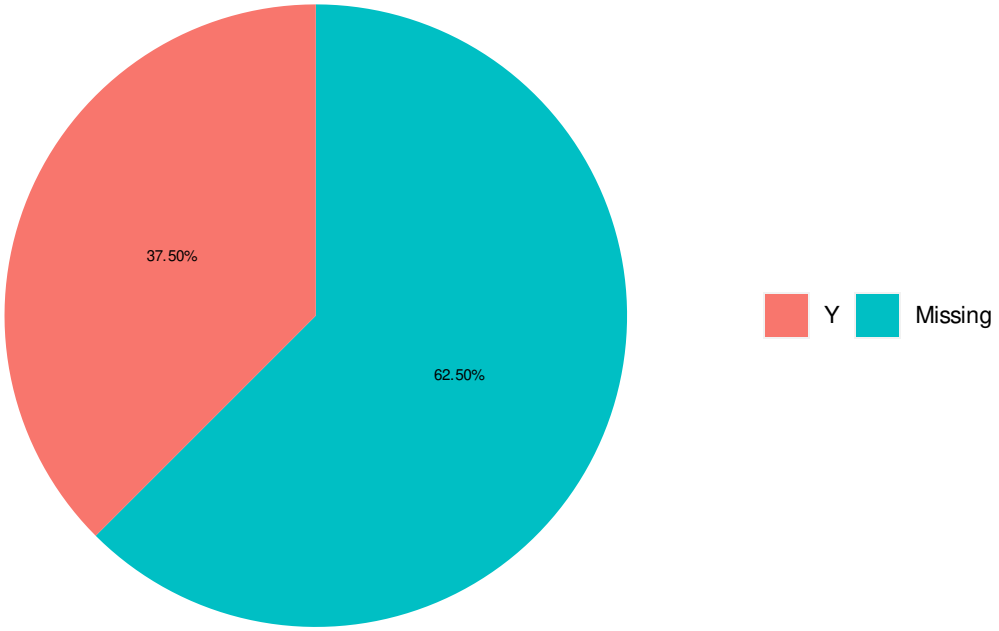


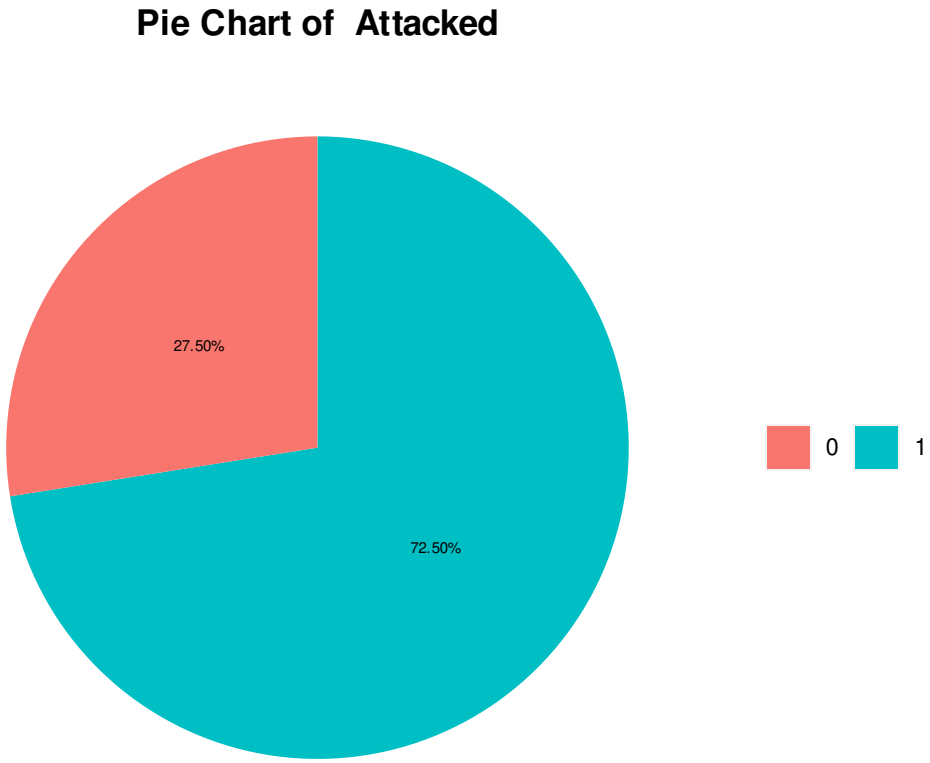
Pie Plots

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

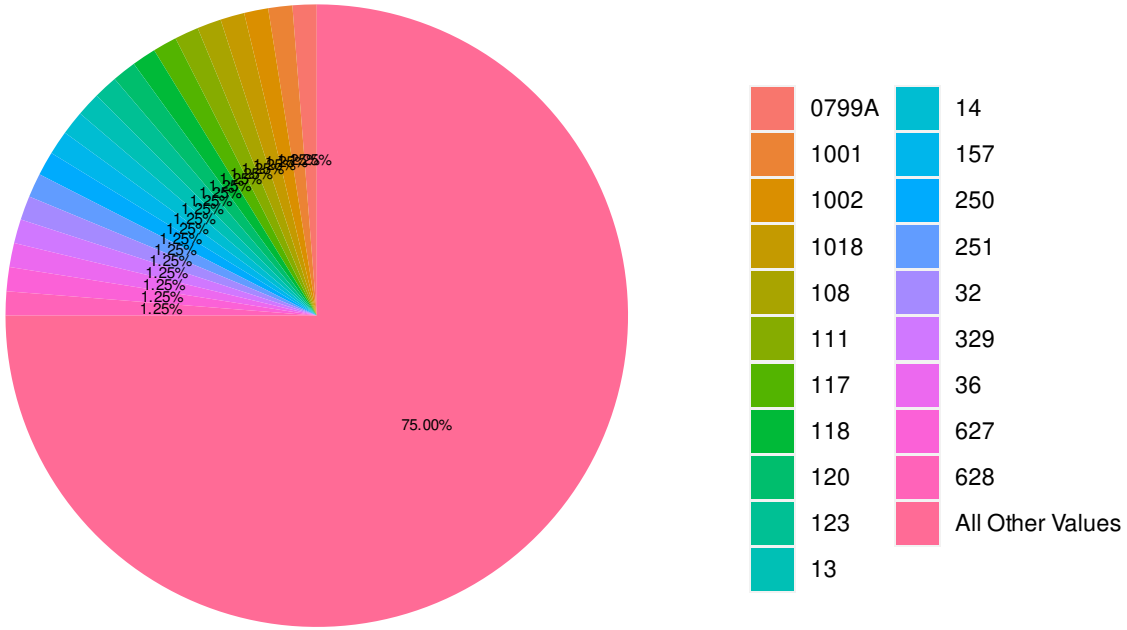


Pie Chart of ant sample

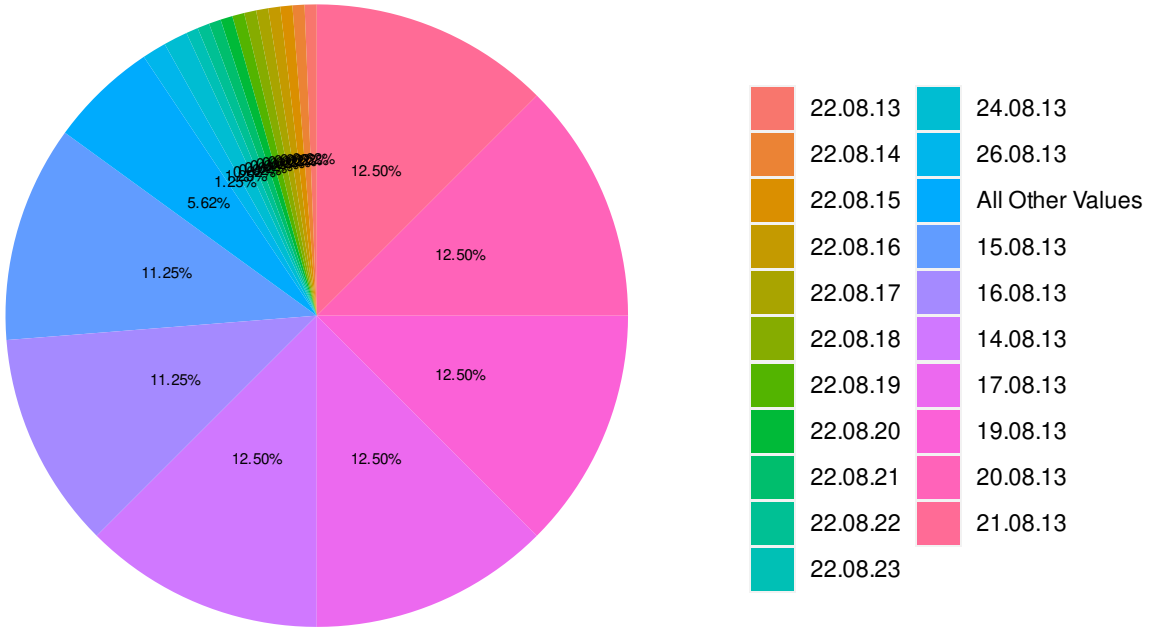


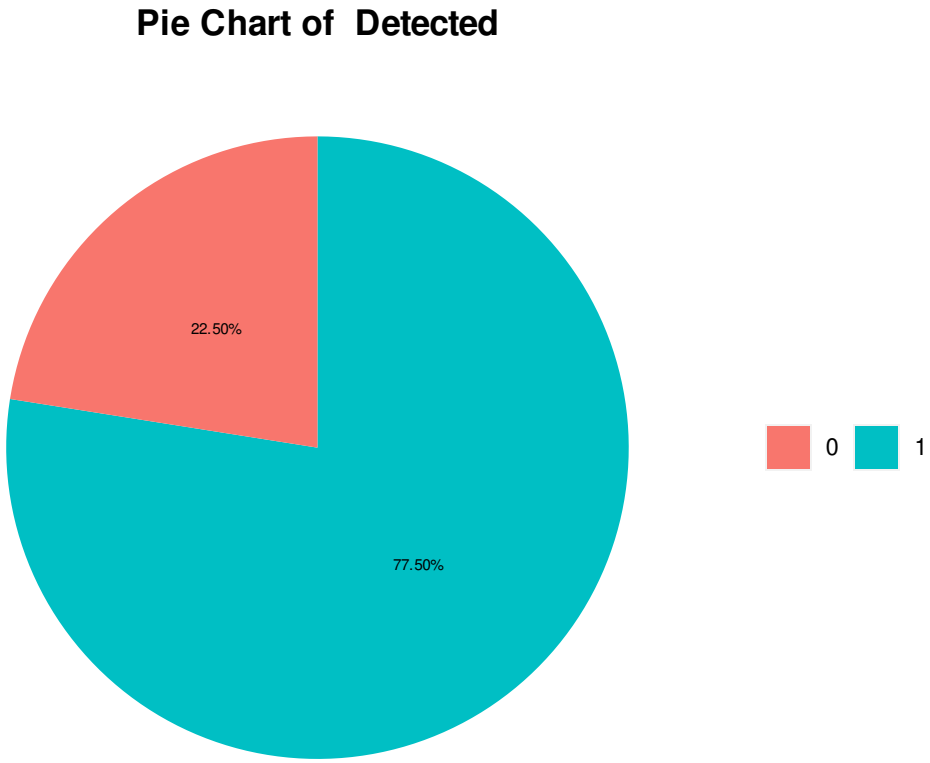

Pie Chart of Attacked
27.50%
72.50%
0
1

Pie Chart of Baiting tree no.

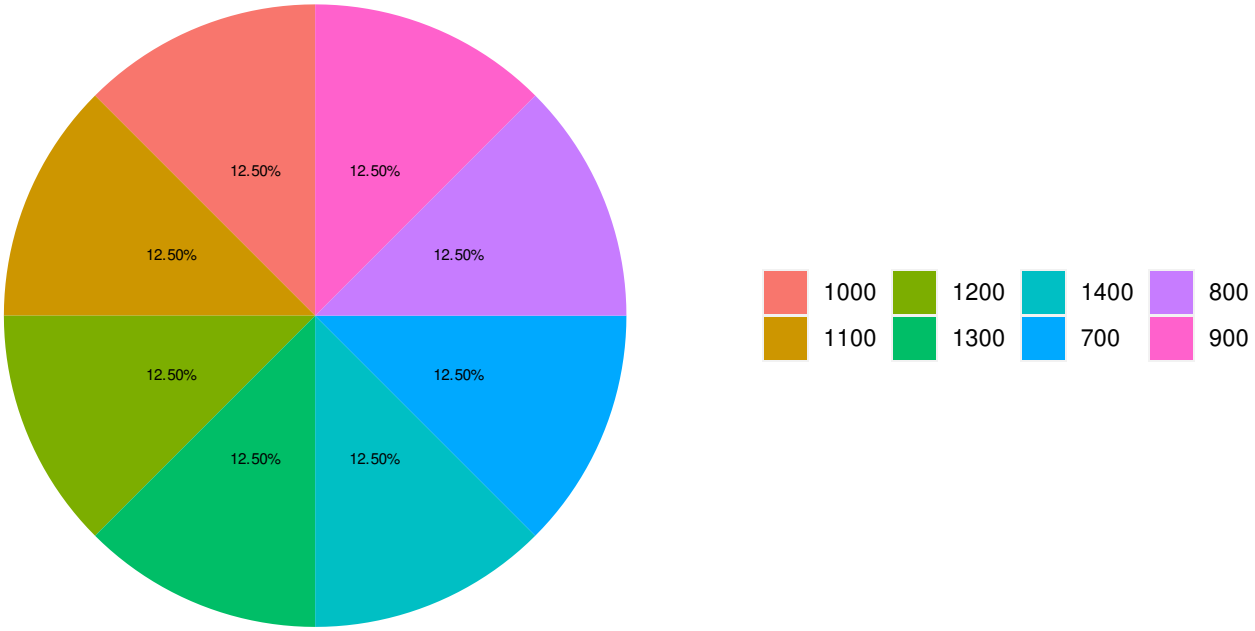


Pie Chart of date

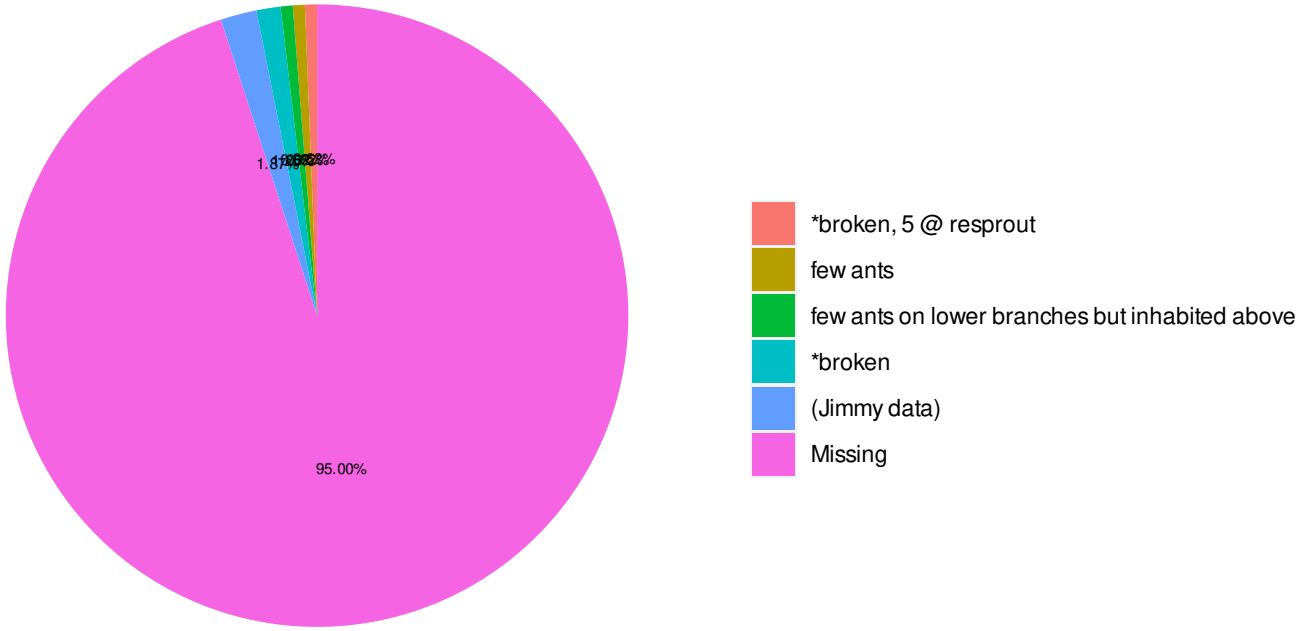



Pie Chart of Detected
22.50%
77.50%
0
1

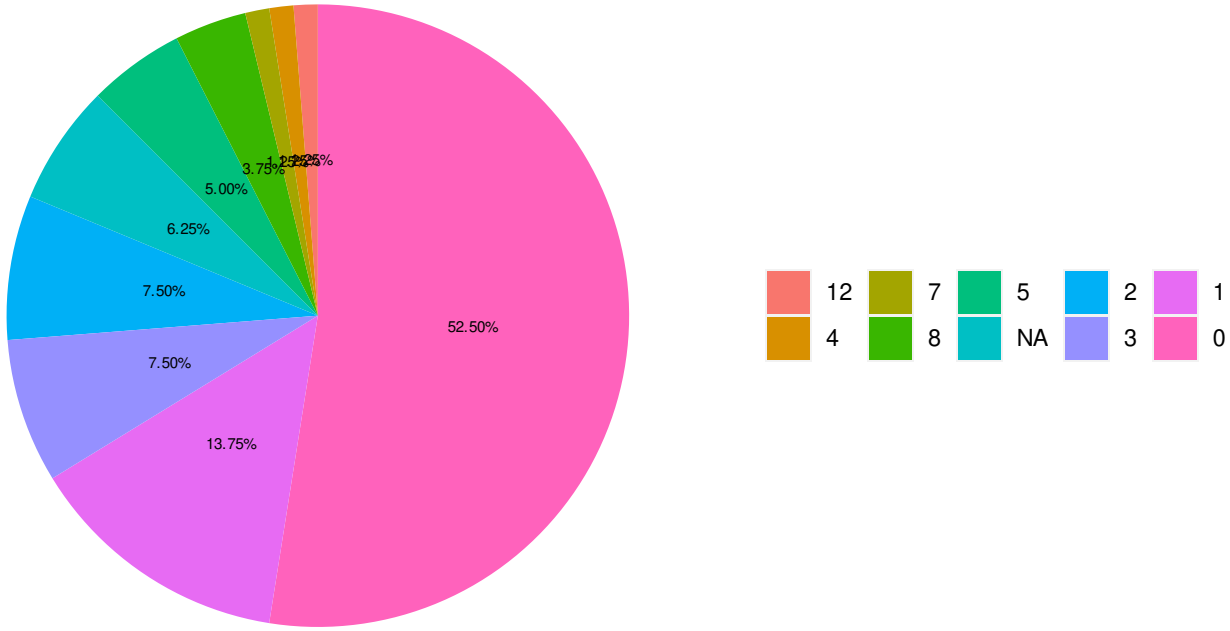
Pie Chart of elevation (m)



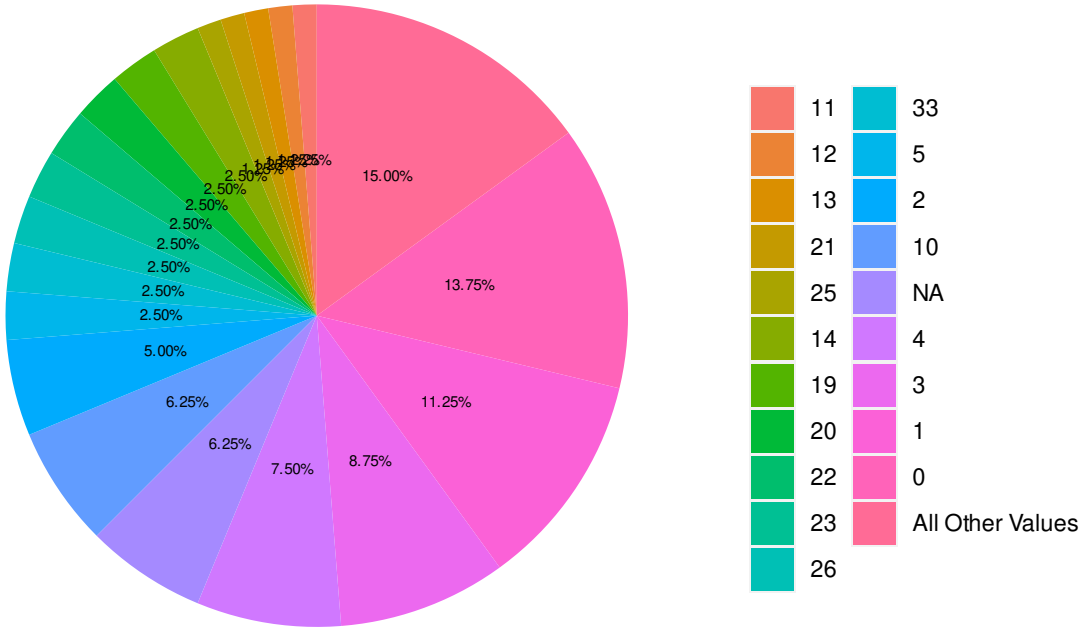
Pie Chart of field notes



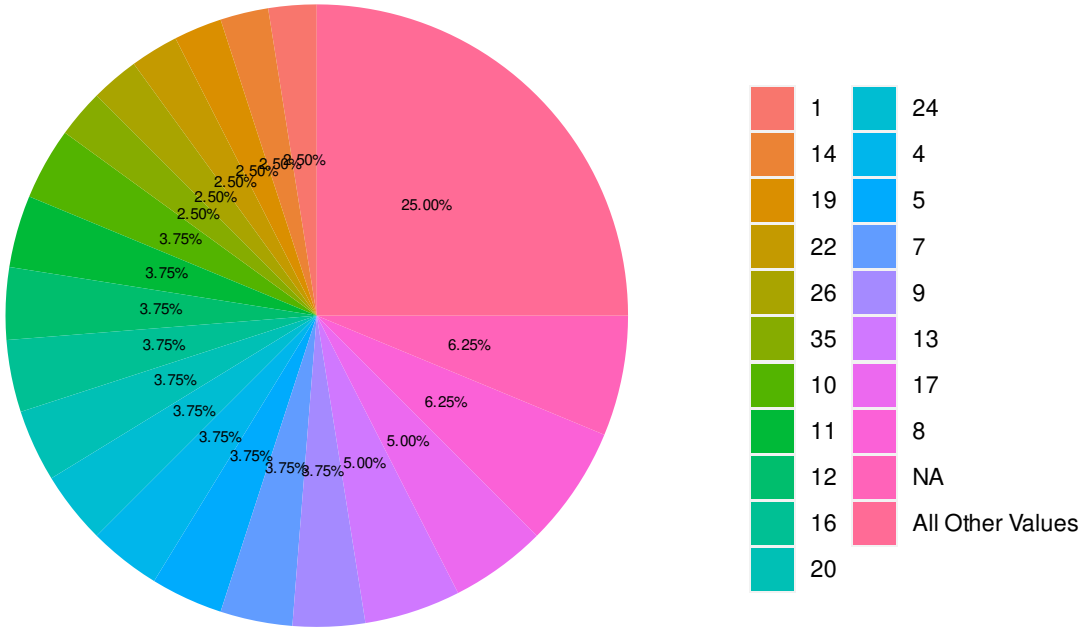
Pie Chart of H: 0



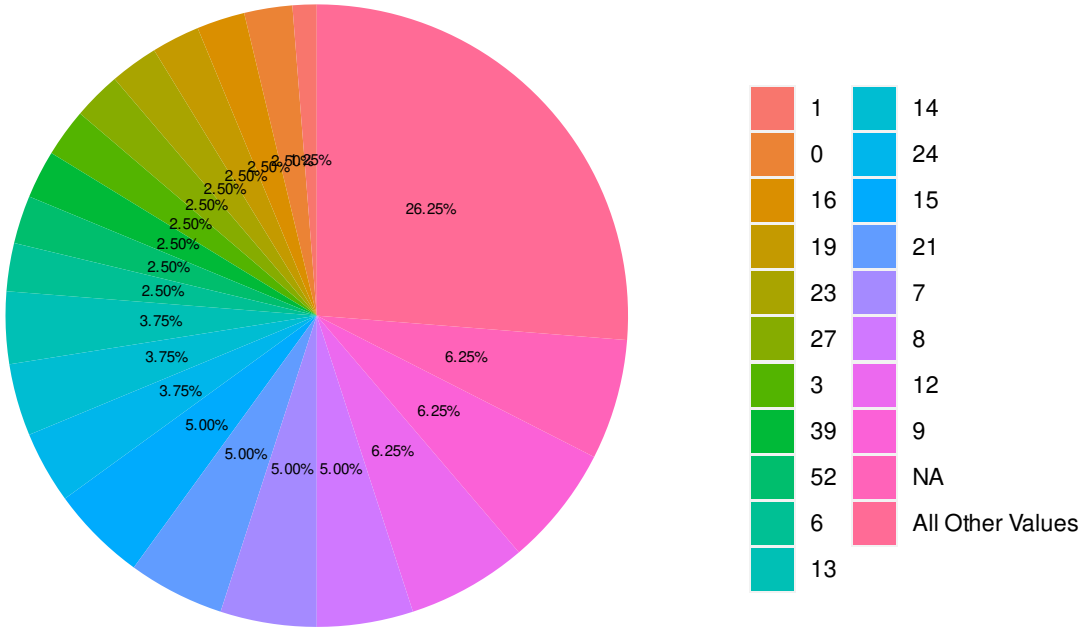
Pie Chart of H: 1-5%



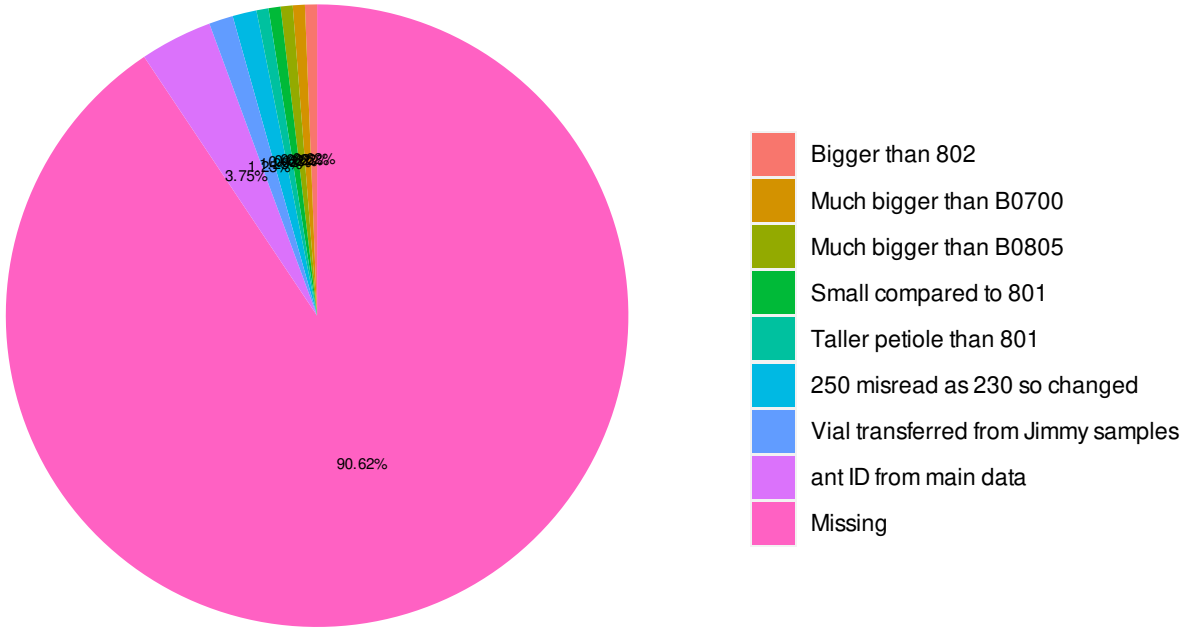
Pie Chart of H: 33+%

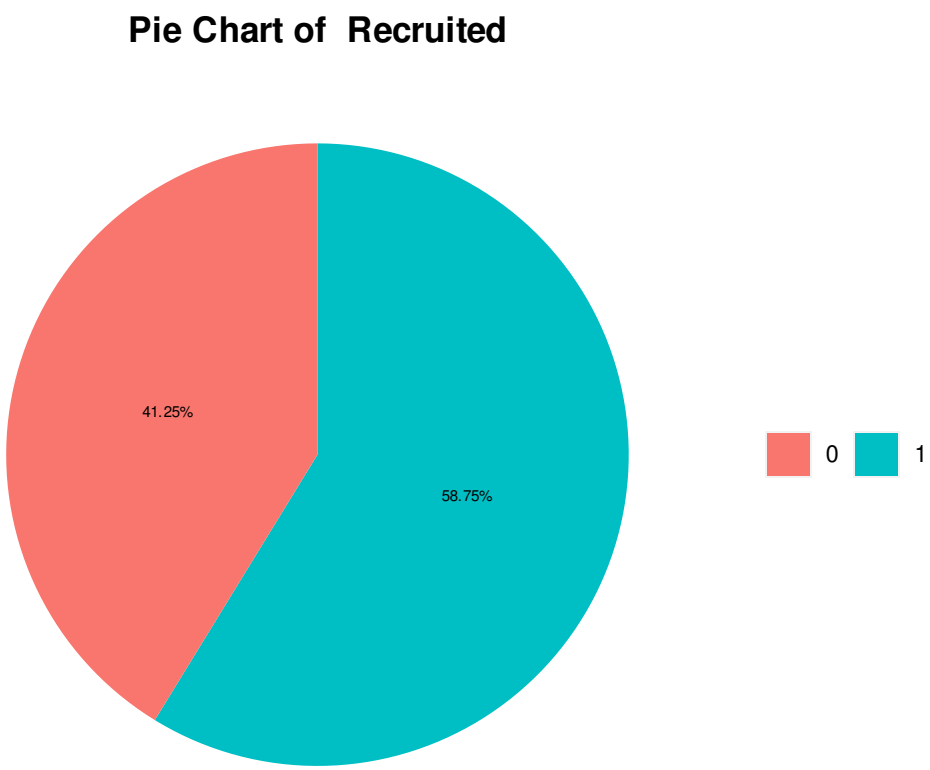


Pie Chart of H: 5-33%

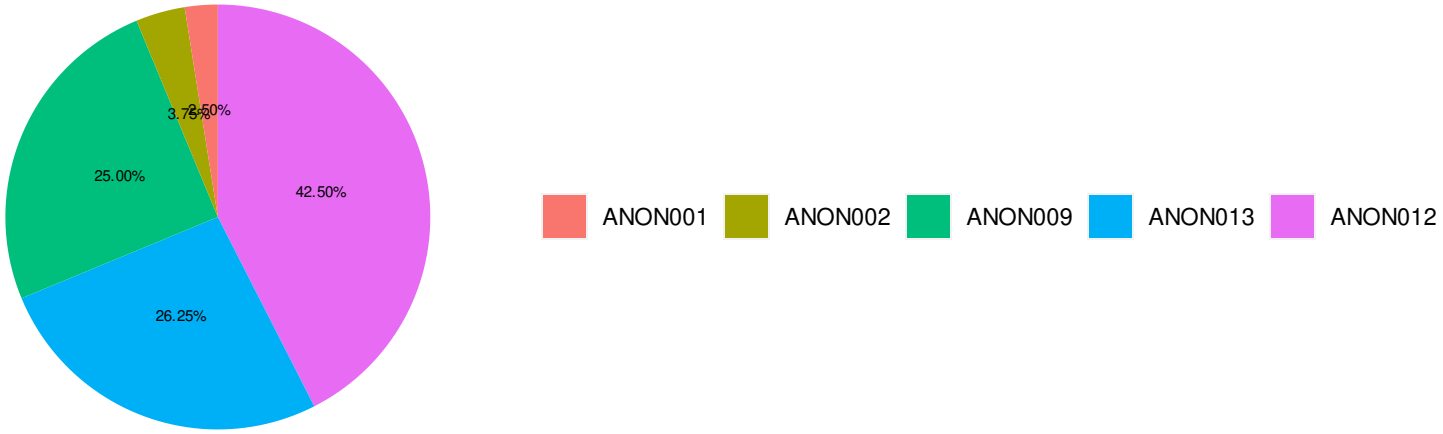


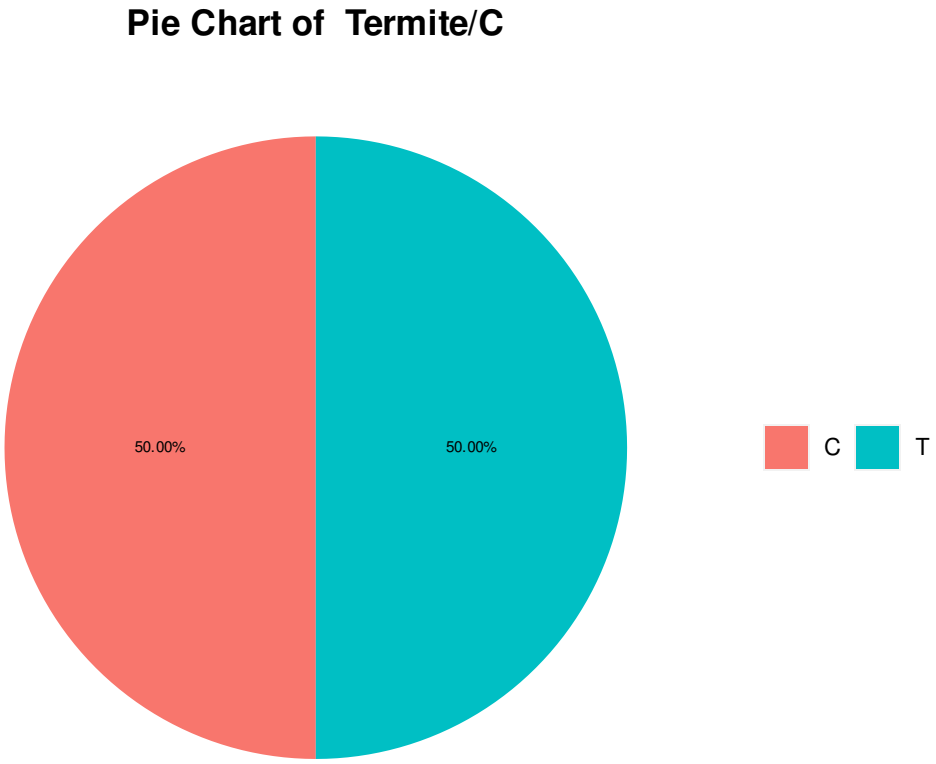
Pie Chart of lab notes



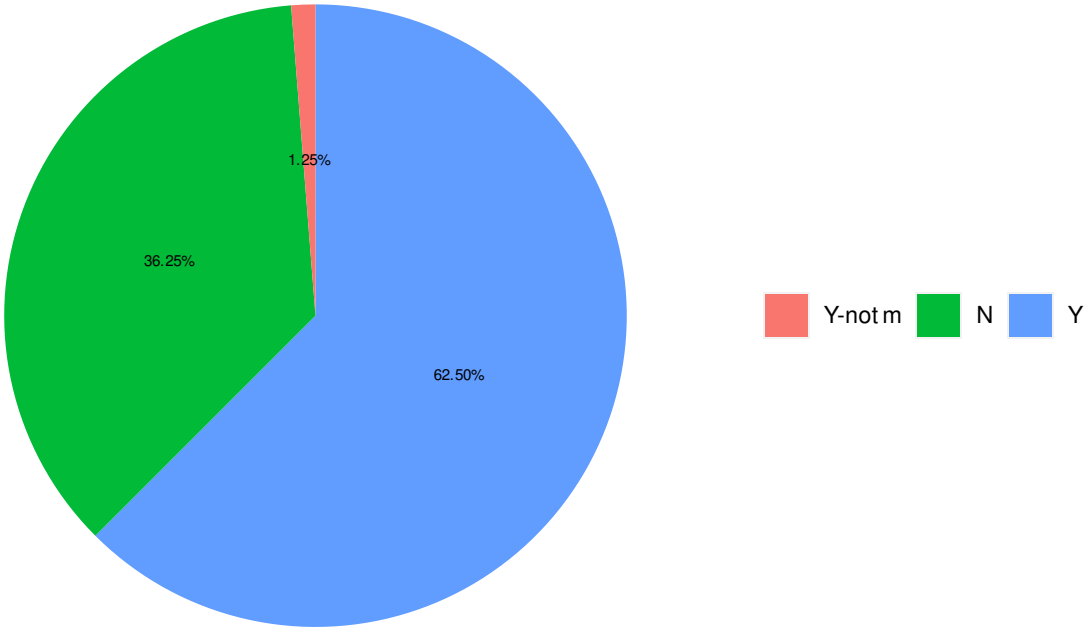

Pie Chart of Recruited
41.25%
58.75%
0
1

Pie Chart of species

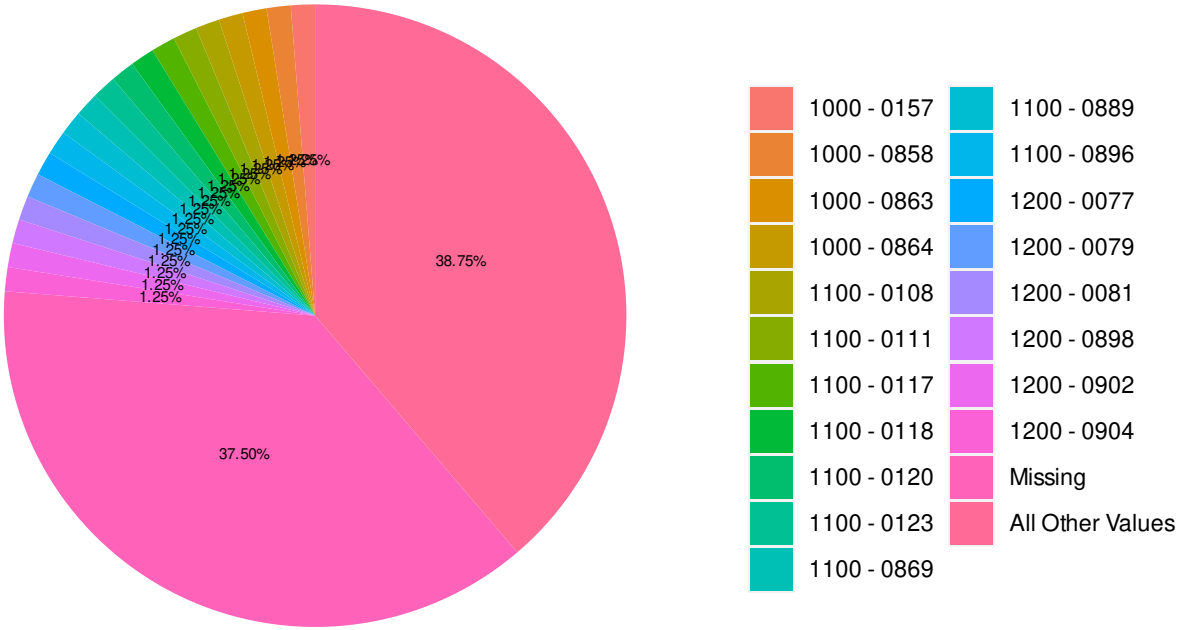



Pie Chart of Termite/C
50.00%
50.00%
C
T

Pie Chart of transect



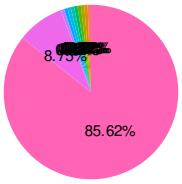
Pie Chart of Tree number



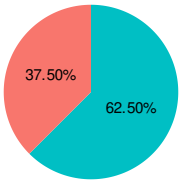
Pie Plots Summary

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

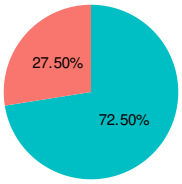
2nd attack stop



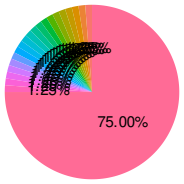
ant sample



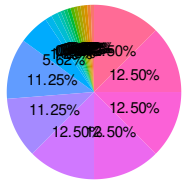
Attacked



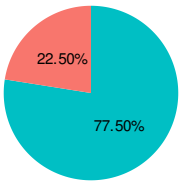
Baiting tree no.



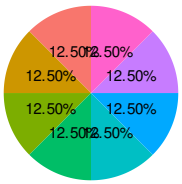
date



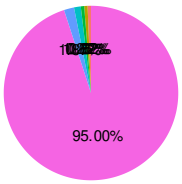
Detected



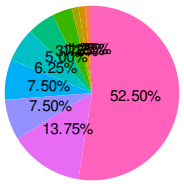
elevation (m)



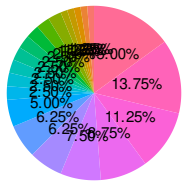
field notes



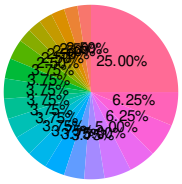
H: 0



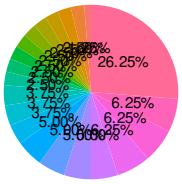
H: 1-5%



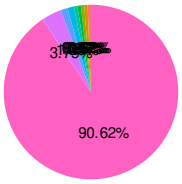
H: 33+%



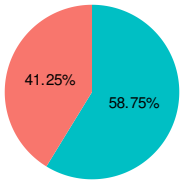
H: 5-33%



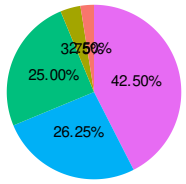
lab notes



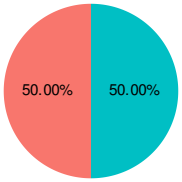
Recruited



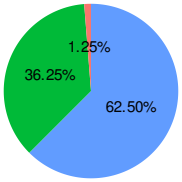
species



Termite/C



transect



Tree number