

Overview

Dataset statistics	
Number of variables	15
Number of observations	712
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	31
Duplicate rows (%)	4.4%
Total size in memory	272.8 KiB
Average record size in memory	392.3 B

Variable types	
Categorical	8
Numeric	4
Boolean	3


Alerts

Dataset has 31 (4.4%) duplicate rows	Duplicates
deck is highly imbalanced (51.4%)	Imbalance
sibsp has 469 (65.9%) zeros	Zeros
parch has 519 (72.9%) zeros	Zeros

Reproduction

Analysis started	2025-04-20 07:13:36.101947
Analysis finished	2025-04-20 07:13:42.270795
Duration	6.17 seconds
Software version	ydata-profiling v4.8.3 (https://github.com/ydataai/ydata-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Titanic%20EDA%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%20%22

Variables

Select Columns 

survived
Categorical

Distinct	2
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	62.1 KiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	712	
Distinct characters	2	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

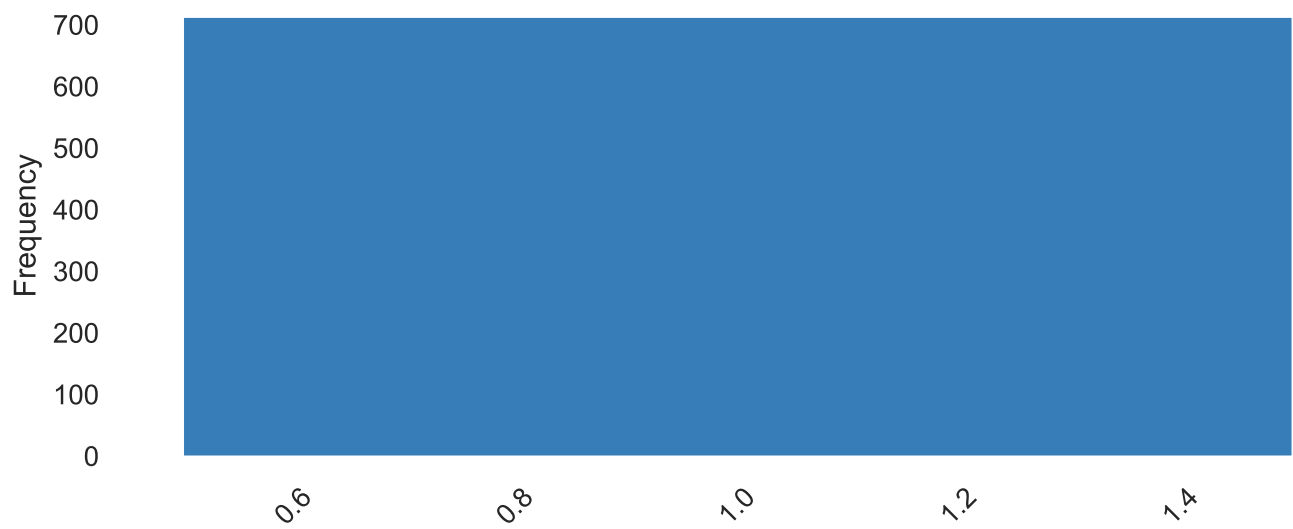
Sample

1st row	0
2nd row	1
3rd row	1
4th row	1
5th row	0

Common Values

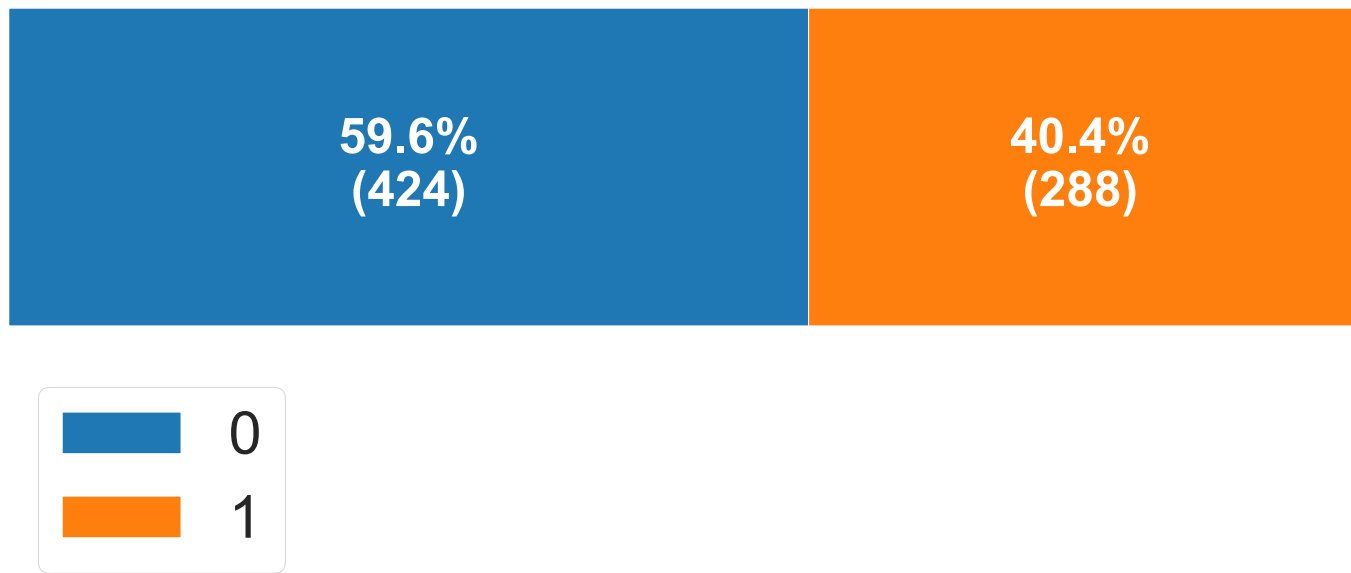
Value	Count	Frequency (%)
0	424	59.6%
1	288	40.4%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
0	424	59.6%
1	288	40.4%

Most occurring characters

Value	Count	Frequency (%)
0	424	59.6%
1	288	40.4%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	712	100.0%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	424	59.6%
1	288	40.4%

Most occurring scripts

Value	Count	Frequency (%)
Common	712	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
0	424	59.6%
1	288	40.4%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	712	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
0	424	59.6%
1	288	40.4%

pclass
Categorical

Distinct	3
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	62.1 KiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	712	
Distinct characters	3	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

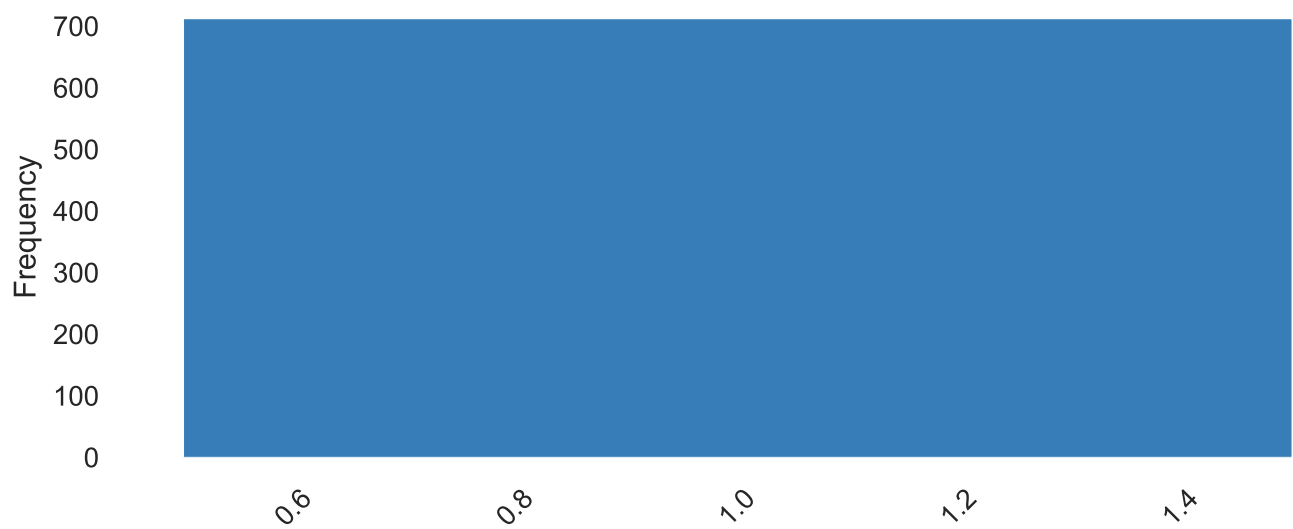
Sample

1st row	3
2nd row	1
3rd row	3
4th row	1
5th row	3

Common Values

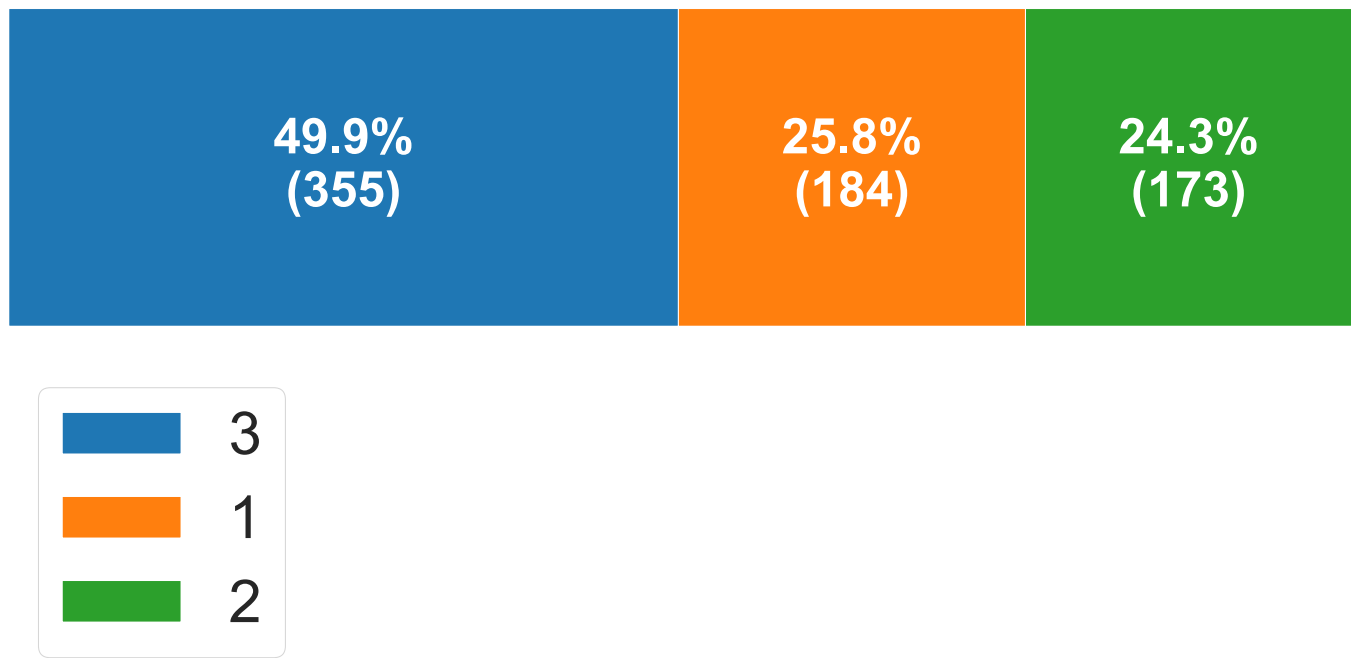
Value	Count	Frequency (%)
3	355	49.9%
1	184	25.8%
2	173	24.3%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
3	355	49.9%
1	184	25.8%
2	173	24.3%

Most occurring characters

Value	Count	Frequency (%)
3	355	49.9%
1	184	25.8%
2	173	24.3%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	712	100.0%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
3	355	49.9%
1	184	25.8%
2	173	24.3%

Most occurring scripts

Value	Count	Frequency (%)
Common	712	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
3	355	49.9%
1	184	25.8%
2	173	24.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	712	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
3	355	49.9%
1	184	25.8%
2	173	24.3%

sex
Categorical

Distinct	2
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	64.6 KiB

Length

Max length	6
Median length	4
Mean length	4.7275281
Min length	4

Characters and Unicode

Total characters	3366	
Distinct characters	5	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

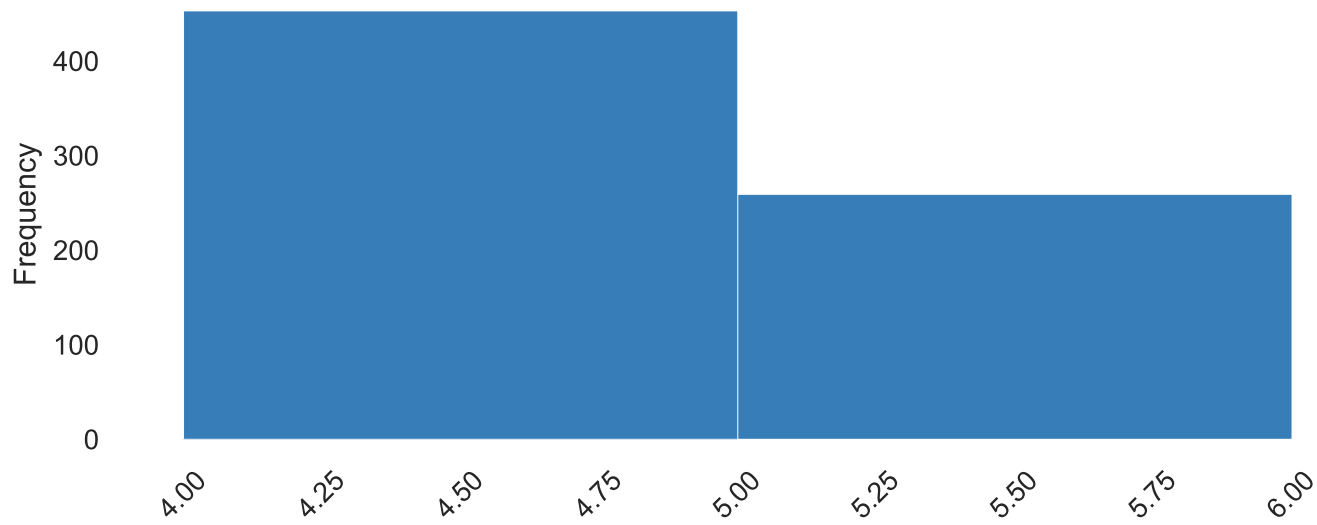
Sample

1st row	male
2nd row	female
3rd row	female
4th row	female
5th row	male

Common Values

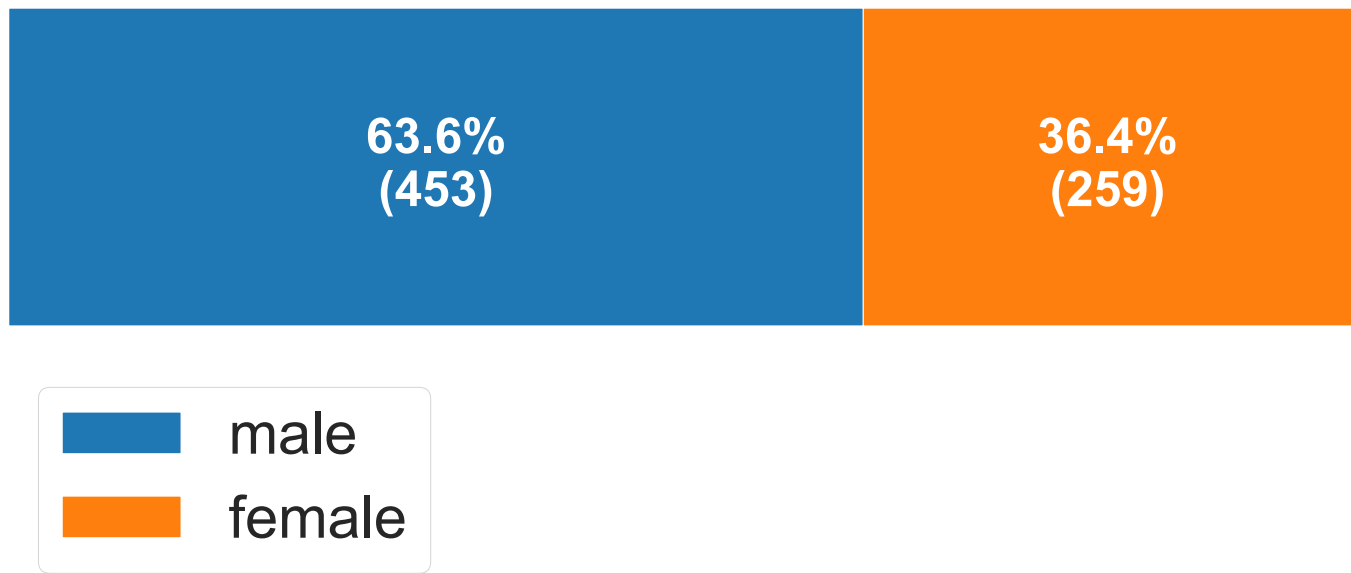
Value	Count	Frequency (%)
male	453	63.6%
female	259	36.4%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
male	453	63.6%
female	259	36.4%

Most occurring characters

Value	Count	Frequency (%)
e	971	28.8%
m	712	21.2%
a	712	21.2%
l	712	21.2%
f	259	7.7%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	3366	100.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	971	28.8%
m	712	21.2%
a	712	21.2%
l	712	21.2%
f	259	7.7%

Most occurring scripts

Value	Count	Frequency (%)
Latin	3366	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	971	28.8%
m	712	21.2%
a	712	21.2%
l	712	21.2%
f	259	7.7%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	3366	100.0%

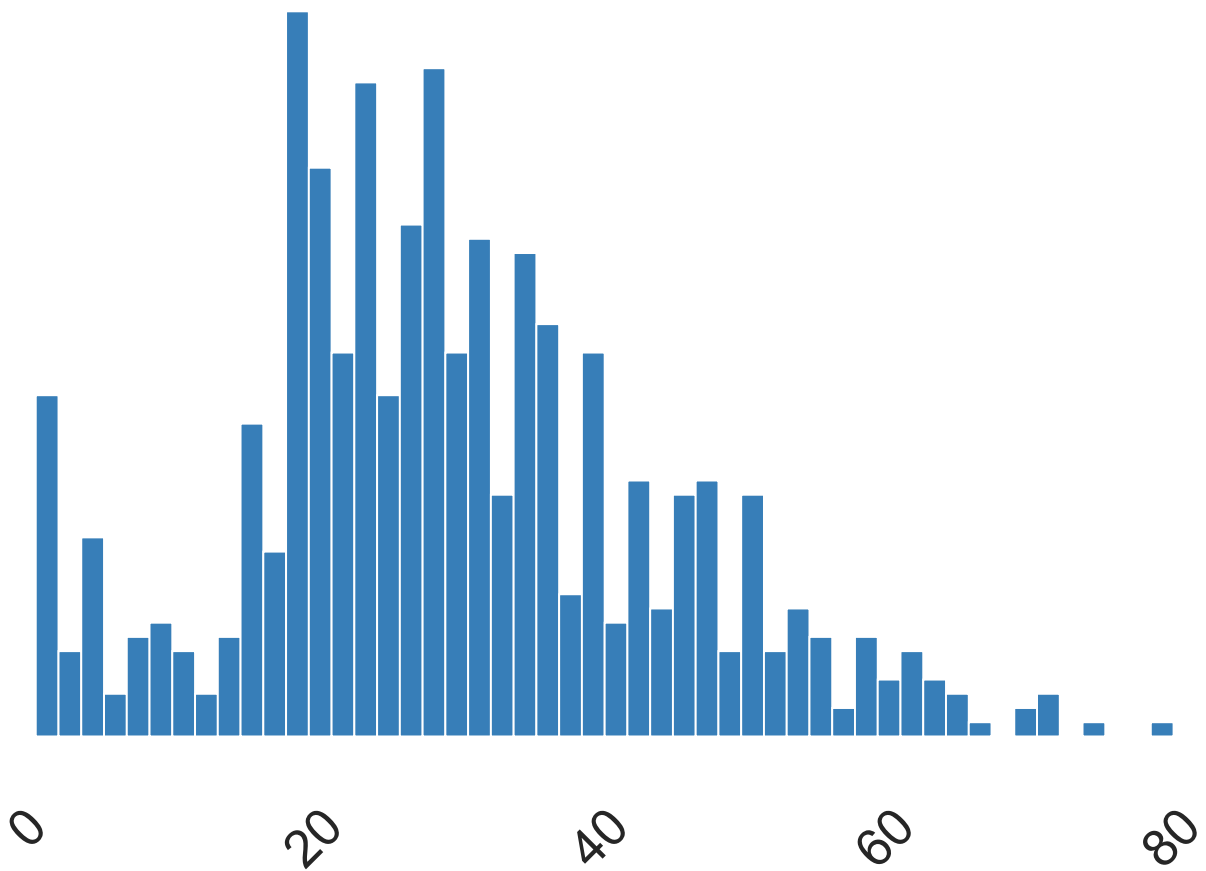
Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	971	28.8%
m	712	21.2%
a	712	21.2%
l	712	21.2%
f	259	7.7%

age
Real number (ℝ)

Distinct	88
Distinct (%)	12.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	29.642093
Minimum	0.42
Maximum	80
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	27.3 KiB

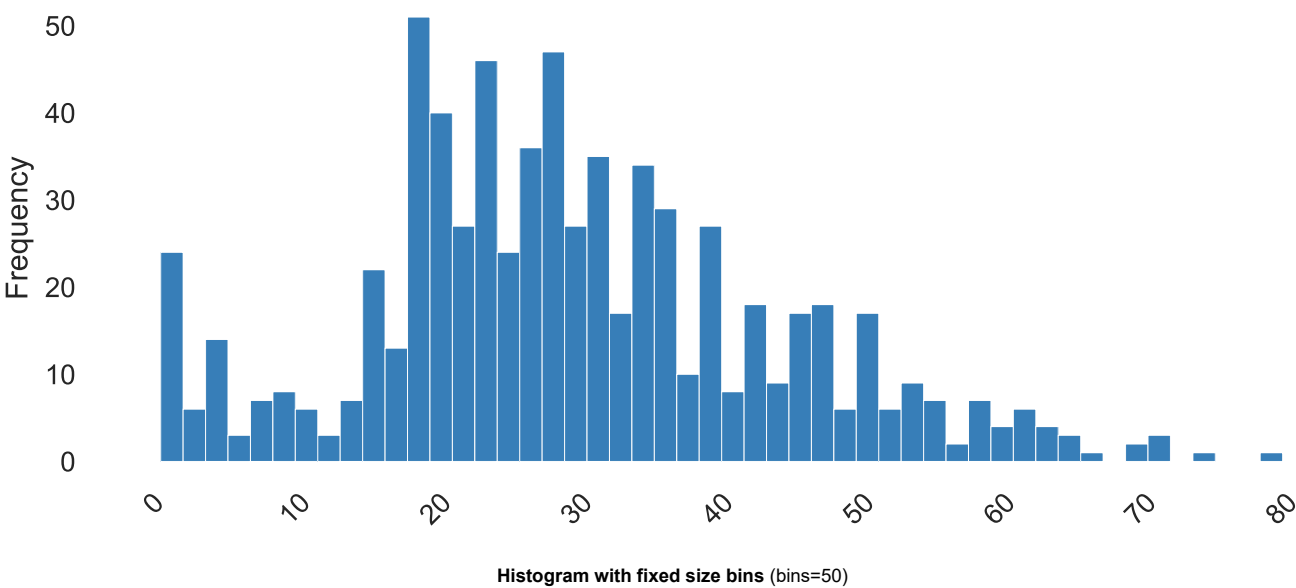


Quantile statistics

Minimum	0.42
5-th percentile	4
Q1	20
median	28
Q3	38
95-th percentile	56
Maximum	80
Range	79.58
Interquartile range (IQR)	18

Descriptive statistics

Standard deviation	14.492933
Coefficient of variation (CV)	0.48893083
Kurtosis	0.18794879
Mean	29.642093
Median Absolute Deviation (MAD)	8.75
Skewness	0.38888062
Sum	21105.17
Variance	210.0451
Monotonicity	Not monotonic



Value	Count	Frequency (%)
24	30	4.2%
22	27	3.8%
18	26	3.7%
19	25	3.5%
28	25	3.5%
30	25	3.5%
21	24	3.4%
25	23	3.2%
36	22	3.1%
29	20	2.8%
Other values (78)	465	65.3%

Value	Count	Frequency (%)
0.42	1	0.1%
0.67	1	0.1%
0.75	2	0.3%
0.83	2	0.3%
0.92	1	0.1%
1	7	1.0%
2	10	1.4%
3	6	0.8%
4	10	1.4%
5	4	0.6%

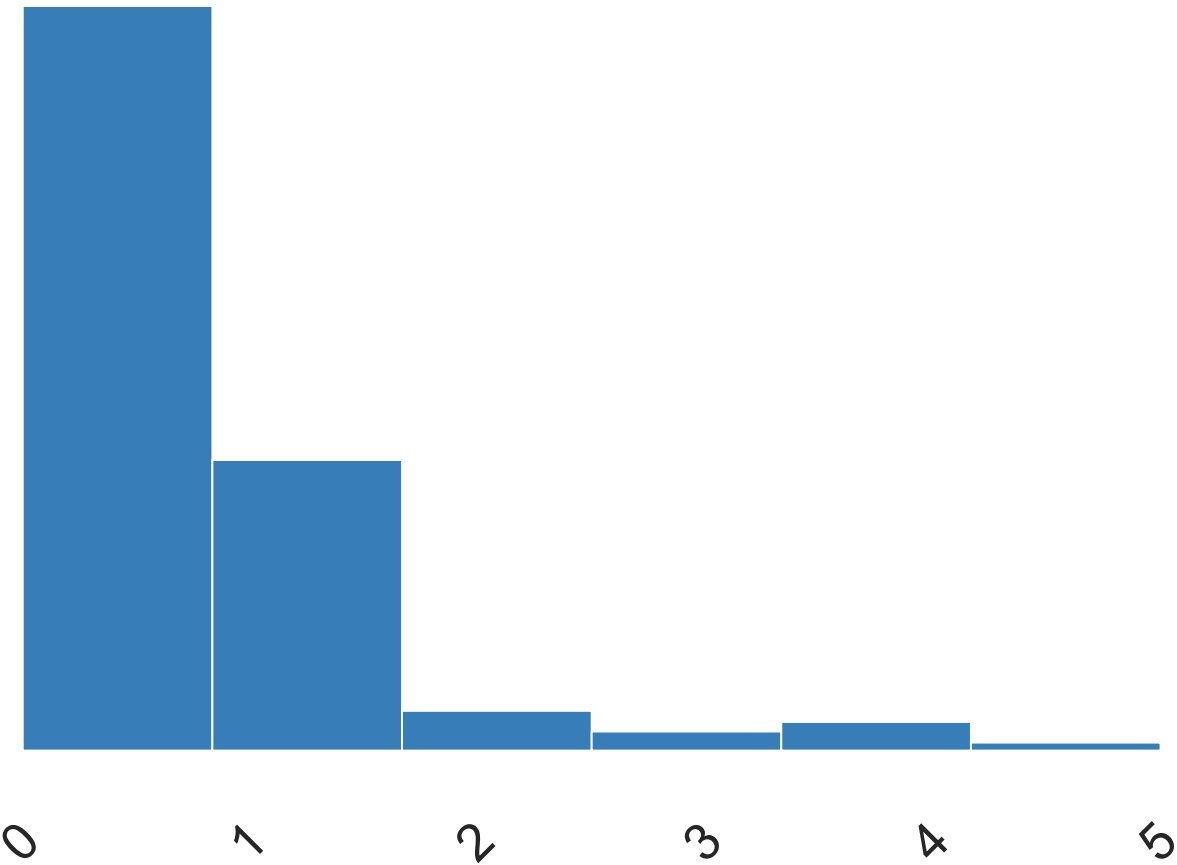
Value	Count	Frequency (%)
80	1	0.1%
74	1	0.1%
71	2	0.3%
70.5	1	0.1%
70	2	0.3%
66	1	0.1%
65	3	0.4%
64	2	0.3%
63	2	0.3%
62	3	0.4%

sibsp

Real number (ℝ)

ZEROS

Distinct	6
Distinct (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	0.51404494
Minimum	0
Maximum	5
Zeros	469
Zeros (%)	65.9%
Negative	0
Negative (%)	0.0%
Memory size	27.3 KiB

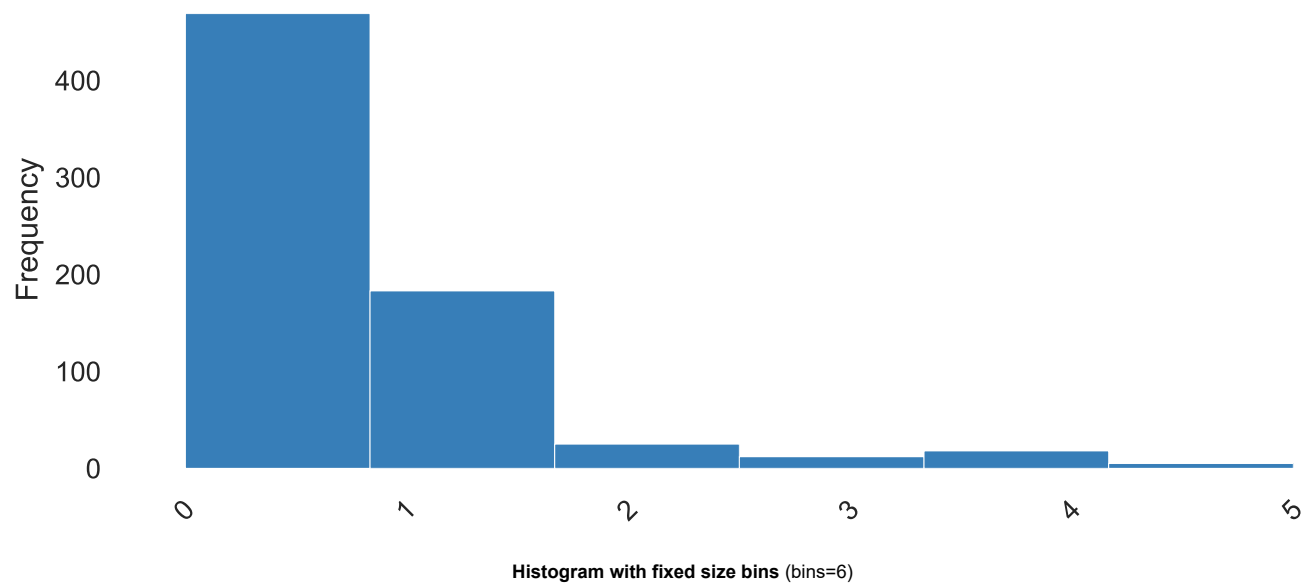


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	1
95-th percentile	2
Maximum	5
Range	5
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	0.93069213
Coefficient of variation (CV)	1.8105268
Kurtosis	7.0181847
Mean	0.51404494
Median Absolute Deviation (MAD)	0
Skewness	2.5151075
Sum	366
Variance	0.86618783
Monotonicity	Not monotonic



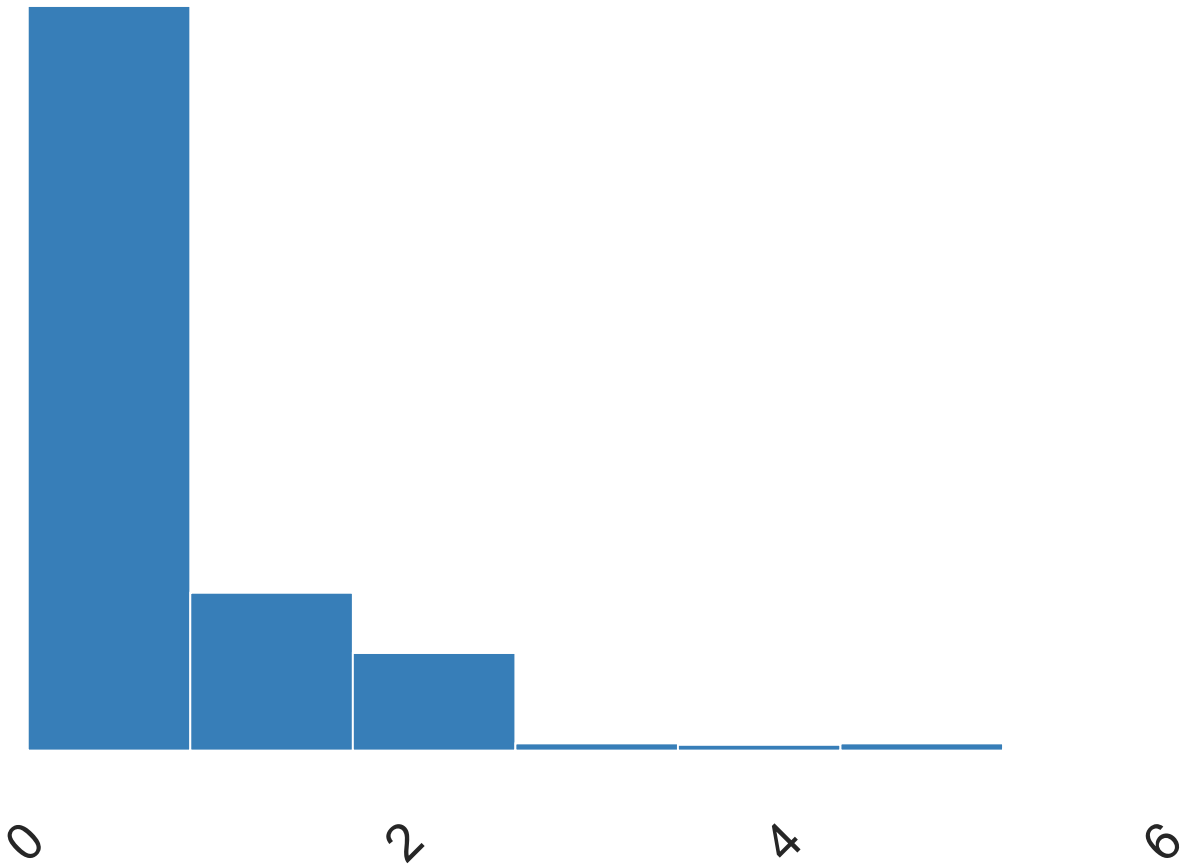
Value	Count	Frequency (%)
0	469	65.9%
1	183	25.7%
2	25	3.5%
4	18	2.5%
3	12	1.7%
5	5	0.7%

Value	Count	Frequency (%)
0	469	65.9%
1	183	25.7%
2	25	3.5%
3	12	1.7%
4	18	2.5%
5	5	0.7%

Value	Count	Frequency (%)
5	5	0.7%
4	18	2.5%
3	12	1.7%
2	25	3.5%
1	183	25.7%
0	469	65.9%

parch
Real number (ℝ)

ZEROS		
Distinct	7	
Distinct (%)	1.0%	
Missing	0	
Missing (%)	0.0%	
Infinite	0	
Infinite (%)	0.0%	
Mean	0.43258427	
Minimum	0	
Maximum	6	
Zeros	519	
Zeros (%)	72.9%	
Negative	0	
Negative (%)	0.0%	
Memory size	27.3 KiB	

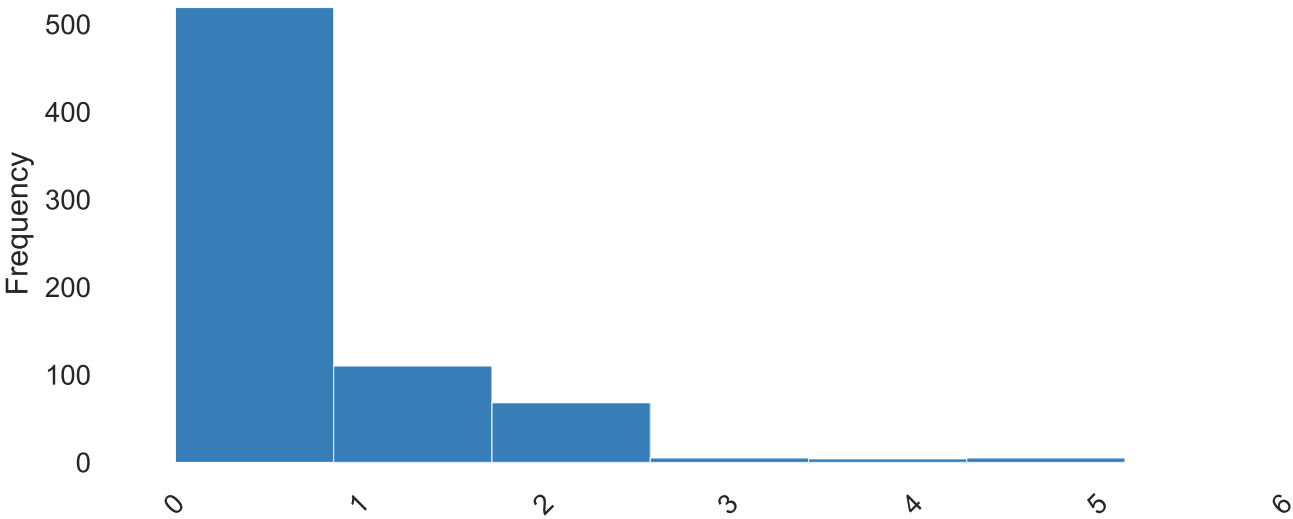


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	1
95-th percentile	2
Maximum	6
Range	6
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	0.85418145
Coefficient of variation (CV)	1.9746013
Kurtosis	8.8220034
Mean	0.43258427
Median Absolute Deviation (MAD)	0
Skewness	2.6141773
Sum	308
Variance	0.72962594
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=7)

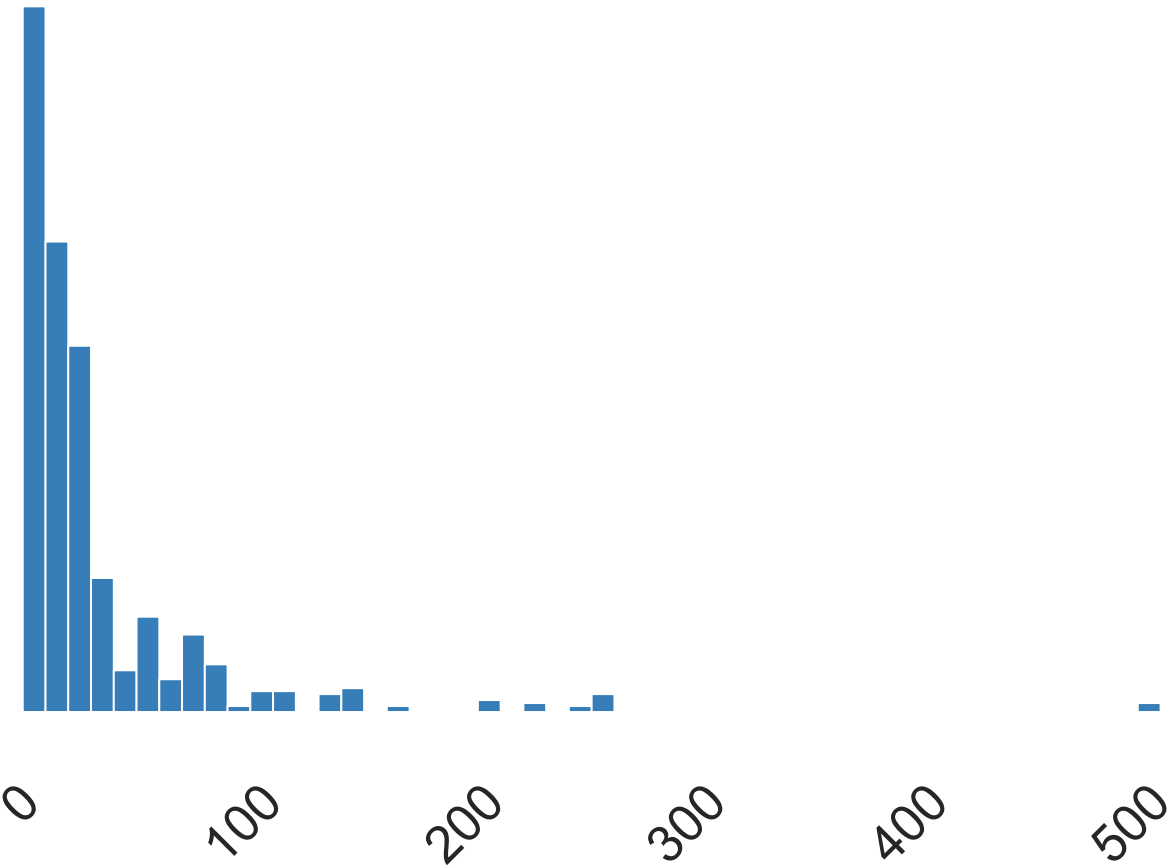
Value	Count	Frequency (%)
0	519	72.9%
1	110	15.4%
2	68	9.6%
5	5	0.7%
3	5	0.7%
4	4	0.6%
6	1	0.1%

Value	Count	Frequency (%)
0	519	72.9%
1	110	15.4%
2	68	9.6%
3	5	0.7%
4	4	0.6%
5	5	0.7%
6	1	0.1%

Value	Count	Frequency (%)
6	1	0.1%
5	5	0.7%
4	4	0.6%
3	5	0.7%
2	68	9.6%
1	110	15.4%
0	519	72.9%

fare
Real number (R)

Distinct	219
Distinct (%)	30.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	34.567251
Minimum	0
Maximum	512.3292
Zeros	7
Zeros (%)	1.0%
Negative	0
Negative (%)	0.0%
Memory size	27.3 KiB

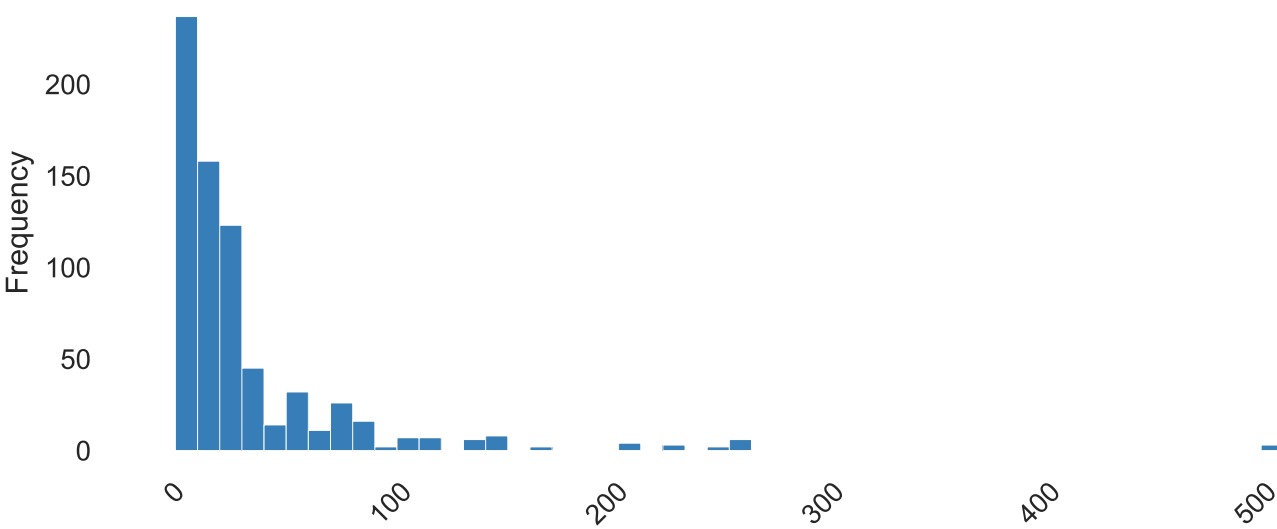


Quantile statistics

Minimum	0
5-th percentile	7.225
Q1	8.05
median	15.64585
Q3	33
95-th percentile	120
Maximum	512.3292
Range	512.3292
Interquartile range (IQR)	24.95

Descriptive statistics

Standard deviation	52.938648
Coefficient of variation (CV)	1.5314683
Kurtosis	31.013377
Mean	34.567251
Median Absolute Deviation (MAD)	8.11045
Skewness	4.6670088
Sum	24611.883
Variance	2802.5005
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
13	41	5.8%
26	30	4.2%
8.05	29	4.1%
10.5	24	3.4%
7.8958	23	3.2%
7.925	18	2.5%
7.75	14	2.0%
7.775	14	2.0%
26.55	13	1.8%
7.8542	13	1.8%
Other values (209)	493	69.2%

Value	Count	Frequency (%)
0	7	1.0%
4.0125	1	0.1%
5	1	0.1%
6.2375	1	0.1%
6.4375	1	0.1%
6.45	1	0.1%
6.4958	2	0.3%
6.75	2	0.3%
6.975	2	0.3%
7.0458	1	0.1%

Value	Count	Frequency (%)
512.3292	3	0.4%
263	4	0.6%
262.375	2	0.3%
247.5208	2	0.3%
227.525	3	0.4%
211.5	1	0.1%
211.3375	3	0.4%
164.8667	2	0.3%
153.4625	3	0.4%
151.55	4	0.6%

embarked
Categorical

Distinct	3
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	62.1 KiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	712	
Distinct characters	3	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

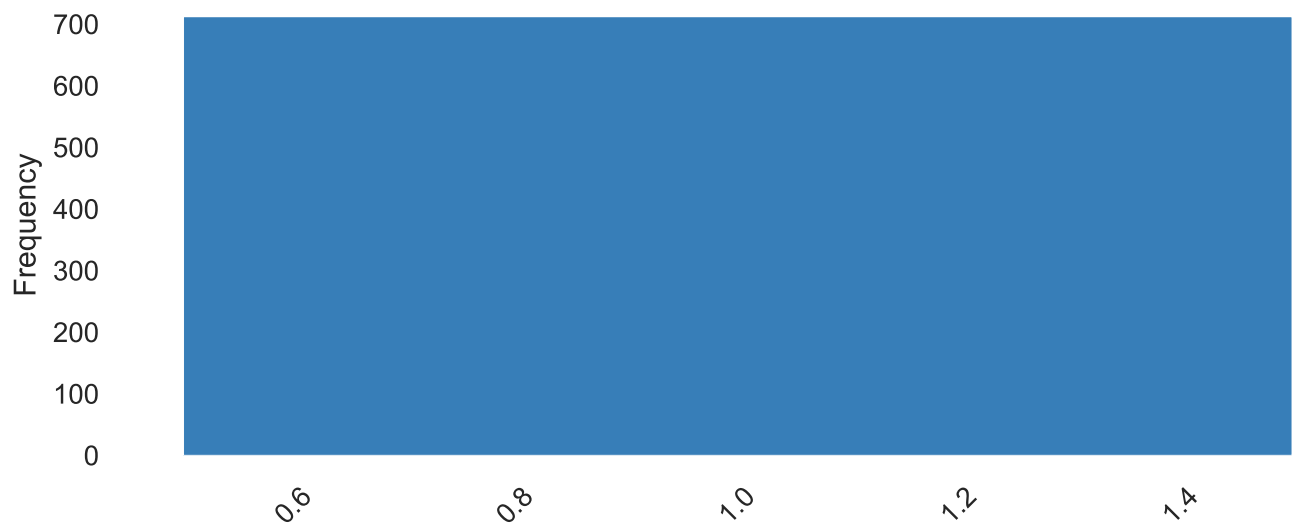
Sample

1st row	S
2nd row	C
3rd row	S
4th row	S
5th row	S

Common Values

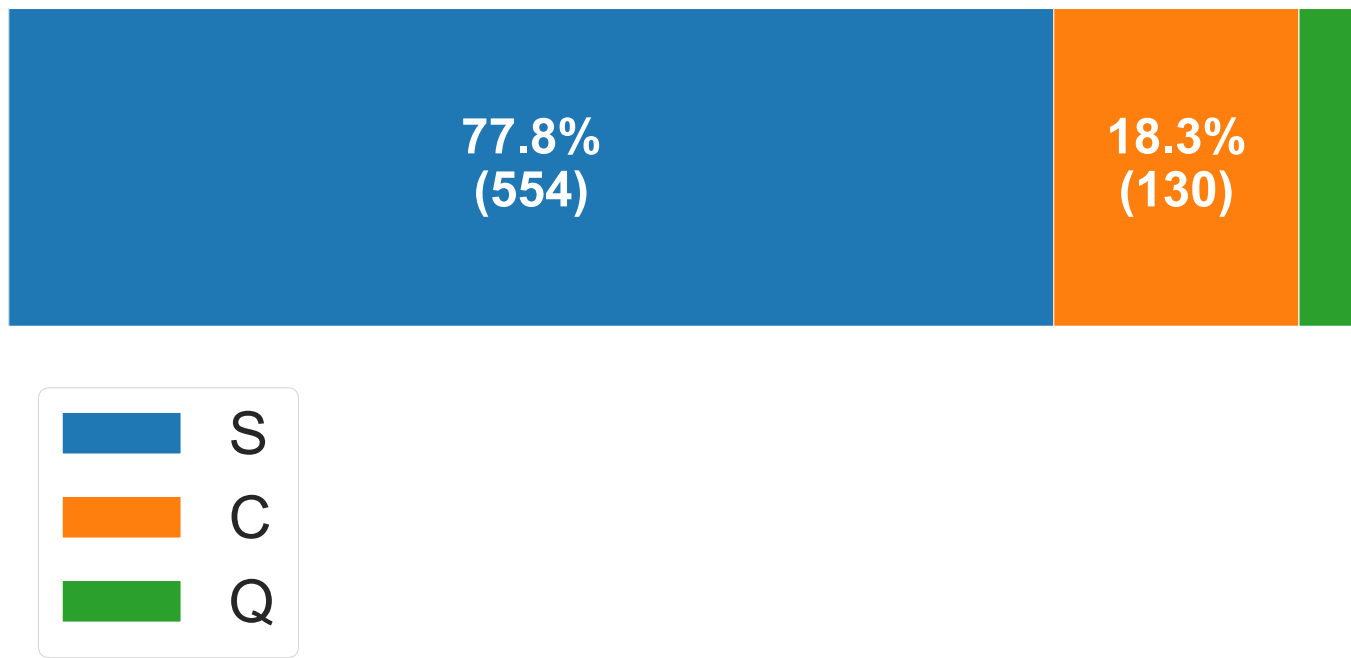
Value	Count	Frequency (%)
S	554	77.8%
C	130	18.3%
Q	28	3.9%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
s	554	77.8%
c	130	18.3%
q	28	3.9%

Most occurring characters

Value	Count	Frequency (%)
S	554	77.8%
C	130	18.3%
Q	28	3.9%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	712	100.0%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
S	554	77.8%
C	130	18.3%
Q	28	3.9%

Most occurring scripts

Value	Count	Frequency (%)
Latin	712	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
S	554	77.8%
C	130	18.3%
Q	28	3.9%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	712	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
S	554	77.8%
C	130	18.3%
Q	28	3.9%

class
Categorical

Distinct	3
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	22.7 KiB

Length

Max length	6
Median length	5
Mean length	5.2429775
Min length	5

Characters and Unicode

Total characters	3733	
Distinct characters	13	
Distinct categories	2 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

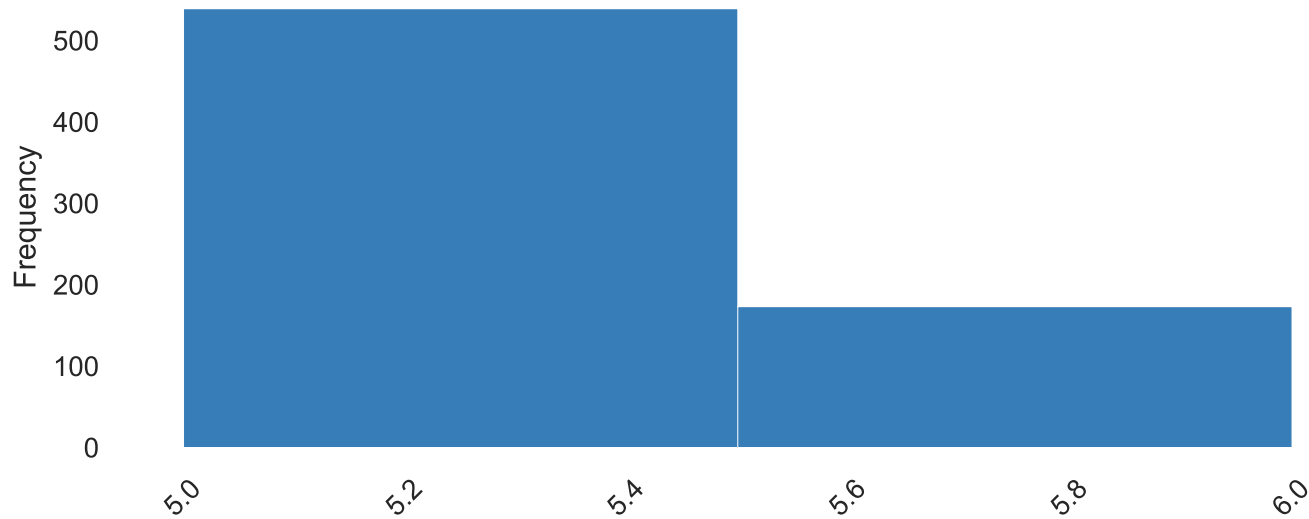
Sample

1st row	Third
2nd row	First
3rd row	Third
4th row	First
5th row	Third

Common Values

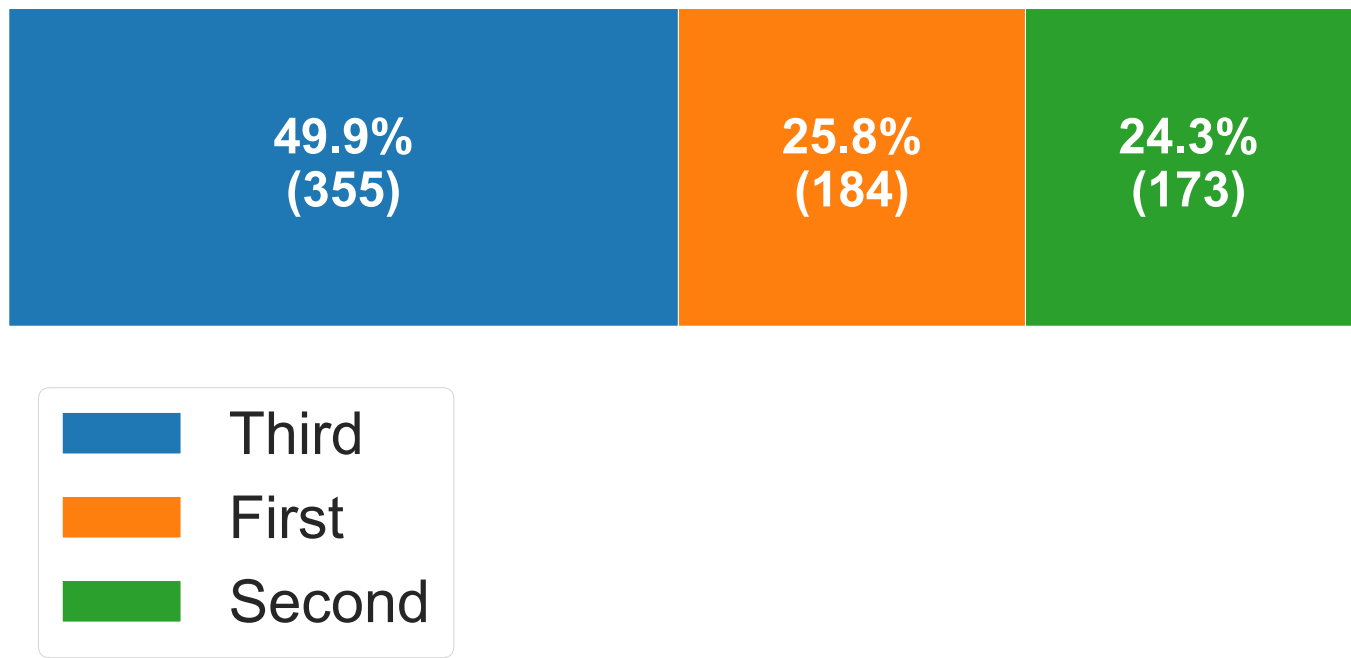
Value	Count	Frequency (%)
Third	355	49.9%
First	184	25.8%
Second	173	24.3%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
third	355	49.9%
first	184	25.8%
second	173	24.3%

Most occurring characters

Value	Count	Frequency (%)
i	539	14.4%
r	539	14.4%
d	528	14.1%
T	355	9.5%
h	355	9.5%
F	184	4.9%
s	184	4.9%
t	184	4.9%
S	173	4.6%
e	173	4.6%
Other values (3)	519	13.9%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	3021	80.9%
Uppercase Letter	712	19.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
i	539	17.8%
r	539	17.8%
d	528	17.5%
h	355	11.8%
s	184	6.1%
t	184	6.1%
e	173	5.7%
c	173	5.7%
o	173	5.7%
n	173	5.7%

Uppercase Letter

Value	Count	Frequency (%)
T	355	49.9%
F	184	25.8%
S	173	24.3%

Most occurring scripts

Value	Count	Frequency (%)
Latin	3733	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
i	539	14.4%
r	539	14.4%
d	528	14.1%
T	355	9.5%
h	355	9.5%
F	184	4.9%
s	184	4.9%
t	184	4.9%
S	173	4.6%
e	173	4.6%
Other values (3)	519	13.9%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	3733	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
i	539	14.4%
r	539	14.4%
d	528	14.1%
T	355	9.5%
h	355	9.5%
F	184	4.9%
s	184	4.9%
t	184	4.9%
S	173	4.6%
e	173	4.6%
Other values (3)	519	13.9%

who
Categorical

Distinct	3
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	64.0 KiB

Length

Max length	5
Median length	3
Mean length	3.8398876
Min length	3

Characters and Unicode

Total characters	2734	
Distinct characters	10	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

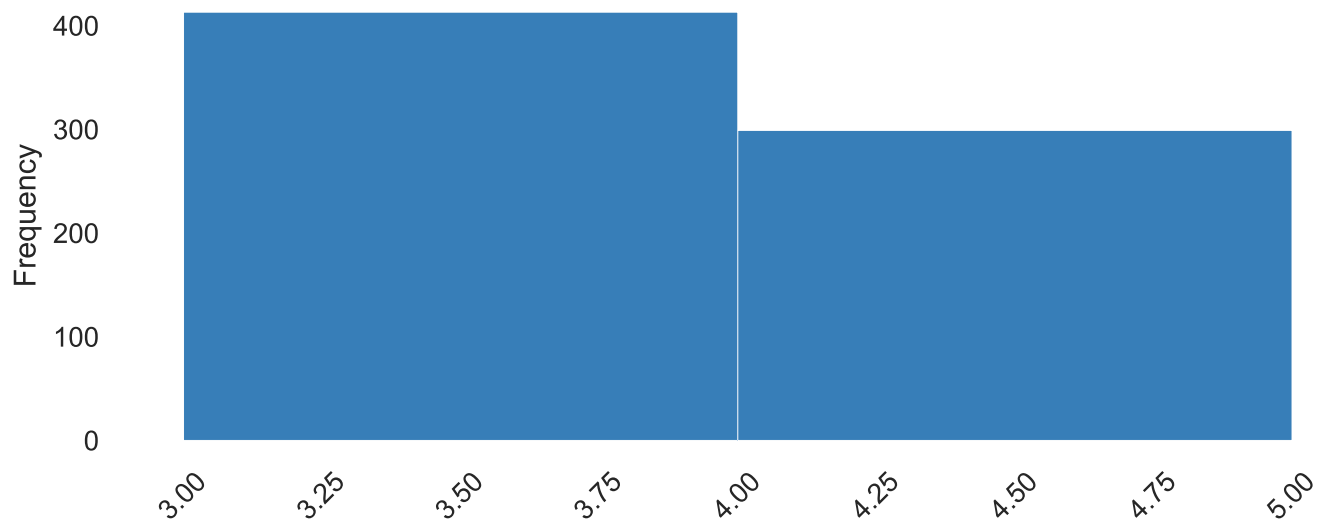
Sample

1st row	man
2nd row	woman
3rd row	woman
4th row	woman
5th row	man

Common Values

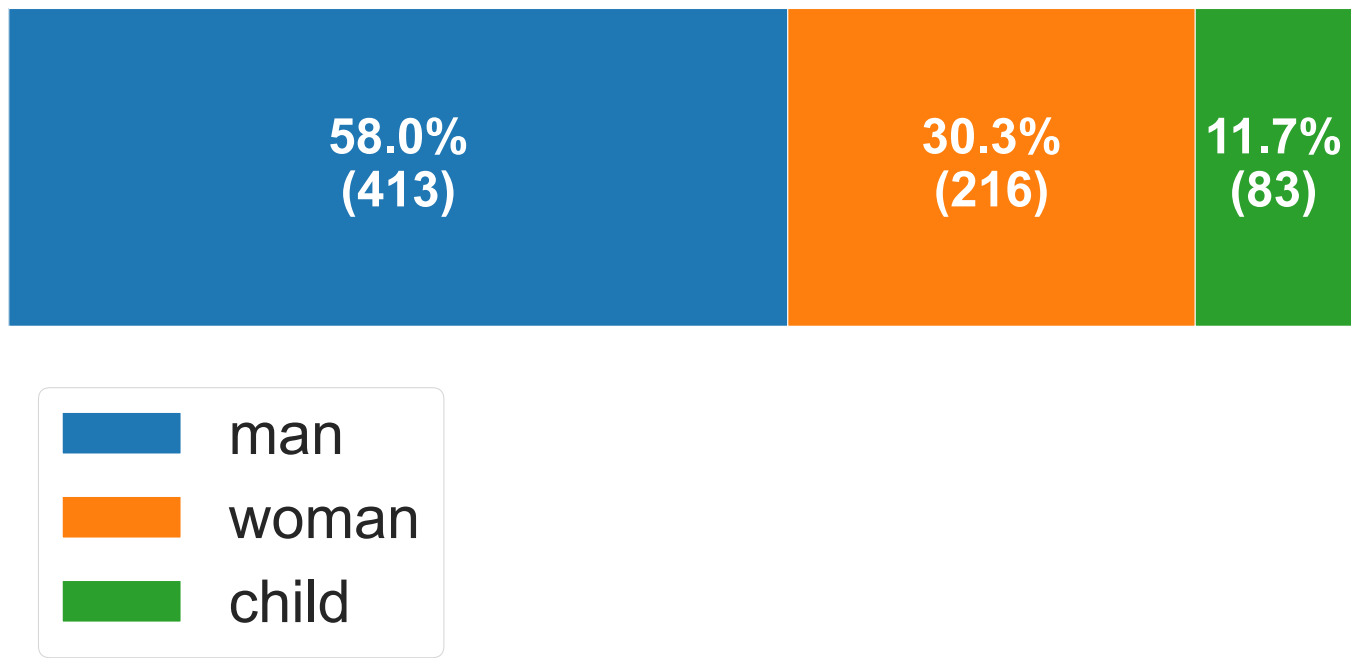
Value	Count	Frequency (%)
man	413	58.0%
woman	216	30.3%
child	83	11.7%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
man	413	58.0%
woman	216	30.3%
child	83	11.7%

Most occurring characters

Value	Count	Frequency (%)
m	629	23.0%
a	629	23.0%
n	629	23.0%
w	216	7.9%
o	216	7.9%
c	83	3.0%
h	83	3.0%
i	83	3.0%
l	83	3.0%
d	83	3.0%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	2734	100.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
m	629	23.0%
a	629	23.0%
n	629	23.0%
w	216	7.9%
o	216	7.9%
c	83	3.0%
h	83	3.0%
i	83	3.0%
l	83	3.0%
d	83	3.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	2734	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
m	629	23.0%
a	629	23.0%
n	629	23.0%
w	216	7.9%
o	216	7.9%
c	83	3.0%
h	83	3.0%
i	83	3.0%
l	83	3.0%
d	83	3.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	2734	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
m	629	23.0%
a	629	23.0%
n	629	23.0%
w	216	7.9%
o	216	7.9%
c	83	3.0%
h	83	3.0%
i	83	3.0%
l	83	3.0%
d	83	3.0%

adult_male
Boolean

Distinct	2
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	22.4 KiB

Value	Count	Frequency (%)
True	413	58.0%
False	299	42.0%



deck
Categorical

IMBALANCE

Distinct	8
Distinct (%)	1.1%
Missing	0
Missing (%)	0.0%
Memory size	23.2 KiB

Length

Max length	7
Median length	7
Mean length	5.4662921
Min length	1

Characters and Unicode

Total characters	3892	
Distinct characters	12	
Distinct categories	2 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

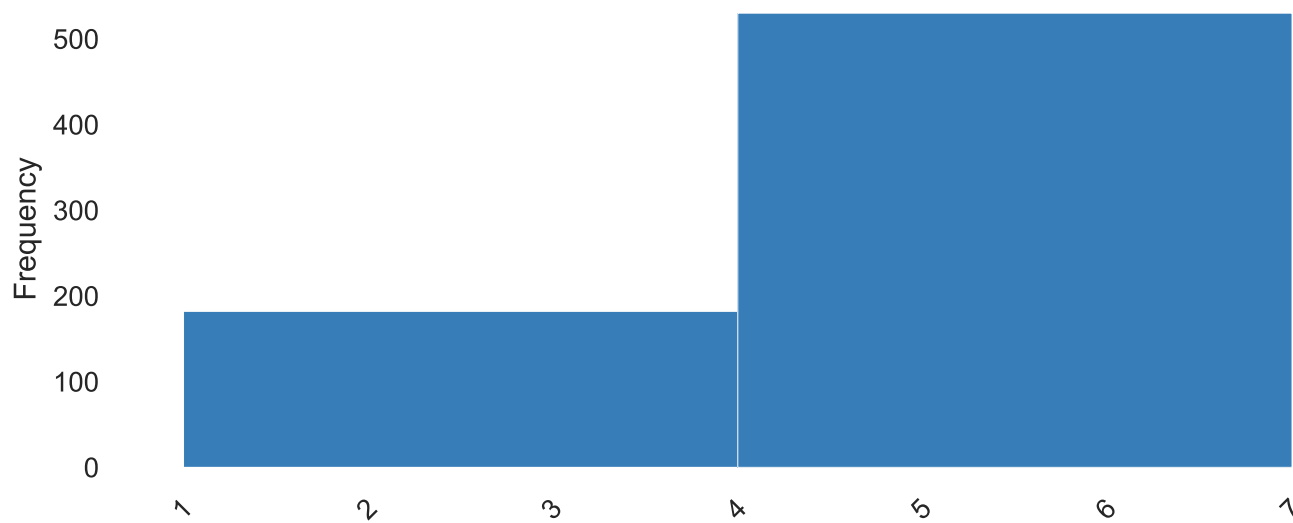
Sample

1st row	Unknown
2nd row	C
3rd row	Unknown
4th row	C
5th row	Unknown

Common Values

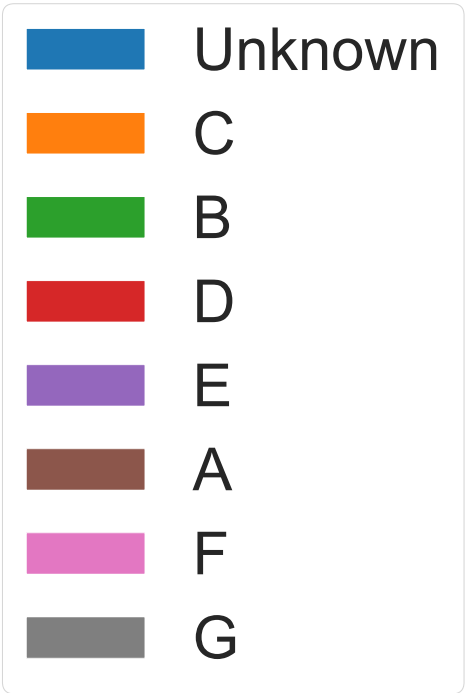
Value	Count	Frequency (%)
Unknown	530	74.4%
C	51	7.2%
B	43	6.0%
D	31	4.4%
E	30	4.2%
A	12	1.7%
F	11	1.5%
G	4	0.6%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
unknown	530	74.4%
c	51	7.2%
b	43	6.0%
d	31	4.4%
e	30	4.2%
a	12	1.7%
f	11	1.5%
g	4	0.6%

Most occurring characters

Value	Count	Frequency (%)
n	1590	40.9%
U	530	13.6%
k	530	13.6%
o	530	13.6%
w	530	13.6%
C	51	1.3%
B	43	1.1%
D	31	0.8%
E	30	0.8%
A	12	0.3%
Other values (2)	15	0.4%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	3180	81.7%
Uppercase Letter	712	18.3%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
U	530	74.4%
C	51	7.2%
B	43	6.0%
D	31	4.4%
E	30	4.2%
A	12	1.7%
F	11	1.5%
G	4	0.6%

Lowercase Letter

Value	Count	Frequency (%)
n	1590	50.0%
k	530	16.7%
o	530	16.7%
w	530	16.7%

Most occurring scripts

Value	Count	Frequency (%)
Latin	3892	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
n	1590	40.9%
U	530	13.6%
k	530	13.6%
o	530	13.6%
w	530	13.6%
C	51	1.3%
B	43	1.1%
D	31	0.8%
E	30	0.8%
A	12	0.3%
Other values (2)	15	0.4%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	3892	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
n	1590	40.9%
U	530	13.6%
k	530	13.6%
o	530	13.6%
w	530	13.6%
C	51	1.3%
B	43	1.1%
D	31	0.8%
E	30	0.8%
A	12	0.3%
Other values (2)	15	0.4%

embark_town
Categorical

Distinct	3
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	68.7 KiB

Length

Max length	11
Median length	11
Mean length	10.595506
Min length	9

Characters and Unicode

Total characters	7544	
Distinct characters	17	
Distinct categories	2 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

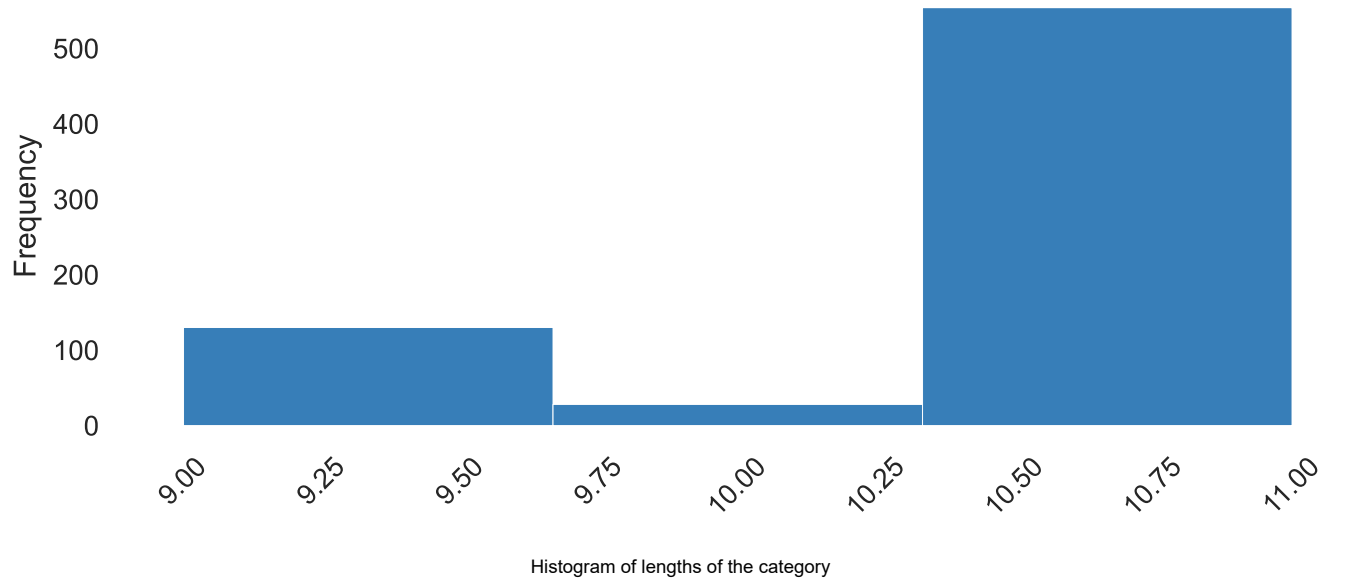
Sample

1st row	Southampton
2nd row	Cherbourg
3rd row	Southampton
4th row	Southampton
5th row	Southampton

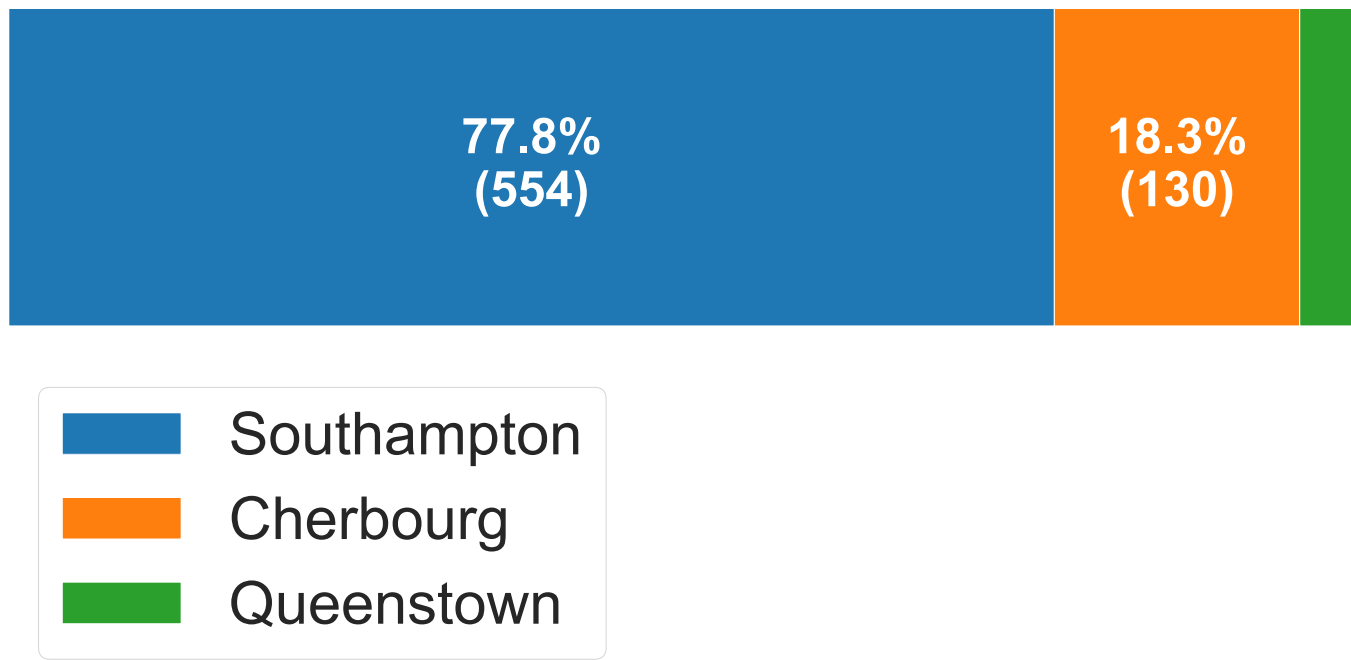
Common Values

Value	Count	Frequency (%)
Southampton	554	77.8%
Cherbourg	130	18.3%
Queenstown	28	3.9%

Length



Common Values (Plot)



Value	Count	Frequency (%)
southampton	554	77.8%
cherbourg	130	18.3%
queenstown	28	3.9%

Most occurring characters

Value	Count	Frequency (%)
o	1266	16.8%
t	1136	15.1%
u	712	9.4%
h	684	9.1%
n	610	8.1%
p	554	7.3%
S	554	7.3%
m	554	7.3%
a	554	7.3%
r	260	3.4%
Other values (7)	660	8.7%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	6832	90.6%
Uppercase Letter	712	9.4%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
o	1266	18.5%
t	1136	16.6%
u	712	10.4%
h	684	10.0%
n	610	8.9%
p	554	8.1%
m	554	8.1%
a	554	8.1%
r	260	3.8%
e	186	2.7%
Other values (4)	316	4.6%

Uppercase Letter

Value	Count	Frequency (%)
S	554	77.8%
C	130	18.3%
Q	28	3.9%

Most occurring scripts

Value	Count	Frequency (%)
Latin	7544	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
o	1266	16.8%
t	1136	15.1%
u	712	9.4%
h	684	9.1%
n	610	8.1%
p	554	7.3%
S	554	7.3%
m	554	7.3%
a	554	7.3%
r	260	3.4%
Other values (7)	660	8.7%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	7544	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
o	1266	16.8%
t	1136	15.1%
u	712	9.4%
h	684	9.1%
n	610	8.1%
p	554	7.3%
S	554	7.3%
m	554	7.3%
a	554	7.3%
r	260	3.4%
Other values (7)	660	8.7%

alive
Boolean

Distinct	2
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	22.4 KiB

Value	Count	Frequency (%)
False	424	59.6%
True	288	40.4%



alone

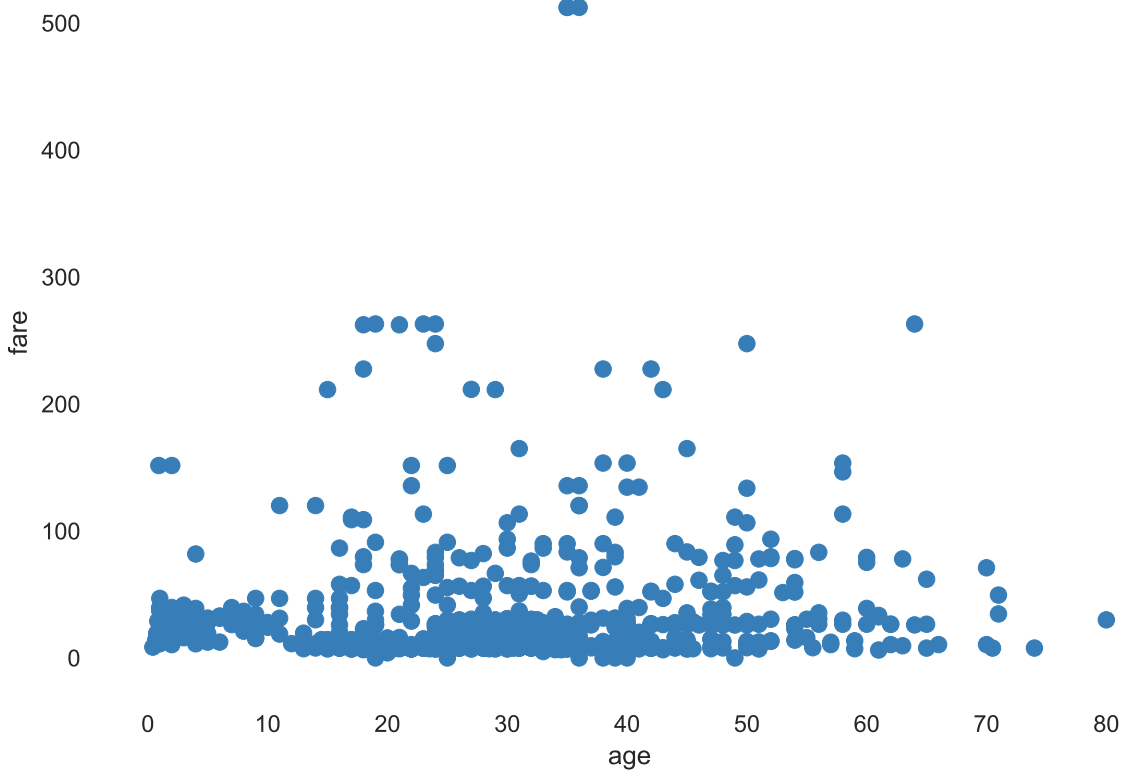
Boolean

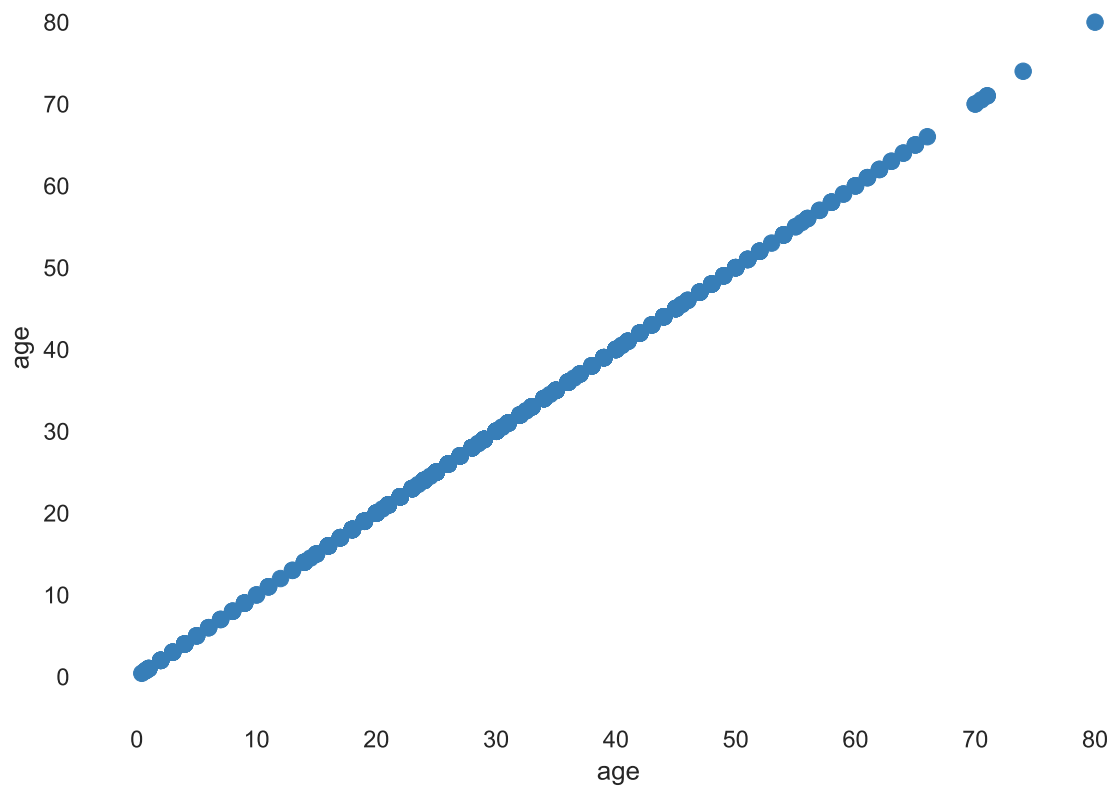
Distinct	2
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	22.4 KiB

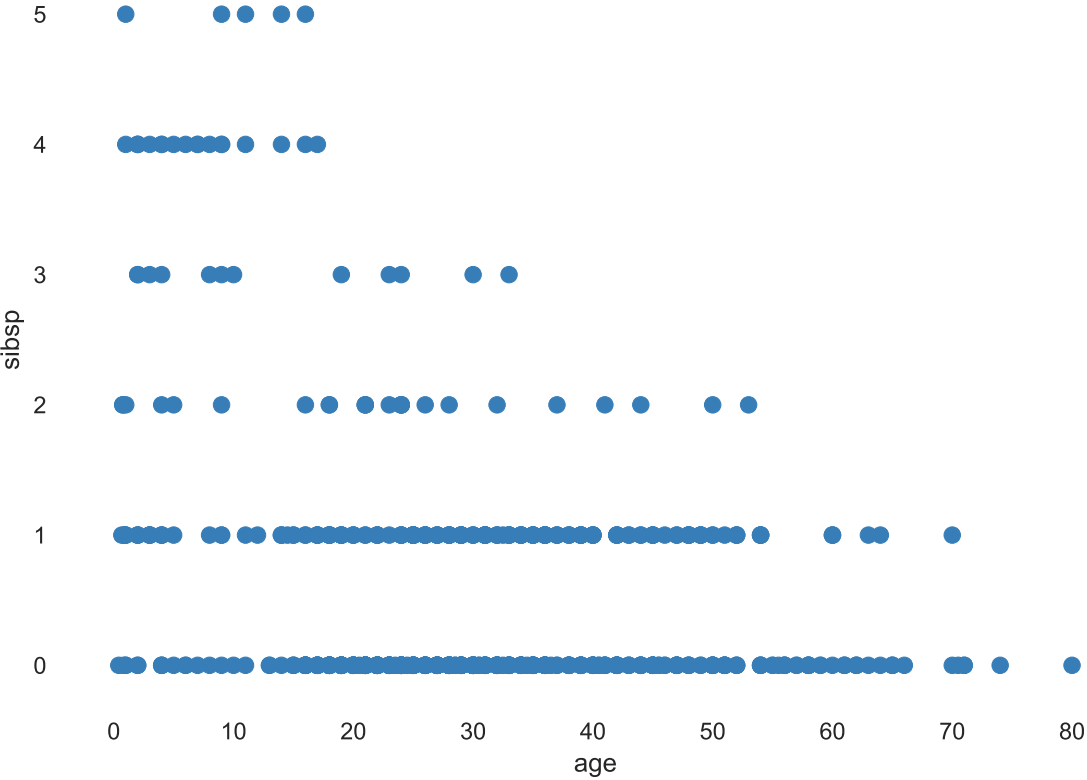
Value	Count	Frequency (%)
True	402	56.5%
False	310	43.5%

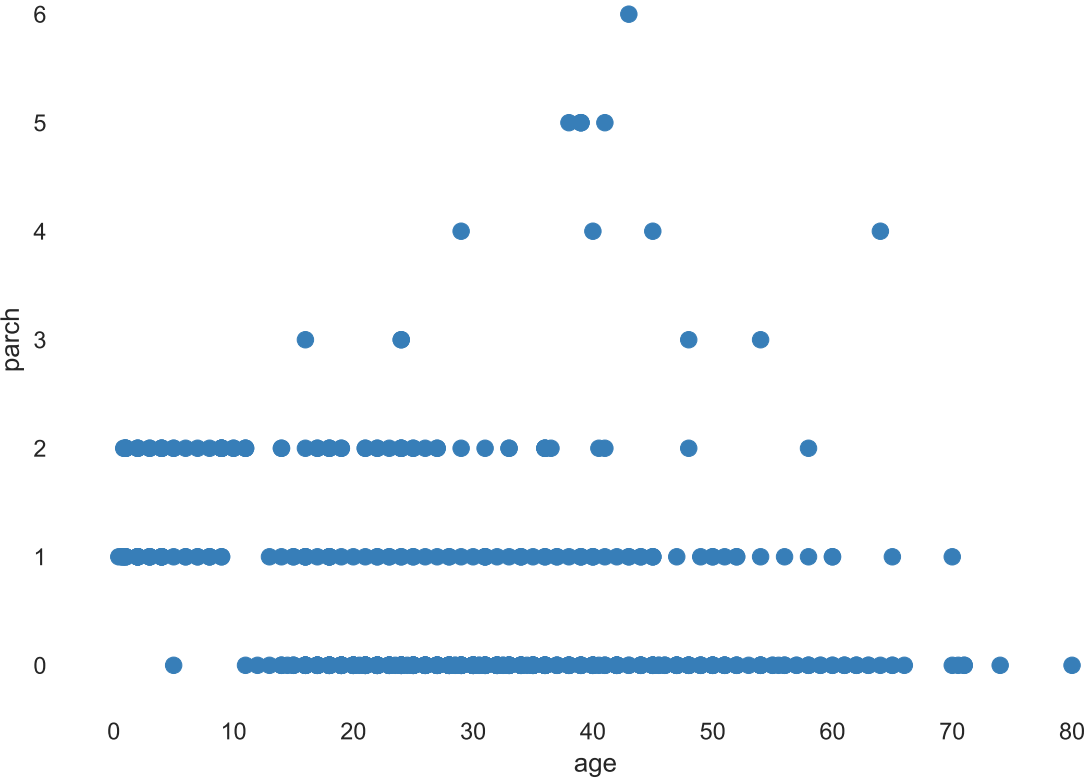


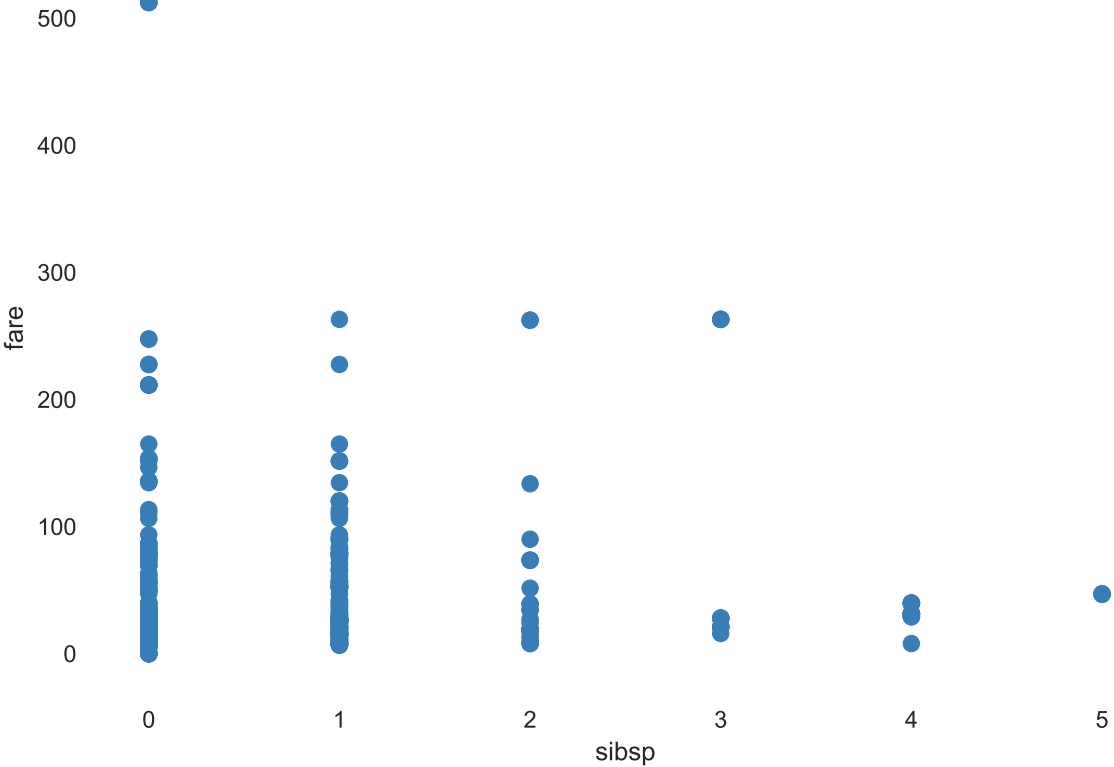
Interactions

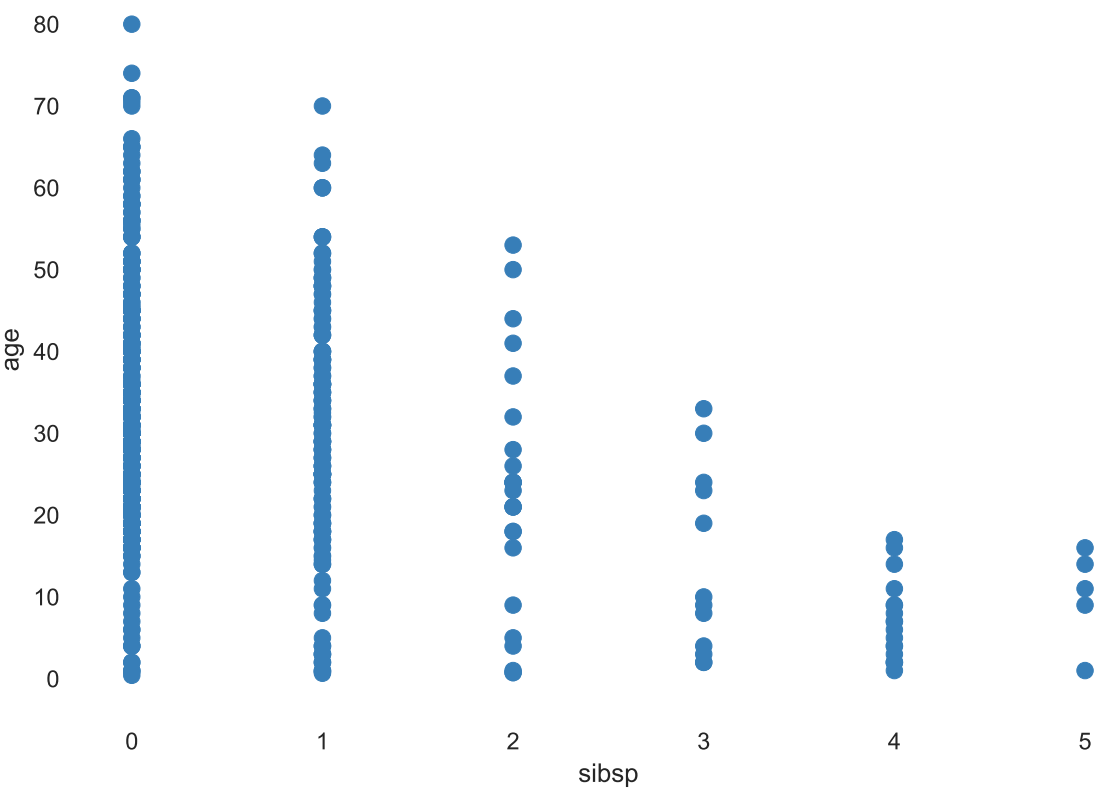


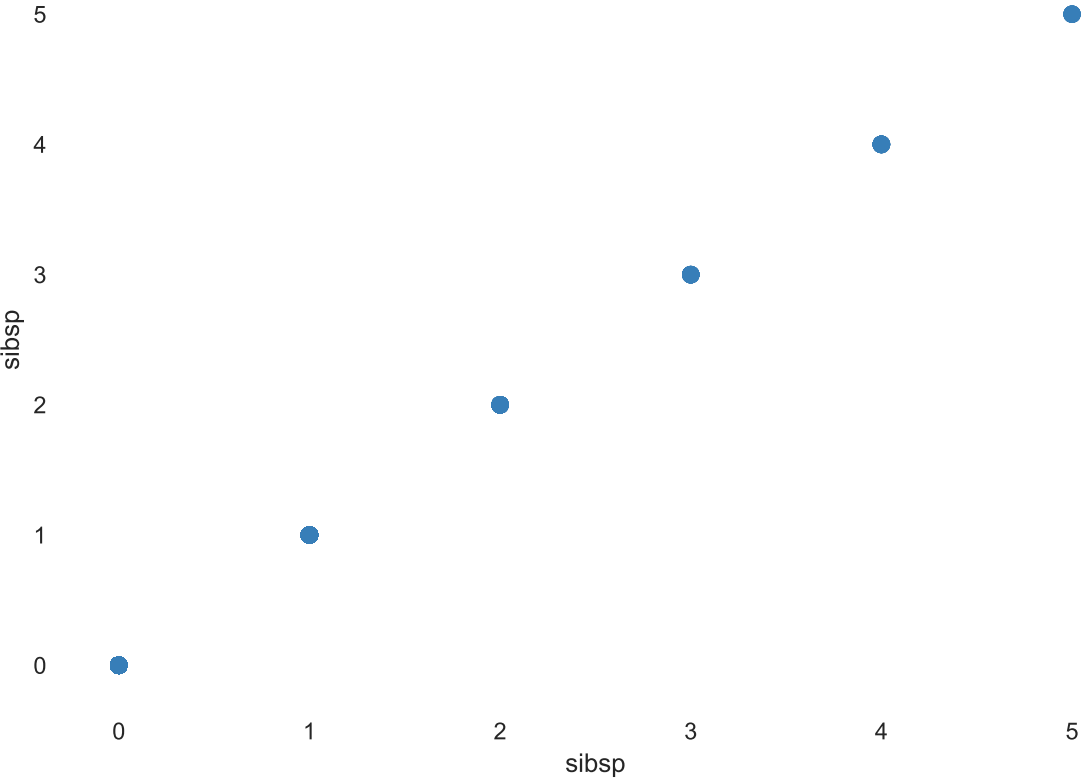


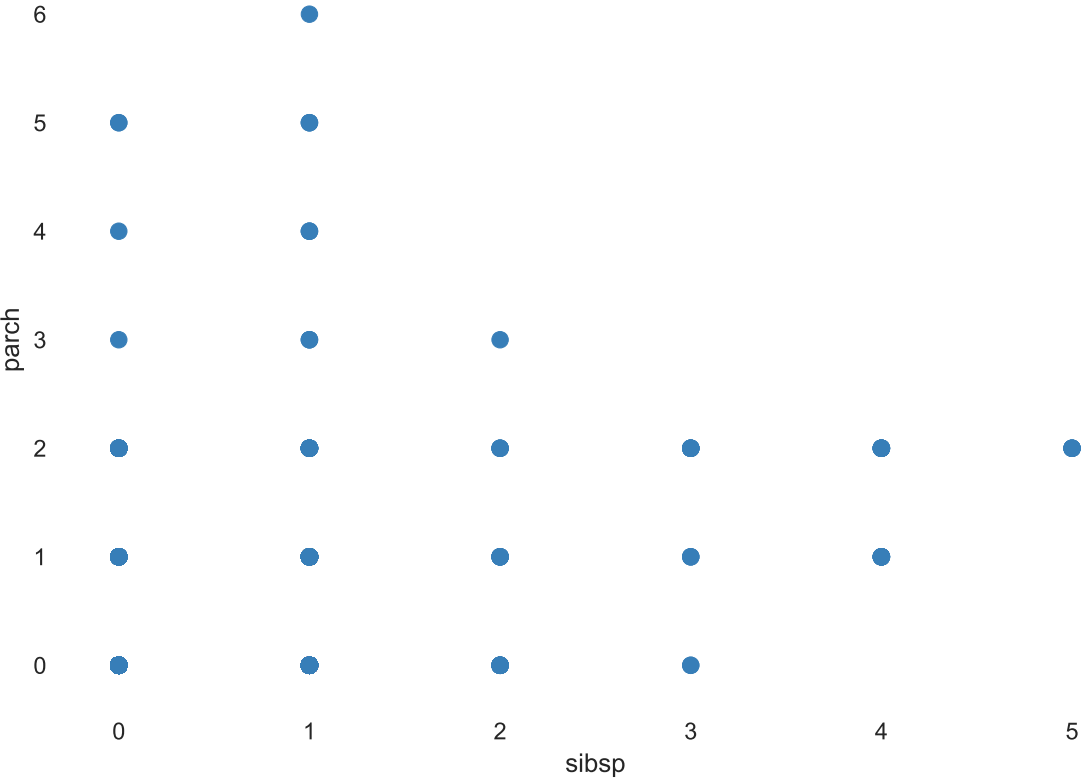


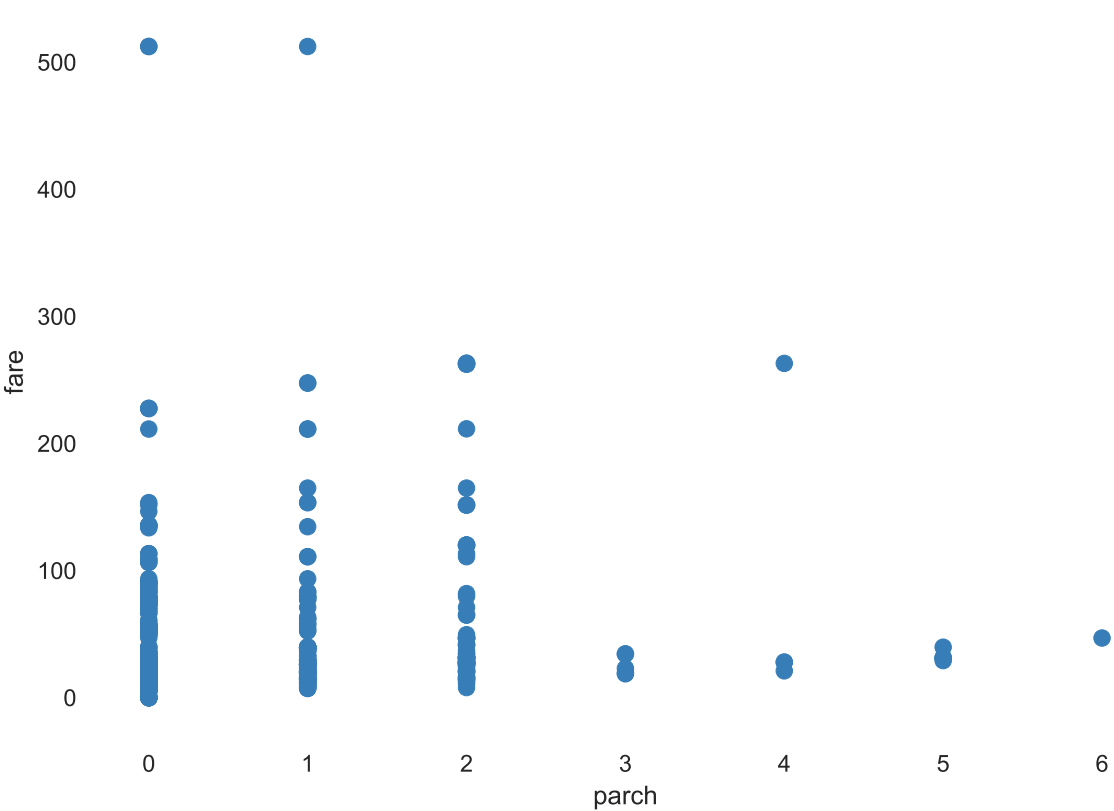


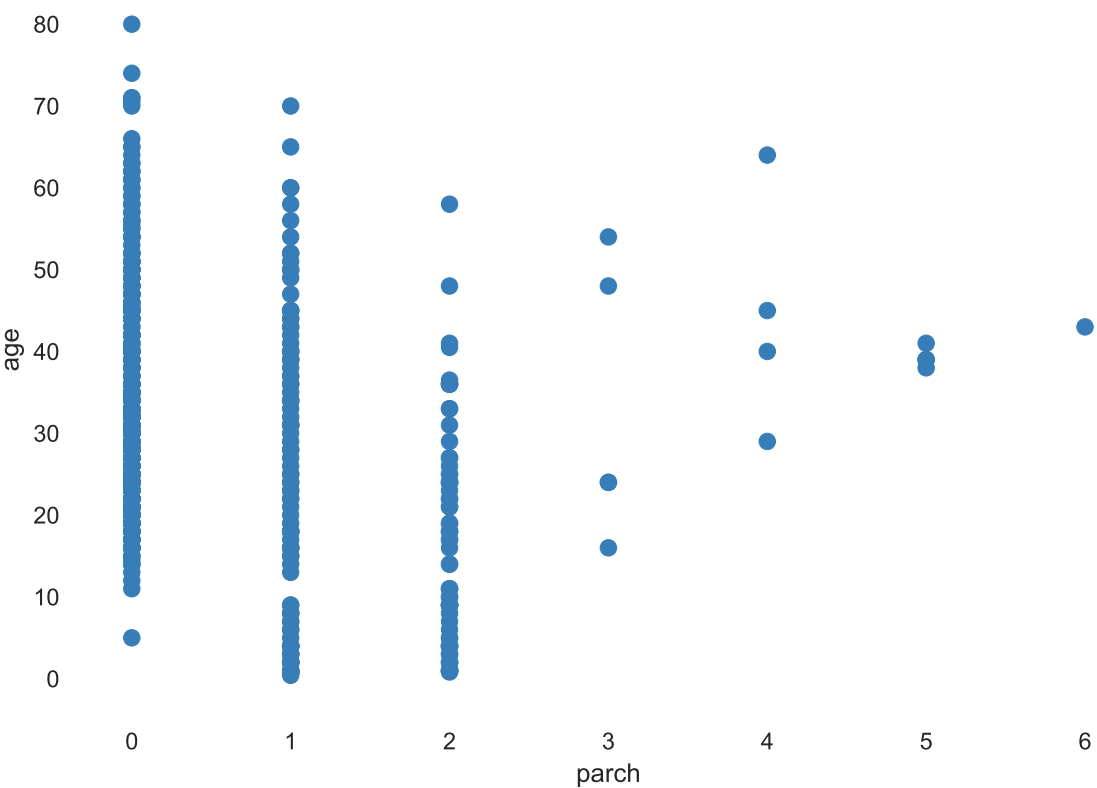


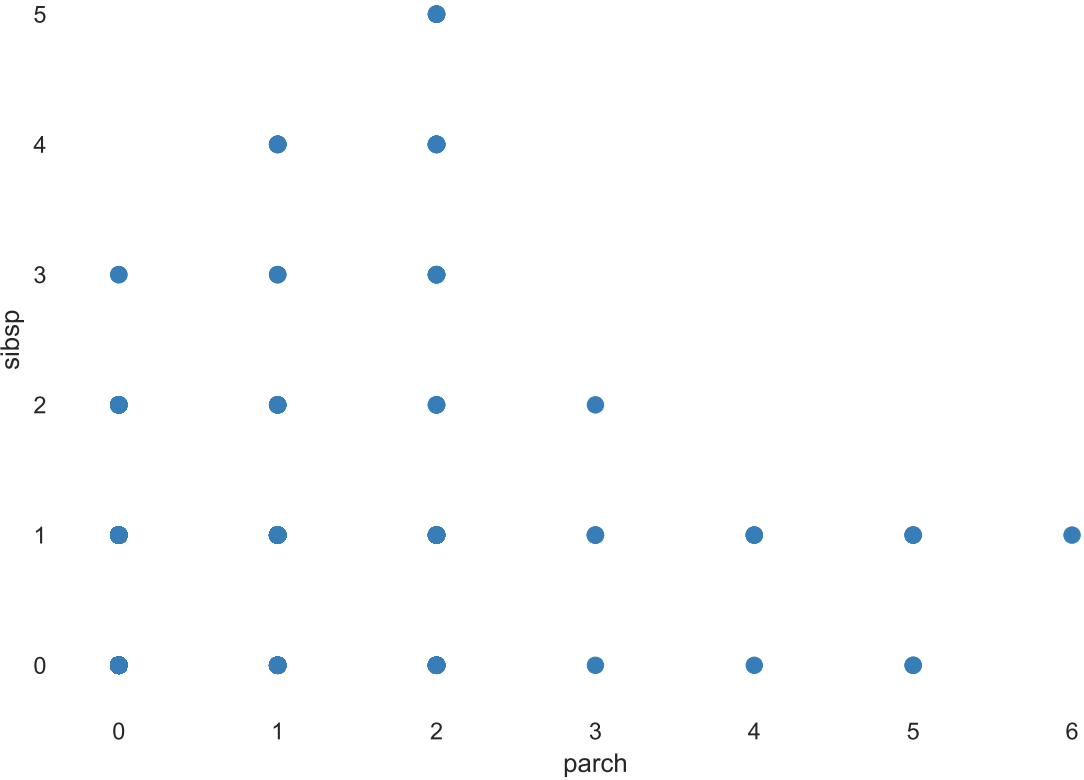


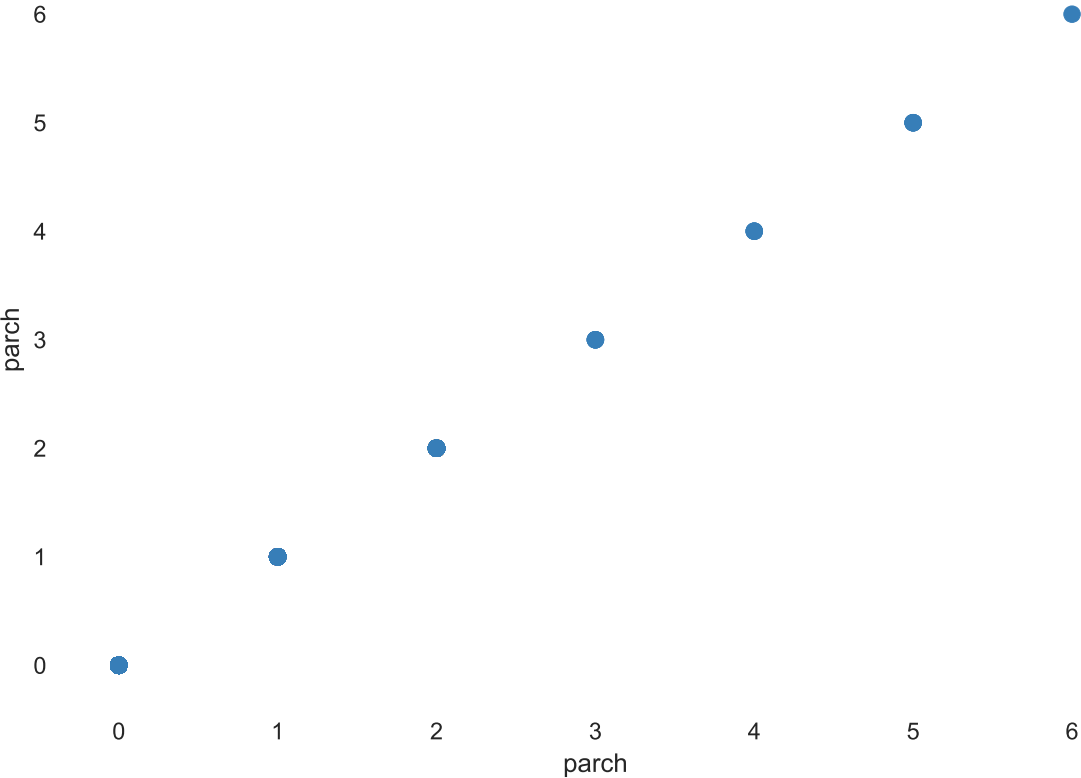


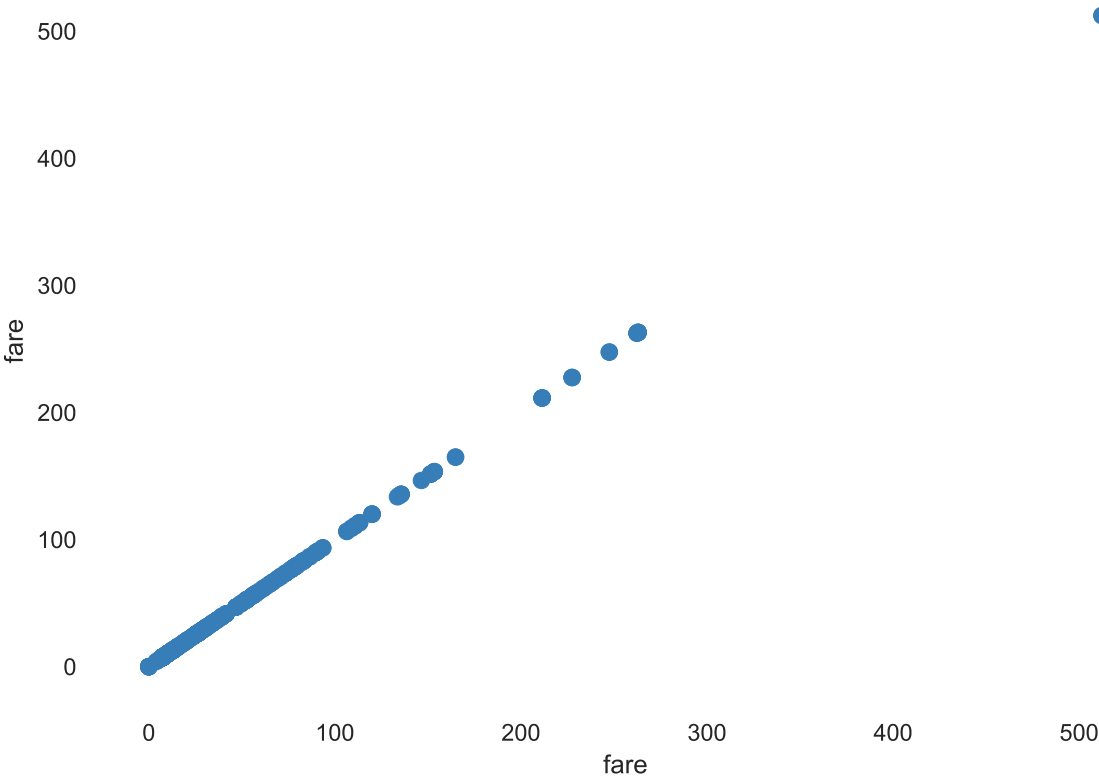


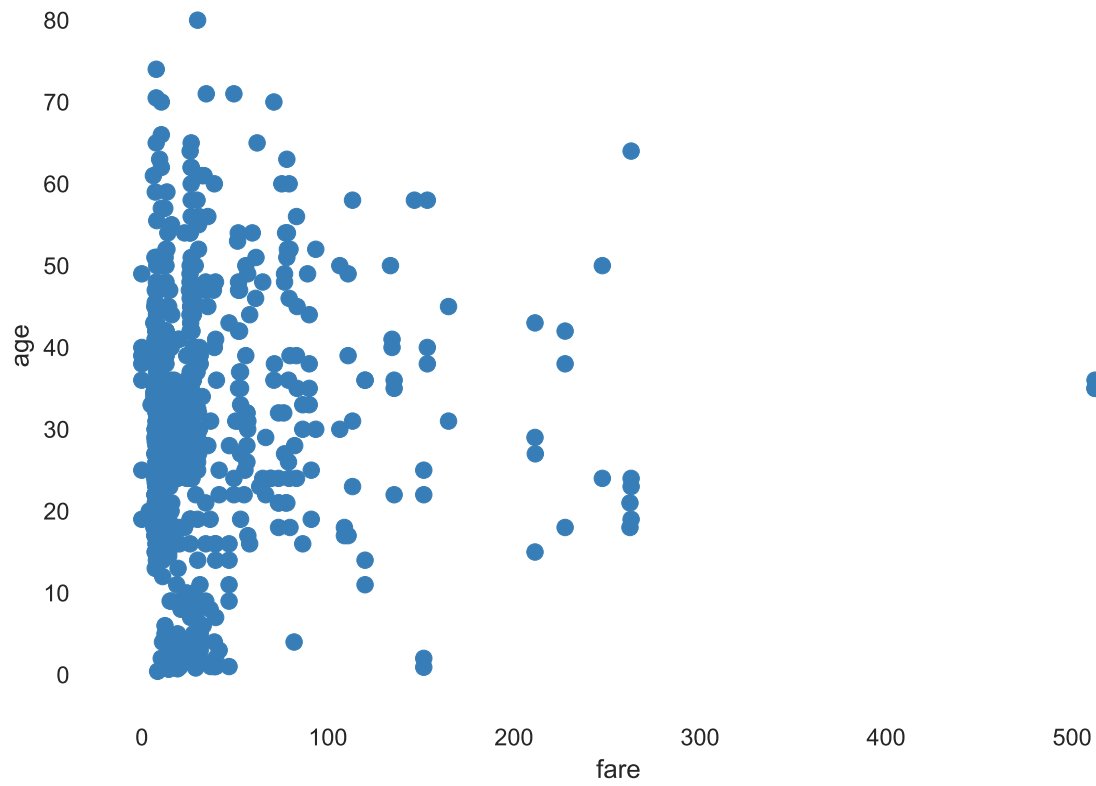


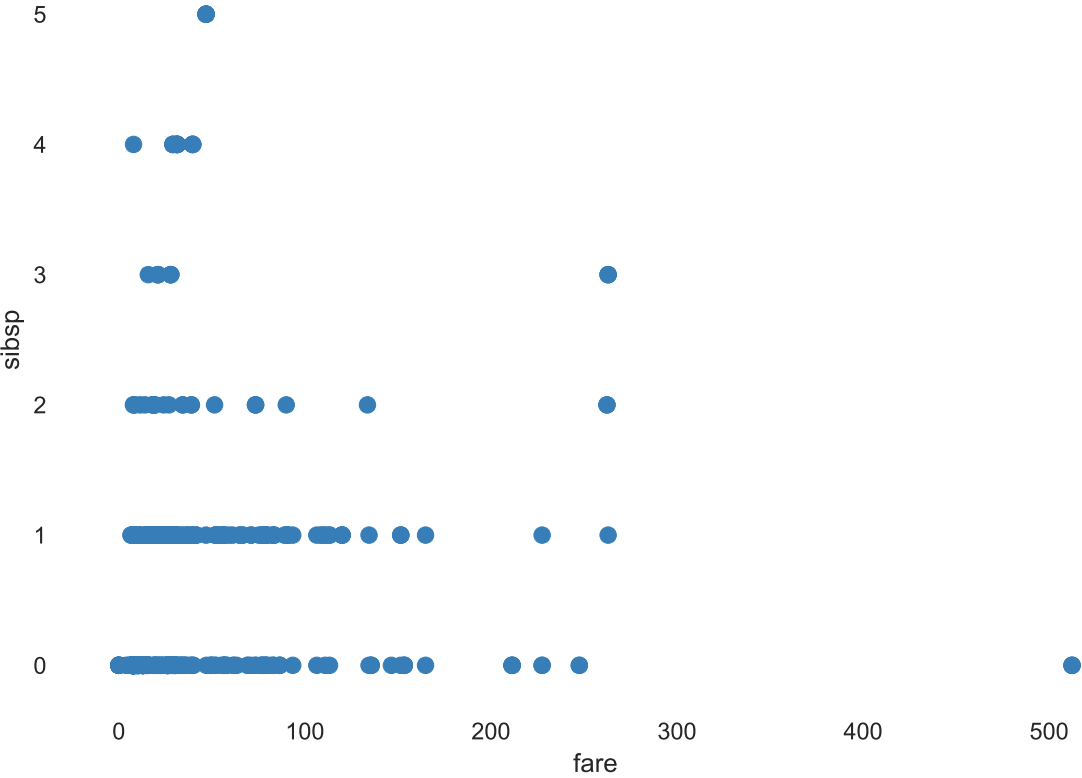


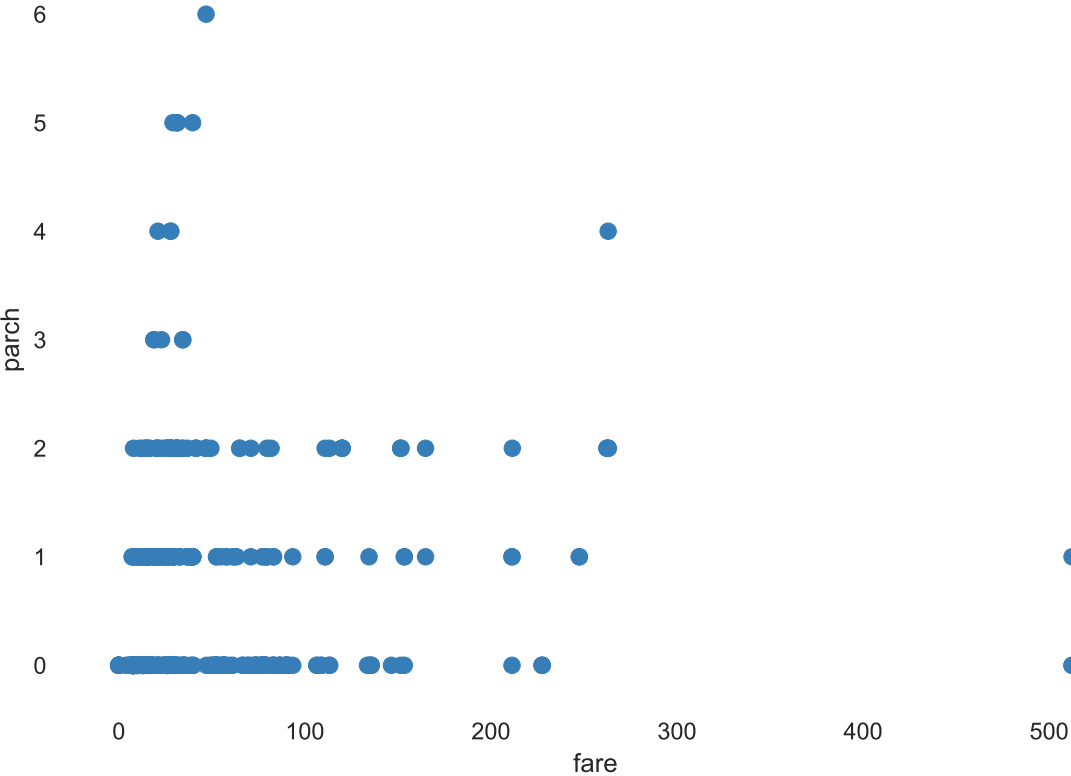




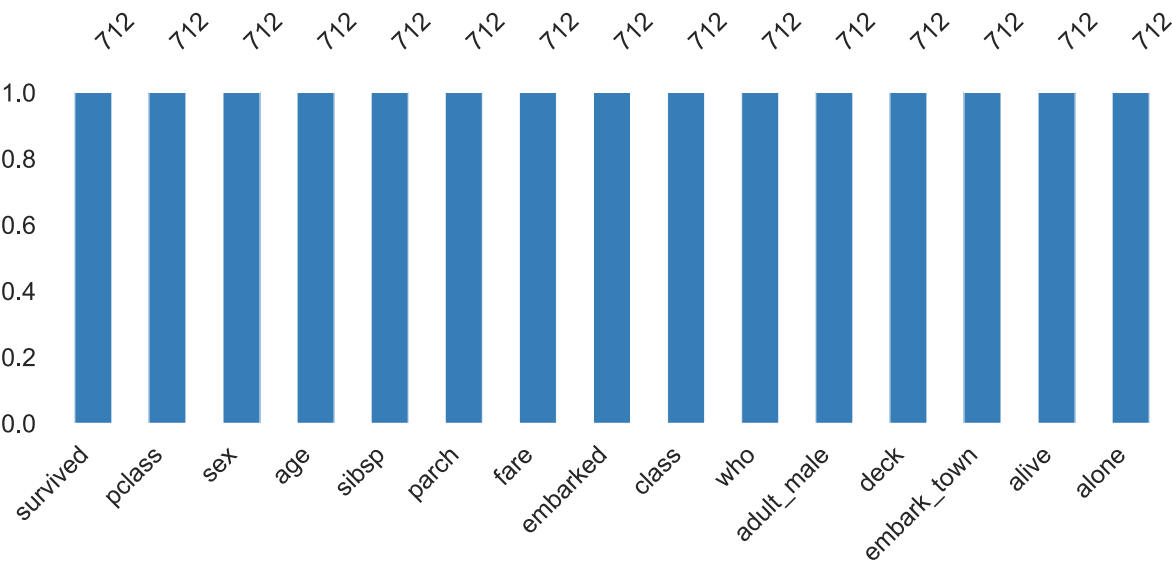




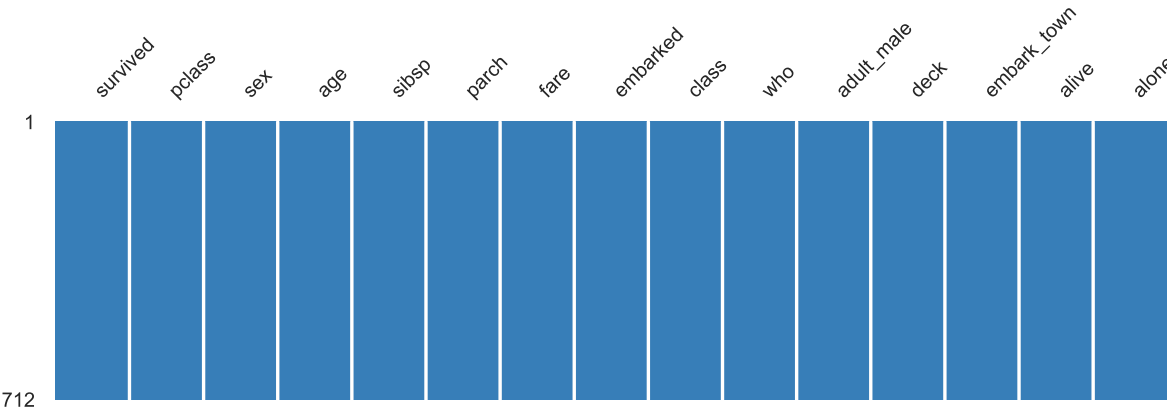




Missing values



A simple visualization of nullity by column.



Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.

Sample

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	Unknown	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	Unknown	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	Unknown	Southampton	no	True
6	0	1	male	54.0	0	0	51.8625	S	First	man	True	E	Southampton	no	True
7	0	3	male	2.0	3	1	21.0750	S	Third	child	False	Unknown	Southampton	no	False
8	1	3	female	27.0	0	2	11.1333	S	Third	woman	False	Unknown	Southampton	yes	False
9	1	2	female	14.0	1	0	30.0708	C	Second	child	False	Unknown	Cherbourg	yes	False
10	1	3	female	4.0	1	1	16.7000	S	Third	child	False	G	Southampton	yes	False

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
880	1	2	female	25.0	0	1	26.0000	S	Second	woman	False	Unknown	Southampton	yes	False
881	0	3	male	33.0	0	0	7.8958	S	Third	man	True	Unknown	Southampton	no	True
882	0	3	female	22.0	0	0	10.5167	S	Third	woman	False	Unknown	Southampton	no	True
883	0	2	male	28.0	0	0	10.5000	S	Second	man	True	Unknown	Southampton	no	True
884	0	3	male	25.0	0	0	7.0500	S	Third	man	True	Unknown	Southampton	no	True
885	0	3	female	39.0	0	5	29.1250	Q	Third	woman	False	Unknown	Queenstown	no	False
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	Unknown	Southampton	no	True
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	B	Southampton	yes	True
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	C	Cherbourg	yes	True
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	Unknown	Queenstown	no	True



Duplicate rows

Most frequently occurring

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone	# duplicat
1	0	2	male	23.0	0	0	13.0000	S	Second	man	True	Unknown	Southampton	no	True	3
2	0	2	male	25.0	0	0	13.0000	S	Second	man	True	Unknown	Southampton	no	True	3
4	0	2	male	30.0	0	0	13.0000	S	Second	man	True	Unknown	Southampton	no	True	3
8	0	3	male	17.0	0	0	8.6625	S	Third	man	True	Unknown	Southampton	no	True	3
9	0	3	male	19.0	0	0	7.8958	S	Third	man	True	Unknown	Southampton	no	True	3
0	0	2	male	18.0	0	0	11.5000	S	Second	man	True	Unknown	Southampton	no	True	2
3	0	2	male	27.0	0	0	13.0000	S	Second	man	True	Unknown	Southampton	no	True	2
5	0	2	male	34.0	0	0	13.0000	S	Second	man	True	Unknown	Southampton	no	True	2
6	0	2	male	34.0	1	0	21.0000	S	Second	man	True	Unknown	Southampton	no	False	2
7	0	2	male	39.0	0	0	13.0000	S	Second	man	True	Unknown	Southampton	no	True	2



Report generated by YData (https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report).