

# Overview

Brought to you by YData ([https://ydata.ai/?utm\\_source=opensource&utm\\_medium=ydataprofiling&utm\\_campaign=report](https://ydata.ai/?utm_source=opensource&utm_medium=ydataprofiling&utm_campaign=report))

## Dataset statistics

Number of variables	8
Number of observations	887
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	152.6 KiB
Average record size in memory	176.2 B

## Variable types

Categorical	3
Text	1
Numeric	4

## Alerts

Sex is highly overall correlated with Survived	High correlation
Survived is highly overall correlated with Sex	High correlation
Name has unique values	Unique
Siblings/Spouses Aboard has 604 (68.1%) zeros	Zeros
Parents/Children Aboard has 674 (76.0%) zeros	Zeros
Fare has 15 (1.7%) zeros	Zeros

## Reproduction

Analysis started	2024-09-08 22:37:15.984208
Analysis finished	2024-09-08 22:37:18.291705
Duration	2.31 seconds
Software version	ydata-profiling vv4.10.0 ( <a href="https://github.com/ydataai/ydata-profiling">https://github.com/ydataai/ydata-profiling</a> )
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Titanic%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%20%22siblings%22%3A%20%22604%22%2C%20%22parents%22%3A%20%22674%22%2C%20%22fare%22%3A%20%2215%22%7D%7D)

# Variables

Select Columns 

Survived  
Categorical

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Sex)

Distinct	2
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	43.4 KiB

Length

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

Characters and  
Unicode

Total characters	887	Unique	0	?	1st row	0
Distinct characters	2	Unique (%)	0.0%		2nd row	1
Distinct categories	1				3rd row	1
Distinct scripts	1				4th row	1
Distinct blocks	1				5th row	0
						?
						?
						?

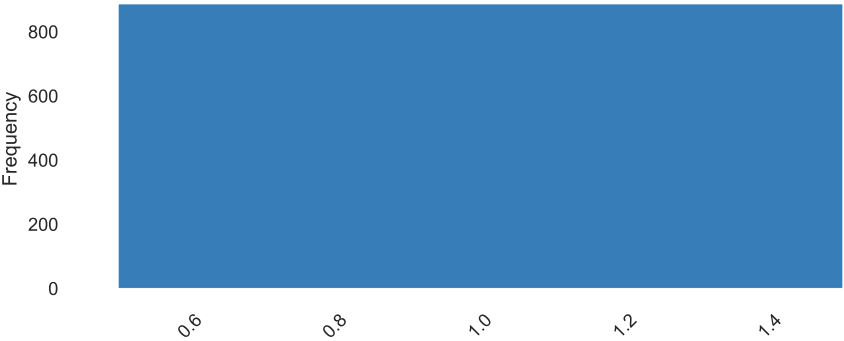
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Common Values

Value	Count	Frequency (%)
0	545	61.4%

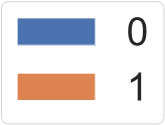
Value	Count	Frequency (%)
1	342	38.6%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
0	545	61.4%
1	342	38.6%

Most occurring characters

Value	Count	Frequency (%)
0	545	61.4%

Value	Count	Frequency (%)
1	342	38.6%

## Most occurring categories

Value	Count	Frequency (%)
(unknown)	887	100.0%

## Most frequent character per category

*(unknown)*

Value	Count	Frequency (%)
0	545	61.4%
1	342	38.6%

## Most occurring scripts

Value	Count	Frequency (%)
(unknown)	887	100.0%

## Most frequent character per script

*(unknown)*

Value	Count	Frequency (%)
0	545	61.4%
1	342	38.6%

## Most occurring blocks

Value	Count	Frequency (%)
(unknown)	887	100.0%

Most frequent character per block

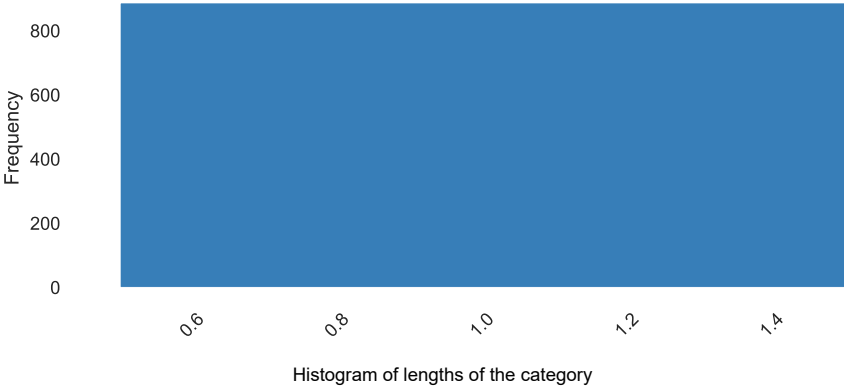
(unknown)

Value	Count	Frequency (%)
0	545	61.4%
1	342	38.6%

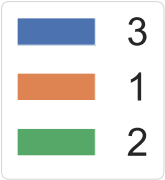
Pclass  
Categorical

Distinct	3
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	43.4 KiB

Length		Characters and Unicode		Unique		Sample	
Max length	1	Total characters	887	Unique	0 ?	1st row	3
Median length	1			Unique (%)	0.0%	2nd row	1
Mean length	1	Distinct characters	3			3rd row	3
Min length	1					4th row	1
		Distinct categories	1			5th row	3 ?
		Distinct scripts	1				
		Distinct blocks	1				



Common Values (Plot)



Value	Count	Frequency (%)
3	487	54.9%
1	216	24.4%
2	184	20.7%

Most occurring characters

Value	Count	Frequency (%)
3	487	54.9%
1	216	24.4%

Value	Count	Frequency (%)
2	184	20.7%

## Most occurring categories

Value	Count	Frequency (%)
(unknown)	887	100.0%

## Most frequent character per category

*(unknown)*

Value	Count	Frequency (%)
3	487	54.9%
1	216	24.4%
2	184	20.7%

## Most occurring scripts

Value	Count	Frequency (%)
(unknown)	887	100.0%

## Most frequent character per script

*(unknown)*

Value	Count	Frequency (%)
3	487	54.9%
1	216	24.4%
2	184	20.7%

## Most occurring blocks



Value	Count	Frequency (%)
(unknown)	887	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
3	487	54.9%
1	216	24.4%
2	184	20.7%

Name  
Text

UNIQUE	
Distinct	887
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Memory size	64.5 KiB



Length		Characters and Unicode		Unique		Sample			
Max length	81	Total characters	22461	Unique	887	?	1st row	Mr. Owen Harris Braund	
Median length	50			Unique (%)	100.0%	2nd row	Mrs. John Bradley (Florence Briggs Thayer) Cummings	?	category)
Mean length	25.322435			Distinct characters	58	3rd row	Miss. Laina Heikkinen	?	code)
Min length	11	Distinct categories	1	(https://en.wikipedia.org/wiki/Unicode_character_category)				?	category)
		Distinct scripts	1	(https://en.wikipedia.org/wiki/Script_(Unicode_block))				?	code)
		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.						5th row	Mr. William Henry Allen
Value	Count	Frequency (%)							
mr	513	14.5%							
miss	182	5.1%							
mrs	125	3.5%							
william	62	1.8%							
john	44	1.2%							
master	40	1.1%							



Value	Count	Frequency (%)
henry	33	0.9%
james	24	0.7%
charles	23	0.7%
george	22	0.6%
Other values ...	2468	69.8%

## Most occurring characters

Value	Count	Frequency (%)
	2649	11.8%
r	1914	8.5%
e	1654	7.4%
a	1618	7.2%
i	1287	5.7%
s	1281	5.7%
n	1273	5.7%
M	1105	4.9%
l	1039	4.6%
o	986	4.4%
Other values (48)	7655	34.1%

## Most occurring categories

Value	Count	Frequency (%)
(unknown)	22461	100.0%

## Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	2649	11.8%
r	1914	8.5%
e	1654	7.4%
a	1618	7.2%
i	1287	5.7%
s	1281	5.7%
n	1273	5.7%
M	1105	4.9%
l	1039	4.6%
o	986	4.4%
Other values (48)	7655	34.1%

## Most occurring scripts

Value	Count	Frequency (%)
(unknown)	22461	100.0%

## Most frequent character per script

*(unknown)*

Value	Count	Frequency (%)
	2649	11.8%
r	1914	8.5%
e	1654	7.4%
a	1618	7.2%
i	1287	5.7%
s	1281	5.7%
n	1273	5.7%
M	1105	4.9%

Value	Count	Frequency (%)
l	1039	4.6%
o	986	4.4%
Other values (48)	7655	34.1%

## Most occurring blocks

Value	Count	Frequency (%)
(unknown)	22461	100.0%

## Most frequent character per block

*(unknown)*

Value	Count	Frequency (%)
	2649	11.8%
r	1914	8.5%
e	1654	7.4%
a	1618	7.2%
i	1287	5.7%
s	1281	5.7%
n	1273	5.7%
M	1105	4.9%
l	1039	4.6%
o	986	4.4%
Other values (48)	7655	34.1%

Sex

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Survived)

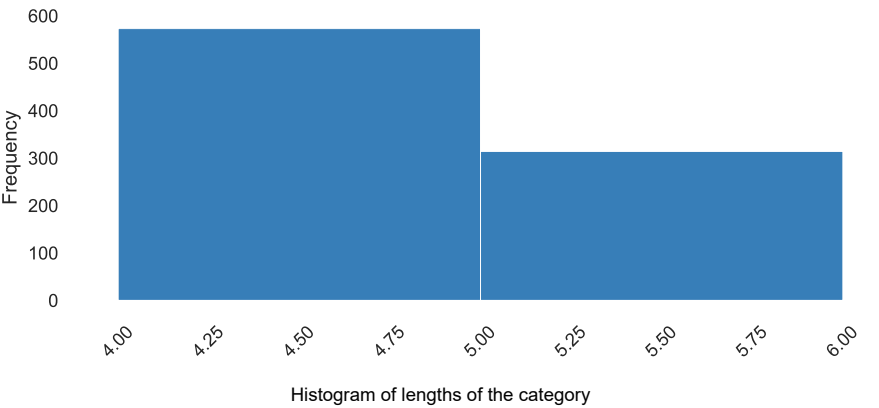
Distinct	2
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	46.7 KiB

Length		Characters and Unicode		Unique		Sample	
Max length	6	Total characters	4176	Unique	0 ?	1st row	male
Median length	4	Distinct characters	5	Unique (%)	0.0%	2nd row	female
Mean length	4.7080045	Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			3rd row	female
Min length	4	Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			4th row	female
		Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)			5th row	male ?
		The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.					

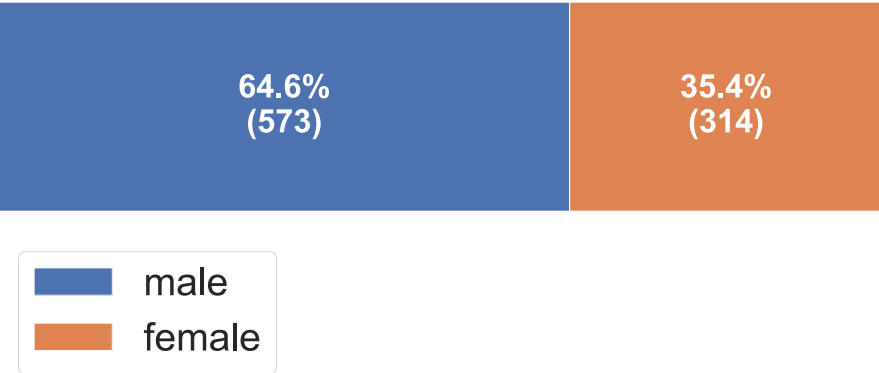
Common Values

Value	Count	Frequency (%)
male	573	64.6%
female	314	35.4%

Length



Common Values (Plot)



Value	Count	Frequency (%)
male	573	64.6%
female	314	35.4%

Most occurring characters

Value	Count	Frequency (%)
e	1201	28.8%
m	887	21.2%
a	887	21.2%
l	887	21.2%

Value	Count	Frequency (%)
f	314	7.5%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	4176	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
e	1201	28.8%
m	887	21.2%
a	887	21.2%
l	887	21.2%
f	314	7.5%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	4176	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
e	1201	28.8%
m	887	21.2%
a	887	21.2%
l	887	21.2%



Value	Count	Frequency (%)
f	314	7.5%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	4176	100.0%

Most frequent character per block

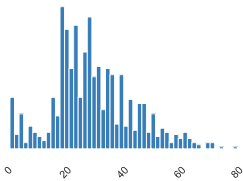
(unknown)

Value	Count	Frequency (%)
e	1201	28.8%
m	887	21.2%
a	887	21.2%
l	887	21.2%
f	314	7.5%

Age

Real number ( $\mathbb{R}$ )

Distinct	89	Minimum	0.42
Distinct (%)	10.0%	Maximum	80
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	29.471443	Memory size	7.1 KiB

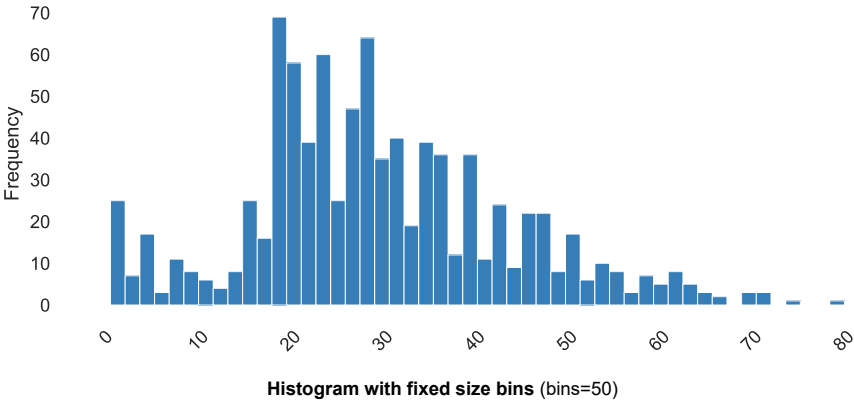


Quantile statistics

Minimum	0.42
5-th percentile	5
Q1	20.25
median	28
Q3	38
95-th percentile	55.85
Maximum	80
Range	79.58
Interquartile range (IQR)	17.75

Descriptive statistics

Standard deviation	14.121908
Coefficient of variation (CV)	0.47917261
Kurtosis	0.29255909
Mean	29.471443
Median Absolute Deviation (MAD)	8
Skewness	0.44718857
Sum	26141.17
Variance	199.4283
Monotonicity	Not monotonic



Value	Count	Frequency (%)
22	39	4.4%
28	37	4.2%
18	36	4.1%
24	34	3.8%
21	34	3.8%
30	33	3.7%
19	33	3.7%
27	26	2.9%
23	25	2.8%
29	25	2.8%
Other values (79)	565	63.7%

Value	Count	Frequency (%)
0.42	1	0.1%
0.67	1	0.1%
0.75	2	0.2%
0.83	2	0.2%
0.92	1	0.1%
1	7	0.8%
2	11	1.2%
3	7	0.8%
4	11	1.2%
5	6	0.7%

Value	Count	Frequency (%)
80	1	0.1%
74	1	0.1%
71	2	0.2%

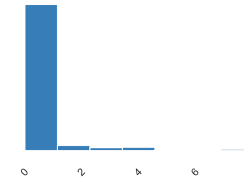
Value	Count	Frequency (%)
70.5	1	0.1%
70	2	0.2%
69	1	0.1%
66	2	0.2%
65	3	0.3%
64	3	0.3%
63	2	0.2%

## Siblings/Spouses Aboard

Real number ( $\mathbb{R}$ )

ZEROS

<b>Distinct</b>	7	<b>Minimum</b>	0
<b>Distinct (%)</b>	0.8%	<b>Maximum</b>	8
<b>Missing</b>	0	<b>Zeros</b>	604
<b>Missing (%)</b>	0.0%	<b>Zeros (%)</b>	68.1%
<b>Infinite</b>	0	<b>Negative</b>	0
<b>Infinite (%)</b>	0.0%	<b>Negative (%)</b>	0.0%
<b>Mean</b>	0.5253664	<b>Memory size</b>	7.1 KiB

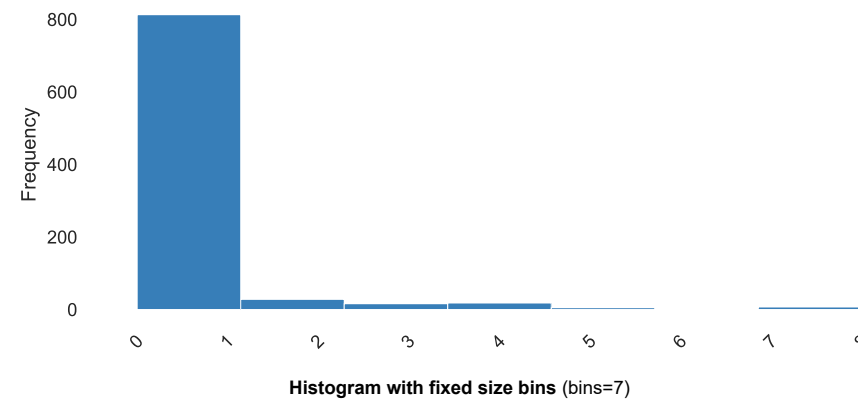


## Quantile statistics

<b>Minimum</b>	0
<b>5-th percentile</b>	0
<b>Q1</b>	0
<b>median</b>	0
<b>Q3</b>	1
<b>95-th percentile</b>	3
<b>Maximum</b>	8
<b>Range</b>	8
<b>Interquartile range (IQR)</b>	1

## Descriptive statistics

<b>Standard deviation</b>	1.1046686
<b>Coefficient of variation (CV)</b>	2.1026631
<b>Kurtosis</b>	17.797537
<b>Mean</b>	0.5253664
<b>Median Absolute Deviation (MAD)</b>	0
<b>Skewness</b>	3.6867598
<b>Sum</b>	466
<b>Variance</b>	1.2202926
<b>Monotonicity</b>	Not monotonic



Value	Count	Frequency (%)
0	604	68.1%
1	209	23.6%
2	28	3.2%
4	18	2.0%
3	16	1.8%
8	7	0.8%
5	5	0.6%

Value	Count	Frequency (%)
0	604	68.1%
1	209	23.6%
2	28	3.2%
3	16	1.8%
4	18	2.0%
5	5	0.6%
8	7	0.8%

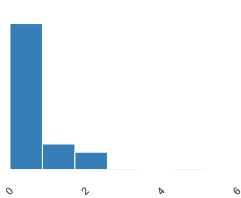
Value	Count	Frequency (%)
8	7	0.8%
5	5	0.6%
4	18	2.0%
3	16	1.8%
2	28	3.2%
1	209	23.6%
0	604	68.1%

Parents/Children Aboard

Real number ( $\mathbb{R}$ )

ZEROS

Distinct	7	Minimum	0
Distinct (%)	0.8%	Maximum	6
Missing	0	Zeros	674
Missing (%)	0.0%	Zeros (%)	76.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	0.38331454	Memory size	7.1 KiB

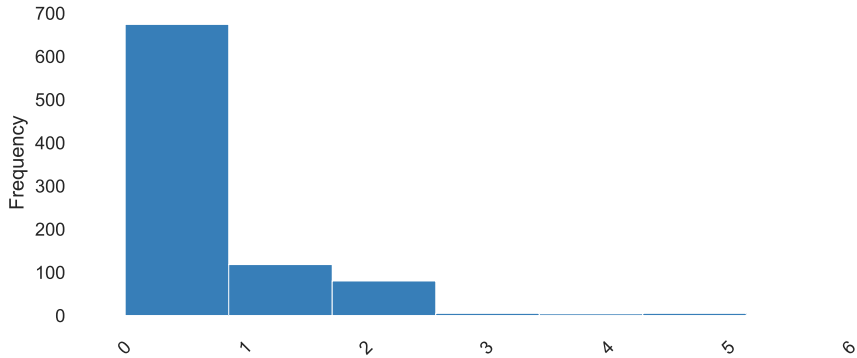


Quantile statistics

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

Descriptive statistics

Standard deviation	0.80746591
Coefficient of variation (CV)	2.1065361
Kurtosis	9.7230659
Mean	0.38331454
Median Absolute Deviation (MAD)	0
Skewness	2.7411981
Sum	340
Variance	0.65200119
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=7)

Value	Count	Frequency (%)
0	674	76.0%
1	118	13.3%
2	80	9.0%
5	5	0.6%
3	5	0.6%
4	4	0.5%
6	1	0.1%

Value	Count	Frequency (%)
0	674	76.0%
1	118	13.3%
2	80	9.0%
3	5	0.6%
4	4	0.5%
5	5	0.6%
6	1	0.1%

Value	Count	Frequency (%)
6	1	0.1%
5	5	0.6%
4	4	0.5%
3	5	0.6%
2	80	9.0%
1	118	13.3%
0	674	76.0%

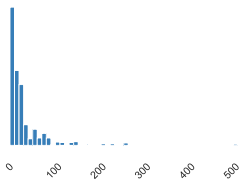


Fare

Real number (R)

ZEROS

Distinct	248	Minimum	0
Distinct (%)	28.0%	Maximum	512.3292
Missing	0	Zeros	15
Missing (%)	0.0%	Zeros (%)	1.7%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	32.30542	Memory size	7.1 KiB

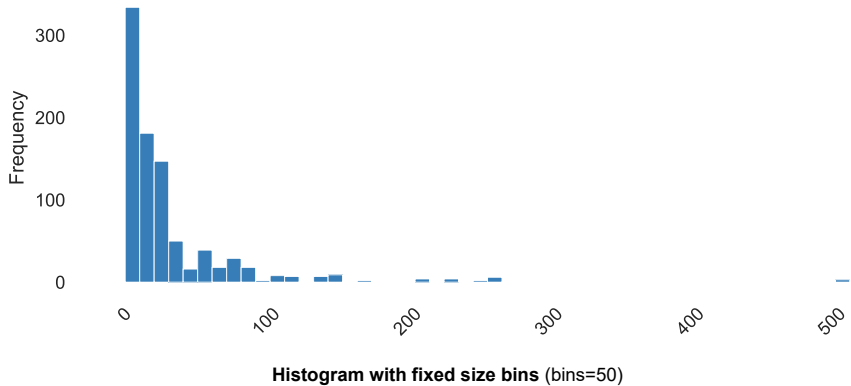


Quantile statistics

Minimum	0
5-th percentile	7.225
Q1	7.925
median	14.4542
Q3	31.1375
95-th percentile	112.55749
Maximum	512.3292
Range	512.3292
Interquartile range (IQR)	23.2125

Descriptive statistics

Standard deviation	49.78204
Coefficient of variation (CV)	1.5409811
Kurtosis	33.264605
Mean	32.30542
Median Absolute Deviation (MAD)	6.9584
Skewness	4.7776714
Sum	28654.908
Variance	2478.2515
Monotonicity	Not monotonic



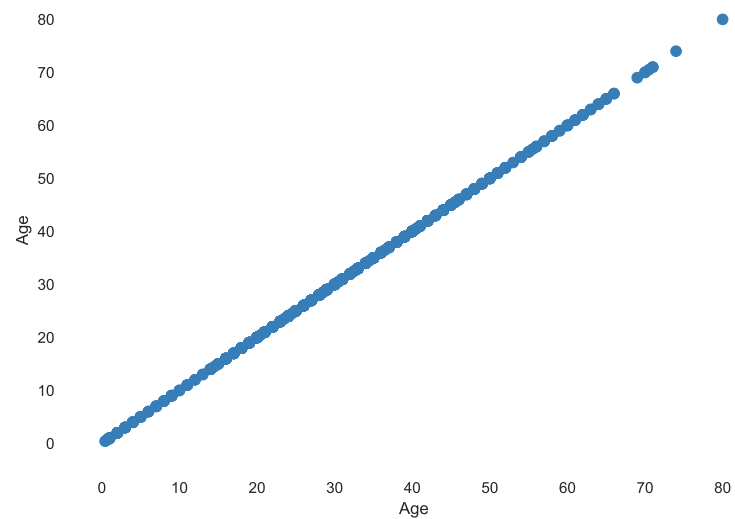
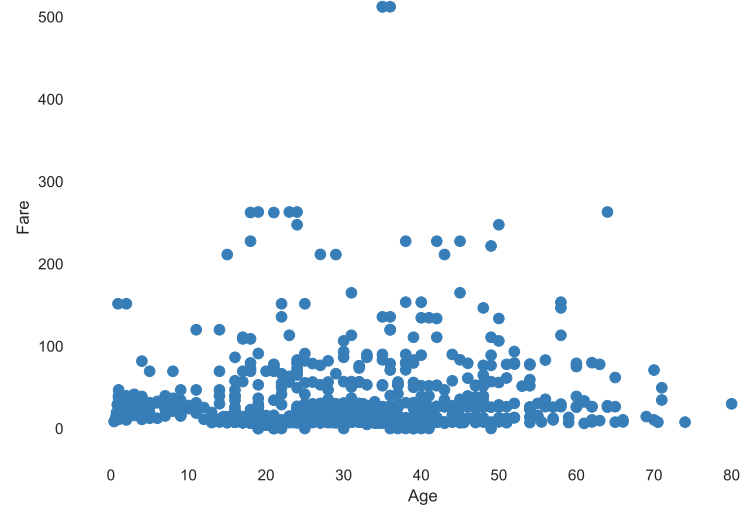
Value	Count	Frequency (%)
8.05	43	4.8%
13	42	4.7%
7.8958	36	4.1%
7.75	33	3.7%
26	31	3.5%
10.5	24	2.7%
7.925	18	2.0%
7.775	16	1.8%
7.2292	15	1.7%
26.55	15	1.7%
Other values (238)	614	69.2%

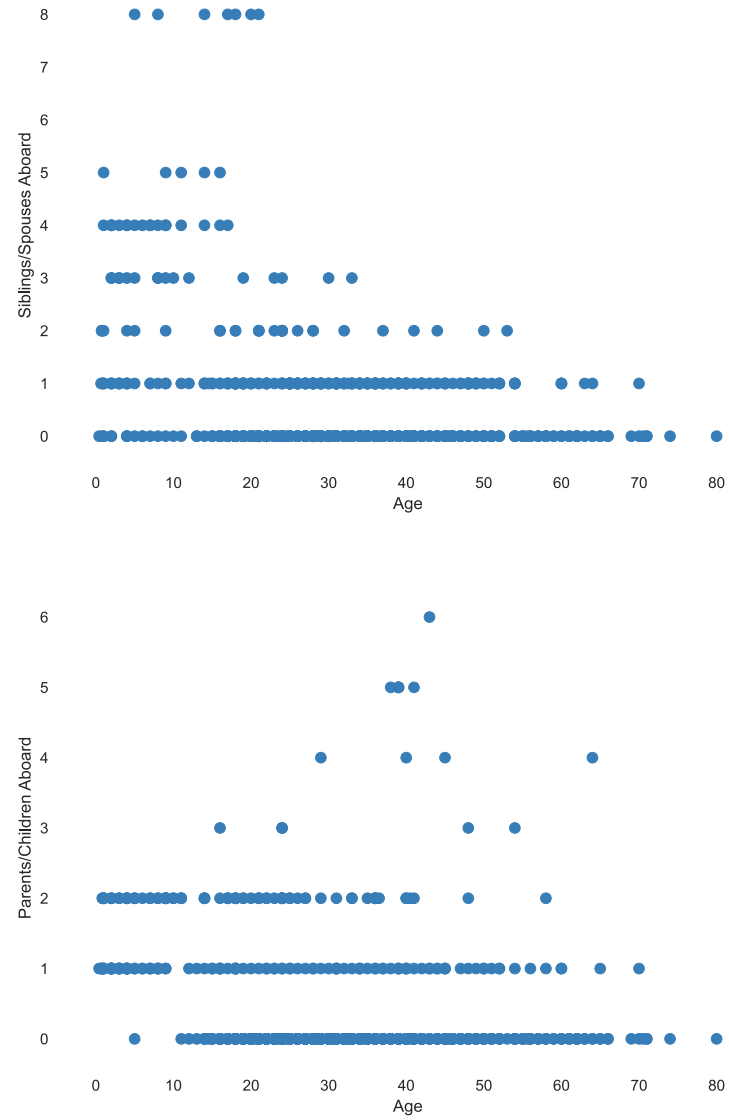
Value	Count	Frequency (%)
0	15	1.7%
4.0125	1	0.1%
5	1	0.1%
6.2375	1	0.1%
6.4375	1	0.1%
6.45	1	0.1%
6.4958	2	0.2%
6.75	2	0.2%
6.8583	1	0.1%
6.95	1	0.1%

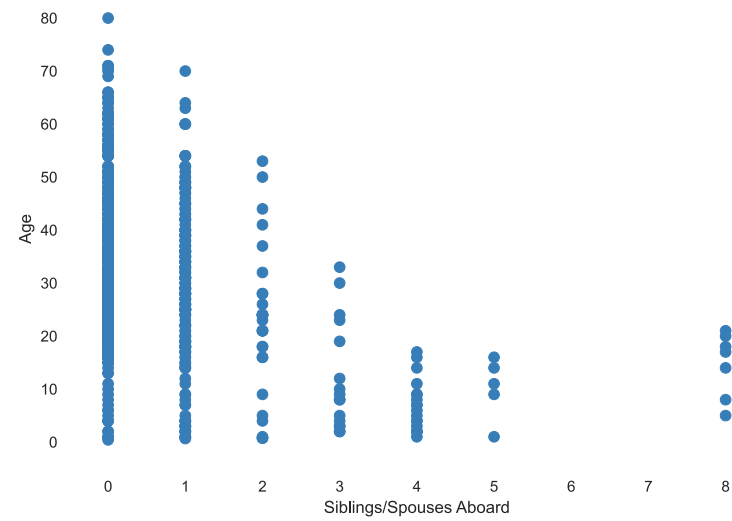
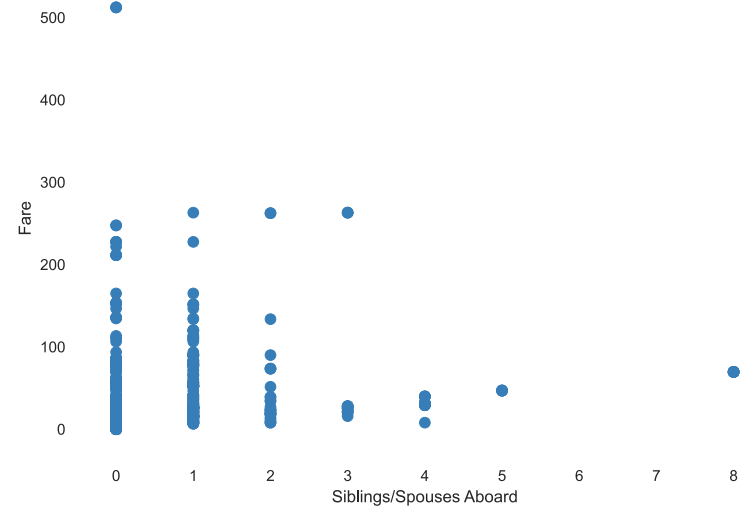
Value	Count	Frequency (%)
512.3292	3	0.3%
263	4	0.5%
262.375	2	0.2%

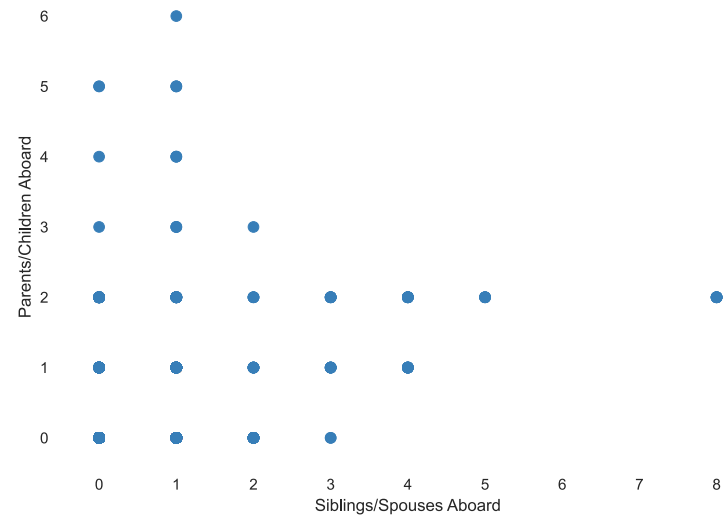
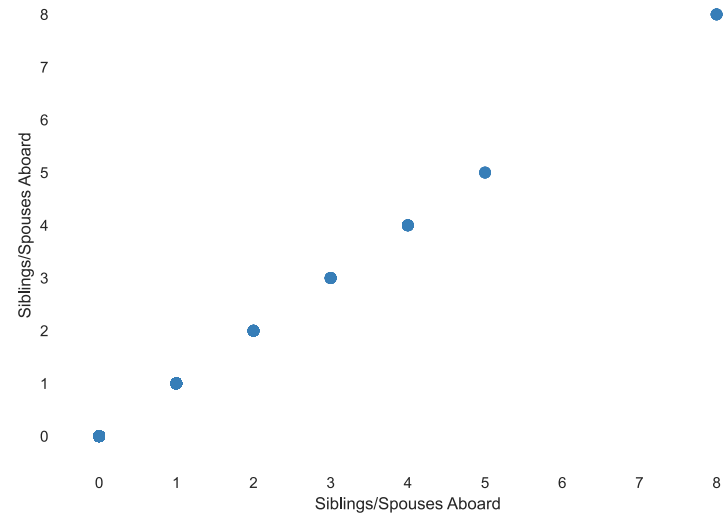
Value	Count	Frequency (%)
247.5208	2	0.2%
227.525	4	0.5%
221.7792	1	0.1%
211.5	1	0.1%
211.3375	3	0.3%
164.8667	2	0.2%
153.4625	3	0.3%

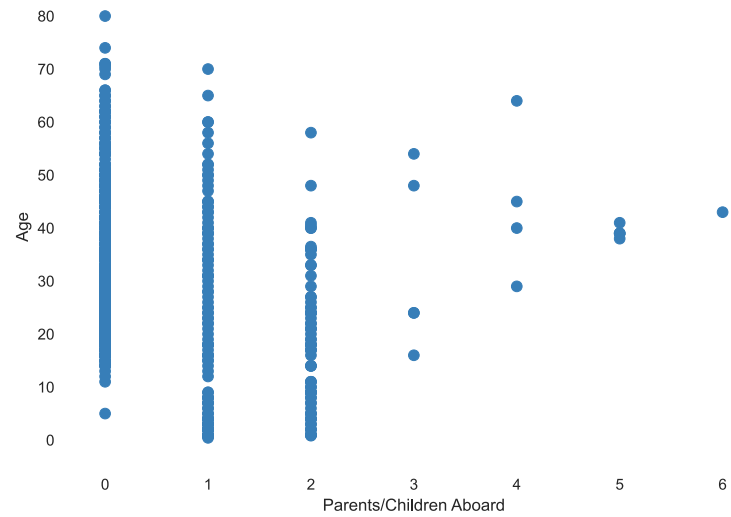
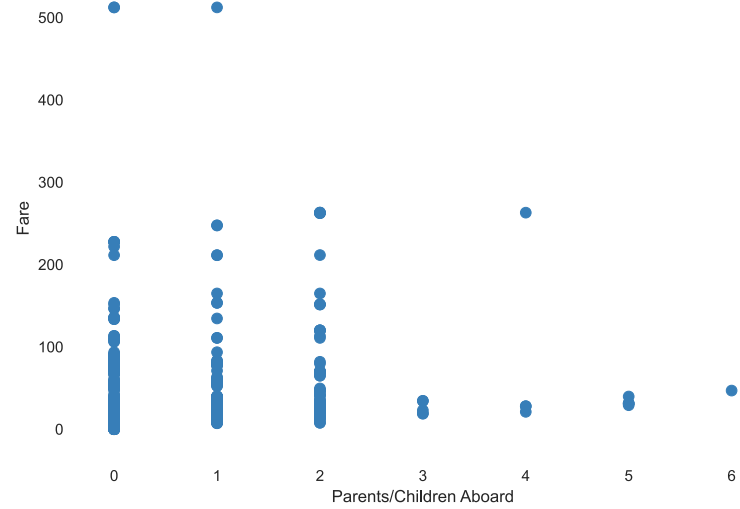
# Interactions



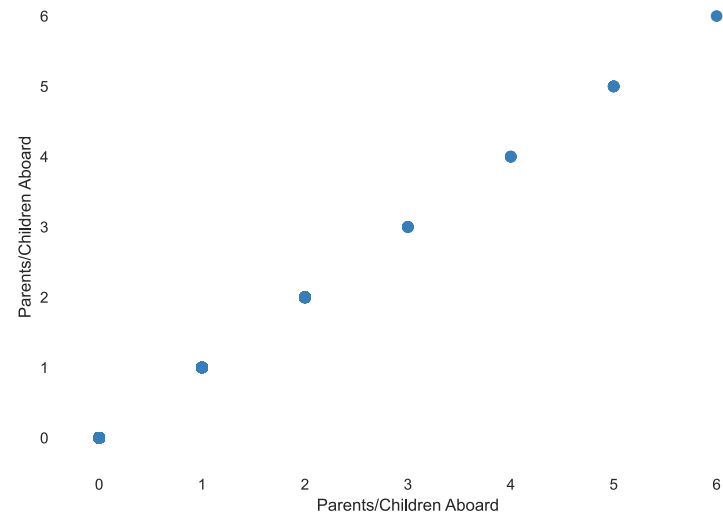
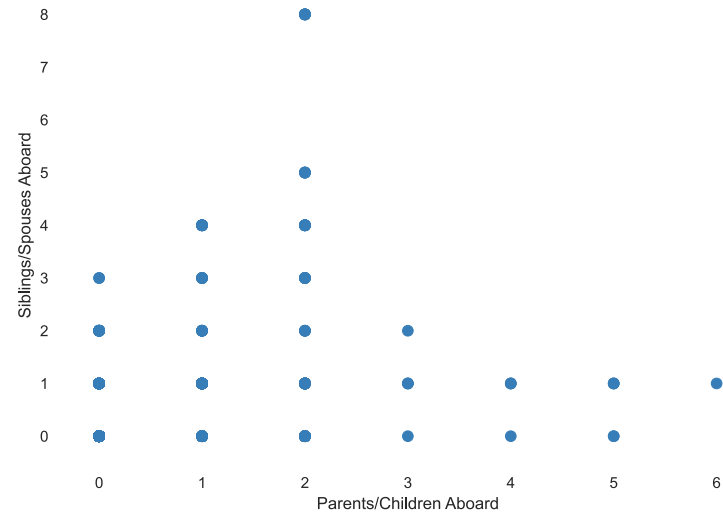


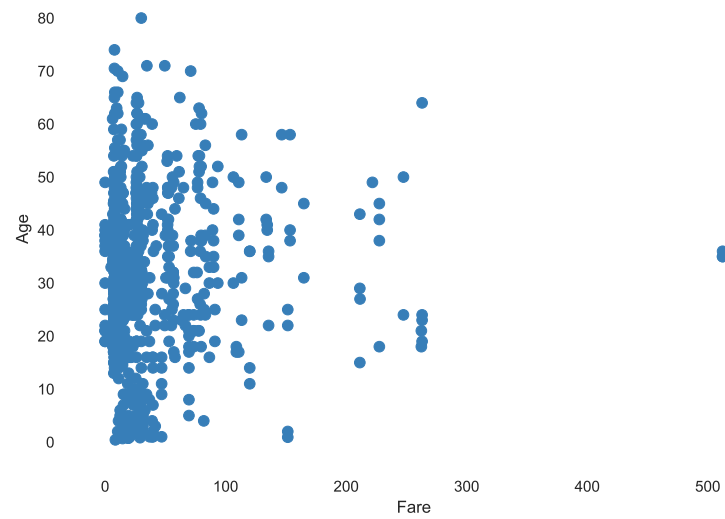
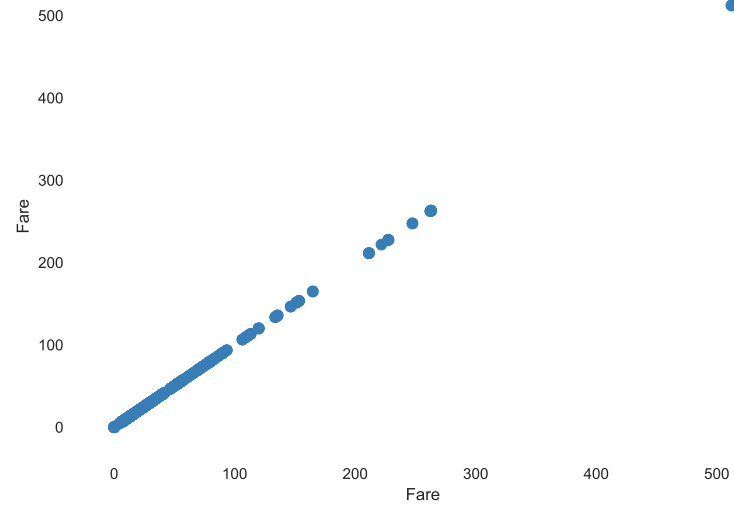


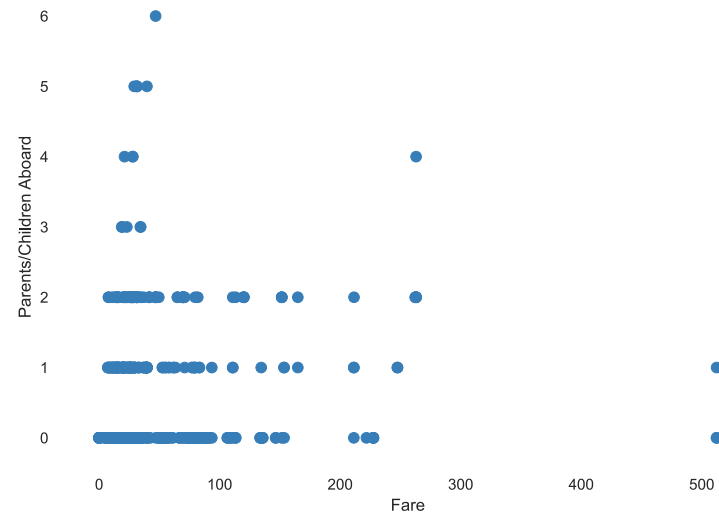
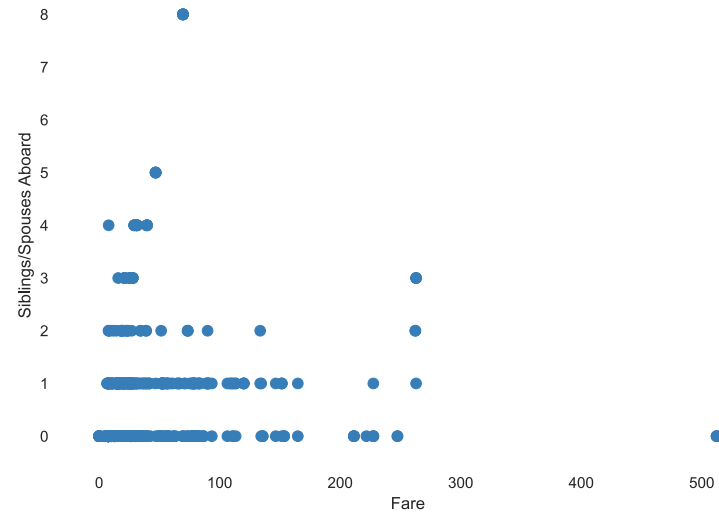




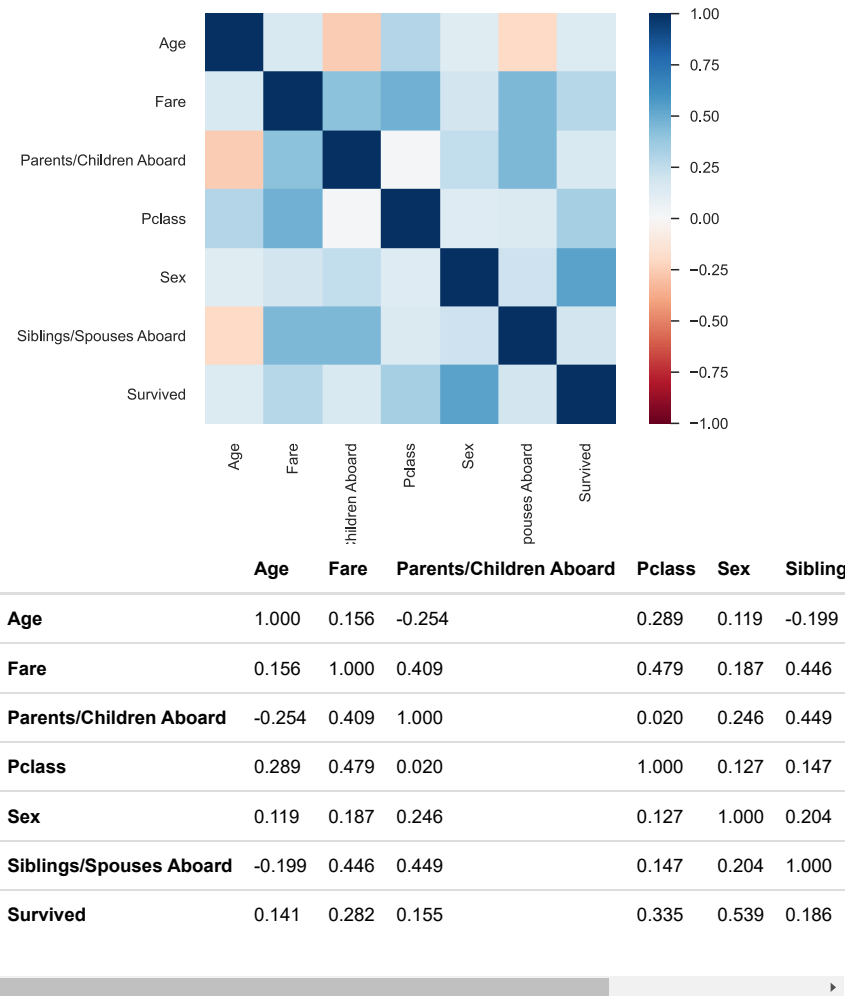




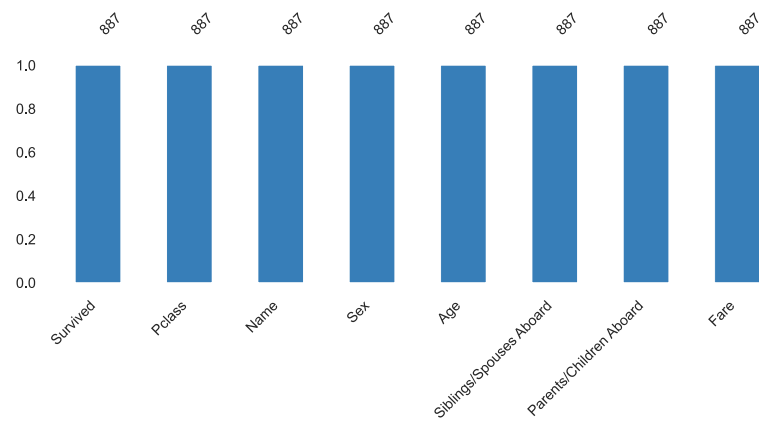




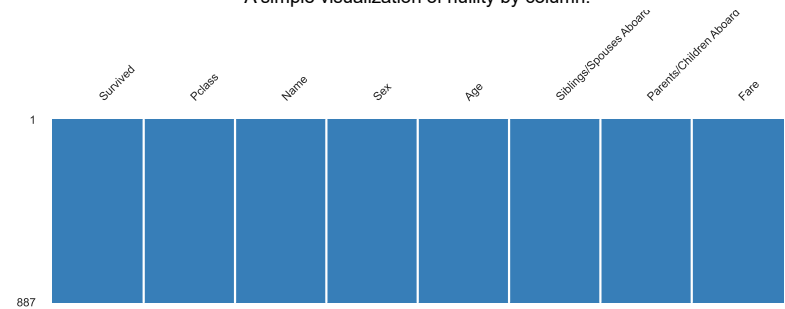
## Correlations



## 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	Name	Sex	Age	Siblings
0	0	3	Mr. Owen Harris Braund	male	22.0	1
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cumings	female	38.0	1
2	1	3	Miss. Laina Heikkinen	female	26.0	0
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1
4	0	3	Mr. William Henry Allen	male	35.0	0
5	0	3	Mr. James Moran	male	27.0	0

	Survived	Pclass	Name	Sex	Age	Sibling:
6	0	1	Mr. Timothy J McCarthy	male	54.0	0
7	0	3	Master. Gosta Leonard Palsson	male	2.0	3
8	1	3	Mrs. Oscar W (Elisabeth Vilhelmina Berg) Johnson	female	27.0	0
9	1	2	Mrs. Nicholas (Adele Achem) Nasser	female	14.0	1

	Survived	Pclass	Name	Sex	Age	Siblings/Spouses Ab
877	0	3	Mr. Johann Markun	male	33.0	0
878	0	3	Miss. Gerda Ulrika Dahlberg	female	22.0	0
879	0	2	Mr. Frederick James Banfield	male	28.0	0
880	0	3	Mr. Henry Jr Sutehall	male	25.0	0
881	0	3	Mrs. William (Margaret Norton) Rice	female	39.0	0
882	0	2	Rev. Juozas Montvila	male	27.0	0
883	1	1	Miss. Margaret Edith Graham	female	19.0	0
884	0	3	Miss. Catherine Helen Johnston	female	7.0	1
885	1	1	Mr. Karl Howell Behr	male	26.0	0
886	0	3	Mr. Patrick Dooley	male	32.0	0



Report generated by YData ([https://ydata.ai/?utm\\_source=opensource&utm\\_medium=pandasprofiling&utm\\_campaign=report](https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report)).