Overview

Dataset statistics

Number of variables	11
Number of observations	45716
Missing cells	29682
Missing cells (%)	5.9%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	3.8 MiB
Average record size in memory	88.0 B

Variable types

Categorical	4
Numeric	7

Alerts

`name` has a high cardinality: 45716 distinct values	High cardinality
`recclass` has a high cardinality: 466 distinct values	High cardinality
`id` is highly overall correlated with `year`	High correlation
`year` is highly overall correlated with `id`	High correlation
`reclat` is highly overall correlated with `reclong` and 2 other fields (reclong, Latitude, Longitude)	High correlation
`reclong` is highly overall correlated with `reclat` and 2 other fields (reclat, Latitude, Longitude)	High correlation
`Latitude` is highly overall correlated with `reclat` and 2 other fields (reclat, reclong, Longitude)	High correlation
`Longitude` is highly overall correlated with `reclat` and 2 other fields (reclat, reclong, Latitude)	High correlation
`nametype` is highly imbalanced (98.2%)	Imbalance
`fall` is highly imbalanced (83.6%)	Imbalance
`reclat` has 7315 (16.0%) missing values	Missing
`reclong` has 7315 (16.0%) missing values	Missing
`Latitude` has 7315 (16.0%) missing values	Missing
`Longitude` has 7315 (16.0%) missing values	Missing
`mass (g)` is highly skewed (γ1 = 76.91011732)	Skewed
`name` is uniformly distributed	Uniform
`name` has unique values	Unique
`id` has unique values	Unique
`reclat` has 6438 (14.1%) zeros	Zeros
`reclong` has 6214 (13.6%) zeros	Zeros
`Latitude` has 6438 (14.1%) zeros	Zeros
`Longitude` has 6214 (13.6%) zeros	Zeros

Reproduction

Analysis started	2022-12-22 01:38:16.348012
Analysis finished	2022-12-22 01:38:34.263781
Duration	17.92 seconds
Software version	pandas-profiling vv3.6.0 (https://github.com/pandas-profiling/pandas-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Pandas%20Proling%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%20%22creator%22%3A%20%22%22%2C%20%22author%22%3A%20

Variables

Select Columns

name

Categorical

HIGH CARDINALITY UNIFORM UNIQUE

Distinct	45716
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Memory size	357.3 KiB

Length

Max length	28
Median length	25
Mean length	17.784605
Min length	2

Characters and Unicode

Total characters	813041	
Distinct characters	96	
Distinct categories	8 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	2 (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	45716	?
Unique (%)	100.0%	

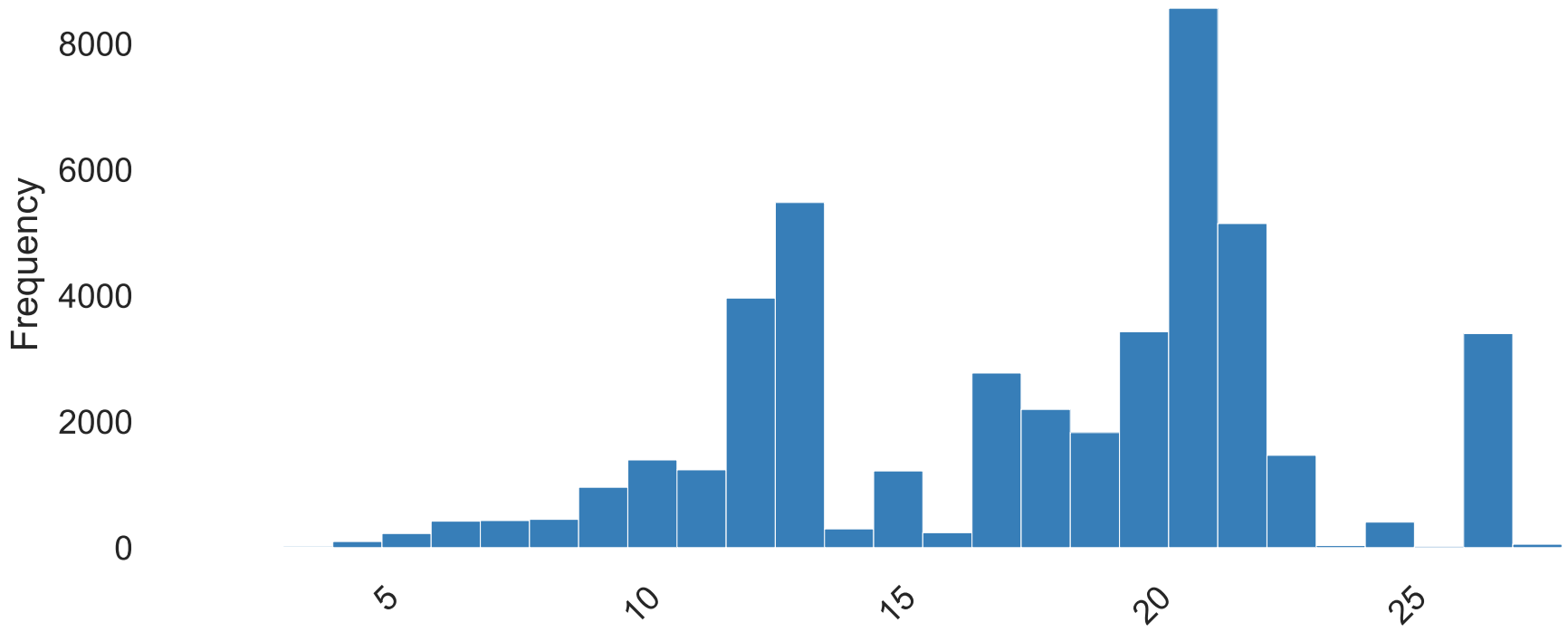
Sample

1st row	Aachen
2nd row	Aarhus
3rd row	Abee
4th row	Acapulco
5th row	Achiras

Common Values

Value	Count	Frequency (%)
Aachen	1	< 0.1%
Northwest Africa 7459	1	< 0.1%
Northwest Africa 7404	1	< 0.1%
Northwest Africa 7407	1	< 0.1%
Northwest Africa 7408	1	< 0.1%
Northwest Africa 741	1	< 0.1%
Northwest Africa 7410	1	< 0.1%
Northwest Africa 7412	1	< 0.1%
Northwest Africa 7414	1	< 0.1%
Northwest Africa 742	1	< 0.1%
Other values (45706)	45706	> 99.9%

Length



Histogram of lengths of the category

Value	Count	Frequency (%)
yamato	7269	5.7%
range	6575	5.2%
africa	4502	3.6%
northwest	4499	3.5%
hills	3995	3.2%
queen	3445	2.7%
alexandra	3444	2.7%
mountains	3004	2.4%
al	2663	2.1%
grove	2496	2.0%
Other values (37726)	84847	66.9%

Most occurring characters

Value	Count	Frequency (%)
	81029	10.0%
a	72706	8.9%
e	48162	5.9%
n	38388	4.7%
0	34943	4.3%
r	33095	4.1%
i	32654	4.0%
l	31870	3.9%
t	30896	3.8%
o	30423	3.7%
Other values (86)	378875	46.6%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	440887	54.2%
Decimal Number	205415	25.3%
Uppercase Letter	84929	10.4%
Space Separator	81029	10.0%
Close Punctuation	294	< 0.1%
Open Punctuation	294	< 0.1%
Dash Punctuation	97	< 0.1%
Other Punctuation	96	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
a	72706	16.5%
e	48162	10.9%
n	38388	8.7%
r	33095	7.5%
i	32654	7.4%
l	31870	7.2%
t	30896	7.0%
o	30423	6.9%
s	20969	4.8%
m	12393	2.8%
Other values (39)	89331	20.3%

Uppercase Letter

Value	Count	Frequency (%)
A	14110	16.6%

Value	Count	Frequency (%)
M	11173	13.2%
R	7599	8.9%
Y	7327	8.6%
N	5796	6.8%
H	5676	6.7%
G	4682	5.5%
L	4630	5.5%
D	3777	4.4%
Q	3478	4.1%
Other values (21)	16681	19.6%

Decimal Number

Value	Count	Frequency (%)
0	34943	17.0%
9	24444	11.9%
8	22179	10.8%
1	21986	10.7%
2	19839	9.7%
7	19347	9.4%
3	17379	8.5%
4	16001	7.8%
5	14812	7.2%
6	14485	7.1%

Other Punctuation

Value	Count	Frequency (%)
'	67	69.8%
.	29	30.2%

Space Separator

Value	Count	Frequency (%)
	81029	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	294	100.0%

Open Punctuation

Value	Count	Frequency (%)
(	294	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	97	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	525816	64.7%
Common	287225	35.3%

Most frequent character per script

Latin

Value	Count	Frequency (%)
a	72706	13.8%
e	48162	9.2%
n	38388	7.3%
r	33095	6.3%
i	32654	6.2%
l	31870	6.1%
t	30896	5.9%
o	30423	5.8%
s	20969	4.0%
A	14110	2.7%
Other values (70)	172543	32.8%

Common

Value	Count	Frequency (%)
	81029	28.2%
0	34943	12.2%
9	24444	8.5%
8	22179	7.7%
1	21986	7.7%
2	19839	6.9%
7	19347	6.7%

Value	Count	Frequency (%)
3	17379	6.1%
4	16001	5.6%
5	14812	5.2%
Other values (6)	15266	5.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	812557	99.9%
None	484	0.1%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	81029	10.0%
a	72706	8.9%
e	48162	5.9%
n	38388	4.7%
0	34943	4.3%
r	33095	4.1%
i	32654	4.0%
l	31870	3.9%
t	30896	3.8%
o	30423	3.7%
Other values (58)	378391	46.6%

None

Value	Count	Frequency (%)
é	204	42.1%
ş	125	25.8%
Ö	63	13.0%
á	11	2.3%
ö	11	2.3%
ä	10	2.1%
ñ	8	1.7%

Value	Count	Frequency (%)
ó	8	1.7%
ü	8	1.7%
ã	5	1.0%
Other values (18)	31	6.4%

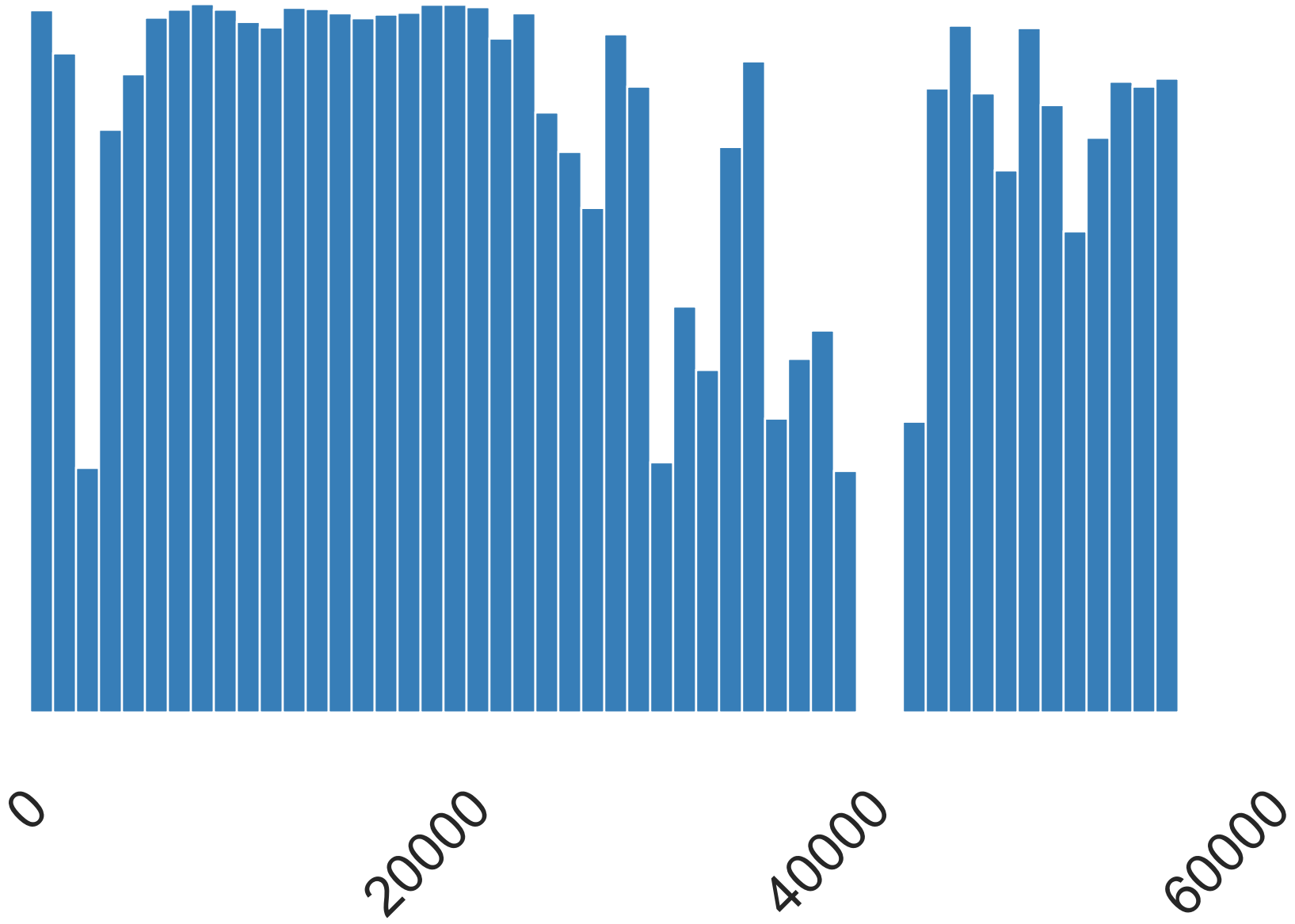
id

Real number (ℝ)

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

Distinct	45716
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	26889.735

Minimum	1
Maximum	57458
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	357.3 KiB

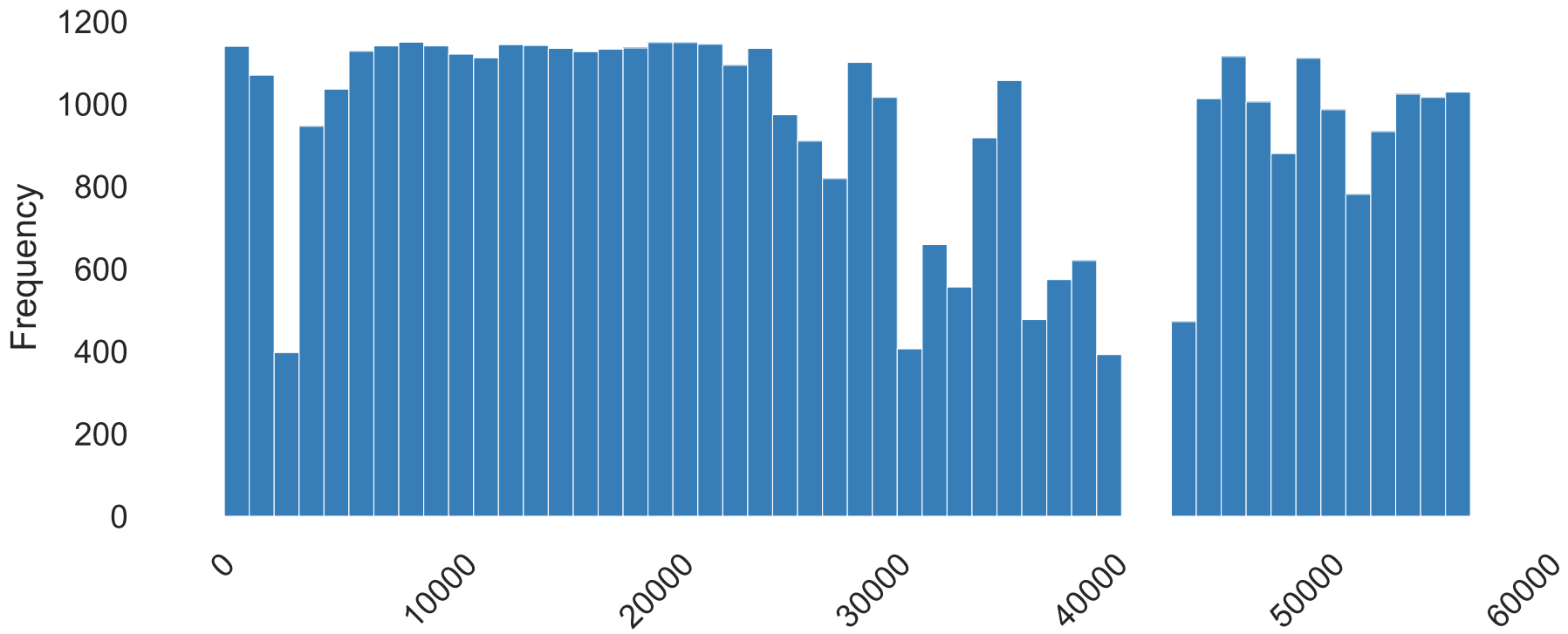

0
20000
40000
60000

Quantile statistics

Minimum	1
5-th percentile	2434.75
Q1	12688.75
median	24261.5
Q3	40656.75
95-th percentile	54891.25
Maximum	57458
Range	57457
Interquartile range (IQR)	27968

Descriptive statistics

Standard deviation	16860.683
Coefficient of variation (CV)	0.62703046
Kurtosis	-1.1602565
Mean	26889.735
Median Absolute Deviation (MAD)	13264
Skewness	0.26654597
Sum	1.2292911×10^{9}
Variance	2.8428263×10^{8}
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
1	1	< 0.1%
56115	1	< 0.1%
55773	1	< 0.1%
55774	1	< 0.1%
55775	1	< 0.1%
17820	1	< 0.1%
55776	1	< 0.1%
55777	1	< 0.1%
55778	1	< 0.1%
17821	1	< 0.1%
Other values (45706)	45706	> 99.9%

Value	Count	Frequency (%)
1	1	< 0.1%
2	1	< 0.1%
4	1	< 0.1%
5	1	< 0.1%
6	1	< 0.1%
7	1	< 0.1%
8	1	< 0.1%
9	1	< 0.1%
10	1	< 0.1%
11	1	< 0.1%

Value	Count	Frequency (%)
57458	1	< 0.1%
57457	1	< 0.1%
57456	1	< 0.1%
57455	1	< 0.1%
57454	1	< 0.1%
57453	1	< 0.1%
57436	1	< 0.1%
57435	1	< 0.1%
57434	1	< 0.1%
57433	1	< 0.1%

nametype
Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	357.3 KiB

Length

Max length	6
Median length	5
Mean length	5.0016406
Min length	5

Characters and Unicode

Total characters	228655	
Distinct characters	9	
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%	

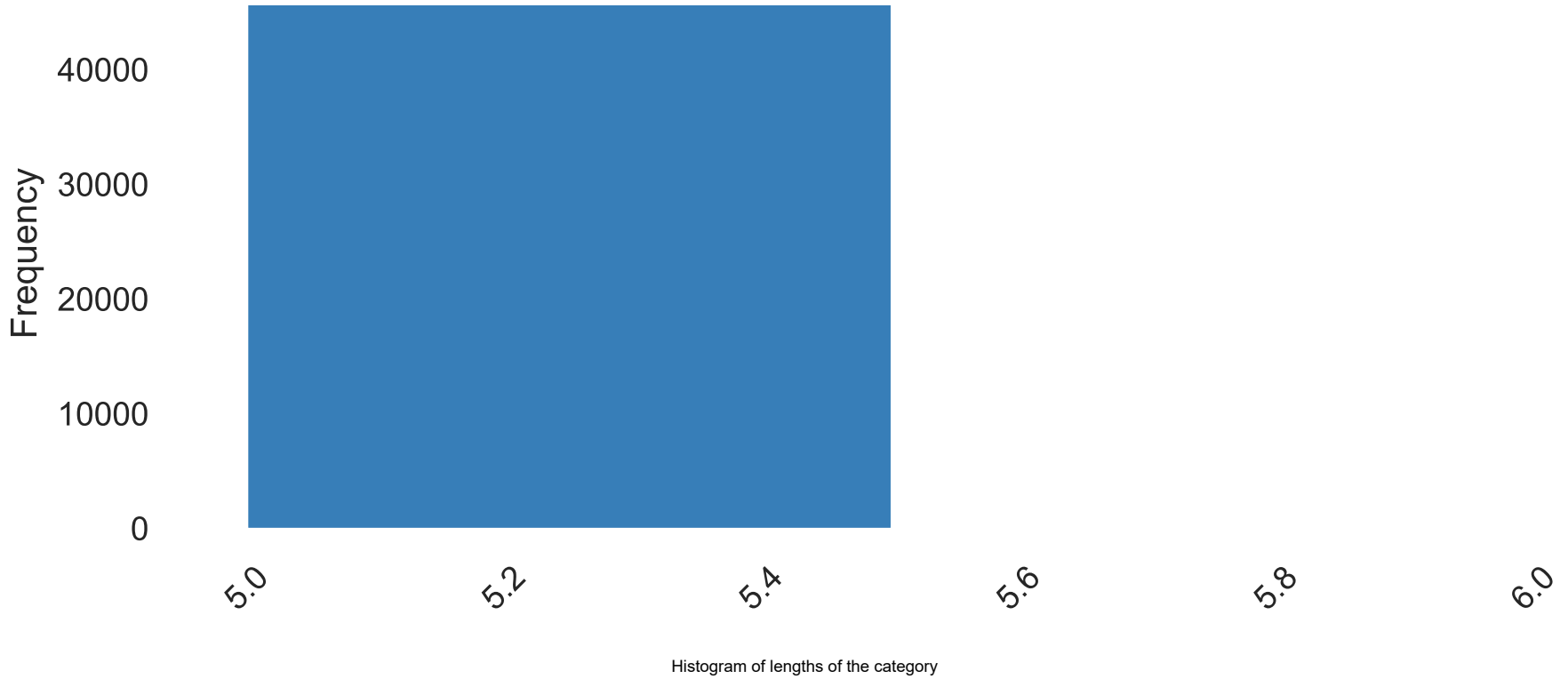
Sample

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

Common Values

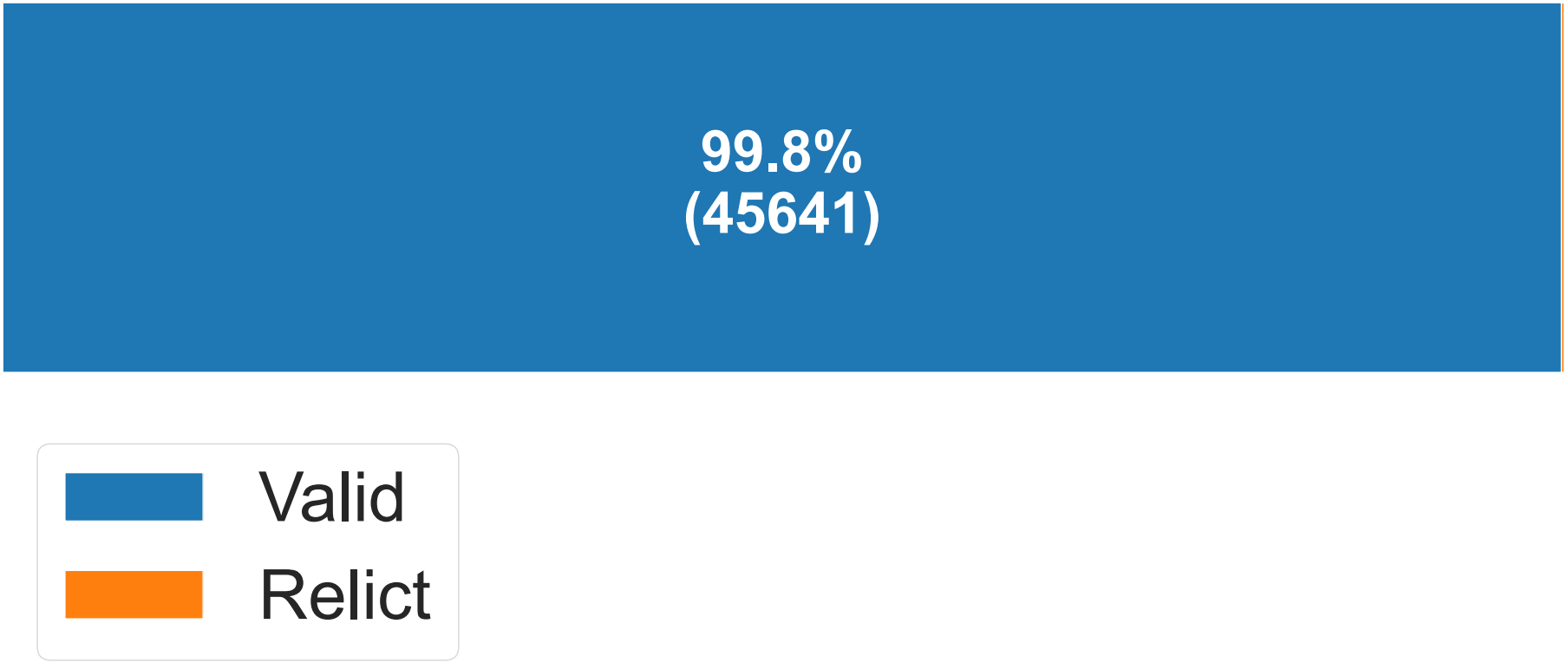
Value	Count	Frequency (%)
Valid	45641	99.8%
Relict	75	0.2%

Length



Histogram of lengths of the category

Common Values (Plot)


99.8%
(45641)
Valid
Relict

Value	Count	Frequency (%)
valid	45641	99.8%
relict	75	0.2%

Most occurring characters

Value	Count	Frequency (%)
l	45716	20.0%
i	45716	20.0%
V	45641	20.0%
a	45641	20.0%
d	45641	20.0%
R	75	< 0.1%
e	75	< 0.1%
c	75	< 0.1%
t	75	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	182939	80.0%
Uppercase Letter	45716	20.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
l	45716	25.0%
i	45716	25.0%
a	45641	24.9%
d	45641	24.9%
e	75	< 0.1%
c	75	< 0.1%
t	75	< 0.1%

Uppercase Letter

Value	Count	Frequency (%)
V	45641	99.8%
R	75	0.2%

Most occurring scripts

Value	Count	Frequency (%)
Latin	228655	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
l	45716	20.0%
i	45716	20.0%
V	45641	20.0%
a	45641	20.0%
d	45641	20.0%
R	75	< 0.1%
e	75	< 0.1%
c	75	< 0.1%
t	75	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	228655	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
l	45716	20.0%
i	45716	20.0%
V	45641	20.0%
a	45641	20.0%
d	45641	20.0%
R	75	< 0.1%
e	75	< 0.1%
c	75	< 0.1%
t	75	< 0.1%

recclass
Categorical

Distinct	466
Distinct (%)	1.0%
Missing	0
Missing (%)	0.0%
Memory size	357.3 KiB

Length

Max length	26
Median length	2
Mean length	3.0524762
Min length	1

Characters and Unicode

Total characters	139547	
Distinct characters	62	
Distinct categories	9 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	145	?
Unique (%)	0.3%	

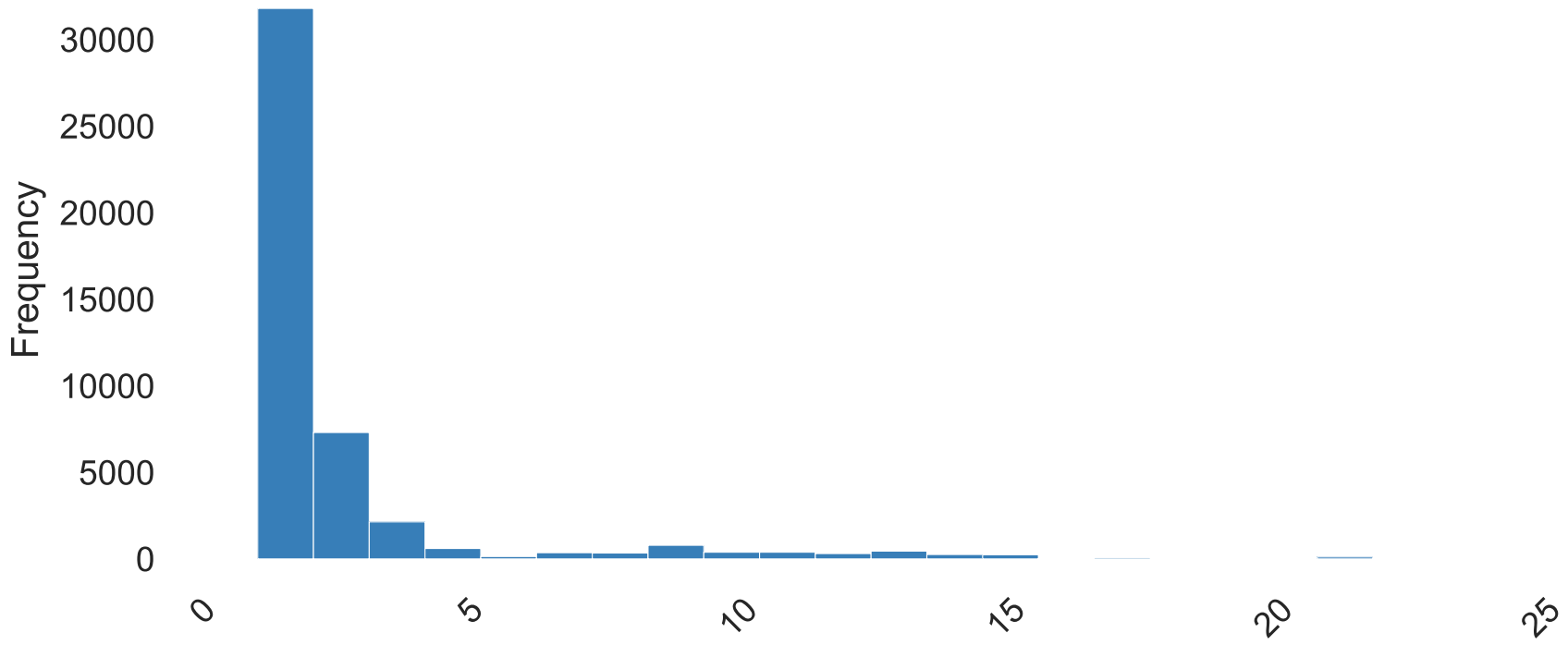
Sample

1st row	L5
2nd row	H6
3rd row	EH4
4th row	Acapulcoite
5th row	L6

Common Values

Value	Count	Frequency (%)
L6	8285	18.1%
H5	7142	15.6%
L5	4796	10.5%
H6	4528	9.9%
H4	4211	9.2%
LL5	2766	6.1%
LL6	2043	4.5%
L4	1253	2.7%
H4/5	428	0.9%
CM2	416	0.9%
Other values (456)	9848	21.5%

Length



Histogram of lengths of the category

Value	Count	Frequency (%)
l6	8339	17.6%
h5	7164	15.1%
l5	4817	10.2%
h6	4529	9.6%
h4	4223	8.9%
ll5	2766	5.8%
ll6	2046	4.3%
l4	1256	2.7%
iron	1070	2.3%
h4/5	428	0.9%
Other values (434)	10707	22.6%

Most occurring characters

Value	Count	Frequency (%)
L	28461	20.4%
H	18392	13.2%
5	16417	11.8%
6	16128	11.6%
4	6928	5.0%
e	3971	2.8%
i	3833	2.7%
r	3648	2.6%
t	3326	2.4%
3	3277	2.3%
Other values (52)	35166	25.2%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	57780	41.4%
Decimal Number	44109	31.6%
Lowercase Letter	29916	21.4%
Other Punctuation	3293	2.4%
Dash Punctuation	1834	1.3%
Space Separator	1747	1.3%
Math Symbol	320	0.2%
Open Punctuation	274	0.2%
Close Punctuation	274	0.2%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	3971	13.3%
i	3833	12.8%
r	3648	12.2%
t	3326	11.1%
n	2520	8.4%
o	2457	8.2%
c	1765	5.9%
u	1468	4.9%
a	1408	4.7%
l	1015	3.4%
Other values (12)	4505	15.1%

Uppercase Letter

Value	Count	Frequency (%)
L	28461	49.3%
H	18392	31.8%
I	2753	4.8%
C	1785	3.1%
E	1259	2.2%
A	984	1.7%
M	913	1.6%
B	754	1.3%
O	542	0.9%
V	350	0.6%
Other values (10)	1587	2.7%

Decimal Number

Value	Count	Frequency (%)
5	16417	37.2%
6	16128	36.6%
4	6928	15.7%
3	3277	7.4%
2	646	1.5%
7	251	0.6%
8	216	0.5%
9	111	0.3%
1	100	0.2%
0	35	0.1%

Other Punctuation

Value	Count	Frequency (%)
/	1174	35.7%
.	1064	32.3%
,	1031	31.3%

Value	Count	Frequency (%)
?	24	0.7%

Math Symbol

Value	Count	Frequency (%)
~	319	99.7%
<	1	0.3%

Dash Punctuation

Value	Count	Frequency (%)
-	1834	100.0%

Space Separator

Value	Count	Frequency (%)
	1747	100.0%

Open Punctuation

Value	Count	Frequency (%)
(	274	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	274	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	87696	62.8%
Common	51851	37.2%

Most frequent character per script

Latin

Value	Count	Frequency (%)
L	28461	32.5%
H	18392	21.0%
e	3971	4.5%
i	3833	4.4%
r	3648	4.2%
t	3326	3.8%
l	2753	3.1%
n	2520	2.9%
o	2457	2.8%
C	1785	2.0%
Other values (32)	16550	18.9%

Common

Value	Count	Frequency (%)
5	16417	31.7%
6	16128	31.1%
4	6928	13.4%
3	3277	6.3%
-	1834	3.5%
	1747	3.4%
/	1174	2.3%

Value	Count	Frequency (%)
.	1064	2.1%
,	1031	2.0%
2	646	1.2%
Other values (10)	1605	3.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	139547	100.0%

Most frequent character per block

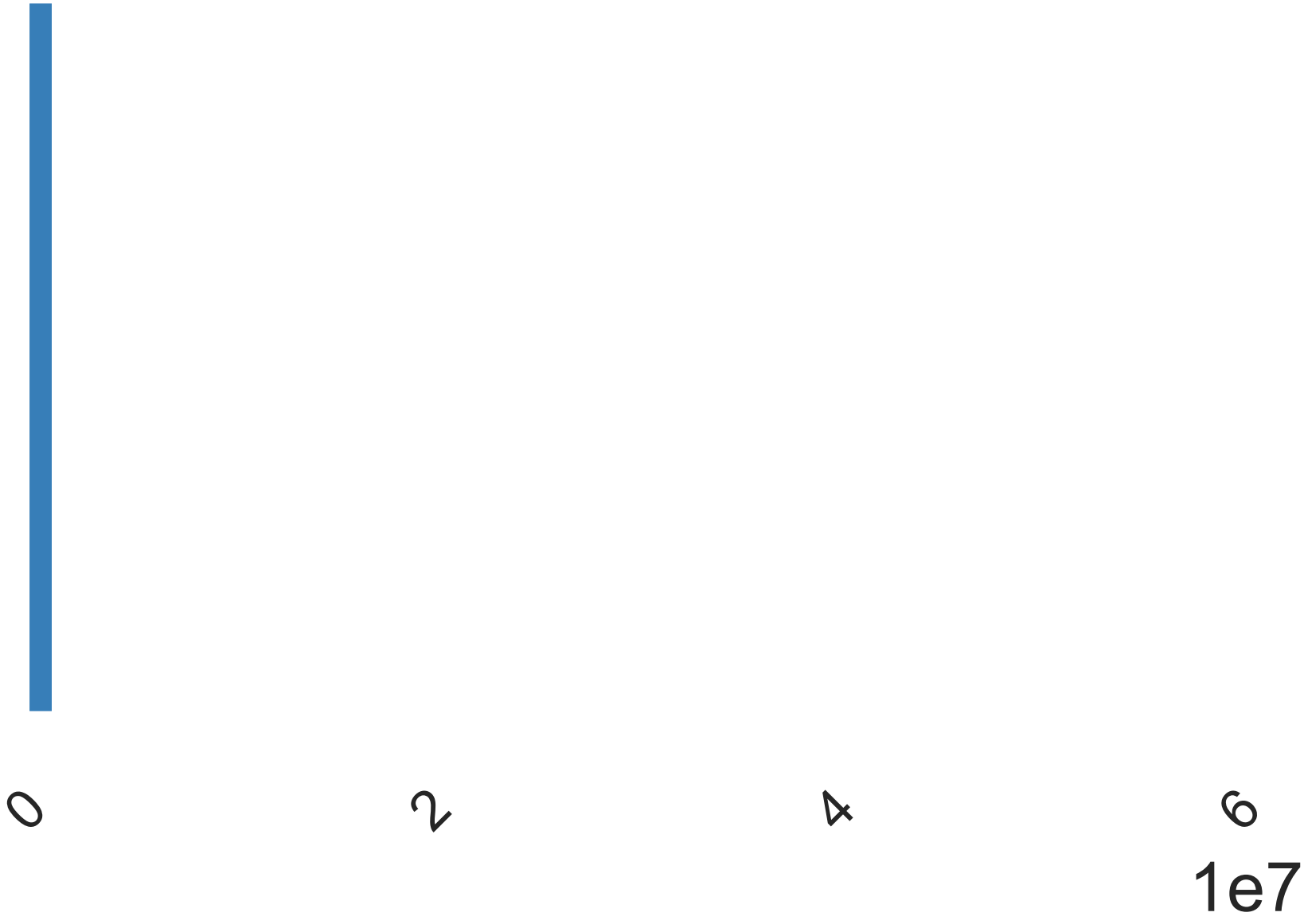
ASCII

Value	Count	Frequency (%)
L	28461	20.4%
H	18392	13.2%
5	16417	11.8%
6	16128	11.6%
4	6928	5.0%
e	3971	2.8%
i	3833	2.7%
r	3648	2.6%
t	3326	2.4%
3	3277	2.3%
Other values (52)	35166	25.2%

mass (g)
Real number (ℝ)

Distinct	12576
Distinct (%)	27.6%
Missing	131
Missing (%)	0.3%
Infinite	0
Infinite (%)	0.0%
Mean	13278.079

Minimum	0
Maximum	60000000
Zeros	19
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	357.3 KiB

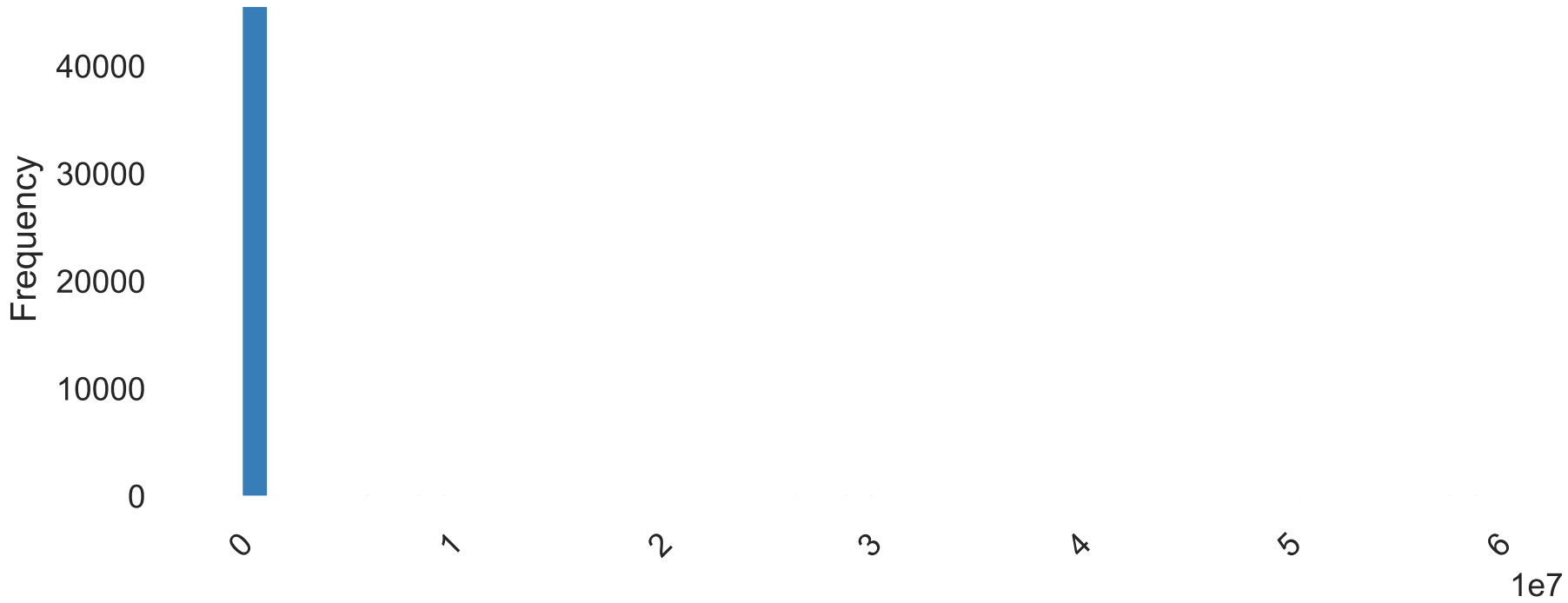

0
2
4
6
1e7

Quantile statistics

Minimum	0
5-th percentile	1.1
Q1	7.2
median	32.6
Q3	202.6
95-th percentile	4000
Maximum	60000000
Range	60000000
Interquartile range (IQR)	195.4

Descriptive statistics

Standard deviation	574988.88
Coefficient of variation (CV)	43.303621
Kurtosis	6796.9162
Mean	13278.079
Median Absolute Deviation (MAD)	30.5
Skewness	76.910117
Sum	6.0528121×10^{8}
Variance	3.3061221×10^{11}
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
1.3	171	0.4%
1.2	140	0.3%
1.4	138	0.3%
2.1	130	0.3%
2.4	126	0.3%
1.6	120	0.3%
0.5	119	0.3%
1.1	116	0.3%
3.8	114	0.2%
0.7	111	0.2%
Other values (12566)	44300	96.9%
(Missing)	131	0.3%

Value	Count	Frequency (%)
0	19	< 0.1%
0.01	2	< 0.1%
0.013	1	< 0.1%
0.02	1	< 0.1%
0.03	1	< 0.1%
0.04	1	< 0.1%
0.05	1	< 0.1%
0.06	1	< 0.1%
0.07	3	< 0.1%
0.08	2	< 0.1%

Value	Count	Frequency (%)
60000000	1	< 0.1%
58200000	1	< 0.1%
50000000	1	< 0.1%
30000000	1	< 0.1%
28000000	1	< 0.1%
26000000	1	< 0.1%
24300000	1	< 0.1%
24000000	1	< 0.1%
23000000	1	< 0.1%
22000000	1	< 0.1%

fall

Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	357.3 KiB

Length

Max length	5
Median length	5
Mean length	4.9757853
Min length	4

Characters and Unicode

Total characters	227473	
Distinct characters	7	
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%	

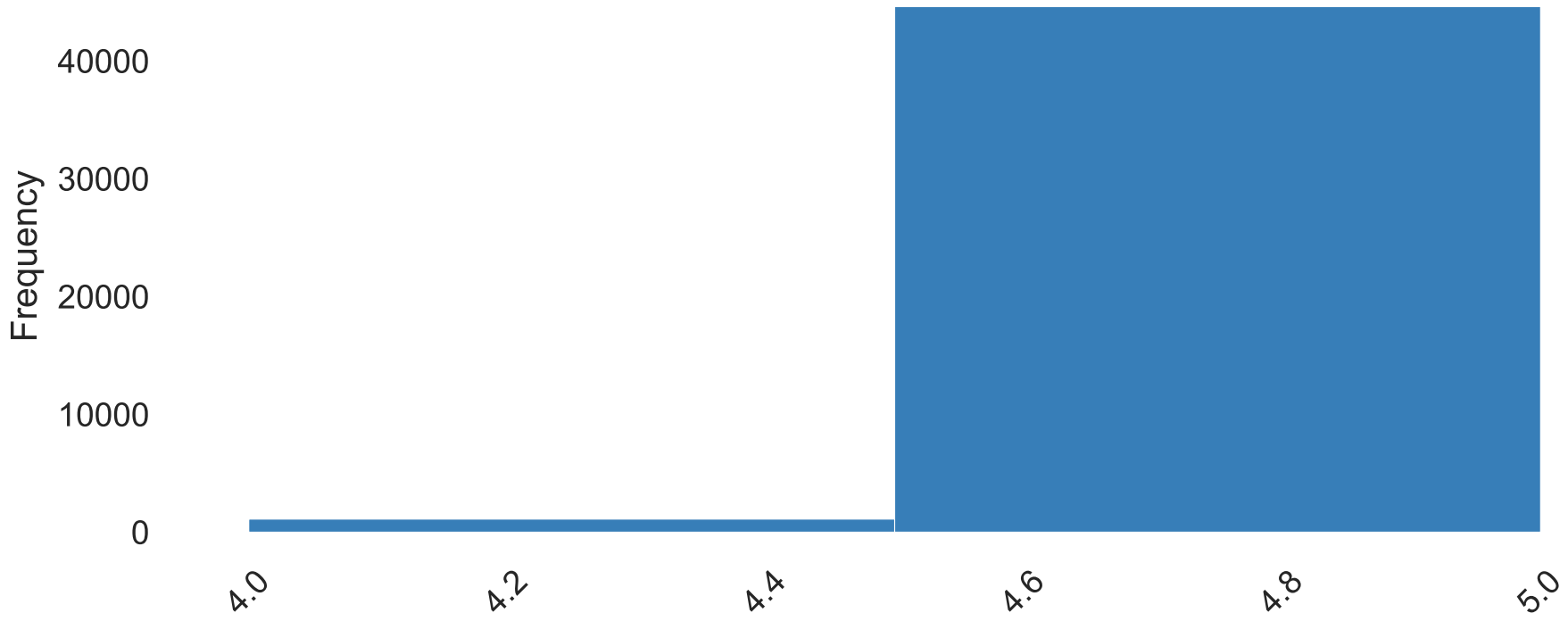
Sample

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

Common Values

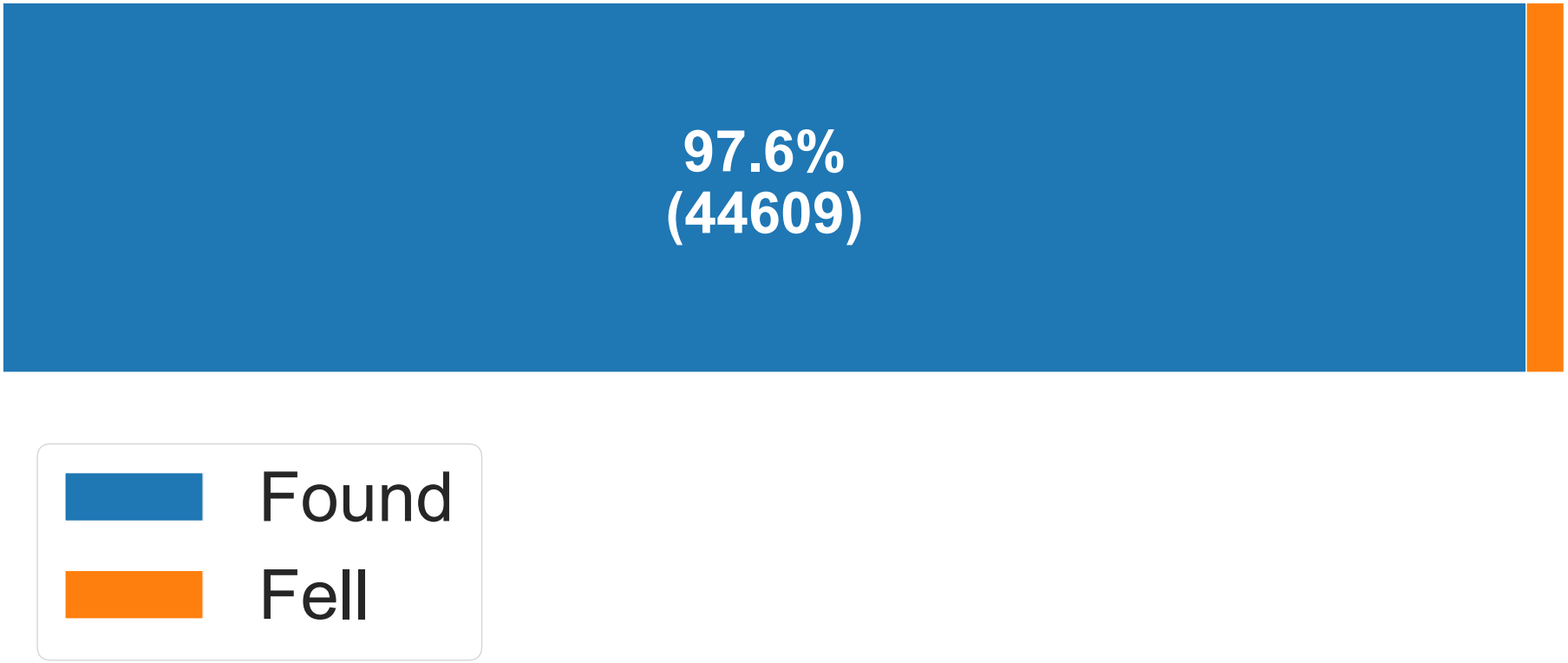
Value	Count	Frequency (%)
Found	44609	97.6%
Fell	1107	2.4%

Length



Histogram of lengths of the category

Common Values (Plot)


97.6%
(44609)
Found
Fell

Value	Count	Frequency (%)
found	44609	97.6%
fell	1107	2.4%

Most occurring characters

Value	Count	Frequency (%)
F	45716	20.1%
o	44609	19.6%
u	44609	19.6%
n	44609	19.6%
d	44609	19.6%
l	2214	1.0%
e	1107	0.5%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	181757	79.9%
Uppercase Letter	45716	20.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
o	44609	24.5%
u	44609	24.5%
n	44609	24.5%
d	44609	24.5%
l	2214	1.2%
e	1107	0.6%

Uppercase Letter

Value	Count	Frequency (%)
F	45716	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	227473	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
F	45716	20.1%
o	44609	19.6%
u	44609	19.6%
n	44609	19.6%
d	44609	19.6%
l	2214	1.0%
e	1107	0.5%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	227473	100.0%

Most frequent character per block

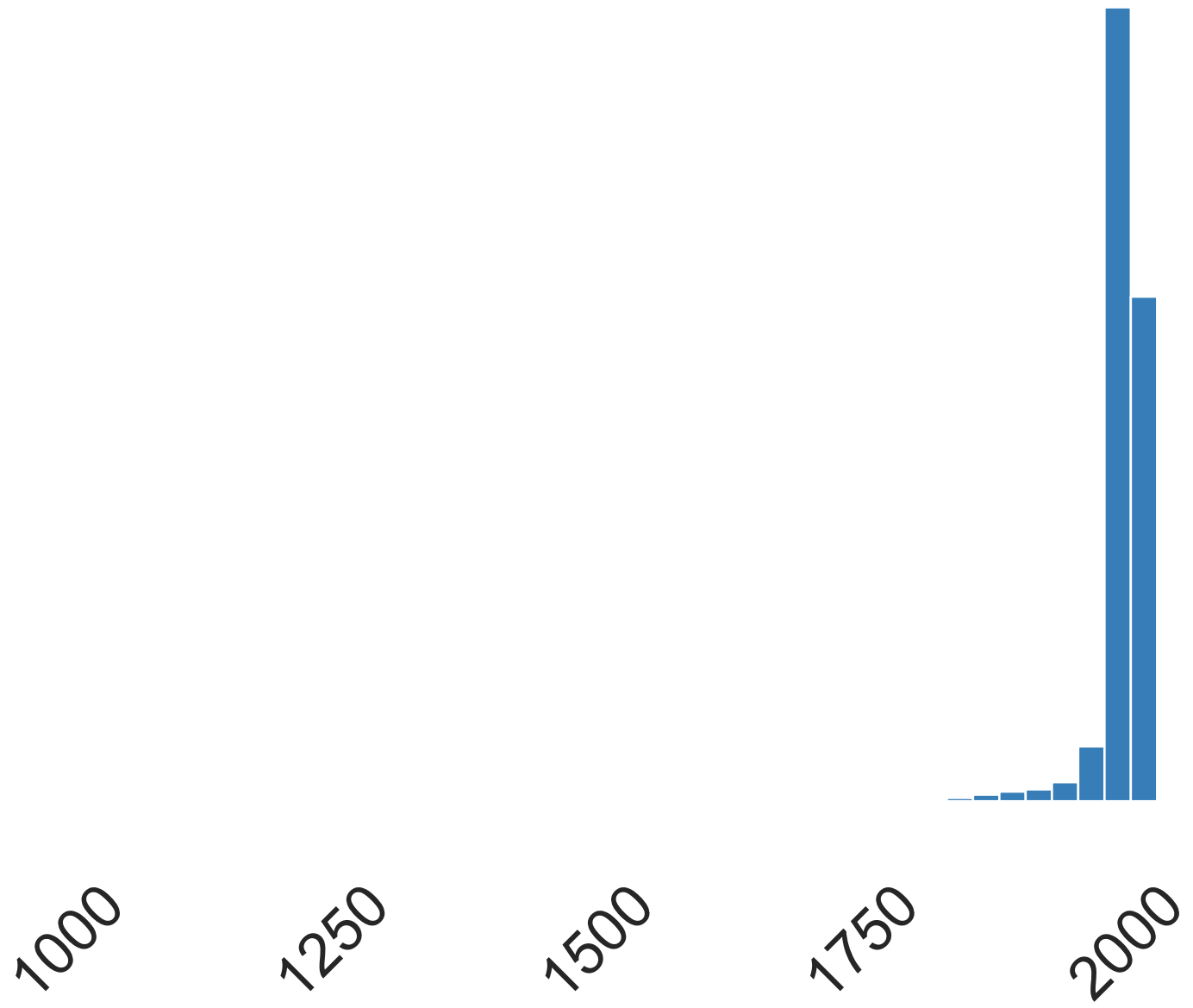
ASCII

Value	Count	Frequency (%)
F	45716	20.1%
o	44609	19.6%
u	44609	19.6%
n	44609	19.6%
d	44609	19.6%
l	2214	1.0%
e	1107	0.5%

year
Real number (ℝ)

Distinct	265
Distinct (%)	0.6%
Missing	291
Missing (%)	0.6%
Infinite	0
Infinite (%)	0.0%
Mean	1991.8288

Minimum	860
Maximum	2101
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	357.3 KiB

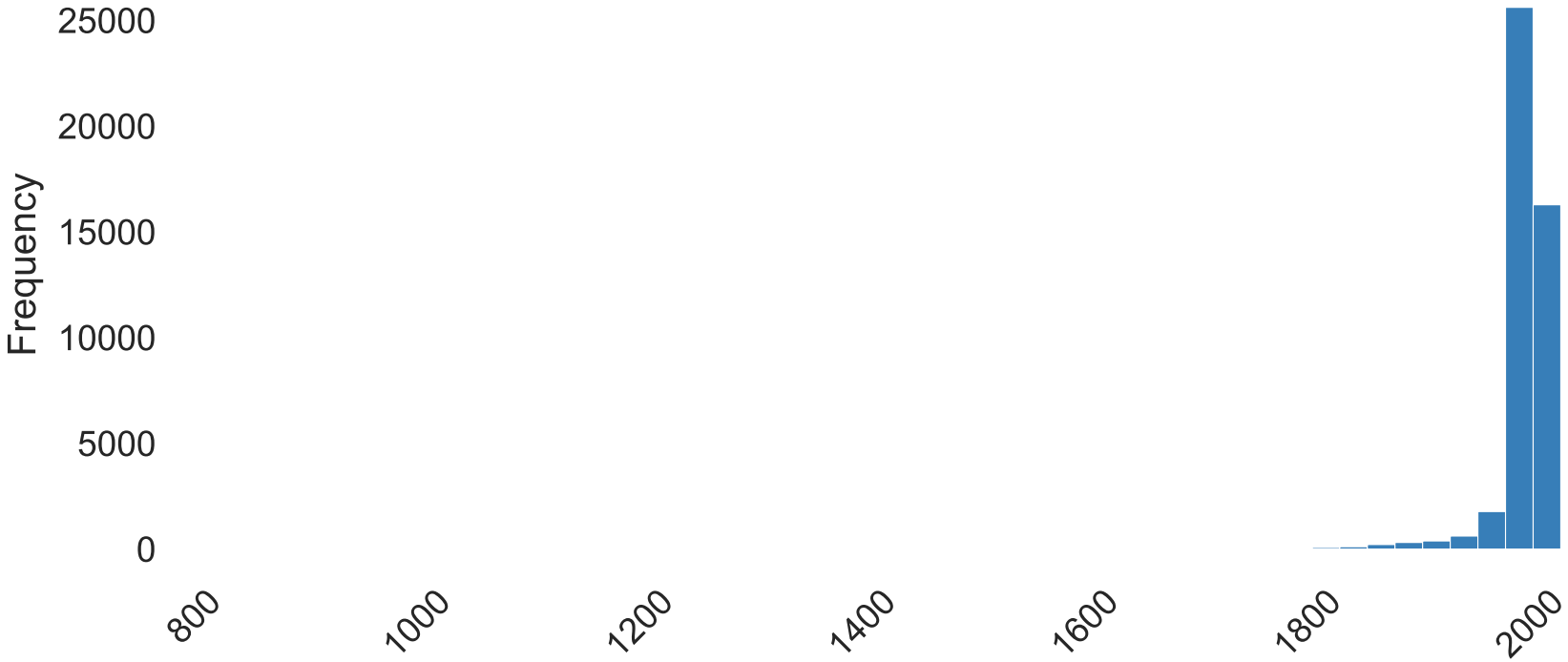

1000
1250
1500
1750
2000

Quantile statistics

Minimum	860
5-th percentile	1969
Q1	1987
median	1998
Q3	2003
95-th percentile	2009
Maximum	2101
Range	1241
Interquartile range (IQR)	16

Descriptive statistics

Standard deviation	25.052766
Coefficient of variation (CV)	0.012577771
Kurtosis	215.99969
Mean	1991.8288
Median Absolute Deviation (MAD)	8
Skewness	-8.8248153
Sum	90478824
Variance	627.64109
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
2003	3323	7.3%
1979	3046	6.7%
1998	2697	5.9%
2006	2456	5.4%
1988	2296	5.0%
2002	2078	4.5%
2004	1940	4.2%
2000	1792	3.9%
1997	1696	3.7%
1999	1691	3.7%
Other values (255)	22410	49.0%

Value	Count	Frequency (%)
860	1	< 0.1%
920	1	< 0.1%
1399	1	< 0.1%
1490	1	< 0.1%
1491	1	< 0.1%
1495	1	< 0.1%
1519	1	< 0.1%
1575	1	< 0.1%
1583	1	< 0.1%
1600	1	< 0.1%

Value	Count	Frequency (%)
2101	1	< 0.1%
2013	11	< 0.1%
2012	234	0.5%
2011	713	1.6%
2010	1005	2.2%
2009	1497	3.3%
2008	957	2.1%
2007	1189	2.6%
2006	2456	5.4%
2005	875	1.9%

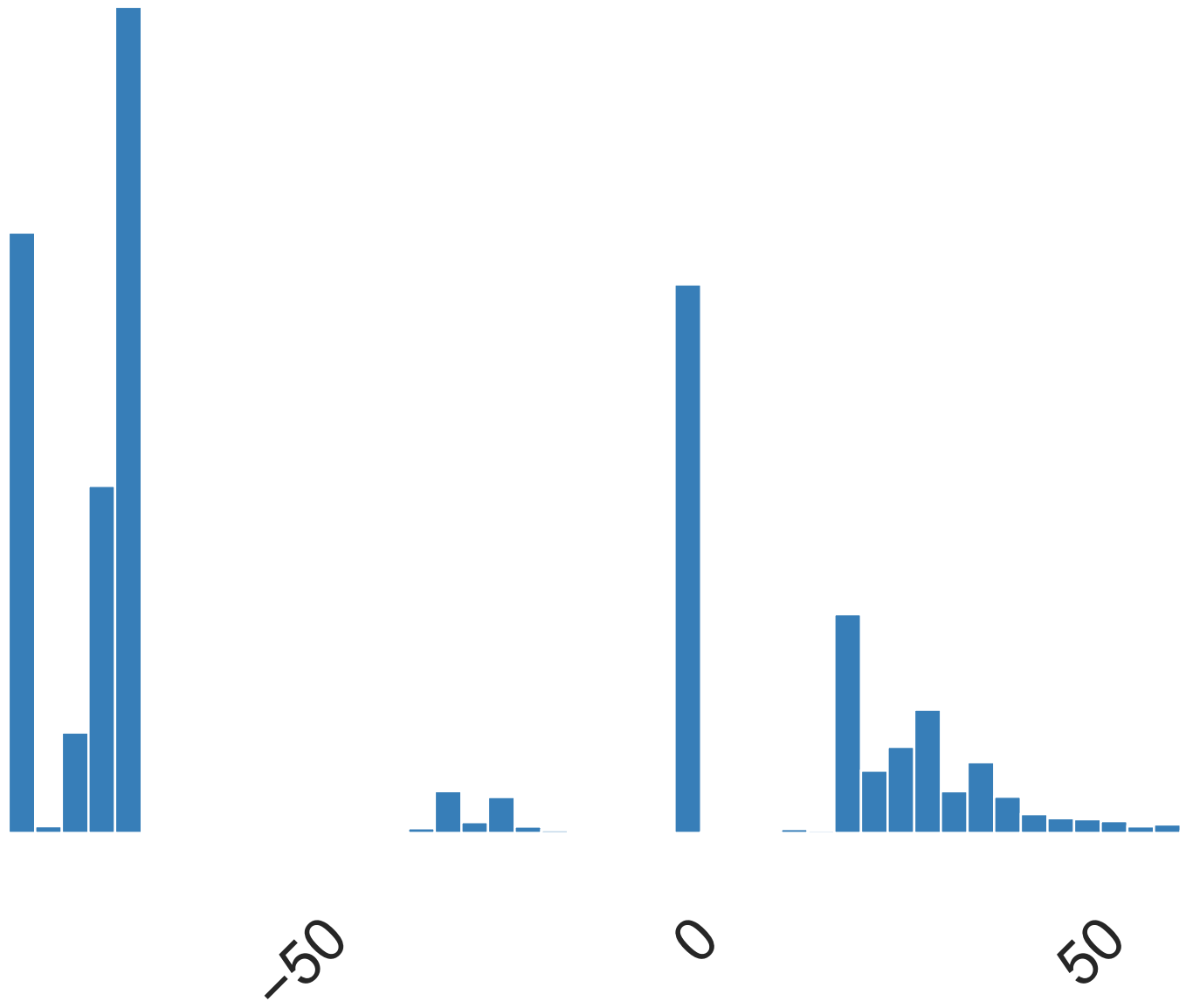
reclat

Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: reclong, Latitude, Longitude) MISSING ZEROS

Distinct	12738
Distinct (%)	33.2%
Missing	7315
Missing (%)	16.0%
Infinite	0
Infinite (%)	0.0%
Mean	-39.12258

Minimum	-87.36667
Maximum	81.16667
Zeros	6438
Zeros (%)	14.1%
Negative	23413
Negative (%)	51.2%
Memory size	357.3 KiB

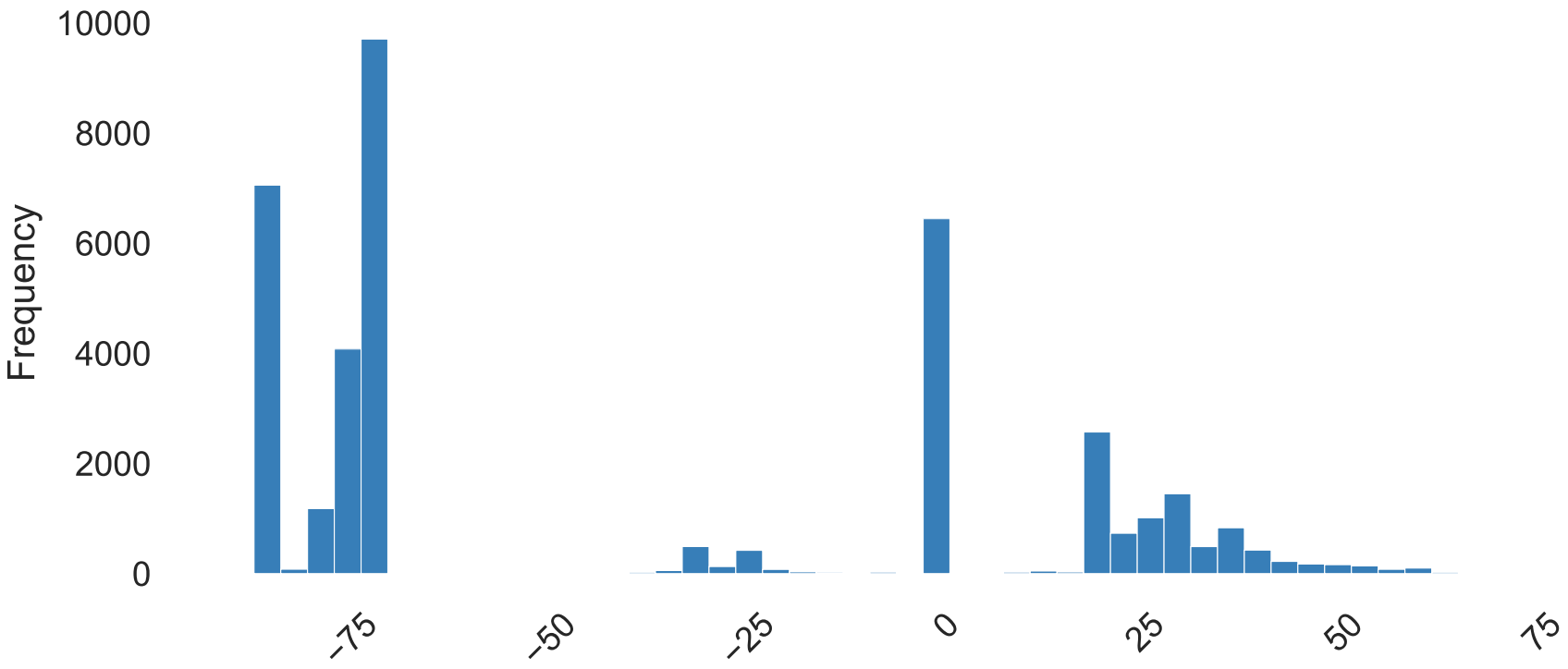

−50
0
50

Quantile statistics

Minimum	-87.36667
5-th percentile	-84.35516
Q1	-76.71424
median	-71.5
Q3	0
95-th percentile	34.49058
Maximum	81.16667
Range	168.53334
Interquartile range (IQR)	76.71424

Descriptive statistics

Standard deviation	46.378511
Coefficient of variation (CV)	-1.1854666
Kurtosis	-1.4768361
Mean	-39.12258
Median Absolute Deviation (MAD)	12.76421
Skewness	0.49161039
Sum	-1502346.2
Variance	2150.9663
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
0	6438	14.1%
-71.5	4761	10.4%
-84	3040	6.6%
-72	1506	3.3%
-79.68333	1130	2.5%
-76.71667	680	1.5%
-76.18333	539	1.2%
-84.21667	263	0.6%
-86.36667	226	0.5%
-86.71667	217	0.5%
Other values (12728)	19601	42.9%
(Missing)	7315	16.0%

Value	Count	Frequency (%)
-87.36667	4	< 0.1%
-87.03333	3	< 0.1%
-86.93333	3	< 0.1%
-86.71667	217	0.5%
-86.56667	17	< 0.1%
-86.54488	1	< 0.1%
-86.5379	1	< 0.1%
-86.53734	1	< 0.1%
-86.53725	1	< 0.1%
-86.53035	1	< 0.1%

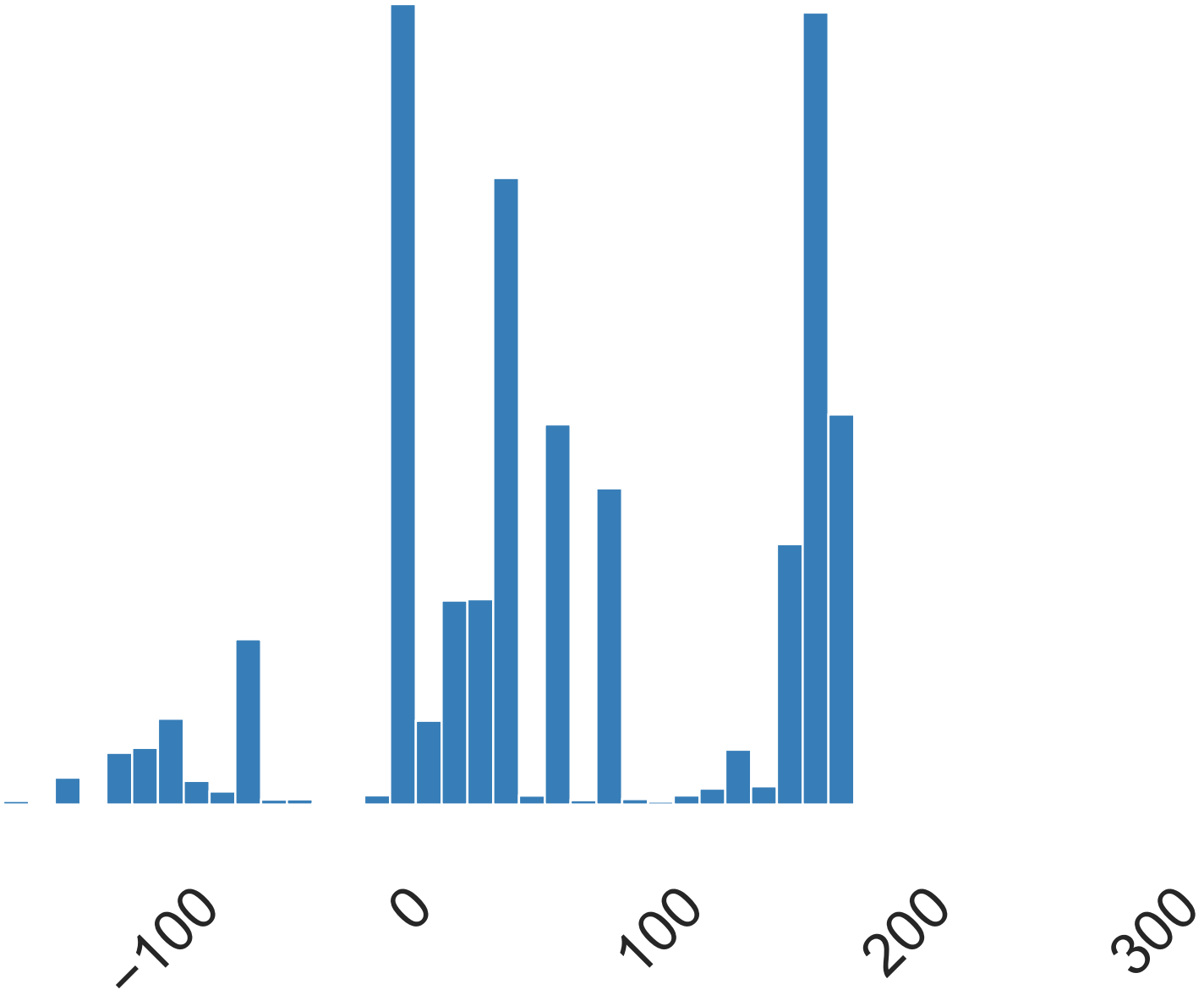
Value	Count	Frequency (%)
81.16667	1	< 0.1%
76.53333	1	< 0.1%
76.13333	1	< 0.1%
72.88333	1	< 0.1%
72.68333	1	< 0.1%
70.73333	1	< 0.1%
70	1	< 0.1%
69.1	1	< 0.1%
68	1	< 0.1%
67.88333	1	< 0.1%

reclong

Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: reclat, Latitude, Longitude) MISSING ZEROS

Distinct	14640
Distinct (%)	38.1%
Missing	7315
Missing (%)	16.0%
Infinite	0
Infinite (%)	0.0%
Mean	61.074319
Minimum	-165.43333
Maximum	354.47333
Zeros	6214
Zeros (%)	13.6%
Negative	4052
Negative (%)	8.9%
Memory size	357.3 KiB

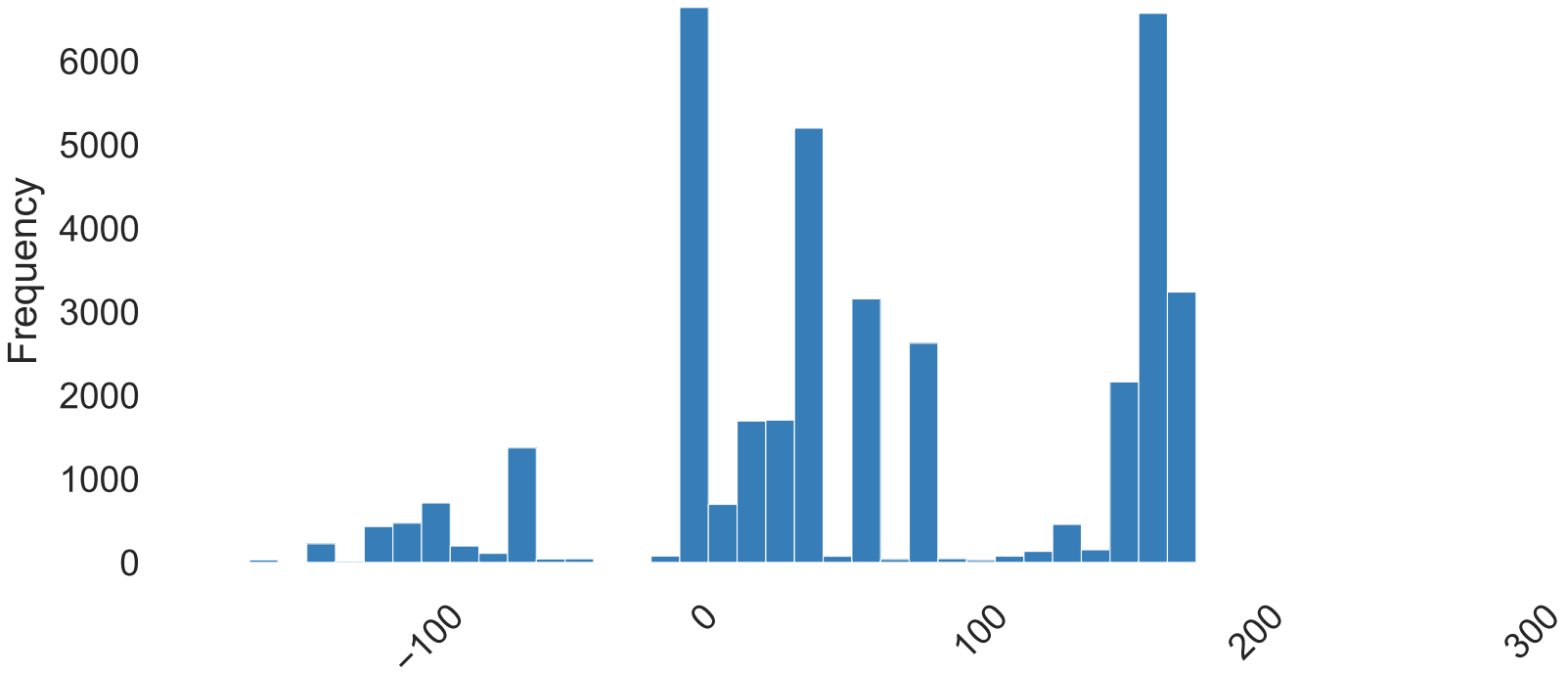

−100
0
100
200
300

Quantile statistics

Minimum	-165.43333
5-th percentile	-90.36556
Q1	0
median	35.66667
Q3	157.16667
95-th percentile	168
Maximum	354.47333
Range	519.90666
Interquartile range (IQR)	157.16667

Descriptive statistics

Standard deviation	80.647298
Coefficient of variation (CV)	1.3204781
Kurtosis	-0.7311811
Mean	61.074319
Median Absolute Deviation (MAD)	39.53972
Skewness	-0.17449637
Sum	2345314.9
Variance	6503.9867
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
0	6214	13.6%
35.66667	4985	10.9%
168	3040	6.6%
26	1506	3.3%
159.75	657	1.4%
159.66667	637	1.4%
157.16667	542	1.2%
155.75	473	1.0%
160.5	263	0.6%
-70	228	0.5%
Other values (14630)	19856	43.4%
(Missing)	7315	16.0%

Value	Count	Frequency (%)
-165.43333	9	< 0.1%
-165.11667	17	< 0.1%
-163.16667	1	< 0.1%
-162.55	1	< 0.1%
-157.86667	1	< 0.1%
-157.78333	1	< 0.1%
-149.5	4	< 0.1%
-148.55	2	< 0.1%
-148	3	< 0.1%
-146.26667	1	< 0.1%

Value	Count	Frequency (%)
354.47333	1	< 0.1%
178.2	1	< 0.1%
178.08333	1	< 0.1%
175.73028	1	< 0.1%
175.13333	1	< 0.1%
175	185	0.4%
174.50043	1	< 0.1%
174.4	1	< 0.1%
172.7	1	< 0.1%
172.6	1	< 0.1%

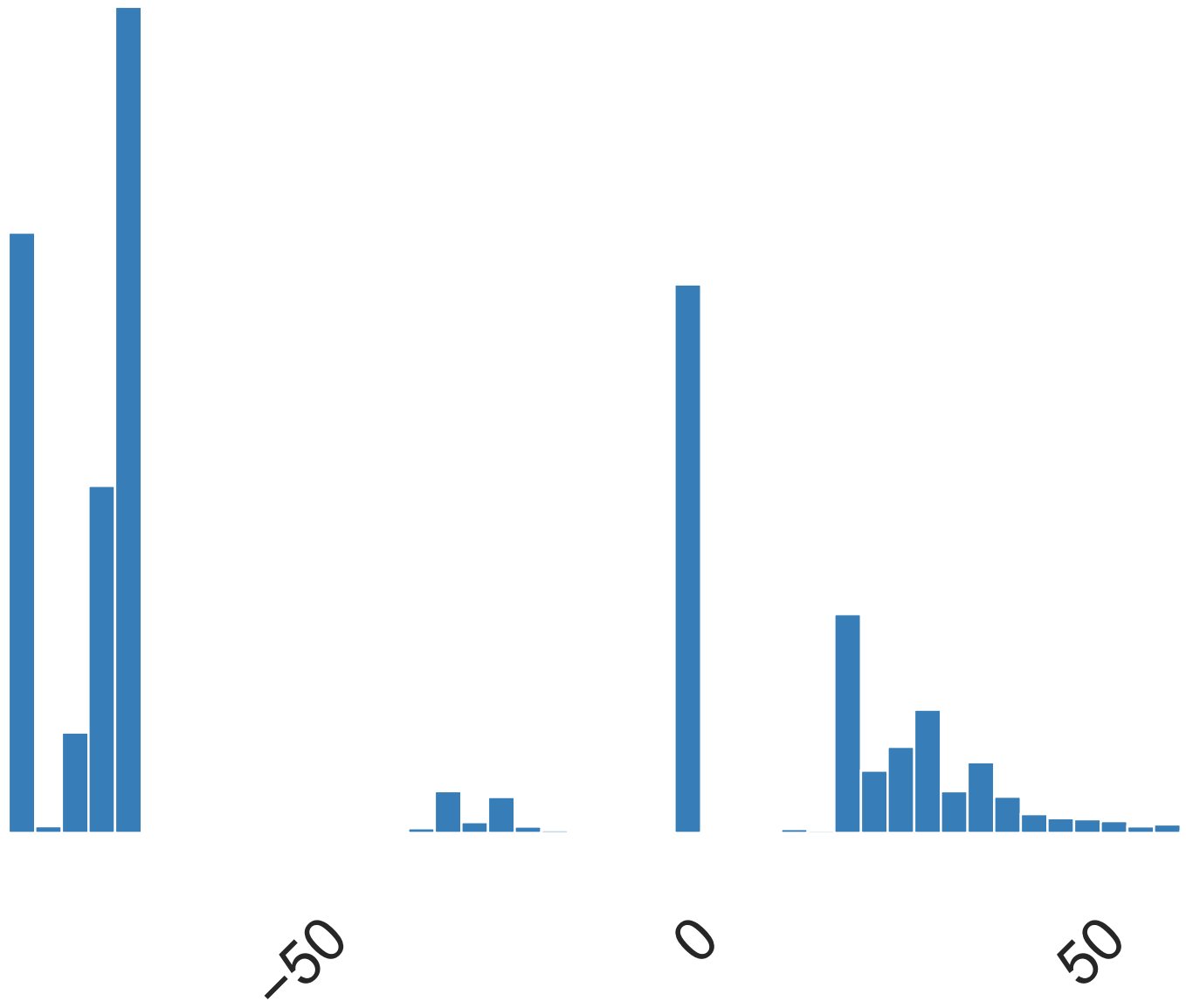
Latitude

Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: reclat, reclong, Longitude) MISSING ZEROS

Distinct	12738
Distinct (%)	33.2%
Missing	7315
Missing (%)	16.0%
Infinite	0
Infinite (%)	0.0%
Mean	-39.12258

Minimum	-87.36667
Maximum	81.16667
Zeros	6438
Zeros (%)	14.1%
Negative	23413
Negative (%)	51.2%
Memory size	357.3 KiB

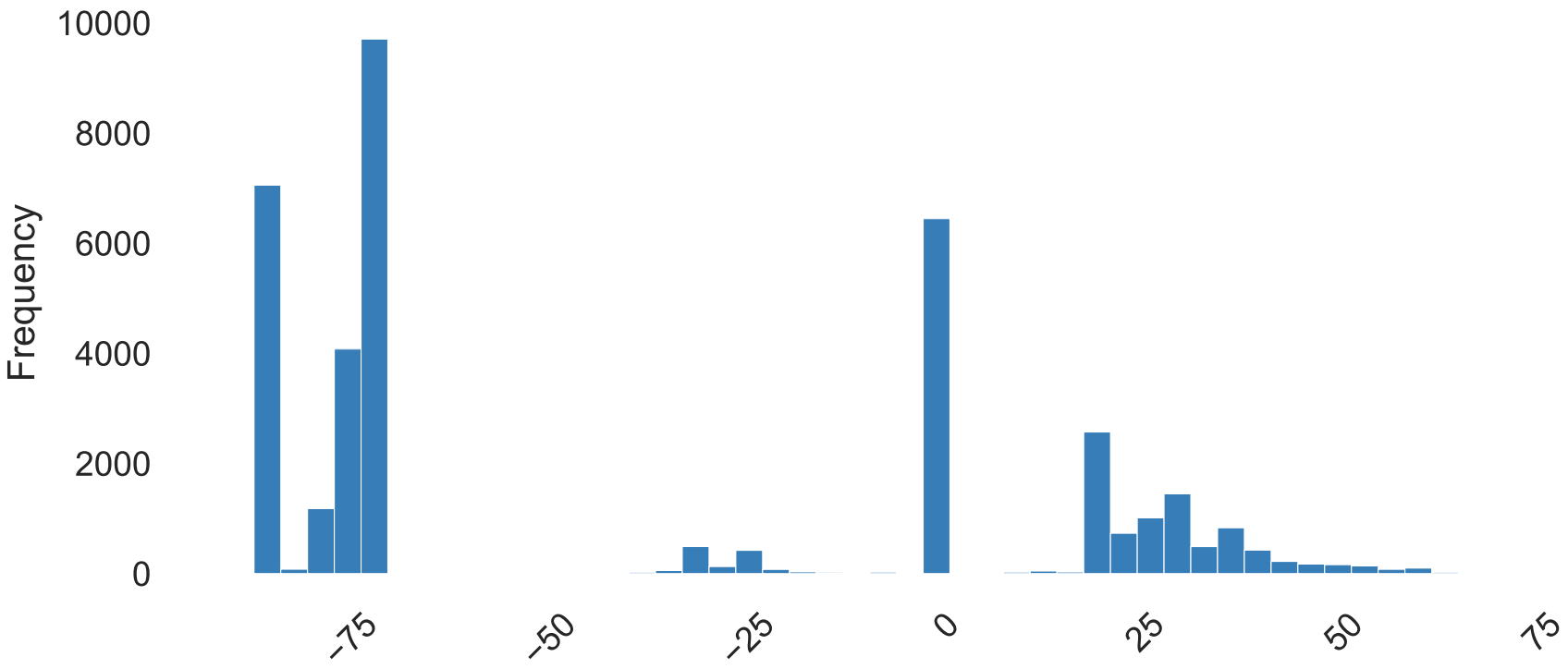

−50
0
50

Quantile statistics

Minimum	-87.36667
5-th percentile	-84.35516
Q1	-76.71424
median	-71.5
Q3	0
95-th percentile	34.49058
Maximum	81.16667
Range	168.53334
Interquartile range (IQR)	76.71424

Descriptive statistics

Standard deviation	46.378511
Coefficient of variation (CV)	-1.1854666
Kurtosis	-1.4768361
Mean	-39.12258
Median Absolute Deviation (MAD)	12.76421
Skewness	0.49161039
Sum	-1502346.2
Variance	2150.9663
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
0	6438	14.1%
-71.5	4761	10.4%
-84	3040	6.6%
-72	1506	3.3%
-79.68333	1130	2.5%
-76.71667	680	1.5%
-76.18333	539	1.2%
-84.21667	263	0.6%
-86.36667	226	0.5%
-86.71667	217	0.5%
Other values (12728)	19601	42.9%
(Missing)	7315	16.0%

Value	Count	Frequency (%)
-87.36667	4	< 0.1%
-87.03333	3	< 0.1%
-86.93333	3	< 0.1%
-86.71667	217	0.5%
-86.56667	17	< 0.1%
-86.54488	1	< 0.1%
-86.5379	1	< 0.1%
-86.53734	1	< 0.1%
-86.53725	1	< 0.1%
-86.53035	1	< 0.1%

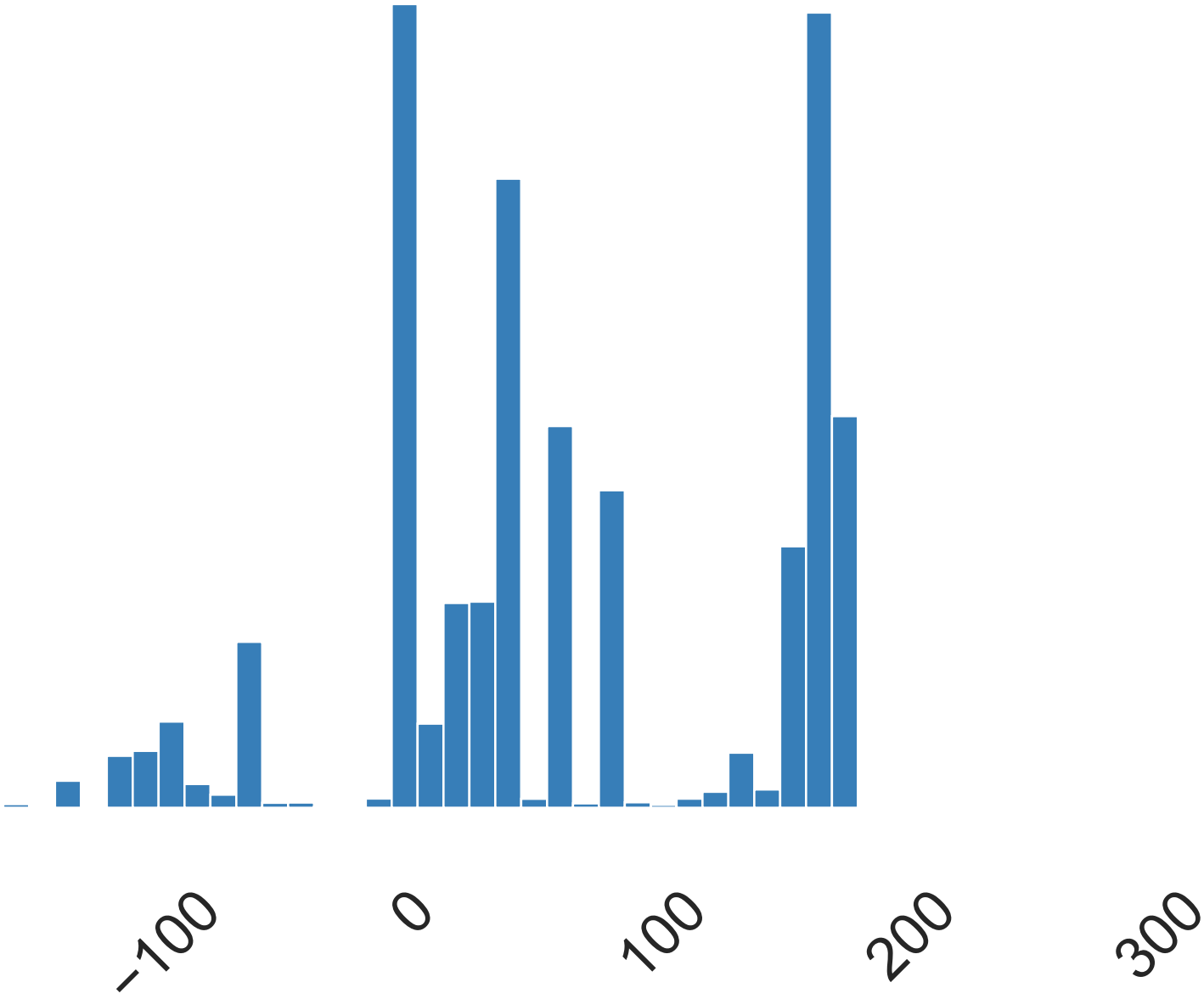
Value	Count	Frequency (%)
81.16667	1	< 0.1%
76.53333	1	< 0.1%
76.13333	1	< 0.1%
72.88333	1	< 0.1%
72.68333	1	< 0.1%
70.73333	1	< 0.1%
70	1	< 0.1%
69.1	1	< 0.1%
68	1	< 0.1%
67.88333	1	< 0.1%

Longitude

Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: reclat, reclong, Latitude) MISSING ZEROS

Distinct	14640
Distinct (%)	38.1%
Missing	7315
Missing (%)	16.0%
Infinite	0
Infinite (%)	0.0%
Mean	61.074319
Minimum	-165.43333
Maximum	354.47333
Zeros	6214
Zeros (%)	13.6%
Negative	4052
Negative (%)	8.9%
Memory size	357.3 KiB

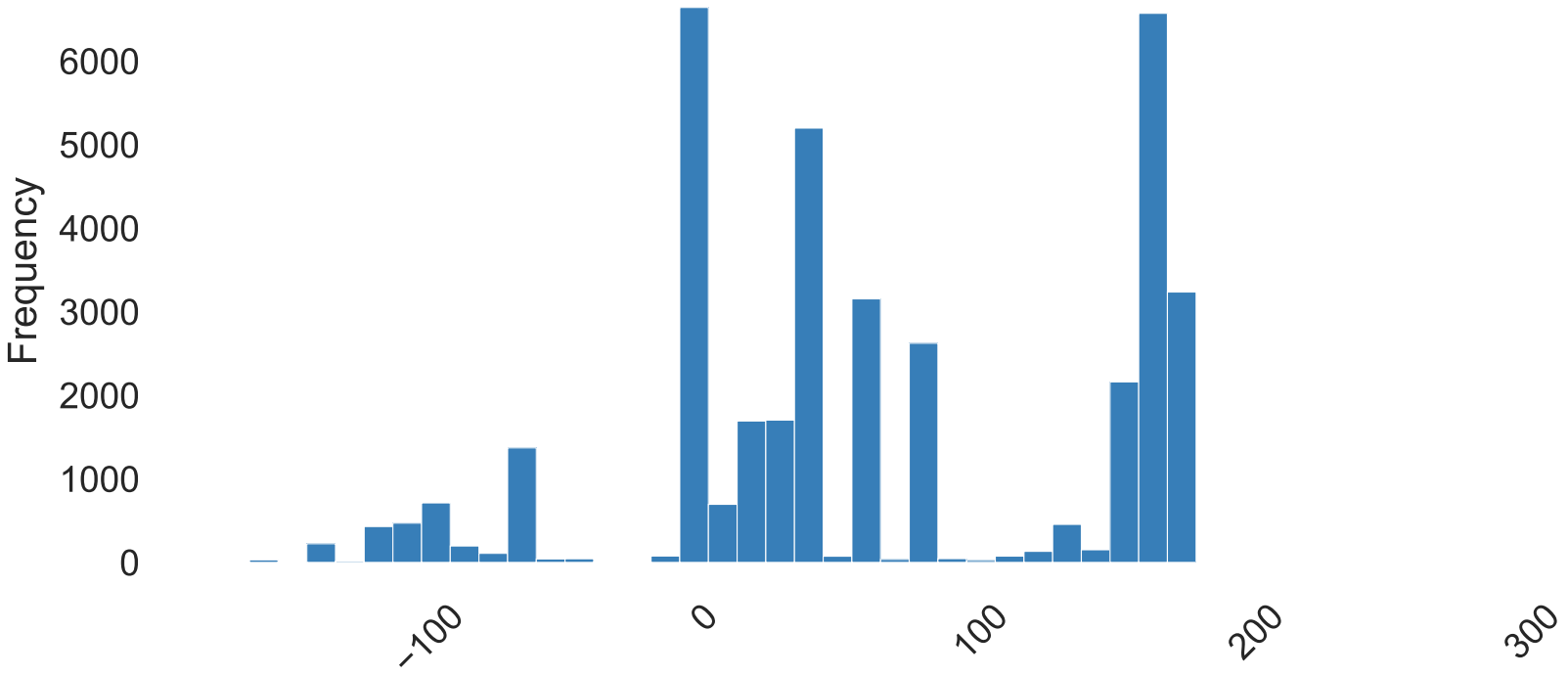

−100
0
100
200
300

Quantile statistics

Minimum	-165.43333
5-th percentile	-90.36556
Q1	0
median	35.66667
Q3	157.16667
95-th percentile	168
Maximum	354.47333
Range	519.90666
Interquartile range (IQR)	157.16667

Descriptive statistics

Standard deviation	80.647298
Coefficient of variation (CV)	1.3204781
Kurtosis	-0.7311811
Mean	61.074319
Median Absolute Deviation (MAD)	39.53972
Skewness	-0.17449637
Sum	2345314.9
Variance	6503.9867
Monotonicity	Not monotonic



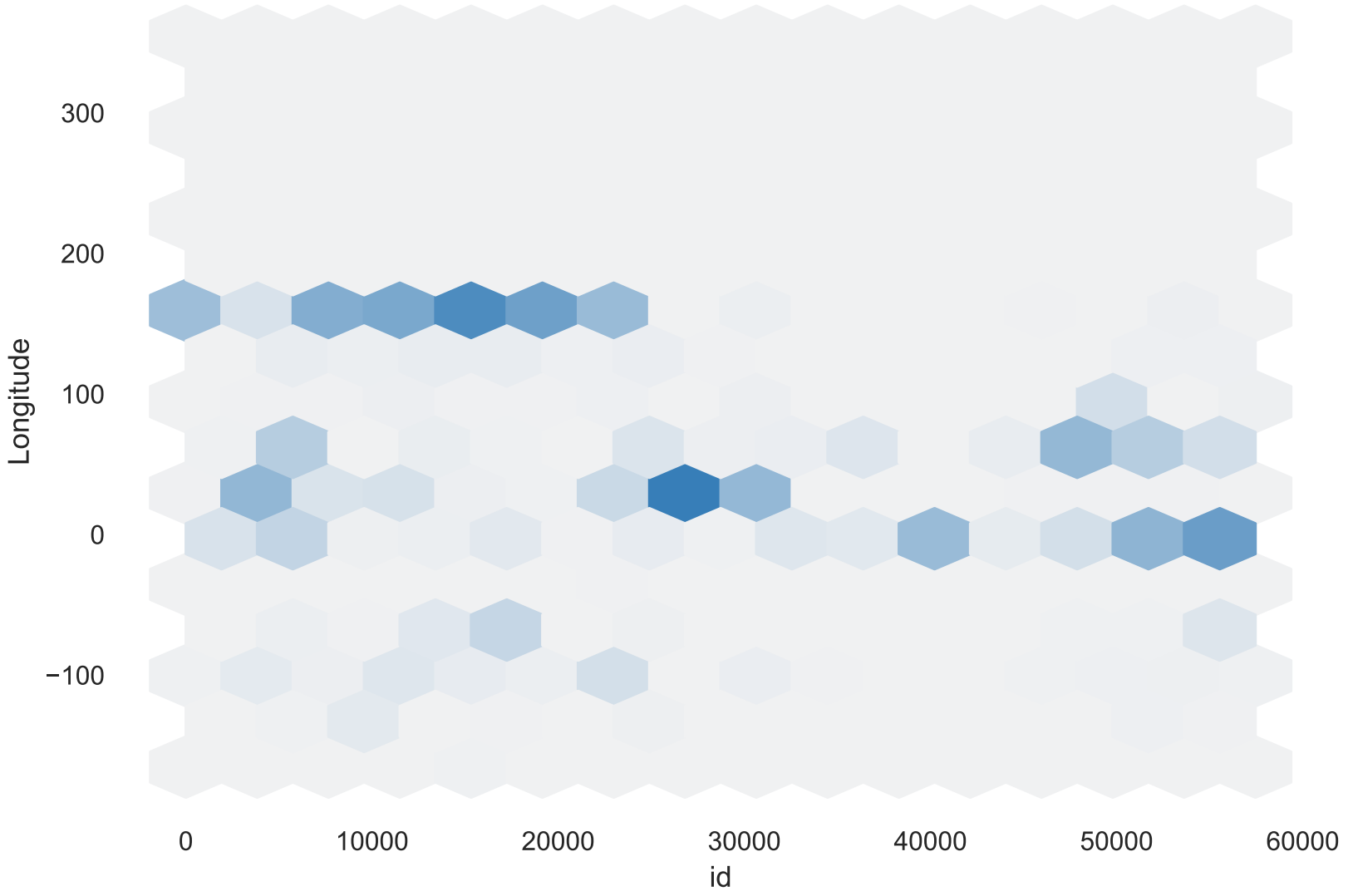
Histogram with fixed size bins (bins=50)

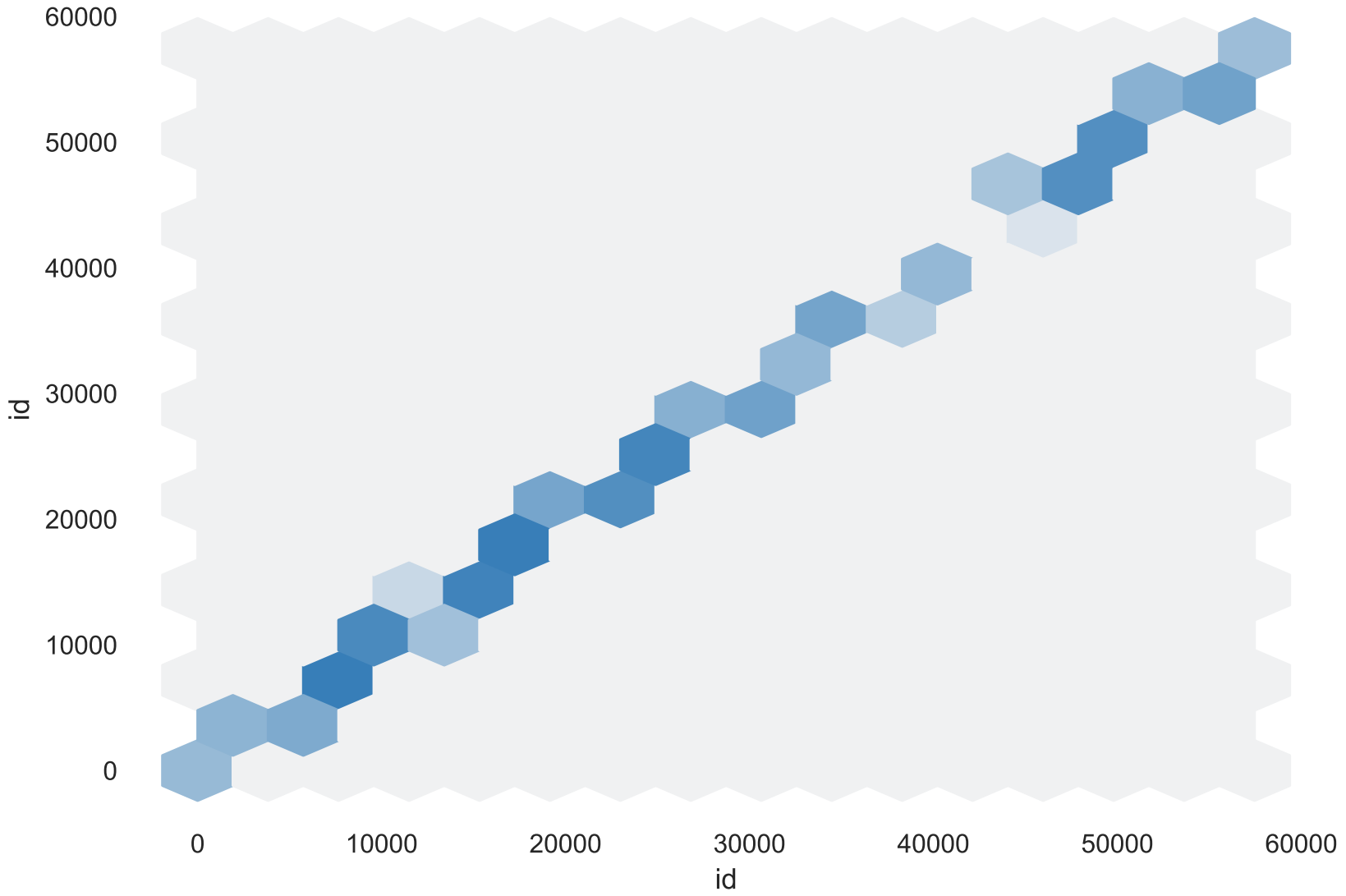
Value	Count	Frequency (%)
0	6214	13.6%
35.66667	4985	10.9%
168	3040	6.6%
26	1506	3.3%
159.75	657	1.4%
159.66667	637	1.4%
157.16667	542	1.2%
155.75	473	1.0%
160.5	263	0.6%
-70	228	0.5%
Other values (14630)	19856	43.4%
(Missing)	7315	16.0%

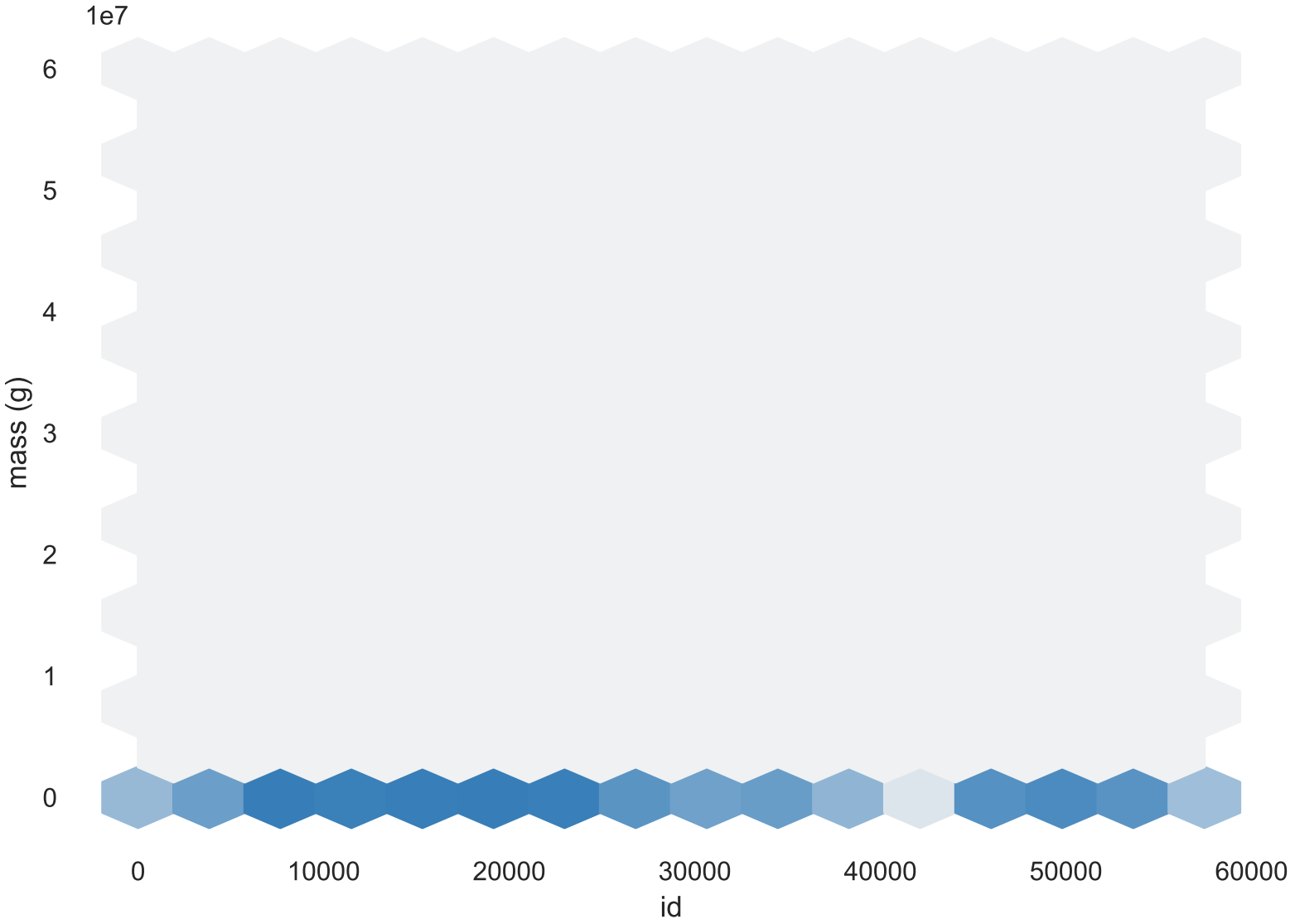
Value	Count	Frequency (%)
-165.43333	9	< 0.1%
-165.11667	17	< 0.1%
-163.16667	1	< 0.1%
-162.55	1	< 0.1%
-157.86667	1	< 0.1%
-157.78333	1	< 0.1%
-149.5	4	< 0.1%
-148.55	2	< 0.1%
-148	3	< 0.1%
-146.26667	1	< 0.1%

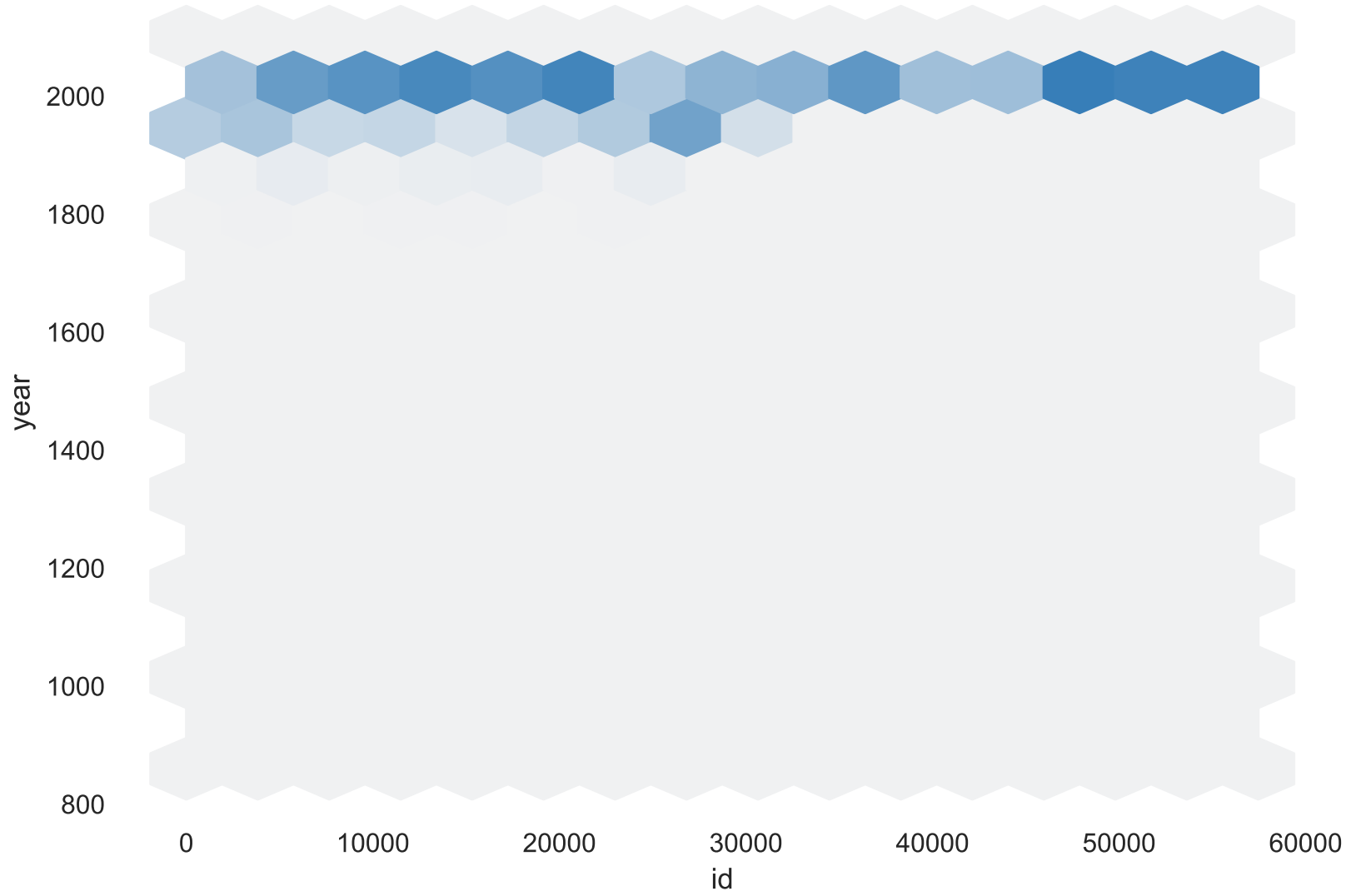
Value	Count	Frequency (%)
354.47333	1	< 0.1%
178.2	1	< 0.1%
178.08333	1	< 0.1%
175.73028	1	< 0.1%
175.13333	1	< 0.1%
175	185	0.4%
174.50043	1	< 0.1%
174.4	1	< 0.1%
172.7	1	< 0.1%
172.6	1	< 0.1%

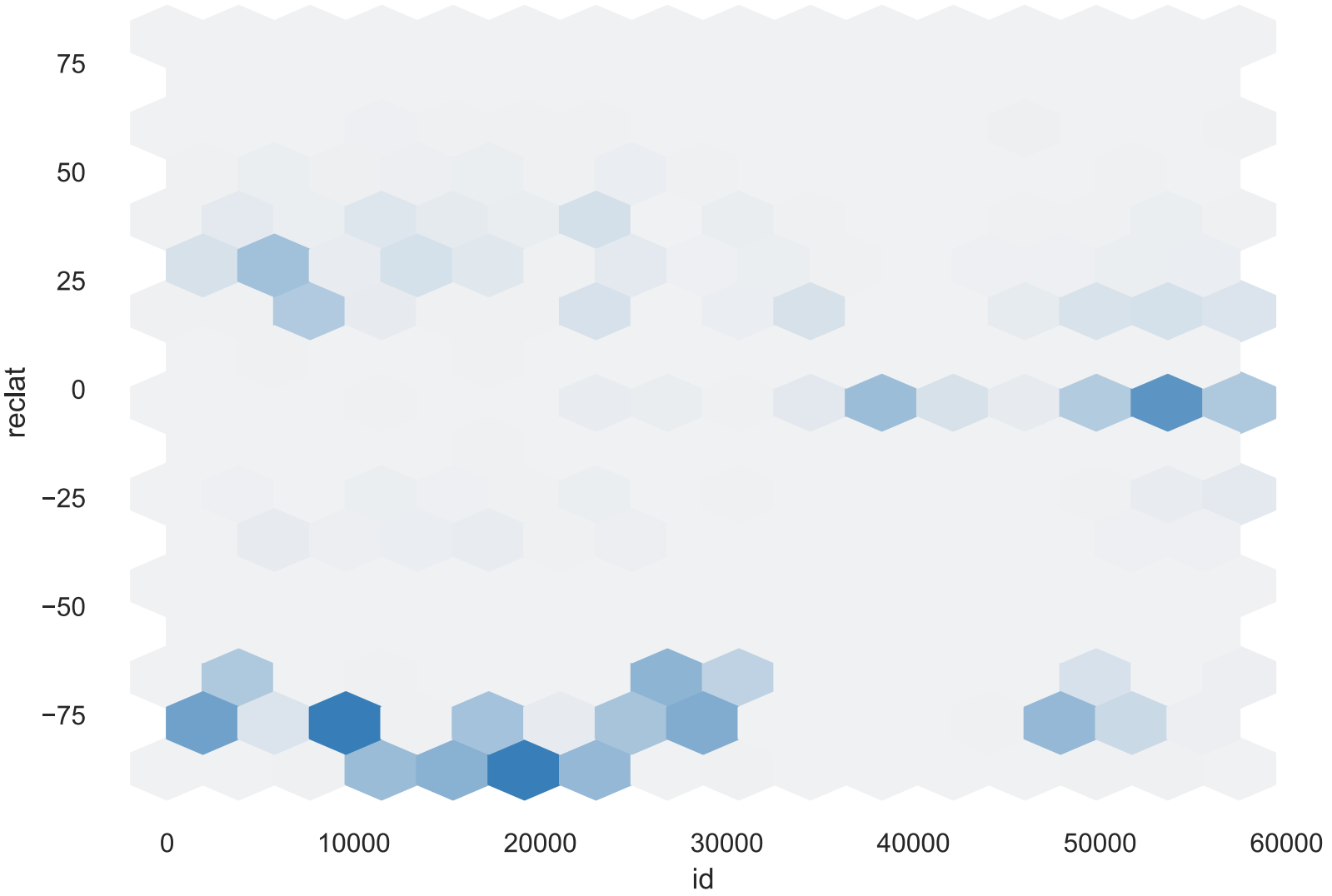
Interactions

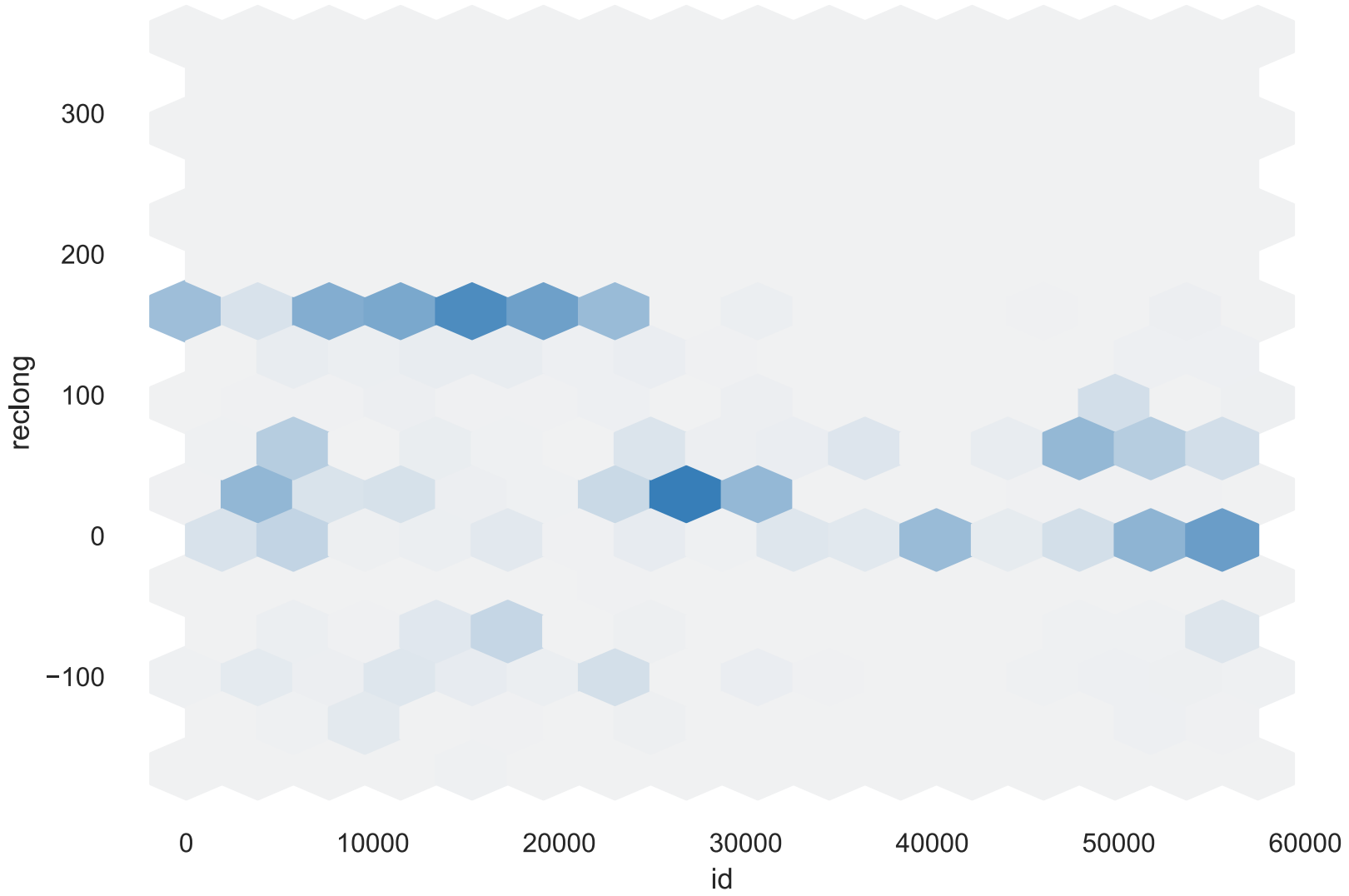

300
200
100
0
−100
Longitude
0
10000
20000
30000
40000
50000
60000
id

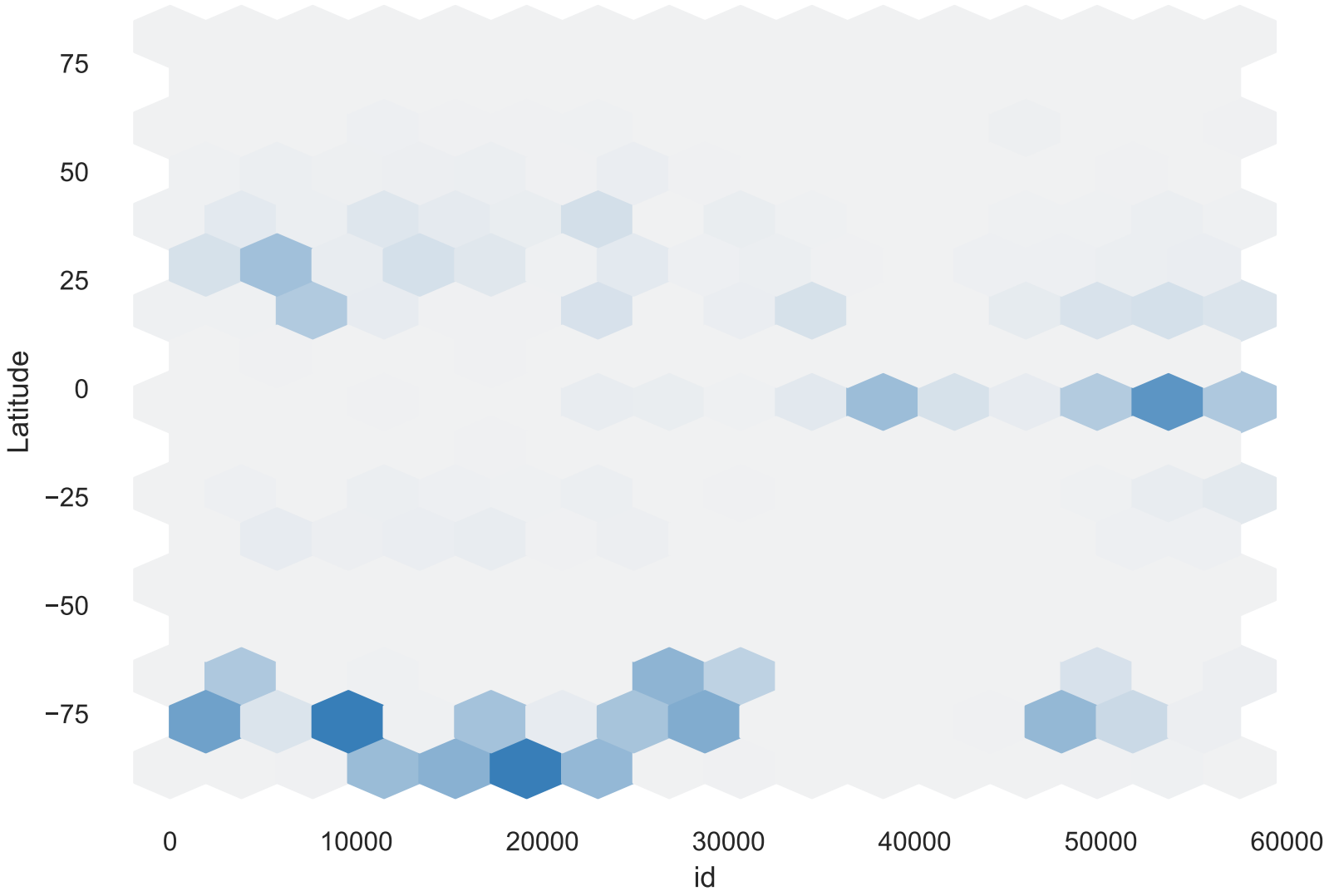

60000
50000
40000
30000
20000
10000
0
id
0
10000
20000
30000
40000
50000
60000
id

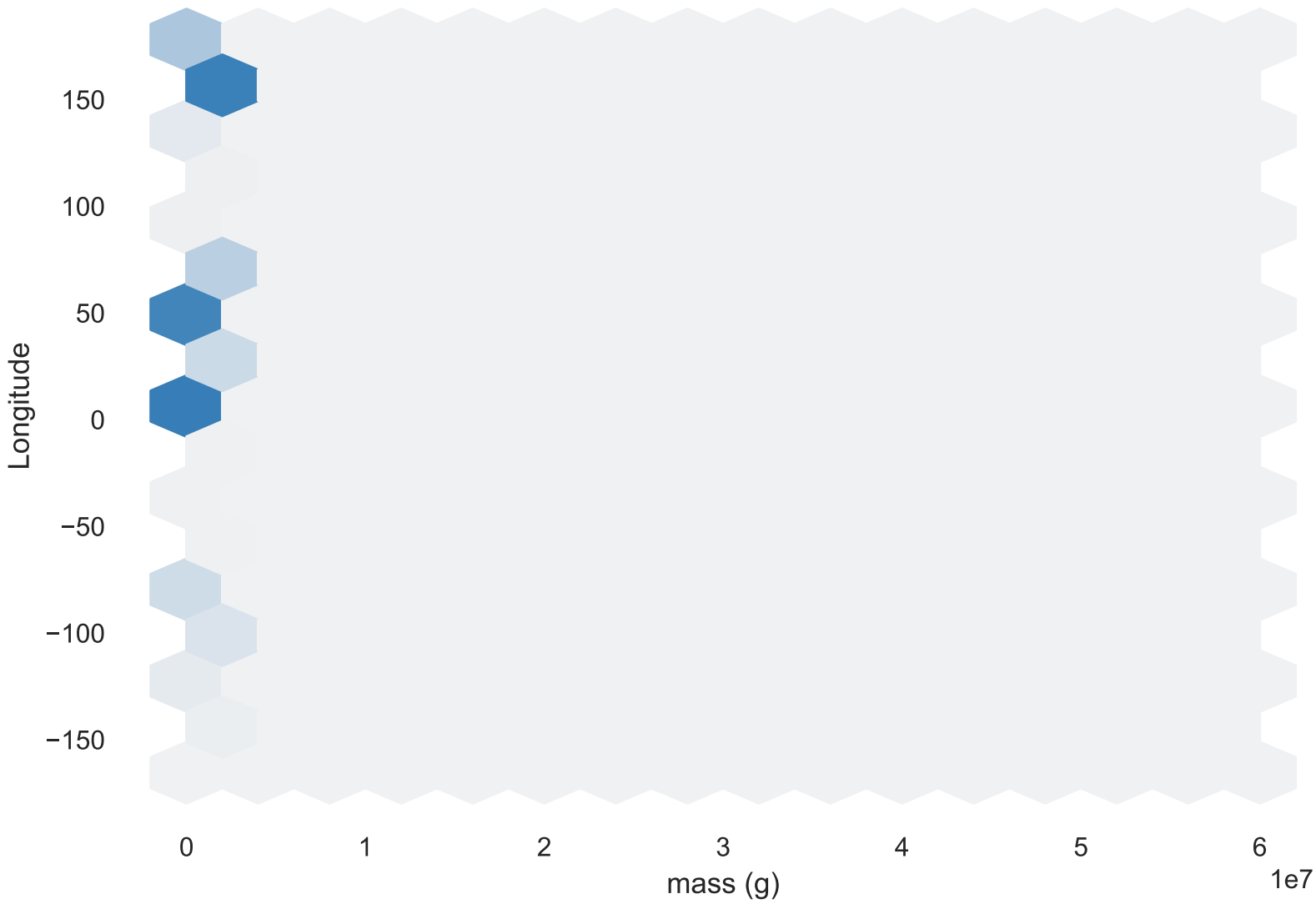

1e7
6
5
4
3
2
1
0
mass (g)
0
10000
20000
30000
40000
50000
60000
id

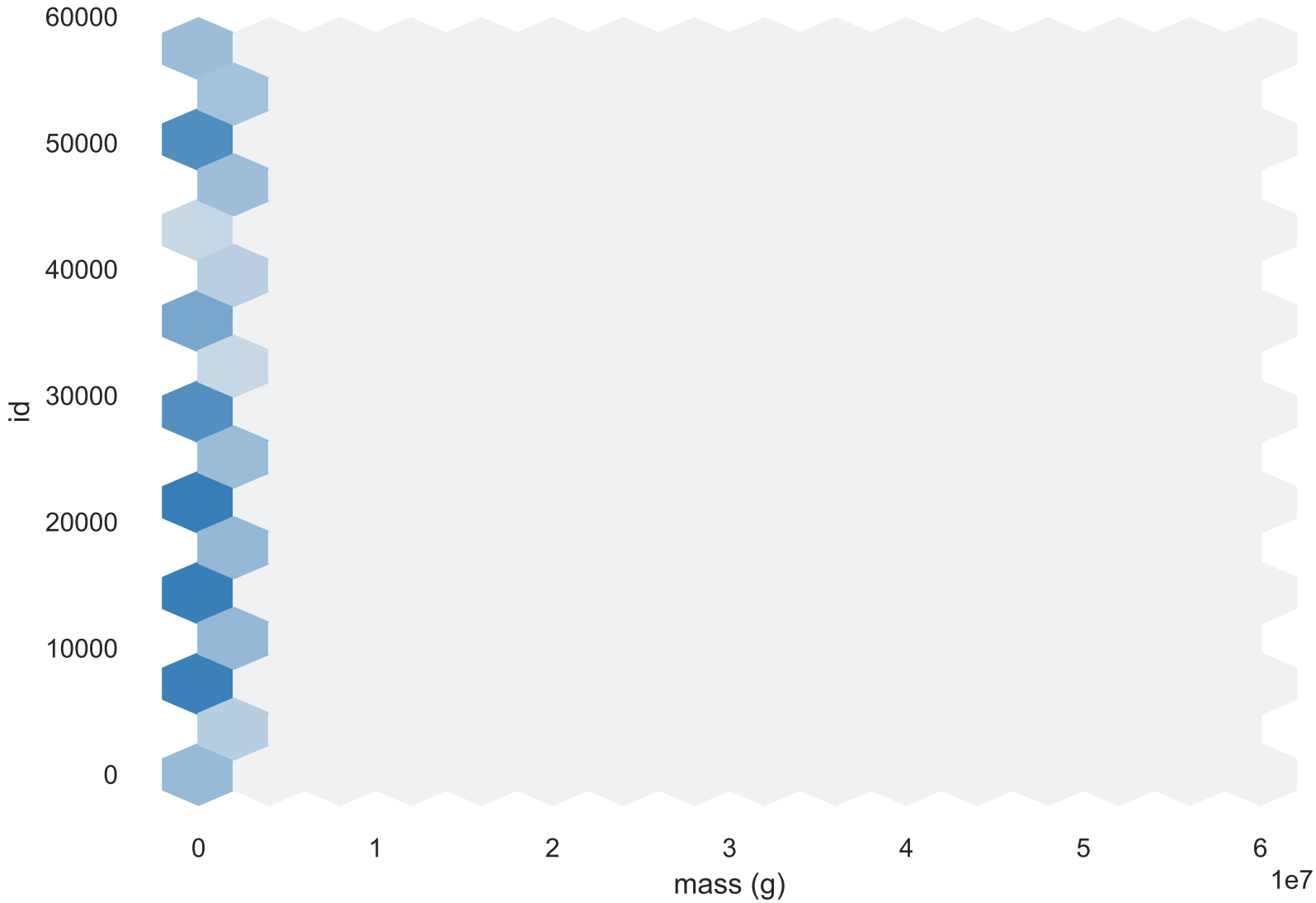

2000
1800
1600
1400
1200
1000
800
year
0
10000
20000
30000
40000
50000
60000
id

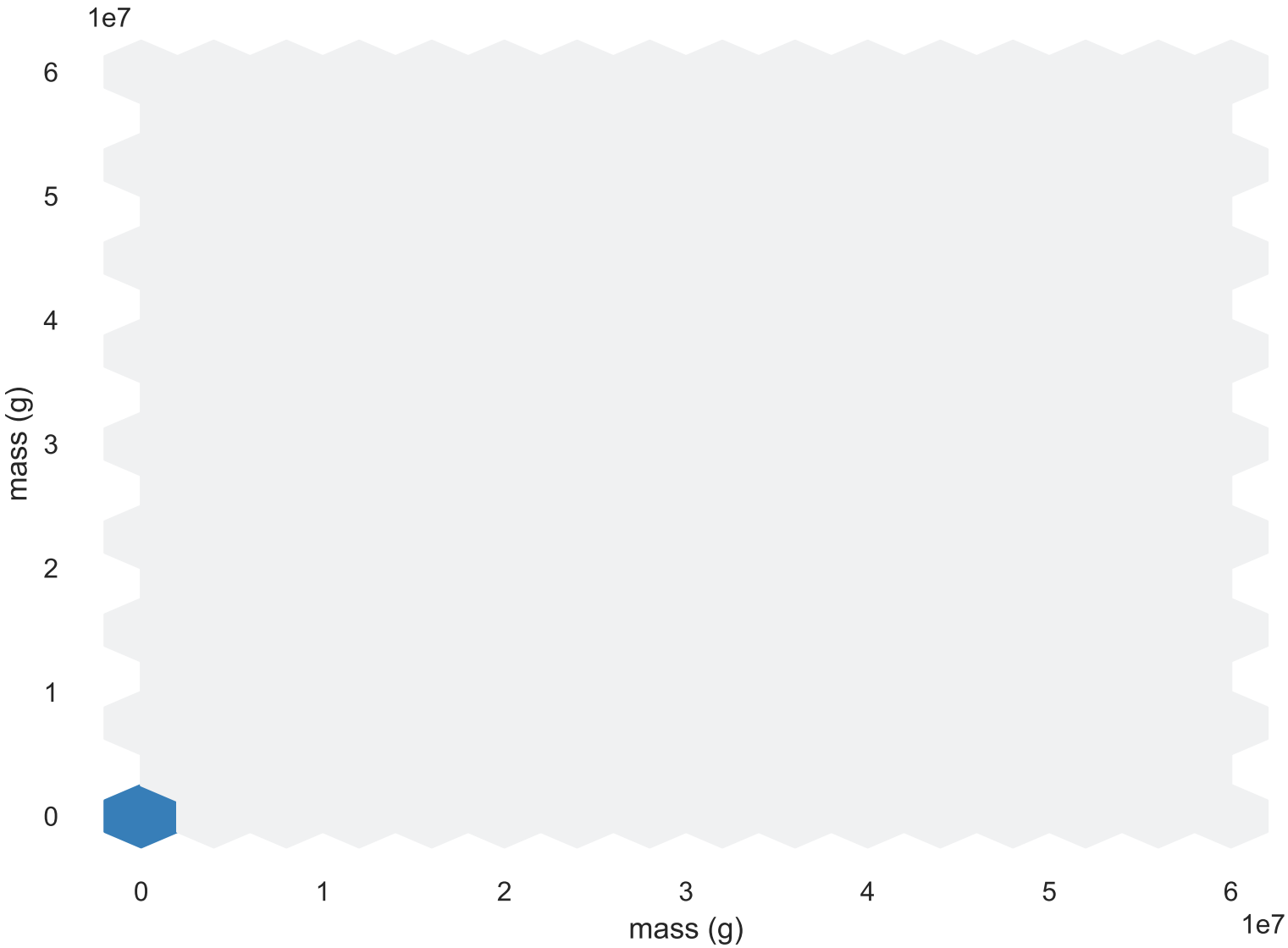

75
50
25
0
−25
−50
−75
reclat
0
10000
20000
30000
40000
50000
60000
id

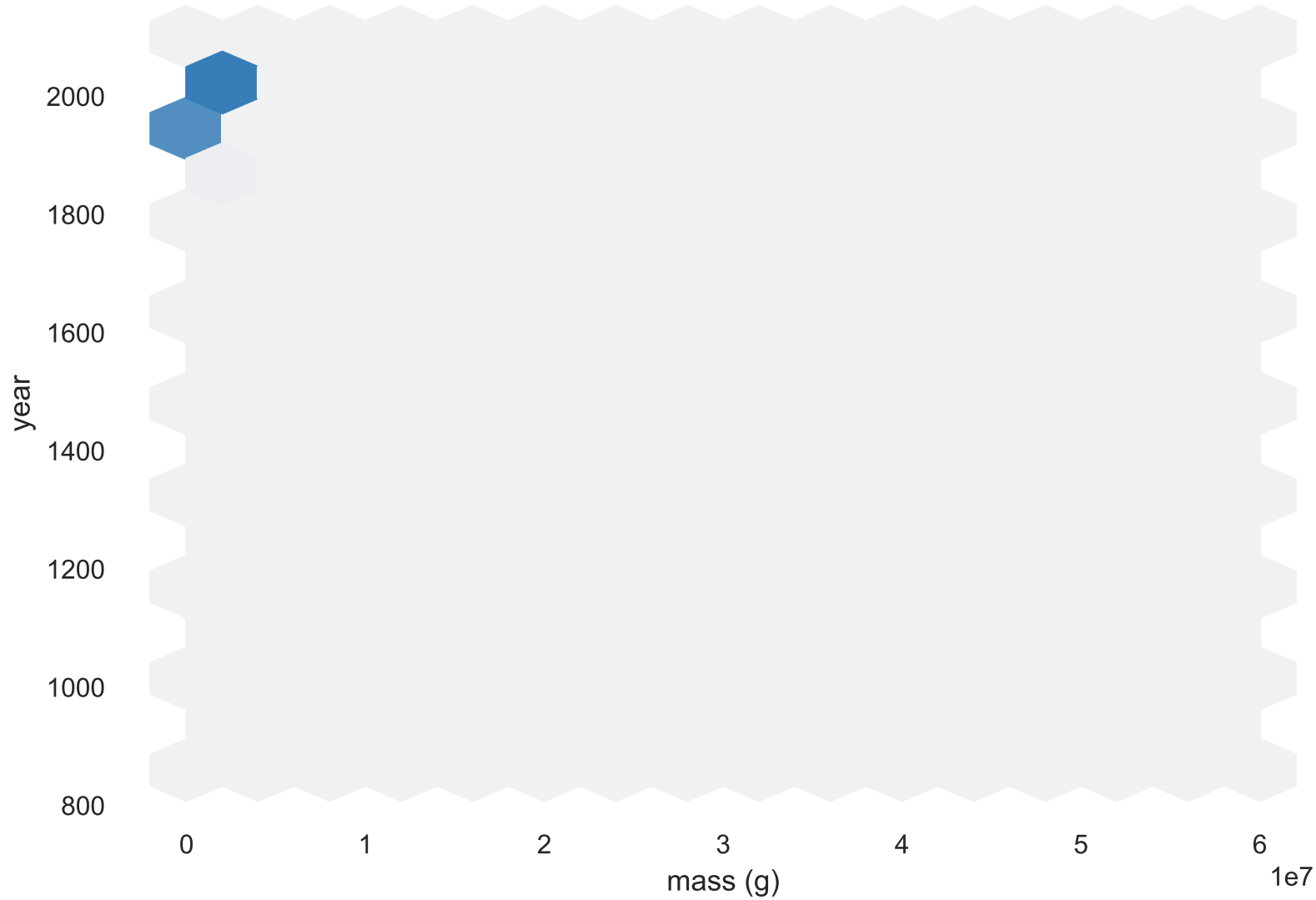

300
200
100
0
−100
reclong
0
10000
20000
30000
40000
50000
60000
id

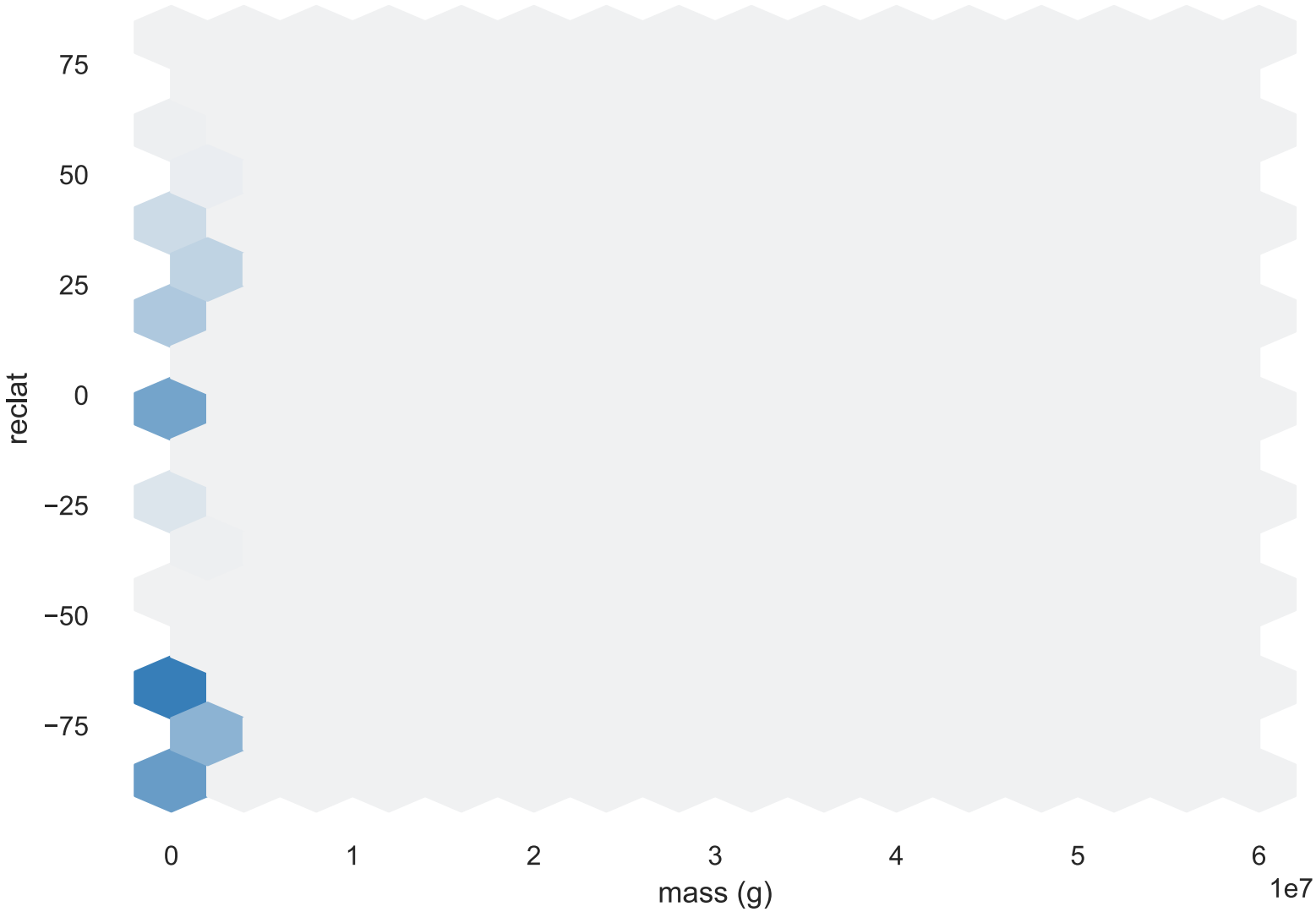

75
50
25
0
−25
−50
−75
Latitude
0
10000
20000
30000
40000
50000
60000
id

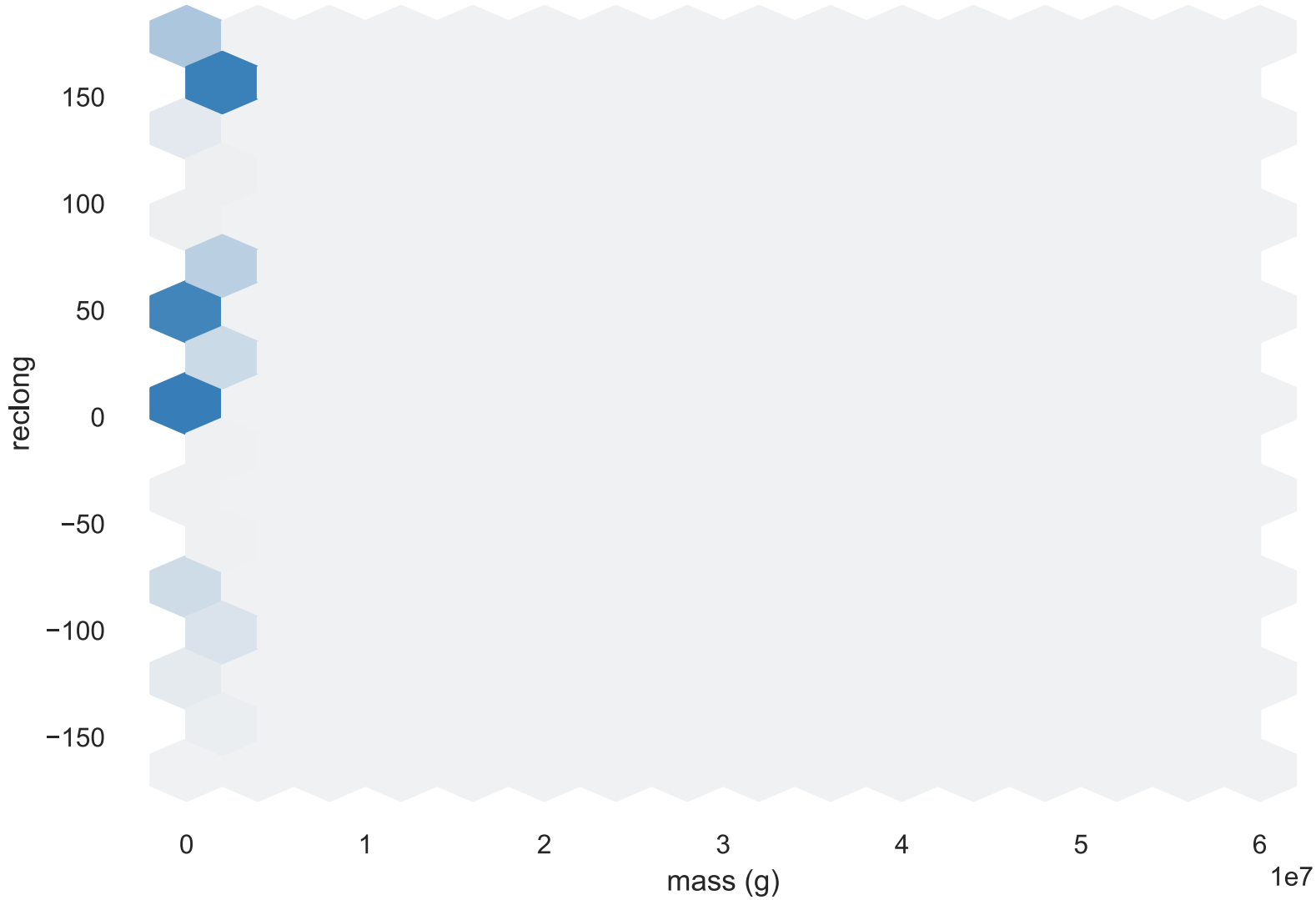

150
100
50
0
−50
−100
−150
Longitude
0
1
2
3
4
5
6
mass (g)
1e7

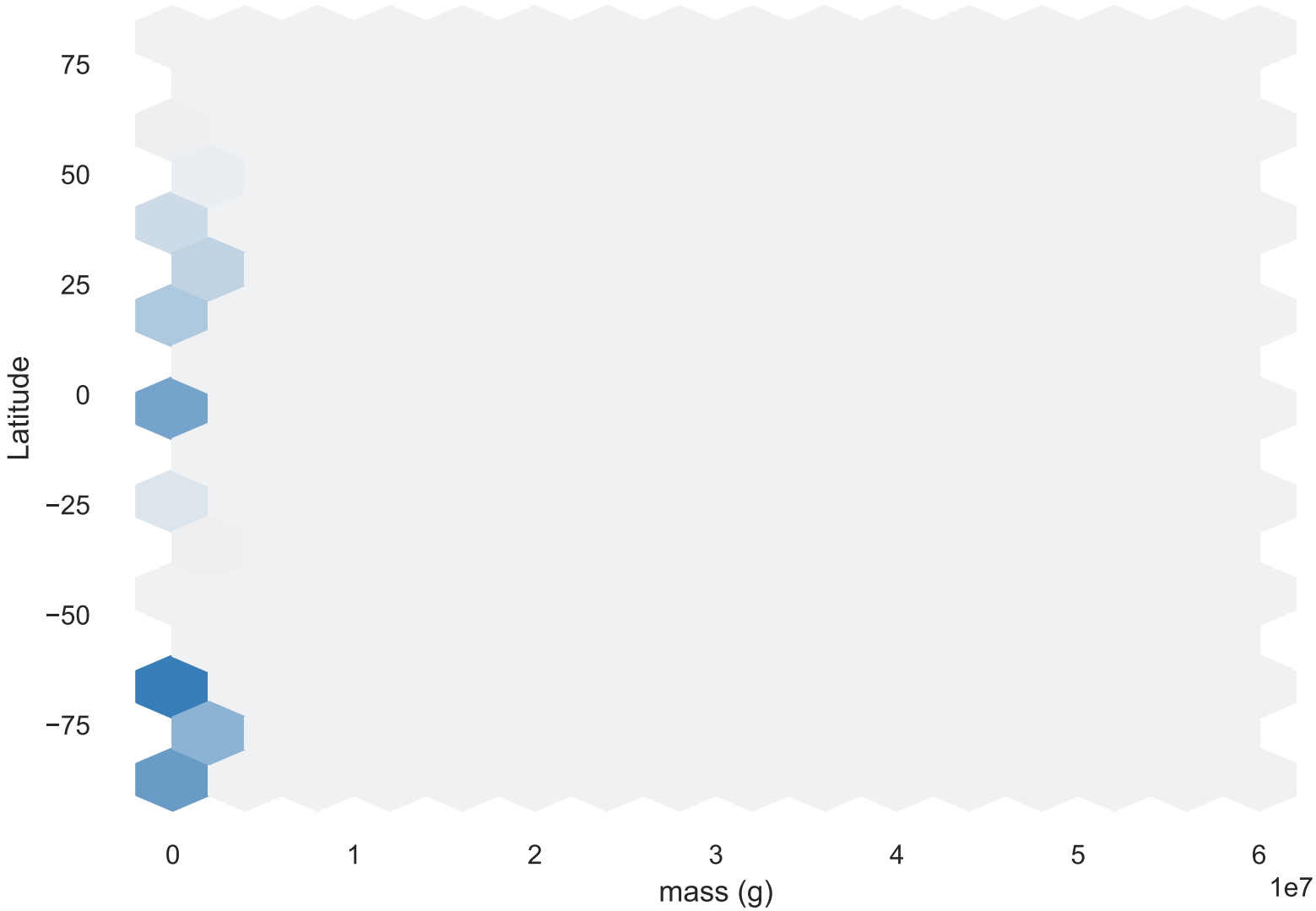

60000
50000
40000
30000
20000
10000
0
id
0
1
2
3
4
5
6
1e7
mass (g)

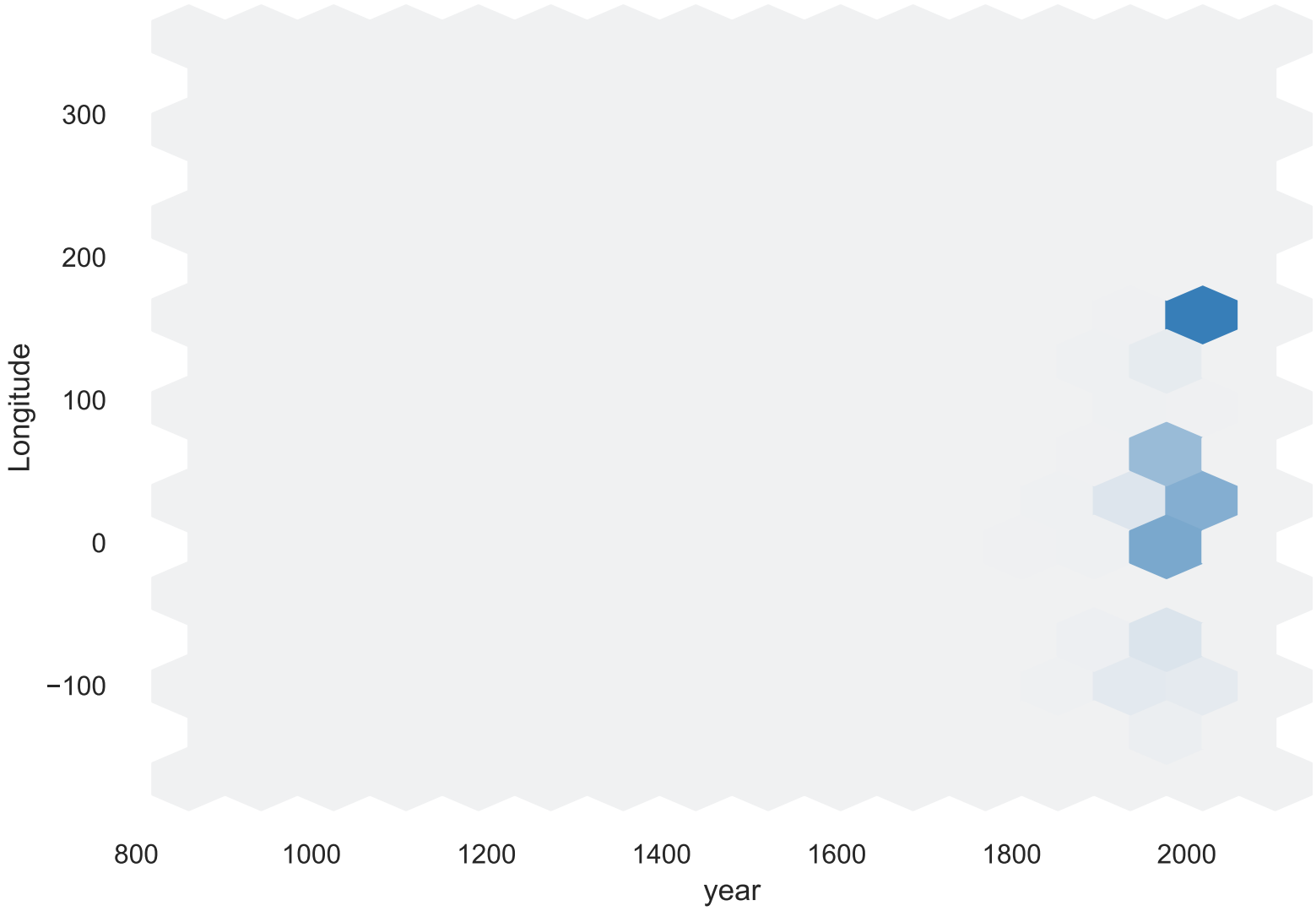

1e7
6
5
4
3
2
1
0
mass (g)
0
1
2
3
4
5
6
1e7
mass (g)

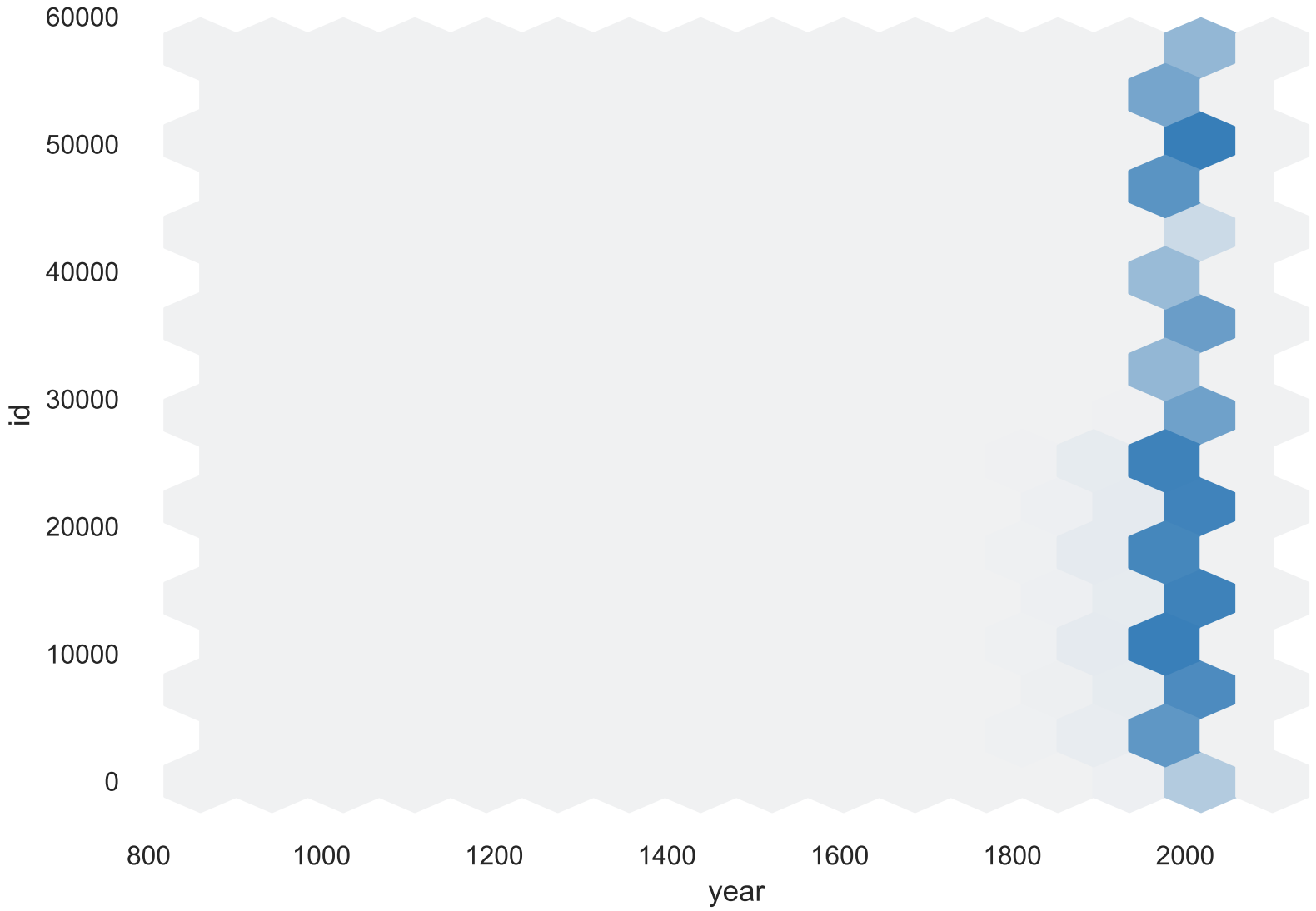

2000
1800
1600
1400
1200
1000
800
year
0
1
2
3
4
5
6
1e7
mass (g)

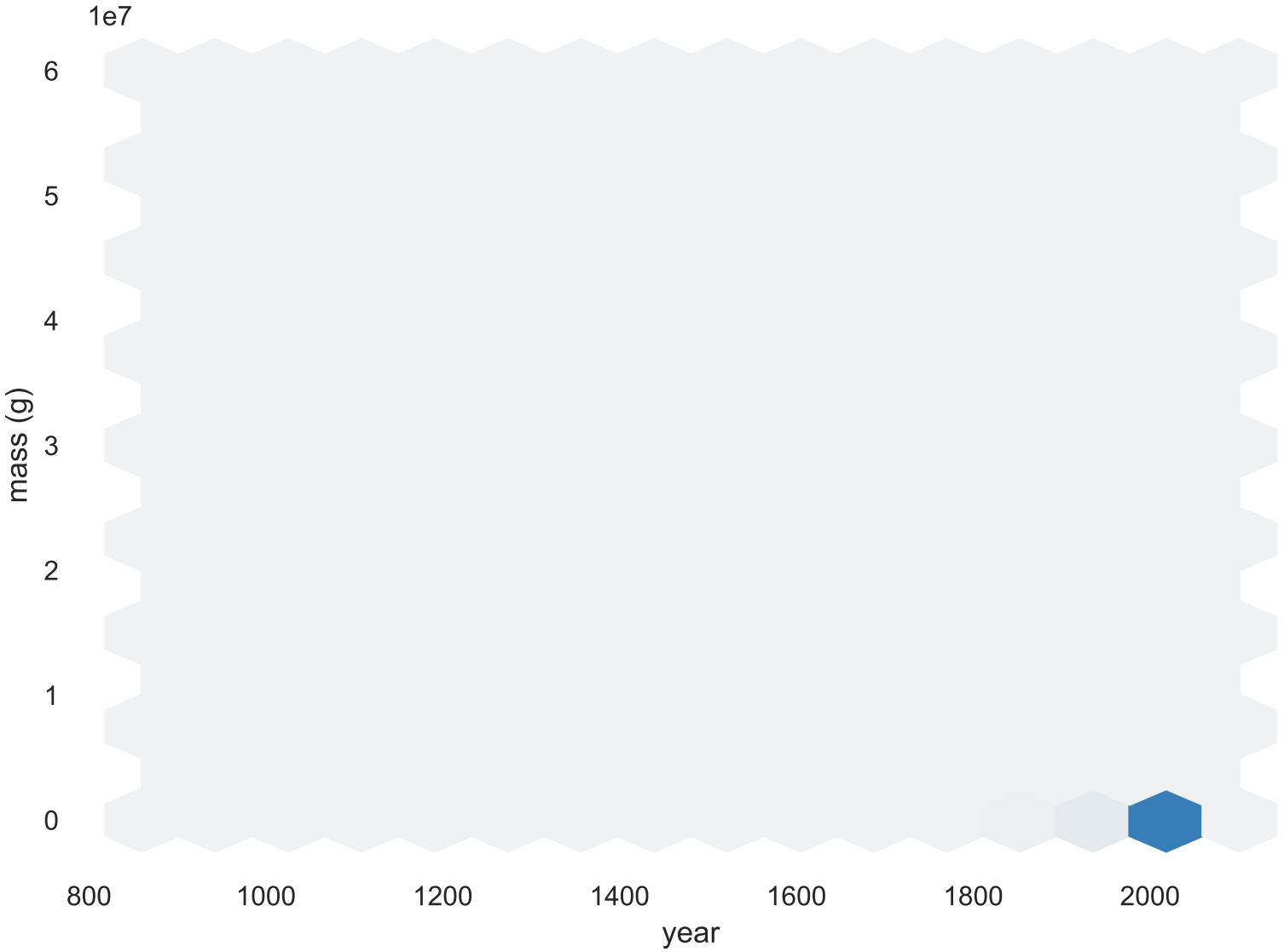

75
50
25
0
−25
−50
−75
reclat
0
1
2
3
4
5
6
1e7
mass (g)

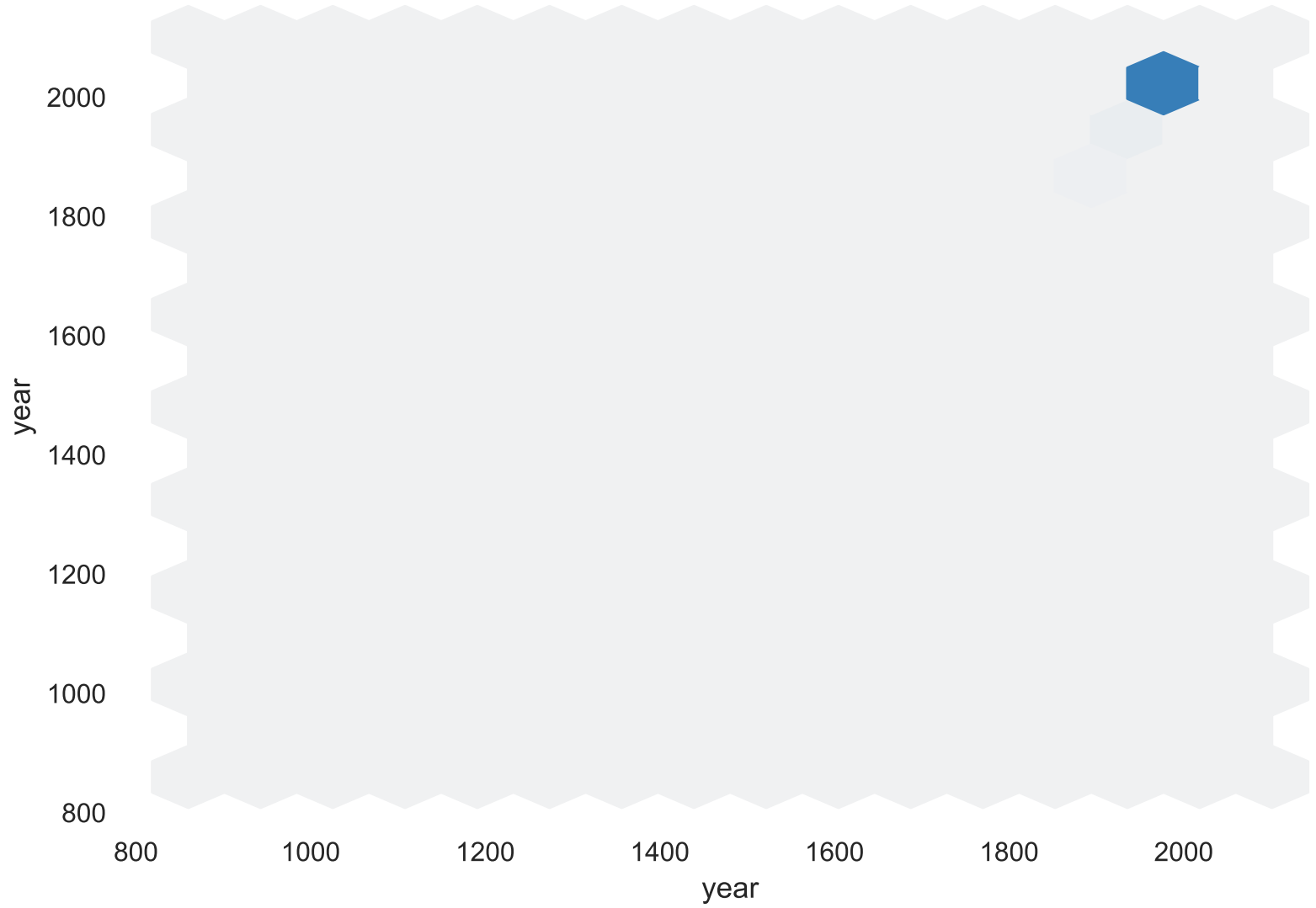

150
100
50
0
−50
−100
−150
reclong
0
1
2
3
4
5
6
1e7
mass (g)

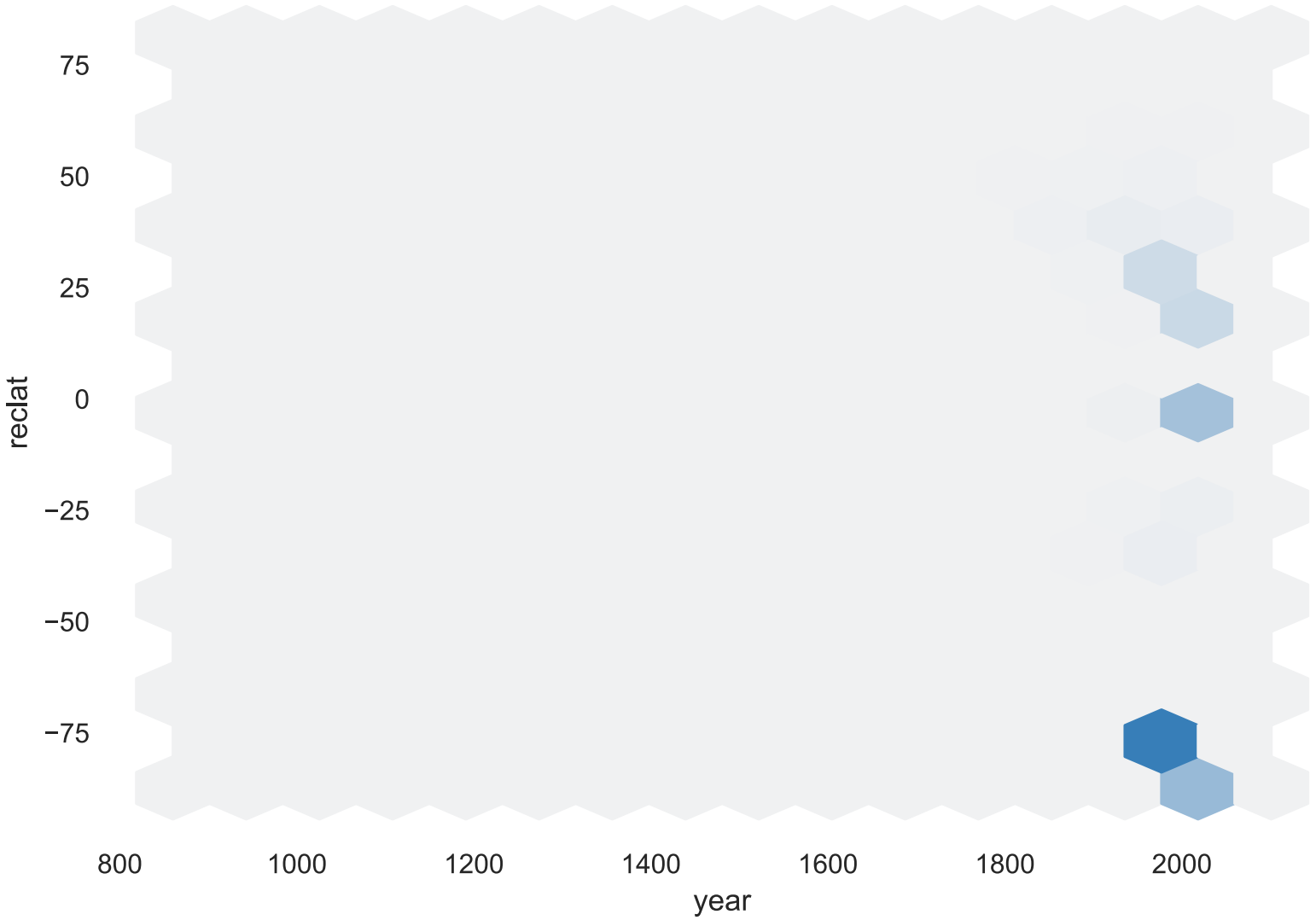

75
50
25
0
−25
−50
−75
Latitude
0
1
2
3
4
5
6
1e7
mass (g)

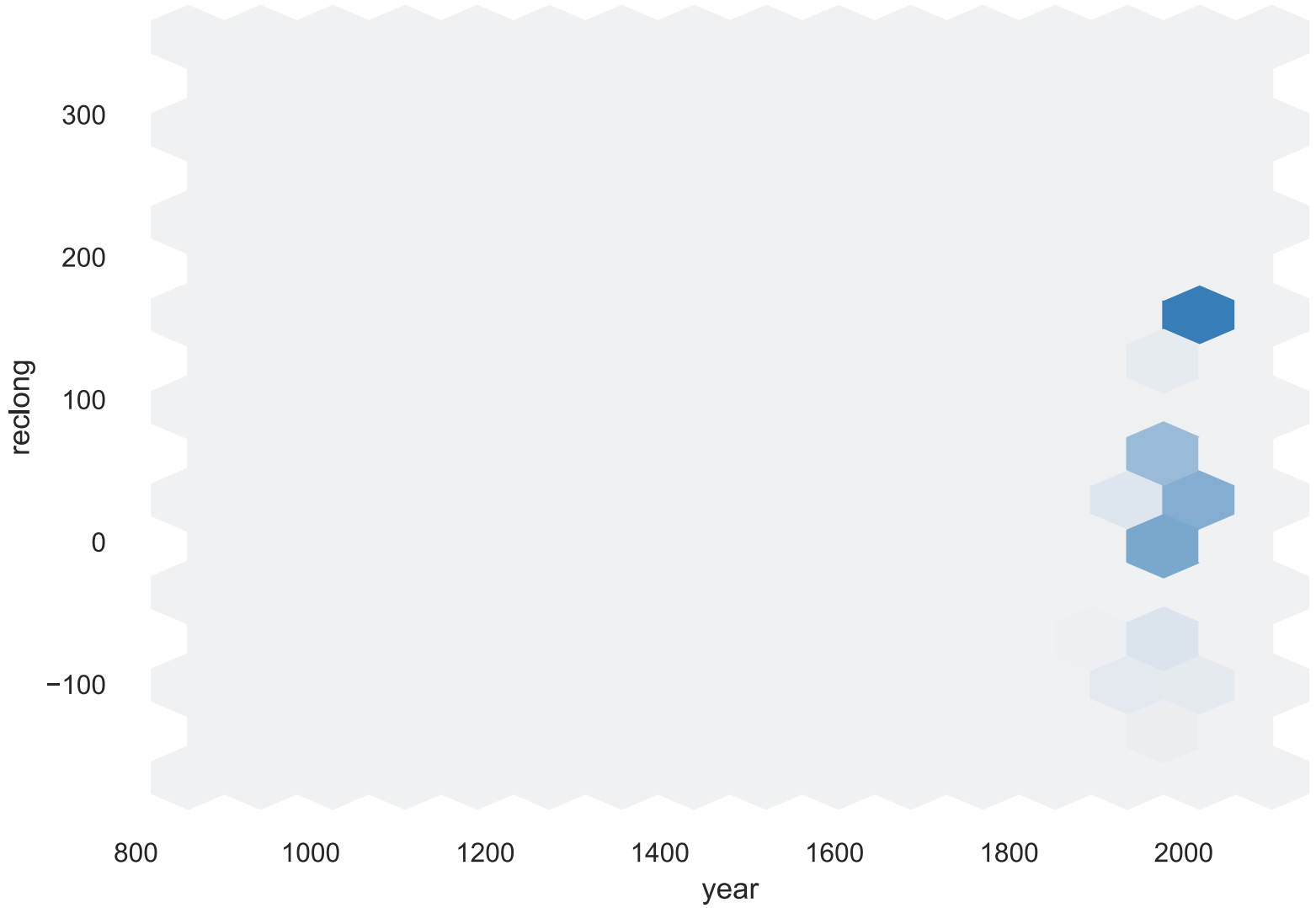

300
200
100
0
−100
Longitude
800
1000
1200
1400
1600
1800
2000
year

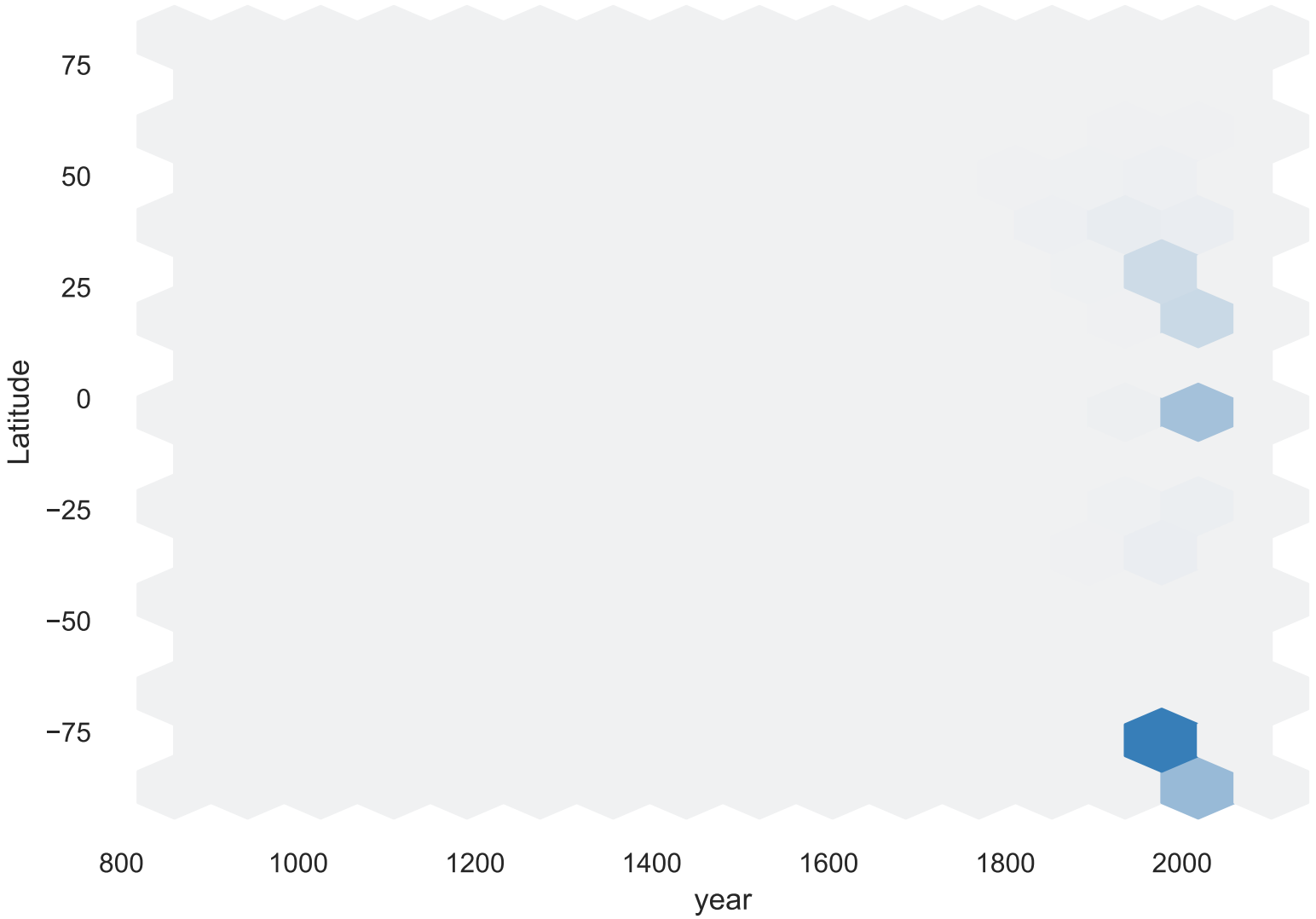

60000
50000
40000
30000
20000
10000
0
id
800
1000
1200
1400
1600
1800
2000
year

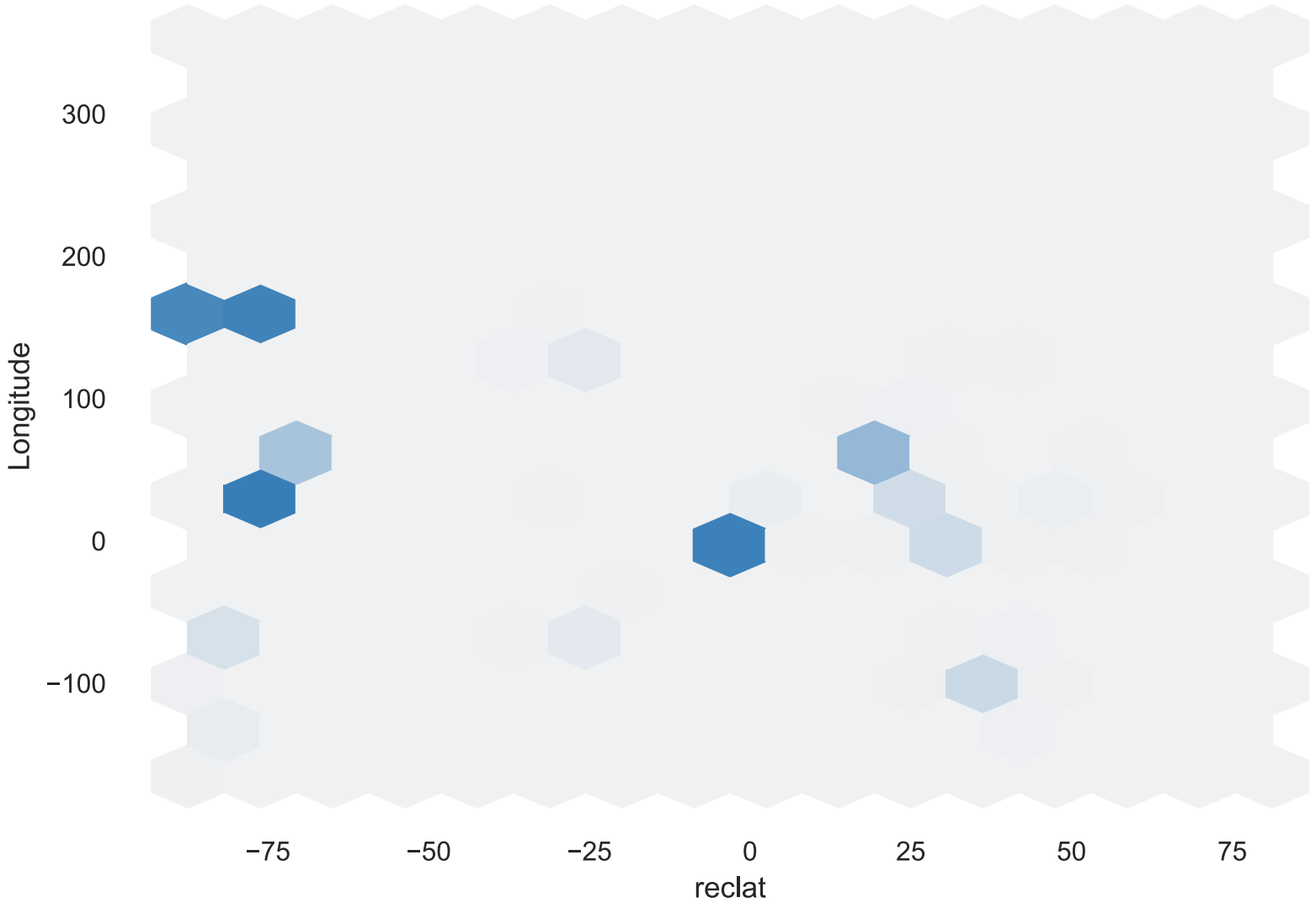

1e7
6
5
4
3
2
1
0
mass (g)
800
1000
1200
1400
1600
1800
2000
year

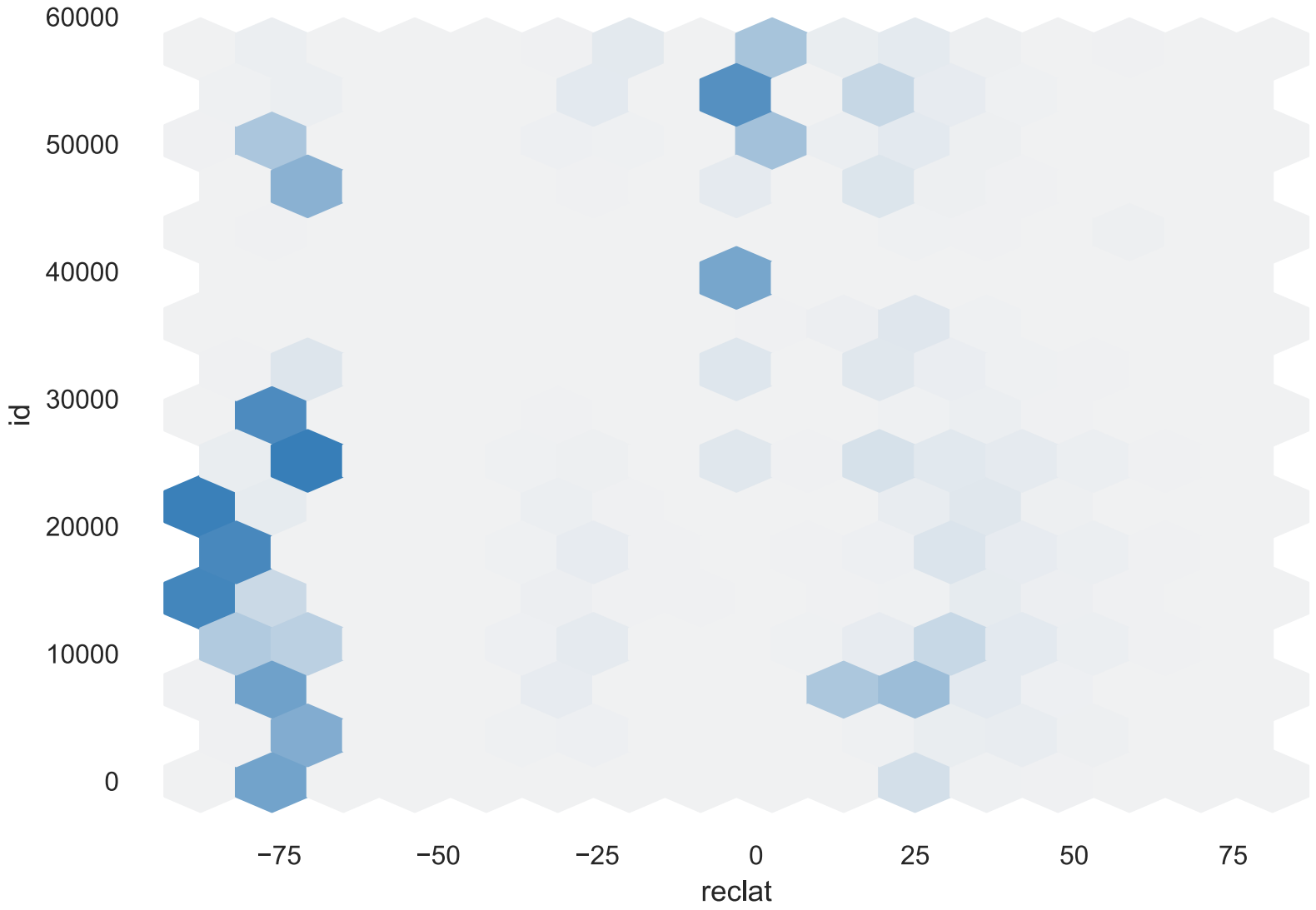

2000
1800
1600
1400
1200
1000
800
year
800
1000
1200
1400
1600
1800
2000
year

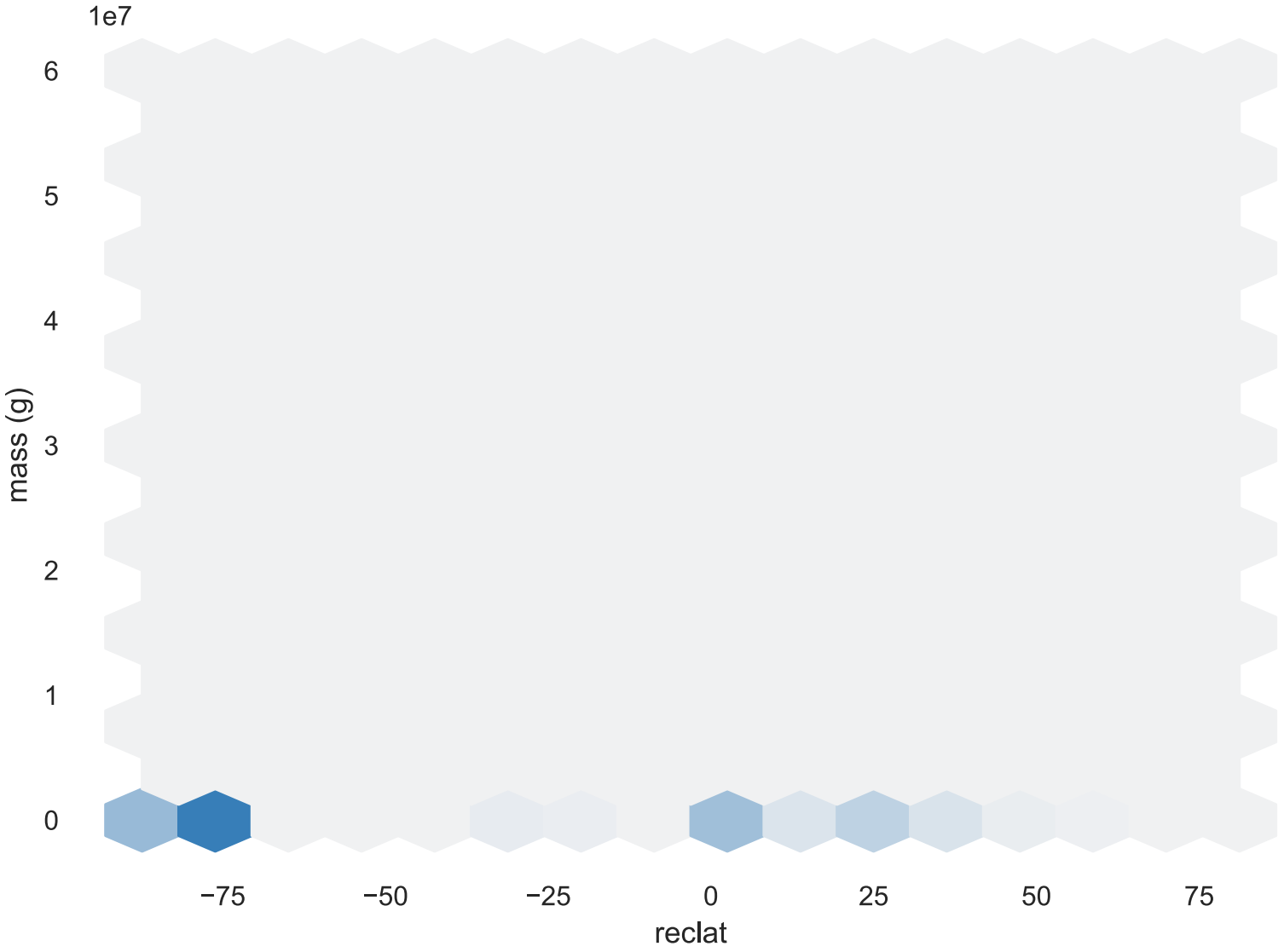

75
50
25
0
−25
−50
−75
reclat
800
1000
1200
1400
1600
1800
2000
year

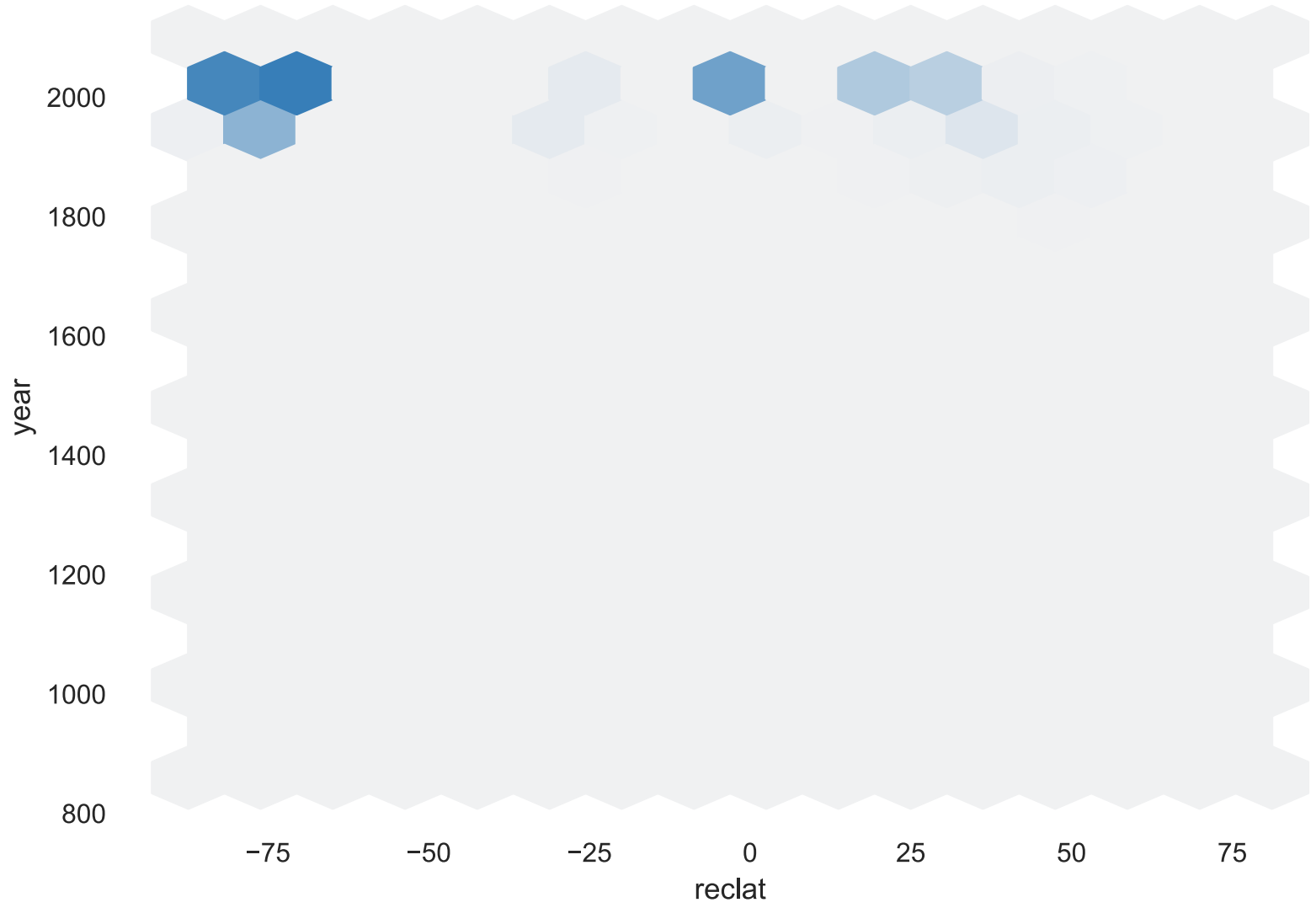

300
200
100
0
−100
reclong
800
1000
1200
1400
1600
1800
2000
year

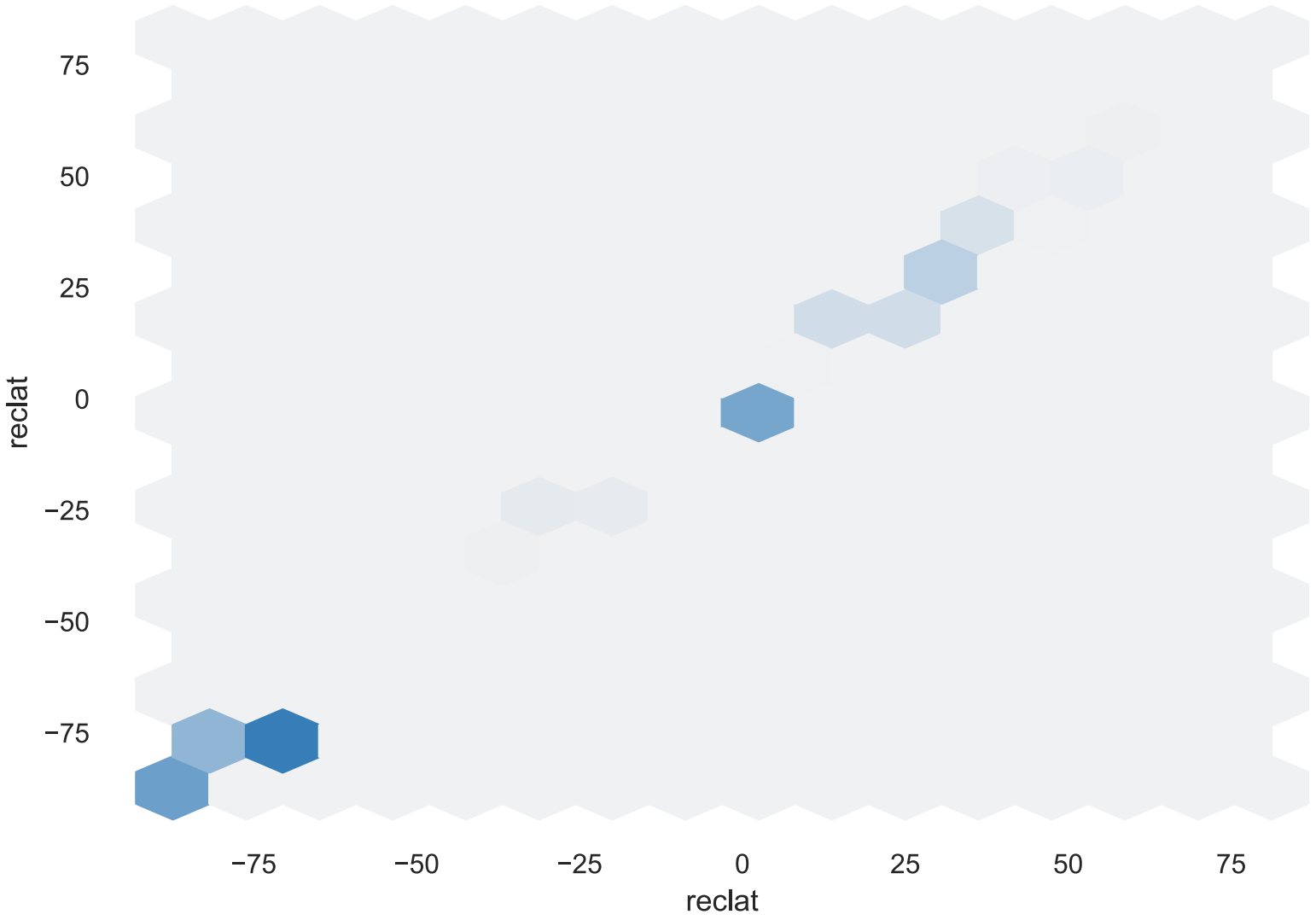

75
50
25
0
−25
−50
−75
Latitude
800
1000
1200
1400
1600
1800
2000
year

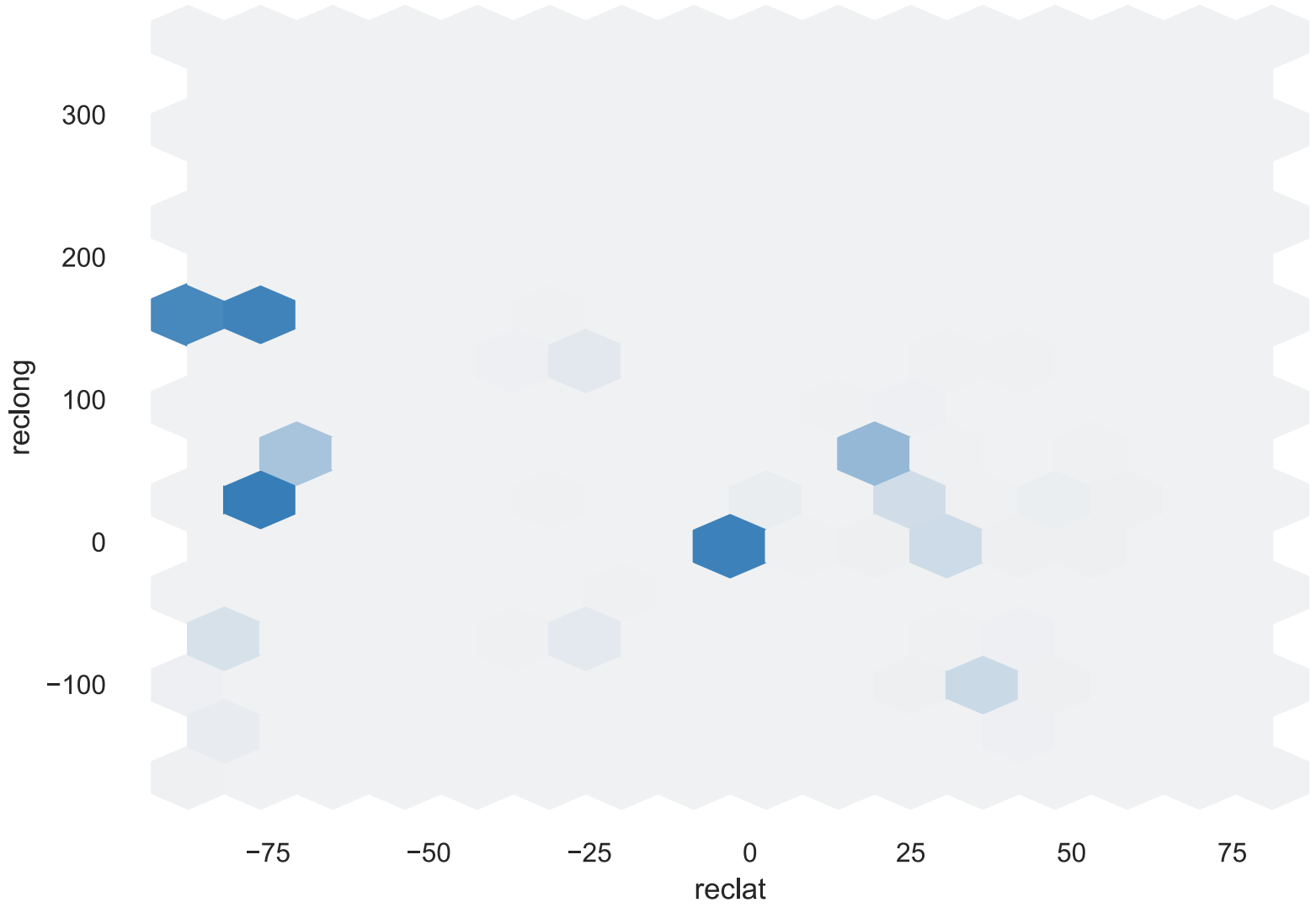

300
200
100
0
−100
Longitude
−75
−50
−25
0
25
50
75
reclat

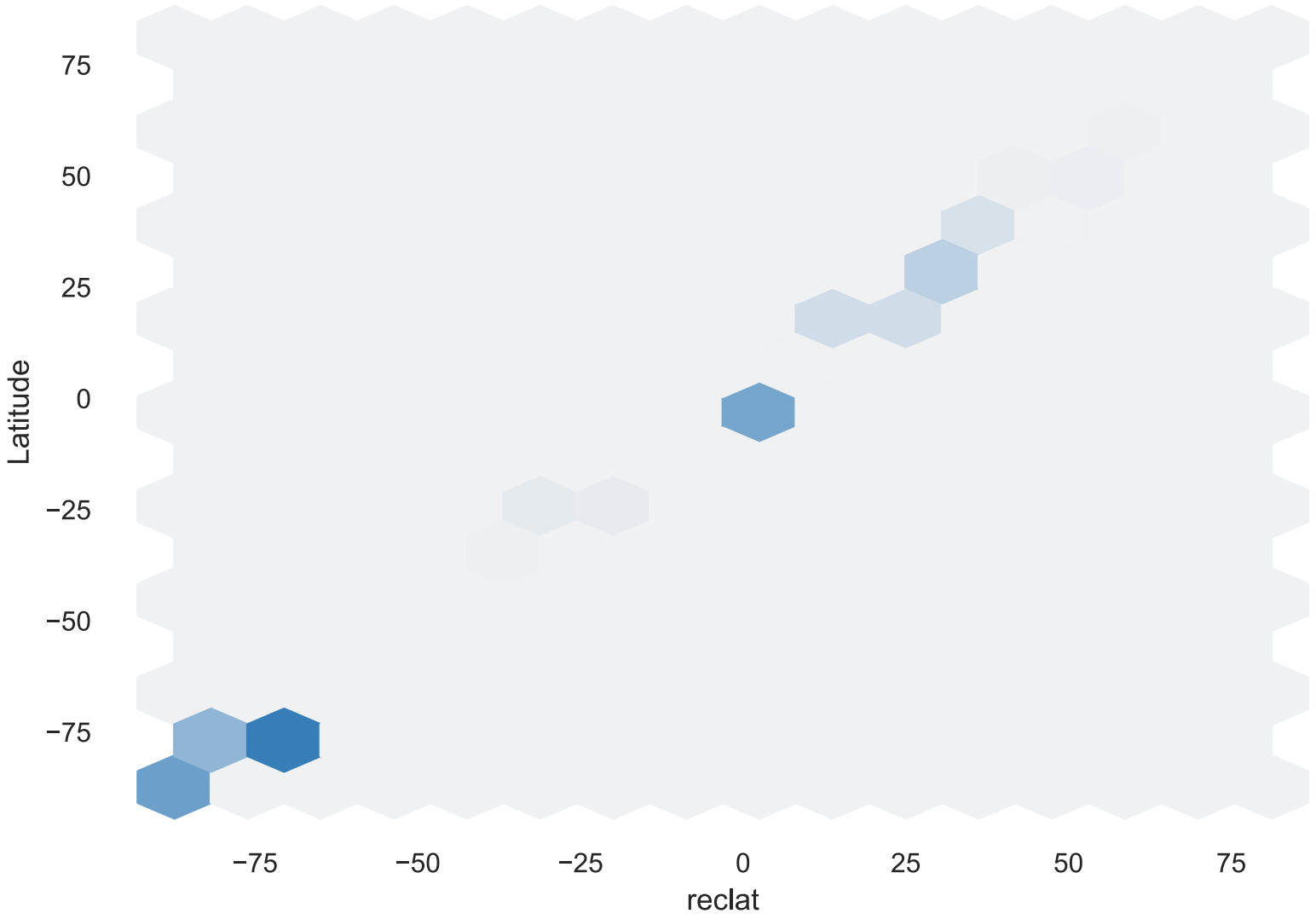

60000
50000
40000
30000
20000
10000
0
id
−75
−50
−25
0
25
50
75
reclat

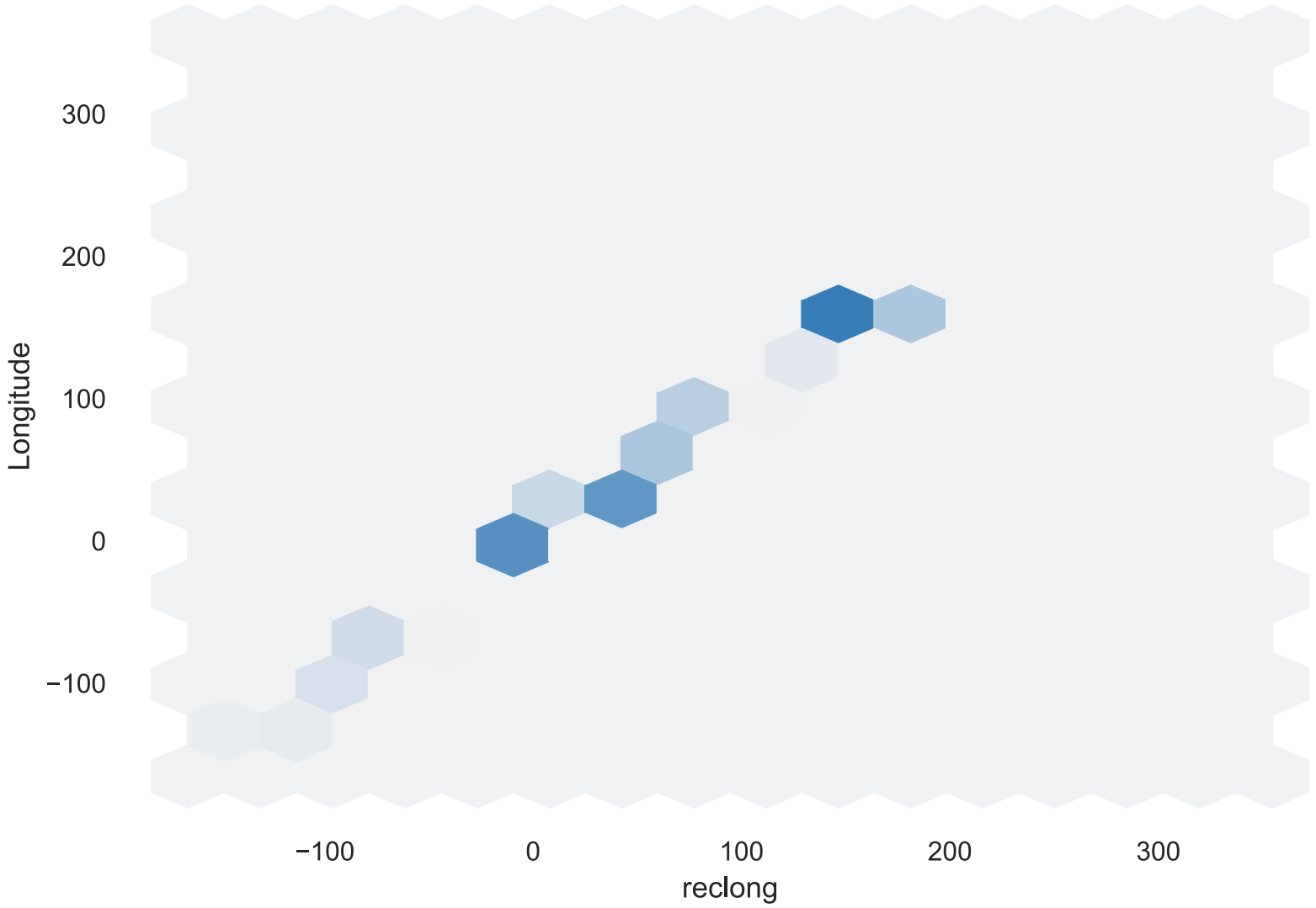

1e7
6
5
4
3
2
1
0
mass (g)
−75
−50
−25
0
25
50
75
reclat

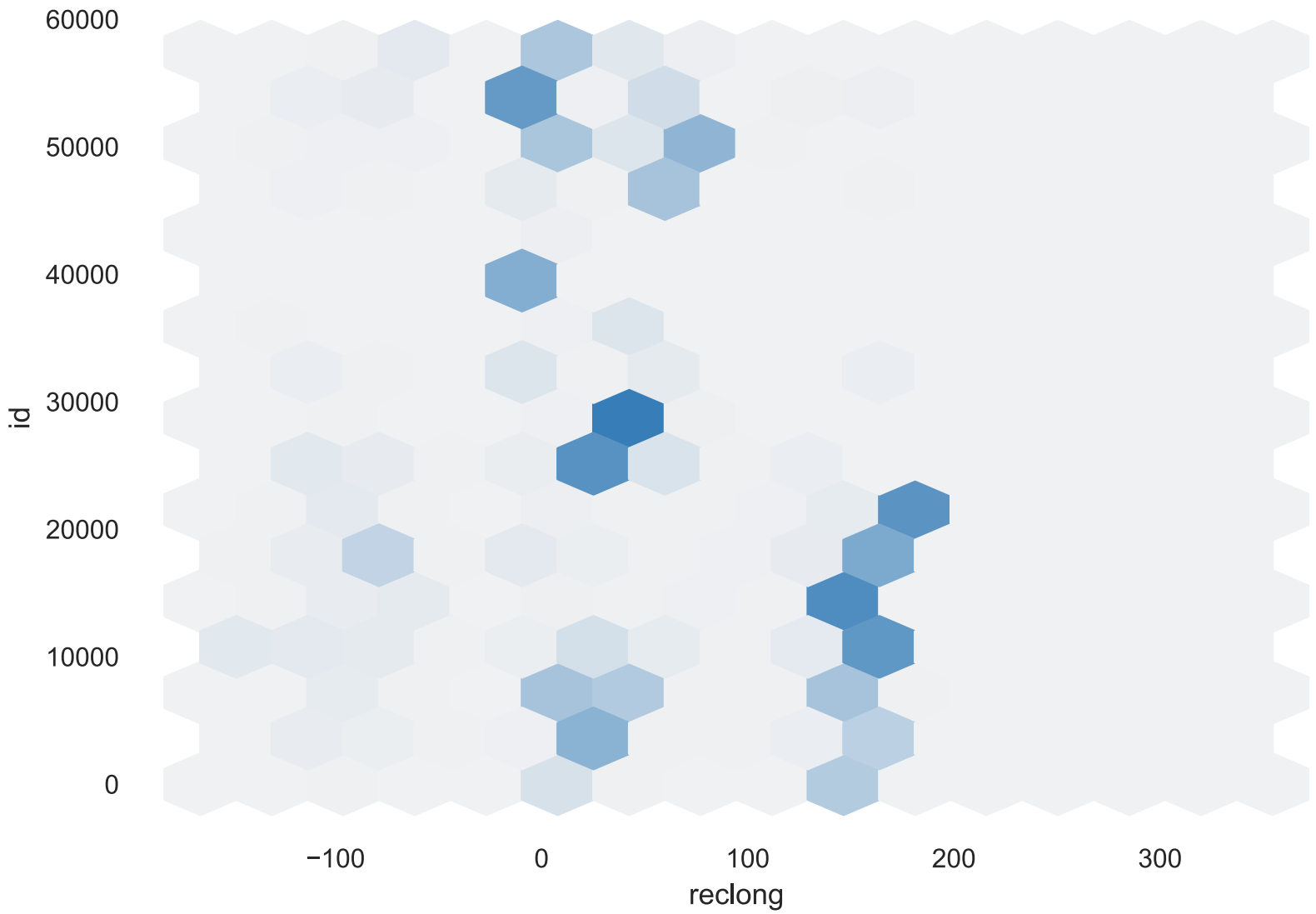

2000
1800
1600
1400
1200
1000
800
year
−75
−50
−25
0
25
50
75
reclat

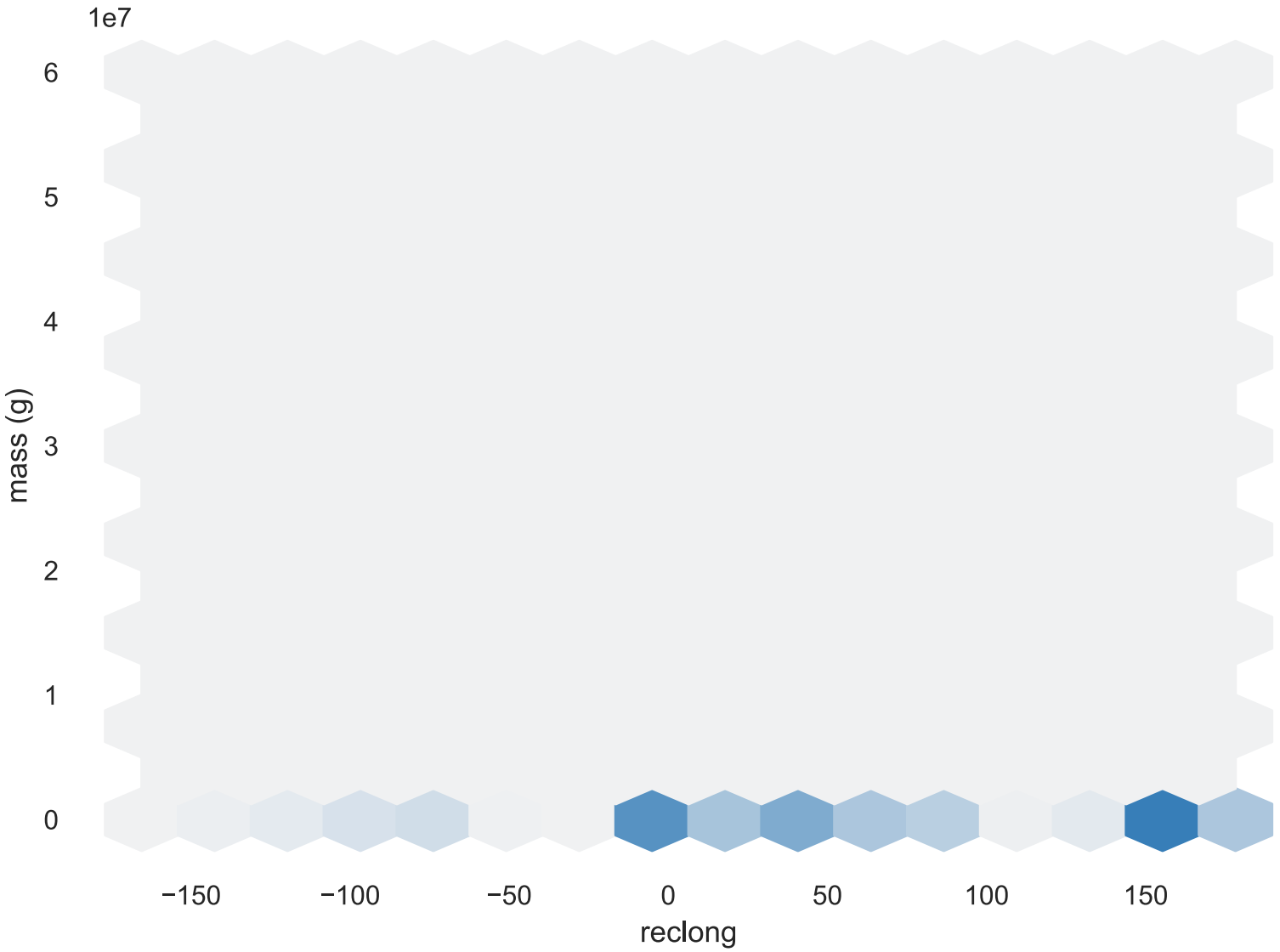

75
50
25
0
−25
−50
−75
reclat
−75
−50
−25
0
25
50
75
reclat

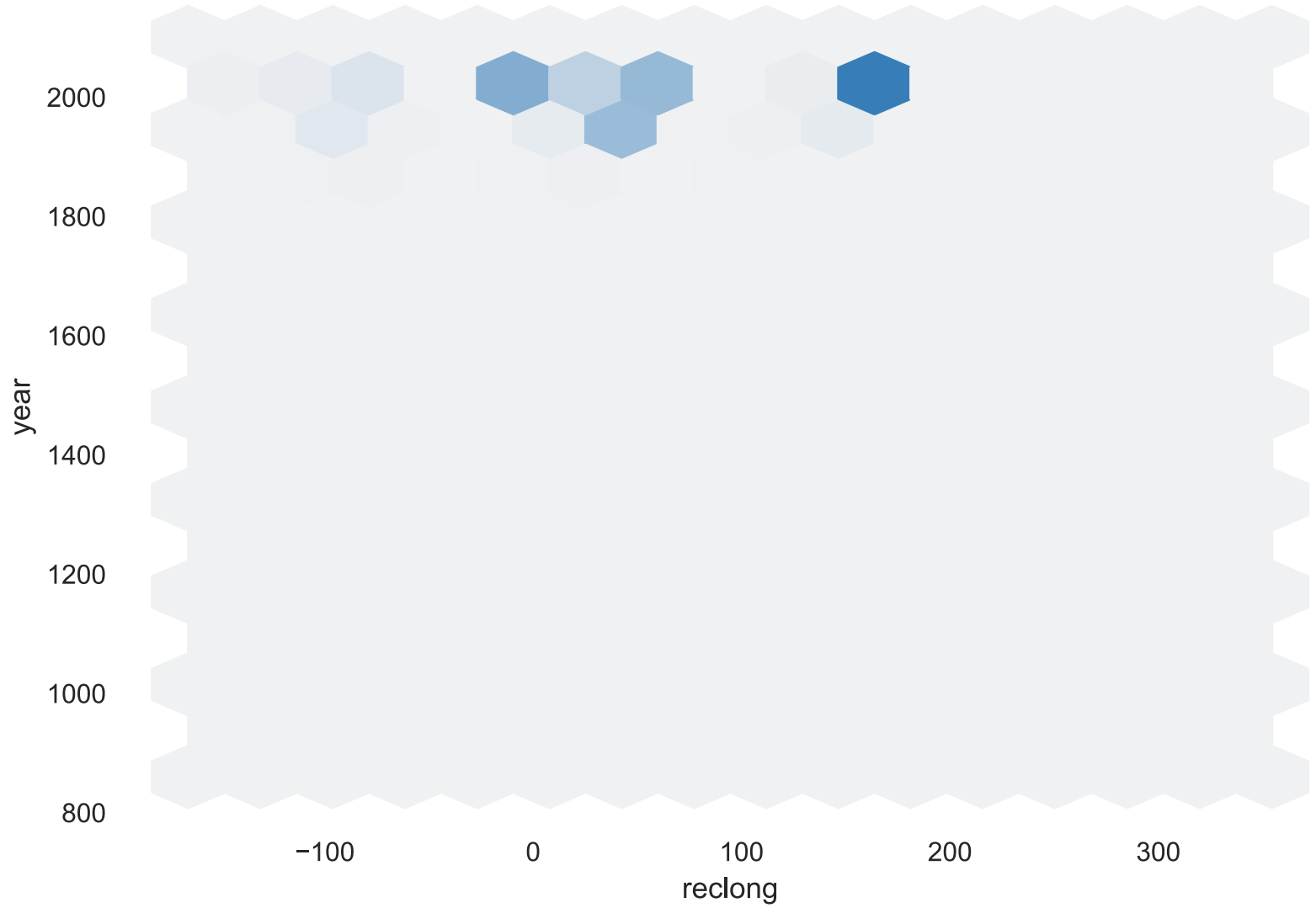

300
200
100
0
−100
reclong
−75
−50
−25
0
25
50
75
reclat

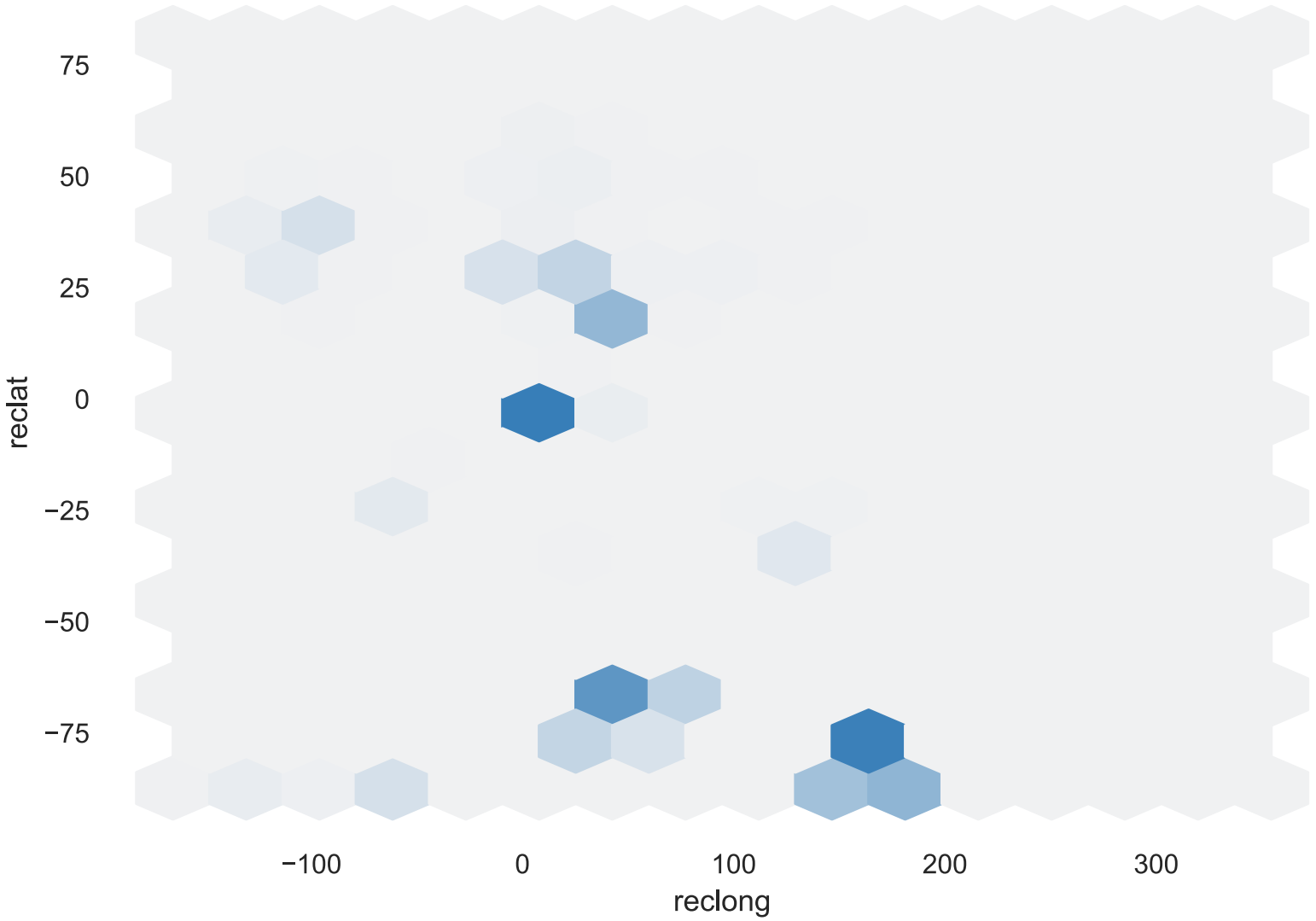

75
50
25
0
−25
−50
−75
Latitude
−75
−50
−25
0
25
50
75
reclat

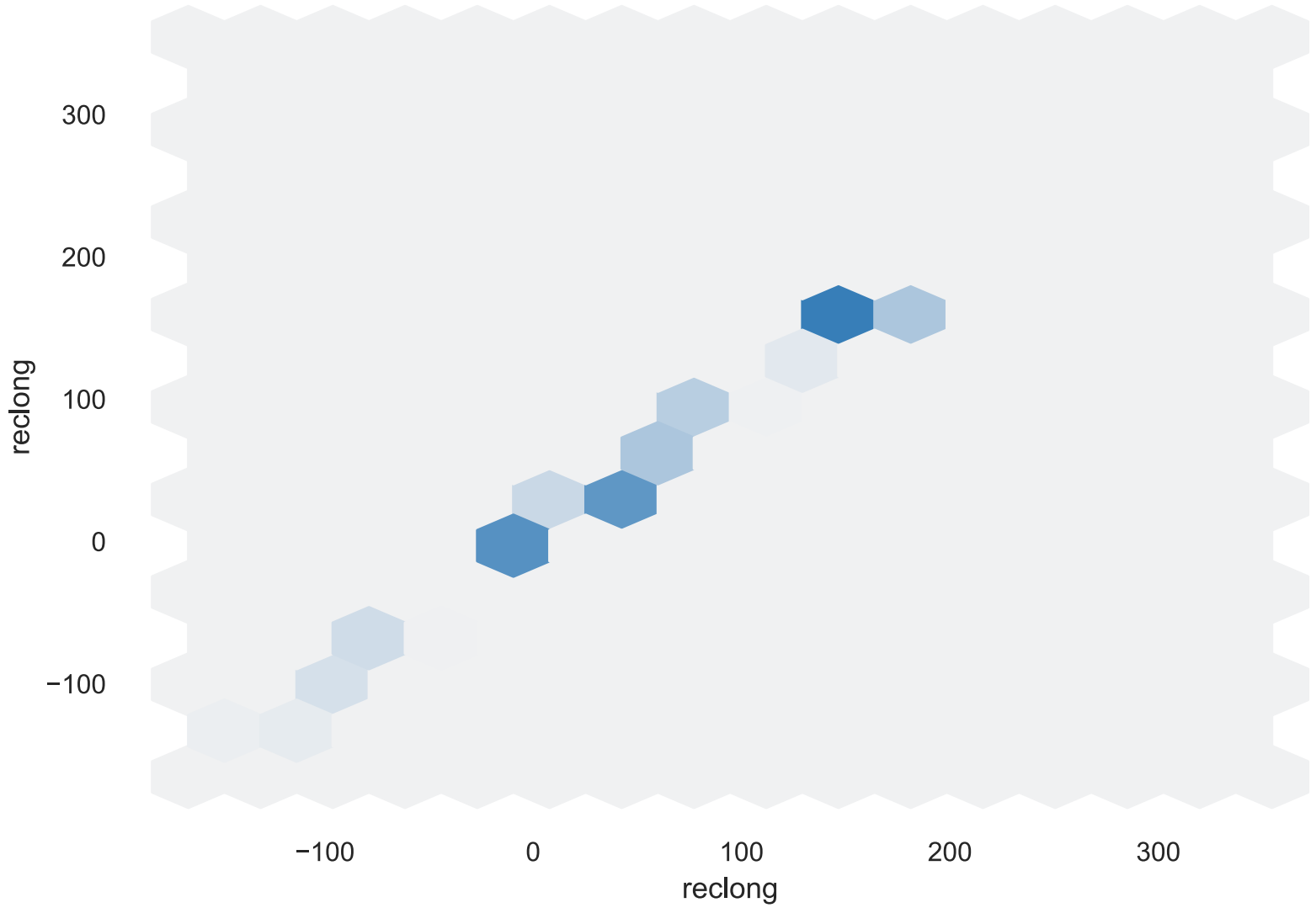

300
200
100
0
−100
Longitude
−100
0
100
200
300
reclong

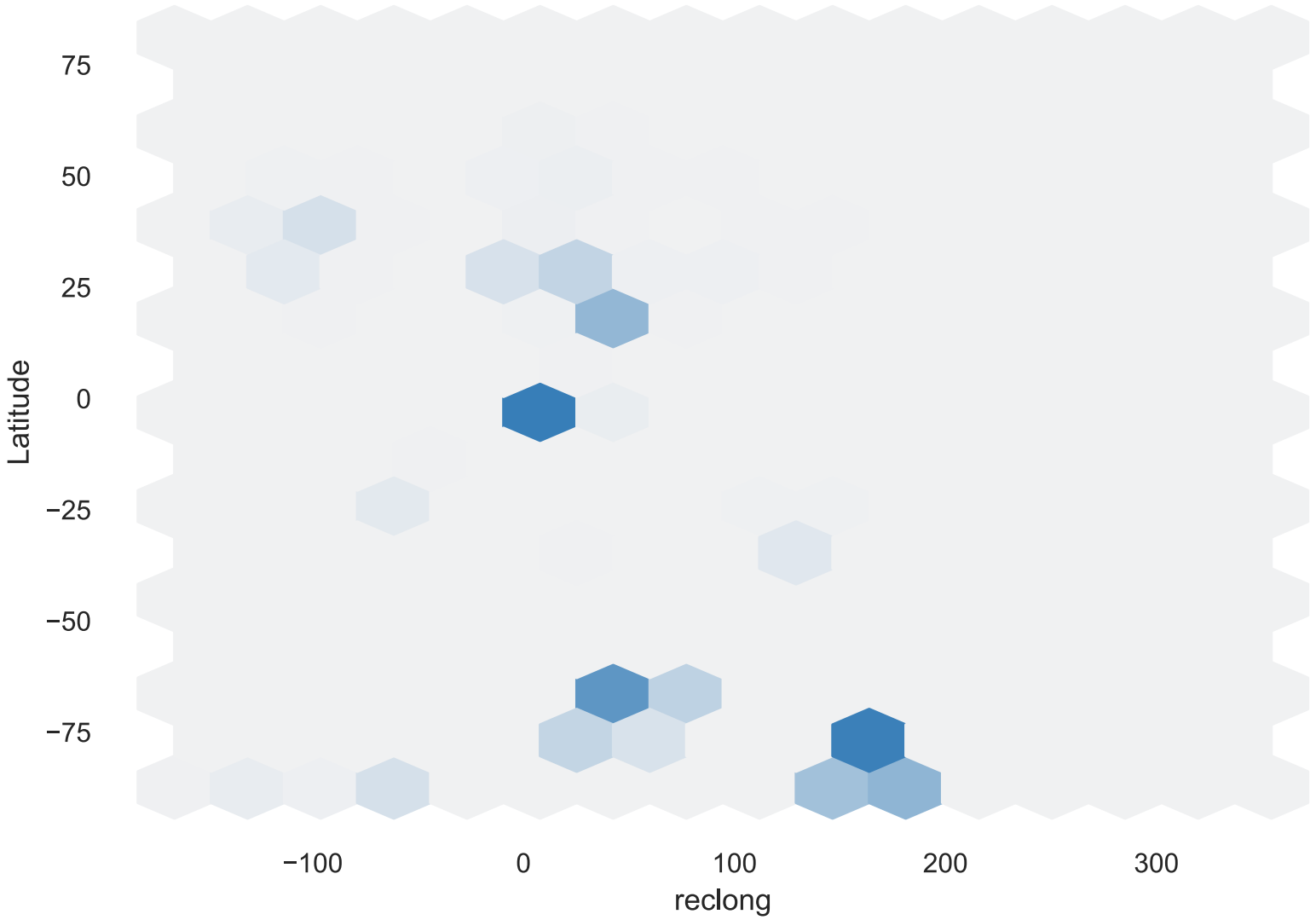

60000
50000
40000
30000
20000
10000
0
id
−100
0
100
200
300
reclong

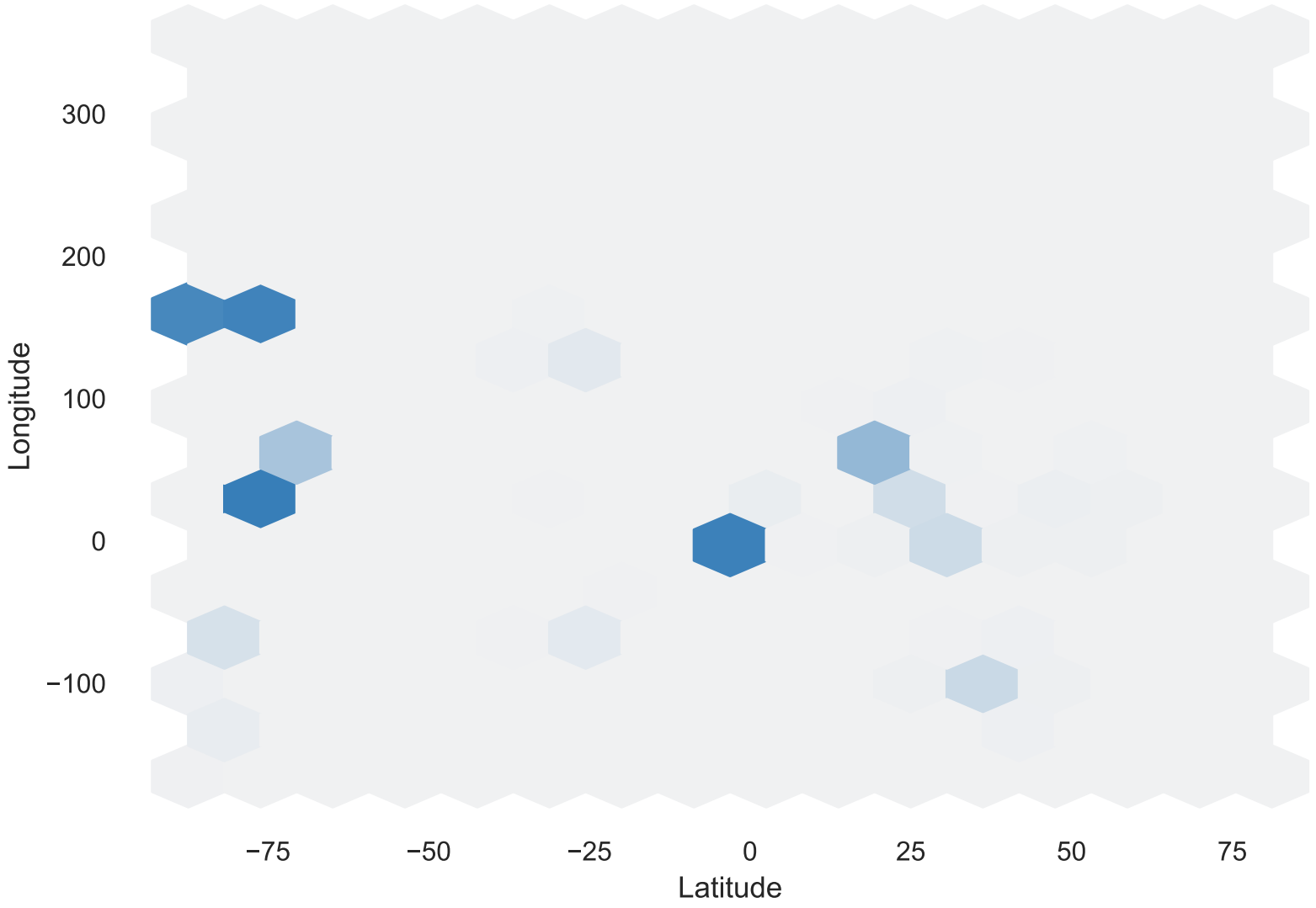

1e7
6
5
4
3
2
1
0
mass (g)
−150
−100
−50
0
50
100
150
reclong

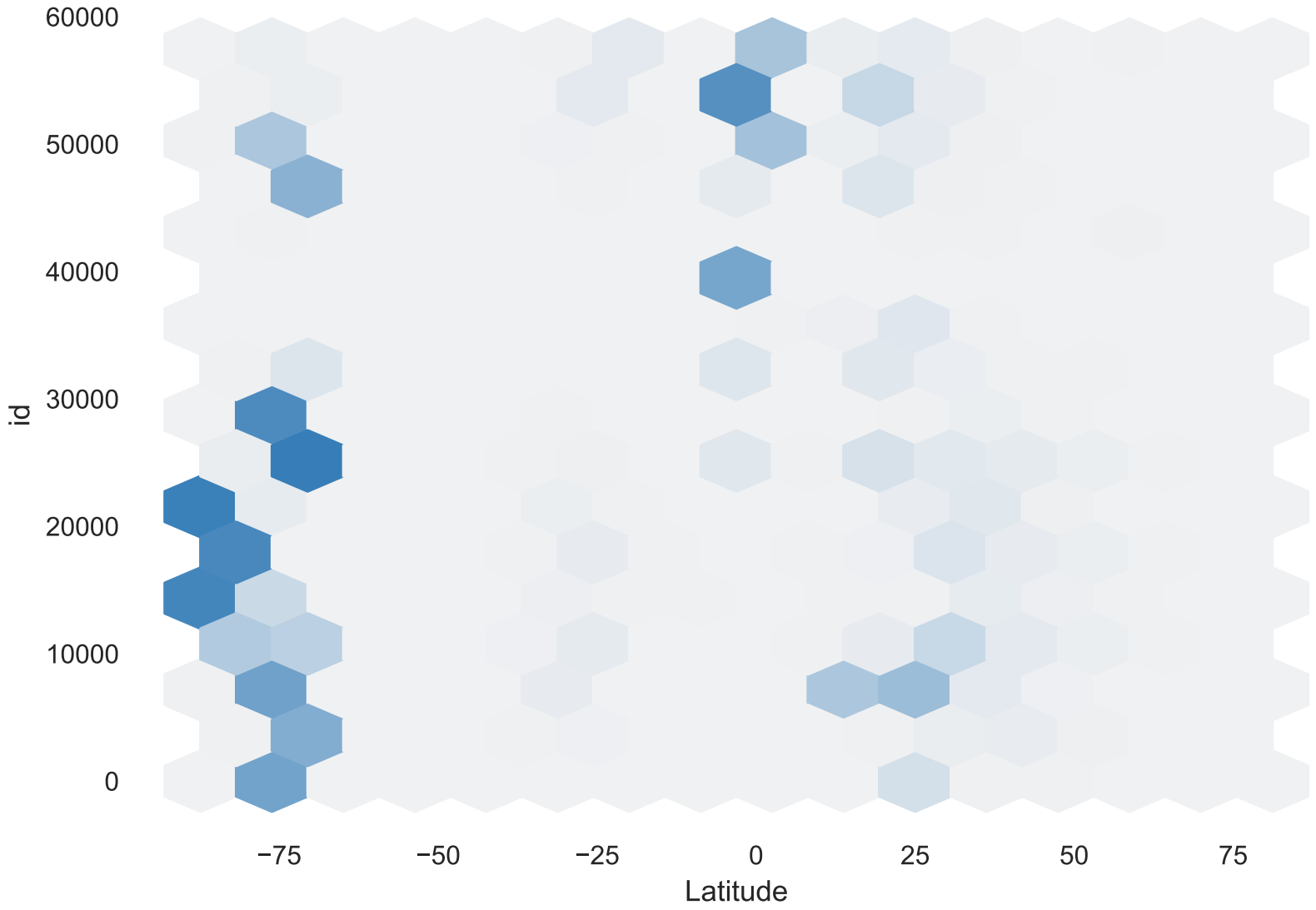

2000
1800
1600
1400
1200
1000
800
year
−100
0
100
200
300
reclong

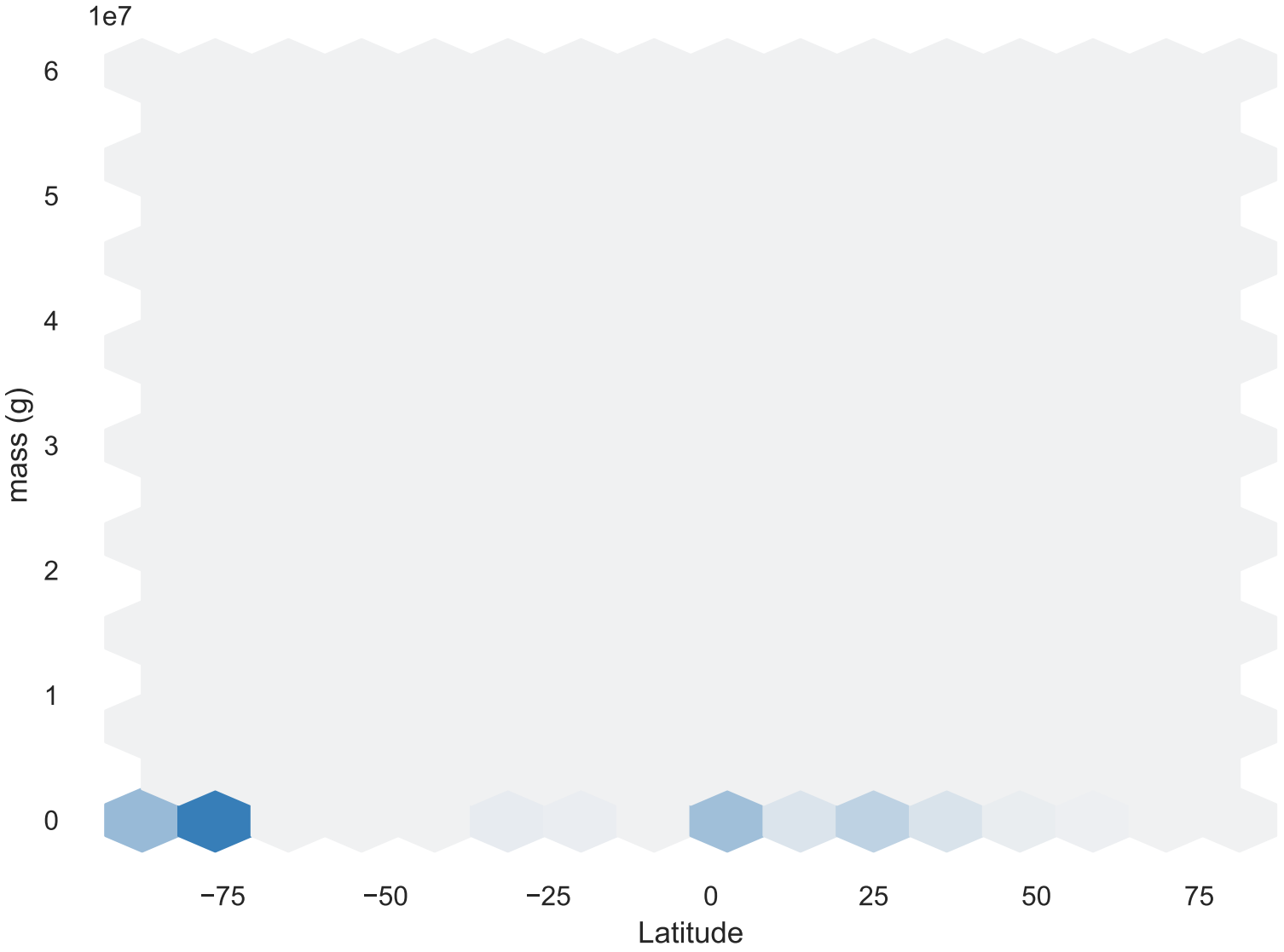

75
50
25
0
−25
−50
−75
reclat
−100
0
100
200
300
reclong

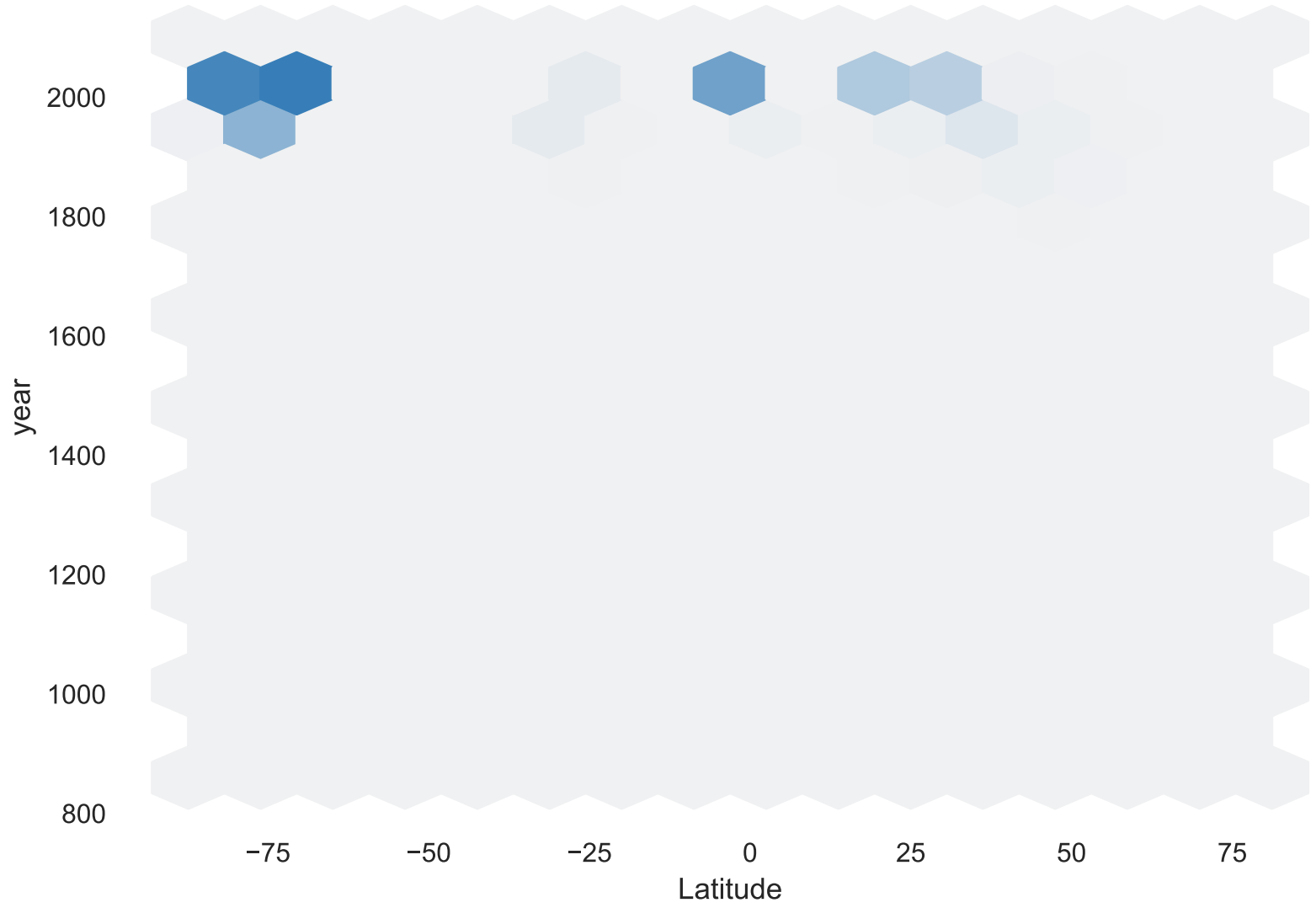

300
200
100
0
−100
reclong
−100
0
100
200
300
reclong

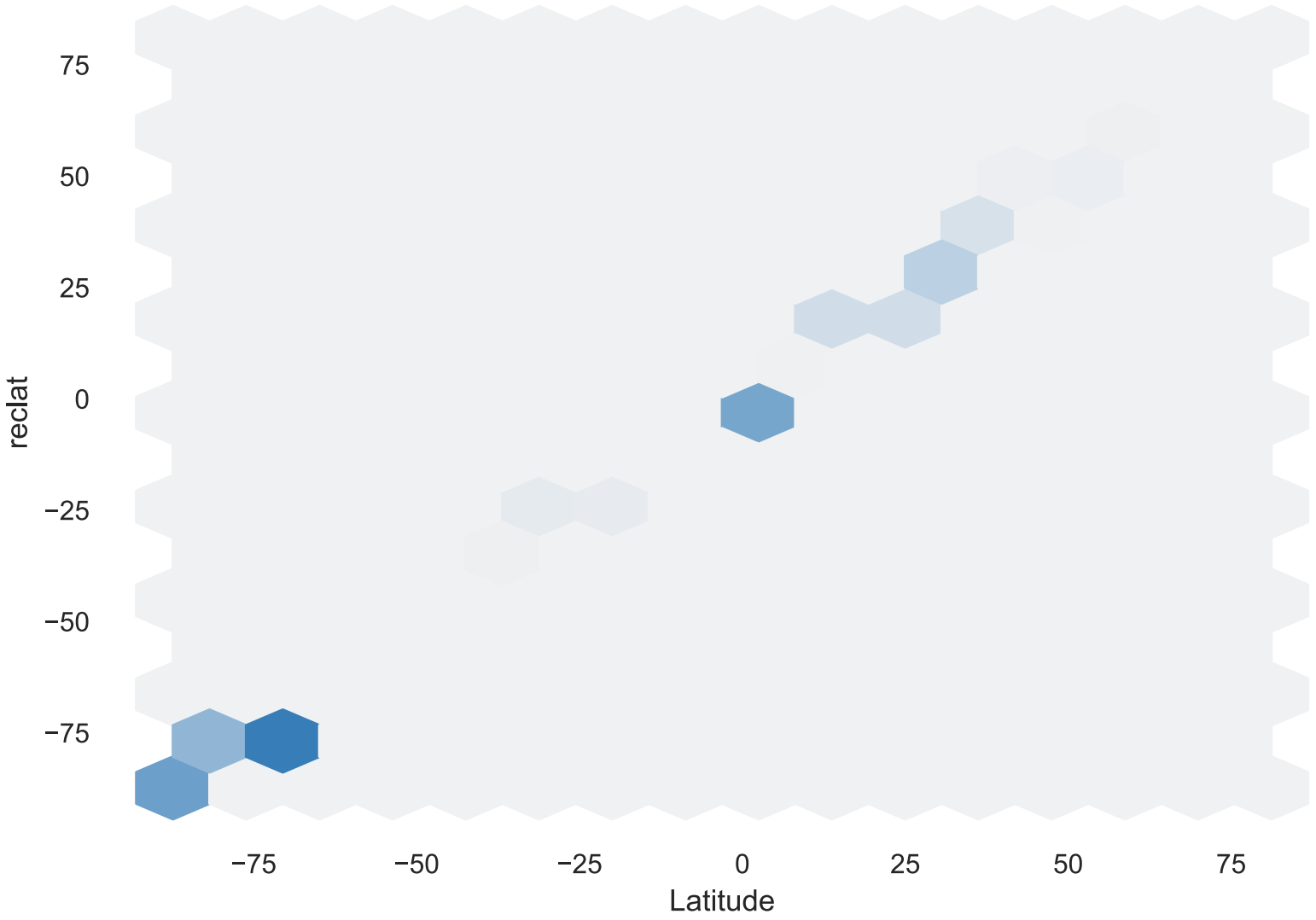

75
50
25
0
−25
−50
−75
Latitude
−100
0
100
200
300
reclong

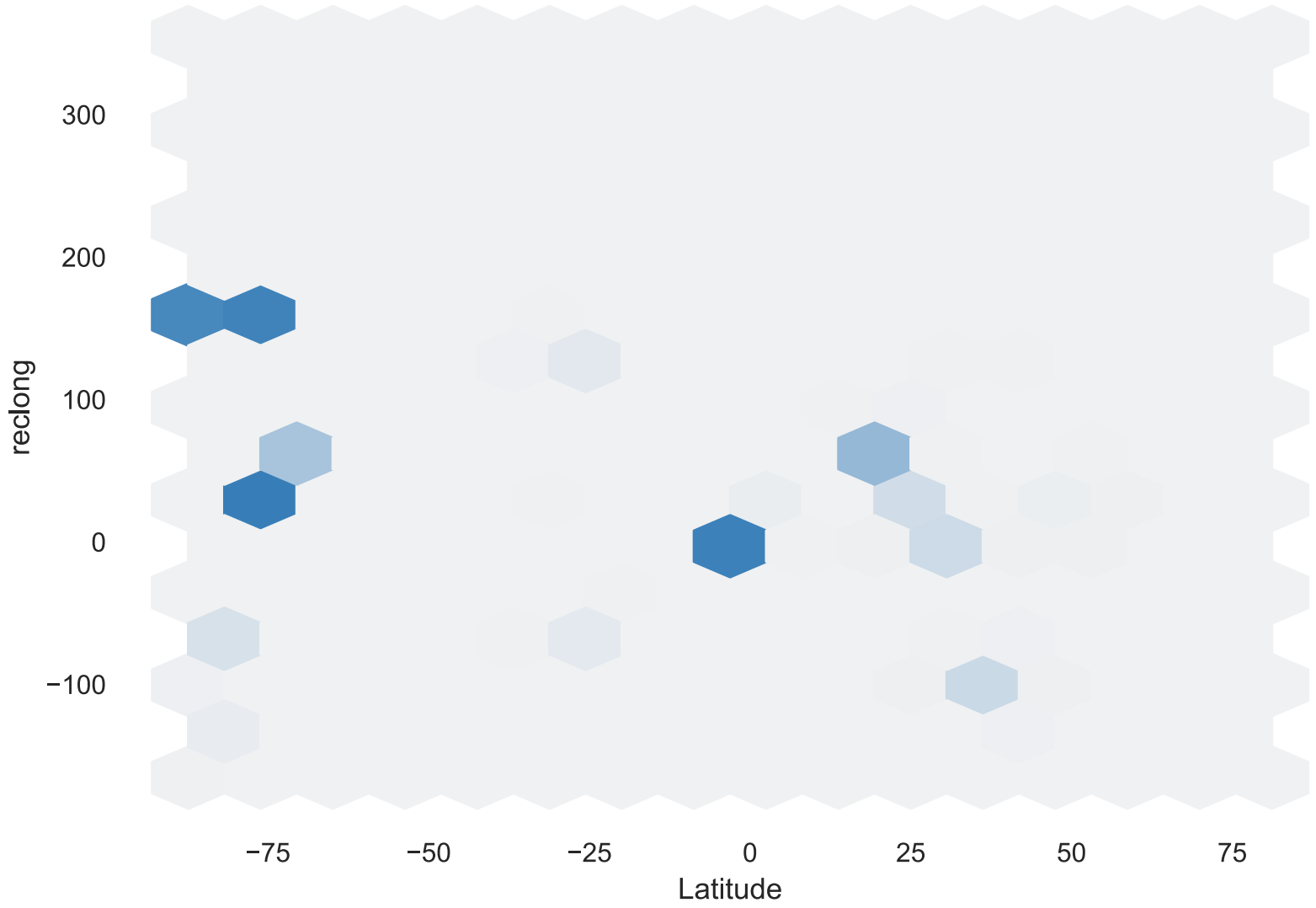

300
200
100
0
−100
Longitude
−75
−50
−25
0
25
50
75
Latitude

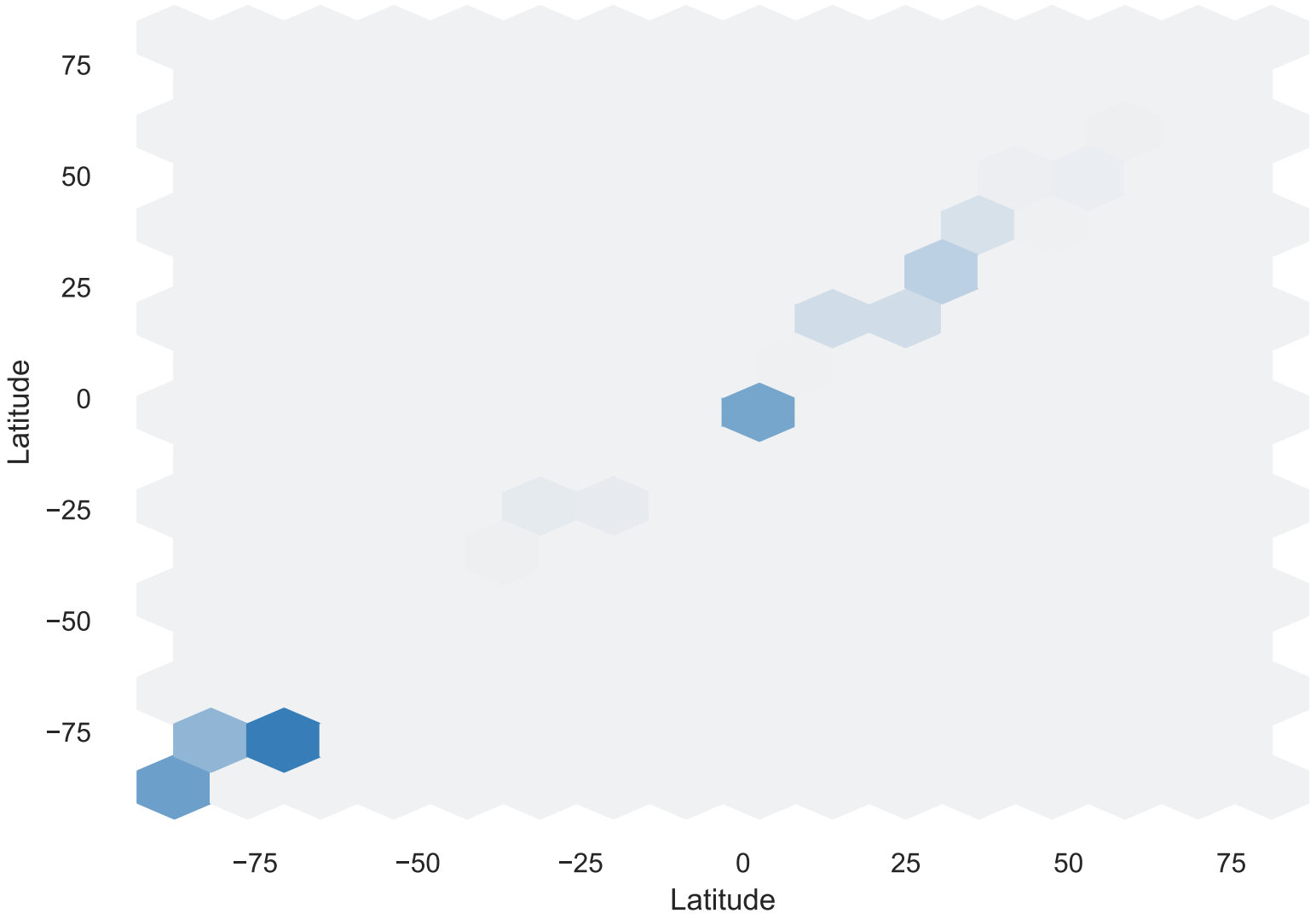

60000
50000
40000
30000
20000
10000
0
id
−75
−50
−25
0
25
50
75
Latitude

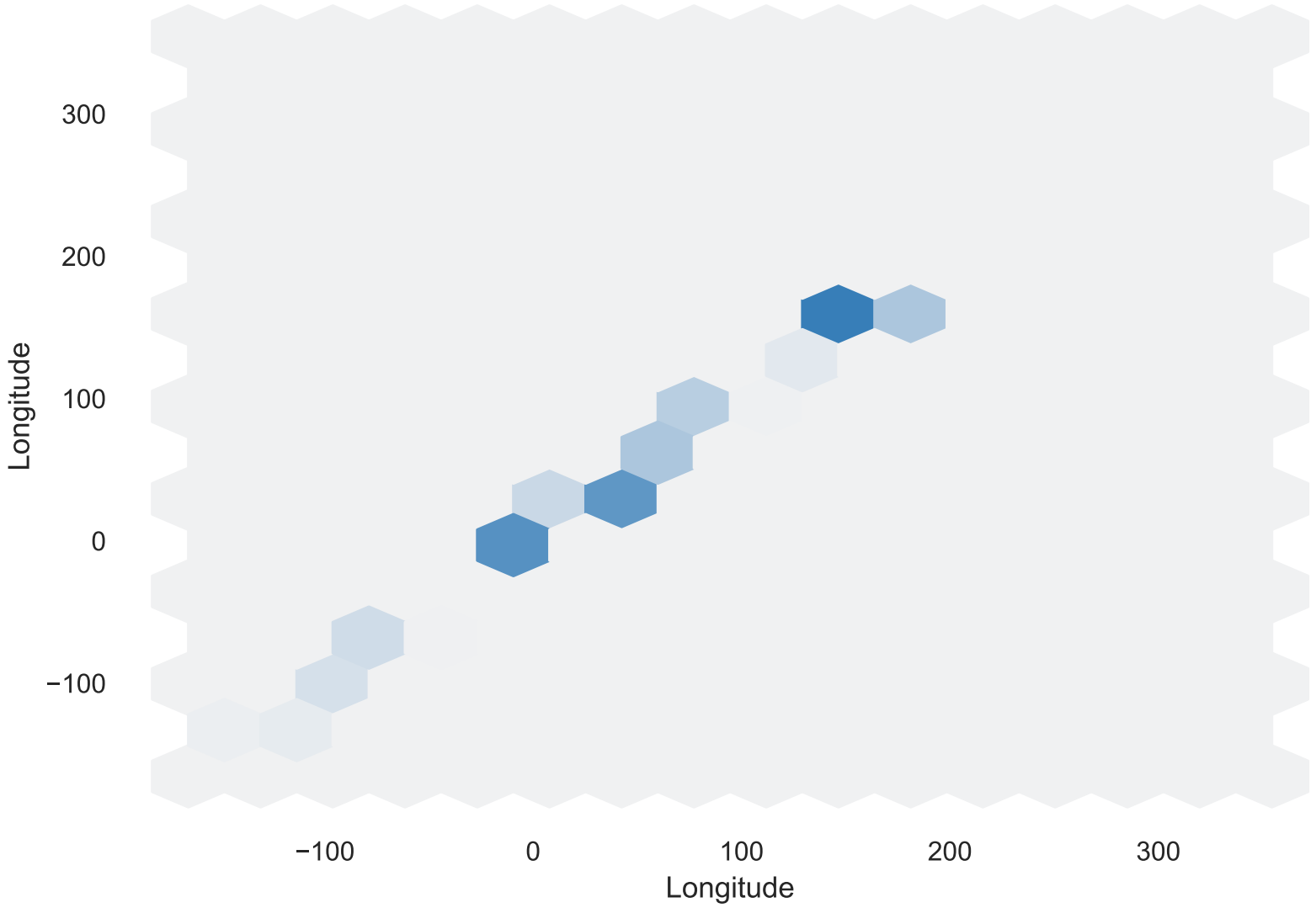

1e7
6
5
4
3
2
1
0
mass (g)
−75
−50
−25
0
25
50
75
Latitude

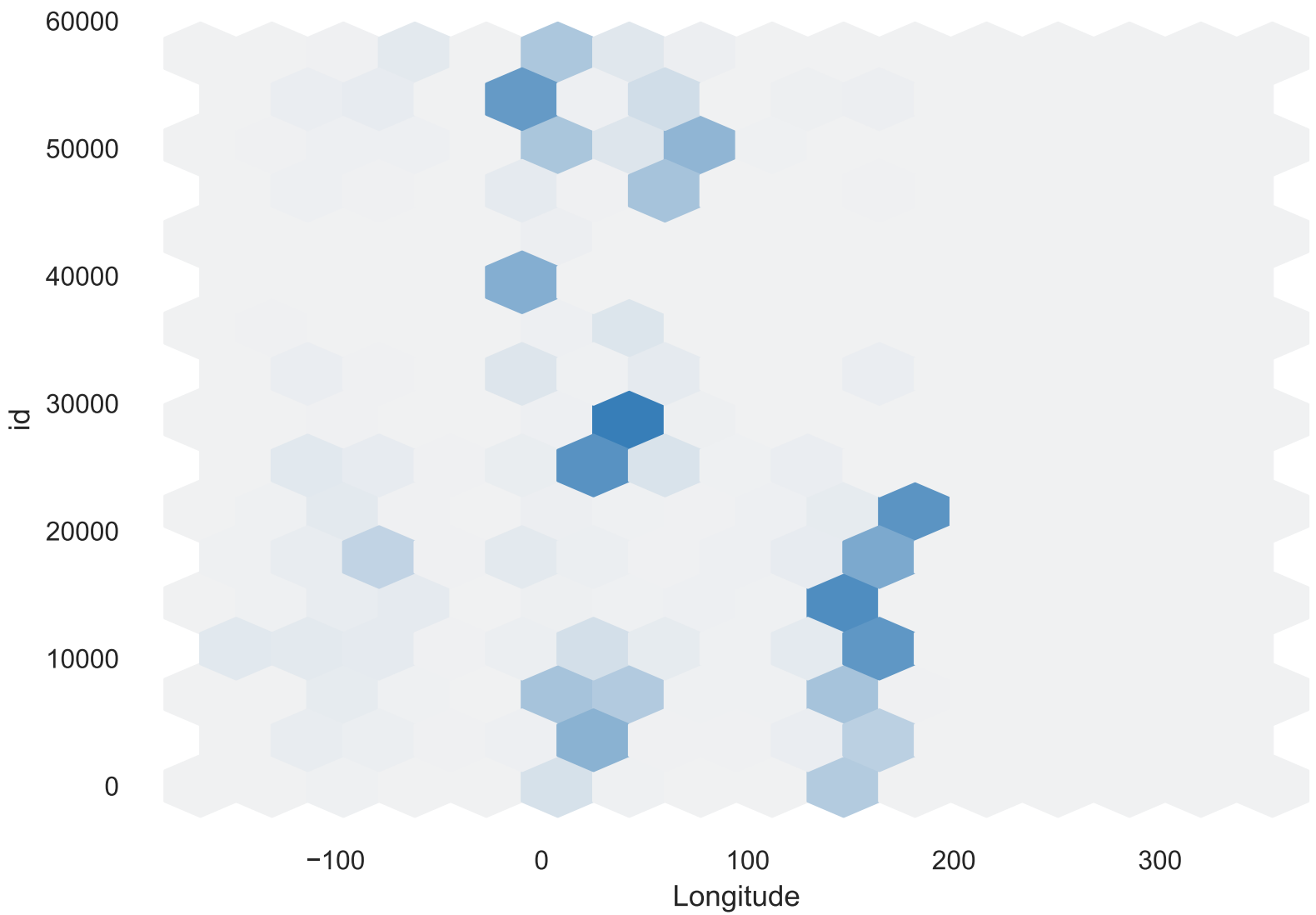

2000
1800
1600
1400
1200
1000
800
year
−75
−50
−25
0
25
50
75
Latitude

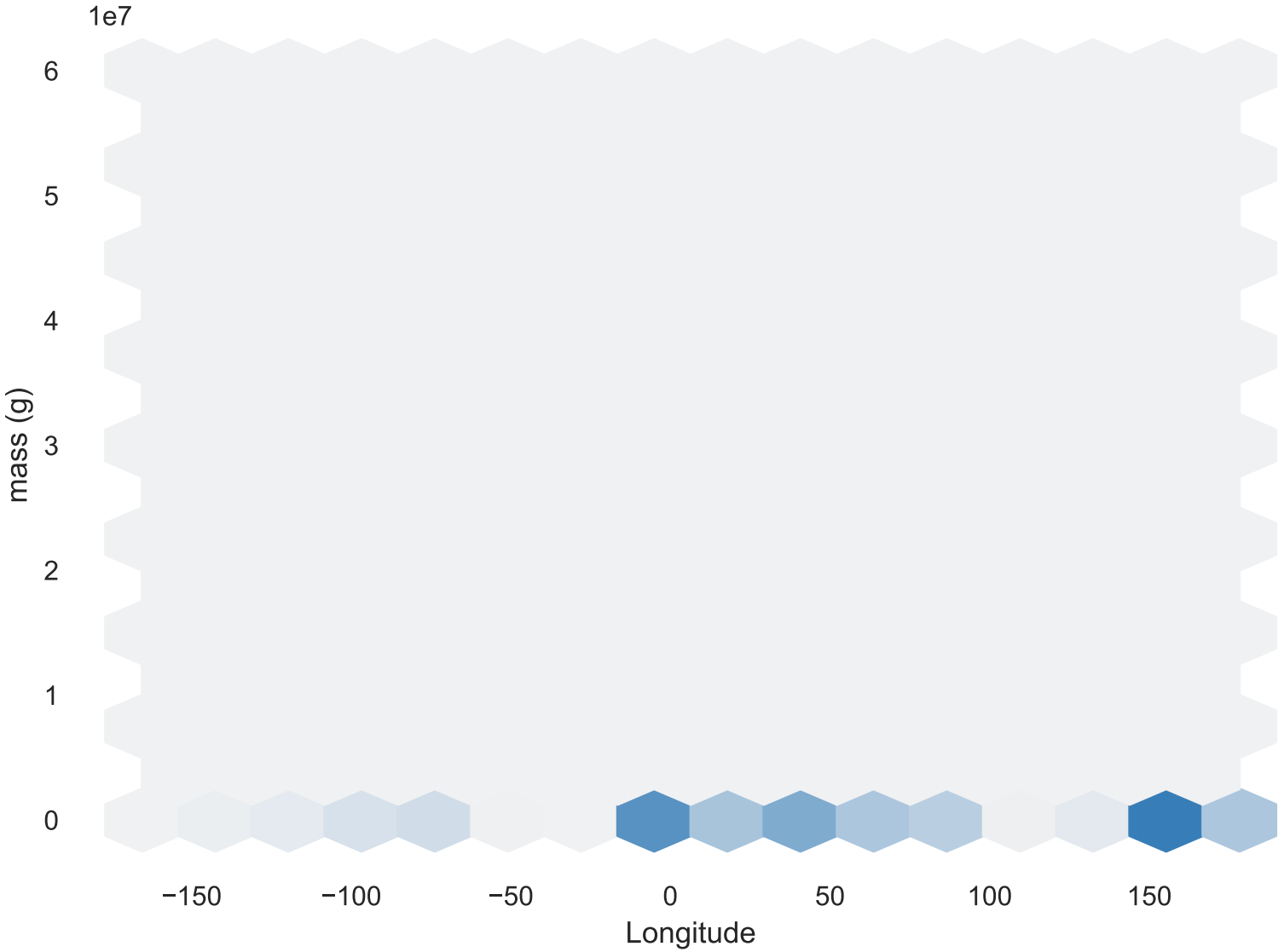

75
50
25
0
−25
−50
−75
reclat
−75
−50
−25
0
25
50
75
Latitude

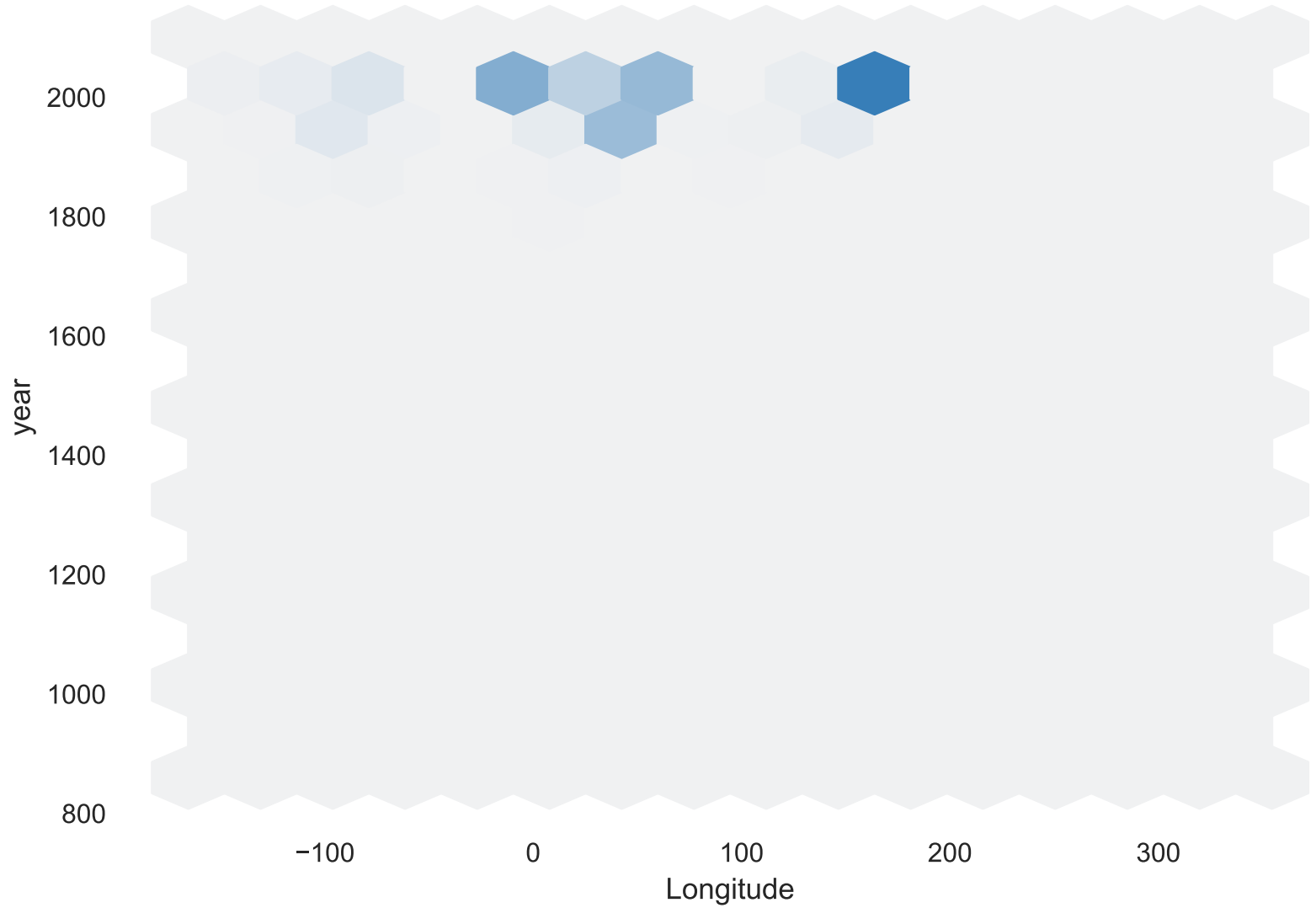

300
200
100
0
−100
reclong
−75
−50
−25
0
25
50
75
Latitude

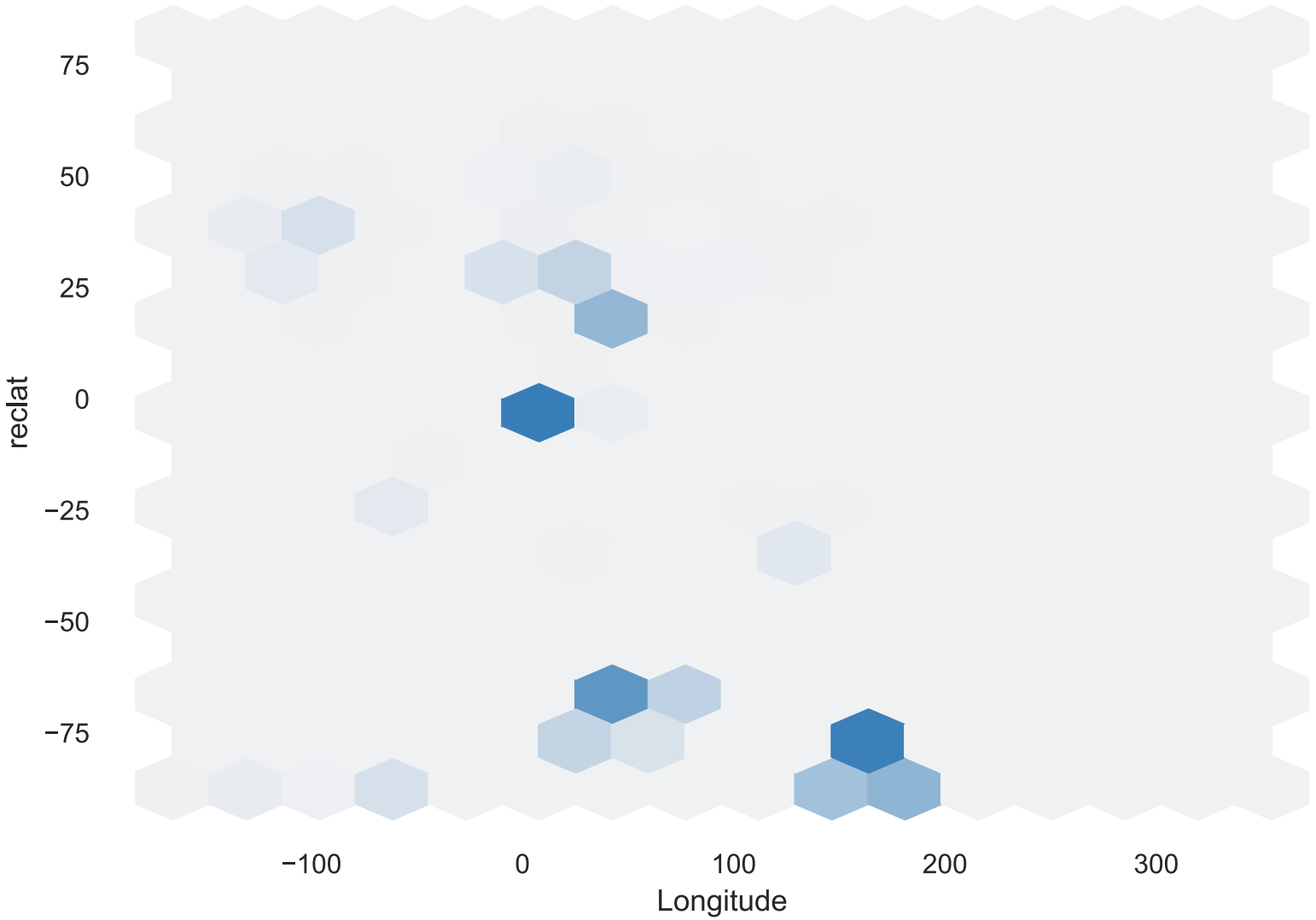

75
50
25
0
−25
−50
−75
Latitude
−75
−50
−25
0
25
50
75
Latitude

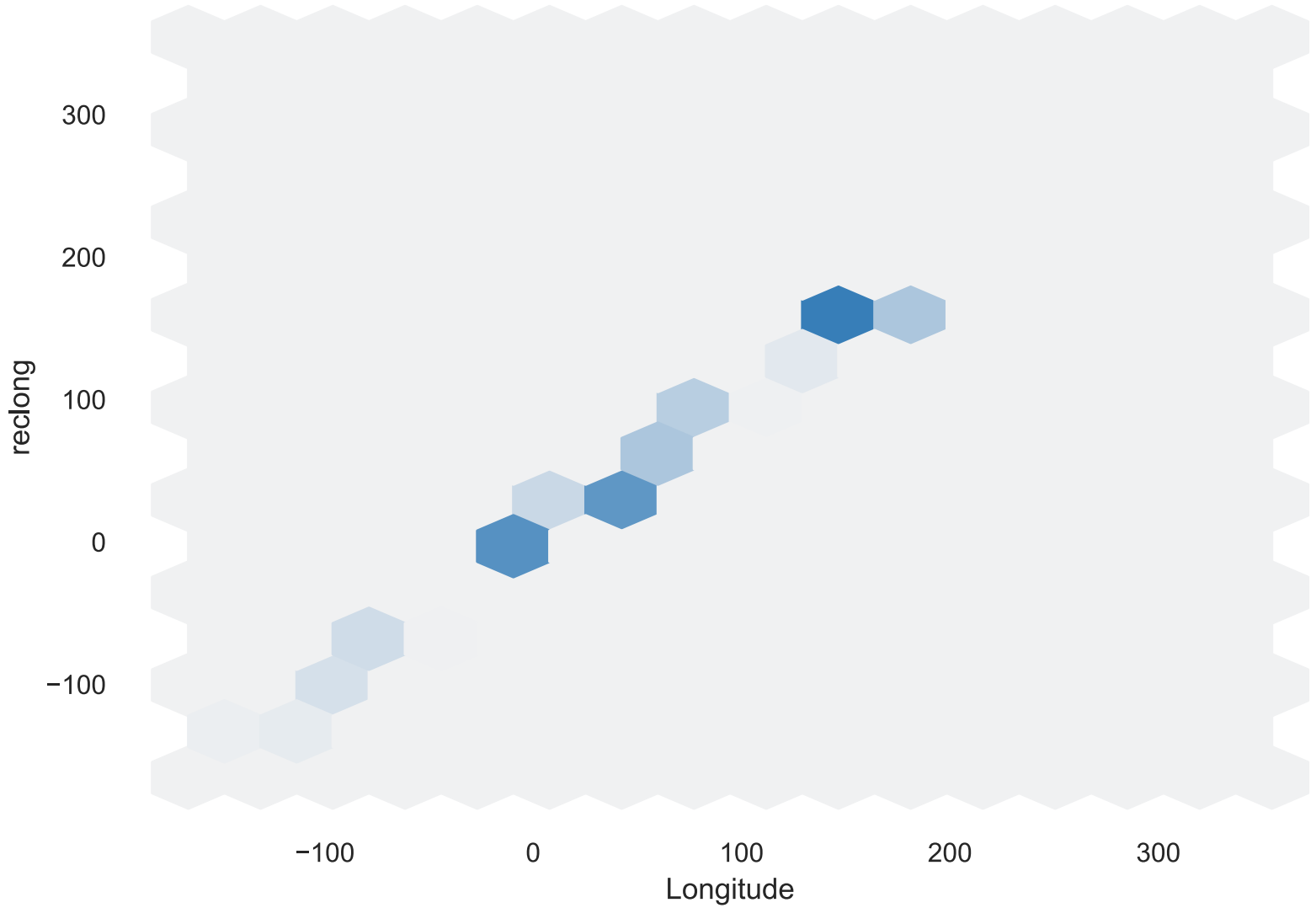

300
200
100
0
−100
Longitude
−100
0
100
200
300
Longitude

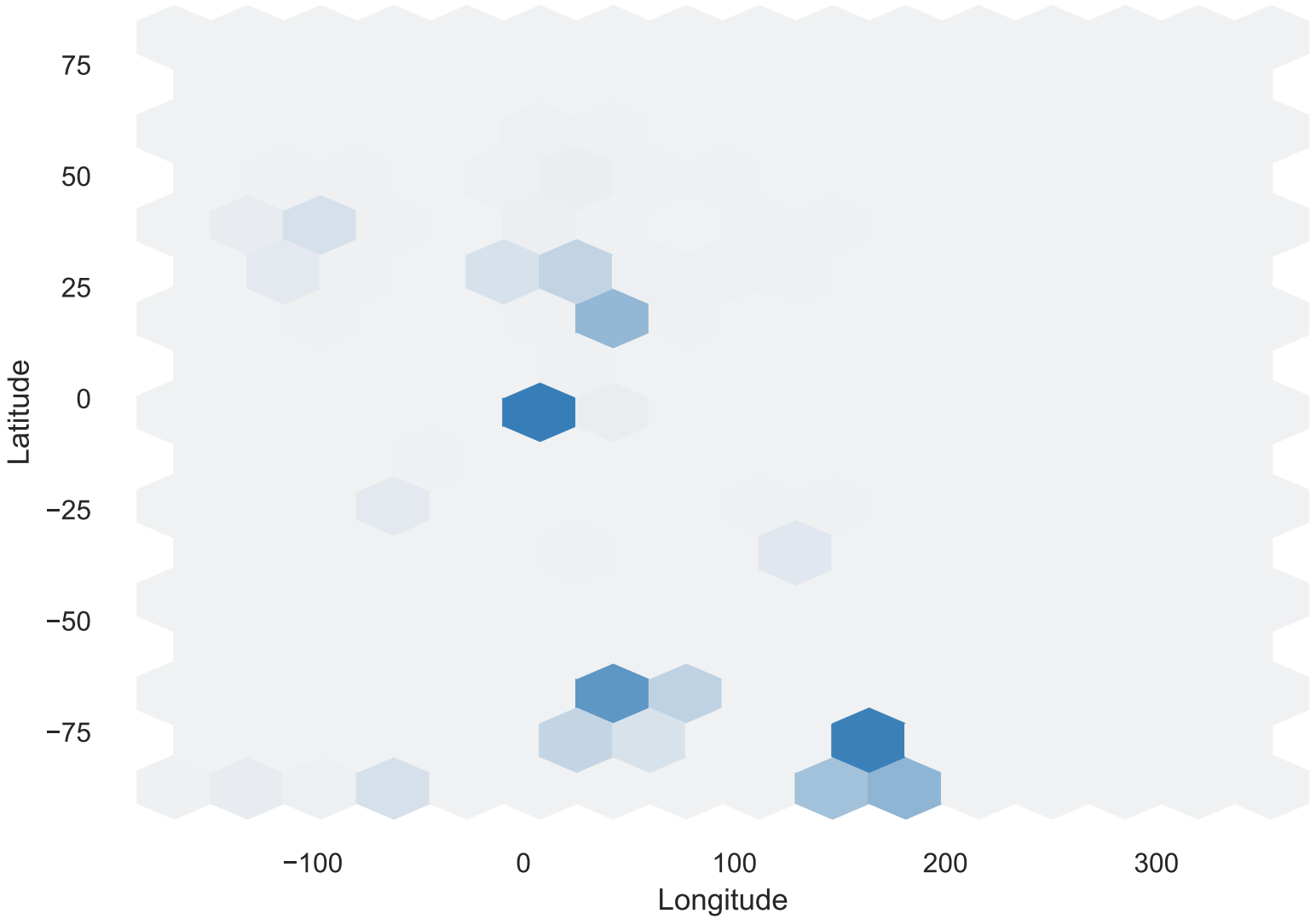

60000
50000
40000
30000
20000
10000
0
id
−100
0
100
200
300
Longitude


1e7
6
5
4
3
2
1
0
mass (g)
−150
−100
−50
0
50
100
150
Longitude

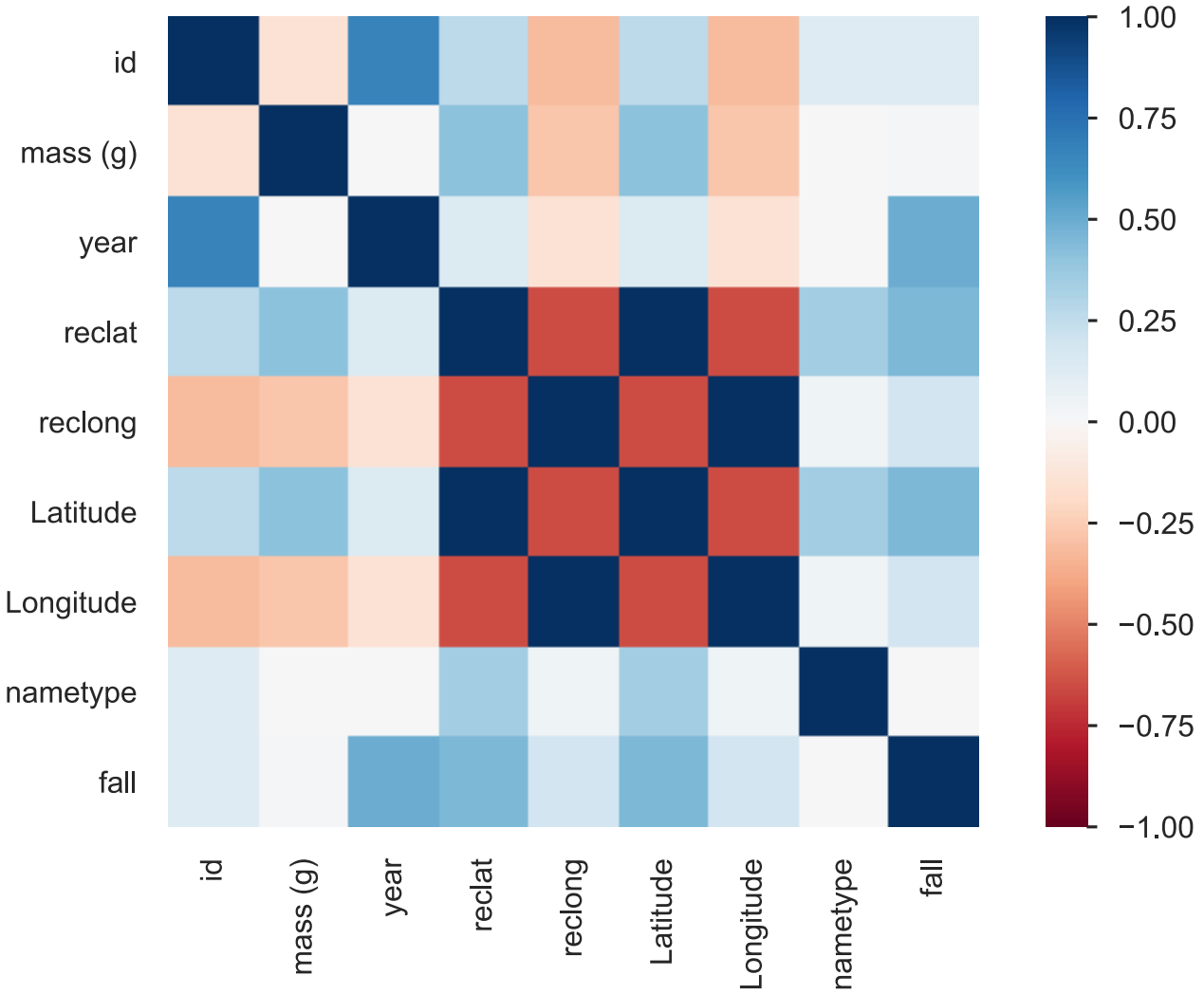

2000
1800
1600
1400
1200
1000
800
year
−100
0
100
200
300
Longitude


75
50
25
0
−25
−50
−75
reclat
−100
0
100
200
300
Longitude


300
200
100
0
−100
reclong
−100
0
100
200
300
Longitude

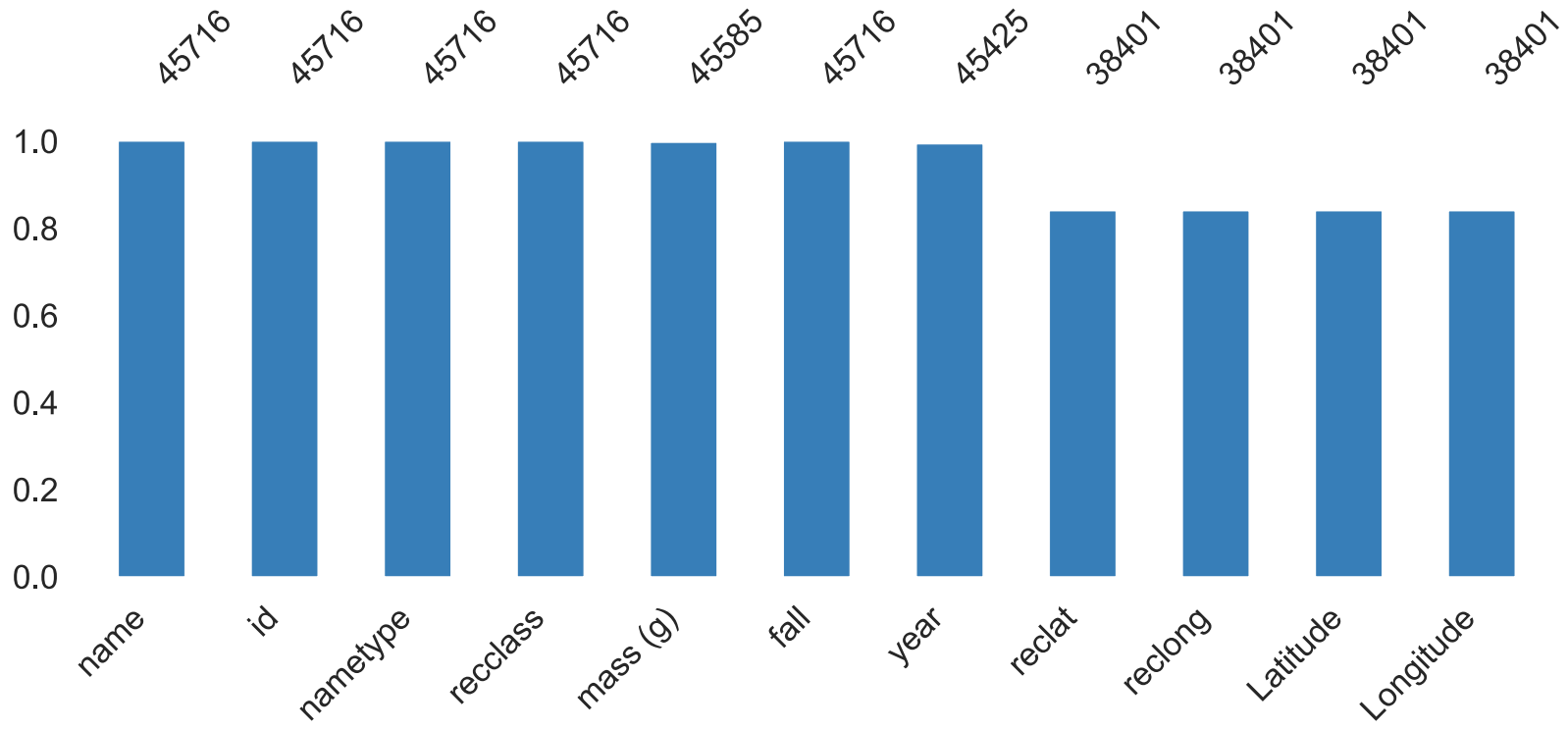

75
50
25
0
−25
−50
−75
Latitude
−100
0
100
200
300
Longitude

Correlations

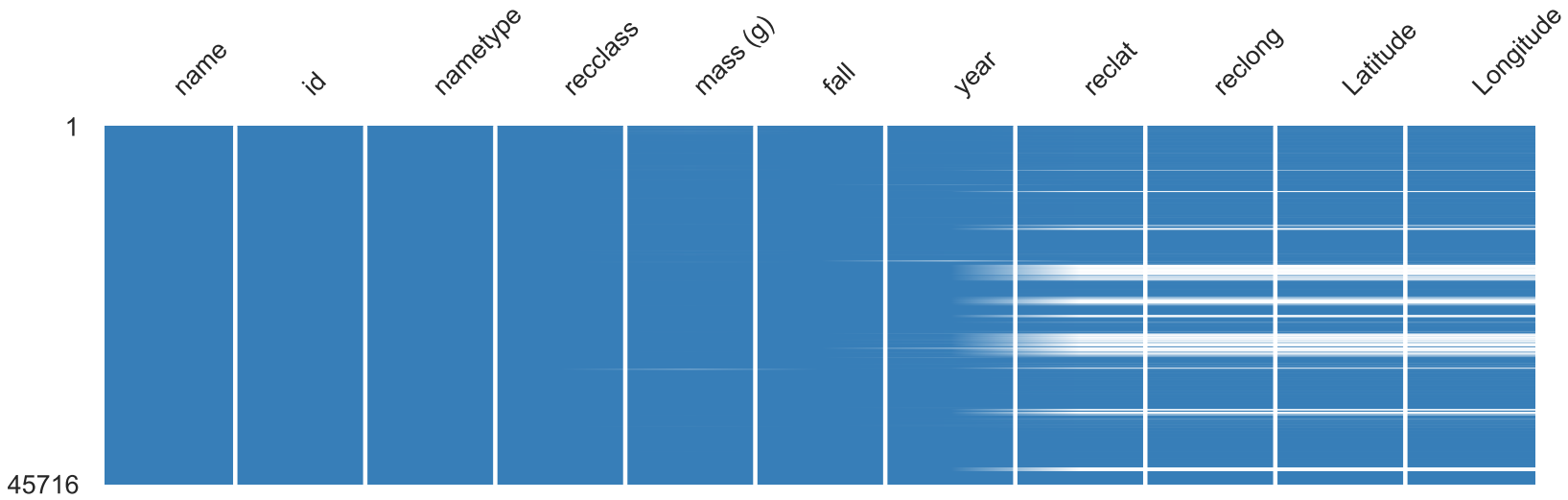

id
mass (g)
year
reclat
reclong
Latitude
Longitude
nametype
fall
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

	id	mass (g)	year	reclat	reclong	Latitude	Longitude	nametype	fall
id	1.000	-0.141	0.667	0.262	-0.317	0.262	-0.317	0.130	0.125
mass (g)	-0.141	1.000	-0.002	0.409	-0.281	0.409	-0.281	0.000	0.012
year	0.667	-0.002	1.000	0.140	-0.144	0.140	-0.144	0.002	0.494
reclat	0.262	0.409	0.140	1.000	-0.650	1.000	-0.650	0.350	0.449
reclong	-0.317	-0.281	-0.144	-0.650	1.000	-0.650	1.000	0.044	0.194
Latitude	0.262	0.409	0.140	1.000	-0.650	1.000	-0.650	0.350	0.449
Longitude	-0.317	-0.281	-0.144	-0.650	1.000	-0.650	1.000	0.044	0.194
nametype	0.130	0.000	0.002	0.350	0.044	0.350	0.044	1.000	0.000
fall	0.125	0.012	0.494	0.449	0.194	0.449	0.194	0.000	1.000

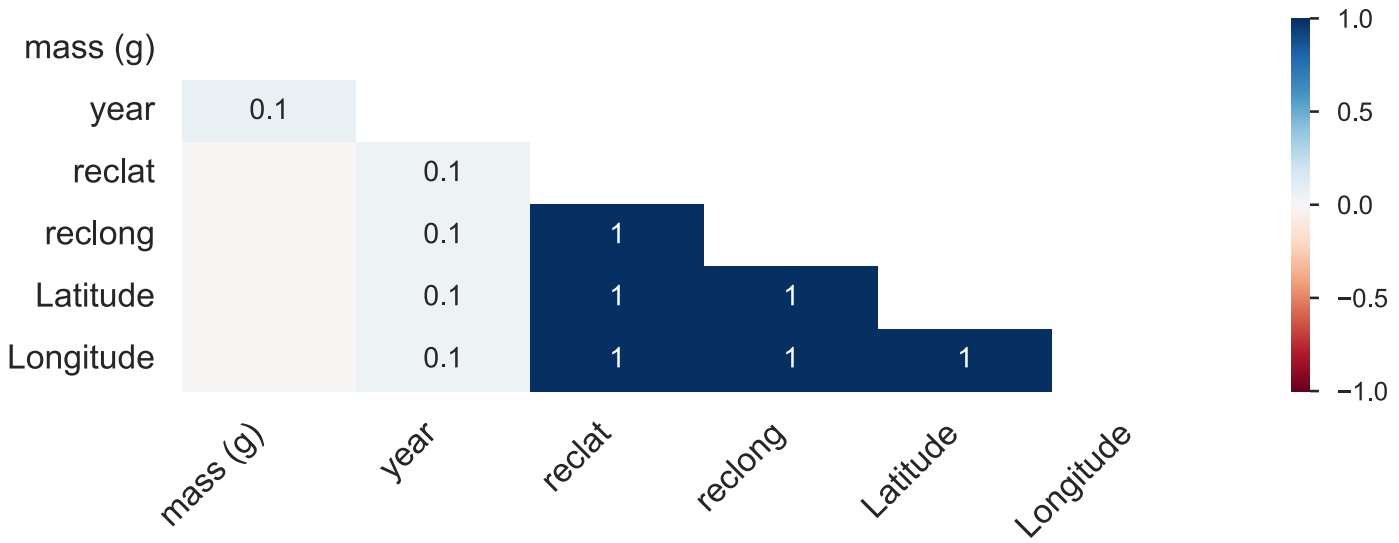
Missing values



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.



The correlation heatmap measures nullity correlation: how strongly the presence or absence of one variable affects the presence of another.

Sample

	name	id	nametype	recclass	mass (g)	fall	year	reclat	reclong	Latitude	Longitude
0	Aachen	1	Valid	L5	21.0	Fell	1880.0	50.77500	6.08333	50.77500	6.08333
1	Aarhus	2	Valid	H6	720.0	Fell	1951.0	56.18333	10.23333	56.18333	10.23333
2	Abee	6	Valid	EH4	107000.0	Fell	1952.0	54.21667	-113.00000	54.21667	-113.00000
3	Acapulco	10	Valid	Acapulcoite	1914.0	Fell	1976.0	16.88333	-99.90000	16.88333	-99.90000
4	Achiras	370	Valid	L6	780.0	Fell	1902.0	-33.16667	-64.95000	-33.16667	-64.95000
5	Adhi Kot	379	Valid	EH4	4239.0	Fell	1919.0	32.10000	71.80000	32.10000	71.80000
6	Adzhi-Bogdo (stone)	390	Valid	LL3-6	910.0	Fell	1949.0	44.83333	95.16667	44.83333	95.16667
7	Agen	392	Valid	H5	30000.0	Fell	1814.0	44.21667	0.61667	44.21667	0.61667
8	Aguada	398	Valid	L6	1620.0	Fell	1930.0	-31.60000	-65.23333	-31.60000	-65.23333
9	Aguila Blanca	417	Valid	L	1440.0	Fell	1920.0	-30.86667	-64.55000	-30.86667	-64.55000

	name	id	nametype	recclass	mass (g)	fall	year	reclat	reclong	Latitude	Longitude
45706	Zerkaly	31354	Valid	H5	16000.0	Found	1956.0	52.13333	81.96667	52.13333	81.96667
45707	Zhaoping	54609	Valid	Iron, IAB complex	2000000.0	Found	1983.0	24.23333	111.18333	24.23333	111.18333
45708	Zhigansk	30405	Valid	Iron, IIIAB	900000.0	Found	1966.0	68.00000	128.30000	68.00000	128.30000
45709	Zhongxiang	30406	Valid	Iron	100000.0	Found	1981.0	31.20000	112.50000	31.20000	112.50000
45710	Zillah 001	31355	Valid	L6	1475.0	Found	1990.0	29.03700	17.01850	29.03700	17.01850
45711	Zillah 002	31356	Valid	Eucrite	172.0	Found	1990.0	29.03700	17.01850	29.03700	17.01850
45712	Zinder	30409	Valid	Pallasite, ungrouped	46.0	Found	1999.0	13.78333	8.96667	13.78333	8.96667
45713	Zlin	30410	Valid	H4	3.3	Found	1939.0	49.25000	17.66667	49.25000	17.66667
45714	Zubkovsky	31357	Valid	L6	2167.0	Found	2003.0	49.78917	41.50460	49.78917	41.50460
45715	Zulu Queen	30414	Valid	L3.7	200.0	Found	1976.0	33.98333	-115.68333	33.98333	-115.68333

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