Final project in Bayesian Statistics

Faculty of Sciences

Daniel Felipe Villa Rengifo

CC. 1005087556

UNIVERSIDAD NACIONAL DE COLOMBIA

Solar Flare Data Set

UNIVERSIDAD NACIONAL DE COLOMBIA

The database is maintained by the University of California at Irvine (UCI) as a survey of solar flares of a given class (we will only focus on two classes for this study) that occur every 24 hours, i.e., each row represents one survey of solar flares presented per day.

Program start year: 1989

Solar Flare Data Set

UNIVERSIDAD NACIONAL DE COLOMBIA

This database has 2698 *(1389 original obs.)* observations and 10 variables that are represented as:

1. Class code (modified Zurich class) (A,B,C,D,E,F,H)

2. Code of the size of the largest spot (X,R,S,A,H,K)

3. Code for the distribution of the stain (X,O,I,C)

4. Activity (1 = reduced, 2 = no change)

5. Evolution (1 = decay, 2 = no growth, 3 = growth)

Solar Flare Data Set

UNIVERSIDAD NACIONAL DE COLOMBIA

6. Activity code of previous 24-hour eruption
(1 = nothing as large as an M1, 2 = an M1, 3 = more activity than an M1).

7. Historically complex (1 = Yes, 2 = No)

8. Region became historically complex in this solar disk pass (1 = yes, 2 = no)

9. Area (1 = small, 2 = large)

10. Area of largest spot (1 = <=5, 2 = >5)

Solar Flare Data Set

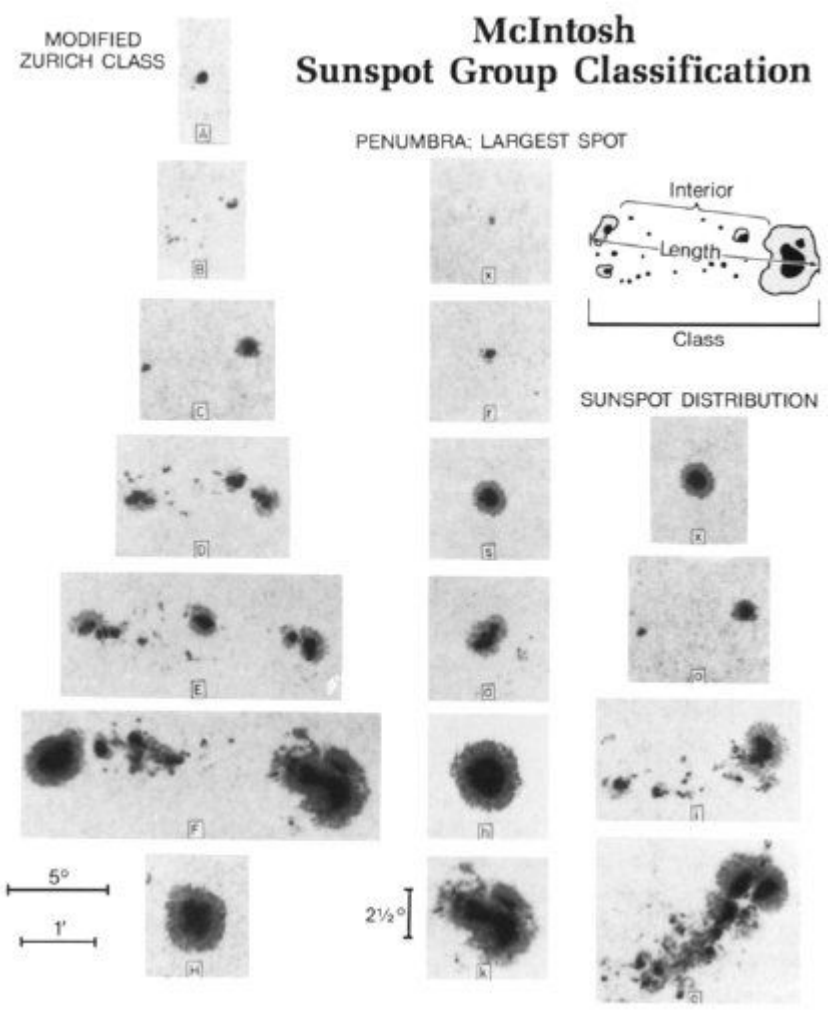
UNIVERSIDAD NACIONAL DE COLOMBIA

class	penumbra	distribution	activity	evolution	flare24.activity	history	sun.disk	area	largest.spot	c.class	m.class	x.class
C	S	O	1	2	1	1	2	1	2	0	0	0
D	S	O	1	3	1	1	2	1	2	0	0	0
C	S	O	1	3	1	1	2	1	1	0	0	0

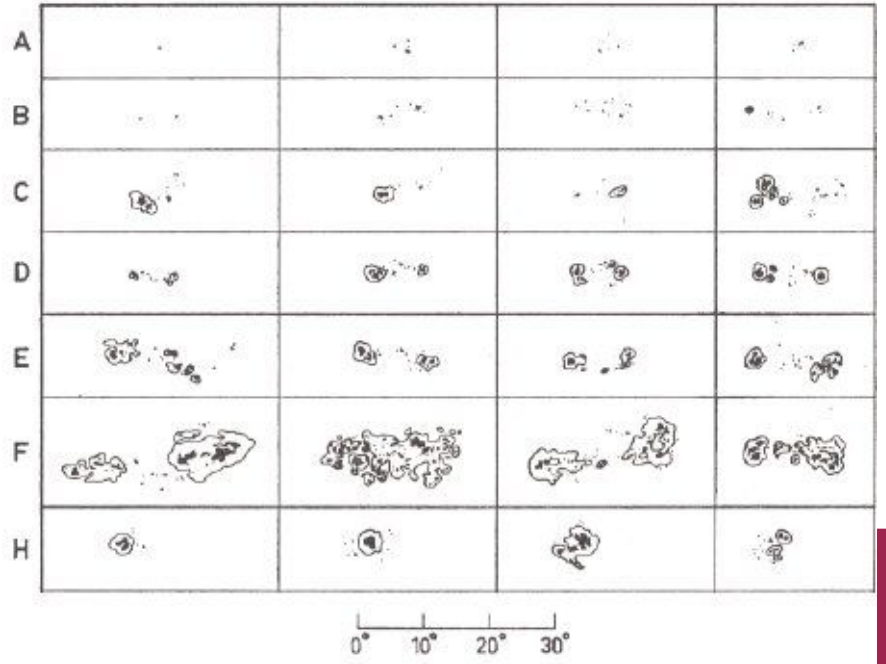
Solar Flare Data Set

UNIVERSIDAD NACIONAL DE COLOMBIA

n	class	penumbra	distribution	activity	evolution	flare24.activity	history	sun.disk	area	largest.spot	class_flare
1	F	X	C	1	3	1	1	1	1	1	**M**
2	C	A	I	1	3	1	1	2	1	1	**M**
3	B	H	I	1	3	1	1	2	1	1	**M**
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
2696	D	X	I	1	3	1	1	2	1	2	**X**
2697	B	A	I	1	3	1	1	2	1	1	**X**
2698	D	S	X	1	3	1	1	2	2	2	**X**



UNIVERSIDAD
NACIONAL
DE COLOMBIA



Objective

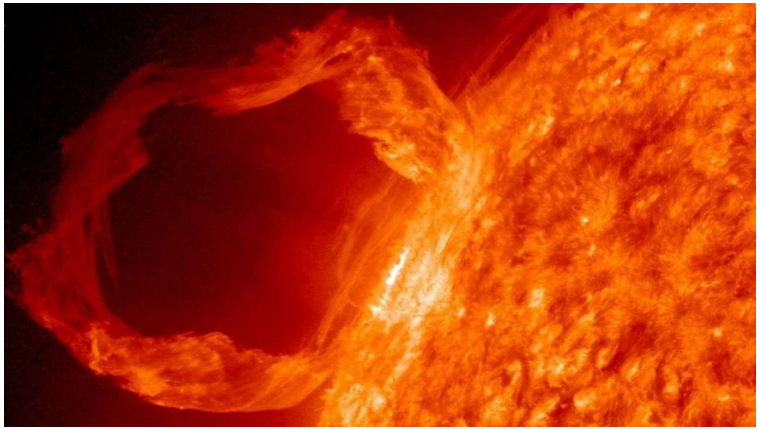
UNIVERSIDAD NACIONAL DE COLOMBIA

To propose a model to predict one of the two most obvious cases of solar flares (Class M or X), that is, based on study variables that help us classify and quantify the area, evolution, history, as well as its shape and aspects, to obtain a possibility that an eruption of that class will occur given these characteristics.

issues of interest

UNIVERSIDAD NACIONAL DE COLOMBIA

Will the variable that gathered data prior to taking the eruption record have a large influence?



Is the Zurich method of classifying eruptions adequate for determining the type?

Do the rows that became complex after 24 hours affect the model's prediction accuracy?

Selected Variables

UNIVERSIDAD NACIONAL DE COLOMBIA

Because some variables are less influential (NASA, visibleearth), we will select those that are physically significant, plus the addition of a new variable that indicates the type of eruption.

Selected Variables

UNIVERSIDAD NACIONAL DE COLOMBIA

1. Class code (modified Zurich class) (A,B,C,D,E,F,H)

2. Code of the size of the largest spot (X,R,S,A,H,K)

3. Code for the distribution of the stain (X,O,I,C)

4. Activity (1 = reduced, 2 = no change)

5. Evolution (1 = decay, 2 = no growth, 3 = growth)

6. Activity code of previous 24-hour eruption
(1 = nothing as large as an M1, 2 = an M1, 3 = more activity than an M1).

7. Historically complex (1 = Yes, 2 = No)

8. Region became historically complex in this solar disk pass
(1 = yes, 2 = no)

Proposed Model

UNIVERSIDAD
NACIONAL
DE COLOMBIA

To provide a solution and to predict this type of phenomenon in space, a binary logistic model was chosen.

UNIVERSIDAD NACIONAL DE COLOMBIA

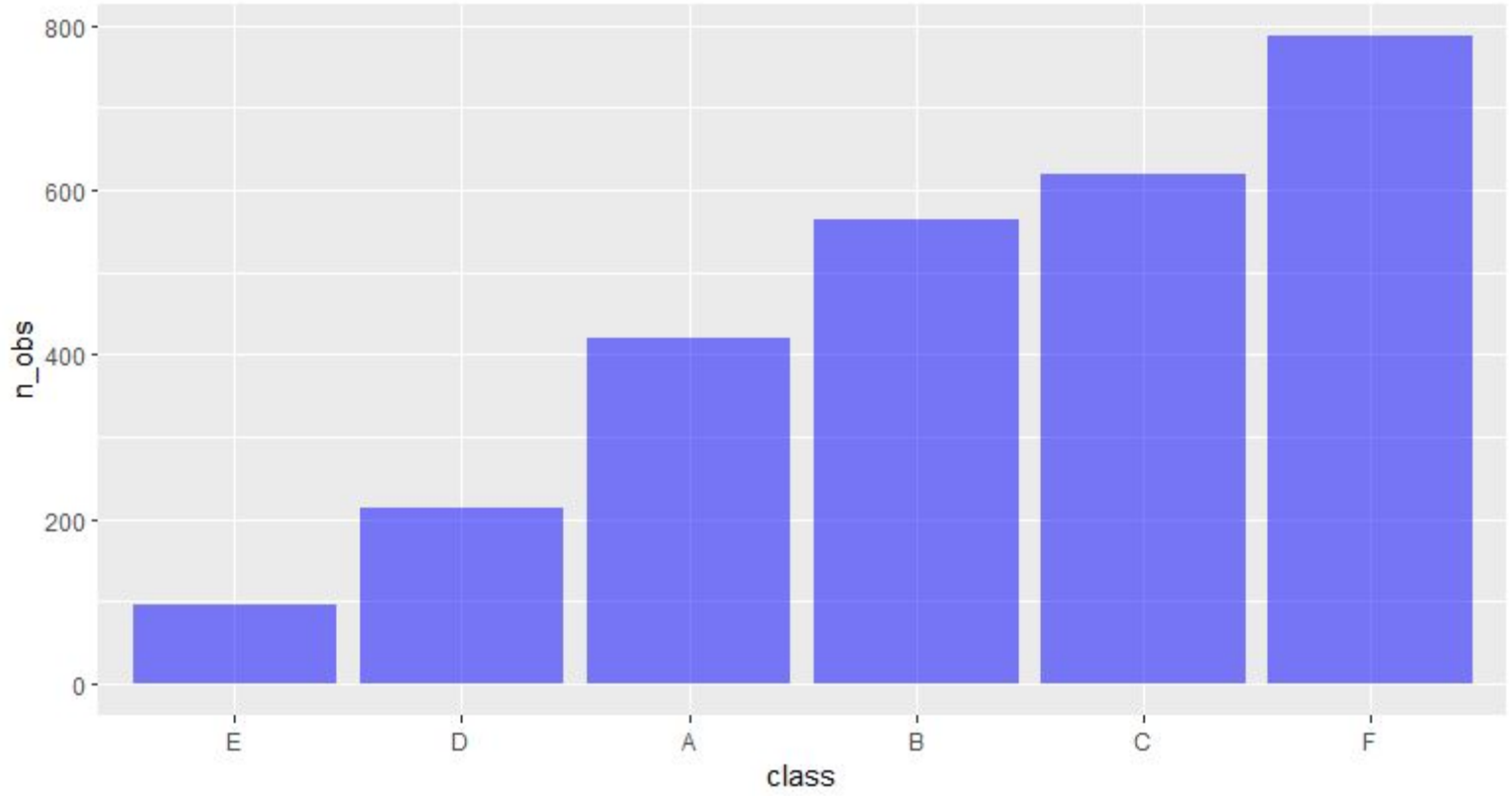
$$y_i | \pi_i \sim Bernoulli(p)$$

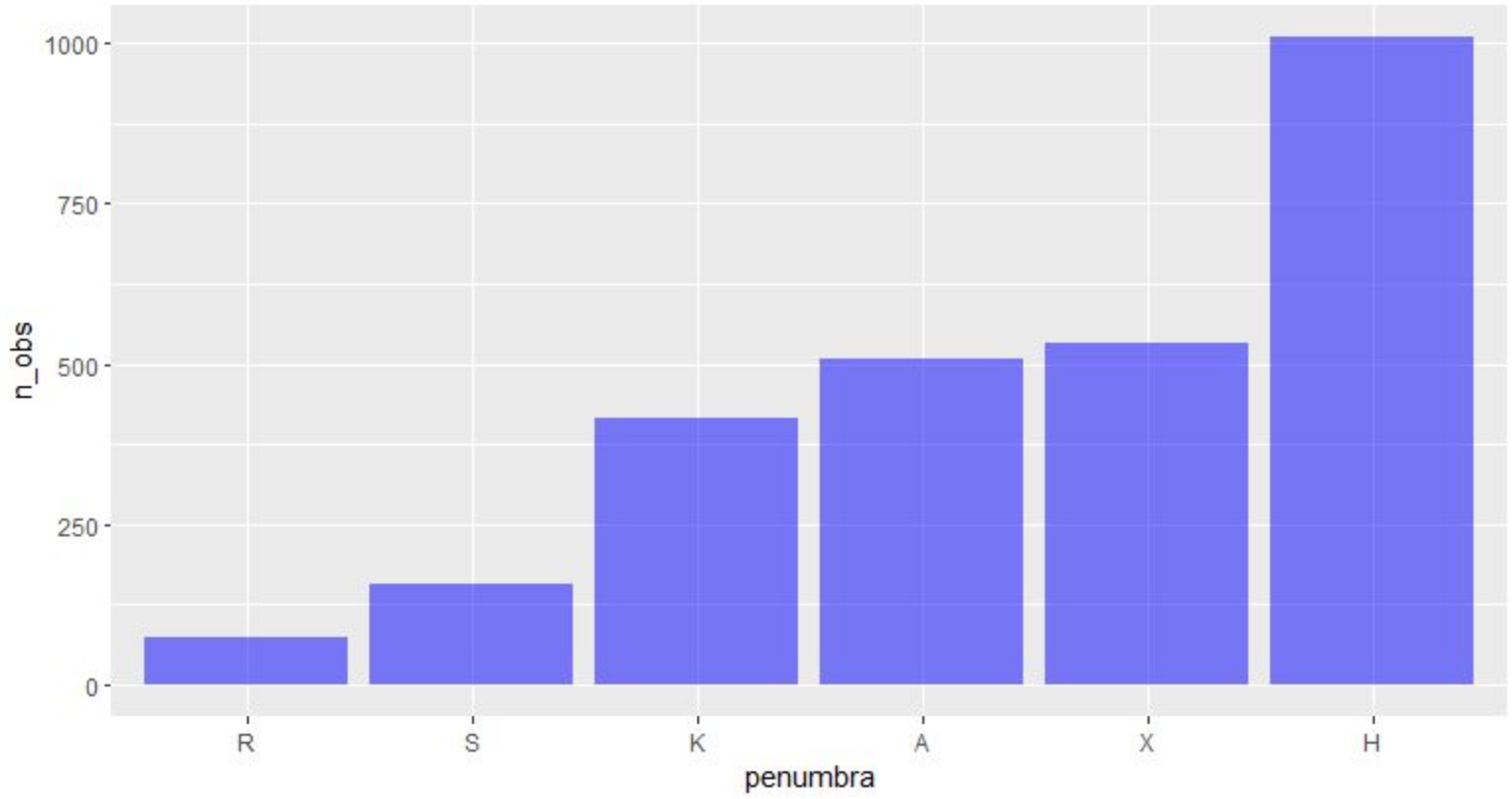
$$Y = \begin{cases} 1 & \text{If it is a class X eruption} \\ 0 & \text{If it is a class M eruption} \end{cases}$$

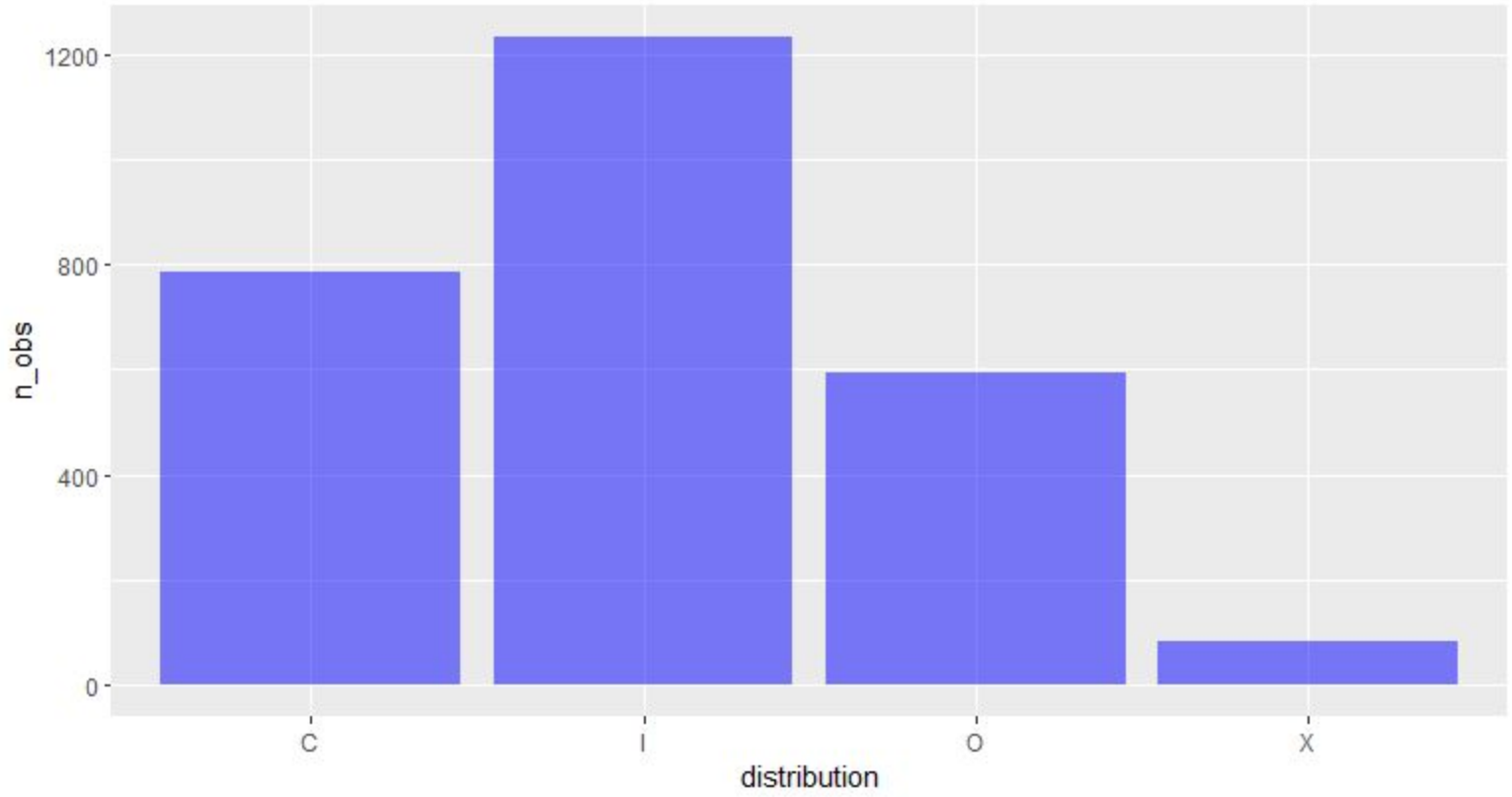
$$\pi_i = P\{y_i = 1\} = \frac{exp(x_i^t * \beta)}{1 + exp(x_i^t * \beta)}$$

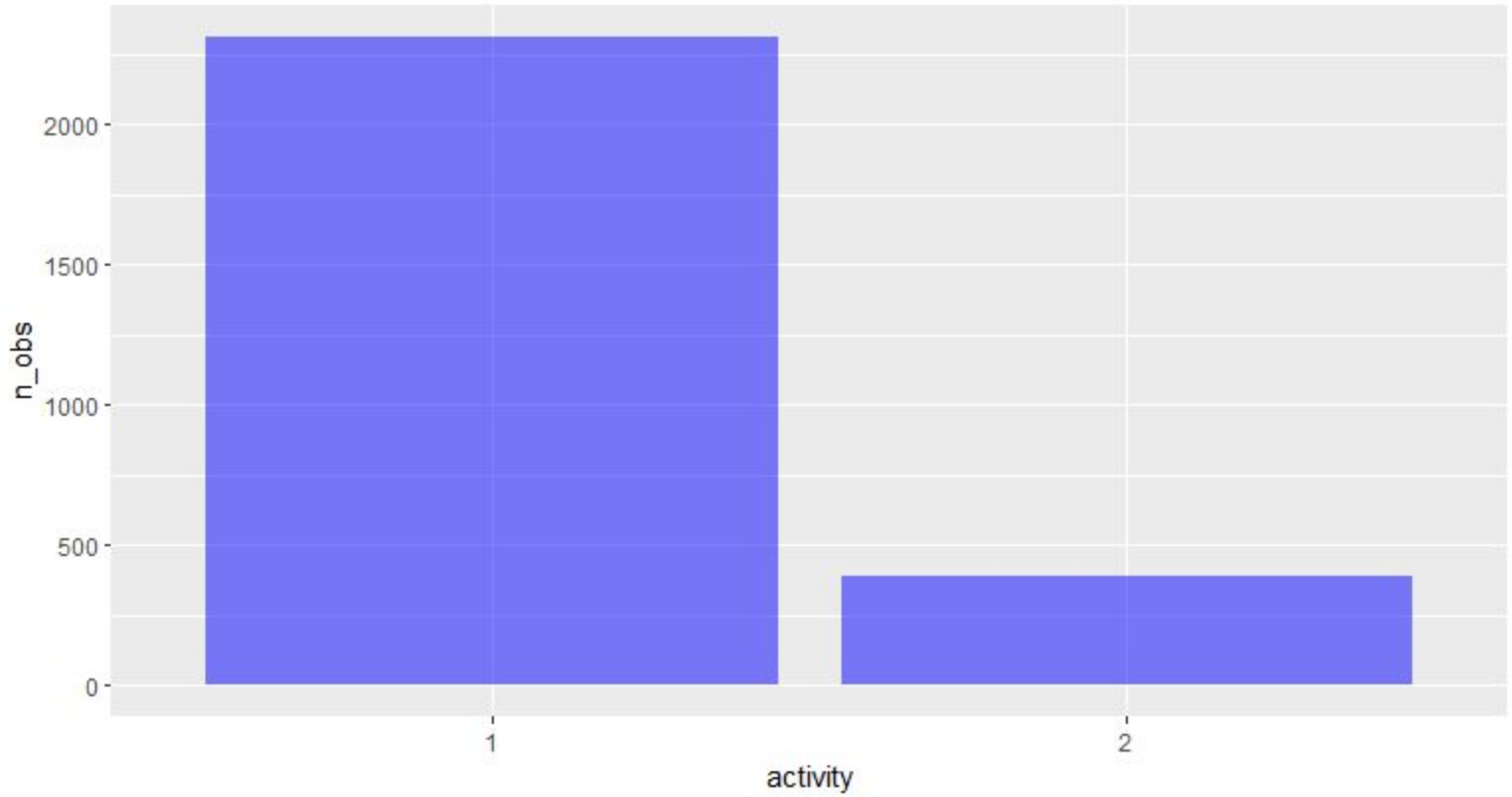
Descriptive Statistics

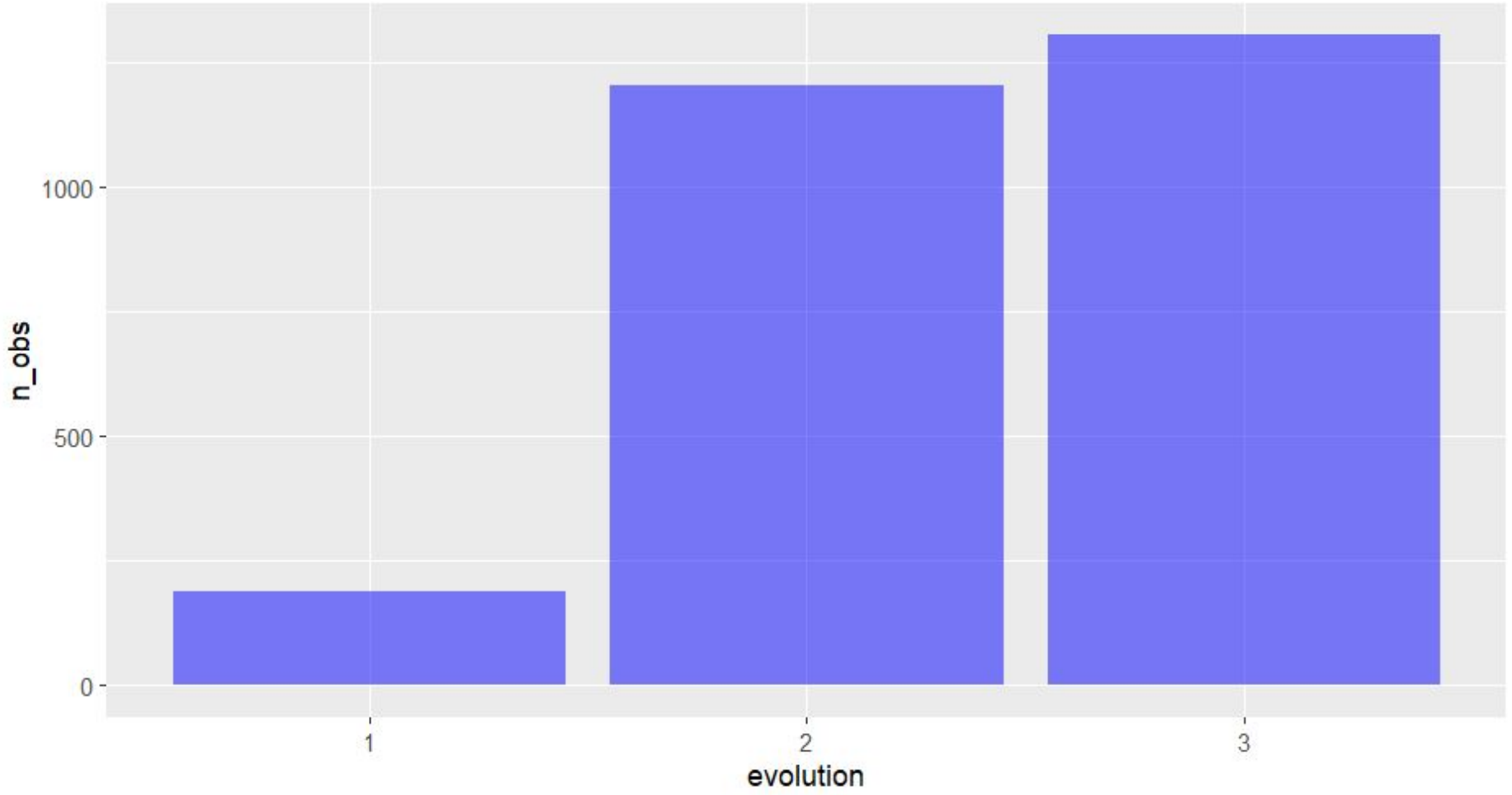
UNIVERSIDAD
NACIONAL
DE COLOMBIA

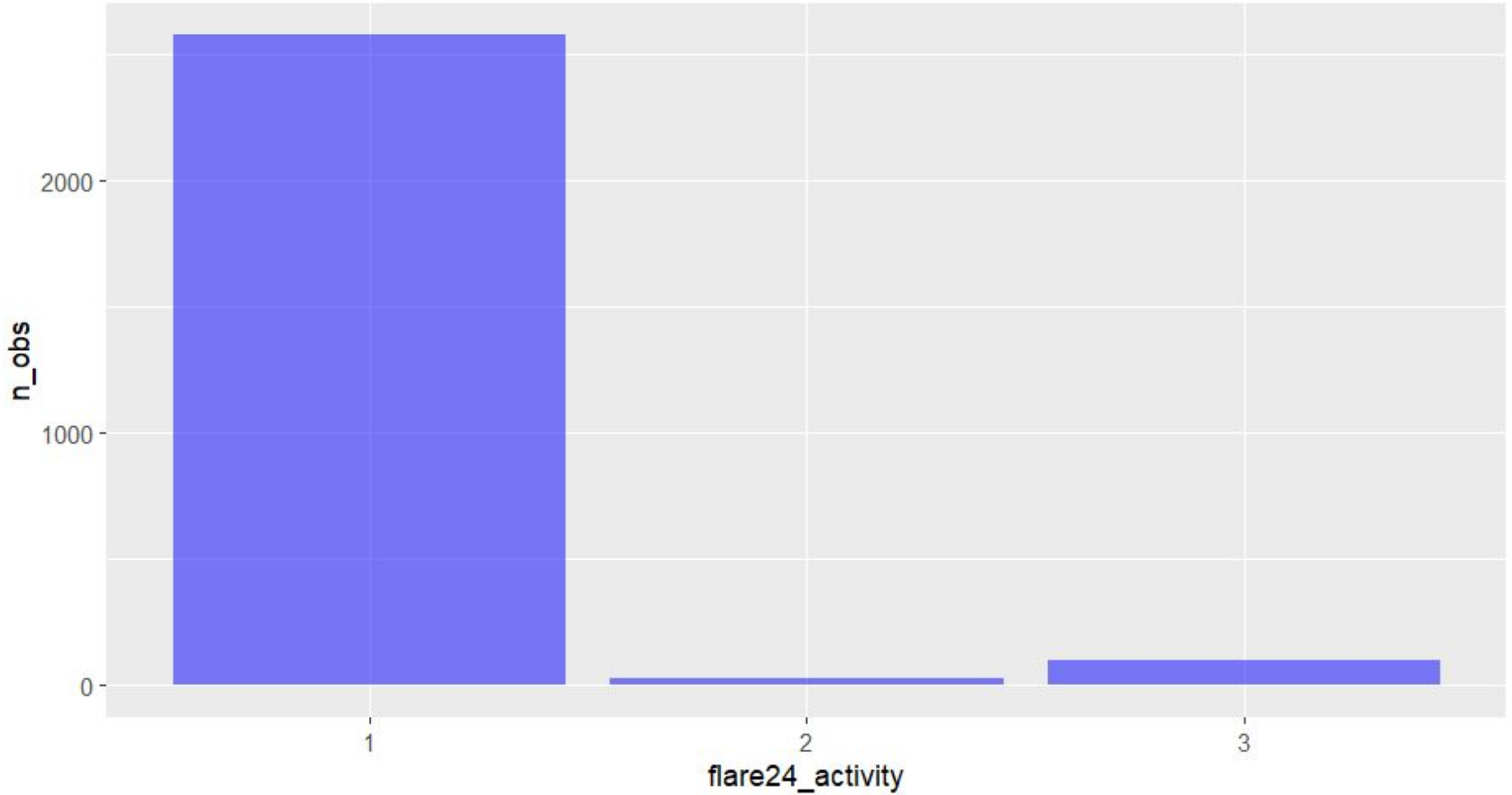

800
600
400
200
0
n_obs
E
D
A
B
C
F
class

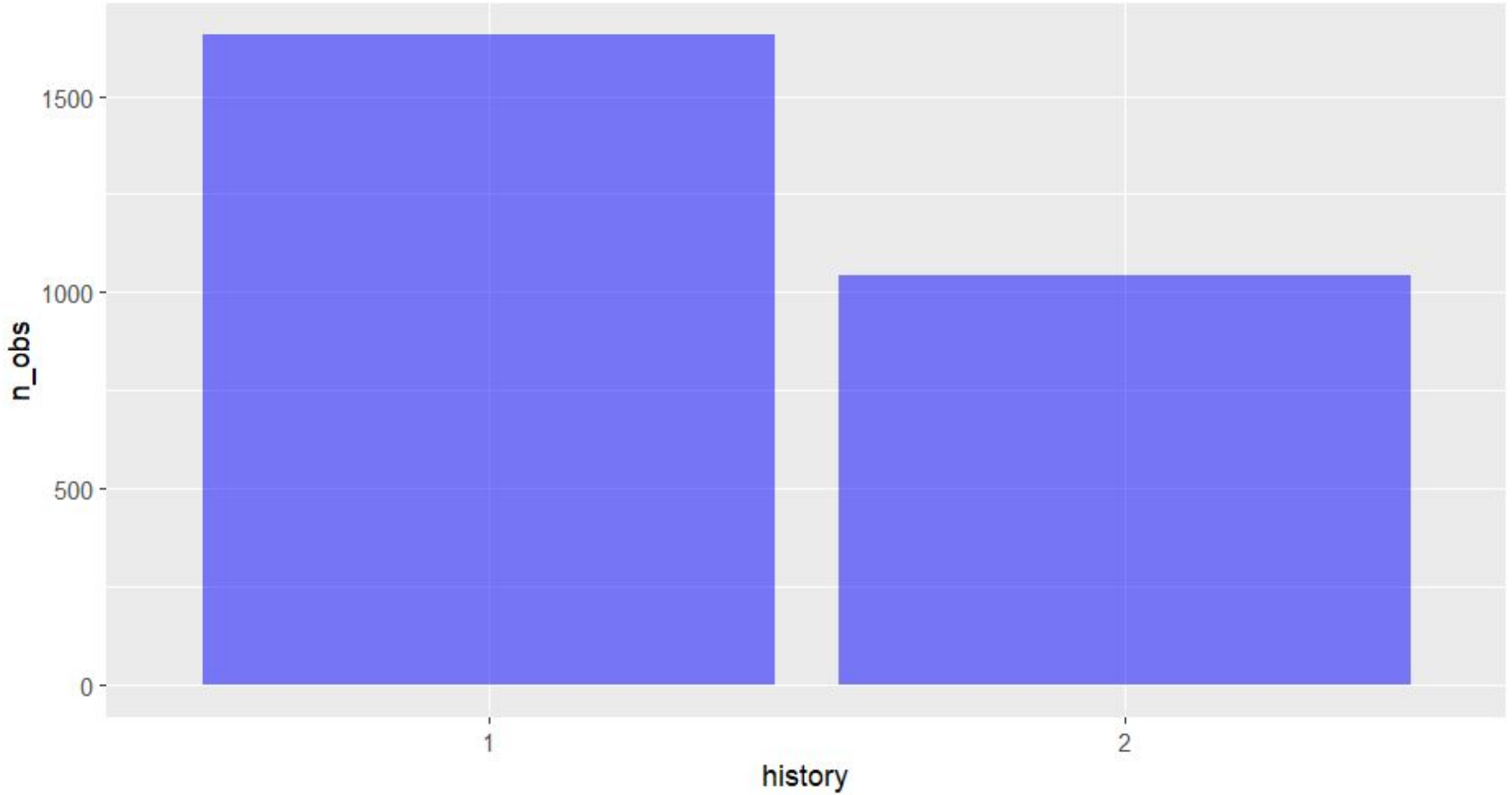

1000
750
500
250
0
n_obs
R
S
K
A
X
H
penumbra

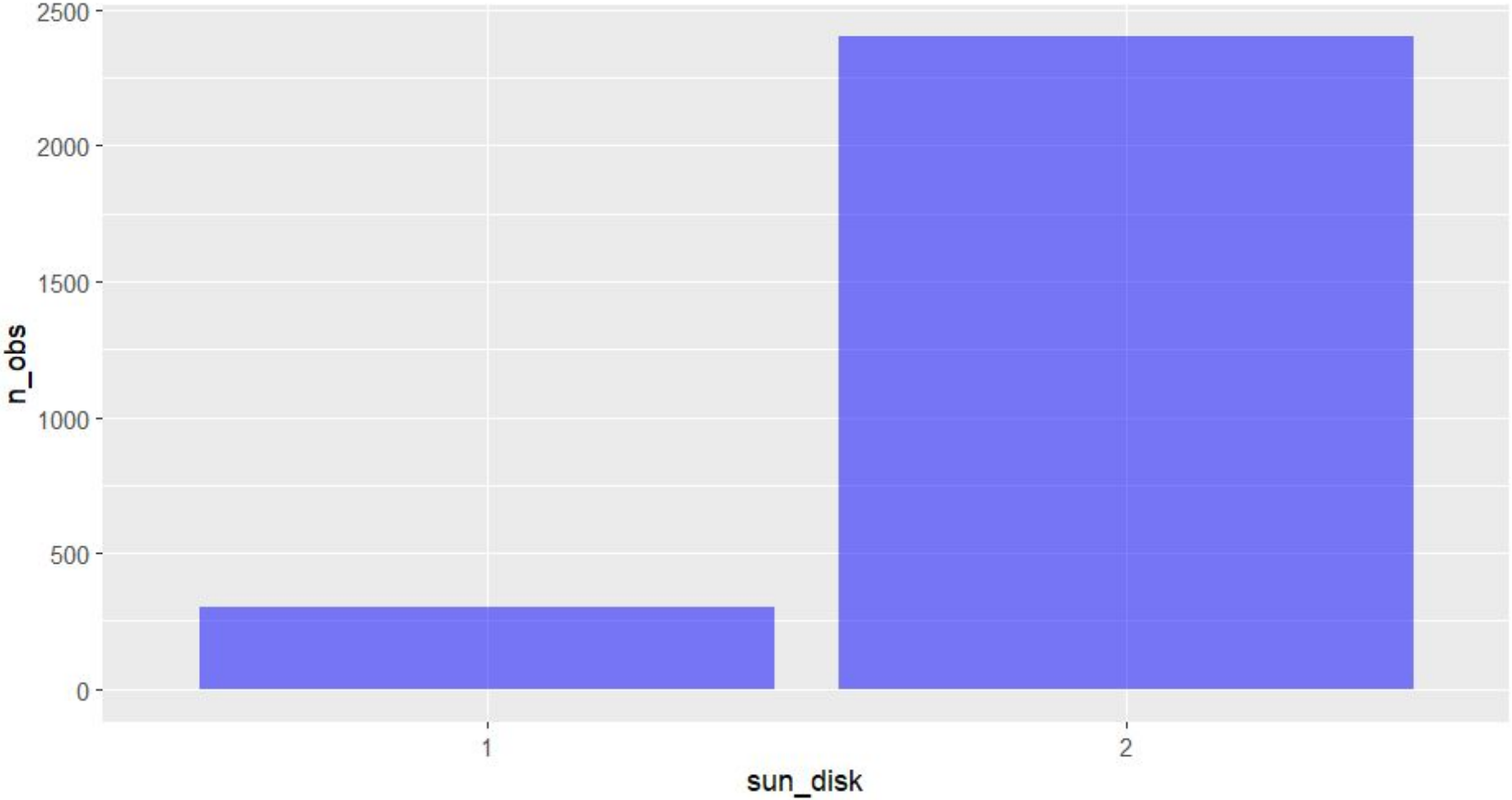

1200
800
400
0
n_obs
C
I
O
X
distribution

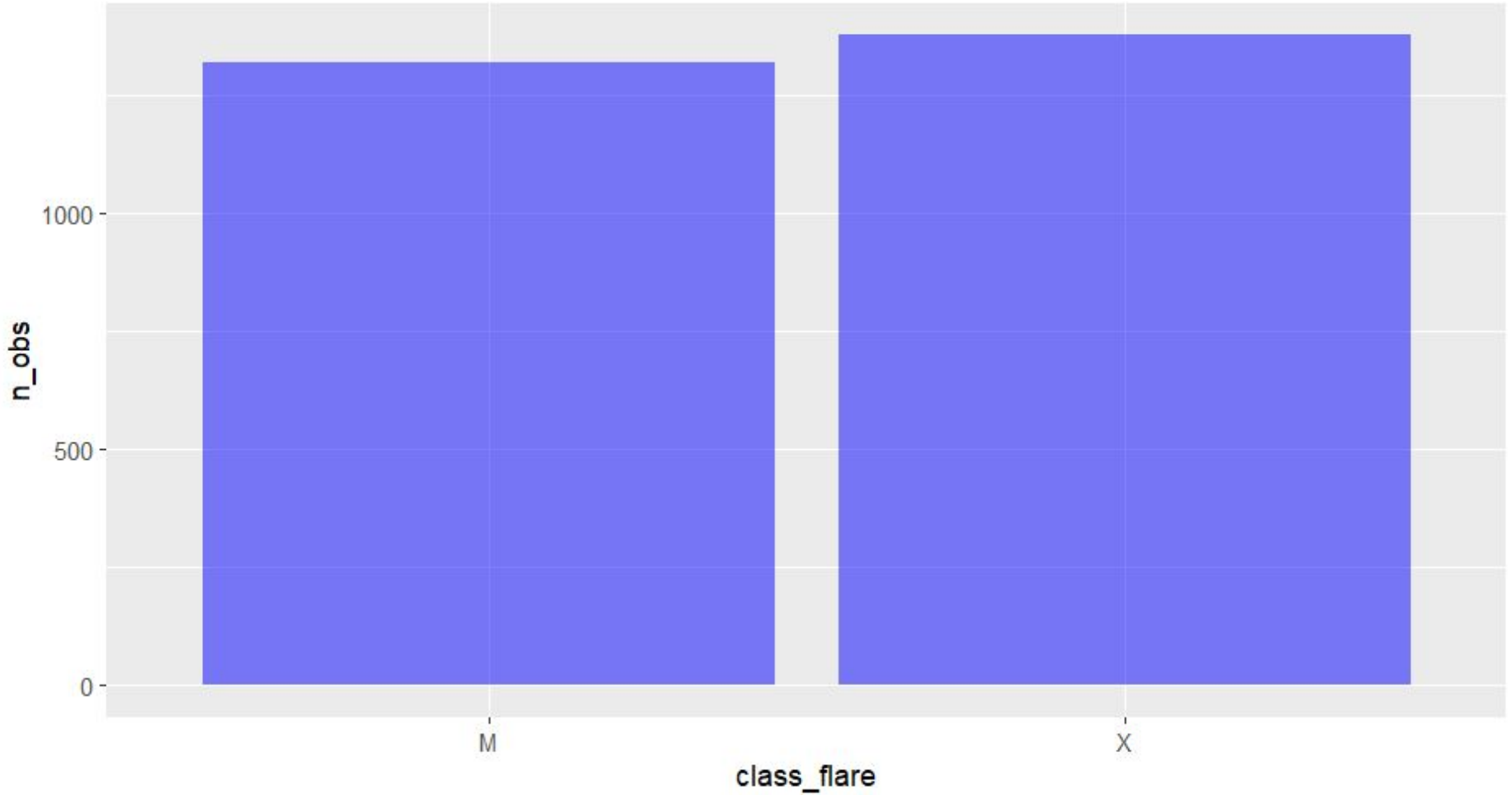

2000
1500
1000
500
0
n_obs
1
2
activity


1000
500
0
n_obs
1
2
3
evolution


2000
1000
0
n_obs
1
2
3
flare24_activity


1500
1000
500
0
n_obs
1
2
history


2500
2000
1500
1000
500
0
n_obs
1
2
sun_disk


n_obs
1000
500
0
M
X
class_flare