

# Cluster Analysis: Bullying in School

Laura Hu

## Data description

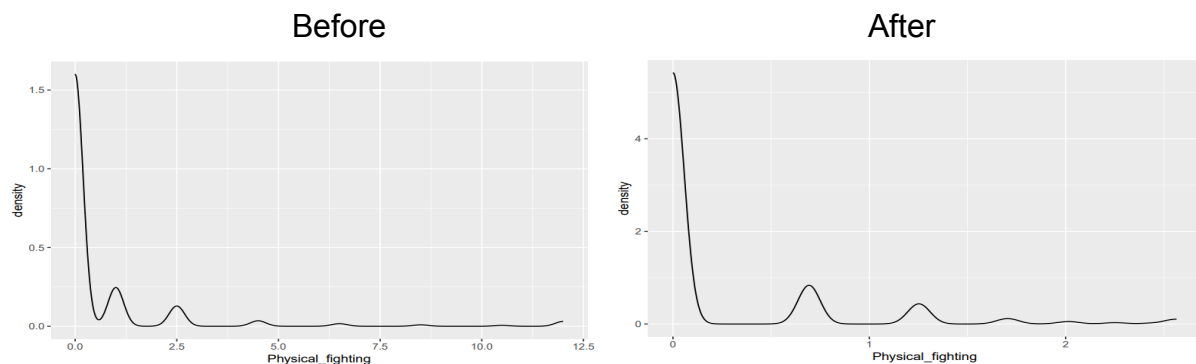
The Global School-Based Student Health Survey (GSHS) is a school-based survey that uses a self-administered questionnaire to obtain data on young people's health behavior and protective factors related to the leading causes of morbidity and mortality. The survey was conducted in Argentina in 2018. A total of 56,981 students participated.

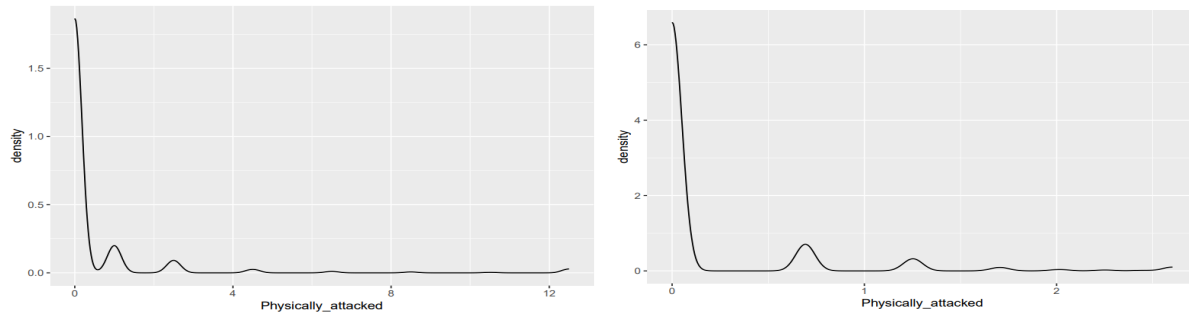
There are 15 variables in the data set: **Bullied in school, Bullied Outside of School, Cyber Bullied, Felt Lonely, Most of the Time or Always Felt Lonely, Missed School Without Permission, Missed Classes or School Without Permission, Close Friends, Close Friends 3 or More, Physical Fighting, Physical Fighting 12 or More, Physically Attacked, Physically Attacked 12 or More, Sex, and Custom Age.**

## Initial Observation

Attempting to fix the Skewness of Physical\_fighting/Physically\_attacked vs. density graphs

We noticed that “Physical\_attacked” and “Physical\_fighting” were heavily skewed to the right, so we attempted a log transformation:





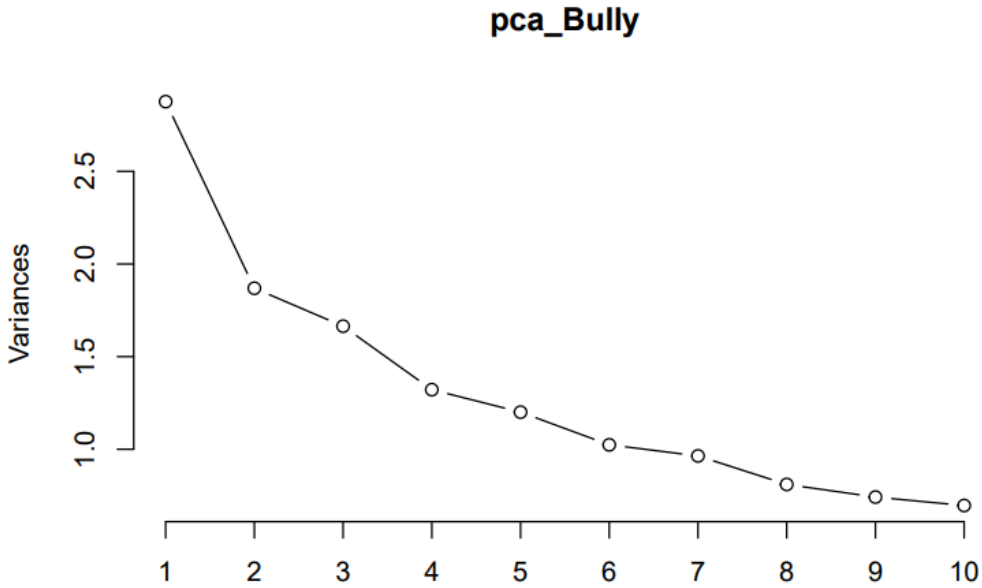
Although the data looks better overall, it is *still* heavily right-skewed.

## Performing Initial PCA

We ran PCA to see how we can visualize the data set, and the result is shown below in the image. From the scree plot, we can see that we need at least 10 variables to explain 90% of the data. Each PCA has its individual characteristics, for example: PCA1 tends to feel lonely and feels lonely most of the time. PC2 tends to miss school and class without permission, and PC3 has more close friends than the rest. PC4 involves more than average physical altercation and is mostly male. However, since the data set is so big and involves a lot of characteristics, we want to further investigate by performing a clustering analysis.

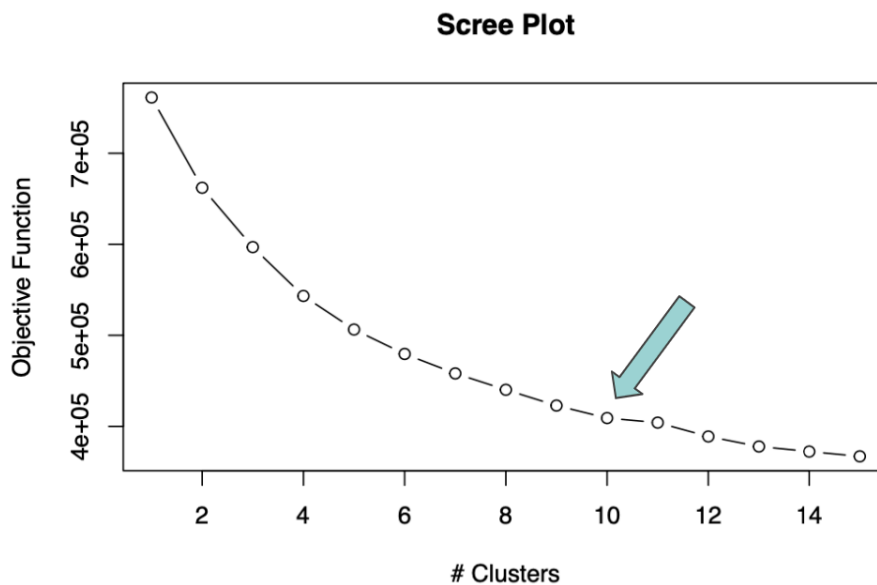
Rotation (n x k) = (15 x 15):

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Bullied.in.School	0.24997777	-0.029246806	0.28821512	-0.24044969	0.3267742	0.194766615	-0.01508237	0.28002144
Bullied.outside.school	0.26468066	-0.093340246	0.29743925	-0.23385187	0.3626926	0.002716212	0.18023144	0.08074576
Cyber.Bullied	0.26695550	-0.062079984	0.28014756	-0.11317110	0.3671536	-0.045104283	0.08394220	-0.23917387
Custom_Age	0.08576274	-0.094331165	-0.15082282	0.17749779	0.0773615	-0.470717437	0.76150449	-0.16925107
Sex	-0.14546009	-0.173443448	-0.14270407	-0.40748166	-0.2050733	0.065015741	0.34504384	0.68008797
Physically_attacked	0.19444699	-0.140492866	0.13756470	-0.39786734	-0.3198931	-0.222641194	-0.16592126	-0.29218380
Physical_fighting	0.10704942	-0.236417004	0.03169120	-0.44412041	-0.3851082	-0.244239899	-0.06983857	-0.18572134
Felt_lonely	0.43259149	0.102571515	0.17523221	0.31064271	-0.2135505	-0.175799686	-0.07126837	0.20816660
Close_friends	-0.29687618	-0.336278840	0.47599676	0.21103472	-0.1220414	0.059951778	0.07375270	-0.02459409
Miss_school_no_permission	0.18780665	-0.560024290	-0.29097724	0.14857972	0.1158258	0.045280353	-0.16551244	0.02885278
Other_students_kind_and_helpful	-0.23234155	-0.041113619	0.05398743	0.04618095	0.1849421	-0.634675264	0.22460462	0.25605982
Parents_understand_problems	-0.26189854	0.003254297	0.00813472	-0.11531992	0.3335802	-0.380460010	0.31895323	0.10902073
Most_of_the_time_or_always_felt_lonely	0.40344348	0.098676345	0.16722040	0.28133136	-0.2708173	-0.175372402	-0.11031115	0.35123683
Missed_classes_or_school_without_permission	0.18597815	-0.556374574	-0.29867686	0.15599825	0.1215022	0.068552923	-0.14511366	0.03525482
Close_friends_3_or_more	-0.29654135	-0.337259348	0.47417359	0.19893917	-0.1438524	0.073705233	0.07465445	0.00125397
	PC9	PC10	PC11	PC12	PC13	PC14	PC15	
Bullied.in.School	-0.306747596	-0.28417395	-0.399773086	0.486995876	0.0229077080	3.306381e-03	-0.0069812529	
Bullied.outside.school	0.123440513	-0.06784477	-0.097285367	-0.760335168	0.0092331931	5.153095e-04	0.0042883465	
Cyber.Bullied	0.348742815	0.39323413	0.462153260	0.377105406	0.0193727868	-2.522318e-03	0.0085696288	
Custom_Age	-0.258150921	-0.06564507	-0.095882574	0.102477332	0.0521447177	-1.018619e-03	0.0056845524	
Sex	0.061945370	0.11457373	0.322275916	0.053850434	-0.1015651423	-1.523387e-02	-0.0149146756	
Physically_attacked	-0.270351898	-0.48843720	0.434071946	-0.018487684	-0.0010721657	2.289927e-03	0.0001141428	
Physical_fighting	0.161809561	0.41043824	-0.537598567	0.052958505	0.0095391342	2.205251e-02	-0.0057302217	
Felt_lonely	-0.050457480	0.09340304	0.005417042	-0.012215705	-0.7327351402	-2.571669e-02	0.0347251800	
Close_friends	-0.049236984	0.01414281	0.004583787	0.001791527	-0.0403149075	6.268699e-04	-0.7063042214	
Miss_school_no_permission	-0.008782242	-0.01611432	0.007641678	-0.007250224	0.0290700814	-7.066626e-01	-0.0004327308	
Other_students_kind_and_helpful	0.487364692	-0.35916205	-0.082596487	0.112787615	0.0006716818	1.112578e-02	0.0077729983	
Parents_understand_problems	-0.590014204	0.41753044	0.109886278	-0.106842926	-0.0486299763	7.765997e-03	0.0052401594	
Most_of_the_time_or_always_felt_lonely	-0.073886662	0.14870383	0.081940065	-0.039324394	0.6650814701	3.007111e-02	-0.0410208362	
Missed_classes_or_school_without_permission	-0.009258911	-0.02408446	0.030248857	-0.007295122	-0.0278208392	7.057873e-01	0.0030485501	
Close_friends_3_or_more	-0.055333130	0.01834570	0.003225209	0.005292301	0.0323133417	-9.653407e-05	0.7054929224	



### Performing Cluster Analysis

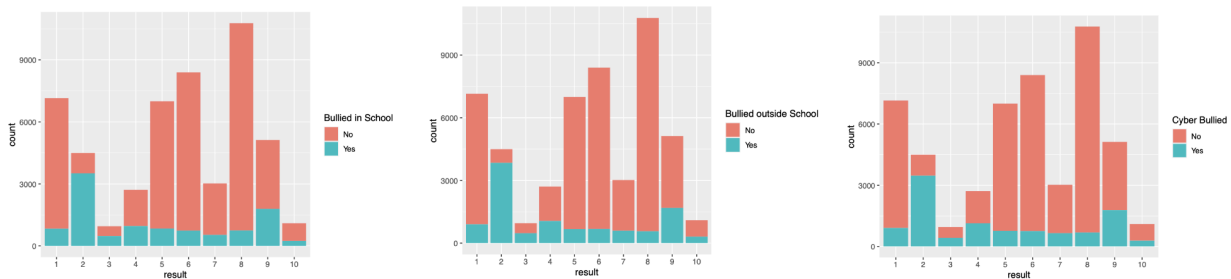
We picked the cluster range from 1-15 and utilized a scree plot to see which cluster could give us the best result. It turns out that the scree plot does not produce a very good result on which cluster can perform the best, and we decided the most appropriate cluster is cluster 10. This number of clusters also matches the result in PCA.



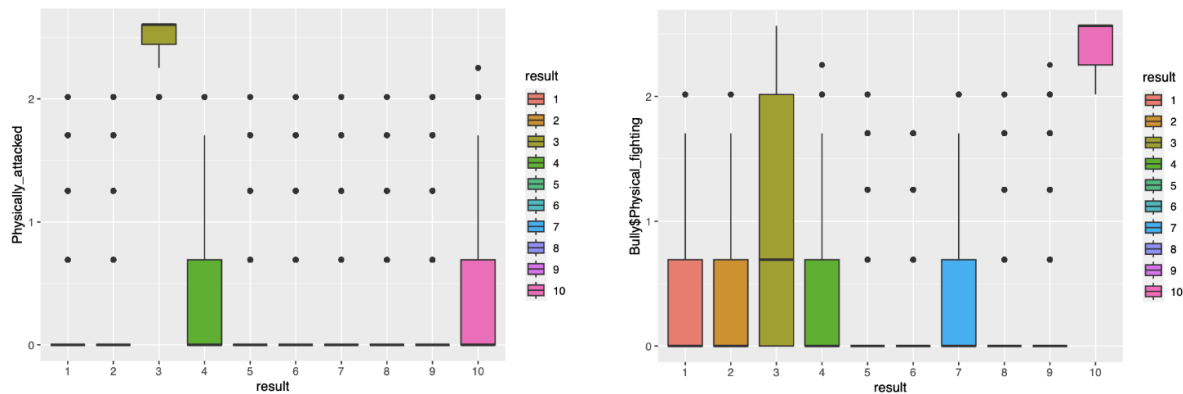
## Cluster Description by Variables

We first looked at each cluster in the context of each variable. Since variables such as “close friends” and “3 or more close friends” provide similar information to the clusters, we decided to analyze them together.

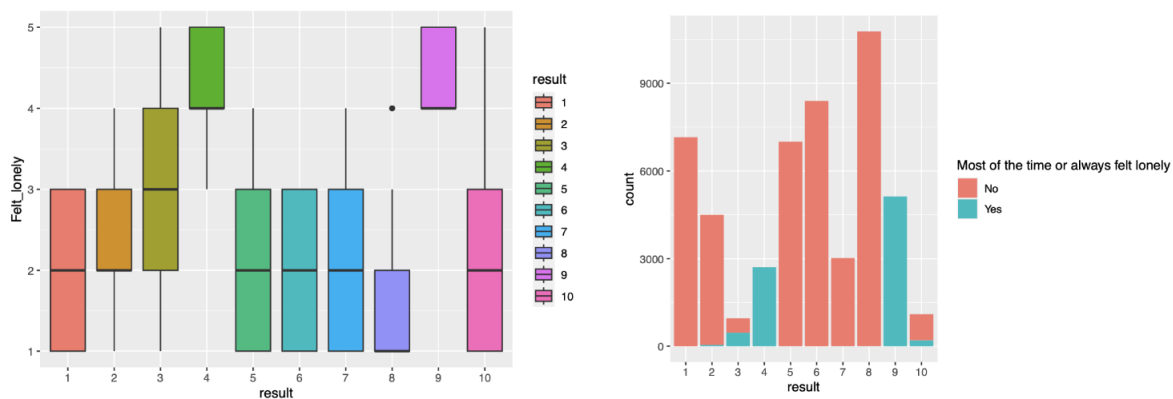
First, we have the overall response to their bullying experience. As we can see from the graph below, every cluster has a mixed response on their experience with three types of bullying. Clusters 2 and 3 have experienced the most bullying, and clusters 4 and 9 have experienced some bullying



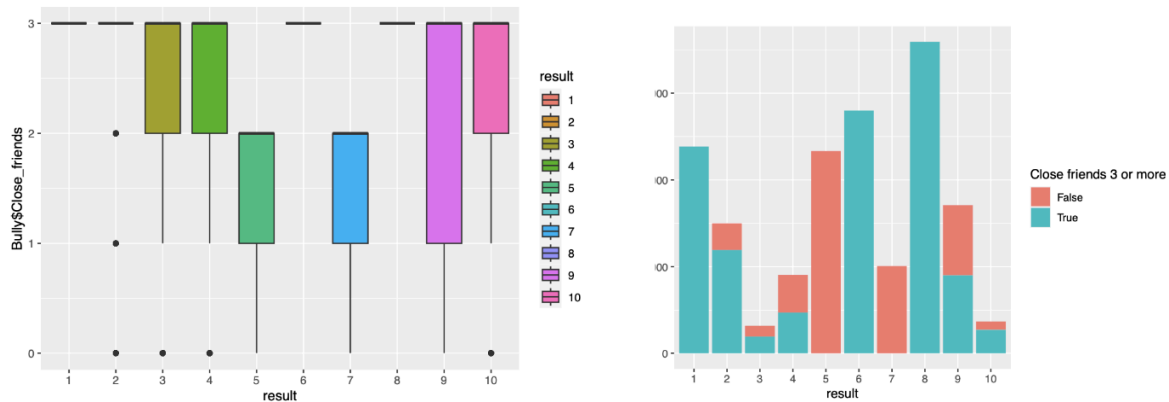
Then we have the frequency of physical altercations. We grouped variables “Physically attacked” and “Physical fighting” together because these are the two types of physical altercation in the survey. As we can see, cluster 3 reported a high frequency of being physically attacked, while clusters 4 and 10 had relatively higher frequency for smaller groups. Cluster 3 also has a range of students that were involved in fights. Cluster 10 reported the highest amount of physical fighting.



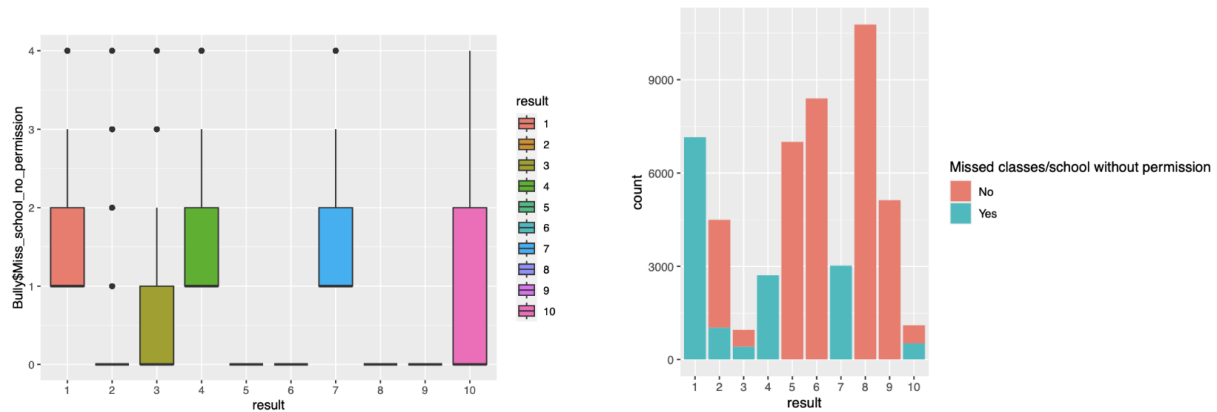
Thirdly, we investigated the variables “Felt lonely” and “most of the time or always felt lonely”. As we can see Clusters 4 and 9 reported a high frequency of feeling lonely, in fact, they feel lonely always or most of the time. Cluster 3 also reported a relatively higher frequency, and about half of them feel lonely always or most of the time.



Fourthly, we investigated “Close friends” and “3 or more close friends” because we should be able to see similar results in both variables. As we can see, Most clusters reported that they have at least 3 close friends or more. Clusters 5 and 7 have a lower number of close friends and it’s reflected in the bar graph. Half of cluster 9 has less than 3 close friends

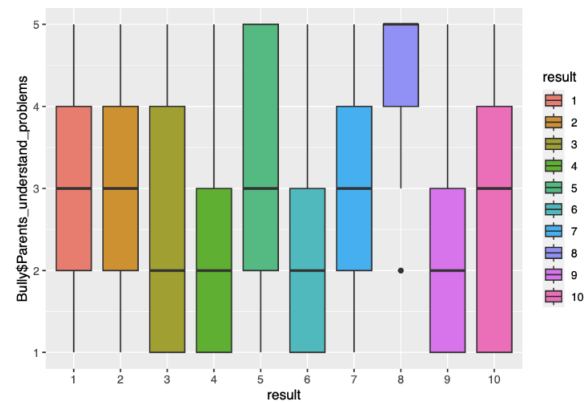
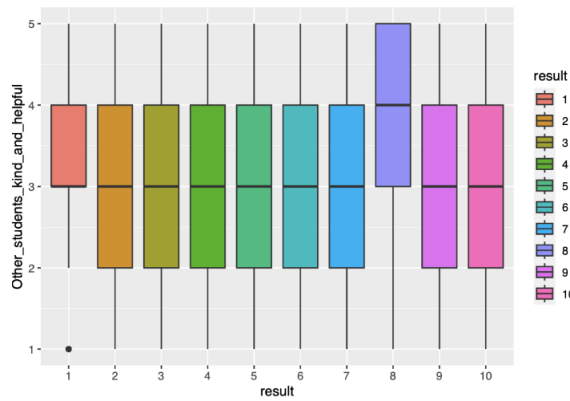


Then we looked at the variables “Miss school no permission” and “Miss school/class without permission”. Note that “Miss school no permission” tells us how often they skipped school, and “Miss school/class without permission” is an indicator variable that tells us if they have missed school/class or not. Clusters 1, 4, and 7 all have reported that they missed school without permission before. The same result can be reflected in the bar graph. Half of cluster 10 have reported that they missed school/class without permission before.

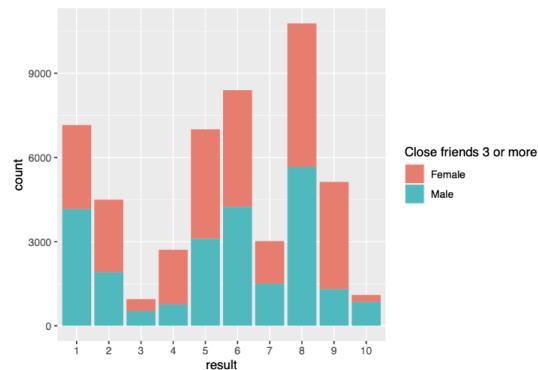
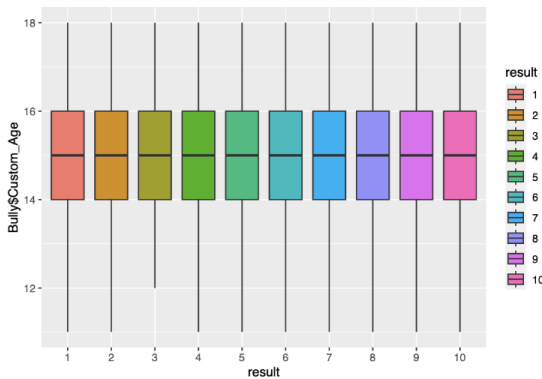


We also considered students’ attitudes towards others by looking at the variables “other students kind and helpful” and “parents understand problems”. We found that Most clusters have similar attitudes towards others being kind and helpful + their parents. Cluster 8 reported a more positive attitude towards other students than the rest of the group. Clusters 4, 6, and 9 have lower opinions on their parents' understanding of

the problem. Although the median of cluster 3 is also low, they have a wider range of answers.



Finally, we looked at students' gender and age. We found an even distribution of age in every cluster, the median of the group is around 15. The gender distribution in every cluster is also fairly even. Although there are more male students in cluster 10, and relatively more female students in cluster 4 and 9.



## Conclusion

Each cluster is labeled and interpreted in the table below.

Cluster Num	Cluster size	Characteristics	Name
1	7145	friendly students that have missed school/class without permission at least once	"Here for the homies"
2	4499	Most bullied but with no outstanding characteristics as to why they were bullied	"Most bullied"
3	960	Half of cluster 3 are bullied. They are physically attacked the most, involved in quite a few of physical fights, and felt lonely sometimes.	"Most attacked"
4	2714	More than average (half) of the cluster is female and has experienced bullying. Some have been physically attacked and involved in physical fights. Always felt lonely but has at least 3 close friends. Have at least missed school/class once. Don't think parents understand their problems.	"Edgy & Aggressive"
5	7000	Lower number of close friends but do not feel lonely most of the time.	"Sigma behavior"
6	8397	A lot of close friends, rarely feel lonely, most of them do not think	"Probably popular"



		parents understand the problem.	
7	3024	Low number of close friends, tend to miss class/school without permission	"School sucks"
8	10774	Have high opinion of other students and believe parents understand their problems. Have lots of friends and rarely feel lonely.	"Normies"
9	5127	More than half of them are female. A little less than half of the cluster experienced bullying. All of them feel lonely most of the time. Only half of them have more than 3 close friends. Most of them don't think parents understand their problems.	"Need more friends"
10	1104	Got into the most physical fights, More than average (half) of the cluster have missed school/class without permission.	"Bullies"