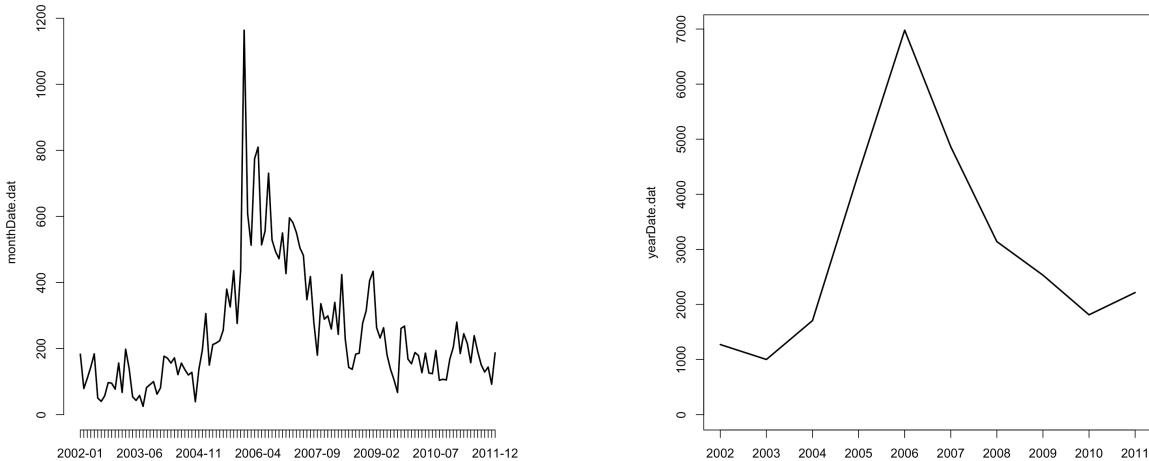


Web Forum Analysis

ThreadID	AuthorID	Date	Time	WC	Analytic	Cloud	Authentic
Min. : 4466	Min. : -1	Length:30000	Length:30000	Min. : 0.0	Min. : 0.00	Min. : 0.00	Min. : 0.00
1st Qu.:237554	1st Qu.: 38226	Class :character	Class :character	1st Qu.: 27.0	1st Qu.:39.45	1st Qu.:42.08	1st Qu.: 9.25
Median :332155	Median : 79334	Mode : character	Mode : character	Median : 60.0	Median :64.56	Median :60.76	Median :28.97
Mean :372186	Mean : 82005			Mean : 103.6	Mean :60.22	Mean :58.76	Mean :37.10
3rd Qu.:491287	3rd Qu.:116303			3rd Qu.:126.0	3rd Qu.:84.15	3rd Qu.:79.76	3rd Qu.:161.18
Max. :877240	Max. :252144			Max. :7418.0	Max. :99.00	Max. :99.00	Max. :99.00
Tone	ppron	i	we	you	shee	they	posemo
Min. : 0.00	Min. : 0.000	Min. : 0.000	Min. : 0.0000	Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.000
1st Qu.:14.06	1st Qu.: 4.170	1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000
Median :25.77	Median : 7.180	Median : 2.110	Median : 0.0000	Median : 0.000	Median : 0.000	Median : 0.000	Median : 2.440
Mean :44.42	Mean : 7.567	Mean : 3.144	Mean : 0.9516	Mean : 1.378	Mean : 0.662	Mean : 1.432	Mean : 3.662
3rd Qu.:80.64	3rd Qu.:10.420	3rd Qu.: 4.710	3rd Qu.: 1.2300	3rd Qu.: 1.830	3rd Qu.: 0.000	3rd Qu.: 2.140	3rd Qu.: 4.460
Max. :99.00	Max. :66.670	Max. :50.000	Max. :33.3300	Max. :50.000	Max. :33.3300	Max. :37.500	Max. :100.000
negemo	anx	anger	sad	focuspast	focusspresent	focussfuture	
Min. : 0.000	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000	Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.000
1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000
Median : 1.380	Median : 0.0000	Median : 0.0000	Median : 0.0000	Median : 2.630	Median : 10.42	Median : 0.000	
Mean : 2.155	Mean : 0.2461	Mean : 0.9234	Mean : 0.2902	Mean : 3.414	Mean : 10.58	Mean : 1.144	
3rd Qu.: 3.120	3rd Qu.: 0.0000	3rd Qu.: 1.2300	3rd Qu.: 0.0000	3rd Qu.: 5.000	3rd Qu.: 13.64	3rd Qu.: 1.650	
Max. :100.000	Max. :50.0000	Max. :100.0000	Max. :100.0000	Max. :100.000	Max. :100.00	Max. :100.00	Max. :40.000

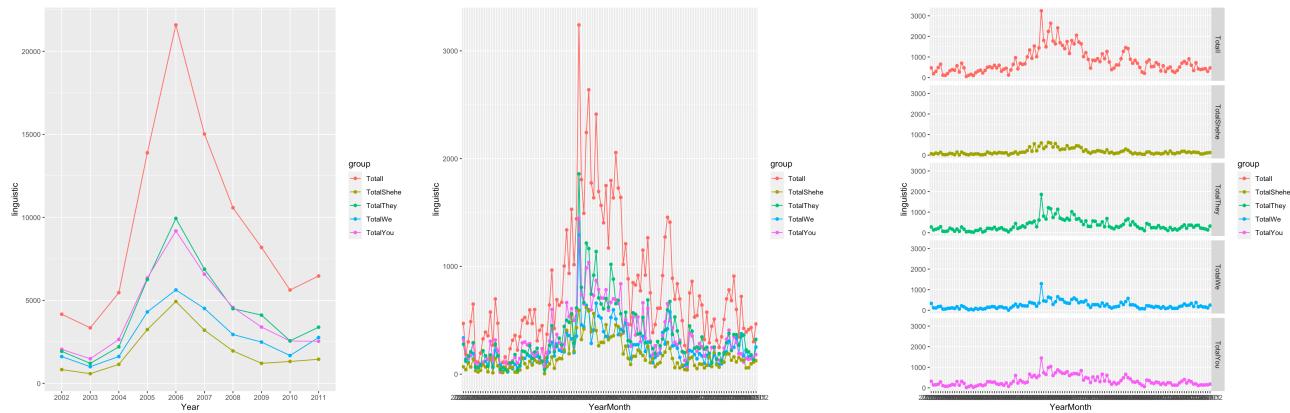
a)

1. In order to get a more accurate reading, all data with the word count being 0 was omitted from calculations as they were taken as redundant posts and it would affect the datas results in a way it may cause it to be unreliable. At first glance, we are able to identify that the web forum peaked at 2006 and it started to fade over the years. But as of 2011, there is a slight increase, which could signify a rise in its popularity once more.

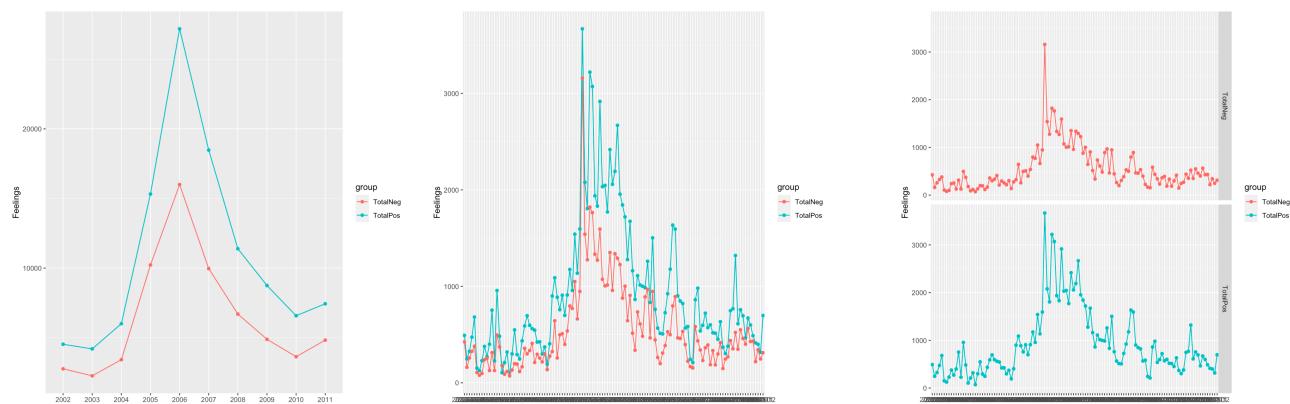


This is not an uncommon trend with webform pages as the internet is ever growing, there are more competitors out there. This could mean that the webform may just not be as interesting in comparison to its competitors as the Data can tell us that the Forums height was short lived as it was not able to maintain and keep its users from staying on the webform posting and writing on threads. Once it peaked in 2006, it started to go down from there. Only being able to see a proper increase in users once more from 2010 to 2011. When we go into more detail, we will be able to notice that through the months it can be seen that the activity on the forum can be quite unstable as it tends to have steep fluctuation points. One of the only period that the data was slightly stable was during late 2010. When it was nearly a flat line with a small bump of increase in user activity.

2. When looking at the graph it can be seen that Authors tend to talk about themselves in their posts more than talking about others. It can also be seen that whenever authors posts they don't really talk much about other individuals as the SheHe graph can be seen relatively flat over the months in comparison to the other graphs. It is also evident in the yearly graph where it shows that it has the least number of activities. Even at the I graphs lowest point, the SheHe graph barely edges over it. There is also a similar trend going on with the You and They graphs. This could plausibly mean that users were asking others about a group of individuals. In the summary sheet above, it can also be seen that the values of the You Graph and They Graph have near similar values, which proves the point more. As seen on the previous graphs regarding activity online by users, the trend can be seen on the Linguistics graphs as well. With the peak being at 2006 and gradually going down.



When looking at the yearly values and trend of both the positive and negative emotions, one may easily say that users tend to be happier and be in a positive mood. But if we are to look closer into the months, it says a different story. The graph overlaps a lot and it can even be seen that whenever the positive emotions go up the negative emotions are close behind. The shocking part of the graph was during the month between early 2007 to early 2008 where there was a significant difference between both emotions. This could mean that that period could have been a good period for the users of the webforum.



b)

1. Before analyzing the Data, I considered rows with both PosEmo and NegEmo with a 0 value to be invalid Data. This is due to the fact that it will cause the Data to have some discrepancies as no emotions were displayed in those posts. The Tables displayed below are the results of the Top 10 best ranking positive emotions and negative emotions ranking. Both emotions were tallied up by grouping them by the years and by their threadID's. Upon doing so I was able to find the emotional difference of the specific thread. Which then allowed me to rank each thread according to the emotional difference. A positive emotion may be high but it does not necessarily mean the negative emotion is not as high. Thus I needed to get the difference in order for me to rank the threads in a more efficient manner. The results can be understood as the higher the difference the better mood that thread would be in, while the lower it is, it would mean their moods are not the best. The formula was a simple Positive - Negative, thus there are some values which are negative.

As it can be seen, the top 10 happiest and saddest threads tend to vary as the years go by, topics change, trends change. As humans, we tend to like talking about different subjects and don't tend to stay too long on a specific topic. These results were not a shocker and was expected. The years were chosen in a way the latest data being 2011, the peak year data being 2006 and lastly choosing 2010 to determine if any of the thread would still make it to the top 10 threads the following year, none of them made it back to the top 10 in 2011 thus proving my point even more. (Left Column 2011 statistics, Middle Column 2010 and lastly the Column on the right is 2006)

ThreadID	Year	positive	negative	EmotionDiff	Rankings
1	773564	2011	167.37	83.38	83.99
2	772248	2011	122.26	84.32	37.94
3	792644	2011	100.06	66.10	33.96
4	853260	2011	87.41	60.73	26.68
5	812061	2011	58.65	41.21	17.44
6	846788	2011	48.05	31.25	16.80
7	822123	2011	22.69	7.06	15.63
8	813754	2011	62.36	47.16	15.20
9	852162	2011	21.79	6.75	15.04
10	805973	2011	26.60	13.29	13.31

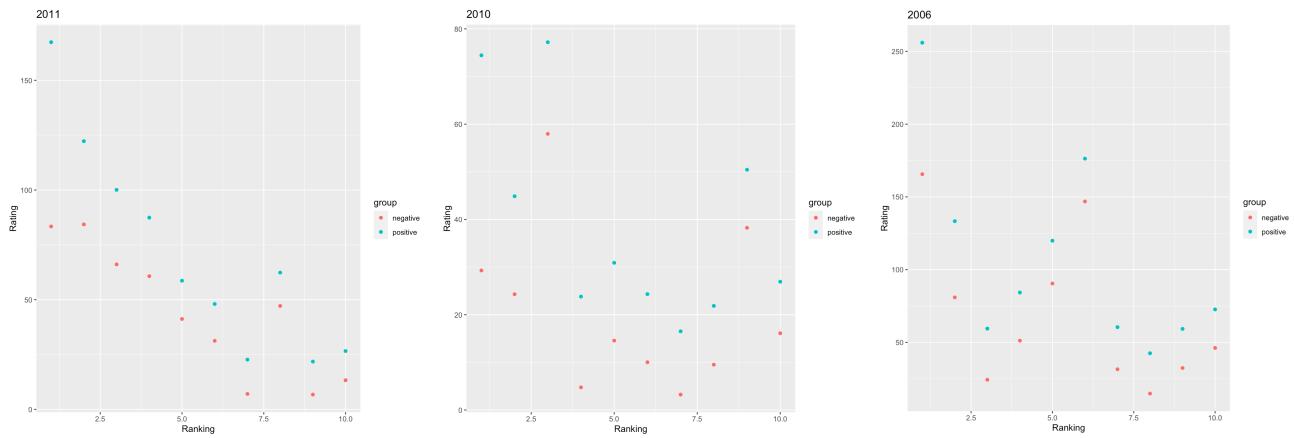
ThreadID	Year	positive	negative	EmotionDiff	Rankings
1	472752	2010	74.45	29.29	45.16
2	411294	2010	44.86	24.31	20.55
3	127115	2010	77.20	57.97	19.23
4	705067	2010	23.81	4.76	19.05
5	694830	2010	30.90	14.57	16.33
6	708828	2010	24.33	10.03	14.30
7	647112	2010	16.52	3.24	13.28
8	722527	2010	21.86	9.52	12.34
9	741695	2010	50.43	38.25	12.18
10	566639	2010	26.93	16.12	10.81

ThreadID	Year	positive	negative	EmotionDiff	Rankings
1	283958	2006	256.02	165.63	90.39
2	191868	2006	133.30	80.92	52.38
3	322645	2006	59.44	24.28	35.16
4	275515	2006	84.25	51.16	33.09
5	300350	2006	119.87	90.44	29.43
6	249001	2006	176.29	146.86	29.43
7	343416	2006	60.43	31.50	28.93
8	321213	2006	42.51	14.79	27.72
9	291742	2006	59.25	32.36	26.89
10	314216	2006	72.64	46.18	26.46

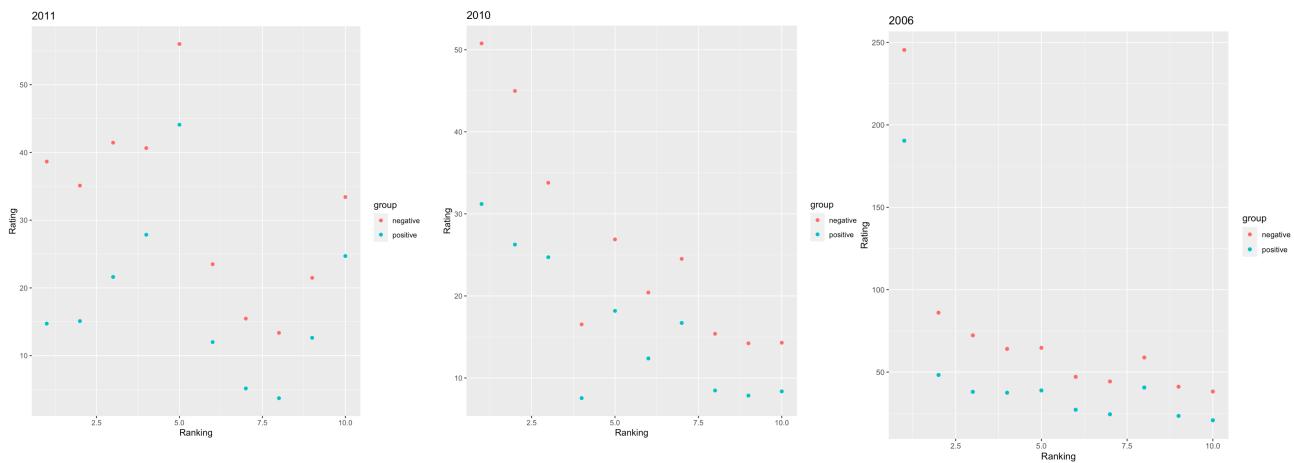
ThreadID	Year	positive	negative	EmotionDiff	Rankings
1	804569	2011	14.73	38.66	-23.93
2	847225	2011	15.10	35.11	-20.01
3	824091	2011	21.62	41.45	-19.83
4	833249	2011	27.85	40.65	-12.80
5	838674	2011	44.10	56.02	-11.92
6	784596	2011	12.00	23.50	-11.50
7	785890	2011	5.16	15.47	-10.31
8	795256	2011	3.72	13.36	-9.64
9	800540	2011	12.63	21.49	-8.86
10	805697	2011	24.71	33.42	-8.71

ThreadID	Year	positive	negative	EmotionDiff	Rankings
1	737764	2010	31.20	50.77	-19.57
2	675493	2010	26.26	44.98	-18.72
3	688803	2010	24.72	33.79	-9.07
4	675970	2010	7.55	16.53	-8.98
5	696538	2010	18.18	26.89	-8.71
6	655543	2010	12.38	20.42	-8.04
7	727690	2010	16.71	24.52	-7.81
8	674404	2010	8.48	15.39	-6.91
9	681711	2010	7.86	14.23	-6.37
10	679043	2010	8.37	14.30	-5.93

ThreadID	Year	positive	negative	EmotionDiff	Rankings
1	309286	2006	190.39	245.46	-55.07
2	325032	2006	48.28	86.04	-37.76
3	295652	2006	38.04	72.33	-34.29
4	342682	2006	37.52	64.09	-26.57
5	218736	2006	38.88	64.75	-25.87
6	338037	2006	27.17	47.14	-19.97
7	297192	2006	24.43	44.33	-19.90
8	328850	2006	40.64	58.87	-18.23
9	268154	2006	23.42	41.14	-17.72
10	258532	2006	20.78	38.24	-17.46



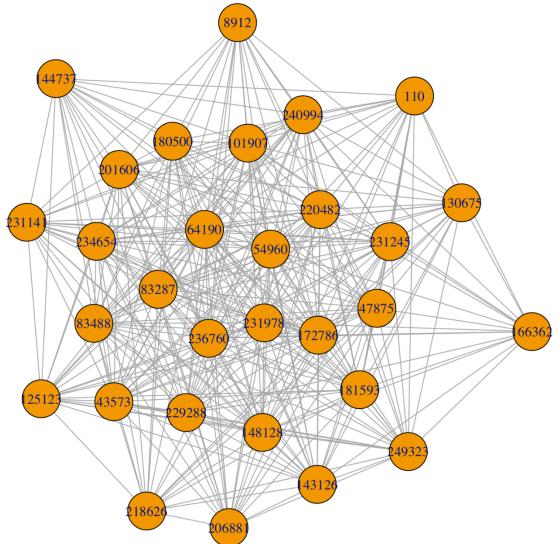
You may refer to the scatter plot diagrams in this page. X axis being the rankings of the Threads with both of its emotional rating (Y axis) on the same X axis. The graphs further proves my point that the higher a specific emotional rating is, does not necessarily mean that it is leaning more towards that specific emotion.



c)

1. First off in order to retrieve 30 users, I dropped all the Authors with the AuthorID -1, assuming that they may be guest users of the web forum. Upon doing so, I grouped up the Authors together along side the threads they've posted on and on a specific year, being 2011. Upon doing so, I calculated the number of threads they were involved in and retrieved the top 30 from there. This is with the assumption that those who were involved in more threads would be able to connect more with others thus building their network. Upon doing so I've concluded with the network shown below.

AuthorID	Year	NoThreads
1	54960	2011
2	83488	2011
3	236760	2011
4	220482	2011
5	83287	2011
6	234654	2011
7	240994	2011
8	201606	2011
9	166362	2011
10	43573	2011
11	148128	2011
12	64190	2011
13	181593	2011
14	172786	2011
15	8912	2011
16	231245	2011
17	206881	2011
18	180500	2011
19	229288	2011
20	125123	2011
21	143126	2011
22	231141	2011
23	231978	2011
24	47875	2011
25	101907	2011
26	130675	2011
27	218626	2011
28	249323	2011
29	110	2011
30	144737	2011



2. Upon looking at the list of Authors, an assumption was made that the authors who posted in more threads would likely have the strongest connections in the network as they would be placing themselves in more places, allowing them to be able to make more connections.

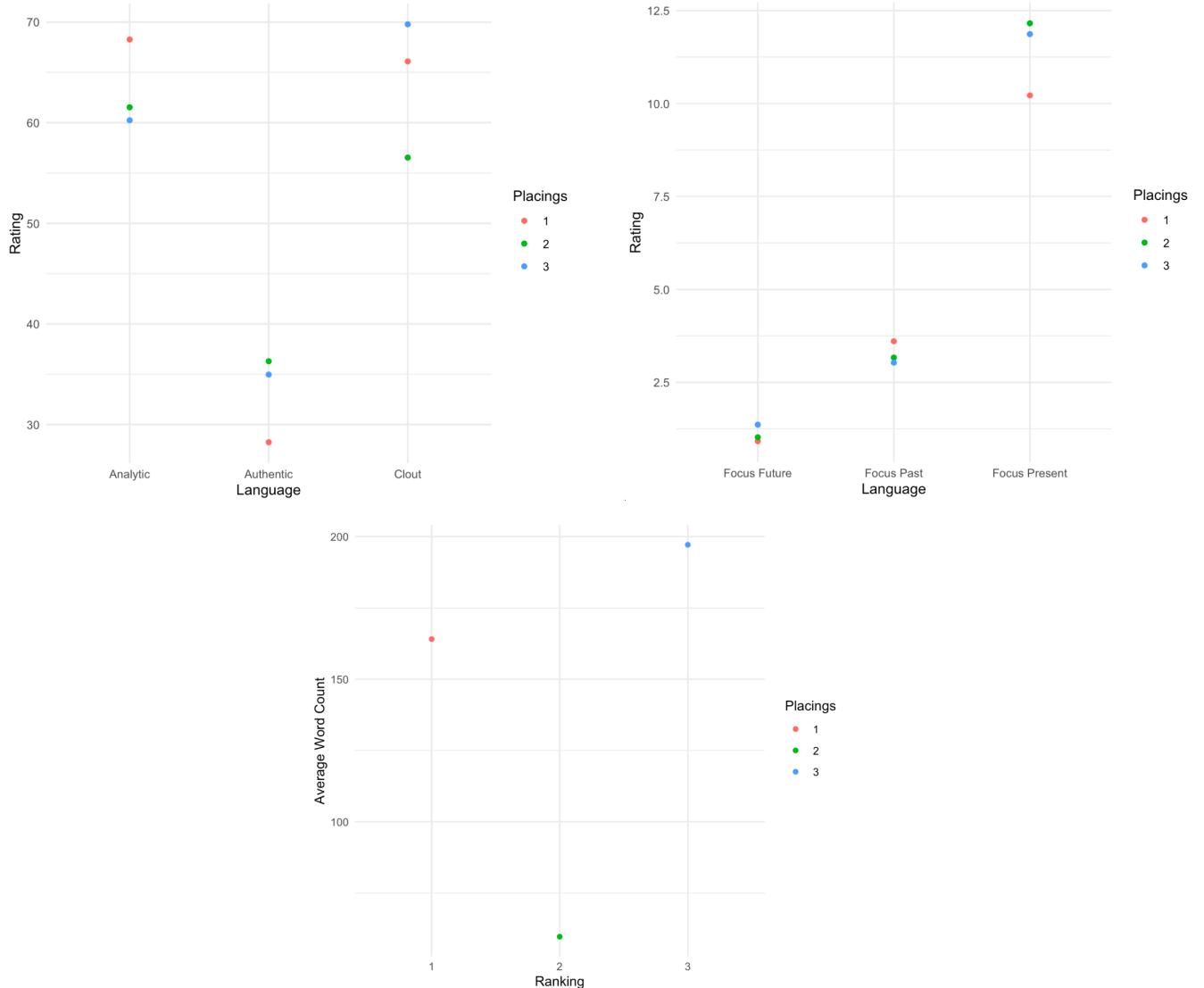
	AuthorID	Year	NoThreads
1	54960	2011	134
2	83488	2011	126
3	236760	2011	94
4	220482	2011	45
5	83287	2011	40

In order to verify whether the claim was true, I measured the degree of each author and the results show that my assumption was partially wrong. The Author with the strongest

54960	83488	236760	220482	83287	234654	240994	201606	166362	43573	148128	64190	181593	172786	8912	231245	206881
29	25	25	28	27	24	23	24	16	25	25	28	26	28	17	24	20
180500	229288	125123	143126	231141	231978	47875	101907	130675	218626	249323	110	144737				
21	23	20	23	20	25	24	26	19	18	22	18	17				

connection may have been the author who posted in the most threads but the next 2 did not have a stronger connection as comparison to Author “220482” who only posted in 45 threads. This could possibly be due to the fact that the authors may be the only ones posting in specific threads, thus not making as much connections to the other Authors.

3 users were taken into account. First being the Author with the strongest connection, next an author who has connections in the mid level and lastly the Author with the least connections in the network. The mean value of the linguistics variables were taken into account to be able to analyze the data in a fair manner.



As it can be seen in the graph based on the average word count in each post, the authors do not necessarily need to write more in each thread to have stronger connections in the network. The top author tends to post about 164 words per post while the middle author tend to only post 59 words per post.

The top author tends to be more analytical and more past focused in terms of how they post on the webforms as compared to the other authors and not so much on present focus and authenticity.

In terms of clout, the author with the least connection uses more clout as maybe they are trying to make their points known to others. The top author is just behind close. This could signify that they are trying to use their influence as a platform for them to influence others.

Appendix

Assumptions:

- AuthorID with Value -1 are guest users
- Remove data with 0 word count as it is taken as an invalid post which brings down data accuracy
- Assumptions that pronouns in the 2000's weren't as acknowledged as the late 2010's
- Remove columns with both posemo and negemo having 0 as they were considered as invalid datas. This is due to the fact that no emotion was shown and it would cause some discrepancies with the data results.
- Web Forum is a news forum