

1. Word Cloud Text Analysis:

1.1. About the Text Variables:

Text variable analysis can make sense when we analyze them as a whole. In this report about variable "V2", we analyze it regarding word frequency and weighted word frequency. These analysis give the reader a better notion about the keywords provided on all instances in the data. Sometimes, these variables arise from survey respondents answers, it is clear that such analysis can be complemented with the grouping of answers. Thus, to provide a better look at this variable behavior and influence in other possible variables in the dataset, grouping is advisable. If you have doubts or need further analysis, please contact us and to ask for help, please feel free to use our ticket system.

1.2. Results:

Variable	word.freq	freq	word.tfidf	weight.tf.idf
V2	research	7038.000	research	543.455
V2	collaborative	3228.000	collaborative	360.292
V2	the	2114.000	science	302.241
V2	science	1613.000	engineering	279.371
V2	systems	1486.000	the	270.227
V2	development	1315.000	career	253.502
V2	career	1314.000	systems	245.685
V2	phase	1274.000	development	212.384
V2	analysis	1158.000	phase	210.974
V2	engineering	1105.000	mathematics	192.799
V2	dissertation	1098.000	workshop	190.984
V2	education	914.000	analysis	185.121
V2	studies	912.000	program	180.950
V2	sbir	906.000	dynamics	177.227
V2	dynamics	902.000	theory	176.988
V2	workshop	881.000	studies	176.593
V2	acquisition	826.000	sger	170.719
V2	technology	821.000	technology	170.159

Variable	word.freq	freq	word.tfidf	weight.tf.idf
V2	new	820.000	sbir	169.180
V2	using	815.000	education	168.816
V2	study	814.000	computer	158.356
V2	program	812.000	conference	155.013
V2	system	808.000	fellowship	154.647
V2	high	774.000	applications	143.530
V2	materials	771.000	postdoctoral	140.663
V2	conference	771.000	new	140.597
V2	sger	725.000	high	136.085
V2	theory	724.000	materials	135.520
V2	data	723.000	modeling	134.203
V2	applications	721.000	design	131.247
V2	modeling	680.000	project	130.641
V2	mathematics	668.000	cooperative	130.401
V2	cooperative	655.000	physics	127.085
V2	molecular	618.000	dissertation	127.082
V2	structure	615.000	study	126.988
V2	control	608.000	system	125.126
V2	international	606.000	molecular	125.105
V2	design	602.000	advance	122.679
V2	evolution	597.000	data	122.450
V2	itr	592.000	chemistry	121.810
V2	doctoral	581.000	learning	119.506
V2	information	575.000	control	118.892
V2	support	566.000	support	118.196
V2	university	517.000	international	117.821
V2	chemistry	510.000	award	117.654
V2	computer	507.000	using	115.164
V2	learning	484.000	information	114.878

Variable	word.freq	freq	word.tfidf	weight.tf.idf
V2	models	482.000	structure	113.973
V2	effects	466.000	center	113.471
V2	rei	465.000	undergraduate	111.149
V2	integrated	461.000	digital	110.612
V2	project	459.000	quantum	108.984
V2	processes	450.000	evolution	108.883
V2	physics	450.000	problems	108.340
V2	approach	431.000	reu	105.831
V2	methods	430.000	models	105.616
V2	environmental	427.000	university	104.972
V2	center	417.000	processes	104.684
V2	undergraduate	414.000	methods	104.274
V2	model	410.000	scholarships	103.882
V2	role	408.000	rei	100.929
V2	laboratory	399.000	acquisition	99.785
V2	problems	397.000	laboratory	99.668
V2	novel	392.000	optical	98.711
V2	biology	392.000	site	97.079
V2	site	380.000	biology	95.419
V2	computational	375.000	curriculum	94.632
V2	optical	370.000	doctoral	93.984
V2	interactions	367.000	planning	91.442
V2	reu	364.000	effects	90.525
V2	quantum	358.000	geometry	89.223
V2	networks	353.000	novel	89.062
V2	experimental	352.000	grant	88.717
V2	based	346.000	integrated	88.267
V2	digital	345.000	nonlinear	86.967
V2	chemical	344.000	national	86.263
V2	field	334.000	structures	85.913

Variable	word.freq	freq	word.tfidf	weight.tf.idf
V2	planning	333.000	biological	84.310
V2	biological	332.000	computational	81.995
V2	grant	329.000	model	80.573
V2	synthesis	326.000	networks	79.659
V2	performance	326.000	interactions	79.510
V2	computing	322.000	experimental	79.307
V2	investigation	321.000	approach	79.217
V2	advanced	321.000	computing	78.351
V2	characterization	320.000	performance	78.195
V2	properties	319.000	leadership	78.020
V2	climate	315.000	microbial	77.984
V2	processing	314.000	ner	77.919
V2	carbon	311.000	environmental	77.811
V2	complex	310.000	role	76.547
V2	change	308.000	theoretical	73.710
V2	nonlinear	304.000	nsf	73.675
V2	transport	304.000	network	73.331
V2	ocean	300.000	mechanisms	73.241
V2	geometry	296.000	field	72.878
V2	simulation	292.000	ocean	72.518
V2	environment	291.000	mathematical	72.515
V2	fellowship	291.000	climate	72.398
V2	sciences	291.000	processing	72.310

1.3. Word Cloud (with Word Freq.) is:

Suppose we have a set of English text documents and wish to rank which document is most relevant to the query, "the white SUV". A simple way to start out is by eliminating documents that do not contain all three words "the", "white", and "SUV", but this still leaves many documents. To further distinguish them, we might count the number of times each term occurs in each document; the number of times a term occurs in a document is called its term frequency. However, in the case where the length of documents varies greatly, adjustments are often made.

