

Latent Topic Modeling on Twitter Data

Method: CA

Correspondence Analysis (CA) is a multivariate graphical technique designed to explore relationships among categorical variables. The outcome from correspondence analysis is a graphical display of the rows and columns of a contingency table that is designed to permit visualization of the salient relationships among the variable responses in a low-dimensional space. Such a representation reveals a more global picture of the relationships among row-column pairs which would otherwise not be detected through a pairwise analysis.

Calculate CA:

- Step 1: Compute row and column averages
- Step 2: Compute the expected values
- Step 3: Compute the residuals
- Step 4: Plotting labels with similar residuals close together
- Step 5: Interpreting the relationship between row and column labels

How to Interpret Correspondence Analysis Plots

Correspondence analysis does not show us which rows have the highest numbers, nor which columns have the highest numbers. It instead shows us the relativities.

- The further things are from the origin, the more discriminating they are.
- Look at the length of the line connecting the row label to the origin. Longer lines indicate that the row label is highly associated with some of the column labels (i.e., it has at least one high residual).
- Look at the length of the label connecting the column label to the origin. Longer lines again indicate a high association between the column label and one or more row labels.
- Look at the angle formed between these two lines. Really small angles indicate association. 90 degree angles indicate no relationship. Angles near 180 degrees indicate negative associations.

Dataset

##	9	10	11	12	13	14	15	16
## #actu	33	1887	1719	0	2248	2178	0	2046
## #ada	5442	8017	25570	8335	7913	4790	4459	5966
## #airdrop	11441	36136	65027	47022	44009	28802	31422	20915
## #altcoin	8576	27302	34043	33968	35554	23094	29489	21702
## #amazon	0	985	706	0	0	0	0	0
## #bch	2256	5532	0	0	5837	6713	0	4335
## #binanc	2967	10488	9240	10857	8285	8518	11244	7510
## #bitcoin	48217	203875	217191	222528	181728	169374	217601	146159
## #bitcoincash	805	3329	3176	3487	2468	2912	3852	7995
## #bittrex	0	0	0	0	0	0	0	1327
## #blockchain	25360	104908	121667	126158	119212	101804	120474	82617
## #bounti	5951	18956	32387	27493	29709	15770	22262	14486
## #btc	28617	95589	119714	95818	91273	76517	91924	64283
## #bts	0	0	0	9725	0	0	0	4577
## #busi	29	1283	1041	0	1989	0	0	1566
## #cambridgeanalyt	0	0	0	5004	1658	0	0	0
## #cambridgeanalytica	0	0	780	13947	4700	4102	6515	1851
## #coin	544	2564	0	0	12546	4176	0	0

## #coinbas	674	0	0	0	0	1647	2644	0
## #congress	0	0	0	0	0	0	2309	0
## #costarica	0	0	0	0	0	0	0	0
## #crowdfund	1649	0	0	0	0	0	9194	4170
## #crowdsal	0	4511	0	7858	4876	5252	9604	4020
## #crypto	23912	84945	101526	96566	88437	81127	102777	75307
## #cryptocurr	25722	97272	114420	117667	107539	95315	116701	86840
## #cryptonew	0	2919	2865	8174	2009	6030	8484	0
## #cybersecur	2396	8491	9800	12639	9689	11301	15782	8265
## #dash	2861	8558	9124	8832	6340	8412	4518	5950
## #data	0	0	0	3958	2986	2558	3721	1867
## #date	0	0	0	0	0	0	0	0
## #deletefacebook	0	0	0	5732	3050	1797	2762	0
## #digitalmarket	0	895	710	0	0	0	0	1356
## #digitizecoin	0	0	0	0	0	0	4733	0
## #earn	0	0	0	0	8515	0	0	0
## #elsalvador	0	0	0	0	0	0	0	0
## #energytoken	5921	8852	0	0	0	0	0	0
## #eo	4481	9439	0	4646	6158	4802	0	4102
## #erc20	1385	0	5729	6245	7331	4942	4706	4441
## #escort	0	0	0	0	0	0	0	5084
## #etc	1643	6167	0	0	0	0	0	0
## #eth	28904	90509	118505	96421	101394	82375	90647	65281
## #ether	4201	17129	19586	20965	23653	24212	24726	14193
## #ethereum	37313	145661	163302	160486	153445	141294	157889	123366
## #f8	0	0	0	0	0	0	0	0
## #facebook	949	50748	41961	130670	86932	76678	128772	63718
## #facebookdatabreach	0	0	0	0	1635	0	2683	0
## #facebookdataleak	0	0	0	0	0	0	2801	0
## #facebookg	0	0	0	3196	0	0	0	0
## #faitsdiv	33	1882	1717	0	2241	2173	0	2037
## #fintech	1109	4166	3392	3953	2660	2524	3768	1911
## #follow	18	0	0	0	0	0	0	0
## #free	2102	8614	2947	0	11960	0	0	0
## #freecoin	0	0	0	0	0	0	9535	0
## #freetoken	1882	6550	7427	8803	18341	10097	11549	6243
## #game	0	0	0	0	4687	0	0	0
## #gdpr	0	0	0	0	0	0	0	1499
## #geoloc	0	0	0	4655	8009	9285	10856	4415
## #girl	0	0	0	0	0	0	0	5119
## #googl	42	2779	2532	4691	5555	3372	4228	3154
## #grab	0	0	9192	0	0	0	0	0
## #gto	3015	0	0	0	0	0	0	0
## #guatemala	0	0	0	0	0	0	0	0
## #hiphop	21	1249	950	0	0	0	0	0
## #ico	22618	85294	106509	107006	96788	85156	99988	71690
## #info	35	1918	1747	2435	2295	2217	0	2101
## #instagram	153	7778	5601	8764	7736	7760	8403	7887
## #invest	2549	10826	4243	10145	8789	10133	11408	6770
## #justic	35	1962	1773	2452	2316	2263	2288	2205
## #linkedin	23	1091	948	0	0	0	0	1384
## #litecoin	6944	24524	28035	22353	22159	18549	19239	13606
## #live	1444	0	0	0	8888	0	0	0
## #ltc	10626	24905	26691	24203	20939	17617	18332	18170

## #maga	0	0	0	0	0	0	2999	0
## #mana	0	0	14506	0	0	0	0	0
## #market	36	2763	1954	3032	4668	2457	2865	2591
## #markzuckerberg	0	0	0	3952	0	1805	7903	0
## #média	33	1917	1760	0	2271	2192	0	2047
## #mexico	0	0	0	0	0	0	0	0
## #money	759	3616	3129	0	11018	1775	0	0
## #music	46	2458	2201	3045	2798	2846	3027	2820
## #neo	6536	17747	12554	11168	10211	8381	7989	5928
## #new	734	11244	6719	5867	6333	10280	5774	9363
## #nowplay	0	0	0	0	0	0	0	0
## #np	0	0	0	0	0	0	0	0
## #opirium	0	0	0	0	0	4354	4372	0
## #panama	0	0	0	0	0	0	0	0
## #pdata	0	0	0	0	0	4269	4318	0
## #pinterest	0	924	745	0	0	0	0	0
## #privaci	0	0	0	5047	4843	3767	6267	3097
## #qanon	0	0	0	0	0	0	2491	0
## #radiocapitol	33	1882	1717	0	2241	2173	0	2037
## #rippl	10406	30151	21378	20985	19602	17416	19468	13612
## #rt	0	0	0	0	0	0	0	0
## #smm	0	1249	922	0	0	1509	0	1194
## #snapchat	20	1406	984	0	1697	1602	0	1631
## #socialmedia	84	5468	4141	8751	6959	6267	8052	6012
## #socialmediamarket	20	0	0	0	0	0	0	0
## #soundcloud	31	1424	1162	0	0	0	0	1183
## #submityourmus	21	1158	885	0	0	0	0	0
## #tge	0	0	0	0	0	0	5112	4757
## #tinder	0	0	0	0	0	0	0	0
## #token	6253	25055	27857	26829	41958	27994	27825	14562
## #tokensal	5635	21400	22542	21919	21075	22293	27843	27337
## #trade	3952	12218	12189	10300	9258	9860	10287	8306
## #travel	0	0	0	0	0	0	0	0
## #tron	3465	7741	0	4996	0	4132	6167	0
## #trump	0	0	0	3805	0	0	0	0
## #trx	8392	19648	34732	21853	17709	13922	17372	13047
## #twitter	160	8199	6384	11196	9690	8526	9834	8848
## #uniqu	33	1887	1720	2527	2248	2251	0	2045
## #usa	21	0	0	0	0	0	0	0
## #usd	0	0	0	0	2241	0	0	0
## #verg	2493	6675	0	0	0	0	0	0
## #video	0	0	0	0	0	0	0	0
## #vip	0	0	0	0	0	0	0	0
## #webradio	33	1882	1718	0	2241	2173	0	2039
## #whatsapp	0	1146	836	0	0	0	0	0
## #xem	0	0	0	0	0	0	0	0
## #xlm	5289	11806	5957	5749	5579	4759	0	3891
## #xmr	1428	0	0	0	0	0	0	0
## #xrb	3144	6450	0	0	0	0	0	0
## #xrp	11169	33533	29413	25704	23416	20931	21598	17065
## #xvg	6037	10978	33533	15046	9724	6645	8805	9520
## #youtub	84	3921	2951	4276	3684	3824	4257	4265
## #zuckerberg	0	0	0	4287	1763	2614	18855	1838
## 000facebook	0	0	0	4391	2785	2602	3727	1675

## 000guarium	0	0	0	0	0	0	0	1311
## account	0	0	0	3637	1793	1493	2533	0
## ad	41	1610	1252	2824	2556	2127	2790	1640
## airdrop	4272	8952	17883	10533	7212	6779	7201	4515
## al	0	0	0	0	0	0	2312	1263
## analytica	0	0	0	5653	1681	1826	2917	0
## app	0	0	0	0	0	0	0	0
## bitcoin	2711	14780	14408	13655	7537	7178	11675	5826
## blockchain	679	3830	3165	9959	6873	5983	8528	5830
## btc	7909	28256	28966	32636	25442	21654	27668	20829
## busi	19	957	765	0	0	0	0	0
## buy	0	0	3049	3092	2252	1773	2703	2144
## call	0	0	0	0	2079	0	0	0
## cambridg	0	0	0	5210	0	1697	2915	0
## ce	25	1120	1013	0	0	0	0	1296
## ceo	0	0	0	0	0	0	2327	0
## chanc	0	0	10243	0	0	0	0	0
## check	23	0	0	0	0	0	0	0
## compani	0	0	0	2474	0	0	0	0
## con	24	0	727	0	0	0	0	0
## congress	0	0	0	0	0	0	2865	0
## crypto	1098	10822	5945	10172	8295	7665	9818	6544
## cryptocurr	956	5092	5437	5222	3318	3026	8716	2091
## data	0	0	1377	13700	8052	6269	11498	3051
## date	0	0	0	0	0	0	0	0
## dato	0	0	0	2483	0	0	0	0
## day	0	935	0	0	0	0	0	0
## de	113	6790	8049	24821	15463	13950	22702	12976
## delet	0	0	0	3964	1851	0	0	0
## di	0	0	0	3647	0	0	0	0
## earn	2223	5665	0	0	0	0	0	0
## el	28	1370	1071	4103	2548	2433	4281	2327
## en	120	5497	4486	9932	7427	7699	9991	7677
## eth	3241	11422	11191	13177	8228	6775	9134	6817
## exchang	0	3077	0	2945	0	1639	2622	0
## facebook	94	5427	4555	15673	10760	8853	15829	7125
## fb	0	0	0	2720	0	0	2346	0
## featur	0	0	0	0	0	0	0	0
## follow	38	1233	1023	0	0	0	0	1291
## free	5808	16399	25969	12939	5688	6081	4254	1327
## friend	1510	0	11701	0	0	0	0	0
## goal	0	0	9690	0	0	0	0	0
## hear	0	0	0	0	0	0	3143	0
## hour	0	0	0	0	0	1838	2616	0
## ico	0	7126	2803	8051	0	4069	8521	5162
## invest	0	2545	0	0	0	0	0	0
## join	6079	19880	19144	14644	12512	10303	11544	9214
## la	80	3471	2896	9927	6200	5685	7844	4939
## las	23	0	0	0	0	0	0	0
## le	0	0	0	2617	1695	0	2817	1396
## les	19	0	0	0	0	0	0	0
## link	0	0	8352	5435	0	0	0	0
## live	18	917	808	0	0	0	0	0
## los	0	0	0	0	0	1652	0	0

## mani	0	0	10111	0	0	0	0	0
## mark	0	0	0	2944	0	0	6420	0
## market	822	4974	4242	4148	3030	2923	3997	1837
## media	33	1383	1208	3226	2058	1885	3156	1705
## million	0	0	0	0	0	1853	0	0
## mine	0	2674	2724	0	0	0	0	0
## moment	26	1072	964	0	0	0	0	1193
## news	0	1296	827	0	1615	1580	0	0
## offer	0	0	5691	0	0	0	0	0
## page	66	2644	2094	3646	2885	2674	3505	2364
## para	20	1126	805	0	1709	1636	0	1655
## peopl	18	0	10731	3429	1921	2004	3763	0
## person	0	0	0	2847	0	0	0	0
## platform	0	4424	0	4653	4636	4337	4483	3521
## por	19	0	0	0	0	0	0	0
## post	32	1751	1211	2450	2007	1685	2311	1629
## price	1243	10972	6949	7200	9158	4847	7262	3282
## privaci	0	0	0	2567	2799	2136	4267	1673
## project	2154	11358	21611	14398	12914	12225	13507	9354
## question	0	0	0	0	0	0	3826	0
## radiocapitol	25	997	922	0	0	0	0	0
## reach	0	0	9778	0	0	0	0	0
## refer	0	0	5695	0	0	0	0	0
## regist	0	0	13344	4883	0	0	0	0
## scandal	0	0	0	3077	1805	1632	0	0
## secur	564	0	0	0	0	0	0	0
## sell	0	0	0	0	0	0	0	1423
## senat	0	0	0	0	0	0	3569	0
## share	0	0	14946	6956	0	1833	2849	0
## social	60	2131	1766	5031	3234	2631	4526	2540
## start	0	2745	7559	0	0	0	0	1323
## su	0	0	0	0	0	0	0	1254
## sur	0	914	0	0	0	0	0	0
## telegram	557	0	0	0	0	0	0	0
## time	28	0	7462	7755	1627	0	2645	0
## tip	19	0	0	0	0	0	0	0
## token	5349	17632	45257	23837	16045	13554	16261	9327
## trade	730	3604	3056	2933	0	1966	0	1375
## user	0	0	0	4479	2942	3965	5084	2144
## video	18	1236	766	0	0	0	0	0
## world	0	0	0	3240	0	0	0	0
## zuckerberg	0	0	0	4690	1625	2486	11162	1224
##	17	18						
## #actu	2078	1070						
## #ada	5662	3024						
## #airdrop	25835	12552						
## #altcoin	22645	11037						
## #amazon	1223	0						
## #bch	8765	3517						
## #binanc	11870	5286						
## #bitcoin	158495	75997						
## #bitcoincash	11788	7675						
## #bittrex	0	0						
## #blockchain	99271	49005						

## #bounti	12209	5916
## #btc	76877	35862
## #bts	0	0
## #busi	1457	712
## #cambridgeanalyt	0	0
## #cambridgeanalytica	1307	826
## #coin	0	0
## #coinbas	0	0
## #congress	0	0
## #costarica	0	1258
## #crowdfund	4572	0
## #crowdsal	4487	0
## #crypto	87376	42214
## #cryptocurr	91921	43949
## #cryptonew	1773	0
## #cybersecur	4301	0
## #dash	7772	2968
## #data	1233	0
## #date	0	832
## #deletefacebook	0	0
## #digitalmarket	1218	0
## #digitizecoin	0	0
## #earn	0	0
## #elsalvador	0	5021
## #energytoken	0	0
## #eo	10170	5725
## #erc20	5897	3412
## #escort	2142	5796
## #etc	0	0
## #eth	77083	37647
## #ether	14613	7550
## #ethereum	128290	63917
## #f8	0	1318
## #facebook	57189	32773
## #facebookdatabreach	0	0
## #facebookdataleak	0	0
## #facebookkg	0	0
## #faitsdiv	2069	1067
## #fintech	6752	1578
## #follow	0	0
## #free	0	0
## #freecoin	0	0
## #freetoken	13292	5198
## #game	0	0
## #gdpr	0	0
## #geoloc	4390	0
## #girl	0	1818
## #googl	3154	1144
## #grab	0	0
## #gto	0	0
## #guatemala	0	5024
## #hiphop	1102	0
## #ico	66002	30532
## #info	2118	1077

## #instagram	7987	3791
## #invest	7421	1640
## #justic	2266	1125
## #linkedin	1390	0
## #litecoin	16692	8097
## #live	0	0
## #ltc	19869	9654
## #maga	0	0
## #mana	0	0
## #market	2902	1449
## #markzuckerberg	0	0
## #média	2093	1075
## #mexico	2147	5094
## #money	0	0
## #music	2985	1497
## #neo	11323	4541
## #new	12422	7881
## #nowplay	0	1820
## #np	6086	5036
## #opirium	0	0
## #panama	0	4941
## #pdata	0	0
## #pinterest	0	0
## #privaci	1785	1170
## #qanon	0	0
## #radiocapitol	2069	1065
## #rippl	19950	9200
## #rt	0	4098
## #smm	0	0
## #snapchat	1688	725
## #socialmedia	5996	3210
## #socialmediamarket	1045	0
## #soundcloud	1213	0
## #submityourmus	0	0
## #tge	5659	2672
## #tinder	0	767
## #token	14649	7684
## #tokensal	22074	9379
## #trade	9931	4516
## #travel	2145	1876
## #tron	6643	2890
## #trump	0	0
## #trx	18464	9152
## #twitter	8827	3895
## #uniqu	2073	1067
## #usa	2307	5215
## #usd	0	0
## #verg	0	0
## #video	6392	1499
## #vip	0	5005
## #webradio	2069	1067
## #whatsapp	0	1342
## #xem	3921	0
## #xlm	5308	2706

## #xmr	0	0
## #xrb	0	0
## #xrp	26585	12577
## #xvg	5656	3131
## #youtub	4256	2146
## #zuckerberg	0	712
## 000facebook	1191	1052
## 000guarium	0	0
## account	0	0
## ad	1670	772
## airdrop	6496	2585
## al	0	0
## analytica	0	0
## app	0	820
## bitcoin	8137	3793
## blockchain	7076	1368
## btc	27933	14644
## busi	0	0
## buy	2452	1038
## call	0	0
## cambridg	0	0
## ce	1241	669
## ceo	0	0
## chanc	0	0
## check	0	0
## compani	0	0
## con	1020	0
## congress	0	0
## crypto	7519	1661
## cryptocurr	2740	1309
## data	1886	1282
## date	0	1409
## dato	0	0
## day	0	0
## de	11409	11511
## delet	0	0
## di	0	0
## earn	0	0
## el	1834	1198
## en	6994	3685
## eth	8523	4759
## exchang	0	0
## facebook	5951	3817
## fb	0	0
## featur	0	815
## follow	1173	675
## free	1846	0
## friend	0	0
## goal	0	0
## hear	0	0
## hour	0	0
## ico	0	0
## invest	0	0
## join	8760	2836

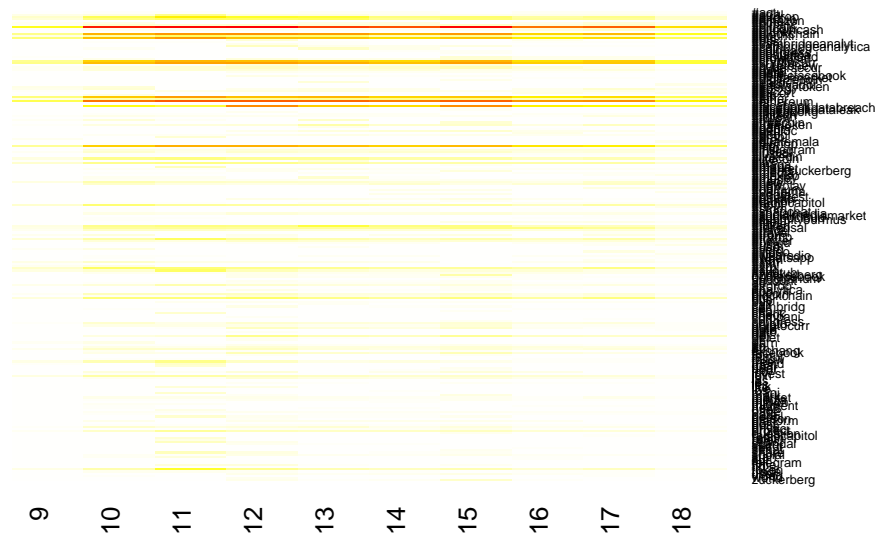
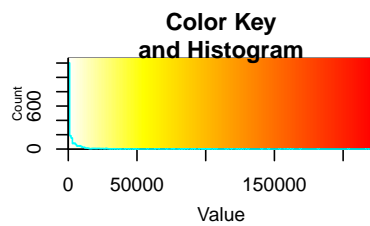
## la	4419	2675
## las	0	0
## le	1059	834
## les	0	725
## link	0	0
## live	0	0
## los	0	0
## mani	0	0
## mark	0	0
## market	4216	1311
## media	1494	672
## million	0	0
## mine	0	0
## moment	1192	0
## news	0	703
## offer	0	0
## page	2321	1154
## para	1425	1028
## peopl	0	0
## person	0	0
## platform	0	0
## por	0	0
## post	1722	898
## price	4641	1921
## privati	0	780
## project	8053	3245
## question	0	0
## radiocapitol	1091	0
## reach	0	0
## refer	0	0
## regist	0	0
## scandal	0	0
## secur	0	0
## sell	0	0
## senat	0	0
## share	0	0
## social	2228	1072
## start	0	0
## su	0	0
## sur	0	0
## telegram	0	0
## time	0	0
## tip	0	0
## token	13797	4653
## trade	1859	0
## user	1281	895
## video	1042	0
## world	0	0
## zuckerberg	0	763

- Research Question

- Do we see new words appearing on a particular week

Analysis

Heatmap

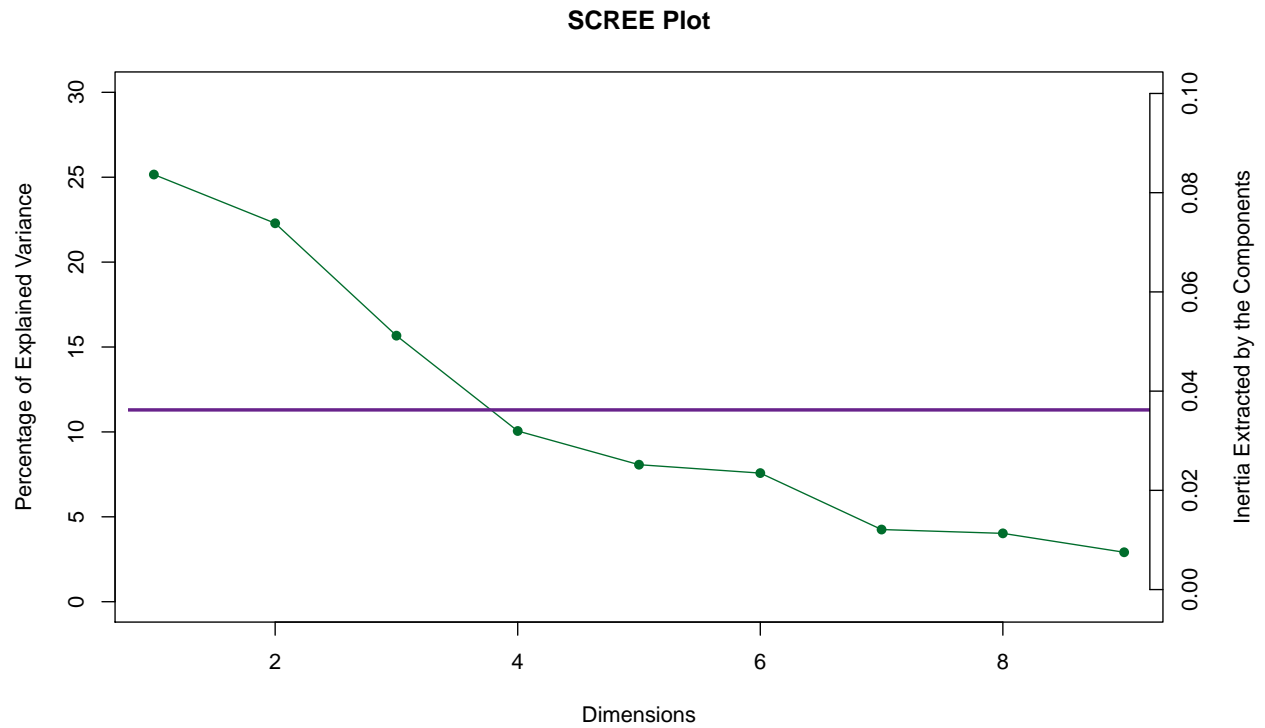


Scree Plot

Gives amount of information explained by corresponding component. Gives an intuition to decide which components best represent data in order to answer the research question.

P.S. The most contribution component may not always be most useful for a given research question.

```
PTCA4CATA::PlotScree(ev = resCA$sym$ExPosition.Data$eigs,
                      #p.ev = we_data_inf$Inference.Data$components$p.vals,
                      title = 'SCREE Plot',
                      plotKaiser = TRUE
)
```

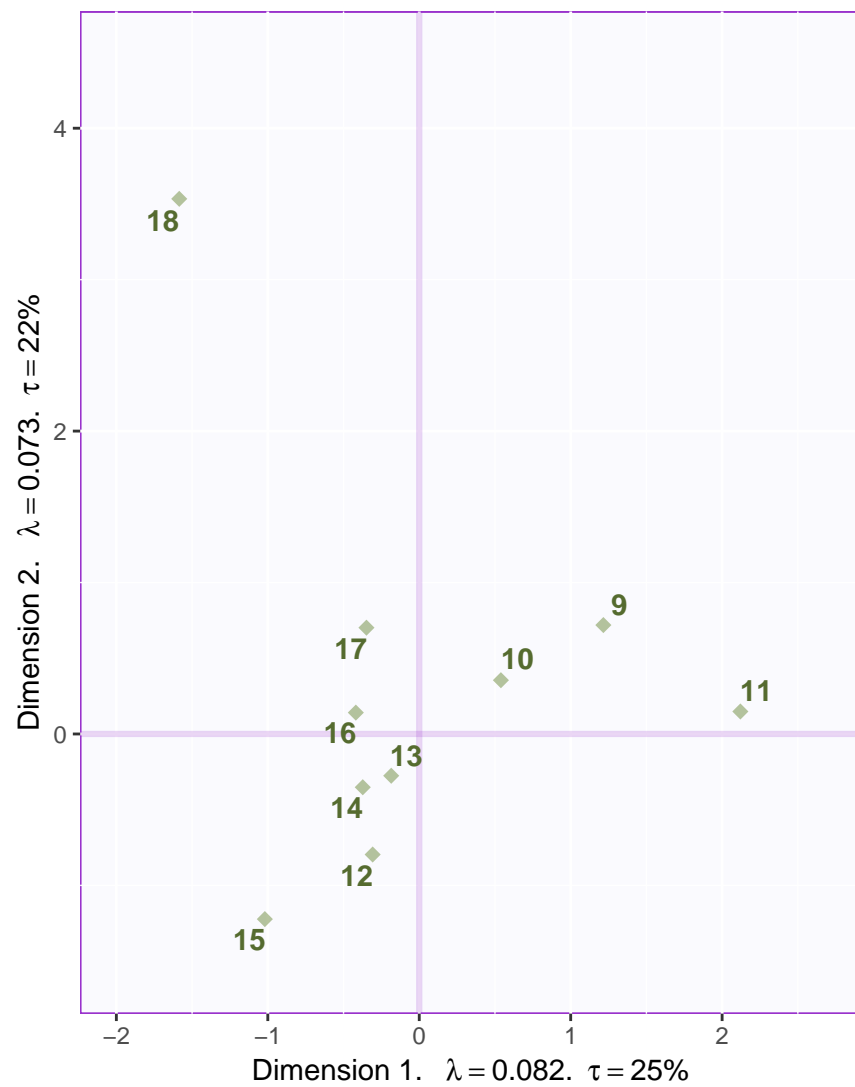


Factor Scores

Asymmetric Plot

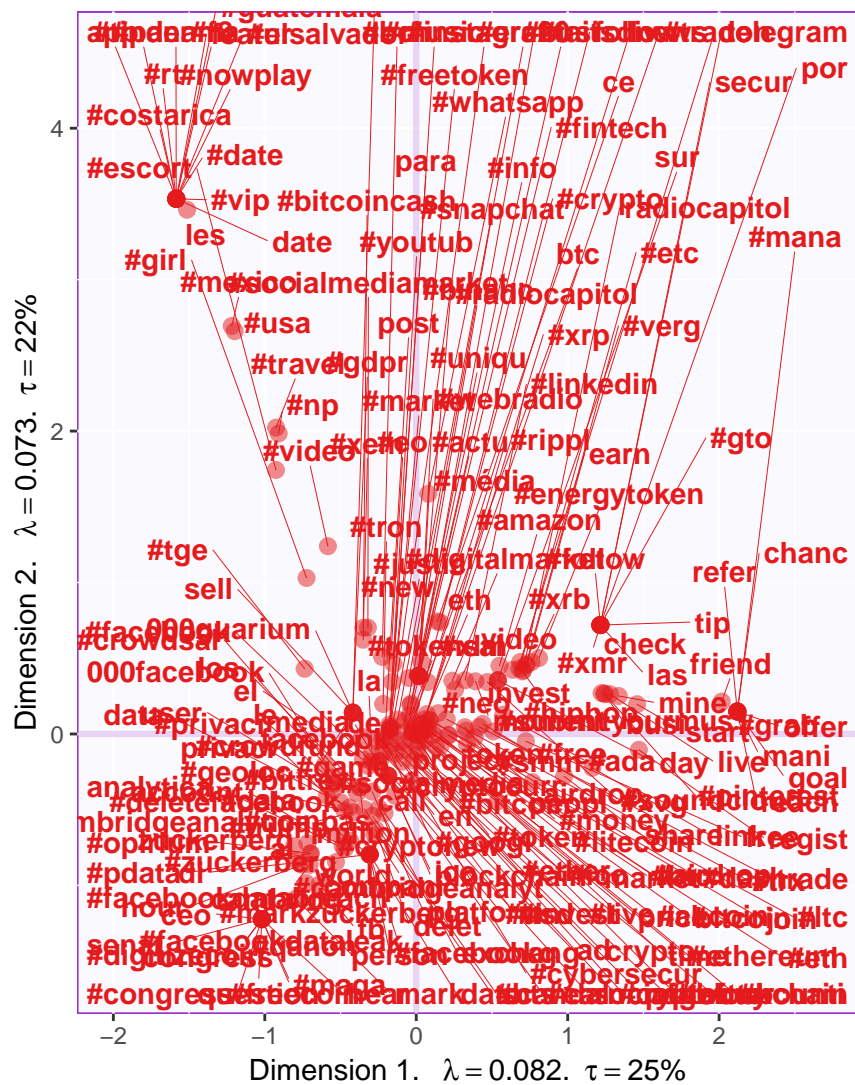
```
#### Asymmetric Plot

map.IJ.asym <- asymMap$baseMap + asymMap$J_labels +
  asymMap$J_points + labels4CA #+ legend$zeMap_dots + legend$zeMap_text
print(map.IJ.asym)
```



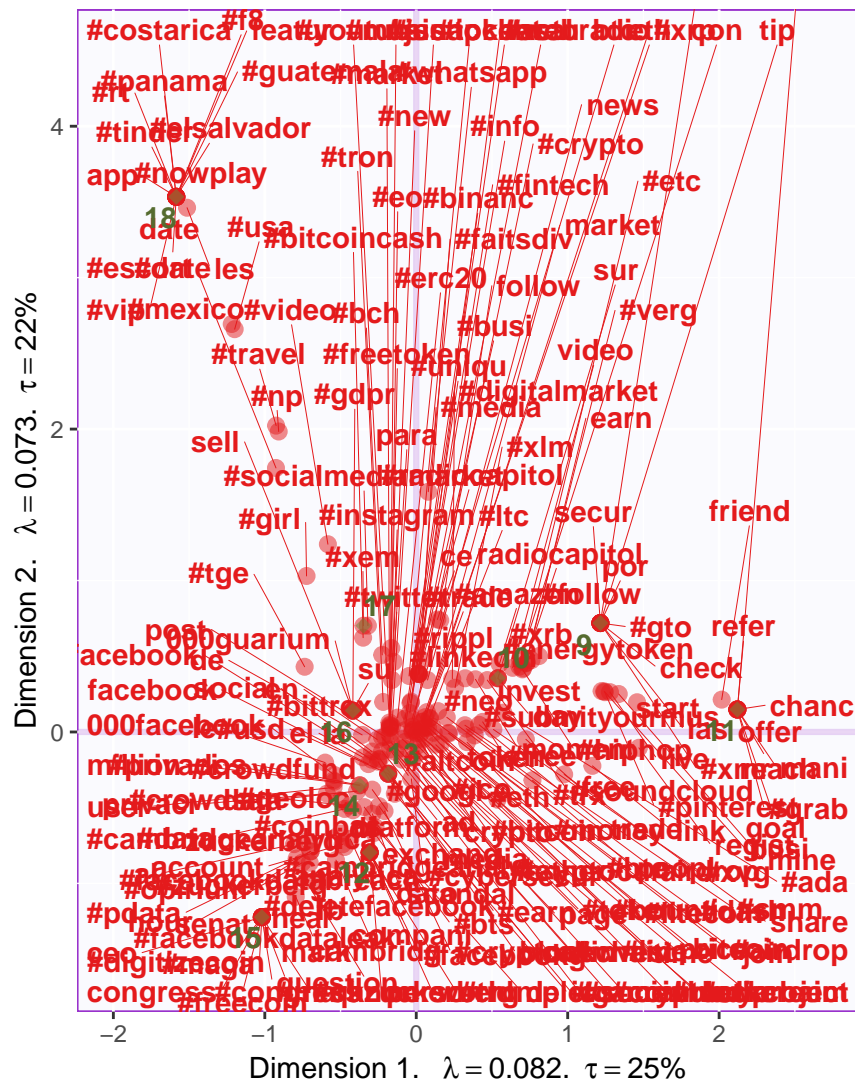
```
#### Asymmetric Plot
```

```
map.IJ.asym <- asymMap$baseMap + asymMap$I_labels +
  asymMap$I_points + labels4CA #+ legend$zeMap_dots + legend$zeMap_text
print(map.IJ.asym)
```



```
#### Asymmetric Plot
```

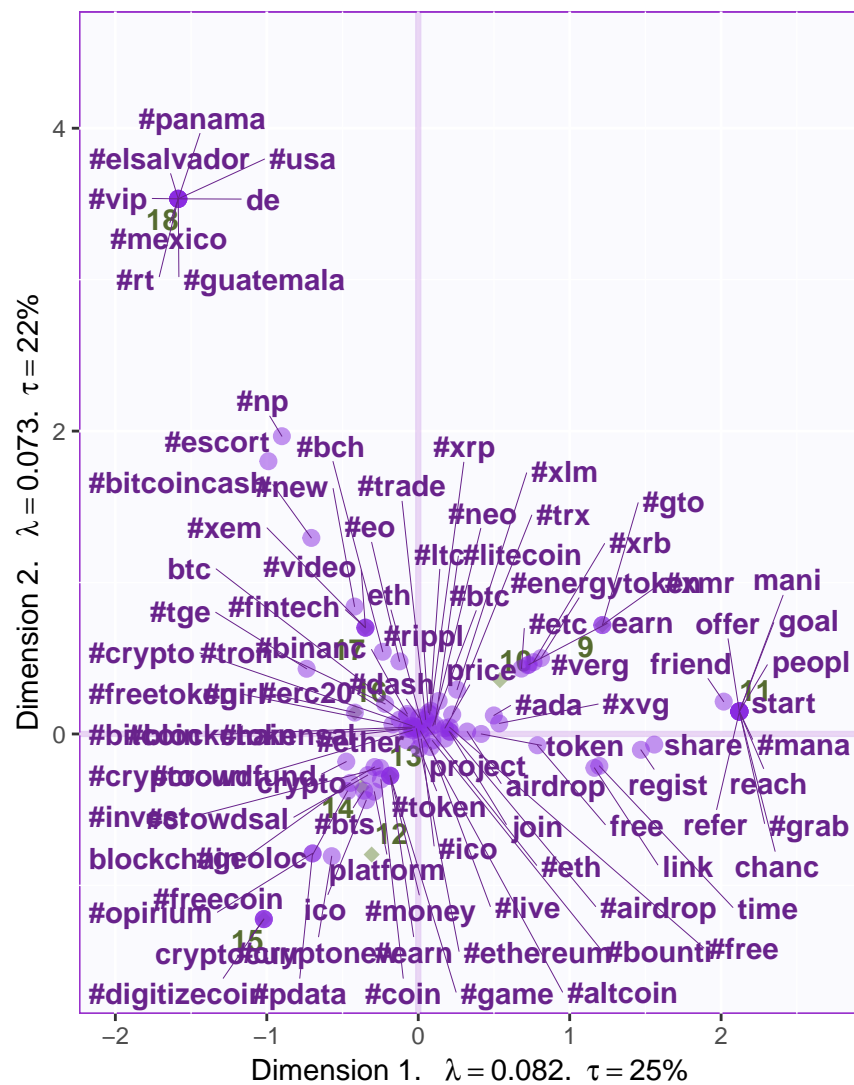
```
map.IJ.asym <- asymMap$baseMap + asymMap$I_labels +
  asymMap$I_points + asymMap$J_labels +
  asymMap$J_points + labels4CA #+ legend$zeMap_dots + legend$zeMap_text
print(map.IJ.asym)
```



```
tweets_eth_supp <- supplementaryRows(tweets_eth, resCA.asym)
asymMap_supp <- createFactorMapIJ(tweets_eth_supp$fii, Fj,
                                   #col.points.i = color4Authors,
                                   #col.labels.i = color4Authors
                                   )

map.IJ.asym <- asymMap$baseMap + #asymMap$I_labels +
  #asymMap$I_points +
  asymMap$J_labels +
  asymMap$J_points + labels4CA + #+ legend$zeMap_dots + legend$zeMap_text
  asymMap_supp$I_points + asymMap_supp$I_labels

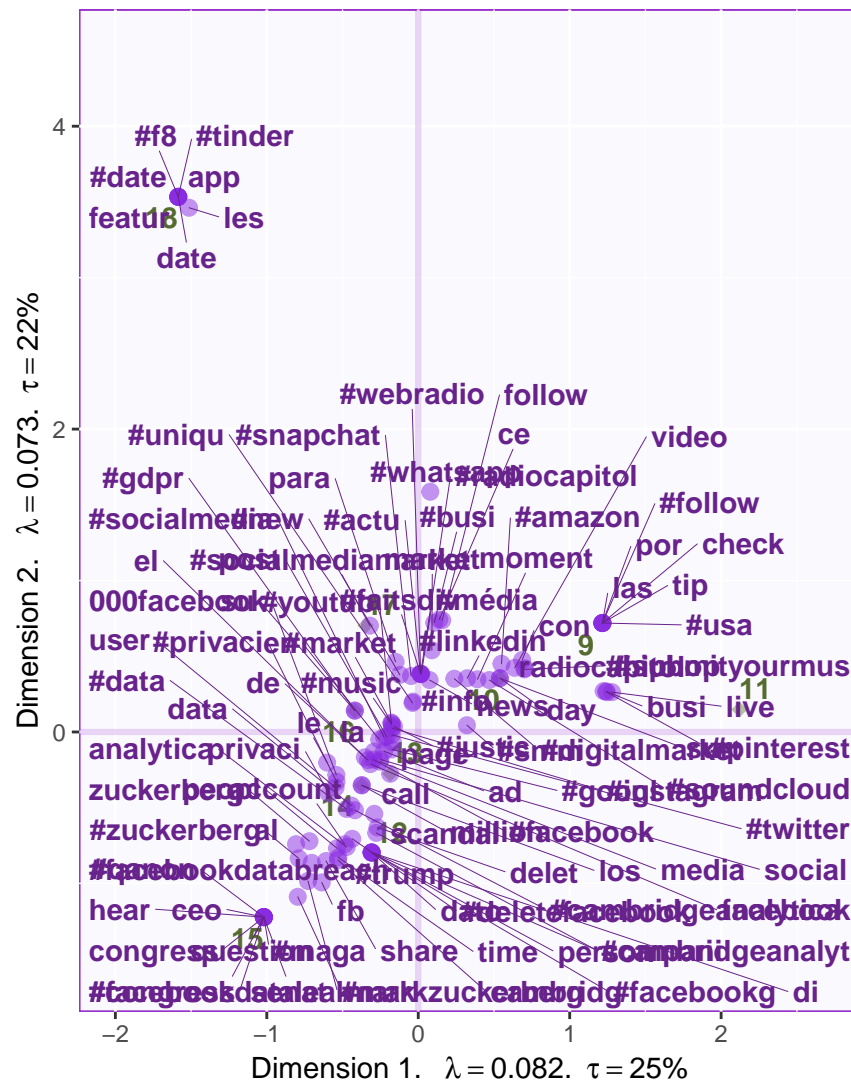
print(map.IJ.asym)
```



```
tweets_facebook_supp <- supplementaryRows(tweets_facebook, resCA.asym)
asymMap_supp <- createFactorMapIJ(tweets_facebook_supp$fii, Fj,
                                   #col.points.i = color4Authors,
                                   #col.labels.i = color4Authors
                                   )

map.IJ.asym <- asymMap$baseMap + #asymMap$I_labels +
  #asymMap$I_points +
  asymMap$J_labels +
  asymMap$J_points + labels4CA + #+ legend$zeMap_dots + legend$zeMap_text
  asymMap_supp$I_points + asymMap_supp$I_labels

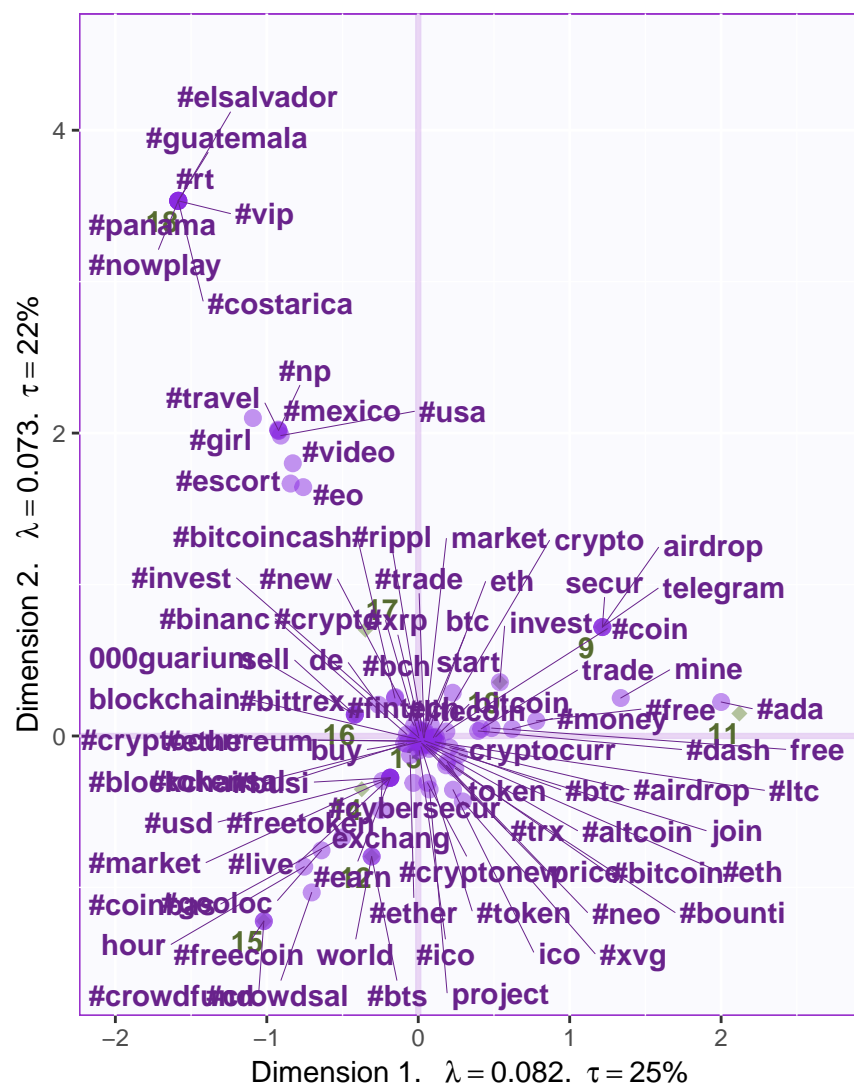
print(map.IJ.asym)
```



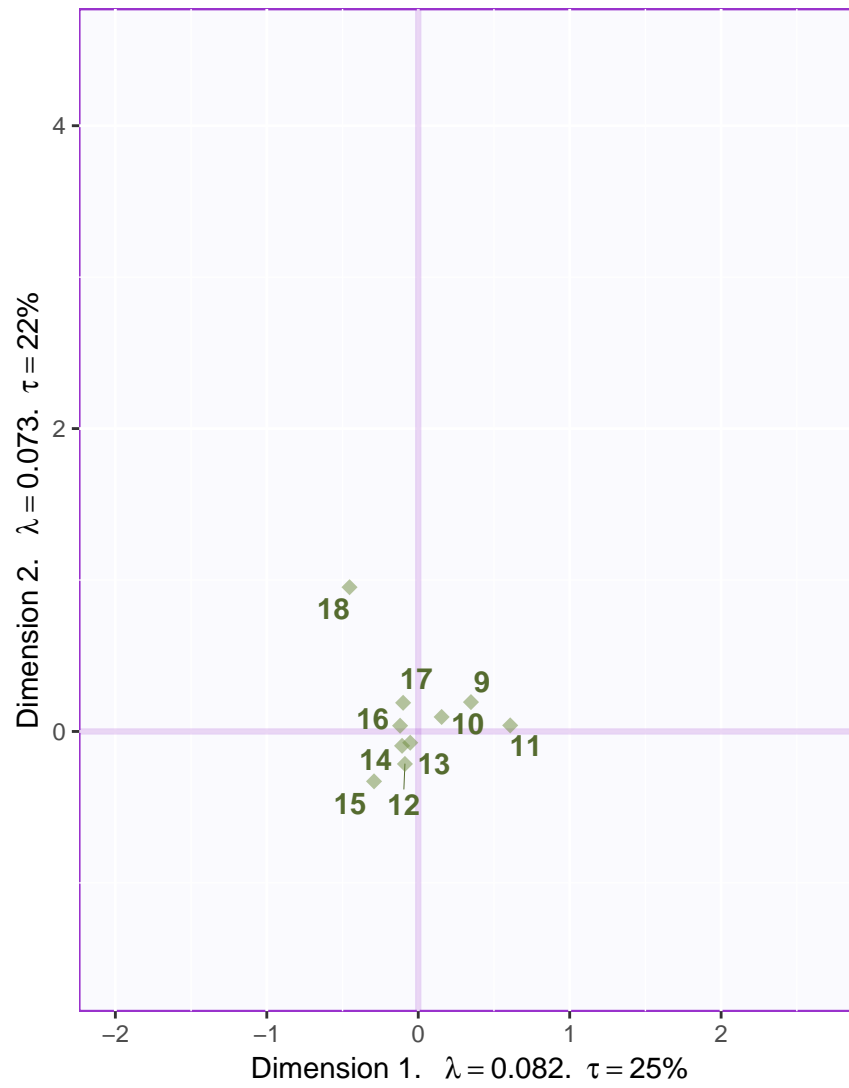
```
tweets_bitcoin_supp <- supplementaryRows(tweets_bitcoin, resCA.asym)
asymMap_supp <- createFactorMapIJ(tweets_bitcoin_supp$fii, Fj,
                                   #col.points.i = color4Authors,
                                   #col.labels.i = color4Authors
                                   )

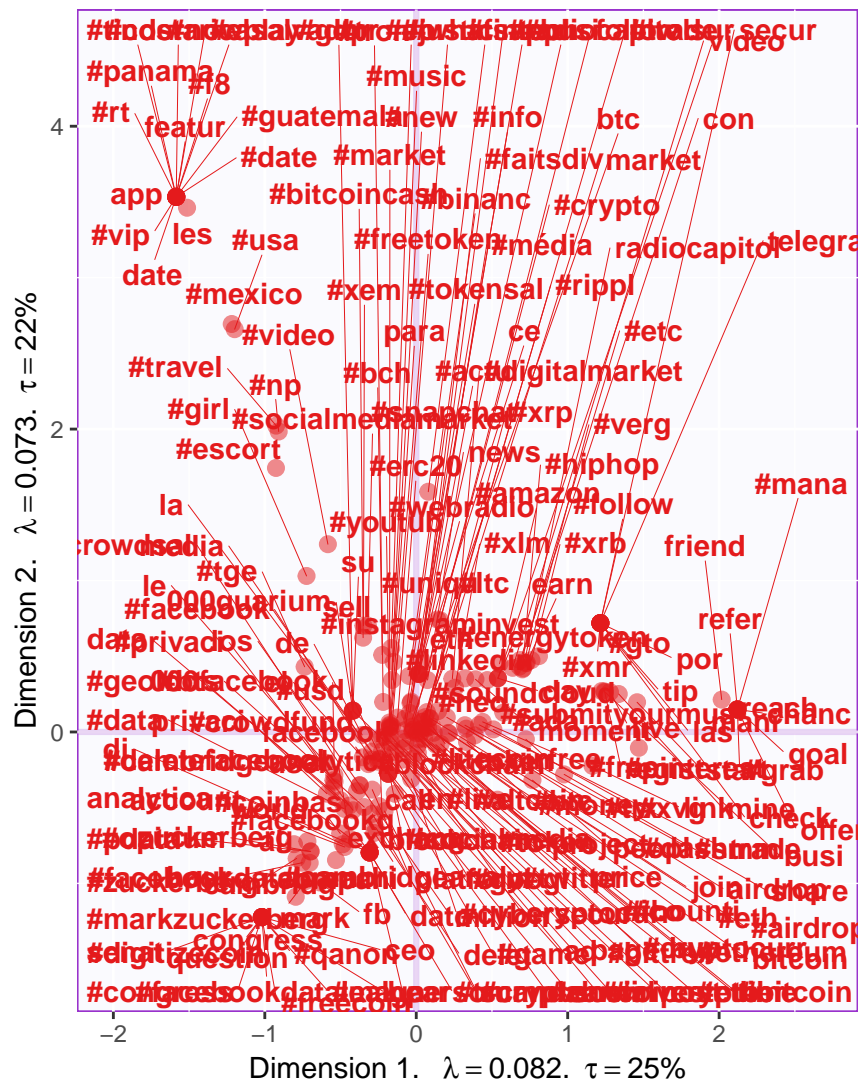
map.IJ.asym <- asymMap$baseMap + #asymMap$I_labels +
  #asymMap$I_points +
  asymMap$J_labels +
  asymMap$J_points + labels4CA + #+ legend$zeMap_dots + legend$zeMap_text
  asymMap_supp$I_points + asymMap_supp$I_labels

print(map.IJ.asym)
```

Symmetric Plot

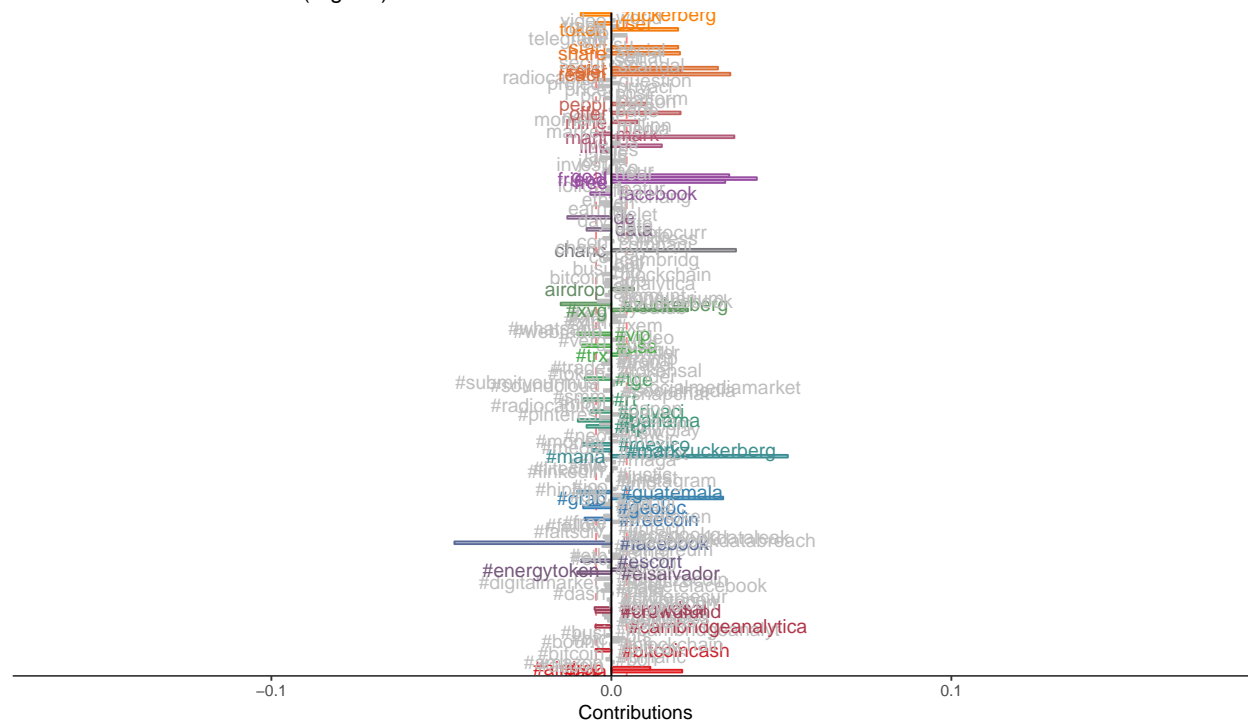




Most Contributing Variables

```
PTCA4CATA::PrettyBarPlot2(ctr.I[,1],
  threshold = 1 / NROW(ctr.I),
  font.size = 4,
  color4bar = gplots::col2hex(color4I),
  color4ns = 'grey',
  main = 'Observations: Contributions (Signed)',
  ylab = 'Contributions', ylim = c(1.2*min(ctr.I),
    1.2*max(ctr.I) ),
  horizontal = FALSE )
```

Observations: Contributions (Signed)



```
PTCA4CATA::PrettyBarPlot2(ctr.J[,1],
  threshold = 1 / NROW(ctr.J),
  font.size = 4,
  color4bar = gplots::col2hex(color4J),
  color4ns = 'grey',
  main = 'Observations: Contributions (Signed)',
  ylab = 'Contributions', ylim = c(1.2*min(ctr.J),
    1.2*max(ctr.J) ),
  horizontal = FALSE )
```

Observations: Contributions (Signed)

