

phase 1 - combined or separate?

January 18, 2019

```
In [157]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [158]: df_all = pd.read_table('./data/all208_composition.txt', sep='\t',
                                header=None, index_col=0)
df_all[1] = df_all[1].apply(lambda x: x[54:61])
df_all.columns= list(range(-1, df_all.shape[1]-1))
df_all.set_index(-1, inplace=True)
df_all.sort_values(by=-1, inplace=True)
df_all.head(12)
```

```
Out [158]:
```

	0	1	2	3	4	5	6	\
-1								
AT-2014	0.000681	0.000939	0.138623	0.001119	0.435915	0.001198	0.265989	
AT-2014	0.000534	0.069386	0.000835	0.000877	0.283020	0.000940	0.404743	
AT-2014	0.000673	0.000927	0.001052	0.001105	0.467584	0.001183	0.003337	
AT-2014	0.000551	0.000758	0.000861	0.000904	0.422957	0.000968	0.123983	
AT-2014	0.019388	0.274380	0.000804	0.000844	0.366707	0.000904	0.087482	
AT-2014	0.011104	0.148211	0.000897	0.000942	0.609253	0.001009	0.139735	
AT-2014	0.000319	0.345536	0.000498	0.000523	0.461252	0.006410	0.013278	
AT-2014	0.000649	0.012803	0.108200	0.001065	0.831982	0.036870	0.003217	
AT-2014	0.354021	0.000647	0.000734	0.000771	0.136751	0.000826	0.002330	
AT-2014	0.000599	0.055790	0.000936	0.000983	0.756967	0.001053	0.002970	
AT-2015	0.576504	0.000758	0.000861	0.051426	0.311809	0.000968	0.033043	
AT-2015	0.281617	0.000844	0.000957	0.001006	0.504219	0.001077	0.025517	
	7	8	9	10	11			
-1								
AT-2014	0.001451	0.000483	0.101337	0.001506	0.050759			
AT-2014	0.001138	0.000379	0.001016	0.236553	0.000579			
AT-2014	0.507816	0.000477	0.013630	0.001488	0.000729			
AT-2014	0.071903	0.354046	0.021255	0.001217	0.000596			
AT-2014	0.001095	0.000365	0.000977	0.246497	0.000557			
AT-2014	0.001222	0.000407	0.001091	0.085509	0.000621			
AT-2014	0.170302	0.000226	0.000606	0.000705	0.000345			

AT-2014	0.001382	0.000460	0.001233	0.001435	0.000703
AT-2014	0.009624	0.491855	0.000893	0.001039	0.000509
AT-2014	0.001275	0.000425	0.111069	0.001324	0.066607
AT-2015	0.001172	0.000391	0.001047	0.001217	0.020805
AT-2015	0.001304	0.000434	0.001164	0.181197	0.000663

```
In [165]: df_combined = pd.read_table('./data/allCombined_composition.txt',sep='\t',
                                     header=None, index_col=0)
df_combined[1] = df_combined[1].apply(lambda x: x[63:70])
df_combined.columns= list(range(-1,df_combined.shape[1]-1))
df_combined.set_index(-1, inplace=True)
df_combined.sort_values(by=-1, inplace=True)
df_combined.head(12)
```

```
Out[165]:
```

	0	1	2	3	4	5	6	\
-1								
AT-2014	0.000026	0.000023	0.000086	0.002035	0.000022	0.171250	0.000075	
AT-2015	0.000027	0.000023	0.000087	0.000051	0.000023	0.361874	0.314097	
AT-2016	0.000024	0.000021	0.000077	0.000045	0.000020	0.191976	0.002752	
AT-2017	0.000018	0.000016	0.387900	0.000034	0.000015	0.186532	0.000726	
AT-2018	0.000701	0.000016	0.168142	0.000034	0.000015	0.454743	0.018500	
SE-2014	0.000030	0.000027	0.000099	0.000057	0.000026	0.145340	0.000087	
SE-2015	0.000022	0.000019	0.000070	0.000041	0.410357	0.024722	0.001680	
SE-2016	0.000021	0.000018	0.000068	0.000040	0.000018	0.049227	0.000060	
SE-2017	0.000021	0.000018	0.000068	0.006366	0.000018	0.094535	0.000060	
SE-2018	0.000020	0.000018	0.000066	0.000039	0.403378	0.067262	0.000058	
TE-2014	0.000015	0.000013	0.000594	0.400341	0.000012	0.011783	0.000588	
TE-2015	0.000018	0.445274	0.002786	0.000034	0.000015	0.000372	0.001416	

	7	8	9	10	11
-1					
AT-2014	0.383139	0.000071	0.441189	0.002024	0.000060
AT-2015	0.000037	0.000072	0.270265	0.000039	0.053405
AT-2016	0.331110	0.000064	0.473821	0.000035	0.000054
AT-2017	0.000025	0.000048	0.424620	0.000026	0.000041
AT-2018	0.021206	0.012347	0.321493	0.002760	0.000042
SE-2014	0.000042	0.400051	0.454127	0.000044	0.000069
SE-2015	0.000030	0.000057	0.562922	0.000032	0.000049
SE-2016	0.000029	0.000056	0.408107	0.000031	0.542326
SE-2017	0.000029	0.000056	0.525498	0.000031	0.373299
SE-2018	0.000028	0.000055	0.500571	0.000798	0.027706
TE-2014	0.000020	0.000039	0.586540	0.000021	0.000033
TE-2015	0.000025	0.000048	0.549942	0.000027	0.000042

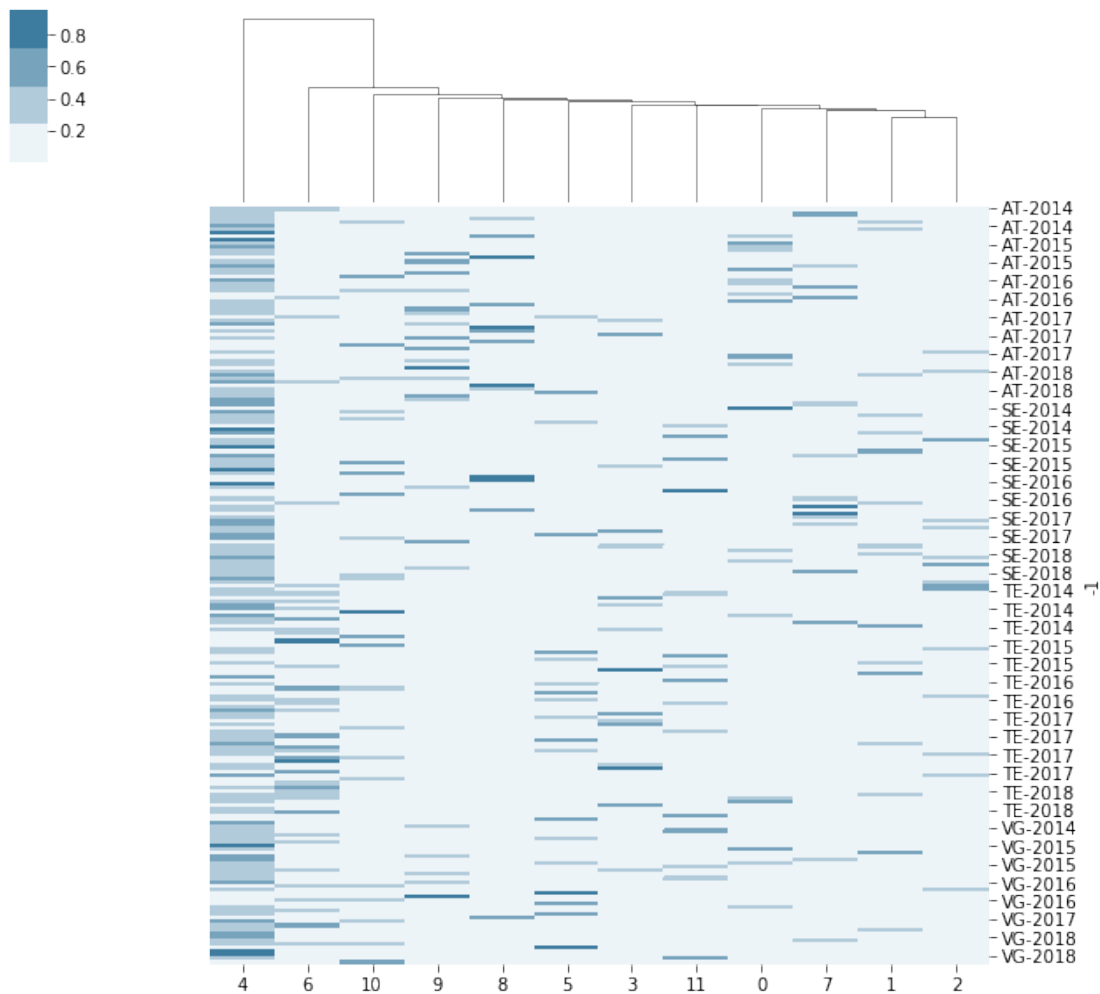
```
In [123]: cmap = sns.choose_cubehelix_palette(as_cmap=False)
cmap = sns.choose_light_palette(as_cmap=False)
```

```
interactive(children=(IntSlider(value=9, description='n_colors', max=16, min=2), FloatSlider(v
```

```
interactive(children=(IntSlider(value=179, description='h', max=359), IntSlider(value=49, desc:
```

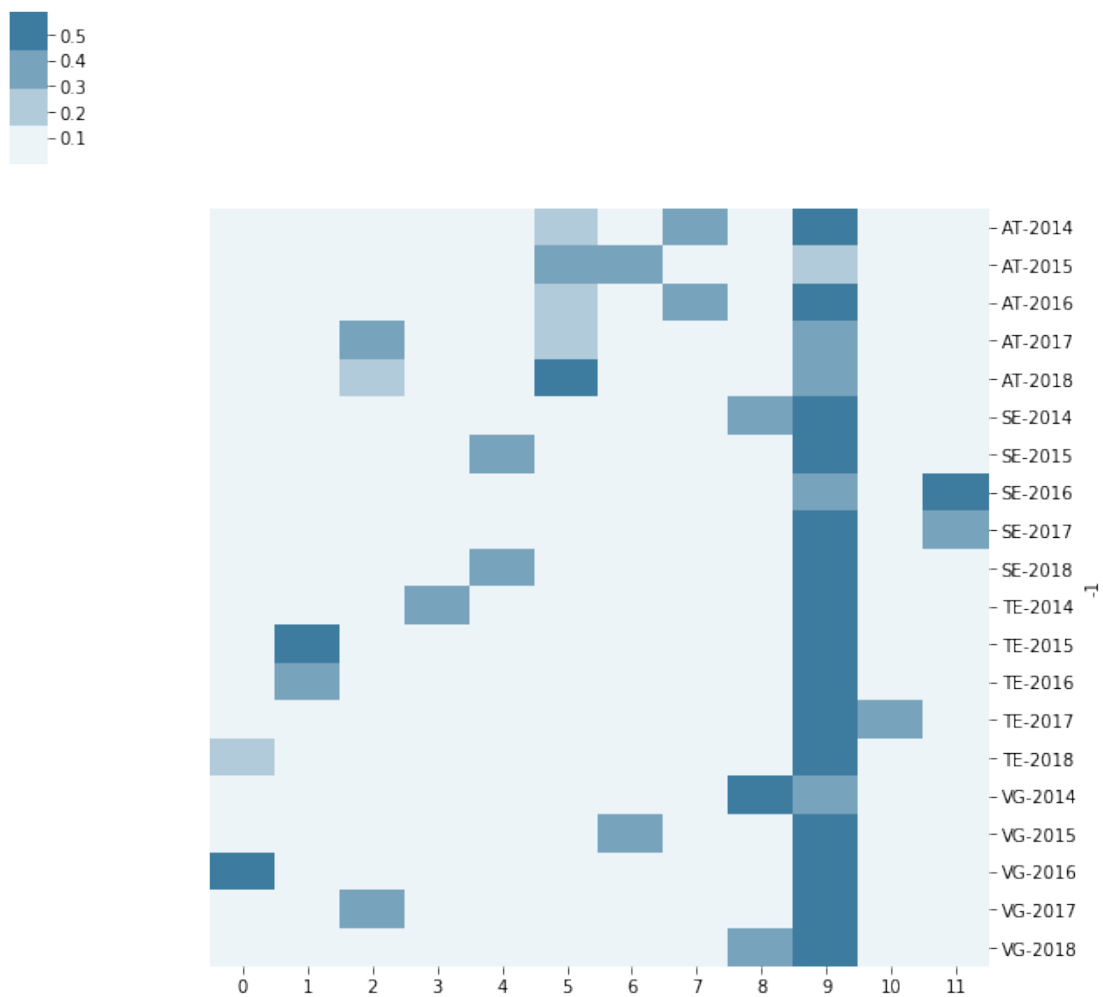
```
In [163]: sns.clustermap(df_all,row_cluster=False,col_cluster=True, cmap=cmap )
          sns.heatmap
```

```
Out[163]: <function seaborn.matrix.heatmap(data, vmin=None, vmax=None, cmap=None, center=None,
```



```
In [166]: sns.clustermap(df_combined,row_cluster=False,col_cluster=False, cmap=cmap )
          sns.heatmap
```

```
Out[166]: <function seaborn.matrix.heatmap(data, vmin=None, vmax=None, cmap=None, center=None,
```



In []: