

Similarity Score Graphs For Consensus Spectras

Kiana Choubak

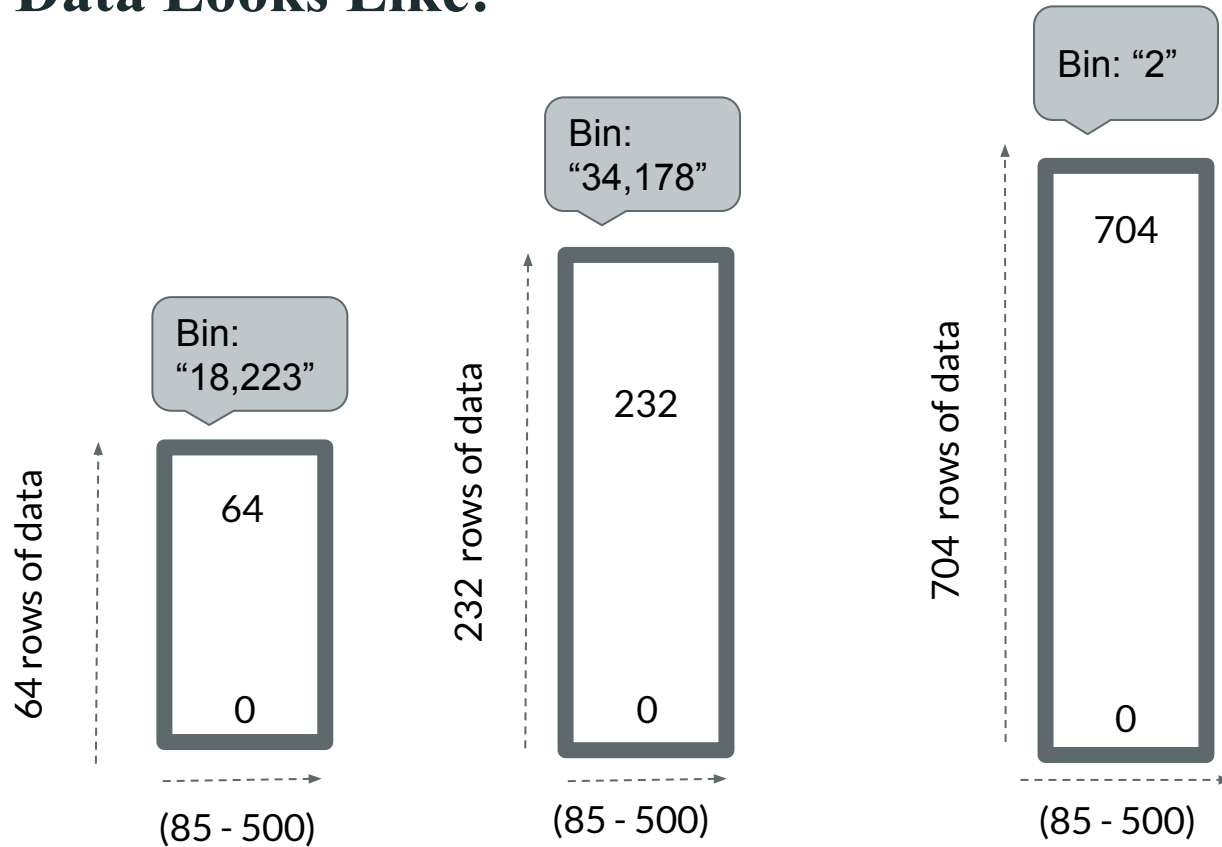
- Data set to compute the similarity scores for consensus spectra with different reference spectrum
- {'18,223': 64, '2': 704, '34,178': 232}

```
data = read_csv('data.csv')
data.sample(frac=1)
```

	spectra_id	bin_id	spectra
13	5,672,698	2	85:364.0 88:10.0 92:45.0 93:1.0 97:28.0 98:2.0...
391	26,758	2	85:1320.0 86:7638.0 87:2416.0 88:1930.0 89:977...
39	5,676,551	2	85:673.0 86:91.0 89:212.0 90:22.0 95:34.0 97:5...
127	5,686,330	18,223	85:358.0 86:330.0 87:93.0 88:71.0 89:119.0 92:...
920	5,785,764	2	85:969.0 86:176.0 88:71.0 90:19.0 93:3.0 94:70...
...
819	107,960	2	85:1451.0 86:9212.0 87:2790.0 88:2271.0 89:555...
501	52,147,333	34,178	85:706.0 86:4060.0 87:814.0 88:574.0 89:467.0 ...
243	22,192	2	85:1158.0 86:6672.0 87:2109.0 88:1668.0 89:605...
877	16,096,210	34,178	85:734.0 86:51.0 87:31.0 89:83.0 90:46.0 93:28...
728	5,892,071	2	86:3051.0 87:844.0 88:636.0 89:549.0 90:149.0 ...

1000 rows × 3 columns

How the Data Looks Like:



Calculate Consensus Spectra:



Count of Spectra (85-500)

Intensity Value

85	86	87	88	89	90	...	500
(R_1)	(R_1)	(R_1)	(R_1)	(R_1)	(R_1)	(R_1)	(R_1)
$(R_1 + R_2)$	$(R_1 + R_2)$	$(R_1 + R_2)$	$(R_1 + R_2)$	$(R_1 + R_2)$	$(R_1 + R_2)$	$(R_1 + R_2)$	$(R_1 + R_2)$
$(R_1 + R_2 + R_3)$	$(R_1 + R_2 + R_3)$	$(R_1 + R_2 + R_3)$	$(R_1 + R_2 + R_3)$	$(R_1 + R_2 + R_3)$	$(R_1 + R_2 + R_3)$	$(R_1 + R_2 + R_3)$	$(R_1 + R_2 + R_3)$
$(R_1 + R_2 + R_3 + R_4)$	$(R_1 + R_2 + R_3 + R_4)$	$(R_1 + R_2 + R_3 + R_4)$	$(R_1 + R_2 + R_3 + R_4)$	$(R_1 + R_2 + R_3 + R_4)$	$(R_1 + R_2 + R_3 + R_4)$	$(R_1 + R_2 + R_3 + R_4)$	$(R_1 + R_2 + R_3 + R_4)$
...
$(R_1 + R_2 + \dots + R_n)$	$(R_1 + R_2 + \dots + R_n)$	$(R_1 + R_2 + \dots + R_n)$	$(R_1 + R_2 + \dots + R_n)$	$(R_1 + R_2 + \dots + R_n)$	$(R_1 + R_2 + \dots + R_n)$	$(R_1 + R_2 + \dots + R_n)$	$(R_1 + R_2 + \dots + R_n)$



Compute The Score:

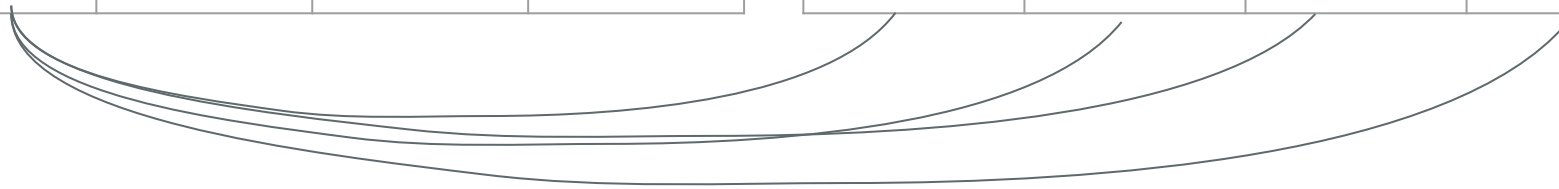
- Intensity consensus is a dictionary with bin number as the key, the value is a 2 dimensional array.
 - Each row represents a consensus spectrum
- Used Cosine Similarity Method to compute the score

Intensity (500 - 85 + 1 = 416 columns for all three bins)

<u>Column 0</u>	<u>Column 1</u>	<u>...</u>	<u>Column n</u>

Consensus Spectra

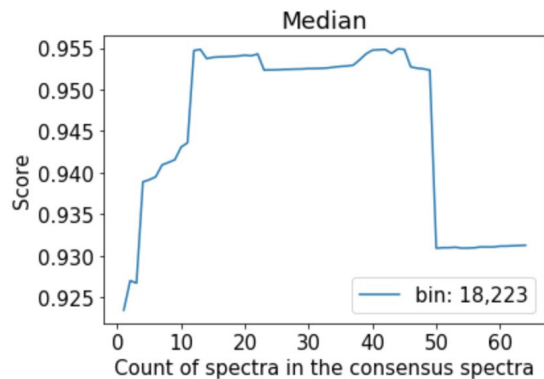
<u>Column 0</u>	<u>Column 1</u>	<u>...</u>	<u>Column n</u>



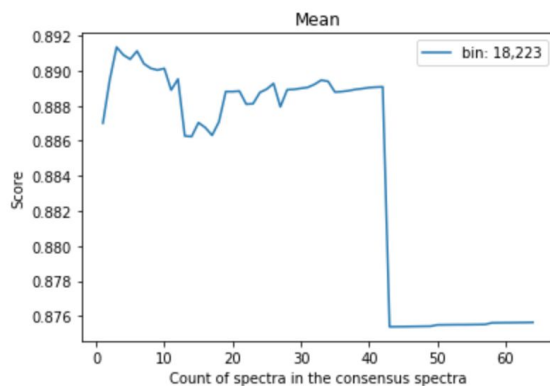
Similarity Score Graph:

- **Y-axis: Mean, Median, Middle Value - Score** = Comparing consensus spectra vs different intensity values (different reference intensity values) at the same time
- **X-axis:** Count of Consensus Spectra

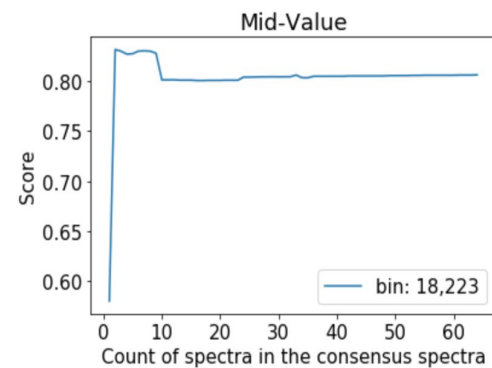
```
1 spectra_out.similarity_plot(metric = "median")
```



```
1 spectra_out.similarity_plot(metric = "mean")
```

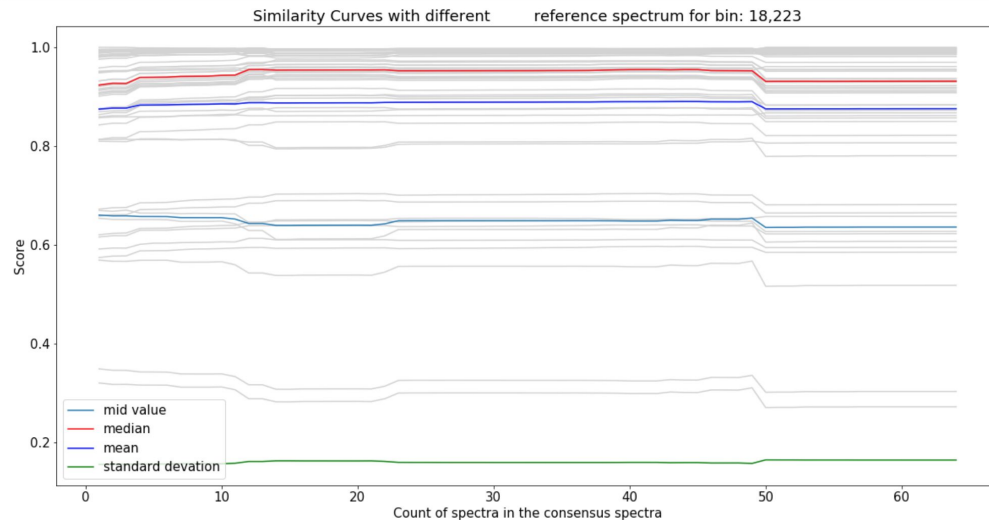
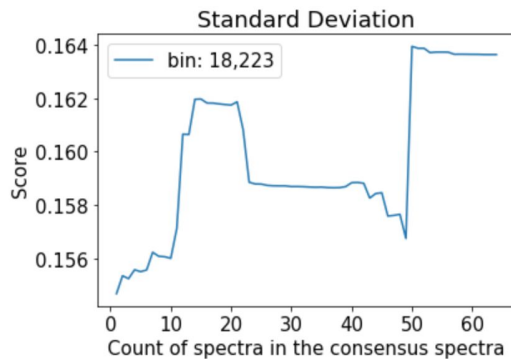


```
1 spectra_out.similarity_plot()
```

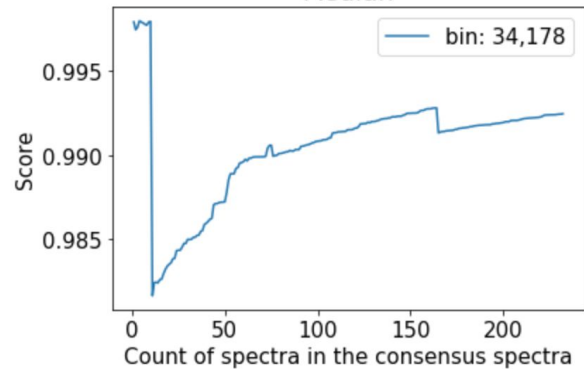


```
1 spectra_out.plot_metrics()
```

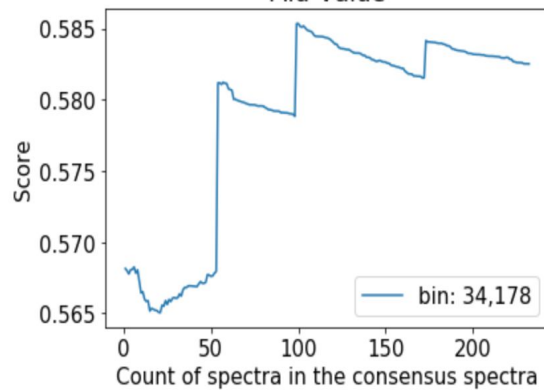
```
1 spectra_out.similarity_plot(metric = "std")
```



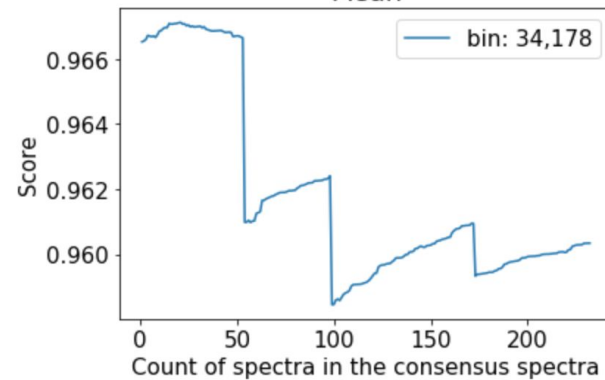
Median



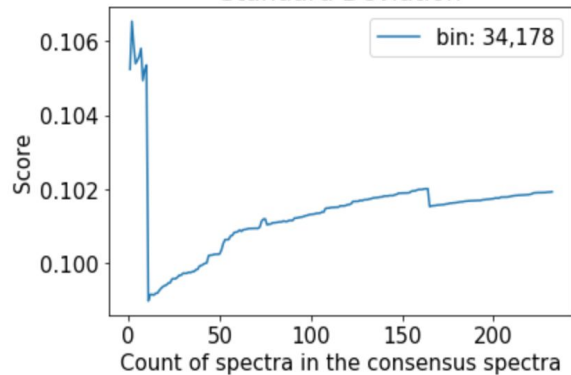
Mid-Value



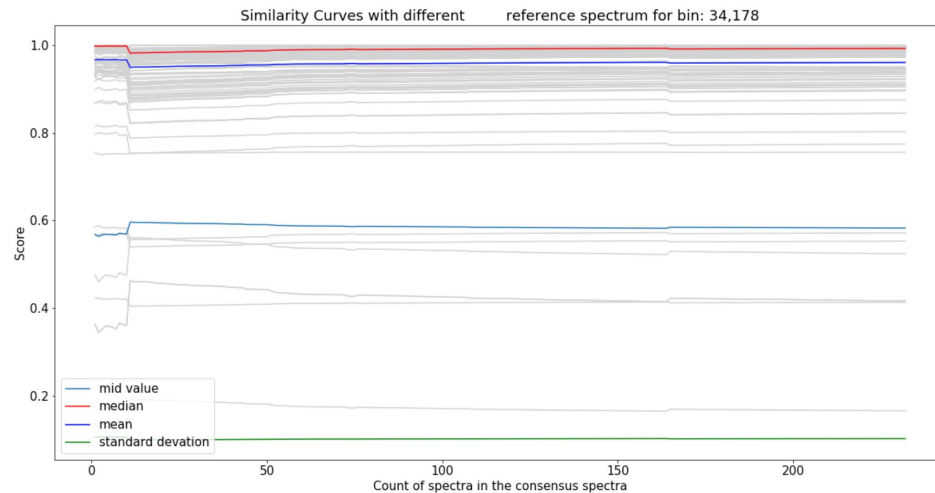
Mean

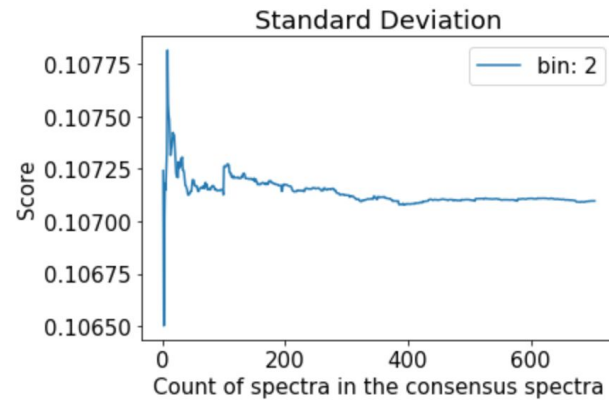
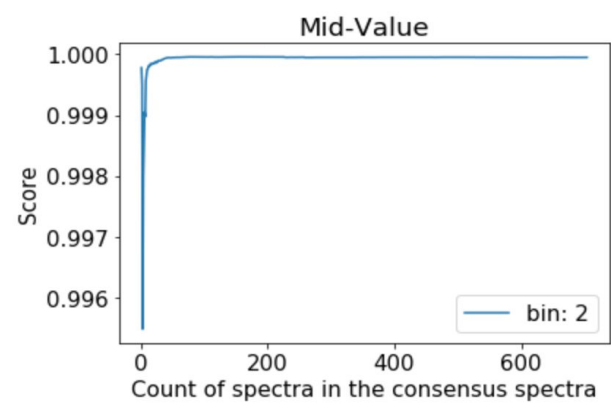
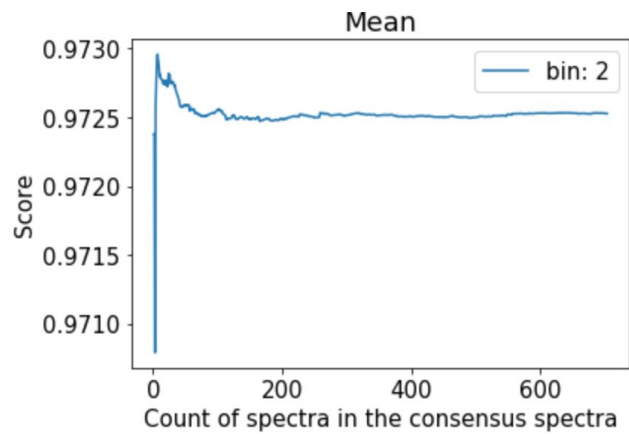
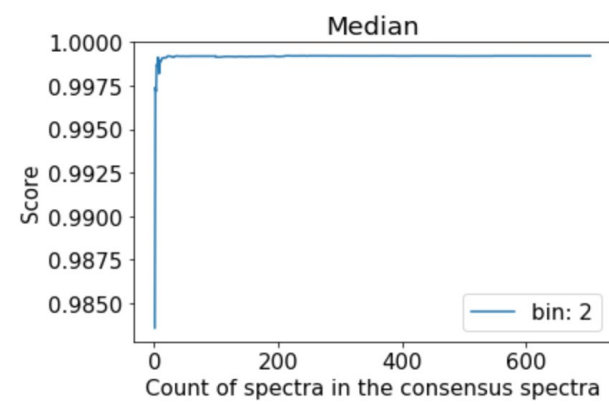


Standard Deviation



```
1 spectra_out.plot_metrics('34,178')
```





```
1 spectra_out.plot_metrics('2')
```

