

Quantifying Common Word Variance Through Rainbow Triangle Graph Decomposition of the Common Word 'the'

Alexandra Arabio, Alicia Carriquiry, Danica Ommen

Project Background

Writing Exemplars: Handwritten samples were collected by CSAFE and added to an open-source repository of over 250 writers ranging in age from 18-60 years old. We use the common phrase prompt to examine the common word 'the'. Participants provided nine writing samples in three sessions. The word 'the' appears three times in the prompt, for a total of 27 occurrences per writer.

Decomposition Process: Our goal is to construct numerical variables that allow us to quantify the shape of a word. Our approach originated from a graph theory question of the decomposition of Kneser Graphs at set size fifteen. We first identify landmarks in the word to create a complete graph of the word. These landmarks could be the beginning and the end pixels in the word, the highest and lowest pixels, and the pixels at the intersection of two strokes. We place a node at each landmark and color-code the nodes by landmark type.

Triangulation Process: Rainbow triangles are formed using the same principles as the Kneser Graph decomposition process. To Construct a triangle, we draw edges between three nodes of differing colors, and where edges are unique in a single triangle.

Methodology

The R package 'handwriter' is used to place the nodes at their indicating landmarks in the word sample.

Node Placement: The common word 'the' (represented by the 6th, 9th, and 13th words within the common phrase prompt) are cropped from the phrase for individual Analysis. Nodes are placed on the individual words to identify specific landmark characteristics, as seen in Figure 1.

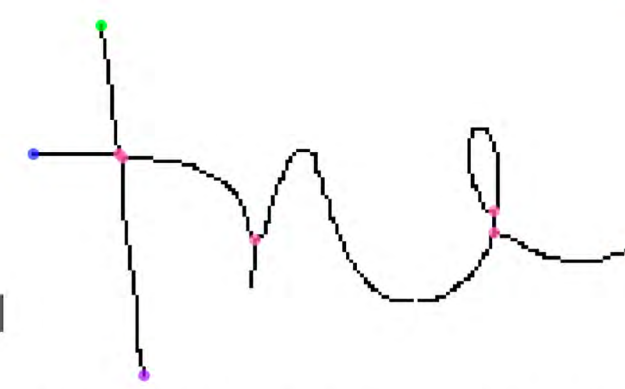


Figure 1. Example of writer 1's 6th common word with nodes.

Node Color Placement Code:

- Blue = left most pixel
- Orange = right most pixel
- Purple = lowest most pixel
- Green = highest most pixel
- Pink = 2+ intersecting or overlapping pixels

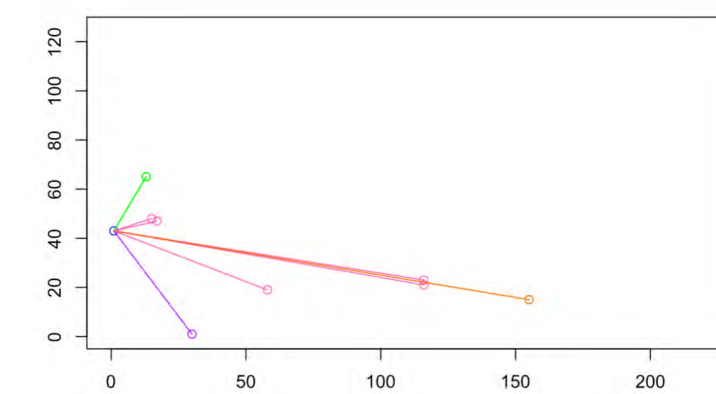
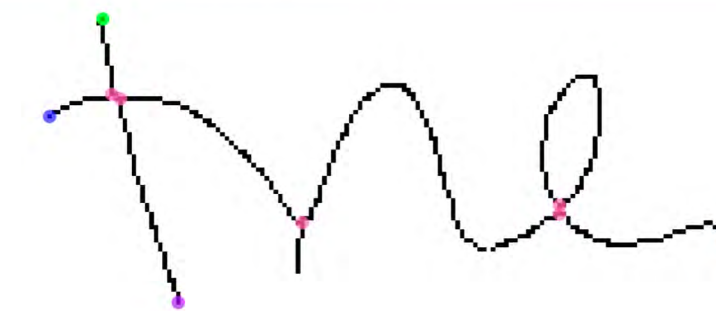
Decomposition and Triangulation: After the nodes are placed on the exemplar, the Blue Node is decomposed to form connecting line segments to each subsequent node within the sample. This is demonstrated in Data & Results Row 2. Next, the triangulation process begins by locating the Pink Node furthest to the left of the sample and forming a connecting edge to the Green Node. This same process is then continued to connect the second Pink Node from the left to the Purple Node and once more for the third Pink Node from the left to the Orange Node. This is demonstrated in Data & Results Row 3.

Quantitative Variables: Once the triangulation process is complete, we measure the area, perimeter, enclosing edge lengths, and angles from each triangle in the word. These variables are then compiled in a data frame for statistical analysis.

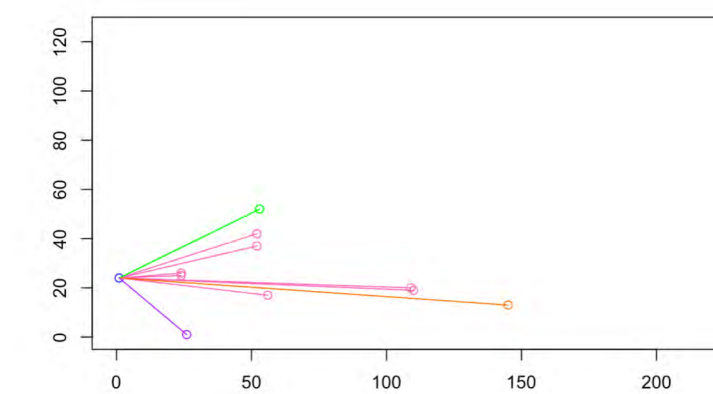
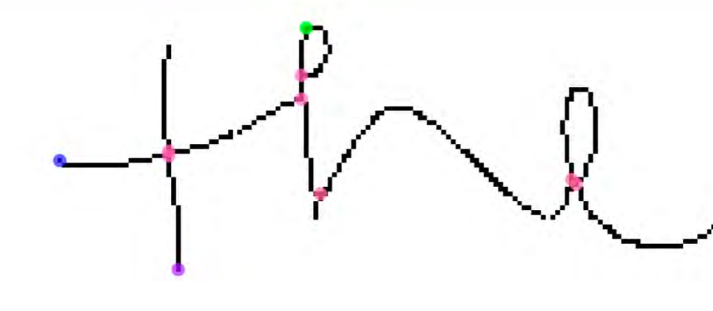
Discussion

Summarization: Through the rainbow triangulation process we have been able to obtain possible variation thresholds for each writer. Using these thresholds, we can form a better understanding of the likelihood of an unknown exemplar being written by a certain known writer. We aim to be able to use this information to help exclude possible writers from being the writer in a closed sample set scenario. Aside from the quantifying measurements this method provides, it also provides a clear visual representation of the variations for interpretation. As we continue this project, we hope to be able to form a likelihood ratio for writership and perform more statistical analysis on the data. In the future we also hope to begin analyzing signature samples and compare results with current analysis in the field.

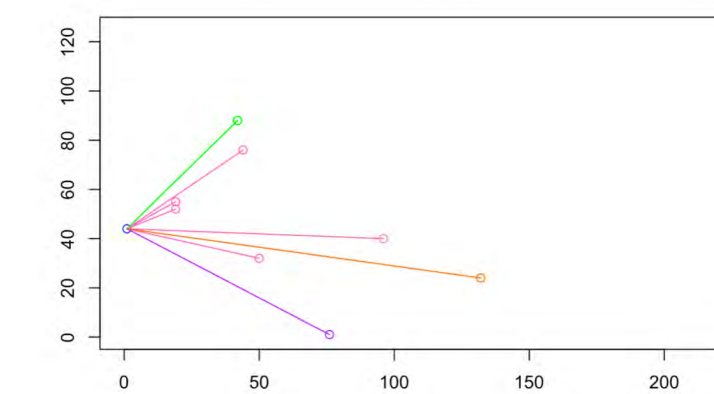
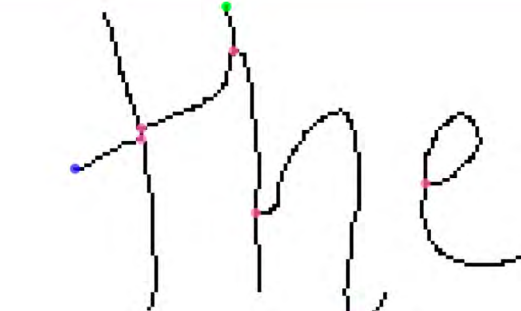
Data & Results



Min.	1 st Q	Median	Mean	3 rd Q	Max.
0.00	16.49	51.04	62.18	116.73	156.52



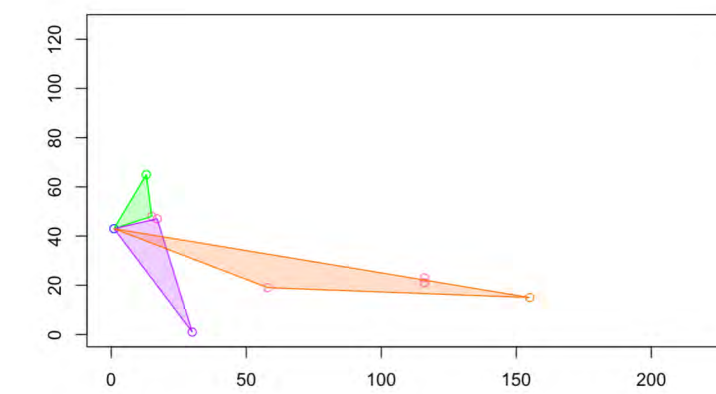
Min.	1 st Q	Median	Mean	3 rd Q	Max.
0.00	28.53	54.08	60.26	83.57	144.42



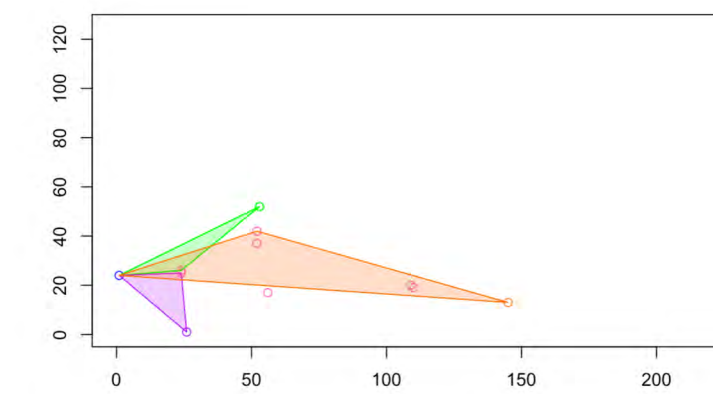
Min.	1 st Q	Median	Mean	3 rd Q	Max.
0.00	21.10	53.60	57.67	86.45	132.52

Row 1: Singular writing exemplars taken of 'the' from writer 1, 3, and 6 respectively from left to right

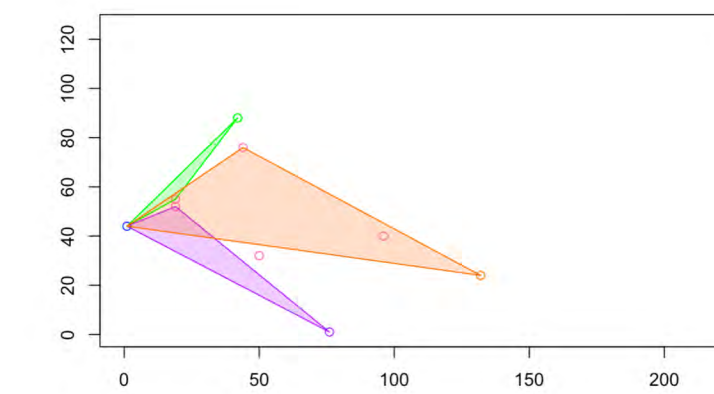
Row 2: Decomposition results from the Row 1 exemplars above with summary statistics of the edge lengths in the table below



Green Length	Purple Length	Orange Length	Green Area	Purple Area	Orange Area
17.12	47.8	97.08	124	394	1050

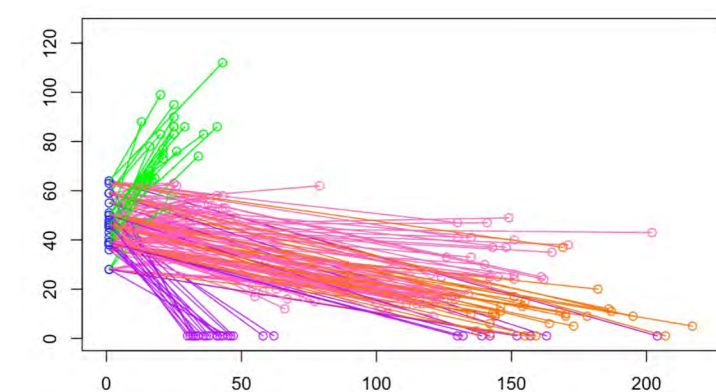


Green Length	Purple Length	Orange Length	Green Area	Purple Area	Orange Area
38.95	24.08	97.42	270	277	1576

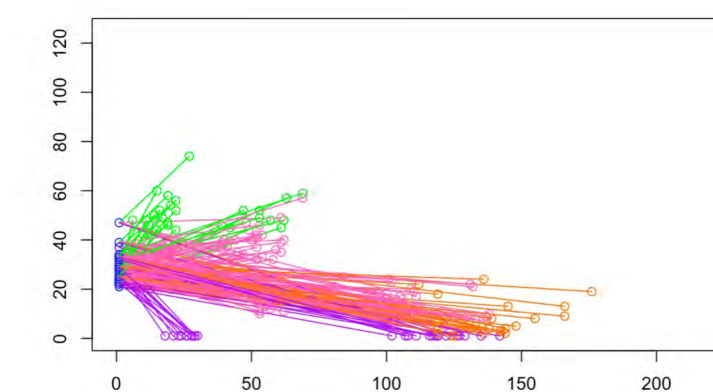


Green Length	Purple Length	Orange Length	Green Area	Purple Area	Orange Area
40.22	76.49	102.20	170	687	2526

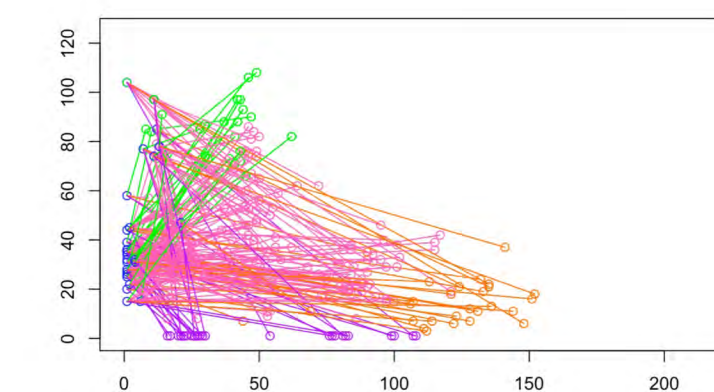
Row 3: Triangulation results from the Row 2 decompositions with quantitative measures of length of enclosing edge used to form each triangle and the area of each triangle in the table below



Min.	1 st Q	Median	Mean	3 rd Q	Max.
0.00	29.07	57.38	76.92	129.91	219.67

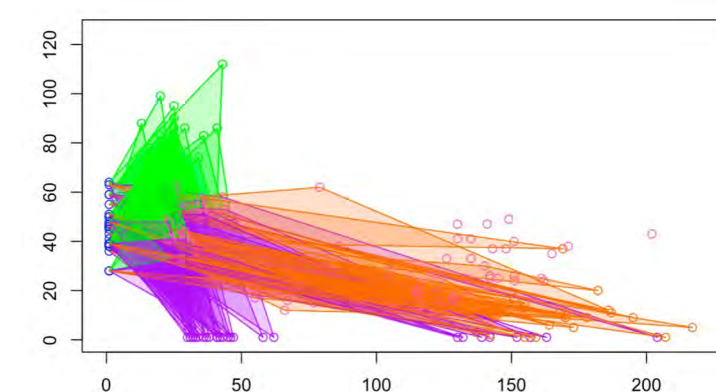


Min.	1 st Q	Median	Mean	3 rd Q	Max.
0.00	22.01	53.29	63.46	108.85	175.41

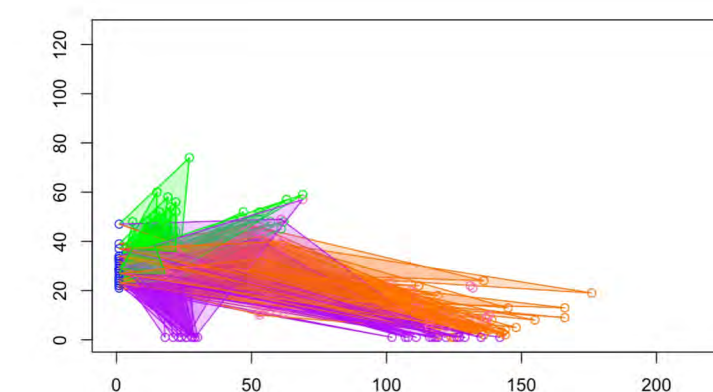


Min.	1 st Q	Median	Mean	3 rd Q	Max.
0.00	24.19	51.52	56.26	83.76	176.67

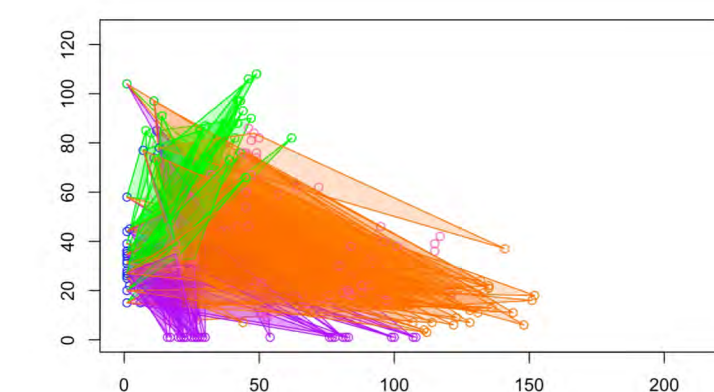
Row 4: Decomposition results from all 27 exemplars taken from writer 1, 3, and 6 above with summary statistics of the compiled edge lengths in the table below



	Min	1 st Q	Median	Mean	3 rd Q	Max.
Length	11.40	22.02	28.64	29.60	36.51	54.04
Length	32.28	43.95	50.01	69.18	101.04	170.79
Length	23.77	47.99	89.36	78.50	100.16	120.35
Area	124	266	416	430	548	1086
Area	394	549	714	739	869	1619
Area	85	396	602	741	939	2324



	Min	1 st Q	Median	Mean	3 rd Q	Max.
Length	8.00	20.25	23.35	28.10	35.69	51.86
Length	21.5	27.65	76.32	69.72	105.36	120.50
Length	62.0	84.12	90.05	89.16	95.52	116.73
Area	72	140	189	203	242	427
Area	18	191	256	391	363	1498
Area	22	769	1095	984	1342	1576



	Min	1 st Q	Median	Mean	3 rd Q	Max.
Length	5.385	40.4	46.615	48.83	56.87	82.88
Length	22.20	30.6	43.10	51.36	72.78	90.87
Length	33.84	76.54	91.00	87.51	99.88	133.03
Area	0.0	149	308	318	473	905
Area	25	269	334	394	446	1829
Area	156	896	1590	1781	2525	4884

Row 5: Triangulation results from all 27 exemplars taken from writer 1, 3, and 6 above with summary statistics of the quantitative measures of enclosing edge lengths used to form each triangle and the area of each triangle in the table below

All segment length measurements are displayed by pixel count and all triangle area measurements are displayed by pixel count squared within the data & results of this project.