Vision Performance Institute

Technical Report

Letter Structure and Legibility

John R. Hayes, Yu-Chi Tai, David Glabe, Vinsunt Donato, James Sheedy

Purpose
To evaluate the legibility of a letter through examination of the physical structure of the letter. The problem was approached by evaluating individual attributes of a letter. Also to test whether images that contain most of their information in a few underlying latent structures are more legible than more complex letters requiring more latent structures to recognize.
Method
30 subjects performed a distance threshold legibility task of black and white letters to establish subjective relative legibility as experienced by the visual system. The subjects identified the center target letter (a,c,e,m,n,o,r,s,v,w) of 3 letters presented. The fonts were Baskerville, Bodoni, Centaur, Consolas, DIN, Futura, Garamon, Georgia, Helvetica, Rockwell, and Verdana. Two objective evaluations of character traits were also determined. Specific attributes were measured for each character, and the set of attributes varied for each letter. For example, the letter “a” attributes were height of letter, width of letter, max width of main stroke, min width of main stroke, serifs, opening size, max bowl width, and max bowl height. Regression analysis for each letter across fonts determined the salient characteristics for legibility. In this method there are only a few characteristics that are common across letters. An alternative objective representation of structure was created by decomposing the original letter image into an ordered set of linear components (singular value decomposition SVD). Stepwise regression analyses on summary characteristics of eigenvalues predicted legibility. Measures included the first eigenvalue, sum of first 2, 5, 10, or 20 eigenvalues, and the slope of the first 5 or 10 eigenvalues. Pixel density was also included in this model. SVD summary measures were common across all letters and fonts.
Results
The stepwise regression for letter attributes ranged from R^2=.66 (“a”) to R^2=.81 (“w”). Removing the between subjects effects from the model revealed a purer estimate of the letter attributes yielding a range of attribute

contribution from R^2=.40 (“n”) to R^2=.68 (“s”). Ability of the SVD eigenvalues to predict relative legibility of unknown fonts was tested with a jackknife procedure in which 11 stepwise regression models excluded a different font each time. The models were then used to predict the relative legibility of the letters within excluded fonts. The R^2 values for the models ranged from .46 to .52. The regression of the predicted letter legibility on the actual legibility of the excluded fonts was significant at R^2=.42. The first eigenvalue, sum of the first 5 eigenvalues and the squared density were the most common significant regression components.

Conclusion

Both individual letter attributes and the decomposition of the graphical image of the letter can provide useful information to the font designer for developing legible fonts.

A simple perusal of Microsoft Word's font selection illustrates the vast variety of available fonts. Most font designs have evolved from a font designer's artistic sensibilities or their general impressions of readability. The actual design and instant rendering of a font on a computer display is surprisingly complex. Given the broad range of font styles our objective was to determine if there are individual letter characteristics that facilitate letter legibility. In this study we subjectively measured the relative legibility of a set of characters across a set of fonts, and also used two approaches to objectively evaluate letter characteristics. The subjective relative legibility measures are used to validate the objective evaluations.

The first objective approach is to evaluate a letter by tabulating and measuring numerous character attributes. Some letters have more attributes than others. Only a few attributes are common across different letters, e.g maximum and minimum heights and widths of the character. Some are common among subsets of letters, e.g. size of opening (c, e). Some are unique, e.g. cap opening height (e). While most letters have generally the same shape across fonts, some offer challenges even in this respect, e.g. a, a.

The second objective approach was to take the image of a font, where each pixel is represented by an entry in a matrix, and express the matrix as a series of eigen-fonts (matrices) and eigenvalues (numbers). Singular Value Decomposition (SVD) is a method to reduce the storage requirements of pictures[1]. It utilizes Principle Component Analysis to extract visual features, allowing the feature contributing the greatest to the overall picture to be extracted first and assigned the highest eigenvalue. Typically, the first few components are able to provide a reasonable approximation of the picture. The more vectors stored the better the reproduction, but the later components usually provide little detectable information. In utilizing this method, we treated each letter in a specific font as a pictorial object and hypothesized that the more information contained in the first few components of a letter, the more easily the visual system would extract the visual feature of that letter. Therefore, the fewer components required to achieve satisfactory reproduction of a letter, the better the legibility. SVD is based on a theorem from linear algebra which says that a rectangular matrix A can be broken down into the product of three matrices - an orthogonal matrix U, a diagonal matrix S, and the transposition of an orthogonal matrix V .

$$(1) \quad \mathbf{A} = \mathbf{USV}^T = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathbf{T}} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathbf{T}} + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^{\mathbf{T}} + \ldots$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}$; $\mathbf{V}^T\mathbf{V} = \mathbf{I}$; the columns of $\mathbf{U}$ are orthonormal eigenvectors of $\mathbf{AA}^T$, the columns of $\mathbf{V}$ are orthonormal eigenvectors of $\mathbf{A}^T A$, and $\mathbf{S}$ is a diagonal matrix containing the square roots of eigenvalues, σ_1, σ_2 …. , from $\mathbf{U}$ or $\mathbf{V}$ in descending order. The matrix $\mathbf{U}$ has dimensions m x m, and the matrix $\mathbf{V}^{\mathbf{T}}$ has dimensions of m x n. The quantity $\mathbf{u}_i\mathbf{v}_i^{\mathbf{T}}$ represents the i^{th} eigen-font. Computationally, a larger i contains more information about the letter because there are more factors summed to more closely approximates $\mathbf{A}$.

Method

To establish the relative legibility as judged by the visual system, 30 subjects performed a distance threshold legibility task of black and white letters. The subject identified the center target letter (a,c,e,m,n,o,r,s,v,w) of 3 letters presented. The fonts were Baskerville, Bodoni, Centaur, Consolas, DIN, Frutiger, Futura, Garamon, Georgia, Helvetica, Rockwell, and Verdana. Specific character attributes were measured for each character, and the set of attributes varied for each letter. For example, the letter “a” attributes were height of letter, width of letter, max width of main stroke, min width of main stroke, serifs, opening size, max bowl width, and max bowl height. A mixed model backward stepwise regression analysis for each letter across fonts determined the salient characteristics for legibility. In this method there are only a few characteristics that are common across letters. An alternative representation of structure was created by decomposing the original letter image into an ordered set of linear components (singular value decomposition SVD). Stepwise regression analyses on summary characteristics of eigenvalues predicted legibility. Measures included the first eigenvalue, sum of first 2, 5, 10, or 20 eigenvalues, and the slope of the first 5 or 10 eigenvalues. In addition, the density of pixel information is also computed for each letter/font combination[2]. The density measurement was for a single letter without regard to spacing between letters. The more black pixels in the hardcopy letter, the higher the density. SVD summary measures were common across all letters and fonts. Singular Value Decompositions were derived using a MatLab program (program code in Appendix B). Statistical analyses were carried out with SPSS (version 17).

Results

First we report singular value decomposition (SVD) as a method designed to quickly evaluate the legibility of text. Second we review the results of the specific attributes of a font in terms of features of the physical shape.

Singular Value Decomposition

Figure 1a illustrates the decreasing amount of information contained in successive components comprising the letter. The possible number of eigenvalues greater than zero can be over 100. Figure 1b demonstrates that most of the variance in eigenvalues occurred within the first 3 to 5 components. Samples of reproduced letters are provided in Figure 2.

Stepwise regression models were employed to test the ability of SVD eigenvalues to predict the measured relative legibility. Table 1 presents the inter-correlations between relative legibility and the other variables. All SVD and density variables were significantly correlated with observed relative legibility (RL). Table 2 contains summary values for the variables.

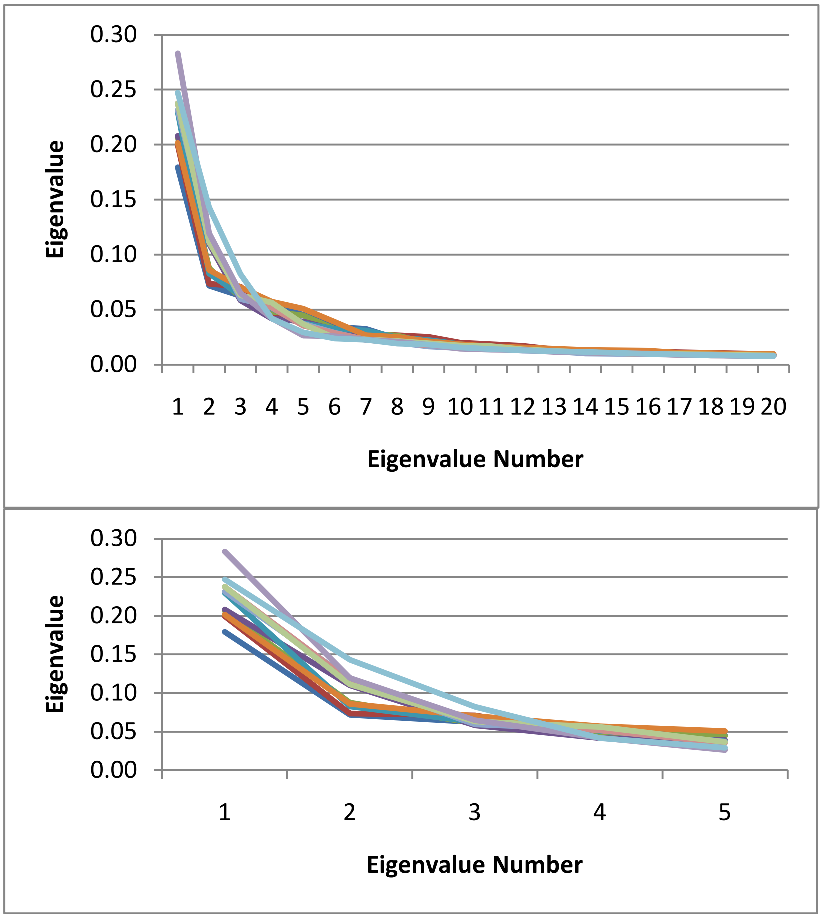


Figure 1a, b. The singular value decomposition of the letter "a" for 11 fonts. 1a Eigenvalues up to 20. 1b. Same as 1a except only first 5 eigenvalues.

Figure 2. Step-wise summation of the first five cumulative eigenvectors.

Table 1. Intercorrelation matrix for Relative Legibility (RL), SVD statistics and density. E1 was the first eigenvalue. Sum(x) was the sum of the first x eigenvalues. Slope(x) was the slope of the first x eigenvalues. Recip(x) was the slope of the first x eigenvalues reciprocals (1/eigenvalue).

	RL	E1	Sum2	Sum5	Sum10	Sum20	Slope5	Slope10	Recip5	Recip10	Density	Density2
RL	1	.338**	.395**	.375**	.331**	.281**	-.387**	-.399**	.447**	.356**	.664**	.672**
E1	.338**	1	.977**	.960**	.957**	.942**	-.986**	-.973**	.555**	.731**	.342**	.333**
Sum2	.395**	.977**	1	.982**	.965**	.938**	-.995**	-.996**	.655**	.774**	.384**	.373**
Sum5	.375**	.960**	.982**	1	.991**	.969**	-.968**	-.992**	.580**	.761**	.313**	.311**
Sum10	.331**	.957**	.965**	.991**	1	.991**	-.949**	-.971**	.492**	.708**	.238*	.248**
Sum20	.281**	.942**	.938**	.969**	.991**	1	-.923**	-.942**	.419**	.647**	.156	.177
Slope5	-.387**	-.986**	-.995**	-.968**	-.949**	-.923**	1	.990**	-.666**	-.770**	-.402**	-.386**
Slope10	-.399**	-.973**	-.996**	-.992**	-.971**	-.942**	.990**	1	-.649**	-.788**	-.381**	-.369**
Recip5	.447**	.555**	.655**	.580**	.492**	.419**	-.666**	-.649**	1	.605**	.583**	.537**
Recip10	.356**	.731**	.774**	.761**	.708**	.647**	-.770**	-.788**	.605**	1	.447**	.425**
Density	.664**	.342**	.384**	.313**	.238*	.156	-.402**	-.381**	.583**	.447**	1	.975**
$Density^2$	.672**	.333**	.373**	.311**	.248**	.177	-.386**	-.369**	.537**	.425**	.975**	1

** Pearson correlation p<.01; * p<.05

Table 2. Summary Statistics for the various SVD/Density measures

	Minimum	Maximum	Mean	Std. Deviation

E1	.14	.41	.24	.08
Sum2	.23	.58	.37	.09
Sum5	.40	.75	.53	.09
Sum10	.53	.84	.64	.08
Sum20	.63	.92	.74	.08
Slope5	-.09	-.02	-.05	.02
Slope10	-.03	-.01	-.02	.01
Recip5	3.46	8.84	6.16	1.35
Recip10	4.04	12.59	6.81	1.90
Density	.05	.29	.14	.06

A stepwise regression approach of mean RL on the above SVD and density variables across all font and letter combinations (n=110) revealed a model that included the first eigenvalue (beta = -.562, p=.02), sum of the first five eigenvalues (beta = .716, p = .004) and square of density (beta = 7.1, p<.001). This model accounted for a significant proportion of the variance of the relative legibility for the 30 subjects (R^2 = .51, p<.001). Figure 2 plots the observed relative legibility on the model predicted legibility values.

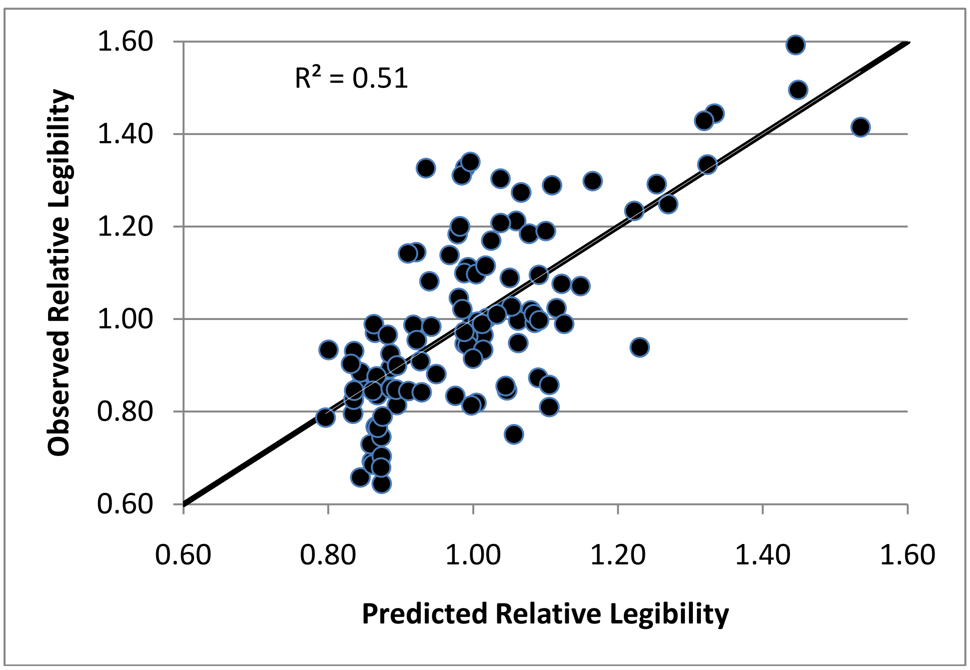


Figure 3. Observed relative legibility relative to the predicted legibility derived from the all the data (n=110 letter by font combinations). The line illustrates the 1:1 correspondence between the two measures.

A project objective was to be able to use this method to estimate the unknown relative legibility of new fonts or fonts in new situations. To this end we repeated our stepwise regression 11 times. (Note: Frutiger was inadvertently left out of the SVD analysis). Each time one font was not included and its relative legibility was predicted based on the regression from the other fonts. A summary of the models is presented in Table 3. The

square of density and the sum of the first five eigenvalues were the most frequently selected. The first eigenvalue was usually in the model. The jackknife is illustrated in Figure 4. There was still a highly significant portion of the variance accounted for by the models (R^2 = .42). The individual font/letter predicted values are in Appendix A.

The regression prediction model for all fonts was:
Relative legibility = .371 + 7.112 $Density^2$ + .1.546 Sum5 + -1.438 Eigenvalue1.
By convention, any time a squared variable is entered in a regression equation, the corresponding linear variable is included. For this sample the model is:
Relative legibility = .301 + .867 Density + 4.434 $Density^2$ + 1.597 Sum5 – Eigenvalue 1 1.509
(R^2 = .49)

Table 3. Individual stepwise regression models excluding a specific font for each run.

Excluded Font	Variables	R^2
Baskerville	$Density^2$, Sum5, E1	.52
Bodoni	$Density^2$, Sum5, E1	.49
Centaur	$Density^2$, Sum5, E1	.50
Consolas	$Density^2$, Sum5	.53
DIN	$Density^2$, Sum5	.48
Futura	$Density^2$, Sum5, E1	.54
Garamond	$Density^2$, Sum5, E1	.50
Georgia	$Density^2$, Sum5, E1	.50
Helvetica	$Density^2$, Sum5, E1	.50
Rockwell	$Density^2$, Sum5, E1	.50
Verdana	$Density^2$, Sum20	.46
All Fonts	$Density^2$, Sum5, E1	.51

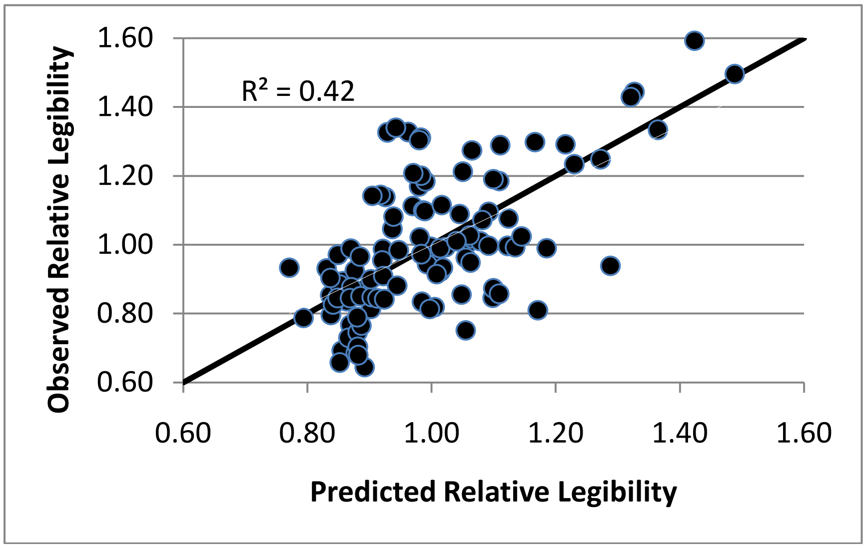


Figure 4.Observed relative legibility of the excluded font relative to the predicted legibility derived from the other 10 fonts (n=100 letter by font combinations). The line illustrates the 1:1 correspondence between the two measures.

Figures 5 and 6 demonstrate the ability of the regression prediction model to discriminate average relative legibility across both fonts and letters.

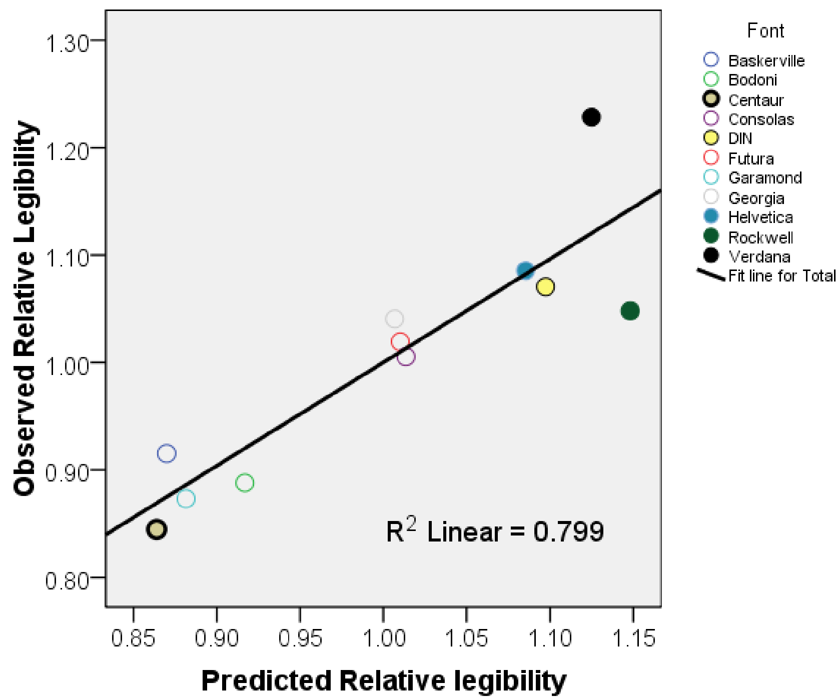


Figure 5. Square of density, sum of the first 5 eigenvalues, and the first eigenvalue were used to predict observed relative legibility for fonts according to the regression model.

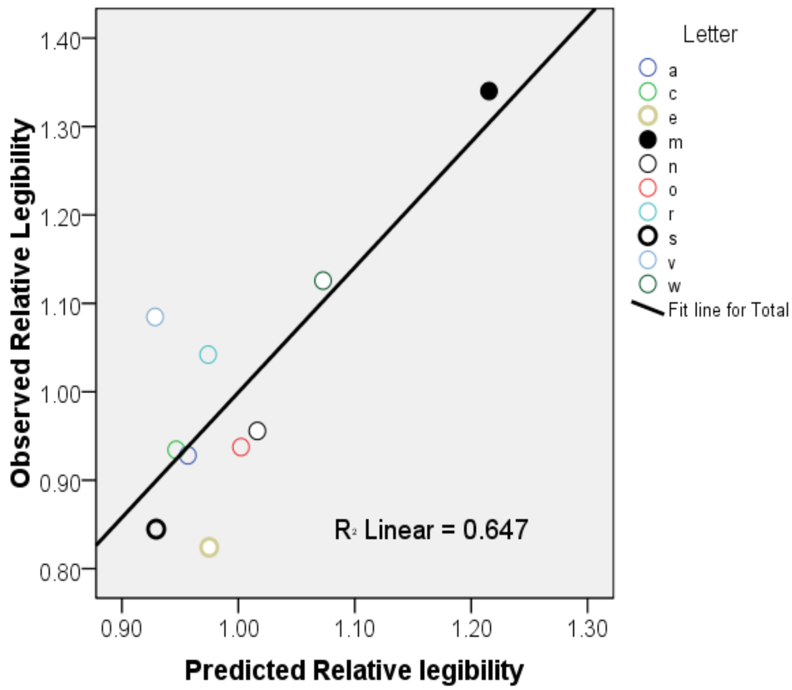


Figure 6. Square of density, sum of the first 5 eigenvalues, and the first eigenvalue were used to predict observed relative legibility for letters according to the regression model.

Conclusions regarding SVD

Our hypothesis was that the more information contained in the early eigenvalues, the better the legibility of the letter. The correlations (table 1) show that both sum of the first five eigenvalues and the first eigenvalue were positively correlated with relative legibility. The more total variance accounted for in the first five eigenvalues, the better the legibility. In the regression model the first eigenvalue had a negative coefficient. This resulted in the effect of large initial eigenvalue moderating the effect of the sum of the first five, but the principle was still the same. A font that has a simpler structure with most of the information contained in the first few eigenvalues is more legible. Density of the letter was a key component of predicting relative legibility. The more black pixels in the hardcopy letter, the more legible it was. This was the case in the context of the fonts that were used in the study and may or may not generalize to other characteristics of fonts, e.g. bolding the font.

We have demonstrated that a simple decomposition of letter shape for a font with the additional information of pixel density can predict relative legibility. This suggests utility as a post-design measure of relative legibility.

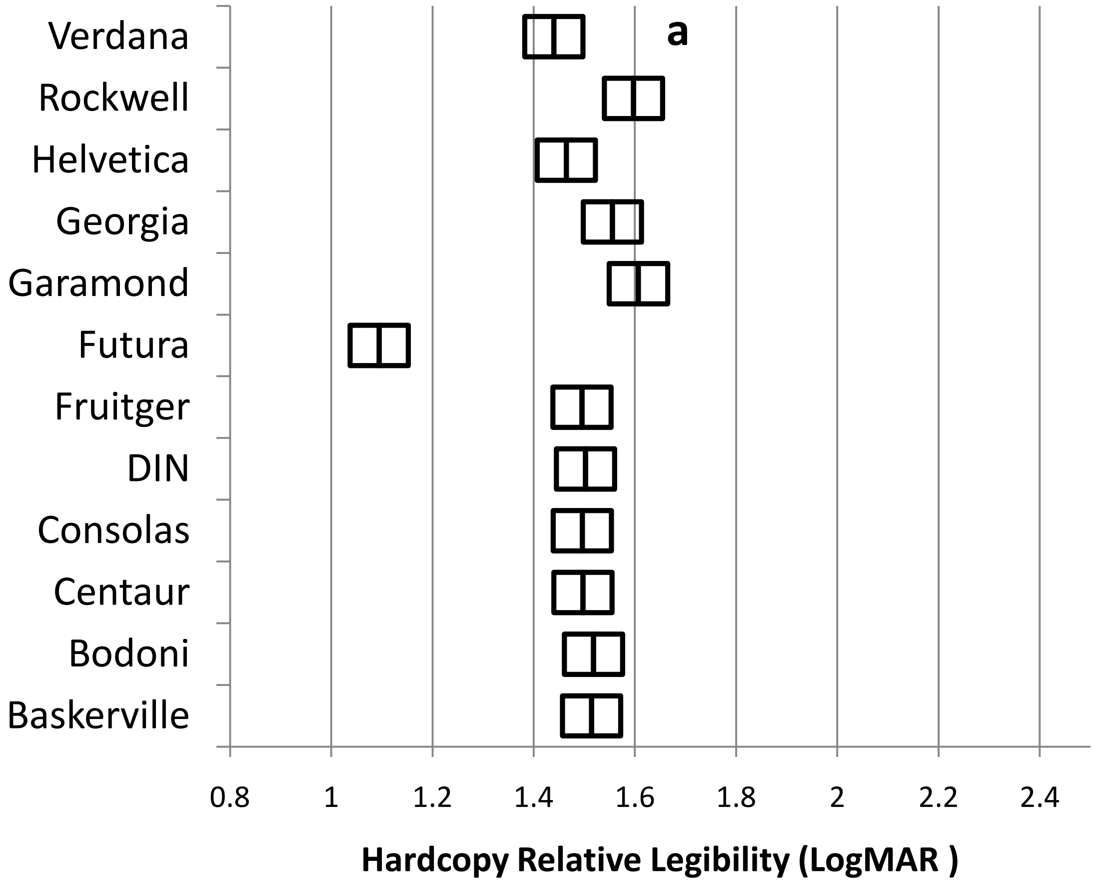
Letter Attributes

In the previous section we were able to use the same predictive variables across all fonts and letters. The challenge in this section was that letters are partially comprised of unique features. Letter attributes must be analyzed one letter at a time. We only have 12 different fonts to test each letter. Further, the relationship between legibility and the attribute is not necessarily linear. This creates many variables with few observations.

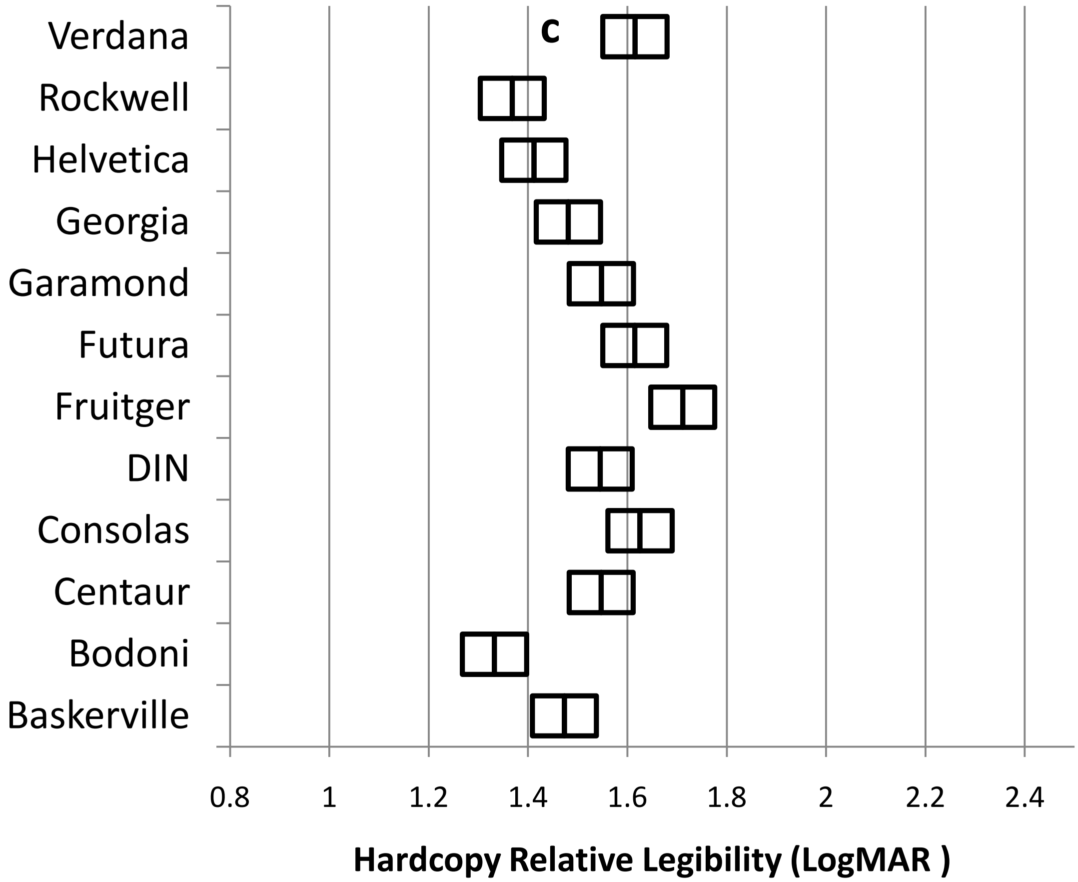
Individual attribute analysis

Individual letter analysis.

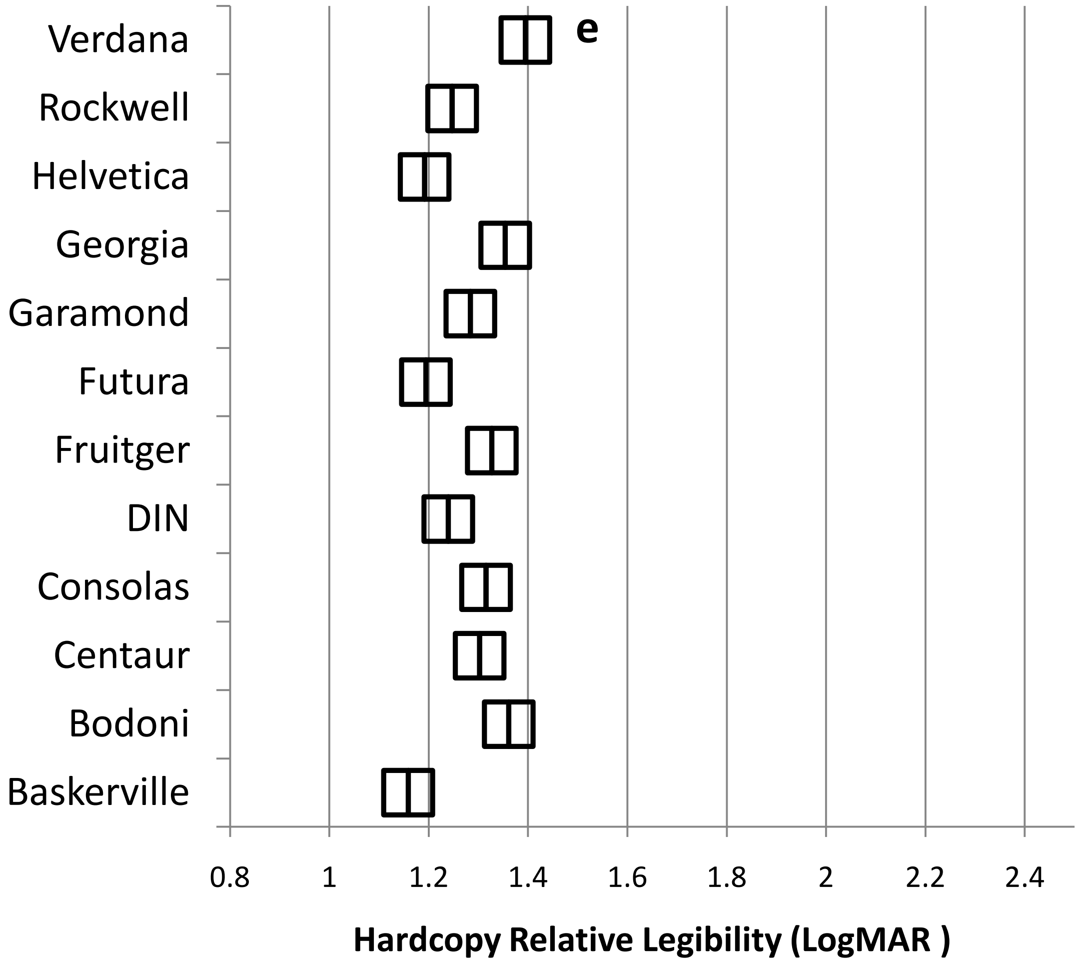
One way ANOVA with repeated measures across 30 subjects provided an estimate for confidence intervals. Non-overlapping 84% confidence intervals identify fonts that are significantly different from each other at an unadjusted $p<.05$[3]. Using the mixed model analysis of variance framework a backward stepwise regression was manually performed to identify the characteristics that accounted for the greatest amount of independent variance in relative legibility measured in LogMAR. Of the 38 individual letter characteristics distributed across all 10 letters, only a few were common in all the letters. All the characteristics were included in the model and then the least significant ($p>.1$) were eliminated one by one until only significant variables were remaining. Height and width of the font was included in all models to control for size. The data in the figures are the least square mean legibility from the final model. There is quite a bit of legibility variability between the fonts that is described by the letter characteristics.



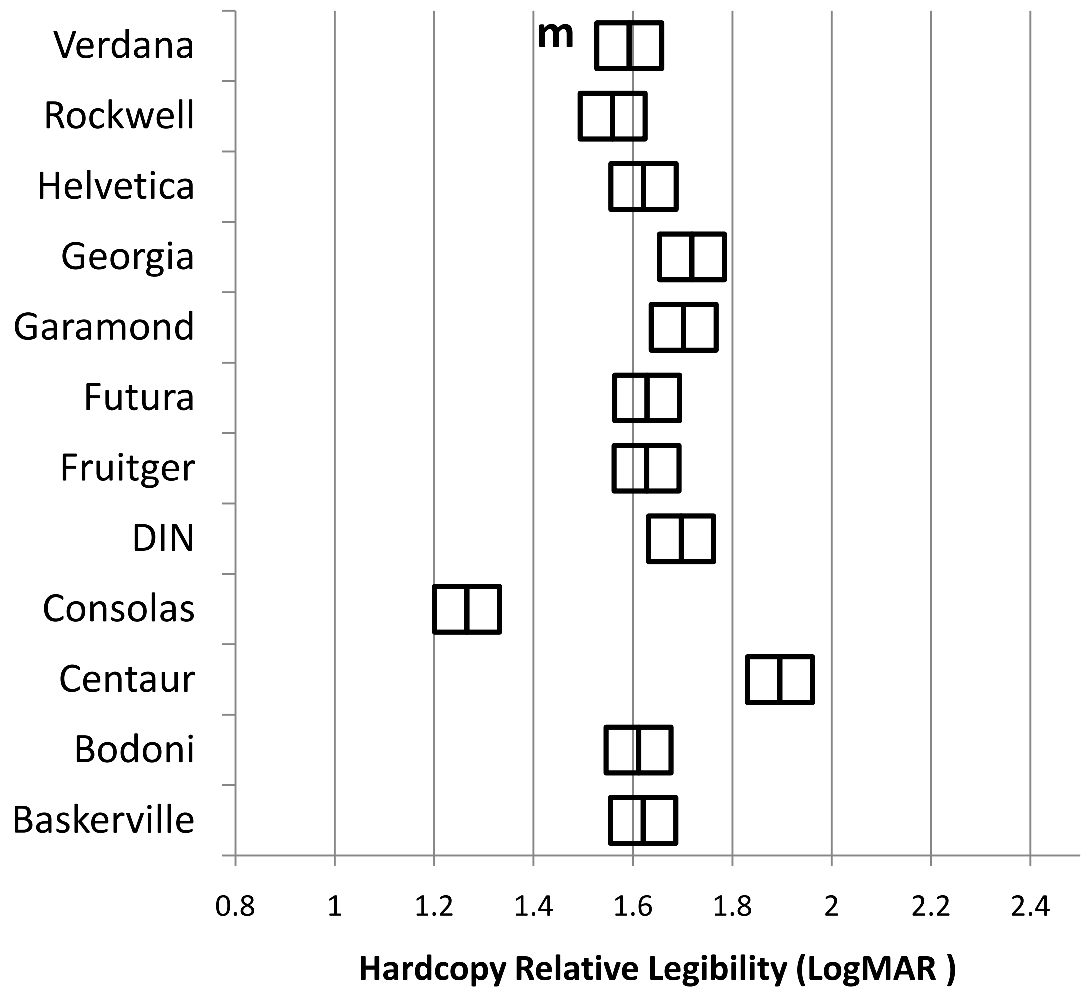
a	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.225	0.062	223.439	19.827	0.000	1.103	1.347
Max Height of Letter	0.000	0.000	326.000	1.454	0.147	0.000	0.001
Max Width of Letter	0.001	0.000	326.000	3.995	0.000	0.000	0.001
MS minimum width	-0.002	0.000	326.000	-4.409	0.000	-0.002	-0.001
max bowl height	-0.002	0.000	326.000	-8.631	0.000	-0.002	-0.001



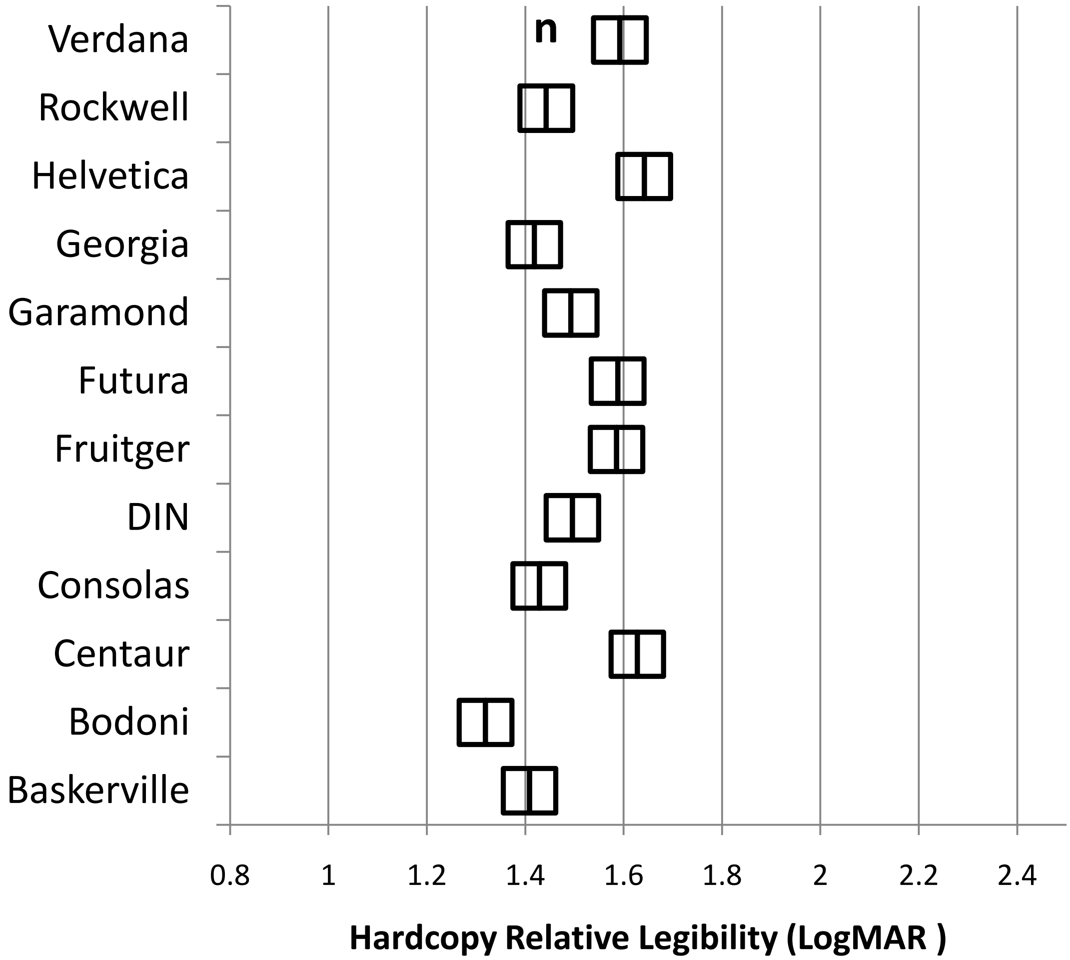
C	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.559	0.066	159.258	23.740	0.000	1.430	1.689
Max Height of Letter	0.000	0.000	326.000	0.471	0.638	-0.001	0.001
Max Width of Letter	-0.001	0.000	326.000	-1.506	0.133	-0.001	0.000
MS Width Ratio (Max/Min)	-0.018	0.006	326.000	-2.952	0.003	-0.030	-0.006
Opening size	0.001	0.000	326.000	3.856	0.000	0.000	0.001



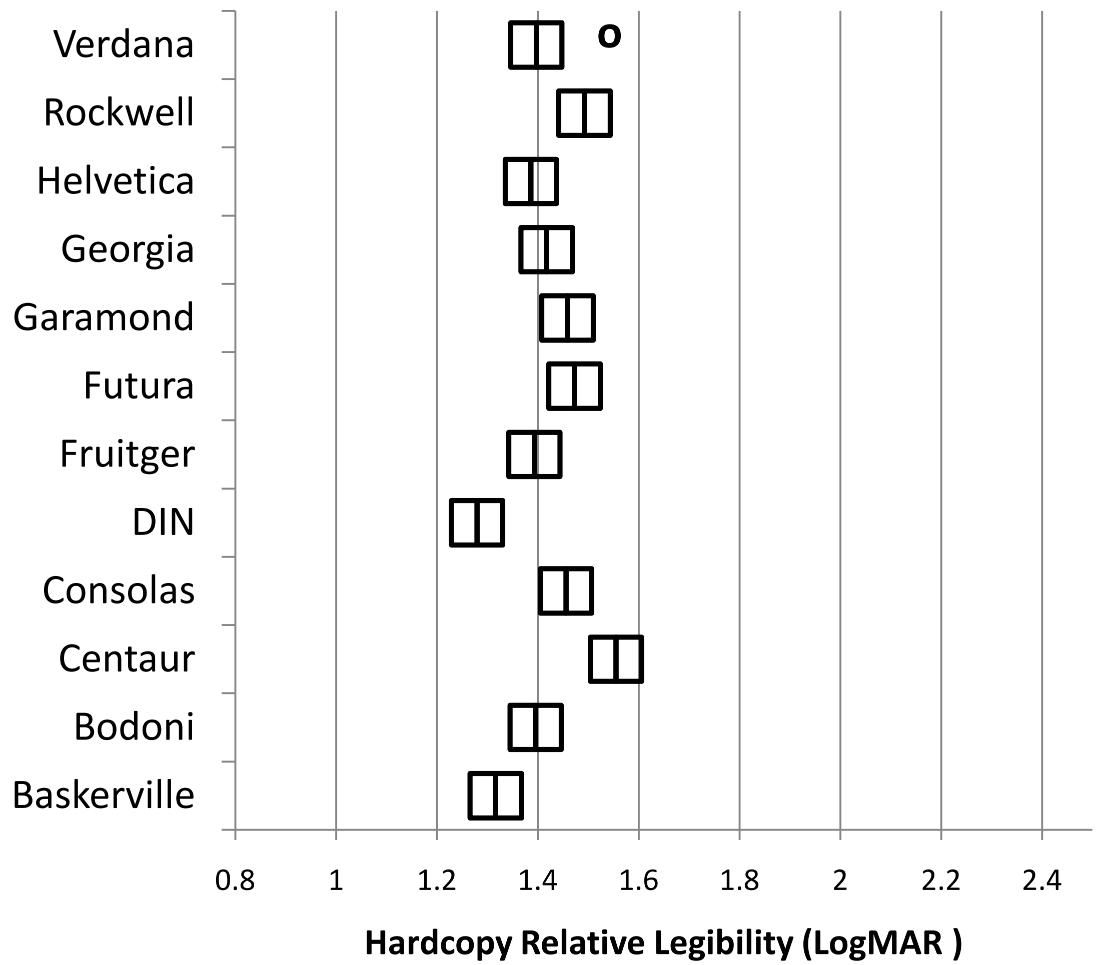
e	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	7.261	0.996	322.585	7.288	0.000	5.301	9.220
Max Height of Letter	0.001	0.001	322.000	0.563	0.574	-0.002	0.004
Max Width of Letter	-0.006	0.001	322.000	-6.450	0.000	-0.008	-0.004
MS minimum width	0.031	0.006	322.000	5.431	0.000	0.020	0.043
Opening size	0.004	0.001	322.000	5.442	0.000	0.003	0.006
Max vertical dim. of stroke	-0.011	0.002	322.000	-6.420	0.000	-0.015	-0.008
(Max vertical dimension of stroke)/(max width of stroke)	-1.195	0.192	322.000	-6.212	0.000	-1.573	-0.816
Distance, bottom of letter to cross-stroke	0.009	0.002	322.000	4.549	0.000	0.005	0.013
Cap opening height	0.016	0.003	322.000	5.673	0.000	0.011	0.022



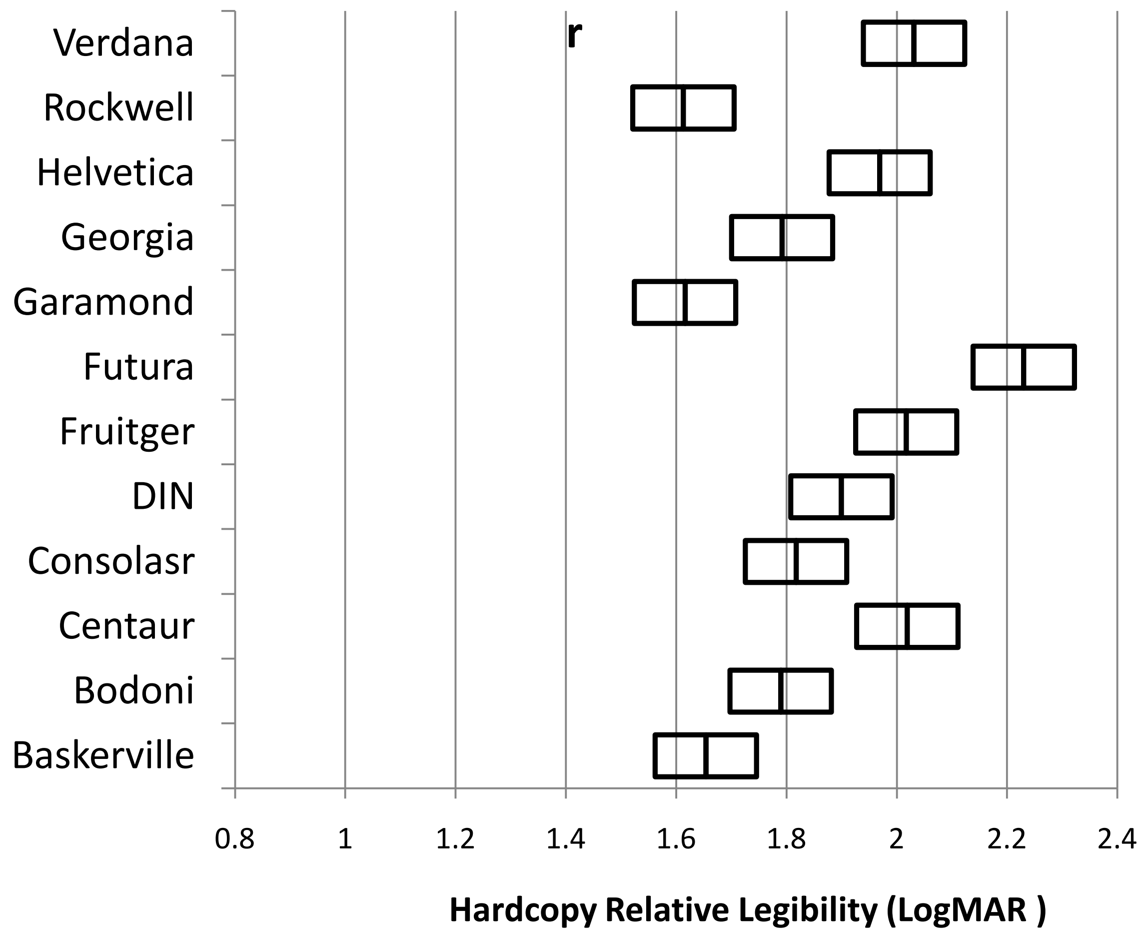
m	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.773	0.077	138.236	22.894	0.000	1.620	1.927
Max Height of Letter	-0.001	0.000	323.000	-4.081	0.000	-0.001	0.000
Max Width of Letter	-0.001	0.001	323.000	-0.952	0.342	-0.002	0.001
MS Max Width	0.005	0.002	323.000	2.814	0.005	0.001	0.008
MS minimum width	-0.006	0.002	323.000	-3.681	0.000	-0.009	-0.003
MS Width Ratio (Max/Min)	-0.055	0.016	323.000	-3.439	0.001	-0.086	-0.023
Opening size	0.003	0.001	323.000	2.421	0.016	0.001	0.005
Serif	0.719	0.395	323.000	1.820	0.070	-0.058	1.497



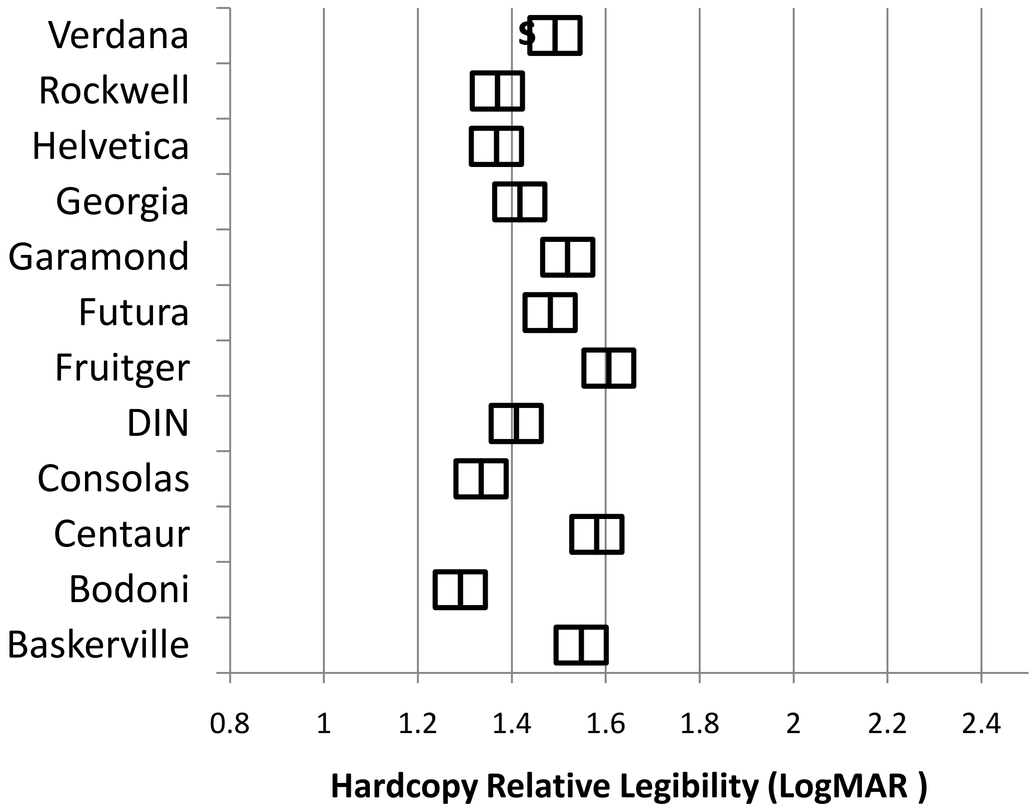
n	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.724	0.063	165.017	27.157	0.000	1.599	1.849
Max Height of Letter	0.000	0.000	323.000	1.126	0.261	0.000	0.001
Max Width of Letter	-0.002	0.000	323.000	-3.407	0.001	-0.002	-0.001
MS Max Width	0.002	0.001	323.000	2.433	0.016	0.000	0.004
MS minimum width	-0.004	0.001	323.000	-3.467	0.001	-0.006	-0.002
MS Width Ratio (Max/Min)	-0.063	0.019	323.000	-3.413	0.001	-0.100	-0.027
Opening size	0.002	0.000	323.000	5.081	0.000	0.001	0.003
Serif	0.832	0.202	323.000	4.117	0.000	0.434	1.229



O	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.478	0.059	102.498	25.255	0.000	1.362	1.594
Max Height of Letter	-0.001	0.000	324.000	-4.922	0.000	-0.001	-0.001
Max Width of Letter	-0.009	0.003	324.000	-2.608	0.010	-0.015	-0.002
MS Max Width	0.019	0.007	324.000	2.858	0.005	0.006	0.032
MS Width Ratio (Max/Min)	-0.038	0.008	324.000	-4.590	0.000	-0.055	-0.022
max bowl width	0.009	0.003	324.000	2.866	0.004	0.003	0.016
max bowl height	0.000	0.000	324.000	1.980	0.049	0.000	0.001

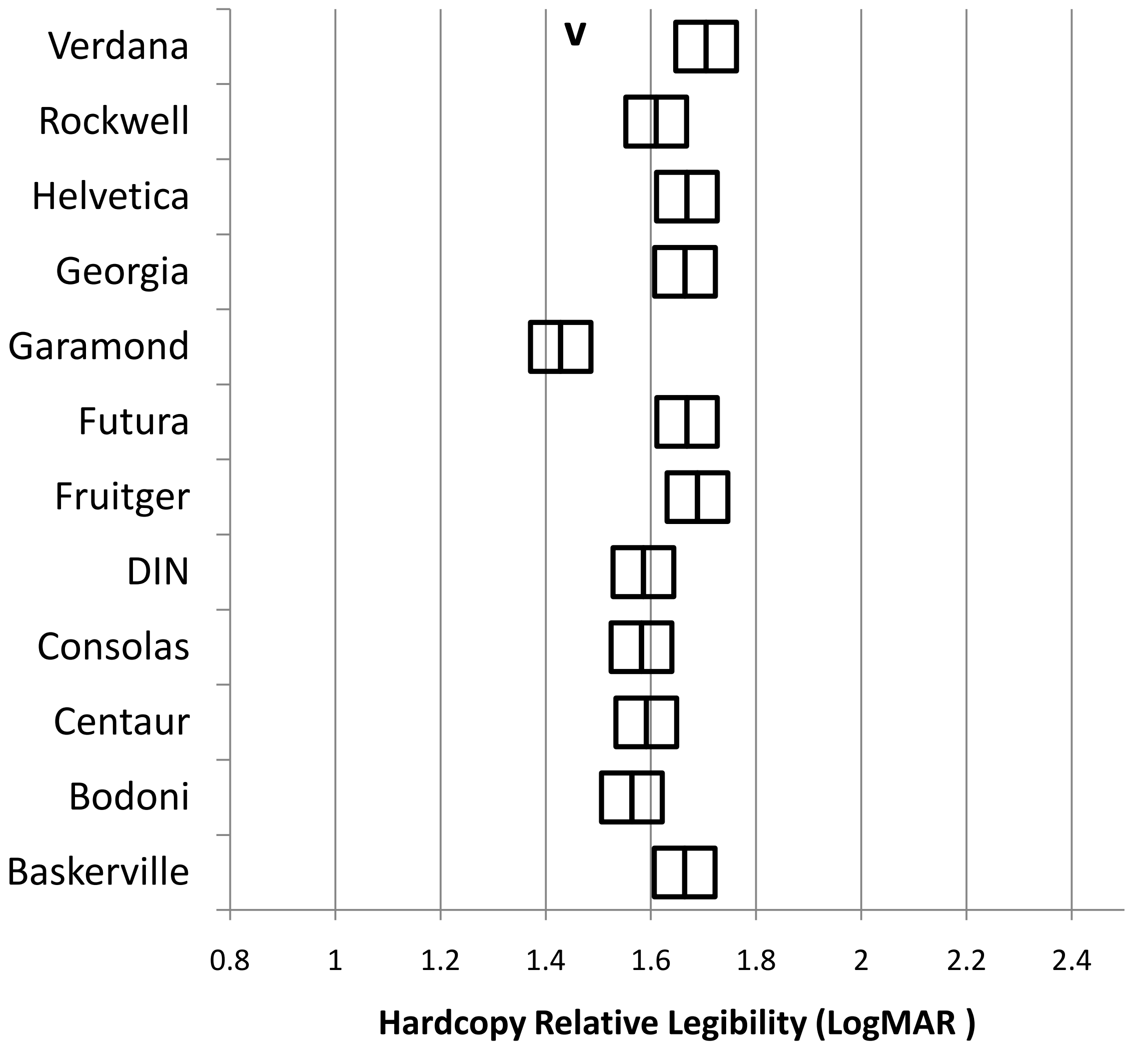


r	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	2.443	0.149	304.345	16.357	0.000	2.149	2.736
Max Height of Letter	0.000	0.000	321.000	1.062	0.289	0.000	0.001
Max Width of Letter	0.014	0.005	321.000	2.866	0.004	0.004	0.024
MS Max Width	0.163	0.025	321.000	6.591	0.000	0.114	0.212
MS minimum width	-0.170	0.026	321.000	-6.448	0.000	-0.222	-0.118
length of horizontal stroke	-0.015	0.005	321.000	-2.960	0.003	-0.024	-0.005
min width of horizontal stroke	-0.035	0.011	321.000	-3.150	0.002	-0.057	-0.013
Width ratio (max/min)	-0.210	0.065	321.000	-3.228	0.001	-0.339	-0.082
width of horizontal stroke at attachment to main stroke	0.023	0.009	321.000	2.557	0.011	0.005	0.041
Serif	-2.885	0.836	321.000	-3.453	0.001	-4.529	-1.241



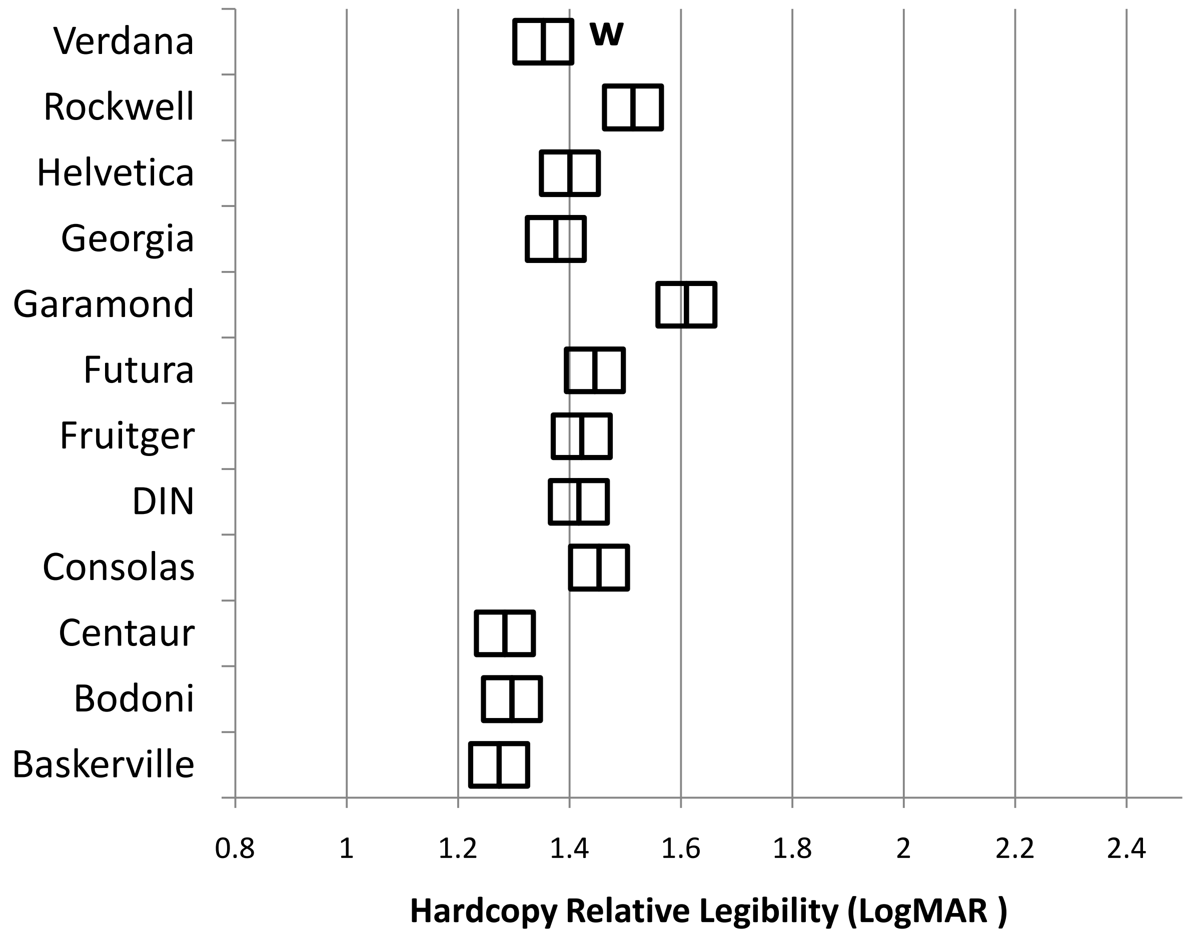
S	Estimate	Std. Error	df	t	Sig.	95% Confidenc Interval	
						Lower Bound	Upp Bou
Intercept	-3.559	1.002	321.116	-3.553	0.000	-5.530	1.5
Max Height of Letter	-0.001	0.001	320.000	-2.504	0.013	-0.002	0.0
Max Width of Letter	0.001	0.000	320.000	1.532	0.127	0.000	0.0
Serif	-0.094	0.055	320.000	-1.721	0.086	-0.202	0.0
MS Max Width	-0.015	0.003	320.000	-5.295	0.000	-0.021	0.0
MS minimum width	-0.013	0.003	320.000	-4.611	0.000	-0.019	0.0
MS Width Ratio (Max/Min)	-0.245	0.059	320.000	-4.169	0.000	-0.361	0.1
Max width of stroke perpendicular to point of tangency of vertical dimension, upper curve	0.045	0.009	320.000	5.156	0.000	0.028	0.0

Max vertical dimension of stroke, upper curve	-0.015	0.004	320.000	- 4.112	0.000	- 0.022	0.0
Ratio of previous two parameters	1.717	0.345	320.000	4.977	0.000	1.038	2.3
Max vertical dimension of stroke, lower curve	0.011	0.003	320.000	4.412	0.000	0.006	0.0



V	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.717	0.072	128.472	23.987	0.000	1.575	1.859
Max Height of Letter	0.000	0.000	324.000	-1.312	0.191	-0.001	0.000
Max Width of Letter	-0.001	0.000	324.000	-4.288	0.000	-0.001	0.000
Serif	0.577	0.131	324.000	4.405	0.000	0.320	0.835

MS Max Width	0.002	0.001	324.000	2.268	0.024	0.000	0.003
MS Width Ratio (Max/Min)	-0.062	0.017	324.000	-3.585	0.000	-0.096	-0.028
Opening size	0.001	0.000	324.000	4.582	0.000	0.001	0.002

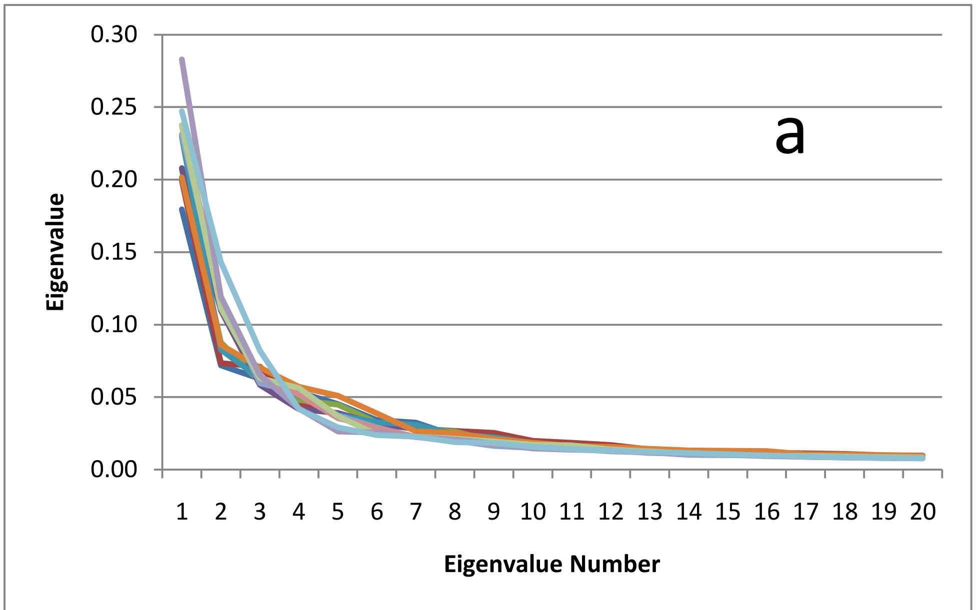


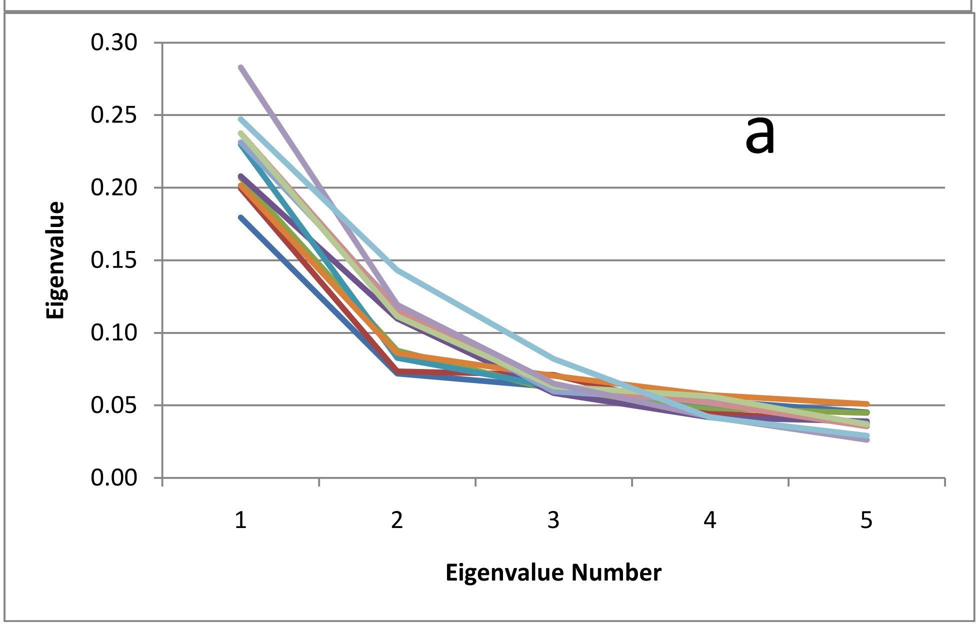
W	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.069	0.091	281.278	11.770	0.000	0.890	1.247
Max Height of Letter	0.002	0.000	323.000	5.935	0.000	0.002	0.003
Max Width of Letter	0.001	0.000	323.000	5.527	0.000	0.000	0.001
Serif	0.326	0.080	323.000	4.060	0.000	0.168	0.484
MS Max Width	-0.009	0.002	323.000	-5.997	0.000	-0.012	-0.006
MS Width Ratio (Max/Min)	0.080	0.020	323.000	4.011	0.000	0.041	0.119
L upper opening size	0.002	0.000	323.000	5.021	0.000	0.001	0.003
lower opening size	-0.004	0.001	323.000	-7.525	0.000	-0.005	-0.003

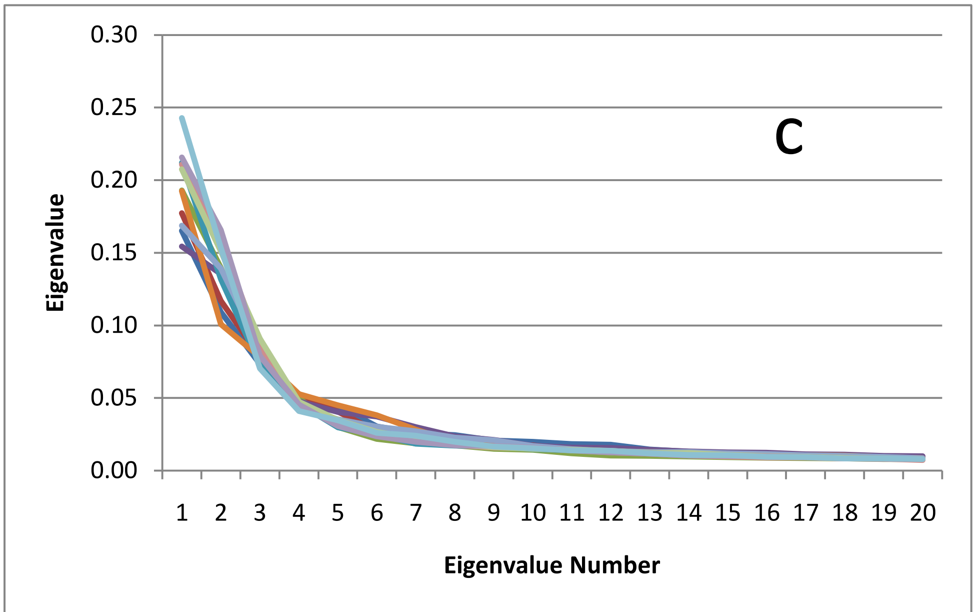
Conclusion for individual letter attributes.

We have demonstrated significant relationships between individual letter attributes and relative legibility. The statistical limitations in our dataset require caution in the interpretation of the results. We need the advice of the font designers to inform us on whether this information is helpful in the design process. Further method validity could be tested by modifying fonts with poor legibility according to the suggested improvements to determine if a causal relationship between the attributes and legibility. Replication in other measures of legibility and fonts will help determine if these findings are robust.

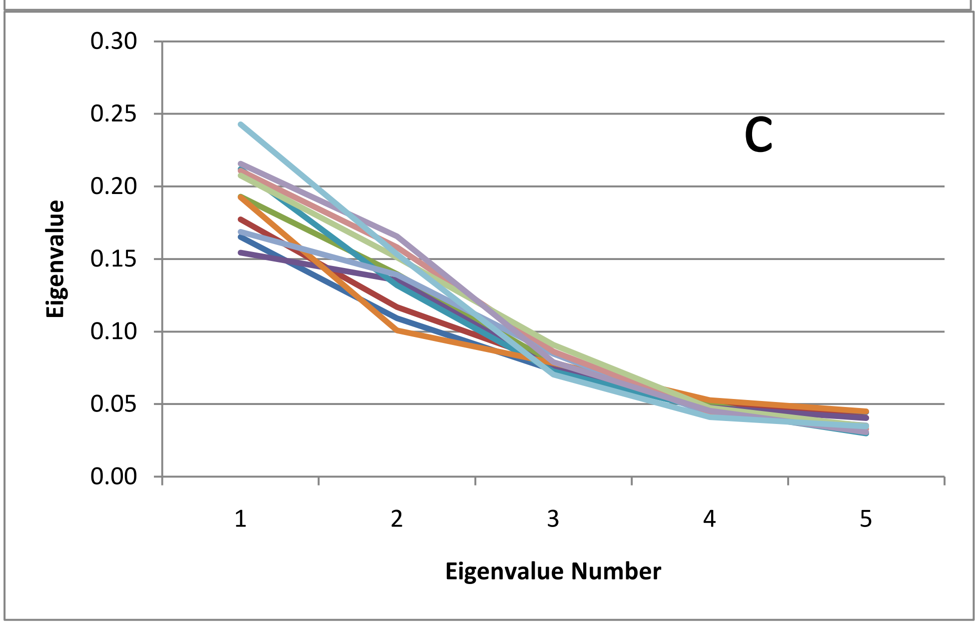
Appendix A. SVD details



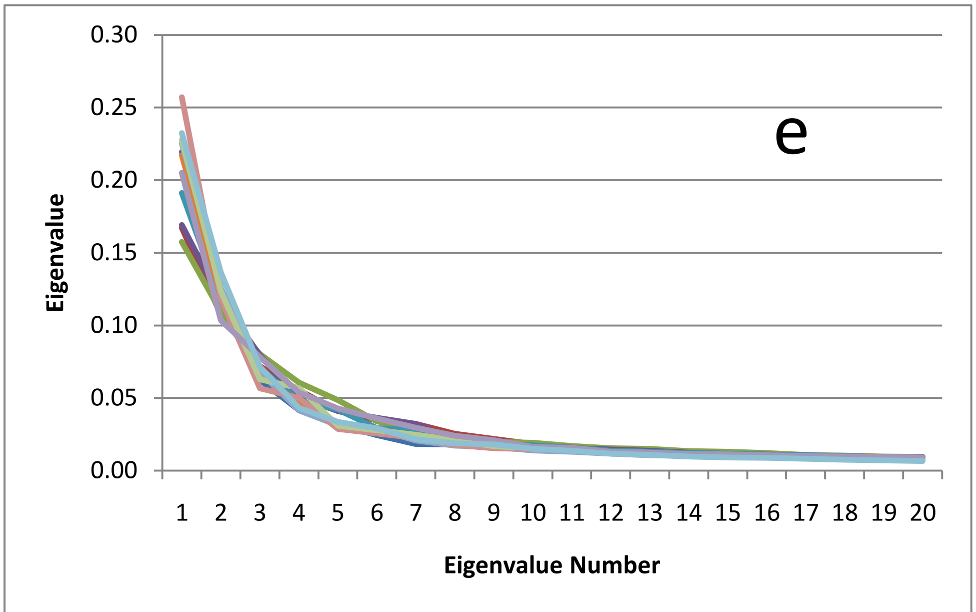


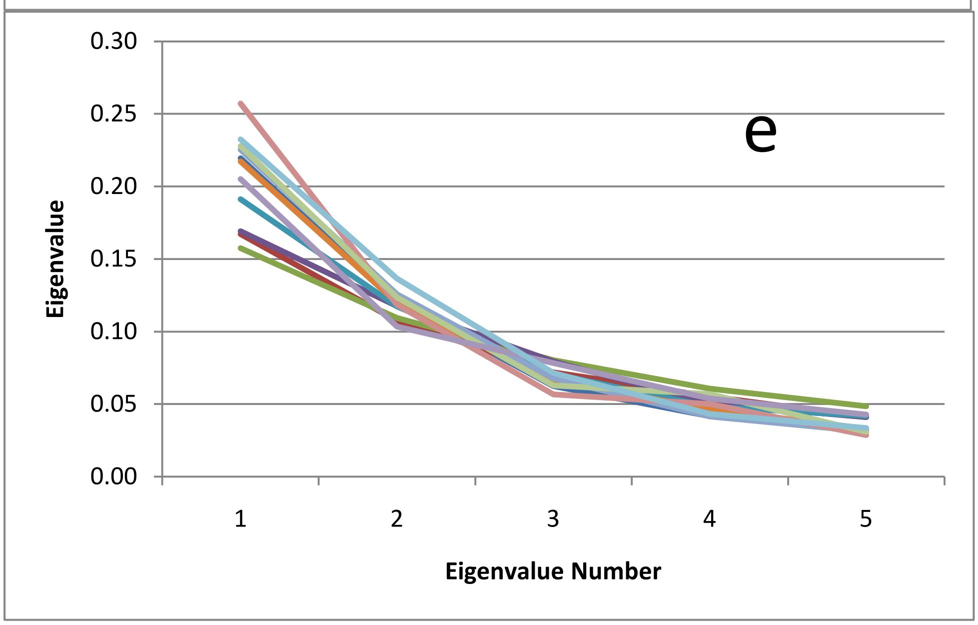


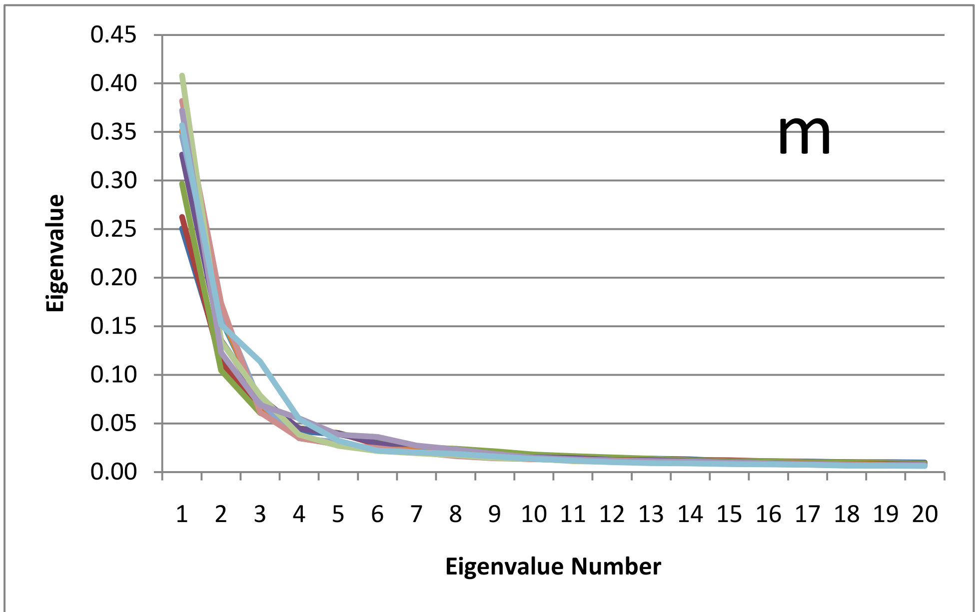
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Eigenvalue Number



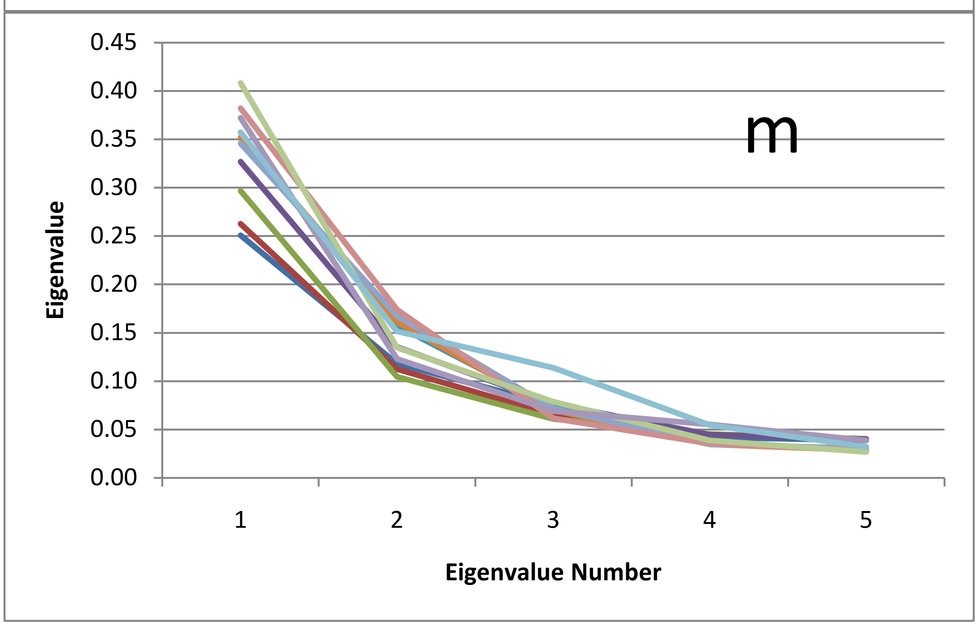
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
C
1 2 3 4 5
Eigenvalue Number


e
Eigenvalue
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Eigenvalue Number

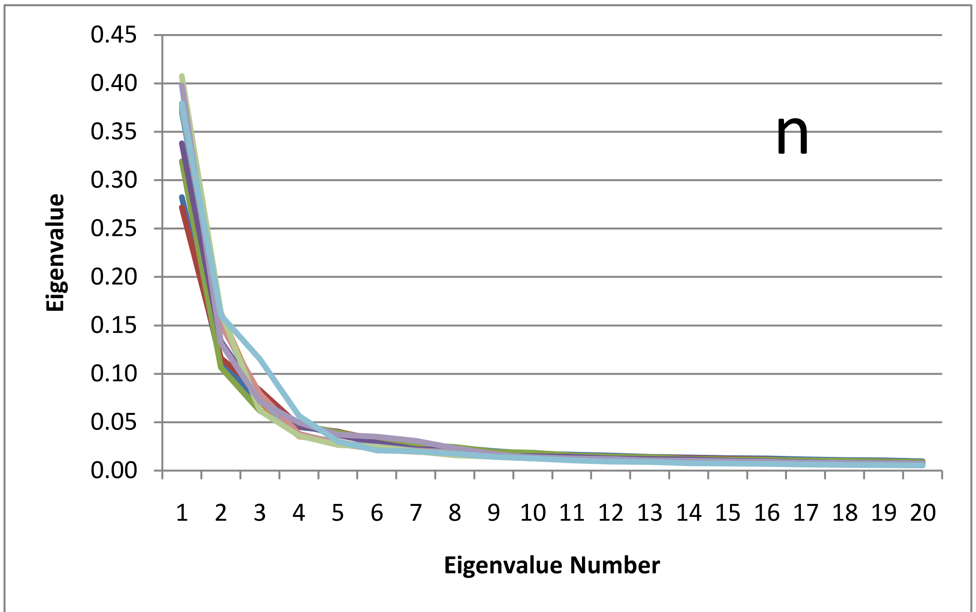

e
Eigenvalue
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5
Eigenvalue Number

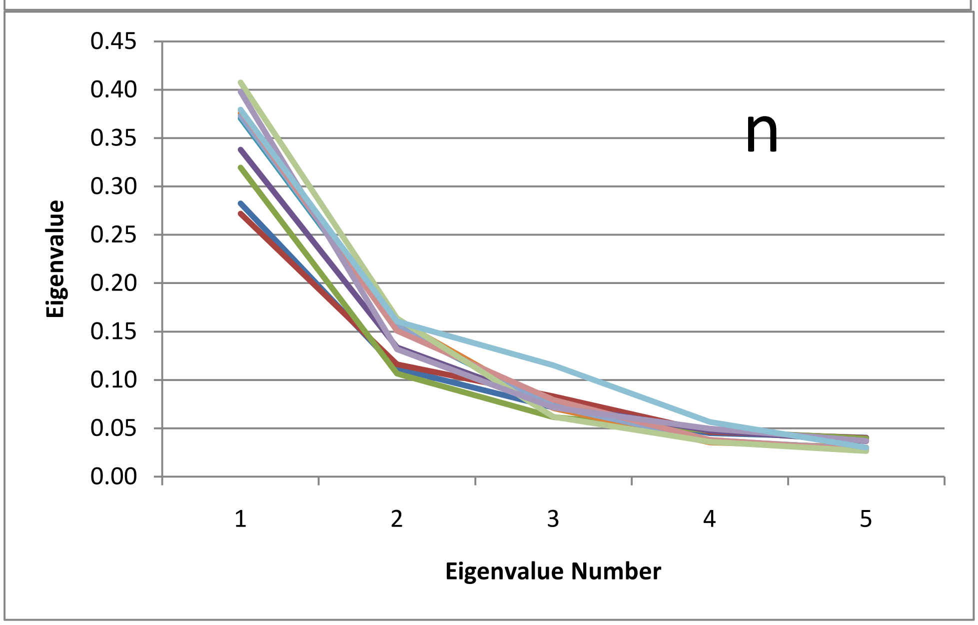


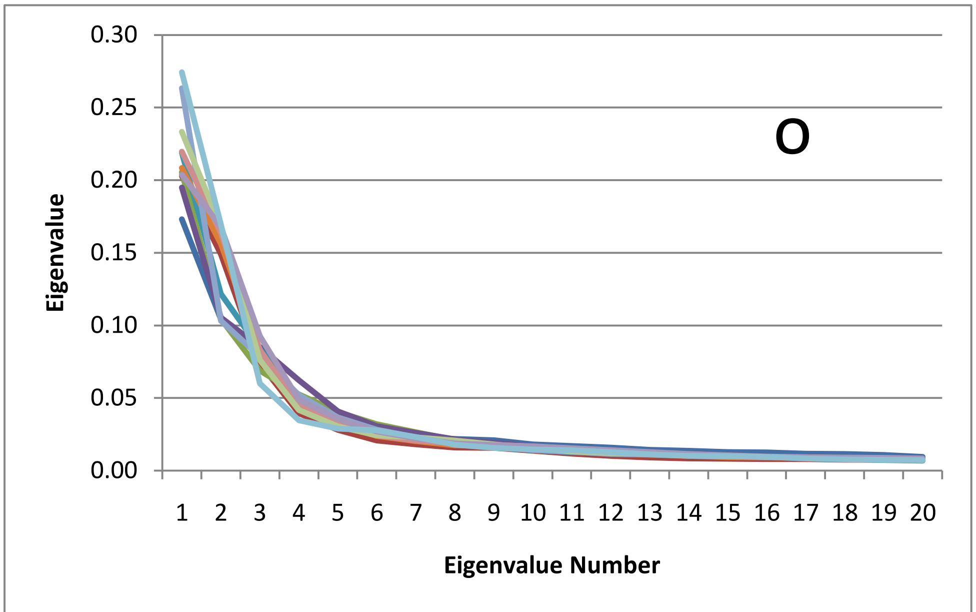
m
Eigenvalue
Eigenvalue Number

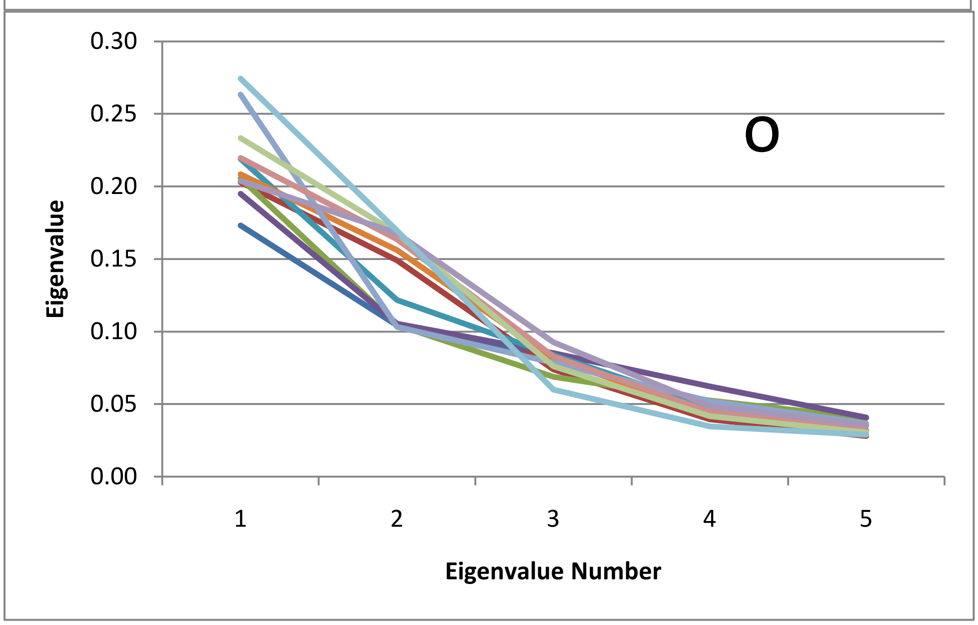


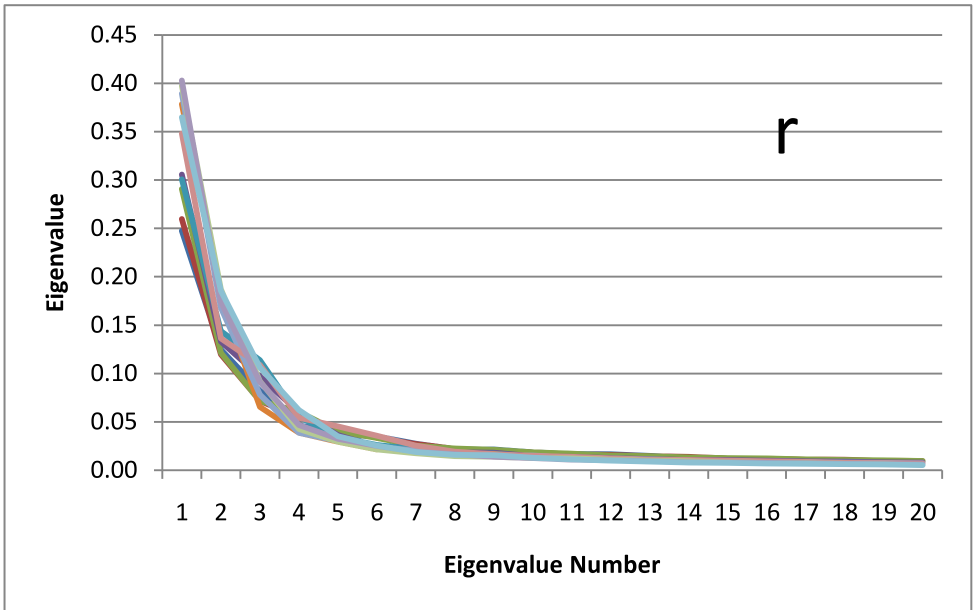
m
Eigenvalue
Eigenvalue Number


0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
n
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Eigenvalue Number

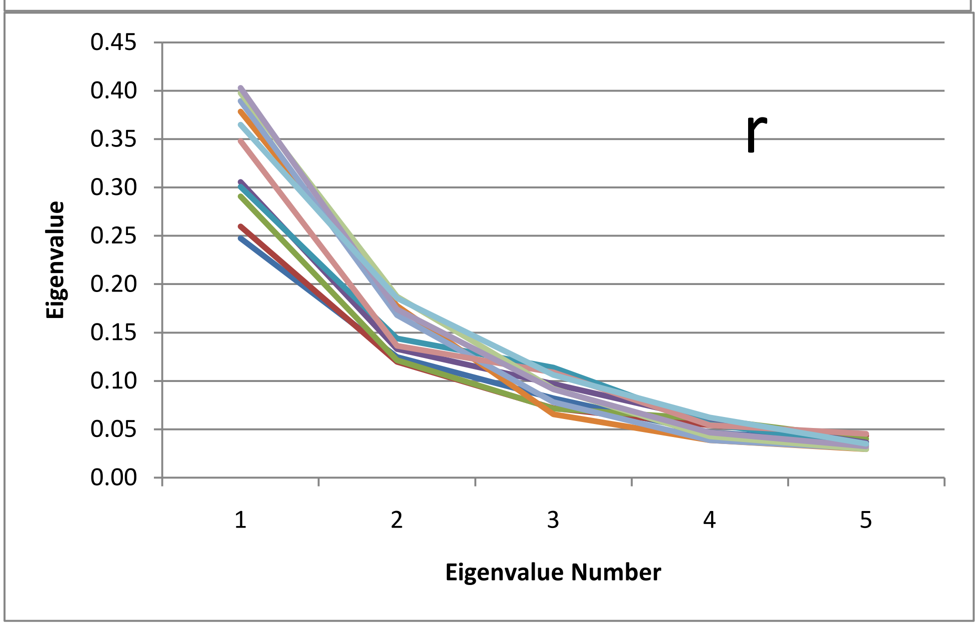

0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
n
1 2 3 4 5
Eigenvalue Number


O
Eigenvalue
Eigenvalue Number

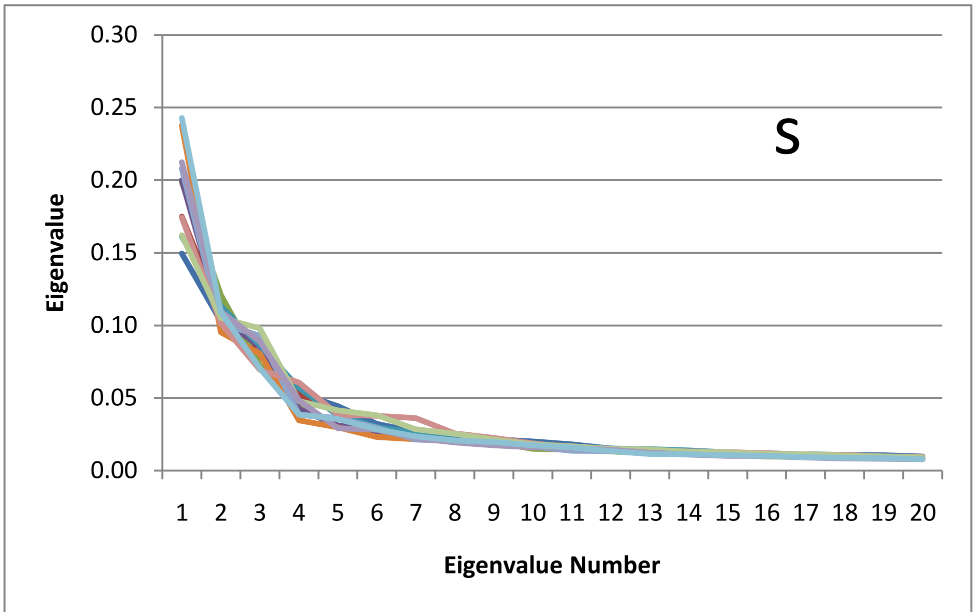

O
Eigenvalue
Eigenvalue Number

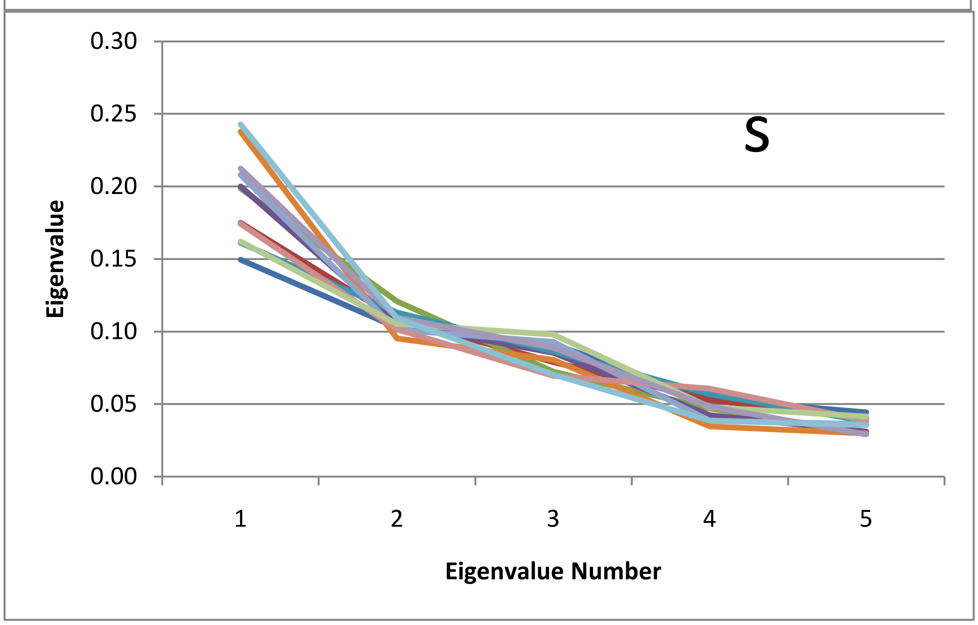


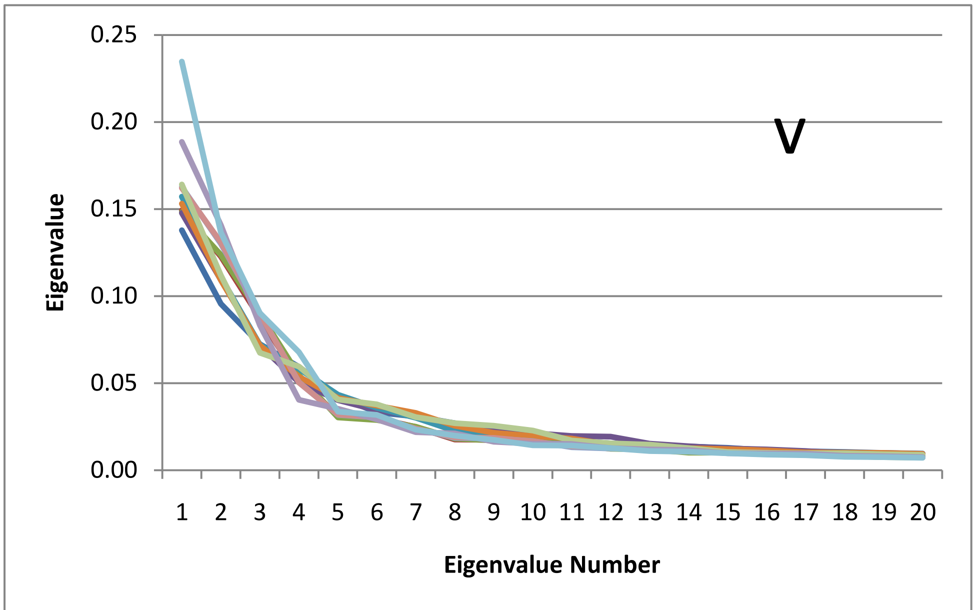
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
r
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Eigenvalue Number

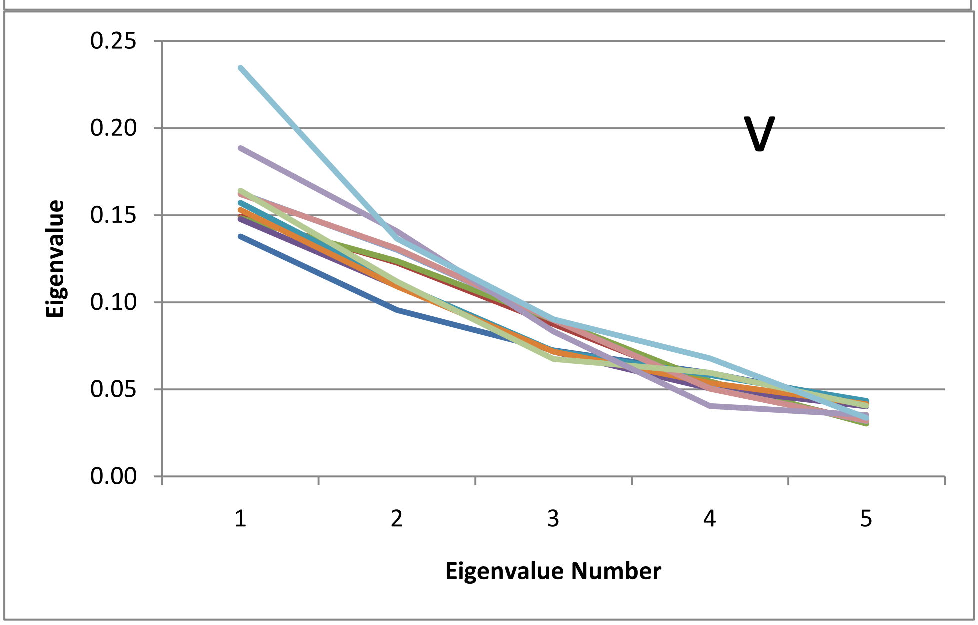


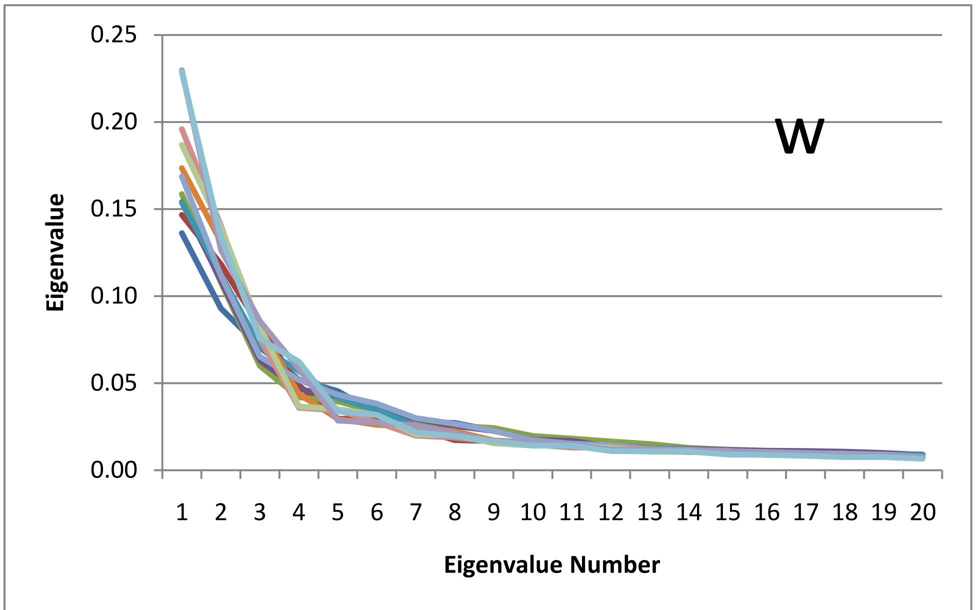
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
r
1 2 3 4 5
Eigenvalue Number

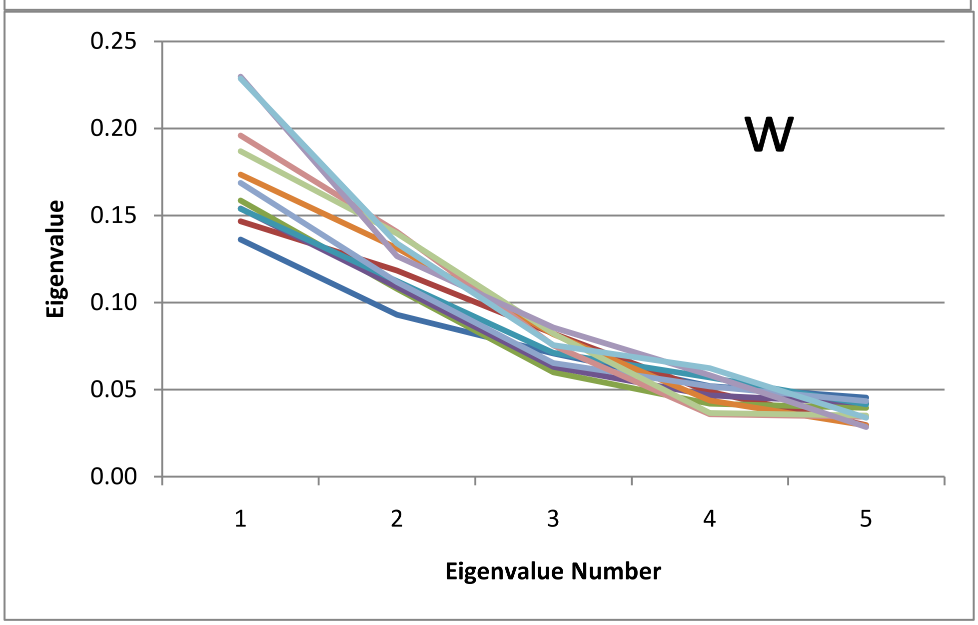

0.30
0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
S
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Eigenvalue Number


0.30
0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
S
1 2 3 4 5
Eigenvalue Number


0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
V
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Eigenvalue Number


0.25
0.20
0.15
0.10
0.05
0.00
Eigenvalue
V
1 2 3 4 5
Eigenvalue Number


W
Eigenvalue
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Eigenvalue Number


W
Eigenvalue
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5
Eigenvalue Number

Appendix B: MatLab programs for computing singular value decomposition

FontSVDScript.m

```
%This script performs a Principal Component Analysis of a font (SVD of
%Font)  Note that it requires a matlab function file called fontsvd.m to
%perform the analysis.

%Read all the jpg files
f=dir('*.bmp')

%Creat a Matrix Y
Y=[]

%For each of the jpg files perform a singular value decomposition
for i=1:length(f)
   X=f(i).name;
   Y=[Y ' ' X]
   Final =fontsvd(X,i);
end;

%Read a file called evals.csv. This file is created in the function
%fontsvd.
Result=dlmread('evals.csv');
Result=Result';

%Write the data to an excel file
xlswrite('evals.xls',Result);
```

FontSVD.m

```
%This function performs an SVD on a jpg image file; it is designed to be
%used with the matlab file fontsvdscript.m

function [q]=fontsvd(letter,i)

%read in the image
X1 = imread(letter);
figure(1), image(X1)

%convert the image from RGB format to GrayScale
%X2 = i(X1)
X2 = X1;
%convert the image from GrayScale to a Matrix of Numbers
X = im2double(X2)

%Find the size of the Matrix
[m,n]=size(X);

%Recenter the data so that the 0's are on the outside and the 1's are on
%the inside
recenter=ones(m,n)-X;
X=recenter;

%Find the Row Mean of X
m1=mean(X,1);
```

```
%Find the Column Mean of X
m2=mean(X,2);

%Subtract the row mean, column mean and total mean from the data
%Note that the most current version of the program does not use mean
%subtraction.  To implement mean subtraction delete the % on line 34 and add
%an % to the front of line 36
%X3=X-repmat(m1,m,1)-repmat(m2,1,n)+mean(X(:));

X3=X

%Singular Value Decomposition
[U3,S3,V3]=svd(X3,'econ');
S3
%Normalization
q=diag(S3)./sum(diag(S3));

%Take the transpose of the matrix q
q=q'

%Write the data to a file called evals.csv
dlmwrite('evals.csv',q,'-append');
```

References

1. Trefethen LN, Bau D. *Numerical Linear Algebra.* Philadelphia, PA: Society for Industrial and Applied Mathematics; 1997.

2. Zimmerman A, Sheedy J, Subbaram M, Hayes J. Font legibility - effects of pixel density and smoothing. *Optom Vis Sci*. 2003;80:195.

3. Payton ME, Greenstone MH, Schenker N. Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance?. *J Insect Sci*. 2003;3:34.