Course Work Cover Sheet - the School of Computing

Your name: Kyle Harrison

Matriculation number: 11000 9870

Module code:AC42001

Assignment title (see module guide): Lab Report 3

Assignment number (see module guide):2

Date submitted: 3/02/2014

Date due to be submitted: 3/02/2014

Number of hours you spent on this item of coursework:

Numbers of hours you were advised to spend (see module guide):

Word count (if applicable): 2,163

Your course work may be used by the School of Computing for demonstration purposes with other classes and/or members of the public, unless you indicate otherwise.

Tick the box at the right if you do *not* wish your work to be used in this way.

I certify that this assignment represents all my own work and that no parts of it have been copied and that no collusion has taken place with any other person. I have also read and understood the School of Computing's regulations as regards course work and I certify that this course work was performed according to those regulations.

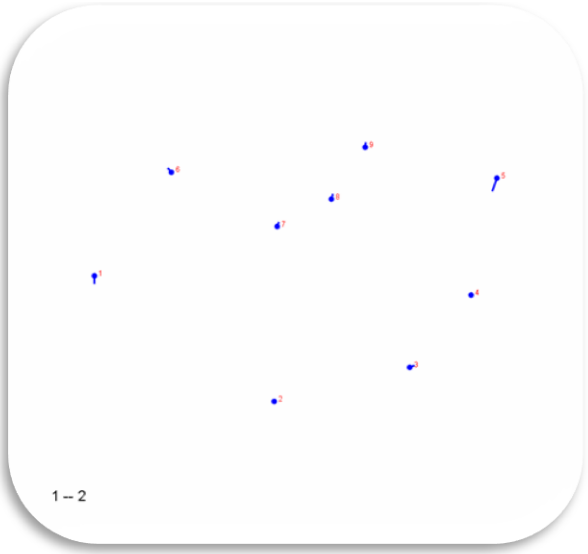
I certify that any files submitted with this assignment have been virus checked and have no associated viruses.

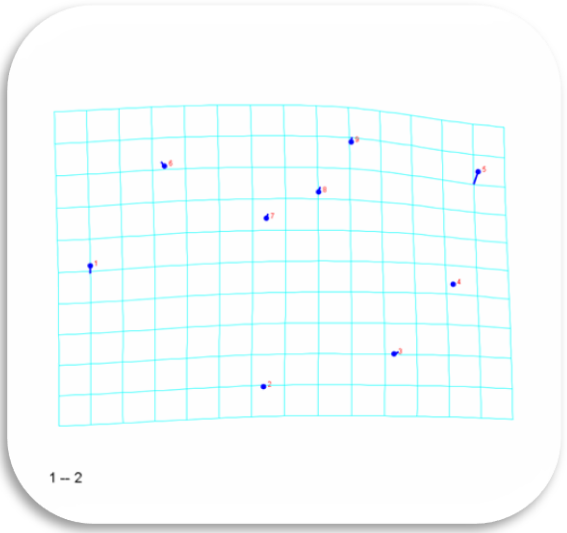
Signed: Kyle harrison

Lab Report 2

HANDS RESULTS

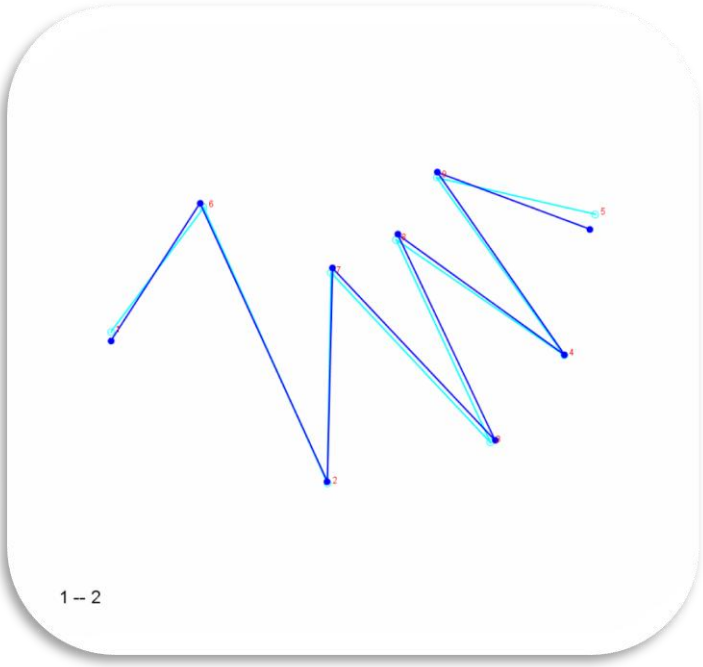
Lollipop graph & Transformation Grid





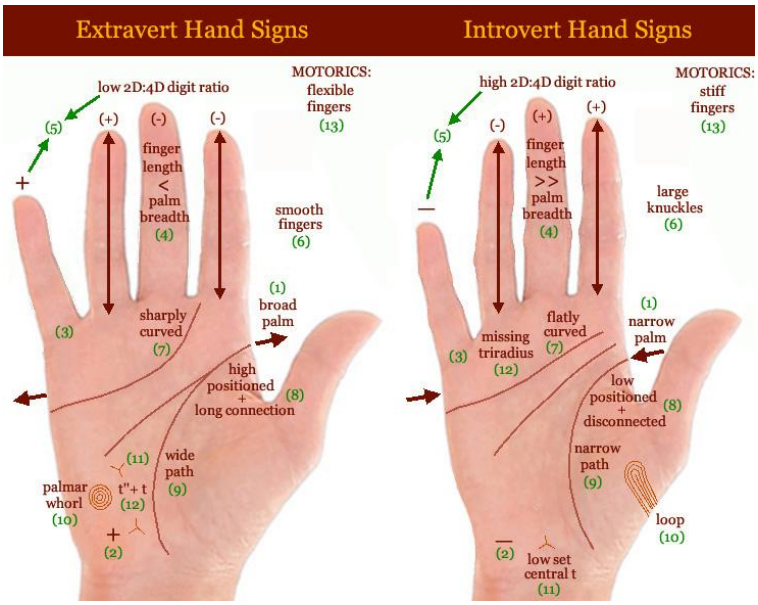
The lollipop graph shows that the maximum variance in the graph is in the position of the little finger. The only noticeable changes in the transformation grid is the distortion around the tip of the little finger. The transformation grid shows a deformation of the grid around the little finger. This can be attributed to the position of the tip of the little finder from the landmarks however it is possible it is due to error as land marking the tip of the little finger could be preformed incorrectly.

Hands Wireframe



In the wireframe the variance can easily be seen and identified as the little finger as the thumb is positioned on the left of the wireframe image.

Assumptions : Low 2D:4D digit ratio + lonk pinky point to Extraversion in females

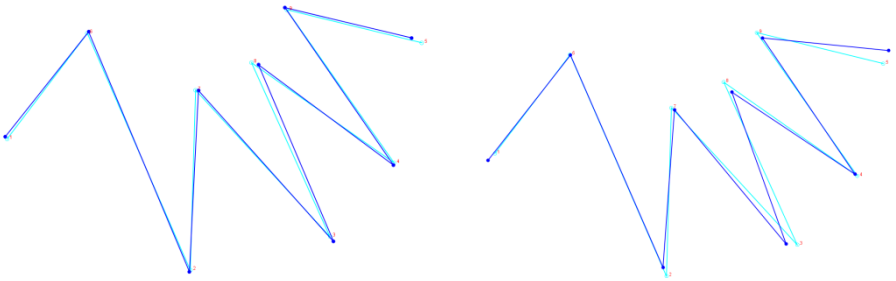


After various digit ratio studies found conflicting results related to Extraversion & sensation seeking (most pointed into the direction of a negative correlation), a new Dutch 2012 study suggest that other finger lengths require to be taken into account in order to find more clear results!

The (unpublished) results from a female only suggest that low 2D:4D digit ratio + long pinky finger are much more often found in the hands of 'extroverts' (8 out of 40 hands = 20%), while this combination is rarely seen in the hands of 'introverts (0 out of 28 hands = 0%).

Additionally, a high 2D:4D digit ratio with short pinky finger is much more common in the hands of 'introverts' (5 out 28 hands = 18%), compared to the hands of 'extroverts' (3 out of 40 hands = 7.5%).[1]

Canonical variate 1 vs Canonical variate 2

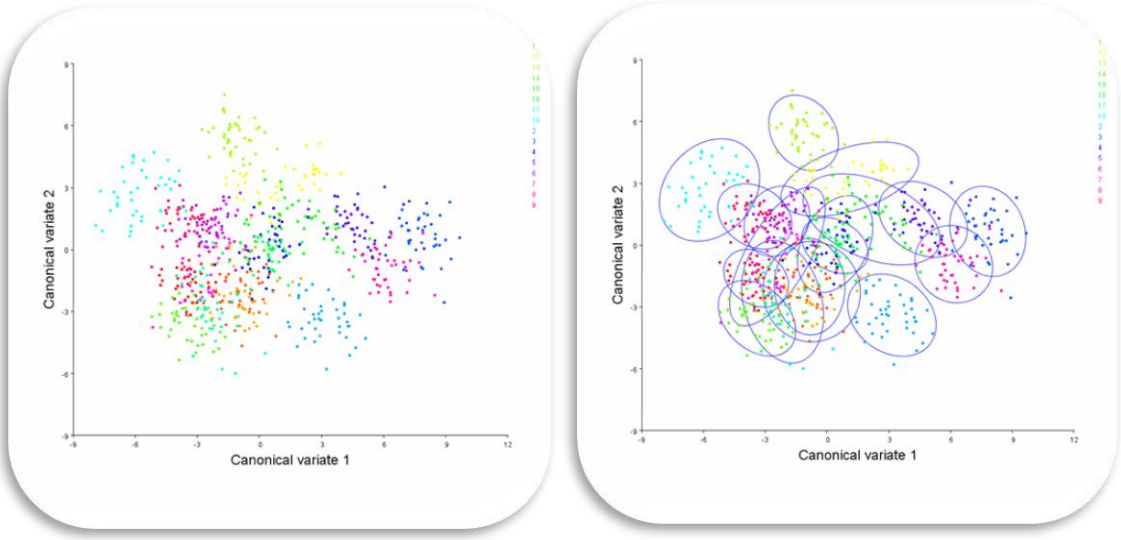


cv1 **CV2**

Canonical variate 2 shows that the little finger is indeed extends further outwards and the index finger is longer making the 2nd and 4th digit shorter in comparison. For this reason CV2 may show females who are more likely to be introverts or extroverts.

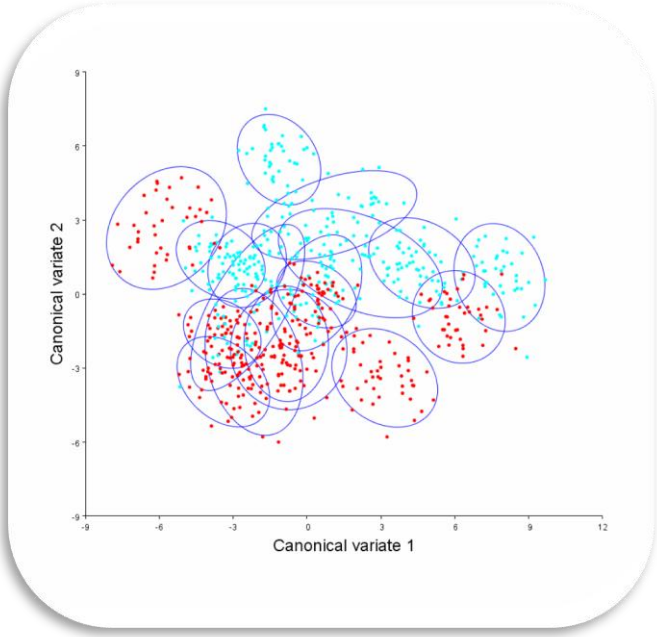
[1] http://fingerlengthdigitratio.wordpress.com/tag/pinky/

Canonical variate 1 vs Canonical variate 2 - Confidence Ellipses - Individual



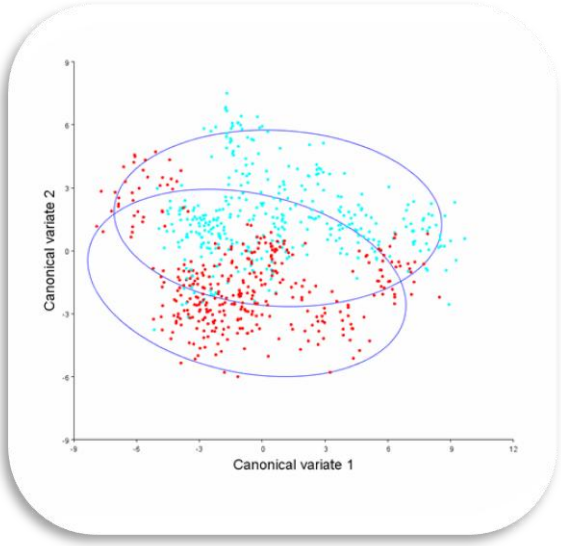
Individuals can be identified by their hands as shown by the clustering of colours to identify individuals. The data suggests approx 5 individuals can be identified by their hands however these could also be out liars as they also provide the largest variance or they could be attributed towards noise in the dataset. Typically the majority of individuals are clustered somewhere in the middle with a large degree of overlap..

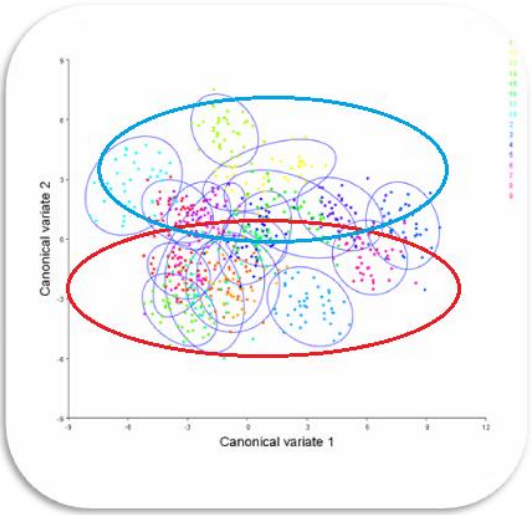
CV1 & CV2 - Confidence Ellipses - Colored by sex - Grouped by Indivdual.



The graph shows male and female hands. The hands contain very little overlap however two clusters of female hands appear to diverge from the concentrated cluster and overlap with a few male hands. These could be attributed to nose, errors in land marking/ classification, or even simply females that have the maximum variance in hand sizes along PC1.

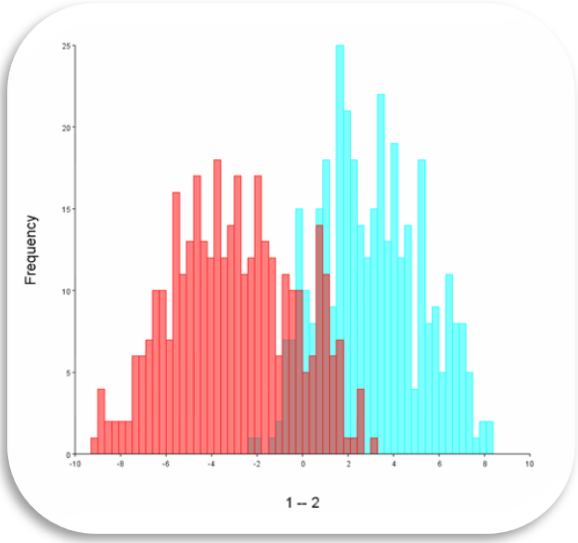
CV1 & CV2- Confidence Ellipses - Colored by sex - Grouped by Sex.

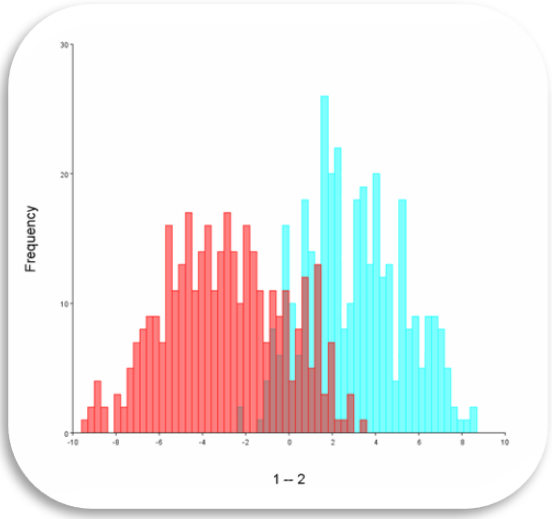




There is a very clear definition between both male and female results even though there is some overlap in the confidence ellipses. Ignoring the 2 female hands with the most variance you can clearly separate the hands into male/female categories. Although individuals cannot be clearly identified by their hands from this dataset, their gender can be inferred. Ideally a smaller overlap seen in the second graph would provide much higher success rates.

Discriminant Scores **Cross Validated Scores**





The discriminate scores shows what has been computed for each person to allow for interpretation of Between-group differences. Each discriminate function measures something unique and different. It also allows for a reduced error rate and multiple dependant variables. In the discriminant graph, it can be seen that there is very little overlap at scale zero when comparing it to the rest of the graph. Male/female can clearly be identified with a reasonable degree of confidence. The discrimiant function depicts the overlap we see in the confidence ellipses when grouping by sex. Identifying the two female clusters on both extremes of CV1 & CV2 and determining if any there is any error in this data may allow for the discrimiant function /scores to achieve a higher successes rate.

Cross validation scores show the lead one out method does not improve the outcome. This may be attributed to the noise generated by the two females in the data set. As one will always exist in the data there will be a higher degree of overlap between both groups.

Canonical Variant analysis Results

	Eigenvalues	% Variance	Cumulative %
1.	12.79035435	40.990	40.990
2.	5.61674360	18.000	58.991
3.	4.47755427	14.350	73.340
4.	2.34291672	7.509	80.849
5.	1.71691197	5.502	86.351
6.	1.39819849	4.481	90.832
7.	1.20326674	3.856	94.688
8.	0.80328539	2.574	97.263
9.	0.30254291	0.970	98.232
10.	0.23165155	0.742	98.975
11.	0.14214730	0.456	99.430
12.	0.09132624	0.293	99.723
13.	0.06492283	0.208	99.931
14.	0.02153003	0.069	100.000

The first three results make up approx 75% of the maximum variance, beyond the 8th result, the variance starts to become noise.

PCA Results

	Eigenvalues	% Variance	Cumulative %
1	0.001064	41.01	41.01
2	0.000343	13.198	54.208
3	0.000262	10.101	64.31
4	0.000242	9.316	73.625
5	0.000174	6.701	80.326
6	0.000147	5.673	85.999
7	0.000111	4.26	90.259
8	7.75E-05	2.987	93.247
9	4.65E-05	1.791	95.038
10	4.19E-05	1.614	96.652
11	3.17E-05	1.221	97.873
12	2.51E-05	0.967	98.839
13	1.71E-05	0.658	99.497
14	1.31E-05	0.503	100

Discussion

From discriminant function:

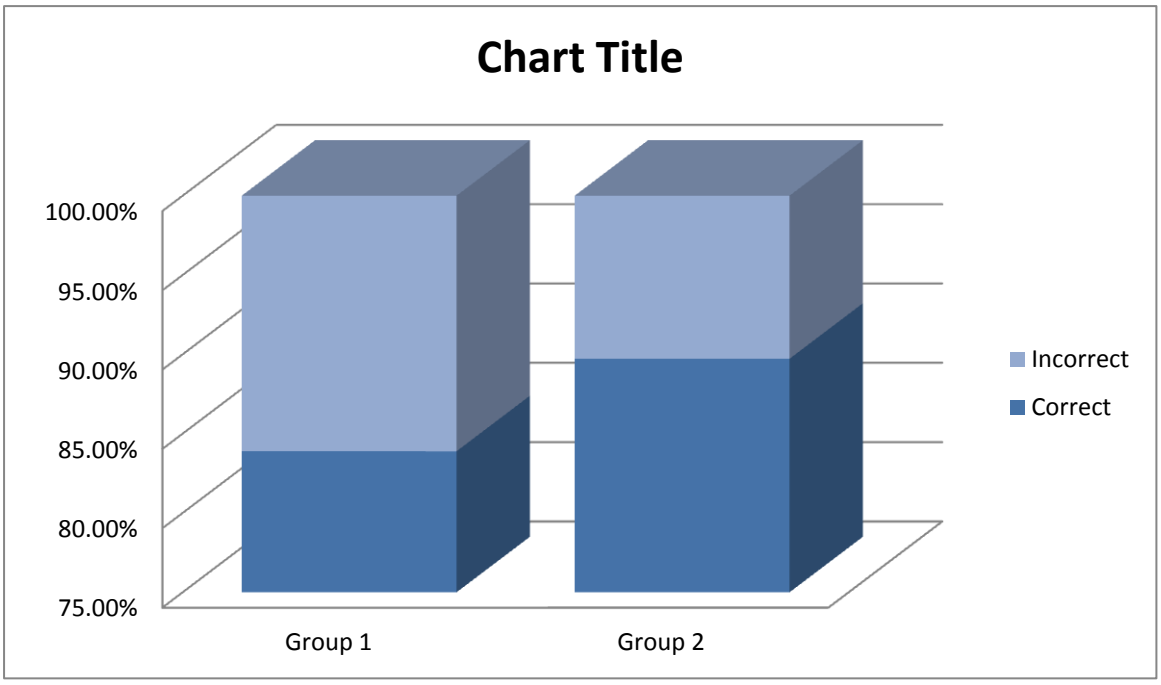
Group	Group 1	Group 2	Total
Group 1	304	56	360
Group 2	34	326	360

Group	Group 1	Group 2	Total
Group 1	**84.44%**	15.56%	100%
Group 2	9.44%	**90.56%**	100%

From cross-validation:

Group	Group 1	Group 2	Total
Group 1	302	58	360
Group 2	37	323	360

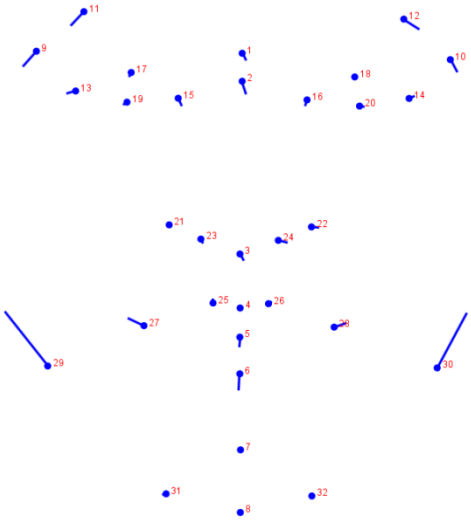
Group	Group 1	Group 2	Total
Group 1	**83.89%**	16.11%	100%
Group 2	10.28%	**89.72%**	100%

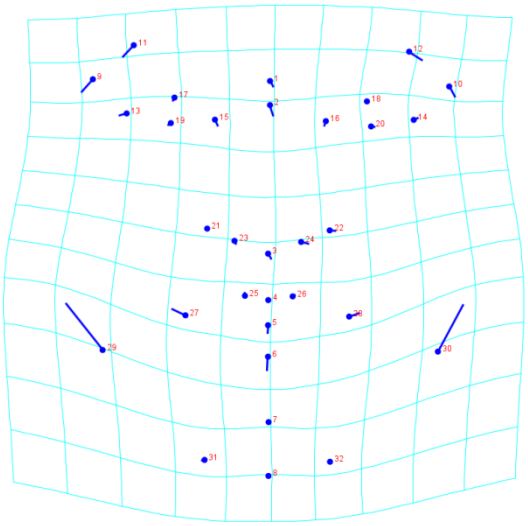


Group 1 shows 84.44% chances of being correct whilst Group 2 shows a 90.56% change of being correct. Cross validation shows that Group 1 has a 83.89% chance of being correct and group p2 has a 89.72% chance of being correct. For both groups this translates to roughly 9 out of 10 which provided a high degree of classifying hands correctly. This allows for male & female hands to be identified reliably and the identification of a cluster of hands belong to one individual to be carried out also.

FACES RESULTS

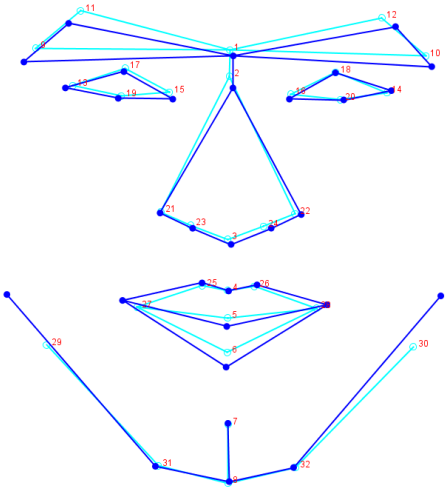
Lollipop graph and transformation grid





The lollipop graph shows the landmarks at the cheeks extending forward. It would be expected that the most lateral point at the angle of the mandible on the left and right would be much higher and wider on a female face than a male face producing a large degree of variance in this direction as the landmarks on male faces would be in a lower and narrower position.

Wireframe

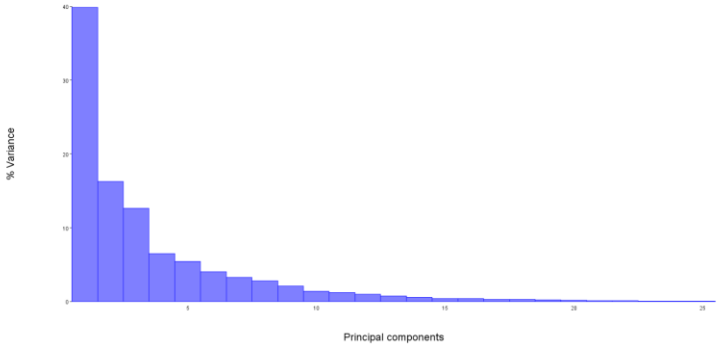


The dark blue line in the wireframe shows a female face whilst the light blue line shows a male face. Comparing this to the image on the right the male face is much lower but not as wide, where as the female face has higher landmarks and the face is slighltly wider maching the results in the wireframe.

[2] http://castleknockcosmetics.ie/wp2/wp-content/uploads/2011/11/m-vs-f-eyebrows.jpg

Principcal components analysis

Principcal compoent 1 shows the maximum variance in jaw line and highlights the difference between genders as it contributes nearly 40% of the variance with PC2 at 16% and PC3 at 12%



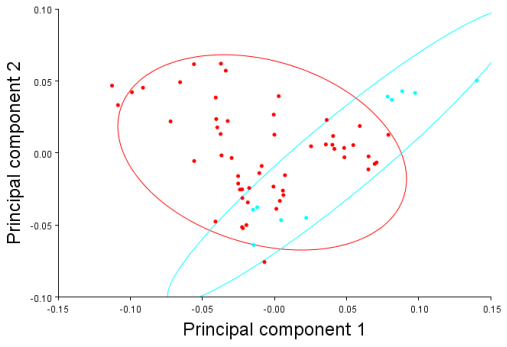
Principal Component Analysis: PCA: CovMatrix, CombinedSetFull, Procrustes coordinates

	Eigenvalues	% Variance	Cumulative %
1.	0.00388723	39.886	39.886
2.	0.00158425	16.255	56.141
3.	0.00123314	12.653	68.794
4.	0.00063133	6.478	75.272
5.	0.00053153	5.454	80.726
6.	0.00039176	4.020	84.745
7.	0.00031618	3.244	87.989
8.	0.00027500	2.822	90.811
9.	0.00020488	2.102	92.913
10.	0.00013568	1.392	94.306
11.	0.00011698	1.200	95.506
12.	0.00009875	1.013	96.519
13.	0.00007333	0.752	97.271
14.	0.00005575	0.572	97.843
15.	0.00004088	0.419	98.263

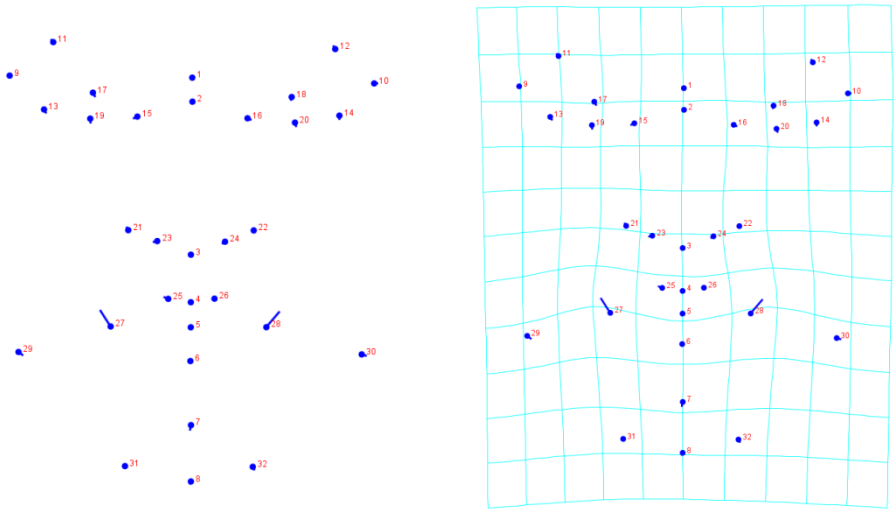
The first three principal components contribute nearly 70% of the overall variance of the face however everything up to 12 could be used to identify faces however after this point the variance becomes much lower and starts to be effected by noise.

PCA Graph Results

The graph shows some clustering for individuals however for the most part the faces are spread out over the graph.

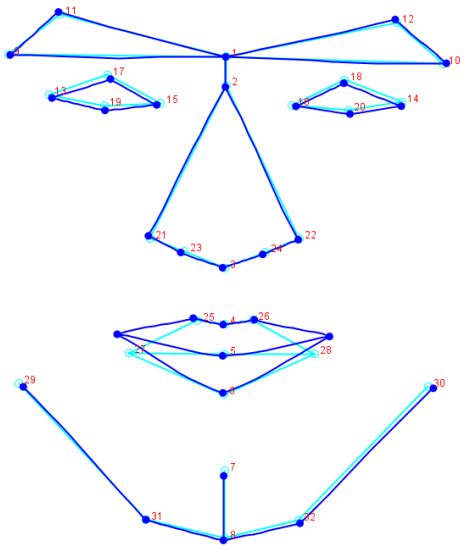


CVA - Lollipop graph and Transformation grid



The lollipop graph shows that the largest variance between the two faces is around the mouth where the most lateral upper and lower parts of the lips meet. This would indicate that a facial change in the mouth has been made such as changing from a neutral face to smiling/disgust.

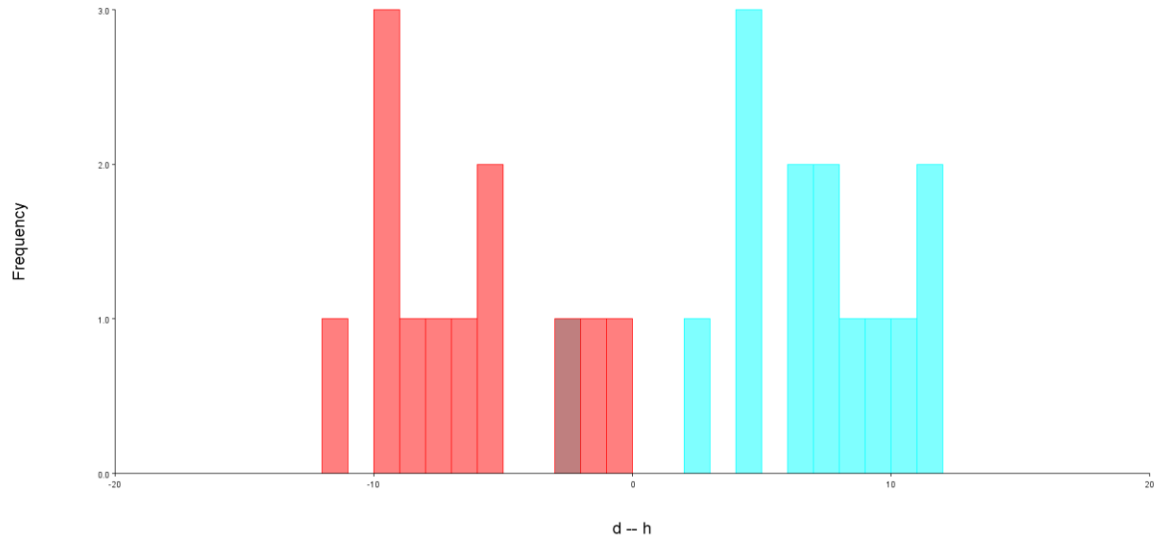
Wireframe



In the wireframe, although there is a difference between the two faces it is not a clear definition between the two (happy & disgust). This could be attributed to the fact that each individual had to produce an emotion without genuinely experiencing it. If the class consisted of individuals that were trained in this exercise then the data may more closely reflect this.

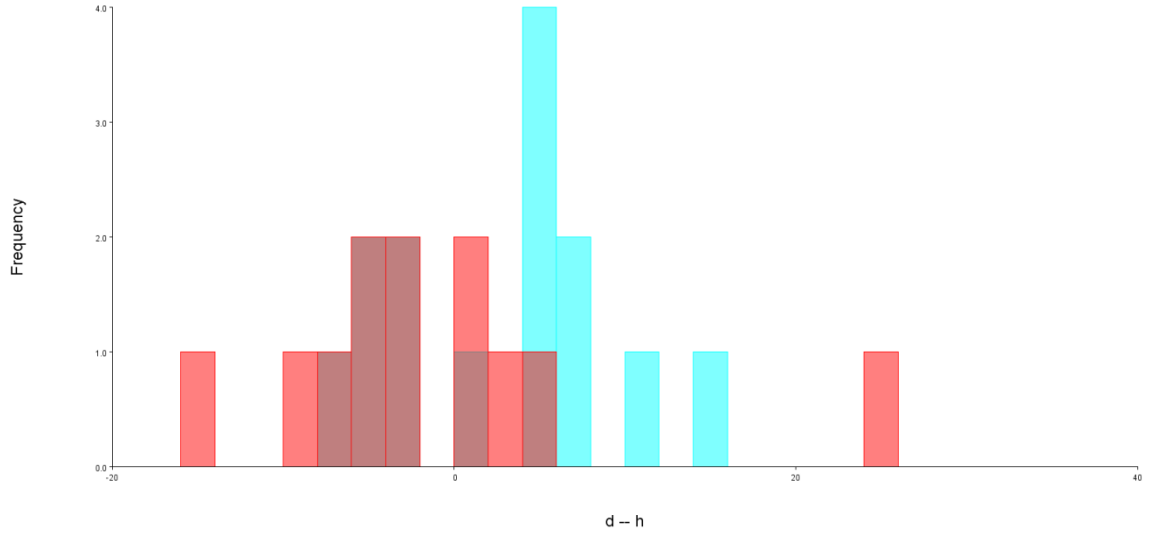
As individuals interpreted the disgust face in different ways there is no way to differentiate between these as some disgust faces were almost identical to neutral where as the class was capable of producing a happy face as it's a standard expression that can be recreated easily.

Discriminant Scores



The discriminate scores show very little overlap in the faces and identifies a clear difference between the two. With a high degree of confidence a happy face can be distinguished from a neutral face.

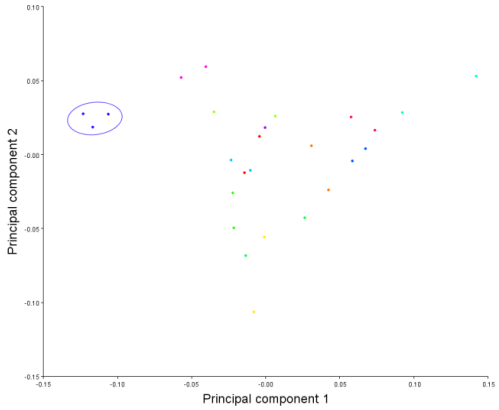
Cross-Validation Scores



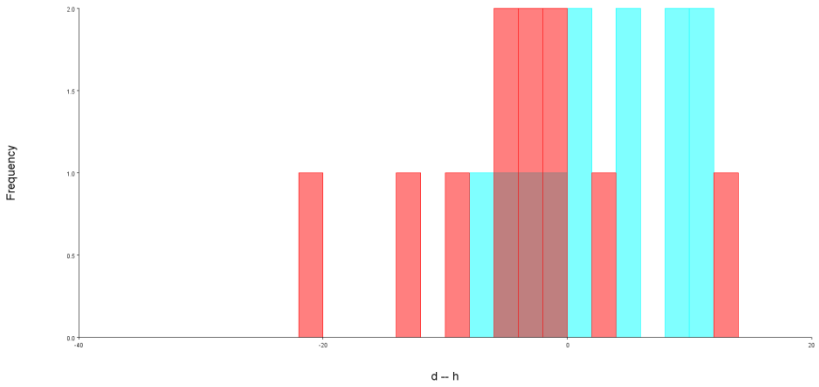
However when cross validation is run the results change dramatically and it no longer remains as simple to differentiate between the two. One disgust face on the far right of the data for a disgust faces make it difficult to separate happy from disgust.

Using a scale of zero, Anything greater than a frequency of 1 is a happy face and anything below zero with a frequency of 1 or greater is a disgust face. If you used zero as the threshold then you could get some wrong as there is an overlap of disgust faces in the happy portion of the graph. There is also an out liar to the far right of the chart.

Out liars in the dataset



The out liar has been identified from the PCA and confirmed by viewing both disgust and happy faces on the share drive. The photo on the left is a disgust face whilst the photo on the right is a happy face however there is only minor movement between the two images which is lost during procrustes fit. Visually the two images are almost identical causing what should be a happy face to appear as a disgust face. Removing Nathan from the dataset produces the following Cross validation scores.



The cross validation shows a clear definition between happy and disgust. On a scale of zero anything greater is happy, whilst anything less is disgust. The frequency in the centre of the graph is also much higher allowing for overlaps to be easily distinguished from one another however both happy and disgust faces overlap and there is a high margin for error in detecting faces.

Classification/misclassification tables

From discriminant function:

Group	Group 1	Group 2	Total
Group 1	**12**	**0**	**12**
Group 2	**1**	**13**	**14**

Group	Group 1	Group 2	Total
Group 1	**100%**	0%	100%
Group 2	92.85%	**7.16%**	100%

From cross-validation:

Group	Group 1	Group 2	Total
Group 1	**7**	**5**	**12**
Group 2	**5**	**9**	**14**

Group	Group 1	Group 2	Total
Group 1	**58.3%**	41.67%	100%
Group 2	35.71	**64.29%**	100%

FACES DISCUSSION

The discriminate function and cross validation function present two very different sets of results. In the discriminate function it shows that with a very high degree of confidence faces can be identified either 9 out of 10 times, or all of the time. The cross validation shows that faces can be identified correctly a little over half the time leaving a large margin for error. Once again this could be contributed to the fact that during the photographing session the class was asked to produce a face they each interpreted differently which could be confused at times with a happy face. From personal experience, i myself found it difficult to produce a disgust face at first and proceeded to smile involuntary at first before making up a random face. From reviewing the data set it appears some of the class had difficulties making a disgust face also and instead opted for a neutral face which more closely resembles a happy face than the polar opposite. If the emotions being expressed were genuine then the accuracy may improve however the ranges of faces for disgusted faces is very wide and harder to distinguish than happy. The following image shows a range of discusted faced from Google, each very different to each other.

 [3]

However the happy faces are conform to a standard and are easier to detect and distinguish from another data set which is reflected in this data.

[3] http://goo.gl/W3tB3m