Data Face
Translating Data into Abstract Face and Explore its Limitations and Advantages

Thesis Presented by

Suyuan Ji

To The College of Arts, Media and Design
In Partial Fulfillment of the Requirements for the Degree of Master of Fine Arts
in Information Design & Visualization

Northeastern University
Boston, Massachusetts
April 2018

Thesis Advisor: Nathan Felde

Readers: Shuishan Yu

Respondent: Kristian Kloeckl

Thesis Committee: Miso Kim, Pedro Cruz

TABLE OF CONTENTS

PREFACE

This thesis began with the interest in the alteration of the eye shape as a cultural activity.

The aim of that research was to explore people's perceptions indicated by eye shape and movement according to different demographic factors.

Visual perception is the ability to interpret the surrounding environment by processing information that is contained in visible light.

Then, my thesis topic moved to the face as a frame for data visualization.

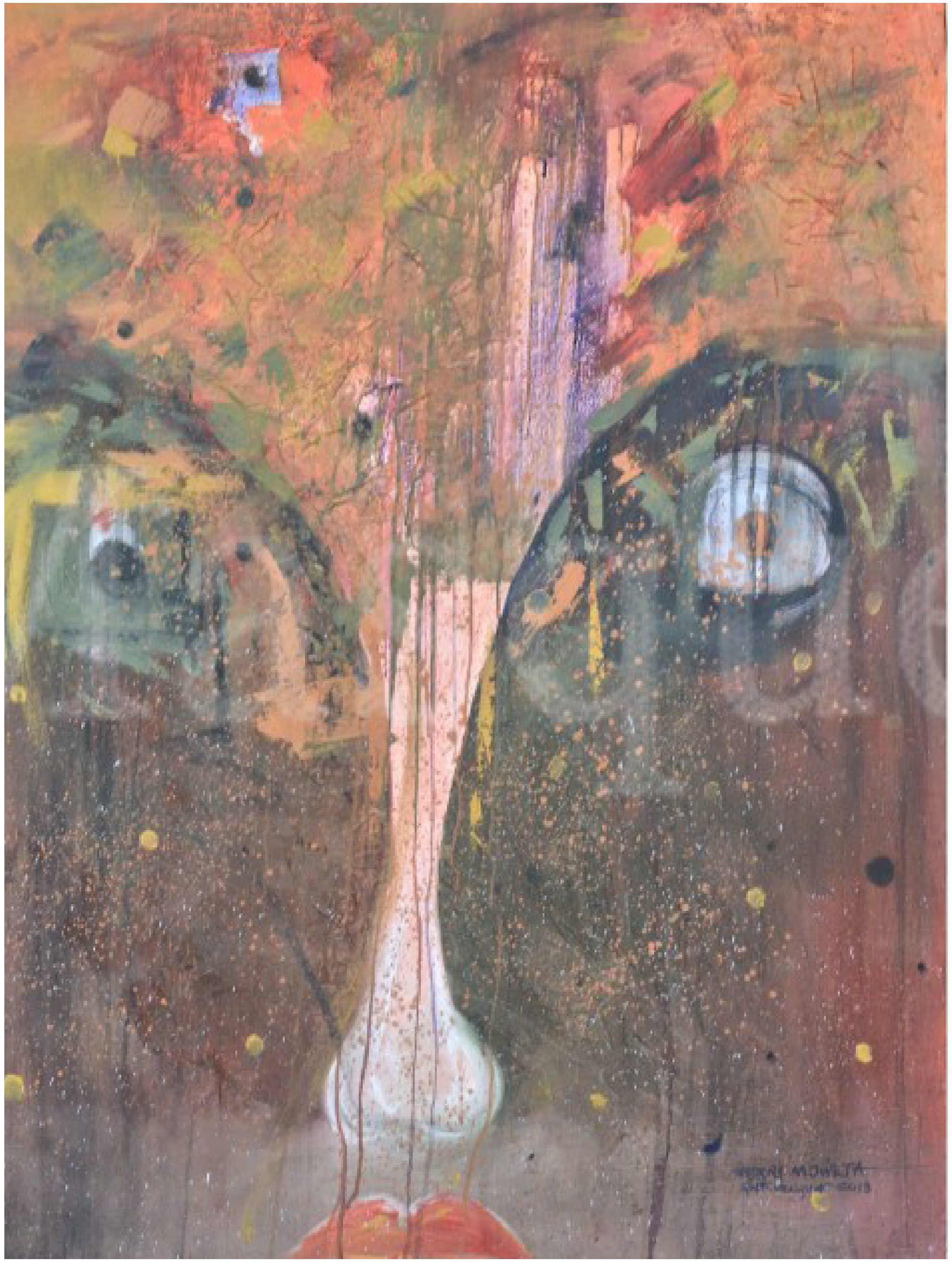


Figure.1. Abstract face.
Artist: Henri Moweta
http://www.masque.com.ng/face-2013-

ABSTRACT

This thesis attempts to explore how facial components can present data. From semiotic perspectives, biological components of a face can form unique symbols and signs to distinguish data. In this thesis, the process of creating a code using a face explains the relationship between signs and signification in terms of abstract faces conveying different information to express emotions.

My thesis applies a few cases to explain how the abstract face transforms different datasets to express the world, such as in the manners of Chernoff faces. Thorough testing, there was consistent design created to investigate the idea in this thesis; experiments have been conducted to examine whether the results are consistent with the findings of previous studies.

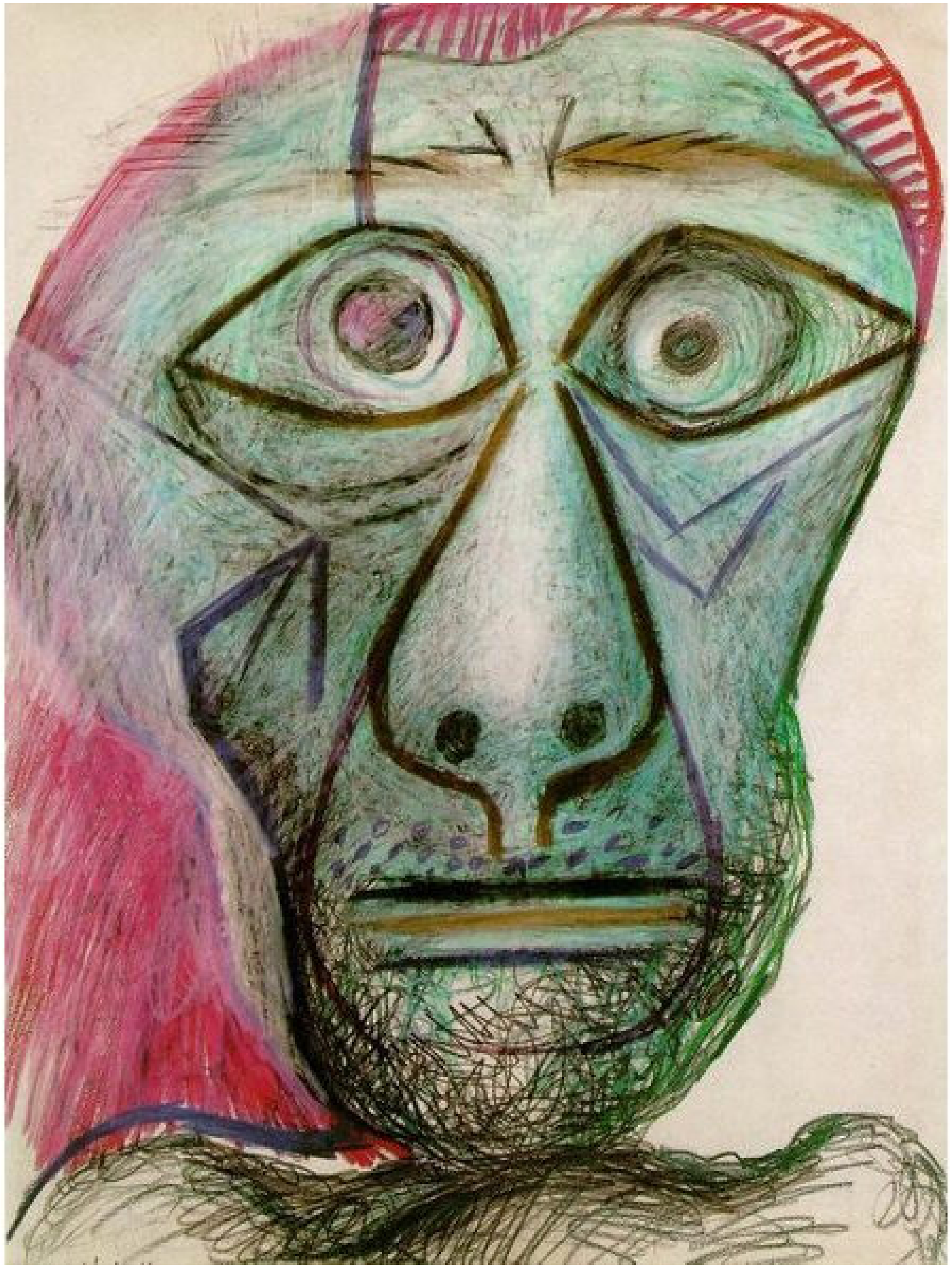


Figure.2. Pablo Picasso: Self-Portrait
Facing Death (1972)
http://arts.pallimed.org

ACKNOWLEDGEMENTS

I would like to express my most profound gratitude to my advisor Professor Nathan Felde for his generous assistance while scrutinizing and examining data, information, and theories throughout this study. I am also in debt for his patience, time, and guidance offered by his excellent advice.

Name and Signature:

Suyuan Ji

I would like to thank my friends: Shanshan Cao, Hayden.Sanders and Xiaoke Hua, who are willing to help me collect valuable data towards research. Finally, I would like to express my appreciation to my beloved parents for their support.

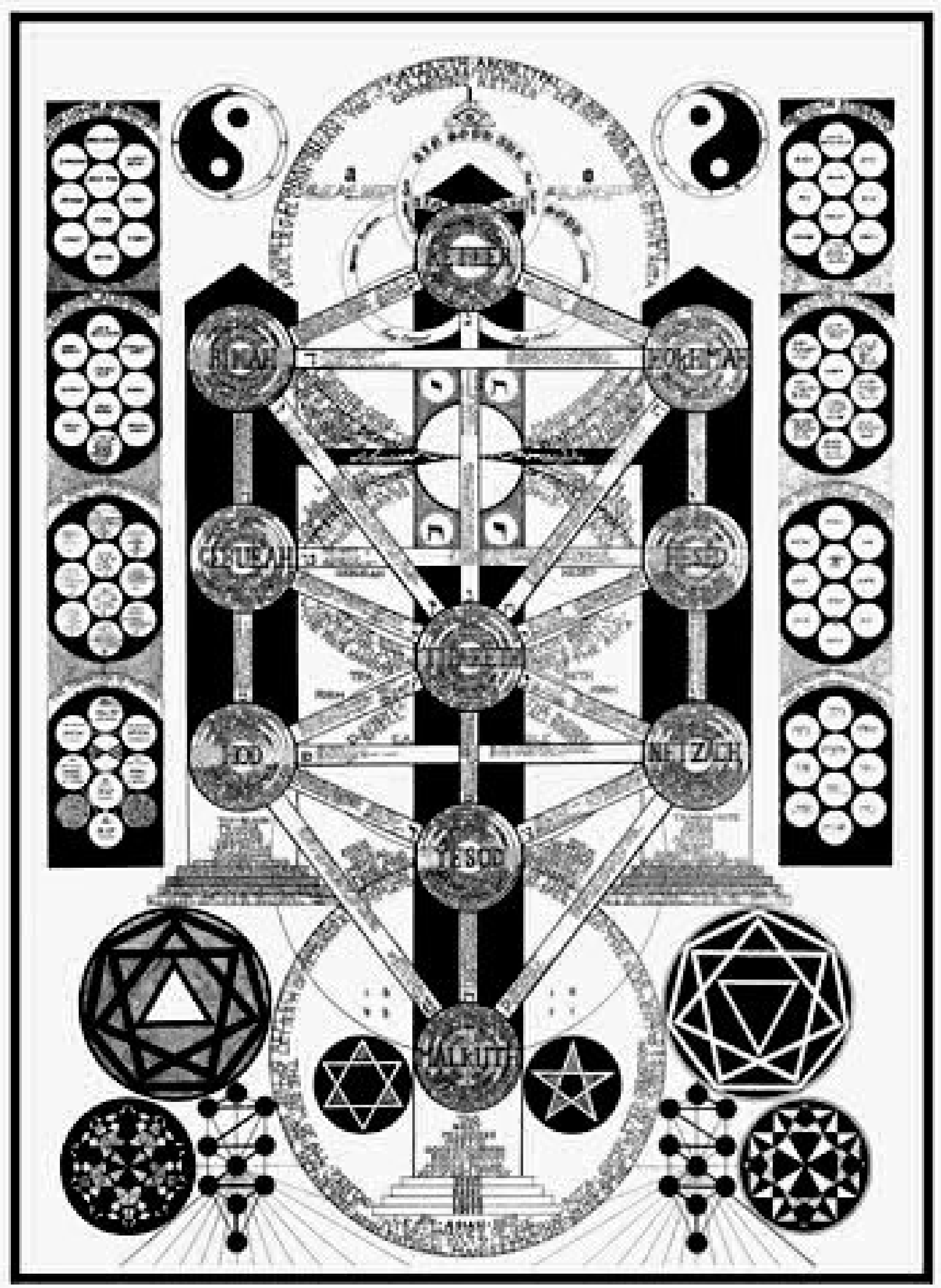


Figure.3. The Ten Sephiroth.The ancient symbol meant to explain the relationship between the material and nonmaterial worlds. https://zhuanlan.zhihu.com/p/26005176

INTRODUCTION

When people open a database, create a table, or add a few fields in the format of graphics, a multi-platform application with a modern-looking approach may be used for descriptive interpretation. That phenomenon is exactly what this paper intends to discern in the field of information visualization.

Accurately speaking, it is to create a model by using faces made of data to represent and analyze the dataset, which will be concerning the human face. In other words, it means that by using face parameters, a cluster of messages is conveyed by visualizing data and applying a code to present data by face elements.

Overview

(Appendix 1) aims to articulate the dimension between independent symbols and knowable as such, through human understanding (Eco, Umberto. 1986).

Therefore, considerations may come to the concept of abstract face, which means using faces as data to present collected information. It signifies that many kinds of facial features can be used to display different data. Just as the different emotions in the face convey different information, the faces made of data will also become a tool for representing different meanings.

Through this thesis, some cases will be used to prove accuracy and versatility. The biological components of the human face, such as the mouth and ears, which are represented by data in the abstract face, will be studied to be imitated better in the model. For example, the organ that is most prominent in the human face can match with the most crucial data.

However, many previous researches focus on using an individual face to present data in specific way. For example, crying faces represent sadness; smiling faces represent happiness. This example is the principal cornerstone of how faces of data can represent holistic meaning for datasets rather than have an individual purpose. The whole face is unique configuration, not simply a set of parts.

The Motivation and Purpose of Study

Few previous studies open up this theme to consider how much data a face can carry; it may make great arguments broader and more concrete of how to present information by each biological component on the human face. This paper will be of great significance to the research in this field by showing how the cluster of data can be classified to convey information that highlights on a selected category from any number of quantifiers provided. It no longer focuses on emotional effects regarding facial expression shown on the faces. This thesis will make up for the gap in literature through exploration of the relationship between facial expressions and symbolic forms to determine the relevant cultural implications of facial recognition. Meanwhile, the results of the paper may indicate correlations between human perceptions and facial expressions from semiotic perspectives. Biological components of a face present more significance in the way that the data is transmitted. Ultimately, signs such as expression of a face can add new shades of connotation to the graphic data.

Therefore, I will be studying the significance of data made of a face mainly by exploring how much data a face can carry and how a face can be a visual tool to read multidimensional data. It can be more specific to show the correspondence between faces made of data and dimensions in the dataset. In doing so, it can demonstrate how to establish the measurements of a face. For example, how to fit into the shape of head, how to measure the chin, and how to identify the length of nose or eyes of the faces, as the criterion of a face becomes the vital issue to express different aspects of data appropriately. In a word, the final map has certain features to determine the priority we offer to different kinds of data regarding various facial elements ,such as the mask and biological components' configuration.

Roughly speaking, this paper focuses on solving and answering the following questions:

1. How much data can a face carry?

2. How can a "DataFace" visualize multi-dimensional data?

3. From a semiotic perspective, how can one build up relations between signs and meaning in terms of an abstract facial indication?

CHAPTER 1: State of the Art

Research Background and Context

Contemporary science is seeing the emergence of a new data economy with data as its fundamental unit of exchange (Beynon. D, 2002). While using data in the digital age provides many potential advantages to suitably present individual or aggregate data, there are important social and ethical concerns evident. For example, the dataset for crime rates in Massachusetts (Appendix 1) could be presented in the form of bar chart or histogram. However, with the ongoing development of information technology, many researchers consider human facial expression recognition as a more valuable and relevant subject. Multivariate data in the elements a human face is usually regarded as an adequate tool to present psychological and emotional factors. The individual parts of a face (such as eyes, ears, mouth and nose) are combined to know a person and identify a human.

Presently, the idea of using faces explores recognition methods to identify human faces, especially regarding the perception of minor changes. It is because the features of the face can be different in perceived importance. However, the characteristics of the faces vary; what makes variables hard to be mapped is that biological components of faces have been found to carry various significant weight. Many experts start combining presentation of the development of data from human beings with the submission of an ethnographic, demographic, and geographic factor. They begin using the data of the faces to demonstrate results from dataset analysis flexibly. This thesis develops DataFace to formulate and produce an application of a "norm" for a dataset that is practical, flexible, and operational in various research fields. DataFace could accurately represent results derived from any standard dataset to make sure all data is physically pooled and presented within the multivariable data of the human face.

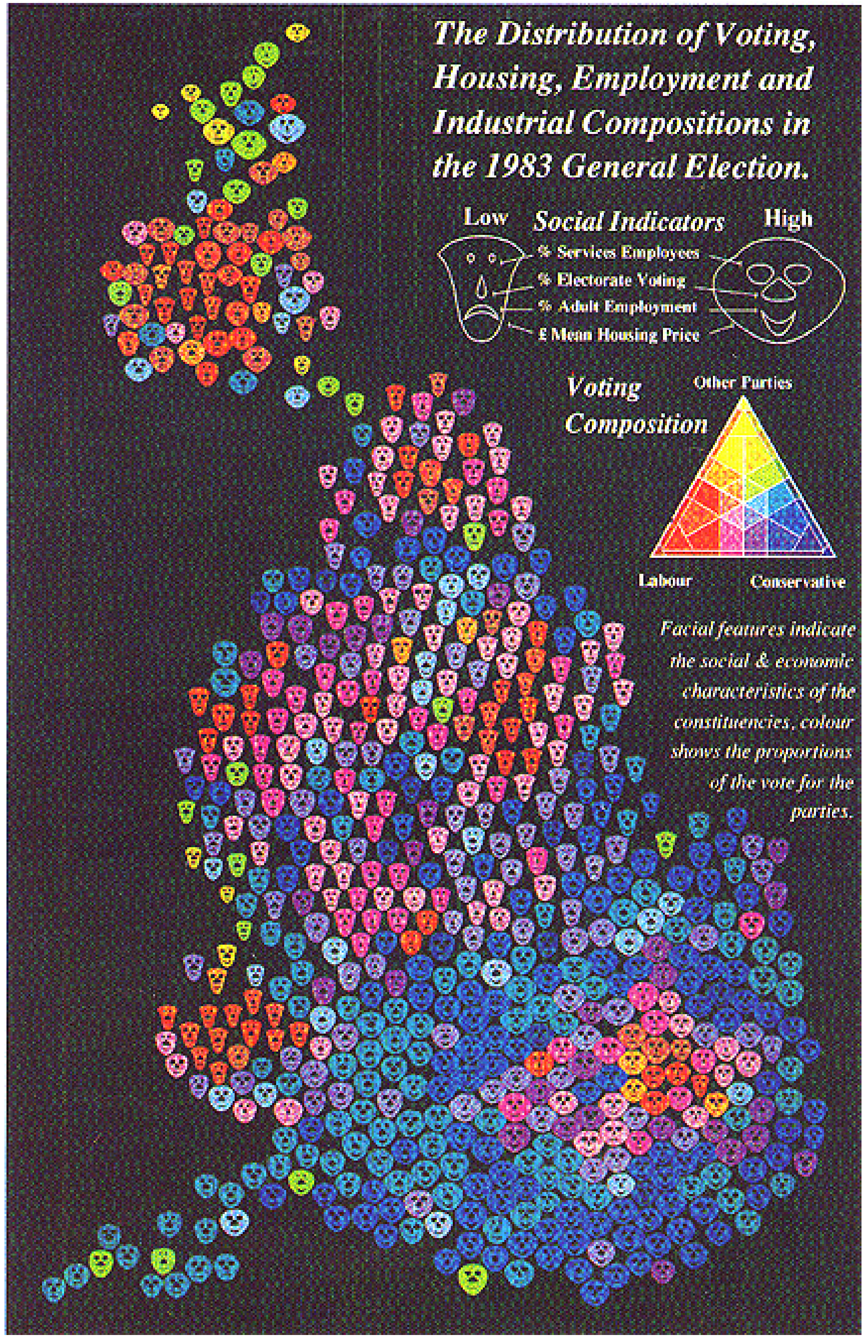


Figure.4. One by Daniel Dorling, 1995
https://cartastrophe.wordpress.com

1.1 CHERNOFF FACES

Herman Chernoff created Chernoff faces in 1973 and opened the door for researchers to examine faces and recognition of minor changes. The goal of Chernoff faces is to show numerous variables at once via facial features like lips, eyes, and nose size (Herman Chernoff, 1973). The features of Chernoff faces display multiple variables at once by positioning parts of the human face to recognize small differences when they represent data.

Describing data with faces rather than by abstract geometric shapes, Chernoff faces can be easily identified. Parts of the faces, such as ears, hair, eyes, and nose, are based on numbers in a dataset. The variables mapped to the features can be plotted on a standard X-Y graph, and then the faces themselves represent the rest of the dimensions for each item. (Fig.5.)

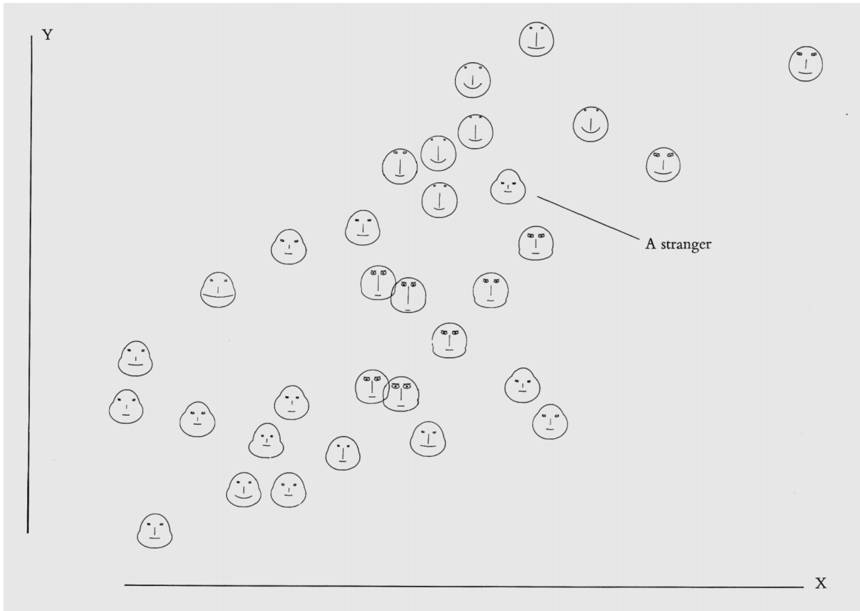


Figure.5. Tufte, E. R. (2001). The visual display of quantitative information. Cheshire, Conn: Graphics Press. Chicago (Author-Date, 15th ed.)
https://www.quora.com/What-are-the-best-examples-of-Chernoff-faces-used-to-visualize-data

Chernoff faces have successfully added new shades of connotation to explain the relationship between semiotics and significance. It also creates the process of transferring data and establishing meaning from a source to a receiver.

Moreover, its priorities are to study signification first, and communication second as it encompasses signs in some medium and sensory modality throughout the Chernoff faces' application of semiotics. Therefore, many concepts can be shared and, in each field, it can be emphasized differently.

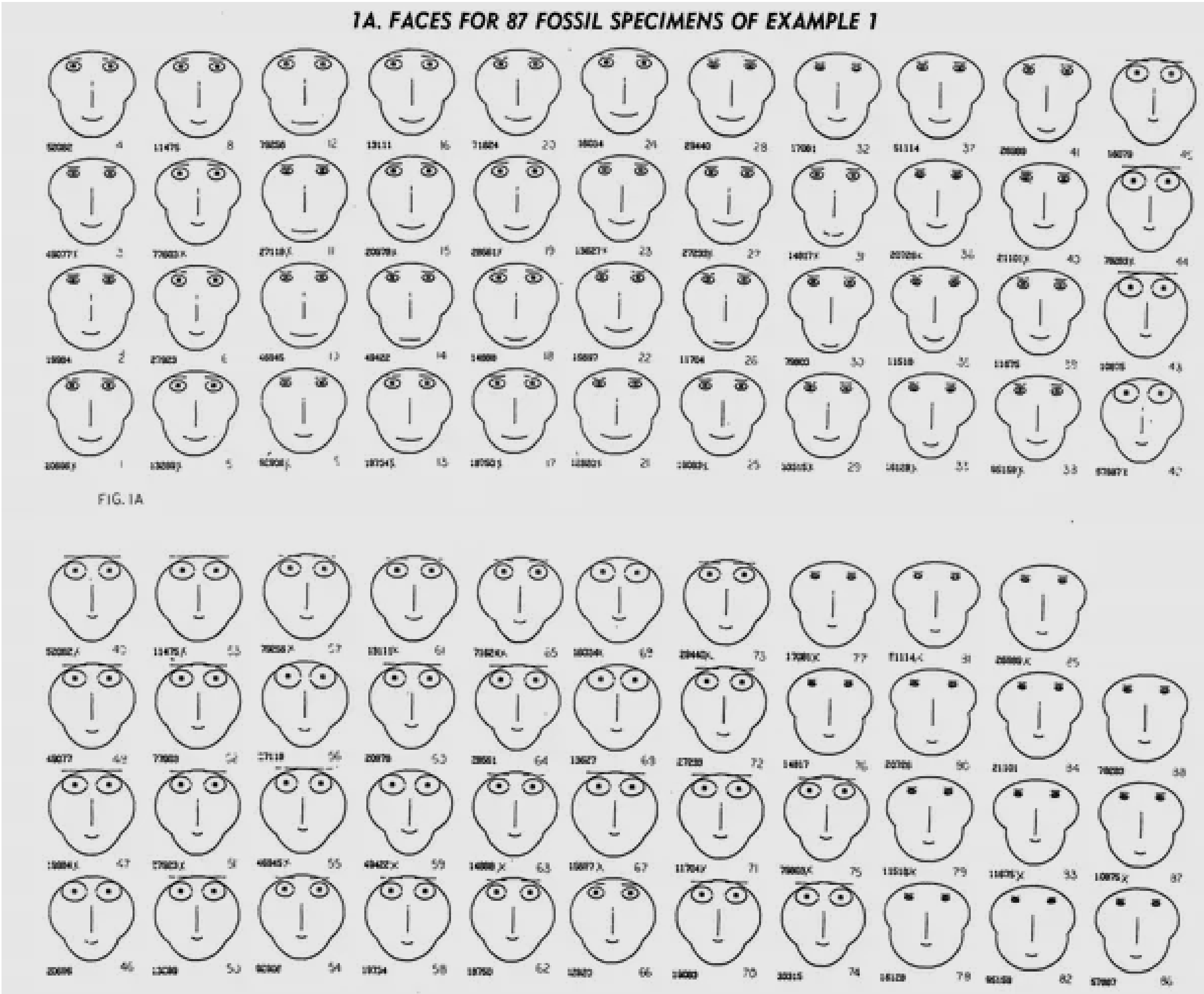


Figure.6. Chernoff, Herman. "The Use of Faces to Represent Points in K-Dimensional Space Graphically." Journal of the American Statistical Association 68.342 (1973): 361.
https://www.quora.com/What-are-the-best-examples-of-Chernoff-faces-used-to-visualize-data

Chernoff faces "reduce well, maintaining legibility even with individual areas of .05 square inches as shown...with cartoon faces and even numbers becoming data measures, we would appear to have reached the limit of a graphical economy of presentation, imagination, and let it be admitted, eccentricity" (Herman Chernoff, 1973).

It implies "asymmetrical" Chernoff faces create the practice of motif classification in images, using iconography as a manner of understanding meaning. The vertical symmetry of a face has been transformed to identify the 18 variables that specify the left being one set of data. With an iconographic arrangement, Chernoff faces "use a different set of data for the right side of the face, allowing one face to depict 36 different measurements".

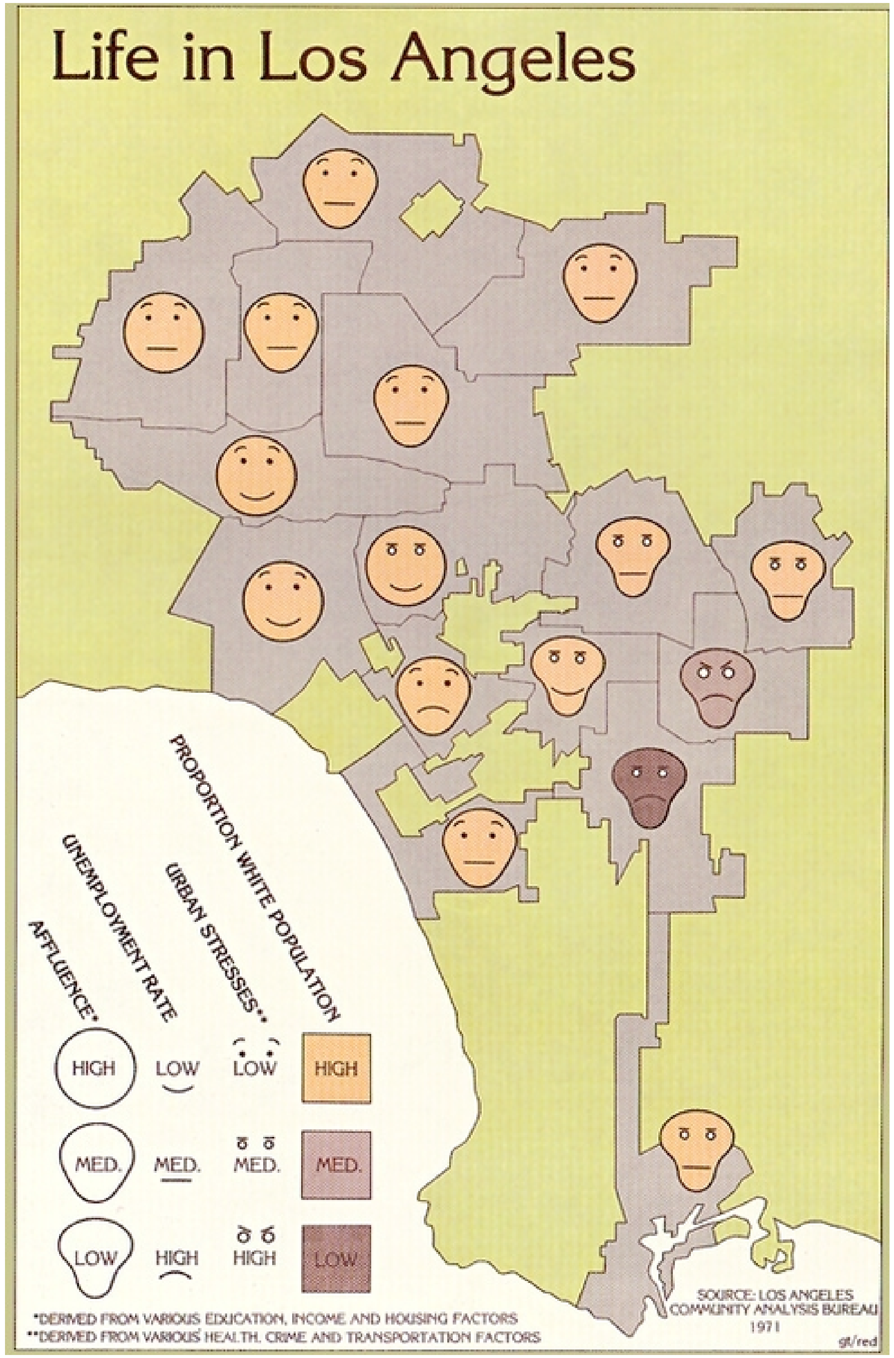


Figure.7. Eugene Turner - Life in Los Angeles (1977)
https://cartastrophe.wordpress.com

1.2 FACE RECOGNITION

The dataset of Families In the Wild (FIW) uses the observation research method to explore facial recognition among visual kinship relationship. (Fig.8)The sample groups consist of 200 family images in complex hierarchical family relationships in order to compare with the existing kinship image datasets. The multiple variables include facial features, biological components comparison, and genealogical factors. The researchers use advanced information technology to conduct the automatic photo library management and genealogical analysis, which is applied in social media. The results are obtained once the statistical analysis is completed. The findings show that given sufficient training data and appropriate training scheme, both the kinship verification and family recognition can improve. The advantage of FIW is that the research method used in the investigation is suitable to collect primary data and information among the sampling groups, throughout which the researchers can use the effective platform to distinguish the kinship relations in terms of a similar feature in facial expression.

Figure.8. Families in the Wild.
https://www.northeastern.edu/rise/presentations/families-in-the-wild-large-scale-image-database-and-evaluations/

Moreover, all data from the images selected in the investigation has been processed by the Smile app as metadata labeled with a name and number for each image. The labeling tool can help the researchers recognize each biological components of a face so as to compare the kinship among the images. The training process uses fine-tuned VGG-Face model to conduct the visual recognition, which can achieve better results in facial identification. It can focus on several variables such as eyes size and mouth shape and targets the objectives to conduct the exact measurement. To reduce the influence of object measurement accuracy, the researchers use visualization app to adjust the size of images and assure the feature dimensions are clearly expressed.

In my thesis, I consider using this method for measuring the value in facial expression during the empirical investigation. The quantitative analysis can be conducted with the numbering value measured by the visualization app, which is more accurate than a manual operation.

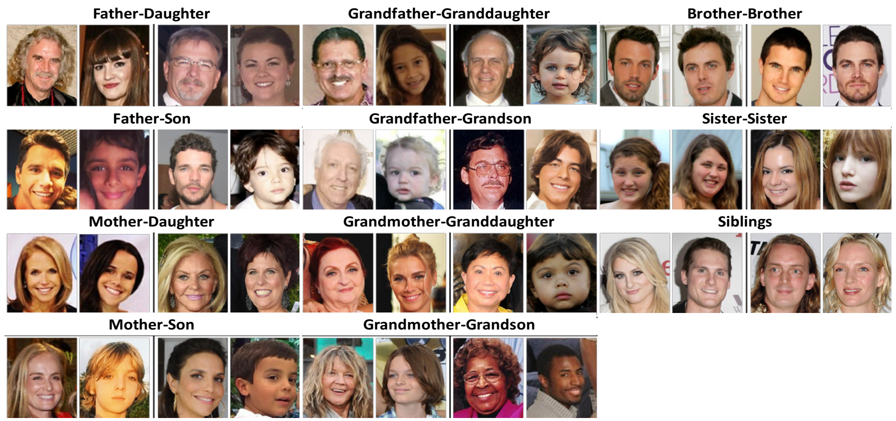


Fig 2 Sample pairs for the 11 kinship relations provided by the FIW database.

Table 2 No. pairs in FIW and other kinship image collections.

Pair-Type	KFW-II	Sibling Face	Group Face	Family 101	FIW (Ours)
Brother-Brother	--	232	40	--	**35,000**
Sister-Sister	--	211	32	--	**35,000**
Siblings	--	277	53	--	**60,000**
Father-Daughter	250	--	69	>147	**35,000**
Father-Son	250	--	69	>213	**35,000**
Mother-Daughter	787	2,589	—	Yes	**35,000**
Mother-Son	101	607	14,816	Yes	**35,000**
gFather-gDaughter	--	--	--	--	**300**
gFather-gSon	--	--	--	--	**300**
gMother-gDaughter	--	--	--	--	**300**
gMother-gSon	--	--	--	--	**300**
Total	1,000	720	395	> 607	**271,200**

Table 3 Comparison of FIW with related datasets.

Dataset	No. Family	No. People	No. Faces	Age Varies	Family Structure	Highlights
CornellKin [5]	150	300	300	No	No	Parent-child pairs.
UB KinFace-I [8]	90	180	270	Yes	No	Parent-child pairs. Parents' 139 images at various ages.
UB KinFace-II [8]	200	400	600	Yes	No	Parent-child pairs. Parents' 139 images at various ages.
KFW-I [6]	—	1,066	1,066	No	No	Parent-child pairs.
KFW-II [6]	—	2,000	2,000	No	No	Parent-child pairs.
TSKinFace [9]	787	2,589	—	Yes	Yes	Two parents-child pairs for tri-verification.
Family101 [7]	101	607	14,816	Yes	Yes	Family structured, variations in age and ethnicity.
FIW(Ours)	**1,000**	**10,676**	**26,725**	Yes	Yes	A corpus of 1k family trees that provides both depth and breadth and multi-task evaluation offerings.

Table 4 Verification accuracy scores for FIW.

	F-D	F-S	M-D	M-S	SIBS	B-B	S-S	GF-D	GF-S	GM-D	GM-S	Avg.
HOG	55.0	55.7	54.9	54.7	57.2	58.0	57.3	55.0	59.0	**66.3**	59.0	57.5
HOG PCA	56.6	57.2	56.8	56.1	59.5	60.6	59.2	54.0	59.0	65.3	60.0	58.6
VGGFace	63.5	62.6	66.0	63.0	72.1	70.7	70.5	68.9	69.1	63.4	**74.0**	67.6
VGGFace PCA	63.8	63.3	65.3	63.3	73.6	71.8	72.4	**71.4**	**69.7**	60.3	71.7	67.9
LBP	54.0	54.5	54.0	54.1	55.8	56.8	55.3	65.0	62.0	63.0	54.3	57.2
LBP PCA	55.0	55.8	56.0	55.6	57.7	58.3	56.6	69.3	66.3	64.7	58.3	59.4
Fine-Tuned	**65.0**	63.1	66.5	64.5	73.7	72.4	71.7	-	-	-	-	68.1
Fine-Tuned PCA	64.8	**63.7**	**66.9**	**64.8**	**74.4**	**72.7**	**72.9**	-	-	-	-	**68.6**

Table 5 Classification accuracy scores for FIW.

Fold	VGG-Face	VGG-Face (fine-tuned)
1	9.6	**10.9**
2	14.5	**14.8**
3	11.6	**12.5**
4	12.7	**14.8**
5	13.1	**13.5**
Avg.	12.3	**13.3**

Figure.9. Families in the wild.
https://www.northeastern.edu/rise/presentations/families-in-the-wild-large-scale-image-database-and-evaluations/

"Intuitive Thinking About Race" project developed an innovative method to explore the race biases perceived by instinctive thinking and implicit memory. The evaluation for instinctive thinking derived from the studying of essentialism that leads to the formation of stereotypes and prejudice. The primary data is collected from all participants' answers from their recognition towards the races shown in 60 images that contain three races: black, white and Asian. The racial biases from facial recognition show the gaps between the essentialism scale and implicit memory. in other words, the results indicate that the participants who make more facial recognition errors will have higher essentialism scores. Additionally, it shows that negative feelings will be associated with higher essentialist thinking as well as greater errors in facial recognition. The significance of this study is that facial recognition not only depends on the facial characteristics shown on the facial emotion, but also on the feelings perceived by the viewers (Fig. 10).

Finally, the researchers use the concept of implicit memory to compare the instinctive thinking because it seems to be a cognitive skill acquired through learning and knowing things that human beings may not know they have learned.

The essentialism scale can measure and probe which types of race have been activated by implicit memory. The findings of this research offer me a new idea that facial recognition could be greatly influenced by the viewers' perception due to cultural differences and other factors.

For example, a crying face may present when a disaster happened for a certain population, but represent less disaster for others, due to ideological gender differences in ideology.

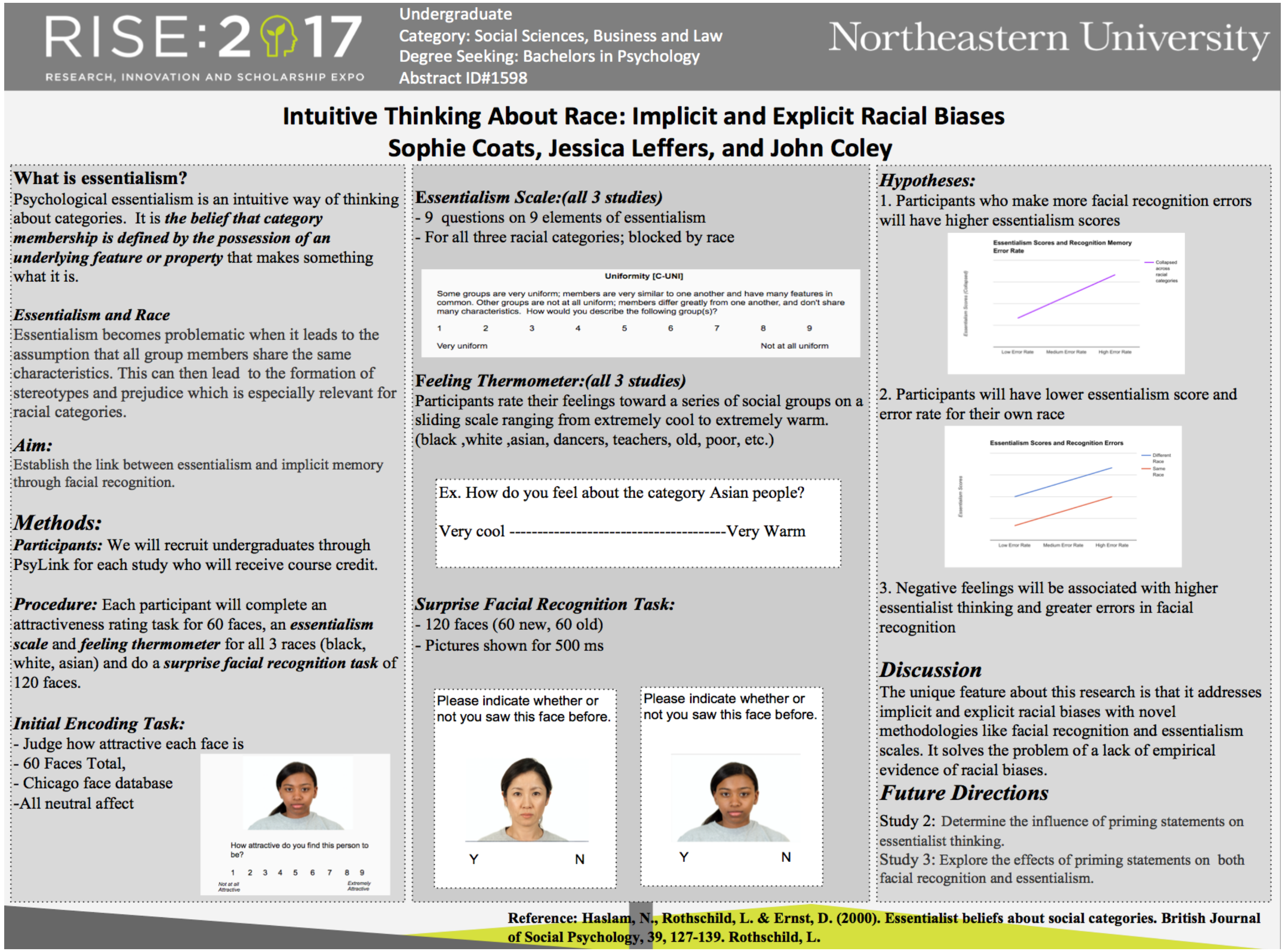
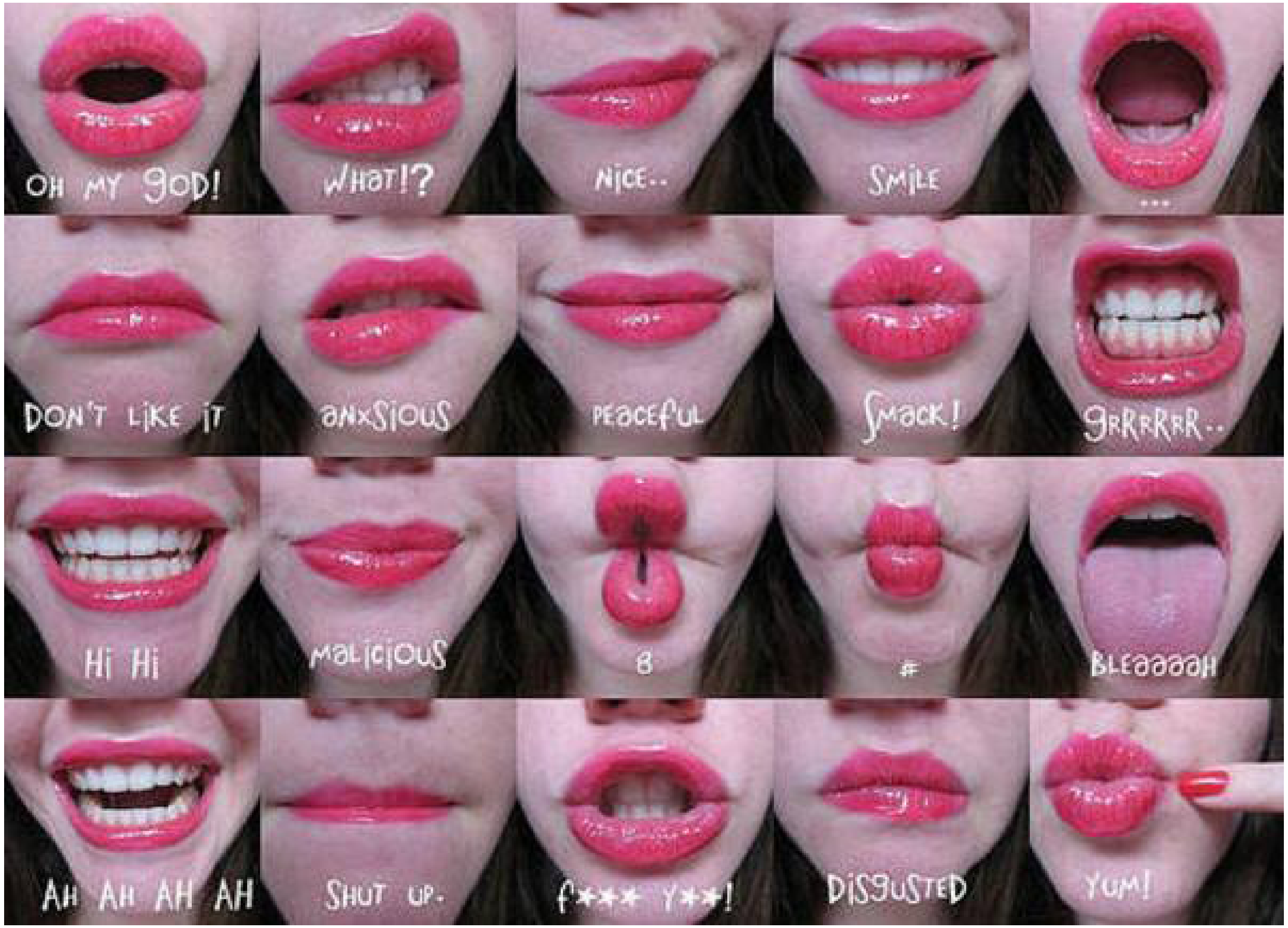


Figure.10. Intuitive Thinking About Race: Implicit and Explicit Racial Biases. https://www.northeastern.edu/rise/



For example, the mouth plays a key role in producing facial expressions and speech articulation. (Fig.11.)

Figure.11. Microexpression.
www.duitang.com

MICROEXPRESSION AND EMOTION

Microexpression

With this thesis, I attempt trace each feature of a face that is shown in a microexpression. Previous research has defined microexpression as an involuntary facial expression used to reveal all signs of how people are feeling that they may not consciously recognize (Haggard, E. A., & Isaacs, K. S., 1966).
Unlike regular facial expressions, microexpression reactions are hard to be controlled because it usually happens or lasts only 1⁄25 to 1⁄15 of a second, but it is true expression without any disguise of non-verbal communication (Haggard et al., 1966).

"Traditionally, seven emotions have been demonstrated in microexpression to show anger, disgust, fear, happiness, sadness, and surprise." Ekman, Paul (1992) created special rules to indicate the causal relations between culture-specific prescriptions and emotions that reflected on microexpression, explaining how cultural differences might conceal the universal effect of expression. Ekman's encoding scheme has been widely used in scientific research to judge the merits and faults of a person from primary emotions. However, it is argued that microexpression may contain mixture emotions such as disgust and anger that may not be classified solely as one variable.

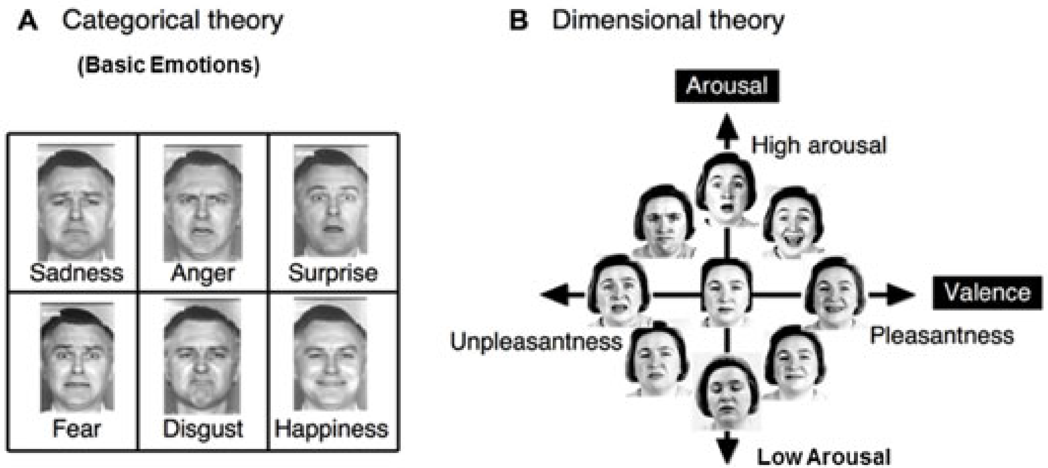


Figure.12. Contrast of the two theories.
James Russell was the first to suggest the circumplex model, which classifies emotion into two dimensions: valence and arousal4. Information about emotion classification is gained by pinpointing the reaction in this 3-dimensional space.
http://www.hcdi.net/face-value-theories-of-emotion-and-their-application-to-neuromarketing/

Ekman's encoding scheme has been expanded, with the range of emotions placed into two groups, positive and negative feelings that involve multiple variables. Amusement, embarrassment, anxiety, guilt, pride, relief, contentment, pleasure, and shame have been added to the list of emotions (Ekman, Paul, 1999). The significance of the coding system is to identify facial expression through a measurement of the muscle movement that produces the facial expressions. For example, the relaxation or contraction of each individual muscle and assigned unit can demonstrate the separate action of a face to show the degree of pain or happiness that people may not express themselves. In the age of information, this coding system is now being designed to detect and categorize facial movements to help researchers recognize different physiological attributes of facial expressions. (Fig.13.)

This thesis will leave the interpretation of this coding system but attempt to utilize this data to examine other works. In other words, as we cannot expect all audience members to be face-reading experts, we will apply facial movements with a coding system to explain the exact movements that the face can perform, as well as what parts of the face produces them. In this manner, the face coding system becomes a valuable tool for the interpretation of dataset, by which it can bridge the communication gap between different datasets and the expressions of a face.

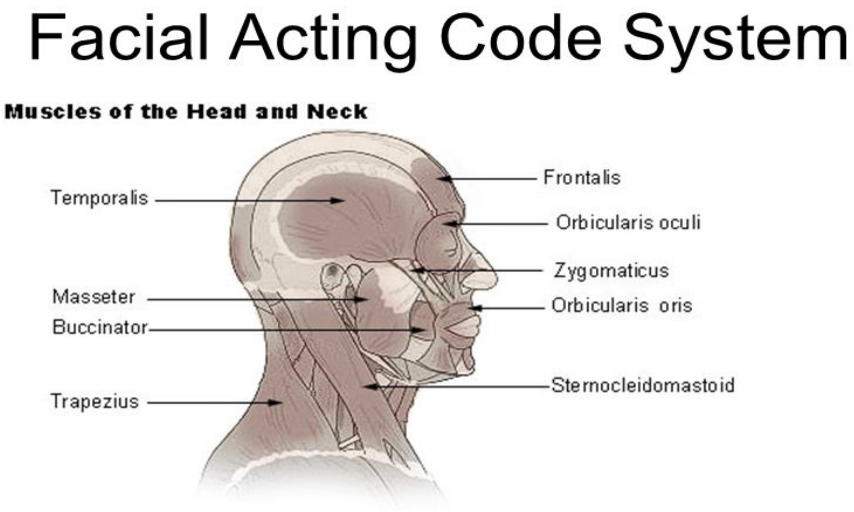


Figure.13. FACS example.
http://mplab.ucsd.edu/grants/project1/research/face-detection.html

1.Facial Acting Code System originally developed by Ekman and Friessen in '70s
2.Based on the physiology of facial muscles

More accurate representations of facial behaviors have also been used in the science of facial recognition. Data acquisition derived from facial recognition helps researchers capture essential information, such as a localization of missing people. However, computational face identification is still a challenging problem that it mainly relies on a large database to compare specific facial parts. It not only requires plenty of time to collect thousands of faces, but also requires highly sophisticated algorithms to obtain acceptable results. In addition, the way that people describe faces are different for the entire population. It may result in the phenomenon of the race bias due to the differences in the perception of facial features. In other words, it means that race or ethnicity may influence the accuracy of face verification.

I will use these facts to capitalize on the linguistic description of the face to make computational face recognition efficiency. Thusly, the specific facial regions and movement concerning the parts of a face have pivotal meaning in this study. For example, the size and color of the eye as important facial features shows the information contained differently in each culture or crime research (such as in response to the category of larceny, robbery, and rape). Nonimportant facial features will be excluded from the classification process. However, as how to modify our data of a face to fit the social setting is still a problem in this paper, I attempt to use six innate facial expressions to display all variables in a varying dataset. Nevertheless, the challenge is that an isolated preliterate figure in a sequence diagram may not fully reveal the taxonomy of dataset.

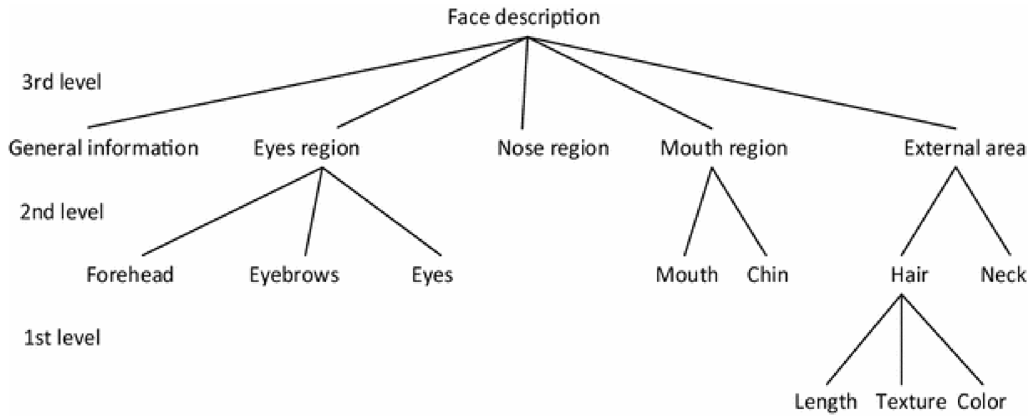


Figure.14. Three-level AHP realized for the selected features. For the clarity of presentation, only the part regarding to the hair from the first AHP level is depicted. Similarly, only a few chosen features of the second level are presented. https://link.springer.com/article/10.1007/s00500-016-2305-9

In relation to classification, Paweł Karczmarek (2016) developed a process of estimating the importance of features considered in face recognition by making use of the analytic hierarchy process (AHP) (Fig. 14). This method provides an effective approach to comparing the facial feature at three levels of hierarchy, each of which containing an individual meaning that can be consequently formed. We may use this method to realize a comprehensive weighting of social meanings. For example, raised brows and colorful eye may indicate an increase in value.

Emotion

The "Emotional Face Perception is Altered Depending on the Presence of Absence of Context" project also highlights the gaps between facial recognition and perceiving emotion. To distinguish the facial expression portrayed by descriptive genre and the facial characteristics exposure to stimuli, 52 participants were asked to rate of faces and situations shown in the images (which played by actors). Seven emotional words (anger, fear, sadness, disgust, neutral, surprise, and happiness) were defined as multiple independent variables. The results show that that ratings of faces and situations together were more similar when participants selected from a list of 7 emotion. However, the significant differences are that nearly 50% of faces showed a “categorical” shift after partnership with context. This empirical investigation shows that there are casual relations between the face recognition and human beings perception of facial characteristics. The comparison research method has been used in this investigation throughout dividing sampling groups into two: one block refers to participants’ viewing the face alone, and another block refers to perceivers’ viewing facial characteristics with the help of context information. It is very useful to conduct the quantitative analysis once the primary data has been collected. (Fig.15.)

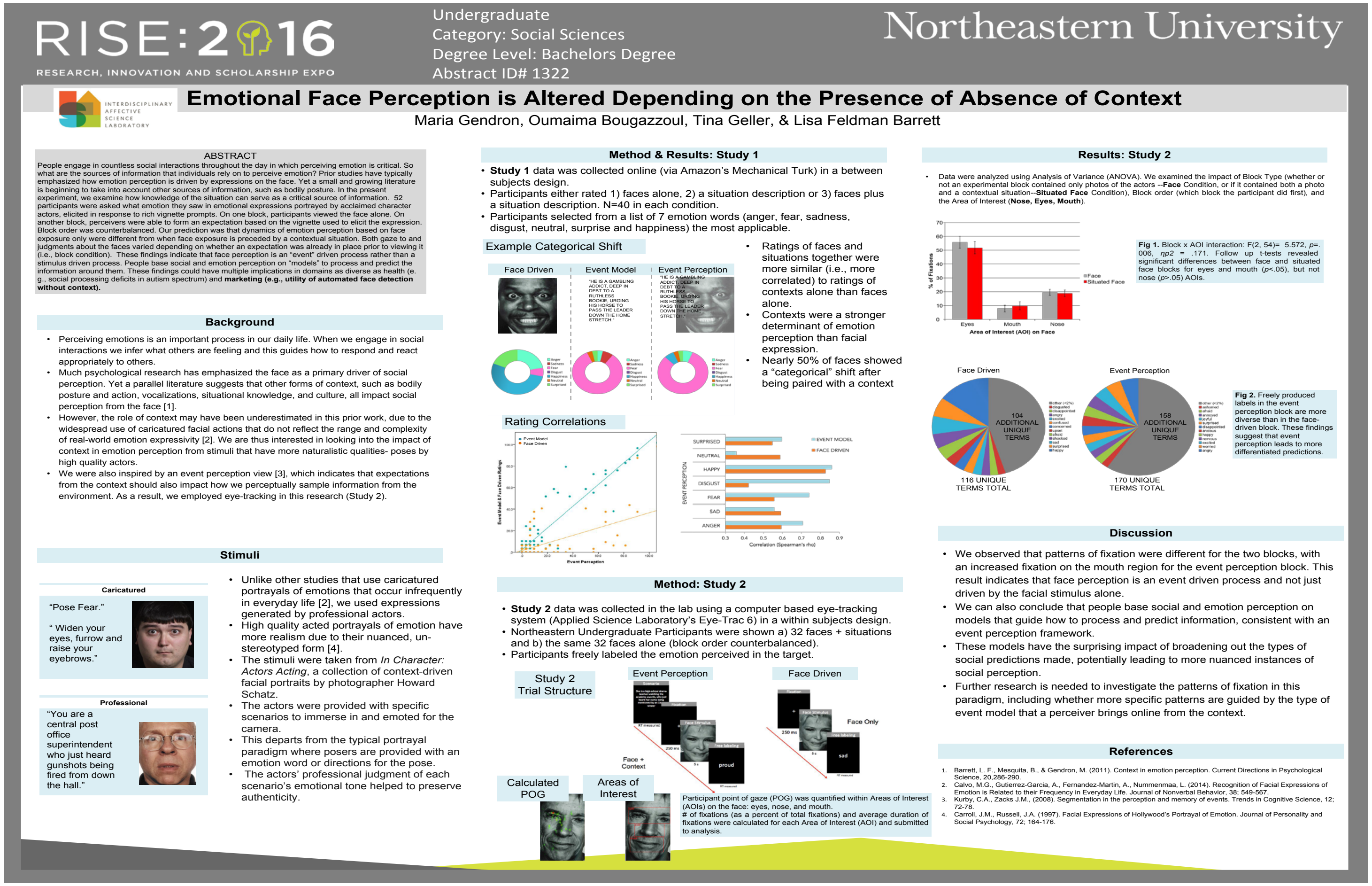


Figure.15. Emotional Face Perception is Altered Depending on the Presence of Absence of Context.
https://www.northeastern.edu/rise/

Moreover, this study has used a computer-based eye-tracking system to collect data. Participant point of gaze (POG) help measures multiple variables on the face: eyes, nose, and mouth. This generates quantitative data for further analysis, which can help researchers obtain more validity and increase the value of results.

To further learning, I would consider using this software and system to support my data analysis in the measurement of facial recognition.

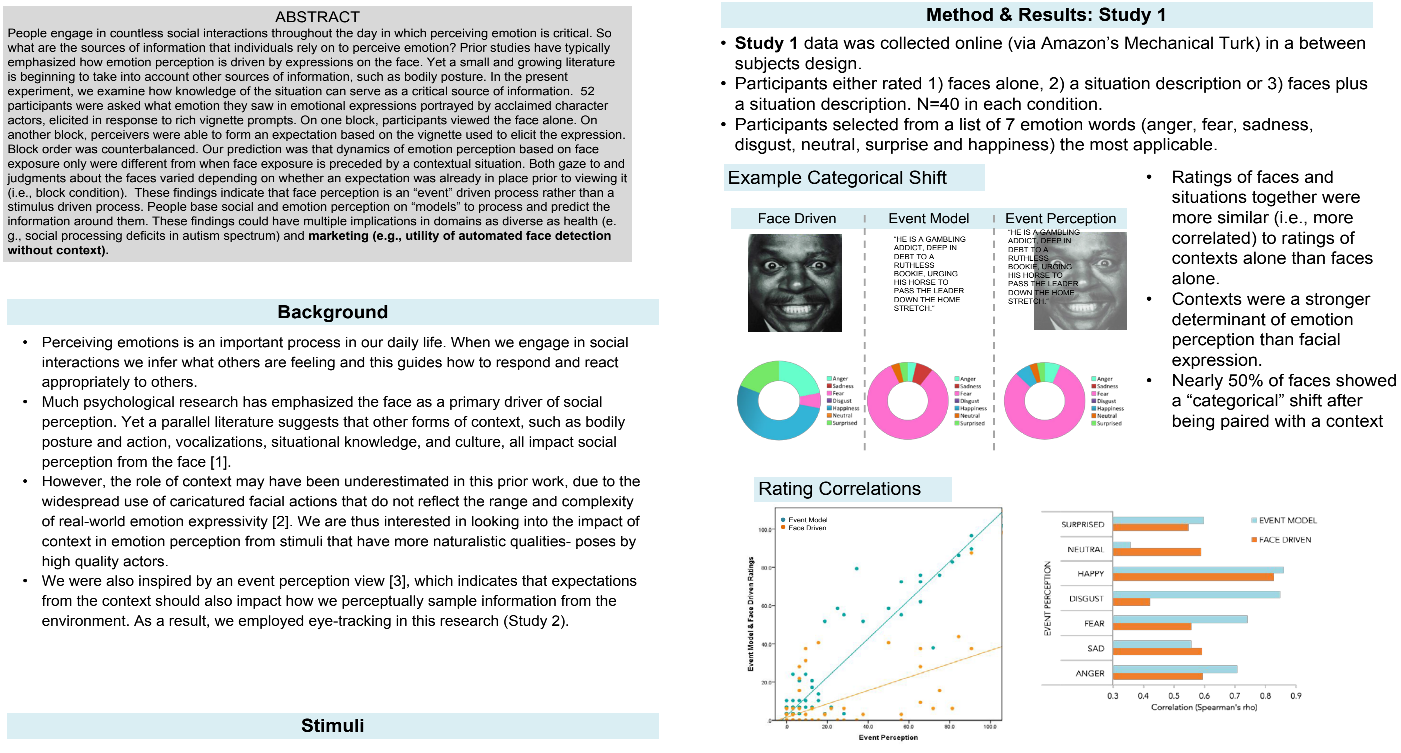


Figure.16. Emotional Face Perception is Altered Depending on the Presence of Absence of Context. https://www.northeastern.edu/rise/

EMOJI

The field of graphical interaction between human being and machine allows for designers to draw points, lines, arcs, etc. in for a digital presentation of varying symbolic language. ASCII art is a graphic design technique with a total of 95 printable characters to be created with any text editor. Therefore, it is usually considered the most common free-form language utilized for the creation of text pictures with text symbols. In the age of information, both designers and users always encounter those ASCII-painted pics somewhere on Internet and they are widely used in place of graphic marks. Initially, ASCII art was for creating visual art in typewriters to highlight advertising effects on newspapers or handouts. Later, became useful for assisting in the printing of large banners, or an e-mail when imbedding an image was not possible.

However, the widespread usage of ASCII art started the transformation of communication in bulletin board systems (BBS) from one-way to two-way communication. The use of text characters to represent images "began to appear in the underground online art groups of the period." It did not become increasingly popular until the 1990s because the techniques of the fixed-width font were replaced by variable-width fonts; this applies incrementally in graphical browsing and "Multi-User Dungeon."

In modern society, ASCII art can be more readily used in graphics works where the transmission of pictures is not possible, as well as "used within the source code of computer programs for representation of a company or product logos, and flow control or other diagrams" (Fig.17.). It means that ASCII art can generate more images and visual text in terms of specific code. For example, the entire source code of a program is a piece of ASCII art that can enter the C language to yield more visual text and images. Moreover, ASCII art can be used in 2D platform multiplayer shooter game such as "0verkill" in terms of linking with the AAlib library.

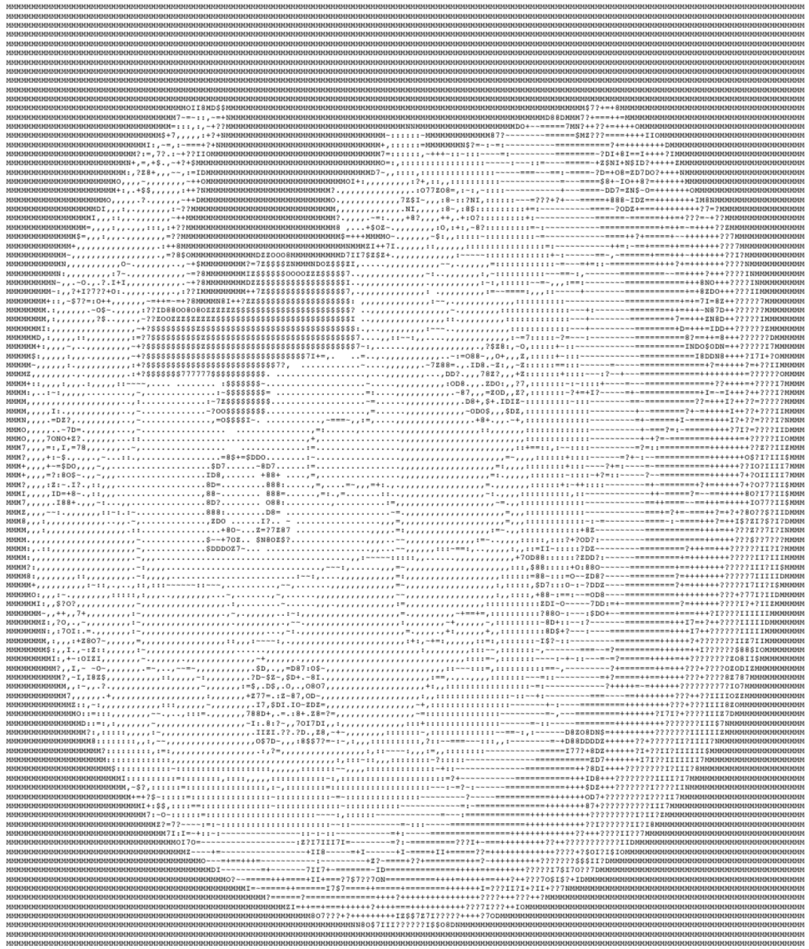


Figure.17. ASCII Art of Wikipedia Logo
https://en.wikipedia.org/wiki/ASCII_art#/media/File:Wikipedia-Ascii.png

Therefore, different techniques could be used in ASCII art, regarding various forms of individual letter characters and visual features, to obtain different artistic effects. Notably, it can combine two or more characters to express emotion in-text. Traditionally, some "face characters" were derived from one-line ASCII art to present an 'emoticon,' 'smile,' or 'smiley' in animation design (such as kao-moji). “More complex examples use several lines of text to draw large symbols or more complex figures.” They are derived from "ASCII face" (Fig. 18.).

From a historical perspective, many designers took inspiration from natural phenomena such as weather forecasts that used symbols to show weather, or animals. Emoji is the typical semiotic information used in electronic messages and Web pages. It carries many iconic images like emoticons to show specific expressions—weather, human emotion, and other subjects and objects. In the age of information, it becomes the most popular application used in various computer and mobile operation systems. The first set of 176 12×12-pixel emoji was created as part of i-mode's messaging features to help facilitate electronic communication and serve as a distinguishing characteristic from other services(Fig. 19).

crying	(;_;) (T_T) (ToT) (T0T) (>_<) (/_ ;)
bowing/I’m sorry/please	m(_ _)m <(_ _)> m(._.)m ≦(_)≧Gomen
running away	...(・・) ((((((^^;) ・・・(((^__^) ((へ^^)へ にげろ！
memo	φ(. .)めもめも ¢(−−;memo. φ(・_・)memo
embarrassed/ surprised	(*_*) (@_@) (+_+) (*o*) (?_?) (=_=) (−_−) (゜_゜;)
perverse	(・・)(_ _)
I can’t hear you.	≪(−_−) ≫kikoenai.
smoking	(−.−)y−。oO (−.−;)y−~~~ (−。−)y− −t(^o^)。o 0
sleeping	(−_−)zzz (−.−)Zzzzz.... (_ _)。oO
I’m sleepy./yawning	('o')。oOO (−_ゞごしごし (~0~)ねむい ヽ(^O^)yawn ＼(~o~)／yawn

Figure.18. Kao-moji(Japanese Emoticons)
http://www2.tokai.or.jp/yuki/kaomoji/

Initially, 180 emoji had been created based on human being facial expressions observed by Japanese designers to represent emotion and other aspects of life. By the end of 2015, hundreds and thousands of emoji were introduced to the market; containing many implications and significances in terms of encoding.

All emoji characters were encoded in the Unicode Standard in version 6.0 released in October 2010 (and in the related international standard ISO/IEC 10646). With more countries participating in ISO/IEC JTC1/SC2/WG2, various new characters were added during the consensus-building process. It allows emoji to be able to apply in many fields and areas outside of Japan. Emoji characters vary slightly between platforms within the limits in meaning defined by the Unicode specification, as companies have tried to provide artistic presentations of ideas and objects (Negishi, Mayumi, 2015). However, designers try to explore more culture-specific meanings with specific emoji which were not displayed in the original glyphs.

For example, some emoji have been utilized in English-language communities rather than in other languages, which means designers can use them to show different social meaning conveyed by original semiotic information. Negishi, Mayumi, (2015) explained that Unicode manuals sometimes provide notes on auxiliary meanings of an object to guide designers on how emoji may be used.

In practice, most designers attempt to explore more above the original information than expected to broaden the scope of their design work.

Figure.19. The set includes 176 images created with just six colors on a 12x12 pixel grid for Japanese pagers in 1999.http://www.theverge.com/2016/10/26/13424976/moma-nyc-art-original-emoji-acquisition

1.5 MASK AND FACE RECOGNITIONS

There is a gap between human facial expression and perceived emotion. It is because people usually engage in countless social interactions throughout the day, but they may be reluctant to show their real emotion to others. It means that the sources of information perceived by human beings may not be consistent with the consequential facial expressions. Prior studies explained that human beings may conceal their facial emotion purposely at the mean time; meanwhile, their body posture may expose the emotions that their facial expression opposes. It seems like people wear "the mask" to cover up their social pressures, abuse, or harassment. Masking could be impacted greatly by social and cultural environmental factors, such as authoritarian parents, rejection, and emotional or sexual abuse. However, people may not notice that they are wearing a mask under certain behavior. Traditionally, masking is closely related to concealing a negative emotion based on some contextual factors because some emotions such as anger, jealousy, or rage, would not be considered socially acceptable.

In general, masking negative emotions differ under varying factors such as gender and ethnicity.

For example, it is generally assumed that girls must act nicer than men, so that female tends to mask their negative emotions more often than males. Masking also differs due to cultural differences. In other words, the effects of masking one's negative emotions are different from culture to culture. Ultimately, the dynamics of emotional perception based on facial exposure may be compromised by masking the emotion. Thus, the judgments about the faces varied, not only depending on whether an expectation was already in place but also in masking condition.

In my thesis, the negative and positive emotions should be considered based on a range of factors to obtain an adequate external response. In the face experiment, data collection from facial expression should pay more attention to the effects of negative emotion, which may be perceived

by people in different understanding. Realistically, how to express emotions and feelings can be displayed in various forms of semiotics. For example, masks occur throughout the world, which tends to share many characteristics and significances. It is usually related to ritual, magical, cultural and religious throughout providing various formations of biological parts of a face.

Ritual masks, as representations of the human face, are extremely revealing of the two fundamental aspects of the human psychological condition: "firstly, the repression of a cooperative, instinctive self or soul; and secondly, the extremely angry state of the unjustly condemned conscious thinking egocentric intellect" (Caplan, Jane, 2000). It means that mask is to present significance in a special communicative way, representing identification, social status, or other social nuances.

Similar to a mask, a tattoo is also a form of body or facial modification to represent symbolisms and emotion. For example, the tattoo on Tyson's face his facial tattoo becomes indelible and a permanent representation of his social status and identity, which makes him distinct from others (Appendix 3). With a unique symbol added to the face, such as mask or tattoo, semiotics have expanded their original meanings to indicate particular meaning, significance, and feeling. Thus, biological parts of a face can be conveyed by organic components in various forms of abstract faces for identification in unique ways.

CHAPTER 2: Conceptual Framing

When I selected my research topic, I was deeply inspired by the data matrix used in the Chernoff method to present social meanings regarding a facial feature (like a shape of the nose or shape of the eyes). Each variable is to correspond with a special motion in a face, referring to map data from a visual structure. Most importantly, Chernoff method was arbitrary to present a face that had no correspondence to the underlying semantics of the data (Chernoff, H. 1971). It created an effective way to understand pattern, although individual faces and facial components do not readily convey anything about the data in the specific mapping used. However, the empathic visualization algorithm (EVA) has been developed through the extension of the type of data visualization by Chernoff, which encouraged many researchers exploiting the idea to understand faces to interpret information encoded into facial features.

EVA focuses on creating an automatic mapping from semantically key features of the data to present psychologically crucial features of the corresponding visual structure, such as a face. Many researchers have developed further based on EVA to represent certain meanings and states, without knowing any details of what the facial expressions represented. For example, given a specific data matrix from observations on variables, the mapping from data can be gleaned from the faces by knowing of measurement of a component part on the face. The measurements can be size, shape or other elements (Fig. 20).

Figure.20. The empathic visualization algorithm (EVA) .

However, a single glance at the visual structure has mainly highlighted emotional elements corresponding to the variables of the underlying data analysis. Concerning interpretation of the visual structure, I attempt to explore cluster Chernoff faces by grouping in other types of forms that represent similarities in the data abstracted from original dataset and through knowledge of the relationships between data features and significant aspects of the visual structure.

This visual structure is defined as DataFace to represent and analyze the data by knowing about the component parts of the face. For example, frowning represents high crime rate, blinking eye represents changes of crime rate and so on.

The significance of this research is that most previous studies use relevant facial elements to express emotional effects. This paper attempts to exploit the mapping structure by defining the variables of components parts on the face. The causal relations between biological factors and abstract face will be combined to present socially conditioned, cultural, and technological trends, in many fields. The features of biological factors can represent various facial expressions such as frowning and turning up the mouth, by which the symbolic forms determines the relevant social meanings of dataset.

It is assumed that DataFace is to be represented by a visual structure, by which people can immediately and transparently interpret the meaning of this structure at a global level (Fig. 21). Given from naturalistic elements, a human face can present numerous global characteristics that can describe the demographic, geographic, or economic factors, all in the expressions of a face.

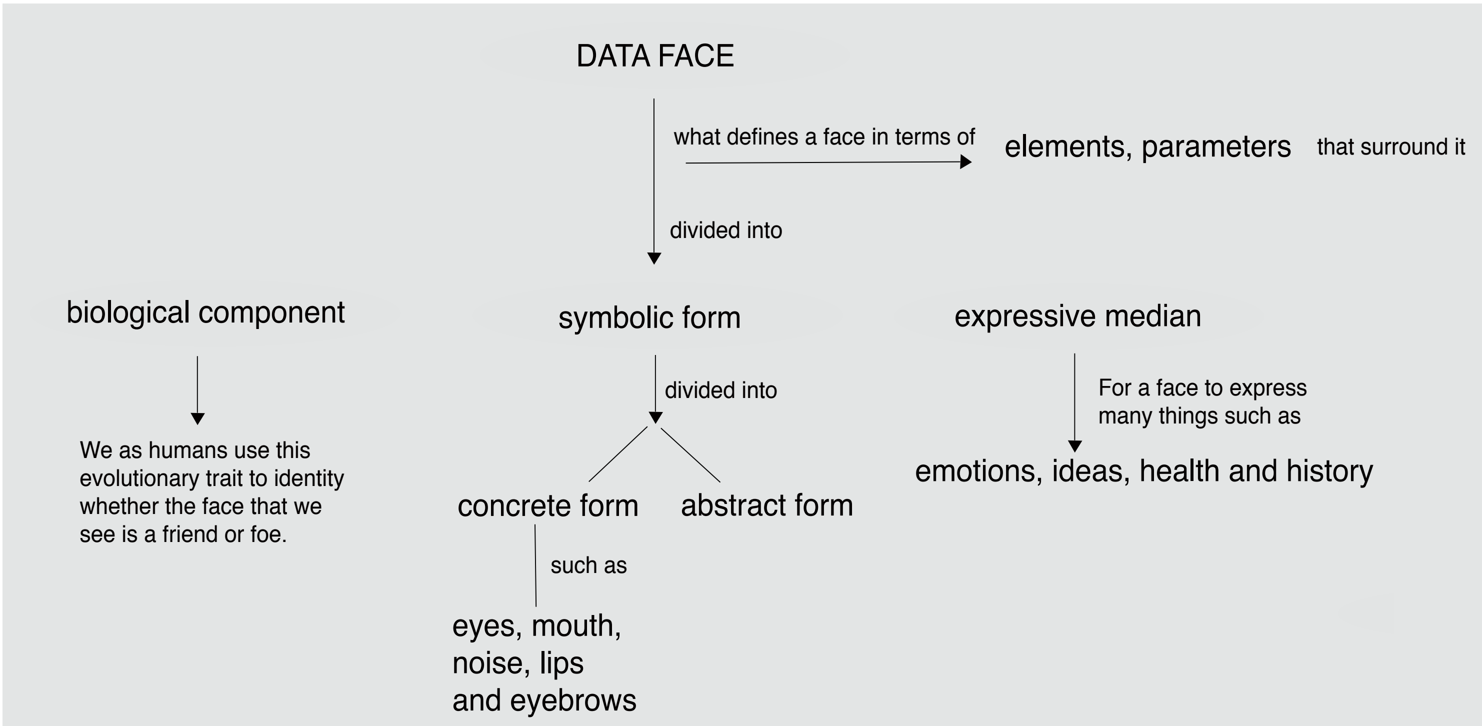


Figure.21. DataFace conceptual model.

DATAFACE SYSTEM MAP

The face has these elements, and these elements have variables. A face can be the biological fact to an image of a face, a representation of an image. It can go from a physical scan to the most abstract.

5 basics expressions establish micro- expression and expression. If we took the face, the face can have an expression, the expression comes from face's elements. The elements consist of eyes, nose, cheeks, lips and chin establish this up front.

This includes the interpretation of a mask. A face can function as a mask, where expression can deceive what we are imagining about the individual's emotional or physical states (Fig. 22).

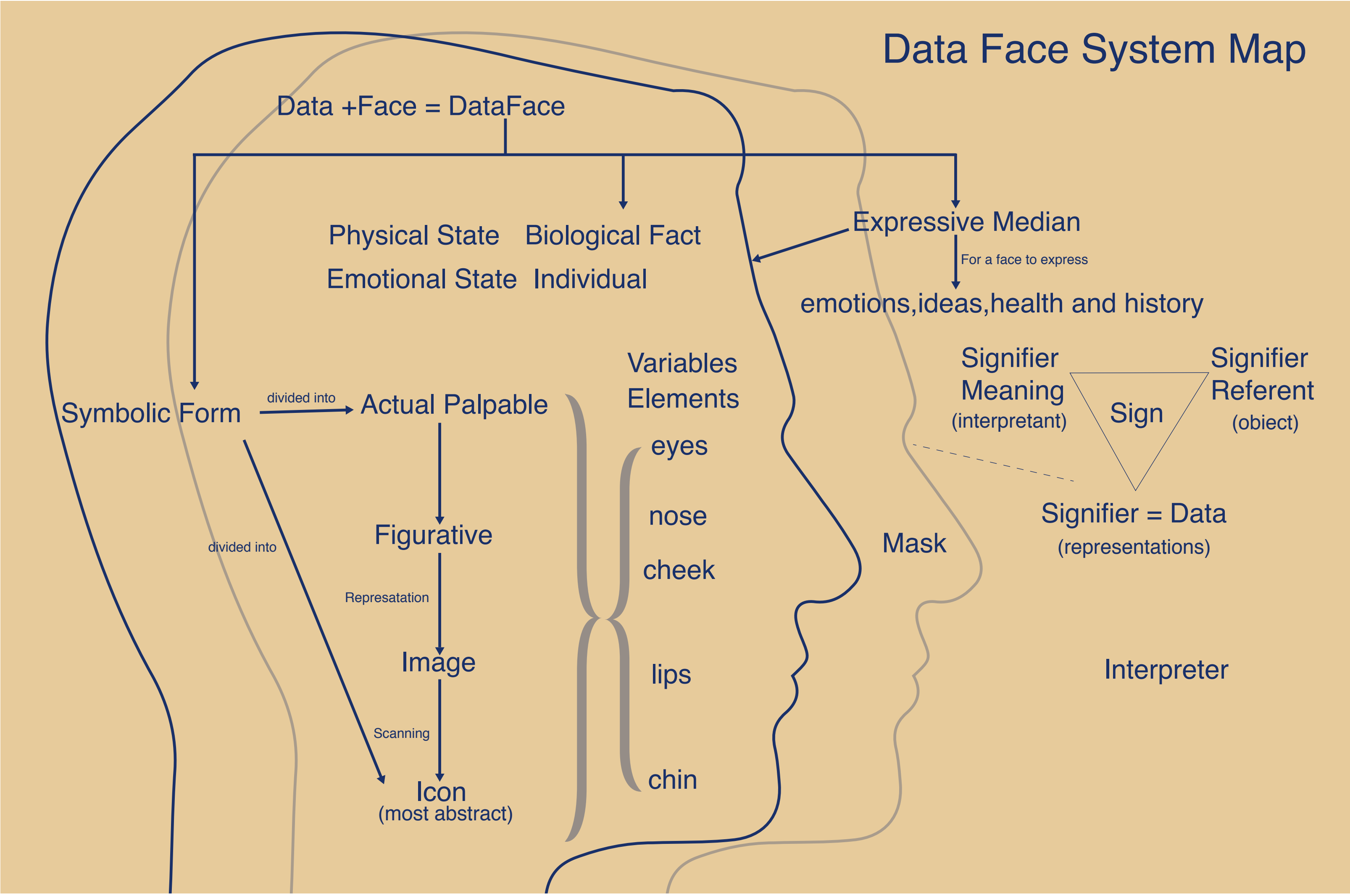


Figure.22. DataFace system map.

THREE STEPS IN THE INTERPRETATION OF

The thesis of meaning-making is defined as assigned processes to express meaningful communication among signs, indication, designation, likeness, symbols, a meaning of language, and signification (Michael, 1999). It contains a significant part of communications originating from both linguistic and non-linguistic sign systems.

The Italian semiotician Umberto Eco explained that every cultural phenomenon may be studied as a form of communication, in which relevant cultural elements can adapt to a semiotic niche in the world. In Semiotics and the Philosophy of Language, Eco stated that semiotic theories are implicit in the work of most, perhaps all, major thinkers. In practice, it has been used frequently in many fields, such as medical research and design work.

Deely (2006) has classified three sections in the interpretation of signs, found in the field of "observant of signs" to denote the branch of semiotics; each section compasses human understanding.

The first section presents the nature of things, which contains their own features and specific manner of operation. The second refers to a voluntary agent to obtain the information from the result. The third refers to media, the ways or means by which the information is communicated. These three sorts were initially rooted in linguistic communication to form a "meaningful world" of objects through observation of non-linguistic animals. The information conveyed by an action of signs can be clearly identified in the semiotic stage.

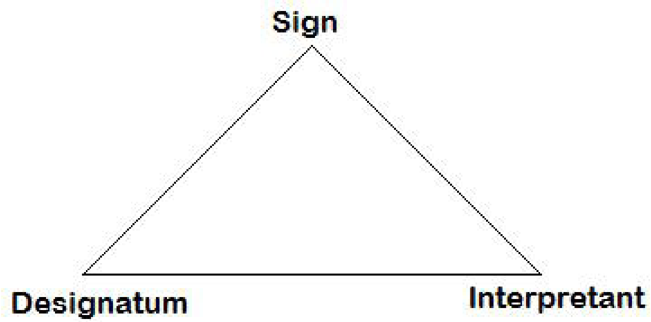


Figure.23. Pierce's semiotic triangle.
https://chhg.wikispaces.com/Semiotics,+Performances,+Airport+Art

Contemporary science is seeing the emergence of a new data economy, which can be defined as the new signs or symbols in semiotic studies. The data, as fundamental unit of exchange, contains three sections including the nature of things, the voluntary agent, and the media that transfer the meaning in the world (Beynon, 2002). While using data in the digital age provides many potential advantages in presenting individual or aggregate data, there are becoming important social and ethical concerns. The process of sign system formed by data carries meaning that depends on the use of codes, derived from words, body movements, or expression.

In relation to the way the data is transmitted, signs such as words or expression of a face can add new shades of connotation to the data. For example, the face as a sign can signify the signified and significance in terms of biological components on a face.

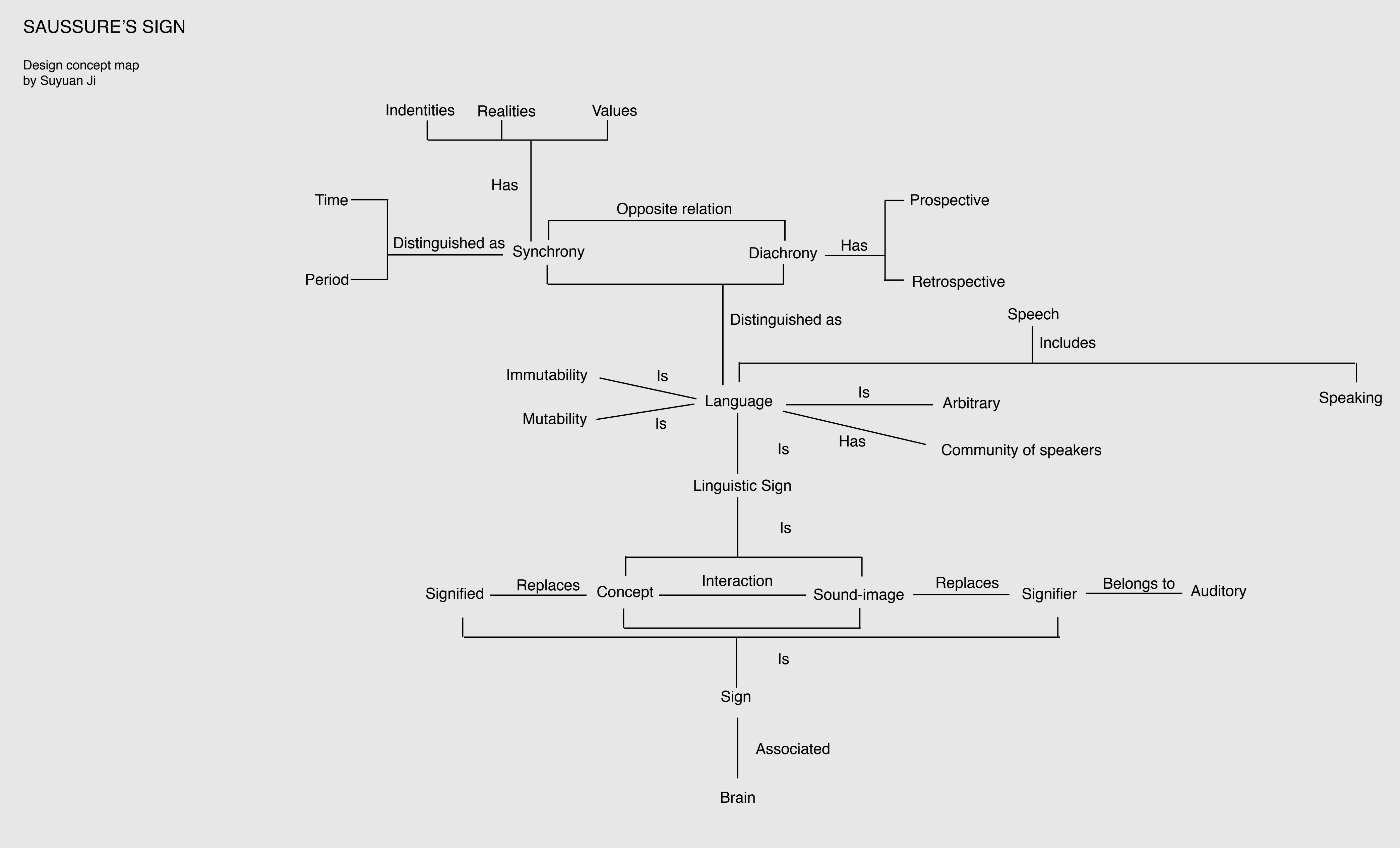


Figure.24. Sign concept map.

BIOLOGICALLY COMPONENTS

Initially, the idea of using faces is to explore the recognition method used to identify human faces, regarding the perception of minute changes in the biological components of the face. However, the features of the faces vary, which makes variables hard to map as anatomical components of the face have been found to carry significant weight. To measure the depth, width, and length at each point of the parts of the face appears to be the appropriate process to form the meaning through signs. However, it is insufficient to define the biological components on a face as a single sign, as it is not only related to semiotic features, but also iconography to interpret the content of images formed by biological components of a face. Unlike semiotic that focus on signs, iconography is to study the "identification, description, and the particular compositions demonstrated on the images."

Face, as the production of an image, can be regarded as "an iconography" to identify a distinctive depiction of a subject based on the content of the image, such as the color of an eye on a face and the shape of a mouth. When referring to the interpretation of the content of images, genres are immediately recognizable through their iconography, and motifs on a face become associated with a specific genre through repetition.

Therefore, it requires a "motifs-maker" or "face designer" to elaborate on the practice of identification and classification of biological components of a face to appropriately use iconography to understand meaning (Fig.25). Panofsky codified a powerful approach to iconography, and defined iconography as "the branch of the history of art which concerns itself with the subject matter or meaning of works of art, as opposed to forming." With the widespread usage of the coding method, iconography as medium allows researchers to move to the idea, considered as a sign signified to deal with an expression that biological components contain connotation (Fig. 26).

Figure.25. Panofsky made important contributions to the study of iconography, including his interpretation of Jan van Eyck's Arnolfini Portrait (1434, pictured).
https://en.wikipedia.org/wiki/Arnolfini_Portrait

Figure.26. Holbein's The Ambassadors is a complex work whose iconography remains the subject of debate.

Hans Holbein the Younger （1497 / 1498-1543）

Jean de Dinteville, French Ambassador to the court of Henry VIII of England, and Georges de Selve, Bishop of Lavaur. The painting is famous for containing, in the foreground, at the bottom, a spectacular anamorphic, which, from an oblique point of view, is revealed to be a human skull. An Armenian vishapagorg rug is on the table.
https://en.wikipedia.org/wiki/The_Ambassadors_(Holbein)

COMPONENTS IN THE PROCESS OF INTERPRETATION

This thesis focuses on the logical dimensions to establish the median in communication between biological components of a face and significance. The idea can be divided into five sources to make sense of the data formation.

TRANSLATION

As semiotics can generalize the definition of a sign, the range of sign systems and sign relations can be broadened in their widest analogical or metaphorical sense. For example, Chernoff faces illustrating the necessary features of signs to show the difference laid between separate components rather than subjects. Therefore, the translation of some differences regarding subjects can be deeply concerned by semiotics analysis.

TRANSMISSION

The process of transmission refers to meaning-makers employing and integrating methods to develop cognitive information with conceptual and textual analysis as well as experimental investigations.

RECEPTION

While semiotic is conceived as philosophical logic studied in terms of signs that are not always artificial (Michael, 1999). The semiotic addresses not only the external communication mechanism but the internal representation machine, which implies the whole inquiry process should be integrated rather than focuses on sign process or the manner of operation. For example, as signs can be icons, indices, and symbols, subdivides may contain similar or different significances perceived in each case. Thus, signifier and signified are not fixed.

EVALUATION

The importance of interpretation of signs is that it provides the "meaningful world" of objects that may not able to carry any meaning in its original formation. For example, the estimative powers of interpretation for the objects rooted in the biological components of a face forms the non-linguistic communication expand human understanding in terms of usage of signs. Thus, a sign relation can be established successfully.

COMPARISON

Technological advances including semiotic stage expansion and iconographic arrangement, data is allowed to build up huge collections of photographs, or index. For example, available online dataset can be defined as the sender in this study, whereas more variables can be received to establish notion opened the way to understanding an action of signs beyond the realm of a face. The conception can be depicted by the case of Chernoff faces.

CHAPTER 3: Methodologies

This thesis and complementary experimentation seeks to explore the limitations and advantages of data faces: sets of data described by facial components. These statistical illustrations intended for data representation convey various forms of information, conveying these datasets into emotional expressions relevant to the data represented. The methodology of this study relies on experiments designed to investigate the respective advantages and disadvantages of data faces in this use: positive and negative consequences such as emotional effects on interpreting the statistics presented by these data faces. For example, as biological components may form symbols and signs, and human nature designates a method of survival related to the recognition of facial expressions and components, experimental subjects may experience emotional influence when studying the data represented by data faces.

Researchers may begin to understand these affects through observance, experimentation, and data analysis; such is the method of this inquiry.

After the creation of unique data faces, examining and representing various pieces of data obtained through governmental databases and associated with numerous cities and/or neighborhoods in the state of Massachusetts, three posters were mathematically designed and designated with the three unique sets of data: crime statistics in the most populous Massachusetts cities, 311 requests in the city of Boston, and weather patterns associated with the most populous Massachusetts cities.

The creation of these data faces, with a formation process explored in the next section of information entitled The State of the Art, is critical to the methodological success of this mathematical, informational, and psychological examination of the assets and liabilities presented in data face utilization. Ultimately, this thesis expresses a desire to examine experimental results in relation to findings of previous studies concluded with data face methods such as Chernoff faces. In turn, a part of the methodology of this study is to examine, compare, and contrast scientific explanations provided by these experiments and prior experiments completed by erstwhile scientists.

In other words, a desire to understand the relative differences in this form of data face versus previous forms drives the experimental process described in this study. The codes created by data faces, codes representing statistical value, are the centralized idea to be examined: whether these statistical, biological codes hold advantages or disadvantages over other forms of data representation, such as bar graphs or tables.

This is an idea examined through survey of participants and supplemental data analysis with the information and feedback provided by the respondents. Through this method of observation and examination, comprehension of its limitations and advantages in practice can progress. In addition, the use of three varying posters can further assist in statistical analysis due to the vast physical, visual, differences in each poster—physical differences ranging from typical, more natural faces to that of obscure or abstract facial components. This data analysis related to participant response and explanation is vital in the methodological intents and purposes set forth by this thesis. Questions formulated for the respondents depend upon considerations of influence, cognition, and perception recognized by the experimental procedure.

CHAPTER 4: Experiment

TESTS
TEST A:MICROEXPRESSION

Step 1: I found 2 sets of graphics, containing 6 basic expressions (Fig.28).

Step 2: I created an online survey and found 100 people to match the emotion to the face.

Step 3: I collected the survey results and analyzed the data.

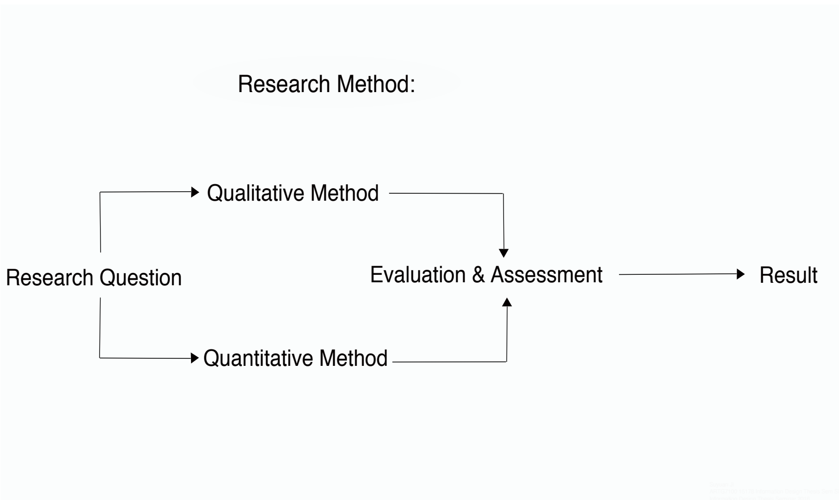


Figure.27. DataFace research method.

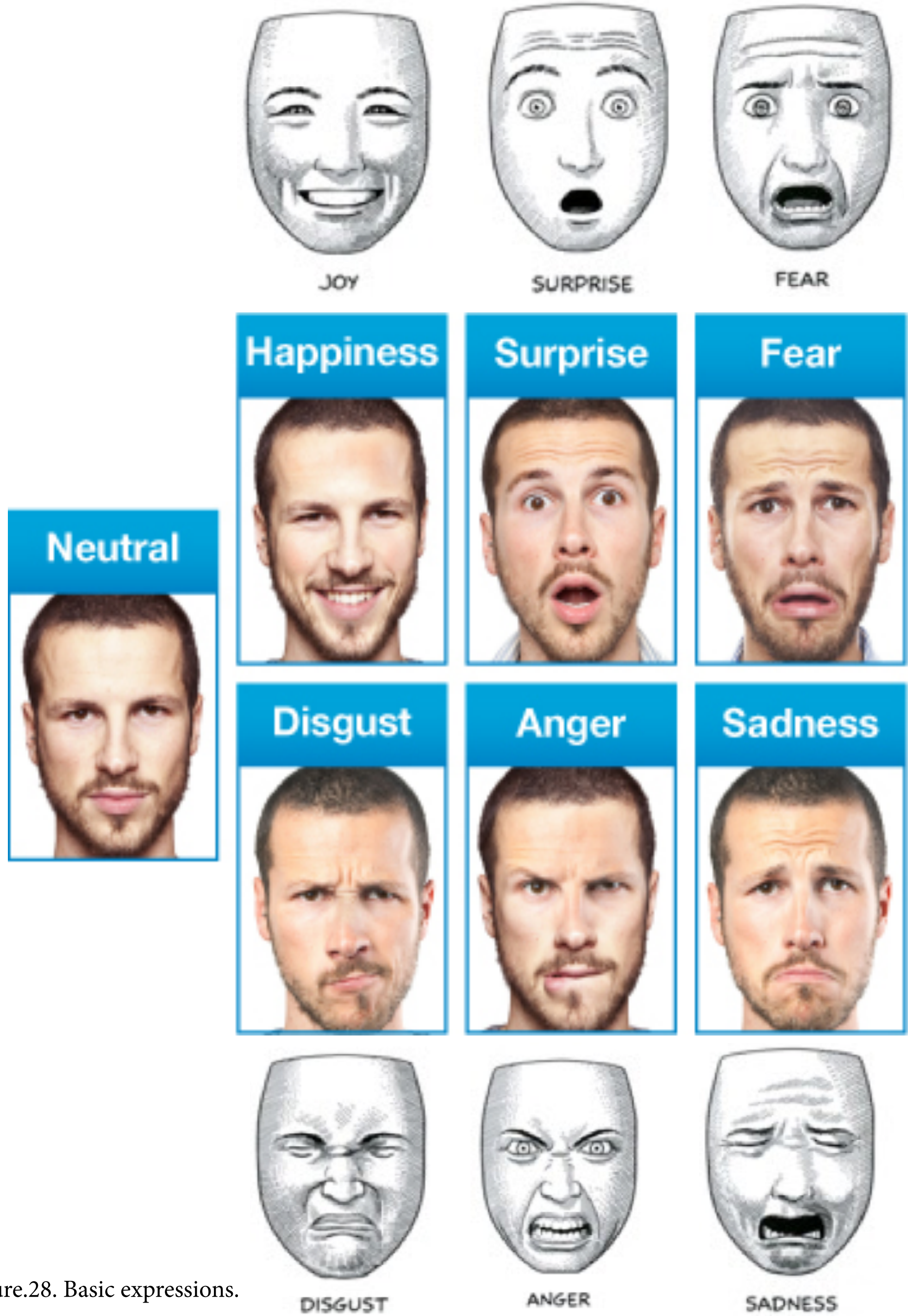


Figure.28. Basic expressions.

Test A aims to prove that people can easily find out the relationship between semiotics and significance. Table 1 is made according to the 100 collected surveys. The 90% all correct rate shows that most people can understand the process of carrying meaning which depends on the use of codes derived from the individual expression they make to show emotion.

Throughout application of semiotics (with six expressions including anger, disgust, fear, joy, sadness and surprise) different semiotics have encompassed signs in sensory modality. Different symbols have broadened the range of sign system and sign relations. The dimensions in an anger face has expanded eye and mouth size, in which the maximal anger expression shows the widest metaphorical sense. The dimensions in the disgust face has shrunk the size of the mouth and eye to present the analogical sense of that emotion in the human face. The dimensions in the fear face has expanded eye and mouth size, as well as face shape, by which the expression shows the meaning of fear. The dimensions in the joy face has a difference in the shape of the eye and to represent the joy expression with a metaphorical sense. The dimensions in the sadness face has an expanded size and shape, in which the maximal expression shows the metaphorical sadness. The dimensions in the surprise face has an expanded eye and mouth shape, in which the maximal surprise expression can be observed.

The result on the following excel table 1 :

	The number of people	rate
All wrong	0	0%
Correct 1 of 6	1	1%
Correct 2 of 6	1	1%
Correct 3 of 6	1	1%
Correct 4 of 6	2	2%
Correct 5 of 6	5	5%
All correct	90	90%

TEST B: PERSONALITY

Step 1: Creation of the survey (see，appendix 3) and find 100 people to respond.（participants should choose 1 facial feature and 1 personality they own).

Step 2: A collection of the survey result and analysis.

Table 2 shows that the test aims to explore the casual relations between semiotic signs and significance that can be observed by people. Five parts of a face have been highlighted to show the related meanings based on changes in size and shape.

All participants are required make one choice for each unit and they will be told that there is no relationship between the main feature of the part of a face and the correspondent interpretation. In reality, each unit (both columns in “main feature” and “interpretation”) is regarded as a valid answer if they tick both in the same row.

Table 2: Symbols and significance

Please tick the box according to your own personality and facial feature.

	Facial feature	☑
The Forehead (Please check only one)	A HIGH forehead	☐
	A HIGH and WIDE forehead	☐
	A HIGH and NARROW fore-head	☐
	A LOW forehead	☐
	A FLAT forehead	☐
The Eyes (Please check only one)	Round Eyes	☐
	Oval Eyes	☐
	Slanted Eyes	☐
	Open Eyes	☐
	Narrow Eyes	☐
The Nose (Please check only one)	A LARGE nose	☐
	A SMALL nose	☐
	A THIN nose	☐
	A WIDE nose	☐
	A LONG nose	☐
	A SHORT nose	☐
The Lips (Please check only one)	Large lips	☐
	Small lips	☐
	Curving lips (upward)	☐
	Curving lips (downward)	☐
The Chin (Please check only one)	A LONG chin	☐
	A SHORT chin	☐
	A POINTY chin	☐
	A FORWARD chin	☐
	A RECEDING chin	☐
	A CLEFT in a chin	☐

	personality	☑
Personality (Please tick only five according to your personality)	intellectual	☐
	open-minded	☐
	inquisitive and analytical	☐
	practical and down-to-earth	☐
	good concentration	☐
	trusting nature	☐
	good natured	☐
	take pleasure in self-love	☐
	open, friendly	☐
	suspicious nature	☐
	open-minded, and aggressive	☐
	quiet	☐
	generally irritated	☐
	generally careless	☐
	careful	☐
	generally happy	☐
	generous	☐
	selfish	☐
	generally happy	☐
	generally unhappy	☐
	stick to what they say	☐
	changes mind often	☐
	out-going, positive	☐
	stubborn	☐
	generally impatient	☐
	self-centered	☐

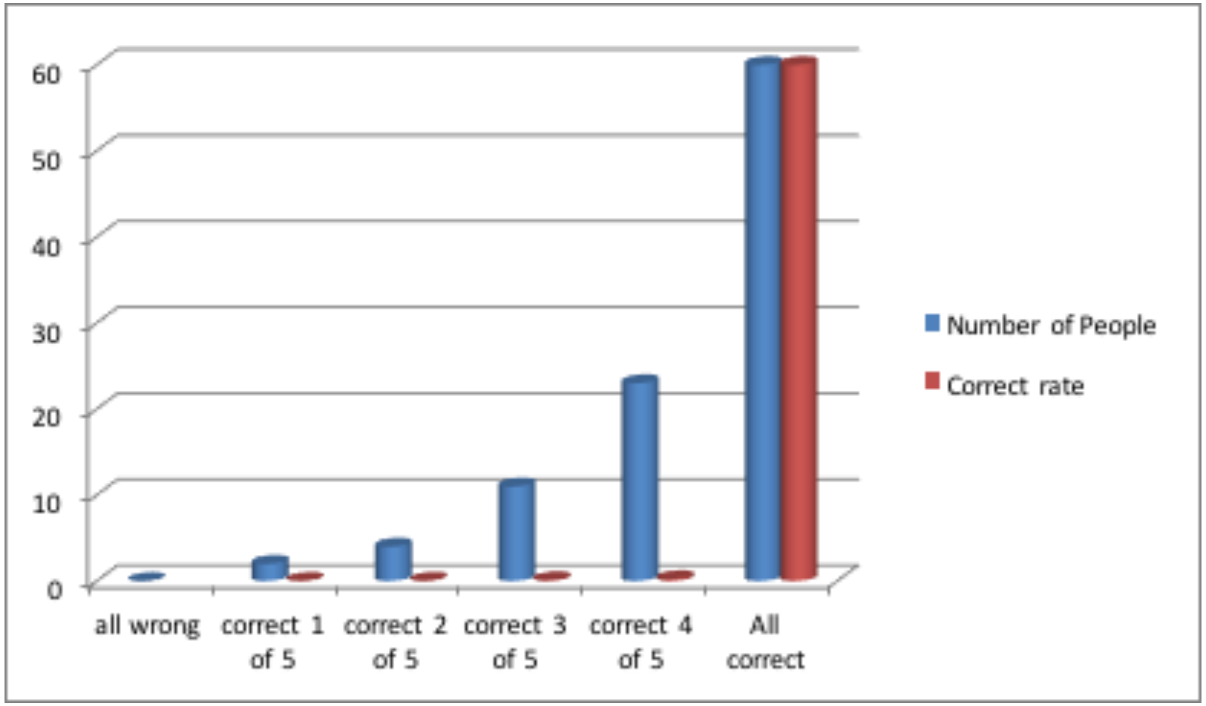
Source： http://fortunesnow.com/newsite/library/faces.asp

Table 3 presents the results from data analysis in the form of an Excel document. Most participants agree that biological parts of a face can carry more significance, over 60%. No participant believes that biological parts of a face can carry zero significance. There are crucial differences to show that more than 23% people agree that most biological parts of a face can carry significance such as eyes and lips, whereas they do not believe that the chin can carry any meaning for emotions. There are 11% people agree that some parts of a face can carry significance, but some cannot.

For example, except for the eye and lip, other parts of a face including the forehead, nose and chin cannot show any feeling and emotion to make communication with others. In addition, nearly 4% participants believe that only eye can invisibly communicate among people, whereas other parts of a face are unable to show any meaning in communication. Only 2% people think it is possible for one of five parts to carry information or significance. No one agrees that there is zero communication between people regarding the biological parts of a face.

Although the data collection is conducted in the empirical investigation, the results may be more valuable if the data size could be larger, for example, the involvement of a 1000 participants. Because of the small size of the sampling group, the findings may be biased. In addition, as the data only focuses on investigating five parts of a face, some other biological parts of a face may be explored to garner more meaning and significance in future research.

Table 3: Results from data analysis(from survey website)



TEST C: MASSACHUSETTS DATAFACE

THE CASE OF MASSACHUSETTS' CRIME RATE

The dataset for Crime rates for Massachusetts (Appendix 1) could be presented in the form of bar chart or histogram. However, with the ongoing development of information technology, human facial expression recognition can be regarded as most valuable and important subject to form the significance of facial semiotics.

The signs and symbols of a face show the anthropological dimensions; For example, every criminal phenomenon can be studied as communication in terms of the semiotics of a face which focuses on the logical dimensions of the science. Multivariate data in the shape of a human face examines areas belonging and show the relevant information of criminals. The individual part of the face such as eyes, ears, mouth and nose are linked to know a person and identify types of their semiotic to the niche in the criminal sciences.

DATAFACE SKETCH

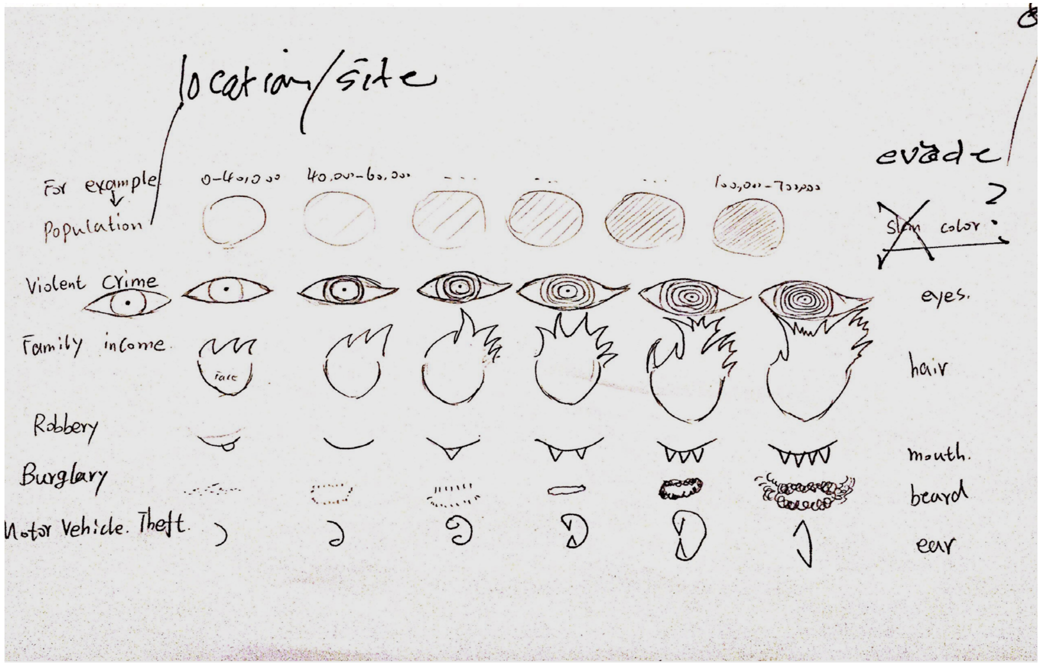


Figure.29. Sketch 1.

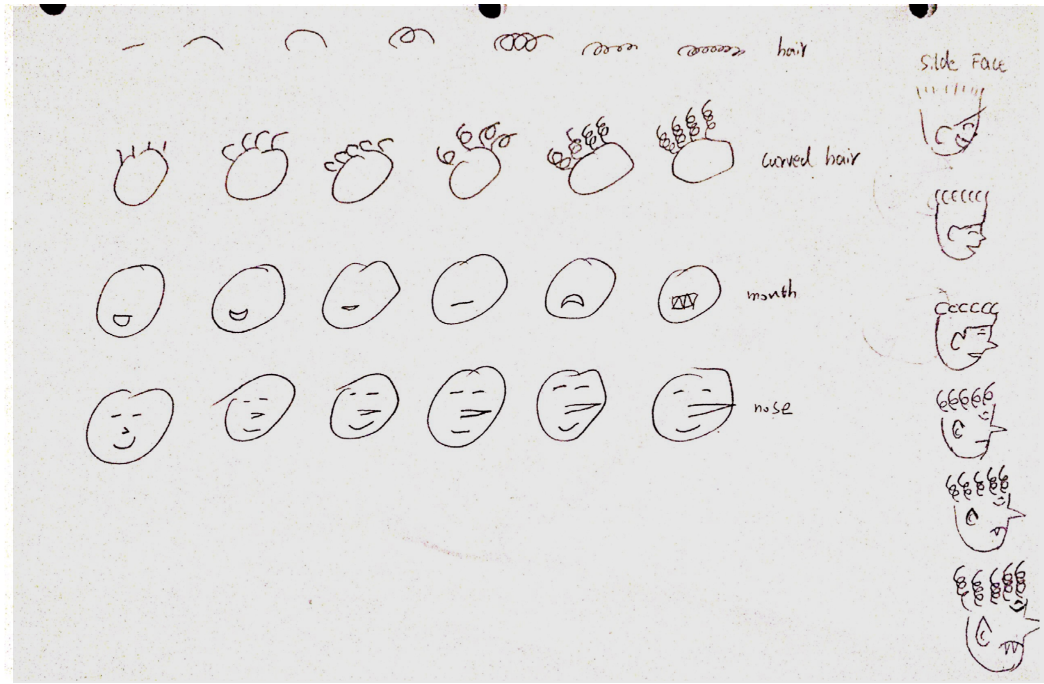


Figure.30. Sketch 2.

Abstract lines and dots are used to present different faces.

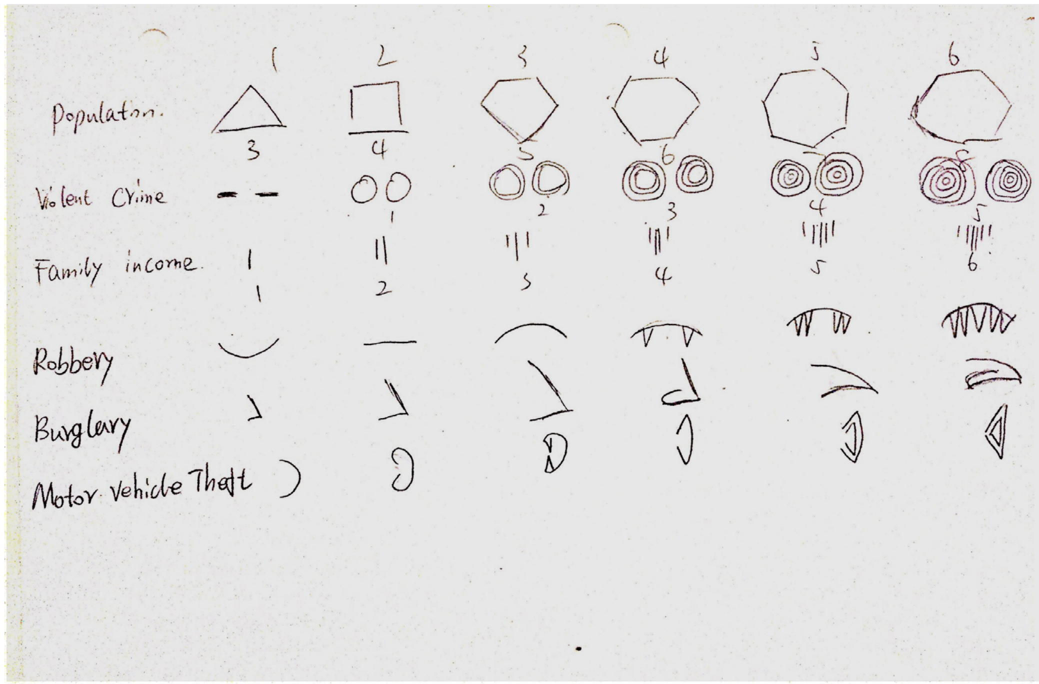


Figure.31. Sketch 3.

Use the nose as an example, the smaller the section between data, the simpler the nose will be. As the section become larger, the nose will be more complex, even like nightmarish.

DataFace Design Process

Data

	City	Robbery
5	City	Robbery
6	Abington	7
7	Acton	2
8	Acushnet	0
9	Adams	3
10	Agawam	3
11	Amesbury	2
12	Amherst	3
13	Andover	1
14	Arlington	10
15	Ashburnham	1
16	Ashby	1
17	Ashland	0
18	Athol	3
19	Attleboro	16
20	Auburn	7
21	Avon	4
22	Ayer	1
23	Barnstable	30
24	Barre	0
25	Becket	0
26	Bedford	3
27	Belchertown	1
28	Bellingham	8
29	Belmont	2
30	Berkley	1
31	Berlin	0
32	Bernardston	1
33	Beverly	9
34	Billerica	8
35	Blackstone	2

Figure.32. Massachusetts data concerning robbery.Complete data in the appendix 4.

Data in Line Graph

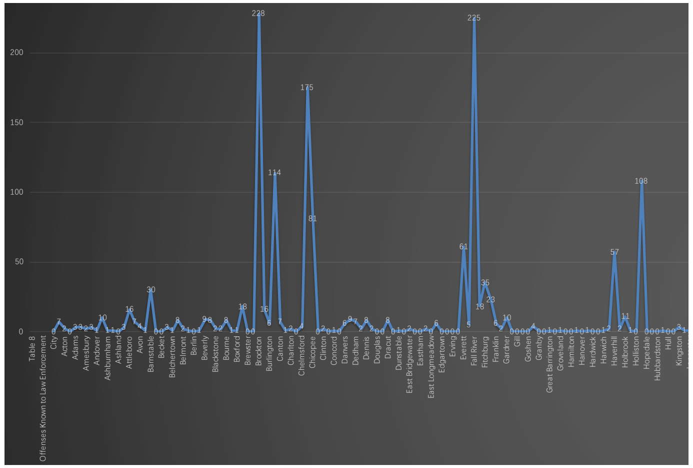


Figure.33. Robbery data in a line graph.
Complete graph in the appendix 5.

Design DataFace

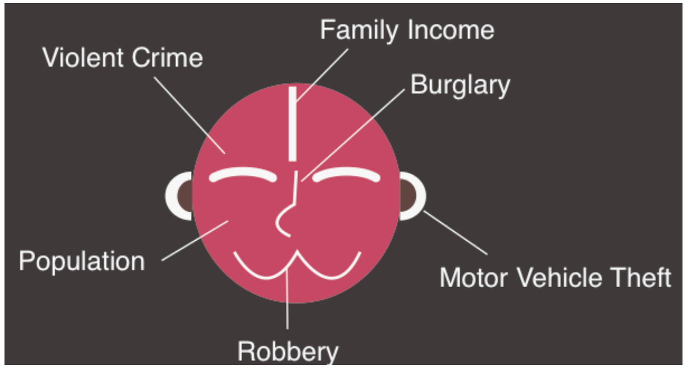


Figure.34. Massachusetts data face 1.

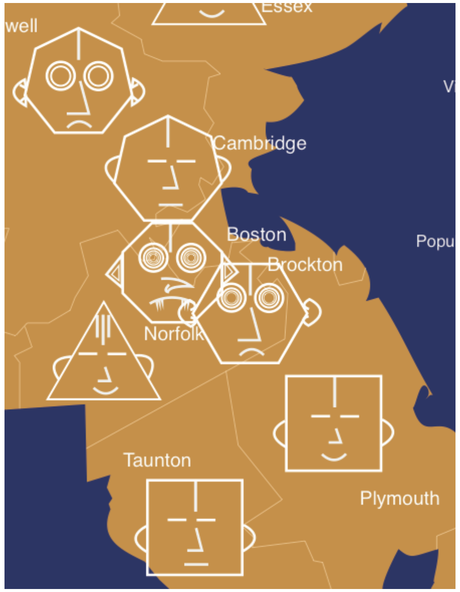


Figure.35. Final version.

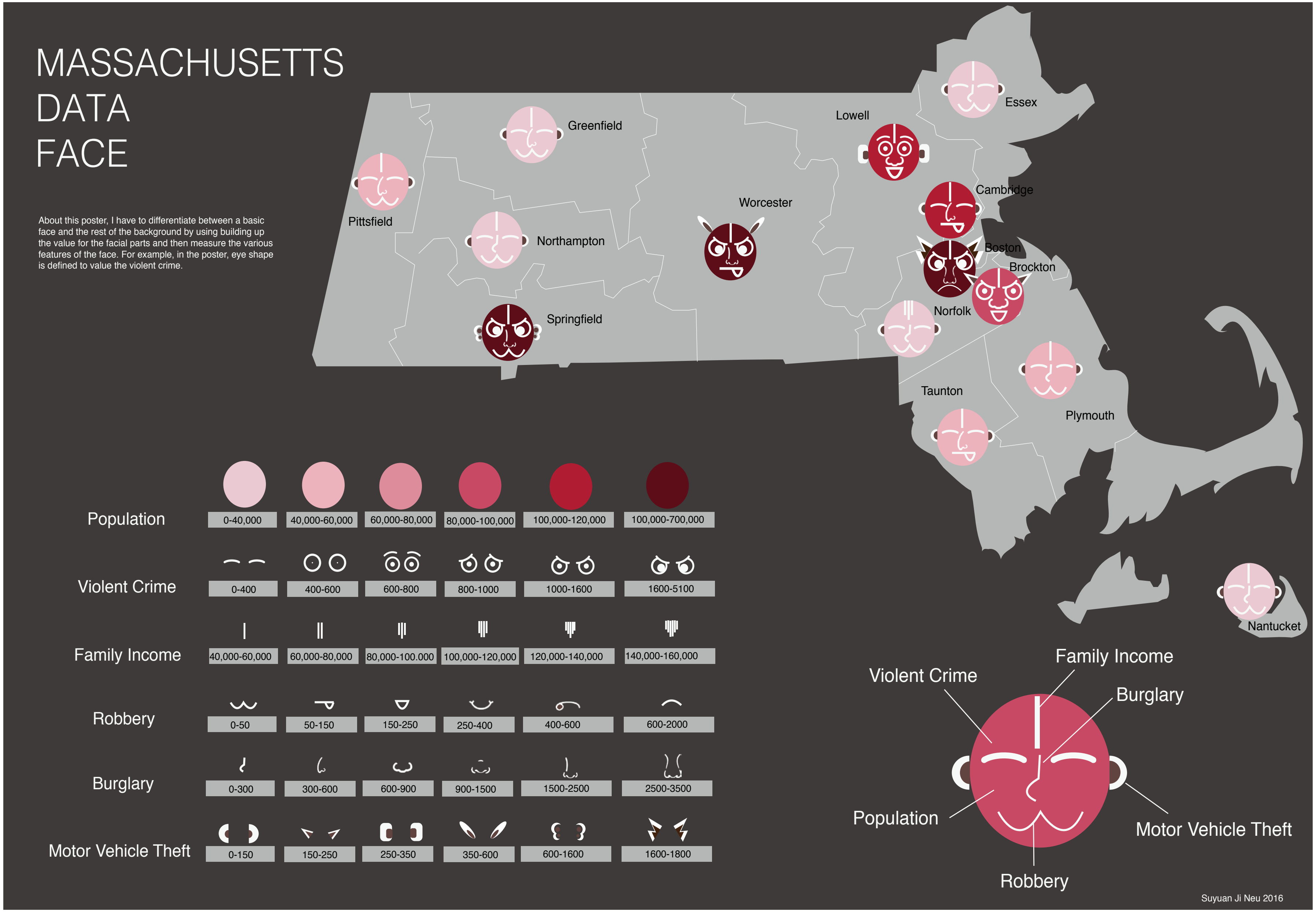


Figure.36. Massachusetts data face 1.

This poster was designed based on my thesis topic at the beginning of this project.
Based on data from a Massachusetts data base, the study use different faces of human origin to display the column of the data.

However, as the study continued, the skin color difference was avoided to consider of racism. Similarly, which information is initial should also be considered(Fig.36).

The description of demographic factors in criminal research is popular because they graphically represent the program logic regarding a serious standard of geometric symbols and lines connected according to innovative thinking of a designer. The relationship between perception towards data made of face and semiotic significance can be found in the following areas (Fig. 37).

1. Line: this is the abstraction symbol to represent the population density; this is an apt way of showing which areas have large population and which have less by using a longer or shorter line to represent the statistic.

2. Eye: the symbol of eye can reflect the criminal rate; in other words, it can be used to show the high criminal rate in terms of eyes' wearing circles (change in eye size). This is based on human perception, as eyes become bigger when people are in panic.

3. Hair: to represent family income; the more lines used for describing the hair, the higher family income.

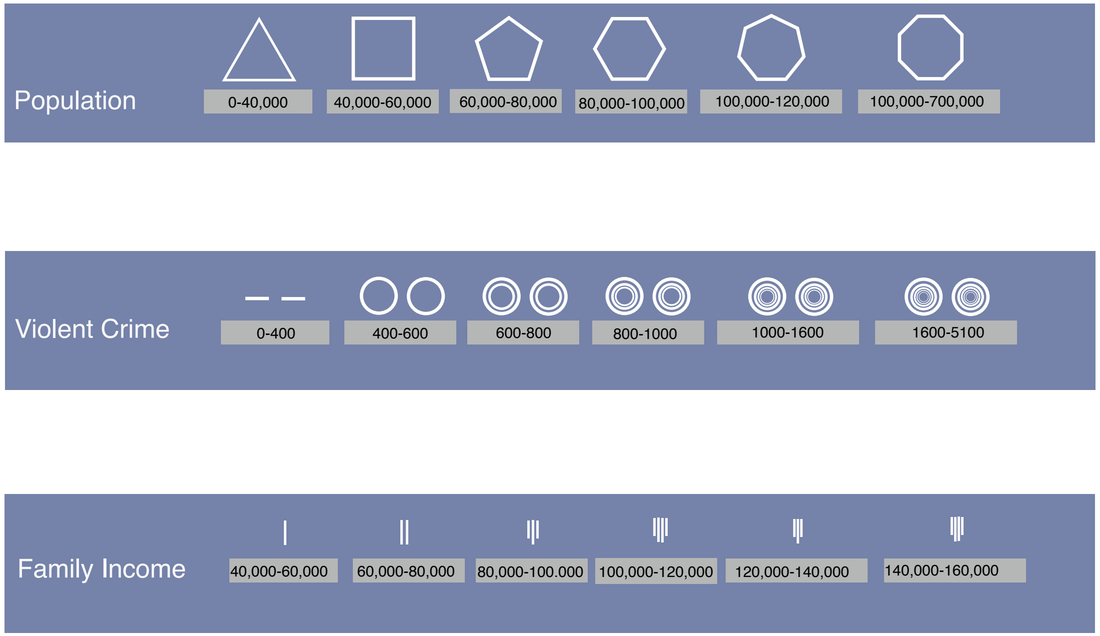


Figure.37. Facial features.

4. Mouth: To indicate the frequency of robbery using an arc. It states that people interact, in which, when they are happy, the mouth represents the action of smiling boosting their feelings of joy; increased robberies mean decreased joy.

5. Nose: To represent the number of burglary cases in criminal investigation. The more burglary case will be, the higher the nose is, which looks increasingly evil.

6. Ear: To represent vehicle theft cases in the criminal map. The harper and more lines on the symbols of the ear, the more vehicle theft cases occur.

Component parts of a face is not merely idiosyncratic marks used in the practical cases, they can contain other meanings to present other significances, such as emotion and identity.

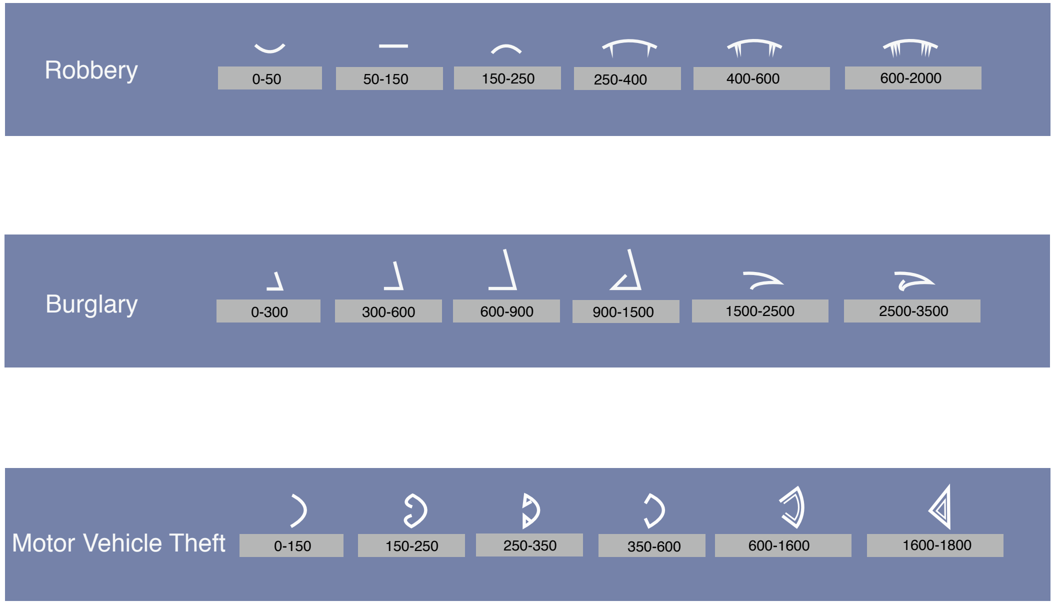


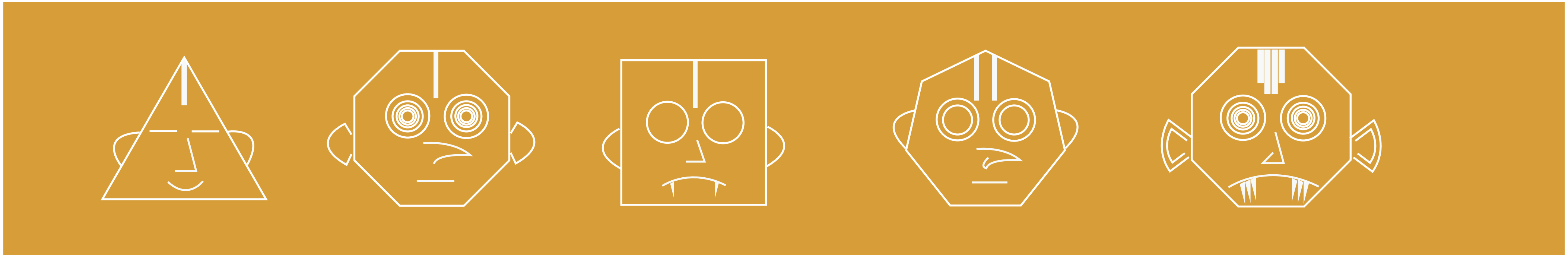
Figure.38. Facial features.

Determining the Priority of Signs Used in the Data

In computer science, a priority queue is a useful tool to classify abstract data while processing each element in a project. Each aspect has "priority" associated with specific features. Ultimately, certain features represented by different kinds of data are determined first to outline various facial elements on a face. Although there are few studies to illustrate a clear priority in determining the variables shown in the data.

Massachusetts DataFace has put eye size and mouth as the highest priority, resulting in a carriage of significant weight among other variables presented by other individual parts (such as the shape of ears and nose). By handling each variable differently, the features of the faces varying in perceived importance can be carefully chosen concerning a specific algorithm. The way to map the priority with variables has been plotted on five example faces to represent multivariate data in a manner that is easily discernible by the viewer (Fig. 39).

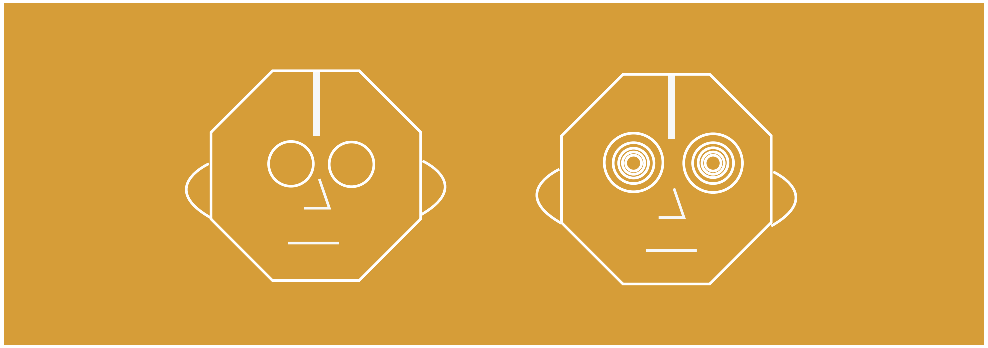
The acute changes in facial characteristics are determined by two-dimensional line drawings that contain a variety of facial features. In other words, these two different dimensions in the highest priority can show the most valuable information in a multidimensional dataset. The higher and less priority can be identified with more elements to visualize data. For example, the Massachusetts DataFace that possessed up to thirty-six distinct facial parameters.



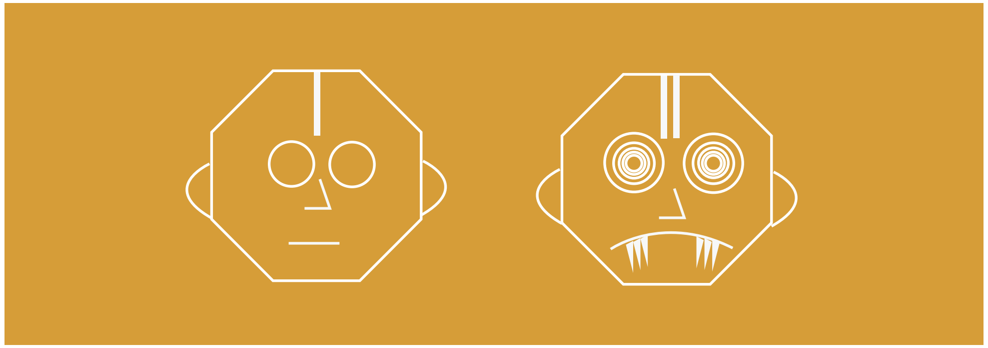
Determining the priority among five variables
(eye size and mouth as the highest priority)
Figure.39

The Massachusetts DataFace map presented different characters and exhibited a noticeable change, allowing viewers to easily recognize distinct faces with even small and subtle changes (Fig.40). It shows that facial recognition is significantly far from humans recognizing data solely dependent on the biological feature. Faces are a specific class of stimuli to show a unique process that contains emotional elements as effective means for visualization (as faces).

The DataFace can deeply explore linkage between the signs shown on a face and social-emotional factors. Sergerie. K. and Lepage. M. (2005) stated that people have negative stimuli from viewing visual or textual information that could yield better memory because human emotion can exert a modulatory role in episodic memory. From a medical perspective, as facial expressions are powerful emotional stimuli, the encoding of faces can integrate personal memory, and psychological processes occud in a viewing of visual or textual information. For example, it can be explained that people may relate to a peaceful situation when they view a calm face, whereas what lousy things may happen is related to a specific face (Fig. 41).



Samll eyes (the face on the left)
Figure.40



Combination of small eyes and slanted mouth
Figure.41

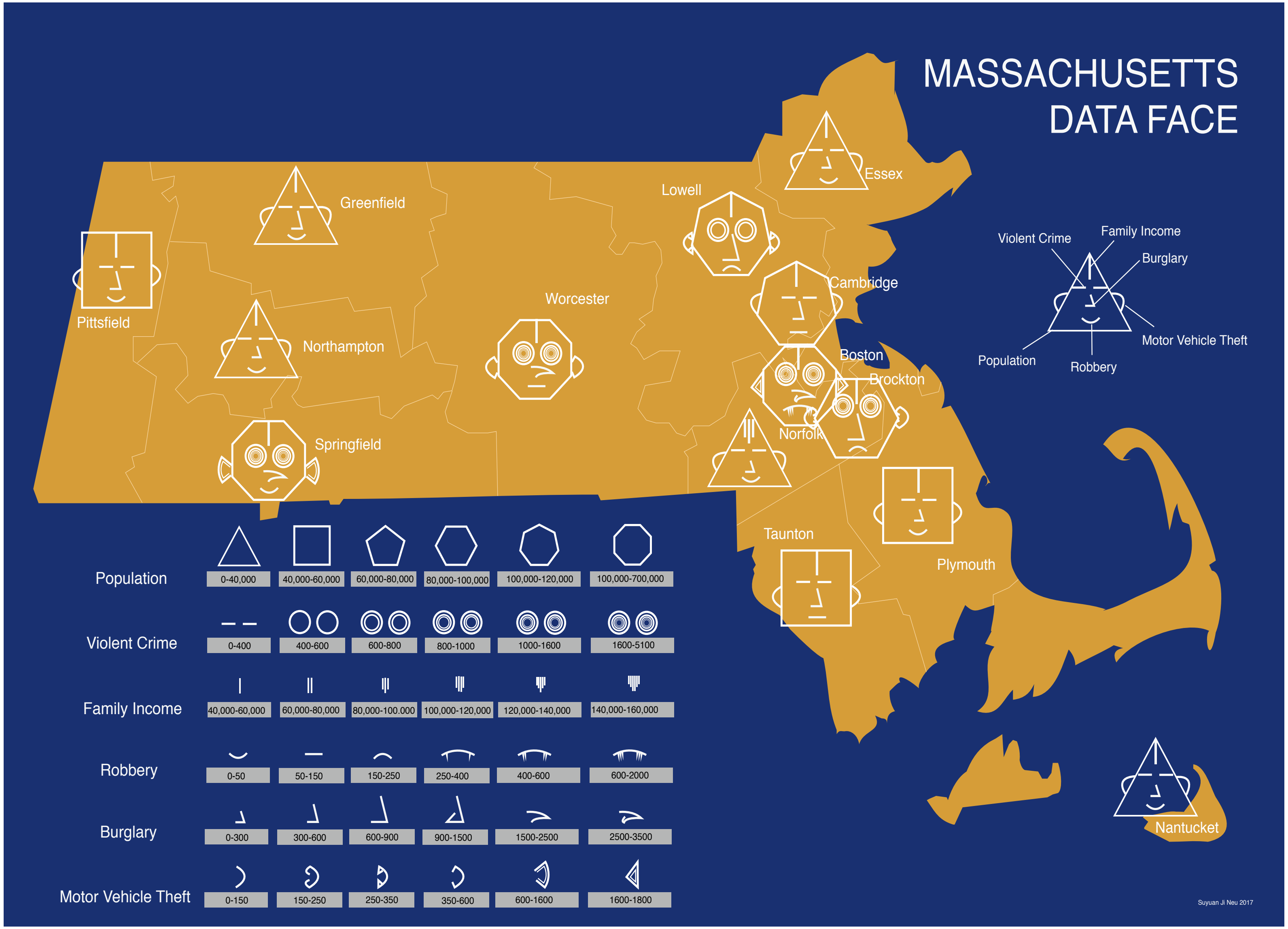


Figure.42. Final version.

This is the poster after edit(Fig.42).
The color is based on Seal of Massachusetts. Every face was re-designed. Lines were used to display the complexity of faces. When people look at this poster, the first thing appears would be the map of Massachusetts with different faces.

The explanations of the differences of faces would also be shown on the poster. Different data combinations will result in different faces. With the variety of data from different cities, the exhibition of data will also change, which will make data more interesting and more artistic.

Variation Between Two Posters

1. As human being perception of facial characteristics play important role in facial recognition, social and culture factors must be considered. For example, stimuli that have more naturalistic qualities to indicate negative feelings should be consistency with emotional perception in the context of culture.

2. In relations to an event's perceptional view, viewers' expectation from the context of design should be considered as the main influencer from the environment. For example, the background colors for these two posters are gray and blue respectively; I chose the colors of Massachusetts seal to revise the final version in order to represent the corresponding dataset(Fig.41). We can employ eye-tracking in this research in order to explore the effects of the negative feelings.

3. Like other studies, the facial recognition in these two posters has applied the way of indicating biological parts of a face, which can use a computer-based system to help researchers to measure the variables such as the size of eyes and nose. For example, the length of a nose can be defined by a unique dataset.

4. As the negative feelings may influence the effects of facial recognition, in my thesis, I will use lines and dots instead of red color to explore the rate of criminals; A color factor may be linked with the concept of race discrimination.

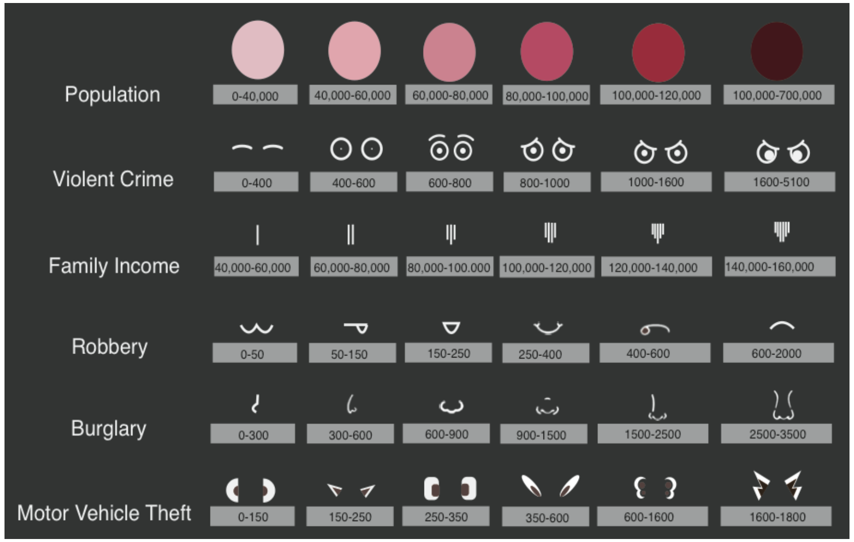


Figure.43. Facial features 1.

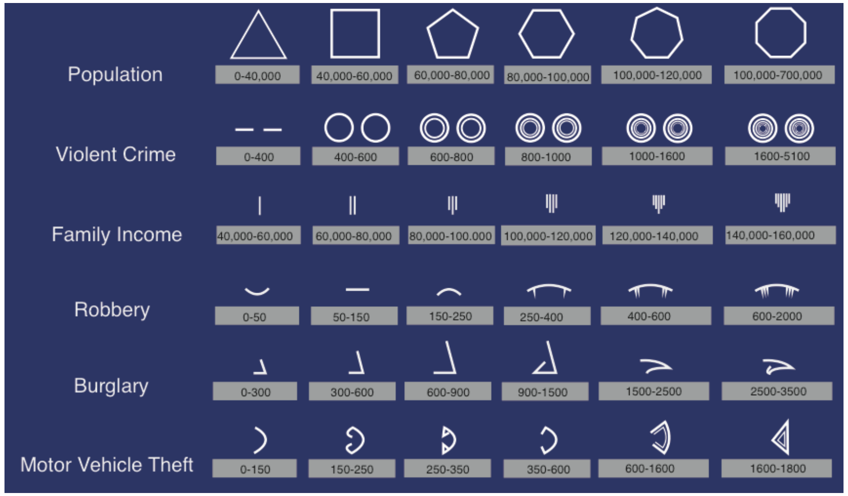


Figure.44. Facial features 2.

Research Method

DataFace has been proposed as a tool for scientific and information visualization. However, the effectiveness of this form of visualization is still open to speculation. DataFace, it is suggested, make use of humans' apparently inherent ability to recognize faces and small changes in facial characteristics. Limited research has been conducted to assess how well DataFace makes use of this ability. So far, it is still unclear how humans recognize faces and whether or not a specific set of rules governs the process.

To test the effectiveness of different features in DataFace, we conducted a controlled experiment with 3 posters, and each of them conclude twenty subjects. In doing some research on the validity of my claim to the DataFace, I use interview methodology to examine people perception towards graphic symbols that present the soul of signs system formed by facial components. All investigator was selected randomly and no matter where they lived before or now since they will be just show the poster without cities name and explanation of the features.

In this thesis, the data in posters is collected from different kinds of official websites sources and the posters are going to be used in interview research method.

The question design is considered as a key role in gathering data for results. The interview is specifically in an attempt to determine whether the interviewee correctly understand the significance of the symbols in DataFace. Total interviewees will not be told the explanation of these posters and just ask them to make choices.

This investigation aims to explore people perception, which focuses on interpreting visual impacts brought along by the symbols in Offenses Known to Law Enforcement and see whether they can read the minds and emotion of the DataFace. As the study is based on culture requires people's opinion and views, a complete understanding to explain peoples expressions in descriptive text format should use the qualitative method for interpretation.

Interviewees will answer the questions according to three posters that represent the different contents of information including weather condition, rate of criminal, and rate of service requests in Massachusetts. The below part that interpret the symbols in the poster will be covered during interviewing. It will discover the degree of people perception for the effectiveness of the overall poster, in which it provides a better understanding of the data rather than if they are explained individually. This method will examine the feasibility of the DataFace.

Poster Comments

In this experiment, I try to generate DataFace to be a tool for scientific and information visualization. In order to address this issue, we have conducted a survey, which tested the effectiveness of several features characteristic DataFace.

This experiment employed a five-factor, within-subjects design. The five factors, systematically varied to create the test conditions, were the target feature, as well as the number of faces in the display. Each of four kinds of facial features were distributed into five levels separately. The order of presentation of different faces was made according to the dataset. The increments of symbolic form are associated with the related variable values.

All the faces were unknown since the data was collected from different official website. For example, the data from poster 1 comes from “FBI Official Website," and poster 2 is based on “Boston 311 Service Request Official Website, and poster 3 uses the data from “National Weather Service Official Website” (for detailed dataset see the Appendix). As a result, there are a total of twenty distinct facial parameters in one poster.

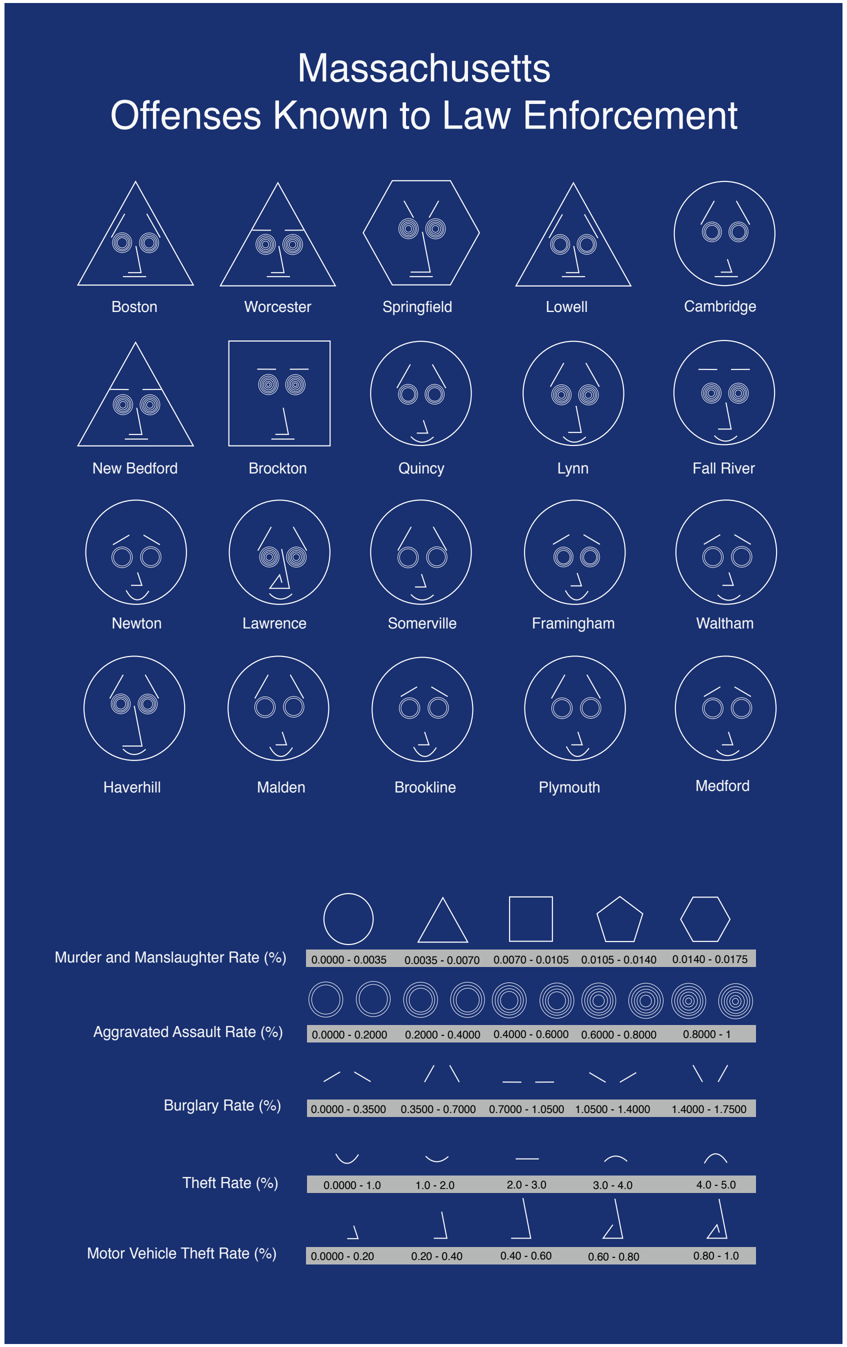


Figure.45. poster 1 - Original data faces used to describe the amount of occurring, known criminal offenses in Massachusetts cities.

Definition

Formula-1: to calculate the five different interval values:
Suppose the x= the max rate; y = min rate
d(common difference)=(x-y)/5
interval 1=[x,x+d]
interval 2=[x+d,x+2d]
interval 3=[x+2d,x+3d]
interval 4=[x+3d,x+4d]
interval 5=[x+4d,y]

The five target feature factors took on five values corresponding to five different intervals of variables:

• Facial features:

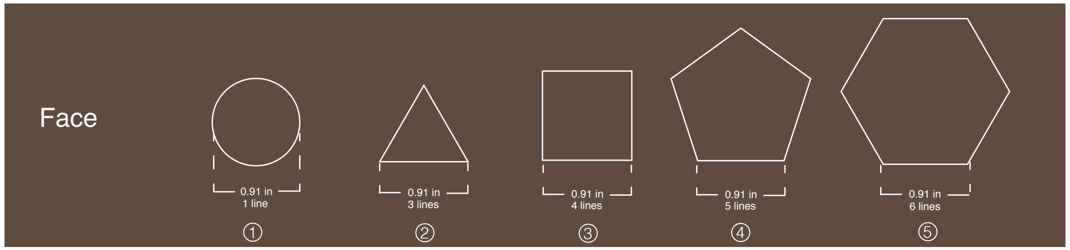


Figure.47. Face features.

Five particular facial features represent five different interval values.
Facial parameters ranged across the number of lines.
The length of all lines is 0.91 inches from circle (1 line) to hexagon (6 lines).

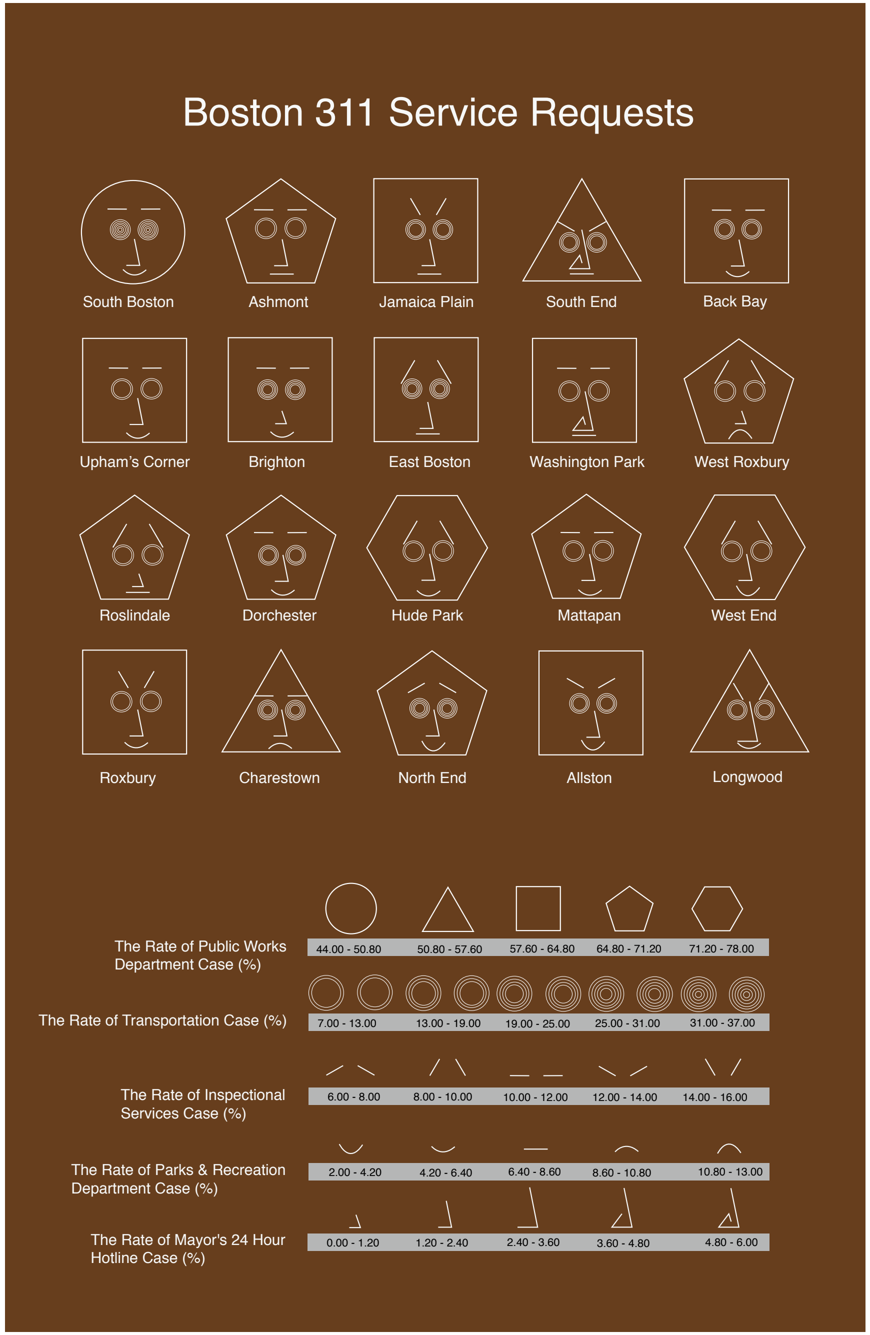


Figure.46. poster 2 - Original data faces used to describe the rate of 311 service requests in Boston, Massachusetts.

• Eyes:

Perception of eye size is distinguished into five different style.

Comments:
1. D (The diameter of each largest circle) = 0.45 inches
2. d -increment: (the distance between each circle) = 0.0375 inches
Difference: The number of the circles increases from two to six.

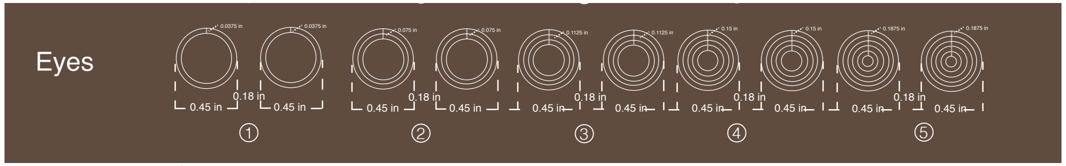


Figure.49. Eyes features.

• Eyebrows:

The eyebrow slants in five angles to represent five levels.

Comments:
1. L (the length of each line) = 0.24 inches
2. r (the shortest distance between each circle) = 0.19 inches
3. a (the increment of degree) = 30 degree
Difference: The angle of the line decreases from 60 degrees to 0.

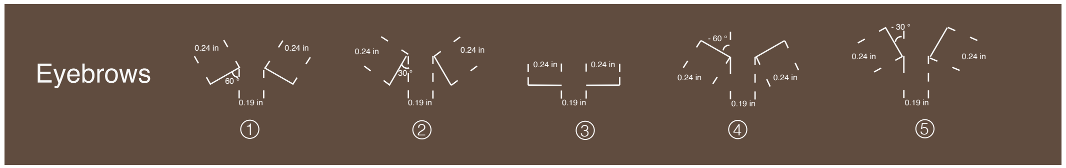


Figure.50. Eyebrows features.

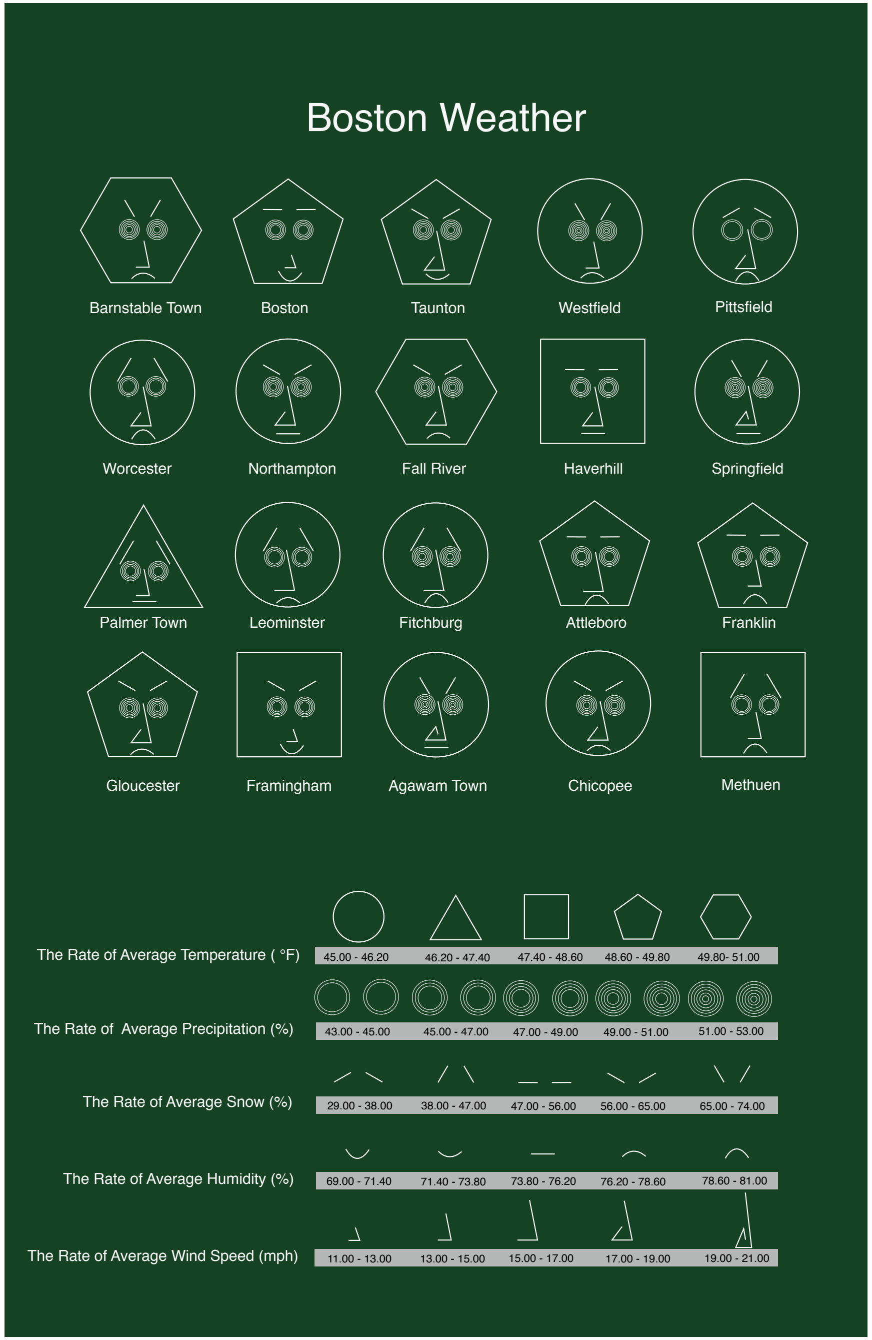


Figure.48. poster 3 - Original data faces used to describe the weather patterns in cities of Massachusetts.

• Mouth:

comments:
1. All the curves of mouths are different parabolas that share the same chord that is perpendicular to its direction.
2. h equals to the distance between parabolic vertex and chord.
3. h of parabolas that going downwards is negative.

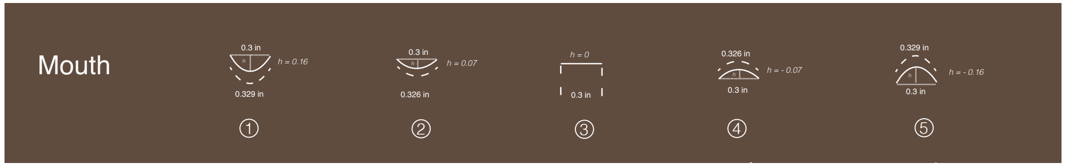


Figure.51. Mouth features.

• Nose:

Comments:
h (the increment of height) = 0.17 inches
L (the length of the under horizontal line) = 0.19 inches
Difference:
The height of the nose raises from 0.17 inches to 0.51inches. Lines were added in last two symbols to enhance the differences between levels.

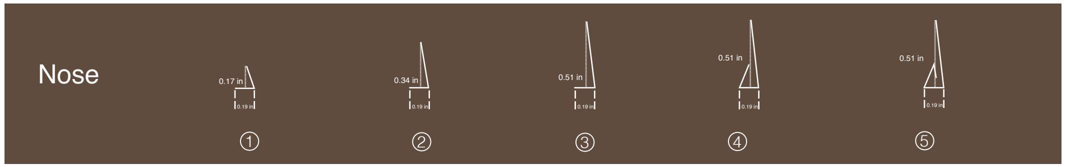


Figure.52. Nose features.

Facial Configuration of five different features (according to the dataset)

Example 1:
Poster 1 - Massachusetts Offenses Known to Law Enforcement
Sample size = 20 cities with biggest population in Massachusetts
Twenty cities, ranked as the twenty (20) largest city populations of Massachusetts have been chosen as the samples to be compared because the crime rate, based on a large population, is more representative and accurate.

Variable Definition:
Facial outline = Murder and Manslaughter Rate
Eyes = Aggravated Assault Rate
Eyebrows = Burglary Rate
Mouth = Theft Rate
Nose = Motor Vehicle Theft Rate

Research Question

This experiment consists of a five-question survey conducted in coordination with the three different posters of data faces, with the sixty data faces to be considered in study. Each question consists of two parts with the aim to understand the psychological reasoning behind the answers provided by each subject. The data obtained through the answers provided by the participants was then analyzed to provide an understanding of the limitations of data faces. Each of the five questions were specifically formulated to enhance the understanding of psychological factors associated with the limitations of data faces.

The first question is posed after a glance at the three data face posters. Participants were asked which data face appeared to represent the best city to live, as well as the worst city to live. It was then followed by a question of what the participant found appealing about the data face selected as the best city, as well as what was found to be unappealing about the data face chosen as the worst city.

This question is posed to discover gaps found in data interpretations between facial symbols and perceived emotions. For example, while a data face may appeal to the interviewee, this appeal is not based on set fact or data but an internal compulsion.

The second question asked by the conductor was designed to understand the participant's perception of similarities found within the presented data faces. The respondents were asked which cities, as represented by data face, they perceived to be the most similar and which was perceived as the most dissimilar. The conductor then asked the participant to explain the corresponding aspects of the data face that caused the lead to their conclusions. This question intends to cite examples that create perceived similarities and seek the reasonings behind this perception.

The third question of the series is closely related to the independent variables set in the posters. It helps to determine whether or not there is an implication that the data provided by data faces can be considered cognitive without the assistance of textual information. The participants were asked to describe the differences—explained by perceived emotions, such as fear, distaste, or happiness—found in the three posters. They were then asked to pinpoint the areas of the data faces that created these emotions to exist.

The fourth question posed to the participants is correlated to each of the five characteristics found in the posters— facial shape, eyes, mouth, nose, and eyebrows. The conductor asked the participants which two features felt the most obvious, as well as which two were the most easily overlooked. They then explained why they chose their answers. This question examines the aspects of data representation that become the most and least influential in data faces and why. This understanding can lead to further advancement in the effectiveness of data faces.

The fifth and final question serves as the control question of the study, completely disregarding the emotional impact of shapes and features found in data faces. The participants were shown graphs depicting the exact information represented in the data faces and were then asked if their answers to questions one and two remained the same. The participants were then asked to explain why they either remained the same or changed, whether it was due to a misunderstanding of the data or an overlooked feature of the data faces.

Results

Participants consistently found the facial shape and the eyes of a data face to be the most apparent characteristics of the graph. Meanwhile, participants observed the nose and the mouth to be the most easily undistinguished characteristics. Similarities were reported by respondents for noses or mouths with different length, curvature, or angle, while more apparent characteristics—such as the eyes and facial shape—were more easily differentiated.

The mouths posed an issue for differentiation between data faces. There were five available variations for mouths—two smiling, a straight face, and two frowning. However, the two smiling mouths and two frowning mouths were frequently confused to be the same data and representation. The slight changes in curvature appeared to become lost to the eye unless the graph was adequately studied and judged by the respondents before answering the questions proposed by the conductor.

Respondents were more likely to choose data faces that appeared more pleasant, or happy, as compared to more abstract or angry-looking faces. Interviewees reported the more obscure data faces to be the most unpleasant cities and conditions with minimal study to the posters, while the opposite occurred for faces appearing to be more pleasing to the eye. Participants were able to put emotions to each data face—whether they felt saddened, enlightened, frightful, or angered—and included this recognition in their process of determining their answers to the survey. This was not recognized by respondents until the data faces were compared with a different form of graphing the represented data.

There were a few issues apparent when the data faces were compared to the corresponding charts. Rates of severity were disregarded in examination by the respondents. This suggests that participants of this study misjudged the rates and percentages due to the facial features, though not proven by data obtained through this experiment.

Therefore, while the data faces may be considered cognitive without the use of textual information, there still appears to be a loss of information when translated by a data face.

Discussion

Participants of the study frequently judged the data faces due to their expressive emotions, while there was a general displeasure with more abstract facial features. Similarly, participants were able to associate emotions with each data face. These effects of emotion that were perceived by respondents concerning the studied data faces may be due to an appeal to pathos—emotional changes and influences due to the psychological forces found in the facial characteristics posed by the data representation.

Respondents were likely to report lines of different lengths, angles, curvature, or trajectories as similar. This effect may be due to an inability of the eye to differentiate minute differences found in line length or angle. In other words, respondents may not notice minor differences in line lengths or angle, as they appear to be the same or nearly the same to the eye. As a result, a loss of information occurred in terms of the severity of rating. However, with closer study and familiarization, these effects and limitations can be negated.

Respondents reported that the most obvious facial features—reported as the facial shape and the eyes—were the most influential in their decision-making. Meanwhile, respondents understood that they appeared to have difficulty distinguishing issues with judging the curvature of the mouth and therefore deciphering their statistical message. Perhaps the most efficient manner of adjusting for this phenomenon is further exaggerated mouths and facial expressions. However, this study did not focus on truly abstract facial features and expressions, therefore this limitation must be explored further.

While posters of data faces may not be considered an inadequate means of data representation, the graphs must still be studied thoroughly by respondents to maintain an understanding of the data. Otherwise, participants appeared to take the graphs for a face value and allow their expression to affect their opinions. Such phenomena defeat the purpose of data faces, as they are intended to be an effective way to study statistical value.

More principal factors described should be utilized in the most obvious features of data faces—such as face shape, eyes, or hair—to reduce the loss of information provided to the respondent. While abstract faces do not naturally appeal to those questioned in this experiment, more drastic differences may also prevent a loss of information provided. A small increase in the angle of the mouth, or slight changes in length of eyebrow or nose lines, appear to be lost in translation when applied to data faces.

Limitations

There are a few evident limitations to this thesis. In turn, further study and experimentation should be completed for additional conclusions. This experiment was able to conclude that there are limitations to data faces due to a neglect of the participants to note minor differences in curvature or line length. However, it was unable to calculate the likelihood of mistakes. Sample size and the amount of information translated into data faces are the two most apparent limitations of this study.

The largest limitation for this experiment exists in sample size. The number of respondents available limits the results of this experiment, as only twenty respondents participated in the study. Further experimentation would be more thorough with an increased number of available participants to create a more accurate sample size and to assist in an accurate representation of mistakes. This is necessary to calculate the limitations in the use of data faces represented in a poster format.

Similarly, additional features on the data faces—such as ears or hair—should be implemented to understand the effects of abstract faces completely. This is a second factor that should be examined, but was considerably limited by this study. As the amount of information available to be translated is a large benefit of the use of data faces, it is all the more important to closely examine all aspects of their use. Further study should include use of data faces with varying amounts of data and style in order to accomplish this scrutiny on effectiveness.

CHAPTER 4: Conclusion

The implementation of data faces began in 1973, evolving into use today. They are regarded as a method of data visualization that allows for multiple unique combinations of data and allows for substantial amounts of data to be translated. As each facial feature corresponds with a dataset, such as precipitation in figure three, multiple values can be expressed in a single graph without becoming messy. Characteristics from color difference, shape, facial features, or hair can create nearly endless possibilities for converting data into a visual form. Nevertheless, there appear to be limitations to the effectiveness of their representations.

This experiment consisted of a five-question survey relating to three posters of data faces, with twenty faces on each poster. Respondents were asked to both answer the questions posed by the conductor and explain their reasoning correlating to their answers given. The questions examined aspects such as overlooked characteristics, independent variables, facial recognition, and cognition without textual information. The data provided by the participants was then analyzed to assist in explaining the emotional and psychological issues associated with facial recognition. These issues became apparent in the experiment, evidenced by graphs correlating with each data face.

This thesis found that data faces are a useful tool for data representation when utilized to the best of their abilities—such as inputting the most definitive or important statistics into the most prominent facial features. However, the answers of participants were found to be predisposed by emotional factors influencing the effectiveness of these data faces. Participants reported emotions associated with each data face and found to drastically favor the more pleasantly appearing expressions. Similarly, there was an apparent inability to distinguish minute changes in angle, curvature, or line length, perhaps due to a lack of emphasis on the importance of differential characteristics.

When the identical data is represented by both a data face and a graph, there was evidence of a slight loss of information in terms of severity. This loss appeared in the mouth in particular—perhaps due to a lack of recognition by the respondents concerning differences in curvature. In other words, both smiling mouths and both frowning mouths were frequently confused for one another. When presented with a graph of a different nature representing the same city with the same data, however, respondents understood their mistakes and began closer study of features.

Facial recognition and recognition of expression certainly plays a factor in deciphering the data provided by data faces. Every respondent in this study was able to easily associate an emotion with each data face and admittedly allowed this to assist in choosing their ideal locations. This may be considered a limitation with the use of data faces. However, with an understanding of this information, augmented data faces can visually present information with increased accuracy and effectiveness amongst observers.

Further experiments studying the effectiveness of data faces should be conducted with a larger sample size and the addition of more facial characteristics (such as hair). While the findings of this study are not conclusive, the results assist in directing further aspects of study, such as those with larger datasets and more in-depth data face creation. Ultimately, additional study is required to comprehend the psychological factors associated with facial data representation entirely.

REFERENCES

Beynon. D (2002). Information Systems: An introduction to informatics in organisations. Basingstoke, UK: Palgrave Macmillan.

Bernhard Flury and Hans Riedwyl (December 1981). "Graphical Representation of Multivariate Data by Means of Asymmetrical Faces" (PDF). Journal of the American Statistical Association. American Statistical Association. 76 (376): 757–765.

Białostocki, Jan, Iconography, Dictionary of The History of Ideas, Online version, University of Virginia Library, Gale Group, 2003

Caesar, Michael (1999). Umberto Eco: Philosophy, Semiotics, and the Work of Fiction. Wiley-Blackwell. p. 55.

Caplan, Jane (2000): Written on the Body: the Tattoo in European and American History, (ed.) Princeton University Press

Deely, John. (2006), "On 'Semiotics' as Naming the Doctrine of Signs",Semiotica 158.1/4 (2006), 1–33.

Eco, Umberto. (1986) Semiotics and the Philosophy of Language. Bloomington: Indiana University Press.

Karine Sergerie and Martin Lepage, (2005), A face to remember: Emotional expression modulates prefrontal activity during memory formation, NeuroImage 24(2):580-5. PubMed.

Herman Chernoff (1973). "The Use of Faces to Represent Points in K-Dimensional Space Graphically" (PDF). Journal of the American Statistical Association. American Statistical Association.

Mark W. Jones., (2001), Facial reconstructing using volumetric data, University of Wales Swansea, London.

Negishi, Mayumi. "Meet Shigetaka Kurita, the Father of Emoji". Wall Street Journal. RetrievedAugust 16, 2015.

APPENDIX

Appendix 1: New list in Aristotle's categories

Categories	Aristotle's Term	Greek	Examples
Substance/Essence	"substance" "this" "what-it-is"	*ousia* *tode ti* *ti esti*	man, horse Socrates "Socrates is a man"
Quantity	How much	*poson*	four-foot, five-foot
Quality	What sort	*poion*	white, literate
Relation	related to what	*pros ti*	double, half, greater
Location	Where	*pou*	in the Lyceum, in the marketplace
Time	When	*pote*	yesterday, last year
Position	Being situated	*keisthai*	lies, sits
Habit	Having, possession	*echein*	is shod, is armed
Action	Doing	*poiein*	cuts, burns
Passion	Undergoing	*paschein*	is cut, is burned

Appendix 2: Massachusetts information
Source from: https://www.neighborhoodscout.com/ma/crime/


Massachusetts Information

POPULATION
6,794,422

DATA | FIND A REAL ESTATE AGENT

Massachusetts Crime Data — Crime Data FAQ

Annual Crimes

VIOLENT	PROPERTY	TOTAL
26,562	114,871	141,433
annual crimes per 1,000 residents		
3.91	16.91	20.82

Violent Crime per 1,000 residents

Appendix 3: Tyson's Tattoo

Appendix 4: Massachusetts data.

1	City	Robbery	Violent crime	Murder and nonnegligent manslaughter	Rape (revised definition)1	Rape (legacy definition)2	Aggravated assault	Property crime	Burglary	Larceny-theft	Motor vehicle theft	Arson3
2	Abington	7	48	0	5		36	297	82	201	14	1
3	Acton	2	8	0	1		5	178	31	144	3	0
4	Acushnet	0	19	0	2		17	74	27	40	7	2
5	Adams	3	21	0	3		15	157	40	109	8	1
6	Agawam	3	41	0	12		26	270	105	130	35	1
7	Amesbury	2	42	0	3		37	239	40	191	8	2
8	Amherst	3	70	0	23		44	367	76	274	17	5
9	Andover	1	20	0	3		16	314	39	255	20	0
10	Arlington	10	55	3	7		35	516	126	365	25	3
11	Ashburnhar	1	5	0	2		2	65	20	44	1	0
12	Ashby	1	7	0	1		5	43	23	18	2	0
13	Ashland	0	18	0	2		16	206	91	109	6	0
14	Athol	3	53	0	7		43	154	55	97	2	4
15	Attleboro	16	108	0	16		76	942	171	702	69	2
16	Auburn	7	47	0	4		36	530	77	437	16	1
17	Avon	4	10	0		0	6	252	29	220	3	
18	Ayer	1	12	0	1		10	137	51	84	2	0
19	Barnstable	30	302	0	18		254	1,191	289	835	67	8
20	Barre	0	26	0	6		20	53	23	27	3	1
21	Becket	0	4	0		0	4	27	20	7	0	
22	Bedford	3	6	0	1		2	74	11	62	1	0
23	Belchertown	1	21	0	4		16	167	41	121	5	0
24	Bellingham	8	35	0	3		24	291	55	225	11	3
25	Belmont	2	24	0	4		18	241	56	172	13	3
26	Berkley	1	9	0	0		8	45	21	23	1	0
27	Berlin	0	0	0	0		0	14	0	14	0	0
28	Bernardston	1	4	0	0		3	15	2	11	2	0
29	Beverly	9	88	0	7		72	574	77	469	28	1
30	Billerica	8	48	0	7		33	468	88	352	28	5
31	Blackstone	2	17	0	3		12	84	20	60	4	0
32	Bolton	2	9	0	0		7	44	18	25	1	0
33	Boston	1,868	5,037	39		279	2,851	17,853	3,096	13,147	1,610	
34	Bourne	8	91	0	14		69	465	215	235	15	4
35	Boxborough	1	8	0	0		7	24	4	17	3	0
36	Boxford	1	4	0	0		3	34	9	23	2	0
37	Braintree	18	83	0	5		60	771	97	648	26	1
38	Brewster	0	7	0	0		7	132	50	78	4	1
39	Brimfield	0	1	0	0		1	21	9	12	0	0
40	Brockton	228	1,162	9	87		838	3,189	865	2,100	224	20
41	Brookline	16	85	0	6		63	676	107	549	20	1
42	Burlington	6	33	0	3		24	501	32	453	16	0
43	Cambridge	114	361	2	24		221	2,907	401	2,402	104	10
44	Canton	7	59	0	3		49	188	43	140	5	1
45	Carver	1	35	0	6		28	185	56	123	6	2
46	Charlton	2	22	0	4		16	113	33	77	3	1
47	Chatham	0	11	0	0		11	123	20	101	2	1
48	Chelmsford	4	22	0	1		17	488	49	422	17	0
49	Chelsea	175	458	5	28		250	1,234	209	874	151	4
50	Chicopee	81	272	1	22		168	1,563	459	995	109	4
51	Chilmark	0	1	0	0		1	8	1	7	0	0
52	Clinton	2	10	0	0		8	40	9	31	0	0
53	Cohasset	0	9	0	0		9	69	20	46	3	0
54	Concord	1	17	0	1		15	184	20	162	2	0
55	Dalton	0	5	0	2		3	69	20	48	1	2
56	Danvers	6	51	1	3		41	761	54	684	23	0
57	Dartmouth	9	49	0	5		35	914	157	732	25	1
58	Dedham	7	15	0	3		5	518	47	446	25	0
59	Deerfield	2	7	0	0		5	86	16	67	3	0
60	Dennis	8	80	0	13		59	505	211	276	18	3
61	Dighton	2	10	0		1	7	30	14	16	0	
62	Douglas	0	10	0	1		9	58	33	21	4	0
63	Dover	0	2	0	0		2	36	5	31	0	0
64	Dracut	8	20	0	4		8	494	75	392	27	0
65	Dudley	0	34	0	3		31	90	39	48	3	0
66	Dunstable	1	7	0	1		5	22	6	15	1	0
67	Duxbury	0	2	0	0		2	59	16	43	0	0
68	East Bridgev	2	31	0	0		29	200	33	158	9	0
69	East Brookfi	0	5	0	0		5	16	3	11	2	0
70	Eastham	0	11	0	0		11	100	16	82	2	1
71	Easthampto	2	19	0	3		14	159	35	118	6	2
72	East Longme	0	28	0	3		25	297	58	227	12	1
73	Easton	6	33	0	5		22	246	67	160	19	0
74	Edgartown	0	8	0	0		8	129	32	89	8	0
75	Egremont	0	1	0		0	1	23	2	21	0	
76	Erving	0	6	0	1		5	22	6	15	1	2
77	Essex	0	0	0		0	0	29	7	19	3	
78	Everett	61	202	1	19		121	898	186	621	91	0
79	Fairhaven	5	64	1	7		51	364	66	285	13	0

Appendix 4: Massachusetts data.

79	Fairhaven	5	64	1	7		51	364	66	285	13	0
80	Fall River	225	944	0	73		646	2,364	698	1,443	223	32
81	Falmouth	18	102	2	11		71	899	467	392	40	2
82	Fitchburg	35	307	2	33		237	1,060	275	705	80	7
83	Framinghar	23	172	0	4		145	1,025	217	735	73	3
84	Franklin	6	7	0	0		1	171	31	138	2	0
85	Freetown	2	21	0	3		16	137	54	74	9	2
86	Gardner	10	93	0	13		70	571	190	366	15	2
87	Georgetowr	0	3	0	1		2	64	17	46	1	0
88	Gill	0	1	0	0		1	12	5	7	0	0
89	Gloucester	0	22	0	9		13	383	33	345	5	0
90	Goshen	0	0	0	0		0	2	1	1	0	0
91	Grafton	4	22	0	2		16	121	40	76	5	0
92	Granby	0	9	0	0		9	75	13	58	4	0
93	Granville	0	2	0		0	2	8	4	4	0	
94	Great Barrin	1	30	0	8		21	124	27	88	9	1
95	Groton	0	8	0	2		6	86	35	51	0	0
96	Groveland	1	1	0	0		0	17	6	11	0	0
97	Halifax	0	8	0	2		6	77	32	41	4	0
98	Hamilton	0	3	0	0		3	50	8	41	1	0
99	Hampden	1	3	0	0		2	55	18	33	4	0
100	Hanover	0	2	0	2		0	211	34	171	6	0
101	Hanson	1	19	0	3		15	86	23	57	6	0
102	Hardwick	0	5	0	1		4	31	8	20	3	0
103	Harvard	0	4	0	2		2	37	23	14	0	3
104	Harwich	1	24	0	2		21	251	120	124	7	0
105	Hatfield	2	5	0	1		2	10	4	6	0	0
106	Haverhill	57	414	2	25		330	1,618	431	1,052	135	5
107	Hingham	2	17	0	2		13	382	84	287	11	0
108	Holbrook	11	39	0		0	28	183	74	100	9	
109	Holden	1	8	0	1		6	96	21	71	4	0
110	Holliston	0	14	0	0		14	75	25	44	6	1
111	Holyoke	108	419	4	42		265	2,664	484	2,067	113	9
112	Hopedale	0	18	0	2		16	57	23	32	2	1
113	Hopkinton	0	3	0	1		2	93	6	86	1	1
114	Hubbardsto	0	6	0	0		6	44	20	21	3	0
115	Hudson	1	16	0	2		13	209	33	169	7	0
116	Hull	0	42	1	5		36	141	42	95	4	2
117	Ipswich	0	10	0	0		10	122	20	96	6	1
118	Kingston	3	20	0	3		14	191	28	152	11	0
119	Lakeville	1	10	1	2		6	183	62	115	6	0
120	Lancaster	1	8	0	3		4	80	30	48	2	1
121	Lawrence	283	776	1		19	473	2,273	433	1,012	828	
122	Lee	0	7	0	3		4	98	23	72	3	0
123	Leicester	0	13	0	1		12	250	17	223	10	1
124	Lenox	0	4	0	1		3	120	35	83	2	1
125	Leominster	34	263	1	25		203	1,235	244	939	52	5
126	Lexington	3	17	0	0		14	232	45	185	2	2
127	Lincoln	1	6	1	1		3	37	8	29	0	1
128	Littleton	0	9	0	4		5	92	15	75	2	1
129	Longmeado	3	14	0	2		9	156	27	121	8	1
130	Lowell	192	625	4	40		389	3,379	759	2,300	320	18
131	Ludlow	8	32	0	7		17	320	78	232	10	0
132	Lunenburg	1	22	0	5		16	210	62	145	3	0
133	Lynn	189	814	2	39		584	2,398	510	1,635	253	7
134	Lynnfield	1	2	0	0		1	133	14	111	8	0
135	Malden	79	205	1	3		122	945	240	618	87	3
136	Manchester-	0	9	0	1		8	28	5	23	0	0
137	Mansfield	3	47	0	8		36	352	144	196	12	2
138	Marblehead	0	17	0	1		16	159	21	134	4	0
139	Marion	1	7	0	1		5	69	9	60	0	0
140	Marlboroug	17	156	0	18		121	756	119	605	32	2
141	Marshfield	3	54	0	5		46	229	29	186	14	2
142	Mashpee	3	45	0	4		38	284	61	214	9	2
143	Mattapoiset	1	6	0	1		4	66	16	45	5	0
144	Maynard	0	13	0	0		13	99	14	79	6	0
145	Medfield	0	12	0	2		10	58	16	42	0	0
146	Medford	29	112	0	10		73	898	160	672	66	2
147	Medway	0	1	0	0		1	111	20	89	2	0
148	Melrose	4	21	0	3		14	262	69	181	12	0
149	Mendon	0	5	0	0		5	57	17	39	1	0
150	Merrimac	1	11	0	0		10	56	24	29	3	0
151	Methuen	30	106	0	4		72	1,127	177	830	120	1
152	Middleboro	7	70	1	6		56	443	126	300	17	1
153	Middleton	0	11	0	0		11	75	4	70	1	0
154	Milford	7	58	0	13		38	522	56	449	17	1
155	Millbury	5	38	0	7		26	275	58	206	11	4
156	Millville	1	2	0	1		0	32	5	27	0	0
157	Milton	9	21	1	1		10	391	95	285	11	0
158	Monson	0	27	0	3		24	101	35	53	13	1
159	Montague	3	45	0	4		38	222	72	147	3	2
160	Monterey	0	0	0	0		0	4	1	2	1	0

160	Monterey	0	0	0	0		0	4	1	2	1	0
161	Nahant	0	8	0	1		7	18	4	13	1	0
162	Nantucket	3	41	0	8		30	330	29	286	15	0
163	Natick	3	36	1	2		30	544	62	468	14	2
164	New Bedfor	256	1,039	6	100		677	3,437	916	2,186	335	20
165	Newburypor	3	31	0	3		25	177	15	154	8	0
166	Newton	18	74	0	11		45	807	205	584	18	4
167	Norfolk	0	0	0	0		0	41	13	28	0	0
168	North Adam	8	133	1	21		103	585	175	395	15	1
169	Northampto	14	130	0	29		87	800	126	647	27	4
170	North Ando	0	10	0	0		10	479	26	440	13	0
171	North Attleb	2	2	0	0		0	406	31	366	9	0
172	Northborou	4	18	0	1		13	182	38	140	4	0
173	Northbridge	1	38	0	15		22	308	93	210	5	2
174	North Brook	0	34	0	3		31	29	12	15	2	0
175	North Readi	3	12	0	0		9	134	17	112	5	0
176	Norton	2	8	0	0		6	90	36	53	1	0
177	Norwell	0	12	0	1		11	122	28	89	5	0
178	Norwood	9	18	0	1		8	386	85	281	20	2
179	Oak Bluffs	0	18	0	0		18	137	26	107	4	0
180	Orange	3	25	0	5		17	166	52	110	4	4
181	Orleans	1	9	0	1		7	148	24	123	1	1
182	Oxford	4	27	0	1		22	212	42	160	10	1
183	Palmer	5	38	0	8		25	210	88	111	11	0
184	Paxton	0	4	0	2		2	25	10	14	1	0
185	Peabody	24	115	0	14		77	1,061	121	877	63	2
186	Pelham	0	0	0	0		0	5	4	1	0	0
187	Pembroke	3	33	0	2		28	159	38	114	7	0
188	Pepperell	1	22	0	1		20	126	32	89	5	1
189	Pittsfield	29	111	0	31		51	1,352	481	827	44	3
190	Plainville	1	2	0	0		1	136	31	101	4	0
191	Plymouth	18	184	0	9		157	806	236	540	30	1
192	Plympton	1	7	0	1		5	42	17	24	1	0
193	Princeton	0	7	0	1		6	32	14	17	1	0
194	Provincetow	0	17	0	1		16	105	17	85	3	1
195	Quincy	87	361	0	35		239	1,647	461	1,113	73	4
196	Randolph	31	176	0	16		129	536	126	374	36	2
197	Raynham	9	26	0	2		15	395	68	310	17	1
198	Reading	3	10	0	0		7	184	51	122	11	0
199	Rehoboth	1	13	0	1		11	143	51	83	9	0
200	Revere	82	317	0	19		216	1,161	182	813	166	8
201	Rochester	0	3	0	1		2	50	18	28	4	0
202	Rockport	1	6	0	1		4	8	5	3	0	0
203	Rowley	1	6	0	1		4	24	7	14	3	0
204	Rutland	1	25	1	3		20	52	11	37	4	0
205	Salem	22	128	0	3		103	1,238	135	1,058	45	8
206	Salisbury	5	45	1	7		32	178	57	113	8	0
207	Sandwich	1	31	0	8		22	259	70	181	8	1
208	Saugus	15	80	0	10		55	648	84	527	37	2
209	Scituate	3	24	0	3		18	135	26	105	4	0
210	Seekonk	4	33	1	0		28	351	50	280	21	0
211	Sharon	0	1	0	0		1	76	26	46	4	0
212	Sherborn	1	3	0	1		1	39	13	23	3	2
213	Shirley	0	11	0	0		11	44	20	23	1	0
214	Shrewsbury	2	6	0	0		4	269	54	204	11	0
215	Somerset	2	28	0	3		23	263	45	208	10	0
216	Somerville	51	191	0	18		122	1,525	426	998	101	21
217	Southampto	1	4	0	2		1	46	9	36	1	0
218	Southborou	0	2	0	1		1	58	31	24	3	0
219	Southbridge	14	82	0	11		57	407	149	236	22	10
220	South Hadle	4	38	0	4		30	276	97	167	12	1
221	Southwick	1	8	0	0		7	122	41	70	11	2
222	Spencer	1	29	0	4		24	161	34	116	11	0
223	Springfield	598	1,673	22	89		964	7,113	2,360	4,018	735	56
224	Sterling	1	12	0	3		8	72	21	48	3	0
225	Stockbridge	0	4	0	0		4	35	15	20	0	0
226	Stoneham	2	11	0	0		9	217	52	154	11	1
227	Stoughton	15	80	1	9		55	532	74	422	36	0
228	Stow	1	4	0	0		3	24	7	17	0	0
229	Sturbridge	5	30	0	7		18	175	25	146	4	2
230	Sudbury	0	0	0	0		0	94	12	80	2	0
231	Sutton	0	4	0	1		3	106	46	55	5	0
232	Swampscott	5	26	0	1		20	266	58	204	4	0
233	Swansea	7	34	1	5		21	350	54	279	17	2
234	Taunton	52	209	0	15		142	962	433	485	44	5
235	Templeton	2	22	0	2		18	81	21	57	3	0
236	Tewksbury	9	141	0	14		118	502	68	398	36	2
237	Townsend	2	7	0	2		3	117	60	57	0	0
238	Truro	0	5	0	0		5	35	12	23	0	0
239	Tyngsboro	2	16	0	0		14	129	26	97	6	1
240	Upton	1	7	0	1		5	60	22	35	3	0
241	Uxbridge	2	21	0	3		16	176	52	115	9	2
242	Wakefield	1	49	0	11		37	313	66	233	14	0

Appendix 4: Massachusetts data.

242	Wakefield	1	49	0	11		37	313	66	233	14	0
243	Wales	0	1	0	0		1	10	4	4	2	0
244	Walpole	3	11	0	1		7	321	36	279	6	0
245	Waltham	28	141	1	22		90	952	212	694	46	6
246	Ware	3	42	0	2		37	238	36	198	4	5
247	Wareham	21	163	0	20		122	806	222	554	30	12
248	Warren	1	33	0	2		30	60	17	39	4	0
249	Watertown	5	53	0	6		42	461	61	388	12	0
250	Wayland	0	5	0	0		5	42	16	24	2	0
251	Webster	21	122	0	9		92	426	85	319	22	8
252	Wellesley	1	15	0	0		14	202	54	142	6	0
253	Wellfleet	0	4	0	0		4	38	13	24	1	2
254	Westboroug	3	14	0	3		8	149	15	131	3	0
255	West Boylstc	0	5	0	4		1	109	23	83	3	0
256	West Bridge	4	23	0	1		18	142	23	111	8	2
257	Westfield	25	103	1	20		57	661	174	464	23	3
258	Westford	1	14	0	3		10	121	26	93	2	3
259	Westhamptc	0	0	0		0	0	5	3	2	0	
260	Westminster	0	9	0	0		9	189	17	169	3	0
261	West Newbu	0	6	0	1		5	24	6	18	0	0
262	Weston	0	3	0	0		3	65	14	49	2	0
263	Westport	2	27	0	3		22	198	64	124	10	1
264	West Spring	40	163	0	22		101	1,321	250	985	86	6
265	West Tisbury	0	1	0	0		1	26	8	16	2	0
266	Westwood	4	20	0	4		12	111	17	87	7	0
267	Weymouth	27	195	1	23		144	784	145	604	35	4
268	Whately	0	6	0	0		6	17	3	11	3	0
269	Whitman	3	53	0	8		42	217	43	164	10	1
270	Wilbraham	1	17	0	4		12	235	52	170	13	1
271	Williamsburg	0	4	0	0		4	36	12	22	2	0
272	Williamstow	1	14	0	3		10	216	43	173	0	0
273	Wilmington	6	29	0	5		18	253	65	180	8	2
274	Winchendor	2	50	0	4		44	217	19	189	9	5
275	Winchester	2	10	0	0		8	194	25	167	2	0
276	Winthrop	4	58	0	2		52	211	82	117	12	2
277	Woburn	15	83	0	12		56	581	91	469	21	1
278	Worcester	483	1,750	9	22		1,236	6,239	1,916	3,924	399	7
279	Wrentham	0	8	0	1		7	262	48	206	8	0
280	Yarmouth	8	202	0	17		177	668	200	446	22	2

Appendix 5: Robbery data in line graph.