Visualization of Time Series Data by Statistical Shape Analysis on Fertility Rate and Education in Indonesia

Yukari Shirota
Department of Management, Faculty of Economics, Gakushuin University, Tokyo, Japan
Email: yukari.shirota@gakushuin.ac.jp

Alfan Presekal and Riri Fitri Sari
Computer Engineering, Department of Electrical Engineering, Universitas Indonesia, Depok, Indonesia
Email: presekal@ui.ac.id, riri@ui.ac.id

Abstract—**Visualization is very effective when we analyze the time series data. In the paper, we shall illustrate change of the whole trend on time series data as shapes, as well as local changes. The method we used is the statistical shape analysis which can extract separately the Affine and non-Affine transformation parts from the time change deformation. The method is helpful to see a local movement of each landmark data, compared to other neighbors. In the paper, we shall conduct the Indonesia province comparison, concerning the total fertility rate and the education status between 2007 and 2012. From the visualization, we can easily understand the time series changes.**

Index Terms—**statistical shape analysis, Affine/non-Affine transformation, partial warp eigenvectors, fertility rates, education level**

I. Introduction

Visualization is very useful for showing the changes in global trend particularly when we analyze the time series data. In this paper, we shall show a new method of the visualization. The data we used is the total fertility rate and education status in Indonesia province data. We employed a new method of "statistical shape analysis." This is one of eigenvector based analyses such as principal component analysis (PCA) [1] and singular value decomposition (SVD) [2], [3].

This eigenvector based analysis is reliable, compared to the deep neural network models, especially when we analyze non-image data. The eigenvector based analysis will always offer the same result. So far as the input data is reliable, the result can be trusted. Therefore, we shall always first analyze the data by the eigenvector based method before we try to analyze the data by machine learning methods such as Random Forest Algorithm [1]. Another reason of the use of statistical shape analysis in this paper is that the shape analysis is most suitable to analyze and visualize this province comparison result. The statistical shape analysis method was originally developed for image analyses in biological or medical fields [4]-[7]. However, we found that the power of this analysis method could also be implemented in economics and socio-science. Subsequently, as one of the pioneer of the application of this method to economics data analysis, we have reported the results in some publications [8]-[14].

Using this statistical analysis, we can measure the change in the shape of an object, so-called deformation. The problem on transforming data sets in different size, orientation and shape of an object into a coordinate system is a complex task, using a coordinate system called register marks or landmarks. By this analysis, however, we can quantify the shape of an object by eliminating the information of location, rotation, and scale.

Though we shall describe the analysis results concerning Indonesia provinces in this paper, our purpose here is to propose the statistical shape analysis for big time series data analyses. To grasp the whole trend change of the big data, the Affine transformation of the statistical shape analysis is effective to extract of the essence.

Indonesia is an emerging country and is a very important economy from business point of view. We would like to analyze the relationship between total fertility rates and an education status by provinces in Indonesia. The Total Fertility Rate (TFR) is the number of children that would be born to a woman over her lifetime. The data we used is from the Statistics Bureau of Indonesia (https://www.bps.go.id/). The original website name is "Badan Pusat Statistik Indonesia" (BPS), a non-department government agency which directly report to the president.

The BPS is instituted by Law Number 16, 1997 on Statistics; Government Regulation Number 51, 1999 on Statistics Undertakings [8]. Therefore, this website's data is very reliable and suitable for researches. In Section 2, using the shape analysis, we will analyze the total trend of the change by provinces. In Section 3, the local change specific to some provinces will be explained, using the non-Affine transformation part of the shape analysis results. Finally we will conclude the paper.

Manuscript received January 10, 2019; revised May 7, 2019.

doi: 10.12720/jait.10.2.60-65

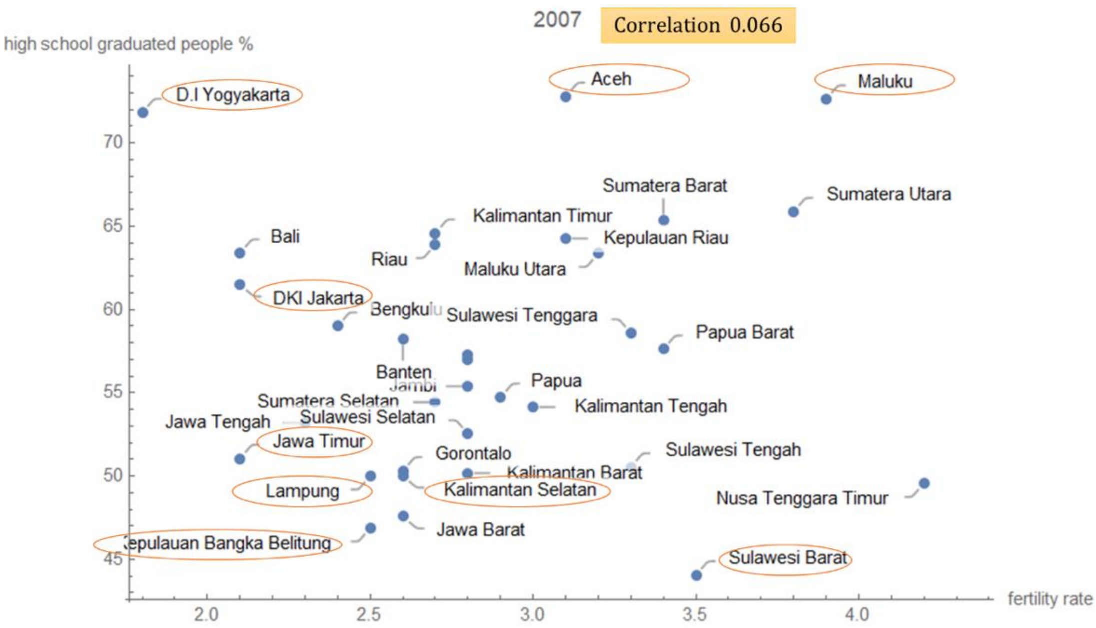

Figure 1. Relationship between fertility rate and senior high school graduated people % in 2007.

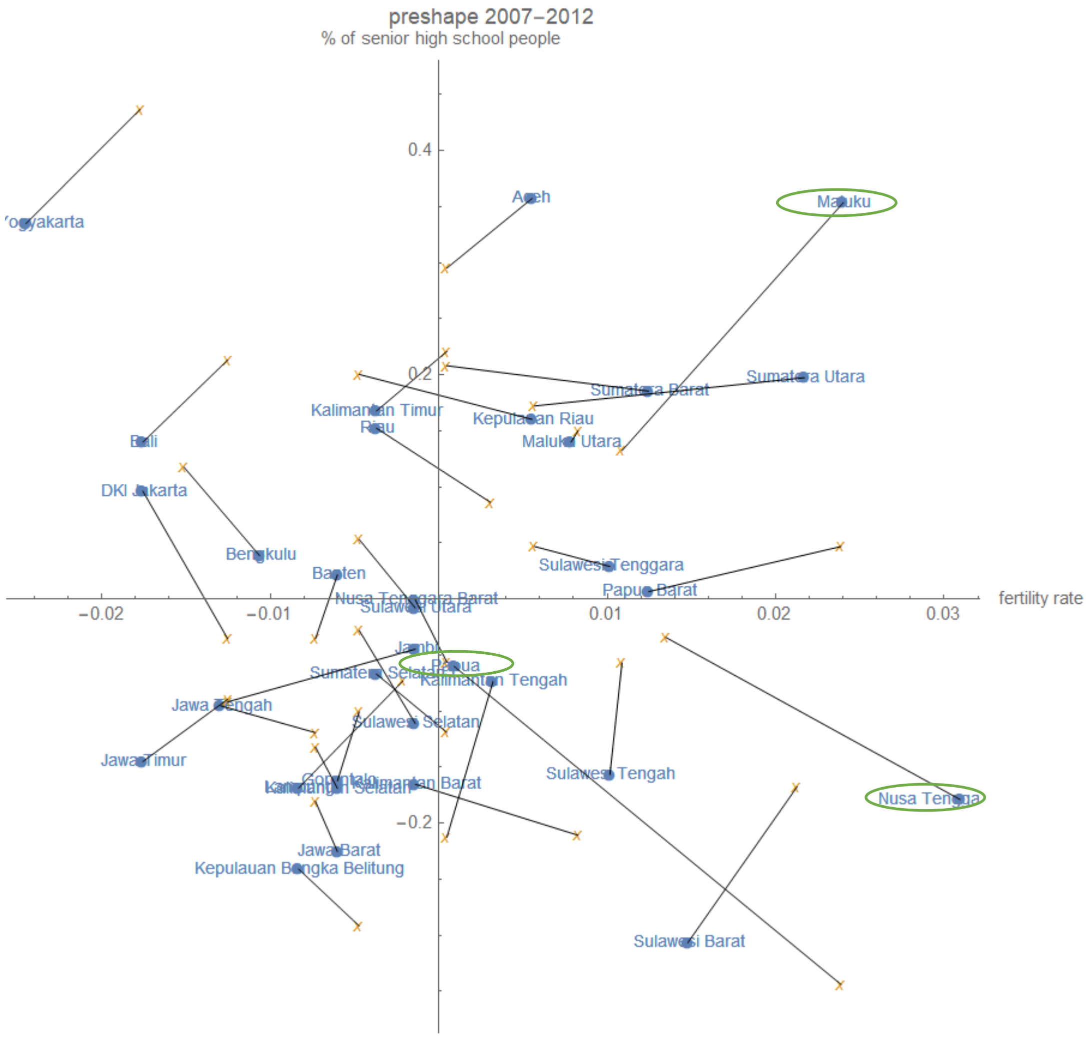

Figure 2. Change between pre-shapes in 2007 and 2012 on fertility rate and senior high school graduated people %.

II. Total Trend of the Change

In this section, we shall conduct the shape analysis on a deformation between in 2007 and in 2012. Fig. 1 shows the relationship between Total Fertility Rates (TFR) and senior high school graduated people percentages in 2007. In the later, we use “high school %” as its abbreviation. Concerning TFRs in Fig. 1, the largest figure/point is 4.2 children in Nusa Tenggara Timur and the smallest figure is 1.8 children in D. I. Yogyakarta. In Yogyakarta, there are many universities, therefore approximately one out of two persons in this city is a university graduated person. We can say that Yogyakarta city has the highest education level in Indonesia. The second smallest ones are Bali, DKI Jakarta, and Jawa Timur of which the figure is 2.1 children. Concerning the high school %, Fig. 1 shows that Aceh and Maluku have the highest percentage which is over 72%. The Pearson correlation coefficient between TFRs and high school % is 0.066. In the year 2012 data, the correlation coefficient is -0.234, which shows a weak negative relationship. Then, we can not say that there is a strong relationship between TFRs and the high school %.

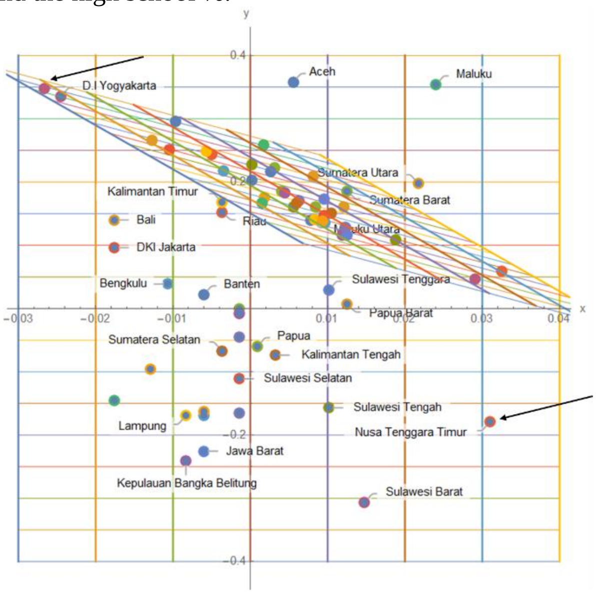

Figure 3. The original 2007 pre-shape and one after the Affine transformation.

Then we conduct the shape analysis. In the shape analysis, the given coordinates are standardized using the centroid size [15]. The standardized data is called pre-shapes. The pre-shape consists of a set of dimensionless data. In this case, a landmark in a pre-shape corresponds to a province. Therefore, in this case, a pre-shape consists of 33 province data. In Fig. 2, the 2007 pre-shape is depicted by a circle mark and the 2012 pre-shape is depicted by a cross mark. Seeing pre-shapes figures, we can see the relative position of each province among all provinces. As shown in the Fig. 2, Papua’s movement is remarkable, compared to the movement of other provincess. In Papua, the TFR has largely increases and the high school percentage has relatively decreases. Let’s see the raw data of high school %. The decline provinces are only Maluku (-4.31%) and Papua (-4.71%). However, when we see the pre-shape difference in Fig. 2, we find that D.K.I. Jakarta, Riau, and Aceh have negative value of directed changes. This means that D.K.I. Jakarta’s high school % increased, but others progress were still greater. The advantage of the statistical shape analysis is that we can extract a relative growth speed like this. Concerning TFRs, we can see that TFRs in Maluku and Nusa Tenggara Timur largely decrease, compared to others.

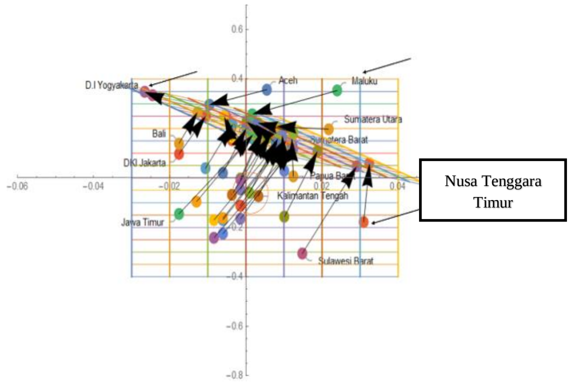

Figure 4. The original 2007 pre-shape and one after the Affine transformation.

Let us extract and see the Affine transformation part first (See Fig. 3). In the 2007 original pre-shape, the transformation grids are orthogonal. The pre-shape after the Affine transformation has the skewed transformation grids as shown in Fig. 3. In the figure, the same colored point means the same province. In the shape analysis, the Affine-transformation part expresses the total movement trend of the deformation. On the other hand, the non-Affine transformation part expresses a local movement which occurs in a local specific area or a province. Let us compare the shape analysis with a linear regression method [1]. A linear regression illustrates the total trend only. On the other hand, the shape analysis can extract non-Affine transformation part, too.

We analyze the Affine transformation part shown in Fig. 3 and 4. In Fig. 4, the same provinces are connected by arrows. The arrow in the figure depicts the change of a province. As shown there, the high school % (the vertical axis) totally increase and a homogenization can be seen. Many provinces of which high school % were lower in 2007 have become concentrated to the center by this Affine transformation skew, which means that an educational level homogenization. It is a reasonable change during the five years in Indonesia. Concerning the TFR, two provinces on both sides, Yogyakarta and Nusa Tenggara Timur do not move the positions. Then we cannot say a homogenization of TFRs.

III. Local Time Series Change

In this section, we shall analyze the local movement from 2007 to 2012. The local movement can be illustrated by non-Affine transformation of pre-shapes (See in Fig. 5). As shown there, the local movement is complicated. In Fig. 6, the 2012 pre-shape is shown, as an addition to the Affine transformation and the non-Affine transformation. For example, let’s see Maluku. Maluku has a large TFR decline in the Affine transformation (See Fig. 4). However, as the local TFR movement is positive as shown in Fig. 5, finally the TFR change became a small decrease (See Fig. 6). As another example, let's see Papua changes on high school %. The change of Papua in the Affine transformation was positive (See Fig. 4). However, Papua's high school % decreased largely. Then the local movement of Papua has a large negative value as shown in Fig. 5. Finally, the change of Papua is not a large negative as shown in Fig. 6.

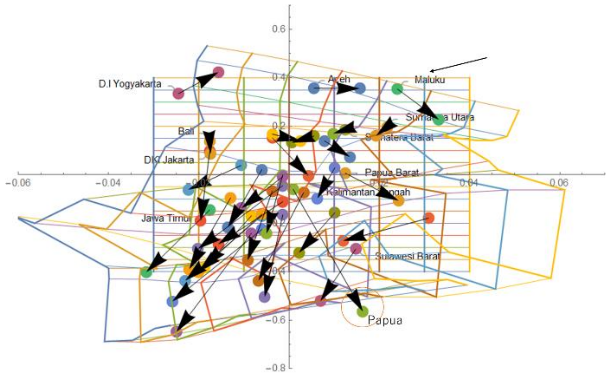

Figure 5. The original 2007 pre-shape and one after the non-Affine transformation.

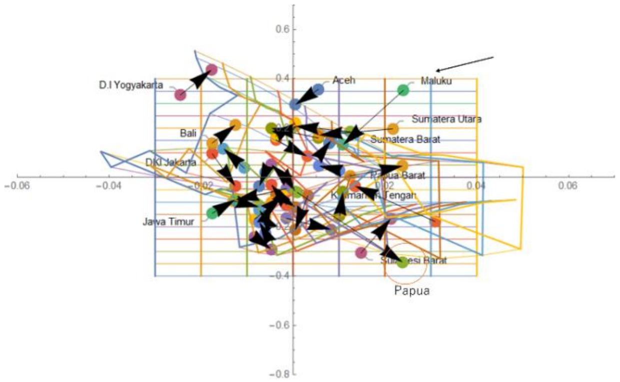

Figure 6. The original 2007 pre-shape and one after both Affine and non-Affine transformations.

In the theory of the shape analysis, a non-Affine transformation can be decomposed to a set of partial warps. In this province data, the number of landmarks is 33. This is because there was no data of Kalimantan Utara data in 2007. Therefore, we can get the 30 (=33 – 3) partial warps. An individual partial warp has its amplitude which is an eigenvalue of the correspondent principal eigenvector. The 33 eigenvalues are shown in Fig. 7. Seeing the eigenvalues, we can see that the first partial warp is a dominant one because the amplitude is still larger than others. Then, we shall see the first partial warp.

In Fig. 8, the partial warp #1 and the original 2007 pre-shape are illustrated. There is an arrow which depicts a delta (difference) part of each province in the partial warp #1. The sum of the delta parts from #1 to #30 becomes the total non-Affine transformation delta part of the province. The total non-Affine transformation is shown in Fig. 5. In other words, we can decompose the non-Affine transformation to the 30 partial warps.

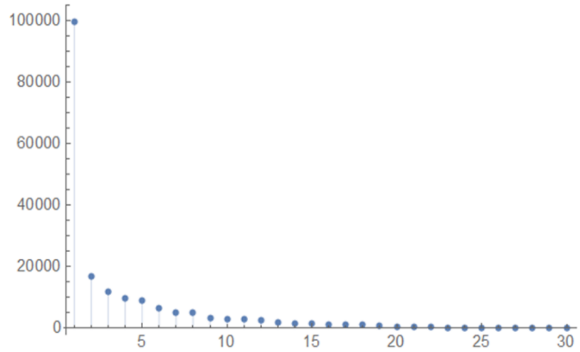

Figure 7. Eigenvalues of the 30 principal warps.

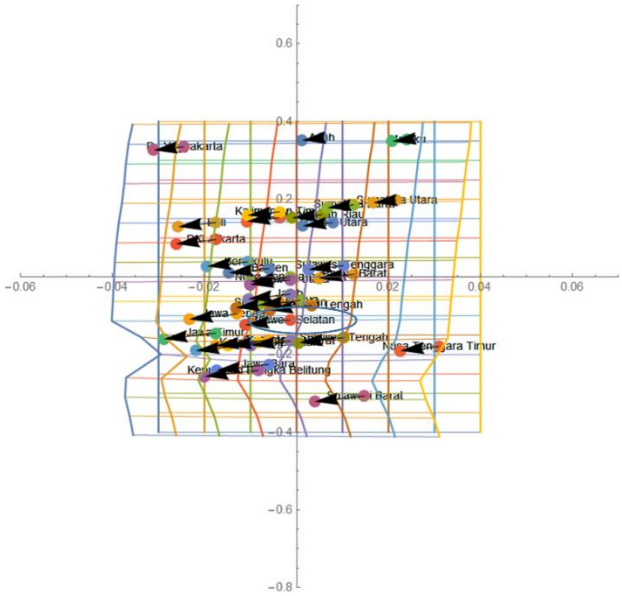

Figure 8. The original 2007 pre-shape and one after the partial warp #1.

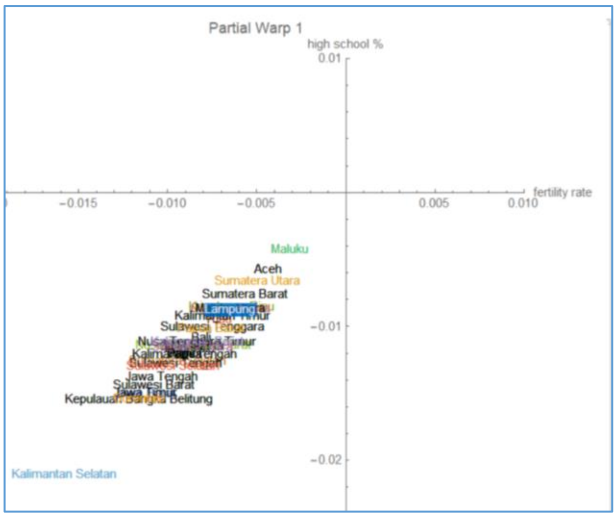

Figure 9. Differences of the partial warp #1.

In the partial warp #1, every province move shows the negative direction on the TFR axis. The largest negative change occurred in Kalimantan Selatan (See Fig. 8). Fig. 9 only shows differences of the province data. There we can see that Kalimantan Selatan has the largest negative change on high school %. All the provinces have the negative values on both TFRs and high school %.

As mentioned previously, the total trend of TFRs has not change, because Yogyakarta and Nusa Tenggara Timur do not move in terms of their the positions. Those local movements with the reverse direction, however, can be extracted in the partial warp #1 as shown in Fig. 8. Seeing the transformation grids, we can find that the TFR change of the high school % lower group is larger.

In the high school % lower group, there are two different local movements; The TFR decrease of Lampung is smaller than other neighbors, which makes the reverse directional transformation grid movement (See Fig. 8 and 9). The high school % movement in the partial warp #1 shows the decrease as well as one of TFRs. The partial warp #1 illustrates declines of both TFRs and the high school %.

Let's see the partial warp #2 in Fig. 10 and 11. From the scales of the both axes, we can see that the movement happens only on the high school %. The largest decline happened in Kalimantan Selatan.

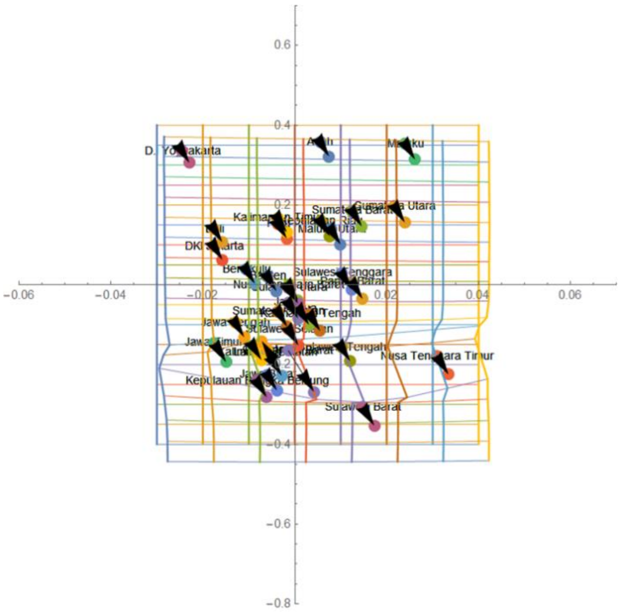

Figure 10. The original 2007 pre-shape and one after the partial warp #2.

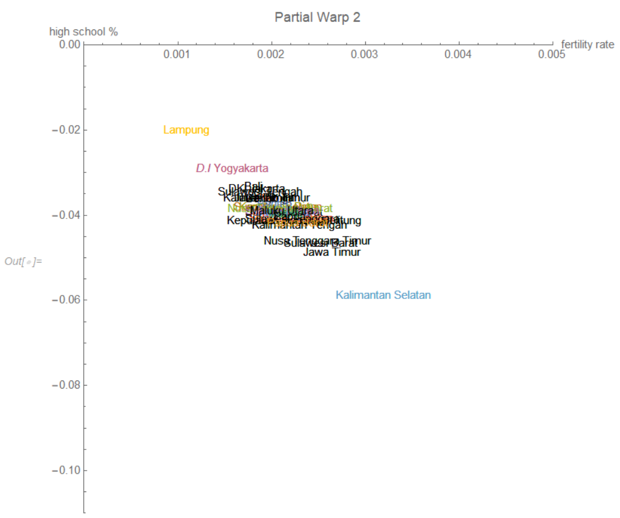

Figure 11. Differences of the partial warp #2.

IV. Conclusion

This paper presented a visualization of time series data changes by the statistical shape analysis. We analyzed and visualized the deformation of the relationship between fertility rates and percentages of senior high school graduated people by provinces in Indonesia. The target data were 2007 and 2012. Advantages of the shape analysis is that we can compare two shapes with the same landmarks relatively, because the shape is standardized by pre-shaping. Then we can get the relative position of the landmark among all the landmarks. By the shape analysis, from the deformation, an Affine transformation part and a non-Affine transformation part can be separately extracted and the non-Affine transformation part can be decomposed of a set of partial warps. The Affine transformation part expresses the total trend of the deformation and the non-Affine transformation part expresses the local change of a landmark. Seeing an individual partial warp, a remarkable growth/decline compared to others can be extracted. The shape analysis method is an excellent formal model to express the local movements on the time series data change.

Acknowledgment

The research was partly supported by funds from the Gakushuin University Research Institute for Economics and Management as the research project in 2018.

References

[1] J. Friedman, R. Tibshirani, and T. Hastie, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-Color Illustrations*, Springer, 2013.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] Y. Shirota and B. Chakraborty, "Visual explanation of eigenvalues and math process in latent semantic analysis," *Information Engineering Express, Information Engineering Express*, vol. 2, no. 1, pp. 87-96, 2016.

[4] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*, J. Wiley Chichester, 1998.

[5] K. Mardia, F. Bookstein, and J. Kent, "Alcohol, babies and the death penalty: Saving lives by analysing the shape of the brain," *Significance*, vol. 10, no. 3, pp. 12-16, 2013.

[6] K. V. Mardia, "Statistical approaches to three key challenges in protein structural bioinformatics," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 62, no. 3, pp. 487-514, 2013.

[7] M. Kanti, "Geometry-Driven statistics and its cutting edge applications: Celebrating four decades of leeds workshops," in *Proc. 33rd Leeds Annual Statistical Research Workshop*, 2015, vol. 33.

[8] C. Apriono, R. F. Sari, Y. Yano, and Y. Shirota, "Economic indicator evaluation based on shape deformation analysis of Indonesian Provinces statistics," *Gakushuin Economics Papers*, vol. 54, no. 3, pp. 1-22, 2017.

[9] Y. Shirota, R. F. Sari, T. Widiyani, and T. Hashimoto, "Visually do statistical shape analysis! As TUTORAL," in *Proc. IEEE International conference on Data Science and Advanced Analytics*, Tokyo, Japan, 2017.

[10] T. Widiyani, Y. Shirota, and R. F. Sari, "Morphometrics analysis on craniofacial differences of ancient humans," in *Proc. IEEE ICACOMIT*, 2017, pp. 22-27.

[11] S. Yukari, H. Takako, and S. Riri Fitri, "Visualization of time series statistical data by shape analysis (GDP ratio changes among

Asia countries),” *Journal of Physics: Conference Series*, vol. 971, no. 1, p. 012013, 2018.
[12] Y. Shirota, R. F. Sari, and T. Hashimoto, “Time series analysis on Indonesia Provinces economic development,” in *Proc. 5th Conference on Management and Sustainability in Asia*, Hiroshima, Japan, 2018, pp. 12-19.
[13] A. Presekal, R. F. Sari, and Y. Shirota, “Local change analysis of correlation of education level to GDP in Indonesia,” *Gakushuin Economics Papers*, vol. 55, no. 3, pp. 61-77, 2018.
[14] Y. Shirota, A. Presekal, and R. F. Sari, “Visualization of time series change on GDP per electricity by provinces in Indonesia,” in *Proc. of IES-KCIC 2018, Bali, Indonesia*, Oct. 2018, pp. 333-336.
[15] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis with Applications in R*, second ed., J. Wiley Chichester, 2016.

Yukari Shirota is a DSc. Prof. of Faculty of Economics, Gakushuin University. She graduated from the Department of Information Science, Faculty of Science, the University of Tokyo, Japan and then received a D.Sc. in computer science in 1998. As a researcher in the private sector, she conducted research for 13 years and then in 2001 she was involved in Faculty of Economics, Gakushuin University, Tokyo as Associate Professor. In 2002, she has become Professor, Faculty of Economics, Gakushuin University. In 2006 to 2007, she stayed at University of Oxford, Oxford, UK as an academic visitor. Research fields are visualization of data on the web, data visualization, social media analysis, and visual education methods for business mathematics.
For over 19 years, she has developed visual teaching materials for business mathematics and statistics as follows:
-https://www-cc.gakushuin.ac.jp/~20010570/mathABC/SELECTED/
-https://www-cc.gakushuin.ac.jp/~20010570/
-https://www-cc.gakushuin.ac.jp/~20010570/VDStat/
-https://www-cc.gakushuin.ac.jp/~20010570/mathABC/SELECTED/ShapeAnalysis/
-https://www-cc.gakushuin.ac.jp/~20010570/private/MAXIMA/
Prof. Shirota is a Board Member of the Japanese Operations Management and Strategy Association.

Alfan Presekal is a lecturer at the Electrical Engineering Department, Faculty of Engineering, Universitas Indonesia (UI). He received a BSc in Computer Engineering from University of Indonesia, a Master degree from Department of Computing Imperial Collage, London-UK with specialism in Secure Software System. He served as Director of Business Unit in Department of Electrical Engineering Universitas Indonesia from 2017 to 2018. Currently he is also working as research assistant in Internet of Things and Cyber Security Research Group. Alfan also active as officer at IEEE Indonesia Section. More info available at: profile.presekal.com

Riri Fitri Sari is a Professor of Computer Engineering at the Electrical Engineering Department, Faculty of Engineering, Universitas Indonesia (UI). She graduated with a BSc in Electrical Engineering from UI, a Master in Human Resources Management from the Atmajaya University Jakarta and an MSc in Software Systems and Parallel Processing from the Department of Computer Science, University of Sheffield, UK, funded by British Council Chevening Award, and a PhD in Computer Networks from the School of Computing, University of Leeds, UK. Her current main teaching and research area includes Computer Network, Internet of Things (IoT), Cloud Computing, Vehicle Ad Hoc Networks, and ICT implementation. She was the IT Director of the University of Indonesia from 2006-2014. She has also been the chairperson of UI GreenMetric Rankings of World Universities since 2010. Prof Riri Fitri Sari is a Senior Member of Institute of Electronic and Electrical Engineers (IEEE).