

Multivariate Analysis of University Student Engagement through Visualization Techniques

Erika Yunuen Morales Mateos¹, María Arely López Garrido¹,
José Alberto Hernández Aguilar², Carlos Alberto Ochoa Ortiz³,
Oscar Alberto González González¹

¹ Universidad Juárez Autónoma de Tabasco Cunduacán, Tabasco,
Mexico

² Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos,
Mexico

³ Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Chihuahua,
Mexico

{erika.morales, arely.lopez, alberto.ochoa}@uacj.mx, oscar.gonzalez@dais.ujat.mx,
jose_hernandez@uaem.mx

Abstract. The objective of this work is to apply visualization techniques for the representation of multivariate analysis results. The techniques were used to publicize the results of student engagement at universities of careers in technology at a university in the south of México. To measure student engagement, the UWES-S instrument was used, composed of vigor, dedication and absorption dimensions. The graphs presented in this paper are: Graphics Matrices, GGE Biplot and Chernoff faces, which were developed with the R language. It concludes on the utility that visualization techniques have in the multivariate analysis, given that in this it is studio was able to identify how students interact (genotypes) with the environment (variables of student engagement), identifying that the majority presents a stable level of student engagement. With the Chernoff faces, it was possible to visualize through the facial features in detail the presence of student engagement as a result medium level.

Keywords: multivariate analysis, visualization techniques, biplot charts, Chernoff faces, student engagement.

1 Introduction

Populations are commonly characterized by a simultaneous manifestation of many processes, where results at best can be measured, which is why it is necessary to have methods capable of considering at the same time, several characteristics to define in an integral way to an object of study [1]. In this way, the need to apply multivariate data analysis arises.

Similarly, there are techniques of visuals for multivariate data that allow exposing what happens in a set of data that is represented by several dimensions. The presentation of data using visual techniques helps to understand the behavior of data with a look, even complex data, so that information analysts can save time when making judgments about patterns, trends, variability, among others, of data. One of the most important advantages of using visual techniques to represent data is that they can display features that otherwise would have been complex or impossible [2].

Visualization techniques for this type of data challenge the combination of statistics and computer science. For the development of this type of visualizations for multivariate data the R language is recommended, which is an integrated collection of software services for data handling, calculation and visualizations through graphics. Software R is free under terms of the GNU General Public License of the Free Software Foundation as source code [3].

There are research areas where it is practical to use multivariate analysis and above all the representation of results through visual techniques, as it facilitates the interpretation of results for decision making.

One of the areas of research where this type of representations is useful is the educational one, an example being the universities that are nowadays in charge of covering their quality indicators operating with educational models that require the integral formation of the students to increase the quality of their educational process, so it is necessary to know the information of their students, which can be interpreted in the best way, especially by the academic authorities. Taking into account this new model where the student is responsible for his own learning, where he is the one who builds the knowledge, who learns, it is necessary to know the different aspects such as that of psychological linkage with studies [4] or also called academic engagement, that is, how dedicated, engaging and vigorous students are.

There are some works where graphic representations have been developed to know the behavior of the data in educational areas at the license level, such as radar graphs where it is possible to visualize the comparison of the data in a fast way, facilitating the interpretation of the data through colored lines that form polygons [5, 6]. Likewise, research has been found related to academic performance and the elements that influence university students where data analysis techniques have been applied [7, 8] or techniques for evaluating the results obtained [9].

The objective of this work is to develop multivariate analysis of student engagement presenting the results through visualization techniques such as a graphical correlation matrix, GGE Biplots type graphs and Chernoff Face graphs. These graphs allow explicit visualization of the associations between variables, analysis where the main components can be observed, as well as the identification of the characteristics of the cases involved in this study.

2 Materials and the Proposed Method

2.1 Data Description

The purpose of this research is to present the results of a multivariate analysis through visualization techniques on the student engagement that students of two undergraduate

degrees and technologies present in a sample population belonging to a computer and systems faculty in a university in the south of Mexico. The bachelor's degrees considered for this study were the bachelor's degree in administrative informatics and bachelor's degree in computer systems. The sample was non-probabilistic, directed and for convenience [10], an Utrecht Work Engagement Scale for Students (UWES-S), survey was applied to a group of 32 students, 16 of each undergraduate degree, who agreed to answer the questionnaire on a voluntary basis, the period of application of the survey was February-August 2015.

2.2 Scale of Welfare in the Academic Context (Utrecht Work Engagement Scale for Students, UWES-S)

The engagement at work is defined as [11]: "a positive psychological state characterized by high levels of energy and vigor, dedication and enthusiasm for work, as well as total absorption and concentration in work activity." Because the studies are considered as an activity in which responsibilities are acquired and one has an objective, the same concepts and instruments of engagement at work for the student engagement were used, with some slight variations. The Utrecht Work Engagement Scale for Students (UWES-S), questionnaire was developed to measure student engagement, which consists of 17 items that indicate [12]: "vigor is evaluated by six items, which refer to high levels of energy and resilience, willingness to devote efforts, not fatigue easily, and persistence in the face of difficulties.

The dedication is evaluated through five items that refer to the meaning or meaning of the work, to feel enthusiastic and proud of its work, and to feel inspired and challenged by the work. Absorption is evaluated by six items that refer to being happily immersed in their work and have difficulty leaving it, in such a way that time passes quickly and one forgets everything around them.

The responses of the instrument are measured according to a Likert scale, where zero means "never" and six "always", so scores range from zero to six for each dimension that makes up student engagement. The original internal consistency obtained by the authors for the UWES-S version of 17 items in Dutch students was for vigor, dedication and absorption of 0.63, 0.81 and 0.72 respectively, thus meeting the criterion of superiority to 0.60 for an instrument of recent development. The content of the questionnaire can be consulted in the Handbook of UWES [12].

2.3 Data Analysis

A descriptive analysis was carried out to determine the minimum, maximum, mean and standard deviation values, corresponding to the dimensions that make up student engagement: vigor, dedication and absorption, as well as student engagement as a variable.

Table 1. Descriptive statistical population study sample.

Variable	Minimum	Maximum	Mean	Standard Deviation
Vigor	1.8	5.7	4.141	1.0073
Dedication	1.8	6.0	4.906	1.0656
Absorption	1.2	5.8	4.125	1.1334
Engagement student	1.7	5.8	4.394	0.9990

To carry out the multivariate analysis and the representation of visual results, the R language was used, which has been considered as one of the most powerful and versatile software, especially in the construction of graphics. The R language has a variety of packages that allow programming at a higher level, complex analyzes and custom graphics design [13].

A correlational analysis was applied, presenting the results in a graphical way, in order to know how associated the variables of the student engagement are. Another way to know the characteristics of the cases presented was through the implementation of the GGE Biplot multivariate analysis, where two elements called genotypes (G) and another of genotype-environment interaction (IGA) are used, this technique facilitates the visual identification of genotypes and environments for their interpretation [13].

In the same way, an analysis was developed through the technique of Chernoff faces where the characteristics of the student engagement are represented. Chernoff proposed a representation method for multivariate data, the use of faces in the form of caricatures to represent data having several dimensions [14]. In Chernoff faces the values of the variables involved in the analysis correspond to facial features, this technique aims to improve the user's ability to detect and understand the importance of the phenomenon and serve as a mnemonic mechanism to remember the main conclusions [14].

3 Results

A descriptive analysis was carried out, which can be observed in Table 1, with a sample of 32 students of two careers in technologies, it is possible to observe the values obtained in minimum and maximum, average, and standard deviation for the student engagement and its dimensions. The minimum value obtained in the cases is observed in the absorption with 1.2, the maximum value is presented in the dedication with 6.0. The values shown in the mean of the dimensions from highest to lowest are for dedication 4,906, vigor 4,141 and absorption 4,125, finally the average of this population shows in student engagement is 4.394.

Table 2 presents a scale that was developed to qualify the student engagement, this table with established values allows to evaluate the observations of a case or group of cases 10. Other methods can also be used to calculate levels of student engagement with the UWES-S, consult Schaufeli and Bakker [12]. Using the scale can be interpreted for this population sample that the average vigor 4,141 and absorption 4,125 is average, that is, the students feel moderately vigorous and absorbed, on the other

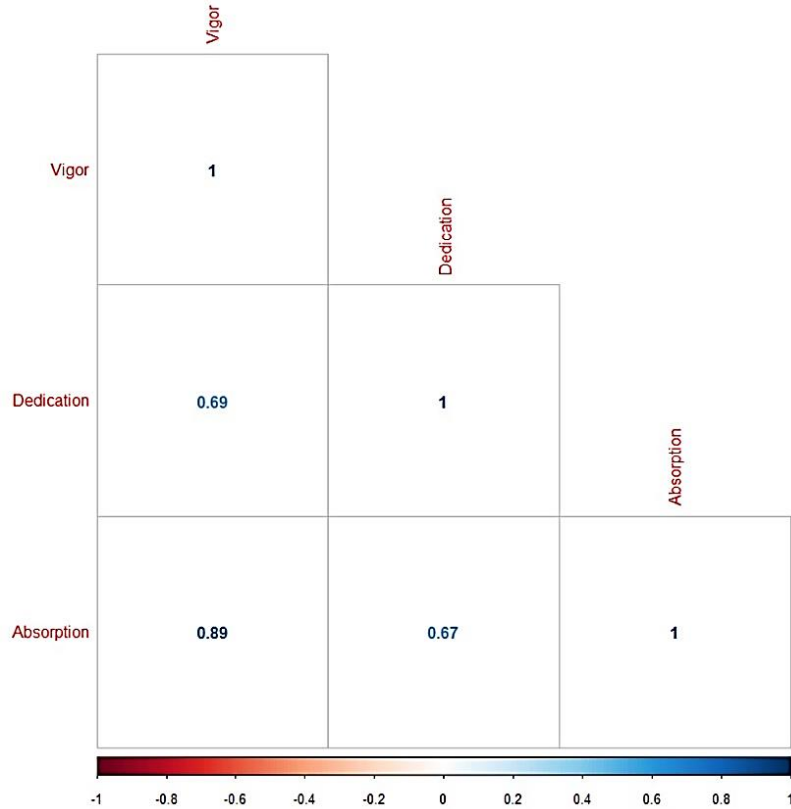


Fig. 1. Correlation matrix chart of the dimension’s vigor, dedication and absorption of the student engagement. Source: (Self-realization, 2017).

Table 2. Scale proposed to measure the UWES-S.

Category	UWES-S	Vigor	Dedication	Absorption
Very Low	Score < 2.20	Score < 2.80	Score < 2.50	Score < 2.70
Under	2.20 ≤ score < 3.30	2.80 ≤ score < 3.80	2.50 ≤ score < 3.50	2.70 ≤ score < 3.60
Middle	3.30 ≤ score < 4.70	3.80 ≤ score < 5.20	3.50 ≤ score < 4.50	3.60 ≤ score < 4.70
High	4.70 ≤ score ≤ 6.00	5.20 ≤ score ≤ 6.00	4.50 ≤ score ≤ 6.00	4.70 ≤ score < 6.00

hand, the dedication has a value of 4,906, considered high, so the students are enthusiastic, motivated and proud of their studies.

3.1 Visual Representation Technique of Correlational Analysis

A correlation analysis was carried out presenting its results through a matrix chart, which measures the degree of association that exists between the dimensions that make

up the student engagement of the sample population of this study. The method used is the Pearson correlation coefficient, which represents the linear relationship between two quantitative variables [15], the level of significance is 0.01. The magnitude of the coefficient indicates the degree of relation between the variables, so if this data is closer to one the correlation is greater, on the contrary, it is smaller if this data is closer to zero [16].

In Figure 1 shows that the resulting correlations between the variables force, dedication and absorption are close to one, so there is an important degree of association between these variables, it is concluded that the instrument that has been selected to measure student engagement is appropriate for the sample population.

3.2 Visual Representation Technique of GGE BIPLOTS

Genotype plus Genotype-by-Environment (GGE) Biplot is another visual technique that allows to represent the type of multivariate analysis [17]. The GGE biplot is constructed from the first two principal components. The first component, when highly correlated with the main genotype effect, represents the proportion of the yield that is due only to the characteristics of the genotype.

The second component represents the part of the yield due to the genotype-environment interaction. The genotypes close to each other in the GGE biplot present similar patterns of response across the environments. The environments close to each other, given by the acute angle between their vectors, indicate positive environmental association, that is, similar patterns of response in the relative behavior of a set of genotypes. The absence of association between environments is given by the right angle between vectors and the negative association by the obtuse angle [18].

In Figure 2, we can see the called Genotypes (G), which represent the students, each case, where they are identified by a label that is integrated by the degree program and a unique additional number, just as it is visualized to the interaction Genotype-by-Environment (GE), formed by the variables, in this case student engagement and its dimensions of vigor, dedication and absorption. The graph shows the first two main components, derived from the decomposition of the combined effects of G + GE [13]. The first main component explains 91.14% and the second main component is 5.64%, so the variability explained by the two axes is greater than 50% [13], so that at 96.78% the interpretations shown are considered valid.

The GGE Biplots are interpreted in terms of distances between elements and orientation of the axes, so from the graph one can know approximately the values of all variables for each case [13]. Degrees programs considered for this study were labeled in the graphics as LIA for the bachelor's degree in administrative informatics and LSC for computer systems.

In Figure 2, all genotypes (G) are observed, most of them agglomerate in a group, observing a significant relation of all these genotypes with their environment (GE), that is, with all the variables that integrate the engagement student. On the other hand, genotypes (G) that are far from the group, which are LIA2, LIA10, LIA6 and LSC9, are observed in the opposite direction to the variables of the student's engagement, reason why it is deduced that they present low levels of vigor, absorption, dedication and consequently also under student engagement.

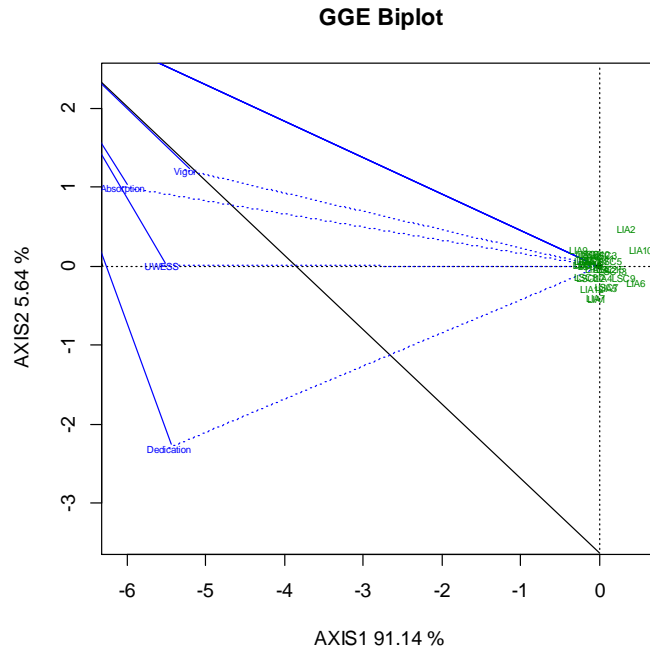


Fig. 2. Graph GGE Biplot of the dimension’s vigor, dedication, absorption and UWES-S, obtained from the students. (Source: Self-realization, 2017).

Table 3. Effect of variables in R Language.

Modified Item	Variables
Height of face	Vigor
Width of face	Dedication
Structure of face	Absorption
Width of ear	UWES-S
Height of eyes	Bachelor’s degree
Height of hair	Gender

3.3 Chernoff Faces Visual Representation Technique

In the representation across the faces of Chernoff each dimension of the data is determined by the size, location, or shape of some facial component of the caricature [14]. In this study, we developed this type of visualization technique to show the characteristics of student engagement taking into account the sample population.

Using the face technique of Chernoff, a facial feature was assigned to each of the variables, the variables corresponding to vigor, dedication, absorption, UWES-S student engagement were represented, and two variables were added that correspond to career and gender of the student, in total six variables.

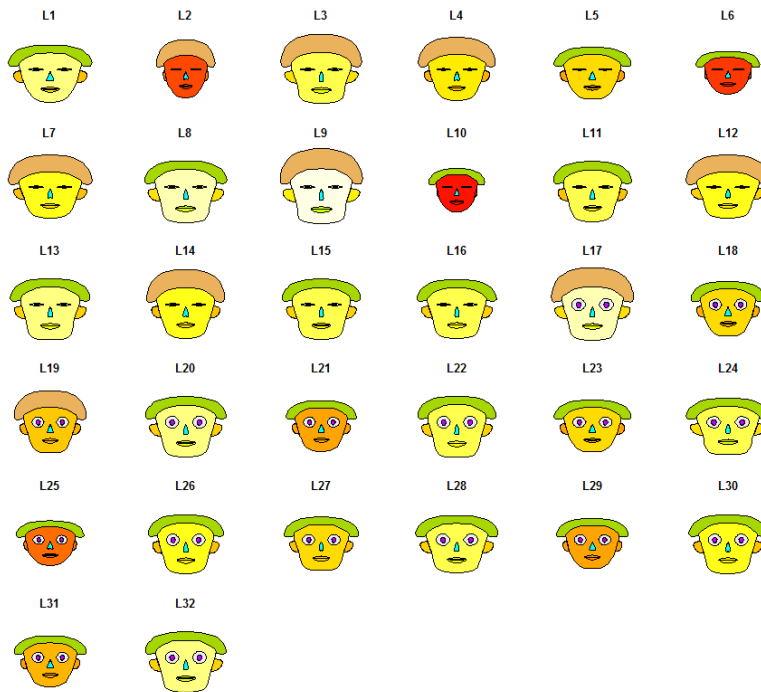


Fig. 3. Representation of Chernoff faces of the dimension's vigor, dedication, absorption and UWES-S, obtained from the students, Source: (Self-realization, 2017).

Table 3 shows the correspondence between the facial feature (Modified Item) and the variables that integrate student engagement (Variables). Examples of these representations are: the variable Vigor that is represented by the height of the face; the student engagement variable, identified as UWES-S, which is represented by the width of the ear; and the variable gender that is represented by the height of the hair.

In Figure 3, the results of the application of the Chernoff faces technique are presented, all cases, 32 figures of both bachelor's degrees: administrative informatics and computer systems. An identifier that begins with the letter L followed by a unique number was used.

One of the representative cases is the student of L3, where the vigor is represented by the height of the face, which is considered of the high ones in comparison with the others, reason why considers a student with a high level of vigor; the dedication that is identified by the width of the face, is remarkably wide, that is, has a high level of dedication; the absorption is determined by the structure of the face, which is very large, notably square, so that the student has a high level of absorption; the student engagement UWES-S, is represented by the ears, which are very wide in this case, so the student has a high level of student engagement.

For additional variables such as bachelor's degree, this is especially identified by the height of the eyes, for almost closed eyes corresponds to the bachelor's degree in

administrative informatics, for wide eyes corresponds bachelor's degree in computer systems; for the case of the variable gender, the height of the hair represents whether it is male or female, very high corresponds to female and short corresponds to male. In this case of the student of L3, it is identified as a part of the degree program in administrative informatics and is a woman.

Another of the cases in Figure 2, where it is observed that its characteristics show remarkable differences, is the L25, considering Table 3, it can be concluded that given its facial features, it has a low level of vigor, dedication, absorption and consequently a low student engagement, in the same way it is observed that it is a man student in computer systems.

4 Conclusions

The use of visual techniques to represent analysis of multivariate data presents advantages such as the rapid identification of the differences between the cases and the rapid interpretation of the same. In this paper, we have gone into the search for new forms of multivariate visualization, since a great amount of information needs to be analyzed from multiple dimensions, three forms of representation of results were selected for this type of studies, the development of correlation charts, GGE Biplot, and Chernoff faces. The R language was used for the implementation of these techniques, since it was found to be one of the most complete languages in the area of visual techniques deployment, since it allows the development of high level graphics.

The multivariate visualizations were developed to give information of the engagement in university students, conformed in three dimensions: vigor, dedication and absorption; variables such as undergraduate and gender were also considered. The correlation technique used allowed us to visualize how closely related the variables of the engagement are. Regarding the visualization through GGE Biplot, this contributed to visualize the relation between the cases according to the variables that make up the student's engagement, or it could also be said according to this technique: the positioning of genotypes in the environment.

When using the Chernoff faces method, a direct translation of the actual data into facial features is done, for each of the cases that integrated the sample. The techniques used for the visualization of multivariate analysis in this work, allow observing the information from different perspectives, contributing to the understanding of the data in a faster way, since it is possible to simplify the tedious and complex, reason why it is recommendable carry out this kind of visual transformations. It is proposed to continue the studies related to the development of visual techniques in the R language to represent data that facilitate the understanding of the results. The next step is to combine analysis of the type of main components with a discriminant analysis, both represented in the same graph.

The analysis of main components allows transforming the set of original variables into another reduced set of variables, the components, which conserve most of the information about the variability of the data. Discriminant analysis is used to separate the elements into groups or known classes.

By combining both methods, a more complete description will be obtained for the analysis of data, suitable when there are many variables correlated with each other and a classification criterion to associate the variables.

References

1. Hazarmabeth, I.: Métodos estadísticos exploratorios y confirmatorios para análisis de datos. Facultad de Estudios Superiores Zaragoza, Universidad Nacional Autónoma de México (2013)
2. Correa, J. C.: Gráficos Estadísticos con R. Universidad Nacional-Sede Medellín (2002)
3. R Development Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing, <http://www.R-project.org> (2016)
4. Casuso, M. J.: Estudio del Estrés, Engagement y Rendimiento Académico en Estudiantes Universitarios de Ciencias de la Salud. Disertación doctoral publicada, Universidad de Málaga, España (2011)
5. Morales, E.Y., Hernández, J.A., Ochoa, C.A., Lopez, M.A.: Comparison Represented in the Form of Radar of University Student Engagement in Degrees in Technologies. In: Sidorov, G., Ritter, G., Serra, J. Cortes, U.(eds.), Research in Computing Science, 122 (2016)
6. López, M.A., Hernández, J.A., Ochoa, C.A., Morales, E.Y., González, C.: Comparative Study of Learning Strategies of Bachelor Students in Nursing. In: Sidorov, G., Ritter, G., Serra, J. Cortes, U.(eds.) Research in Computing Science, 122 (2016)
7. Quintana, M., Trinidad, J.C., Morales, S.J., Landassuri, V.M.: Análisis Comparativo de Algoritmos de Minería de Datos para Predecir la Deserción Escolar. In: Santos, E.A., Chimal, J.C., Cabrera, L., Castillo, E. (eds.). Research in Computing Science, 67, Special issue Advances in Computer Science, pp. 51–61 (2013)
8. Gudino-Penalosa, F., González-Mendoza, M., Mora-Vargas, J.: Uso de técnicas de agrupamiento en la clasificación de estilos de aprendizaje. In: Miranda, S. (ed.), Research in Computing Science, 95, Minería de textos, lógica y ontologías, pp. 135–146 (2015)
9. Cruz, C., Reyes, L.A., Hernández, G., Sánchez-Cervantes, J.L.: Diseño de una ontología para el proceso de evaluación de las asignaturas técnico científicas del Instituto Tecnológico de Orizaba. In: Sidorov, G., Ritter, G., Serra, J. Cortes, U.(eds.), Research in Computing Science, 128 (2016)
10. Hernández, R., Fernández, C., Baptista, M.: Metodología de la investigación. McGraw-Hill Interamericana (2010)
11. Salanova, M., Schaufeli, W.B.: El Engagement de los empleados un reto emergente para la dirección de recursos humanos. Estudios Financieros, 261, pp. 109–138 (2004)
12. Schaufeli, W.B., Bakker, A.: Utrecht Work Engagement Scale (UWES). Escala de Engagement en el trabajo de Utrecht, Occupational Health Psychology Unit: Utrecht University (2003)
13. Guisande, C., Vaamonde, A.: Gráficos estadísticos y mapas con R. España. Ediciones Díaz de Santos (2013)
14. Raciborski, R.: The Stata Journal. Graphical representation of multivariate data using Chernoff faces. 9(3) pp. 374–378 (2009)
15. Cuñer, N.: Escala de Inteligencia para Niños de Wechsler, WISC-IV. Diccionario de Psicometría (2013)
16. Aquije, R., Ramírez, R., Castillo, H.: Guía para la aplicación del análisis multivariado a las encuestas de hogares. Lima, Perú: INEI (2002)
17. Dumble, S., Frutos, E., Galindo, P.: Package'GGEbiplots (2017)

18. Ibañez, M.A., Cavanagh, M. M. Bonamico, N.C., Di Renzo, M.A.: Análisis gráfico mediante biplot del comportamiento de híbridos de maíz. *RIA*, 35(3), pp. 83—93, INTA (2006)