2020 HAWAII UNIVERSITY INTERNATIONAL CONFERENCES
ARTS, HUMANITIES, SOCIAL SCIENCES, & EDUCATION JANUARY 6 - 8, 2020
HAWAII PRINCE HOTEL WAIKIKI, HONOLULU, HAWAII

NOVEL EDUCATIONAL METHODOLOGY OF PLAYING VIDEO GAMES

CHEN, MASON
STANFORD UNIVERSITY ONLINE HIGH SCHOOL
PALO ALTO, CALIFORNIA

CHEN, CHARLES
MORRILL LEARNING CENTER
SANTA CLARA, CALIFORNIA

Mr. Mason Chen
Stanford University Online High School
Palo Alto, California

Dr. Charles Chen
Morrill Learning Center
Santa Clara, California

Novel Educational Methodology of Playing Video Games

This paper would introduce a novel educational methodology of playing a strategic video game- Empire Four Kingdoms. In modern Big Data World, Artificial Intelligence algorithm is powerful for engineering problem solving to discover the complicated science or/and mechanics. Separating Statistics from Math can draw practical decision and conduct risk assessment. Several modern Artificial Intelligence algorithms are adopted to build a statistical model of optimizing the troop design and understanding the Military Science. Environmental Science and Natural Resources (Stone, Wood, and Food) are also addressed to educate the importance of protecting our world against Technology Advancing.

I. Project Introduction

This paper will demonstrate how to use Statistical Modeling to help student study the Environmental Science when playing the video games. There are three sessions to be included in the project introduction: (1) Introduce Empire Four Kingdoms Video Game, (2) Environmental Science, and (3) Build a Powerful Troop to Win the Battle.

1.1 Introduce Empire Four Kingdoms Video Game

Playing video game is becoming a critical portion of social activities for most middle school and high school students. However, parents are worrying that kids may play video games too much and most video games may not help develop their critical thinking and teamwork concept. The objective of this project is to convert playing video games to become conducting Projects. Students can learn Environmental Science and Statistics while playing video games. Authors have searched several video games and picked the Empire Four Kingdoms video game not based on the commercial rating. But based on the potential of applying statistical data-driven and engineering problem-solving approach due to its embedded database which can record 7 characteristics of each military unit. The main challenges of playing this video game are: (1) make your dream a reality and become a king and castle lord with the Empire, (2) Produce new resources and build your small castle into a mighty fortress, (3) recruit a powerful army to conquer more land for your kingdom and defend it against enemy attacks, and (4) form alliances with friends to defeat your enemies or fight epic battles against countless other players. Your strategy determines whether you become a legendary king or remain a simple peasant [1].

1.2 Environmental Science

Environmental resource is very critical in the early stage to choose the location of your main castle and the first few operation sites (supporting resources to main castle). There are three types of natural resources: Stone, Wood and Food. Each castle and operation site have its own combination of three natural resources. The functions of each resource in game are described below:

Stone can be obtained through establishing a Stone Quarry. Stone is used in construction, creating tools for castle siege or defense. The abundance of the resource can be seen in the form of rock formations.
Wood is used in construction, creating tools for castle siege or defense through constructing a Woodcutter building. A high wood production area has lush forests around it.
Food is for feeding your citizens, soldiers. This seems to be, among the 3 basic resources, the most important and hard to maintain resource due to just by having citizens and soldiers around, because they need to eat, this resource continuously diminishes at a rate equal to how many your population + the soldiers are.

The castle type of your first few castles may impact the Kingdom's personality and growing strategies to survive in the neighbor surrounding by weaker or mighty kingdoms.

1.3 Understand the Military Science

To expand your Kingdom territory, building a powerful troop is the highest priority. What kind of Troops needed at each Phase? More than 40 types of Troop Units are available: some are for defense and some are for attack. There are seven characteristics for each troop unit: (Melee/Range Defense, Melee/Range Attack, Travel Speed, Looting Capacity, Food Consumption) [2]. For example, Deathly Horror & Demon Horror units are some of the most powerful units you can get in the game. They can be obtained when you buy a lot of rubies or attain a rather high position in a tournament or event. These units are mostly hotly contested since having them in an attacking force greatly increases your chances of success. A summary of 7 characteristics across most popular troops is shown in Table 1.

Table 1: summary of 7 troop characteristics

	Melee Attack	Melee Defense	Ranged Attack	Ranged Defense	Looting Capacity	Travel Speed	Food Consumption
Archer	0	53	10	55	9	25	2
Bowman	0	8	24	24	13	75	2
Composite Bowman	0	74	16	159	2	25	3
Crossbowman	0	6	36	16	22	73	2
Crossbowman of the Elite Guard	0	14	129	23	29	28	4
Crossbowman of the Kingsguard	0	14	121	23	29	28	4
Deathy Horror	0	15	162	24	40	28	5
Demon Horror	185	19	0	5	35	28	5
Elite Crossbowman	0	14	129	23	29	28	4
Elite Kight	140	18	0	5	33	28	4
Flame Bearer	14	170	0	68	3	25	3

II. Multivariate Statistics

To build a powerful troop, building a diverse and powerful troop to balance the overall strength across all 7 characteristics is a tough challenging. This section would apply several Multivariate Statistics techniques including: (1) Multivariate-Correlation Analysis, (2) K-Means Clustering Analysis, and (3) Hierarchical Clustering Analysis, (4) Clustering Algorithms, and (5) Parallel Plot.

2.1 Multivariate-Correlation Analysis

Multi-Correlation analysis was done to discover the affinity patterns among 7 characteristics as shown in Figure 1.

Correlations

	Melee Attack	Melee Defense	Ranged Attack	Ranged Defense	LootingCapacity	Travel Speed	Food Consumption
Melee Attack	1.0000	-0.2591	-0.5620	-0.4615	0.3408	-0.1502	0.4160
Melee Defense	-0.2591	1.0000	-0.3562	0.5717	-0.4613	-0.1325	-0.3920
Ranged Attack	-0.5620	-0.3562	1.0000	-0.1044	0.2682	-0.0841	0.3950
Ranged Defense	-0.4615	0.5717	-0.1044	1.0000	-0.4755	-0.1554	-0.3625
LootingCapacity	0.3408	-0.4613	0.2682	-0.4755	1.0000	-0.1385	0.7727
Travel Speed	-0.1502	-0.1325	-0.0841	-0.1554	-0.1385	1.0000	-0.4937
Food Consumption	0.4160	-0.3920	0.3950	-0.3625	0.7727	-0.4937	1.0000

Figure 1: Multivariate Correlation Analysis of seven characteristics

Two strong positive correlation pairs were observed.

- Melee Defense and Ranged Defense: better units for Defense on both Melee and Ranged
- Looting Capacity and Food Consumption: stronger looting units will also consume more Foods which will indicate the limitation of sending more looting units during attack

Several negative correlation pairs were observed:

- Melee Attack and Ranged Attack: any attack unit can only perform well on either Melee or Ranged, but not BOTH
- Melee Attack and Food Consumption: stronger Melee Attack units also consume more Foods
- Melee Defense and Looting Capacity: stronger Melee Defense units will loot less
- Ranged Defense and Looting Capacity: stronger Ranged Defense units will loot less
- Travel Speed and Looting Capacity: faster troop will loot less

2.2 K-Means Clustering Analysis

The objective of this section is grouping troop units which serve similar functions with similar capabilities. Previous multi-correlation analysis has indicated some clustering possibility. Under battle constraint, depending on castle size, the game only allows limited troop unit types, and quantities in each battle which has forced the Kingdom to simplify their troop types but to balance their 7 characteristics in diverse. We need both “Defense” and “Attack” Troops in each battle. Clustering Analysis can identify the vital few clusters which can represent the entire population. The challenges are how to cluster the populations and how many clusters needed? If too few clusters, the troop won’t be diverse to possess all 7 characteristics sufficiently. If too many clusters, some cluster size may be too small and not truly representing any characteristic. Number of clusters is always very critical any cluster analysis. K-Means Cluster analysis is chosen to study the effect of cluster numbers. Though, K-Means method may not be powerful enough. The initial seeding of each cluster center may impact the result significantly (not a robust method).

In Figure 2, K-Means analysis of 5-8 clusters are compared side by side. Apparently, the clustering patterns are highly depending on the number of clusters specified. It’s hard to tell, based on Clustering patterns, which cluster number is optimal to present the clustering patterns more effectively. Though, it seems a clear boundary between 5-6 clusters and 7-8 clusters.

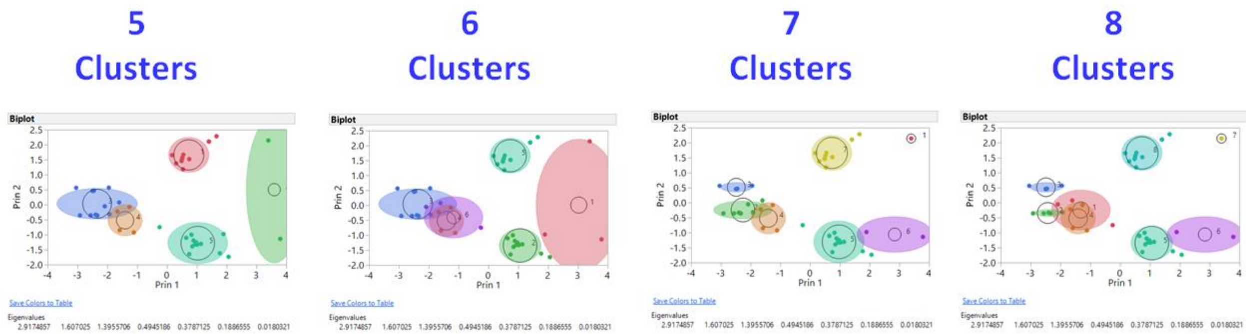

Figure 2: K-Means Clustering Analysis of 5-8 Clusters

2.3 Hierarchical Clustering Analysis

Artificial Intelligence clustering analysis can help discover the affinity grouping patterns among 40+ troops units to help build the best troops based on kingdom strategy. Hierarchical Clustering Analysis (HCA) [4] was used to further analyze the troop patterns. In data mining and statistics, hierarchical clustering (also called hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types [5]:

- Agglomerative: This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
- Divisive: This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

In the general case, the computing time of the Agglomerative approach is faster than the Divisive approach. Optimal efficient agglomerative methods have been developed to significantly improve the computing algorithm for large data sets [6,7]. The main objective of this analysis was to search for the degree of similarity among troops based on 7 characteristics. The Agglomerative approach can identify a clustering pattern faster and more accurately. The Divisive approach may not split the 40+ troop units into clusters which are more concentrated on the bottom level efficiently. Therefore, the authors chose the Agglomerative approach. This approach builds the hierarchy from the individual elements by progressively merging clusters based on a defined distance metric (Euclidean distance). The distance is calculated by the discrepancy of scores among the seven characteristics. This HCA approach can pair the troop unit with similar score patterns and use clustering to group units. JMP statistical software was used to calculate the closest distance (the affinity) among all potential pairs, and grouped the first pair, at the strongest affinity (based on their similar score pattern). The linkage criterion determines the distance between sets of observations as a function of the pairwise distances between observations [8, 9, 10].

Authors used Data Mining Cluster and Dendrogram to group the similar troops units into eight clusters as shown in Figure 3. Then, Kingdom can select the best troop unit from each Cluster to build a more diverse troop which can have sufficient power across all 7 attributes. In this paper, the top 3 attributes are: Higher Melee/Range Defense Power, and Lower Food Consumption. Based Scree Plot, 8 clusters were chosen to present the affinity behaviors among 40+ troop types associated with 7 characteristics.

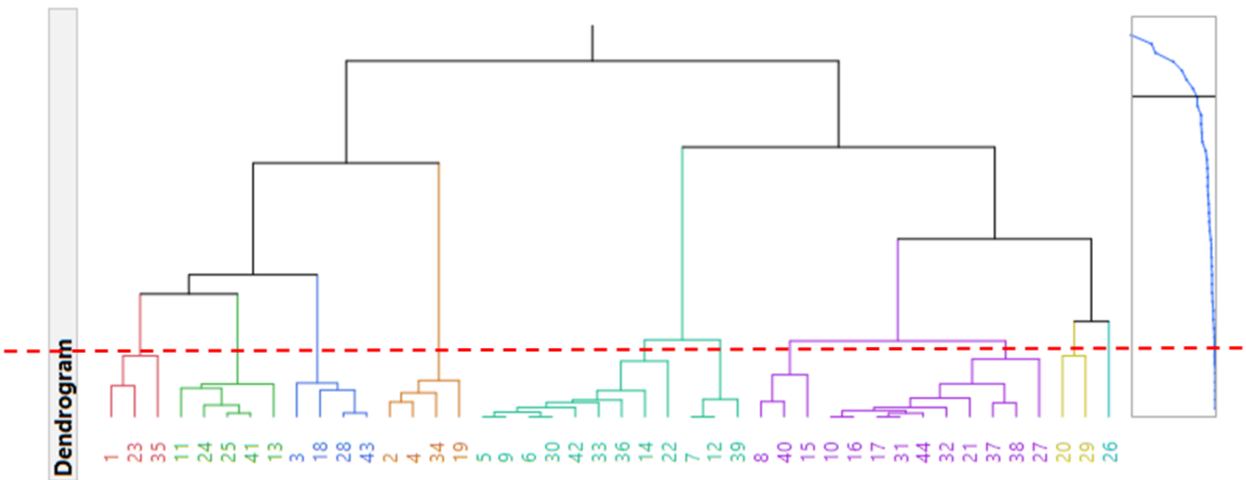

Figure 3. Cluster Analysis of Troop Types

Each eight clusters identified has its own identity across 7 troop characteristics as shown in Figure 4. For example, 1st cluster is about average across all 7 characteristics. 2nd and 3rd cluster are stronger on the Melee/Ranged Defense. The 4th cluster has faster speed and less food consumption. The remaining four clusters are good at attack and looting. Based on the kingdom strategy set previously, the 2nd-4th clusters may fit to the kingdom strategy better than the other clusters.

Cluster	Count	Melee Attack	Melee Defense	Ranged Attack	Ranged Defense	Looting Capacity	Travel Speed	Food Consumption	Characteristic
1	3	20.3	43.0	8.7	37.0	12.0	27.0	2.7	About Average
2	5	14.6	153.0	0.0	61.4	13.6	28.0	3.0	Melee Defense + Ranged Defense
3	4	0.0	66.3	16.8	139.3	13.0	27.5	3.0	Ranged Defense + Melee Defense
4	4	16.0	19.5	15.0	13.5	20.3	74.0	2.0	Travel Speed + Less Food
5	12	0.0	14.9	134.8	23.5	32.7	29.3	4.3	Ranged Attack
6	13	145.7	20.8	0.0	8.8	32.2	28.5	4.2	Melee Attack
7	2	139.5	14.5	0.0	3.5	80.0	29.0	5.0	Melee Attack + Looting
8	1	0.0	8.0	144.0	14.0	90.0	28.0	6.0	Ranged Attack +Looting

Figure 4: Cluster Summary of Troop Characteristic

2.4 Clustering Algorithms

Previous 2.4 clustering method has identified three car stage groups separated. The clustering patterns were identified based on clustering distance algorithm of calculating the dissimilarity of 7 troop characteristics among 40+ troop types. This section will study the mathematics of various clustering distance algorithms. There are several cluster algorithms: (1) Average, (2) Centroid, (3) Ward, (4) Single, and (5) Complete (Citation). Will these 5 different clustering algorithms have the same results? If different, how to select which algorithm to explore the clustering patterns best? In Figure 9, three existing clusters (Green, Yellow, Red) are going to join next. Which two clusters should bond first? The joining sequence is determined by the clustering distance algorithms. Centroid, Single, and Complete algorithms are compared show in Figure 5. The Centroid algorithm connects Green cluster and Yellow cluster through the purple line connecting the two cluster means (purple triangles). The Single algorithm groups Green cluster and Red cluster by the closest points between these two clusters. The Complete algorithm groups Yellow cluster and Green cluster by the farthest points between these two clusters. Depending on which distance algorithm chosen, the clustering sequence and pattern may be different. We must dive into the mathematical calculations for each clustering distance algorithm and understand the benefits and limitations of each algorithm to choose the best algorithm to draw reliable clustering patterns and results.

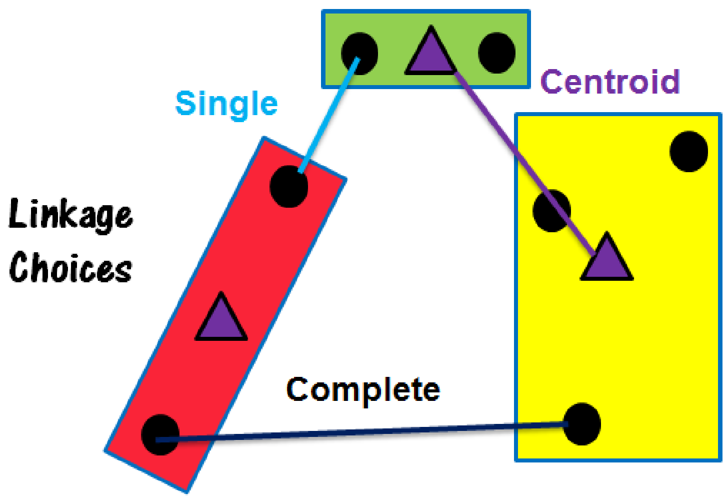

Figure 5: Diagram of the Centroid, Single, and Complete Clustering Methods

We will compare five major clustering distance algorithms [11-12]. The calculations of the five different clustering algorithms are shown in Figure 6. The first algorithm is Average which is the distance pair divided by the number of distances. Since the Average algorithm compares the average distances, it typically joins smaller and similar variances. The 2nd algorithm is centroid which calculates the distance between the cluster means [13-14]. Among five algorithms mentioned, Centroid is the most robust algorithm to outliers. The 3rd algorithm Ward uses the ANOVA sum/mean of squares (between divided by within). The Ward algorithm is Centroid divided by the degree of freedom [15-16]. Ward joins smaller numbers of observations and which is the most sensitive to outliers. The 4th Single uses the minimum distance, and therefore typically, joining larger variances/larger number [17]. Clusters (favor in Single algorithm) are large, elongated or irregular. Those clusters may have shorter distances with other similar clusters than with small-sized clusters. The last algorithm Complete joins clusters based on the farthest distance. It is more sensitive to moderate outliers and, very different from single algorithm. Complete algorithm normally joins smaller variances/smaller numbers of clusters. How will these algorithms impact the clustering patterns?

Average Linkage Distance for the average linkage cluster method is:

$$D_{KL} = \sum_{i \in C_K} \sum_{j \in C_L} \frac{d(x_i, x_j)}{N_K N_L}$$

Average

Centroid Method Distance for the centroid method of clustering is:

$$D_{KL} = \left\| \overline{x_K} - \overline{x_L} \right\|^2$$

Ward's Distance for Ward's method is:

$$D_{KL} = \frac{\left\| \overline{x_K} - \overline{x_L} \right\|^2}{\frac{1}{N_K} + \frac{1}{N_L}}$$

ANOVA

Single Linkage Distance for the single linkage cluster method is:

$$D_{KL} = \min_{i \in C_K} \min_{j \in C_L} d(x_i, x_j)$$

Minimum

Complete Linkage Distance for the Complete linkage cluster method is:

$$D_{KL} = \max_{i \in C_K} \max_{j \in C_L} d(x_i, x_j)$$

Figure 6: JMP Clustering Distance Algorithms [18]

2.5 Parallel Plot

Based on Hierarchical Clustering Analysis, 8 clusters were identified. Parallel Plot can further display the major characteristics of each cluster as shown in Figure 7. The first cluster is closer to the average of 7 troop characteristic. The second cluster is particularly much stronger on Melee Defense while the third cluster is stronger on the Ranged Defense. The fourth cluster is good on Travel Speed. The bottom four clusters are better on Attack and Looting.

Figure 7: Parallel Plot of 7 Troop Characteristics.

III. Build Predictive Model and Optimize Troops

This section will use Statistical Transfer Function to optimize the troops power on Melee/Ranged Defense Power, and Control Food Consumption.

3.1 Build Transfer Functions

Raw characteristics of top 15 troops units (2 each from first seven clusters and 1 from the 8th cluster) are summarized in Table 2. Transfer functions of total castle troop power can be calculated by the coefficients listed in the Table if quantity of each unit type provided.

Table 2: summary of characteristics across top 15 troop units.

Unit	Qty	Melee Attack Index	Melee Defense Index	Ranged Attack Index	Ranged Defense Index	Looting Index	Speed Index	Food Index
Archer	1	0	53	10	55	9	9	2
Renegade Bowmaster	1	0	71	16	53	15	15	3
Renegade Katana Warrior	1	14	165	0	66	16	16	3
Renegade Lancer	1	13	150	0	63	16	16	3
Renegade Spear Thrower	1	0	68	14	139	15	15	3
Veteran Longbowman	1	0	67	17	134	14	14	3
Maceman	1	38	38	0	6	32	32	2
Spearman	1	26	26	0	8	14	14	2
Renegade Arrow Thrower	1	0	18	148	27	25	25	4
Veteran Deathly Horror	1	0	17	175	26	50	50	5
Khan Guards	1	175	54	0	50	32	32	5
Renegade Sai Warrior	1	160	20	0	7	29	29	4
Marauder	1	113	18	0	4	70	70	4
Renegade Swashbuckler	1	166	11	0	3	90	90	6
Renegade Sai Ripper	1	0	8	144	14	90	90	6

3.2 Set Kingdom Strategy and Optimal Desirability Functions

Depending on the Kingdom strength and strategy, troops should be customized to best support the Castle situations. This paper will focus on attacking nearby weaker Neighbors by building stronger attacking troops with relatively weaker defense but higher economical looting return. Food consumption is expecting higher to support stronger attacking units and looting units. Based on Kingdom Strategy, the JMP Desirability Functions are set to emphasize more on Defense and Food Consumption by setting LSL/USL and Importance as shown in Table 3.

Table 3: JMP Design Optimization Specifications and Desirability (Importance) Setting

	LSL	USL	Importance
Attack	**1000**		**2**
Defense	500		1
Loot	**1000**		**2**
Speed	500		1
Food		**150**	**2**

Through JMP Software DOE optimization as shown in Table 4, the Kingdom could achieve the power requirements across all 7 Characteristics.

Table 4: Optimal Castle Troop Power Performance

Attack	Defense	Looting	Speed	Food
2855	1675	1900	1750	205

IV Results and Conclusions

The innovative educational methodology is demonstrated through an Empire Video Game. Environmental science and natural resources (Stone, Wood, Food) are discussed. To build a mighty troop while protecting the environment, several Artificial Intelligence algorithms are adopted to discover the troop characteristic affinity pattern among 40+ melee and ranged troops. Mathematics of clustering algorithm and Statistics are applied to find the optimal Troop design to meet the project scope and objectives. This modern method can be commonly extended to various educational systems from K-G12, Higher Education to Industrial Education across most Scientific and Engineering fields.

References

[1] https://www.goodgamestudios.com/games/empire-four-kingdoms-app/

[2] https://www.appgamer.com/empire-four-kingdoms/strategy-guide/special-units#pPhoto[guideimages]/2/

[3] Merrilea J. Mayo, Video Games: A Route to Large-Scale STEM Education? *Science* 02 Jan 2009: Vol. 323, Issue 5910, pp. 79-82, DOI: 10.1126/science.1166900

[4] Michael R. Anderberg. Cluster analysis for applications. Academic Press, New York, 1973. ISBN 0120576503.

[5] Rokach, Lior, and Oded Maimon. "Clustering methods." Data mining and knowledge discovery handbook. Springer US, 2005. 321-352

[6] R. Sibson. “Slink: an optimally efficient algorithm for the single-link cluster method”. The Computer Journal, British Computer Society. 16 (1): 30-34, 1973.

[7] D. Defays. “An efficient algorithm for a complete-link method”. The Computer Journal, British Computer Society. 20 (4): 364-366.

[8] Szekely, G.J. and Rizzo, M.L. “Hierarchical Clustering vis Joint Between-Within Distances: Extending Ward’s Minimum Variance Method”. Journal of Classification 22, 151-183, 2005.

[9] Ward, Joe H. “Hierarchical Grouping to Optimize an Objective Function”. Journal of the American Statistical Association. 58(301): 236-244.

[10] Ma, et al. "Segmentation of multivariate mixed data via lossy data coding and compression." IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(9) (2007): 1546-1562.

[11] Milligan, G.W. (1980), “An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms,” Psychometrika, 45, 325–342.

[12] Hartigan, J.A. (1981), “Consistence of Single Linkage for High–Density Clusters,” Journal of the American Statistical Association, 76, 388–394.

[13] Florek, K., Lukaszewicz, J., Perkal, J., and Zubrzycki, S. (1951a), “Sur La Liaison et la Division des Points d’un Ensemble Fini,” Colloquium Mathematica, 2, 282–285.

[14] Jardine, N. and Sibson, R. (1971), Mathematical Taxonomy, New York: John Wiley and Sons.

[15] McQuitty, L.L. (1957), “Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies,” Educational and Psychological Measurement, 17, 207–229.

[16] Sokal, R.R. and Michener, C.D. (1958), “A Statistical Method for Evaluating Systematic Relationships,” University of Kansas Science Bulletin, 38, 1409–1438.

[17] Sneath, P.H.A. (1957) “The Application of Computers to Taxonomy,” Journal of General Microbiology,17, 201–226.

[18] “Correlations and Multivariate Techniques.” Multiple Linear Regression