



Selección de Instituciones Comparables- UPRM

Oficina de Planificación,
Investigación y Mejoramiento
Institucional

Enero 2019



1 RESUMEN EJECUTIVO

El Recinto Universitario de Mayagüez de la Universidad de Puerto Rico se esfuerza continuamente por mantener una educación superior de alta calidad. Como parte de los objetivos establecidos en el Plan Estratégico 2012-2022, se desarrollaron e integraron métricas institucionales que nos permiten medirnos y compararnos con instituciones que ejercen una función similar a la nuestra. Es por esto que la identificación de instituciones comparables (“peer institutions”), como un punto de referencia, resulta en un componente fundamental de este esfuerzo de avalúo institucional.

En 2013, el Recinto actualizó la lista de instituciones comparables utilizando exclusivamente la herramienta de comparación de la fundación Carnegie, que permite filtrar instituciones con características idénticas o parecidas a las nuestras. La lista actual de instituciones comparables es la siguiente:

1. Ball State University (Indiana)
2. Bowling Green State University-Main Campus (Ohio)
3. Central Michigan University
4. East Tennessee State University
5. Georgia Southwestern University
6. Northern Arizona University
7. Sam Houston State University (Texas)
8. University of Louisiana at Lafayette
9. University of Mississippi – Main Campus
10. Wright State University – Ohio
11. North Dakota State University – Main Campus

En esta ocasión, la oficina de Planificación, Investigación y Mejoramiento Institucional (OPIMI) tuvo como objetivo actualizar esta lista utilizando una metodología estadística. Por lo tanto, en este ejercicio se utilizó análisis de conglomerados (“análisis de cluster”) para la identificación de instituciones comparables a nuestro Recinto. El análisis de cluster se define como un análisis exploratorio que busca agrupar un conjunto de datos



tratando de lograr la máxima similitud entre los elementos de cada grupo y la mayor diferencia entre grupos.

Este estudio requirió la selección de un grupo inicial de instituciones que compartan características similares al Recinto utilizando las herramientas de filtro de la fundación Carnegie y IPEDS (ej., clasificación Carnegie, tamaño, sector, entre otros), se identificaron las variables a ser utilizadas en el análisis y finalmente se determinó el nuevo grupo de instituciones comparables al Recinto utilizando análisis de cluster.

Luego de este análisis, OPIMI recomienda las siguientes 8 instituciones públicas como la nueva lista de instituciones comparables:

1. California State University-Chico
2. California State University-Dominguez Hills
3. California State University-East Bay
4. CUNY Bernard M Baruch College
5. CUNY Brooklyn College
6. CUNY City College
7. CUNY Queens College
8. Western Washington University



2 METODOLOGÍA

Para llevar a cabo el estudio, se desarrollaron los siguientes pasos:

1. Selección de grupo inicial de instituciones similares al Recinto basado en características del servicio del Sistema Integrado de datos de Educación Post-Secundaria (IPEDS) y clasificación Carnegie.
2. Selección de las variables a utilizar para llevar a cabo el análisis.
3. Estandarización de las variables cuantitativas a utilizar en el análisis de cluster.
4. Análisis de cluster.
5. Identificación de las instituciones que resultaron dentro del mismo cluster que el Recinto.

2.1 Selección Inicial de Instituciones

El grupo de instituciones a utilizarse en el análisis de cluster fue seleccionado utilizando la herramienta de agrupamiento del IPEDS Data Center y la herramienta de comparación de la fundación Carnegie. Todas las instituciones de educación superior que reciben apoyo económico del gobierno federal bajo Título IV, están obligados a proveer a IPEDS vasta información acerca de sus unidades. Esta información puede encontrarse fácilmente a través del sistema de la página electrónica. La Tabla 1 a continuación, resume los criterios utilizados en la herramienta del IPEDS Data Center para filtrar las instituciones. En algunas de las dimensiones utilizadas se determinó incluir instituciones con clasificación distinta a nuestro Recinto para así ampliar el tamaño de la muestra.

Tabla 1: Criterios de comparación para la selección del grupo inicial de instituciones utilizando IPEDS Data Center. Otras opciones fueron añadidas para ampliar el tamaño de la muestra.

Criterio	Clasificación IPEDS para UPRM	Categorías incluidas en la búsqueda
Select	Title IV participating	First Look Universe, Title IV participating
State Or Other Jurisdiction	N/A	All 50 states
Sector	Public, 4-year or above	Public, 4-year or above
Degree-Granting Status	Degree-Granting	Degree-Granting



Criterio	Clasificación IPEDS para UPRM		Categorías incluidas en la búsqueda
Highest Degree Offered	Doctor's research/scholarship	Degree-	Doctor's Degree (Other), Doctor's Degree- (Other), research/scholarship, Degree- research/scholarship, Master's Degree, Bachelor's degree
Institutional Category	Degree-Granting, Primarily Baccalaureate or Above		Degree-Granting, Primarily Baccalaureate or Above
Carnegie Classification	Master's Colleges and Universities (Medium Programs)		Master's Colleges and Universities (Larger Programs), Master's Colleges and Universities (Medium Programs)
Degree of Urbanization	City (Small)		City (Medium), City (Small), Suburban (Large), Suburban (Medium), Suburban (Small)
Institutional Size	10,000- 19,999		10,000- 19,999
Has Full-Time First-Time Undergraduates	Yes		Yes
All Programs Offered Completely Via Distance Education	No		No

De este filtro inicial, utilizando el IPEDS Data Center, resultaron seleccionadas un total de 19 instituciones.

Además, se utilizó la herramienta de comparación de instituciones de la fundación Carnegie para verificar si pudiera haber alguna otra institución que se pudiera añadir al análisis. La fundación Carnegie utiliza la información provista por IPEDS, NSF y el "College Board" para clasificar a las instituciones de educación superior en distintas categorías. En las dimensiones seleccionadas, se determinó incluir exclusivamente instituciones de idéntica clasificación a la nuestra. Las dimensiones seleccionadas están descritas en la Tabla 2.



Tabla 2: Criterios de comparación para la selección del grupo inicial de instituciones utilizando la herramienta de la fundación Carnegie.

Renjón de Clasificación Carnegie	Clasificación Carnegie para UPRM	Categorías incluidas en la búsqueda
Programas Subgraduados	Grados Profesionales + Artes & Ciencias, alguna coexistencia graduados	Grados Profesionales + Artes & Ciencias, alguna coexistencia graduados
Perfil Matricula	Muy alta proporción de estudiantes subgraduados	Muy alta proporción de estudiantes subgraduados
Level	4-years or above	
Control	Public	
Basic	Master college & Universities: Larger Programs	
Size and Setting	Four-year, Large, primarily nonresidential	

De este filtro, resultaron seleccionadas 6 instituciones públicas que no estaban contenidas en la lista de universidades de IPEDS, por lo tanto, se añadieron a esta lista, quedando así un total de 25 instituciones en la lista final.

2.2 Variables Utilizadas en el Análisis de Cluster

Las variables utilizadas en el análisis de cluster fueron seleccionadas por la oficina de OPIMI. La información de cada institución fue descargada a través del sistema de la página electrónica de IPEDS. Las variables seleccionadas son las siguientes:

1. Total estudiantes matriculados en agosto 2017
2. Matrícula de estudiantes subgraduados en agosto 2017
3. Matrícula de estudiantes graduados en agosto 2017
4. Cantidad de estudiantes "Full-time-equivalent" (FTE)
5. Estudiantes subgraduados con beca Pell
6. Aportaciones del estado por estudiante (State appropriations/student)



2.3 Análisis de Cluster

El análisis de cluster fue realizado utilizando el programa estadístico R Studio. Antes de llevar a cabo el análisis, fue necesario realizar un pre-procesamiento de los datos. Una vez completado el pre-procesamiento, se ejecutó el análisis de cluster y finalmente se verificó el ajuste de los grupos usando un análisis de componentes principales (PCA).

Pre-procesamiento de los datos

Como parte del pre-procesamiento, todas las variables fueron estandarizadas debido a la diferencia en escalas. Este paso es particularmente importante ya que la variación en escalas puede afectar los resultados del análisis. Los elementos de cada variable fueron estandarizados restándole su promedio y dividiendo por la desviación estándar, según ilustrado en la Ecuación 1.

$$x' = \frac{x - \bar{x}}{\sigma} \quad (1)$$

Análisis de Cluster

El método de cluster utilizado en este estudio fue cluster jerárquico. Este método es considerado el más apropiado ya que el tamaño del conjunto de datos utilizado es pequeño. La función de cluster jerárquico de R Studio utiliza distancias euclidianas para medir la similitud de las observaciones. La Figura 1 muestra el dendograma resultante del análisis de cluster realizado. Un dendograma es una representación gráfica que muestra los grupos que se forman al crear los conglomerados de observaciones según su nivel de similitud. Del análisis de cluster se puede determinar que las 25 instituciones fueron divididas en 3 grupos (ver Figura 1). Nuestro recinto está identificado en el dendograma con el número 25, que está encasillado en el primer grupo (cluster).

Peer Institutions Cluster Analysis

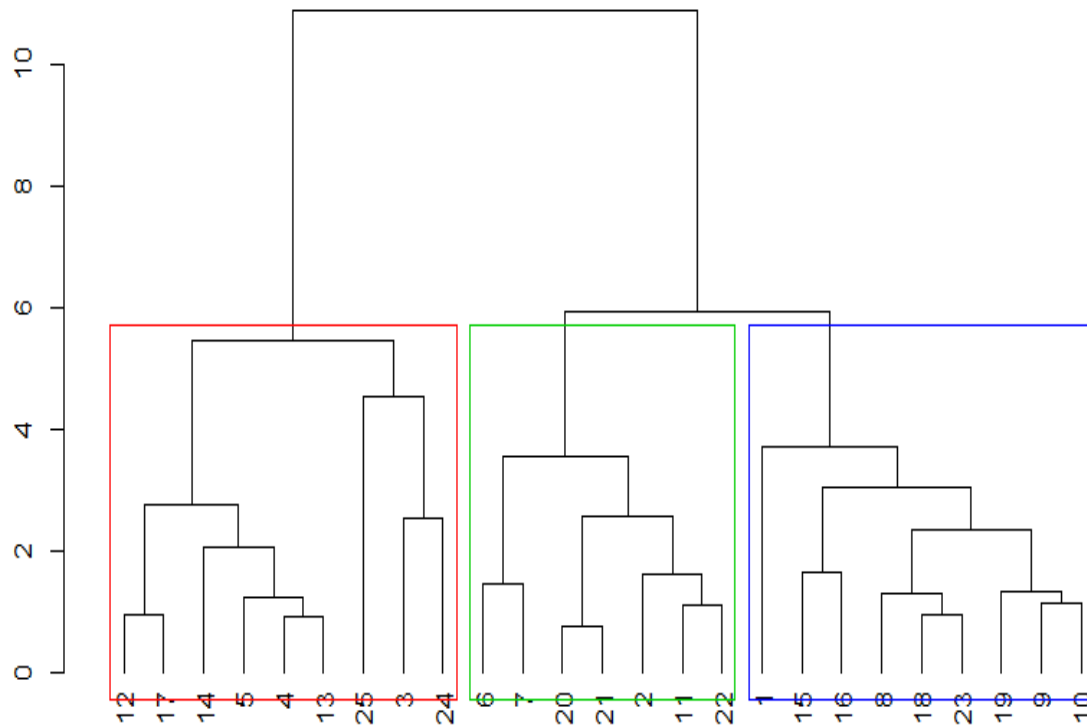


Figura 1: Dendrograma resultante del análisis de cluster jerárquico. El recinto está identificado con el número 25, encasillado en el primer grupo (color rojo).

Análisis de Componentes Principales (PCA)

Luego de llevar a cabo el análisis de cluster, se corrió un análisis de componentes principales (PCA) para verificar en donde se ubicaban los 3 grupos. PCA se utiliza para encontrar una combinación lineal de variables que puedan separar o clasificar dos o más grupos. En este análisis los primeros dos componentes principales explican aproximadamente 76% de la variación en el estudio, así que graficarlos nos puede dar una buena idea de donde están ubicados los grupos (cluster) y cuan cercanos están entre ellos. El gráfico de dispersión en la Figura 2 muestra que hay buena separación entre los 3 grupos resultantes del análisis de cluster; donde nuestro recinto está en el grupo 3 (color verde).

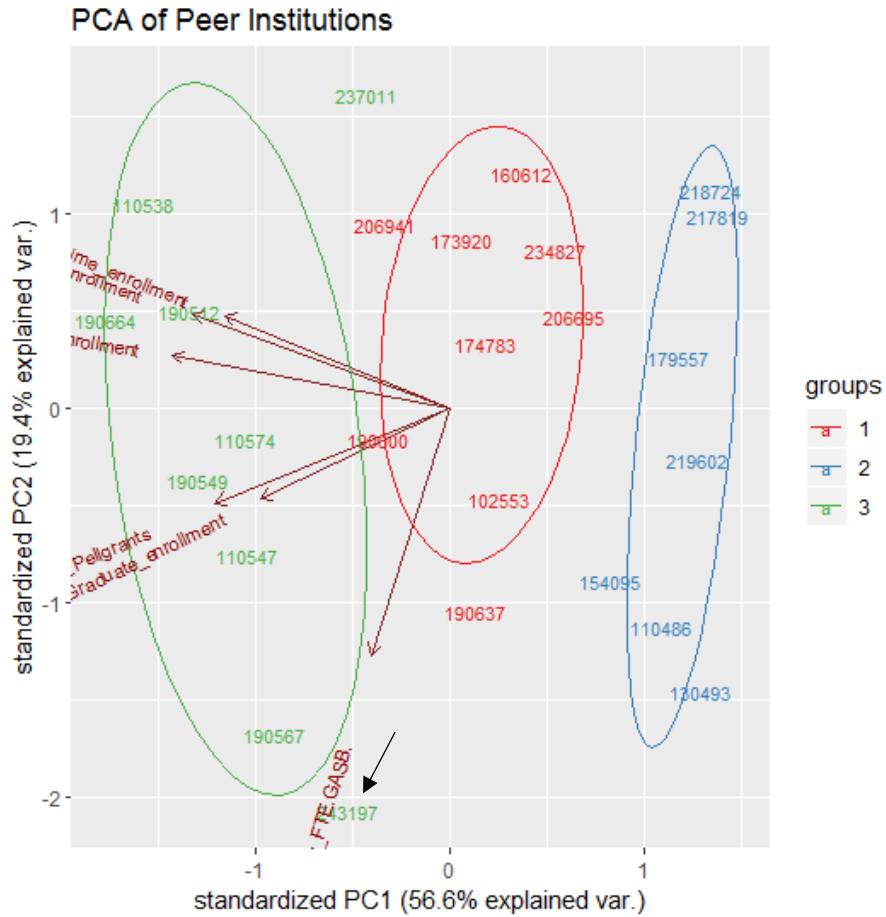


Figura 2: Gráfico de dispersión de las instituciones por grupos (clusters). Las instituciones están identificadas por su número de identificación de IPEDS. El recinto está ubicado en el grupo 3 (color verde), marcado con una flecha.



3 RESULTADOS

Utilizando cluster jerárquico se determinó que las 25 instituciones evaluadas (incluyendo el recinto) se dividen mejor en 3 grupos. Nuestro recinto resultó ser más estadísticamente similar al grupo 3 donde se encuentran las siguientes 8 universidades:

1. California State University-Chico
2. California State University-Dominguez Hills
3. California State University-East Bay
4. CUNY Bernard M Baruch College
5. CUNY Brooklyn College
6. CUNY City College
7. CUNY Queens College
8. Western Washington University

La Tabla 3 resume los valores del recinto para las variables evaluadas en el análisis de cluster en comparación con los valores promedios del grupo 3. Además, para cada una de las 8 instituciones del grupo 3 se verificó si ofrecen programas de ingeniería y/o agricultura ya que esto es una característica peculiar de nuestro recinto. La Tabla 4 resume los programas de ingeniería y agricultura que ofrecen estas instituciones.

Tabla 3: Valores del recinto para cada variable evaluada en el análisis de cluster comparados con los valores promedios del grupo 1.

Variables utilizadas	UPRM	Promedio Grupo 3
Total de estudiantes	13,481	16,887
Estudiantes a tiempo completo	12,378	12,744
Matrícula subgraduada	12,486	14,527
Matrícula graduada	995	2359
Aportaciones del estado/estudiante	\$15,522	\$8,037
Número de estudiantes con beca Pell	7,826	6,780



Tabla 4: Resumen de los programas de ingeniería o agricultura ofrecidos en las instituciones en el grupo 1.

IPEDS ID	Nombre de la institución	Colegios de Ingeniería o Agricultura	Programas ofrecidos en Ingeniería o Agricultura
110538	California State University-Chico	College of Agricultural	Agriculture, Agricultural Business, Animal Science
		College of Engineering, Computer Science, and Construction	Computer Animation & Game Development (CAGD), Civil Engineering (CIVL), Computer Engineering (CMPE), Computer Information Systems (CINS), Computer Science (CSCI), Concrete Industry Management (CIMT), Construction Management (CMGT), Electrical Engineering (EENG), Mechanical Engineering (MENG), Mechatronic Engineering (MECA), Sustainable Manufacturing (SFMT), Computer Science (CSCI), Electrical & Computer Engineering (EECE)
110547	California State University-Dominguez Hills		
110574	California State University-East Bay	School of Engineering	Computer Engineering, Industrial Engineering, Engineering Management, Construction Management, Construction Management
190512	CUNY Bernard M Baruch College		
190549	CUNY Brooklyn College		
190567	CUNY City College	Grove School of Engineering	Biomedical engineering, Chemical engineering, Civil engineering, Computer science, Electrical engineering, Mechanical engineering
190664	CUNY Queens College		
237011	Western Washington University	College of Science and Engineering	Electrical Engineering BS, Manufacturing Engineering BS, Plastics and Composites Engineering BS



4 RECOMENDACIONES

Basado en los resultados del análisis de cluster realizado junto con los criterios de evaluación descritos anteriormente, OPIMI recomienda las 8 instituciones contenidas en el primer grupo del análisis de cluster como instituciones comparables a nuestro recinto. Estas instituciones son:

1. California State University-Chico
2. California State University-Dominguez Hills
3. California State University-East Bay
4. CUNY Bernard M Baruch College
5. CUNY Brooklyn College
6. CUNY City College
7. CUNY Queens College
8. Western Washington University



5 REFERENCIAS

Teeter, D. J and Brinkman, P.T. (1987). Peer institutional studies/institutional comparisons. Primer for Institutional Research, J. Muffo and G. McLaughlin (eds), Association for Institutional Research: Tallahassee.

Terenzini, P. T., Hartmark, L., Lorang, W. G., & Shirley, R. C. (1980). A conceptual and methodological approach to the identification of peer institutions. *Research in Higher Education*, 12, 347-364



6 APÉNDICE-CÓDIGO DE R

```

install.packages("BBmisc")
install.packages("dplyr")
install.packages("dendextend")
library(dendextend)
library(BBmisc)
library(dplyr)

data = read.csv("Documents/PeerInstitutions_RawData.csv")

#Data pre-process----
rows = which(is.na(data$StateAppro_Est))
data = data[-rows, -c(6:7)] #delete rows (institutions) with missing data and
unused columns

NormVars = normalize(data[,3:ncol(data)]) # data frame of standarized variabl
es  $x(s) = X_i - \bar{X} / \text{Std dev}$ 

# Hierarchical Clustering----
set.seed(1)
d=dist(NormVars, method = "euclidean") #distance between institutions. Distan
ce matrix for clustering
hclustfit=hclust(d, method = "ward.D2")
plot(hclustfit)

dend = as.dendrogram(hclustfit)
plot(dend, main= "Peer Institutions Cluster Analysis")
rect.hclust(hclustfit, k=3, border = 2:6)

#Cut the tree to create only five clusters (from hclust plot)
groups=cutree(hclustfit, k=3)
table(data$InstName,groups)

DataCluster = mutate(data, Cluster = groups) #Data frame with cluster number
count(DataCluster, Cluster)

PeerInsts = DataCluster[which(DataCluster$Cluster==3),] # UPRM is in cluster
3
Clust1 = DataCluster[which(DataCluster$Cluster==1),]
Clust2 = DataCluster[which(DataCluster$Cluster==2),]

AvgClust1 = apply(Clust1 [,3:8],2,mean)
AvgClust2 = apply(Clust2[,3:8],2,mean)
AvgClust3 = apply(PeerInsts [,3:8],2,mean)

rbind(AvgClust1,AvgClust2,AvgClust3)

```



```
# PCA Analysis -----
install.packages("devtools")
library(devtools)
install_github("vqv/ggbiplot")
library(ggbiplot)

row.names(data)=data$InstName
data.pca = prcomp(data[,c(3:7)], center = TRUE, scale. = TRUE) #PCA
summary(data.pca)
str(data.pca)

ggbiplot(data.pca, var.axes = FALSE, ellipse = TRUE, labels = data$UnitID,
          groups = as.factor(DataCluster$Cluster)) +
  scale_color_brewer(palette = "Set1") +
  ggtitle("PCA of Peer Institutions")

pca.plot = ggbiplot(data.pca, ellipse = TRUE, labels = data$UnitID,
                    groups = as.factor(DataCluster$Cluster)) +
  scale_color_brewer(palette = "Set1") +
  ggtitle("PCA of Peer Institutions")
```