



Forbes 2000

Análisis del conjunto de datos



¿Qué veremos?

- Introducción: Forbes
- Análisis de los datos
 - Pre-procesamiento de datos
 - Análisis de componentes principales
 - Clustering (agrupamiento)
 - Clasificación
 - Regresión

¿Qué es?

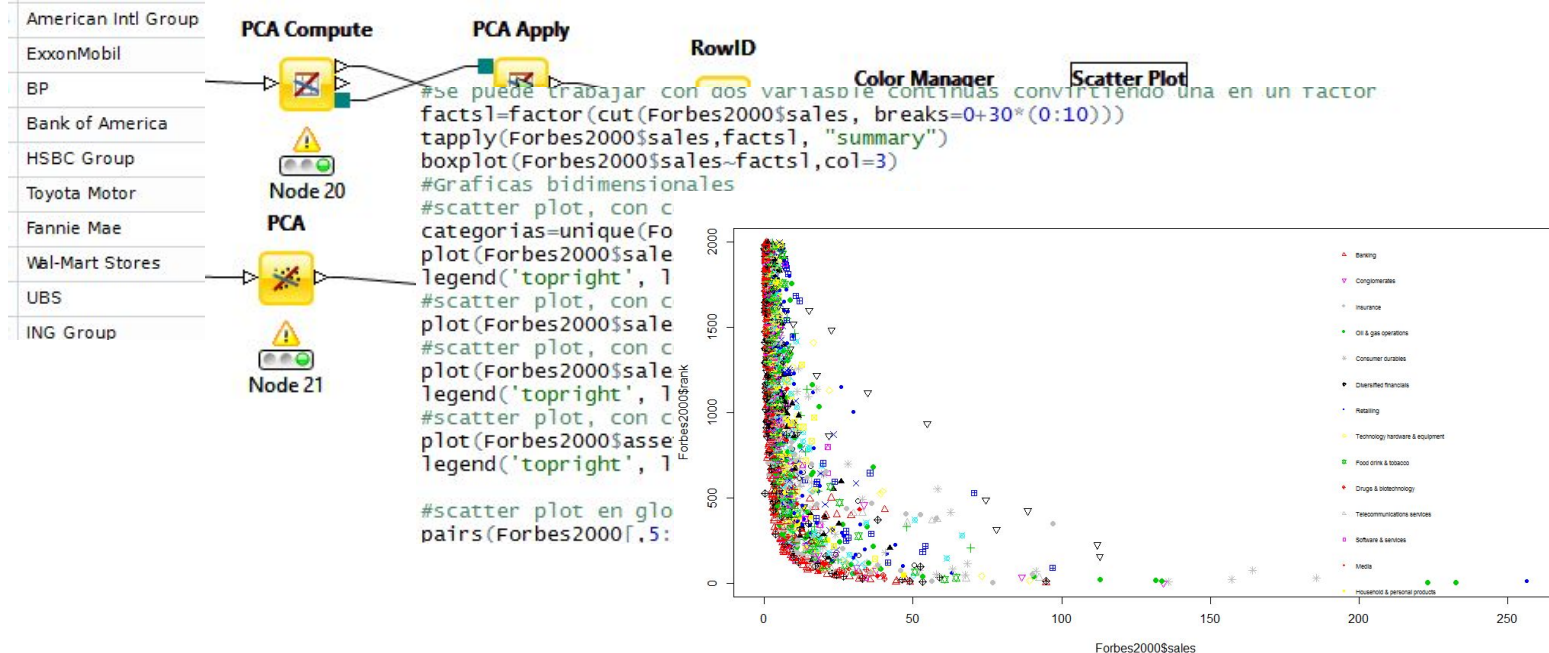
world's biggest
companies



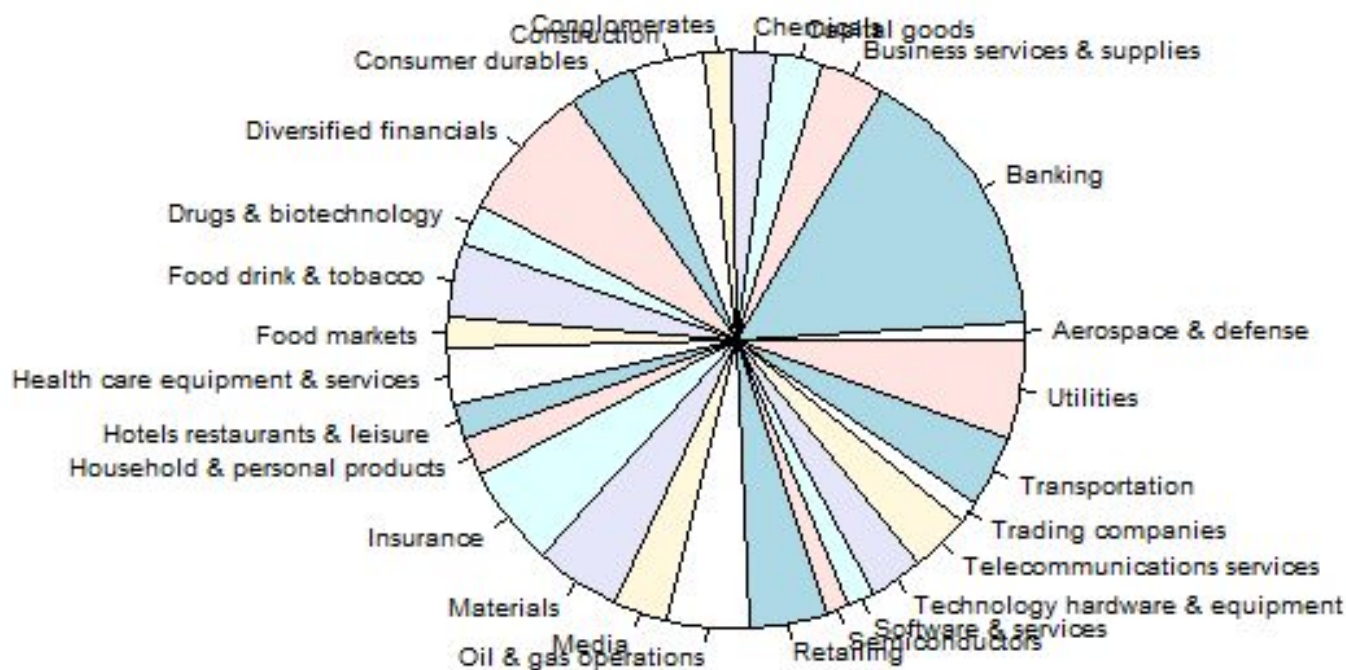
The Global 2,000

¿Qué haremos?

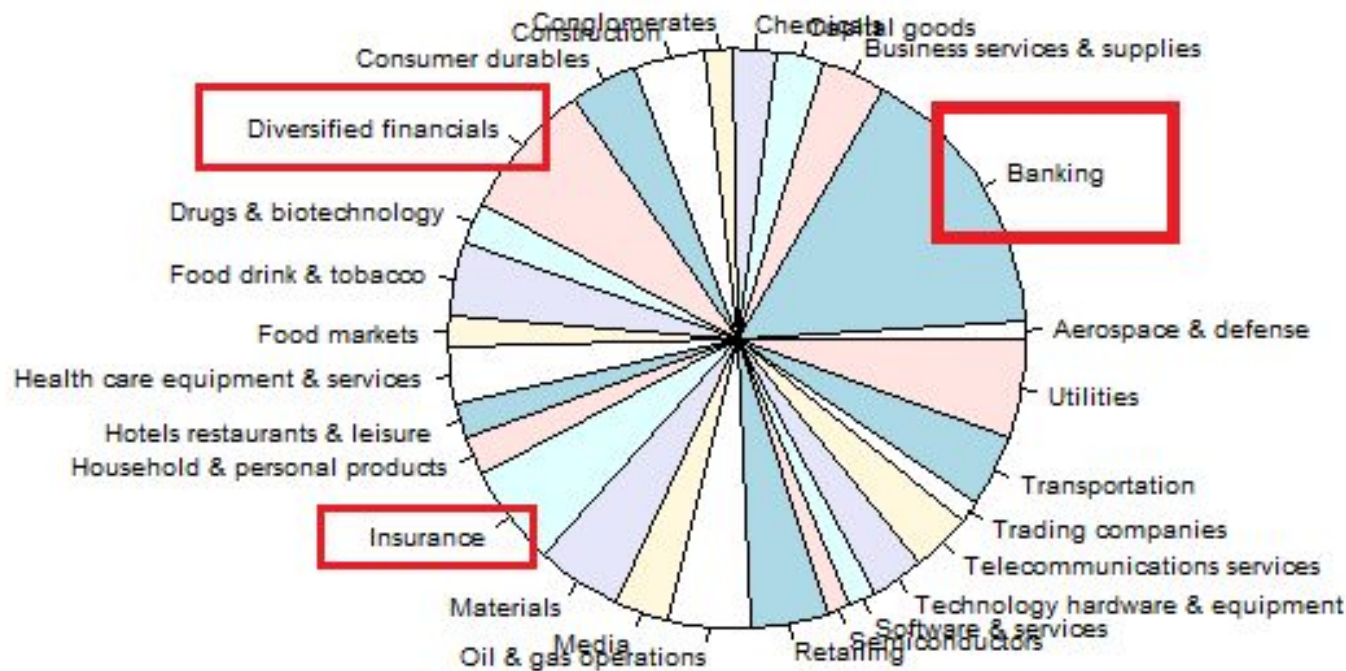
name	country	category	sales	profits
Citigroup	United States	Banking	94.71	17.85
General Electric	United States	Conglomerates	134.19	15.59



¿Cómo se reparten las categorías?



¿Cómo se reparten las categorías?



¿Cómo se reparten las categorías?

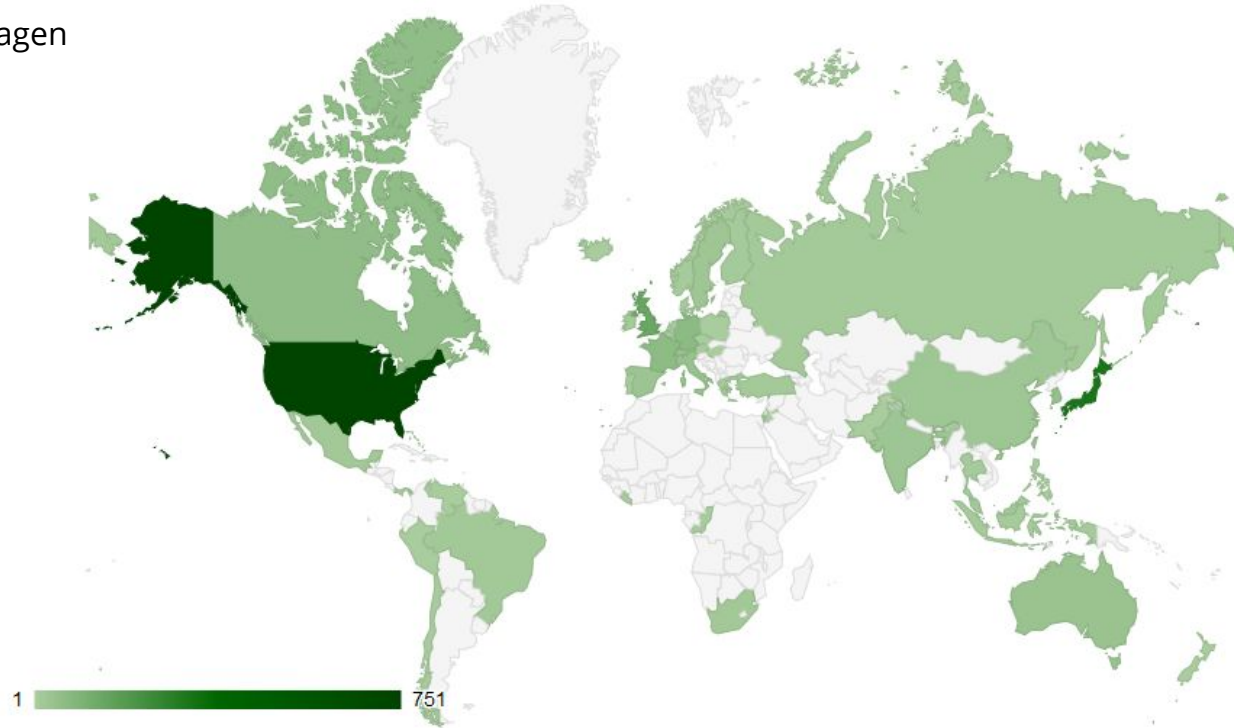
Aerospace & defense	Banking
19	313
Business services & supplies	Capital goods
70	53
Chemicals	Conglomerates
50	31
Construction	Consumer durables
79	74
Diversified financials	Drugs & biotechnology
158	45
Food drink & tobacco	Food markets
83	33
Health care equipment & services	Hotels restaurants & leisure
65	37
Household & personal products	Insurance
44	112
Materials	Media
97	61
Oil & gas operations	Retailing
90	88
Semiconductors	Software & services
26	31
Technology hardware & equipment	Telecommunications services
59	67
Trading companies	Transportation
25	80
Utilities	
110	

¿Cómo se reparten geográficamente?

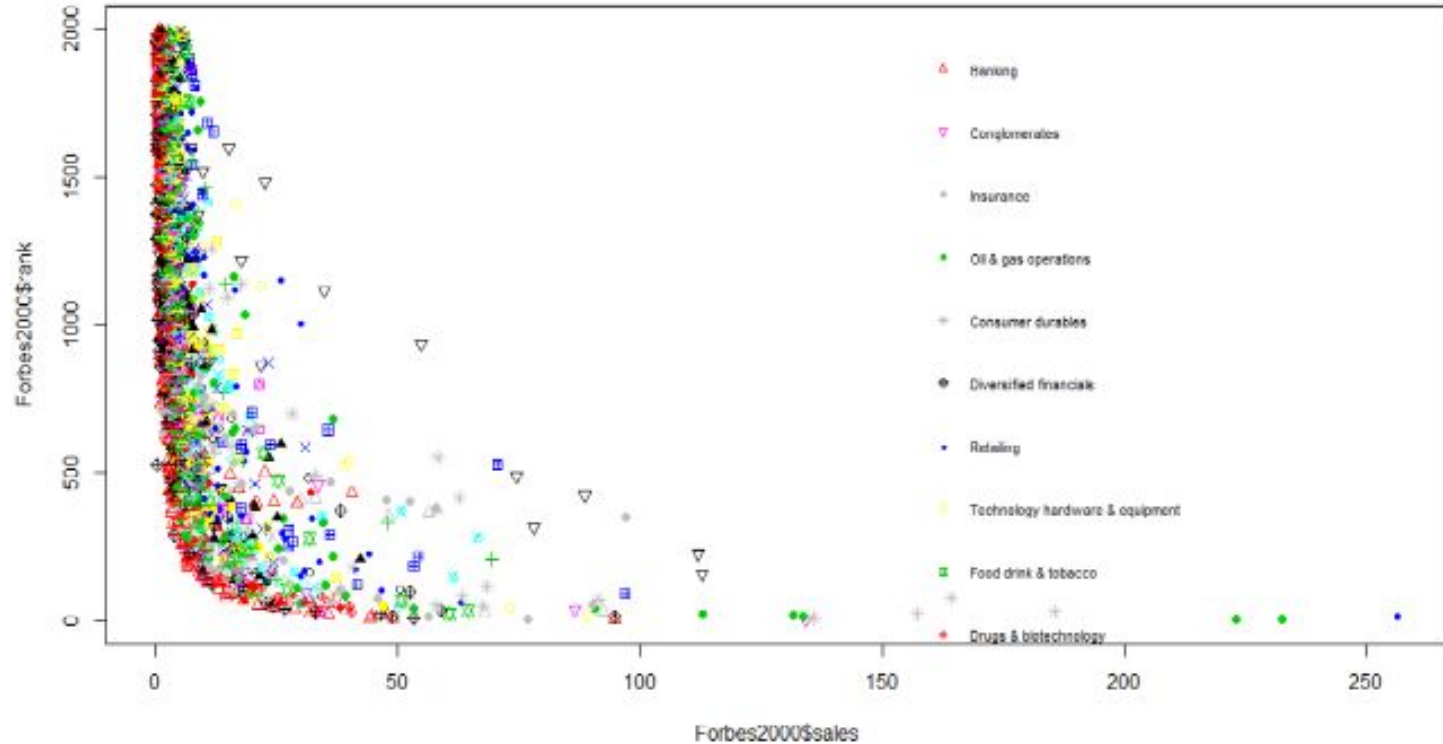
Click en la imagen

para el mapa

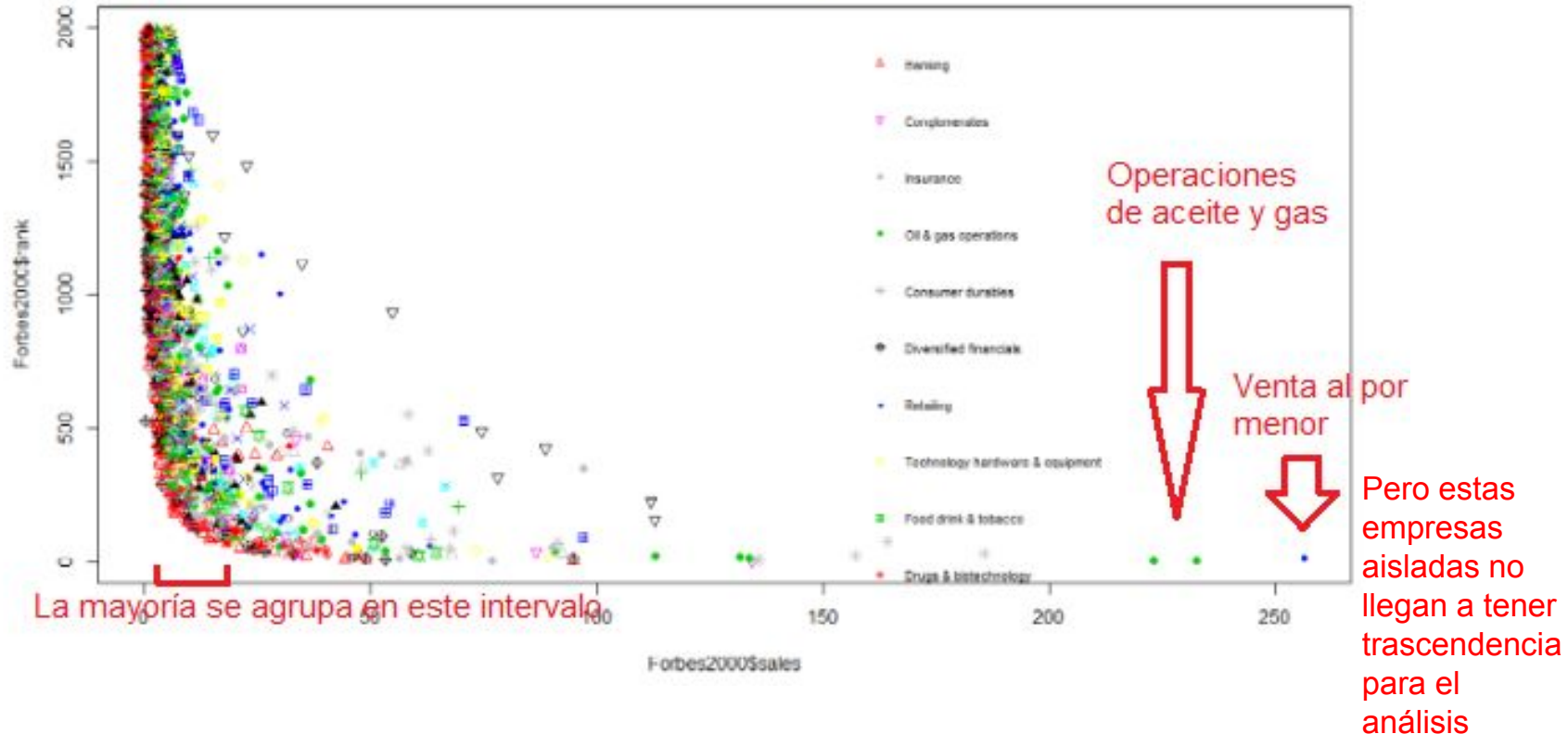
interactivo



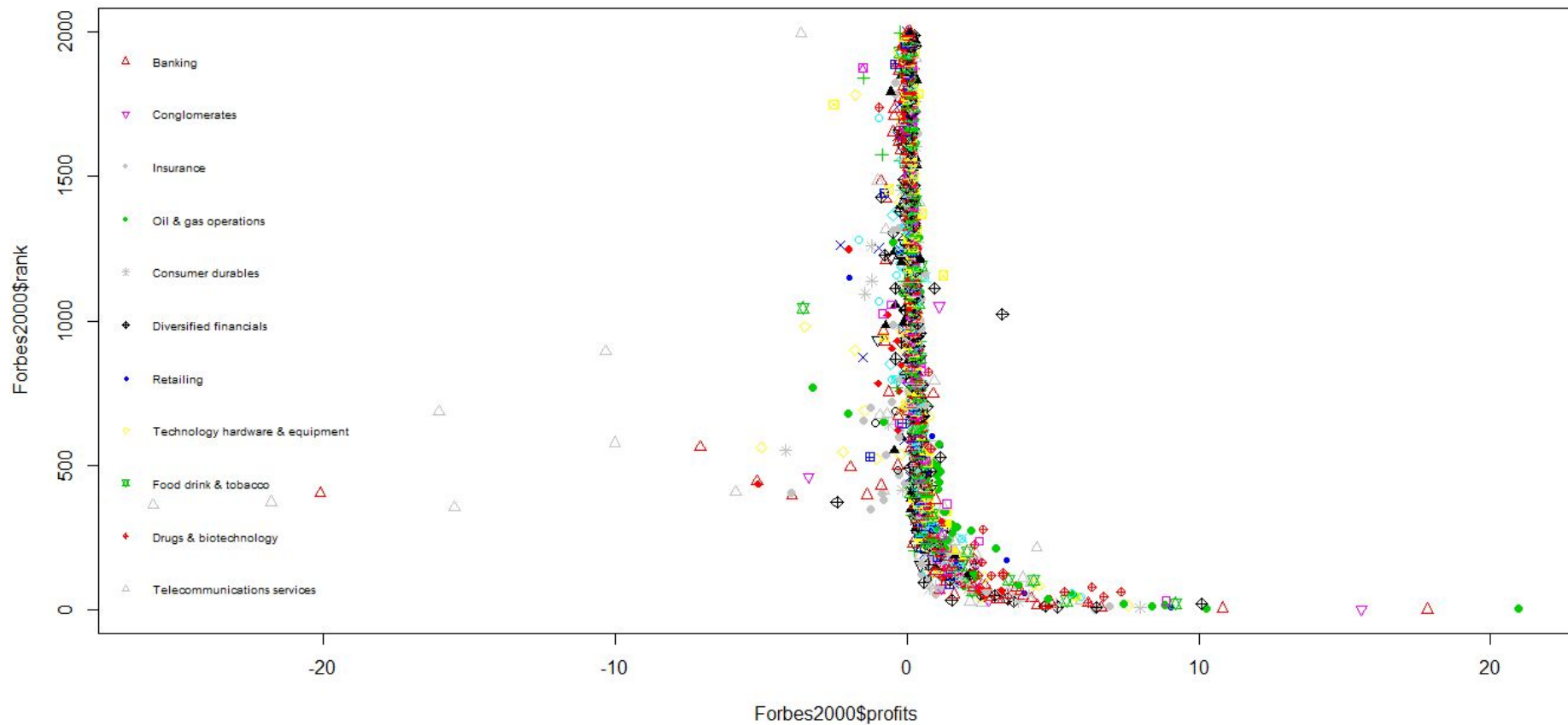
Comparamos varias componentes: ventas



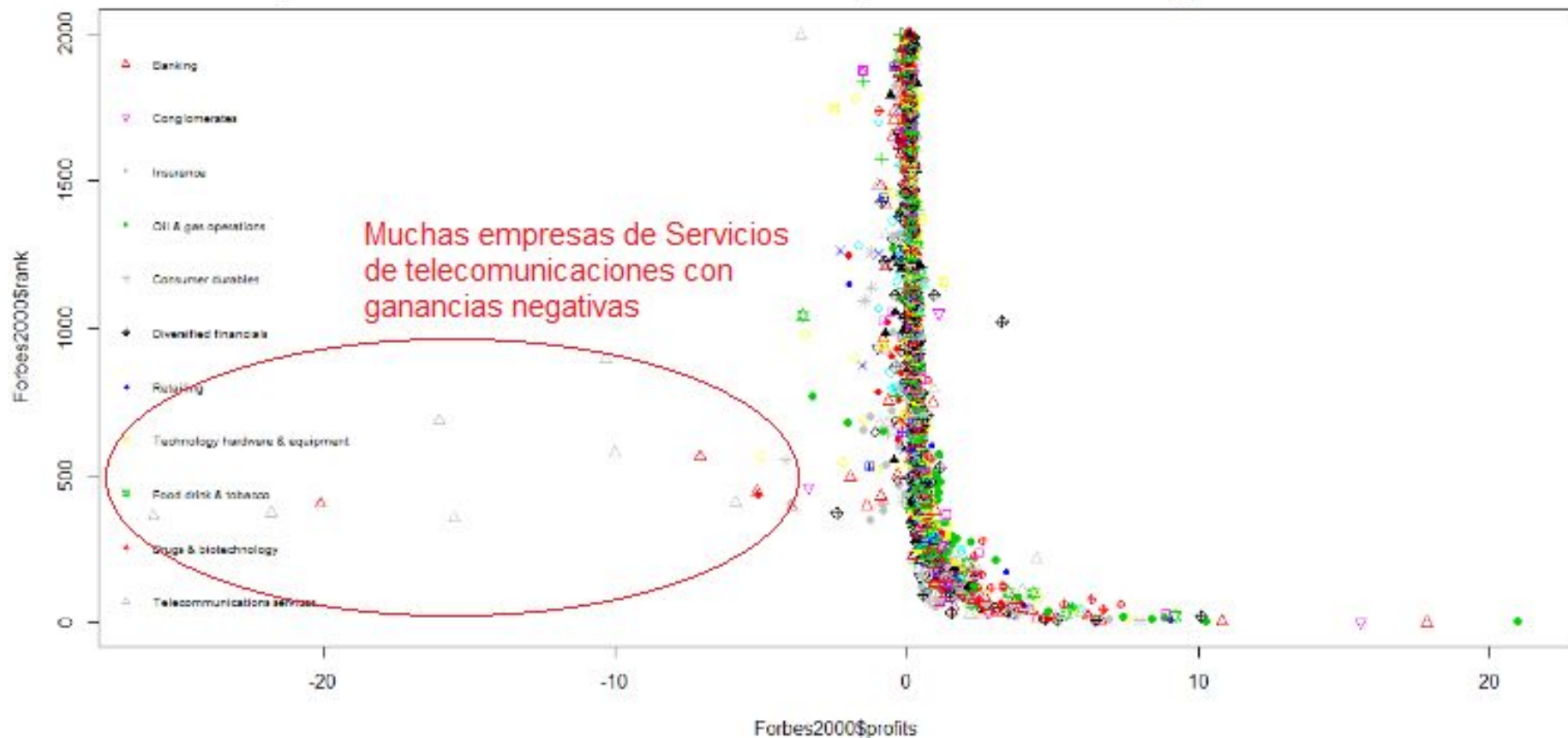
Comparamos varias componentes: ventas



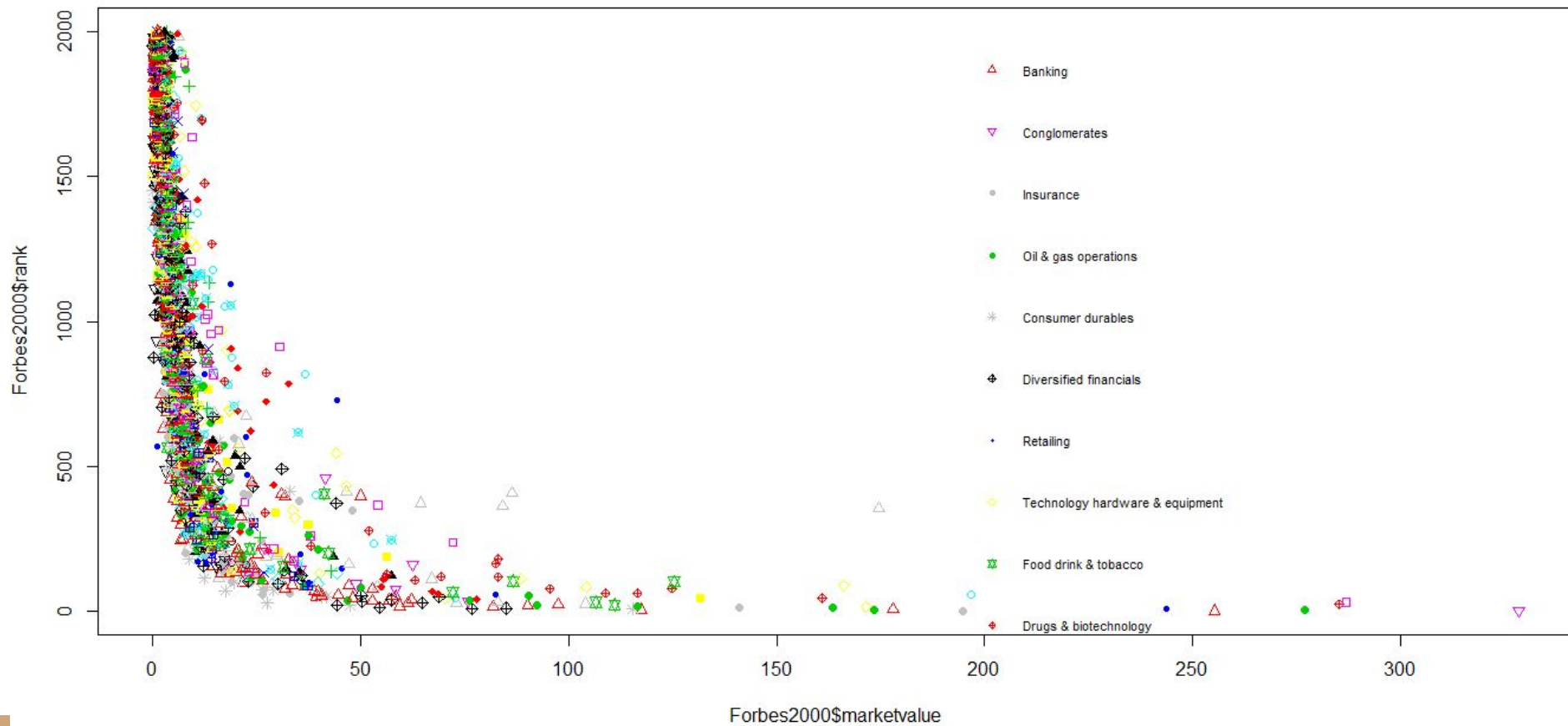
Comparamos varias componentes: ganancias



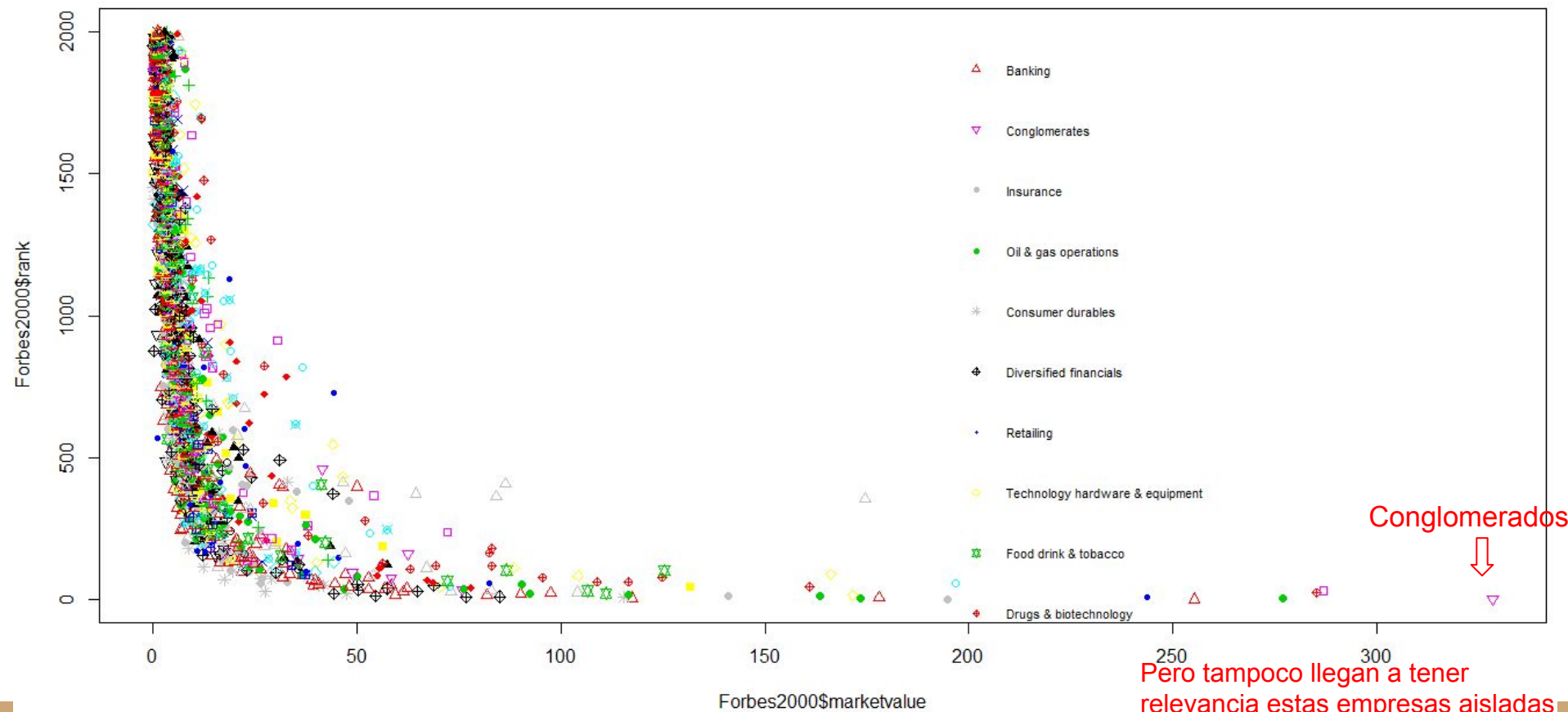
Comparamos varias componentes: ganancias



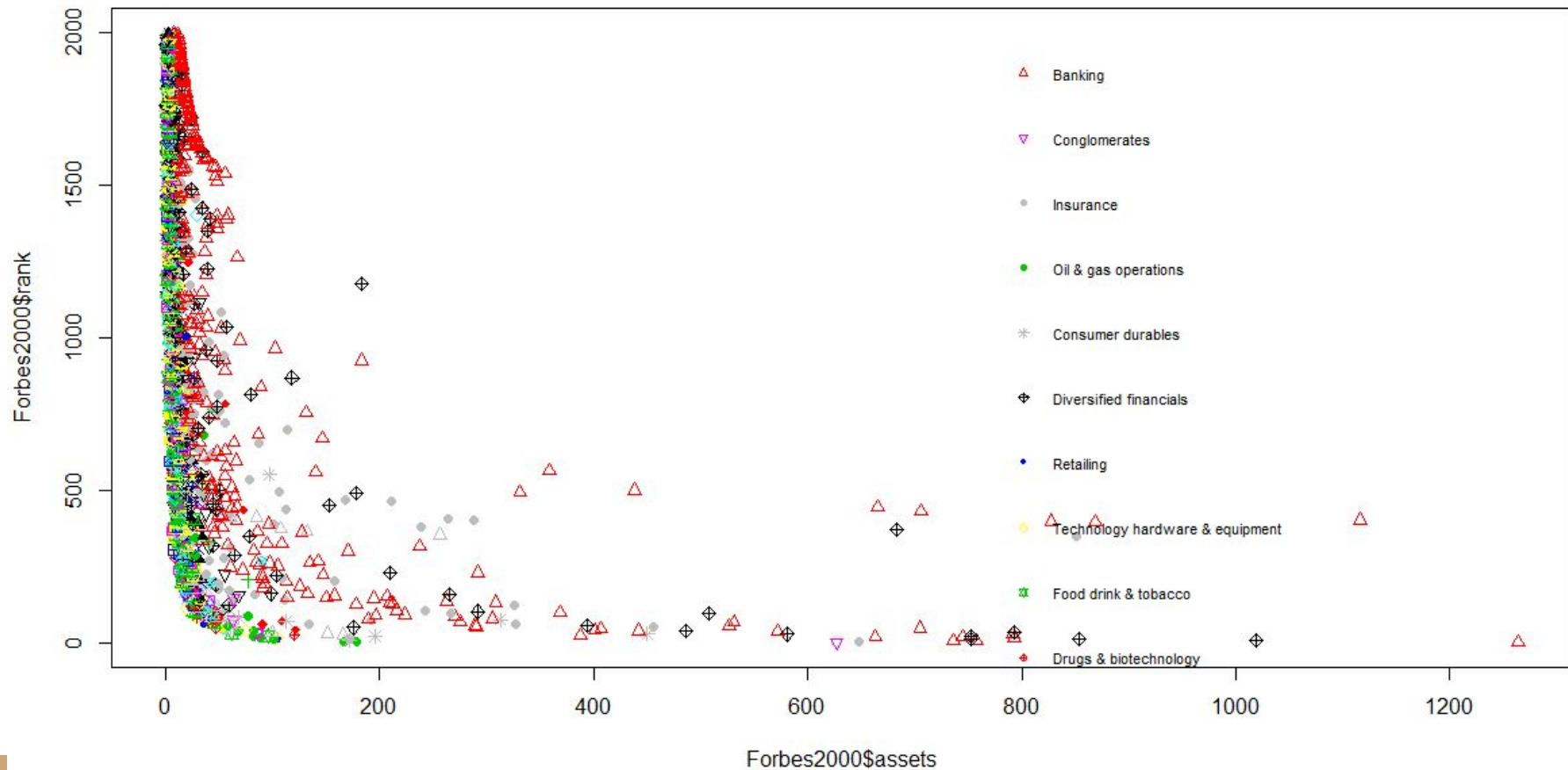
Comparamos varias componentes: valor de mercado



Comparamos varias componentes: valor de mercado



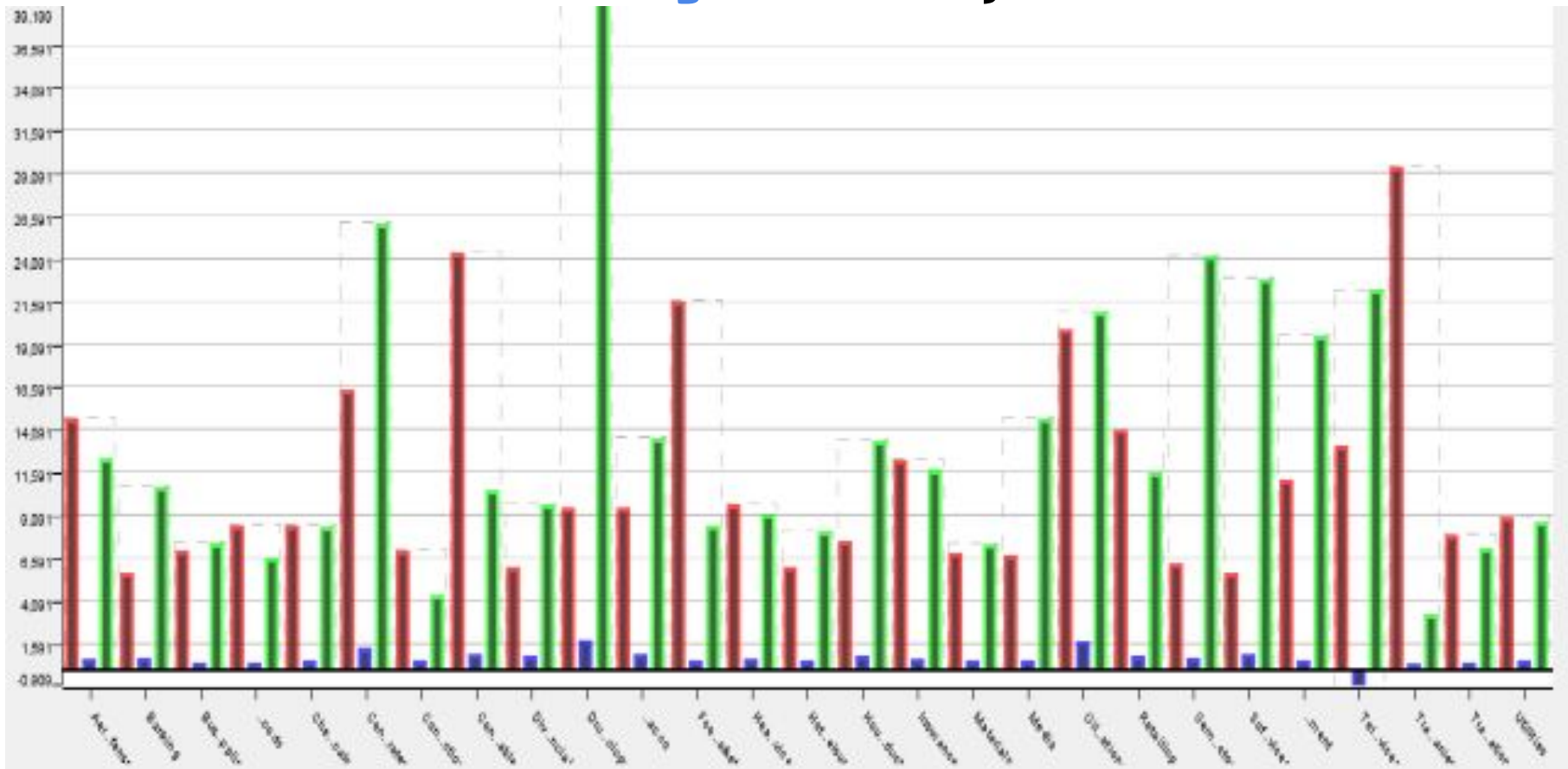
Comparamos varias componentes: bienes



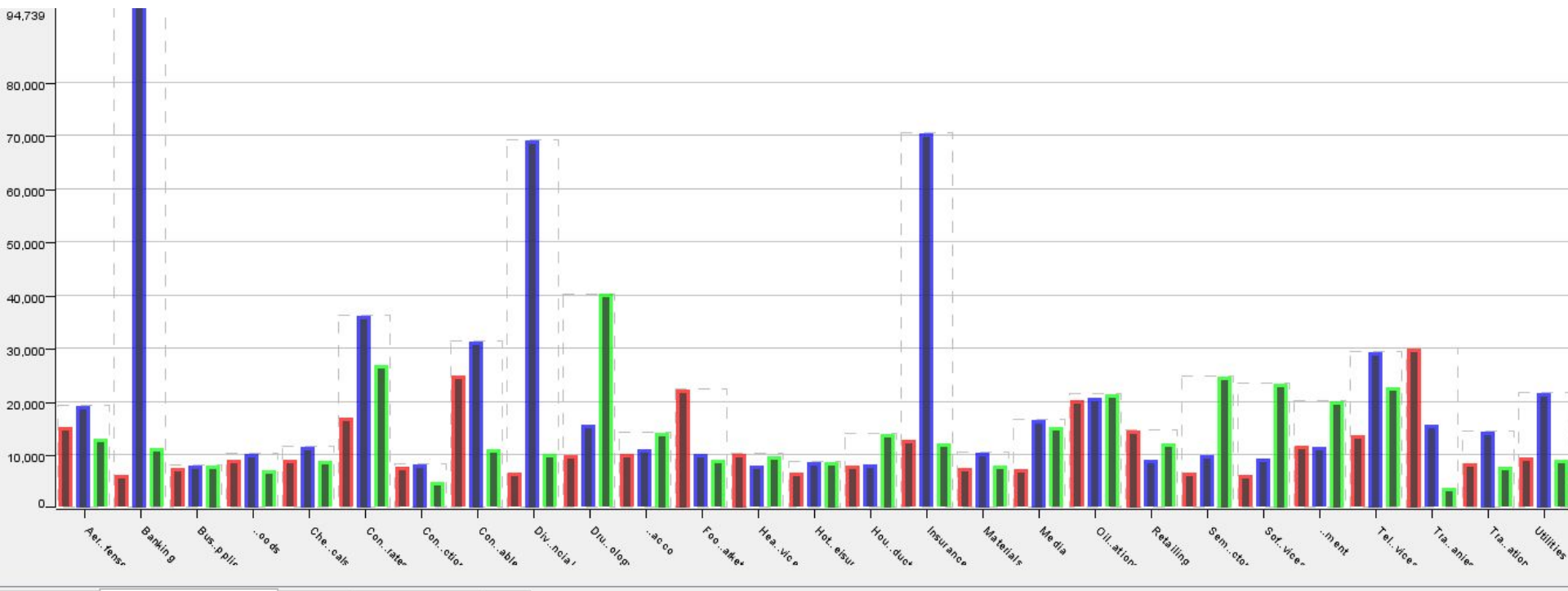
Comparamos varias componentes: bienes



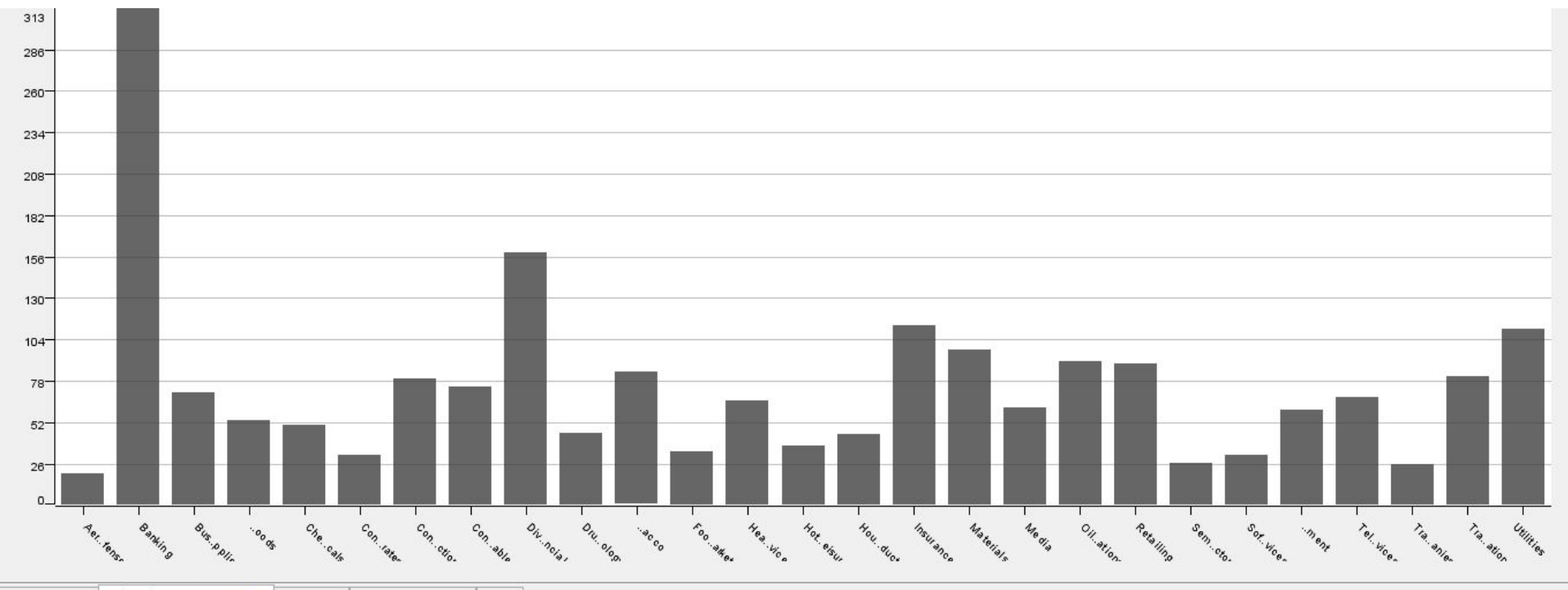
Promedio de **ventas**, **ganancias** y **valor de mercado**



Promedio de **ventas**, **bienes** y **valor de mercado**



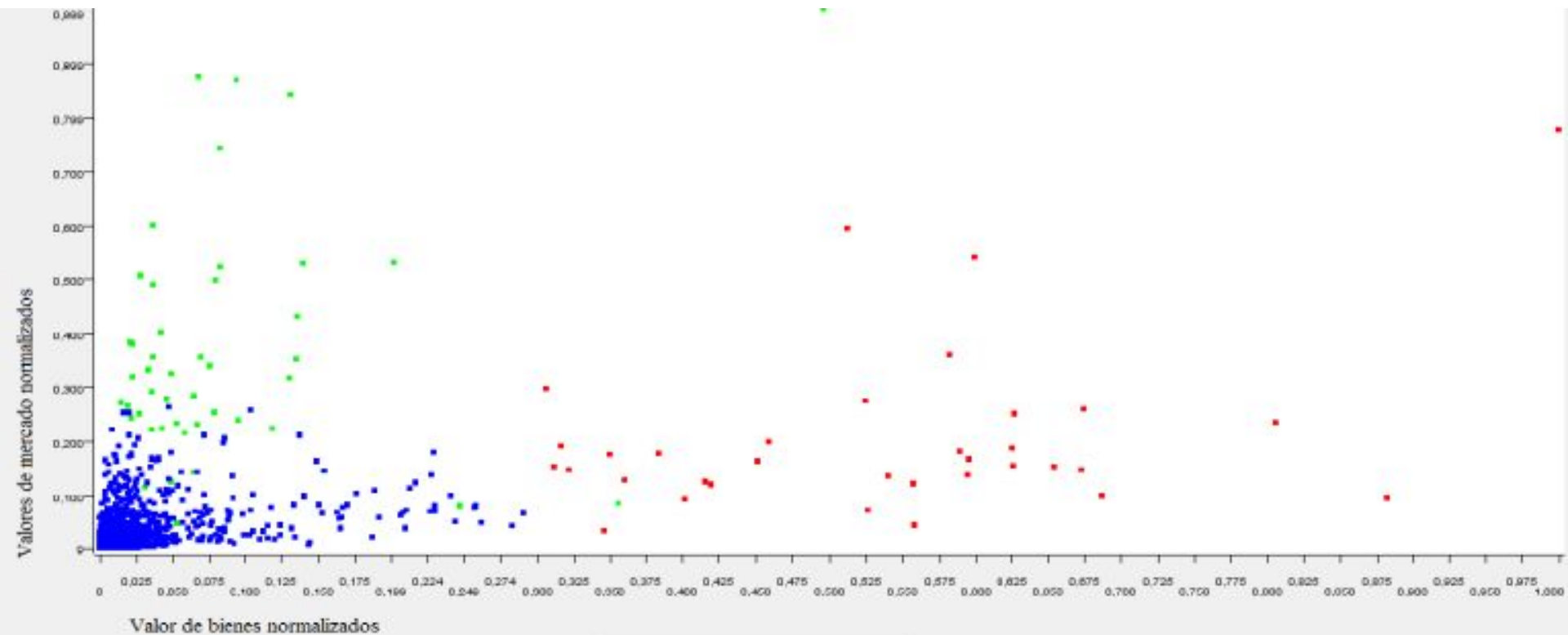
Empresas por categoría



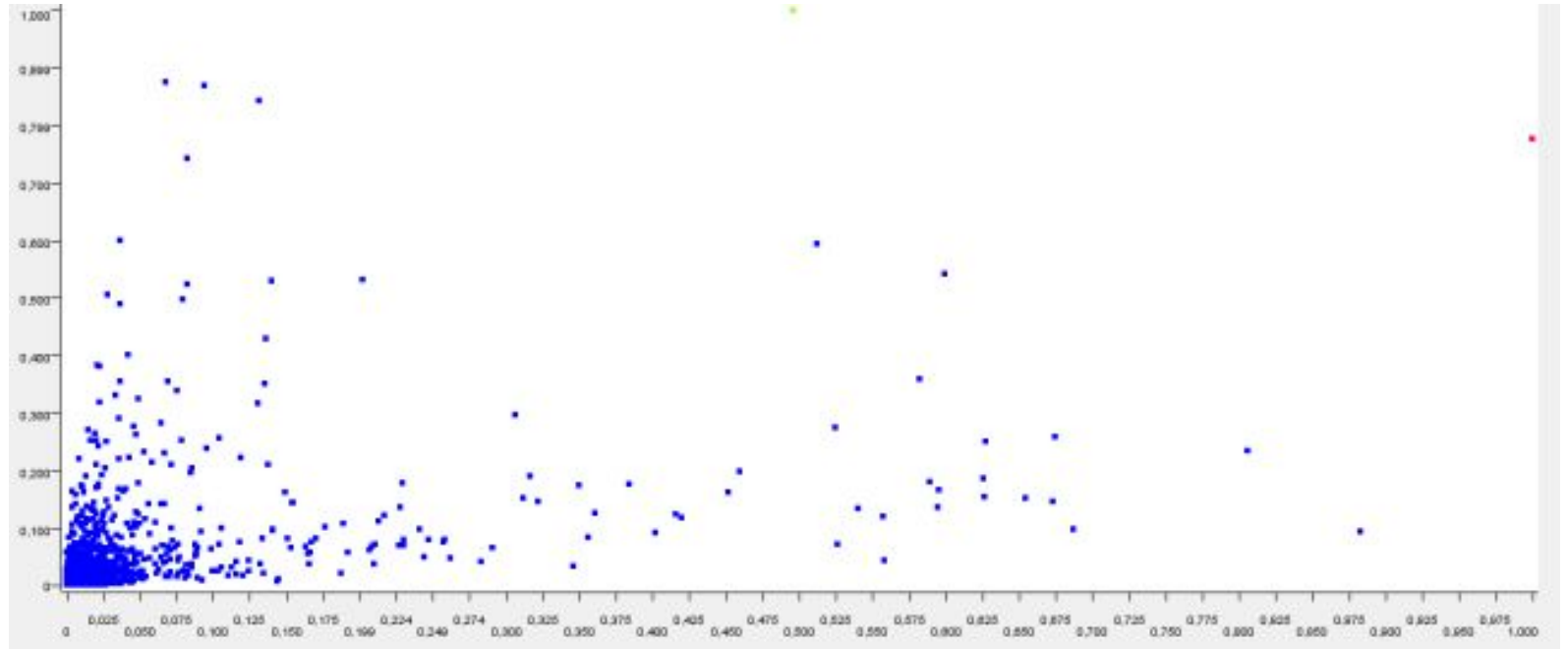
Podemos decir que ...

Como se vio al principio, las empresas más numerosas son las del sector **Bancario, Finanzas diversas y Aseguradoras**. Con algunas gráficas de dispersión y las de promedio (sobre todo en bienes) lo confirmamos, y podemos decir que estas 3 son las que tienen mayor “peso” .

Clustering: k-means

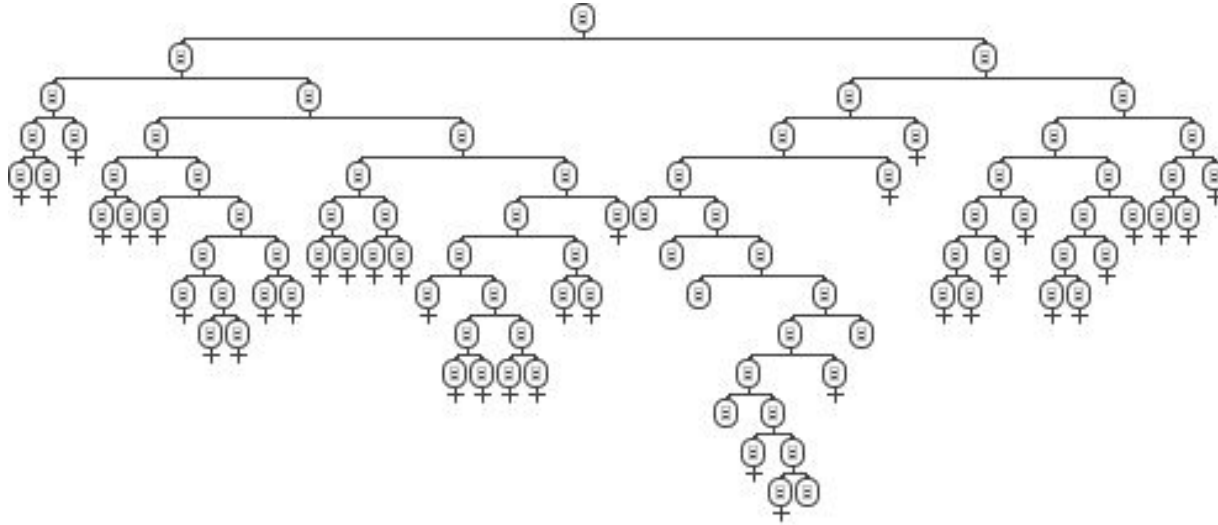


Clustering: jerárquico



El clustering jerárquico no nos da un resultado que aporte algo. Nos quedamos con el resultado del clustering usando k-means.

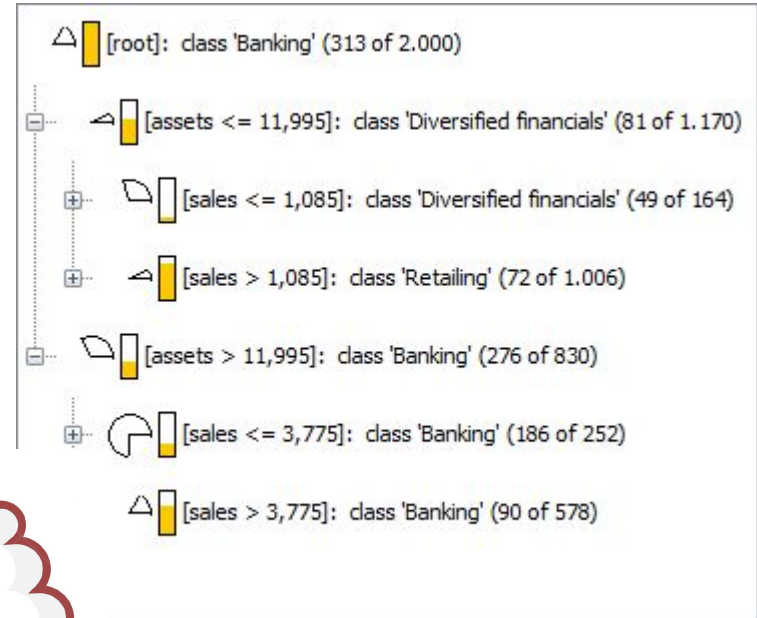
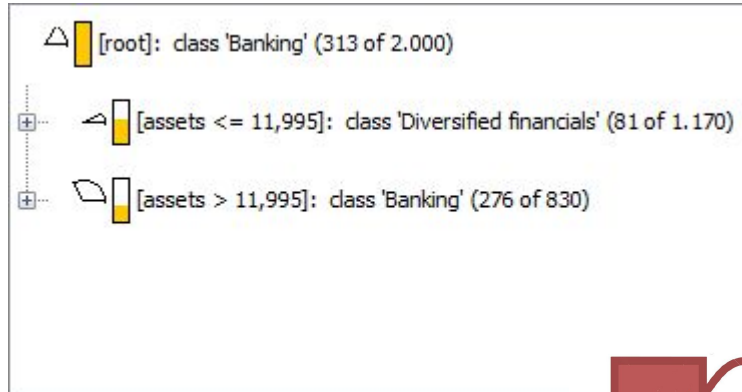
Clasificación: árboles de decisión



Sabiendo los datos no podemos predecir de qué categoría sería una empresa, ya que cada empresa tiene valores muy dispares.

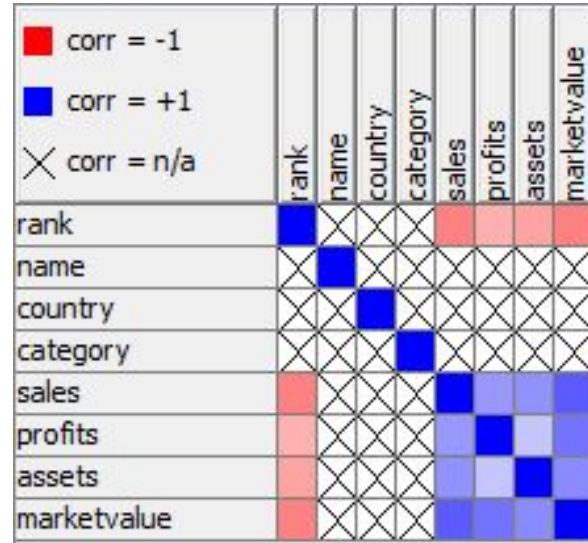
Clasificación: árboles de decisión

Mostramos los 2 primeros niveles de nodos:



Regresión: Correlación

No vemos una correlación lineal clara, pero podemos decir que ningún país o categoría influye en el ranking.



¿Preguntas?

Puedes mandarme [un correo](#), o mediante [un issue](#)