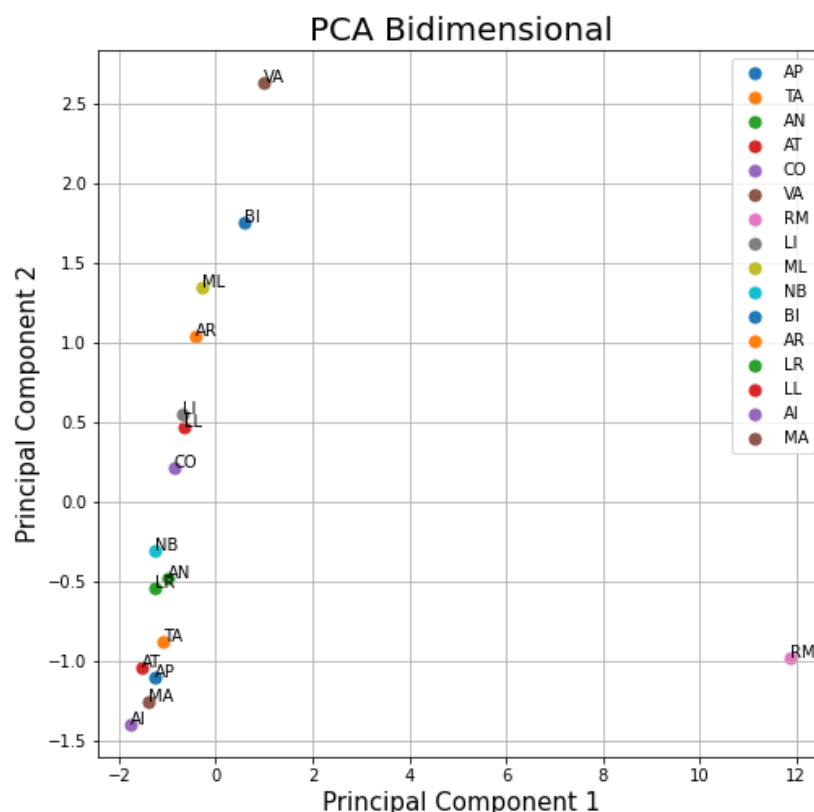


**PREGUNTA 5 Tarea 1: Gabriel Díaz Vásquez 21.123.354-0**

R. Trabajaremos con una base de datos realizada por la Encuesta nacional de Empleo (ENE), dónde podemos apreciar datos de varios trimestres pero en particular crearemos una sub base de datos con los resultados del último trimestre separados por regiones. Nos fijamos que existe una cantidad considerable de variables como puede ser "Directores, gerentes y administradores", "Profesionales, científicos e intelectuales", "Personal de apoyo administrativo", entre otros. Lo que para el análisis de datos vienen a ser dimensiones, tener varias dimensiones hace que el estudio sea más preciso, sin embargo, como sabemos solo podemos graficar y observar hasta 3 dimensiones, nuestro objetivo es poder trabajar con datos de manera eficaz y más simplificada, para esto existe el análisis de componentes principales y de esta manera reducir dimensiones, conservando la mayor parte de información importante.

En nuestro ejercicio haremos el análisis para 2 y 3 dimensiones.

N=2



Podemos apreciar que se habla acerca de 2 componentes principales donde un grupo es completamente notorio y forman cierto patrón a excepción de la Región Metropolitana la cual viene a ser un punto lejano que escapa del común del estudio.

```

      Region ... No sabe No responde 3
0      Región de Arica y Parinacota (AP) ... 0.08
1      Región de Tarapacá (TA) ... 0.40
2      Región de Antofagasta (AN) ... 0.22
3      Región de Atacama (AT) ... 0.12
4      Región de Coquimbo (CO) ... 0.00
5      Región de Valparaíso (VA) ... 0.32
6      Región Metropolitana de Santiago (RM) ... 8.42
7      Región del Libertador General Bernardo O'Higgi... ... 0.00
8      Región del Maule (ML) ... 0.00
9      Región del Bío-Bío (NB) ... 0.00
10     Región del Biobío (BI) ... 0.28
11     Región de La Araucanía (AR) ... 0.00
12     Región de Los Ríos (LR) ... 0.27
13     Región de Los Lagos (LL) ... 0.00
14     Región de Aysén del General Carlos Ibáñez del ... ... 0.0
15     Región de Magallanes y de la Antártica Chilena... ... 0.00

[16 rows x 13 columns]
(16, 13)
[0.82649642 0.11300372]

      Variable      PC1      PC2
0      Población ocupada (total) -0.039806 0.826461
1      Directores gerentes y administradores 0.311115 -0.150298
2      Profesionales científicos e intelectuales 0.316521 -0.056650
3      Técnicos y profesionales de nivel medio 0.316973 -0.008537
4      Personal de apoyo administrativo 0.315796 -0.080878
5      Trabajadores de los servicios y vendedores de... 0.316944 0.024278
6      Agricultores y trabajadores calificados agrop... 0.220358 0.489897
7      Artesanos y operarios de oficios 0.316095 0.065163
8      Operadores de instalaciónes máquinas y ensamb... 0.315450 0.048704
9      Ocupaciones elementales 0.314251 0.075278
10     Otros no identificados 0.240304 0.022329
11     No sabe No responde 3 0.310328 -0.176572

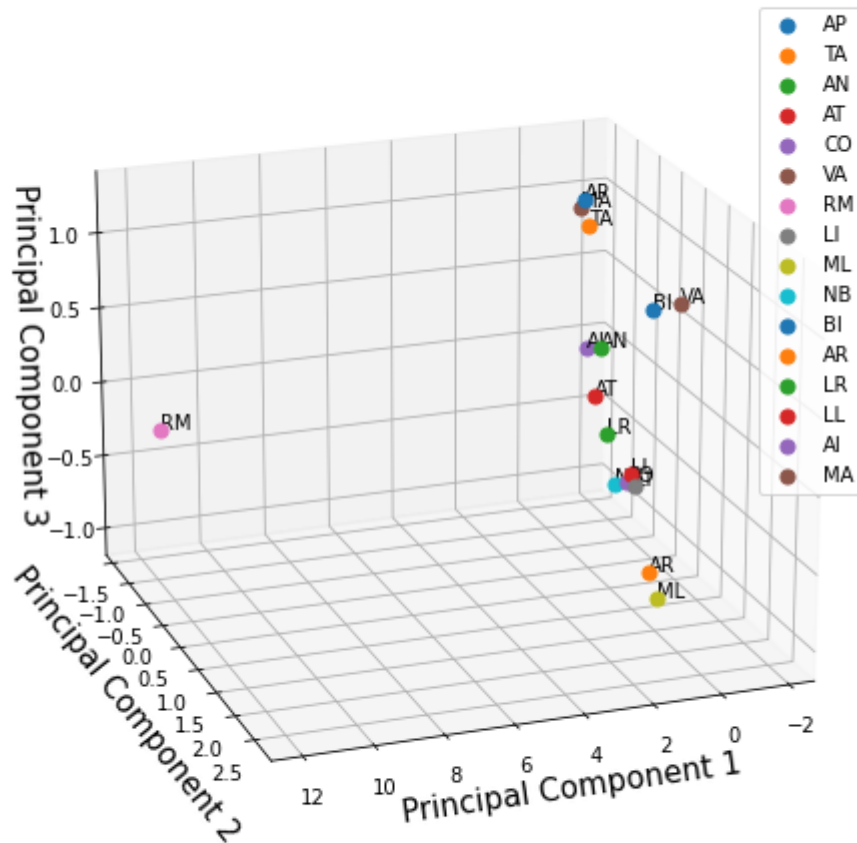
```

Podemos observar la varianza “[0.82649642 0.11300372]” que nos dice que tanta información se atribuye a cada componente principal donde vemos fuerte superioridad en la primera componente a diferencia de la segunda.

Y de manera acumulada seguimos teniendo el 0.93950014 que expresado es aprox 94%. De esta manera bajamos las dimensiones a 2 componentes manteniendo gran parte de la información.

N = 3

## PCA tridimensional



Donde al igual que en la parte anterior vemos que se puede ver otro cambio en los patrones de los datos de manera que algunos tienen cierto nivel en esta nueva componente 3 lo cual era poco posible de visualizar en solo 2 componentes. Y que de igual manera influye en la distribución de los datos donde podemos seguir visualizando y analizando gracias a PCA.

```
[16 rows x 13 columns]
(16, 13)
```

Warning

Figures now render in the Plots pane by default. To make them also appear inline in the Console, uncheck "Mute Inline Plotting" under the Plots pane options menu.

```
[0.82649642 0.11300372 0.04594487]
```

|    | Variable   | ... | PC3       |
|----|--|-----|-----------|
| 0  | Poblacion ocupada (total)                        | ... | 0.235794  |
| 1  | Directores gerentes y administradores            | ... | -0.080554 |
| 2  | Profesionales cientificos e intelectuales        | ... | -0.022438 |
| 3  | Tecnicos y profesionales de nivel medio          | ... | 0.014800  |
| 4  | Personal de apoyo administrativo                 | ... | -0.046465 |
| 5  | Trabajadores de los servicios y vendedores de... | ... | 0.004940  |
| 6  | Agricultores y trabajadores calificados agrop... | ... | -0.461526 |
| 7  | Artesanos y operarios de oficios                 | ... | -0.009936 |
| 8  | Operadores de instalacione maquinas y ensamb...  | ... | -0.032538 |
| 9  | Ocupaciones elementales                          | ... | -0.094059 |
| 10 | Otros no identificados                           | ... | 0.843400  |
| 11 | No sabe No responde 3                            | ... | -0.025915 |

```
[12 rows x 4 columns]
```

Y aquí la Varianza la vemos como “[0.82649642 0.11300372 0.04594487]” donde vemos que sigue dominando la componente 1 mientras la componente 3 viene a ser la nueva menor componente que aún así se lleva cierta cantidad de los datos, y el total de los datos acumulados aplicados con 3 componentes es del 0.98544501 lo que en porcentaje es aprox 99%.