

# Análisis de Componentes Principales

## GRUPO 6

- Francisco Sendot
- Lucia Digon
- Martín E. Zahnd
- Juan Ignacio Fernández Dinardo



# Conjunto de Datos y Objetivo

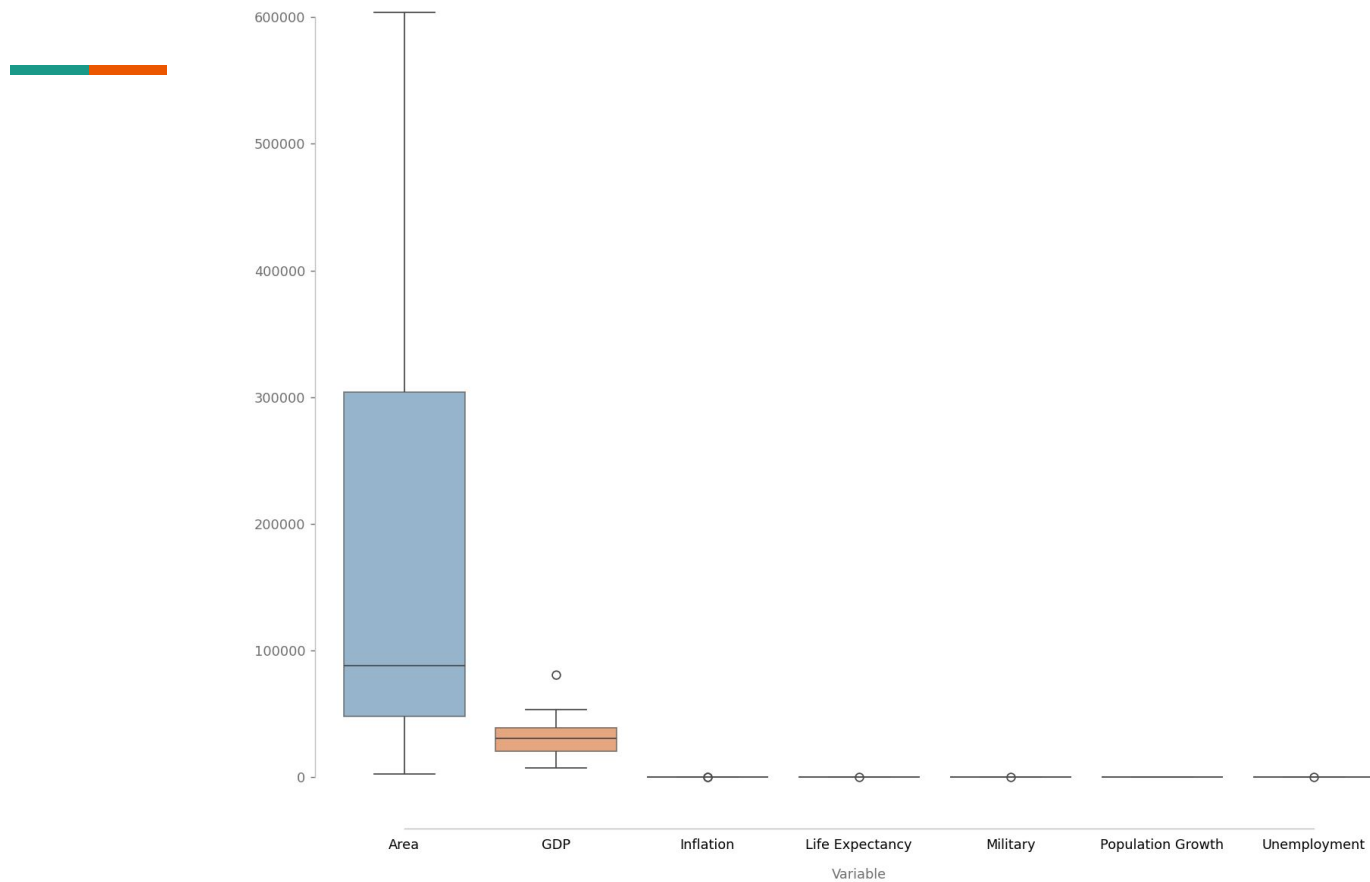
Se posee un conjunto de datos correspondiente a características **geográficas, económicas y sociales** de 28 países de Europa

- País (*Country*)
- Área (*Area*)
- PBI (*GDP*)
- Inflación (*Inflation*)
- Esperanza de Vida (*Life.expect*)
- Defensa (*Military*)
- Crecimiento Poblacional (*Pop.growth*)
- Desempleo (*Unemployment*)

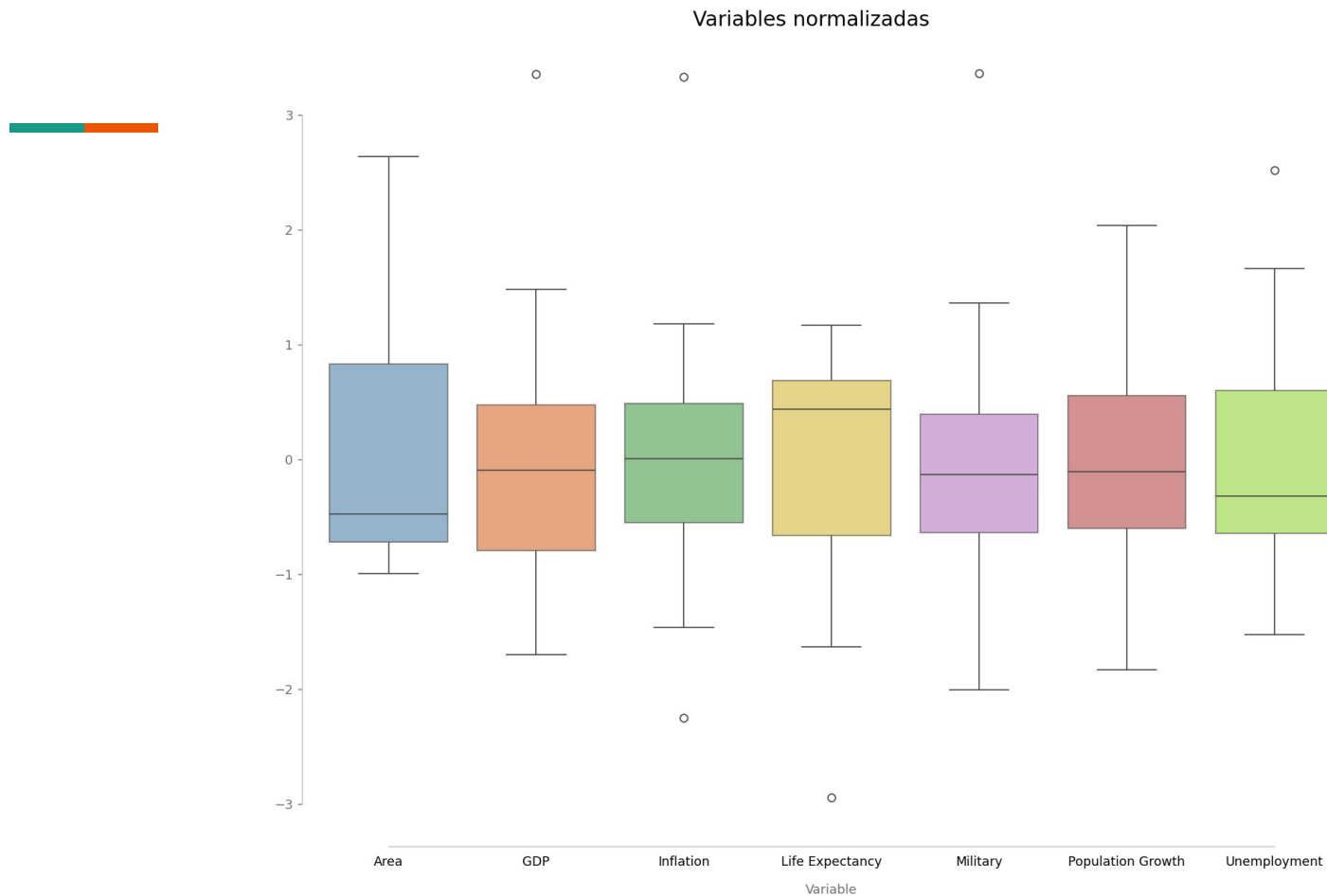
Se busca ordenar dichas variables a través de un índice con el objetivo de **proponer nuevas políticas públicas**

# Boxplot: Sin Normalizar

Variables sin normalizar



# Boxplot: Normalizado



# Boxplot

## Análisis



- Analizando el boxplot en su versión sin normalizar, está claro que el Área tiene una escala mucho mayor al resto de las variables, lo cual generaría un impacto desproporcionado en el análisis de la PCA.
- Una vez normalizadas, el impacto que genera cada variable en el análisis se observa más parejo, que es lo buscado. También se elimina el sesgo generado por la diferencia de escalas, unidades y métodos de medición aplicados a cada variable.
- El proceso de normalización vuelve más interpretables las cargas de las variables, es decir, los coeficientes de contribución que tiene cada variable sobre las componentes principales.

# Primera Componente



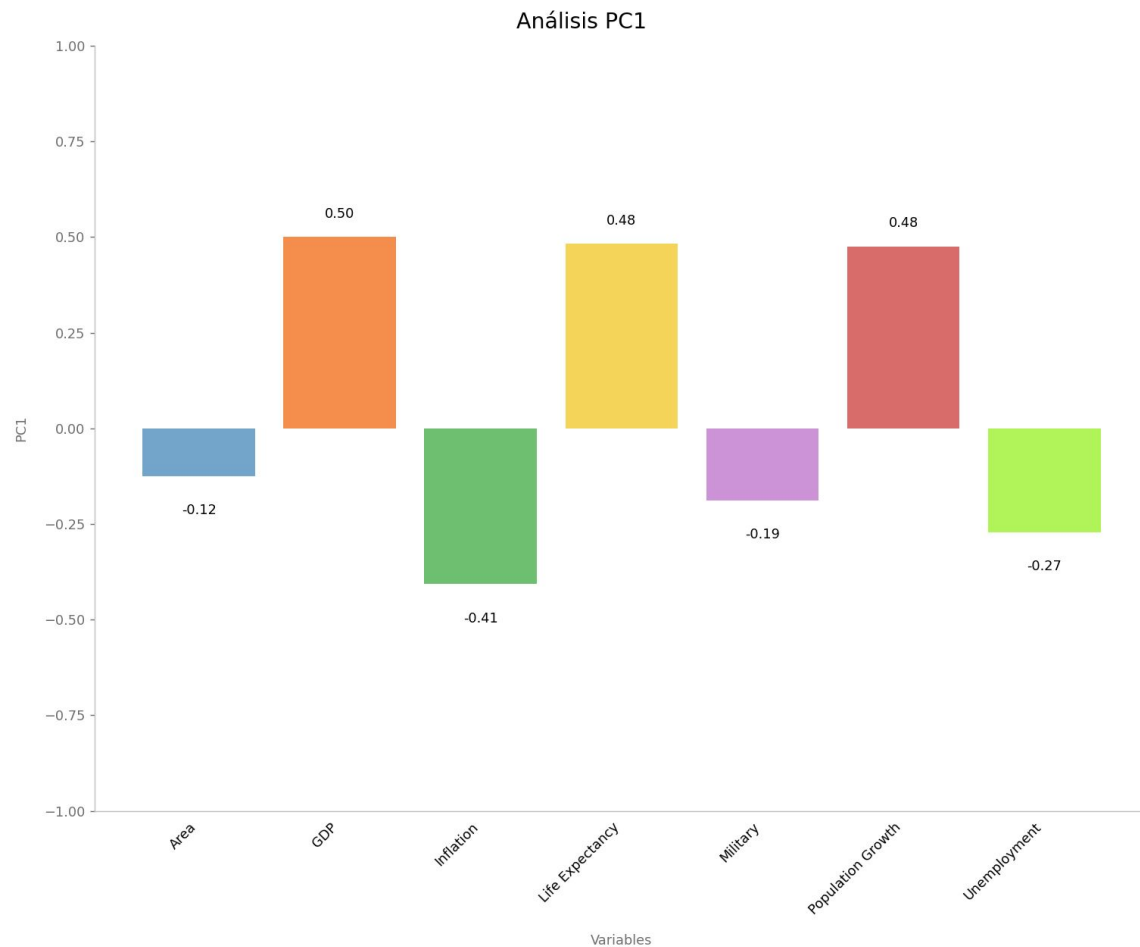
# Primera Componente

## Análisis Preliminar



- Al realizar el análisis de la primera componente, en una primera impresión vemos que Ucrania es un outlier con un score mayor por 1,08, en módulo, que el siguiente país en la lista: Luxemburgo. El tercer puesto se lo lleva Suiza.
- Sin embargo, muchos países tienen un score menor a cero, y otros uno mayor.
- ¿Qué significa esto?

# Primera Componente: Variables





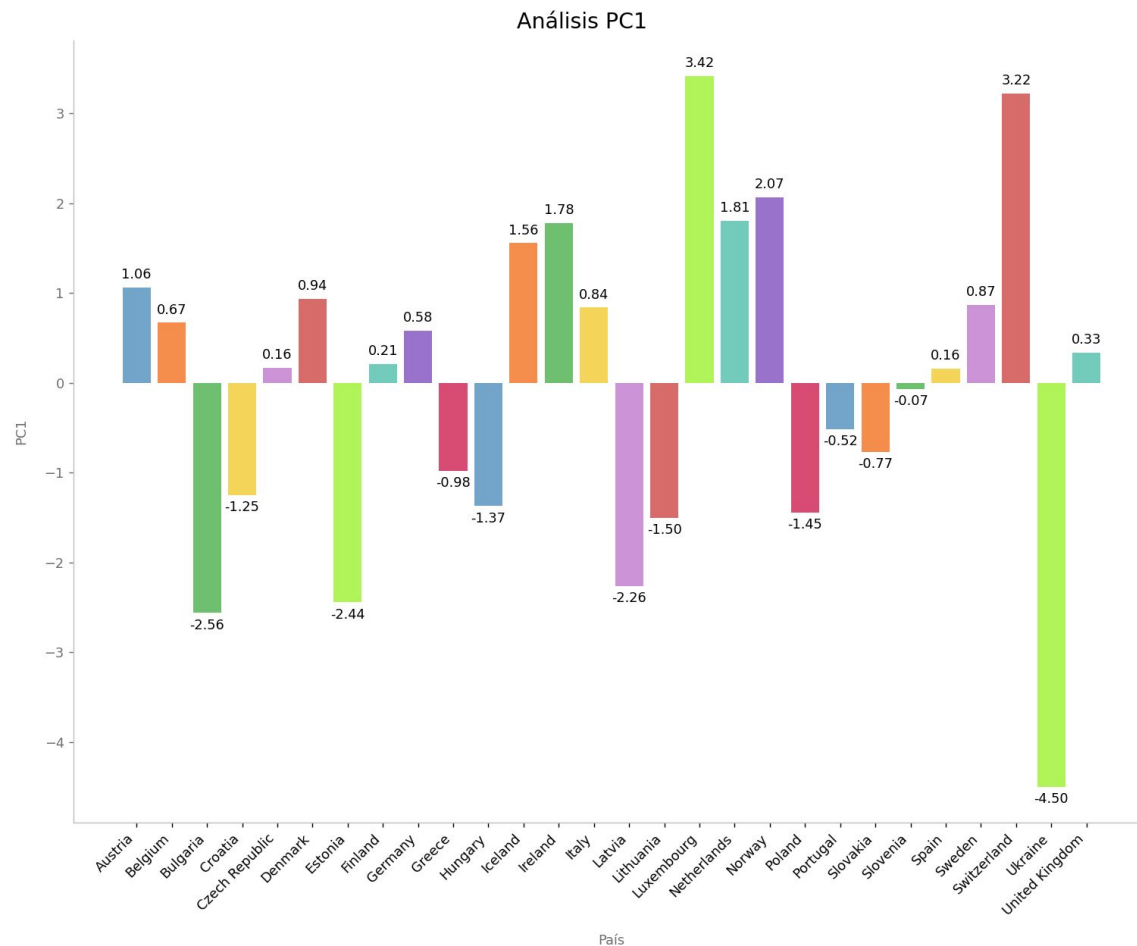
# Primera Componente

## Análisis Preliminar



- Observando cómo la PC1 agrupa las variables, notamos que la inflación, el desempleo, el gasto militar y, en menor medida, el área forman un conjunto de características, que llamaremos "baja calidad de vida"; mientras que el PBI, la esperanza de vida y el crecimiento poblacional forman otro, al que llamaremos "alta calidad de vida".

# Primera Componente



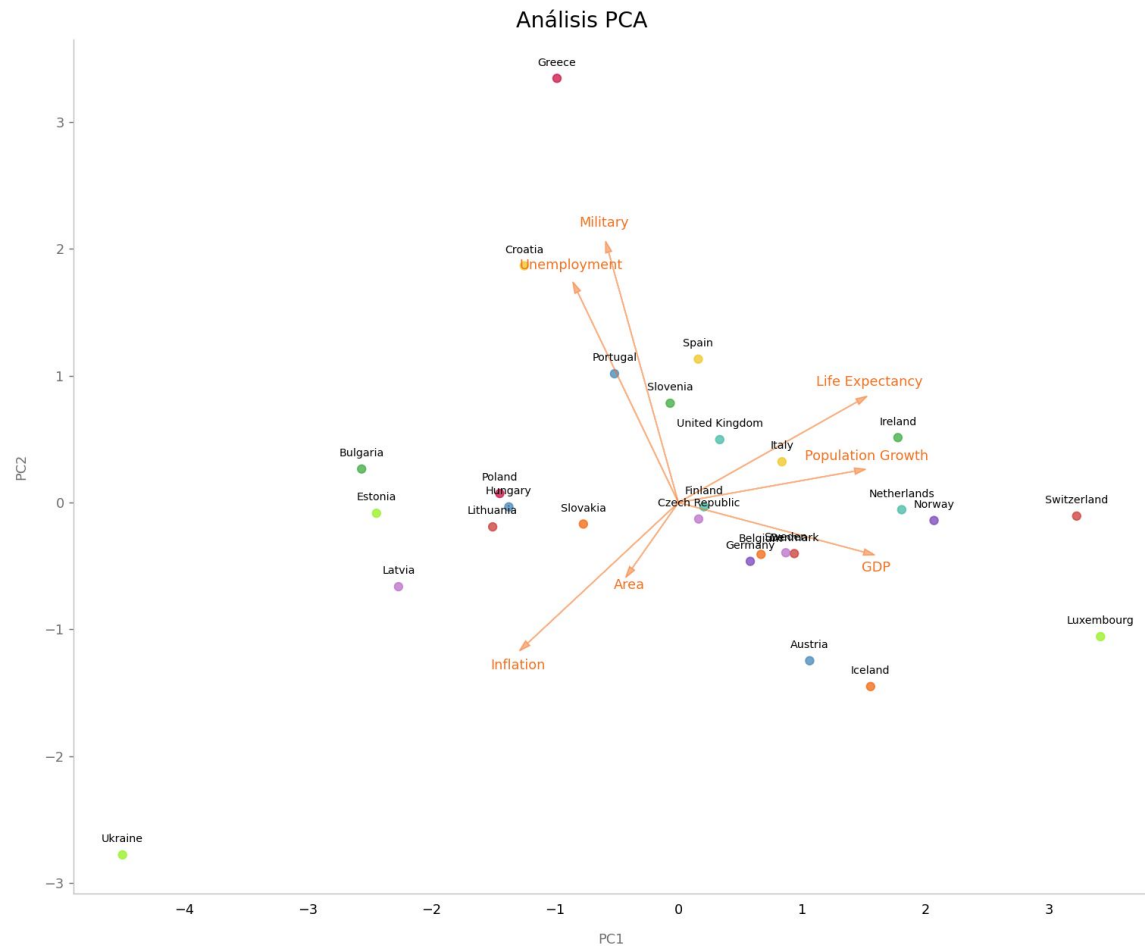
# Primera Componente

## Análisis



- Regresando al gráfico de barras, comprendemos que aquellos países con un score menor a cero son los que se ven mayormente afectados por factores del primer grupo (inflación, desempleo, gasto militar, y área), y los que tienen score positivo pertenecen al segundo grupo.
- Esta lectura nos permite identificar a países como Ucrania, Bélgica, Estonia, etc, como países en los que la inflación y las otras tres variables del grupo poseen una alto peso; mientras que países como Luxemburgo, Suiza, Irlanda y Noruega gozan de lo que denominamos una "mayor calidad de vida" que su contraparte.
- Cabe destacar que un "puntaje negativo" no implica que esa variable "es mala". Los conjuntos de variables podrían tener signo opuesto y los resultados del análisis serían los mismos.

# Biplot: PCA con 2 Componentes



# Biplot: PCA con 2 Componentes

## Análisis



- Viendo, por último, el biplot, encontramos a Ucrania muy alejado de cualquier otro grupo, en particular, se encuentra en el mismo sentido que la inflación, con muy ligera tendencia hacia desempleo. Además, inflación y área se encuentran muy próximos, lo cual sugiere que estas dos variables están muy relacionadas.
- Algo similar ocurre con desempleo y gasto militar, ambas poseen una correlación, mostrándonos que Grecia es el outlier en gasto en defensa, y además posee una alta tasa de desempleo, aunque el país con mayor tasa de desempleo es Croacia, el más alineado con esta variable (y probablemente el outlier en esta categoría el boxplot).
- También podemos observar patrones opuestos: a mayor inflación, menor esperanza de vida, y menor crecimiento de la población. Y grupos intermedios, como Austria e Islandia, que poseen un GDP mayor al de gran parte de los países analizados, pero también una inflación muy superior a la de países con GDP similar como Alemania y Bélgica.