

## Lab ACP

Edwin Sanchez    Stephanie Tamayo    Andres Felipe Torres    Fredy  
Urrea    Sergio Velasquez    Manuel Espitia

2025-05-18

# Introducción

## Carga de los datos

Organización de los datos y renombrar variables

```
CIUDADES<-readxl::read_excel("ciudades original-filtrado-con etiquetas (1).xlsx",  
                             sheet = 1)  
RHINT<-CIUDADES[,c("CIUDADES", "RH_1", "RH_2", "RH_5", "RH_6", "RH_7",  
                  "RH_8", "RH_9", "RH_10", "RH_11", "RH_12", "RH_13",  
                  "RH_14", "RH_15", "RH_16", "INT_66", "INT_67",  
                  "INT_68", "INT_69", "INT_70")]  
colnames(RHINT)[2:20] <- c("PC", "TCP", "AA", "CBPS", "CBES", "RAP", "CC",  
                          "CPT", "CS", "MI", "DCFS", "VI", "H", "S", "BCPIB",  
                          "CI", "TOE", "DP", "DM")
```

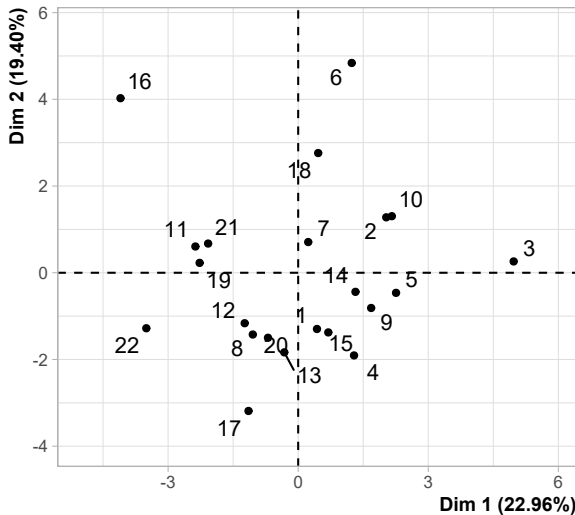
# Punto 1

## Estandarización de datos

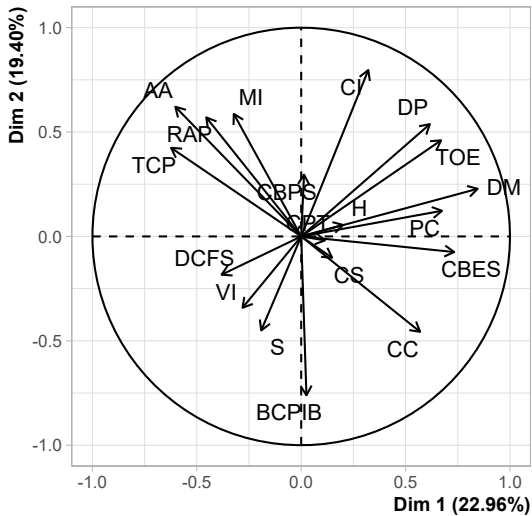
```
RHINTce <- scale(RHINT[,2:20], scale = TRUE, center = TRUE)
colnames(RHINTce) <- colnames(RHINT)[2:20]
rownames(RHINTce) <- RHINT$CIUDADES
```

Hacemos el PCA de nuestros datos con todas las variables que nos toca en nuestro grupo 6 (RH+INT) y observamos nuestros datos con el círculo unitario con las dos componentes que capturan mayor varianza de nuestros datos.

PCA graph of individuals



PCA graph of variables



## Selección de Componentes Principales:

Aplicamos el criterio de Kaiser (autovalor  $> 1$ ), reteniendo solo los componentes más informativos.

- 6 componentes cumplieron el criterio, explicando  $\sim 80\%$  de la varianza total.
- El componente 7 (6.2% adicional) podría omitirse para simplificar el modelo sin perder información crítica.

Varianza explicada:

- Componente 1: 23.0% (autovalor: 4.36)
- Componente 2: 19.4% (autovalor: 3.69)
- Componente 3: 12.4% (autovalor: 2.36)

Total acumulado (3 primeros): 54.8%

## Tabla de varianza explicada con índice personalizado

**Table 1:** Tabla 1. Varianza explicada por los componentes principales

|        | Autovalor | Varianza (%) | Varianza Acumulada (%) |
|--------|-----------|--------------|------------------------|
| comp 1 | 4.36      | 22.96        | 22.96                  |
| comp 2 | 3.69      | 19.40        | 42.36                  |
| comp 3 | 2.36      | 12.43        | 54.78                  |
| comp 4 | 2.02      | 10.65        | 65.43                  |
| comp 5 | 1.48      | 7.77         | 73.21                  |
| comp 6 | 1.27      | 6.69         | 79.90                  |
| comp 7 | 1.17      | 6.18         | 86.08                  |

*Note:*

Nota: Se muestran los primeros 7 componentes que explican el 86.1 % de la varianza total

<sup>1</sup> Criterio de selección: Autovalor > 1 (Kaiser)

<sup>2</sup> Fecha de análisis: 28/05/2025

### Componente 1 (22.96% varianza):

- Variables: Población (0.54), Tasa de Ocupación (0.45), Analfabetismo (0.49).
- Interpretación: Eje de desarrollo urbano y acceso a servicios básicos.

### Componente 2 (19.40% varianza):

- Variables: Cobertura educativa (0.98), Internacionalización (0.97).
- Interpretación: Capital humano y apertura económica.

### Componente 3 (12.43% varianza):

- Variables: Calidad educativa (0.97), Diversificación de mercados (0.94).
- Interpretación: Calidad educativa y diversificación productiva.

### Hallazgos Principales:

- Los 3 primeros componentes resumen el 54.78% de la varianza total.
- Variables educativas y económicas forman dimensiones independientes.
- Correlaciones  $>0.7$  en componentes 2 y 3 permiten simplificar el modelo sin pérdida crítica de información.



Table 2: Matriz de correlaciones entre variables y componentes principales

|             | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 |
|-------------|--------|--------|--------|--------|--------|--------|
| <b>PC</b>   | 0.67   | 0.12   | 0.45   | -0.17  | 0.39   | -0.23  |
| <b>TCP</b>  | -0.62  | 0.42   | 0.42   | -0.29  | 0.25   | 0.11   |
| <b>AA</b>   | -0.60  | 0.62   | 0.26   | -0.32  | -0.12  | 0.04   |
| <b>CBPS</b> | 0.01   | 0.30   | 0.31   | 0.69   | -0.49  | -0.03  |
| <b>CBES</b> | 0.73   | -0.07  | 0.50   | -0.03  | 0.16   | 0.00   |
| <b>RAP</b>  | -0.45  | 0.57   | 0.46   | 0.19   | -0.11  | -0.25  |
| <b>CC</b>   | 0.57   | -0.46  | 0.35   | -0.01  | 0.02   | 0.19   |
| <b>CPT</b>  | 0.11   | -0.02  | -0.52  | 0.08   | -0.15  | 0.31   |
| <b>CS</b>   | 0.15   | -0.10  | 0.39   | -0.47  | -0.29  | 0.23   |
| <b>MI</b>   | -0.32  | 0.59   | 0.16   | -0.25  | 0.27   | 0.54   |
| <b>DCFS</b> | -0.38  | -0.18  | 0.30   | 0.59   | 0.52   | 0.10   |
| <b>VI</b>   | -0.28  | -0.34  | 0.24   | 0.65   | 0.31   | 0.03   |
| <b>H</b>    | 0.20   | 0.06   | -0.08  | 0.30   | -0.03  | 0.70   |

## Dimensión 1 (X) - Eje principal

- Variables clave: MP (0.94), PC (0.94)
- Interpretación: Eje “tamaño económico-demográfico” donde ciudades más pobladas concentran mayor actividad económica.

## Dimensión 2 (Y) - Eje secundario

- Variables débiles: DM (0.22), TCP (0.12)
- Interpretación: “Dinamismo básico” menos estructurado, vinculado a crecimiento incremental.

## Hallazgo crítico:

- El desarrollo económico (X) no garantiza mejoras en indicadores sociales (Y), requiriendo políticas diferenciadas.

## Punto 2

Utilizar la función PCA del paquete FactoMiner para realizar un ACO con todas las variables específicas del grupo utilizando como ilustrativas las variables de Recursos Humanos.

### Preparación de datos

Seleccionamos las variables activas (internacionalización) y las variables

```
base_acp <- RHINT[, c("BCPIB", "CI", "TOE", "DP", "DM",  
                      "PC", "TCP", "AA", "CBPS", "CBES", "RAP", "CC",  
                      "CPT", "CS", "MI", "DCFS", "VI", "H", "S")]
```

### Análisis de Componentes Principales (ACP)

```
res.pca <- PCA(base_acp, scale.unit = TRUE, quanti.sup = 6:19, graph = FALSE)
```

## Resultados

### a) Valores propios

```
eig.val <- round(res.pca$eig, 3)
kable(eig.val, format = "latex", booktabs = TRUE,
      caption = "Valores propios y varianza explicada") %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

**Table 3:** Valores propios y varianza explicada

|        | eigenvalue | percentage of variance | cumulative percentage of variance |
|--------|------------|------------------------|-----------------------------------|
| comp 1 | 3.110      | 62.199                 | 62.199                            |
| comp 2 | 1.162      | 23.238                 | 85.437                            |
| comp 3 | 0.476      | 9.520                  | 94.957                            |
| comp 4 | 0.252      | 5.043                  | 100.000                           |
| comp 5 | 0.000      | 0.000                  | 100.000                           |

## b) Correlaciones de variables activas

```
kable(round(res.pca$var$cor, 3), format = "latex",  
      booktabs = TRUE,  
      caption = "Correlación de variables activas con los factores") %>%  
kable_styling(latex_options = c("striped", "hold_position"))
```

**Table 4:** Correlación de variables activas con los factores

|       | Dim.1  | Dim.2  | Dim.3  | Dim.4  | Dim.5 |
|-------|--------|--------|--------|--------|-------|
| BCPIB | -0.699 | 0.672  | 0.218  | 0.113  | 0     |
| CI    | 0.924  | -0.352 | 0.104  | -0.103 | 0     |
| TOE   | 0.802  | 0.285  | 0.523  | -0.037 | 0     |
| DP    | 0.876  | 0.178  | -0.192 | 0.404  | 0     |
| DM    | 0.596  | 0.688  | -0.327 | -0.253 | 0     |

### c) Correlaciones de variables ilustrativas

```
kable(round(res.pca$quant1.sup$cor, 3), format = "latex",  
      booktabs = TRUE,  
      caption = "Correlación de variables ilustrativas con los factores") %>%  
kable_styling(latex_options = c("striped", "hold_position"))
```

**Table 5:** Correlación de variables ilustrativas con los factores

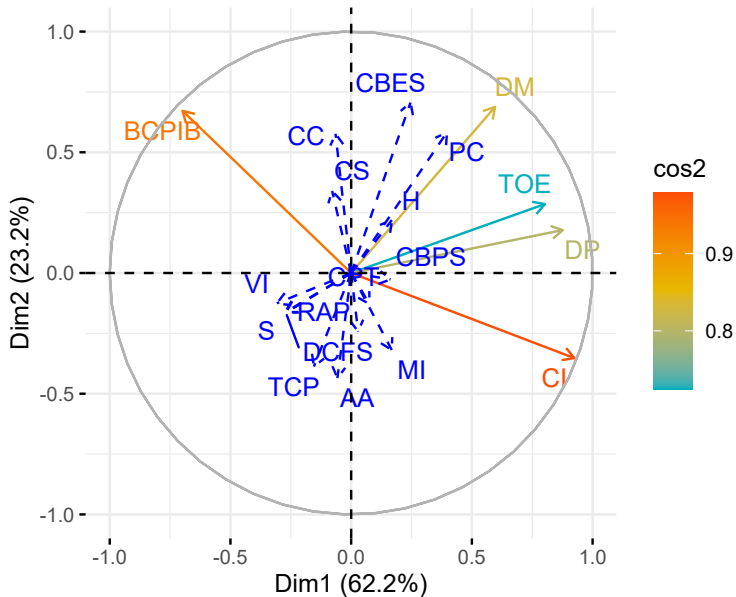
|      | Dim.1  | Dim.2  | Dim.3  | Dim.4  | Dim.5 |
|------|--------|--------|--------|--------|-------|
| PC   | 0.397  | 0.583  | 0.062  | -0.436 | 0     |
| TCP  | -0.154 | -0.387 | -0.040 | -0.305 | 0     |
| AA   | -0.057 | -0.435 | -0.067 | -0.231 | 0     |
| CBPS | 0.163  | -0.026 | -0.277 | 0.112  | 0     |
| CBES | 0.246  | 0.706  | -0.038 | -0.462 | 0     |
| RAP  | 0.029  | -0.242 | -0.200 | -0.066 | 0     |
| CC   | -0.065 | 0.577  | -0.155 | -0.334 | 0     |
| CPT  | 0.076  | -0.128 | -0.164 | -0.040 | 0     |
| CS   | -0.079 | 0.340  | -0.066 | 0.086  | 0     |
| MI   | 0.168  | -0.319 | 0.163  | -0.313 | 0     |
| DCFS | -0.270 | -0.165 | 0.103  | -0.038 | 0     |
| VI   | -0.308 | -0.125 | -0.017 | -0.108 | 0     |
| H    | 0.169  | 0.219  | 0.094  | 0.094  | 0     |
| S    | -0.273 | -0.162 | 0.137  | 0.127  | 0     |

## Visualizaciones

### a) Mapa de variables

```
fviz_pca_var(res.pca, col.var = "cos2",  
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
             repel = TRUE, title = "Variables activas e ilustrativas")
```

## Variables activas e ilustrativas

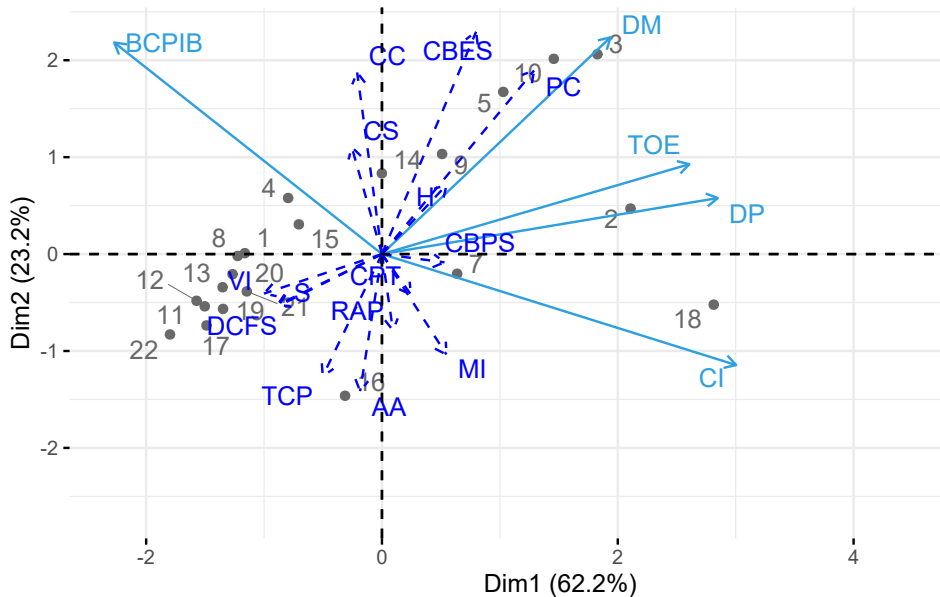




## b) Biplot

```
fviz_pca_biplot(res.pca, repel = TRUE,  
  col.var = "#2E9FDF", col.ind = "#696969",  
  title = "Biplot: ciudades y variables")
```

## Biplot: ciudades y variables



## Punto 3.

Utilizar la funcion PCA del paquete FactoMiner para realizar un ACP de las variables de recursos humanos (RH) utilizando como ilustrativas las que le correspondieron al grupo

### Carga de archivos y ejecucion del ACP

```
# Estandrizacion y asignacion de nombres de la filas
```

```
RHINTce <- scale(RHINT[,2:20], scale = TRUE, center = TRUE)
```

```
colnames(RHINTce) <- colnames(RHINT)[2:20]
```

```
rownames(RHINTce) <- RHINT$CIUDADES
```

```
require(FactoMineR)
```

```
#ACP sin grafica
```

```
pcaRHINTce <- PCA(RHINTce, quanti.sup = c(15:19), graph = F, ncp = 6)
```

```
#pcaRHINTce
```

**Table 6:** Valores propios, porcentaje de varianza y varianza acumulada

|         | Valor propio | % Varianza | % Acumulado |
|---------|--------------|------------|-------------|
| comp 1  | 3.5715       | 25.5104    | 25.5104     |
| comp 2  | 2.2390       | 15.9932    | 41.5036     |
| comp 3  | 2.0716       | 14.7971    | 56.3007     |
| comp 4  | 1.6027       | 11.4479    | 67.7486     |
| comp 5  | 1.2583       | 8.9880     | 76.7366     |
| comp 6  | 1.1918       | 8.5128     | 85.2494     |
| comp 7  | 0.7712       | 5.5087     | 90.7581     |
| comp 8  | 0.4614       | 3.2959     | 94.0539     |
| comp 9  | 0.3276       | 2.3401     | 96.3941     |
| comp 10 | 0.2153       | 1.5376     | 97.9316     |
| comp 11 | 0.1288       | 0.9199     | 98.8515     |
| comp 12 | 0.0942       | 0.6726     | 99.5241     |
| comp 13 | 0.0538       | 0.3845     | 99.9086     |
| comp 14 | 0.0128       | 0.0914     | 100.0000    |

## Análisis de Varianza Explicada (ACP):

- 6 primeras componentes (valores propios  $>1$ ) explican 85.25% de varianza
- 1ra componente: 25.51% varianza (mayor influencia)
- 2 componentes (41.5% varianza) permiten análisis bidimensional básico
- 90% varianza requiere 8 componentes (complejidad multidimensional)
- Conclusión: Reducción dimensional efectiva, pero se requieren múltiples componentes para capturar la complejidad de los datos.

**Table 7:** Correlaciones variables factor primeras 6 dimensiones

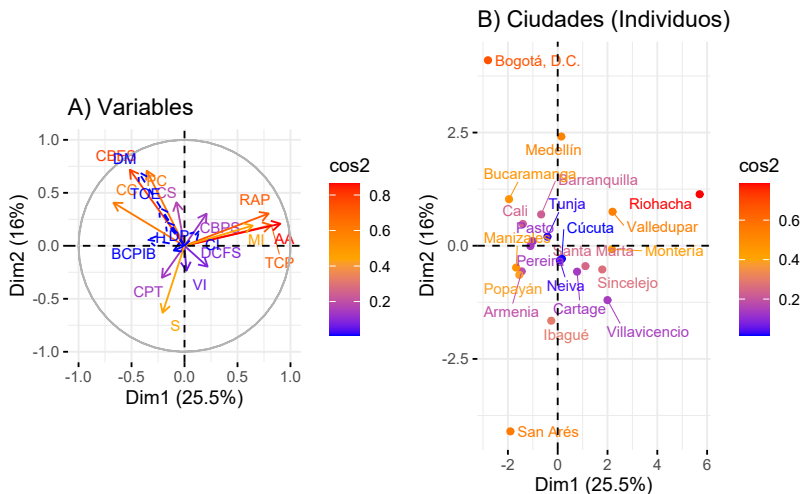
|      | Dim.1   | Dim.2   | Dim.3   | Dim.4   | Dim.5   | Dim.6   |
|------|---------|---------|---------|---------|---------|---------|
| PC   | -0.3569 | 0.7088  | -0.0370 | 0.2907  | 0.0014  | 0.4134  |
| TCP  | 0.8193  | 0.1945  | -0.1064 | 0.3789  | 0.1926  | 0.0511  |
| AA   | 0.9042  | 0.2083  | -0.2736 | -0.0455 | 0.0122  | 0.0645  |
| CBPS | 0.2085  | 0.3028  | 0.6300  | -0.5551 | -0.1747 | -0.1542 |
| CBES | -0.5160 | 0.7179  | 0.1057  | 0.0908  | 0.1345  | 0.2576  |
| RAP  | 0.7936  | 0.3049  | 0.2618  | -0.0864 | -0.2436 | 0.1866  |
| CC   | -0.6709 | 0.4063  | 0.1403  | 0.1353  | 0.2379  | -0.0390 |
| CPT  | -0.2128 | -0.3014 | -0.1846 | -0.5536 | 0.3631  | 0.4263  |
| CS   | -0.0764 | 0.4103  | -0.2834 | 0.2444  | 0.0393  | -0.7590 |
| MI   | 0.6433  | 0.1904  | -0.2198 | 0.0607  | 0.6197  | 0.0650  |
| DCFS | 0.2180  | -0.1962 | 0.7679  | 0.4088  | 0.2387  | 0.1102  |
| VI   | 0.0181  | -0.2347 | 0.8007  | 0.3024  | 0.0763  | -0.0669 |
| H    | -0.1010 | -0.0445 | 0.1945  | -0.3387 | 0.6877  | -0.3074 |
| S    | -0.2064 | -0.6319 | -0.2376 | 0.5359  | 0.0577  | 0.1202  |

### Análisis de Correlaciones Principales (6 primeras dimensiones):

- D1: Analfabetismo↑ + Educación↓
- D2: Educación↑ vs Secuestros↑
- D3: Violencia género (Delitos 0.77, VIF 0.80)
- D5-D6: Salud y violencia vinculadas

## Variables y objetos graficados en distintas dimensiones

# Variables y ciudades en las Dimensiones 1 y 2

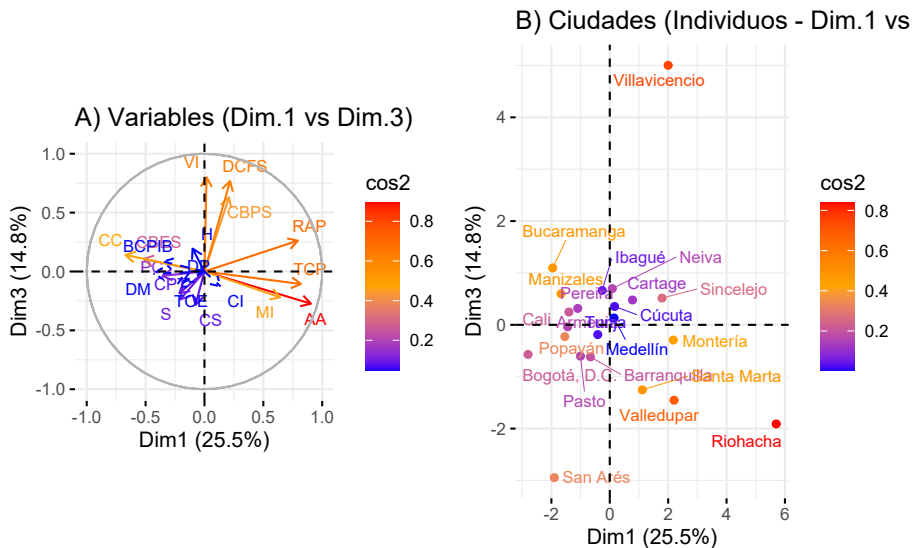


**Figure 1:** Figura 1. Variables y ciudades en las Dimensiones 1 y 2

## Análisis de Ciudades y Variables (Dimensiones 1 y 2 - 41.5% varianza total)

- Eje X (25.5%): Educación↑ vs Secuestros↑ → Bogotá/Medellín vs Cúcuta
- Eje Y (16%): Violencia↑ vs Educación↓ → Villavicencio/Santa Marta vs Armenia
- Disparidad regional: Bogotá/Medellín vs. Cúcuta/San Andrés en desarrollo y seguridad
- Relación educación-mercados: CBES y DM correlacionan en ciudades desarrolladas
- Variables suplementarias: BCPIB y DM ayudan a contextualizar los ejes principales





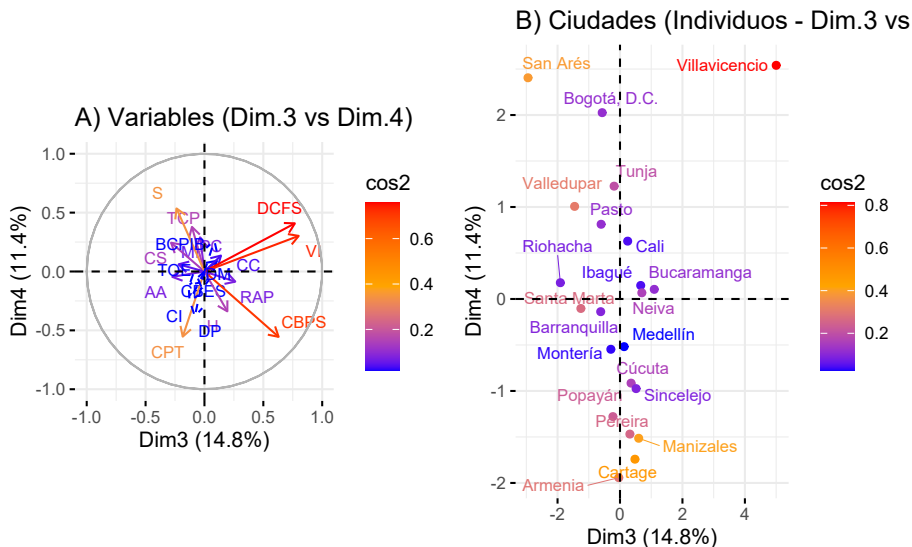
**Figure 2:** Figura 1.1 Variables y ciudades en las Dimensiones 1 y 3

## Análisis Ciudades-Variables (Dim 1 y 3 - 40.3% varianza)

- Eje X: Crecimiento (AA/TCP) vs Educación (CC/CBES)
- Eje Y: Violencia familiar (VI/DCFS) vs Seguridad (CS/S)

Ciudades con alto crecimiento muestran dos patrones:

- Con alta violencia (Villavicencio)
- Con baja violencia (Riohacha)
- Infraestructura educativa se asocia con menor violencia específica
- Variables suplementarias (DM/PC) validan vínculo entre diversificación económica y desarrollo



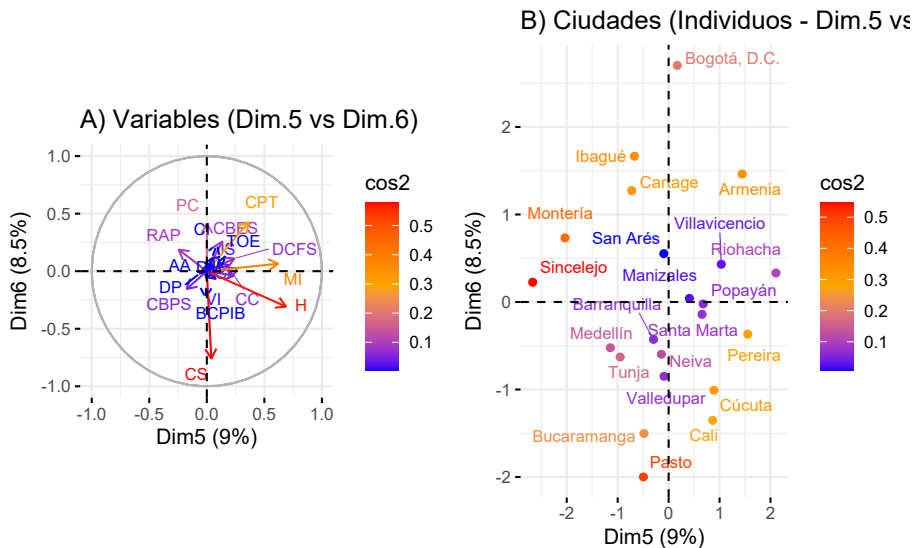
**Figure 3:** Figura 1. Variables y ciudades en las Dimensiones 3 y 4

## Análisis Ciudades-Variables (Dim 3 y 4 - 26.2% varianza)

- Eje Y (D3): Violencia (VI/DCFS) vs Educación básica (CBPS)
- Eje X (D4): Secuestros/Salud vs Homicidios/Capacitación
- Paradoja educativa: Alta cobertura escolar (CBPS) coexiste con violencia familiar/sexual (VI/DCFS)

Dos modelos de seguridad:

- Ciudades con violencia doméstica (Villavicencio)
- Ciudades con secuestros pero baja violencia familiar (Bogotá)
- Formación laboral crítica: Baja capacitación (CPT) vinculada a homicidios (H) en ciudades como Cúcuta



**Figure 4:** Figura 4. Variables y ciudades en las Dimensiones 5 y 6

## Análisis Dimensiones 5 y 6 (17.5% varianza total)

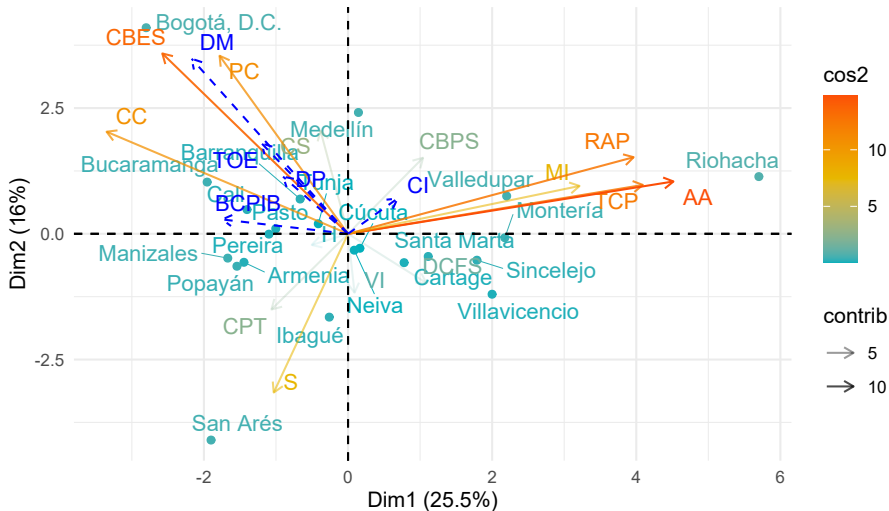
- Paradoja capitalina: Bogotá combina altos índices violentos con fuerte capacitación laboral

Dos modelos exitosos:

- Ciudades con servicios consolidados (Pasto/Medellín)
- Ciudades con apuesta educativa (Montería/San Andrés)
- Foco crítico: Cúcuta/Cali requieren urgentes mejoras en cobertura sanitaria
- Dim5: Salud vs Violencia (H/MI)
- Dim6: Formación laboral + Vulnerabilidades
- Bogotá: Violencia + Capacitación
- Cúcuta/Cali: Urgen mejoras en salud
- Pasto/Medellín: Modelo a replicar

## Biplot en las Dimensiones 1 y 2

Biplot ACP - Dimensiones 1 y 2



## Biplot Dimensiones 1 y 2

El biplot (41.5% varianza total) revela contrastes socioeconómicos entre ciudades:

- Dimensión 1 (25.5%): Seguridad vs. Desarrollo
  - (+) Riohacha/Valledupar: altos Homicidios (H) y Mortalidad Infantil (MI)
  - (-) Bogotá/Medellín: mejor Cobertura Salud (CS) y Educación Superior (CBES)
- Dimensión 2 (16%): Excelencia educativa
  - (+) Bogotá/Medellín: destacan en Calidad Colegios (CC) y CBES
  - (-) San Andrés/Ibagué: menores indicadores educativos
- Variables suplementarias (internacionalización) correlacionan levemente con ciudades desarrolladas.



Biplot ACP - Dimensiones 1 y 3

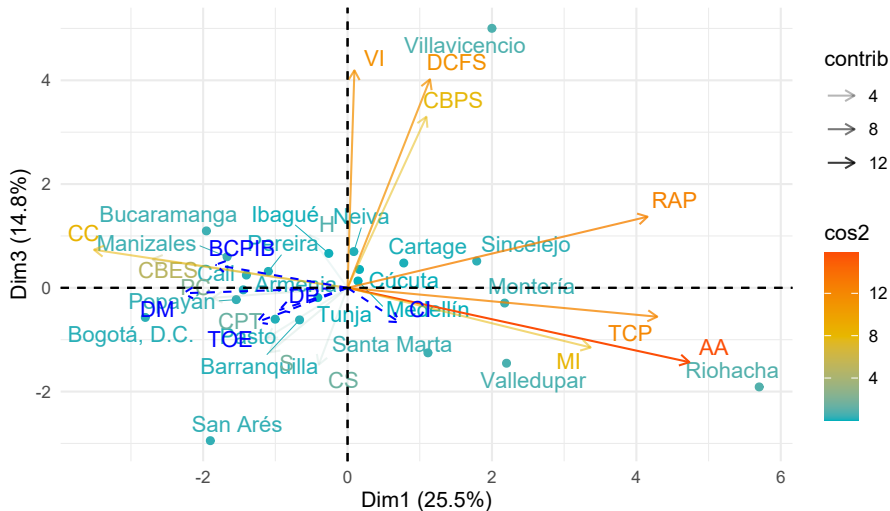
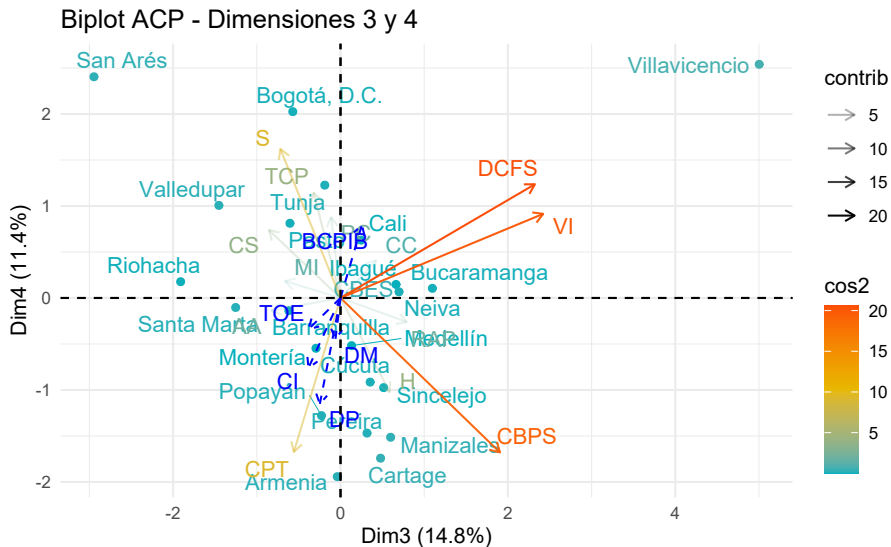


Figure 6: Figura 4.1 biplot Dimensiones 1 y 3

## Biplot Dimensiones 1 y 3

- Ejes principales (40.3% varianza total):
- Dimensión 1 (25.5%): Contrasta dinamismo demográfico (TCP, AA) con calidad educativa (CC, CBES). Ciudades como Villavicencio muestran alto crecimiento pero desafíos en alfabetización, mientras Bogotá y Pasto destacan por su infraestructura formativa.
- Dimensión 3 (14.8%): Revela una tensión entre violencia (VI, DCFS) y escolarización básica (CBPS), con ciudades como Bucaramanga combinando buena cobertura educativa con focos de violencia doméstica.
- Hallazgos clave:
  - Cuatro perfiles urbanos identificados: desde ciudades con crecimiento acelerado y violencia (Villavicencio) hasta aquellas con educación sólida y baja criminalidad (Bogotá).
  - Variables como Cobertura en Salud (CS) y Secuestros (S) requieren análisis en otras dimensiones para su plena interpretación.
  - Las variables suplementarias (ej: diversificación de mercados) matizan patrones, sugiriendo que factores económicos complementan estas dimensiones.
- Conclusión: Este plano destaca la necesidad de políticas que equilibren crecimiento poblacional, calidad educativa y seguridad, invitando a profundizar en dimensiones adicionales para un diagnóstico integral.



**Figure 7:** Figura 4. biplot Dimensiones 3 y 4

## Biplot Dimensiones 3 y 4

- Dimensión 3 (14.8%):
  - Enfoque: Violencia intrafamiliar (VI) y delitos sexuales (DCFS)
  - Ciudades críticas: Bucaramanga (alta incidencia)
  - Ciudades destacadas: Armenia, Cartago y Manizales (baja incidencia + buena cobertura educativa CBPS)
- Dimensión 4 (11.4%):
- . Variables económicas: BCPIB y TCP con influencia moderada
- Caso atípico: Villavicencio (dinámicas no explicadas por estas dimensiones)
- Conclusiones:
  - Patrones claros de violencia de género vs. desempeño educativo
  - Oportunidad para políticas focalizadas en seguridad ciudadana y equidad educativa
  - Necesidad de análisis adicional para casos atípicos

Biplot ACP - Dimensiones 5 y 6

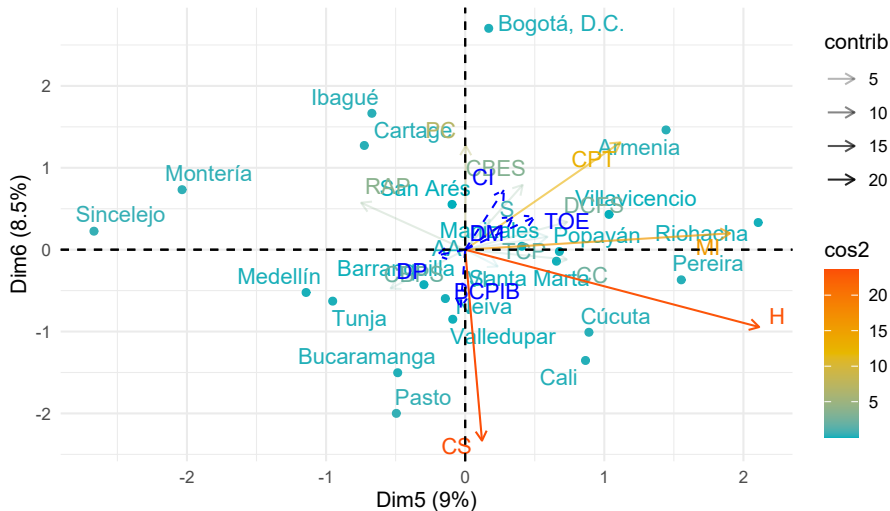


Figure 8: Figura 4. biplot Dimensiones 5 y 6

## Biplot Dimensiones 5 y 6

- Dimensión 5 (9%):
  - Variable clave: Homicidios (H)
  - Ciudades críticas: Riohacha, Pereira, Cúcuta (alta violencia letal)
- Dimensión 6 (8.5%):
  - Variable moderada: Secuestros (S) (impacto negativo)
  - Ciudades estables: Bogotá, Ibagué, Cartago (posición neutra)
- Patrón destacado:
  - Ciudades con alta incidencia de homicidios se agrupan claramente
  - Variables de bienestar (CS, CC) tienen influencia secundaria