

LABORATORIO_3

Sergio Alejandro Buitrago Melo
María Paula Camargo Rincón
Sergio Alejandro Gaona Díaz

María Alejandra Cabra Arias
Ana María Chacón Bello
Juan Esteban García Mute

2025-05-14

Laboratorio #2

Librerías y preliminares:

```
#librerías necesarias  
library(readxl)  
library(FactoMineR)  
library(ggplot2)  
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(dplyr)
```

```
##
```

```
## Adjuntando el paquete: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(tibble)
```

```
library(sqldf)
```

```
## Cargando paquete requerido: gsubfn
```

```
## Cargando paquete requerido: proto
```

```
## Cargando paquete requerido: RSQLite
```

```
#base de datos
ciudades <- read_excel("ciudades.xlsx")
```

Se realizarán los distintos ACP con las variables de RH e INFRA.

Table 1: Variables RH

Nombres	Abreviaturas RH
Población de la ciudad	RH_1
Tasa de Crecimiento de la Población	RH_2
Tasa de Ocupación	RH_3
Tasa Global de Participación	RH_4
Analfabetismo Absoluto	RH_5
Cobertura bruta en primaria y secundaria	RH_6
Cobertura bruta en educación superior	RH_7
Relación alumno-profesor	RH_8
Calidad de los colegios	RH_9
Capacitación para el trabajo	RH_10
Cobertura en salud	RH_11
Mortalidad Infantil	RH_12
Delitos contra la formación sexual	RH_13
Violencia Intrafamiliar	RH_14
Homicidios	RH_15
Secuestros	RH_16

Table 2: Variables INFRA

Nombres	Abreviaturas INFRA
Líneas telefónica por cada 10.000 habitantes	INFRA_25
Energía Eléctrica por cada 10.000 habitantes	INFRA_26
Acueducto por cada 10.000 habitantes	INFRA_27
Alcantarillado por cada 10.000 habitantes	INFRA_28
Tarifa media de energía en la industria	INFRA_29
Gas Natural por cada 10.000 habitantes	INFRA_30
Número de camas por cada 10.000 habitantes	INFRA_31
Número de salas de hospitales y clinicas por cada 10.000 habitantes	INFRA_32
Carga aérea	INFRA_33
Distancia en Kilómetros ponderada a los principales mercados (Bogotá, Medellín, Cali y Barranquilla)	INFRA_34
% de red vial pavimentada	INFRA_35
% de vías pavimentadas sobre total de la superficie	INFRA_36
Número de clientes conmutados y dedicados de internet por cada 10.000 habitantes	INFRA_37
Proveedores de servicio de Internet por cada 10.000 habitantes	INFRA_38

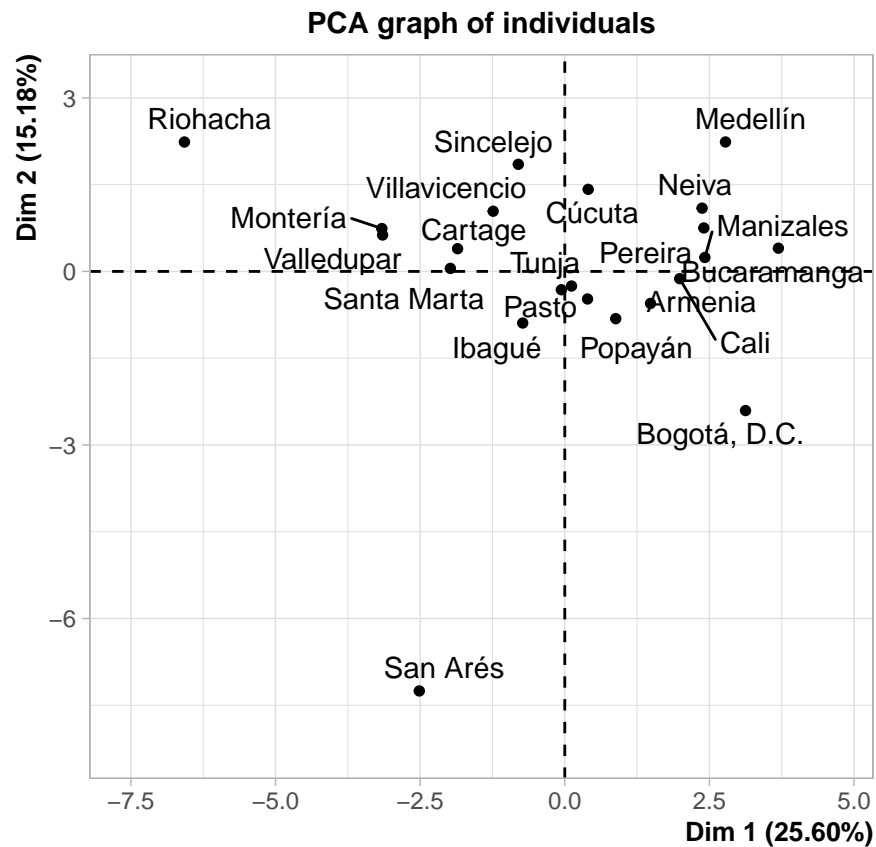
Base de datos con las variables requeridas:

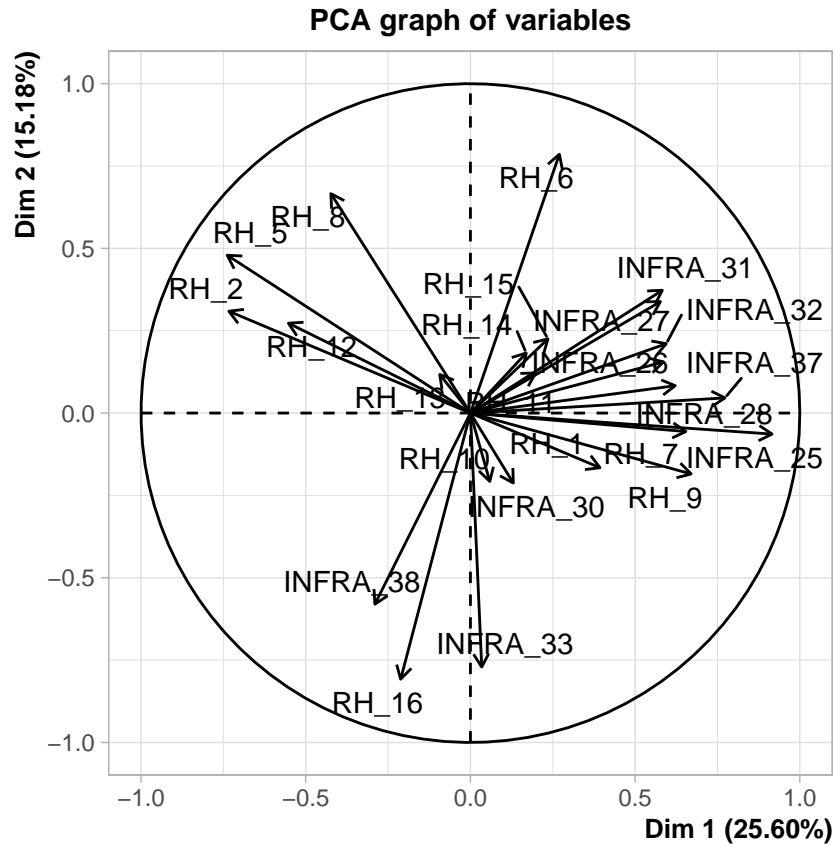
```
#base de datos RH+INFRA
ciudadest<-ciudades[,c(1:15,21:30)]
# Ahora convierte esa columna en rownames
ciudadest <- column_to_rownames(ciudadest, var = "CIUDADES")
```

Primer punto Utilizar la función PCA del paquete FactoMiner para realizar un ACP con todas las variables que le correspondieron.

```
#ACP todas las variables que le corresponden####
acp1<-PCA(ciudadest,ncp = 7)
```

```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```





Contiene los auto valores correspondientes a los componentes principales
`acp1$eig`

##	eigenvalue	percentage of variance	cumulative percentage of variance
## comp 1	6.143979350	25.59991396	25.59991
## comp 2	3.644318791	15.18466163	40.78458
## comp 3	2.950521777	12.29384074	53.07842
## comp 4	2.723578886	11.34824536	64.42666
## comp 5	1.735004053	7.22918355	71.65585
## comp 6	1.639276649	6.83031937	78.48616
## comp 7	1.134367266	4.72653028	83.21269
## comp 8	0.952165621	3.96735675	87.18005
## comp 9	0.845039806	3.52099919	90.70105
## comp 10	0.568206619	2.36752758	93.06858
## comp 11	0.496018133	2.06674222	95.13532
## comp 12	0.366954836	1.52897848	96.66430
## comp 13	0.306467492	1.27694788	97.94125
## comp 14	0.226007063	0.94169610	98.88294
## comp 15	0.100129109	0.41720462	99.30015
## comp 16	0.057199198	0.23832999	99.53848
## comp 17	0.042294765	0.17622819	99.71471
## comp 18	0.031711599	0.13213166	99.84684
## comp 19	0.021900878	0.09125366	99.93809
## comp 20	0.010023059	0.04176275	99.97985
## comp 21	0.004835049	0.02014604	100.00000

Siguiendo el criterio de los valores propios mayores a 1, se obtienen 7 componentes principales, de las cuales explican un 83% de la varianza total.

Por otro lado la suma de los valores propios de estos componente es igual al número de variables que estamos manejando.

```
#al sumar la columna de los eigenvalue da 24
sum(acp1$eig[,1])
```

```
## [1] 24
```

```
#coordenadas de las variables
acp1$var$coord
```

```
##          Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## RH_1      0.39326416 -0.16582617  0.79808707  0.234443904  0.03870311
## RH_2     -0.73318056  0.30910863  0.24834687  0.295451362 -0.13622339
## RH_5     -0.73829548  0.47923668  0.28171147 -0.101211477 -0.14542959
## RH_6      0.27009928  0.78569503 -0.14262801  0.065820983  0.19208678
## RH_7      0.65391553 -0.05609103  0.61819704  0.185758272  0.09978721
## RH_8     -0.42353084  0.66615482  0.24434105  0.168470217  0.15409973
## RH_9      0.67000755 -0.18481789  0.18727088  0.164122673  0.03335124
## RH_10     0.05867078 -0.20747753 -0.18407648 -0.559808670  0.50690189
## RH_11     0.19745370  0.12168754  0.22078089  0.004902492 -0.55512379
## RH_12    -0.55248164  0.27277774  0.30421092  0.051980537  0.06233117
## RH_13    -0.09305287  0.11713539 -0.33640969  0.821218569  0.34941540
## RH_14     0.16938965  0.18224626 -0.50837225  0.632713120  0.02554166
## RH_15     0.23534090  0.22514611 -0.15862851 -0.043439229  0.33033865
## RH_16    -0.21174190 -0.80781025 -0.24491096  0.184368443  0.01087989
## INFRA_25  0.91574462 -0.06385140  0.15085539 -0.019583543  0.18640140
## INFRA_26  0.58559310  0.15503623 -0.13338967 -0.193634112  0.45484585
## INFRA_27  0.57675110  0.33826657 -0.09625905 -0.481188041 -0.33679031
## INFRA_28  0.62174339  0.08309287 -0.14615670 -0.423228096 -0.42681853
## INFRA_30  0.13008484 -0.21238227 -0.17355059  0.489006179 -0.40760158
## INFRA_31  0.58251848  0.37273449 -0.47121912  0.247924906 -0.09799203
## INFRA_32  0.59448410  0.21046931 -0.36289112  0.340194841 -0.31102191
## INFRA_33  0.03432868 -0.77093458  0.37373834  0.338405521 -0.00279060
## INFRA_37  0.77264888  0.04596188  0.37454936  0.190432249  0.10077022
## INFRA_38 -0.28993642 -0.58012722 -0.55526987 -0.185891591 -0.03510750
##          Dim.6      Dim.7
## RH_1     -0.186970829  0.09711721
## RH_2      0.187059555  0.14579669
## RH_5      0.038259715  0.07840561
## RH_6     -0.161775467 -0.12790827
## RH_7      0.032767999  0.11098763
## RH_8     -0.281247035 -0.23116490
## RH_9      0.195707854  0.16811725
## RH_10    -0.004230834  0.39212288
## RH_11     0.636168752 -0.26133285
## RH_12     0.313790688  0.49403548
## RH_13    -0.003822078  0.05054391
## RH_14    -0.177522243  0.18840110
## RH_15     0.550233297  0.34119418
```

```
## RH_16      0.139008709 -0.13241458
## INFRA_25   0.043531483  0.03401149
## INFRA_26  -0.064865708 -0.35067018
## INFRA_27  -0.257353066  0.20863439
## INFRA_28  -0.223418106  0.22175271
## INFRA_30  -0.514782517  0.29924306
## INFRA_31   0.229991839  0.10928353
## INFRA_32   0.294938030 -0.08347409
## INFRA_33  -0.064945043  0.04349121
## INFRA_37   0.102692039 -0.14001836
## INFRA_38   0.156578722 -0.04739415
```

```
#correlaciones variable factor
acpl$var$cor
```

```
##          Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## RH_1      0.39326416 -0.16582617  0.79808707  0.234443904  0.03870311
## RH_2     -0.73318056  0.30910863  0.24834687  0.295451362 -0.13622339
## RH_5     -0.73829548  0.47923668  0.28171147 -0.101211477 -0.14542959
## RH_6      0.27009928  0.78569503 -0.14262801  0.065820983  0.19208678
## RH_7      0.65391553 -0.05609103  0.61819704  0.185758272  0.09978721
## RH_8     -0.42353084  0.66615482  0.24434105  0.168470217  0.15409973
## RH_9      0.67000755 -0.18481789  0.18727088  0.164122673  0.03335124
## RH_10     0.05867078 -0.20747753 -0.18407648 -0.559808670  0.50690189
## RH_11     0.19745370  0.12168754  0.22078089  0.004902492 -0.55512379
## RH_12    -0.55248164  0.27277774  0.30421092  0.051980537  0.06233117
## RH_13    -0.09305287  0.11713539 -0.33640969  0.821218569  0.34941540
## RH_14     0.16938965  0.18224626 -0.50837225  0.632713120  0.02554166
## RH_15     0.23534090  0.22514611 -0.15862851 -0.043439229  0.33033865
## RH_16    -0.21174190 -0.80781025 -0.24491096  0.184368443  0.01087989
## INFRA_25  0.91574462 -0.06385140  0.15085539 -0.019583543  0.18640140
## INFRA_26  0.58559310  0.15503623 -0.13338967 -0.193634112  0.45484585
## INFRA_27  0.57675110  0.33826657 -0.09625905 -0.481188041 -0.33679031
## INFRA_28  0.62174339  0.08309287 -0.14615670 -0.423228096 -0.42681853
## INFRA_30  0.13008484 -0.21238227 -0.17355059  0.489006179 -0.40760158
## INFRA_31  0.58251848  0.37273449 -0.47121912  0.247924906 -0.09799203
## INFRA_32  0.59448410  0.21046931 -0.36289112  0.340194841 -0.31102191
## INFRA_33  0.03432868 -0.77093458  0.37373834  0.338405521 -0.00279060
## INFRA_37  0.77264888  0.04596188  0.37454936  0.190432249  0.10077022
## INFRA_38 -0.28993642 -0.58012722 -0.55526987 -0.185891591 -0.03510750
##          Dim.6      Dim.7
## RH_1     -0.186970829  0.09711721
## RH_2      0.187059555  0.14579669
## RH_5      0.038259715  0.07840561
## RH_6     -0.161775467 -0.12790827
## RH_7      0.032767999  0.11098763
## RH_8     -0.281247035 -0.23116490
## RH_9      0.195707854  0.16811725
## RH_10    -0.004230834  0.39212288
## RH_11     0.636168752 -0.26133285
## RH_12     0.313790688  0.49403548
## RH_13    -0.003822078  0.05054391
## RH_14    -0.177522243  0.18840110
## RH_15     0.550233297  0.34119418
```

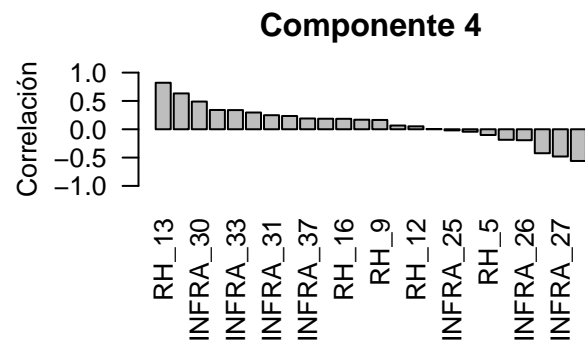
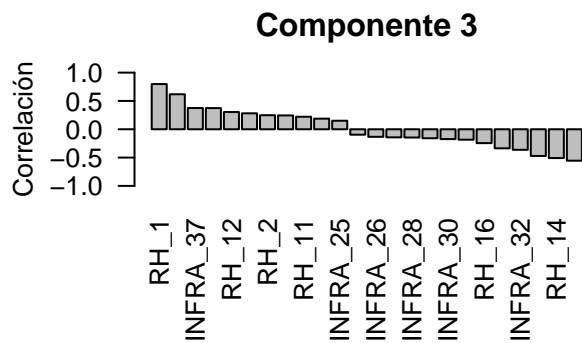
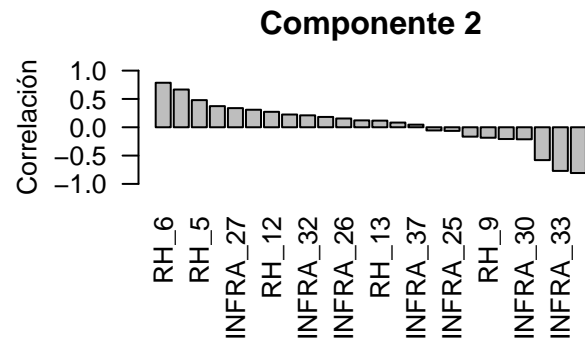
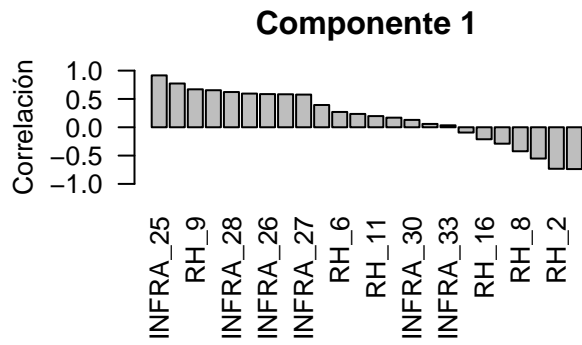
```
## RH_16      0.139008709 -0.13241458
## INFRA_25   0.043531483  0.03401149
## INFRA_26  -0.064865708 -0.35067018
## INFRA_27  -0.257353066  0.20863439
## INFRA_28  -0.223418106  0.22175271
## INFRA_30  -0.514782517  0.29924306
## INFRA_31   0.229991839  0.10928353
## INFRA_32   0.294938030 -0.08347409
## INFRA_33  -0.064945043  0.04349121
## INFRA_37   0.102692039 -0.14001836
## INFRA_38   0.156578722 -0.04739415
```

```
all.equal(acp1$var$coord, acp1$var$cor)
```

```
## [1] TRUE
```

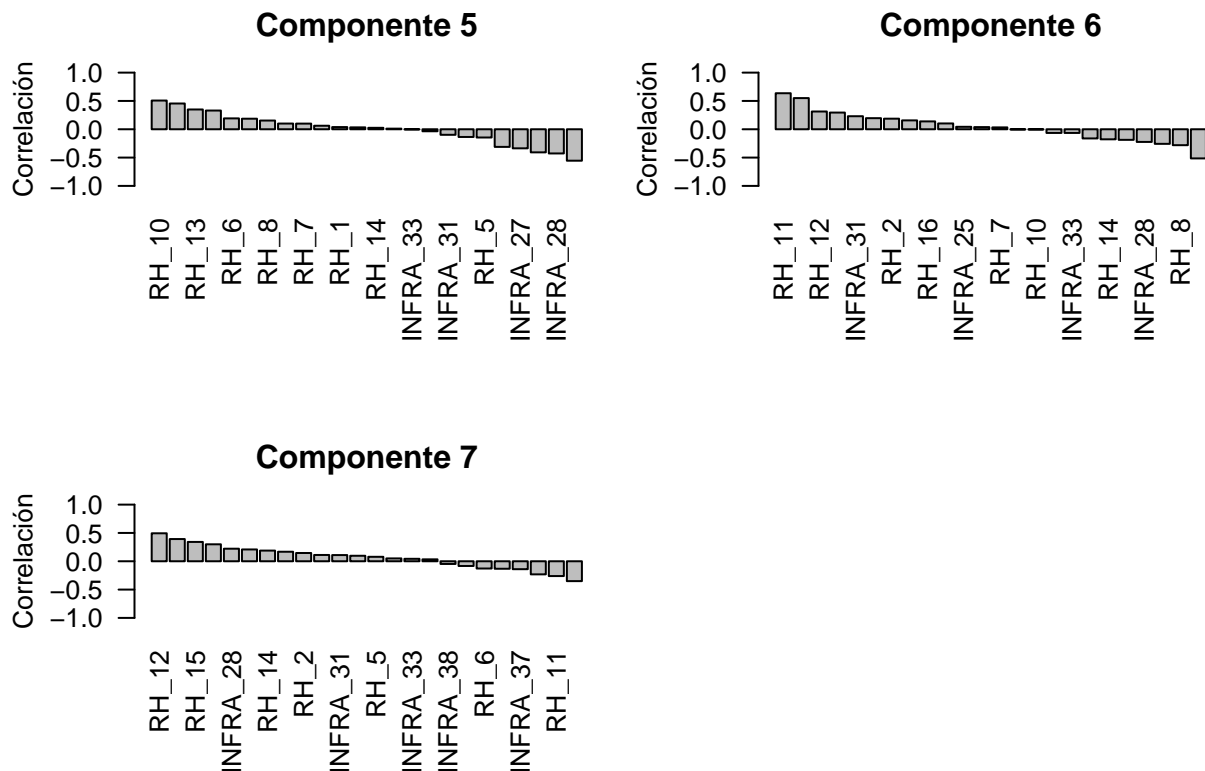
```
# Configurar gráficos para componentes 1-4 (2x2)
par(mfrow = c(2, 2), mar = c(7, 4, 3, 1)) # márgenes ajustados para etiquetas

# Graficar componentes 1-4
for (i in 1:4) {
  barplot(acp1$var$cor[order(acp1$var$cor[, i], decreasing = TRUE), i],
    las = 2, # Etiquetas verticales
    main = paste("Componente", i),
    ylab = "Correlación",
    ylim = c(-1, 1),
    col = "gray") # Todas las barras en gris
}
```



```
# Configurar gráficos para componentes 5-7 (2x2, último espacio vacío)
par(mfrow = c(2, 2), mar = c(7, 4, 3, 1))

# Graficar componentes 5-7
for (i in 5:7) {
  barplot(acp1$var$cor[order(acp1$var$cor[, i], decreasing = TRUE), i],
    las = 2,                                     # Etiquetas verticales
    main = paste("Componente", i),
    ylab = "Correlación",
    ylim = c(-1, 1),
    col = "gray")                               # Todas las barras en gris
}
```

La Dimensión 1 (Dim.1) es la que explica el mayor número de variables, mayoritariamente en sentido positivo. Entre ellas se encuentran variables de Recursos Humanos como Cobertura bruta en educación superior (RH_7), Calidad de los colegios (RH_9) y variables de infraestructura como Líneas telefónicas por cada 10.000 habitantes (INFRA_25), Energía eléctrica por cada 10.000 habitantes (INFRA_26), Acueducto (INFRA_27), Alcantarillado (INFRA_28), Número de camas hospitalarias (INFRA_31), Número de salas de hospitales y clínicas (INFRA_32) y Clientes de internet por cada 10.000 habitantes (INFRA_37). Esto sugiere que Dim.1 representa un eje de desarrollo urbano. Sin embargo, esta misma dimensión está fuertemente asociada en sentido negativo con Tasa de crecimiento de la población (RH_2), Analfabetismo absoluto (RH_5) y Cobertura en salud (RH_12), lo que sugiere que esta dimensión también distingue entre desarrollo urbano-educativo y carencias sociales estructurales.

La Dimensión 2 (Dim.2) se relaciona positivamente con variables como Cobertura en salud (RH_6) y Cobertura bruta en educación primaria y secundaria (RH_8), indicando un eje centrado en bienestar educativo y sanitario formal. En contraste, se asocia negativamente con Secuestros (RH_16) y variables de infraestructura como Carga aérea (INFRA_33) y Clientes de internet por cada 10.000 habitantes (INFRA_38), reflejando una posible oposición entre cobertura social básica y contextos de violencia o desconexión digital.

La Dimensión 3 (Dim.3) está explicada principalmente por Población de la ciudad (RH_1), lo que sugiere que capta una dimensión demográfica, y posiblemente de tamaño urbano o regional.

La Dimensión 4 (Dim.4) recoge información relevante sobre Violencia intrafamiliar (RH_14) y Delitos contra la formación sexual (RH_13) en sentido positivo, mientras que se asocia negativamente con Capacitación para el trabajo (RH_10), lo cual sugiere que esta dimensión refleja un eje de vulnerabilidad social y violencia doméstica contrapuesta a oportunidades de formación laboral.

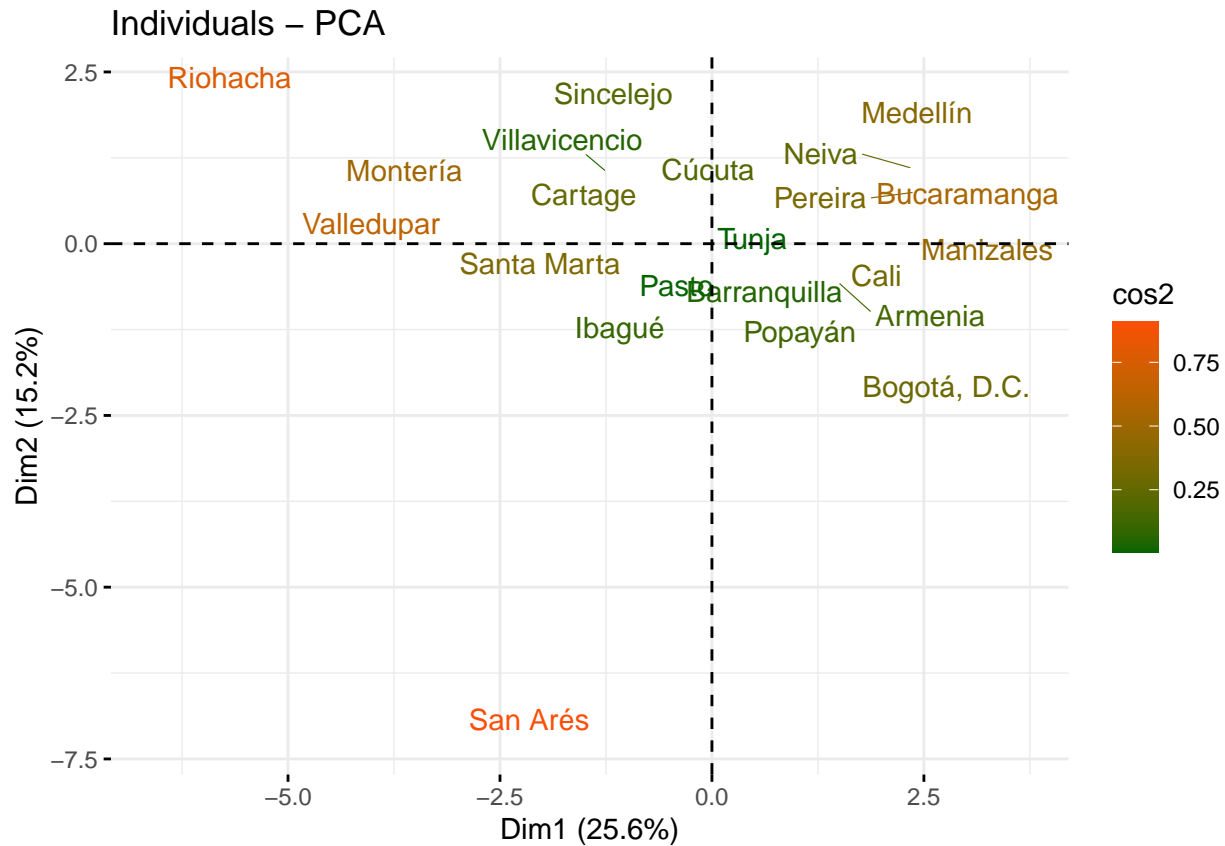
Finalmente, la Dimensión 6 (Dim.6) explica de forma positiva a Relación alumno-profesor (RH_11) y Delitos contra la formación sexual (RH_15), pero se asocia negativamente con Gas natural por cada 10.000 habitantes (INFRA_30), indicando un eje que contrasta calidad docente y riesgo social con acceso desigual a servicios.

```
#cosenos cuadrados de las variables
acp1$ind$cos2
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## Armenia	0.1206007640	0.0168217787	0.021779358	0.3019426841	0.2704638532
## Barranquilla	0.0207042054	0.0303278245	0.221422032	0.0111212344	0.1113213530
## Bogotá, D.C.	0.1794727575	0.1063446968	0.565927253	0.0635298004	0.0000340343
## Bucaramanga	0.5427631819	0.0064187423	0.003692434	0.1149305948	0.0798545780
## Cali	0.3067947356	0.0012145212	0.015892489	0.0532729664	0.0380979951
## Cartage	0.2475537126	0.0111368877	0.009736382	0.0395270649	0.1064509824
## Cúcuta	0.0143461211	0.1760205372	0.037801400	0.0934816370	0.0128602994
## Ibagué	0.0417699395	0.0628401251	0.039752977	0.0013292354	0.0652762285
## Manizales	0.4604777790	0.0045658914	0.131313338	0.0884271025	0.1732593714
## Medellín	0.2435365819	0.1580542500	0.184148380	0.0330012039	0.0338069406
## Montería	0.4781261740	0.0262715389	0.016876090	0.0143838717	0.0389391166
## Neiva	0.2030603529	0.0430952808	0.130305739	0.0058727771	0.2806353428
## Pasto	0.0003247286	0.0092764049	0.012439226	0.0030805639	0.1709345690
## Pereira	0.3206768091	0.0313302849	0.022421798	0.0252986966	0.0890436286
## Popayán	0.0724306380	0.0622295718	0.188675624	0.1526564059	0.0363637558
## Riohacha	0.6968575220	0.0806839002	0.070740650	0.0001321617	0.0012093339
## San Arés	0.0974063184	0.8102345341	0.025555710	0.0015368945	0.0016235918
## Santa Marta	0.3572210137	0.0002494462	0.006624299	0.0592938844	0.0129169731
## Sincelejo	0.0315000405	0.1674191579	0.118486166	0.0224166858	0.1445611772
## Tunja	0.0011011321	0.0051722166	0.090891738	0.0002022689	0.1522900451
## Valledupar	0.6180701531	0.0247042873	0.085100124	0.0003829994	0.1594315528
## Villavicencio	0.0291402485	0.0205999297	0.107140924	0.7545828937	0.0765873993
##	Dim.6	Dim.7			
## Armenia	2.735963e-03	8.819147e-02			
## Barranquilla	7.571316e-04	3.946223e-04			
## Bogotá, D.C.	3.386988e-02	2.193132e-02			
## Bucaramanga	1.631924e-02	9.188270e-03			
## Cali	9.401429e-02	1.374251e-02			
## Cartage	3.315980e-01	7.089978e-03			
## Cúcuta	1.150180e-01	1.290014e-02			
## Ibagué	3.851752e-01	2.506309e-02			
## Manizales	1.894157e-02	8.690660e-04			
## Medellín	5.720053e-03	2.106959e-01			
## Montería	1.089441e-01	2.305072e-01			
## Neiva	8.523734e-02	9.877403e-02			
## Pasto	1.339590e-01	7.011103e-02			
## Pereira	2.533450e-01	2.850348e-03			
## Popayán	7.200734e-03	6.817677e-02			
## Riohacha	6.193703e-02	4.796880e-02			
## San Arés	9.919473e-03	1.264737e-02			
## Santa Marta	4.206732e-05	8.464990e-03			
## Sincelejo	2.568069e-01	2.861832e-02			
## Tunja	3.616425e-02	7.332659e-02			
## Valledupar	2.481085e-02	9.329342e-04			
## Villavicencio	2.363053e-04	4.819520e-06			

Gràfics primer punto

```
# Gráfico de individuos
G11<-fviz_pca_ind(acp1, geom = 'text',
  col.ind = "cos2", # calidad de representación
  gradient.cols = c("darkgreen", "#FC4E07"),
  repel = TRUE
)
G11
```



En el gráfico podemos observar que destacan 4 ciudades, San Andrés, Riohacha, Valledupar y Bucaramanga, lo que da a entender que la mayor parte de variabilidad de esas ciudades esta siendo explicada por las componentes 1 y 2 (Dim1 y Dim2), además también se puede ver por cuál de las dimensiones parecen ser mejor representadas, por ejemplo, Valledupar y Bucaramanga parecen estar mejor representados por la componente 1 que San Andrés y Riohacha, los cuales parecen estar mejor representados por la componente 2

```
cos2_ind_1 <- as.data.frame(acp1$ind$cos2)
cos2_ind_1_dim1 <- cos2_ind_1 %>% select(Dim.1,Dim.2) %>% mutate(cos2_total = rowSums(across(everything)))
arrange(desc(cos2_total))
```

A continuación se podrán ver los valores de los primeros 4 \cos^2 totales más altos:

Ciudad	\cos^2 Dim.1	\cos^2 Dim.2	\cos^2 total
San Andrés	0.09	0.81	0.90
Riohacha	0.69	0.08	0.77

Ciudad	\cos^2 Dim.1	\cos^2 Dim.2	\cos^2 total
Valledupar	0.62	0.02	0.64
Bucaramanga	0.54	0.00	0.54

Como se mencionó anteriormente hay ciudades que destacan por componentes.

```
# Crear data frame con la dimensión de mayor cos2 por ciudad
tabla_max_cos2 <- cos2_ind_1 %>%
  as.data.frame() %>%
  mutate(
    ciudad = rownames(.),
    mejor_componente = colnames(cos2_ind_1)[max.col(., ties.method = "first")],
    max_cos2 = apply(., 1, max)
  ) %>%
  select(ciudad, mejor_componente, max_cos2)
tabla_max_cos2
```

```
##               ciudad mejor_componente  max_cos2
## Armenia          Armenia          Dim.4 0.3019427
## Barranquilla     Barranquilla     Dim.3 0.2214220
## Bogotá, D.C.     Bogotá, D.C.     Dim.3 0.5659273
## Bucaramanga      Bucaramanga      Dim.1 0.5427632
## Cali             Cali             Dim.1 0.3067947
## Cartage          Cartage          Dim.6 0.3315980
## Cúcuta           Cúcuta           Dim.2 0.1760205
## Ibagué           Ibagué           Dim.6 0.3851752
## Manizales        Manizales        Dim.1 0.4604778
## Medellín         Medellín         Dim.1 0.2435366
## Montería         Montería         Dim.1 0.4781262
## Neiva            Neiva            Dim.5 0.2806353
## Pasto            Pasto            Dim.5 0.1709346
## Pereira          Pereira          Dim.1 0.3206768
## Popayán          Popayán          Dim.3 0.1886756
## Riohacha         Riohacha         Dim.1 0.6968575
## San Arés         San Arés         Dim.2 0.8102345
## Santa Marta      Santa Marta      Dim.1 0.3572210
## Sincelejo        Sincelejo        Dim.6 0.2568069
## Tunja            Tunja            Dim.5 0.1522900
## Valledupar       Valledupar       Dim.1 0.6180702
## Villavicencio    Villavicencio    Dim.4 0.7545829
```

A continuación una tabla que dice cuantas veces la componente represento mejor al individuo, es decir la componente que tenía el \cos^2 más alto entre todas las componentes.

```
tabla_max_cos2 %>% group_by(mejor_componente) %>% summarise(conteo=n()) %>% arrange(desc(conteo))

## # A tibble: 6 x 2
##   mejor_componente conteo
##   <chr>           <int>
## 1 Dim.1             9
## 2 Dim.3             3
```

```
## 3 Dim.5          3
## 4 Dim.6          3
## 5 Dim.2          2
## 6 Dim.4          2
```

Mejor componente	Conteo
Dim.1	9
Dim.2	2
Dim.3	3
Dim.4	2
Dim.5	2
Dim.6	3

Podemos ver que la componente 1 que explica aproximadamente el 25.6% de variabilidad de la población es la misma que logra representar mejor a 9 ciudades de 22 en total.

Aunque las componentes 3,5,6 individualmente expliquen un porcentaje menor que las componentes 1 y 2, en la tabla se evidencia que hay individuos que están mejor explicados y representados en estos componentes menores. A continuación se mostrarán las ciudades que estan siendo representadas por cada componente de forma descendente es decir, que la ciudad con mayor coseno dentro de este componente aparece enlistada en el primer lugar, la que le sigue en el segundo y así sucesivamente.

```
tabla_max_cos2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.1") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 9 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>             <dbl>
## 1 Riohacha   Dim.1                 0.697
## 2 Valledupar Dim.1                 0.618
## 3 Bucaramanga Dim.1                 0.543
## 4 Montería   Dim.1                 0.478
## 5 Manizales   Dim.1                 0.460
## 6 Santa Marta Dim.1                 0.357
## 7 Pereira    Dim.1                 0.321
## 8 Cali       Dim.1                 0.307
## 9 Medellín   Dim.1                 0.244
```

```
tabla_max_cos2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.2") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 2 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>             <dbl>
## 1 San Arés    Dim.2                 0.810
## 2 Cúcuta      Dim.2                 0.176
```

```
tabla_max_cos2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.3") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 3 x 3
```

```
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>        <chr>              <dbl>
## 1 Bogotá, D.C. Dim.3              0.566
## 2 Barranquilla Dim.3              0.221
## 3 Popayán      Dim.3              0.189
```

```
tabla_max_cos2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.4") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 2 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>        <chr>              <dbl>
## 1 Villavicencio Dim.4              0.755
## 2 Armenia       Dim.4              0.302
```

```
tabla_max_cos2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.5") %>% arrange(desc(max_cos2))
```

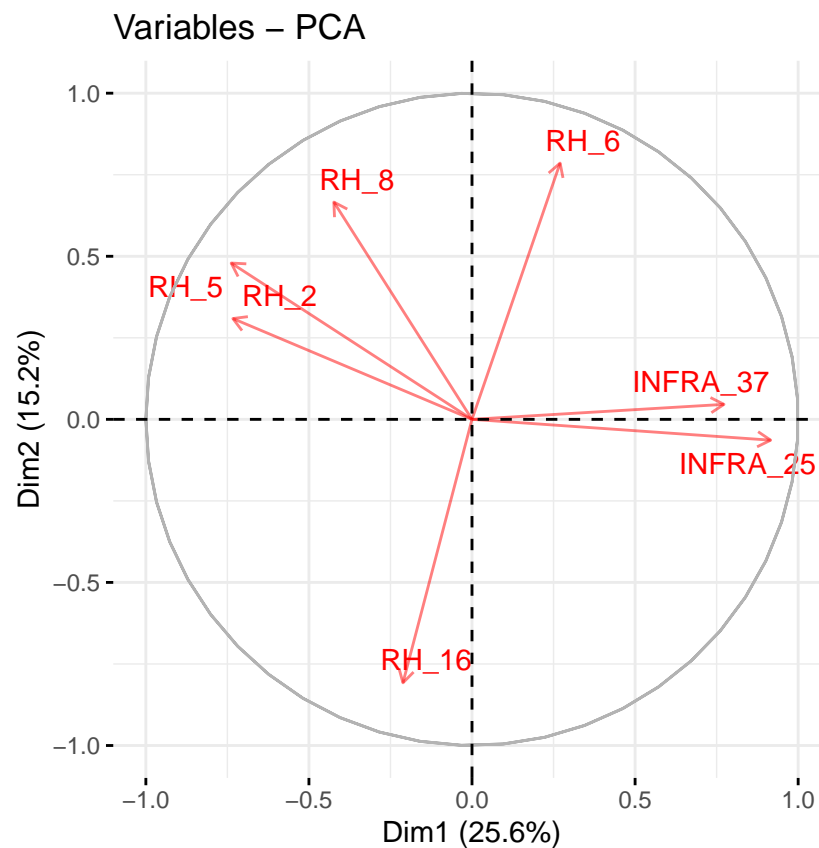
```
## # A tibble: 3 x 3
## # Groups:   mejor_componente [1]
##   ciudad mejor_componente max_cos2
##   <chr>  <chr>              <dbl>
## 1 Neiva  Dim.5              0.281
## 2 Pasto  Dim.5              0.171
## 3 Tunja  Dim.5              0.152
```

```
tabla_max_cos2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.6") %>% arrange(desc(max_cos2))
```

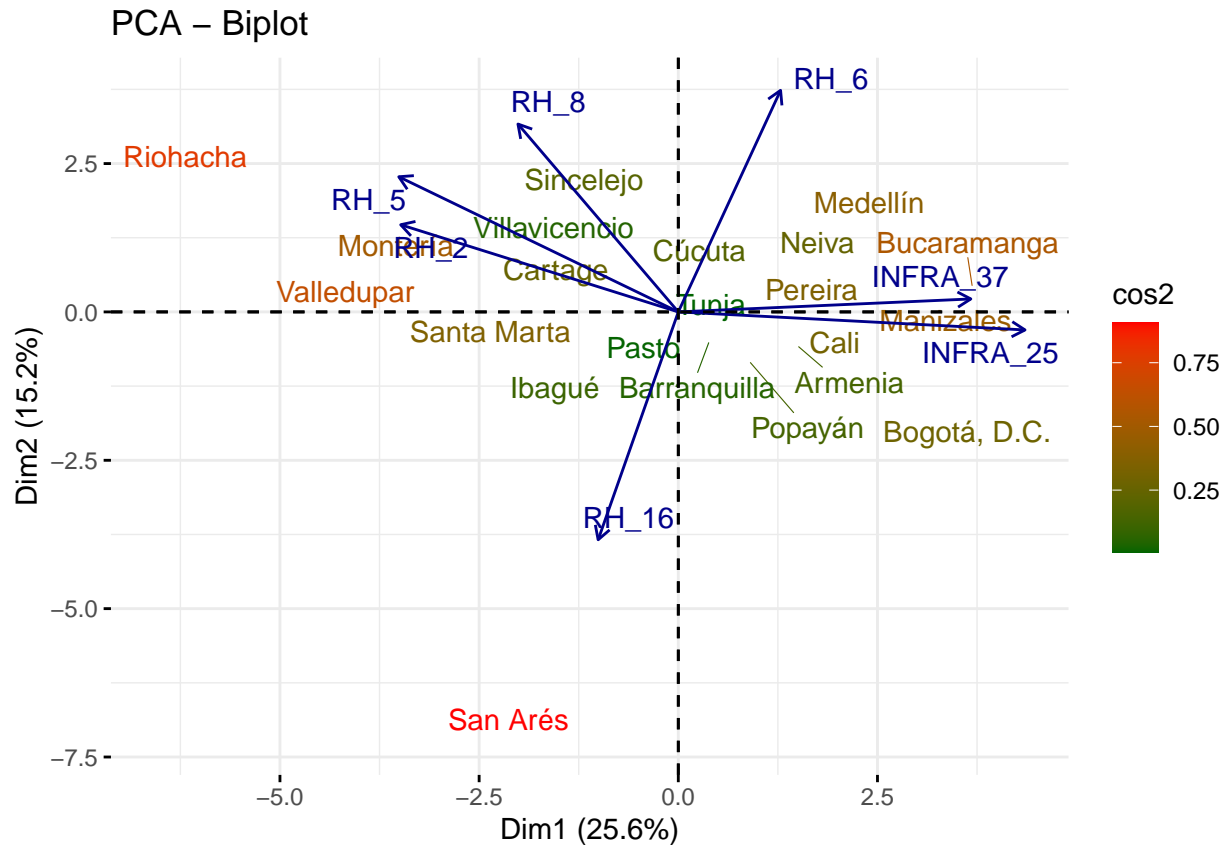
```
## # A tibble: 3 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>        <chr>              <dbl>
## 1 Ibagué       Dim.6              0.385
## 2 Cartage      Dim.6              0.332
## 3 Sincelejo    Dim.6              0.257
```

Dimensión	Ciudades
Dim.1	Riohacha, Valledupar, Bucaramanga, Montería, Manizales, Santa Marta, Pereira, Cali, Medellín
Dim.2	San Andrés, Cúcuta.
Dim.3	Bogotá, Barranquilla, Popoyán.
Dim.4	Villavicencio, Armenia.
Dim.5	Neiva, Pasto, Tunja
Dim.6	Ibagué, Cartagena, Sincelejo.

```
G12<-fviz_pca_var(acp1,
  col.var = 'red',
  alpha.var = 0.5,
  select.var = list(contrib = 7),
  repel = TRUE,
  col.quanti.sup = 'darkgrey'
)
G12
```



```
# 6. Biplot (variables + individuos)
G13<-fviz_pca_biplot(acp1,
  col.ind = "cos2",
  gradient.cols = c("darkgreen", "red"),
  col.var = "darkblue",
  select.var = list(contrib = 7),
  geom = "text",
  repel = T,
  col.quanti.sup = 'darkgrey'
)
G13
```



Al observar el plano, tres ciudades se destacan por su lejanía respecto al origen: Bogotá (3), San Andrés (17) y Riohacha (16), lo cual indica que presentan características particulares frente a las variables consideradas.

San Andrés (17) se ubica en el tercer cuadrante, con valores negativos en ambas dimensiones. Esto sugiere limitaciones importantes en infraestructura y educación (por ejemplo, baja población, baja cobertura educativa y débil acceso a internet y telefonía), explicables por su condición insular.

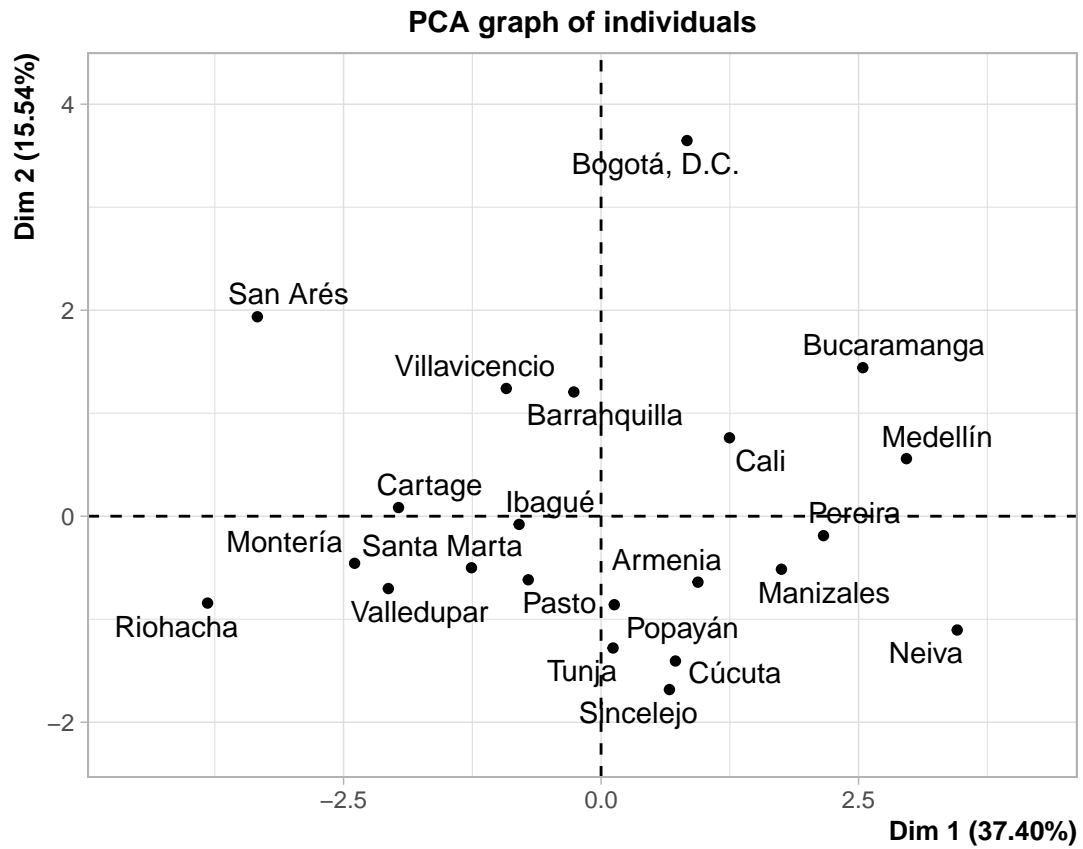
Bogotá (3) aparece en el cuarto cuadrante, asociada positivamente con variables como INFRA_25 (líneas telefónicas) e INFRA_37 (internet), lo que refleja su alta conectividad y desarrollo tecnológico, coherente con su papel como capital.

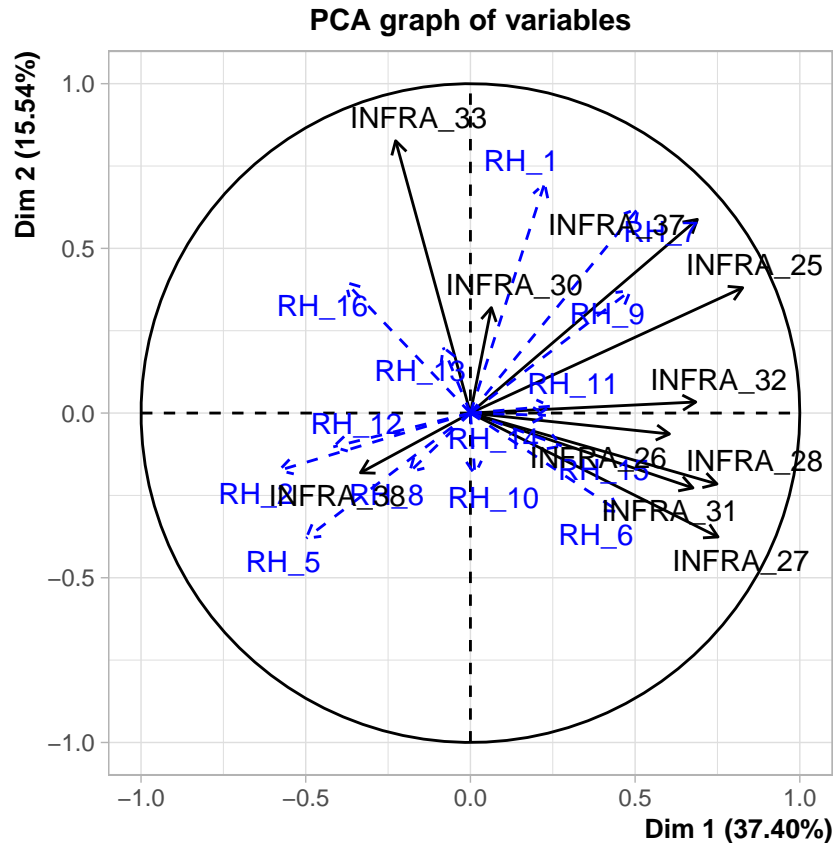
Riohacha (16) se encuentra en el segundo cuadrante, con una fuerte relación negativa respecto a las variables de infraestructura tecnológica, lo que la posiciona como una de las ciudades más rezagadas en este aspecto.

Segundo punto Utilizar la función PCA del paquete FactoMiner para realizar un ACP de las variables específicas (INFRA) utilizando como ilustrativas las variables de Recursos Humanos (RH).

Solo las variables INFRA_25 a INFRA_38 (columnas 15 a 24) serán activas. Las variables RH (1 a 14) no se usan para construir los ejes principales, pero se proyectan en el plano factorial para ser interpretadas.

```
acp2<-PCA(ciudadest, quanti.sup = c(1:14))
```



```
# Contiene los auto valores correspondientes a los componentes principales
acp2$eig
```

```
##      eigenvalue percentage of variance cumulative percentage of variance
## comp 1  3.73989557          37.3989557          37.39896
## comp 2  1.55392842          15.5392842          52.93824
## comp 3  1.44763393          14.4763393          67.41458
## comp 4  1.10618953          11.0618953          78.47647
## comp 5  1.02559925          10.2559925          88.73247
## comp 6  0.52558475           5.2558475          93.98831
## comp 7  0.28734095           2.8734095          96.86172
## comp 8  0.19940455           1.9940455          98.85577
## comp 9  0.08526143           0.8526143          99.70838
## comp 10 0.02916163           0.2916163          100.00000
```

Siguiendo el criterio de los valores propios mayores a 1, se obtienen 5 componentes principales, de las cuales explican un 88% de la varianza total.

Por otro lado la suma de los valores propios de estos componente es igual al número de variables que estamos manejando.

```
#al sumar la columna de los eigenvalue da 10
sum(acp2$eig[,1])
```

```
## [1] 10
```

```
#coordenadas de las variables
acp2$var$coord
```

```
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## INFRA_25  0.82657707  0.38050074 -0.2133359  0.08213467  0.16151727
## INFRA_26  0.60426012 -0.06343287 -0.5120243 -0.03094806  0.33777561
## INFRA_27  0.75167757 -0.37613407  0.1569072  0.48413257 -0.03703933
## INFRA_28  0.74857470 -0.21533667  0.3277062  0.48501239  0.14985208
## INFRA_30  0.06352942  0.32015548  0.7698038  0.08783533 -0.27664414
## INFRA_31  0.67559759 -0.22729549  0.2420077 -0.54809721 -0.07556691
## INFRA_32  0.68414019  0.03360157  0.3554094 -0.51667260  0.05895429
## INFRA_33 -0.22701374  0.82664856  0.2180315  0.17665952  0.29193043
## INFRA_37  0.68896104  0.58803751 -0.2401200 -0.08573810 -0.07411593
## INFRA_38 -0.33381722 -0.18124393  0.3539277 -0.12340378  0.82774211
```

```
#correlaciones variable factor
acp2$var$cor
```

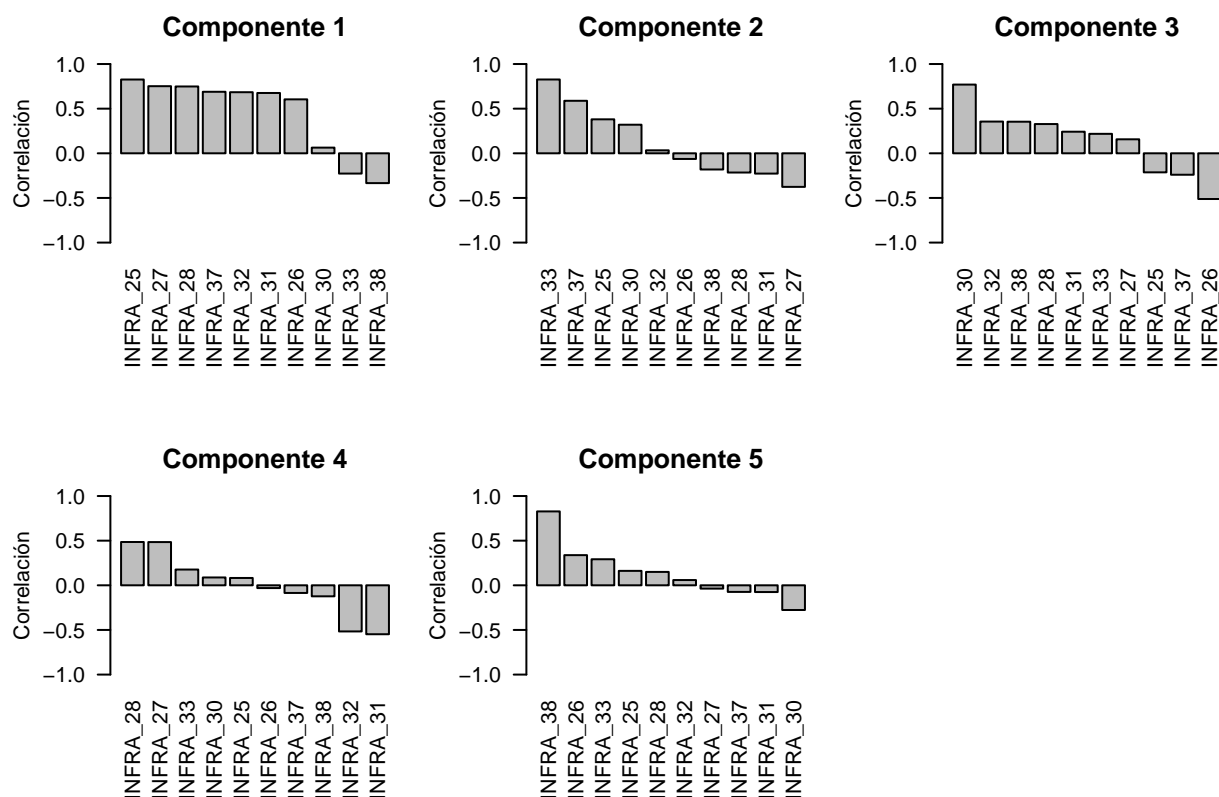
```
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## INFRA_25  0.82657707  0.38050074 -0.2133359  0.08213467  0.16151727
## INFRA_26  0.60426012 -0.06343287 -0.5120243 -0.03094806  0.33777561
## INFRA_27  0.75167757 -0.37613407  0.1569072  0.48413257 -0.03703933
## INFRA_28  0.74857470 -0.21533667  0.3277062  0.48501239  0.14985208
## INFRA_30  0.06352942  0.32015548  0.7698038  0.08783533 -0.27664414
## INFRA_31  0.67559759 -0.22729549  0.2420077 -0.54809721 -0.07556691
## INFRA_32  0.68414019  0.03360157  0.3554094 -0.51667260  0.05895429
## INFRA_33 -0.22701374  0.82664856  0.2180315  0.17665952  0.29193043
## INFRA_37  0.68896104  0.58803751 -0.2401200 -0.08573810 -0.07411593
## INFRA_38 -0.33381722 -0.18124393  0.3539277 -0.12340378  0.82774211
```

```
#aqui tambien son iguales
all.equal(acp2$var$coord,acp2$var$cor)
```

```
## [1] TRUE
```

```
# Configurar gráficos para componentes 1 a 5 en una matriz 2x3
par(mfrow = c(2, 3), mar = c(7, 4, 3, 1)) # 2 filas, 3 columnas

# Iterar sobre los 5 componentes
for (i in 1:5) {
  barplot(acp2$var$cor[order(acp2$var$cor[, i], decreasing = TRUE), i],
    las = 2,                                # Etiquetas verticales
    main = paste("Componente", i),
    ylab = "Correlación",
    ylim = c(-1, 1),
    col = "gray")                          # Barras en gris
}
```



La Dimensión 1 (Dim.1) explica fuertemente, en sentido positivo, variables relacionadas con el acceso a servicios públicos básicos y cobertura urbana. Entre estas se encuentran Líneas telefónicas por cada 10.000 habitantes (INFRA_25), Energía Eléctrica por cada 10.000 habitantes (INFRA_26), Acueducto por cada 10.000 habitantes (INFRA_27), Alcantarillado por cada 10.000 habitantes (INFRA_28), Número de camas por cada 10.000 habitantes (INFRA_31), Número de salas de hospitales y clínicas por cada 10.000 habitantes (INFRA_32) y Acceso a gas natural por cada 10.000 habitantes (INFRA_30). Estas asociaciones indican que Dim.1 refleja principalmente un eje de dotación básica de infraestructura urbana, condiciones de salubridad y conectividad.

Por su parte, la Dimensión 2 (Dim.2) está explicada positivamente por Carga aérea (INFRA_33), lo que sugiere que esta dimensión recoge información sobre infraestructura de transporte y conectividad nacional o internacional.

La Dimensión 3 (Dim.3) es definida principalmente por Acueducto por cada 10.000 habitantes (INFRA_27), y en menor medida por Clientes de internet por cada 10.000 habitantes (INFRA_38), lo que podría representar una dimensión entre infraestructura esencial y brecha digital.

Finalmente, la Dimensión 5 (Dim.5) está explicada mayoritariamente por Clientes de internet por cada 10.000 habitantes (INFRA_38), indicando que este eje representa un componente tecnológico o digital.

```
#cosenos cuadrados de las variables
acp2$var$cos2
```

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## INFRA_25	0.683229645	0.144780812	0.04551221	0.0067461047	0.026087830
## INFRA_26	0.365130298	0.004023729	0.26216891	0.0009577821	0.114092360
## INFRA_27	0.565019166	0.141476835	0.02461986	0.2343843481	0.001371912
## INFRA_28	0.560364077	0.046369880	0.10739136	0.2352370211	0.022455645

```
## INFRA_30 0.004035987 0.102499532 0.59259786 0.0077150444 0.076531981
## INFRA_31 0.456432105 0.051663240 0.05856772 0.3004105567 0.005710358
## INFRA_32 0.468047795 0.001129065 0.12631586 0.2669505712 0.003475609
## INFRA_33 0.051535239 0.683347846 0.04753774 0.0312085843 0.085223378
## INFRA_37 0.474667317 0.345788119 0.05765760 0.0073510226 0.005493171
## INFRA_38 0.111433937 0.032849363 0.12526482 0.0152284928 0.685157007
```

```
var_contrib <- data.frame(acp2$var$contrib)
```

```
top5_vars <- rownames(var_contrib)[order(var_contrib$Dim.1 + var_contrib$Dim.2, decreasing = TRUE)][1:5]
```

```
sup_vars <- rownames(acp2$quanti.sup$coord)
```

```
all_vars <- c(top5_vars, sup_vars)
```

```
s2<-acp2$eig
```

```
sum(s2[,1])
```

```
## [1] 10
```

Gráficos punto#2:

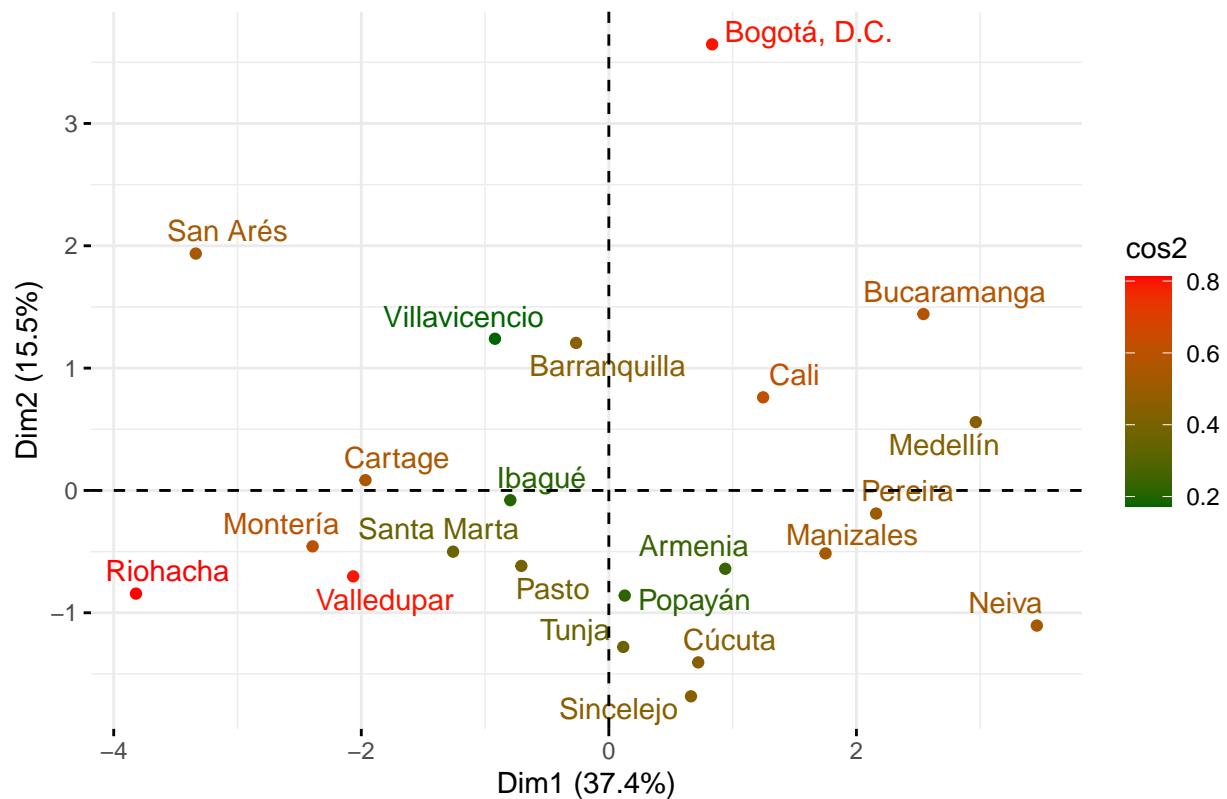
```
# Gráfico de individuos
```

```
G21<-fviz_pca_ind(acp2,
  col.ind = "cos2", # calidad de representación
  gradient.cols = c("darkgreen", "red"),
  repel = TRUE
```

```
)
```

```
G21
```

Individuals – PCA



En el gráfico podemos observar que destacan 3 ciudades, Bogotá,D.C., Riohacha y Valledupar, lo que da a entender que la mayor parte de variabilidad de esas ciudades esta siendo explicada por las componentes 1 y 2 (Dim1 y Dim2), además también se puede ver por cuál de las dimensiones parecen ser mejor representadas, por ejemplo, Valledupar y Riohacha parecen estar mejor representados por la componente 1 que San Bogotá,D.C., la cual parece estar mejor respresenta por la componente 2

```
cos2_ind_2 <- as.data.frame(acp2$ind$cos2)
cos2_ind_2_dim1 <- cos2_ind_2 %>% select(Dim.1,Dim.2) %>% mutate(cos2_total = rowSums(across(everything
  arrange(desc(cos2_total)))
```

A continuación se podrán ver los valores de los primeros 3 \cos^2 totales más altos:

Ciudad	\cos^2 Dim.1	\cos^2 Dim.2	\cos^2 total
Riohacha	0.77	0.04	0.81
Bogotá, D.C.	0.04	0.75	0.79
Valledupar	0.71	0.08	0.79

Como se mencionó anteriormente hay ciudades que destacan por componentes.

```
# Crear data frame con la dimensión de mayor cos2 por ciudad
tabla_max_cos2_2 <- cos2_ind_2 %>%
  as.data.frame() %>%
  mutate(
    ciudad = rownames(.),
    mejor_componente = colnames(cos2_ind_2)[max.col(., ties.method = "first")],
    max_cos2 = apply(., 1, max)
  ) %>%
  select(ciudad, mejor_componente, max_cos2)
tabla_max_cos2_2
```

```
##              ciudad mejor_componente  max_cos2
## Armenia          Armenia          Dim.5 0.3406184
## Barranquilla    Barranquilla          Dim.2 0.4314598
## Bogotá, D.C.    Bogotá, D.C.          Dim.2 0.7576747
## Bucaramanga     Bucaramanga          Dim.1 0.4394121
## Cali            Cali            Dim.1 0.4458738
## Cartage         Cartage          Dim.1 0.5546474
## Cúcuta          Cúcuta          Dim.2 0.3564270
## Ibagué          Ibagué          Dim.4 0.2941408
## Manizales       Manizales          Dim.1 0.4933737
## Medellín        Medellín          Dim.3 0.5211847
## Montería        Montería          Dim.1 0.5791771
## Neiva           Neiva            Dim.1 0.4834851
## Pasto           Pasto            Dim.5 0.2329005
## Pereira         Pereira          Dim.1 0.5033667
## Popayán         Popayán          Dim.2 0.2133196
## Riohacha        Riohacha          Dim.1 0.7747932
## San Arés        San Arés          Dim.1 0.4003900
## Santa Marta     Santa Marta          Dim.1 0.3111148
## Sincelejo       Sincelejo          Dim.3 0.4362759
## Tunja           Tunja            Dim.2 0.3555548
## Valledupar      Valledupar          Dim.1 0.7136357
## Villavicencio   Villavicencio      Dim.4 0.6366388
```

A continuación una tabla que dice cuantas veces la componente represento mejor al individuo, es decir la componente que tenía el \cos^2 más alto entre todas las componentes.

```
tabla_max_cos2_2 %>% group_by(mejor_componente) %>% summarise(conteo=n()) %>% arrange(desc(conteo))
```

```
## # A tibble: 5 x 2
##   mejor_componente conteo
##   <chr>           <int>
## 1 Dim.1             11
## 2 Dim.2              5
## 3 Dim.3              2
## 4 Dim.4              2
## 5 Dim.5              2
```

Mejor componente	Conteo
Dim.1	11
Dim.2	5
Dim.3	2
Dim.4	2
Dim.5	2

Podemos ver que la componente 1 que explica aproximadamente el 37.4% de variabilidad de la población es la misma que logra representar mejor a 11 ciudades de 22 en total, es decir a la mitad. Podemos ver que en los siguientes componentes va disminuyendo la cantidad de individuos que son mejor representados dentro de estos a comparación de los demás.

```
tabla_max_cos2_2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.1") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 11 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>           <dbl>
## 1 Riohacha   Dim.1             0.775
## 2 Valledupar Dim.1             0.714
## 3 Montería   Dim.1             0.579
## 4 Cartage    Dim.1             0.555
## 5 Pereira    Dim.1             0.503
## 6 Manizales  Dim.1             0.493
## 7 Neiva      Dim.1             0.483
## 8 Cali       Dim.1             0.446
## 9 Bucaramanga Dim.1          0.439
## 10 San Arés   Dim.1             0.400
## 11 Santa Marta Dim.1          0.311
```

```
tabla_max_cos2_2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.2") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 5 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>           <dbl>
```

```
## 1 Bogotá, D.C. Dim.2          0.758
## 2 Barranquilla Dim.2          0.431
## 3 Cúcuta      Dim.2          0.356
## 4 Tunja       Dim.2          0.356
## 5 Popayán     Dim.2          0.213
```

```
tabla_max_cos2_2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.3") %>% arrange(desc(
```

```
## # A tibble: 2 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>              <dbl>
## 1 Medellín  Dim.3              0.521
## 2 Sincelejo Dim.3              0.436
```

```
tabla_max_cos2_2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.4") %>% arrange(desc(
```

```
## # A tibble: 2 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>              <dbl>
## 1 Villavicencio Dim.4          0.637
## 2 Ibagué       Dim.4          0.294
```

```
tabla_max_cos2_2 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.5") %>% arrange(desc(
```

```
## # A tibble: 2 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>              <dbl>
## 1 Armenia   Dim.5              0.341
## 2 Pasto     Dim.5              0.233
```

Dimensión	Ciudades
Dim.1	Riohacha, Valledupar, Montería, Cartagena, Pereira, Manizales, Neiva, Cali, Bucaramanga, San Andrés.
Dim.2	Bogotá, Barranquilla, Cúcuta, Tunja, Popayán.
Dim.3	Medellín, Sincelejo.
Dim.4	Villavicencio, Ibagué.
Dim.5	Armenia, Pasto.

```
G22<-fviz_pca_var(acp2,
  col.var = "darkblue",
  gradient.cols = c("darkgreen", "#E7B800", "red"),
  alpha.var = 0.5,
  repel = TRUE,
```

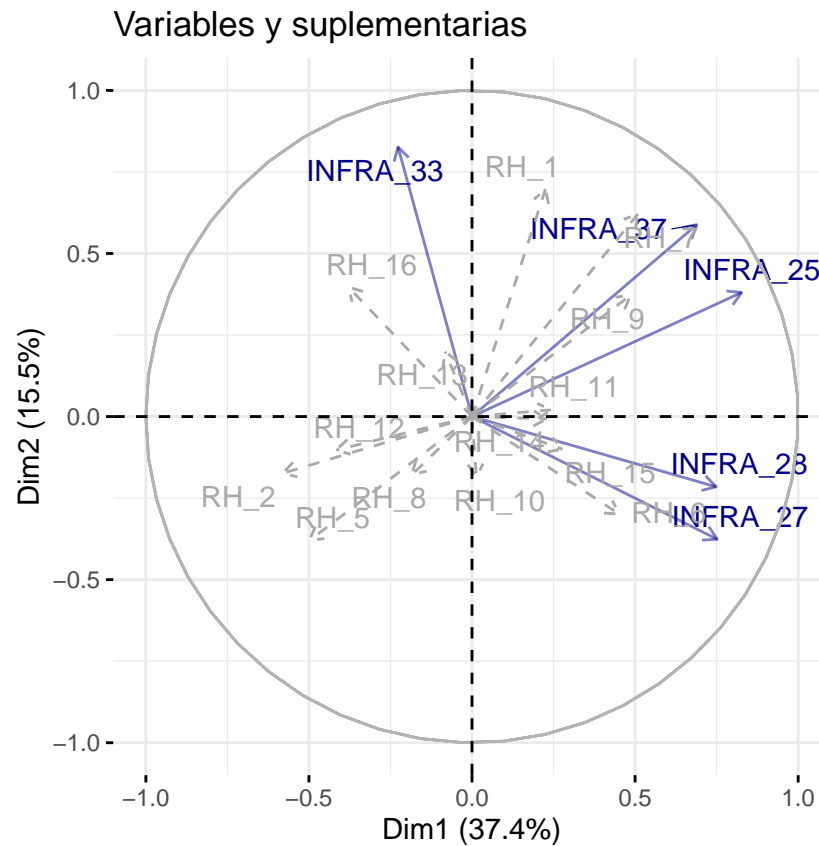


```

col.quant.sup = 'darkgrey',
select.var = list(name =c(all_vars)),
title = 'Variables y suplementarias'
)

```

G22

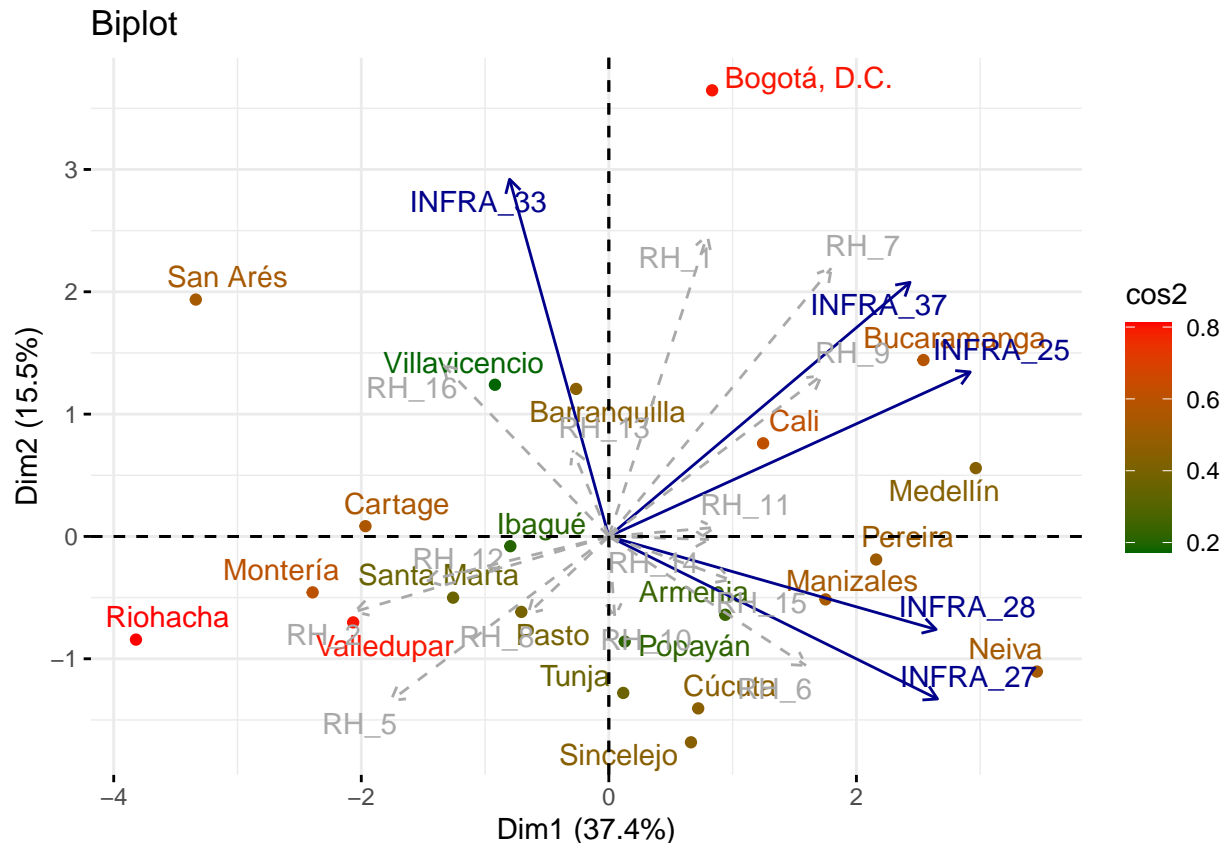


```

G23<-fviz_pca_biplot(acp2,
  col.ind = "cos2",
  gradient.cols = c("darkgreen", "red"),
  col.var = "darkblue",
  repel = T,
  col.quant.sup = 'darkgrey',
  select.var = list(name =c(all_vars)),
  title = 'Biplot'
)

```

G23



Al revisar las coordenadas 3 individuos resaltan al ojo, Bogotá (3), San Andrés (17) y Riohacha (16) por ser los individuos más alejados del origen de nuestro plano, esto nos indica singularidades y posibles aspectos a destacar de estas ciudades.

Al revisar a San Andrés vemos que este guarda una relación negativa con las variables INFR_27/28 (acueducto y alcantarillado cada 10000 habitantes) esto explicable fácilmente gracias a su baja población y su geografía de isla que dificulta la construcción de infraestructura pública como alcantarillado.

Por el lado de la capital vemos su alto posicionamiento en la dimensión 2, y a pesar que existe una relación destacable con INFR_37/25 (clientes de internet y líneas telefónicas cada 10000 habitantes), el rasgo más importante de este individuo es su alta posición en el eje Y, explicada en parte gracias a la variable de carga aérea (INFR_33) donde Bogotá es líder seguido de cerca de San Andrés.

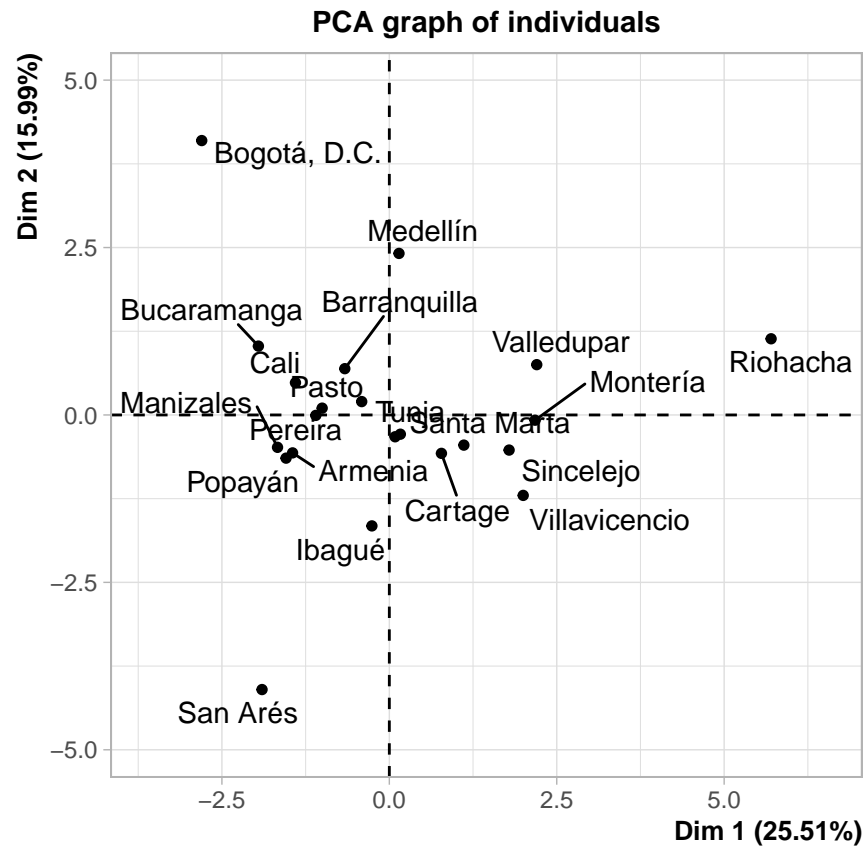
En el tercer cuadrante encontramos a Riohacha, el cual es representado de manera negativa con respecto a nuestras variables de telecomunicaciones siendo este el individuo mas destacable del lado negativo del eje X

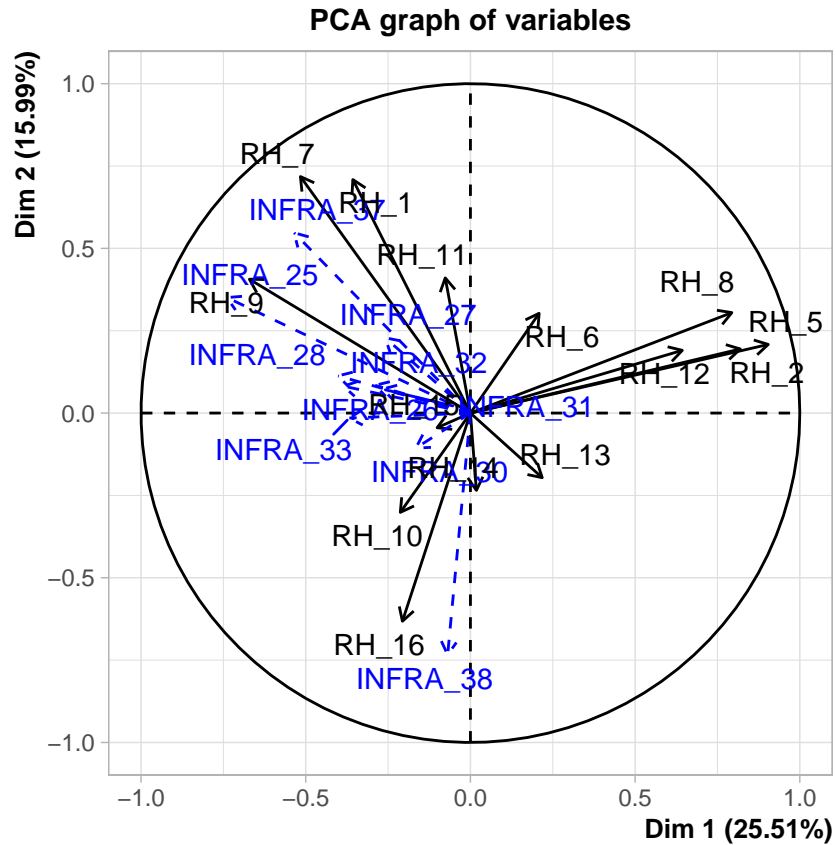
Tercer punto Utilizar la función PCA del paquete FactoMiner para realizar un ACP de las variables de Recursos Humanos (RH) utilizando como ilustrativas las que le correspondieron.

Solo las variables RH (1 a 14) serán activas. Las variables INFR_25 a INFR_38 (columnas 15 a 24) no se usan para construir los ejes principales, pero se proyectan en el plano factorial para ser interpretadas.

```
acp3<-PCA(ciudadest, quanti.sup = c(15:24), ncp = 6)
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```





```
#contiene los auto valores correspondientes a los componentes principales
acp3$eig
```

```
##      eigenvalue percentage of variance cumulative percentage of variance
## comp 1  3.57145227      25.51037333      25.51037
## comp 2  2.23904563      15.99318305      41.50356
## comp 3  2.07159426      14.79710185      56.30066
## comp 4  1.60270922      11.44792298      67.74858
## comp 5  1.25832699       8.98804991      76.73663
## comp 6  1.19179090       8.51279212      85.24942
## comp 7  0.77121137       5.50865263      90.75808
## comp 8  0.46142015       3.29585818      94.05393
## comp 9  0.32761648       2.34011768      96.39405
## comp 10 0.21525724       1.53755171      97.93160
## comp 11 0.12879153       0.91993949      98.85154
## comp 12 0.09416088       0.67257769      99.52412
## comp 13 0.05382938       0.38449557      99.90862
## comp 14 0.01279373       0.09138382     100.00000
```

Siguiendo el criterio de los valores propios mayores a 1, se obtienen 6 componentes principales, de las cuales explican un 85% de la varianza total.

Por otro lado la suma de los valores propios de estos componente es igual al número de variables que estamos manejando.

```
#al sumar la columna de los eigenvalue da 14
sum(acp3$eig[,1])
```

```
## [1] 14
```

```
#coordenadas de las variables
acp3$var$coord
```

```
##          Dim.1      Dim.2      Dim.3      Dim.4      Dim.5      Dim.6
## RH_1 -0.35693798  0.70877743 -0.03696321  0.29065674  0.001408176  0.41343738
## RH_2  0.81930164  0.19450143 -0.10637763  0.37891761  0.192633367  0.05114370
## RH_5  0.90417056  0.20834627 -0.27356525 -0.04550233  0.012177864  0.06451571
## RH_6  0.20854441  0.30278075  0.62998230 -0.55512354 -0.174660976 -0.15421996
## RH_7 -0.51603493  0.71791876  0.10571827  0.09082143  0.134489129  0.25757944
## RH_8  0.79360603  0.30489947  0.26180205 -0.08636297 -0.243592220  0.18660305
## RH_9 -0.67088638  0.40633864  0.14033448  0.13526355  0.237862034 -0.03899100
## RH_10 -0.21282189 -0.30137196 -0.18458694 -0.55364752  0.363134774  0.42626576
## RH_11 -0.07643766  0.41032125 -0.28342113  0.24444158  0.039336194 -0.75902571
## RH_12  0.64325590  0.19044110 -0.21975096  0.06071786  0.619726462  0.06498233
## RH_13  0.21798586 -0.19615992  0.76788389  0.40875032  0.238711523  0.11023355
## RH_14  0.01809994 -0.23469219  0.80072702  0.30237281  0.076254577 -0.06686807
## RH_15 -0.10102305 -0.04451304  0.19453606 -0.33874643  0.687721548 -0.30741848
## RH_16 -0.20644844 -0.63189077 -0.23757449  0.53590502  0.057669955  0.12015151
```

```
#correlaciones variable factor
acp3$var$cor
```

```
##          Dim.1      Dim.2      Dim.3      Dim.4      Dim.5      Dim.6
## RH_1 -0.35693798  0.70877743 -0.03696321  0.29065674  0.001408176  0.41343738
## RH_2  0.81930164  0.19450143 -0.10637763  0.37891761  0.192633367  0.05114370
## RH_5  0.90417056  0.20834627 -0.27356525 -0.04550233  0.012177864  0.06451571
## RH_6  0.20854441  0.30278075  0.62998230 -0.55512354 -0.174660976 -0.15421996
## RH_7 -0.51603493  0.71791876  0.10571827  0.09082143  0.134489129  0.25757944
## RH_8  0.79360603  0.30489947  0.26180205 -0.08636297 -0.243592220  0.18660305
## RH_9 -0.67088638  0.40633864  0.14033448  0.13526355  0.237862034 -0.03899100
## RH_10 -0.21282189 -0.30137196 -0.18458694 -0.55364752  0.363134774  0.42626576
## RH_11 -0.07643766  0.41032125 -0.28342113  0.24444158  0.039336194 -0.75902571
## RH_12  0.64325590  0.19044110 -0.21975096  0.06071786  0.619726462  0.06498233
## RH_13  0.21798586 -0.19615992  0.76788389  0.40875032  0.238711523  0.11023355
## RH_14  0.01809994 -0.23469219  0.80072702  0.30237281  0.076254577 -0.06686807
## RH_15 -0.10102305 -0.04451304  0.19453606 -0.33874643  0.687721548 -0.30741848
## RH_16 -0.20644844 -0.63189077 -0.23757449  0.53590502  0.057669955  0.12015151
```

```
#aqui tambien son iguales
all.equal(acp3$var$coord,acp3$var$cor)
```

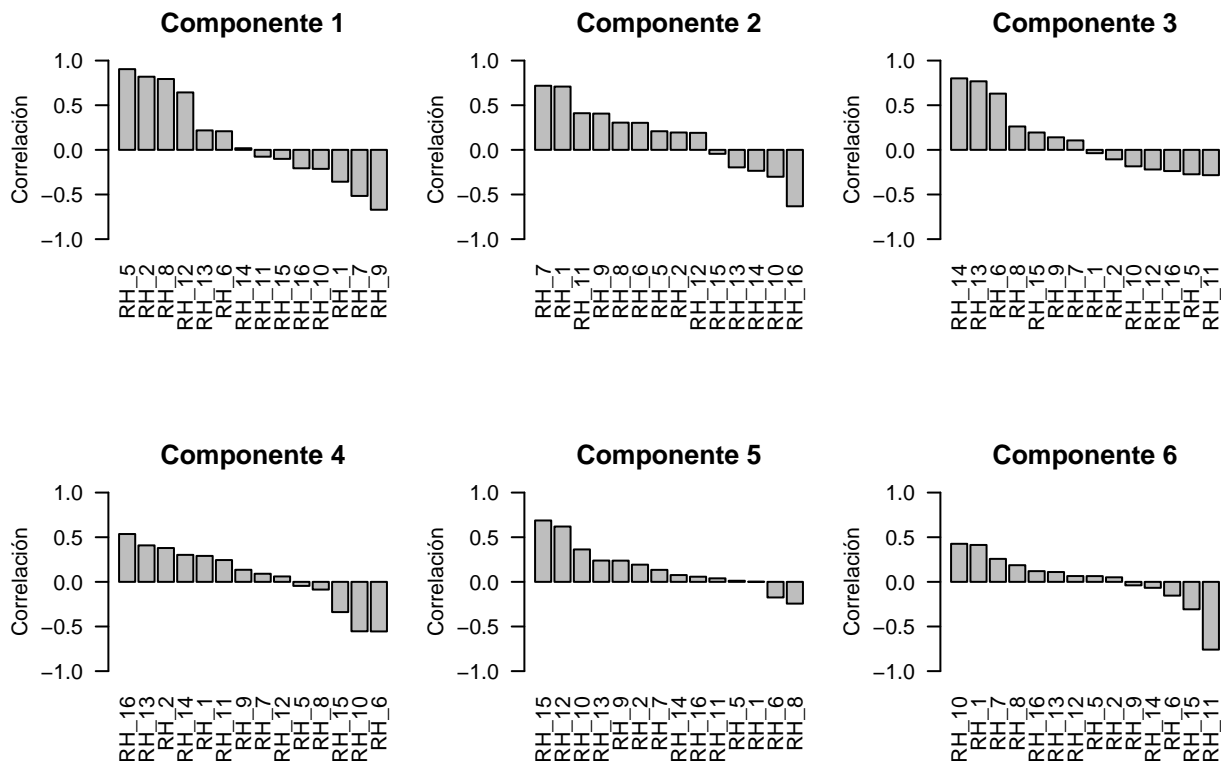
```
## [1] TRUE
```

```
# Configurar gráficos para componentes 1 a 6 (2 filas, 3 columnas)
par(mfrow = c(2, 3), mar = c(7, 4, 3, 1)) # Márgenes ajustados para etiquetas
```

```

# Iterar sobre los 6 componentes
for (i in 1:6) {
  barplot(acp3$var$cor[order(acp3$var$cor[, i], decreasing = TRUE), i],
    las = 2,                                # Etiquetas de variable verticales
    main = paste("Componente", i),
    ylab = "Correlación",
    ylim = c(-1, 1),
    col = "gray")                          # Barras en gris
}

```



La Dimensión 1 (Dim.1) explica fuertemente, en sentido positivo, a variables como Tasa de Crecimiento de la Población (RH_2), Analfabetismo Absoluto (RH_5), Cobertura bruta en educación superior (RH_7) y Capacitación para el trabajo (RH_10). Estas asociaciones positivas sugieren que Dim.1 representa un eje relacionado con indicadores de formación y desarrollo humano. Sin embargo, esta misma dimensión también presenta una fuerte correlación negativa con Cobertura bruta en primaria y secundaria (RH_4), lo que significa que Dim.1 también puede estar recogiendo contradicciones entre niveles educativos básicos y superiores, o entre cantidad y calidad educativa.

La Dimensión 2 (Dim.2) se encuentra positivamente relacionada con variables como Población de la ciudad (RH_1), Cobertura en salud (RH_6), Calidad de los colegios (RH_9) y Delitos contra la formación sexual (RH_15), lo que indica que esta dimensión podría capturar una interacción entre densidad urbana, acceso a salud, y presencia de conflictos sociales o culturales. También destaca que Homicidios (RH_14) tiene su mayor correlación con esta dimensión, pero en sentido negativo, reforzando la idea de que Dim.2 distingue entre bienestar urbano y contextos de violencia.

La Dimensión 3 (Dim.3) se ve explicada por Cobertura bruta en primaria y secundaria (RH_4), Cobertura en

salud (RH_6) y Relación alumno–profesor (RH_11), reflejando un eje centrado en aspectos de infraestructura educativa y eficiencia del sistema escolar.

Por su parte, la Dimensión 4 (Dim.4) presenta su principal correlación negativa con Calidad de los colegios (RH_8), lo que podría señalar un factor que separa calidad percibida de cobertura o acceso.

La Dimensión 5 (Dim.5) está explicada por Violencia Intrafamiliar (RH_13), representando una dimensión social crítica vinculada al entorno familiar.

```
#cosenos cuadrados de las variables
acp3$var$cos2
```

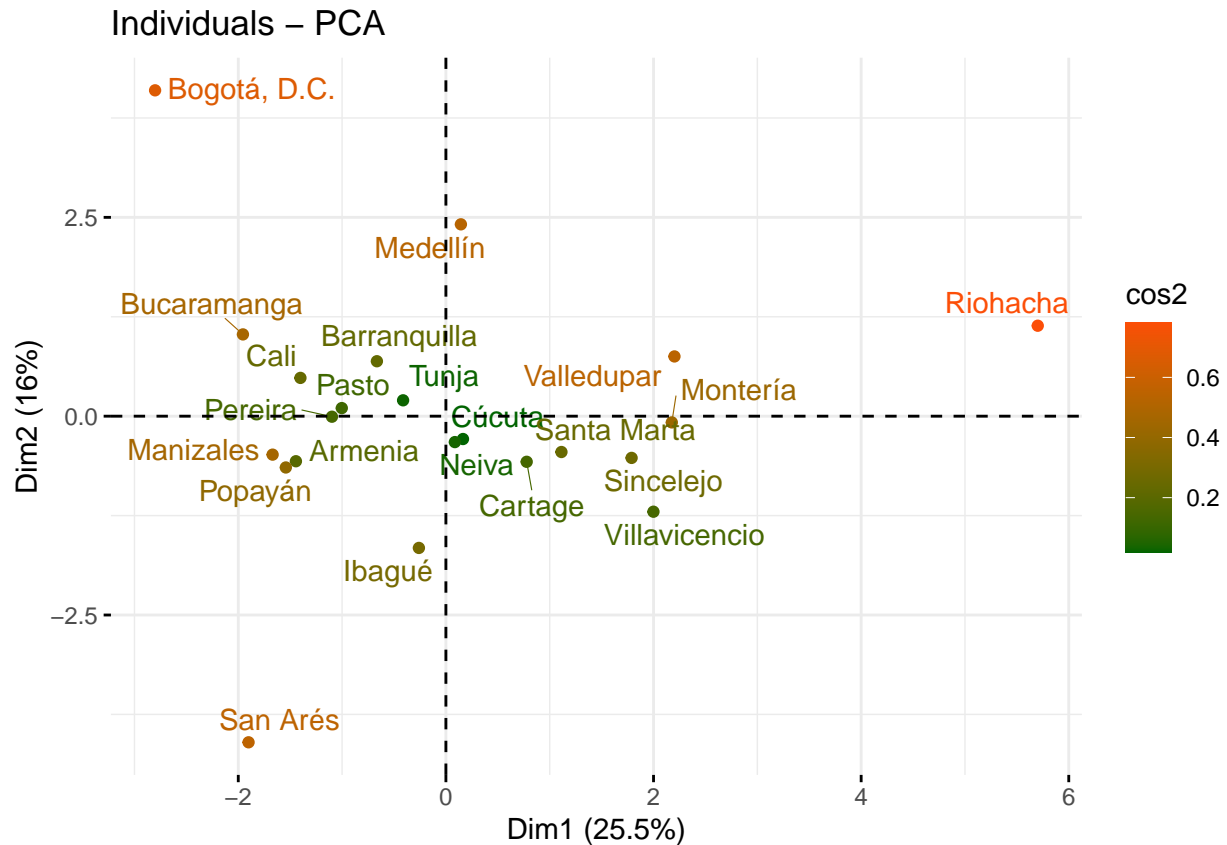
```
##          Dim.1      Dim.2      Dim.3      Dim.4      Dim.5      Dim.6
## RH_1  0.1274047208 0.50236545 0.001366279 0.084481342 1.982960e-06 0.170930470
## RH_2  0.6712551732 0.03783081 0.011316201 0.143578559 3.710761e-02 0.002615678
## RH_5  0.8175244042 0.04340817 0.074837947 0.002070462 1.483004e-04 0.004162277
## RH_6  0.0434907701 0.09167619 0.396877693 0.308162144 3.050646e-02 0.023783796
## RH_7  0.2662920502 0.51540735 0.011176352 0.008248532 1.808733e-02 0.066347168
## RH_8  0.6298105273 0.09296368 0.068540314 0.007458562 5.933717e-02 0.034820698
## RH_9  0.4500885389 0.16511109 0.019693767 0.018296227 5.657835e-02 0.001520298
## RH_10 0.0452931572 0.09082506 0.034072337 0.306525573 1.318669e-01 0.181702494
## RH_11 0.0058427156 0.16836353 0.080327536 0.059751685 1.547336e-03 0.576120031
## RH_12 0.4137781496 0.03626781 0.048290484 0.003686659 3.840609e-01 0.004222703
## RH_13 0.0475178359 0.03847871 0.589645668 0.167076823 5.698319e-02 0.012151436
## RH_14 0.0003276077 0.05508042 0.641163760 0.091429319 5.814760e-03 0.004471339
## RH_15 0.0102056570 0.00198141 0.037844280 0.114749141 4.729609e-01 0.094506122
## RH_16 0.0426209586 0.39928594 0.056441641 0.287194189 3.325824e-03 0.014436386
```

```
var_contrib3 <- data.frame(acp3$var$contrib)

top5_vars3 <- rownames(var_contrib3)[order(var_contrib3$Dim.1 + var_contrib3$Dim.2, decreasing = TRUE)]
sup_vars3 <- rownames(acp3$quanti.sup$coord)
all_vars3 <- c(top5_vars3, sup_vars3)
```

Gráficos punto #3

```
# Gráfico de individuos
G31<-fviz_pca_ind(acp3,
  col.ind = "cos2", # calidad de representación
  gradient.cols = c("darkgreen", "#FC4E07"),
  repel = TRUE
)
G31
```



En el gráfico podemos observar que destacan 4 ciudades, San Andrés, Bogotá,D.C., Riohacha y Valledupar, lo que da a entender que la mayor parte de variabilidad de esas ciudades esta siendo explicada por las componentes 1 y 2 (Dim1 y Dim2), además también se puede ver por cuál de las dimensiones parecen ser mejor representadas, por ejemplo, Valledupar y Riohacha parecen estar mejor representados por la componente 1 que San Bogotá,D.C., la cual parece estar mejor respresenta por la componente 2

```
cos2_ind_3 <- as.data.frame(acp3$ind$cos2)
cos2_ind_3_dim1 <- cos2_ind_3 %>% select(Dim.1,Dim.2) %>% mutate(cos2_total = rowSums(across(everything)))
arrange(desc(cos2_total))
```

A continuación se podrán ver los valores de los primeros 3 \cos^2 totales más altos:

Ciudad	\cos^2 Dim.1	\cos^2 Dim.2	\cos^2 total
Riohacha	0.75	0.03	0.78
Bogotá, D.C.	0.21	0.45	0.66
San Andrés	0.10	0.45	0.55
Valledupar	0.48	0.05	0.53

Como se mencionó anteriormente hay ciudades que destacan por componentes.

```
# Crear data frame con la dimensión de mayor cos2 por ciudad
tabla_max_cos2_3 <- cos2_ind_3 %>%
  as.data.frame() %>%
  mutate(
```



```

ciudad = rownames(.),
mejor_componente = colnames(cos2_ind_3)[max.col(., ties.method = "first")],
max_cos2 = apply(., 1, max)
) %>%
select(ciudad, mejor_componente, max_cos2)
tabla_max_cos2_3

```

```

##              ciudad mejor_componente  max_cos2
## Armenia          Armenia          Dim.4 0.2951413
## Barranquilla      Barranquilla      Dim.2 0.1162845
## Bogotá, D.C.      Bogotá, D.C.      Dim.2 0.4555949
## Bucaramanga       Bucaramanga       Dim.1 0.3668279
## Cali              Cali              Dim.1 0.2093450
## Cartage           Cartage           Dim.4 0.4398617
## Cúcuta            Cúcuta            Dim.6 0.1725695
## Ibagué            Ibagué            Dim.6 0.2901214
## Manizales         Manizales         Dim.1 0.4275987
## Medellín          Medellín          Dim.2 0.5261027
## Montería          Montería          Dim.1 0.4293558
## Neiva             Neiva             Dim.3 0.1613908
## Pasto             Pasto             Dim.6 0.4667174
## Pereira           Pereira           Dim.5 0.2756790
## Popayán           Popayán           Dim.1 0.3276161
## Riohacha          Riohacha          Dim.1 0.7524178
## San Arés          San Arés          Dim.2 0.4532104
## Santa Marta       Santa Marta       Dim.3 0.2679472
## Sincelejo         Sincelejo         Dim.5 0.5422884
## Tunja             Tunja             Dim.4 0.1950723
## Valledupar        Valledupar        Dim.1 0.4813210
## Villavicencio     Villavicencio     Dim.3 0.6446866

```

A continuación una tabla que dice cuantas veces la componente represento mejor al individuo, es decir la componente que tenía el \cos^2 más alto entre todas las componentes.

```

tabla_max_cos2_3 %>% group_by(mejor_componente) %>% summarise(conteo=n()) %>% arrange(desc(conteo))

```

```

## # A tibble: 6 x 2
##   mejor_componente conteo
##   <chr>           <int>
## 1 Dim.1             7
## 2 Dim.2             4
## 3 Dim.3             3
## 4 Dim.4             3
## 5 Dim.6             3
## 6 Dim.5             2

```

Mejor componente	Conteo
Dim.1	7
Dim.2	4
Dim.3	3

Mejor componente	Conteo
Dim.4	3
Dim.5	2
Dim.6	3

Podemos ver que la componente 1 que explica aproximadamente el 25.5% de variabilidad de la población es la misma que logra representar mejor a 7 ciudades de 22 en total. Podemos ver que en los siguientes componentes va disminuyendo la cantidad de individuos que son mejor representados dentro de estos a comparación de los demás, aunque podemos ver que a pesar de que el componente 6 explica menos variabilidad de la población que el componente 5, logra explicar y representar más individuos que este.

```
tabla_max_cos2_3 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.1") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 7 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>             <dbl>
## 1 Riohacha   Dim.1                 0.752
## 2 Valledupar Dim.1                 0.481
## 3 Montería   Dim.1                 0.429
## 4 Manizales  Dim.1                 0.428
## 5 Bucaramanga Dim.1                 0.367
## 6 Popayán    Dim.1                 0.328
## 7 Cali       Dim.1                 0.209
```

```
tabla_max_cos2_3 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.2") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 4 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>             <dbl>
## 1 Medellín   Dim.2                 0.526
## 2 Bogotá, D.C. Dim.2                 0.456
## 3 San Arés   Dim.2                 0.453
## 4 Barranquilla Dim.2                 0.116
```

```
tabla_max_cos2_3 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.3") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 3 x 3
## # Groups:   mejor_componente [1]
##   ciudad      mejor_componente max_cos2
##   <chr>      <chr>             <dbl>
## 1 Villavicencio Dim.3                 0.645
## 2 Santa Marta   Dim.3                 0.268
## 3 Neiva         Dim.3                 0.161
```

```
tabla_max_cos2_3 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.4") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 3 x 3
```

```
## # Groups:   mejor_componente [1]
##   ciudad mejor_componente max_cos2
##   <chr>   <chr>             <dbl>
## 1 Cartage Dim.4             0.440
## 2 Armenia Dim.4             0.295
## 3 Tunja   Dim.4             0.195
```

```
tabla_max_cos2_3 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.5") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 2 x 3
## # Groups:   mejor_componente [1]
##   ciudad mejor_componente max_cos2
##   <chr>   <chr>             <dbl>
## 1 Sincelejo Dim.5             0.542
## 2 Pereira   Dim.5             0.276
```

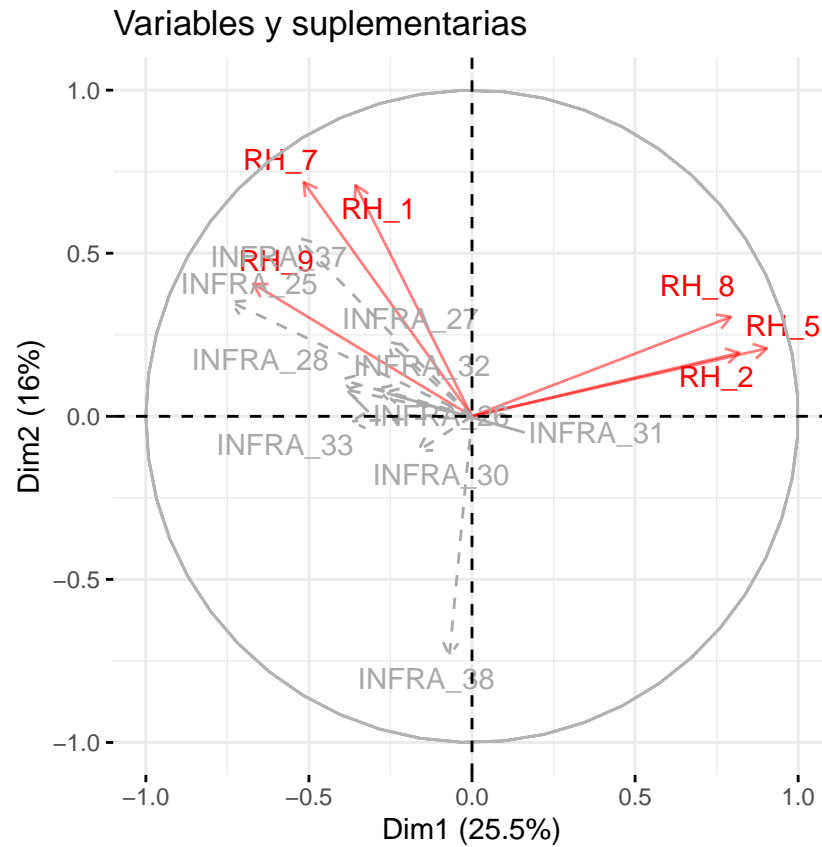
```
tabla_max_cos2_3 %>% group_by(mejor_componente) %>% filter(mejor_componente=="Dim.6") %>% arrange(desc(max_cos2))
```

```
## # A tibble: 3 x 3
## # Groups:   mejor_componente [1]
##   ciudad mejor_componente max_cos2
##   <chr>   <chr>             <dbl>
## 1 Pasto   Dim.6             0.467
## 2 Ibagué Dim.6             0.290
## 3 Cúcuta Dim.6             0.173
```

Dimensión	Ciudades
Dim.1	Riohacha, Valledupar, Montería, Manizales,
Dim.2	Bucaramanga, Popayán. Medellín, Bogotá, San Andrés, Barranquilla.
Dim.3	Villavicencio, Santa Marta, Neiva.
Dim.4	Cartagena, Armenia, Tunja.
Dim.5	Sincelejo, Pereira.
Dim.6	Pasto, Ibagué, Cúcuta.

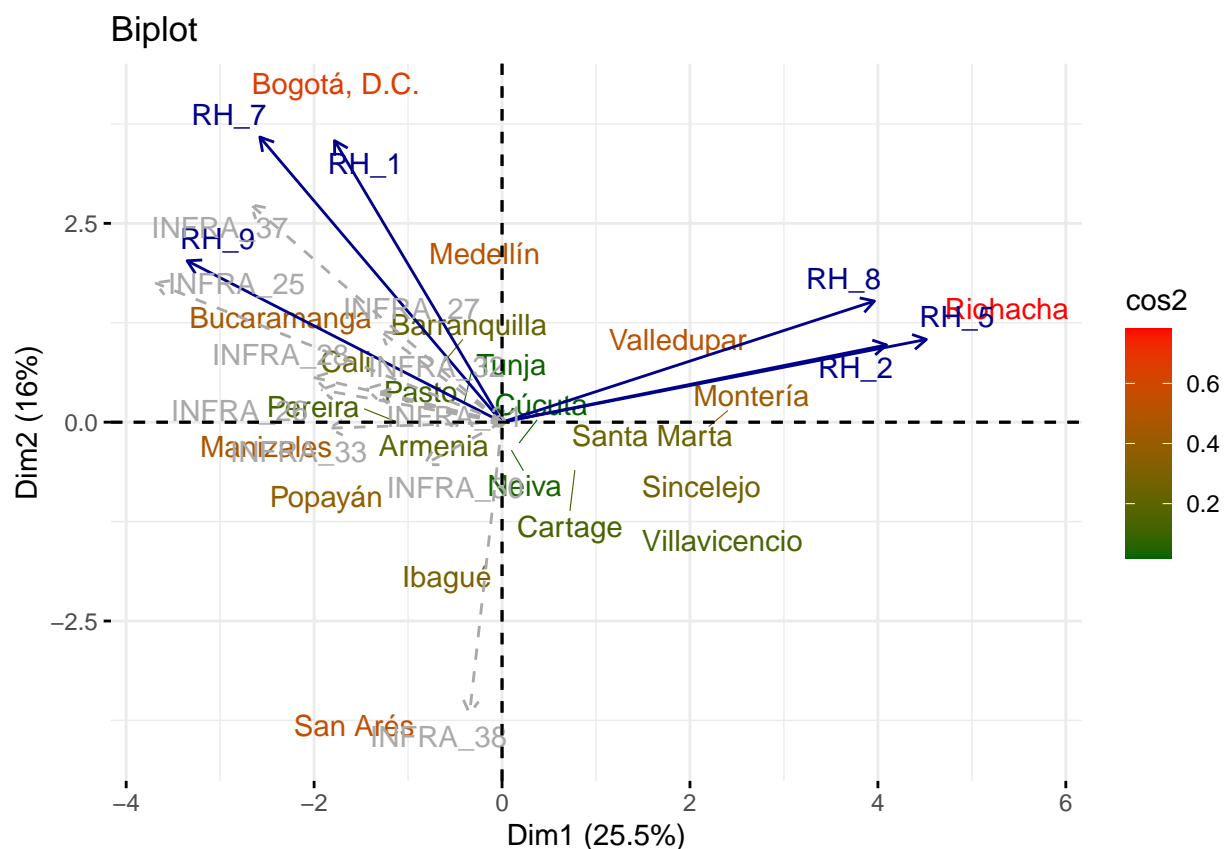
```
G32<-fviz_pca_var(acp3,
  col.var = 'red',
  alpha.var = 0.5,
  repel = TRUE,
  col.quanti.sup = 'darkgrey',
  select.var = list(name =c(all_vars3)),
  title = 'Variables y suplementarias'
)

G32
```



```
G33<-fviz_pca_biplot(acp3,
  col.ind = "cos2",
  gradient.cols = c("darkgreen", "red"),
  col.var = "darkblue",
  repel = T,
  col.quant.sup = 'darkgrey',
  select.var = list(name =c(all_vars3)),
  geom = "text",
  title = 'Biplot'
)
```

G33



Para este punto destacan nuevamente tres ciudades por su ubicación alejada del origen: Bogotá (3), San Andrés (17) y Riohacha (16).

San Andrés (17) se ubica en el tercer cuadrante con valores negativos en ambas dimensiones. Esta posición sugiere un bajo desempeño general, especialmente en infraestructura digital (INFRA_38, internet) y telecomunicaciones. Esto es coherente con su geografía insular y limitada escala urbana.

Bogotá (3) se encuentra en el segundo cuadrante, fuertemente asociada a variables como RH_7 (educación superior) y RH_1 (población), reflejando su liderazgo en aspectos demográficos y educativos.

Riohacha (16) se sitúa en el cuadrante superior derecho, destacándose por su fuerte asociación con RH_5 (analfabetismo) y RH_2 (crecimiento poblacional), lo que sugiere desafíos importantes en términos de desarrollo humano, a pesar de un crecimiento demográfico activo.

Comparación (PUNTO 4)

Al comparar los tres análisis, podemos ver que:

La dimensión o eje que más aporta varianza a los análisis se encuentra cuando las variables de recursos humanos son tomadas como suplementarias. Esto nos indica que las variables específicas de infraestructura son más representativas, debido a la gran varianza de estas variables entre las ciudades. Además, es el análisis con la menor cantidad de componentes principales, con solo cinco (de las cuales cuatro tienen correlaciones considerables con otros componentes principales).

Naturalmente, se puede apreciar que ciudades como Neiva, Pereira o Manizales, si bien resaltan de manera positiva con respecto a variables de infraestructura, están realmente cercanas al promedio al analizar las variables de recursos humanos. En el caso contrario, Riohacha es un dato a resaltar a través de los tres

gráficos, siempre ubicada y relacionada de manera cercana a las variables de crecimiento de la población y analfabetismo absoluto (aunque estas sean suplementarias en el segundo gráfico).

De igual manera, podemos resaltar cómo secuestros y cobertura bruta en primaria y secundaria son dos variables que solo vemos presentes como componentes principales en el gráfico combinado de ambos conjuntos de variables (además de aparecer con una correlación negativa altamente considerable). Al realizar el análisis solamente con variables de recursos humanos, estas dos variables son “reemplazadas” por las variables de población, cobertura en educación superior y calidad de los colegios, con una correlación a resaltar entre las tres.

En general, si bien la manera más fiel de realizar un análisis de las variables de infraestructura es al usar exclusivamente las variables de infraestructura, al emparejarlas con las variables de recursos humanos aún podemos encontrar correlaciones interesantes y ciudades atípicas con respecto al conjunto de variables.

Es de notar que a pesar de que se mantienen este tipo de diferencias hay ciudades que resaltan sin importar de que punto sea, como lo son el ejemplo de riohacha, bogotá y san andrés los cuales presentan diferencias importantes con respecto al promedio de las otras ciudades (están muy alejadas siempre del centro). Esto se puede deber a que si se piensa desde la realidad, la infraestructura puede estar muy relacionada con el acceso a recursos humanos, por lo que tiene sentido que se mantengan ciertas ciudades sobresalientes

las conclusiones pueden ser que siempre es necesario saber de lo que hablan nuestras variables para encontrar las formas mas efectivas de concentrar la información, también de la importancia de agrupar y hacer casos para ver cuales son las cosas que mas afectan y como terminan en los resultados finales

Cosenos cuadrados

De todas las componentes a pesar de que estamos tratando diferentes, hubo ciudades que se mantuvieron en la misma componente (no necesariamente en el mismo orden):

```
library(dplyr)

tabla_max_cos2 <- tabla_max_cos2 |>
  rename(componente_pca1 = mejor_componente, cos2_pca1 = max_cos2)

tabla_max_cos2_2 <- tabla_max_cos2_2 |>
  rename(componente_pca2 = mejor_componente, cos2_pca2 = max_cos2)

tabla_max_cos2_3 <- tabla_max_cos2_3 |>
  rename(componente_pca3 = mejor_componente, cos2_pca3 = max_cos2)

pca_comparada <- tabla_max_cos2 |>
  inner_join(tabla_max_cos2_2, by = "ciudad") |>
  inner_join(tabla_max_cos2_3, by = "ciudad")
pca_comparada <- pca_comparada |>
  mutate(
    mismo_componente = componente_pca1 == componente_pca2 &
      componente_pca1 == componente_pca3
  )
```

Dimensión	Ciudad	Valor en el primer punto	Valor en el segundo punto	Valor en el tercer punto
	Riohacha	0.6968575	0.7747932	0.7524178
	Valledupar	0.6180702	0.7136357	0.4813210

Dimensión	Ciudad	Valor en el primer punto	Valor en el segundo punto	Valor en el tercer punto
Dim.1	Montería	0.4781262	0.5791771	0.4293558
	Manizales	0.4604778	0.4933737	0.4275987
	Bucaramanga	0.5427632	0.4394121	0.3668279
	Cali	0.3067947	0.4458738	0.2093450
Dim.6	Ibagué	0.3851752	Dim.4: 0.2941408	0.2901214

En primer lugar, se evidencia que ciudades como Riohacha y Valledupar mantienen valores de coseno cuadrado consistentemente altos en los tres puntos de análisis, lo que sugiere que su posición relativa frente a los componentes principales es relativamente estable y que su estructura de datos es bien capturada en todos los casos. Esto coincide con el hecho de que Riohacha, por ejemplo, ya había sido identificada como una ciudad atípica en las proyecciones de los ejes principales por su relación con variables de crecimiento poblacional y analfabetismo.

Por el contrario, otras ciudades como Cali y Bucaramanga presentan valores de coseno cuadrado comparativamente más bajos, especialmente en el tercer análisis, lo que podría interpretarse como una representación menos robusta en esos componentes. Esto sugiere que, dependiendo del conjunto de variables incluidas, la explicación de su variabilidad requiere más componentes principales o que su relación con las variables no es tan marcada.

En el caso particular de Ibagué, se observa que en el segundo punto de análisis su componente dominante cambia (Dim.4), y su coseno cuadrado es más bajo en comparación con otras ciudades, lo que indicaría cierta variabilidad en su representación según el enfoque de análisis (infraestructura, recursos humanos o combinado).

En general, los cosenos cuadrados altos implican que el modelo de componentes principales captura de forma satisfactoria la información de la ciudad en ese componente específico. El hecho de que varias ciudades mantengan altos valores en las tres perspectivas refuerza la idea de que ciertos patrones estructurales en los datos se preservan incluso al variar el conjunto de variables analizadas. Esto complementa la interpretación anterior de que existen ciudades con características particulares que sobresalen de manera consistente frente al promedio, como es el caso de Riohacha.