

Nombre: David Plazas Escudero Código: 201710005101 Nota:

Profesor: Francisco Zuluaga

Grupo: 001

Mayo 2 de 2019

3) Utilizando la base de datos **USArrests** de R, CARGUE EL ARCHIVO con este comando **data(USArrests)** y con el comando **?USArrests** obtiene informacion de la base de datos

1. Calcule los componentes principales a partir de la matriz de correlacion
Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	1.5748783	0.9948694	0.5971291	0.41644938
Proportion of Variance	0.6200604	0.2474413	0.0891408	0.04335752
Cumulative Proportion	0.6200604	0.8675017	0.9566425	1.00000000

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4
Murder	0.536	0.418	0.341	0.649
Assault	0.583	0.188	0.268	-0.743
UrbanPop	0.278	-0.873	0.378	0.134
Rape	0.543	-0.167	-0.818	

2. Utilizando 3 criterios determine cuantos componentes retener: Se deben retener los 2 primeros componentes principales, ya que dos de tres criterios lo sugieron.

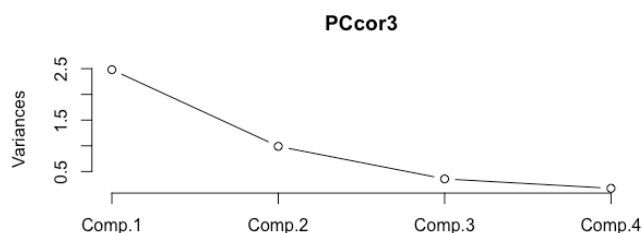
1. 80% del acumulado: Se retienen los dos primeros componentes principales, ya que ambos acumulan más del 80% de la varianza.

	Comp.1	Comp.2	Comp.3	Comp.4
Cumulative Proportion	0.6200604	0.8675017	0.9566425	1.00000000

2. Por encima del valor propio promedio: El valor propio promedio es de 1, sólo el primer componente principal se retendría.

	Comp.1	Comp.2	Comp.3	Comp.4
EigenValues	2.4802416	0.9897652	0.3565632	0.1734301

3. Scree plot: Se retendrían los 2 primeros componentes principales, ya que la línea que une al 1 y al 2 tiene una pendiente muy distinta a las demás.



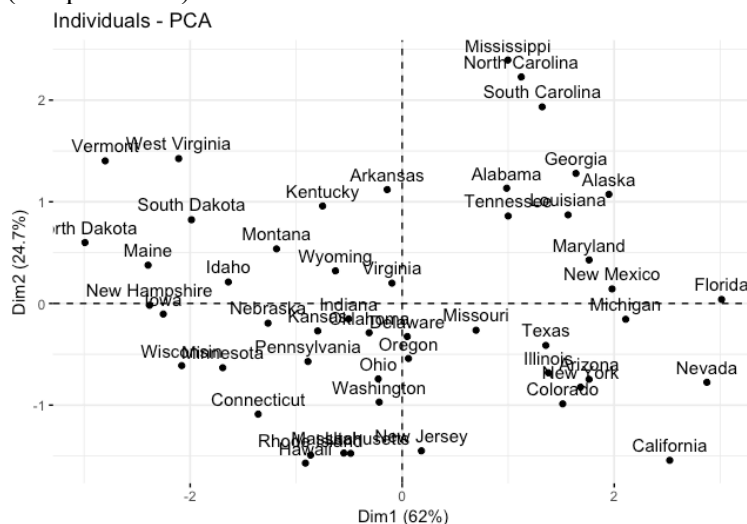
3. Si el primer componente lo interpretamos como un índice de violencia. Cuales son la dos ciudades con mayor índice de violencia? . obtenga las puntuaciones de todas las ciudades y ordene de mayor a menor:

Claramente se ve que los dos estados con mayor índice de violencia son Florida y Nevada. Puntuaciones de mayor a menor:

Florida	Nevada	California	Michigan	New Mexico	Alaska	Maryland	Arizona
100.000000	97.691232	91.856114	84.936540	82.798128	82.300465	79.190389	79.186961
New York	Georgia	Louisiana	Colorado	Illinois	Texas	South Carolina	North Carolina
77.845037	77.124097	75.884102	75.047251	72.788366	72.392504	71.819434	68.533233
Tennessee	Mississippi	Alabama	Missouri	New Jersey	Oregon	Delaware	Virginia
66.474425	66.420352	66.238369	61.429615	52.850127	50.811184	50.621138	48.222507
Arkansas	Washington	Ohio	Oklahoma	Massachusetts	Indiana	Utah	Wyoming
47.471743	46.214795	46.063893	44.634857	41.731016	41.409714	40.658642	39.345438
Kentucky	Kansas	Rhode Island	Pennsylvania	Hawaii	Montana	Nebraska	Connecticut
37.323340	36.556979	35.443107	35.032738	34.629071	30.086480	28.751242	27.202419
Idaho	Minnesota	South Dakota	Wisconsin	West Virginia	Iowa	New Hampshire	Maine
22.520657	21.640064	16.726161	15.195183	14.714418	12.298861	10.129610	9.914564
Vermont	North Dakota						
3.177442	0.000000						

4. Haga una grafica de los dos primeros componentes principales.(como interpreta esta grafica relacionadola con el punto 3

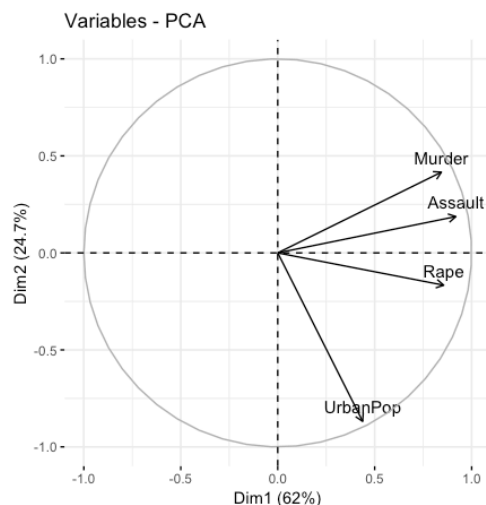
En la gráfica se puede ver que, para el primer componente principal (Dim 1, eje x), los estados de Florida y Nevada son los que mayor valor tienen, y efectivamente son las que hallamos con mayor índice de violencia (componente 1).



5. Interprete el círculo de correlación:

Se puede ver que hay una alta influencia en las variables de asalto, asesinato y violación que contribuyen al primer componente principal (índice de violencia), ya que su flecha tiene gran magnitud en dirección del primer componente principal. Por otra parte, como los ángulos entre éstas variables son relativamente pequeños, indica que las variables tienen una correlación significativa.

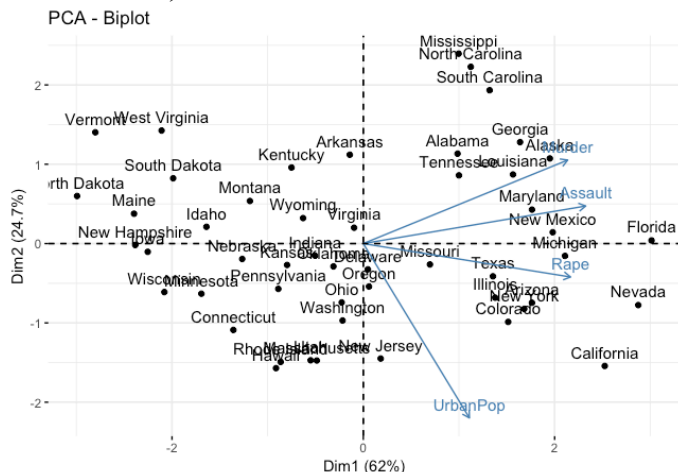
Por último, nótese que la variable población urbana, tiene muy poca contribución al primer componente principal (índice de violencia) y que tiene mayor impacto en el segundo componente principal. Además, tiene un ángulo relativamente grande respecto a las demás variables, lo que indica una correlación baja.



6. Interprete el BIPLLOT

Del biplot se pueden ver los valores en los que las variables tienen mayor valor (o menor). Por ejemplo, si se ve el caso de estados como California o New Jersey, éstos son estados que tienen mucha población, ya que tienen una alta proporción de su valor en la variable población urbana; mientras que estados como West Virginia o South Dakota tienen bajos valores. Por otro lado, por ejemplo, se puede ver que estados como Maryland o New Mexico tienen un alto valor de asaltos, ya que su distancia a la línea que representa la variable asaltos es muy baja.

En general, entre más cerca esté un “individuo” de una línea que representa una variable, quiere decir que tiene un mayor valor para esta variable. Por ejemplo, Florida es un estado que tienen mayor valor para las variables asalto, violación e incluso asesinatos.



- 4) Los datos de la pagina 502 del libro presentan información preliminar sobre un estudio de contaminación. **Tabla 14.13 Para el análisis omita y1.(diox en el archivo adjunto)**

Archivo : **dioxido.txt**

recuerde que con este comando los nombres de la ciudad aparecen etiquetados en el grafico

```
rownames(dioxido)<-c("phoenix","little Rock","San Francisco","Denver","Harford","Wilmington","Washington",
"Jacksonville","Miami","Atlanta","Chicago","Indianapolis","Des Moines","Wichita",
"Louisville","New Orleans","Baltimore","Detroit","Minneapolis","Kansas City","St Louis",
"Omaha","Albuquerque","Albany","Buffalo","Cincinnati","Cleveland","Columbus",
"Philadelphia","Pittsburgh","Providence","Memphis","Nashville","Dallas","Hosuton","Salt Lake City",
"Norfolk","Richmond","Seatle","Charleston","Milwaukee")
```

```
#      temp      :      temperatura      anual      promedio      en      grados      F
#      emp      :      Numero      de      empresas      con      un      numero      de      trabajadores      mayores      a      20
#      pob      :      Tamaño      de      la      población      en      miles
#      viento   :      promedio      anual      de      la      velocidad      del      viento      en      millas      por      hora
#      precip   :      precipitación      anual      promedio      en      pulgadas
#      dias     :      Numero      promedio      de      dias      con      precipitaciones      por      año
```

1. Calcule los componentes principales a partir de la matriz de correlacion

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.4811699	1.2251019	1.1815108	0.8719905	0.33829237	0.185698720
Proportion of Variance	0.3656441	0.2501458	0.2326613	0.1267279	0.01907362	0.005747336
Cumulative Proportion	0.3656441	0.6157899	0.8484511	0.9751790	0.99425266	1.000000000

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
temp	0.329	0.135	0.670	0.307	0.559	0.133
emp	-0.612	0.170	0.270	-0.136		0.704
pob	-0.578	0.226	0.348		-0.695	
viento	-0.353	-0.135	-0.297	0.869	0.114	
precip	-0.617	0.511	0.171	-0.569		
dias	-0.237	-0.709		-0.312	0.580	

2. Utilizando 3 criterios determine cuantos componentes retener.

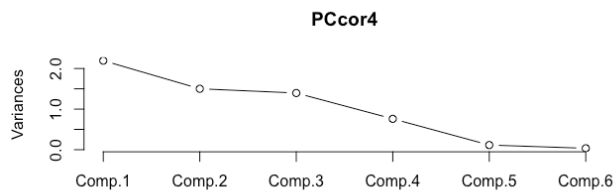
1. 80% De la varianza: Se deberían retener los primeros 3 componentes principales.

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Cumulative Proportion	0.3656441	0.6157899	0.8484511	0.9751790	0.99425266	1.000000000

2. Valor propio promedio: El valor propio promedio también es de 1. Con esto, debería retener los primeros 3 componentes principales, ya que su varianza es mayor a la varianza promedio.

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
	2.19386437	1.50087474	1.39596767	0.76036747	0.11444173	0.03448401

3. Scree Plot: A partir del scree plot, se deberían retener 2, 3 o, tal vez, 5 componentes principales, ya que en estos puntos se presenta un quiebre en la pendiente significativo, no obstante, es más adecuado retener 3 ya que a partir de allí la pendiente es muy parecida en los demás componentes principales.



3. Si el primer componente lo interpretamos como un índice de calidad de vida. Cuales son la dos ciudades que tiene mejor calidad de vida? . obtenga las puntuaciones de todas las ciudades y ordene de mayor a menor:

A partir de las puntuaciones, se puede observar que las ciudades de Phoenix y Little Rock son las que tienen mayor índice de calidad de vida.

phoenix	little Rock	Miami	New Orleans	Charleston	Albuquerque
100.00000	90.79626	89.92989	89.03814	88.77629	88.60071
Jacksonville	Richmond	Wilmington	Salt Lake City	Nashville	Atlanta
86.51727	85.88679	83.93013	82.97659	82.96296	79.49023
Norfolk	Memphis	Albany	Cincinnati	San Francisco	Louisville
79.39085	79.25189	78.83119	78.47246	78.38040	77.53460
Harford	Denver	Wichita	Omaha	Des Moines	Kansas City
75.25248	75.09631	75.00069	74.29446	74.28124	74.26284
Columbus	Washington	Dallas	Providence	St Louis	Indianapolis
74.13404	73.05300	72.86014	72.03044	69.59178	69.36145
Pittsburgh	Seattle	Hosuton	Baltimore	Buffalo	Milwaukee
69.21361	67.43916	67.11543	67.10692	57.60320	57.27259
Minneapolis	Cleveland	Detroit	Philadelphia	Chicago	
56.05273	53.08140	48.59554	41.53770	0.00000	

4. Si el segundo componente principal lo interpretamos como un índice de tiempo Humedo y el tercero como un indicador de tipo de clima, Cuales son las ciudades mas húmedas y las mas calurosas

Tiempo Húmedo (2do componente principal): La dos ciudades más húmedas son Phoenix y Alburquerque.

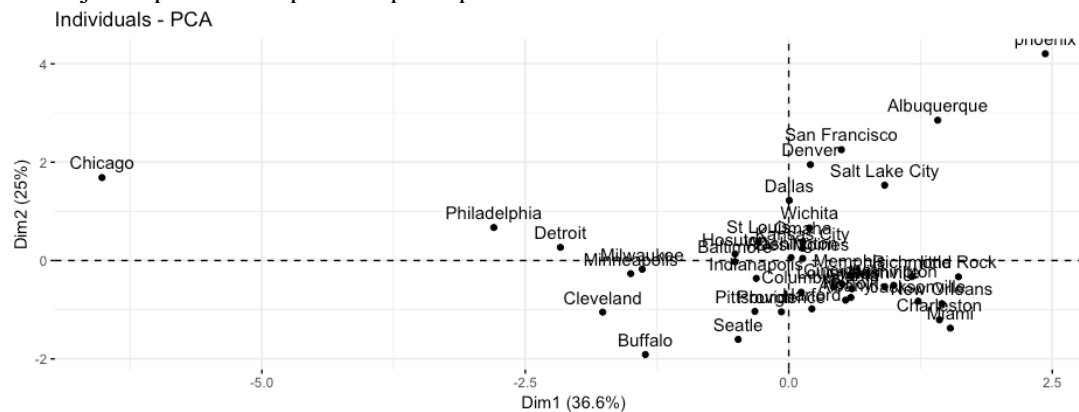
phoenix	Albuquerque	San Francisco	Denver	Chicago	Salt Lake City
100.000000	77.912397	68.104098	63.137732	58.828737	56.300194
Dallas	Philadelphia	Wichita	St Louis	Omaha	Detroit
51.236286	42.233881	42.097239	37.563608	37.329994	35.630171
Kansas City	Hosuton	Washington	Des Moines	Baltimore	Milwaukee
35.333693	33.442802	32.221924	31.965221	30.851225	28.355088
Minneapolis	Memphis	Richmond	little Rock	Indianapolis	Cincinnati
26.892454	26.157615	25.906891	25.842784	25.319561	23.386852
Wilmington	Nashville	Louisville	Atlanta	Columbus	Norfolk
23.018270	22.554885	22.510490	21.849088	20.732047	18.980385
Albany	Jacksonville	New Orleans	Harford	Pittsburgh	Providence
18.089985	17.737789	16.916070	15.200656	14.347369	14.193822
Cleveland	Charleston	Miami	Seattle	Buffalo	
14.098116	11.528688	8.781031	5.038425	0.000000	

Tipo de Clima (3er componente principal): Las dos ciudades más calientes son Miami y Chicago.

Miami	Chicago	New Orleans	Hosuton	Jacksonville	Philadelphia
100.0000000	91.8770215	85.9028315	85.6489184	83.2551055	72.3551053
Memphis	Atlanta	Dallas	phoenix	Nashville	little Rock
65.7994073	63.1453500	62.7284513	60.7706387	60.0571667	59.5596079
Richmond	Louisville	Baltimore	Washington	Cincinnati	St Louis
52.0344432	49.0262190	48.6260426	48.3910025	46.5591052	43.7839757
Charleston	Norfolk	Detroit	San Francisco	Kansas City	Indianapolis
42.5085407	41.9134095	36.9233328	34.6420901	34.0136500	33.9506154
Columbus	Wilmington	Harford	Seattle	Pittsburgh	Cleveland
30.9061829	30.6298945	27.0722788	26.9053684	23.5936292	23.5892317
Providence	Wichita	Omaha	Denver	Albuquerque	Albany
20.3971982	14.4051353	12.1266455	11.0600431	10.6339442	9.5696311
Salt Lake City	Des Moines	Milwaukee	Minneapolis	Buffalo	
4.3751085	2.7800282	1.8828346	0.5648954	0.0000000	

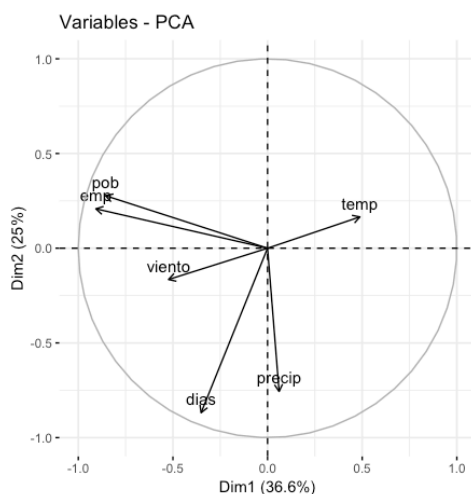
5. Haga una grafica de los dos primeros componentes principales.(como interpreta esta grafica relacionadola con el punto 3

En la gráfica se ve que Phoenix es la ciudad con mayor valor en los 2 primeros componentes principales, al estar más alejado del origen en ambas direcciones. En la tarea anterior se vio que Albuquerque era la segunda ciudad más caliente y acá se evidencia al ser la segunda que más distancia tiene en el eje del segundo componente principal. Por otra parte, en el ranking obtenido para el primer componente, se vio que Chicago era la ciudad con menor índice de calidad de vida, y el efecto, ya que es la que tiene mayor distancia negativa en el eje del primer componente principal.



6. Interprete el círculo de correlacion:

En el círculo de correlación se puede ver las relación de las variables con los componentes principales y entre ellas: Por ejemplo, se puede ver que la variable temperatura está estrechamente correlacionada con la variable viento, pero con correlación negativa (relación inversamente proporcional), por otra parte, la variable temperatura no está correlacionada significativamente con la variable precipitación, al tener un ángulo “grande” entre ellos. Por otra parte, respecto a los componentes principales, se puede ver claramente que la variable precipitación está ligada estrechamente al segundo componente principal, ya que está en dirección de este eje. Así consecutivamente.



7. Interprete el BIPLLOT

En el BIPLOT se puede ver la relación de los individuos y las variables. Por ejemplo, Phoenix es una ciudad con poca precipitación y pocos días de precipitación. Por ejemplo, Detroit y Philadelphia son ciudad con altos valores de población y altos valores de empresas con trabajadores mayores a 20 años. Por ejemplo, Cleveland registra altos valores para la velocidad del viento y muy bajos para la temperatura.

