

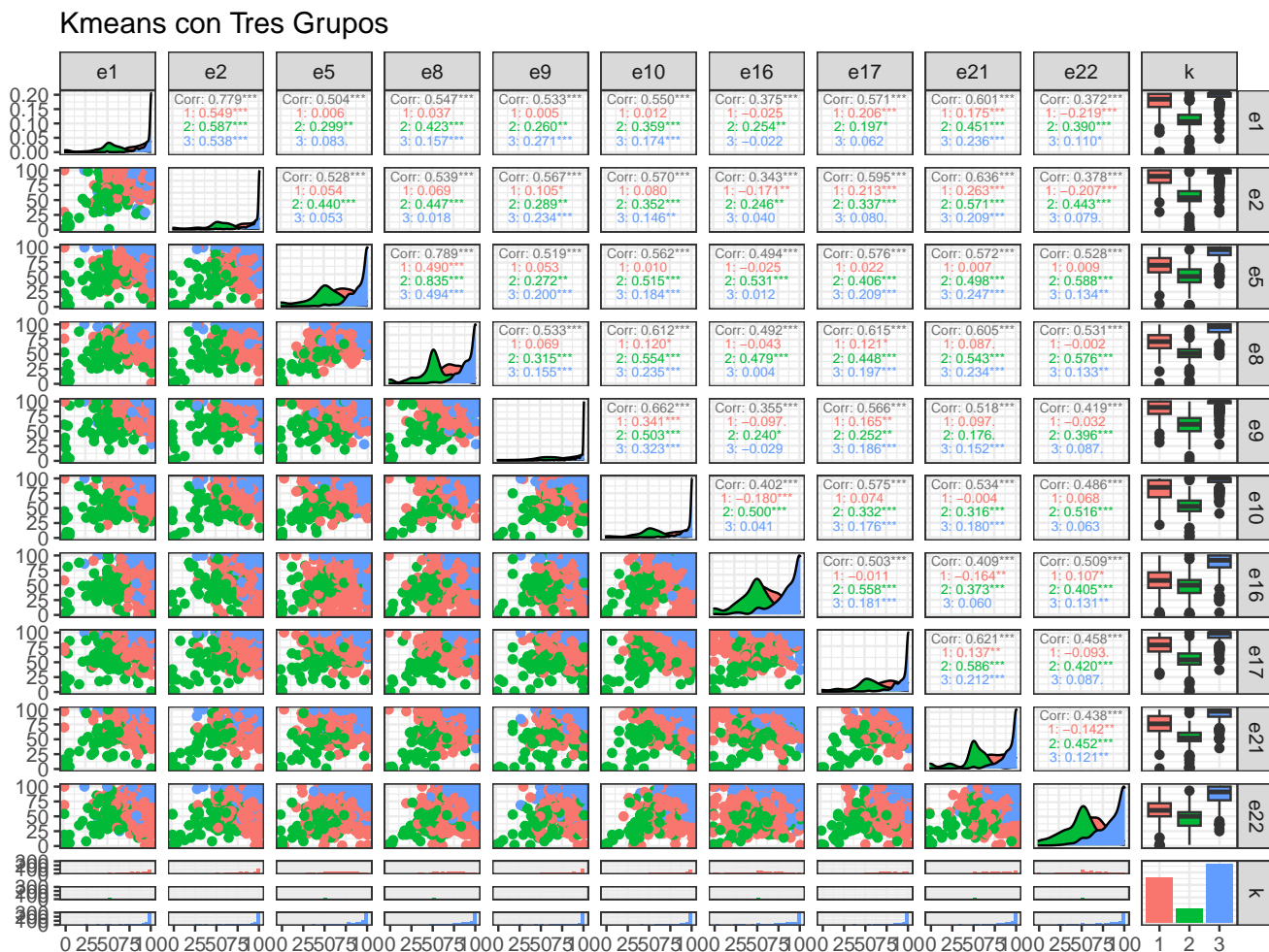
4. Análisis de conglomerados

El objetivo del análisis es identificar grupos de clientes para focalizar la publicidad de Oddjob Airways, a partir de una encuesta resumida en `Dat4ExB.csv`, y cuyas respuestas van de 1 a 100 (100 es que la persona considera que un aspecto es crucial en el servicio, mientras que 1 corresponde a que no lo es). Estos aspectos son puntualidad (e1), servicio según lo ofrecido (e2), experiencia placentera (e5), comodidad (e8), seguridad (e9), estado del avión (e10), comida adecuada (e16), hospitalidad (e17), viajar de forma sencilla (e21) y entretenimiento a bordo (e22). Como primer paso vamos a considerar que las variables son continuas, entonces dado ese supuesto obtendremos algunos grupos considerando el método k-means.

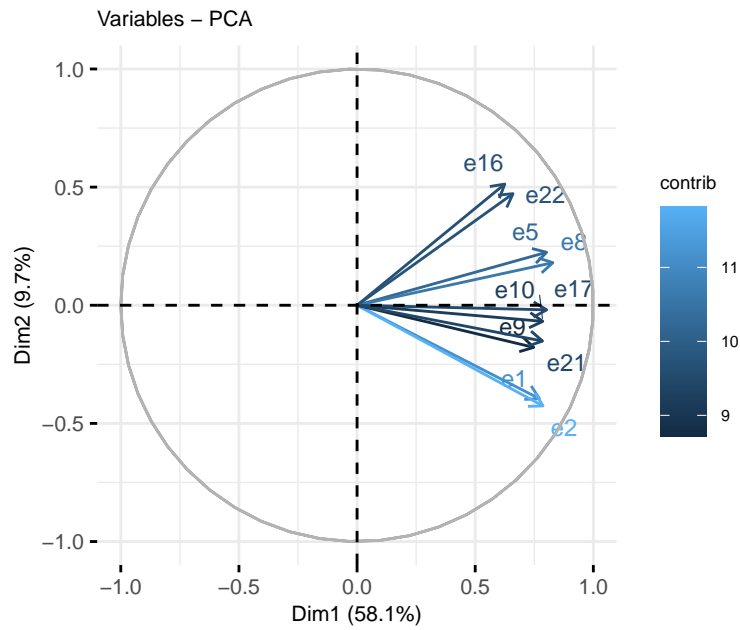
Aún cuando el indicador de **Average Silhouette width** y los indicadores de **Connectivity** y **Dunn** muestran que el número óptimo de clusters es de 2, el indicador de **Hubert statistic values** muestra que deben de ser 3, y el indicador de **Dindex values** que deben de ser 5. Por lo que no hay un consenso indiscutible del número de clusters a considerar como óptimos. (Chunks `clValid`, `fviz_nbclust_kmeans_silhouette` y `NbClust`, líneas 66, 78 y 87).

Se decidió tomar al menos tres aspectos generales del servicio detectados en las variables: puntualidad y servicio según lo ofrecido; seguridad y estado del avión; y comodidad, experiencia, entretenimiento, hospitalidad y comida. Podemos focalizar la publicidad de la empresa en 3 grupos de clientes con base a estos tres aspectos.

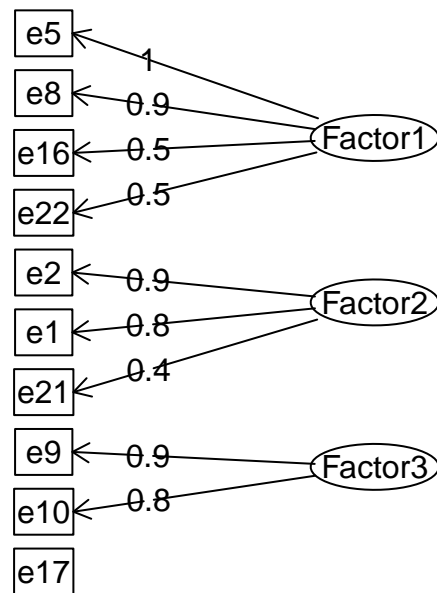
En la siguiente Gráfica podemos observar la asignación de clusters para tres grupos, y las correlaciones entre los aspectos: las correlaciones más altas son entre puntualidad (e1) y servicio acorde a lo ofrecido (e2) por un lado, seguridad (e9) y avión en buen estado (e10) por otro, y por otro lado experiencia placentera (e5) con comodidad (e8).



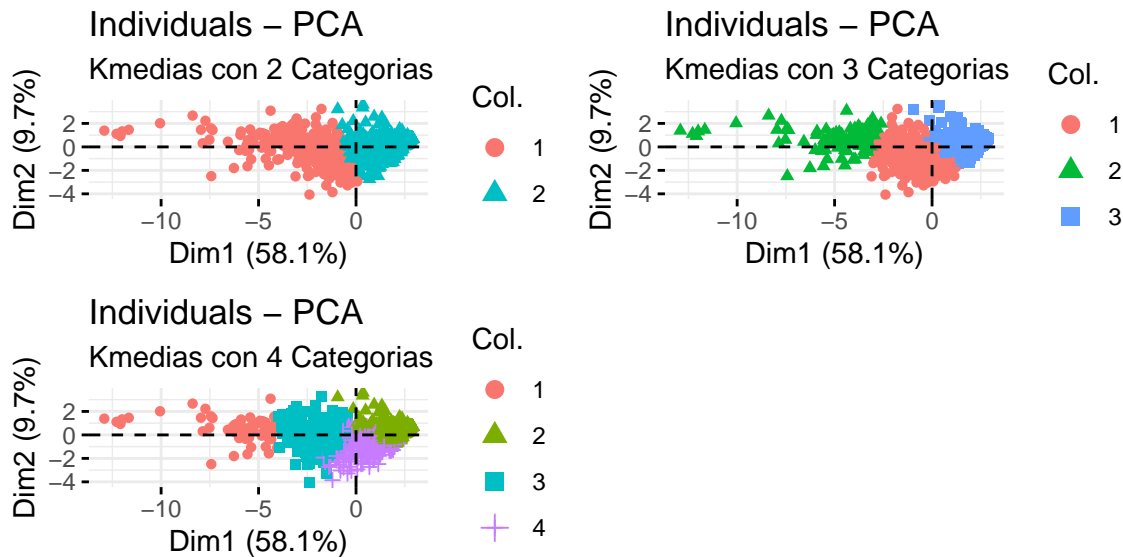
En los siguientes resultados auxiliares a este análisis, tenemos que la primera y segunda componente principal conservan una varianza de 58.1% y 9.7% respectivamente. Además si consideramos 3 factores, tenemos consistencia en lo planteado con los grupos. En ambos casos podemos observar e1 y e2 muy correlacionados o en el mismo factor; e5, e8, e16 y e22 por otra parte; y e9 y e10 por otra parte.



Factor Analysis



En la primera gráfica siguiente, vemos que a la derecha se encuentran los clientes potenciales con buenas expectativas en general en todas las preguntas, y a la izquierda los de regular y mala, según el primer componente principal.

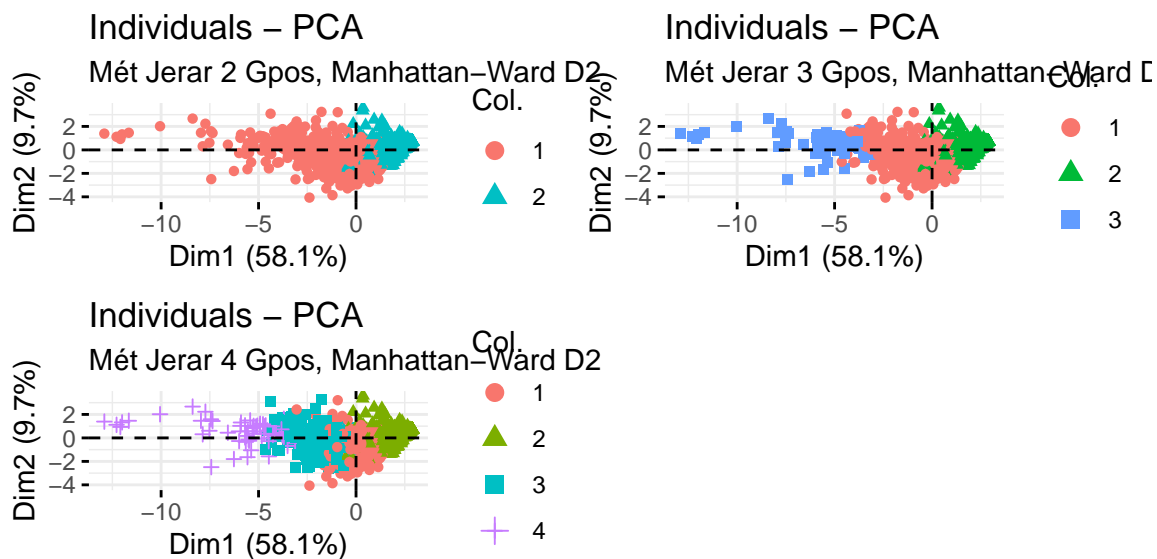


De acuerdo al método silhouette, se nos sugiere tomar dos grupos, pero en nuestro caso para mejorar la interpretación decidimos que es recomendable usar más de 2. Adicionalmente podemos ver que no cambia mucho la interpretación si nos quedamos con 3 grupos o con 4, pues cuando agrupamos en 4 grupos, el grupo 4 combina parte de los grupos 1 y 3.

Observando los componentes principales, podemos decir que es mejor focalizar la publicidad en 3 grupos de clientes: los que esperan puntualidad y un servicio acorde a lo contratado; los que esperan seguridad y buen mantenimiento y estado del avión; y los que esperan comodidad, experiencia, hospitalidad y entretenimiento. Tal como se decidió agrupar desde un inicio.

Metodo Jerarquico Aglomerativo

Para esto vamos a tomar que las variables son continuas como se hizo anteriormente y tomando tanto las escalas dadas como haciendo transformaciones. Además agregaremos las disimilaridades entre clientes y clusters.



En esta figura podemos observar las mismas comparaciones que realizamos en el ejercicio 1 donde se puede ver que el resultado obtenido en este caso aplicando el método aglomerativo resultó ser muy similar al obtenido con K-means. En esta ocasión los 3 grupos de clientes son: los que esperan puntualidad y servicio acorde a lo ofrecido, los que esperan seguridad y buen estado y mantenimiento del avión, y los que esperan experiencia placentera, comodidad, hospitalidad y entretenimiento.

Modificaciones y uso de Componentes principales

Importance of components:

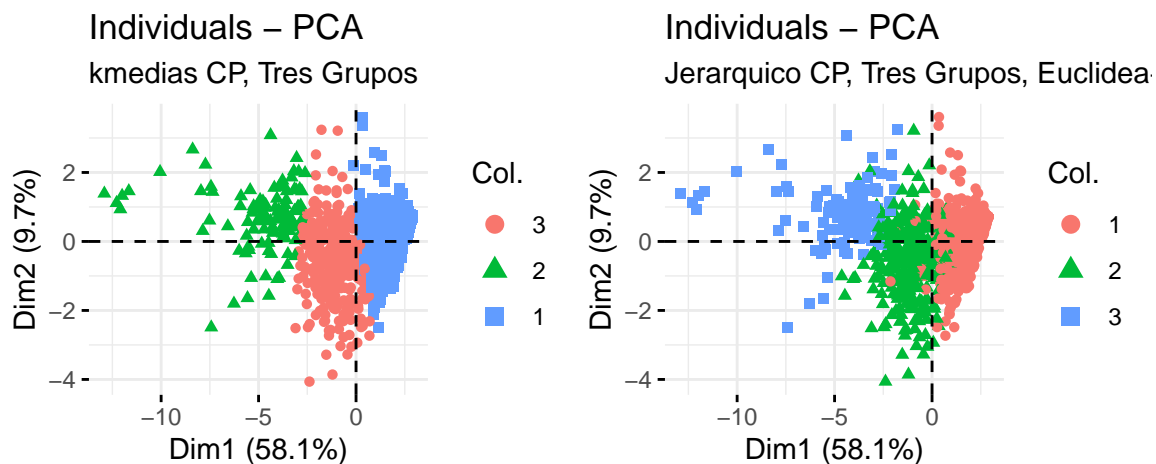
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
## Standard deviation	2.4101	0.98298	0.79440	0.75653	0.70253	0.65595	0.58946
## Proportion of Variance	0.5808	0.09663	0.06311	0.05723	0.04935	0.04303	0.03475
## Cumulative Proportion	0.5808	0.67747	0.74057	0.79781	0.84716	0.89019	0.92493

	PC8	PC9	PC10
## Standard deviation	0.57595	0.47730	0.43717
## Proportion of Variance	0.03317	0.02278	0.01911
## Cumulative Proportion	0.95811	0.98089	1.00000

Primero vamos a hacer el proceso de K-means con las 4 componentes principales que se escogieron.

Análogamente a los ejercicios anteriores vamos a probar usando clusters jerarquicos y conservando las disimilaridades que se usaron en el caso anterior

Obtuvimos que los mejores modelos a usar para 3 clusters fueron Euclidean, Minkowski y Ward D2.



Conclusiones

Como pudimos ver a lo largo de todo este análisis y aplicando distintos métodos de evaluación como fue usar K-means, algoritmos de jerarquía y componentes principales decidimos conservar el de Componentes principales ya que además de permitirnos conservar las variables que conservan mayor información dadas las originales y así poder reducir el estudio a estas los resultados obtenidos fueron mas cercanos a lo que deseamos, por ejemplo, la clusterización que obtuvimos con la primer componente fue mejor, lo mismo pasó para la segunda componente. Hablando en términos mas generales tenemos que el primer grupo tiene mayor promedio en todas las respuestas, seguido por el segundo grupo y por ultimo se queda el tercer grupo.

Finalmente, creemos que el modelo a utilizar para focalizar la publicidad al publico siempre dependerá en gran medida de el numero de la cantidad de publico que quiera alcanzar la empresa y conforme a esto lanzar los distintos tipos de publicidad, ya que nosotros decidimos tomar 3 clasificaciones sobre 2 0 4, esto con el fin de mantener un equilibrio entre las preferencias de todos los clientes que buscan seguridad, puntualidad y un buen trato por parte de los trabajadores, cosas que sin duda son fundamentales para que la empresa logre atraer nuevos clientes potenciales que le den una gran importancia a estos criterios.