



**Escuela de Gobierno y
Transformación Pública
Tecnológico de Monterrey**

TAREA # 2, PARTE 1

Ciencia de datos

Entrega: 10 de marzo del 2021

Guillermo Alberto García Candanosa
A01034958@itesm.mx

Instrucciones

Descarga el archivo zippeado y utiliza el archivo de R que le acompaña. Este será tu kit de inicio para que construyas sobre el mismo.

Similar al ejercicio que vimos en clase, tu tarea consiste en elaborar un índice de ingreso para la zona metropolitana de Monterrey. Por favor, considera los siguientes pasos intermedios:

1. Revisa la metodología que emplea el AMAI (con base en encuestas y no datos censales) para estimar nivel socioeconómico de los hogares de México. <http://nse.amai.org/niveles-socio-economicos-amai/> (Enlaces a un sitio externo.)
2. Busca el diccionario de datos del censo de la tarea anterior
3. Trata de replicar la selección de variables del AMAI. Puedes considerar alguna variable adicional o eliminar la que consideres. El objetivo es que generes un índice de ingreso para la zona metropolitana y sea acertado. ¿Hay alguna nueva variable del censo nuevo que consideres que valga la pena incluir y sea un buen indicador de ingreso? Ejemplo: TV por streaming o consolas de videojuegos.
4. Selecciona las AGEBS de los municipios de la zona metropolitana. De acuerdo con INEGI, son los siguientes: Apodaca, El Carmen, García, San Pedro Garza García, General Escobedo, Guadalupe, Juárez, Monterrey, Salinas Victoria, San Nicolás de los Garza, Santa Catarina, Santiago, Cadereyta Jiménez, General Zuazua, Pesquería.
5. Utiliza PCA para generar tu índice. Realiza una selección de variables de tal forma que la primera componente explique más del 65% de la varianza en los datos.
6. Reporta en esta tarea tu selección final de variables y el porcentaje de la varianza explicada por la primera componente.
7. Construye un mapa de leaflet con tu índice para la Zona Metropolitana
8. Sube tu archivo .html con el índice y un documento .pdf con el listado de variables que incluiste y la varianza explicada.

Selección de variables

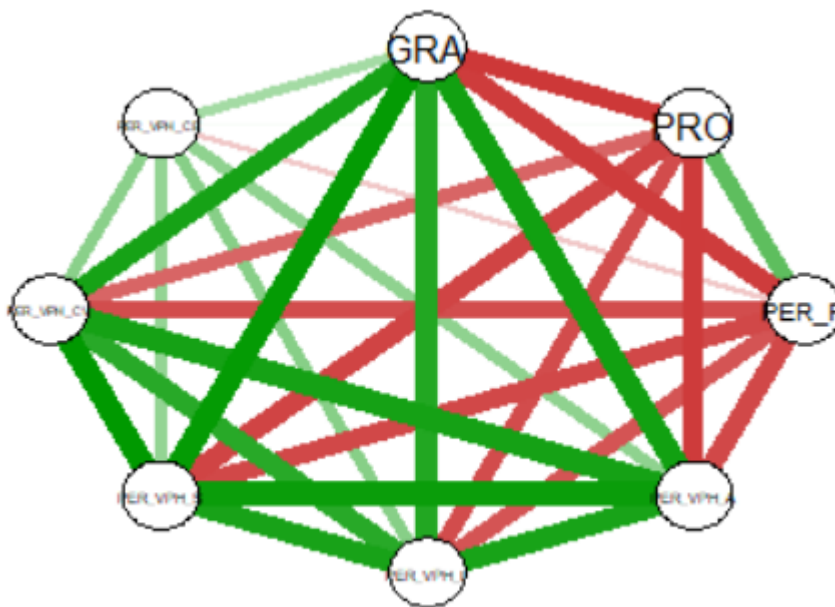
Las variables que seleccioné para llevar a cabo este ejercicio fueron las siguientes:

- **GRAPROES**: Grado promedio de escolaridad.
- **PRO_OCUP_C**: Promedio de ocupantes por cuarto en viviendas particulares habitadas.
- **PER_PSINDER**: Porcentaje de la población sin afiliación a servicios de salud. Variable creada a partir de la división de:
 - **PSINDER**: Población sin afiliación a servicios de salud.
 - **POBTOT**: Población total.
- **PER_VPH_AUTOM**: Porcentaje de las viviendas particulares habitadas que disponen de un automóvil o camioneta.
 - **VPH_AUTOM**: Viviendas particulares habitadas que disponen de automóvil o camioneta.
 - **TVIVPARHAB**: Total de viviendas particulares habitadas.

- **PER_VPH_INTER:** Porcentaje de las viviendas particulares habitadas que disponen de internet.
 - **VPH_INTER:** Viviendas particulares habitadas que disponen de internet.
 - **TVIVPARHAB:** Total de viviendas particulares habitadas.
- **PER_VPH_SPMVPI:** Porcentaje de las viviendas particulares habitadas que disponen de servicio de películas, música o videos de paga por internet.
 - **VPH_SPMVPI:** Viviendas particulares habitadas que disponen de servicio de películas, música o videos de paga por internet.
 - **TVIVPARHAB:** Total de viviendas particulares habitadas.
- **PER_VPH_CVJ:** Porcentaje de las viviendas particulares habitadas que disponen de consola de videojuegos.
 - **VPH_CVJ:** Viviendas particulares habitadas que disponen de consola de videojuegos.
 - **TVIVPARHAB:** Total de viviendas particulares habitadas.
- **PER_VPH_CEL:** Porcentaje de las viviendas particulares habitadas que disponen de consola de teléfono celular.
 - **VPH_CEL:** Viviendas particulares habitadas que disponen de teléfono celular.
 - **TVIVPARHAB:** Total de viviendas particulares habitadas.

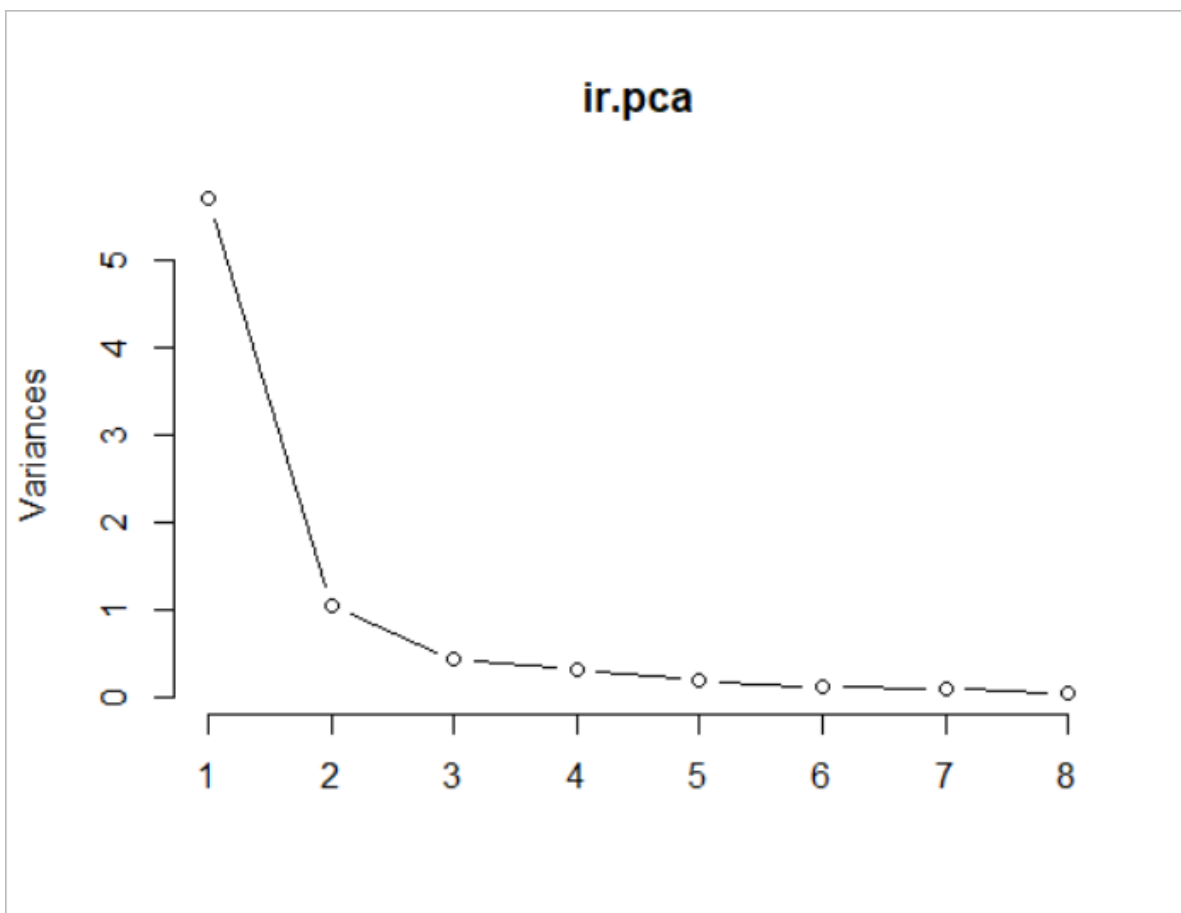
Resultados

Se determinó el grado de asociación de las variables con la siguiente gráfica:



La primera componente, correspondiente a mi índice, alcanzó a explicar un 71.04% de la varianza.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	2.38400	1.03030	0.66090	0.56967	0.44611	0.35961	0.33710	0.22751
Proportion of Variance	0.71040	0.13270	0.05460	0.04057	0.02488	0.01617	0.01420	0.00647
Cumulative Proportion	0.71040	0.84310	0.89770	0.93828	0.96316	0.97933	0.99350	1.00000



De acuerdo con la siguiente visualización, la variable que se explica en menor medida por la primera componente es la PER_VPH_CEL.

