Tarea 4

**Fecha de entrega: 14 de julio de 2022, 10am.**

I - INSTRUCCIONES

Para completar esta tarea, por favor sigan los siguientes pasos.

1. **Mantengan grupos de la tarea 3.**
2. **Inserten sus respuestas en este documento de Word**.
3. **Una vez tengan todo el material, salven el documento en formato PDF.**
4. **Suban el documento de PDF y sus códigos de R a través de Módulo 7.**

II - IDENTIFICACIÓN

1. **Miembros del grupo:**

|  |  |
| --- | --- |
| Nombres y apellidos: |  |
|  |  |

1. **Cumplimiento con el código de conducta del curso**[[1]](#footnote-1) **(Marque con una X)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  | **Yes** | **No** | | Certificamos que nuestro trabajo se hizo bajo cumplimiento del Código de conducta del curso. |  |  | |

III – Datos de criminalidad y seguridad pública en Chicago (5 puntos)

Estás empezando tu programa de pasantías en el Gobierno de la Ciudad de Chicago. Tu supervisor muestra una lista de los data sets del **Portal de Datos de Chicago** que debes descargar para elaborar el reporte anual de seguridad ciudadana:

* Límites de Distritos Policiales ([Descárgalo aquí](https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Districts-current-/fthy-xz3r)).
* Crímenes en 2022 ([Descárgalo aquí](https://data.cityofchicago.org/Public-Safety/Crimes-2022/9hwr-2zxp)).
* Estaciones Policiales ([Descárgalo aquí](https://data.cityofchicago.org/Public-Safety/Police-Stations-Map/gkur-vufi)).

El pasante anterior adelantó algo del trabajo, y te dejó una lista de los tasks que faltan por completar:

1. Genere una base de datos con la tasa de arresto por tipo de crimen (variable *Primary Type*). Presente una tabla o gráfico de barras legible y detallado donde se puedan ver la cantidad de reportes, la cantidad de arrestos y la tasa de arrestos por los siguientes conceptos: *“THEFT”, “NARCOTICS”, “DECEPTIVE PRACTICE.”*
2. Genere una base de datos con la tasa de arrestos totales por Distrito Policial. Presente un mapa coloreado según la tasa de arresto (valores claros deberían corresponder a bajas tasas de arresto, y viceversa).
3. Genere una base de datos o “*data frame”* que contenga la tasa de arresto de cada uno de los delitos del primer punto (“*THEFT”, “NARCOTICS”, “DECEPTIVE PRACTICE*”) en cada uno de los Distrito Policial. Luego, junte esta base de datos con los polígonos de cada Distrito Policial y grafique la distribución de la tasa de arresto en tres mapas, uno para cada delito. *Pista: Los comandos de dplry filter() y left\_join() le pueden ser de mucha ayuda.*
4. Genere un mapa donde se aprecie la ubicación exacta de cada estación policial y el límite de cada Distrito Policial. *Pistas: Deberá usar la función st\_as\_sf() para convertir el data frame de estaciones policiales en un objeto sf que pueda graficar con el paquete tmap, para ello deberá haber separado la columna LOCATION en una columna con la Longitud y otra con la Latitud. Antes de graficar, asegúrese que todos los objetos sf tengan mismo el sistema CRS, preferiblemente este "+proj=longlat +datum=NAD83".*

IV – Juego de Tronos (5 Puntos):

Si, leíste bien. El análisis de redes es sumamente útil para aislar comunidades dentro de un sistema amplio de interconexiones, y una gran excusa para charlar sobre una de mis series favoritas. Tras un ejercicio de “*text mining”* sobre los cinco libros de Game Of Thrones, unos fanáticos lograron construir un mapa de co-apariciones de los personajes. La manera en que lo logran es contando la cantidad de veces que los nombres de dos personajes son nombrados en con menos de quince palabras de por medio. Para reforzar lo aprendido en clase le pediré que me ayude a identificar a los personajes más centrales de los libros (según criterio de co-aparición) y a segmentar a los personajes en comunidades de co-aparición. Siga los siguientes pasos:

1. Lo primero que debes hacer es acceder a [kaggle.com](https://www.kaggle.com/code/mmmarchetti/game-of-thrones-network-analysis) y descargar los datos de los cinco libros de Game Of Thrones. Una vez descargados, levanta el primer y el quinto libro en R.
2. Concéntrate primero en los datos del libro 1. Crea un objeto tipo *igraph* con el partiendo de los datos que levantaste, luego calcula la centralidad tipo *“strength”* de cada personaje (es cómo la centralidad degree, pero ponderada por el peso de cada enlace). Responda: ¿Cuántos personajes hay? ¿Cuál es el personaje más central de la red?
3. Imprime una red donde: El tamaño de cada nodo sea su centralidad “strength”, el grosor de cada enlace sea la frecuencia de co-aparición entre dos personajes y el color corresponda a su respectiva comunidad
4. Repite el proceso anterior para el libro 5. ¿Cuántos personajes hay? ¿Cuántos de los que aparecieron en el libro 1 aparecen en el 5? ¿Cuál es el personaje más central en el libro 5?

III – Todo es data: Échale un vistazo al curso

Durante el semestre, usted gentilmente nos compartía su opinión acerca de los materiales, formatos, y aspectos generales del curso, todo de forma completamente anónima. Con el objetivo de que practique sus habilidades de text-mining, subimos al M7 un csv. con las respuestas a la pregunta “¿Qué otro comentario tienes en relación con el curso y los instructores?”.

1. Simplifica el texto (remueve *stop words*, signos de puntuación, ponlo todo en mayúsculas), y después crea un *data.frame* que muestre cada palabra y la cantidad de veces que fue mencionada.
2. Usa nueva data para producir un gráfico de barras con las palabras más comunes en el archivo. Vuelve a usarla para producir una nube de palabras.

V – PROYECTO FINAL (6 puntos)

Adjunta a tu entrega un archivo .pptx con los primeros 5 slides de su presentación del proyecto (recuerden que tienen que presentar por 10 minutos en la última clase). Estos slides deben cubrir la pregunta de investigación, hechos estilizados que estimulan su pregunta, su hipótesis sobre la respuesta a la pregunta, los problemas metodológicos para establecer el efecto causal deseado entre causa y consecuencia, su estrategia empírica sugerida y los datos que podrían usarse para implementar esa estrategia.

1. Si bien invitamos la colaboración entre estudiantes, cada estudiante debe entregar respuestas individuales y originales a las asignaciones. En Módulo 7 y por el grupo de WhatsApp del curso los estudiantes podrán interactuar y hacer preguntas basadas en código, pero está prohibido compartir códigos enteros para responder a preguntas o a asignaciones por cualquier vía. Cualquier violación será tratada con la mayor severidad. [↑](#footnote-ref-1)