Características geográficas de los hallazgos de fosas clandestinas obervadas por la prensa en México: Una aproximación desde la estadística y la minería de textos.

Jorge Ruiz Reyes

22/04/2021

# Introducción

El fenómeno de la desaparición de personas y hallazgos de fosas clandestinas es uno de los problemas públicos más apremiantes en México. Sus hallazgos han sido recurrentes y extendidos en la mayoría de las entidades del país, durante un periodo de quince años.[[1]](#footnote-21) De acuerdo con cifras oficiales, actualmente existen más de 82 mil personas desaparecidas en el país. Asimismo, entre 2018 y 2020, la Comisión Nacional de Búsqueda (CNB) ha registrado un total de XXX fosas clandestinas en XXX entidades del país.

A pesar de la escala, existen tres retos alrededor del fenómeno de fosas clandestinas que deben ser atendidos para lograr implementar políticas públicas adecuadas en materia de búsqueda, exhumación e identificación de personas: 1) estimar la verdadera magnitud del problema; 2) superar los problemas de intercambio de información e identificación de cuerpos de personas que permancen sin identificar y; 3) conocer los patrones cuantitativos y cualitativos de los hallazgos de fosas que nos permitan generar nuevas búsquedas.

Este artículo presentará una aproximación que nos permita seguir generando respuestas respecto al tercer punto. Para hacerlo se realizará un ejercicio de análisis de texto cuantitativo sobre notas de prensa que han registrado hallazgos de fosas clandestinas en el periodo 2009-2019. El objetivo es identificar las caracterísiticas geográficas de los sitios descritos por las y los periodistas, y reconocer posibles diferencias y similitudes entre municipios y regiones. Con esto se pretende mostrar posibles nuevas rutas que sigan aportando al estudio de la desaparición y a la pregunta sobre dónde buscar a las personas desaparecidas.

# Análisis

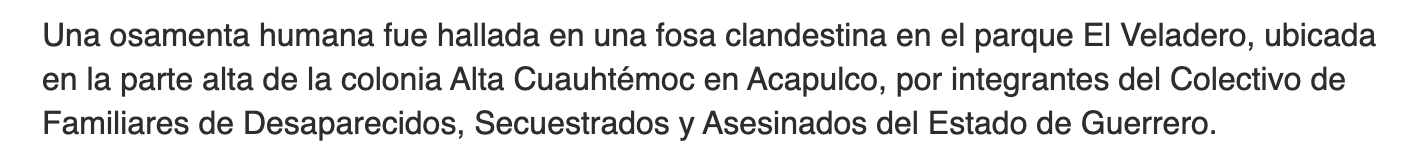
Los hallazgos de fosas clandestinas han sido registrados y observados en México por fuentes fuentes oficiales, prensa escrita nacional y local, así como por colectivos de familiares en búsqueda. Esta información cuenta con características específicas y se ve condicionada por la capacidad política, económica o geográfica que tienen los acotores para documentarla [2 BALL Y MANRIQUE]. Sabemos que existen municipios y zonas que solo son observadas por la prensa, mientras que otras zonas solo son documentadas por las autoridades o por colectivos de familiares. No solo esto, sino que la información que se produce de las fuentes es cualitativamente distinta, ya que los objetivos de su producción tiene fines distintos [3 PDH DC]. Esto significa que tenemos información dispersa y fragmentada, que responde a las capacidades que tenemos de documentar los hechos.

Dicha fragmentación o sesgos en la información son un reto, sin embargo son comunes en contextos de violencia extendida como el mexicano. Sobre todo porque, en términos estadísticos, no son muestras representativas de los hechos de la violencia [4 Ball de bigness]. A pesar de esto, estas fuentes contienen información que es valiosa porque nos permiten conocer los hechos y preparar estretategias de búsqueda y exhumación de personas.

En México han exisitido diversos estudios que han documentado los hallazgos con miras a aproximarse a la magnitud de los hechos. Es decir, a responder cuántas fosas clandestinas se han localizado y cuántos cuerpos de personas se han exhumado en estados y municipios. [5,6 PDH Y QUINTO ELEMENTO] Sin embargo, estas fuentes contienen datos que no han sido aprovechados y que podrían ayudarnos a responder preguntas como ¿cuáles son las características geográficas de los sitios donde se localizan fosas clandestinas?, ¿existen patrones específicos con base en los municipios y estados del país donde se localizan?

Las preguntas anteriores pueden responderse a través desde diversas aproximaciones. Por ejemplo, analizando fichas de documentación de los sitios de hallazgo generadas por colectivos de familiares, equipos forenses independientes o autoridades. El problema es que esta información puede ser limitada a ciertas zonas, encontrarse en diferentes formatos o inclusive puede ser negada por parte de las autoridades. A pesar de esto, fuentes como la prensa escrita también pueden darnos este tipo de descripiciones, como se puede observar en la siguiente imagen.

## Imagen 1



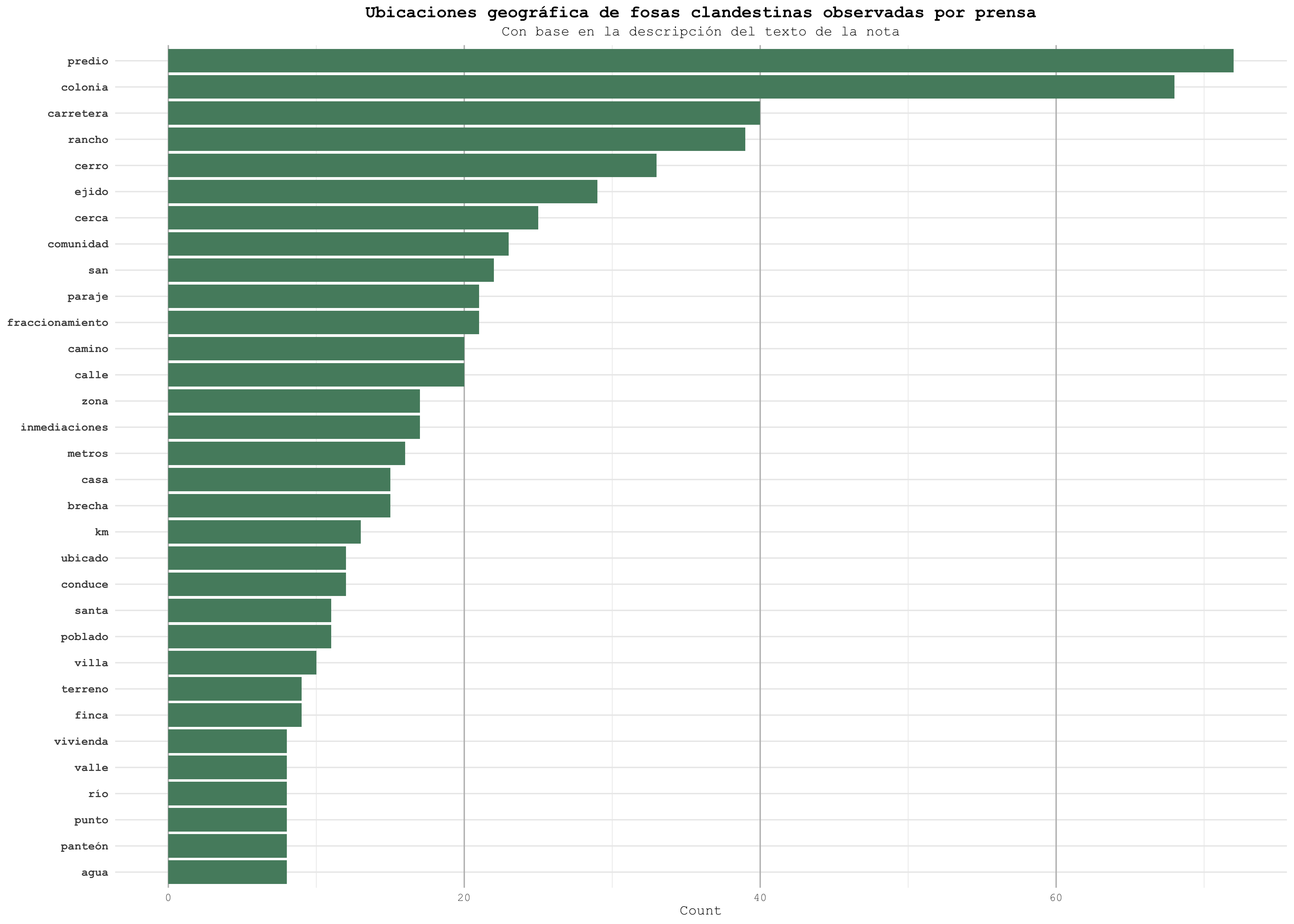
La **Imagen 1** recupera la descripción de una fosa clandestinas localizada en Acapulco, Guerrero, en noviembre de 2020 [<https://www.elsoldeacapulco.com.mx/policiaca/localizan-osamenta-humana-en-fosa-clandestina-en-el-parque-veladero-acapulco-estado-noticia-guerrero-violencia-6027154.html>]. Como se puede identificar, este fragmento no solo nos da información de cuántas fosas y osamentas se localizaron, sino que también nos da detalles geográficos, como el parque “El Veladero” y la colonia “Alta Cuauhtémoc”. Este tipos de datos son comunes en las descripciones de las fosas que se han hecho por la prensa desde el año 2007 en el país.

Si tomamos un número mayor de notas y extraemos el texto de éstas, entonces se pueden generar análisis cuantitativos de texto que nos permitan conocer las características de estos sitios en las diferentes regiones del país. Esta aproximación metodológica es conocida como *minería de textos*, la cual integra técnicas provenientes de la estadística, el aprendizaje de máquina (machine learning) y la lingüistica computacional para descrubir patrones de datos no estructurdos en forma de texto [<https://www.sciencedirect.com/science/article/pii/B9780123814791000137>].

Utilizando la base de datos de fosas clandestinas generada por el Programa de Derechos Humanos de la Universidad Iberoamericana Ciudad de México (PDH Ibero), se tomaron las descripciones geográficas de 639 sitios que reportaron un total de 1491 en 320 municipios del país entre 2007 y 2019. Cada uno de los sitios tuvo por lo menos un reporte prensa, sin embargo, existen hallazgos donde podían existir más de diez reportes, dada la atención mediática de los hechos.[[2]](#footnote-27) Lo anterior permitió tener una cantidad suficiente de texto con la que se exploró de manera preliminar, la posibilidad de extraer información cualitativa sobre las caracerísticas de las fosas que son observadas por la prensa.

Si tomamos en cuenta los 320 municipios donde se han localizado fosas reportadas por la prensa, poder ver que las zonas geográficas se componen por las siguientes características:

## Figura 1



La **figura 1** muestra las palabras (ubicaciones) que se mencionan mayor frecuencia por las notas de prensa que reportaron hallazgos de fosas entre los años 2007-2019. Los predios (72 menciones), colonias (68 menciones), carretera (40 menciones), rancho (39 menciones) y cerro (33 menciones) son las cinco palabras que se retoman con mayor frecuencia dentro de las descripciones de las y los periodistas que redecataron las notas. Sin embargo, la figura anterior también nos permite reconocer que los hallazgos de fosas se dan tanto en zonas rurales, como en zonas urbanas de los municipios de México.

Lo anterior indica las dificultades y las diferentes estrategias que deben ser planeadas para los procesos de exhumación y recuperación de restos de personas. Sobre todo porque no existe un solo sitio característico para iniciar una exhumación, ya que ésta puede ocurrir en una brecha o cuerpo de agua en una zona rural, así como un fraccionamiento en una ciudad con una densidad poblacional alta.

Consecuentemente, con la minería de textos podemos encontrar otros patrones que nos permitan conocer si hay sitios característicos o “distintivos” donde se localizan fosas clandestinas en estados del país, en comparación con otros.

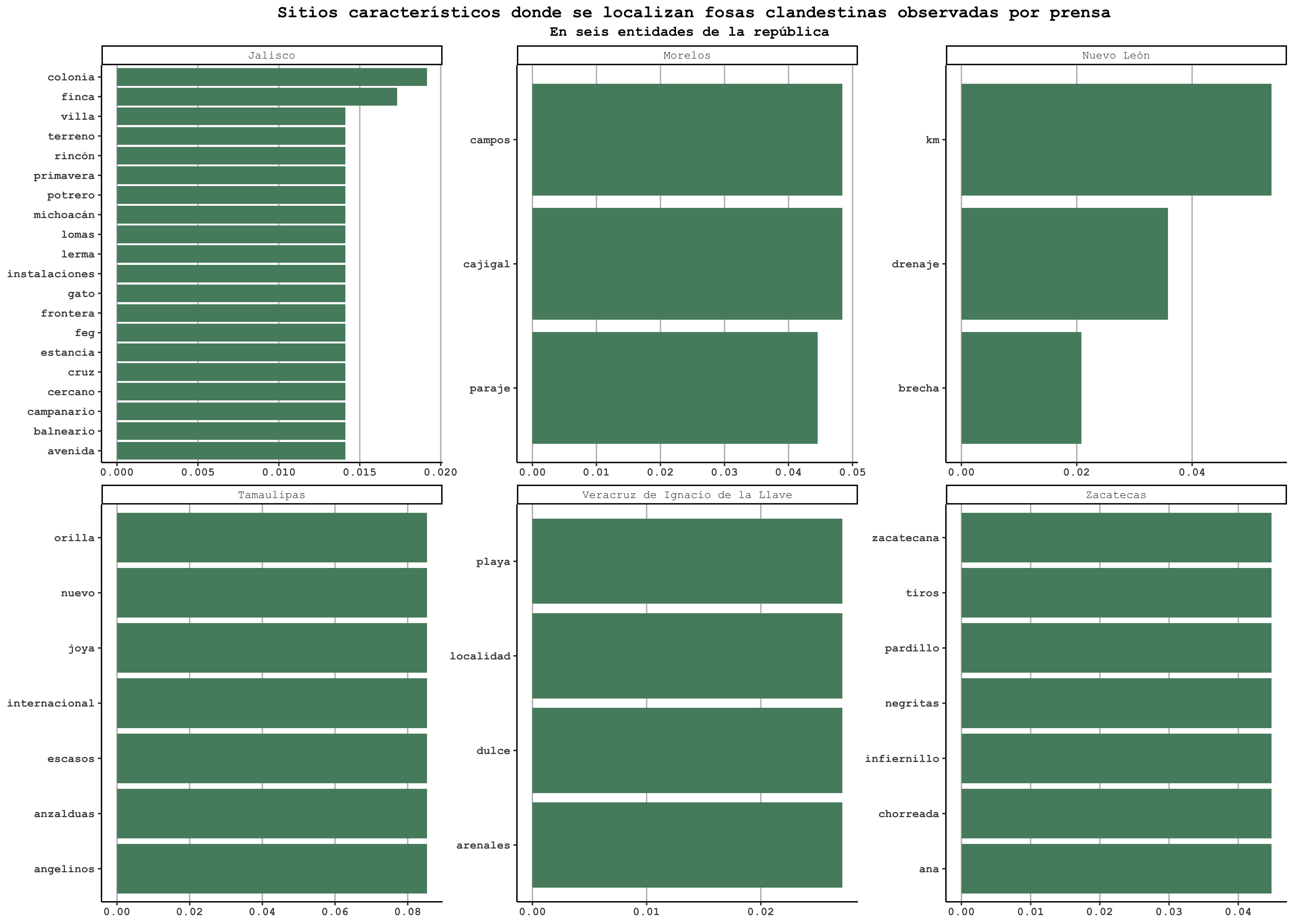
Calculando una medida estadística conocida como *Frecuencia de término - frecuecia inversa de término*(TF-IDF por sus siglas en inglés), podemos conocer la relevancia que tiene una palabra para un subconjunto de textos, con base una colección de textos mayor. Específicamente, se multiplican dos métricas: 1) la frecuencia de término donde se identifica la frecuencia de la palabra en el subconjunto y 2) la frecuencia inversa de documento donde se divide el número total de documentos en la colección, por el número de subconjuntos que contienen el término, y tomando el logarítimico. La fórmula puede verse como:

donde:

y donde:

En nuestro caso, agruparemos el texto de las notas con base a los estados del país, donde cada descripción geográfica en un estado es un subcojunto de las descripciones geográficas de todos los estaos (D). Posteriormente, analizamos si una palabra se menciona con frecuencia en un determinado estado (t) o si no se menciona (d). De esta forma, mientras más frecuente sea una palabra en un estado, pero no en otro, podemos establecer que es una palabra, o en nuestro caso, un sitio distintivo donde se localizan fosas en comparación con otros estados.

Los resultados de esta medida pueden observarse a continuación:



En la **Figura 2** podemos ver cuáles son los sitios característicos donde se localizan fosas clandestinas observadas por la prensa en México, tomando seis entidades como ejemplo.

1. 1 FOSAS VIOLENCIA [↑](#footnote-ref-21)
2. El análisis se realizó utilizando el lenguaje de programación R. Los datos y el código pueden consultarse en este repositorio: <https://github.com/JorgeRuRe/fosas-GIZ-CICR> [↑](#footnote-ref-27)