



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

---

**FACULTAD DE INGENIERÍA**

**Sistemas de Información Geográfica  
en Lenguaje R aplicados a la base de  
datos de la ENCODAT 2016 - 2017.**

**TESIS**

Que para obtener el título de  
**Ingeniero Geomático**

**P R E S E N T A**

Eduardo Díaz Cruz

**DIRECTOR(A) DE TESIS**

Ing. Ana Lilia Salas Alvarado



Ciudad Universitaria, Cd. Mx., 2022



“No midas el trabajo hasta que el día termine  
y la labor esté hecha.”

Elizabeth Barrett Browning



## Índice

Introducción .....	8
Objetivos.....	9
Objetivo general.....	9
Objetivos particulares.....	9
Capítulo 1. Ingeniería Geomática .....	10
1.1 De Ingeniería Topográfica y Geodésica a Ingeniería Geomática. ....	10
1.2 Disciplinas que componen la Ingeniería Geomática. ....	11
1.2.1 Topografía.....	11
1.2.2 Fotogrametría.....	11
1.2.3 Geodesia.....	12
1.2.4 Cartografía. ....	12
1.2.5 Sistema Global de Navegación por Satélite (GNSS).....	12
1.2.6 Percepción Remota.....	12
1.2.7 Sistemas de Información Geográfica (SIG). ....	13
Capítulo 2. Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco .....	14
2.1 Encuesta Nacional de Adicciones (ENA) 2008 y 2011.....	15
2.1.1 Encuesta Nacional de Adicciones, ENA 2008.....	16
2.1.2 Encuesta Nacional de Adicciones, ENA 2011.....	18
2.2. Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017. .....	19
Capítulo 3. Vínculo de la Información con Lenguaje R.....	28
3.1 Base de datos ENCODAT 2016 – 2017, manipulación con lenguaje R.....	29
3.1.1 Clasificación de los datos recopilados en la encuesta.....	30
3.1.2 Administración geoespacial. ....	33
Capítulo 4. Filtro de datos correspondientes a la ENCODAT 2016 – 2017 y métodos de presentación gráfica. ....	36

4.1 Composición de la base de datos georreferenciados. ....	36
4.2 Filtro lógico.....	42
4.2.1 Tidyverse. ....	43
4.3 Bibliotecas de manipulación vectorial, Simple Features (sf). ....	46
4.4 Métodos de presentación en R.....	47
4.4.1 Rayshader. ....	49
Capítulo 5. Análisis conjunto de los resultados filtrados de la encuesta.....	50
5.1 El mapeo del uso de sustancias ilegales.....	50
5.2 Distribución de los casos a Nivel Nacional.....	51
Capítulo 6. Percepción y prevención del riesgo ante sustancias ilegales en la principal zona metropolitana del país. El nivel educativo como factor.....	70
6.1 Inclusión en programas de prevención en la ZMVM.....	72
6.1.1 Variable ds9 – Grado de estudios.....	73
6.2 Percepción de riesgo.....	75
6.2.1 Variable ts18 – Marihuana. ....	76
6.2.2 Variable ts23 – Tabaco. ....	79
6.2.3 Variable ts24 – Alcohol. ....	82
Capítulo 7. Conclusiones.....	85
Anexos.....	87
Código – Casos de consumo de sustancias a nivel nacional. ....	87
Código – Cartografía 3D .....	91
Código – Niveles de percepción.....	93
Código – Zonas metropolitanas.....	97
Mapa 1, Casos de Consumo de Marihuana en Hombres a Nivel Nacional .....	99
Mapa 2, Casos de Consumo de Cocaína en Hombres a Nivel Nacional .....	100
Mapa 3, Casos de Consumo de Crack en Hombres a Nivel Nacional .....	101
Mapa 4, Casos de Consumo de Alucinógenos en Hombres a Nivel Nacional .....	102
Mapa 5, Casos de Consumo de Inhalables en Hombres a Nivel Nacional .....	103
Mapa 6, Casos de Consumo de Heroína en Hombres a Nivel Nacional .....	104

Mapa 7, Casos de Consumo de Anfetaminas en Hombres a Nivel Nacional .....	105
Mapa 8, Casos de Consumo de Marihuana en Mujeres a Nivel Nacional .....	106
Mapa 9, Casos de Consumo de Cocaína en Mujeres a Nivel Nacional .....	107
Mapa 10, Casos de Consumo de Crack en Mujeres a Nivel Nacional .....	108
Mapa 11, Casos de Consumo de Alucinógenos en Mujeres a Nivel Nacional .....	109
Mapa 12, Casos de Consumo de Inhalables en Mujeres a Nivel Nacional .....	110
Mapa 13, Casos de Consumo de Heroína en Mujeres a Nivel Nacional .....	111
Mapa 14, Casos de Consumo de Anfetaminas en Mujeres a Nivel Nacional .....	112
Referencias.....	113

## Introducción

En México, desde el año 1988 y hasta el año 2011 se aplicó de forma constante la Encuesta Nacional de Adicciones (ENA) para conocer el panorama general del uso de sustancias legales e ilegales a nivel estatal, en el año 2016 y con la participación del Instituto Nacional de Psiquiatría Ramón De la Fuente Muñiz, la Secretaría de Salud y el Instituto Nacional de Salud Pública se llevó a cabo la Encuesta Nacional de Consumo de Drogas Alcohol y Tabaco 2016 – 2017 (ENCODAT) con el fin de conocer a nivel localidad el comportamiento del uso de sustancias. La ENCODAT 2016 – 2017 recopiló en dos cuestionarios, hogar e individual, los datos de consumo de alrededor de 1500 personas por entidad federativa distribuidas a lo largo de las localidades de cada estado para contener un total aproximado de 56 000 observaciones a nivel nacional donde se recolectaron datos referentes al consumo de alcohol, tabaco, marihuana, cocaína, crack, inhalables, etc.

La ENCODAT 2016 – 2017 a diferencia de las anteriores ENA's registró en la base de datos la ubicación a nivel localidad de cada individuo entrevistado, lo que permite tener a una escala territorial menor una concentración de datos capaz de ser moldeada con fines cartográficos, lo cual representa un área de oportunidad para el aporte de información. Este tipo de instrumentos implican la recopilación de una gran cantidad de datos, que al ser tan grande el análisis numérico sumado a una representación cartográfica puede resultar muy complejo, sin embargo, la evolución tecnológica en lo que al desarrollo de software se refiere ha presentado un avance importante, al mismo tiempo que los lenguajes de bajo nivel orientado a objetos, como el lenguaje R, han incorporado aplicativos para el análisis de información con una referencia geoespacial.

Según lo mencionado, en el presente trabajo se realizó el tratamiento (limpieza, filtrado, estadísticas, etc.) de la base datos de la ENCODAT 2016 – 2017 mediante el lenguaje de programación R, también se elaboraron representaciones cartográficas usando las bibliotecas GDAL y OGR, que contienen funciones de análisis espacial también integradas en los softwares SIG más distribuidos. La generación de la cartografía se desarrolló tomando información vectorial ya existente de los Estados Unidos Mexicanos comprendida en el Marco Geoestadístico del INEGI, información de la delimitación de las zonas metropolitanas de México del sitio web de la CONAPO y la ubicación puntual de las localidades a lo largo del territorio de la plataforma de CONABIO, de forma que este estudio permitirá conocer la relación cada vez más estrecha entre los sistemas de información geográfica y los lenguajes de programación.

Así mismo, este trabajo contribuye en demostrar la amplitud de las aplicaciones que la Ingeniería Geomática oferta a diferentes disciplinas, que en este caso particular se refiere al tratamiento y representación cartográfica de datos sobre el consumo de sustancias, lo que complementa a las investigaciones desarrolladas por los organismos de salud.

## Objetivos

### Objetivo general.

Utilizar el lenguaje de programación R para implementar un sistema de información geográfica con el cual sea posible extraer y manipular los datos contenidos en la base de datos de la ENCODAT 2016-2017 para generar cartografía y estadísticas.

### Objetivos particulares.

Obtener los casos puntuales de consumo por sustancia (alcohol, tabaco, crack, etc.) a nivel nacional y en la principal zona metropolitana del país relacionando la media de edad del inicio de consumo con el desarrollo y distribución de los casos.

Determinar los espacios clave donde la población que ha consumido alguna sustancia alguna vez ha estado expuesto a pláticas de prevención.

Analizar la impresión y las diferencias en la percepción de riesgo que han dejado las campañas de prevención en grupos clasificados por nivel educativo incluyendo en la perspectiva a las personas que concluyeron sus estudios o aquellas que continúan estudiando.

Mostrar la versatilidad de aplicaciones que tiene la Ingeniería Geomática en distintas especialidades para el diagnóstico de problemáticas.

## Capítulo 1. Ingeniería Geomática

### 1.1 De Ingeniería Topográfica y Geodésica a Ingeniería Geomática.

La Universidad Nacional Autónoma de México imparte la licenciatura en Ingeniería Geomática desde el año 2005 siendo Ingeniería Topográfica y Geodésica el precursor de la carrera con su entrada en vigor en el año 1994 y la creación del plan de estudios original en 1947.

La labor de un ingeniero topógrafo es vista como una tarea preliminar en cualquier obra de infraestructura, un trabajo de planeación y supervisión durante la ejecución de la misma y la elaboración de los resultados gráficos en cada etapa de un proyecto civil. Como perfil un ingeniero topógrafo hace uso de las disciplinas fundamentales de la Ingeniería Geomática para poder desarrollar trabajos de cartografía, levantamientos topográficos e inspeccionar el comportamiento de una obra, estructura o vía terrestre durante y tiempo después de que se ha dado por terminada su etapa constructiva además de brindar la supervisión necesaria tras un desastre natural o cualquier tipo de siniestro; el Ingeniero Geomático además de compartir las labores de un Ingeniero Topógrafo hace uso de nuevas tecnologías en equipo, software e información para presentar proyectos cartográficos más robustos además de incorporar estas herramientas en temas sociales, económicos, de seguridad e incluso de entretenimiento haciendo a la Ingeniería Geomática una labor interdisciplinaria.

En la búsqueda de agilizar tareas y priorizar tiempos la topografía se ha mantenido a la vanguardia valiéndose de herramientas y disciplinas especializadas para poder aterrizar sobre la superficie terrestre los planes de construcción de las obras más ambiciosas, ejemplo de esto es la intervención de la fotogrametría que tras su propio proceso de planeación y ejecución entrega la posibilidad de adelantar un trabajo topográfico por días o semanas, identificar zonas de riesgo para la población o mapear extensas áreas devastadas por la naturaleza.

Con el paso del tiempo y la entrada de nuevas tecnologías la Ingeniería Topográfica ha evolucionado y forma parte fundamental de lo que ahora conocemos como Ingeniería Geomática. El término fue acuñado en la década de los ochenta por la introducción de la computación y nuevas tecnologías que también estaban impulsando a la topografía; se define como Geomática al sistema multidisciplinario para seleccionar técnicas e instrumentos para reunir, almacenar, integrar, modelar, transformar y distribuir información georreferenciada a voluntad. (Gomasasca, 2009)

Con el avance de las tecnologías y la constante actualización de los medios de captura de información en disciplinas como la Percepción Remota, los Sistemas de Posicionamiento

Global y los Sistemas de Información Geográfica se han diversificado las aplicaciones y con esto se ha advertido la profunda necesidad de contar con los datos estadísticos georreferenciados de infinidad de acontecimientos sociales, económicos, políticos, de seguridad o salud con el fin de que cada disciplina o agencia pueda hacer uso de esta información para plantear una estrategia de acción o tener el panorama completo de las variables que influyen en el comportamiento del entorno.

Se puede decir entonces que la Ingeniería Geomática recopila los datos en campo o a distancia y aplica procesos que complementan los análisis que pertenecen a diversas disciplinas para sintetizar en su presentación cartográfica lo que también resultan ser las características de la superficie terrestre, la obra y el comportamiento humano.

## 1.2 Disciplinas que componen la Ingeniería Geomática.

### 1.2.1 Topografía.

La topografía (topos – lugar y grafos – descripción) se ha entendido, de manera histórica, como el arte de representar la superficie de la tierra; ubicar y determinar gráficamente todos los elementos (naturales y artificiales) que componen el medio. (Antonio G. M., 1994)

Se utiliza topografía para determinar la altitud de relieves, la graduación de una pendiente, el cálculo de volúmenes, la distancia entre dos puntos con o sin visibilidad entre ellos y en general para el levantamiento de sitios de interés (terreno natural, naves industriales, terracerías, minas etc.) como trabajo preliminar de obra civil, arqueología y de diversas disciplinas.

### 1.2.2 Fotogrametría.

La fotogrametría como herramienta de la Ingeniería Geomática se ha utilizado para facilitar y agilizar trabajos de análisis territorial a gran escala; su significado etimológico se deriva de los vocablos photós – luz, gramma - trazo y metrón – medir, como “medir sobre fotos” según la Sociedad Americana de Fotogrametría y Teledetección (ASPRS).

Los métodos fotogramétricos se distinguen por el uso de fotografías con características especializadas para determinar las dimensiones y particularidades del terreno capturado, una de las cualidades de la fotogrametría es permitir el levantamiento de extensas áreas, muchas veces de difícil acceso para el trabajo en campo, en un periodo de tiempo excepcionalmente corto en comparación con un levantamiento topográfico.

### 1.2.3 Geodesia.

La palabra geodesia tiene origen en las etimologías griegas geo – tierra y daiein – dividir las cuales complementan su significado como “división de la tierra”.

La esencia de la geodesia es el análisis y determinación de la forma, dimensiones de la tierra y su campo de gravedad. Los estudios de la geodesia han servido para establecer la orientación de la tierra en el espacio y definir los sistemas de referencia terrestre para la localización de puntos sobre la superficie. (Sevilla de Lerma, Historias de Matemáticas, Introducción Histórica a la Geodesia, 2012)

### 1.2.4 Cartografía.

Se puede definir a la cartografía (khartes – mapa y graphein – escribir) como la representación de la superficie terrestre basada en reglas específicas (Gomasasca, 2009). Según la Asociación Cartográfica Internacional se considera que la cartografía es el arte y el uso de tecnologías en la elaboración de mapas y el estudio de esos como documentos científicos.

### 1.2.5 Sistema Global de Navegación por Satélite (GNSS)

El acrónimo GNSS (Global Navigation Satellite System) es el término con el que se denominan los distintos sistemas de navegación por satélite que proporcionan el dato preciso de coordenadas geográficas de un punto en cualquier parte del mundo en cualquier momento. (Berné Valero, Anquela Julián, & Garrido Villén, 2014)

Ejemplos de estos sistemas son: GPS (EUA), GLONASS (Rusia), BEIDOU (China), GALILEO (UE); siendo los dos primeros los únicos con una cobertura global operativa.

### 1.2.6 Percepción Remota.

La percepción remota fundamenta su estudio en la manipulación de la información satelital capturada por sensores remotos; de un modo parecido a la fotogrametría, pero a mayor escala.

La percepción remota hace uso de plataformas satelitales para obtener imágenes de la superficie terrestre con la diferencia de que los datos recabados se analizan dentro del rango de longitudes de onda en el del espectro electromagnético debido a que cada elemento

natural o artificial que compone el terreno absorbe y emite parte de la radiación incidente sobre él en diferente proporción, lo cual depende de sus características estructurales, químicas y cromáticas (Gomasca, 2009).

### 1.2.7 Sistemas de Información Geográfica (SIG).

Un sistema de información geográfica es una herramienta que permite el análisis y manipulación de información georreferenciada a través de coordenadas espaciales o geográficas (Star & Estes, 1990).

Un SIG, incluye tanto elementos de información digital como física, hardware y software dedicados que con el análisis debido permite contener, relacionar, manipular y generar de manera gráfica la representación de cualquier tipo de variable levantada en campo a la que se le ha asignado una georreferencia, que es la característica clave en un análisis de Ingeniería Geomática.

## Capítulo 2. Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco

En México se han realizado investigaciones epidemiológicas acerca del consumo de drogas desde la década de los setenta utilizando una metodología uniforme para cada tipo de población estudiada lo que ha permitido forjar un conocimiento global del consumo en distintos escenarios como el hogar, escuela y los centros de tratamiento (Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; Instituto Nacional de Salud Pública; Secretaría de Salud., 2012).

Como parte del fomento a la prevención del consumo de drogas la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC) estableció en 1987 el día 26 de junio de cada año como el "Día internacional de la lucha contra el uso indebido y el tráfico ilícito de drogas", además de alentar a los países a implementar campañas de prevención basadas en evidencia científica, en la salud pública y sensible a los derechos humanos. A su vez, como parte de las campañas de prevención se destacan, el día 15 de noviembre "Día Nacional contra el Uso Nocivo del Alcohol"; el día 31 de mayo "Día Mundial Sin Tabaco".

La Organización Mundial de la Salud define a las drogas como cualquier sustancia psicoactiva que puede modificar la percepción, estado de ánimo, cognición, conducta o funciones motoras del organismo (OMS, 1969); independientemente existen términos como droga de abuso que hace referencia a una sustancia de uso no médico con efectos psicoactivos y de sencillo autoadministración que hace que la definición de la OMS resulte un tanto imprecisa ya que esta incluye fármacos de prescripción, plantas, hongos y otras sustancias químicas, de manera crítica la diferencia entre una droga y un fármaco está matizada por tintes instrumentales y sociales, de este modo algunas sustancias pueden ser consideradas drogas o fármacos según el contexto, por ejemplo: los esteroides son un fármaco de prescripción pero que de ser usados en gimnasios con el fin de mejorar el rendimiento físico se considerarían drogas (Fernando, 2016).

La clasificación de las drogas en estatus legales e ilegales es un tema de controversia a nivel internacional, al que aluden restricciones con peso en temas religiosos, culturales, de seguridad y derecho; en México a nivel nacional las drogas consideradas legales son: el alcohol, tabaco y los inhalables (Río Arronte, Fundación, 2020).

## 2.1 Encuesta Nacional de Adicciones (ENA) 2008 y 2011.

La Encuesta Nacional de Adicciones (ENA) es un instrumento de análisis estadístico que se ha aplicado desde el año 1988 con el fin de generar estrategias y definir políticas para afrontar el problema del consumo de sustancias adictivas.

Antes de la ENA 2008 se realizaron cinco encuestas nacionales, 1988, 1993, 1998 y 2002 de las cuales las primeras tres contienen solo datos urbanos, no es sino hasta la encuesta 2002 que se incluyeron por primera vez datos rurales, los cuatro sondeos anteriores tienen representatividad nacional con los cuales se desglosaron datos regionales, norte sur y centro.

La Encuesta Nacional de Adicciones ENA 2008 consideró el estudio de 50 688 viviendas de todo el país y la entrevista se realizó directamente en el hogar a un adulto entre 18 y 65 años, así como a un adolescente de entre 12 y 17 años; por otra parte, el muestreo de la ENA 2011 se llevó a cabo en 17 500 hogares seleccionados de forma aleatoria, dividiendo el país en 8 regiones, tomando una muestra de 1 994 personas por región y repitiendo los rangos de edad de la ENA 2008.

Cabe resaltar que la ENA 2008 así como la ENA 2011 cuentan con representación urbano-rural lo cual permite representar información a nivel estatal.

REGIÓN	ESTADOS
Norcentral	Coahuila, Chihuahua, Durango
Noroccidental	Baja California, Baja California Sur, Sonora, Sinaloa
Nororiental	Nuevo León, Tamaulipas, San Luis Potosí
Occidental	Zacatecas, Aguascalientes, Jalisco, Colima y Nayarit
Centro	Puebla, Tlaxcala, Morelos, Estado de México, Hidalgo, Querétaro, Guanajuato
Ciudad de México	Distrito Federal
Centro Sur	Veracruz, Oaxaca, Guerrero, Michoacán,
Sur	Yucatán, Quintana Roo, Campeche, Chiapas, Tabasco

Tabla 1: Integración estatal de cada una de las ocho regiones. Fuente: ENA 2011.

### 2.1.1 Encuesta Nacional de Adicciones, ENA 2008.

Metodología de recolección de los datos.

La Encuesta Nacional de Adicciones ENA 2008 que tiene representatividad nacional y estatal consideró como localidad rural a aquellas con un número de habitantes menor o igual a 2500 y como unidad urbana a aquellas con más de 2500 habitantes; la selección de las muestras se realizó en distintas escalas: áreas geoestadísticas básicas (AGEB), manzanas o segmentos y viviendas; mientras que las unidades de análisis consideradas fueron hogar, adolescente y adulto.

Como antecedente de las encuestas realizadas entre 1988 y 2002 se encontraron diferencias importantes entre los indicadores de interés acordes con el tamaño de la localidad por lo que para la ENA 2008 se decidió estratificar la muestra en tres sectores diferentes:

Estrato	Descripción
I. Ciudad o área metropolitana	<ul style="list-style-type: none"><li>• Ciudades y áreas metropolitanas</li><li>• Ciudades de 100 000 y más habitantes y/o capitales de estado</li></ul>
II. Complemento urbano	<ul style="list-style-type: none"><li>• Localidades de 2 500 a 99 999 habitantes</li></ul>
III. Rural	<ul style="list-style-type: none"><li>• Localidades con menos de 2 500 habitantes</li></ul>

Tabla 2: Estratificación por urbanidad. Fuente: ENA 2008.

Considerado en el esquema de selección un AGEb, Área Geoestadística Básica es la extensión territorial que corresponde a la subdivisión de las Áreas Geoestadística Municipales (AGEM) y dependiendo de las características poblacionales se clasifican en rurales o urbanos (INEGI, 2010).

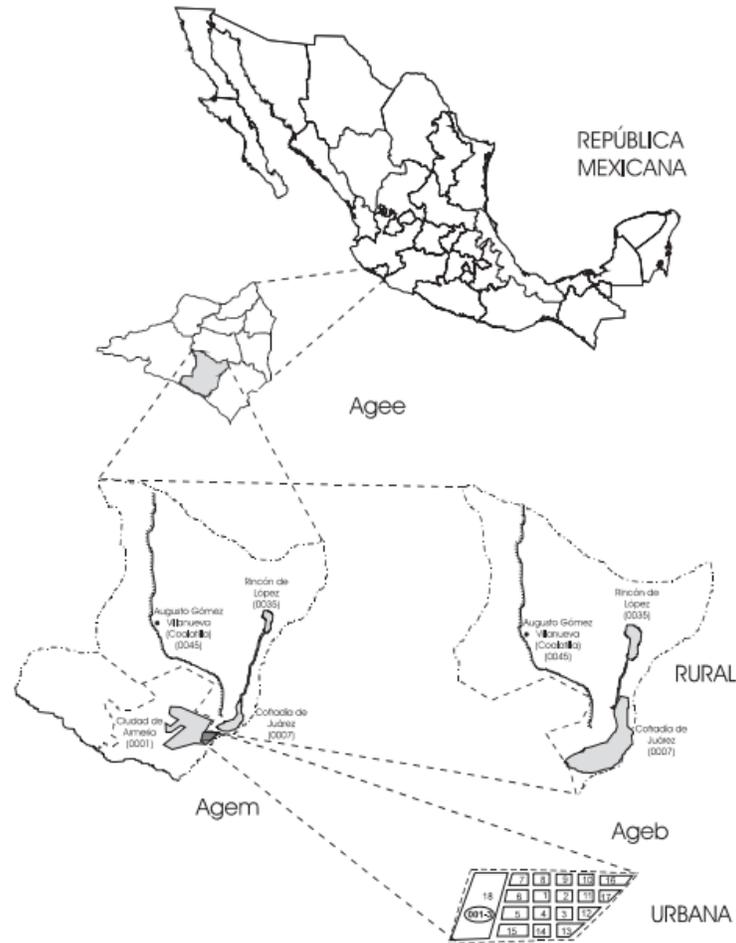


Figura 1: Conformación del Marco Geoestadístico Nacional. Fuente: Manual de Cartografía Estadística, INEGI 2010.

El esquema de selección de la ENA 2008 eligió para los estratos I y II (urbano) un número constante de AGEB según la cantidad de viviendas particulares, de estos AGEB se definieron 6 manzanas igualmente proporcionales al número de viviendas y de manera aleatoria se eligieron 6 viviendas de cada manzana y por cada domicilio se nominó un adulto y un adolescente al que se les aplicó el cuestionario correspondiente. En el estrato III (rural) la probabilidad de seleccionar un AGEB fue proporcional al número de viviendas particulares habitadas en él, de este modo por cada localidad o grupo de localidades, contenidas en el AGEB, se usó una muestra de tres grupos de doce viviendas llamados segmentos.

La logística de campo consistió en dos fases, la primera (tamizaje) consistió en la elaboración de croquis, listado de las viviendas seleccionadas, la aplicación de la encuesta hogar, prueba piloto y posteriormente la selección aleatoria de los adultos y adolescentes; en la segunda fase se aplicó la encuesta a los residentes preseleccionados.

El cuestionario de la ENA 2008 recabó información con respecto a los siguientes temas:

- Datos sociodemográficos
- Consumo de tabaco, alcohol, drogas.
- Dependencia y abuso de drogas, alcohol.
- Exposición al uso de drogas.
- Consumo de drogas en el entorno familiar.
- Percepción acerca de las drogas y de los usuarios.
- Exposición a la prevención de drogas.
- Tratamiento por consumo de drogas.

A su vez el cuestionario incluyó evaluaciones sobre aspectos de salud mental, sintomatología depresiva, intento suicida, violencia de género, conducta antisocial, migración y la relación del consumo de drogas y el comportamiento sexual con algunas enfermedades como hepatitis o VIH/SIDA.

### 2.1.2 Encuesta Nacional de Adicciones, ENA 2011.

Metodología de recolección de los datos.

La metodología de recolección de datos de la ENA 2011 en comparativa con la metodología aplicada en la ENA 2008 mantuvo tres rubros en la estratificación por urbanidad al mismo tiempo que agregó una estratificación para AGEB's.

Estrato	Definición
Rural	Localidades con menos de 2500 habitantes.
Urbano	Localidades no incluidas en los estratos rural y metropolitano Capitales de los estados
Metropolitano	Localidades con más de 100 mil habitantes. Localidades con 2500 o más habitantes de los municipios que formaban las áreas metropolitanas en el año 2000.

Tabla 3: Estratificación por urbanidad. Fuente: ENA 2011.

Siguiendo el mismo método de la ENA 2008 la Encuesta Nacional de Adicciones 2011 tomó para el estrato urbano y metropolitano una muestra de seis domicilios en seis manzanas seleccionadas de manera aleatoria dentro de cada AGEB; en el caso del estrato rural se seleccionaron 3 localidades dependiendo de su tamaño y se construyeron pseudo-

manzanas integradas por alrededor de cincuenta viviendas de las cuales se seleccionó de manera aleatoria 12 viviendas.

El cuestionario individual se aplicó a la persona seleccionada aleatoriamente, dentro del grupo de edad 12 – 17 años para adolescentes y 18 – 65 años para adultos recabando información en los rubros:

- Datos sociodemográficos
- Consumo de tabaco
- Tolerancia social
- Percepción de riesgo
- Consumo y dependencia de drogas (médicas e ilegales)
- Consumo y dependencia de alcohol
- Consumo de alcohol y drogas en periodos establecidos
- Tratamiento por consumo de alcohol o drogas
- Escala de salud mental K6
- Embarazo
- Violencia por parte de la pareja
- Conducta antisocial, inseguridad y delincuencia
- Percepción social en relación al uso de drogas
- Conocimientos sobre VIH/SIDA
- Migración hacia Estados Unidos

En ambas encuestas la coordinación estuvo a cargo de la Secretaría de Salud a través de la Subsecretaría de Prevención y Promoción de la Salud, la Comisión Nacional Contra las Adicciones (CONADIC), del Centro Nacional para la Prevención y el Control de las Adicciones (CENADIC), la participación del Instituto Nacional de Psiquiatría “Dr. Ramón de la Fuente Muñiz” y del Instituto Nacional de Salud Pública.

## 2.2. Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017.

Así como la Encuesta Nacional de Adicciones 2008, 2011 y las anteriores ENA's 1988, 1993, 1998 y 2002 la ENCODAT 2016 – 2017 permite conocer la prevalencia del consumo en la población de 12 a 65 años, la información de la encuesta pertenece a 55 907 hogares a lo largo del país con una muestra aproximada de 1600 individuos por estado: La ENCODAT 2016 – 2017 presenta los datos recabados de población que no usa ni ha usado drogas,

sustancias de uso médico con potencial psicoactivo usadas fuera de prescripción, sustancias ilegales, inhalables y las variaciones del uso por grupos de población.

El propósito planteado por la ENCODAT 2016 – 2017 fue estimar a nivel nacional y entidad federativa el consumo de drogas médicas y no médicas, evaluar las tendencias en el consumo e identificar los grupos de población con mayor riesgo a presentar problemas debido al consumo de drogas; las unidades de análisis se mantuvieron constantes a las anteriores ENA clasificando a los adolescentes en el grupo de edad 12 – 17 años, a los adultos en el rango de 18 – 65 años y la variable hogar como el conjunto de personas relacionadas por algún parentesco o no, que habitualmente duermen en una misma vivienda y viven bajo un ingreso común.

De acuerdo con los resultados del Censo de Población y Vivienda 2010 la estratificación de localidades según urbanidad se mantuvo invariable, definiendo:

#### Estrato I. Ciudad o área metropolitana

- Capital estatal
- Localidades con más de 100 mil habitantes.
- Localidades con 2500 habitantes de los municipios conformantes del área metropolitana.

#### Estrato II. Complemento urbano

- Localidades no incluidas en los estratos rural y metropolitano.

#### Estrato III. Rural

- Localidades con menos de 2 500 habitantes.

La Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco en el cuestionario individual recopiló información de índole:

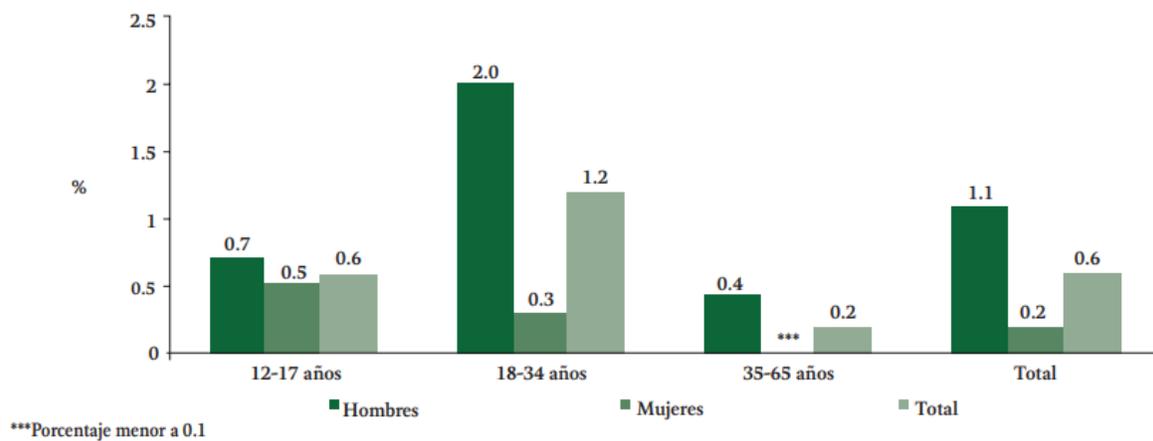
- Sociodemográfica
- Tabaco
- Exposición a Drogas
- Consumo de drogas medicas o ilegales
- Inyección de drogas
- Dependencia y dificultades relacionadas al consumo de drogas
- Alcohol y dificultades relacionadas
- Prevención hacia el consumo
- Malestar emocional
- Igualdad de género
- Intento suicida
- Percepción de drogas

- Migración internacional
- Uso de internet, entre otras.

### Resultados principales, ENCODAT 2016 – 2017.

Con los datos recolectados en el cuestionario de la ENCODAT 2016 – 2017 se establecen factores que influyen en el consumo de sustancias tales como la exposición a la prevención del consumo de drogas, la percepción del riesgo y la tolerancia social ante el consumo, variables de las cuales se pueden estimar las prevalencias y tendencias de uso, la edad de inicio, las dificultades relacionadas al consumo y la necesidad de atención o de tratamiento.

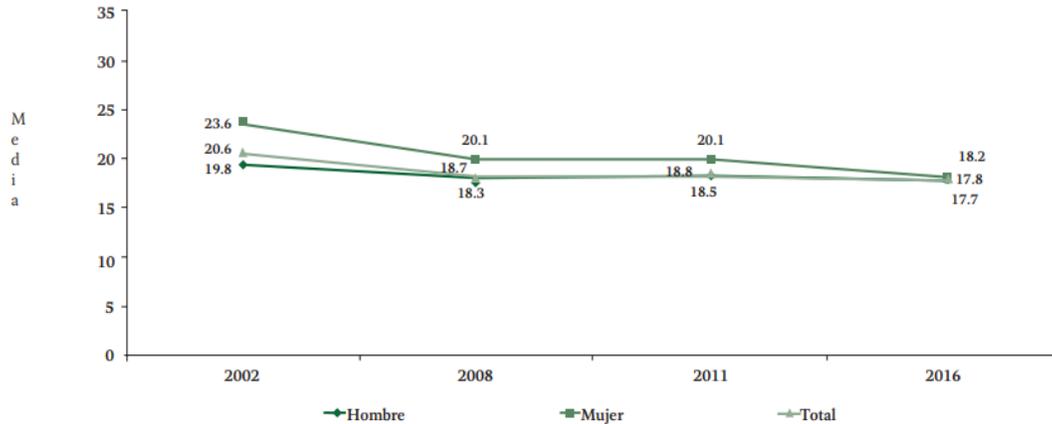
En el caso de la población de 12 a 17 años la dependencia a sustancias es de 0.6% de la población; en el rango de edad de 18 a 34 años es de 1.2% y finalmente en el rango de 35 a 65 años se tiene nuevamente 0.6% en dependencia a alguna sustancia definiendo a la población masculina de entre 18 a 34 años como grupo de mayor dependencia al consumo de cualquier sustancia.



Gráfica 1: Dependencia a drogas en la población de 12 a 65 años. Fuente: Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017.

### Tendencias en la edad de inicio del consumo de drogas.

Con los resultados obtenidos por las ENA 2002, 2008 y 2011 se observan disminuciones en la edad de inicio en el consumo a lo largo de los años, 20.6, 18.7, 18.8 años (respectivamente), en comparación con la media de 17.8 años en 2016.



Grafica 2: Tendencias de la media de edad de inicio de consumo de drogas en la población de 12 a 65 años. Fuente: Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017.

### Tendencias regionales del consumo de drogas.

Recordando los datos de integración estatal de cada una de las ocho regiones, tabla 1, para la representación de los resultados del consumo de sustancias de la ENA 2011 el territorio nacional se dividió en regiones, mismas que se replicaron para la representación en la ENCODAT 2016 – 2017, siguiendo esa guía el incremento en el consumo de drogas ilegales en la población total presentó un crecimiento de 9.2% a 13.7% en la región Occidental, 7.3% a 10.9% en la región Norcentral, 5% a 9.3% en la región sur y de 6.2% a 9.1% en la región Centro.

	2011				2016			
	Cualquier droga		Drogas ilegales		Cualquier droga		Drogas ilegales	
	%	IC 95%						
<b>Total</b>								
Norcentral	2.0	1.321-2.773	1.6	0.926-2.320	3.8*	3.042-4.461	3.3*	2.629-4.013
Noroccidental	2.9	2.055-3.773	2.8	1.945-3.660	3.6	2.984-4.259	3.5	2.846-4.108
Nororiental	3.1	1.966-4.275	2.4	1.348-3.474	2.7	2.011-3.345	2.4	1.722-2.989
Occidental	2.0	1.365-2.627	1.5	0.921-2.014	4.5*	3.111-5.827	4.1*	2.807-5.491
Centro	1.5	0.837-2.131	1.2	0.591-1.784	2.6	2.047-3.080	2.4*	1.913-2.920
Ciudad de México	2.1	1.186-2.976	1.7	0.874-2.575	2.9	1.710-4.107	2.7	1.561-3.929
Centro Sur	1.5	0.717-2.200	1.3	0.604-2.069	2.6	2.072-3.132	2.1	1.669-2.608
Sur	0.8	0.311-1.194	0.6	0.145-0.963	2.2*	1.736-2.631	2.0*	1.536-2.412
Nacional	1.8	1.529-2.120	1.5	1.229-1.782	2.9*	2.670-3.216	2.7*	2.417-2.947
<b>Hombres</b>								
Norcentral	3.2	1.765-4.610	3.0	1.588-4.340	5.7	4.431-6.946	5.3	4.043-6.516
Noroccidental	5.3	3.557-7.034	5.3	3.557-7.034	6.2	4.978-7.334	6.0	4.812-7.135
Nororiental	5.3	3.132-7.520	4.5	2.424-6.630	3.7	2.502-4.884	3.4	2.267-4.607
Occidental	3.2	2.018-4.316	2.7	1.619-3.811	6.8*	4.340-9.315	6.6*	4.095-9.058
Centro	2.1	1.000-3.223	1.7	0.684-2.670	4.1	3.111-5.119	4.0*	3.015-5.016
Ciudad de México	3.3	1.694-4.923	2.6	1.141-4.115	4.9	2.623-7.208	4.7	2.448-6.997
Centro Sur	2.7	1.270-4.228	2.7	1.270-4.228	4.1	3.106-5.054	3.5	2.593-4.413
Sur	1.0	0.224-1.767	0.9	0.102-1.603	3.5*	2.703-4.360	3.4*	2.569-4.218
Nacional	3.0	2.435-3.516	2.6	2.126-3.147	4.6*	4.139-5.160	4.4*	3.897-4.906
<b>Mujeres</b>								
Norcentral	1.0	0.403-1.501	0.3	---	1.9	1.152-2.607	1.4	0.770-2.086
Noroccidental	0.5	0.031-0.984	0.3	---	1.1	0.646-1.509	1.0	0.548-1.393
Nororiental	1.0	0.119-1.835	0.4	---	1.7	1.007-2.400	1.3	0.713-1.919
Occidental	0.9	0.238-1.541	0.3	---	2.2	0.962-3.525	1.9	0.619-3.098
Centro	0.9	0.170-1.650	0.7	0.031-1.448	1.2	0.726-1.577	1.0	0.619-1.303
Ciudad de México	1.0	0.030-1.902	0.9	---	1.1	0.289-1.856	0.9	0.194-1.676
Centro Sur	0.3	---	0.1	---	1.3	0.761-1.749	0.9	0.497-1.292
Sur	0.5	0.064-0.976	0.3	---	0.9	0.530-1.289	0.6	0.290-0.974
Nacional	0.7	0.470-1.023	0.4	0.200-0.692	1.3*	1.113-1.585	1.1*	0.867-1.286

\*Incremento significativo.

Tres guiones indican que la precisión del estudio no permite detectar la prevalencia de la conducta.

Tabla 4: Tendencias regionales del consumo de drogas en el último año. Población de 12 a 65 años.

Fuente: Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017.

### Tendencias y variaciones estatales del consumo de drogas.

Al observar la información a nivel estatal, para el consumo de cualquier droga alguna vez se registró un incremento de 5.7% en 2008 a 10.3% en 2016; por estado el mayor crecimiento proporcional se presentó, con respecto a 2008, fueron Jalisco de 5.3% a 16%, Colima 2.8% a 12.4% y Coahuila de 2.6% a 11.4%.

Estado	2008		2016	
	%	IC 95%	%	IC 95%
Aguascalientes	5.3	3.942-6.715	11.8*	9.723-13.932
Baja California	9.6	7.008-12.183	13.8	11.803-15.754
Baja California Sur	8.0	6.217-9.689	11.1	9.099-13.201
Campeche	5.3	3.576-7.059	11.8*	9.328-14.337
Chiapas	2.2	1.040-3.312	5.8*	4.170-7.455
Chihuahua	8.2	6.137-10.167	12.2	10.149-14.174
Ciudad de México	8.5	7.008-9.938	10.6	8.355-12.906
Coahuila	3.6	2.380-4.848	11.4*	9.215-13.499
Colima	2.8	1.764-3.758	12.4*	10.185-14.700
Durango	8.2	6.366-10.093	10.5	8.583-12.396
Estado de México	4.8	3.491-6.078	10.0*	7.936-12.150
Guanajuato	6.9	5.562-8.276	10.0	8.091-11.836
Herrero	5.2	3.487-6.829	7.8	5.930-9.586
Hidalgo	9.2	6.855-11.544	7.9	6.155-9.710
Jalisco	5.3	3.905-6.746	16.0*	12.831-19.255
Michoacán	5.6	3.794-7.328	9.2*	7.500-10.892
Morelos	4.6	3.298-5.840	7.5	5.556-9.518
Nayarit	7.6	5.324-9.848	13.2*	10.916-15.512
Nuevo León	4.4	3.032-5.782	11.4*	9.002-13.719
Oaxaca	3.5	2.317-4.704	6.9*	5.319-8.572
Puebla	3.6	2.497-4.712	7.9*	5.608-10.126
Querétaro	6.8	5.133-8.400	12.3*	10.105-14.490
Quintana Roo	11.2	9.355-13.046	15.5	12.315-18.719
San Luis Potosí	3.4	2.302-4.430	7.7*	6.112-9.340
Sinaloa	6.6	4.253-8.921	9.4	7.726-11.170
Sonora	5.5	4.015-6.926	10.9*	8.849-12.866
Tabasco	6.4	4.407-8.474	10.8*	8.957-12.544
Tamaulipas	11.1	8.360-13.763	10.3	8.662-11.942
Tlaxcala	3.0	2.002-3.988	7.1*	5.385-8.893
Veracruz	3.3	2.053-4.544	9.2*	7.329-11.142
Yucatán	4.7	3.146-6.324	11.4*	8.814-13.897
Zacatecas	6.4	4.915-7.871	9.9	7.749-11.971
<b>Nacional</b>	5.7	5.360-6.082	10.3*	9.788-10.813

\*Incremento significativo.

Tabla 5: Incidencia Acumulada del uso de cualquier droga. Población de 12 a 65 años. Fuente:  
Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017.

Dificultades relacionadas al consumo de drogas.

Las dificultades ligadas al consumo fueron evaluadas en una escala de 0 a 10 donde cero representa ninguna dificultad y diez una dificultad muy grave; los datos indican cómo en la medida que la población muestra una mayor problemática o dependencia al consumo se presenta una mayor dificultad ante ciertas situaciones cotidianas.

Dificultades*	Hombres						Mujeres					
	12-17 años		18-34 años		35-65 años		12-17 años		18-34 años		35-65 años	
	$\bar{X}$	D.E	$\bar{X}$	D.E	$\bar{X}$	D.E	$\bar{X}$	D.E	$\bar{X}$	D.E	$\bar{X}$	D.E
	<b>En las tareas domésticas</b>											
Consumidores con Dependencia	3.9	3.2	3.8	2.8	3.6	2.7	2.7	2.4	3.7	3.2	3.3	1.8
Consumidores Regulares	1.9	2.2	1.7	2.1	1.1	1.6	1.4	1.8	2.7	3.9	1.0	2.0
Consumidores Experimentales	1.2	1.8	1.0	1.4	0.7	1.4	0.9	1.2	0.6	1.5	0.4	0.7
	<b>En su capacidad para trabajar o estudiar</b>											
Consumidores con Dependencia	3.4	2.9	4.0	3.3	3.4	2.7	3.8	3.6	3.1	2.9	2.7	1.8
Consumidores Regulares	2.2	2.8	2.1	2.6	1.5	2.1	1.5	2.0	1.9	2.3	0.4	0.7
Consumidores Experimentales	1.4	2.3	1.8	3.1	1.0	1.5	1.3	2.1	1.1	2.3	0.6	1.4
	<b>Para iniciar o mantener relaciones estrechas</b>											
Consumidores con Dependencia	2.8	2.9	3.6	3.0	3.9	3.1	3.2	3.3	2.4	2.6	0.4	2.0
Consumidores Regulares	2.0	2.4	1.9	2.5	1.2	1.9	1.6	2.5	2.1	2.2	0.8	1.5
Consumidores Experimentales	1.3	2.1	1.7	2.7	1.3	1.8	1.1	2.0	0.8	2.0	1.1	2.9
	<b>En su vida social</b>											
Consumidores con Dependencia	3.8	3.4	4.2	3.1	4.2	3.2	3.1	3.2	2.8	3.1	0.4	2.0
Consumidores Regulares	2.3	2.9	2.2	2.8	1.5	2.1	1.4	2.4	1.8	1.8	0.6	1.9
Consumidores Experimentales	1.6	2.5	1.8	2.6	1.5	1.9	1.6	2.2	0.7	2.0	1.0	2.4
	<b>Días perdidos en los últimos 12 meses**</b>											
Consumidores con Dependencia	13.1	40.2	38.5	93.2	22.9	67.8	13.6	32.7	23.0	35.7	15.7	7.9
Consumidores Regulares	15.0	62.4	5.5	36.6	1.2	8.8	0.8	1.7	0.4	1.8	6.9	48.3
Consumidores Experimentales	2.2	17.8	8.1	38.8	3.7	15.2	16.1	71.3	0.4	1.7	0.2	0.9

\*Media de puntuación de una escala de 0 a 10.

\*\*Media de días perdidos debido al consumo de drogas.

Tabla 6: Dificultades relacionadas al consumo de drogas. Población de 12 a 65 años. Fuente: Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017.

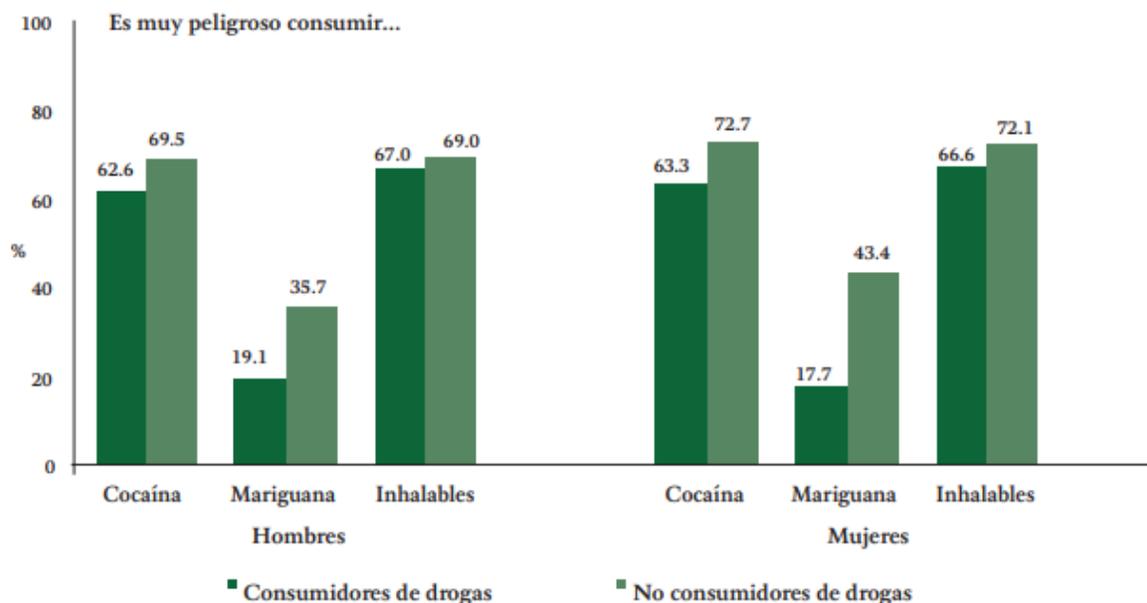
Como consecuencia de los problemas ocasionados por el consumo de drogas se desprenden los días perdidos, esta tendencia presenta al rango de edad de 18 – 34 años como quienes más días perdidos reportaron, 38.5 días para hombres y 23 días para mujeres; seguido por el grupo de 35 – 65 años con 22.9 y 15.7 días respectivamente y en último lugar los días perdidos en el rango de edad 12 – 17 años con una media de 13.1 días para hombres y 13.6 para mujeres siendo este grupo de edad los que menos días perdidos reportaron.

#### Tratamiento y necesidades de atención.

La información recabada por la ENCODAT 2016 -2017 en el campo de la solicitud de atención especializada indicó que de los consumidores de drogas el 8% ha acudido a tratamiento; el 21.4% de los consumidores de drogas en los últimos doce meses y el 24.5% de las personas con dependencia en el último año completaron el tratamiento con algún profesional de la salud. Del mismo modo los datos indicaron que en la población de 12 a 17 años el 0.6% dependiente y el 5.8% sin dependencia requieren de algún tipo de ayuda por consumo.

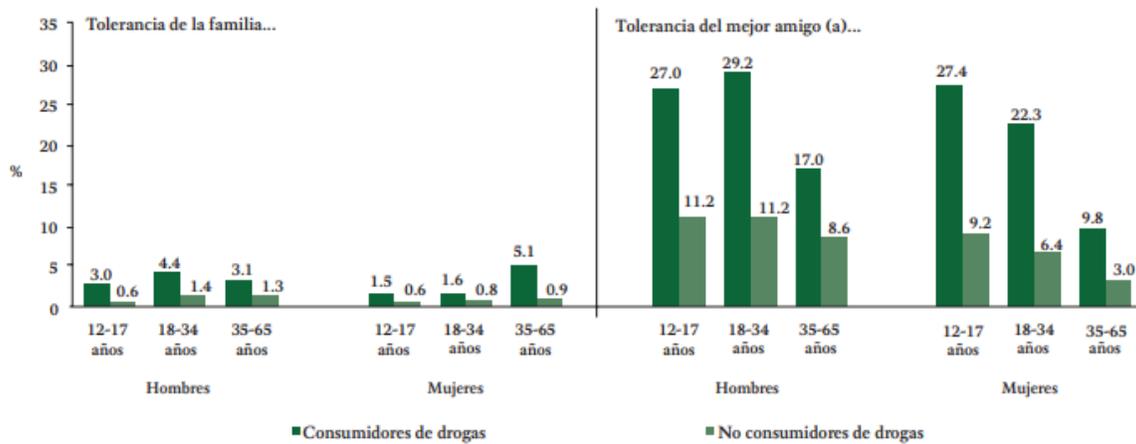
#### Percepción de riesgo y tolerancia social.

En cuestión social que alguna droga sea ofrecida, en venta o regalada, involucra el entorno que rodea al individuo. Por sexo, los hombres no consumidores reportan porcentajes menores en la percepción de riesgo que las mujeres siendo 69.5% / 72.7% en cocaína, 35.7% / 43.4% mariguana, 69% / 72.1% inhalables respectivamente.



Gráfica 3: Percepción de riesgo ante el consumo de drogas. Población de 12 a 65 años. Fuente: Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017.

En la población de 12 a 17 años que ha consumido drogas se observa una baja tolerancia de la familia hacia el consumo y aún más baja en la población no consumidora, el grupo consumidor de entre los 18 a los 34 años considera que su familia es tolerante ante el consumo (4.4% hombres y 1.6% mujeres) mientras que la parte que no consume tiene un porcentaje menor (1.4 hombres y 0.8% mujeres); en el grupo de mayor edad, de 35 a 65 años se reporta que la tolerancia de la familia hacia el consumo es de 3.1% en hombres y 5.1% en mujeres.



Gráfica 4: Tolerancia social hacia el consumo de drogas. Fuente: Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco ENCODAT 2016-2017.

Por sexo, se observa que los hombres expuestos a drogas, el 35% ha consumido cualquier sustancia, completamente diferente en la población no expuesta donde solo ha consumido el 3.8%. En el caso de las mujeres quienes han estado expuestas el 25.1% ha consumido mientras que de las no expuestas tan solo el 1.4% ha consumido cualquier droga.

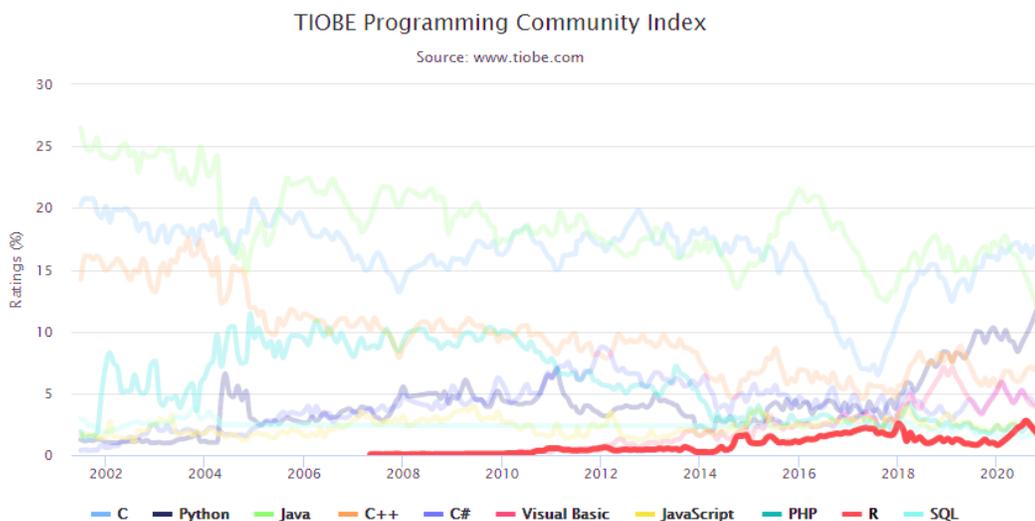
En el último caso, la exposición a la prevención demuestra que quienes no han estado expuestos a programas de prevención y que tampoco han participado en ellos, dentro del rango de 35 – 65 años, duplican a quienes si han participado en programas preventivos; demostrando el impacto positivo que tienen los programas de prevención.

### Capítulo 3. Vínculo de la Información con Lenguaje R

Los datos recopilados en las distintas encuestas son una representación general del total de variables que el sector salud toma en cuenta al entrevistar a un usuario de sustancias para así realizar estudios en características específicas. La mayor parte de los datos relacionados a la ENA 2008 y ENA 2011 se encuentra en bases de datos que tienen un formato especializado, en este caso ".sav" que corresponde a SPSS Statistics, mientras que los archivos digitales de la ENCODAT 2016-2017 se encuentran en formato ".csv" para poder manipularlos a través de Microsoft Access o limitadamente en Microsoft Excel; en ambos casos el software para la manipulación es de licencia restringida o software de pago. El hecho de que la información se encuentre en estos formatos no implica que el análisis o la manipulación de los resultados deban realizarse sobre un software específico, aunque se debe dejar claro que el uso de software privado es una gran ventaja, ya sea por las funciones implementadas, la interfaz de usuario intuitiva o el soporte técnico al que se tiene acceso.

Lenguaje R.

Se considera a R un lenguaje de programación porque permite dar instrucciones, usando código, al ordenador para que realice tareas específicas. El lenguaje R se posicionó como uno de los lenguajes más populares en apenas 10 años, situándose a finales de 2020 como el número 9 según *TIOBE Programming Community Index*, R es distribuido de manera gratuita bajo la Licencia Pública General de GNU por lo que es de uso libre y de código abierto lo que significa que es posible examinar el código que ejecuta R o implementar versiones con necesidades particulares sin importar si es con un fin académico, personal o comercial.



Gráfica 5: Popularidad de los lenguajes de programación, 2002-2020. Fuente: www.tiobe.com

El lenguaje R además de ser un código de programación es una herramienta en la elaboración de informes debido a que contiene una gran colección de bibliotecas para el análisis de datos y estadística, virtudes que lo convierte en un instrumento clave en el ambiente *Big Data*.

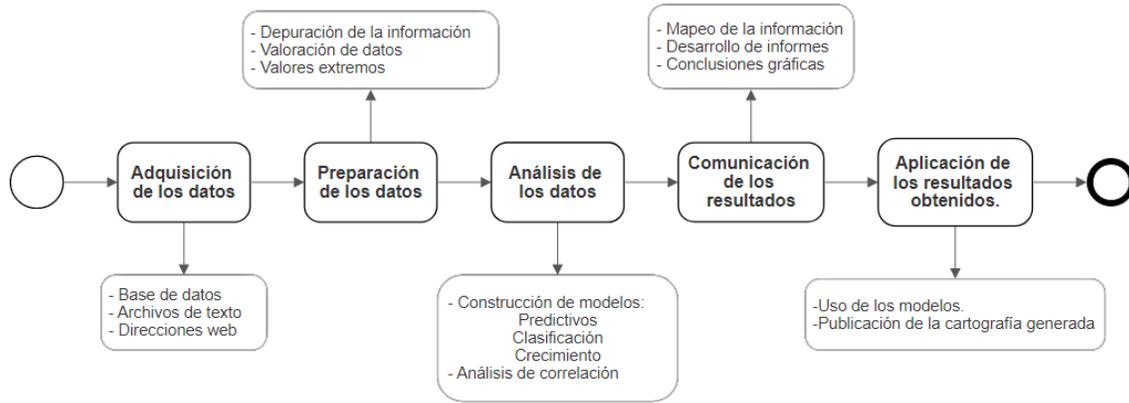


Figura 2: Fases del análisis de datos estadísticos. Fuente: Elaboración propia.

El lenguaje R tiene una base de desarrollo en el lenguaje S que es un lenguaje de licencia cerrada, creado y comercializado por AT&T Bell; el código fuente de R fue diseñado por Ross Ihaka y Robert Gentleman, está escrito en C y menormente en Fortran (Paradis, 2002, pág. 3); como se mencionó, el lenguaje R posee varias funciones para análisis estadísticos y la generación de gráficos que pueden ser visualizados de manera inmediata o ser guardados en distintos formatos, característica que ha hecho que sea considerado un lenguaje dedicado a la estadística cuando en realidad se trata de un lenguaje interpretado (como Java o Python), orientado a objetos y bastante flexible.

### 3.1 Base de datos ENCODAT 2016 – 2017, manipulación con lenguaje R.

La base de datos de la ENCODAT 2016 – 2017 se compone de dos archivos “.csv” que corresponden al cuestionario hogar e individual, el cuestionario hogar se concentró en datos sociodemográficos obtenidos de un mayor de edad por cada vivienda con conocimiento sobre las condiciones sociales mientras que el cuestionario individual estuvo dirigido a ambos grupos de edad, adolescentes y adultos que habitaran en la misma vivienda.

La base de datos correspondiente al cuestionario individual de la ENCODAT 2016 – 2017 se compone de 56 877 observaciones con un total de 1 715 variables que incluyen, por mencionar algunas, entidad estatal, municipio, localidad, sexo, edad, estado civil, religión, etc. Por otra parte, la base de datos correspondiente al cuestionario hogar incluye 55 907

observaciones con 48 variables donde se describen las condiciones físicas de la vivienda, percepción económica, estado de beneficencia en programas sociales, etc. En la Figura 3 se proporciona una muestra de la estructura de la base de datos.

ID_PERS	entidad	desc_ent	municipi	desc_mun	locali	desc_loc	ds2	ds3	ds4a	ds4b	ds4c	ds5	ds5esp
1	01001000103220050011	1	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	1	48	26	5	1968	2
2	01001000103220050071	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	30	12	6	1986	2
3	01001000103220050121	1	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	1	39	99	99	9999	2
4	01001000103220050171	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	42	27	5	1974	2
5	01001000103220050321	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	47	30	3	1969	2
6	01001000103220120041	1	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	63	99	99	9999	2
7	01001000103220120081	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	25	99	99	9999	2
8	01001000103220120121	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	44	26	5	1972	2
9	01001000103220120123	3	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	1	14	4	12	2001	2
10	01001000103220120151	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	63	30	11	1952	2
11	01001000103220120191	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	1	21	30	5	1995	2
12	01001000103220160011	5	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	1	31	9	12	1984	2
13	01001000103220160011	7	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	13	20	12	2002	2
14	01001000103220160041	4	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	23	25	6	1993	2
15	01001000103220160121	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	1	46	25	1	1970	2
16	01001000103220160221	2	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	50	10	11	1965	2
17	01001000103220220021	1	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	1	34	29	1	1982	2
18	01001000103220220021	4	01 AGUASCALIENTES	1	001 AGUASCALIENTES	1	0001 AGUASCALIENTES	2	13	8	9	2002	2

  Muestra general de las variables    
   Nombre clave de la variable    
   Variable con opción de especificar  
  Observación

Figura 3: Vista parcial de la base de datos de la ENCODAT 2016 – 2017. Fuente: Elaboración propia.

### 3.1.1 Clasificación de los datos recopilados en la encuesta.

El proceso para la clasificación y almacenamiento de información en la base de datos correspondiente al cuestionario individual de la ENCODAT 2016-2017 se realizó a través de variables con nombre clave para cada uno de los rubros del cuestionario, algunos de estos rubros incluyen a su vez preguntas que contienen una respuesta digital o descrita por el entrevistado en el caso de que no figure dentro de las opciones preestablecidas; el manejo de este tipo respuestas para un análisis SIG implica dos escenarios, una primer situación donde la respuesta contenida en las opciones es editada para hacer de un dígito una descripción textual con el fin de no tener que recurrir a la definición de las variables en cada ocasión que se necesite realizar una búsqueda y una segunda instancia donde la respuesta se encuentre fuera de las opciones y por ende sea necesario hacer nula la observación debido a la particularidad en el sondeo.

Por ejemplo, en el caso de la variable “ds9” que corresponde a la pregunta “¿Cuál fue el último grado que ha completado o completó en la escuela?” y tiene un rango de opciones del 1 al 9 sería necesario adaptar la respuesta a un texto que permita agilizar la búsqueda al

mismo tiempo que se pueda conocer la totalidad de la consulta, es decir, remplazar el dígito "5" por una descripción textual "BachilleratoIC" que representaría Bachillerato Inconcluso.

<b>ds9</b>	<b>1</b>	<b>Primaria incompleta (1 a 5 años)</b>
	2	<b>Primaria completa (6 años)</b>
	3	<b>Secundaria incompleta (1 a 2 años)</b>
	4	<b>Secundaria completa o equivalente (3 años)</b>
	5	<b>Bachillerato incompleto (1 a 2 años)</b>
	6	<b>Bachillerato completo o equivalente (aprox. 3 años)</b>
	7	<b>Estudios Universitarios incompletos (1 a 3 años)</b>
	8	<b>Estudios Universitarios completos (4 a 5 años)</b>
	9	<b>Estudios de Posgrado (2 a 4 años)</b>
	99	<b>No contesta</b>

Tabla 7: Descripción de la variable "ds9". Fuente: Definición de las variables ENCODAT 2016 - 2017.

En un análisis SIG con una cantidad moderada de observaciones realizar este proceso no implica mayor problema, en cambio, durante un análisis de más de 50 000 observaciones, como es el caso de la ENCODAT 2016 – 2017 realizar la edición de las variables para tener una descripción específica por observación implicaría una extensión en el tiempo de trabajo considerable aunado a extender innecesariamente el tamaño de los archivos; llevar un dígito o un carácter a una cadena de texto implica una mayor cantidad de memoria para el tratamiento y también en el almacenamiento de los archivos.

Los datos espaciales recopilados por la ENCODAT 2016-2017 (cuestionario hogar e individual) registra la dirección completa de cada entrevistado, numero, calle, colonia, municipio, etc. Mientras que, en la base de datos, por observación, únicamente se asentó la entidad federativa, municipio, localidad y un número de identificación personal que no incluye información sensible o directa sobre la vivienda del entrevistado, esto para respetar los términos de privacidad establecidos para la encuesta que también incluyó un consentimiento informado y en el caso de los menores de edad la autorización de los padres (Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; Instituto Nacional de Salud Pública, Comisión Nacional Contra las Adicciones; Secretaría de Salud, 2017, pág. 42)

**1. IDENTIFICACIÓN GEOGRÁFICA**

ENTIDAD FEDERATIVA \_\_\_\_\_

MUNICIPIO O DELEGACIÓN \_\_\_\_\_

LOCALIDAD \_\_\_\_\_

CLAVE DE AGEB..... \_\_\_\_\_

MANZANA..... \_\_\_\_\_

**2. FOLIO DEL HOGAR**

FOLIO: \_\_\_\_\_

NÚMERO PROGRESIVO DE LA VIVIENDA \_\_\_\_\_

**3. DIRECCIÓN DE LA VIVIENDA**

\_\_\_\_\_

(CALLE, AVENIDA, CALLEJÓN, CARRETERA, CAMINO, BOULEVARD, KM.)

NÚMERO EXTERIOR    NÚMERO INTERIOR    \_\_\_\_\_ (COLONIA, FRACCIONAMIENTO, BARRIO, UNIDAD HABITACIONAL)    CÓDIGO POSTAL \_\_\_\_\_

Figura 4: Registro de identificación geográfica. Fuente: Cuestionario Hogar, ENCODAT 2016 – 2017.

Tomando en cuenta la metodología seguida por la ENCODAT 2016-2017 cada una de las 56 877 observaciones cuenta con un número único de reconocimiento, el identificador de persona, número que consta de veintidós dígitos de los cuales, para los fines de este análisis se utilizan nueve y se ajustan como:

- Dos dígitos por entidad federativa.
- Tres dígitos para la clave municipal, y finalmente
- Cuatro dígitos para la localidad.

ID_PERS	entidad	desc_ent	munici	desc_mun	locali	desc_loc
37434	22	22 QUER_TARO	6	006 CORREGIDORA	1	0001 EL PUEBLITO

Figura 5: Descripción del número de identificación personal. Fuente: Vista parcial de la base de datos de la ENCODAT 2016 – 2017 en RStudio.

La finalidad del número de identificación personal es mantener cada una de las observaciones como única, considerando que se realizaron dos entrevistas por domicilio a dos grupos de edad diferentes siempre que la vivienda cumpliera con esta característica.

### 3.1.2 Administración geoespacial.

Las tecnologías de la información a la par de incorporar herramientas para recolectar, guardar, relacionar y distribuir datos han contribuido a la expansión de los horizontes de la Ingeniería Geomática como disciplina especializada pues es conocido que los estudios territoriales, climáticos, de flora, fauna, de urbanismo, etc. Han estado en manos de disciplinas que no son ajenas a los sistemas de información geográfica y que hacen uso de estas metodologías para presentar planes de desarrollo sustentable con aplicación en distintas áreas profesionales. Los estudios previos acerca del consumo de drogas (ENA's) se han concentrado y mostrado resultados a nivel regional, pero sin contundencia sobre el problema a pequeña escala debido a que dirigen el análisis a resultados globales en regiones establecidas, sur, centro y norte del país, etc.; mientras que al contar con registros a nivel localidad es viable condicionar el análisis y enfrentarlo a distintos factores en territorios pequeños, la selección de estos factores podría ser dentro del contexto de la ENCODAT y su respectiva base de datos o bien ser un elemento ajeno perteneciente a una base de datos distinta.

En un análisis GIS el proceso para convertir datos en información corresponde a la necesidad de comprender las componentes de la información geográfica, se trata, básicamente, de una componente espacial y su respectiva componente temática, en el caso de los datos recopilados por la ENCODAT 2016 -2017 podemos decir que la componente geográfica consiste en la ubicación de cada uno de los casos a lo largo del territorio nacional, mientras que la componente temática consta de las variables relacionadas a la encuesta; generalmente la componente espacial responde a la pregunta "¿dónde?" mientras que la componente temática describe el "qué" (Olaya, 2011, pág. 65).

La representación de la superficie terrestre ha sido una labor milenaria que ha sido considerada tanto un saber empírico como un oficio y es por estas cualidades que ha entregado obras de arte y representaciones gráficas diseñadas bajo la guía de ecuaciones matemáticas que hacen figurar, en forma o escala el reflejo de la realidad, este tipo de cálculos ha permitido que la cartografía proporcione, ahora de manera digital, una referencia tangible a aquellas formas y objetos que se han tomado de la realidad para localizarlas geográficamente a través de la computación. Los métodos cartográficos han hecho que la representación de la faz de la tierra pueda tener una vista global si se habla de coordenadas geográficas (latitud y longitud) o coordenadas cartesianas cuando se proyecta el globo terráqueo sobre un plano.

Proyecciones cartográficas existen en una variedad según las características que se quiere representar en la cartografía de acuerdo a las necesidades del cartógrafo. Estas se dividen en dos ramas: proyecciones por técnicas de proyección y proyecciones por características geométricas, de las primeras se pueden mencionar las proyecciones planas, cónicas o

cilíndricas mientras que, de las segundas se obtienen proyecciones conformes (mantienen la forma del terreno), equivalentes (mantienen el área del terreno proyectado) y finalmente, las equidistantes (conservan las distancias en una escala única). Es importante mencionar estos detalles ya que como se verá en el Capítulo 4 un objeto espacial necesita tener un sistema de referencia coordinado para poder mostrar con exactitud la ubicación de las características geográficas.

Los códigos EPSG, *European Petroleum Survey Group*.

La información geográfica digital que se encuentra en línea, en plataformas tales como CONABIO, CONAPO, INEGI, etc. tiene el atributo de encontrarse georreferenciada de manera global a un Datum, NAD27, WGS84, ITRF92 o estar proyectada bajo parámetros específicos para la nación en UTM (Universal transversa de Mercator), CCL (Cónica Conforme de Lambert.) u otra.

Cada uno de estos sistemas de referencia o proyecciones cuenta con un código EPSG, que es una codificación convencional que permite georreferenciar, ubicar, identificar o reproyectar las geometrías según los objetivos que se persigan, de manera homogénea para distintos usuarios de diversas nacionalidades y en distintos softwares especializados para este tipo de tareas.

El conjunto de datos de parámetros geodésicos EPSG es una colección de definiciones para distintos sistemas de referencia coordinados que pueden ser globales, nacionales o locales. (IOGP, 2020)

De manera única y según las características de la proyección para México los códigos EPSG corresponden de la siguiente manera para coordenadas geográficas:

- EPSG: 4326 – WGS84.
- EPSG: 4267 – NAD27.
- EPSG: 4483 – ITRF92.

Y cartesianas si se trata de una proyección:

EPSG: 6362 México ITRF92 / LLC (Cónica Conforme de Lambert)

Debido a las particularidades de la proyección Universal Transversa de Mercator (UTM) se tiene un código EPSG distinto para cada zona UTM a nivel global; de modo que el código correspondiente a la Zona Metropolitana del Valle de México es:

Sistema de Referencia	Código EPSG
México ITRF2008/ UTM Zona 14N	6369

Independientemente de la referencia geográfica que es posible asignar mediante software GIS o código en R con la biblioteca "*sf*" esta investigación no persigue fines de medición territorial por lo que no se hará énfasis en este tema.

## Capítulo 4. Filtro de datos correspondientes a la ENCODAT 2016 – 2017 y métodos de presentación gráfica.

La totalidad de los datos recopilados por la ENCODAT 2016 – 2017 representan una cantidad importante de variables y observaciones que concentran los hechos y características del consumo de sustancias, legales o ilegales a lo largo de las 32 entidades federativas que conforman al país; cada observación a su vez se compone de variables donde se presentan particularidades únicas por individuo y su experiencia en el consumo, percepción o exposición a las sustancias.

Como se vio en el Capítulo 3 la base de datos de la ENCODAT 2016 – 2017 contiene identificadores puntuales para cada observación mismos que, para un análisis geoespacial, necesitan ser filtrados, transportados y depurados para generar información en los niveles geográficos esperados, el filtro y la manipulación de estos datos se ejecuta por medio de lenguaje R en la interfaz RStudio con el objetivo de generar nuevas reservas de datos según el análisis que se realiza. Estas reservas generadas serán el instrumento con el que se elaborará cartografía haciendo uso de la colección de bibliotecas para presentación gráfica disponibles en lenguaje R.

### 4.1 Composición de la base de datos georreferenciados.

Dentro de los archivos base de la colección de la ENCODAT 2016 – 2017 se encuentra uno denominado “Definición de variables encodat 2016\_2017” que incluye la descripción de las variables que conforman la base de datos según el cuestionario que se aplicó a cada integrante seleccionado por vivienda.

Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la  
ENCODAT 2016 - 2017.

Variable	Ubicación	Etiqueta	Nivel de medida	Papel	Ancho de columna	Alineación	Formato de	Formato de	Valores
ID_PERS	1	<ninguno>	Nominal	Entrada	24	Izquierda	A22	A22	
entidad	2	Entidad	Nominal	Entrada	7	Izquierda	A2	A2	
desc_ent	3	Descripción de la Entidad	Nominal	Entrada	22	Izquierda	A50	A50	
munici	4	Municipio	Nominal	Entrada	3	Izquierda	A3	A3	
desc_mun	5	Descripción del Municipio	Nominal	Entrada	11	Izquierda	A50	A50	
locali	6	Localidad	Nominal	Entrada	4	Izquierda	A4	A4	
desc_loc	7	Descripción de la Localidad	Nominal	Entrada	13	Izquierda	A50	A50	
ds2	8	Marque el sexo del entrevistado	Nominal	Entrada	8	Derecha	F6	F6	
ds3	9	¿Qué edad tiene usted?	Nominal	Entrada	8	Derecha	F6	F6	
ds4a	10	Dégame por favor, su fecha de nacimiento: día, mes y año. DIA	Nominal	Entrada	8	Derecha	F6	F6	
ds4b	11	Dégame por favor, su fecha de nacimiento: día, mes y año. MES	Nominal	Entrada	8	Derecha	F6	F6	

Figura 6: Vista parcial de la descripción parcial de las variables que componen la base de datos de la ENCODAT 2016 – 2017. Fuente: Definición de variables encodat 2016\_2017.

Siguiendo la guía que entrega la descripción de las variables es posible determinar qué variables se pueden seleccionar y separar en una nueva tabla para comenzar el análisis. La selección de variables para la representación cartográfica temática es el paso fundamental ya que de la totalidad de datos recabados se extraen solo aquellos que se consideran con una relevancia importante para la presentación gráfica en turno de modo que se genere información precisa conforme los datos manipulados.

En este listado de variables prioritarias se encuentra la referencia geográfica de las observaciones, es decir, la ubicación puntual de cada entrevistado según el nivel espacial del análisis que se necesita realizar; de manera general las variables básicas que se almacenan y requieren en cada tabla son las siguientes:

ID_Pers	Identificador único por observación.
ent	Entidad Federativa de residencia
mun	Municipio de residencia.
loc	Localidad de residencia.

Tabla 7: Selección de las variables de referencia geográfica. Fuente: Definición de variables encodat 2016\_2017.

Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la  
ENCODAT 2016 - 2017.

A esta selección de datos referenciados según el análisis se pueden anexar nuevas variables como edad, sexo, estado civil, etc. Con la finalidad de completar el esquema de selección correspondiente al análisis temático que se realice, paralelamente la información cartográfica digital del país utiliza claves para representar la división política a nivel nacional, municipal y local, esquema similar que utilizó la ENCODAT 2016 – 2017 en la digitalización de los datos recabados.

Clave Entidad	Nombre Entidad
01	Aguascalientes
02	Baja California
03	Baja California Sur
04	Campeche
05	Coahuila de Zaragoza
06	Colima
07	Chiapas
08	Chihuahua
09	Ciudad de México
10	Durango
11	Guanajuato
12	Guerrero
13	Hidalgo
14	Jalisco
15	México
16	Michoacán de Ocampo
17	Morelos
18	Nayarit
19	Nuevo León
20	Oaxaca
21	Puebla
22	Querétaro
23	Quintana Roo
24	San Luis Potosí
25	Sinaloa
26	Sonora
27	Tabasco
28	Tamaulipas
29	Tlaxcala
30	Veracruz de Ignacio de la Llave
31	Yucatán
32	Zacatecas

Tabla 8: Clave única por Entidad Federativa. Fuente: Presupuesto de Egresos de la Federación 2012, Anexo: Catálogo de Claves de Entidades Federativas y Municipios.

Asimismo, cada entidad federativa cuenta con una clave de tres dígitos que corresponde a cada municipio.

Clave Entidad	Nombre Entidad	Clave Municipio	Nombre del Municipio
03	Baja California Sur	001	Comondú
03	Baja California Sur	002	Mulegé
03	Baja California Sur	003	La Paz
03	Baja California Sur	008	Los Cabos
03	Baja California Sur	009	Loreto

Tabla 9: Clave única municipal. Fuente: Presupuesto de Egresos de la Federación 2012, Anexo: Catálogo de Claves de Entidades Federativas y Municipios.

Y adicionalmente se utiliza una clave de cuatro dígitos para referirse a las localidades que existen dentro de un municipio.

Clave Entidad	Nombre Entidad	Clave Municipio	Nombre del Municipio	Clave Localidad	Nombre Localidad
03	Baja California Sur	001	Comondú	0001	Ciudad Constitución
03	Baja California Sur	001	Comondú	0003	Puerto Adolfo Lopez Mateos
03	Baja California Sur	001	Comondú	0011	El Aguajito Uno
03	Baja California Sur	001	Comondú	0012	El Aguajito
03	Baja California Sur	001	Comondú	0020	Andachires
03	Baja California Sur	001	Comondú	...	...

Tabla 10: Clave única por localidad. Fuente: Presupuesto de Egresos de la Federación 2012, Anexo: Catálogo de Claves de Entidades Federativas y Municipios.

Una vez definida la base geográfica el tratamiento de la base de datos consiste en seleccionar por rubro el grupo de variables representativas del cuestionario que para los

finde de esta tesis serán inicialmente las sustancias ilegales y algunos factores sociodemográficos; además se deben establecer los parámetros, medidas y nivel geográfico con el fin de diseñar un análisis adecuado para los datos y su derivada presentación gráfica.

Las variables seleccionadas de interés significativo por rubro son:

#### Sociodemográfico.

ds2	Sexo
ds3	¿Cuál es su edad?
ds6	¿Cuál es su estado civil?
ds7	¿Cuál es su religión?
ds8	¿Actualmente estudia?
ds9	¿Cuál es su último grado de estudios?
ds23	¿Alguno de sus hijos ha tenido problemas por consumir drogas o alcohol?

Tabla 11: Selección de las variables sociodemográficas. Fuente: Definición de variables encodat 2016\_2017.

#### Sustancias Ilegales.

di1a	¿Ha usado, tomado o probado marihuana?
di1b	¿Ha usado, tomado o probado cocaína?
di1c	¿Ha usado, tomado o probado crack?
di1d	¿Ha usado, tomado o probado alucinógenos?
di1e	¿Ha usado, tomado o probado inhalables?
di1f	¿Ha usado, tomado o probado heroína?
di1g	¿Ha usado, tomado o probado estimulantes anfetamínicos?
di1h	¿Ha usado, tomado o probado otras drogas?
di4a	¿Qué edad tenía cuando uso marihuana por primera vez?
di4b	¿Qué edad tenía cuando uso cocaína por primera vez?
di4c	¿Qué edad tenía cuando uso crack por primera vez?
di4d	¿Qué edad tenía cuando uso alucinógenos por primera vez?
di4e	¿Qué edad tenía cuando uso inhalables por primera vez?
di4f	¿Qué edad tenía cuando uso heroína por primera vez?
di4g	¿Qué edad tenía cuando uso estimulantes anfetamínicos por primera vez?

Tabla 12: Selección de las variables de consumo de drogas. Fuente: Definición de variables encodat 2016\_2017.

Percepción de riesgo.

ts18	Consumir marihuana es... No es peligroso, peligroso, muy peligroso.
ts19	Consumir heroína es... No es peligroso, peligroso, muy peligroso.
ts20	Consumir cocaína es... No es peligroso, peligroso, muy peligroso.
ts21	Consumir inhalables es... No es peligroso, peligroso, muy peligroso.
ts22	Consumir metanfetaminas es... No es peligroso, peligroso, muy peligroso.
ts23	Consumir tabaco es... No es peligroso, peligroso, muy peligroso.
ts24	Consumir alcohol frecuentemente es... No es peligroso, peligroso, muy peligroso.

Tabla 13: Selección de las variables, percepción de riesgo. Fuente: Definición de variables encodat 2016\_2017.

La selección elaborada para cada uno de los rubros en el cuestionario de la ENCODAT 2016 – 2017 se puede reducir o ampliar según el interés que se tenga para cada parte del cuestionario (alcohol, tabaco, drogas, exposición a la prevención, etc.) haciendo uso del nombre clave de la variable y transportando los datos almacenados en la base a una nueva tabla siguiendo un proceso de filtrado lógico del cual, con los resultados, se elabora la cartografía o representación estadística.

El filtro de las variables para el proceso cartográfico o estadístico genera datos que se reúnen de manera independiente en un objeto y mediante un proceso de unión se hacen en nueva base datos ligada a un archivo vectorial que posee información geométrica o puntual georreferenciada y en norma con los parámetros geográficos de manipulación mediante lenguaje de bajo nivel o software GIS.

## 4.2 Filtro lógico.

Las herramientas y protocolos de trabajo en Software GIS se encuentran basados en funciones que pertenecen a las bibliotecas GDAL y OGR (las más utilizadas, pero no únicas); GDAL en conjunto con OGR son una biblioteca de acceso a datos geoespaciales, GDAL para formatos ráster y OGR en formatos vectoriales (puntos, líneas y polígonos). Como instrumento independiente estas bibliotecas están diseñadas en el lenguaje de programación C++ lo que ha permitido que sean distribuidas a otras plataformas de código como lo son C, Python, Java y R, por mencionar algunos; como ejemplo de esta múltiple distribución se pueden mencionar los intérpretes más conocidos del lenguaje Python que pertenecen a software GIS como ArcGis y Qgis con ArcPy y PyQ respectivamente. GDAL/OGR se encuentran dentro del programa *Open Source License* que es parte de *Open Source Geospatial Foundation* (OSGeo) lo que significa que es posible hacer uso, modificar y diseñar sobre estas funciones de manera gratuita con fines educativos o de lucro.

El uso de estas bibliotecas en el proceso de filtrado es esencial para representar los datos contenidos en la base de datos de la ENCODAT 2016 – 2017 ya que consisten en una observación única por entrevistado de modo que, tomando en cuenta el proceso de intervención y la forma en que se almacenaron los datos para obtener información a distintos niveles de población es necesario contabilizar de manera agrupada los datos, a nivel localidad, municipal y estatal, de tal forma que la representación cartográfica sea adecuada en proporción y escala para los tipos de variables que se presentan.

Para realizar un filtrado exitoso en un conjunto tan extenso de variables como lo es la base de datos de la ENCODAT 2016 – 2017 es importante conocer los niveles de medición estadística y cuál es su impacto en las características de los datos recolectados para cada una de las variables.

La estadística descriptiva contempla dos tipos de variables, cuantitativas y cualitativas, la primera corresponde a aquellas observaciones para las que se posee una unidad de medición; se les considera cuantitativa continua si es posible representar fracciones, proporciones o porcentajes y son consideradas cuantitativas discretas en el caso que solo contabilicen unidades enteras, un ejemplo de una variable continua es la cantidad de agua contenida en un recipiente medida en litros, mientras que una variable discreta podría ser el número de cigarrillos consumidos en un día.

Por otra parte, las variables cualitativas se caracterizan por categorizar a los individuos en relación a sus atributos o características que los excluyen de otros casos u observaciones (sexo, etnia, etc.).

El proceso de filtrado en lenguaje R consiste en utilizar funciones de agrupamiento y/o selección para aquellas variables u observaciones que nos interesa presentar; para el manejo

de base de datos en lenguaje R se cuenta con una gran variedad de bibliotecas que ofrecen funciones especializadas como `dplyr`, `tidyr`, `readr`, `prrr`, `forcats`, etc. que pertenecen a un compendio más grande conocido como "Tidyverse" enfocado al manejo masivo de datos.

Cuando se habla de manejo lógico se hace referencia a lo que en programación se conoce como lenguaje de bajo nivel, que es indicar instrucciones precisas a la computadora mediante código para acceder a operaciones matemáticas, lógicas, cíclicas, de edición o guardado.

#### 4.2.1 Tidyverse.

##### Biblioteca Readr.

Las funciones de la biblioteca `readr` consisten en importar datos contenidos en archivos de texto plano o archivos rectangulares que poseen diversas características de guardado, almacenando la imagen del archivo en un objeto. Las funciones recurrentes de la biblioteca son:

`read_csv()`: Despliega en una tabla el contenido de un archivo de texto específico separado por comas.

```
> library(readr)
> read_csv("../...dataframe.csv", col_names = TRUE)
```

Figura 7: Ejemplo, carga de archivos mediante la biblioteca `readr`. Fuente: Elaboración propia.

`read_csv()`: Abre un archivo de texto específico separado por tabulaciones y lo asocia a un objeto.

`read_delim()`: Importa a un objeto de tipo tabular el contenido de un archivo de texto plano especificando el tipo de separador que se utilizó en él.

`read_table()`: Toma de un fichero de texto tabular las columnas donde se almacenan los datos separados por un espacio en blanco.

Según la función empleada es posible agregar argumentos que faciliten el proceso de análisis como es el caso del argumento `cols()` donde se especifica mediante el nombre de la variable (columna) el tipo de dato asociado, entero, doble, factor, carácter, etc.

Algunas de las funciones del paquete "readr" fueron concebidas bajo el concepto de importar `dataframes` con características particulares ya que el código base de R también

contiene funciones de importación para archivos .csv o de texto plano .txt, como es el caso de *read.csv()* o *read.table()* respectivamente.

```
> E1617 <- read.csv(file = "../encodat1617.csv", sep = ",")
```

Figura 8: Ejemplo, función base para la importación de archivos ".csv" Fuente: Elaboración propia.

### Biblioteca Dplyr.

La principal contribución del paquete *dplyr* no consiste en agregar nuevas funciones, sino que radica en la incorporación de una gramática "verbal" paralela a las funciones base para la manipulación de *dataframes*; independientemente la biblioteca *dplyr* agrega un operador "%>%" conocido como "pipe", que es utilizado para concatenar múltiples acciones en una sola línea, "pipeline".

Las funciones principales o recurrentes de la biblioteca *dplyr* son:

*select()*: Devuelve la selección de un conjunto de columnas pertenecientes a un mismo *dataframe*.

*filter()*: Entrega un conjunto de observaciones (filas) de acuerdo a una o varias condiciones lógicas.

*rename()*: Renombra variables - columnas de un *dataframe*.

*mutate()*: Agrega nuevas columnas al *dataframe* según condiciones lógicas u operaciones.

*summarise()*: Genera resúmenes estadísticos de distintas variables en el *dataframe*.

*group\_by()*: Agrupa un conjunto de observaciones (filas) que comparten valores en una o más columnas o que cumplen con cierta condición lógica.

```
> Drogas_Edad_M <- E1617%>% filter(ds2 == 2) %>% group_by(CVEGEO) %>% summarise(Entrevistadas=n())
```

Figura 9: Ejemplo, Pipeline para el filtro de observaciones. Fuente: Elaboración propia.

### Biblioteca Tidy.

La biblioteca o paquete *tidyr* está pensado para manipular la forma de un *dataframe* entre formatos "anchos" y "largos", en el formato "ancho" (formato estipulado para la base de datos de la ENCODAT 2016 – 2017) cada fila es un tema, observación o individuo donde se

tienen múltiples variables de observación; el formato largo es aquel donde cada columna es una variable y cada fila es una observación, este tipo de ajustes al formato están ideados para tener un acceso más visual a los datos contenidos.

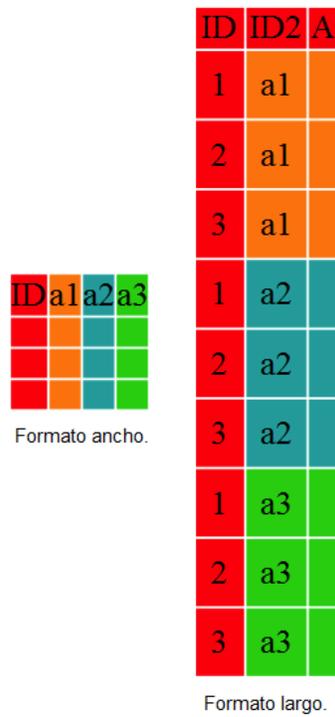


Figura 10: Formatos tidy. Fuente: <https://swcarpentry.github.io/r-novice-gapminder-es/14-tidy/>

Algunas de las funciones recurrentes de la biblioteca *tidyr* son:

*separate()*: Separa una columna en múltiples columnas según el método indicado.

*complete()*: Completa un *dataframe* que tenga combinaciones de valores perdidos.

*drop\_na()*: Elimina columnas que contienen valores perdidos.

*replace\_na()*: Reemplaza valores perdidos con valores indicados o generados mediante procesos lógicos o matemáticos.

### 4.3 Bibliotecas de manipulación vectorial, Simple Features (sf).

La biblioteca *simple features* es una actualización independiente a lo que en lenguaje R alguna vez fue la biblioteca *sp*, que poseía características de ambientes como *rgdal* y *rgeos* (GDAL y OGR) pero que durante el proceso de trabajo entregaba resultados incompletos o ineficientes.

El paquete *sf* hace énfasis en la manera de representar la geometría del mundo real a través de una computadora siguiendo la filosofía de que un grupo de características pertenecen a un objeto particular y este a su vez forma parte de un objeto más grande donde ambos describen cualidades físicas y tienen atributos de geolocalización.

La geometría más simple de representación geográfica es un punto, el cual posee coordenadas en dos, tres o hasta cuatro dimensiones. La geometría bidimensional básica se refiere a un punto al que se le asocia norte y este, latitud o longitud, que conocemos como XY; si se le asocia, además, altitud a las propiedades espaciales se tiene un objeto tridimensional XYZ, mientras que un objeto de cuatro dimensiones es aquel que posee características como tiempo de medición o la medida de error de las coordenadas, XYZM.

Como se mencionó en el capítulo 3, existe una variedad de sistemas de referencia coordenados que permiten detallar la ubicación de un punto sobre la faz de la tierra, propiedad de la cual la biblioteca "*sf*" no es ajena pues a cada objeto espacial le asigna un CRS o sistema de referencia coordenado que puede estar asociado a un datum como en el caso del WGS84 o cartesiano si se habla de una proyección como lo podría ser UTM o CCL, todo dependiendo de las características contenidas en el objeto espacial.

Cuando se trabaja con archivos vectoriales existentes las funciones recurrentes de la biblioteca *simple features* son:

*st\_read()*: Lee y almacena un archivo espacial como un objeto espacial, "*sf object*"

```
> DPLoc<-st_read("../loc2010gw.shp")
```

Figura 11: Ejemplo, carga de un archivo en formato ".shp". Fuente: Elaboración propia.

*st\_crs()*: Identifica el CRS del archivo espacial solicitado.

*st\_as\_sf()*: Crea un objeto espacial "*sf object*" desde un *dataframe* no geoespacial.

*st\_transform()*: Convierte del CRS original de un objeto espacial a un sistema de referencia coordenado diferente.

#### 4.4 Métodos de presentación en R.

Las bibliotecas de presentación gráfica en lenguaje R funcionan de manera independiente al tipo de dato que se quiere presentar, es decir, se puede elaborar un código que muestre resultados estadísticos (histogramas, gráficas, etc.) de cualquier tema siempre que se involucren colecciones de datos almacenados en *dataframes* que son objetos donde se depositan reproducciones del archivo de datos original.

Para hacer uso de elementos vectoriales es necesario asociar la biblioteca gráfica con la biblioteca de manejo espacial de manera que sea posible trasladar y reducir los datos a la selección deseada además de poder ligarlos a la tabla de características geométricas.

La presentación gráfica en lenguaje R se logra a partir del comando "plot()" que es una función base cuando se trabaja con una interfaz de usuario como RStudio, pero paralelamente existen bibliotecas como "ggplot2" que profundizan en los detalles de presentación todo a través de una línea concatenada de funciones; algunas de las funciones utilizadas conjuntamente entre la biblioteca "ggplot2" y "sf" para realizar una presentación cartográfica son:

*ggplot()*: Función base para hacer una presentación.

*geom\_sf()*: Determina el origen de los datos a presentar mediante instrucciones de filtro o selección lógica.

*aes()*: Establece las características estéticas de los elementos presentados.

*geom\_sf\_text()*: Permite el uso de datos de tipo carácter que se encuentran en el *dataframe* para mostrar etiquetas, estableciendo un proceso lógico de selección.

*labs()*: Configura las características visuales y de contenido de las etiquetas o simbología contenida.

```
ggplot()+  
  geom_sf(data = DPMun_EH %>% filter(CVE_ENT=="09"), aes(fill=di4a))+  
  scale_fill_viridis_b(option = "viridis", direction = -1)+  
  geom_sf(data = DPLoc_H %>% filter(CVE_EDO=="09"), aes(size=di1a), alpha=.7, col="red")+
```

Figura 12: Ejemplo, código parcial para generar una presentación cartográfica. Fuente: Elaboración propia.

Un objeto espacial del tipo "sf" puede ser un archivo con formato ".shp" que contenga las características geométricas de un área determinada, en este caso y para este análisis se trabajó con los archivos vectoriales de división política a nivel estatal, municipal y local de los Estados Unidos Mexicanos. Estos archivos fueron tomados de Geoportal de CONABIO (división política estatal y municipal) y del Marco Geoestadístico Integrado 2019 de INEGI (Información puntual de localidades) además de la delimitación de las Zonas Metropolitanas del país rescatada de CONAPO.

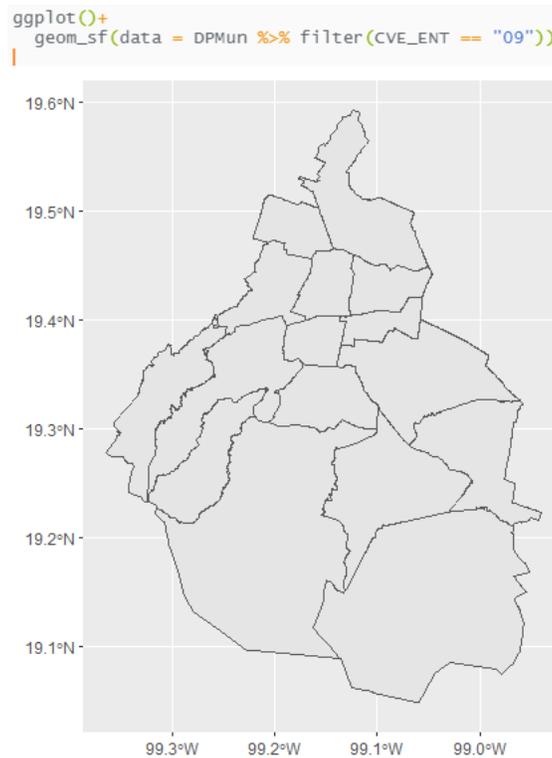


Figura 13: Línea de código para la presentación de la división política de la CDMX. Fuente: Elaboración propia.

Como se mencionó previamente la función `st_crs()` devuelve la información perteneciente al sistema de referencia coordinado de un objeto espacial, en este caso el archivo en formato ".shp" correspondiente a la división política a nivel municipal.

```
> st_crs(DPMun)
Coordinate Reference System:
  User input: WGS 84
  wkt:
  GEOGCRS["WGS 84",
    DATUM["World Geodetic System 1984",
      ELLIPSOID["WGS 84",6378137,298.257223563,
        LENGTHUNIT["metre",1]]],
    PRIMEM["Greenwich",0,
      ANGLEUNIT["degree",0.0174532925199433]],
    CS[ellipsoidal,2],
    AXIS["latitude",north,
      ORDER[1],
      ANGLEUNIT["degree",0.0174532925199433]],
    AXIS["longitude",east,
      ORDER[2],
      ANGLEUNIT["degree",0.0174532925199433]],
    ID["EPSG",4326]]
> |
```

Figura 14: Sistema de referencia coordenado del objeto DPMun correspondiente al archivo ".shp" de la división política a nivel municipal. Fuente: Elaboración propia.

#### 4.4.1 Rayshader.

La biblioteca "rayshader" es una herramienta independiente del Tidyverse y de la biblioteca Simple Features "sf" pero que en conjunto con estas permite elaborar presentaciones 3D de objetos 2D como lo son las presentaciones "ggplot()" y la asociación de ambas, "ggplot() + geom\_sf()"; por lo que las funciones recurrentes de la biblioteca son:

*plot\_gg()*: Transforma un objeto ggplot2 en 3D.

*render\_snapshot()*: Presenta al objeto dentro de una visualización 3D.

*save\_png()*: Guarda el objeto seleccionado en una imagen de formato PNG.

## Capítulo 5. Análisis conjunto de los resultados filtrados de la encuesta.

Puesto que parte de los objetivos principales de este trabajo es demostrar el proceso de análisis SIG mediante un lenguaje de programación, en este caso el lenguaje R, es importante saber de buena tinta que las herramientas de almacenamiento y elaboración de información vectorial están disponibles en código y que se puede prescindir de una plataforma o software para llevar a cabo los distintos procesos de exploración vectorial, es por esto que el análisis de uso de sustancias ilegales, marihuana, cocaína, crack, heroína, etc. Es un análisis que mostrará a diferentes niveles de territorio el número total de casos positivos de consumo ligado a variables como la edad y el promedio de edad de inicio de consumo entre los entrevistados, teniendo en cuenta que esta medida tendrá variaciones según el nivel geográfico, así como el tipo de sustancia.

### 5.1 El mapeo del uso de sustancias ilegales.

Como se mencionó en el Capítulo 3 los datos recopilados por el cuestionario individual de la ENCODAT 2016 – 2017 tienen representatividad a nivel localidad, la media de observaciones tomadas por entidad federativa ronda los 1 500 entrevistados de los cuales no todos han consumido drogas o presentan algún tipo de dependencia a cualquier sustancia ilegal, alcohol o tabaco; con el tipo de datos recopilados es posible derivar diversos factores que influyen en el consumo de sustancias ya sean legales o ilegales así como comparar estos datos con cifras de percepción de riesgo, exposición a la prevención de tal modo que manipulando estos datos de manera paralela al análisis territorial sea posible replicarlos tanto a una escala pequeña como lo es un municipio o a nivel nacional.

La ENCODAT 2016 -2017 y las antecesoras ENA en sus resultados principales incluyen un apartado de las sustancias ilegales más consumidas por región, norcentral, noroccidental, etc., como se menciona en el Capítulo 2; la representación cartográfica del uso de sustancias de este análisis toma en cuenta a nivel estatal el total de casos reportados de consumo para presentar estos datos como resultado sintetizado de la base de datos del cuestionario individual.

## 5.2 Distribución de los casos a Nivel Nacional.

Para generar la cartografía correspondiente al consumo de drogas ilegales a nivel nacional se utilizó la capa de información vectorial con la división política del territorio nacional a la cual se le aplicó un proceso para determinar el centroide de cada polígono asociado a cada uno de los estados con el fin de generar una capa de elementos puntuales compatibles con las geometrías poligonales, el proceso de generación de los baricentros se realiza con la función `st_centroid()` que pertenece a la biblioteca `simple features` mencionada en el capítulo anterior.

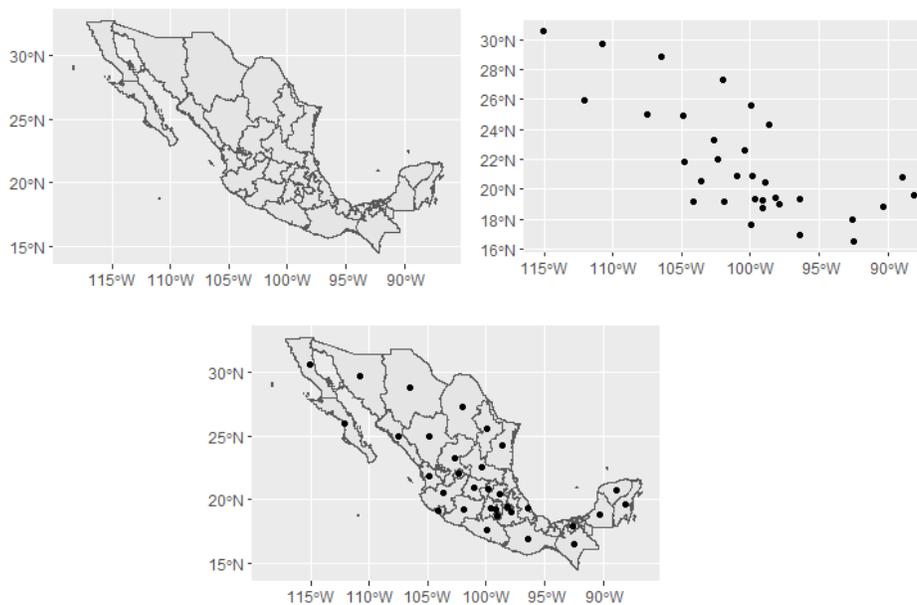


Figura 15: Capa correspondiente a la División Política de los Estados Unidos Mexicanos (arriba izquierda). Centroides generados mediante la función `st_read()` (arriba derecha). Presentación de ambas capas (abajo centro). Fuente: Elaboración propia.

Elaboración cartográfica.

El proceso de análisis pre cartográfico consistió en tomar las observaciones totales de la ENCODAT 2016 – 2017 sometiéndolas a un proceso de filtraje por categorías y condiciones lógicas para agruparlas por nivel territorial ejerciendo el método siguiente:

```
mar <- E1617 %>% filter(dia == 1 & ds2 == 1) %>% group_by(entidad) %>% summarise(dia=n())
```

Figura 16: Sección de código para el filtro de variables. Fuente: Elaboración propia.

Donde por cada variable que correspondiera a una sustancia ilegal (di1a - marihuana), se consideró que se tratara de un caso positivo de consumo "di1a == 1" contemplando al mismo tiempo que el sexo de la observación se tratase del sexo masculino "ds2 == 1" y/o femenino "ds2 == 2" agrupando por entidad el conteo de los casos que cumplieran con las condiciones mencionadas lo que arrojó como resultado distintos vectores (el objeto "mar") que contienen la condensación de los resultados al nivel solicitado.

	entidad	di1a
1	1	45
2	2	45
3	3	38
4	4	25
5	5	36
6	6	32
7	7	18
8	8	37
9	9	28
10	10	40

Figura 17: Vista parcial del vector "mar" con los casos positivos de consumo de marihuana, variable "di1a", agrupados por entidad. Fuente: Elaboración propia.

Con el uso de la función *left\_join()* que pertenece a la biblioteca *dplyr* se creó un nuevo *dataframe* el cual contiene la unión de todos los casos de consumo a nivel estatal.

```
Drogas_Edad_H <- Drogas_Edad_H %>% left_join(var1)
```

Figura 18: Función *left\_join()* para unir objetos con un campo en común, "var1" corresponde al nombre del vector que será unido.

	CVE_ENT	Entrevistados	di1a	di1b	di1c	di1d	di1e	di1f	di1g
1	1	732	118	34	12	6	16	1	9
2	2	843	163	53	12	5	9	5	31
3	3	802	113	51	15	9	9	3	17
4	4	706	99	36	13	NA	8	1	1
5	5	728	98	40	11	6	20	1	5

Figura 19: Tabla general de casos positivos de sustancias agrupadas por entidad. Fuente: Elaboración propia.

Asimismo, se unieron las capas vectoriales, poligonales y puntuales con las tablas generadas durante el análisis de los datos.

Para elaborar la cartografía de casos a nivel nacional se unió la tabla de casos de consumo positivos con la capa vectorial de información puntual generada anteriormente (Centroides.shp); además se realizó un proceso de filtraje para determinar la media de inicio de consumo de todos los casos positivos según el tipo de sustancia involucrada para que esta información se anexara a los atributos de la capa de información poligonal y se mostrara a la par de los casos de consumo.

En el primer mapa de la cartografía generada y también guía para la estructura, visualización e información contenida en la cartografía subsecuente se aprecian tres características esenciales:

1. Media de edad de inicio de consumo, distintiva por tipo de sustancia.
2. Casos de consumo por entidad federativa, particular por tipo de sustancia.
3. Etiquetas con el nombre de los estados que cumplen la condición establecida específica por mapa.



Mapa 1: Casos de Consumo de Marihuana en Hombres a Nivel Nacional, ver anexo amplificado.

Variables: "di1a" - ¿Ha usado, tomado o probado marihuana?, "di4a" - ¿Qué edad tenía cuando uso marihuana por primera vez?

En el mapa número 1, referente a los casos de consumo de marihuana en hombres se puede apreciar que en 16 estados se tiene registro de más de 100 consumidores estando dentro de estos estados la Ciudad de México y Nuevo León contenedores de las principales zonas metropolitanas del país.

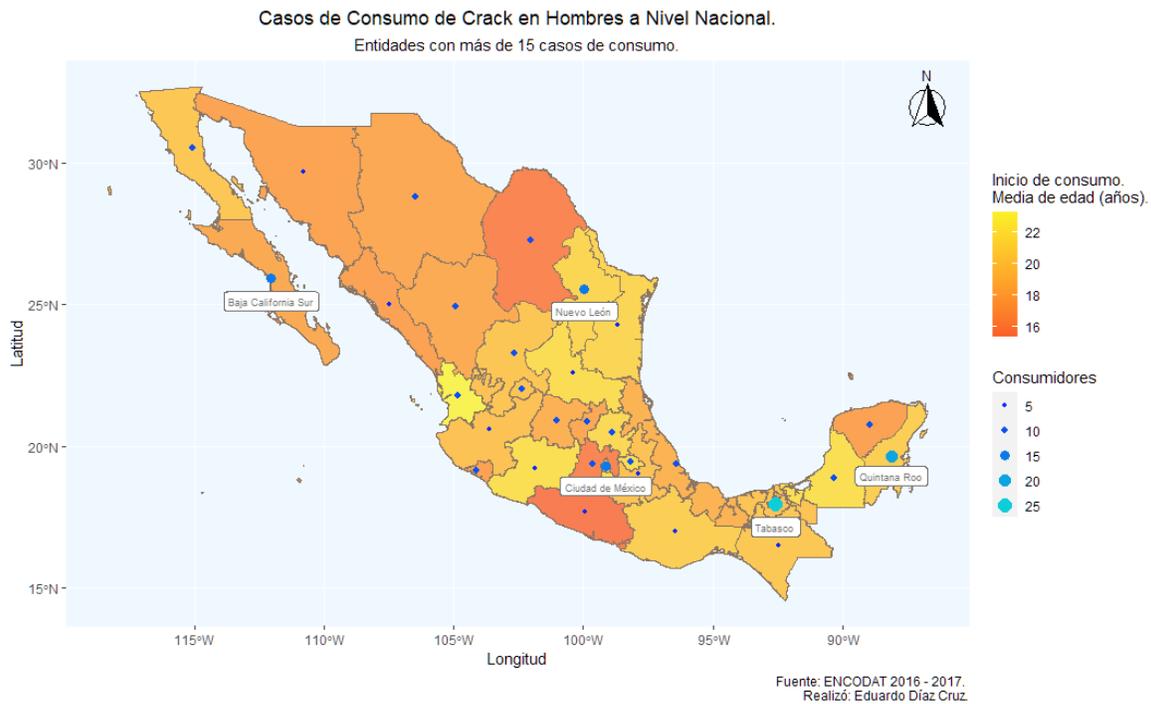
Por otra parte, en el Estado de Morelos se localiza el menor valor para la media de edad en el inicio de consumo de marihuana, 16.5 años, caso inverso para los estados de Veracruz y Nayarit con un valor de media de edad de 18.5 años, cabe resaltar el comportamiento del número de consumidores en el estado de Morelos ya que a una edad temprana de inicio de consumo se esperarían una mayor cantidad de consumidores.



Mapa 2: Casos de Consumo de Cocaína en Hombres a Nivel Nacional, ver anexo amplificado.

Variables: "di1b" - ¿Ha usado, tomado o probado cocaína?, "di4b" - ¿Qué edad tenía cuando uso cocaína por primera vez?

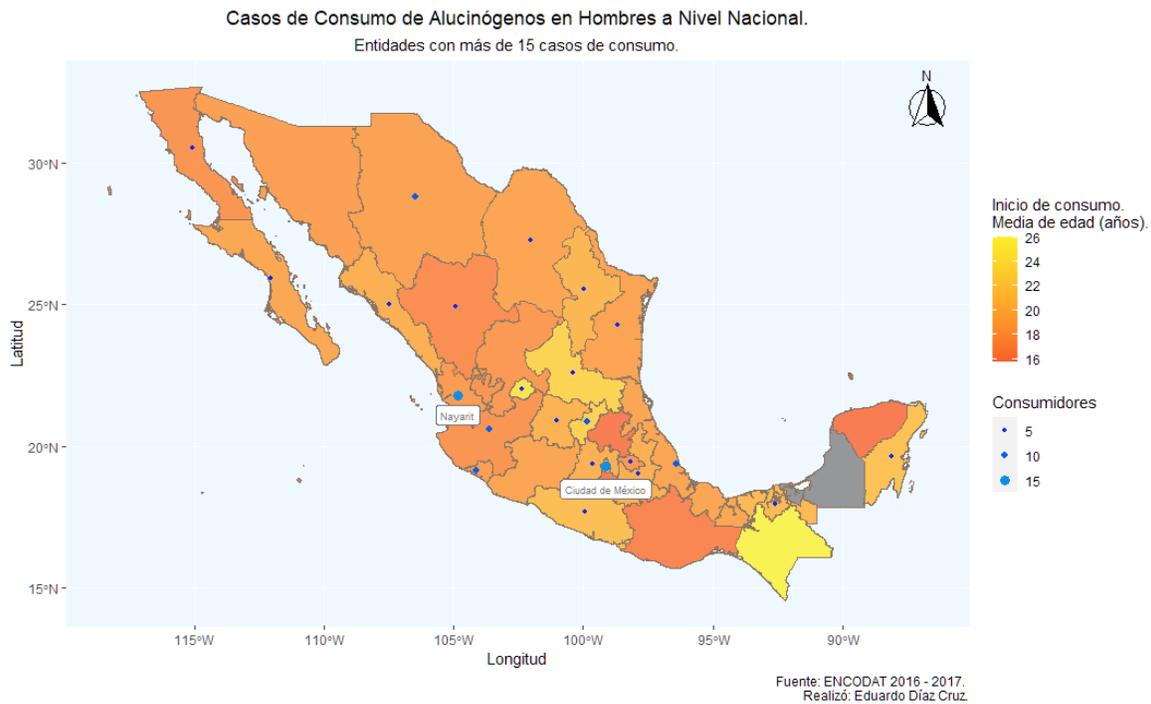
En el caso de la edad de inicio en el consumo de cocaína, se puede apreciar que el valor más bajo es de 18 años en el Estado de Puebla, seguido por el Estado de Oaxaca y Yucatán con una media de 19 años, hecho que contrasta con el Estado de Zacatecas donde la media de edad se localiza por encima de los 22 años. En cuanto a la cantidad de casos se observa una evidente proliferación de consumo en la zona noroeste del país con un par de registros importantes en la zona sureste del país, Quintana Roo y Tabasco, dejando a la zona centro con evidencia de consumo presente pero no superior a los cincuenta casos.



Mapa 3: Casos de Consumo de Crack en Hombres a Nivel Nacional, ver anexo amplificado.

Variables: "di1c" - ¿Ha usado, tomado o probado crack?, "di4c" - ¿Qué edad tenía cuando uso crack por primera vez?

Con la información contenida en el Mapa 3 es posible observar que la media de edad de inicio de consumo tiene el valor mínimo en los estados de Coahuila, Guerrero y México mientras que el único estado con una media de edad alta es Nayarit a pesar de que la cantidad de casos resgistrados esta dentro de la moda para el resto del país con 10 o menos casos. Por otra parte los estados con la mayor cantidad de casos de consumo son Quintana Roo, Tabasco, Baja California Sur y nuevamente la Ciudad de México y el Estado de Nuevo León.



Mapa 4: Casos de Consumo de Alucinógenos en Hombres a Nivel Nacional, ver anexo amplificado.

Variables: "di1d" - ¿Ha usado, tomado o probado alucinógenos?, "di4d" - ¿Qué edad tenía cuando uso alucinógenos por primera vez?

En cuanto al uso de alucinógenos es importante señalar que durante la aplicación de los cuestionarios se exentó a aquellos individuos pertenecientes a grupos étnicos por razones de susceptibilidad o sensibilidad en cuanto a los temas por lo que las observaciones aquí descritas son todas acerca de individuos que han consumido en experimentación o uso continuo y fuera de fines religiosos; en cuanto a las observaciones se puede notar que en gran parte del territorio el acercamiento o la edad de auge de las sustancias alucinógenas se encuentra por debajo de los 20 años y es la Ciudad de México y el Estado de Nayarit los estados que reportan un mayor número de casos, superior a 15.



Mapa 5: Casos de Consumo de Inhalables en Hombres a Nivel Nacional, ver anexo amplificado.

Variables: "di1e" - ¿Ha usado, tomado o probado inhalables?, "di4e" - ¿Qué edad tenía cuando uso inhalables por primera vez?

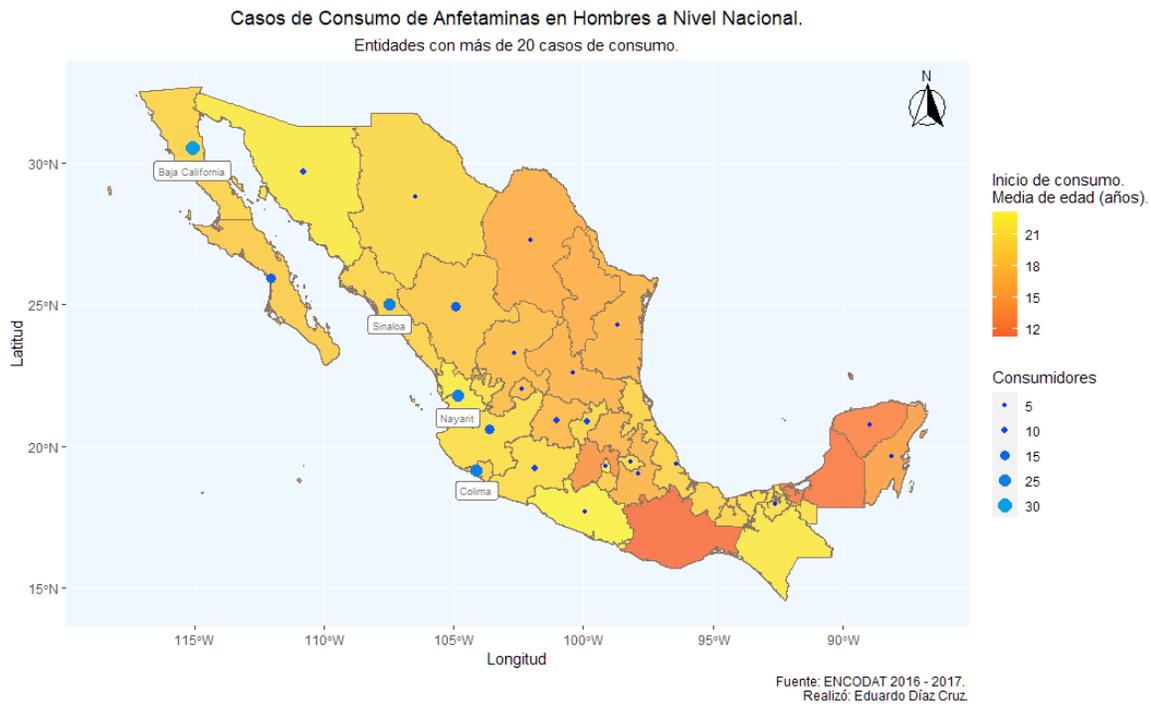
En el caso de los inhalables debe resaltarse que se ha encontrado, hasta ahora, el valor más bajo en cuanto a edad de inicio de consumo siendo esta menor de 15 años y con incidencia en nueve estados del país mientras que los estados restantes se encuentran por debajo de la media de 18 años, datos que matizan la baja edad de inicio de consumo; al mismo tiempo es importante señalar que son nueve los estados que presentan más de 15 casos de consumo entre los entrevistados y teniendo nuevamente a dos de los tres estados contenedores de las principales zonas metropolitanas del país con un número de consumidores dentro de los valores máximos para la colección de datos.



Mapa 6: Casos de Consumo de Heroína en Hombres a Nivel Nacional, ver anexo amplificado.

Variables: "di1f" - ¿Ha usado, tomado o probado heroína?, "di4f" - ¿Qué edad tenía cuando uso heroína por primera vez?

En el Mapa 6, referente a los casos de consumo de Heroína en hombres, se puede observar la distribución de los casos de consumo en la zona oeste y noroeste del país, dejando a varios estados centrales y del sur sin casos de consumo (color gris). También se puede apreciar un caso único en el estado de nuevo león donde solo se tiene una observación y la edad de inicio de consumo de esta se encuentra en el rango 5-10 años.



Mapa 7: Casos de Consumo de Anfetaminas en Hombres a Nivel Nacional, ver anexo amplificado.

Variables: "di1g" - ¿Ha usado, tomado o probado estimulantes anfetamínicos?, "di4g" - ¿Qué edad tenía cuando uso estimulantes por primera vez?

Con el Mapa 7 puede distinguirse una distribución de los casos en la zona oeste y noroeste del país, muy similar al Mapa 6, teniendo el máximo de casos de consumo (31) en el estado de Baja California. En cuanto a los valores de la media de edad para el inicio de consumo se observa que los estados de Oaxaca y Campeche registran el valor más bajo, seguidos del Estado de México y Yucatán a la par que ninguno de estos presenta una cantidad superior a los diez casos de consumo, datos que contrastan con el concepto de que a una menor edad de contacto habrían más casos de consumo.

La cartografía generada hasta este punto ha sido en dos dimensiones, sin embargo, cabe mencionar que existen procesos de presentación cartográfica mucho más atractivos y que resultan más intuitivos en proporción inversa a la información que despliegan; en software comercial como ArcGis se utiliza ArcScene como herramienta de manipulación 3D, el cual permite elaborar formas, contornos y adaptar diseños 2D a geometrías tridimensionales las cuales entregan como resultado una cartografía más instintiva; en software libre, de forma paralela, existen extensiones para QGis como "Qgis2threejs" que permite asignar valores de "altura" a un objeto vectorial según el contenido de su tabla de atributos para generar la versión en tres dimensiones de estos datos.

Mediante código, en lenguaje R, la modelación 3D es posible a través de la biblioteca "rayshader" la cual no es una herramienta especializada en procesos SIG pero que como parte de un lenguaje de bajo nivel es posible adaptarla para presentar información bidimensional en elementos 3D.

La evolución de los lenguajes de programación y la expansión de herramientas de código abierto han permitido contar con bibliotecas orientadas a los Sistemas de Información de Geográfica entregando opciones para el análisis espacial, intercambio de información entre código y bases de datos, aplicaciones web de localización, etc., sintetizando su funcionamiento en líneas de bajo nivel en lugar de softwares especializados, haciendo de la programación, no solo en R, un elemento robusto para el análisis de datos geospaciales.

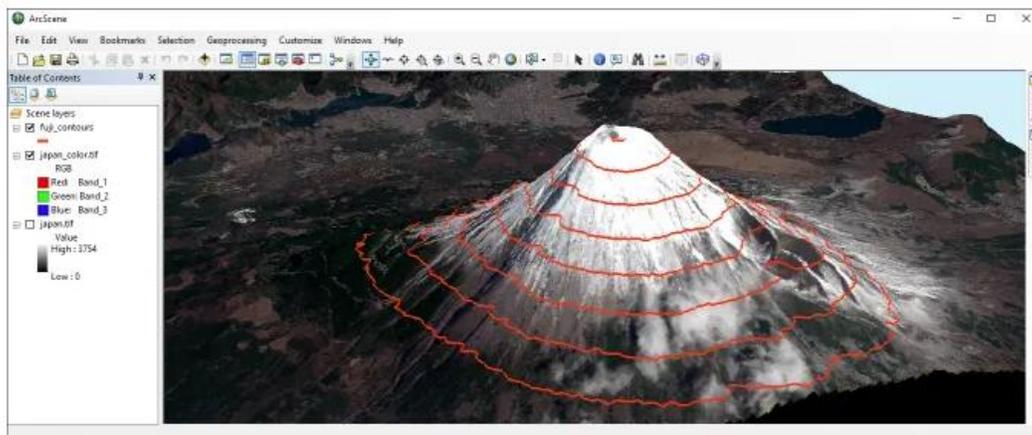


Figura 19: ArcScene, representación 3D del Monte Fuji. Fuente: <https://gisgeography.com/arcscene-arcglobe-3d-gis-software/>

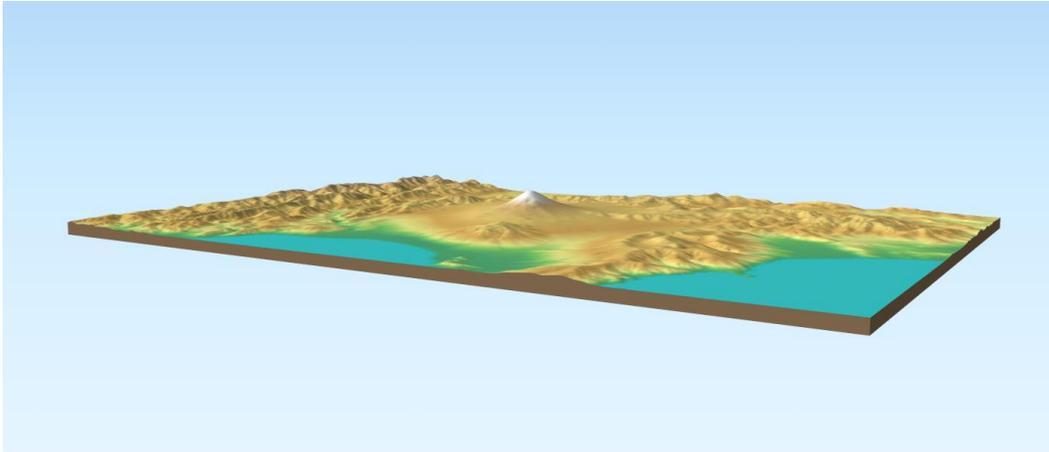
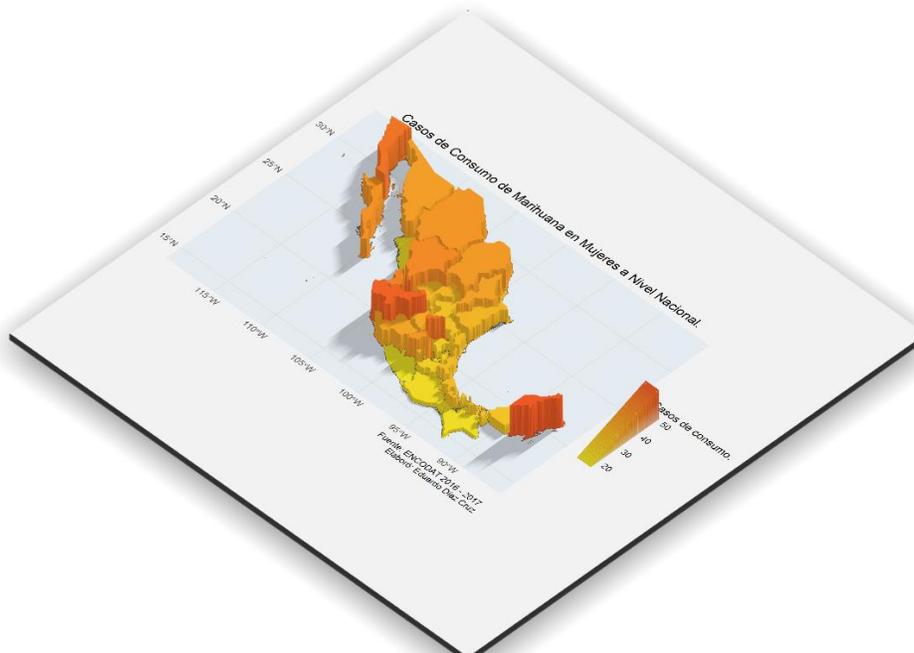


Figura 20: Sombras de relieve del Monte Fuji con datos SRTM3. Fuente:  
<https://qgis2threejs.readthedocs.io/en/docs/Examples.html>

Para la generación de cartografía referente al consumo de marihuana en mujeres a nivel nacional se implementó el uso de la biblioteca *rayshader* la cual, como se ha comentado, permite realizar presentaciones tridimensionales de elementos 2D, entre estos, información vectorial.

La variable de interés en la cartografía siguiente son los casos de consumo a nivel estatal, variable *ds1(a-g)*, marihuana, cocaína, crack, inhalables, etc. Por ende, en la cartografía de esta sección se considerarán dos características principales:

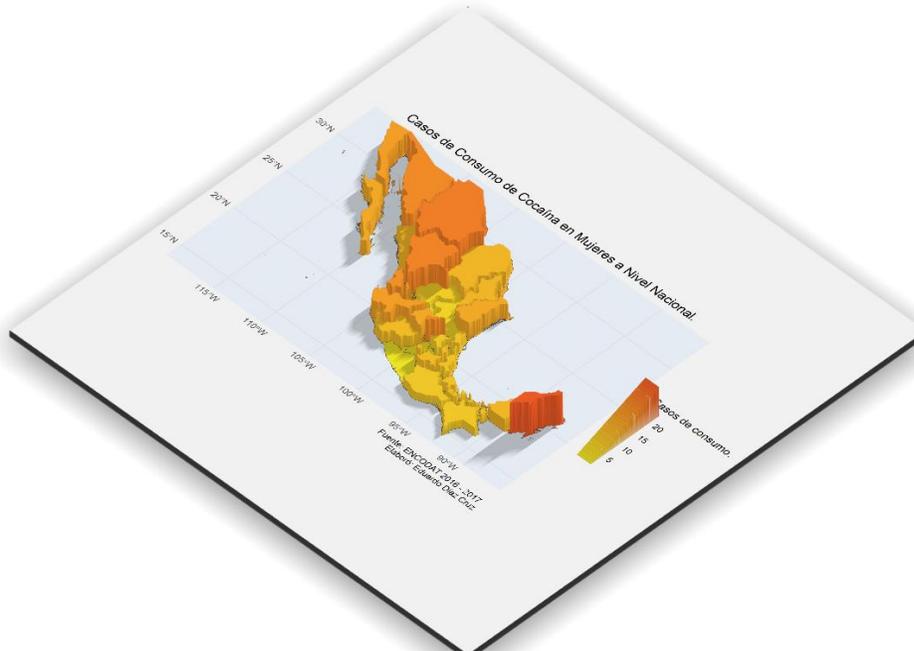
1. Una escala simbólica creciente en colores, de claros a oscuros, que muestran los incrementos de la variable de interés (casos de consumo).
2. La representación vectorial de los estados en elementos 3D que presentan con distintos niveles de altura la diferencia de valores entre los datos registrados, además de contener el color correspondiente al número de casos también especificado en la rampa simbólica.



Mapa 8: Casos de Consumo de Marihuana en Mujeres a Nivel Nacional, ver anexo amplificado.

Variables: "di1a" - ¿Ha usado, tomado o probado marihuana?

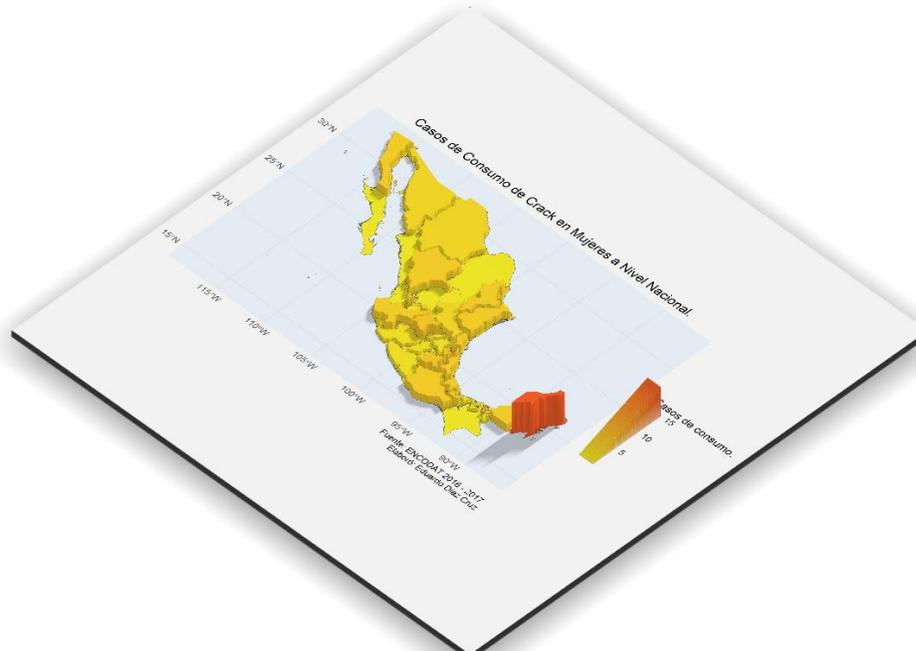
En el mapa 8, relativo a los casos de consumo de marihuana, se aprecia que el Estado de Quintana Roo posee el máximo número de casos de consumo a nivel nacional (51) seguido de los estados de Jalisco, Aguascalientes y Baja California mientras que en el panorama opuesto se encuentra el sur del país, donde la cantidad de casos disminuye considerablemente para los estados de Guerrero, Oaxaca y Chiapas, denotando un comportamiento creciente hacia el centro y el lado este del país.



Mapa 9: Casos de Consumo de Cocaína en Mujeres a Nivel Nacional, ver anexo amplificado.

Variables: "di1b" - ¿Ha usado, tomado o probado cocaína?

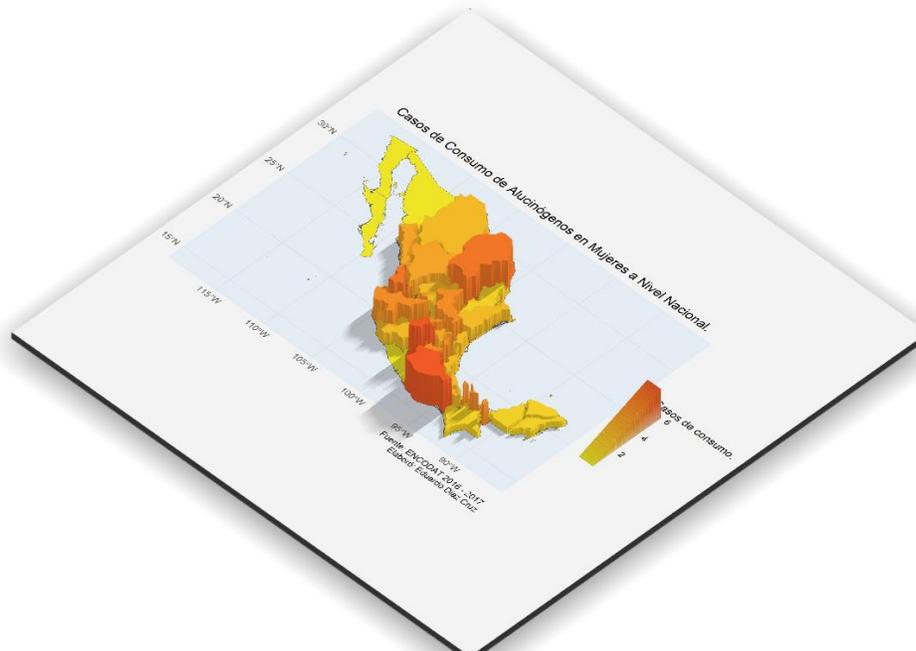
En cuanto al consumo de Cocaína se puede apreciar que el conteo máximo de casos de consumo se encuentra, nuevamente, en el estado de Quintana Roo; esta tendencia disminuye notablemente en el centro del país hasta llegar al estado de Guerrero y México donde se localizan los valores mínimos de consumo, posteriormente se puede apreciar en la cartografía que los valores de casos de consumo vuelven a incrementarse en la zona norte del país, concretamente en los estados de Chihuahua y Durango, con 18 y 17 casos respectivamente.



Mapa 10: Casos de Consumo de Crack en Mujeres a Nivel Nacional, ver anexo amplificado.

Variables: "di1c" - ¿Ha usado, tomado o probado crack?

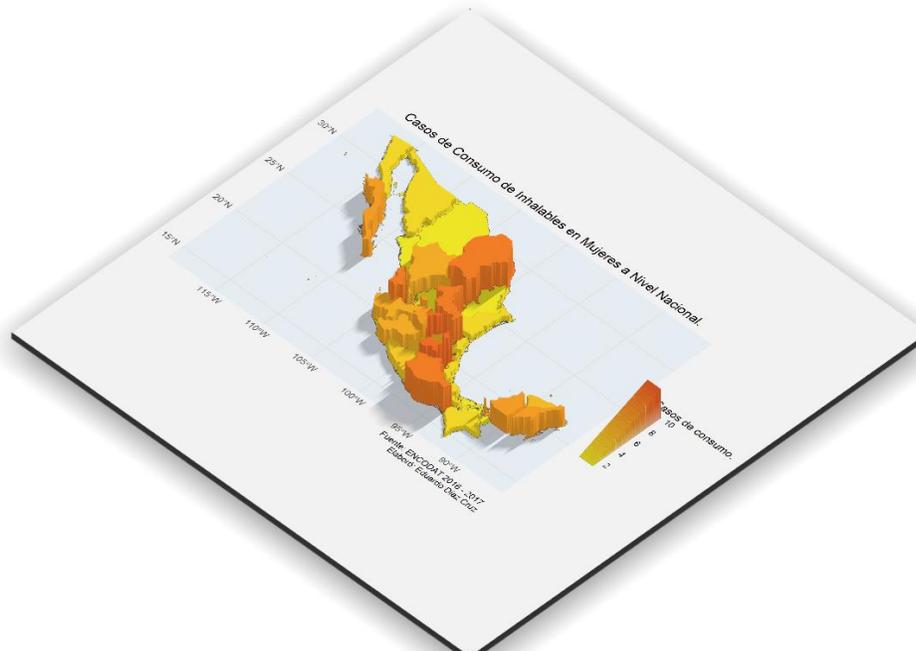
Referente al uso de crack se nota una caída en el número de casos de consumo en la totalidad del territorio sin superar los cuatro casos en Puebla y Guanajuato pero teniendo nuevamente y por tercera vez a Quintana Roo como el estado con más casos de consumo (15) de sustancias en mujeres (marihuana, cocaína y crack).



Mapa 11: Casos de Consumo de Alucinógenos en Mujeres a Nivel Nacional, ver anexo amplificado.

Variables: "di1d" - ¿Ha usado, tomado o probado alucinógenos?

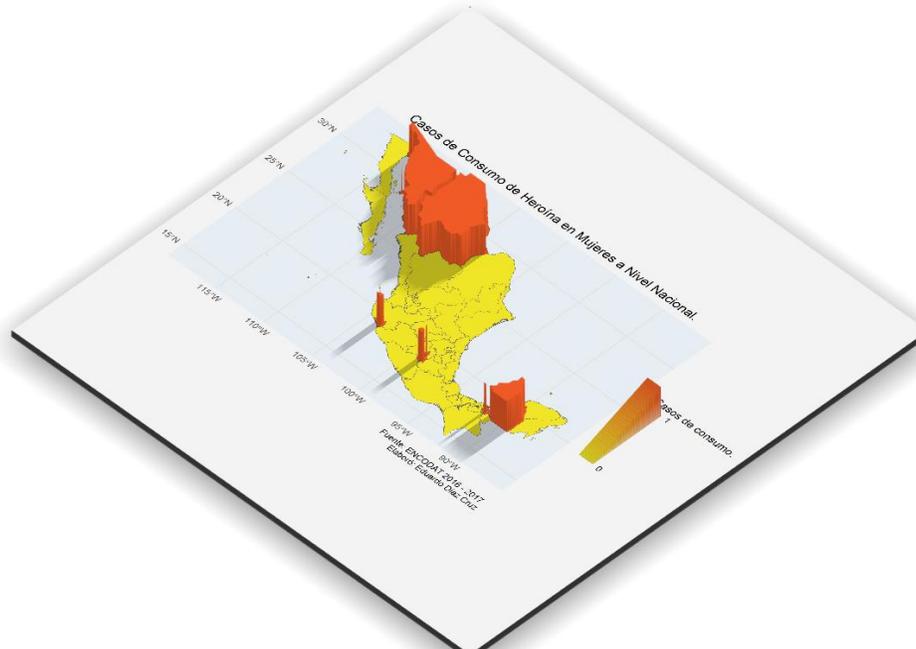
Como se mencionó anteriormente, para la recolección de datos se aplicó un filtro para evitar encuestar a poblaciones que consideren el tema como sensible por lo que en ambos mapas de consumo de alucinógenos (hombres y mujeres) la información desplegada corresponde a individuos fuera de los grupos étnicos y sociales considerados. Por otra parte, en el mapa 11, podemos visualizar que el mayor número de casos de consumo de alucinógenos se encuentra en el estado de Oaxaca, seguido del estado de México, Coahuila y Nayarit, tonos claros según la escala simbólica utilizada. Además, en comparativa, podemos apreciar también que el estado de Quintana Roo no registró valores significativos en el uso de alucinógenos como venía presentando con las sustancias anteriores.



Mapa 12: Casos de Consumo de Inhalables en Mujeres a Nivel Nacional, ver anexo amplificado.

Variables: "di1e" - ¿Ha usado, tomado o probado inhalables?

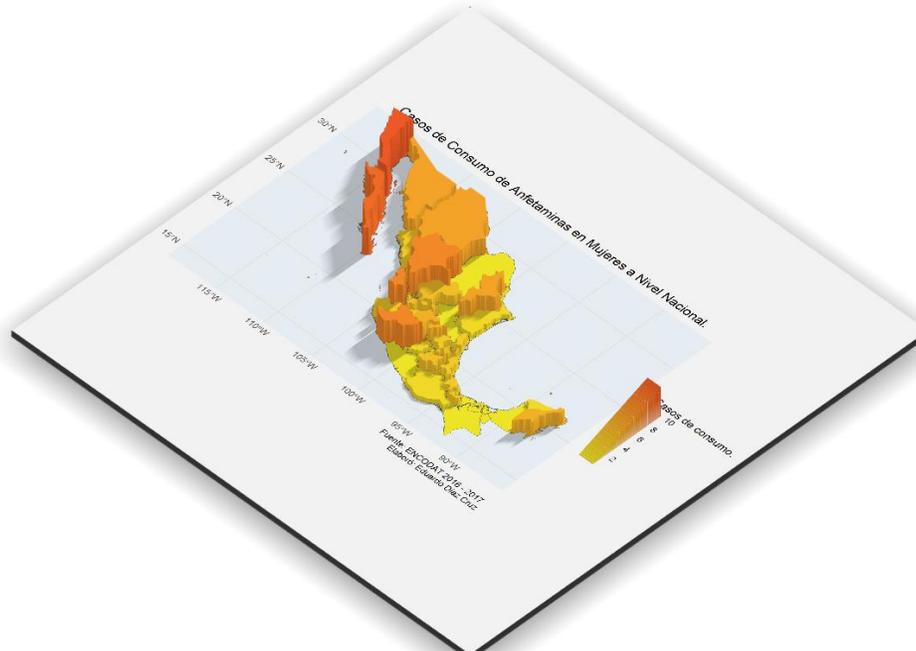
Con relación al consumo de inhalables el estado que registró la mayor cantidad de casos de uso es la Ciudad de México con 10, seguido del estado de Querétaro y el estado de Puebla con 9 casos para ambos. Aunque el comportamiento de los casos máximos se encuentran distribuidos en tres estados es relevante mencionar que los valores en los estados restantes se comportan dentro de la moda estadística (5).



Mapa 13: Casos de Consumo de Heroína en Mujeres a Nivel Nacional, ver anexo amplificado.

Variables: "di1f" - ¿Ha usado, tomado o probado heroína?

Para la variable "di1f" correspondiente al uso de heroína se registraron tan solo 5 casos distribuidos en cinco estados a lo largo de toda la República Mexicana sin alguna relación por ubicación. La baja cantidad de observaciones acerca del consumo de heroína puede ser explicado por las características mismas de los consumidores debido a que el uso de heroína esta relacionado a poblaciones sin hogar fijo, indigentes o con problemas de policonsumo (Fleiz-Bautista, 2019), grupos que se encuentran fuera de los parametros para la aplicación del cuestionario.



Mapa 14: Casos de Consumo de Anfetaminas en Mujeres a Nivel Nacional, ver anexo amplificado.

Variables: "di1g" - ¿Ha usado, tomado o probado estimulantes anfetamínicos?

En el mapa 14, Casos de Consumo de Anfetaminas en Mujeres a Nivel Nacional se aprecia como el uso de este tipo de sustancias tiene una mayor ocurrencia en la zona noroeste del país donde Baja California encabeza la cantidad de casos con 10, seguido de Baja California Sur con 9, teniendo también registros mínimos de uso en Tamaulipas, Veracruz o Yucatán, mientras que en el sur los casos son menores exepctuando al estado de Quintana Roo.

## Capítulo 6. Percepción y prevención del riesgo ante sustancias ilegales en la principal zona metropolitana del país. El nivel educativo como factor.

El consumo de sustancias legales o ilegales ha sido por mucho tiempo tópicos de distintos programas gubernamentales, campañas de promoción anti consumo en televisión y radio además de la continua exposición de estos temas durante la vida académica de estudiantes como mecanismo de prevención, es por esto que en este capítulo se analizará el impacto de las jornadas de prevención en la percepción de riesgo ante sustancias como la marihuana, tabaco y alcohol, siendo el factor principal el nivel educativo de los entrevistados.

Es importante definir el territorio de estudio debido a que gran parte de la población laboralmente activa vive distribuida alrededor de una zona metropolitana, cualquiera que sea su ubicación. Una zona metropolitana se define como el conjunto de dos o más municipios donde se localiza una ciudad de 50 mil o más habitantes cuyas funciones y actividades rebasan el límite del municipio de ubicación original. (INEGI, 2014)

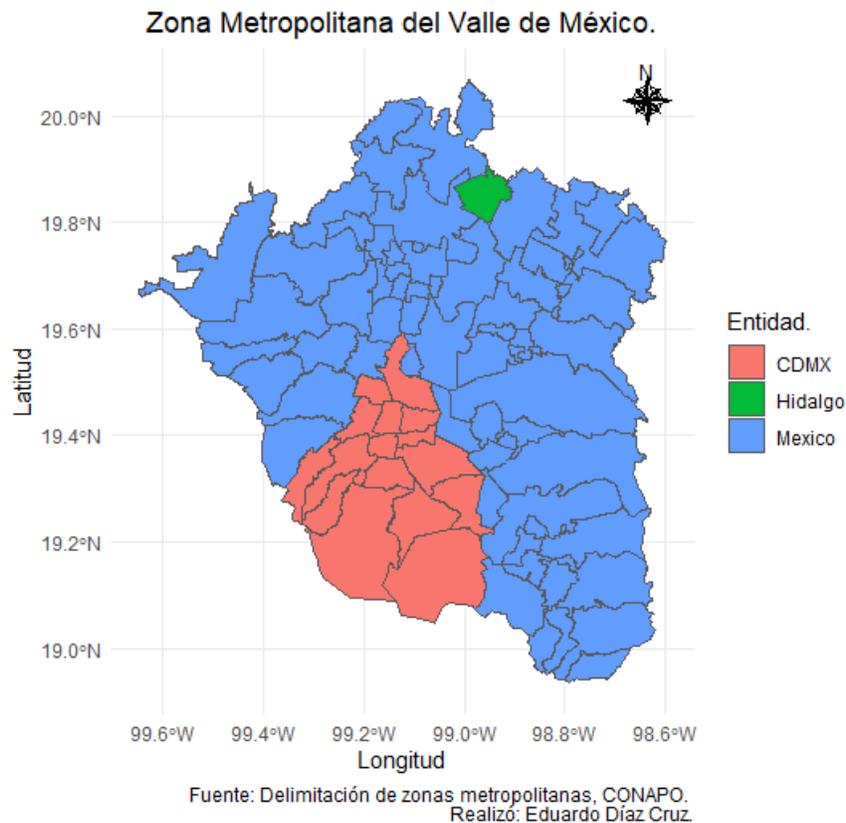
### Zonas metropolitanas de la República Mexicana, 2013

Zona metropolitana	Población <sup>3</sup>		Personal ocupado total <sup>4</sup>		Producción bruta total <sup>4</sup>	
	Absoluto	%	Absoluto	%	Absoluto	%
<b>Total nacional</b>	<b>110 991 953</b>	<b>100.0</b>	<b>21 576 358</b>	<b>100.0</b>	<b>13 984 313 218</b>	<b>100.0</b>
Valle de México	19 768 740	17.8	5 083 414	23.6	3 675 453 780	26.3
Monterrey	4 052 168	3.7	1 314 638	6.1	1 254 493 911	9.0
Guadalajara	4 364 069	3.9	1 142 465	5.3	567 860 604	4.1
Puebla- Tlaxcala	2 684 381	2.4	582 765	2.7	409 924 981	2.9
Toluca	1 911 175	1.7	392 291	1.8	325 740 390	2.3
Saltillo	811 980	0.7	240 059	1.1	324 852 696	2.3
León	1 600 750	1.4	464 124	2.2	270 423 266	1.9
Querétaro	1 085 796	1.0	355 386	1.6	267 694 866	1.9
Tehuantepec	738 739	0.7	29 938	0.1	219 631 638	1.6
San Luis Potosí- Soledad de Graciano Sánchez	160 728	0.1	289 427	1.3	215 176 286	1.5

Tabla 14: Principales zonas metropolitanas del país. Fuente: Las zonas metropolitanas en México, mini monografía INEGI, 2014.

La Zona Metropolitana del Valle de México, ZMVM.

La ZMVM es el centro político y económico de México, de acuerdo con datos de la OCDE (Organización para la Cooperación y el Desarrollo Económico) es la tercera zona metropolitana más grande dentro de la organización y la zona metropolitana más grande del mundo fuera de Asia; abarca casi 7 866 km<sup>2</sup> y comprende las 16 Alcaldías de la CDMX, 59 municipios del Estado de México y un municipio del Estado de Hidalgo, en datos de población y trabajo la ZMVM representa el 17% de los habitantes y el 18% de los empleados a nivel nacional además de generar casi el 23% del PIB (OECD Publishing, 2015).



Mapa 15: Zona Metropolitana del Valle de México. Fuente: Elaboración propia.

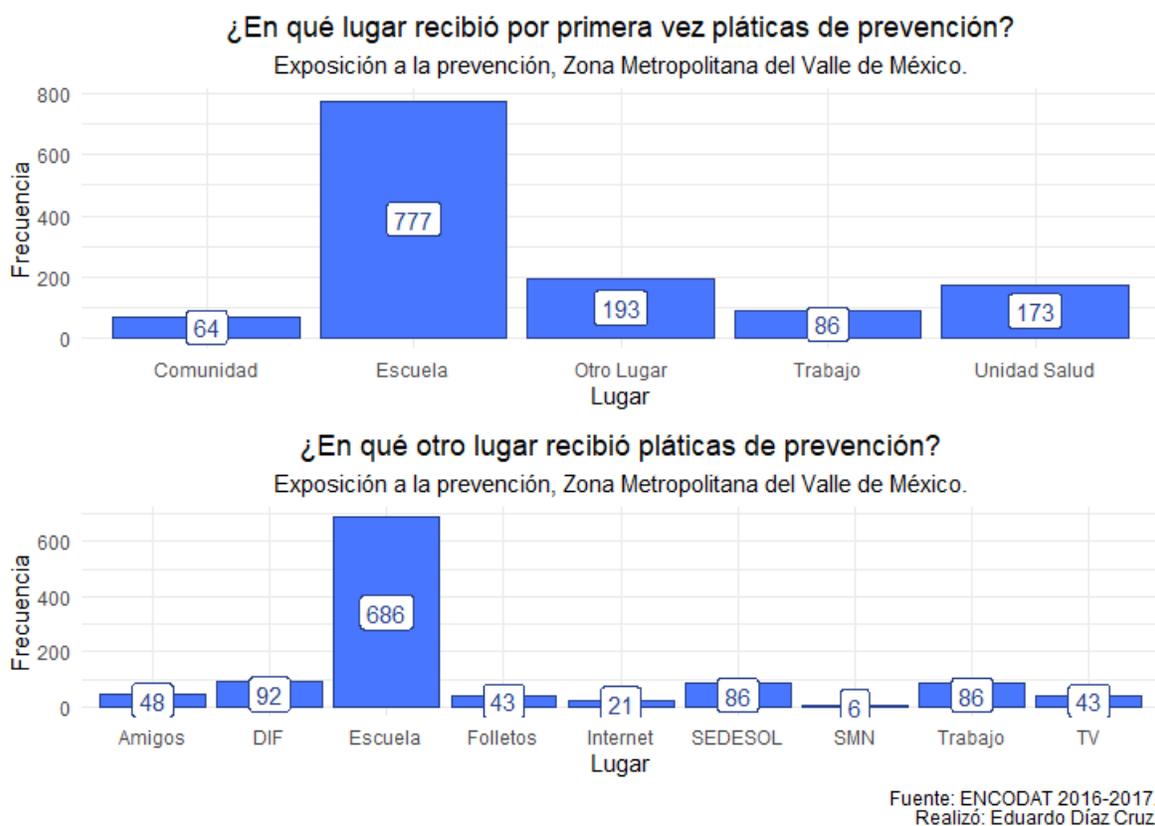
## 6.1 Inclusión en programas de prevención en la ZMVM.

En el cuestionario individual de la ENCODAT 2016 – 2017 en la sección de prevención se encuestó acerca de los lugares donde se recibió por primera vez pláticas de prevención, además de la pregunta subsecuente sobre otros lugares donde se participó en actividades de prevención. Lo anterior corresponde a las variables:

pc3 – ¿En qué lugar recibió por primera vez las pláticas o participó en alguna actividad para prevenir el uso de alcohol o drogas?

pc6[a-i] - ¿En qué otro lugar ha recibido pláticas o ha participado en actividades para prevenir el uso de alcohol o drogas? ¿Quién se la ha proporcionado? a – Televisión o radio, b – Internet o redes sociales, c – Amigos o familiares, d – Difusión Impresa, e – DIF, f – Escuela, g – Trabajo, h – SMN, i – SEDESOL.

Datos con los que se generaron las gráficas siguientes:



Gráfica 6: Inclusión en programas de prevención en la ZMVM. Fuente: Elaboración propia.

Teniendo esta información se puede asegurar que el principal medio de acercamiento para la prevención es mediante pláticas durante la etapa educativa, afirmación que se tomó en cuenta para dirigir un método de depuración con el fin de conocer la perspectiva de riesgo en los entrevistados según tres condiciones relacionadas al nivel educativo:

- Sin ningún grado de educación o educación primaria incompleta.
- Educación media superior, primaria completa a Bachillerato completo.
- Educación superior, estudios universitarios incompletos o más.

### 6.1.1 Variable ds9 – Grado de estudios.

La variable ds9, que corresponde a la pregunta del cuestionario individual de la ENCODAT 2016 – 2017: ¿Cuál fue el último grado que ha cursado o cursó en la escuela? Ofrece un total de diez categorías:

ds9	1	Primaria incompleta (1 a 5 años)
	2	Primaria completa (6 años)
	3	Secundaria incompleta (1 a 2 años)
	4	Secundaria completa o equivalente (3 años)
	5	Bachillerato incompleto (1 a 2 años)
	6	Bachillerato completo o equivalente (aprox. 3 años)
	7	Estudios Universitarios incompletos (1 a 3 años)
	8	Estudios Universitarios completos (4 a 5 años)
	9	Estudios de Posgrado (2 a 4 años)
	99	No contesta

Tabla 15: Valores de la variable ds9. Fuente: Definición de variables ENCODAT 2016 – 2017.

Con las cuales se realizó el filtro para obtener nuevas tablas siguiendo nuestros propósitos y con las condiciones mencionadas anteriormente.

```
#nunca ha ido a la escuela ds8 o primaria incompleta ds9
educacion_nula <- E1617_zmvm %>% filter(ds8==1 | ds9==1)

#ya no estudian _yne
#educación media superior
educacion_ms_yne <- E1617_zmvm %>% filter(ds9>=2 & ds9<=6 & ds8==2)
#educación superior
educacion_s_yne <- E1617_zmvm %>% filter(ds9>=7 & ds8==2)

#aún estudian _ae
#educación media superior
educacion_ms_ae <- E1617_zmvm %>% filter(ds9>=2 & ds9<=6 & ds8==3)
#educación superior
educacion_s_ae <- E1617_zmvm %>% filter(ds9>=7 & ds8==3)
```

Figura 21: Sección de código, condiciones de reducción en la variable "ds9", último grado de estudios y "ds8" ¿Actualmente estudia? Fuente: Elaboración propia.

En el tratamiento de los datos con lenguaje R, se creó el objeto *E1617\_zmvm* que corresponde a las observaciones totales de la ENCODAT 2016 – 2017 únicamente para la Zona Metropolitana del Valle de México del cual se removieron aquellas observaciones con valor *99 – no contesta*, para posteriormente aplicar las condiciones necesarias a la variable *ds9* para generar cinco nuevos objetos correspondientes a los niveles: sin ningún grado de educación o básico incompleto, educación media superior y educación superior, estos dos últimos con las variantes "aún estudia" y "ya no estudia".

El procedimiento de separar las observaciones totales de la ENCODAT 2016 – 2017 para la Zona Metropolitana del Valle de México mediante la variable *ds9* y *ds8* devolvió los datos distribuidos de la siguiente manera:

Grado de estudios.	Obs. Aún estudia.	Obs. Ya no estudia.
Sin ningún grado de educación o básico incompleto.	-	130
Educación media superior.	562	1695
Educación Superior.	95	407

Tabla 16: Resultados de la reducción con las variables *ds8* y *ds9*. Fuente: Elaboración propia.

## 6.2 Percepción de riesgo.

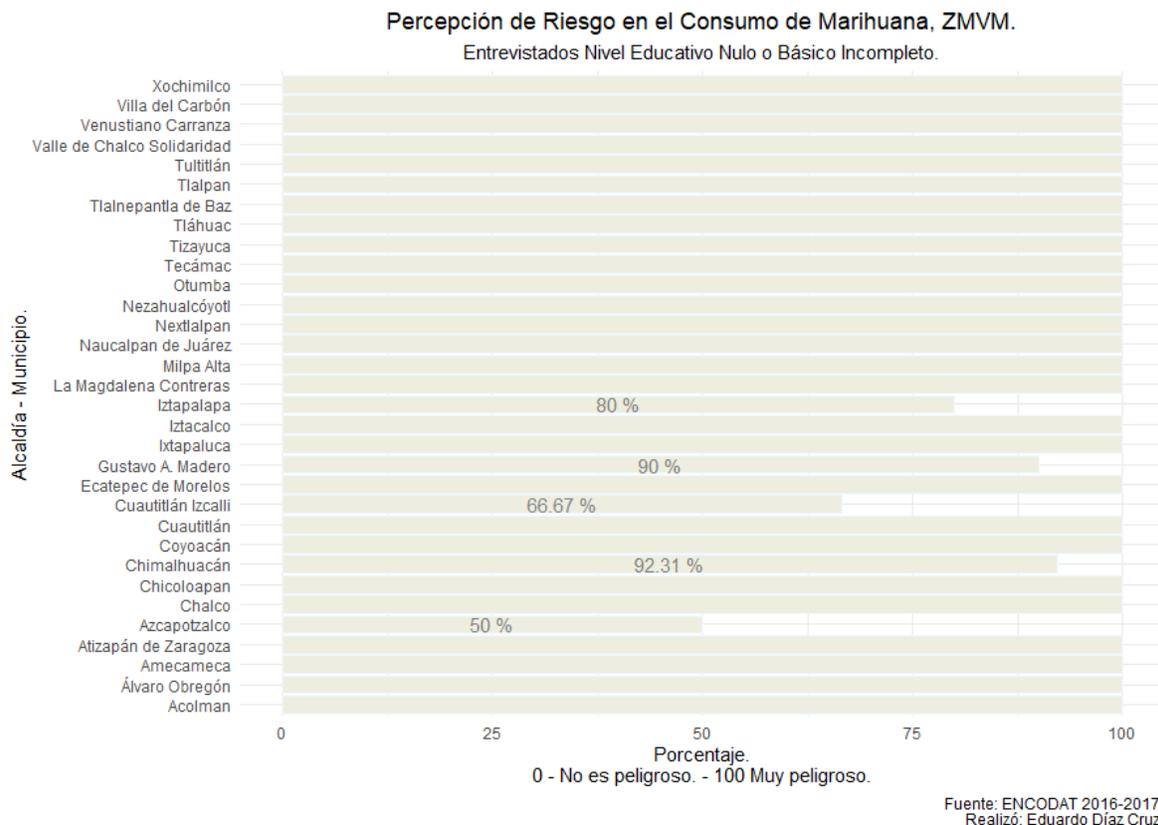
La relevancia del estudio de la percepción de riesgo asociado al consumo de sustancias está basada en la consideración del mismo como un factor de protección frente al inicio y continuidad del consumo por lo que la percepción de riesgo es entendida como el grado en que se atribuye a un determinado tipo de conducta una amenaza para la salud (Becoña, 2007), en este caso el uso de sustancias.

De forma paralela, en informes internacionales, los resultados acerca de la percepción de riesgo entre sexos han encontrado aspectos similares en la apreciación de peligro coincidiendo en que la percepción de riesgo es mayor entre las mujeres que entre los hombres para todas las sustancias psicoactivas y para todas las frecuencias de uso (Armstrong, 2014), lo que deja a un análisis por sustancia (marihuana, tabaco y alcohol) un espacio para generalizar el punto de vista biológico-sexual y concentrar el estudio en factores donde se utilice el nivel educativo al que se ha tenido acceso como variable discrepante, ya que se sabe que el comienzo del uso de sustancias ocurre en su mayor parte durante la adolescencia como un proceso de aprendizaje vinculado a la dinámica cultural, donde las características del individuo y de las sustancias consumidas determinan la interacción que consolida ese aprendizaje. (Berruecos Villalobos, 2007).

De este modo, como objetivo particular, se analizarán los datos relativos a la impresión que han dejado las campañas de prevención en personas que detuvieron o concluyeron sus estudios y en aquellas otras que continúan estudiando.

### 6.2.1 Variable ts18 – Marihuana.

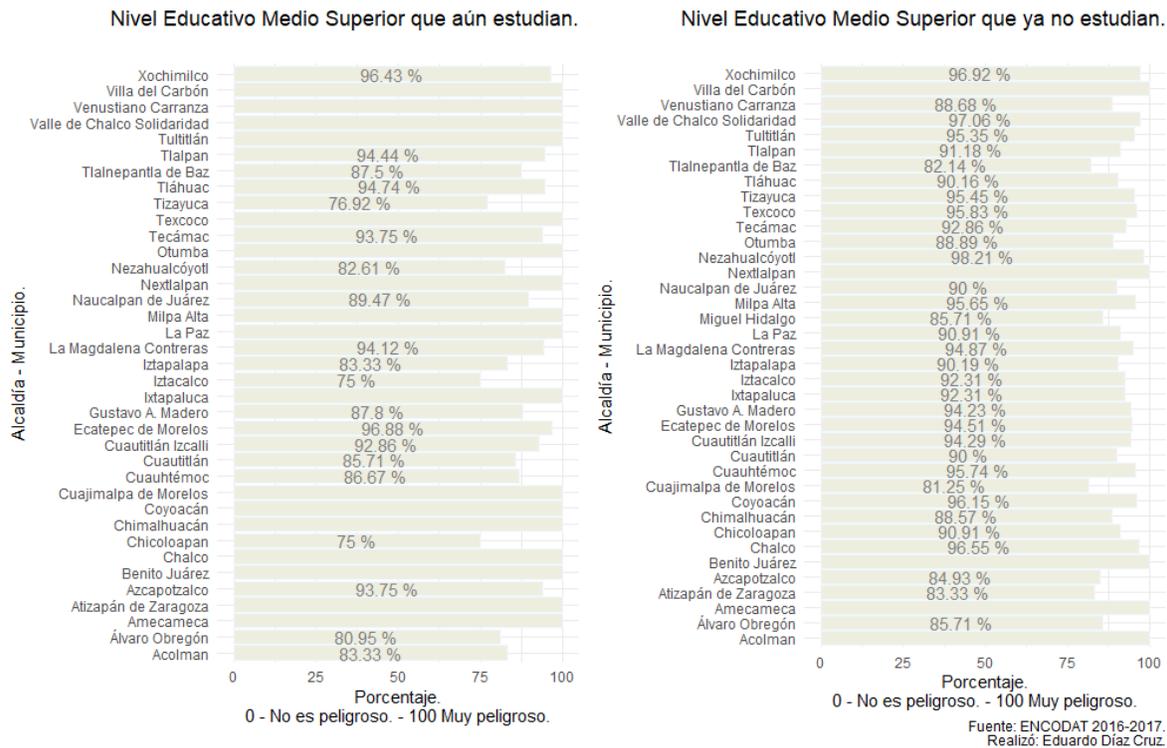
Por mucho tiempo, el consumo de marihuana ha sido considerado el primer paso al consumo de drogas más fuertes y una consecuencia del consumo de otras drogas “blandas” como el alcohol y el tabaco (Secades-Villa, 2015) y aunque no hay una evidencia certera de que lo sea o que cumpla esta función si es conocido que puede resultar un agravante de la dependencia al alcohol (Weinberger, 2016) además de estar asociado a otros trastornos como la dependencia a la nicotina.



Gráfica 7: Percepción de Riesgo en el consumo de Marihuana en la ZMVM; Nivel Educativo Nulo o Básico Incompleto. Fuente: Elaboración propia.

En el grupo de entrevistados de nivel educativo nulo o básico incompleto para la Zona Metropolitana del Valle de México se encontró que gran parte de los entrevistados coinciden en que el consumo de marihuana representa un peligro mayor a la salud, es tan solo en dos municipios del Estado de México (Cuautitlán Izcalli y Chimalhuacán) donde la apreciación de peligro no es una constante entre los encuestados; mientras que para la Ciudad de México existe una opinión dividida respecto al consumo en la Alcaldía Azcapotzalco y una tendencia a pensar que no existe, en gran medida, tal riesgo para la salud en dos alcaldías más, Gustavo A. Madero e Iztapalapa.

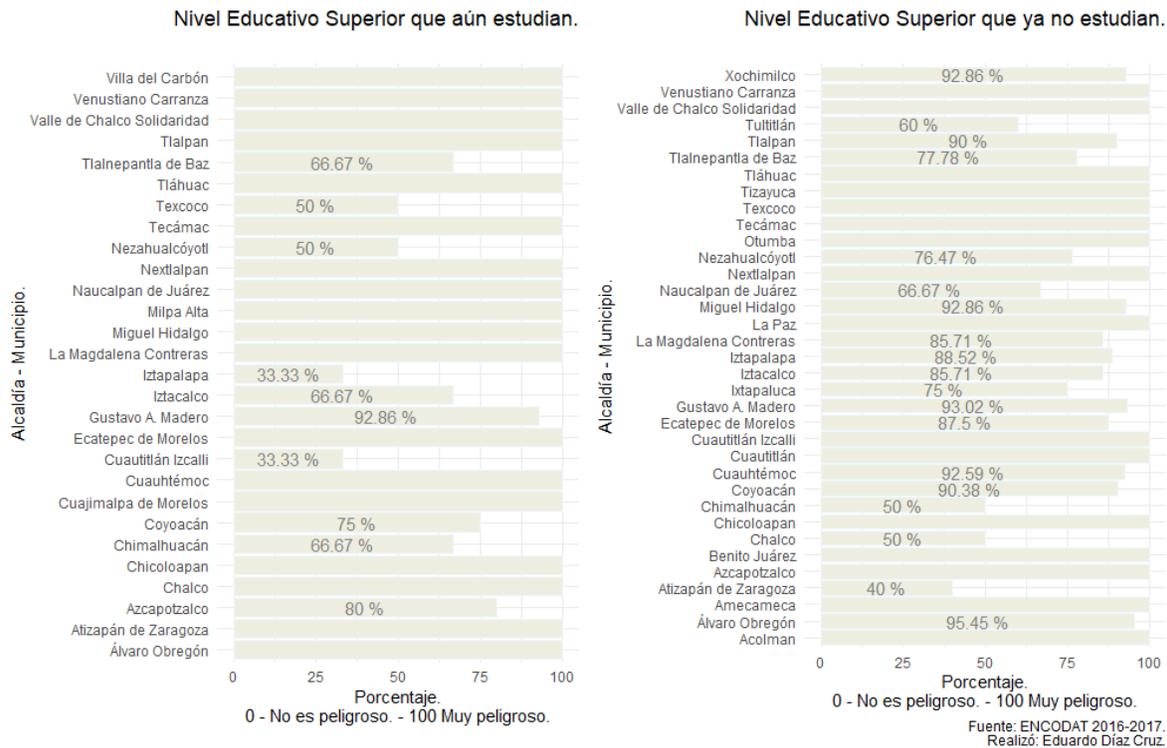
Percepción de Riesgo en el Consumo de Marihuana, ZMVM.



Gráfica 8: Percepción de Riesgo en el consumo de Marihuana en la ZMVM; Nivel Educativo Medio Superior. Fuente: Elaboración propia.

Siguiendo lo establecido y luego de dividir a los grupos de nivel educativo en dos sectores (los que aún estudian y los que ya no estudian) es posible a primera vista notar 1) la variación de opiniones en el grupo de educación medio superior que ya no estudia contra el grupo que aún estudia, 2) en el grupo que aún estudia se puede observar una menor cantidad de alcaldías y/o municipios con observaciones a la vez que también posee una mayor concentración de resultados en donde se considera muy peligroso el consumo de marihuana; para entender esto se calculó la media general de la percepción, es decir, el promedio de la apreciación de peligro de todas las observaciones a nivel municipal, siendo 93.28 % para el nivel medio superior que aún estudia y 92.65 % para el grupo que ya no estudia; con esto se puede afirmar que la percepción de riesgo es mayor entre los individuos que se encuentran estudiando en comparación con los que han concluido sus estudios, así mismo se calculó la media de edad de ambos grupos siendo de 17.39 años para el grupo que estudia y de 40.74 para el grupo que cuenta con educación hasta el nivel medio superior, denotando el impacto que tienen las campañas durante estos niveles educativos a largo plazo pues es en esta etapa de la formación donde las pláticas de prevención son obligatorias y/o rutinarias en contraste con el nivel superior donde las medidas comienzan a ser desacostumbradas.

Percepción de Riesgo en el Consumo de Marihuana, ZMVM.



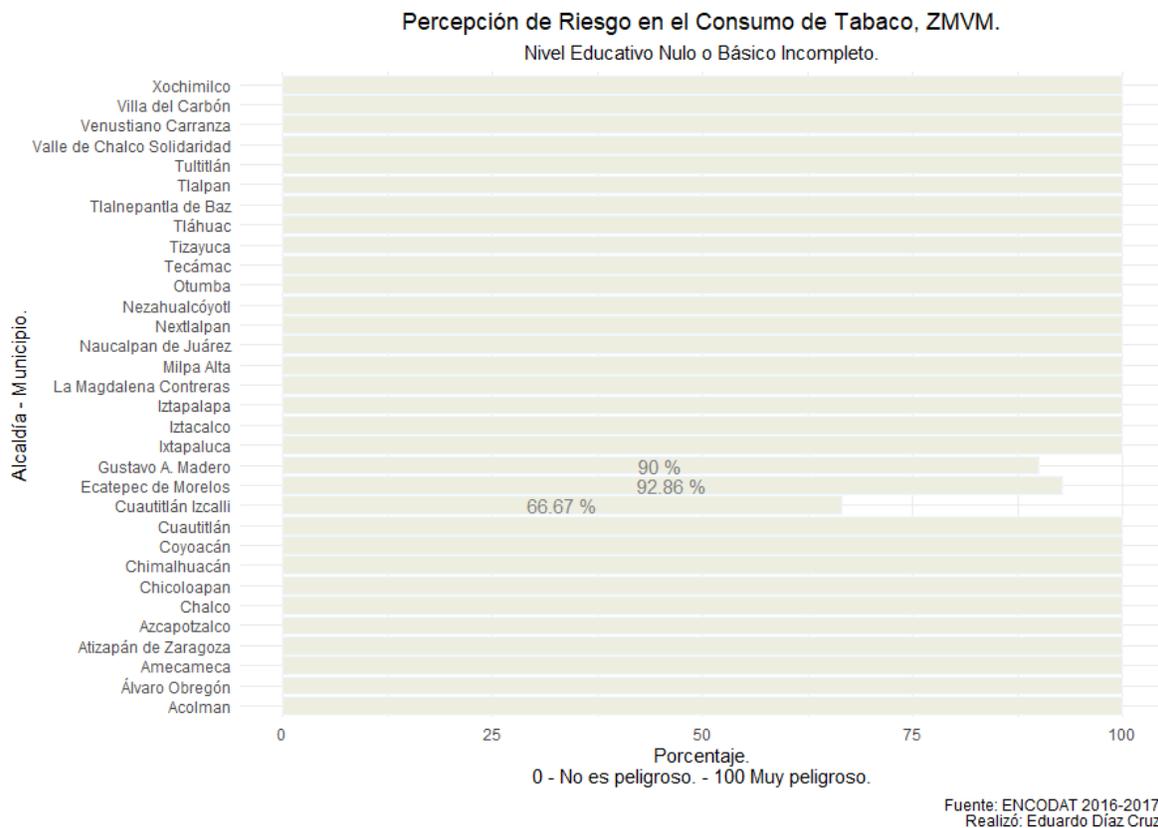
Gráfica 9: Percepción de Riesgo en el consumo de Marihuana en la ZMVM; Nivel Educativo Superior.  
Fuente: Elaboración propia.

En el nivel educativo superior, que comprende de estudios universitarios inconclusos a estudios de posgrado o más, existe un contraste evidente en las opiniones de ambos grupos, teniendo más observaciones en aquellos que han concluido sus estudios en comparación con los que continúan estudiando, 407 y 91 respectivamente, datos que tienen impacto en la cantidad de alcaldías y/o municipios mostrados en la gráfica, aunado a esta disparidad de observaciones es el grupo que aún estudia quien muestra una menor apreciación de riesgo ante el consumo, 86.23 % mientras que la media de los que han concluido sus estudios indica un 88.30 % general.

### 6.2.2 Variable ts23 – Tabaco.

El tabaco, al ser una droga legal, tiende a pasar desapercibida como un problema de abuso de sustancias pues es común observar personas fumando en la calle o en los espacios destinados específicamente a esta actividad. Al mismo tiempo y de una forma similar al alcohol, el consumo de tabaco crece favoreciéndose de factores de riesgo socioculturales como publicidad atractiva en los medios de comunicación y la sociedad de consumo (Vargas Pineda, 2001).

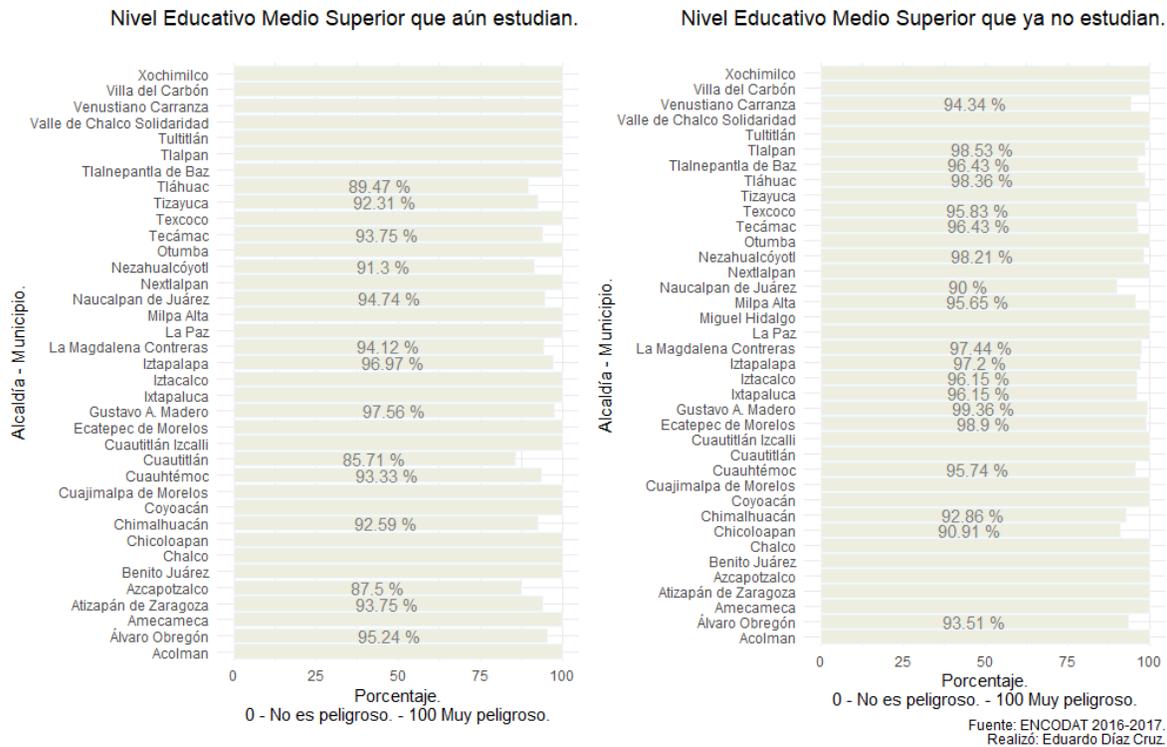
En México, la ENA 2002 determinó que entre la población de 12 a 65 años el 23.5 % de los entrevistados fumaba, mientras que para el 2008 el consumo de cigarrillos en áreas urbanas era del 20.4 %, 29.9 % para hombres y 11.8 % para mujeres, valores que confirman que la prevalencia del consumo de tabaco en hombres duplica o sobrepasa al consumo en mujeres (Waters H, 2010).



Gráfica 10: Percepción de Riesgo en el consumo de Tabaco en la ZMVM; Nivel Educativo Nulo o Básico. Fuente: Elaboración propia.

La percepción de riesgo en el consumo de tabaco en el nivel educativo nulo o básico incompleto demostró que entre los municipios y alcaldías de las que se tienen observaciones, existe una apreciación de riesgo general del 98.42 % siendo casi unánime la noción de peligro asociado al consumo de tabaco en la Ciudad de México, opinión muy parecida al rechazo del consumo de marihuana en el mismo nivel educativo.

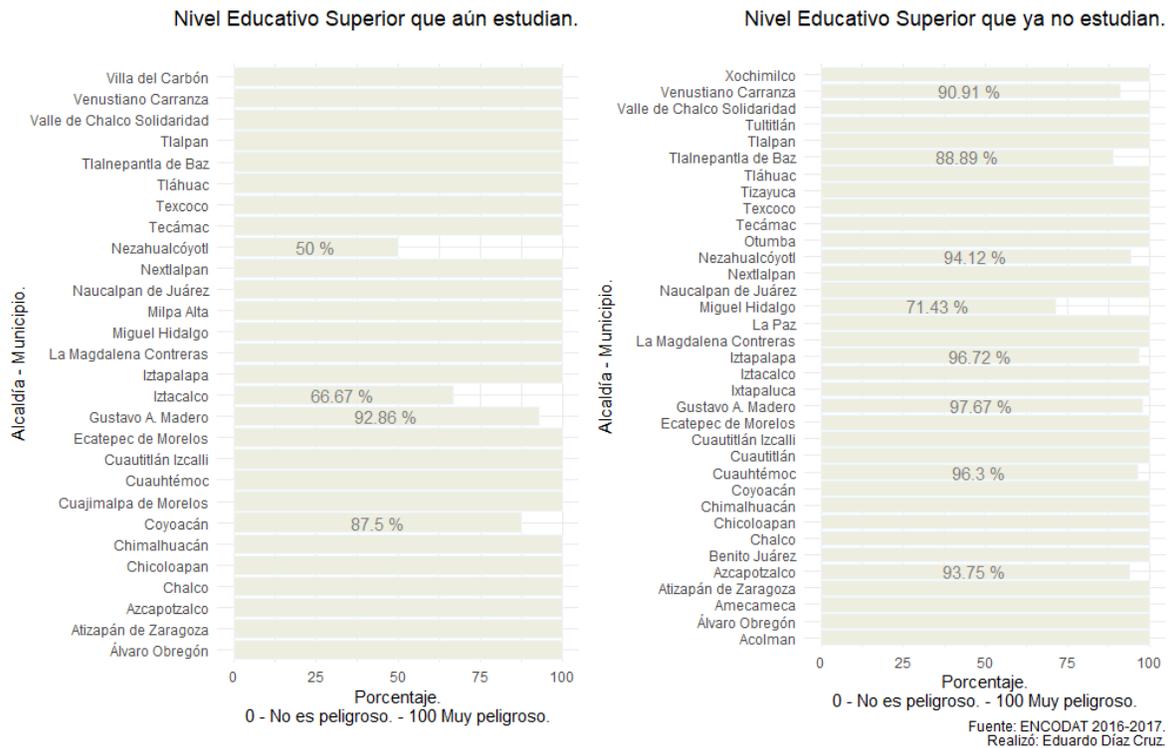
### Percepción de Riesgo en el Consumo de Tabaco, ZMVM.



Gráfica 11: Percepción de Riesgo en el consumo de Tabaco en la ZMVM; Nivel Educativo Medio Superior. Fuente: Elaboración propia.

Es en el grupo de educación media superior donde se encontró el mayor número de observaciones, 561 que aún estudian y 1695 que han concluido y tienen el bachillerato como su último grado educativo. Debido a que la media de edad del grupo que continúa estudiando se encuentra en los 17 años, es importante reiterar que mucha de la literatura sobre el inicio de consumo de sustancias indica que el primer acercamiento ocurre durante la adolescencia y es aquí donde factores como la curiosidad y la necesidad de ser aceptados por el grupo tienen un mayor peso (Molina Prado, Consumo de tabaco, alcohol y drogas en la adolescencia., 2013). Por otra parte, aunque la media de edad del grupo que ya no estudia se encuentra en los 40 años los niveles de apreciación de riesgo, en su mayoría, se encuentran a no más de 5% de diferencia entre grupos.

Percepción de Riesgo en el Consumo de Tabaco, ZMVM.



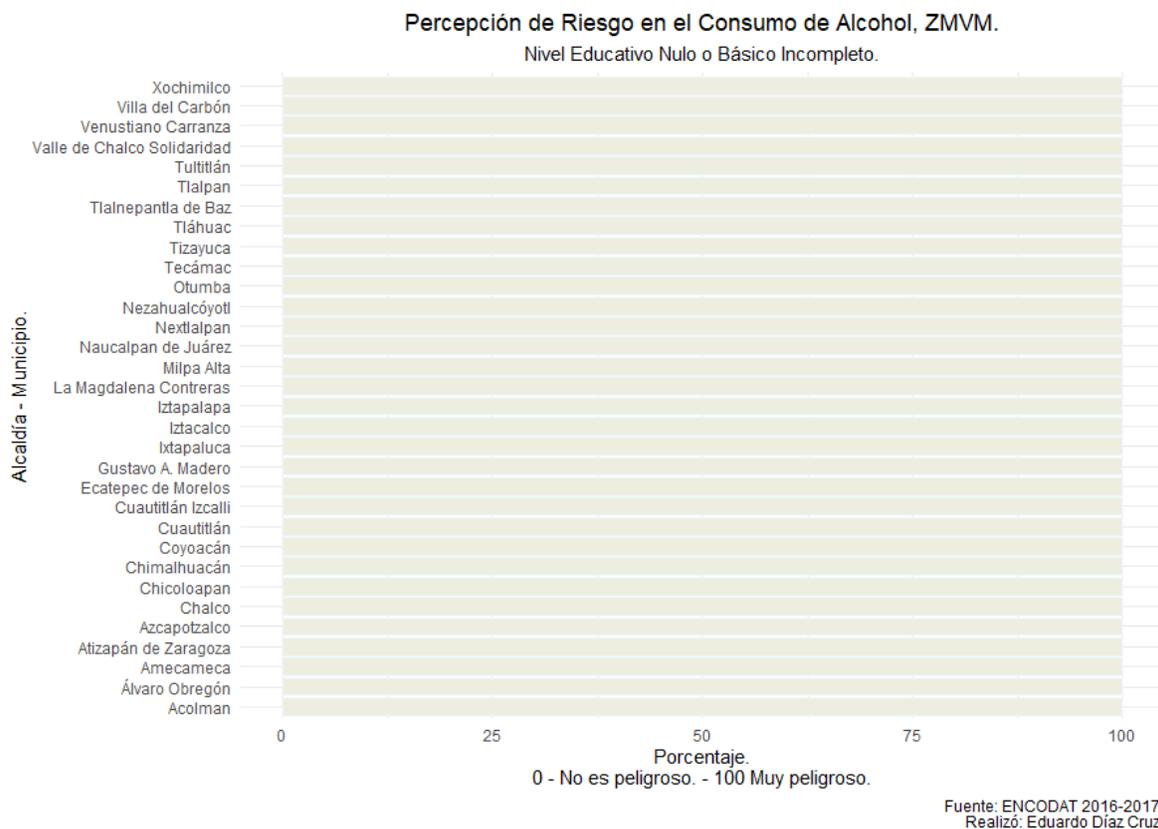
Gráfica 12: Percepción de Riesgo en el consumo de Tabaco en la ZMVM; Nivel Educativo Superior.  
Fuente: Elaboración propia.

En el último segmento de estudio, con media de edad 28 años en los entrevistados que aún estudian y 41 años en los que han detenido sus estudios se puede notar una leve variación en los porcentajes de quienes ya no estudian mientras que entre los que continúan estudiando la variación de opiniones es más evidente, resultado que se puede explicar tomando en cuenta que el estrés está asociado a la alta exigencia académica, factor que se considera una causa de inicio de consumo, seguido de la imitación de profesores y la moda o del uso de este como artículo publicitario o de convivencia (Cheesman Mazariegos & Suárez Lugo, 2015) asociado a la ya mencionada baja apreciación como sustancia adictiva.

Entre los grupos de educación superior se encontraron 91 observaciones para el segmento que aún estudia y se contabilizaron 407 observaciones en el grupo que ha concluido sus estudios, detalle que permite comprender las diferencias porcentuales en los niveles de percepción de riesgo entre ambos conjuntos pues a menor número de observaciones el valor de la media tiende a ceder ante los indicadores de bajo valor.

### 6.2.3 Variable ts24 – Alcohol.

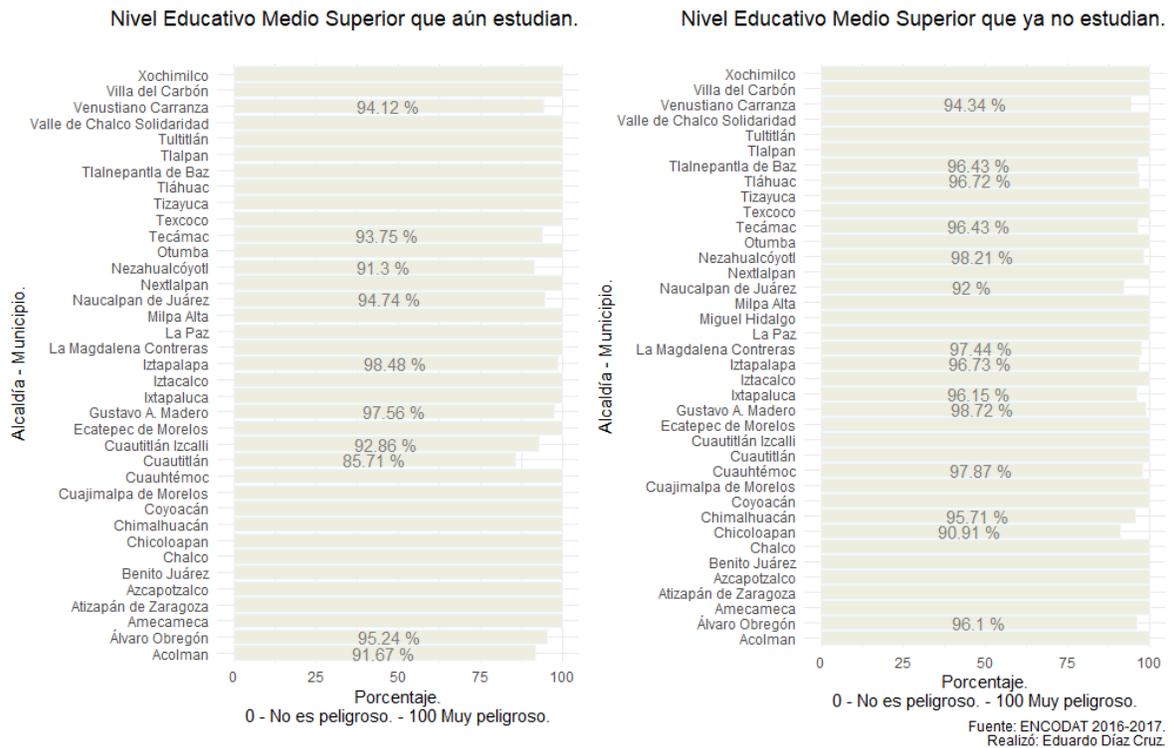
Entre adolescentes, el alcohol es la sustancia más consumida y a esto contribuyen diversos factores sociales como la alta tolerancia y/o aceptación social respecto al consumo, relacionado también al fácil acceso a las diversas presentaciones o métodos de venta (Durá, 2002) además del efecto y el conocimiento de la noción de desinhibición que provoca su consumo durante la convivencia social, aunado a esto, el alcohol se encuentra asociado a las conductas de policonsumo entre tabaco, marihuana y cocaína (Molina Prado, Consumo de tabaco, alcohol y drogas en la adolescencia., 2013).



Gráfica 13: Percepción de Riesgo en el consumo de Alcohol en la ZMVM; Nivel Educativo Nulo o Básico. Fuente: Elaboración propia.

En este grafico se puede apreciar la completa valoración del consumo del alcohol como una conducta dañina para la salud en el grupo de educación básica incompleta o nula para la totalidad de entrevistados en la Zona Metropolitana del Valle de México; este segmento se cuenta con 130 observaciones y la media de edad se localiza en 18.95 años.

Percepción de Riesgo en el Consumo de Alcohol, ZMVM.

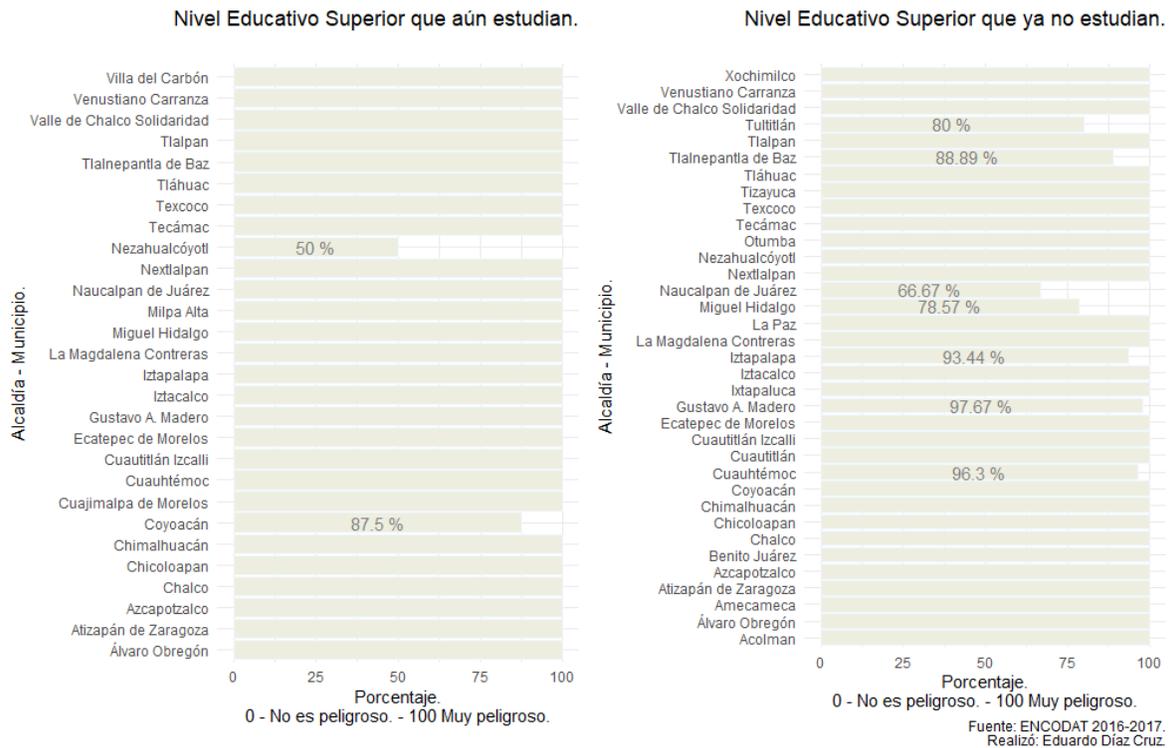


Gráfica 14: Percepción de Riesgo en el consumo de Alcohol en la ZMVM; Nivel Educativo Medio Superior. Fuente: Elaboración propia.

Como se ha mencionado con anterioridad, el impacto de las campañas de prevención durante la etapa educativa tiene un peso mayor en comparación a cuando se ha concluido esta etapa, es este hecho lo que permite rescatar que la percepción de daño a la salud es una idea común en gran parte de los municipios que conforman la ZMVM, como se puede apreciar en la gráfica teniendo menos localidades diferentes al cien por ciento de la apreciación.

En cuanto a la noción general de riesgo entre los grupos la diferencia es de tan solo el .27%, 98.25% para el grupo que aún estudia y 98.52% para el segmento que ha concluido su educación.

Percepción de Riesgo en el Consumo de Alcohol, ZMVM.



Gráfica 15: Percepción de Riesgo en el consumo de Alcohol en la ZMVM; Nivel Educativo Superior.  
Fuente: Elaboración propia.

En el último intervalo de investigación se tiene a los entrevistados de educación superior, que estudian o que han dejado de estudiar, en este segmento la percepción de riesgo general es de 97.76 % y de 97.18 % respectivamente, teniendo una variación acerca de la apreciación de peligro ligeramente menor en el grupo que ha concluido sus estudios, cuya media de edad está en los 41.58 años.

## Capítulo 7. Conclusiones

En este trabajo, con la implementación del lenguaje R para manipular los datos referentes al consumo de sustancias y la información geográfica vectorial existente se logró por medio de la extracción y manipulación intercambiar entre ellos resúmenes estadísticos y características referentes al uso de las sustancias especificadas dándole a los informes y a la cartografía el matiz adecuado según las variables sociodemográficas seleccionadas para cada caso.

A nivel nacional se redujo exitosamente la base de datos a una colección de cifras acerca del uso de sustancias obteniendo para cada entidad federativa un resultado único con el cual es posible, a través de la cartografía generada, comparar el cambio en la cantidad de casos de uso en cada estado y el comportamiento del consumo de las diferentes sustancias a nivel nacional al mismo tiempo de integrar a la cartografía la media del inicio de consumo para cada sustancia, información que permite localizar áreas con un comportamiento similar y otras con un comportamiento diferente pero ligado a su posición geográfica y los territorios colindantes.

En cuanto a la representación de valores de consumo, percepción y prevención a nivel local se encontraron dificultades relacionadas al bajo número de observaciones por localidad debido a la escala de la misma encuesta.

Conforme a lo determinado por la ENCODAT 2016 – 2017 se localizaron y coincidieron los espacios con mayor relevancia en la prevención del consumo entre la población encuestada. Además, utilizando elementos sociodemográficos, como el nivel de estudios y la continuación de la actividad académica, se desarrolló estadística para la principal zona metropolitana del país donde se encontraron casos particulares en la tendencia de la percepción de riesgo que no siempre se mantuvieron de acuerdo a las expectativas en los resultados, pero que sí conservó un estándar reflejado en el comportamiento de la apreciación de peligro entre ambos grupos, los que continúan estudiando y los que han concluido o detenido su periodo formativo, para las tres sustancias examinadas, marihuana, tabaco y alcohol.

En lo relativo al uso de un lenguaje de bajo nivel para administrar tanto información geográfica como la base de datos de la ENCODAT 2016 – 2017 se encontraron facilidades y aplicaciones con una mayor versatilidad de análisis, además de lograr un estudio enfocado únicamente en las variables requeridas con la libertad de eliminar o reducir los datos a la colección deseada para evitar inconvenientes en cuanto a almacenamiento o velocidad de procesamiento y carga, conflictos comunes en el uso de software dedicado y grandes cantidades de datos. Igualmente se demostró la no dependencia en el uso de una interfaz

de usuario y/o programas interactivos de terceros para lograr presentaciones más detalladas o con un nivel de comunicación más intuitivo.

Con la implementación de un sistema de información geográfica basado en un lenguaje de bajo nivel como herramienta de la Geomática se manifestó la mutabilidad de la disciplina para otorgar la referencia y un punto de vista enfocado en la perspectiva geográfica tanto a los resultados como su relación con las áreas adyacentes entregando productos que expongan y amplíen el panorama de las problemáticas.

## Anexos

### Código – Casos de consumo de sustancias a nivel nacional.

```
# Casos de consumo nivel nacional vs edad de inicio de consumo

library(dplyr)
library(tidyr)
library(sf)
library(ggplot2)
library(ggspatial)
library(gridExtra)
library(GLDEX)
sf::sf_use_s2(FALSE)

#Todos los enteros (int) con solo 4 decimales
options(digits=4)

#Se lee el conjunto de datos
E1617<- read.csv("C:/Users/Eduardo/Documents/Tesis/Base de
datos/DB/2016/encodat_bases_individual-
formato_libre/ENCODAT_2016_2017.csv")
#Se edita la columna IDPERS separando clave municipal y local
E1617 <- E1617 %>% separate(ID_PERS,c("CVEGEO","IDPers"),sep = 9)
#Se crea una nueva columna (CVEGEO) con los mismo datos de CVEGEO
E1617$CVEGEO <- E1617$CVEGEO
#La nueva columna se separa a razon de los digitos para las claves
municipales y locales
E1617 <- E1617 %>% separate(CVEGEO,c("CVEGEO","Loc"), sep = 5)
#Se elimina la clave de localidad, no es neceswria en este punto
E1617$Loc <- NULL

#Vector con las variables de interés, edad de consumo por sustancia
vars <- c("di4a", "di4b", "di4c", "di4d", "di4e", "di4f", "di4g")

#Selección de todos los entrevistados de sexo masculino por entidad
Drogas_Edad_H <- E1617%>% filter(ds2 == 1) %>% group_by(entidad) %>%
summarise(Entrevistados=n())
#Calculo de la media de edad al inicio del consumo Hombres (Media a nivel
estatal)
for(x in vars){
  var <- select(E1617, entidad, x) #todas las observaciones de la
variable de interés (di4a marihuana)
  na_obs <- which(!complete.cases(var[,x]))
  var <- var[-na_obs,]
  na_obs2 <- which(var[,x]>65) #valores >65 no son utiles
  var <- var[-na_obs2,]
  var1 <- aggregate(x = var[,x], # Columna específica de datos
                    by = list(var$entidad), # Indicador del grupo
                    FUN = mean) #calcula la media de solo los
consumidores
  names(var1) <- c("entidad", x) #se edita el nombre de la columna
generada
```

Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la  
ENCODAT 2016 - 2017.

```
Drogas_Edad_H <- Drogas_Edad_H %>% left_join(var1) #Se une la columna
generada a la tabla inicial
}

#Selección de todos los entrevistados de sexo femenino por entidad
Drogas_Edad_M <- E1617%>% filter(ds2 == 2) %>% group_by(entidad) %>%
summarise(Entrevistadas=n())
#Calculo de la media de edad al inicio del consumo Mujeres (Media a nivel
estatal)
for(x in vars){
  var <- select(E1617, entidad, x) #todas las observaciones pero solo la
variable di4a (marihuana)
  na_obs <- which(!complete.cases(var[,x]))
  var <- var[-na_obs,]
  na_obs2 <- which(var[,x]>65)
  var <- var[-na_obs2,]
  var1 <- aggregate(x = var[,x],by = list(var$entidad),FUN = mean)
  names(var1) <- c("entidad", x)
  Drogas_Edad_M <- Drogas_Edad_M %>% left_join(var1)
}

#Se carga el archivo a nivel estatal
DPEst <-
st_read("C:/Users/Eduardo/Documents/Tesis/Sig/div_pol_est/dest2019gw.shp"
)
#se renombra el campo en la tabla de edades generadas para unir los
atributos
Drogas_Edad_H <- rename(Drogas_Edad_H,"CVE_ENT" = "entidad")
Drogas_Edad_M <- rename(Drogas_Edad_M,"CVE_ENT" = "entidad")

#se unen las tablas de edad
DPEst$CVE_ENT <- as.numeric(DPEst$CVE_ENT)
DPEst_EH <- DPEst %>% left_join(Drogas_Edad_H)
DPEst_EM <- DPEst %>% left_join(Drogas_Edad_M)

#Análisis de casos de consumo por entidad

#Vector de sustancias
susts <- c("dila","dilb","dilc","dild","dile","dilf","dilg")

#Se crea la base de la tabla de casos de consumo de hombres
Drogas_Hombres <- E1617%>% filter(ds2 ==1) %>% group_by(entidad) %>%
summarise(Entrevistados=n())
#utilizando el vector susts se hace el conteo de casos de consumo
for (x in susts){
  temp <- E1617 %>% filter(!! rlang::sym(x) == 1 & ds2 == 1) %>%
group_by(entidad) %>% summarise(colgen=n())
  names(temp) <- c("entidad", x)
  Drogas_Hombres <- Drogas_Hombres %>% left_join(temp)
}

#Se crea la base de la tabla de casos de consumo de mujeres
Drogas_Mujeres <- E1617%>% filter(ds2 ==2) %>% group_by(entidad) %>%
summarise(Entrevistadas=n())
#utilizando el vector susts se hace el conteo de casos de consumo
for (x in susts){
```

## Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la ENCODAT 2016 - 2017.

```
temp <- E1617 %>% filter( !! rlang::sym(x) == 1 & ds2 == 2 ) %>%
group_by(entidad) %>% summarise(colgen=n())
names(temp) <- c("entidad", x)
Drogas_Mujeres <- Drogas_Mujeres %>% left_join(temp)
}

#Se editan las tablas para poder unir las
Drogas_Hombres <- rename(Drogas_Hombres, "CVE_ENT" = "entidad")
Drogas_Mujeres <- rename(Drogas_Mujeres, "CVE_ENT" = "entidad")

#Centroides
C_DPEst <- st_centroid(DPEst)

#Se une la tabla de casos de consumo a los centroides generados
DPEst_DH <- C_DPEst %>% left_join(Drogas_Hombres)
DPEst_DM <- C_DPEst %>% left_join(Drogas_Mujeres)

#Se cambian los valores NA por 0 para evitar problemas en la escala
DPEst_DH[is.na(DPEst_DH)] <- 0
DPEst_DM[is.na(DPEst_DM)] <- 0

#Plot de mapas

#Escala de colores personalizada para el numero de casos.

#Variable continua (Numero de casos)
labels_var <- pretty(range(DPEst_DH$dilg)) #modificar la variable a una
variable de tipo discreta (niveles)
breaks_var <- c(labels_var, Inf)
DPEst_DH <- DPEst_DH %>%
  mutate(Consumidores = cut(dilg,breaks = breaks_var, labels =
labels_var, right = FALSE)) #modificar la variable
#Paleta de colores y tamaño (Puntos)
colores <- c("#0C26F9", "#0CF9D1") #Colores inicial - final (Oscuro -
Claro)
pal <- colorRampPalette(colores)(length(labels_var))
pal_size <- seq(1, 6, length.out = 7)

#Hombres
ggplot()+
  geom_sf(data = DPEst_EH, aes(fill=dilg), lwd = .5, color =
"peachpuff4", alpha=.8)+
  scale_fill_continuous(high="#FAF029", low="#FA5E28")+ #Paleta de
colores (Fondo)
  geom_sf(data = DPEst_DH %>% filter(Consumidores!=0),
aes(color=Consumidores, size=Consumidores))+
  scale_color_manual(values = pal) +
  scale_size_manual(values = pal_size)+
  geom_sf_label(data = DPEst_DH %>% filter(dilg>=20),
aes(label=NOM_ENT), nudge_x = c(0,0),nudge_y = c(-.8,-.7), size=2.5,
colour="gray45")+
  labs(title = "Casos de Consumo de Anfetaminas en Hombres a Nivel
Nacional.", subtitle = "Entidades con más de 20 casos de
consumo.",fill="Inicio de consumo.\nMedia de edad (años).", x="Longitud",
y="Latitud", caption = "Fuente: ENCODAT 2016 - 2017. \n Realizó: Eduardo
Díaz Cruz.")+#
```

```
  annotation_north_arrow(aes(location="tr"),which_north = "true", height
= unit(1.5, "cm"), width = unit(1.5, "cm"), style =
north_arrow_fancy_orienteeing)+
  theme(plot.title = element_text(hjust = 0.5), plot.subtitle =
element_text(hjust = 0.5),
        panel.background = element_rect(fill = "aliceblue", colour =
"aliceblue",size = 0.5, linetype = "solid"))+
  guides(fill = guide_colorbar(order = 1))

#Imágenes

#p1 <- ggplot()+
# geom_sf(data = DPEst)

#p2 <- ggplot()+
# geom_sf(data = C_DPEst)

#p3 <- ggplot()+
# geom_sf(data = DPEst)+
#geom_sf(data = C_DPEst)

#grid.arrange(p1, p2, p3, ncol=2)
```

## Código – Cartografía 3D

```
# Casos de consumo a nivel nacional, rayshader
library(dplyr)
library(tidyr)
library(sf)
library(ggplot2)
library(ggspatial)
library(gridExtra)
library(rayshader)
library(rayrender)
library(viridis)
#Todos los enteros con solo 4 decimales
options(digits=4)

#Se lee el conjunto de datos
E1617 <- read.csv("C:/Users/Eduardo/Documents/Tesis/Base de
datos/DB/2016/encodat_bases_individual-
formato_libre/ENCODAT_2016_2017.csv")
#Se edita la columna IDPERS separando clave municipal y local
E1617 <- E1617 %>% separate(ID_PERS,c("CVEGEO", "IDPers"), sep = 9)
#Se crea una nueva columna (CVEGEO) con los mismo datos de CVEGEO
E1617$CVEGEO <- E1617$CVEGEO
#La nueva columna se separa a razon de los digitos para las claves
municipales y locales
E1617 <- E1617 %>% separate(CVEGEO,c("CVEGEO", "Loc"), sep = 5)
#Se elimina la clave de localidad, no es necesaria
E1617$Loc <- NULL

#Análisis de casos de consumo por entidad
#Vector de sustancias
susts <- c("dila", "dilb", "dilc", "dild", "dile", "dile", "dile", "dile")

#Se crea la base de la tabla de casos de consumo de hombres
Drogas_Hombres <- E1617%>% filter(ds2 ==1) %>% group_by(entidad) %>%
summarise(Entrevistados=n())
#utilizando el vector susts se hace el conteo de casos de consumo
for (x in susts){
  temp <- E1617 %>% filter(!! rlang::sym(x) == 1 & ds2 == 1) %>%
group_by(entidad) %>% summarise(colgen=n())
  names(temp) <- c("entidad", x)
  Drogas_Hombres <- Drogas_Hombres %>% left_join(temp)
}

#Se crea la base de la tabla de casos de consumo de mujeres
Drogas_Mujeres <- E1617%>% filter(ds2 ==2) %>% group_by(entidad) %>%
summarise(Entrevistadas=n())
#utilizando el vector susts se hace el conteo de casos de consumo
for (x in susts){
  temp <- E1617 %>% filter(!! rlang::sym(x) == 1 & ds2 == 2) %>%
group_by(entidad) %>% summarise(colgen=n())
  names(temp) <- c("entidad", x)
  Drogas_Mujeres <- Drogas_Mujeres %>% left_join(temp)
}
```

Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la  
ENCODAT 2016 - 2017.

```
#Se editan las tablas para poder unir las
Drogas_Hombres <- rename(Drogas_Hombres, "CVE_ENT" = "entidad")
Drogas_Mujeres <- rename(Drogas_Mujeres, "CVE_ENT" = "entidad")

# Se carga el archivo a nivel estatal
DPEst <-
st_read("C:/Users/Eduardo/Documents/Tesis/Sig/div_pol_est/dest2019gw.shp"
)

# Se unen las tablas al .shp de poligonos
DPEst$CVE_ENT <- as.numeric(DPEst$CVE_ENT)
DPEst_DH <- DPEst %>% left_join(Drogas_Hombres)
DPEst_DM <- DPEst %>% left_join(Drogas_Mujeres)

#Se cambian los valores NA por 0 para evitar problemas en la escala
DPEst_DH[is.na(DPEst_DH)] <- 0
DPEst_DM[is.na(DPEst_DM)] <- 0

#Ploteo

# Cartografía 2D
ggshp <- ggplot()+
  geom_sf(data = DPEst_DM,aes(fill = di1d), lwd = 0)+
  scale_fill_continuous(high="#FA5E28", low="#FAF029", breaks =
c(2,4,6))+ #Paleta de colores (Fondo)
  labs(title = "Casos de Consumo de Alucinógenos en Mujeres a Nivel
Nacional.", fill = "Casos de consumo.", caption= "Fuente: ENCODAT 2016 -
2017 \n Elaboró: Eduardo Díaz Cruz")+
  theme_minimal()+
  theme(plot.title = element_text(hjust = 0.5),
        panel.background = element_rect(fill = "aliceblue",colour =
"aliceblue",size = 0.5, linetype = "solid"))
ggshp
# Crear y guardar elemento 3D
#plot_gg(ggshp, width = 8, height = 8, window size = c(1366,768), zoom
= .6, offset_edges = 1.5, sunangle = 0)
#render_snapshot("C:/Users/Eduardo/Documents/Tesis/Imagenes/Cap5/Nacional
_v2_m/m_d", clear = T)

# Parametros render_movie
phivechalf <- 30 + 60 * 1/(1 + exp(seq(-7, 20, length.out = 180)/2))
phivecfull <- c(phivechalf, rev(phivechalf))
thetavec <- -90 + 45 * sin(seq(0, 359, length.out = 360) * pi/180)
zoomvec <- 0.45 + 0.2 * 1/(1 + exp(seq(-5, 20, length.out = 180)))
zoomvecfull <- c(zoomvec, rev(zoomvec))

#render_movie(filename =
"C:/Users/Eduardo/Documents/Tesis/Imagenes/Cap5/Nacional_v2_m/videos/m_f"
, type = "custom", frames = 360, fps = 30, phi = phivecfull, zoom =
zoomvecfull, theta = thetavec)
```

## Código – Niveles de percepción.

```
# Percepción por niveles educativos

library(dplyr)
library(tidyr)
library(ggplot2)
library(ggspatial)
library(viridis)
library(sf)
library(gridExtra)
library(grid)

#Carga de la base de datos ENCODAT 2016 - 2017
ENCODAT1617 <- read.csv("C:/Users/Eduardo/Documents/Tesis/Base de
datos/DB/2016/encodat_bases_individual-
formato_libre/ENCODAT_2016_2017.csv")
E1617 <- ENCODAT1617
E1617 <- E1617 %>% separate(ID_PERS, c("CVEGEO", "IDPers"), sep = 9)
E1617$CVEGEO <- E1617$CVEGEO
E1617 <- E1617 %>% separate(CVEGEO, c("CVEGEO", "Loc"), sep = 5)
E1617$Loc <- NULL
options(digits=4)

#Carga del archivo shp de las zonas metropolitanas
zm <- st_read("C:/Users/Eduardo/Documents/Tesis/Sig/zonas_m/ZM_2010.shp")
#se abre el .shp de zm
zmvm <- zm %>% filter(NOM_SUN=="Valle de Mexico") #Solo la zmvm

# Tomar las claves de los municipios de la zmvm
clv_zmvm <- select(zmvm, CVE_MUN1, NOM_MUN)
clv_zmvm$geometry <- NULL
clv_zmvm <- rename(clv_zmvm, "CVEGEO" = "CVE_MUN1")

E1617_zmvm <- clv_zmvm %>% left_join(E1617) #Todas las observaciones de
la E1617 para la ZMVM
na_obs <- which(!complete.cases(E1617_zmvm$ds9 == 99))
E1617_zmvm <- E1617_zmvm[-na_obs,]

rm(E1617, ENCODAT1617, zm, zmvm)

# Hombres/mujeres adolescentes/adultos
#H_ado_vm <- E1617_zmvm %>% filter(ds2==1 & ds3<18)
#H_adu_vm <- E1617_zmvm %>% filter(ds2==1 & ds3>=18)
#M_ado_vm <- E1617_zmvm %>% filter(ds2==2 & ds3<18)
#M_adu_vm <- E1617_zmvm %>% filter(ds2==2 & ds3>=18)

# Generacion de los df por grado educativo de la zmvm

#nunca ha ido a la escuela ds8 o primaria incompleta ds9
educacion_nula <- E1617_zmvm %>% filter(ds8==1 | ds9==1)

#ya no estudian_yne
#educación media superior
educacion_ms_yne <- E1617_zmvm %>% filter(ds9>=2 & ds9<=6 & ds8==2)
#educación superior
```

Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la  
ENCODAT 2016 - 2017.

```
educacion_s_yne <- E1617_zmvm %>% filter(ds9>=7 & ds8==2)

#aún estudian _ae
#educación media superior
educacion_ms_ae <- E1617_zmvm %>% filter(ds9>=2 & ds9<=6 & ds8==3)
#educación superior
educacion_s_ae <- E1617_zmvm %>% filter(ds9>=7 & ds8==3)

#percepcion riesgo
vect_var <- c("ts18", "ts19", "ts20", "ts21", "ts22", "ts23", "ts24")

#           Ciclo percepcion
#   de los df educacion calcula los entrevistados y los porcentajes de
#   percepción de riesgo
percepcion <- clv_zmvm #se crea un df contenedor vacio, solo con las
#   claves geograficas de los municipios de la zmvm
for (x in vect_var){
  v <- educacion_s_ae %>% group_by(CVEGEO) %>%
    filter(!! rlang::sym(x)>=2) %>% #!! rlang:: sym() utiliza a x como un
#   objeto
    summarise(n())
  names(v)=c("CVEGEO", x)
  percepcion <- percepcion %>% left_join(v)
  na_obs2 <- which(complete.cases(percepcion$ts18))
  percepcion <- percepcion[na_obs2,]
}
entrevistados <- educacion_s_ae %>% group_by(CVEGEO) %>%
summarise(Entrevistados=n())
percepcion <- percepcion %>% left_join(entrevistados)

#           Ciclo porcentajes

for (x in vect_var){
  name=paste(x, "_p", sep = "")
  percepcion <- percepcion %>%
mutate(nom_prov=(percepcion[,x]/percepcion$Entrevistados)*100)
  percepcion <- rename(percepcion, !!name := "nom_prov") #!!name: utiliza
#   la variable como el contenedor de un objeto
}

#           un df por cada grado de estudios, incluyen entrevistados y
#   porcentajes
#percp_educacion_nula <- percepcion
#percp_educacion_ms_ae <- percepcion
#percp_educacion_ms_yne <- percepcion
#percp_educacion_s_ae <- percepcion
#percp_educacion_s_yne <- percepcion

#se eliminan las variables que no se utilizan
rm(clv_zmvm,educacion_ms_ae,educacion_ms_yne,educacion_nula,educacion_s_a
e,educacion_s_yne,entrevistados,v,na_obs,na_obs2,name,vect_var,x)

#           Graficos

graf <- ggplot(percp_educacion_nula, aes(x=NOM_MUN, y=ts23_p))+
  geom_bar(color = "aliceblue", fill = "ivory2",stat = "identity")+
  labs(title = "Percepción de Riesgo en el Consumo de Tabaco, ZMVM.",
```

Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la  
ENCODAT 2016 - 2017.

```
  subtitle = "Nivel Educativo Nulo o Básico Incompleto.",
  x= "Alcaldía - Municipio.", y= "Porcentaje.\n0 - No es peligroso.
- 100 Muy peligroso.",
  caption = "Fuente: ENCODAT 2016-2017.\nRealizó: Eduardo Díaz
Cruz.")+
  geom_text(data = subset(percp_educacion_nula,ts23_p!=100),aes(label =
paste(round((ts23/Entrevistados)*100,2),"%",sep = " ")),
  position = position_stack(vjust = .5), color = "gray50")+
  theme_minimal()+
  theme(plot.title = element_text(hjust = .5), plot.subtitle =
element_text(hjust = .5))+
  coord_flip()

# graf2 solo cuando se trabaja educacion media superior y superior
graf2 <- ggplot(percp_educacion_s_yne, aes(x=NOM_MUN, y=ts24_p))+
  geom_bar(color = "aliceblue", fill = "ivory2",stat = "identity")+
  labs(title = "Nivel Educativo Superior que ya no estudian.",
  subtitle = "",
  x= "Alcaldía - Municipio.", y= "Porcentaje.\n0 - No es peligroso.
- 100 Muy peligroso.",
  caption = "Fuente: ENCODAT 2016-2017.\nRealizó: Eduardo Díaz
Cruz.")+
  geom_text(data = subset(percp_educacion_s_yne,ts24_p!=100),aes(label =
paste(round((ts24/Entrevistados)*100,2),"%",sep = " ")),
  position = position_stack(vjust = .5), color = "gray50")+
  theme_minimal()+
  theme(plot.title = element_text(hjust = 1), plot.subtitle =
element_text(hjust = 1))+
  coord_flip()

grid.arrange(graf,graf2, ncol=2, nrow=1, top =
textGrob("\n",gp=gpar(fontsize=20,font=1)))

# Calcular edades y medias de porcentajes
#media edad y percepcion edu nula
media_nula <- E1617_zmvm %>% filter(ds9>=2 & ds8==3) %>%
summarise(mean(ds3))
media_nula$Marihuana <- percp_educacion_nula %>% summarise(mean(ts18_p))
media_nula$Tabaco <- percp_educacion_nula %>% summarise(mean(ts23_p))
media_nula$Alcohol <- percp_educacion_nula %>% summarise(mean(ts24_p))
media_nula$Total <- percp_educacion_nula %>%
summarise(sum(Entrevistados))

#media edad y percepcion ms_ae
media_ms_ae <- E1617_zmvm %>% filter(ds9>=2 & ds9<=6 & ds8==3) %>%
summarise(mean(ds3))
media_ms_ae$Marihuana <- percp_educacion_ms_ae %>%
summarise(mean(ts18_p))
media_ms_ae$Tabaco <- percp_educacion_ms_ae %>% summarise(mean(ts23_p))
media_ms_ae$Alcohol <- percp_educacion_ms_ae %>% summarise(mean(ts24_p))
media_ms_ae$Total <- percp_educacion_ms_ae %>%
summarise(sum(Entrevistados))
#media edad y percepcion ms_yne
media_ms_yne <- E1617_zmvm %>% filter(ds9>=2 & ds9<=6 & ds8==2) %>%
summarise(mean(ds3))
```

Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la  
ENCODAT 2016 - 2017.

```
media_ms_yne$Marihuana <- percp_educacion_ms_yne %>%  
summarise(mean(ts18_p))  
media_ms_yne$Tabaco <- percp_educacion_ms_yne %>%  
summarise(mean(ts23_p))  
media_ms_yne$Alcohol <- percp_educacion_ms_yne %>%  
summarise(mean(ts24_p))  
media_ms_yne$Total <- percp_educacion_ms_yne %>%  
summarise(sum(Entrevistados))  
#media edad y percepcion s_ae  
media_s_ae <- E1617_zmvm %>% filter(ds9>=7 & ds8==3) %>%  
summarise(mean(ds3))  
media_s_ae$Marihuana <- percp_educacion_s_ae %>% summarise(mean(ts18_p))  
media_s_ae$Tabaco <- percp_educacion_s_ae %>% summarise(mean(ts23_p))  
media_s_ae$Alcohol <- percp_educacion_s_ae %>% summarise(mean(ts24_p))  
media_s_ae$Total <- percp_educacion_s_ae %>%  
summarise(sum(Entrevistados))  
#media edad y percepcion s_yne  
media_s_yne <- E1617_zmvm %>% filter(ds9>=7 & ds8==2) %>%  
summarise(mean(ds3))  
media_s_yne$Marihuana <- percp_educacion_s_yne %>%  
summarise(mean(ts18_p))  
media_s_yne$Tabaco <- percp_educacion_s_yne %>% summarise(mean(ts23_p))  
media_s_yne$Alcohol <- percp_educacion_s_yne %>% summarise(mean(ts24_p))  
media_s_yne$Total <- percp_educacion_s_yne %>%  
summarise(sum(Entrevistados))
```

## Código – Zonas metropolitanas.

```
# Extracción de las zonas metropolitanas del país según CONAPO

library(dplyr)
library(tidyr)
library(sf)
library(ggplot2)
library(ggspatial)
library(gridExtra)

zm <- st_read("C:/Users/Eduardo/Documents/Tesis/Sig/zonas_m/ZM_2010.shp")

zmvm <- zm %>% filter(NOM_SUN == "Valle de México")
zmvm$NOM_ENT[zmvm$NOM_ENT=="México"] <- "Mexico"
zmvm$NOM_ENT[zmvm$NOM_ENT=="Distrito Federal"] <- "CDMX"

zmvm_dg <- st_transform(zmvm, 4326)

ggplot()+
  geom_sf(data = zmvm_dg, aes(fill = NOM_ENT))+
  labs(title = "Zona Metropolitana del Valle de México.",
       fill="Entidad.", x="Longitud", y="Latitud",
       caption = "Fuente: Delimitación de zonas metropolitanas, CONAPO.
\n Realizó: Eduardo Díaz Cruz.")+
  annotation_north_arrow(aes(location="tr"),which_north = "true", height
= unit(1, "cm"), width = unit(1, "cm"),
                        style = north_arrow_nautical)+
  theme_minimal()+
  theme(plot.title = element_text(hjust = 0.5), plot.subtitle =
element_text(hjust = 0.5))

## Zona Metropolitana Monterrey

zmm <- zm %>% filter(CVE_ENT == "19")
zmm$NOM_ENT[zmm$NOM_ENT=="Nuevo León"] <- "Nuevo Leon"

zmm_dg <- st_transform(zmm,4326)

ggplot()+
  geom_sf(data = zmm_dg, aes(fill = NOM_ENT))+
  labs(title = "Zona Metropolitana de Monterrey.",
       fill="Entidad.", x="Longitud", y="Latitud",
       caption = "Fuente: Delimitación de zonas metropolitanas, CONAPO.
\n Realizó: Eduardo Díaz Cruz.")+
  annotation_north_arrow(aes(location="tr"),which_north = "true", height
= unit(1, "cm"), width = unit(1, "cm"),
                        style = north_arrow_nautical)+
  theme_minimal()+
  theme(plot.title = element_text(hjust = 0.5), plot.subtitle =
element_text(hjust = 0.5))

## Zona Metropolitana de Guadalajara

zmg <- zm %>% filter(CVE_ENT == "14")
```

Sistemas de Información Geográfica en Lenguaje R aplicados a la base de datos de la  
ENCODAT 2016 - 2017.

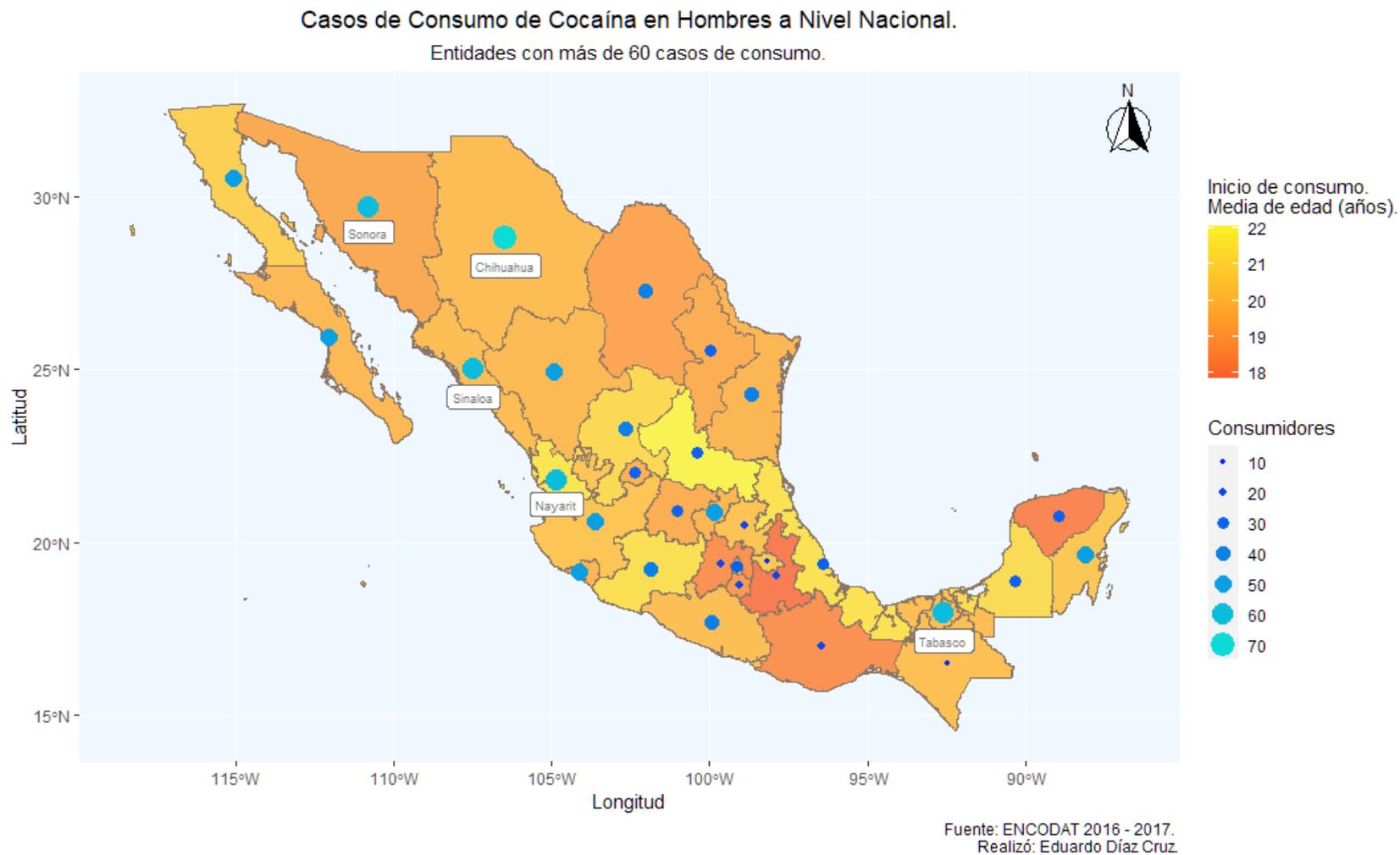
```
zmg <- zmg[!(zmg$NOM_MUN == "Puerto Vallarta"),] #se elimina el municipio  
que no corresponde  
zmg_dg <- st_transform(zmg, 4326)
```

```
ggplot()+  
  geom_sf(data = zmv_dg, aes(fill = NOM_ENT))+  
  labs(title = "Zona Metropolitana de Guadalajara.",  
        fill="Entidad.", x="Longitud", y="Latitud",  
        caption = "Fuente: Delimitación de zonas metropolitanas, CONAPO.  
\n Realizó: Eduardo Díaz Cruz.")+  
  annotation_north_arrow(aes(location="tr"),which_north = "true", height  
= unit(1, "cm"), width = unit(1, "cm"),  
                        style = north_arrow_nautical)+  
  theme_minimal()+  
  theme(plot.title = element_text(hjust = 0.5), plot.subtitle =  
element_text(hjust = 0.5))  
  
#st_write(zmv_dg, "C:/Users/Eduardo/Documents/Tesis/Sig/Shapes_g",  
"ZM_ValleMexico", driver="ESRI Shapefile")  
#st_write(zmm_dg, "C:/Users/Eduardo/Documents/Tesis/Sig/Shapes_g",  
"ZM_Monterrey", driver="ESRI Shapefile")  
#st_write(zmg_dg, "C:/Users/Eduardo/Documents/Tesis/Sig/Shapes_g",  
"ZM_Guadalajara", driver="ESRI Shapefile")  
  
rm(zm, zmg, zmm, zmv)
```

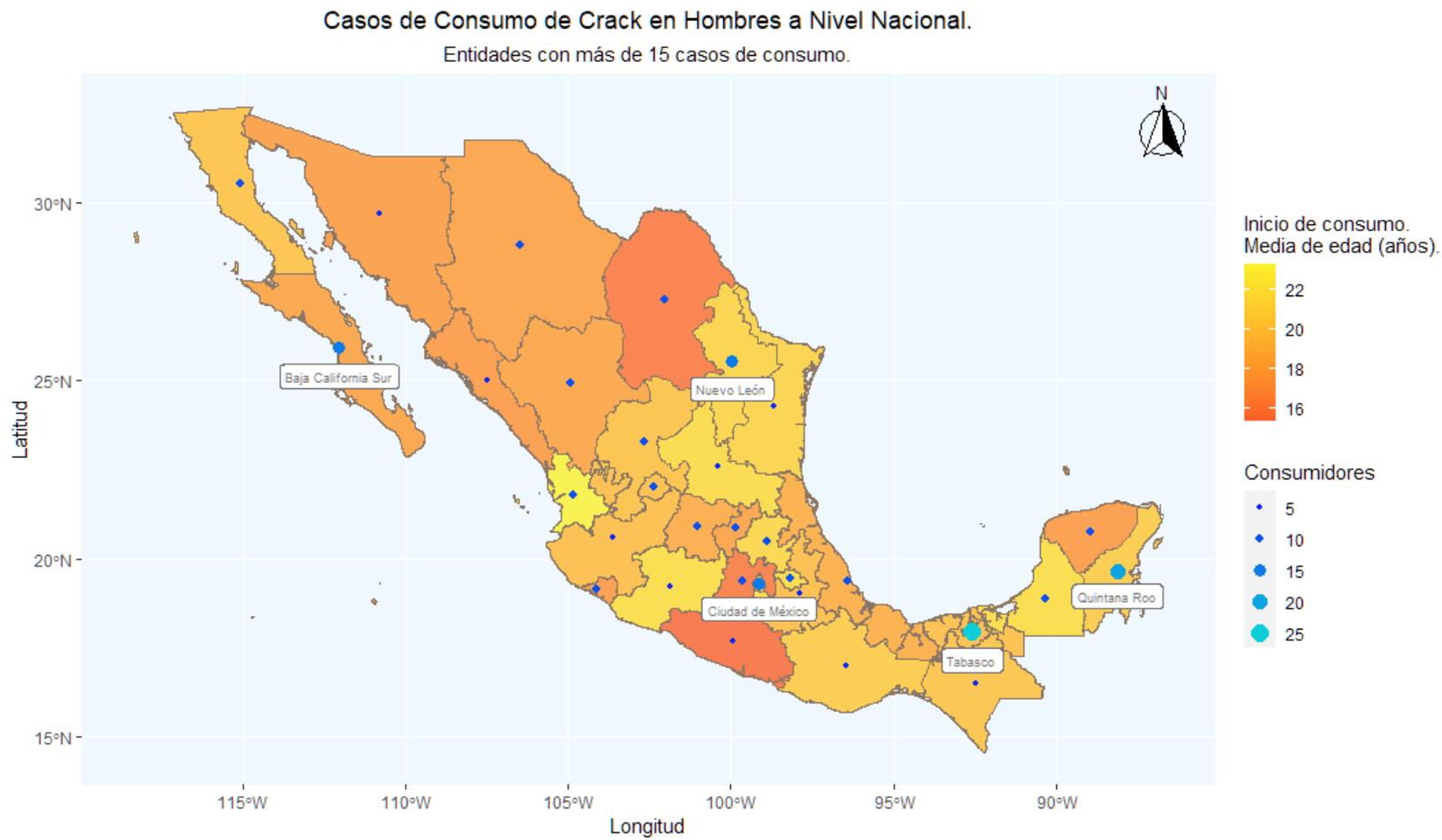
Mapa 1, Casos de Consumo de Marihuana en Hombres a Nivel Nacional



Mapa 2, Casos de Consumo de Cocaína en Hombres a Nivel Nacional

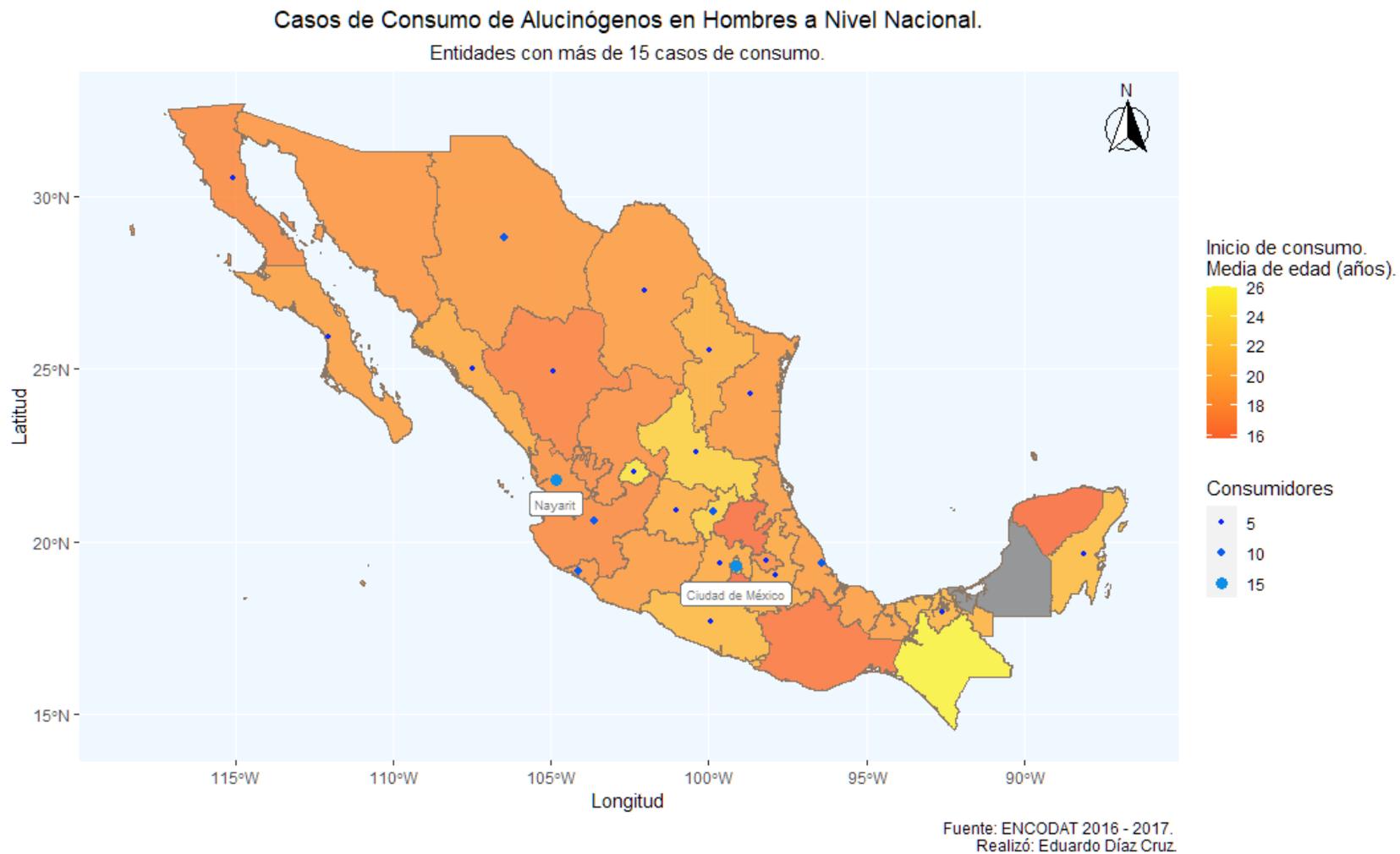


Mapa 3, Casos de Consumo de Crack en Hombres a Nivel Nacional



Fuente: ENCODAT 2016 - 2017.  
Realizó: Eduardo Díaz Cruz.

Mapa 4, Casos de Consumo de Alucinógenos en Hombres a Nivel Nacional



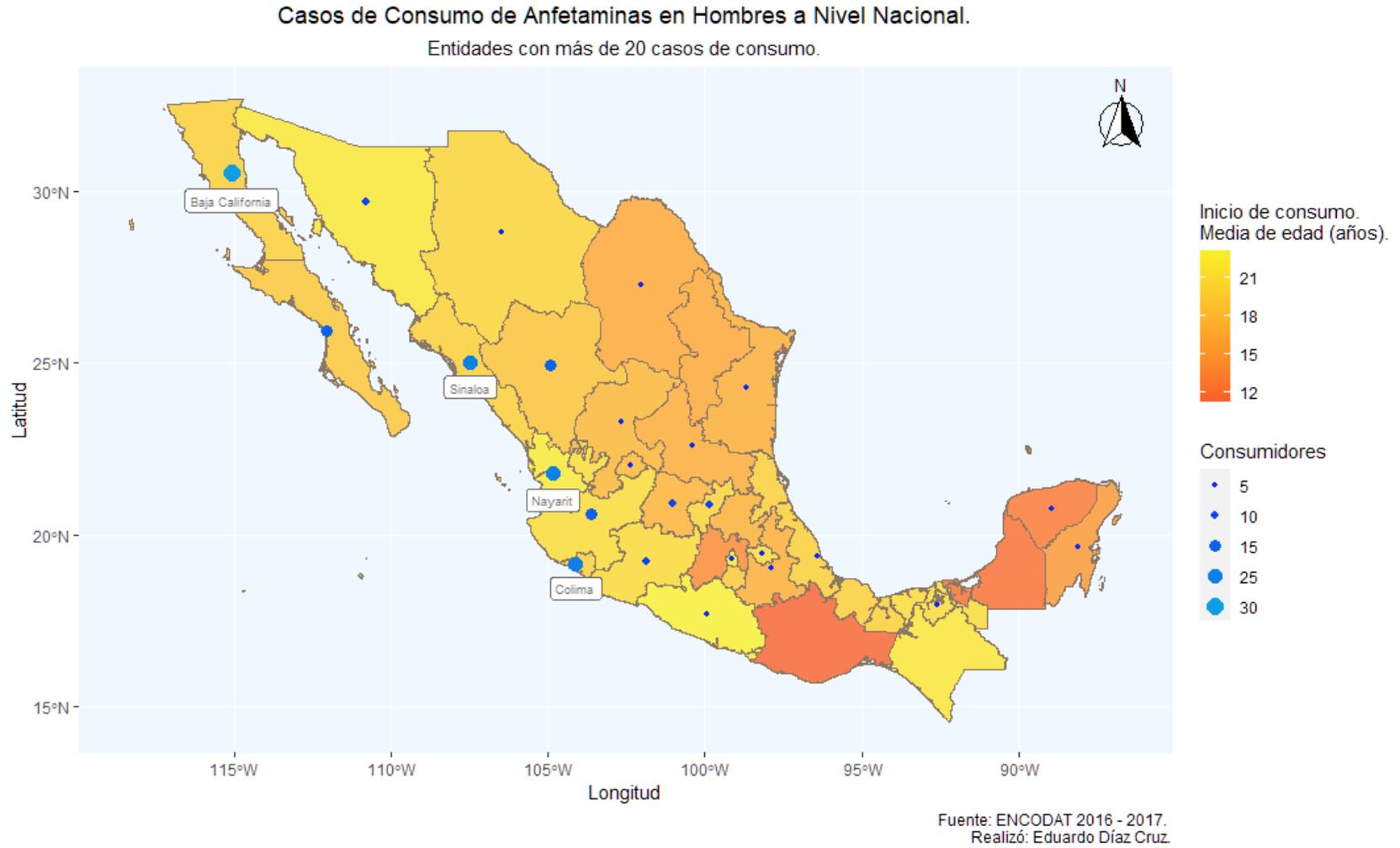
Mapa 5, Casos de Consumo de Inhalables en Hombres a Nivel Nacional



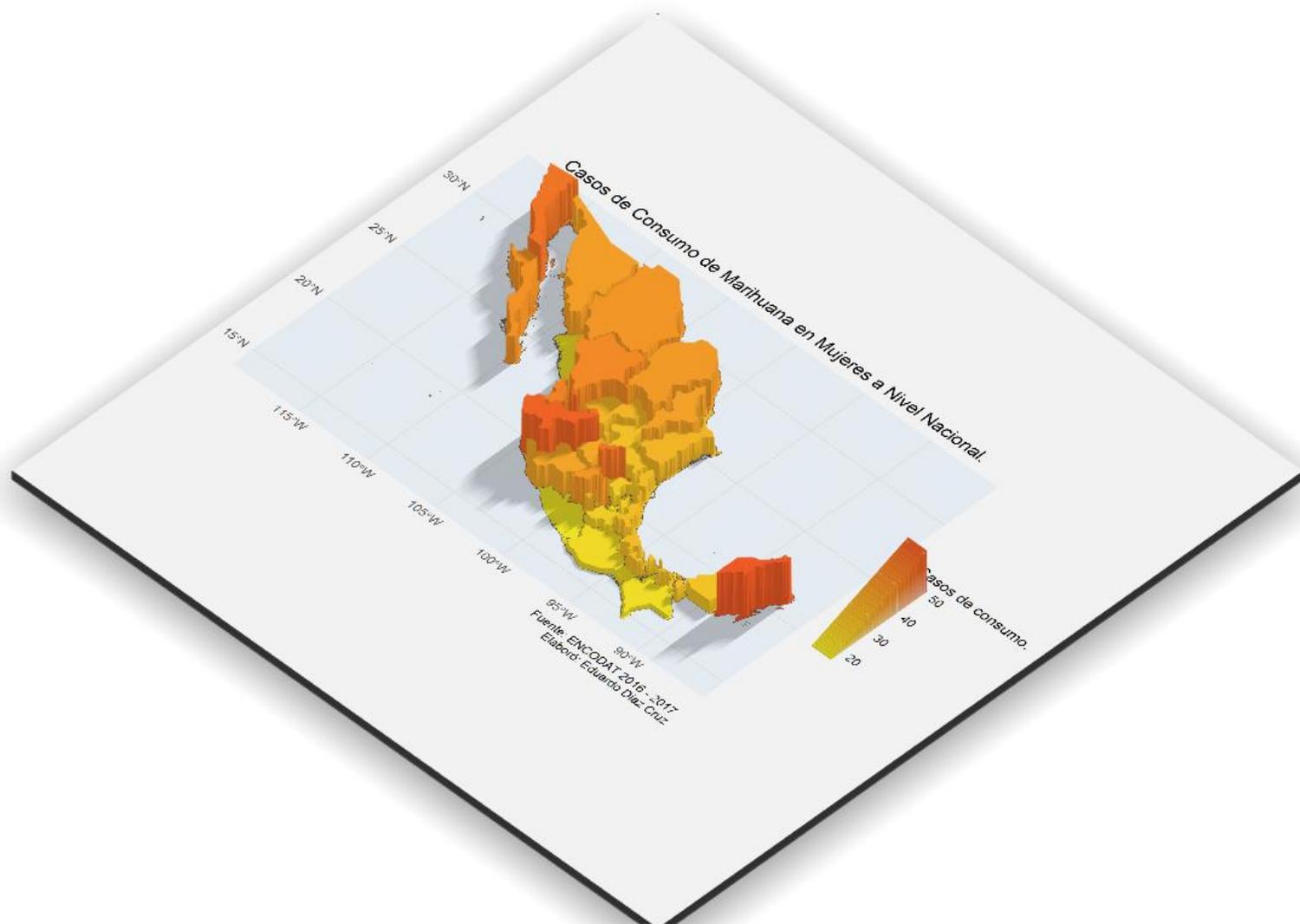
Mapa 6, Casos de Consumo de Heroína en Hombres a Nivel Nacional



Mapa 7, Casos de Consumo de Anfetaminas en Hombres a Nivel Nacional



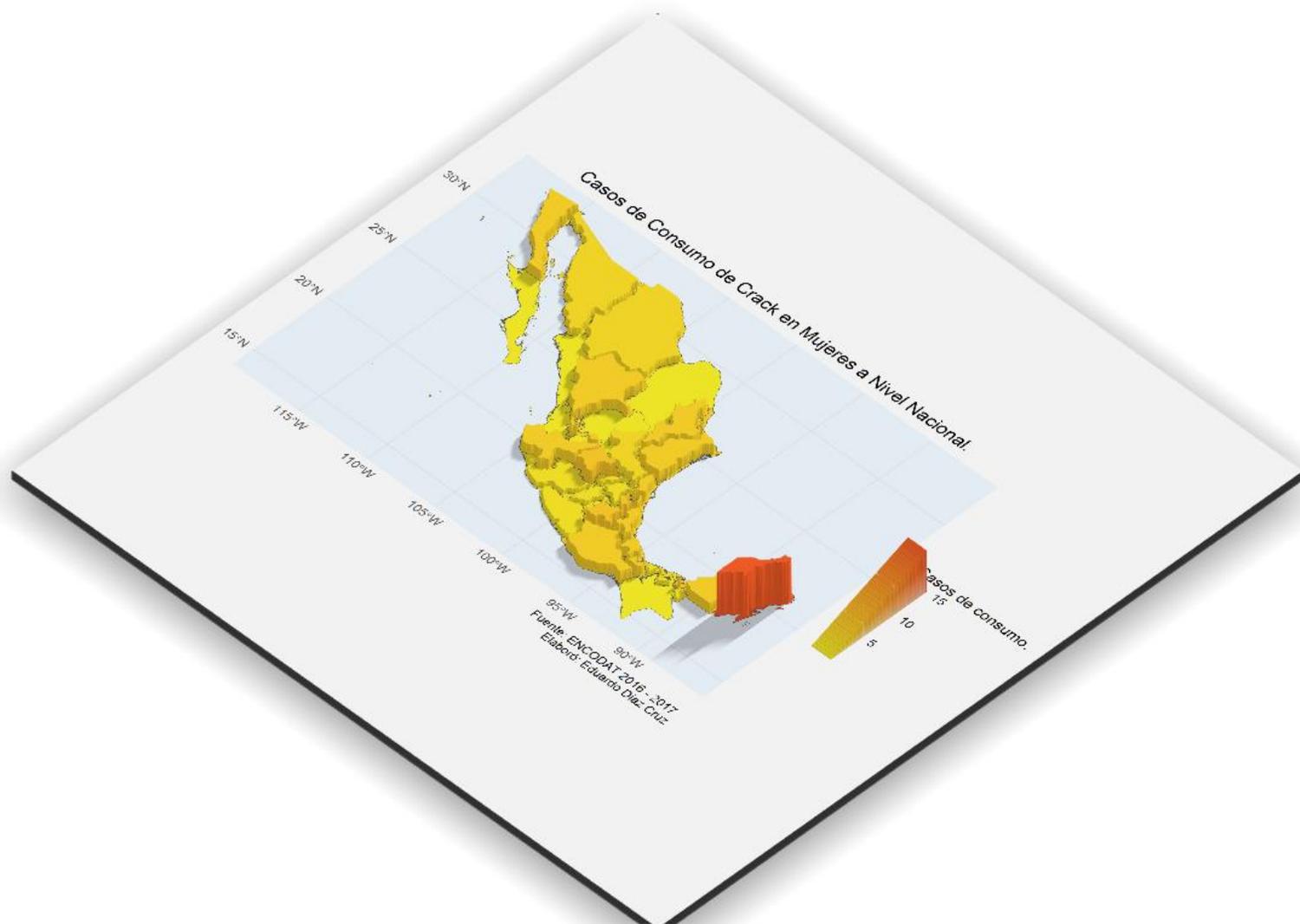
Mapa 8, Casos de Consumo de Marihuana en Mujeres a Nivel Nacional



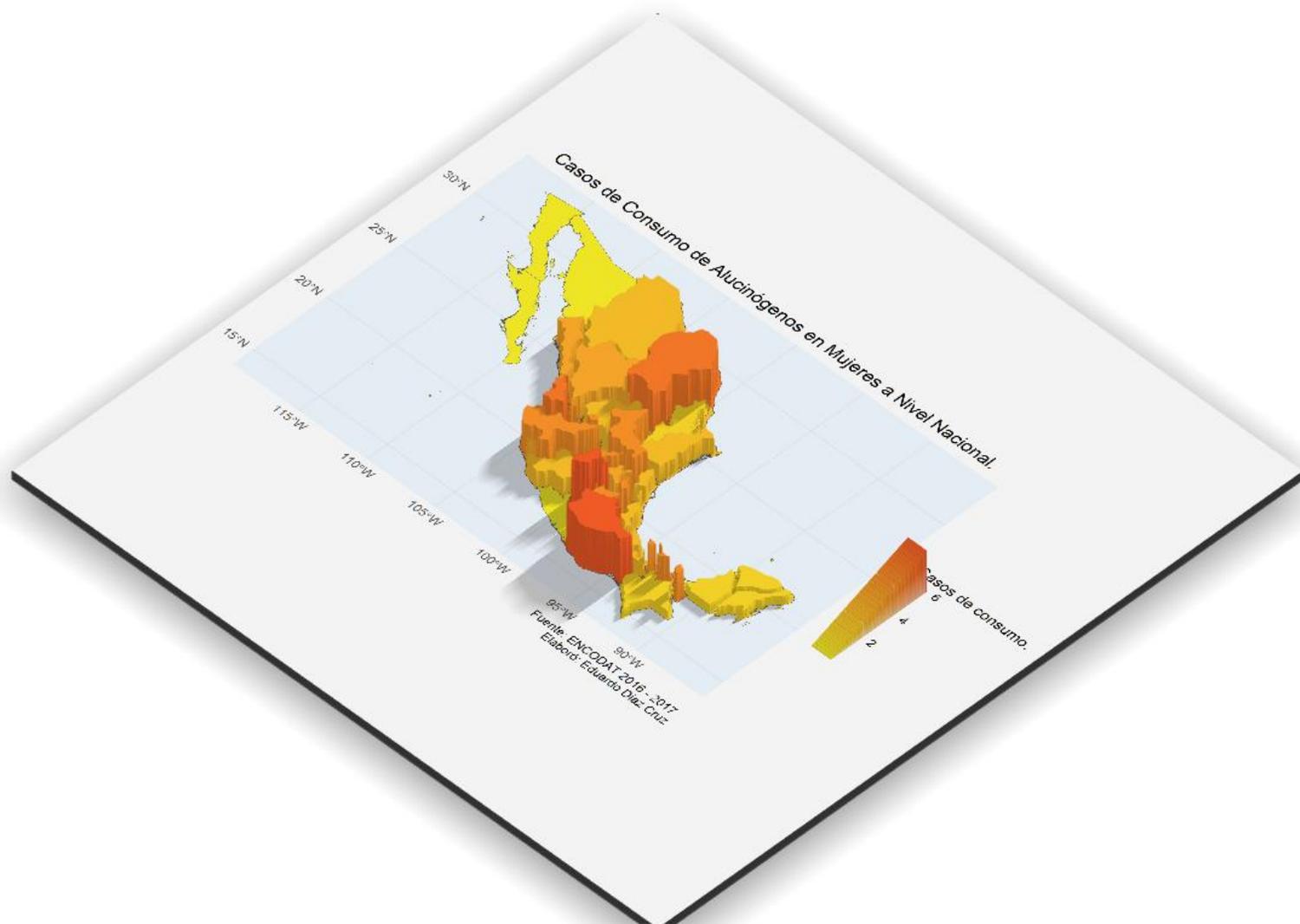
Mapa 9, Casos de Consumo de Cocaína en Mujeres a Nivel Nacional



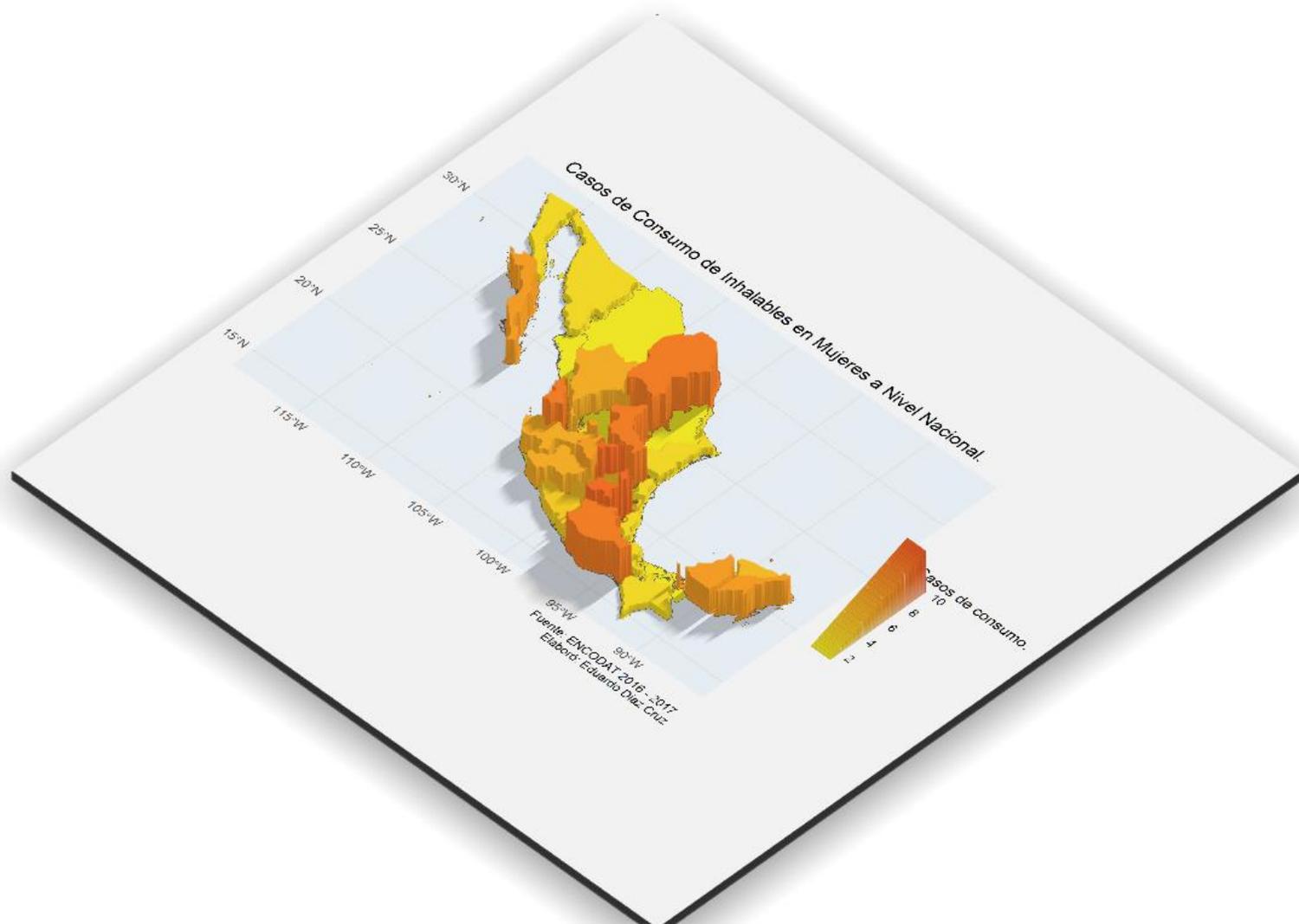
Mapa 10, Casos de Consumo de Crack en Mujeres a Nivel Nacional



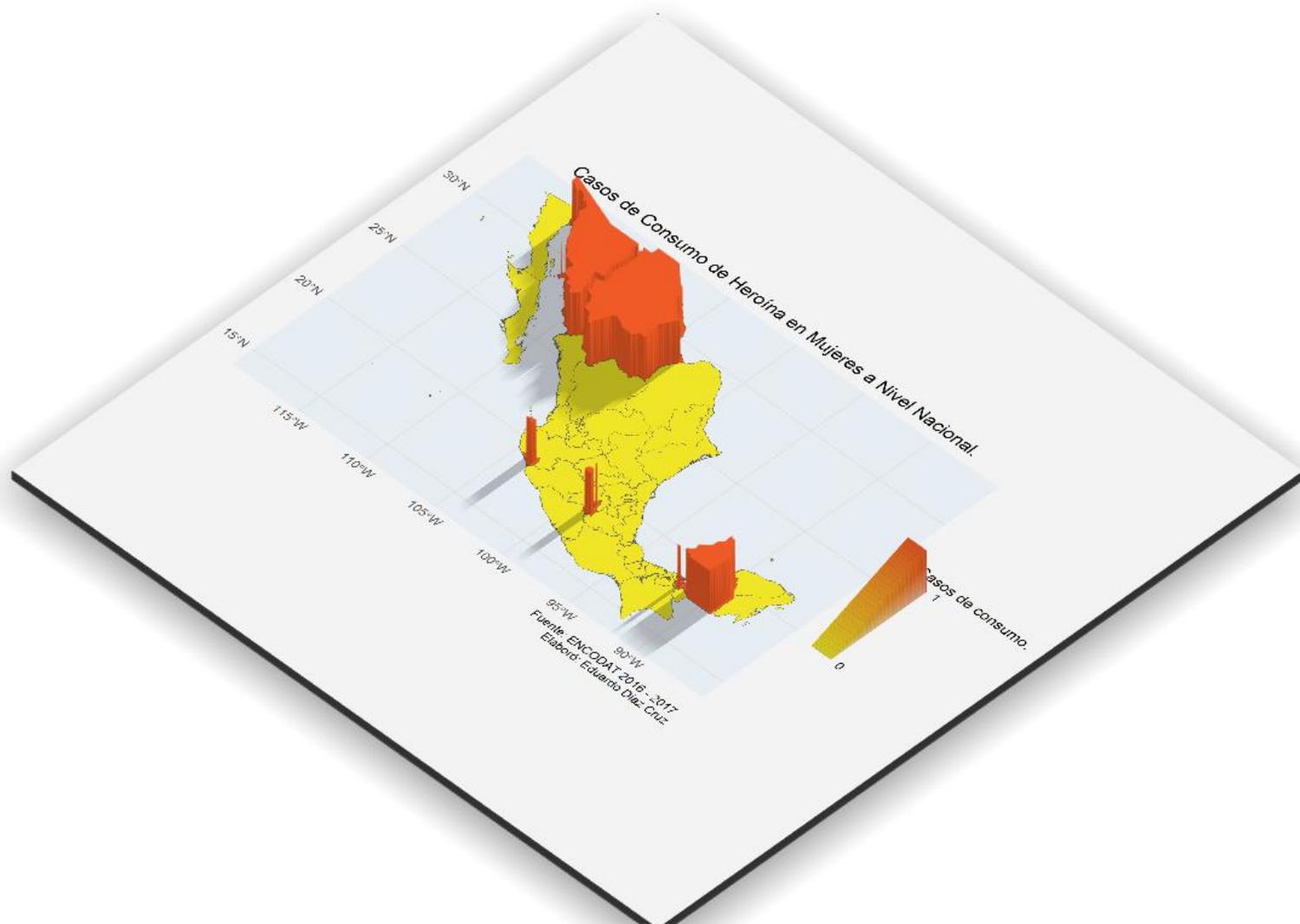
Mapa 11, Casos de Consumo de Alucinógenos en Mujeres a Nivel Nacional



Mapa 12, Casos de Consumo de Inhalables en Mujeres a Nivel Nacional



Mapa 13, Casos de Consumo de Heroína en Mujeres a Nivel Nacional



Mapa 14, Casos de Consumo de Anfetaminas en Mujeres a Nivel Nacional



## Referencias.

1. Antonio, G. M., Campoy, M. R., & Vázquez, F. S. (1994). Topográfica básica para ingenieros. Murcia: Servicio de Publicaciones, Universidad de Murcia.
2. Armstrong, K. A. (2014). Profile of women detected drink driving via Roadside Breath Testing (RBT) in Queensland, Australia, between 2000 and 2011. *Accident Analysis and Prevention*, 67-74.
3. Becoña, E. (2007). Resiliencia y consumo de drogas: una revisión. *Adicciones*, 89-101.
4. Berné Valero, J. L., Anquela Julián, A. B., & Garrido Villén, N. (2014). GNSS GPS: Fundamentos y aplicaciones en Geomática. Valencia, España: Universitat Politecnica de Valencia.
5. Berruecos Villalobos, L. (2007). El consumo de drogas en la ciudad de México. *El Cotidiano*, 105-113.
6. Cheesman Mazariegos, S. S., & Suárez Lugo, N. (2015). Tabaquismo en estudiantes de medicina de la Universidad de San Carlos de Guatemala. *Revista Cubana de Salud Pública*, 18-32.
7. Durá, T. (2002). Consumo de alcohol en alumnos de educación secundaria obligatoria. *Anales del Sistema Sanitario de Navarra*, 281-287.
8. Fernando, C. G. (2016). Drogas: Conceptos Generales, Epidemiología y valoración del consumo. Grupo de Intervención en Drogas semFYC.
9. Fleiz-Bautista, C. D.-G.-V.-Q.-R.-E.-G.-M. (2019). Cuqueando la Chiva: Contextos del consumo de heroína en la frontera norte de México.... México: INPRFM.
10. Gomasasca, M. A. (2009). Basics of Geomatics. Springer Netherlands.
11. INEGI. (2010). Manual de Cartografía Geoestadística. Distrito federal.: INEGI.
12. INEGI. (2014). Delimitación de las zonas metropolitanas. Las zonas metropolitanas en México, mini monografía.
13. Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; Instituto Nacional de Salud Pública, Comisión Nacional Contra las Adicciones; Secretaría de Salud. (2017). Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco 2016-2017: Reporte de Drogas. Ciudad de México, México: Secretaría de Salud.
14. Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; Instituto Nacional de Salud Pública; Secretaría de Salud. (2012). Encuesta Nacional de Adicciones 2011: Reporte de Drogas. México: INPRFM.
15. IOGP. (diciembre de 2020). EPSG Geodetic Parameter Dataset. Obtenido de EPSG Dataset: <https://epsg.org/home.html>
16. Molina Prado, R. (2013). Consumo de tabaco, alcohol y drogas en la adolescencia. *Pediatría Integral*, 205-216.
17. OECD Publishing. (2015). OECD Territorial Reviews: Valle de México, México. OECD Territorial Reviews.

18. Olaya, V. (2011). *Sistemas de Información Geográfica*. Girona, España: Creative Common Atribución.
19. Paradis, E. (2002). *R para principiantes*. Montpellier, Francia: Institut des Sciences de l'Évolution, Université Montpellier II.
20. Río Arronte, Fundación. (20 de diciembre de 2020). Centros de Integración Juvenil, Drogas Legales. Obtenido de Saber sobre drogas: <http://www.cij.gob.mx/SaberSobreDrogas/DrogasLegales.html>
21. Secades-Villa, R. G.-R. (2015). Probability and predictors of the cannabis gateway effect: a national study. *The International journal on drug policy*., 135-142.
22. Sevilla de Lerma, M. J. (2012). *Historias de Matemáticas, Introducción Histórica a la Geodesia*. Pensamiento Matemático, 1.
23. Star, J., & Estes, J. (1990). *Geographic Information Systems: An Introduction*. Prentice-Hall.
24. Vargas Pineda, D. R. (2001). Alcoholismo, Tabaquismo y Sustancias Psicoactivas. *Salud Pública* Vol. 3, 74-88.
25. Waters H, S. d. (2010). *La Economía del Tabaco y los Impuestos al Tabaco en México*. Unión Internacional contra la Tuberculosis y Enfermedades Respiratorias.
26. Weinberger, A. H. (2016). Is cannabis use associated with an increased risk of onset and persistence of alcohol use disorders? A three-year prospective study among adults in the United States. *Drug and alcohol dependence*, 363-367.