



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Estudio sobre incidencias de tráfico en la ciudad de Valencia

Trabajo Fin de Grado

Grado en Ingeniería Informática

Autor: Raúl García Crespo

Tutor: César Ferri Ramírez

Curso: 2018/2019

Resumen

Los accidentes de tráfico representan una de las principales causas de muerte no natural en el mundo, e incluso si no llegan a causar víctimas mortales sus consecuencias pueden afectar a los implicados tanto física como económicamente. Conocer cómo se han producido estos incidentes y cuáles son sus causas más comunes es casi indispensable si se pretende reducir su número.

Este trabajo de Fin de Grado realiza un estudio estadístico de los registros de incidentes de tráfico recogidos por la Policía Local de Valencia con el propósito de extraer patrones y pautas de estos sucesos que puedan ayudar en la elaboración de planes de prevención y de asistencia. Para ello se hace uso del entorno de desarrollo Rstudio y el lenguaje de computación estadística R, el cual nos ofrece también la posibilidad de elaborar gráficas y mapas detallados con los que representar nuestros resultados. Con estas herramientas se han analizado los datos disponibles en base a periodos de tiempo específicos, a condiciones meteorológicas peligrosas, a las circunstancias que hayan provocado el incidente y a su localización. Los patrones que se han obtenido como resultado permiten localizar problemas de tráfico específicos para la ciudad de Valencia.

Palabras clave: ciencia de datos, accidentes, delitos, incidentes, tráfico, análisis, R, RStudio, estadístico.

Abstract

Traffic accidents represent one of the main causes of non-natural death at a global scale and, even if they don't cause mortal victims, their consequences can affect the people involved in them both physically and economically. Knowing how this incidents take place and what their more common causes are is almost indispensable if the objective is to reduce their number.

This essay carries out a statistical study of the records of traffic incidents collected by the Local Police of Valencia with the purpose of obtaining patterns in these events which can help in the elaboration of prevention and assistance plans. For it, we make use of the statistic computing language R and the development environment RStudio, which offers the possibility to create detailed graphics and maps to illustrate our results. With these tools, the available data has been analysed based on specific time frames, dangerous meteorological conditions, the circumstances which cause the incidents and their location. The patterns obtained as result let us identify particular traffic problems in the city of Valencia.

Keywords: accidents, crimes, incidences, traffic, analysis, R, RStudio, statistic

Agradecimientos

Se agradece al Ayuntamiento de Valencia, en especial al cuerpo de la Policía Local de Valencia, su colaboración al haber facilitado datos indispensables para la realización de este estudio.

También se agradece su colaboración a César Ferri Ramírez, tutor de este trabajo, por su supervisión y sus consejos con los que ha ayudado a la realización y múltiples mejoras del estudio que aquí se expone.

Tabla de contenidos

1.	Introducción	9
1.1.	Motivación	9
1.2.	Objetivos	10
1.3.	Estructura del trabajo.....	10
2.	Incidencias en el tráfico urbano	13
3.	Herramientas empleadas.....	16
3.1.	R y RStudio	16
3.2.	Convertor JSON a CSV	17
3.3.	Google Cloud Platform	17
4.	Descripción de los datos de estudio.....	18
4.1.	Datos proporcionados por la Policía Local.....	18
4.2.	AEMET OpenData.....	18
5.	Análisis y depuración de los datos.....	21
5.1.	Unión de tablas.....	21
5.2.	Formato de fechas.....	22
5.3.	Formato de decimales	22
5.4.	Errores de clasificación.....	23
6.	Estudios y resultados.....	24
6.1.	Incidentes en base a periodos temporales.....	24
6.1.1.	Incidentes por mes	24
6.1.2.	Incidentes por hora.....	27
6.2.	Incidentes en base a las condiciones climáticas	31
6.2.1.	Nº de incidentes bajo precipitaciones fuertes.....	31
6.2.2.	Nº de incidentes durante temperaturas máximas extremas	34
6.2.3.	Nº de incidentes durante rachas fuertes de viento.....	36
6.3.	Incidentes en base a otras causas.....	37
6.3.1.	Principales causas de incidentes	37
6.3.2.	Causas de incidentes en los periodos de mayor actividad	40
6.4.	Incidentes según la localización.....	43
6.4.1.	Obtención de las coordenadas	43

6.4.2.	Distribución y densidad de incidentes	46
6.4.3.	Distribución y densidad de incidentes en base a sus causas y circunstancias.....	50
7.	Conclusiones	54
7.1.	Análisis de los resultados	54
7.2.	Propuestas de mejora.....	57
7.3.	Relación con los estudios cursados.....	58
8.	Bibliografía.....	60
9.	Anexo	62
9.1.	Tabla descriptiva del contenido de los archivos “RED-ATESTADOS 20XX” .	62
9.2.	Gráfica de incidentes en función del mes	67
9.3.	Gráficas de incidentes en función de la hora	69
9.4.	Gráficas de cantidad de incidentes según su causa	73
9.5.	Gráficas de causas de incidente por intervalo de tiempo	77
9.6.	Mapas.....	80
9.6.1.	Mapa base de la ciudad de Valencia	80
9.6.2.	Mapa de densidad de los incidentes sucedidos entre 2010 y 2017.....	81
9.6.3.	Mapa de densidad de los incidentes sucedidos entre 2010 y 2017 incluyendo las direcciones de los puntos de máxima densidad.....	82
9.6.4.	Mapa de densidad y puntos de máximos de accidentes causados por colisión entre 2010 y 2017	83
9.6.5.	Mapa de densidad y puntos máximos de los delitos por alcoholemia entre 2010 y 2017.....	84
9.6.6.	Mapa de densidad y puntos máximos de las sanciones administrativas por alcoholemia entre 2010 y 2017	85
9.6.7.	Mapa de densidad y puntos máximos de los accidentes graves entre 2010 y 2017	86
9.6.8.	Mapa de densidad y puntos máximos de los incidentes sucedidos entre las 4:00 y las 5:00.....	87
9.6.9.	Mapa de densidad y puntos máximos de los incidentes sucedidos entre las 12:00 y las 14:00.....	88
9.6.10.	Mapa de densidad y puntos máximos de los incidentes sucedidos entre las 19:00 y las 20:00.....	89
9.7.	Ejemplos de código	90
9.7.1.	Ejemplo de construcción de un gráfico de líneas	90
9.7.2.	Ejemplo de construcción de un gráfico Box-Plot.....	90
9.7.3.	Ejemplo de construcción de un mapa de densidad.....	90

1. Introducción

1.1. Motivación

En la actualidad se producen decenas de accidentes de tráfico al día en toda España, parte de los cuales pueden suponer pérdidas económicas para las víctimas e implicados, e incluso llegar a provocar la pérdida de vidas humanas. Según la Organización Mundial de la Salud, 1.35 millones de personas fallecen al año en todo el mundo como consecuencia de los accidentes de tráfico, teniendo la mayoría de estas edades comprendidas entre los 5 y los 29 años. Solo en España, durante el año 2017 se registraron más de 102000 accidentes con víctimas, 64000 de los cuales se produjeron en vías urbanas. Al mismo tiempo, la principal causa de accidente mortal durante este año fueron las distracciones al volante, seguidas por infringir los límites de velocidad y el consumo de drogas y alcohol, representando estos dos últimos un 50 % de los accidentes mortales. Conocer las principales causas que provocan este tipo de incidentes es crucial si se pretenden elaborar estrategias enfocadas a disminuir el número, no solo de víctimas en carretera, sino también de los tipos de accidentes más frecuentes cuyo coste económico puede suponer un importante porcentaje de gasto público. Como ya hemos dicho, las causas más habituales para los accidentes de tráfico ya son conocidas a nivel de España, sin embargo, cuanto más específicos sean estos datos, mejor enfocados a los problemas reales serán nuestros planes de prevención. Es por ello que un análisis de los distintos incidentes de tráfico en la ciudad de Valencia nos ofrecerá datos concretos que nos indiquen los problemas específicos que existen en sus calles, lo cual facilitará un reparto de recursos eficientes por parte de las fuerzas del orden al asistir o prevenir dichos incidentes.

En conjunto con estos hechos, vivimos en un mundo donde la cantidad de datos que se producen y almacenan diariamente ha superado cualquiera de las expectativas que podríamos haber tenido en el pasado, y la información de los incidentes de tráfico no es una excepción. La Policía Local de Valencia mantiene un registro de cada incidencia de tráfico, ya sea accidente o infracción, que haya requerido la presencia de personal policial y en el que se recogen todos los datos de los vehículos implicados y la situación del incidente. Estos datos se almacenan en caso de que en algún momento sea necesario consultar los detalles de algún incidente por causas judiciales o personales por parte de los implicados. Sin embargo, con la ayuda de las herramientas y el software adecuados, estos datos pueden ser también analizados de forma estadística en busca de patrones que nos muestren el impacto de factores tales como la meteorología, los periodos de máxima actividad de tráfico o la localización. Dichos patrones o pautas nos indicarán las principales fuentes de riesgo para el tráfico de la ciudad.

Siguiendo los principios de otros estudios similares, los cuales se introducen en el apartado 2 , este trabajo trata por lo tanto de averiguar cuáles son los principales problemas para el tráfico en la ciudad de Valencia mediante un correcto tratamiento de

los datos que incluye la obtención de datos fiables a partir de fuentes de confianza, la depuración de dichos datos en preparación para su estudio y el uso de herramientas estadísticas adecuadas para cada uno de los diferentes datos de los que disponemos.

1.2. Objetivos

El principal objetivo de este trabajo es realizar un análisis de datos estadístico de los registros correspondientes a la red de atestados de la Policía Local de Valencia entre los años 2010 y 2017, ofreciendo como resultado tablas, gráficas y mapas que representen los patrones y cambios que se han observado en los datos a lo largo de los diferentes años estudiados. En este estudio se recogen por lo tanto todos los pasos necesarios para alcanzar ese objetivo, incluyendo la adquisición, el tratamiento y la manipulación de los datos y las diferentes herramientas empleadas durante el desarrollo de este. Para lograr dicho objetivo se han establecido varios sub-objetivos específicos centrados en aspectos concretos de nuestro objetivo principal:

- Detectar los periodos temporales para los que se produzcan incidentes con mayor frecuencia.
- Establecer si existe o no una relación entre distintas condiciones meteorológicas y el número de incidentes registrados bajo dichas condiciones.
- Averiguar los motivos principales causantes de la mayoría de incidentes, manteniendo una diferenciación entre delitos y accidentes.
- Localizar los puntos geográficos de la ciudad y su periferia que más riesgo suponen para el tráfico.

Por otro lado y como objetivo secundario se analiza la utilidad de los datos desde los que parte este trabajo, ofreciendo consejo sobre cómo mejorar estos para futuros estudios similares en base a las dificultades que se han encontrado durante el desarrollo de este proyecto. De igual modo, se analiza de forma objetiva la utilidad y la validez de todos los resultados que se obtienen en este trabajo, de nuevo, ofreciendo consejo y posibles mejoras para futuros trabajos similares.

1.3. Estructura del trabajo

Incidentes en el tráfico urbano

Este estudio comienza analizando otros trabajos de índole similar al nuestro, explicando los objetivos de estos, las herramientas que emplean y las conclusiones que se alcanzaron con el fin de que sirvan de guía para nuestro propio trabajo.

Herramientas empleadas

A continuación se detallan todas las herramientas que han sido necesarias para llevar a cabo nuestro estudio, desde los lenguajes y entornos de desarrollo con los que han sido tratados nuestros datos, hasta las librerías y paquetes de funciones de terceros de los que se ha hecho uso, explicando su utilidad en nuestro trabajo.

Descripción de los datos

Una explicación del contenido de los datos disponibles y de los diferentes campos que estos registran, junto con los posibles valores que pueden tomar, nos ayudará a comprender qué tipo de estudios seremos capaces de obtener. Adicionalmente, conocer las fuentes y los métodos mediante los que se han obtenido nuestros datos garantiza la posibilidad de analizar la veracidad y fiabilidad de estos.

Análisis de los datos

Este paso de nuestro trabajo consiste en el análisis y la depuración de la información contenida en los ficheros que vamos a usar como base para nuestro estudio. Este proceso siempre sería necesario en estudios similares, puesto que no es una buena práctica aceptar todos los datos que recibimos como si fueran correctos y suele ser preferible un análisis de los mismos para poder asegurar que no crearan errores más adelante; sin embargo es mucho más necesario en el caso que aquí se nos presenta dado que las fuentes de información principales en las que nos vamos a basar (las tablas Excel con la información de los incidentes de tráfico) han sido rellenadas a mano por el personal de la Policía Local, lo cual supone una potencial fuente de posibles errores humanos en los datos.

Este proceso también incluye la carga de los datos a nuestro entorno de desarrollo, *RStudio*, y cualquier reordenación que sufran las tablas en preparación para su estudio, como por ejemplo el añadido de columnas adicionales o la unión de diferentes tablas.

Estudios y resultados

Una vez garantizada la validez de los datos procederemos a estudiar diferentes factores de los mismos que puedan haber influido en la cantidad y distribución de los accidentes e infracciones de tráfico en la ciudad de Valencia. Los factores elegidos para el estudio son:

- Periodos temporales significativos. En este apartado se estudia cuáles son los periodos de tiempo en los que se produce una mayor cantidad de incidentes, al mismo tiempo que se ofrecen posibles explicaciones para ellos. Se analizan estos periodos a nivel mensual, semanal y diario.
- Condiciones climatológicas. Donde se estudiará el efecto de distintas condiciones meteorológicas adversas que pueden influir tanto sobre los conductores y peatones, como sobre los vehículos y las condiciones de la calzada.
- Circunstancias del incidente. Los factores que aquí se estudiarán serán aquellos que conformen las características propias del incidente como por ejemplo la causa del incidente o si estuvieron involucrados el consumo de alcohol o drogas.

- Localización. Aquí se analizarán aquellos tramos de carretera de la ciudad de Valencia que acumulan mayor cantidad de accidentes de tráfico.

A medida que el trabajo avance y los diferentes factores estudiados aporten sus resultados, puede resultar interesante realizar estudios cruzados con factores de las diferentes categorías que aquí se exponen con el fin de obtener resultados aún más precisos. Un ejemplo de este tipo de estudios cruzados podría ser el de analizar los puntos geográficos de mayor riesgo durante el periodo de tiempo en el que se haya detectado una mayor actividad.

Conclusiones

En este último apartado se recogen los patrones más representativos encontrados en nuestros análisis, se ofrecen propuestas de mejora para futuros trabajos similares y se relacionan los conocimientos aplicados durante el estudio con las asignaturas cursadas durante el Grado en Ingeniería Informática.

2. Incidencias en el tráfico urbano

Datos de accidentes en carreteras a nivel mundial son recogidos y publicados en informes anuales de la Organización Mundial de la Salud como indicadores para la supervisión de su Plan Global de la Década de Acción [18], un proyecto que tiene como objetivo aumentar la seguridad en las carreteras de 110 países. Otros datos contenidos en este tipo de informes son los principales factores de riesgo que influyen en las cifras de muertos y heridos en accidentes de tráfico, siendo algunos de estos superar los límites de velocidad establecidos, conducir bajo la influencia de alcohol y drogas o no usar el cinturón de seguridad en turismos o el casco en el caso de las motocicletas. Cada uno de estos factores se nos presenta acompañado por diferentes estadísticas que nos ofrecen información de la importancia y gravedad de dichos factores en los accidentes de tráfico. Si bien los informes de la Organización Mundial de la Salud pueden ser útiles a la hora de ilustrar la gravedad y el impacto que los accidentes tienen en nuestra vida diaria, sus datos no ofrecen información detallada de estos.

A nivel nacional, en el caso concreto de España, es la Dirección General de Tráfico [16] quien recoge y publica estadísticas anuales del tráfico en su página web. Estos datos accesibles para cualquier ciudadano recogen las cifras anuales de tráfico desde 1998 hasta la actualidad en las carreteras Españolas. En ellas se encuentran el número de accidentes e infracciones totales y el número de heridos o muertos en carreteras. Adicionalmente, también se ofrece la posibilidad de consultar estos datos teniendo en cuenta distintos factores: según la provincia, según el mes u hora del día, según el tipo de accidente o en función de la vía donde tuvo lugar el incidente, entre otras. Junto con los datos se incluye un manual de términos en el que se definen algunas de las categorías según las cuales es posible clasificar los incidentes como por ejemplo cuando se ha considerado que un accidente de tráfico ha tenido víctimas asociadas o en qué circunstancias se considera que una víctima ha sido herida de gravedad. Los datos que aquí se exponen no han sido tratados con herramientas de software estadístico, sino que simplemente se han empleado hojas de cálculo para la contabilización y representación de los incidentes en forma de tablas.

Entre los estudios centrados en zonas urbanas podemos encontrar entre otros los trabajos *Traffic Analysis and Road Accidents: A Case Study of Hyderabad using GIS* [19] y *Spatio-Temporal Analysis on Urban Traffic Accidents: A Case Study of Tehran City, Iran*, [20] los cuales emplean datos de tráfico de los ciudades de Hyderabad, en la India, y Teherán, en Irán, respectivamente. Mientras que el primero centra sus esfuerzos en identificar las principales causas de accidentes en la ciudad, analizando los factores de riesgo de mayor influencia y los puntos geográficos más propensos a accidentes; el segundo se centra en la distribución de los incidentes de tráfico de la ciudad en función del espacio y el tiempo, es decir, analizan en que periodos (horas, días o meses) se producen accidentes con una mayor frecuencia en combinación con una representación



de los puntos y zonas de la ciudad que registran mayor actividad de estos incidentes en dichos periodos. Ambos trabajos hacen uso de diferentes Sistemas de Información Geográfica (GIS por sus siglas en inglés) para la referenciación precisa de los puntos geográficos estudiados. A su vez los autores de ambos trabajos ven en las conclusiones de sus estudios una oportunidad para la elaboración de posibles planes y la creación de leyes de seguridad vial más estrictas centradas en los principales problemas y causas que sus respectivos estudios han hallado como más influyentes en las cifras de accidentes.

Otro estudio titulado *An analysis of traffic incidents on an Australian urban road network* [21], centrado no solo en una ciudad concreta sino en la totalidad de las carreteras urbanas de Australia también plantea su metodología en analizar los periodos de tiempo y lugares de mayor frecuencia de accidentes. Sin embargo, a diferencia de los estudios anteriormente nombrados, este incluye entre sus datos el impacto causado por cada uno de los incidentes que estudia en el resto del tráfico, ya que dispone del tiempo durante el cual las consecuencias de los accidentes bloquearon total o parcialmente las vías en las que se produjeron, alterando en flujo habitual del tráfico. Otra novedad de este trabajo es su consideración del clima como un factor de riesgo para la seguridad vial, ya que este incluye estadísticas relacionadas con los incidentes causados por lluvias e inundaciones, así como otras posibles fuentes de peligro como la presencia de elementos extraños en la calzada o los incendios.

Si nos centramos en la zona de Valencia, en portales web como el del Ayuntamiento de Valencia [17] es posible encontrar datos relacionados con los accidentes de tráfico sucedidos en la ciudad, pero estos tan solo representan un pequeño conjunto de estadísticas entre otros datos relativos al tráfico. Estos datos presentan el mismo tratamiento que aquellos ofrecidos por la DGT, hojas de cálculo con el conteo total de incidentes, sin desarrollar ningún aspecto específico y sin estudiar los puntos de mayor siniestralidad.

No se han hallado otros estudios como los ya comentados que analicen en detalle la información del tráfico disponible de la ciudad de Valencia, con la excepción del Trabajo de Final de Master “Estudio de patrones de accidentes en la ciudad de Valencia” [22] desarrollado por el alumno de Master Universitario en Gestión de la Información, Joaquín Montesinos Muñoz. El objetivo de dicho estudio es el diseño y el desarrollo de una aplicación web destinada a la visualización y predicción de los accidentes de tráfico en Valencia y su periferia para el cual se emplearon datos a disposición del público por parte de la DGT. En un principio este trabajo de Master también incluye los datos en los que se basa nuestro estudio, sin embargo, a causa de un retraso en la entrega de dichos datos por parte de la Policía Local de Valencia, el alumno Joaquín Montesinos los recibió en las últimas etapas de su trabajo por lo que no se pudieron incluir completamente. Nuestro trabajo reutiliza esos datos proporcionados específicamente por las fuerzas de policía para elaborar un informe detallado sobre los incidentes de tráfico, tanto infracciones como accidentes.

Otro estudio, no relacionado directamente con los accidentes sino con los métodos de registro de estos, lo encontramos en *Urban Traffic Accident data collection*

and analysis in Europe: Survey study [23] por Carlos Martínez Pérez. Este estudio parte de diferentes encuestas realizadas por entidades similares a la DGT y otros expertos en tráfico en diferentes países de Europa con el objetivo de diagnosticar la situación de los procedimientos, sistemas y calidad de los registros de accidentes de tráfico llevados a cabo a un nivel local en toda Europa. En él se comparan los datos registrados, el entrenamiento de los agentes y los métodos de almacenamiento para cada país participante en las encuestas, entre otros factores. Como resultados, este estudio encuentra las siguientes dificultades en los métodos y procedimientos de registro de incidentes de tráfico: no existen métodos electrónicos para registrar los accidentes directamente desde el lugar del suceso, la información registrada no se ajusta a las necesidades locales o a las nuevas tecnologías, los sistemas de posicionamiento empleados para situar el accidente no son fiables, discrepancias en el criterio para clasificar la gravedad de los heridos... En el caso concreto de los sistemas de registro empleados en España este estudio ha encontrado que el 32% de los municipios españoles no analizan estadísticamente sus datos de accidentalidad, y los que lo hacen suelen emplear como herramientas las hojas de cálculo.

En cuanto a lo referente al análisis de los datos una vez registrados, el informe *Guía de 'buenas prácticas' para la recogida, gestión y análisis de datos de accidentes de tráfico urbanos* [24] ofrece un conjunto de recomendaciones sobre las herramientas y tratamiento de datos necesarios para la elaboración de análisis estadísticos correctos.

3. Herramientas empleadas

En este apartado se detallan todos los programas, paquetes de funciones y servicios que han sido necesarios para la realización de este estudio, justificando el porqué de su uso.

3.1. R y RStudio

La mayor parte de este trabajo se ha realizado empleando el lenguaje R [versión 3.5.1 (2018-07-02)], tanto para manipular los datos como para crear los gráficos que representen los resultados de este estudio. Se ha escogido este lenguaje frente a otros como Python por su especificación en el análisis estadístico, su inclusión de numerosos modelos estadísticos, que pueden resultar útiles durante la elaboración de este trabajo; su capacidad para construir y diseñar gráficos y mapas detallados, que forman una parte esencial de este estudio; y por la gran extensión del archivo público de paquetes de R, el cual nos ofrece acceso a numerosas funciones adicionales.

Junto con este lenguaje es necesario un entorno de desarrollo para este. Se ha escogido RStudio [1] por ser un entorno de software libre, centrado únicamente en R, y por algunas de sus funcionalidades tales como la ejecución de código directamente desde el editor de código, lo que nos permite ejecutar y comprobar nuestro código de forma ágil y sencilla, la inclusión de documentación en el propio entorno y la administración de múltiples paquetes de funciones adicionales.

Adicionalmente a las funciones básicas que acompañan a la versión de R empleada, también serán necesarios ciertos paquetes de librerías para tener acceso a funcionalidades adicionales:

- **Paquete *xlsx*** versión 0.6.1 [5] Este paquete contiene las funciones que serán necesarias para la importación de nuestras tablas Excel al entorno de desarrollo de R.
- **Paquete *plyr*** versión 1.8.4 [6] Este paquete contiene funciones que nos serán útiles a la hora de fusionar las diferentes tablas de datos de las que disponemos sin perder ninguna información.
- **Paquetes *ggmap* y *ggplot*** versiones 3.0.0 y 3.1.1 respectivamente [7] [8] Estos paquetes contienen funciones que serán necesarias para la elaboración de mapas y la representación de información sobre estos, y para la elaboración de gráficas personalizadas.
- **Paquetes *RCurl* y *RJSONIO*** versiones 1.95-4.12 y 1.3-1.1 respectivamente [9] [10] Estos paquetes nos permiten construir nuestras propias funciones con las que realizar consultas a la API de Google Cloud Platform y obtener las respuestas en formato JSON.
- **Paquete *Hmisc*** versión 4.2-0 [12] Este paquete contiene varias funciones de utilidad que emplearemos durante la fase de estudio y resultados.

- **Paquetes *MASS* y *raster*** versiones 7.3-51.4 y 2.9-5 respectivamente [14] [15] Estos paquetes nos permitirán definir y crear objetos de tipo *kernel* y *raster* necesarios para el cálculo de puntos máximos de un mapa de densidad.

3.2. Convertor JSON a CSV

Algunos de los datos con los que se va a trabajar los recibiremos en formato JSON. Antes de poder introducir estos a nuestro entorno de R será necesario realizar un cambio de formato a CSV, para ello hemos empleado una herramienta de conversión online JSON to CSV [2].

3.3. Google Cloud Platform

Google Cloud Platform [11] nos ofrece acceso a una plataforma de almacenamiento, procesamiento y análisis de datos de gran potencia. También nos ofrece un conjunto de APIs de diferentes servicios ofrecidos por Google. En el caso de este trabajo, emplearemos las APIs de geocodificación y de construcción estática de mapas.



4. Descripción de los datos de estudio

4.1. Datos proporcionados por la Policía Local

La principal fuente de información en la que se va a basar este estudio está comprendida por 8 archivos en formato Excel proporcionados por la Policía Local de Valencia, cada uno correspondiendo con un año entre el 2010 y el 2017, ambos inclusive. Dichos archivos consisten en un registro detallado de cada delito o accidente de tráfico sucedido en el año correspondiente en la ciudad de Valencia y sus alrededores para los que fue necesaria presencia policial. Debido a la gran cantidad de columnas que componen estos registros, la explicación completa de cada uno de los datos registrados y su formato se incluyen en el Anexo 9.1. de este trabajo. De manera más resumida, estos archivos contienen, entre otros datos, información sobre:

- La fecha y hora de los incidentes,
- La localización de los incidentes,
- El número y tipo de vehículos implicados,
- El número de víctimas y su gravedad,
- El tipo de choque o delito que origina el incidente.

Si bien la mayoría de las columnas que se describen en el anexo aparecen en cada uno de los ficheros con los que vamos a trabajar, algunas como por ejemplo “2ª PRUEBA-COND,2” no aparecen en los datos de los primeros años y otras pueden cambiar de nombre de un año para otro; esto debe tenerse en cuenta a la hora de automatizar la construcción de gráficas, pues puede dar lugar a numerosos errores.

4.2. AEMET OpenData

Adicionalmente a los datos ya descritos y para complementar nuestro estudio se han recopilado los datos del clima correspondientes a cada uno de los días para los cuales se dispone de datos sobre el tráfico. Para ello se ha empleado la información de acceso general disponible en AEMET OpenData, una plataforma que ofrece un sistema de difusión de la información de la Agencia Estatal de Meteorología. Antes de poder acceder a los datos, es necesario registrarse en dicha plataforma para poder obtener una *key* que se nos requerirá cada vez que realicemos una consulta.

AEMET OpenData nos permite acceder a diferentes tipos de información:

- Valores climáticos almacenados (Divididas por fecha y/o localización).
- Predicciones climáticas (Divididas por fecha y/o localización).
- Productos del análisis de los valores climáticos como el balance hídrico.
- Avisos de fenómenos meteorológicos adversos.
- Mapas y gráficos.

Una vez hecha la consulta se nos devolverá una URL en caso de que esta se haya resuelto satisfactoriamente. En dicha URL encontraremos los datos solicitados en formato JSON.

Para el desarrollo de este trabajo vamos a emplear los valores climáticos diarios almacenados correspondientes a la ciudad de Valencia entre los años 2010 y 2017. Dado que al realizar una consulta de los datos en un segmento de tiempo el máximo lapso de tiempo permitido para dicho periodo es de un mes, procederemos a realizar manualmente cada una de las consultas correspondientes a cada uno de los meses de cada año estudiado, es decir, un total de 96 consultas para el periodo de 8 años que se va a estudiar. Cada una de las consultas que realicemos nos exigirá las siguientes entradas:

- Provincia.
- Estación meteorológica.
- Fecha inicio.
- Fecha fin.

Las tablas con las que vamos a trabajar se han obtenido con los datos de la provincia de Valencia; la estación 8416Y, por ser la más cercana al centro de la ciudad de Valencia; y el primer y último día de cada mes estudiado.

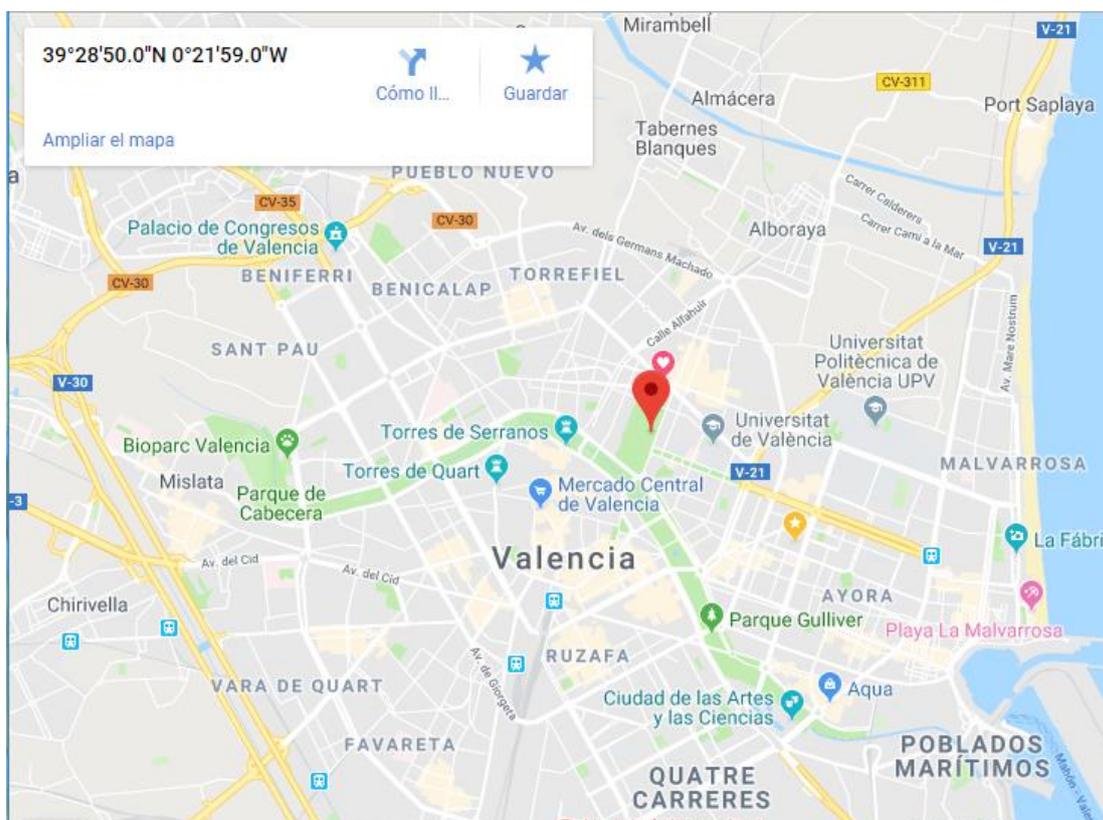


Figura 1: Localización y coordenadas de la estación 8416Y

En las tablas JSON que nos serán devueltas por AEMET OpenData para cada consulta encontraremos los siguientes formatos para los pares identificador-valor:

IDENTIFICADOR	FORMATO DEL VALOR	EXPLICACIÓN
fecha	Expresada en formato dd/mm/aaaa	Día en el que se obtuvieron los datos
indicativo	Cadena de números y letras	Identificador de la estación meteorológica
nombre	Cadena de símbolos	Designación de la estación meteorológica
provincia	Cadena de símbolos	Provincia donde se localiza la estación
altitud	Expresada en m	Altitud de la estación sobre el nivel del mar
tmed	Expresada en C°	Temperatura media
prec	Expresada mm/h	Volumen de las precipitaciones
tmin	Expresada en C°	Temperatura mínima
horatmin	Expresada en formato hh:mm	Hora a la que se alcanza la temperatura mínima
tmax	Expresada en C°	Temperatura máxima
horatmax	Expresada en formato hh:mm	Hora a la que se alcanza la temperatura máximo
dir	Número entero entre 1 y 12 representando la posición de las manecillas en un reloj analógico	Dirección media del viento
velmedia	Expresada en km/h	Velocidad media del viento
racha	Expresada en km/h	Máximo alcanzado por la velocidad del viento
horaracha	Expresada en formato hh:mm	Hora a la que se alcanza la velocidad máxima del viento

Figura 2 Tabla explicativa de los posibles valores para los datos meteorológicos

Del mismo modo que ocurría para las tablas Excel proporcionadas por la Policía Local, algunas medidas de estas consultas no aparecen para todos los años; concretamente todas las medidas relacionadas con la intensidad y dirección del viento solo comienzan a estar registradas a partir del año 2013.

5. Análisis y depuración de los datos

Antes de comenzar a manipular los datos de los que disponemos, resulta necesario asegurarse de que estos siguen unas pautas comunes y que son adecuados para el trabajo que vamos a realizar. Por ejemplo, debemos asegurarnos de que la notación numérica empleada en los datos de origen no crea problemas a la hora de cargar dichos datos en otras herramientas que vayamos a utilizar a lo largo del trabajo. Para ello vamos a analizar los ficheros Excel que nos han sido cedidos con el fin de encontrar incompatibilidades y faltas de normalización, y corregirlos con las herramientas de las que disponemos.

Un posible método de detectar incongruencias en los datos es cargar el fichero Excel en *RStudio* y aplicar el comando *factor()* a cada una de las columnas que queramos analizar. Este comando nos devolverá cada uno de los diferentes valores que aparecen en esa columna. Si observamos estos valores es posible que encontrar, por ejemplo, que una misma información se ha representado de dos formas diferentes en dos filas distintas del fichero, lo cual puede dificultar el análisis estadístico del mismo.

5.1. Unión de tablas

Cuando carguemos nuestras tablas de datos, tanto aquellas en formato Excel como aquellas en CSV, en el entorno de desarrollo de *RStudio* estas pasarán a quedar representadas mediante objetos denominados *dataframes*. Este formato facilita la manipulación de las tablas y posee funciones propias incluidas en R que nos facilitan el acceso a la información que contienen. Una de estas funciones nos ofrece la posibilidad de unir dos *dataframes* diferentes en función de los valores de una columna específica de uno de ellos. Por desgracia, dado que algunos de los ficheros con la información del clima con los que trabajamos poseen más columnas que otros que presentan los datos del mismo año, no nos es posible utilizar esa función. Para solucionar este problema se ha descargado el paquete *plyr*, el cual incluye la función *rbind* que nos permite, mediante el parámetro *fill*, especificar que aquellas columnas que no estén presentes en todas los *dataframes* de entrada sean rellenadas con valores nulos en caso de no poseer un valor.

Una vez hemos conseguido un método para la unión de tablas que nos garantiza que no perderemos ningún tipo de información al hacerlo podemos pasar a decidir de qué modo vamos a relacionar nuestros *dataframes*. Dado que poseemos la información de los incidentes de tráfico separada según el año, podemos unir en una tabla los datos del clima correspondientes a todos los meses de un mismo año, con lo que obtenemos una tabla con las características meteorológicas de cada día de dicho año, y a continuación podremos mezclar la información de dicha tabla con la de los incidentes del año correspondiente en función de la columna “fecha”; es decir, una vez obtenida la tabla con el clima de cada día del año, podemos asociar a cada fila de la tabla de incidentes



la información del clima del día en que se produjo simplemente mezclando la información de ambas tablas. Esto se ha resuelto con la función *merge* incluida en R, para la cual se especifican los *dataframes* de entrada y la columna en base a cuyos valores se va a realizar la fusión de tablas.

5.2. Formato de fechas

El primer problema de formato que encontramos es la forma en la que se ha representado en las tablas de Excel la fecha de cada uno de los accidentes registrados. Dado que tenemos un fichero con los accidentes de todo un año, el procedimiento que emplean para representar esta información es el de añadir dos columnas (DIA y MES), una para representar el día y otra para representar el mes. Este tipo de formato no resulta útil, pues en cada ocasión que se quiera obtener la fecha de un accidente será necesario acceder a dos columnas diferentes de la tabla y combinar ambas junto con el año, que solamente aparece en el nombre del archivo. Para solucionarlo, la opción más simple y efectiva es la de añadir una nueva columna al fichero, en la que representaremos las fechas correspondientes en formato “dd/mm/aa”. Disponemos de dos métodos distintos para conseguirlo: cargar cada una de las tablas a R Studio y mediante un pequeño script que cree una nueva columna combinando la información de las columnas “mes” y “día”, o añadir la nueva columna directamente en el archivo Excel y, empleando la función FECHA() que ya viene definida en Excel por defecto, completar esta columna a partir de la información de las celdas con el mes y el día de cada fila (Dado que se va a modificar el fichero original siempre es recomendable mantener una copia del original en caso de que surja algún problema). En nuestro caso vamos a emplear la segunda opción en base a su simplicidad. El uso de la función FECHA() requiere 3 argumentos (año, mes, día) y automáticamente nos devuelve la fecha en el formato adecuado.

5.3. Formato de decimales

La siguiente posible fuente de problemas que podemos detectar en los datos iniciales procede del formato empleado para representar los números decimales, por ejemplo en las columnas correspondientes a los resultados de las pruebas de alcoholemia. Podemos encontrar estos datos hasta con tres formatos diferentes dentro de un mismo archivo: unas veces la coma decimal viene representado por el símbolo “.”, otras por el símbolo “,” y en unas pocas ocasiones por el símbolo “’”’. De nuevo, para solucionar este problema podemos emplear una función de sustitución de caracteres en R o simplemente emplear la funcionalidad Reemplazar de Excel que sustituye una tira de caracteres por otra o por el vacío. En nuestro caso usaremos otra vez la solución más sencilla, para ello en cada una de las tablas Excel seleccionamos las columnas correspondientes a datos decimales y reemplazamos solo en estas los puntos y las comas flotantes por puntos. Hemos escogido la coma (“,”) como separador decimal ya que al seleccionar el formato de las celdas correspondientes a los números decimales, el modo que Excel propone para representar estos números es empleando “,”.

Las columnas a las que afectan estos cambios son: "HORA", "1ª PRUEBA COND-1", "2ª PRUEBA COND-1", "1ª PRUEBA COND-2", "2ª PRUEBA COND-2".

5.4. Errores de clasificación

En varios apartados de cada registro se clasifica el incidente dentro de una categoría, como en el caso de la columna "TIPO DE ACCIDENTE", empleando para ello una cadena de caracteres. En caso de que los posibles valores para dichas categorías no estén normalizados, por ejemplo creando una lista de opciones en la que seleccionar el valor al registrar el incidente, y deban ser introducidos a mano puede conducir a errores que clasifiquen dos incidentes del mismo tipo con diferentes denominaciones o pueden producirse directamente errores ortográficos. Algunos ejemplos de estos errores serían las entradas de "SALIDA DE VIA" y "SALIDA DE LA VIA" que crearían dos categorías diferentes a la hora de clasificar al conjunto de todos los incidentes o el uso de mayúsculas y minúsculas en "alcance" y "ALCANCE".

Otro problema relacionado con este apartado es el uso de caracteres no alfanuméricos en las clasificaciones, por ejemplo emplear el carácter "?" para los valores desconocidos. Esta práctica puede acarrear problemas a la hora de estudiar los datos si no se asegura que el software empleado para realizar el estudio de los datos acepte tales caracteres.

6. Estudios y resultados

En esta sección se detallan todos los aspectos que han sido estudiados a partir de los *dataframes* obtenidos como resultado de las acciones descritas en el apartado anterior, justificando el porqué de esos aspectos, como se ha realizado su estudio y por último representando (en forma de tabla, gráfico o mapa) los resultados que se han obtenido. Todas las gráficas y mapas que se obtienen como resultado en este apartado pueden encontrarse en el anexo al final de este trabajo a mayor tamaño para facilitar su observación. Se indicará al pie de cada imagen que sea necesario en que apartado del anexo puede encontrarse. De igual modo, en la sección 9.7 del Anexo pueden encontrarse ejemplos de código similares a los empleados para la construcción de las gráficas y mapas.

6.1. Incidentes en base a periodos temporales

Conocer los periodos con mayor concentración de accidentes y detenciones puede ayudar a planificar de antemano la distribución del personal Policial y sanitario con el fin de reducir los tiempos de respuesta en caso de accidentes graves, reduciendo las consecuencias de un choque, en incluso llegando a salvar la vida de los implicados.

6.1.1. Incidentes por mes

mes\añ	2010	2011	2012	2013	2014	2015	2016	2017
Enero	810	775	807	810	855	830	820	883
Febrero	828	745	958	828	885	848	872	855
Marzo	889	774	1031	868	885	868	873	996
Abril	792	685	814	898	734	751	857	846
Mayo	846	860	914	869	895	960	914	921
Junio	764	816	880	878	902	925	988	903
Julio	787	753	731	787	808	848	892	836
Agosto	651	553	582	521	621	561	659	628
Septiembre	778	795	791	832	878	811	889	827
Octubre	814	827	807	891	906	869	920	915
Noviembre	868	753	858	876	947	880	955	900
Diciembre	877	943	845	969	932	927	910	892
TOTAL	9704	9279	10018	10027	10248	10078	10549	10402

Figura 3: Tabla de total de incidentes por mes y año

Un primer estudio interesante sería analizar en qué periodos del año es necesaria una mayor o menor intervención Policial a causa de los incidentes de tráfico, lo cual nos permitiría ampliar o reducir el tamaño del personal de servicio en previsión de un periodo de elevada actividad, o por el contrario identificar aquellos periodos cuando es menos probable que se produzca un accidente, lo que permitiría reducir el gasto en recursos que no van a ser necesarios. La información del número de incidentes por mes puede resumirse en la forma de la tabla de la Figura 2, en la que podemos observar que, bien si no de forma gradual, la cantidad total de incidentes registrados por año ha ido en aumento durante los años de los que disponemos datos. Por otro lado, para la búsqueda de patrones y de observaciones más concretas, este modo de

representación no ofrece mucha ayuda, por lo que hemos optado por ampliar esta información mediante diferentes gráficas.

Para conseguir esto, el primer paso consiste en extraer de los *dataframe* de cada año la información de la columna MES de cada registro, de esta forma obtenemos un vector con una longitud igual a la cantidad de registros del año correspondiente donde cada posición es el número de mes en el que se produjo el registro. A continuación aplicamos la función *table()*, la cual se encuentra entre las funciones básicas de R, al vector que acabamos de extraer lo que nos devuelve una tabla con la cantidad de veces que se repite cada valor del vector, en este caso enteros del 1 al 12. Para representar los datos con el método que utilizamos se requiere que estos se encuentren en formato *dataframe*, por lo que simplemente es necesario aplicar la función básica de R *as.data.frame()* a nuestra tabla y añadir una columna que indique el año correspondiente a cada medida, con lo que obtenemos un *dataframe* como el de la Figura 4.

Var1	Freq	year
1	883	17
2	855	17
3	996	17
4	846	17
5	921	17
6	903	17
7	836	17
8	628	17
9	827	17
10	915	17
11	900	17
12	892	17

Figura 4: Contenido del dataframe con los datos por mes de 2017

Si realizamos el mismo proceso para los datos de cada año y fusionamos los *dataframes* obtenidos ya tenemos los datos en el formato apropiado para representarlos en forma de gráfica. Para ello hemos escogido el paquete *ggplot2* para la representación de los datos en base a su capacidad para personalizar los resultados, lo cual ayuda a que las gráficas resultantes sean más detalladas y al mismo tiempo más fáciles de comprender. A continuación pueden observarse los resultados que hemos obtenido:

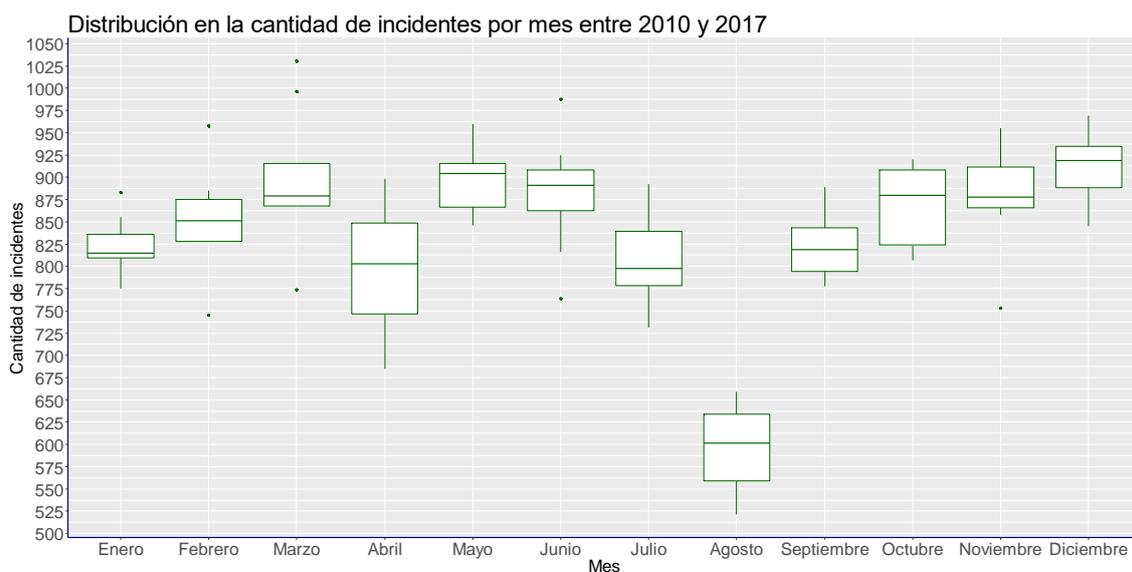


Figura 5: Distribución en la cantidad de incidentes por mes (Anexo 9.2)



Para esta primera representación hemos escogido un gráfico de tipo Box-Plot, el cual nos permite observar la cantidad de incidentes al mismo tiempo que podemos ver la variación de este valor a lo largo de los años. Este tipo de gráficos representa con una caja los dos cuartiles de datos que rodean a la mediana, es decir el 50% de los datos que se encuentran más cerca de la mediana. Los dos cuartiles restantes se representan mediante las líneas que se extiende verticalmente desde la caja, siendo su extensión los valores máximos y mínimos según su posición con respecto a la caja. Los puntos que aparecen más allá de las dos líneas que se extienden desde la caja son valores considerados automáticamente como demasiado extremos para incluirlos en la representación.

En esta gráfica podemos observar que los meses con más cantidad de accidentes se corresponderían con Marzo, Mayo y Diciembre, sin embargo, de entre estos Marzo presenta varios valores extremos que indican que la cantidad de incidentes en este mes tiende a variar entre un año y otro, por lo que no sería correcto afirmar con seguridad que en un año futuro se producirá un pico de incidentes en este mes. Para Mayo y Diciembre sin embargo, sí que podríamos prever que se dará esta situación ya que los datos de los que disponemos nos indican que adicionalmente a ser valores altos, la cantidad de incidentes en estos meses es estable a lo largo de todos los años ya que el tamaño de las cajas no es excesivamente grande y no presentan valores extremos. Entre los valores mínimos encontramos que es claramente durante el mes de Agosto donde se produce la menor cantidad de incidentes cada año y que estas cifras suelen ser estables, entre 525 y 625 incidentes durante los años que estamos estudiando. Esta bajada en la cantidad de incidentes tan significativa con respecto al resto de meses puede tener su explicación en el hecho de que es durante el mes de Agosto cuando la mayor parte de la población de la ciudad se encuentra de vacaciones, lo que supone un menor tráfico a causa de las personas que ya no necesitan desplazarse a su lugar de trabajo.

Si bien un gráfico Box-Plot nos permite observar los valores en torno a los cuales se han registrado incidentes cada año, no nos permite observar la evolución durante dichos años ya que no tenemos forma de identificar un valor en un mes concreto para un año específico. Para ello hemos elaborado una segunda gráfica en la que presentamos las cifras totales de incidentes para cada mes y año, y en la que nos centraremos en analizar la evolución en los meses que hemos destacado durante el estudio del gráfico Box-Plot por ser los más relevantes.

En esta nueva gráfica (Figura 6) podemos observar que para los meses de Mayo y Agosto se ha ido produciendo un aumento de los incidentes conforme a cada año, si bien en el año 2017 se registraron menos incidentes en Mayo que durante 2016 y 2015.

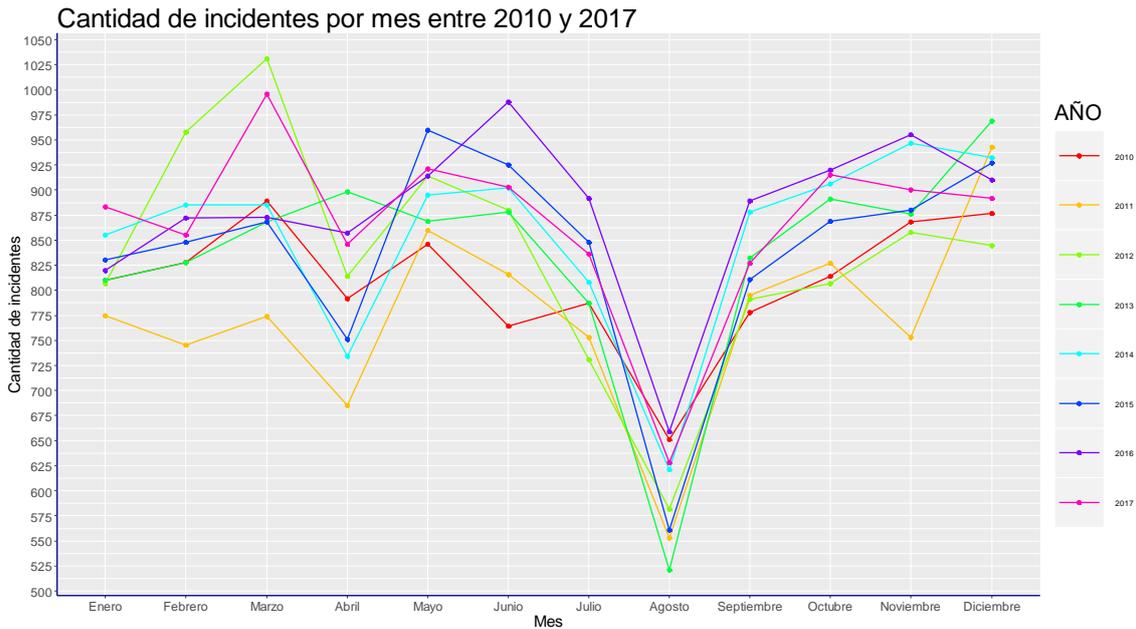


Figura 6: Cantidad de incidentes por mes entre 2010 y 2017

Por otra parte Diciembre presenta un comportamiento diferente ya que la cantidad de incidentes parece tender a estabilizarse en torno a un valor central en los últimos 3 años.

6.1.2. Incidentes por hora

Como ya hemos visto en el apartado anterior los periodos de trabajo y vacaciones de los ciudadanos parecen influir en gran medida en el tráfico y por extensión, en el número de accidentes e infracciones. Es por ello, que al momento de distribuir y representar los datos de los que disponemos en función de la hora se va a hacer una distinción entre los días laborales y los fines de semana ya que, como se puede observar en las gráficas obtenidas, se obtienen dos comportamientos claramente diferentes.

En primer lugar procedemos a extraer los datos correspondientes a la columna HORA de cada registro, pero debemos diferenciar entre aquellos datos que pertenezcan a fines de semana y aquellos que no. Para ello al extraer los datos podemos especificar una condición, la cual será que el valor de la columna "DIA DE LA SEMANA" sea igual a S o D para el caso de los valores de los fines de semana y al contrario para los días entre semana. Sin embargo los datos de los años anteriores a 2015 no poseen la columna "DIA DE LA SEMANA", para solucionarlo en su lugar hemos aplicado la función weekday() de R (la cual nos devuelve el día de la semana a partir de una fecha en formato dd/mm/aaaa) a cada valor de la columna "Fecha" que hemos añadido durante el tratamiento inicial de los datos, y hemos comprobado si esa fecha correspondía al día que estábamos buscando. Los comandos concretos para extraer los datos de esta forma son los siguientes:

```
weekend17 <- year2017$HORA[year2017$DIA.DE.LA.SEMANA == "S" |
year2017$DIA.DE.LA.SEMANA == "D"]
```



```
weekend14 <- year2014$HORA[weekdays(year2014$Fecha) == "sábado" |
weekdays(year2014$Fecha) == "domingo"]
```

A diferencia del apartado anterior, esta vez no vamos a emplear la función *table()* pues ahora tratamos de clasificar valores continuos y no discretos. En su lugar utilizaremos la función *cut2()* del paquete *Hmisc*, la cual nos va a permitir clasificar cada valor de un vector de valores numéricos según entre que 2 valores se encuentren de una secuencia dada. En nuestro caso, vamos a distribuir los valores de la columna *HORA* en intervalos de una hora, por lo que ejecutamos la función *cut2()* con la secuencia 1...24. Como resultado obtendremos una tabla con dos columnas: el intervalo y el número de valores del vector de entrada que se encuentran en ese intervalo. A continuación se procede de igual forma que en el apartado anterior, transformando la tabla en *dataframe* y añadiendo una nueva columna con el valor del año correspondiente. Si realizamos el mismo tratamiento para los datos de fines de semana y entre semana estas son las gráficas que se obtienen:

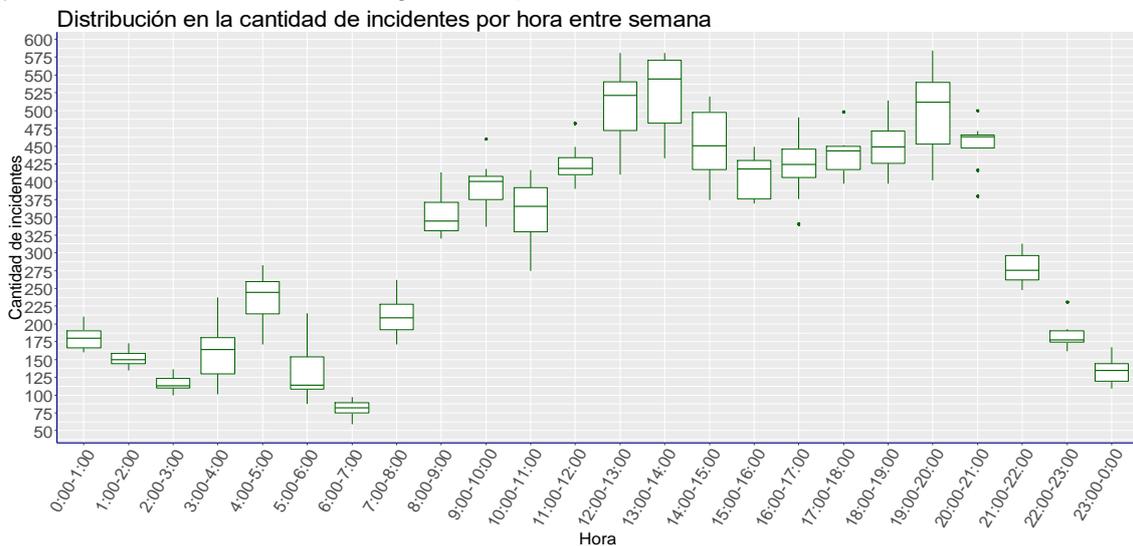


Figura 7: Distribución de la cantidad de incidentes por hora entre semana (Anexo 9.3)

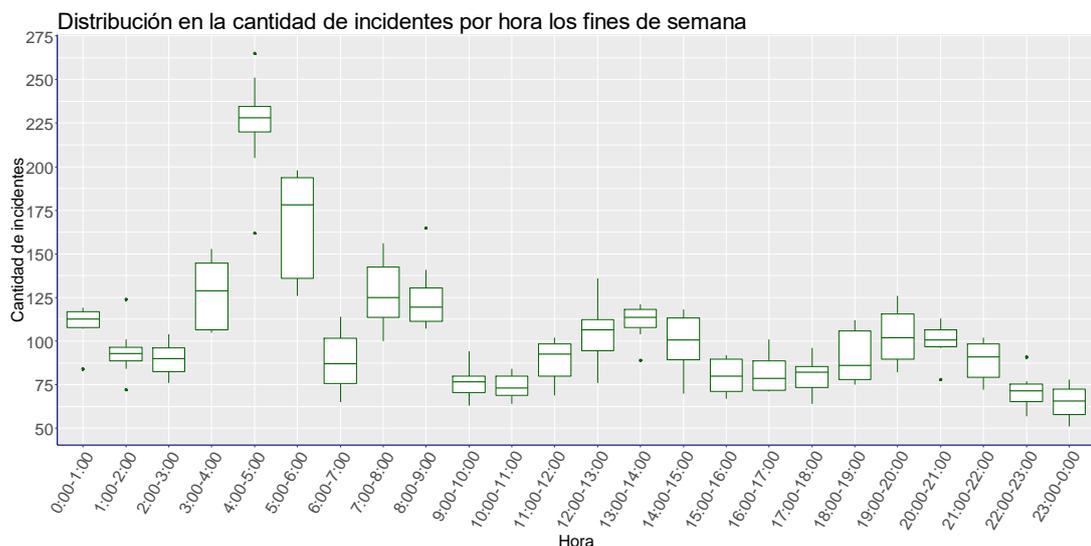


Figura 8: Distribución de la cantidad de incidentes por hora los fines de semana (Anexo 9.3)

Empezaremos comentando la gráfica de los datos recogidos entre semana. Lo primero que cabe destacar es que existen dos franjas de tiempo en las que la cantidad de incidentes difiere en gran medida. En primer lugar tenemos que entre las 22:00 horas y las 7:00 la cantidad de incidentes es mucho menor de media en comparación con la franja entre las 7:00 y las 22:00. A pesar de ello, en la primera franja es posible distinguir un pico en el número de incidentes entre las 4:00 y las 5:00 del que hablaremos en detalle más adelante. En la segunda franja encontramos 2 picos, el primero entre las 12:00 y las 14:00; y el segundo entre las 19:00 y las 20:00. Estos 2 picos coinciden como ya se ha comentado anteriormente con los horarios habituales de salida de la jornada laboral. A pesar de sus altos valores, la cantidad de incidentes de estos picos posee una gran variación, como puede observarse por el tamaño de las cajas y la longitud de las líneas que la representan; lo cual podría significar que no se garantiza que en un año dado el mayor valor del número de incidentes se haya dado en una de estas franjas. Para conocer con mayor precisión esta distribución, y siguiendo el método del apartado anterior, se ha construido también una gráfica en la que se tienen en cuenta los valores en función del año.

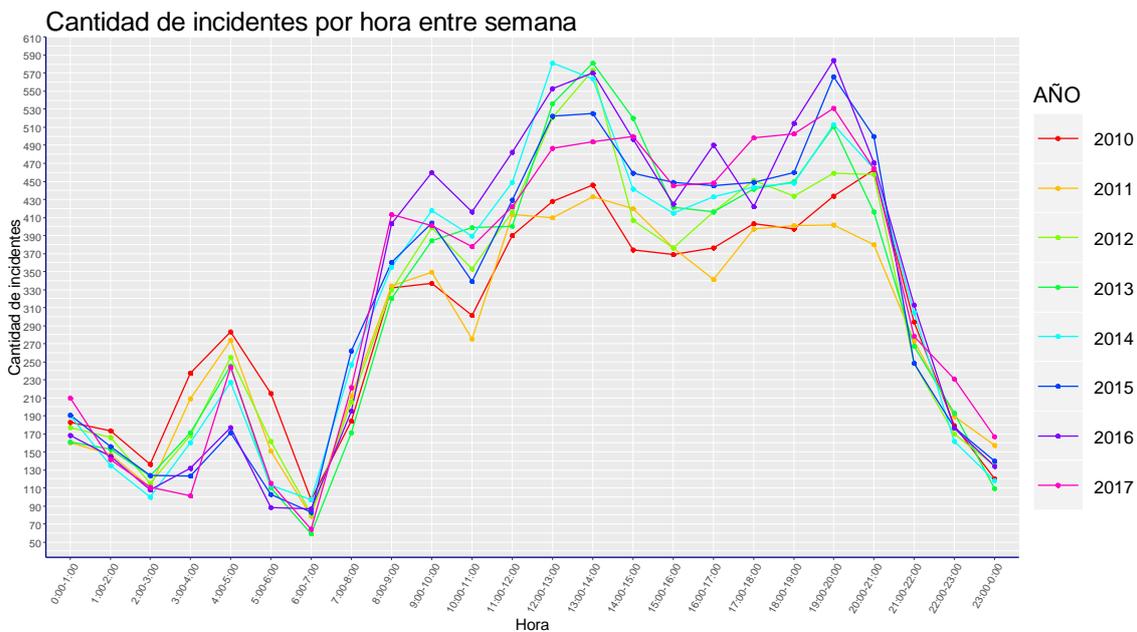


Figura 9: Cantidad de incidentes por hora entre semana entre 2010 y 2017 (Anexo 9.3)

En esta nueva gráfica (Figura 9) observamos que si comparamos los valores en las franjas de 12:00 a 14:00 y de 19:00 a 20:00 con el resto de franjas de un mismo año, estas sí que representan puntos máximos en la gráfica y que la variación de valores se produce entre años. Entre las 19:00 y las 20:00 podemos destacar también que la cantidad de incidentes registrados parece ascender con los años; con la excepción de 2010 y 2017, cada año el número de incidentes ha aumentado hasta alcanzar su máximo en 2016, con una diferencia de casi 200 incidentes con respecto a 2011. Para el periodo entre 12:00 y 14:00 no se observa un comportamiento tan específico, pero sí que de nuevo en el año 2017 ha descendido el número de incidentes con respecto a los 2 años anteriores. Por último, la franja entre 4:00 y 5:00 presenta una distribución diametralmente opuesta a la primera que hemos analizado, puesto que el número de



incidentes ha disminuido cada año, con la excepción del último año del que se disponen datos, en el que se han superado los valores de los 3 años anteriores.

En la gráfica de los fines de semana (Figura 8) encontramos otro patrón al compararlo con los datos de entre semana, pues es entre las 00:00 y las 9:00 donde se producen los picos de mayor registro de incidencias mientras que entre las 9:00 y las 00:00 estos se registros se mantienen más estables y a un menor nivel. De nuevo y al igual que en la anterior gráfica, observamos un máximo de la gráfica en el periodo 4:00-5:00 que además alcanza aproximadamente un valor igual en tan solo 2 días que el valor que se alcanza en el mismo periodo en los días laborables. No se conocen las causas para este comportamiento por lo que será necesario realizar más estudios en torno a él. Si incluimos una representación que tenga en cuenta los valores por año (Figura 10) podemos ver como para el periodo de actividad más estable, entre las 9:00 y las 0:00, es apreciable como las valores de los 3 últimos años suelen dominar como las mayores en casi todas las franjas; mientras que en el periodo entre las 0:00 y las 9:00 no hay un claro patrón discernible. Sí que se repite sin embargo el mismo comportamiento que en los datos de entre semana, con un descenso en los valores para los 3 últimos años en la franja entre 4:00 y 5:00.

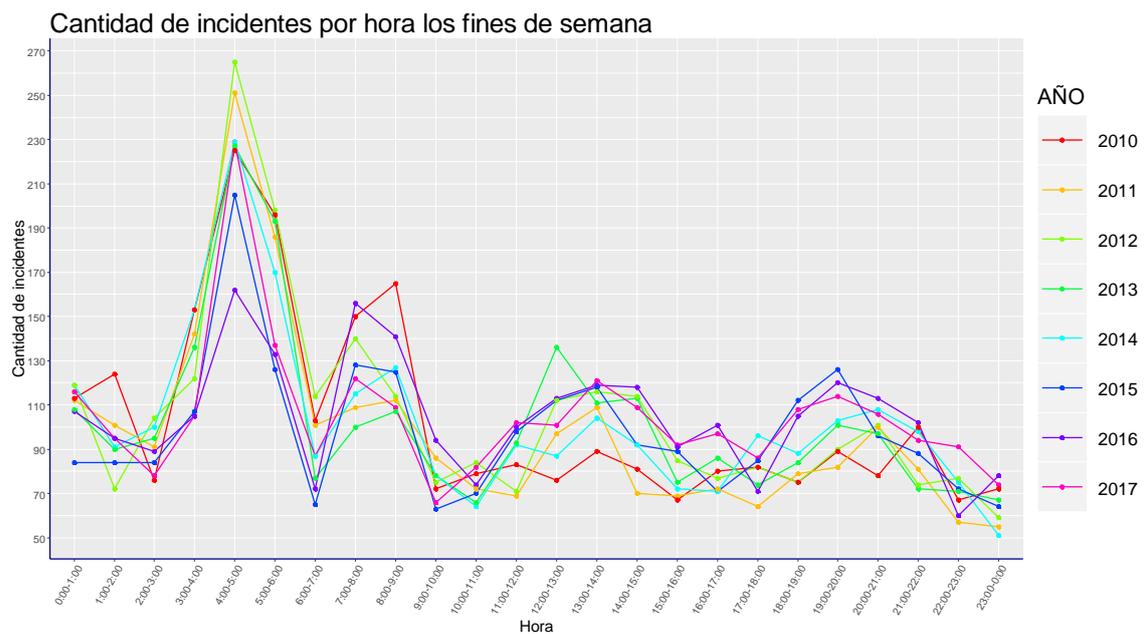


Figura 10: Cantidad de incidentes por hora los fines de semana entre 2010 y 2017 (Anexo 9.3)

6.2. Incidentes en base a las condiciones climáticas

Es innegable que ciertos fenómenos meteorológicos afectan a la capacidad de manejar un vehículo para muchos conductores, ya sea disminuyendo la visibilidad o alterando las condiciones físicas de la calzada. Es por esto que resultaría interesante averiguar si existe una relación entre estos cambios en el comportamiento de los conductores y la cantidad de incidentes que se registran durante estos fenómenos meteorológicos, ya que de este modo el personal responsable de responder a estos incidentes podría estar mejor preparado tan solo con obtener una predicción del tiempo o, por el contrario, ignorar aquellos eventos en los que no sea realmente necesario elevar el nivel de atención.

6.2.1. Nº de incidentes bajo precipitaciones fuertes

Las precipitaciones afectan en dos sentidos a la habilidad para manejar un vehículo; por un lado, si la intensidad de la precipitación es lo bastante elevada, puede reducir en gran medida la visibilidad de los conductores; por el otro, un tramo de calzada húmeda facilita que los vehículos puedan perder el control al realizar alguna maniobra brusca y aumenta la distancia necesaria para el frenado a aproximadamente el doble [3].

Año	Nº de días de precipitación fuerte	Nº de incidentes con precipitación fuerte	Nº de días sin precipitación fuerte	Nº de incidentes sin precipitación fuerte	Media de incidentes con precipitación por día	Media de incidentes habitual por día
2010	40	1160	325	8544	29,00	26,29
2011	35	876	330	8403	25,03	25,46
2012	17	517	349	9502	30,41	27,23
2013	22	679	343	9349	30,86	27,27
2014	27	836	338	9412	30,96	27,85
2015	35	889	330	9189	25,40	27,85
2016	37	1000	329	9549	27,03	29,03
2017	25	722	340	9680	28,88	28,47

Figura 11: Tabla de medias de incidentes en base a precipitaciones

Para averiguar si las precipitaciones afectan al tráfico y a los accidentes vamos a proceder a calcular las medias del número de registros que se producen los días con precipitaciones fuertes y los días que no para cada año. El primer paso será simplemente realizar un conteo del número de registros cuya columna "prec" en nuestro *dataframe* supere el valor de mm/h a partir del cual la precipitación se podría considerar de intensidad fuerte, siendo este valor de 15 mm/h según la Agencia Estatal de Meteorología (AEMET) [4]. Al mismo tiempo también será necesario realizar un conteo de los días en los que se produjeron estas precipitaciones, el cual podemos realizar mediante la misma consulta pero sobre las tablas del clima que hemos obtenido directamente de AEMET Open Data. Con esto y sabiendo el número total de incidentes por año y la cantidad total de días por año, ya podemos calcular las medias necesarias, representadas en forma de tabla en la Figura 11.

Podemos representar gráficamente estos resultados mediante un diagrama Box-Plot o diagrama de caja. Este tipo de diagramas muestran de forma simple la mediana y los cuartiles de un conjunto de datos, con lo que nos permitirá comparar al mismo tiempo los valores y la variación de estos entre los dos conjuntos de medias, las de los días de lluvia y las de los días habituales.

Comparación de la media de incidentes en función de la lluvia

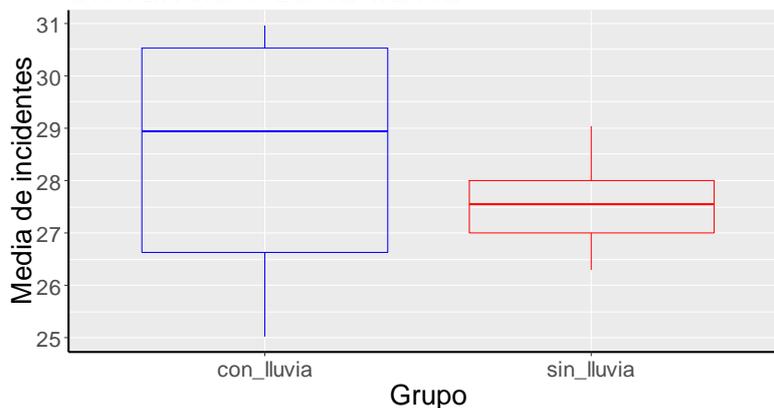


Figura 12: Box-Plot de medias con y sin precipitación

Como podemos observar en la Figura 12, la varianza del número de incidentes en los días de lluvia parece mayor que la del resto de días, pero al mismo tiempo ambos grupos de medias se encuentran en torno a valores muy cercanos. Por desgracia, simplemente representar gráficamente el resultado no nos permite determinar si existe o no una relación entre el factor que estamos estudiando y los datos resultantes, para ello vamos a necesitar otra metodología.

Ya que no buscamos representar gráficamente nuestro resultado para analizar posibles patrones si no averiguar si existe una relación entre que se produzca una precipitación fuerte y el número de incidentes, emplearemos un test estadístico para determinarlo. El propósito de estos test es el de verificar o rechazar una hipótesis de un modelo estadístico, en nuestro caso, que el número de incidentes está relacionado con la intensidad de las precipitaciones. Escoger el test estadístico apropiado depende de los datos con los que lo vayamos a realizar pues existen numerosas variaciones; dado que en este trabajo vamos a proceder a comparar las medias en el número de incidentes para los días en los que se registró una lluvia fuerte y las medias en el número de incidentes del resto de días del año, existen dos test apropiados para verificar nuestra hipótesis: el test t de Student y el test de Welch. A su vez cada uno de estos test solo resulta adecuado bajo ciertas condiciones: la prueba t de Student solo funciona adecuadamente si la varianza de los dos conjuntos de datos proporcionados es similar, mientras que el test de Welch es apropiado para varianzas diferentes. Por lo tanto antes de aplicar alguna de estas dos pruebas sobre nuestros datos, debemos aplicar en primer lugar un test adicional que verifique si la varianza de nuestros dos conjuntos de datos es similar, y una vez conozcamos esto aplicar el test correspondiente.

Una vez calculadas las medias las almacenamos en dos vectores, uno que contenga las medias de los incidentes ocurridos durante días con precipitaciones fuertes y otro con las medias del resto de días, los cuales pasamos como parámetro a la función básica de R `var.test()`, la cual comprueba si ambos conjuntos poseen una varianza similar mediante un test estadístico. Tanto este test, como los test de Student y Welch asumen una hipótesis nula a partir de la cual realizan sus cálculos y ofrecen sus resultados. Uno de estos resultados es la significación estadística denotada por el valor p-value en los resultados del test. Una vez obtenido, este valor se compara con un nivel de significación establecido de antemano al realizar el test, si el p-value es inferior al nivel de significación, la hipótesis nula es rechazada y viceversa.

La prueba que realiza la función `var.test()` asume como hipótesis nula que las varianzas de los datos proporcionados son similares y otorga un valor de 0.05 al nivel de significación. Al introducir las medias que hemos calculado obtenemos el resultado de la Figura 13.

```
> var.test(mediacon11$media, mediasin11$media)

      F test to compare two variances

data:  mediacon11$media and mediasin11$media
F = 4.2695, num df = 7, denom df = 7, p-value = 0.07462
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.854773 21.325833
sample estimates:
ratio of variances
 4.269514
```

Figura 13: Resultado del test de varianzas similares para las medias con y sin precipitación.

Como podemos observar en el resultado, el p-value devuelto es 0.07462, el cual es superior al nivel de significación 0.05 por lo que se acepta la hipótesis nula y se verifica que las varianzas de los dos vectores de medias son similares. De esta manera averiguamos que para verificar si existe una relación entre el número de incidentes y las precipitaciones debemos aplicar el test de Student, el cual se realiza de forma muy similar a la prueba de varianzas mediante la función `t.test()` con el parámetro `var.equal` con valor TRUE, también incluida en las funciones básicas de R.

```
> t.test(mediasin11, mediacon11, var.equal = TRUE)

      Two Sample t-test

data:  mediasin11 and mediacon11
t = -1.0885, df = 14, p-value = 0.2948
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-3.0149669 0.9849669
sample estimates:
mean of x mean of y
 27.43125 28.44625
```

Figura 14: Test t de Student sobre los datos de precipitaciones



La prueba de Student realizada por la función $t.test()$ asume como nivel de significación 0.05 y como hipótesis nula que no existe diferencia entre las medias de los dos conjuntos de datos proporcionados y por lo tanto no existe ningún cambio provocado por el factor que se esté estudiando, la lluvia en nuestro caso, pues ambos conjuntos proporcionan resultados estadísticamente iguales. Como podemos observar en la Figura 14 el p-value resultante es 0.2948, superior al nivel de significación establecido por lo que debemos aceptar la hipótesis nula y deducir así que no existe una relación entre las precipitaciones de intensidad fuerte y el número de incidencias de tráfico en la ciudad de Valencia.

6.2.2. Nº de incidentes durante temperaturas máximas extremas

Al igual que las precipitaciones, las altas temperaturas pueden suponer un riesgo para los conductores, pues pueden provocar golpes de calor, que si ya de por sí pueden resultar peligrosos, el daño que pueden causar se multiplica si la persona que lo sufre se encuentra al control de un vehículo. Es por esto que se va realizar un estudio similar al del apartado anterior aunque en esta ocasión calcularemos las medias de incidentes de aquellos días donde se registraron unas temperaturas máximas peligrosas y las compararemos con la media del resto de días. Para decidir la temperatura umbral a partir de la cual se considera que existe una situación de peligro se van a emplear los umbrales marcados por el Ministerio de Sanidad, Consumo y Bienestar Social, los cuales marcan unas temperaturas umbrales para cada región de España a partir de las cuales se considera que existe un riesgo para la salud, estando marcado en 34°C el umbral para la ciudad de Valencia.

Año	Nº de días de temperaturas altas	Nº de incidentes con altas temperaturas	Nº de días sin altas temperaturas	Nº de incidentes sin altas temperaturas	Media de incidentes con altas temperaturas por día	Media de incidentes sin altas temperaturas por día
2010	4	93	361	9611	23,25	26,62
2011	2	38	363	9238	19,00	25,45
2012	4	82	362	9937	20,50	27,45
2013	2	46	363	9982	23,00	27,50
2014	4	99	361	10149	24,75	28,11
2015	6	167	359	9911	27,83	27,61
2016	2	50	364	10500	25,00	28,85
2017	6	168	359	10234	28,00	28,51

Figura 15: Tabla de medias para los días con y sin temperaturas altas.

Conociendo la temperatura umbral de peligro procedemos de igual forma al apartado anterior, calculando los días y registros por año con y sin temperaturas altas para calcular las medias necesarias y una vez obtenidas, pasamos a representarlas mediante un diagrama Box-Plot:

Comparación de la media de incidentes en función de las temperaturas máximas

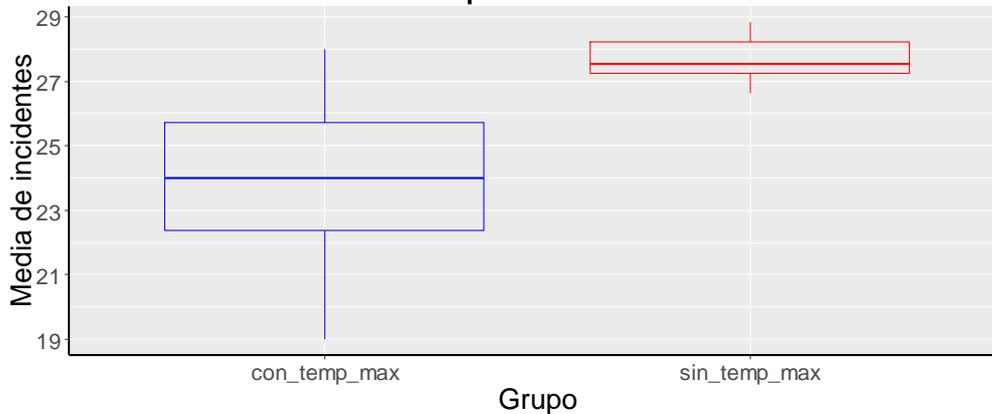


Figura 16: Comparación de medias con y sin altas temperaturas

Tal y como podemos ver en el diagrama Box-Plot esta vez la varianza y la mediana de ambos grupos de medias no es tan similar como lo es para los resultados que tienen en cuenta el volumen de precipitaciones. A simple vista podría parecer que los días con altas temperaturas suceden, de media, menos registros, sin embargo no es posible garantizar la diferencia de varianzas por lo que sigue siendo necesaria la aplicación de un test estadístico. Tal y como se ha procedido en el apartado anterior el primer paso es averiguar si se puede considerar que ambos grupos de medias poseen una varianza similar, aplicando para ello la función `var.test()` con el mismo nivel de significación y la misma hipótesis nula que en el caso anterior.

```
> var.test(mediacontemp$media, mediasintemp$media)

F test to compare two variances

data: mediacontemp$media and mediasintemp$media
F = 8.6518, num df = 7, denom df = 7, p-value = 0.01082
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.732128 43.215078
sample estimates:
ratio of variances
 8.651825
```

Figura 17: Resultado del test de varianzas similares para las medias con y sin altas temperaturas.

En este caso el valor de p-value que hemos obtenido es 0.01082, el cual es menor que el nivel de significación por lo que rechazamos la hipótesis nula y concluimos que las varianzas de los dos grupos de datos son diferentes por lo que habrá que aplicar el test de Welch y no el de Student para determinar si existe una relación entre ambos. Para ello ejecutamos la función `t.test()` pero esta vez el parámetro `var.equal` tiene valor FALSE para especificar que el test que se pretende realizar es el de Welch.

```
> t.test(mediacontemp$media, mediasintemp$media)

      welch Two sample t-test

data:  mediacontemp$media and mediasintemp$media
t = -3.0241, df = 8.5968, p-value = 0.01515
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.3057236 -0.8867764
sample estimates:
mean of x mean of y
 23.91625  27.51250
```

Figura 18: test de Welch sobre los datos con y sin altas temperaturas

En este resultado, el valor obtenido por el test de Welch para el p-value es de 0.01515, de nuevo menor que el nivel de significación de 0.05 por lo que se rechaza la hipótesis nula y se deduce que en este caso si existe una relación entre las altas temperaturas y el número de incidentes de tráfico. Sin embargo, tal y como se puede ver en el diagrama de la Figura 16, la media de registros de los días con altas temperaturas es en conjunto menor que la del resto de días a pesar de que nuestra suposición inicial era la de que las altas temperaturas podrían afectar de manera negativa a los conductores provocando un aumento y no una disminución de los incidentes. Una posible explicación para este comportamiento inesperado de los datos se encuentra en uno de los resultados que también se ha obtenido en este trabajo. En el apartado 6.1.1, donde se estudian los incidentes en función de características temporales, se ha observado que durante los meses de Agosto de cada año descienden de forma sistemática el número de registros. Agosto es también uno de los meses en los que se registran las temperaturas más altas del año, por lo que es posible que la explicación para este supuesto descenso del número de incidentes aquellos días que se alcanzaron altas temperaturas no se halle en el clima si no en los patrones temporales.

6.2.3. Nº de incidentes durante rachas fuertes de viento

Como último estudio de la relación entre el clima y los incidentes de tráfico vamos a comprobar si las rachas de viento causan un aumento de los registros. Si bien la mayoría de los incidentes registrados en nuestros *dataframes* han tenido lugar en el interior de la ciudad de Valencia, donde no se espera que el viento suponga un riesgo potencial debido a que este es bloqueado por los edificios, en los registros también se tienen en cuenta los incidentes que han tenido lugar en las carreteras que rodean la ciudad, donde sí se espera la influencia del viento.

Para determinar cuándo se considera que una racha de viento posee la intensidad necesaria para que sea considerada peligrosa vamos a emplear de nuevo el manual de la Agencia Estatal de Meteorología[4], quienes consideran que una racha de viento es de fuerte intensidad cuando supera los 41 km/h. Por lo que debemos contabilizar el número de registros que se produjeron los días en los que el viento superó esta velocidad para realizar la media de incidentes. Por desgracia, apenas se han registrado vientos de esta magnitud entre los años 2010 y 2017 en la ciudad de Valencia, como máximo podemos encontrar un día o ninguno al año en la mayoría de los casos. Incluso si empleásemos la categoría de fuerza del viento de la AEMET anterior a esta,

que considera las rachas de viento entre 21 y 41 km/h, los datos que dispondríamos para trabajar seguirían sin ser suficientes para obtener unos resultados estadísticamente significativos, por lo que nos es imposible determinar la relación entre el viento y los incidentes de tráfico.

Año	Nº de días de viento fuerte	Nº de incidentes con viento	Media de incidentes viento por día	Media de incidentes habitual por día
2010				26,29
2011				25,46
2012	1	46	46,00	27,23
2013	1	24	24,00	27,27
2014	2	68	34,00	27,85
2015	1	39	39,00	27,85
2016	1	40	40,00	29,03
2017	2	67	33,50	28,47

Figura 19: Tabla de medias de incidentes en función de las fuertes rachas de viento

6.3. Incidentes en base a otras causas

Conocer la causa de los incidentes puede ayudarnos, por ejemplo en el caso de los delitos, a comprobar si las campañas de concienciación y prevención de ciertos tipos de comportamiento al volante están siendo efectivas, al mismo tiempo que nos aportan información sobre que nuevas campañas pueden ser necesarias. Por otra parte, conocer el tipo y la gravedad de los accidentes más comunes puede ayudar a los cuerpos de Policía a responder de forma más efectiva ante ellos.

6.3.1. Principales causas de incidentes

Para este primer apartado hemos recopilado la información de las columnas “MOTIVO QUE ORIGINA EL DELITO” y “TIPO DE ACCIDENTE” de nuestros *dataframes* que contienen todos los registros de cada año, y procedemos de forma similar al apartado 6.1.1. Para ello seleccionamos ambas columnas con la información que deseamos, obtenemos la tabla con el conteo de cada uno de los diferentes elementos que en ella aparecen, convertimos esta tabla a *dataframe*, añadimos las columnas necesarias para indicar el año y repetimos este proceso para los *dataframes* de cada año. Si unimos cada uno *dataframes* obtenidos para cada año en uno solo, ya estamos en disposición de representarlos gráficamente.

Analizar estas gráficas (Figuras 20 y 21) puede ayudarnos a averiguar cuáles fueron los delitos y los tipos de accidente más frecuentes en el periodo de 2010 a 2017, lo cual a su vez nos permite centrarnos en los aspectos realmente relevantes para este estudio de estos datos en la elaboración de otras gráficas más completas.

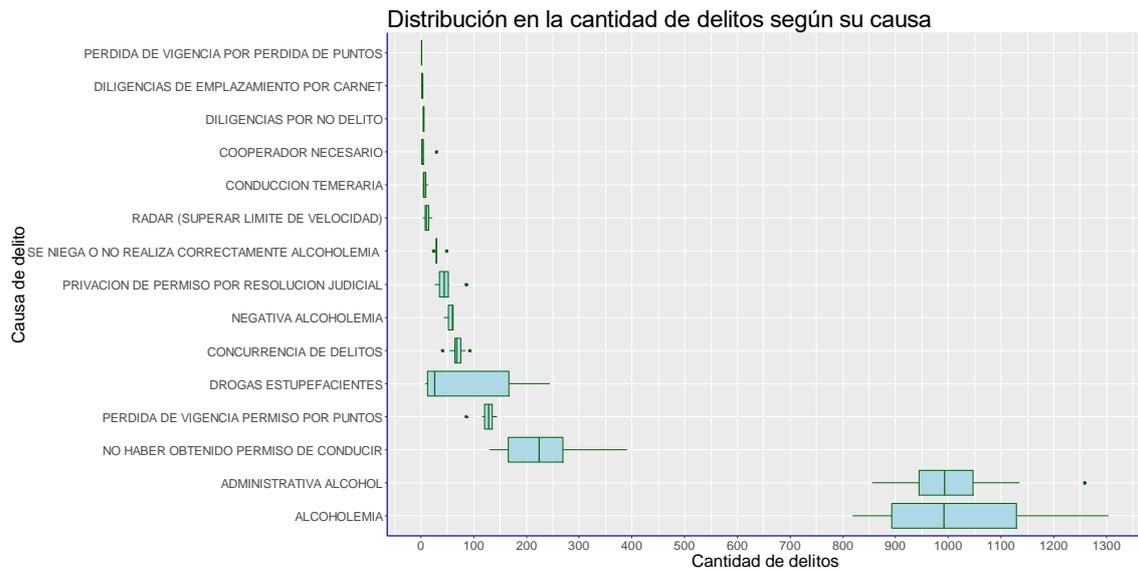


Figura 20: Distribución en la cantidad de accidentes según su tipo (Anexo 9.4)

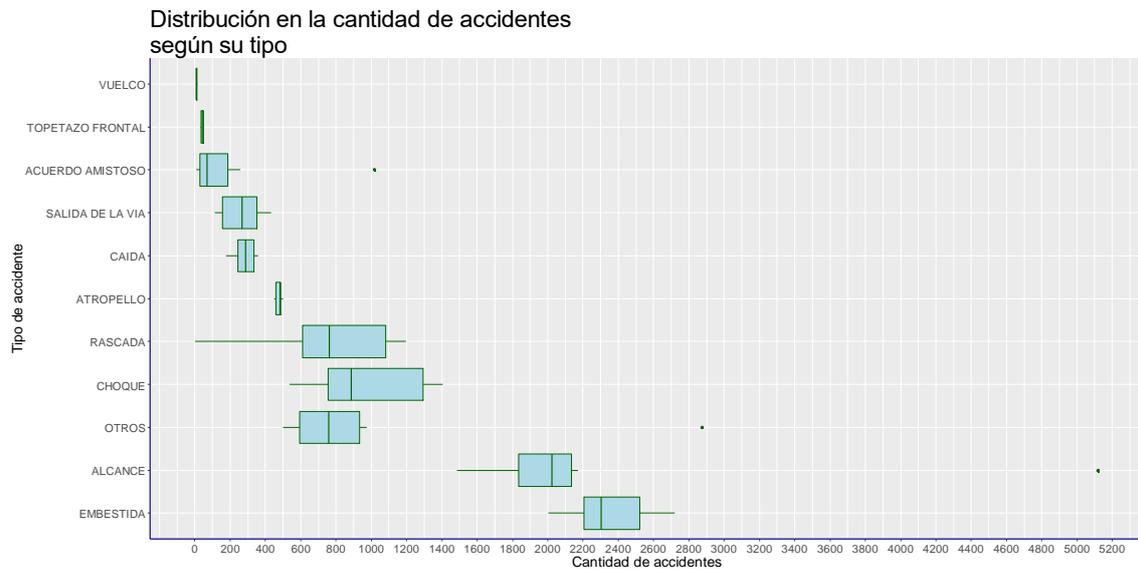


Figura 21: Distribución en la cantidad de accidentes según su tipo por año (Anexo 9.4)

La gráfica de delitos (Figura 20) nos indica que las principales causas de detención fueron “ALCOHOLEMIA”, “ADMINISTRATIVA ALCOHOL”, “NO HABER OBTENIDO PERMISO DE CONDUCCIÓN”, “PERDIDA DE VIGENCIA DEL PERMISO POR PUNTOS” y “DROGAS ESTUPEFACIENTES”; siendo de entre estos los más frecuentes con diferencia los dos primeros. También observamos que existe una considerable variación para los datos relacionados con el alcohol y las drogas, la cual se verá en detalle en la Figura 22.

En la gráfica de accidentes, por otro lado, podemos observar que los 4 tipos de accidente más frecuentes son “EMBESTIDA”, “ALCANCE”, “CHOQUE” y “ATROPELLO” si ignoramos la categoría “OTROS” que no aporta ninguna información. En esta segunda gráfica también detectamos la presencia de valores extraños que se alejan excesivamente de los valores habituales en algunas de las categorías representadas. Adicionalmente, si comparamos ambas gráficas entre sí podemos comprobar que la cantidad de delitos es de un orden menor a la de accidentes, siendo la mediana del delito más frecuente 1000 frente a 2300 del tipo de accidente más frecuente.

Si seleccionamos solo las infracciones y los tipos de accidentes que hemos encontrado como más frecuentes, y añadimos a nuestros datos los valores separados por año, obtendremos otro conjunto de gráficas en el que podemos apreciar la evolución y los cambios de patrones a lo largo del tiempo.

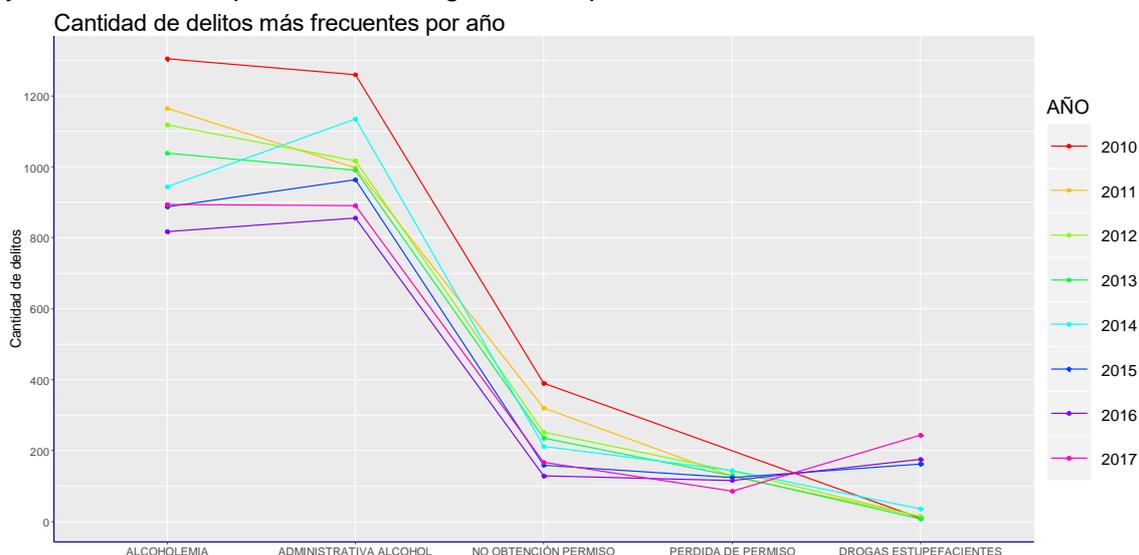


Figura 22: Cantidad de delitos más frecuentes por causa entre 2010 y 2017 (Anexo 9.4)

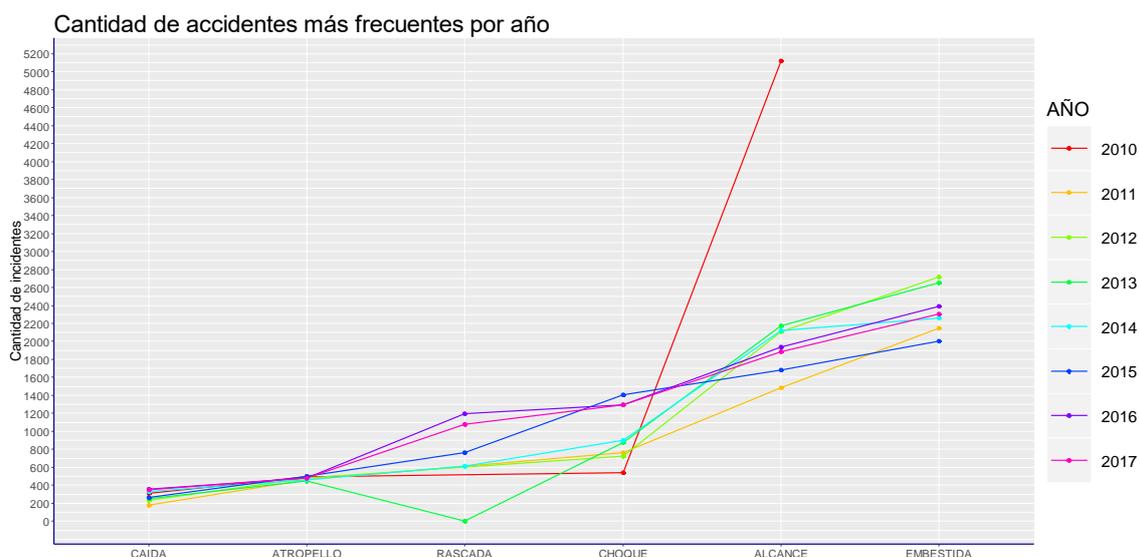


Figura 23: Cantidad de accidentes más frecuentes por tipo entre 2010 y 2017 (Anexo 9.4)



Si analizamos las gráficas que hemos obtenido es posible observar ciertos cambios en el comportamiento desde el 2010. Por ejemplo, en el caso de los delitos, se observa que la cantidad de delitos por superar la tasa de alcoholemia ha disminuido gradualmente cada año, aunque ha aumentado ligeramente en el 2017 con respecto al año anterior. Las sanciones administrativas por alcoholemia también han disminuido de media, si bien no de forma tan gradual con respecto a los 5 primeros años de los que se disponen datos. Por el contrario, los delitos relacionados con el consumo de drogas y estupefacientes han visto un aumento, principalmente entre los años 2015 y 2017. Por otra parte, otros dos delitos que cuyos números también han disminuido son la pérdida del permiso de conducir a causa de la pérdida de puntos en este y haber sido detenido sin haber obtenido el permiso de conducir.

En el caso de los accidentes lo primero que destaca es la gran cantidad de registros con la causa “ALCANCE” que se produjeron durante el año 2010 y que no existan registros de accidentes de “EMBESTIDA” en ese mismo año, probablemente debido a que en un principio no se hacía distinción entre ellos y se clasificaban de forma idéntica. Junto con ello, podemos observar que en accidentes del tipo “ALCANCE” y “EMBESTIDA” la tendencia a lo largo de los años es la de estabilizarse en torno a un valor central en los últimos años mientras que se observa una considerable variación en su número entre años anteriores. Para el caso de los accidentes de tipo “CHOQUE”, “RASCADA” y “CAIDA” se ha detectado un aumento gradual con el paso de los años, mientras que los valores de “ATROPELLO” tienden a ser uniformes sin importar el año en que se produzcan.

6.3.2. Causas de incidentes en los periodos de mayor actividad

Como ya hemos visto en apartados anteriores, al distribuir los incidentes por horas aparecen picos de actividad que coinciden con los horarios habituales de entrada y salida de la jornada laboral y un máximo entre las 4:00 y las 5:00 para el que, por ahora, no tenemos explicación. En este apartado vamos a recopilar las causas de los incidentes que se produjeron durante esos periodos de alta actividad para estudiar el origen de los mismos. Para ello, igual que en el apartado anterior vamos a recopilar la información de las columnas “MOTIVO QUE ORIGINA EL DELITO” y “TIPO DE ACCIDENTE”, pero esta vez combinamos ambas columnas en una a la que denominaremos “CAUSA”. Habitualmente un registro se añade a las tablas del año correspondiente bien porque se trata de un accidente o bien porque se ha cometido un delito, de forma que solo una de las dos columnas de nuestros *dataframes* para un registro dado posee información y no hay ningún problema a la hora de fusionarlas en una única columna de causas. Sin embargo, existen registros cuyo origen es un accidente al que más adelante se han añadido los delitos correspondientes, como el consumo de alcohol, por lo que las dos columnas, “MOTIVO QUE ORIGINA EL DELITO” y “TIPO DE ACCIDENTE”, poseen valor no nulo para el mismo registro. En estos casos y para construir nuestra columna de causas se interpretará como origen del incidente el valor de la columna “TIPO DE ACCIDENTE”. Para conseguir esto basta con cargar primero el contenido de la columna de delitos en un vector vacío y después cargar sobre ese mismo vector el contenido de la columna de accidentes, de forma que en aquellos

registros que ambas columnas posean información prevalece la del accidente, y en aquellos que no simplemente se toma el valor de la columna correspondiente.

A continuación necesitamos seleccionar de nuestro vector "CAUSAS" solamente aquellos incidentes que hayan ocurrido durante el periodo que estamos estudiando. Debido al método que hemos usado para obtener las causas, estas mantienen el mismo orden que los registros en nuestros *dataframes*, lo cual facilita en gran medida nuestra tarea. Mediante la función de R *which()* podemos averiguar que elementos de un vector cumplen una cierta propiedad que nosotros especifiquemos, con lo que podemos averiguar que filas de nuestro *dataframe* de registros poseen un valor de la columna "HORA" dentro del periodo que estamos estudiando:

```
>causas2017_entre_4_y_5 <- causas2017[which(year2017$HORA > 4 &
year2017$HORA<= 5)]
```

Con esto ya tenemos las causas de los incidentes que nos interesan y podemos iniciar el proceso para representarlos: obtener la tabla con el conteo de los elementos, pasarla a formato *dataframe*, añadir una columna con el año correspondiente y unir los *dataframes* obtenidos de cada año, tal y como se describe en el apartado 6.1.1.

Si repetimos este proceso para cada uno de los intervalos de tiempo en los que se producen los máximos que podemos ver en la gráficas de las Figuras 7 y 8 en el apartado 6.6.1., es decir: entre las 4:00 y las 5:00, entre las 12:00 y las 14:00, y entre las 19:00 y las 20:00; podemos construir 3 gráficas que nos indiquen cuales son las causas principales de incidentes para cada intervalo.

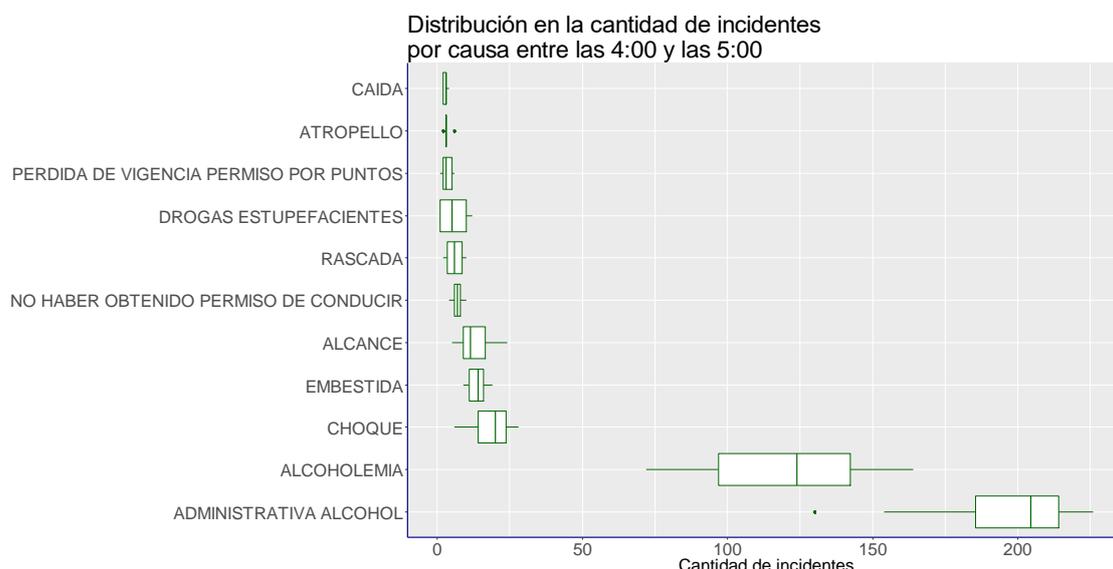


Figura 24: Gráfica de causas entre las 4:00 y las 5:00 (Anexo 9.5)



Estudio sobre incidencias de tráfico en la ciudad de Valencia

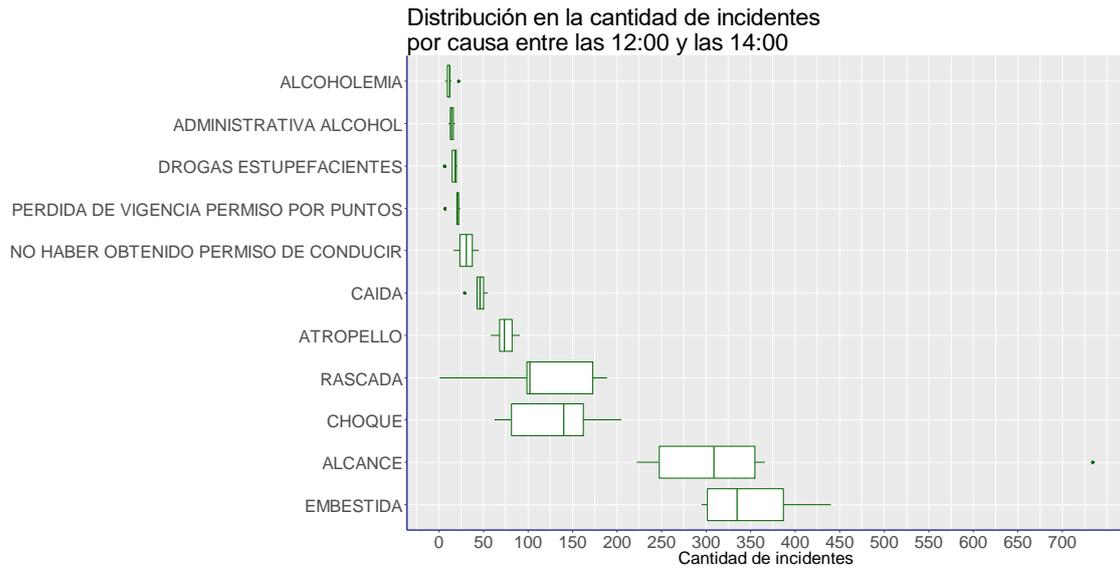


Figura 25: Gráfica de causas entre las 12:00 y las 14:00 (Anexo 9.5)

Si nos centramos en los resultados para el intervalo entre las 4:00 y las 5:00, intervalo en el que hemos detectado un máximo en apartados anteriores para el que no teníamos una explicación, podemos ver que en este se producen incidentes principalmente en torno a dos causas muy similares, los delitos y las sanciones administrativas relacionadas con la alcoholemia. En base a esto podemos proponer que una posible explicación para la cantidad de registros que se producen durante este intervalo sería que es en esta hora cuando el cuerpo de la Policía Local de Valencia establece sus controles de alcoholemia. Si este fuese el caso, este patrón que detectamos cada año sería artificial puesto que no estaría producido por el comportamiento del tráfico y los conductores.

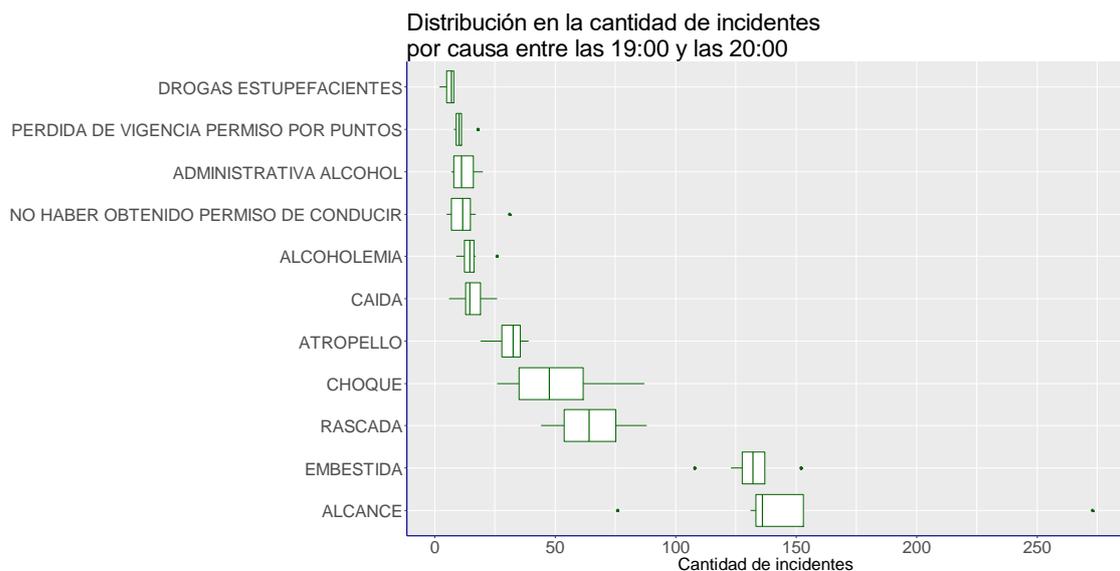


Figura 26: Gráfica de causas entre las 19:00 y las 20:00 (Anexo 9.5)

En el caso de los resultados de los otros dos intervalos de tiempo encontramos que la mayoría de registros se originan por accidentes de diferentes tipos, siendo los más frecuentes “EMBESTIDA” y “ALCANCE”, y que a diferencia del intervalo entre las 4:00 y las 5:00 los delitos tienen muy poca presencia en comparación. Todas las gráficas producidas en este apartado pueden encontrarse en el ANEXO 9.5. A partir de estas gráficas, en el apartado 6.4. Se establecerán cuáles son los puntos de la ciudad de Valencia en los que se producen los tipos de accidentes y delitos más comunes.

6.4. Incidentes según la localización

6.4.1. Obtención de las coordenadas

Cada tabla de registros de incidentes posee como mínimo una columna dedicada a almacenar la localización donde dicho incidente tuvo lugar. Esta información se representa mediante el nombre de una calle o carretera seguido de dos posibles opciones: el número de casa o kilómetro indicando a qué altura de dicha calle o carretera se produjo el incidente; o el nombre de una segunda calle o carretera para indicar que el incidente tuvo lugar en un cruce. Esta representación de los datos resulta útil para que sean comprensibles para cualquier persona que tenga acceso a ellos, por el contrario, a la hora de emplear estos datos con fines estadísticos es necesario aplicar ciertos cambios para que estos puedan resultar útiles. Dado que nuestro objetivo es el de representar el conjunto de todos los incidentes de los que disponemos sobre un mapa de la ciudad de Valencia, el formato más indicado para las localizaciones será el de coordenadas. Para traducir nuestras localizaciones expresadas como un conjunto de caracteres a un par latitud/longitud hemos hecho uso del servicio de geocodificación que nos ofrece Google Cloud Platform.

Para poder hacer uso de este servicio de geocodificación el primer paso será obtener una *key* que nos permita acceder a las API's de Google Cloud Platform a través de RStudio, para lo cual bastará con crear una cuenta en esta plataforma. Una vez obtenida tenemos que registrarla en nuestro entorno de trabajo de forma que las funciones con las que vamos a trabajar puedan acceder a ella, para lo cual hacemos uso de la función `register_google()` del paquete de R `ggmap`, cuyo único argumento será la *key* generada para nosotros por Google Cloud Platform, es decir ejecutamos el siguiente comando:

```
register_google(key = "NUESTRA API KEY")
```

Con esto, otras funciones del paquete `ggmap` ya tienen acceso a nuestra *key* y pueden realizar consultas a la plataforma de Google sin problema, sin embargo, en el caso de que implementemos nuestras propias funciones todavía será necesario que esta *key* sea introducida de nuevo. Entre las funciones de `ggmap` se incluye la función `geocode()`, la cual realiza la geocodificación y nos devuelve las coordenadas de la dirección que se le proporcione como argumento; sin embargo, con el fin de ejemplificar el funcionamiento de esta función y de añadir pequeños cambios que serán necesarios para nuestro conjunto de datos específico, vamos a proceder a construir nuestra propia función de geocodificación.

Los pasos que debe seguir nuestra función son los siguientes: debe construir una URL a partir de la dirección de un incidente, realizar una consulta sobre dicha URL, obtener una respuesta y extraer de dicha respuesta los datos que nos sean de interés. Para construir nuestra URL haremos uso de otra función personalizada para nuestros datos. Esta función secundaria, a la que llamaremos simplemente `url()`, debe obtener como parámetro una dirección y devolver como resultado la URL correspondiente a una petición del servicio de geocodificación de Google Cloud Platform.

```
url <- function(address, return.call = "json") {
  root <- "https://maps.google.com/maps/api/geocode/"
  u <- paste(root, return.call, "?address=", address,
    "&NUESTRA_API_KEY", sep="")
  return(URLEncode(u))
}
```

Tal y como se puede ver en este fragmento de código, la función `url()` parte de la base `"https://maps.google.com/maps/api/geocode/"` a la cual añade los parámetros necesarios para la consulta como son el formato de la respuesta (`return.call`), la dirección (`address`) y nuestra `key`. Una vez se han unido todos los elementos de la URL mediante la función `paste()`, podemos codificar en el formato URL adecuado mediante la función `URLEncode()` del paquete `Rcurl`. Ahora que ya disponemos de un método para construir una URL adecuada ya podemos implementar nuestra función de geocodificación, la cual se muestra en detalle en el siguiente fragmento de código:

```
geocodificar <- function(address) {
  u <- url(address)
  respuesta <- getURL(u)
  x <- fromJSON(respuesta, simplify = FALSE)
  if(x$status=="OK") {
    lat <- x$results[[1]]$geometry$location$lat
    long <- x$results[[1]]$geometry$location$lng
    formatted_address <- x$results[[1]]$formatted_address
    return(c(lat, long, formatted_address))
    Sys.sleep(0.25)
  } else {
    return(c(NA, NA, NA))
  }
}
```

Esta función, a la que hemos denominado `geocodificar`, recibe como parámetro la dirección que se pretende traducir a coordenadas, la cual emplea en primer lugar para fabricar la URL. A continuación, mediante la función `getURL()` realiza la consulta y almacena la respuesta en la variable `respuesta`, la cual se ha recibido en formato JSON por lo que es necesario convertir a `dataframe` mediante la función `fromJSON()`. Una vez disponemos de la respuesta en un formato con el que podemos interactuar mediante comandos de R, comprobamos el estado de la respuesta y en caso de que se haya

producido algún error durante la petición nuestra función devolverá un valor nulo. En caso contrario, procedemos a extraer la información relevante de la respuesta, siendo esta la latitud, la longitud y la dirección formateada, y las devolvemos como resultado de la función. La función `sys.sleep()` que hemos introducido tras el `return()` tiene como objetivo no saturar de peticiones la API de Google, ya que esta limita el número de peticiones por segundo, provocando una pausa de 0.25 segundos entre cada petición. Esto no es necesario si se va a realizar una pequeña cantidad de peticiones, pero dado que disponemos de casi 10000 registros por año debemos asegurarnos de no superar el límite de consultas por segundo.

La función que acabamos de implementar ya es capaz de geocodificar una dirección, sin embargo dicha dirección debe estar expresada como una única cadena de caracteres, y las direcciones de las que disponemos en los registros pueden estar, dependiendo del año al que correspondan, separadas en hasta 3 datos distintos. Para solucionar esto hemos implementado una función lanzadera que unifica los 3 posibles datos en una cadena única, además de añadir información que no se encuentra en los registros como el nombre de la ciudad y el país para que resulte más fácil y precisa la geolocalización. El contenido de la función lanzadera es el siguiente:

```
lanzadera <- function(registro) {  
  address <- paste("calle", as.character( registro[1]), "@", na.  
exclude(registro[2]),"@", na.exclude(registro[3]),",valencia, españa",  
  ", sep = " ")  
  return(geoCode(address))  
}
```

Esta función recibe el contenido de las tres columnas del *dataframe* de registros que contienen la información de la localización (columnas CALLE, Nº y CRUCE del anexo 9.1.) para un único registro, y los combina en una sola cadena añadiendo las palabras “calle” al comienzo y la cadena “valencia, españa” al final. La palabra “calle” al comienzo de la dirección resulta útil para aquellos casos en los que el nombre de la calle que se está geocodificando coincide con el nombre de otra ciudad, por ejemplo “calle de Barcelona”, que en nuestros registros se almacena simplemente como “Barcelona”. Por otra parte, las arrobas que se han intercalado entre los nombres y números de calle actúan como indicadores de comienzo de una nueva calle o dirección; por ejemplo para expresar un cruce se emplea la cadena “calle de Goya @ calle de Quevedo”. En el caso de que no haya una calle de cruce o un número de calle, conseguimos que no influya en nuestra cadena mediante la función `na.exclude()`; por ejemplo si la dirección se compone de una calle y un número, pero no un cruce obtendremos “calle de Goya @ 7 @ , valencia, españa” por lo que no influye en el resultado de ningún modo.

Para poder aplicar la función lanzadera a todas las filas de un *dataframe*, es decir, a todos los registros de un mismo año, hemos empleado la función `apply()` que recibe como entrada las columnas deseadas de un *dataframe* y la función lanzadera, y devuelve una tabla de la misma extensión que el *dataframe* de entrada donde cada fila es el valor de retorno de la función para cada fila de este.

```
coordenadas2017 <- apply(year2017[,31:33], 1, lanzadera)
```

En el ejemplo hemos seleccionado las columnas 31, 32 y 33 del dataframe correspondiente a los registros del año 2017, pero para otros años puede ser necesario modificar estas cifras ya que el número de columnas puede variar.

Si empleamos este método para cada uno de los dataframes de registros de los que disponemos obtenemos las coordenadas de cada incidente, sin embargo, la geocodificación no es perfecta y en algunas ocasiones, debido a un error a la hora de registrar el nombre de la calle por ejemplo, puede que no exista ningún punto que coincida con la localización que encontramos en el registro, en estos casos la API de Google nos devuelve simplemente la coordenada del centro geográfico de la ciudad. Aunque la cantidad de puntos que se ven afectados por este efecto no nos impide trabajar con el resto de coordenadas, sí que pueden influir en los resultados aumentando la densidad de incidentes en el centro de la ciudad, es por ello que en los mapas que hemos desarrollado no se han tenido en cuenta aquellos puntos que apunten al centro de la ciudad. Es poco probable que esta decisión afecte a coordenadas que en efecto apuntan a un incidente ocurrido en esa localización ya que la precisión de las coordenadas es de 7 cifras significativas y solo eliminaremos las coordenadas que coincidan exactamente con latitud = 39.46991 y longitud = -0.3762881.

6.4.2. Distribución y densidad de incidentes

Con las direcciones de cada incidente geocodificadas podemos proceder a representar sus coordenadas sobre un mapa de la ciudad de Valencia. Para lograrlo haremos uso de los paquetes *ggmap* y *ggplot*, los cuales contienen funciones que nos permitirán tanto obtener los mapas que nos sean necesarios, como representar sobre estos los distintos elementos que nos ayuden a visualizar la información.

En primer lugar, hemos obtenido un mapa de la ciudad mediante la función *qmap()*. Esta función hace uso de las APIs de Google Cloud Platform, por lo que es necesario poseer una *key* de acceso a esta plataforma registrada en nuestro entorno de trabajo, tal y como se ha explicado en el apartado anterior. Esta función requiere una dirección en el formato de una cadena de caracteres, un valor para el zoom y, opcionalmente, uno de los posibles tipos de mapa. En nuestro caso, se ha seleccionado como dirección “Valencia, España”, un zoom de valor 13 y el tipo de mapa “Hybrid”.

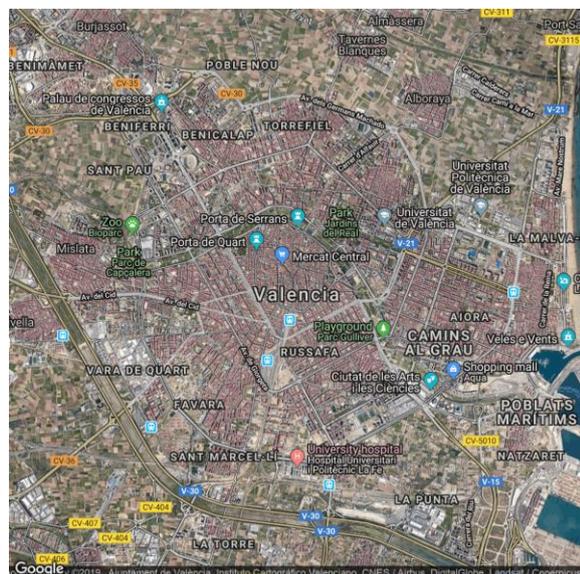


Figura 27: Mapa de la ciudad de Valencia sobre el que se van representar los incidentes. ANEXO 9.6.1.

Sobre este mapa podemos representar directamente cada punto marcado por las coordenadas que hemos obtenido, sin embargo esto nos ofrecería muy poca información útil ya que podría resultar confuso a simple vista localizar las zonas donde se produce una mayor cantidad de incidentes. En su lugar, se ha optado por un mapa de densidad como método de representación de las coordenadas. Este tipo de mapas permiten representar mediante áreas sobre un mapa la concentración de elementos, en nuestro caso, la concentración de incidentes. Este tipo de mapas pueden obtenerse mediante la función `stat_density2d()`, la cual obtiene como entrada un conjunto de coordenadas y elabora de forma automática un mapa de densidad sobre un mapa plano que haya sido construido anteriormente, la posición de las coordenadas y las áreas de densidad se ajusta de forma automática si aplicamos la función sobre un mapa diferente.

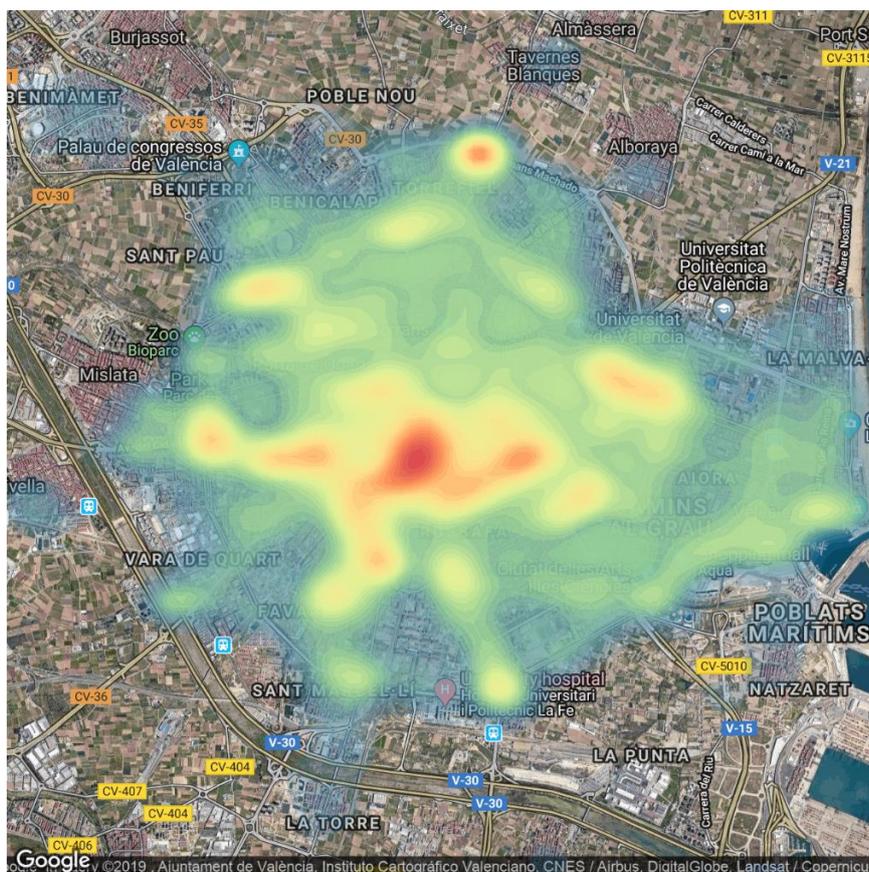


Figura 28: Mapa de densidad de los incidentes sucedidos entre 2010 y 2017 (Anexo 9.6.2.)

Adicionalmente a esta función, podemos usar `geom_density2d()`, la cual representa las líneas que separan las áreas de diferentes densidades para que sea más fácil diferenciarlas. El resultado puede verse en la Figura 28, donde quedan representadas las densidades de todos los incidentes sucedidos entre 2010 y 2017. En este tipo de mapas resulta mucho más fácil apreciar que zonas poseen una mayor frecuencia de incidentes, siendo estas representadas por las zonas de color más intenso y menor transparencia. Sin embargo debido a que las áreas de densidad ocultan parte del mapa, resulta complicado situar con precisión el centro de dichas áreas; para solucionarlo vamos a calcular las coordenadas de los puntos máximos de densidad para conocer su dirección.

Para encontrar los puntos de mayor densidad necesitamos construir, a partir de las coordenadas de los incidentes, una matriz de pesos en la que buscar máximos locales, de forma que cada celda de la matriz corresponda con una sección cuadrada del mapa y cuyo valor sea la densidad de incidentes dentro de dicha sección. Esta acción puede realizarse mediante la función *kde2d()* del paquete MASS y a la que invocamos de la siguiente manera:

```
> x = kde2d(coord$long, coord$lat, n=500, lims = c(-0.424312, -0.323171, 39.440330, 39.501179))
```

Esta función requiere un conjunto de coordenadas, separadas en un vector de longitudes y un vector de latitudes; el tamaño máximo de la matriz cuadrada que se va a crear introduciendo la longitud de uno de los lados en el parámetro n; y por último, 4 puntos que delimiten el espacio real sobre el que se pretende construir la matriz de pesos. Si ejecutamos este segmento de código obtenemos una matriz de pesos, a la cual aplicamos la función *raster()* para convertir dichos pesos en valores de color RGB que nos permitirán representar la matriz mediante una imagen (Figura 29).

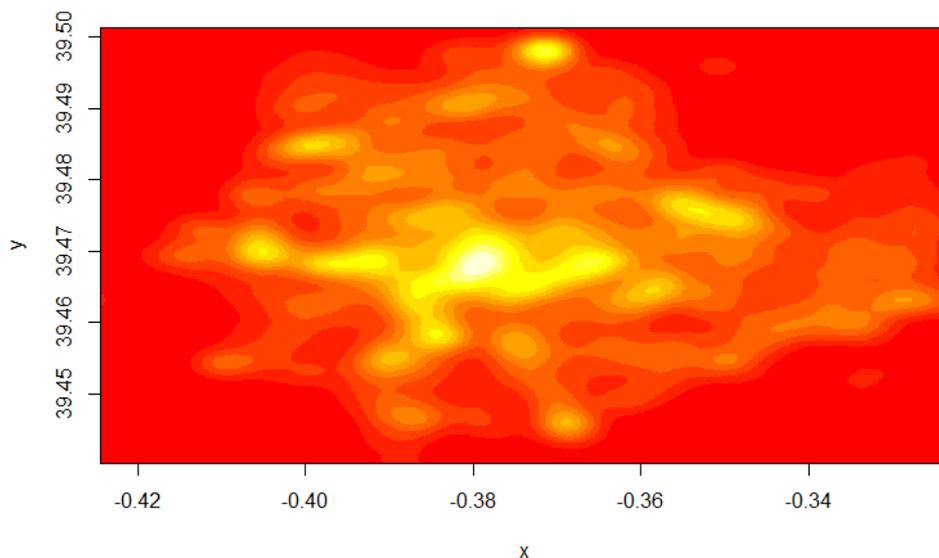


Figura 29: Raster de incidentes en la ciudad de Valencia

Esta transformación resulta necesaria no solo para obtener una representación visual, sino porque el método que se va emplear para hallar los puntos máximos requiere un objeto de tipo *raster* como entrada:

```
> r = raster(x)
> f <- function(x) max(x, na.rm=TRUE)
> w <- matrix(1,51,51)
> localmax <- focal(r, w, fun = f, pad=TRUE, padValue=NA)
> r2 <- r==localmax
> maxXY <- xyFromCell(r2, which(r2==1, cells=TRUE))
```


Si analizamos los resultados obtenidos en este apartado podemos establecer cuáles son las zonas de mayor riesgo con respecto al tráfico de la ciudad de Valencia, sin embargo cabe destacar que las direcciones obtenidas en los puntos de mayor densidad de incidentes no señalan el punto exacto en el que más incidentes se han producido, de hecho podría darse el caso de que no se haya producido ningún incidente en ese punto, puesto que lo que indican es el centro de una zona de gran densidad. Por otra parte este resultado está basado en todos los registros de los 8 años de incidentes que estamos estudiando, sin hacer ningún tipo de diferenciación entre las causas o las circunstancias de dichos incidentes, de modo que puede ofrecernos información sobre qué áreas de la ciudad requieren más presencia policial o un cambio en el forma de dirigir el tráfico pero de un modo muy general. Para conocer datos de carácter más específico, vamos a realizar el mismo proceso pero esta vez sobre conjuntos de registros que en apartados anteriores de este mismo trabajo se haya encontrado que presenten comportamientos destacables, como por ejemplo incidentes sucedidos durante las horas de mayor actividad o incidentes provocados por una causa concreta.

6.4.3. Distribución y densidad de incidentes en base a sus causas y circunstancias.

En primer lugar vamos a analizar las localizaciones de aquellos registros provocados por un choque, alcance o embestida. Se ha decidido unificar los tres tipos de accidentes bajo el mismo análisis ya que básicamente los tres suponen una colisión, solamente se ve alterada la forma en la que se produce.

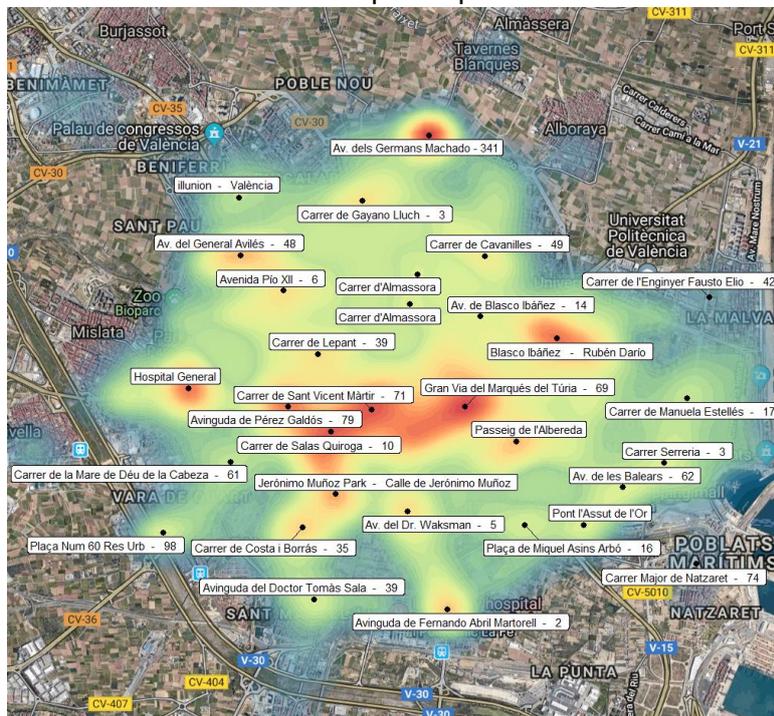


Figura 31: Mapa de densidad y puntos máximos de los accidentes de tipo colisión entre 2010 y 2017 (Anexo 9.6.4)

Si seleccionamos solamente las coordenadas de los incidentes provocados por alguna de estas causas (mediante la función *which()*) como ya se ha empleado en

apartados anteriores) y realizamos el mismo tratamiento que en el apartado anterior obtendremos el mapa disponible en la Figura 31. Si comparamos este mapa con el obtenido a partir de todos los registros, podemos apreciar que ambos son muy similares debido al hecho de que estos tipos de accidentes (choque, embestida y alcance) son la principal causa de incidente en la ciudad de Valencia. Los principales puntos máximos los encontramos en los cruces de las grandes avenidas que atraviesan la ciudad como por ejemplo las Avenidas de Blasco Ibáñez y Pérez Galdós, y la Gran Vía del Marques del Turia. Por otro lado, también destacan en este aspecto aquellos tramos de carretera que actúan como entradas a la ciudad y circunvalaciones en torno a esta donde destacan los puntos de la Avenida de los hermanos Machado, el Hospital General en la Avenida del Cid y la Avenida de Fernando Abril Martorel.

Si en lugar de en las colisiones, nos centramos en los delitos más habituales, es decir, superar la tasa máxima de alcoholemia, nos encontramos con un mapa completamente diferente.

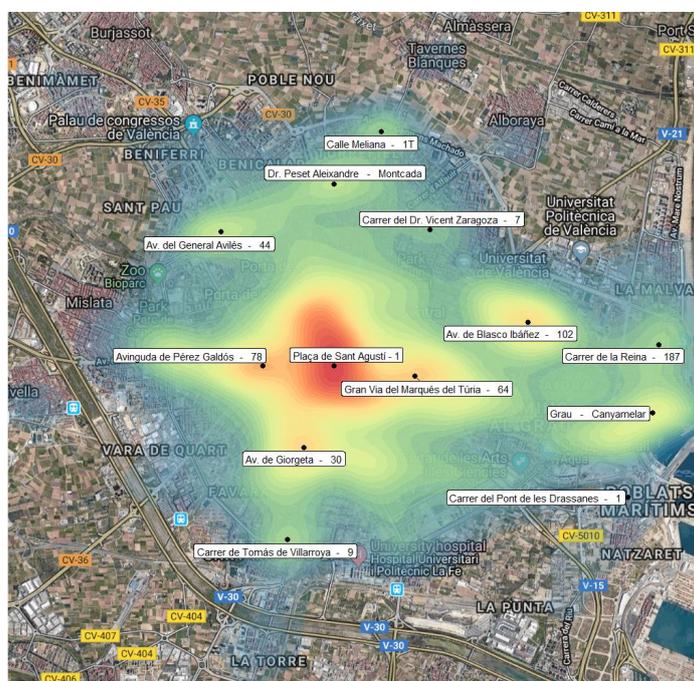


Figura 32: Mapa de densidad y puntos máximos de los delitos por alcoholemia entre 2010 y 2017. (Anexo 9.6.5.)

En aquellos casos en los que el incidente resultó en delito, podemos observar tres focos de actividad principales, uno en centro de la ciudad, otro en la zona de la Avenida de Blasco Ibáñez y un último foco en la parte Este de la ciudad. Si lo comparamos con el mapa que representa las colisiones podemos apreciar que los incidentes que ocurren en los límites de la ciudad siguen apareciendo pero a un nivel menor de intensidad.

Por otra parte, en el caso de las incidencias que solo resultaron en una sanción administrativa seguimos apreciando que la actividad sigue ocurriendo principalmente en el centro de la ciudad, sin embargo han aparecidos nuevos focos en la zona de la Avenida del Cid y cerca del Puerto, mientras que la zona de Blasco Ibáñez ha visto reducida su intensidad

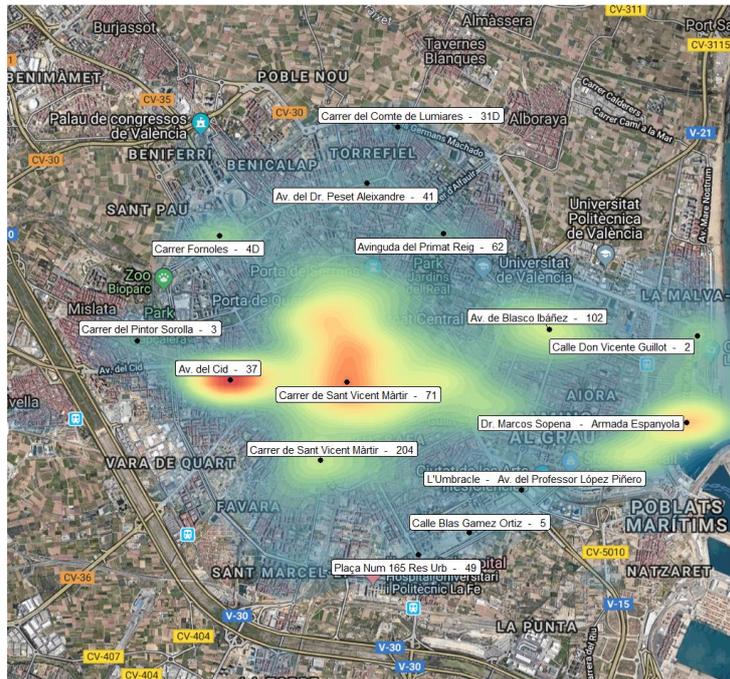


Figura 33: Mapa de densidad y puntos máximos de las sanciones administrativas por alcoholemia entre 2010 y 2017 (Anexo 9.6.6)

Como estudio adicional ajeno a las causas que ya sean analizado en este trabajo se realizado un mapeado de los localizaciones de los accidentes de carácter más grave en busca de posibles puntos negros en las calles de la ciudad de Valencia. Para ello hemos seleccionado aquellos accidentes que registraran víctimas con lesiones graves o muy graves y aquellos en los que se registraron víctimas mortales.

Al representarlos obtenemos los puntos habituales en el centro de la ciudad y las cercanías a la Avenida de Blasco Ibáñez (Calle de Bélgica), pero junto a ellos encontramos puntos claramente distintivos de este tipo de accidentes si los comparamos con el resto de mapas obtenidos. Estos puntos se encontrarían en el cruce de la Gran Vía de Fernando el Católico con el río Turia y en la zona Norte de la ciudad, en el Camino Viejo de Alboraya y en la Avenida de la Constitución.

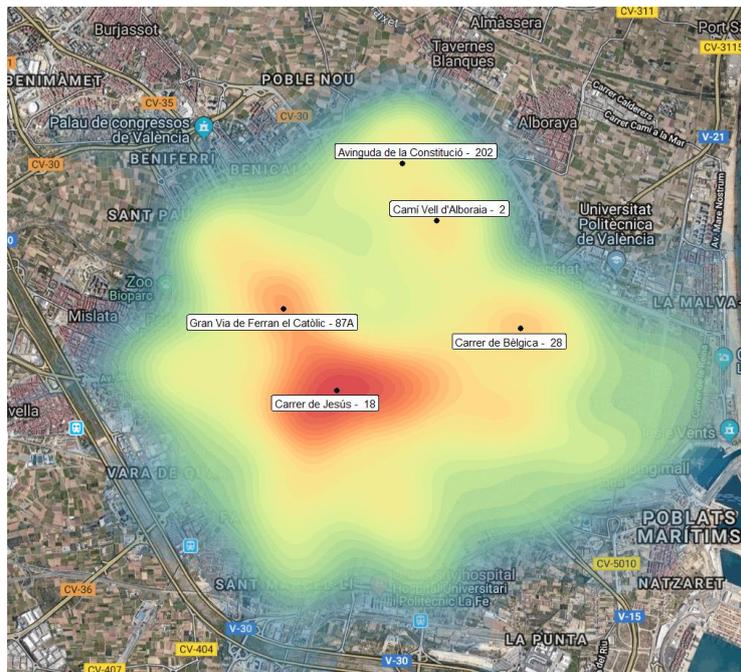


Figura 34: Mapa de densidad y puntos máximos de los accidentes más graves entre 2010 y 2017 (Anexo 9.6.7)

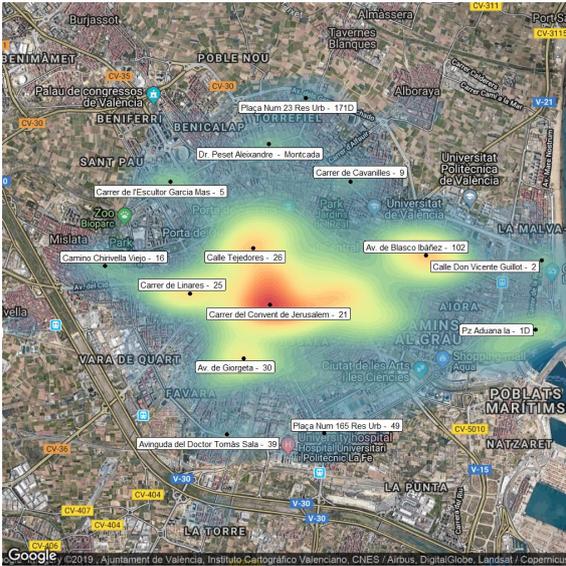


Figura 35: Mapa de densidad y puntos máximos para los incidentes sucedidos entre las 4:00 y las 5:00 (Anexo 9.6.8.)

En estos 3 mapas podemos observar como la frecuencia de los incidentes es diferente en cada uno de ellos, centrándose en el centro de la ciudad y la zona de la Avenida de Blasco Ibáñez durante la madrugada; pero desplazándose hacia las salidas y entradas de la ciudad durante los periodos diarios

Por último, para obtener las zonas donde se concentran la mayoría de incidentes en función de la hora se han vuelto a utilizar las coordenadas correspondientes a todos los incidentes de los que disponemos y a representar únicamente aquellos que tuvieron lugar durante los periodos de máxima actividad que hemos detectado en apartados anteriores, es decir 4:00-5:00, 12:00-14:00 y 19:00-20:00.

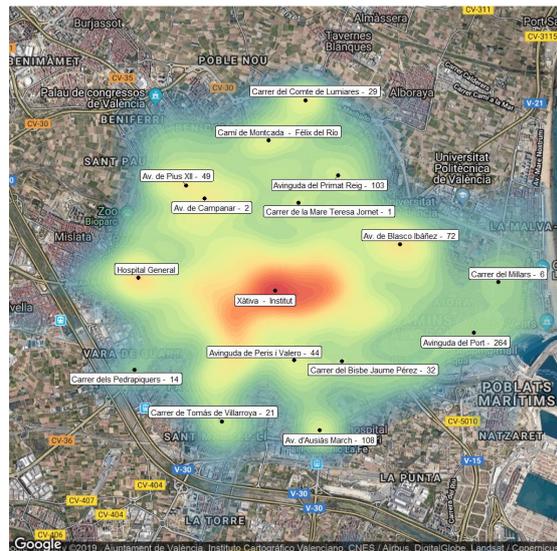


Figura 36: Mapa de densidad y puntos máximos para los incidentes sucedidos entre las 12:00 y las 14:00 (Anexo 9.6.9.)

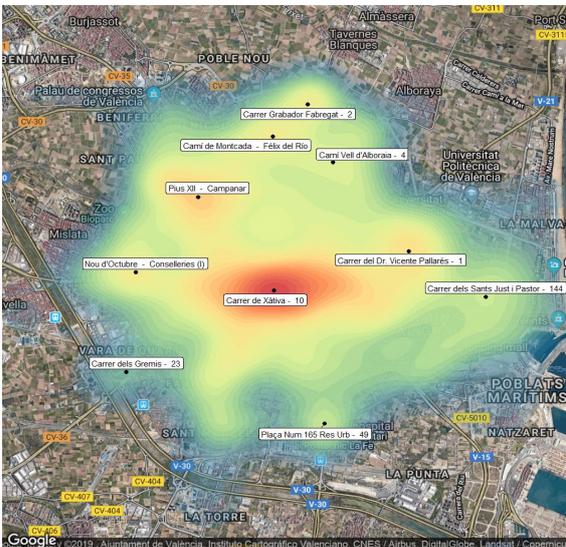


Figura 37: Mapa de densidad y puntos máximos para los incidentes sucedidos entre las 19:00 y las 20:00 (Anexo 9.6.10.)

7. Conclusiones

7.1. Análisis de los resultados

Si observamos los resultados obtenidos en el conjunto de todos los factores que se han estudiado en este trabajo podemos extraer cierta información que puede servir de guía a la hora de orientar la toma de decisiones para mejorar la seguridad en las calles y carreteras de la ciudad de Valencia. En este apartado se realiza una interpretación en detalle de los resultados con el objetivo de obtener datos útiles para este propósito.

En primer lugar y en un ámbito general, a partir de la tabla de Figura 3 en el apartado 6.1.1 podemos observar como la cantidad total de incidentes de tráfico por año ha ido en aumento durante los 8 años en los que se centra este estudio, con excepción de los años 2011 y 2017, donde esta cifra desciende. En conjunto la frecuencia de incidentes va en aumento, pasando de un total de 9279 en su año más bajo a 10549 en el 2016. Estas cifras deben tenerse en cuenta en caso de que en la actualidad se estén llevando a cabo activamente campañas con el objetivo de reducir estas cifras, ya que a simple vista no parecen estar teniendo resultado y pueda ser necesario un cambio de enfoque. Por otra parte estas campañas podrían haber comenzado a ofrecer resultados positivos en los años más recientes, tal y como podría ejemplificar la bajada del número de incidentes en 2017, sin embargo no es posible afirmarlo con los datos disponibles ya que para ello se necesitaría acceso a los datos de años posteriores.

En el caso de los incidentes repartidos por meses en el apartado 6.1 resalta claramente la bajada en el número de incidentes en el mes de agosto para cada uno de los años, lo que indica que este periodo podría ser el indicado para reducir el número de efectivos habituales desplegados por parte de la Policía Local de la ciudad que podrían ser empleados para otras labores, optimizando así el reparto de los recursos disponibles. Por el contrario para los meses de Mayo y Diciembre se recomendaría tomar acciones completamente opuestas y aumentar la cantidad de personal policial disponible durante estos periodos.

Al distribuir nuestros datos en base a periodos de una hora diariamente es cuando surgen algunos de los patrones más interesantes. En los resultados del apartado 6.1.2 es posible observar, en primer lugar, las diferencias entre los días laborables, donde la actividad principal del tráfico sucede durante el día, y los fines de semana, donde los mayores valores se registran de madrugada. Sería posible emplear estos datos de nuevo como indicativo de en qué periodos sería necesario aumentar o disminuir la cantidad de efectivos. Estos periodos se corresponderían principalmente con las horas de entrada y salida del trabajo para la mayoría de los trabajadores, comenzando con el periodo de entrada entre las 8:00 y las 10:00, y siguiendo con los periodos de salida entre las 12:00 y 14:00 y entre las 19:00 y 20:00. En menor medida también sería recomendable tener en cuenta el periodo correspondiente entre las 4:00

y las 5:00, donde se han registrado mayormente delitos relacionados con alcohol y drogas. Si tenemos en cuenta los datos en función de los años, es posible observar que mientras que en los periodos influenciados por los horarios laborales el número de incidentes aumenta cada año siguiendo el comportamiento general de la Figura 3, en el periodo entre las 4:00 y las 5:00 ha descendido lo que podría indicar un éxito en las campañas de prevención y concienciación relacionadas con la conducción bajo los efectos del alcohol, que como ya se ha comentado, es la principal causa de incidentes durante este periodo.

Es en el apartado donde se analizan los efectos de las condiciones meteorológicas en el cual nos encontramos los resultados menos significativos de este trabajo. Tras analizar 3 factores distintos (las precipitaciones, las temperaturas y la velocidad del viento) se ha alcanzado la conclusión de que ninguno de estos posee un impacto directo en el número de incidentes registrados. En el caso de las precipitaciones, los test estadísticos llevados a cabo ofrecen como resultado que la diferencia en la cantidad media de incidentes registrados los días de mayor intensidad en las precipitaciones y el resto no es lo suficiente significativa como para asumir que existe una relación entre ambos. Para el caso de las temperaturas los test si indican una relación, sin embargo los resultados muestran un comportamiento contrario al que se planteaba como hipótesis inicial, lo que nos lleva a concluir que los datos empleados para este estudio concreto se hallan influenciados por otros factores como el periodo anual.

Si nos centramos en las principales causas para todos los incidentes registrados y observamos los resultados obtenidos en el apartado 6.3 vemos que los incidentes provocados por algún tipo de accidente son mucho más frecuentes que aquellos registrados por haber cometido un delito. Esto hace que sea más difícil tomar medidas preventivas ya que la naturaleza de los accidentes los hace impredecibles, sin embargo en el apartado 6.4 se ha construido un mapa a partir de las coordenadas de cada incidente registrado donde se muestran las localizaciones de las zonas donde se tienden a registrar un mayor número de accidentes, lo que resulta útil para conocer en que partes de la ciudad es más probable que pueda ser necesaria una presencia policial o sanitaria, reduciendo de esta manera el tiempo de respuesta necesario para reaccionar ante un posible incidente.

Entre las causas de los incidentes también destaca el hecho de que los accidentes de tipo “EMBESTIDA” y “ALCANCE” son los más frecuentes, a pesar de que podría pensarse que los accidentes del tipo “RASCADA” fuesen los más frecuentes debido a su menor gravedad. Esto puede deberse al hecho de que para la mayoría de los incidentes reales provocados por esta causa no se solicita o no se requiere la presencia de personal policial ya que no alteran el flujo habitual del tráfico. Por otro lado, en el caso de los delitos, los más frecuentes son aquellos relacionados con el alcohol y las drogas, como ya se ha comentado, y aquellos relativos a la obtención o retirada del permiso de conducir; mientras que otros delitos, como el superar el límite de velocidad permitido son prácticamente despreciables a causa de la poca presencia que tienen en nuestros datos. Destacamos de nuevo el hecho de que mientras los delitos por alcoholemia en conductores disminuyen, los delitos por consumo de drogas aumentan,

lo que sugiere una necesidad en nuevas campañas dirigidas a la prevención del consumo de este tipo de sustancias al volante; al mismo tiempo, se deberían mantener las actuales campañas contra la alcoholemia pues, a pesar de estar ofreciendo resultados positivos, las cifras que se han obtenido durante este trabajo todavía podrían ser consideradas motivo de preocupación.

Para finalizar, cuando observamos las localizaciones y mapas que se han obtenido durante los estudios cuyos resultados pueden verse en el apartado 6.4 detectamos que algunas zonas de la ciudad son más propensas a la hora de producirse incidentes determinados. Por ejemplo, mientras que de madrugada la mayoría de incidentes se produjeron principalmente en zonas céntricas de la ciudad y en los barrios cercanos a la costa, coincidiendo con lo que ya se había podido ver en anteriores gráficas con la localización de la mayoría de los delitos por alcohol; durante las horas de salida de la jornada laboral las zonas más afectadas por incidentes de tráfico se encuentran mucho más cercanas a los límites de la ciudad, en las principales vías de entrada y salida de Valencia o en las principales calles que conducen a estas. Conocer esto permitiría centrar esfuerzos en aquellas zonas que resulten más conflictivas durante los periodos de mayor actividad del tráfico.

Los accidentes provocados por colisiones de vehículos se reparten por toda la ciudad, tal y como puede verse en la Figura 31, con focos de densidad dispersos. Sin embargo algunos de estos focos se encuentran especialmente resaltados en nuestros mapas, localizándose a lo largo de las grandes avenidas y principales vías que atraviesan la ciudad desde la Avenida del Cid en el Este hasta la Avenida de Blasco Ibáñez en el Oeste. Fuera de esta norma sin embargo, encontramos un punto con una densidad de accidentes similar aquella registrada en los puntos más céntricos de la ciudad con diferencia de que este punto se halla en el extremo Norte, en la Avenida de los Hermanos Machado. Este punto es especialmente destacable por el hecho de que si bien aparece también como punto máximo en algunos de los otros mapas que hemos realizado, solo alcanza niveles de densidad similares al máximo punto del mapa en el caso de los incidentes provocados por una colisión. Esto nos indica que la localización de este máximo en particular debería de ser cuidadosamente vigilada pues resulta ser especialmente propensa a este tipo de accidentes, los cuales conllevan el mayor riesgo para la vida de los implicados.

A pesar de la distribución que hemos obtenido para las colisiones de tráfico, los puntos donde se han registrado las heridas de mayor gravedad y el mayor número de víctimas mortales siguen centrándose principalmente en el centro de la ciudad. Por otra parte, es en este único mapa donde durante la construcción de nuestros mapas ha surgido un punto máximo que se desplaza de los habituales puntos comunes presentes en otros mapas. Dicho punto, localizado en la Gran Vía de Fernando el Católico solo aparece representado de nuevo en el mapa correspondiente a los incidentes sucedidos durante el periodo entre las 4:00 y las 5:00, pero también se pueden encontrar zonas de gran densidad pero sin máximos en esa zona en los mapas de delitos por alcoholemia. En conjunto, toda esta información sugiere que el alcohol es la principal causa de los accidentes más graves registrados en este punto, lo que lo convierte en un punto clave para la realización de controles Policiales.

7.2. Propuestas de mejora

Durante el desarrollo de este trabajo se han encontrado algunas dificultades, la mayoría de ellas como consecuencia de un registro ineficiente de los datos desde los que se parte. En este apartado se ofrecen alternativas y mejoras para ayudar en la realización de estudios similares en el futuro, con el propósito de disminuir la cantidad de preprocesos necesarios antes de poder trabajar con los datos.

La mayoría de los procesos descritos en el apartado 5 podrían ser innecesarios si se estableciese una normalización a la hora de introducir un nuevo registro en las tablas de atestados por parte de la Policía Local. En primer lugar deberían establecerse formatos adecuados para la información que se va a almacenar, siendo el primer ejemplo de este problema las columnas correspondientes a la representación de la hora a la que se produjo cada incidente en las tablas Excel en las que se basa este trabajo; en ellas cada hora se almacene en la forma de número decimal, donde la parte entera corresponde a las horas y la decimal a los minutos. Este formato no resulta adecuado para este tipo de datos y puede dar origen a problemas y errores ya que, por ejemplo, permite que se pueda almacenar un valor mayor a 59 en la parte decimal o un valor mayor a 23 en la entera, lo cual no tendría ningún sentido en este contexto. Si se va a seguir empleando Excel como archivo para almacenar los registros de atestados este posee un formato especialmente adecuado para representar horas, por lo que la solución para este problema consistiría simplemente en aplicar dicho formato a la columna correspondiente a las horas, de forma que tan solo se puedan registrar siguiendo la norma adecuada.

Para otros datos que sí requieran almacenarse en formato decimal también es necesario establecer una norma común para todos los registros, ya que en los datos que hemos recibido estos aparecían representados en hasta 3 formatos diferentes: separando la parte decimal con coma, punto y coma flotante. Si en un futuro se van a realizar más estudios partiendo de este tipo de datos, lo más recomendable sería emplear tan solo el punto como elemento separador, puesto que es la sintaxis más habitual en la mayoría de lenguajes de programación y en programas de análisis.

Por otra parte, en datos para los que no existe un formato preestablecido podría ser necesario construir uno personalizado antes de comenzar a completar los registros, de manera que cada año se mantuviese de la misma manera. Un ejemplo de estos datos son los diferentes tipos de accidentes y delitos bajo los que se pueden clasificar cada registro. Dado que no existe una lista de la que seleccionar el tipo de accidente más adecuado a la situación que ha tenido lugar, en varias ocasiones dos registros similares han empleado dos designaciones diferentes para un mismo tipo de accidente, lo que más adelante puede dificultar el estudio al clasificar cada registro según sus causas. La solución pasaría por decidir de antemano una lista cerrada de los posibles tipos de accidentes y delitos que se pueden producir y, en el momento de introducir esa información en nuestros registros, seleccionar una opción de dicha lista en lugar de tener que introducirla por escrito.

A pesar de cualquier normalización que se aplique a los datos siempre puede surgir la necesidad de introducir cambios de un año para otro, incluyendo una nueva columna con nuevos datos por ejemplo. En este trabajo recomendamos que cualquiera de estos cambios sea compatible con la información de archivos de años anteriores, manteniendo el mismo nombre para las columnas que sean comunes entre distintos años y no alterando el orden de estas en la medida de lo posible. Esto facilitará la automatización de las tareas de extracción de los datos necesarios para futuros estudios, permitiendo aplicar los mismos métodos a las tablas de registros todos los años sin tener que crear un método específico para cada año como ha ocurrido durante este trabajo.

En cuanto a los estudios realizados en este trabajo también se proponen alternativas y estudios adicionales para profundizar en algunos aspectos de los resultados encontrados. Como se ha podido comprobar en varias secciones del trabajo, la duración de la jornada laboral habitual de los ciudadanos tiene un importante impacto en el flujo del tráfico y la cantidad de incidentes registrados, por lo que podría ser conveniente realizar un estudio centrado en las fechas en las que produjeron eventos y periodos de vacaciones o fiestas de importancia, en preparación para eventos similares en años posteriores.

Si se tuviese acceso a las localizaciones y periodos durante los que se realizaron tareas de construcción en las calles y carreteras que rodean a la ciudad y que hayan obligado a desviar parte del tráfico, podría estudiarse su impacto en otras zonas de la ciudad cuyos resultados podrían emplearse en la optimización de este tipo de eventos, que pueden llegar a generar grandes atascos y poseen un gran impacto en el flujo de vehículos. De igual modo, si en los datos de los que disponemos se incluyesen datos relacionados con cada accidente indicando si al asistir a este fue necesario cerrar algún carril o cortar alguna vía de tráfico, podrían también realizarse estudios similares.

Por otra parte, los datos del tráfico empleados tan solo registran las situaciones en las que la policía estuvo presente, por lo que delitos tales como superar la velocidad máxima permitida, que se detectan de forma automática por radares, no se ven representados. Si se desea realizar un análisis centrado en los delitos e infracciones podrían ser necesarios datos adicionales.

7.3. Relación con los estudios cursados

La mayoría de los conocimientos aplicados en este trabajo se han obtenido de asignaturas como Estadística y Análisis Matemático, las cuales son de carácter obligatorio en el Grado de Ingeniería Informática. En ellas se estudiaron conceptos como los test estadísticos y se introdujo a gráficos de carácter más avanzado como los Box-Plot, que han resultado de considerable utilidad la hora de representar los resultados de este estudio.

Por otra parte, si bien en ninguna de las asignaturas del Grado que he cursado se ha empleado R o RStudio, lo que ha supuesto el tener que aprender este lenguaje y el manejo de este entorno desde cero, los conceptos generales de otros lenguajes

similares empleados durante las prácticas de asignaturas como Percepción y Aprendizaje Automático han facilitado mucho esta tarea. Adicionalmente, durante estas dos asignaturas también se nos introdujo a la interpretación de resultados y la construcción de gráficas con estos de forma práctica.

8. Bibliografía

- [1] RStudio Team (2016). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA URL: <<http://www.rstudio.com/>>
- [2] JSON-CSV.COM. *JSON to CSV Converter*. Consultado: 12/04/19. URL: <<https://json-csv.com/>>
- [3] Dirección General de Tráfico. *Consejos estacionales*. Consulta: 20/04/19. URL: <<http://www.dgt.es/es/el-trafico/recomendaciones/consejos-estacionales/>>
- [4] Agencia Estatal de Meteorología. *Manual de uso de términos meteorológicos, Edición 2015*. URL: <http://www.aemet.es/documentos/es/eltiempo/prediccion/comun/Manual_de_uso_de_terminos_met_2015.pdf>
- [5] Adrian A. Dragulescu and Cole Arendt (2018). *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*. R package version 0.6.1. URL: <<https://CRAN.R-project.org/package=xlsx>>
- [6] Hadley Wickham (2011). *The Split-Apply-Combine Strategy for Data Analysis*. Journal of Statistical Software, 40(1), 1-29. URL: <<http://www.jstatsoft.org/v40/i01/>>
- [7] D. Kahle and H. Wickham. *ggmap: Spatial Visualization with ggplot2*. The R Journal, 5(1), 144-161. URL: <<http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>>
- [8] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [9] Duncan Temple Lang and the CRAN team (2019). *RCurl: General Network (HTTP/FTP/...) Client Interface for R*. R package version 1.95-4.12. URL: <<https://CRAN.R-project.org/package=RCurl>>
- [10] Duncan Temple Lang and Jonathan Wallace (2018). *RJSONIO: Serialize R Objects to JSON, JavaScript Object Notation*. R package version 1.3-1.1. URL: <<https://CRAN.R-project.org/package=RJSONIO>>
- [11] Google LLC. Google Cloud Platform. URL: <<https://cloud.google.com/>>
- [12] Frank E Harrell Jr, with contributions from Charles Dupont and many others (2019). *Hmisc: Harrell Miscellaneous*. R package version 4.2-0. URL: <<https://CRAN.R-project.org/package=Hmisc>>

- [13] Ministerio de Sanidad, Consumo y Bienestar Social. *Temperaturas Umbrales por Capitales de Provincia*. Página web del Ministerio de Sanidad, Consumo y Bienestar Social. URL: <http://www.msrebs.gob.es/ciudadanos/saludAmbLaboral/planAltasTemp/2016/tempeUmbrales.htm>
- [14] Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- [15] Robert J. Hijmans (2019). *Raster: Geographic Data Analysis and Modeling*. R package versión 2.9-5. URL: <https://CRAN.R-project.org/package=raster>
- [16] Dirección General de Tráfico. *Anuario estadístico de accidentes*. Portal Web DGT. Consultado: 18/06/2019. URL: <http://www.dgt.es/es/seguridad-vial/estadisticas-e-indicadores/publicaciones/anuario-estadistico-accidentes/>
- [17] Ayuntamiento de Valencia. *Estadísticas de tránsito y transporte*. Págin Web Ayuntamiento de Valencia. Consultado: 18/06/19. URL: <http://www.valencia.es/ayuntamiento/catalogo.nsf/IndiceAnuario?readForm&lang=1&capitulo=3&tema=7&bdOrigen=ayuntamiento/estadistica.nsf&idApoyo=58FB3C7A3D56E414C1257DD40057EB6C>
- [18] World Health Organization. *Road traffic injuries*. Consultado: 26/07/19. URL: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- [19] M Bhagyaiah and B Shrinagesh. *Traffic Analysis and Road Accidents: A Case Study of Hyderabad using GIS*. 2014 IOP Conf. Ser.: Earth Environ. Sci. 20012026
- [20] Aghasi, N. (2018) *Spatio-Temporal Analysis on Urban Traffic Accidents: A Case Study of Tehran City, Iran*. *Journal of Geographic Information System*, 10, 603-642. doi: 10.4236/jgis.2018.105032.
- [21] Ahmad Tavassoli Hojati, Phil Charles, Luis Ferreira, Mohamad Raduan bin Kabit. *An analysis of traffic incidents on an Australian urban road network*. The University of Western Australia.
- [22] Montesinos Muñoz, J. (2018). *Estudio sobre patrones de accidentes en la ciudad de València*. URL: <http://hdl.handle.net/10251/111763>
- [23] Carlos Martínez Pérez. *Urban Traffic Accident data collection and analysis in Europe: Survey study*. Instituto de Tráfico y Seguridad Vial (INTRAS), Universitat de València. URL: <https://www.uv.es/sau/ppts/CMARTINEZ.pdf>
- [24] Dña. Elena López de Cózar, Jaime Sanmartín Arce, Mauricio Chisvert Perales. *Guía de 'buenas prácticas' para la recogida, gestión y análisis de datos de accidentes de tráfico urbanos*. Instituto de Tráfico y Seguridad Vial (INTRAS), Universitat de València. URL: https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/projects_sources/sau_deliverable_4_buenas_practicas.pdf

9. Anexo

9.1. Tabla descriptiva del contenido de los archivos “RED-ATESTADOS 20XX”

Nombre de la columna	Posibles valores de los datos	Explicación
REGISTRO	Número entero no negativo	Número que actúa como identificador del suceso dentro del año correspondiente.
DIA DE LA SEMANA	Uno de los siguientes caracteres: (L,M,X,J,V,S,D)	Día de la semana en la que se produjo el suceso
DIA	Número entero en el intervalo 1..31	Día del mes en el que se produjo el suceso
MES	Número entero en el intervalo 1..12	Mes del año en el que se produjo el suceso
HORA	Número decimal	Hora a la que se produjo el suceso. La parte entera representa la hora y la decimal, los minutos.
FUGA	Un 1 o un espacio en blanco	Indica si el responsable del accidente se dio a la fuga. Representado con un 1 en caso afirmativo y con nada en caso contrario.
FUGA FINALIZADA	Un 1, un espacio en blanco o la cadena “SD”	Indica si el responsable fue localizado tras su fuga. Marcado como un 1 en caso positivo. Dado que en la descripción del archivo no se especifica “SD” como posible valor a esta columna, se desconoce su significado.
INMOVILIZACIÓN	Un 1 o un espacio en blanco	Indica si el vehículo responsable del delito o accidente fue inmovilizado.

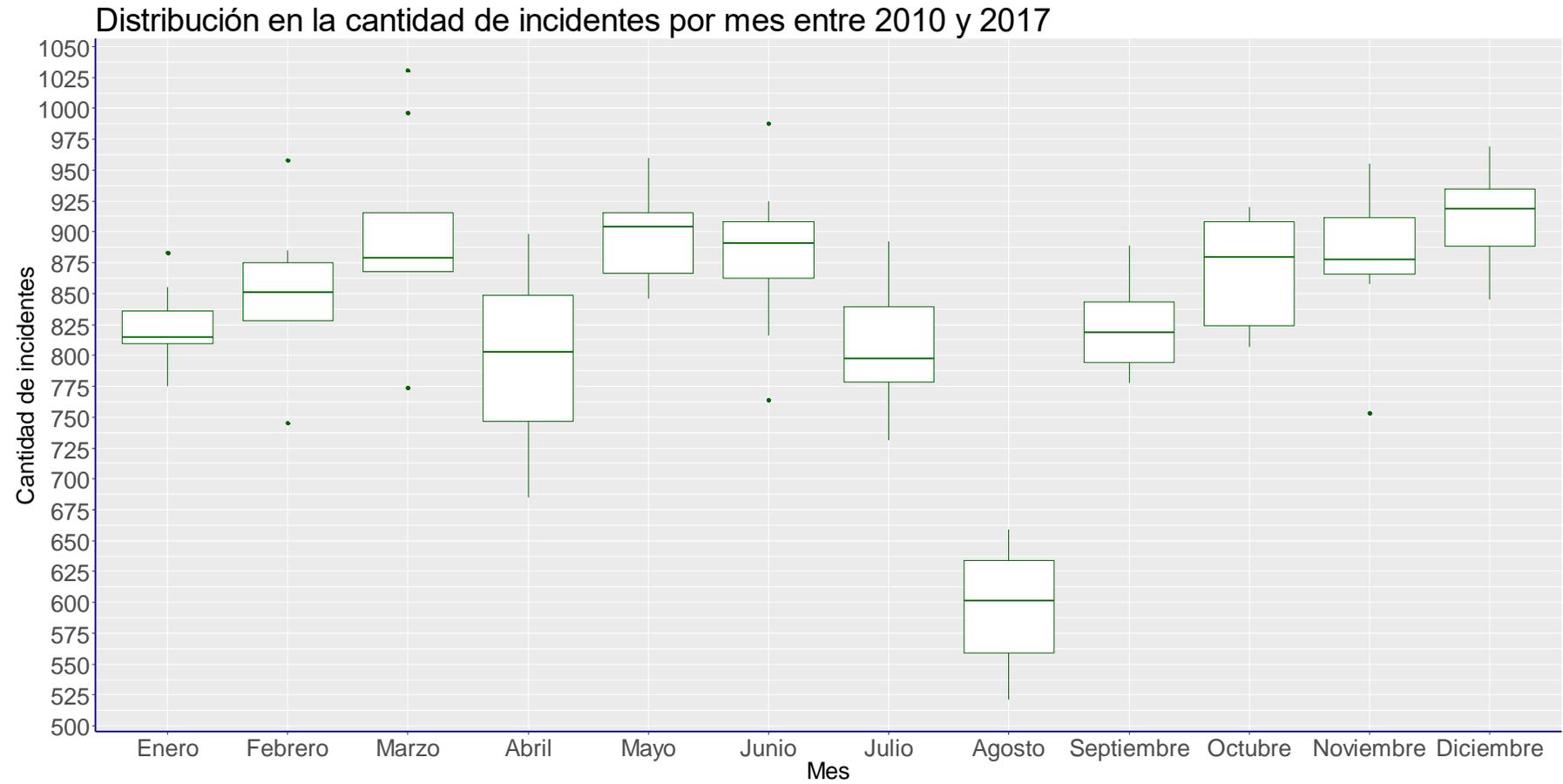
FECHA INMOVILIZACIÓN	El día y el mes separados por un guion	Fecha en la que se produjo la inmovilización del vehículo.
DESINMOVILIZACIÓN	Un 1 o un espacio en blanco	Indica si el vehículo responsable del delito o accidente fue desinmovilizado.
FECHA DESINMOVILIZACIÓN	El día y el mes separados por un guion	Fecha en la que se produjo la desinmovilización del vehículo.
RESEÑA EN C.N.P.	Un 1 o un espacio en blanco	Indica si el registro del accidente originó una reseña en el C.N.P., si el autor fue investigado, detenido o puesto en libertad, respectivamente.
INVESTIGADO	Un 1 o un espacio en blanco	
DETENIDO	Un 1 o un espacio en blanco	
PUESTA LIBERTAD	Un 1 o un espacio en blanco	
TRASLADO CEDRO	Un 1 o un espacio en blanco	
CONDUCCION TEMERARIA	Un 1 o un espacio en blanco	Indica si se produjo un juicio rápido por la causa correspondiente.
PERMISO	Un 1 o un espacio en blanco	
COOPERADOR	Un 1 o un espacio en blanco	
DROGAS	Un 1 o un espacio en blanco	
J. R. ALCOHOLEMIA	Un 1 o un espacio en blanco	
ALCOHOLEMIA ADMTVA.	Un 1 o un espacio en blanco	Indica si el delito lleva aparejado un accidente de circulación.
ATTD. X DELITO+ACCTE. (penal)	Un 1 o un espacio en blanco	
ACCTE. SIN VICTIMAS	Un 1 o un espacio en blanco	Indica si no se produjo ninguna víctima en el accidente.

DILIG. PREVENCIÓN	Un 1 o un espacio en blanco	Indica si fue necesario la hospitalización de alguna víctima.
ACCTE. CON VÍCTIMAS	Un 1 o un espacio en blanco	Indica si se produjo alguna víctima en el accidente.
RADAR PENAL	Un 1 o un espacio en blanco	Indica si el vehículo fue detectado por un radar.
INTERV. CON MENORES	Un 1 o un espacio en blanco	Indica si fue necesaria la intervención con menores.
CALLE	Nombre de la calle	Indica la calle en la que se produjo el accidente
Nº	Número entero o espacio en blanco	Indica en que punto de la calle o en que cruce se produjo el accidente. Solo se completa una de las dos columnas por accidente.
CRUCE	Nombre de la calle o espacio en blanco	
MOTIVO QUE ORIGINA EL DELITO O INFRACCIÓN	Nombre del delito	Indica, si lo hubo, el delito que originó el accidente.
TIPO ACCIDENTE	Nombre del tipo de accidente	Indica el tipo de accidente.
OCUPANTES	Número entero o espacio en blanco	Indica el número total de ocupantes de todos los vehículos implicados, incluido el conductor
PEATON	Número entero o espacio en blanco	Indica el número de peatones implicados.
DAÑOS MUNICIPALES	Un 1 o espacio en blanco	Indica si se han producido daños municipales.
ACCIDENTE LABORAL	Un 1 o espacio en blanco	Indica si se trata de un accidente laboral
UNIDAD 1	Nombre del tipo de vehículo	Indica el tipo de vehículo responsable del accidente.
UNIDAD 2	Nombre del tipo de vehículo o espacio en blanco	Indica el tipo del segundo vehículo involucrado en el accidente si lo hubiera.
SIN SEGURO	Un 1 o espacio en blanco	Indica si el vehículo responsable está asegurado.

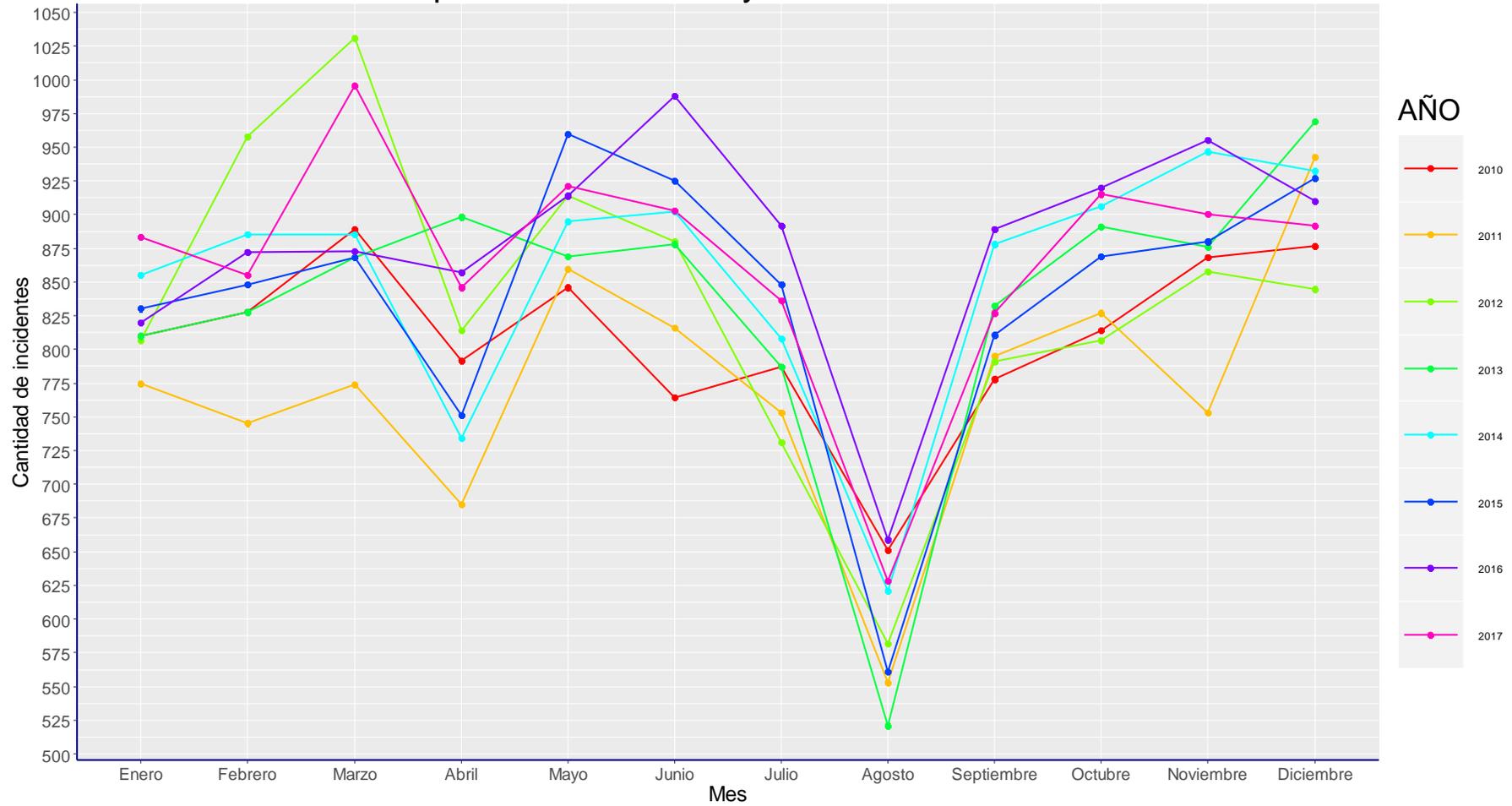
ALCOHOLEMIA	Un 1 o espacio en blanco	Indica si se realizó la prueba de alcoholemia a alguno de los conductores.
1ª PRUEBA-COND,1	Número decimal	Indica el resultado de la primera prueba de alcoholemia al conductor del primer vehículo en mg/l
2ª PRUEBA-COND,1	Número decimal	Indica el resultado de la segunda prueba de alcoholemia al conductor del primer vehículo en mg/l
1ª PRUEBA-COND,2	Número decimal	Indica el resultado de la primera prueba de alcoholemia al conductor del segundo vehículo en mg/l
2ª PRUEBA-COND,2	Número decimal	Indica el resultado de la segunda prueba de alcoholemia al conductor del segundo vehículo en mg/l
SE NIEGA O NO CORRECTA	Un 1 o espacio en blanco	Indica si uno de los conductores se negó a realizar la prueba de alcoholemia o el resultado no fue correcto.
EXT. SANGRE		Indica si se realizó una extracción de sangre a uno de los conductores para la prueba de drogas.
DROGAS NEGATIVAS		Indica si se produjo un resultado positivo en la prueba de drogas.
RADAR Km/h		Indica la velocidad a la que el vehículo responsable fue detectado por un radar.
PAÍS DE ORIGEN	Nombre de país	Indica el país de origen del conductor responsable del accidente o delito.
HERIDOS	Número entero o espacio en blanco	Indica el número de heridos, muertos y lesiones producidos durante el accidente.
MUERTOS		
LESIONES LEVES		
LESIONES GRAVES		

LESIONES MUY GRAVES		
SAMU-SVB	Nombre de equipo médico o espacio en blanco	Indica que los lesionados fueron atendidos por el equipo médico correspondiente.
HOSPITAL	Nombre del hospital o espacio en blanco	Indica el hospital a l que fueron trasladados los heridos.
UNIDAD	Código de la unidad de Policía Local o espacio en blanco	Indica el código que identifica a la unidad de policía encargada de gestionar el delito o accidente.
TURNO	“MAÑANA”, “TARDE” o “NOCHE”	Indica el turno de al que pertenecían la unidad encargada de gestionar el delito o accidente.
UNIDAD 3	Nombre del tipo de vehículo o espacio en blanco	Indica el tipo del resto de vehículos implicados, si los hubiera.
UNIDAD 4		
UNIDAD 5		
UNIDAD 6		
UNIDAD 7		
UNIDAD 8		

9.2. Gráfica de incidentes en función del mes

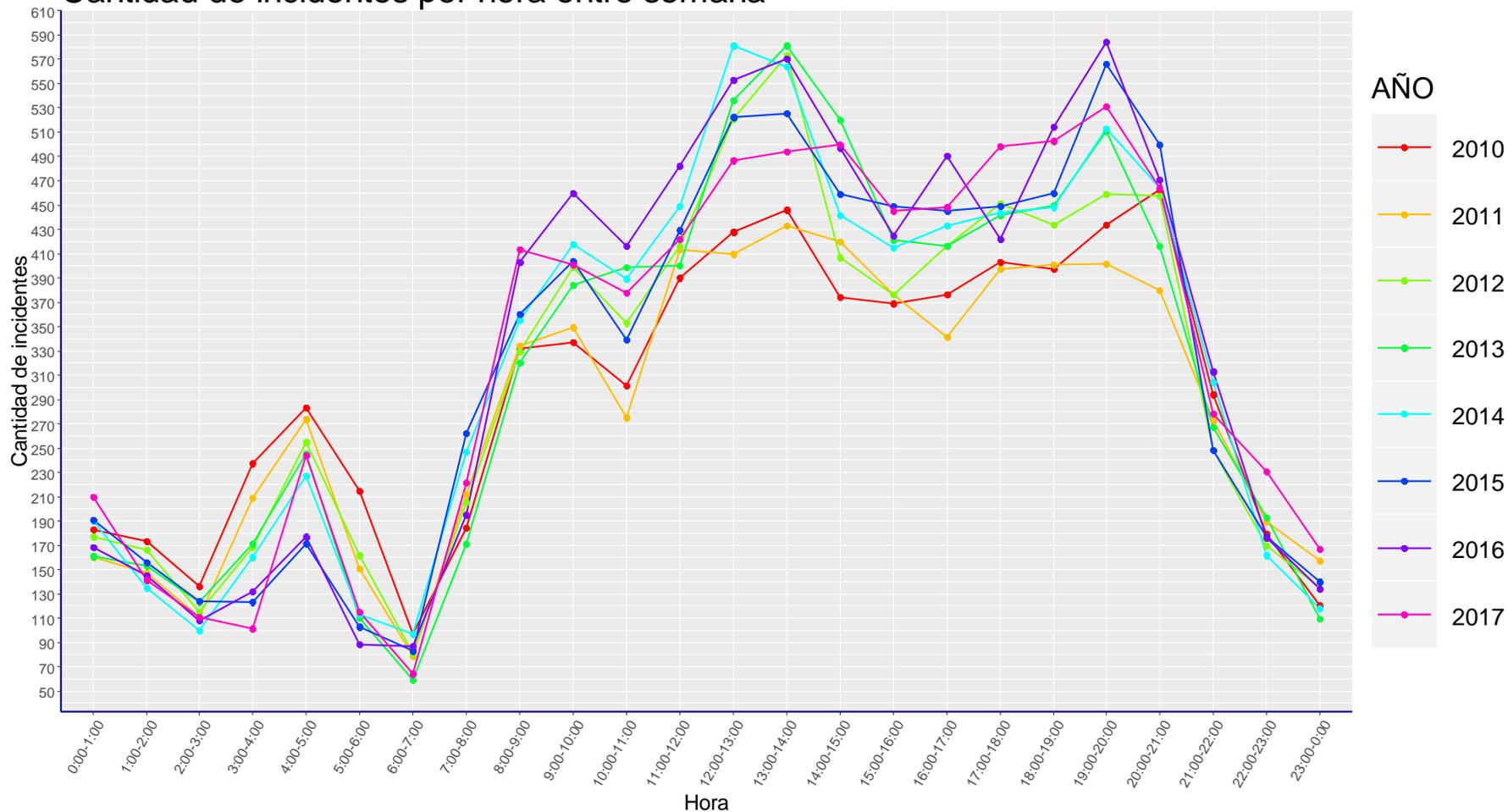


Cantidad de incidentes por mes entre 2010 y 2017

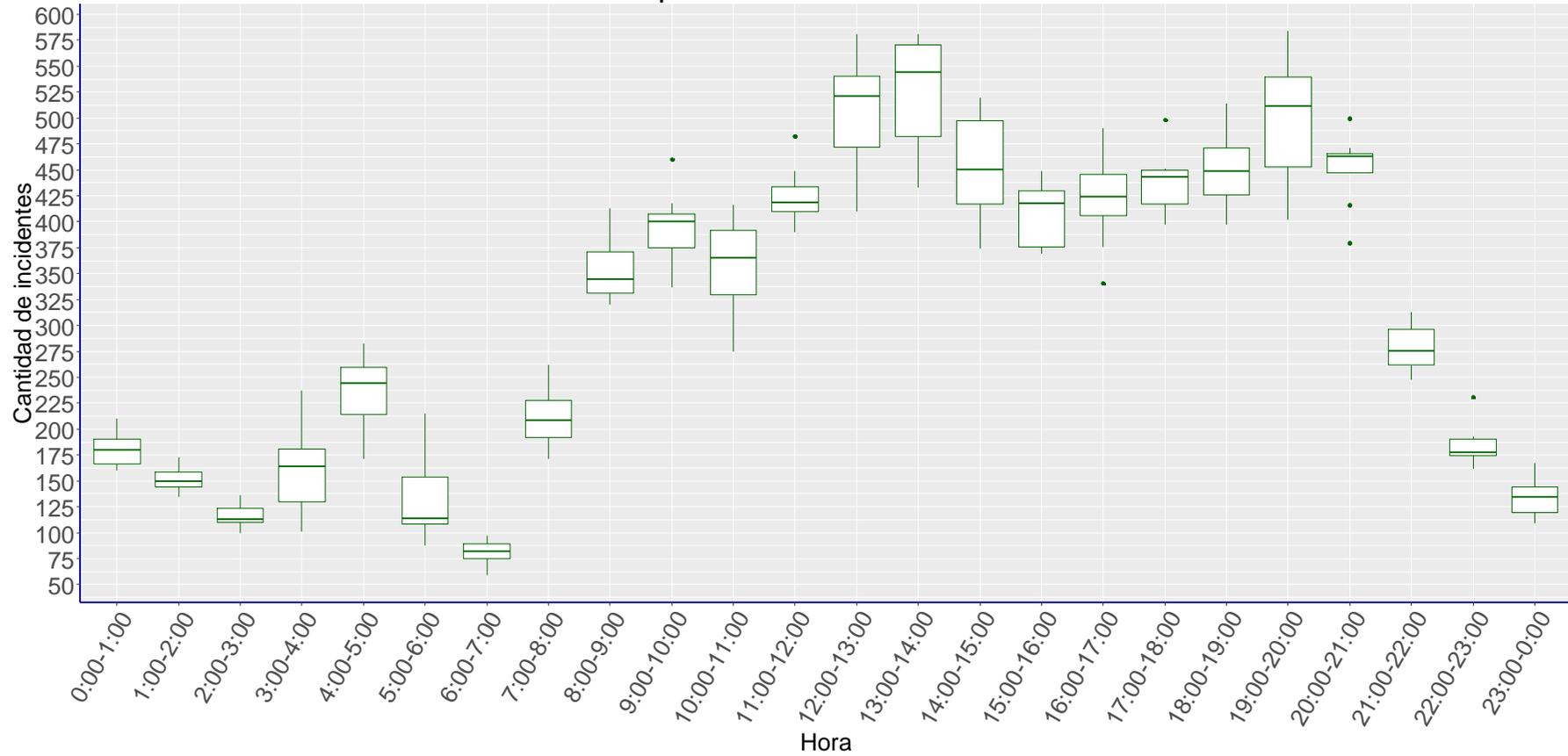


9.3. Gráficas de incidentes en función de la hora

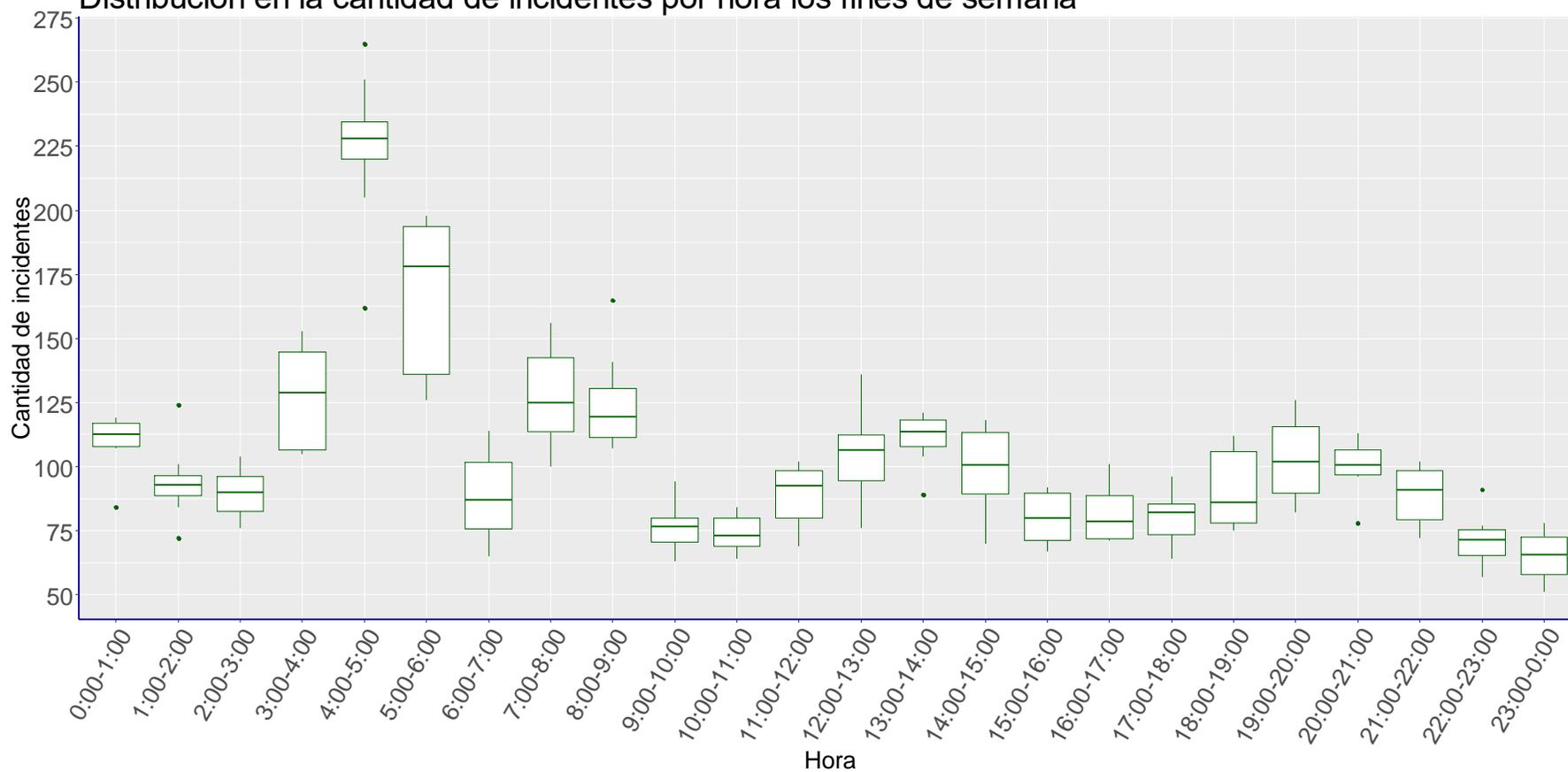
Cantidad de incidentes por hora entre semana



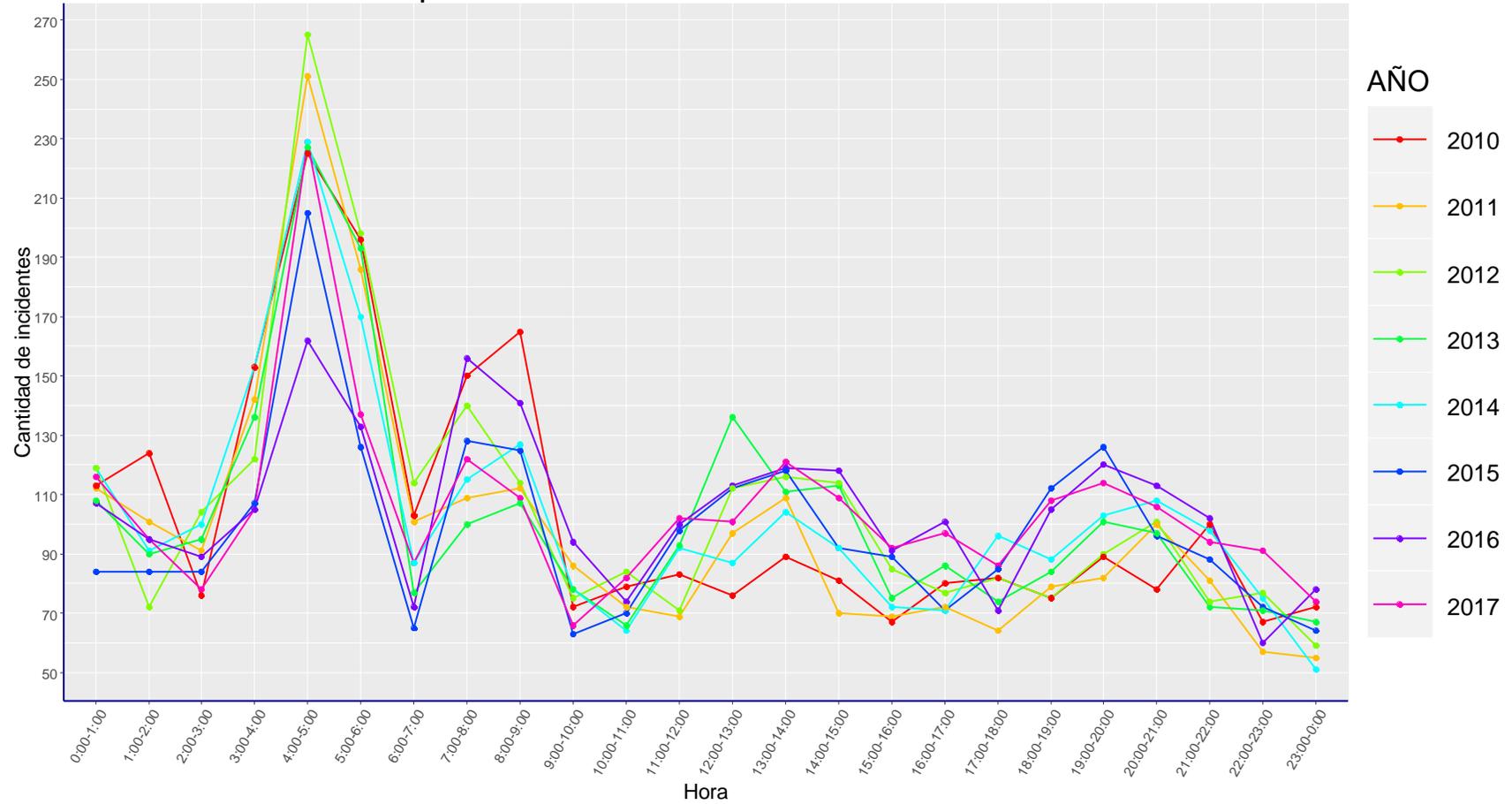
Distribución en la cantidad de incidentes por hora entre semana



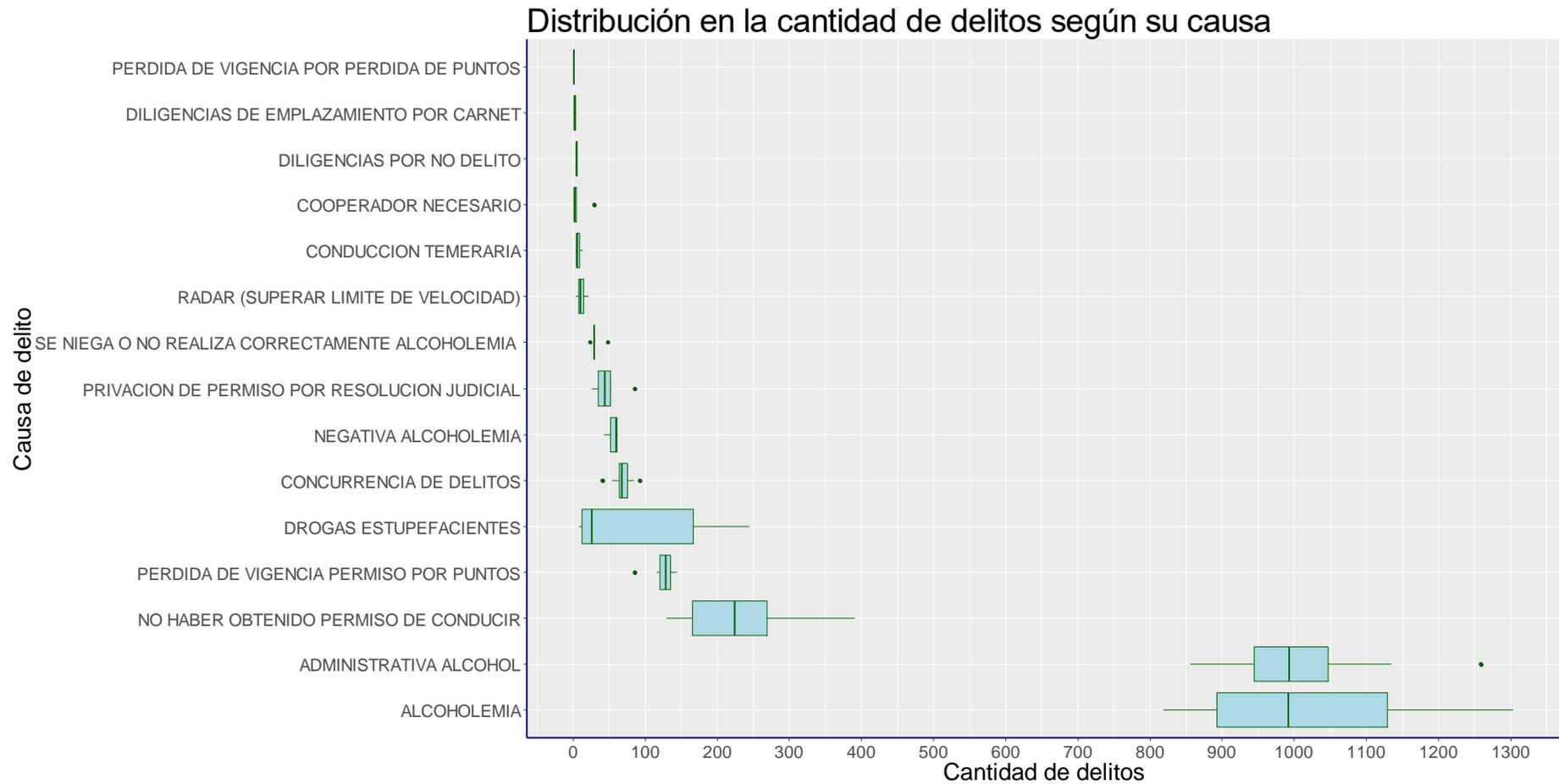
Distribución en la cantidad de incidentes por hora los fines de semana



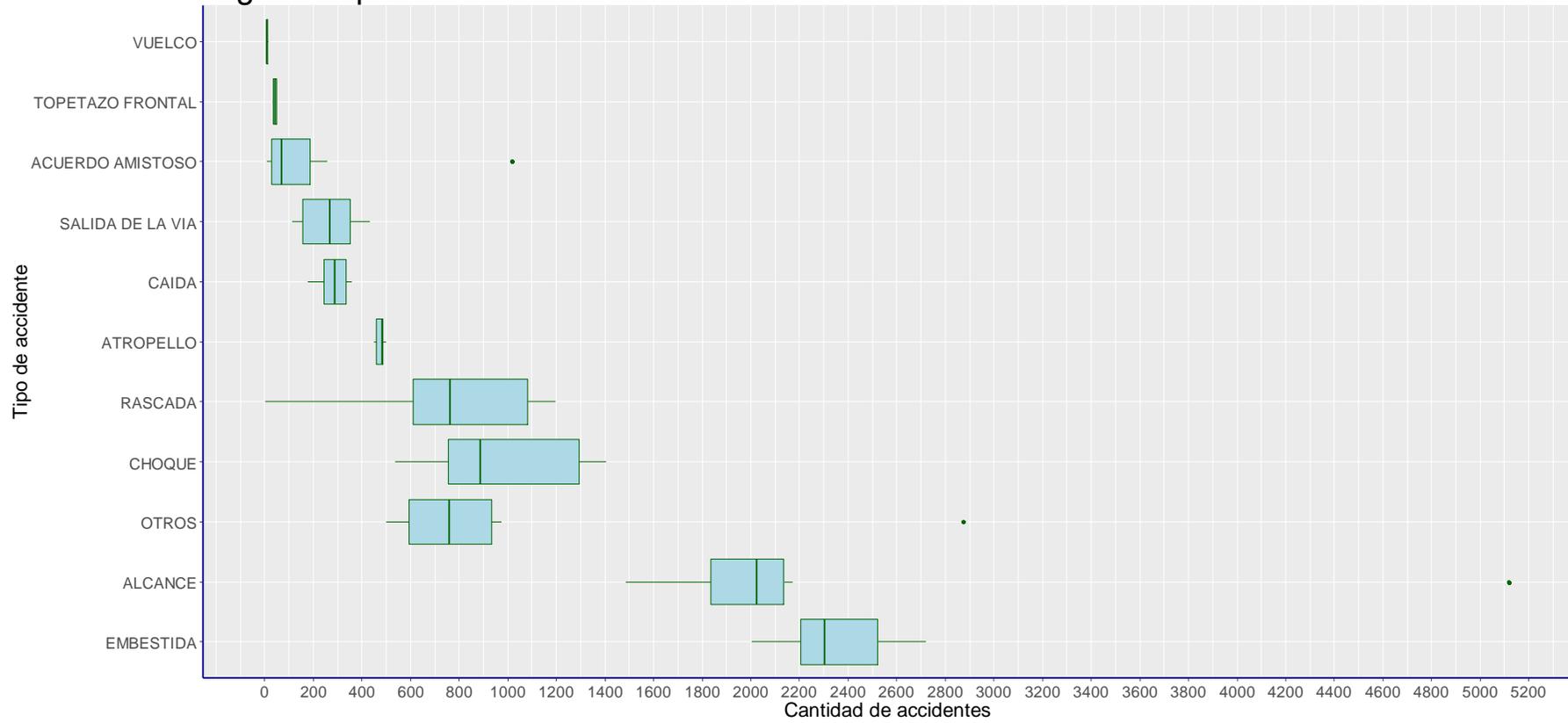
Cantidad de incidentes por hora los fines de semana



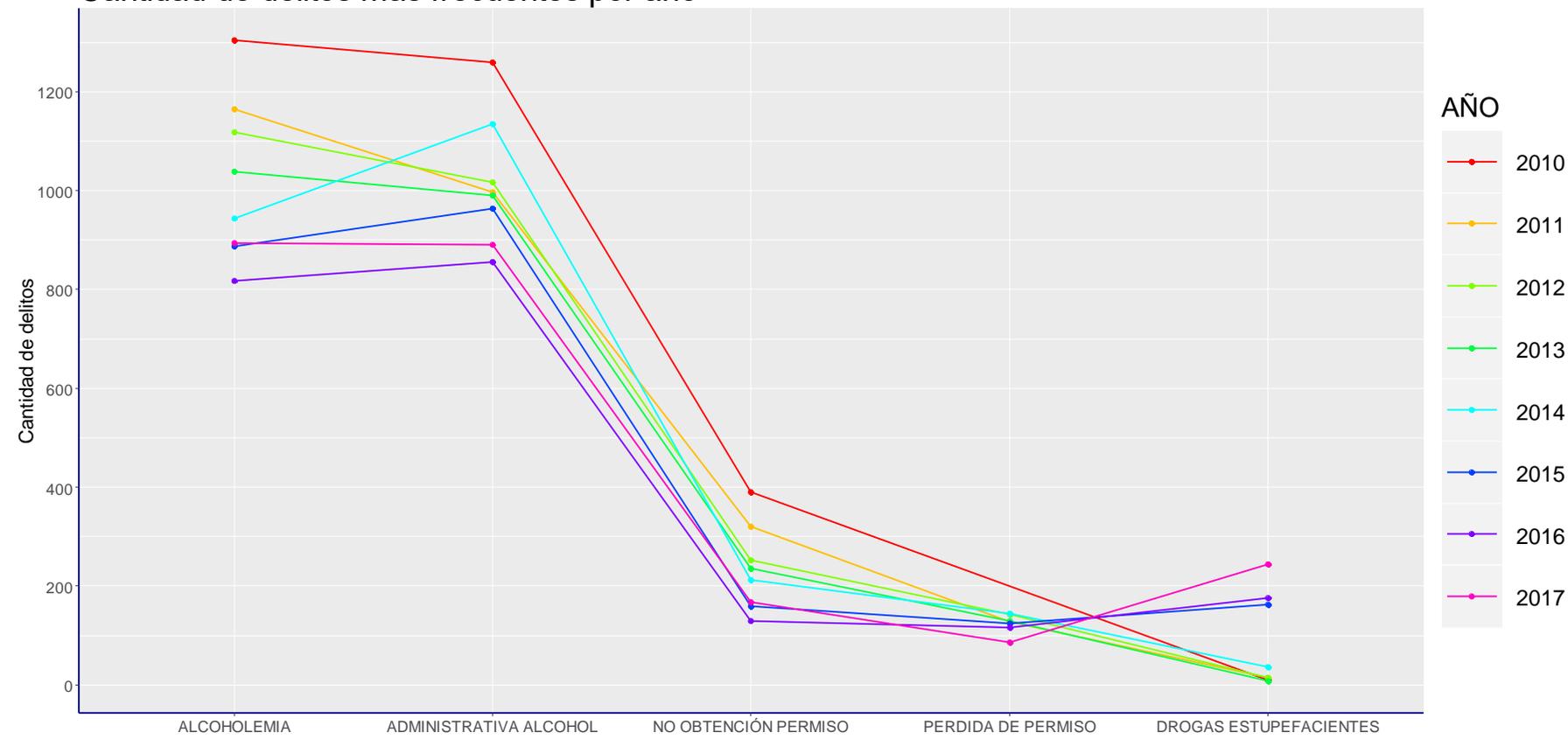
9.4. Gráficas de cantidad de incidentes según su causa



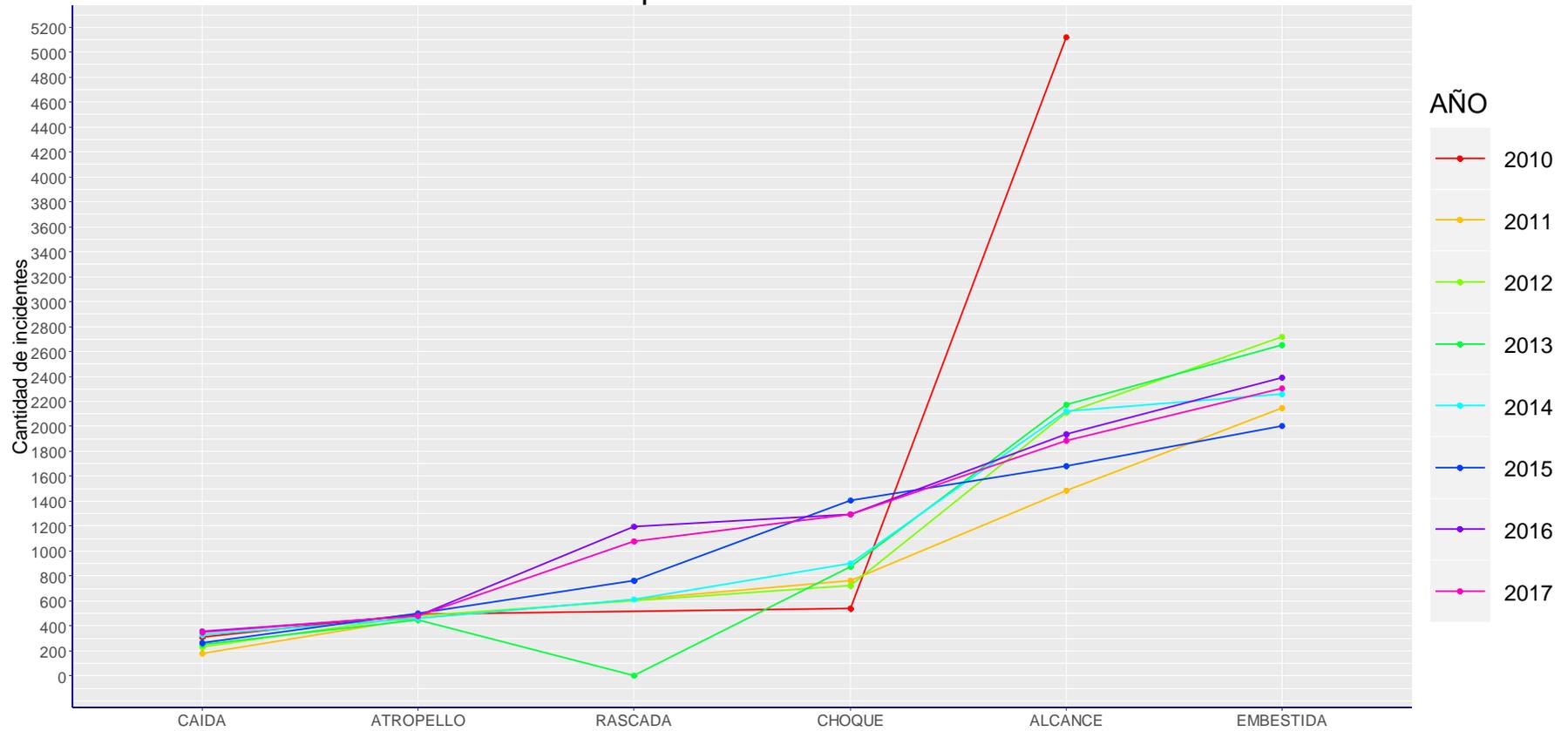
Distribución en la cantidad de accidentes según su tipo



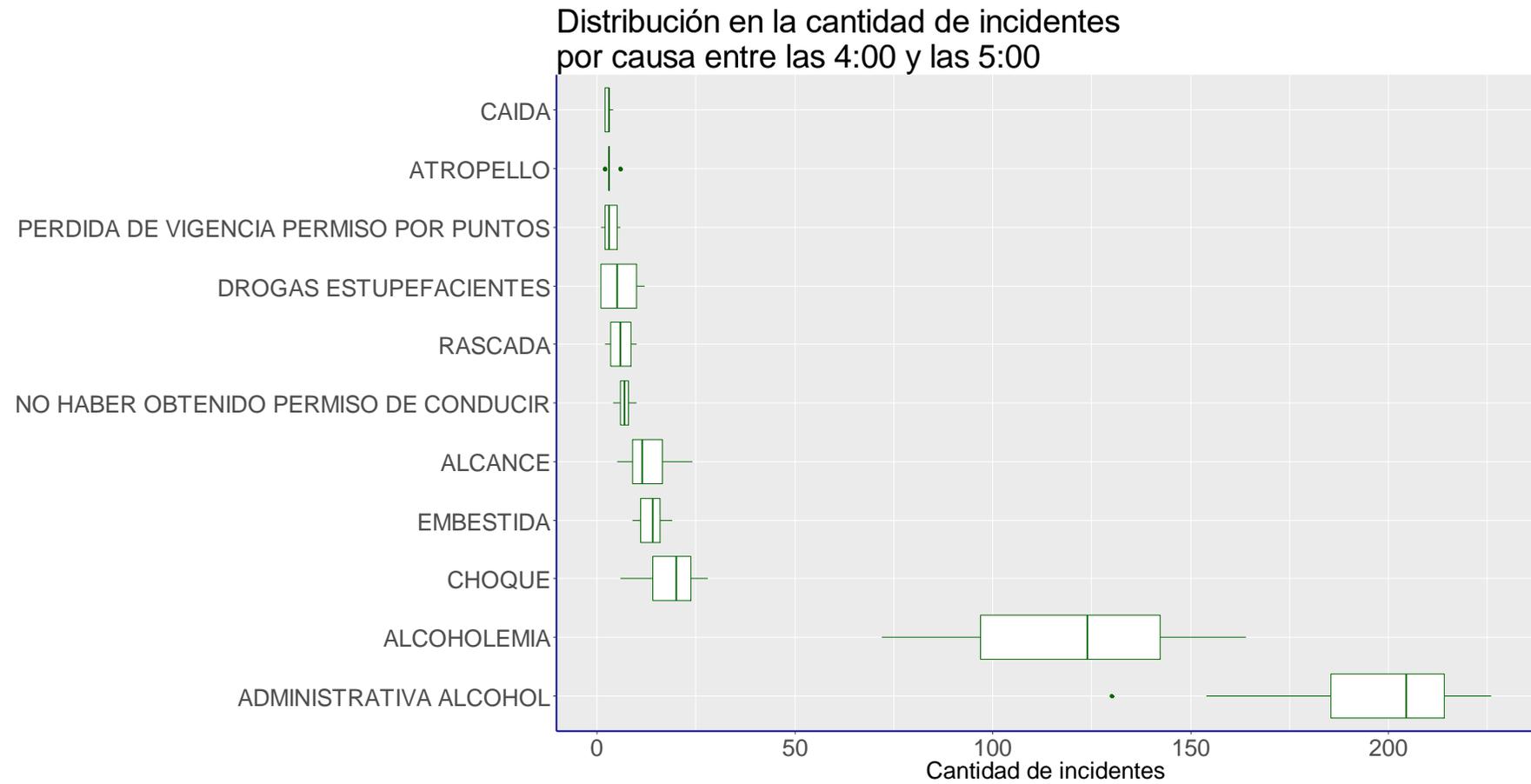
Cantidad de delitos más frecuentes por año



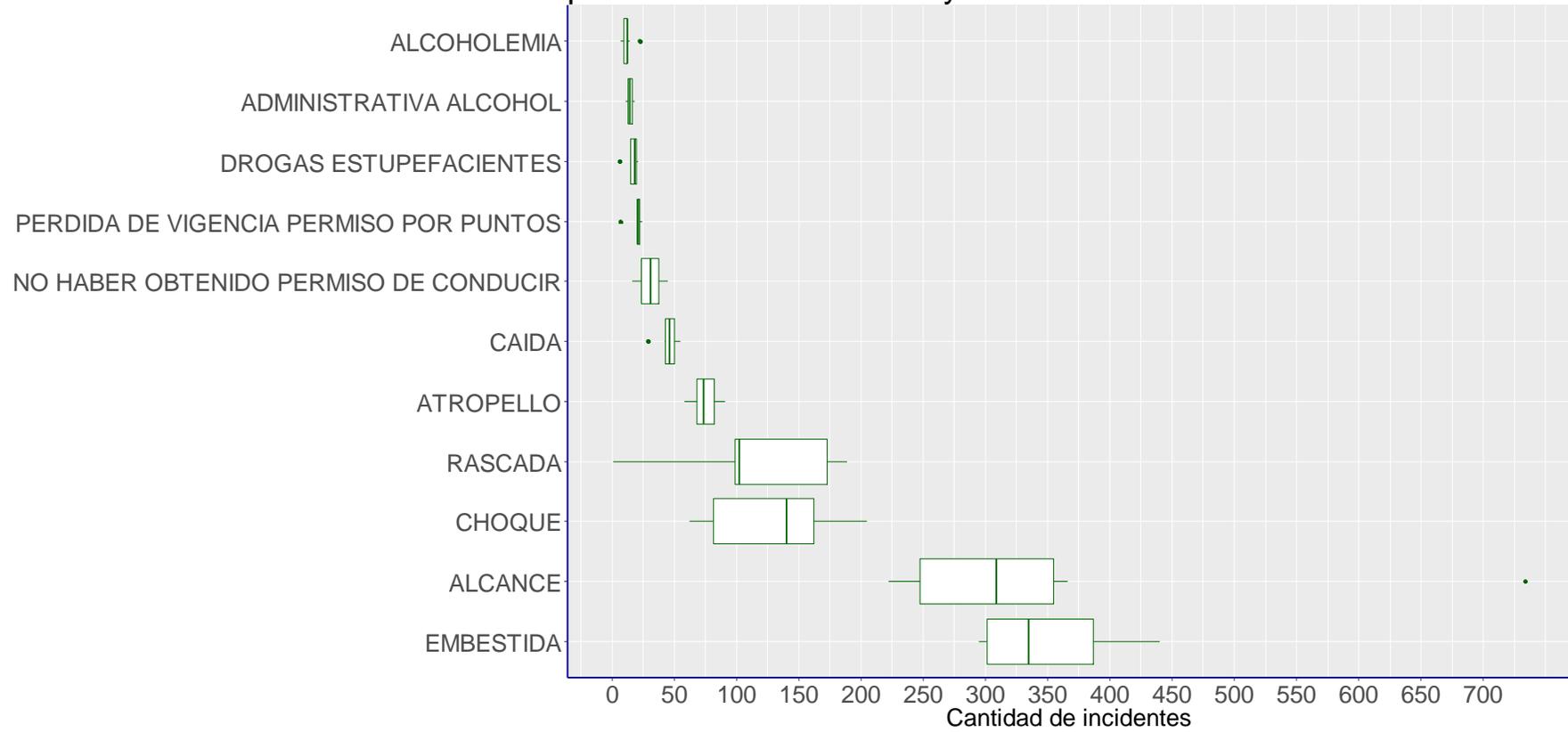
Cantidad de accidentes más frecuentes por año



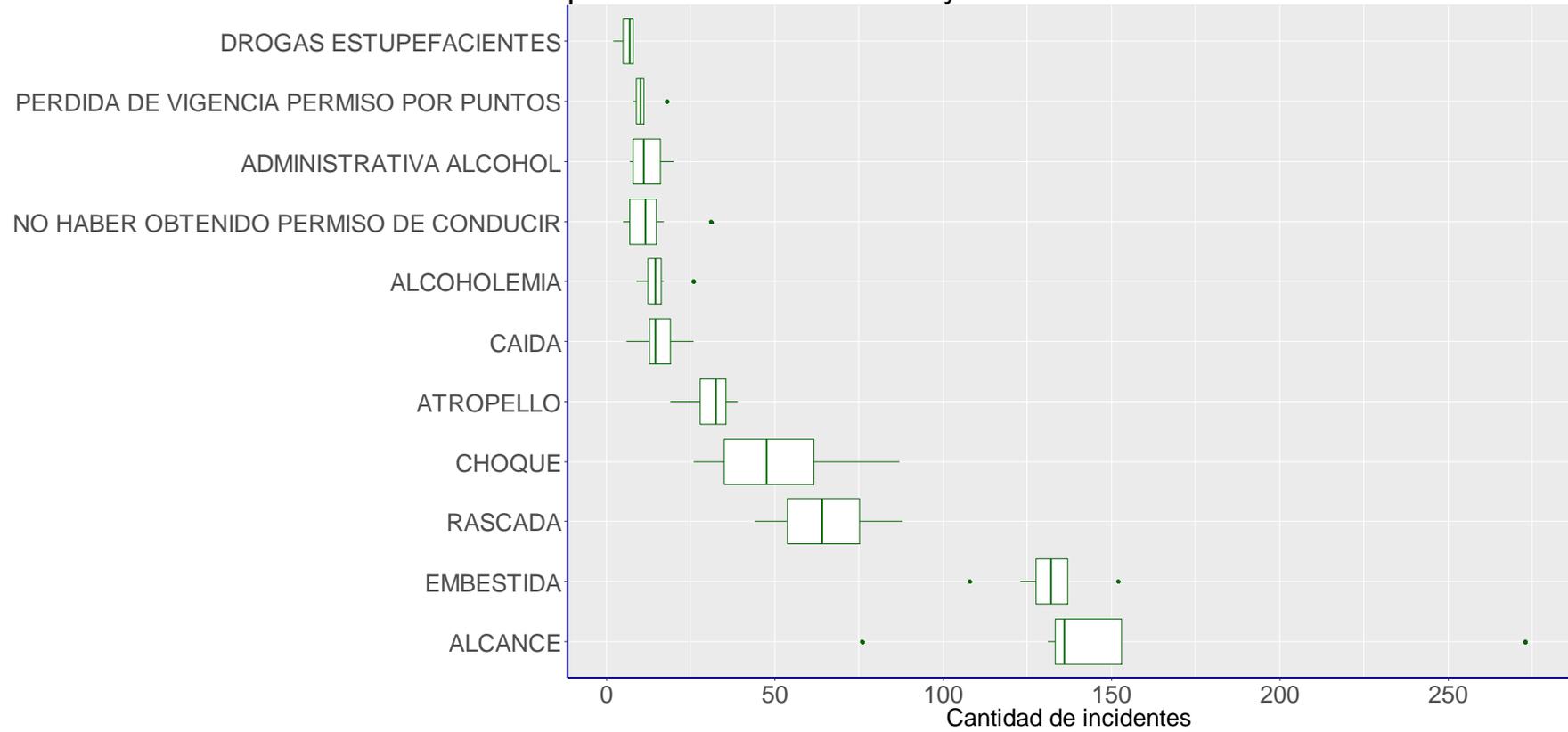
9.5. Gráficas de causas de incidente por intervalo de tiempo



Distribución en la cantidad de incidentes por causa entre las 12:00 y las 14:00

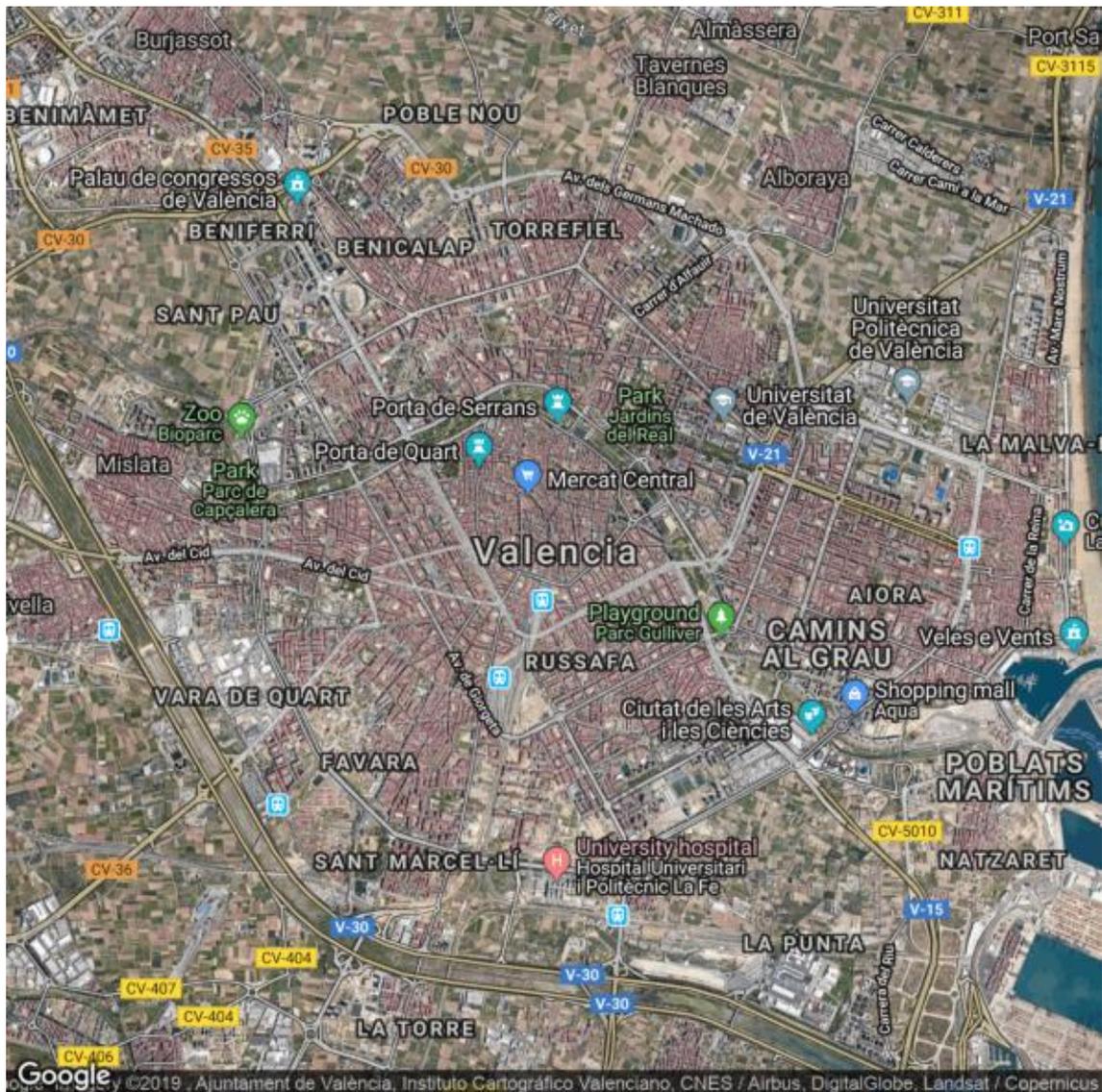


Distribución en la cantidad de incidentes por causa entre las 19:00 y las 20:00

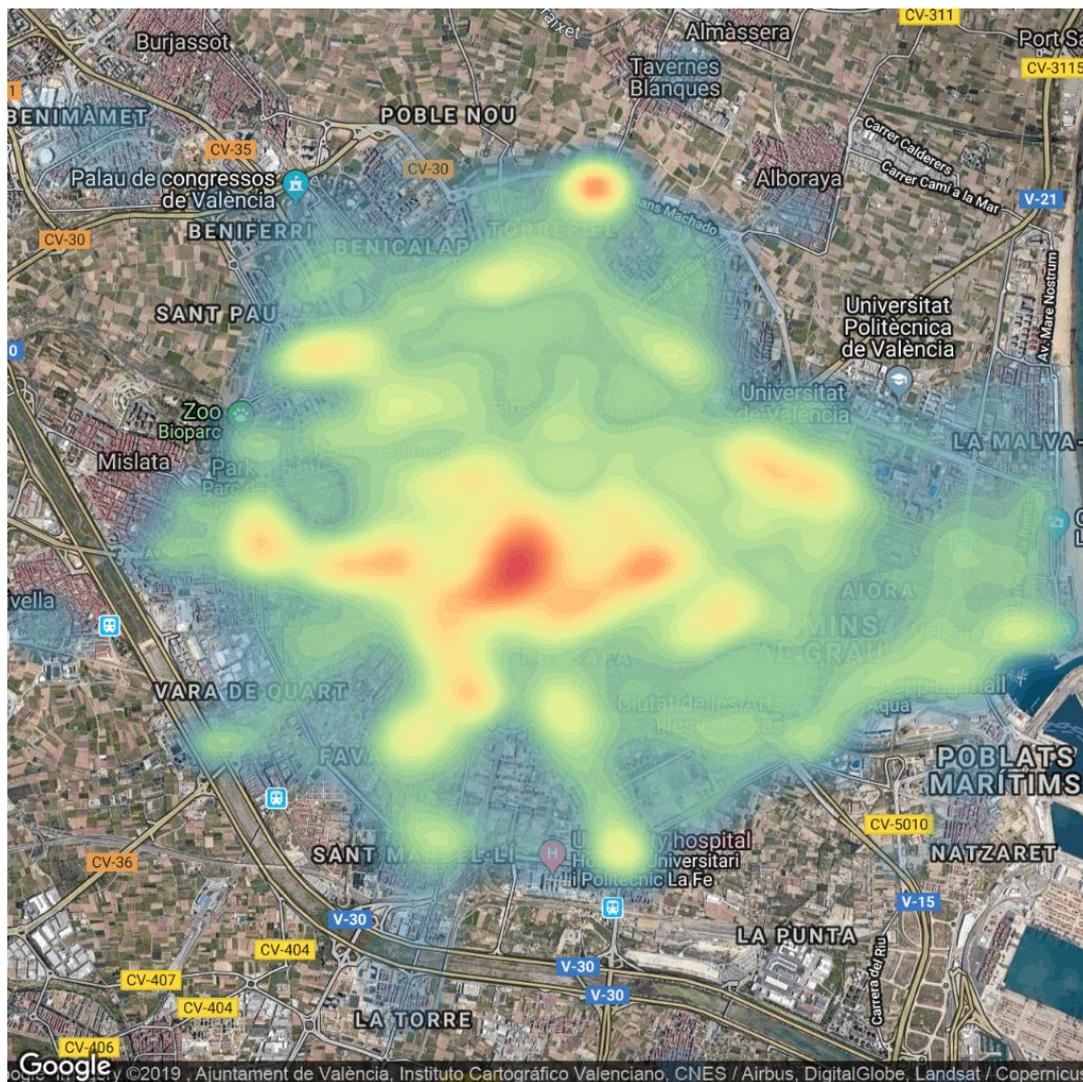


9.6. Mapas

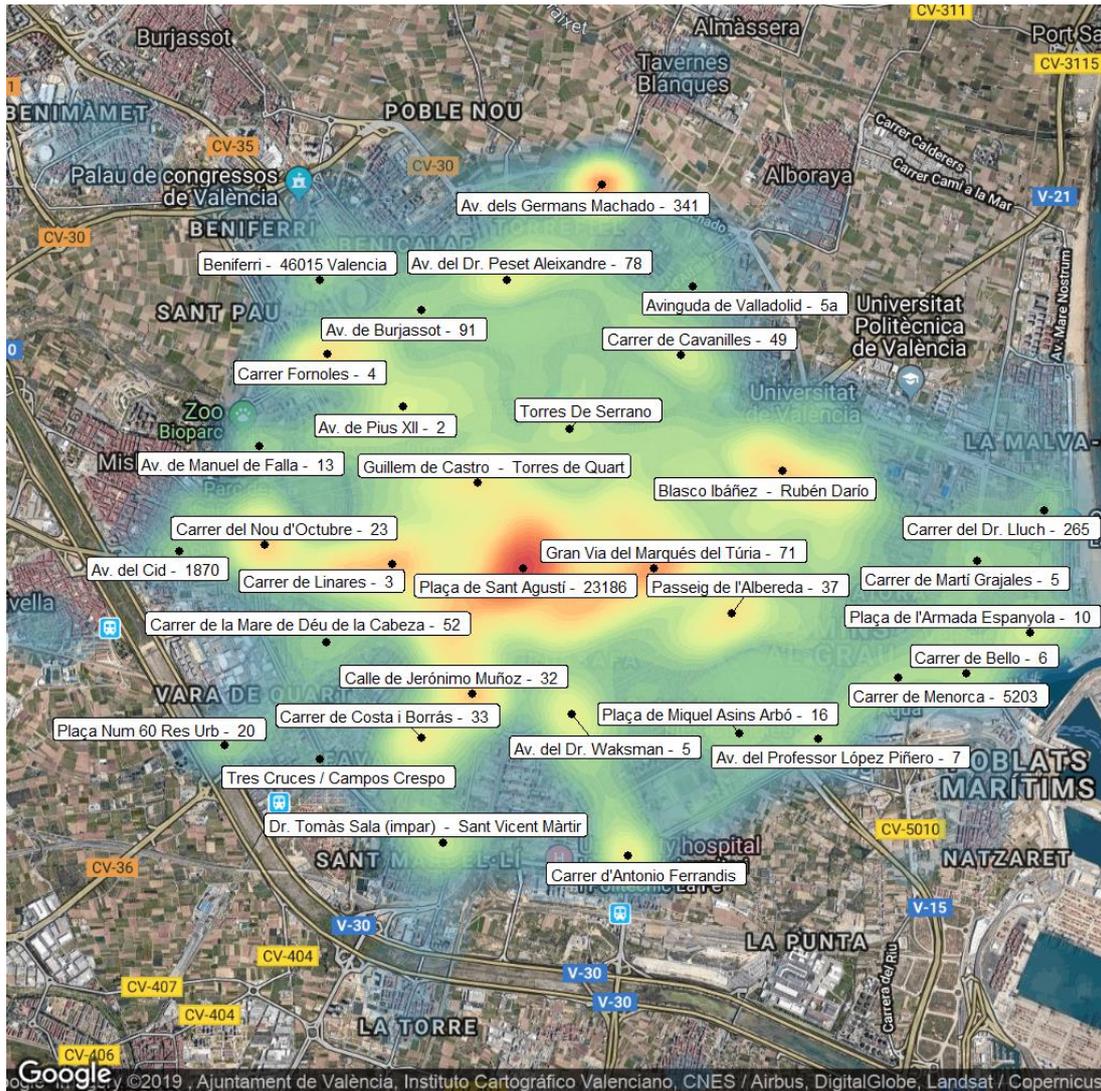
9.6.1. Mapa base de la ciudad de Valencia



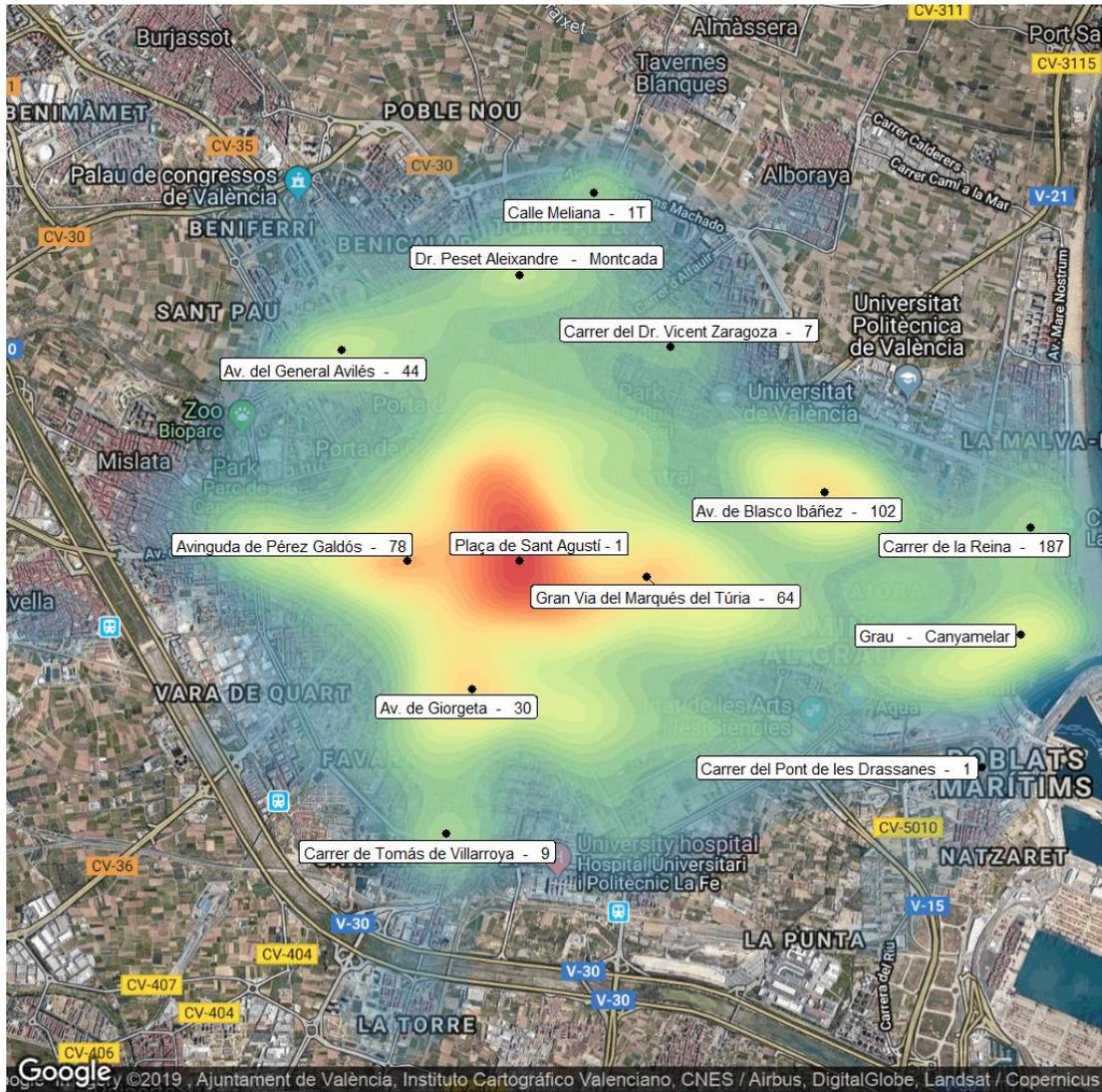
9.6.2. Mapa de densidad de los incidentes sucedidos entre 2010 y 2017



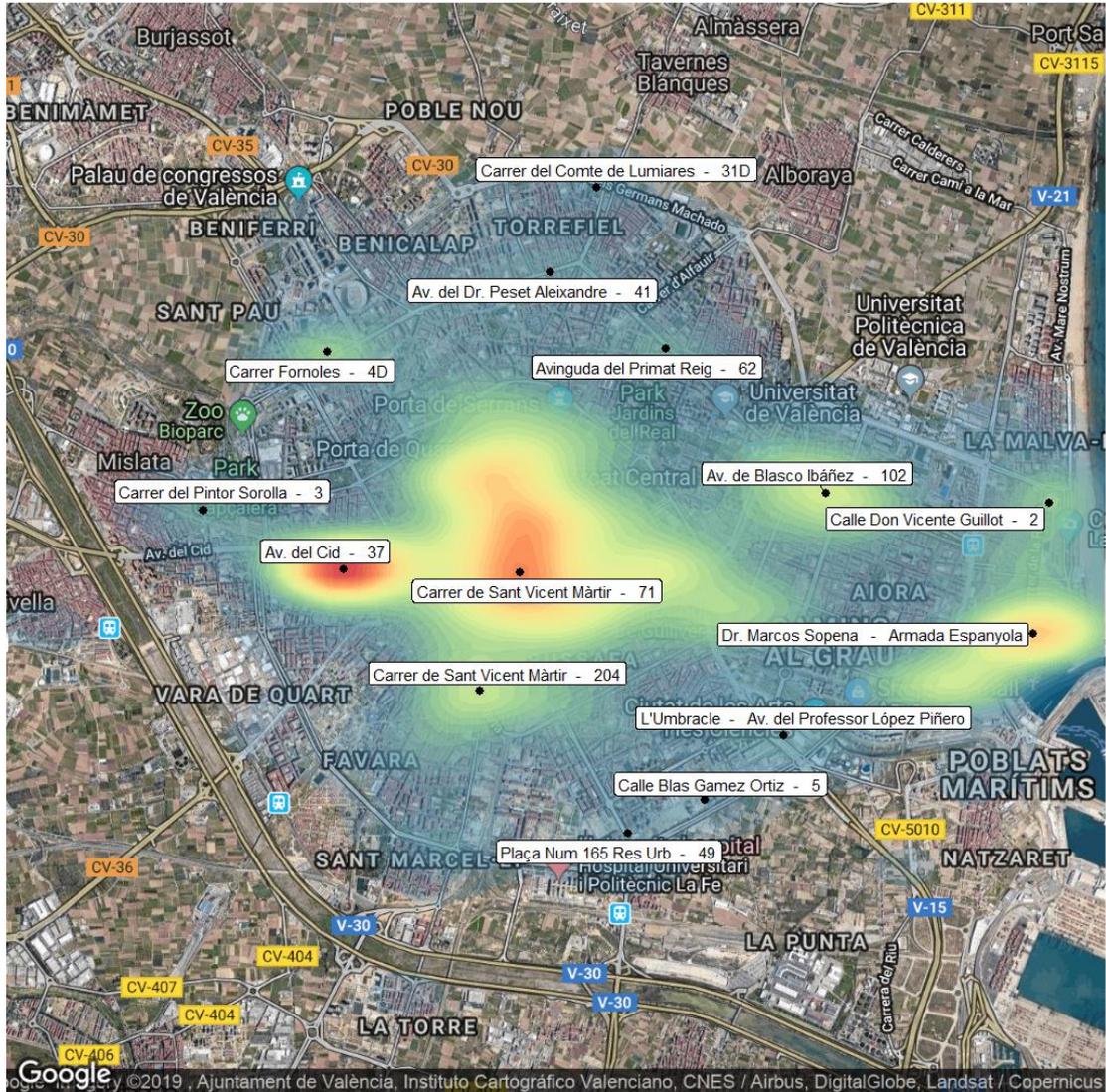
9.6.3. Mapa de densidad de los incidentes sucedidos entre 2010 y 2011 incluyendo las direcciones de los puntos de máxima densidad



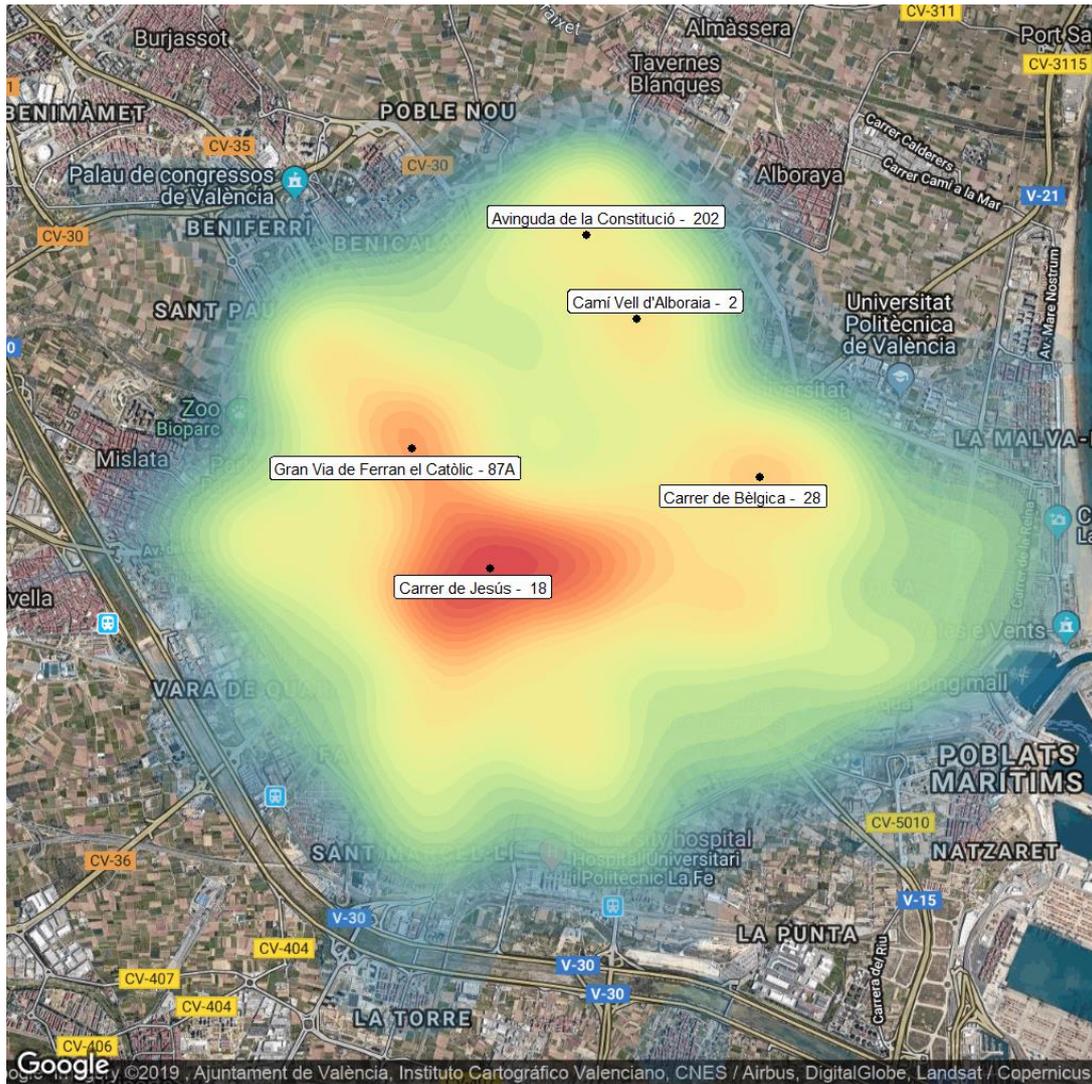
9.6.5. Mapa de densidad y puntos máximos de los delitos por alcoholemia entre 2010 y 2017.



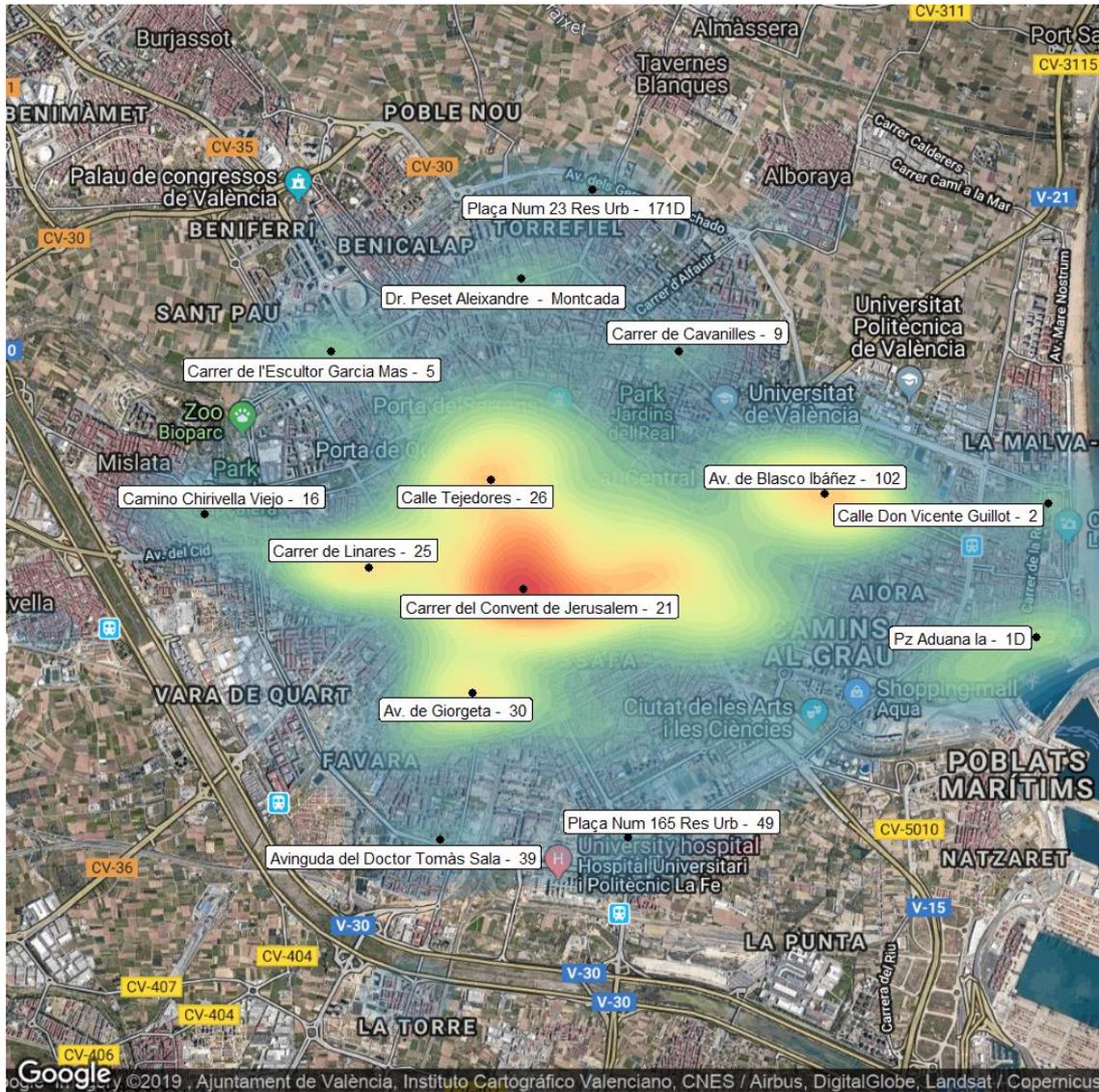
9.6.6. Mapa de densidad y puntos máximos de las sanciones administrativas por alcoholemia entre 2010 y 2017



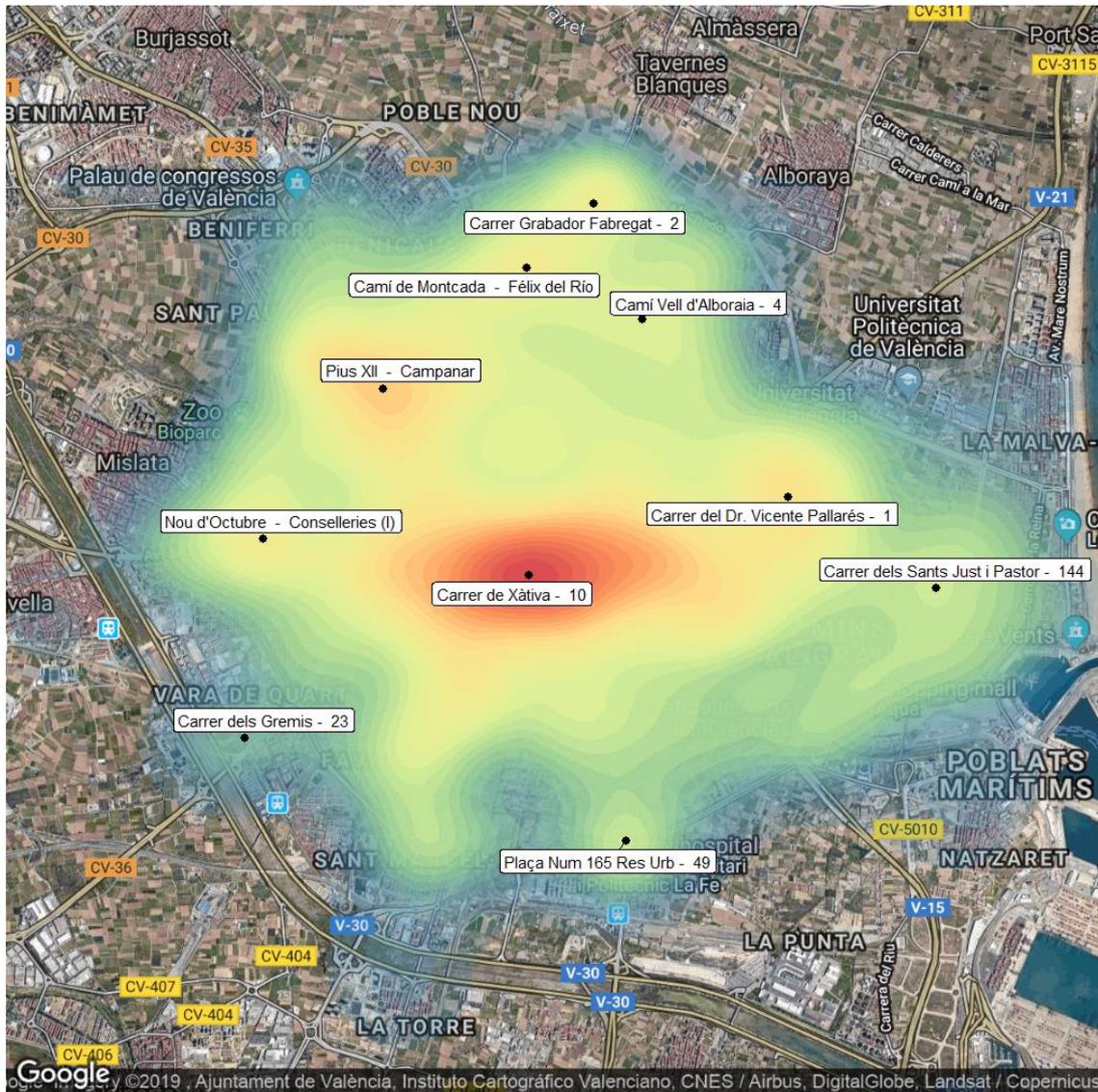
9.6.7. Mapa de densidad y puntos máximos de los accidentes graves entre 2010 y 2017



9.6.8. Mapa de densidad y puntos máximos de los incidentes sucedidos entre las 4:00 y las 5:00



9.6.10. Mapa de densidad y puntos máximos de los incidentes sucedidos entre las 19:00 y las 20:00



9.7. Ejemplos de código

9.7.1. Ejemplo de construcción de un gráfico de líneas

```
x <- ggplot(meses.df, aes(x=Var1, y=Freq, group = year,
  colour=as.factor(year) )) + geom_line(size = 1)+
  scale_color_manual(values=rainbow(8)) + geom_point(size=2) +
  scale_x_discrete(labels=meses) +
  scale_y_continuous(breaks=seq(500,1050,25))+ labs(colour =
  "AÑO", title = "Cantidad de incidentes por mes entre 2010 y
  2017") + xlab("Mes") + ylab("Cantidad de incidentes") +
  theme( axis.line = element_line(colour = "darkblue", size =
  1, linetype = "solid"), axis.text = element_text(size = 14),
  axis.title = element_text(size = 18), title =
  element_text(size = 25))
```

9.7.2. Ejemplo de construcción de un gráfico Box-Plot

```
ggplot(entres, aes(x=Var1, y = Freq))+ geom_boxplot(color="dark gre
  en") + scale_x_discrete( labels=horas) + scale_y_continuous(breaks=
  seq(0,350,25))+ labs(colour = "AÑO", title = "Distribución en la ca
  ntidad de incidentes por mes entre 2010 y 2017") + xlab("Mes") + yl
  ab("Cantidad de incidentes") + theme( axis.line = element_line(coło
  ur = "darkblue", size = 1, linetype = "solid"), axis.text = element
  _text(size = 22), axis.title = element_text(size = 22), title = ele
  ment_text(size = 25))
```

9.7.3. Ejemplo de construcción de un mapa de densidad

```
map + stat_density2d(aes(x = long, y = lat, fill = ..level.., alpha
  =..level..), size = 0.2, bins = 10, data = coordenadas[which(motivo
  == "ALCOHOLEMIA"),], geom = "polygon") + geom_density2d(data =
  coordenadas[which(motivo == "ALCOHOLEMIA"),], aes(x = long, y =
  lat), size = 0.3) + scale_fill_distiller(palette = 'spectral') +
  scale_alpha(range = c(0.2, 1))
```