

Siniestros de tránsito en Uruguay

Bruno Bellagamba y Lucía Quintero

Junio 2018

Introducción

El siguiente trabajo se enmarca dentro de la materia “Nuevas tecnologías para el análisis estadístico de datos” de la carrera Licenciatura en Estadística.

Para el mismo se elige trabajar con los datos de los siniestros de tránsito ocurridos en Uruguay en el período 2013-2017 obtenidos de la página web de la Unidad Nacional de Seguridad Vial (UNASEV) dado que el tema representa una problemática actual para la cual se busca generar información de interés para entender mejor las características de la siniestralidad del país.

El principal objetivo del trabajo es aplicar las técnicas aprendidas en el curso para realizar el análisis exploratorio de los datos seleccionados, lo cuál también implica incluir visualizaciones apropiadas y una buena interpretación de las mismas. Como objetivos adicionales se busca que el trabajo sea completamente reproducible y esté disponible en GitHub para colaboraciones futuras, y poder comunicar los principales resultados del mismo mediante una aplicación shiny.

Para alcanzar los objetivos planteados se utilizan diversas herramientas de R destacándose el uso del paquete *tidyverse*, en particular los paquetes *dplyr* y *ggplot*. También cabe destacar el uso de *rmarkdown* para la realización del presente informe y el paquete *shiny* para realizarla aplicación con los principales resultados.

La metodología para realizar el análisis exploratorio consiste en realizar preguntas de interés y responderlas a través del uso de las herramientas anteriormente mencionadas.

Datos

Como se menciona en la sección anterior, los datos seleccionados corresponden a los siniestros de tránsito ocurridos en Uruguay en el período 2013-2017. Los mismos se obtuvieron de la página web de la UNASEV (*ver enlace en la sección Bibliografía*)

La base cuenta con 149.377 observaciones correspondientes cada una a un siniestro de tránsito, y 10 variables que representan detalles de los mismos como el lugar, gravedad, fecha y hora y tipo de siniestro. Para poder realizar un mejor análisis se agregaron las variables Año y Mes, totalizando 12 variables para trabajar.

A partir de una primera aproximación con los datos se observa que la mayoría de las variables son cualitativas (Calle, Tipo de siniestro, Gravedad, Día de la semana, Departamento, Localidad), y las restantes cuantitativas (Año, Mes, Hora, longitud, latitud). En principio no se detecta la presencia de datos faltantes, pero si de una observación con la descripción “SIN DATOS” en la variable Tipo de siniestro, la cuál se decide sacar.

Para el análisis se decide trabajar con las variables Tipo de siniestro, Gravedad, Día de la semana, Departamento, Hora, Año y Mes.

Análisis Exploratorio

El análisis exploratorio intenta responder diversas preguntas que se generan a partir de la primera visualización de los datos, como por ejemplo:

- ¿Cómo varía la cantidad de siniestros a través de los años del período analizado?
- ¿Hay un “comportamiento” temporal para los siniestros, ya sea mensual, diario o por hora?
- ¿Qué relación hay entre los tipos de siniestros y la gravedad de los mismos?

- A partir de la información por departamento, ¿podemos caracterizarlos?

Comenzamos con un análisis temporal explorando los datos a través de los años y meses, incluyendo también los días de la semana y hora de los siniestros.

En primer lugar nos planteamos cuantificar el número de siniestros en los diferentes años de análisis para luego realizar un gráfico donde pretendemos ver la variación de los mismos.

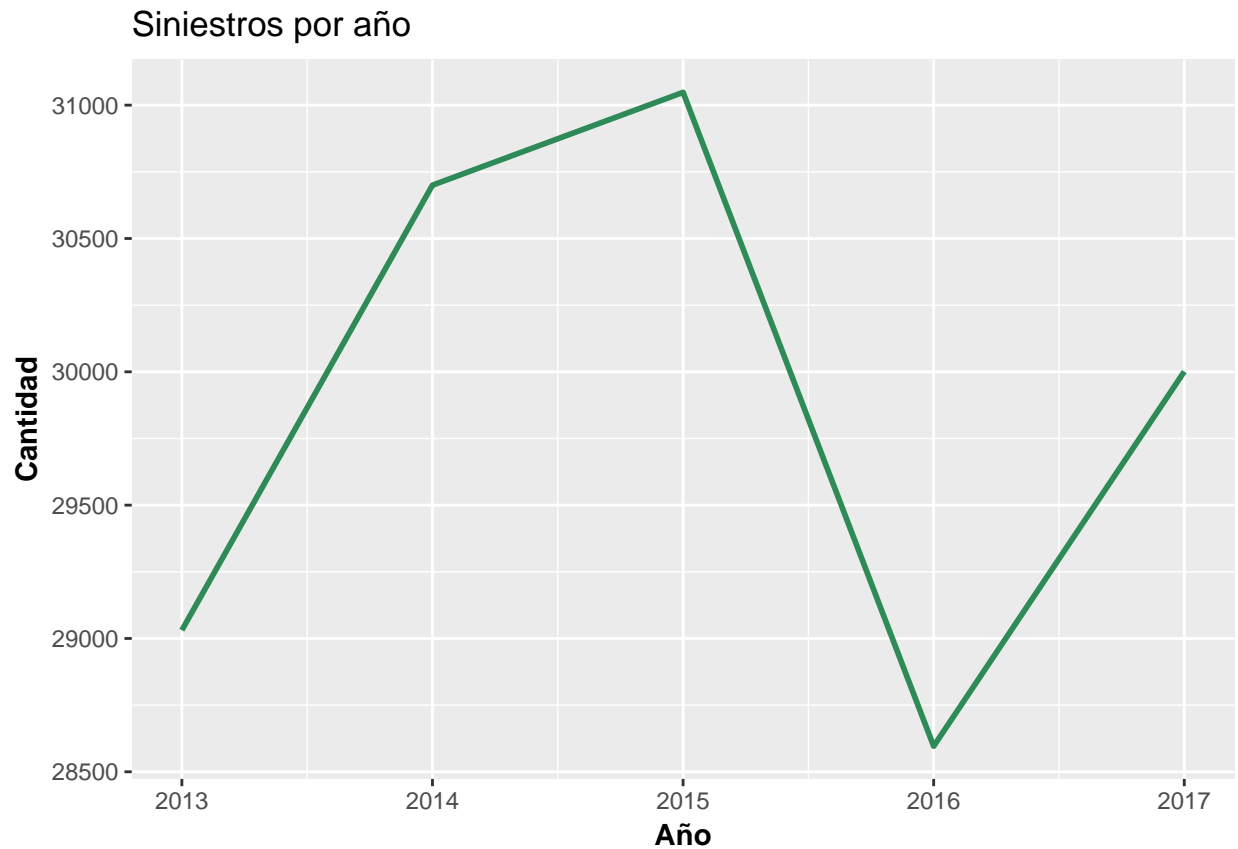


Figure 1: A partir del gráfico podemos observar que del año 2013 al 2015 hubo un incremento sostenido en la cantidad de siniestros habiendo un descenso en el año 2016 para volver a crecer en el 2017, aunque por debajo de los años 2014 y 2015. De todas formas no se observa una variación importante a lo largo del período.

Continuando con la exploración, queremos ver si existe alguna tendencia mensual en la cantidad de siniestros. Para esto realizamos un gráfico de los siniestros por mes para cada año.

También nos interesa ver si hay un comportamiento semanal, para esto en primer lugar vamos a graficar los siniestros por día de la semana para cada año y luego para cada mes (acumulado de los 5 años).

Finalmente, vamos a ver que sucede con la hora de los siniestro graficando la cantidad de siniestros por hora para cada año y luego por día de la semana.

El análisis desde el punto de vista temporal de la cantidad de siniestros nos muestra que el año con mayor cantidad de siniestros fue el 2015 con un total de 31.048, mientras que el 2016 fue el de menor cantidad con 28.596 siniestros, siendo la diferencia entre estos valores 2.452 siniestros lo que representa una variación del 7,89% negativa. Respecto a la información mensual, en todos los años se observan aumentos de los siniestros en los meses de marzo y diciembre, éste último logrando los valores más altos todos los años. En el año 2015 se ve una fuerte disminución a partir de mayo hasta noviembre. La información semanal no resulta concluyente, se observa un aumento los fines de semana pero no pareciera ser significativo. En cambio viendo

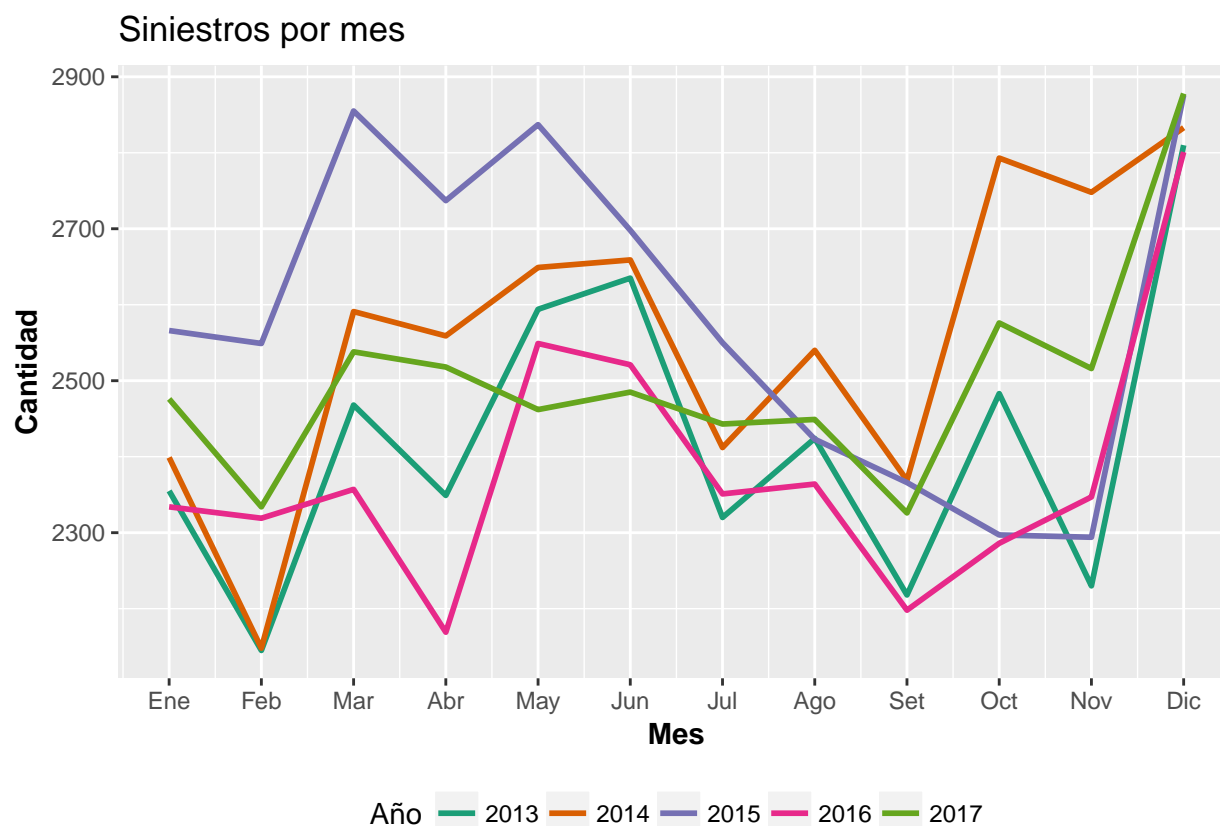


Figure 2: Se puede observar un comportamiento bastante regular en los meses de los diferentes años. Hay dos picos claros en los meses de marzo y diciembre los cuales coinciden con semana de turismo y fin de año respectivamente, algo que es razonable ya que son momentos donde aumenta el movimiento vehicular.

el comportamiento por hora se observa claramente como hay un aumento sostenido en el número de siniestros a partir de las 7-8hs de la mañana para llegar a su punto máximo entre las 17 y 19 horas y luego comenzar a descender. Esto se observa de manera incambiada para todos los años y todos los meses.

Luego de explorar temporalmente nuestro problema de interés pasamos a realizar el análisis a través de las variables Tipo de siniestro y Gravedad, dos variables que a priori parecen ayudar en la caracterización del problema de estudio. temporal

Para comenzar, cuantificamos los siniestros para cada año según su clasificación para poder visualizar los principales motivos de éstos.

De manera análoga, cuantificamos para cada año la gravedad de los siniestros ocurridos.

Para darle más profundidad al análisis, realizamos una visualización de como estas dos variables se distribuyen conjuntamente en cada año para ver que relación existe entre el motivo del accidente y su consecuencia para los involucrados.

A modo de resumen podemos decir que los resultados del análisis de estas variables en el tiempo son constantes, siendo la colisión entre vehículos la principal causa de los siniestros en Uruguay y siendo la gran mayoría con consecuencias leves o sin consecuencias para los involucrados.

Por último buscamos darle al análisis un enfoque territorial por lo que pasamos a visualizar las variables ya vistas por departamento.

Para empezar vamos a cuantificar la cantidad de siniestros por departamento en los distintos años.

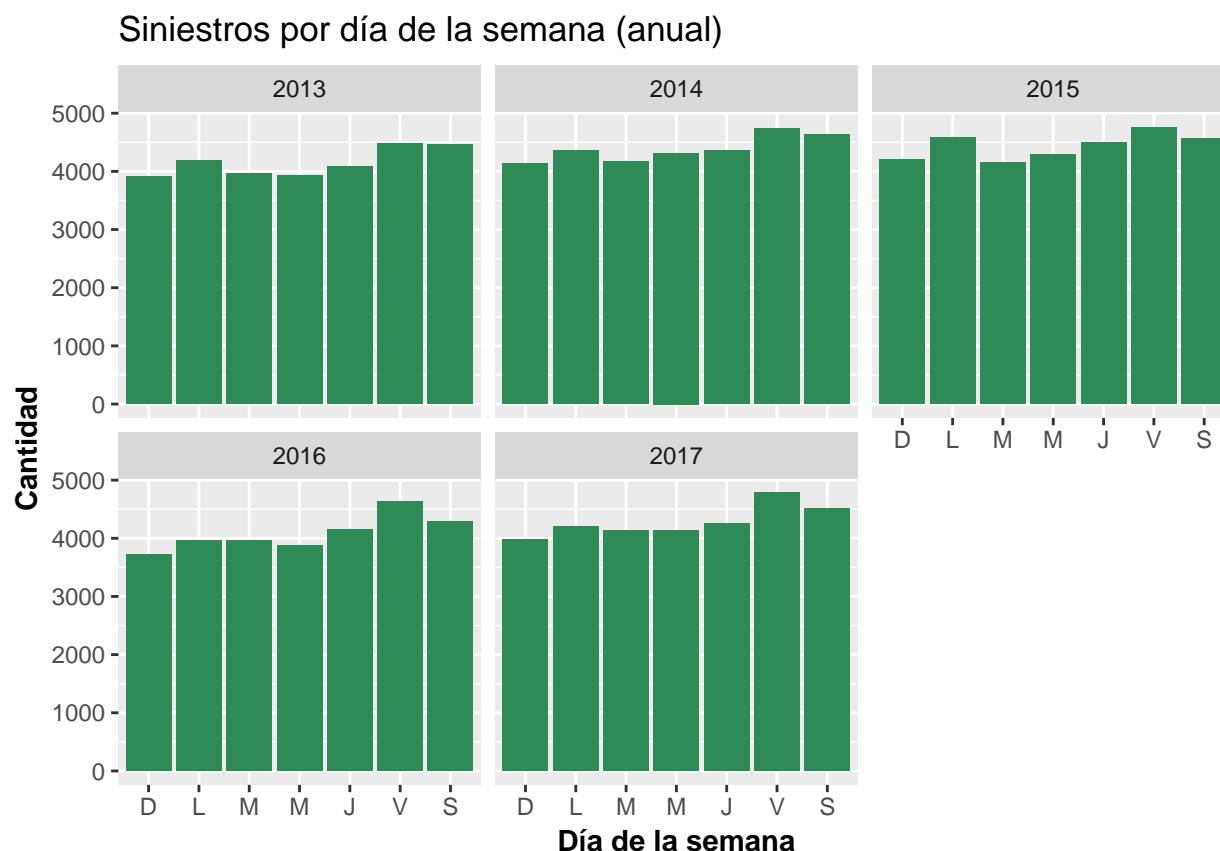


Figure 3: Se observa un incremento en los siniestros los días lunes, viernes y sábados siendo estables el resto de los días. Esto parece concordar con el mayor movimiento que se da los fines de semana.

Continuamos viendo qué relación hay entre el tipo de siniestro y los departamentos.

Para finalizar vamos a observar qué relación existe entre la gravedad de los siniestros y el departamento.

De éste último análisis podemos destacar que la principal causa de los siniestros para todo el país se mantiene así para cada departamento como los resultados para los involucrados en términos de gravedad. Más allá de que Montevideo es el departamento que registra mayor número de atropello de peatones y este es el motivo con mayor consecuencia fatal vemos que San José es el departamento con mayor número de siniestros fatales.

Aplicación Shiny

El principal objetivo de la realización de la aplicación Shiny es tener un primer acercamiento a esta herramienta para mostrar los resultados del presente trabajo. En la misma se incluye un resumen del problema y su contexto, una descripción del contenido y tres paneles con distintas visualizaciones que intentan ser una guía para la comprensión de los resultados del trabajo intentando contemplar todos los puntos del análisis realizado (plano temporal, variables de interés y plano territorial). La interactividad con el usuario está dada en la posibilidad de poder seleccionar las distintas variables en los distintos años del análisis y poder visualizarlas de tres maneras diferentes, en gráficos individuales, combinados y un gráfico mapa.

Mostramos una imagen de la aplicación:

Siniestros por día de la semana (mensual)

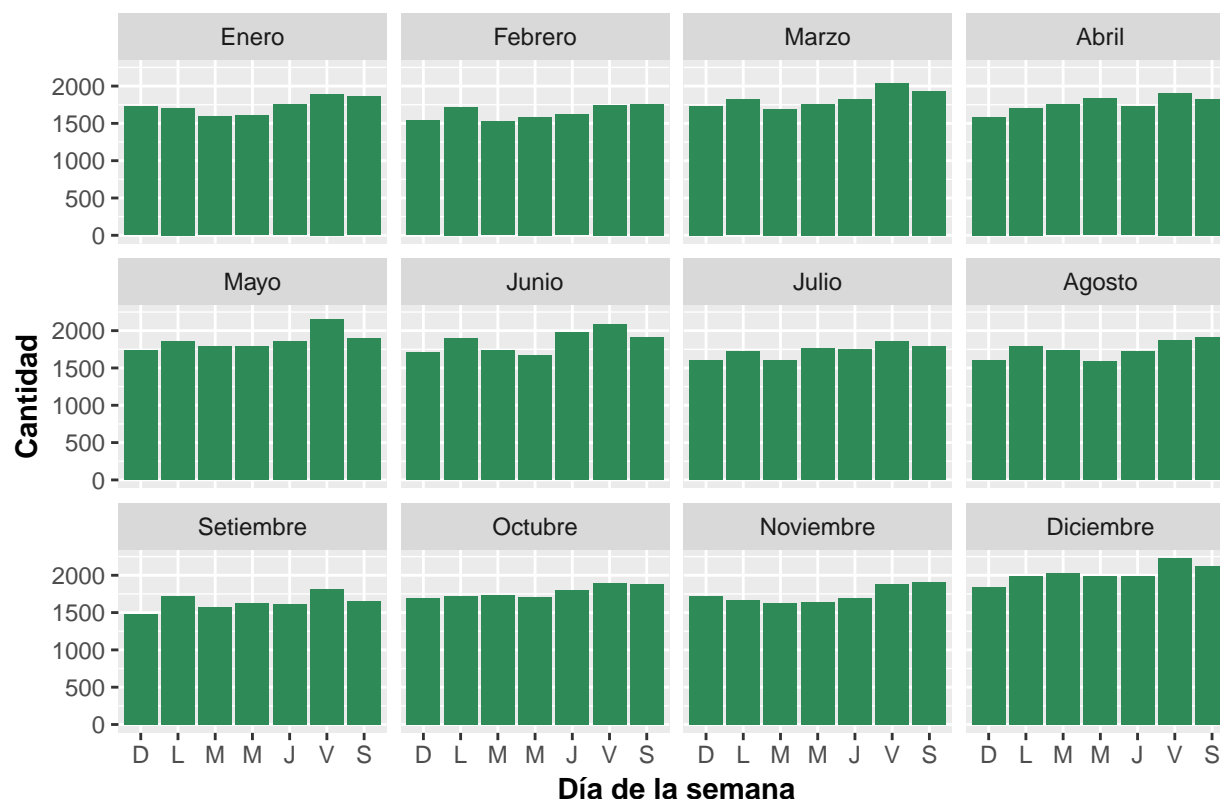


Figure 4: En el gráfico por mes vemos el mismo comportamiento que en el gráfico anual, los lunes, viernes y sábados se ve un aumento en los siniestros. También visualizamos el aumento que hay en diciembre en los siniestros en general.

Comentarios finales

Presentamos las principales conclusiones obtenidas del análisis exploratorio realizado:

- El efecto anual en la cantidad de siniestros no parece aplicarse, pero si se observa una variación en la gravedad de los accidentes a lo largo de los años, presentando una mejoría (menos accidentes graves y más sin lesionados).
- La distribución semanal no tiene un efecto en la cantidad de accidentes pero la mensual si la tiene, dándose la mayor cantidad de accidentes los meses de marzo y diciembre.
- También hay una distribución interesante de los accidentes por hora, observándose el pico entre las horas 17 y 19 independientemente del día de la semana, mes o año.
- Geográficamente, si bien es evidente que en Montevideo se dieron la mayor cantidad de accidentes, el mayor ratio de siniestros cada 100 habitantes se da en Paysandú.
- También se pueden observar las diferencias en el ratio de accidentes según el tipo de siniestro y su gravedad destacando por ejemplo que San José tiene el mayor ratio de accidentes fatales en el país o que en Paysandú se producen la mayor cantidad de colisiones entre vehículos por cada 100 habitantes.

Para finalizar, queremos señalar posibles agregados y continuaciones al presente trabajo:

- Modelar y predecir los siniestros de años futuros utilizando modelización a partir de las series temporales para luego contrastar y evaluar los modelos propuestos con los resultados publicados a futuro.
- Agregar una distribución geográfica más precisa al análisis, llegando al punto de identificación exacta de donde ocurrió el accidente basándose en coordenadas y las calles de ocurrencia.

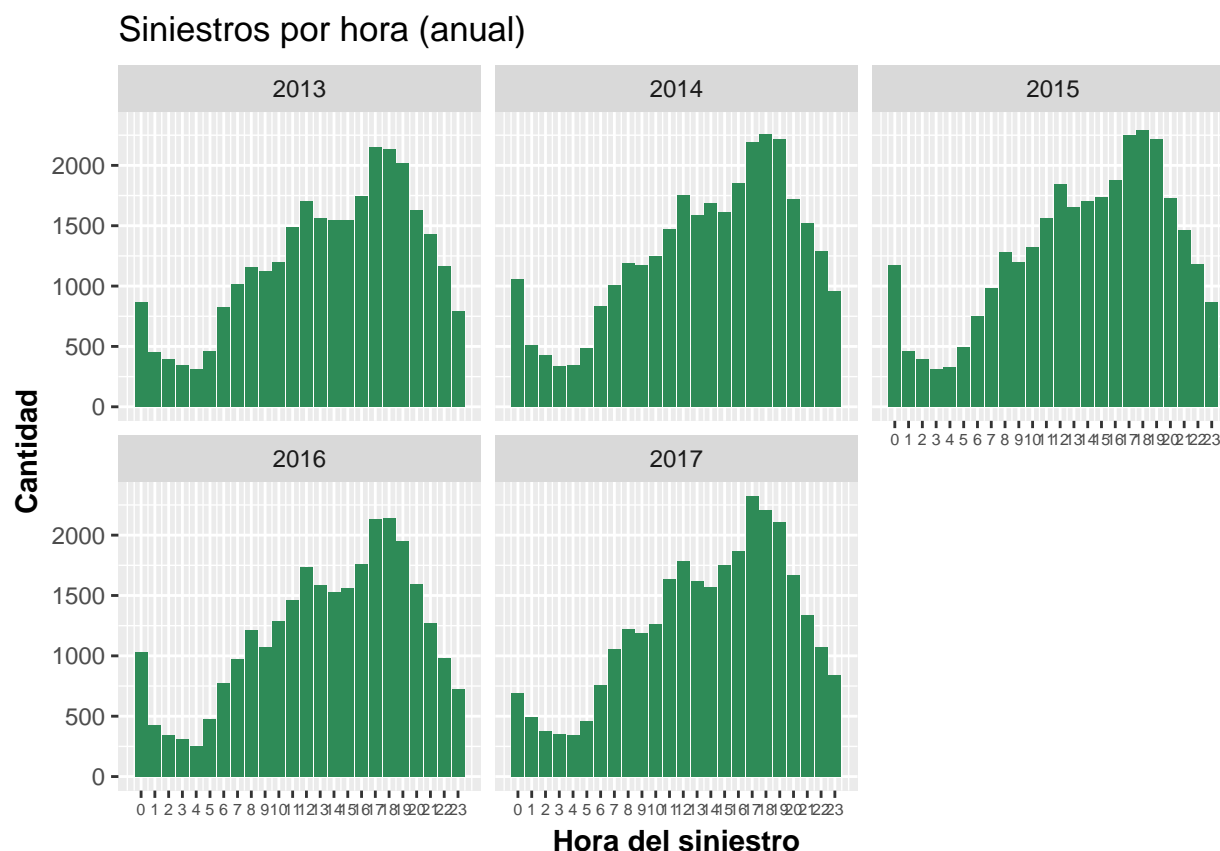


Figure 5: A partir de los gráficos se observa que el comportamiento es muy similar en todos los años siendo el horario de 17 a 19 horas el que presenta mayor cantidad de siniestros.

Apartado

Nos parece pertinente mencionar ciertos puntos que sabemos se deben mejorar para trabajos futuros:

- Solucionar el problema que ambos tenemos para compilar de forma correcta a PDF.
- Mejorar el tamaño de los gráficos y los captions correspondientes, consideramos que deben ser de menor tamaño.
- Lograr que todas las etiquetas de los gráficos sean las pertinentes (gráfico 6 “Siniestros por hora (semanal)”).
- Agregar una breve descripción (caption) a los gráficos en la aplicación Shiny.

Bibliografía

- R for Data Science - Garrett Golemund, Hadley Wickham (<http://r4ds.had.co.nz/>)
- ggplot2 Elegant Graphics for Data Analysis - Hadley Wickham
- Datos: Página web UNASEV (http://aplicaciones.unasev.gub.uy/mapas/AdministracionMapaIndicadores/AdministracionMapaIndicadores_Alta)

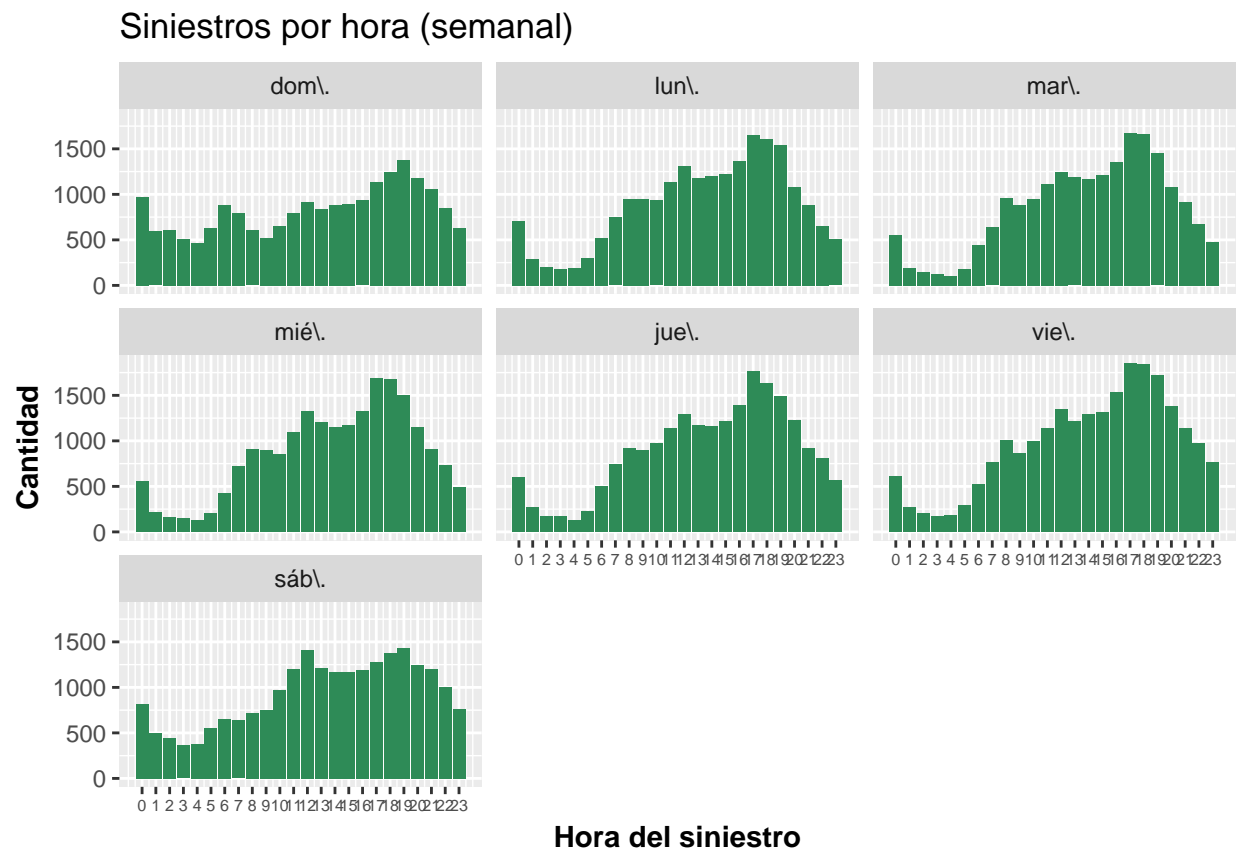


Figure 6: Observamos que el comportamiento es similar en todos los días de la semana viendo al igual que en el gráfico por años que el horarios de 17 a 19 horas es donde ocurrieron la mayor cantidad de siniestros.

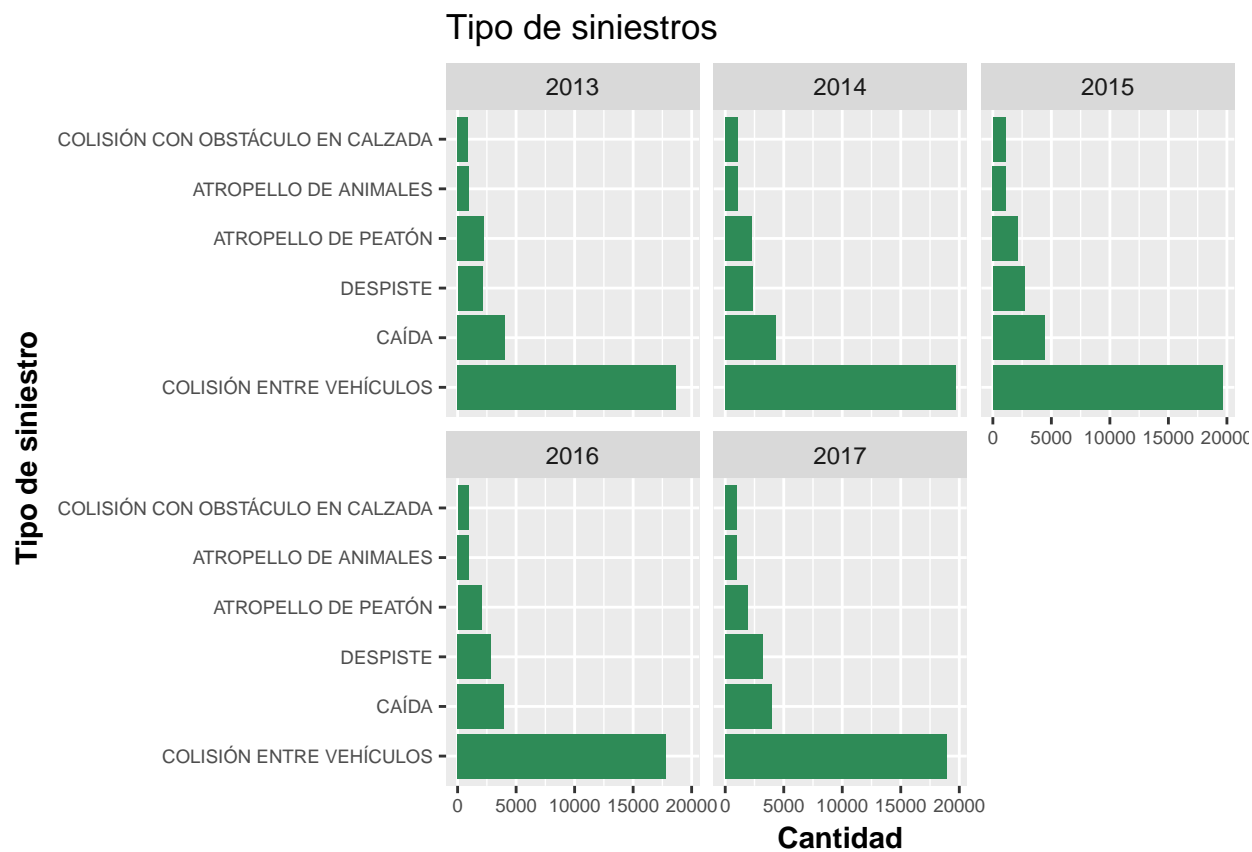


Figure 7: Se observa que la colisión entre vehiculos es la principal causa de los siniestros siguiendo por las caídas y despiestes en menor medida. No se observa una diferencia significativa en las cantidades de siniestros de cada tipo en los diferentes años,

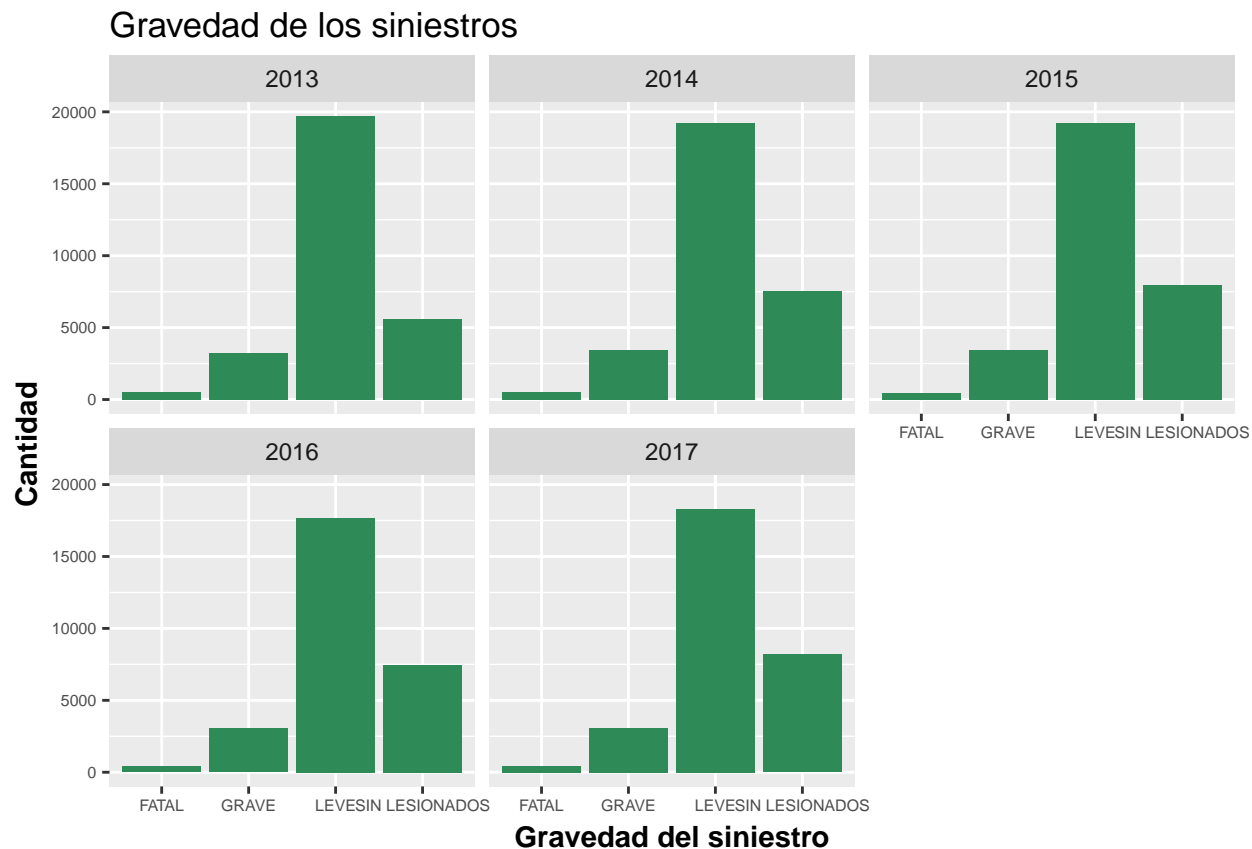


Figure 8: Se puede observar que a lo largo de los años, las cantidades de siniestros fatales y graves se han mantenido casi constantes mientras que los siniestros con gravedad leve han disminuído y aumentado el número de accidentes sin lesionados aunque levemente. También destacamos que la mayoría de los siniestros tienen consecuencias leves o sin lesionados.

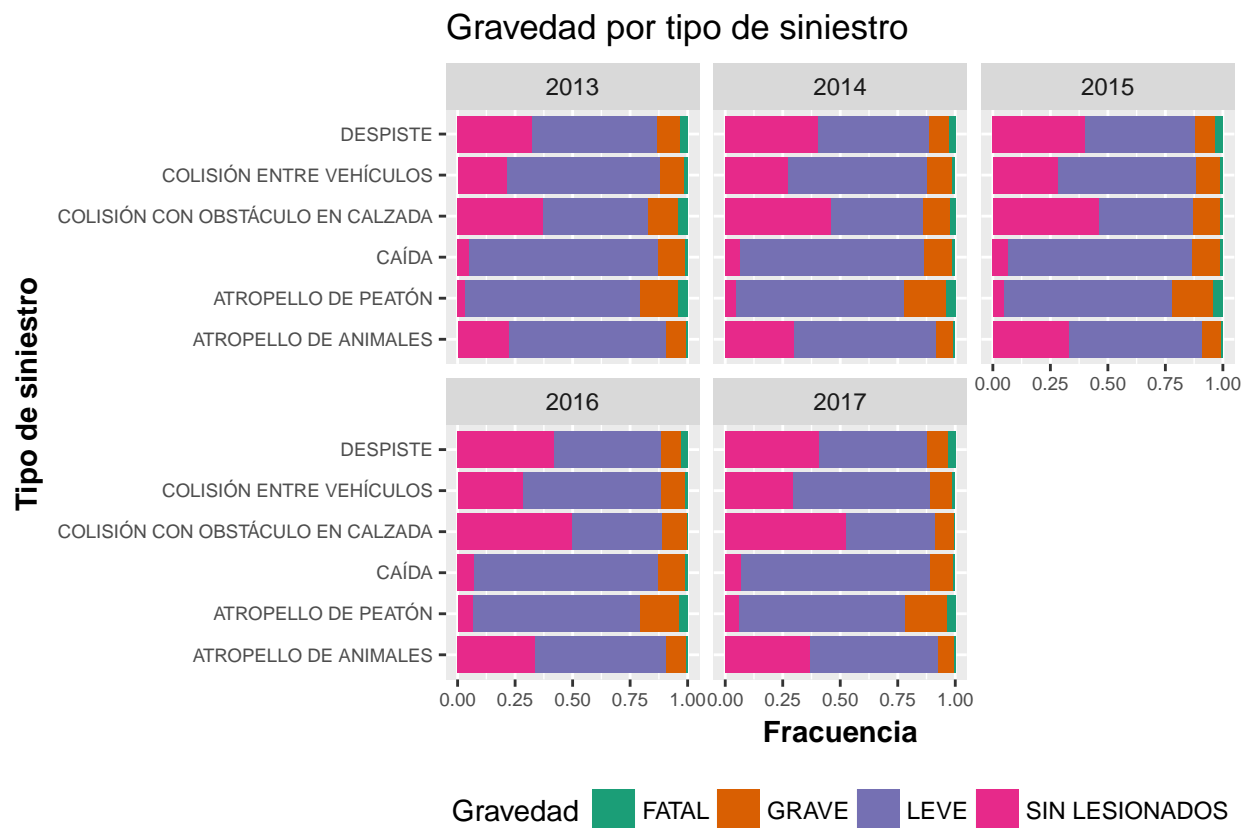


Figure 9: Como era esperable, los accidentes con menor proporción de no lesionados son los correspondientes a los atropellos de peatón siendo estos los tipos de siniestros con más resultados fatales. Algo quizás no tan esperado es que las colisiones con obstáculos tengan la mayor proporción de resultados sin lesionados.

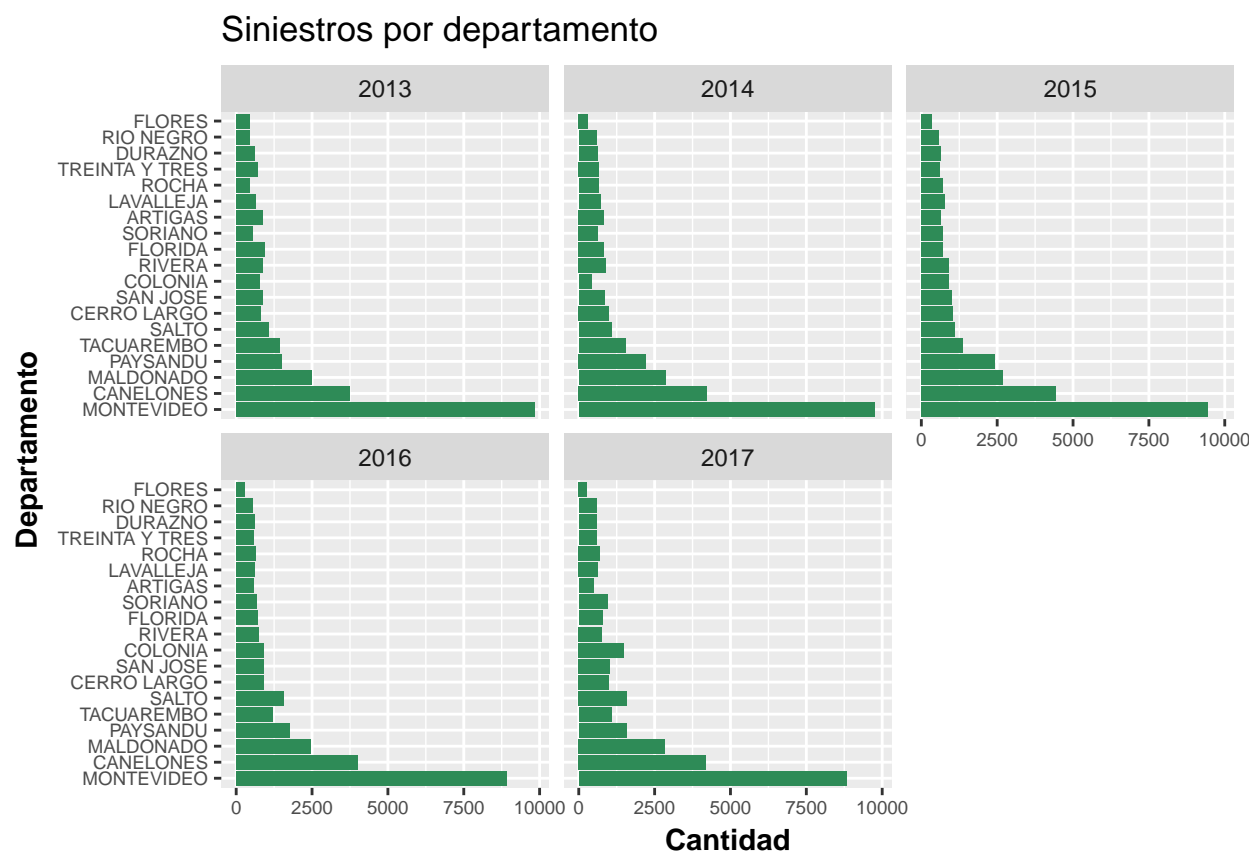


Figure 10: Se puede observar que Montevideo es claramente el departamento con mayor cantidad de siniestros, un correcto análisis de esta situación sería tomar en cuenta la población de cada departamento. Vemos que no hay grandes cambios en el correr de los años.

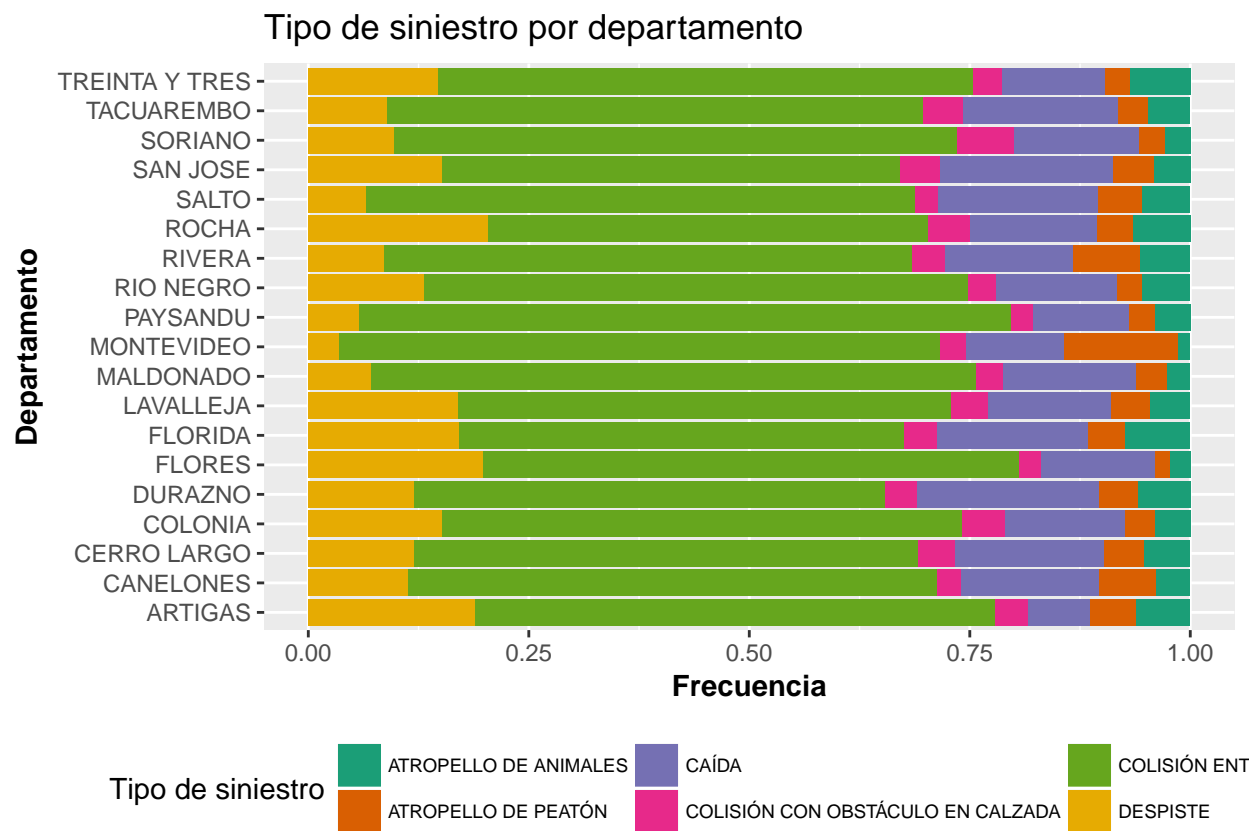


Figure 11: Se puede ver que en todos los departamentos la mayoría de los siniestros son causados por la colisión entre vehículos. También se observa que Montevideo es el que mayor número de atropello de peatones tiene.

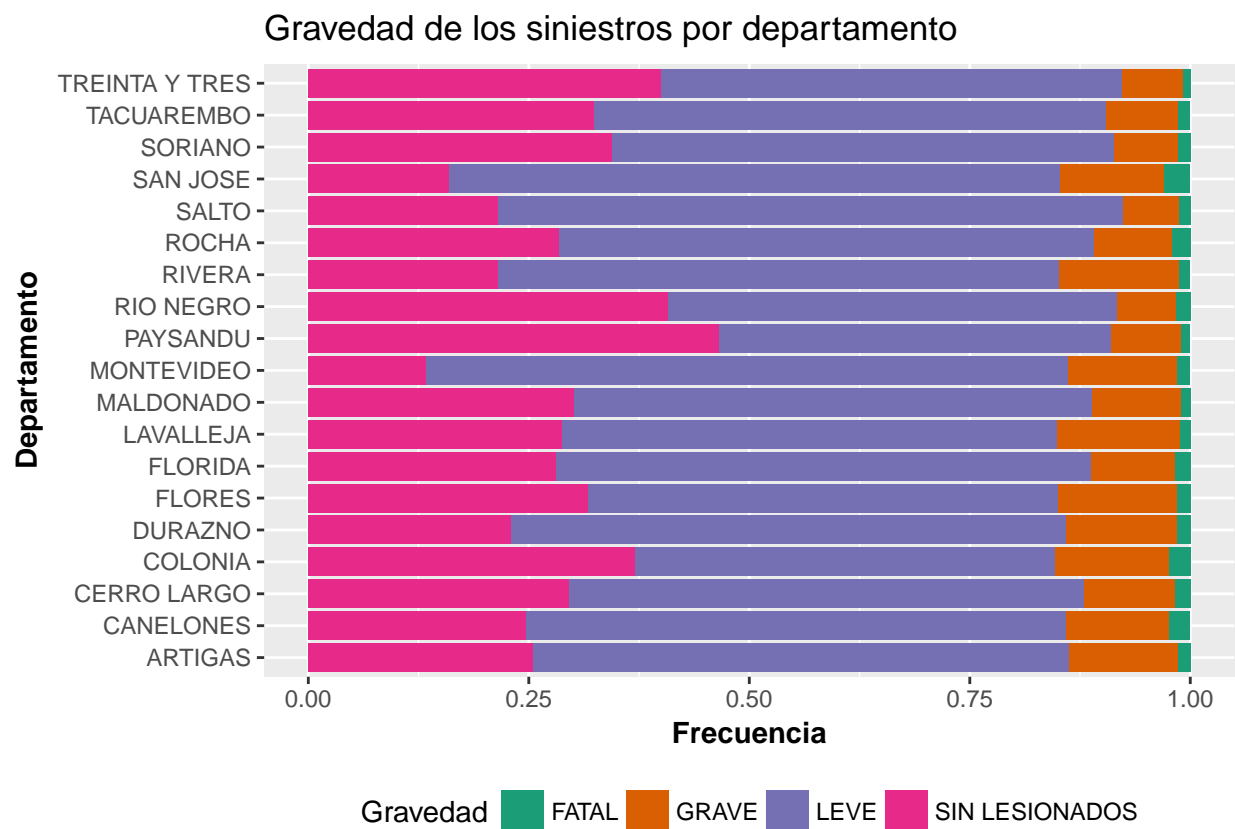


Figure 12: Claramente la mayoría de los siniestros para todos los departamentos son sin lesionados o lesionados leves. San José es quien tiene la mayor cantidad de siniestros fatales.

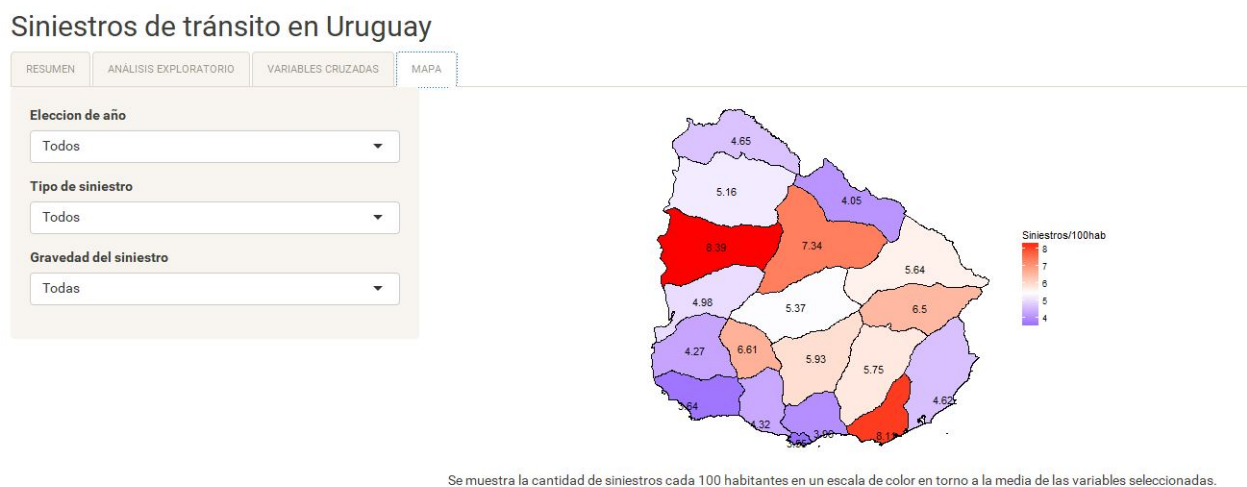


Figure 13: