

Proyecto 1  
Visualización de Información  
Prof. Armando Arce Orozco  
Primer Semestre 2024  
Hecho por:  
Isaac Cordero Tapia c2022193583  
Santiago Moreno Granados c2022040858

## Contenido

Introducción.....	2
<b>Proyecto 1</b> .....	2
.....	2
<b>Análisis exploratorio de datos</b> .....	3
<b>Tipos de gráficas</b> .....	3
<b>Generación de página Web</b> .....	4
<b>Consideraciones generales</b> .....	4
Gráficas Unidimensionales .....	6
Gráfica Bidimensional .....	7
Gráfica de Facetas .....	7
Imagen Compuestada .....	8
Conclusiones .....	8

## Introducción

En este proyecto se va a realizar una observación detallada de una serie de datos acerca de los accidentes de tránsito registrados en Costa Rica esto con el fin de poder encontrar patrones y distintos escenarios que nos pueden arrojar los datos al analizarlos con una herramienta como lo es R, esto con el fin de poder representar de una manera más fácil de comprender todos los datos recolectados, ya que manejar grandes cantidades de datos sin que exista alguna gráfica es extremadamente complicado.

## Proyecto 1

El propósito de este proyecto es realizar un análisis exploratorio (EDA) de un conjunto de datos de interés. Para ello se deberá publicar una página Web en donde se describa el EDA realizado, incorporando una serie de visualizaciones que

describen el comportamiento de los datos y los patrones identificados en los mismos.

## **Análisis exploratorio de datos**

El análisis exploratorio de datos (EDA), que a menudo hace uso de técnicas de visualización de datos, es una herramienta utilizada por los científicos de datos para examinar, evaluar y resumir grandes conjuntos de datos. Facilita a los científicos de datos la búsqueda de patrones, la identificación de anomalías, la comprobación de hipótesis o la validación de suposiciones, ayudándoles a gestionar de forma óptima las fuentes de datos para obtener las respuestas necesarias.

EDA ofrece un mejor conocimiento de las variables del conjunto de datos y de las interacciones entre ellas. Se utiliza sobre todo para ver qué pueden revelar los datos más allá del trabajo formal de modelización o comprobación de hipótesis. También puede ayudar a determinar la idoneidad de los métodos estadísticos que se piensan utilizar para el análisis de datos. Los enfoques EDA fueron creados por primera vez en la década de 1970 por el matemático estadounidense John Tukey y siguen siendo un enfoque popular en el proceso de descubrimiento de datos.

## **Tipos de gráficas**

- Gráfica unidimensionales (una sola variable): se deben crear tres o más gráficas en las que se muestre la distribución de los datos de una sola variable. Se debe describir en el documento dicha gráfica, así como cualquier tendencia detectada en la misma. Nótese que la idea no es mostrar la

distribución de cualquier variable, sino escoger aquellas que presentan un patrón interesante.

- Gráfica bidimensionales (dos variables): se deben crear dos o más gráficas en las que se muestre la combinación de dos variables. Se deben escoger aquellas combinaciones de variables en las que se detecte algún interrelación entre las variables. Igualmente se debe describir en el documento dicha gráfica, así como cualquier tendencia detectada en la misma.
- Gráfica multidimensional: Se deberá elaborar al menos un tipo de gráfica multidimensional en donde se incorporen los valores de al menos 5 variables al mismo tiempo. Dichas variables deben ser seleccionados cuidadosamente de forma que sean fáciles de observar las tendencias o casos particulares de los datos.
- Facetas: Se debe incorporar al menos una gráfica de facetas en el análisis. Dicha gráfica podrá involucrar tres o más variables que presenten algún comportamiento interesante.
- Imagen compuestada: Adicionalmente se debe incluir en el informe al menos una gráfica compuesta de otras gráficas.
- Interacción: Todas las gráficas que se presenten en la página Web deberán contar con capacidades de interacción. Para ello se pueden utilizar las librerías: Plotly, Giraffe, Bokeh, etc.

## Generación de página Web

Para generar la página Web se utilizará un *notebook* de R (rnotebook). Note que se deben ocultar los segmentos de código en la página generada. La página generada debe ser subida al TecDigital y debe quedar publicada en algún sitio público (github, netlify, etc.)

## Consideraciones generales

- Todo el desarrollo del proyecto debe realizarse en lenguaje R.
- Se deberá generar una documentación formal, en formato pdf, en donde se describan las diferentes etapas del desarrollo del proyecto, las decisiones de diseño que se tomaron, los mecanismos de programación (en R) utilizados, y

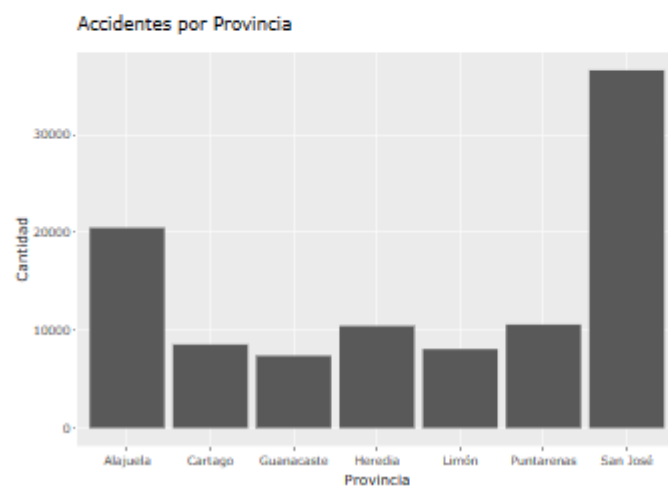
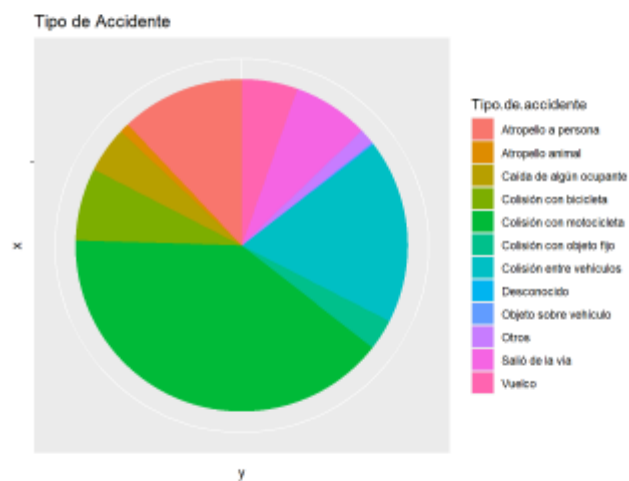
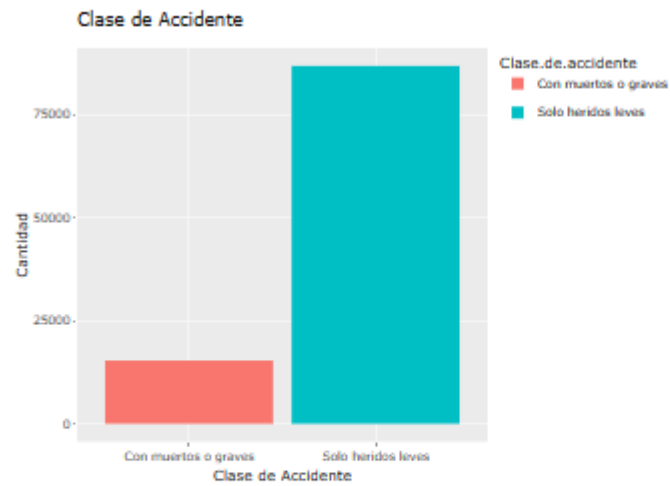
los resultados de las diferentes pruebas al programa. Dicha documentación deberá incluir al menos las siguientes secciones:

- Introducción
  - Descripción del problema (este enunciado)
  - Definición de fuentes de datos
  - Descripción detallada y explicación de las secciones principales del documento.
  - Conclusiones
- El proyecto puede realizarse en grupos de a los más dos estudiantes. No se permite la copia entre grupos de estudiantes.

Para el proyecto vamos a utilizar datos sobre los accidentes ocurridos en Costa Rica y que cuenta con el tipo de accidente, clima, heridos, hora, e incluso año. Y para poder realizar las gráficas vamos a tener que analizar dichos datos y para poder plasmarlos en una gráfica y que se puedan ver lo más claro posible, esto con la finalidad de hacer que el usuario final sea capaz de entender los datos sencillamente y que simplemente con ver las gráficas sepa de que tratan.

# Gráficas Unidimensionales

Gráficas Unidimensionales:

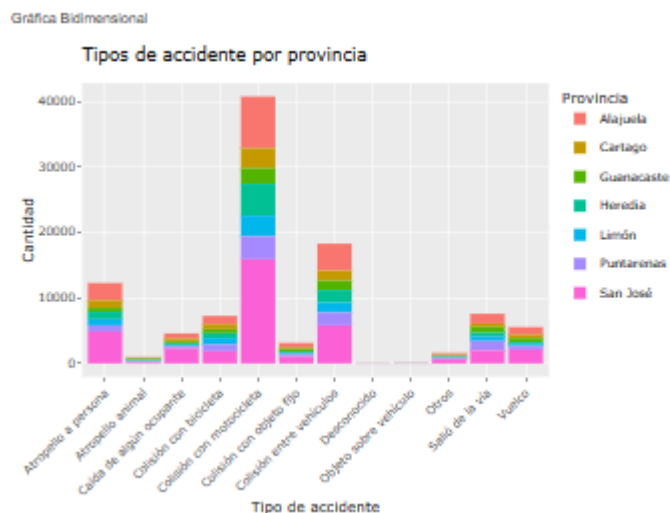


En las gráficas unidimensionales utilizamos dos gráficas de barras y una gráfica conocida como pie chart. En la primera gráfica podemos ver la diferencia en cuánto a la cantidad de accidentes con heridos que son leves en comparación a los accidentes con muertos o heridos de gravedad.

En la segunda gráfica podemos apreciar la comparación de los distintos Tipos de accidente y además podemos ver de una forma bastante clara cuál es el tipo de accidente que se presenta con más frecuencia.

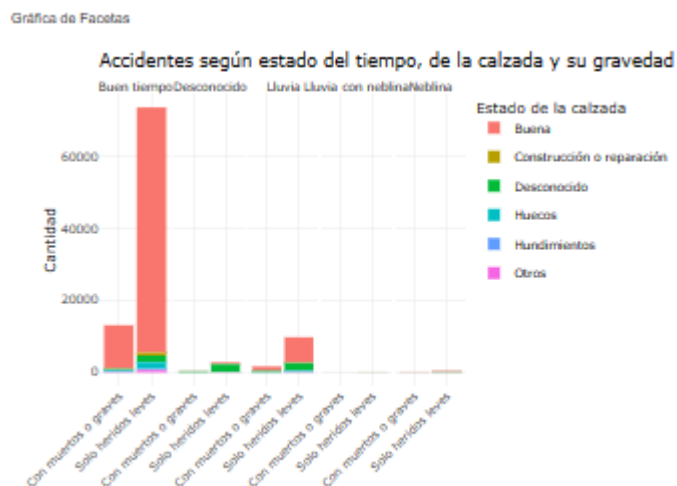
En la última gráfica se ven las provincias que más accidentes tienen en los años que se han documentado.

## Gráfica Bidimensional



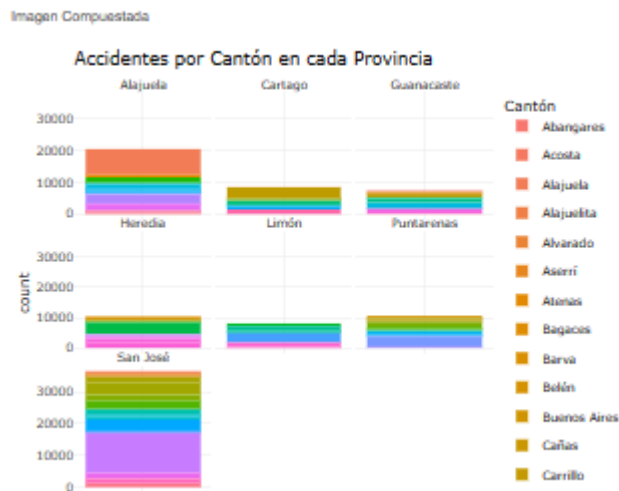
En este tipo de gráfica podemos apreciar los tipos de accidente y en qué provincia se han producido, por lo que podemos ver dos cosas claramente, las cuáles serían el accidente más común y en qué provincia se producen más de ese tipo.

## Gráfica de Facetas



En esta gráfica vemos los accidentes clasificados por su tipo de heridos y además logramos observar en qué estado de calzada se han dado más accidentes. También sabemos por lógica que se van a dar muchos más accidentes en calzadas en buen estado debido a que hay más circulación de vehículos por esas calles.

## Imagen Compuestada



Aquí podemos apreciar la cantidad de accidentes registrados en cada cantón de cada provincia y nuevamente podemos ver en qué cantón se registran más accidentes.

## Conclusiones

Organizar datos y mostrarlos en una gráfica no es una tarea sencilla ya que tenemos que depurar mucha información y además necesitamos que los datos que hayamos sacado puedan tener una relación, esto con el objetivo de poder dar una gráfica con sentido y que además los datos de esta puedan ser fácilmente vistos por el usuario. Las gráficas no solo deben de ser observadas, sino que también se deben de entender los datos que estas contienen, por lo que se debe de hacer muy bien la gráfica para que esta pueda ser entendida correctamente, ya que de nada sirve una gráfica que se vea muy bien estéticamente pero que al tener tantos colores y tantos datos entrelazados entre sí no se puedan lograr entender estos mismos. Al analizar durante este proyecto los datos de accidentes nos hemos dado cuenta de la importancia de estos puntos, como también a diferenciar ciertos factores que hacen que una gráfica se vea bien y sea funcional.