

Accidentalidad en Medellín

Brahian Cano Urrego
Daniel Alexander Cano Cuartas
Sebastian Pino Sánchez
Yeison Yovany Ocampo Naranjo

12 de febrero de 2018

Introducción

La movilidad como muchos otros factores son importantes en la economía, en las relaciones interpersonales, pero sobre todo en la vida diaria, puesto que en la actualidad gastamos mucho más tiempo del que se debería en el transporte nuestros desplazamientos, muchos de estos debidos a una serie de **accidentes de tránsito** que encontramos a nuestro paso; éstos no solo producen malestar entre las personas que se movilizan por las calles de una ciudad sino que también causan fallecimientos, lesiones, inseguridad entre los conductores y transeúntes. Por estas y muchas otras razones las organizaciones gubernamentales necesitan encontrar los factores que motiven o aumenten

Análisis Descriptivo

Con base a la información suministrada por la alcaldía y especialmente por la Secretaría de Movilidad de Medellín que recopiló datos sobre varios factores que pueden en donde se presentaron accidentes de tránsito, éstos fueron recopilados durante un período de tres años los cuales fueron: 2015, 2016 y 2017 en sus bases de datos libres que se encuentra en la página colombiana de **DATOS ABIERTOS**. Haremos un análisis estadístico *enfocado* en los factores: **BARRIO, Tipo de Accidente, Hora de los accidentes**, lo anterior para tratar de determinar si hay alguna relación entre estos factores con la accidentalidad de la ciudad de Medellín y sus corregimientos, además de analizar el alto número de accidentes que ocurrieron estos años en particular.



Figure 1: Alto flujo vehicular



Figure 2: Accidentalidad

Motivación

Para ninguna persona es un secreto que el número de accidentes ocurridos en la ciudad no solo es un gasto adicional de recursos de la Secretaría de Movilidad de Medellín para atender de manera oportuna y rápida los incidentes, además de esto genera gran congestión en las diferentes vías de la ciudad principalmente en las vías arterias que conectan gran parte de la población, por otro lado incrementa la inconformidad de las personas y reduciría la percepción de calidad de gestión de dicha secretaría, mencionando además la mala imagen que se presenta de la ciudad a nivel tanto nacional como internacional.

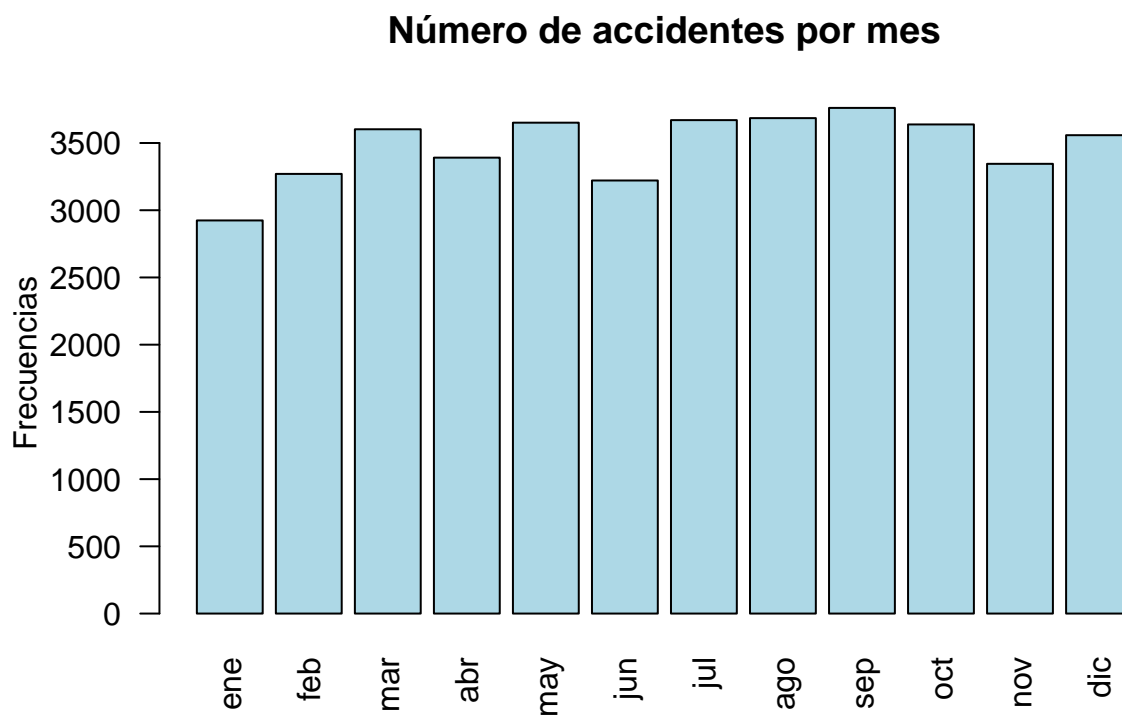
Requisitos

Basados en los comentarios anteriores hablaremos acerca de los requerimientos que solicita el contratante el cual será El sub secretario de movilidad, entre sus expectativas está la de poder entender mediante una aplicación entender el comportamiento de los meses, las horas y el tipo de accidente para diferentes años que tenemos posibles registros, así se consideran las siguientes necesidades a resolver:

- Que el usuario pueda filtrar por año
- Que pueda escoger el barrio de interés.
- Que pueda escoger la franja horaria de interés
- Que pueda seleccionar el tipo de accidente

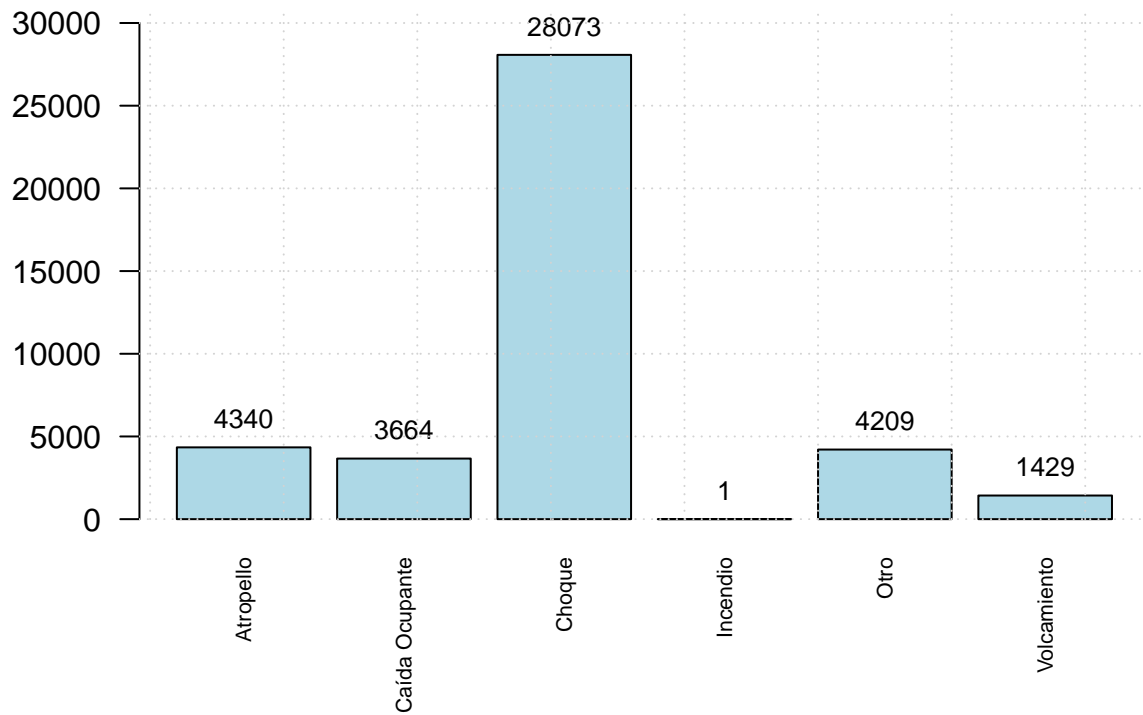
Una de las características principales de este trabajo es la capacidad de predecir la probabilidad de sufrir cualquier tipo de lesión o muerte bajo las características de **HORA, DISEÑO y BARRIO** para determinar si en un accidente ocurrido hay una persona que tuvo una lesión o una muerte con respecto a los accidentes donde solo hubo daños.

Para el año 2015



En el gráfico de barras no se puede apreciar una diferencia significativa, en el número de accidentes por mes pero podemos observar que la frecuencia más alta se encuentra en septiembre con 3761 registros de accidentes de tránsito y la más baja en enero con 2924 incidentes o accidentes de tránsito. Por otro lado observamos que los meses de febrero, junio y noviembre también poseen un número de accidentes menor, esto es cercano a las temporadas vacacionales

tipo de Accidentes(2015)



Claramente se puede observar que el tipo de accidente que predomina para este año es el de los choques, es el accidente más común en la movilidad y es mucho más frecuente que por ejemplo los atropellos que es la segunda categoría con más frecuencia, además cabe aclarar que la clase de accidente **“Incendio”** solo contó con un registro.

En la siguiente gráfica se presentará los 10 barrios con menor accidentalidad y los 10 barrios más accidentados respectivamente

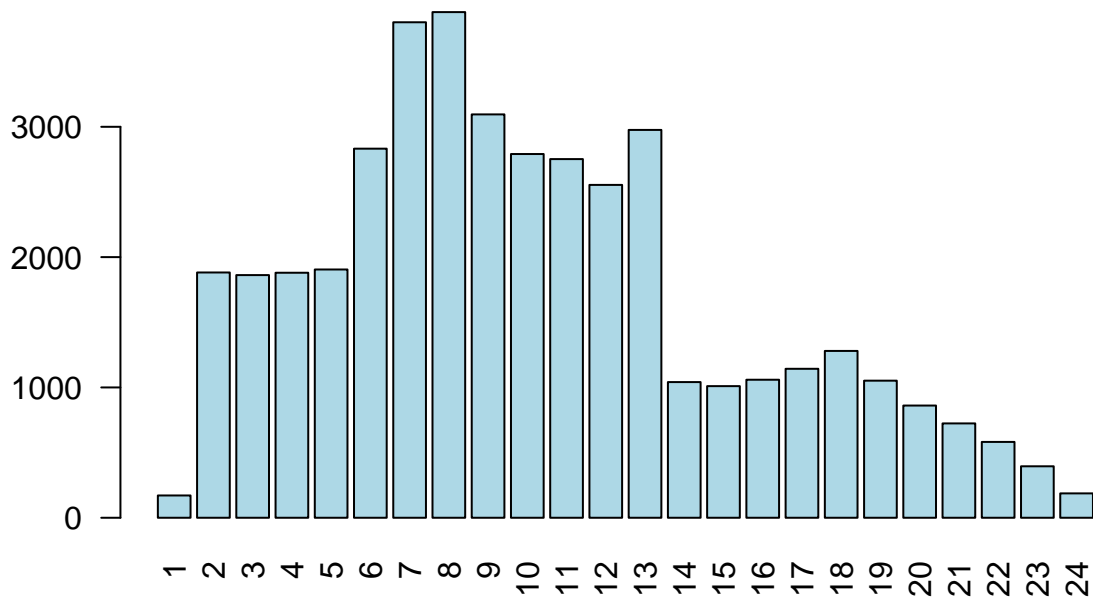
##			
##	Aures No.2	Barrios de Jesús	
##	1	1	
##	El Vergel	La Oculta	
##	1	1	
##	La Sierra	Moscú No.1	
##	1	1	
##	Piedras Blancas - Matasano	Piedras Blancas Represa	
##	1	1	
##	Potrerito	Santo Domingo Savio No.1	
##	1	1	
##			
##	Naranjal	Villa Nueva	Campo Amor
##	633	635	739
##	Barrio Colón	Los Conquistadores	Perpetuo Socorro
##	744	779	800
##	San Benito	Guayaquil	Caribe
##	855	881	894
##	La Candelaria		

##

1062

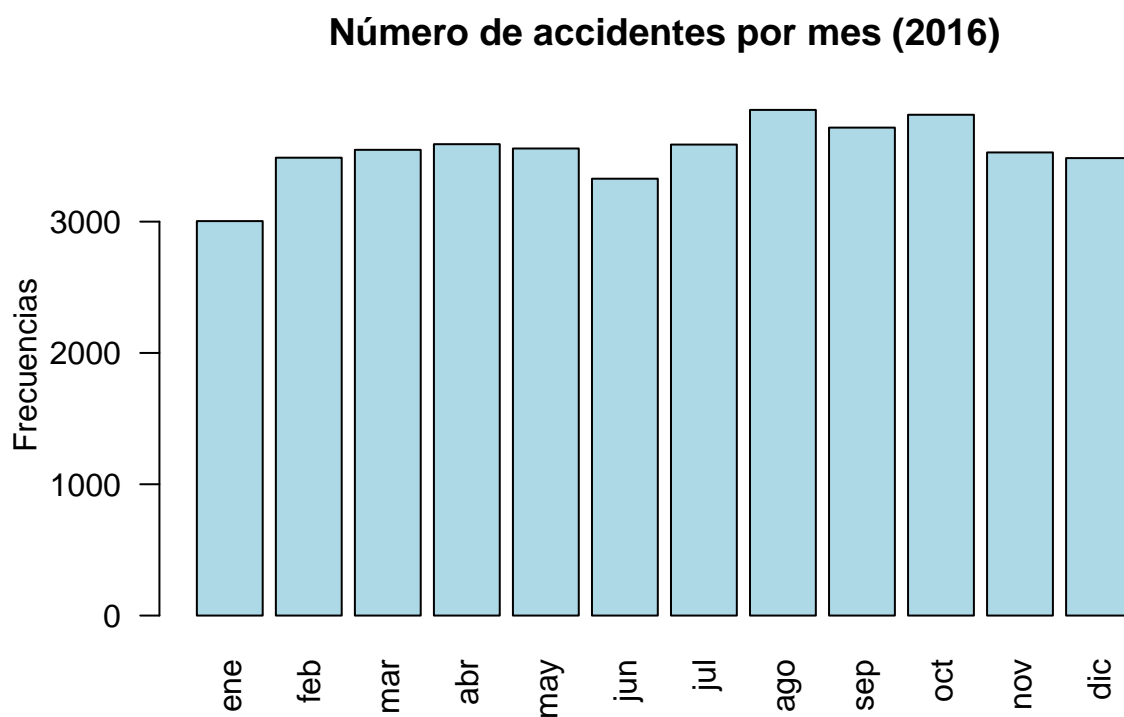
Como podemos observar en la tabla anterior que todos los barrio que poseen el menor registro de accidentes solo tienen escazamente un registro, por otro lado podemos apreciar que el barrio que mayor accidentalidad tiene es **La Candelaria** con 1062 seguido de **Caribe** con 894 choques, además todos los anteriores están a sectores altamente concurridos por conductores, caso que el los de poca tasa de incidencia no existen vías o zonas muy recurrentes.

frecuencia de la accidentalidad por hora 2015



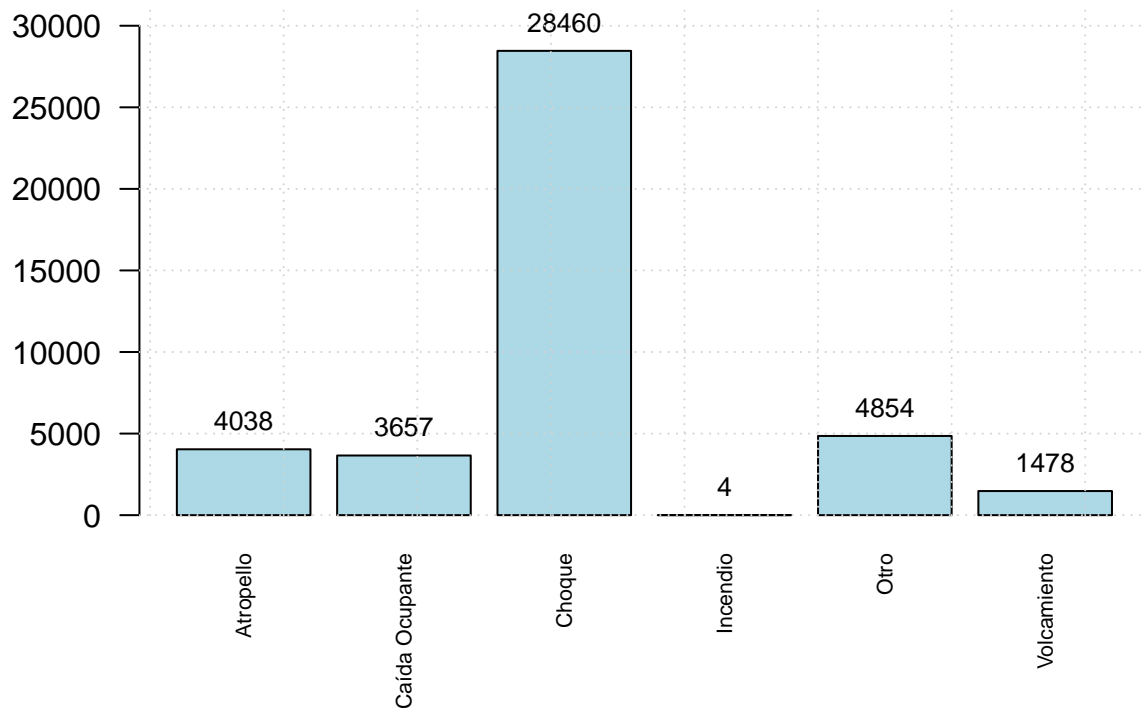
En este gráfico podemos ver un acercamiento al comportamiento de la accidentalidad por horas, donde podemos destacar los picos de frecuencia dentro del rango de 6 AM a 1 PM, y resaltamos también que los horarios con menos accidentalidad se ubican entre las 2 PM y la 1 AM, además observamos que el número de accidentes empieza a disminuir a partir de las 6 PM

Para el año 2016



Podemos apreciar en el gráfico de barras para cada uno de los meses y se nota que no hay una diferencia significativa entre los meses, es decir tiene cierta distribución uniforme, sin embargo se puede observar que los meses enero y junio son los meses de menor accidentalidad. Ésto podría deberse a la temporada vacacional de las personas más específicamente todas aquellas entidades públicas educativas.

tipo de Accidentes(2016)



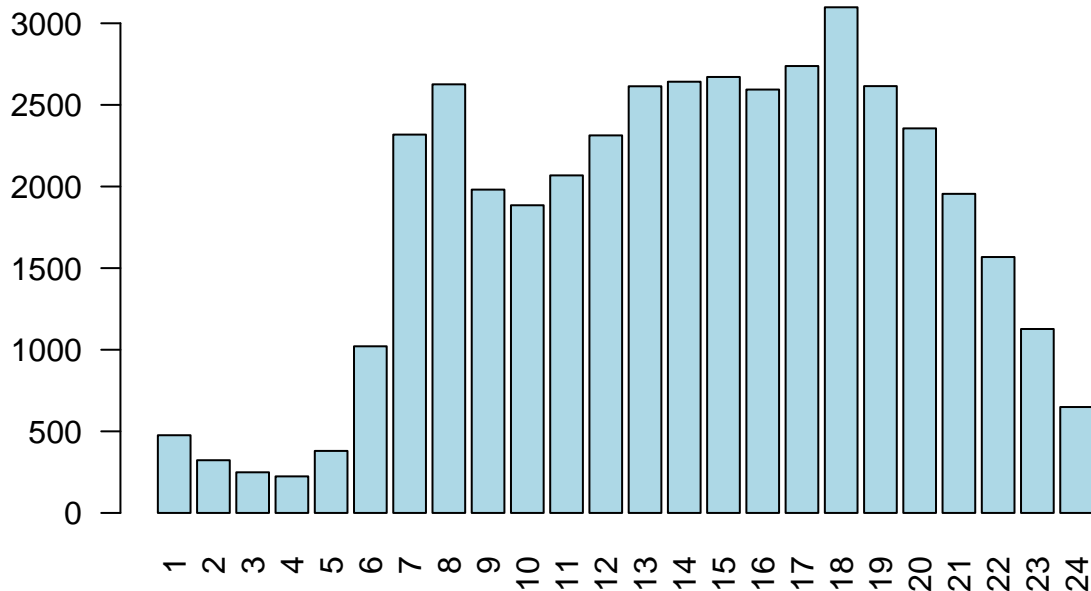
- Por otro lado observamos en el siguiente gráfico, el cual es el número de accidentes por cada tipo de accidente observamos en gran medida que los choque son el tipo de accidente más usual con un número de 28.460 casos, además el número de accidentes por incendios es un número muy bajo ya que solo cuenta con un total de 4 incidentes; cabe destacar que la siguiente clase con menor número de accidentes es la de volcamientos con un total de 1478 incidentes de este tipo.

En la siguiente gráfica se presentará los 10 barrios con menor accidentalidad y los 10 barrios más accidentados respectivamente

##			
##	B. Cerro El Volador	Aures No.2	La Oculta
##	0	1	1
##	Santa Rosa de Lima Suburbano	Chacaltaya	Versalles No.1
##	1	1	1
##	Versalles No.2	Villa Lilliam	Yolombo
##	1	1	1
##	Berlín		
##	2		
##			
##	Carlos E. Restrepo	Santa Fé	Barrio Colón
##	618	693	700
##	San Benito	Guayaquil	Los Conquistadores
##	726	757	814
##	Campo Amor	Caribe	Perpetuo Socorro
##	836	859	870
##	La Candelaria		
##	946		

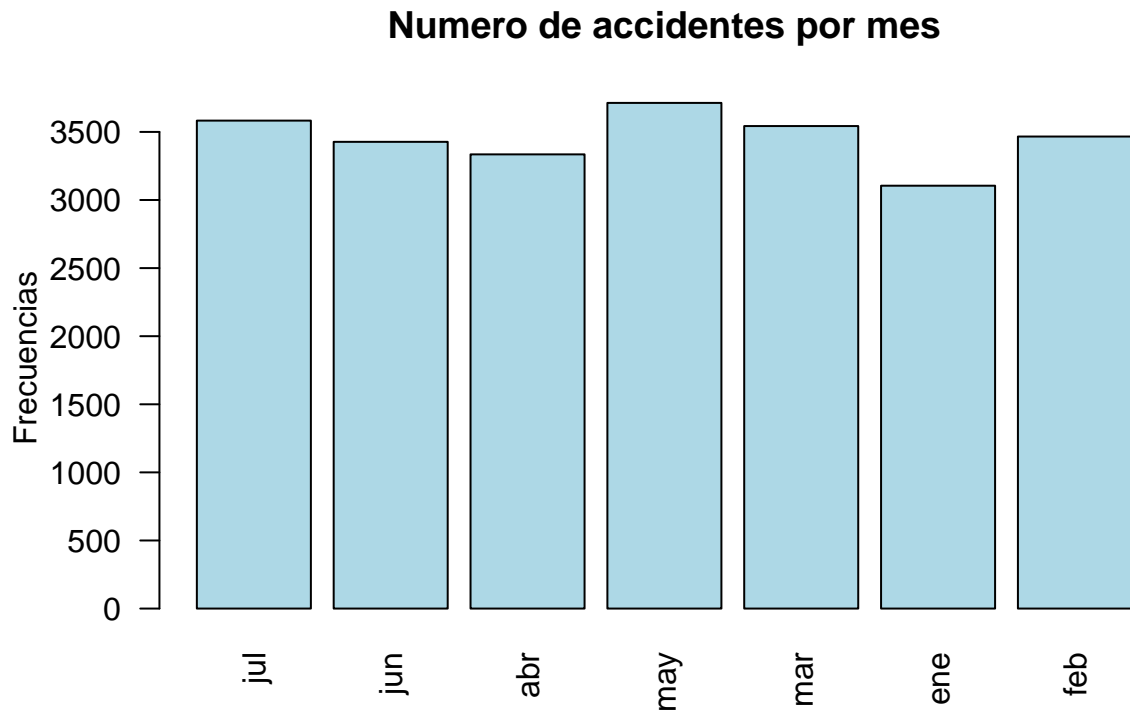
Como podemos observar en la tabla anterior que todos los barrio que poseen el menor registro de accidentes solo tienen escasamente uno a dos registro, por otro lado podemos apreciar que el barrio que mayor accidentalidad tiene es La Candelaria con 948 seguido de Perpetuo Socorro con 875, además todos los anteriores están a sectores altamente concurridos por conductores, caso que el los de poca tasa de incidencia no existen vias o zonas muy recurrentes.

frecuencia de la accidentalidad por hora 2016



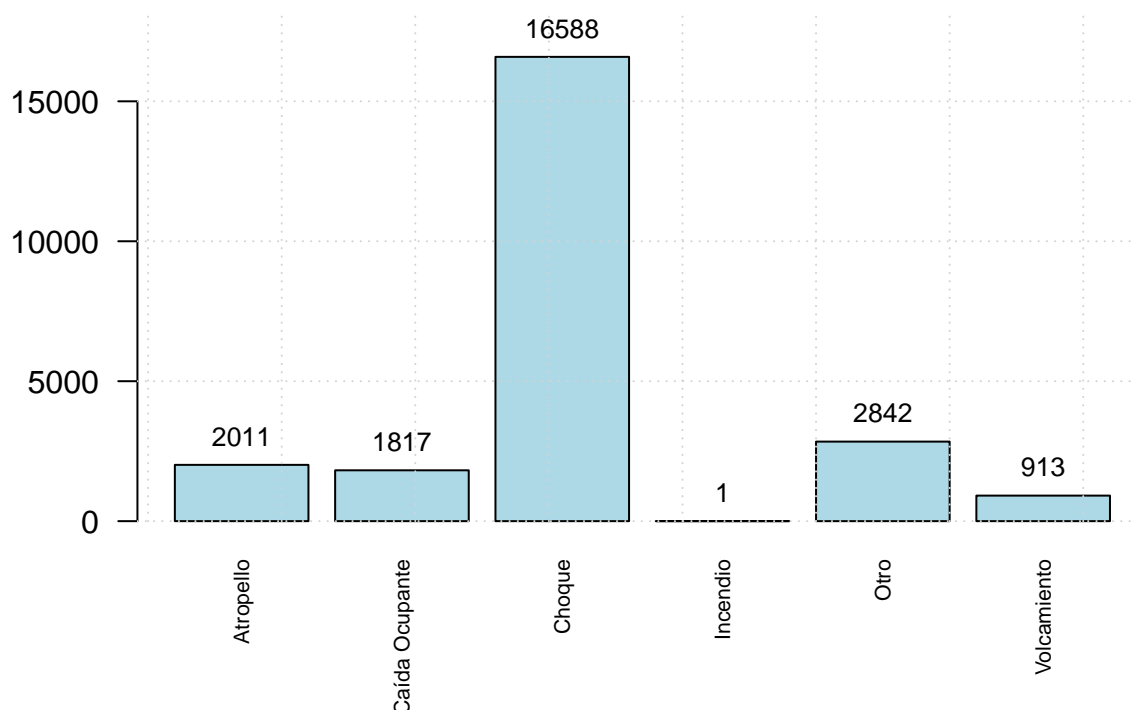
Para la frecuencia de accidentalidad por hora notamos que hay una baja frecuencia para las primeras 4-5 horas del día, sin embargo a partir de las 6 AM encontramos un rápido crecimiento del número de accidentes, especialmente en la franja de la mañanas hasta las 9 AM hay un crecimiento muy rápido. Además observamos que la hora que presenta mayor problema es las 18 horas, equivalente a las 6 PM. Cabe resaltar que en la franja de la noche a partir de las 7 PM empieza a decrecer rápidamente el número de accidentes a medida que transcurre la noche.

Para el año 2017



Podemos apreciar en el gráfico de barras de cada uno de los meses del año 2017 y se observa que no hay una diferencia significativa entre los meses, es decir tiene cierta distribución uniforme, ésto contrario a la creencia de que la temporada decembrina aumenten el número de accidentes o las temporadas vacacionales que se esperaría que el número de accidentes disminuya debido a la masiva salida de autos de la ciudad, sin embargo se puede observar que los meses enero y abril son los meses de menor accidentalidad.

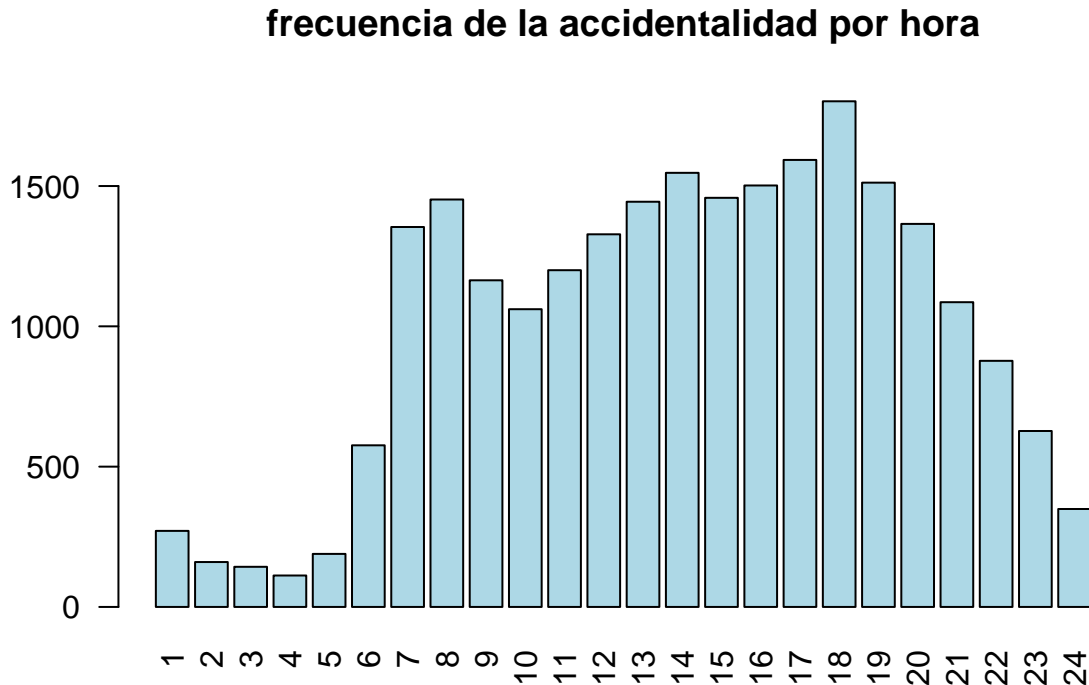
tipo de Accidentes(2017)



En este caso observando en el siguiente gráfico, el cual es el número de accidentes por cada tipo de accidente, observamos en gran medida que los choques son el tipo de accidente más usual con un número de 29.193 casos, además el número de accidentes por incendios es un número muy bajo ya que solo cuenta con un total de 4 incidentes, también choque y atropello (C y A) con un registro. Cabe destacar que la siguiente clase con mayor número de accidentes es la de otros con un total de 1478 incidentes de este tipo.

```
##
## Area De Expansion Belen Rincon                      El Vergel
##                               1                      1
##                Piedra Gorda          Santo Domingo Savio No.1
##                               1                      1
##                Suburbano el Tesoro          Suburbano Palmitas
##                               1                      1
##                Travesías                Villa Lilliam
##                               1                      1
##                               6001                AUC1
##                               2                      2
##
##
## Terminal de Transporte          Guayaquil          Carlos E. Restrepo
##                344                351                363
##                Santa Fé          Los Conquistadores          Barrio Colón
##                404                415                425
##                Campo Amor          Perpetuo Socorro          Caribe
##                449                493                494
##                La Candelaria
##                526
```

Como podemos observar en la tabla anterior que todos los barrio que poseen el menor registro de accidentes solo tienen escasamente un registro, por otro lado podemos apreciar que el barrio que mayor accidentalidad tiene es **La Candelaria** con 926 seguido de Caribe con 876, además todos los anteriores están a cercanos o asociados a vías arteriales y a sectores altamente concurridos por conductores, caso que el los de poca tasa de incidencia no existen vías o zonas muy recurrentes.



En este caso la frecuencia de accidentalidad por hora se nota que hay una baja frecuencia para las primeras 4-5 horas del día, sin embargo a partir de las 6 AM encontramos un rápido crecimiento del número de accidentes, especialmente en la franja de la mañanas hasta las 8 AM hay un crecimiento muy rápido. Además observamos que las horas que presenta mayor problema son las horas pico (7-8 AM y 6-7 PM). Cabe resaltar que en la franja de la noche a partir de las 8 PM empieza a decrecer rápidamente el número de accidentes a medida que transcurre la noche.

Inferencias sobre la gravedad de los accidentes

Ajustando un modelo.

Se ha decidido ajustar un modelo logit donde nos va a interesar la probabilidad de que dado que ya se accidentó una persona o hubo un reporte de accidente, ¿Cuál sería la probabilidad de que en el accidente halla estado involucrada la lesión o muerte de una persona considerando nuestra clase de interés como (Muerto o Herido), esto lo realizamos usando la variable **GRAVEDAD** Y restringiendo a el valor de 0 los choques y el valor de 1 en caso de que en el accidente halla involucrada una persona ya sea en una lesión o una muerte. Con esta nueva variable que se nombró **var_resp** se han considerado hacer una selección de variables empíricas que se relacionan con el lugar del accidente donde luego de un análisis descriptivo se consideraron importantes las variables **HORA**, **DISEÑO** y **CLASE** para el diseño de un modelo no

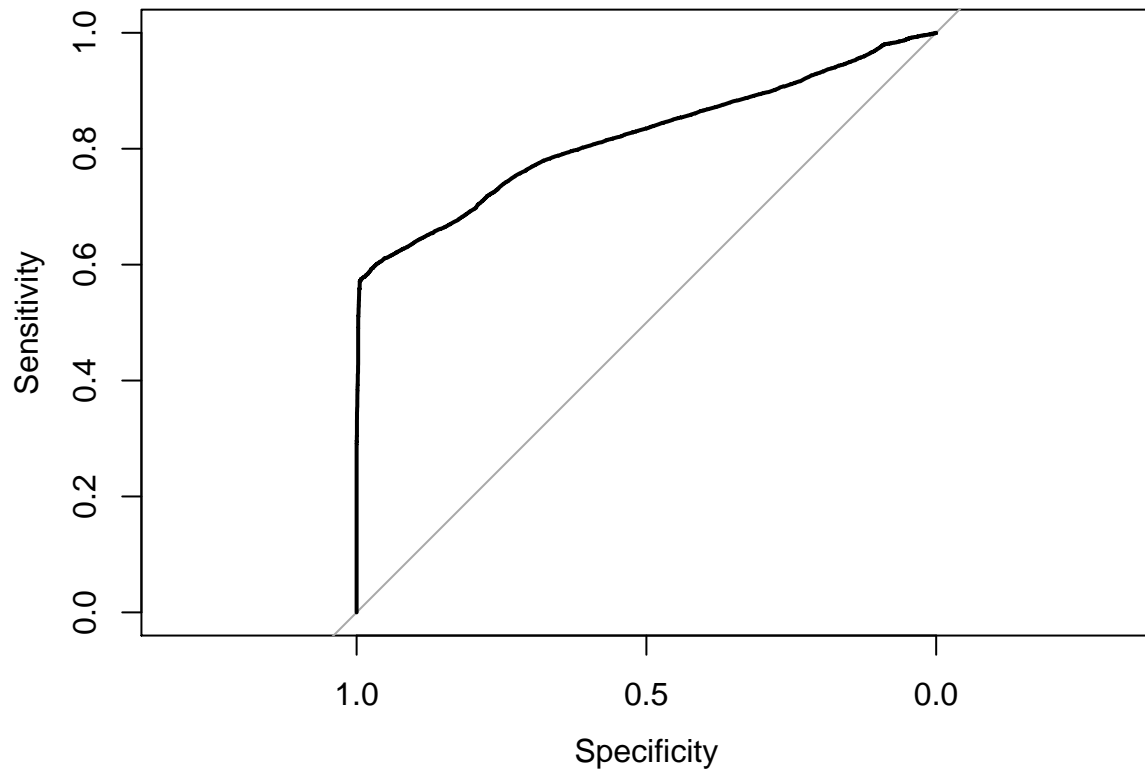
lineal y encontrar la probabilidad asociada a un accidente con ciertas características asociadas al accidente y tratando de predecir así su probabilidad de lesión o muerte en un accidente.

El modelo ajustado se presenta a continuación:

```
##
## Call:
## glm(formula = var_res ~ CLASE + HORA + DISENO, family = "binomial",
##      data = base3)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4346  -0.8645   0.0001   0.2600   1.8351
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.631e+05  5.592e+04  -4.704 2.55e-06 ***
## CLASECaída Ocupante  6.864e-02  2.099e+02   0.000 0.999739
## CLASEChoque       -1.934e+01  1.445e+02  -0.134 0.893551
## CLASEIncendio     -3.677e+01  6.524e+03  -0.006 0.995503
## CLASEOtro         -1.395e+01  1.445e+02  -0.097 0.923113
## CLASEVolcamiento  -1.516e+01  1.445e+02  -0.105 0.916449
## HORA              -4.232e-06  8.996e-07  -4.705 2.54e-06 ***
## DISENOGlorieta    -5.351e-01  2.435e-01  -2.198 0.027973 *
## DISENOInterseccion  5.598e-01  2.269e-01   2.467 0.013634 *
## DISENOLote o Predio -2.608e-01  2.424e-01  -1.076 0.281842
## DISENOPaso a Nivel -2.861e-01  6.961e-01  -0.411 0.681056
## DISENOPaso Elevado  7.309e-01  2.870e-01   2.546 0.010889 *
## DISENOPaso Inferior 1.210e+00  3.498e-01   3.458 0.000544 ***
## DISENOPontón       4.717e-01  7.952e-01   0.593 0.553061
## DISENOPuente       -9.624e-02  3.197e-01  -0.301 0.763372
## DISENOTramo de via  5.482e-03  2.257e-01   0.024 0.980623
## DISENOTunel        7.197e-01  1.432e+00   0.503 0.615311
## DISENOVia peatonal -2.365e+00  2.492e+00  -0.949 0.342664
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 33314  on 24171  degrees of freedom
## Residual deviance: 21527  on 24154  degrees of freedom
## AIC: 21563
##
## Number of Fisher Scoring iterations: 17
```

Se encontraron que las variables consideradas son útiles para el modelo, en particular *CLASE* al estar incluida en el modelo se logró aumentar la tasa de buena clasificación en aproximadamente un 20% y las demás variables poseen al menos un nivel significativo dentro del modelo.

A continuación haremos el análisis para validar la bondad del ajuste del modelo usando la curva **ROC**



Area under the curve: 0.8192

Encontramos un área bajo la curva y un gráfico que indica el ajuste. Observe que la curva ROC nos dió un valor de casi el 82% el cual es un valor considerablemente bueno para realizar las estimaciones de las probabilidades, en particular para pronosticar la probabilidad de sufrir alguna lesión o muerte en un accidente presentado con ciertas características.

Probabilidad= 0.421

Como podemos observar en los resultados anteriores hemos dado el ejemplo de un caso en particular donde un sujeto (accidente) con las siguientes características: **se accidentó a las 7PM**, ____ donde el accidente fue un choque____ y **ocurre en una intersección** los da una probabilidad de que en el accidente halla un herido o muerto es del 42.1%.

Conclusiones

- Frente a la creencia tradicional de que el mes de un año en particular pudiera tener mayor número de accidentes se ve desmentida con gran fuerza ya que la distribución aproximada según las gráficas en cada uno de los años es bastante uniforme, por otro lado se pueden tener creencias respecto a que la temporada vacacional la cual se comprende en los meses mayormente de *enero, junio, diciembre*, sin embargo también se ve desmentido dado que en algunos casos dicho mes tiene un número de accidentes altos.
- Se observó que en el tipo de accidente que predomina es aquel que solo involucra los choques, por otro lado notamos que el suceso de que el accidente sea un incendio es realmente escaso que aunque el número de volcamientos es considerable no es una de las categorías más grandes
- La tasa de mortalidad de los accidente es extremadamente baja, encontrando solamente casos en el año 2017, esto puede ser un buen indicador de que por lo menos si se están tomando las medidas preventivas adecuadamente y la seguridad de los vehículos ha mejorado considerablemente y la rigurosidad de las nuevas medidas de tránsito han permitido reducir el número de muertos.
- Se resalta que los valores en las probabilidades en ciertas categorías es muy alto, en algunos casos 100% aproximadamente, pero esto se debe a las características del accidente.
- Podemos apreciar en este modelo que la **clase de accidente** es realmente importante para determinar

si hay una lesión o muerte dado el accidente.

- Encontramos que la categoria **otro** (De la cual no se tiene mayor información) tiene una probabilidad de que una persona sufra una lesión o muerte muy alta, lo cual nos conlleva a pensar qué tipo de accidentes son los que comprenden esta categoria, ya que falta especificación.
- Se encontró que al ocurrir un accidente la probabilidad de tener una lesión o muerte en un atropello es aproximadamente del 100% lo cual es un claro resultado esperado, además de que se notó que según las características los choques son los que presentan una “menor” probabilidad de lesión o muerte.
- Se encontró que la hora en que se transitan las personas si es un factor que influye sobre el número de accidentes que ocurren, en especial en aquellas horas que conocemos como horas pico ya que hay un desplazamiento masivo de personas estas horas conocidas como horas críticas , sobre ésto se podrían implementar estrategias que diversificarán o mejorarán las horas en el que la gente se desplaza.
- A partir de que ocurre un accidente podemos hacer que en el 50% de los casos se verá involucrada la integridad física de una de las personas, es decir, en el 50% de los casos habrá un muerto o un herido en el incidente.

Recomendaciones:

- Se recomienda que al tomar los datos el usuario solo deba ingresar una de las opciones, puesto que al digitar o escribir se cometieron graves errores y eso impide los cálculos y las demás estimaciones usadas.
- Se recomienda seguir una estricta recolección de datos ya que estos nos permitirán comprender mejor aún el problema y tomar medidas de contingencia para disminuir este alto número de accidentes.
- Se recomienda dar una descripción más clara y completa de las variables que usan, especialmente de las categorias de ciertas variables ya que ciertas clases no son claras y no permiten tener una comprensión mejor al problema.