UNIVERSIDAD DEL VALLE DE GUATEMALA

Data Science

Sección 10

Ing. Lynette García

Lab 8 - Visualización

Análisis exploratorio

Diego Sevilla 17238

Rodrigo Samayoa 17332

Alejandro Tejada 17584

Guatemala, 08 de octubre de 2020

Datos INE

Limpieza de datos

La limpieza de los datos consistió en tener un set de tablas para los años 2014, 2015, 2016 y 2017 con la misma forma y datos para así poder predecir resultados para el año 2018.

Las variables escogidas para trabajar fueron las siguientes: año, mes y dia de la semana de ocurrencia, horas de ocurrencia, departamento de ocurrencia, tipo y color de vehículo.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
num_corr	año_ocu	dia_ocu	hora_ocu	g_hora	g_hora_5	mes_ocu	día_sem_(	mupio_oc	depto_oc	zona_ocu	tipo_veh	marca_ve	color_veh	modelo_veh	g_modelo	tipo_eve
1	2018	1	16	3	2	1	1	115	1	99	4	32	5	9999	99	2
2	2018	1	12	3	2	1	1	2207	22	99	1	69	2	9999	99	1
3	2018	1	7	2	1	1	1	2102	21	99	1	999	6	9999	99	2
4	2018	1	22	4	3	1	1	1903	19	99	4	999	5	9999	99	2
5	2018	1	23	4	3	1	1	1903	19	99	4	27	5	9999	99	2
6	2018	1	1	1	1	1	1	501	5	99	1	69	2	2003	4	1

Se eliminaron columnas innecesarias para el análisis:

```
#Limpiamos 2016
DB2016$núm_corre <- NULL
DB2016$día_ocu <- NULL
DB2016$mupio_ocu <- NULL
DB2016$zona_ocu <- NULL
DB2016$marca_veh <- NULL
DB2016$modelo_veh <- NULL
DB2016$g_modelo_veh <- NULL
DB2016$tipo_eve <- NULL
DB2016$área_geo_ocu <- NULL
names(DB2016)[names(DB2016) == "año_ocu"] <- "anio_ocu"
names(DB2016)[names(DB2016) == "día_sem_ocu"] <- "dia_sem_ocu"
str(DB2016)
```

Y se obtiene la forma:

```
> str(DB2016)
'data.frame':   7964 obs. of  9 variables:
 $ anio_ocu   : num  2016 2016 2016 2016 2016 ...
 $ hora_ocu   : Factor w/ 25 levels "0","1","10","11",..: 13 13 6 5 5 20 13 3 10 14 ...
 $ mes_ocu    : Factor w/ 12 levels "Enero","Febrero",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ dia_sem_ocu: Factor w/ 7 levels "Lunes","Martes",..: 5 5 5 5 5 5 5 5 5 5 ...
 $ depto_ocu  : Factor w/ 22 levels "Guatemala","El Progreso",..: 1 1 1 1 1 1 1 22 20 19 ...
 $ tipo_veh   : Factor w/ 20 levels "Automóvil","Camioneta",..: 2 1 1 4 2 1 1 20 4 20 ...
 $ color_veh  : Factor w/ 18 levels "Rojo","Blanco",..: 9 5 9 3 9 6 18 18 18 18 ...
 $ g_hora     : Factor w/ 5 levels "00:00 a 05:59",..: 1 1 3 3 3 1 1 2 3 4 ...
 $ g_hora_5   : Factor w/ 4 levels "Mañana","Tarde",..: 1 1 2 2 2 1 1 1 2 3 ...
 - attr(*, "variable.labels")= Named chr  "Número de correlativo" "Día de ocurrencia" "Año de ocurren
cia" "Hora de ocurrencia" ...
  ..- attr(*, "names")= chr  "núm_corre" "día_ocu" "año_ocu" "hora_ocu" ...
 - attr(*, "codepage")= int 1252
```

En el caso de los datos de 2015 se realizaron condiciones de reemplazo debido a que las variables estaban representadas con números a diferencia de las otras tablas que tenían los nombre de cada variable en cuestión:

Antes:

```
> str(DB2015)
'data.frame':   6854 obs. of  9 variables:
 $ anio_ocu   : int  2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
 $ mes_ocu    : int  1 1 1 1 1 1 1 1 1 1 ...
 $ dia_sem_ocu: int  4 4 4 4 4 4 4 4 4 4 ...
 $ hora_ocu   : int  16 22 2 9 1 8 17 10 20 7 ...
 $ g_hora     : int  3 4 1 2 1 2 3 2 4 2 ...
 $ g_hora_5   : int  2 3 1 1 1 1 2 1 3 1 ...
 $ depto_ocu  : int  1 1 1 1 1 1 1 6 6 11 ...
 $ tipo_veh   : int  4 4 3 4 1 4 4 4 3 3 ...
 $ color_veh  : int  5 5 6 5 4 4 1 5 1 2 ...
```

```
#mes
DB2015$mes_ocu <- as.factor(ifelse(DB2015$mes_ocu ==  1 , "Enero", ifelse(
  DB2015$mes_ocu ==  2, "Febrero", ifelse(
    DB2015$mes_ocu ==  3, "Marzo",   ifelse(
      DB2015$mes_ocu ==  4, "Abril",   ifelse(
        DB2015$mes_ocu ==  5, "Mayo",    ifelse(
          DB2015$mes_ocu ==  6, "Junio",   ifelse(
            DB2015$mes_ocu ==  7, "Julio",   ifelse(
              DB2015$mes_ocu ==  8, "Agosto",  ifelse(
                DB2015$mes_ocu ==  9, "Septiembre",    ifelse(
                  DB2015$mes_ocu ==  10, "Octubre",    ifelse(
                    DB2015$mes_ocu ==  11, "Noviembre",  "Diciembre"
                  ))))))))))))
```

** para cada variable revisar .R
Después

```
> str(DB2015)
'data.frame':   6854 obs. of  9 variables:
 $ anio_ocu   : num  2015 2015 2015 2015 2015 ...
 $ mes_ocu    : Factor w/ 12 levels "Abril","Agosto",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ dia_sem_ocu: Factor w/ 7 levels "Domingo","Jueves",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ hora_ocu   : Factor w/ 24 levels "0","1","2","3",..: 17 23 3 10 2 9 18 11 21 8 ...
 $ g_hora     : Factor w/ 4 levels "00:00 a 05:59",..: 3 4 1 2 1 2 3 2 4 2 ...
 $ g_hora_5   : Factor w/ 3 levels "Mañana","Noche",..: 3 2 1 1 1 1 3 1 2 1 ...
 $ depto_ocu  : Factor w/ 22 levels "Alta Verapaz",..: 7 7 7 7 7 7 7 18 18 15 ...
 $ tipo_veh   : Factor w/ 19 levels "Automóvil","Bicicleta",..: 15 15 17 15 1 15 15 15 17 17 ...
 $ color_veh  : Factor w/ 15 levels "Amarillo","Anaranjado",..: 12 12 15 12 9 9 13 12 13 5 ...
```

De esta forma, con todas las tablas trabajadas se realizó un merge con los datos de cada año normalizados y así comenzar con el análisis y agrupamiento.

Posteriormente, debido al análisis de clustering todas las variables correspondientes se sustituyen por valores numéricos de acuerdo a un código proporcionado por la fuente de datos.

Exploración de los datos

1. Haga un resumen de las variables numéricas e investigue si siguen una distribución normal y tablas de frecuencia para las variables categóricas, escriba lo que vaya encontrando.

2. Cruce las variables que considere que son las más importantes para hallar los elementos clave que lo pueden llevar a comprender lo que está causando el problema encontrado.

Las variables con las que se está trabajando son en su mayoría de categoría cualitativa y no cuantitativa por lo cual se deben convertir a cuantitativas para hacer un clustering efectivo.

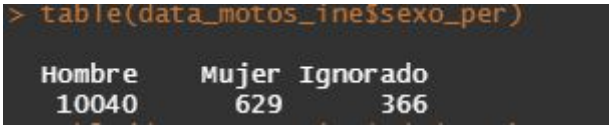
Tablas de frecuencias

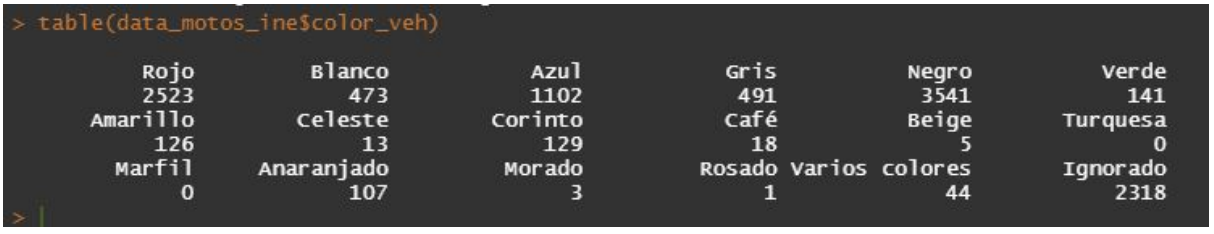
Se van a observar 7 variables, las cuales van a estar descritas a continuación con una descripción de cada una de las variables.

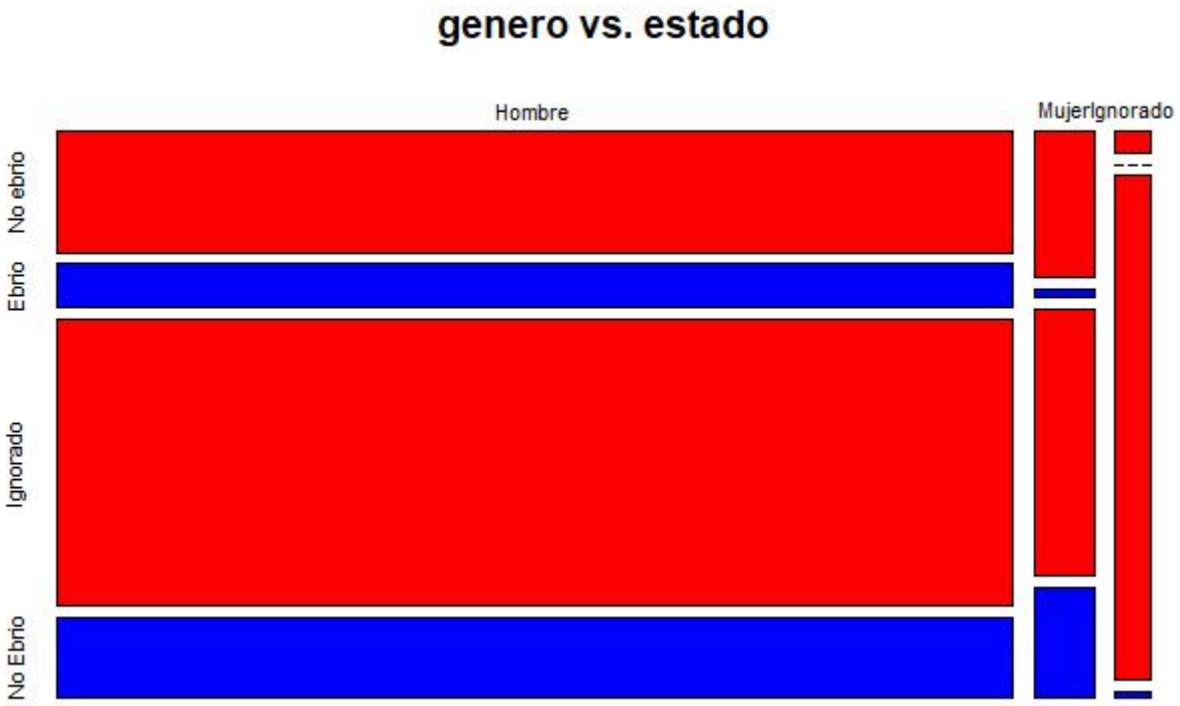
Variables a observar	Descripción
Sexo	Nos indica el sexo de la persona involucrada en el accidente, esta se representa con M para los hombres y F para las mujeres.
Hora	Esta nos indica la hora en la cual ocurrió el accidente, al principio la variable estaba por hora de acontecimiento en 4 dígitos, se cambió a que fuera por segmentos de 6 horas para determinar de mejor manera los intervalos de tiempo donde ocurren la mayoría de los accidentes.
Tipo de Vehículo	El tipo de vehículo estaba descrito por automóvil, motocicleta, camioneta, bus y muchos otros más, sin embargo para poder hacer el clustering de manera adecuada se cambió por un código numérico.
Mes	Los meses estaban escritos por su nombre y se cambió a que fuera su valor numérico del calendario, esto ya que se sospecha que la mayoría de accidentes deberían ocurrir durante épocas festivas, en su mayoría diciembre.

Año	Esta nos sirve para clasificar los años de los accidentes y para poder hacer un grupo de los años 2014-2017 para entrenamiento con Cross-Validation y proyectar para 2018 como nuestro año de testing, la razón es que no se puede tener mucha certeza de los valores de 2019 ya que son muy recientes por cuál quedó descartado ese año de momento.
Estado	Esta variable nos indica el estado de la persona que estaba implicada en el accidente, indicando si estaba ebria o no o si se encontraba bajo algún efecto.
Departamento	Dado que queremos saber en qué lugar ocurren la mayoría de accidentes para poder tomar más precauciones en tales áreas se escogió esta variable. Esta se encontraba con los nombres de los departamentos en los cuales ocurrió el accidente, sin embargo fueron cambiadas a variables numéricas para poder ser clasificadas en un cluster.

Revisión de datos en general







En esta gráfica podemos apreciar un estimado de las proporciones de los hombres y mujeres accidentados con estado de ebriedad o no. En el caso de los casos donde se ignora el sexo podria ignorarse dicha proporcion.

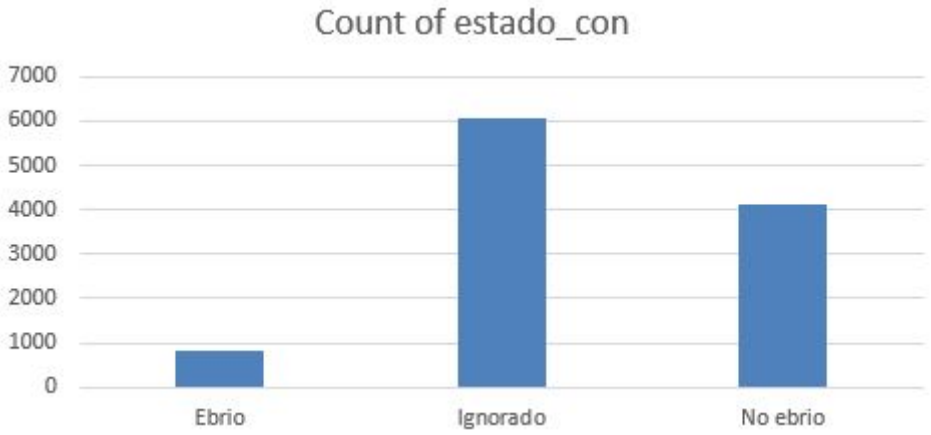
Es interesante observar que la mayoría de accidentes no suceden en estado de ebriedad, este es el caso tanto para las mujeres como para los hombres

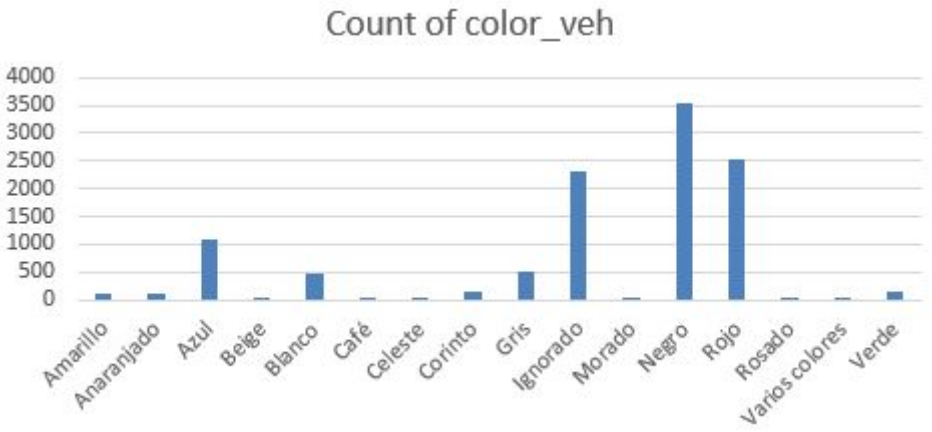
```
> chisq.test(generoEstado)

        Pearson's Chi-squared test

data:  generoEstado
X-squared = 295.1, df = 6, p-value < 0.00000000000000022
```

Chi cuadrado nos indica qué





Datos SAT

Limpieza de los datos

Se realizó la lectura de datos para los años 2016 a 2019.

```
##2016
data012016 <- read.delim("dataSAT2016/web_imp_08012016.txt", sep = "|",header=TRUE,row.names = NULL)
data022016 <- read.delim("dataSAT2016/web_imp_08022016.txt", sep = "|",header=TRUE,row.names = NULL)
data032016 <- read.delim("dataSAT2016/web_imp_08032016.txt", sep = "|",header=TRUE,row.names = NULL)
data042016 <- read.delim("dataSAT2016/web_imp_08042016.txt", sep = "|",header=TRUE,row.names = NULL)
data052016 <- read.delim("dataSAT2016/web_imp_08052016.txt", sep = "|",header=TRUE,row.names = NULL)
data062016 <- read.delim("dataSAT2016/web_imp_08062016.txt", sep = "|",header=TRUE,row.names = NULL)
data072016 <- read.delim("dataSAT2016/web_imp_08072016.txt", sep = "|",header=TRUE,row.names = NULL)
data082016 <- read.delim("dataSAT2016/web_imp_08082016.txt", sep = "|",header=TRUE,row.names = NULL)
data092016 <- read.delim("dataSAT2016/web_imp_08092016.txt", sep = "|",header=TRUE,row.names = NULL)
data102016 <- read.delim("dataSAT2016/web_imp_08102016.txt", sep = "|",header=TRUE,row.names = NULL)
data112016 <- read.delim("dataSAT2016/web_imp_08112016.txt", sep = "|",header=TRUE,row.names = NULL)
data122016 <- read.delim("dataSAT2016/web_imp_08122016.txt", sep = "|",header=TRUE,row.names = NULL)

dataSat2016 <- rbind(data012016,
                     data022016,
                     data032016,
```

Luego de leer todos los datos se unieron y renombraron las columnas por un problema con num.row qué causo el corrido de los nombres de las columnas inesperadamente.

```r
names(dataSat)[names(dataSat) == "row.names"] <- "Pais.de.Proveniencia_"
names(dataSat)[names(dataSat) == "Pais.de.Proveniencia"] <- "Aduana.de.Ingreso_"
names(dataSat)[names(dataSat) == "Aduana.de.Ingreso"] <- "Fecha.de.la.Poliza_"
names(dataSat)[names(dataSat) == "Fecha.de.la.Poliza"] <- "Partida.Arancelaria_"
names(dataSat)[names(dataSat) == "Partida.Arancelaria"] <- "Modelo.del.Vehiculo_"
names(dataSat)[names(dataSat) == "Modelo.del.Vehiculo"] <- "Marca_"
names(dataSat)[names(dataSat) == "Marca"] <- "Linea_"
names(dataSat)[names(dataSat) == "Linea"] <- "Centimetros.Cubicos_"
names(dataSat)[names(dataSat) == "Centimetros.Cubicos"] <- "Distintivo_"
names(dataSat)[names(dataSat) == "Distintivo"] <- "Tipo.de.Vehiculo_"
names(dataSat)[names(dataSat) == "Tipo.de.Vehiculo"] <- "Tipo.de.Importador_"
names(dataSat)[names(dataSat) == "Tipo.de.Importador"] <- "Tipo.Combustible_"
names(dataSat)[names(dataSat) == "Tipo.Combustible"] <- "Asientos_"
names(dataSat)[names(dataSat) == "Asientos"] <- "Puertas_"
names(dataSat)[names(dataSat) == "Puertas"] <- "Tonelaje_"
names(dataSat)[names(dataSat) == "Tonelaje"] <- "Valor.CIF_"
names(dataSat)[names(dataSat) == "Valor.CIF"] <- "Impuesto_"
str(dataSat)
View(dataSat)
```

Estructura de los datos

```
> str(data_motos_)
'data.frame':   610101 obs. of  13 variables:
 $ Pais.de.Proveniencia_: chr  "ITALIA" "ALEMANIA REP. FED." "ALEMANIA REP. FED." "ALEMANIA REP. FED." ..
 $ Aduana.de.Ingreso_   : chr  "EL CARMEN" "PUERTO BARRIOS" "PUERTO BARRIOS" "PUERTO BARRIOS" ...
 $ Fecha.de.la.Poliza_  : chr  "26/01/2016" "12/01/2016" "12/01/2016" "12/01/2016" ...
 $ Modelo.del.Vehiculo_ : int  2009 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
 $ Marca_               : chr  "APRILIA" "BMW" "BMW" "BMW" ...
 $ Linea_               : chr  "SCARABEO 100" "F 800 GS" "F 800 GS" "F 700 GS" ...
 $ Centimetros.Cubicos_ : int  96 798 798 798 798 798 798 999 1170 798 ...
 $ Distintivo_          : chr  "LIVIANO" "LIVIANO" "LIVIANO" "LIVIANO" ...
 $ Tipo.de.Vehiculo_    : chr  "MOTO" "MOTO" "MOTO" "MOTO" ...
 $ Tipo.de.Importador_  : chr  "OCASIONAL" "DISTRIBUIDOR" "DISTRIBUIDOR" "DISTRIBUIDOR" ...
 $ Valor.CIF_           : num  2755 77220 77220 64675 64675 ...
 $ Impuesto_            : num  331 9266 9266 7761 7761 ...
 $ Año                  : chr  "2016" "2016" "2016" "2016" ...
```

```
> head(data_motos_)
   Pais.de.Proveniencia_ Aduana.de.Ingreso_ Fecha.de.la.Poliza_ Modelo.del.Vehiculo_ Marca_       Linea_ Centimetros.Cubicos_ Distintivo_
76               ITALIA          EL CARMEN          26/01/2016                 2009 APRILIA SCARABEO 100                  96     LIVIANO
95    ALEMANIA REP. FED.      PUERTO BARRIOS          12/01/2016                 2016     BMW     F 800 GS                 798     LIVIANO
96    ALEMANIA REP. FED.      PUERTO BARRIOS          12/01/2016                 2016     BMW     F 800 GS                 798     LIVIANO
97    ALEMANIA REP. FED.      PUERTO BARRIOS          12/01/2016                 2016     BMW     F 700 GS                 798     LIVIANO
98    ALEMANIA REP. FED.      PUERTO BARRIOS          12/01/2016                 2016     BMW     F 700 GS                 798     LIVIANO
99    ALEMANIA REP. FED.      PUERTO BARRIOS          12/01/2016                 2016     BMW     F 700 GS                 798     LIVIANO
   Tipo.de.Vehiculo_ Tipo.de.Importador_ Valor.CIF_ Impuesto_  Año
76              MOTO           OCASIONAL     2754.97    330.60 2016
95              MOTO         DISTRIBUIDOR   77219.82   9266.38 2016
96              MOTO         DISTRIBUIDOR   77219.82   9266.38 2016
97              MOTO         DISTRIBUIDOR   64674.99   7761.00 2016
98              MOTO         DISTRIBUIDOR   64674.99   7761.00 2016
99              MOTO         DISTRIBUIDOR   65323.98   7838.88 2016
```

Resumen de los datos

```
> summary(data_motos_)
 Pais.de.Proveniencia_ Aduana.de.Ingreso_ Fecha.de.la.Poliza_ Modelo.del.Vehiculo_    Marca_             Linea_          Centimetros.Cubicos_
 Length:610101         Length:610101      Length:610101       Min.   :1900         Length:610101      Length:610101      Min.   :   0.0
 Class :character      Class :character   Class :character    1st Qu.:2017         Class :character   Class :character   1st Qu.: 125.0
 Mode  :character      Mode  :character   Mode  :character    Median :2018         Mode  :character   Mode  :character   Median : 125.0
                                                              Mean   :2018                                               Mean   : 151.7
                                                              3rd Qu.:2019                                               3rd Qu.: 150.0
                                                              Max.   :2020                                               Max.   :6000.0
                                                                                                                         NA's   :5
 Distintivo_        Tipo.de.Vehiculo_  Tipo.de.Importador_   Valor.CIF_          Impuesto_           Año
 Length:610101      Length:610101      Length:610101       Min.   :    860   Min.   :   103.3   Length:610101
 Class :character   Class :character   Class :character    1st Qu.: 318184   1st Qu.: 38182.1   Class :character
 Mode  :character   Mode  :character   Mode  :character    Median : 445150   Median : 53418.0   Mode  :character
                                                           Mean   : 773687   Mean   : 92876.3
                                                           3rd Qu.: 796619   3rd Qu.: 95594.2
                                                           Max.   :7849242   Max.   :941909.1
```

Tablas de frecuencias Variables categóricas

País de proveniencia

	Var1	Freq
1	CHINA	401179
2	INDIA	177778
3	JAPON	13483
4	BRASIL	7512
5	TAIWAN	2298
6	ESTADOS UNIDOS	1621
7	INDONESIA	1434
8	TAILANDIA	1020
9	ALEMANIA REP. FED.	939
10	COLOMBIA	605
11	AUSTRIA	568
12	ITALIA	540
13	HONDURAS	321
14	SUIZA	141
15	FRANCIA	119
16	ESPANA	102
17	COREA DEL SUR	89

Aduana de ingreso

	Var1	Freq
1	PUERTO QUETZAL	561296
2	EXPRESS AEREO	13841
3	EL CARMEN	12310
4	SANTO TOMAS DE CASTILLA	6278
5	TECUN UMAN	5416
6	PEDRO DE ALVARADO	3410
7	PUERTO BARRIOS	2629
8	SAN CRISTOBAL	2149
9	ADUANA INTEGRADA AGUA CALIENTE	1706
10	ADUANA INTEGRADA EL FLORIDO	322
11	G8, CENTRALSA	265
12	ADUANA INTEGRADA CORINTO	104
13	G1, INTEGRADA	93
14	G3, ALPASA	90
15	G5, CEALSA	83
16	G4, ALSERSA	72
17	LA MESILLA	10

Modelo

	Var1	Freq
1	2018	161940
2	2017	147985
3	2019	145388
4	2020	81165
5	2016	53099
6	2005	1973
7	2006	1936
8	2015	1880
9	2007	1826
10	2004	1651
11	2003	1444
12	2008	1323
13	2009	1146
14	2013	870
15	2014	834
16	2012	784
17	2002	776

Marca

	Var1	Freq
1	SUZUKI	116873
2	HONDA	111081
3	ITALIKA	79672
4	BAJAJ	59460
5	FREEDOM	53446
6	YAMAHA	41490
7	SERPENTO	40551
8	HERO	22050
9	MOVESA	20208
10	HAOJUE	14135
11	TVS	7588
12	ASIA HERO	7334
13	MRT	3913
14	AVANTI	2674
15	OSBORNE	2590
16	KYMCO	2402
17	KTM	1956

Linea de produccion

	Var1	Freq
1	GN125F	51555
2	NAVI 110	19864
3	XR150L	19130
4	GTK125	15288
5	CGL125	15000
6	PULSAR 135 LS	12279
7	FIRE125	11778
8	AN125HK	11078
9	EN125-2A	8656
10	CS125	8268
11	AX100	8190
12	FIRE 150	7497
13	PULSAR 125 NS	7115
14	XTZ125E	6650
15	GIXXER	6618
16	125Z	6406
17	UNICORN 160	6288

Tipo de distribuidor

	Var1	Freq
1	OCASIONAL	513546
2	DISTRIBUIDOR	96555

Las variables a analizar son:

- país de proveniencia
- aduana de ingreso
- fecha de poliza
- modelo del vehiculo
- Marca
- linea
- centrimetros cubicos
- valor CIF
- impuesto.

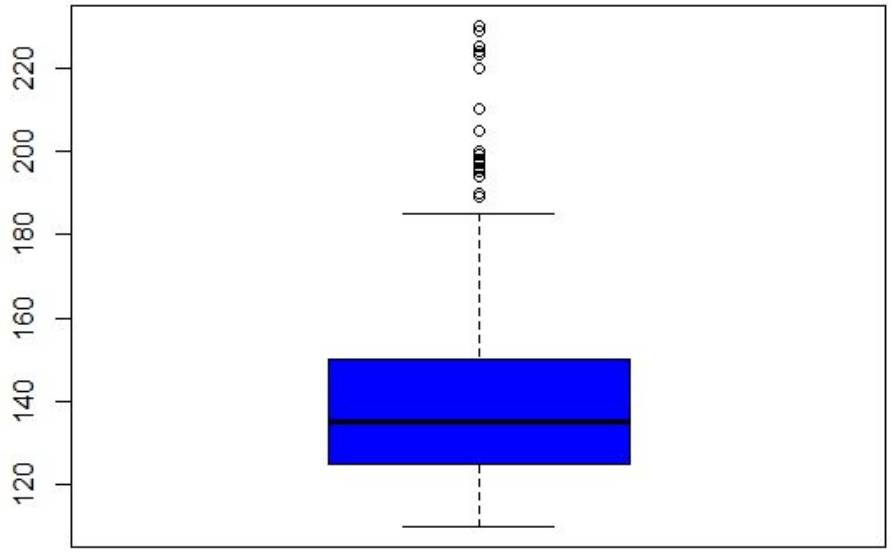
Variables a observar	Descripción
país de proveniencia	País de donde viene la motocicleta
aduana de ingreso	Aduana de ingreso al país
modelo del vehículo	El año del modelo
Marca	Marca del vehículo
Línea	Linea de produccion

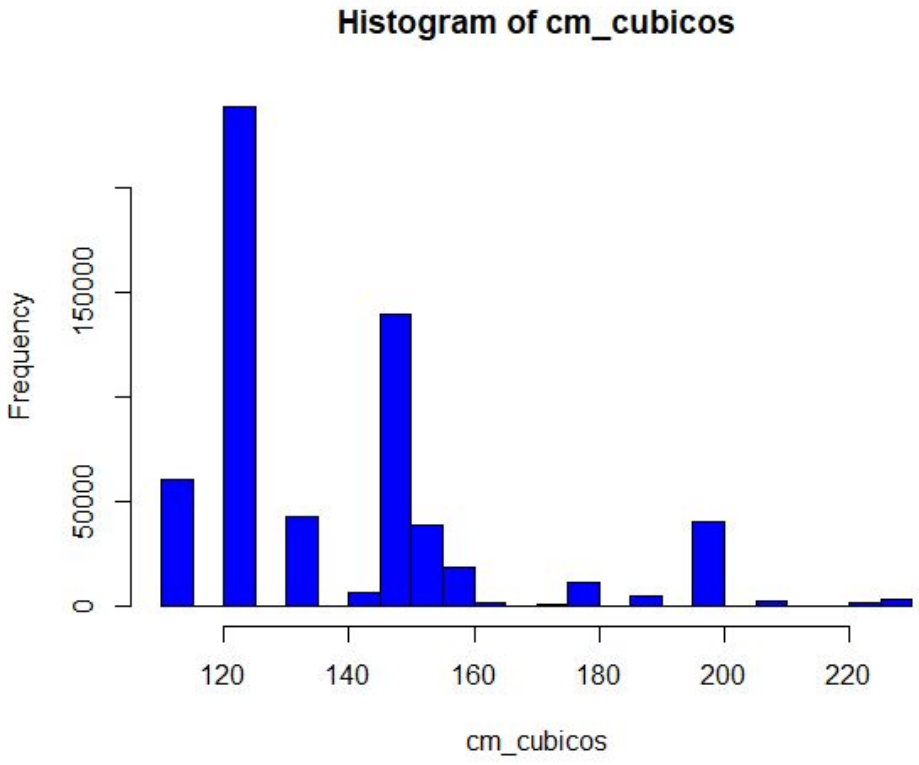
Centímetros cúbicos	Tamaño del motor en centímetros cúbicos
Valor CIF	El **valor CIF** es el **valor** real de las mercancías durante el despacho aduanero, el cual abarca tres conceptos: **costo** de las mercancías en el país de origen, **costo** del seguro y **costo** del flete hasta el puerto de destino.
Impuesto	Es el impuesto a pagar en la aduana

Gráfico de introducción de los datos

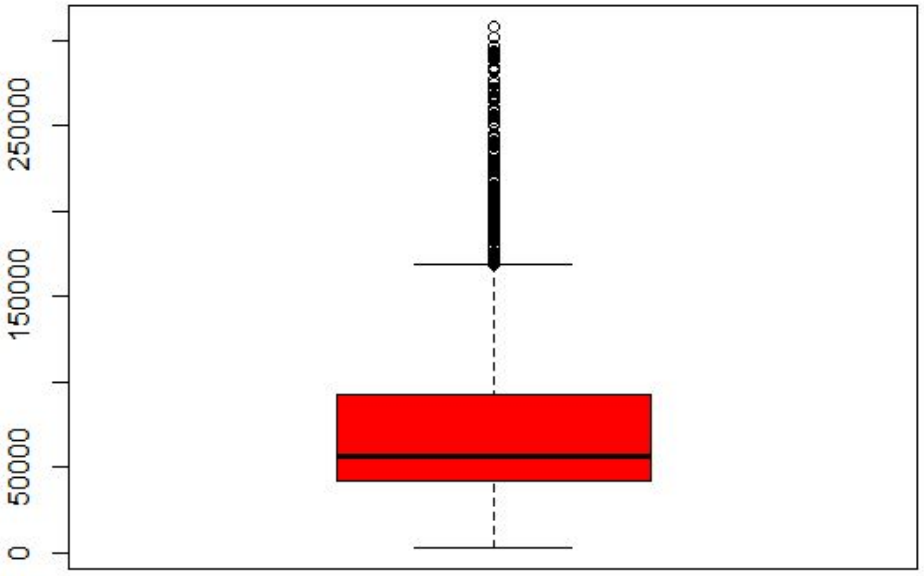
Analisis variables cuantitativas

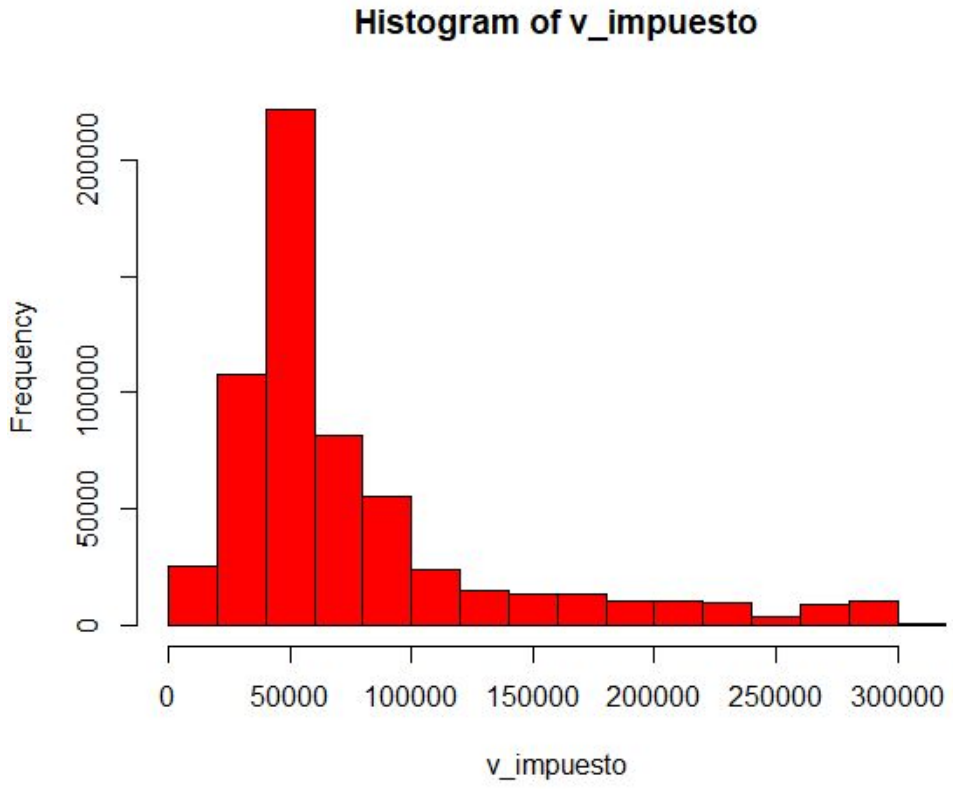
Centímetros cúbicos





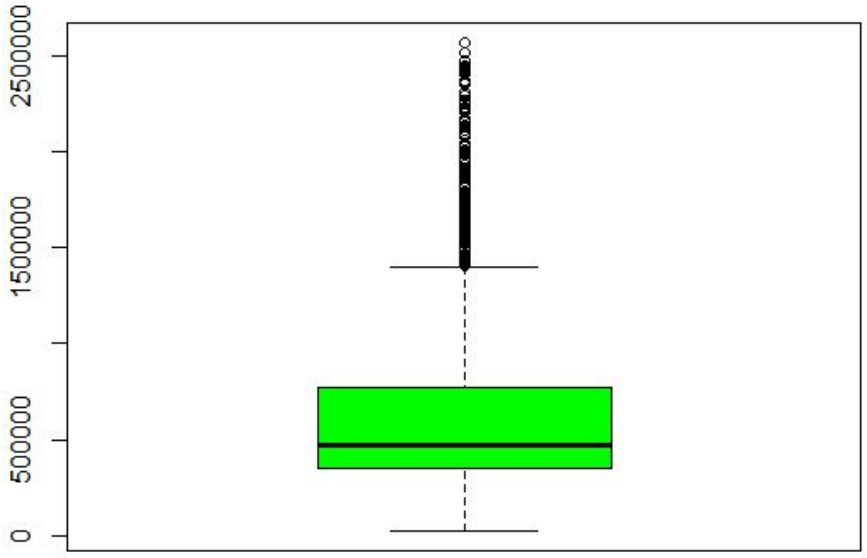
Impuesto

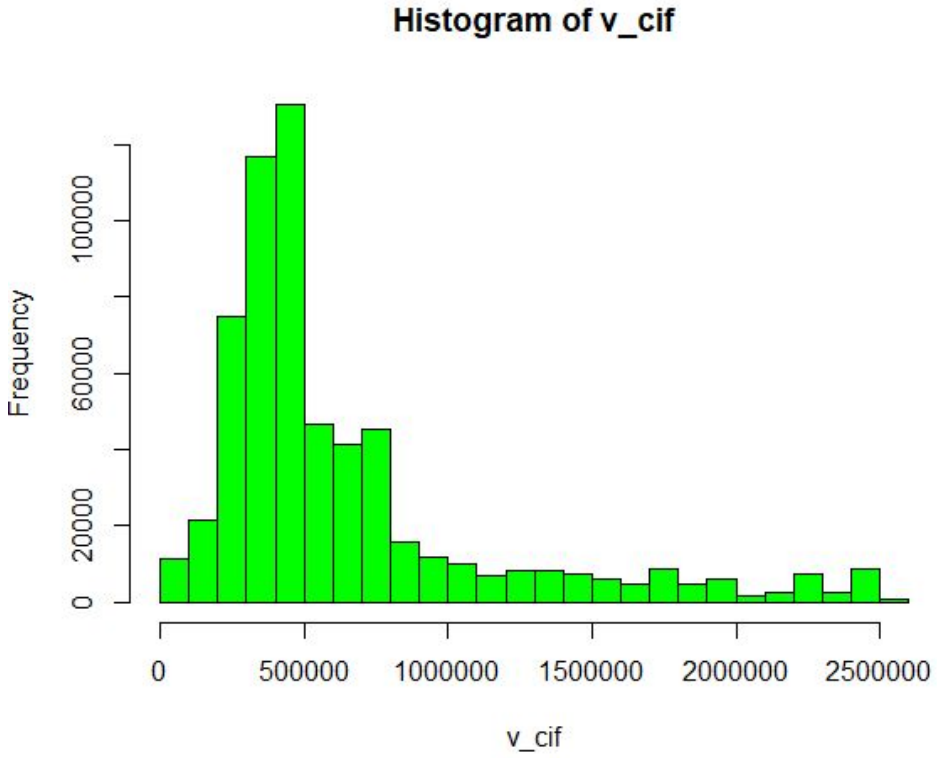




Se puede observar que existe una asimetría positiva en la campana de gauss del histograma de impuestos. La mayoría de los datos están entre Q.400,000 y Q.500,000.

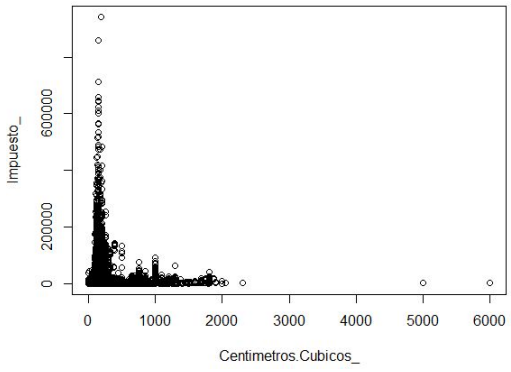
Valor CIF

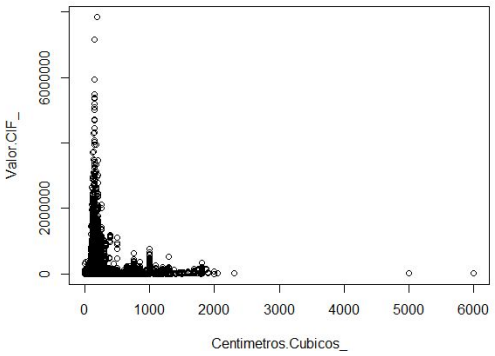




Se puede observar que existe una asimetría positiva en la campana de gauss del histograma valor cif. La mayoría de los datos están entre 400,000 y 500,000.

Correlaciones





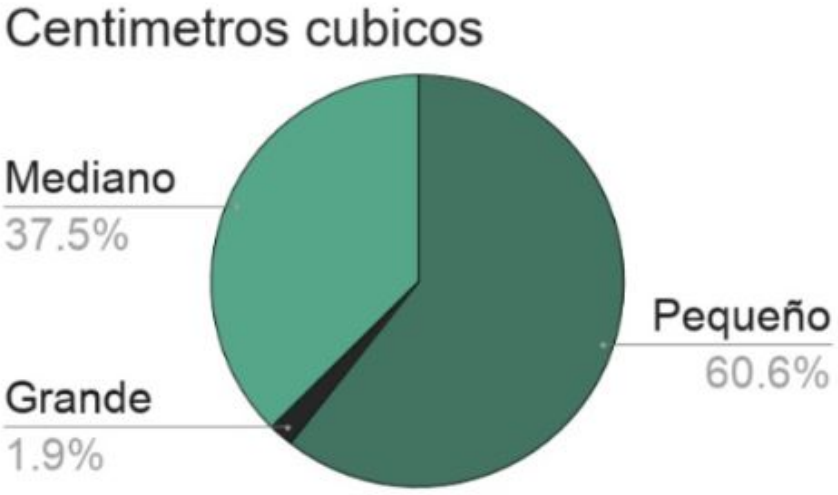
Las correlaciones muestran que no hay relación entre el tamaño del motor y el pago de impuestos de aduana o el valor cif.

Porcentajes del tamaño de motor de motos que entran a guatemala a través de las aduanas

```
> summary(data_motos_$Centimetros.Cubicos_)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    0.0   125.0   125.0   151.7   150.0  6000.0       5
>
```

Podemos observar que la mayoría de motos usadas en Guatemala son de tamaño pequeño mediano.

Está es una grafica de pie con el tamaño de motos que entran al país



En este caso se observa que casi no hay mercado de motocicletas de 500 cm3 en adelante, porque en la ciudad es más conveniente tener una motocicleta mediana o pequeña para transportarse con facilidad en los pueblos y ciudades.

Conclusiones

1. Los porcentajes de tamaño de motor para las motocicletas que entran al país por medio de las aduanas son
2. Los impuestos pagados en aduanas no tiene correlación con el tamaño de motor de las motocicletas
3. El Valor Cif engloba los impuestos pagados en el ingreso de cualquier motocicleta en el país. Es decir sus correlaciones son casi perfectas.
4. La distribución de los datos no cumple exactamente con las características de una distribución normal, esto puede observarse
5.

Referencias

- https://stackoverflow.com/questions/13871614/replacing-values-from-a-column-using-a-condition-in-r
- https://bookdown.org/rdpeng/exdata/exploratory-graphs.html
- https://stats.stackexchange.com/questions/376291/calculate-percentage-of-each-sub-category-in-r-programming
- https://portal.sat.gob.gt/portal/descarga/11569/acuerdos-y-tablas-de-valores-2019/34160/tabla-iva-importacion-e-iprima-2019.pdf