

DATOS DE PRESENTACIÓN

MAESTRIA EN CIENCIAS EN INFORMÁTICA

ALUMNO : CARLOS MARTINEZ O.

MATERIA : ALMACENAMIENTO DE DATOS Y SU ADMINISTRACIÓN

PROFESOR : M.EN C. EDUARDO RENÉ RODRÍGUEZ ÁVILA

TAREA : 7

RESUMEN

EL presente informe tiene como objetivo analizar una década de registros en archivos acerca de las muertes ocasionadas por las Tormentas. Los informes fueron tomados desde el repositorio de la NATIONAL CLIMATIC DATA CENTER, generada por la NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION responsable de la conservación, supervisión, evaluación y facilitar el acceso público de la información del clima.

DESCRIPCIÓN DE LOS DATOS

Esta sección describe la información contenida en el archivo consolidado en el periodo de 1 década (2000 al 2010). El archivo contiene la información de las muertes ocurridas a causa de las Tormentas.

fatality_id : Número de identificación asignado por la National Weather Service en un evento de tormenta.

event_id : Número de identificación asignado por la NWC para distinguir una parte dentro del evento de tormenta.

fatality_type : Tipo de Deceso (D = Ocasionada directamente por el evento, I = Ocasionada indirectamente por el evento).

fatality_date : Fecha y hora del evento (Formato: 01/14/2000 12:00:00).

fatality_age : Edad de las víctimas.

fatality_sex : Sexo de las víctimas.

fatality_location: Ubicación de las muertes.

OBTENCIÓN DE DATOS

Esta sección tiene el objetivo de visualizar el archivo y el tipo de información que contiene.

Obtención de las dimensiones del archivo

```
## REGISTROS TOTALES: 6915
## COLUMNAS TOTALES : 11
```

Obtención de las columnas del archivo

```
## [1] "FAT_YEARMONTH"      "FAT_DAY"            "FAT_TIME"
## [4] "FATALITY_ID"        "EVENT_ID"           "FATALITY_TYPE"
## [7] "FATALITY_DATE"      "FATALITY_AGE"       "FATALITY_SEX"
## [10] "FATALITY_LOCATION" "EVENT_YEARMONTH"
```

Obtención del resumen del archivo

```
## FAT_YEARMONTH      FAT_DAY      FAT_TIME      FATALITY_ID
## Min.   :200001     Min.   : 1.0    Min.   : 0      Min.   : 2
## 1st Qu.:200309     1st Qu.: 9.0    1st Qu.: 0      1st Qu.: 5124
## Median :200508     Median :18.0    Median : 0      Median : 19531
## Mean   :200544     Mean   :17.1    Mean   : 159    Mean   : 218284
## 3rd Qu.:200803     3rd Qu.:27.0    3rd Qu.: 0      3rd Qu.: 21810
## Max.   :201012     Max.   :31.0    Max.   :1200    Max.   :1008716
##
##      EVENT_ID      FATALITY_TYPE      FATALITY_DATE
## Min.   : 445      D:6083          08/28/2005 11:00:00: 982
## 1st Qu.:159356     I: 832          02/05/2008 00:00:00: 54
## Median :5263586                    05/04/2003 12:00:00: 36
## Mean   :3330676                    11/10/2002 12:00:00: 34
## 3rd Qu.:5470157                    09/13/2008 00:00:00: 34
## Max.   :5535195                    08/03/2006 12:00:00: 31
##                                     (Other)          :5744
##      FATALITY_AGE      FATALITY_SEX      FATALITY_LOCATION
## Min.   : 0.00      F :2388      Vehicle/Towed Trailer:1536
## 1st Qu.: 31.00      M :4405      Other :1185
## Median : 51.00      NA's: 122    Outside/Open Areas :1182
## Mean   : 49.97                    Permanent Home :1178
## 3rd Qu.: 70.00                    In Water : 859
## Max.   :102.00                    Mobile/Trailer Home : 402
## NA's   :454                    (Other) : 573
##      EVENT_YEARMONTH
## Min.   :200001
## 1st Qu.:200308
## Median :200508
## Mean   :200540
## 3rd Qu.:200802
## Max.   :201012
## NA's   :82
```

ANÁLISIS DE DATOS

Esta sección analiza los datos del archivo para detectar datos importantes y datos anómalos de cada columna.

Análisis de los datos de la columna EDAD

- Obtenemos un Resumen de "EDAD"

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.00   31.00   51.00   49.97   70.00   102.00   454
```

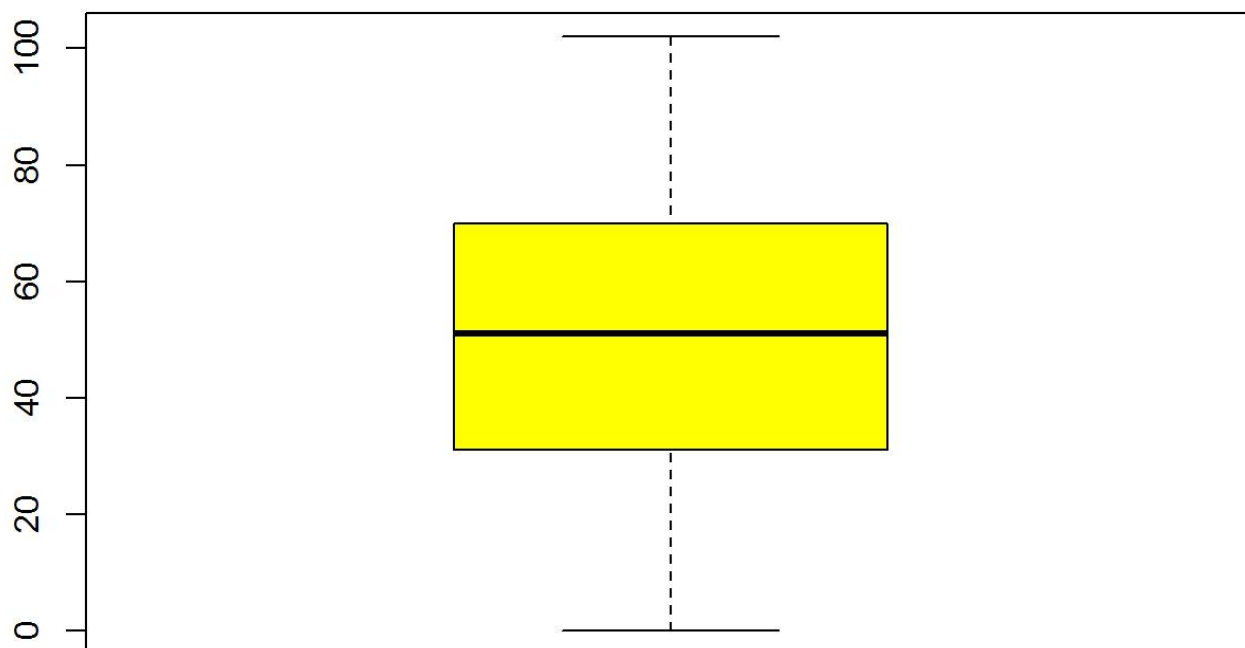
- Obtenemos la Media de los valores perdidos (NA's) de "EDAD"

```
## [1] 0.06565437
```

```
## La proporción de valores perdidos (NA's) de la columna EDAD es relativamente baja
```

- Graficamos los valores de "EDAD" (Gráfica de caja)

GRÁFICA DE CAJA DE 'EDAD'



Análisis de los datos de la columna SEXO

- Obtenemos un Resumen de “SEXO”

```
##      F      M NA's
## 2388 4405  122
```

- Obtenemos la Media de los valores perdidos (NA's) de “SEXO”

```
## [1] 0.01764281
```

```
## La proporcion de valores perdidos (NA's) de la columna SEXO es relativamente baja
```

Análisis de los datos de la columna LOCALIDAD

- Obtenemos un Resumen de “LOCALIDAD”

```
##      Ball Field      Boat
##      31             74
##      Business      Camping
##      41             59
##      Golfing       In Water
##      23            859
##      Long Span Roof  Mobile/Trailer Home
##      27            402
##      Other         Outside/Open Areas
##      1185           1182
##      Permanent Home      School
##      1178             14
##      Under Tree    Vehicle/Towed Trailer
##      173           1536
## Heavy Equipment/Construction      Boating
##      14             67
##      Permanent Structure      Telephone
##      48             1
##      Church
##      1
```

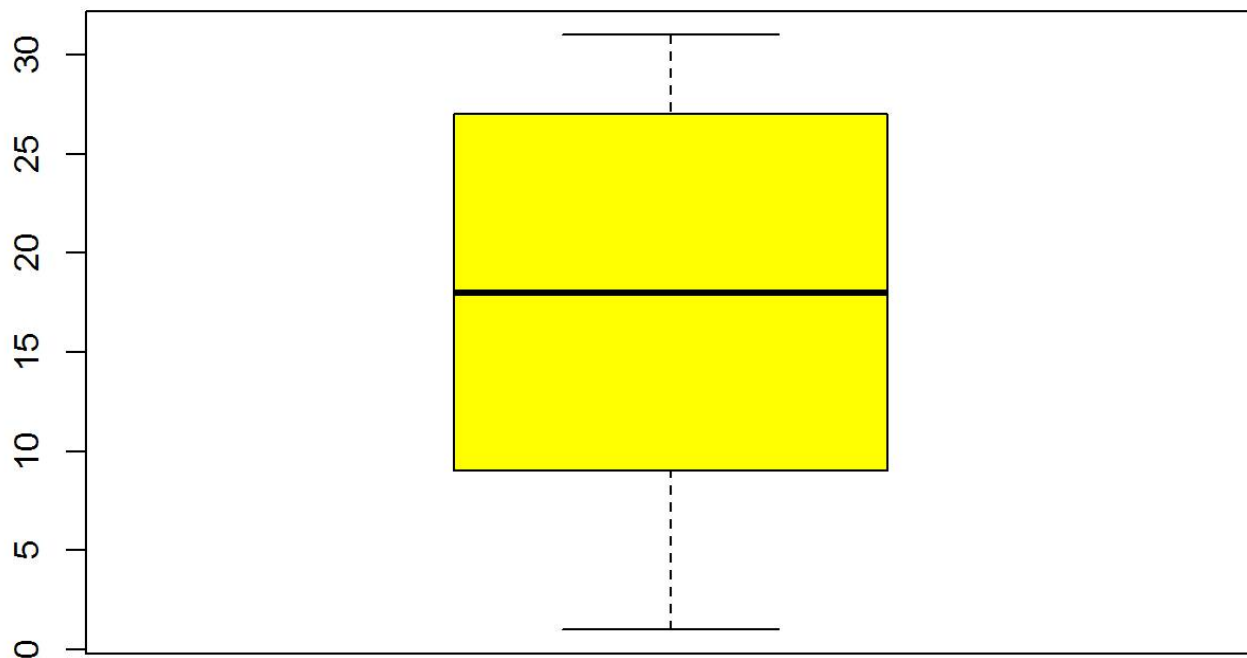
```
## Como observamos no hay valores perdidos (NA's) de la columna LOCALIDAD
```

Análisis de los datos de la columna DIA

- Obtenemos un Resumen de “DIA”

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.0      9.0     18.0    17.1    27.0    31.0
```

```
## Como observamos no hay valores perdidos (NA's) de la columna DIA
```

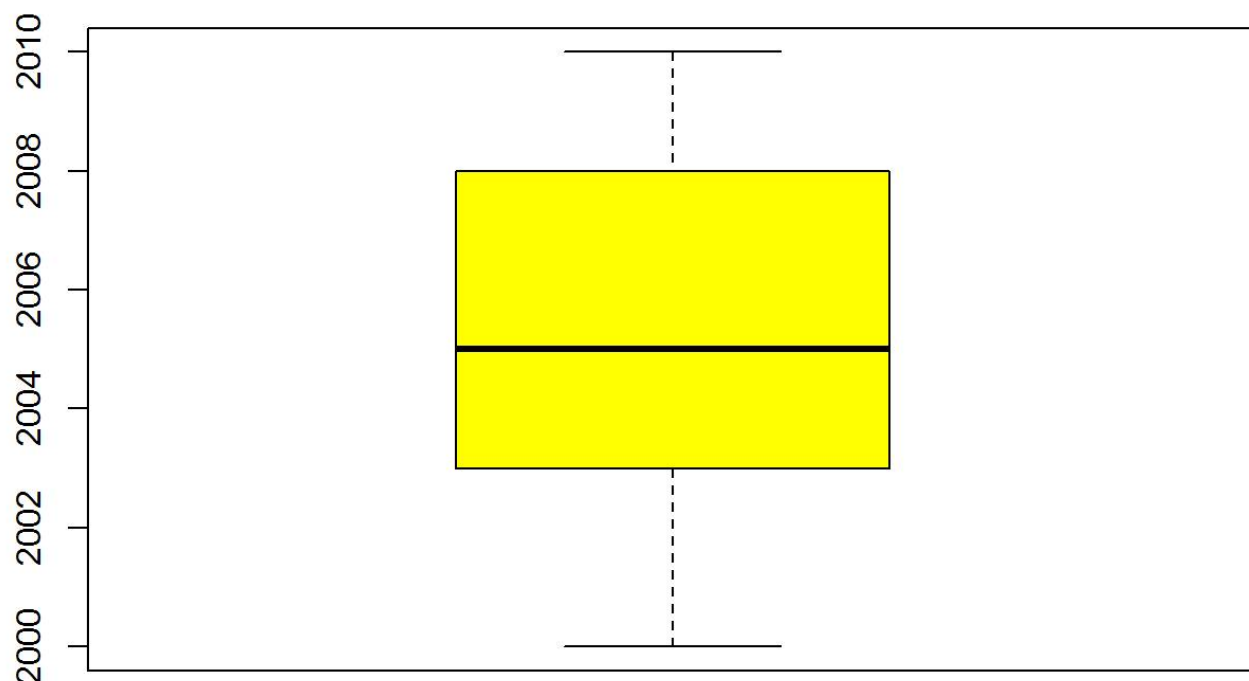
- Graficamos los valores de “DIA” (Gráfica de caja)**GRÁFICA DE CAJA DE 'DIA'****Análisis de los datos de la columna AÑO-MES**

****_Separamos el Mes y el Año de la columna AÑO-MES****

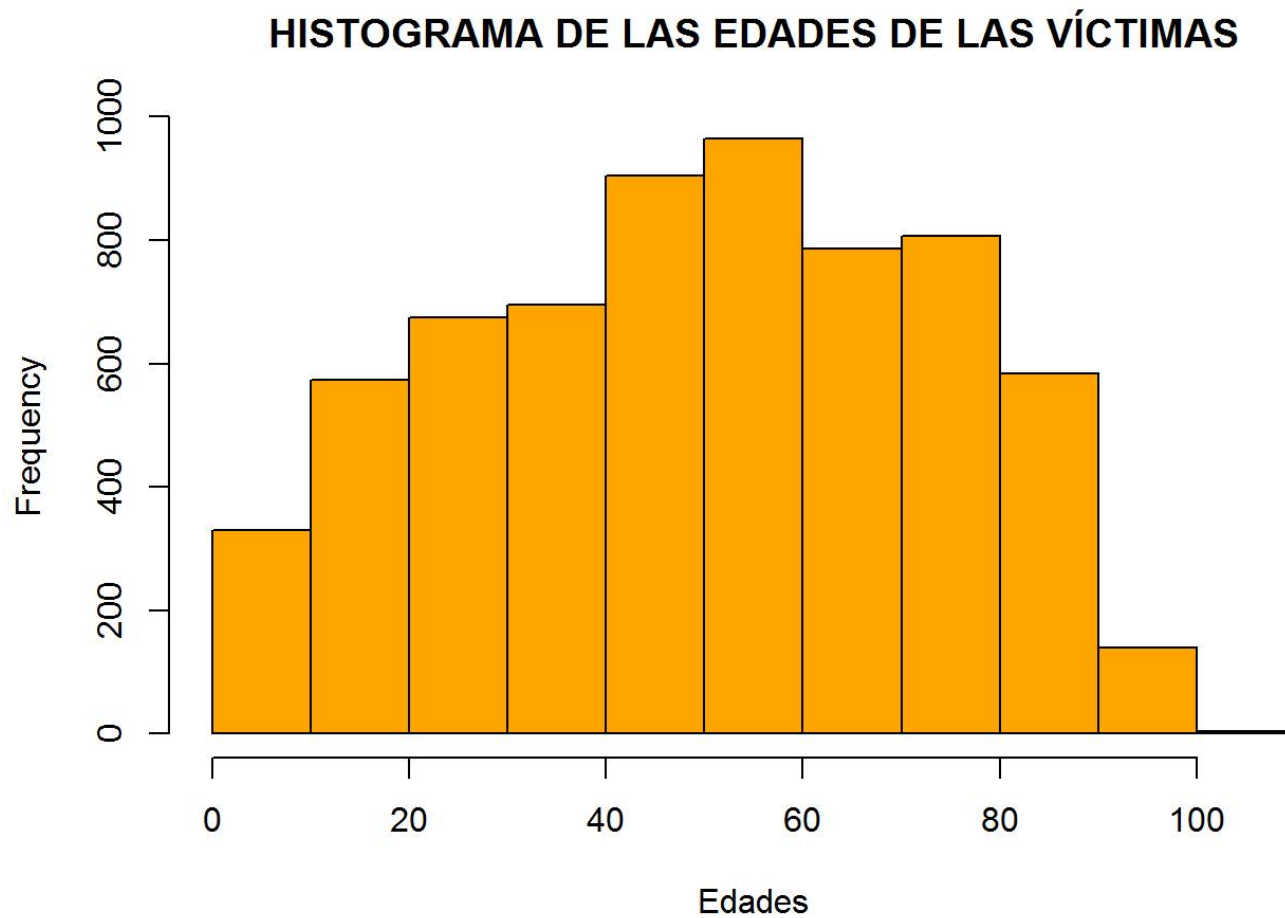
- Obtenemos un Resumen de “AÑO”

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	2000	2003	2005	2005	2008	2010

Como observamos no hay valores perdidos (NA's) de la columna AÑO

- Graficamos los valores de "AÑO" (Gráfica de caja)**GRÁFICA DE CAJA DE 'AÑO'****RESULTADO DE LOS DATOS**

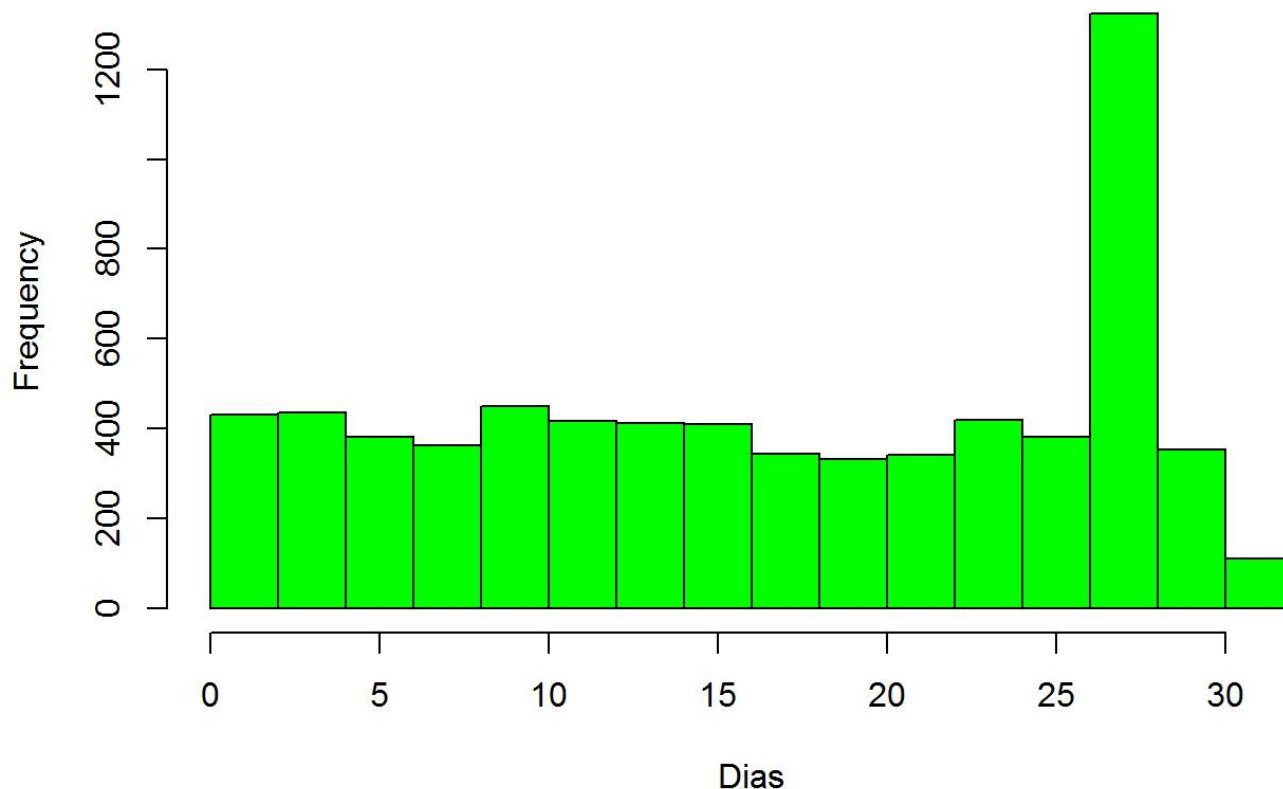
Esta sección vamos a revisar los datos resultantes de las extracciones que realizamos en las secciones anteriores.

- HISTOGRAMA DE LOS EDADES DE LAS VÍCTIMAS

Como resultado observamos que el mayor rango de edad donde ocurren los decesos están entre los 40 y 60 años

- HISTOGRAMA DE LOS DIAS DE LOS TORMENTAS

HISTOGRAMA DE LAS DIAS EN QUE OCURREN LAS TORMENTAS



Como resultado observamos que lo días donde ocurren las Tormentas están entre los días 26 y 28 de cada mes, nos podemos dar cuenta que es muy notable y frecuente los eventos en estos días.

CONCLUSIONES

EL lenguaje de programación R nos ofrece una gran cantidad de herramientas para automatizar nuestro análisis exploratorio de datos. En mi caso, ya que actualmente estoy empezando a desarrollar mi tema de tesis, me encuentro en el punto de que puedo obtener información, archivos y bases de datos gratuitas de repositorios de internet, pero que simple vista es mucha información y no se que podría servirme y que no. En ésta práctica podemos realizar procesos para visualizar en un panorama general la información de aquellos archivos, poder detectar rápidamente cuáles de ellos contienen información en un primer plano se encuentra completa, por ejemplo en esta práctica vimos rápidamente aquellos campos que tenías datos perdidos. Además R ofrece la facilidad de que nuestros procesos se ejecuten automáticamente y se presenten en archivos finales como word, pdf y a través de un explorador web, ésto nos ahorra tiempo en el diseño y nos ayuda en las presentaciones y avances de nuestros trabajos e investigaciones.

BIBLIOGRAFÍA

[1] R MARKDOWN, Reference guide. [Online]. Available: <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf> (<https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>) [Accessed: 11-Ene-2016].

[2] R MARKDOWN, Reference guide. [Online]. Available: <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf> (<https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>) [Accessed: 11-Ene-2016].

[3] Markdown & Pygments Lexers Cheat Sheet. [Online]. Available: <http://joedicastro.com/pages/markdown.html#mark4> (<http://joedicastro.com/pages/markdown.html#mark4>) [Accessed: 11-ENE-2016].

[4] RUG Barcelona. [Online]. Available: <https://rugbcn.wordpress.com/2012/10/08/963/> (<https://rugbcn.wordpress.com/2012/10/08/963/>) [Accessed: 11-Ene-2016].