Preparación

## PREPARACIÓN

### FASE I: Clasificación entre: ALIVE, DEATH, NO-DATA y MD/GR

* **En esta fase, se ha realizado una clasificacion de los datos dados según4 resultados finales:**
  + Fallecidos o con discapacidades severas (SD-D)
  + Con discapacidad moderada o buena recuperacion (MR-GR)
  + Vivos (pero sin resultados finales)
  + Sin datos
* **Para este procesado se han tenido en cuenta principalmente las siguientes variables:**
  + EO\_Outcome
  + EO\_Symptoms
  + TH\_Outcome
  + TH\_Symptoms
  + GOS5
  + GOS8

#### Cuando las variables de GOS5 y GOS8 **tienen** datos

* Si las filas ya contenian datos en las columnas de GOS5 y GOS8, directamente se han clasificado -segun estas variables-. De lo contrario, se ha tenido que analizar las otras variables.

head(datos.modelo[,c(17,18,27,28,29,30)])

## EO\_Outcome EO\_Symptoms TH\_Outcome TH\_Symptoms GOS5 GOS8  
## 1 4 1 NA NA <NA> <NA>  
## 2 4 3 NA NA <NA> MD+  
## 3 4 2 NA NA SD\* <NA>  
## 4 4 2 NA NA <NA> GR+  
## 5 4 1 NA NA <NA> <NA>  
## 6 4 2 NA NA <NA> SD-

#### Cuando las variables de GOS5 y GOS8 **no tienen** datos

* Si las variables de Outcome contenian el valor de 1 (fallecimiento) o las variables de Symptoms contenian el valor de 6, directamente esas filas del dataset pasaban a clasificarse como fallecidos.

## EO\_Outcome EO\_Symptoms TH\_Outcome TH\_Symptoms GOS5 GOS8 TH\_Cause  
## 22 1 6 NA NA <NA> <NA> NA  
## 38 1 6 NA NA <NA> <NA> NA  
## 50 1 6 NA NA <NA> <NA> NA  
## 55 1 6 NA NA <NA> <NA> NA  
## 61 1 6 NA NA <NA> <NA> NA  
## 85 1 6 NA NA <NA> <NA> NA

* Si las variables de Outcome contenian el valor de 4 (alta) y las de Symptoms el valor de 1, entonces se han clasificado como “Vivos (pero sin resultados finales)”.

## EO\_Outcome EO\_Symptoms TH\_Outcome TH\_Symptoms GOS5 GOS8 TH\_Cause  
## 1 4 1 NA NA <NA> <NA> NA  
## 5 4 1 NA NA <NA> <NA> NA  
## 18 4 1 NA NA <NA> <NA> NA  
## 20 4 1 NA NA <NA> <NA> NA  
## 36 4 1 NA NA <NA> <NA> NA  
## 51 4 1 NA NA <NA> <NA> NA

* Se clasificaran como “Sin datos” todas aquellas filas que no contengan valores ni en las columnas de Symptomps. Se tienen en cuenta los transferidos a otros hospitales.

## EO\_Outcome EO\_Symptoms TH\_Outcome TH\_Symptoms GOS5 GOS8 TH\_Cause  
## 384 NA NA NA NA <NA> <NA> NA  
## 417 NA NA NA NA <NA> <NA> NA  
## 985 NA NA NA NA <NA> <NA> NA  
## 997 NA NA NA NA <NA> <NA> NA  
## 2270 NA NA NA NA <NA> <NA> NA  
## 2292 NA NA NA NA <NA> <NA> NA

* Si los Symptoms son de 4 o de 5 (Discapacidad Severa), entonces se clasificaran como “Fallecidos o con discapacidades severas”.

## EO\_Outcome EO\_Symptoms TH\_Outcome TH\_Symptoms GOS5 GOS8 TH\_Cause  
## 160 5 5 NA NA <NA> <NA> NA  
## 241 5 5 NA NA <NA> <NA> NA  
## 317 5 5 NA NA <NA> <NA> NA  
## 336 5 5 NA NA <NA> <NA> NA  
## 357 5 5 NA NA <NA> <NA> NA  
## 573 5 5 NA NA <NA> <NA> NA

* Así mismo, si las variables de Outcome contenían el valor de 4 y las de Symptoms el valor de 9, significa que el paciente ha sido dado de alta, pero no se tiene ningún dato sobre el estado final, por lo tanto, se han incluido en la clasificación de “Vivos (pero sin resultados finales)”.

## EO\_Outcome EO\_Symptoms TH\_Outcome TH\_Symptoms GOS5 GOS8 TH\_Cause  
## 409 4 9 NA NA <NA> <NA> NA  
## 1000 4 9 NA NA <NA> <NA> NA  
## 4859 4 9 NA NA <NA> <NA> NA

* Se han visto 3 elementos de NODATA, cuyos pacientes obtienen un estado de symptoma 4, por lo que se envía a estado de fallecido, son datos anómalos.

## EO\_Outcome EO\_Symptoms TH\_Outcome TH\_Symptoms GOS5 GOS8 TH\_Cause  
## 52 2 4 NA 4 <NA> <NA> 3  
## 699 2 4 NA 4 <NA> <NA> NA  
## 3025 2 4 NA 4 <NA> <NA> 1

#### DATOS FINALES:

* Fallecidos o con discapacidades severas

## [1] 3559

* Con discapacidad moderada o buena recuperación

## [1] 5997

* Vivos (pero sin resultados finales)

## [1] 127

* Sin datos

## [1] 86

* Con NA

## [1] 239

### FASE II: Clasificación entre: ESCANEADOS y NO ESCANEADOS

* En primer lugar, se han encontrado ciertos datos anómalos, en los que aparecen datos escaneados (1) y no tienen los datos del escáner, entonces deberíamos ponerlo como no escaneado (2).

## EO\_Head.CT.scan EO\_1.or.more.PH EO\_Subarachnoid.bleed  
## 201 1 NA NA  
## 314 1 NA NA  
## 1277 1 NA NA  
## 3234 1 NA NA  
## 3687 1 NA NA  
## 4256 1 NA NA  
## EO\_Obliteration.3rdVorBC EO\_Midline.shift..5mm EO\_Non.evac.haem  
## 201 NA NA NA  
## 314 NA NA NA  
## 1277 NA NA NA  
## 3234 NA NA NA  
## 3687 NA NA NA  
## 4256 NA NA NA  
## EO\_Evac.haem  
## 201 NA  
## 314 NA  
## 1277 NA  
## 3234 NA  
## 3687 NA  
## 4256 NA

* A continuación se van a clasificar los datos como:
  + Escaneados
  + No escaneados
  + En analisis
* Si el “Outcome” es 2 (el paciente se ha transferido a otro hospital), se ha escaneado en dicho hospital (“TH\_SCAN”) y no se tiene ninguna información en los escáneres, se clasificaran como “En análisis”.

## EO\_Outcome TH\_Head.CT.scan TH\_1.or.more.PH TH\_Subarachnoid.bleed  
## 52 2 <NA> NA NA  
## 128 2 <NA> NA NA  
## 135 2 <NA> NA NA  
## 188 2 <NA> NA NA  
## 193 2 <NA> NA NA  
## 207 2 <NA> NA NA

* Sobre el dataset **NO ESCANEADO**: Si el “Outcome” es 2 (el paciente se ha transferido a otro hospital) y no se ha realizado ningún escáner pero si contiene datos en el escáner, entonces se clasificara como “Escaneado”.

## EO\_Outcome TH\_Head.CT.scan TH\_1.or.more.PH TH\_Subarachnoid.bleed  
## 201 2 1 2 2  
## 217 2 1 2 1  
## 257 2 1 1 2  
## 314 2 1 1 2  
## 318 2 1 2 2  
## 1184 2 1 2 2

* Sobre el dataset **NO ESCANEADO**: Nos hemos dado cuenta que existen datos anómalos, que contienen varios escáneres, pero sin embargo, no se indica como escaneado, son los registros: 2628,3276,3279,8469,8655, etc. (En total son 12)

## EO\_Head.CT.scan EO\_1.or.more.PH EO\_Subarachnoid.bleed  
## 2628 2 2 2  
## 3276 2 2 2  
## 3279 2 2 2  
## 3720 2 2 2  
## 7286 2 2 2  
## 8469 2 2 2

* Sobre el dataset **EN ANALISIS**: Nos hemos dado cuenta de que existen datos anómalos. Para las variables de los pacientes que se han transferido a otro hospital (TH), existen variables de escáner (“TH\_Head.CT.scan”) que se encuentran vacías, junto con el resto de variables del escáner en particular. Por lo tanto, se ha asignado el valor de 2 a la variable de escáner (“TH\_Head.CT.scan”) y se han incluido en los escaneados, puesto que en todos ellos, en la variable “EO\_Head.CT.scan” si que existe un valor de 1 (escaneados) y no se han encontrado mas anomalías en dichos datos.

## EO\_Head.CT.scan EO\_1.or.more.PH EO\_Outcome TH\_Head.CT.scan  
## 681 1 2 2 <NA>  
## 1639 1 2 2 <NA>  
## 5743 1 2 2 <NA>  
## 8434 1 2 2 <NA>  
## 8972 1 2 2 <NA>  
## TH\_1.or.more.PH TH\_Subarachnoid.bleed  
## 681 NA NA  
## 1639 NA NA  
## 5743 NA NA  
## 8434 NA NA  
## 8972 NA NA

* Sobre el dataset **ESCANEADO**: Se van a eliminar todas las filas que no tengan información en el “TH\_Major.EC.injury” y en el “EO\_Major.EC.injury”

## EO\_Outcome TH\_Major.EC.injury  
## 76 2 NA  
## 90 2 NA  
## 315 2 NA  
## 361 2 NA  
## 510 2 NA  
## 565 2 NA

* Sobre el dataset **ESCANEADO**: Comprobamos que las variables: “EO\_Cause”" y “EO\_Symptoms”, no contengan valores nulos.

## EO\_Cause EO\_Major.EC.injury  
## 177 3 2  
## 211 NA 1  
## 242 NA 2  
## 255 2 2  
## 293 NA 2  
## 321 NA 1

* Sobre el dataset **ESCANEADO**: Comprobamos que las variable: “EO\_Outcome”" no se encuentre nula.(En total son 2 registros).

## EO\_Cause EO\_Outcome  
## 9036 3 NA  
## 9333 3 NA

* Sobre el dataset **ESCANEADO**: Comprobamos que existe un valor anómalo (que se sale del rango) en un registro en la columna de “EO\_Major.EC.Injury”. Este valor lo cambiaremos a positivo -> 1

## EO\_Cause EO\_Major.EC.injury  
## 3862 2 -1

#### DATOS FINALES:

* Vivos y escaneados

## [1] 4157

* Vivos y no escaneados

## [1] 1535

* Vivos en análisis

## [1] 305

* Fallecidos y escaneados

## [1] 2829

* Fallecidos y no escaneados

## [1] 439

* Fallecidos en análisis

## [1] 291

### FASE III: Eliminación y centralización de variables

* Se va a centralizar las variables de “PUPIL\_REACT\_LEFT”" y “PUPIL\_REACT\_RIGHT”"

## PUPIL\_REACT\_LEFT PUPIL\_REACT\_RIGHT ESTADOESCANER  
## 1 1 1 SCANEADO  
## 2 1 1 SCANEADO  
## 3 1 1 SCANEADO  
## 4 1 1 SCANEADO  
## 5 1 1 SCANEADO  
## 6 1 1 SCANEADO

+ Both reactive

## [1] 5662

+ No response unilateral

## [1] 497

+ No response

## [1] 634

+ Unable to assess

## [1] 193

* Ahora vamos a ver si podemos prescindir o aunar las variables de “EO\_Cause” y “TH\_Cause”. Para ello veremos en que caso, ambas variables difieren:

## EO\_Cause TH\_Cause  
## 2307 1 3  
## 2813 3 1  
## 3285 2 3  
## 4021 3 1

* Como se puede observar, podríamos prescindir de la variable “TH\_Cause”, puesto que recoge la misma información que “EO\_Cause”.
* A continuación vamos a aunar todas las variables del escáner. Si un paciente ha sido transferido a otro hospital y se han realizado los escáneres en dicho hospital, entonces, se mantendrán los últimos valores del escáner. En caso contrario, se usaran los primeros resultados de escáner obtenidos en el primer. Además, eliminaremos todas las variables que no se utilicen.

## sex age cause ec eye motor verbal pupils phm sah oblt mdls hmt outcome  
## 1 0 11 1 1 1 5 1 1 2 2 2 2 2 0  
## 2 0 14 1 2 1 2 1 1 1 2 2 2 1 1  
## 3 0 14 1 2 2 5 1 1 2 2 2 2 1 1  
## 4 0 14 1 2 2 5 2 1 2 2 2 2 2 0  
## 5 0 14 3 2 4 6 4 1 2 1 2 2 2 0  
## 6 0 15 1 2 1 5 1 1 2 2 2 2 2 1