

# TPMineria

Bodean, Emiliano & Ojeda, Zacarías

## 1 Resumen

## 2 Introducción

## 3 Procesamiento de Datos

### 3.1 Lectura de Datasets

Cargamos los datasets originales

```
sentencias_1c <- read_csv("./sentencias_1c.csv") %>%  
  filter(is.na(mat) | toupper(mat)=="C") %>% # solo incluir materia civil  
  select(-mat)
```

```
organismos <- read_csv("./organismos.csv")
```

### 3.2 Inspección y “limpieza” de los datos

Inspección de los datos

```
sentencias_1c %>%  
  mostrar(caption = "Sentencias Primera Instancia Original")
```

Table 1. Sentencias Primera Instancia Original

nro	tproc	as	econ	finicio	fdesp	fvenc1	fvenc2	fres	tres	justiciales	reccap	lep
12858	RESTRICCIONES A LA CAPACIDAD	S	0	15/09/2015	28/11/2017	22/12/2017	19/02/2018	29/12/2017	7	1	0	jdofam0002gch
11852/5	INCIDENTE	S	1	15/04/2015	19/09/2017	04/10/2017	19/10/2017	29/12/2017	6	3	0	jdofam0002gch
12237	ORDINARIO FILIACION E INDEMNIZACION DE DAÑOS	S	1	03/12/2014	12/10/2017	13/12/2017	19/03/2018	29/12/2017	7	1	0	jdofam0002gch
14440	MEDIDA CAUTELAR (FAMILIA)	S	0	21/04/2017	29/11/2017	18/12/2017	02/02/2018	29/12/2017	7	1	0	jdofam0002gch
11507	ORDINARIO DAÑOS Y PERJUICIOS	S	1	13/03/2014	30/11/2017	02/02/2018	06/04/2018	29/12/2017	7	2	0	jdofam0002gch
8133	ORDINARIO FILIACION E INDEMNIZACION DE DAÑOS	S	1	17/06/2010	16/03/2017	17/05/2017	28/07/2017	29/12/2017	7	2	0	jdofam0002gch

```
organismos %>%  
  mostrar(caption = "Organismos")
```

Quitamos los tipos de procesos “Monitorios”, ya que son de mero trámite y no interesan en el análisis

```
sentencias_1c <- sentencias_1c %>%  
  filter(!grepl("MONITORIO", tproc))
```

Table 2. Organismos

X1	organismo	organismo_descripcion	email_oficial	fuero	circunscripcion	localidad	categoria	tipo	materia
1	jdocco0000dia	Jdo Civ y Com Lab	jdccyc-dia@jusertreros.gov.ar	Civil y Comercial	Diamante	Diamante	NA	jdo	cco lab
2	jdocco0000fed	Jdo Civ y Com Lab Fam	jdccyc-fcion@jusertreros.gov.ar	Civil y Comercial	Federación	Federación	NA	jdo	cco fam lab
3	jdocco0000fcl	Jdo Civ y Com Lab	jdccyc-fcl@jusertreros.gov.ar	Civil y Comercial	Federal	Federal	NA	jdo	cco lab
4	jdocco0000ssa	Jdo Civ y Com Lab Fam	jdccyclab-ssdor@jusertreros.gov.ar	Civil y Comercial	San Salvador	San Salvador	NA	jdo	cco fam lab
5	jdocco0000tal	Jdo Civ y Com -comp.Laboral	jdccyc-tala@jusertreros.gov.ar	Civil y Comercial	Tala	Rosario del Tala	NA	jdo	cco lab
6	jdocco0000vic	Jdo Civ y Com -comp.Laboral	jdccyc-vic@jusertreros.gov.ar	Civil y Comercial	Victoria	Victoria	NA	jdo	cco lab

Calcula duracion como Fecha de Resolucion - Fecha de inicio

```

sentencias_1c <- sentencias_1c %>%
  mutate(finicio = lubridate::dmy(finicio)) %>%
  mutate(fres = lubridate::dmy(fres)) %>%
  mutate(duracion = fres - finicio)

sentencias_1c %>%
  mostrar(caption = "Sentencias con duración")

```

Table 3. Sentencias con duración

nro	tproc	as	con	finicio	fdesp	fvenc1	fvenc2	fres	tres	justiciales	reccap	lep	duracion
12858	RESTRICCIONES A LA CAPACIDAD	S	0	2015-09-15	28/11/2017	22/12/2017	19/02/2018	2017-12-29	7	1	0	jdofam0002gch	836 days
11852	INCIDENTE	S	1	2015-04-15	19/09/2017	04/10/2017	19/10/2017	2017-12-29	6	3	0	jdofam0002gch	989 days
12237	ORDINARIO FILIACION E INDEMNIZACION DE DAÑOS	S	1	2014-12-03	12/10/2017	13/12/2017	19/03/2018	2017-12-29	7	1	0	jdofam0002gch	1122 days
14440	MEDIDA CAUTELAR (FAMILIA)	S	0	2017-04-21	29/11/2017	18/12/2017	02/02/2018	2017-12-29	7	1	0	jdofam0002gch	252 days
11507	ORDINARIO DAÑOS Y PERJUICIOS	S	1	2014-03-13	30/11/2017	02/02/2018	06/04/2018	2017-12-29	7	2	0	jdofam0002gch	1387 days
8133	ORDINARIO FILIACION E INDEMNIZACION DE DAÑOS	S	1	2010-06-17	16/03/2017	17/05/2017	28/07/2017	2017-12-29	7	2	0	jdofam0002gch	2752 days

Eliminamos filas que tienen datos invalidos de fecha (datos nulos o futuros por error de tipeo) Reemplazamos los datos NA de reccap por cero.

```

sentencias_1c <- sentencias_1c %>%
  filter(!is.na(finicio)) %>%
  filter(!is.na(fres)) %>%
  filter(fres < '2018-09-01', finicio < '2018-09-01')

```

Calcula los cuartiles de duración por cada tipo de proceso (tproc), y se clasifica en rapido / normal / demorado si duracion es mayor a media de tproc utilizando one hot encoding.

```

demora <- sentencias_1c %>%
  group_by(tproc) %>%
  summarise(techo_rapido=quantile(duracion, probs=0.25),
            piso_demorado=quantile(duracion, probs=0.75))

sentencias_1c <- sentencias_1c %>%
  left_join(demora, by="tproc") %>%
  mutate(rapido = duracion <= techo_rapido) %>%

```

```

mutate(normal = duracion > techo_rapido & duracion < piso_demorado) %>%
mutate(demorado = duracion >= piso_demorado) %>%
select(-duracion, -techo_rapido, -piso_demorado) # quitando columnas temporales

sentencias_1c %>%
  mutate(tproc = str_trunc(tproc, 20))%>%
  mostrar(caption = "Agregando columnas demora")

```

Table 4. Agregando columnas demora

nro	tproc	as	ccon	finicio	fdesp	fvenc1	fvenc2	fres	tres	justiciables	reccap	iep	rapido	normal	demorado
12858	RESTRICCIONES A L...	S	0	2015-09-15	28/11/2017	22/12/2017	19/02/2018	2017-12-29	7	1	0	jdofam0002gch	FALSE	TRUE	FALSE
11852/5	INCIDENTE	S	1	2015-04-15	19/09/2017	04/10/2017	19/10/2017	2017-12-29	6	3	0	jdofam0002gch	FALSE	FALSE	TRUE
12237	ORDINARIO FILIACL...	S	1	2014-12-03	12/10/2017	13/12/2017	19/03/2018	2017-12-29	7	1	0	jdofam0002gch	FALSE	TRUE	FALSE
14440	MEDIDA CAUTELAR (...)	S	0	2017-04-21	29/11/2017	18/12/2017	02/02/2018	2017-12-29	7	1	0	jdofam0002gch	FALSE	FALSE	TRUE
11507	ORDINARIO DAÑOS Y...	S	1	2014-03-13	30/11/2017	02/02/2018	06/04/2018	2017-12-29	7	2	0	jdofam0002gch	FALSE	TRUE	FALSE
8133	ORDINARIO FILIACL...	S	1	2010-06-17	16/03/2017	17/05/2017	28/07/2017	2017-12-29	7	2	0	jdofam0002gch	FALSE	FALSE	TRUE

Agrega datos de organismos para tenerlos separados por columna, actualmente se encuentra en columna iep.

```

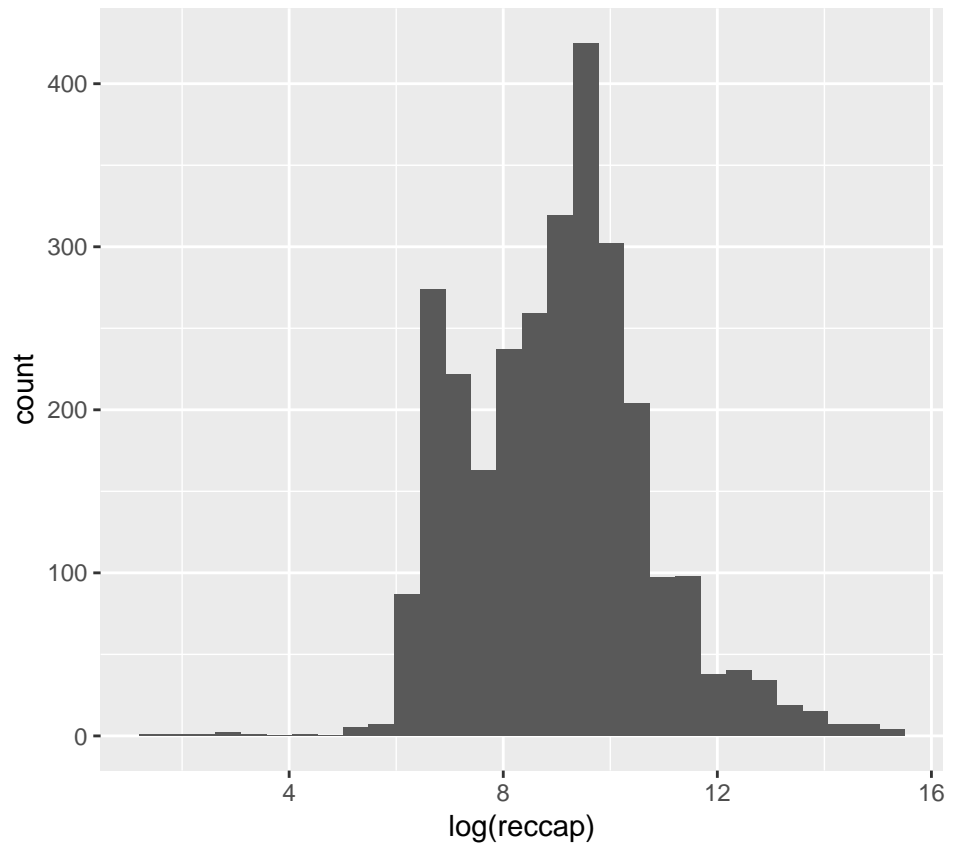
organismos <- organismos %>%
  select(organismo, circunscripcion, localidad, materia)

sentencias_1c <- sentencias_1c %>%
  left_join(organismos, by = c('iep'='organismo'))

```

Exploremos la variable capital reclamado para definir los rangos

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Calculamos los cuartiles para ver si nos sirven para parametrizar (reccap)

```
print('1º Curtil:')
## [1] "1º Curtil:"
quantile(pull(sentencias_1c[, 'reccap']), .25, na.rm = TRUE)
## 25%
## 0

print('2º Curtil:')
## [1] "2º Curtil:"
quantile(pull(sentencias_1c[, 'reccap']), .50, na.rm = TRUE)
## 50%
## 0

print('3º Curtil:')
```

```
## [1] "3º Curtil:"

quantile(pull(sentencias_1c[, 'reccap']), .75, na.rm = TRUE)

## 75%
## 0

#View(sentencias_1c)
```

Como los curtiles son cero, elimino los ceros y vuelvo a calcular los cuartiles.

```
#reccap_not_cero <- which(sentencias_1c$reccap != 0)

print('1º Curtil:')

## [1] "1º Curtil:"

quantile(which(sentencias_1c$reccap != 0), .25)

## 25%
## 3181.5

print('2º Curtil:')

## [1] "2º Curtil:"

quantile(which(sentencias_1c$reccap != 0), .50)

## 50%
## 6271.5

print('3º Curtil:')

## [1] "3º Curtil:"

quantile(which(sentencias_1c$reccap != 0), .75)

## 75%
## 8948.75

capmedio <- mean(pull(sentencias_1c[, 'reccap']))

sentencias_1c <- sentencias_1c %>%
  mutate(reccap_0 = reccap == 0) %>%
  mutate(reccap_1 = (reccap < quantile(which(sentencias_1c$reccap != 0), .25)) & (reccap != 0))
  mutate(reccap_2 = (reccap >= quantile(which(sentencias_1c$reccap != 0), .25)) & (reccap != 0))
  mutate(reccap_3 = (reccap >= quantile(which(sentencias_1c$reccap != 0), .50)) & (reccap != 0))
  mutate(reccap_4 = (reccap >= quantile(which(sentencias_1c$reccap != 0), .75)))
```

Separo la columna justiciables en 6 rangos para poder aplicar apriori.

```

sentencias_1c <- sentencias_1c %>%
  mutate(justiciales0_1 = justiciables < 2) %>%
  mutate(justiciales2_3 = (justiciables > 1) & (justiciables < 4)) %>%
  mutate(justiciales4_5 = (justiciables > 3) & (justiciables < 6)) %>%
  mutate(justiciales6_7 = (justiciables > 5) & (justiciables < 8)) %>%
  mutate(justiciales8_9 = (justiciables > 7) & (justiciables < 10)) %>%
  mutate(justiciales10_N = justiciables > 9)

```

Separamos Localidad y Circunscripcion en columnas.

```

sentencias_1c <- sentencias_1c %>%
  mutate(localidad = as.factor(localidad))

```

Convertimos columna tproc en categórica, esto es requerido por el algoritmo

```

sentencias_1c <- sentencias_1c %>%
  mutate(tproc = as.factor(tproc)) %>%
  mutate(circunscripcion = as.factor(circunscripcion)) %>%
  mutate(materia = as.factor(materia))

```

Tomamos solamente las columnas tipo booleanos y categóricas.

```

sentencias_final <- sentencias_1c %>%
  select(-nro, -as, -ccon, -finicio, -fres, -fdesp, -fvenc1, -fvenc2, -tres, -justiciabl

sentencias_final %>%
  mostrar(caption = "Tabla final a utilizar en el algoritmo apriori")

```

Table 5. Tabla final a utilizar en el algoritmo apriori

tipo	capdo	normal	denunciado	circunscripcion	materia	frecap_0	frecap_1	frecap_2	frecap_3	frecap_4	justiciab0_1	justiciab0_2	justiciab0_3	justiciab0_4	justiciab0_5	justiciab0_6	justiciab0_7	justiciab0_8	justiciab0_9	justiciab0_10_N
RESTRICCIONES A LA CAPACIDAD	FALSE	TRUE	FALSE	Guadalupechitl	famipen	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
INCIDENTE	FALSE	FALSE	TRUE	Guadalupechitl	famipen	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
ORDINARIO FILIACION E INDEMNIZACION DE DAÑOS	FALSE	TRUE	FALSE	Guadalupechitl	famipen	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
MEDEIA CATELILAR (FAMILIA)	FALSE	FALSE	TRUE	Guadalupechitl	famipen	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
ORDINARIO DAÑOS Y PERJUICIOS	FALSE	TRUE	FALSE	Guadalupechitl	famipen	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
ORDINARIO FILIACION E INDEMNIZACION DE DAÑOS	FALSE	FALSE	TRUE	Guadalupechitl	famipen	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

## 4 Generación de Reglas

```

rules <- apriori(sentencias_final, parameter = list(supp=0.001, conf=0.5, minlen=2), app

## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##           0.5    0.1    1 none FALSE                TRUE         5    0.001    2
## maxlen target  ext

```

```

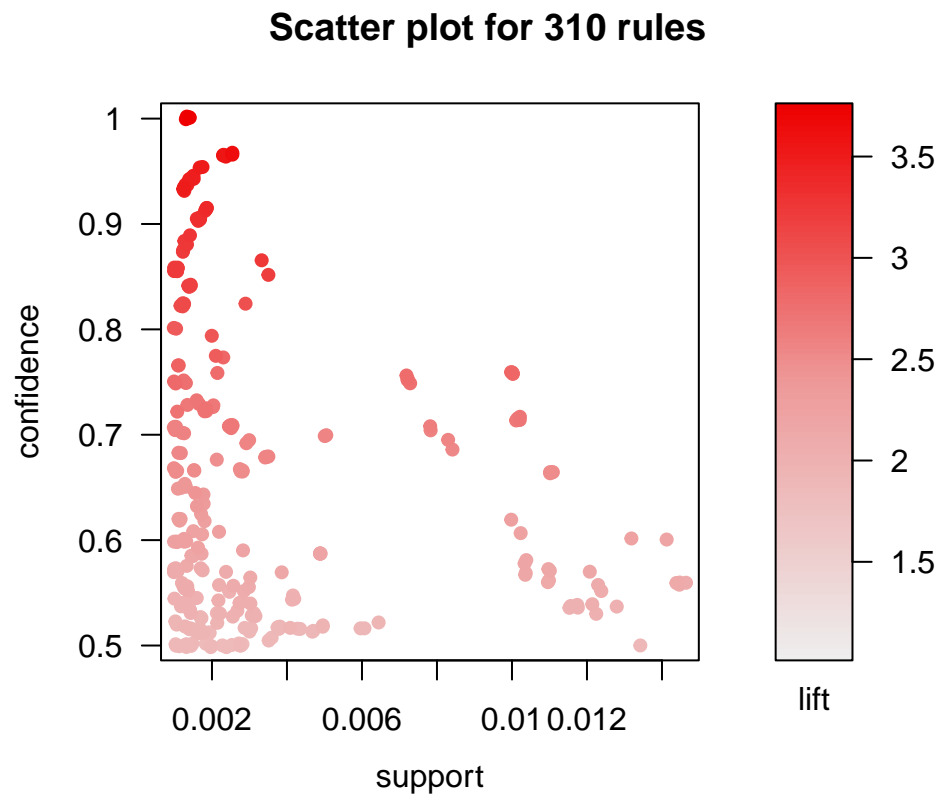
##      10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE FALSE TRUE      2      TRUE
##
## Absolute minimum support count: 11
##
## set item appearances ...[2 item(s)] done [0.00s].
## set transactions ...[275 item(s), 11576 transaction(s)] done [0.01s].
## sorting and recoding items ... [105 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 done [0.00s].
## writing ... [310 rule(s)] done [0.00s].
## creating S4 object ... done [0.01s].

summary(rules)

## set of 310 rules
##
## rule length distribution (lhs + rhs):sizes
##      2      3      4      5      6
##      2  47 126 107  28
##
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      2.000   4.000   4.000   4.361   5.000   6.000
##
## summary of quality measures:
##      support      confidence      lift      count
## Min.      :0.001037   Min.      :0.5000   Min.      :1.837   Min.      : 12.0
## 1st Qu.:0.001296   1st Qu.:0.5385   1st Qu.:2.004   1st Qu.: 15.0
## Median :0.001814   Median :0.6197   Median :2.321   Median : 21.0
## Mean      :0.003352   Mean      :0.6674   Mean      :2.474   Mean      : 38.8
## 3rd Qu.:0.003455   3rd Qu.:0.7582   3rd Qu.:2.840   3rd Qu.: 40.0
## Max.      :0.014599   Max.      :1.0000   Max.      :3.749   Max.      :169.0
##
## mining info:
##      data ntransactions support confidence
## sentencias_final      11576      0.001      0.5

```

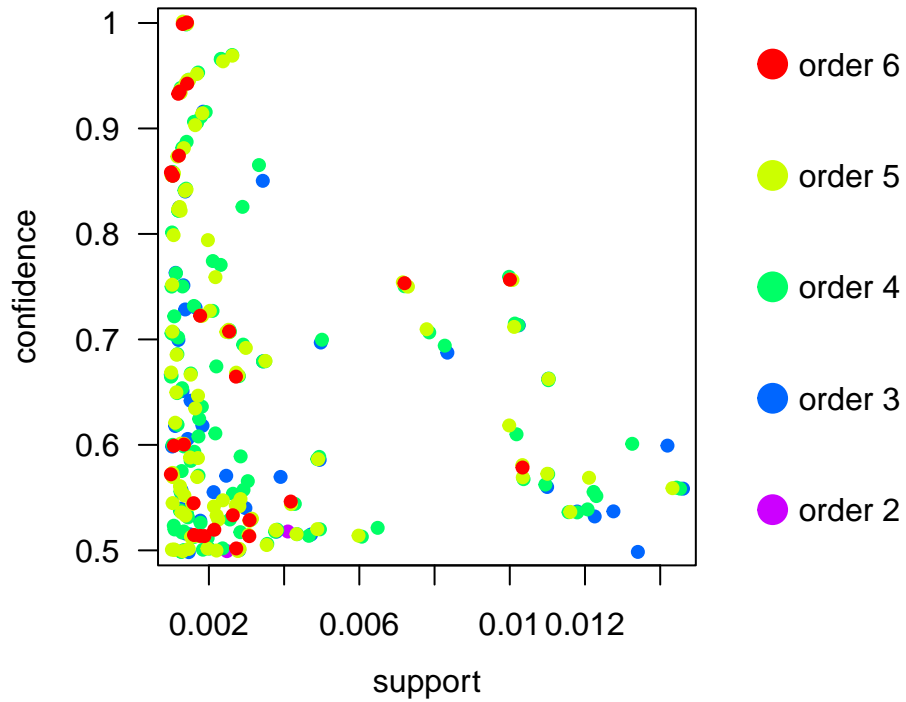
Graficamos



## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



## Two-key plot



Realizamos un inspect de las primeras reglas, vemos que hay reglas que no son de interes.

```
inspect(rules[1:8])
```

##	lhs	rhs	support	confidence
## [1]	{circunscripcion=San Salvador}	=> {rapido}	0.002505183	0.5000000 1.
## [2]	{justiciales10_N}	=> {demorado}	0.004060124	0.5164835 1.
## [3]	{tproc=EJECUTIVO,			
##	circunscripcion=San Salvador}	=> {rapido}	0.001295784	0.5555556 2.
## [4]	{circunscripcion=San Salvador,			
##	materia=paz}	=> {rapido}	0.001814098	0.6176471 2.
## [5]	{circunscripcion=San Salvador,			
##	reccap_0}	=> {rapido}	0.002418798	0.5714286 2.
## [6]	{circunscripcion=San Salvador,			
##	justiciales2_3}	=> {rapido}	0.002159641	0.5555556 2.
## [7]	{tproc=DIVORCIO POR MUTUO CONSENTIMIENTO,			
##	circunscripcion=Guaileguay}	=> {rapido}	0.001641327	0.9047619 3.
## [8]	{materia=cco,			

```
##      justiciables10_N}                                => {demorado} 0.002764340 0.5000000 1.
```

Volvemos a visualizar ordenas por soporte y confianza

```
top.confidence <- sort(rules, decreasing = TRUE, na.last = NA, by = "confidence")
inspect(top.confidence[1:8])
```

	lhs	rhs	support	confidence	lift	count
## [1]	{tproc=ACCION DE AMPARO, circunscripcion=Nogoyá, justiciables0_1}	=> {demorado}	0.001382170	1.0000000	3.748705	16
## [2]	{tproc=ACCION DE AMPARO, circunscripcion=Nogoyá, materia=paz, justiciables0_1}	=> {demorado}	0.001382170	1.0000000	3.748705	16
## [3]	{tproc=ACCION DE AMPARO, circunscripcion=Nogoyá, reccap_0, justiciables0_1}	=> {demorado}	0.001382170	1.0000000	3.748705	16
## [4]	{tproc=EJECUTIVO, circunscripcion=Paraná, reccap_1, justiciables0_1}	=> {rapido}	0.001295784	1.0000000	3.673754	15
## [5]	{tproc=ACCION DE AMPARO, circunscripcion=Nogoyá, materia=paz, reccap_0, justiciables0_1}	=> {demorado}	0.001382170	1.0000000	3.748705	16
## [6]	{tproc=EJECUTIVO, circunscripcion=Paraná, materia=paz, reccap_1, justiciables0_1}	=> {rapido}	0.001295784	1.0000000	3.673754	15
## [7]	{tproc=ACCION DE AMPARO, circunscripcion=Nogoyá, materia=paz}	=> {demorado}	0.002591569	0.9677419	3.627779	30
## [8]	{tproc=ACCION DE AMPARO, circunscripcion=Nogoyá, materia=paz, reccap_0}	=> {demorado}	0.002591569	0.9677419	3.627779	30

```
top.support <- sort(rules, decreasing = TRUE, na.last = NA, by = "support")
inspect(top.support[1:8])
```

	lhs	rhs	support	confidence	lift	count
## [1]	{tproc=APREMIO, reccap_1}	=> {demorado}	0.01459917	0.5577558	2.090862	169

```

## [2] {tproc=APREMIO,
##      reccap_1,
##      justiciables2_3}      => {demorado} 0.01451279 0.5600000 2.099275 168
## [3] {tproc=APREMIO,
##      materia=paz,
##      reccap_1}             => {demorado} 0.01442640 0.5585284 2.093758 167
## [4] {tproc=APREMIO,
##      materia=paz,
##      reccap_1,
##      justiciables2_3}      => {demorado} 0.01434001 0.5608108 2.102314 166
## [5] {tproc=APREMIO,
##      reccap_0}             => {rapido} 0.01416724 0.6007326 2.206944 164
## [6] {circunscripcion=Uruguay,
##      justiciables2_3}      => {demorado} 0.01338977 0.5000000 1.874352 155
## [7] {tproc=APREMIO,
##      reccap_0,
##      justiciables2_3}      => {rapido} 0.01321700 0.6023622 2.212931 153
## [8] {tproc=EJECUTIVO,
##      circunscripcion=Guaileguaychú} => {demorado} 0.01278507 0.5362319 2.010175 148

```

## 5 Resultados / Discusión

...

## 6 Conclusiones

...

## 7 Referencias

...