

# TPMineria

Zacarías F. Ojeda

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")
```

Inspección de los datos

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

Table 1. Sentencias Primera Instancia Original

nro	tproc	as	ccon	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")
```

Table 2. Organismos

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

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))
```

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	econ	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
11852/5	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")

```

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

**Table 4.** Agregando columnas demora

nro	tproc	as	econ	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

```

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

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

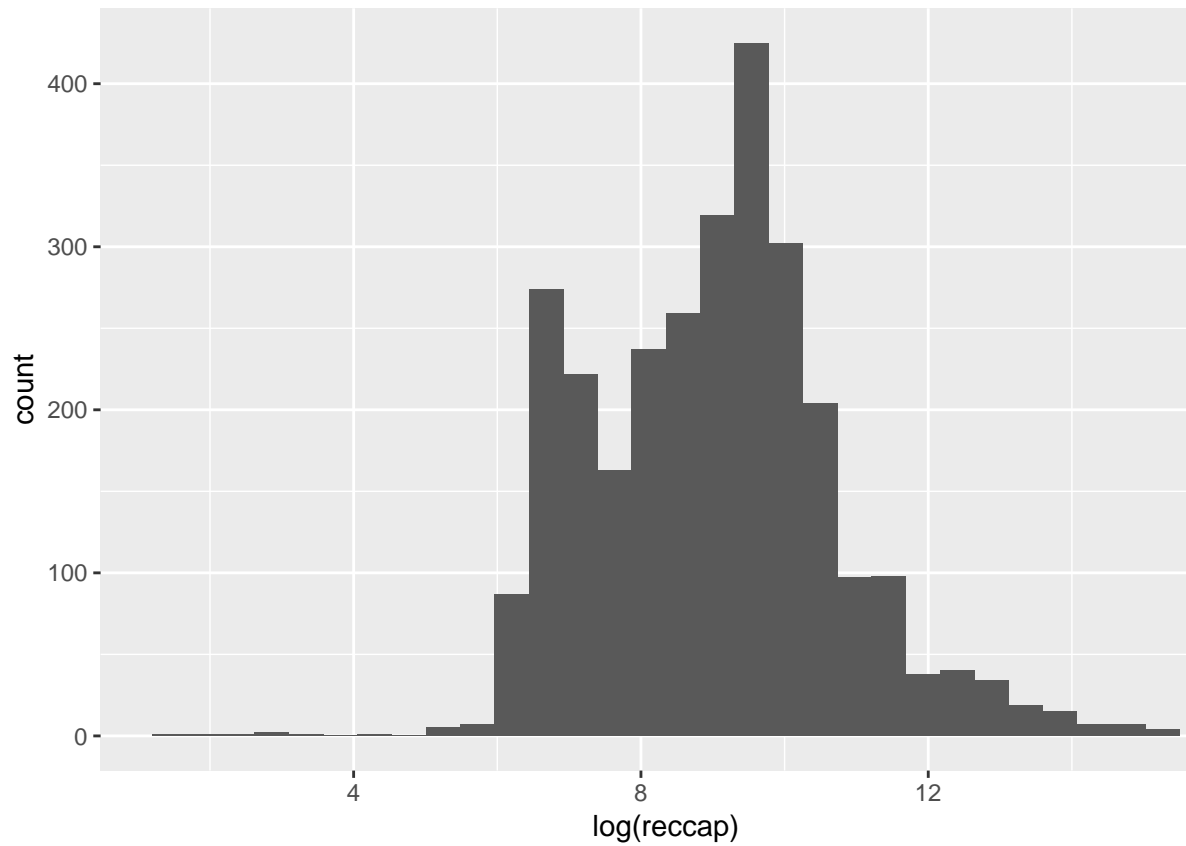
Exploremos la variable capital reclamado para definir los rangos

histograma <- sentencias_1c %>%
  ggplot() +
  geom_histogram(aes(x=log(reccap)))

histograma

## `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:"
```

```
## [1] "3º Curtil:"

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

## 75%
## 0

#View(sentencias_1c)
```

Como dos de 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 4 rangos para poder aplicar apriori.

```
sentencias_1c <- sentencias_1c %>%
  mutate(justiciables0_1 = justiciables < 2) %>%
  mutate(justiciables2_3 = (justiciables > 1) & (justiciables < 4)) %>%
  mutate(justiciables4_5 = (justiciables > 3) & (justiciables < 6)) %>%
  mutate(justiciables6_7 = (justiciables > 5) & (justiciables < 8)) %>%
  mutate(justiciables8_9 = (justiciables > 7) & (justiciables < 10)) %>%
  mutate(justiciables10_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))
```

Tomos 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

[illegible]

```
rules <- apriori(sentencias_final, parameter = list(supp=0.01, conf=0.7, minlen=2), appe

## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.7    0.1   1 none FALSE              TRUE        5    0.01    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: 115
##
## set item appearances ...[2 item(s)] done [0.00s].
## set transactions ...[275 item(s), 11576 transaction(s)] done [0.00s].
## sorting and recoding items ... [46 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 ... [8 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].

summary(rules)

## set of 8 rules
##
## rule length distribution (lhs + rhs):sizes
## 3 4 5 6
## 1 3 3 1
##
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 3.0 4.0 4.5 4.5 5.0 6.0
##
## summary of quality measures:
## support confidence lift count
## Min. :0.01002 Min. :0.7134 Min. :2.674 Min. :116.0
## 1st Qu.:0.01002 1st Qu.:0.7147 1st Qu.:2.679 1st Qu.:116.0
## Median :0.01006 Median :0.7367 Median :2.762 Median :116.5
## Mean :0.01009 Mean :0.7362 Mean :2.760 Mean :116.8
## 3rd Qu.:0.01013 3rd Qu.:0.7582 3rd Qu.:2.842 3rd Qu.:117.2
## Max. :0.01019 Max. :0.7582 Max. :2.842 Max. :118.0
##
## mining info:
## data ntransactions support confidence
## sentencias_final 11576 0.01 0.7

inspect(rules[1:8])

## lhs rhs support confidence lift count
## [1] {circunscripcion=Uruguay, reccap_1} => {demorado} 0.01019350 0.7151515 2.680892 118
## [2] {tproc=APREMIO, circunscripcion=Uruguay, reccap_1} => {demorado} 0.01002073 0.7581699 2.842155 116

```

```

## [3] {circunscripcion=Uruguay,
##      materia=paz,
##      reccap_1} => {demorado} 0.01019350 0.7151515 2.680892 118
## [4] {circunscripcion=Uruguay,
##      reccap_1,
##      justiciables2_3} => {demorado} 0.01010712 0.7134146 2.674381 117
## [5] {tproc=APREMIO,
##      circunscripcion=Uruguay,
##      materia=paz,
##      reccap_1} => {demorado} 0.01002073 0.7581699 2.842155 116
## [6] {tproc=APREMIO,
##      circunscripcion=Uruguay,
##      reccap_1,
##      justiciables2_3} => {demorado} 0.01002073 0.7581699 2.842155 116
## [7] {circunscripcion=Uruguay,
##      materia=paz,
##      reccap_1,
##      justiciables2_3} => {demorado} 0.01010712 0.7134146 2.674381 117
## [8] {tproc=APREMIO,
##      circunscripcion=Uruguay,
##      materia=paz,
##      reccap_1,
##      justiciables2_3} => {demorado} 0.01002073 0.7581699 2.842155 116

```

```

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

```

	lhs	rhs	support	confidence	lift	count
## [1]	{tproc=APREMIO, circunscripcion=Uruguay, reccap_1}	=> {demorado}	0.01002073	0.7581699	2.842155	116
## [2]	{tproc=APREMIO, circunscripcion=Uruguay, materia=paz, reccap_1}	=> {demorado}	0.01002073	0.7581699	2.842155	116
## [3]	{tproc=APREMIO, circunscripcion=Uruguay, reccap_1, justiciables2_3}	=> {demorado}	0.01002073	0.7581699	2.842155	116
## [4]	{tproc=APREMIO, circunscripcion=Uruguay, materia=paz, reccap_1, justiciables2_3}	=> {demorado}	0.01002073	0.7581699	2.842155	116
## [5]	{circunscripcion=Uruguay, reccap_1}	=> {demorado}	0.01019350	0.7151515	2.680892	118



```
## [6] {circunscripcion=Uruguay,
##      materia=paz,
##      reccap_1} => {demorado} 0.01019350 0.7151515 2.680892 118
## [7] {circunscripcion=Uruguay,
##      reccap_1,
##      justiciables2_3} => {demorado} 0.01010712 0.7134146 2.674381 117
## [8] {circunscripcion=Uruguay,
##      materia=paz,
##      reccap_1,
##      justiciables2_3} => {demorado} 0.01010712 0.7134146 2.674381 117
```

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

	lhs	rhs	support	confidence	lift	count
## [1]	{circunscripcion=Uruguay, reccap_1}	=> {demorado}	0.01019350	0.7151515	2.680892	118
## [2]	{circunscripcion=Uruguay, materia=paz, reccap_1}	=> {demorado}	0.01019350	0.7151515	2.680892	118
## [3]	{circunscripcion=Uruguay, reccap_1, justiciables2_3}	=> {demorado}	0.01010712	0.7134146	2.674381	117
## [4]	{circunscripcion=Uruguay, materia=paz, reccap_1, justiciables2_3}	=> {demorado}	0.01010712	0.7134146	2.674381	117
## [5]	{tproc=APREMIO, circunscripcion=Uruguay, reccap_1}	=> {demorado}	0.01002073	0.7581699	2.842155	116
## [6]	{tproc=APREMIO, circunscripcion=Uruguay, materia=paz, reccap_1}	=> {demorado}	0.01002073	0.7581699	2.842155	116
## [7]	{tproc=APREMIO, circunscripcion=Uruguay, reccap_1, justiciables2_3}	=> {demorado}	0.01002073	0.7581699	2.842155	116
## [8]	{tproc=APREMIO, circunscripcion=Uruguay, materia=paz, reccap_1, justiciables2_3}	=> {demorado}	0.01002073	0.7581699	2.842155	116

```
library(arulesViz)
```

```
## Loading required package: grid
```

```
plot(rules, method = "grouped")
```

