

# Análise de Sobrevivência em estudos do Direito

Fernando Corrêa

8 de agosto de 2016

# Introdução

- ▶ O que é análise de sobrevivência?
- ▶ Quando utilizamos análise de sobrevivência?
- ▶ Que tipo de perguntas um modelo de sobrevivência responde?
- ▶ Por que um modelo de sobrevivência é especial?

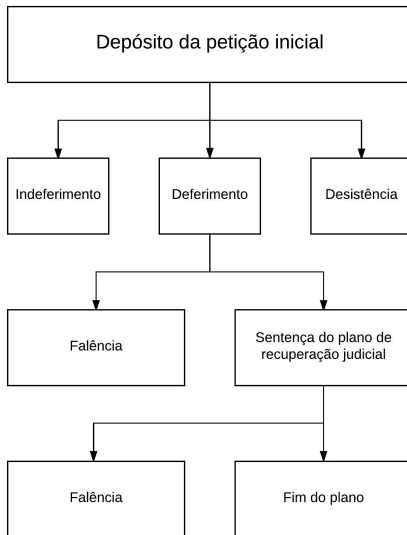
# Tempos no direito

- ▶ No direito, é interessante estudar o tempo entre muitos eventos
- ▶ Sentenças
- ▶ Decisões
- ▶ Distribuição de processos
- ▶ Fim do processo

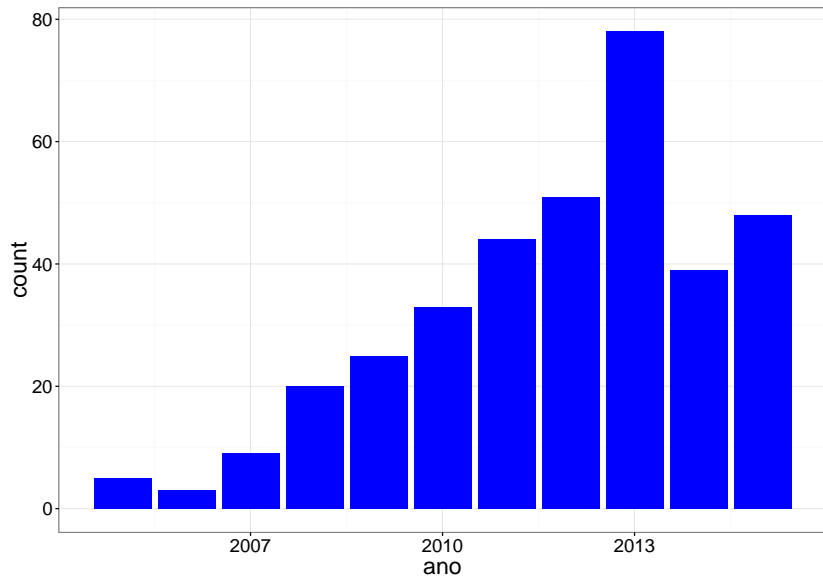
## Exemplo trabalhado durante a apresentação

- ▶ Interesse na duração do processo de recuperação judicial
- ▶ Trabalho a ser apresentado no Encontro de Pesquisa Empírica em Direito (EPED) de 2016
- ▶ Desejamos estudar a duração do processo por fases

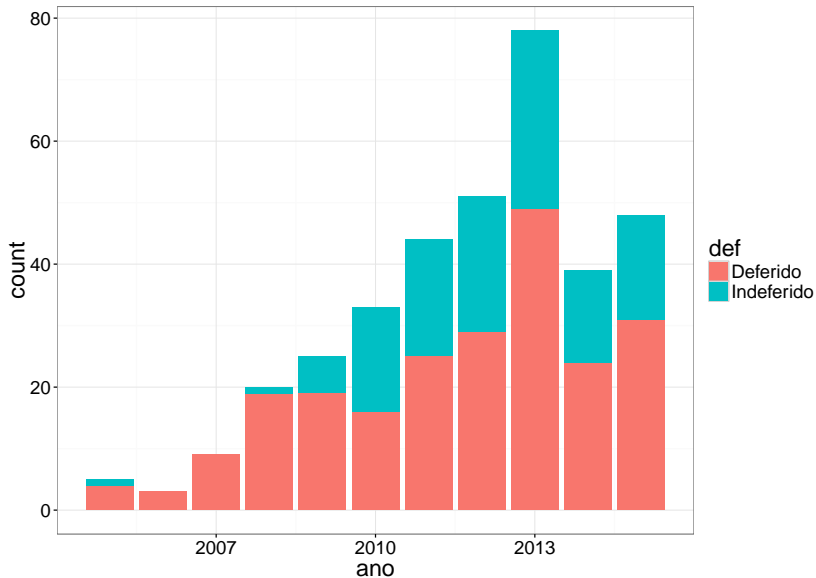
# Exemplo trabalhado durante a apresentação



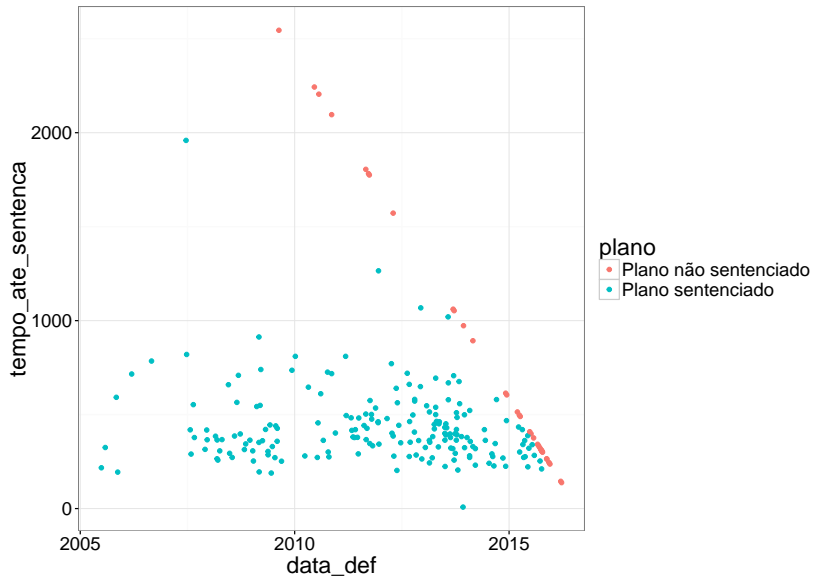
## Principais resultados dessa pesquisa



# Principais resultados dessa pesquisa

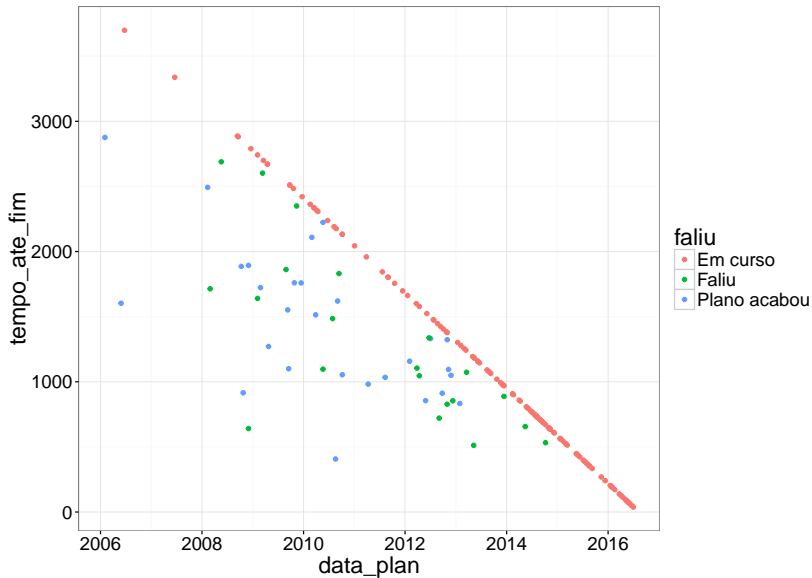


# Principais resultados dessa pesquisa





# Principais resultados dessa pesquisa



## Principais resultados dessa pesquisa

Estados	Proporção
Desistiu	6%
Extinto	2%
Faliu 1	7%
Faliu 2	10%
Indeferido	62%
Plano acabou	13%

- ▶ O tempo mediano até uma decisão de deferimento/indeferimento é de 39 dias
- ▶ O tempo mediano até a sentença do plano é de 1.25 anos
- ▶ Falências ocorreram mais rápido nos períodos mais recentes.
- ▶ Mais de 80% dos processos levam mais de 5 anos entre a sentença do plano e o fim do plano.

# Análise de sobrevivência - rudimentos

- ▶ Estudar duração, em dias, dos períodos entre os eventos
- ▶ Como descrever o comportamento de uma quantidade aleatória, incerta?
- ▶ Como descrever o tempo, incerto, de duração de um período?

# Análise de sobrevivência - rudimentos

- ▶ Função de probabilidades:

$p(x_i)$  = Probabilidade do evento ocorrer no instante  $x_i$

- ▶ Função de distribuição de probabilidades:

$F(x_j)$  = Probabilidade do evento ocorrer até o instante  $x_j$  =

$$\sum_{x_i \leq x_j} p(x_i)$$

# Análise de sobrevivência - rudimentos

- ▶ Descrever tempos
- ▶ Função de sobrevivência

$S(x_j)$  = Probabilidade do evento ocorrer

depois do instante  $x_j = 1 - F(x_j) = \sum_{x_i > x_j} p(x_i)$

# Análise de sobrevivência

- ▶ Supondo que o evento ainda não ocorreu até  $x_0$ , como qualificar a plausibilidade da sua ocorrência no futuro próximo?

# Análise de sobrevivência

- ▶ Supondo que o evento ainda não ocorreu até  $x_0$ , como qualificar a plausibilidade da sua ocorrência no futuro próximo?
- ▶ Probabilidade condicional

$$P(X = x_0 | T \geq x_0) = \frac{p(x_0)}{S(x_0)}$$

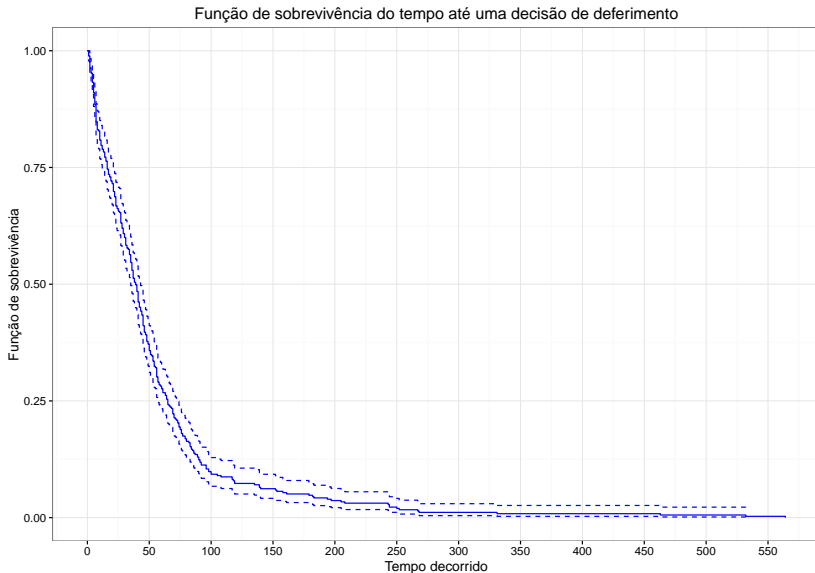
## Análise de sobrevivência - rudimentos

- ▶ Tempo até o deferimento de um processo de recuperação judicial

```
recuperacoes2 %>%  
  survfit(Surv(tempo_ate_def) ~ 1, data = .) %>%  
  GGally::ggsurv(plot.cens = F,  
                 surv.col = c('blue'))+  
  theme_bw()+  
  ylab('Função de sobrevivência')+  
  xlab('Tempo decorrido')+  
  ggtitle('Função de sobrevivência do tempo  
          até uma decisão de deferimento')+  
  scale_x_continuous(breaks = seq(0,600,50))
```



# Análise de sobrevivência - rudimentos

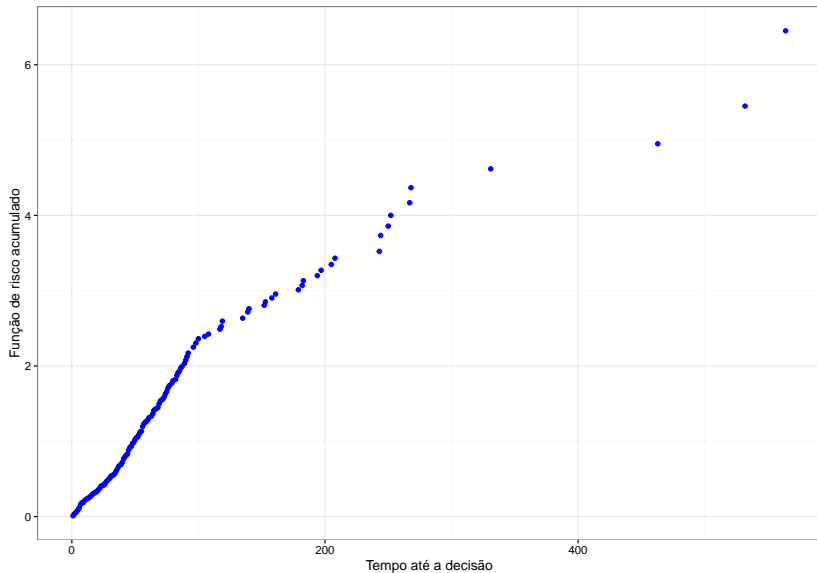


# Análise de sobrevivência - rudimentos

- Função de risco do tempo de deferimento

```
recuperacoes2 %>%  
coxph(Surv(tempo_ate_def)~1,  
      data=.) %>%  
basehaz() %>%  
ggplot(aes(x = time, y = hazard))+  
geom_point(color = 'blue')+  
xlab('Tempo até a decisão')+  
ylab('Função de risco')+  
theme_bw()
```

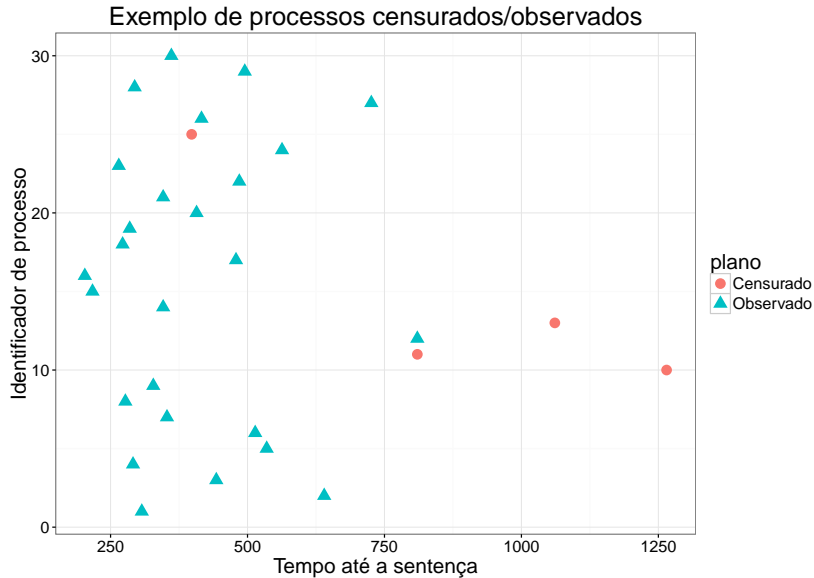
# Análise de sobrevivência - rudimentos



# Análise de sobrevivência - Censura

- ▶ Sempre encontramos informações incompletas
- ▶ Informação parcial é útil

# Análise de sobrevivência - Censura



# Análise de sobrevivência - Censura

- ▶ Censura à direita
- ▶ Tipo mais comum
- ▶ Se ignorada subestima o tempo até o evento

# Análise de sobrevivência - Censura

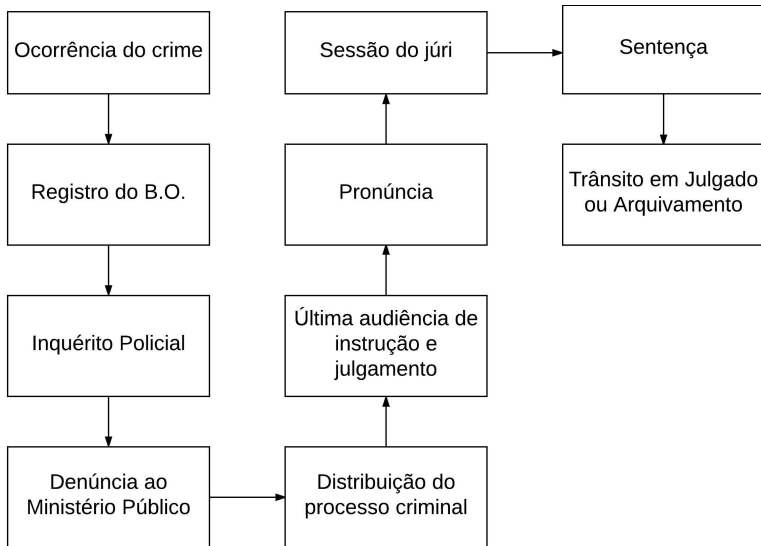
- ▶ Censura à esquerda
- ▶ Tipo mais raro
- ▶ O tempo de acompanhamento é maior do que o tempo entre a ocorrência do evento de interesse e hoje
- ▶ Se ignorada subestima o tempo até o evento

# Análise de sobrevivência - Exemplo

- ▶ Projeto SENASP
- ▶ Interesse em avaliar o tempo de duração das fases do processo de homicídio
- ▶ Nossas observações ora começavam no inquérito, ora na distribuição



# Análise de sobrevivência - Exemplo



## Análise de sobrevivência - Censuras no R

- ▶ Censuras são importantes
- ▶ Como armazenamos informações sobre censuras nas nossas bases de dados?

n_processo	tempo_ate_sentenca	censura_sentenca
0031947-20.2010.8.26.0100	726	1
1037458-06.2015.8.26.0100	276	1
1036831-02.2015.8.26.0100	398	0
0034968-04.2010.8.26.0100	301	1
1069200-83.2014.8.26.0100	277	1
1077328-58.2015.8.26.0100	325	0
1083770-40.2015.8.26.0100	237	0
0111886-83.2009.8.26.0100	421	1
1005882-29.2014.8.26.0100	522	1
1083052-14.2013.8.26.0100	324	1

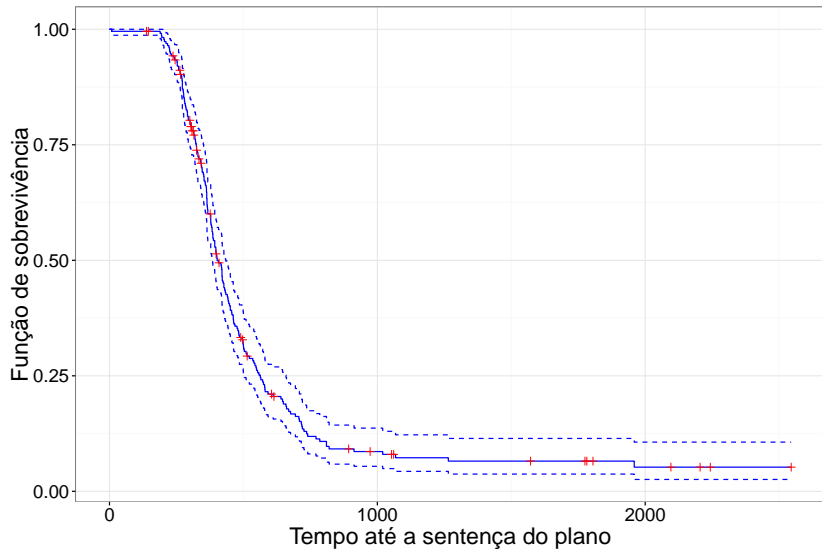
# Análise de sobrevivência - Censuras no R

- ▶ Censuras são importantes
- ▶ Como estimamos risco e sobrevivência sabendo que existem censuras?

```
fit_sentenca <- recuperacoes2 %>%  
  filter(def == 1) %>%  
  survfit(Surv(tempo_ate_sentenca, censura_sentenca) ~ 1,  
          data = .)
```

```
ggsurv(fit_sentenca, surv.col = 'blue')+  
  theme_bw()+  
  theme(text = element_text(size = 20))+  
  xlab('Tempo até a sentença do plano')+  
  ylab('Função de sobrevivência')
```

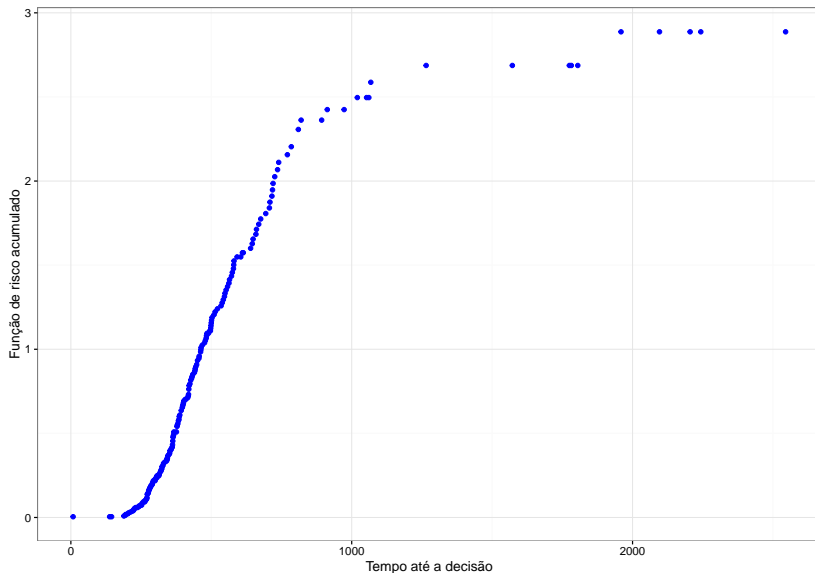
## Análise de sobrevivência - Censuras no R



## Análise de sobrevivência - Censuras no R

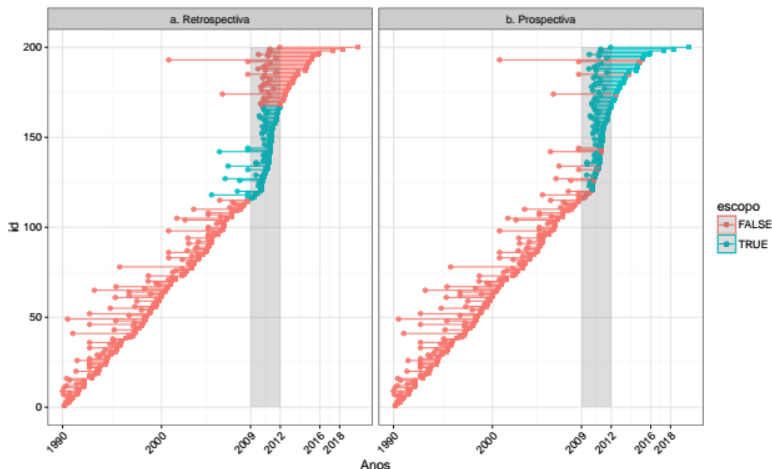
```
recuperacoes2 %>%  
filter(def == 1) %>%  
coxph(Surv(tempo_ate_sentenca, censura_sentenca)~1,data=.)  
basehaz() %>%  
ggplot(aes(x = time, y = hazard))+  
geom_point()+  
xlab('Tempo até a decisão')+  
ylab('Função de risco')+  
theme_bw()
```

# Análise de sobrevivência - Censuras no R



# Análise de sobrevivência - aspectos computacionais e metodológicos

- ▶ Como conduzir um estudo de análise de sobrevivência?
- ▶ Prospectivo x retrospectivo



# Análise de sobrevivência - aspectos computacionais e metodológicos

- ▶ Como conduzir um estudo prospectivo?
- ▶ De que forma virão as informações?
- ▶ Devemos transformá-las para qual formato?

n_processo	primeira_mov	data_def	data_plan	data_fal
1	2009-04-16	2009-06-05	2010-08-24	NA
2	2015-03-19	2015-03-30	NA	NA
3	2010-12-21	NA	NA	NA
4	2014-07-21	2014-09-05	2015-08-17	NA
5	2007-05-02	2007-07-27	2008-09-18	NA



# Análise de sobrevivência - aspectos computacionais e metodológicos

- ▶ É necessário partir de uma lista de processos

```
d_rjs %>%  
  with(nuprocesso) %>%  
  cpo_pg(path = 'data-raw/cpopg',  
         tj = 'TJSP')  
  
list.files('data-raw/cpopg', full.names = T) %>%  
  grep(pattern = '.html', fixed = T, value = T) %>%  
  parse_cpopg()  
  
htmls_processos %>%  
  grep(pattern = '.rds', fixed = T, value = T) %>%  
  plyr::ldply(function(x){readRDS(x)}) ->  
  dados_rjs
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

- ▶ O produto final desta fase é uma lista de movimentações, uma para cada processo

data_mov	titulo
2014-02-20	Arquivado Definitivamente no Arquivo Geral
2013-08-09	Trânsito em Julgado às partes - com Baixa
2012-12-17	Certidão de Publicação Expedida
2012-12-14	Remetido ao DJE
2012-12-06	Despacho

# Análise de sobrevivência - aspectos computacionais e metodológicos

```
encontra_deferimento <- function(vtxt){  
  k = 1  
  deferimento <- 'ensejar |  
    entendo presentes os pressupostos legais|  
    está em ordem a documentação|  
    hei por bem determinar| defiro|  
    concedida'  
  
  pei <- 'deferimento da recuperação judicial|  
    desta recuperação judicial|  
    petição inicial|  
    processamento|  
    pedido de recuperação judicial|  
    recuperação judicial'  
  
  indeferimento <- ' indefiro| indeferida'
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
defs <- character(length(vtxt))
for(texto in vtxt){

  sentencas <- unlist(strsplit(texto, '.', fixed = T))
  tem_def <- grepl(pei,sentencas, ignore.case = T)*
    grepl(deferimento,sentencas, ignore.case = T)
  tem_indef <- -grepl(pei,sentencas, ignore.case = T)*
    grepl(indeferimento,sentencas, ignore.case = T)

  defs[k] <- ifelse(min(tem_indef)< 0,min(tem_indef),max(
names(defs)[k] <- sentencas[which(tem_def == 1)[1]]
  k <- k + 1
}
return(defs)
}
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
encontra_plano <- function(vtxt){  
  k = 1  
  deferimento <- 'plano de recuperação aprovado|  
HOMOLOGO O PLANO para conceder a RECUPERAÇÃO JUDICIAL|  
aprovação da alteração do plano de recuperação judicial|  
homologo o plano de recuperação judicial|  
homologo a recuperação judicial|  
concedo a recuperação judicial|  
aprovo o plano de recuperação judicial'  
  deferimento <- gsub('\n ', '', deferimento)  
  pei <- deferimento  
  negacao <- 'não'  
  defs <- character(length(vtxt))
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
for(texto in vtxt){  
  
  sentencas <- unlist(strsplit(texto, '.', fixed = T))  
  tem_def <- grepl(pei,sentencas, ignore.case = T)*grepl(  
  
  defs[k] <- max(tem_def)  
  names(defs)[k] <- sentencas[which(tem_def == 1)[1]]  
  k <- k + 1  
}  
return(defs)  
}
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
encontra_falencia <- function(vtxt){  
  k = 1  
  deferimento <- ' foi decretada|  
  é caso de|  
  convolo a recuperação|  
  decreto |  
  houve a convolação|convolo '  
  pei <- ' decretada a falência|  
  convolação em falência|  
  falência|  
  convolação da recuperação judicial em falência'
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
encontra_desistencia <- function(vtxt){  
  
  deferimento <- 'homologo|concedo '  
  pei <- ' desistência| pedido de desistência'  
  
  encontra_extint <- function(vtxt){  
  
    deferimento <- 'julgo |decreto '  
    pei <- ' extinto| extincao| extinta a ação'
```



## Análise de sobrevivência - aspectos computacionais e metodológicos

```
encontra_fim <- function(vtxt){  
  deferimento <- 'determino|  
  decreto|  
  DECLARO ENCERRADA a recuperação judicial |  
  DECLARO que o plano de recuperação judicial foi cumprido  
  pei <- 'encerramento da recuperação judicial|  
  DECLARO ENCERRADA a recuperação judicial |  
  DECLARO que o plano de recuperação judicial foi cumprido
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
defs <- encontra_deferimento(movs_recuperacoes$mov)
planos <- encontra_plano(movs_recuperacoes$mov)
desist <- encontra_desistencia(movs_recuperacoes$mov)
extint <- encontra_extincao(movs_recuperacoes$mov)
fim <- encontra_fim(movs_recuperacoes$mov)
fals <- encontra_falencia(movs_recuperacoes$mov)
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
computa_estado_final <- function(  
  primeira_mov, data_def,  
  data_desist, data_extint,  
  data_plan, data_fal,  
  data_fim, def){  
  
  tempos <- c(data_def - primeira_mov,  
              data_desist - primeira_mov,  
              data_extint - primeira_mov,  
              data_plan - data_def,  
              data_fal - data_def,  
              data_fal - data_plan,  
              data_fim - data_plan)
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
classes <- c('Deferido',  
            'Desistiu',  
            'Extinto',  
            'Sentença do plano',  
            'Faliu 1',  
            'Faliu 2',  
            'Plano acabou')  
  
n1 = ifelse(  
  is.infinite(min(tempos[1:3])),  
  NA,  
  which.min(tempos[1:3]))  
  
final = ifelse(def == -1 & !is.na(def),  
              'Indeferido',  
              classes[n1][1])
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
if(final == 'Deferido'){  
  
  n2 = 3+ifelse(  
    is.infinite(min(tempos[4:5])),  
    NA,  
    which.min(tempos[4:5]))  
  
  final = ifelse(  
    is.na(n2),  
    'Deferido',  
    classes[n2][1])  
  
}
```

## Análise de sobrevivência - aspectos computacionais e metodológicos

```
if(final == 'Sentença do plano') {  
  
  n3 = 5+ifelse(  
    is.infinite(min(tempos[6:7])),  
    NA,  
    which.min(tempos[6:7]))  
  
  final = ifelse(  
    is.na(n3),  
    'Sentença do plano',  
    classes[n3][1])  
  
}  
  
return(final)
```

# Análise de sobrevivência - conclusões

- ▶ Muita mineração de texto é necessária
- ▶ Cuidados metodológicos são muito importantes
- ▶ O primeiro passo na análise de sobrevivência é a modelagem das fases processusi