

Jurimetrics

José de Jesus Filho Julio Trecenti

Brussels, July 06

What is jurimetrics?

- ▶ Statistics in the law
- ▶ Jurimetrics Brazilian Association



Jurimetrics: spheres

- ▶ One litigation
 - ▶ Compensating for the loss of a chance
- ▶ Many litigations
 - ▶ Predictive models
 - ▶ Decision making
 - ▶ Topic models
- ▶ Public policy
 - ▶ Public policy impact
- ▶ Court administration
 - ▶ litigation complexity
 - ▶ court productivity

Orgs

- ▶ github.com/abjur: ABJ's org
- ▶ github.com/courtsbr: github organisation
- ▶ github.com/decrypтр: breaking CAPTCHAs using R

Packages

- ▶ `tjsp` and `esaj` packages: web scraping.
- ▶ `jurisMiner`: analyze judicial text data.
- ▶ ...

Download: usage

```
library(tjsgp)
s <- cjsg_session()
parms <- cjsg_parms(s, livre = 'erro medico')
cjsg(s, parms, path = 'data-raw', max_pag = 10)
```

Parsing: usage

```
files <- dir('data-raw', full.names = TRUE)
d_cjsg <- parse_cjsg(files)
```

```
tibble::glimpse(d_cjsg, width = 50)
## Observations: 200
## Variables: 8
## $ n_processo      <chr> "1013781-16.2015.8.26...
## $ comarca         <chr> "Santos", "São Paulo"...
## $ data_julgamento <chr> "15/06/2017", "13/06/...
## $ data_registro   <chr> "15/06/2017", "15/06/...
## $ ementa          <chr> "EMBARGOS DE DECLARAÇ...
## $ orgao_julgador   <chr> "5ª Câmara de Direito...
## $ relatora        <chr> "Fábio Podestá", "Val...
## $ classe_assunto   <chr> "Embargos de Declaraç..."
```

Download and parse full decision

```
code <- d_cjsg$cd_acordao[1]  
tjsp::download_pdf(code, 'data-raw')
```


Clean text data

- ▶ regex... regex...
- ▶ future: thesaurus

Topic models

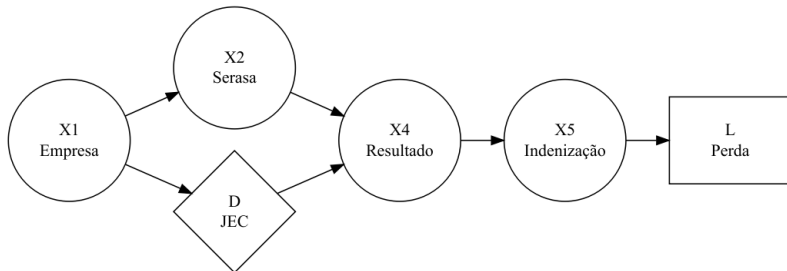
- ▶ unsupervised learning of litigation topics

Predictive models

- ▶ Using bag of words + LASSO, RF, NN etc.

Influence diagrams: decision

- bayesian network with decision node.



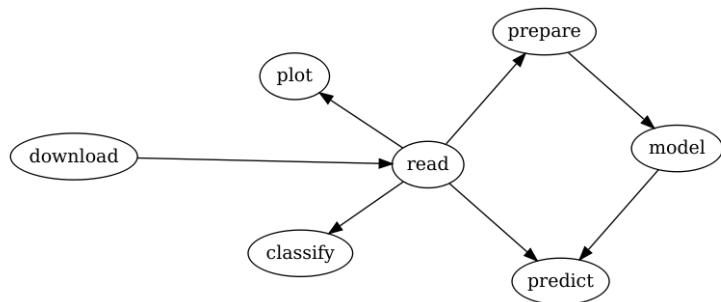
Breaking captchas

```
library(decryptr)  
read_captcha('data-raw/captcha.png') %>% plot()
```



```
captchaSajAudio::decifrar('data-raw/captcha.mp3')  
## [1] "emcdz"
```

Breaking CAPTCHAs - API



Future work: tech

- ▶ Better scrapers
- ▶ Public API

Future work: research

- ▶ Most relevant arguments

Links

- ▶ Github:

- ▶ ABJ: <https://github.com/abjur>
- ▶ decryptr: <https://github.com/decryptr>
- ▶ courtsbr: <https://github.com/courtsbr>
- ▶ Jose: <https://github.com/jjesusfilho>
- ▶ Julio: <https://github.com/jtrecenti>

- ▶ Sites (in portuguese):

- ▶ ABJ Blog: <http://abj.org.br>
- ▶ Brazil Blog Aggregator: <http://brbloggers.com.br>
- ▶ Our R blog: <http://curso-r.com>

- ▶ Personal:

- ▶ <http://jjesusfilho.info>
- ▶ <http://jtrecenti.com>