Programas

```
load("data/programas.RData")

library(tidytext)
```

```
## Warning: package 'tidytext' was built under R version 3.6.1
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(stringr)
library(scales)
library(readxl)
library(ggplot2)

stop_words_es <-read.csv("data/stop-words-spanish.csv", stringsAsFactors = FALSE, fileEncoding = "utf-8
sentiments <- read.csv("data/sentimientos.csv", stringsAsFactors = FALSE, fileEncoding = "utf-8")

tidy_prog = function(txt){
text_df <- tibble(line = 1:length(txt), text = txt)
text_df=text_df%>%
  unnest_tokens(word, text) %>%
  mutate(word=gsub("[[:punct:]]", "", word)) %>%
  mutate(word=gsub("[[:digit:]]","", word)) %>%
  anti_join(stop_words_es) %>%
  filter(!word %in% c("aa","","as"))
}

for (i in partidos){
  assign(paste0("tidy_",i), tidy_prog(get(i)))
}
```

```
## Joining, by = "word"
```

```
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
## Joining, by = "word"
```

```
prog = function(prog, name){
  aux = data.frame("text"=prog, "partido"=name)
}
```

```
for (i in partidos){
  assign(paste0("prog_",i), prog(get(i),i))
}
programas <- bind_rows(prog_ca, prog_fa,prog_pc,prog_pg, prog_pi, prog_pn, prog_up)
```

```
## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector
```

```
tidy_books<-  bind_rows(mutate(tidy_ca, partido = "ca"),
          mutate(tidy_pc, partido = "pc"),
          mutate(tidy_pg, partido = "pg"),
          mutate(tidy_pi, partido = "pi"),
          mutate(tidy_pn, partido = "pn"),
```

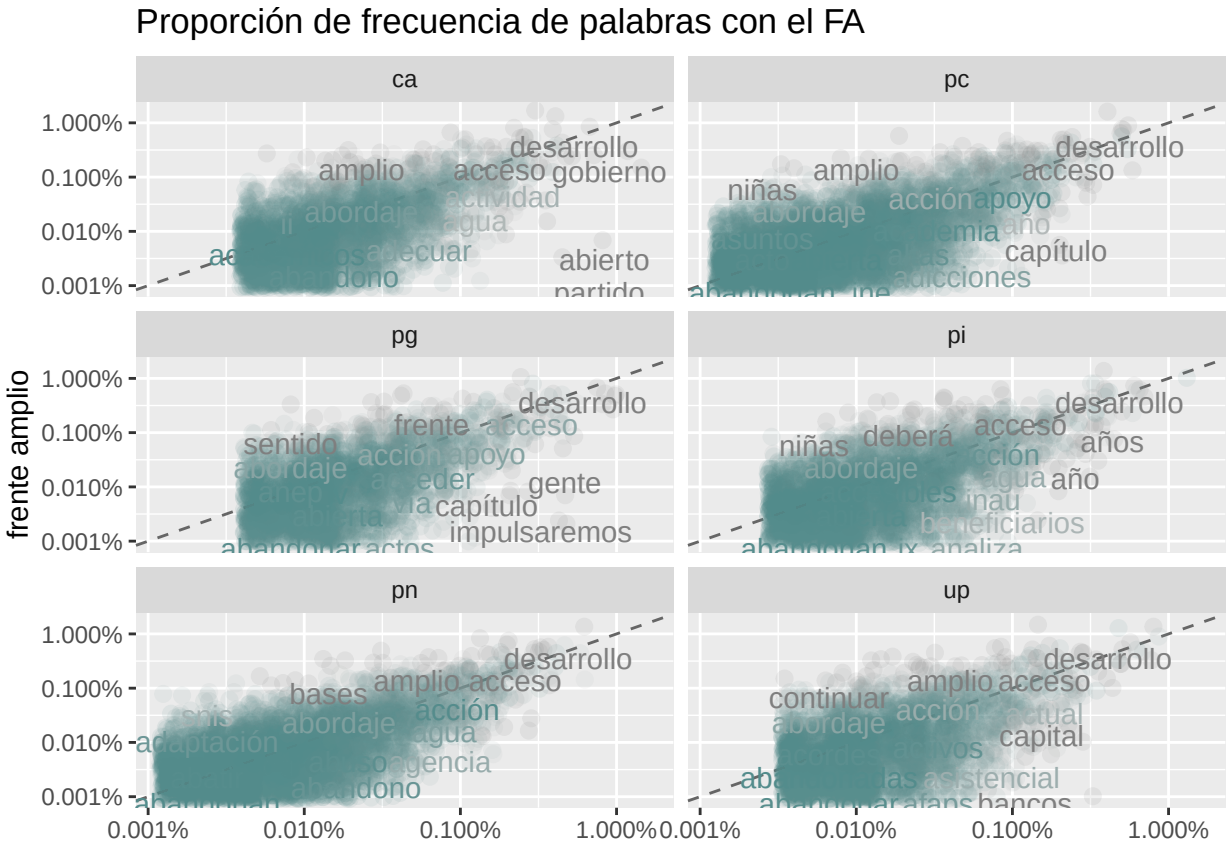
```
          mutate(tidy_up, partido = "up"),
          mutate(tidy_fa, partido = "fa"))

frequency <- bind_rows(mutate(tidy_ca, partido = "ca"),
                       mutate(tidy_pc, partido = "pc"),
                       mutate(tidy_pg, partido = "pg"),
                       mutate(tidy_pi, partido = "pi"),
                       mutate(tidy_pn, partido = "pn"),
                       mutate(tidy_up, partido = "up"),
                       mutate(tidy_fa, partido = "fa")) %>%
#  mutate(word = str_extract(word, "[a-z']+")) %>%
  count(partido, word) %>%
  group_by(partido) %>%
  mutate(proportion = n / sum(n)) %>%
  select(-n) %>%
  spread(partido, proportion) %>%
#  melt(id.vars="word", variable.name="partido",value.name = "frec" )
  gather(partido, proportion, c("ca","pc","pg","pi","pn","up"))
```

#comparacion contra el FA de los partidos de proporcion de palabras

```
ggplot(frequency, aes(x = proportion, y = `fa`, color = abs(`fa` - proportion))) +
  geom_abline(color = "gray40", lty = 2) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.3, height = 0.3) +
  geom_text(aes(label = word), check_overlap = TRUE, vjust = 1.5) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  scale_color_gradient(limits = c(0, 0.001), low = "darkslategray4", high = "gray75") +
  facet_wrap(~partido, ncol = 2) +
  theme(legend.position="none") +
  labs(title="Proporción de frecuencia de palabras con el FA", y = "frente amplio", x = NULL)
```

```
## Warning: Removed 93244 rows containing missing values (geom_point).
```

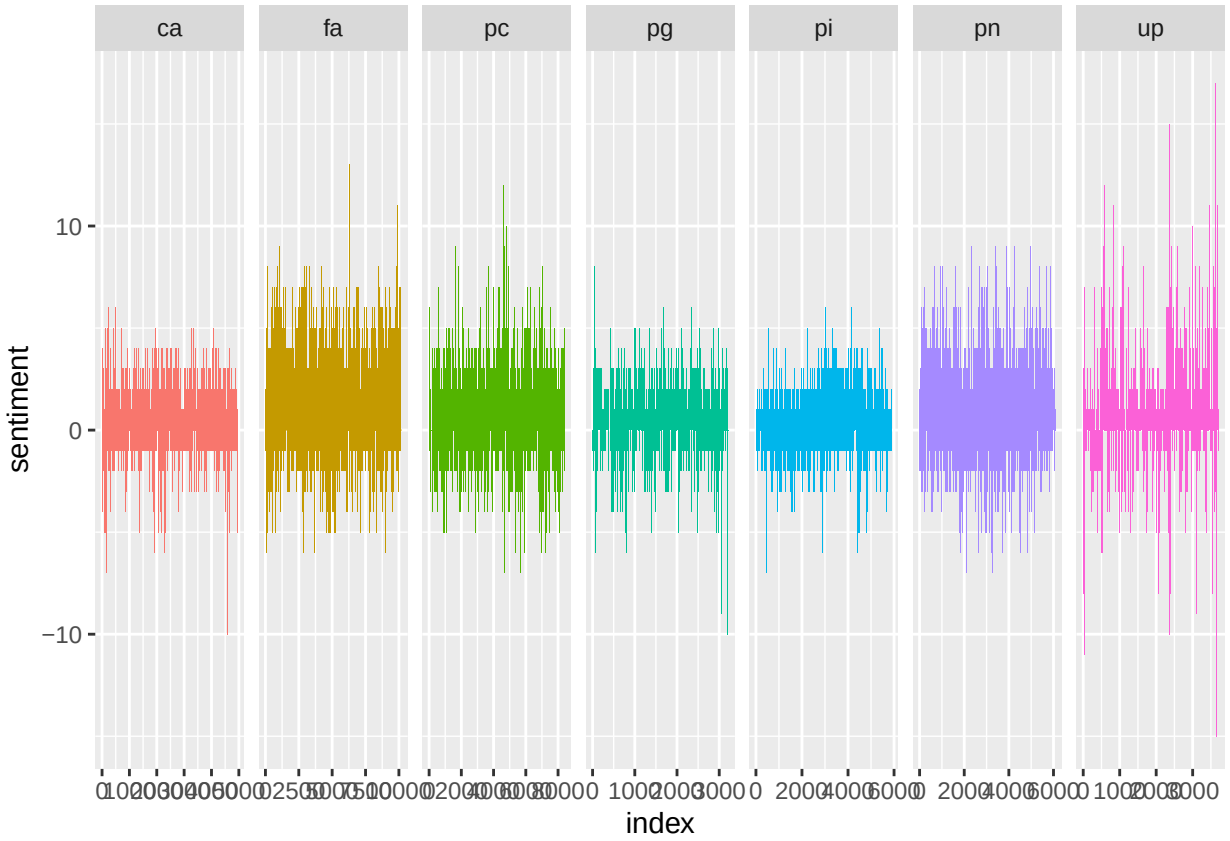
```
## Warning: Removed 93244 rows containing missing values (geom_text).
```




```
sent=sentiments %>%
#  filter(Joy==1) %>%
  filter(Spanish..es.!="NO TRANSLATION") %>%
  rename(word=Spanish..es., positive=Positive, negative=Negative) %>%
  select(word, positive,negative) %>%
  gather(sentiment,value,positive:negative) %>%
  filter(value==1)

sent_fa=tidy_books %>%
  inner_join(sent) %>%
  count(partido, index = line %/% 1, sentiment) %>%
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative)
```

```
## Joining, by = "word"
```

```
ggplot(sent_fa, aes(index, sentiment, fill = partido)) +
  geom_col(show.legend = FALSE)+
  facet_grid(~partido, scales="free")
```



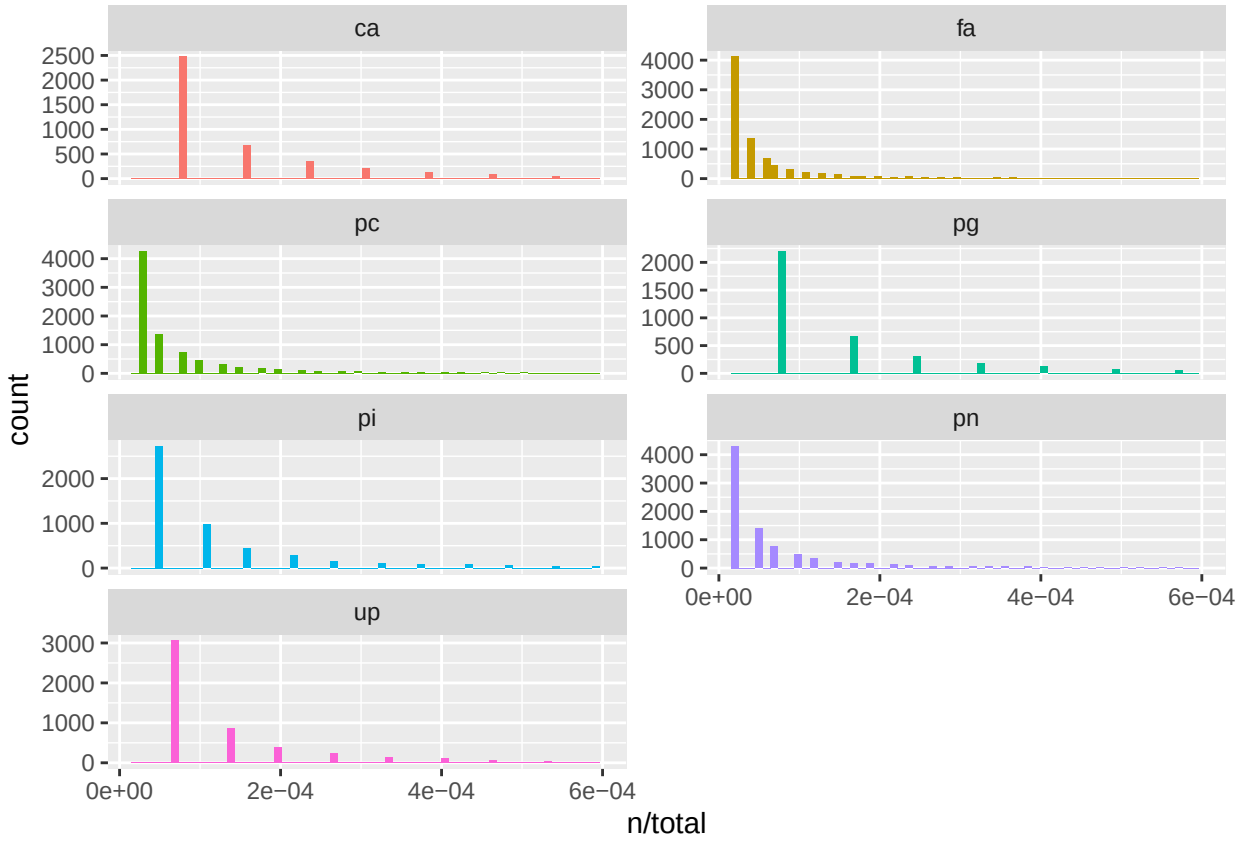

```
book_words <- tidy_books %>%
  select(partido,word) %>%
  count(partido, word, sort = TRUE)

total_words <- book_words %>%
  group_by(partido) %>%
  summarize(total = sum(n))

book_words <- left_join(book_words, total_words)
```

```
## Joining, by = "partido"
```

```
freq_by_rank <- book_words %>%
  group_by(partido) %>%
  mutate(rank = row_number(),
         `term frequency` = n/total)

ggplot(book_words, aes(n/total, fill = partido)) +
  geom_histogram(show.legend = FALSE, bins=60) +
  xlim(NA, 0.0006) +
  facet_wrap(~partido, ncol = 2, scales = "free_y")
```

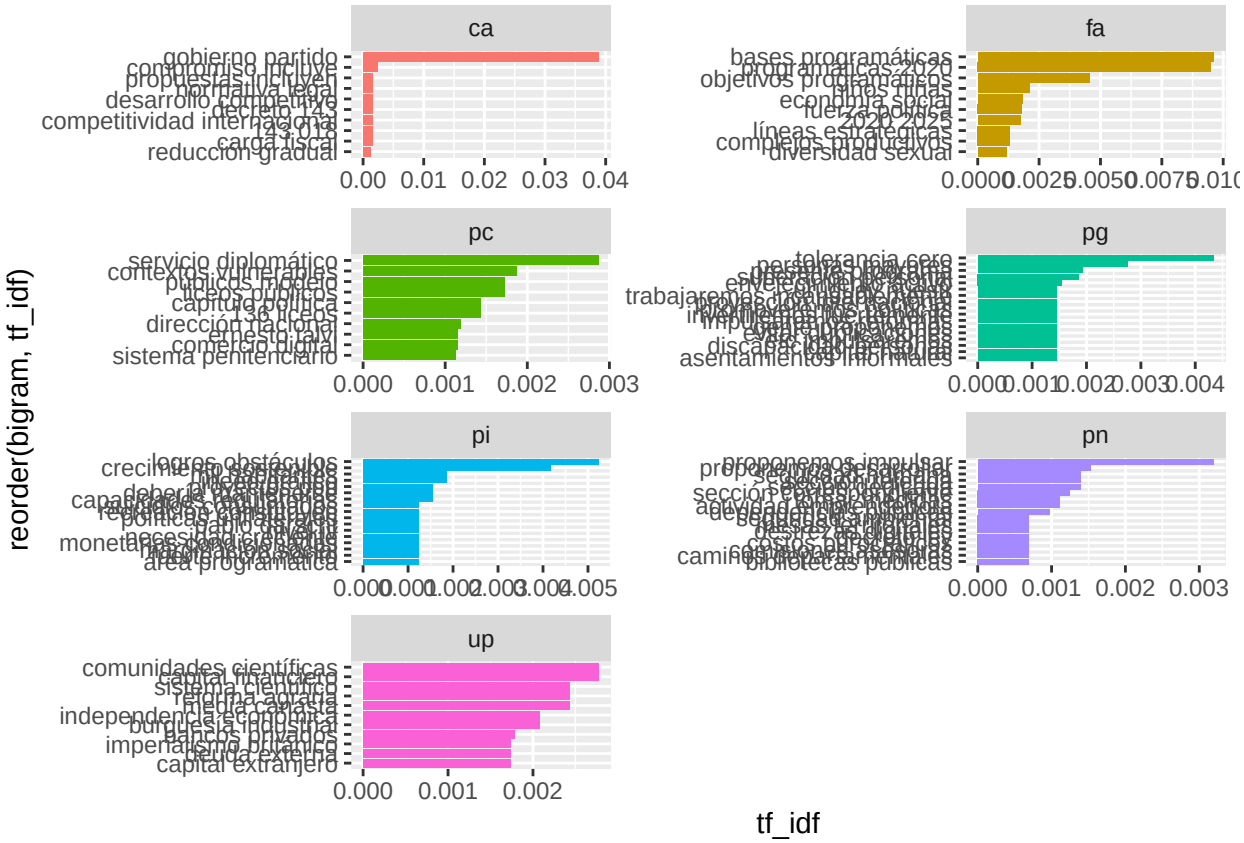
```
## Warning: Removed 2153 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 7 rows containing missing values (geom_bar).
```




```
programas_bigrams <- programas %>%
  unnest_tokens(bigram, text, token = "ngrams", n = 2, to_lower = TRUE)


bigrams_separated <- programas_bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ")

bigrams_filtered <- bigrams_separated %>%
  filter(!word1 %in% stop_words_es$word) %>%
  filter(!word2 %in% stop_words_es$word) %>%
  filter(!word1 %in% c("aa","","as","pág")) %>%
  filter(!word2 %in% c("aa","","as","pág")) %>%
  filter(!(partido=="fa" & word1=="frente")) %>%
  filter(!(partido=="fa" & word1=="amplio")) %>%
  filter(!(partido=="fa" & word2=="frente")) %>%
  filter(!(partido=="fa" & word2=="amplio")) %>%
  filter(!(partido=="ca" & word1=="cabildo")) %>%
  filter(!(partido=="ca" & word1=="abierto")) %>%
  filter(!(partido=="ca" & word2=="cabildo")) %>%
  filter(!(partido=="ca" & word2=="abierto")) %>%
  filter(!(partido=="pc" & word1=="partido")) %>%
  filter(!(partido=="pc" & word1=="colorado")) %>%
  filter(!(partido=="pc" & word2=="partido")) %>%
  filter(!(partido=="pc" & word2=="colorado")) %>%
  filter(!(partido=="pi" & word1=="partido")) %>%
  filter(!(partido=="pi" & word1=="independiente")) %>%
```

```
  filter(!(partido=="pi" & word2=="partido")) %>%
  filter(!(partido=="pi" & word2=="independiente")) %>%
  filter(!(partido=="pn" & word1=="partido")) %>%
  filter(!(partido=="pn" & word1=="nacional")) %>%
  filter(!(partido=="pn" & word2=="partido")) %>%
  filter(!(partido=="pn" & word2=="nacional")) %>%
  filter(!(partido=="up" & word1=="unidad")) %>%
  filter(!(partido=="up" & word1=="popular")) %>%
  filter(!(partido=="up" & word2=="unidad")) %>%
  filter(!(partido=="up" & word2=="popular"))


# new bigram counts:
bigram_counts <- bigrams_filtered %>%
#  group_by(partido) %>%
  count(word1, word2, sort = TRUE)

nombres_partidos = c("cabildo abierto","frente amplio","partido colorado", "partido independiente",
                     "partido nacional","unidad popular")

bigram_tf_idf <- bigrams_filtered %>%
  unite(bigram, word1, word2, sep = " ") %>%
  filter(!bigram %in% nombres_partidos) %>%
  count(partido, bigram) %>%
  bind_tf_idf(bigram, partido, n) %>%
  arrange(desc(tf_idf))


bigram_tf_idf %>%
  group_by(partido) %>%
  top_n(10, tf_idf) %>%
  ungroup() %>%
  ggplot(aes(reorder(bigram,tf_idf),tf_idf, fill = partido)) +
  geom_bar(stat="identity") +theme(legend.position = "none")+
  facet_wrap(~partido, ncol = 2, scales = "free")+
  coord_flip()
```




```
library(igraph)
```

```
## Warning: package 'igraph' was built under R version 3.6.1
```

```
##
## Attaching package: 'igraph'
```

```
## The following object is masked from 'package:tidyr':
##
##     crossing
```

```
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

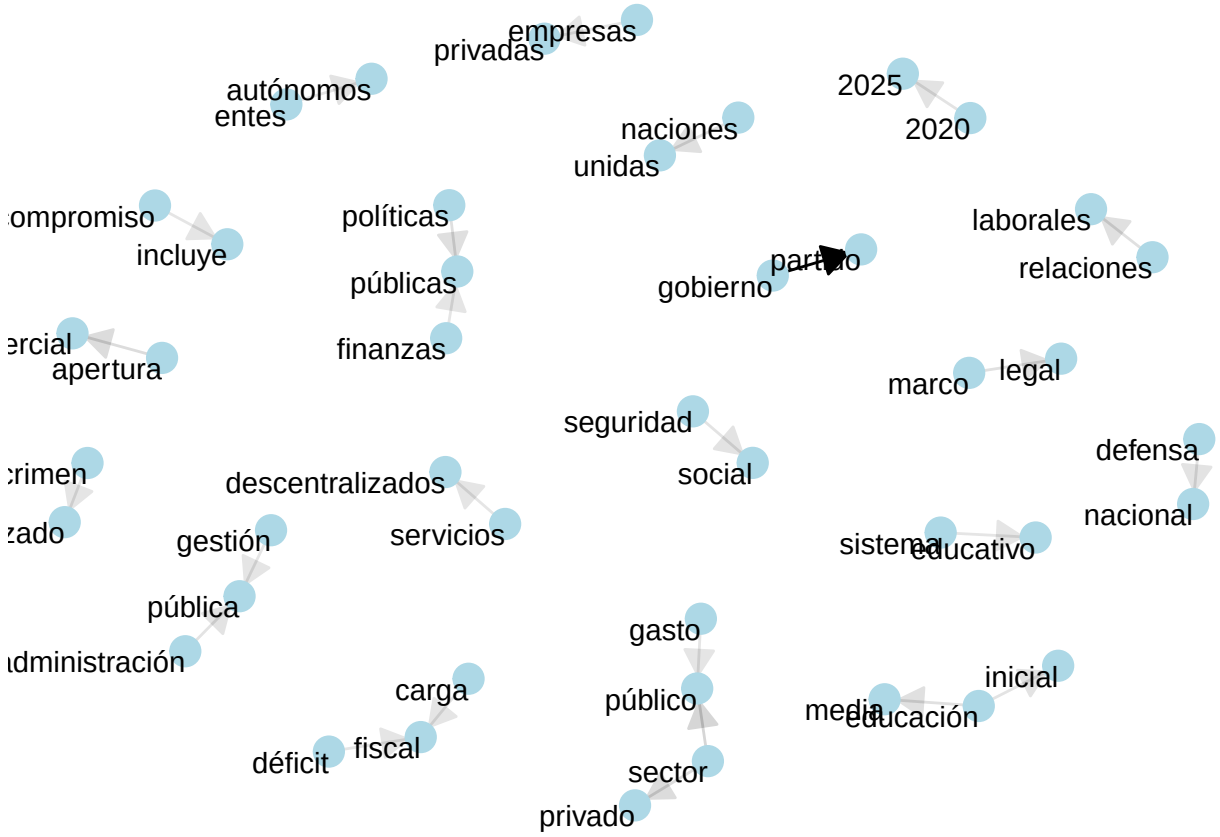
```
## The following object is masked from 'package:base':
##
##     union
```

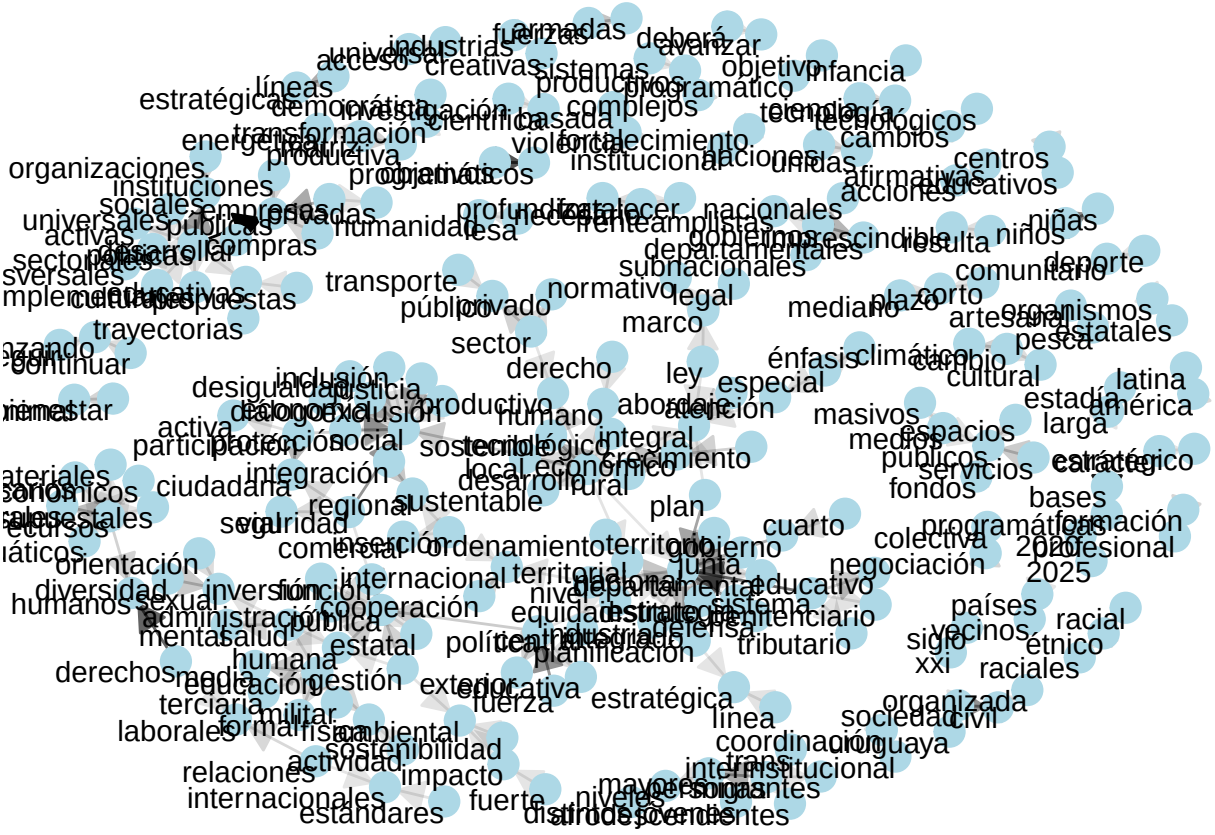
```
library(ggraph)
```

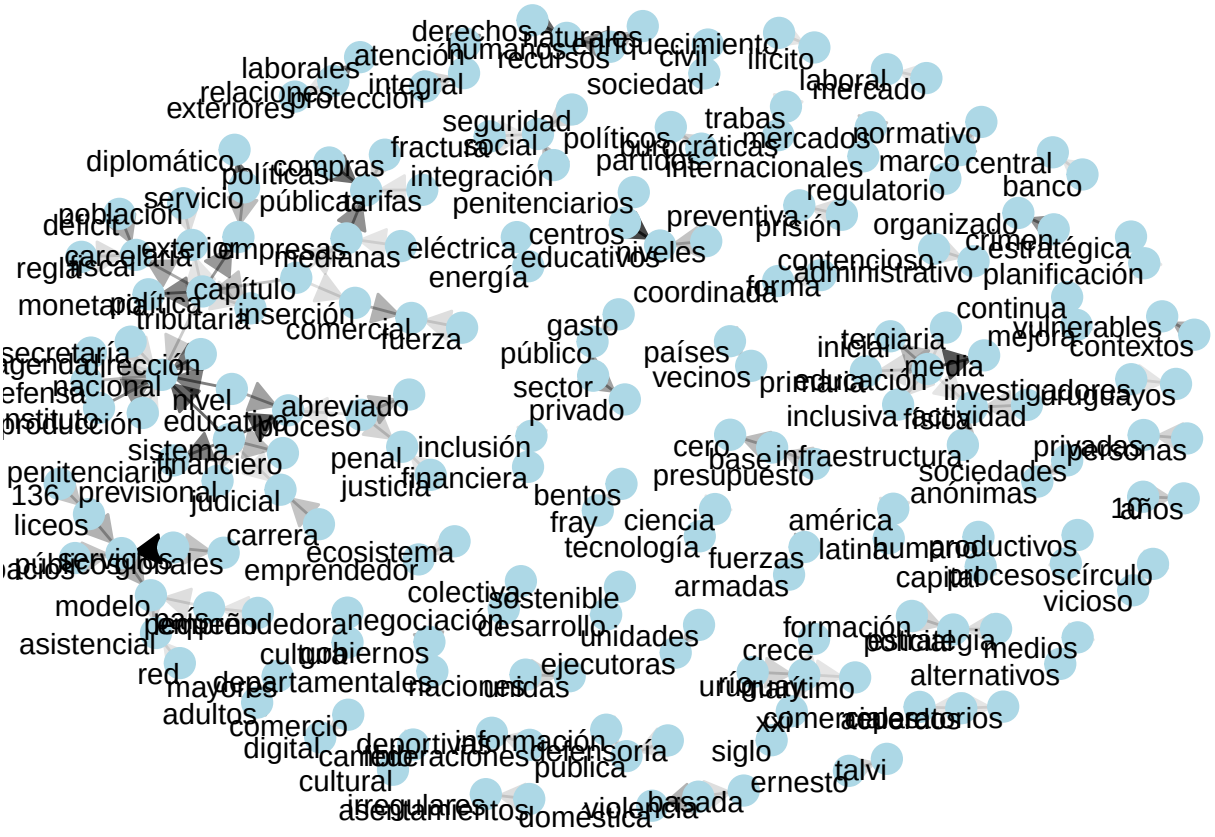
```
## Warning: package 'ggraph' was built under R version 3.6.1
```

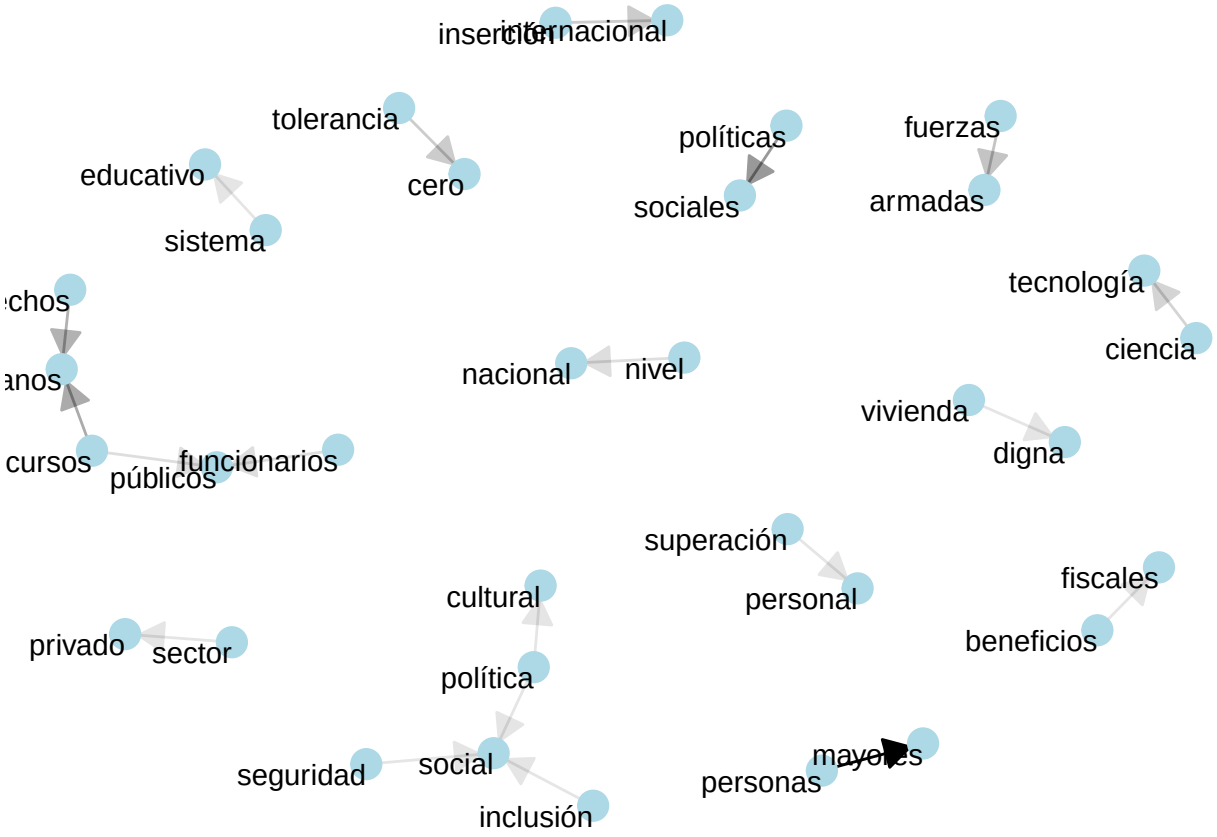
```
set.seed(2017)

for (i in partidos){
```

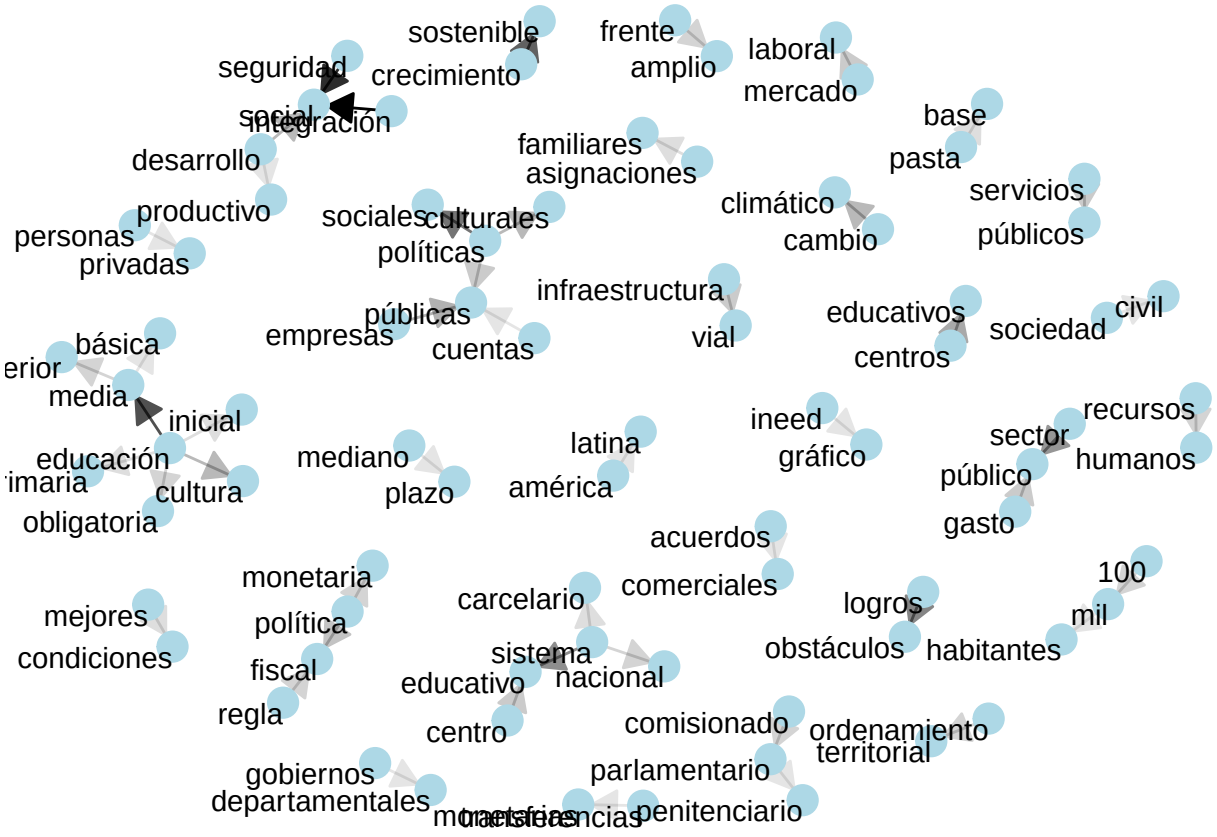
```
bigram_graph <- bigrams_filtered %>%
  filter(partido==i) %>%
  select(-partido) %>%
  count(word1, word2, sort = TRUE) %>%
  filter(n>5) %>%
  graph_from_data_frame()

a <- grid::arrow(type = "closed", length = unit(.15, "inches"))

fig=ggraph(bigram_graph, layout = "fr") +
  geom_edge_link(aes(edge_alpha = n), show.legend = FALSE,
                 arrow = a, end_cap = circle(.07, 'inches')) +
  geom_node_point(color = "lightblue", size = 5) +
  geom_node_text(aes(label = name), vjust = 1, hjust = 1) +
  theme_void()
print(fig)
}
```



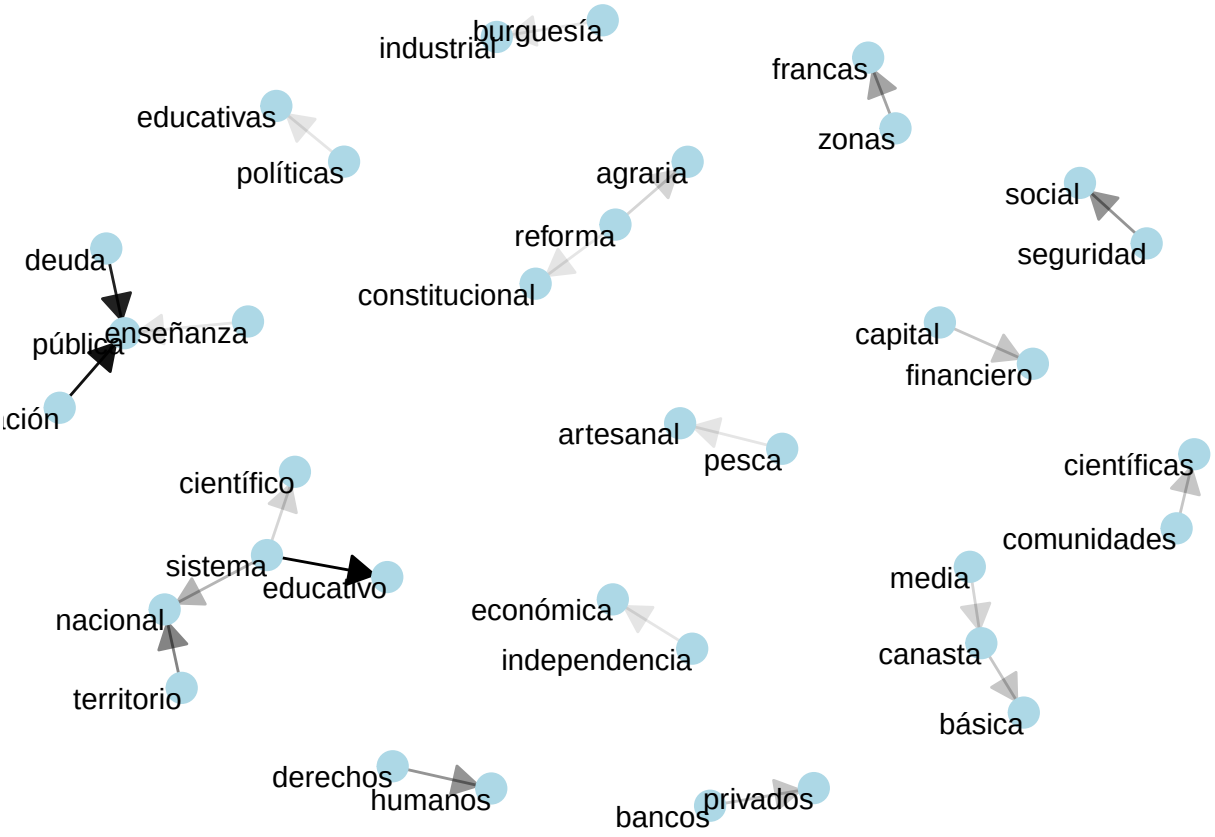

fuerzas
armadas
deberá
avanzar
acceso
universal
industrias
creativas
sistemas
objetivo
infancia
líneas
estratégicas
democrática
producción
programático
complejos
tecnología
investigación
científica
basada
tecnológicos
transformación
energética
violencia
fortalecimiento
cambios
productiva
organizaciones.
objetivos
institucional
acciones
unidas
centros
afirmativas
instituciones
acciones
educativos
sociales
empresas
privadas
profundidad
fortalecer
nacionales
universales
públicas
humanidad
mesa
frente
ampliasta
niñas
activas
compras
departamentales
resulta
niños
sectoriales
subnacionales
deporte
transporte
normativo
legal
comunitario
públicos
privado
mediano
plazo
corto
marco
artesanal
organismos
trayectorias
sector
pesca
estatales
continuar
derecho
énfasis
climático
cambio
ley
especial
cultural
latina
desigualdad
inclusión
justicia
abordaje
atención
estadía
américa
productivo
humano
masivos
larga
participación
protección
social
integral
medios
espacios
activa
sostenible
tecnológico
crecimiento
públicos
servicios
integración
económico
desarrollo
rural
fondos
ciudadanía
regional
sustentable
plan
bases
cuarto
colectiva
programáticas
formación
comercial
ordenamiento
territorial
gobierno
junta
profesional
orientación
negociación
2025
diversidad
inversión
internacional
territorial
departamental
educativo
humanos
sexual
cooperación
equidad
sistema
países
racial
administración
pública
penitenciario
vecinos
mental
salud
estatal
política
defensa
tributario
siglo
étnico
humana
planificación
xxi
raciales
derechos
media
gestión
exterior
educación
estratégica
organizada
terciaria
militar
fuerza
sociedad
civil
laborales
formalización
ambiental
línea
coordinación
uruguaya
sostenibilidad
actividad
impacto
trans
institucional
relaciones
internacionales
fuerte
estándares


derechos
naturales
humanos
atención
recursos
civil
ilícito
laborales
integral
sociedad
laboral
relaciones
protección
exteriores
mercado
seguridad
trabas
fractura
social
políticos
burocráticas
mercados
normativo
diplomático
compras
partidos
internacionales
marco
central
políticas
integración
regulatorio
banco
servicio
públicas
tarifas
penitenciarios
población
preventiva
prisión
organizado
déficit
centros
crimen
exterior
empresas
eléctrica
educativos
contencioso
estratégica
carcelaria
medianas
fiscal
energía
administrativo
planificación
regla
capítulo
coordinada
forma
monetaria
política
continua
tributaria
inserción
comercial
fuerza
gasto
terciaria
mejora
vulnerables
público
países
inicial
contextos
secretaría
vecinos
media
dirección
primaria
educación
sector
investigadores
nacional
privado
inclusiva
física
actividad
producción
abreviado
proceso
sistema
inclusión
cero
privadas
penal
base
infraestructura
personas
financiero
presupuesto
sociedades
penitenciario
financiera
anónimas
previsional
justicia
bentos
judicial
fray
ciencia
américa
liceos
carrera
tecnología
latina
humano
productivos
ecosistema
fuerzas
capital
procesos
círculo
emprendedor
armadas
vicioso
colectiva
sostenible
modelo
negociación
desarrollo
formación
unidades
crece
estrategia
medios
asistencial
cultura
gobiernos
ejecutoras
alternativos
red
departamentales
naciones
unidas
mayores
adultos
comercio
digital
información
defensoría
siglo
cultural
pública
ernesto
talvi
asentamientos
irregulares
violencia
doméstica
basada


inserción
internacional
tolerancia
cero
educativo
sistema
políticas
sociales
fuerzas
armadas
tecnología
ciencia
chos
anos
cursos
públicos
funcionarios
nacional
nivel
vivienda
digna
superación
personal
fiscales
beneficios
cultural
privado
sector
política
seguridad
social
inclusión
personas


sostenible
frente
laboral
seguridad
crecimiento
amplio
mercado
social
integración
base
familiares
pasta
desarrollo
asignaciones
servicios
climático
productivo
sociales
culturales
cambio
públicos
personas
políticas
privadas
infraestructura
educativos
civil
públicas
vial
básica
empresas
cuentas
sociedad
centros
erior
media
recursos
ineed
inicial
sector
latina
educación
mediano
gráfico
humanos
imaria
plazo
américa
público
cultura
obligatoria
gasto
acuerdos
100
monetaria
comerciales
carcelario
logros
mejores
política
mil
condiciones
sistema
obstáculos
habitantes
fiscal
educativo
nacional
regla
centro
comisionado
ordenamiento
territorial
gobiernos
parlamentario
departamentales
penitenciario


animal
bienestar
negociación
corto
central
mental
colectiva
plazo
regla
ventanilla
única
administración
salud
mediano
fiscal
docente
situación
pública
formación
tomar
frente
información
política
profesional
medidas
amplio
laborales
obra
exterior
orientadas
referida
gobiernos
relaciones
sección
correspondiente
departamentales
nº
ley
alto
vigente
normativa
micro
menores
dedicada
regulatorio
legal
marco
orgánica
rendimiento
pequeñas
ingresos
policial
normativo
personal
nivel
medianas
mayores
años
técnico
experiencia
empresas
tarifas
adultos
protegidas
internacional
inserción
avanzar
áreas
armadas
necesario
europea
tenencia
fuerzas
políticas
específicos
unión
responsables
siguientes
acciones
desarrollar
programas
sociales
atención
impulsar
proponemos
inteligencia
penitenciarias
especial
policía
redes
unidades
gasto
crear
científica
problemático
artificial
reguladoras
énfasis
transporte
consumo
metropolitana
nacionales
crimen
aéreo
área
naciones
física
rutas
código
organizado
económico
unidas
educación
penal
públicos
desarrollo
condiciones
educativo
sistema
integración
ciudadana
infraestructura
mejores
administrativo
contencioso
social
seguridad
vial
prácticas
presupuestales
protección
internacionales
derechos
humana
organismos
recursos
humanos
productiva
emprendedora
naturales
actividad
servicio
uruguaya
sociedad
privadas
económica
personas


burguesía
industrial
francas
educativas
zonas
políticas
agraria
social
reforma
seguridad
deuda
constitucional
enseñanza
capital
pública
financiero
ción
artesanal
pesca
científicas
científico
comunidades
sistema
media
educativo
económica
nacional
independencia
canasta
básica
territorio
derechos
humanos
privados
bancos