

# ‘Media and wolves: Content analysis of French local/national newspapers’

Olivier Gimenez

12/8/2016 (updated: 2020-11-03)

## Introduction

Il s’agit d’une analyse de contenu des corpus loups Le Monde et Nice Matin constitués par Marie Chandelier. On adopte une Approche non-supervisée (on ne définit pas les topics à l’avance, ils sont déterminés statistiquement) récente connue sous le nom de **structural topic modeling** : voir Westgate et al. (2015) pour une introduction et Roberts et al. (2015) pour une implémentation dans R.

Quelques définitions : un **topic** est un mélange de mots où chaque mot a une probabilité d’appartenir au topic. Un **document** (article) est un mélange de topics. La **prévalence** est la proportion d’un document associée à un topic, le **contenu** réfère aux mots utilisés dans un topic.

L’originalité ici est qu’on va essayer d’expliquer la variabilité dans la prévalence et le contenu en fonction de covariables, à savoir l’année (de 1993 à 2014) et le journal (Le Monde vs. Nice Matin).

On commence d’abord par mettre en forme les deux corpus en préparation des analyses statistiques. Chaque article est stocké dans un fichier texte. Tous les articles du Monde sur la période considérée sont dans un répertoire, idem pour Nice Matin.

Une fois le corpus mis au propre, on ajuste un modèle STM. Ici on considère 15 topics et un effet i) de l’interaction de l’année et du journal sur la prévalence, et ii) du journal sur le contenu.

## Création du corpus

On construit d’abord le jeu de données. On charge tous le package nécessaire.

```
library(tm)
library(gdata)
library(stm)
```

On lit les fichiers texte du monde qui sont dans le répertoire spécifié et en fait un corpus:

```
(base_monde <- VCorpus(DirSource("MONDE_thematique_principale", encoding="utf-8"),
  readerControl = list(reader = readPlain, language = "fr", load = TRUE)))

## <<VCorpus>>
## Metadata: corpus specific: 0, document level (indexed): 0
## Content: documents: 120
```

```
#names(base_monde)
#lapply(base_monde, as.character)
```

On récupère les métadonnées puis ajoute les année (1993-2014) et journal (lemonde/nicematin):

```
dat_monde <- read.xls("metada_principale_monde_fonction_ratio.xlsx", header = T)
dim(dat_monde)
```

```
## [1] 123 2
```

```
# supprime les 3 dernieres lignes qui sont bizarres
dat_monde <- dat_monde[-c(121,122,123),]
# on vire les modalités bizarres, puis on convertit année en numérique
dat_monde$annee <- factor(dat_monde$annee)
dat_monde$annee <- as.numeric(levels(dat_monde$annee))[dat_monde$annee] # astuce trouvée http://tinyurl
metadata <- data.frame(year = dat_monde$annee, title = rep('LeMonde',120))
meta(base_monde, tag = c("year","title")) <- metadata
#meta(base_monde)
#base_monde
```

On lit les fichiers texte de Nice Matin qui sont dans le répertoire spécifié et en fait un corpus. A noter : j'ai supprimé les 2 articles de 1992 pour faire coïncider la période avec Le Monde.

```
(base_nice <- VCorpus(DirSource("NICE_thematique_principale", encoding = "UTF-8"),
  readerControl = list(reader=readPlain, language = "fr", load = TRUE)))
```

```
## <<VCorpus>>
## Metadata: corpus specific: 0, document level (indexed): 0
## Content: documents: 744
```

```
#names(base_nice)
#lapply(base_nice, as.character)
```

On récupère les métadonnées puis ajoute année (1993-2014) et journal (lemonde/nicematin):

```
dat_nice <- read.xls("metadata-principale-nice-matin-fonction-ratio.xlsx",
  header = T)
head(dat_nice)
```

```
##          id annee
## 1 LOU-NM-1992-05-02-va 1992
## 2 LOU-NM-1992-11-29-ttes 1992
## 3 LOU-NM-1993-01-09-ttes 1993
## 4 LOU-NM-1993-04-02-ttes 1993
## 5 LOU-NM-1993-04-14-ttes 1993
## 6 LOU-NM-1993-04-15-am 1993
```

```
dim(dat_nice)
```

```
## [1] 746 2
```

```

# on supprime les deux premières lignes qui correspondent à 2 articles en 1992
dat_nice <- dat_nice[-c(1,2),]
metadata <- data.frame(year = dat_nice$annee,
                      title = rep('NiceMatin', 744))
meta(base_nice,tag = c("year","title")) <- metadata
#meta(base_nice)
#base_nice

```

On joint joint les deux bases:

```

base <- c(base_monde, base_nice)
#base
#str(base)
#meta(base)

```

On jette un coup d'oeil à la base:

```

#inspect(base)
#lapply(base, as.character)
base2 <- base

```

On applique tout un tas de traitements aux textes

```

# remplace funny caracteres par un espace
replace_chars <- content_transformer (function(x) gsub("<e0>", "à", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("<ea>", "ê", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("<e9>", "é", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("<e8>", "è", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("<ab>", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("<bb>", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\u0092", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\u0085", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\u0093", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\u0094", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\u0096", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\u0095", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\u0091", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\'", " ", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("\u009c", "œ", x))

```

```

base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("patous","patou", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("ã", "ê", x))
base2 <- tm_map(base2, replace_chars)
replace_chars <- content_transformer (function(x) gsub("ä", "é", x))
base2 <- tm_map(base2, replace_chars)

# capitales deviennent minuscules
base2 <- tm_map(base2, content_transformer(tolower))
#lapply(base2[1:10], as.character)

# aujourd'hui -> aujourdhui
replace_chars <- content_transformer (function(x) gsub("aujourd'hui","aujourd'hui", x))
base2 <- tm_map(base2, replace_chars)

# remplace les apostrophes par des espaces
replace_chars <- content_transformer (function(x) gsub("'", " ", x))
base2 <- tm_map(base2, replace_chars)
#lapply(base2[1:4], as.character)

#Remplace la ponctuation par des espaces (garde les tirets)
base2 <- tm_map(base2, content_transformer(removePunctuation), preserve_intra_word_dashes=T)
#lapply(base2[1:5], as.character)

#retire les articles, prépositions et autres mots analogues
base2 <- tm_map(base2, content_transformer(removeWords), stopwords("french"))
#lapply(base2[1:5], as.character)

#retire les nombres
base2 <- tm_map(base2, content_transformer(removeNumbers))

# tronque à la racine
base2 <- tm_map(base2, stemDocument, language="fr")

#retire les espaces inutiles
base2 <- tm_map(base2, content_transformer(stripWhitespace))
#lapply(base2[100:500], as.character)
#lapply(base2[800:864], as.character)

```

On crée un doc qu'on peut analyser par la suite, une matrice en gros:

```

dtm <- DocumentTermMatrix(base2)
str(dtm)

## List of 6
## $ i      : int [1:152559] 1 1 1 1 1 1 1 1 1 1 ...
## $ j      : int [1:152559] 272 484 544 648 744 772 941 953 1048 1461 ...
## $ v      : num [1:152559] 1 1 2 1 1 1 1 1 1 1 ...
## $ nrow   : int 864
## $ ncol   : int 17289
## $ dimnames:List of 2
## ..$ Docs : chr [1:864] "1993-04-16.txt" "1994-01-10.txt" "1994-08-01.txt" "1994-08-24.txt" ...

```

```
## ..$ Terms: chr [1:17289] "---" "-abri" "-at" "-boutist" ...
## - attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
## - attr(*, "weighting")= chr [1:2] "term frequency" "tf"
```

On récupère le nombre de mots par texte

On calcule la proportion du mot loup

On obtient liste des mots présents au moins à 50 reprises

```
findFreqTerms(dtm,50)
```

```
## [1] "-dessus"          "abattag"          "abattr"
## [4] "abattu"           "abord"            "abruzz"
## [7] "accept"           "action"           "activité"
## [10] "actuel"           "administr"        "administratif"
## [13] "adult"           "affair"           "affirm"
## [16] "affût"           "afin"             "agent"
## [19] "agit"            "agneau"           "agneaux"
## [22] "agricol"         "agriculteur"      "agricultur"
## [25] "aid"             "aider"            "ailleur"
## [28] "ainsi"           "air"              "ajout"
## [31] "alain"           "aller"            "alor"
## [34] "alp"             "alpag"            "alpes--haute-prov"
## [37] "alpes-maritim"   "alpha"            "alpin"
## [40] "altitud"         "aménag"           "ami"
## [43] "analys"          "anim"             "animali"
## [46] "animaux"         "anné"             "annonc"
## [49] "annoncé"         "août"             "appel"
## [52] "approch"         "après"            "après-midi"
## [55] "arc"             "arrêté"           "arriv"
## [58] "arrivé"          "assemblé"         "assez"
## [61] "associ"          "assur"            "attaqu"
## [64] "attaqué"         "atteint"          "attend"
## [67] "attent"          "aucun"            "augment"
## [70] "aujourd'hui"     "auprè"            "aussi"
## [73] "autant"          "authion"          "automn"
## [76] "autoris"         "autorisé"         "autorité"
## [79] "autour"          "autr"             "avanc"
## [82] "avant"           "avenir"           "avoir"
## [85] "avril"           "battu"            "baudin"
## [88] "beaucoup"        "bell"             "belvédèr"
## [91] "berger"          "bergeri"          "bern"
## [94] "bernard"         "besoin"           "bête"
## [97] "bêtes"           "bien"             "biodiversité"
## [100] "blanc"           "blessé"           "boi"
## [103] "bon"             "bonn"             "boréon"
## [106] "bout"           "brebi"            "bruno"
## [109] "caban"          "cadavr"           "cadr"
## [112] "camp"           "campagn"          "cani"
## [115] "canidé"         "canjuer"          "canton"
## [118] "captur"         "car"              "carcass"
## [121] "cas"            "caus"             "cell"
## [124] "celui"          "cent"             "centain"
```

## [127]	"centr"	"central"	"cepend"
## [130]	"certain"	"ceux"	"chacun"
## [133]	"chambr"	"chamoi"	"chaqu"
## [136]	"charg"	"chargé"	"chass"
## [139]	"chasseur"	"chef"	"cheptel"
## [142]	"chevreuil"	"chez"	"chien"
## [145]	"chiffr"	"chose"	"christian"
## [148]	"cinq"	"cinquant"	"clôture"
## [151]	"cœur"	"cohabit"	"col"
## [154]	"colèr"	"comité"	"comm"
## [157]	"commenc"	"comment"	"commiss"
## [160]	"commune"	"comport"	"compt"
## [163]	"concern"	"concerné"	"conclus"
## [166]	"condit"	"conduit"	"confirm"
## [169]	"connaiss"	"connaître"	"connu"
## [172]	"conseil"	"constat"	"cont"
## [175]	"continu"	"contr"	"contrair"
## [178]	"convent"	"côté"	"coup"
## [181]	"coupl"	"cour"	"coût"
## [184]	"craint"	"création"	"danger"
## [187]	"daniel"	"date"	"débat"
## [190]	"début"	"décembr"	"décidé"
## [193]	"décision"	"découvert"	"découvrir"
## [196]	"défendr"	"défens"	"défenseur"
## [199]	"dégâts"	"déjà"	"demand"
## [202]	"demandé"	"demeur"	"département"
## [205]	"départemental"	"depui"	"député"
## [208]	"dernier"	"dernier"	"dès"
## [211]	"désormai"	"destiné"	"destruct"
## [214]	"détermin"	"deux"	"devant"
## [217]	"développ"	"dévoré"	"devrait"
## [220]	"dialogu"	"différent"	"difficil"
## [223]	"difficulté"	"dimanch"	"dire"
## [226]	"direct"	"directeur"	"directric"
## [229]	"disparaître"	"disparit"	"disparu"
## [232]	"dispositif"	"dit"	"dix"
## [235]	"dizain"	"doit"	"doivent"
## [238]	"domestiqu"	"dommag"	"donc"
## [241]	"donné"	"donner"	"dont"
## [244]	"dossier"	"dout"	"douz"
## [247]	"droit"	"durabl"	"durant"
## [250]	"écologi"	"écologist"	"économiqu"
## [253]	"édition"	"effectué"	"effet"
## [256]	"efficac"	"également"	"égorgé"
## [259]	"élevag"	"éleveur"	"élimin"
## [262]	"ell"	"élus"	"enclo"
## [265]	"encor"	"endroit"	"enfant"
## [268]	"enfin"	"enquêt"	"ensembl"
## [271]	"ensuit"	"entr"	"entré"
## [274]	"environ"	"environn"	"envisag"
## [277]	"équilibér"	"éradic"	"errant"
## [280]	"espac"	"espagn"	"espèc"
## [283]	"estim"	"estimé"	"estiv"
## [286]	"estrosi"	"etat"	"état"

## [289]	"être"	"étude"	"euro"
## [292]	"europ"	"européen"	"européenn"
## [295]	"exempl"	"exist"	"expliqu"
## [298]	"exploit"	"face"	"facil"
## [301]	"façon"	"fair"	"fait"
## [304]	"famill"	"faun"	"faut"
## [307]	"favor"	"fdsea"	"fédération"
## [310]	"femell"	"fête"	"février"
## [313]	"fil"	"film"	"fin"
## [316]	"final"	"foi"	"foir"
## [319]	"fond"	"font"	"forêt"
## [322]	"form"	"fort"	"franc"
## [325]	"françai"	"français"	"franco"
## [328]	"fusil"	"gard"	"gaston"
## [331]	"gen"	"gendarm"	"gendarmeri"
## [334]	"général"	"gestion"	"gibier"
## [337]	"gouvern"	"grâce"	"grand"
## [340]	"gros"	"group"	"habit"
## [343]	"haut"	"hautes-alp"	"hauteur"
## [346]	"hectar"	"hélicoptèr"	"heur"
## [349]	"hier"	"histoir"	"hiver"
## [352]	"homm"	"hor"	"huit"
## [355]	"humain"	"idé"	"imag"
## [358]	"import"	"important"	"imposs"
## [361]	"indemnis"	"indemnisé"	"indiqu"
## [364]	"individus"	"inform"	"initi"
## [367]	"instal"	"installé"	"instant"
## [370]	"intérêt"	"isèr"	"issu"
## [373]	"itali"	"italien"	"italienn"
## [376]	"jamai"	"janvier"	"jean"
## [379]	"jean-claud"	"jean-pierr"	"jeudi"
## [382]	"jeun"	"jour"	"journé"
## [385]	"juillet"	"juin"	"jusqu"
## [388]	"just"	"kilomètr"	"liberté"
## [391]	"lieu"	"lieuten"	"lieux"
## [394]	"limit"	"lire"	"livr"
## [397]	"local"	"locaux"	"loi"
## [400]	"loin"	"long"	"longtemp"
## [403]	"lor"	"lorsqu"	"loui"
## [406]	"loup"	"louv"	"louveteaux"
## [409]	"louveteri"	"lundi"	"lupus"
## [412]	"lynx"	"mai"	"mainten"
## [415]	"mair"	"maison"	"mal"
## [418]	"mâle"	"malgré"	"manièr"
## [421]	"manifest"	"mar"	"mardi"
## [424]	"massif"	"mathieu"	"matin"
## [427]	"meilleur"	"membr"	"menac"
## [430]	"menacé"	"mené"	"mercantour"
## [433]	"mercredi"	"mesur"	"métier"
## [436]	"mètres"	"mettr"	"meut"
## [439]	"michel"	"mieux"	"milieu"
## [442]	"mill"	"millier"	"million"
## [445]	"ministèr"	"ministr"	"mis"
## [448]	"mise"	"mission"	"mme"

## [451]	"moi"	"moin"	"moment"
## [454]	"mond"	"montagn"	"mort"
## [457]	"mouflon"	"mouton"	"mouvement"
## [460]	"moyen"	"moyenn"	"nation"
## [463]	"national"	"natur"	"naturel"
## [466]	"naturell"	"nécessair"	"neig"
## [469]	"nice"	"niveau"	"nom"
## [472]	"nombr"	"nombreus"	"nombreux"
## [475]	"non"	"nord"	"notam"
## [478]	"nourrir"	"nouveau"	"nouveaux"
## [481]	"nouvell"	"novembr"	"nuit"
## [484]	"objectif"	"objet"	"observ"
## [487]	"occas"	"octobr"	"œuvr"
## [490]	"offic"	"officiel"	"oncf"
## [493]	"ongulé"	"opérat"	"oppos"
## [496]	"ordr"	"organis"	"organisé"
## [499]	"origin"	"our"	"outr"
## [502]	"ouvert"	"ouvertur"	"ovin"
## [505]	"parc"	"parfoi"	"parl"
## [508]	"parlementair"	"parler"	"parmi"
## [511]	"part"	"parti"	"particip"
## [514]	"particulièr"	"partir"	"pass"
## [517]	"passag"	"passé"	"passer"
## [520]	"passion"	"pastoral"	"pastoralism"
## [523]	"patou"	"pâturag"	"pay"
## [526]	"pendant"	"perdu"	"périod"
## [529]	"permet"	"permett"	"permettr"
## [532]	"permi"	"personn"	"pert"
## [535]	"petit"	"peu"	"peur"
## [538]	"peut"	"peutêtr"	"peuvent"
## [541]	"philipp"	"photo"	"pierr"
## [544]	"place"	"plaint"	"plan"
## [547]	"plateau"	"plein"	"plupart"
## [550]	"plus"	"plusieur"	"plutôt"
## [553]	"point"	"polémiqu"	"politiqu"
## [556]	"popul"	"port"	"porté"
## [559]	"posit"	"possibilité"	"possibl"
## [562]	"pourquoi"	"pourrait"	"pourront"
## [565]	"poursuit"	"pourtant"	"pouvoir"
## [568]	"pratiqu"	"précédent"	"précise"
## [571]	"prédateur"	"prédation"	"préfectur"
## [574]	"préfet"	"prélèvement"	"premier"
## [577]	"premièr"	"prend"	"prendr"
## [580]	"près"	"présenc"	"présent"
## [583]	"présenté"	"président"	"presqu"
## [586]	"pression"	"preuv"	"printemp"
## [589]	"pris"	"prise"	"prix"
## [592]	"probabl"	"problèm"	"procédur"
## [595]	"prochain"	"produit"	"profess"
## [598]	"programm"	"proi"	"projet"
## [601]	"propo"	"propos"	"proposit"
## [604]	"propriétair"	"protect"	"protégé"
## [607]	"protéger"	"protocol"	"proximité"
## [610]	"public"	"pui"	"puisqu"



## [613]	"pyréné"	"qualité"	"quand"
## [616]	"quant"	"quatr"	"quelqu"
## [619]	"question"	"quoi"	"quota"
## [622]	"racont"	"raison"	"randonneur"
## [625]	"rapid"	"rapport"	"rare"
## [628]	"réalis"	"réalisé"	"réalité"
## [631]	"récemment"	"recensé"	"recherch"
## [634]	"regard"	"région"	"régional"
## [637]	"régulat"	"régulier"	"réintroduc"
## [640]	"réintroduit"	"rencontr"	"rendr"
## [643]	"rendu"	"répons"	"représent"
## [646]	"respect"	"respons"	"rest"
## [649]	"résultat"	"retour"	"retrouv"
## [652]	"retrouvé"	"réunion"	"revenu"
## [655]	"rien"	"risqu"	"rôle"
## [658]	"rout"	"roya"	"saint-martin-vésubi"
## [661]	"saison"	"sait"	"sall"
## [664]	"samedi"	"sanglier"	"sauvag"
## [667]	"savoi"	"savoir"	"scientifiqu"
## [670]	"second"	"secteur"	"selon"
## [673]	"semain"	"sembl"	"sen"
## [676]	"sept"	"septembr"	"servic"
## [679]	"seul"	"seulement"	"siècl"
## [682]	"simpl"	"site"	"situat"
## [685]	"situé"	"six"	"société"
## [688]	"soir"	"solut"	"sort"
## [691]	"sorti"	"souhait"	"souligh"
## [694]	"sous"	"sous-préfet"	"soutien"
## [697]	"souvent"	"spécialist"	"spectacl"
## [700]	"subi"	"sud"	"suit"
## [703]	"suivi"	"suivr"	"sujet"
## [706]	"sûr"	"surtout"	"surveil"
## [709]	"survi"	"suscept"	"syndicat"
## [712]	"tant"	"tard"	"techniqu"
## [715]	"tel"	"tell"	"temp"
## [718]	"tend"	"tenu"	"term"
## [721]	"terrain"	"territoir"	"tête"
## [724]	"têtes"	"tiné"	"tir"
## [727]	"tirer"	"total"	"toujour"
## [730]	"tourism"	"tous"	"tout"
## [733]	"toutefoi"	"trace"	"train"
## [736]	"transhum"	"traqu"	"travail"
## [739]	"travaux"	"traver"	"trent"
## [742]	"trentain"	"très"	"tribun"
## [745]	"troi"	"trop"	"troupeau"
## [748]	"troupeaux"	"trouv"	"trouvé"
## [751]	"trouver"	"tué"	"tuée"
## [754]	"tuer"	"utilis"	"vacheri"
## [757]	"vallé"	"var"	"venc"
## [760]	"vendredi"	"venir"	"venu"
## [763]	"venus"	"ver"	"véritabl"
## [766]	"vert"	"vésubi"	"vétérinair"
## [769]	"veut"	"viand"	"victim"
## [772]	"vie"	"viennent"	"vient"

## [775]	"vill"	"villag"	"vingt"
## [778]	"vingtain"	"visit"	"visiteur"
## [781]	"vite"	"vivant"	"vivr"
## [784]	"voie"	"voir"	"voisin"
## [787]	"vont"	"vosg"	"vrai"
## [790]	"vraiment"	"vue"	"zone"

*# On supprime mots rares*

```
dtm2 <- removeSparseTerms(dtm, 0.95)
```

```
findFreqTerms(dtm2, 50)
```

## [1]	"-dessus"	"abattag"	"abattr"
## [4]	"abattu"	"abord"	"abruzz"
## [7]	"accept"	"action"	"activité"
## [10]	"actuel"	"administr"	"administratif"
## [13]	"adult"	"affair"	"affirm"
## [16]	"afin"	"agent"	"agit"
## [19]	"agneau"	"agneaux"	"agricol"
## [22]	"agriculteur"	"agricultur"	"aid"
## [25]	"ailleur"	"ainsi"	"ajout"
## [28]	"aller"	"alor"	"alp"
## [31]	"alpag"	"alpes--haute-prov"	"alpes-maritim"
## [34]	"alpha"	"alpin"	"altitud"
## [37]	"aménag"	"ami"	"analys"
## [40]	"anim"	"animali"	"animaux"
## [43]	"anné"	"annonc"	"annoncé"
## [46]	"août"	"appel"	"approch"
## [49]	"aprè"	"après-midi"	"arrêté"
## [52]	"arriv"	"arrivé"	"assemblé"
## [55]	"assez"	"associ"	"assur"
## [58]	"attaqu"	"attaqué"	"atteint"
## [61]	"attend"	"attent"	"aucun"
## [64]	"augment"	"aujourd'hui"	"auprè"
## [67]	"aussi"	"autant"	"autumn"
## [70]	"autoris"	"autorisé"	"autorité"
## [73]	"autour"	"autr"	"avanc"
## [76]	"avant"	"avenir"	"avoir"
## [79]	"avril"	"battu"	"beaucoup"
## [82]	"bell"	"berger"	"bergeri"
## [85]	"bern"	"bernard"	"bête"
## [88]	"bêtes"	"bien"	"blanc"
## [91]	"blessé"	"bon"	"bonn"
## [94]	"boréon"	"bout"	"brebi"
## [97]	"bruno"	"caban"	"cadavr"
## [100]	"cadr"	"camp"	"campagn"
## [103]	"cani"	"canidé"	"captur"
## [106]	"car"	"carcass"	"cas"
## [109]	"caus"	"cell"	"celui"
## [112]	"cent"	"certain"	"centr"
## [115]	"cepend"	"certain"	"ceux"
## [118]	"chacun"	"chambr"	"chamoi"
## [121]	"chaqu"	"charg"	"chargé"
## [124]	"chass"	"chasseur"	"chef"
## [127]	"cheptel"	"chez"	"chien"

## [130]	"chiffr"	"chose"	"christian"
## [133]	"cinq"	"cinquant"	"clôture"
## [136]	"cœur"	"cohabit"	"colèr"
## [139]	"comm"	"commenc"	"comment"
## [142]	"commiss"	"commune"	"comport"
## [145]	"compt"	"concern"	"concerné"
## [148]	"conclus"	"condit"	"conduit"
## [151]	"confirm"	"connaiss"	"connaître"
## [154]	"connu"	"conseil"	"constat"
## [157]	"continu"	"contr"	"contrair"
## [160]	"convent"	"côté"	"coup"
## [163]	"coupl"	"cour"	"coût"
## [166]	"craint"	"création"	"danger"
## [169]	"date"	"débat"	"début"
## [172]	"décembr"	"décidé"	"décision"
## [175]	"découvert"	"défendr"	"défens"
## [178]	"défenseur"	"dégâts"	"déjà"
## [181]	"demand"	"demandé"	"demeur"
## [184]	"département"	"départemental"	"depuis"
## [187]	"député"	"dernier"	"dernière"
## [190]	"dès"	"désormai"	"destiné"
## [193]	"destruct"	"deux"	"devant"
## [196]	"développ"	"dévoré"	"devrait"
## [199]	"dialogu"	"différent"	"difficil"
## [202]	"difficulté"	"dimanch"	"dire"
## [205]	"direct"	"directeur"	"directric"
## [208]	"disparit"	"disparu"	"dispositif"
## [211]	"dit"	"dix"	"dizain"
## [214]	"doit"	"doivent"	"domestiqu"
## [217]	"dommag"	"donc"	"donné"
## [220]	"donner"	"dont"	"dossier"
## [223]	"dout"	"droit"	"durant"
## [226]	"écologi"	"écologist"	"économiqu"
## [229]	"édition"	"effectué"	"effet"
## [232]	"efficac"	"également"	"égorgé"
## [235]	"élevag"	"éleveur"	"élimin"
## [238]	"ell"	"élus"	"enclo"
## [241]	"encor"	"enfant"	"enfin"
## [244]	"enquêt"	"ensembl"	"ensuit"
## [247]	"entr"	"entré"	"environ"
## [250]	"environn"	"envisag"	"équilibér"
## [253]	"éradic"	"errant"	"espace"
## [256]	"espagn"	"espèc"	"estim"
## [259]	"estimé"	"estiv"	"estrosi"
## [262]	"etat"	"état"	"être"
## [265]	"étude"	"euro"	"europ"
## [268]	"européen"	"européenn"	"exempl"
## [271]	"exist"	"expliqu"	"exploit"
## [274]	"face"	"facil"	"façon"
## [277]	"fair"	"fait"	"famill"
## [280]	"faun"	"faut"	"favor"
## [283]	"fédération"	"femell"	"février"
## [286]	"fil"	"fin"	"final"
## [289]	"foi"	"fond"	"font"

## [292]	"forêt"	"form"	"fort"
## [295]	"franc"	"françai"	"français"
## [298]	"fusil"	"gard"	"gendarm"
## [301]	"gendarmeri"	"général"	"gestion"
## [304]	"gibier"	"gouvern"	"grâce"
## [307]	"grand"	"gros"	"group"
## [310]	"habit"	"haut"	"hautes-alp"
## [313]	"hauteur"	"hectar"	"heur"
## [316]	"hier"	"histoir"	"hiver"
## [319]	"homm"	"hor"	"huit"
## [322]	"humain"	"idé"	"imag"
## [325]	"import"	"important"	"imposs"
## [328]	"indemnis"	"indemnisé"	"indiqu"
## [331]	"individus"	"inform"	"initi"
## [334]	"instal"	"installé"	"instant"
## [337]	"intérêt"	"isèr"	"issu"
## [340]	"itali"	"italien"	"jamai"
## [343]	"janvier"	"jean"	"jeudi"
## [346]	"jeun"	"jour"	"journé"
## [349]	"juillet"	"juin"	"jusqu"
## [352]	"kilomètr"	"liberté"	"lieu"
## [355]	"lieuten"	"lieux"	"limit"
## [358]	"lire"	"local"	"loi"
## [361]	"loin"	"long"	"longtemp"
## [364]	"lor"	"lorsqu"	"loup"
## [367]	"louv"	"louveteaux"	"louveteri"
## [370]	"lundi"	"lupus"	"lynx"
## [373]	"mai"	"mainten"	"mair"
## [376]	"maison"	"mal"	"mâle"
## [379]	"malgré"	"manière"	"manifest"
## [382]	"mar"	"mardi"	"massif"
## [385]	"matin"	"meilleur"	"membr"
## [388]	"menac"	"menacé"	"mené"
## [391]	"mercantour"	"mercredi"	"mesur"
## [394]	"métier"	"mètres"	"mettr"
## [397]	"meut"	"michel"	"mieux"
## [400]	"milieu"	"mill"	"million"
## [403]	"ministèr"	"ministr"	"mis"
## [406]	"mise"	"mission"	"moi"
## [409]	"moin"	"moment"	"mond"
## [412]	"montagn"	"mort"	"mouflon"
## [415]	"mouton"	"mouvement"	"moyen"
## [418]	"nation"	"national"	"natur"
## [421]	"naturel"	"naturell"	"nécessair"
## [424]	"neig"	"nice"	"niveau"
## [427]	"nom"	"nombr"	"nombreus"
## [430]	"nombreux"	"non"	"nord"
## [433]	"notam"	"nourrir"	"nouveau"
## [436]	"nouveaux"	"nouvell"	"novembr"
## [439]	"nuit"	"objectif"	"objet"
## [442]	"observ"	"occas"	"octobr"
## [445]	"œuvr"	"offic"	"officiel"
## [448]	"oncf"	"opérat"	"oppos"
## [451]	"ordr"	"organis"	"organisé"

## [454]	"origin"	"our"	"outr"
## [457]	"ouvert"	"ouvertur"	"ovin"
## [460]	"parc"	"parfoi"	"parl"
## [463]	"parlementair"	"parler"	"parmi"
## [466]	"part"	"parti"	"particip"
## [469]	"particulièr"	"partir"	"pass"
## [472]	"passag"	"passé"	"passer"
## [475]	"passion"	"pastoral"	"pastoralism"
## [478]	"patou"	"pâturag"	"pay"
## [481]	"pendant"	"perdu"	"périod"
## [484]	"permet"	"permett"	"permettr"
## [487]	"permi"	"personn"	"pert"
## [490]	"petit"	"peu"	"peur"
## [493]	"peut"	"peutêtr"	"peuvent"
## [496]	"photo"	"pierr"	"place"
## [499]	"plaint"	"plan"	"plein"
## [502]	"plupart"	"plus"	"plusieur"
## [505]	"plutôt"	"point"	"polémiqu"
## [508]	"politiqu"	"popul"	"port"
## [511]	"porté"	"posit"	"possibilité"
## [514]	"possibl"	"pourquoi"	"pourrait"
## [517]	"pourront"	"poursuit"	"pourtant"
## [520]	"pouvoir"	"pratiqu"	"précédent"
## [523]	"précise"	"prédateur"	"prédation"
## [526]	"préfectur"	"préfet"	"prélèvement"
## [529]	"premier"	"premièr"	"prend"
## [532]	"prendr"	"près"	"présenc"
## [535]	"présent"	"présenté"	"président"
## [538]	"presqu"	"pression"	"preuv"
## [541]	"printemp"	"pris"	"prise"
## [544]	"prix"	"probabl"	"problèm"
## [547]	"procédur"	"prochain"	"produit"
## [550]	"profess"	"programm"	"projet"
## [553]	"propos"	"proposit"	"propriétair"
## [556]	"protect"	"protégé"	"protéger"
## [559]	"proximité"	"public"	"pui"
## [562]	"puisque"	"pyréné"	"qualité"
## [565]	"quand"	"quant"	"quatr"
## [568]	"quelqu"	"question"	"quoi"
## [571]	"racont"	"raison"	"rapid"
## [574]	"rapport"	"réalis"	"réalisé"
## [577]	"réalité"	"récemment"	"recensé"
## [580]	"recherch"	"région"	"régional"
## [583]	"régulat"	"régulier"	"réintroduc"
## [586]	"réintroduit"	"rencontr"	"rendr"
## [589]	"rendu"	"répons"	"représent"
## [592]	"respect"	"respons"	"rest"
## [595]	"résultat"	"retour"	"retrouv"
## [598]	"retrouvé"	"réunion"	"revenu"
## [601]	"rien"	"risqu"	"rôle"
## [604]	"rout"	"roya"	"saint-martin-vésubi"
## [607]	"saison"	"sait"	"samedi"
## [610]	"sanglier"	"sauvag"	"savoir"
## [613]	"scientifiqu"	"second"	"secteur"

## [616]	"selon"	"semain"	"sembl"
## [619]	"sen"	"sept"	"septembr"
## [622]	"servic"	"seul"	"seulement"
## [625]	"siècl"	"simpl"	"site"
## [628]	"situat"	"situé"	"six"
## [631]	"société"	"soir"	"solut"
## [634]	"sort"	"sorti"	"souhait"
## [637]	"soulign"	"sous"	"soutien"
## [640]	"souvent"	"spécialist"	"spectacl"
## [643]	"subi"	"sud"	"suit"
## [646]	"suivi"	"suivr"	"sujet"
## [649]	"sûr"	"surtout"	"surveil"
## [652]	"survi"	"suscept"	"syndicat"
## [655]	"tant"	"tard"	"techniqu"
## [658]	"tel"	"tell"	"temp"
## [661]	"tend"	"tenu"	"term"
## [664]	"terrain"	"territoir"	"tête"
## [667]	"têtes"	"tiné"	"tir"
## [670]	"tirer"	"total"	"toujour"
## [673]	"tous"	"tout"	"toutefoi"
## [676]	"trace"	"train"	"travail"
## [679]	"traver"	"trent"	"trentain"
## [682]	"très"	"tribun"	"troi"
## [685]	"trop"	"troupeau"	"troupeaux"
## [688]	"trouv"	"trouvé"	"trouver"
## [691]	"tué"	"tuée"	"tuer"
## [694]	"utilis"	"vallé"	"var"
## [697]	"vendredi"	"venir"	"venu"
## [700]	"venus"	"ver"	"véritabl"
## [703]	"vésubi"	"vétérinair"	"veut"
## [706]	"viand"	"victim"	"vie"
## [709]	"viennent"	"vient"	"vill"
## [712]	"villag"	"vingt"	"vingtain"
## [715]	"visit"	"visiteur"	"vite"
## [718]	"vivant"	"vivr"	"voie"
## [721]	"voir"	"voisin"	"vont"
## [724]	"vrai"	"vraiment"	"vue"
## [727]	"zone"		

```
dtm2 = dtm
```

On extrait base pour le monde et nice matin (après traitement et avant traitement)

```
lemonde2 <- subset(base2,meta(base2)$title=='LeMonde')
nicematin2 <- subset(base2,meta(base2)$title=='NiceMatin')
lemonde <- subset(base,meta(base)$title=='LeMonde')
nicematin <- subset(base,meta(base)$title=='NiceMatin')
```

On lit corpus entier pour analyse via package STM

```
out <- readCorpus(dtm2, type = "Matrix")
docs <- out$documents # textes
vocab <- out$vocab # mots
```

```
meta <-meta(base)
#head(meta)
```

## Analyse de contenu

On analyse le jeu de données via l'approche topic modeling. On analyse les corpus Le Monde et Nice Matin ensemble, avec le titre comme covariable sur la prévalence et le contenu (on ne distingue pas le type épisodique/thématique). On ajuste un modèle stm avec 15 topics. J'ai essayé 5, 10, 20, 25 et il semble que ça soit la solution la plus interprétable. On sauvegarde les résultats car l'ajustement prend beaucoup de temps.

```
poliblogPrevFit_15 <- stm(documents = docs,
                        vocab = vocab,
                        K = 15,
                        prevalence =~ title + s(year),
                        content =~ title,
                        max.em.its = 75,
                        data = meta,
                        init.type = "Spectral")
save(poliblogPrevFit_15, file = 'stm_marie.RData')
```

On charge le résultat de l'analyse.

```
load('stm_marie.RData')
wolf_stm <- poliblogPrevFit_15
```

On affiche les mots associés à chaque topic, d'abord dans l'ordre, puis par importance ie fréquence dans corpus :

```
labelTopics(wolf_stm)
```

```
## Topic Words:
## Topic 1: vittel, exhum, hantis, découvert, belvédér, muséum, cadavr
## Topic 2: défil, gap, gaymard, reçus, lepelti, herv, serg
## Topic 3: mytholog, ironis, chauffeur, der, dysfonction, cirqu, scandinav
## Topic 4: prompt, fui, bël, nancy, voulez-, solitud, raid
## Topic 5: marboutin, éventr, descendu, barengo, achemin, caill, désalter
## Topic 6: tenac, documentair, roumain, primair, scolaire, film, battent
## Topic 7: viseur, jean-david, expérimental, réfé, abel, olin, intensifi
## Topic 8: grandjean, prosper, gardes-moniteur, impass, simon, guth, émul
## Topic 9: anem, mir, dupont, relativis, chevalli, lobby, artus
## Topic 10: paradox, blaureau, récolt, léger, réapprendr, poudreux, paradoxal
## Topic 11: porch, bastien, petite-fill, bardeau, dîn, august, puc
## Topic 12: sédentair, bio, poguntk, guigon, producteur, folklor, cabinet
## Topic 13: veillent, régress, soixante-dix, assaut, tueur, docil, obéiss
## Topic 14: boss, mongol, là-haut, toundr, julien, authent, vaporis
## Topic 15: pastoraloup, bénévol, héliportag, wwf, chevr, programm, estiv
##
## Covariate Words:
## Group LeMonde: épaul, valoir, savoi, kilometr, radio, perçu, vis-àv
```

```

## Group NiceMatin: destin, noir, âge, édit, semain, retir, poursuivre
##
## Topic-Covariate Interactions:
## Topic 1, Group LeMonde: morpholog, exod, france-nature-environ, aout, dépêchesl, réappropri, réveill
## Topic 1, Group NiceMatin: vomissur, toxicolog, loub, loubet, lingosti, raticid, bousiéri
##
## Topic 2, Group LeMonde: émotionnel, guetteur, nomin, bouteil, intoler, cooper, infrastructur
## Topic 2, Group NiceMatin: tract, géraldin, cannel, félix, massen, automobil, colmian
##
## Topic 3, Group LeMonde: géraldin, vallet, énerv, éprouv, attrist, sec, crim
## Topic 3, Group NiceMatin: fléchet, niemei, trappeur, hirsch, colli, arkétal, osman
##
## Topic 4, Group LeMonde: covarel, foix, madres, dino, vint, madam, exil
## Topic 4, Group NiceMatin: bombardi, piral, terrass, tabourd, stourz, bargemon, bess
##
## Topic 5, Group LeMonde: boucley, signatair, haute-marn, parametr, lumin, guet, atteign
## Topic 5, Group NiceMatin: verdeg1, briasq, colet, dozol, jonathan, fabron, audibergu
##
## Topic 6, Group LeMonde: pum, prioritaire, conquer, tardiv, mitig, réform, bérang
## Topic 6, Group NiceMatin: club, lycéen, varg, resurg, barnab, atou, bigarad
##
## Topic 7, Group LeMonde: désamorc, efforc, goutt, fassent, boucli, ramass, darmstaedt
## Topic 7, Group NiceMatin: candon, brunelot, chery, mouri, lucéram, lamy, prorog
##
## Topic 8, Group LeMonde: cantabr, resurg, encher, somiedo, égorgeur, profondeur, équivoqu
## Topic 8, Group NiceMatin: attentat, italicus, désamorc, fusion, barbu, rénov, guadeloup
##
## Topic 9, Group LeMonde: adet, rigon, cannel, atav, fusion, soupçonnent, rejoign
## Topic 9, Group NiceMatin: savornin, danon, germain, turb, signatair, jospin, triomph
##
## Topic 10, Group LeMonde: diabl, instig, doutent, destructeur, sculpteur, couron, gent
## Topic 10, Group NiceMatin: dobrem, chaînon, rabattr, fog, franz-olivi, giesbert, estéron
##
## Topic 11, Group LeMonde: raréfact, brandebourg, chant, divert, dûment, inépuis, certificat
## Topic 11, Group NiceMatin: muséograph, profondeur, clairi, ludden, sculpteur, réhabl, sensat
##
## Topic 12, Group LeMonde: ain, sensat, moloss, nomad, filiat, extraire, usé
## Topic 12, Group NiceMatin: ballarello, danseur, prieur, vassalo, forain, authi, jean-mario
##
## Topic 13, Group LeMonde: génom, lauren, taberlet, grenoblois, intensif, cascad, radios
## Topic 13, Group NiceMatin: vallet, destructeur, éprouv, rein, photo-électr, radar, galin
##
## Topic 14, Group LeMonde: tison, varg, thorac, facial, clairi, ressurg, postérieur
## Topic 14, Group NiceMatin: fran, casse-dall, mohawk, rédac, tois, truc, x-ray
##
## Topic 15, Group LeMonde: chevreau, christin, aveyron, babouin, boiteux, chimpanz, croat
## Topic 15, Group NiceMatin: stag, etou, st-martin-vésub, giordan, numer, écovolontair, engagé
##

```

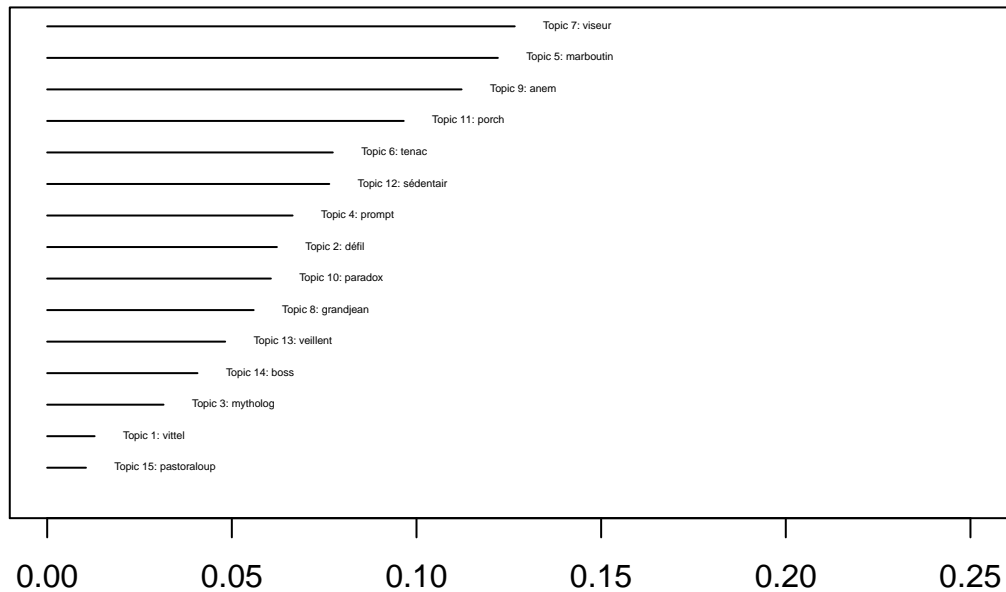
```

plot(wolf_stm,
     type = "summary",
     n = 1,
     text.cex = 0.3)

```



## Top Topics



## Expected Topic Proportions

On cherche les 15 topics avec le plus de poids. On jette un coup d'oeil sur GitHub à la fonction `plot.STM` qu'on adapte pour nos besoins :

```
x <- wolf_stm
model <- x
contentcov <- length(model$beta$logbeta) != 1
type <- "summary"
n <- 1
topics <- NULL
labeltype <- "prob"
frexw <- .5
custom.labels <- NULL
topic.names <- NULL
if(!is.null(custom.labels)) labeltype <- "custom"
if(is.null(n)) n <- switch(type,
                           summary=3,
                           labels=20,
                           perspectives=25,
                           hist=3)
if(type!="perspectives" & is.null(topics)) topics <- 1:model$settings$dim$K

if(labeltype!="custom"){
  if(type != "perspectives") {
    lab <- labelTopics(model, topics=topics, n = n, frexweight=frexw)
    if(contentcov) {
      lab <- lab$topics
    } else {
      lab <- lab[[labeltype]]
    }
  }
}
```

```

} else {
  lab <- custom.labels
  if(length(lab)!=length(topics)) lab <- rep_len(lab, length.out=length(topics))
}

if(!is.null(topic.names)) topic.names <- rep_len(topic.names, length.out=length(topics))

frequency <- colMeans(model$theta[,topics]) # expected topic prop de pour les 100 premiers topics
rank <- order(frequency, decreasing=TRUE)
rank[1:15] # les 15 premiers topics

```

```
## [1] 7 5 9 11 6 12 4 2 10 8 13 14 3 1 15
```

```
frequency[rank[1:15]] # et leur poids
```

```
## [1] 0.12660439 0.12203429 0.11220653 0.09654183 0.07733947 0.07638318
## [7] 0.06645444 0.06220570 0.06056881 0.05591196 0.04819539 0.04070124
## [13] 0.03150671 0.01284486 0.01050120
```

```
sum(frequency[rank[1:15]]) # et leur poids
```

```
## [1] 1
```

J'en profite pour refaire le graphe des Top Topics en ne prenant que les 15 premiers topics classés par ordre croissant de l'Expected Topic Proportion:

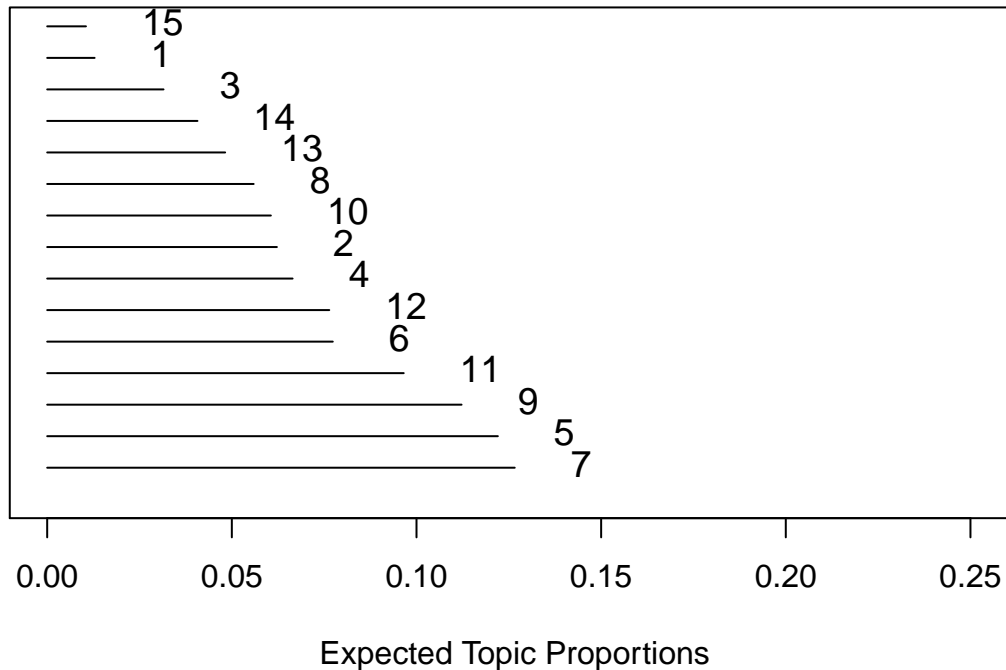
```

xlim <- c(0,min(2*max(frequency), 1))
ylim <- c(0,15)
main <- "Top Topics"
xlab <- "Expected Topic Proportions"
ylab <- ""

plot(c(0,0),
     type = "n",
     xlim = xlim,
     ylim = ylim,
     main = main,
     yaxt = "n",
     ylab = ylab,
     xlab = xlab)
for(i in 1:15) {
  lines(c(0,frequency[rank[i]]), c(i, i))
  text(frequency[rank[i]] + .01, i , rank[i], pos = 4, cex = 1.2)
}

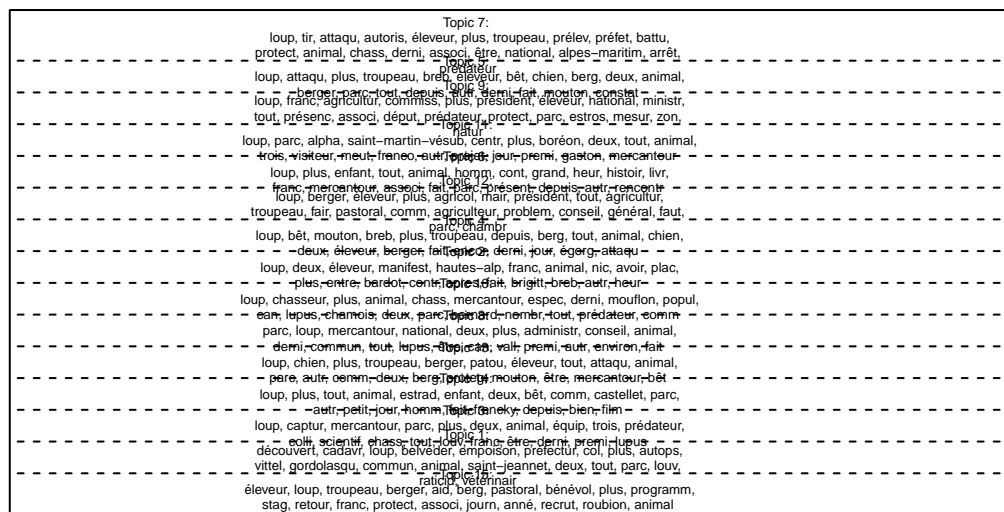
```

## Top Topics



Les mots indicatifs pour topic listes :

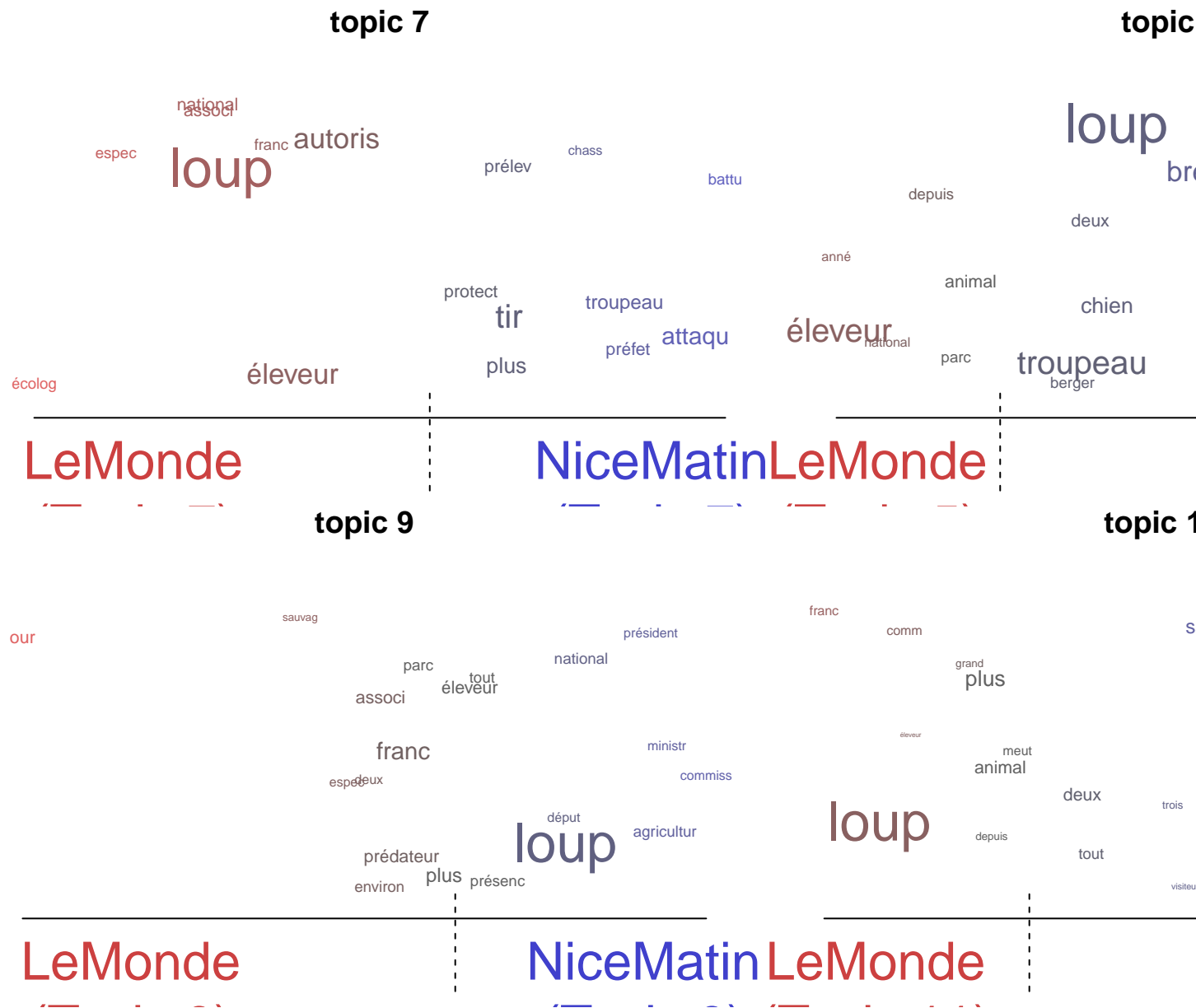
```
#ppi <- 300
#name.fig <- 'label_topic.png'
#png(name.fig, width = 6*ppi, height = 6*ppi, res = ppi)
plot(wolf_stm, type = "labels", topics = rank[1:15], text.cex = 0.4)
```



```
#dev.off()
```

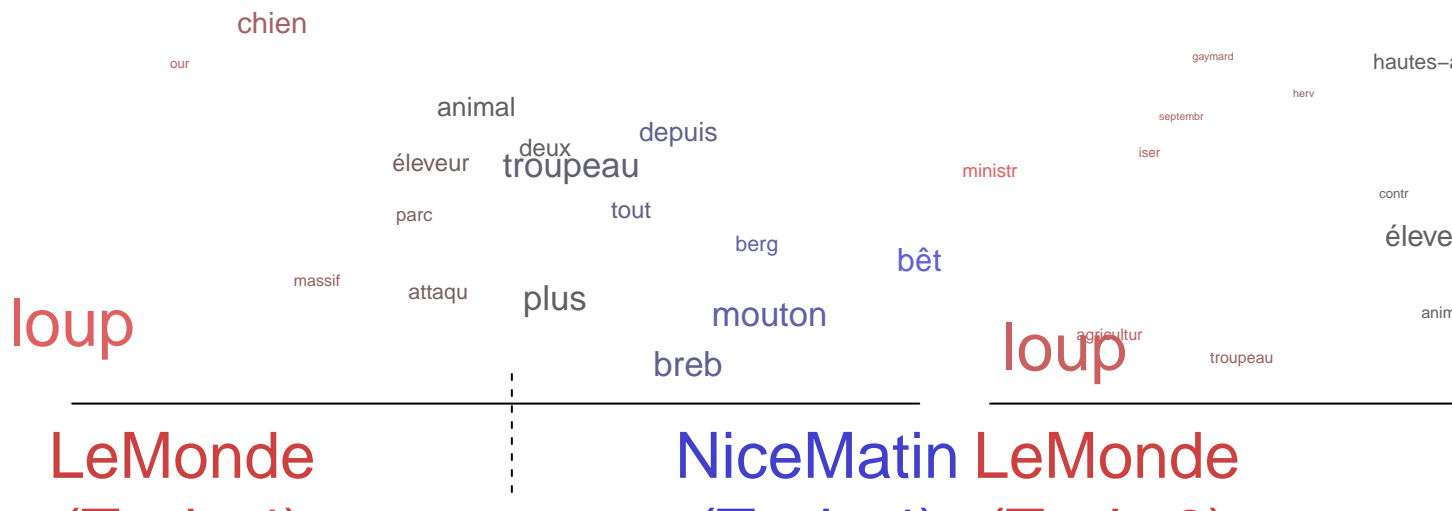
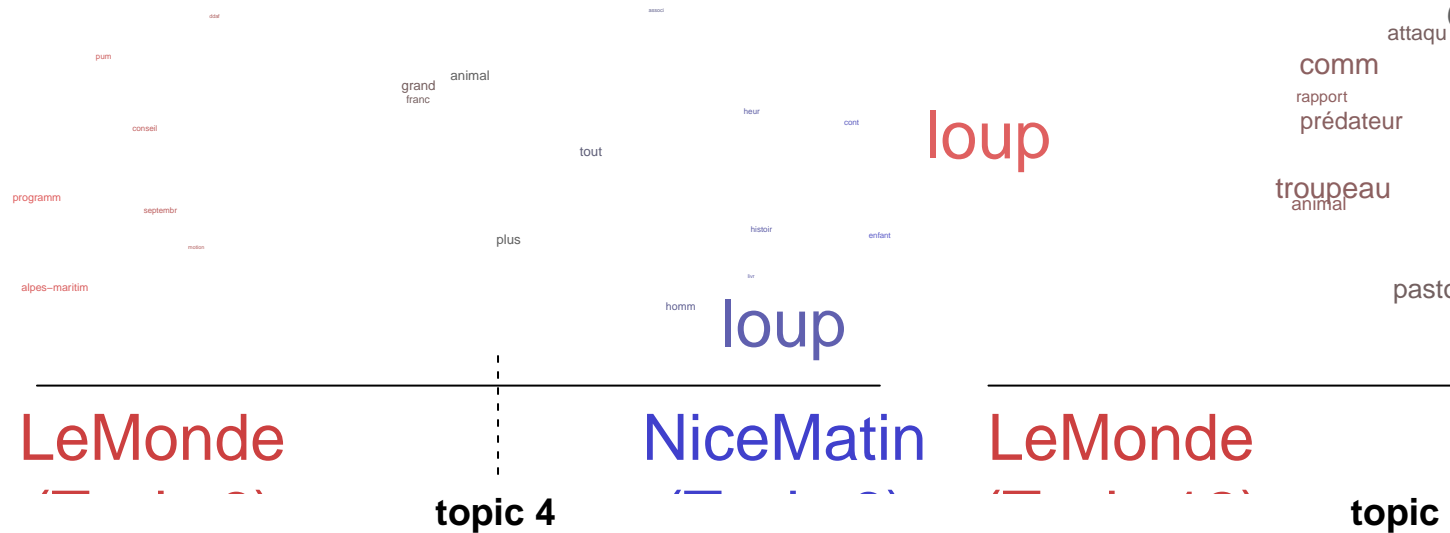
La différence dans le contenu d'un topic en fonction de la covariable contenu (ici le journal)

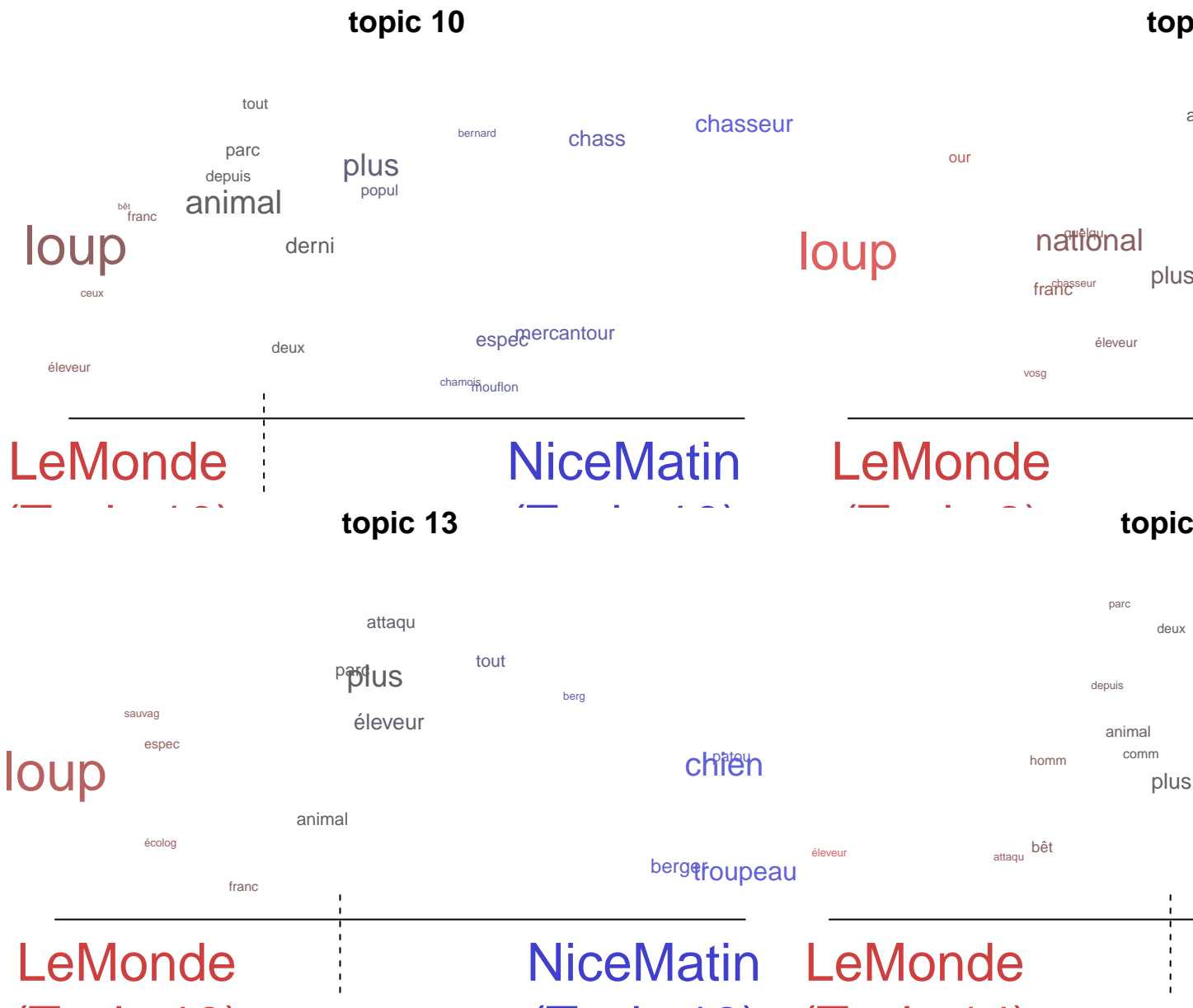
```
# Make a 6x6 inch image at 300dpi
#pi <- 300
for (i in rank[1:15]){
  #name.fig <- paste('diff_topic',i,'.png',sep='')
  #png(name.fig, width = 6*ppi, height = 6*ppi, res = ppi)
  plot(wolf_stm, type = "perspectives", topics = i, main = paste("topic ",i,sep=''))
  #dev.off()
}
```

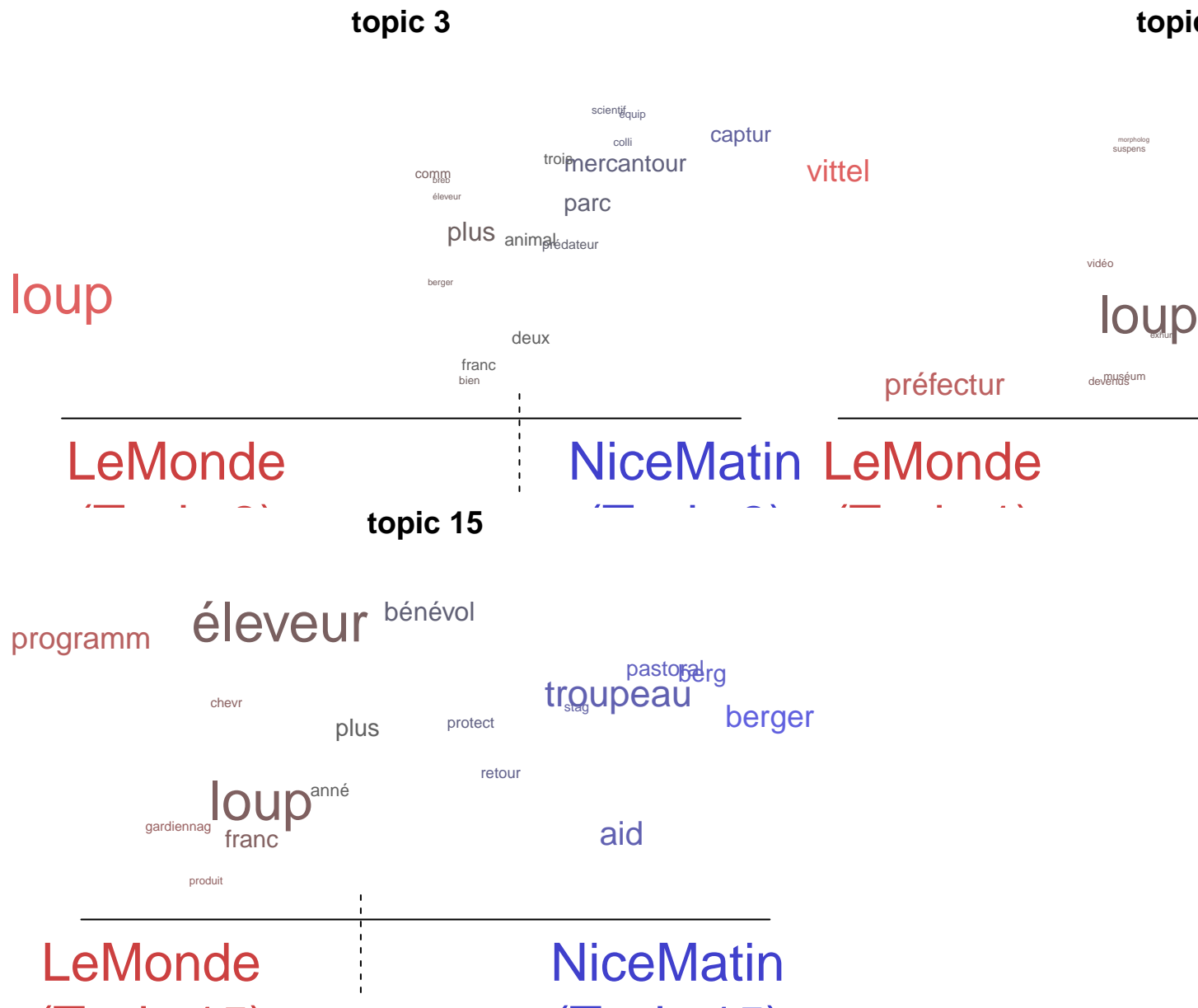


topic 6

topic 1







On regarde l'évolution temporelle de la proportion du corpus que le topic occupe (en rouge, Nice Matin ; en bleu, Le Monde):

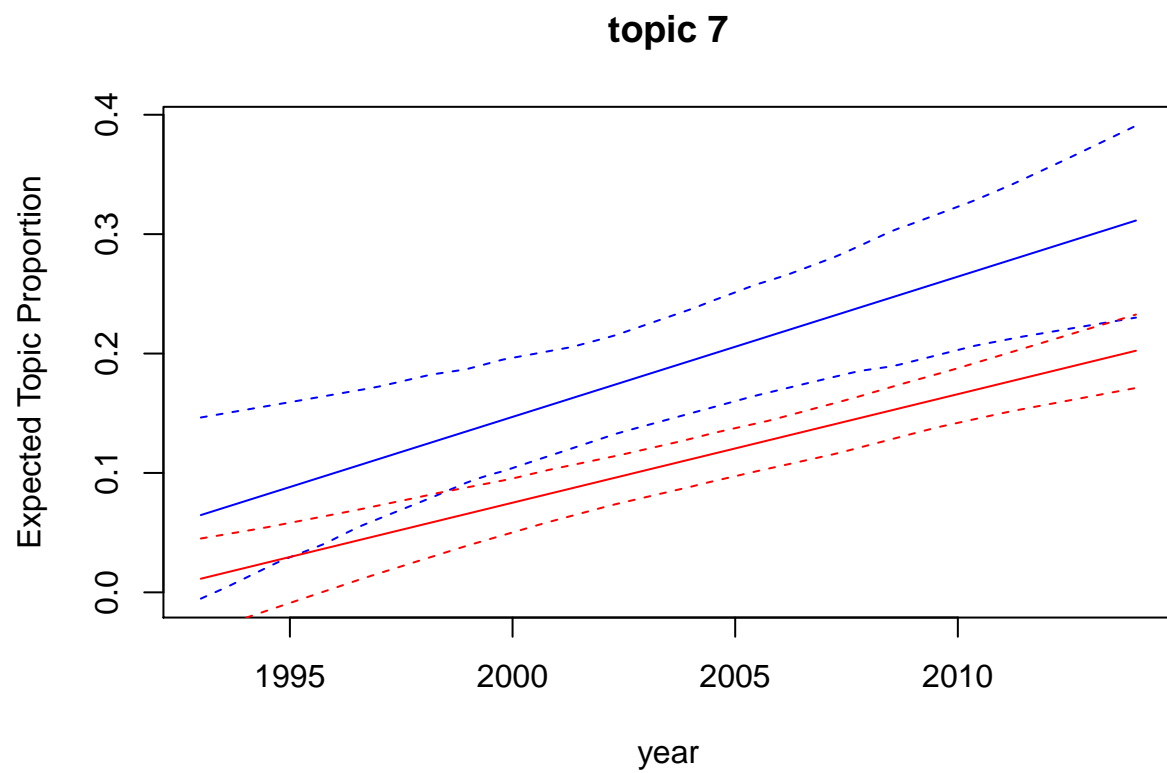
```
#ppi <- 300
for (i in rank[1:15]){
  #name.fig = paste('trend_topic_add', i, '.png', sep='')
  #png(name.fig, width = 6*ppi, height = 6*ppi, res = ppi)
  prep <- estimateEffect(c(i) ~ title * year,
                        wolf_stm,
                        metadata = meta,
                        uncertainty = "None")

  prep$parameters
  plot.estimateEffect(prep,
                    covariate = "year",
                    model = wolf_stm,
```

```

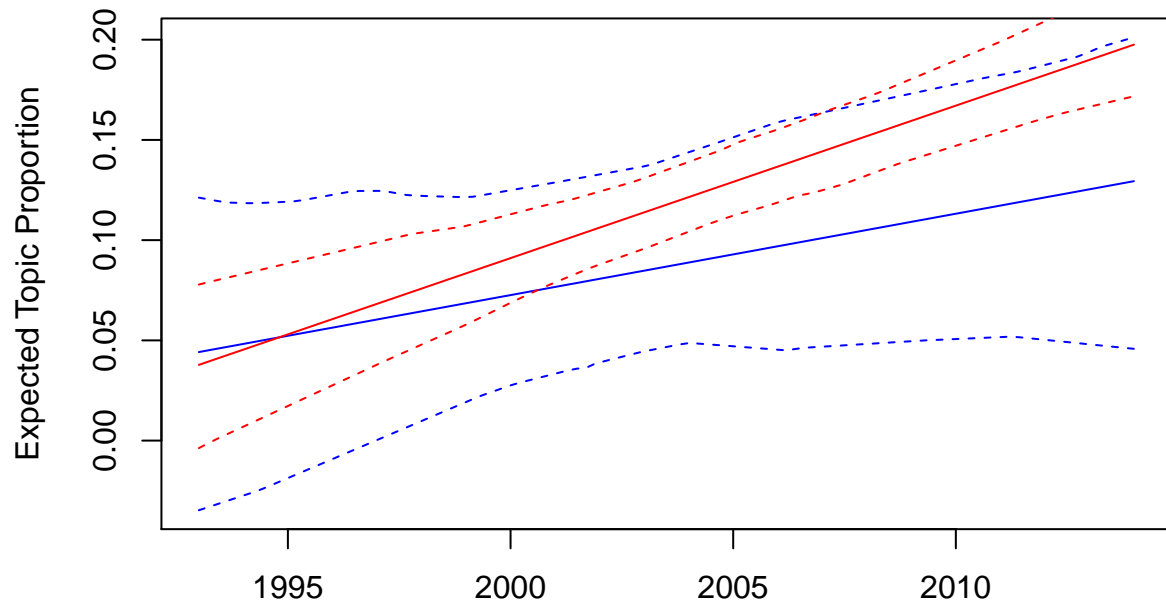
        method = "continuous",
        xlab = "year",
        moderator = "title",
        moderator.value = "LeMonde",
        linecol = "blue",
        printlegend = F,
        main = paste("topic",i))
plot.estimateEffect(prep,
                    covariate = "year",
                    model = wolf_stm,
                    method = "continuous",
                    xlab = "year",
                    moderator = "title",
                    moderator.value = "NiceMatin",
                    linecol = "red",
                    add = T,
                    printlegend = F)
#dev.off()
}

```

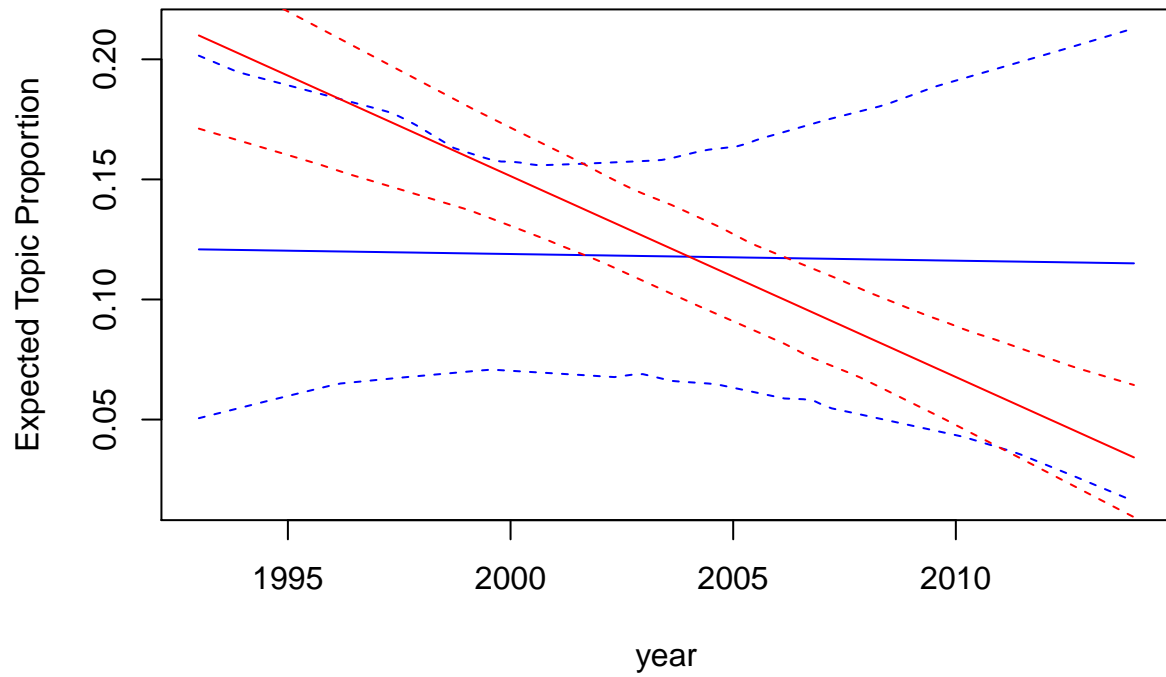




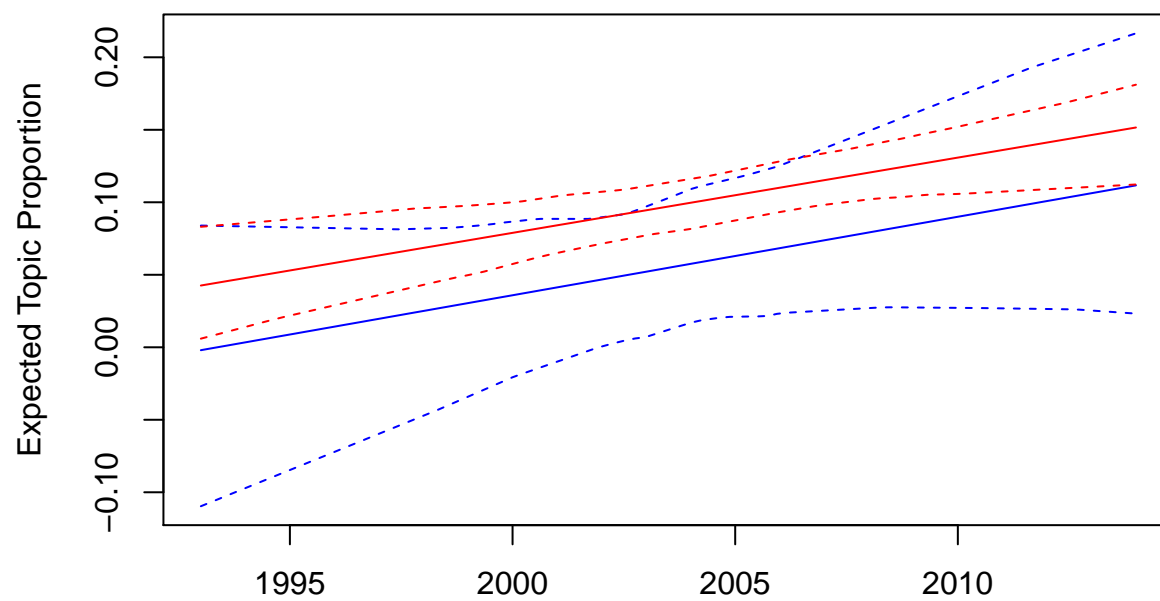
**topic 5**



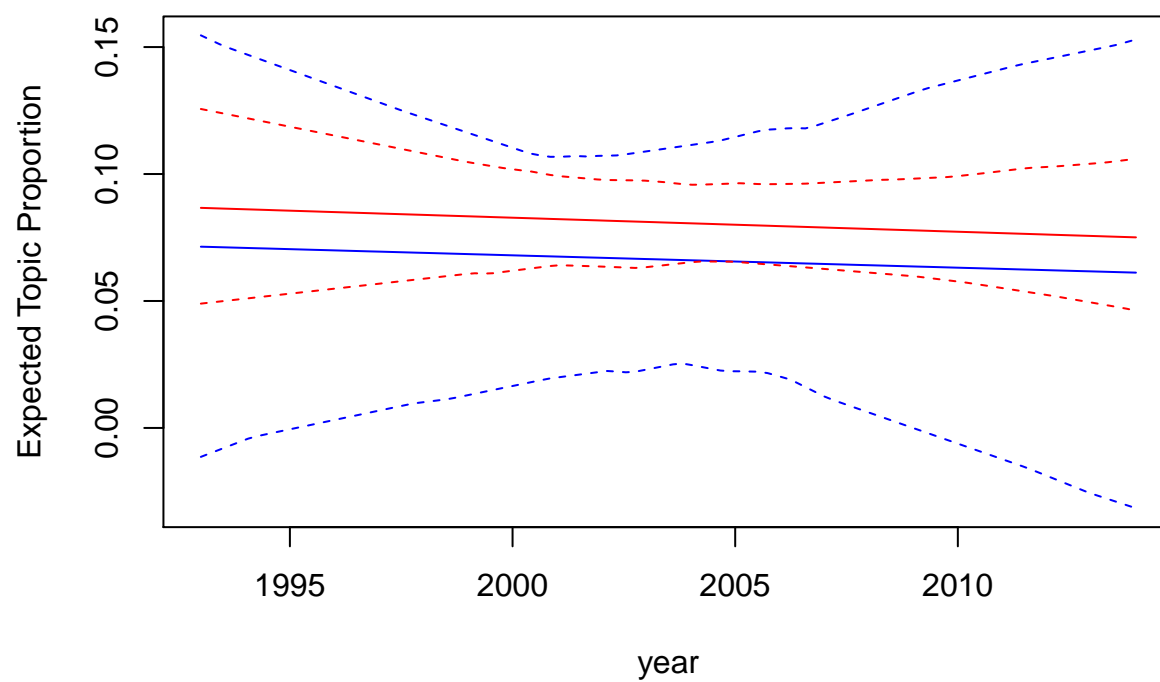
**topic 9**



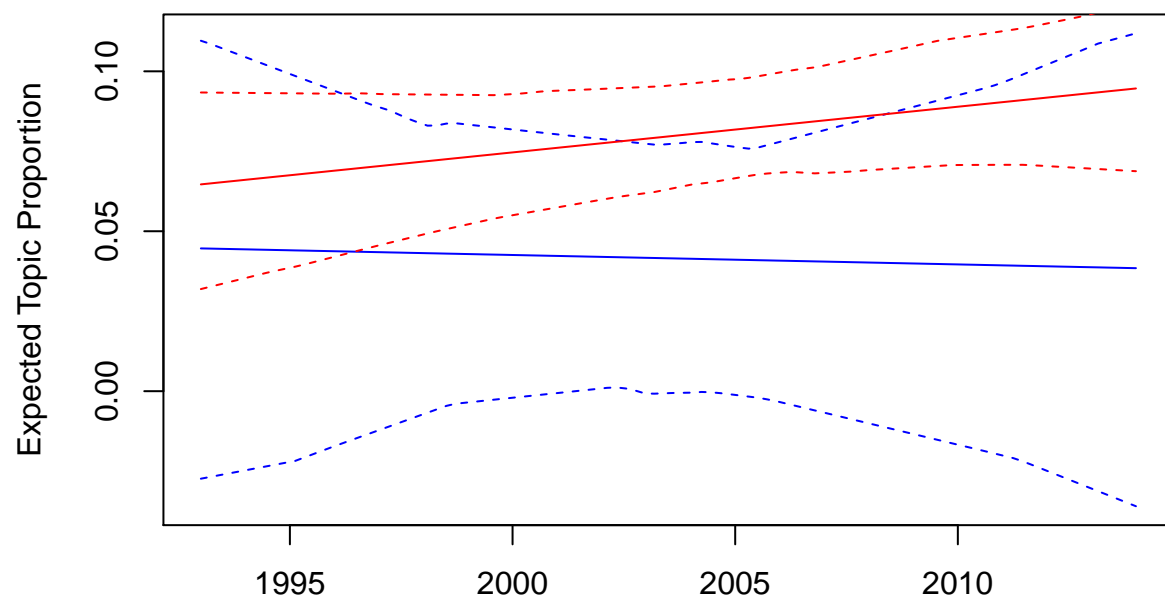
**topic 11**



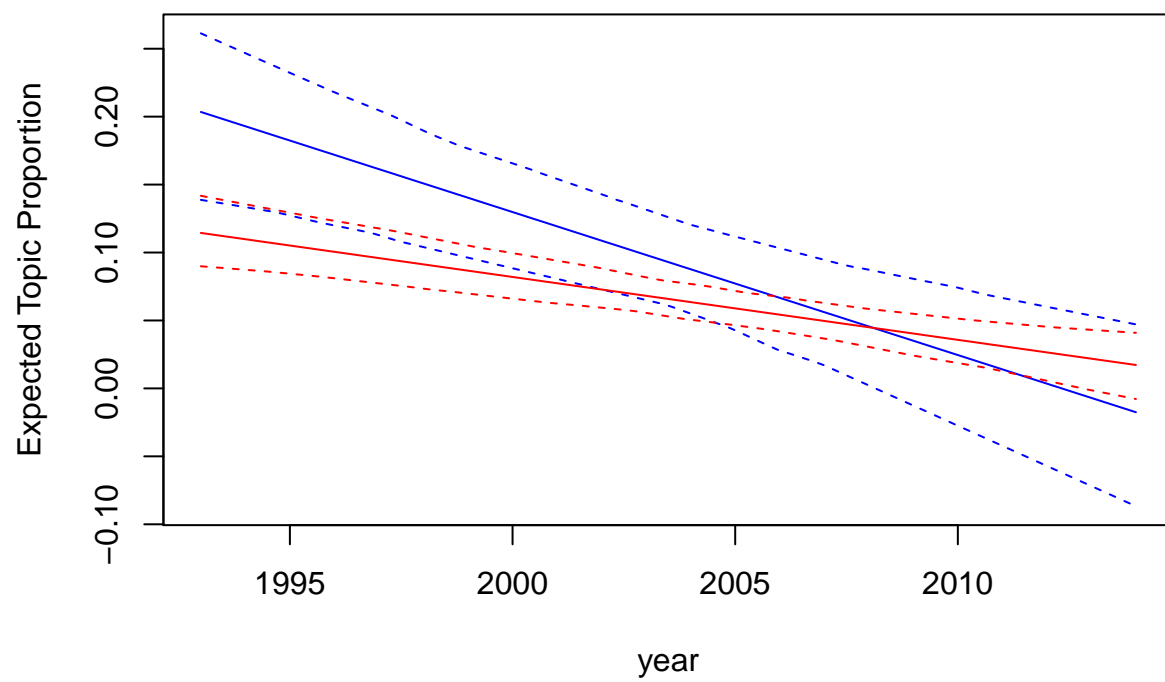
**topic 6**



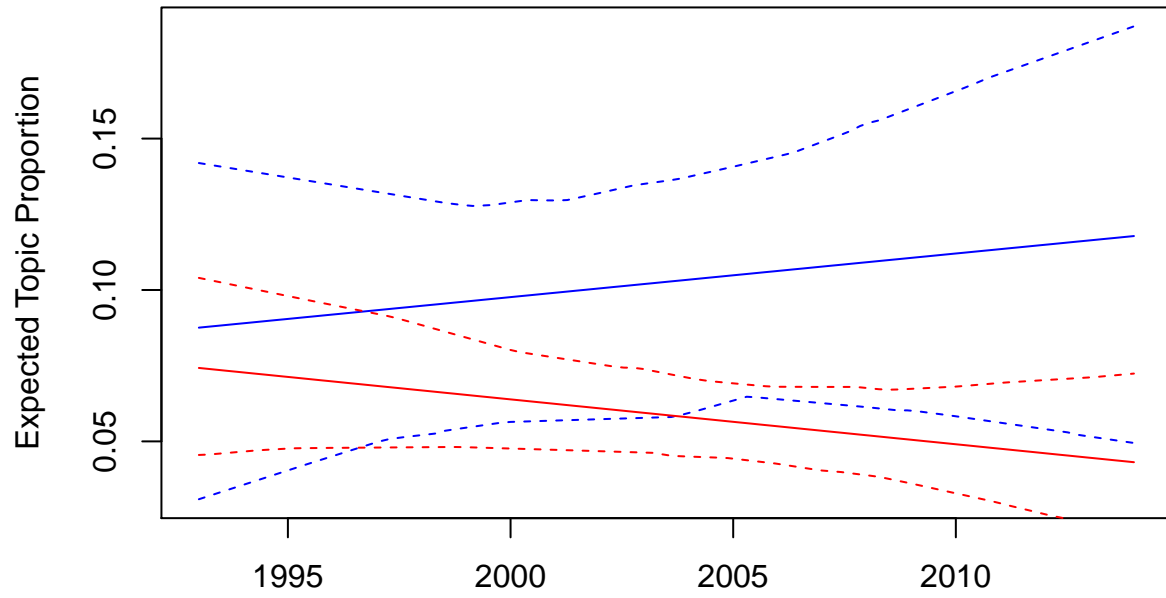
**topic 12**



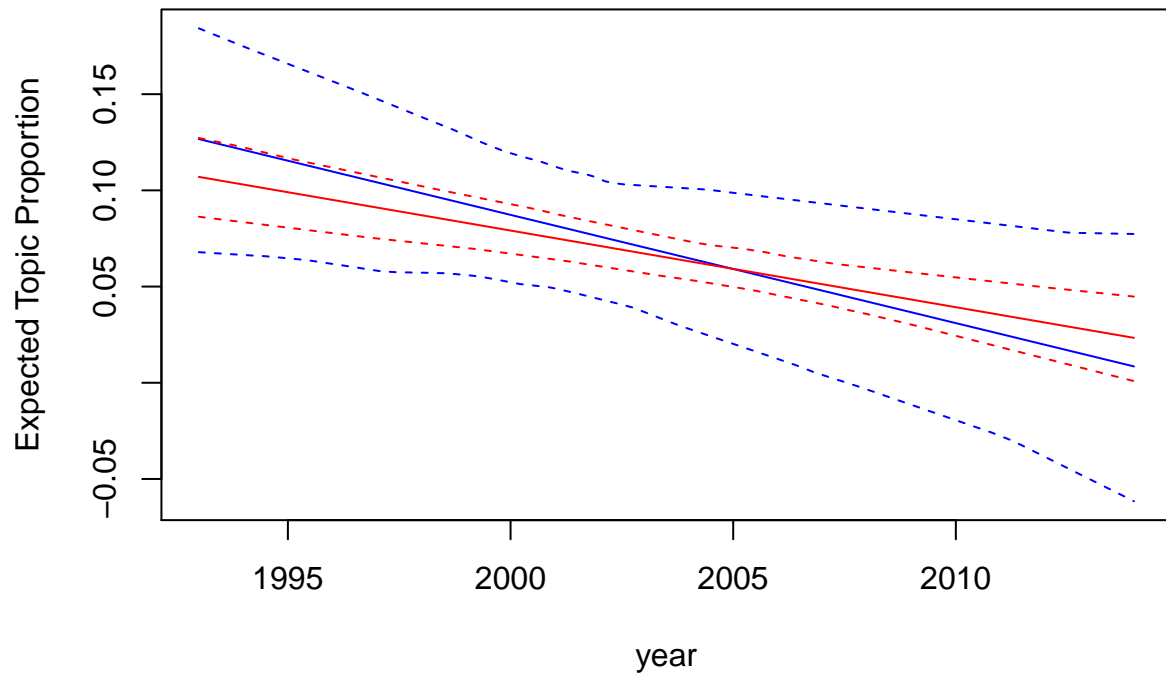
**topic 4**



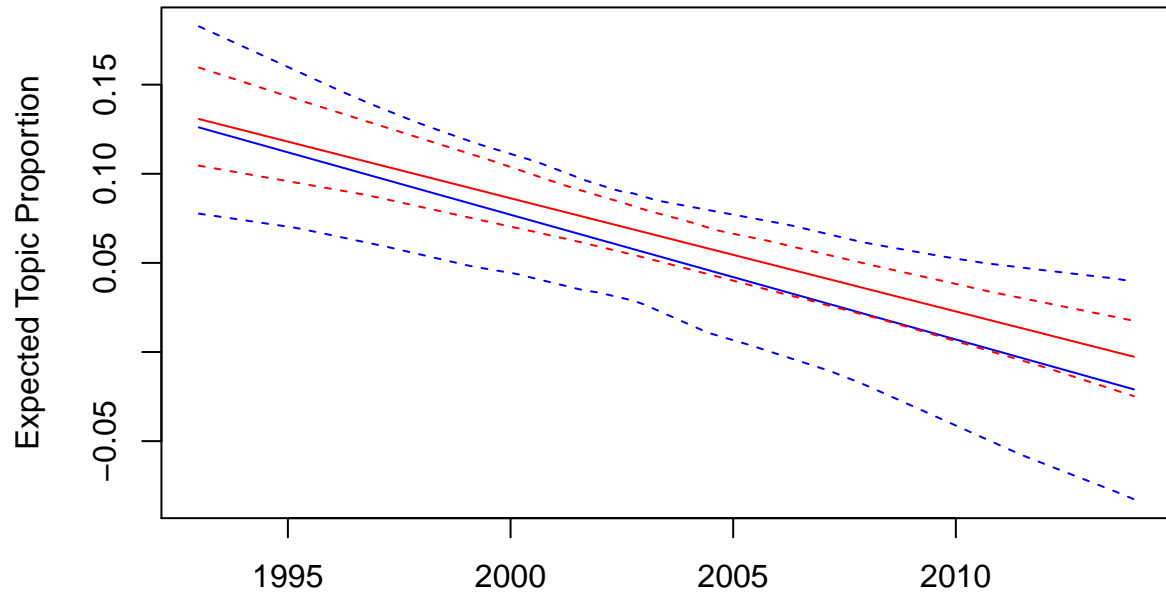
**topic 2**



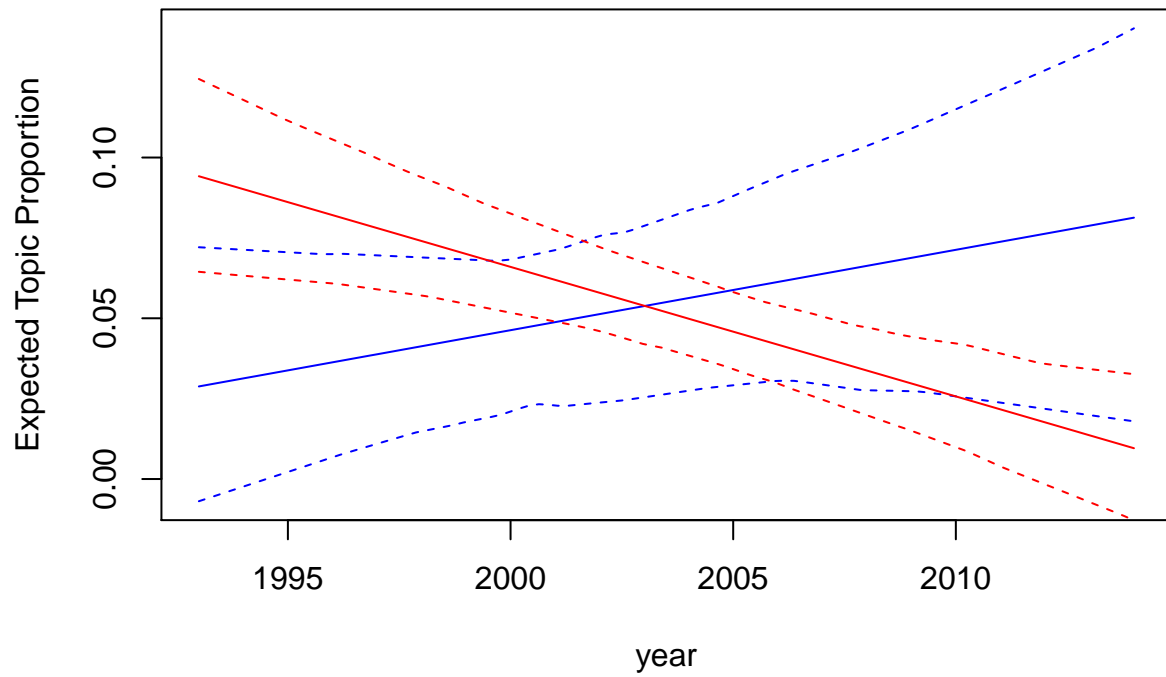
**topic 10**



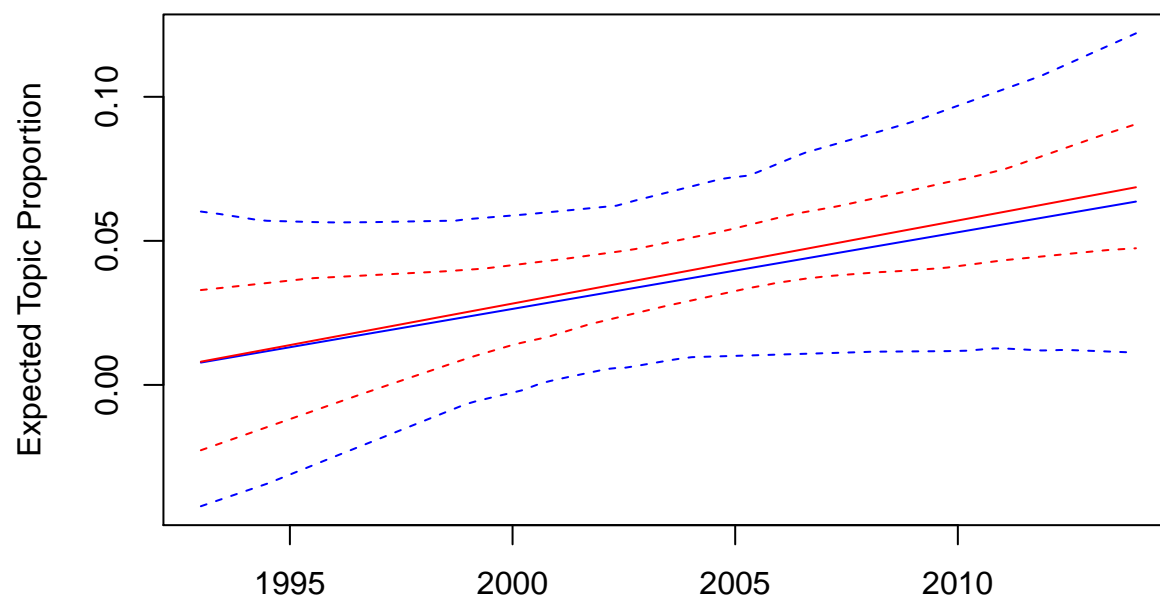
**topic 8**



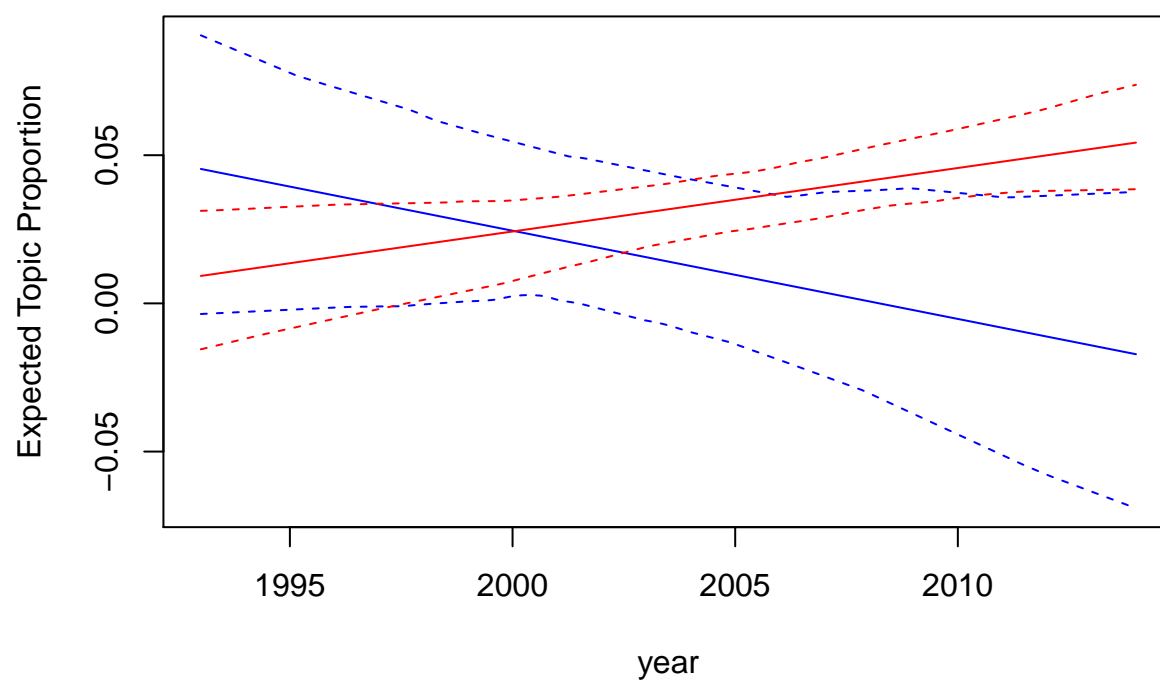
**topic 13**



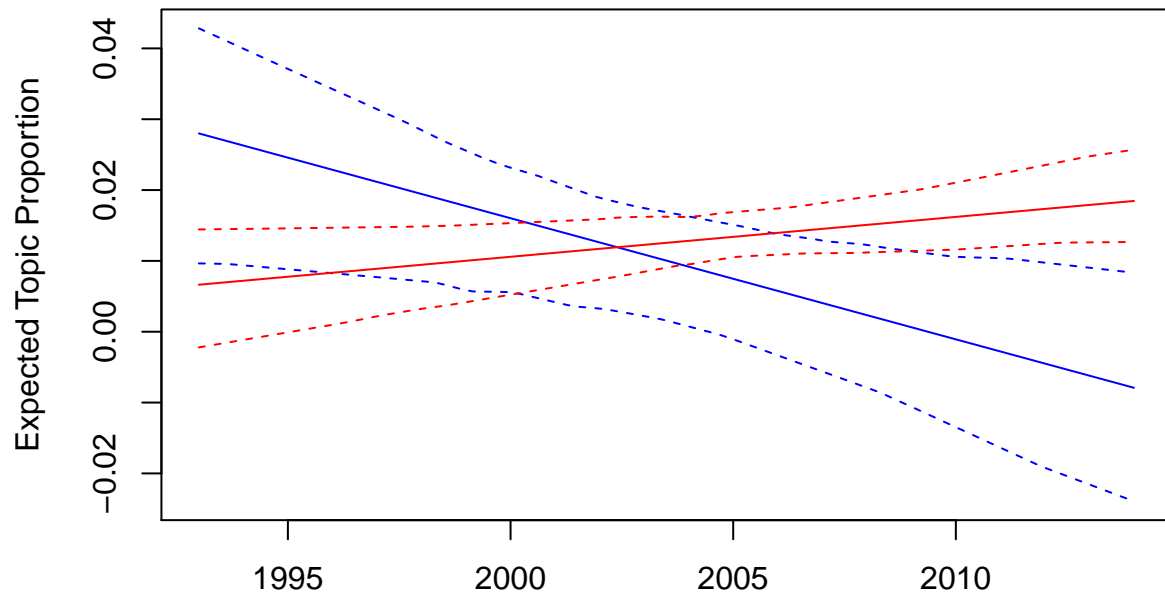
**topic 14**



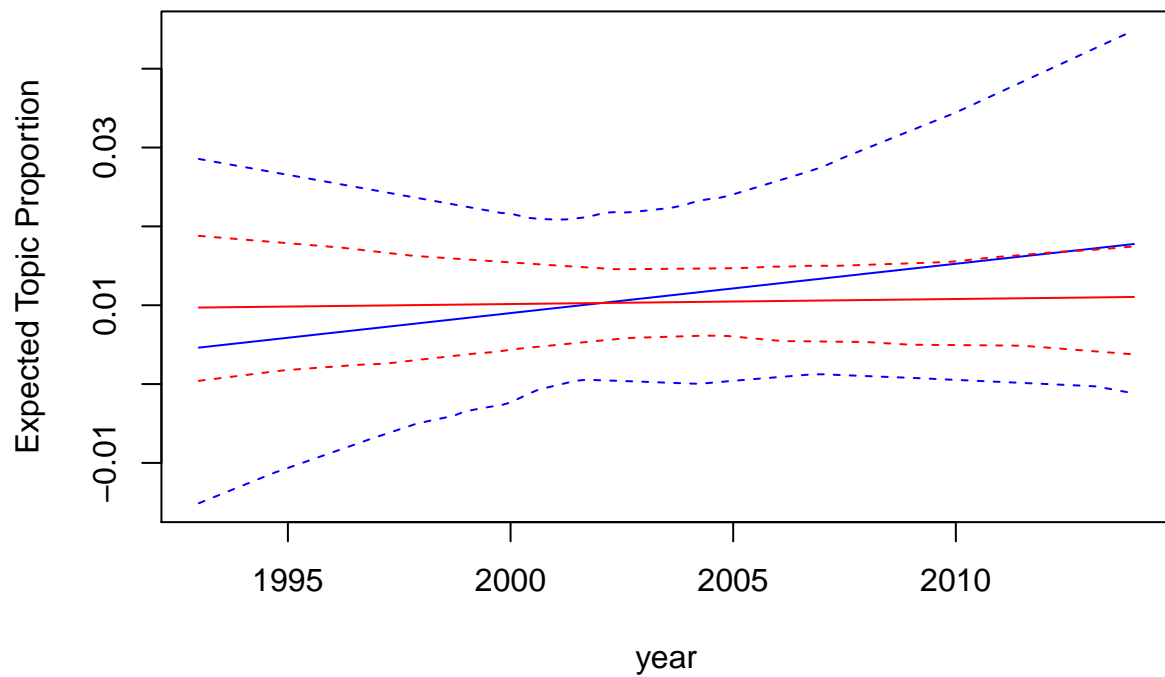
**topic 3**



**topic 1**

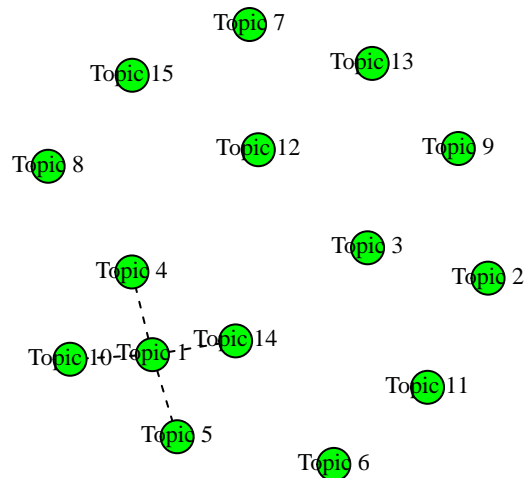


**topic 15**



On regarde GRAPHICAL NETWORK DISPLAY of how closely related topics are to one another, (i.e., how likely they are to appear in the same document):

```
mod.out.corr <- topicCorr(wolf_stm)
plot.topicCorr(mod.out.corr)
```



On détermine les topics les plus importants et on les interprète. Pour ce faire, on récupère les documents qui sont très corrélés avec le topic qui nous intéresse:

On transforme d'abord la base en un vecteur de caractères où chaque composante est un doc:

```

texts <- rep(NA,length(base))
for (i in 1:length(base)){
  temp <- as.character(lapply(base[i], as.character))
  texts[i] <- temp
}

```

On cherche et affiche les 5 documents les plus associés aux topics:

```

for (i in rank[1:15]){
  thoughts <- findThoughts(wolf_stm, texts = texts, n = 5, topics=i)$docs[[1]]
  # pour améliorer la lisibilité, rajoute des séparateurs entre les 5 documents
  thoughts <- gsub("c[:(:]","-----",thoughts)
  write(thoughts,paste("topic",i,".txt",sep=' '))
}

```