Untitled

Liza Wood

2/15/2021

```r
library(tidytext)
library(data.table)
library(dplyr)
```

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

## The following objects are masked from 'package:data.table':
##
##     between, first, last

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

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

```r
library(topicmodels)
library(textstem)
```

```
## Loading required package: koRpus.lang.en

## Loading required package: koRpus

## Loading required package: sylly

## For information on available language packages for 'koRpus', run
##
##    available.koRpus.lang()
##
## and see ?install.koRpus.lang()
```

```r
library(ldatuning)
library(ggplot2)
library(stringr)
library(lme4)
```

```
## Loading required package: Matrix
```

```r
library(equatiomatic)
```

```r
# MODEL-----
#combine users/gamma with the user data, then run a model

topic18 <- users %>% filter(topic == 18, document != "")
df.18 <- left_join(topic18, nodes, by = c("document" = "screen_name"))
```

```r
fit18 <- lmer(topic18$gamma ~ (1|community) + (1|id_) + (1|location_gen), data = df.18)
summary(fit18) # community explains most of the variance; fixef only .029
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: topic18$gamma ~ (1 | community) + (1 | id_) + (1 | location_gen)
##    Data: df.18
##
## REML criterion at convergence: -39165.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.1727 -0.3133 -0.1520 -0.0413 10.4411
##
## Random effects:
##  Groups       Name        Variance  Std.Dev.
##  id_          (Intercept) 1.597e-05 0.003996
##  community    (Intercept) 6.263e-04 0.025027
##  location_gen (Intercept) 9.600e-05 0.009798
##  Residual                 8.232e-03 0.090731
## Number of obs: 20012, groups:  id_, 15; community, 12; location_gen, 11
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 0.030088   0.008045    3.74
```

```r
coef(fit18) # community 3, 4 and 6; europe; even across types
```

```
## $id_
##                    (Intercept)
##                     0.02884117
## civil_society       0.03234369
## company             0.02643794
## company_employee    0.02946038
## expert              0.02866813
## farm                0.03004825
## farmer              0.02582564
## govt                0.03059457
## govt_employee       0.02929523
## media               0.02940739
## media_employee      0.03278963
## org                 0.03460366
## org_employee        0.03026691
## university          0.02711581
## university_employee 0.03562450
##
## $community
##      (Intercept)
## 1   0.027335929
## 2   0.025161691
## 3   0.054118999
## 4   0.056478754
## 5   0.010993583
## 6   0.089809801
## 7   0.022625735
```

```
## 8  0.010929309
## 9  0.010463854
## 10 0.009094929
## 11 0.027165253
## 12 0.016880488
##
## $location_gen
##               (Intercept)
##               0.03049783
## Africa        0.01436198
## Asia          0.02193869
## Europe        0.04459716
## Midwest       0.02411097
## North America 0.02862297
## Northeast     0.03167232
## Oceania       0.03762864
## South         0.02084028
## South America 0.03890645
## West          0.03779285
##
## attr(,"class")
## [1] "coef.mer"
```

```
ranef(fit18)
```

```
## $id_
##                        (Intercept)
##                       -1.247022e-03
## civil_society          2.255499e-03
## company               -3.650257e-03
## company_employee      -6.278175e-04
## expert                -1.420062e-03
## farm                  -3.994008e-05
## farmer                -4.262558e-03
## govt                   5.063784e-04
## govt_employee         -7.929610e-04
## media                 -6.808021e-04
## media_employee         2.701436e-03
## org                    4.515471e-03
## org_employee           1.787146e-04
## university            -2.972385e-03
## university_employee    5.536305e-03
##
## $community
##     (Intercept)
## 1  -0.002752265
## 2  -0.004926503
## 3   0.024030805
## 4   0.026390560
## 5  -0.019094610
## 6   0.059721608
## 7  -0.007462459
## 8  -0.019158885
## 9  -0.019624339
## 10 -0.020993265
```

```
## 11 -0.002922941
## 12 -0.013207706
##
## $location_gen
##              (Intercept)
##               0.0004096359
## Africa       -0.0157262176
## Asia         -0.0081495081
## Europe        0.0145089635
## Midwest      -0.0059772284
## North America -0.0014652216
## Northeast     0.0015841280
## Oceania       0.0075404458
## South        -0.0092479131
## South America 0.0088182555
## West          0.0077046601
##
## with conditional variances for "id_" "community" "location_gen"
```

```r
topic19 <- users %>% filter(topic == 19, document != "")
df.19 <- left_join(topic19, nodes, by = c("document" = "screen_name"))

fit19 <- lmer(topic19$gamma ~ (1|community) + (1|id_) + (1|location_gen), data = df.19)
summary(fit19) # community explaining much more of the variance here; a little more populat fixef .09
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: topic19$gamma ~ (1 | community) + (1 | id_) + (1 | location_gen)
##    Data: df.19
##
## REML criterion at convergence: -25626.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7485 -0.2343 -0.1171 -0.0325  7.5098
##
## Random effects:
##  Groups       Name        Variance  Std.Dev.
##  id_          (Intercept) 0.0014968 0.03869
##  community    (Intercept) 0.0184536 0.13584
##  location_gen (Intercept) 0.0001354 0.01164
##  Residual                 0.0161407 0.12705
## Number of obs: 20012, groups:  id_, 15; community, 12; location_gen, 11
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.07747    0.04086   1.896
```

```r
coef(fit19) # media much more likely to be talking about this kind of cc; community 4 and 6; northeast
```

```
## $id_
##                  (Intercept)
##                   0.06330501
## civil_society     0.06053098
## company           0.05827832
## company_employee  0.06938191
```

```
## expert                  0.06392645
## farm                    0.07344214
## farmer                  0.06494468
## govt                    0.06928779
## govt_employee           0.05403054
## media                   0.15341266
## media_employee          0.17371109
## org                     0.06257557
## org_employee            0.07446201
## university              0.06237540
## university_employee     0.05832693
##
## $community
##    (Intercept)
## 1   0.02717804
## 2   0.03375091
## 3   0.01623421
## 4   0.16561234
## 5   0.01572385
## 6   0.48881334
## 7   0.03730971
## 8   0.02886663
## 9   0.04079609
## 10  0.02713009
## 11  0.02282647
## 12  0.02535153
##
## $location_gen
##                (Intercept)
##                 0.07920326
## Africa          0.07500140
## Asia            0.08063882
## Europe          0.07580085
## Midwest         0.05992955
## North America   0.06951409
## Northeast       0.09586131
## Oceania         0.07185835
## South           0.07594830
## South America   0.07216116
## West            0.09621000
##
## attr(,"class")
## [1] "coef.mer"
```

```
ranef(fit19)
```

```
## $id_
##                    (Intercept)
##                    -0.014161092
## civil_society      -0.016935123
## company            -0.019187775
## company_employee   -0.008084190
## expert             -0.013539648
## farm               -0.004023958
## farmer             -0.012521418
```

```
## govt                 -0.008178313
## govt_employee         -0.023435557
## media                  0.075946561
## media_employee         0.096244994
## org                   -0.014890527
## org_employee          -0.003004084
## university            -0.015090699
## university_employee   -0.019139171
##
## $community
##     (Intercept)
## 1  -0.05028806
## 2  -0.04371519
## 3  -0.06123189
## 4   0.08814624
## 5  -0.06174225
## 6   0.41134724
## 7  -0.04015639
## 8  -0.04859947
## 9  -0.03667001
## 10 -0.05033601
## 11 -0.05463963
## 12 -0.05211457
##
## $location_gen
##               (Intercept)
##                0.001737158
## Africa        -0.002464695
## Asia           0.003172725
## Europe        -0.001665253
## Midwest       -0.017536552
## North America -0.007952011
## Northeast      0.018395211
## Oceania       -0.005607747
## South         -0.001517796
## South America -0.005304944
## West           0.018743904
##
## with conditional variances for "id_" "community" "location_gen"
```

```r
topic38 <- users %>% filter(topic == 38, document != "")
df.38 <- left_join(topic38, nodes, by = c("document" = "screen_name"))


fit38 <- lmer(topic38$gamma ~ (1|community) + (1|id_) + (1|location_gen), data = df.38)
summary(fit38) # int is .02; community still most explanatory, then location then id
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: topic38$gamma ~ (1 | community) + (1 | id_) + (1 | location_gen)
##    Data: df.38
##
## REML criterion at convergence: -42488
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3957 -0.1501 -0.0325  0.0082 11.3591
```

```
##
## Random effects:
##  Groups      Name         Variance  Std.Dev.
##  id_         (Intercept) 0.0000334 0.005779
##  community   (Intercept) 0.0020295 0.045050
##  location_gen (Intercept) 0.0002254 0.015014
##  Residual                 0.0069618 0.083438
## Number of obs: 20012, groups:  id_, 15; community, 12; location_gen, 11
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  0.02318    0.01395   1.661
```

```
coef(fit38) # community 11
```

```
## $id_
##                   (Intercept)
##                    0.01573600
## civil_society      0.02148478
## company            0.02180781
## company_employee   0.02134249
## expert             0.02882518
## farm               0.02325001
## farmer             0.02059511
## govt               0.03064558
## govt_employee      0.02154408
## media              0.01994809
## media_employee     0.01887489
## org                0.03120178
## org_employee       0.02308853
## university         0.02866606
## university_employee 0.02070304
##
## $community
##    (Intercept)
## 1  0.022307988
## 2  0.019514451
## 3  0.162577639
## 4  0.007736572
## 5  0.002012404
## 6  0.003826320
## 7  0.007132415
## 8  0.004703530
## 9  0.003972481
## 10 0.005255258
## 11 0.031929344
## 12 0.007202331
##
## $location_gen
##            (Intercept)
##             0.023099599
## Africa     -0.009708484
## Asia        0.052472727
## Europe      0.028969495
## Midwest     0.020278707
```

```
## North America  0.020635669
## Northeast      0.018429767
## Oceania        0.023121247
## South          0.020038914
## South America  0.034026006
## West           0.023626194
##
## attr(,"class")
## [1] "coef.mer"
```

```
ranef(fit38)
```

```
## $id_
##                        (Intercept)
##                       -7.444890e-03
## civil_society         -1.696116e-03
## company               -1.373085e-03
## company_employee      -1.838408e-03
## expert                 5.644282e-03
## farm                   6.911576e-05
## farmer                -2.585790e-03
## govt                   7.464687e-03
## govt_employee         -1.636814e-03
## media                 -3.232809e-03
## media_employee        -4.306001e-03
## org                    8.020889e-03
## org_employee          -9.236420e-05
## university             5.485161e-03
## university_employee   -2.477859e-03
##
## $community
##       (Intercept)
## 1   -0.0008729068
## 2   -0.0036664434
## 3    0.1393967445
## 4   -0.0154443224
## 5   -0.0211684909
## 6   -0.0193545742
## 7   -0.0160484794
## 8   -0.0184773646
## 9   -0.0192084133
## 10  -0.0179256361
## 11   0.0087484497
## 12  -0.0159785632
##
## $location_gen
##                   (Intercept)
##                  -8.129600e-05
## Africa           -3.288938e-02
## Asia              2.929183e-02
## Europe            5.788601e-03
## Midwest          -2.902187e-03
## North America    -2.545225e-03
## Northeast        -4.751128e-03
## Oceania          -5.964735e-05
```

```
## South          -3.141980e-03
## South America  1.084511e-02
## West            4.452993e-04
##
## with conditional variances for "id_" "community" "location_gen"
```

```
#extract_eq(fit18)
```

$$\text{climatetopic}_i \sim N\left(\alpha_{j[i],k[i],l[i]}, \sigma^2\right)$$
$$\alpha_j \sim N\left(\mu_{\alpha_j}, \sigma^2_{\alpha_j}\right), \text{ for role j} = 1, \ldots, \text{J}$$
$$\alpha_k \sim N\left(\mu_{\alpha_k}, \sigma^2_{\alpha_k}\right), \text{ for community k} = 1, \ldots, \text{K}$$
$$\alpha_l \sim N\left(\mu_{\alpha_l}, \sigma^2_{\alpha_l}\right), \text{ for location l} = 1, \ldots, \text{L}$$