

Cultural distances: Wikipedia data

Contents

Introduction	1
Load libraries	1
All domains	2
Load data	2
LMER models	4
Tests within domains	11
LMER models	11
Mantel tests	19
Data prep	19
Tests	21
References	22

Introduction

We compare cultural distances between societies with linguistic similarities between societies, controlling for shared history in two ways.

The first test uses mixed effects modelling. The pairing of the language family of each language (according to Glottolog) is used as a random effect. That means that the model can capture the likelihood that two languages from the Indo-European language family will be more similar to each other than two languages from different language families. The same is done with geographic area according to Autotyp.

The second test controls for history using distances from a phylogenetic tree. The tree comes from Bouckaert et al. (2012). Patristic distances between languages are used as a measure of historical distance between societies in a Mantel test. Note that the Mantel test assumes a strict distance metric, which is not necessarily the case with this data, but there are few other ways to deal with continuous pairwise distances.

Load libraries

```
library(ape)
library(ecodist)
library(lme4)
library(sjPlot)
library(ggplot2)
library(igraph)
library(lattice)
```

Parameters:

```
datasetName = "wikipedia"
lingDistancesFile = "../data/FAIR/wikipedia-by-language-pair-mean-rho.csv"
lingDistancesByDomainFile = "../results/EA_distances/All_Domains_with_ling.csv"
```

All domains

Load data

Read the cultural distances:

```
cult = read.csv("../results/EA_distances/CulturalDistances_Long.csv", stringsAsFactors = F)
names(cult) = c("l1", "l2", "cult.dist")
cultLangs = unique(c(cult$Var1, cult$Var2))
```

Add language family:

```
l = read.csv("../data/FAIR_langauges_glottol_xdid.csv", stringsAsFactors = F)
g = read.csv("../data/glottolog-languoid.csv/languoid.csv", stringsAsFactors = F)
l$family = g[match(l$glotto, g$id),]$family_pk
l$family = g[match(l$family, g$pk),]$name
```

Read the semantic distances

```
ling = read.csv(lingDistancesFile, stringsAsFactors = F)
```

Combine the linguistic and cultural distances

```
cult$l1.iso2 = l[match(cult$l1, l$Language2),]$iso2
cult$l2.iso2 = l[match(cult$l2, l$Language2),]$iso2

fairisos = unique(c(ling$l1, ling$l2))
cultisos = unique(c(cult$l1.iso2, cult$l2.iso2))

cult = cult[(cult$l1.iso2 %in% fairisos) & (cult$l2.iso2 %in% fairisos),]
ling = ling[(ling$l1 %in% cultisos) & (ling$l2 %in% cultisos),]

matches = sapply(1:nrow(ling), function(i){
  which(cult$l1.iso2==ling$l1[i] & cult$l2.iso2==ling$l2[i])
})

ling$cult.dist = cult[matches,]$cult.dist
ling$cult.dist.center = scale(ling$cult.dist)
cdc.s = attr(ling$cult.dist.center, "scaled:scale")
cdc.c = attr(ling$cult.dist.center, "scaled:center")
ling$cult.dist.center = as.numeric(ling$cult.dist.center)

ling$family1 = l[match(ling$l1, l$iso2),]$family
ling$family2 = l[match(ling$l2, l$iso2),]$family
ling$area1 = l[match(ling$l1, l$iso2),]$autotyp.area
ling$area2 = l[match(ling$l2, l$iso2),]$autotyp.area

fgroup = cbind(ling$family1, ling$family2)
fgroup = apply(fgroup, 1, sort)
ling$family.group = apply(fgroup, 2, paste, collapse=":")
agroup = cbind(ling$area1, ling$area2)
agroup = apply(agroup, 1, sort)
ling$area.group = apply(agroup, 2, paste, collapse=":")

ling$rho.center = scale(ling$rho)
```

Each observation is now associated with a language family pair:

```
head(ling[,c("l1", "l2", "rho", 'family.group')])
```

```
##      l1 l2      rho      family.group
## 1  ab ba 0.07191187  Abkhaz-Adyge:Turkic
## 2  ab be 0.10607846  Abkhaz-Adyge:Indo-European
## 3  ab bg 0.10662759  Abkhaz-Adyge:Indo-European
## 4  ab bn 0.03015938  Abkhaz-Adyge:Indo-European
## 8  ab cs 0.12521903  Abkhaz-Adyge:Indo-European
## 12 ab el 0.14504456  Abkhaz-Adyge:Indo-European
```

And the same is true for area:

```
tail(ling[,c("l1", "l2", "rho", 'area.group')])
```

```
##      l1 l2      rho      area.group
## 1365 te tr 0.2725814  Greater Mesopotamia:Indic
## 1366 te tt 0.1897841      Indic:Inner Asia
## 1367 te uk 0.2897670      Indic:Inner Asia
## 1368 tr tt 0.2489276  Greater Mesopotamia:Inner Asia
## 1369 tr uk 0.3311116  Greater Mesopotamia:Inner Asia
## 1370 tt uk 0.2492871      Inner Asia:Inner Asia
```

LMER models

Mixed effects model, predicting Linguistic similarity from cultural distances, with random intercept for family and area and random slope for cultural distance for family and area.

We compare a null model to a model with a fixed effect for cultural distance, with random intercepts for family and area, and random slopes for cultural distance by both.

```
m0 = lmer(
  rho.center ~ 1 +
    (1 + cult.dist.center | family.group) +
    (1 + cult.dist.center | area.group),
  data = ling
)
m1 = lmer(
  rho.center ~ 1 +
    cult.dist.center +
    (1 + cult.dist.center | family.group) +
    (1 + cult.dist.center | area.group),
  data = ling
)
anova(m0,m1)

## refitting model(s) with ML (instead of REML)

## Data: ling
## Models:
## m0: rho.center ~ 1 + (1 + cult.dist.center | family.group) + (1 +
## m0:      cult.dist.center | area.group)
## m1: rho.center ~ 1 + cult.dist.center + (1 + cult.dist.center | family.group) +
## m1:      (1 + cult.dist.center | area.group)
##      Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m0  8 1200.9 1234.9 -592.43  1184.9
## m1  9 1189.8 1228.1 -585.91  1171.8 13.05      1 0.0003032 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cultural distance is significantly correlated with Linguistic similarity. Here are the model estimates:

```
summary(m1)

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## rho.center ~ 1 + cult.dist.center + (1 + cult.dist.center | family.group) +
##      (1 + cult.dist.center | area.group)
##      Data: ling
##
## REML criterion at convergence: 1177.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6023 -0.6672  0.1129  0.6520  4.7446
##
## Random effects:
##      Groups      Name              Variance Std.Dev. Corr
## family.group (Intercept)      0.3930387 0.62693
##      cult.dist.center 0.0512354 0.22635 0.01
```

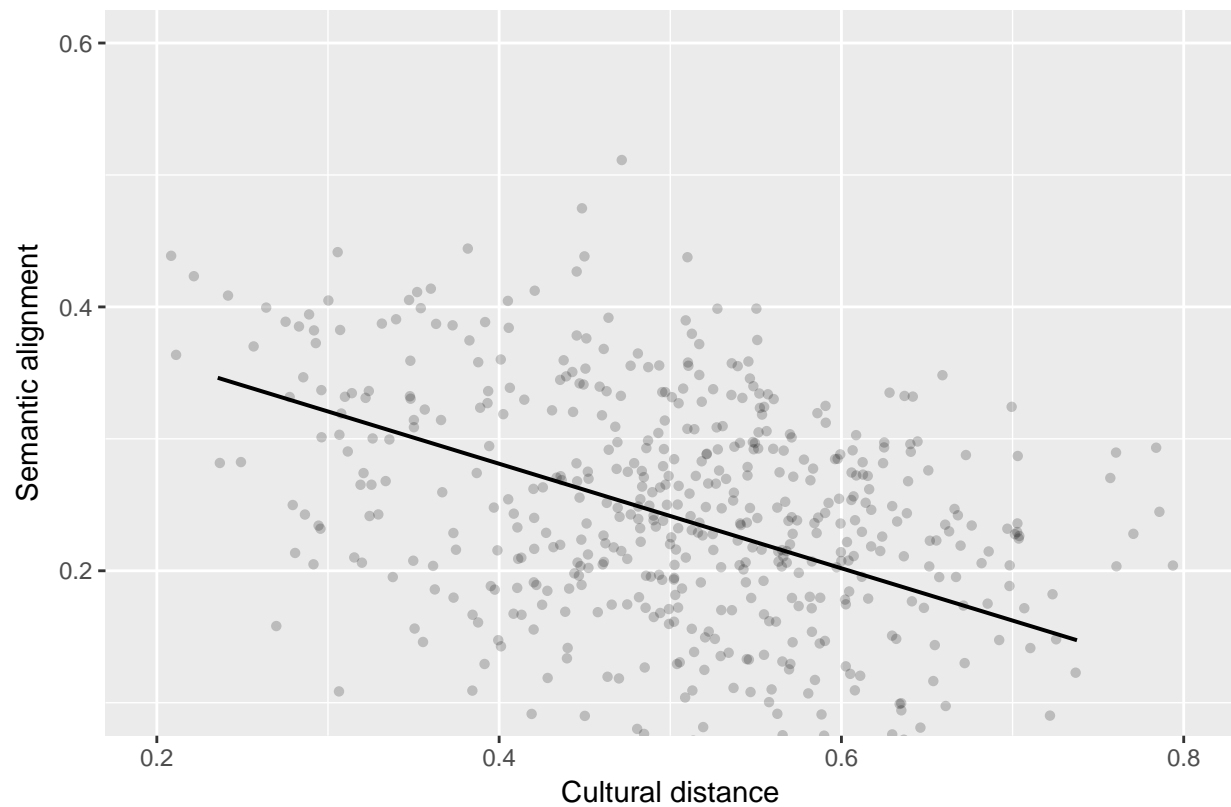
```
## area.group (Intercept) 0.0154283 0.12421
## cult.dist.center 0.0003434 0.01853 -1.00
## Residual 0.4771413 0.69075
## Number of obs: 522, groups: family.group, 32; area.group, 15
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) -0.54145 0.13723 -3.946
## cult.dist.center -0.40161 0.07766 -5.171
##
## Correlation of Fixed Effects:
## (Intr)
## clt.dst.cnt -0.171
```

Plot the estimates, rescaling the variables back to the original units:

```
gx = sjp.lmer(m1, 'pred', 'cult.dist.center',
             prnt.plot = F)
```

```
## Warning: package 'bindrcpp' was built under R version 3.3.2
```

```
gx$plot$data$y = gx$plot$data$y *
  attr(ling$rho.center, "scaled:scale") +
  attr(ling$rho.center, "scaled:center")
gx$plot$data$res.p.y = gx$plot$data$res.p.y *
  attr(ling$rho.center, "scaled:scale") +
  attr(ling$rho.center, "scaled:center")
gx$plot$data$x = gx$plot$data$x *
  cdc.s + cdc.c
gx = gx$plot + coord_cartesian(ylim=c(0.1,0.6),
                               xlim=c(0.2,0.8)) +
  xlab("Cultural distance") +
  ylab("Semantic alignment") +
  ggtitle("")
gx
```



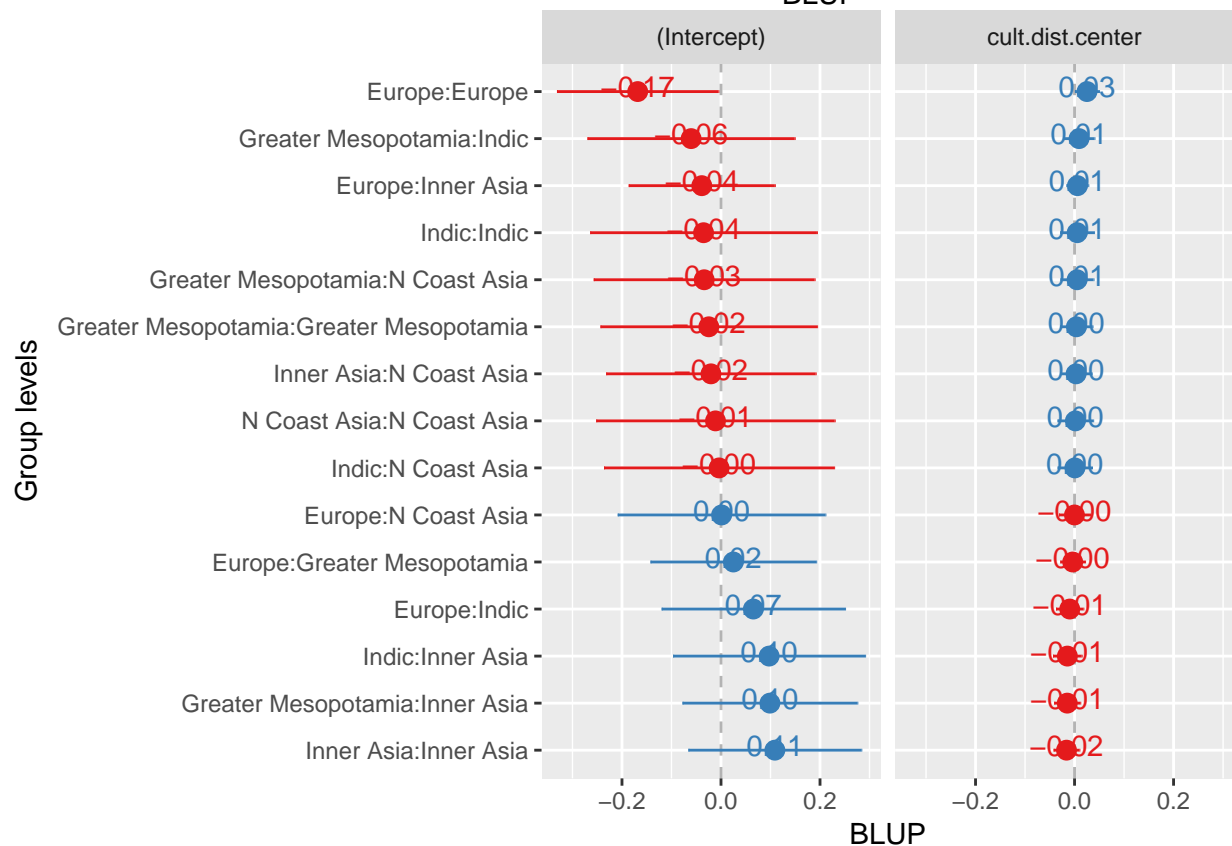
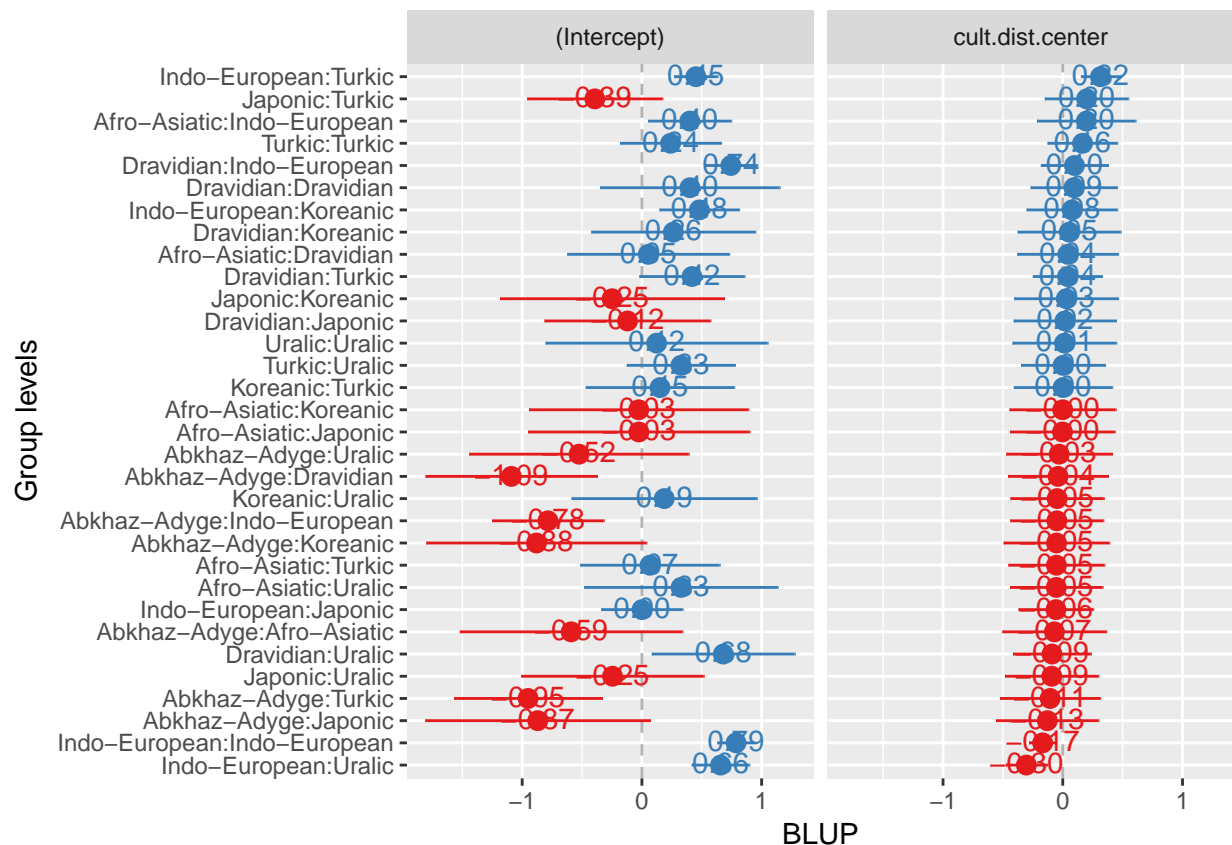
```
pdf(paste0("../results/stats/",datasetName,"/CulturalDistance_Rho_Graph.pdf"),
    height=2.5, width=2.5)
gx
dev.off()
```

```
## pdf
## 2
```

Plot the random effects:

```
sjp.lmer(m1,'re', sort.est = "cult.dist.center")
```

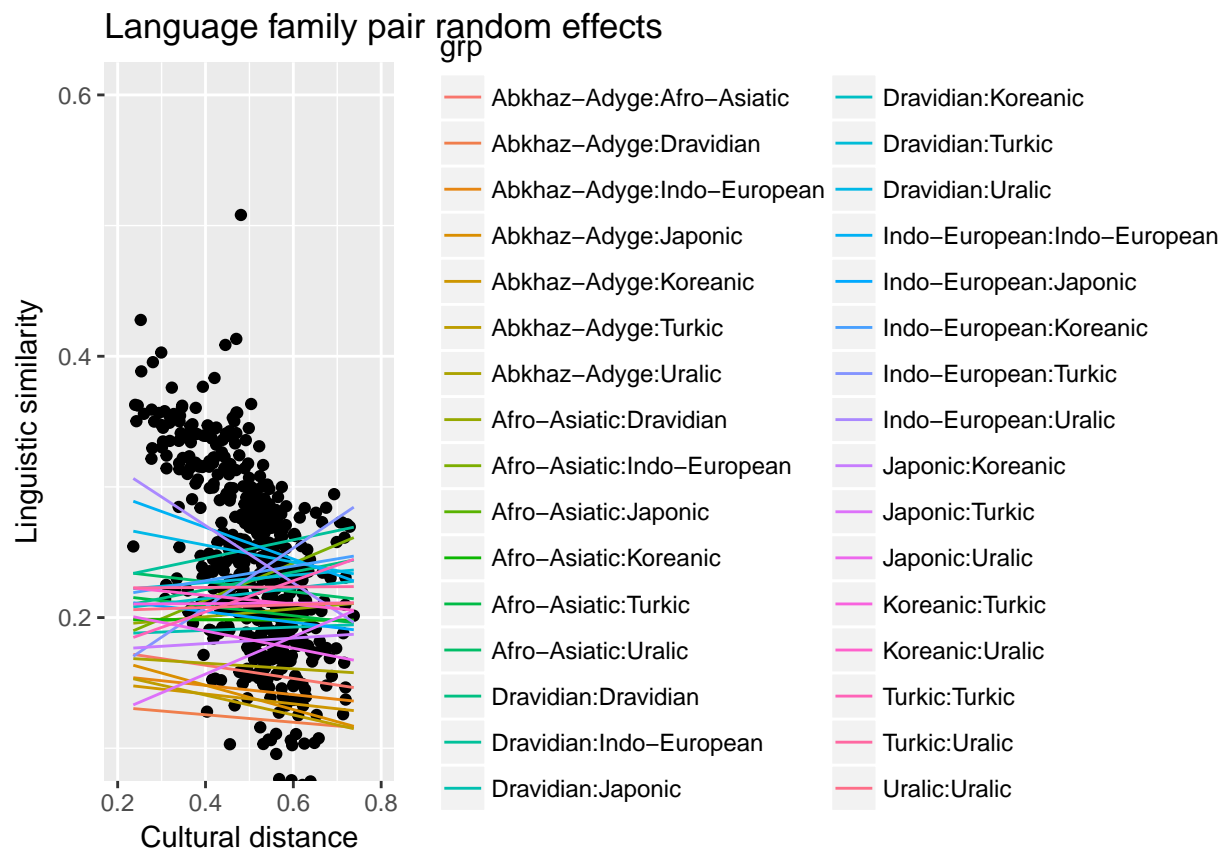
```
## Plotting random effects...
## Plotting random effects...
```



```

px = sjp.lmer(m1, 'rs.ri', prnt.plot = F)
dx = px$plot[[1]]$data
dx$x = dx$x * cdc.s + cdc.c
dx$y = dx$y *
  attr(ling$rho.center, "scaled:scale") +
  attr(ling$rho.center, "scaled:center")
ggplot(dx, aes(x, y)) +
  geom_point(data=ling,
             mapping=aes(x=as.numeric(cult.dist),
                        y=as.numeric(rho))) +
  geom_line(mapping = aes(colour=grp)) +
  xlab("Cultural distance")+
  ylab("Linguistic similarity") +
  ggtitle("Language family pair random effects") +
  coord_cartesian(ylim=c(0.1,0.6),
                 xlim=c(0.2,0.8))

```



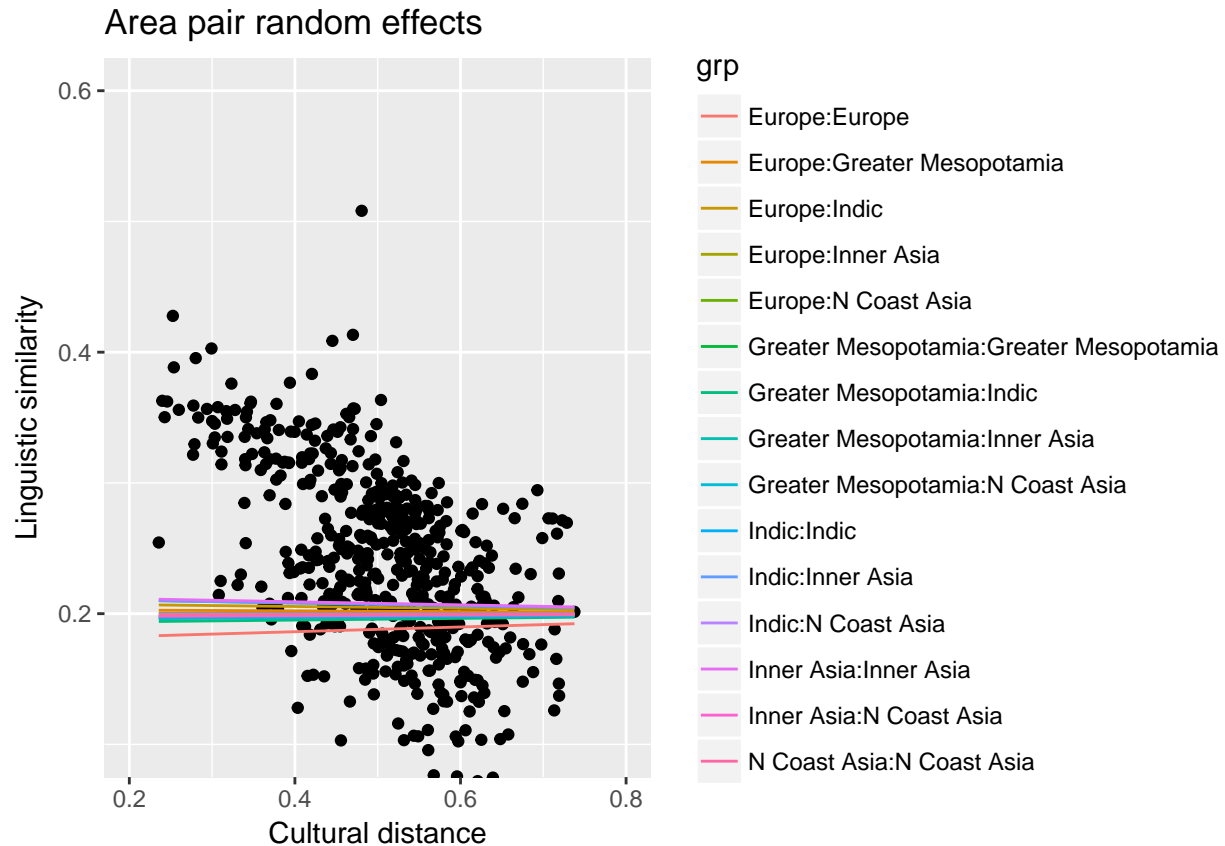
```

dx = px$plot[[2]]$data
dx$x = dx$x * cdc.s + cdc.c
dx$y = dx$y *
  attr(ling$rho.center, "scaled:scale") +
  attr(ling$rho.center, "scaled:center")
ggplot(dx, aes(x, y)) +
  geom_point(data=ling,
             mapping=aes(x=as.numeric(cult.dist),
                        y=as.numeric(rho))) +

```



```
geom_line(mapping = aes(colour=grp)) +
xlab("Cultural distance")+
ylab("Linguistic similarity") +
ggtitle("Area pair random effects") +
coord_cartesian(ylim=c(0.1,0.6),
                xlim=c(0.2,0.8))
```



Note that the random slopes for family are set to 0. We can check whether taking language family out makes a difference:

```
m0b = lmer(
  rho.center ~ 1 +
    (1 + cult.dist.center | area.group),
  data = ling
)
m1b = lmer(
  rho.center ~ 1 +
    cult.dist.center +
    (1 + cult.dist.center | area.group),
  data = ling
)
anova(m0b,m1b)
```

```
## refitting model(s) with ML (instead of REML)
## Data: ling
## Models:
## m0b: rho.center ~ 1 + (1 + cult.dist.center | area.group)
```

```
## m1b: rho.center ~ 1 + cult.dist.center + (1 + cult.dist.center | area.group)
##      Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m0b  5 1297.2 1318.4 -643.58  1287.2
## m1b  6 1276.7 1302.2 -632.32  1264.7 22.509      1 2.091e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(m1b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## rho.center ~ 1 + cult.dist.center + (1 + cult.dist.center | area.group)
##      Data: ling
##
## REML criterion at convergence: 1271.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7595 -0.6905  0.1326  0.7236  4.4159
##
## Random effects:
##      Groups      Name                Variance Std.Dev. Corr
## area.group (Intercept)      0.08603  0.2933
##                cult.dist.center 0.02655  0.1630  0.37
## Residual                    0.62455  0.7903
## Number of obs: 522, groups: area.group, 15
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   -0.12192   0.09138  -1.334
## cult.dist.center -0.58404   0.06716  -8.697
##
## Correlation of Fixed Effects:
##              (Intr)
## clt.dst.cnt 0.184
```

The model is numerically almost exactly the same, so the p-value from the model comparison is just lower.

Tests within domains

Load distances for specific domains and match up to language family and area:

```
ling.dom = read.csv(
  lingDistancesByDomainFile,
  stringsAsFactors = F)

ling.dom = ling.dom[!is.na(ling.dom$cult.dist),]

ling.dom$family1 = l[match(ling.dom$l1, l$iso2),]$family
ling.dom$family2 = l[match(ling.dom$l2, l$iso2),]$family
ling.dom$area1 = l[match(ling.dom$l1, l$iso2),]$autotyp.area
ling.dom$area2 = l[match(ling.dom$l2, l$iso2),]$autotyp.area

# Paste language family names together,
# but order shouldn't matter, so sort first
fgroup = cbind(ling.dom$family1, ling.dom$family2)
fgroup = apply(fgroup, 1, sort)
ling.dom$family.group = apply(fgroup, 2, paste, collapse=":")

agroup = cbind(ling.dom$area1, ling.dom$area2)
agroup = apply(agroup, 1, sort)
ling.dom$area.group = apply(agroup, 2, paste, collapse=": ")
```

Center the data:

```
ling.dom$cult.dist.center = scale(ling.dom$cult.dist)
ling.dom$rho.center = scale(ling.dom$rho)
```

LMER models

Test whether random slopes are warranted for family:

```
mD0 = lmer(
  rho.center ~ 1 +
    (1 | family.group) +
    (1 | area.group) +
    (1 | imputed_semantic_domain),
  data = ling.dom)
mD1 = lmer(
  rho.center ~ 1 +
    (1 | family.group) +
    (0 + cult.dist.center | family.group) +
    (1 | area.group) +
    (1 | imputed_semantic_domain),
  data = ling.dom)
mD2 = lmer(
  rho.center ~ 1 +
    (1 + cult.dist.center | family.group) +
    (1 | area.group) +
    (1 | imputed_semantic_domain),
  data = ling.dom)
```

```
anova(mD0,mD1,mD2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: ling.dom
## Models:
## mD0: rho.center ~ 1 + (1 | family.group) + (1 | area.group) + (1 |
## mD0: imputed_semantic_domain)
## mD1: rho.center ~ 1 + (1 | family.group) + (0 + cult.dist.center |
## mD1: family.group) + (1 | area.group) + (1 | imputed_semantic_domain)
## mD2: rho.center ~ 1 + (1 + cult.dist.center | family.group) + (1 |
## mD2: area.group) + (1 | imputed_semantic_domain)
##      Df   AIC   BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
## mD0  5 10311 10342 -5150.6    10301
## mD1  6 10293 10331 -5140.7    10281 19.9538      1 7.933e-06 ***
## mD2  7 10289 10333 -5137.7    10275  5.9424      1  0.01478 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Random slopes (and the correlation coefficient) for family improves the fit of the model.

Test the same for area:

```
mD3 = lmer(
  rho.center ~ 1 +
    (1 + cult.dist.center | family.group) +
    (1 | area.group) +
    (0 + cult.dist.center | area.group) +
    (1 | imputed_semantic_domain),
  data = ling.dom)
mD4 = lmer(
  rho.center ~ 1 +
    (1 + cult.dist.center | family.group) +
    (1 | area.group) +
    (1 | imputed_semantic_domain),
  data = ling.dom)
anova(mD2,mD3,mD4)
```

```
## refitting model(s) with ML (instead of REML)

## Data: ling.dom
## Models:
## mD2: rho.center ~ 1 + (1 + cult.dist.center | family.group) + (1 |
## mD2: area.group) + (1 | imputed_semantic_domain)
## mD4: rho.center ~ 1 + (1 + cult.dist.center | family.group) + (1 |
## mD4: area.group) + (1 | imputed_semantic_domain)
## mD3: rho.center ~ 1 + (1 + cult.dist.center | family.group) + (1 |
## mD3: area.group) + (0 + cult.dist.center | area.group) + (1 |
## mD3: imputed_semantic_domain)
##      Df   AIC   BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
## mD2  7 10289 10333 -5137.7    10275
## mD4  7 10289 10333 -5137.7    10275 0.0000      0      1.000
## mD3  8 10291 10341 -5137.6    10275 0.1353      1      0.713
```

Random slopes for area do not improve the fit of the model.

Test random slopes for domain:

```

mdom1 = lmer(
  rho.center ~ 1 +
    (1 + cult.dist.center | family.group) +
    (1 | area.group) +
    (1 | imputed_semantic_domain),
  data = ling.dom)
mdom2 = lmer(
  rho.center ~ 1 +
    (1 + cult.dist.center | family.group) +
    (1 | area.group) +
    (1 + cult.dist.center | imputed_semantic_domain),
  data = ling.dom)
anova(mdom1,mdom2)

## refitting model(s) with ML (instead of REML)

## Data: ling.dom
## Models:
## mdom1: rho.center ~ 1 + (1 + cult.dist.center | family.group) + (1 |
## mdom1:      area.group) + (1 | imputed_semantic_domain)
## mdom2: rho.center ~ 1 + (1 + cult.dist.center | family.group) + (1 |
## mdom2:      area.group) + (1 + cult.dist.center | imputed_semantic_domain)
##      Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mdom1  7 10289 10333 -5137.7    10275
## mdom2  9 10280 10336 -5130.9    10262 13.619      2  0.001103 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Random slope for domains significantly improves model.

Now we test the main effect of cultural distance:

```

mD5 = lmer(
  rho.center ~ 1 +
    (1 + cult.dist.center | family.group) +
    (1 | area.group) +
    (1 + cult.dist.center | imputed_semantic_domain),
  data = ling.dom)
mD6 = update(mD5, ~.+cult.dist.center)
anova(mD5,mD6)

## refitting model(s) with ML (instead of REML)

## Data: ling.dom
## Models:
## mD5: rho.center ~ 1 + (1 + cult.dist.center | family.group) + (1 |
## mD5:      area.group) + (1 + cult.dist.center | imputed_semantic_domain)
## mD6: rho.center ~ (1 + cult.dist.center | family.group) + (1 | area.group) +
## mD6:      (1 + cult.dist.center | imputed_semantic_domain) + cult.dist.center
##      Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mD5  9 10280 10336 -5130.9    10262
## mD6 10 10281 10343 -5130.6    10261 0.6717      1  0.4124

```

Summary of the final model, with random effects plot:

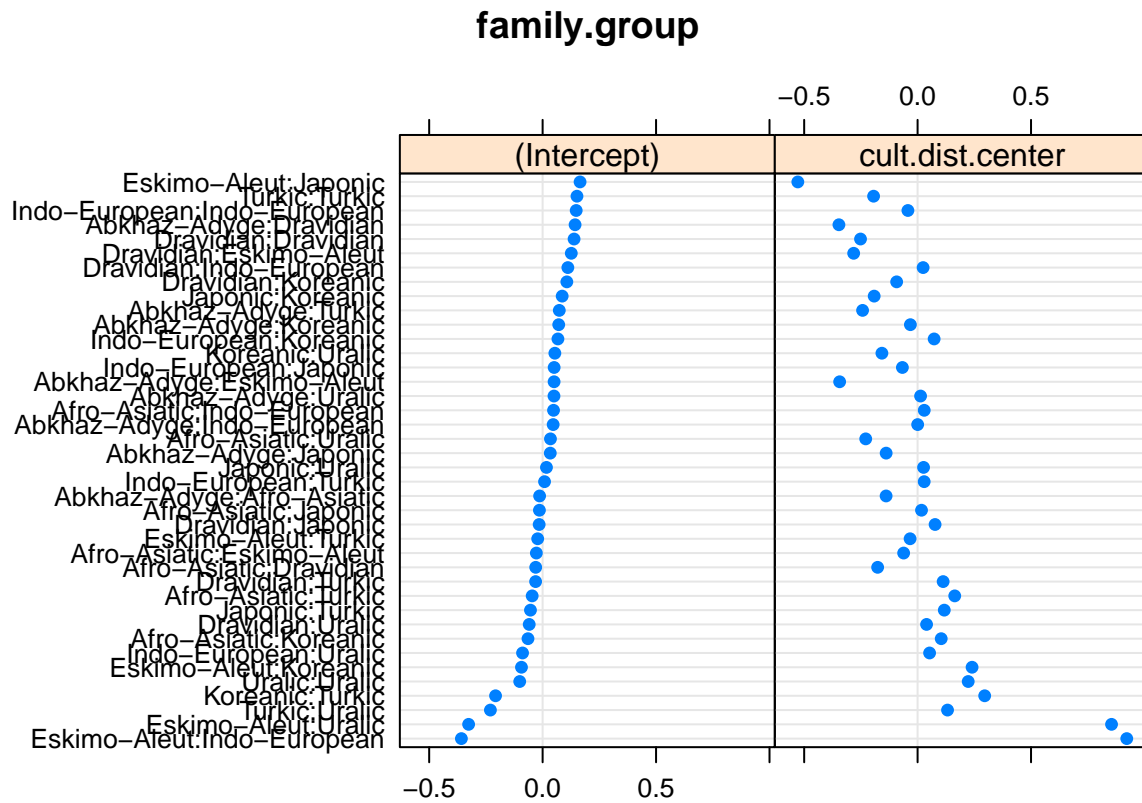
```
summary(mD6)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

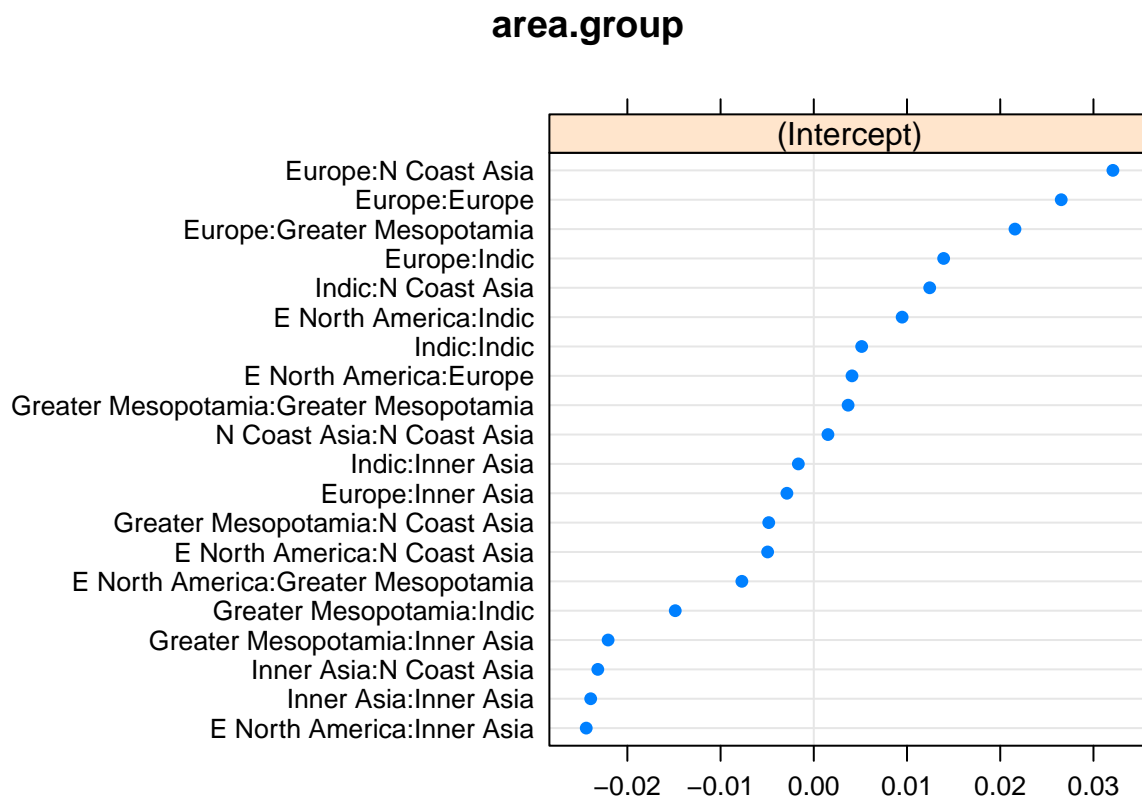
```

## Formula:
## rho.center ~ (1 + cult.dist.center | family.group) + (1 | area.group) +
##      (1 + cult.dist.center | imputed_semantic_domain) + cult.dist.center
## Data: ling.dom
##
## REML criterion at convergence: 10268
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.2667 -0.3978  0.0508  0.4604  3.6029
##
## Random effects:
## Groups              Name                Variance Std.Dev. Corr
## family.group        (Intercept)         0.031677 0.17798
##                     cult.dist.center 0.122253 0.34965 -0.76
## area.group          (Intercept)         0.001677 0.04095
## imputed_semantic_domain (Intercept)     0.027232 0.16502
##                     cult.dist.center 0.006224 0.07889 -0.33
## Residual                        0.930120 0.96443
## Number of obs: 3673, groups:
## family.group, 40; area.group, 20; imputed_semantic_domain, 7
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   -0.05603   0.07714  -0.726
## cult.dist.center -0.06561   0.07764  -0.845
##
## Correlation of Fixed Effects:
##              (Intr)
## clt.dst.cnt -0.438
dotplot(ranef(mD6))
## $family.group

```

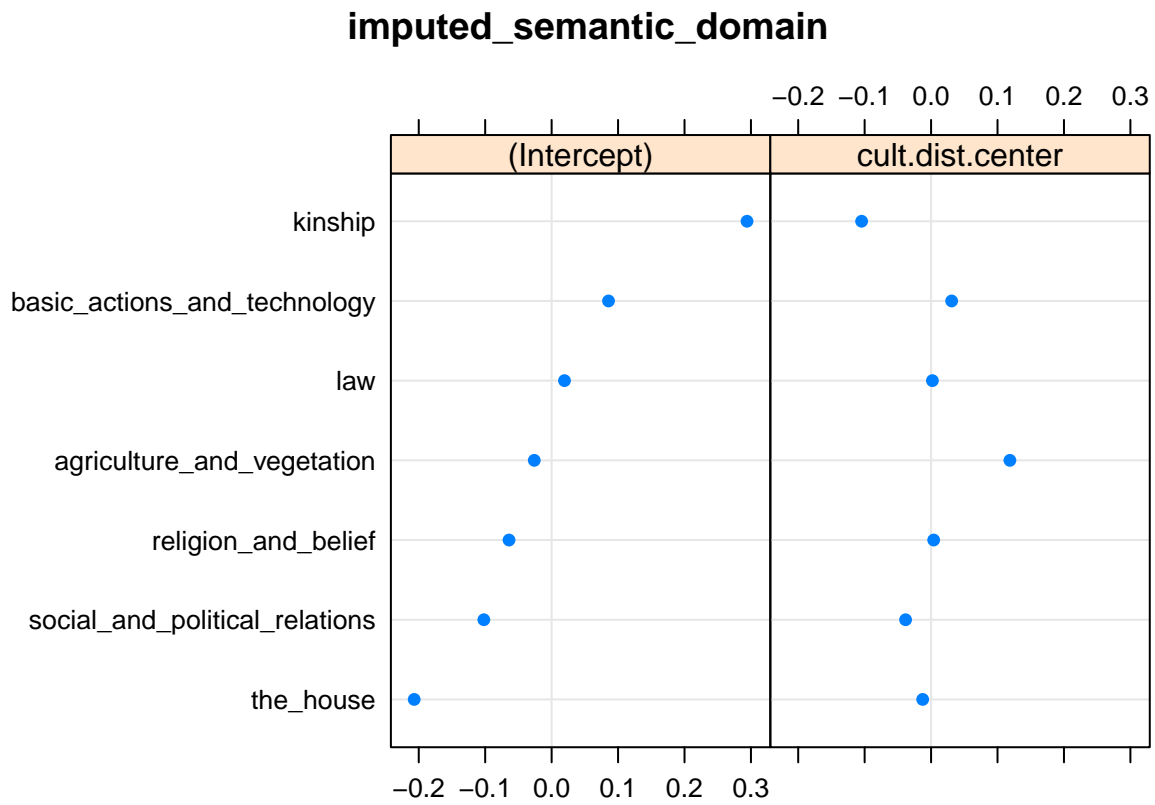


```
##
## $area.group
```



```
##
```

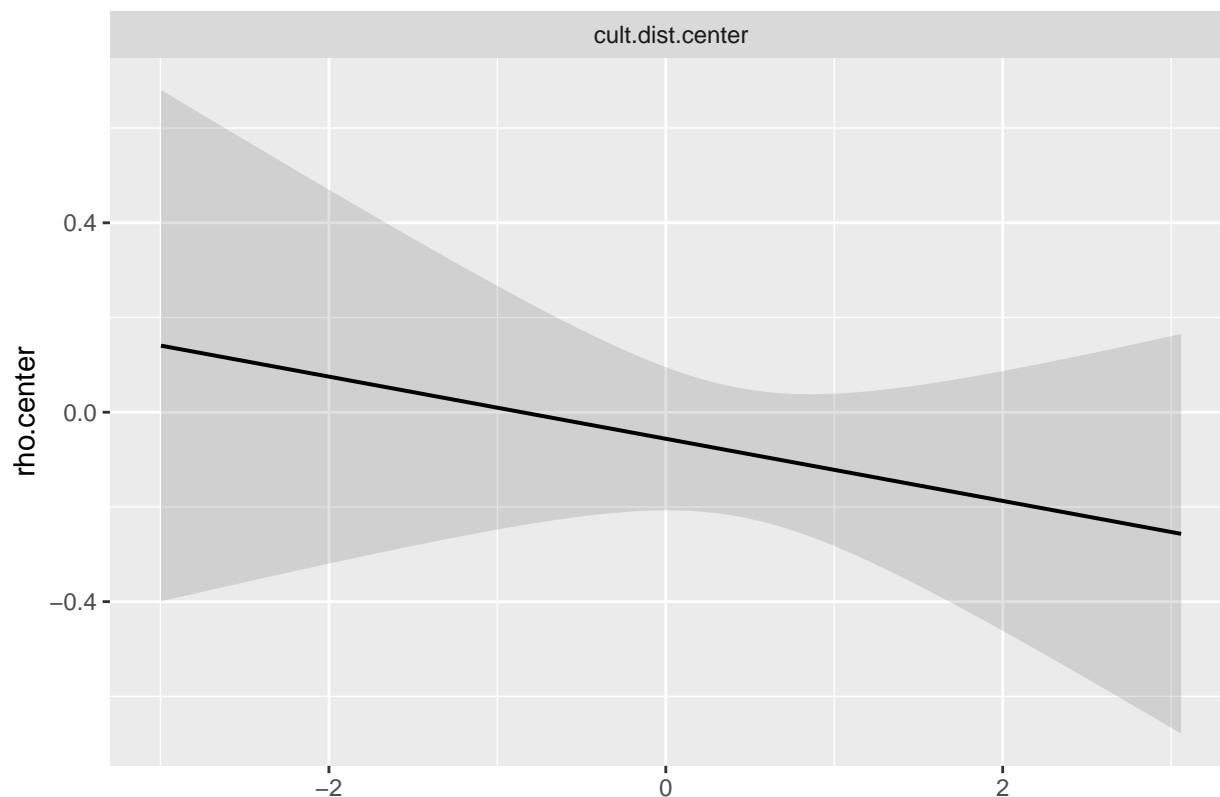
```
## $imputed_semantic_domain
```



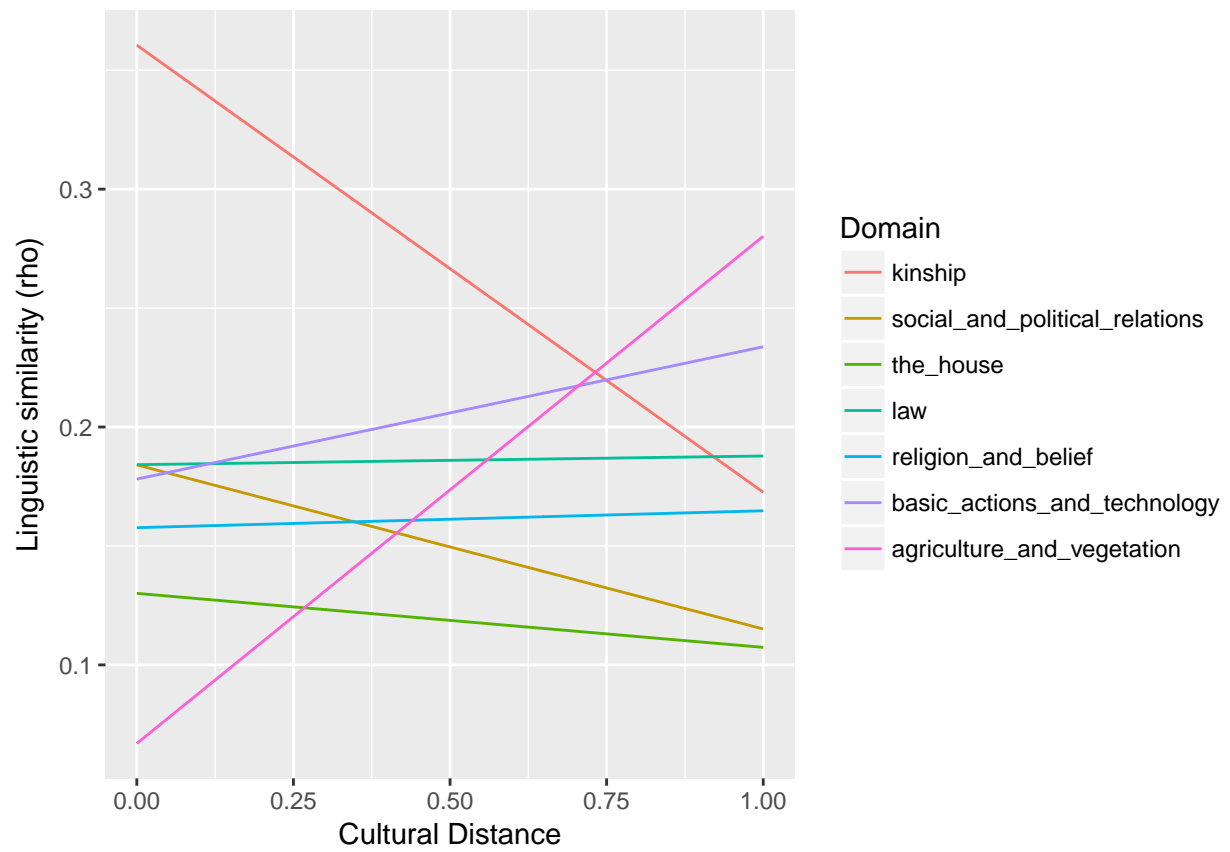
Plot the predicted relationships for each domain. The domains in the legend are sorted by the slope for cultural distance (greatest negative slope to greatest positive slope):

```
sjp.lmer(mD6,'eff', show.ci = T)
```

Marginal effects of model predictors



```
dom.order = ranef(mD6)$imputed_semantic_domain
dom.order = rownames(dom.order[order(dom.order$cult.dist.center),])
px = sjp.lmer(mD6,'rs.ri', show.ci = T, prnt.plot = F)
pdx = px$plot[[2]]$data
pdx$Domain = factor(pdx$grp, levels = dom.order)
pdx$x = pdx$x *
  attr(ling.dom$cult.dist.center,"scaled:scale") +
  attr(ling.dom$cult.dist.center,"scaled:center")
pdx$y = pdx$y *
  attr(ling.dom$rho.center,"scaled:scale") +
  attr(ling.dom$rho.center,"scaled:center")
ggplot(pdx,
  aes(x,y,colour=Domain)) +
  geom_line() +
  xlab("Cultural Distance") +
  ylab("Linguistic similarity (rho)")
```



“Agriculture and vegetation” seems to be working differently from “Kinship”.

Mantel tests

Read the historical distances for Indo-European, based on the phylogenetic distances.

Data prep

Load historical distances:

```
hist = read.csv("../data/trees/IndoEuropean_historical_distances.csv", stringsAsFactors = F)
hist = hist[!duplicated(hist[,1]),!duplicated(hist[,1])]
rownames(hist) = hist[,1]
hist = hist[,2:ncol(hist)]
hist.m = as.matrix(hist)
colnames(hist.m) = rownames(hist.m)
hist.m = hist.m/max(hist.m)
```

Read the cultural distances as a matrix:

```
cult.m = read.csv("../results/EA_distances/CulturalDistances.csv", stringsAsFactors = F)
rownames(cult.m) = cult.m[,1]
cult.m = cult.m[,2:ncol(cult.m)]
```

Convert the linguistic similarities to a matrix. This uses `igraph` to make an undirected graph from the long format with `rho` as the edge weights, then output a matrix of adjacencies.

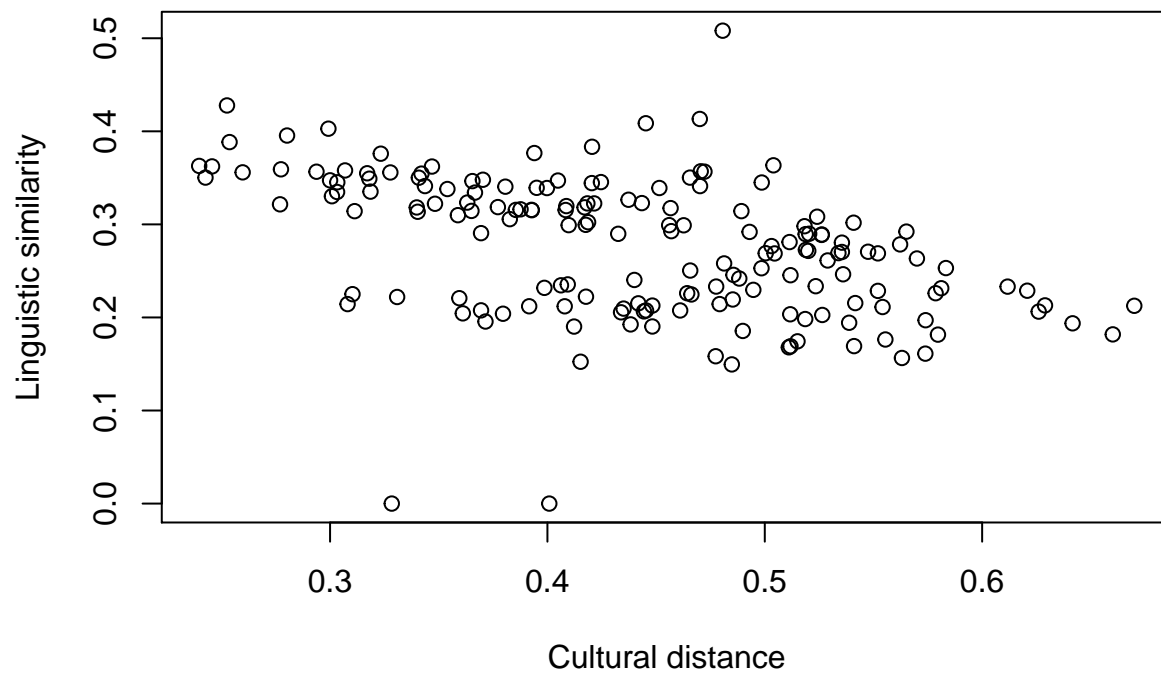
```
grph <- graph.data.frame(ling[,c("l1", 'l2', 'rho')], directed=FALSE)
# add value as a weight attribute
ling.m = get.adjacency(grph, attr="rho", sparse=FALSE)
rownames(ling.m) = 1[match(rownames(ling.m),l$iso2),]$Language2
colnames(ling.m) = 1[match(colnames(ling.m),l$iso2),]$Language2
```

Match the distance matrices

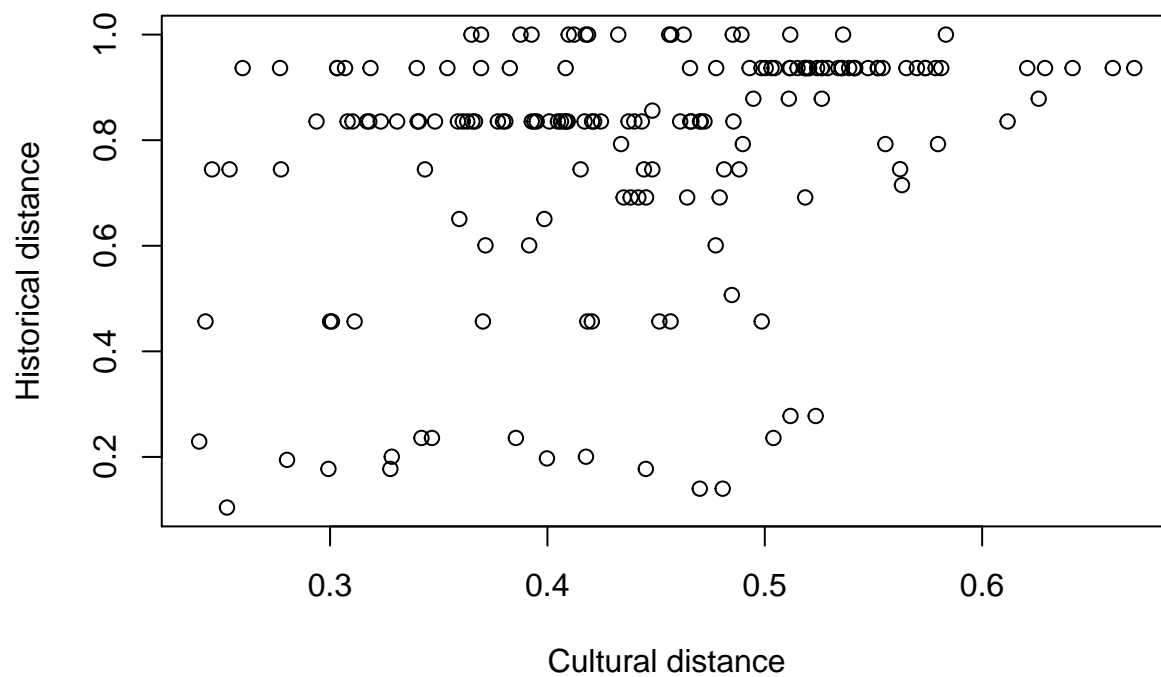
```
in.analysis = intersect(rownames(ling.m),rownames(cult.m))
in.analysis = intersect(in.analysis, rownames(hist.m))
cult.m2 = cult.m[in.analysis,in.analysis]
ling.m2 = ling.m[in.analysis,in.analysis]
hist.m2 = hist.m[in.analysis,in.analysis]
```

Note that there are only 19 languages with data on linguistic, cultural and historical distance.

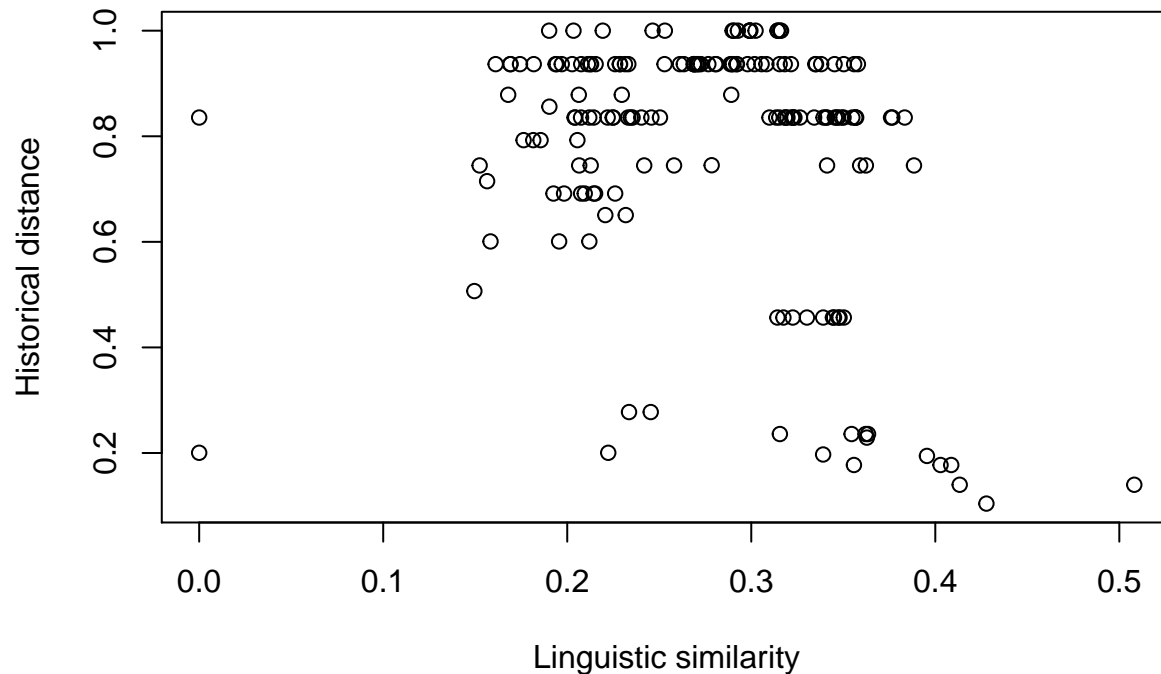
```
plot(as.dist(cult.m2),as.dist(ling.m2),
     xlab="Cultural distance",
     ylab="Linguistic similarity")
```



```
plot(as.dist(cult.m2),as.dist(hist.m2),
     xlab="Cultural distance",
     ylab="Historical distance")
```



```
plot(as.dist(ling.m2),as.dist(hist.m2),
     xlab="Linguistic similarity",
     ylab="Historical distance")
```



Tests

```
set.seed(1498)
```

Correlation between cultural and linguistic distances:

```
ecodist::mantel(as.dist(cult.m2) ~
  as.dist(ling.m2),
  nperm = 100000)
```

```
##      mantelr      pval1      pval2      pval3  llim.2.5% ulim.97.5%
## -0.4528007  0.9908500  0.0091600  0.0094600 -0.5569239 -0.3098959
```

Correlation between cultural and historical distances:

```
ecodist::mantel(as.dist(cult.m2) ~
  as.dist(hist.m2),
  nperm = 100000)
```

```
##      mantelr      pval1      pval2      pval3  llim.2.5% ulim.97.5%
##  0.3243830  0.0128500  0.9871600  0.0137800  0.2286808  0.4280939
```

Correlation between linguistic and historical distances:

```
ecodist::mantel(as.dist(ling.m2) ~
  as.dist(hist.m2),
  nperm = 100000)
```

```
##      mantelr      pval1      pval2      pval3  llim.2.5% ulim.97.5%
## -0.26095278  0.96603000  0.03398000  0.05103000 -0.46686747 -0.08320129
```

Run a mantel test comparing the Linguistic similaritys to the cultural distances, controlling for the historical distance between languages:

```
ecodist::mantel(as.dist(ling.m2)~
               as.dist(cult.m2) +
               as.dist(hist.m2),
               nperm = 100000)
```

```
##      mantelr      pval1      pval2      pval3  llim.2.5% ulim.97.5%
## -0.4031667  0.9845500  0.0154600  0.0176800 -0.5356183 -0.2683118
```

Run a mantel test comparing the linguistic similarities to the historical similarities, controlling for cultural distance:

```
ecodist::mantel(as.dist(ling.m2)~
               as.dist(hist.m2) +
               as.dist(cult.m2),
               nperm = 100000)
```

```
##      mantelr      pval1      pval2      pval3  llim.2.5% ulim.97.5%
## -0.1352527  0.8422400  0.1577700  0.3105800 -0.3828450  0.1082516
```

References

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, 337(6097), 957-960.