Proximity measures in the `proxy` package for R

David Meyer

September 2, 2012

1 Similarities

```
Aliases: Jaccard, binary, Reyssac, Roux
Type   : binary
Formula: a / (a + b + c)

Aliases: Kulczynski1
Type   : binary
Formula: a / (b + c)

Aliases: Kulczynski2
Type   : binary
Formula: [a / (a + b) + a / (a + c)] / 2

Aliases: Mountford
Type   : binary
Formula: 2a / (ab + ac + 2bc)

Aliases: Fager, McGowan
Type   : binary
Formula: a / sqrt((a + b)(a + c)) - 1 / 2 sqrt(a + c)

Aliases: Russel, Rao
Type   : binary
Formula: a / n

Aliases: simple matching, Sokal/Michener
Type   : binary
Formula: (a + d) / n

Aliases: Hamman
Type   : binary
Formula: ([a + d] - [b + c]) / n

Aliases: Faith
```

```
Type   : binary
Formula: (a + d/2) / n

Aliases: Tanimoto, Rogers
Type   : binary
Formula: (a + d) / (a + 2b + 2c + d)

Aliases: Dice, Czekanowski, Sorensen
Type   : binary
Formula: 2a / (2a + b + c)

Aliases: Phi
Type   : binary
Formula: (ad - bc) / sqrt[(a + b)(c + d)(a + c)(b + d)]

Aliases: Stiles
Type   : binary
Formula: log(n(|ad-bc| - 0.5n)^2 / [(a + b)(c + d)(a + c)(b + d)])

Aliases: Michael
Type   : binary
Formula: 4(ad - bc) / [(a + d)^2 + (b + c)^2]

Aliases: Mozley, Margalef
Type   : binary
Formula: an / (a + b)(a + c)

Aliases: Yule
Type   : binary
Formula: (ad - bc) / (ad + bc)

Aliases: Yule2
Type   : binary
Formula: (sqrt(ad) - sqrt(bc)) / (sqrt(ad) + sqrt(bc))

Aliases: Ochiai
Type   : binary
Formula: a / sqrt[(a + b)(a + c)]

Aliases: Simpson
Type   : binary
Formula: a / min{(a + b), (a + c)}

Aliases: Braun-Blanquet
Type   : binary
Formula: a / max{(a + b), (a + c)}
```

```
Aliases: cosine, angular
Type   : metric
Formula: xy / sqrt(xx * yy)

Aliases: eJaccard, extended_Jaccard
Type   : metric
Formula: xy / (xx + yy - xy)

Aliases: fJaccard, fuzzy_Jaccard
Type   : metric
Formula: sum_i (min{x_i, y_i} / max{x_i, y_i})

Aliases: correlation
Type   : metric
Formula: xy / sqrt(xx * yy) for centered x,y

Aliases: Chi-squared
Type   : nominal
Formula: sum_ij (o_i - e_i)^2 / e_i

Aliases: Phi-squared
Type   : nominal
Formula: [sum_ij (o_i - e_i)^2 / e_i] / n

Aliases: Tschuprow
Type   : nominal
Formula: sqrt{[sum_ij (o_i - e_i)^2 / e_i] / n / sqrt((p - 1)(q - 1))}

Aliases: Cramer
Type   : nominal
Formula: sqrt{[Chi / n)] / min[(p - 1), (q - 1)]}

Aliases: Pearson, contingency
Type   : nominal
Formula: sqrt{Chi / (n + Chi)}

Aliases: Gower
Type   : NA
Formula: Sum_k (s_ijk * w_k) / Sum_k (d_ijk * w_k)
```

2 Dissimilarities

```
Aliases: Euclidean, L2
Type   : metric
```

```
Formula: sqrt(sum_i (x_i - y_i)^2))

Aliases: Mahalanobis
Type   : metric
Formula: sqrt((x - y) Sigma^(-1) (x - y))

Aliases: Bhjattacharyya
Type   : metric
Formula: sqrt(sum_i (sqrt(x_i) - sqrt(y_i))^2))

Aliases: Manhattan, City-Block, L1, taxi
Type   : metric
Formula: sum_i |x_i - y_i|

Aliases: supremum, max, Tschebyscheff, Chebyshev
Type   : metric
Formula: max_i |x_i - y_i|

Aliases: Minkowski, Lp
Type   : metric
Formula: (sum_i (x_i - y_i)^p)^(1/p)

Aliases: Canberra
Type   : metric
Formula: sum_i |x_i - y_i| / |x_i + y_i|

Aliases: Wave, Hedges
Type   : metric
Formula: sum_i (1 - min(x_i, y_i) / max(x_i, y_i))

Aliases: divergence
Type   : metric
Formula: sum_i (x_i - y_i)^2 / (x_i + y_i)^2

Aliases: Kullback, Leibler
Type   : metric
Formula: sum_i [x_i * log((x_i / sum_j x_j) / (y_i / sum_j y_j)) / sum_j x_j)]

Aliases: Bray, Curtis
Type   : metric
Formula: sum_i |x_i - y_i| / sum_i (x_i + y_i)

Aliases: Soergel
Type   : metric
Formula: sum_i |x_i - y_i| / sum_i max{x_i, y_i}
```

```
Aliases: Levenshtein
Type   : other
Formula: Number of insertions, edits, and deletions between to strings

Aliases: Podani, discordance
Type   : metric
Formula: 1 - 2 * (a - b + c - d) / (n * (n - 1))

Aliases: Chord
Type   : metric
Formula: sqrt(2 * (1 - xy / sqrt(xx * yy)))

Aliases: Geodesic
Type   : metric
Formula: arccos(xy / sqrt(xx * yy))

Aliases: Whittaker
Type   : metric
Formula: sum_i |x_i / sum_i x - y_i / sum_i y| / 2

Aliases: Hellinger
Type   : metric
Formula: sqrt(sum_i (sqrt(x_i / sum_i x) - sqrt(y_i / sum_i y)) ^ 2)
```