HW2_P1

Naveen Nagarajan

3/22/2021

R Markdown

Problem

FE582 – Assignment 2 Spring 2021

The data provided in the files contains several quantitative and categorical variables associate with each ticker. Please select a subset of 100 tickers from each file and use data for a specific year (ex: 2013). Use a small number of quantitative variables (10 or 12) out of ~76 columns available (example: After Tax ROE, Cash Ratio, Current Ratio, Operating Margin, Pre-Tax Margin, Pre-Tax ROE, Profit Margin, Quick Ratio, Total Assets, Total Liabilities, Earnings Per Share, etc. . .). The categorical variables available are GICS Sector, GICS Sub Industry, and possibly HQ Address (although this is sparse data for the 100 tickers subset selected). Next, you have to apply several distance and similarity functions to find the extreme values for distance and similarities between the subset of tickers that you chose. For each of the following cases, please define the function that allows you to calculate the quantity required, calculate the values for all ticker pairs, and rank the pairs by calculated value of distance or similarity, and report the top and bottom 10 values for each case: a) LpNorm for p = 1 b) LpNorm for p = 2 c) LpNorm for p = 3 d) LpNorm for p = 10 e) Minkovski distance (assign different weights for the feature components in the Lp-norm based on your assessment on the importance of the features) f) Match-Based Similarity Computation (use a small number of equi-depth buckets, ex: 3) g) Mahalanobis distance h) Similarity: overlap measure i) Similarity: inverse frequency j) Similarity: Goodall k) Overall similarity between tickers by using mixed type data (choose a lamda value for calculation) l) Overall normalized similarity between tickers by using mixed type data (choose a lamda value for calculation)

Summary of Quantitative dataset chosen

a) Lp-norm for p-1

```
lpNormDistance1 <- lpNormDistance(filteredSecQuantMatrix, p = 1, upper = TRUE)
displayRanks(filteredSecuritiesIdxQuant, rankAMatrix(lpNormDistance1), lpNormDistance1,
    showDs = FALSE, sameRankShow = 3)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  0 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : AKAM CTXS  distance:  808752000 index:  ( 13 , 75 )"
## [1] "Rank : 3  pair : AME DPS  distance:  1299270000 index:  ( 19 , 91 )"
## [1] "Rank : 4  pair : AME CHD  distance:  1520048000 index:  ( 19 , 57 )"
## [1] "Rank : 5  pair : CTXS DLR  distance:  1597747000 index:  ( 75 , 88 )"
## [1] "Rank : 6  pair : CMG DPS  distance:  1663066000 index:  ( 66 , 91 )"
## [1] "Rank : 7  pair : AME CMG  distance:  1781090000 index:  ( 19 , 66 )"
```

```
## [1] "Rank : 8  pair : ALLE CHD  distance:  1786200000 index:  ( 17 , 57 )"
## [1] "Rank : 9  pair : AMG DNB  distance:  1920000000 index:  ( 20 , 89 )"
## [1] "Rank : 10  pair : ALLE DNB  distance:  1941500000 index:  ( 17 , 89 )"
## [1] "######Bottom 10########"
## [1] "Rank : 4941  pair : BAC DAL  distance:  926029000000 index:  ( 37 , 80 )"
## [1] "Rank : 4942  pair : AAL BAC  distance:  926375000000 index:  ( 1 , 37 )"
## [1] "Rank : 4943  pair : AMZN BAC  distance:  927507000000 index:  ( 24 , 37 )"
## [1] "Rank : 4944  pair : APA BAC  distance:  928163000000 index:  ( 28 , 37 )"
## [1] "Rank : 4945  pair : ADM BAC  distance:  946430000000 index:  ( 4 , 37 )"
## [1] "Rank : 4946  pair : BAC DUK  distance:  946624000000 index:  ( 37 , 92 )"
## [1] "Rank : 4947  pair : ANTM BAC  distance:  955602600000 index:  ( 26 , 37 )"
## [1] "Rank : 4948  pair : AIG BAC  distance:  1000945000000 index:  ( 9 , 37 )"
## [1] "Rank : 4949  pair : BAC CVS  distance:  1022143000000 index:  ( 37 , 76 )"
## [1] "Rank : 4950  pair : BAC CVX  distance:  1082265000000 index:  ( 37 , 77 )"
```

b) Lp-norm for p-2

```
lpNormDistance2 <- lpNormDistance(filteredSecQuantMatrix, p = 2, upper = TRUE)
displayRanks(filteredSecuritiesIdxQuant, rankAMatrix(lpNormDistance2), lpNormDistance2,
    showDs = FALSE, sameRankShow = 3)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  0 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : AKAM CTXS  distance:  369958819.551582 index:  ( 13 , 75 )"
## [1] "Rank : 3  pair : AME DPS  distance:  523267577.304384 index:  ( 19 , 91 )"
## [1] "Rank : 4  pair : CTXS DLR  distance:  596597661.474632 index:  ( 75 , 88 )"
## [1] "Rank : 5  pair : CMG DPS  distance:  755604964.202856 index:  ( 66 , 91 )"
## [1] "Rank : 6  pair : ALLE CHD  distance:  809137268.947612 index:  ( 17 , 57 )"
## [1] "Rank : 7  pair : AME CHD  distance:  813609766.569945 index:  ( 19 , 57 )"
## [1] "Rank : 8  pair : AME CMG  distance:  815634029.05102 index:  ( 19 , 66 )"
## [1] "Rank : 9  pair : ALK AME  distance:  848796222.575242 index:  ( 15 , 19 )"
## [1] "Rank : 10  pair : AIV CTXS  distance:  871322921.561805 index:  ( 10 , 75 )"
## [1] "######Bottom 10########"
## [1] "Rank : 4941  pair : BAC EFX  distance:  619081939608.482 index:  ( 37 , 98 )"
## [1] "Rank : 4942  pair : BAC DLR  distance:  619105000682.513 index:  ( 37 , 88 )"
## [1] "Rank : 4943  pair : AEE BAC  distance:  619125968482.99 index:  ( 6 , 37 )"
## [1] "Rank : 4944  pair : BAC CVX  distance:  619130080187.516 index:  ( 37 , 77 )"
## [1] "Rank : 4945  pair : AIV BAC  distance:  619139081257.625 index:  ( 10 , 37 )"
## [1] "Rank : 4946  pair : BAC CXO  distance:  619141024577.74 index:  ( 37 , 78 )"
## [1] "Rank : 4947  pair : AWK BAC  distance:  619167943911.02 index:  ( 34 , 37 )"
## [1] "Rank : 4948  pair : BAC COG  distance:  619168525956.46 index:  ( 37 , 72 )"
## [1] "Rank : 4949  pair : AIG BAC  distance:  622486411670.97 index:  ( 9 , 37 )"
## [1] "Rank : 4950  pair : BAC CVS  distance:  624851552398.648 index:  ( 37 , 76 )"
```

c) Lp-norm for p-3

```
lpNormDistance3 <- lpNormDistance(filteredSecQuantMatrix, p = 3, upper = TRUE)
displayRanks(filteredSecuritiesIdxQuant, rankAMatrix(lpNormDistance3), lpNormDistance3,
    showDs = FALSE, sameRankShow = 3)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  0 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : AKAM CTXS  distance:  309404435.494929 index:  ( 13 , 75 )"
## [1] "Rank : 3  pair : AME DPS  distance:  417627249.313972 index:  ( 19 , 91 )"
## [1] "Rank : 4  pair : CTXS DLR  distance:  454614071.535045 index:  ( 75 , 88 )"
## [1] "Rank : 5  pair : CMG DPS  distance:  615432777.296685 index:  ( 66 , 91 )"
## [1] "Rank : 6  pair : ALLE CHD  distance:  667122763.443879 index:  ( 17 , 57 )"
## [1] "Rank : 7  pair : AME CMG  distance:  667801383.173314 index:  ( 19 , 66 )"
## [1] "Rank : 8  pair : ALK AME  distance:  683729852.356055 index:  ( 15 , 19 )"
## [1] "Rank : 9  pair : AKAM DLR  distance:  706086953.933701 index:  ( 13 , 88 )"
## [1] "Rank : 10  pair : AIV CTXS  distance:  711783156.991522 index:  ( 10 , 75 )"
## [1] "######Bottom 10########"
## [1] "Rank : 4941  pair : AN BAC  distance:  588262500837.65 index:  ( 25 , 37 )"
## [1] "Rank : 4942  pair : ALK BAC  distance:  588314820536.736 index:  ( 15 , 37 )"
## [1] "Rank : 4943  pair : BAC EFX  distance:  588321420164.488 index:  ( 37 , 98 )"
## [1] "Rank : 4944  pair : AIV BAC  distance:  588338342383.713 index:  ( 10 , 37 )"
## [1] "Rank : 4945  pair : BAC CHTR  distance:  588373934709.902 index:  ( 37 , 60 )"
## [1] "Rank : 4946  pair : BAC DLR  distance:  588380102488.914 index:  ( 37 , 88 )"
## [1] "Rank : 4947  pair : AWK BAC  distance:  588404439531.361 index:  ( 34 , 37 )"
## [1] "Rank : 4948  pair : AEE BAC  distance:  588419602502.545 index:  ( 6 , 37 )"
## [1] "Rank : 4949  pair : BAC COG  distance:  588421962769.079 index:  ( 37 , 72 )"
## [1] "Rank : 4950  pair : BAC CXO  distance:  588430681134.642 index:  ( 37 , 78 )"
```

d) Lp-norm for p-10

```
lpNormDistance10 <- lpNormDistance(filteredSecQuantMatrix, p = 10, upper = TRUE)
displayRanks(filteredSecuritiesIdxQuant, rankAMatrix(lpNormDistance10), lpNormDistance10,
    showDs = FALSE, sameRankShow = 3)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  0 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : AKAM CTXS  distance:  268506417.231126 index:  ( 13 , 75 )"
## [1] "Rank : 3  pair : CTXS DLR  distance:  350657303.648576 index:  ( 75 , 88 )"
## [1] "Rank : 4  pair : AME DPS  distance:  353277923.840989 index:  ( 19 , 91 )"
## [1] "Rank : 5  pair : CMG DPS  distance:  504435368.402998 index:  ( 66 , 91 )"
## [1] "Rank : 6  pair : AME CMG  distance:  551685681.491926 index:  ( 19 , 66 )"
## [1] "Rank : 7  pair : ALLE CHD  distance:  581105316.147757 index:  ( 17 , 57 )"
## [1] "Rank : 8  pair : AKAM DLR  distance:  592118707.799895 index:  ( 13 , 88 )"
## [1] "Rank : 9  pair : AMT AWK  distance:  594721252.927908 index:  ( 23 , 34 )"
## [1] "Rank : 10  pair : AIV CTXS  distance:  603834946.599024 index:  ( 10 , 75 )"
## [1] "######Bottom 10########"
## [1] "Rank : 4941  pair : BAC CTL  distance:  581835135807.123 index:  ( 37 , 73 )"
## [1] "Rank : 4942  pair : AIV BAC  distance:  581842732947.109 index:  ( 10 , 37 )"
## [1] "Rank : 4943  pair : ALK BAC  distance:  581856146712.942 index:  ( 15 , 37 )"
## [1] "Rank : 4944  pair : AN BAC  distance:  581887741712.188 index:  ( 25 , 37 )"
## [1] "Rank : 4945  pair : BAC DLR  distance:  581910271318.694 index:  ( 37 , 88 )"
## [1] "Rank : 4946  pair : AWK BAC  distance:  581926147804.527 index:  ( 34 , 37 )"
## [1] "Rank : 4947  pair : BAC COG  distance:  581942193806.618 index:  ( 37 , 72 )"
## [1] "Rank : 4948  pair : AEE BAC  distance:  581958146767.1 index:  ( 6 , 37 )"
## [1] "Rank : 4949  pair : BAC CHTR  distance:  581960143982.684 index:  ( 37 , 60 )"
## [1] "Rank : 4950  pair : BAC CXO  distance:  581963126141.479 index:  ( 37 , 78 )"
```

e) Minkovski distance (assign different weights for the feature components in the Lp-norm based on your assessment on the importance of the features)

```
minkowski <- lpNormDistance(filteredSecQuantMatrix, p = 10, upper = TRUE, dictRelevance = dict(list(`1`
    `2` = 0.5, `3` = 0.8, `4` = 0.1, `5` = 1, `6` = 0.8, `7` = 0.2, `8` = 0.8, `9` = 0.1,
    `10` = 1, `11` = 0.8)))
displayRanks(filteredSecuritiesIdxQuant, rankAMatrix(minkowski, fctr = 1), minkowski,
    showDs = FALSE, sameRankShow = 3)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  0 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : AKAM CTXS  distance:  268506417.231126 index:  ( 13 , 75 )"
## [1] "Rank : 3  pair : CTXS DLR  distance:  350657303.648576 index:  ( 75 , 88 )"
## [1] "Rank : 4  pair : AME DPS  distance:  353277923.840989 index:  ( 19 , 91 )"
## [1] "Rank : 5  pair : CMG DPS  distance:  504435368.402998 index:  ( 66 , 91 )"
## [1] "Rank : 6  pair : AME CMG  distance:  551685681.491926 index:  ( 19 , 66 )"
## [1] "Rank : 7  pair : ALLE CHD  distance:  581105316.147757 index:  ( 17 , 57 )"
## [1] "Rank : 8  pair : AKAM DLR  distance:  592118707.799895 index:  ( 13 , 88 )"
## [1] "Rank : 9  pair : AMT AWK  distance:  594721252.927908 index:  ( 23 , 34 )"
## [1] "Rank : 10  pair : AIV CTXS  distance:  603834946.599024 index:  ( 10 , 75 )"
## [1] "######Bottom 10########"
## [1] "Rank : 4941  pair : BAC CTL  distance:  581835135807.123 index:  ( 37 , 73 )"
## [1] "Rank : 4942  pair : AIV BAC  distance:  581842732947.109 index:  ( 10 , 37 )"
## [1] "Rank : 4943  pair : ALK BAC  distance:  581856146712.942 index:  ( 15 , 37 )"
## [1] "Rank : 4944  pair : AN BAC  distance:  581887741712.188 index:  ( 25 , 37 )"
## [1] "Rank : 4945  pair : BAC DLR  distance:  581910271318.694 index:  ( 37 , 88 )"
## [1] "Rank : 4946  pair : AWK BAC  distance:  581926147804.527 index:  ( 34 , 37 )"
## [1] "Rank : 4947  pair : BAC COG  distance:  581942193806.618 index:  ( 37 , 72 )"
## [1] "Rank : 4948  pair : AEE BAC  distance:  581958146767.1 index:  ( 6 , 37 )"
## [1] "Rank : 4949  pair : BAC CHTR  distance:  581960143982.684 index:  ( 37 , 60 )"
## [1] "Rank : 4950  pair : BAC CXO  distance:  581963126141.479 index:  ( 37 , 78 )"
```

f) Match-Based Similarity Computation (use a small number of equi-depth buckets, ex: 3)

```
pSelctMatrix <- pSelectDistance(filteredSecQuantMatrix, p = 1, upper = TRUE, kd = 3)
# forcing NA for 0 distances, because some of the rows didn't match on any of
# fields bucket resulting in 0 distance rk1<-rankAMatrix(pSelctMatrix,fctr = -1)
# as.vector(which(rk1==4837,arr.ind = T))
pSelctMatrix[pSelctMatrix == 0] <- NA
displayRanks(filteredSecuritiesIdxQuant, rankAMatrix(pSelctMatrix, fctr = -1), pSelctMatrix,
    showDs = FALSE, sameRankShow = 10)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  11 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : ALLE DNB  distance:  9.43063309460863 index:  ( 17 , 89 )"
## [1] "Rank : 3  pair : D EIX  distance:  9.40611095513475 index:  ( 79 , 99 )"
## [1] "Rank : 4  pair : ARNC BHI  distance:  8.70537656525627 index:  ( 31 , 41 )"
## [1] "Rank : 5  pair : ALLE CHD  distance:  8.65959012601853 index:  ( 17 , 57 )"
## [1] "Rank : 6  pair : ADM CAT  distance:  8.56258613430367 index:  ( 4 , 50 )"
## [1] "Rank : 7  pair : AAL ARNC  distance:  8.4937582632173 index:  ( 1 , 31 )"
```

```
## [1] "Rank : 8  pair : AAL BHI  distance:  8.4829745577607 index:  ( 1 , 41 )"
## [1] "Rank : 9  pair : ABT CAT  distance:  8.39779427385182 index:  ( 3 , 50 )"
## [1] "Rank : 10  pair : ALB ALLE  distance:  8.13220499059124 index:  ( 14 , 17 )"
## [1] "######Bottom 10########"
## [1] "Rank : 4827  pair : CVX DVA  distance:  0.00956846203763795 index:  ( 77 , 93 )"
## [1] "Rank : 4828  pair : CBG CVX  distance:  0.00934294162192084 index:  ( 52 , 77 )"
## [1] "Rank : 4829  pair : AON APC  distance:  0.00831798587359089 index:  ( 27 , 29 )"
## [1] "Rank : 4830  pair : CTSH CVX  distance:  0.00495640201927494 index:  ( 74 , 77 )"
## [1] "Rank : 4831  pair : AIG EMN  distance:  0.00483991064780342 index:  ( 9 , 100 )"
## [1] "Rank : 4832  pair : ALK CVX  distance:  0.00481872418540619 index:  ( 15 , 77 )"
## [1] "Rank : 4833  pair : BLL C  distance:  0.00337185329861112 index:  ( 44 , 49 )"
## [1] "Rank : 4834  pair : AIZ CVX  distance:  0.00208942828602465 index:  ( 11 , 77 )"
## [1] "Rank : 4835  pair : C ECL  distance:  0.00191786024305551 index:  ( 49 , 96 )"
## [1] "Rank : 4836  pair : AMG CVX  distance:  0.000757228086278094 index:  ( 20 , 77 )"
```

g) Mahalanobis distance

```
mahalanobisDistanceMatrix <- mahalanobisDistance(filteredSecQuantMatrix, upper = TRUE)
displayRanks(filteredSecuritiesIdxQuant, rankAMatrix(mahalanobisDistanceMatrix, fctr = 1),
    mahalanobisDistanceMatrix, showDs = FALSE, sameRankShow = 10)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  0 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : CHD DNB  distance:  0.130838165079741 index:  ( 57 , 89 )"
## [1] "Rank : 3  pair : AME DPS  distance:  0.162585331202262 index:  ( 19 , 91 )"
## [1] "Rank : 4  pair : ALLE DNB  distance:  0.168819990099324 index:  ( 17 , 89 )"
## [1] "Rank : 5  pair : ALLE CHD  distance:  0.174801925722648 index:  ( 17 , 57 )"
## [1] "Rank : 6  pair : AME CHD  distance:  0.180006216789242 index:  ( 19 , 57 )"
## [1] "Rank : 7  pair : CHD CMG  distance:  0.1900583376053 index:  ( 57 , 66 )"
## [1] "Rank : 8  pair : AKAM CTXS  distance:  0.199142043284983 index:  ( 13 , 75 )"
## [1] "Rank : 9  pair : DNB EFX  distance:  0.199720884600913 index:  ( 89 , 98 )"
## [1] "Rank : 10  pair : AMG DPS  distance:  0.211011931789655 index:  ( 20 , 91 )"
## [1] "######Bottom 10########"
## [1] "Rank : 4941  pair : AIG C  distance:  12.6843595058411 index:  ( 9 , 49 )"
## [1] "Rank : 4942  pair : AIG CVX  distance:  12.7789325574695 index:  ( 9 , 77 )"
## [1] "Rank : 4943  pair : BK C  distance:  12.855097920768 index:  ( 43 , 49 )"
## [1] "Rank : 4944  pair : AIG BAC  distance:  13.0920158573914 index:  ( 9 , 37 )"
## [1] "Rank : 4945  pair : APC BAC  distance:  13.2725035840845 index:  ( 29 , 37 )"
## [1] "Rank : 4946  pair : APC C  distance:  13.2752100260823 index:  ( 29 , 49 )"
## [1] "Rank : 4947  pair : APC CVX  distance:  13.508620318013 index:  ( 29 , 77 )"
## [1] "Rank : 4948  pair : C CVX  distance:  13.5168206089448 index:  ( 49 , 77 )"
## [1] "Rank : 4949  pair : BAC CVX  distance:  13.6579136282132 index:  ( 37 , 77 )"
## [1] "Rank : 4950  pair : BAC C  distance:  13.7343548321117 index:  ( 37 , 49 )"
```

##Categorical Variables

h) Similarity: overlap measure

```
overlapSimMtx <- similarityfn(dataMatrix = securitiesCategoricalMtx, upper = TRUE,
    type = "overlap")
```

```
displayRanks(securitiesCategoricalFiltered, rankAMatrix(replace(overlapSimMtx, overlapSimMtx ==
    0, NA), fctr = -1), overlapSimMtx, showDs = FALSE, sameRankShow = 10)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : BK C  distance:  3 index:  ( 43 , 49 )"
## [1] "Rank : 1  pair : APA COG  distance:  3 index:  ( 28 , 72 )"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  3 index:  ( 85 , 86 )"
## [1] "Rank : 4  pair : ABBV ABT  distance:  2 index:  ( 2 , 3 )"
## [1] "Rank : 4  pair : AIG AIZ  distance:  2 index:  ( 9 , 11 )"
## [1] "Rank : 4  pair : AAL ALK  distance:  2 index:  ( 1 , 15 )"
## [1] "Rank : 4  pair : AIG ALL  distance:  2 index:  ( 9 , 16 )"
## [1] "Rank : 4  pair : ALXN AMGN  distance:  2 index:  ( 18 , 21 )"
## [1] "Rank : 4  pair : AMG AMP  distance:  2 index:  ( 20 , 22 )"
## [1] "Rank : 4  pair : AJG AON  distance:  2 index:  ( 12 , 27 )"
## [1] "Rank : 4  pair : APA APC  distance:  2 index:  ( 28 , 29 )"
## [1] "Rank : 4  pair : AIG AXP  distance:  2 index:  ( 9 , 35 )"
## [1] "Rank : 4  pair : AIZ AXP  distance:  2 index:  ( 11 , 35 )"
## [1] "######Bottom 10########"
## [1] "Rank : 94  pair : AEE AEP  distance:  1 index:  ( 6 , 7 )"
## [1] "Rank : 94  pair : AFL AIG  distance:  1 index:  ( 8 , 9 )"
## [1] "Rank : 94  pair : AFL AIZ  distance:  1 index:  ( 8 , 11 )"
## [1] "Rank : 94  pair : AFL AJG  distance:  1 index:  ( 8 , 12 )"
## [1] "Rank : 94  pair : AIG AJG  distance:  1 index:  ( 9 , 12 )"
## [1] "Rank : 94  pair : AIZ AJG  distance:  1 index:  ( 11 , 12 )"
## [1] "Rank : 94  pair : ADS AKAM  distance:  1 index:  ( 5 , 13 )"
## [1] "Rank : 94  pair : AFL ALL  distance:  1 index:  ( 8 , 16 )"
## [1] "Rank : 94  pair : AIZ ALL  distance:  1 index:  ( 11 , 16 )"
## [1] "Rank : 94  pair : AJG ALL  distance:  1 index:  ( 12 , 16 )"
```

i) Similarity: inverse frequency

```
inverseSimMtx <- similarityfn(dataMatrix = securitiesCategoricalMtx, upper = TRUE,
    type = "inverse")
displayRanks(securitiesCategoricalFiltered, rankAMatrix(replace(inverseSimMtx, inverseSimMtx ==
    0, NA), fctr = -1), inverseSimMtx, showDs = FALSE, sameRankShow = 10)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  3767.36111111111 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : CHD CL  distance:  2900 index:  ( 57 , 63 )"
## [1] "Rank : 3  pair : ALB ECL  distance:  2777.77777777778 index:  ( 14 , 96 )"
## [1] "Rank : 3  pair : DD EMN  distance:  2777.77777777778 index:  ( 81 , 100 )"
## [1] "Rank : 5  pair : AMT BXP  distance:  2704.08163265306 index:  ( 23 , 48 )"
## [1] "Rank : 5  pair : AMT DLR  distance:  2704.08163265306 index:  ( 23 , 88 )"
## [1] "Rank : 7  pair : CHK CVX  distance:  2656.25 index:  ( 58 , 77 )"
## [1] "Rank : 7  pair : BWA DLPH  distance:  2656.25 index:  ( 47 , 87 )"
## [1] "Rank : 7  pair : CHK DVN  distance:  2656.25 index:  ( 58 , 94 )"
## [1] "Rank : 10  pair : ARNC BA  distance:  2551.02040816327 index:  ( 31 , 36 )"
## [1] "Rank : 10  pair : CMI DOV  distance:  2551.02040816327 index:  ( 67 , 90 )"
## [1] "Rank : 10  pair : DAL EFX  distance:  2551.02040816327 index:  ( 80 , 98 )"
## [1] "Rank : 10  pair : DNB EFX  distance:  2551.02040816327 index:  ( 89 , 98 )"
## [1] "######Bottom 10########"
```

```
## [1] "Rank : 425  pair : AFL AIG  distance:  25 index:  ( 8 , 9 )"
## [1] "Rank : 425  pair : AFL AIZ  distance:  25 index:  ( 8 , 11 )"
## [1] "Rank : 425  pair : AFL AJG  distance:  25 index:  ( 8 , 12 )"
## [1] "Rank : 425  pair : AIG AJG  distance:  25 index:  ( 9 , 12 )"
## [1] "Rank : 425  pair : AIZ AJG  distance:  25 index:  ( 11 , 12 )"
## [1] "Rank : 425  pair : AFL ALL  distance:  25 index:  ( 8 , 16 )"
## [1] "Rank : 425  pair : AIZ ALL  distance:  25 index:  ( 11 , 16 )"
## [1] "Rank : 425  pair : AJG ALL  distance:  25 index:  ( 12 , 16 )"
## [1] "Rank : 425  pair : AFL AMG  distance:  25 index:  ( 8 , 20 )"
## [1] "Rank : 425  pair : AIG AMG  distance:  25 index:  ( 9 , 20 )"
```

j) Similarity: Goodall

```
goodallSimMtx <- similarityfn(dataMatrix = securitiesCategoricalMtx, upper = TRUE,
    type = "goodall")
displayRanks(securitiesCategoricalFiltered, rankAMatrix(replace(goodallSimMtx, goodallSimMtx ==
    0, NA), fctr = -1), goodallSimMtx, showDs = FALSE, sameRankShow = 10)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  2.9923 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : APA COG  distance:  2.9886 index:  ( 28 , 72 )"
## [1] "Rank : 3  pair : BK C  distance:  2.9503 index:  ( 43 , 49 )"
## [1] "Rank : 4  pair : CHD CL  distance:  1.9971 index:  ( 57 , 63 )"
## [1] "Rank : 5  pair : ALB ECL  distance:  1.996 index:  ( 14 , 96 )"
## [1] "Rank : 5  pair : DD EMN  distance:  1.996 index:  ( 81 , 100 )"
## [1] "Rank : 7  pair : AMT BXP  distance:  1.9947 index:  ( 23 , 48 )"
## [1] "Rank : 7  pair : AMT DLR  distance:  1.9947 index:  ( 23 , 88 )"
## [1] "Rank : 9  pair : AIV BXP  distance:  1.9942 index:  ( 10 , 48 )"
## [1] "Rank : 9  pair : AIV CCI  distance:  1.9942 index:  ( 10 , 53 )"
## [1] "Rank : 9  pair : BXP CCI  distance:  1.9942 index:  ( 48 , 53 )"
## [1] "Rank : 9  pair : AKAM CTXS  distance:  1.9942 index:  ( 13 , 75 )"
## [1] "Rank : 9  pair : AKAM EBAY  distance:  1.9942 index:  ( 13 , 95 )"
## [1] "Rank : 9  pair : CTXS EBAY  distance:  1.9942 index:  ( 75 , 95 )"
## [1] "######Bottom 10########"
## [1] "Rank : 425  pair : AFL AIG  distance:  0.96 index:  ( 8 , 9 )"
## [1] "Rank : 425  pair : AFL AIZ  distance:  0.96 index:  ( 8 , 11 )"
## [1] "Rank : 425  pair : AFL AJG  distance:  0.96 index:  ( 8 , 12 )"
## [1] "Rank : 425  pair : AIG AJG  distance:  0.96 index:  ( 9 , 12 )"
## [1] "Rank : 425  pair : AIZ AJG  distance:  0.96 index:  ( 11 , 12 )"
## [1] "Rank : 425  pair : AFL ALL  distance:  0.96 index:  ( 8 , 16 )"
## [1] "Rank : 425  pair : AIZ ALL  distance:  0.96 index:  ( 11 , 16 )"
## [1] "Rank : 425  pair : AJG ALL  distance:  0.96 index:  ( 12 , 16 )"
## [1] "Rank : 425  pair : AFL AMG  distance:  0.96 index:  ( 8 , 20 )"
## [1] "Rank : 425  pair : AIG AMG  distance:  0.96 index:  ( 9 , 20 )"
```

Merged dataset

k) Overall similarity between tickers by using mixed type data (choose a lambda value for calculation)

```
overallSimilarityMtx <- overallSimilarity(mahalanobisDistanceMatrix, goodallSimMtx,
    0.8)
displayRanks(mergedQuantAndCatDS, rankAMatrix(overallSimilarityMtx, fctr = -1), overallSimilarityMtx,
    showDs = FALSE, sameRankShow = 3)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  1.39846 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : AKAM CTXS  distance:  1.06598365031222 index:  ( 13 , 75 )"
## [1] "Rank : 3  pair : DNB EFX  distance:  1.06282176685298 index:  ( 89 , 98 )"
## [1] "Rank : 4  pair : AEE CMS  distance:  1.03633347621952 index:  ( 6 , 68 )"
## [1] "Rank : 5  pair : AMT DLR  distance:  0.91026933543739 index:  ( 23 , 88 )"
## [1] "Rank : 6  pair : ALLE DNB  distance:  0.880530990551605 index:  ( 17 , 89 )"
## [1] "Rank : 7  pair : AIV BXP  distance:  0.874066748976519 index:  ( 10 , 48 )"
## [1] "Rank : 8  pair : AMT BXP  distance:  0.864596368372274 index:  ( 23 , 48 )"
## [1] "Rank : 9  pair : CMI DOV  distance:  0.859630271650241 index:  ( 67 , 90 )"
## [1] "Rank : 10  pair : AME DNB  distance:  0.850185009786169 index:  ( 19 , 89 )"
## [1] "######Bottom 10########"
## [1] "Rank : 4941  pair : BA BAC  distance:  0.0601072198909198 index:  ( 36 , 37 )"
## [1] "Rank : 4942  pair : APA C  distance:  0.05999032213971 index:  ( 28 , 49 )"
## [1] "Rank : 4943  pair : BK CVX  distance:  0.0595379811582895 index:  ( 43 , 77 )"
## [1] "Rank : 4944  pair : APC BK  distance:  0.0594308968088498 index:  ( 29 , 43 )"
## [1] "Rank : 4945  pair : AMGN BK  distance:  0.0594282914071192 index:  ( 21 , 43 )"
## [1] "Rank : 4946  pair : AIG CVX  distance:  0.0580596498795055 index:  ( 9 , 77 )"
## [1] "Rank : 4947  pair : APC BAC  distance:  0.0560518338837267 index:  ( 29 , 37 )"
## [1] "Rank : 4948  pair : APC C  distance:  0.0560412069971873 index:  ( 29 , 49 )"
## [1] "Rank : 4949  pair : C CVX  distance:  0.0551084856354198 index:  ( 49 , 77 )"
## [1] "Rank : 4950  pair : BAC CVX  distance:  0.0545780266067457 index:  ( 37 , 77 )"
```

l) Overall normalized similarity between tickers by using mixed type data (choose a lambda value for calculation)

```
overallNormSimilarityMtx <- overallNormSimilarity(mahalanobisDistanceMatrix, goodallSimMtx,
    0.8)
displayRanks(mergedQuantAndCatDS, rankAMatrix(overallNormSimilarityMtx, fctr = -1),
    overallNormSimilarityMtx, showDs = FALSE, sameRankShow = 3)
```

```
## [1] "######Top 10########"
## [1] "Rank : 1  pair : DISCA DISCK  distance:  6.53731008281393 index:  ( 85 , 86 )"
## [1] "Rank : 2  pair : AKAM CTXS  distance:  5.19591759364172 index:  ( 13 , 75 )"
## [1] "Rank : 3  pair : DNB EFX  distance:  5.18665558846981 index:  ( 89 , 98 )"
## [1] "Rank : 4  pair : AEE CMS  distance:  5.01259459945434 index:  ( 6 , 68 )"
## [1] "Rank : 5  pair : ALLE DNB  distance:  4.78698549590141 index:  ( 17 , 89 )"
## [1] "Rank : 6  pair : AME DNB  distance:  4.59692928806898 index:  ( 19 , 89 )"
## [1] "Rank : 7  pair : CHD DPS  distance:  4.56462697167637 index:  ( 57 , 91 )"
## [1] "Rank : 8  pair : AME EFX  distance:  4.52768948953189 index:  ( 19 , 98 )"
## [1] "Rank : 9  pair : AME DOV  distance:  4.50718061095405 index:  ( 19 , 90 )"
## [1] "Rank : 10  pair : CHD DNB  distance:  4.43068026851635 index:  ( 57 , 89 )"
```

```
## [1] "######Bottom 10########"
## [1] "Rank : 4941  pair : BA BAC  distance:  0.376450191678412 index:  ( 36 , 37 )"
## [1] "Rank : 4942  pair : APA C  distance:  0.375718063642386 index:  ( 28 , 49 )"
## [1] "Rank : 4943  pair : BK CVX  distance:  0.372885062058404 index:  ( 43 , 77 )"
## [1] "Rank : 4944  pair : APC BK  distance:  0.372214395141094 index:  ( 29 , 43 )"
## [1] "Rank : 4945  pair : AMGN BK  distance:  0.372198077567553 index:  ( 21 , 43 )"
## [1] "Rank : 4946  pair : AIG CVX  distance:  0.36362630588448 index:  ( 9 , 77 )"
## [1] "Rank : 4947  pair : APC BAC  distance:  0.351051398613148 index:  ( 29 , 37 )"
## [1] "Rank : 4948  pair : APC C  distance:  0.350984842657275 index:  ( 29 , 49 )"
## [1] "Rank : 4949  pair : C CVX  distance:  0.345143229352631 index:  ( 49 , 77 )"
## [1] "Rank : 4950  pair : BAC CVX  distance:  0.341820976162677 index:  ( 37 , 77 )"
```