Government Employment Figures by Region

The data set is available from the website LMIP Gov AU.

It contains regional figures for population counts of employment per industry.

```
data <- read.csv("data/employment/SA4_regions_feb2017.csv", header=TRUE)

# need to fix the numbers
for(i in 4:9) {
  data[,i] <- as.numeric(gsub(",", "", data[,i]))
}
head(data)
```

```
##   Employment.Region State.Territory
## 1    Capital Region         NSW/ACT
## 2    Capital Region         NSW/ACT
## 3    Capital Region         NSW/ACT
## 4    Capital Region         NSW/ACT
## 5    Capital Region         NSW/ACT
## 6    Capital Region         NSW/ACT
##                                    Industry
## 1          Agriculture, Forestry and Fishing
## 2                                     Mining
## 3                                Manufacturing
## 4 Electricity, Gas, Water and Waste Services
## 5                                Construction
## 6                             Wholesale Trade
##   Employment.by.Industry...Total Employed.Full.Time Employed.Part.Time
## 1                           7200               5100               2100
## 2                            800                800                  0
## 3                           9500               7900               1600
## 4                           2500               2500                  0
## 5                          24400              21900               2500
## 6                           4000               3000               1000
##   Employed...Male Employed...Female Five.year.growth.by.Industry
## 1            5800              1300                        -4300
## 2             500               300                          300
## 3            7500              2000                        -1800
## 4            2200               300                         -400
## 5           22600              1800                         -500
## 6            2800              1200                        -1200
##   Employment.Distribution....
## 1                         2.2
## 2                         0.2
## 3                         3.0
## 4                         0.8
## 5                         7.6
## 6                         1.2
```

```
str(data)
```

```
## 'data.frame':    836 obs. of  10 variables:
##  $ Employment.Region            : Factor w/ 44 levels "Adelaide North",..: 9 9 9 9 9 9 9 9 9 9 ...
```

```
##  $ State.Territory               : Factor w/ 9 levels "","NSW","NSW/ACT",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ Industry                      : Factor w/ 19 levels "Accommodation and Food Services",..: 3 12 11
##  $ Employment.by.Industry...Total: num  7200 800 9500 2500 24400 4000 24700 22500 11800 3800 ...
##  $ Employed.Full.Time            : num  5100 800 7900 2500 21900 3000 12000 8000 9100 3000 ...
##  $ Employed.Part.Time            : num  2100 0 1600 0 2500 1000 12700 14600 2700 800 ...
##  $ Employed...Male               : num  5800 500 7500 2200 22600 2800 10800 11200 9800 2100 ...
##  $ Employed...Female             : num  1300 300 2000 300 1800 1200 14000 11400 1900 1700 ...
##  $ Five.year.growth.by.Industry  : num  -4300 300 -1800 -400 -500 -1200 1100 1900 1900 -2700 ...
##  $ Employment.Distribution....   : num  2.2 0.2 3 0.8 7.6 1.2 7.7 7 3.7 1.2 ...
```

Initially we will investigate the industry by region, the field of interest is the employment by industry total. The data is currently in long format we need to convert it to wide format.

```
data1 <- data[,c(1:4)]

colnames(data1) <- c("Region", "State", "Industry", "TotalCount")
data1 <- data1[data1$Region != "Australia",]

head(data1)
```

```
##           Region   State                                    Industry
## 1 Capital Region NSW/ACT          Agriculture, Forestry and Fishing
## 2 Capital Region NSW/ACT                                     Mining
## 3 Capital Region NSW/ACT                                Manufacturing
## 4 Capital Region NSW/ACT Electricity, Gas, Water and Waste Services
## 5 Capital Region NSW/ACT                                  Construction
## 6 Capital Region NSW/ACT                              Wholesale Trade
##   TotalCount
## 1       7200
## 2        800
## 3       9500
## 4       2500
## 5      24400
## 6       4000
```

```
temp <- reshape(data1, idvar=c("Region", "State"), timevar=c("Industry"), direction="wide")


temp$Location <- paste(temp$State, temp$Region, sep=" ")
names(temp)
```

```
##  [1] "Region"
##  [2] "State"
##  [3] "TotalCount.Agriculture, Forestry and Fishing"
##  [4] "TotalCount.Mining"
##  [5] "TotalCount.Manufacturing"
##  [6] "TotalCount.Electricity, Gas, Water and Waste Services"
##  [7] "TotalCount.Construction"
##  [8] "TotalCount.Wholesale Trade"
##  [9] "TotalCount.Retail Trade"
## [10] "TotalCount.Accommodation and Food Services"
## [11] "TotalCount.Transport, Postal and Warehousing"
## [12] "TotalCount.Information Media and Telecommunications"
## [13] "TotalCount.Financial and Insurance Services"
## [14] "TotalCount.Rental, Hiring and Real Estate Services"
## [15] "TotalCount.Professional, Scientific and Technical Services"
```

```
## [16] "TotalCount.Administrative and Support Services"
## [17] "TotalCount.Public Administration and Safety"
## [18] "TotalCount.Education and Training"
## [19] "TotalCount.Health Care and Social Assistance"
## [20] "TotalCount.Arts and Recreation Services"
## [21] "TotalCount.Other Services"
## [22] "Location"
```

```
colnames(temp) <- c("Region",
                    "State",
                    "AGRIC_FRST_FISH",
                    "MINING",
                    "MANUF",
                    "UTILITIES",
                    "CONSTR",
                    "WSALE_TRADE",
                    "RETAIL_TRADE",
                    "ACC_FOOD_SRV",
                    "TRNS_POST_WHOUSE",
                    "INFO_MEDIA_TELEC",
                    "FIN_INS_SRV",
                    "RENT_HIRE_RE_SRV",
                    "PROF_SCI_TECH_SRV",
                    "ADM_SUP_SRV",
                    "PADMIN_SAFETY",
                    "EDU_TRAIN",
                    "HEALTH_SOC_ASSIST",
                    "ARTS_REC_SRV",
                    "OTHER_SRV",
                    "Location")
df1 <- data.frame(Location=temp$Location,
                  temp[,4:ncol(temp)-1])
head(df1)
```

```
##                     Location AGRIC_FRST_FISH MINING MANUF UTILITIES CONSTR
## 1     NSW/ACT Capital Region            7200    800  9500      2500  24400
## 20          NSW Central West           11700   8200  5200      2600   7800
## 39        NSW Far West Orana           10500   1300  2600       400   4500
## 58                NSW Hunter            7200  17200 20900      4700  30100
## 77 NSW Illawarra South Coast             400   2900 13000      1800  22700
## 96       NSW Mid North Coast            9500    400  6600      2800  16100
##    WSALE_TRADE RETAIL_TRADE ACC_FOOD_SRV TRNS_POST_WHOUSE INFO_MEDIA_TELEC
## 1         4000        24700        22500             11800              3800
## 20        2900         9900         5600              3800               400
## 39        1700         6900         5000               500                 0
## 58        7100        30700        28000             13500              2800
## 77        3100        23300        16800             10500              2900
## 96        3300        21500        14000              4000               700
##    FIN_INS_SRV RENT_HIRE_RE_SRV PROF_SCI_TECH_SRV ADM_SUP_SRV
## 1         5900             4600             33500        9100
## 20        2400              500              4300        1300
## 39        1300                0               600        2300
## 58        9800             4000             18000       12200
## 77        5400             3500             10700        7600
## 96        1700             1900              8500        4600
```

```
##    PADMIN_SAFETY EDU_TRAIN HEALTH_SOC_ASSIST ARTS_REC_SRV OTHER_SRV
## 1          77200     23600              36400         8000     10700
## 20          7900      8200              15800         1800      5300
## 39          3600      6100               7700          600      2800
## 58         17400     26700              45600         6800     13600
## 77         10900     23000              32600         2700      5800
## 96         11300     13100              24600         1400      6500
```

Looking initially at the data we can check if it is multivariate normal,

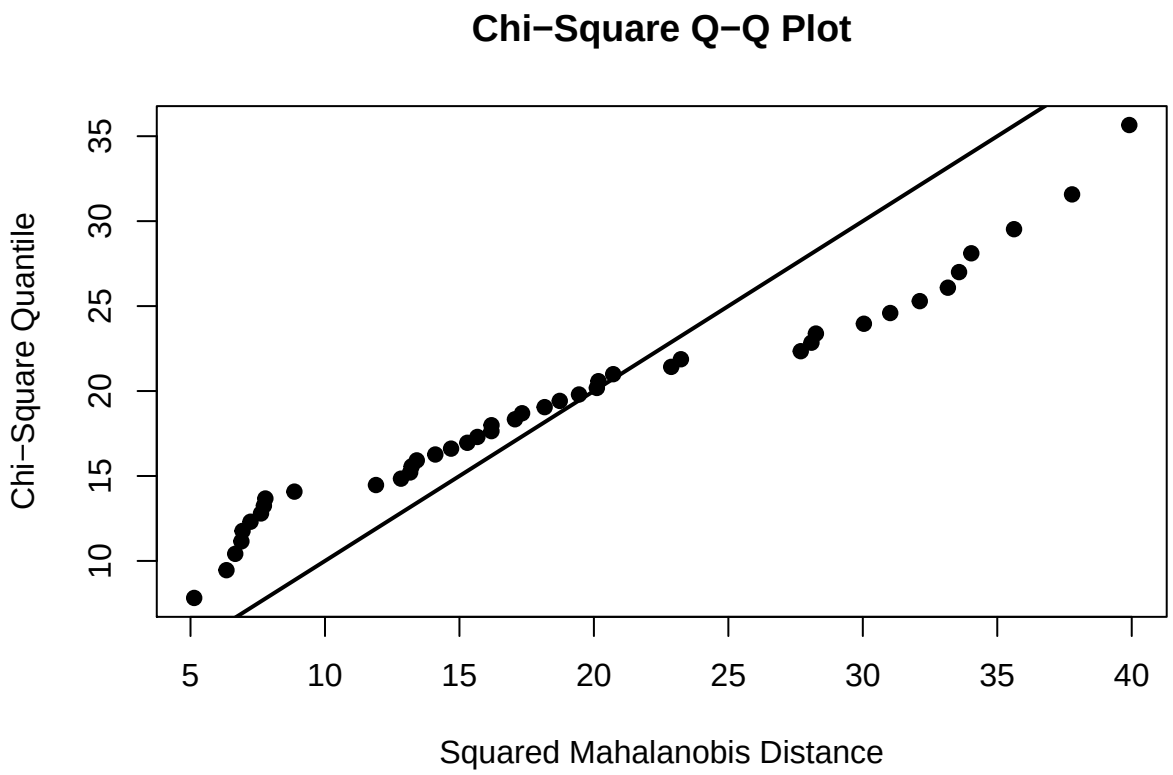
```
require(MVN)
```

```
## Loading required package: MVN
```

```
##
## This data.table install has not detected OpenMP support. It will work but slower in single threaded r
```

```
## sROC 0.1-2 loaded
```

```
X <- df1[,2:ncol(df1)]
X <- scale(X)
mardiaTest(X, qqplot=TRUE)
```




```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : X
##
##    g1p            : 285.222
##    chi.skew       : 2044.091
##    p.value.skew   : 3.269859e-33
##
##    g2p            : 456.9211
```

```
##    z.kurtosis     : 6.722636
##    p.value.kurt   : 1.784661e-11
##
##    chi.small.skew : 2201.714
##    p.value.small  : 4.51671e-46
##
##    Result          : Data are not multivariate normal.
## ---------------------------------------
```

```
hzTest(X)
```

```
##   Henze-Zirkler's Multivariate Normality Test
## ---------------------------------------------
##   data : X
##
##   HZ      : 1.026877
##   p-value : 0
##
##   Result  : Data are not multivariate normal.
## ---------------------------------------------
```

```
roystonTest(X)
```

```
##   Royston's Multivariate Normality Test
## ---------------------------------------------
##   data : X
##
##   H       : 68.46346
##   p-value : 1.764763e-14
##
##   Result  : Data are not multivariate normal.
## ---------------------------------------------
```

The test results suggest that the data is not multivariate normal, and this is also reinforced by the qqplot. However we can still perform ordination and some analysis.

Initially inspecting the data through principle components.

```
row.names(X) <- df1$Location
df1.prcomp <- princomp(X, cor=TRUE)
df1.prcomp
```

```
## Call:
## princomp(x = X, cor = TRUE)
##
## Standard deviations:
##     Comp.1     Comp.2     Comp.3     Comp.4     Comp.5     Comp.6
## 3.88669332 1.09781697 0.92408420 0.82408389 0.68134731 0.50312674
##     Comp.7     Comp.8     Comp.9    Comp.10    Comp.11    Comp.12
## 0.35455678 0.31044448 0.26588623 0.18540309 0.17999963 0.15488755
##    Comp.13    Comp.14    Comp.15    Comp.16    Comp.17    Comp.18
## 0.12072621 0.10609013 0.10174910 0.08643939 0.07003998 0.06000035
##    Comp.19
## 0.04785478
##
##  19  variables and  43 observations.
```

```
df1.prcomp$loadings
```

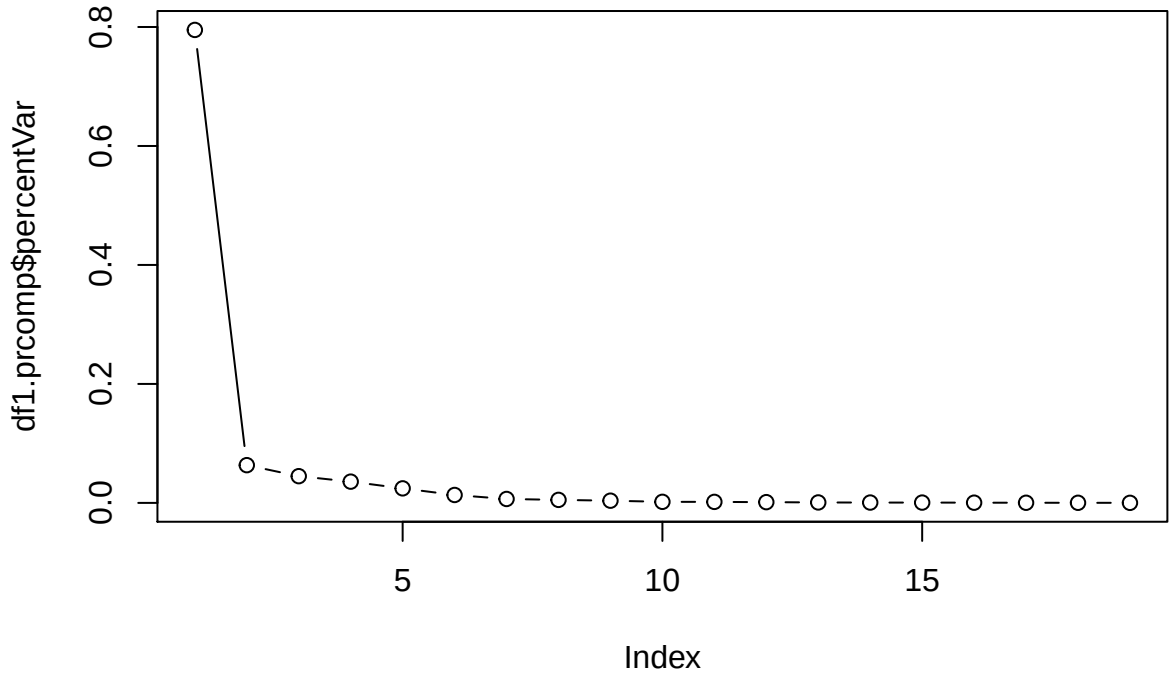
```
##
## Loadings:
##                   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
## AGRIC_FRST_FISH    0.124         0.818 -0.501  0.160 -0.155
## MINING                  -0.830  0.115  0.307 -0.208 -0.286         0.174
## MANUF             -0.222        -0.251 -0.502               -0.177  0.190
## UTILITIES         -0.224 -0.302        -0.167         0.554 -0.159 -0.264
## CONSTR            -0.245 -0.126 -0.120 -0.163        -0.111  0.378  0.200
## WSALE_TRADE       -0.247               -0.243 -0.132        -0.211  0.356
## RETAIL_TRADE      -0.254                                     0.104 -0.122
## ACC_FOOD_SRV      -0.250         0.131  0.101                0.208 -0.386
## TRNS_POST_WHOUSE  -0.228        -0.266 -0.209        -0.587 -0.193 -0.516
## INFO_MEDIA_TELEC  -0.239  0.256  0.145  0.115 -0.191        -0.176
## FIN_INS_SRV       -0.238  0.234  0.159        -0.203 -0.233 -0.169  0.105
## RENT_HIRE_RE_SRV  -0.244  0.126         0.148        -0.143  0.591
## PROF_SCI_TECH_SRV -0.241  0.163  0.215  0.215               -0.221  0.216
## ADM_SUP_SRV       -0.254                                     0.133
## PADMIN_SAFETY     -0.197               0.224  0.895 -0.102 -0.117  0.207
## EDU_TRAIN         -0.252                                    0.212         0.128
## HEALTH_SOC_ASSIST -0.253                                    0.238  0.165
## ARTS_REC_SRV      -0.244         0.173  0.216         0.120 -0.328 -0.347
## OTHER_SRV         -0.248 -0.164        -0.179                0.174  0.112
##                   Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15
## AGRIC_FRST_FISH
## MINING
## MANUF              0.380          0.308          -0.207  -0.259
## UTILITIES         -0.591          0.102                           -0.125
## CONSTR            -0.178  0.243  -0.668                  -0.326
## WSALE_TRADE              0.212           0.221   0.259          -0.170
## RETAIL_TRADE       0.301 -0.120  -0.220  -0.155  -0.212   0.232  -0.710
## ACC_FOOD_SRV       0.363                          0.629  -0.255
## TRNS_POST_WHOUSE  -0.184 -0.252
## INFO_MEDIA_TELEC          0.329          -0.265           0.171  -0.217
## FIN_INS_SRV       -0.262 -0.224          -0.260   0.159           0.113
## RENT_HIRE_RE_SRV  -0.186          0.410   0.383  -0.329
## PROF_SCI_TECH_SRV -0.158          0.110   0.257          -0.146  -0.136
## ADM_SUP_SRV        0.105          0.153  -0.620  -0.317  -0.132   0.347
## PADMIN_SAFETY
## EDU_TRAIN          0.127 -0.441  -0.341   0.248  -0.127   0.380   0.304
## HEALTH_SOC_ASSIST        -0.444                   0.160  -0.129
## ARTS_REC_SRV       0.233  0.399  -0.131   0.300  -0.254           0.260
## OTHER_SRV                 0.294   0.176           0.305   0.668   0.239
##                   Comp.16 Comp.17 Comp.18 Comp.19
## AGRIC_FRST_FISH
## MINING
## MANUF             -0.396                   0.161
## UTILITIES                  0.155
## CONSTR            -0.105                   0.121
## WSALE_TRADE        0.489   0.110          -0.474
## RETAIL_TRADE       0.202          -0.204   0.175
## ACC_FOOD_SRV      -0.138   0.293   0.106
## TRNS_POST_WHOUSE          -0.118   0.222
```

```
## INFO_MEDIA_TELEC  -0.524  -0.161   0.325  -0.303
## FIN_INS_SRV       -0.121   0.147  -0.666
## RENT_HIRE_RE_SRV           0.119  -0.116  -0.156
## PROF_SCI_TECH_SRV  0.138           0.375   0.642
## ADM_SUP_SRV        0.415   0.149   0.229
## PADMIN_SAFETY
## EDU_TRAIN                  0.373   0.184  -0.191
## HEALTH_SOC_ASSIST         -0.743          -0.179
## ARTS_REC_SRV       0.128  -0.239  -0.297
## OTHER_SRV                 -0.148           0.290
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var  0.053  0.053  0.053  0.053  0.053  0.053  0.053  0.053
## Cumulative Var  0.053  0.105  0.158  0.211  0.263  0.316  0.368  0.421
##                Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15
## SS loadings     1.000   1.000   1.000   1.000   1.000   1.000   1.000
## Proportion Var  0.053   0.053   0.053   0.053   0.053   0.053   0.053
## Cumulative Var  0.474   0.526   0.579   0.632   0.684   0.737   0.789
##                Comp.16 Comp.17 Comp.18 Comp.19
## SS loadings      1.000   1.000   1.000   1.000
## Proportion Var   0.053   0.053   0.053   0.053
## Cumulative Var   0.842   0.895   0.947   1.000
```

```
df1.prcomp$var <- df1.prcomp$sdev^2
```

The amount of variance explained per component

```
total <- sum(df1.prcomp$var)
df1.prcomp$percentVar <- df1.prcomp$var / total
plot(df1.prcomp$percentVar, type="b")
```



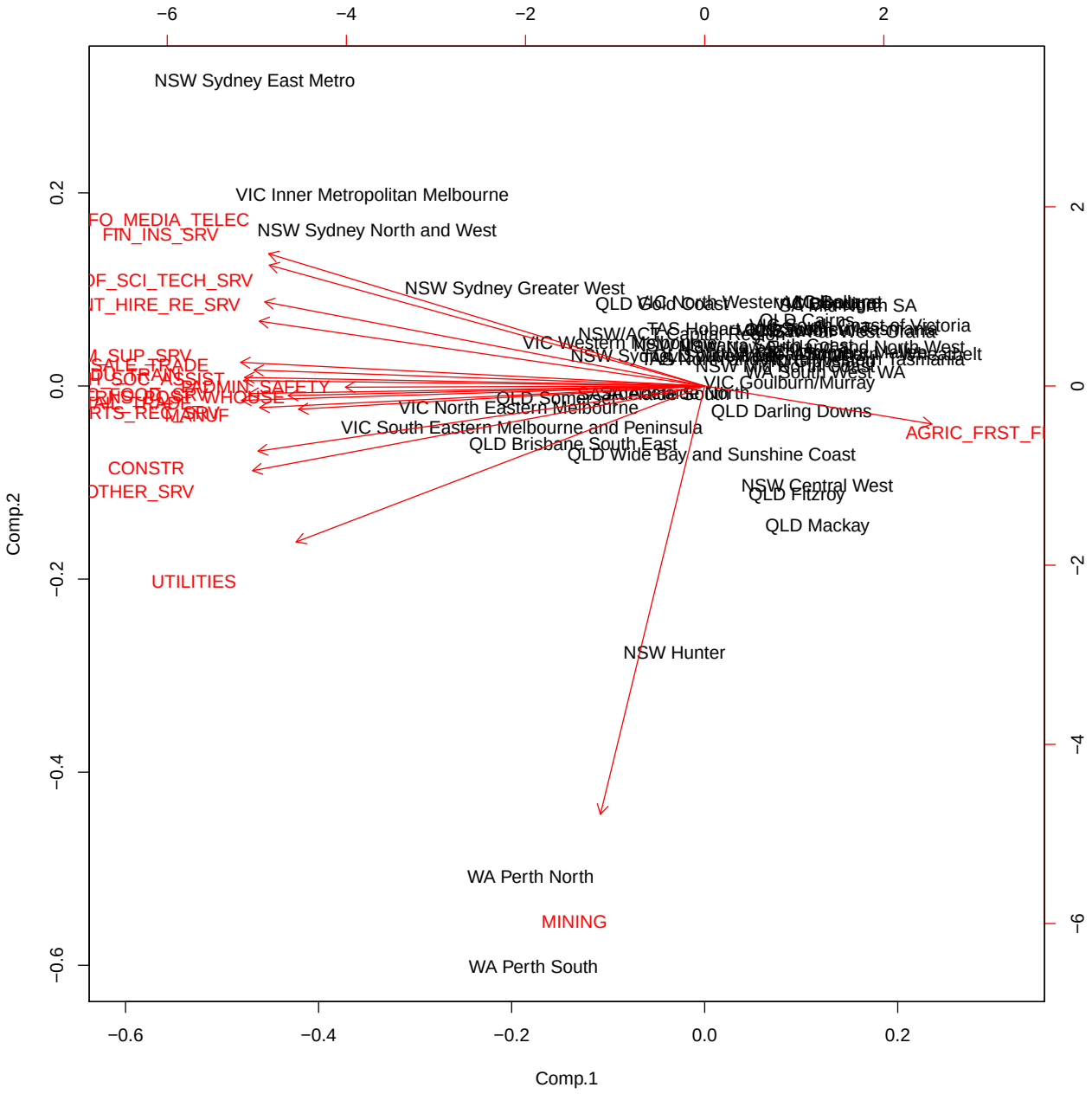

```
data.frame(component=1:length(df1.prcomp$var), variance=df1.prcomp$var, percent=round(df1.prcomp$percen
```

```
##         component     variance percent
## Comp.1          1 15.106384947  0.7951
## Comp.2          2  1.205202104  0.0634
## Comp.3          3  0.853931608  0.0449
## Comp.4          4  0.679114257  0.0357
## Comp.5          5  0.464234160  0.0244
## Comp.6          6  0.253136518  0.0133
## Comp.7          7  0.125710513  0.0066
## Comp.8          8  0.096375778  0.0051
## Comp.9          9  0.070695489  0.0037
## Comp.10        10  0.034374305  0.0018
## Comp.11        11  0.032399865  0.0017
## Comp.12        12  0.023990153  0.0013
## Comp.13        13  0.014574817  0.0008
## Comp.14        14  0.011255115  0.0006
## Comp.15        15  0.010352880  0.0005
## Comp.16        16  0.007471769  0.0004
## Comp.17        17  0.004905599  0.0003
## Comp.18        18  0.003600042  0.0002
## Comp.19        19  0.002290080  0.0001
```

We note that the first component explains 99% of the variance, however in the biplot it is difficult to decipher.

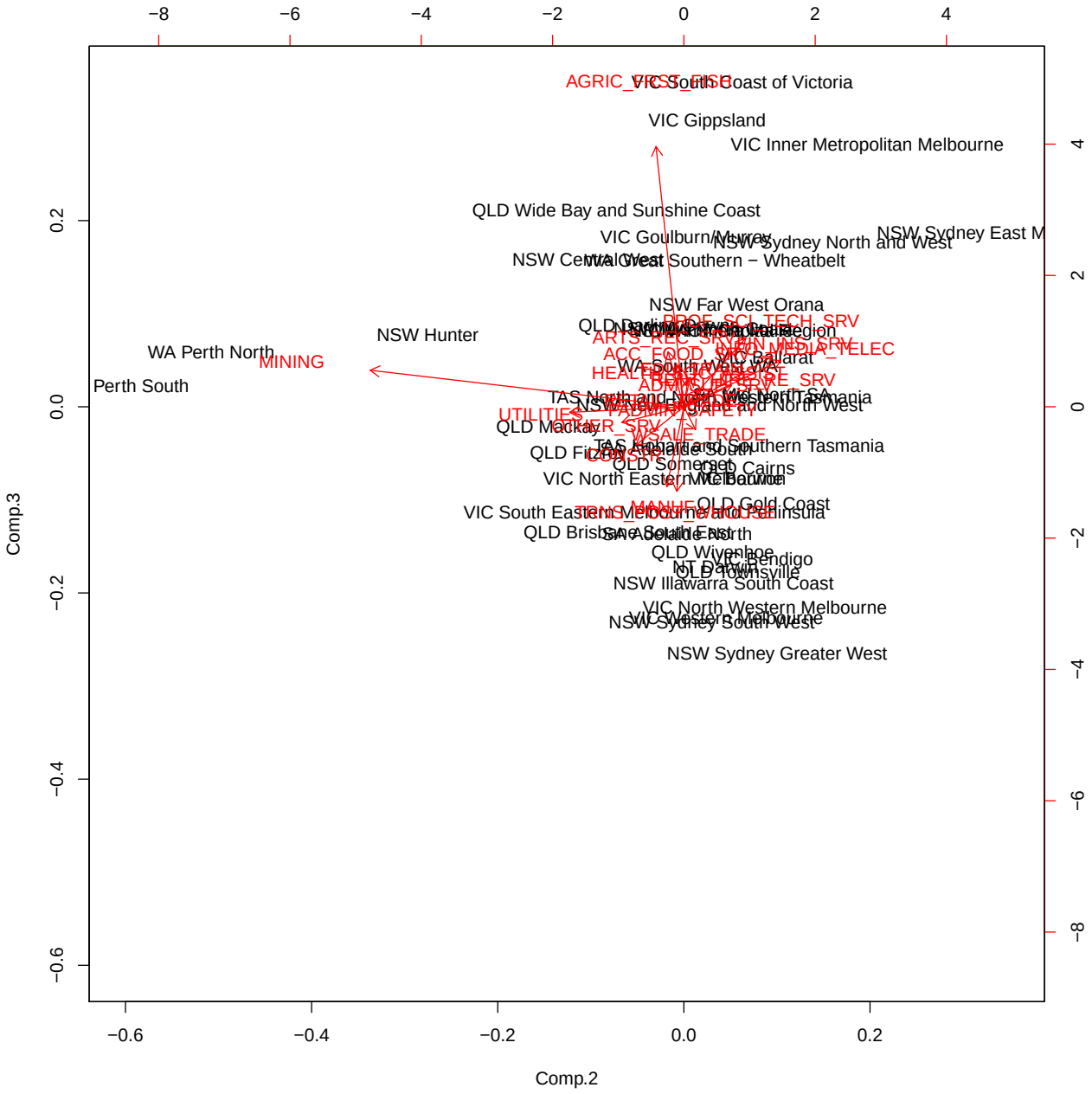
Printing the biplot for industries, there are a large number of regions hence we can move the viewport in order to get a better view of the ordination.

```
biplot(df1.prcomp, choices=c(1,2))
```



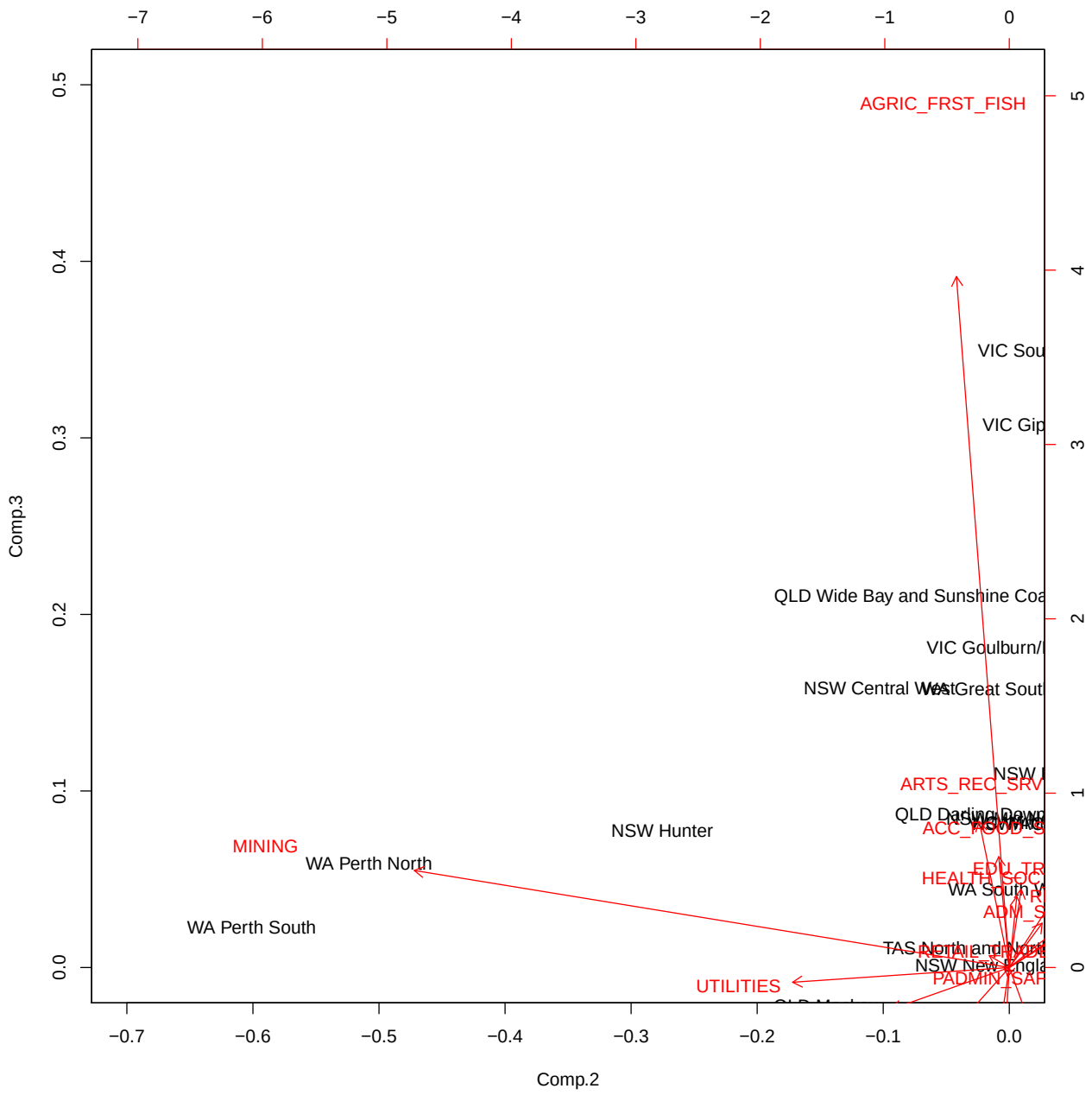
While the 1st component provides the most variation explained, it is difficult to visualise, the 2nd and 3rd components provide a separate axes which permits some simplication in the visualisation.

```
biplot(df1.prcomp, choices=c(2,3), expand=1)
```

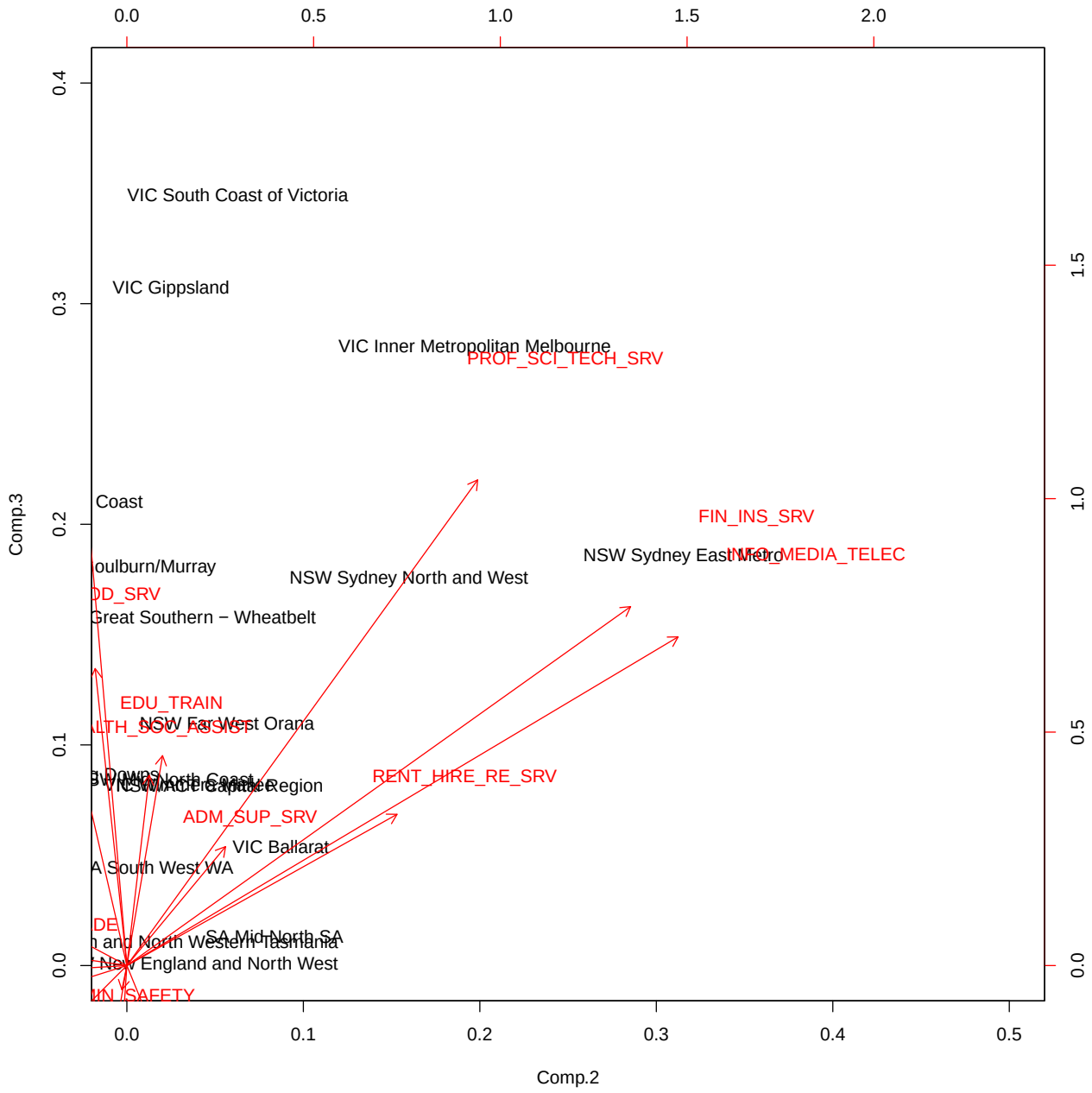


Each segment of the axes is drawn separately in the series below.

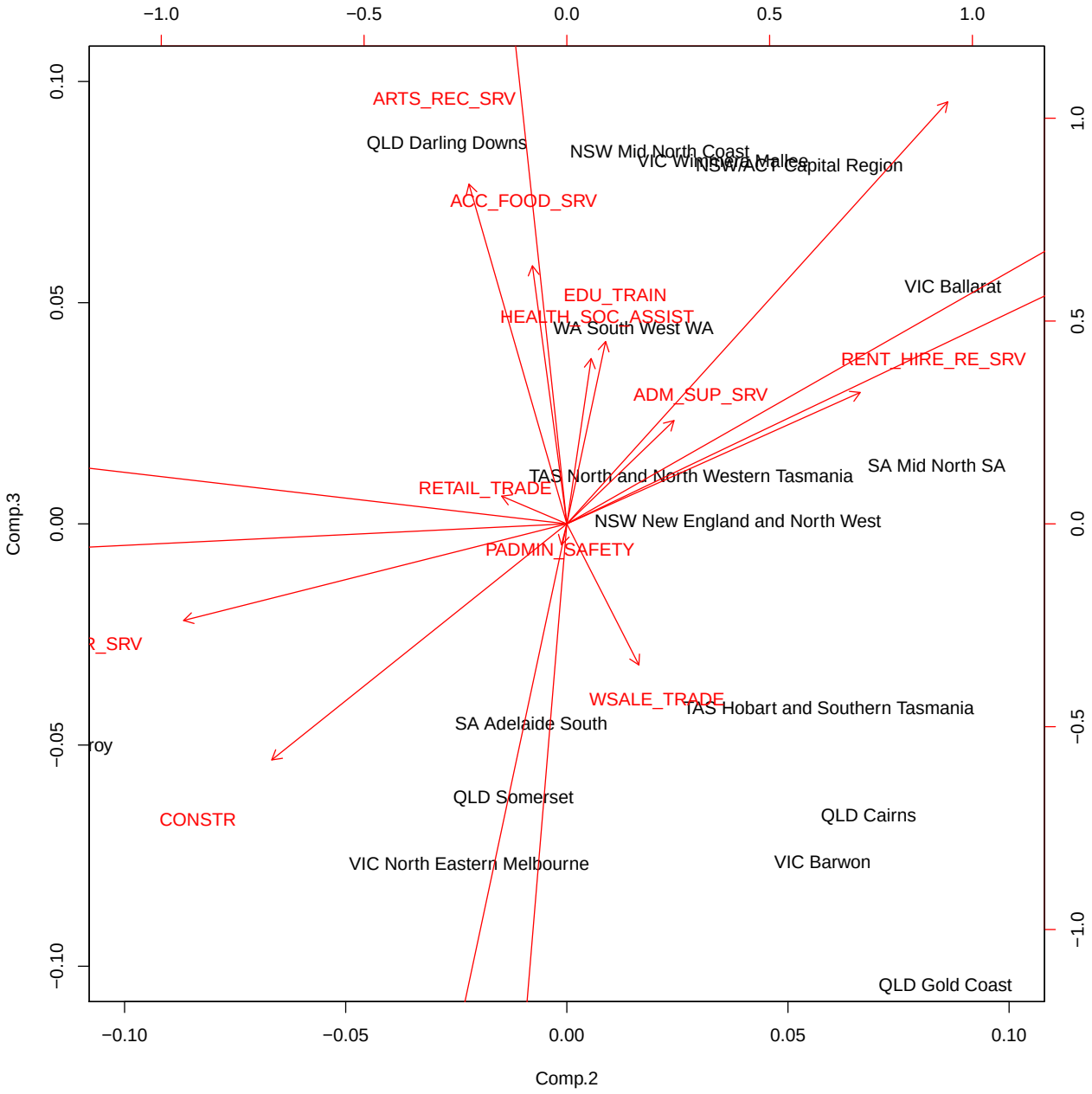
```
biplot(df1.prcomp, choices=c(2,3), expand=1.4, xlim=c(-0.7, 0.0), ylim=c(0.0, 0.5))
```



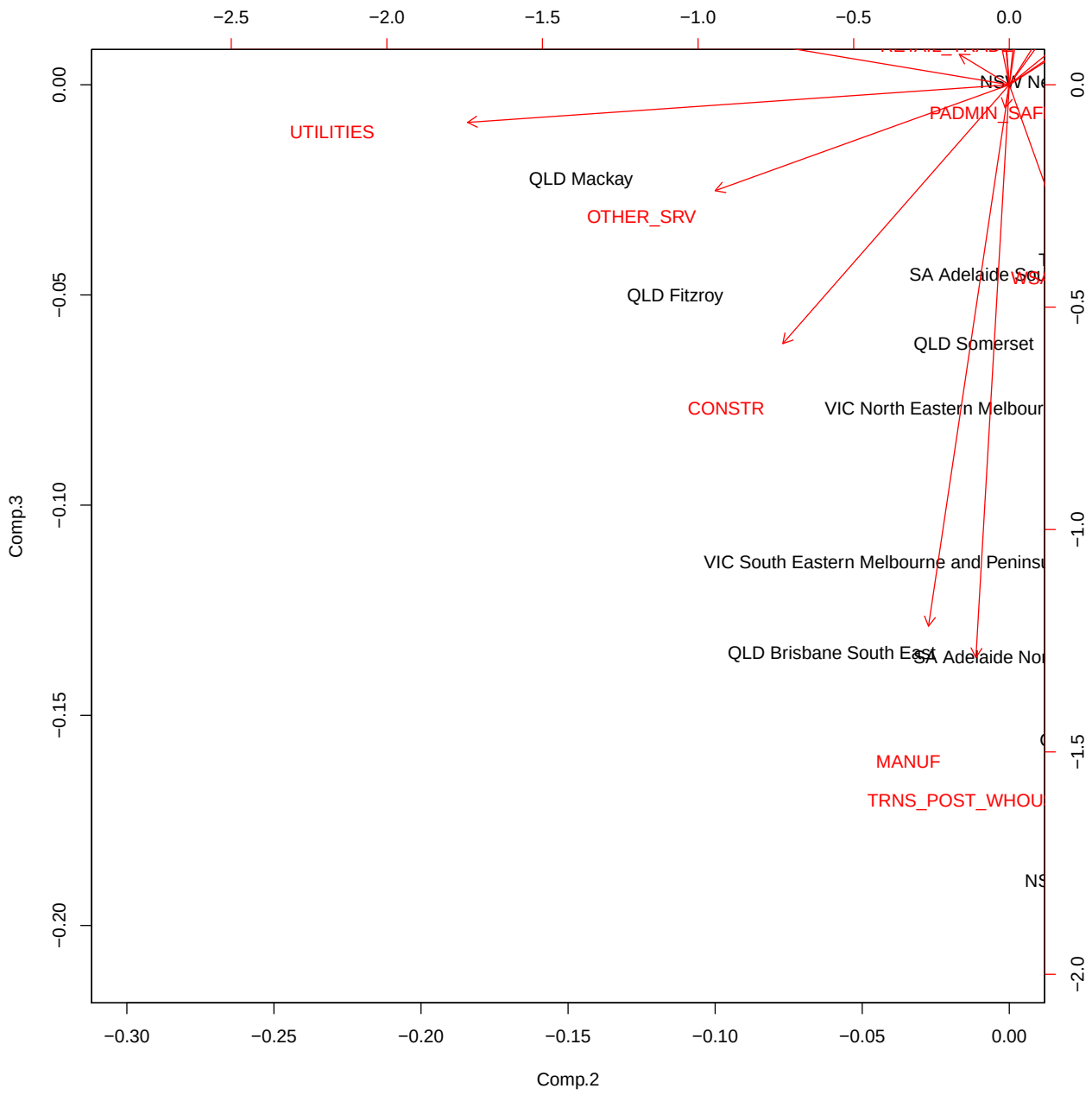

```
biplot(df1.prcomp, choices=c(2,3), expand=3, xlim=c(0.0, 0.5), ylim=c(0.0, 0.4))
```



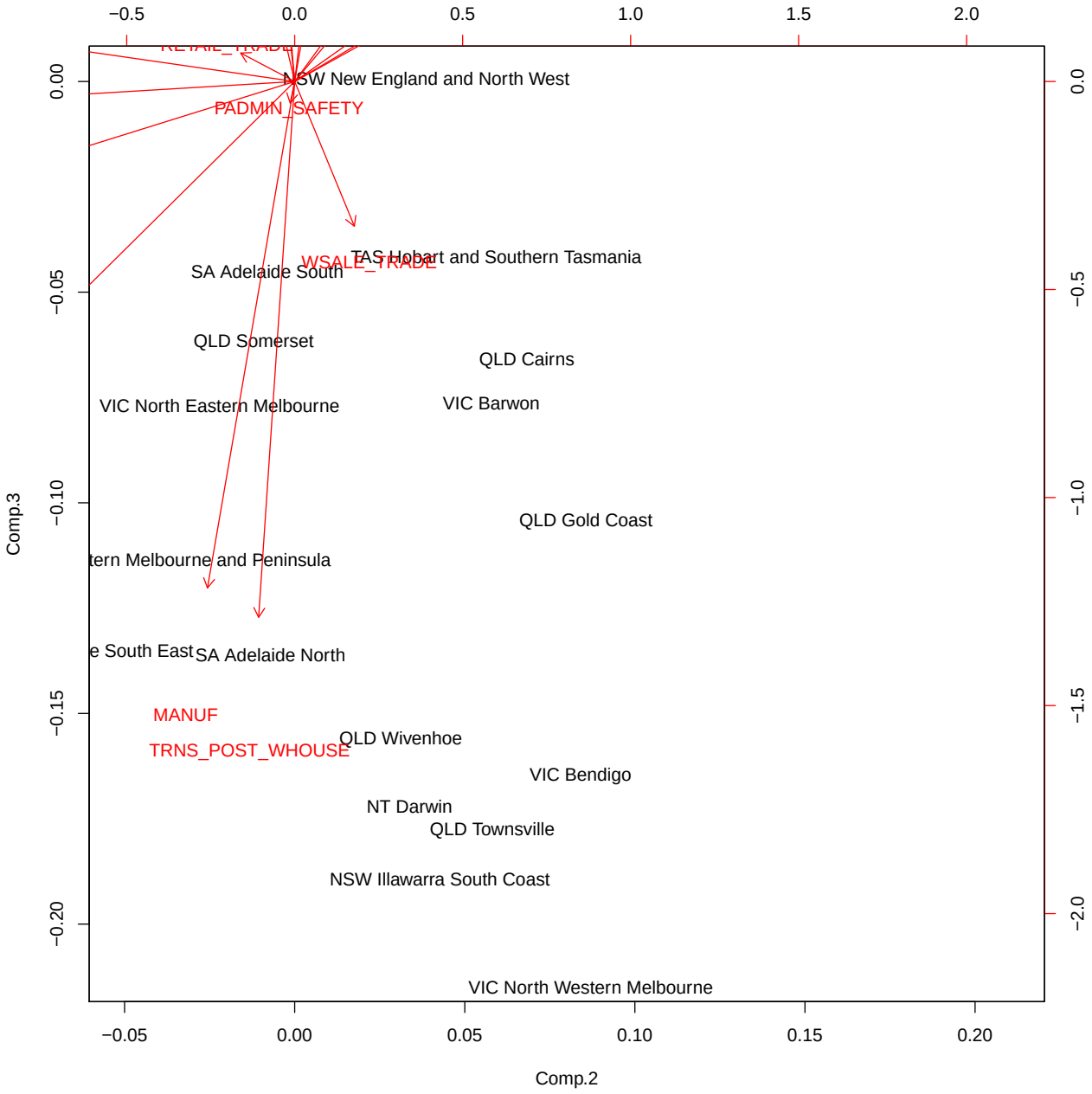

```
biplot(df1.prcomp, choices=c(2,3), expand=1.3, xlim=c(-0.1, 0.1), ylim=c(-0.1, 0.1))
```




```
biplot(df1.prcomp, choices=c(2,3), expand=1.5, xlim=c(-0.3, 0.0), ylim=c(-0.21, 0.0))
```




```
biplot(df1.prcomp, choices=c(2,3), expand=1.4, xlim=c(-0.05, 0.21), ylim=c(-0.21, 0.0))
```


−0.5
0.0
0.5
1.0
1.5
2.0
NSW New England and North West
PADMIN_SAFETY
TAS Hobart and Southern Tasmania
WSALE_TRADE
SA Adelaide South
QLD Somerset
QLD Cairns
VIC North Eastern Melbourne
VIC Barwon
QLD Gold Coast
tern Melbourne and Peninsula
e South East
SA Adelaide North
MANUF
TRNS_POST_WHOUSE
QLD Wivenhoe
VIC Bendigo
NT Darwin
QLD Townsville
NSW Illawarra South Coast
VIC North Western Melbourne
0.00
−0.05
−0.10
−0.15
−0.20
0.0
−0.5
−1.0
−1.5
−2.0
Comp.3
−0.05
0.00
0.05
0.10
0.15
0.20
Comp.2