36-315: Statistical Graphics and Visualization

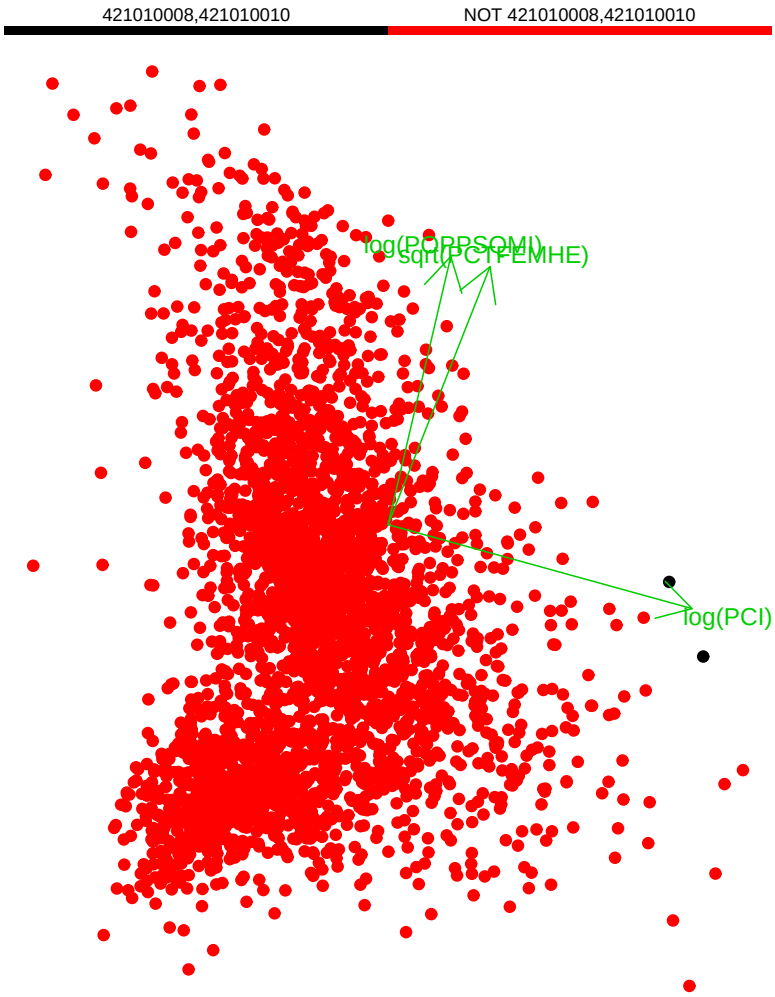
Handout 23
Date: April 16, 2003

Visualizing high-dimensional data through projections

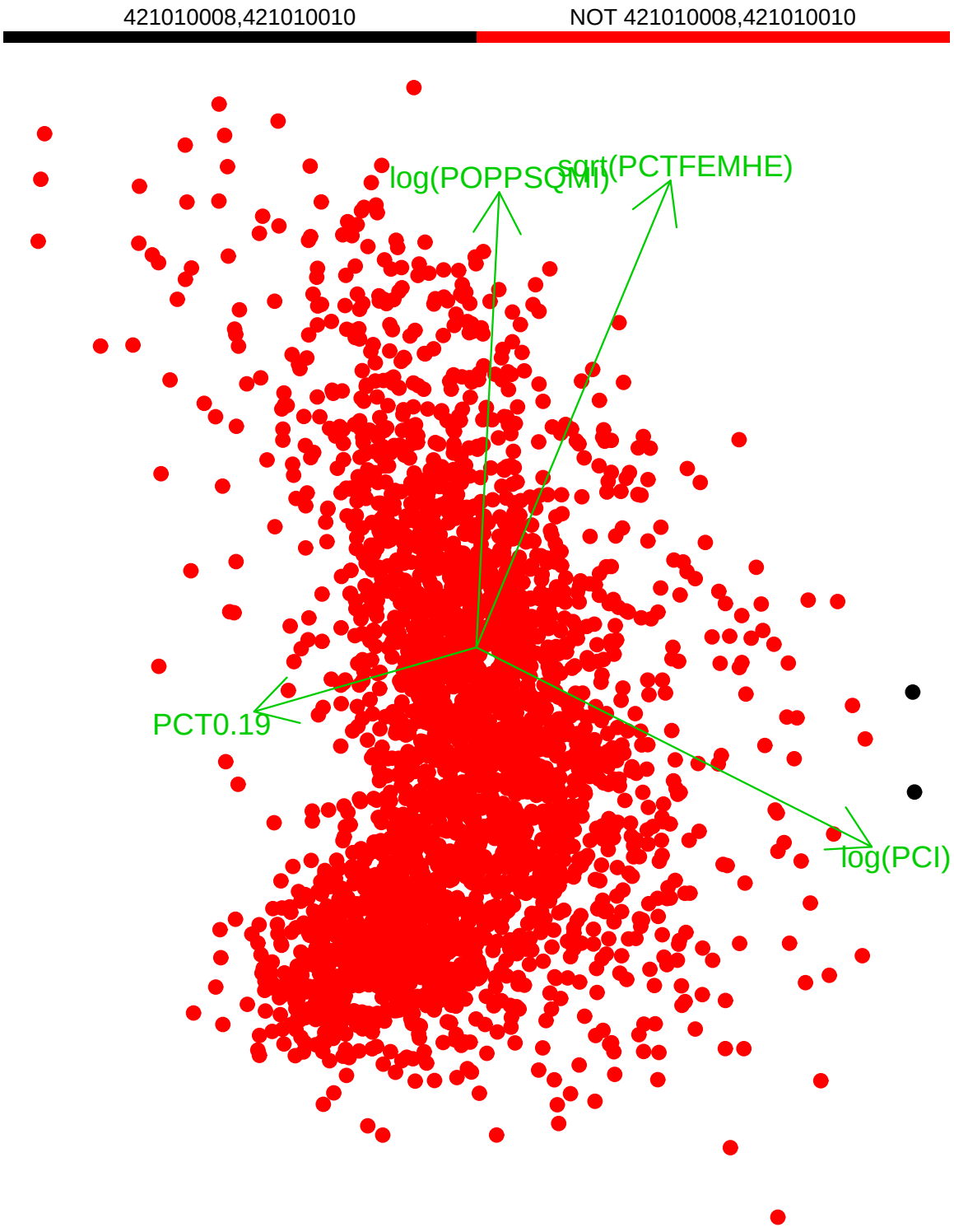
Discriminative projection—Separating two or more specific groups.

Regression projection—Separating all values of a response variable.

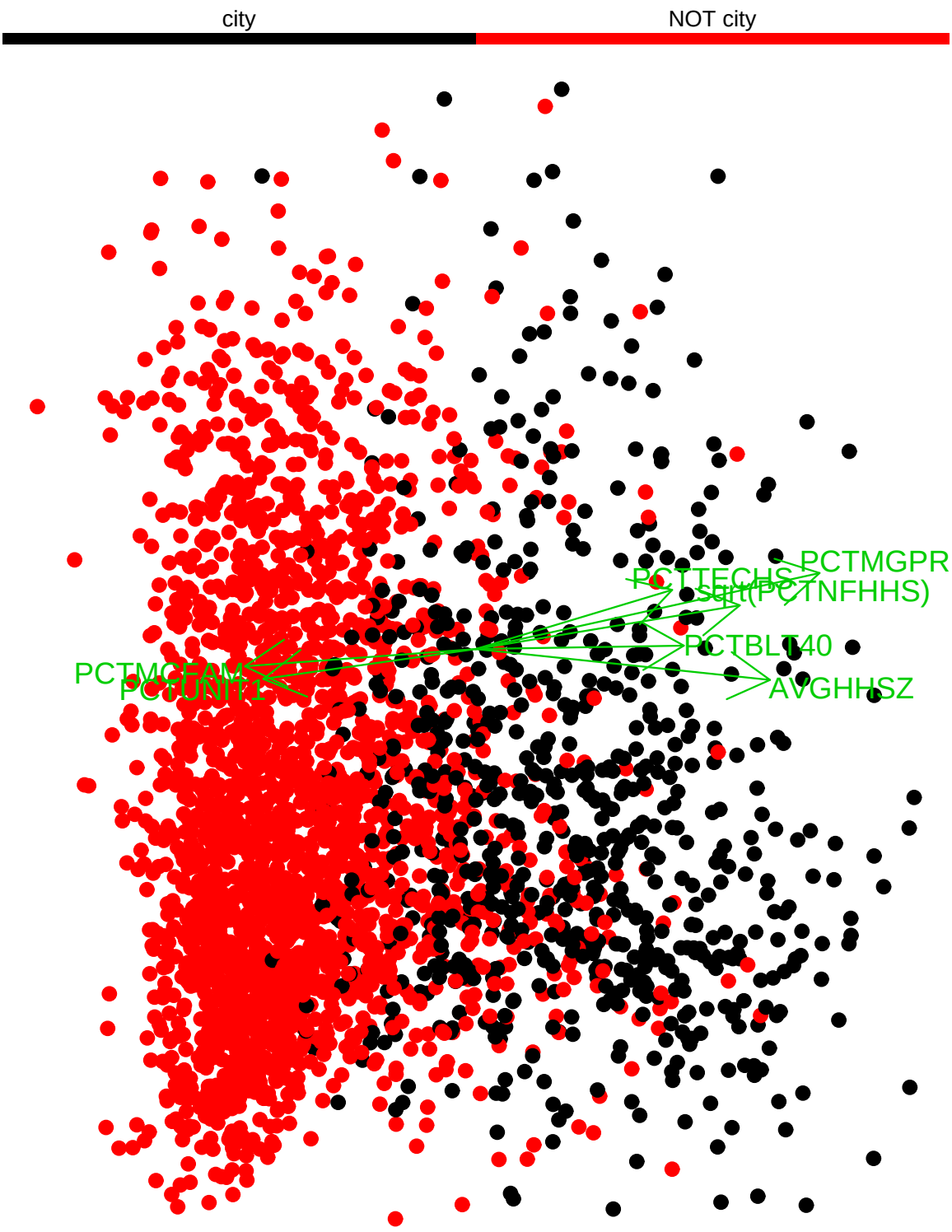
PCA projection—Spreading out all of the data as much as possible.



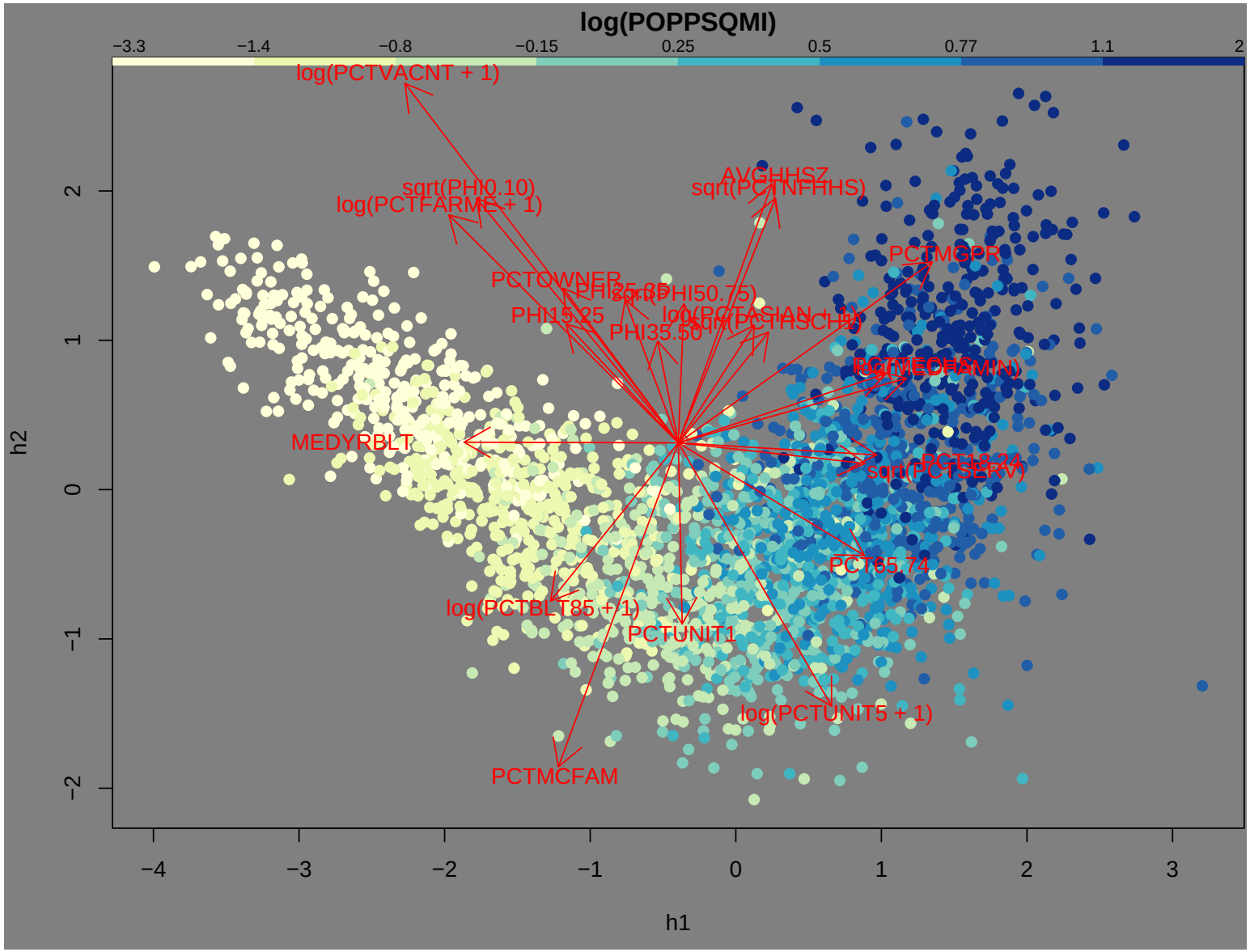
Discriminative projection for outliers (3 dimensions)



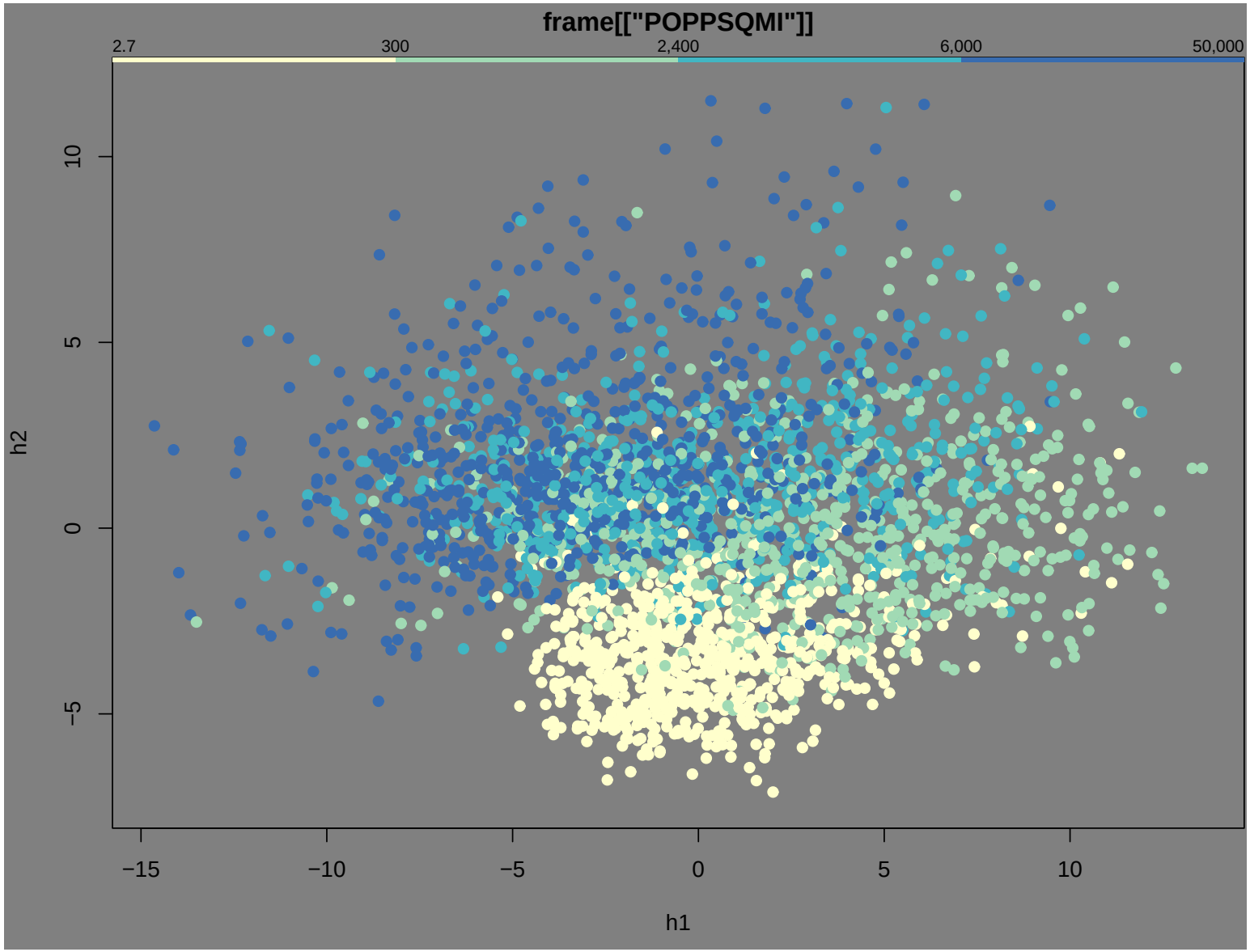
Discriminative projection for outliers (4 dimensions)



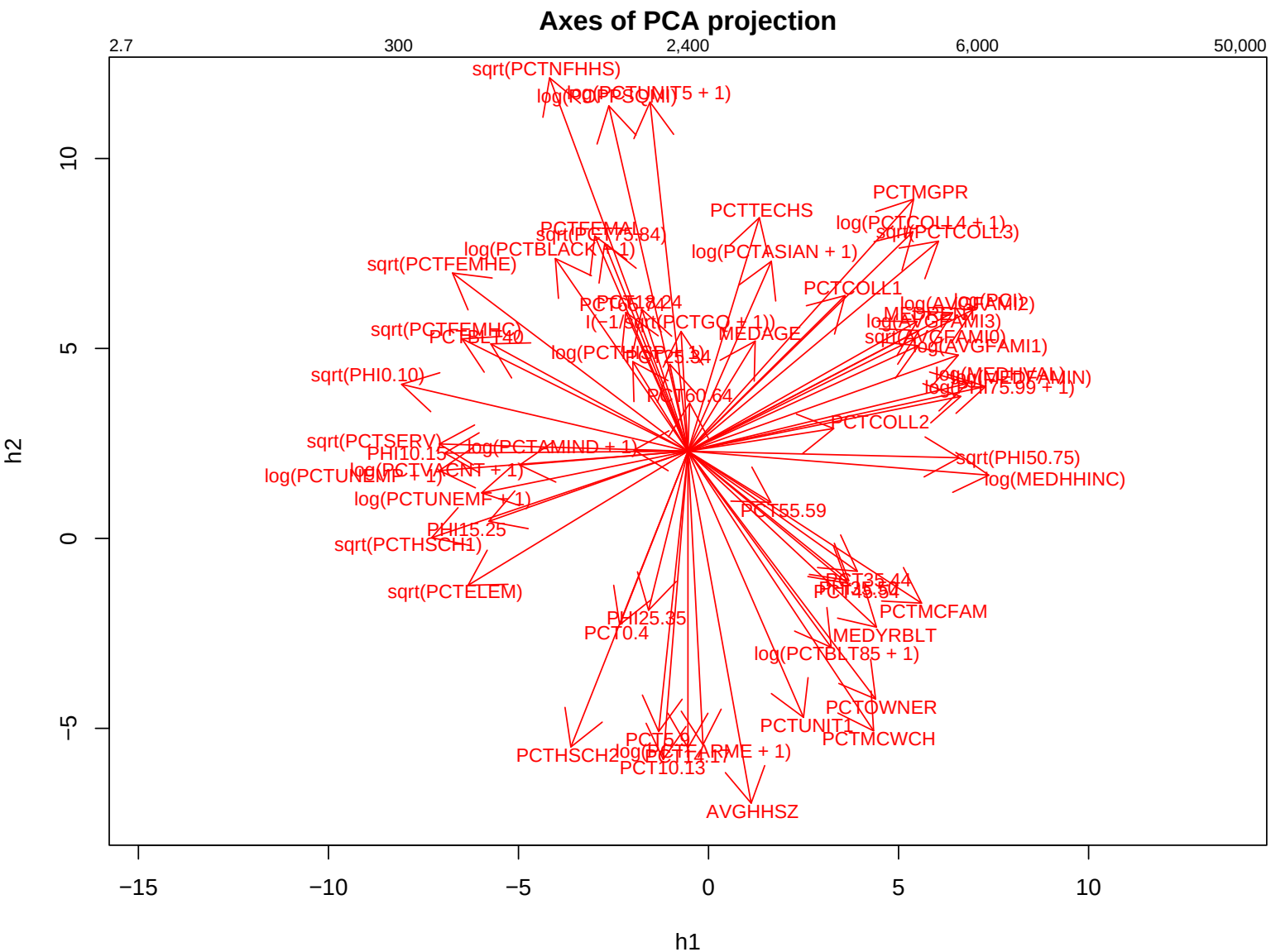
Discriminative projection of all "city" tracts (62 dimensions)

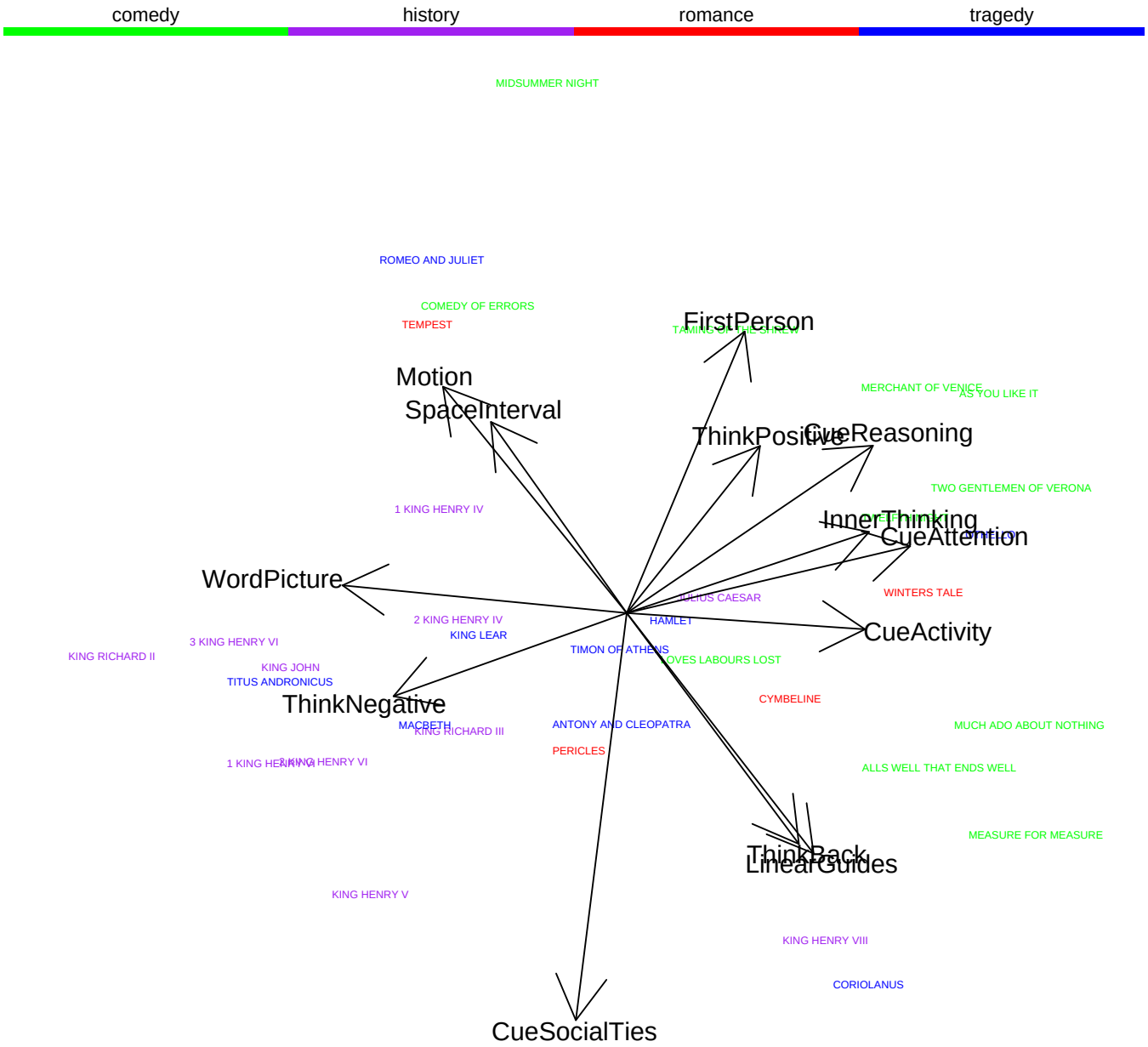


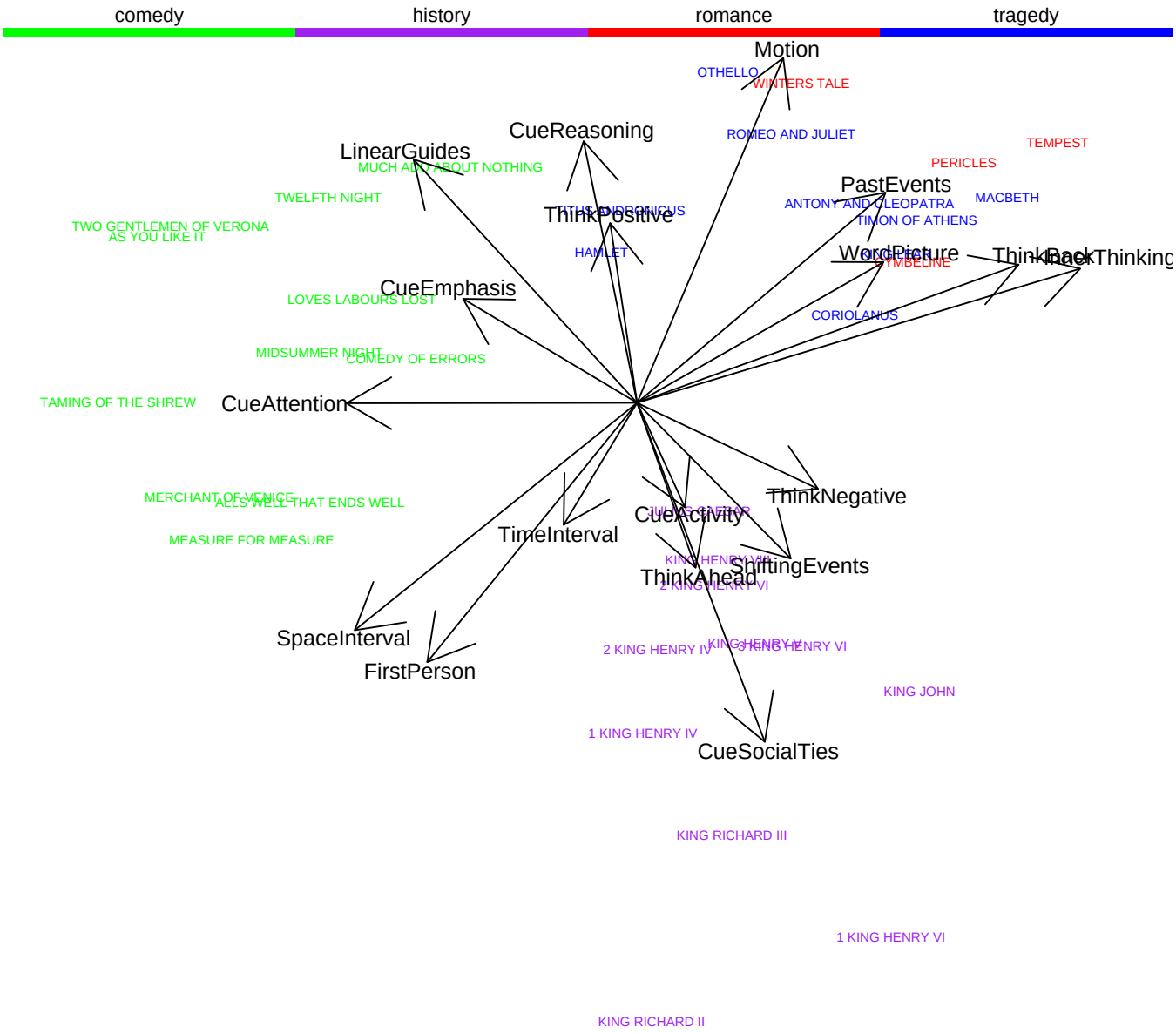
Regression projection for POPPSQMI



PCA projection of all 3,000 tracts in 62 dimensions


Axes of PCA projection
2.7
300
2,400
6,000
50,000
h1
h2
−15
−10
−5
0
5
10
sqrt(PCTNFHHS)
log(PCTUNIT5 + 1)
PCTMGPR
PCTTECHS
log(PCTCOLL4 + 1)
sqrt(PCTCOLL3)
log(PCTASIAN + 1)
PCTFEMAL
sqrt(PCT75.84)
log(PCTBLACK + 1)
sqrt(PCTFEMHE)
PCTCOLL1
log(AVGFAMI2)
log(AVGFAMI3)
sqrt(AVGFAMI0)
log(AVGFAMI1)
MEDAGE
sqrt(PCTFEMHC)
PCTBLT40
sqrt(PHI0.10)
PCT60.64
PCTCOLL2
sqrt(PCTSERV)
PHI10.15
log(PCTAMIND + 1)
log(PCTVACNT + 1)
log(PCTUNEMF + 1)
sqrt(PHI50.75)
log(MEDHHINC)
PCT55.59
PHI15.25
sqrt(PCTHSCH1)
sqrt(PCTELEM)
PCT35.44
PCTMCFAM
PHI25.35
PCT0.4
MEDYRBLT
log(PCTBLT85 + 1)
PCTOWNER
PCTUNIT1
PCTMCWCH
PCTHSCH2
PCT10.13
AVGHHSZ


comedy
history
romance
tragedy
MIDSUMMER NIGHT
ROMEO AND JULIET
COMEDY OF ERRORS
TEMPEST
FirstPerson
TAMING OF THE SHREW
Motion
SpaceInterval
MERCHANT OF VENICE
AS YOU LIKE IT
ThinkPositive
CueReasoning
TWO GENTLEMEN OF VERONA
1 KING HENRY IV
InnerThinking
CueAttention
WordPicture
WINTERS TALE
JULIUS CAESAR
2 KING HENRY IV
HAMLET
KING LEAR
CueActivity
3 KING HENRY VI
KING RICHARD II
TIMON OF ATHENS
LOVES LABOURS LOST
KING JOHN
TITUS ANDRONICUS
ThinkNegative
CYMBELINE
MACBETH
ANTONY AND CLEOPATRA
MUCH ADO ABOUT NOTHING
KING RICHARD III
PERICLES
1 KING HENRY VI
2 KING HENRY VI
ALLS WELL THAT ENDS WELL
MEASURE FOR MEASURE
ThinkBack
LinearGuides
KING HENRY V
KING HENRY VIII
CORIOLANUS
CueSocialTies


comedy
history
romance
tragedy
Motion
OTHELLO
WINTERS TALE
CueReasoning
ROMEO AND JULIET
TEMPEST
LinearGuides
PERICLES
MUCH ADO ABOUT NOTHING
PastEvents
TWELFTH NIGHT
ANTONY AND CLEOPATRA
MACBETH
ThinkPositive
TIMON OF ATHENS
TWO GENTLEMEN OF VERONA
AS YOU LIKE IT
HAMLET
WordPicture
ThinkBackThinking
CueEmphasis
LOVES LABOURS LOST
CORIOLANUS
MIDSUMMER NIGHT
COMEDY OF ERRORS
TAMING OF THE SHREW
CueAttention
MERCHANT OF VENICE
ALLS WELL THAT ENDS WELL
ThinkNegative
CueActivity
TimeInterval
MEASURE FOR MEASURE
ShiftingEvents
ThinkAhead
SpaceInterval
FirstPerson
2 KING HENRY IV
KING JOHN
1 KING HENRY IV
CueSocialTies
KING RICHARD III
1 KING HENRY VI
KING RICHARD II

The Shakespeare data is 35 plays in 19 rhetorical dimensions. The data was collected by David Kaufer at CMU.

The first plot is a PCA projection. The second plot is a discriminative projection designed to separate the three largest categories. Romances were excluded from the separation, in order to see where they fell naturally.