80k terms

all
nonprofits

kaiser — a particlar nonprofit

"comuniti" — a particular word

take subset of rows
and columns

~5k terms (exact # not important)

10K
"large"
nonprofits

want to reduce this from
5k dimensions to 10
dimensions
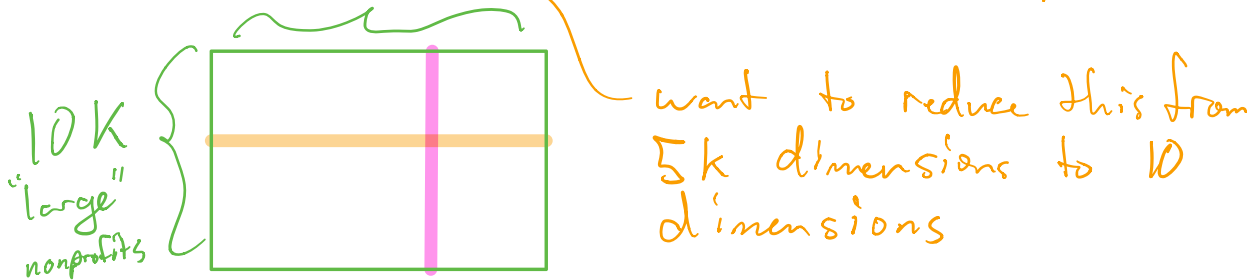
transpose (so columns are dimensions,
as expected by software)

kaiser

10K nonprofits

5k
terms
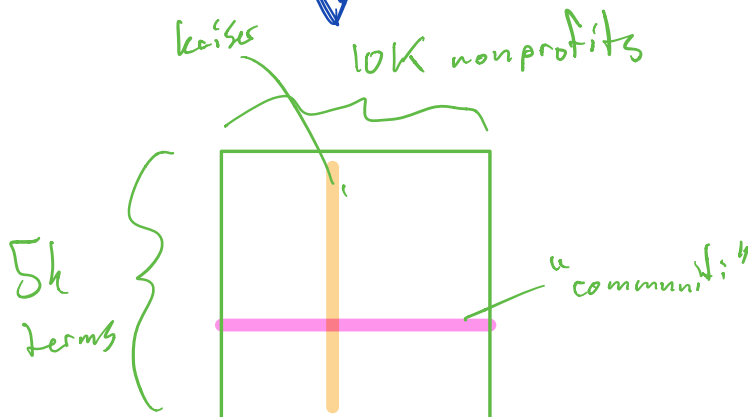
"communiti"
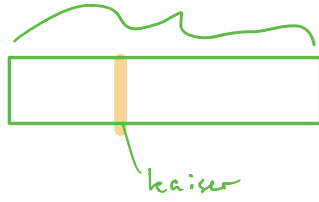
fit PCA, and use it
~~~ transform data

to ~~view~~ ~~~~ ~~~~

10K nonprofits

10 transformed dimensions

kaiser

"communfi" is gone

Kmeans, find 3 clusters

assignments

one for
Kaiser

$$[1, 1, 2, 1, 3, | ... , 3, 1]$$

10K assignments to clusters
(integers 1, 2, 3)