The best number of dimensions for PCA is dataset dependent (1) Branched linear manifold (2) Branched random walk 1.00 Sanity Distance 0.75 Delta method VST  $\log(x+1/(4\alpha))$  $a\cosh(2\alpha x + 1)$ 0.50 log(x+1)/log(CPM + 1) 0.25 السيايي \_\_\_\_\_ 1.00 Sanity Distance Rand. Quant. Resid Mean KNN Recall 0.75 Residual VST Pearson Resid 0.50 0.25 1.00 Sanity Distance 0.75 Sanity MAP Sanity 0.50 0.25 0.00 2 5 10 20 400800 full 2 400800 full 5 50 100 10 20 50 100 data data **#PCA Dimensions #PCA Dimensions**