PRECISION HEALTH
ANALYSIS BOOTCAMP

Omics data exploration

Amrit Singh, PhD

Department of Anesthesiology, Pharmacology and Therapeutics, UBC

Centre for Heart Lung Innovation

August 05, 2022 | 09:00-11:00

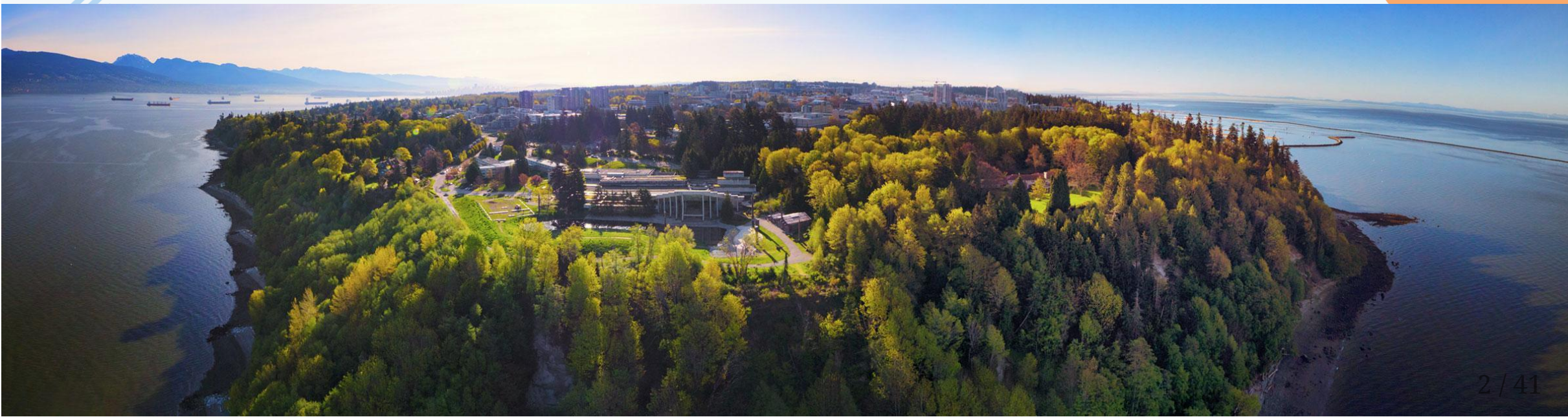
lab

code

asingh_22g

PRECISION HEALTH
ANALYSIS BOOTCAMP

We would like to begin by acknowledging that the land on which we gather is the traditional, ancestral, and unceded territory of the xwməθkwəy̓əm (Musqueam) People.



PRECISION HEALTH
ANALYSIS BOOTCAMP

Copyright Information

Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

This is a human-readable summary of (and not a substitute for) the license. Disclaimer.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Read more here:
https://creativecommons.org/licenses/by-sa/4.0/

Learning outcomes

1. Appreciate the importance exploratory data analysis (EDA) of omics data as a pre-requiste for any research study

2. Use R/Python to apply standard EDA techniques to omics data to explain patterns in data

3. Differentiate between dimension reduction techniques and know when to use what.

RStudio setup

on local machine

- install docker

```
docker --version
docker login
```

build docker image and push to DockerHub

- update DockerHub username in Makefile

```
git clone https://github.com/Phillip-a-richmond/Precis
cd Workshops/omics_exploration
make build
make push
```

instructions are for Mac

on sockeye

```
ssh <cwl>@sockeye.arc.ubc.ca
mkdir -p /scratch/tr-precisionhealth-1/Workshops/Stude
cp -R /project/tr-precisionhealth-1/PrecisionHealthVir
```

pull docker image

```
sh get_rstudio.sh
```

- update email in job script

```
qsub run_rstudio.sh
```

J Clin Oncol. 2007 Feb 10;25(5):517-25.

An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer.

Dressman HK[1], Berchuck A, Chan G, Zhai J, Bild A, Sayer R, Cragun J, Clarke J, Whitaker RS, Li L, Gray J, Marks J, Ginsburg GS, Potti A, West M, Nevins JR, Lancaster JM.

Nat Genet. 2007 Feb;39(2):226-31. Epub 2007 Jan 7.

Common genetic variants account for differences in gene expression among ethnic groups.

Spielman RS[1], Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG.

Science. 2010 Jul 1;2010. doi: 10.1126/science.1190532. Epub 2010 Jul 1.

Genetic signatures of exceptional longevity in humans.

Sebastiani P[1], Solovieff N, Puca A, Hartley SW, Melista E, Andersen S, Dworkis DA, Wilk JB, Myers RH, Steinberg MH, Montano M, Baldwin CT, Perls TT.

RETRACTED ARTICLE

See: Retraction Notice

J Clin Oncol. 2008 Mar 1;26(7):1186-7; author reply 1187-8. doi: 10.1200/JCO.2007.15.1951.

Run batch effects potentially compromise the usefulness of genomic signatures for ovarian cancer.

Baggerly KA, Coombes KR, Neeley ES.

J Clin Oncol. 2007 Feb 10;25(5):517-25.

An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer.

Dressman HK[1], Berchuck A, Chan G, Zhai J, Bild A, Sayer R, Cragun J, Clarke J, Whitaker RS, Li L, Gray J, Marks J, Ginsburg GS, Potti A, West M, Nevins JR, Lancaster JM.

Nat Genet. 2007 Jul;39(7):807-8; author reply 808-9.

On the design and analysis of gene expression studies in human populations.

Akey JM, Biswas S, Leek JT, Storey JD.

Nat Genet. 2007 Feb;39(2):226-31. Epub 2007 Jan 7.

Common genetic variants account for differences in gene expression among ethnic groups.

Spielman RS[1], Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG.

RETRACTED ARTICLE

See: Retraction Notice

Science. 2010 Nov 12;330(6006):912. doi: 10.1126/science.330.6006.912-b.

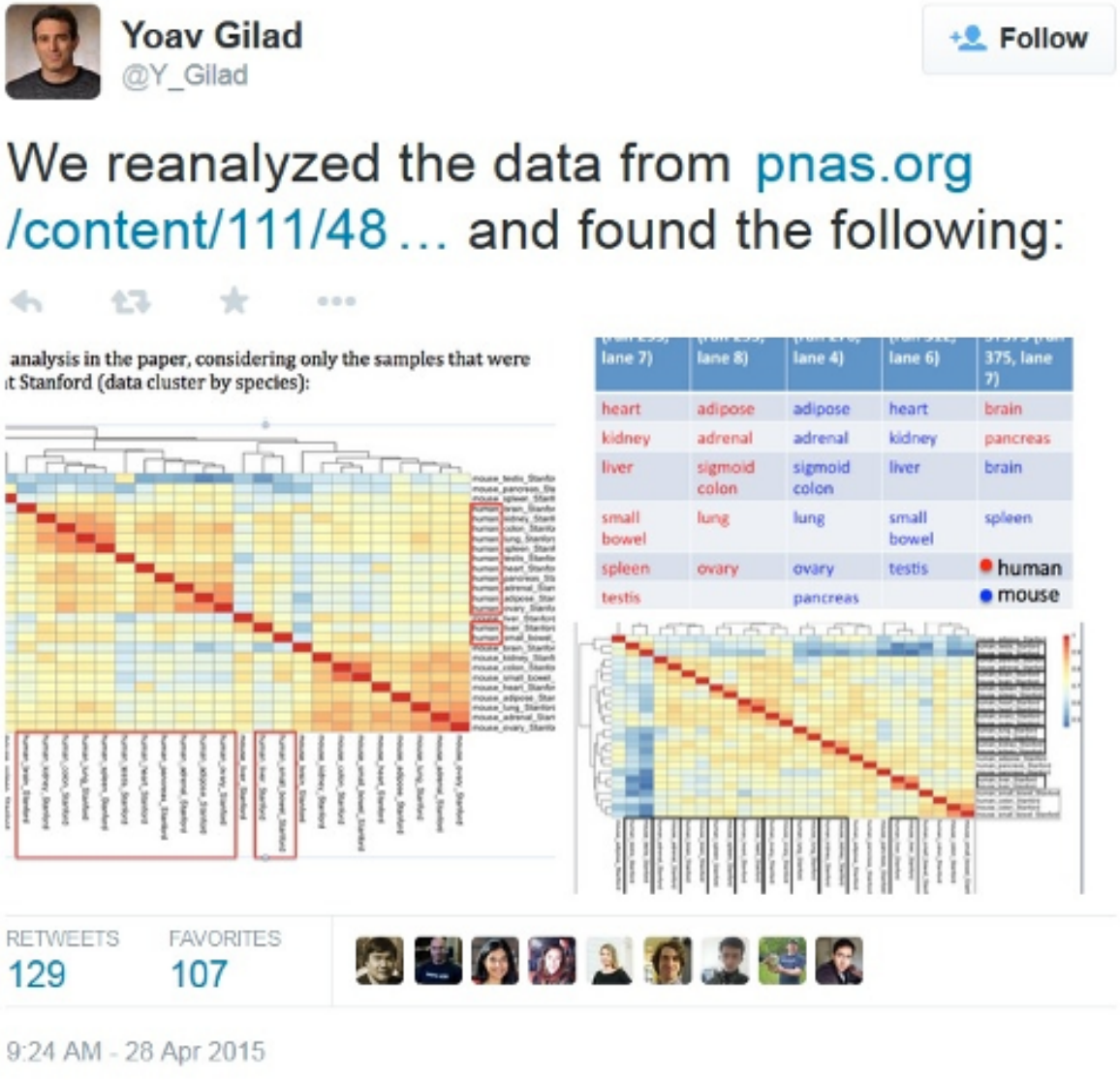
Editorial expression of concern.

Alberts B.

Science. 2010 Jul 1;2010. doi: 10.1126/science.1190532. Epub 2010 Jul 1.

Genetic signatures of exceptional longevity in humans.

Sebastiani P[1], Solovieff N, Puca A, Hartley SW, Melista E, Andersen S, Dworkis DA, Wilk JB, Myers RH, Steinberg MH, Montano M, Baldwin CT, Perls TT.



Nature 521, 397 (28 May 2015)

F1000Research 2015, 4:121

Gene names

Ziemann *et al. Genome Biology* (2016) 17:177
DOI 10.1186/s13059-016-1044-7

Genome Biology

COMMENT Open Access

Gene name errors are widespread in the scientific literature

Mark Ziemann[1], Yotam Eren[1,2] and Assam El-Osta[1,3*]

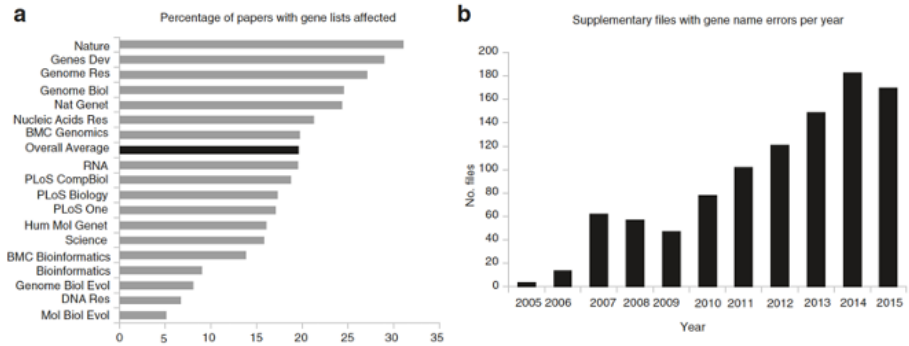
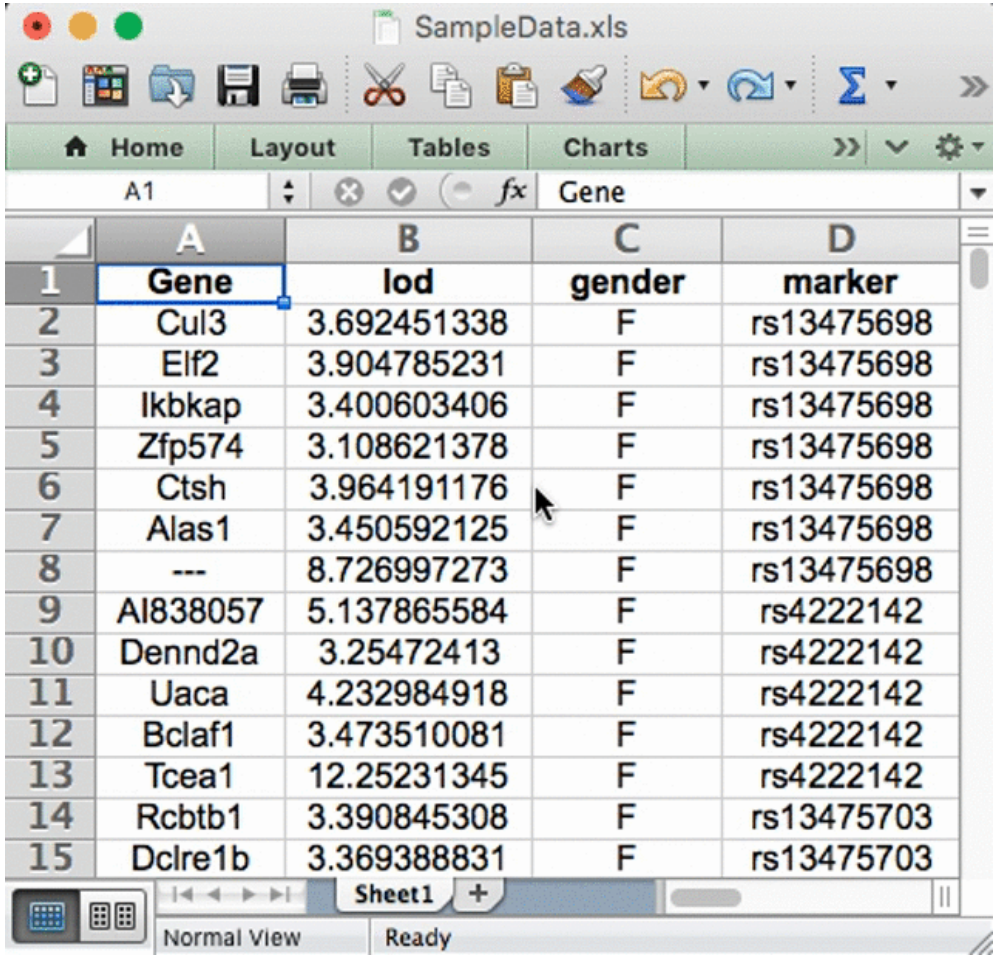


Fig. 1 Prevalence of gene name errors in supplementary Excel files. **a** Percentage of published papers with supplementary gene lists in Excel files affected by gene name errors. **b** Increase in gene name errors by year



	A	B	C	D
1	Gene	lod	gender	marker
2	Cul3	3.692451338	F	rs13475698
3	Elf2	3.904785231	F	rs13475698
4	Ikbkap	3.400603406	F	rs13475698
5	Zfp574	3.108621378	F	rs13475698
6	Ctsh	3.964191176	F	rs13475698
7	Alas1	3.450592125	F	rs13475698
8	---	8.726997273	F	rs13475698
9	AI838057	5.137865584	F	rs4222142
10	Dennd2a	3.25472413	F	rs4222142
11	Uaca	4.232984918	F	rs4222142
12	Bclaf1	3.473510081	F	rs4222142
13	Tcea1	12.25231345	F	rs4222142
14	Rcbtb1	3.390845308	F	rs13475703
15	Dclre1b	3.369388831	F	rs13475703

What does processed omics data look like?

	RTN2	NDRG2	CCDC113	FAM63A	ACADS	GMDS	HLA-H	SEMA4A	ETS2	LIMD2	NME3	ZEB1	CDCP1	GIYD2	RTKN2
A0U4	2.2186948	7.385044	4.8601805	3.800083	3.4547735	5.455150	7.815700	7.481899	6.076811	5.430468	5.2380449	4.100027	6.256089	4.484376	2.8195301
A04T	1.0909136	7.436516	3.6648151	5.139812	2.4096918	5.173432	6.717064	7.111433	5.103906	4.187376	4.4212733	3.299597	6.613511	2.956674	2.7540724
A04D	3.2641428	9.645246	3.9300116	5.762041	3.0452661	5.038067	4.364824	7.004776	4.696697	5.462888	4.3580441	2.612997	5.932216	3.480355	4.9640939
A18P	3.4750011	4.823776	4.3747137	5.640225	3.0702128	3.582233	6.824283	6.756933	5.275035	2.993024	4.5993150	5.466487	6.729030	4.256978	5.0483954
A15R	5.8295426	6.122818	4.0586415	6.802337	3.3391137	2.615647	2.957603	6.721796	5.379089	1.559730	5.8103253	4.932753	1.184280	4.747195	1.9019414
A08A	5.3197574	5.670118	4.2134599	6.421754	6.4357258	4.211985	7.336078	6.699267	4.992061	4.062846	7.0127259	4.543252	5.822906	5.302962	1.4923767
A18S	4.8859746	5.190614	5.9812017	6.839739	3.8077087	4.467859	5.178437	6.672839	4.973980	3.764294	5.3857955	4.007190	6.445439	4.217784	0.9838633
A0JL	3.2438595	9.339442	3.4400301	4.850530	3.0060187	5.033427	6.413001	6.662359	5.558211	4.895874	3.1591973	4.464115	3.619750	6.352563	4.1278116
A0FL	6.1428496	10.260165	1.4415576	5.961550	3.8505413	6.642912	7.246420	6.645500	4.672800	4.298330	5.4708339	3.103197	4.511900	4.810870	3.2014550
A0XU	3.3205574	7.341238	4.6208587	4.955763	2.6294123	5.699021	8.452830	6.612863	5.481314	4.172537	3.5895416	3.316380	5.608396	3.981899	3.1792076
A128	1.9570125	3.858092	4.2362158	1.840127	3.2048642	5.958135	7.820970	6.610413	4.279613	5.864948	3.7418811	4.544412	7.869354	4.215449	2.2112487
A0DA	4.7707976	8.748061	4.3054015	5.307480	3.2399091	4.236539	6.909727	6.591109	5.858016	3.766283	4.2593544	4.800017	6.052282	4.006990	2.0696264
A08T	4.4925749	5.924680	3.5655753	6.133809	4.1936049	5.025717	5.546672	6.568790	5.224105	2.716533	5.7761450	5.685821	6.495370	5.305507	3.1441736
A12V	2.8137145	8.556629	2.2430796	4.719815	3.4467922	4.414810	7.652525	6.548990	5.884915	6.035908	5.4373028	4.771115	6.645278	4.732800	3.5446923
A0B3	3.3520618	5.098404	0.5932056	5.217585	3.8851534	5.917886	8.043341	6.532893	6.309117	4.113873	4.7864632	4.257365	7.265190	4.635484	3.5271329
A0E0	0.6743217	4.561989	2.2716035	6.097113	4.3603055	4.424637	4.850381	6.481614	6.421863	3.904824	5.4166787	2.944633	4.290746	5.399474	3.7677046
A140	3.3710825	4.767956	4.2663599	7.680580	4.9585346	2.306479	7.066051	6.481121	4.512836	3.627944	5.6017226	5.687797	2.137567	3.858729	2.9491950

- data pre-processing is specific for each omics data modality and is beyond the scope of this workshop.

What are the characteristics of omics data?

Terminology review

Term	Synonym	Description
Variable	feature, dimension	a measurable quantity that describes an observation's attributes. *e.g.* age, sex, gene, protein abundance, single nucleotide variants, operational taxonomic units, pixels etc.
Observation	sample, observation, assay, array	A single entity belonging to a larger grouping. *e.g.* patients, subjects, participants, cells, mice.
component	factors, latent variable	variables that are not directly observed but computed from mathematical models

Make a fake dataset with some outliers

- 100 samples (observations), 1000 genes (variables)
- add 5 outlier samples

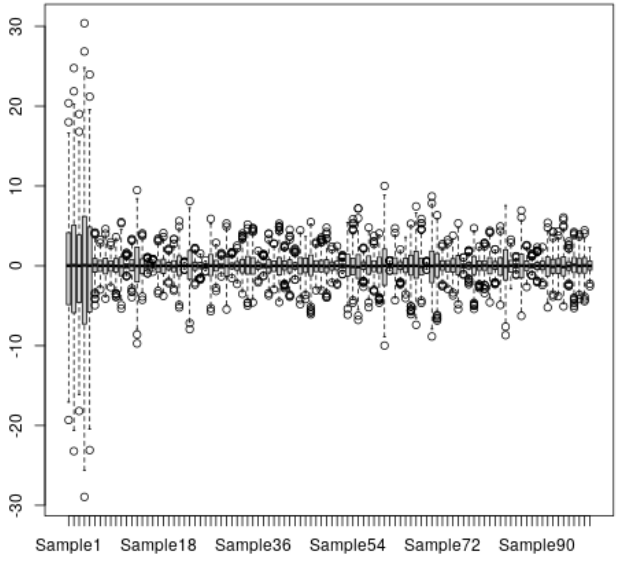
```
##              Gene1      Gene2      Gene3       Gene4     Gene5
## Sample1   8.761927  5.5730852  0.3153502  5.15195428 2.0470396
## Sample2  10.809552  6.5221590  1.0018223  6.24235523 2.5544775
## Sample3   8.071760  5.3596590 -0.1006712  4.81865334 1.8672498
## Sample4  12.974535  8.4508884  0.1229828  7.69270306 3.0149176
## Sample5  10.075766  6.9090601 -0.5050180  6.08531829 2.3128528
## Sample6   1.220478 -1.6066933  4.1766805 -0.04844715 0.4810837
## Sample7   1.661820  0.4868416  1.0486437  0.79384131 0.4351325
## Sample8   2.280169  0.7838800  1.2378511  1.12647839 0.5875122
## Sample9   1.338251  0.6975058  0.3147202  0.73747200 0.3252923
## Sample10  1.838302 -0.8692086  3.6014356  0.42558301 0.5970971
```

- Cholesky Decomposition

How can we identify which samples/variables are outliers?

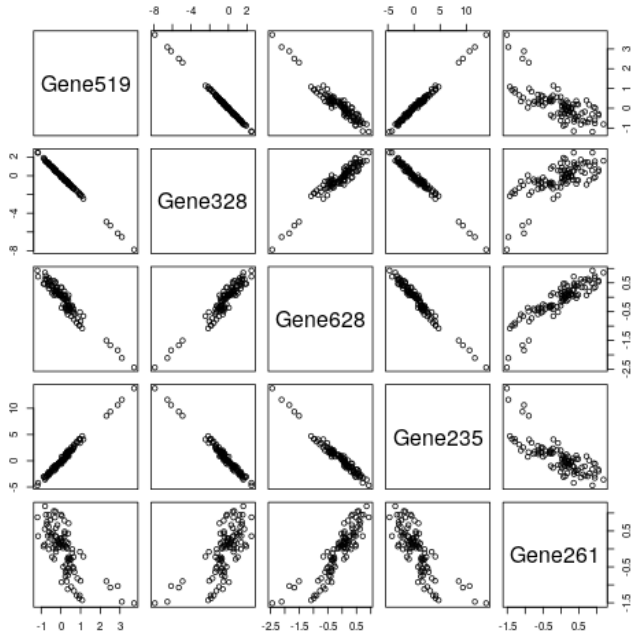
boxplots

```
boxplot(t(X))
```



SPLOM plots (Scatter plot matrix)

```
pairs(X[,sample(ncol(X), 5)])
```



How can we identify which samples/variables are outliers?

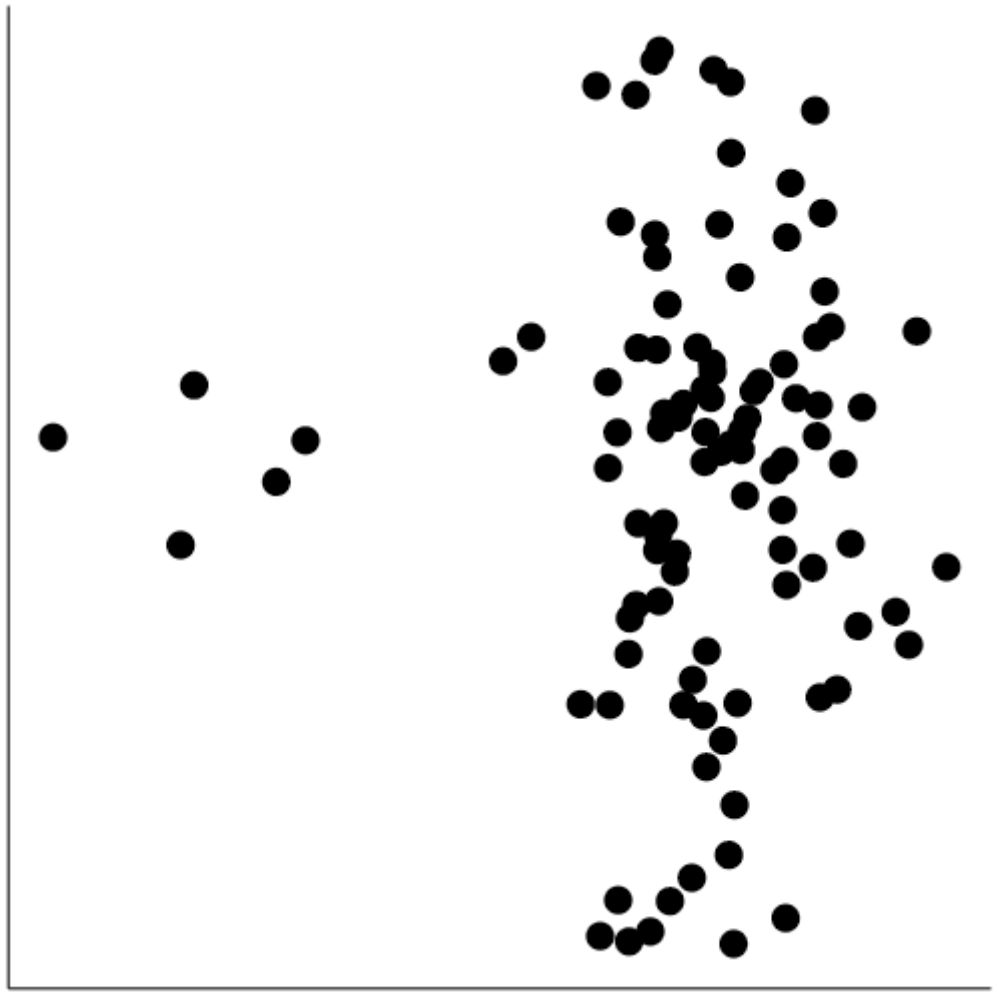
Sample correlation matrix

```
par(mar=c(1,1,1,1))
gplots::heatmap.2(cor(t(X)), margins = c(1,1), trace =
```

other heatmap functions

- ComplexHeatmap
- NMF::aheatmap()
- mixOmics::cim()
- seaborn.heatmap()
- scanpy.pl.heatmap()

How to get something like this?



Dimension reduction (DR) methods

Method	Input Data	Method Class	Nonlinear
PCA	continuous data	unsupervised	
CA	categorical data	unsupervised	
MCA	categorical data	unsupervised	
PCoA (cMDS)	distance matrix	unsupervised	
NMDS	distance matrix	unsupervised	
Isomap	continuous*	unsupervised	✓
Diffusion Map	continuous*	unsupervised	✓
Kernel PCA	continuous*	unsupervised	✓
t-SNE	continuous/distance	unsupervised	✓
Barnes–Hut t-SNE	continuous/distance	unsupervised	✓
LDA	continuous (X and Y)	supervised	
PLS (NIPALS)	continuous (X and Y)	supervised	
NCA	distance matrix	supervised	✓
Bottleneck NN	continuous/categorical	supervised	✓
STATIS	continuous	multidomain	
DiSTATIS	distance matrix	multidomain	

EDUCATION

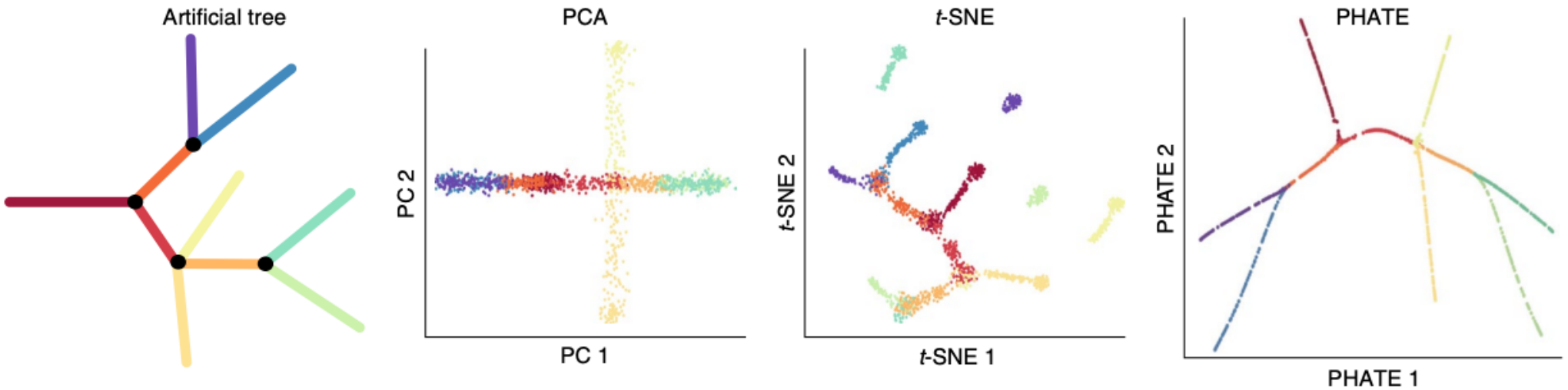
Ten quick tips for effective dimensionality reduction

Lan Huong Nguyen[1], **Susan Holmes**[2]*

1 Institute for Mathematical and Computational Engineering, Stanford University, Stanford, California, United States of America, **2** Department of Statistics, Stanford University, Stanford, California, United States of America

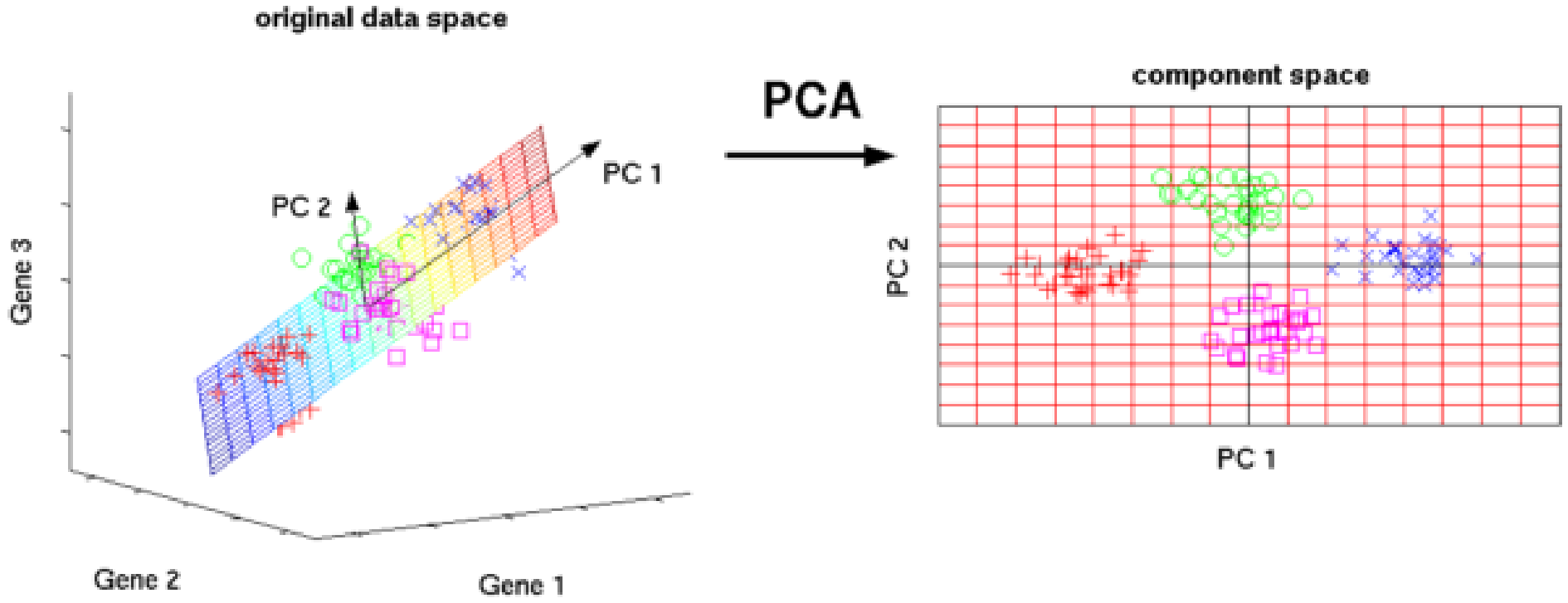
* susan@stat.stanford.edu

1) Choose an appropriate DM method



Nature Biotechnology volume 37, pages1482–1492 (2019)

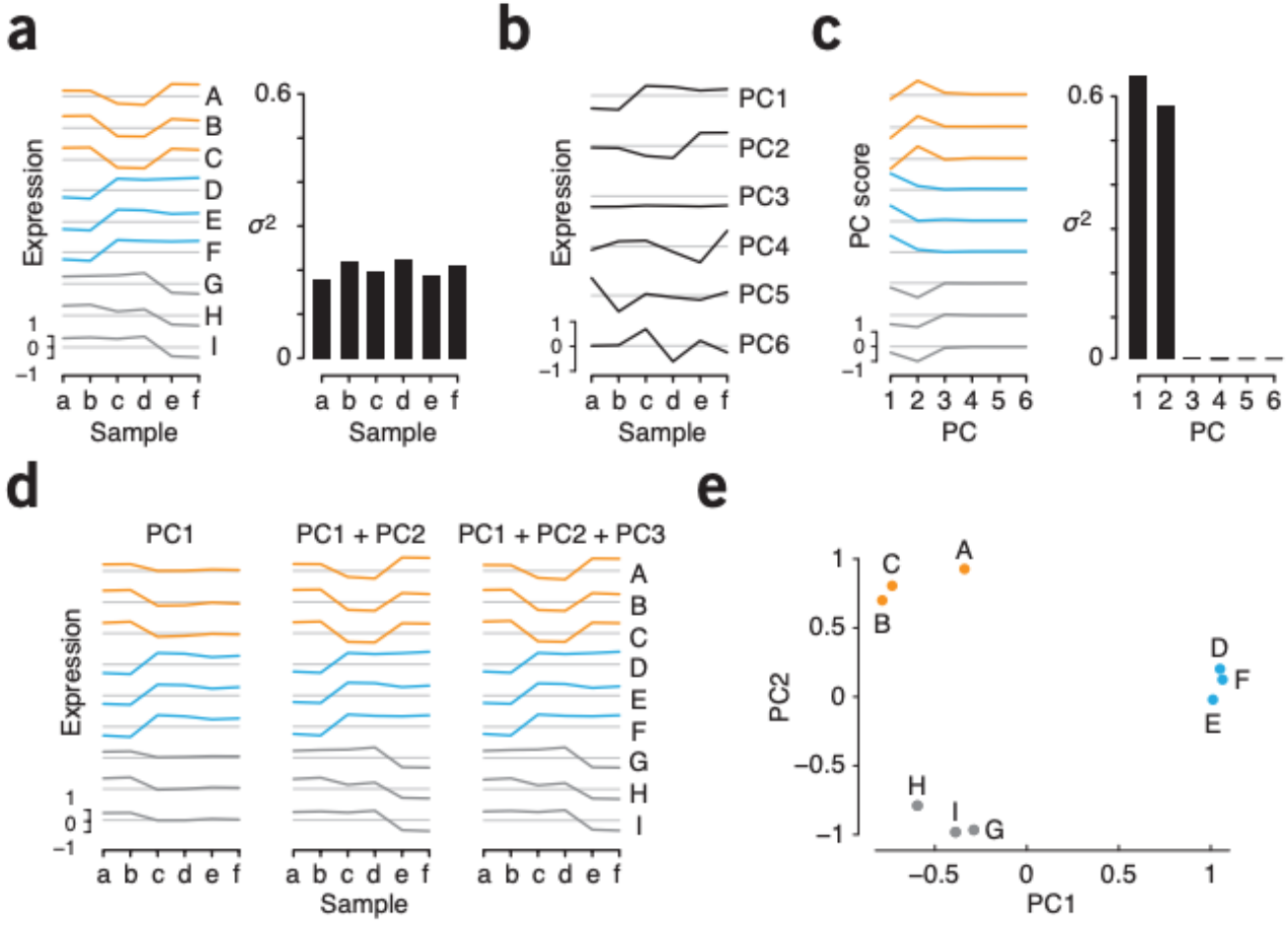
Principal Component Analysis (PCA): $PC_{nxk} = X_{nxp}V_{pxk}$



- transform X which contains 3 variables (Gene 1-3) into PC which contain 2 variables (PC1-2)

Genius blog

Principal Component Analysis (PCA): $PC_{nxk} = X_{nxp}V_{pxk}$



Nature Methods volume 14, pages 641–642 (2017)

Singular Value Decomposition: $X_{nxn} = U_{nxk} D_{kxk} V^T_{kxp}$

```
?prcomp
methods(prcomp)
```

```
## [1] prcomp.default* prcomp.formula*
## see '?methods' for accessing help and source code
```

```
getAnywhere(prcomp.default)
```

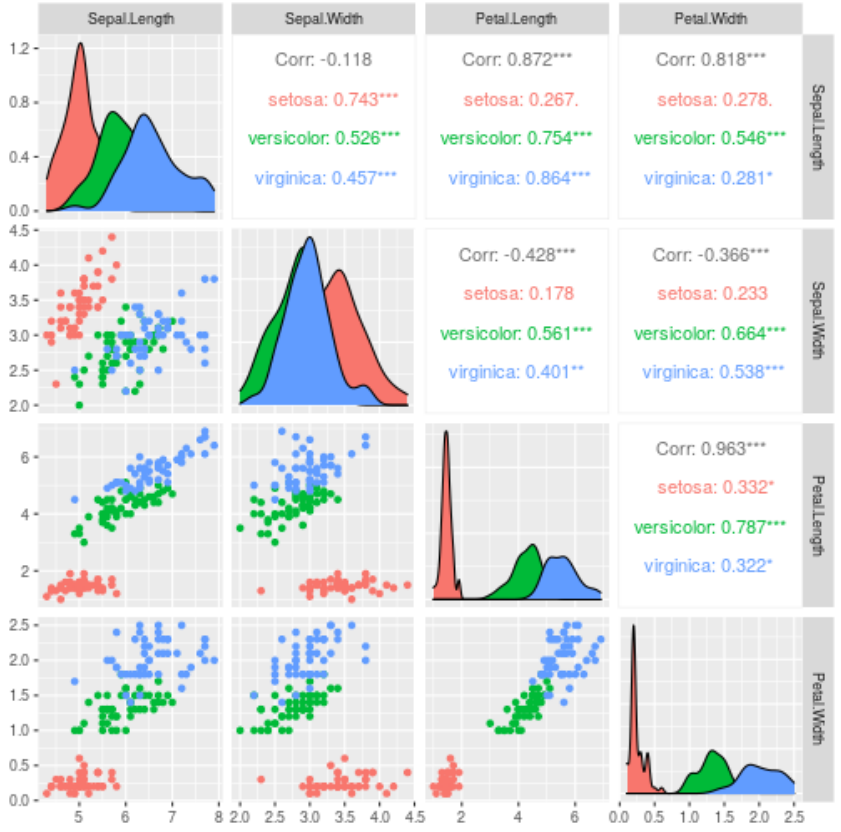
> s <- svd(x, nu = 0, nv = k)
>
> r$x <- x %*% s$v

Properties

- D is a diagnoal matrix containing the variances associated with each PC.
- the proportion of variation explained is a useful measure of how well PCA summarizes your data.
- assumes that data is centered (subtract the mean value of each variable from all values of that variable)
- PCs are uncorrelated

Edgar Anderson's Iris Data

?iris: This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.



Apply PCA to IRIS

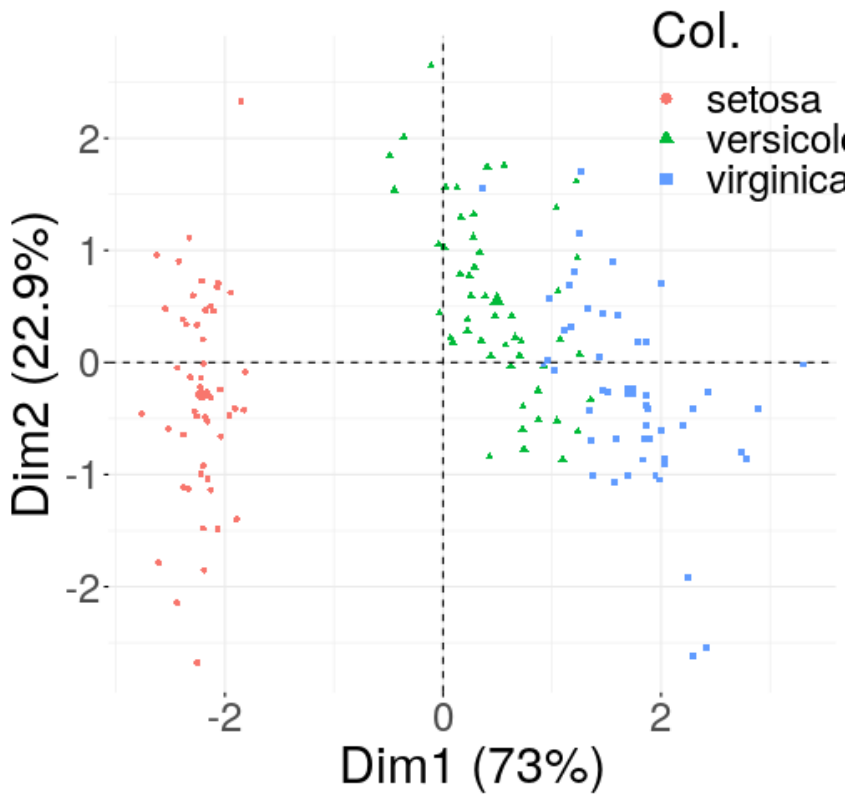
```
X <- iris[, setdiff(colnames(iris), "Species")]
y <- iris$Species

pca.res <- prcomp(X, rank. = 2, scale. = TRUE, center = TRUE)
100*(pca.res$sdev^2)/sum(pca.res$sdev^2)
```

```
## [1] 72.9624454 22.8507618  3.6689219  0.5178709
```

> first 2 PCs explain 95% of the variation in the data. 😮

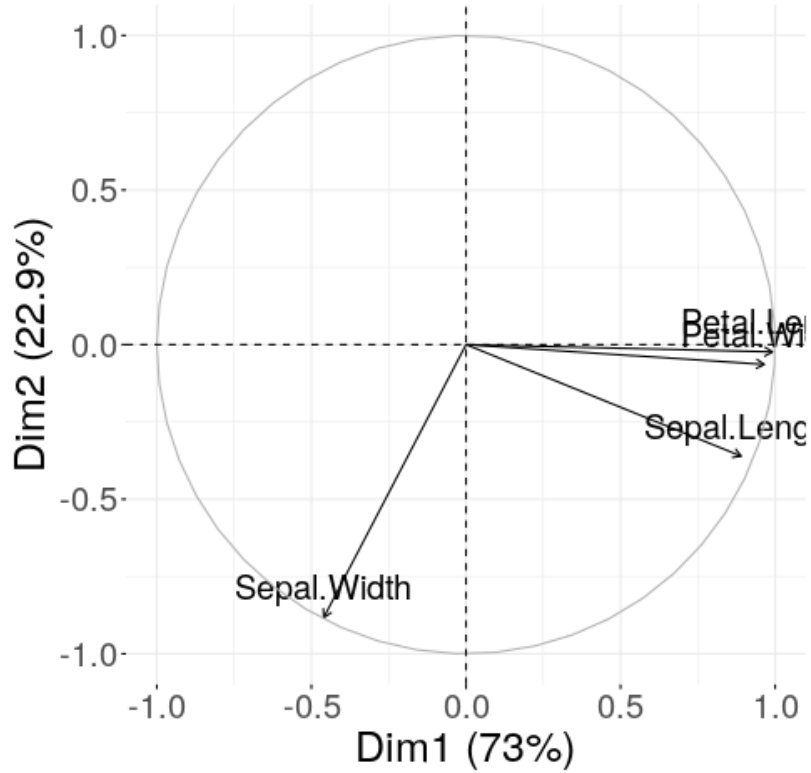
Plot samples: Component plot

```
fviz_pca_ind(pca.res, geom = "point", col.ind = y)  +
  ggtitle("") +
  theme(text = element_text(size = 30)) +
  theme(legend.position = c(0.9, 0.9))
```



Plot variables: Correlation circle

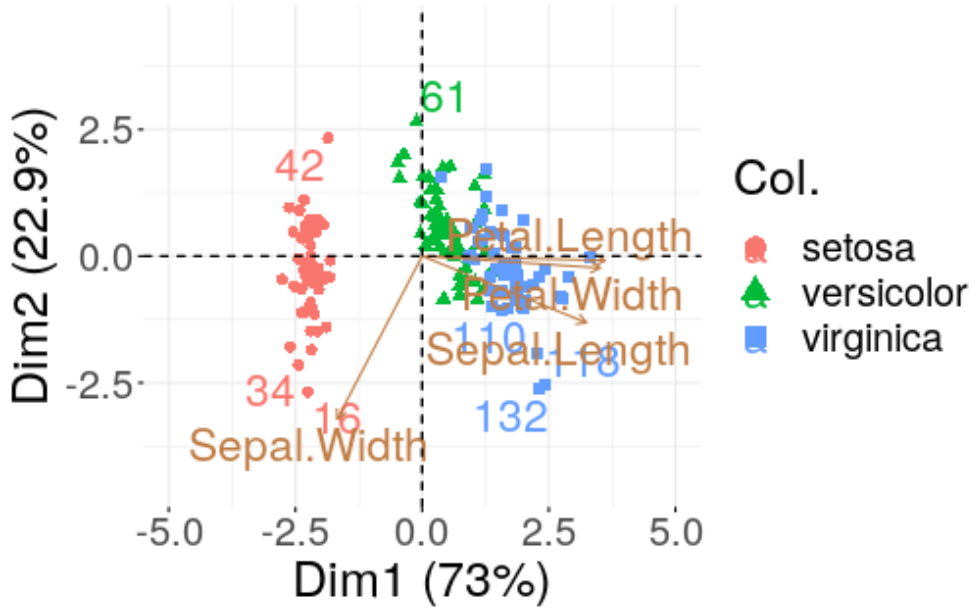
```
fviz_pca_var(pca.res, labelsize = 7) +
  ggtitle("") + theme(text = element_text(size = 25))
```



BioData Min. 2012 Nov 13;5(1):19.

biplot: plot observations and variables

```
p <- fviz_pca_biplot(pointsize = 2.5,
  pca.res, col.ind = y,
  repel = TRUE,  label="all",
  labelsize = 8, col.var = "#c07d44") +
  coord_fixed() + ggtitle("") +
  ylim(-4.5, 4.5) + xlim(-5, 5) +
  theme(text = element_text(size = 25))
```



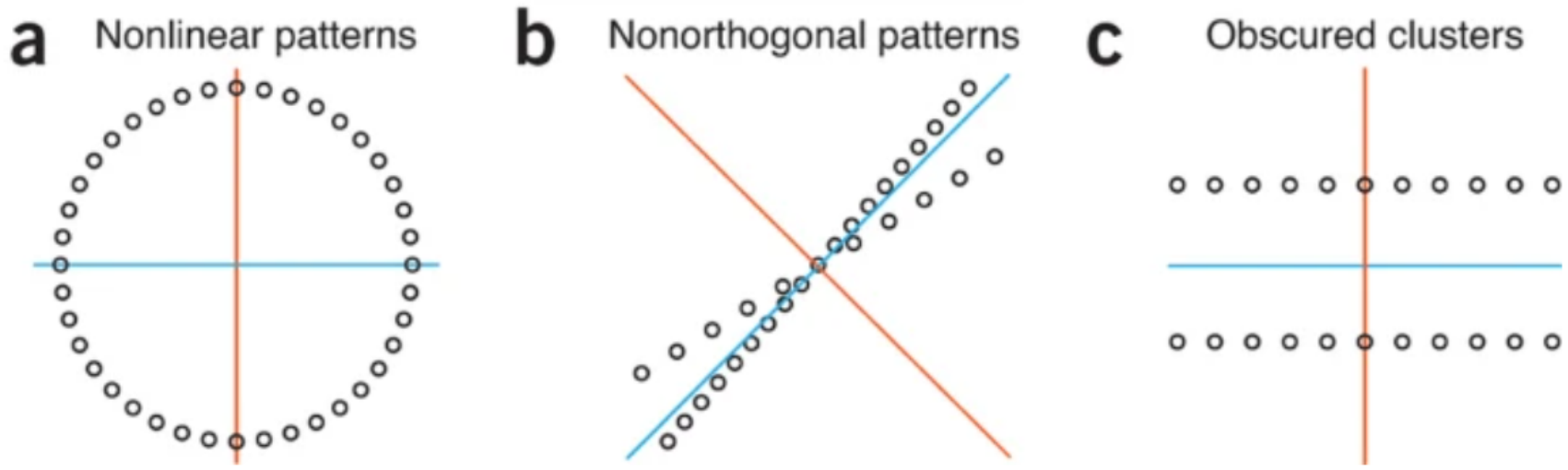
Sanity check

Sepal length is great for flower 132 as compared to flower 42

```
rbind(iris[42,], iris[132,]) %>%
  mutate(flower = rownames(.)) %>%
  gt() %>%
  tab_style(
    style = list(
      cell_fill(color = "lightcyan"),
      "font-variant: small-caps;"
    ),
    locations = cells_body(columns = Sepal.Length)
  )
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	flower
4.5	2.3	1.3	0.3	setosa	42
7.9	3.8	6.4	2.0	virginica	132

PCA fails in non-linear settings: *eg.* single cell omics



(**a**–**c**) Limitations of PCA are that it may miss nonlinear data patterns (**a**); structure that is not orthogonal to previous PCs may not be well characterized (**b**); and PC1 (blue) may not split two obvious clusters (**c**). PC2 is shown in orange.

Nature Methods volume 14, pages 641–642 (2017)

t-SNE (t-distributed Stochastic Neighbourhood Embedding)

Goal: preserve distances between points in a neighbourhood (controlled by the perplexity) in a lower dimensional space.

```
?tsne
```

- P = pairwise probabilities between observations given expected number of neighbours using a Gaussian distribution
- Q = pairwise probabilities of randomly generated observations in lower dimensions using a t-distribution with 1 degree of freedom
- use Kullback-Leibler (KL) divergence loss to minimize loss and update lower bound (Q)
- update randomly generated point using gradient descent using derivative of KL loss

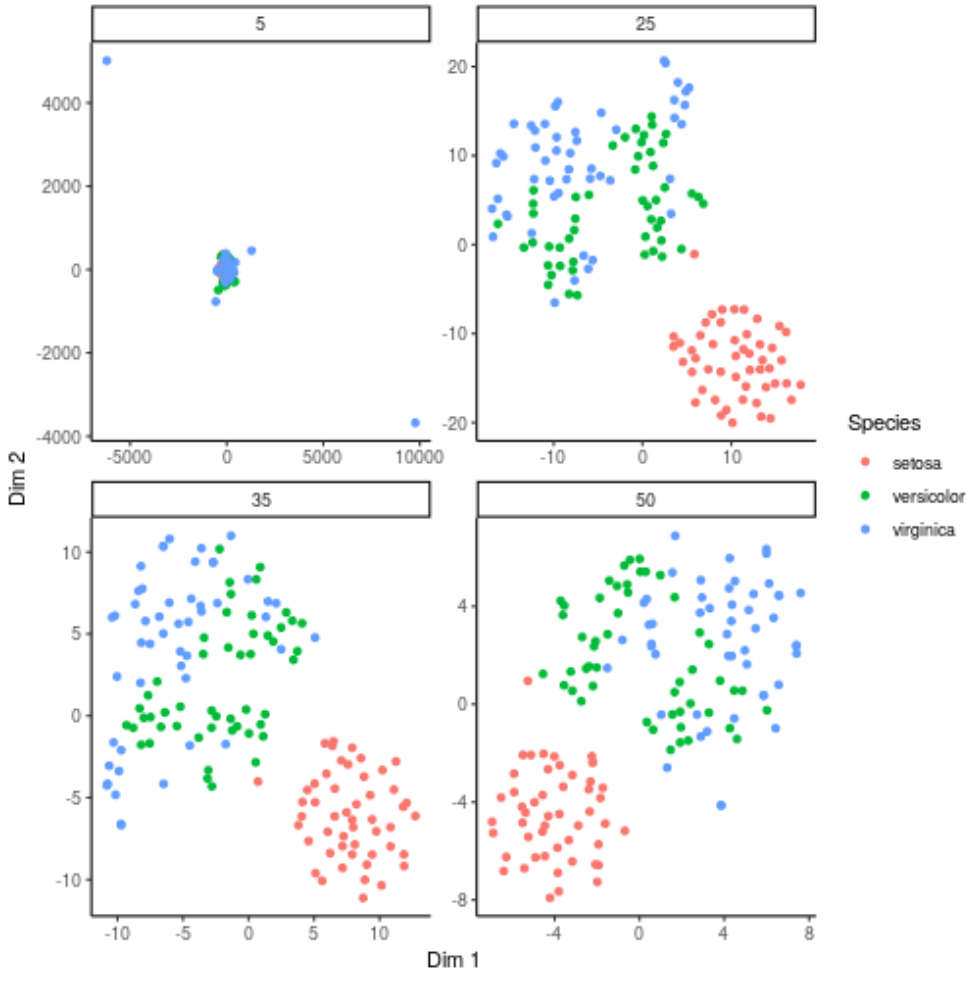
Key takeaways

1. Value of perplexity affects clustering (recommended range between 5-50)
2. cluster size (spread of points in a cluster) CANNOT be interpreted
3. distance between clusters CANNOT be interpreted

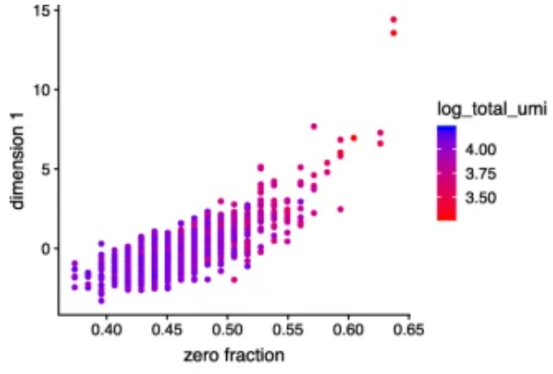
How to use t-SNE effectively StatQuest: t-SNE, Clearly Explained tSNE math explained

Apply tSNE to IRIS

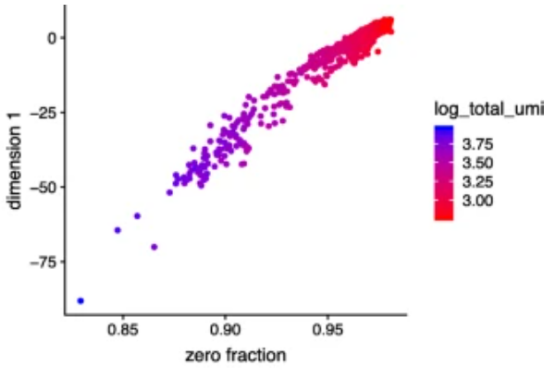
```
neighbours <- c(5, 25, 35, 50)

tsne_all <- lapply(neighbours, function(i){
  dims <- as.data.frame(tsne(iris[, setdiff(colnames(i
  dims$perplexity <- factor(i)
  dims$Species <- iris$Species
  dims
}) %>%
  do.call(rbind, .)

p <- tsne_all %>%
  ggplot(aes(x = V1, y = V2, color = Species)) +
  geom_point()+
  facet_wrap(~perplexity, scales = "free") +
  xlab("Dim 1") +
  ylab("Dim 2") +
  theme_classic()
```



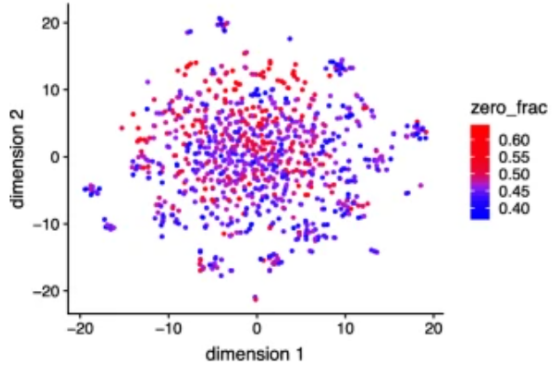
Apply PCA+tSNE to single cell data



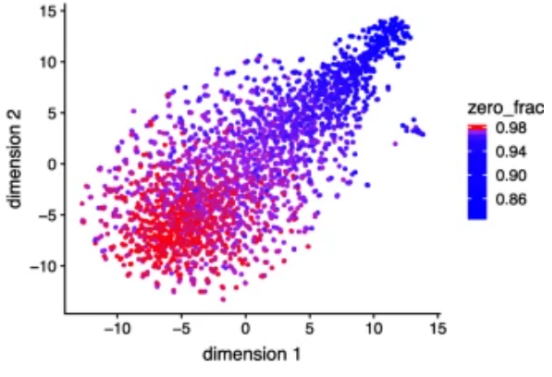
(a) Technical replicates, PCA



(b) Biological replicates, PCA



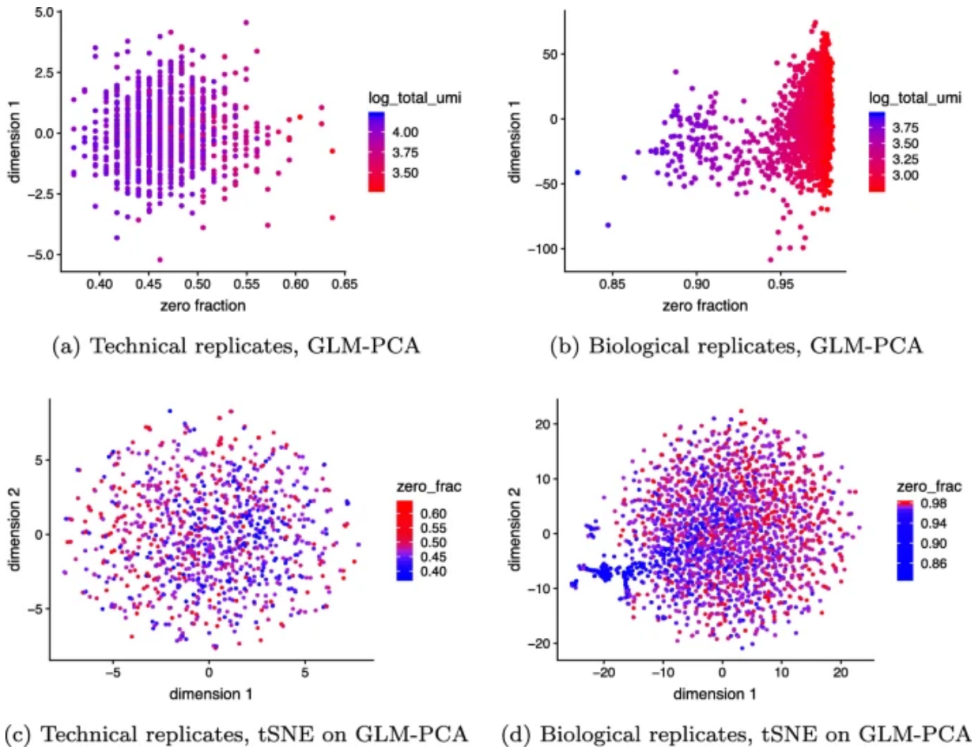
(c) Technical replicates, tSNE on PCA



(d) Biological replicates, tSNE on PCA

Genome Biol. 2019 Dec 23;20(1):295.

GLM-PCA



(a) Technical replicates, GLM-PCA

(b) Biological replicates, GLM-PCA

(c) Technical replicates, tSNE on GLM-PCA

(d) Biological replicates, tSNE on GLM-PCA

Genome Biol. 2019 Dec 23;20(1):295.

SCTransform in Seurat:

In particular, two recent studies proposed to use generalized linear models (GLMs), where cellular sequencing depth was included as a covariate, as part of scRNA-seq preprocessing workflows. Our ***sctransform*** *approach utilizes the Pearson residuals from negative binomial regression as input to standard dimensional reduction techniques, while* ***GLM-PCA*** *focuses on a generalized version of principal component analysis (PCA) for data with Poisson-distributed errors.*

Genome Biol. 2019;20(1):296.

Genome Biol. 2022; 23: 27.

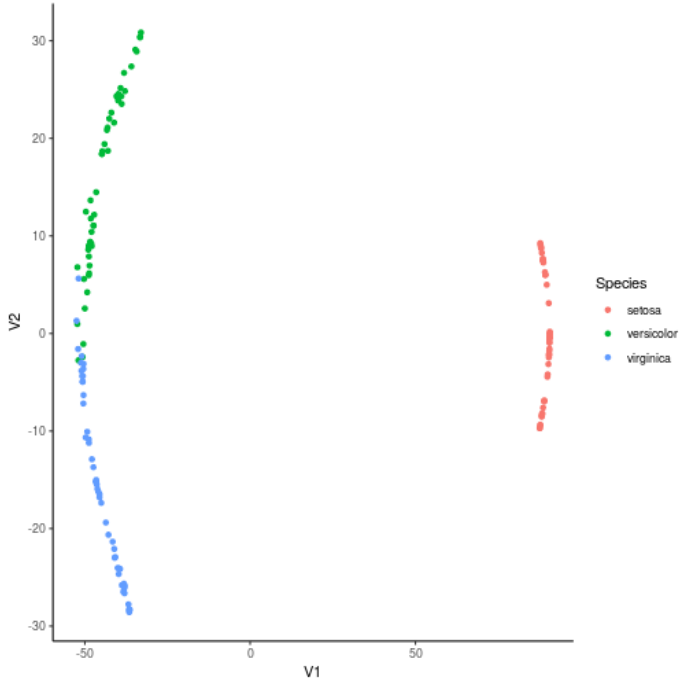
PHATE

```
write.table(iris, "iris.txt", sep="\t", row.names = F)
```

```
import os
import pandas as pd
import numpy as np
import phate # https://github.com/KrishnaswamyLab/PHAT

# import data
data = pd.read_csv("iris.txt", sep="\t")

# run PHATE
phate_op = phate.PHATE()
data_phate = phate_op.fit_transform(data.iloc[:,1:4])
```

```
## Calculating PHATE...
##   Calculating graph and diffusion operator...
## Calculating KNN search...
## Calculating affinities...
##   Calculating optimal t...
##     Automatically selected t = 19
##   Calculated optimal t in 0.02 seconds.
##   Calculating diffusion potential...
##   Calculating metric MDS...
```

```
phate %>%
  mutate(Species = iris$Species) %>%
  ggplot(aes(x = V1, y = V2, color = Species)) +
  geom_point() +
  theme_classic()
```



2) Preprocess input data

Methods	Year	Method strategy	Platform	Input	Available URL	Version	References
PCA	1987	Linear	R	Counts	R Package Seurat	3.1.0	Jolliffe, 2002
ICA	2001	Linear	R	Counts	R Package Seurat	3.1.0	Liebermeister, 2002
ZIFA	2015	Model-based	Python	Counts	https://github.com/epierson9/ZIFA	0.1	Pierson and Yau, 2015
GrandPrix	2017	Non-linear	Python	1,000 highly genes	https://github.com/ManchesterBioinference/GrandPrix	0.1	Ahmed et al., 2019
t-SNE	2008	Non-linear	R	Counts	R Package Rtsne	0.15	Maaten and Hinton, 2008
UMAP	2018	Non-linear	R/Python	Counts	https://github.com/lmcinnes/umap	0.3.1	McInnes et al., 2018
DCA	2019	Neural network	Python	1,000 Highly genes	https://github.com/theislab/dca	0.2.2	Eraslan et al., 2019
scvis	2018	Neural network	Python	PCA-100	https://bitbucket.org/jerry00/scvis-dev	0.1.0	Ding et al., 2018
VAE	2019	Neural network	Python	Counts	https://github.com/greenelab/CZI-Latent-Assessment/tree/master/single_cell_analysis	NA	Hu and Greene, 2019
SIMLR	2017	Ensemble method	R	Counts	https://github.com/BatzoglouLabSU/SIMLR	1.6.0	Wang et al., 2017

PCA:

- data must be centered for applying PCA (shifts origin to center of the data cloud)
- scale variable if they have different units

PHATE: Also works with counts

Front Genet. 2021 Mar 23;12:646936.

3) Handle categorical input data

Categorical data: Multiple Correspondence Analysis (MCA) (ade4::mcoa())

- code the data are 0s and 1s; one variable (with multiple categoies) can be coded with multiple columns (this results in an inflation of the variance) --> eigenvalue correction is applied

Multiple Correspondence Analysis

Mixed data: Multiple Factor Analysis (MFA) (FactoMineR::MFA())

- applies PCA to numerical data, and MCA to categorical data and combines the resulting components
- optimal scaling

FactoMineR

4) Use DR methods for similiarity/dissimilarity matrices

- choose metric that is appropriate for your data vegan:
  - binary: Manhattan
  - sparse: Jaccard
  - other distances: Euclidean, gower

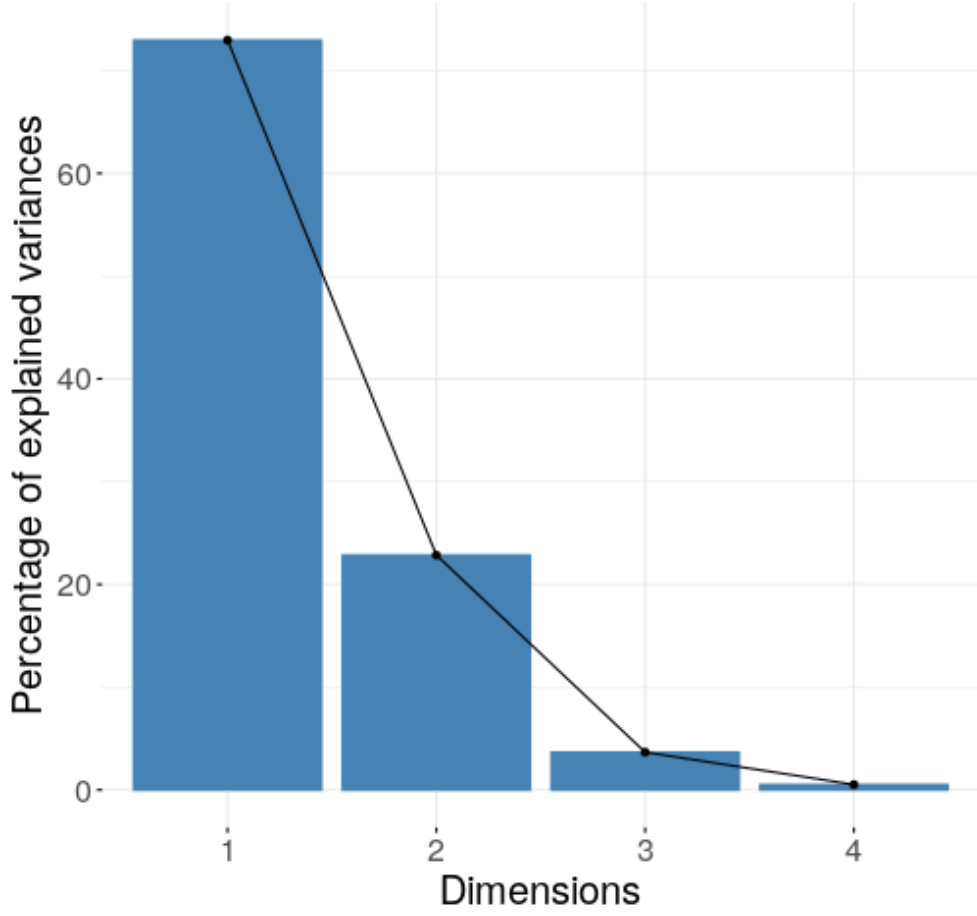
5) Decide on number of dimensions to retain

PCA: scree plot (elbow point)

Other

- t-SNE: use # of dims that minimizes the KL divergence loss

```
fviz_eig(pca.res) + ggtitle("") +
  theme(text = element_text(size = 20))
```



6) Apply the correct aspect ratio to plots

```
fviz_pca_ind(pca.res, geom = "point") + coord_fixed() +
  ggtitle("") + xlim(NA, 5) +
  theme(text = element_text(size = 30))
```

8) Find the hidden signal

- **PCA**: linear method, reduces noise by transforming data, fits ellipsoid to data (models global structure) - can apply to any data
- **tSNE**: non-linear method, models local structure (distance between clusters is meaningless) - commonly used for single cell data
- **PHATE**: non-linear method: models both local and global structure; useful to model differentiation trajectories

9) Favor multidomain data

- STATIS and DISTATIS are generalization of PCA to multiple data matrices (more on the on unsupervised data integration)

10) Robustness of results and quantify uncertainties

- test different parameters to see how clustering changes (*e.g.* perplexity in t-SNE)
- outliers can affect DR methods; robust variants are available (*e.g.* robust PCA)

10 Tips for effective DR

1) Choose an appropriate DM method

2) Preprocess input data

3) Handle categorical input data

4) Use DR methods for similiarity/dissimilarity matrices

5) Decide on number of dimensions to retain

6) Apply the correct aspect ratio to plots

7) What do the new dimensions mean?

8) Find the hidden signal

9) Favor multidomain data

10) Robustness of results and quantify uncertainties

PRECISION HEALTH
ANALYSIS BOOTCAMP

THANK YOU!

August 05, 2022 | 09:00-11:00

lab
code
asingh_22g