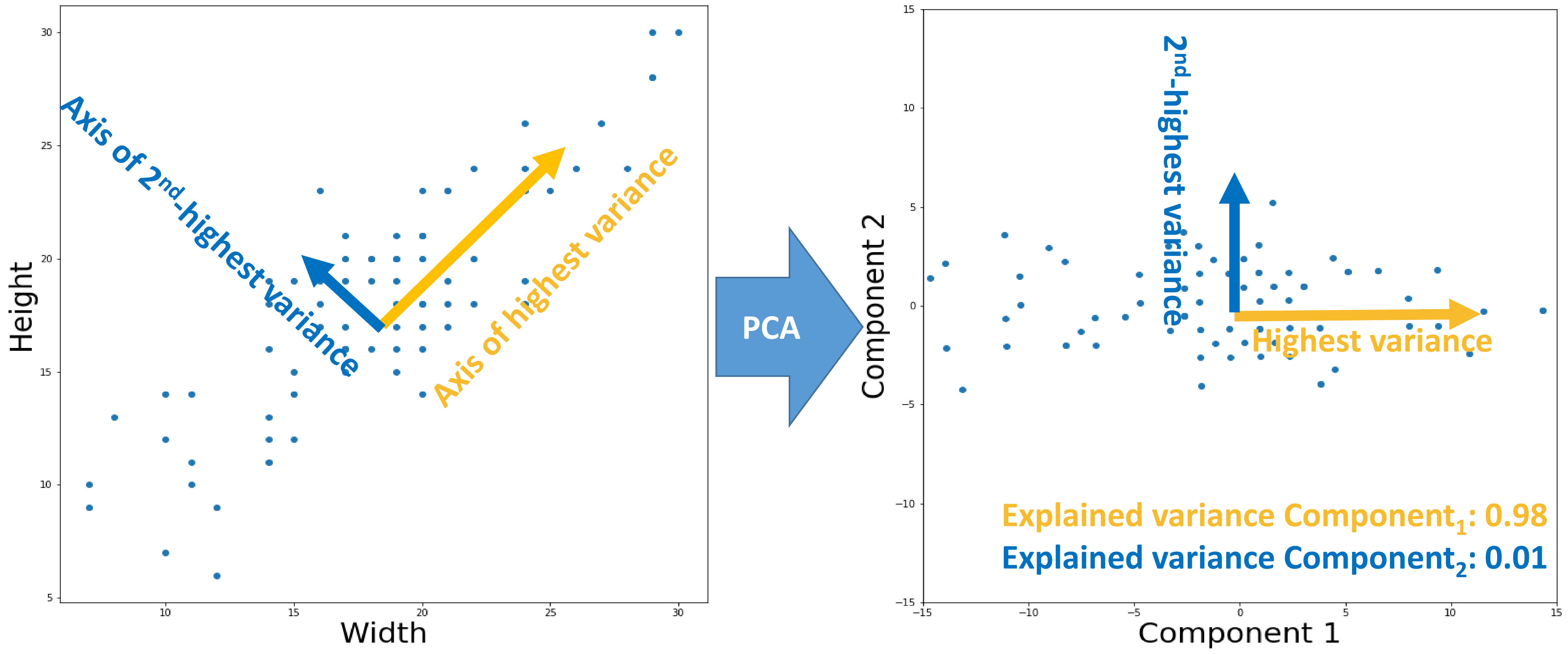
Recap day 4

Robert Haase, Marcelo Zoccoler, Johannes Müller

October 2022

Dimensionality reduction and clustering



Explained variance Component$_1$: 0.98

Explained variance Component$_2$: 0.01

PCA: Linear transformation and identifies axes that explain most of the variance in the data

→ If the explained variances of component$_1$ and component$_2$ are [0.32, 0.2], we…

Need more components	Cry	Will find no groups in the data

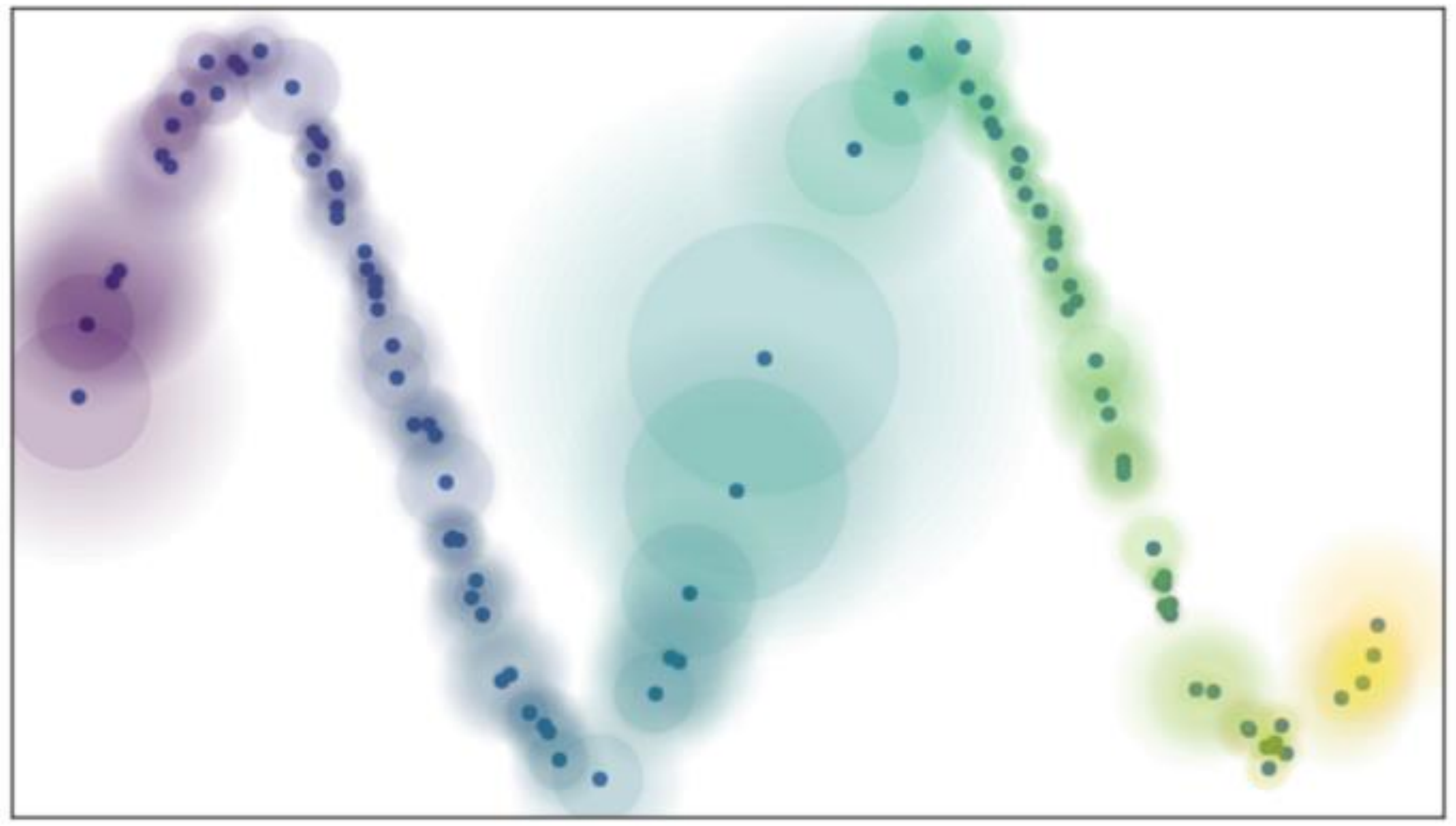
→ If we add more data to a PCA-transformed set of data, we …

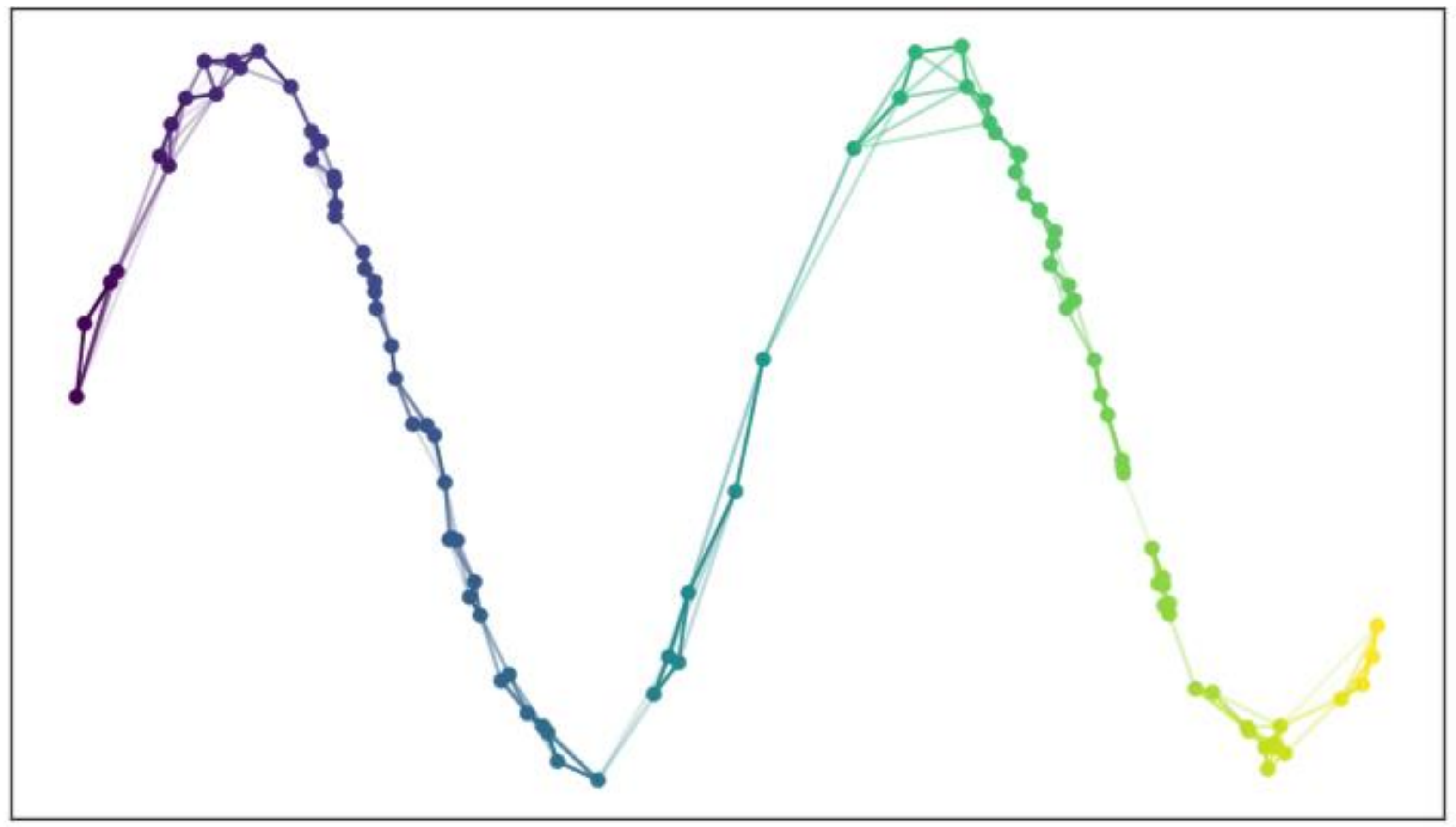
Need to re-run the PCA	Can use the same PCA components	Cannot use PCA at all

→ We can measure meaningful differences between data point in terms of component$_1$ and component$_2$

True	False

UMAP





UMAP:

→ Link neighboring points by introducing point-wise distance metrics

→ "Relax" neighborhood graph to be able to display it in fewer dimensions while preserving its topology

We can measure meaningful differences between data points in terms of UMAP_0 and UMAP_1

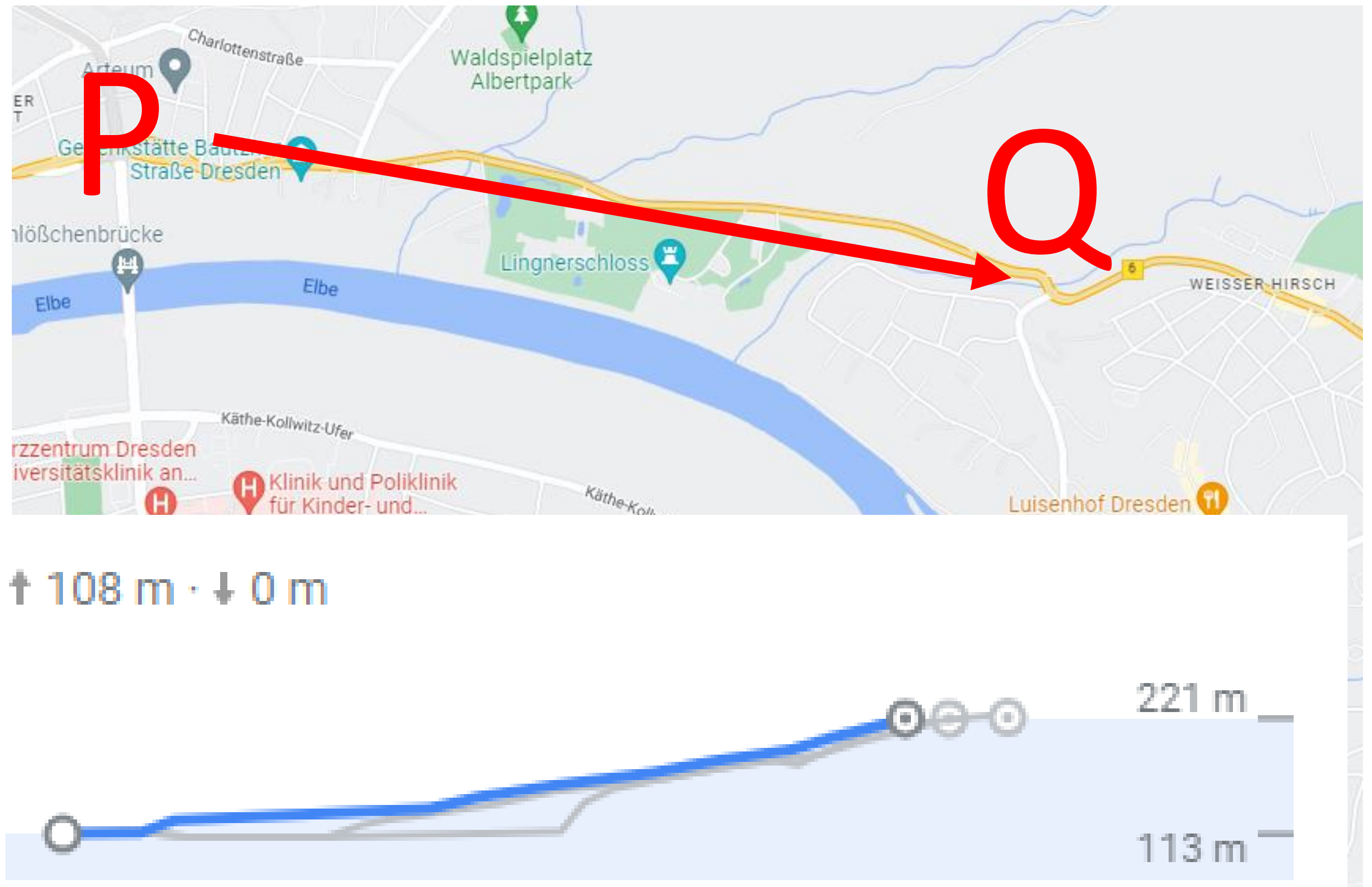
True | False

If we add new points to an existing UMAP projection, we can use the same projection

True | False

Running a UMAP twice will give the same result

True | False

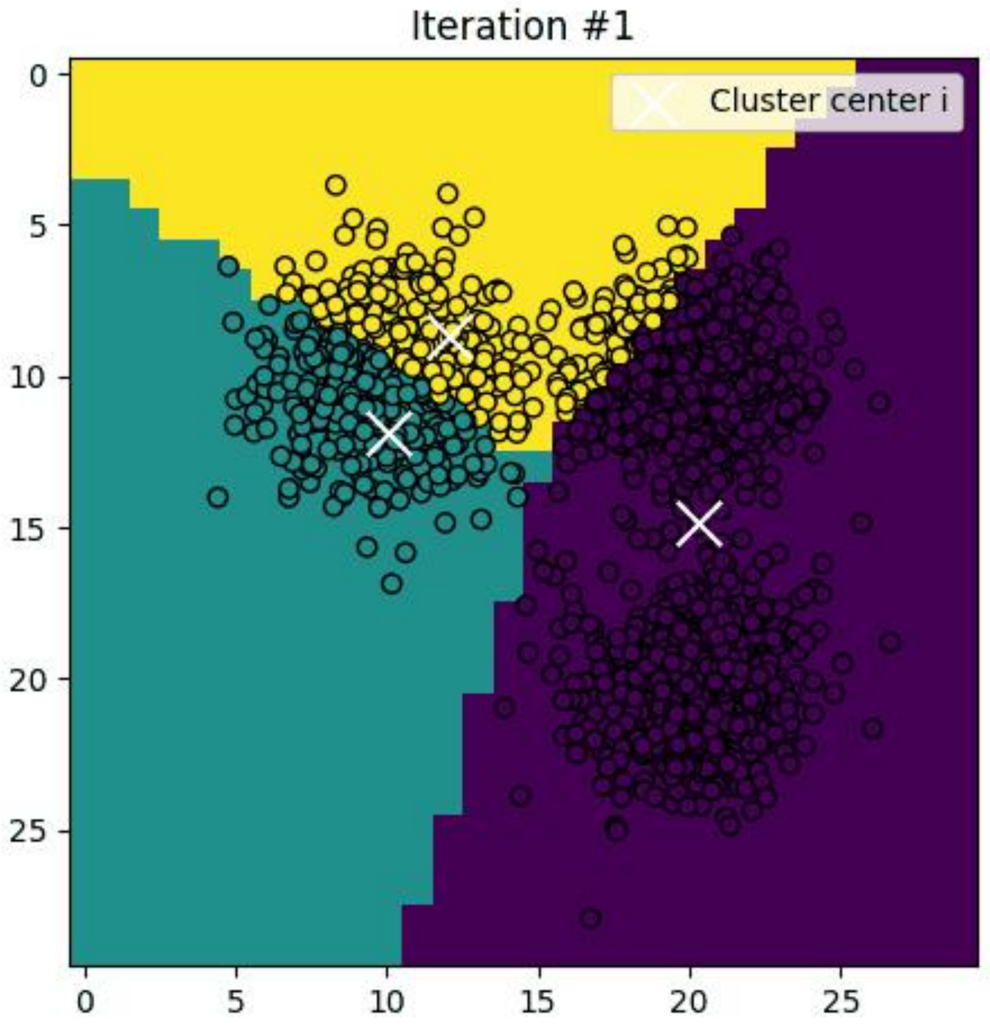


Source: https://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html

@haesleinhuepf, @jm_mightypirate

October 2022

Clustering



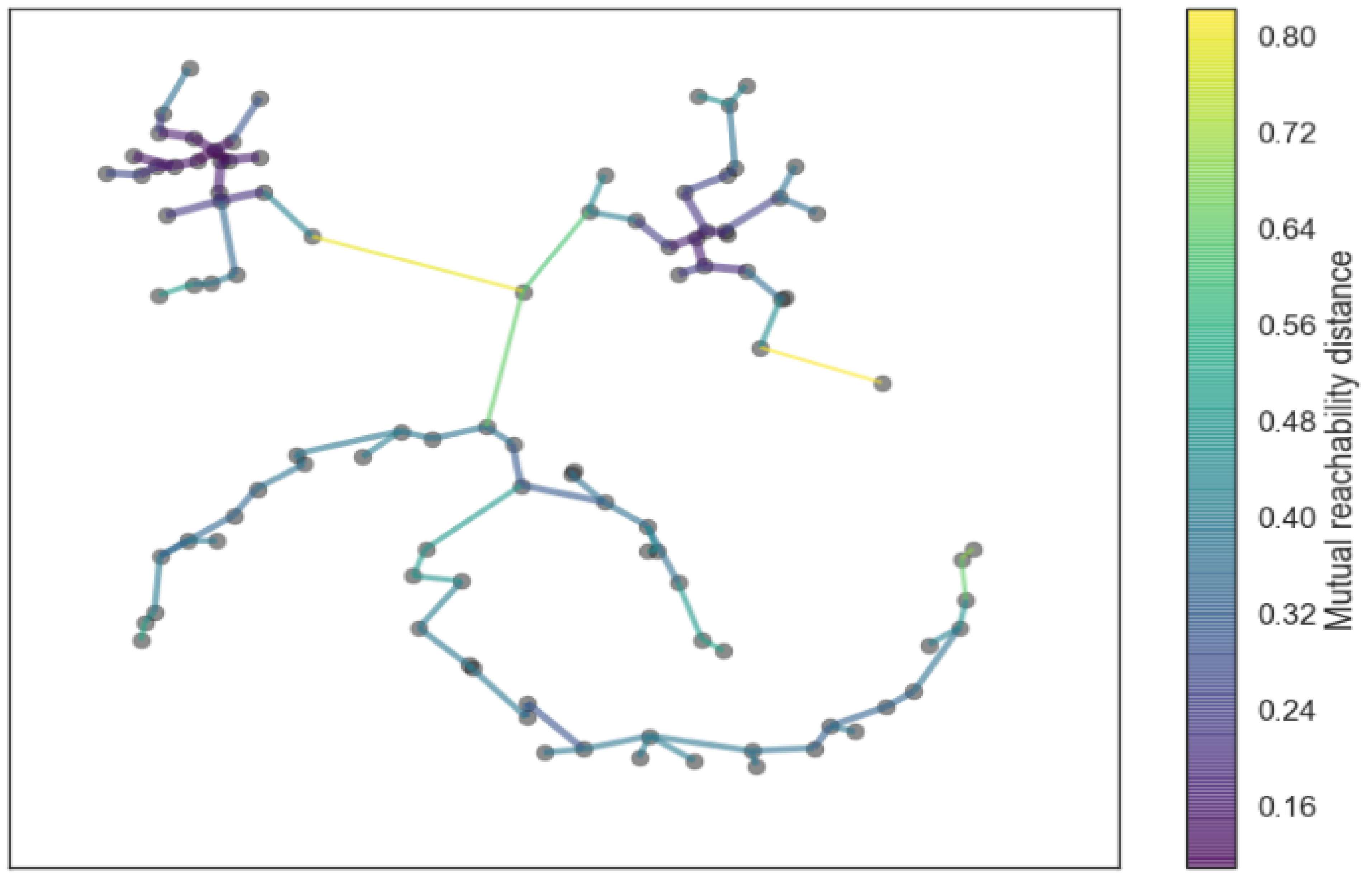
K-Means clustering:

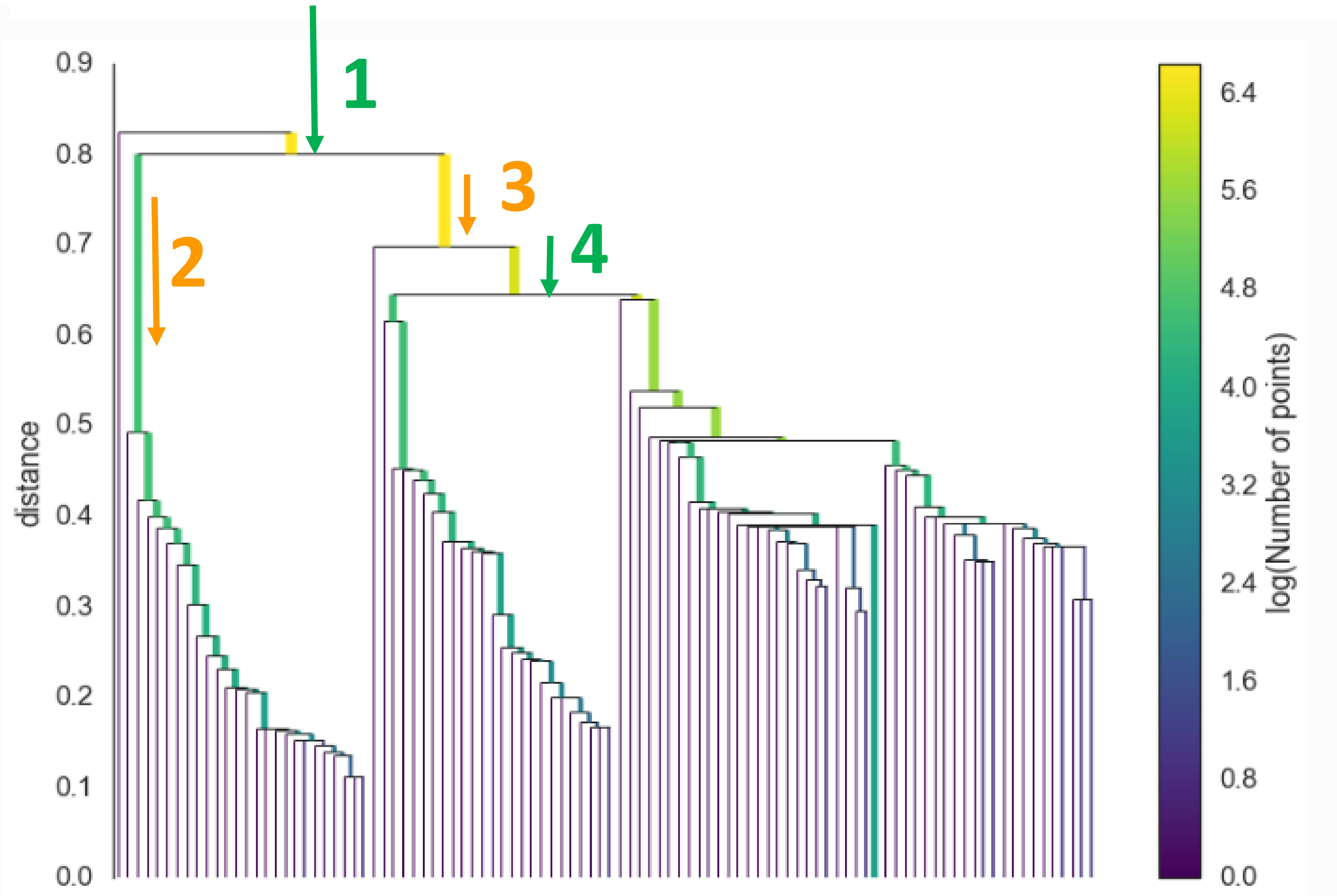
If we add more data to the pool, we can infer the cluster of a new point from the previously determined clusters

True

False

HDBSCAN: If we add more data to the pool, we can infer the cluster of a new point from the previously determined clusters





True

False

An important parameter is…

The point density

Data units

Minimal cluster size

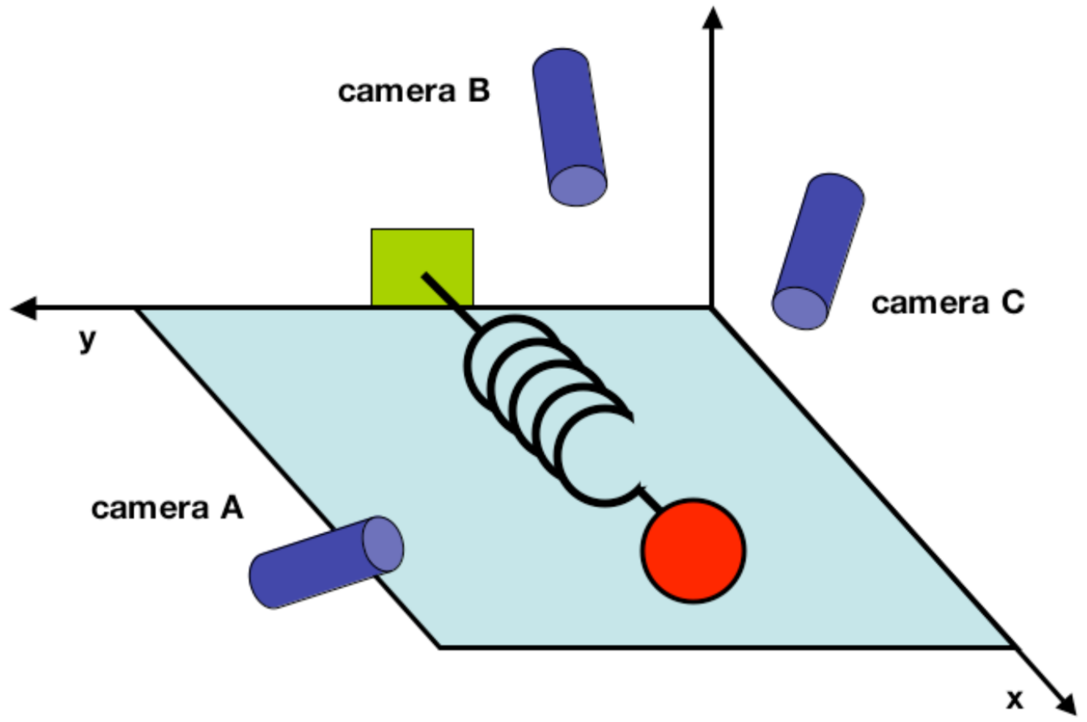
Source: https://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html

PCA

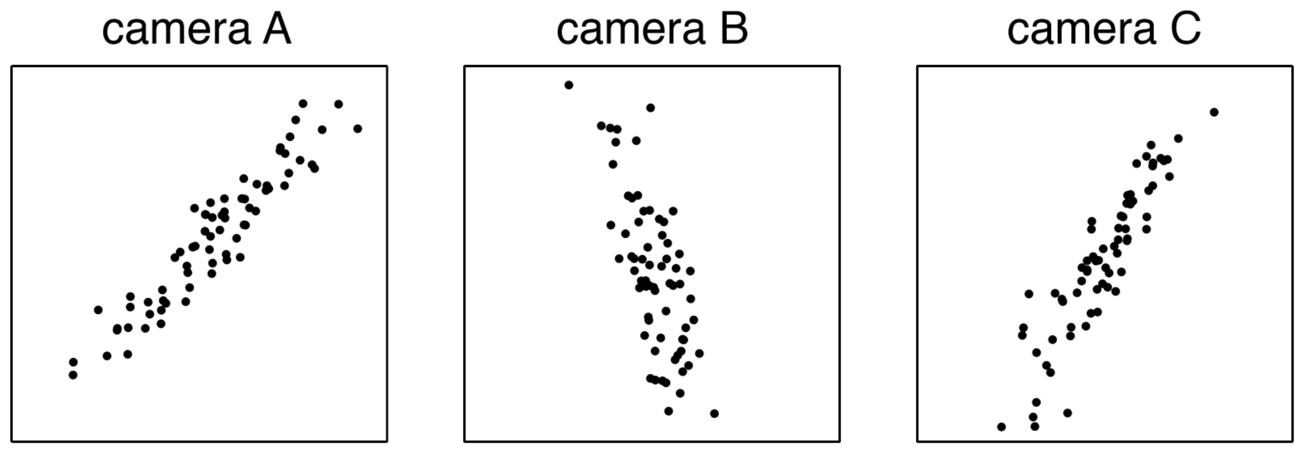
A Tutorial on Principal Component Analysis

Jonathon Shlens

https://arxiv.org/abs/1404.1100



$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} \\ \vdots \\ \mathbf{x_m} \end{bmatrix}$$



covariance matrix $\mathbf{C_X}$.

$$\mathbf{C_X} \equiv \frac{1}{n}\mathbf{X}\mathbf{X}^T.$$

Is there another basis, which is a linear combination of the original basis, that best re-expresses our data set?

$$\mathbf{PX} = \mathbf{Y}$$

SVD is a more general method for change of basis.

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$$

$\mathbf{C_X}$ captures the covariance between all possible pairs of measurements. The covariance values reflect the noise and redundancy in our measurements.

- In the diagonal terms, by assumption, large values correspond to interesting structure.
- In the off-diagonal terms large magnitudes correspond to high redundancy.

$$\begin{aligned}
\mathbf{C_Y} &= \frac{1}{n}\mathbf{Y}\mathbf{Y}^T \\
&= \frac{1}{n}(\mathbf{PX})(\mathbf{PX})^T \\
&= \frac{1}{n}\mathbf{P}\mathbf{X}\mathbf{X}^T\mathbf{P}^T \\
&= \mathbf{P}(\frac{1}{n}\mathbf{X}\mathbf{X}^T)\mathbf{P}^T \\
\mathbf{C_Y} &= \mathbf{P}\mathbf{C_X}\mathbf{P}^T
\end{aligned}$$

A symmetric matrix is diagonalized by a matrix of its orthonormal eigenvectors.

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^T$$

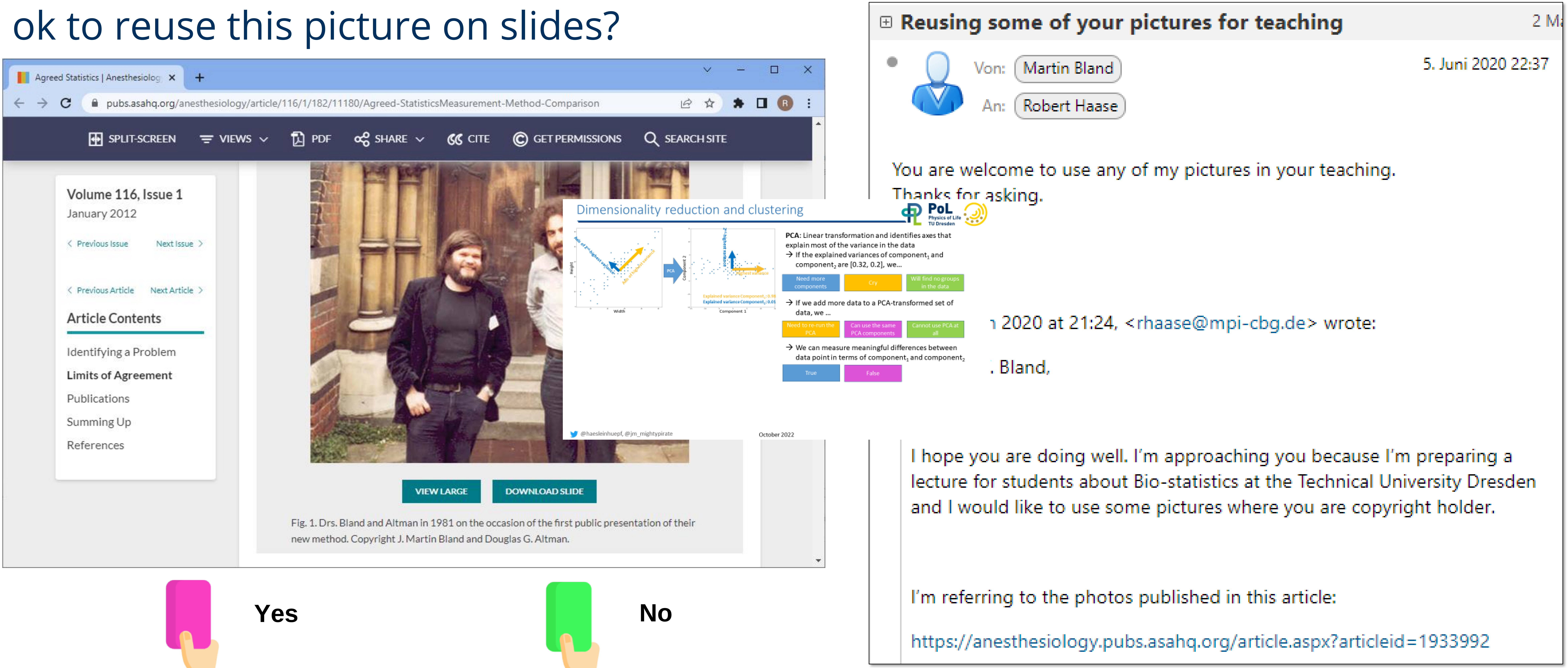
$$\begin{aligned}
\mathbf{C_Y} &= \mathbf{P}\mathbf{C_X}\mathbf{P}^T \\
&= \mathbf{P}(\mathbf{E}^T\mathbf{D}\mathbf{E})\mathbf{P}^T
\end{aligned}$$

If we select P as the eigenvectors of C$_x$: $\mathbf{P} \equiv \mathbf{E}^T$

$$\begin{aligned}
&= \mathbf{P}(\mathbf{P}^T\mathbf{D}\mathbf{P})\mathbf{P}^T \\
&= (\mathbf{P}\mathbf{P}^T)\mathbf{D}(\mathbf{P}\mathbf{P}^T) \\
&= (\mathbf{P}\mathbf{P}^{-1})\mathbf{D}(\mathbf{P}\mathbf{P}^{-1}) \\
\mathbf{C_Y} &= \mathbf{D}
\end{aligned}$$

Quiz

It's ok to reuse this picture on slides?



Agreed Statistics | Anesthesiolog ×

pubs.asahq.org/anesthesiology/article/116/1/182/11180/Agreed-StatisticsMeasurement-Method-Comparison

Volume 116, Issue 1
January 2012

Article Contents

Identifying a Problem
Limits of Agreement
Publications
Summing Up
References

Fig. 1. Drs. Bland and Altman in 1981 on the occasion of the first public presentation of their new method. Copyright J. Martin Bland and Douglas G. Altman.

Dimensionality reduction and clustering

PCA: Linear transformation and identifies axes that explain most of the variance in the data

→ If the explained variances of component$_1$ and component$_2$ are [0.32, 0.2], we...

Need more components | Cry | Will find no groups in the data

→ If we add more data to a PCA-transformed set of data, we ...

Need to re-run the PCA | Can use the same PCA components | Cannot use PCA at all

→ We can measure meaningful differences between data point in terms of component$_1$ and component$_2$

True | False

@haesleinhuepf, @jm_mightypirate

October 2022

Reusing some of your pictures for teaching

2 M

Von: Martin Bland

An: Robert Haase

5. Juni 2020 22:37

You are welcome to use any of my pictures in your teaching.

Thanks for asking.

ı 2020 at 21:24, <rhaase@mpi-cbg.de> wrote:

. Bland,

I hope you are doing well. I'm approaching you because I'm preparing a lecture for students about Bio-statistics at the Technical University Dresden and I would like to use some pictures where you are copyright holder.

I'm referring to the photos published in this article:

https://anesthesiology.pubs.asahq.org/article.aspx?articleid=1933992

Yes

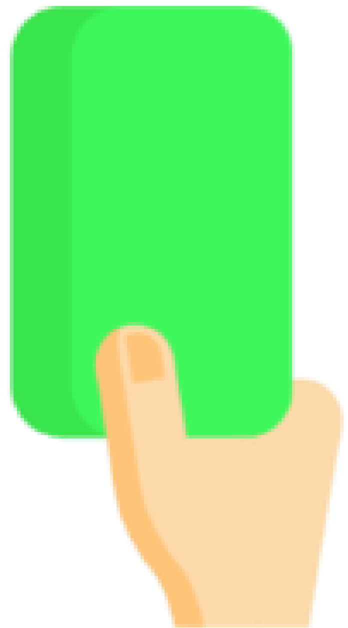
No

Source:
https://anesthesiology.pubs.asahq.org/article.aspx?articleid=1933992

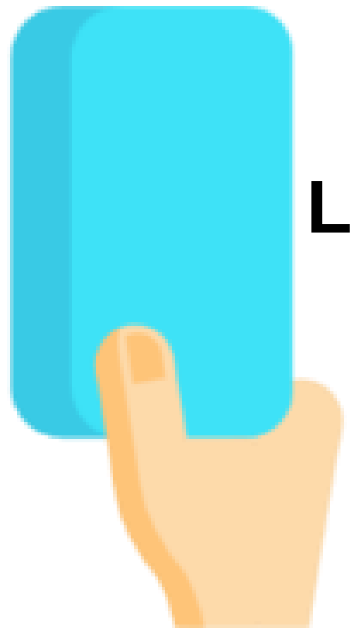
Quiz

It's ok to reuse code from this repository if…

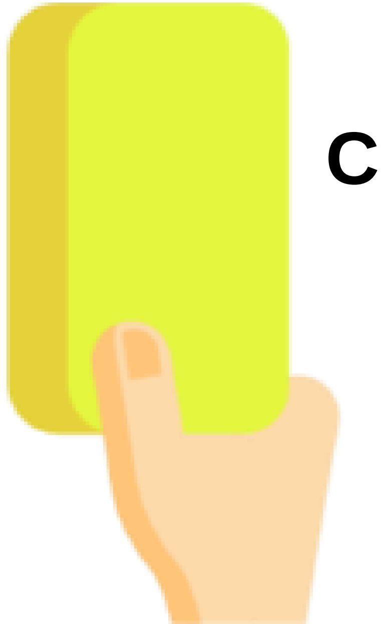
Mention author



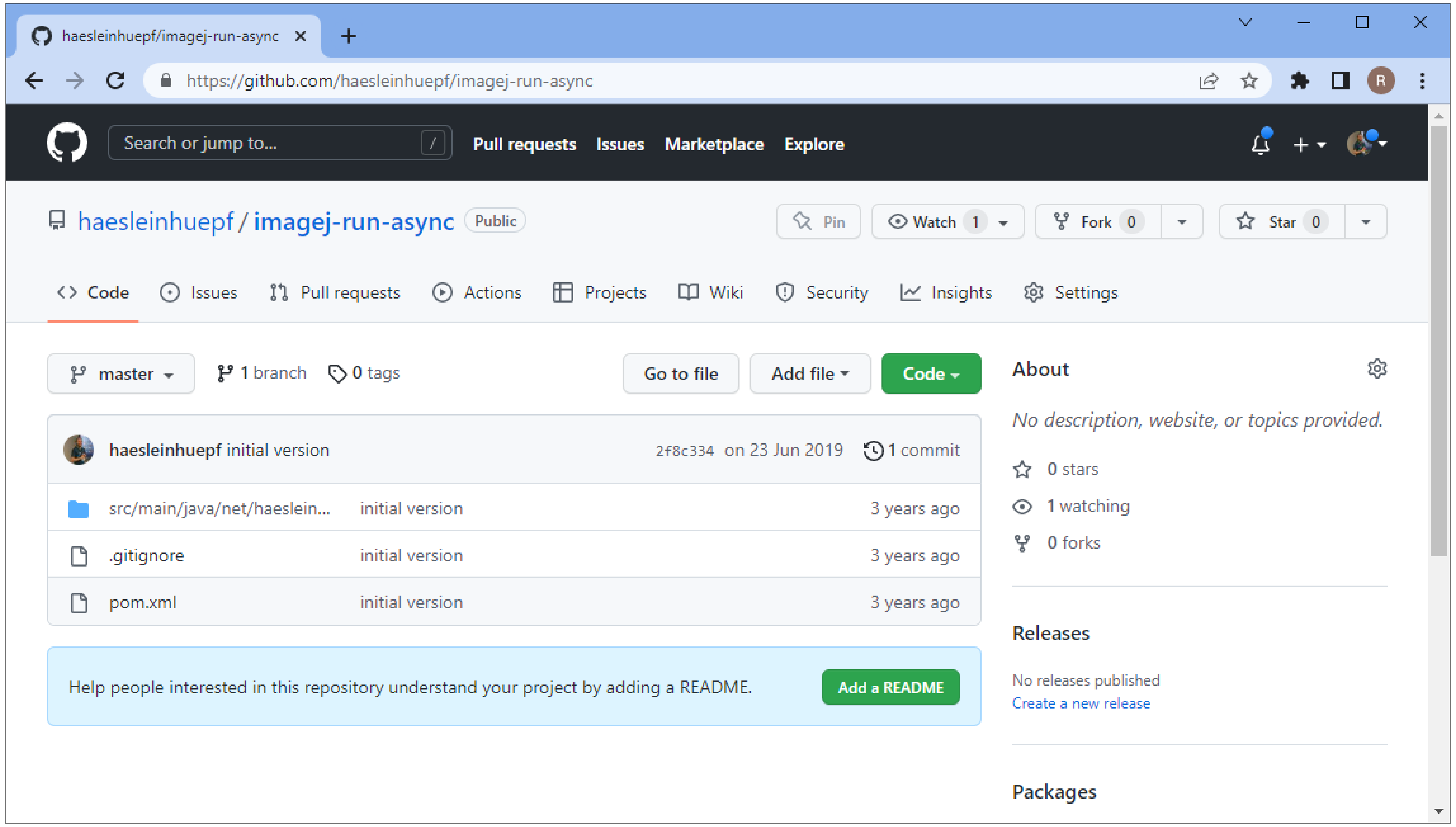
Ask the authors



Link to the repository



Copy the copyright statement



haesleinhuepf / imagej-run-async Public

1 branch 0 tags

haesleinhuepf initial version 2f8c334 on 23 Jun 2019 1 commit

src/main/java/net/haeslein…	initial version	3 years ago
.gitignore	initial version	3 years ago
pom.xml	initial version	3 years ago

Help people interested in this repository understand your project by adding a README.

About

No description, website, or topics provided.

0 stars

1 watching

0 forks

Releases

No releases published

Create a new release

Packages

Quiz

Physics of Life TU Dresden

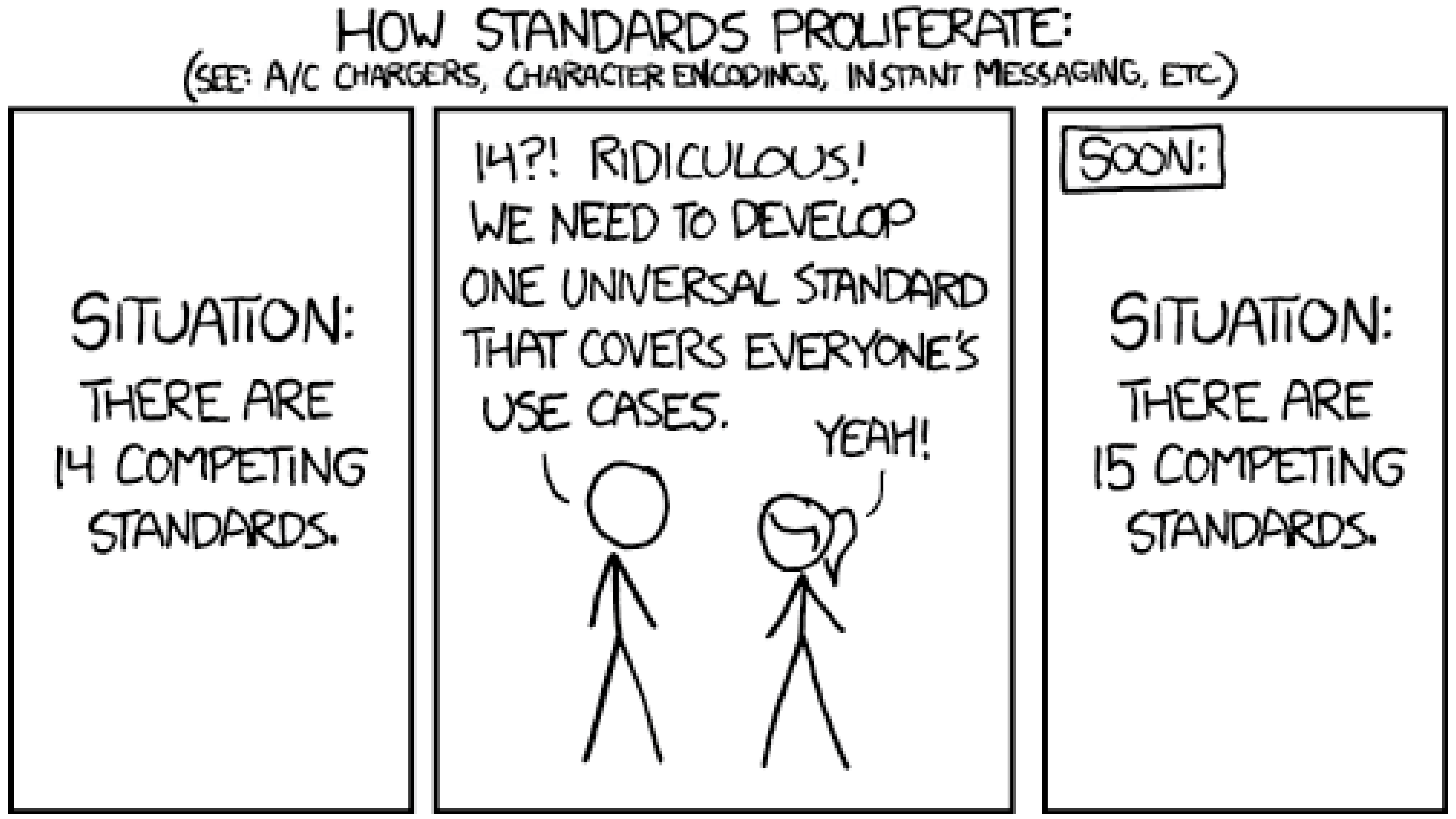
It's ok to reuse Figures from XKCD.com if...

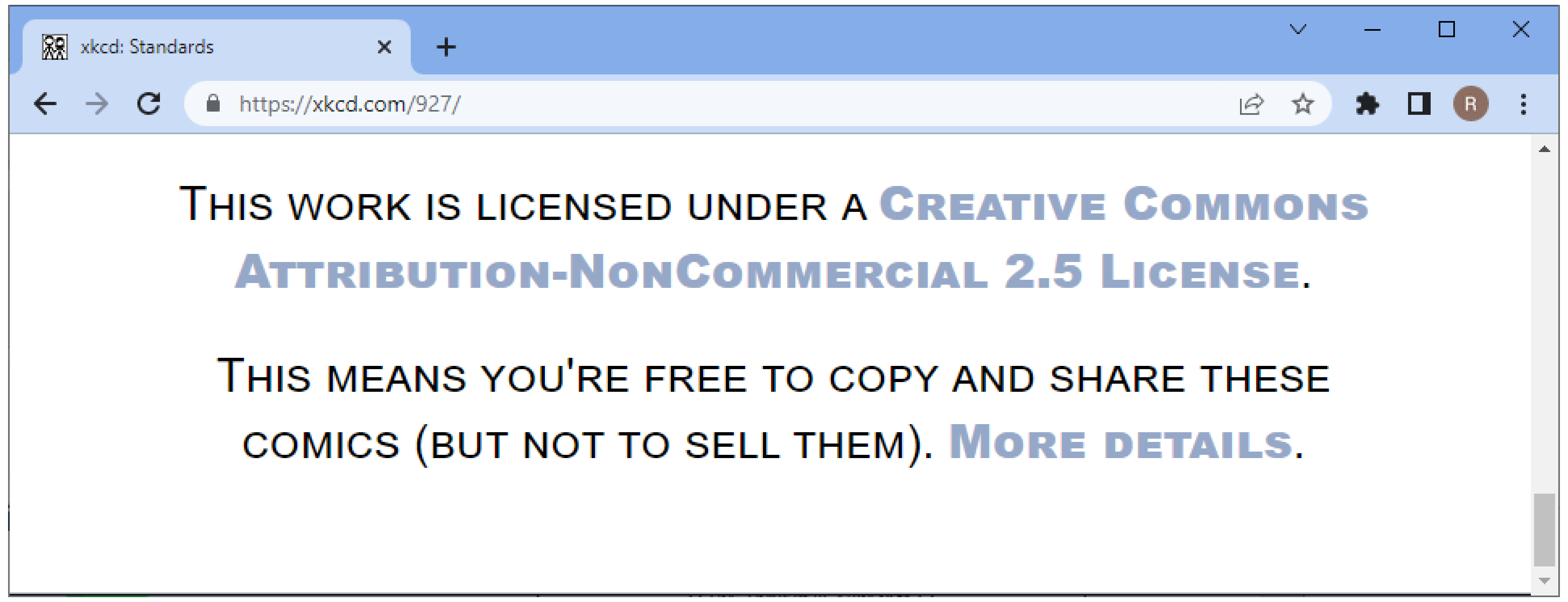
🟪 **Mention author**

🟩 **Ask the authors**

🟦 **Link to the license**

🟨 **Copy the copyright statement**


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.



THIS WORK IS LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL 2.5 LICENSE.

THIS MEANS YOU'RE FREE TO COPY AND SHARE THESE COMICS (BUT NOT TO SELL THEM). MORE DETAILS.

https://xkcd.com/927/
The Figure is licensed by XKCD.com under a Creative Commons Attribution-NonCommercial 2.5 License.