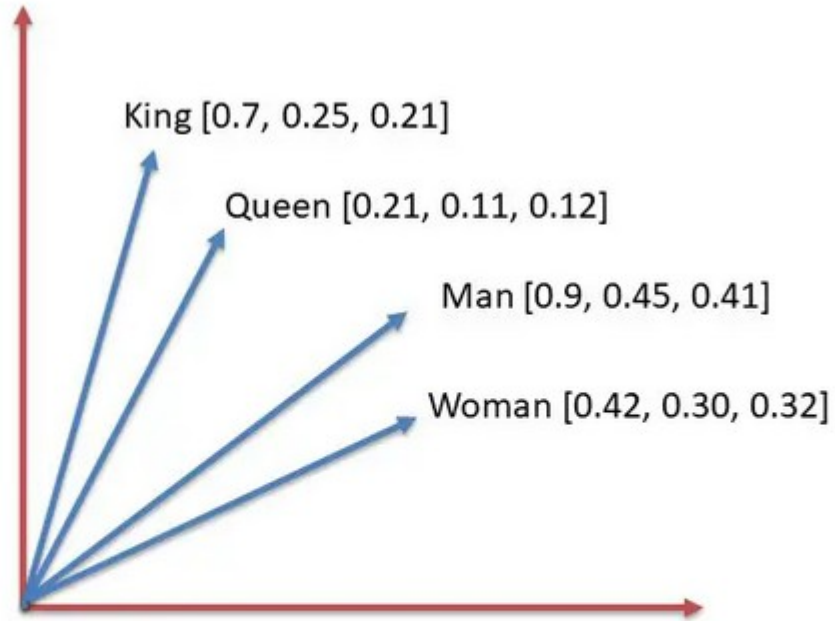


# Using Word Embeddings to analyse differences and shifts of semantics over time, space, and corpora

Stefan Resch

# What are word embeddings?



$$\begin{aligned}\text{King} - \text{Man} + \text{Woman} &= [0.7, 0.25, 0.21] - [0.9, 0.45, 0.41] + [0.42, 0.30, 0.32] \\ &= [0.22, 0.1, 0.12] \approx \text{Queen}\end{aligned}$$

# Stack

- Static word embeddings
  - word2vec
  - fastText
  - GloVe
- Contextual word embeddings
  - BERT families
- postgresql + pgvector

# Visualizations and clusterings

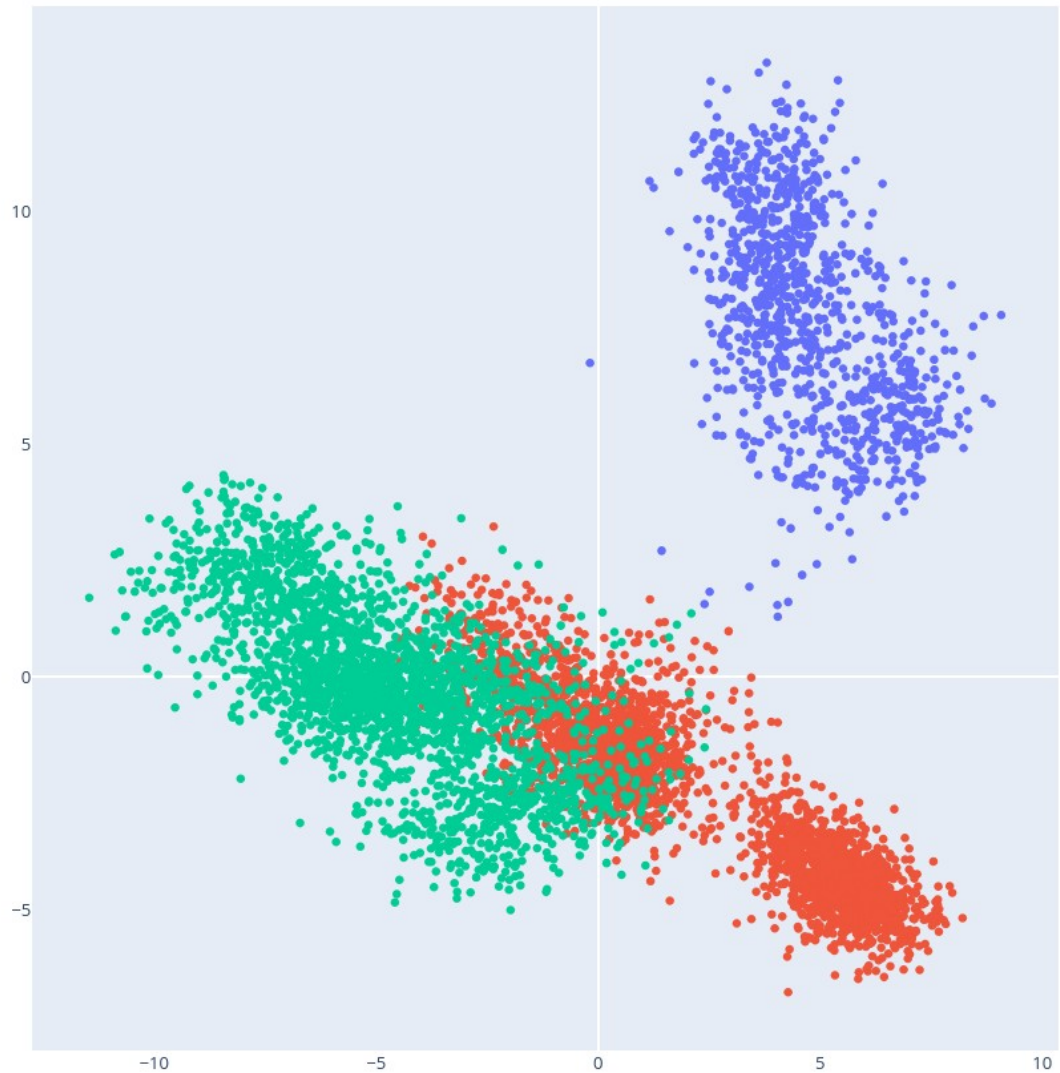
- visualizations
  - Parallel coordinats
  - t-SNE
  - PCA
  - UMAP
- clustering
  - k-means
  - HDBSCAN

# Parallel coordinates

dimensions with highest variation to the left, lowest to the right

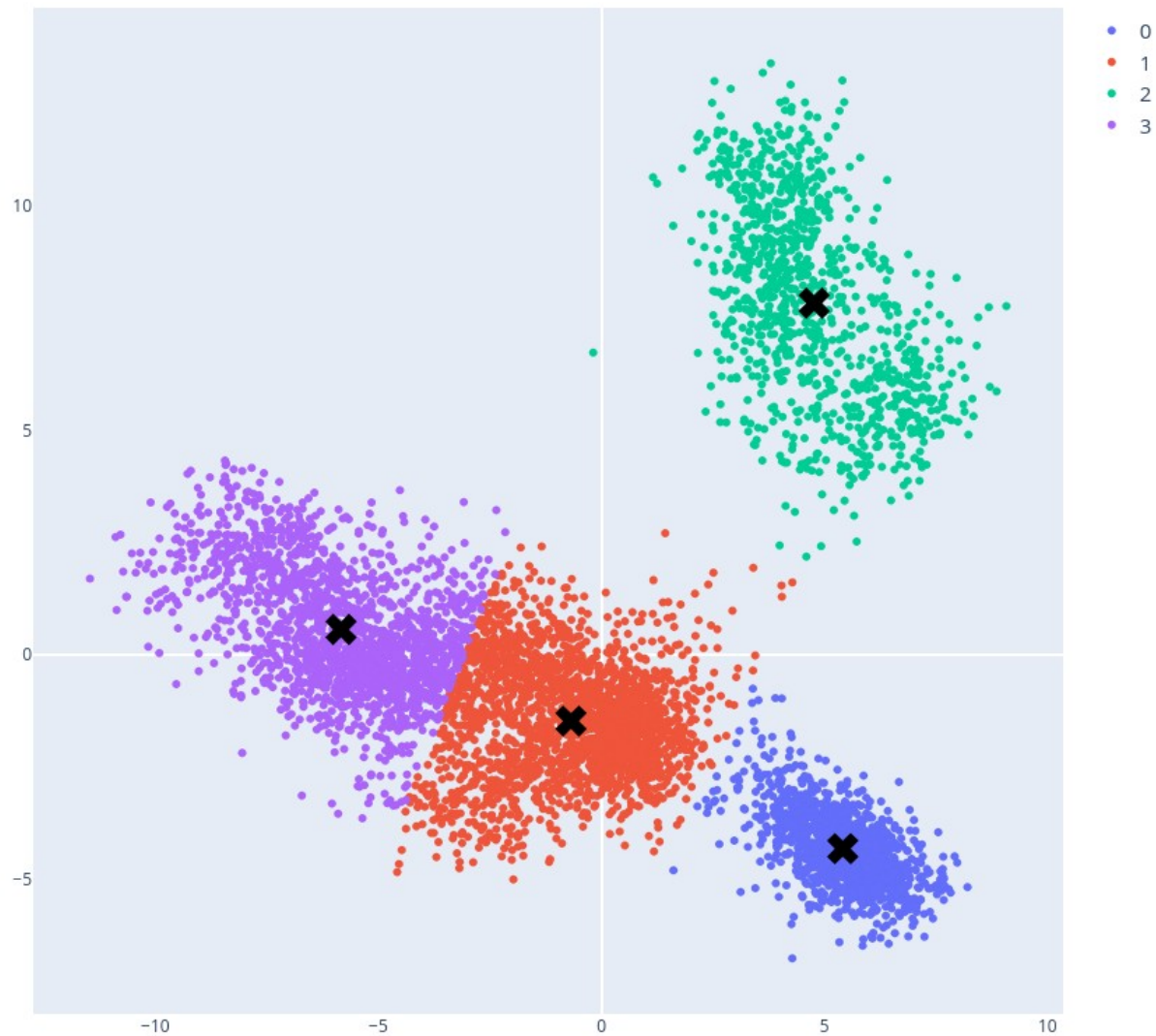


PCA reduction. No clustering



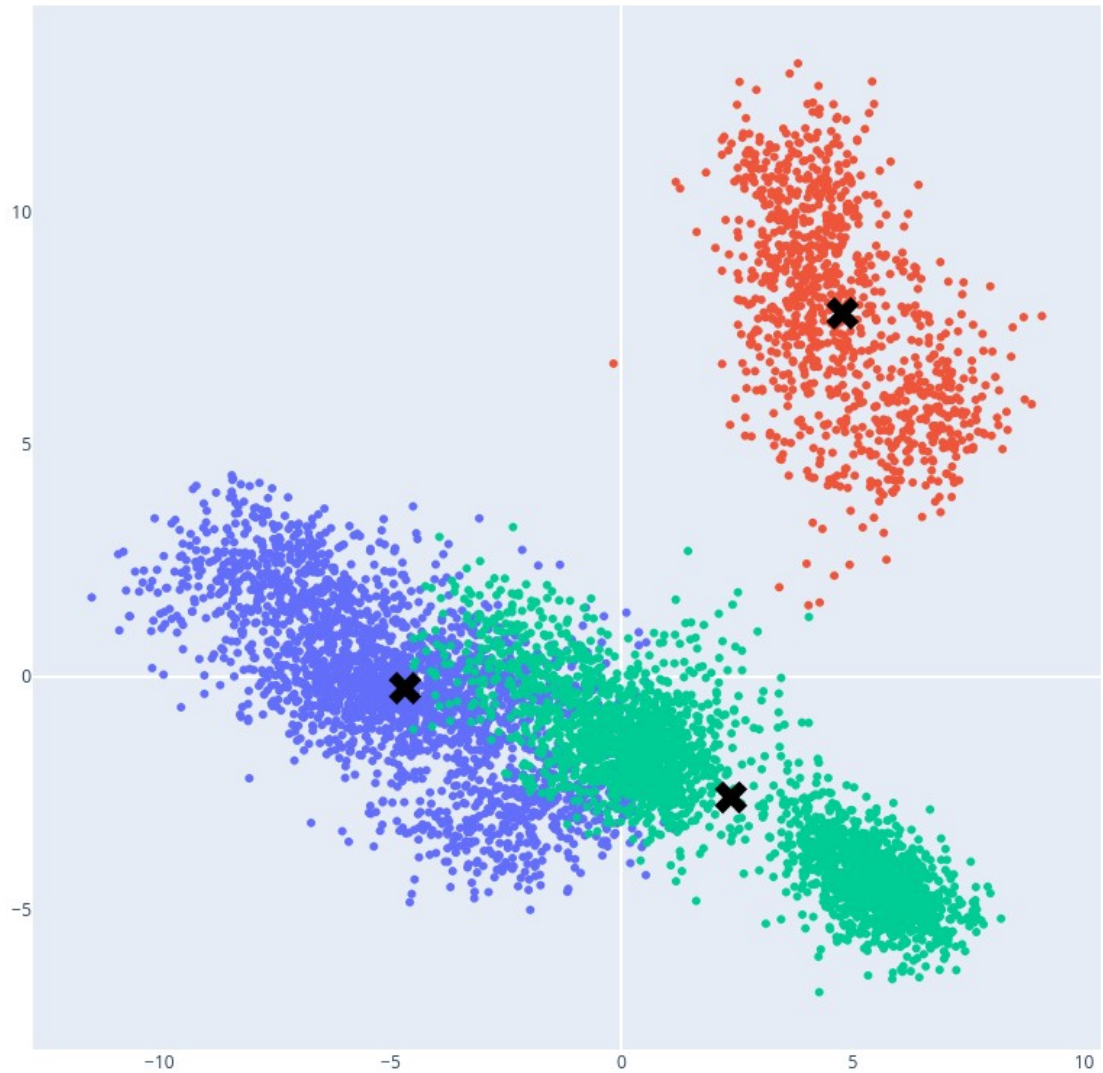
- Auto
- Frau
- Mann

PCA reduces high dimensional data to two dimensions by picking the two dimensions with the highest variance



K-means clusters by dividing all vectors into groups where the distances to the centers are minimized while group memberships are equalized.

Here, k-means is applied on 2D data

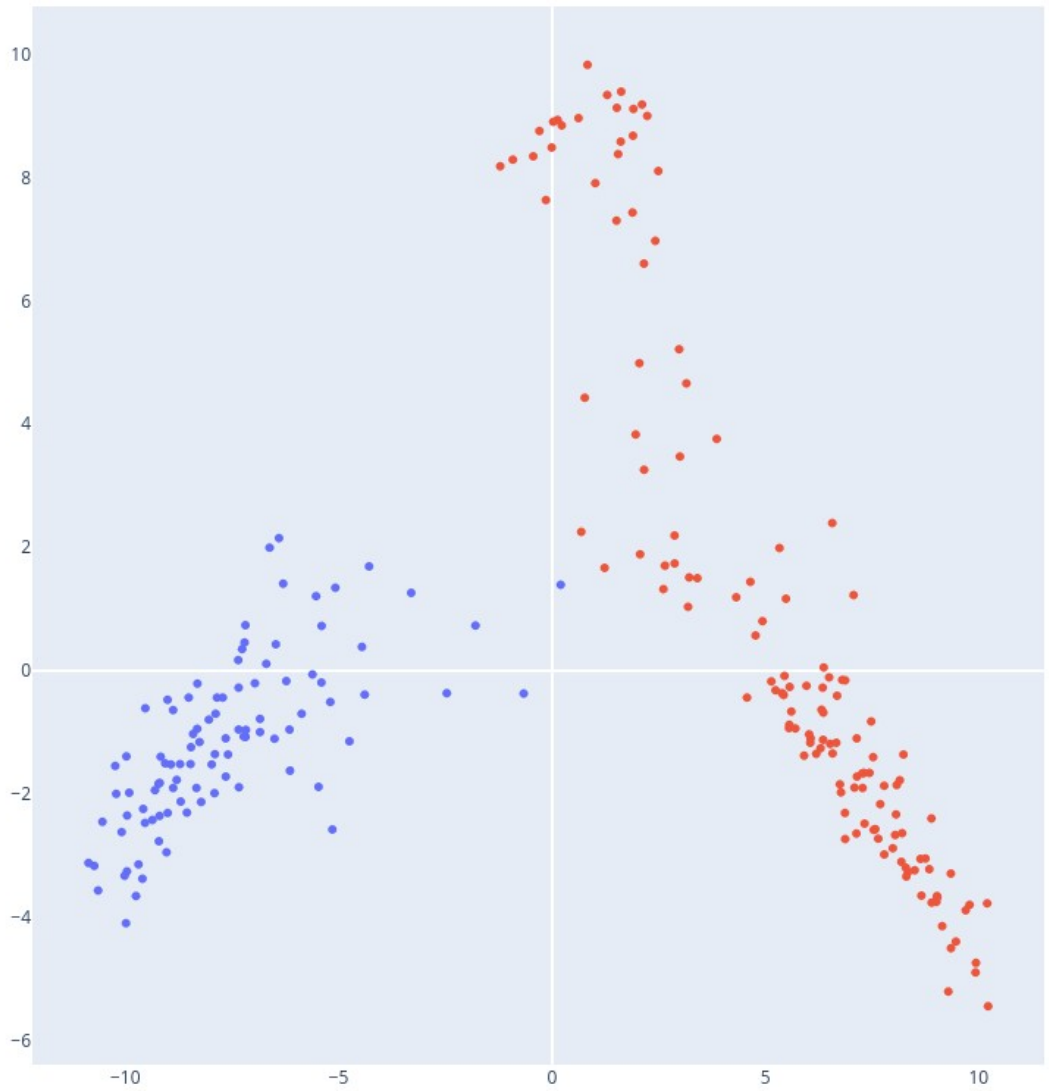


- 0
- 1
- 2

K-means clusters by dividing all vectors into groups where the distances to the centers are minimized while group memberships are equalized.

Here, k-means is applied on 2D data

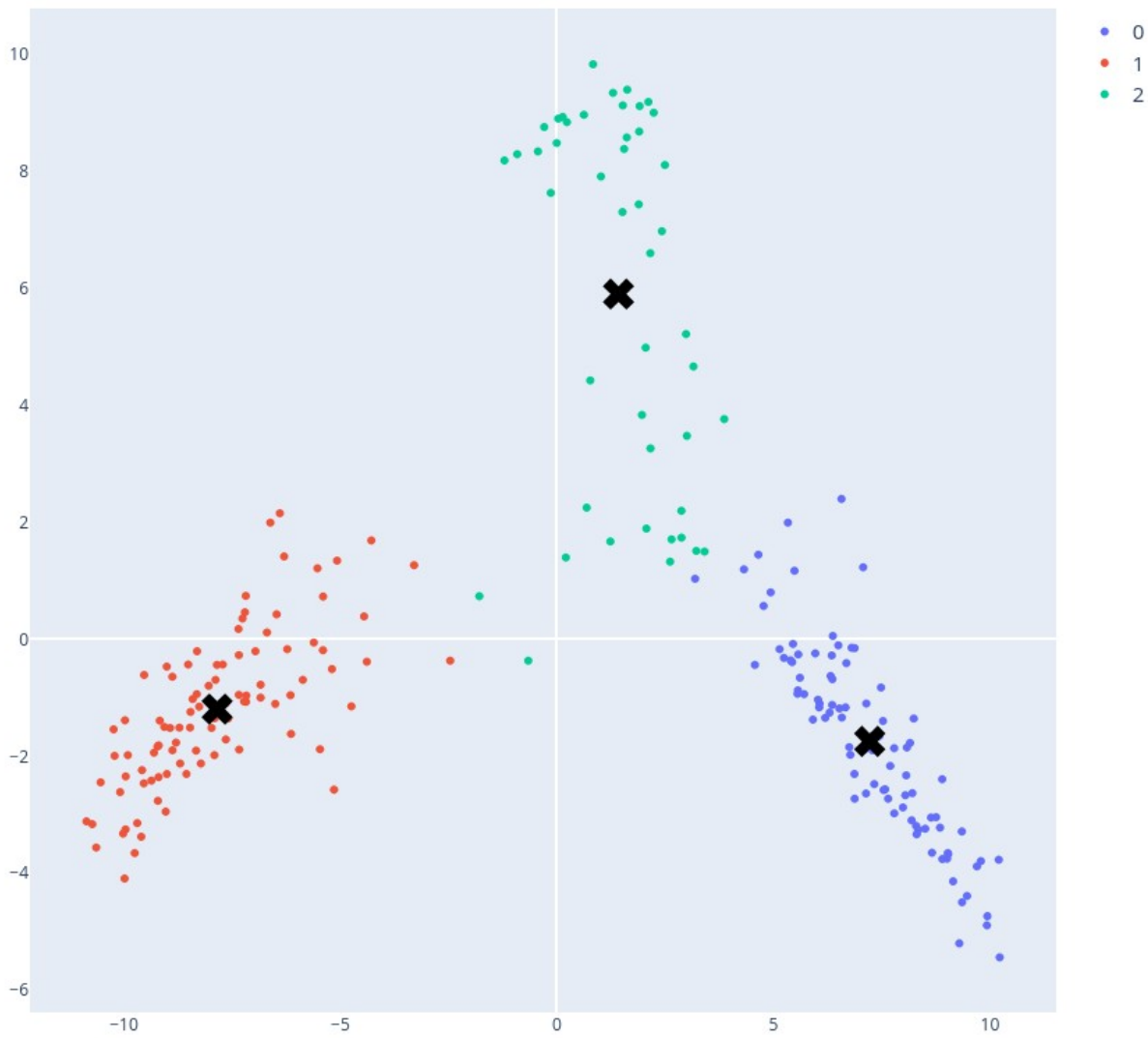
PCA reduction. No clustering



• Gemüse  
• Pilz

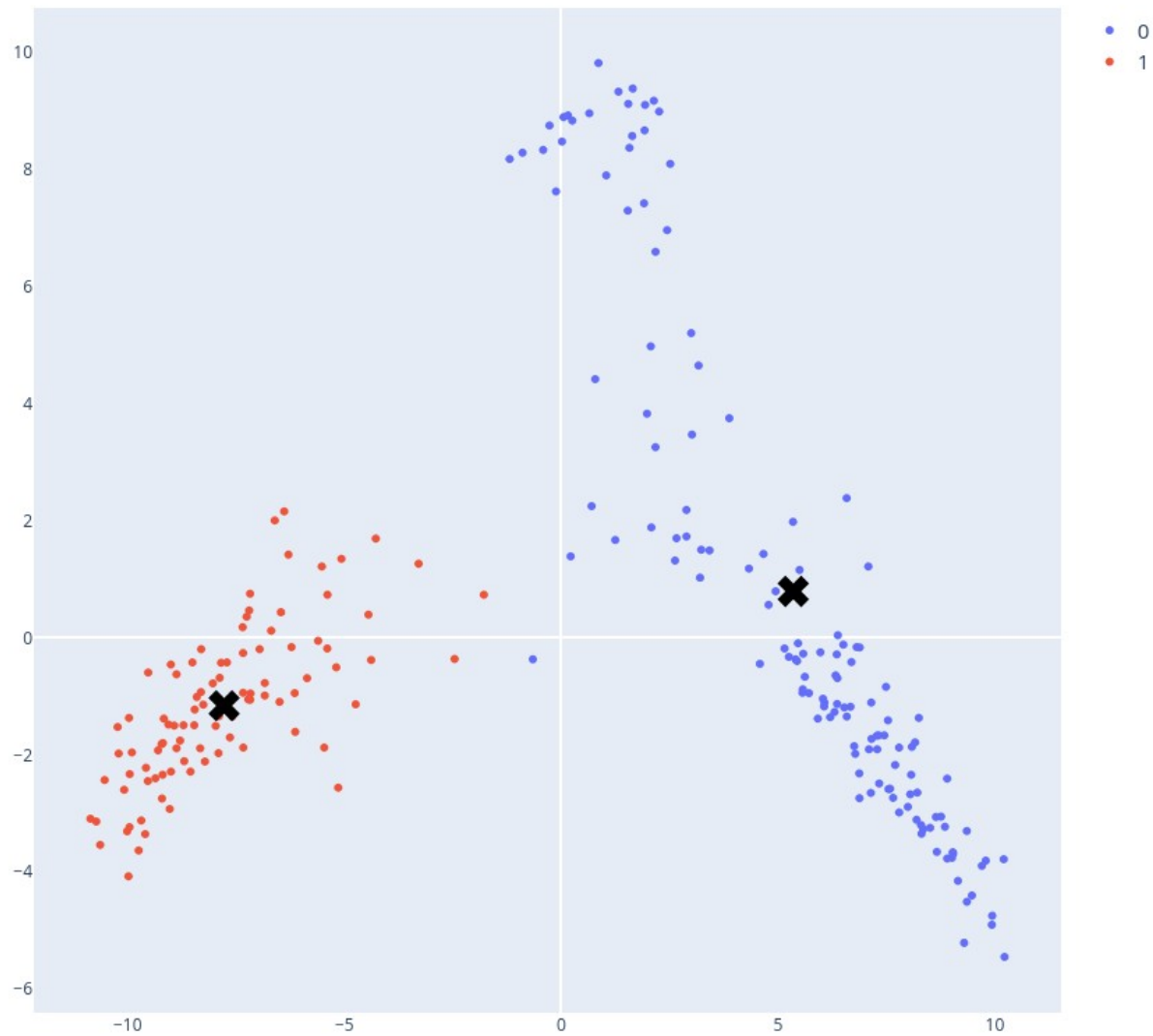
“Gemüse” to the left is mostly well aligned (by human judgements via samples). But “Pilz” has a mushroom cluster to the top and a politician cluster to the lower right.

PCA reduction. K-means clustering on reduced data



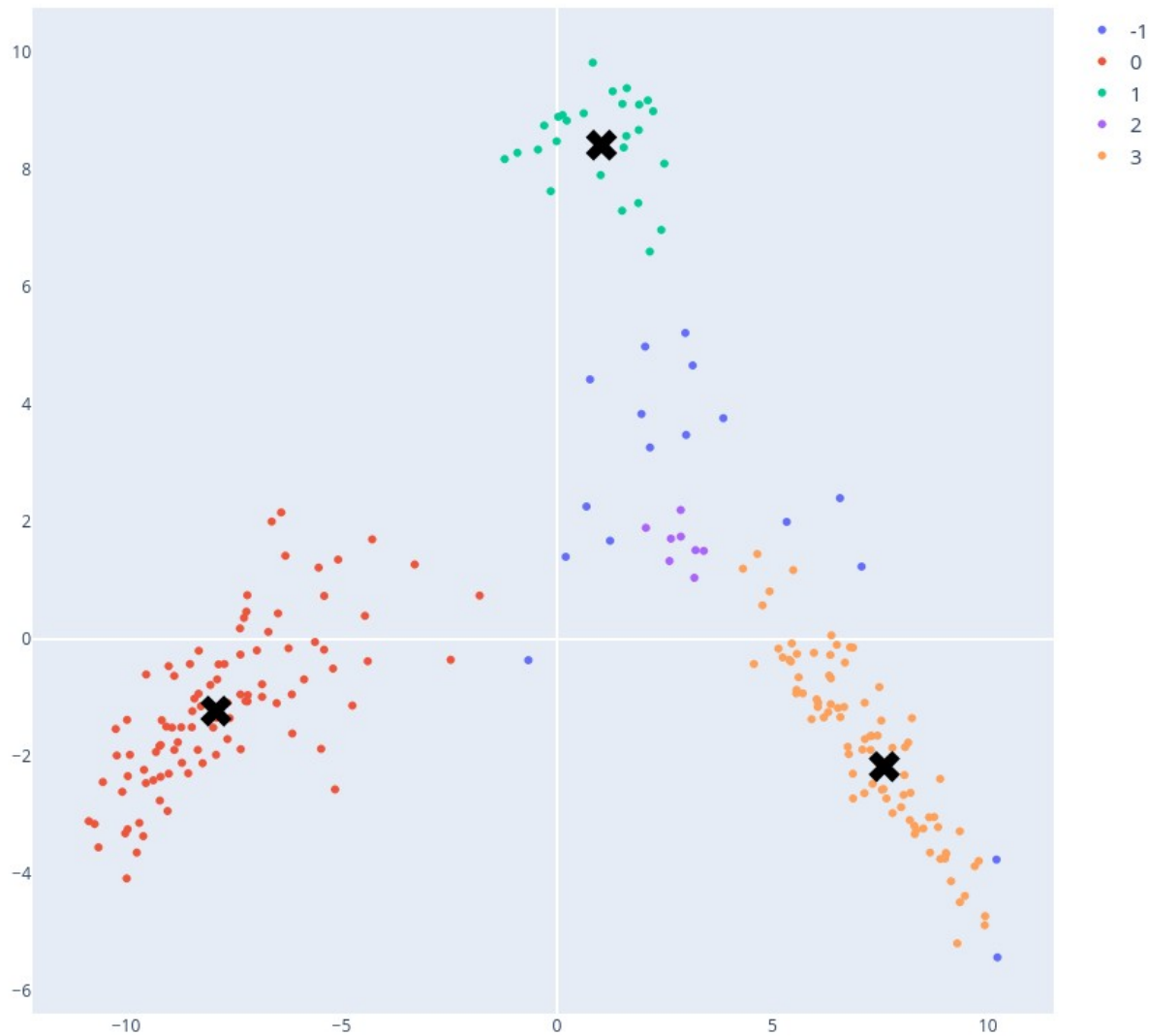
K-means works on 2D data somewhat. But a few samples from “politicians” are misgrouped to the “mushroom” cluster.

PCA reduction. K-means clustering on high dimensional data



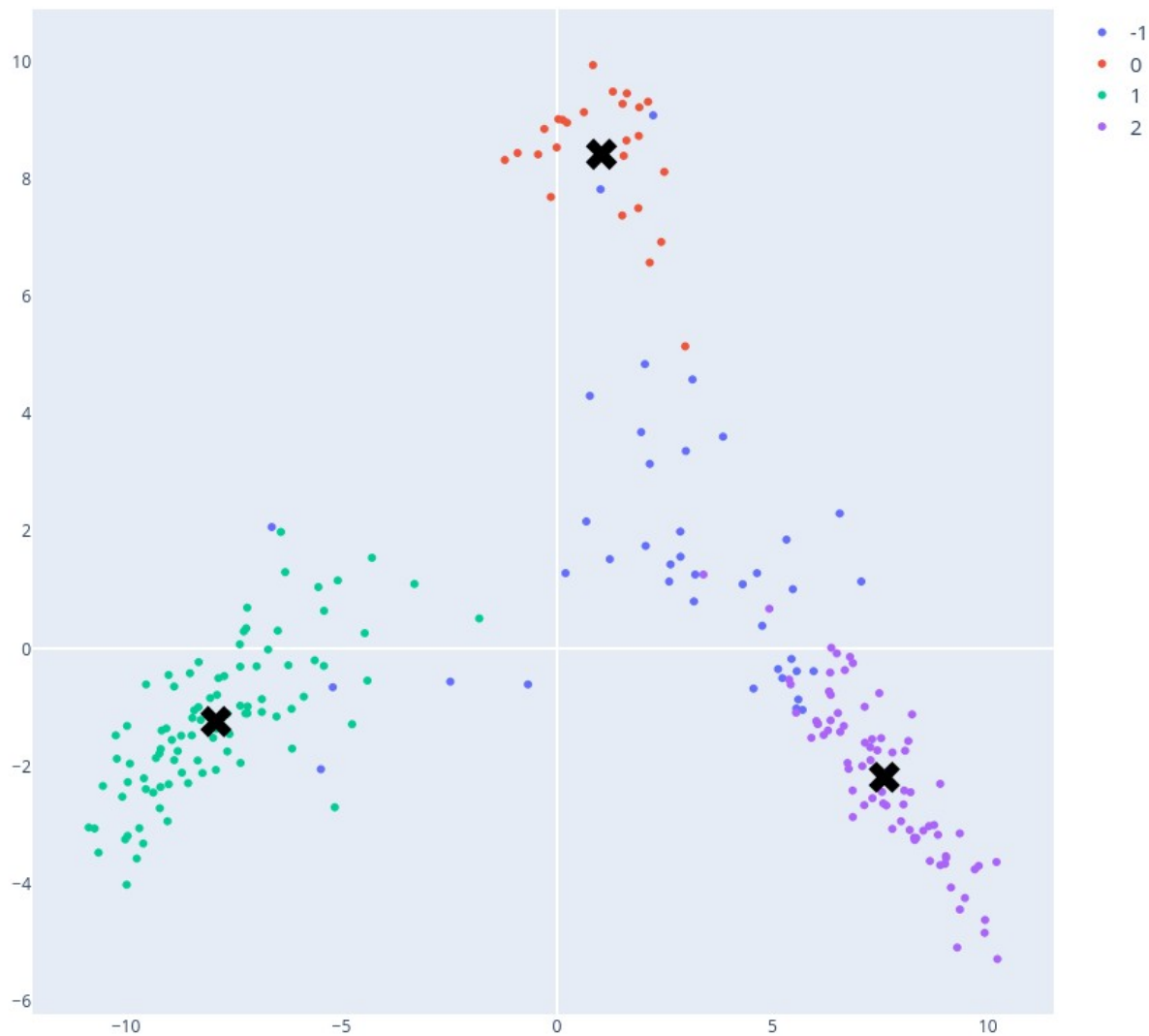
K-means performs worse on high dimensional data. Clusters all “Pilz” into one cluster.

PCA reduction. HDBSCAN clustering on reduced data



HDBSCAN works better, by focusing on varying density, instead of forcing all vectors into equally sized cluster.

Here, HDBSCAN is applied on 2D data.



HDBSCAN works better, by focusing on varying density, instead of forcing all vectors into equally sized cluster.

Here, HDBSCAN is applied on high dimensional data.

# t-SNE: "Leiter"

- cluster: 2864, lemma: Leiter
- cluster: 2865, lemma: Leiter

## Cluster 2864

Entlang des flach abfallenden Sandstrandes sind Kiosk, Sanitärbereiche, Umkleiden, Liegeflächen und ein Nichtschwim...

»Erdbestattung ist schädlich für die Umwelt, das ist korrekt«, sagt Andreas Waldherr, Leiter der Bestattung Kärnten, »bes...

Er entließ den langjährigen Leiter der Nationalgalerie in Prag, den Kunsthistoriker Jiri Fajt.

Erfreulich ist aber, dass wir hinten die Null halten konnten.“ Tschurndorfs Sportlicher Leiter Andreas Fellingner war mit de...

Er ist Gründer und Leiter des Arnold-Schönberg-Chors.

## Cluster 2865

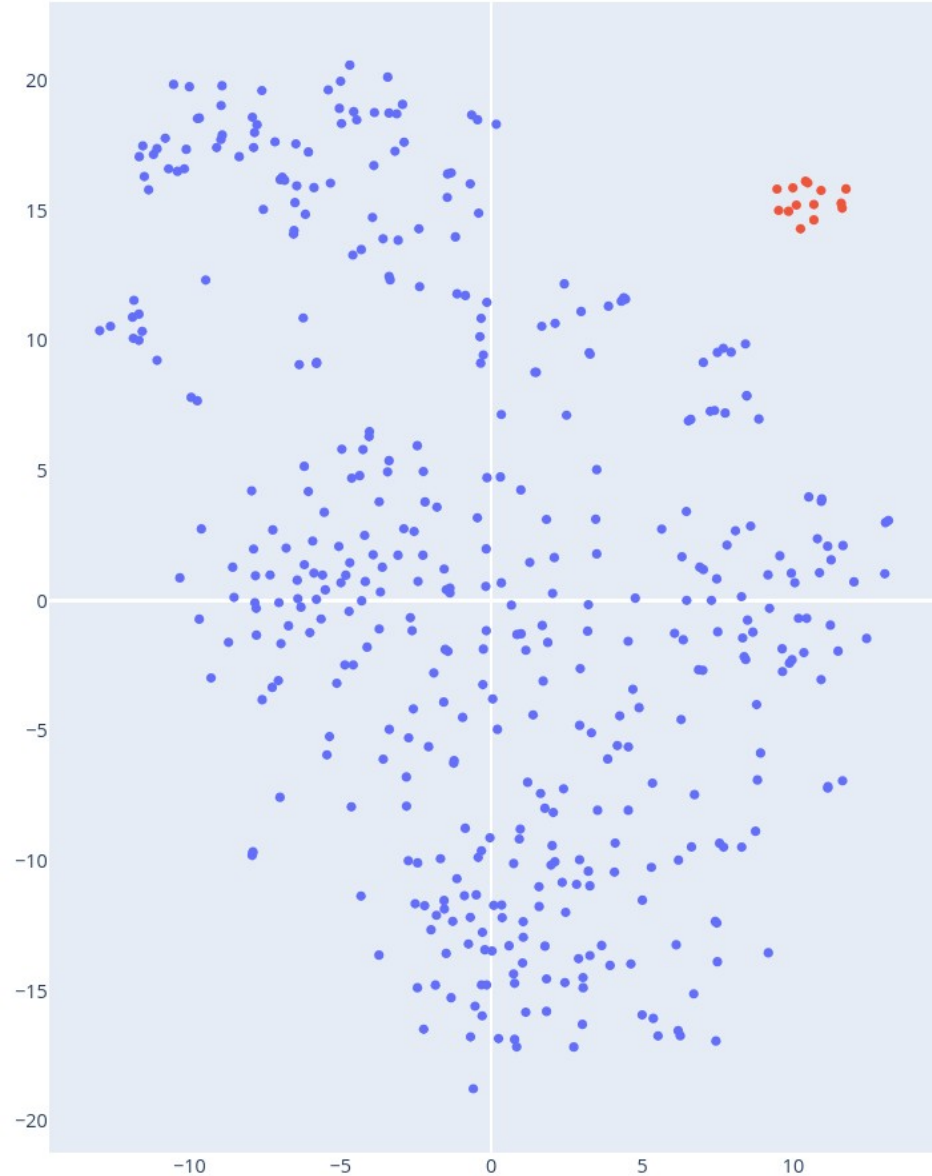
Der 25-Jährige, der seine Leiter an den Mast gelehnt hatte, stürzte mit ihm um und landete auf dem Boden.

Der Frühpensionist war beim Dachbodenausbau vom Gerüst auf die Alu-Leiter gestiegen, die wegrutschte.

Dort stellte der Täter seine Leiter auf, kletterte hoch und bedrohte von oben die Mitarbeiter mit einer Pistole.

Ein 83-jähriger Altbauer ist in Pettenbach (Bezirk Kirchdorf) beim Apfelpflücken von der Leiter gestürzt.

Ein Zimmerer (37) aus Weißensee stürzte Montag bei Schalungsarbeiten im Rohbaugebäude der Kläranlage Steinfeld von einer Leiter u



# PCA: "Leiter"

- cluster: 2864, lemma: Leiter
- cluster: 2865, lemma: Leiter

## Cluster 2864

Entlang des flach abfallenden Sandstrandes sind Kiosk, Sanitärbereiche, Umkleiden, Liegeflächen und ein Nichtschwim...

»Erdbestattung ist schädlich für die Umwelt, das ist korrekt«, sagt Andreas Waldherr, Leiter der Bestattung Kärnten, »bes...

Er entließ den langjährigen Leiter der Nationalgalerie in Prag, den Kunsthistoriker Jiri Fajt.

Erfreulich ist aber, dass wir hinten die Null halten konnten.“ Tschurndorfs Sportlicher Leiter Andreas Fellingner war mit de...

Er ist Gründer und Leiter des Arnold-Schönberg-Chors.

## Cluster 2865

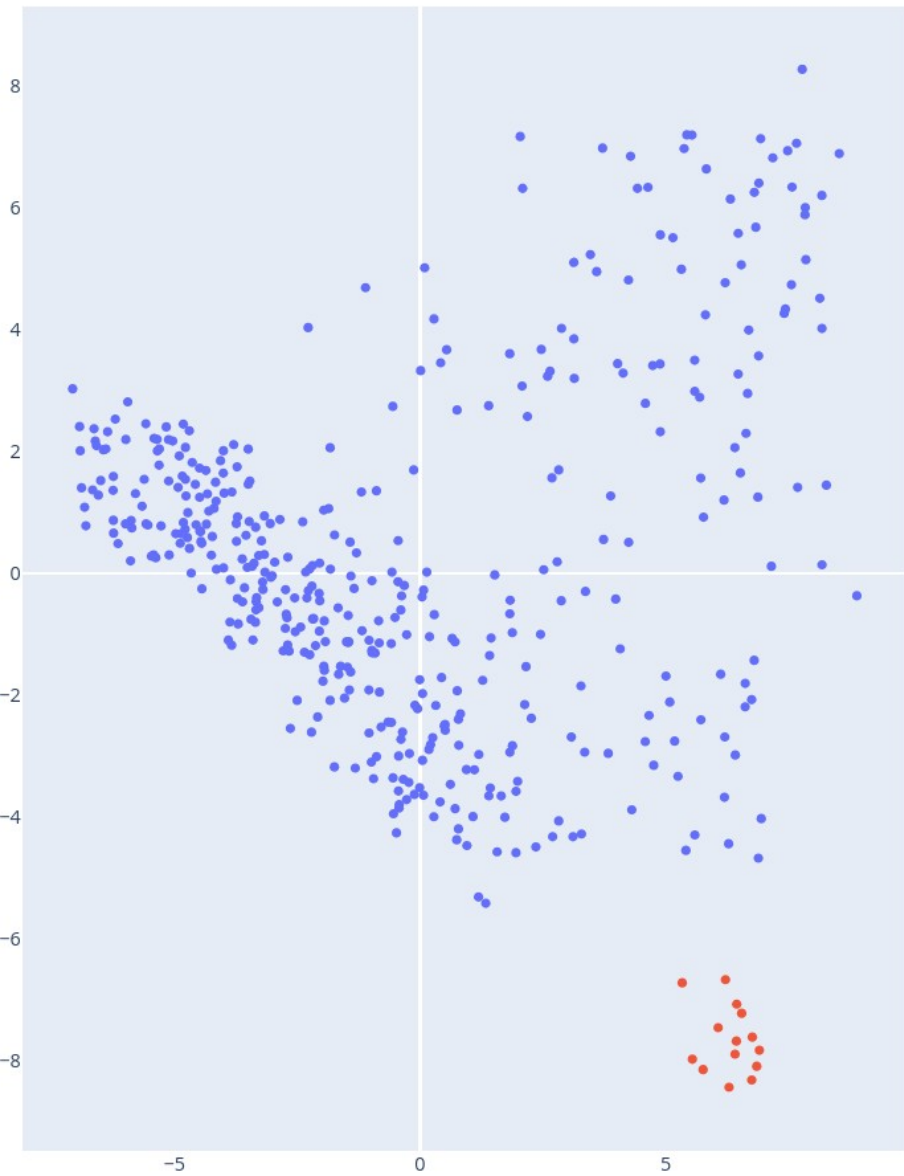
Der 25-Jährige, der seine Leiter an den Mast gelehnt hatte, stürzte mit ihm um und landete auf dem Boden.

Der Frühpensionist war beim Dachbodenausbau vom Gerüst auf die Alu-Leiter gestiegen, die wegrutschte.

Dort stellte der Täter seine Leiter auf, kletterte hoch und bedrohte von oben die Mitarbeiter mit einer Pistole.

Ein 83-jähriger Altbauer ist in Pettenbach (Bezirk Kirchdorf) beim Apfelpflücken von der Leiter gestürzt.

Ein Zimmerer (37) aus Weißensee stürzte Montag bei Schalungsarbeiten im Rohbaugebäude der Kläranlage Steinfeld von einer Leiter u



# UMAP: "Leiter"

- cluster: 2864, lemma: Leiter
- cluster: 2865, lemma: Leiter

## Cluster 2864

Entlang des flach abfallenden Sandstrandes sind Kiosk, Sanitärebereiche, Umkleiden, Liegeflächen und ein Nichtschwim...

»Erdbestattung ist schädlich für die Umwelt, das ist korrekt«, sagt Andreas Waldherr, Leiter der Bestattung Kärnten, »bes...

Er entließ den langjährigen Leiter der Nationalgalerie in Prag, den Kunsthistoriker Jiri Fajt.

Erfreulich ist aber, dass wir hinten die Null halten konnten.“ Tschurndorfs Sportlicher Leiter Andreas Fellingner war mit de...

Er ist Gründer und Leiter des Arnold-Schönberg-Chors.

## Cluster 2865

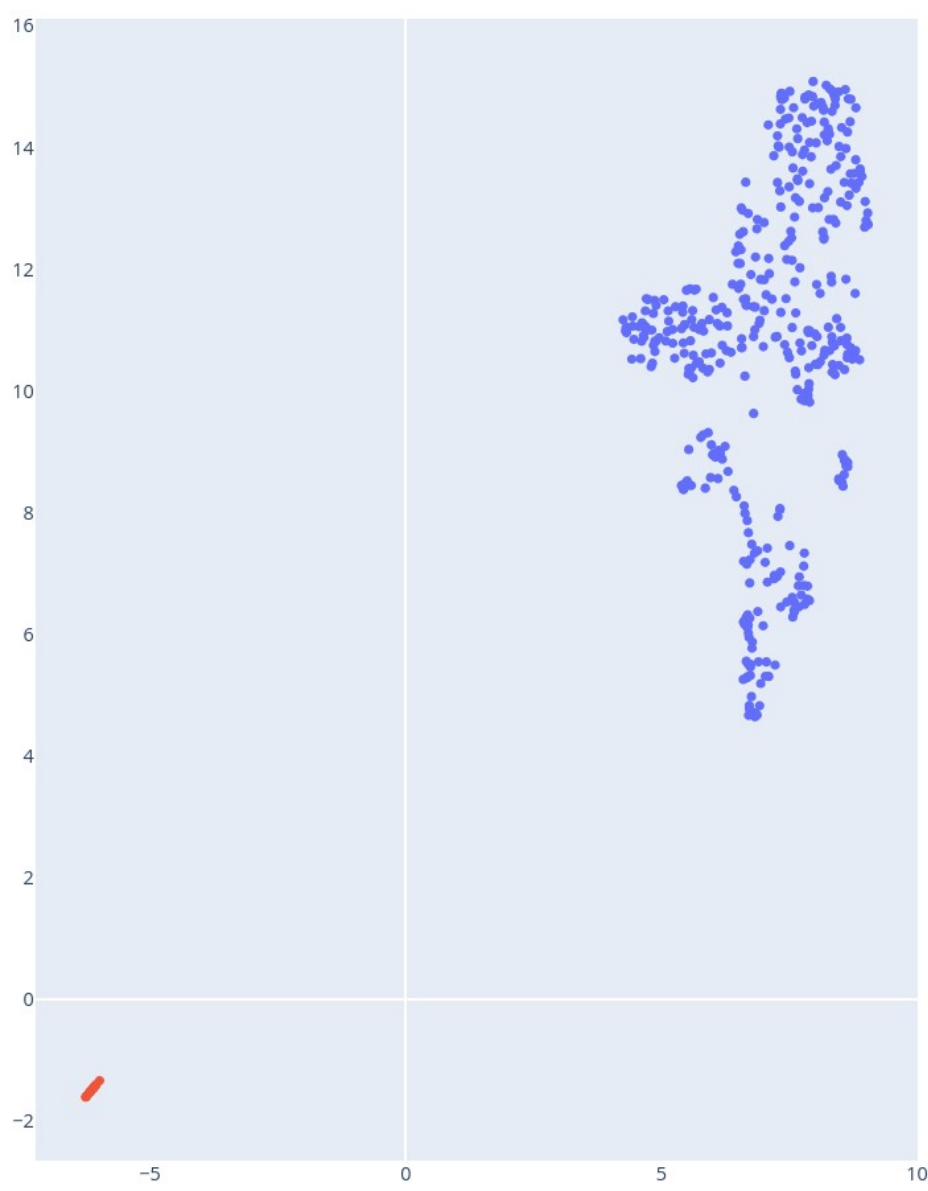
Der 25-Jährige, der seine Leiter an den Mast gelehnt hatte, stürzte mit ihm um und landete auf dem Boden.

Der Frühpensionist war beim Dachbodenausbau vom Gerüst auf die Alu-Leiter gestiegen, die wegrutschte.

Dort stellte der Täter seine Leiter auf, kletterte hoch und bedrohte von oben die Mitarbeiter mit einer Pistole.

Ein 83-jähriger Altbauer ist in Pettenbach (Bezirk Kirchdorf) beim Apfelpflücken von der Leiter gestürzt.

Ein Zimmerer (37) aus Weißensee stürzte Montag bei Schalungsarbeiten im Rohbaugebäude der Kläranlage Steinfeld von einer Leiter u...



# t-SNE:

## "Pilz", "Gemüse"

- cluster: 10379, lemma: Gemüse
- cluster: 10380, lemma: Gemüse
- cluster: 8212, lemma: Pilz
- cluster: 8213, lemma: Pilz
- cluster: 8214, lemma: Pilz

### Cluster 10379

Der Frost hat vor allem Kulturen wie Wein, Obst und Gemüse zugesetzt.

Der Klagenfurter Benediktinermarkt – er feiert heuer seinen 75. Geburtstag – ist viel mehr als nur Bauernbrot und buntes Gemüse.

Die Seminarbäuerin Sylvia Schilcher gibt Tipps und Tricks zum Einkochen von Obst und Gemüse.

### Cluster 10380

Ist das Gemüse fertig, streuen wir das Mehl gleichmäßig darüber, rösten alles unter ständigem Rühren durch, gießen mit zwei gro...

Schon am Vortag röstet er das Gemüse und die Lammknochen für seinen Fond an.

Und so geht's: Sud aus Rotkraut (färbt blau), Roten Rüben, Spinat oder Zwiebelschalen machen, indem man das Gemüse lange k...

### Cluster 8212

Utl.: Pilz und Jerusalem kritisieren Auslagerung an "Fonds Soziales Wien" und Umstellung auf Individualförderung =

Utl.: Pilz warf Staatsanwalt Klackl Ermittlungen unter Umgehung der Immunität vor - Klackl: Pilz war kein Beschuldigter im Sinn d...

Utl.: Kindische Manöver von Amon und Pilz geben Untersuchungen der Lächerlichkeit preis =

### Cluster 8213

Dabei handle es sich um das Ergebnis jenes seit Tagen andauernden Prozesses, der die Struktur der Liste Pilz – neben dem Klub ...

Da gebe ich dem Herrn Pilz recht, weil das war ja das, was wir... was wir zuerst schon angesprochen haben, unter Schwarz-Blau e...

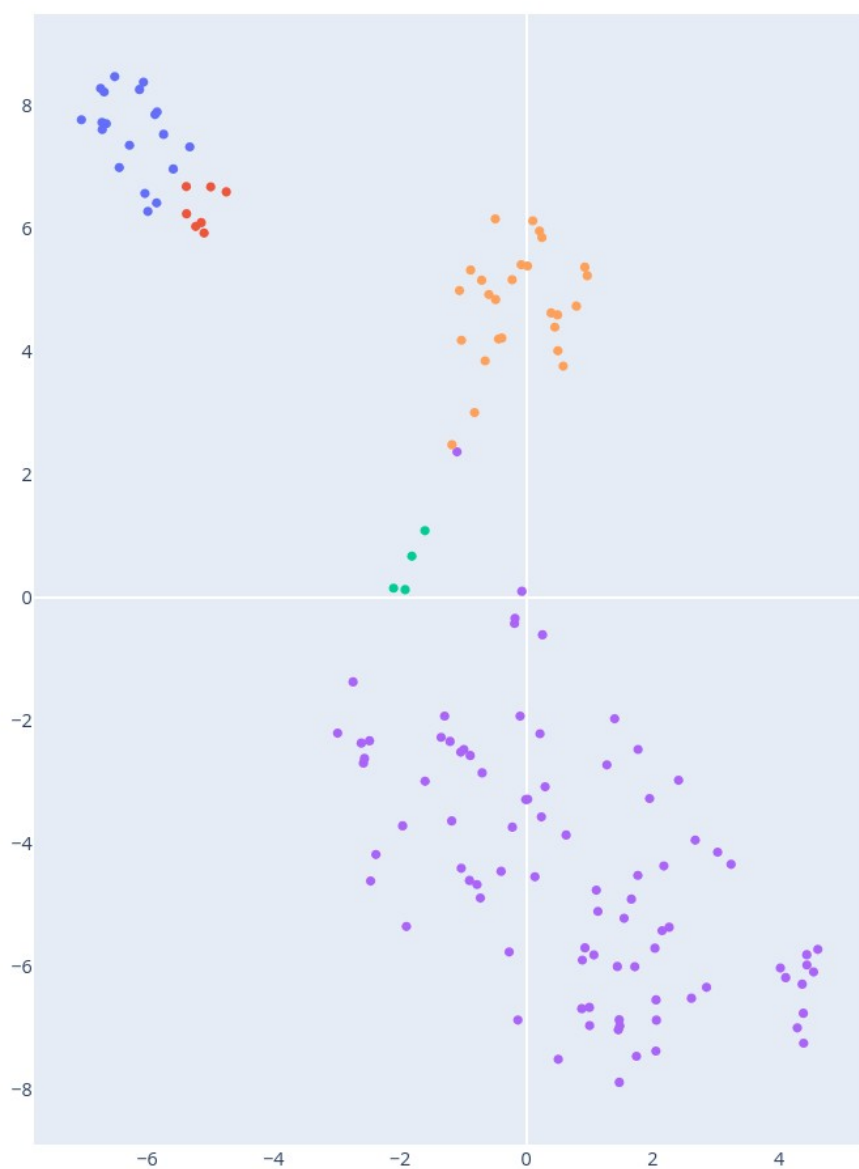
"Das politische Ende des Peter Pilz" von Karl Ettinger, 26. 6. Einen Politiker wie Peter Pilz auf diese s. g. "demokratische" Art zu "d...

### Cluster 8214

Denn der Pilz besteht aus drei Komponenten: Das Schwammerl, das wir kennen, ist nur der Fruchtkörper.

Der Grund für diese strenge Maßnahme liegt in der Einschleppungsgefahr von gefährlichen Schadorganismen, z. B. Viren, Pilzen ...

Die Kalbsvögel auf der Tomaten-Pilz-Mischung anrichten.



# PCA:

## "Pilz", "Gemüse"

- cluster: 10379, lemma: Gemüse
- cluster: 10380, lemma: Gemüse
- cluster: 8212, lemma: Pilz
- cluster: 8213, lemma: Pilz
- cluster: 8214, lemma: Pilz

### Cluster 10379

Der Frost hat vor allem Kulturen wie Wein, Obst und Gemüse zugesetzt.

Der Klagenfurter Benediktinermarkt – er feiert heuer seinen 75. Geburtstag – ist viel mehr als nur Bauernbrot und buntes Gemüse.

Die Seminarbäuerin Sylvia Schilcher gibt Tipps und Tricks zum Einkochen von Obst und Gemüse.

### Cluster 10380

Ist das Gemüse fertig, streuen wir das Mehl gleichmäßig darüber, rösten alles unter ständigem Rühren durch, gießen mit zwei gro...

Schon am Vortag röstet er das Gemüse und die Lammknochen für seinen Fond an.

Und so geht's: Sud aus Rotkraut (färbt blau), Roten Rüben, Spinat oder Zwiebelschalen machen, indem man das Gemüse lange k...

### Cluster 8212

Utl.: Pilz und Jerusalem kritisieren Auslagerung an "Fonds Soziales Wien" und Umstellung auf Individualförderung =

Utl.: Pilz warf Staatsanwalt Klackl Ermittlungen unter Umgehung der Immunität vor - Klackl: Pilz war kein Beschuldigter im Sinn d...

Utl.: Kindische Manöver von Amon und Pilz geben Untersuchungen der Lächerlichkeit preis =

### Cluster 8213

Dabei handle es sich um das Ergebnis jenes seit Tagen andauernden Prozesses, der die Struktur der Liste Pilz – neben dem Klub ...

Da gebe ich dem Herrn Pilz recht, weil das war ja das, was wir... was wir zuerst schon angesprochen haben, unter Schwarz-Blau e...

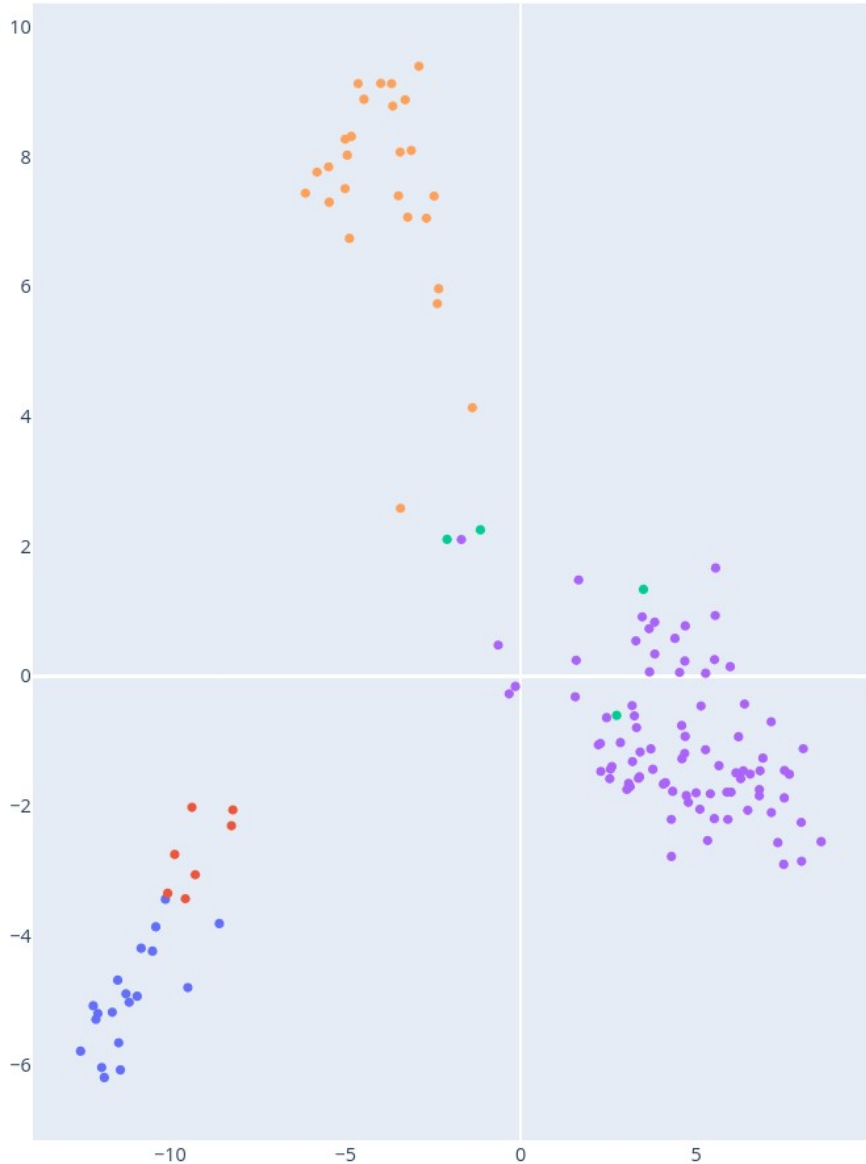
"Das politische Ende des Peter Pilz" von Karl Ettinger, 26. 6. Einen Politiker wie Peter Pilz auf diese s. g. "demokratische" Art zu "d...

### Cluster 8214

Denn der Pilz besteht aus drei Komponenten: Das Schwammerl, das wir kennen, ist nur der Fruchtkörper.

Der Grund für diese strenge Maßnahme liegt in der Einschleppungsgefahr von gefährlichen Schadorganismen, z. B. Viren, Pilzen ...

Die Kalbsvögel auf der Tomaten-Pilz-Mischung anrichten.



# UMAP:

## "Pilz", "Gemüse"

- cluster: 10379, lemma: Gemüse
- cluster: 10380, lemma: Gemüse
- cluster: 8212, lemma: Pilz
- cluster: 8213, lemma: Pilz
- cluster: 8214, lemma: Pilz

### Cluster 10379

Der Frost hat vor allem Kulturen wie Wein, Obst und Gemüse zugesetzt.

Der Klagenfurter Benediktinermarkt – er feiert heuer seinen 75. Geburtstag – ist viel mehr als nur Bauernbrot und buntes Gemüse.

Die Seminarbäuerin Sylvia Schilcher gibt Tipps und Tricks zum Einkochen von Obst und Gemüse.

### Cluster 10380

Ist das Gemüse fertig, streuen wir das Mehl gleichmäßig darüber, rösten alles unter ständigem Rühren durch, gießen mit zwei gro...

Schon am Vortag röstet er das Gemüse und die Lammknochen für seinen Fond an.

Und so geht's: Sud aus Rotkraut (färbt blau), Roten Rüben, Spinat oder Zwiebelschalen machen, indem man das Gemüse lange k...

### Cluster 8212

Utl.: Pilz und Jerusalem kritisieren Auslagerung an "Fonds Soziales Wien" und Umstellung auf Individualförderung =

Utl.: Pilz warf Staatsanwalt Klackl Ermittlungen unter Umgehung der Immunität vor - Klackl: Pilz war kein Beschuldigter im Sinn d...

Utl.: Kindische Manöver von Amon und Pilz geben Untersuchungen der Lächerlichkeit preis =

### Cluster 8213

Dabei handle es sich um das Ergebnis jenes seit Tagen andauernden Prozesses, der die Struktur der Liste Pilz – neben dem Klub ...

Da gebe ich dem Herrn Pilz recht, weil das war ja das, was wir... was wir zuerst schon angesprochen haben, unter Schwarz-Blau e...

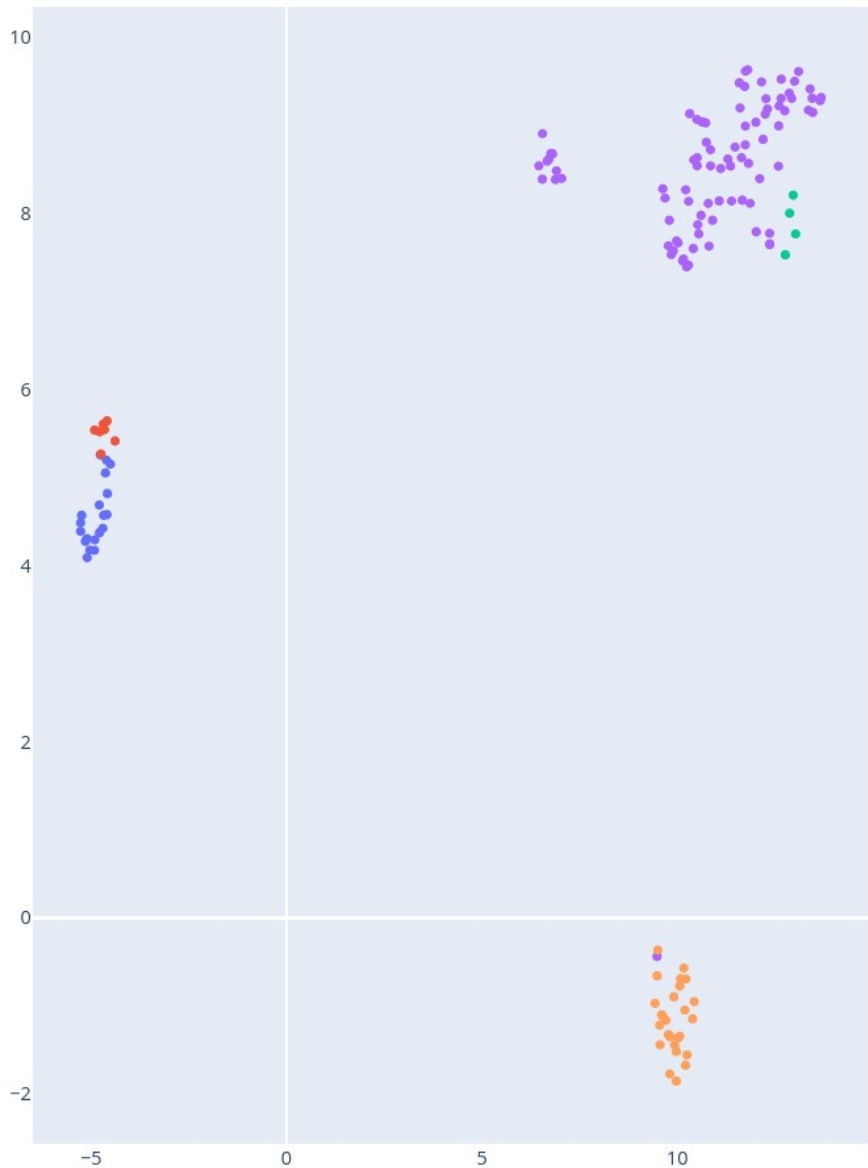
"Das politische Ende des Peter Pilz" von Karl Ettinger, 26. 6. Einen Politiker wie Peter Pilz auf diese s. g. "demokratische" Art zu "d...

### Cluster 8214

Denn der Pilz besteht aus drei Komponenten: Das Schwammerl, das wir kennen, ist nur der Fruchtkörper.

Der Grund für diese strenge Maßnahme liegt in der Einschleppungsgefahr von gefährlichen Schadorganismen, z. B. Viren, Pilzen ...

Die Kalbsvögel auf der Tomaten-Pilz-Mischung anrichten.



# T-SNE: "Ball"

- cluster: 3637, lemma: Ball
- cluster: 3638, lemma: Ball
- cluster: 3639, lemma: Ball
- cluster: 3640, lemma: Ball
- cluster: 3641, lemma: Ball

## Cluster 3637

Wir haben im Verein sehr strenge Compliance Regeln und unsere Finanzgebarung hält jeder Prüfung Stand. Auch der Life Ball 20...

"Dieses einmalige Projekt des "Life Ball Shops" halten wir für eine sehr gute Idee, um einerseits das Thema HIV und AIDS nicht n...

Am 10. Juni öffnet das Wiener Rathaus die Pforten zum großen Life Ball 2017 - schon gestern wurde die "Life Ball"-Woche mit de...

## Cluster 3638

Schlierbach Ball 14:00 Veranstaltungssaal, Stiftskeller, Klosterstraße 1 Ball des Seniorenbundes Schlierbach

St. Florian am Inn Ball 20:00 Landhotel, Sankt Florian am Inn 18 Feuerwehrball

Ulrichsberg Ball 19:30 Pfarrzentrum, Markt 35 Rotkreuz Ball

## Cluster 3639

Sabitzer und Arnautovic machen über die Seiten Dampf, Schöpf ist als Zehner für die Bälle aus der Tiefe zuständig.

Mödling wirkte unkoordiniert und zweikampfschwach. 80 Prozent (!) der Bälle gingen verloren.

Nach dem Seitenwechsel hatte Willendorf nur noch wenig entgegensetzen. „Wir haben die Bälle im Spielaufbau zu einfach verl...

## Cluster 3640

Der will auch immer den Ball, kickt aber gleichermaßen genial und systemdienlich.

Die große Ehre, den diesjährigen Ball mit dem Auftanz um 20.30 Uhr in der Dogana zu eröffnen, haben heuer 55 Paare der Jungb...

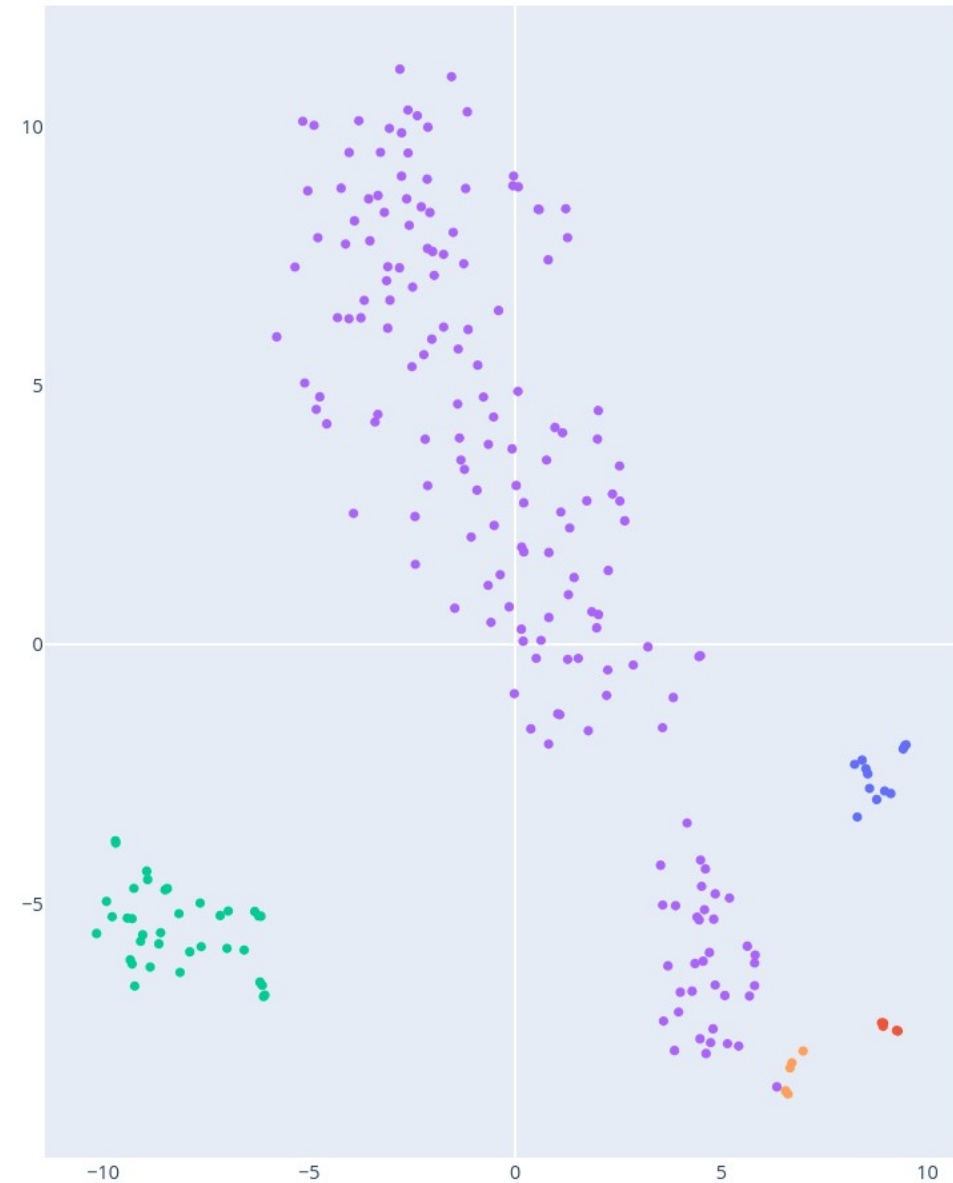
Die Handelskammer veranstalte ihren Ball im Waldorf Astoria heuer zum 48. Mal und sei sich des Erfolgs ihres Konzepts sicher.

## Cluster 3641

BALL Salzburger Kongresshaus, 13.00, Ball des Salzburger Pensionistenbundes.

Grafenegg: Ball des Piaristengymnasiums, Motto: "American Beauty", Schloss, 20 Uhr.

Krumau: Ball der Feuerwehr Preinreichs am Sa ab 20 Uhr im Gashaus Hauer.



# PCA: "Ball"

- cluster: 3637, lemma: Ball
- cluster: 3638, lemma: Ball
- cluster: 3639, lemma: Ball
- cluster: 3640, lemma: Ball
- cluster: 3641, lemma: Ball

## Cluster 3637

Wir haben im Verein sehr strenge Compliance Regeln und unsere Finanzgebarung hält jeder Prüfung Stand. Auch der Life Ball 20...

"Dieses einmalige Projekt des "Life Ball Shops" halten wir für eine sehr gute Idee, um einerseits das Thema HIV und AIDS nicht n...

Am 10. Juni öffnet das Wiener Rathaus die Pforten zum großen Life Ball 2017 - schon gestern wurde die "Life Ball"-Woche mit de...

## Cluster 3638

Schlierbach Ball 14:00 Veranstaltungssaal, Stiftskeller, Klosterstraße 1 Ball des Seniorenbundes Schlierbach

St. Florian am Inn Ball 20:00 Landhotel, Sankt Florian am Inn 18 Feuerwehrball

Ulrichsberg Ball 19:30 Pfarrzentrum, Markt 35 Rotkreuz Ball

## Cluster 3639

Sabitzer und Arnautovic machen über die Seiten Dampf, Schöpf ist als Zehner für die Bälle aus der Tiefe zuständig.

Mödling wirkte unkoordiniert und zweikampfschwach. 80 Prozent (!) der Bälle gingen verloren.

Nach dem Seitenwechsel hatte Willendorf nur noch wenig entgegenzusetzen. „Wir haben die Bälle im Spielaufbau zu einfach verl...

## Cluster 3640

Der will auch immer den Ball, kickt aber gleichermaßen genial und systemdienlich.

Die große Ehre, den diesjährigen Ball mit dem Auftanz um 20.30 Uhr in der Dogana zu eröffnen, haben heuer 55 Paare der Jungb...

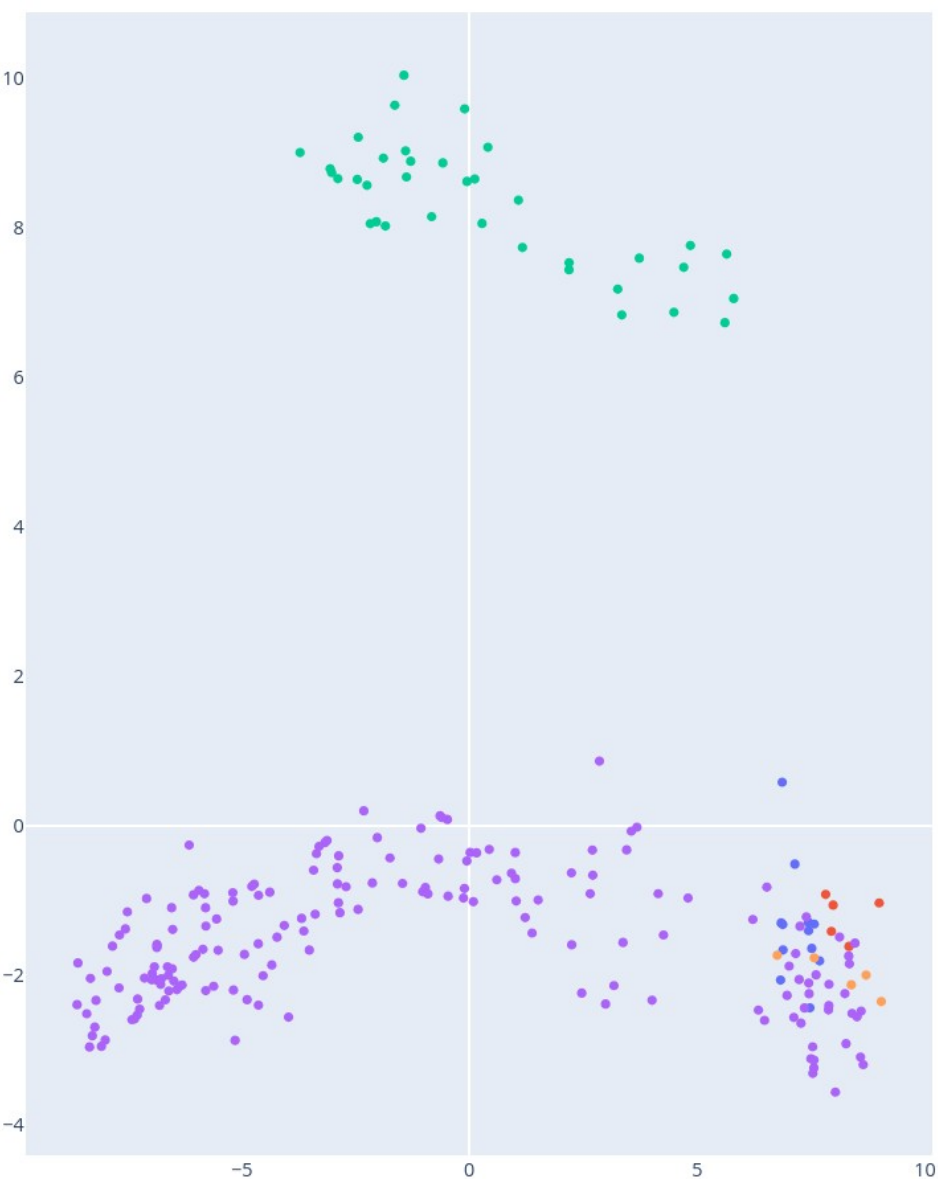
Die Handelskammer veranstalte ihren Ball im Waldorf Astoria heuer zum 48. Mal und sei sich des Erfolgs ihres Konzepts sicher.

## Cluster 3641

BALL Salzburger Kongresshaus, 13.00, Ball des Salzburger Pensionistenbundes.

Grafenegg: Ball des Piaristengymnasiums, Motto: "American Beauty", Schloss, 20 Uhr.

Krumau: Ball der Feuerwehr Preinreichs am Sa ab 20 Uhr im Gashaus Hauer.



# UMAP: "Ball"

- cluster: 3637, lemma: Ball
- cluster: 3638, lemma: Ball
- cluster: 3639, lemma: Ball
- cluster: 3640, lemma: Ball
- cluster: 3641, lemma: Ball

## Cluster 3637

Wir haben im Verein sehr strenge Compliance Regeln und unsere Finanzgebarung hält jeder Prüfung Stand. Auch der Life Ball 20...

"Dieses einmalige Projekt des "Life Ball Shops" halten wir für eine sehr gute Idee, um einerseits das Thema HIV und AIDS nicht n...

Am 10. Juni öffnet das Wiener Rathaus die Pforten zum großen Life Ball 2017 - schon gestern wurde die "Life Ball"-Woche mit de...

## Cluster 3638

Schlierbach Ball 14:00 Veranstaltungssaal, Stiftskeller, Klosterstraße 1 Ball des Seniorenbundes Schlierbach

St. Florian am Inn Ball 20:00 Landhotel, Sankt Florian am Inn 18 Feuerwehrball

Ulrichsberg Ball 19:30 Pfarrzentrum, Markt 35 Rotkreuz Ball

## Cluster 3639

Sabitzer und Arnautovic machen über die Seiten Dampf, Schöpf ist als Zehner für die Bälle aus der Tiefe zuständig.

Mödling wirkte unkoordiniert und zweikampfschwach. 80 Prozent (!) der Bälle gingen verloren.

Nach dem Seitenwechsel hatte Willendorf nur noch wenig entgegensetzen. „Wir haben die Bälle im Spielaufbau zu einfach verl...

## Cluster 3640

Der will auch immer den Ball, kickt aber gleichermaßen genial und systemdienlich.

Die große Ehre, den diesjährigen Ball mit dem Auftanz um 20.30 Uhr in der Dogana zu eröffnen, haben heuer 55 Paare der Jungb...

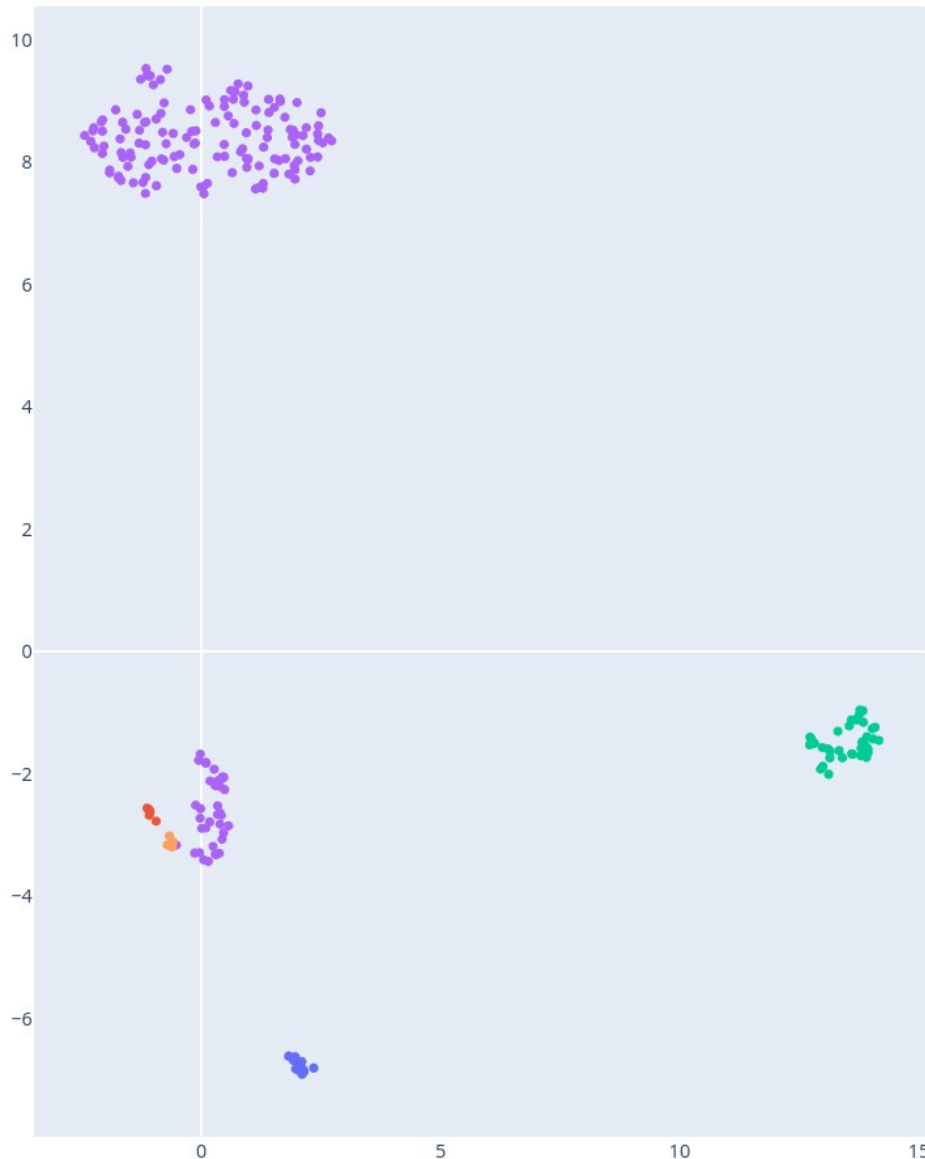
Die Handelskammer veranstalte ihren Ball im Waldorf Astoria heuer zum 48. Mal und sei sich des Erfolgs ihres Konzepts sicher.

## Cluster 3641

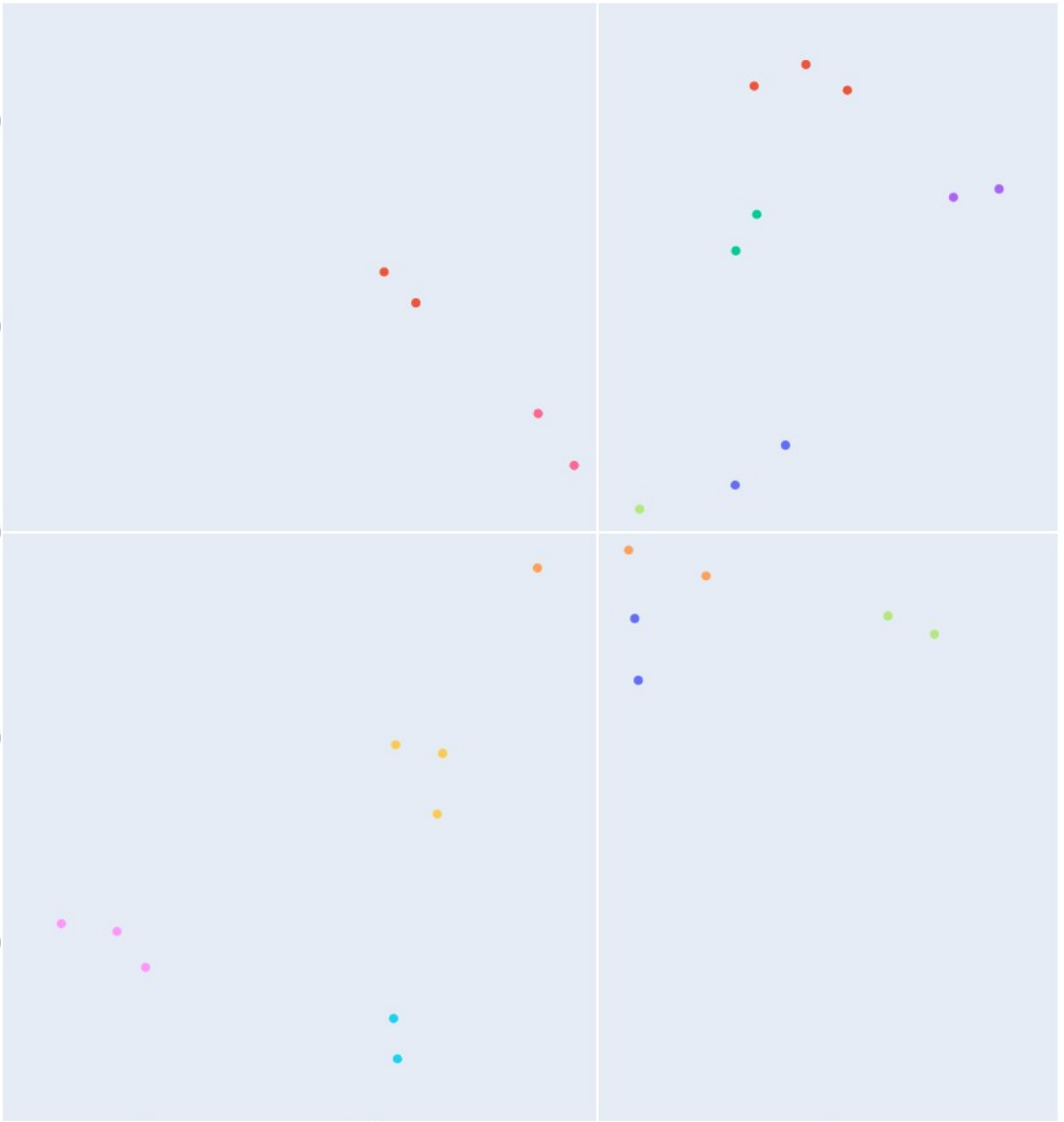
BALL Salzburger Kongresshaus, 13.00, Ball des Salzburger Pensionistenbundes.

Grafenegg: Ball des Piaristengymnasiums, Motto: "American Beauty", Schloss, 20 Uhr.

Krumau: Ball der Feuerwehr Preinreichs am Sa ab 20 Uhr im Gashaus Hauer.



# Several Cluster Centers



- Arzt
- Auto
- Bus
- Flugzeug
- Frau
- Gemüse
- Kind
- Mann
- Obst
- Pilz
- Politiker
- Techniker

## Outlook

- Scale up on HPC (CLIP, MUSICA)
- Improve clustering parameters
- Evaluate against existing hand-picked data sets and GermaNet
- Extend contextual data (time, region, newspaper)