

General stats about the dataset

How many samples

10 000

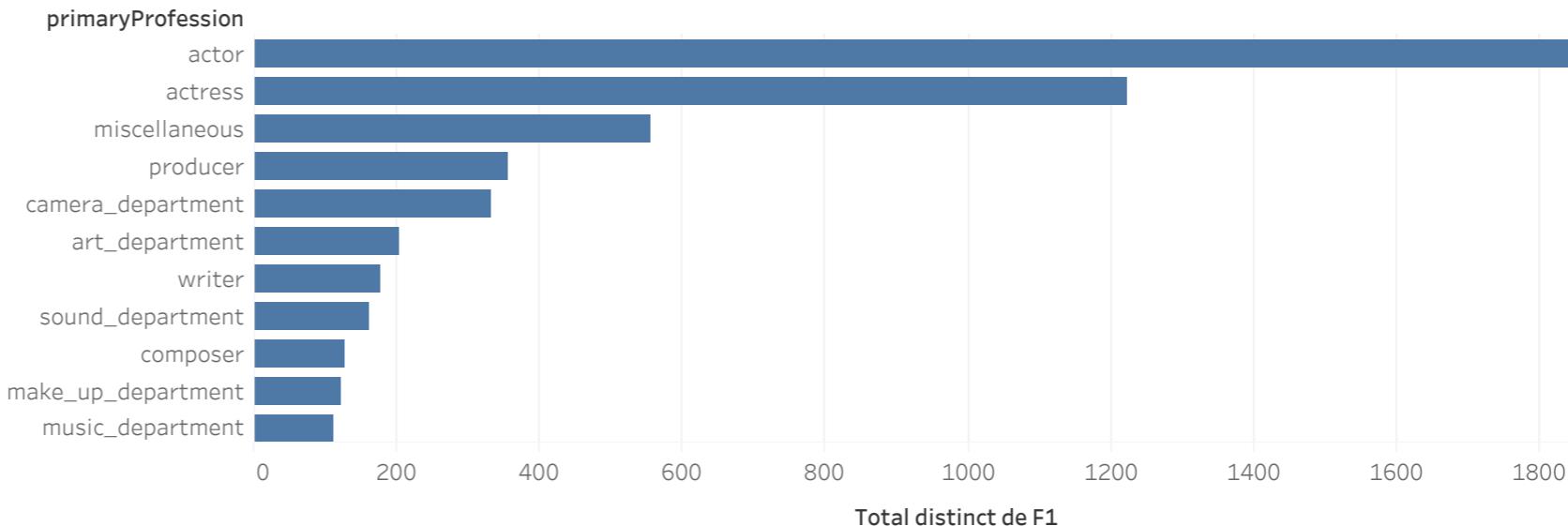
How many names

8 106

How many surnames

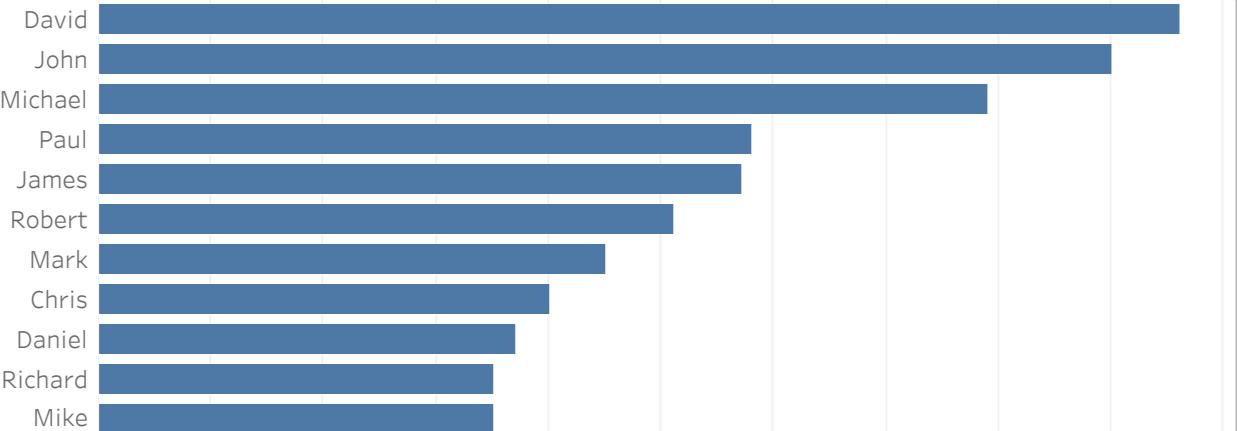
4 422

main professions



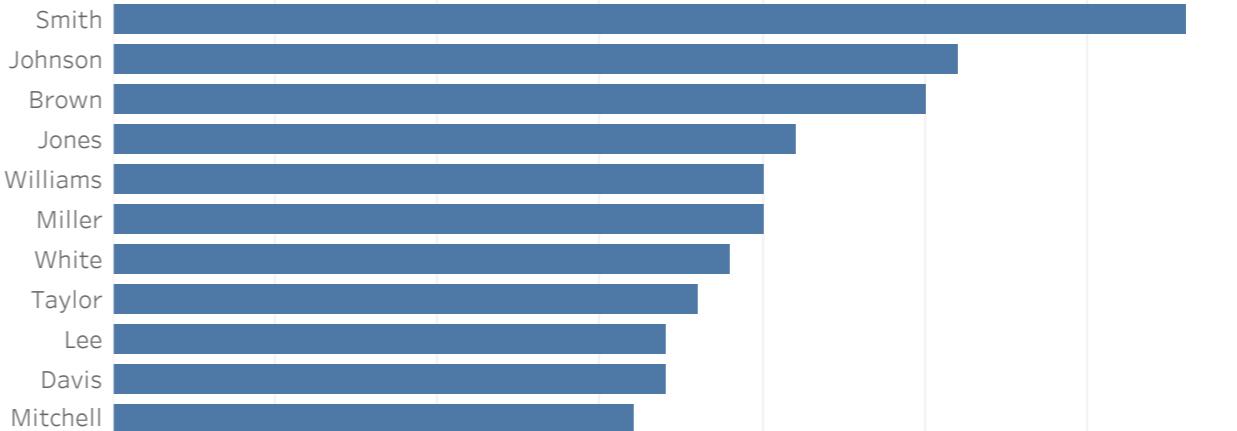
Top prenom

prenom



Top nom

nom



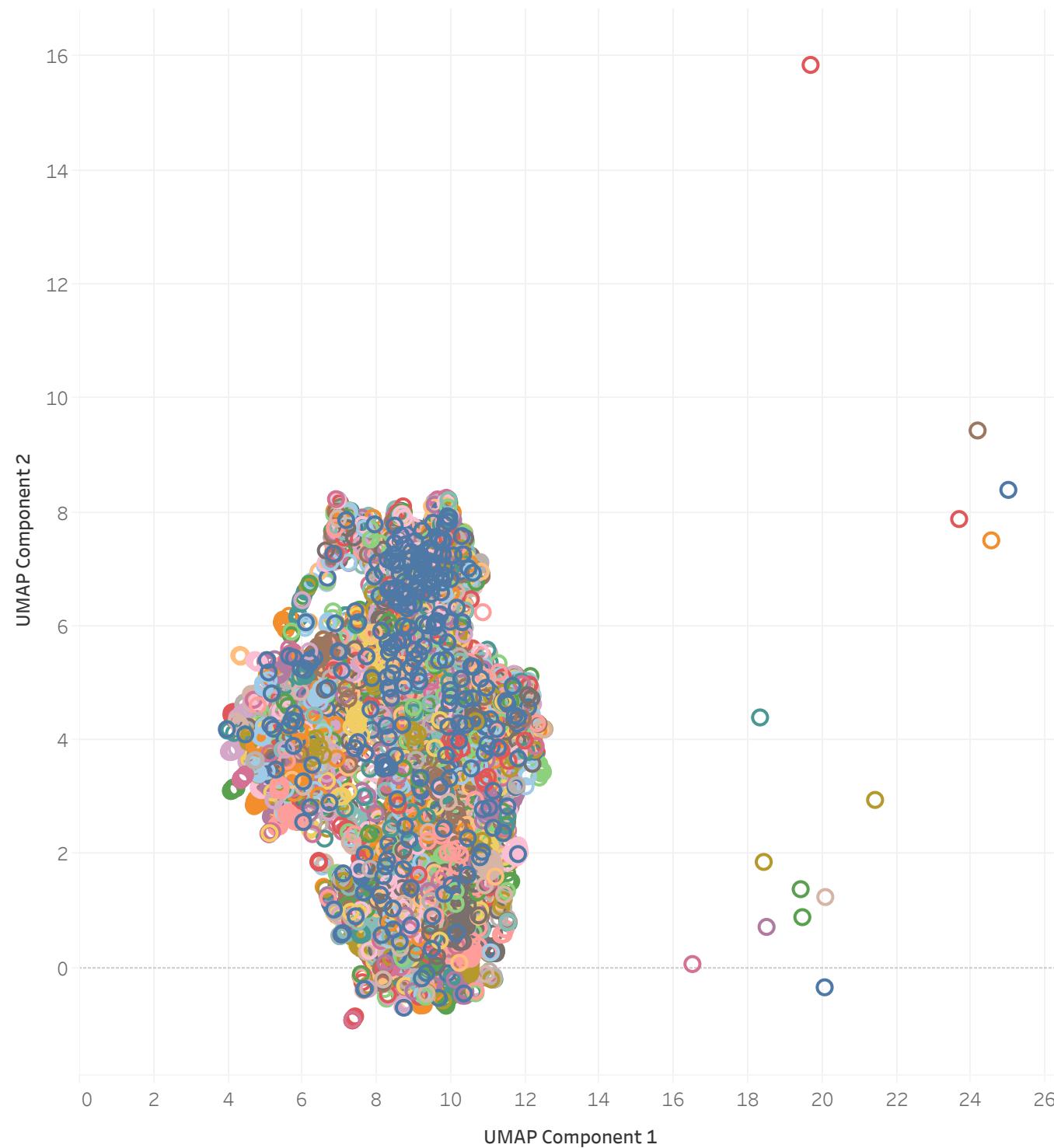
Dataset extract

General dataset overview

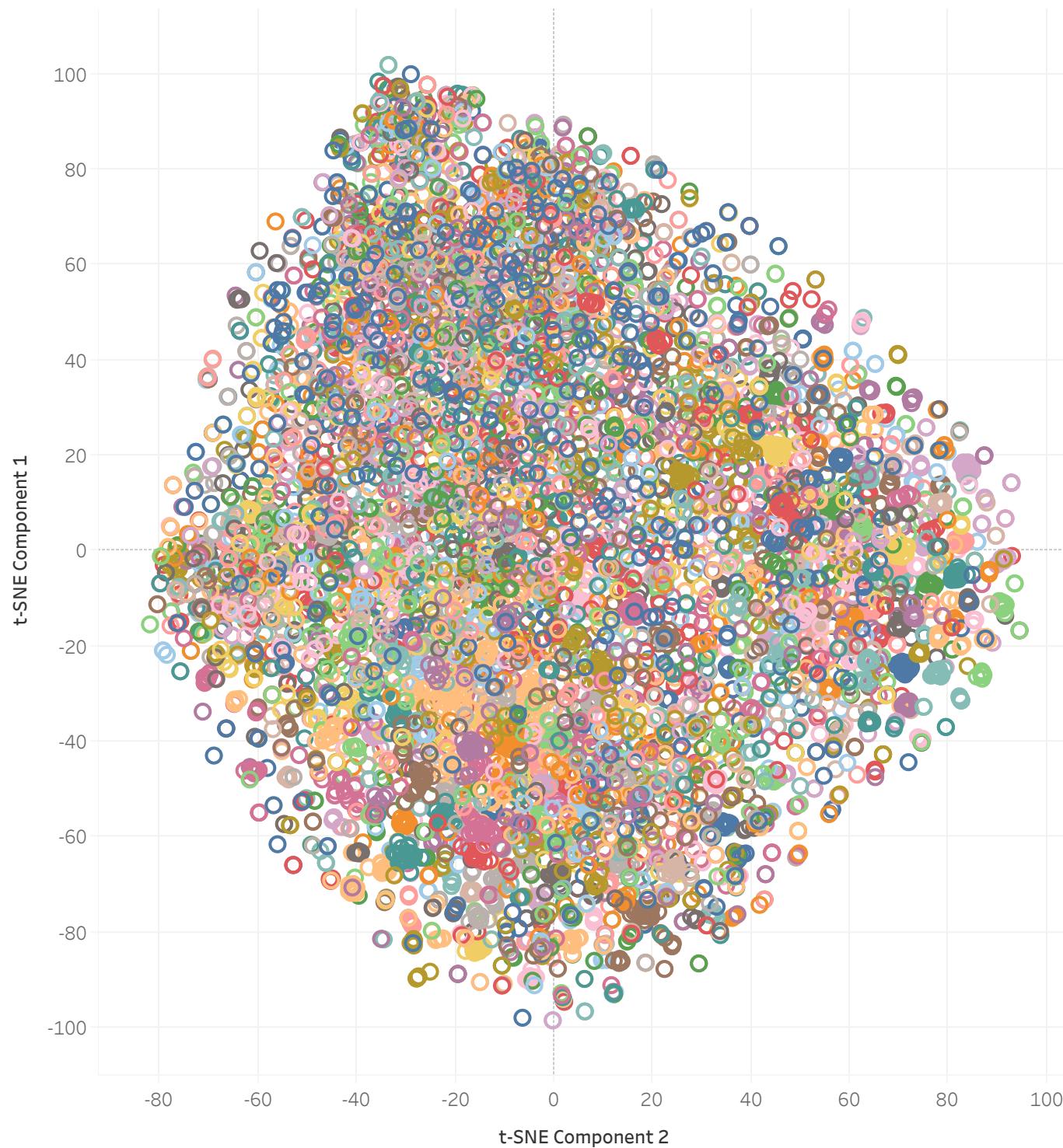
primaryName	prenom	nom	primaryProfession
318		318	actor
'Dissay' Mohamed Rifau	'Dissay'	Mohamed Rifau	location_management
'Pep' Jose del Rio	'Pep'	Jose del Rio	Null
A. Brock	A.	Brock	camera_department
A. Markin	A.	Markin	art_department,miscella..
A.C. Boarug	A.C.	Boarug	Null
A.J. Grimsley	A.J.	Grimsley	actress
A.K. Shah	A.K.	Shah	Null
A.V. Loeben	A.V.	Loeben	actor
Aalika Sheikh	Aalika	Sheikh	actress
Aarika Cornelius	Aarika	Cornelius	actress
Aarne Lindén	Aarne	Lindén	actor
Aarón Emeterio Fernandez	Aarón	Emeterio Fernandez	Null
Aaron Frankie Lewis	Aaron	Frankie Lewis	director
Aaron Heike	Aaron	Heike	music_department
Aaron James	Aaron	James	director,camera_departm..
Aaron Kelly	Aaron	Kelly	composer
Aaron Nocedal	Aaron	Nocedal	camera_department
Aaron Parks	Aaron	Parks	actor,music_department
Aaron Schuett	Aaron	Schuett	Null
Aaron Thomas	Aaron	Thomas	Null
Aaron Tudisco	Aaron	Tudisco	producer,writer
Aaron Walker	Aaron	Walker	actor
Aate Kähkönen	Aate	Kähkönen	actor
Abantica Das	Abantica	Das	art_department
Abbe Opher	Abbe	Opher	Null
Abbie Walsh	Abbie	Walsh	costume_department
Abby Baird	Abby	Baird	make_up_department
Abdellah Mountassir	Abdellah	Mountassir	actor
Abdelmounaime Ennokra	Abdelmounaime	Ennokra	sound_department,editor
Abderrahim Goumri	Abderrahim	Goumri	Null
Abdesalam Belkaziz	Abdesalam	Belkaziz	costume_department

UMAP / TSNE v1 - sentence-transformers/all-MiniLM-L6-v2

UMAP scatterplot

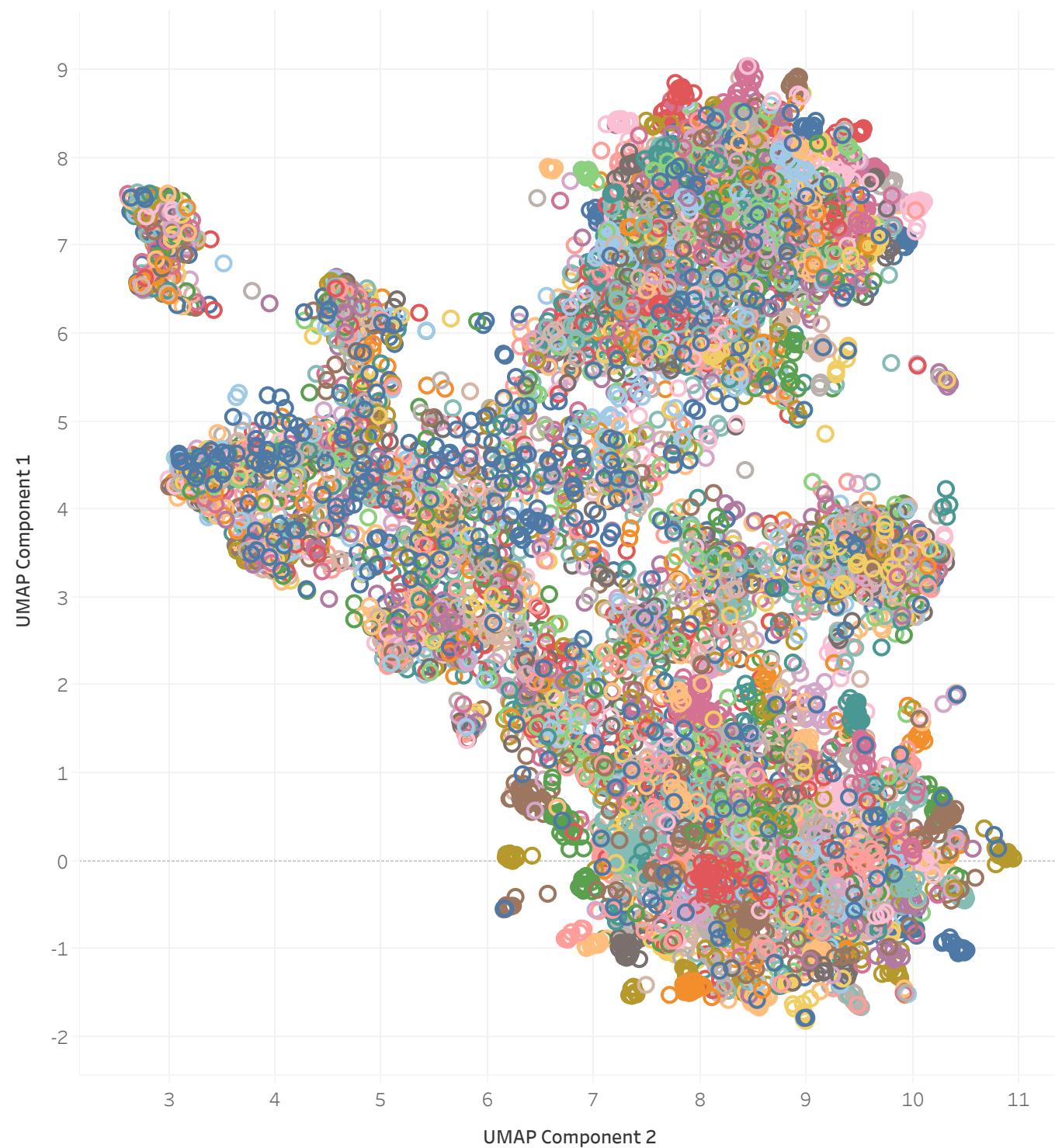


TSNE scatterplot

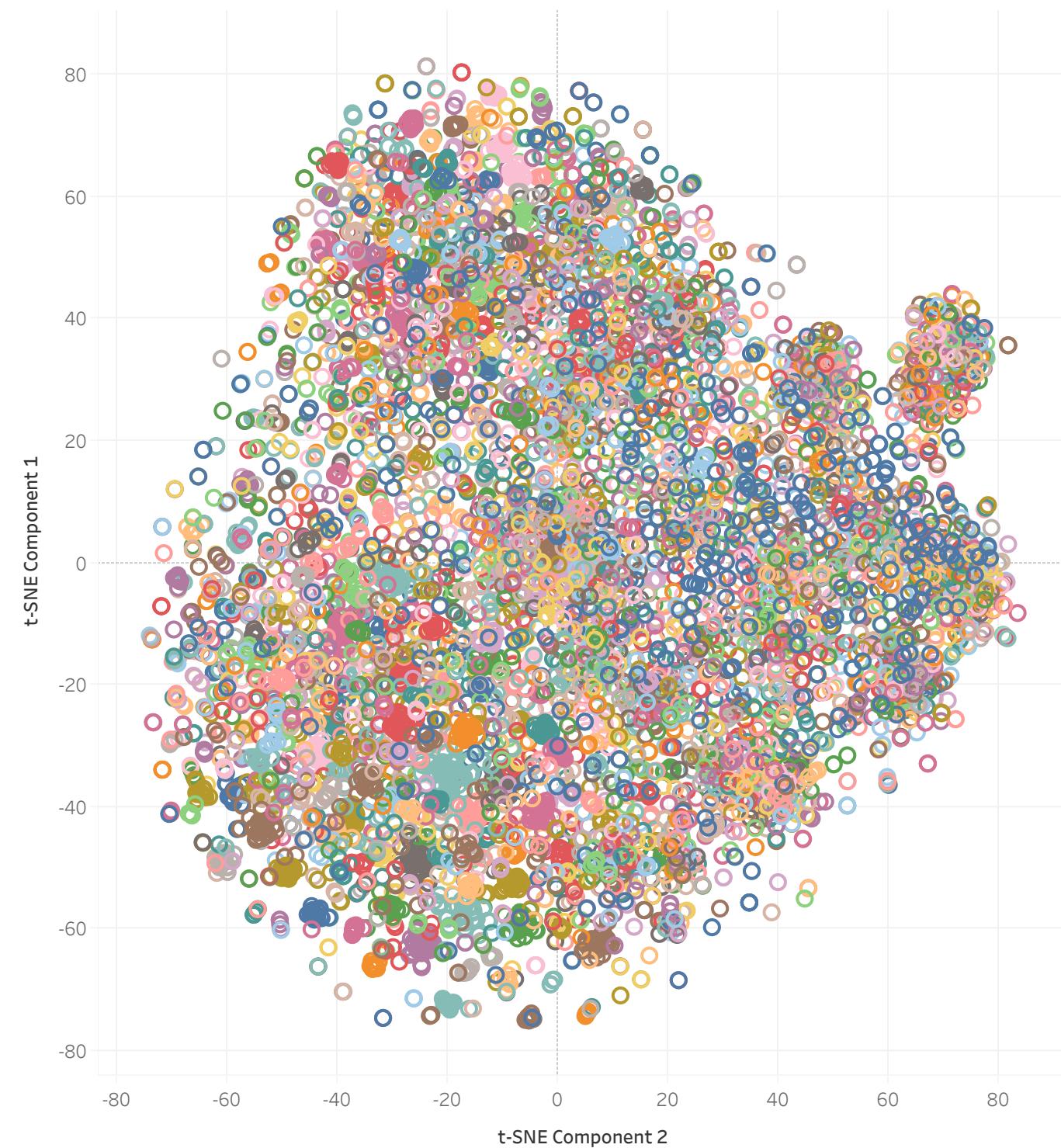


UMAP/TSNE v2 - sentence-transformers/all-mppnet-base-v2

UMAP v2

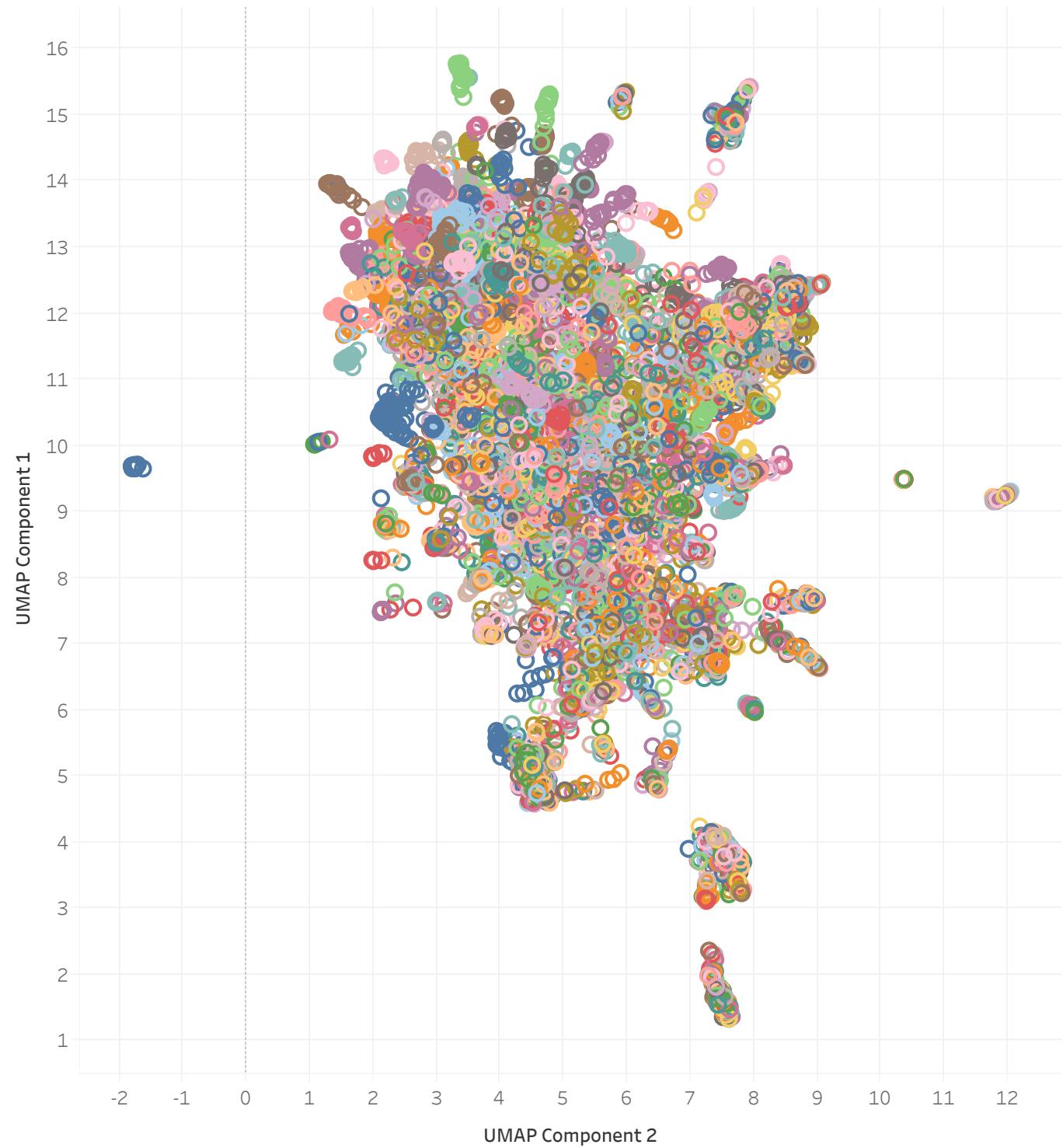


TSNE v2

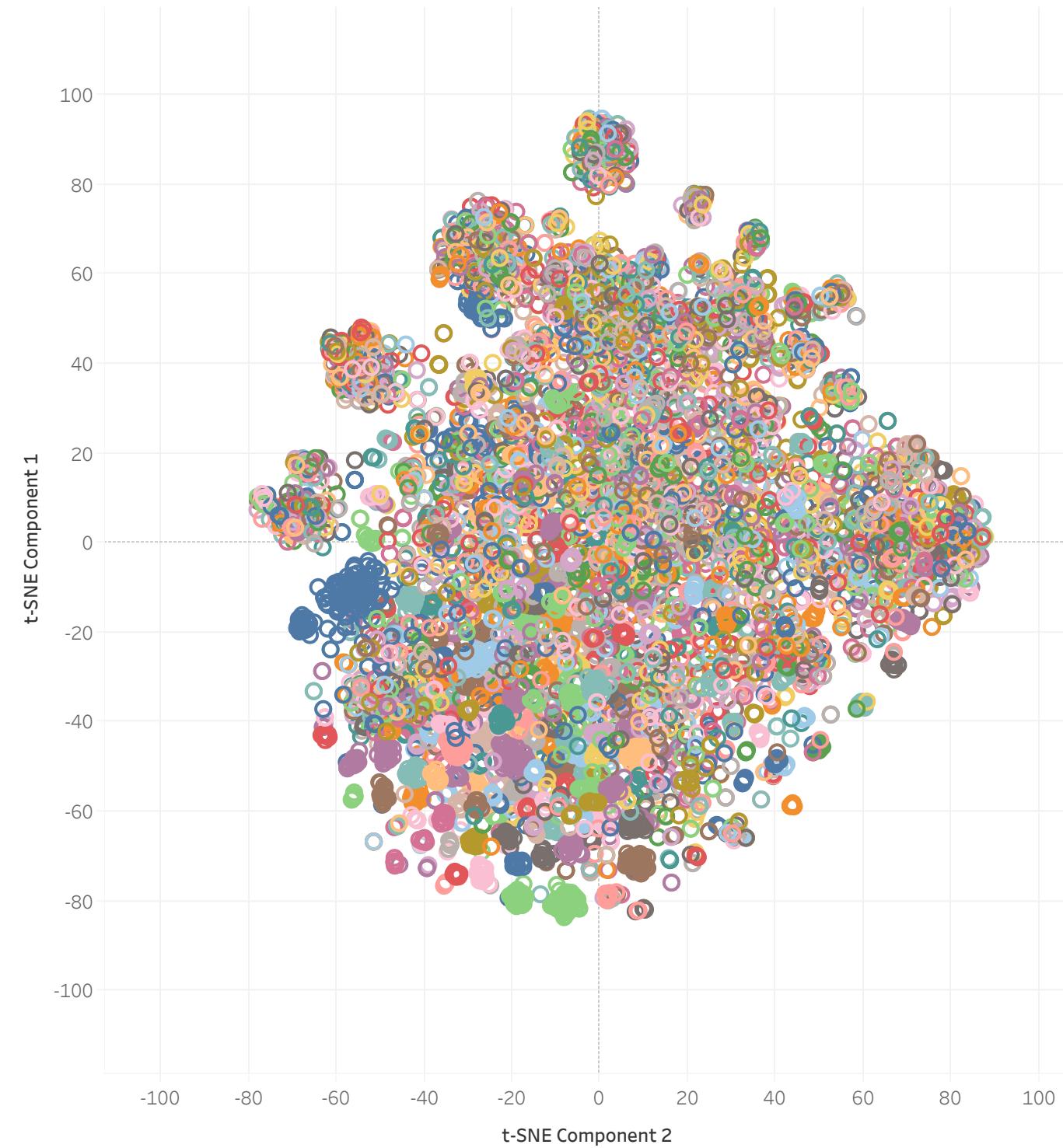


UMAP/TSNE v3 - facebook/bart-large

UMAP v3

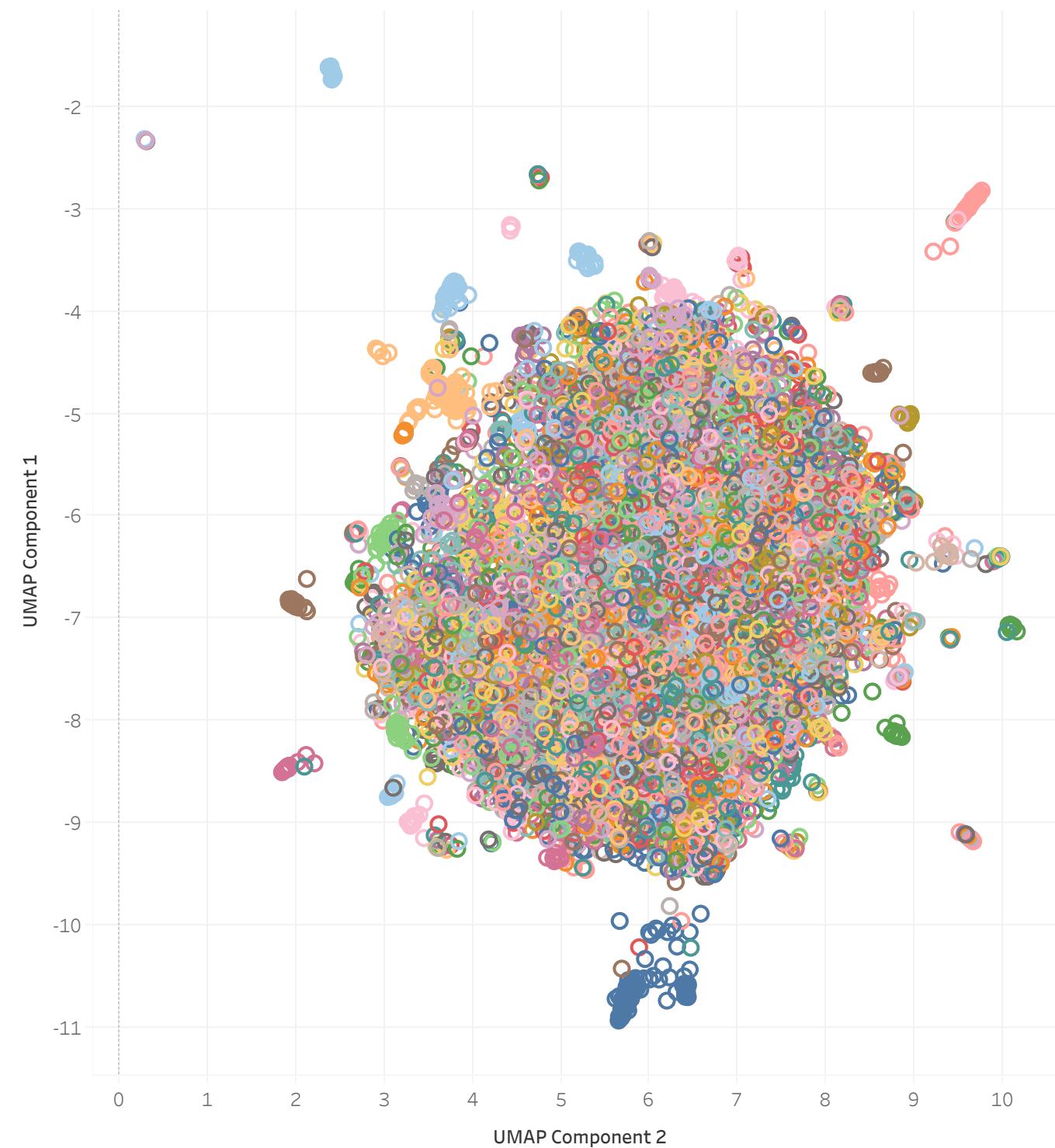


TSNE v3



UMAP/TSNE v4 - facebook/fasttext-en-vectors

UMAP v4



TSNE v4

