

Word Embeddings

Brian Ferrell
SCMA 669

What are they?

Text Data



Text Data



Text Data

Text Data

```
train_df['Text']
```

This study has two primary research questions: ...
There is only one site (VCU). We will communicate ...
A large number of uninsured patients utilize t...
The PI will directly supervise all activities ...
Career Decision-Making Self Efficacy of high s...

This is a retrospective cohort study using sec...
Using email correspondence through secure work...
Older foster youth (18 and up) will be asked t...
More than 70% of adults living in the U.S. are...
I will solely be responsible for supervising t...

```
ext, Length: 2127, dtype: object
```

WEB

Why?

Because...

- Computers cannot understand words



Because...

- Computers cannot understand words
- Brings meaning to how words are being used



Because...

- Computers cannot understand words
- Brings meaning to how words are being used
- It is super fun



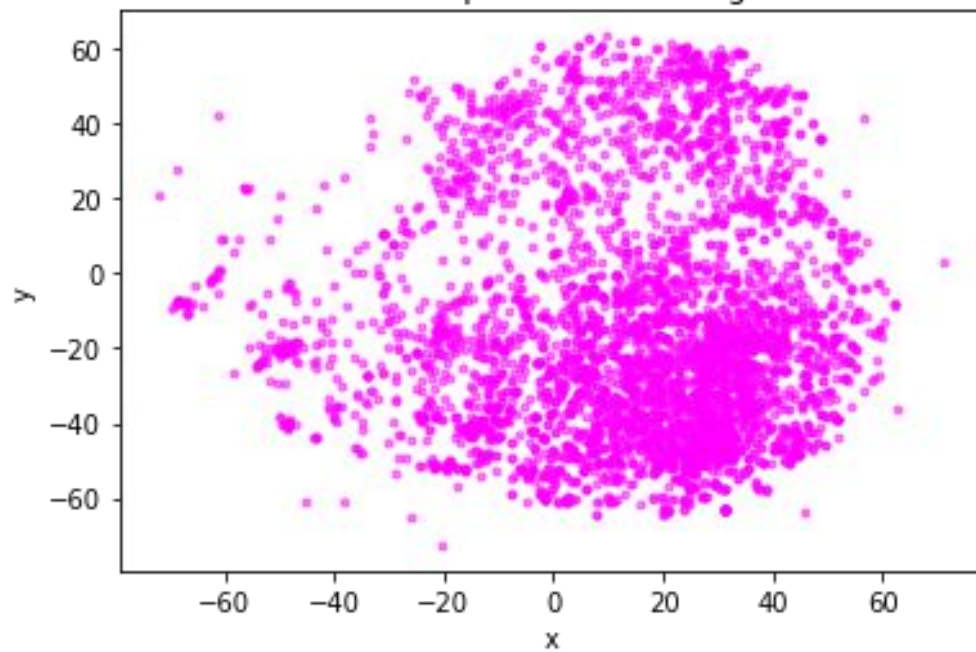
How?

Glove Embeddings

- Download
Pre-trained model
- Train Algorithm
- Output weights

```
array([[ -7.4810913e-04,  2.6995107e-04, -5.0375756e-04, ...,  
        -2.2509256e-03, -5.6318432e-04,  2.5097989e-03],  
       [ 9.3710169e-02,  7.6061952e-01, -4.4982275e-01, ...,  
        -9.2357421e-01,  3.1074095e-01, -4.5782194e-02],  
       [-2.2945511e-01,  1.4638431e-01,  9.3409792e-02, ...,  
        -6.1984915e-01,  1.4767970e-01, -6.6302657e-01],  
       ...,  
       [-9.2641616e-01,  2.8046867e-01,  3.9904031e-01, ...,  
        1.0227690e+00, -5.1031756e-01,  8.0794466e-01],  
       [-1.4626870e-03,  2.7000540e-04, -1.1074090e-04, ...,  
        1.7447599e-04, -9.1576058e-04, -9.9028624e-04],  
       [-4.0838900e-03, -6.6572954e-03,  2.2188339e-03, ...,  
        1.6201815e-03, -2.3363857e-03, -2.3705274e-05]], dtype=float32)
```

Scatter plot of Embeddings

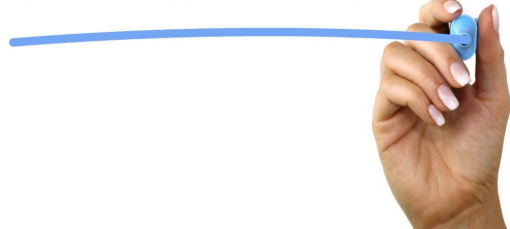




THANK YOU!
Google

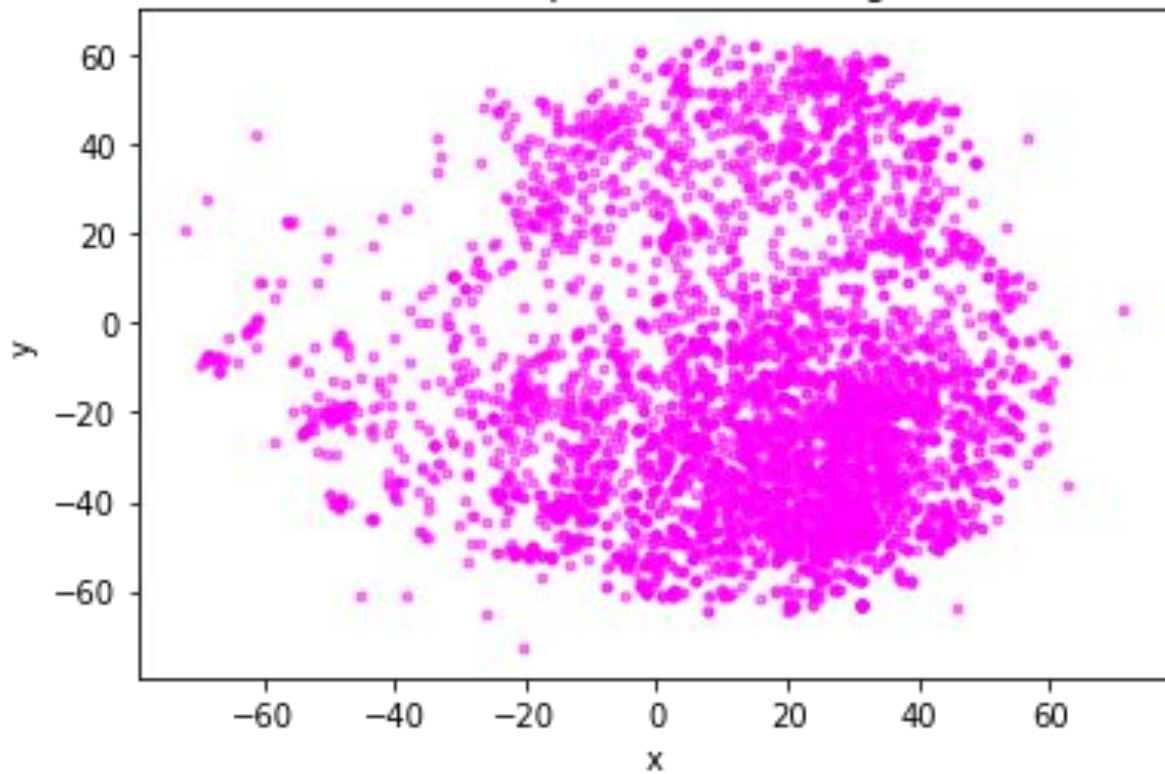
***This is where we thank
Google***

THANK YOU

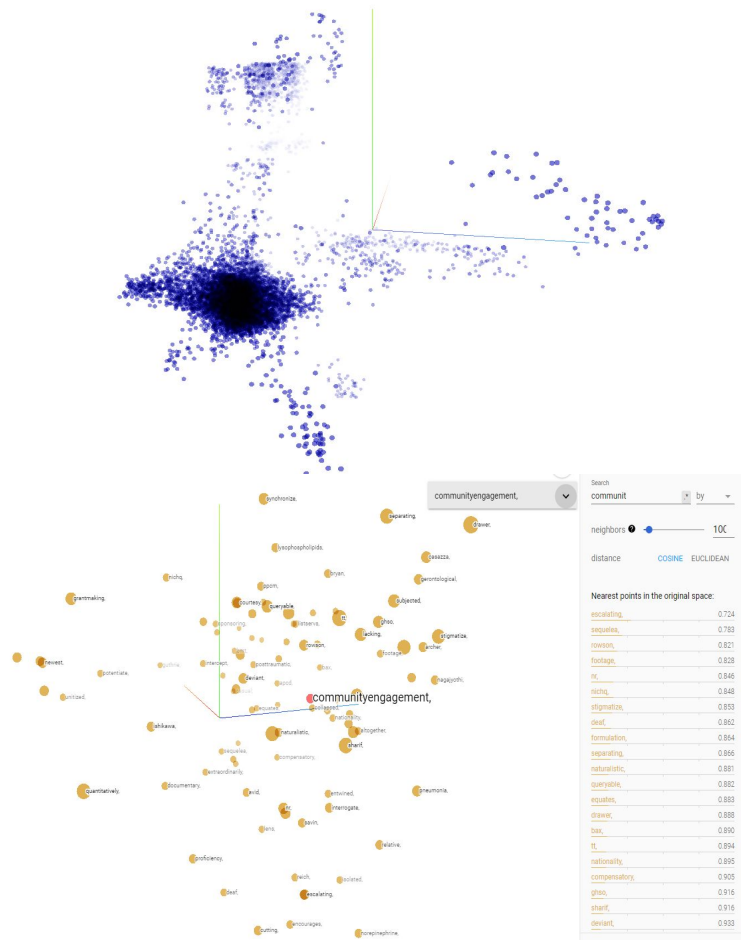


It can turn

Scatter plot of Embeddings

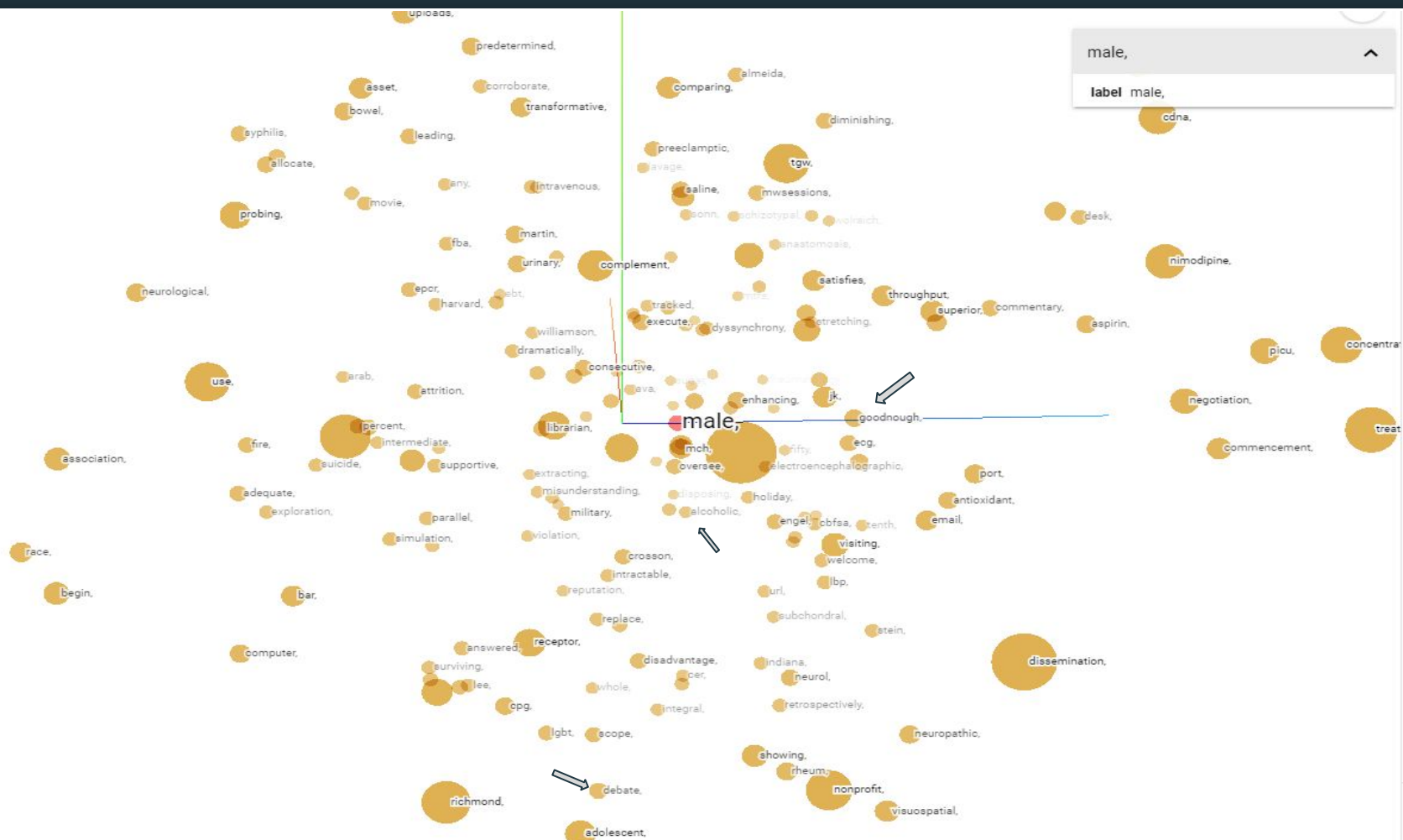


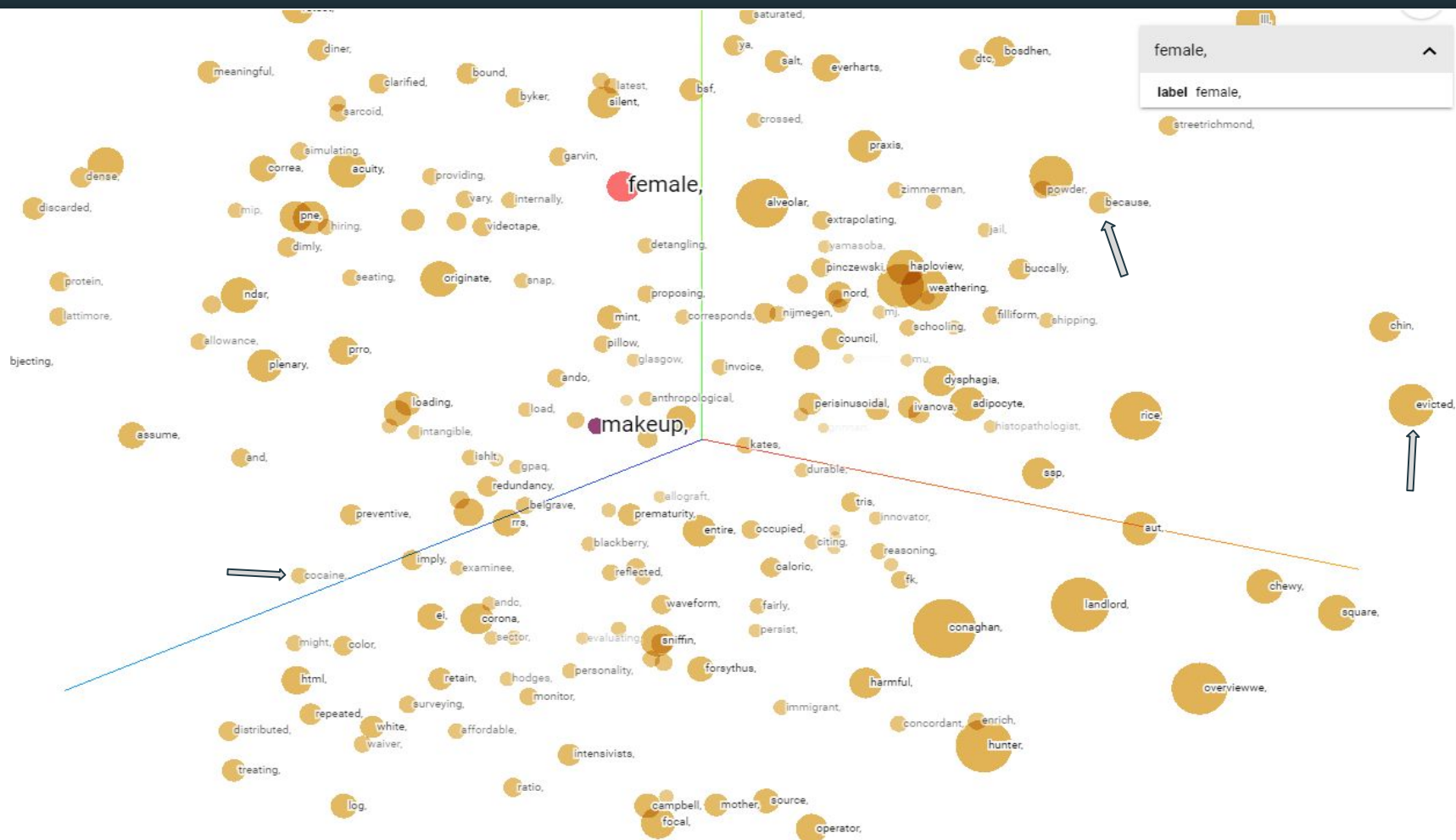
Into...



My Data







?

pottery

eaten

clean

crops

trees

marine

factories

lit

brands

travel

leather

aging

expensive

shops

yeast

soil

production

flavor

fast

soup

cheap

modern

drinks

shipping

beer

grain

information

fresh

farm

chicken

potato

furniture

causing

storage

horse

dog

basic

fruit

fish

additive

tourist

living

shoes

countless

people

but

fried

beef

milk

produce

traditional

dogs

maintenance

put

dried

potatoes

pet

cattle

wheat

animals

wood

pork

chairs

vegetables

waste

rice

salt

toxic

consumed

concern

money

honey

sheep

cheese

beans

liver

bread

variety

buy

survival

local

ingredients

eat

peasant

food

small

suitable

meal

foods

chains

sp

protection

bacterial

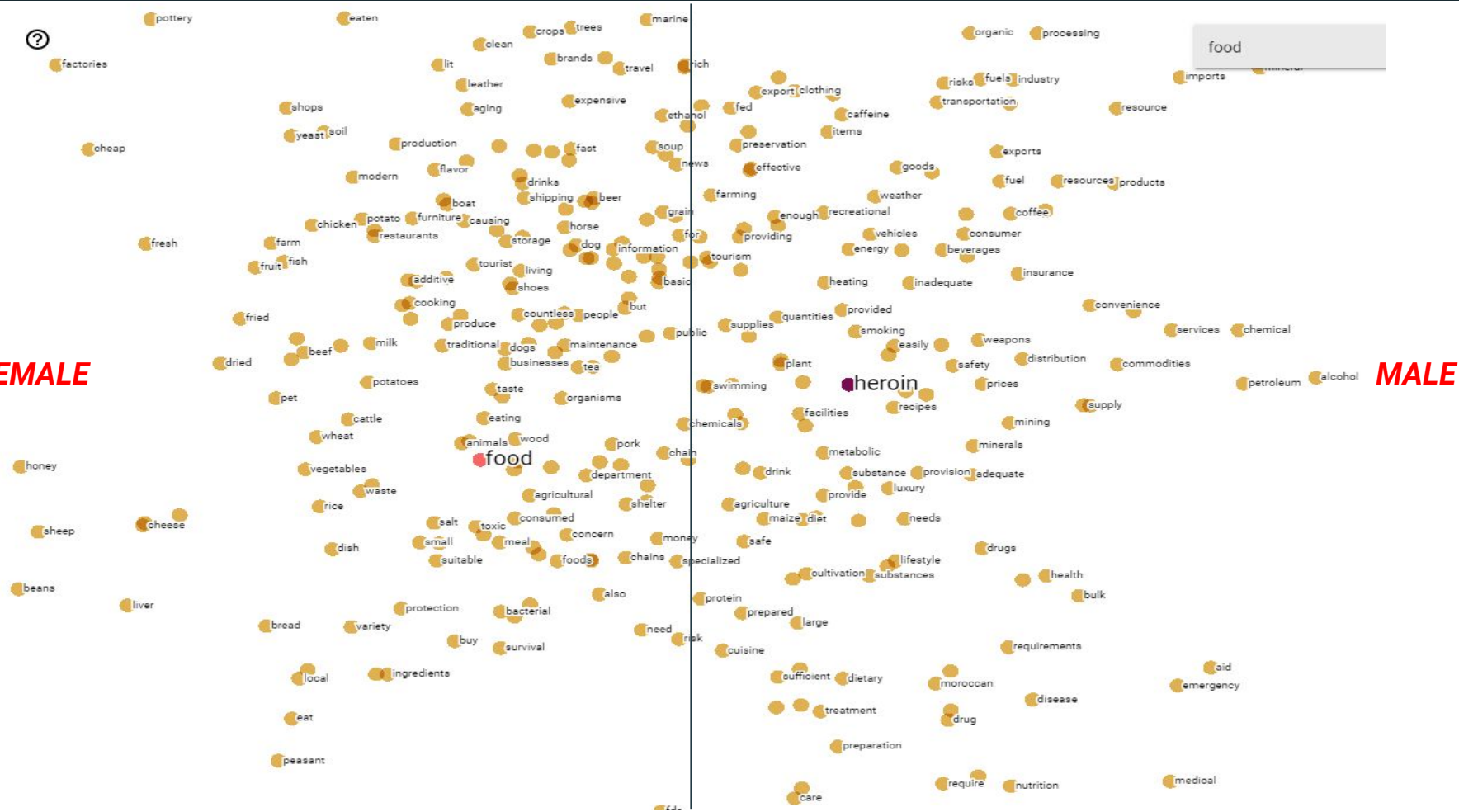
also

need

ri

FEMALE

MALE

**MALE**

Conclusion

- Just shows your data
- Biased
- Doesn't actually mean men eat petroleum

Thank You!