Learning Objectives
By the end of this lab, students should be able to:
• Understand the need for visualizing high-dimensional embeddings.
• Load pre-trained word embeddings.
• Apply t-SNE for dimensionality reduction.
• Plot embeddings in 2-D space.
• Interpret observed clusters and relationships.
Lab Outcomes (LOs)
After completing this lab, students will be able to:
• Explain what t-SNE does and when to use it.
• Select meaningful words for visualization.
• Convert vectors to 2-D representations.
• Generate scatter plots of word embeddings.
• Identify clusters formed by semantic similarity.
• Prepare a concise analysis report.

```
!pip install gensim
import gensim.downloader as api
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.manifold import TSNE
```

```
Collecting gensim
  Downloading gensim-4.4.0-cp312-cp312-manylinux_2_24_x86_64.manylinux_2_28_x86_64.whl.metadata (8.4 kB)
Requirement already satisfied: numpy>=1.18.5 in /usr/local/lib/python3.12/dist-packages (from gensim) (2.0.2)
Requirement already satisfied: scipy>=1.7.0 in /usr/local/lib/python3.12/dist-packages (from gensim) (1.16.3)
Requirement already satisfied: smart_open>=1.8.1 in /usr/local/lib/python3.12/dist-packages (from gensim) (7.5.0)
Requirement already satisfied: wrapt in /usr/local/lib/python3.12/dist-packages (from smart_open>=1.8.1->gensim) (2.1.1)
Downloading gensim-4.4.0-cp312-cp312-manylinux_2_24_x86_64.manylinux_2_28_x86_64.whl (27.9 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 27.9/27.9 MB 55.3 MB/s eta 0:00:00
Installing collected packages: gensim
Successfully installed gensim-4.4.0
```

```
print("Loading pre-trained word embeddings...")
model = api.load("glove-wiki-gigaword-100")  # 100-dimensional GloVe
print("Vocabulary size:", len(model.key_to_index))
```

```
Loading pre-trained word embeddings...
[==================================================] 100.0% 128.1/128.1MB downloaded
Vocabulary size: 400000
```

```
word = "computer"
print(f"Vector for '{word}':\n", model[word])
```

```
Vector for 'computer':
 [-1.6298e-01  3.0141e-01  5.7978e-01  6.6548e-02  4.5835e-01 -1.5329e-01
  4.3258e-01 -8.9215e-01  5.7747e-01  3.6375e-01  5.6524e-01 -5.6281e-01
  3.5659e-01 -3.6096e-01 -9.9662e-02  5.2753e-01  3.8839e-01  9.6185e-01
  1.8841e-01  3.0741e-01 -8.7842e-01 -3.2442e-01  1.1202e+00  7.5126e-02
  4.2661e-01 -6.0651e-01 -1.3893e-01  4.7862e-02 -4.5158e-01  9.3723e-02
  1.7463e-01  1.0962e+00 -1.0044e+00  6.3889e-02  3.8002e-01  2.1109e-01
 -6.6247e-01 -4.0736e-01  8.9442e-01 -6.0974e-01 -1.8577e-01 -1.9913e-01
 -6.9226e-01 -3.1806e-01 -7.8565e-01  2.3831e-01  1.2992e-01  8.7721e-02
  4.3205e-01 -2.2662e-01  3.1549e-01 -3.1748e-01 -2.4632e-03  1.6615e-01
  4.2358e-01 -1.8087e+00 -3.6699e-01  2.3949e-01  2.5458e+00  3.6111e-01
  3.9486e-02  4.8607e-01 -3.6974e-01  5.7282e-02 -4.9317e-01  2.2765e-01
  7.9966e-01  2.1428e-01  6.9811e-01  1.1262e+00 -1.3526e-01  7.1972e-01
 -9.9605e-04 -2.6842e-01 -8.3038e-01  2.1780e-01  3.4355e-01  3.7731e-01
 -4.0251e-01  3.3124e-01  1.2576e+00 -2.7196e-01 -8.6093e-01  9.0053e-02
 -2.4876e+00  4.5200e-01  6.6945e-01 -5.4648e-01 -1.0324e-01 -1.6979e-01
  5.9437e-01  1.1280e+00  7.5755e-01 -5.9160e-02  1.5152e-01 -2.8388e-01
  4.9452e-01 -9.1703e-01  9.1289e-01 -3.0927e-01]
```

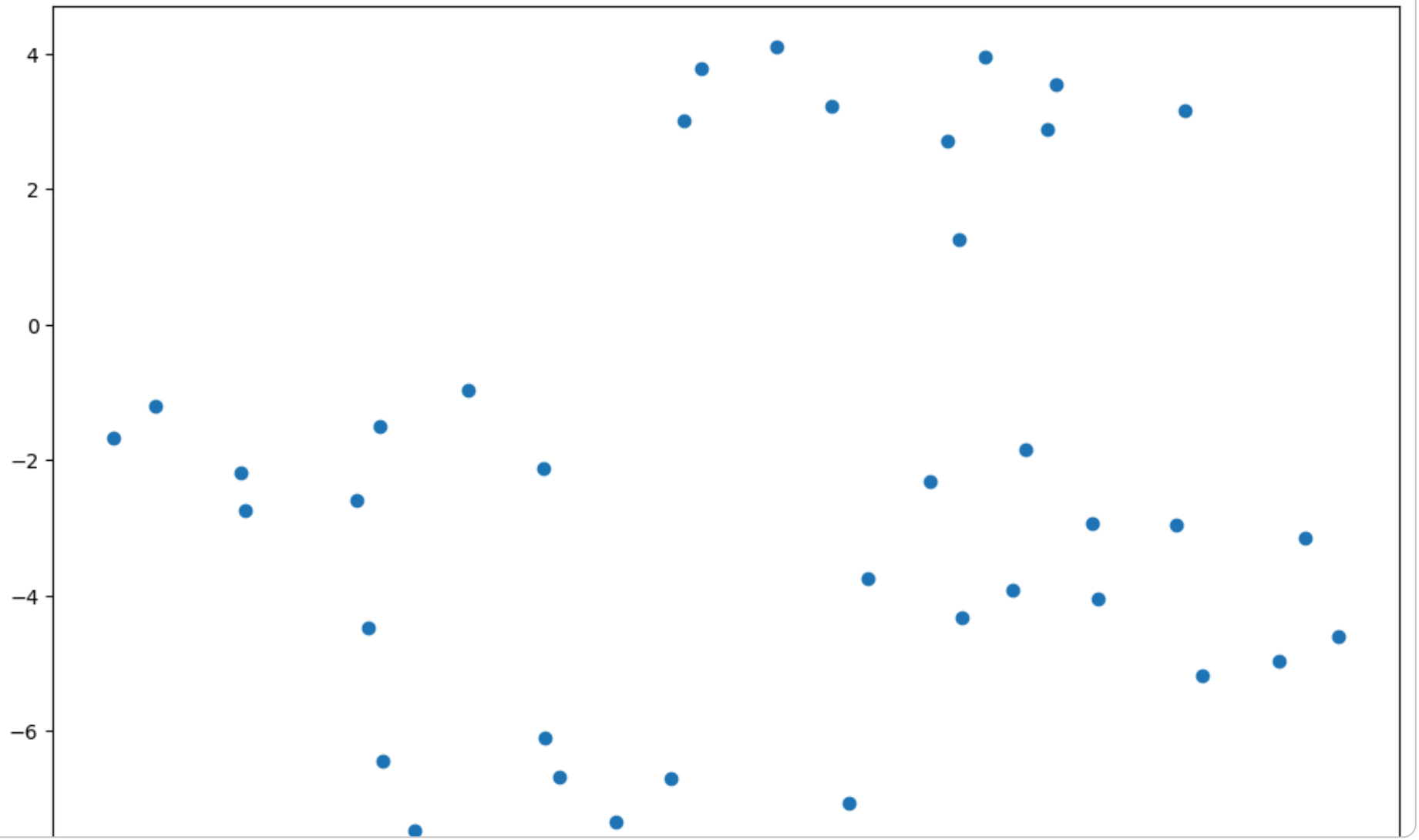
```
vectors = np.array([model[w] for w in words])
```

```
print("Running t-SNE...")
tsne = TSNE(n_components=2, random_state=42, perplexity=15, n_iter=1000)
word_embeddings_2d = tsne.fit_transform(vectors)
```

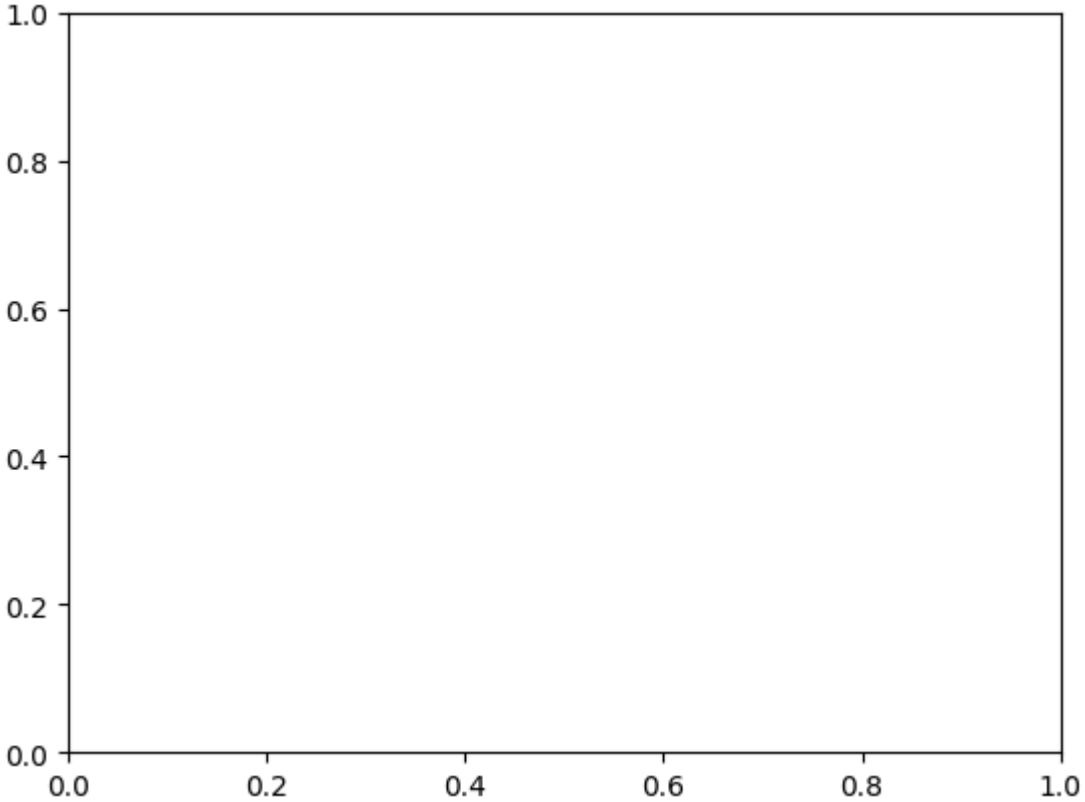
```
Running t-SNE...
/usr/local/lib/python3.12/dist-packages/sklearn/manifold/_t_sne.py:1164: FutureWarning: 'n_iter' was renamed to 'max_ite
  warnings.warn(
```

```
plt.figure(figsize=(12,8))
plt.scatter(word_embeddings_2d[:,0], word_embeddings_2d[:,1])
```

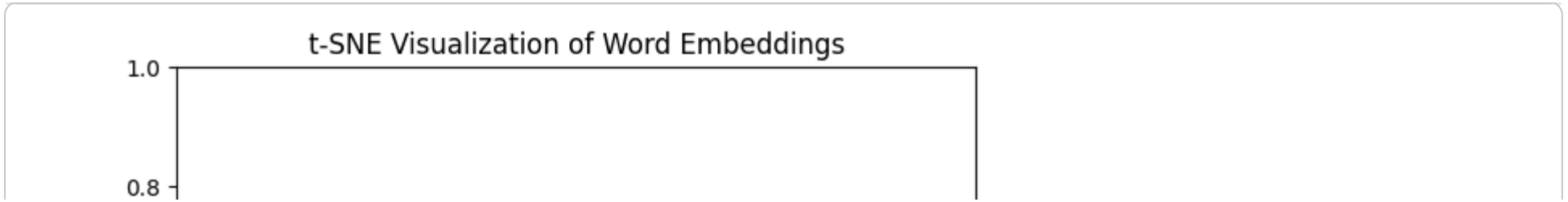
```
<matplotlib.collections.PathCollection at 0x7e44bbca8920>
```




```
for i, word in enumerate(words):
    plt.annotate(word, (word_embeddings_2d[i,0], word_embeddings_2d[i,1]))
```




```
plt.title("t-SNE Visualization of Word Embeddings")
plt.xlabel("Dimension 1")
plt.ylabel("Dimension 2")
plt.show()
```


t-SNE Visualization of Word Embeddings
1.0
0.8