Homework Assignment 6 [30 pts]

STAT437 Unsupervised Learning – Spring 2025

Due: Friday, March 7 on Canvas at 11:59pm CST.

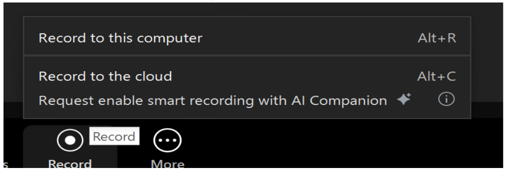
Video Question [1 point]: Make a 3-4 minute video explaining what you did in **questions 1-4.**

IMPORTANT Video Element of ALL Homework Assignments:

- In order to receive points for each video submission, you need to do **ALL** of the following.
  - Have your camera on.
  - Show your FULL screen in Zoom (not just a particular application).
  - We should be able to hear the audio. Make sure to turn your mic on.
  - You should give a good faith attempt to answer the prompt.
  - Your video meet the minimum time requirement.
  - It should not sound like you are just reading off a script.
  - It's ok if your video recording is not the most eloquent. What's important is that you are putting together YOUR authentic thoughts on your particular understanding of the assignment and the lecture content.

How to Submit Videos:

- You should record your videos in your UIUC Zoom client.
- You should record your videos To the Cloud.


Record to this computer Alt+R
Record to the cloud Alt+C
Request enable smart recording with AI Companion
Record
Record
More

- You can find your recording link at https://illinois.zoom.us/recording/.
- Click on the corresponding video and Copy shareable link to paste the link in Canvas.

Problems	Points
1.1	0.25
1.2	0.25
1.3	1
1.4	1
2.1	1.5
2.2	2
2.3	2
2.4	1.5
3.1	1.5
3.2	1.5
3.3	1.5
3.4	0.5
3.5	0.75
3.6.1.	1
3.6.2	0.5
3.6.3	1
3.6.4	0.5
4.1.1	0.75
4.1.2	0.75
4.2.1	0.75
4.2.2	0.75
4.3.1	0.75
4.3.2	1
4.4	1
5	2.5
6	2.5

Questions 1-4: See Jupyter notebook

Question 5: The dataset below is comprised of 10 professors in the UIUC Statistics department. Suppose we'd like to cluster this dataset using k-modes with k=2 clusters. The *current* cluster modes are shown below. What would be the *new cluster modes* found in the next iteration of the k-modes clustering algorithm. Show your work.

- *Hint: If there's a tie in the cluster assignment step, assign the person to the mode with the highest index.*
- *Hint: If there's a tie in the centroid/mode update step, select the attribute value with the highest alphabetical order (ie. A>B).*

	PhD	**Sex**	**Generation**
Tori Ellison	Operations Research	Female	millennial
Karle Flanagan	Statistics Education	Female	millennial
Kelly Findley	Statistics Education	Male	millennial
Julie Deeke	Statistics	Female	millennial
Chris Kinson	Statistics	Male	millennial
Jeff Douglas	Statistics	Male	boomer
Bo Li	Statistics	Female	Gen X
Steve Culpepper	Educational Psychology	Male	Gen X
Dave Zhao	Statistics	Male	millennial
Vimal Rao	Educational Psychology	Male	millennial

	PhD	**Sex**	**Generation**
Mode 1	Statistics	Male	millennial
Mode 2	Operations Research	Female	millennial

Question 6: In the page below 4 categorical datasets are listed. Each of these datasets was put into a Hamming distance matrix. Each hamming distance matrix was then used as input into the t-SNE algorithm producing the following 4 sets of t-SNE plots shown in the pages below.

Match each of the 4 datasets 1-4 to the corresponding t-SNE plot sets A-D.

Hint: Some points may be completely overlapping in some of the t-SNE plots below.

Explanations not required, but may help for partial credit if you get it wrong.

Dataset 1

pet	fav jonas bro	is a hotdog a sandwich ?
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	joe	yes

Dataset 2

pet	fav jonas bro	is a hotdog a sandwich ?
cat	joe	no
cat	joe	no
cat	joe	no
cat	joe	no
cat	joe	no
cat	joe	not_sure
cat	joe	not_sure
cat	joe	not_sure
cat	joe	yes
cat	joe	yes
cat	joe	yes
cat	kevin	no
cat	kevin	no
cat	kevin	not_sure
cat	kevin	not_sure
cat	kevin	not_sure
cat	kevin	yes
cat	kevin	yes
cat	kevin	yes
cat	kevin	yes
cat	kevin	yes
cat	kevin	yes
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	yes
cat	nick	yes
dog	joe	no
dog	joe	no
dog	joe	no
dog	joe	no
dog	joe	not_sure
dog	joe	not_sure
dog	joe	yes
dog	joe	yes
dog	joe	yes
dog	joe	yes
dog	joe	yes
dog	joe	yes
dog	kevin	no
dog	kevin	no
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	not_sure
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	kevin	yes
dog	nick	no
dog	nick	no
dog	nick	no
dog	nick	no
dog	nick	no
dog	nick	not_sure
dog	nick	not_sure
dog	nick	not_sure
dog	nick	yes
dog	nick	yes
dog	nick	yes
dog	nick	yes
fish	joe	no
fish	joe	no
fish	joe	no
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	yes
fish	joe	yes
fish	joe	yes
fish	kevin	no
fish	kevin	no
fish	kevin	no
fish	kevin	not_sure
fish	kevin	not_sure
fish	kevin	not_sure
fish	kevin	not_sure
fish	kevin	not_sure
fish	kevin	not_sure
fish	kevin	not_sure
fish	kevin	yes
fish	kevin	yes
fish	kevin	yes
fish	nick	no
fish	nick	no
fish	nick	no
fish	nick	no
fish	nick	no
fish	nick	not_sure
fish	nick	not_sure
fish	nick	not_sure
fish	nick	yes
fish	nick	yes

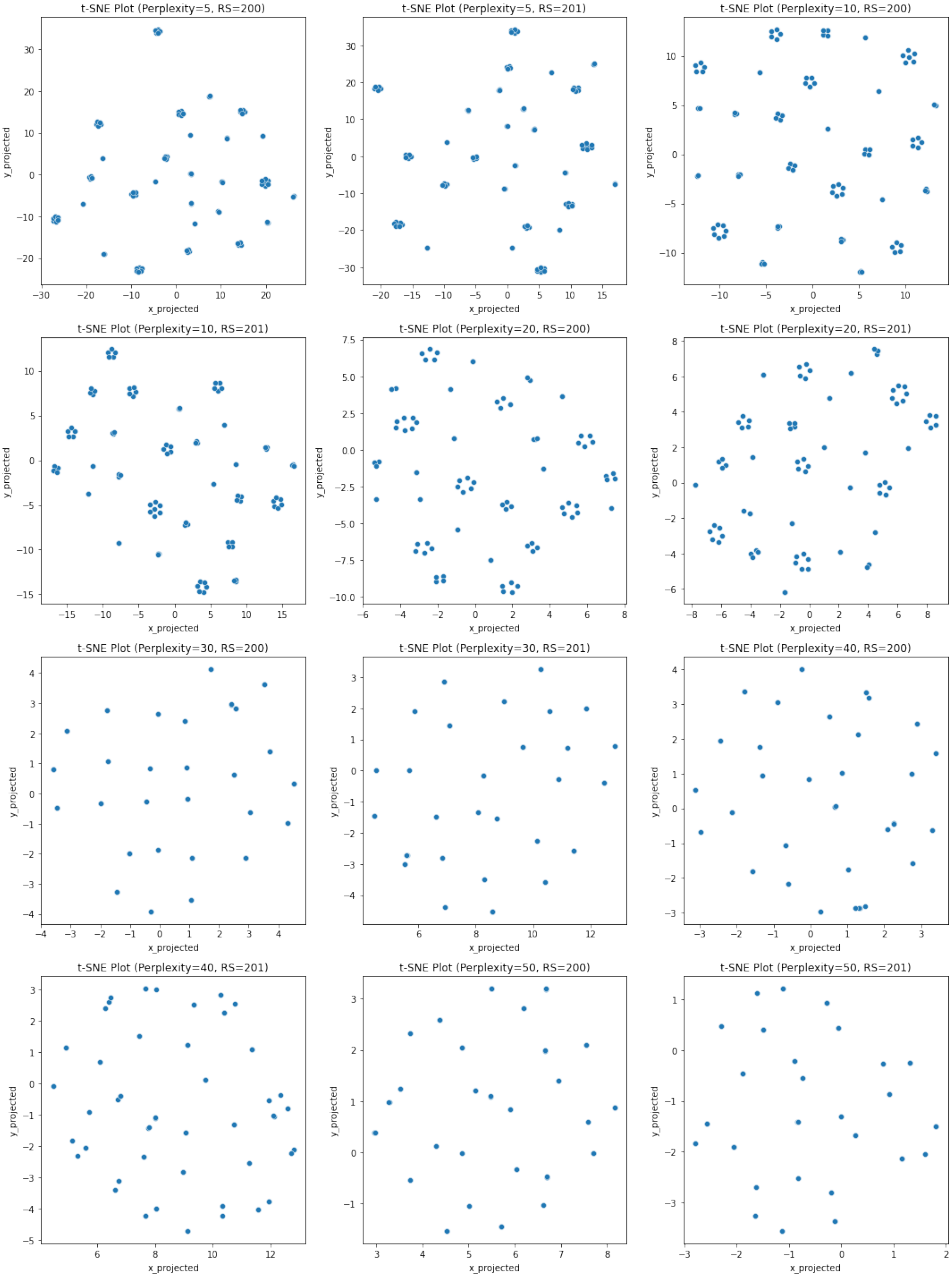
Dataset 3

pet	fav jonas bro	is a hotdog a sandwich ?
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	no
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
cat	nick	not_sure
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure

Dataset 4

pet	fav jonas bro	is a hotdog a sandwich?
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	kevin	no
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
cat	nick	yes
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
dog	kevin	no
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure
fish	joe	not_sure

Plot Set A



t-SNE Plot (Perplexity=5, RS=200)

t-SNE Plot (Perplexity=5, RS=201)

t-SNE Plot (Perplexity=10, RS=200)

t-SNE Plot (Perplexity=10, RS=201)

t-SNE Plot (Perplexity=20, RS=200)

t-SNE Plot (Perplexity=20, RS=201)

t-SNE Plot (Perplexity=30, RS=200)

t-SNE Plot (Perplexity=30, RS=201)

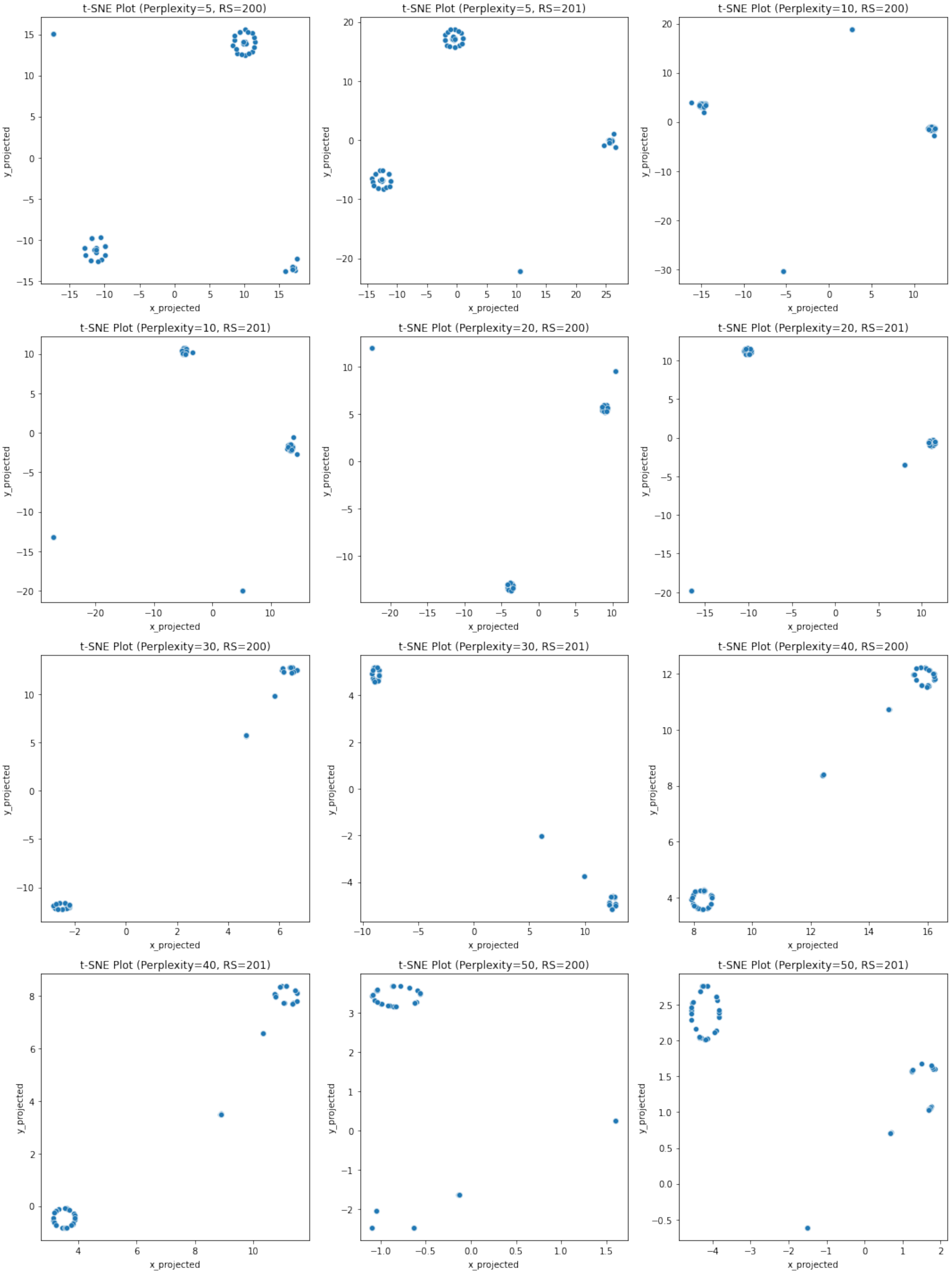
t-SNE Plot (Perplexity=40, RS=200)

t-SNE Plot (Perplexity=40, RS=201)

t-SNE Plot (Perplexity=50, RS=200)

t-SNE Plot (Perplexity=50, RS=201)

Plot Set B



t-SNE Plot (Perplexity=5, RS=200)

t-SNE Plot (Perplexity=5, RS=201)

t-SNE Plot (Perplexity=10, RS=200)

t-SNE Plot (Perplexity=10, RS=201)

t-SNE Plot (Perplexity=20, RS=200)

t-SNE Plot (Perplexity=20, RS=201)

t-SNE Plot (Perplexity=30, RS=200)

t-SNE Plot (Perplexity=30, RS=201)

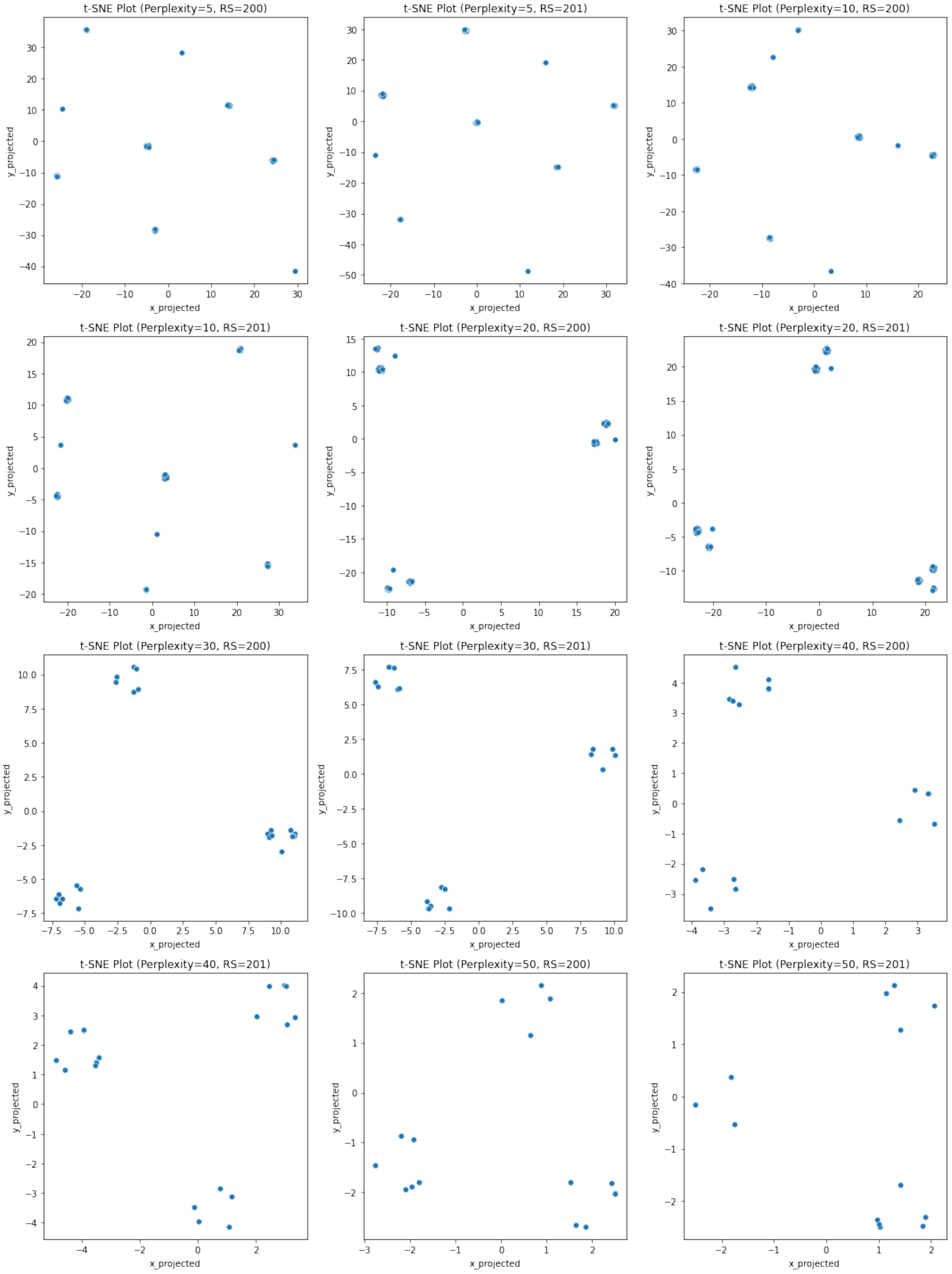
t-SNE Plot (Perplexity=40, RS=200)

t-SNE Plot (Perplexity=40, RS=201)

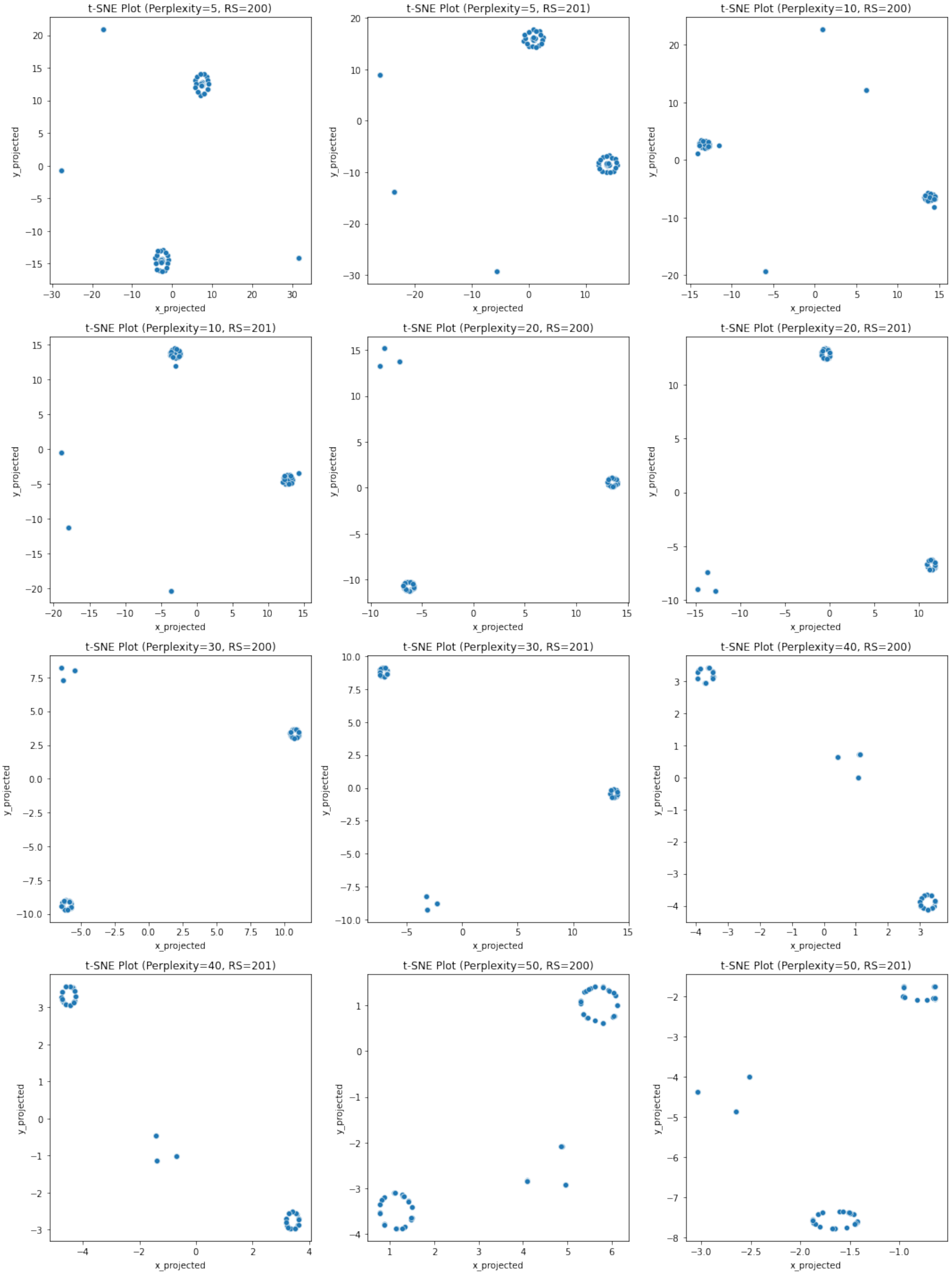
t-SNE Plot (Perplexity=50, RS=200)

t-SNE Plot (Perplexity=50, RS=201)

Plot Set C


t-SNE Plot (Perplexity=5, RS=200)
y_projected
x_projected
30
20
10
0
−10
−20
−30
−40
−20
−10
0
10
20
30
t-SNE Plot (Perplexity=5, RS=201)
y_projected
x_projected
30
20
10
0
−10
−20
−30
−40
−50
−20
−10
0
10
20
30
t-SNE Plot (Perplexity=10, RS=200)
y_projected
x_projected
30
20
10
0
−10
−20
−30
−40
−20
−10
0
10
20
t-SNE Plot (Perplexity=10, RS=201)
y_projected
x_projected
20
15
10
5
0
−5
−10
−15
−20
−20
−10
0
10
20
30
t-SNE Plot (Perplexity=20, RS=200)
y_projected
x_projected
15
10
5
0
−5
−10
−15
−20
−10
−5
0
5
10
15
20
t-SNE Plot (Perplexity=20, RS=201)
y_projected
x_projected
20
15
10
5
0
−5
−10
−20
−10
0
10
20
t-SNE Plot (Perplexity=30, RS=200)
y_projected
x_projected
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
t-SNE Plot (Perplexity=30, RS=201)
y_projected
x_projected
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
t-SNE Plot (Perplexity=40, RS=200)
y_projected
x_projected
4
3
2
1
0
−1
−2
−3
−4
−3
−2
−1
0
1
2
3
t-SNE Plot (Perplexity=40, RS=201)
y_projected
x_projected
4
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
t-SNE Plot (Perplexity=50, RS=200)
y_projected
x_projected
2
1
0
−1
−2
−3
−2
−1
0
1
2
t-SNE Plot (Perplexity=50, RS=201)
y_projected
x_projected
2
1
0
−1
−2
−2
−1
0
1
2

Plot Set D



t-SNE Plot (Perplexity=5, RS=200)

t-SNE Plot (Perplexity=5, RS=201)

t-SNE Plot (Perplexity=10, RS=200)

t-SNE Plot (Perplexity=10, RS=201)

t-SNE Plot (Perplexity=20, RS=200)

t-SNE Plot (Perplexity=20, RS=201)

t-SNE Plot (Perplexity=30, RS=200)

t-SNE Plot (Perplexity=30, RS=201)

t-SNE Plot (Perplexity=40, RS=200)

t-SNE Plot (Perplexity=40, RS=201)

t-SNE Plot (Perplexity=50, RS=200)

t-SNE Plot (Perplexity=50, RS=201)