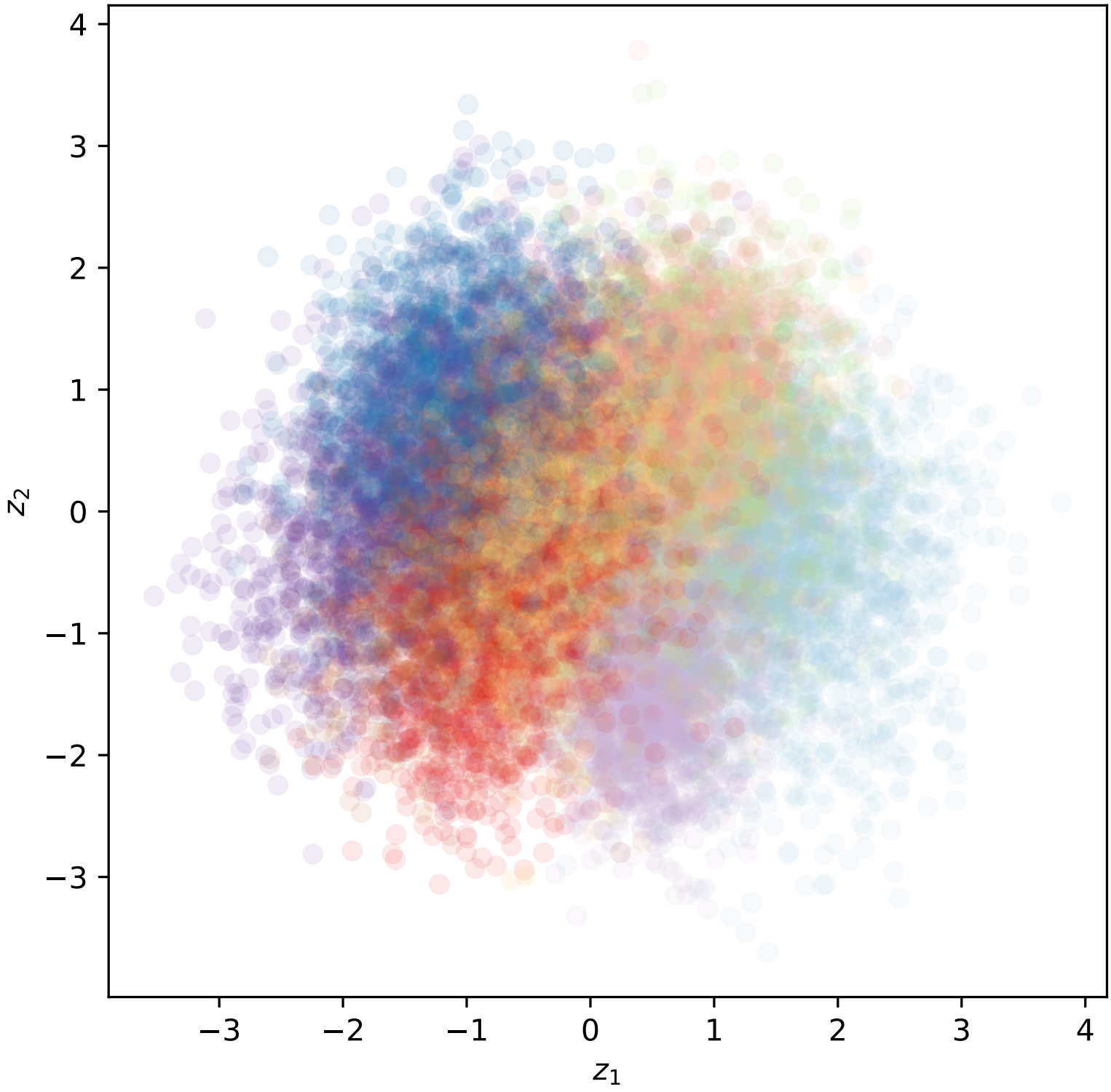
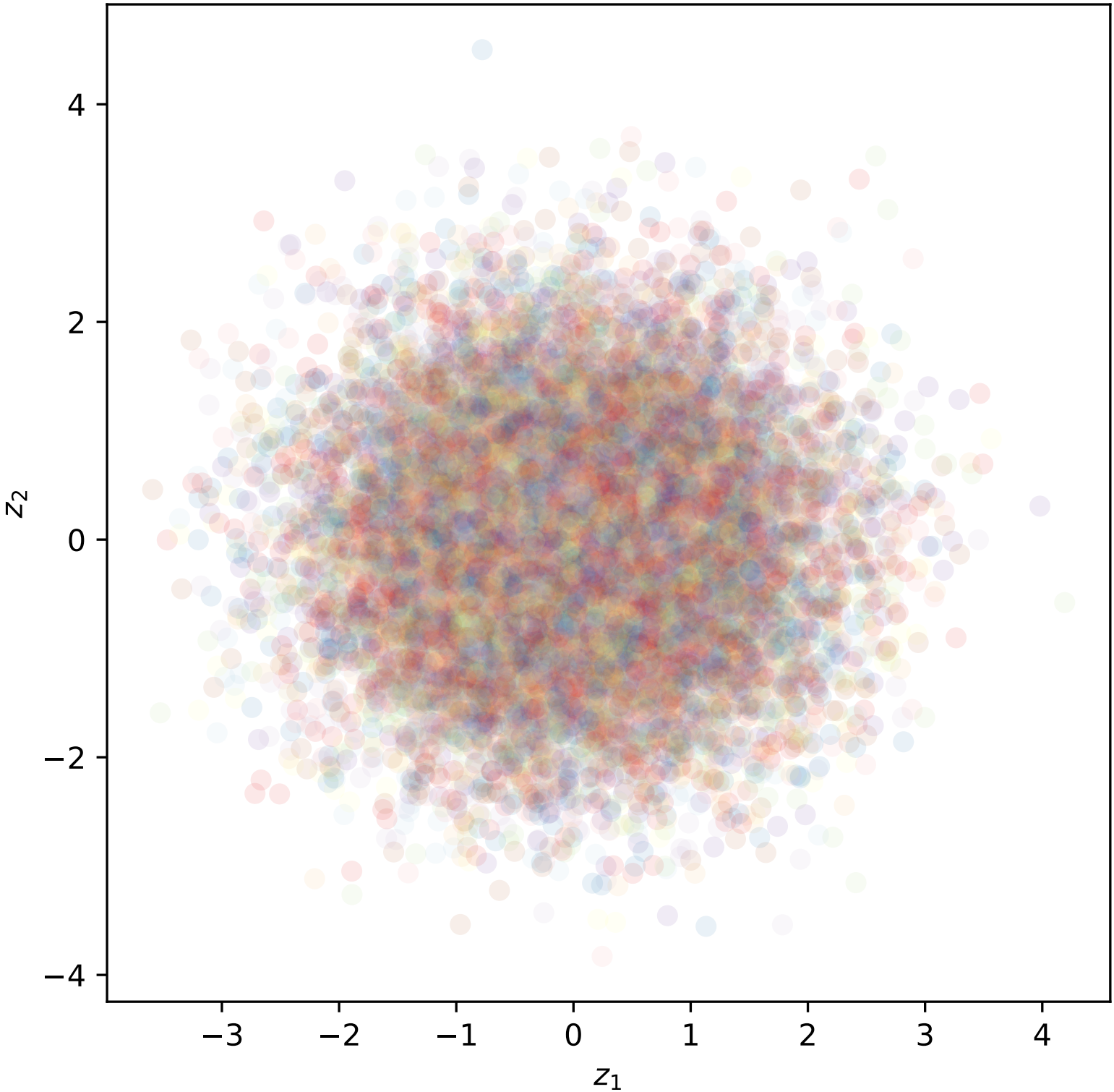
/62

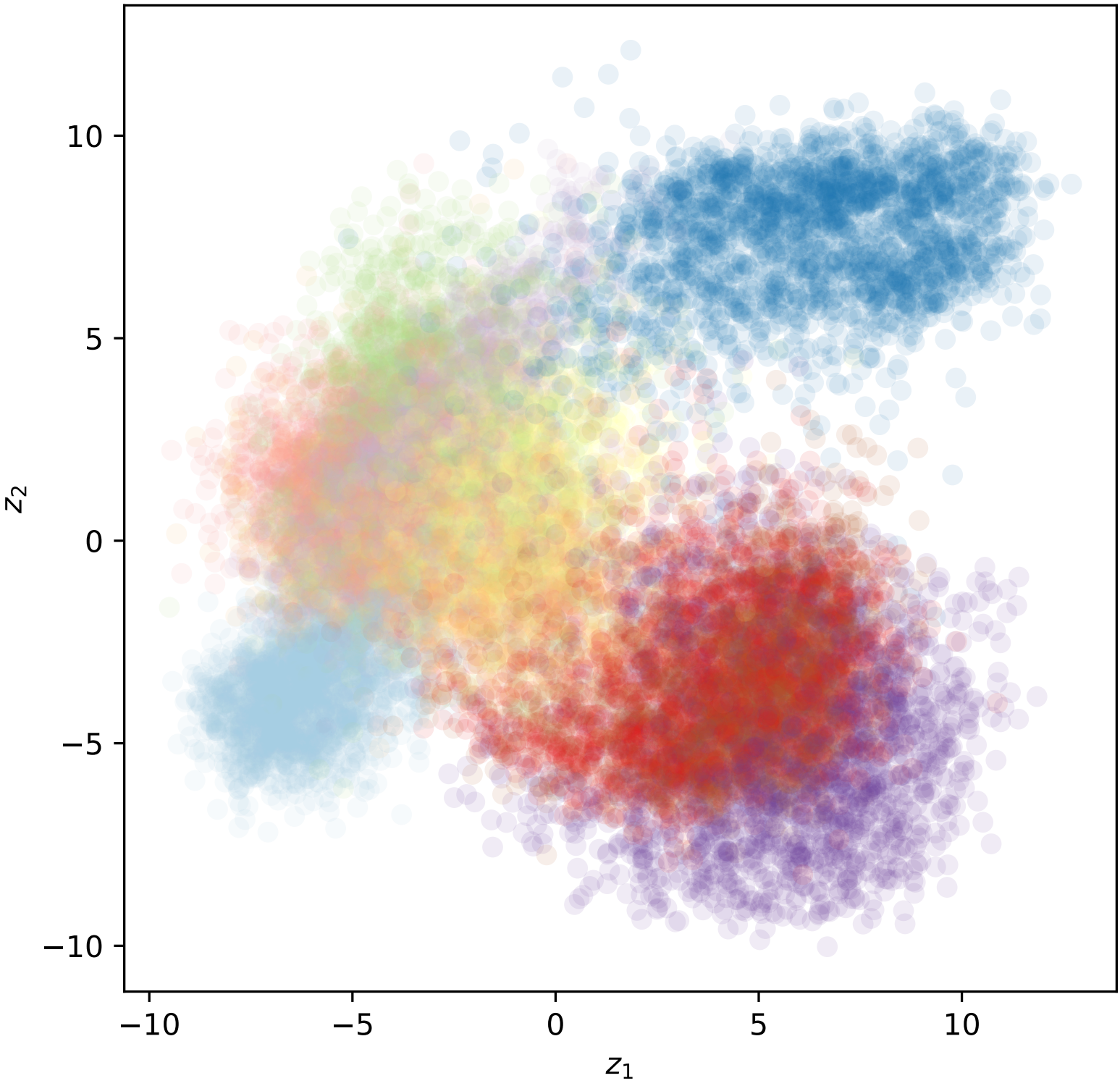
26

A semantically meaningful latent space

The KL-term enforces simplicity in the latent space, encouraging learned semantic structure and *disentanglement*


4
3
2
1
0
−1
−2
−3
z_2
−3
−2
−1
0
1
2
3
4
z_1


4
2
0
−2
−4
z_2
−3
−2
−1
0
1
2
3
4
z_1


10
5
0
−5
−10
z_2
−10
−5
0
5
10
z_1

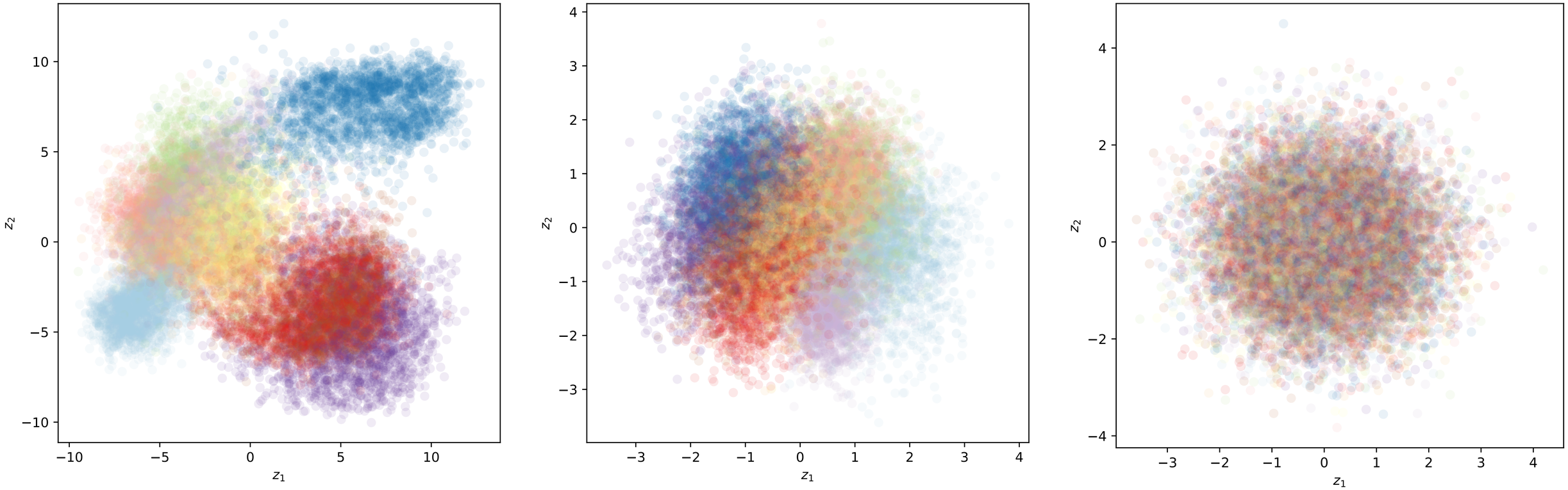
More latent regularization

Pure reconstruction

A semantically meaningful latent space

The KL-term enforces simplicity in the latent space, encouraging learned semantic structure and *disentanglement*

Pure reconstruction ⟷ *More latent regularization*



Rate-distortion theory: *neural compression*

Autoencoding is a form of (neural) compression!

$$\text{ELBO} = \left\langle \log p_\vartheta(x \mid z) \right\rangle_{q_\varphi} - D_{\text{KL}} \left(q_\varphi(z \mid x) \,\|\, p(z) \right)$$