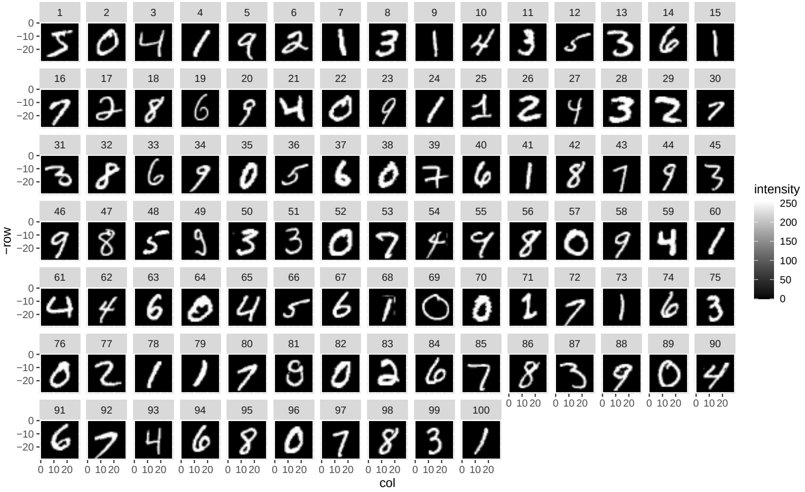
Principal Components Analysis

Toby Dylan Hocking

Motivation: MNIST digits data



Set of digits is represented as a matrix

- Each digit image in MNIST data set is a matrix of 28×28 pixel intensity values, $x_i \in \{0, \ldots, 255\}^{784}$.
- Each of the images is a row in the data matrix.
- Each of the columns is a pixel.
- All images on last slide represented by a data matrix with $n = 100$ rows/images and $p = 784$ columns/pixels.

Background/motivation: dimensionality reduction

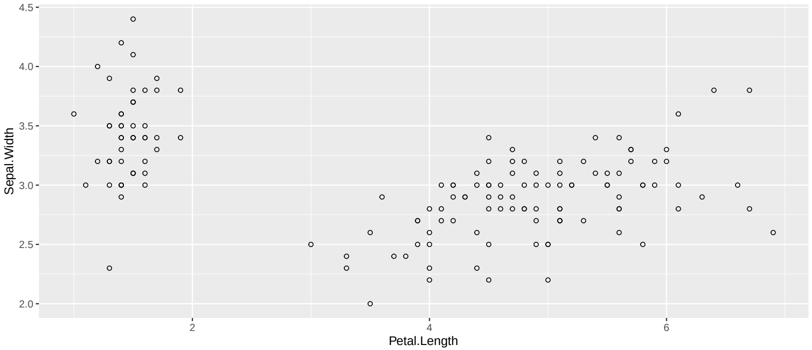
- High dimensional data are difficult to visualize.
- For example each observation/example in the MNIST data is of dimension 28 x 28 $= 784$ pixels.
- We would like to map each observation into a lower-dimensional space for visualization / understanding patterns in the data.

Example: 2d iris data

- Simpler example: iris.
- One row for each flower (only 6 of 150 shown below).
- One column for each measurement/dimension.

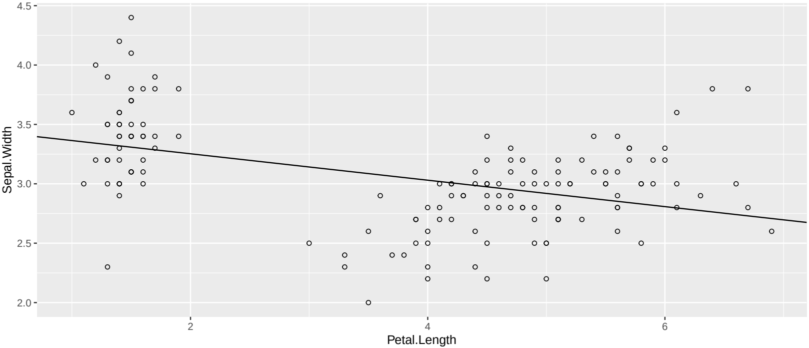
```
##   Sepal.Width Petal.Length
## 1         3.5          1.4
## 2         3.0          1.4
## 3         3.2          1.3
## 4         3.1          1.5
## 5         3.6          1.4
## 6         3.9          1.7
```

Example: 2d iris data



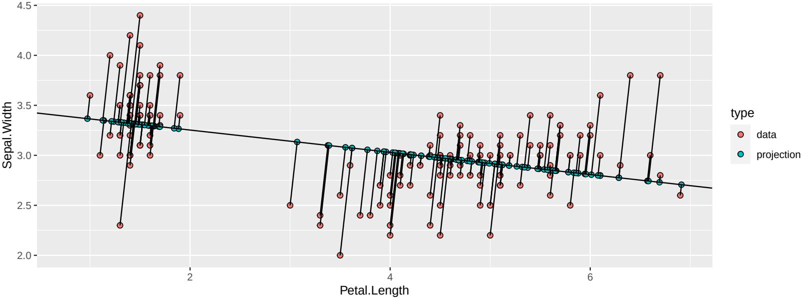
Project 2d data onto 1d subspace (line)

Why this line?



Principal Components Projection

The first principal component is the line which minimizes the reconstruction error, squared distance between projection and data.



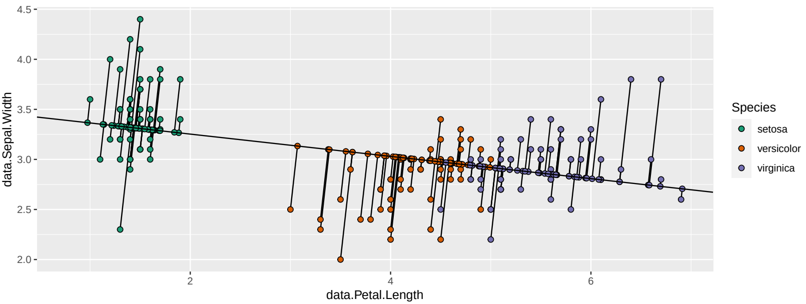
Mathematical representation

Each of the n inputs $x_i \in \mathbb{R}^p$ where p is the input dimension, $p = 2$ for iris in previous slides, or $p = 784$ for the images of digits MNIST data.

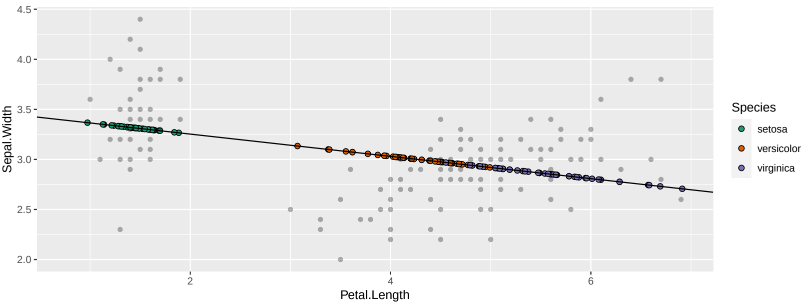
$$\min \sum_{i=1}^{n} ||x_i - \mu - V_q \lambda_i||^2.$$

- $\mu \in \mathbb{R}^p$ is mean vector.
- $V_q \in \mathbb{R}^{p \times q}$ is an orthogonal matrix (each column is an orthogonal unit vector).
- $\lambda_i \in \mathbb{R}^q$ is a vector of principal components (contribution of each unit vector).

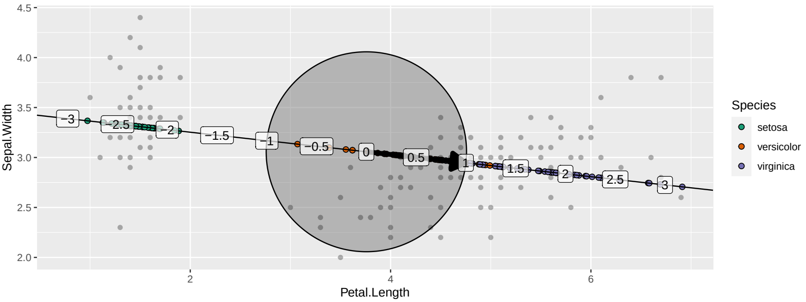
Map label onto projection



Map label onto projection

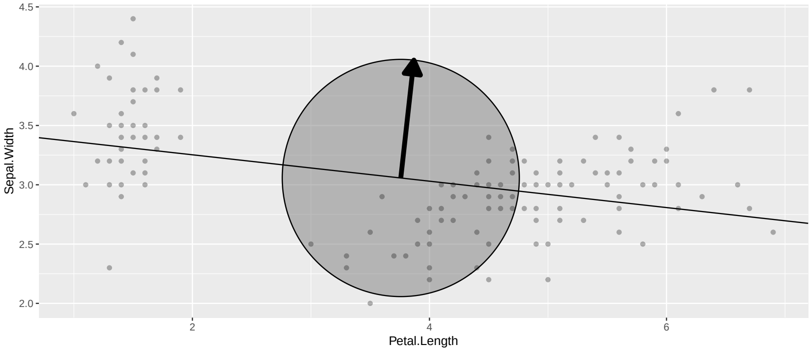


Principal component 1, amount along projection

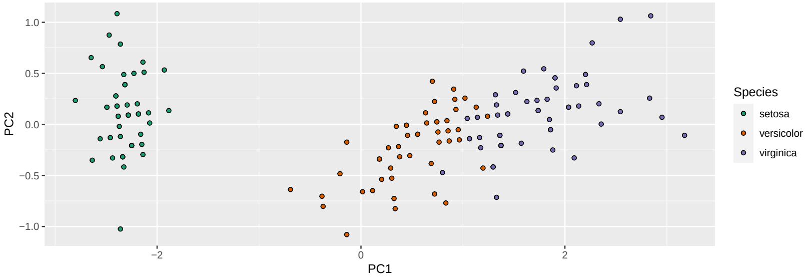


- 0 represents mean μ of data.
- $0 \rightarrow 1$ is an orthogonal unit vector, a principal direction, first column of V_q.
- Numbers in white boxes represent principal components, λ_i.

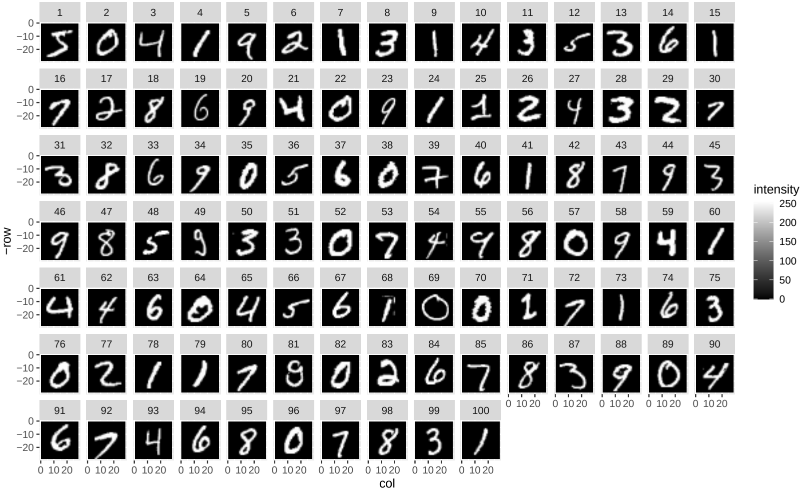
Principal component 2



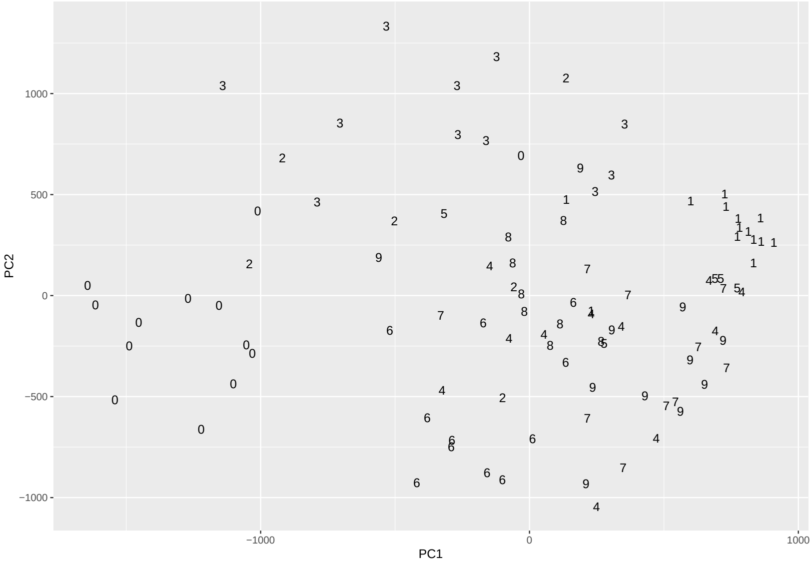
Re-plot using PC units



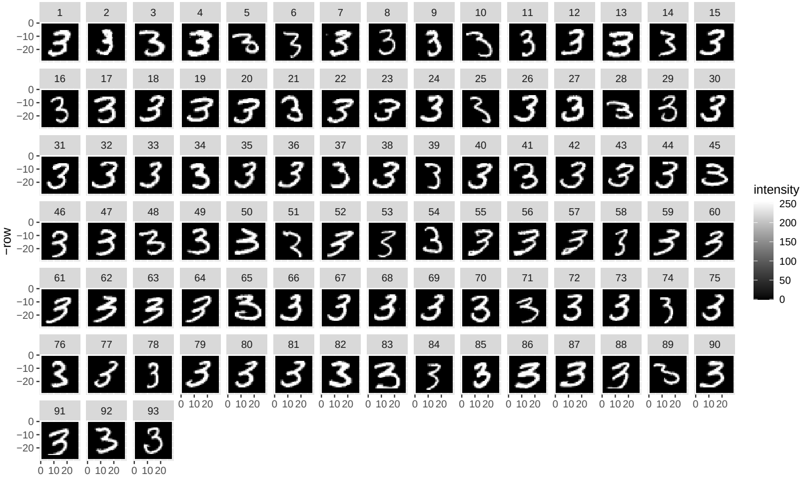
MNIST digits data



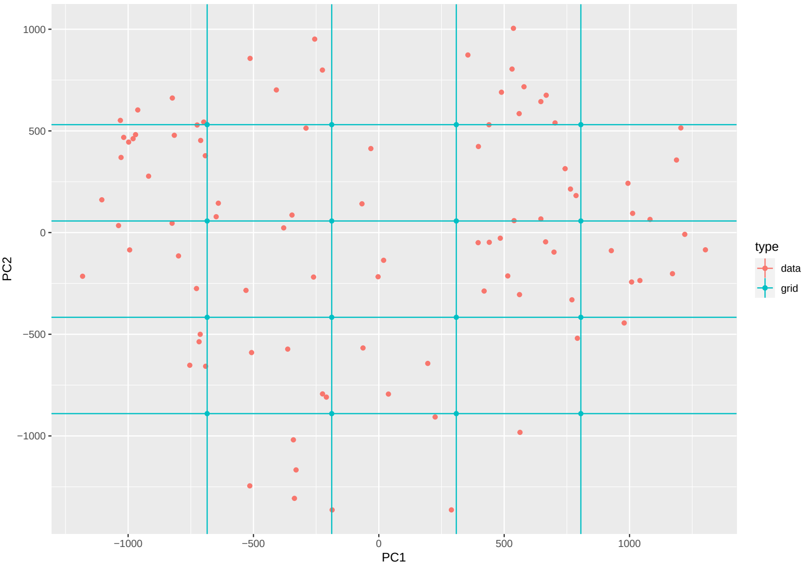
PCA with MNIST digit data



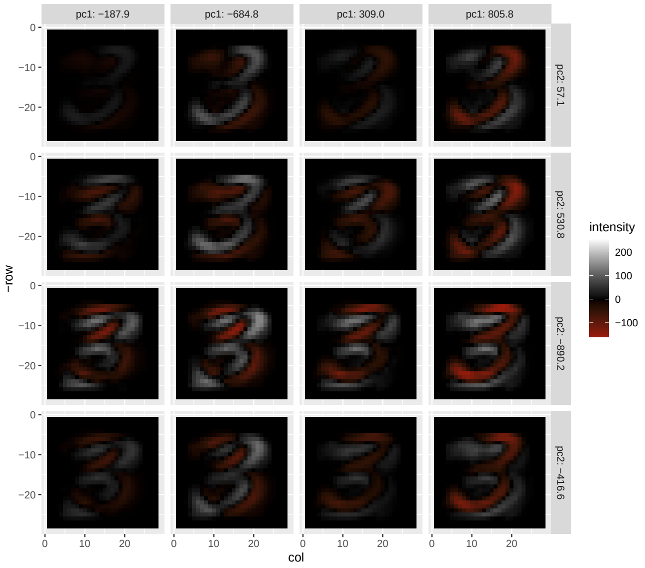
Another PCA on just one digit class



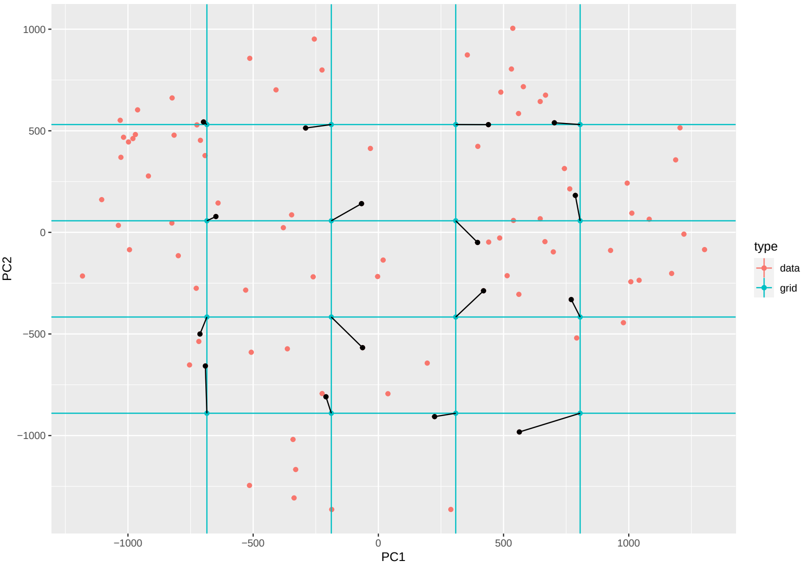
Mapping onto first two PCs



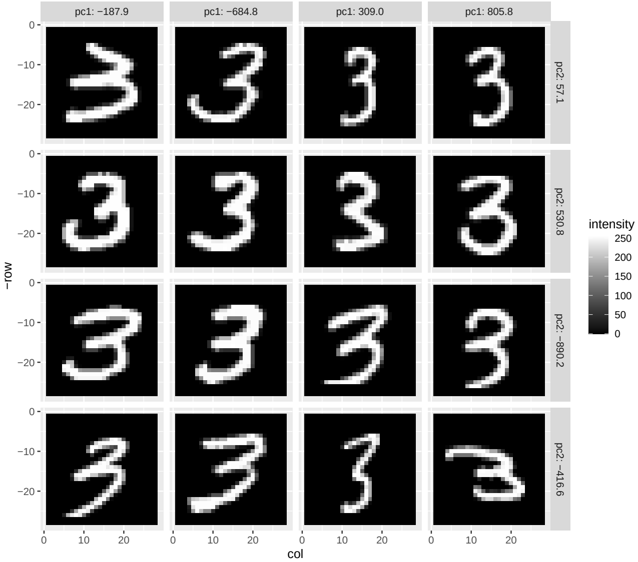
Reconstruction at grid points



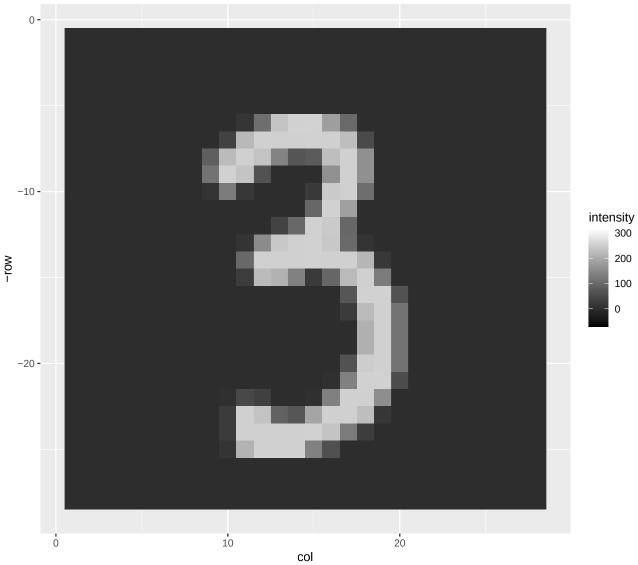
Highlight closest data point to each grid point



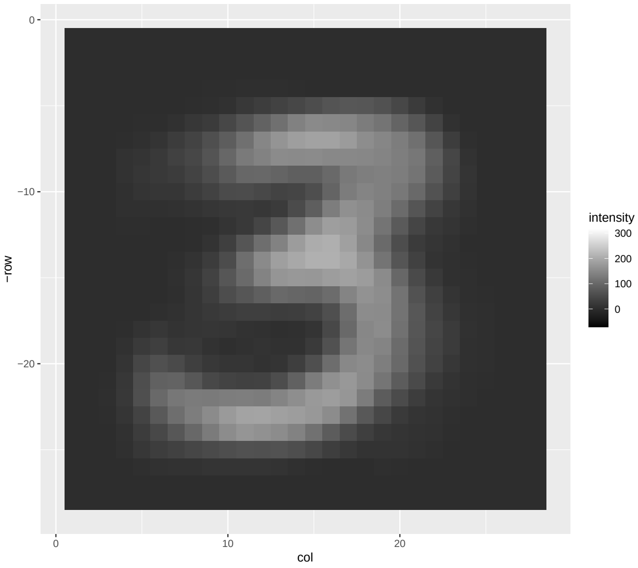
Digits highlighted



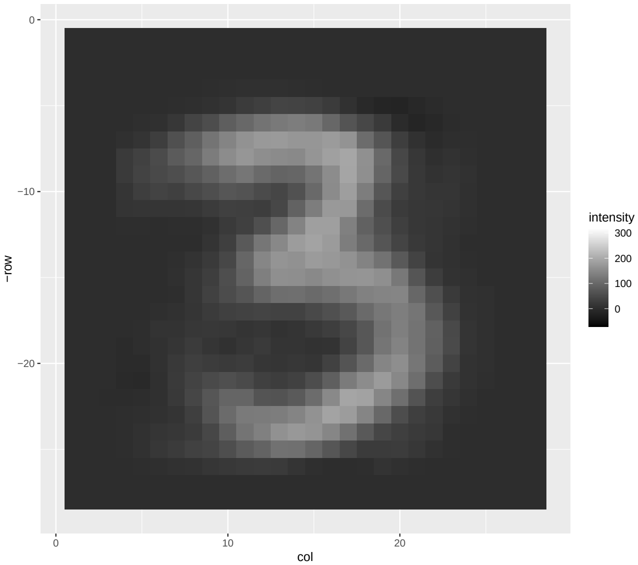
One digit



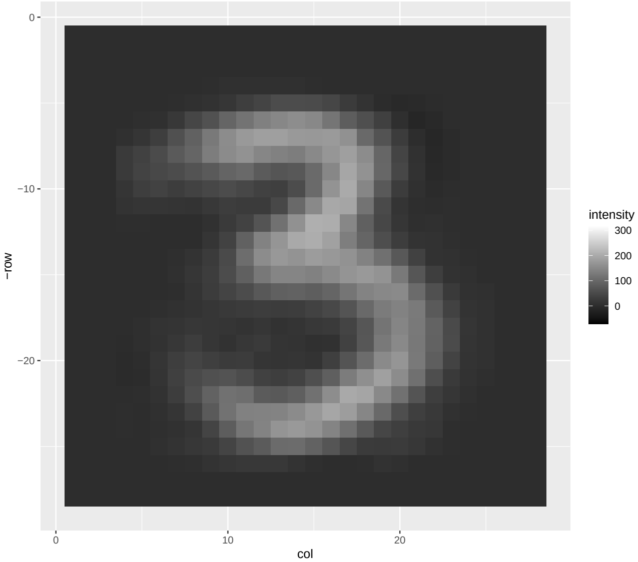
Reconstruction with no components (mean)



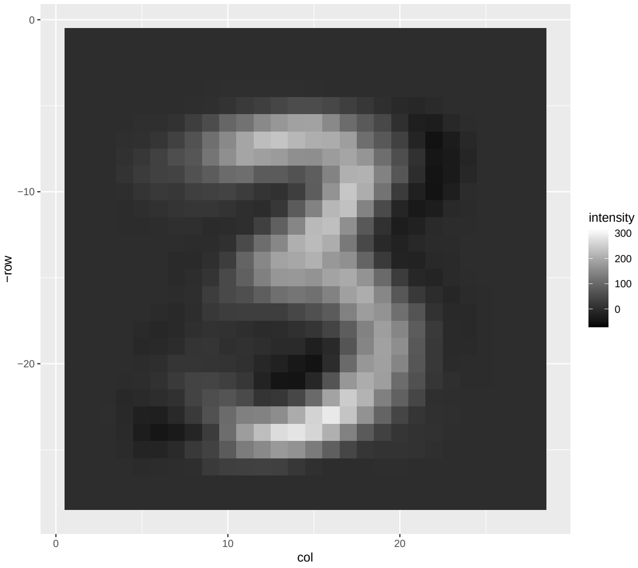
Reconstruction with one PC



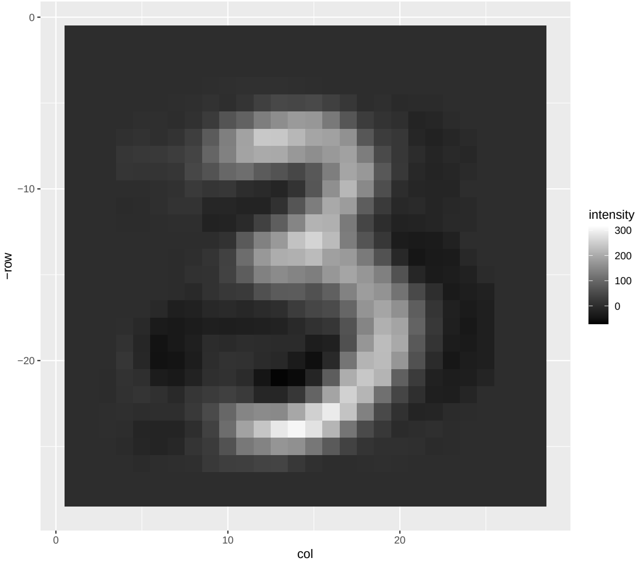
Reconstruction with 2 PCs



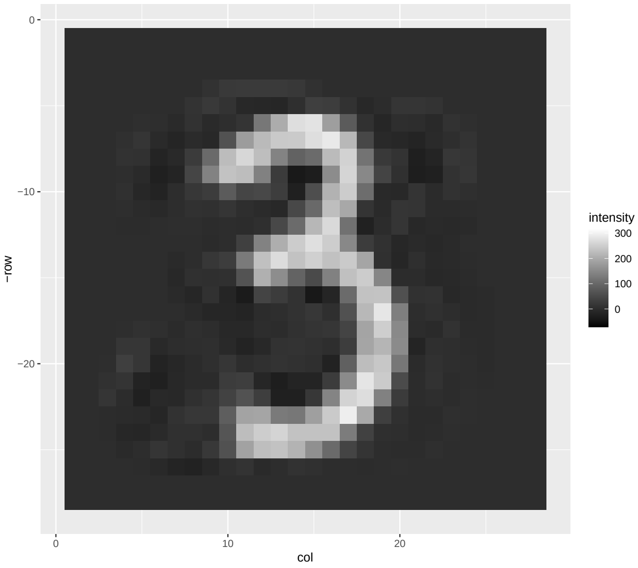
Reconstruction with 5 PCs



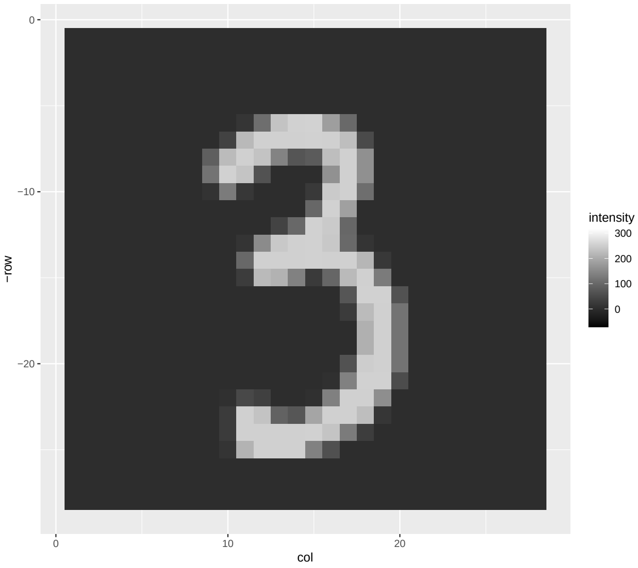
Reconstruction with 10 PCs



Reconstruction with 50 PCs



Reconstruction with all PCs



How to compute PCA?

SVD = Singular Value Decomposition (many algorithms available to compute).

$$X = UDV^T$$

- $X \in \mathbb{R}^{n \times p}$ data matrix.
- $U \in \mathbb{R}^{n \times p}$ orthogonal matrix.
- $D \in \mathbb{R}^{p \times p}$ diagonal matrix.
- $V \in \mathbb{R}^{p \times p}$ orthogonal matrix.
- The V_q we want for PCA is the first q columns of V.
- The columns of UD are the principal components, λ_i values.

Possible exam questions

- When is the max number of principal components equal to the number of rows of the data matrix?
- When is the max number of principal components equal to the number of columns of the data matrix?