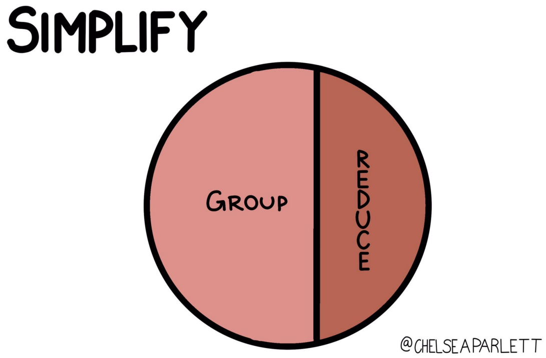
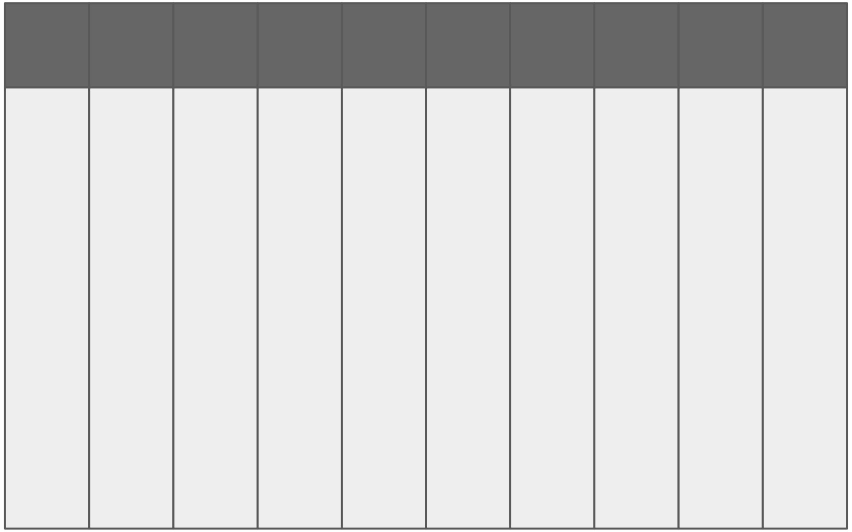


Principal Components Analysis (PCA)

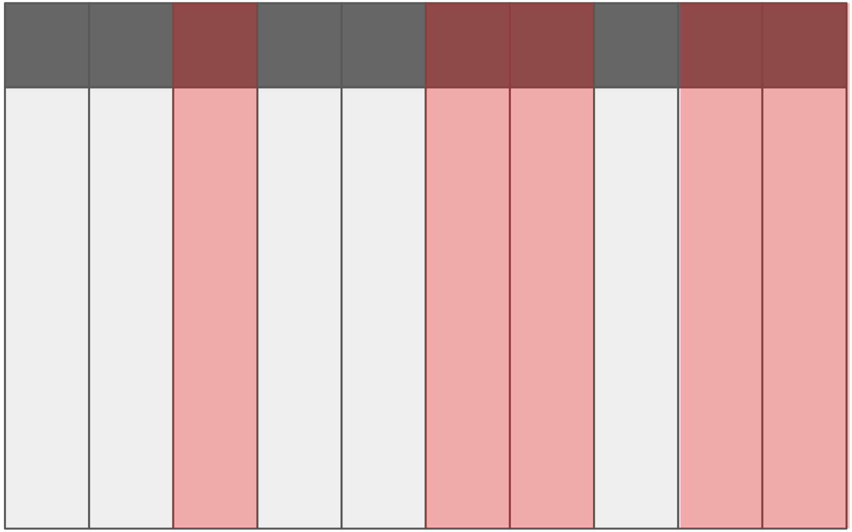
Dr. Chelsea Parlett-Pelleriti

PCA Overview

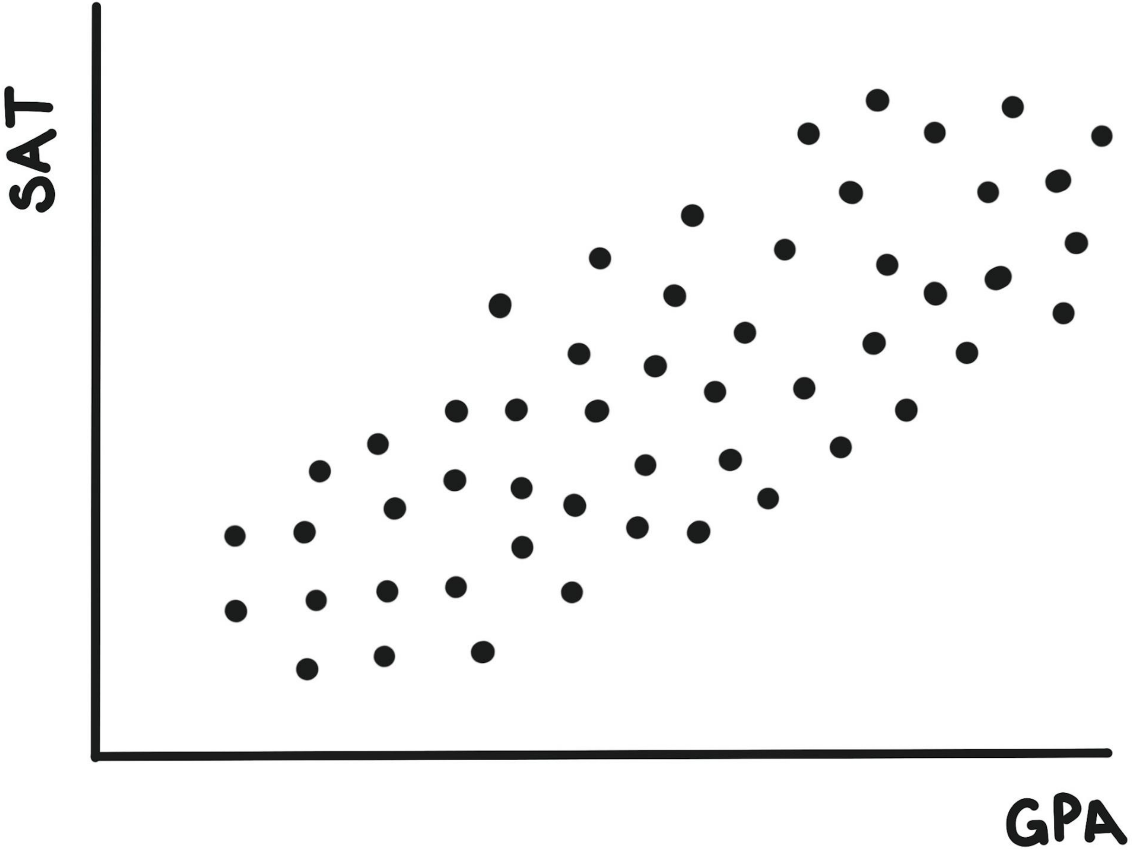
Dimensionality Reduction



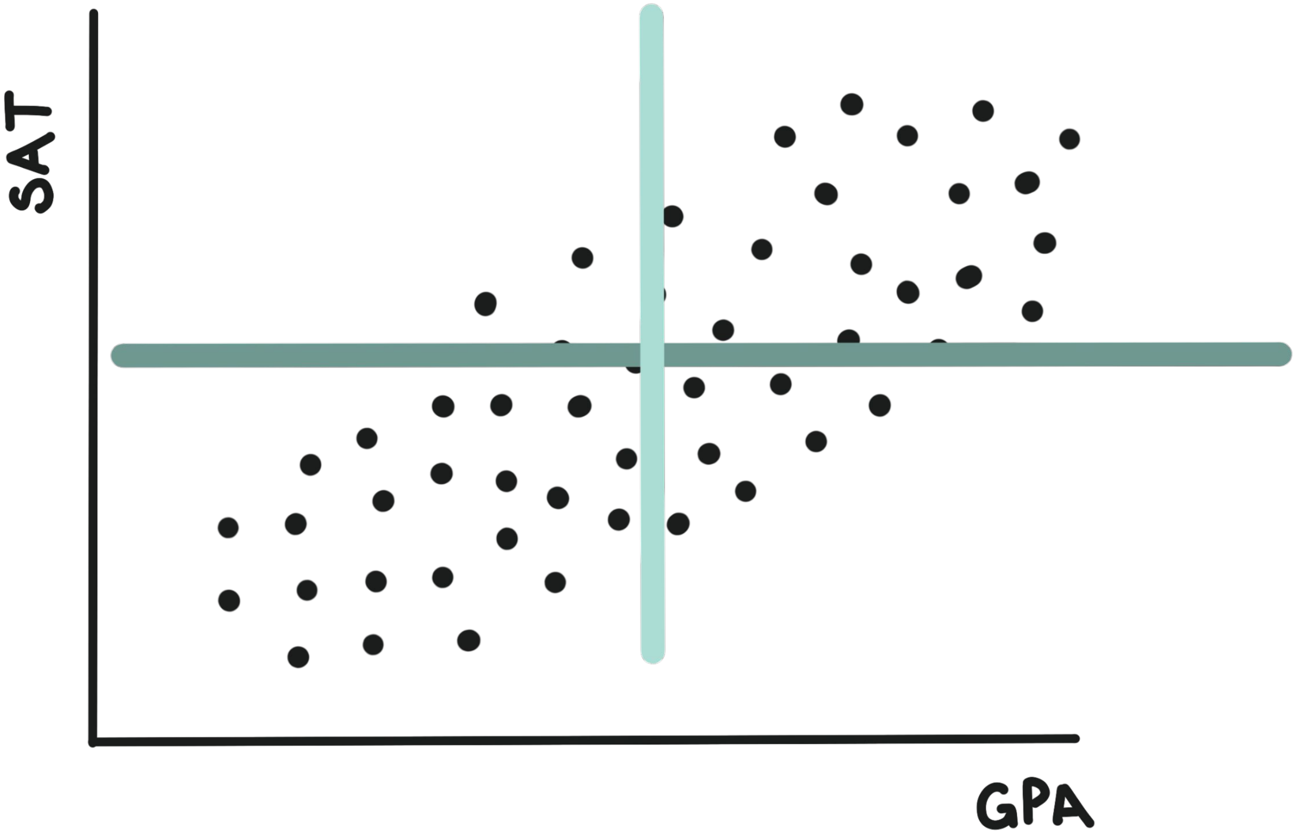
Dimensionality Reduction



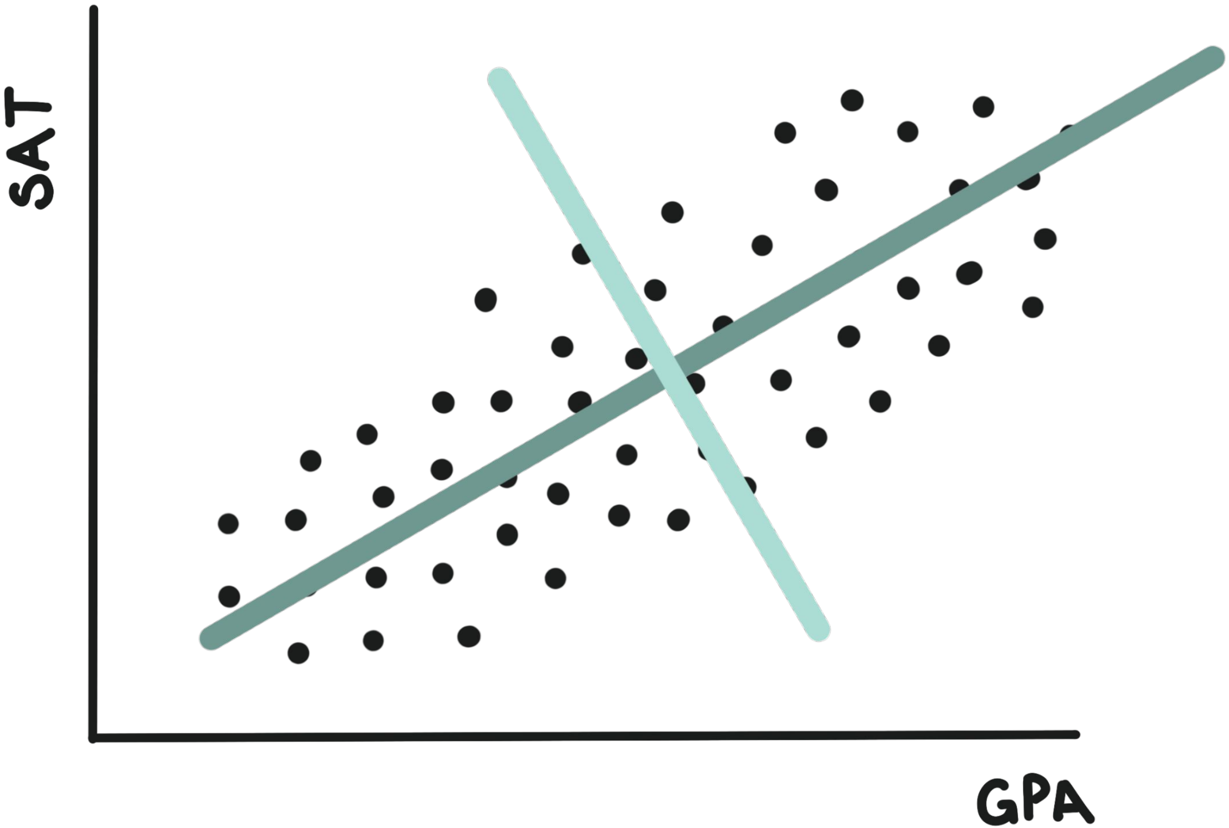
Variation



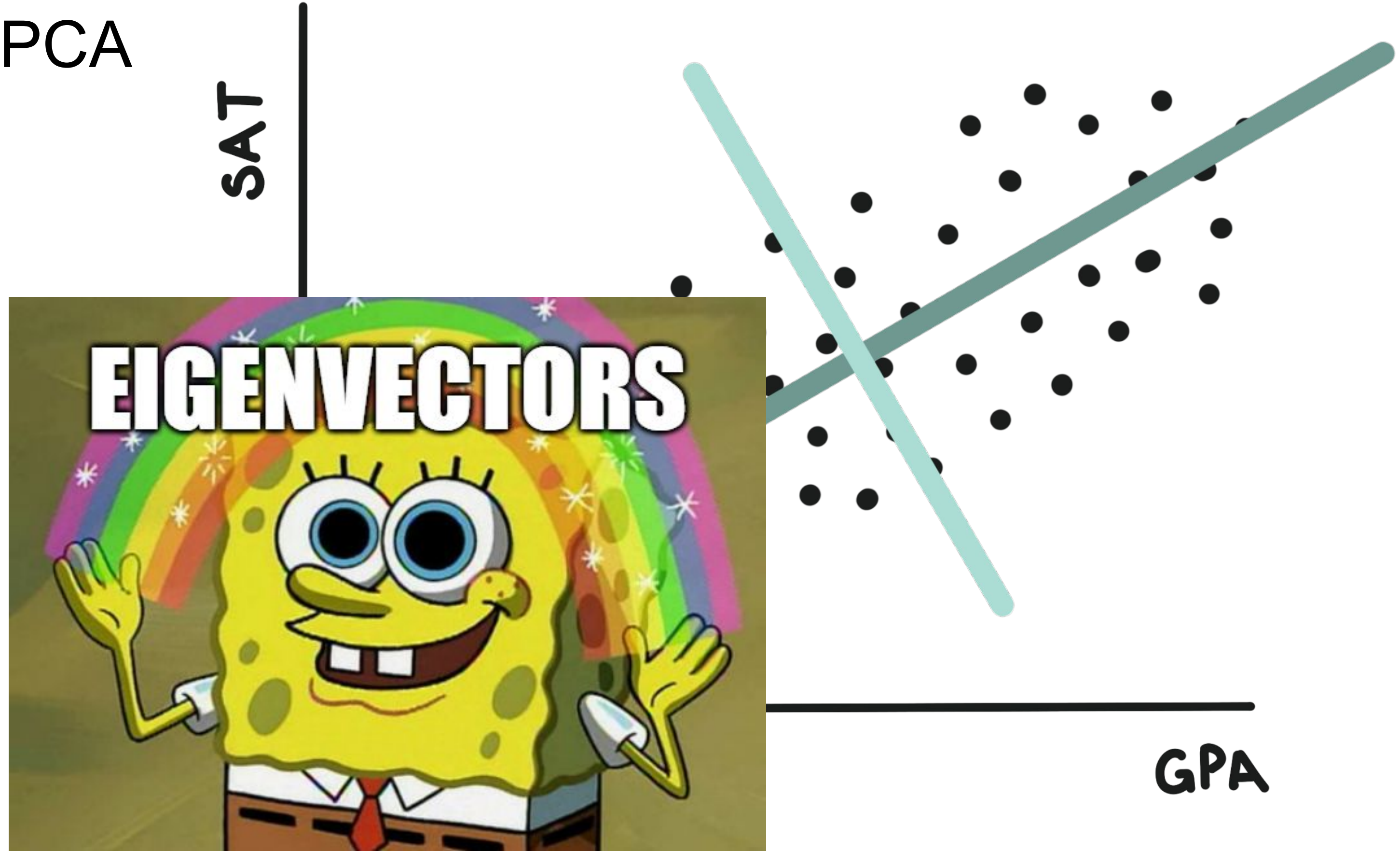
PCA



PCA

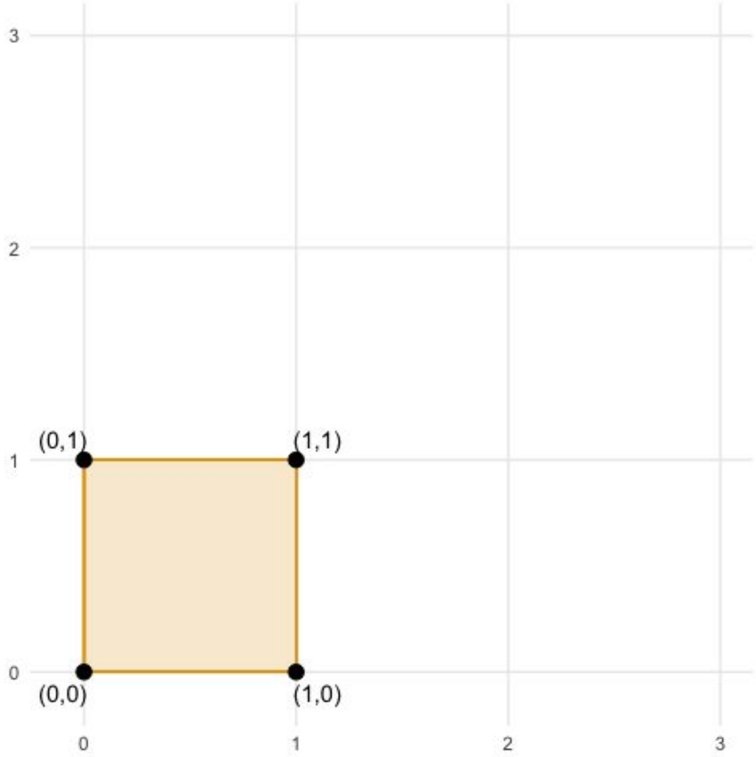


PCA



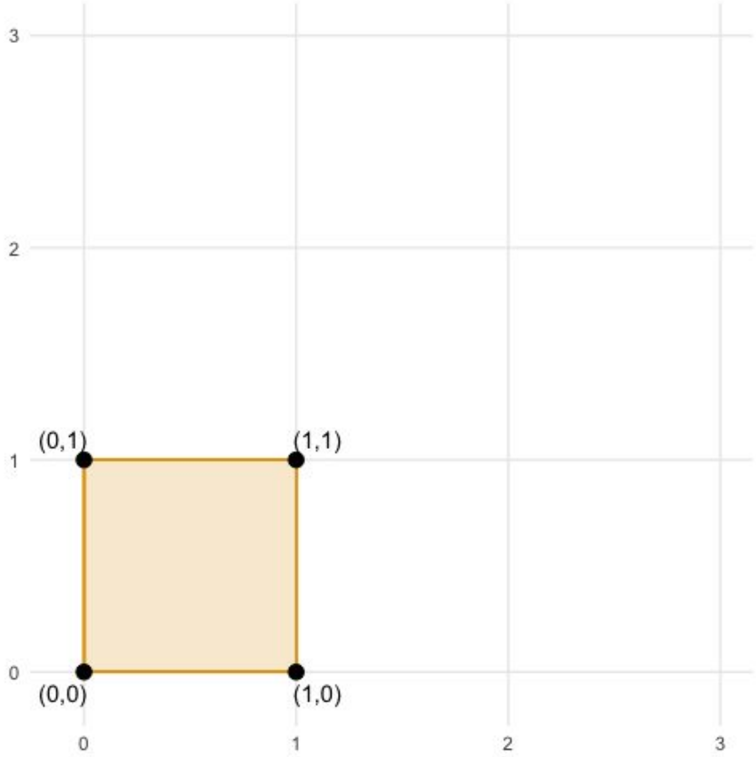
Eigendecomposition

Eigendecomposition



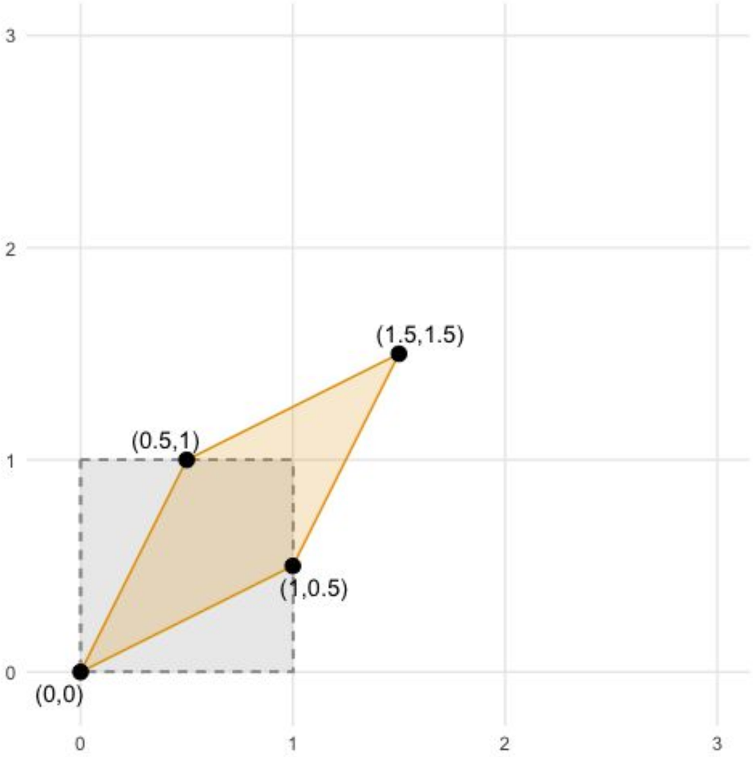
Eigendecomposition

$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



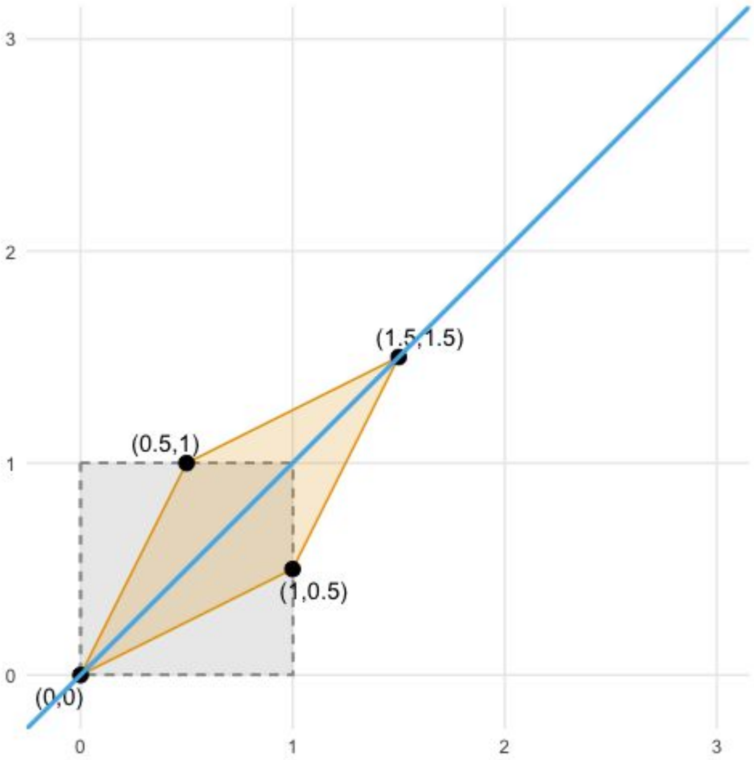
Eigendecomposition

$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



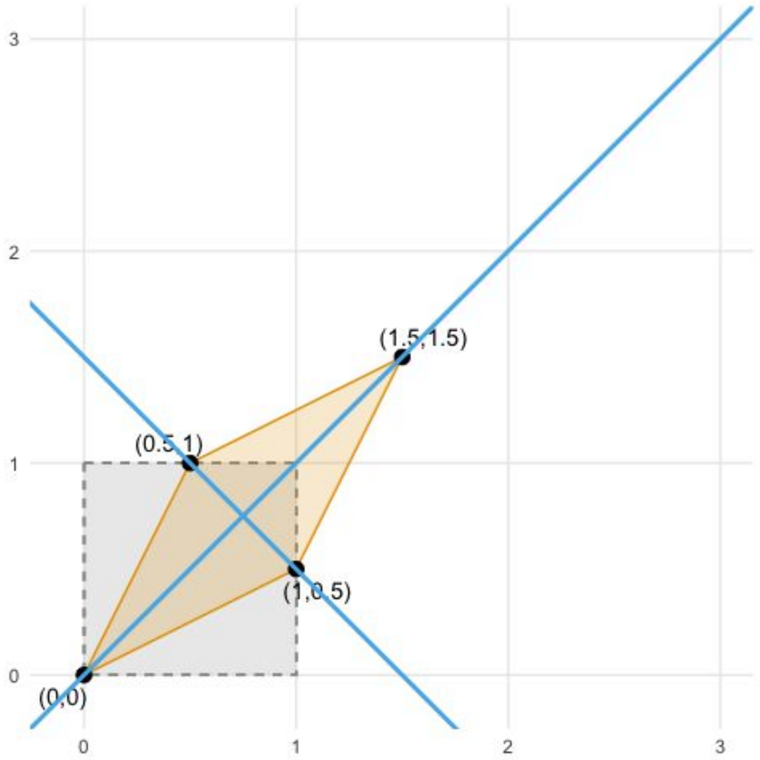
Eigendecomposition

$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



Eigendecomposition

$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



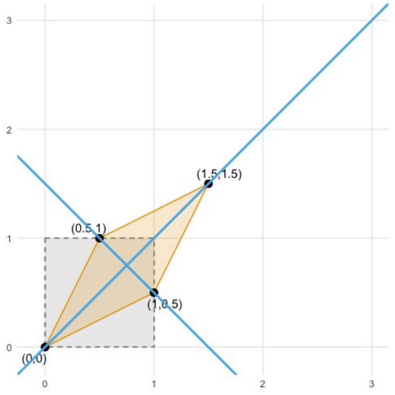
Eigendecomposition

1. What are the directions of stretch and squish?

2. How much do we stretch and squish?

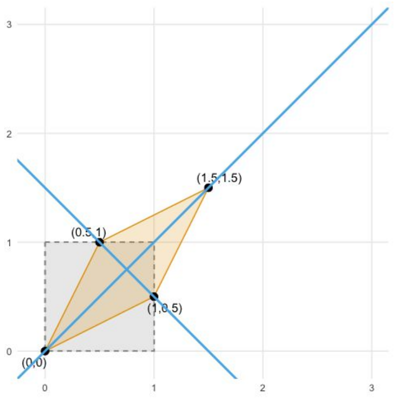
Eigendecomposition

$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad Ax = \lambda x$$



$$|A - \lambda I| = 0$$

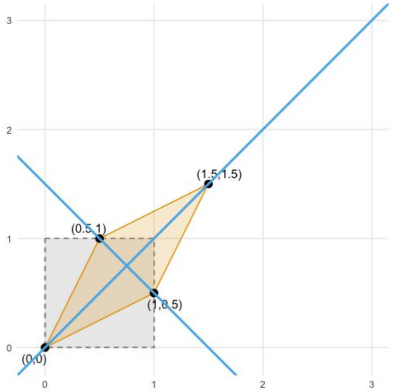
Eigendecomposition



$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad Ax = \lambda x$$

$$|A - \lambda I| = 0$$

$$|A - \lambda I| = 0$$

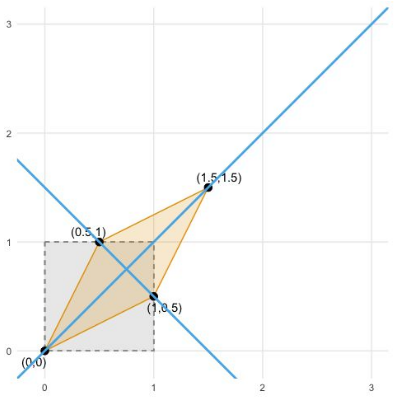


$$\begin{bmatrix} 1-\lambda & 0.5 \\ 0.5 & 1-\lambda \end{bmatrix}$$

$$(1-\lambda)(1-\lambda) - (0.5)(0.5)$$

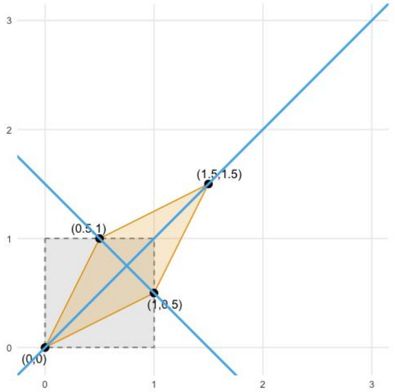
$$|A - \lambda I| = 0$$

$$\begin{bmatrix} 1-\lambda & 0.5 \\ 0.5 & 1-\lambda \end{bmatrix}$$



$$(1 - \lambda)(1 - \lambda) - (0.5)(0.5)$$

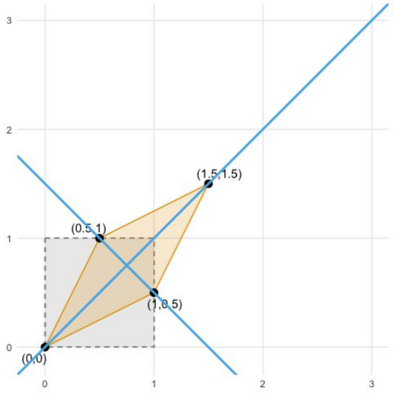
$$\lambda^2 - 2\lambda + 0.75$$



$$(\lambda - 1.5)(\lambda - 0.5)$$

$$\lambda = 1.5, 0.5$$

$$\lambda^2 - 2\lambda + 0.75$$

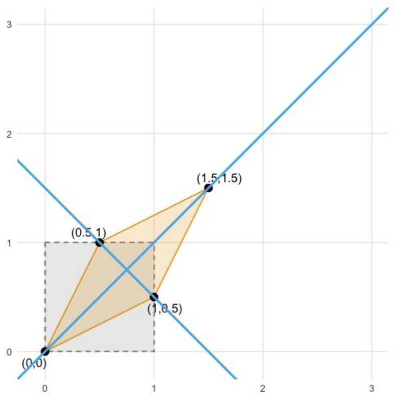


$$(\lambda - 1.5)(\lambda - 0.5)$$

$$\lambda = 1.5, 0.5$$

$$\lambda^2 - 2\lambda + 0.75$$

$$(\lambda - 1.5)(\lambda - 0.5)$$

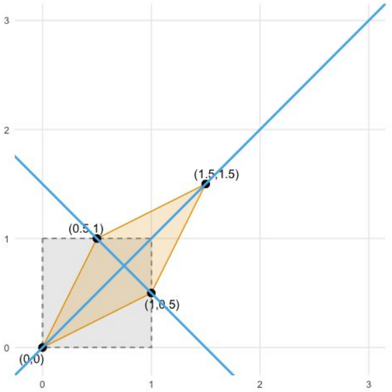


$$\lambda = 1.5, 0.5$$

$$\lambda = 1.5, 0.5$$

$$\begin{bmatrix} 1-\lambda & 0.5 \\ 0.5 & 1-\lambda \end{bmatrix}$$

$$\begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix} \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$$



$$\lambda = 1.5, 0.5$$

$$\begin{bmatrix} 1-\lambda & 0.5 \\ 0.5 & 1-\lambda \end{bmatrix}$$

$$\begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix} \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$$

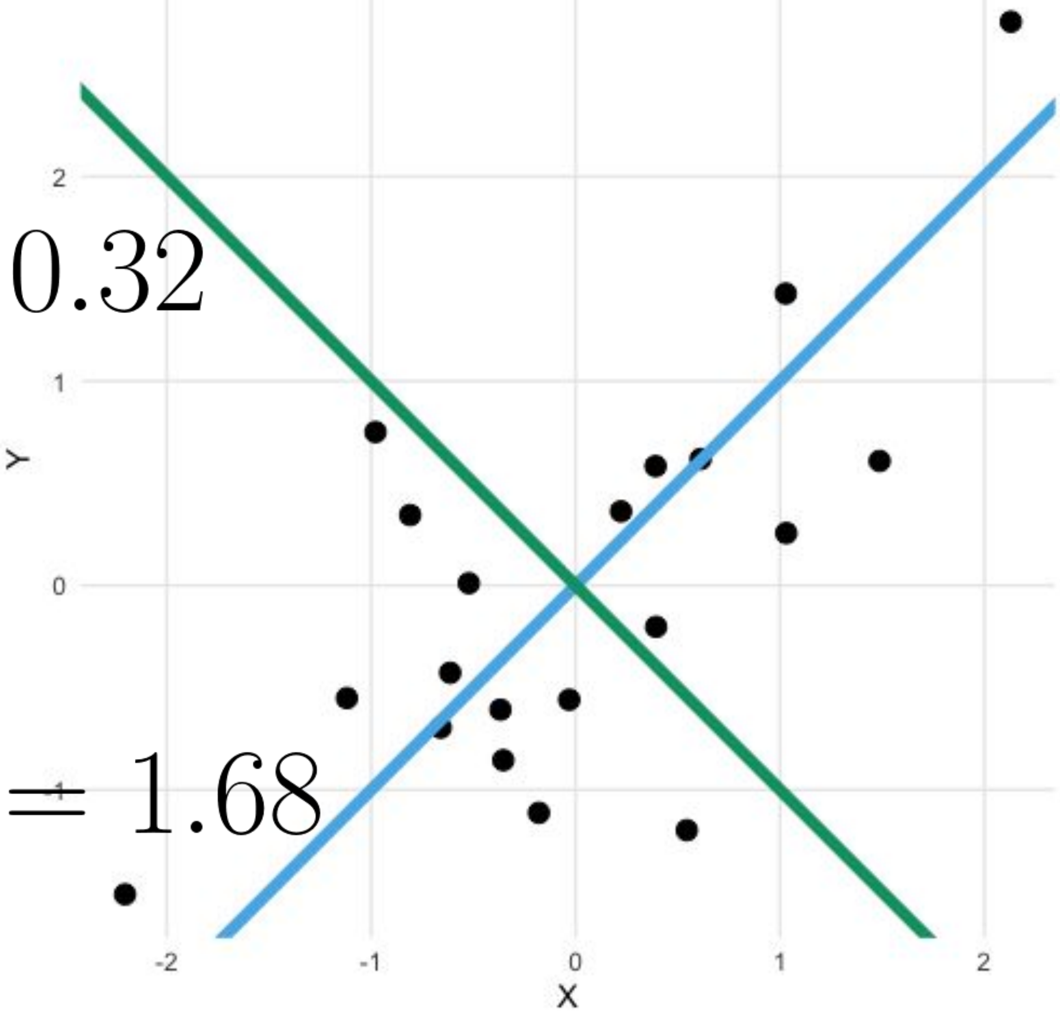
Principal Component Analysis

PCA

$$cov(\mathbf{X}) = \begin{bmatrix} 1 & 0.68 \\ 0.68 & 1 \end{bmatrix}$$

PCA

$$\lambda_2 = 0.32$$

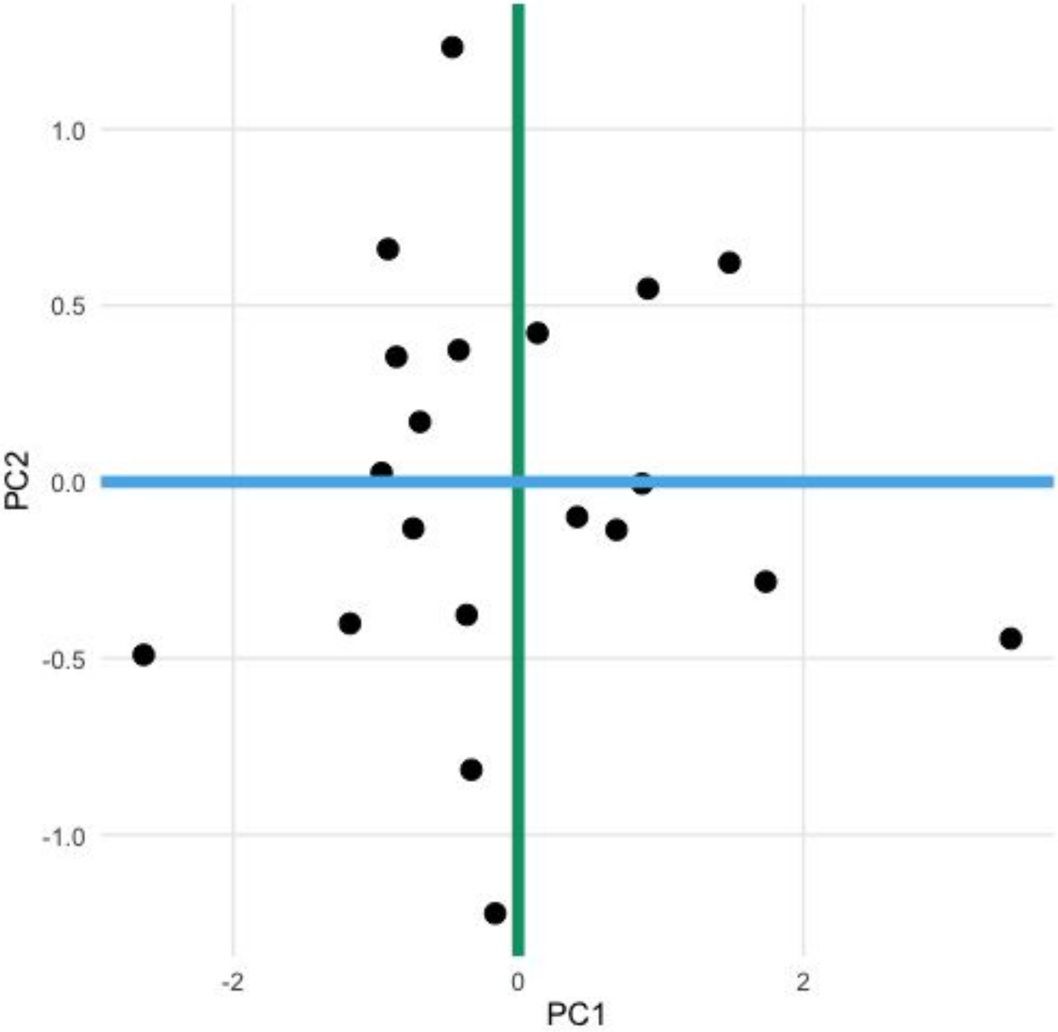


$$\lambda_1 = 1.68$$

PCA

$$\text{eigenvec}_1 = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}$$

$$\text{eigenvec}_2 = \begin{bmatrix} 0.71 \\ -0.71 \end{bmatrix}$$



Loadings

Weights

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.314	0.145	-0.676	-0.347	-0.241	0.494	0.018
Education	0.237	0.444	-0.401	0.240	0.622	-0.357	0.103
Age	0.484	-0.135	-0.004	-0.212	-0.175	-0.487	-0.657
Residence	0.466	-0.277	0.091	0.116	-0.035	-0.085	0.487
Employ	0.459	-0.304	0.122	-0.017	-0.014	-0.023	0.368
Savings	0.404	0.219	0.366	0.436	0.143	0.568	-0.348
Debt	-0.067	-0.585	-0.078	-0.281	0.681	0.245	-0.196
Credit cards	-0.123	-0.452	-0.468	0.703	-0.195	-0.022	-0.158

https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/multivariate/how-to/principal-components/interpret-the-results/all-statistics-and-graphs/#principal-components-pc

Component Scores

New scores

PC1	PC2
-0.809722735	0.08256115
-3.123826392	-0.04680691
-0.009346561	0.97125244
0.812390987	0.16654000
0.041430260	-0.59984809
-2.147002620	-0.60889442
1.413295183	-0.31382581
-1.688064279	0.19483917
-1.633290817	0.04825966
-0.456644805	0.82142855
-0.010438739	0.03708409
0.566679507	-0.15038116
0.512723998	-1.18058477
-0.755004514	-0.74959709
-2.550657933	0.21442421
0.201041435	-0.24101236
1.657943766	0.89216807
0.918123342	-0.57273573
-0.279086907	-0.70387691
0.467917136	-1.47696045

Components

Component Loadings

```
loadings2 = pd.DataFrame(p.components_,
    columns = ["A", "B", "C"])
loadings2.index = ["PC1", "PC2", "PC3"]
loadings2 = loadings2.T
loadings2
```

✓ 0.3s

	PC1	PC2	PC3
A	-0.696795	-0.104304	0.709646
B	-0.680594	-0.216177	-0.700043
C	-0.226426	0.970767	-0.079642

Component Scores

```
pscores = pd.DataFrame(p.transform(d))
pscores
```

✓ 0.3s

	0	1	2
0	0.799567	-1.420019	-0.121829
1	1.340988	-0.363126	0.010270
2	-1.828154	0.980559	-0.018594
3	-2.943051	1.111052	0.583680
4	2.404853	1.158168	0.342818

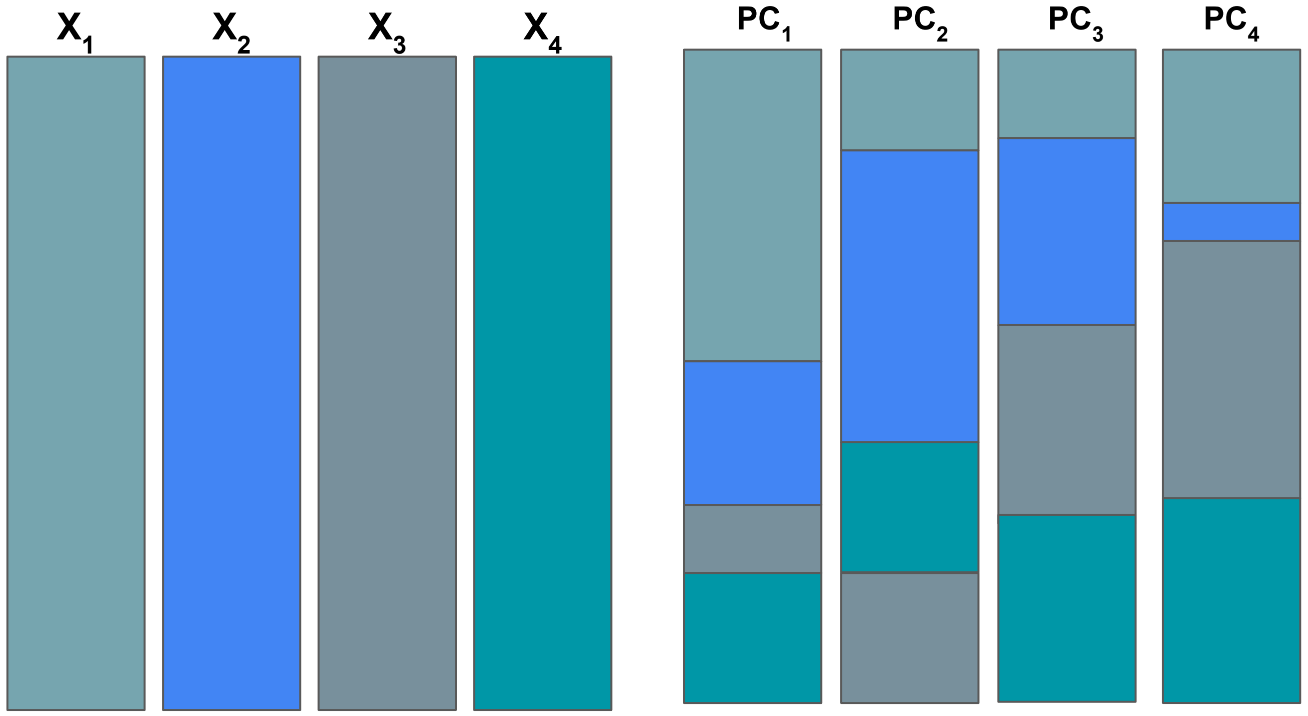
Original Data

	A	B	C
0	-0.495414	-0.150791	-1.547534
1	-0.889168	-0.840230	-0.654650
2	1.158441	1.046401	1.369631
3	2.349085	1.355365	1.700785

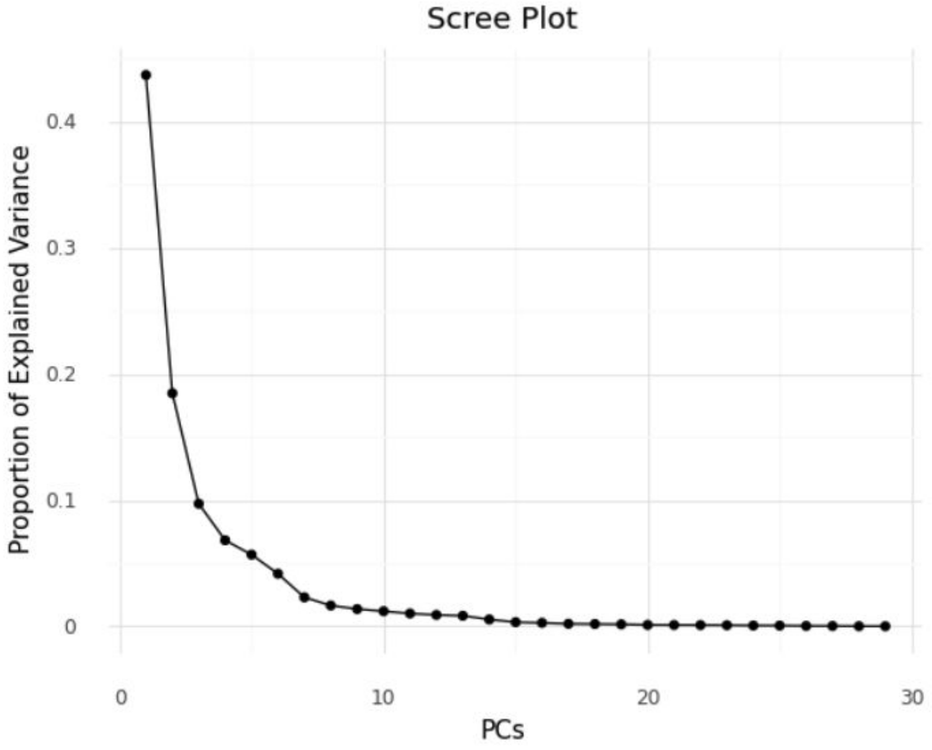
(-0.696795*-0.495414) +
(-0.680594*-0.150791) +
(-0.226426*-1.547534) ≈ 0.799

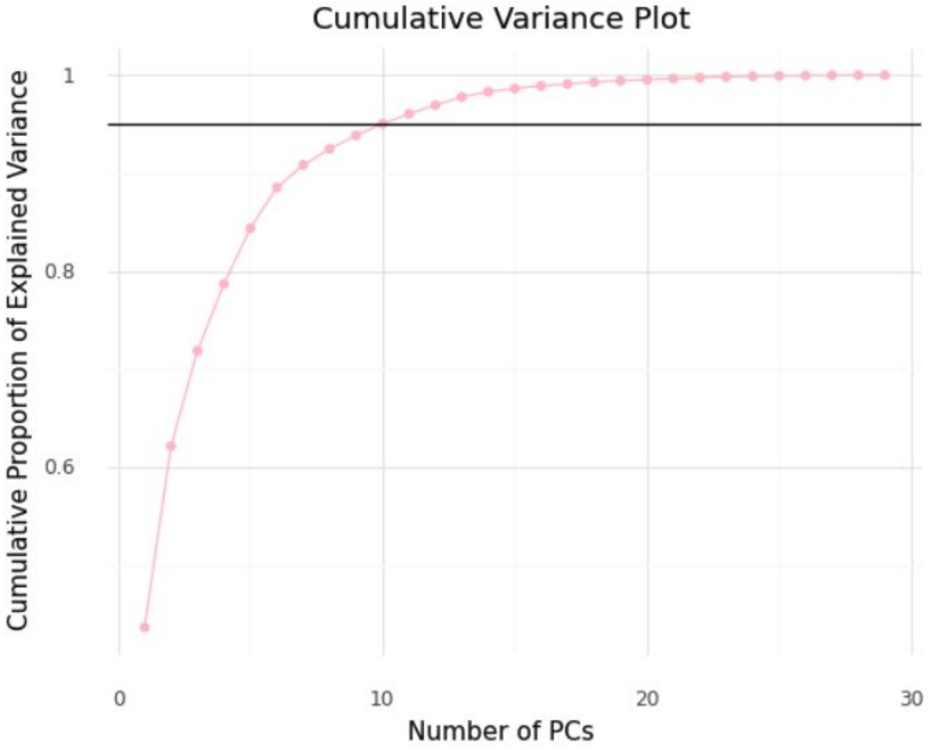
Dimensionality Reduction

PCA

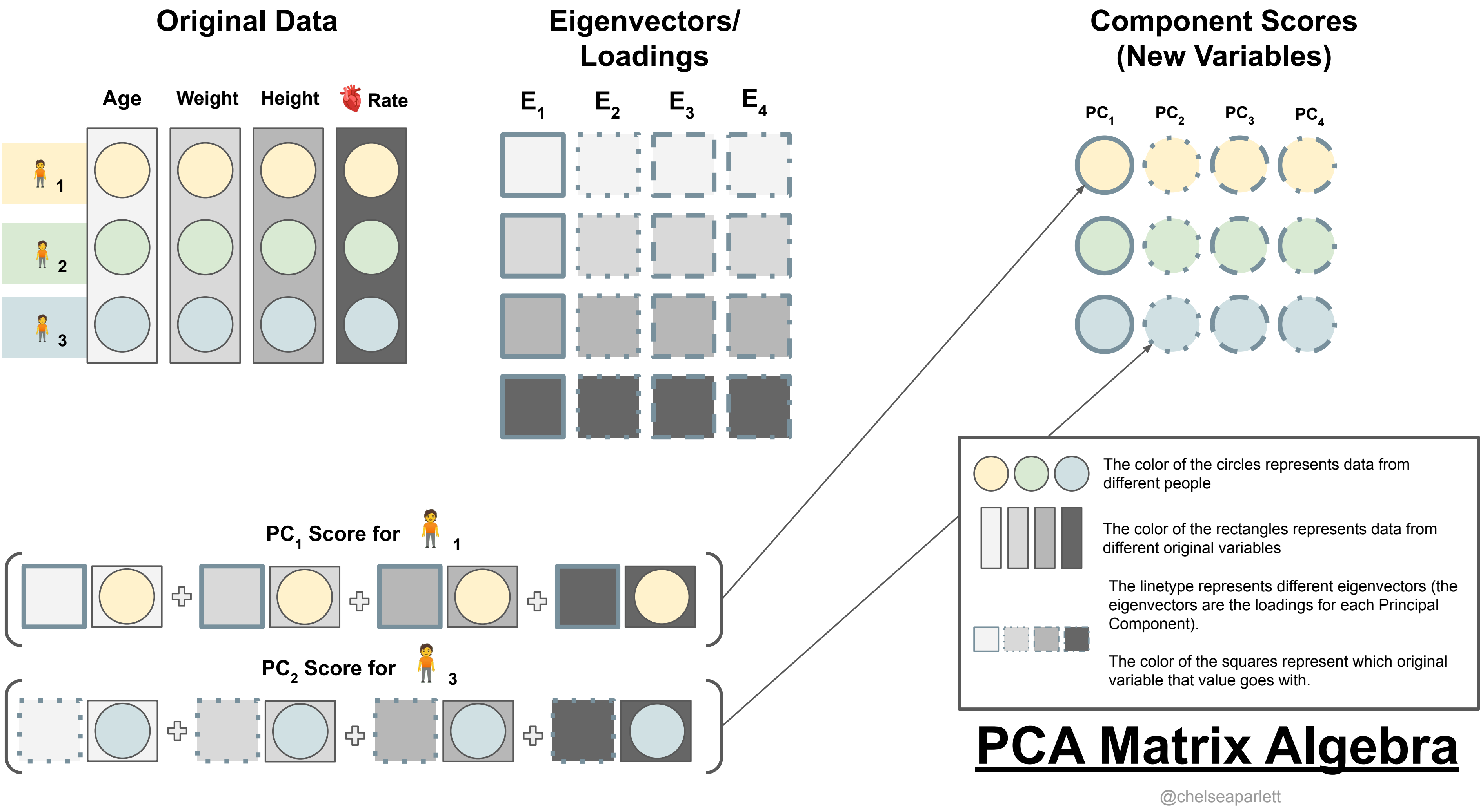


Scree and Cumulative Variance Plots





Recap


Original Data
Eigenvectors/ Loadings
Component Scores (New Variables)
Age
Weight
Height
Rate
1
2
3
E1
E2
E3
E4
PC1
PC2
PC3
PC4
PC1 Score for 1
PC2 Score for 3
The color of the circles represents data from different people
The color of the rectangles represents data from different original variables
The linetype represents different eigenvectors (the eigenvectors are the loadings for each Principal Component).
The color of the squares represent which original variable that value goes with.
PCA Matrix Algebra
@chelseaparlett

Why PCA?

- **Dimensionality Reduction**
- (Factor Analysis: Understanding which variables go together)