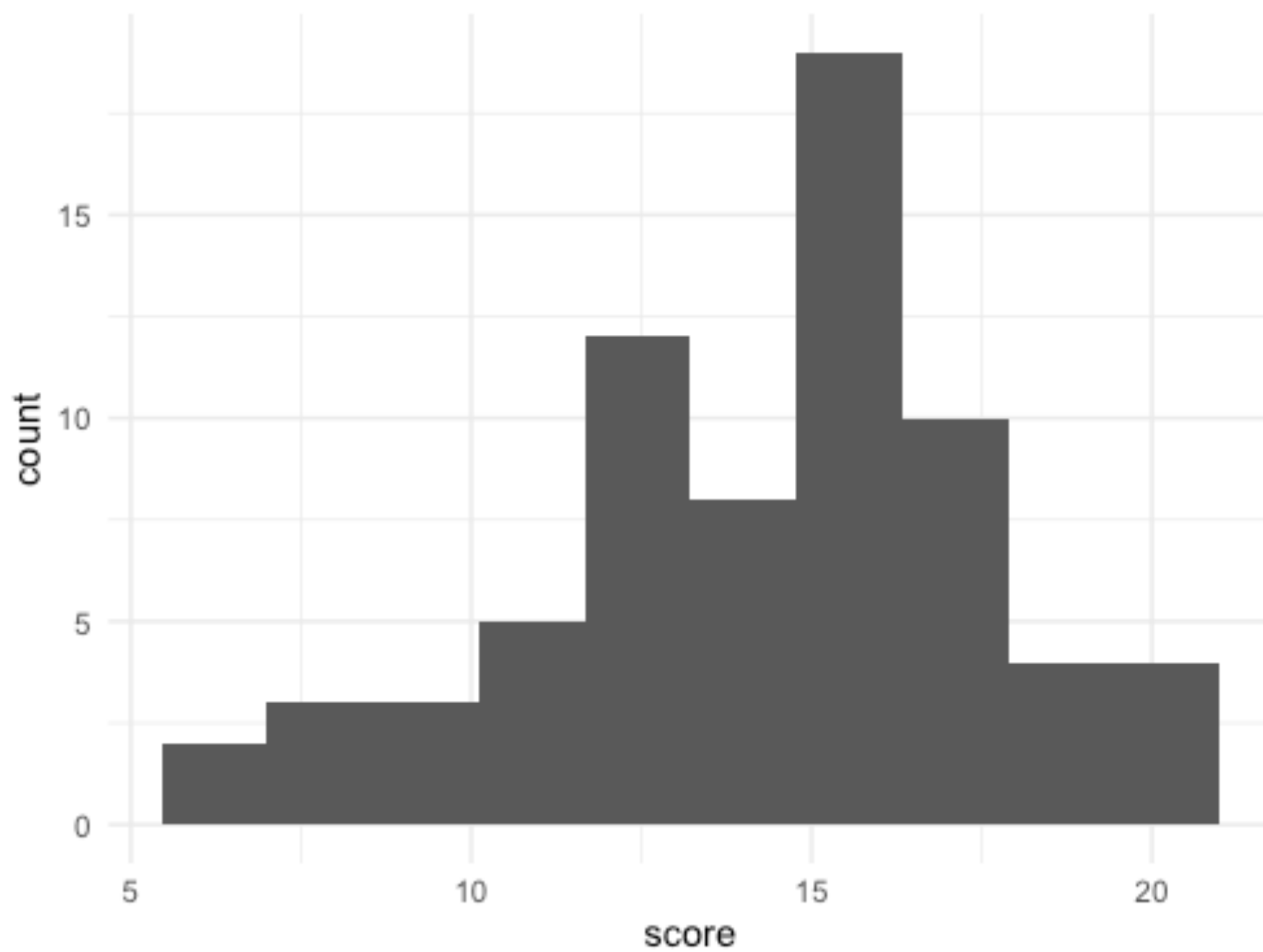


HUDK 4050: CORE METHODS IN EDM

Today

- Troubleshooting is a feature not a bug
- Test distribution
- Dimension reduction
- Work on A2

Check-In Test



Essential Problem

- Dimensionality Reduction
 - Feature selection: select a subset of dimensions
 - Feature extraction: transform lots of dimensions into fewer dimensions
- Why?
 - As a form of insight
 - Avoid “Curse of Dimensionality”

Curse of Dimensionality

Sparsity: The more dimensions that we add, the more comparisons we are missing

	Stats	Cog Psy
Amy	3	2
Chen	2	2
Asif	1	3

Possible Combinations

3 - 3
3 - 2
3 - 1
2 - 3
2 - 2
2 - 1
1 - 3
1 - 2
1 - 1

Curse of Dimensionality

Sparsity: The more dimensions that we add, the more comparisons we are missing

	Stats	Cog Psy	Socio- logy	Crit Theory	Wood- work	Data Sci	Music	Design
Amy	3	2	1	1	3	2	2	2
Chen	2	2	2	3	1	3	2	3
Asif	1	3	3	7	3	2	1	1

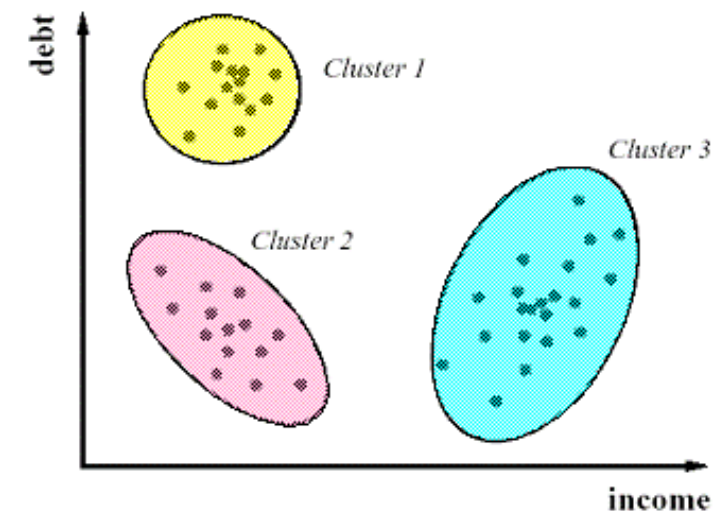
How to reduce dimensions?

Mean, median, mode

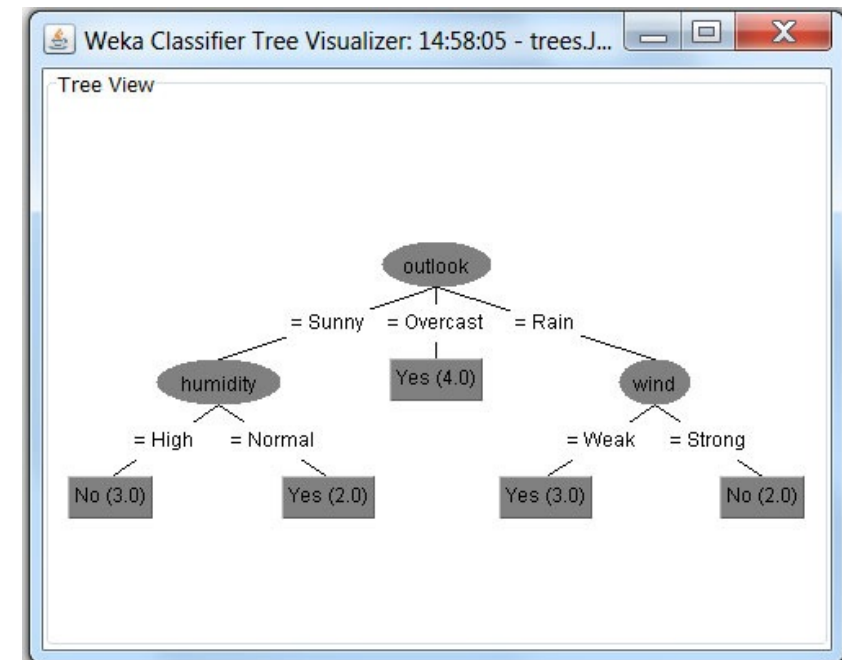
Principle Component
Analysis

Google

Cluster Analysis



Decision Tree



Dimensionality Reduction

Cluster Analysis

Grouping stuff

By Variables

ID	Var1	Var2	Var3
A			
B			
C			
D			

ID	Var2
A	
B	
C	
D	

Selection

ID	Var2+3
A	
B	
C	
D	

Extraction

By People



ID	Var1	Var2	Var3
A			
C			

ID	Var1	Var2	Var3
B			
D			