

Data Madness

Xavier Weber and Jeroen Vermazeren



Topic Introduction

- MovieLens Dataset: Movies , User Ratings
- Research Questions:
 1. How to cluster similar movies according to their genre?
 2. How to determine which group of users would like a movie?

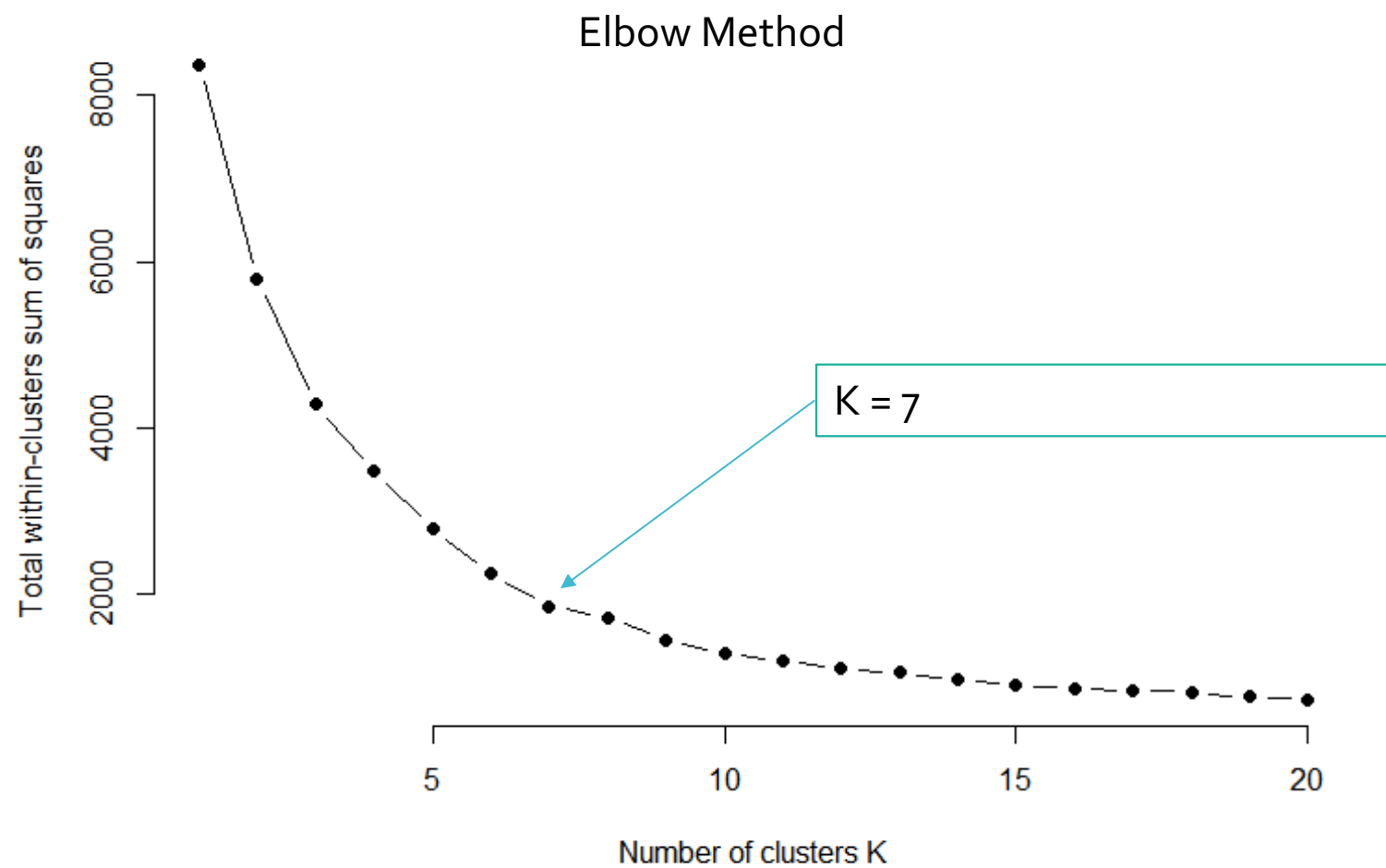
Approach

- Clustering using K-Means
- Construct User Groups using Mean Rating per Cluster

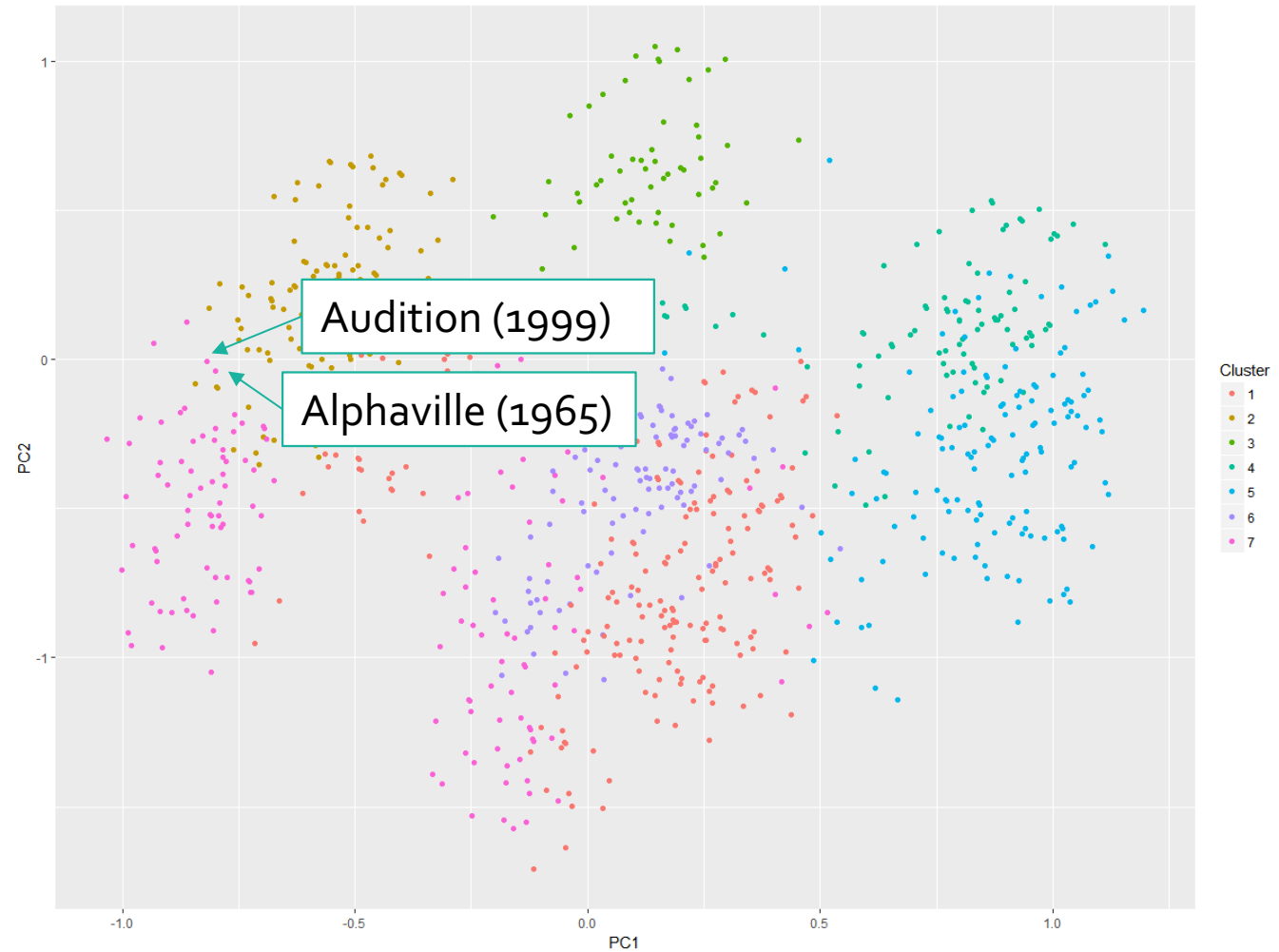
Principal Component Analysis

- Make Data Suitable for K-Means
- Reduce Data Dimensions

Elbow



K-Means



movieId	title	Romance	Drama	Thriller	Horror	Mystery	Sci-Fi
680	Alphaville (1965)	1	1	1	0	1	1
27317	Audition (1999)	1	1	1	1	1	0

User Group

userId	Cluster	Mean Rating
28	7	4.71
46	7	5.00
71	7	5.00
89	7	4.62
91	7	4.60
113	7	4.60
145	7	4.57
147	7	4.60
169	7	4.72
....	7	>4.5

Conclusion

- **Research Questions :**
 1. How to cluster similar movies according to their genre?
 2. How to determine which group of users would like a movie?
- **Practical Use:**
 - Find popular movies for Netflix