



Project 1 Clustering

Angela Ji, Matthew Bloom, Kenneth
Marengo, Abhinav Ratnagiri



Problem Description

- NCAA “clusters” colleges and universities into different conferences. 5 of the 10 FBS conferences are known as the Power 5 football athletic conferences.
- Power 5
 - Atlantic Coast Conference (ACC)
 - Big Ten Conference (BIG)
 - Big 12 Conference
 - Pac-12 Conference
 - Southeastern Conference (SEC)

Data Collection

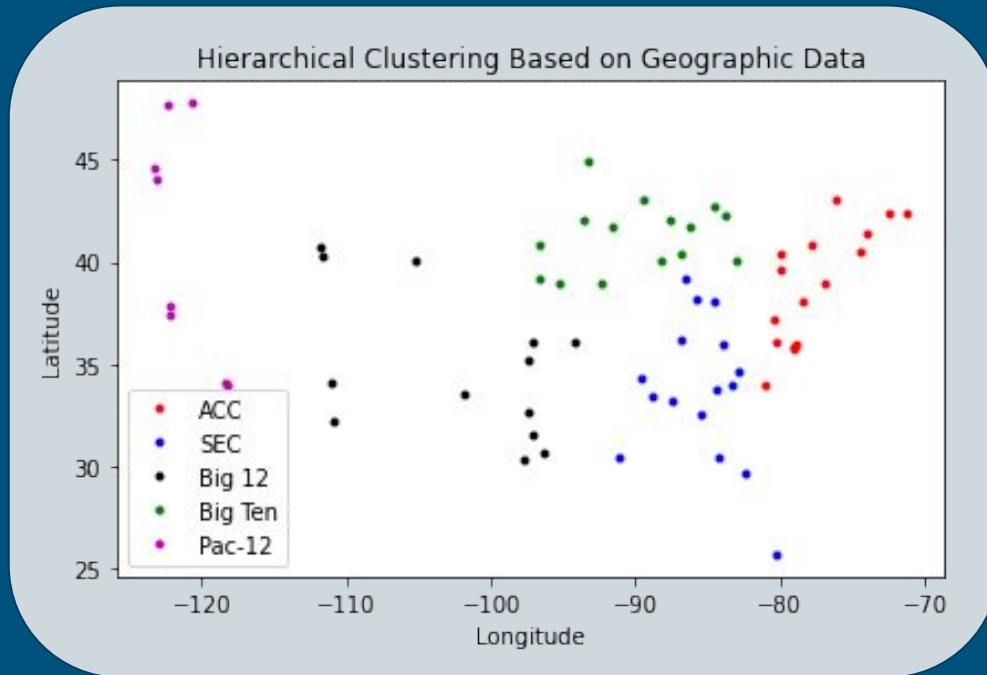
- Name
- Geographic Location
 - Longitude and Latitude
- Conference
- # of NFL Players
- Home City Population (thousands)
- Graduation rate (2011)
- Student Population

Methods of Clustering

- Hierarchical Clustering (complete linkage)
 - Geographic (latitude and longitude)
 - # of NFL Players
 - Student Population
 - Others
- K-means
 - Relatively unsuccessful - algorithmic failure for our data
- Method for categorizing



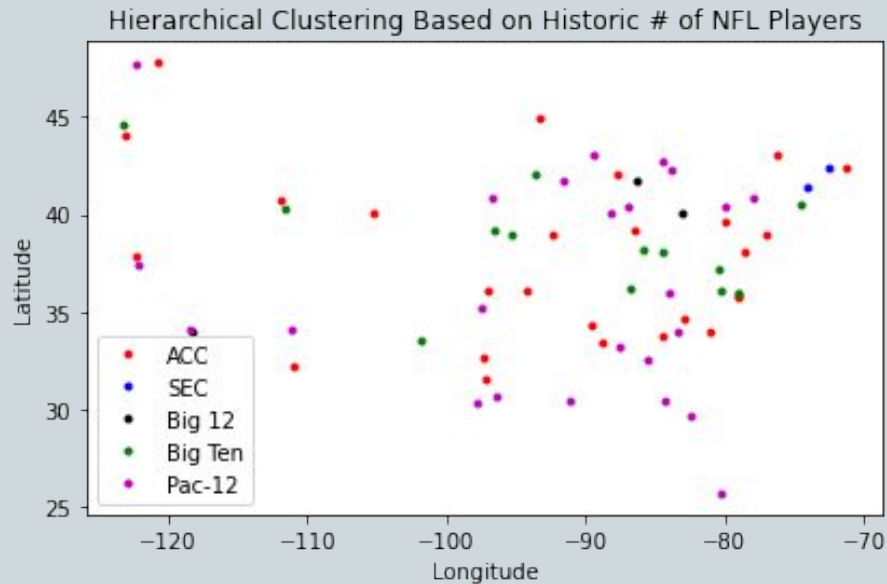
Geographic Data



Part I: Location Data

Location Data Confusion Matrix	ACC predicted	Big 12 predicted	Big Ten predicted	Pac-12 predicted	SEC predicted
ACC actual	9/14= 0.648	0	0	0	5
Big 12 actual	1	6/10= 0.600	3	0	0
Big Ten actual	3	0	10/14= 0.714	0	1
Pac-12 actual	0	4	0	8/12= 0.750	0
SEC actual	1	2	1	0	10/14= 0.714

Historical NFL Player Data



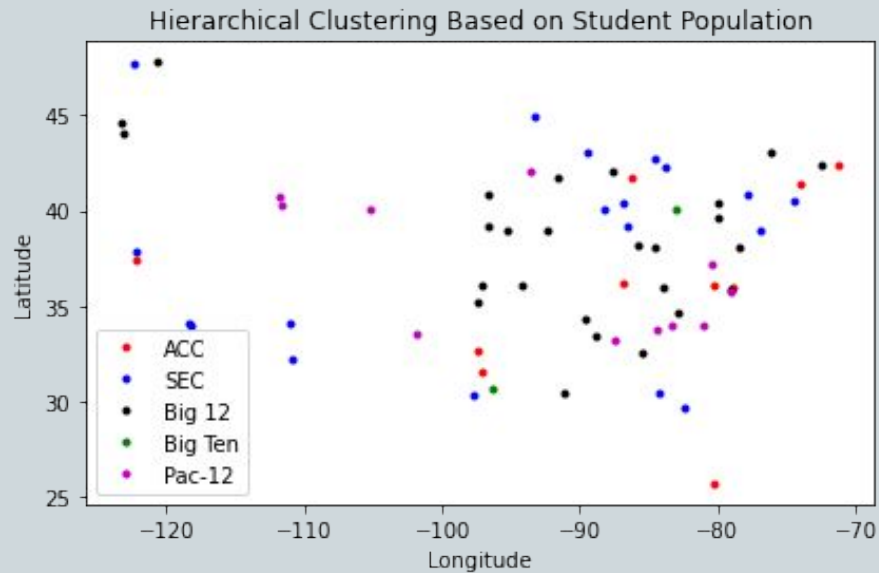
Results of Clustering

- Location Data: $43/64=67.2\%$ success rate
- NFL players: $16/64=25\%$ success rate
- Student Population: $19/64=29.7\%$ success rate

Independent Schools:

- **BYU:** Big 12
- **US Military Academy:** ACC
- **UMass-Amherst:** ACC
- **Notre Dame:** Big Ten

Part II: Student Population Data



New Conferences

ACC	Big 12	Big Ten	Pac-12	SEC
NC State	Boston College	Florida State	Virginia	tOSU
Virginia Tech	Wake Forest	UT Austin	Pitt	Texas A&M
Georgia Tech	Duke	Purdue	Syracuse	
Iowa State	Miami (FL)	Maryland	UNC	
Texas Tech	TCU	Minnesota	Clemson	
BYU	Baylor	UIUC	Louisville	
Utah	Notre Dame	Indiana	Kansas	
UCo Boulder	Army	Penn State	Oklahoma	
South Carolina	Stanford	Michigan	West Virginia	
Georgia	Vanderbilt	Michigan State	Oklahoma State	
Alabama		

Conclusions & Discussion

- Location clearly strongest correlation
 - Adding variables was worse - alternate variables? weights?
- Student population could make some conferences more competitive
 - Downside: travel
 - Inequality of attendance
- Consolidating conferences?
 - Pac-12 and BIG don't fit on a slide!