Project 2: Principal Components Analysis and Dimension Reduction

Parker Herrin

MSDS 6372-401

3/21/2016

World Series Champions (1965 – 2015)

American League WS Champs		National League WS Champs	
2015 KC Royals	2013 Boston Red Sox	2014 SF Giants	2012 SF Giants
2009 NY Yankees	2007 Boston Red Sox	2011 STL Cardinals	2010 SF Giants
2005 CHI White Sox	2004 Boston Red Sox	2008 PHI Phillies	2006 Cardinals
2002 Anaheim Angels	2000 NY Yankees	2003 Florida Marlins	2001 ARI Diamondbacks
1999 NY Yankees	1998 NY Yankees	1997 Florida Marlins	1995 ATL Braves
1996 NY Yankees	1993 TOR Blue Jays	1990 Cincinnati Reds	1988 LA Dodgers
1992 TOR Blue Jays	1991 MIN Twins	1986 NY Mets	1982 STL Cardinals
1989 OAK Athletics	1987 MIN Twins	1981 LA Dodgers	1981 PHI Phillies
1985 KC Royals	1984 Detroit Tigers	1979 Pittsburgh Pirates	1976 Cincinnati Reds
1983 Baltimore Orioles	1978 NY Yankees	1975 Cincinnati Reds	1971 Pittsburgh Pirates
1977 Yankees	1974 OAK Athletics	1969 NY Mets	1967 STL Cardinals
1973 OAK Athletics	1972 OAK Athletics	1965 LA Dodgers	
1970 Baltimore Orioles	1968 Detroit Tigers		
1966 Baltimore Orioles			

Selected Variables for Analysis

Batting	Pitching	Fielding	Team
Total Runs	WHIP	Fielder's Age	Run Differential
RBI	FIP	Range Factor / G	September Win PCT
BB	ERA	Completed Games	
OBP	ERA Plus		
OPS Plus	BB / 9		
	Pitcher's Age		
	H / 9		

Scree Plot by PC Variance

Variances

Cumulative Proportion:
0.785
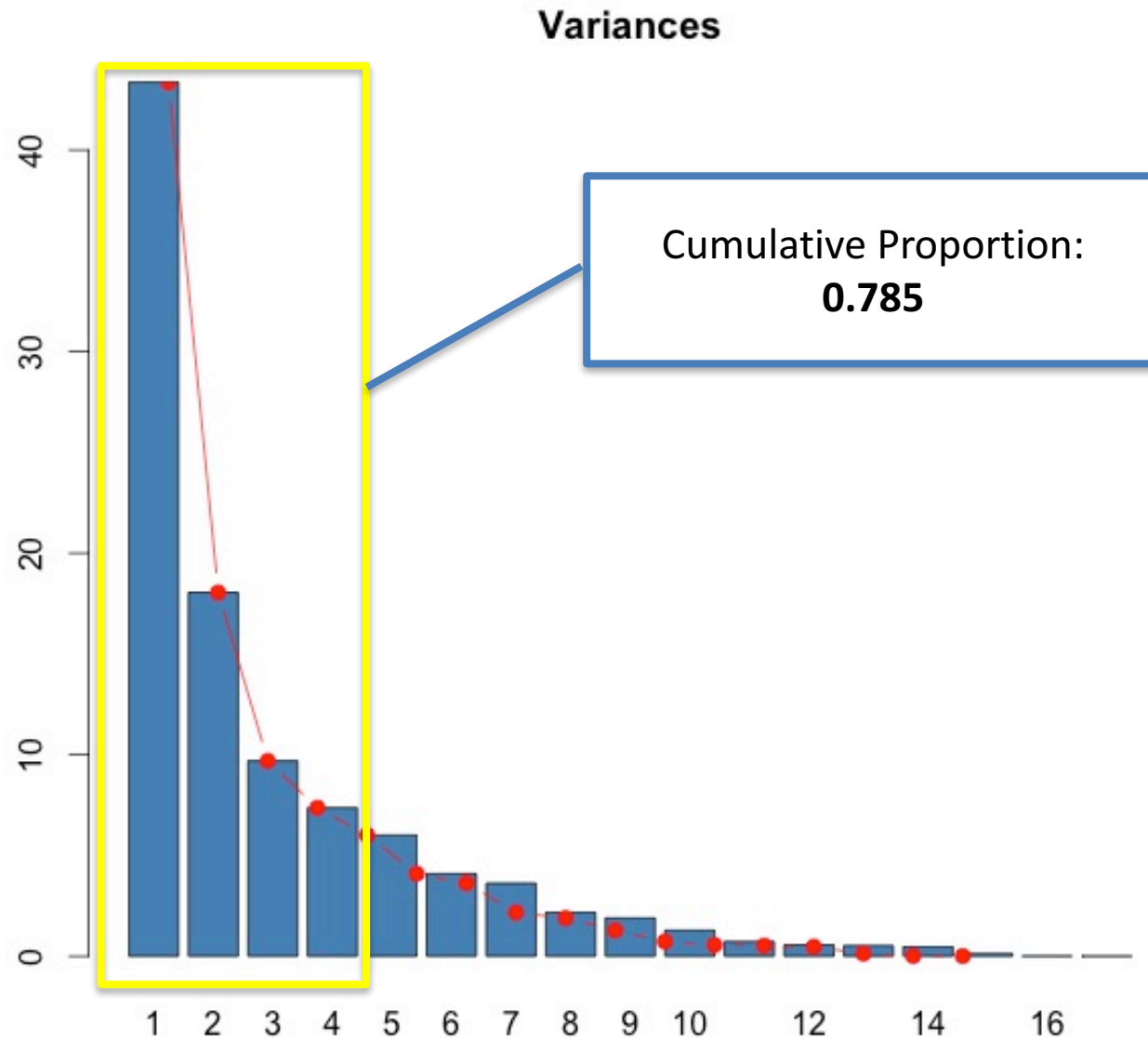
Selected Components: Summary

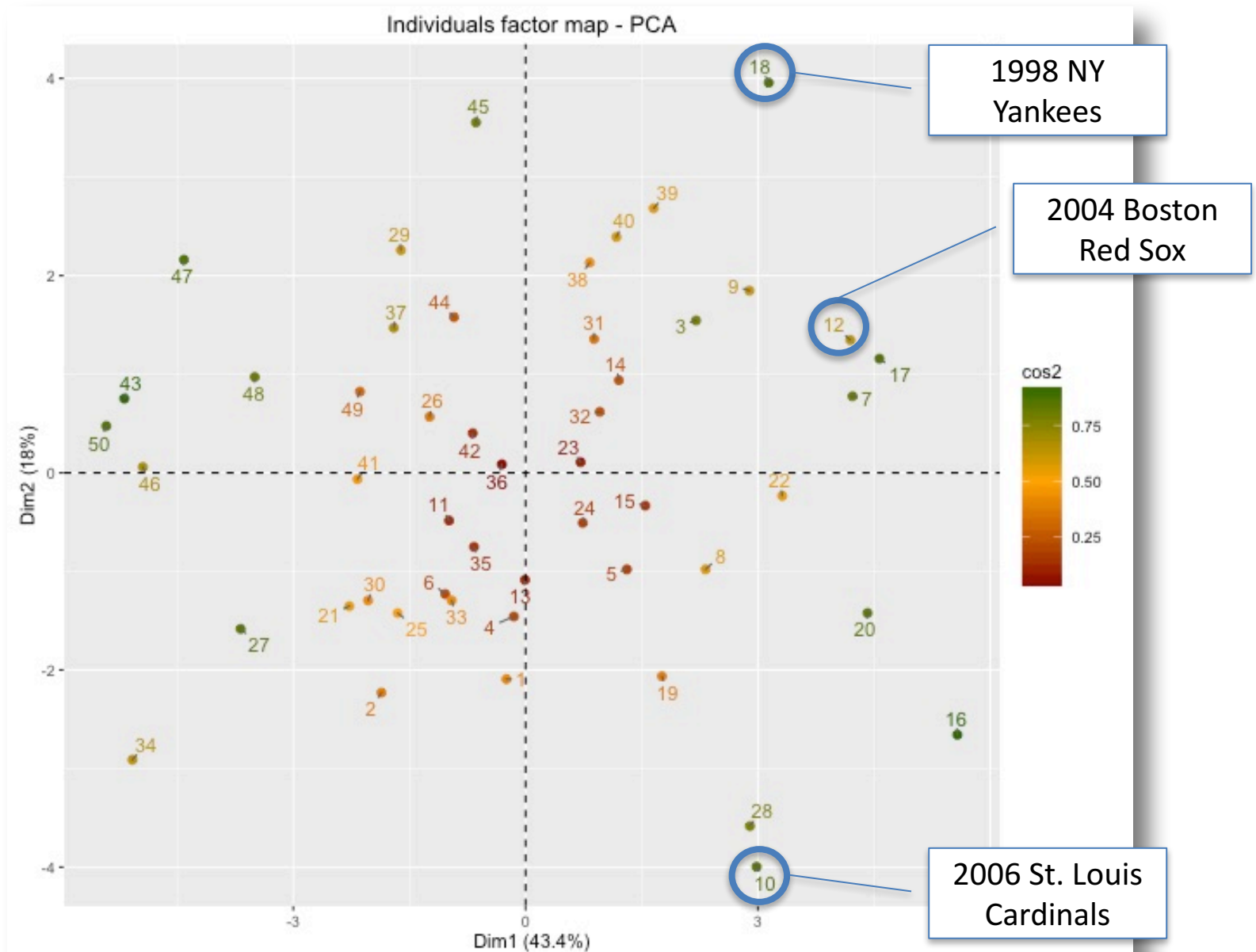| Component #1	Component #2	Component #3	Component #4
Runs (Batting)	OPS + (Batting)	P.Age (Pitching)	OPS + (Batting)
RBIs (Batting)	Sep. WinPCT (Team)	ERA + (Pitching)	ERA+ (Pitching)
OBP (Batting)	Run Differential (Team)	BB9 (Pitching)	CG (Fielding)
ERA (Pitching)		F.Age (Fielding)	
FIP (Pitching)		CG (Fielding)	
WHIP (Pitching)		RF / G (Fielding)	

Variable Factor Map: Dimension 1,2
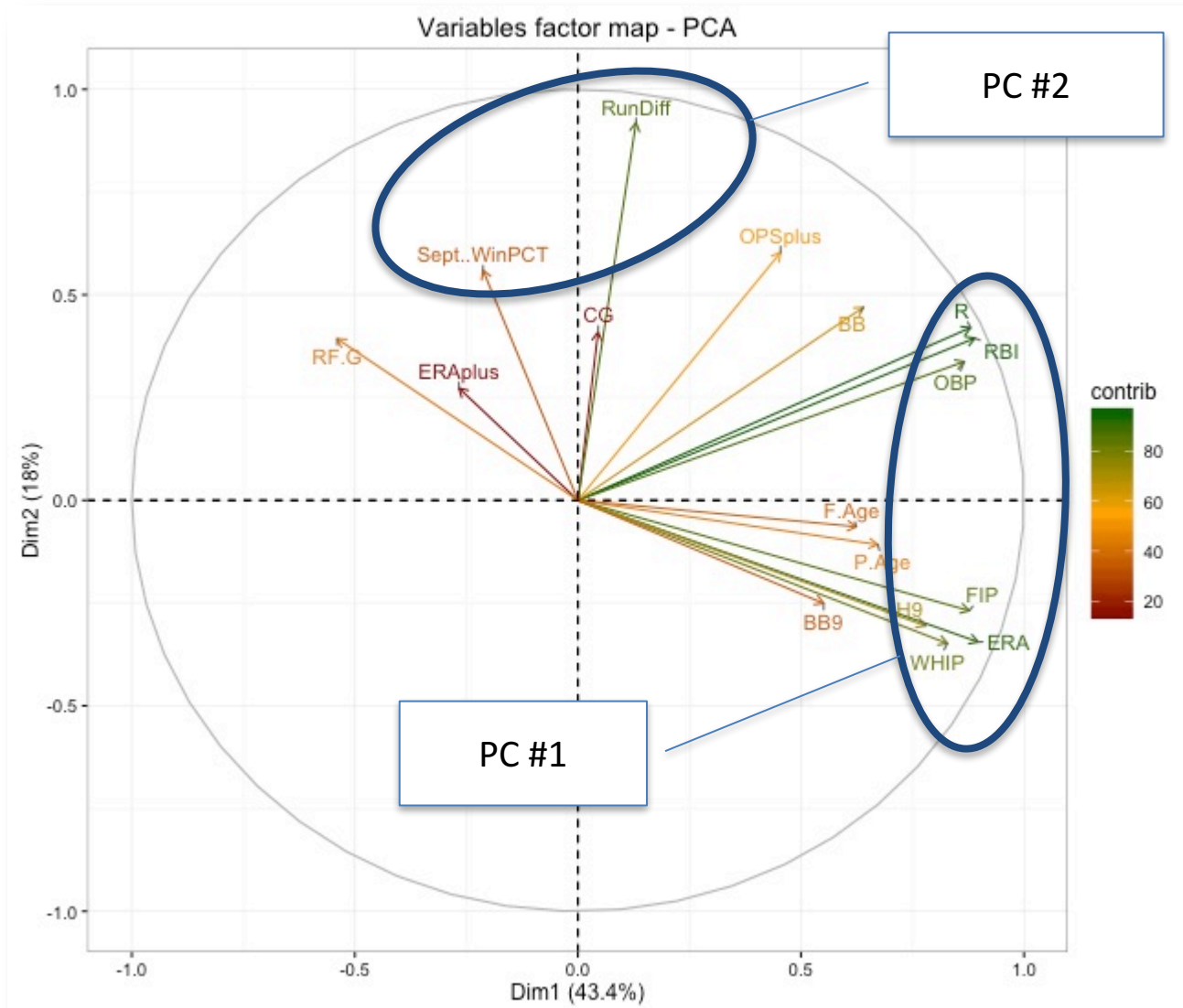
Variable Factor Map: Dimension 3,4
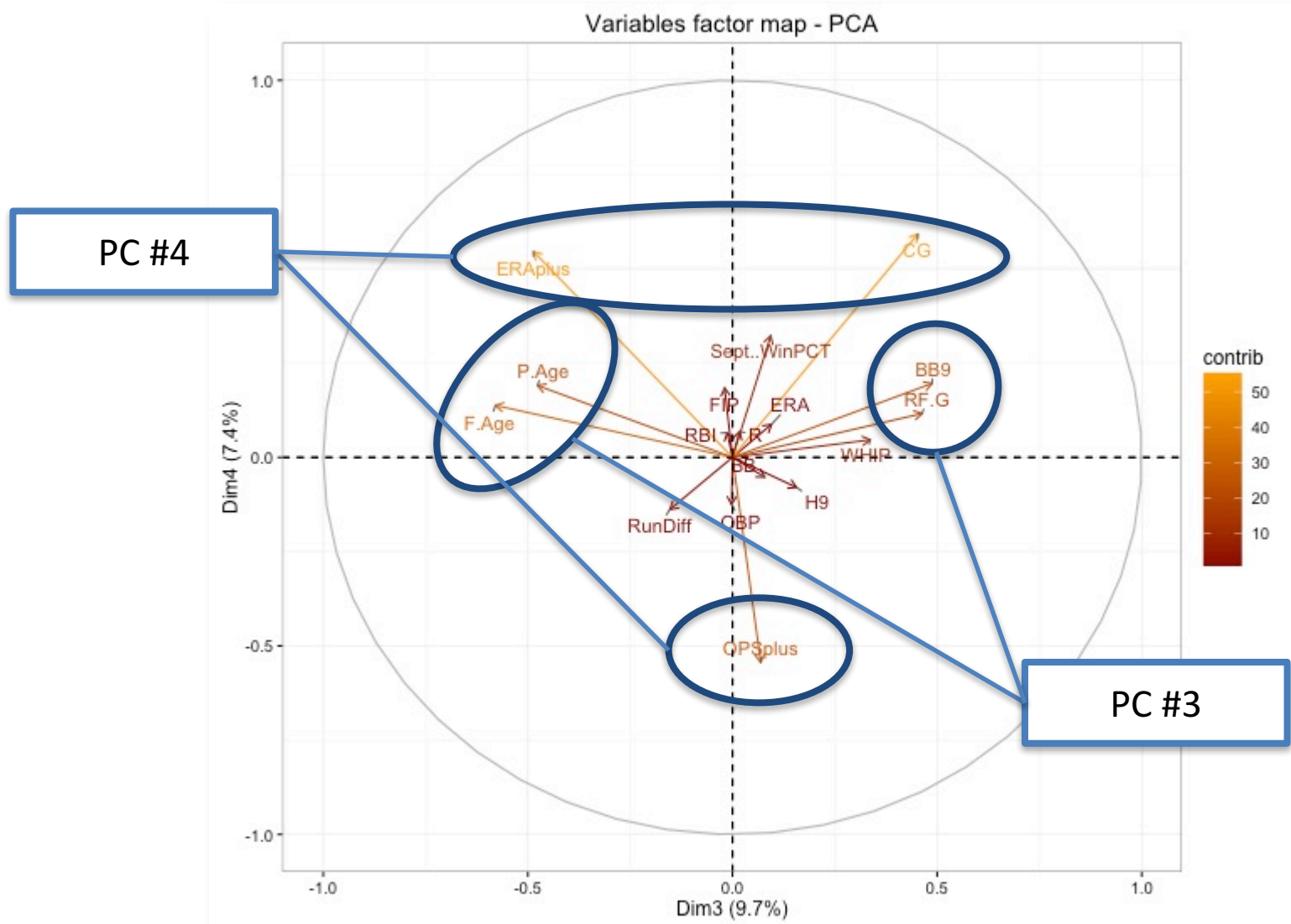
Individual Factor Map: Dimension 1,2

Regression Model: PCA

Component	Estimate	Standard Error	t Value	Pr (> |t|)
(Intercept)	95.440	0.5619	169.864	< .0001
Off/Def	-0.3223	0.2069	-1.557	0.12639
Team	-3.5987	0.3208	-11.219	< .0001
Age/Def	-0.5856	0.4377	-1.338	0.18765
OPS+/ERA+/CG	1.5308	0.5020	3.050	0.00383

Regression Model: All Variables

Parameter	Estimate	p-Value
(Intercept)	146.960384	< .0001***
Runs	0.100925	0.10302
RBI	0.015332	0.82076
BB	0.005025	0.61381
OBP	-7.616645	0.93600
OPSplus	-0.182203	0.18877
P_Age	-0.210268	0.63271
ERA	-14.942152	0.00273**
ERAplus	-0.023090	0.84149

Parameter	Estimate	p-Value
FIP	3.224192	0.25170
WHIP	193.771060	0.05103
H9	-23.955958	0.03158*
BB9	-22.982588	0.03999*
F_Age	-2.022688	0.00479**
CG	0.008108	0.34113
RF_G	-0.658873	0.86045
SeptWinPCT	14.053706	0.02429*
RunDiff	-0.022374	0.32540