MA3505 Multivariate Statistics Project 1

April 26, 2016

1 Introduction and exploratory data analysis for the variables.

2 Analysis to answer each research question

2.1 Question 1

2.2 Question 2

2.3 Question 3

2.3.1 Cleveland

From running variance inflation factor we get the following

age	sex	cp	trestbps	htn	chol	cigs	years
2.070591	2.379469	1.683710	2.935706	1.734144	1.326342	2.346224	2.315459
fbs	famhist	restecg	ekgmo	ekgday	ekgyr	dig	prop
1.281244	1.291443	1.338021	14.903816	3.357399	78.992867	1.296383	1.679766
nitr	pro	diuretic	thaldur	thaltime	met	thalach	thalrest
1.546570	1.415979	1.480903	9.549788	1.422540	10.328475	2.868773	1.713892
tpeakbps	tpeakbpd	trestbpd	exang	xhypo	oldpeak	slope	rldv5e
2.829387	2.173463	2.785971	1.734917	1.870852	2.831028	2.291928	1.557587
ca	thal	cmo	cday	cyr	lmt	ladprox	laddist
1.841289	2.051953	15.389866	3.413846	80.511913	1.401270	1.496650	1.526869
cxmain	om1	rcaprox	rcadist				
1.543251	1.789705	1.764053	1.835745				

Here we see the variables, ekgmo, ekgyr, cmo and cyr are collinear with other variables in the model.

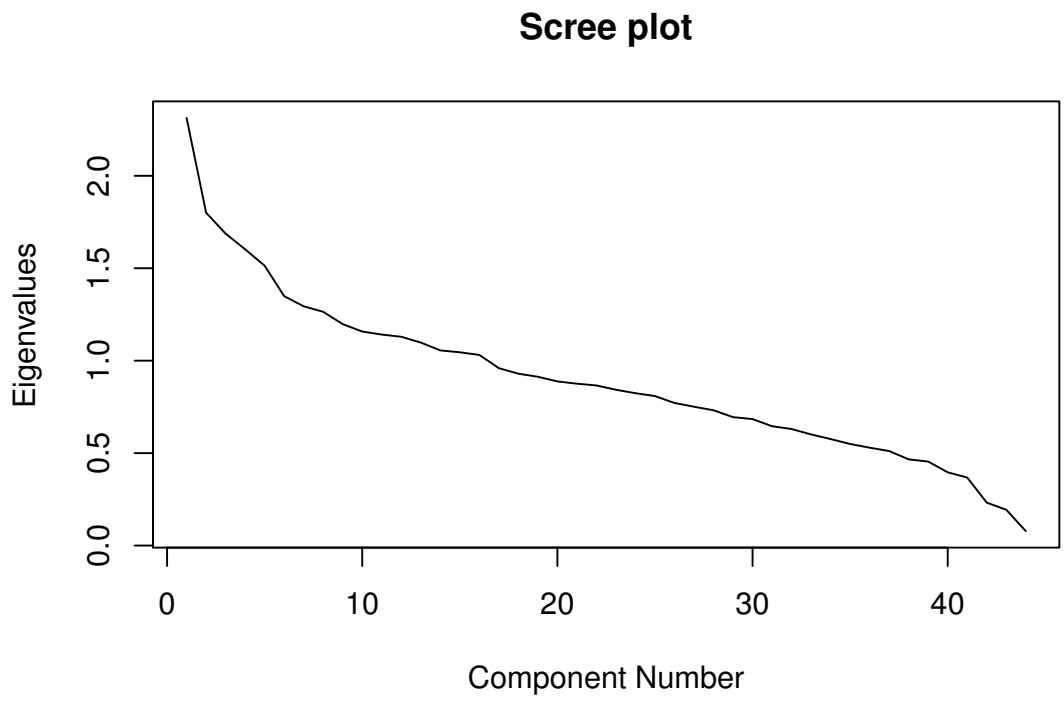

Figure 1: Screeplot for PCA of Cleveland

2.3.2 Hungary

2.3.3 Longbeach

2.3.4 Switzerland

2.4 Question 4

3 Summary