MA3505 Multivariate Statistics Project 1

April 27, 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.

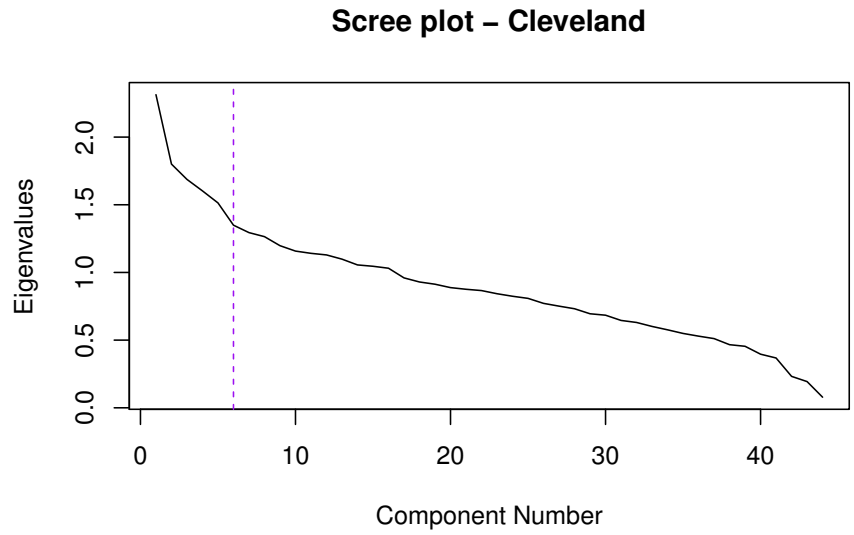

Figure 1: Scree plot for PCA of Cleveland

From the scree plot in Figure 1 we see that we keep 6 components.

We have the loadings of each components as follows.

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
age	0.192		−0.196			0.167		−0.122	0.163	0.372
sex		−0.195	0.306	0.193					−0.303	
cp	0.208							0.384		−0.116
trestbps	0.133	−0.144	−0.297	0.222	0.107	0.119		−0.149		
htn			0.222		−0.189	0.390		0.117		
chol			−0.184					0.184	0.213	0.222
cigs		−0.200	0.181	0.231	−0.292		−0.128		−0.214	
years		−0.189	0.145	0.223	−0.330	0.138			−0.156	
fbs			−0.128	0.143		−0.214	0.132			0.129
famhist					0.123	0.140		0.133	0.162	−0.136
restecg		−0.103	−0.132				−0.128	0.238		
ekgmo		−0.244		−0.433	−0.109		0.220	−0.161	−0.144	
ekgday					0.384	0.326	0.255	0.298	−0.109	
ekgyr		0.414		0.193			0.268		−0.129	−0.212
dig		0.105				0.195	−0.112	−0.150	−0.230	
prop	0.102	0.107			0.162	−0.263	−0.247	0.105		0.173
nitr	0.142	0.107		−0.128			−0.141		−0.180	0.131
pro		0.236				−0.115		0.154	−0.222	
diuretic					0.128	−0.417			−0.136	0.199
thaldur	−0.301	−0.109	0.237	0.125	0.184	−0.149				
thaltime			0.154		0.153			−0.191	0.359	−0.189
met	−0.295	−0.137	0.228	0.135	0.181	−0.167				
thalach	−0.298	−0.172					0.130	0.101	0.106	−0.141
thalrest			−0.229		−0.254		0.221	0.145	0.194	−0.113
tpeakbps		−0.211	−0.236	0.297				−0.211		
tpeakbpd		−0.167	−0.330	0.142			−0.128	0.161	−0.148	−0.159
trestbpd		−0.222	−0.314	0.112	0.130		−0.106		−0.167	−0.194
exang	0.224			−0.100					−0.207	−0.157
xhypo	0.104	0.153		−0.229		−0.102			−0.113	
oldpeak	0.280				0.185			−0.219	0.150	−0.159
slope	0.232				0.230		−0.120	−0.217		−0.263
rldv5e				0.126	0.127	0.166		−0.295		
ca	0.213			0.113	−0.124		0.283		0.211	0.251

thal	0.231	−0.163	0.167	0.102					−0.143	
cmo		−0.243		−0.433	−0.116		0.209	−0.162	−0.122	
cday					0.391	0.294	0.280	0.243	−0.135	
cyr		0.415		0.195			0.261		−0.130	−0.218
lmt	0.130	−0.106					0.132			−0.126
ladprox	0.183		0.147				−0.234	0.150	0.132	
laddist	0.206		0.107			−0.114	0.254			
cxmain	0.189		0.150	0.104					0.111	0.201
om1	0.249					−0.108	0.202			
rcaprox	0.191					−0.237	0.162	0.151	0.181	−0.291
rcadist	0.196		0.103				0.183	−0.130		0.250

2.3.2 Hungary

Figure 2: Scree plot for PCA of Hungary

From the scree plot in Figure 2 we see that we keep 6 components.

We have the loadings of each components as follows.

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
age	0.162		0.184		−0.253		−0.105	−0.185	−0.114	0.137
sex		−0.144	−0.164	−0.236					0.232	
painloc	0.143		−0.261	−0.112	0.116					0.214
painexer	0.212		−0.335			−0.146				
relrest	0.228		−0.332					−0.159		−0.108
cp	0.229		−0.357			−0.163		−0.115		
trestbps	0.213		0.179	−0.290	0.113			−0.214		0.198
htn		−0.160	0.101			0.247		0.125	0.230	0.214
chol				−0.139	−0.207	−0.105		0.186	−0.529	
fbs		−0.170		−0.127	−0.193	0.168		0.265	−0.304	−0.195
restecg									−0.135	0.304
ekgmo	−0.191		−0.177		−0.276		0.344	−0.380		
ekgday			−0.108		−0.254	0.479	−0.110		0.255	
ekgyr	0.126	−0.312	0.189			−0.181	0.237		0.101	−0.326
prop	0.132	−0.253		0.301	0.117		0.112			0.194
nitr		−0.286		0.402						0.163
pro		−0.309		0.355					−0.105	0.119

diuretic			0.129	−0.101	0.149		0.275	0.123	0.377	0.281
proto	−0.312	−0.277	−0.136	−0.121						
thaldur	−0.305	−0.277	−0.135	−0.130						
thaltime	−0.303	−0.270	−0.138	−0.128			−0.111			0.113
met	−0.306	−0.227					−0.192			
thalach	−0.259			−0.135	0.126		0.376	0.177	−0.142	
thalrest				−0.126			0.539	0.233	−0.141	0.269
tpeakbps		−0.225		−0.292		0.170		−0.264		
tpeakbpd		−0.191	0.271	−0.231		0.134		−0.199		
trestbpd	0.157		0.155	−0.301	0.137			−0.207	−0.150	0.256
exang	0.237		−0.216	−0.140						−0.175
oldpeak		0.113		−0.168	0.217	0.146			0.259	−0.200
slope	0.156		−0.245			0.280		0.287		
rldv5	−0.147	0.126			0.439	0.283		−0.130	−0.232	−0.109
rldv5e	−0.128			0.116	0.444	0.289		−0.179	−0.172	−0.129
cmo	−0.175		−0.194		−0.284	0.103	0.297	−0.395		
cday		−0.123			−0.229	0.387	0.133	0.142		−0.211
cyr	0.130	−0.316	0.179			−0.181	0.222			−0.331

2.3.3 Longbeach

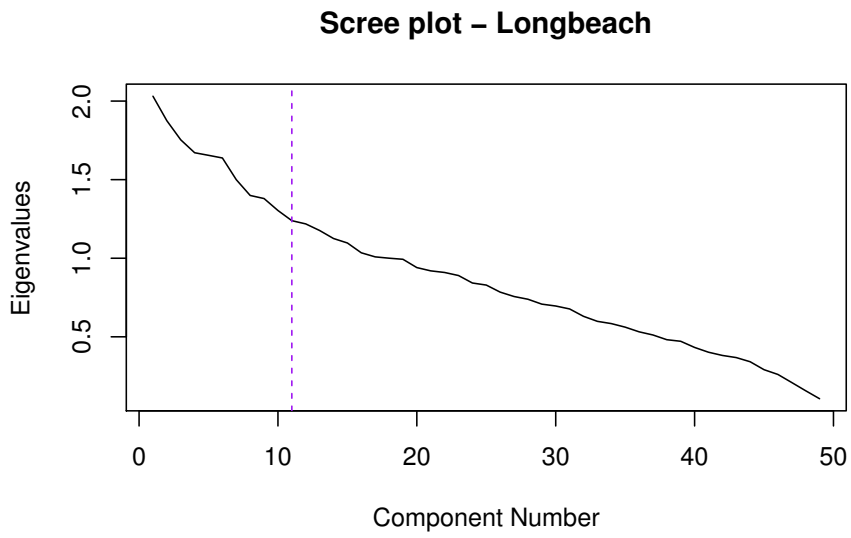

Figure 3: Scree plot for PCA of Longbeach

From the scree plot in Figure 3 we see that we keep 11 components.

We have the loadings of each components as follows.

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
age	−0.204		0.197			−0.182		−0.205		
sex			−0.106	−0.215			0.147			0.146
painloc		−0.133	0.209			0.255		0.136	−0.105	
painexer	−0.192	−0.223			0.288	0.188	−0.162			
relrest	−0.181	−0.203			0.114	0.314		0.197		
cp	−0.181	−0.276			0.216	0.288	−0.128	0.184		
trestbps	−0.196		0.310	−0.133		−0.202				
htn			0.313	−0.134	−0.219					0.106
chol			0.118	0.166	0.175	−0.156	−0.232			−0.194
smoke	0.160		−0.191	−0.334						−0.186

cigs			−0.250	−0.320						0.240
years	0.155	−0.117	−0.133	−0.344					0.103	
fbs		0.111	0.196			−0.168	−0.154			0.315
famhist				−0.124	−0.235			0.159		−0.316
restecg	0.125		0.132					−0.128	0.257	0.229
ekgmo		−0.189	−0.134				0.178		−0.289	
ekgday			−0.149				0.357		0.264	0.166
ekgyr	−0.357	0.130	−0.161	−0.127			−0.195			
dig	0.166				−0.111	−0.241		0.111	0.227	−0.183
prop			0.137	−0.106		0.115	0.179	−0.179	0.135	−0.116
nitr			0.123		−0.215	0.181		−0.255	−0.139	
pro				−0.177	−0.253			0.170	−0.112	
diuretic		−0.140	0.162	−0.221			0.106		0.187	−0.109
proto	−0.288	0.135	−0.240		−0.102			−0.171		
thaldur		0.402					0.102	0.153	−0.190	−0.113
met		0.353			0.126		0.131	0.237	−0.172	
thalach		0.151		−0.129	0.323		0.279	0.172		
thalrest					0.349		0.169		0.119	
tpeakbps		0.264	0.229	−0.167			0.154			0.158
tpeakbpd	0.139		0.223		0.124	0.141	0.235	0.144		
trestbpd			0.200	−0.181		−0.154		0.135		
exang	−0.105	−0.260				−0.141	0.135		−0.110	−0.174
xhypo	−0.128				0.153	−0.177			−0.148	
oldpeak	−0.250				0.204	−0.208	0.116			−0.274
rldv5	−0.238			0.274	−0.118		0.280		0.249	−0.137
rldv5e	−0.238			0.262	−0.191		0.255		0.224	−0.153
cmo		−0.244	−0.131	0.117		−0.124	0.210		−0.291	
cday					0.135	0.129	0.160	−0.310		0.299
cyr	−0.352	0.156	−0.160	−0.134			−0.185			
lmt	−0.104						−0.149	0.166		
ladprox		−0.103		−0.127	0.229	−0.220				0.140
laddist						0.123	0.102		−0.233	
diag	−0.105				−0.109			0.384		0.174
cxmain		−0.126				−0.233			−0.118	
ramus		0.103	0.113	−0.139	0.140		−0.114	−0.232		−0.116
om1		−0.105	0.100	−0.144					−0.347	−0.161
om2			0.249	−0.217	0.100	0.133		−0.310		−0.142
rcaprox		−0.139		−0.118		−0.347		−0.113		
rcadist	−0.196				−0.192		0.131		−0.177	0.172

	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15	Comp.16	Comp.17	Comp.18	Comp.19
age	−0.163		−0.136	0.110		0.110		0.185	−0.230
sex		−0.264		0.125			−0.206	−0.150	−0.363
painloc				0.210	0.397			−0.138	−0.199
painexer						−0.246			
relrest	−0.147						0.229	0.114	
cp						−0.138			
trestbps	0.112		−0.129	−0.101					
htn	−0.123		0.241	−0.196	0.100	0.102			
chol		0.180	−0.233				0.113	−0.105	−0.178
smoke	−0.107						0.203		
cigs			−0.160		0.131			−0.118	0.198
years			−0.264				0.156	0.211	
fbs	−0.182	0.233		0.139	0.176			0.135	0.250
famhist		0.269		0.199	0.119		−0.113		0.209
restecg	−0.169					−0.216	−0.139		−0.238
ekgmo	−0.384	0.110	−0.184	−0.205	−0.258				
ekgday	0.168	0.147		0.172	0.161				
ekgyr		0.133				0.168	−0.111		
dig			0.121	−0.146		−0.255	−0.247		−0.129
prop		0.242	−0.265	−0.255	−0.141	−0.265		−0.185	−0.188

nitr				0.322	−0.116			−0.135	
pro		0.168	0.127	0.208	−0.186	−0.188	−0.190		
diuretic	−0.194		0.170	−0.138	0.170	0.322	0.248	0.107	−0.126
proto							−0.156		
thaldur	−0.152				0.114	−0.138	0.111	−0.183	−0.152
met	−0.139				0.171	−0.230	0.215	−0.143	−0.166
thalach				0.135		0.102		0.290	0.111
thalrest		0.183			−0.220	0.151	−0.344		−0.102
tpeakbps	−0.110	−0.145		0.166	−0.242			0.162	−0.112
tpeakbpd	0.102	−0.105		0.146	−0.169	0.353	−0.157		
trestbpd	0.371	0.130	−0.147	−0.210	−0.110		−0.134	−0.124	0.114
exang	0.206	−0.116	−0.215	0.208	0.156	−0.106			
xhypo		0.349		−0.156	0.253	0.155		0.123	
oldpeak		−0.159	0.101		0.100			0.143	0.175
rldv5		−0.105							0.119
rldv5e					−0.108		0.104		
cmo	−0.337		−0.142				−0.116		
cday	0.209	0.127			0.122		0.258	−0.119	
cyr		0.146				0.164			
lmt		−0.226	−0.455	−0.245		0.126			
ladprox		−0.137	0.224				0.117	−0.310	0.236
laddist	0.331	−0.207		−0.191	0.182	−0.167	−0.168	0.349	−0.174
diag		−0.262						−0.327	0.139
cxmain	0.104	0.130		0.137	−0.356	−0.184	0.350	0.191	−0.105
ramus	−0.196	−0.260		0.165	0.104	−0.133	−0.169	0.115	0.279
om1			0.313	−0.280		0.122		−0.129	
om2						−0.257	−0.104		0.209
rcaprox				0.184	0.156			−0.297	
rcadist		0.124	0.214				−0.106		

2.3.4 Switzerland

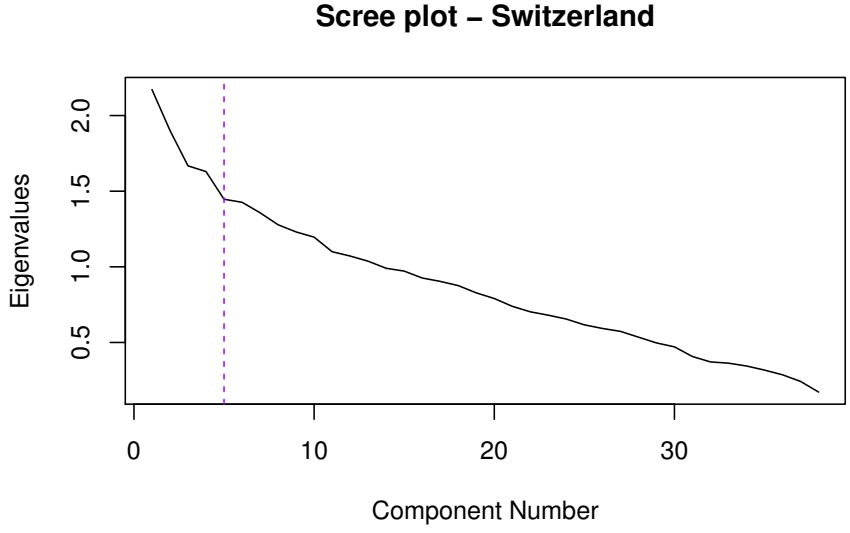

Figure 4: Scree plot for PCA of Switzerland

From the scree plot in Figure 4 we see that we keep 5 components.

We have the loadings of each components as follows.

```
Loadings:
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
age			0.366		0.115		−0.106	0.158	0.220	−0.138
sex		−0.107			−0.238			−0.326	0.118	
painloc	−0.209	−0.265	−0.145	−0.197						0.160
painexer	−0.245	−0.238	−0.183	−0.211			0.160			
relrest	−0.215	−0.193	−0.187	−0.264			0.147	0.144		
cp	−0.214	−0.292	−0.203	−0.250						
trestbps	−0.154	0.126	0.369	−0.192		0.114		0.169	0.116	
restecg		0.110	0.170		−0.182			0.101	0.202	−0.451
ekgmo	−0.337	0.199			−0.120	−0.168				
ekgday		−0.143	−0.219		−0.172	0.420	−0.178	0.222		
ekgyr	0.316	−0.219	0.142	−0.179	0.137		0.171			
dig				0.140		0.167	0.281	0.187		0.342
prop	0.111	−0.227		0.129		−0.201	−0.174		0.204	0.143
nitr	0.137	−0.214				−0.307		0.181	0.332	
pro		−0.111	0.106		−0.198	−0.331	−0.201	0.325		0.104
diuretic	−0.105	−0.119	0.133	0.113	−0.114	−0.201	−0.162	0.173	−0.162	
thaldur	0.231			−0.107	−0.355	−0.161	0.192	−0.166		0.214
thalach	0.184	0.172	−0.191	−0.187	−0.242		0.247	0.232	−0.236	
thalrest		0.269	−0.156				0.222	0.411	−0.151	−0.160
tpeakbps		0.215	0.117	−0.372	−0.176			0.197		
tpeakbpd	−0.116	0.125	0.148	−0.352		0.100			0.189	
trestbpd	−0.206	0.106	0.199	−0.179		0.214	−0.123		0.163	0.152
exang	−0.188					0.153	−0.315		−0.251	−0.117
xhypo			0.114	0.170	0.226	0.229	0.290	0.269		0.182
oldpeak		−0.110		−0.298			0.151		−0.205	−0.187
cmo	−0.345	0.185			−0.128	−0.128		−0.109		
cday		−0.129	−0.189		−0.188	0.381	−0.231	0.228		
cyr	0.275	−0.194	0.136	−0.154	0.202		0.144	−0.138		
lmt		−0.103	−0.142	0.123	0.119	−0.116	0.135		0.166	−0.333
ladprox	−0.150	−0.113		0.186	0.132			0.122		−0.195
laddist		−0.150	0.117		−0.348					0.198
diag		−0.187	0.196		−0.179				−0.323	
cxmain	−0.102	−0.128			−0.190	0.219		−0.208		−0.366
ramus	−0.139	−0.107	0.237	0.129			0.108		−0.311	
om1		−0.186	0.154	0.153	−0.247		0.315		0.124	−0.149
om2		−0.161	0.212						−0.338	
rcaprox	−0.189		0.119		0.159		0.133		0.116	0.156
rcadist	−0.106			0.137	−0.255		0.277		0.220	

2.4 Question 4

3 Summary