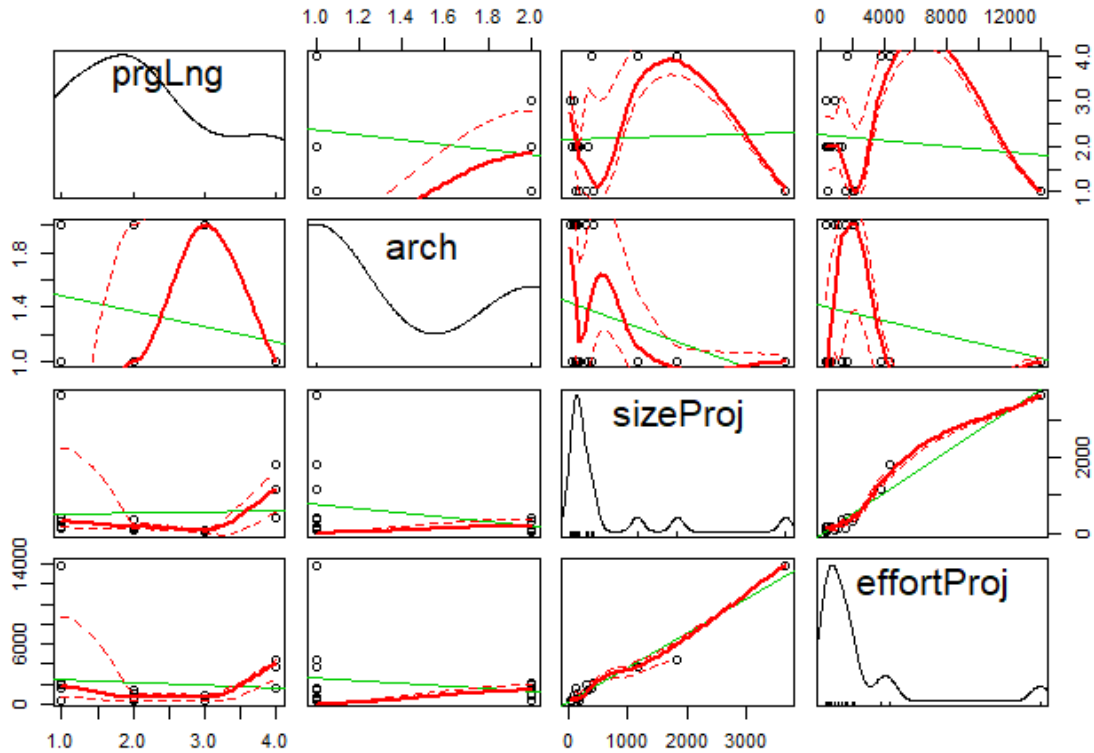


1. LDA Analysis



Scatterplot matrix - red dash lines represent confidence interval.

Results Interpretation

> dataset.lda

Call:

lda(prgLng ~ ., data = data[2:5])

Prior probabilities of groups:

	1	2	3	4
0.2941176	0.4117647	0.1176471	0.1764706	

Group means:

	arch	sizeProj	effortProj
1	1.600000	934.2000	3967.4000
2	1.142857	169.5714	664.7143
3	2.000000	63.5000	641.5000
4	1.000000	1128.6667	3271.0000

Coefficients of linear discriminants:

	LD1	LD2	LD3
arch	2.1254814374	1.544891665	-1.3674561996
sizeProj	-0.0020539101	0.005519467	-0.0027943210
effortProj	0.0007244299	-0.001303567	0.0009804635

Proportion of trace:

	LD1	LD2	LD3
0.6525	0.2394	0.1082	

Results Explanation:

In order to explain 80% of information of dataset, we have to consider classes LD1 and LD2 because these two together make 0.8919, which covers a little bit more than 80% of dataset.

2. PCA Analysis

Summary of PCA

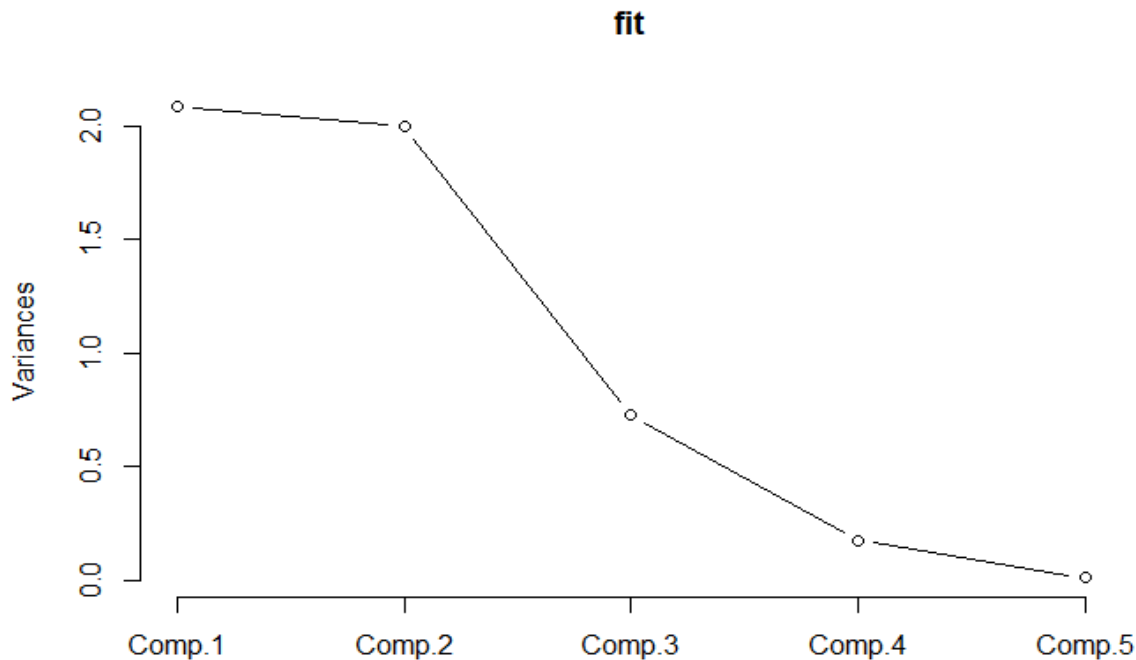
```
> summary(fit) # print variance accounted for,  
Importance of components:
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.4448777	1.4127358	0.8547196	0.41790901	0.106360142
Proportion of Variance	0.4175343	0.3991645	0.1461091	0.03492959	0.002262496
Cumulative Proportion	0.4175343	0.8166988	0.9628079	0.99773750	1.000000000

Explanation:

Important component in our case is Component 2 - 0.8166988, where we can see that Comp.2 extracts 0.8166 of cumulative variance of dataset.

PCA is Interdependent type of method which we use to find out the difference between two classes.



Plot of Components versus Variance.

```
> loadings(fit) # pc loadings
```

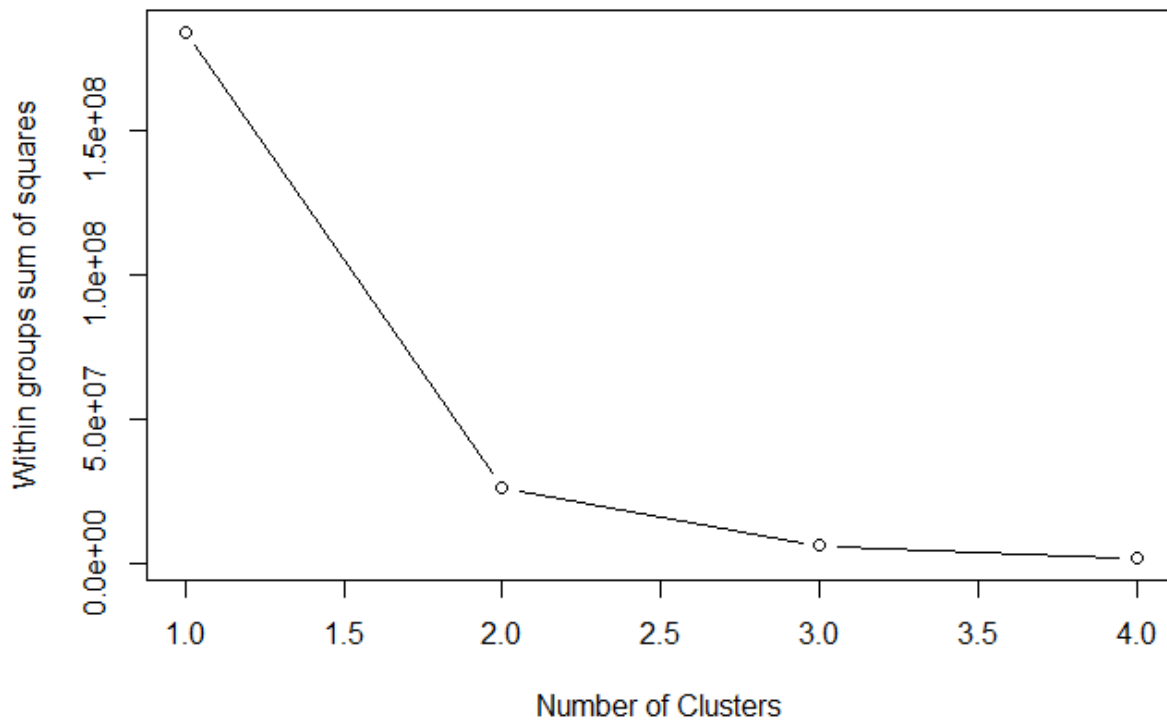
Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
projID	0.138	0.653	-0.152	-0.728	
prgLng	-0.127	-0.485	-0.812	-0.293	
arch	0.398	0.444	-0.543	0.590	
sizeProj	-0.655	0.207	-0.117	0.129	-0.706
effortProj	-0.614	0.314		0.142	0.704

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
SS loadings	1.0	1.0	1.0	1.0	1.0
Proportion var	0.2	0.2	0.2	0.2	0.2
Cumulative Var	0.2	0.4	0.6	0.8	1.0

Variance and it's components.

3. K-means clustering Analysis



Normalised WSS

```
> nwss  
[1] 0.1 0.2 0.3 0.4
```

SWSS

```
> swss=cumsum(nwss) # cumulative WSS  
> swss  
[1] 0.1 0.3 0.6 1.0
```

Important clusters would be 1 and 2.

4. Canonical Correlation Analysis

Summary:

```
> summary(data)
      projID      prgLng      arch      sizeProj      effortProj
Min.   : 1      Min.   :1.000  Min.   :1.000  Min.   : 38.0  Min.   : 262
1st Qu.: 5      1st Qu.:1.000  1st Qu.:1.000  1st Qu.:129.0  1st Qu.: 448
Median : 9      Median :2.000  Median :1.000  Median :182.0  Median :1114
Mean   : 9      Mean   :2.176  Mean   :1.353  Mean   :551.2  Mean   :2093
3rd Qu.:13     3rd Qu.:3.000  3rd Qu.:2.000  3rd Qu.:388.0  3rd Qu.:1947
Max.   :17     Max.   :4.000  Max.   :2.000  Max.   :3656.0  Max.   :13905
```

Correlation between variables X and Y:

```
> cor(x,y) # correlation between two sets of variables
      sizeProj effortProj
prgLng 0.0361900 -0.09275729
arch   -0.3010737 -0.18048327
```

There is no correlation between prgLng and EffortProj.

Cross - Correlation between variables X and Y:

```
> matcor(x, y) # to see the autocorrelations and cross-corrlations
$Xcor
      prgLng      arch
prgLng 1.0000000 -0.2431259
arch   -0.2431259 1.0000000

$Ycor
      sizeProj effortProj
sizeProj 1.0000000 0.9752405
effortProj 0.9752405 1.0000000

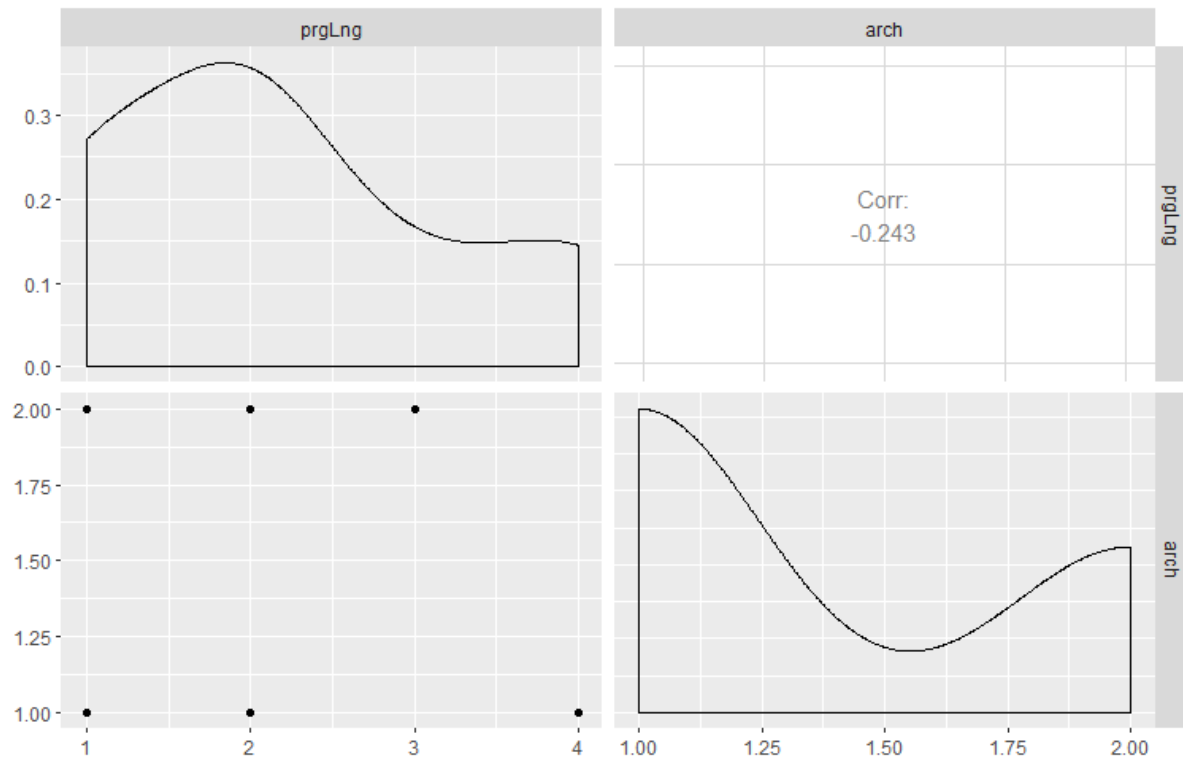
$XYcor
      prgLng      arch      sizeProj      effortProj
prgLng 1.00000000 -0.2431259 0.0361900 -0.09275729
arch   -0.24312591 1.0000000 -0.3010737 -0.18048327
sizeProj 0.03619000 -0.3010737 1.0000000 0.97524048
effortProj -0.09275729 -0.1804833 0.9752405 1.00000000
```

We can see high correlation between sizeProj and effortProj.

Canonical Correlation between variables X and Y:

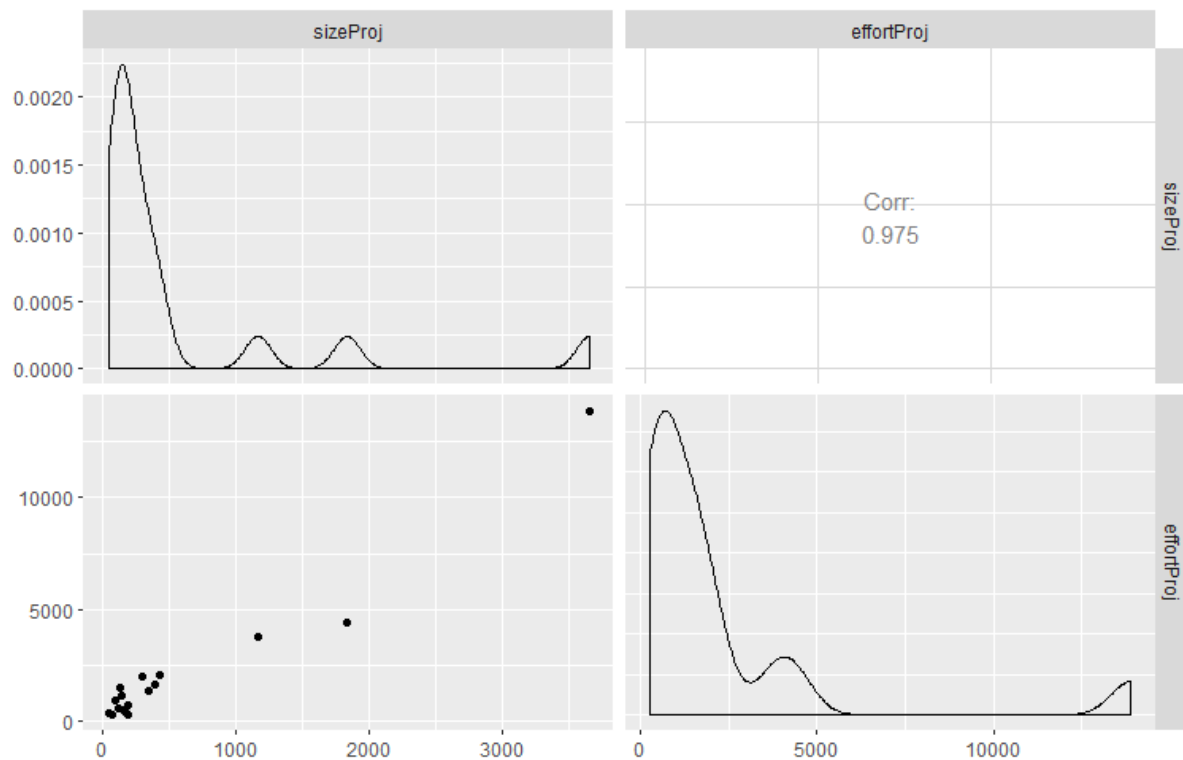
```
> cc1$cor
[1] 0.7240861 0.2218479
```

GGpairs plot for variables of set X



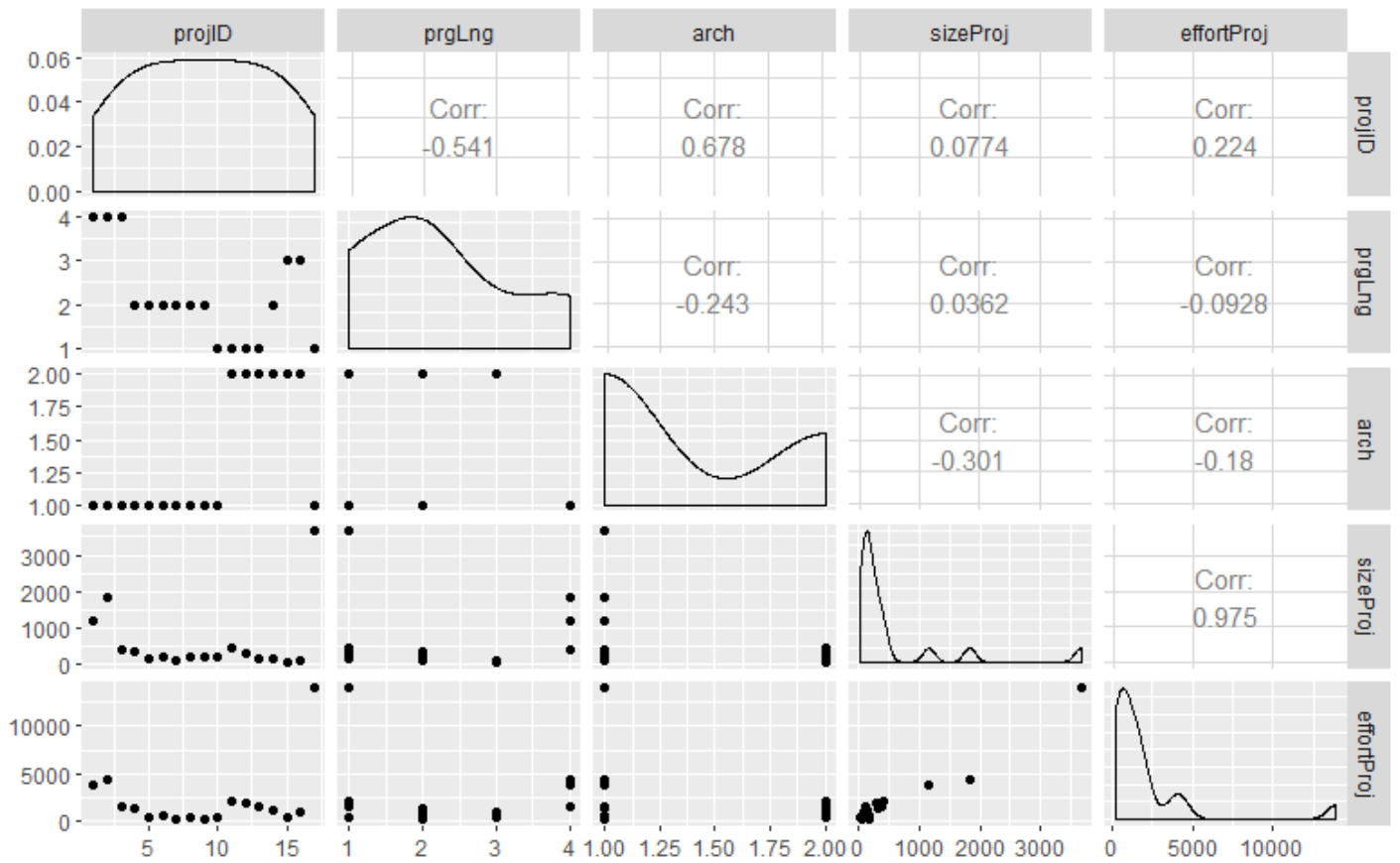
There is no high correlation between variables in set X.

GGpairs plot for variables of set Y.



There is high correlation between variables in set Y.

GGpairs plot for dataset.



From the plot we can see high correlation (0.975) between sizeProj and effortProject. There is also relatively high correlation between Arch and projID.