Principal Component Analysis

Principal component analysis is a technique used to construct composite variable(s) such that composite variable(s) are weighted combination of original or measured variables.

- **Applications**

  - To transform data matrix to new set of variables that are completely independent of each other. These independent variables can be used as input to regression analysis or these could also be used as dependent variables to be explained by other independent variables not included in principal component analysis.
  - To reduce number of variables to a manageable number of components by examining those principal dimensions and ignoring not so central ones.
    1. Products that are currently marketed may only have 3-5 principal components.
    2. Consumer characteristics such as life-cycle, lifestyle and socio-economic status may be formed by combining several measured variables.
  - To identify structure to discover the basic structure underlying a set of measures (assumption is that some measures are redundant or completely overlap with other variables).
    1. Semantic differential scales form three basic dimension to describe emotional responses; Pleasure, activity, and potency.
    2. Likert scales are often used to understand feeling towards objects.
    3. Structure of abilities, verbal and quantitative dimensions.

- **Steps in Principal Component Analysis**

  1. Should principal component analysis be applied to the data?
     - analysis objective, reduce number of variables, underlying structure, etc.
     - Questions, items interrelated. Examine correlation matrix. Low correlations across all measures indicate lack of interrelatedness among variables. This is not particularly useful for principal component analysis.
     - Analysis by variable or object? Compute appropriate correlation matrix.
  2. Some cautions and warnings.
     - Do indicators have measurement error and is there interest in separating it from other components?

- Are variables (objects) related for the sample?
- Are some variables conceptually and / or causally related?

3. Decision about number of components. The maximum number of components that can be extracted is equal to number of variables.
    - Rule based
        - Each component must explain certain amount of variance.
        - A priori criterion.
        - Fixed number of components.
    - Conceptually meaningful
    - Statistical rules
4. Interpretation of Components.
5. Validation of Component analysis.
    - another sample
    - other set of variables in the same data set.
    - Predictive validation
    - Examining Component scores

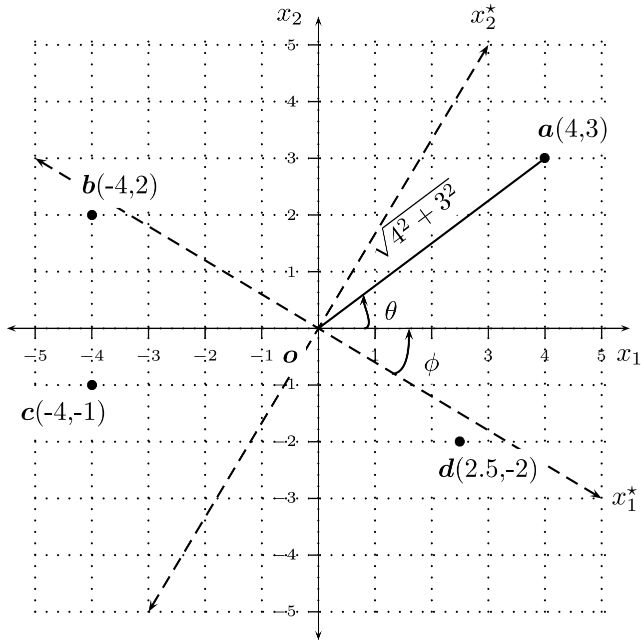
Before we examine process of deriving principal components when there are many variables, we will examine geometry of changing axis and apply that approach to an example with eight observations. We will then proceed to derive generalization to many variables.

Geometry of Axis Transformation

Any point can be represented either in co-ordinate system or polar system. Consider vector $\boldsymbol{a}$ may be represented with co-ordinates (4,3) or $(r\cos\theta, r\sin\theta)$, where r is distance between the point and the origin, or $r = \sqrt{x_1^2 + x_2^2} = \sqrt{4^2 + 3^2}$. Thus, point may be written as $(5\cos\theta, 5\sin\theta)$. Finally, definition of cosine is

$$\cos\theta = \frac{x_1}{r} \quad \text{and} \quad \sin\theta = \frac{x_2}{r}.$$

From these definition, it follows that $\sin^2\theta + \cos^2\theta = 1$.



Suppose both axes are rotated at an angle ϕ. The result of this rotation is new axes, call them $x_1^\star$ and $x_2^\star$. Each point can be projected on to new axes as $r\cos(\theta+\phi)$ and $r\sin(\theta+\phi)$. But from trignometric rules[1], these could be written as

$$\begin{aligned}
\boldsymbol{a}^\star &= \begin{bmatrix} r(\cos\theta\cos\phi - \sin\theta\sin\phi) & r(\cos\theta\sin\phi + \sin\theta\cos\phi) \end{bmatrix} \\
&= \begin{bmatrix} r\cos\theta & r\sin\theta \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \\
&= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}.
\end{aligned}$$

Matrix on the right hand side of equal sign containing ϕ is called a transformation matrix. Note also another interesting pattern. Observations $\boldsymbol{b}$ and $\boldsymbol{d}$ were far from both original axes but both of these points are close to axis $x_1^\star$. It should come as no surprise that if point lies on a particular axis, then the other dimension will be zero and consequently, all variability is concentrated along one axis. Obviously, there has to be some rules governing these new axes. An idea of *Principal Component Analysis* is to construct new axes such that variability is concentrated among few axes. The word *principal* refers to main, primary, fundamental or major component of variables.

- **Illustration: Highly Correlated Variables**

This idea of transforming and constructing new axes will be illustrated using price for gasoline and diesel in eight OECD countries as of August 1999. All prices are in the US dollar and

[1] Note that $\cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi$ and $\sin(\theta+\phi) = \cos\theta\sin\phi + \sin\theta\cos\phi$. This is helpful in subsequent derivation

show considerable variation. Note that the first variable (price of gasoline) has variance of 0.1192 compared to the second variable (price of diesel) has variance of 0.0565. Thus, the total variance explained is $0.1192 + 0.0565$ or 0.1757. Suppose overall mean price is substracted from actual country prices, then resulting set of prices will have mean of zero. Note substracting mean had no influence on variance, covariance or correlation.

	Price per litre		Obs. with Mean 0	
Country	Gasoline	Diesel	Gasoline	Diesel
France	1.09	0.63	0.2158	0.0031
Germany	0.92	0.58	0.0468	-0.0509
Italy	1.07	0.68	0.1908	0.0521
Spain	0.80	0.54	-0.0752	-0.0869
UK	1.37	1.09	0.4978	0.4591
Japan	0.98	0.79	0.1048	0.1601
Canada	0.43	0.39	-0.4443	-0.2379
USA	0.34	0.33	-0.5363	-0.2989
Mean	0.8743	0.6289	0.00	0.00
Variance	0.1192	0.0565	0.1192	0.0565
Covariance	0.0751		0.0751	
correlation	0.9160		0.9160	

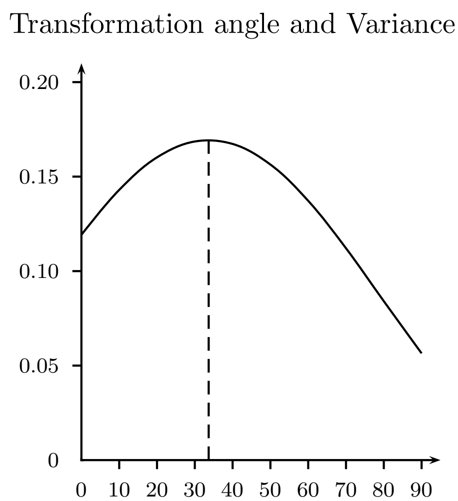
Suppose transformation suggested on previous page is applied to these observations. To accomplish this task one needs to know appropriate angle to transform. Let us start with some arbitrary angle and keep changing until suitable angle can be obtained. For example, when angle is 20 degrees transformed numbers for France are obtained by multiplying following

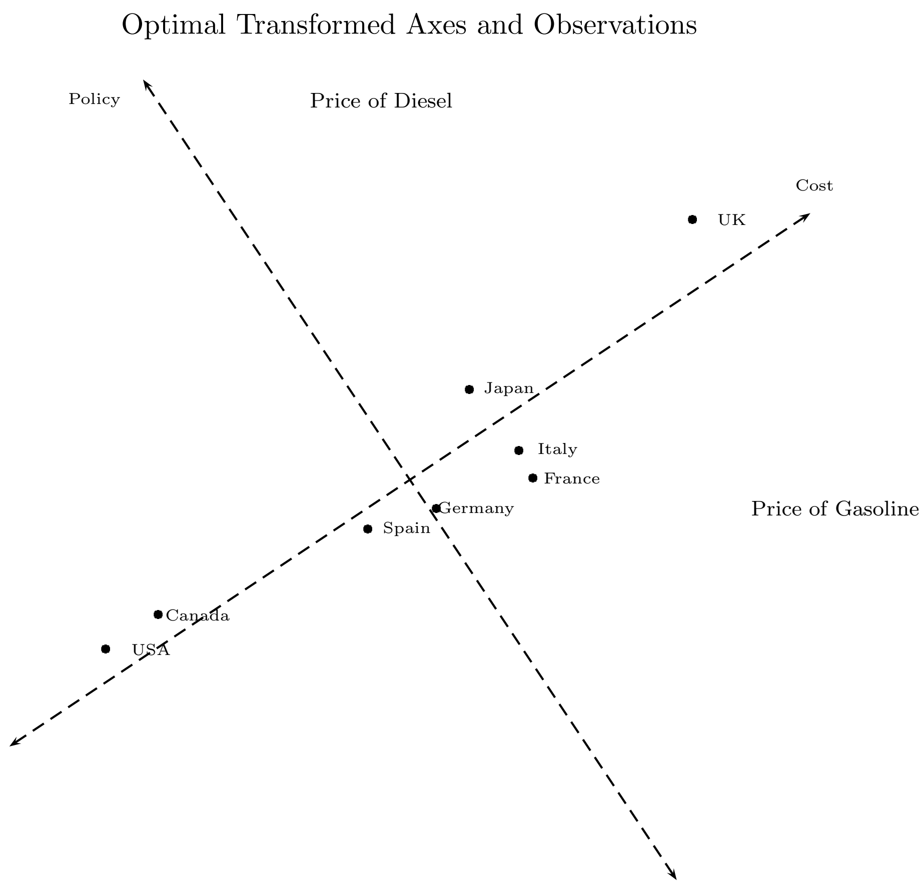
$$\begin{aligned}(x_1^\star, x_2^\star) &= (x_1, x_2)\begin{bmatrix}\cos(20) & -\sin(20)\\ \sin(20) & \cos(20)\end{bmatrix}\\ &= (0.2158, 0.0031)\begin{bmatrix}0.9397 & -0.3420\\ 0.3420 & 0.9397\end{bmatrix}\end{aligned}$$

This process is continued for all observations and then summary measures re-estimated. The key criterion that we wish to maximize is the variation explained by the first axis. A sample of transformations is provided on the next page. Note that variation explained by the first transformed axis increases until 40 degrees and after that angle, it decreases. After changing angle several times, the maximum variance explained by the first axis or component occured at about 33.67 degrees (see figures below). At such point, weight matrix is

$$\begin{bmatrix}0.8322 & -0.5544\\ 0.5544 & 0.8322\end{bmatrix}.$$

Note also that the first new axis account for 96.4% of variation in original variables and correlations among new variables is 0.


Transformation angle and Variance
0.20
0.15
0.10
0.05
0
0 10 20 30 40 50 60 70 80 90


Optimal Transformed Axes and Observations
Policy
Price of Diesel
Cost
UK
Japan
Italy
France
Germany
Price of Gasoline
Spain
Canada
USA

Country	Price per litre		Transformation angle in degrees											
	Gasoline	Diesel	0		20		40		60		80		90	
France	0.2158	0.0031	0.2158	0.0031	0.2038	-0.0709	0.1673	-0.1363	0.1106	-0.1853	0.0405	-0.2119	0.0031	-0.2158
Germany	0.0468	-0.0509	0.0468	-0.0509	0.0265	-0.0638	0.0031	-0.0690	-0.0207	-0.0659	-0.0420	-0.0549	-0.0509	-0.0468
Italy	0.1908	0.0521	0.1908	0.0521	0.1971	-0.0163	0.1796	-0.0827	0.1405	-0.1391	0.0845	-0.1788	0.0521	-0.1908
Spain	-0.0752	-0.0869	-0.0752	-0.0869	-0.1004	-0.0559	-0.1135	-0.0182	-0.1129	0.0217	-0.0986	0.0590	-0.0869	0.0752
UK	0.4978	0.4591	0.4978	0.4591	0.6248	0.2612	0.6764	0.0318	0.6465	-0.2015	0.5386	-0.4105	0.4591	-0.4978
Japan	0.1048	0.1601	0.1048	0.1601	0.1532	0.1146	0.1832	0.0553	0.1910	-0.0107	0.1759	-0.0754	0.1601	-0.1048
Canada	-0.4443	-0.2379	-0.4443	-0.2379	-0.4988	-0.0716	-0.4932	0.1033	-0.4281	0.2658	-0.3114	0.3962	-0.2379	0.4443
USA	-0.5363	-0.2989	-0.5363	-0.2989	-0.6061	-0.0974	-0.6029	0.1157	-0.5270	0.3150	-0.3875	0.4762	-0.2989	0.5363
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Variance	0.1192	0.0565	0.1192	0.0565	0.1602	0.0155	0.1673	0.0084	0.1372	0.0384	0.0841	0.0916	0.0565	0.1192
Covariance	0.0751		0.0751		0.0374		-0.0178		-0.0647		-0.0813		-0.0751	
correlation	0.9160		0.9160		0.7507		-0.4767		-0.8914		-0.9270		-0.9160	
		Cosine	1.0000		0.9397		0.7660		0.5000		0.1736		0.0000	
		sine	0.0000		0.3420		0.6428		0.8660		0.9848		1.0000	

Country	Price per litre		Transformation angle in degrees											
	Gasoline	Diesel	33.4		33.5		33.67		33.7		33.8		33.9	
	Gasoline	Diesel	33.4		33.5		33.67		33.7		33.8		33.9	
France	0.2158	0.0031	0.1818	-0.1162	0.1816	-0.1165	0.1813	-0.1170	0.1812	-0.1171	0.1810	-0.1174	0.1808	-0.1177
Germany	0.0468	-0.0509	0.0110	-0.0682	0.0109	-0.0682	0.0107	-0.0683	0.0107	-0.0683	0.0105	-0.0683	0.0104	-0.0683
Italy	0.1908	0.0521	0.1879	-0.0615	0.1878	-0.0618	0.1876	-0.0624	0.1876	-0.0625	0.1875	-0.0628	0.1874	-0.0631
Spain	-0.0752	-0.0869	-0.1106	-0.0311	-0.1107	-0.0309	-0.1108	-0.0306	-0.1108	-0.0305	-0.1109	-0.0303	-0.1109	-0.0301
UK	0.4978	0.4591	0.6683	0.1093	0.6685	0.1081	0.6688	0.1061	0.6688	0.1058	0.6690	0.1046	0.6692	0.1035
Japan	0.1048	0.1601	0.1756	0.0760	0.1757	0.0757	0.1760	0.0752	0.1760	0.0751	0.1761	0.0748	0.1763	0.0745
Canada	-0.4443	-0.2379	-0.5018	0.0460	-0.5017	0.0468	-0.5016	0.0483	-0.5016	0.0486	-0.5015	0.0495	-0.5014	0.0503
USA	-0.5363	-0.2989	-0.6122	0.0457	-0.6121	0.0467	-0.6120	0.0486	-0.6120	0.0489	-0.6119	0.0500	-0.6118	0.0510
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Variance	0.1192	0.0565	0.169261	0.0064	0.169263	0.0064	0.169264	0.0064	0.169264	0.0064	0.169263	0.0064	0.169262	0.0064
Covariance	0.0751		0.0008		0.0005		0.0000		-0.000075		-0.0004		-0.0007	
correlation	0.9160		0.0233		0.0147		0.0000		-0.0026		-0.0112		-0.0199	
	sine		0.5505		0.5519		0.5544		0.5548		0.5563		0.5577	
	Cosine		0.8348		0.8339		0.8322		0.8320		0.8310		0.8300	

Let us see whether above analysis can be reproduced using SAS.

- **SAS Input for Principal Component Analysis**

```
options nodate nocenter ls=80 ps=60 ;
data gasol;
Input country$ gasol diesel ;
cards;
France  1.090 0.632
Germany 0.921 0.578
Italy   1.065 0.681
Spain   0.799 0.542
UK      1.372 1.088
Japan   0.979 0.789
Canada  0.430 0.391
USA     0.338 0.330
;;;;
proc princomp cov out=new;
var gasol diesel;
proc print;
var gasol diesel prin1 prin2;
proc corr;
var gasol diesel prin1 prin2;
run;
```

- **SAS Output**

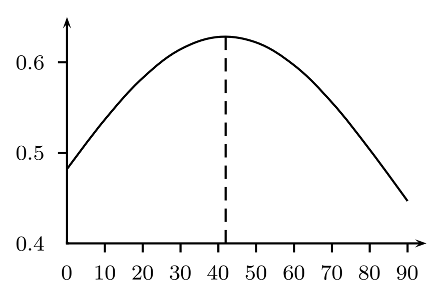
```
Simple Statistics

                 GASOL           DIESEL

Mean      0.8742500000      0.6288750000
StD       0.3452600469      0.2376097867


Covariance Matrix

                   GASOL           DIESEL

GASOL       0.1192045000      0.0751466071
DIESEL      0.0751466071      0.0564584107

Total Variance = 0.1756629107

Eigenvalues of the Covariance Matrix

           Eigenvalue      Difference      Proportion      Cumulative

PRIN1        0.169264        0.162865        0.963574         0.96357
PRIN2        0.006399         .              0.036426         1.00000


Eigenvectors

               PRIN1         PRIN2

GASOL       0.832245      -.554408
DIESEL      0.554408      0.832245
```

```
OBS    GASOL    DIESEL      PRIN1       PRIN2

 1     1.090     0.632     0.18129    -0.11701
 2     0.921     0.578     0.01070    -0.06826
 3     1.065     0.681     0.18765    -0.06237
 4     0.799     0.542    -0.11079    -0.03058
 5     1.372     1.088     0.66879     0.10615
 6     0.979     0.789     0.17595     0.07519
 7     0.430     0.391    -0.50160     0.04833
 8     0.338     0.330    -0.61199     0.04856

Correlation Analysis

   4 'VAR' Variables:  GASOL    DIESEL   PRIN1    PRIN2

                               Simple Statistics

Variable           N        Mean     Std Dev         Sum     Minimum     Maximum

GASOL              8     0.87425     0.34526     6.99400     0.33800     1.37200
DIESEL             8     0.62888     0.23761     5.03100     0.33000     1.08800
PRIN1              8           0     0.41142           0    -0.61199     0.66879
PRIN2              8           0     0.07999           0    -0.11701     0.10615


Pearson Correlation Coefficients  / N = 8

                   GASOL           DIESEL           PRIN1            PRIN2

GASOL            1.00000          0.91601          0.99172         -0.12845

DIESEL           0.91601          1.00000          0.95995          0.28018

PRIN1            0.99172          0.95995          1.00000          0.00000

PRIN2           -0.12845          0.28018          0.00000          1.00000
```

- **Illustration: Moderately Correlated Variables**

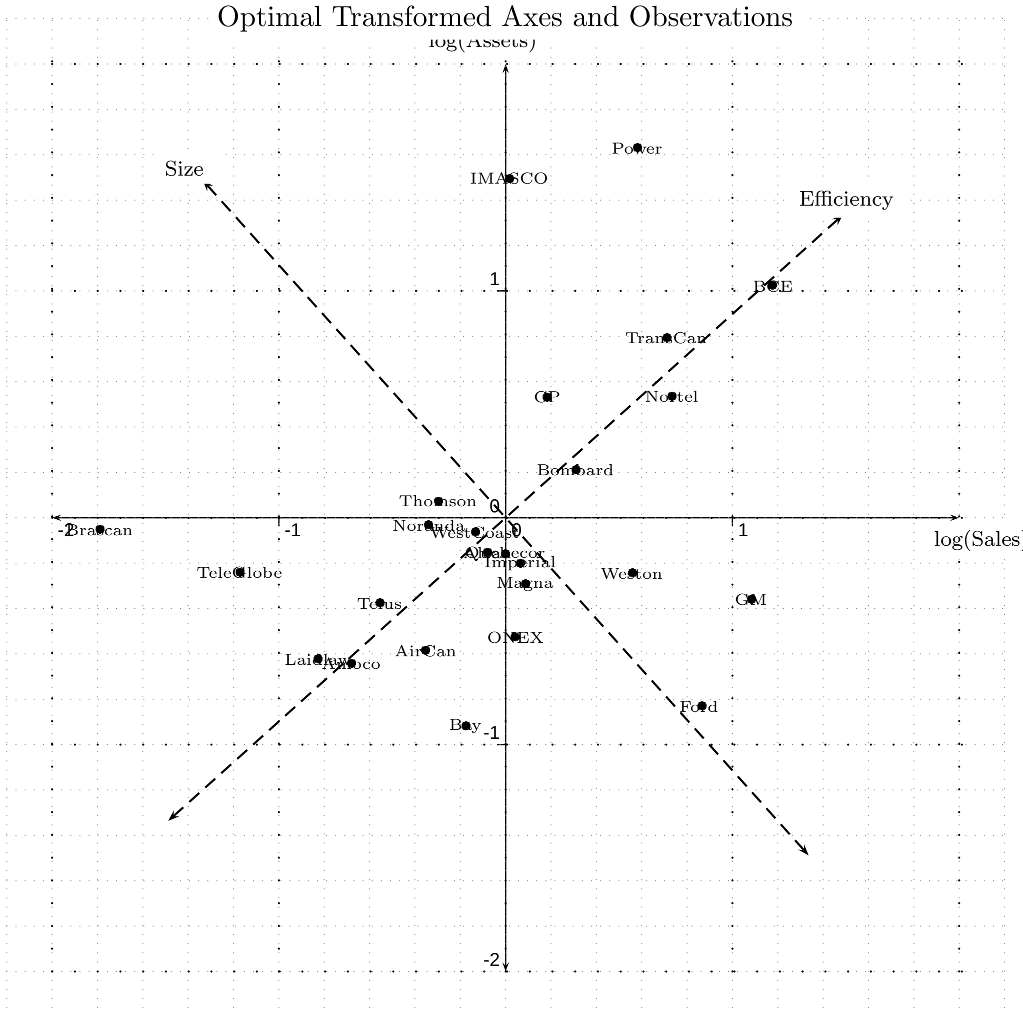
Consider a situation where variables are completely independent. In such situation, deriving principal component is likely to be less interesting. As an illustration, let us consider sales and assets of 25 largest industrial organizations in Canada. There is a modest correlation among these variables. As might one expect, there is slightly higher correlation, when we transform sales and assets in log units. One would repeat procedure that we used for earlier example and find that the optimal angle is 41.92 and at that point the first axis accounts for 67.6% of variation, a modest improvement from 50%.

Transformation angle and Variance



Sales and Assets of 25 Largest Industrial Organization in Canada, 1999

	in $'000		in log units		Obs. with Mean 0	
Company	Sales	Assets	Sales	Assets	Sales	Assets
GM	$24,919	$8,051	17.031	15.901	1.0824	-0.3600
BCE	$27,454	$32,072	17.128	17.283	1.1793	1.0222
FORD	$20,101	$5,029	16.816	15.431	0.8675	-0.8305
NORTEL	$17,575	$19,732	16.682	16.798	0.7333	0.5365
TRANSCAN	$17,155	$25,561	16.658	17.057	0.7091	0.7953
POWER	$15,101	$58,925	16.530	17.892	0.5816	1.6305
WESTON	$14,726	$9,036	16.505	16.017	0.5564	-0.2445
ALCAN	$7,789	$9,901	15.868	16.108	-0.0805	-0.1531
BOMBARD	$11,500	$14,272	16.258	16.474	0.3092	0.2126
CP	$10,151	$19,669	16.133	16.795	0.1844	0.5333
THOMSON	$6,269	$12,447	15.651	16.337	-0.2976	0.0757
MAGNA	$9,191	$8,621	16.034	15.970	0.0850	-0.2916
ONEX	$8,813	$6,820	15.992	15.735	0.0430	-0.5259
IMASCO	$8,584	$51,522	15.965	17.758	0.0167	1.4963
QUEBECOR	$8,425	$9,841	15.947	16.102	-0.0020	-0.1591
IMPERIAL	$9,002	$9,429	16.013	16.059	0.0642	-0.2019
WESTCOAST	$7,376	$10,820	15.814	16.197	-0.1350	-0.0643
BAY	$7,075	$4,604	15.772	15.342	-0.1766	-0.9188
NORANDA	$6,013	$11,175	15.609	16.229	-0.3393	-0.0321
AIRCAN	$5,932	$6,422	15.596	15.675	-0.3528	-0.5860
TELUS	$4,843	$7,923	15.393	15.885	-0.5557	-0.3759
BRASCAN	$1,409	$10,942	14.158	16.208	-1.7903	-0.0531
LAIDLAW	$3,690	$6,185	15.121	15.638	-0.8275	-0.6237
AMOCO	$4,270	$6,076	15.267	15.620	-0.6816	-0.6414
TELEGLOBE	$2,612	$9,072	14.776	16.021	-1.1731	-0.2405
Mean	$10,399	$14,966	15.95	16.26	0.00	0.00
Variance			0.4821	0.4468	0.4821	0.4468
Covariance			0.1565		0.1565	
Correlation	0.3069		0.3371		0.3371	



General Process of Specifying Principal Components

Intuitively, we are trying to identify new axes called *Principal Component* and the resulting co-ordinates that we obtain are called *Principal Component Scores* such that

1. The first components account for the maximum variance in the original variables.
2. Each new component is a linear combination of original variables.
3. The second and subsequent components account for the maximum variance that has not accounted for by the preceding components.
4. All components are uncorrelated to each other.

Above intuitive idea can be written more formally. Suppose we have p variables and number of objects is n which will be subjected to principal component analysis. Each data point is denoted by x_{ij} where i stands for variable ($i = 1, \cdots, p$) and j for object ($j = 1, \cdots, n$). Let is denote transformation matrix a_{li} where l stands for component ($l = 1, \cdots, p$) and it is of size $p \times p$. Finally, if y_{lj} denote component scores, then mathematically $y_{lj} = a_{li} \times x_{ij}$ with a_{li} to be unknowns to be determined by the principal component analysis. In matrix notation, this can be written as $\boldsymbol{Y} = \boldsymbol{AX}$ and idea of independence can be expressed as $\boldsymbol{A}'\boldsymbol{A}$ is an identity matrix, denoted by $\boldsymbol{I}$. Let us see appearance of this for one observation, with observation j.

$$\begin{bmatrix} y_{1j} & y_{2j} & \cdots & y_{pj} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \begin{bmatrix} x_{1j} & x_{2j} & \cdots & x_{pj} \end{bmatrix}.$$

If each element of each row of matrix $\boldsymbol{A}$ is squared and summed for the row, then each sum add to one. Formally, $a_{i1}^2 + a_{i2}^2 + \cdots + a_{ip}^2 = 1$ for $i = 1, \cdots, p$. Moreover, sum of cross products of each pair of rows is zero. Formally, $a_{i1}a_{j1} + a_{i2}a_{j2} + \cdots + a_{ip}a_{jp} = 0$ for all $i \neq j$. To determine matrix $\boldsymbol{A}$, let simplify expression for the variance of $\boldsymbol{Y}$. We know that variance for matrix $\boldsymbol{Y}$ can be written as $\mathcal{E}(\boldsymbol{YY}')$ and we want to maximize variance subject to $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$. We can write maximization problem as follows:

$$\underset{\boldsymbol{A}}{\text{Max}}\ \mathcal{E}(\boldsymbol{YY}') - \boldsymbol{\lambda}(\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{I}),$$

where $\boldsymbol{\lambda}$ is vector of the Langrangian multiplier incorporating constraints on optimatization. Let us see the expected value component. Note that $\mathcal{E}(\boldsymbol{YY}')$ is equal to

$$\begin{aligned} \mathcal{E}(\boldsymbol{YY}') &= \mathcal{E}(\boldsymbol{AX}[\boldsymbol{AX}]') \\ &= \mathcal{E}(\boldsymbol{AXX}'\boldsymbol{A}') \\ &= \boldsymbol{ASA}' \end{aligned}$$

where $\boldsymbol{S}$ is a sample covariance matrix. We can use calculus and solve this optimization. The first order condition then would imply that $(\boldsymbol{S} - \boldsymbol{\lambda I})\boldsymbol{A} = 0$, which is same as obtaining solution to eigenvalues and eigenvectors. Intuitively, we note that eigenvectors are transformation matrix which allow us to create a set of new axes and each axis account for maximum variability that is not accounted for by prior axes. Moreover, eigenvalues are amount of variance that is accounted by that component and we would expect that the first eigenvalue to be the largest and second to be less than or equal to the first one and so on.

Application of Principal Component Analysis

When there are many variables, principal component analysis by itself does not provide insightful picture of nature of correlations among variables. We will illustrate this using two different datasets. First dataset is about six variables related to demographic and land distribution for 273 metropolitan regions of the US. In this case, variables have modest to low correlations among other variables. The second dataset is based on the national records for eight track and field events (100 m, 200 m, etc.) for 192 countries as of November 1999. A quick look at correlation matrix indicates that these records are highly correlated across nations. In both datasets, first one or two components appear to have a large loadings on all variables. All other components have one or two variables with a large eigenvector and remaining variable have low loadings. This analysis suggest that principal component analysis applied by itself or end-analysis, may have limited application.

```
Principal Component Analysis

     273 Observations
       6 Variables
POPUL    -  Number of individuals living in a region
BIRTHS   -  Number of births per 1,000 habitants
AGE65    -  Percent of indiviuals with age more than 65.
AGE35_54 -  Percent of individuals with age 35 to 54.
LESS5    -  Percent of individuals with age less than 5 years.
LANDA    -  Land area in square miles.

Simple Statistics

                 POPUL             BIRTHS              AGE65

Mean        783388.168        14.78168498        12.73882784
StD        1868378.705         2.78436313         3.59937108


              AGE35_54              LESS5              LANDA

Mean       27.95421245        7.153113553        2591.630037
StD         2.07672470        1.056840208        3780.196337


Correlation Matrix

               POPUL      BIRTHS       AGE65    AGE35_54       LESS5       LANDA

POPUL         1.0000      0.1321      -.0638      0.1714      0.1084      0.5391
BIRTHS        0.1321      1.0000      -.5531      -.3451      0.9060      0.2124
AGE65         -.0638      -.5531      1.0000      -.0733      -.5605      -.1417
AGE35_54      0.1714      -.3451      -.0733      1.0000      -.3498      0.1370
LESS5         0.1084      0.9060      -.5605      -.3498      1.0000      0.2049
LANDA         0.5391      0.2124      -.1417      0.1370      0.2049      1.0000


Eigenvalues of the Correlation Matrix

           Eigenvalue      Difference      Proportion      Cumulative
```

PRIN1	2.55970	0.968650	0.426616	0.42662
PRIN2	1.59105	0.633417	0.265174	0.69179
PRIN3	0.95763	0.502689	0.159605	0.85140
PRIN4	0.45494	0.111848	0.075824	0.92722
PRIN5	0.34309	0.249502	0.057182	0.98440
PRIN6	0.09359	.	0.015599	1.00000

Eigenvectors

	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5	PRIN6
POPUL	0.167807	0.624008	0.300440	0.689705	-.126528	0.022104
BIRTHS	0.585471	-.126061	0.024374	0.052982	0.378595	-.703276
AGE65	-.436686	-.016839	0.616809	-.027577	0.653981	0.010839
AGE35_54	-.200957	0.476387	-.680011	0.009317	0.519762	0.004251
LESS5	0.584822	-.140301	0.012273	0.023408	0.364531	0.710435
LANDA	0.236579	0.589755	0.257144	-.721183	-.099652	-.00782

Tract and Field Event National Record Dataset

Principal Component Analysis

192 Observations
8 Variables

```
M100    -  100 m record in seconds
M200    -  200 m record in seconds
M400    -  400 m record in seconds
M800    -  800 m record in seconds
M1500   -  1500 m record in seconds
M5000   -  5000 m record in seconds
M10000  -  10000 m record in seconds
MARA    -  Marathon record in minutes
```

Simple Statistics

	M100	M200	M400	M800
Mean	10.47010417	21.12052083	46.63562500	149.9875000
StD	0.37924559	0.84466403	1.71733436	9.7746967

	M1500	M5000	M10000	MARA
Mean	351.6123958	1400.031823	2972.134688	223.9190104
StD	23.3845744	88.430482	227.904489	19.6120945

Correlation Matrix

	M100	M200	M400	M800	M1500	M5000	M10000	MARA
M100	1.0000	0.9273	0.7907	0.4591	0.4497	0.4512	0.4914	0.3830
M200	0.9273	1.0000	0.8452	0.5047	0.4975	0.5436	0.5632	0.4372
M400	0.7907	0.8452	1.0000	0.6366	0.5977	0.6068	0.5819	0.4718
M800	0.4591	0.5047	0.6366	1.0000	0.6752	0.6061	0.5283	0.3263
M1500	0.4497	0.4975	0.5977	0.6752	1.0000	0.7984	0.7367	0.5353
M5000	0.4512	0.5436	0.6068	0.6061	0.7984	1.0000	0.9057	0.6353
M10000	0.4914	0.5632	0.5819	0.5283	0.7367	0.9057	1.0000	0.6732

MARA	0.3830	0.4372	0.4718	0.3263	0.5353	0.6353	0.6732	1.0000

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	5.19419	4.01211	0.649274	0.64927
PRIN2	1.18208	0.46872	0.147760	0.79703
PRIN3	0.71336	0.35512	0.089170	0.88620
PRIN4	0.35824	0.12646	0.044780	0.93098
PRIN5	0.23178	0.05422	0.028972	0.95996
PRIN6	0.17756	0.09470	0.022195	0.98215
PRIN7	0.08286	0.02293	0.010357	0.99251
PRIN8	0.05993	.	0.007491	1.00000

Principal Component Analysis

Eigenvectors

	PRIN1	PRIN2	PRIN3	PRIN4
M100	0.338087	-.524057	0.156289	-.118454
M200	0.363647	-.460332	0.149546	-.121648
M400	0.378337	-.319605	-.060693	0.167391
M800	0.322035	0.035299	-.692572	0.454735
M1500	0.361797	0.292429	-.282986	-.104725
M5000	0.380203	0.351942	-.004402	-.350109
M10000	0.375377	0.323233	0.165013	-.407610
MARA	0.300422	0.310174	0.602113	0.660815

	PRIN5	PRIN6	PRIN7	PRIN8
M100	0.120772	0.433916	0.098933	0.599956
M200	-.021902	0.135577	-.299426	-.714266
M400	-.024075	-.818463	0.194576	0.121198
M800	-.350748	0.290940	0.019623	-.023493
M1500	0.822274	0.032377	0.077056	-.097051
M5000	-.236167	-.143084	-.671145	0.286886
M10000	-.356179	0.090737	0.635020	-.151296
MARA	0.050567	0.095149	-.051850	0.004815