Stat 5100 Handout #7.1.1 – SAS: Principal Component Regression, Quantile Regression

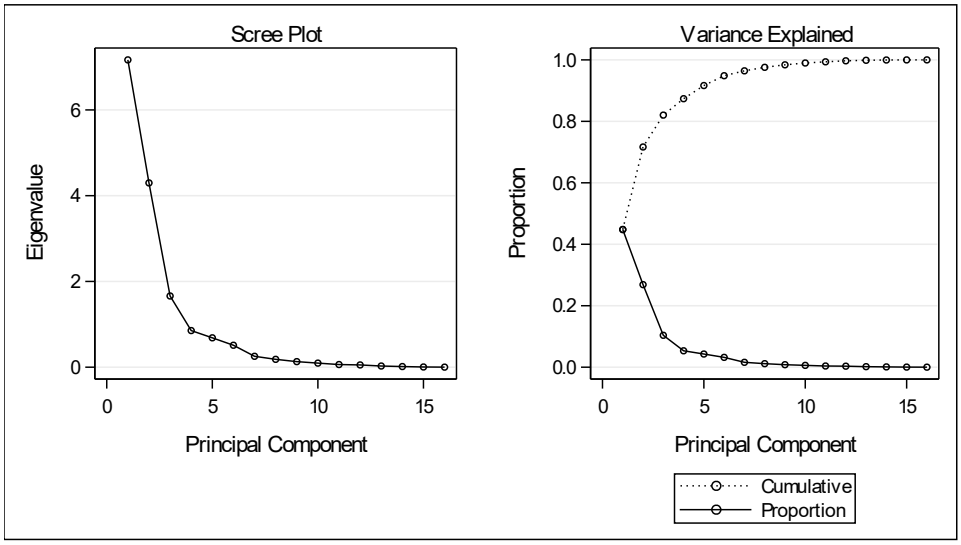
Example: (Baseball, same as Handout 4.1.1 Ex. 2)

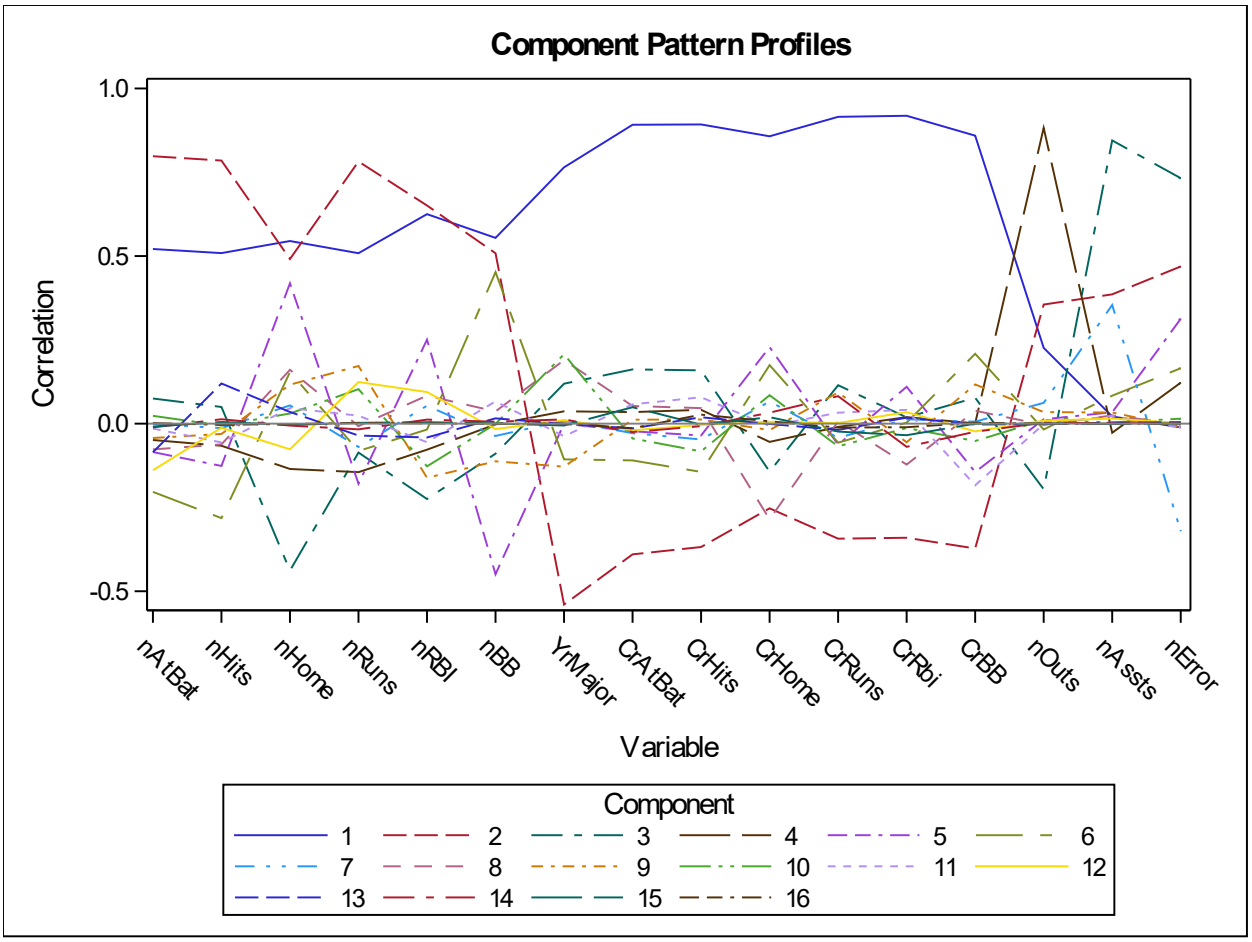
```
data baseball; set sashelp.baseball;

/* check out multicollinearity */
proc reg data=baseball plots=diagnostics;
 model logSalary = nAtBat nHits nHome nRuns nRBI nBB
                   yrMajor crAtBat crHits crHome crRuns crRbi
                   crBB nOuts nAssts nError / vif;
run;
```

Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	Intercept	1	4.30079	0.15849	27.14	<.0001	0
nAtBat	Times at Bat in 1986	1	-0.00235	0.00116	-2.04	0.0429	21.47655
nHits	Hits in 1986	1	0.01340	0.00431	3.11	0.0021	28.44674
nHome	Home Runs in 1986	1	0.00704	0.01120	0.63	0.5300	7.73102
nRuns	Runs in 1986	1	-0.00285	0.00541	-0.53	0.5981	14.54214
nRBI	RBIs in 1986	1	-0.00033415	0.00468	-0.07	0.9432	11.46548
nBB	Walks in 1986	1	0.01145	0.00329	3.48	0.0006	3.96894
YrMajor	Years in the Major Leagues	1	0.07104	0.02256	3.15	0.0018	9.23684
CrAtBat	Career Times at Bat	1	0.00017800	0.00024538	0.73	0.4689	249.85140
CrHits	Career Hits	1	-0.00072565	0.00122	-0.59	0.5530	497.07282
CrHome	Career Home Runs	1	-0.00044509	0.00292	-0.15	0.8791	50.06939
CrRuns	Career Runs	1	0.00149	0.00135	1.10	0.2711	161.01942
CrRbi	Career RBIs	1	0.00015423	0.00126	0.12	0.9024	134.74454
CrBB	Career Walks	1	-0.00128	0.00059287	-2.15	0.0324	20.47714
nOuts	Put Outs in 1986	1	0.00026793	0.00014236	1.88	0.0610	1.25638
nAssts	Assists in 1986	1	0.00024148	0.00040391	0.60	0.5505	2.71651
nError	Errors in 1986	1	-0.00797	0.00797	-1.00	0.3183	2.19559

```
/* Consider principal components */
proc princomp data=baseball standard out=PCout
  plots=patternprofile;
 var nAtBat nHits nHome nRuns nRBI nBB yrMajor crAtBat crHits
     crHome crRuns crRbi crBB nOuts nAssts nError;
run;
```

Eigenvectors									
		Prin1	Prin2	Prin3	Prin4	Prin5	…	Prin15	Prin16
nAtBat	Times at Bat in 1986	0.194633	0.384860	0.058211	-.052847	-.102427		-.126379	0.047791
nHits	Hits in 1986	0.190051	0.378615	0.038601	-.071430	-.152517		0.071796	-.098918
nHome	Home Runs in 1986	0.203519	0.236925	-.340195	-.146328	0.505900		-.055262	-.029052
nRuns	Runs in 1986	0.189951	0.377264	-.066923	-.156591	-.216392		0.036053	0.060174
nRBI	RBIs in 1986	0.233485	0.313965	-.174490	-.084087	0.302800		0.062446	0.026685
nBB	Walks in 1986	0.207015	0.245261	-.069512	-.002478	-.542691		0.009062	-.016386
YrMajor	Years in the Major Leagues	0.285640	-.260301	0.092466	0.039820	0.000112		-.081208	0.018552
CrAtBat	Career Times at Bat	0.333086	-.188121	0.125620	0.037253	-.027497		0.714548	-.388370
CrHits	Career Hits	0.333439	-.177530	0.123382	0.044623	-.042762		-.045866	0.764364
CrHome	Career Home Runs	0.320242	-.121864	-.111153	-.058900	0.276003		0.281744	0.194363
CrRuns	Career Runs	0.341921	-.165545	0.088996	-.010568	-.075900		-.345141	-.331793
CrRbi	Career RBIs	0.343103	-.164207	0.014411	0.021034	0.132858		-.502603	-.304186
CrBB	Career Walks	0.320946	-.179646	0.059861	0.000844	-.174415		-.001737	0.081124
nOuts	Put Outs in 1986	0.084465	0.171405	-.151815	0.954857	0.015633		0.003118	-.002488
nAssts	Assists in 1986	0.008006	0.186094	0.655724	-.029424	0.041027		-.011644	0.007155
nError	Errors in 1986	-.004221	0.226228	0.568323	0.132244	0.377839		0.001090	-.005189

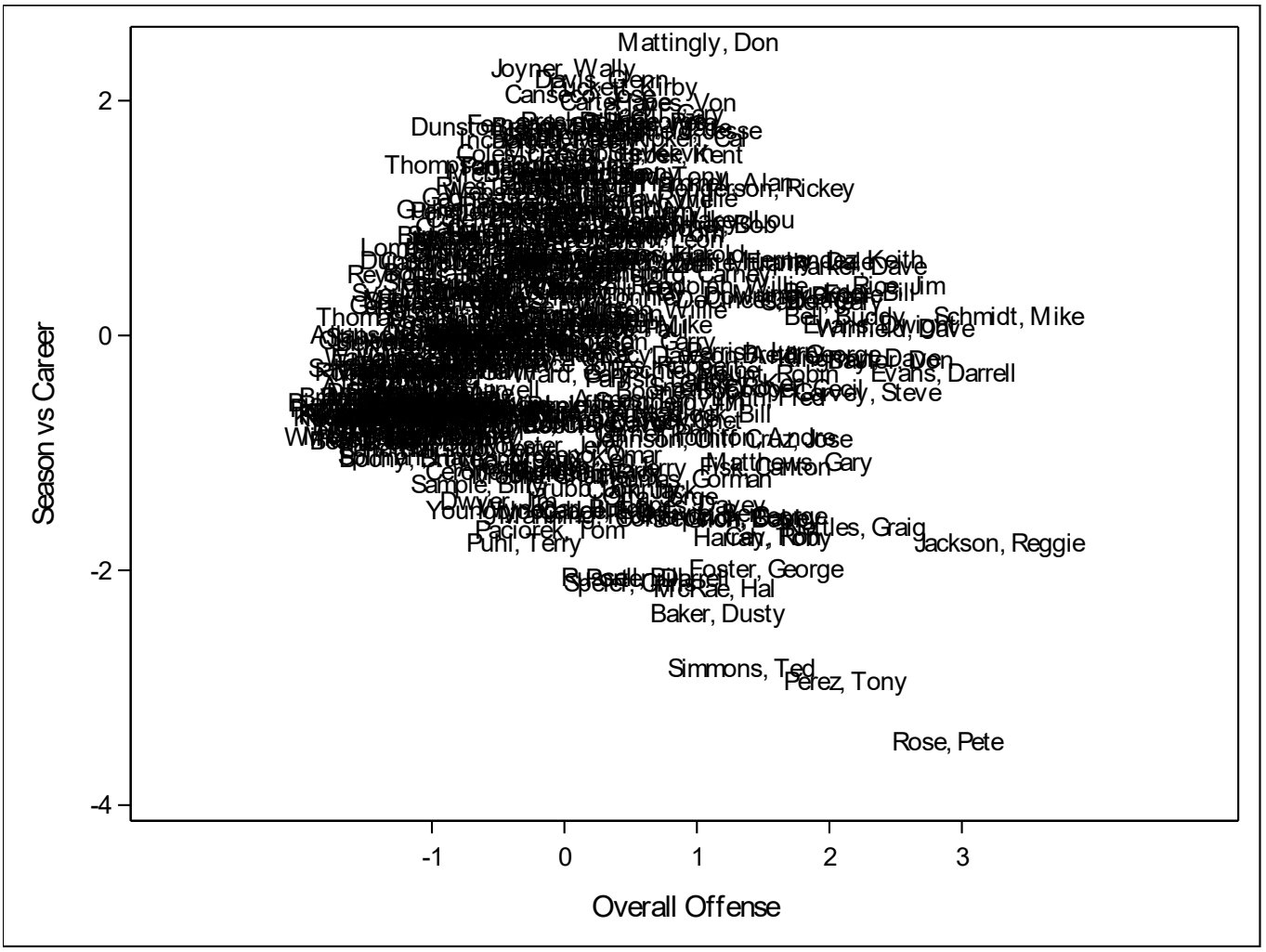





```
data PCout; set PCout;
  label Prin1='Overall Offense'
        Prin2='Season vs Career'
        Prin3='Defense vs Offense'
        Prin4='Outs'
        ;

proc print data=PCout;
  var Name logSalary Prin1-Prin5;
run;
```

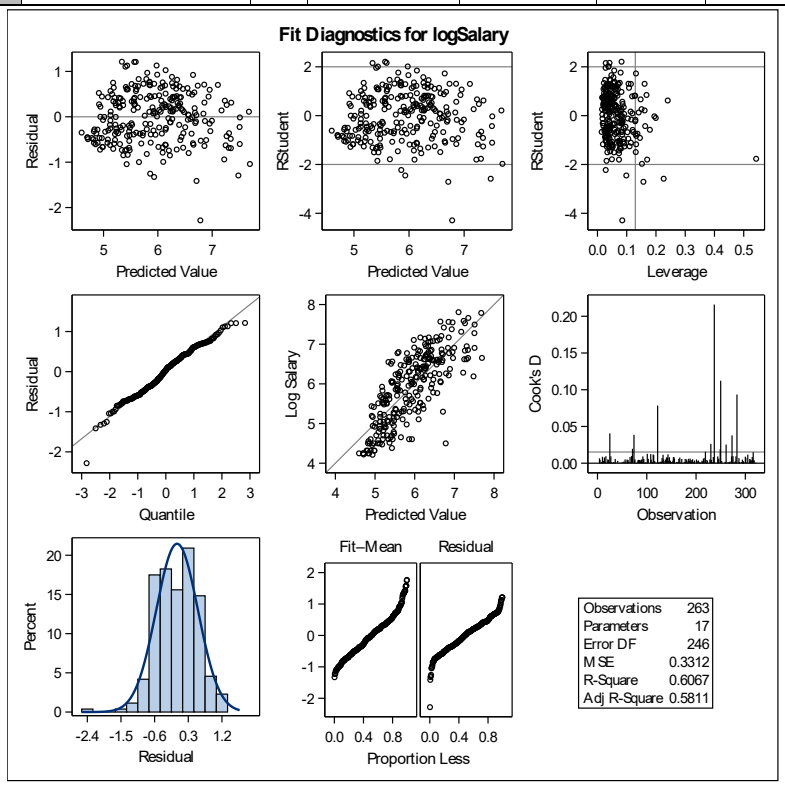
Obs	Name	logSalary	Prin1	Prin2	Prin3	Prin4	Prin5
1	Allanson, Andy	.	-1.27885	0.00278	0.60377	1.26842	1.07501
2	Ashby, Alan	6.16331	0.03647	-0.65768	0.10695	1.78036	0.10054
3	Davis, Alan	6.17379	0.13751	1.27855	-0.51492	1.85527	-0.17923
4	Dawson, Andre	6.21461	1.27982	-0.16949	-0.78343	-0.82638	0.84045
...							

```
proc sgplot data=PCout;
 scatter x=Prin1 y=Prin2 / markerchar=Name
markercharattrs=(size=10);
run;
```




```
/* Principal components regression */
proc reg data=PCout;
  model logSalary = Prin1-Prin16 / vif;
run;
```

Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	Intercept	1	5.90018	0.03603	163.76	<.0001	0
Prin1	Overall Offense	1	0.65596	0.03622	18.11	<.0001	1.01680
Prin2	Season vs Career	1	0.07574	0.03652	2.07	0.0391	1.00894
Prin3	Defense vs Offense	1	0.01048	0.03367	0.31	0.7560	1.00471
Prin4	Outs	1	0.04455	0.03539	1.26	0.2093	1.00666
Prin5		1	-0.10908	0.03529	-3.09	0.0022	1.02209
Prin6		1	-0.10307	0.03473	-2.97	0.0033	1.01054
Prin7		1	0.01280	0.03400	0.38	0.7068	1.00697
Prin8		1	0.07035	0.03679	1.91	0.0570	1.02220
Prin9		1	-0.07326	0.03527	-2.08	0.0388	1.01285
Prin10		1	0.05876	0.03596	1.63	0.1036	1.02031
Prin11		1	0.04691	0.03552	1.32	0.1879	1.01060
Prin12		1	0.03059	0.03409	0.90	0.3704	1.00390
Prin13		1	0.08563	0.03461	2.47	0.0140	1.01154
Prin14		1	0.05602	0.03654	1.53	0.1265	1.01733
Prin15		1	0.01100	0.03592	0.31	0.7596	1.00899
Prin16		1	-0.02885	0.03666	-0.79	0.4321	1.02242



```
/* Quantile regression */
proc quantselect data=baseball plots=coefficients;
 model Salary = nAtBat nHits nHome nRuns nRBI nBB yrMajor
  crAtBat crHits crHome crRuns crRbi  crBB nOuts nAssts nError
   / quantile = 0.1 0.5 0.9 selection=lasso(sh=3);
 partition fraction(validate=0.3);
run;
```

Quantile=0.1

Parameter Estimates

Parameter	DF	Estimate	Standardized Estimate
Intercept	1	-54.330850	0
nHits	1	1.096282	0.104164
nRBI	1	-0.090184	-0.005029
nBB	1	0.593059	0.028266
YrMajor	1	12.100961	0.128828
CrHits	1	-0.017422	-0.025075
CrRuns	1	0.121594	0.091718

Quantile=0.5

Parameter Estimates

Parameter	DF	Estimate	Standardized Estimate
Intercept	1	-106.041638	0
nHits	1	2.984638	0.283588
CrHits	1	0.404477	0.582163

Quantile=0.9

Parameter Estimates

Parameter	DF	Estimate	Standardized Estimate
Intercept	1	-92.518900	0
nRuns	1	5.141574	0.277304
nRBI	1	1.759843	0.098145
nBB	1	3.416767	0.162846
CrHome	1	1.055749	0.210026
CrRuns	1	1.037092	0.782277
CrRbi	1	0.040515	0.029589

```
proc quantreg data=baseball;
 model Salary = nRuns YrMajor CrRuns nHits CrHome CrRbi nOuts /
   quantile= 0.05 to 0.95 by 0.05
   plot=quantplot;
run;
```