

```

# Look at the meta data for the file
> rxGetInfo(songs,getVarInfo=TRUE)
File name: C:\Users\Joseph\Documents\DATA\Million Song\MillionSong_XDF\MS_ABVZ.xdf
Number of observations: 107420
Number of variables: 33
Number of blocks: 4
Variable information:
Var 1: artist.id, Type: character
Var 2: artist.mbid, Type: character
Var 3: artist.playmeid, Type: numeric, Low/High: (-1.0000, 241634.0000)
Var 4: artist.7digitalid, Type: numeric, Low/High: (-1.0000, 817066.0000)
Var 5: artist.name, Type: character
Var 6: artist.latitude, Type: numeric, Low/High: (-51.8006, 70.6958)
Var 7: artist.longitude, Type: numeric, Low/High: (-162.4365, 175.4713)
Var 8: artist.location, Type: character
Var 9: artist.hotttnesss, Type: numeric, Low/High: (0.0000, 1.0825)
Var 10: artist.familiarity, Type: numeric, Low/High: (0.0000, 1.0000)
Var 11: song.id, Type: character
Var 12: track.id, Type: character
Var 13: track.7digitalid, Type: numeric, Low/High: (107.0000, 9090596.0000)
Var 14: release.7digitalid, Type: numeric, Low/High: (17.0000, 823604.0000)
Var 15: audio.md5, Type: character
Var 16: title, Type: character
Var 17: release, Type: character
Var 18: year, Type: numeric, Low/High: (0.0000, 2010.0000)
Var 19: mode, Type: numeric, Low/High: (0.0000, 1.0000)
Var 20: mode.confidence, Type: numeric, Low/High: (0.0000, 1.0000)
Var 21: key, Type: numeric, Low/High: (0.0000, 11.0000)
Var 22: key.confidence, Type: numeric, Low/High: (0.0000, 1.0000)
Var 23: time.signature, Type: numeric, Low/High: (0.0000, 7.0000)
Var 24: time.signature.confidence, Type: numeric, Low/High: (0.0000, 1.0000)
Var 25: analysis.sample.rate, Type: numeric, Low/High: (22050.0000, 22050.0000)
Var 26: start.of.fade.out, Type: numeric, Low/High: (0.5480, 3027.7220)
Var 27: end.of.fade.in, Type: numeric, Low/High: (0.0000, 300.5880)
Var 28: tempo, Type: numeric, Low/High: (0.0000, 284.2080)
Var 29: duration, Type: numeric, Low/High: (0.5481, 3033.5995)
Var 30: loudness, Type: numeric, Low/High: (-56.6610, 3.6340)
Var 31: energy, Type: numeric, Low/High: (0.0000, 0.0000)
Var 32: danceability, Type: numeric, Low/High: (0.0000, 0.0000)
Var 33: song.hotttnesss, Type: numeric, Low/High: (0.0000, 1.0000)

# Look at a summary of the linear model
> summary(mod)
Call:
rxLinMod(formula = form, data = songs, blocksPerRead = 10000)

Linear Regression Results for: duration ~ artist.hotttnesss +
  artist.familiarity + track.7digitalid + release.7digitalid + year +
  mode + mode.confidence + key + key.confidence + time.signature +
  time.signature.confidence + start.of.fade.out + end.of.fade.in +
  tempo + loudness
File name: C:\Users\Joseph\Documents\DATA\Million Song\MillionSong_XDF\MS_ABVZ.xdf
Dependent variable(s): duration
Total independent variables: 16
Number of valid observations: 107382
Number of missing observations: 38

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.148e+00  2.062e-01  15.268 2.22e-16 ***
artist.hotttnesss -8.116e-01  2.890e-01  -2.808 0.004980 **
artist.familiarity  8.848e-01  2.657e-01   3.330 0.000868 ***
track.7digitalid -2.080e-07  1.204e-07  -1.727 0.084248 .
release.7digitalid  8.613e-07  1.340e-06   0.643 0.520318
year          2.300e-04  2.644e-05   8.698 2.22e-16 ***
mode         -2.384e-01  5.455e-02  -4.371 1.24e-05 ***
mode.confidence  2.642e+00  2.052e-01  12.878 2.22e-16 ***
key           9.710e-04  7.064e-03   0.137 0.890673
key.confidence  1.390e-02  1.436e-01   0.097 0.922909
time.signature  9.461e-02  2.167e-02   4.366 1.27e-05 ***
time.signature.confidence  3.766e-01  7.151e-02   5.266 1.40e-07 ***
start.of.fade.out  1.014e+00  2.056e-04  4934.245 2.22e-16 ***
end.of.fade.in  2.301e-01  1.009e-02  22.807 2.22e-16 ***
tempo        -7.379e-03  7.322e-04  -10.079 2.22e-16 ***
loudness     -1.502e-01  4.989e-03  -30.113 2.22e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.156 on 107366 degrees of freedom
Multiple R-squared:  0.9958
Adjusted R-squared:  0.9958
F-statistic: 1.686e+06 on 15 and 107366 DF,  p-value: < 2.2e-16
Condition number: 835.6625

# Look at the results of the stepwise regression
> system.time(rxlm.step <- rxLinMod(form, data = songs,
+                                 blocksPerRead=100000,
+                                 variableSelection = varsel,
+                                 verbose = 1,
+                                 dropMain = FALSE,
+                                 coefLabelStyle = "R"))
Rows Read: 107420, Total Rows Processed: 107420, Total Chunk Time: 0.015 seconds

```

Start: AIC=450745

duration ~ artist.hotttness + artist.familiarity + track.7digitalid + release.7digitalid + year + mode + mode.confidence + key + key.confidence + time.signature + time.signature.confidence + start.of.fade.out + end.of.fade.in + tempo + loudness

DataSet[15, 6]

Row	StepMethod	Step	Df	Sum of Sq	RSS	AIC
[1,.]	-1	key.confidence	1	0.6229	7141486.9076	450742.7197
[2,.]	-1	key	1	1.2567	7141487.5414	450742.7293
[3,.]	-1	release.7digitalid	1	27.4885	7141513.7732	450743.1237
[4,.]	0	<none>	NA	NA	7141486.2847	450744.7104
[5,.]	-1	track.7digitalid	1	198.2853	7141684.5699	450745.6918
[6,.]	-1	artist.hotttness	1	524.6108	7142010.8954	450750.5983
[7,.]	-1	artist.familiarity	1	737.6291	7142223.9137	450753.8011
[8,.]	-1	time.signature	1	1267.9401	7142754.2247	450761.7739
[9,.]	-1	mode	1	1270.6709	7142756.9555	450761.8150
[10,.]	-1	time.signature.confidence	1	1844.6337	7143330.9184	450770.4434
[11,.]	-1	year	1	5031.8840	7146518.1687	450818.3450
[12,.]	-1	tempo	1	6756.8811	7148243.1658	450844.2613
[13,.]	-1	mode.confidence	1	11031.4180	7152517.7027	450908.4548
[14,.]	-1	end.of.fade.in	1	34598.6881	7176084.9727	451261.6928
[15,.]	-1	start.of.fade.out	1	1619433677.0857	1626575163.3704	1033645.6063

Step 1: AIC=450743

duration ~ artist.familiarity + artist.hotttness + end.of.fade.in + key + loudness + mode + mode.confidence + release.7digitalid + start.of.fade.out + tempo + time.signature + time.signature.confidence + track.7digitalid + year

DataSet[14, 6]

Row	StepMethod	Step	Df	Sum of Sq	RSS	AIC
[1,.]	-1	key	1	1.2218	7141488.1294	450740.7381
[2,.]	-1	release.7digitalid	1	27.4462	7141514.3538	450741.1324
[3,.]	0	<none>	NA	NA	7141486.9076	450742.7197
[4,.]	-1	track.7digitalid	1	198.1807	7141685.0882	450743.6996
[5,.]	-1	artist.hotttness	1	524.5906	7142011.4981	450748.6074
[6,.]	-1	artist.familiarity	1	737.5030	7142224.4106	450751.8085
[7,.]	-1	time.signature	1	1269.6637	7142756.5712	450759.8092
[8,.]	-1	mode	1	1271.0649	7142757.9724	450759.8302
[9,.]	-1	time.signature.confidence	1	1844.1855	7143331.0930	450768.4460
[10,.]	-1	year	1	5031.2726	7146518.1802	450816.3452
[11,.]	-1	tempo	1	6772.4943	7148259.4019	450842.5052
[12,.]	-1	mode.confidence	1	27536.9169	7169023.8245	451153.9786
[13,.]	-1	end.of.fade.in	1	34598.1197	7176085.0272	451259.6936
[14,.]	-1	start.of.fade.out	1	1624179486.0477	1631320972.9552	1033956.4554

Step 2: AIC=450741

duration ~ artist.familiarity + artist.hotttness + end.of.fade.in + loudness + mode + mode.confidence + release.7digitalid + start.of.fade.out + tempo + time.signature + time.signature.confidence + track.7digitalid + year

DataSet[14, 6]

Row	StepMethod	Step	Df	Sum of Sq	RSS	AIC
[1,.]	-1	release.7digitalid	1	27.4442	7141515.5736	450739.1508
[2,.]	0	<none>	NA	NA	7141488.1294	450740.7381
[3,.]	-1	track.7digitalid	1	198.1725	7141686.3019	450741.7179
[4,.]	1	key.confidence	1	0.5880	7141487.5414	450742.7293
[5,.]	-1	artist.hotttness	1	524.6142	7142012.7436	450746.6261
[6,.]	-1	artist.familiarity	1	737.3917	7142225.5211	450749.8252
[7,.]	-1	time.signature	1	1269.6840	7142757.8134	450757.8279
[8,.]	-1	mode	1	1338.1752	7142826.3047	450758.8575
[9,.]	-1	time.signature.confidence	1	1844.5312	7143332.6606	450766.4696
[10,.]	-1	year	1	5030.8951	7146519.0245	450814.3578
[11,.]	-1	tempo	1	6771.3002	7148259.4296	450840.5056
[12,.]	-1	mode.confidence	1	27535.7626	7169023.8920	451151.9796
[13,.]	-1	end.of.fade.in	1	34596.9575	7176085.0869	451257.6945
[14,.]	-1	start.of.fade.out	1	1624441188.6894	1631582676.8188	1033971.6807

Step 3: AIC=450739

duration ~ artist.familiarity + artist.hotttness + end.of.fade.in + loudness + mode + mode.confidence + start.of.fade.out + tempo + time.signature + time.signature.confidence + track.7digitalid + year

DataSet[14, 6]

Row	StepMethod	Step	Df	Sum of Sq	RSS	AIC
[1,.]	0	<none>	NA	NA	7141515.5736	450739.1508
[2,.]	1	key	1	1.2199	7141514.3538	450741.1324
[3,.]	1	key.confidence	1	0.5470	7141515.0266	450741.1426
[4,.]	-1	artist.hotttness	1	528.4498	7142044.0235	450745.0964
[5,.]	-1	artist.familiarity	1	734.8160	7142250.3896	450748.1991
[6,.]	-1	time.signature	1	1269.3804	7142784.9541	450756.2359
[7,.]	-1	mode	1	1339.8532	7142855.4268	450757.2953
[8,.]	-1	time.signature.confidence	1	1843.1049	7143358.6786	450764.8607
[9,.]	-1	year	1	5020.6437	7146536.2173	450812.6162
[10,.]	-1	tempo	1	6765.9667	7148281.5404	450838.8378
[11,.]	-1	track.7digitalid	1	12587.2228	7154102.7964	450926.2495
[12,.]	-1	mode.confidence	1	27529.0970	7169044.6707	451150.2909
[13,.]	-1	end.of.fade.in	1	34593.8615	7176109.4351	451256.0588
[14,.]	-1	start.of.fade.out	1	1624515870.8133	1631657386.3869	1033974.5976

Rows Read: 107420, Total Rows Processed: 107420, Total Chunk Time: 0.015 seconds

Elapsed computation time: 0.135 secs.

Regression Results for: duration ~ artist.familiarity + artist.hotttness + end.of.fade.in + loudness + mode + mode.confidence + start.of.fade.out + tempo + time.signature + time.signature.confidence + track.7digitalid + year
Model Name: duration ~ artist.familiarity + artist.hotttness + end.of.fade.in + loudness + mode + mode.confidence +

start.of.fade.out + tempo + time.signature + time.signature.confidence + track.7digitalid + year
Description: duration ~ artist.familiarity + artist.hotttnesss + end.of.fade.in + loudness + mode + mode.confidence +
start.of.fade.out + tempo + time.signature + time.signature.confidence + track.7digitalid + year

Dependent Variable: duration

Total independent variables: 13

Number of valid observations: 107382 (Number excluded for missings: 38)

Coeffs.	Value	Std. Error	t Value	Pr(> t)
(Intercept)	3.1568	0.2006	15.7337	0.0000
artist.familiarity	0.8831	0.2657	3.3238	0.0009
artist.hotttnesss	-0.8145	0.2890	-2.8187	0.0048
end.of.fade.in	0.2300	0.0101	22.8057	0.0000
loudness	-0.1502	0.0050	-30.1322	0.0000
mode	-0.2397	0.0534	-4.4882	0.0000
mode.confidence	2.6572	0.1306	20.3442	0.0000
start.of.fade.out	1.0144	0.0002	4942.0398	0.0000
tempo	-0.0074	0.0007	-10.0858	0.0000
time.signature	0.0945	0.0216	4.3686	0.0000
time.signature.confidence	0.3764	0.0715	5.2640	0.0000
track.7digitalid	-0.0000	0.0000	-13.7565	0.0000
year	0.0002	0.0000	8.6881	0.0000

R-Squared: 0.9958

Residual standard error: 8.1556 (107369 d.f.)

F-statistic: 2107077.4664 (12 and 107369 d.f.), p-value: 0.0000

Condition number: 4.4711

user	system	elapsed
0.20	0.01	0.16