

# Linear\_Regression.R

## Kaustubh Chalke

```
> #Importing the Hepatitis c Dataset
> HCV <- read.csv("C:/Users/Kaustubh Chalke/Documents/Multivariate Analysis/Project/HCV-Egy-Data.csv")
> library(dplyr)
> attach(HCV)
> Survivorship =HCV$Survivorship <- if_else( RNA.EOT>= 400000 , 0,1)
> cbind(data.frame(Survivorship),HCV)
```

	Survivorship	Age	Gender	BMI	Fever	Nausea.Vomting	Headache	Diarrhea	Fatigue...generalized.bone.
ache Jaundice Epigastric.pain									
1	1	56	1	35	2		1	1	1
2	2		2						
2	1	46	1	29	1		2	2	1
2	2		1						
3	0	57	1	33	2		2	2	2
1	1		1						
4	0	49	2	33	1		2	1	2
1	2		1						
5	1	59	1	32	1		1	2	1
2	2		2						
6	1	58	2	22	2		2	2	1
2	2		1						
7	1	42	2	26	1		1	2	2
2	2		2						
8	1	48	2	30	1		1	2	2
1	1		2						
9	1	44	1	23	1		1	2	2
2	1		2						
10	1	45	1	30	2		1	2	2
1	1		2						
11	0	37	2	24	2		1	2	1
2	2		1						
12	1	36	1	22	2		2	1	1
1	1		1						
13	1	45	2	25	2		1	1	1
2	1		2						
14	1	34	1	22	1		2	1	1
2	2		1						
15	0	40	2	32	2		2	2	1
2	1		1						
16	0	58	1	34	2		1	1	1
2	1		1						
17	1	61	1	35	1		2	2	2
1	1		2						
18	1	55	2	24	2		1	2	2
2	2		2						
19	0	56	1	27	1		2	2	2
2	2		2						
20	0	35	2	23	2		2	1	1
1	1		2						
21	0	57	2	23	1		1	2	2
1	1		1						
22	1	33	1	25	2		1	2	2
2	2		2						
23	1	41	1	23	1		2	2	2
2	1		2						
24	1	39	2	29	1		2	1	2
1	1		2						
25	1	33	2	24	1		2	2	2
2	1		2						
26	1	43	2	34	2		2	2	1
1	1		1						
27	0	51	1	34	2		1	2	2
1	1		2						
28	0	39	2	33	2		1	2	1
1	1		2						
29	1	57	2	26	1		2	2	1
1	1		2						
30	1	47	2	29	1		1	2	1
2	1		2						

	WBC	RBC	HGB	Plat	AST.1	ALT.1	ALT4	ALT.12	ALT.24	ALT.36	ALT.48	ALT.after.24.w	RNA.Base
31		1	55	2	33	1		2	2	1			
2	1			2									
32		1	58	2	35	2		2	2	2			
1	1			1									
1	7425	4248807	14	112132	99	84	52	109	81	5	5		5 655330
634536	288194		5	5									
2	12101	4429425	10	129367	91	123	95	75	113	57	123		44 40620
538635	637056	336804	31085										
3	4178	4621191	12	151522	113	49	95	107	116	5	5		5 571148
661346		5	735945	558829									
4	6490	4794631	10	146457	43	64	109	80	88	48	77		33 1041941
449939	585688	744463	582301										
5	3661	4606375	11	187684	99	104	67	48	120	94	90		30 660410
738756	3731527	338946	242861										
6	11785	3882456	15	131228	66	104	121	96	65	73	114		29 1157452
1086852		5	5	5									
7	11620	4747333	12	177261	78	57	113	118	107	84	80		28 325694
1034008	275095	214566	635157										
8	7335	4405941	11	216176	119	112	80	127	45	96	53		39 641129
72050	787295	370605	506296										
9	10480	4608464	12	148889	93	83	55	102	97	122	39		45 591441
757361		5	371090	203042									
10	6681	4455329	12	98200	55	68	72	127	81	125	43		30 1151206
230488	267320	275295	555516										
11	4437	4265042	12	166027	103	124	111	74	53	123	101		33 1023123
103190	731929	448466	59998										
12	6052	4130219	13	144266	75	49	93	52	46	46	59		45 137712
1122999	561438	63145	806204										
13	9279	4116937	13	203003	97	101	66	53	95	55	104		26 936444
536969		5	5	5									
14	5638	4321603	14	141110	120	61	64	51	78	90	113		23 392976
884322	586834	182775	782154										
15	11507	4165603	14	222874	127	122	106	105	88	111	111		36 1133727
1111871	421304	437544	124609										
16	8035	4896464	11	149506	117	53	50	80	120	66	86		34 614951
314296	83690	671490	135145										
17	10843	4165219	10	197640	86	105	70	86	83	87	47		33 900099
721460		5	5	5									
18	8476	4466885	14	163276	53	101	50	95	112	97	68		27 1145310
230993	457882	318363	256415										
19	6599	4448466	15	190642	53	124	62	76	57	46	93		26 506756
359181	743399	405175	162983										
20	4845	4436025	10	111819	115	121	63	127	95	124	93		42 1080499
76404	717159	404314	477719										
21	5925	4031637	15	116558	86	109	118	119	55	103	84		32 169624
786017	669076	531187	282524										
22	9952	4994729	10	109023	84	77	67	81	117	68	42		32 1135200
572747		5	5	5									
23	7961	4595487	14	94733	45	92	103	104	40	115	93		33 293380
440576	53098	18292	187341										
24	7136	4625248	10	211363	70	102	76	58	111	95	58		25 993940
992652	96482	334897	762760										
25	6057	4300774	11	222135	62	91	116	128	41	70	106		43 243433
981370	12504	360015	753383										
26	6648	4529290	15	109871	48	112	99	85	59	87	78		35 955296
540654		5	5	5									
27	11032	4052583	15	94503	41	54	128	64	71	89	87		34 766355
531269	737603	734863	372837										
28	5234	4906158	12	190314	61	120	113	75	88	114	99		43 486467
45990	45578	733292	19572										
29	6038	4763261	13	126721	51	118	98	42	93	53	83		45 285374
186657		5	5	5									
30	5846	4753531	15	104729	120	72	117	126	45	95	49		38 426136
247777	767015	377123	315150										
31	5383	3999388	15	182262	96	49	59	88	62	58	81		41 1194301
928679	29778	124250	244049										
32	7378	3998925	10	201114	57	110	128	96	69	105	72		26 557708
287714	623587	66891	35044										

```

> HCV_lr = select(HCV, RNA.12, RNA.EF, RNA.EOT, RNA.Base)
> #Multiple Regression
> view(HCV)
> # Performing multiple regression on HCV dataset
> fit <- lm(Survivorship~RNA.EF+RNA.EOT+RNA.Base+RNA.12,data=HCV)
> #show the results
> summary(fit)

```

```

Call:
lm(formula = Survivorship ~ RNA.EF + RNA.EOT + RNA.Base + RNA.12,

```

```

data = HCV)

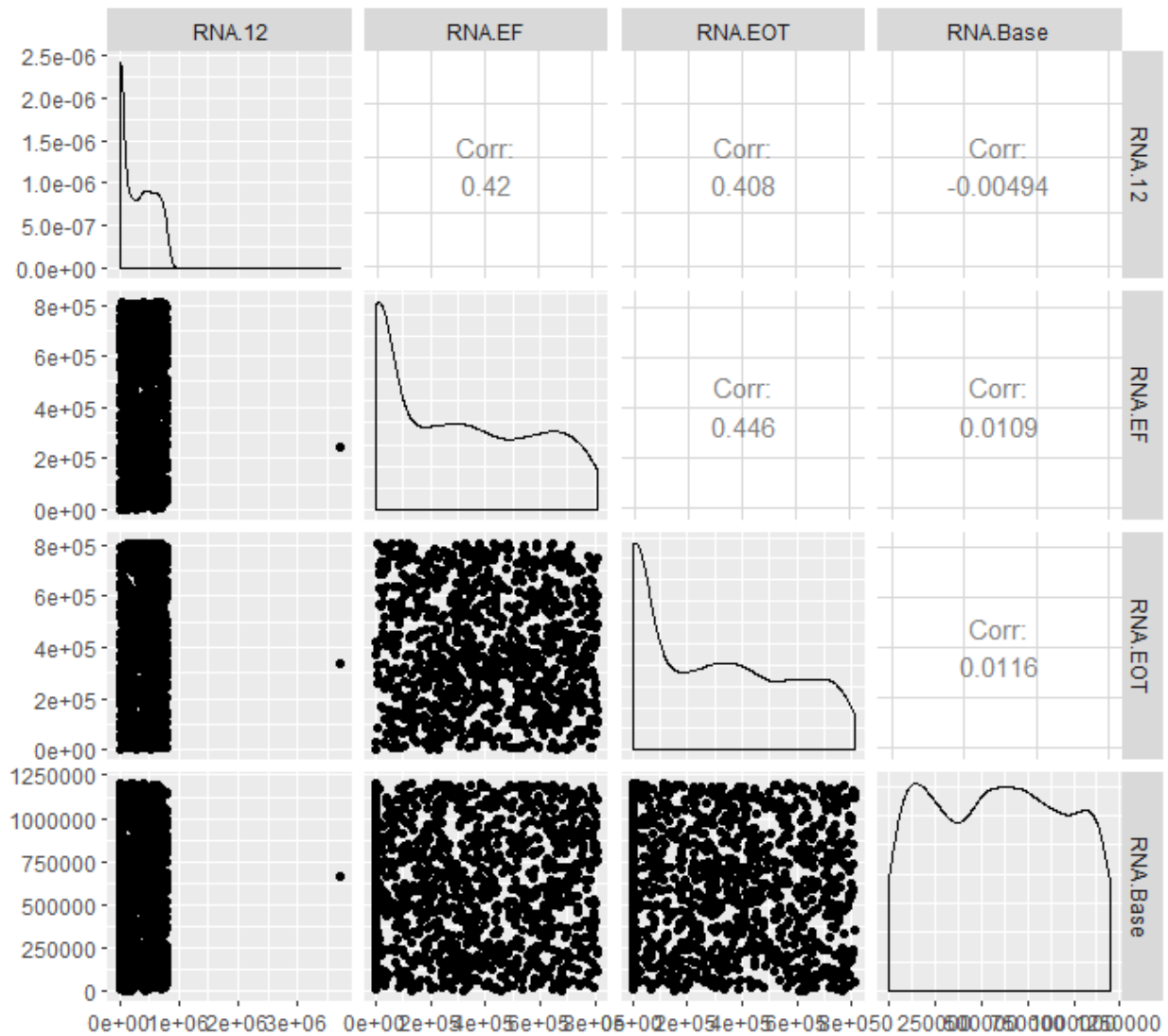
Residuals:
    Min       1Q   Median       3Q      Max
-0.55458 -0.08468 -0.05481  0.13407  0.59337

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.068e+00  1.453e-02  73.482  < 2e-16 ***
RNA.EF       1.451e-07  2.682e-08   5.409  7.47e-08 ***
RNA.EOT      -1.687e-06  2.697e-08 -62.531  < 2e-16 ***
RNA.Base     -1.384e-08  1.737e-08  -0.797    0.426
RNA.12       1.224e-07  2.467e-08   4.963  7.80e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2287 on 1380 degrees of freedom
Multiple R-squared:  0.7703, Adjusted R-squared:  0.7696
F-statistic: 1157 on 4 and 1380 DF, p-value: < 2.2e-16
> #Summary has three sections. Section1: How well does the model fit the data (before Coefficient
> s). Section2: Is the hypothesis supported? (until signif codes). Section3: How well does data fit
> the model (again).
> # Useful Helper Functions
> coefficients(fit)
            (Intercept)      RNA.EF      RNA.EOT      RNA.Base      RNA.12
1.067746e+00  1.450885e-07 -1.686741e-06 -1.383850e-08  1.224221e-07
> #install.packages("GGally", lib="/Library/Frameworks/R.framework/Versions/3.5/Resources/library")
> library(GGally)
> ggpairs(data=HCV_lr, title="HCV Data")

```

## HCV Data



```
> confint(fit, level=0.95)
              2.5 %          97.5 %
(Intercept) 1.039242e+00 1.096251e+00
RNA.EF       9.246772e-08 1.977093e-07
RNA.EOT      -1.739656e-06 -1.633825e-06
RNA.Base     -4.791802e-08 2.024102e-08
RNA.12       7.403556e-08 1.708087e-07
> # Predicted values
> fitted(fit)
      1      2      3      4      5      6      7
8 1.093951141 0.581583066 -0.100425497 -0.046202140 0.978951116 1.051721814 0.827153724 0.603
599640 0.463088785
17 0.700788986 0.395451379 1.145034724 1.054780232 0.939337139 0.383690303 -0.043539588 1.055
283193 0.608158926
19 20 21 22 23 24 25
26 0.491963928 0.527928690 0.292325027 1.052029748 1.066513908 0.611586573 0.567963654 1.054
519349 -0.037988796
```

```

      28      29      30      31      32      33      34
35      36
-0.167439567 1.063790063 0.565364837 0.880695657 1.028626040 -0.011389086 0.506226136 0.011
654539 0.213029450
      37      38      39      40      41      42      43
44      45
0.597629258 0.737527908 1.005760078 0.793516211 0.267095793 0.588765861 0.794159304 -0.016
029038 1.009401378
> residuals(fit)
      1      2      3      4      5      6      7
8      9
-0.093951141 0.418416934 0.100425497 0.046202140 0.021048884 -0.051721814 0.172846276 0.396
400360 0.536911215
      10      11      12      13      14      15      16
17      18
0.299211014 -0.395451379 -0.145034724 -0.054780232 0.060662861 -0.383690303 0.043539588 -0.055
283193 0.391841074
      19      20      21      22      23      24      25
26      27
-0.491963928 -0.527928690 -0.292325027 -0.052029748 -0.066513908 0.388413427 0.432036346 -0.054
519349 0.037988796
      28      29      30      31      32      33      34
35      36
0.167439567 -0.063790063 0.434635163 0.119304343 -0.028626040 0.011389086 0.493773864 -0.011
654539 -0.213029450
      37      38      39      40      41      42      43
44      45
0.402370742 0.262472092 -0.005760078 0.206483789 -0.267095793 0.411234139 0.205840696 0.016
029038 -0.009401378
      46      47      48      49      50      51      52
53      54
0.356215266 -0.022662356 -0.237126633 0.073368272 0.087265310 0.113395942 0.427698389 -0.145
781416 -0.032229682
      55      56      57      58      59      60      61
62      63
-0.058694243 -0.077549428 0.193079578 -0.201812452 -0.020263805 0.004195303 0.353194003 0.241
011417 -0.389892131
      64      65      66      67      68      69      70
71      72
0.058767072 0.048147108 0.170029600 0.452182189 -0.010218911 0.199892619 -0.483167144 0.429
041016 0.265917721
      73      74      75      76      77      78      79
80      81
0.478695135 -0.498119095 -0.377537795 0.147279156 0.341947206 -0.416317016 0.203756691 0.515
857719 0.178512111
      82      83      84      85      86      87      88
89      90
0.157948751 -0.439036477 0.224930967 -0.412856208 0.349722377 -0.047544726 -0.443753655 -0.294
019820 0.399434327
      91      92      93      94      95      96      97
98      99
0.511166604 -0.202967827 -0.058981268 0.128043944 -0.279917451 -0.061377944 0.100981119 0.090
908069 -0.057401881
> #Anova Table
> anova(fit)
Analysis of Variance Table

Response: Survivorship
      Df Sum Sq Mean Sq F value Pr(>F)
RNA.EF 1 29.323 29.323 560.5988 < 2.2e-16 ***
RNA.EOT 1 211.422 211.422 4041.9169 < 2.2e-16 ***
RNA.Base 1 0.039 0.039 0.7417 0.3893
RNA.12 1 1.289 1.289 24.6336 7.796e-07 ***
Residuals 1380 72.184 0.052
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> vcov(fit)
      (Intercept)      RNA.EF      RNA.EOT      RNA.Base      RNA.12
(Intercept) 2.111404e-04 -8.223385e-11 -8.432992e-11 -1.771662e-10 -7.084659e-11
RNA.EF -8.223385e-11 7.195422e-16 -2.398368e-16 -4.605975e-18 -1.929385e-16
RNA.EOT -8.432992e-11 -2.398368e-16 7.276145e-16 -5.055297e-18 -1.804488e-16
RNA.Base -1.771662e-10 -4.605975e-18 -5.055297e-18 3.018065e-16 5.574184e-18
RNA.12 -7.084659e-11 -1.929385e-16 -1.804488e-16 5.574184e-18 6.084031e-16
> cov2cor(vcov(fit))
      (Intercept)      RNA.EF      RNA.EOT      RNA.Base      RNA.12
(Intercept) 1.0000000 -0.210977843 -0.21515200 -0.701828333 -0.19766857
RNA.EF -0.2109778 1.000000000 -0.33146446 -0.009883919 -0.29160511
RNA.EOT -0.2151520 -0.331464455 1.000000000 -0.010787772 -0.27121123
RNA.Base -0.7018283 -0.009883919 -0.01078777 1.000000000 0.01300832
RNA.12 -0.1976686 -0.291605109 -0.27121123 0.013008324 1.00000000
> temp <- influence.measures(fit)

```

```

> temp
Influence measures of
lm(formula = Survivorship ~ RNA.EF + RNA.EOT + RNA.Base + RNA.12, data = HCV) :

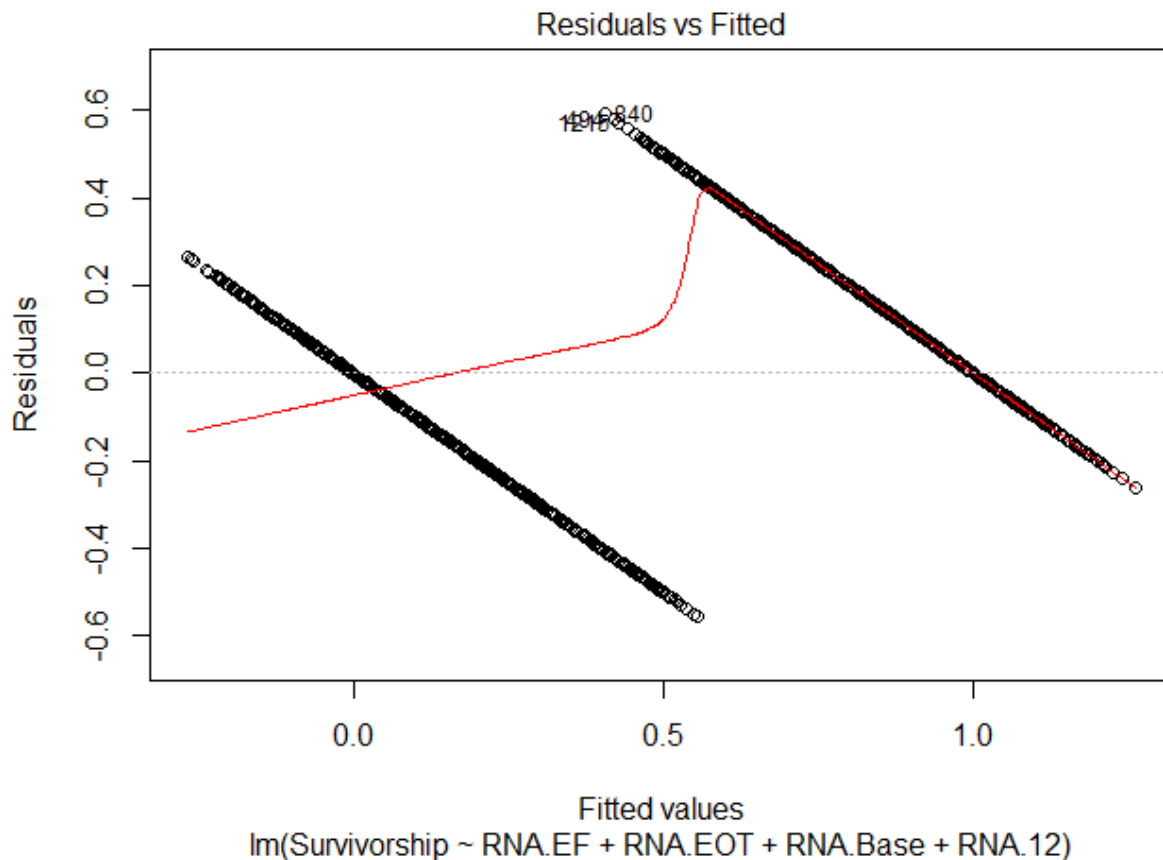
```

	dfb.1	dfb.RNA.EF	dfb.RNA.EO	dfb.RNA.B	dfb.RNA.1	dffit	cov.r	cook.d	hat	inf
1	-0.009242	9.45e-03	0.009317	-2.30e-03	-7.89e-03	-0.01974	1.005	7.80e-05	0.00230	
2	0.070830	-7.91e-02	0.011375	-7.56e-02	8.17e-02	0.13596	0.997	3.69e-03	0.00545	
3	0.000257	1.01e-02	0.022491	-1.23e-03	-2.41e-02	0.03375	1.009	2.28e-04	0.00584	
4	-0.007665	1.34e-03	0.006796	6.85e-03	1.60e-03	0.01308	1.008	3.43e-05	0.00416	
5	-0.007033	-1.24e-02	-0.009916	1.08e-03	3.97e-02	0.03979	1.165	3.17e-04	0.13877	*
6	-0.000415	3.24e-03	0.003318	-9.83e-03	2.59e-03	-0.01422	1.007	4.05e-05	0.00393	
7	0.014511	3.32e-02	-0.016197	-1.55e-02	-8.46e-03	0.04221	1.005	3.56e-04	0.00310	
8	-0.016288	1.09e-02	-0.022960	7.23e-03	7.62e-02	0.09722	0.996	1.89e-03	0.00312	
9	0.041378	-1.07e-02	0.051171	-8.76e-04	-7.25e-02	0.10381	0.986	2.15e-03	0.00194	*
10	-0.031871	4.17e-02	-0.015180	5.55e-02	-1.37e-02	0.07817	1.001	1.22e-03	0.00354	
11	0.033901	8.30e-02	-0.025488	-5.83e-02	-8.88e-02	-0.13291	0.999	3.53e-03	0.00583	
12	-0.014484	-3.89e-02	0.034561	2.19e-02	-1.18e-02	-0.05555	1.010	6.17e-04	0.00756	
13	-0.003268	3.39e-03	0.003466	-6.37e-03	2.79e-03	-0.01260	1.006	3.18e-05	0.00275	
14	0.001621	1.40e-02	-0.010662	-4.01e-03	4.94e-03	0.01895	1.008	7.19e-05	0.00506	
15	0.033820	5.05e-02	-0.033425	-6.98e-02	-2.65e-02	-0.10098	0.997	2.04e-03	0.00359	
16	0.001641	-5.14e-03	0.010957	2.35e-04	-5.55e-03	0.01276	1.008	3.26e-05	0.00445	
17	-0.003767	3.41e-03	0.003490	-5.76e-03	2.83e-03	-0.01239	1.006	3.07e-05	0.00261	
18	-0.037327	-1.90e-02	-0.000724	7.28e-02	3.27e-02	0.09253	0.996	1.71e-03	0.00290	
19	-0.013742	7.33e-02	-0.012156	1.24e-02	-1.07e-01	-0.13290	0.991	3.52e-03	0.00378	
20	0.072959	-8.02e-03	0.014894	-8.68e-02	-8.49e-02	-0.14289	0.988	4.07e-03	0.00379	*
99	-0.005962	3.51e-03	0.003591	-3.06e-03	2.97e-03	-0.01178	1.006	2.78e-05	0.00220	
100	0.033978	-8.45e-03	-0.003932	-2.74e-02	3.14e-04	0.03582	1.004	2.57e-04	0.00253	

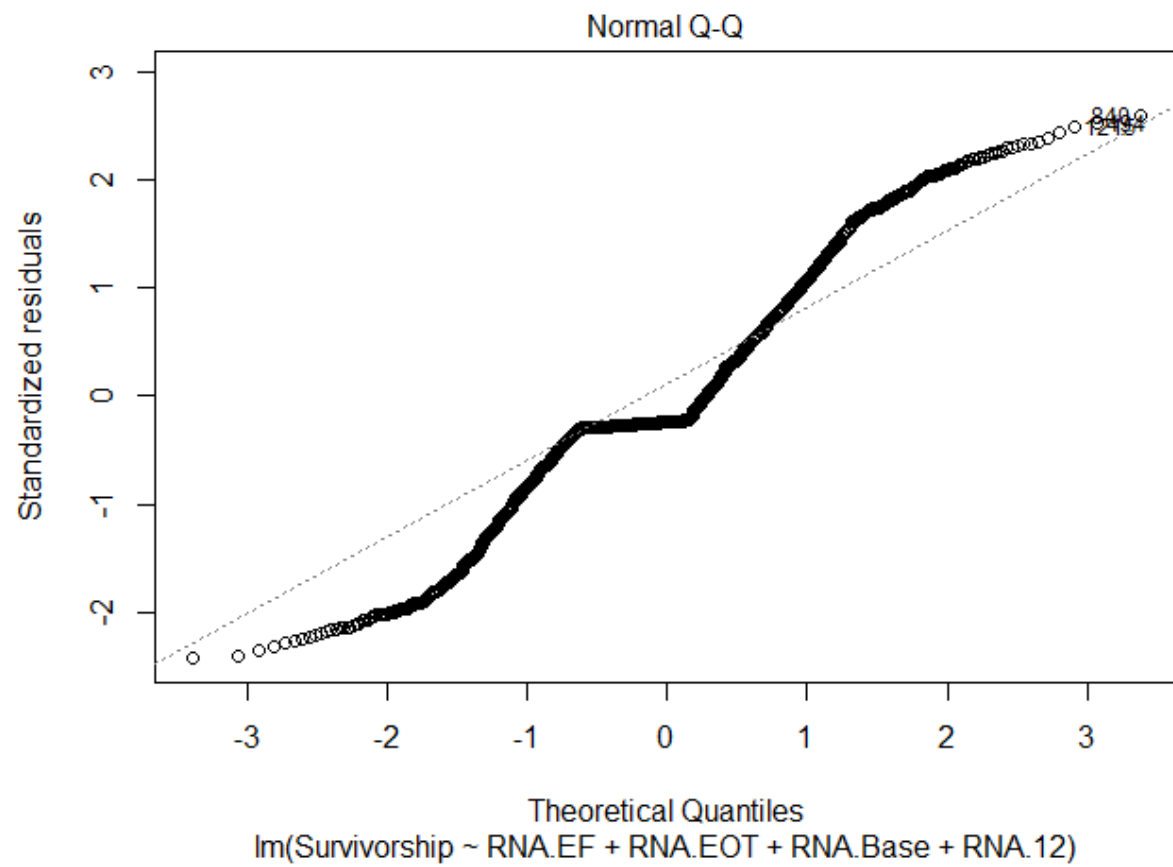
```

[ reached 'max' / getOption("max.print") -- omitted 1285 rows ]
> view(temp)
> #diagnostic plots
> plot(fit)
Hit <Return> to see next plot:

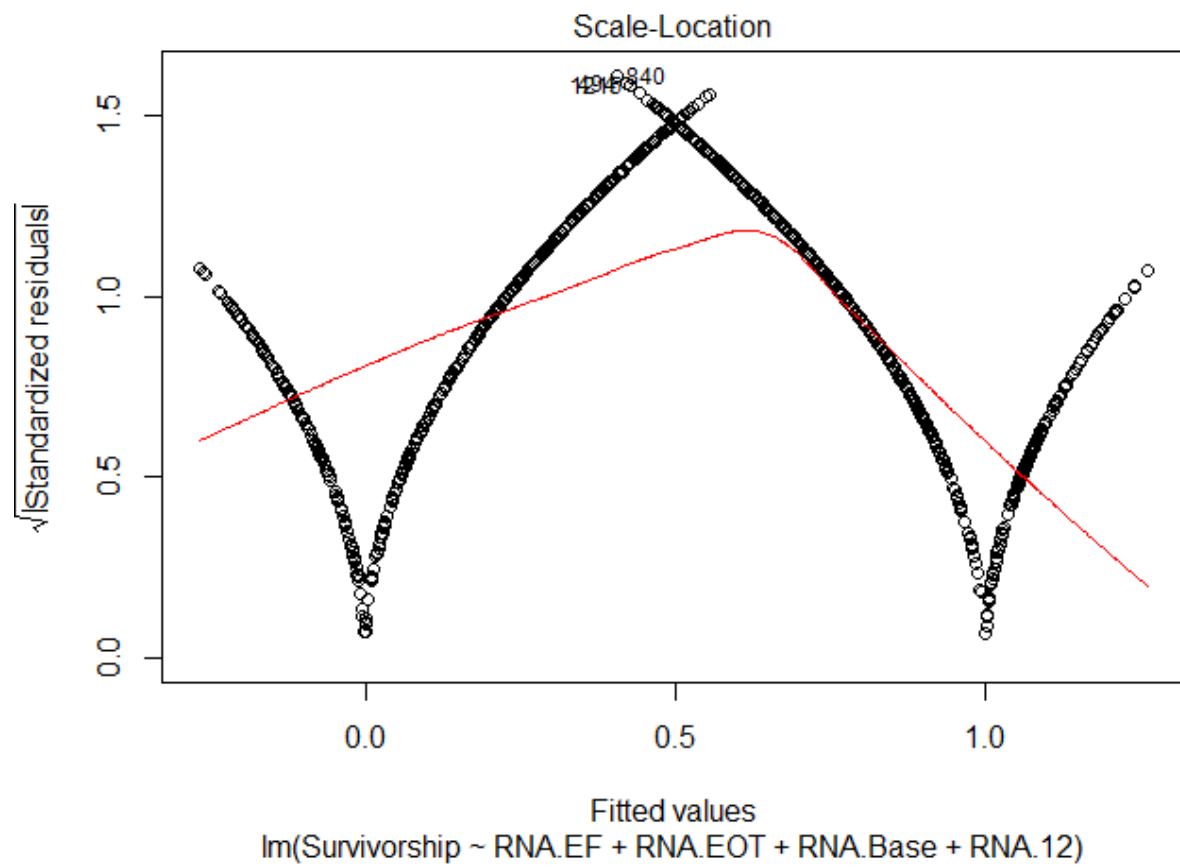
```



Hit <Return> to see next plot:

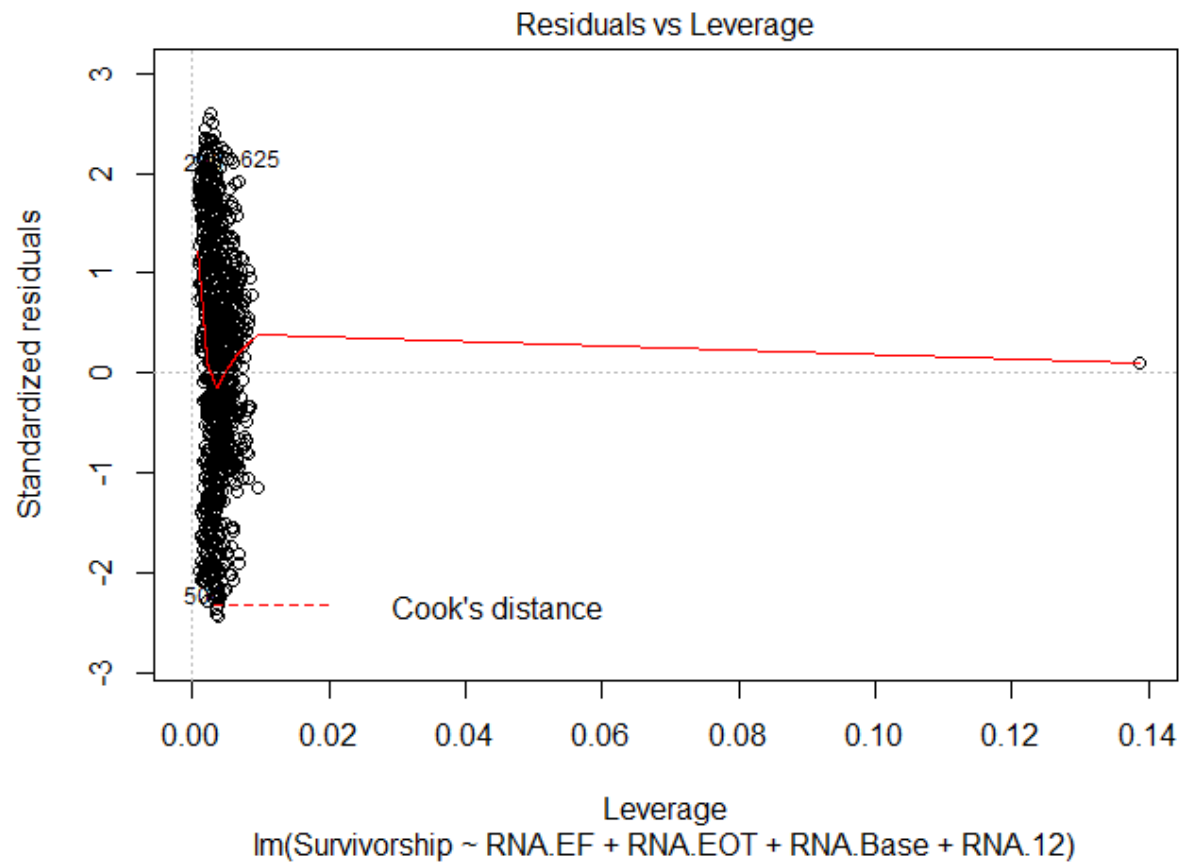


Hit <Return> to see next plot:

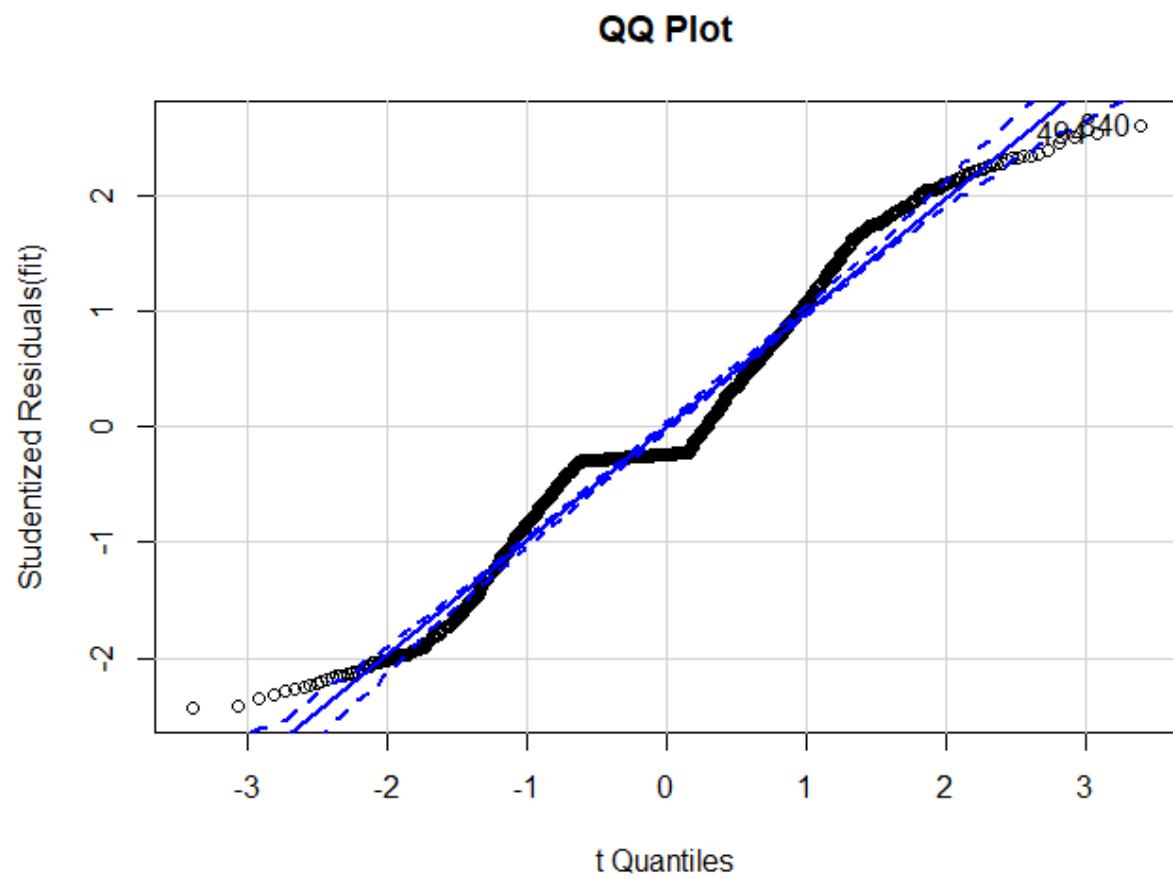


Hit <Return> to see next plot:



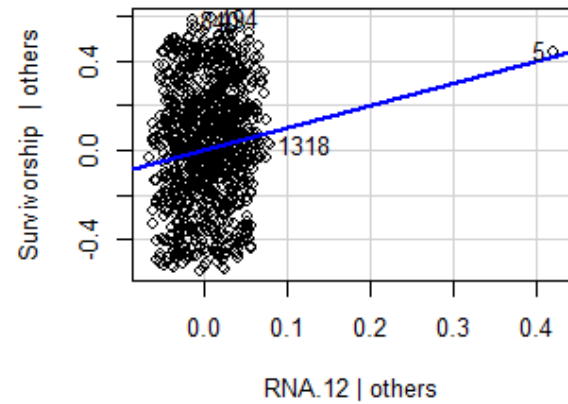
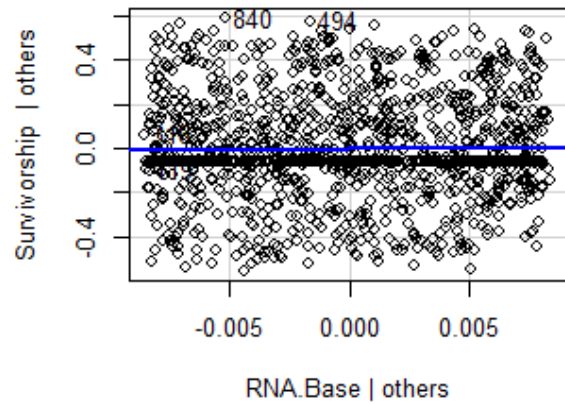
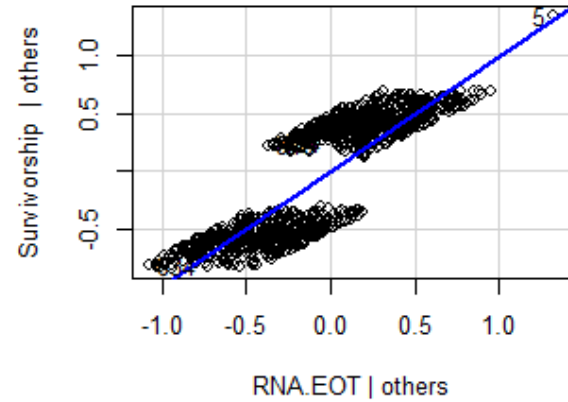
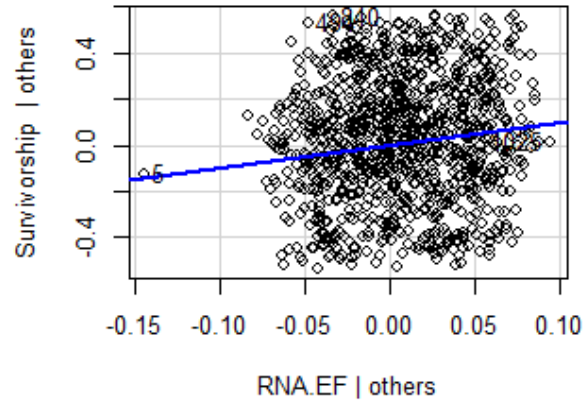


```
> library(car)
> # Assessing Outliers
> outlierTest(fit)
No Studentized residuals with Bonferroni p < 0.05
Largest |rstudent|:
      rstudent unadjusted p-value Bonferroni p
840  2.603482          0.0093273          NA
> qqPlot(fit, main="QQ Plot")
[1] 494 840
```

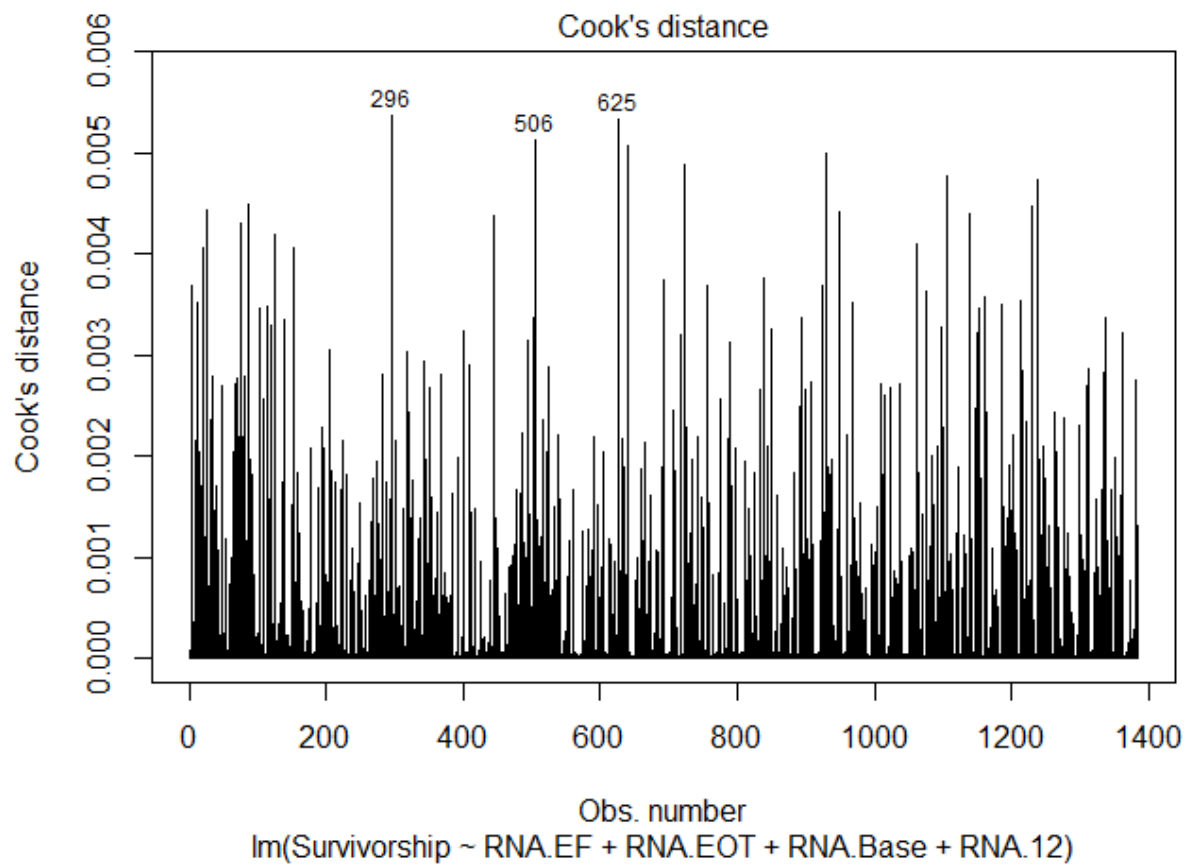


```
> leveragePlots(fit) # leverage plots
```

## Leverage Plots



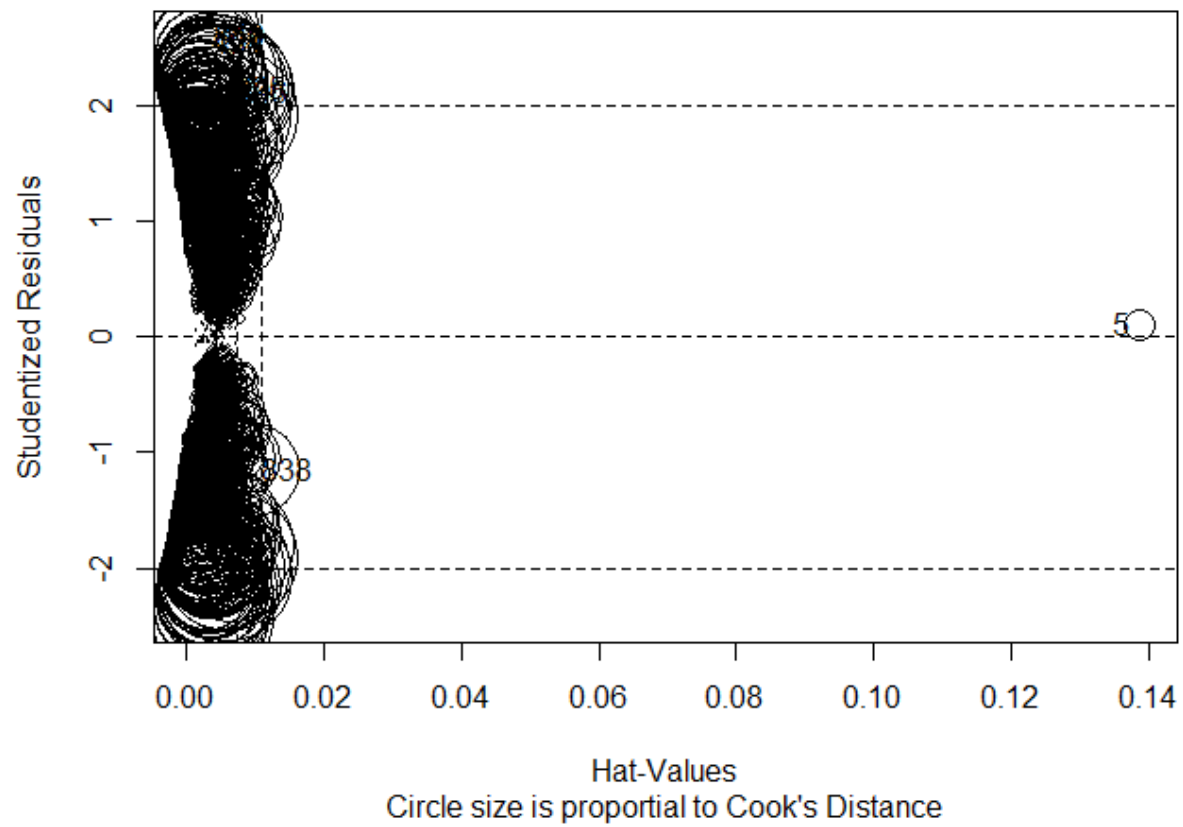
```
> # Influential Observations
> # Cook's D plot
> # identify D values > 4/(n-k-1)
> cutoff <- 4/((nrow(HCV_lr)-length(fit$coefficients)-2))
> plot(fit, which=4, cook.levels=cutoff)
```



```
> # Influence Plot
> influencePlot(fit, id.method="identify", main="Influence Plot", sub="Circle size is proportional
to Cook's Distance")
```

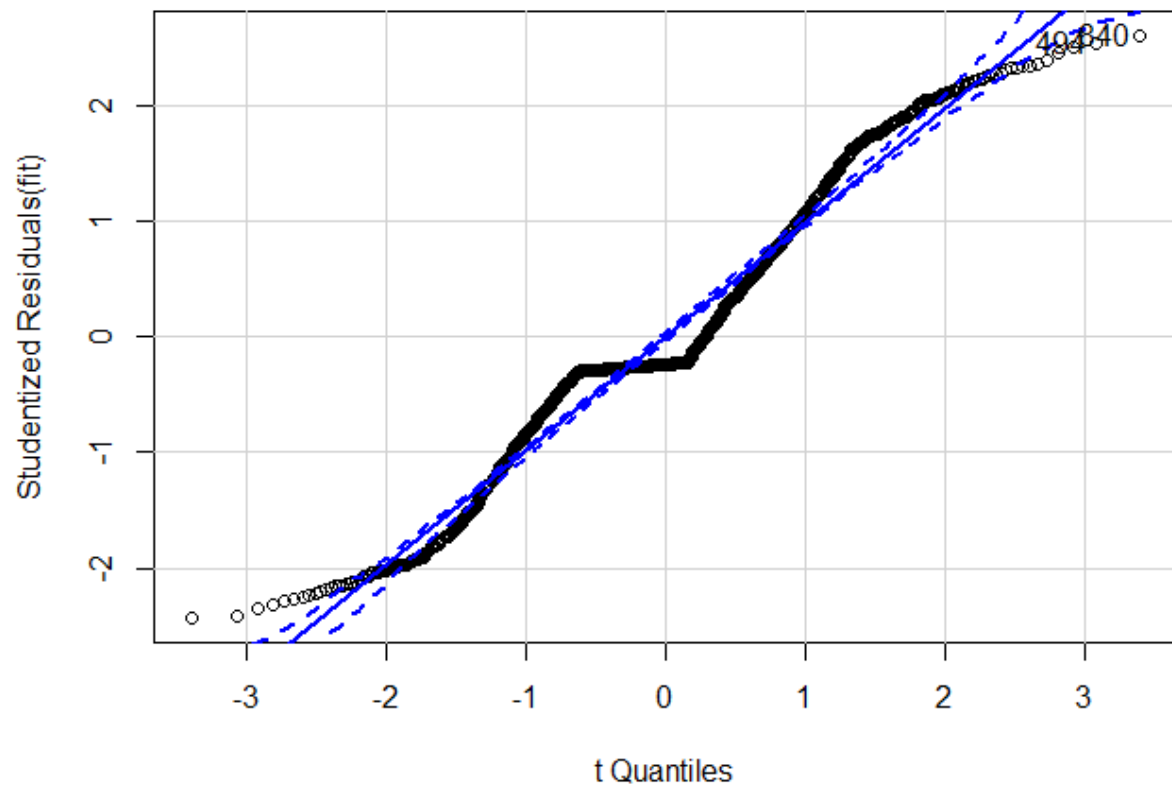
	StudRes	Hat	CookD
5	0.0991361	0.138768759	0.0003169402
296	2.1047833	0.006039832	0.0053705967
494	2.5371963	0.002452653	0.0031530644
625	2.1441686	0.005779939	0.0053316046
838	-1.1507095	0.009669033	0.0025850093
840	2.6034820	0.002786195	0.0037717929

## Influence Plot



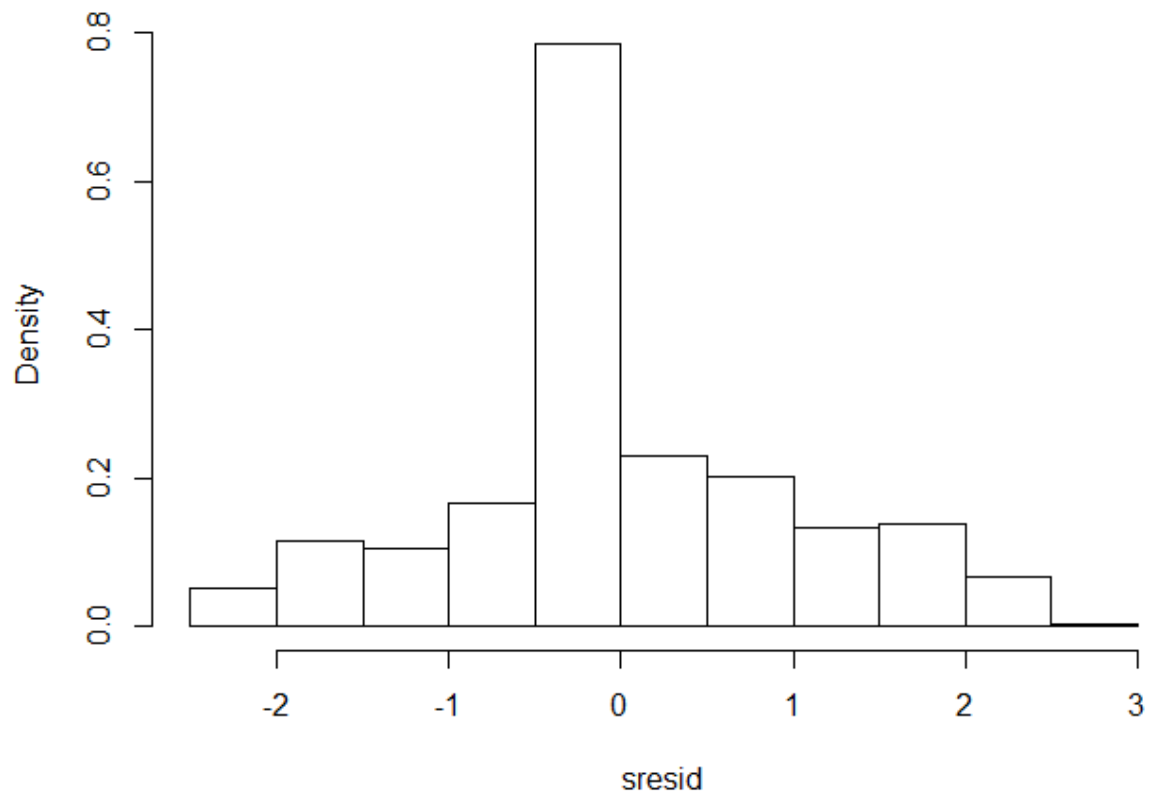
```
> # Normality of Residuals  
> # qq plot for studentized resid  
> qqPlot(fit, main="QQ Plot")  
[1] 494 840
```

QQ Plot



```
> # distribution of studentized residuals
> library(MASS)
> sresid <- studres(fit)
> hist(sresid, freq=FALSE,
+      main="Distribution of Studentized Residuals")
```

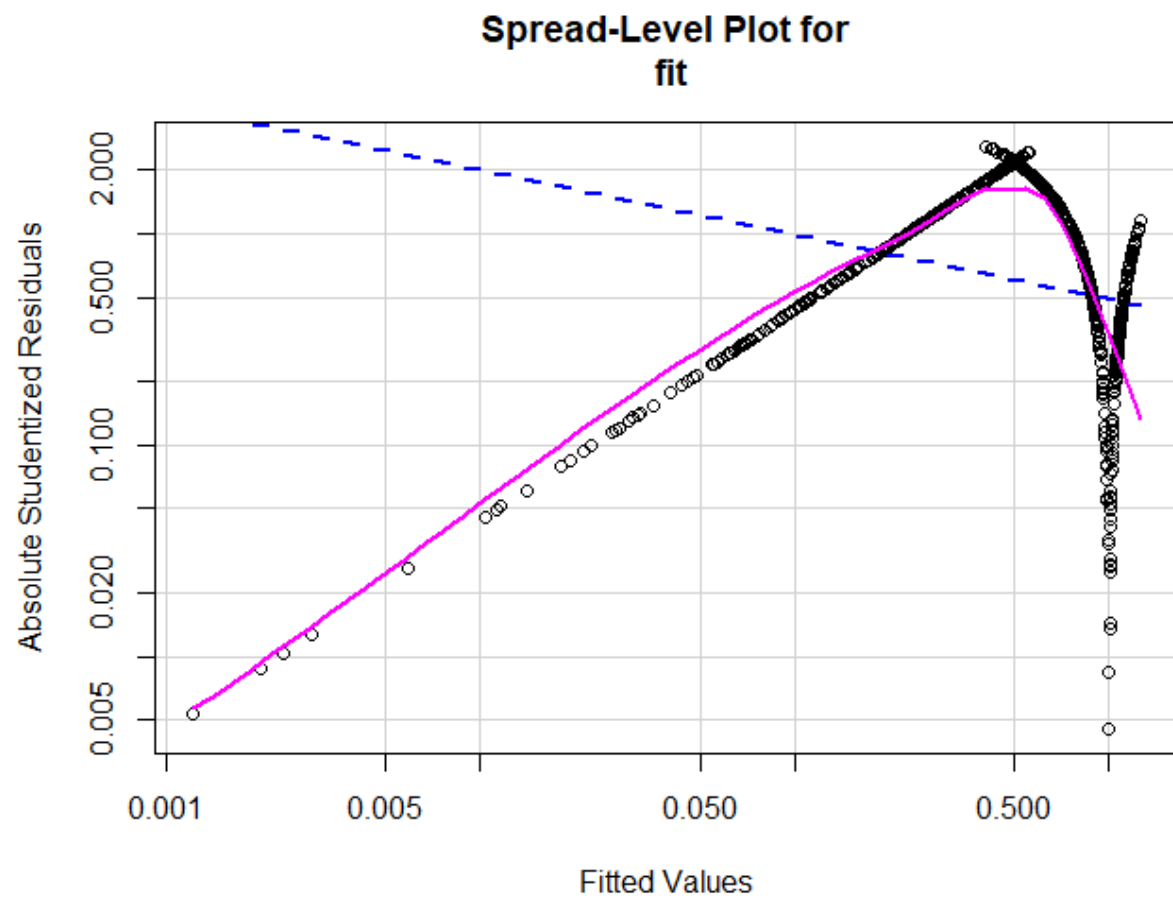
## Distribution of Studentized Residuals



```
> xfit<-seq(min(sresid),max(sresid),length=40)
> yfit<-dnorm(xfit)
> lines(xfit, yfit)
> #Non-constant Error Variance
> # Evaluate homoscedasticity
> # non-constant error variance test
> ncvTest(fit)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 79.34254, Df = 1, p = < 2.22e-16
> # plot studentized residuals vs. fitted values
> spreadLevelPlot(fit)
```

Suggested power transformation: 1.305065

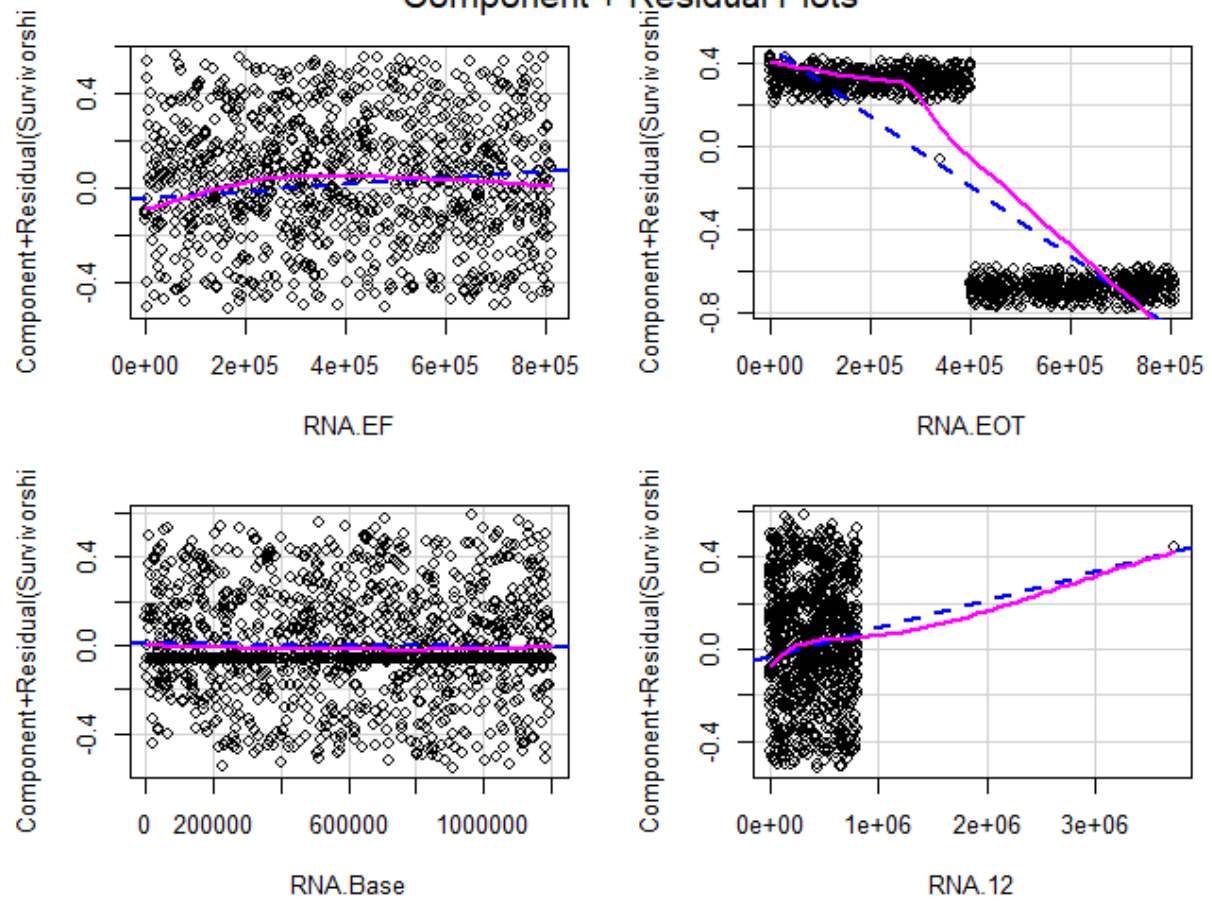
Warning message:  
In spreadLevelPlot.lm(fit) :  
145 negative fitted values removed



```
> #Multi-collinearity
> # Evaluate Collinearity
> vif(fit) # variance inflation factors
RNA.EF RNA.EOT RNA.Base RNA.12
1.364358 1.347477 1.000344 1.310757
> sqrt(vif(fit)) > 2 # problem?
RNA.EF RNA.EOT RNA.Base RNA.12
FALSE FALSE FALSE FALSE
> #Nonlinearity
> # component + residual plot
> crPlots(fit)
```

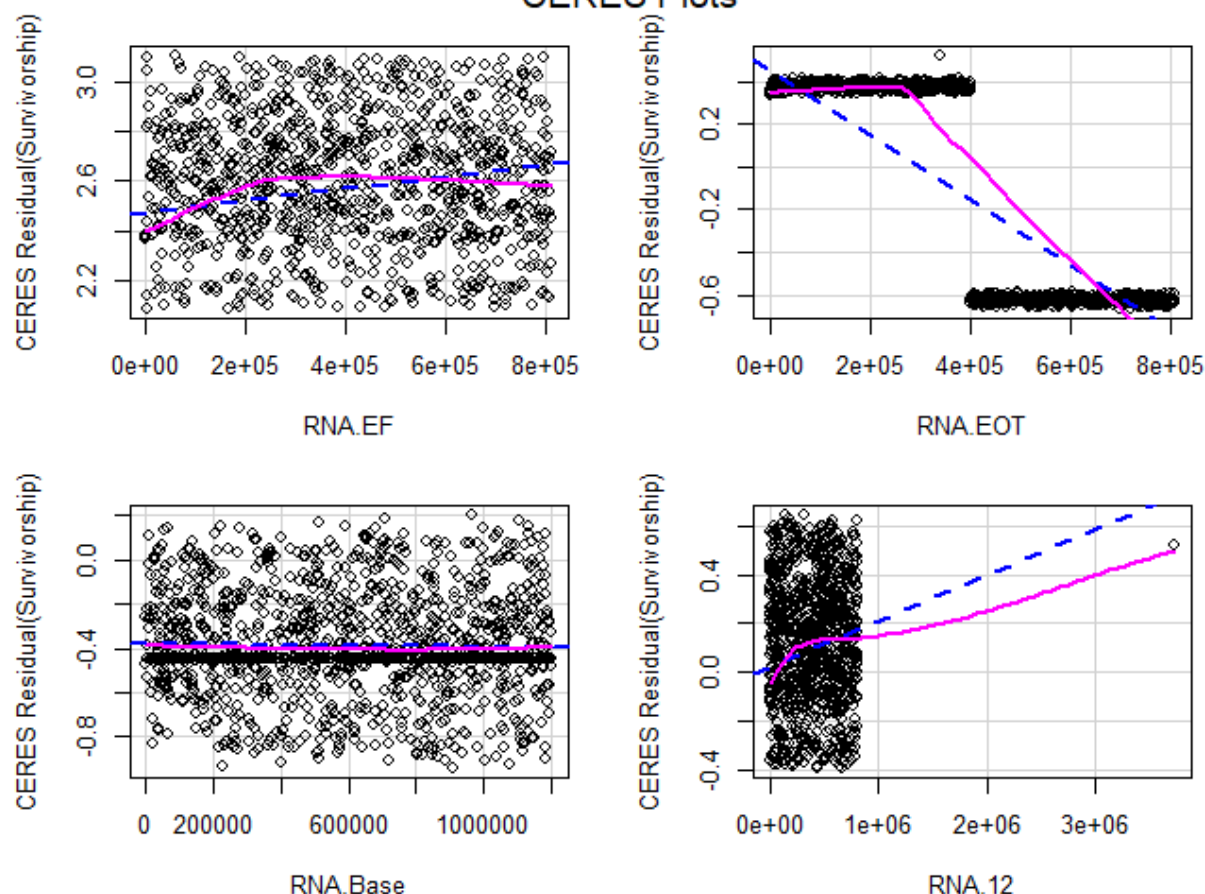


## Component + Residual Plots



```
> # Ceres plots
> ceresPlots(fit)
```

## CERES Plots



```
> #Non-independence of Errors
> # Test for Autocorrelated Errors
> durbinwatsonTest(fit)
lag Autocorrelation D-W Statistic p-value
1 0.05058114 1.898679 0.07
Alternative hypothesis: rho != 0
> # Global test of model assumptions
> install.packages("gvlma", lib="/Library/Frameworks/R.framework/Versions/3.5/Resources/library")
> library(gvlma)
> gvmmodel <- gvlma(fit)
> summary(gvmmodel)
```

```
Call:
lm(formula = Survivorship ~ RNA.EF + RNA.EOT + RNA.Base + RNA.12,
    data = HCV)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.55458 -0.08468 -0.05481  0.13407  0.59337
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.068e+00  1.453e-02  73.482 < 2e-16 ***
RNA.EF       1.451e-07  2.682e-08   5.409 7.47e-08 ***
RNA.EOT     -1.687e-06  2.697e-08 -62.531 < 2e-16 ***
RNA.Base    -1.384e-08  1.737e-08  -0.797  0.426
RNA.12      1.224e-07  2.467e-08   4.963 7.80e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2287 on 1380 degrees of freedom
Multiple R-squared:  0.7703, Adjusted R-squared:  0.7696
F-statistic: 1157 on 4 and 1380 DF, p-value: < 2.2e-16
```

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS  
 USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:  
 Level of Significance = 0.05

Call:  
 gvlma(x = fit)

	Value	p-value	Decision
Global Stat	27.0771	1.918e-05	Assumptions NOT satisfied!
Skewness	8.7671	3.067e-03	Assumptions NOT satisfied!
Kurtosis	0.1197	7.294e-01	Assumptions acceptable.
Link Function	18.0083	2.199e-05	Assumptions NOT satisfied!
Heteroscedasticity	0.1820	6.696e-01	Assumptions acceptable.

> fit

Call:  
 lm(formula = Survivorship ~ RNA.EF + RNA.EOT + RNA.Base + RNA.12,  
 data = HCV)

Coefficients:  
 (Intercept) RNA.EF RNA.EOT RNA.Base RNA.12  
 1.068e+00 1.451e-07 -1.687e-06 -1.384e-08 1.224e-07  
 > summary(fit)

Call:  
 lm(formula = Survivorship ~ RNA.EF + RNA.EOT + RNA.Base + RNA.12,  
 data = HCV)

Residuals:  

Min	1Q	Median	3Q	Max
-0.55458	-0.08468	-0.05481	0.13407	0.59337

Coefficients:  

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.068e+00	1.453e-02	73.482	< 2e-16 ***
RNA.EF	1.451e-07	2.682e-08	5.409	7.47e-08 ***
RNA.EOT	-1.687e-06	2.697e-08	-62.531	< 2e-16 ***
RNA.Base	-1.384e-08	1.737e-08	-0.797	0.426
RNA.12	1.224e-07	2.467e-08	4.963	7.80e-07 ***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2287 on 1380 degrees of freedom  
 Multiple R-squared: 0.7703, Adjusted R-squared: 0.7696  
 F-statistic: 1157 on 4 and 1380 DF, p-value: < 2.2e-16

> fit1 <- fit  
 > fit2 <- lm(Survivorship~RNA.EF+RNA.EOT+RNA.12, data=HCV)  
 > # compare models  
 > anova(fit1, fit2)  
 Analysis of Variance Table

Model 1: Survivorship ~ RNA.EF + RNA.EOT + RNA.Base + RNA.12  
 Model 2: Survivorship ~ RNA.EF + RNA.EOT + RNA.12  

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1380	72.184				
2	1381	72.217	-1	-0.03319	0.6345	0.4258

 > step <- stepAIC(fit, direction="both")  
 Start: AIC=-4081.61  
 Survivorship ~ RNA.EF + RNA.EOT + RNA.Base + RNA.12

	Df	Sum of Sq	RSS	AIC
- RNA.Base	1	0.033	72.217	-4083.0
<none>			72.184	-4081.6
- RNA.12	1	1.289	73.473	-4059.1
- RNA.EF	1	1.530	73.714	-4054.6
- RNA.EOT	1	204.531	276.715	-2222.5

Step: AIC=-4082.98  
 Survivorship ~ RNA.EF + RNA.EOT + RNA.12

	Df	Sum of Sq	RSS	AIC
<none>			72.217	-4083.0
+ RNA.Base	1	0.033	72.184	-4081.6
- RNA.12	1	1.294	73.512	-4060.4
- RNA.EF	1	1.526	73.743	-4056.0
- RNA.EOT	1	204.611	276.828	-2223.9

> step\$anova # display results  
 Stepwise Model Path  
 Analysis of Deviance Table

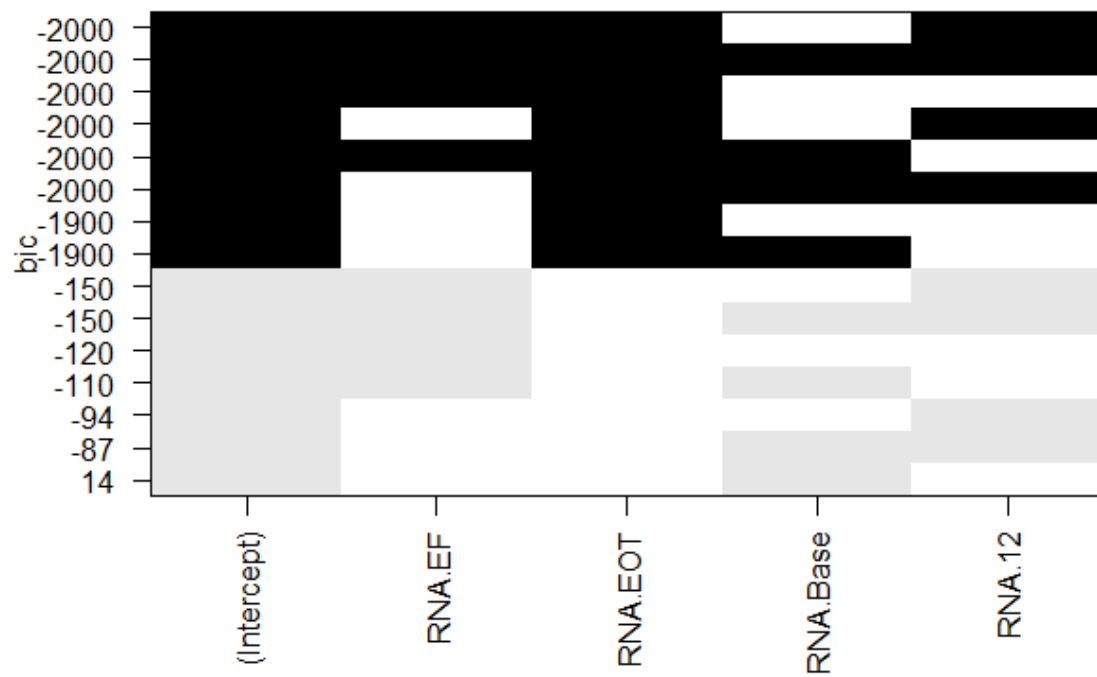
Initial Model:  
Survivorship ~ RNA.EF + RNA.EOT + RNA.Base + RNA.12

Final Model:  
Survivorship ~ RNA.EF + RNA.EOT + RNA.12

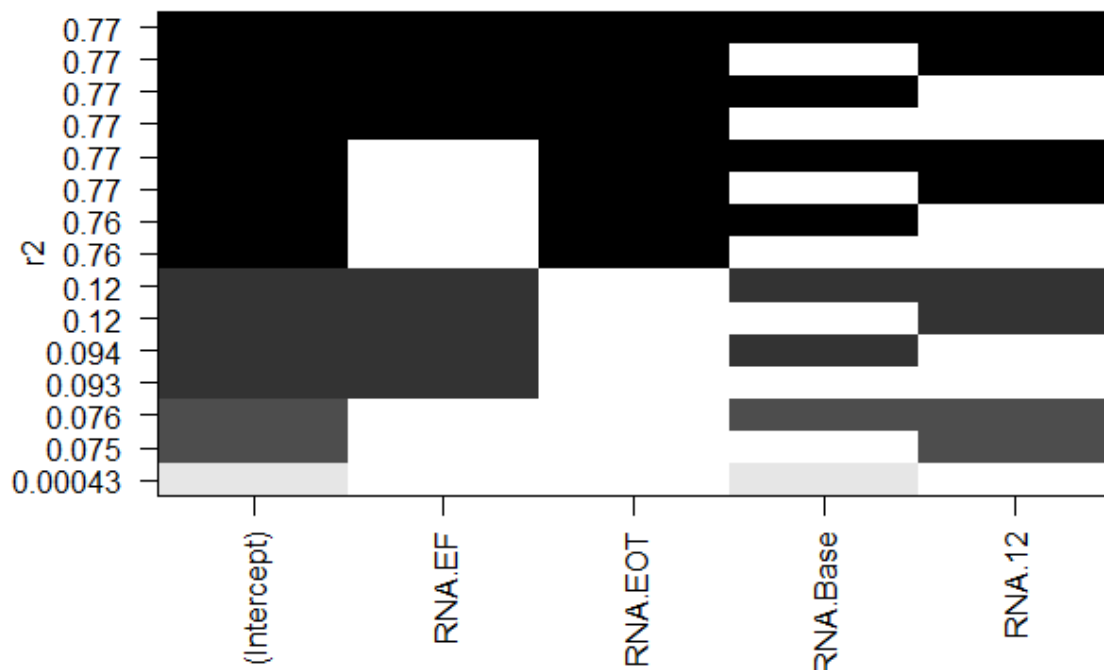
```

      Step Df   Deviance Resid. Df Resid. Dev      AIC
1              1380    72.18418 -4081.615
2 - RNA.Base  1 0.03319039    1381    72.21738 -4082.978
> install.packages("leaps", lib="/Library/Frameworks/R.framework/Versions/3.5/Resources/library")
> library(leaps)
> leaps<-regsubsets(Survivorship~RNA.EF+RNA.EOT+RNA.Base+RNA.12, data=HCV_lr,nbest=10)
> # view results
> summary(leaps)
Subset selection object
Call: regsubsets.formula(Survivorship ~ RNA.EF + RNA.EOT + RNA.Base +
      RNA.12, data = HCV_lr, nbest = 10)
4 Variables (and intercept)
      Forced in Forced out
RNA.EF      FALSE      FALSE
RNA.EOT      FALSE      FALSE
RNA.Base     FALSE      FALSE
RNA.12       FALSE      FALSE
10 subsets of each size up to 4
Selection Algorithm: exhaustive
      RNA.EF RNA.EOT RNA.Base RNA.12
1 ( 1 ) " " " " " "
1 ( 2 ) "*" " " " " " "
1 ( 3 ) " " " " " " "*"
1 ( 4 ) " " " " "*" " "
2 ( 1 ) "*" "*" " " " "
2 ( 2 ) " " "*" " " "*"
2 ( 3 ) " " "*" "*" " "
2 ( 4 ) "*" " " " " "*"
2 ( 5 ) "*" " " "*" " "
2 ( 6 ) " " " " "*" "*"
3 ( 1 ) "*" "*" " " "*"
3 ( 2 ) "*" "*" "*" " "
3 ( 3 ) " " "*" "*" "*"
3 ( 4 ) "*" " " "*" "*"
4 ( 1 ) "*" "*" "*" "*"
> # plot a table of models showing variables in each model.
> # models are ordered by the selection statistic.
> plot(leaps)

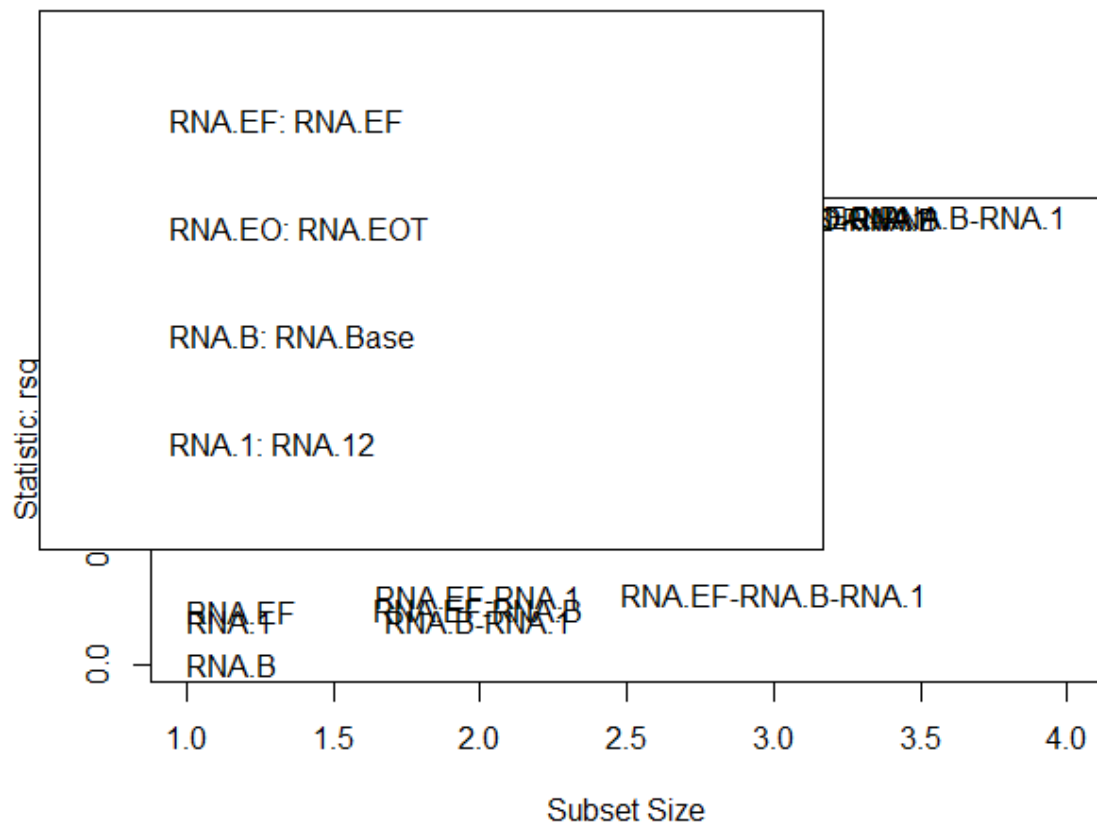
```



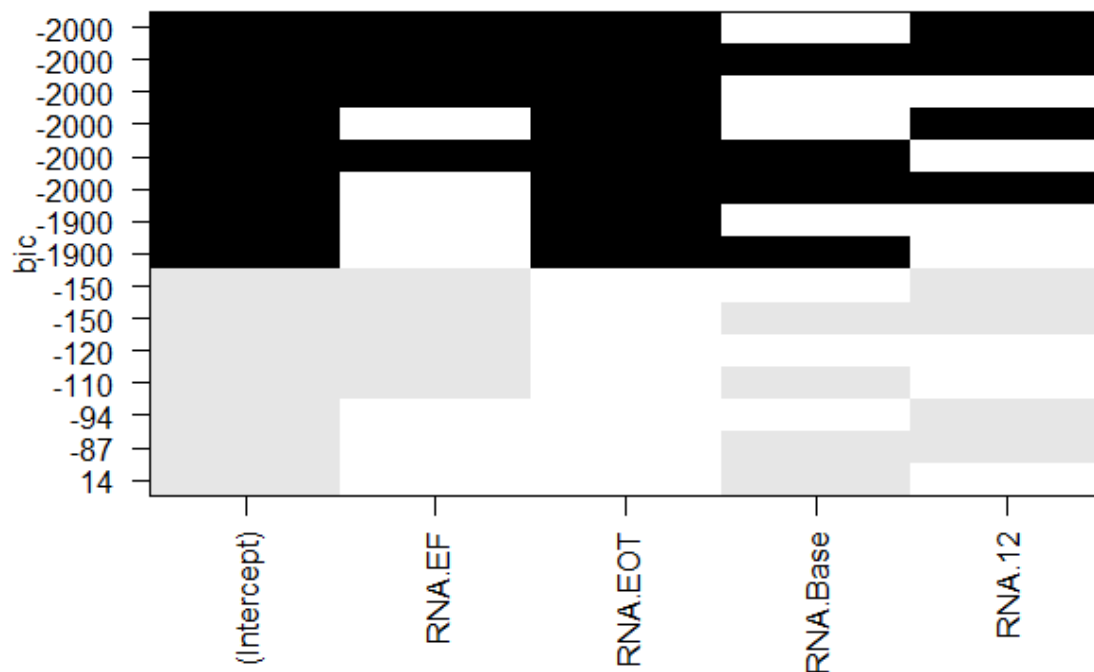
```
> plot(leaps,scale="r2")
```



```
> subsets(leaps, statistic="rsq")
```



```
> # All Subsets Regression
> plot(leaps,scale="bic")
```



```
> summary(leaps)
Subset selection object
Call: regsubsets.formula(Survivorship ~ RNA.EF + RNA.EOT + RNA.Base +
  RNA.12, data = HCV_lr, nbest = 10)
4 variables (and intercept)
      Forced in Forced out
RNA.EF      FALSE      FALSE
RNA.EOT      FALSE      FALSE
RNA.Base     FALSE      FALSE
RNA.12       FALSE      FALSE
10 subsets of each size up to 4
Selection Algorithm: exhaustive
      RNA.EF RNA.EOT RNA.Base RNA.12
1 ( 1 ) " " " " " "
1 ( 2 ) "*" " " " " " "
1 ( 3 ) " " " " " " "*"
1 ( 4 ) " " " " "*" " "
2 ( 1 ) "*" "*" " " " "
2 ( 2 ) " " "*" " " "*"
2 ( 3 ) " " "*" "*" " "
2 ( 4 ) "*" " " " " "*"
2 ( 5 ) "*" " " "*" " "
2 ( 6 ) " " " " "*" "*"
3 ( 1 ) "*" "*" " " "*"
3 ( 2 ) "*" "*" "*" " "
3 ( 3 ) " " "*" "*" "*"
3 ( 4 ) "*" " " "*" "*"
4 ( 1 ) "*" "*" "*" "*"
> view(leaps)
> leaps
Subset selection object
Call: regsubsets.formula(Survivorship ~ RNA.EF + RNA.EOT + RNA.Base +
  RNA.12, data = HCV_lr, nbest = 10)
4 variables (and intercept)
      Forced in Forced out
```



```

RNA.EF      FALSE      FALSE
RNA.EOT     FALSE      FALSE
RNA.Base    FALSE      FALSE
RNA.12      FALSE      FALSE
10 subsets of each size up to 4
Selection Algorithm: exhaustive
> coef(leaps,1:4)
[[1]]
      (Intercept)          RNA.EOT
1.102941e+00 -1.567667e-06

[[2]]
      (Intercept)          RNA.EF
8.104191e-01 -5.437384e-07

[[3]]
      (Intercept)          RNA.12
7.841322e-01 -4.576448e-07

[[4]]
      (Intercept)          RNA.Base
6.683893e-01 -2.775824e-08
> #prediction of the cured
> predict.lm(fit, data.frame(RNA.12 =585688 ,RNA.EF=582301,RNA.EOT= 744463,RNA.Base=1041941) )
1
-0.04620214

```