

Dataset:

```
hospitals <- read.csv("Hospitals.csv")
tempHosp <- hospitals # create a new dataset so as to not modify the original data
```

Model 1 (Number following “Table” or “Plot” indicates model number)

Code:

```
modell1 <- lm(LenStay ~ Age + InfRisk + Culture + XRay + School + Region + Beds + Census +
Nurses + Services, data = tempHosp)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.242682	1.612956	1.390	0.167489
Age	0.079922	0.028266	2.827	0.005668 **
InfRisk	0.439665	0.127298	3.454	0.000812 ***
Culture	0.005546	0.015982	0.347	0.729299
XRay	0.012688	0.007147	1.775	0.078892 .
SchoolYes	0.266644	0.441089	0.605	0.546872
RegionNE	0.812966	0.351406	2.313	0.022744 *
RegionS	-0.345311	0.313506	-1.101	0.273345
RegionW	-1.067594	0.398736	-2.677	0.008672 **
Beds	-0.004851	0.003603	-1.346	0.181224
Census	0.015182	0.004424	3.432	0.000872 ***
Nurses	-0.005891	0.002218	-2.656	0.009203 **
Services	-0.012179	0.013774	-0.884	0.378698

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.231 on 100 degrees of freedom
Multiple R-squared: 0.6299, Adjusted R-squared: 0.5855
F-statistic: 14.18 on 12 and 100 DF, p-value: < 2.2e-16

Anova Table (Type II tests)

Response: LenStay

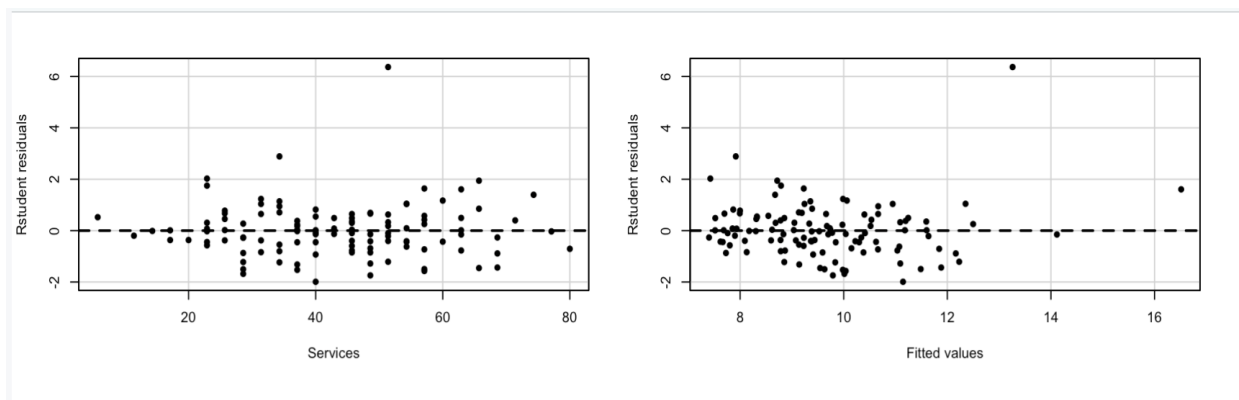
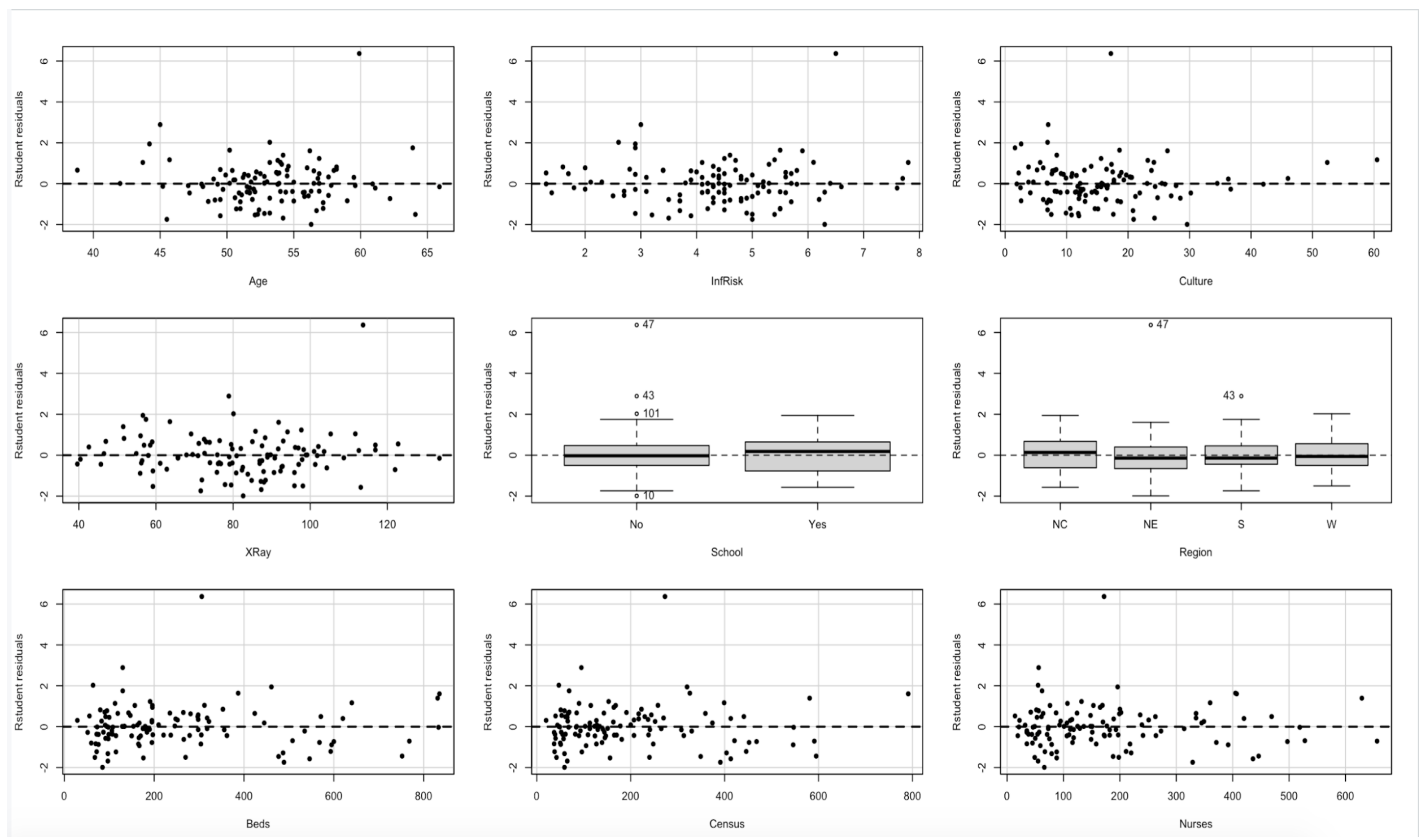
	Sum Sq	Df	F value	Pr(>F)
Age	12.108	1	7.9946	0.0056679 **
InfRisk	18.067	1	11.9290	0.0008119 ***
Culture	0.182	1	0.1204	0.7292992
XRay	4.773	1	3.1517	0.0788924 .
School	0.553	1	0.3654	0.5468722
Region	29.928	3	6.5869	0.0004153 ***
Beds	2.745	1	1.8127	0.1812243
Census	17.840	1	11.7794	0.0008723 ***
Nurses	10.684	1	7.0546	0.0092028 **
Services	1.184	1	0.7818	0.3786983
Residuals	151.451	100		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 1a. Anova output and global F-test output

Code:

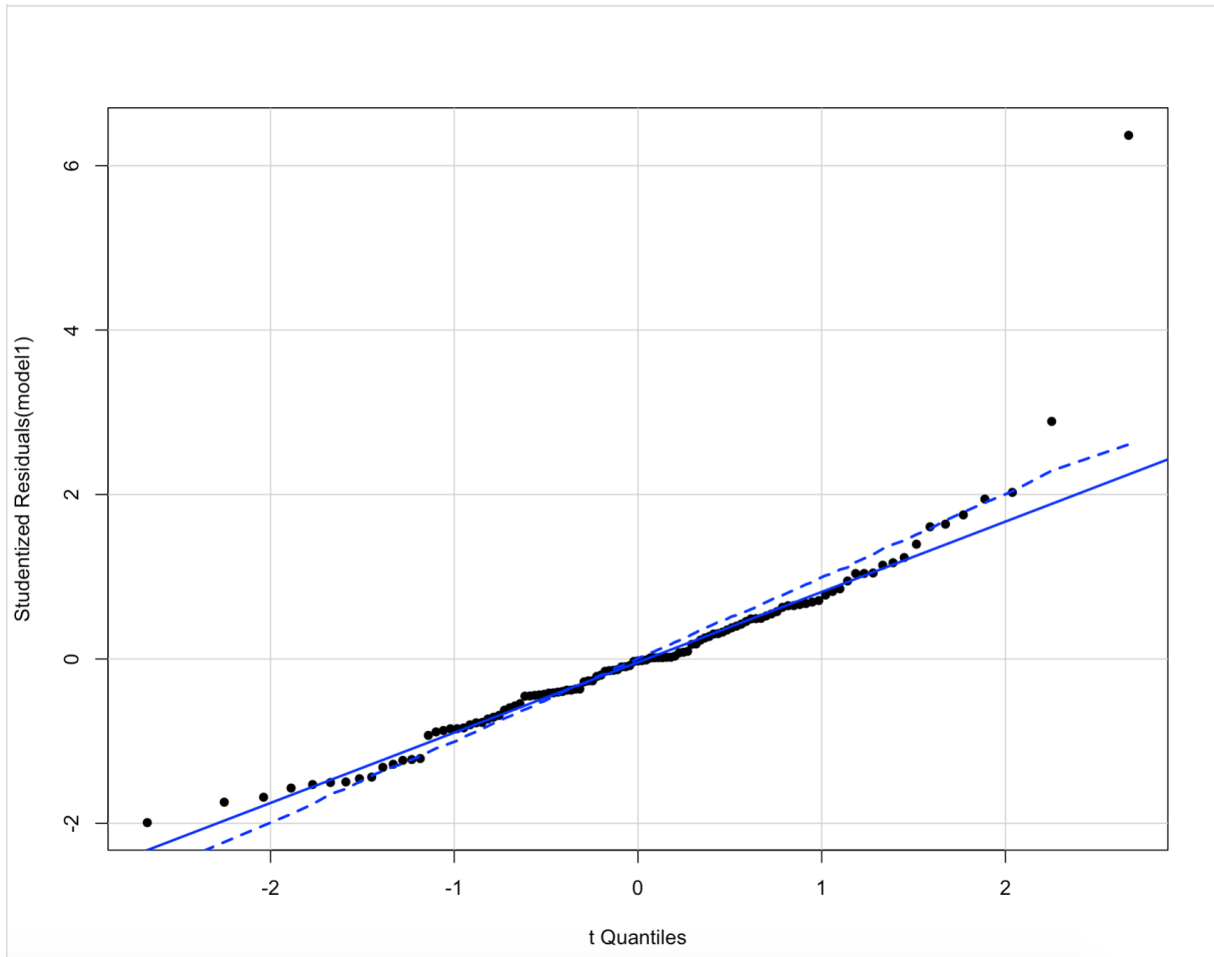
```
summary(modell1)
Anova(modell1)
```



Plot 1a. Residuals vs. Predicted Plots

Code:

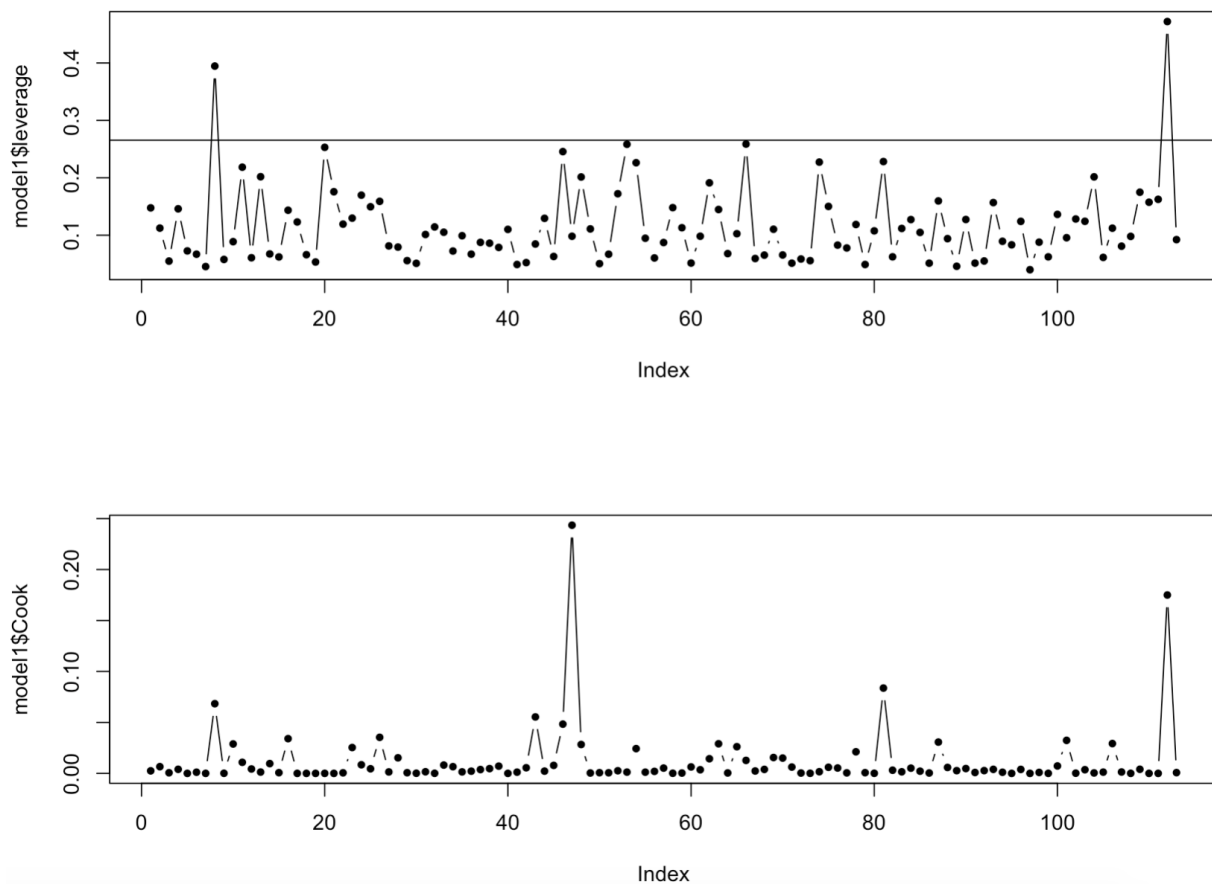
```
residualPlots(model1,type="rstudent",pch=16,quadratic=FALSE,id=FALSE,tests=FALSE)
```



Plot 1b. Normal Probability Plot

Code:

```
qqPlot(model1,envelope=FALSE,pch=16,id=FALSE)
```



Plot 1c. Residuals vs. Leverages plot and Cook's distance plot

Code:

```
#HIGH LEVERAGE
n = nrow(tempHosp)
p = 10
leverage.cutoff <- 3*p/n
model1$leverage <- hatvalues(model1)
plot(model1$leverage,type="b",pch=20)
abline(h = leverage.cutoff)

#COOK'S DIST
model1$Cook <- cooks.distance(model1)
plot(model1$Cook,type="b",pch=20)
```

	Variables	Tolerance	VIF
1	Age	0.85021570	1.176172
2	InfRisk	0.46410296	2.154694
3	Culture	0.50542828	1.978520
4	XRay	0.70608230	1.416265
5	SchoolYes	0.53898638	1.855334
6	RegionNE	0.58231265	1.717291
7	RegionS	0.61921830	1.614939
8	RegionW	0.69356705	1.441822
9	Beds	0.02801183	35.699204
10	Census	0.02923000	34.211423
11	Nurses	0.14173293	7.055523
12	Services	0.30846950	3.241812

```
> cor(tempHosp$Beds, tempHosp$Census)
[1] 0.9809977
> cor(tempHosp$Census, tempHosp$Nurses)
[1] 0.907897
> cor(tempHosp$Beds, tempHosp$Nurses)
[1] 0.9155042
```

Table 1b. VIF output and correlations

Code:

```
ols_vif_tol(model1)
cor(tempHosp$Beds, tempHosp$Census)
cor(tempHosp$Census, tempHosp$Nurses)
cor(tempHosp$Beds, tempHosp$Nurses)
```

Model 2

```
model2 <- lm(LenStay ~ Age + InfRisk + Culture + XRay + School + Region + Census + Nurses  
+ Services, data = tempHosp)
```

	Variables	Tolerance	VIF
1	Age	0.8536011	1.171507
2	InfRisk	0.4804537	2.081366
3	Culture	0.5180458	1.930331
4	XRay	0.7060893	1.416251
5	SchoolYes	0.5568362	1.795860
6	RegionNE	0.6297585	1.587910
7	RegionS	0.6200648	1.612735
8	RegionW	0.6981455	1.432366
9	Census	0.1360935	7.347888
10	Nurses	0.1503007	6.653327
11	Services	0.3316289	3.015419

Table 2a. VIF output

Code:

```
ols_vif_tol(model2)
```

Model 3

Code:

```
tempHosp$NursePerson <- tempHosp$Nurses / tempHosp$Census
```

#We now proceed with a variable that has the ratio for average nurses/average daily hospital patients.

```
model3 <- lm(LenStay ~ Age + InfRisk + Culture + XRay + School + Region + Census + NursePerson + Services, data = tempHosp)
```

	Variables	Tolerance	VIF
1	Age	0.8550171	1.169567
2	InfRisk	0.4786091	2.089388
3	Culture	0.5232812	1.911018
4	XRay	0.6964141	1.435927
5	SchoolYes	0.5567031	1.796290
6	RegionNE	0.6288709	1.590152
7	RegionS	0.6167636	1.621367
8	RegionW	0.6441908	1.552335
9	Census	0.2716812	3.680785
10	NursePerson	0.6928434	1.443327
11	Services	0.3333499	2.999851

Table 3a. VIF output

Code:

```
ols_vif_tol(model3)
```

Subsets Regression Summary											
Model	R-Square	Adj. R-Square	Pred R-Square	C(p)	AIC	SBIC	SBC	MSEP	FPE	HSP	APC
1	0.2846	0.2781	0.2538	76.8216	434.2539	111.6991	442.4361	298.0404	2.6842	0.0240	0.7412
2	0.4604	0.4404	0.4075	33.1433	408.3742	82.4570	424.7385	226.8389	2.0987	0.0188	0.5690
3	0.5232	0.5009	0.4536	18.8440	396.4026	71.0066	415.4943	202.3126	1.8881	0.0169	0.5118
4	0.5551	0.5299	0.4758	12.5643	390.5835	65.6805	412.4026	190.5512	1.7937	0.0161	0.4861
5	0.5823	0.5544	0.4911	7.5011	385.4569	61.2709	410.0034	180.5926	1.7145	0.0154	0.4646
6	0.6013	0.5706	0.5076	4.5516	382.1809	58.7616	409.4548	173.9960	1.6659	0.0149	0.4513
7	0.6088	0.5746	0.5064	4.6016	382.0327	59.1224	412.0340	172.3610	1.6642	0.0149	0.4508
8	0.6111	0.5730	0.5016	6.0025	383.3645	60.7942	416.0931	173.0082	1.6844	0.0151	0.4562
9	0.6111	0.5688	0.4937	8.0000	385.3616	63.0295	420.8177	174.7000	1.7149	0.0154	0.4644

AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
 FPE: Final Prediction Error
 HSP: Hocking's Sp
 APC: Amemiya Prediction Criteria

Table 3b. Best subsets regression summary output

Code:

```

select = ols_step_best_subset(model3)
select

```


Anova Table (Type II tests)

Response: LenStay

	Sum Sq	Df	F value	Pr(>F)	
Age	13.235	1	8.4366	0.0044924	**
InfRisk	28.544	1	18.1960	4.403e-05	***
XRay	7.798	1	4.9709	0.0279270	*
Region	34.977	3	7.4321	0.0001474	***
Census	22.335	1	14.2379	0.0002681	***
NursePerson	13.308	1	8.4834	0.0043869	**
Residuals	163.147	104			

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

Anova Table (Type II tests)

Response: LenStay

	Sum Sq	Df	F value	Pr(>F)	
Age	13.383	1	8.6114	0.004119	**
InfRisk	30.204	1	19.4349	2.561e-05	***
XRay	7.139	1	4.5936	0.034446	*
Region	37.566	3	8.0573	7.122e-05	***
Census	19.434	1	12.5046	0.000610	***
NursePerson	9.409	1	6.0544	0.015535	*
Services	3.072	1	1.9769	0.162732	
Residuals	160.075	103			

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

Table 3c. Global F-tests for best_subsets_modelA and best_subsets_modelB

Code:

```
best_subsets_modelA = lm(LenStay ~ Age + InfRisk + XRay + Region + Census +  
NursePerson, data = tempHosp)
```

```
best_subsets_modelB = lm(LenStay ~ Age + InfRisk + XRay + Region + Census + NursePerson  
+ Services, data = tempHosp)
```

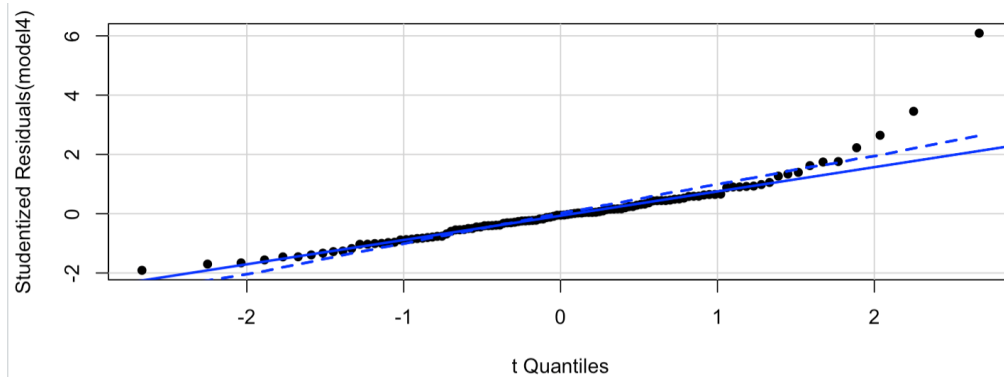
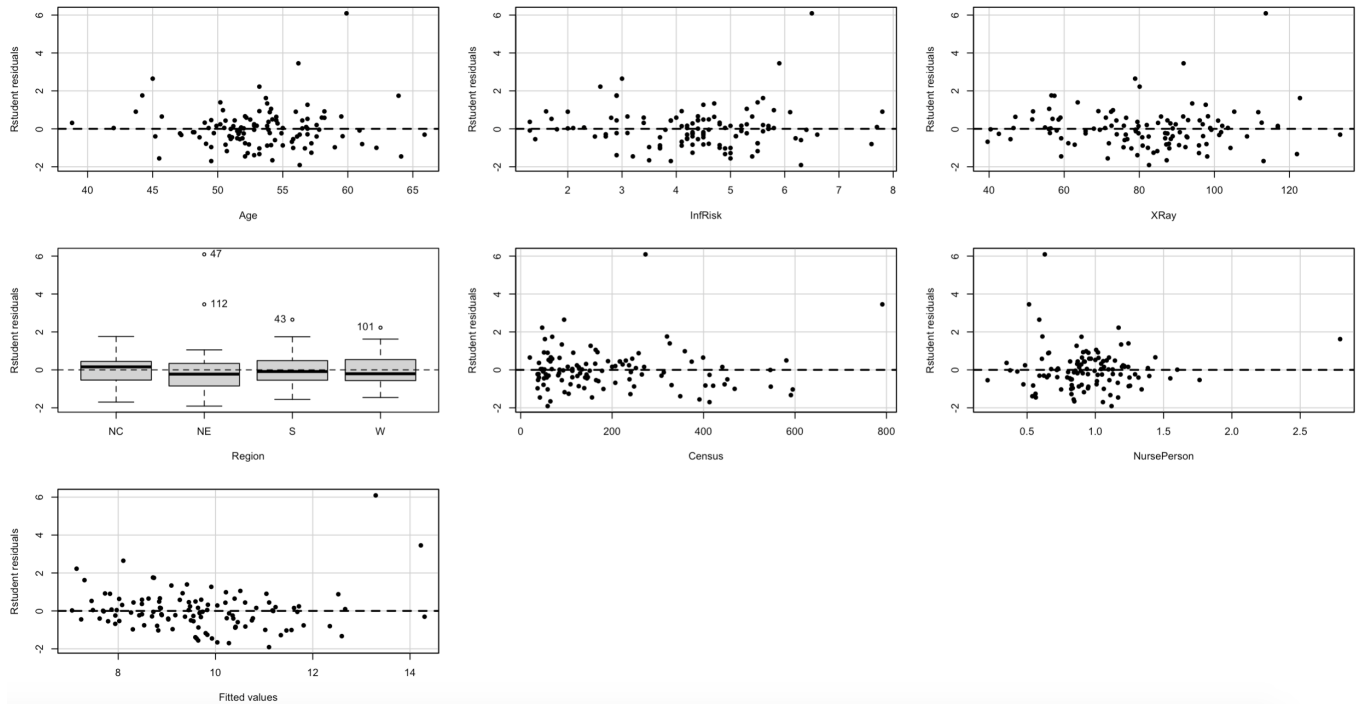
```
Anova(best_subsets_modelA)
```

```
Anova(best_subsets_modelB)
```

Model 4

Code:

```
model4 = lm(LenStay ~ Age + InfRisk + XRay + Region + Census + NursePerson, data =  
tempHosp)
```



Plot 4a. Residuals vs. predicted plot and normal probability plot

Code:

```
residualPlots(model4,type="rstudent",pch=16,quadratic=FALSE,id=FALSE,tests=FALSE)  
qqPlot(model4,envelope=FALSE,pch=16,id=FALSE)
```

Call:

```
lm(formula = LenStay ~ InfRisk + XRay + Region + Census + NursePerson +  
    InfRisk:Census, data = tempHosp)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.4265	-0.7497	-0.2095	0.5389	6.3475

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	8.4236249	0.7768264	10.844	< 2e-16	***
InfRisk	0.0661136	0.1508260	0.438	0.662047	
XRay	0.0162049	0.0069945	2.317	0.022474	*
RegionNE	0.9566034	0.3206044	2.984	0.003549	**
RegionS	-0.2977037	0.3009477	-0.989	0.324851	
RegionW	-0.7881964	0.4006397	-1.967	0.051807	.
Census	-0.0107840	0.0036150	-2.983	0.003556	**
NursePerson	-1.0589416	0.4151215	-2.551	0.012200	*
InfRisk:Census	0.0029142	0.0007344	3.968	0.000133	***

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

Residual standard error: 1.214 on 104 degrees of freedom

Multiple R-squared: 0.6256, Adjusted R-squared: 0.5969

F-statistic: 21.73 on 8 and 104 DF, p-value: < 2.2e-16

Table 4a. Stepwise regression results

Code:

```
step1 <- step(model4,scope=~.^2,direction="both",trace=1,k=log(n))  
summary(step1)
```

	Sum Sq	Df	F value	Pr(>F)	
cInfRisk	29.707	1	21.3935	1.111e-05	***
cXRay	4.478	1	3.2249	0.0755173	.
Region	33.621	3	8.0707	7.148e-05	***
cCensus	19.479	1	14.0279	0.0002999	***
cNursePerson	6.390	1	4.6018	0.0343359	*
Region:cInfRisk	12.941	3	3.1065	0.0298552	*
cInfRisk:cCensus	17.842	1	12.8492	0.0005218	***
Residuals	140.247	101			

	Sum Sq	Df	F value	Pr(>F)	
cInfRisk	37.548	1	26.9433	1.087e-06	***
cXRay	4.829	1	3.4650	0.0655885	.
Region	33.621	3	8.0418	7.393e-05	***
cCensus	20.420	1	14.6532	0.0002242	***
cNursePerson	7.588	1	5.4450	0.0216057	*
Region:cCensus	12.437	3	2.9748	0.0352084	*
cCensus:cInfRisk	13.825	1	9.9202	0.0021506	**
Residuals	140.751	101			

	Sum Sq	Df	F value	Pr(>F)	
cInfRisk	27.446	1	20.0723	1.969e-05	***
cXRay	4.464	1	3.2645	0.0737735	.
Region	33.621	3	8.1961	6.178e-05	***
cCensus	15.877	1	11.6114	0.0009425	***
cNursePerson	9.585	1	7.0099	0.0094063	**
Region:cNursePerson	15.087	3	3.6779	0.0146071	*
cInfRisk:cCensus	18.084	1	13.2259	0.0004367	***
Residuals	138.101	101			

Table 4b. Testing other possible interactions with global F-test

Code:

```
possible1 = lm(LenStay ~ Region:cInfRisk+ cInfRisk + cXRay + Region + cCensus +
cNursePerson + cInfRisk:cCensus, data = tempHosp)
Anova(possible1)
```

```
possible2 = lm(LenStay ~ Region:cCensus+ cInfRisk + cXRay + Region + cCensus +
cNursePerson + cInfRisk:cCensus, data = tempHosp)
Anova(possible2)
```

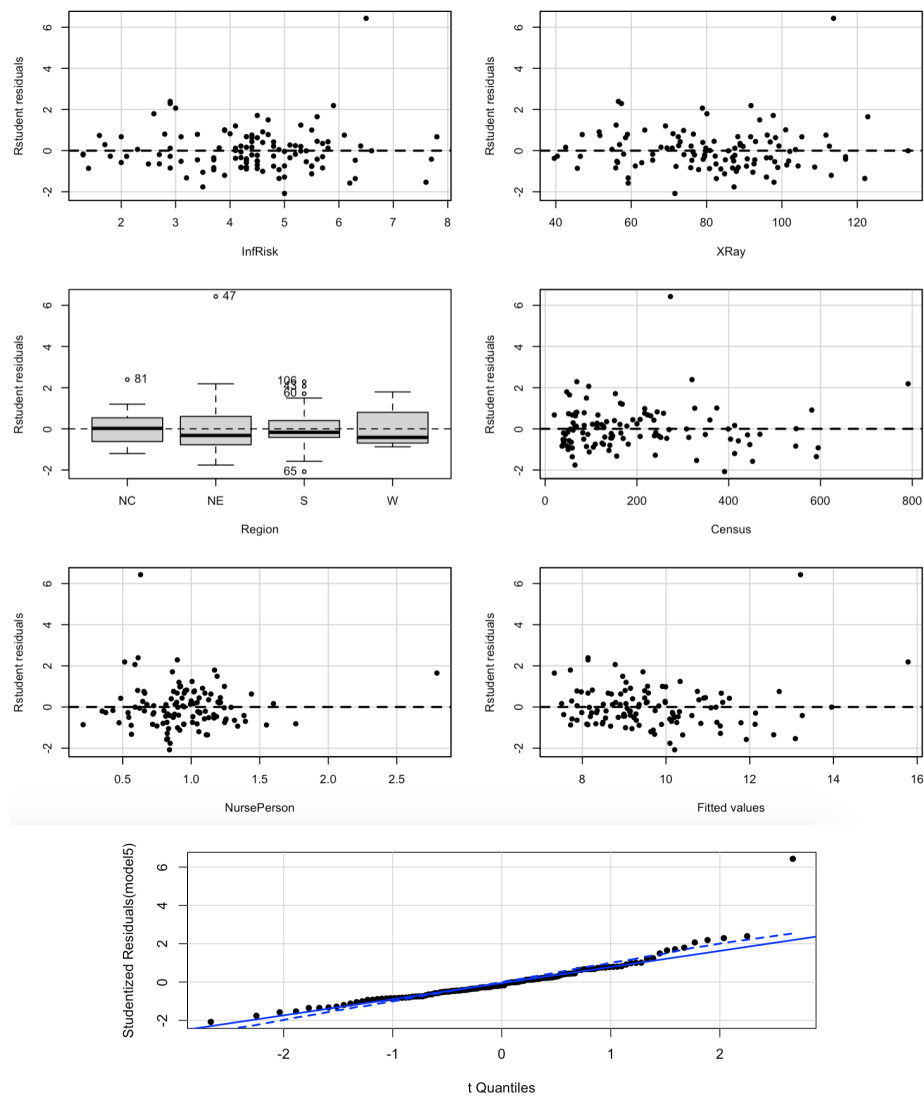
```
possible3 = lm(LenStay ~ Region:cNursePerson+ cInfRisk + cXRay + Region + cCensus +
cNursePerson + cInfRisk:cCensus, data = tempHosp)
Anova(possible3)
```

Note: Anova results are in order of code.

Model 5

Code:

```
model5 <- lm(LenStay ~ InfRisk + XRay + Region + Census + NursePerson +  
  InfRisk:Census, data = tempHosp)
```



Plot 5a. Residuals vs. predicted plots and normal probability plot

Code:

```
residualPlots(model5,type="rstudent",pch=16,quadratic=FALSE,id=FALSE,tests=FALSE)  
qqPlot(model5,envelope=FALSE,pch=16,id=FALSE)
```

	Variables	Tolerance	VIF
1	InfRisk	0.32153156	3.110115
2	XRay	0.71694015	1.394817
3	RegionNE	0.68038495	1.469756
4	RegionS	0.65354166	1.530124
5	RegionW	0.66814518	1.496681
6	Census	0.04256648	23.492664
7	NursePerson	0.73809612	1.354837
8	InfRisk:Census	0.03398529	29.424497

Table 5a. VIF output

Code:

```
ols_vif_tol(model5)
```

Model 6 || Final Model

Code:

```
final_model = lm(LenStay ~ cInfRisk + cXRay + Region + cCensus + cNursePerson +  
cInfRisk:cCensus, data = tempHosp)  
summary(final_model)  
Anova(final_model)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.4265	-0.7497	-0.2095	0.5389	6.3475

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.3932257	0.2198023	42.735	< 2e-16 ***
cInfRisk	0.6238084	0.1134435	5.499	2.75e-07 ***
cXRay	0.0162049	0.0069945	2.317	0.022474 *
RegionNE	0.9566034	0.3206044	2.984	0.003549 **
RegionS	-0.2977037	0.3009477	-0.989	0.324851
RegionW	-0.7881964	0.4006397	-1.967	0.051807 .
cCensus	0.0019070	0.0009055	2.106	0.037621 *
cNursePerson	-1.0589416	0.4151215	-2.551	0.012200 *
cInfRisk:cCensus	0.0029142	0.0007344	3.968	0.000133 ***

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

Residual standard error: 1.214 on 104 degrees of freedom
Multiple R-squared: 0.6256, Adjusted R-squared: 0.5969
F-statistic: 21.73 on 8 and 104 DF, p-value: < 2.2e-16

Anova Table (Type II tests)

Response: LenStay

	Sum Sq	Df	F value	Pr(>F)
cInfRisk	29.707	1	20.1680	1.842e-05 ***
cXRay	7.906	1	5.3677	0.0224738 *
Region	33.621	3	7.6084	0.0001196 ***
cCensus	20.420	1	13.8635	0.0003195 ***
cNursePerson	9.585	1	6.5072	0.0121997 *
cInfRisk:cCensus	23.194	1	15.7462	0.0001334 ***
Residuals	153.188	104		

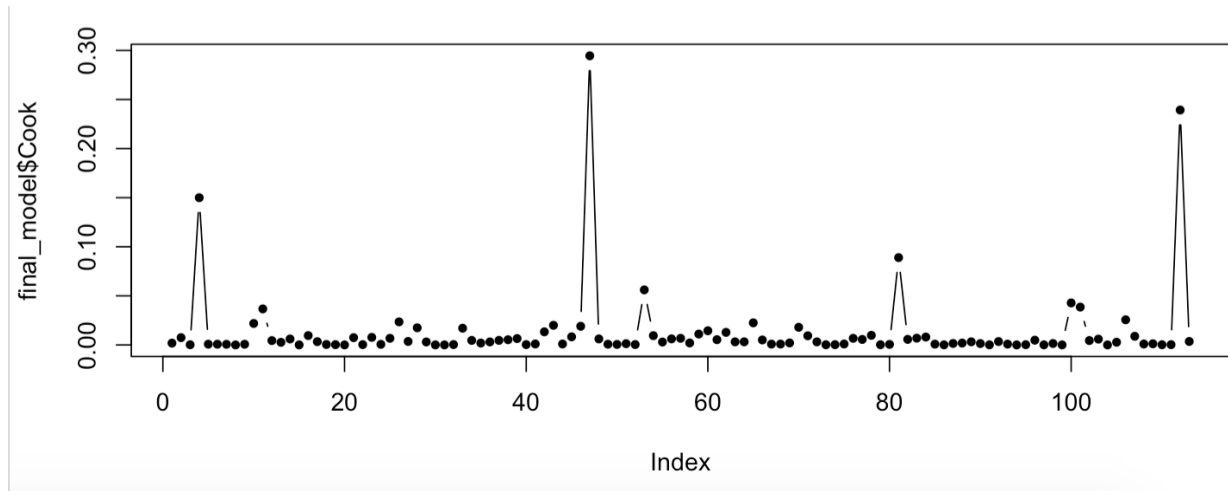
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 6a. Anova results and global F-test

Code:

```
summary(final_model)
```

Anova(final_model)



Plot 6a. Cook's Distance Plot

Code:

```
final_model$Cook <- cooks.distance(final_model)
plot(final_model$Cook,type="b",pch=20)
```


Analysis of Variance Table

Model 1: LenStay ~ cInfRisk + cXRay + Region + cCensus + cNursePerson +
cInfRisk:cCensus

Model 2: LenStay ~ Age + InfRisk + Culture + XRay + School + Region +
Beds + Census + Nurses + Services + InfRisk:Census + NursePerson

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	104	153.19				
2	98	143.13	6	10.057	1.1477	0.3408

Table 6b. Partial F-test comparing reduced model to full model

Code:

```
full_model = lm(LenStay ~ Age + InfRisk + Culture + XRay + School + Region + Beds +  
Census + Nurses + Services + InfRisk:Census + NursePerson, data = tempHosp)
```

```
anova(final_model, full_model)
```