

Breast Tumor Malignancy Based on Biopsy Data

Dorothy Walls, Greg Brown, Abraham Hussein

November 3, 2019

The Dataset

We consider the individuals examined at the University of Wisconsin, which produced 669 digitized images of tumors using fine needle aspiration (FNA) on the breast. In addition to the measurements, we know if the tumor examined was malignant (cancerous) or non-malignant. We will define our binary response variables:

$Y_i = 1$ if the i th individual has a malignant tumor,

$Y_i = 0$ if the i th individual has a non-malignant tumor.

We also consider the following 9 potential predictor variables:

X_1 = radius, or the mean of distances from center to points on the perimeter (out of 10),

X_2 = texture, or the standard deviation of gray-scale values produced from the FNA (out of 10),

X_3 = Perimeter, or the mean of distances from center to points on the perimeter (out of 10),

X_4 = Area, the total area of the tumor (out of 10),

X_5 = Smoothness, the local variation in radius lengths (out of 10),

X_6 = Compactness, which is defined as $\frac{perimeter^2}{area} - 1$ (out of 10),

X_7 = Concavity, the severity of concave portions of the contour for the tumor (out of 10),

X_8 = Concave points, the total number of concave portions of the contour (out of 10),

X_9 = Symmetry, the measure of symmetry in the tumor (out of 10).

Our goal is create a logistic model of our data, one that will attempt to predict the likelihood of a tumor located at the breast developing into breast cancer.

```
cancer <- read.csv("breast-cancer-wisconsin.csv", header = T)
attach(cancer)
names(cancer)
```

```
## [1] "radius"      "texture"      "perimeter"    "area"
## [5] "smoothness"  "compactness"  "concavity"    "concave.points"
## [9] "symmetry"    "class"
```

```
cancer[1:3,]
```

```
##   radius texture perimeter area smoothness compactness concavity
## 1     5       1         1    1         2           1         3
## 2     5       4         4    5         7          10         3
## 3     3       1         1    1         2           2         3
##   concave.points symmetry class
## 1             1         1     0
## 2             2         1     0
## 3             1         1     0
```

```
dim(cancer)
```

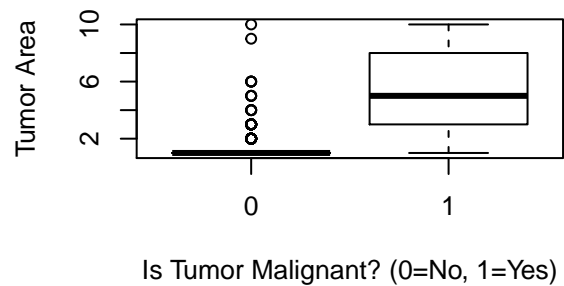
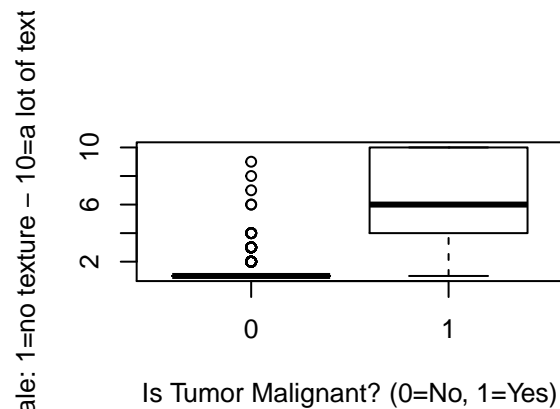
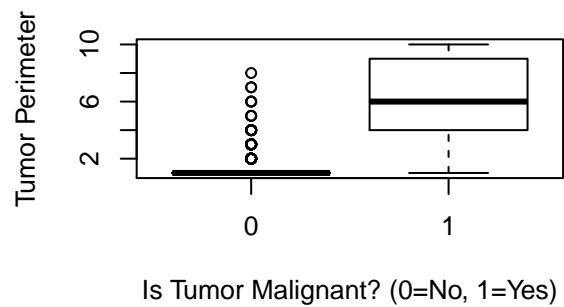
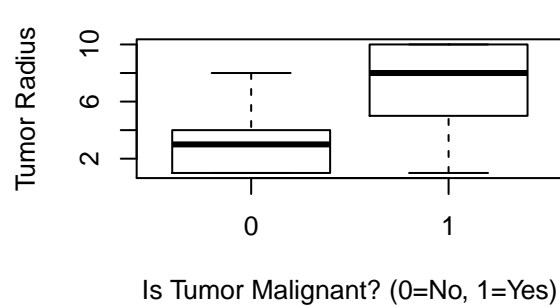
```
## [1] 699 10
```

```
sum(is.na(cancer))
```

```
## [1] 0
```

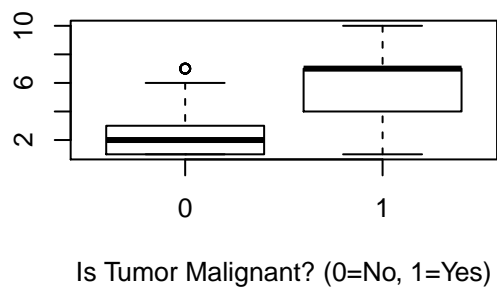
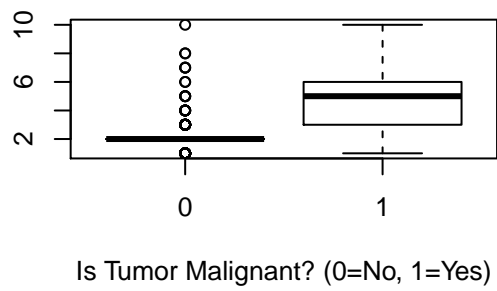
```
n<-nrow(cancer) #The total number of observations.
```

```
malignant<-class #class is a function in R and will cause problems.
```

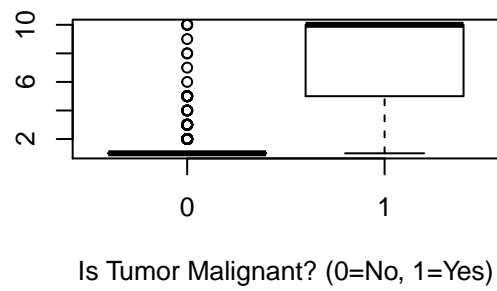


Scale: 1=no texture – 10=a lot of text

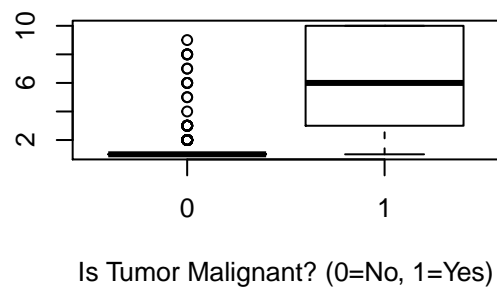
(Scale: 1=not concave – 10=concav(Scale: 1=not smooth – 10=very smc

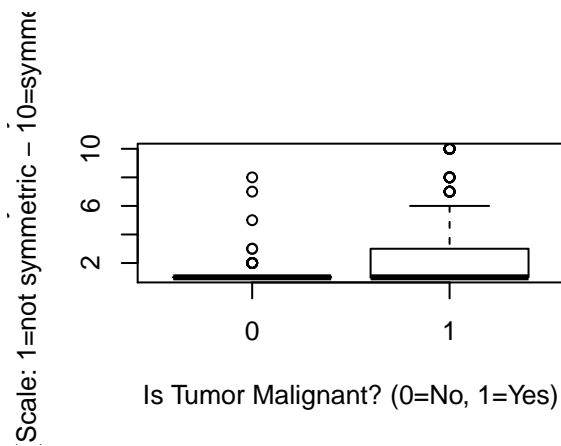


(Scale: 1=fluffy – 10=very dense



Concave Points on the Tumor





From our boxplots, we can begin to visualize the differences between the malignant and non malignant tumors. With the exception of the variable *symmetry*, malignant tumors scores are higher compared to non-malignant tumors; in fact, non malignant tumors tend to have very little spread in terms of score, yet they have a few outliers.

Model Building

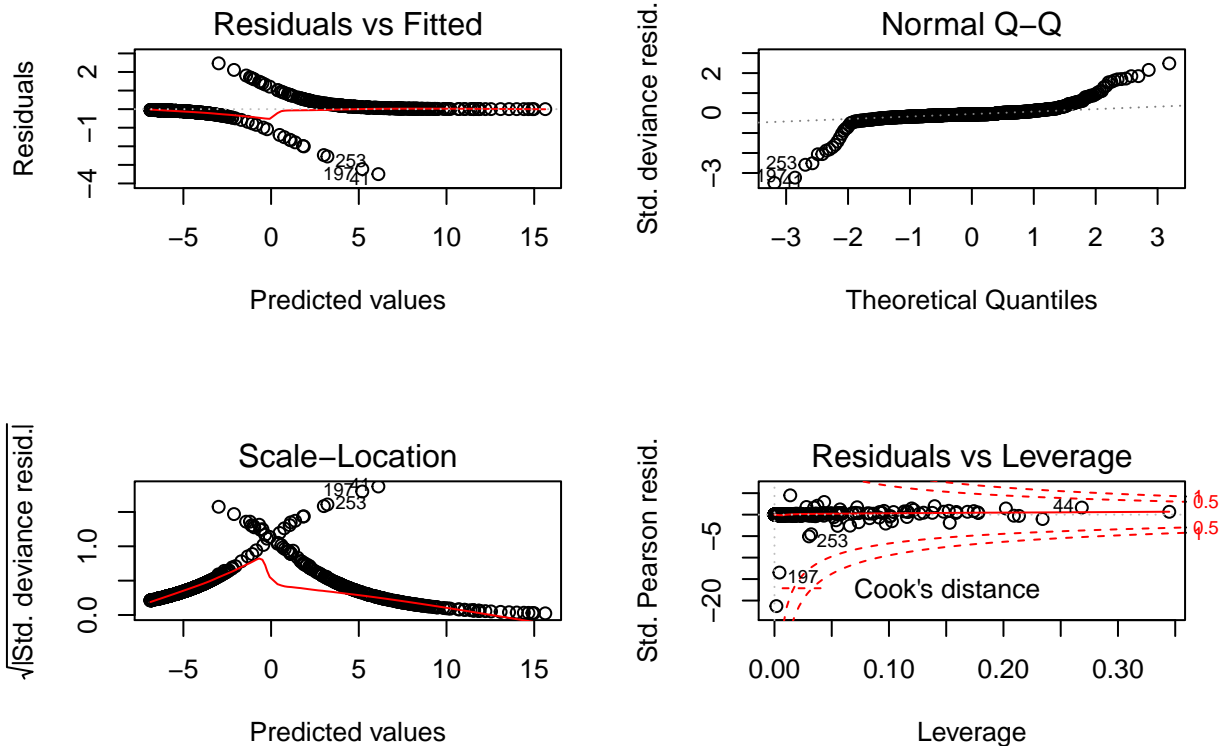
We start our model building process by the most general case; we include every predictor in our model.

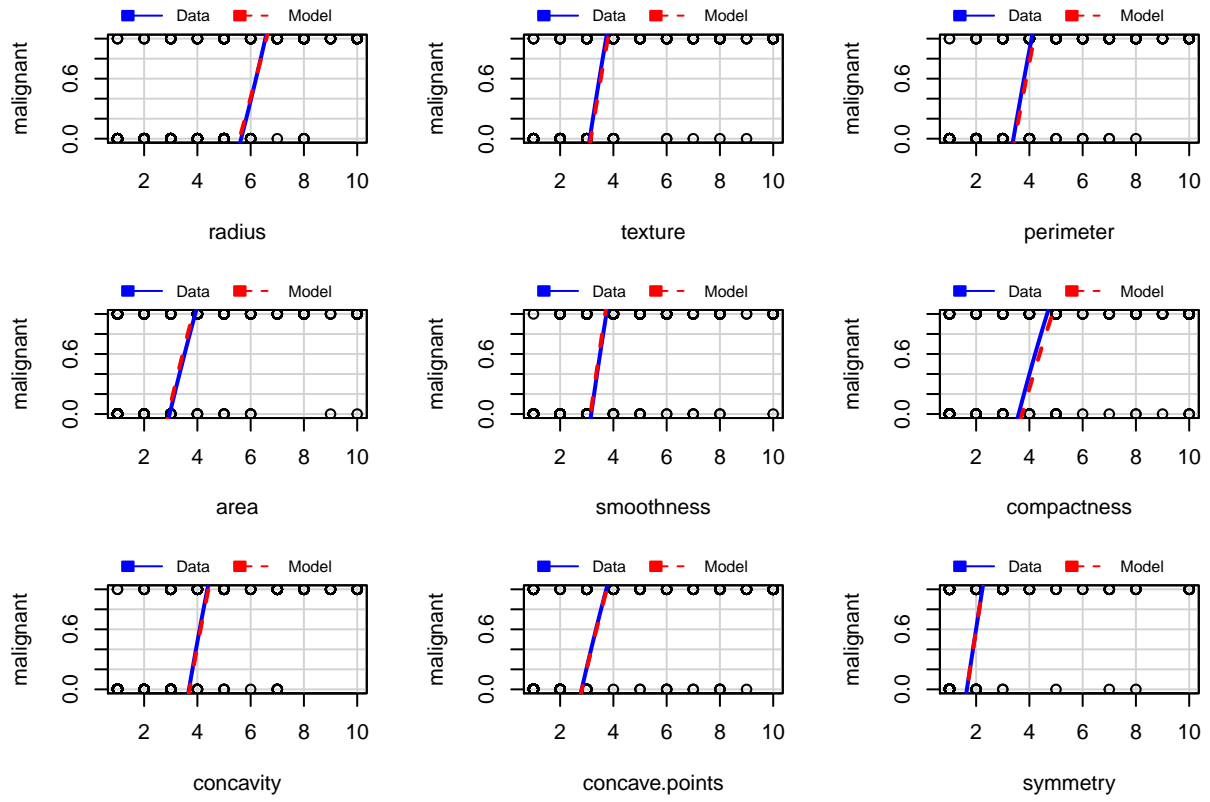
```
m1<-glm(malignant~radius+texture+perimeter+area+smoothness+compactness+concavity
+concave.points+symmetry,family=binomial)
summary(m1)
```

```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      smoothness + compactness + concavity + concave.points + symmetry,
##      family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4978  -0.1342  -0.0718   0.0325   2.4651
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```

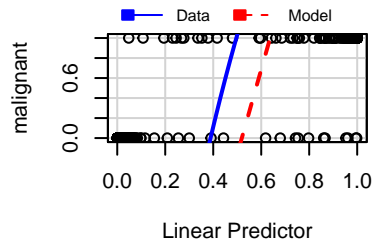
```
## (Intercept)      -9.525617    1.029386   -9.254   < 2e-16 ***
## radius           0.525974    0.128993    4.078  4.55e-05 ***
## texture          0.002369    0.183335    0.013  0.98969
## perimeter        0.355460    0.203985    1.743  0.08141 .
## area            0.238993    0.112556    2.123  0.03373 *
## smoothness       0.068416    0.147221    0.465  0.64213
## compactness      0.351942    0.085565    4.113  3.90e-05 ***
## concavity        0.424393    0.154280    2.751  0.00594 **
## concave.points   0.136815    0.101169    1.352  0.17627
## symmetry         0.543867    0.301147    1.806  0.07092 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 121.74  on 689  degrees of freedom
## AIC: 141.74
##
## Number of Fisher Scoring iterations: 8
```

```
par(mfrow=c(2,2))
plot(m1)
```





Marginal Model Plots



There appears to be a difference between the data and the predicted models.

Can we check if our model `m1` is significant? Formally we have the test

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_9 = 0 \quad vs. \quad H_A : H_0 \text{ is false.}$$

From our summary output, $G^2 = 900.53 - 121.74 = 778.8$, and with 9 degrees of freedom,

$$P\text{-value} = P(\chi_9^2 > 778.8) = 0.$$

Therefore we can reject the null hypothesis and can conclude there is some interaction between the response and our predictors.

```
summary(powerTransform(cbind(radius,texture,perimeter,area,smoothness,compactness,
                             concavity,concave.points,symmetry)))
```

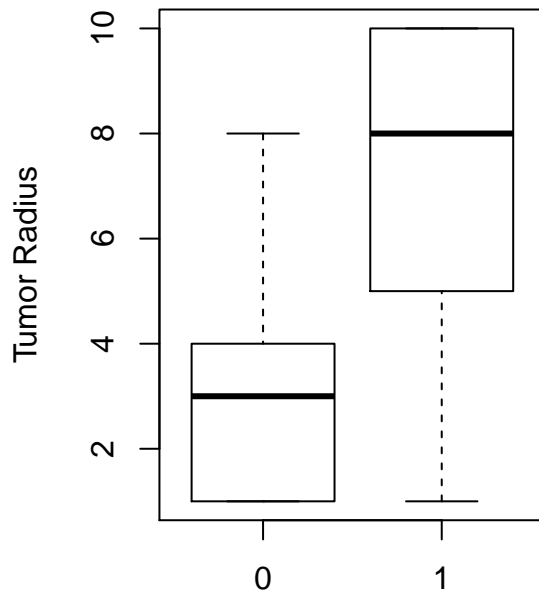
```
## bcPower Transformations to Multinormality
##           Est Power Rounded Pwr Wald Lwr Bnd Wald Up Bnd
## radius      0.5295      0.50    0.4330    0.6259
## texture     -0.4048     -0.33   -0.4976   -0.3120
## perimeter   -0.2434     -0.33   -0.3362   -0.1507
## area        -0.8173     -0.82   -0.9443   -0.6903
## smoothness  -0.2724     -0.33   -0.3623   -0.1826
## compactness -0.6654     -0.67   -0.7796   -0.5511
## concavity    0.1935      0.19    0.1007    0.2864
## concave.points -1.0065    -1.00   -1.1393   -0.8737
## symmetry    -4.1146     -4.11   -4.4597   -3.7694
```

```
##
## Likelihood ratio test that transformation parameters are equal to 0
## (all log transformations)
##                               LRT df      pval
## LR test, lambda = (0 0 0 0 0 0 0 0 0) 2132.159  9 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                               LRT df      pval
## LR test, lambda = (1 1 1 1 1 1 1 1 1) 8521.286  9 < 2.22e-16
```

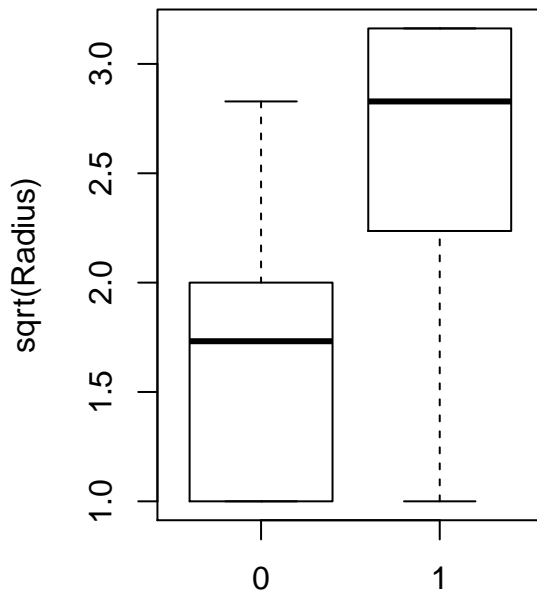
The Box-Cox transformation technique suggest a few transformations; here we opt to:

- i. Take the square root of radius (X_1),
- ii. Raise symmetry to the -4 (X_9),
- iii. Take the inverse of area (X_4),
- iv. Take the inverse of concave points (X_8).

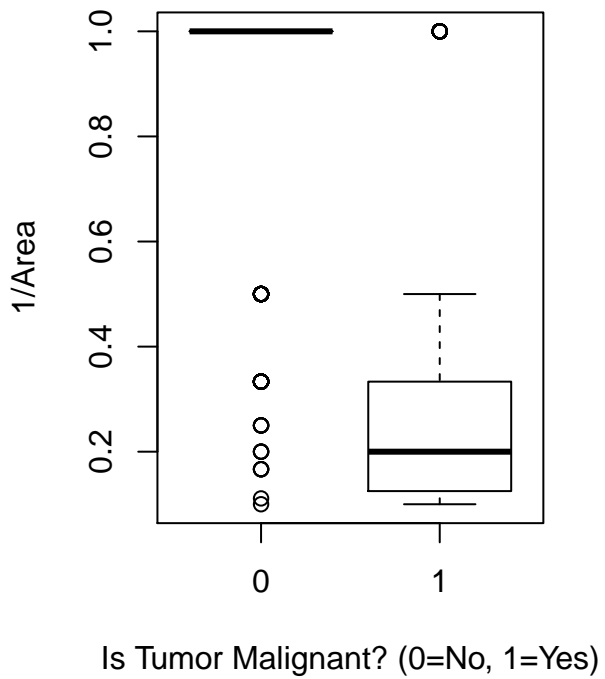
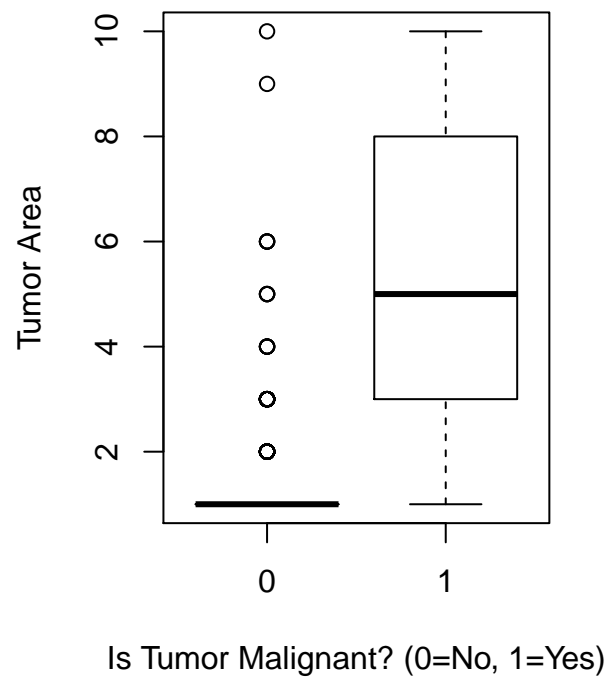
```
nradius<-sqrt(radius)
narea<-(area)^(-1)
nconcave.points<-(concave.points)^(-1)
nsymmetry<-(symmetry)^(-4)
```

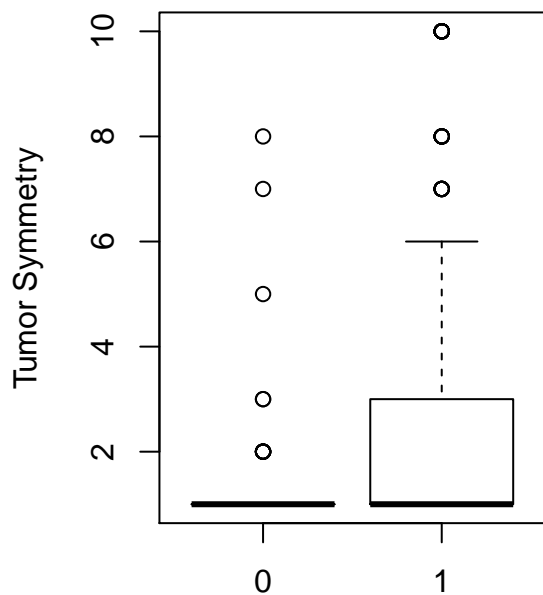


Is Tumor Malignant? (0=No, 1=Yes)

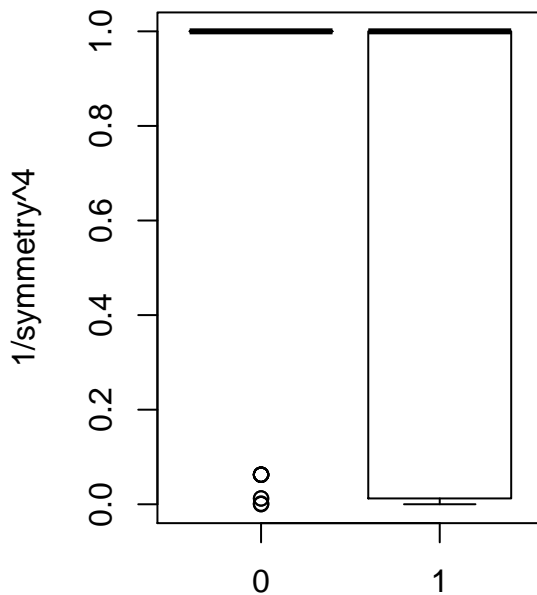


Is Tumor Malignant? (0=No, 1=Yes)





Is Tumor Malignant? (0=No, 1=Yes)



Is Tumor Malignant? (0=No, 1=Yes)

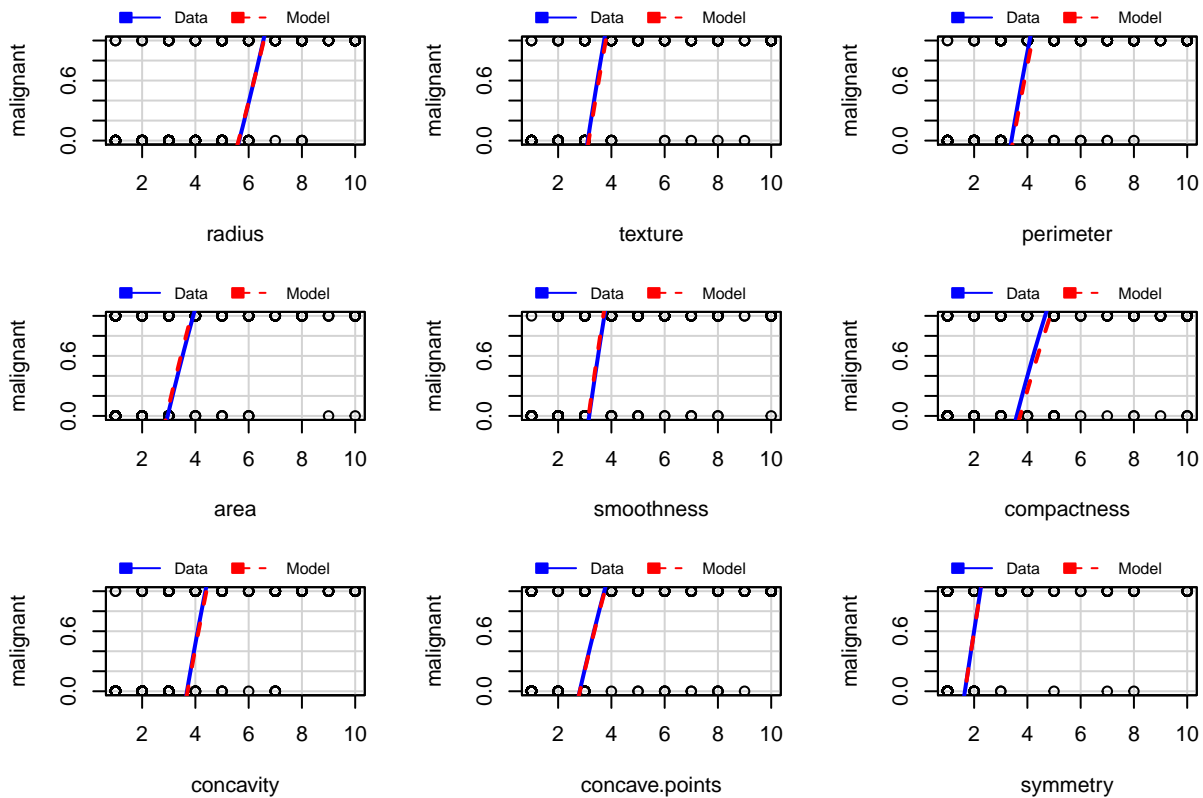
We will examine 4 new models, each with a single transformation.

```
#Added nRadius to m1.
m2<-glm(malignant~radius+texture+perimeter+area+smoothness+compactness+concavity
+concave.points+symmetry+nradius,family=binomial)
summary(m2)
```

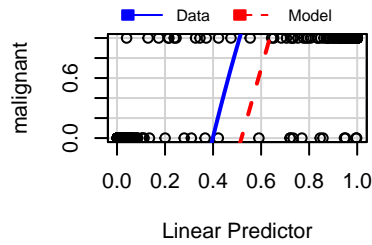
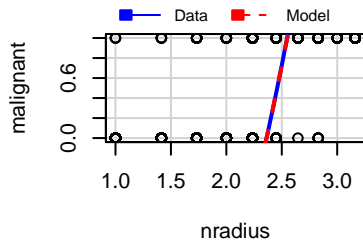
```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      smoothness + compactness + concavity + concave.points + symmetry +
##      nradius, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4906  -0.1326  -0.0869   0.0267   2.5377
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.57025    3.22639  -2.036  0.04171 *
## radius         1.20191    0.75578   1.590  0.11177
## texture       -0.01532    0.18561  -0.083  0.93421
## perimeter      0.37113    0.20902   1.776  0.07580 .
## area           0.24590    0.11452   2.147  0.03177 *
## smoothness     0.08434    0.14943   0.564  0.57247
## compactness    0.34707    0.08522   4.073 4.65e-05 ***
```

```
## concavity      0.42331    0.15401    2.749  0.00599 **
## concave.points 0.14101    0.10169    1.387  0.16554
## symmetry      0.47970    0.28725    1.670  0.09493 .
## nradius       -2.89850    3.13787   -0.924  0.35564
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 120.92  on 688  degrees of freedom
## AIC: 142.92
##
## Number of Fisher Scoring iterations: 8
```

`mmps(m2)`



Marginal Model Plots

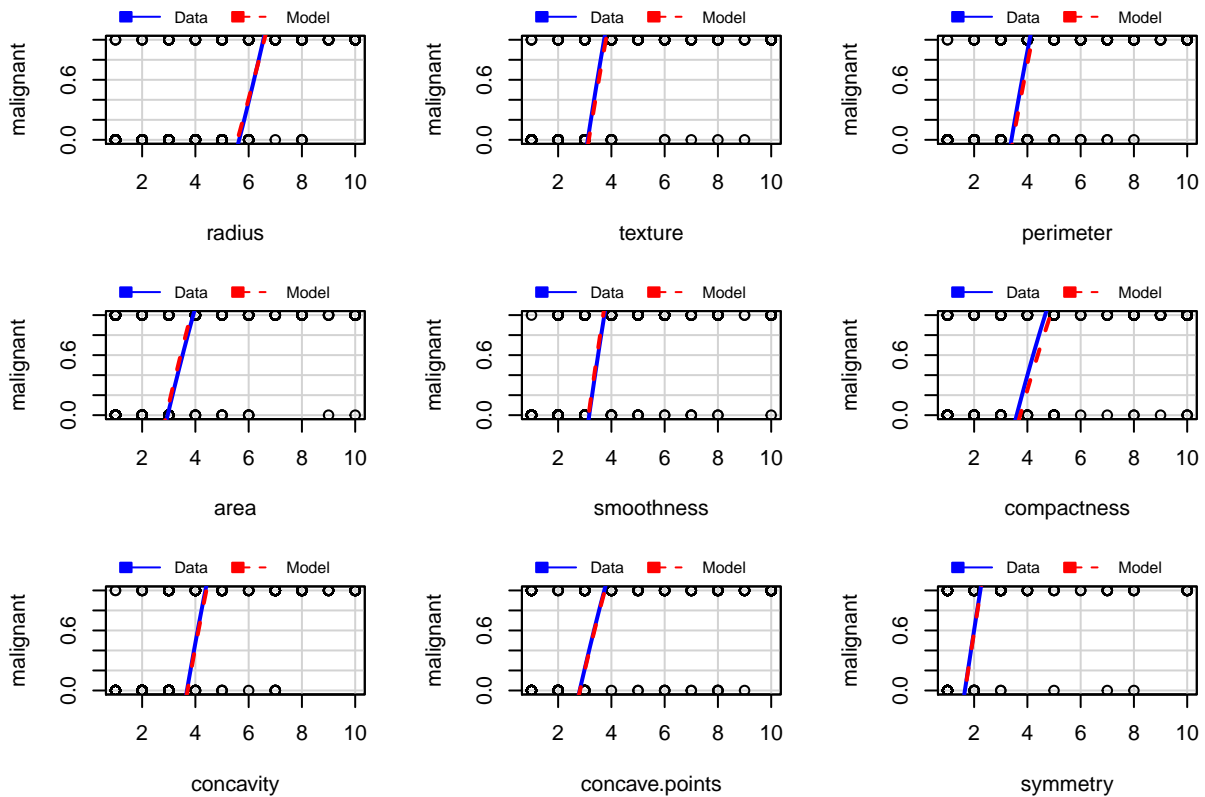


```
#Added nSymmetry to m1.
m3<-glm(malignant~radius+texture+perimeter+area+smoothness+compactness+concavity
+concave.points+symmetry+nsymmetry,family=binomial)
summary(m3)
```

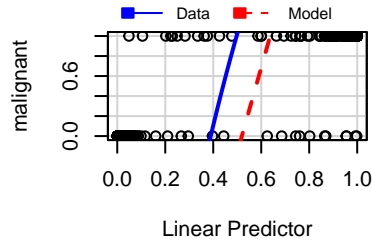
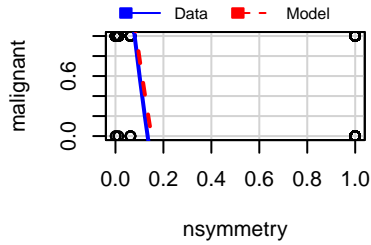
```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      smoothness + compactness + concavity + concave.points + symmetry +
##      nsymmetry, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5113  -0.1356  -0.0703   0.0297   2.4584
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -10.26096    2.29892  -4.463 8.07e-06 ***
## radius         0.54137    0.13638   3.970 7.20e-05 ***
## texture        0.01015    0.18610   0.055 0.95649
## perimeter      0.34290    0.20809   1.648 0.09939 .
## area           0.23672    0.11293   2.096 0.03606 *
## smoothness     0.06815    0.14742   0.462 0.64388
## compactness    0.35618    0.08643   4.121 3.78e-05 ***
## concavity      0.42412    0.15389   2.756 0.00585 **
## concave.points 0.14001    0.10174   1.376 0.16880
```

```
## symmetry          0.69199      0.49536      1.397  0.16243
## nsymmetry         0.53117      1.45670      0.365  0.71538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 121.61  on 688  degrees of freedom
## AIC: 143.61
##
## Number of Fisher Scoring iterations: 8
```

`mmps(m3)`



Marginal Model Plots



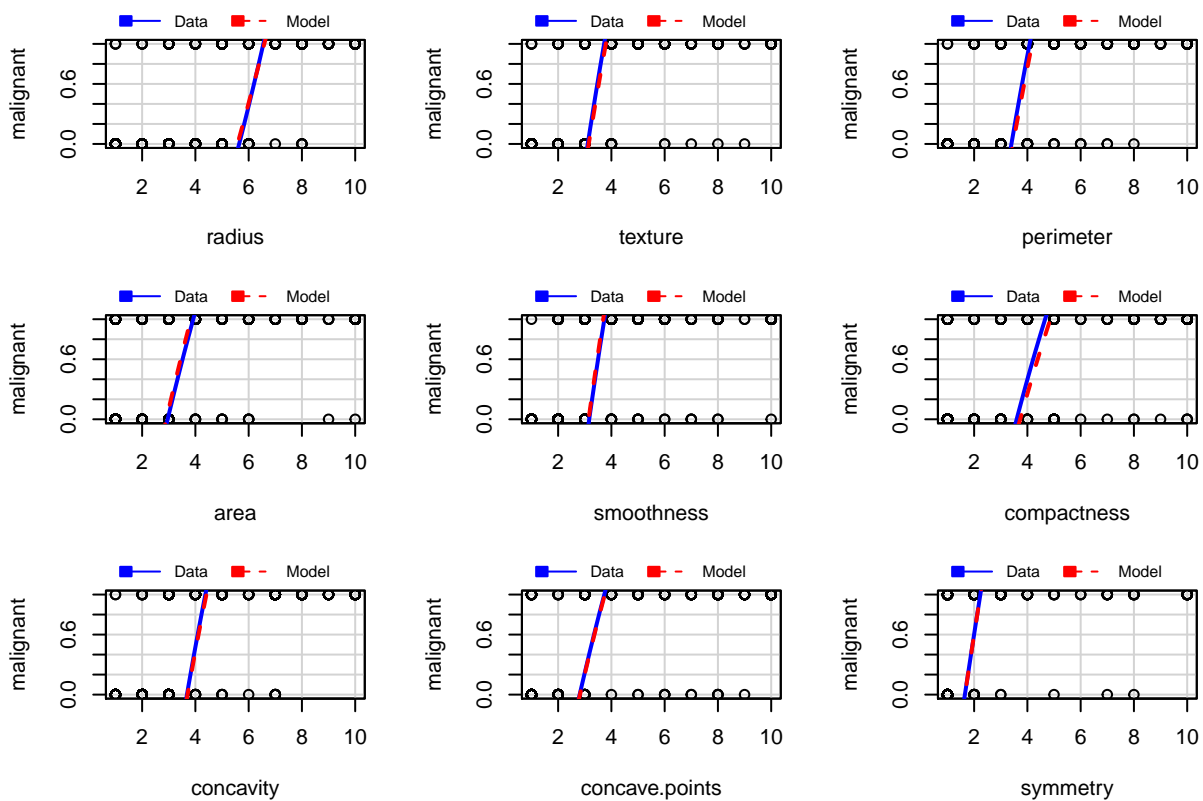
#Added nArea to m1.

```
m4<-glm(malignant~radius+texture+perimeter+area+smoothness+compactness+concavity+concave.points+symmetry+
summary(m4)
```

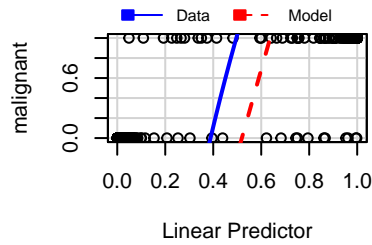
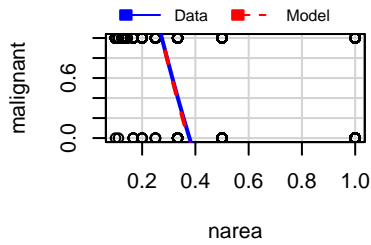
```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      smoothness + compactness + concavity + concave.points + symmetry +
##      narea, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4923  -0.1339  -0.0715   0.0325   2.4579
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -9.463562   1.640873  -5.767 8.05e-09 ***
## radius         0.526474   0.129378   4.069 4.72e-05 ***
## texture        0.002145   0.183485   0.012  0.99067
## perimeter      0.354482   0.205078   1.729  0.08390 .
## area           0.232054   0.182071   1.275  0.20248
## smoothness     0.068586   0.147031   0.466  0.64088
## compactness    0.351498   0.086025   4.086 4.39e-05 ***
## concavity      0.423575   0.155126   2.731  0.00632 **
## concave.points 0.136803   0.101147   1.353  0.17621
## symmetry       0.544700   0.301343   1.808  0.07067 .
```

```
## narea          -0.061565    1.270579   -0.048   0.96135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 121.74  on 688  degrees of freedom
## AIC: 143.74
##
## Number of Fisher Scoring iterations: 8
```

`mmps(m4)`



Marginal Model Plots



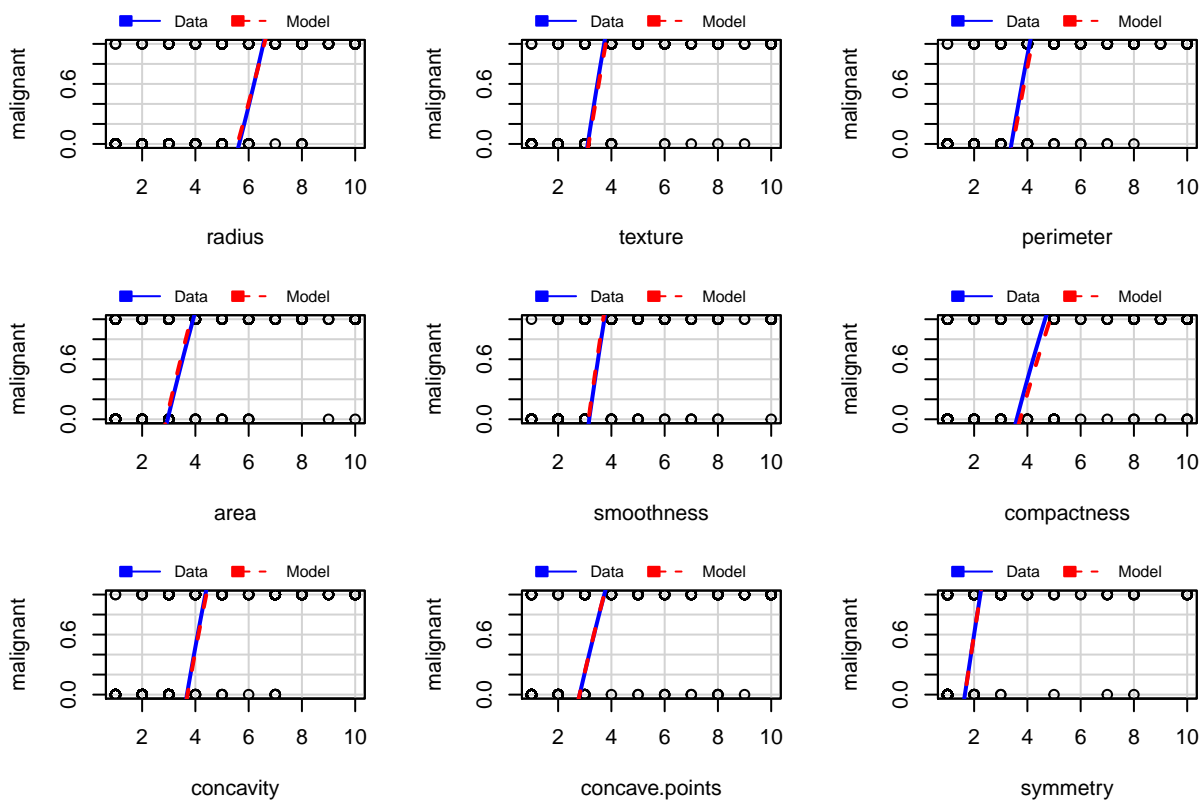
#Added nConcave Points to m1.

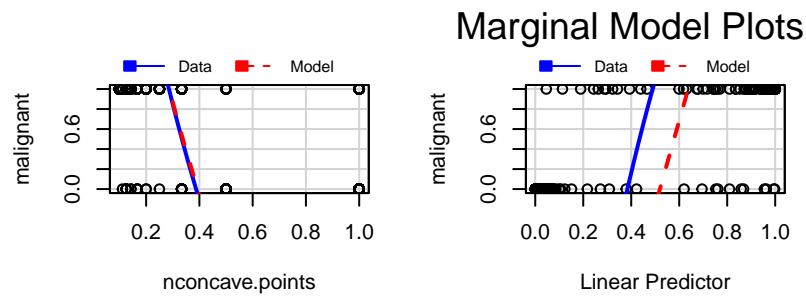
```
m5<-glm(malignant~radius+texture+perimeter+area+smoothness+compactness+concavity+concave.points+symmetry,
summary(m5))
```

```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      smoothness + compactness + concavity + concave.points + symmetry +
##      nconcave.points, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4799  -0.1323  -0.0710   0.0328   2.4815
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -9.154771   1.751317  -5.227 1.72e-07 ***
## radius         0.524501   0.129273   4.057 4.96e-05 ***
## texture        0.006014   0.182529   0.033 0.97372
## perimeter      0.345514   0.206269   1.675 0.09392 .
## area           0.234008   0.113865   2.055 0.03987 *
## smoothness     0.070064   0.147293   0.476 0.63430
## compactness    0.351102   0.085579   4.103 4.08e-05 ***
## concavity      0.424808   0.154213   2.755 0.00588 **
## concave.points 0.099354   0.175555   0.566 0.57143
## symmetry       0.539180   0.304358   1.772 0.07647 .
```

```
## nconcave.points -0.342299  1.320651 -0.259  0.79549
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 121.68  on 688  degrees of freedom
## AIC: 143.68
##
## Number of Fisher Scoring iterations: 8
```

`mmps(m5)`





Even with our transformed predictors, we still see a gap between the data and our model.

Model Validity/Differences

```
anova(m1,m2,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: malignant ~ radius + texture + perimeter + area + smoothness +
## compactness + concavity + concave.points + symmetry
## Model 2: malignant ~ radius + texture + perimeter + area + smoothness +
## compactness + concavity + concave.points + symmetry + nradius
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      689      121.74
## 2      688      120.92  1  0.82727  0.3631
```

```
anova(m1,m3,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: malignant ~ radius + texture + perimeter + area + smoothness +
## compactness + concavity + concave.points + symmetry
## Model 2: malignant ~ radius + texture + perimeter + area + smoothness +
```

```
## compactness + concavity + concave.points + symmetry + nsymmetry
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1 689 121.74
## 2 688 121.61 1 0.13274 0.7156
```

```
anova(m1,m4,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: malignant ~ radius + texture + perimeter + area + smoothness +
## compactness + concavity + concave.points + symmetry
## Model 2: malignant ~ radius + texture + perimeter + area + smoothness +
## compactness + concavity + concave.points + symmetry + narea
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1 689 121.74
## 2 688 121.74 1 0.0023446 0.9614
```

```
anova(m1,m5,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: malignant ~ radius + texture + perimeter + area + smoothness +
## compactness + concavity + concave.points + symmetry
## Model 2: malignant ~ radius + texture + perimeter + area + smoothness +
## compactness + concavity + concave.points + symmetry + nconcave.points
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1 689 121.74
## 2 688 121.68 1 0.066745 0.7961
```

Our ANOVA test suggests that every model is significantly different from the general model.

```
vif(m1)
```

```
## radius texture perimeter area smoothness
## 1.156697 2.764041 2.599061 1.217030 1.413033
## compactness concavity concave.points symmetry
## 1.186902 1.244031 1.308941 1.028365
```

```
range(radius)
```

```
## [1] 1 10
```

```
range(nradius)
```

```
## [1] 1.000000 3.162278
```

```
range(texture)
```

```
## [1] 1 10
```

```
range(perimeter)
```

```
## [1] 1 10
```

```
range(area)
```

```
## [1] 1 10
```

```
range(narea)
```

```
## [1] 0.1 1.0
```

```
range(smoothness)
```

```
## [1] 1 10
```

```
range(compactness)
```

```
## [1] 1 10
```

```
range(concavity)
```

```
## [1] 1 10
```

```
range(concave.points)
```

```
## [1] 1 10
```

```
range(nconcave.points)
```

```
## [1] 0.1 1.0
```

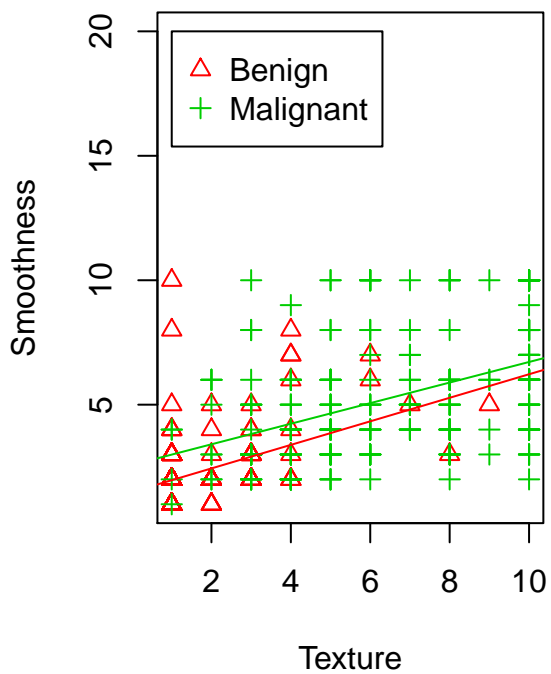
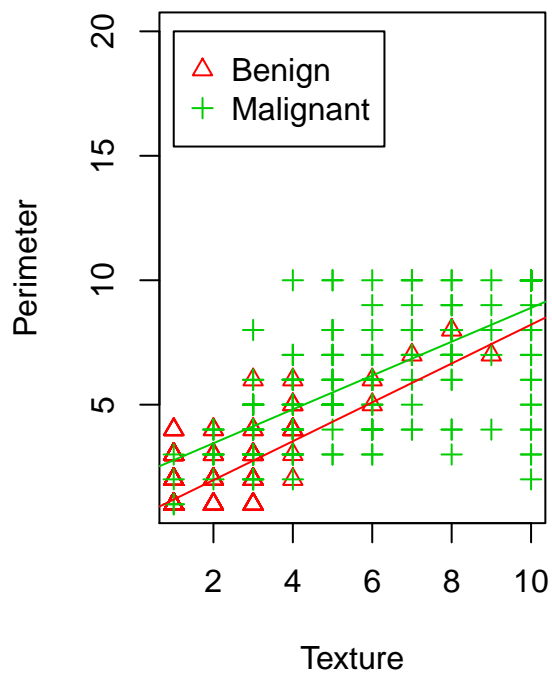
```
range(symmetry)
```

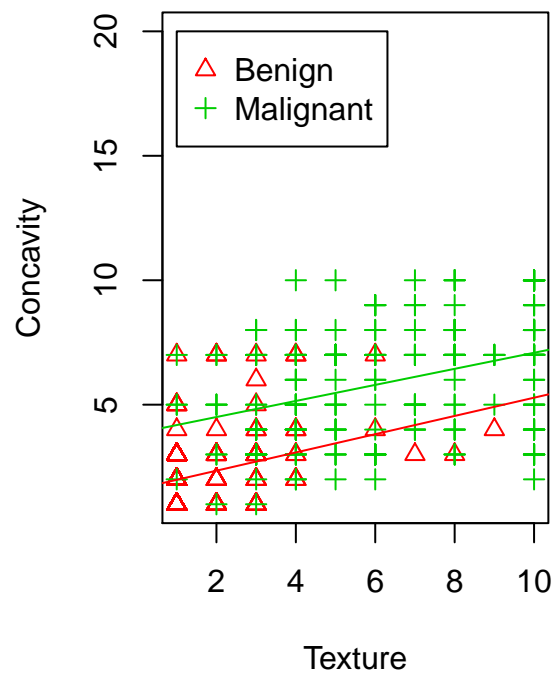
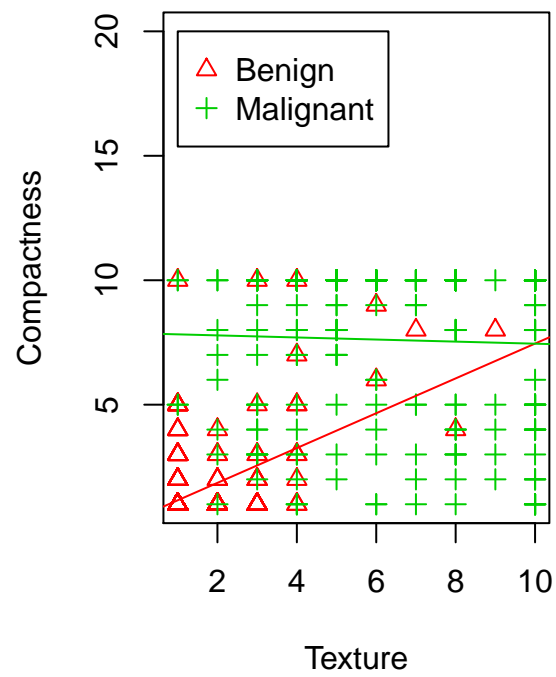
```
## [1] 1 10
```

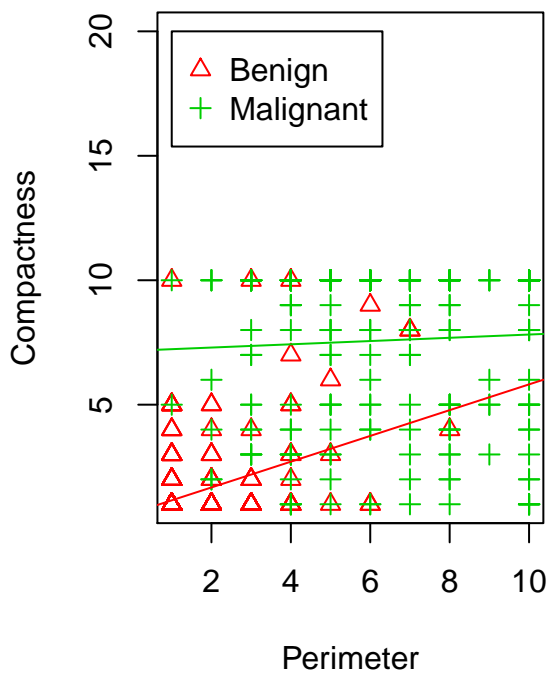
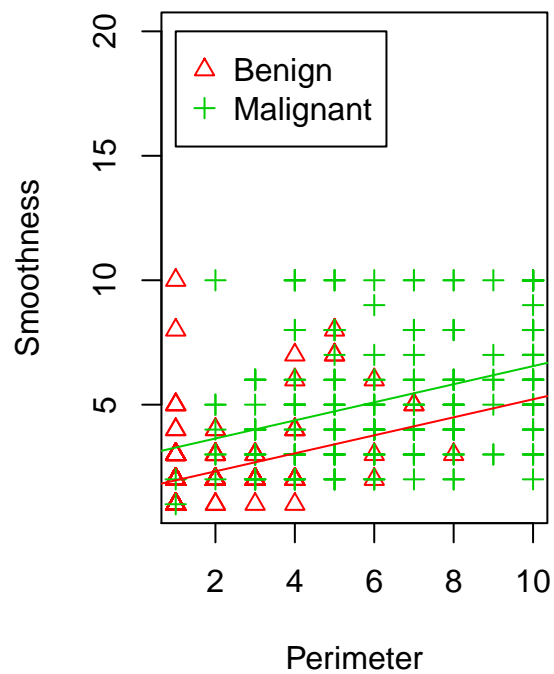
```
range(nsymmetry)
```

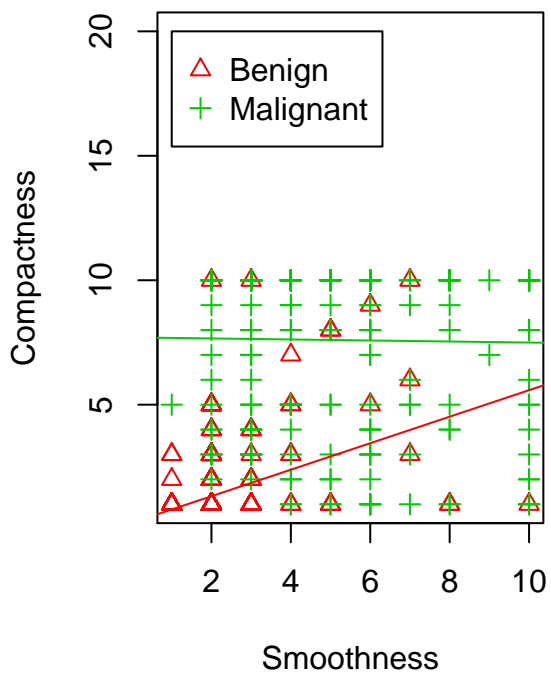
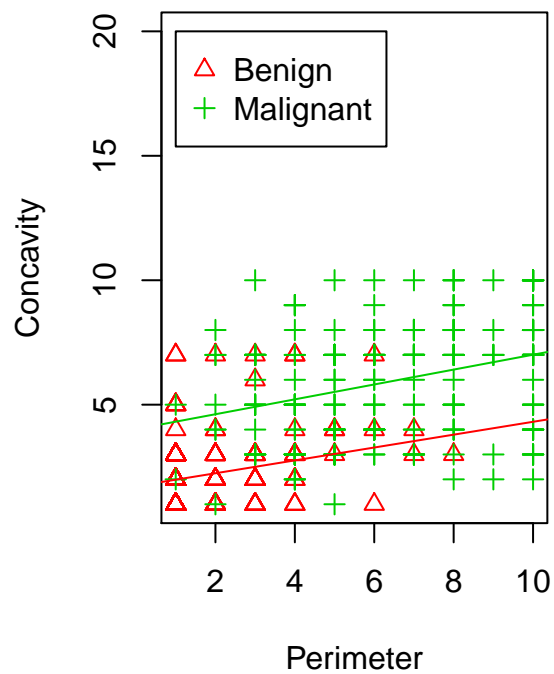
```
## [1] 1e-04 1e+00
```

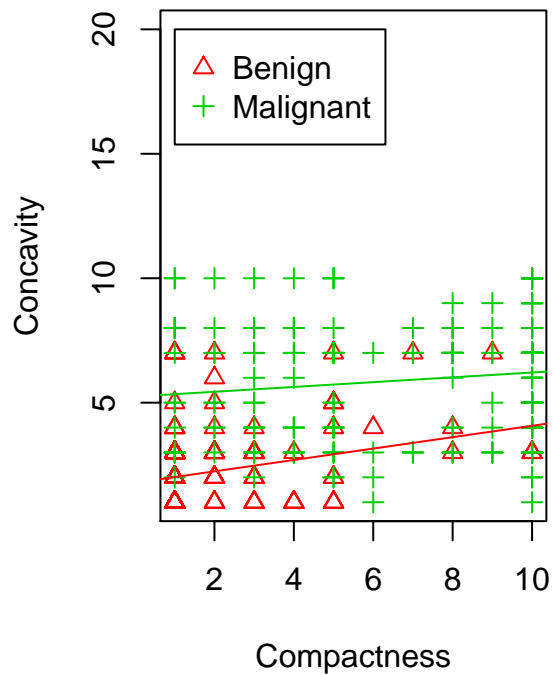
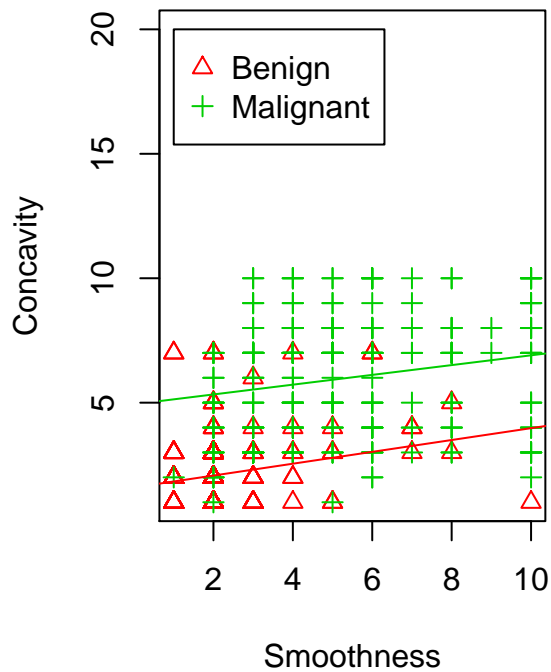
Place holder for pairplots











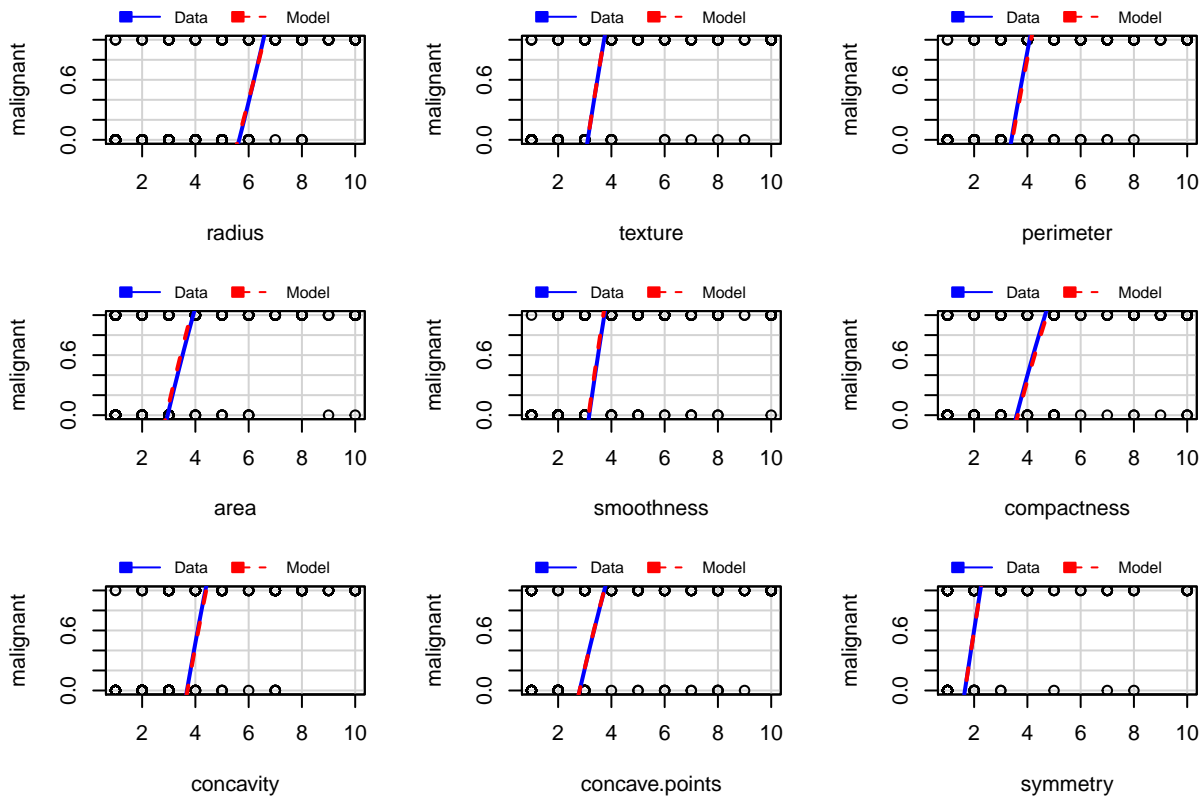
```
#Compactness:Texture
```

```
m6<-glm(malignant~radius+texture+perimeter+area+smoothness+compactness+concavity
+concave.points+symmetry+compactness:texture,family=binomial)
summary(m6)
```

```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      smoothness + compactness + concavity + concave.points + symmetry +
##      compactness:texture, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2118  -0.0979  -0.0523   0.0481   2.2405
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -10.85300    1.27714  -8.498  < 2e-16 ***
## radius         0.47983    0.13258   3.619  0.000295 ***
## texture        0.67659    0.31335   2.159  0.030833 *
## perimeter      0.31355    0.20404   1.537  0.124365
## area           0.21213    0.11120   1.908  0.056429 .
## smoothness    -0.01722    0.14936  -0.115  0.908217
## compactness     0.78840    0.16921   4.659  3.18e-06 ***
## concavity      0.41504    0.16233   2.557  0.010565 *
## concave.points 0.09295    0.10119   0.919  0.358330
```

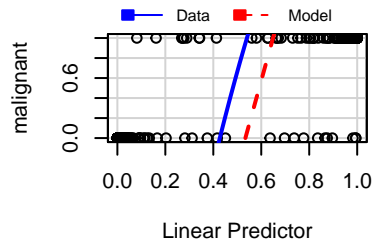
```
## symmetry          0.59576    0.29754    2.002 0.045256 *
## texture:compactness -0.11232    0.03642   -3.084 0.002040 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 112.18  on 688  degrees of freedom
## AIC: 134.18
##
## Number of Fisher Scoring iterations: 8
```

`mmps(m6)`



```
## Warning in mmps(m6): Interactions and/or factors skipped
```

Marginal Model Plots

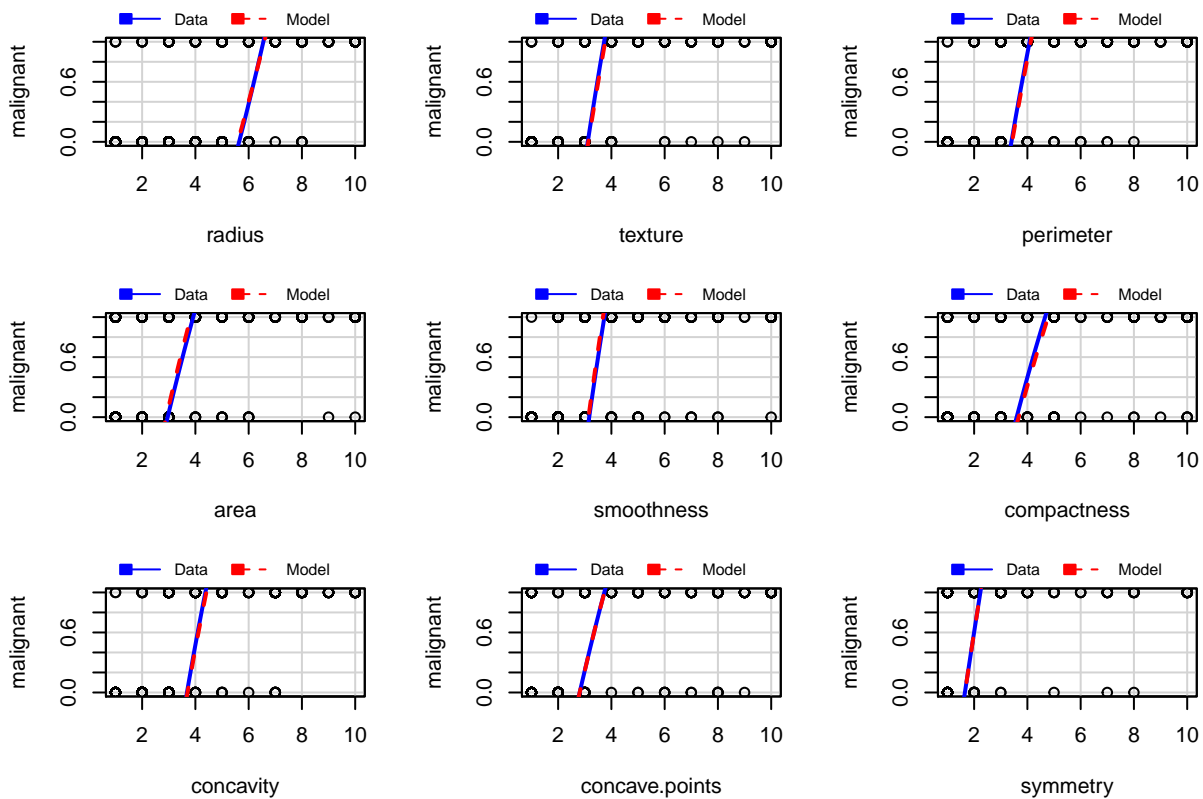


```
#Compactness:Perimeter
m7<-glm(malignant~radius+texture+perimeter+area+smoothness+compactness+concavity
+concave.points+symmetry+compactness:perimeter,family=binomial)
summary(m7)
```

```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      smoothness + compactness + concavity + concave.points + symmetry +
##      compactness:perimeter, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2506  -0.1069  -0.0563   0.0513   2.2676
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -10.53527     1.23001  -8.565 < 2e-16 ***
## radius           0.47039     0.12869   3.655 0.000257 ***
## texture          0.03283     0.18113   0.181 0.856155
## perimeter        0.78647     0.29820   2.637 0.008354 **
## area             0.21769     0.10847   2.007 0.044767 *
## smoothness       0.02474     0.14404   0.172 0.863648
## compactness      0.72698     0.18425   3.946 7.96e-05 ***
## concavity        0.41003     0.15881   2.582 0.009826 **
## concave.points   0.11428     0.10101   1.131 0.257908
```

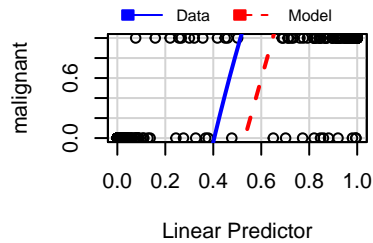
```
## symmetry          0.55054    0.30806    1.787 0.073921 .
## perimeter:compactness -0.09093    0.03900   -2.331 0.019740 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 116.71  on 688  degrees of freedom
## AIC: 138.71
##
## Number of Fisher Scoring iterations: 8
```

```
mmps(m7)
```



```
## Warning in mmps(m7): Interactions and/or factors skipped
```

Marginal Model Plots

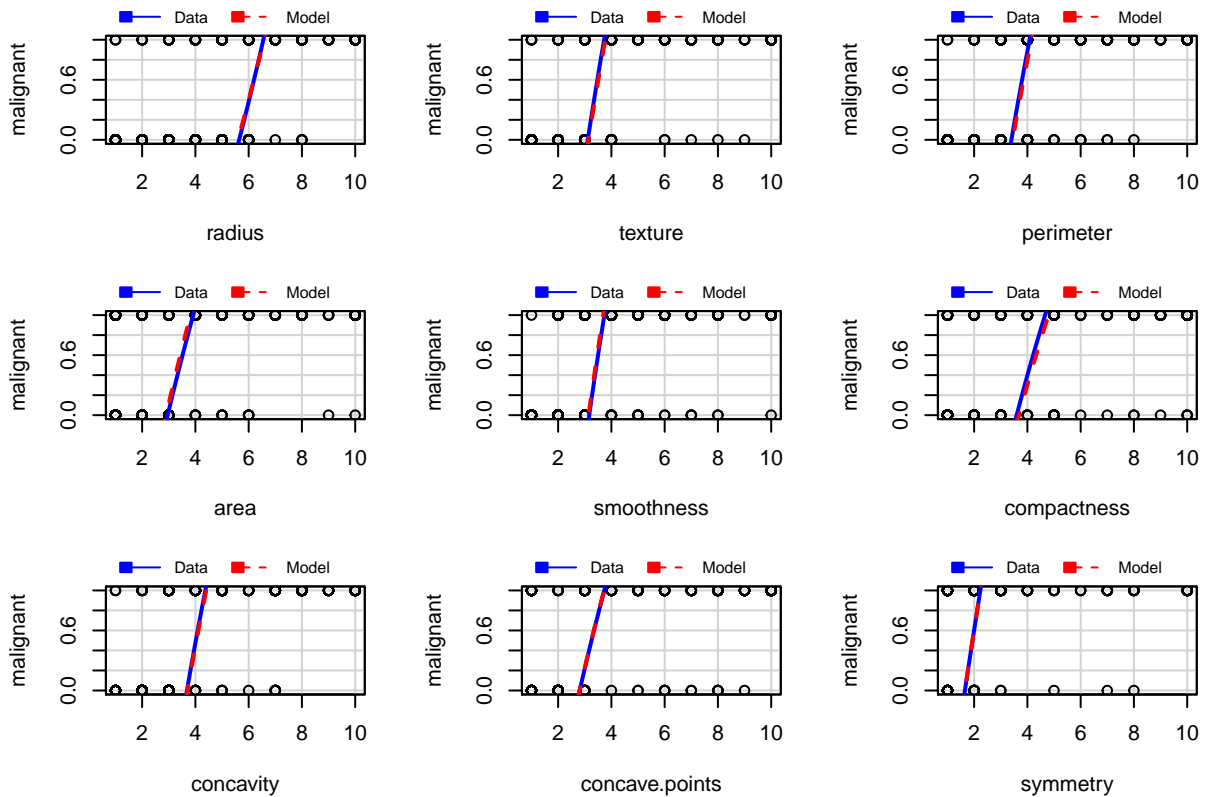


```
#Compactness:Smoothness
m8<-glm(malignant~radius+texture+perimeter+area+smoothness+compactness+concavity
+concave.points+symmetry+compactness:smoothness,family=binomial)
summary(m8)
```

```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      smoothness + compactness + concavity + concave.points + symmetry +
##      compactness:smoothness, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2455  -0.1102  -0.0617   0.0402   2.2131
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -10.72246     1.22479  -8.755  < 2e-16 ***
## radius           0.50085     0.13408   3.735 0.000187 ***
## texture        -0.01097     0.17779  -0.062 0.950787
## perimeter       0.35646     0.20142   1.770 0.076765 .
## area           0.19016     0.11289   1.685 0.092078 .
## smoothness      0.45466     0.20800   2.186 0.028823 *
## compactness     0.79449     0.19267   4.124 3.73e-05 ***
## concavity       0.42106     0.15913   2.646 0.008145 **
## concave.points  0.10912     0.09882   1.104 0.269474
```

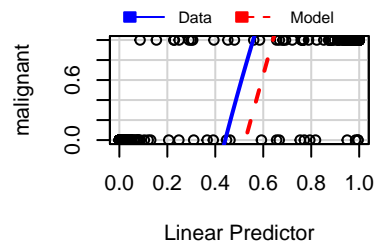
```
## symmetry          0.53508    0.30243    1.769 0.076850 .
## smoothness:compactness -0.09380    0.03433   -2.732 0.006296 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 115.05  on 688  degrees of freedom
## AIC: 137.05
##
## Number of Fisher Scoring iterations: 8
```

```
mmps(m8)
```



```
## Warning in mmps(m8): Interactions and/or factors skipped
```

Marginal Model Plots



Model Selection Process

Backwards Selection

```
backAIC<-step(m6,direction = "backward", k = 2)
```

```
## Start:  AIC=134.18
## malignant ~ radius + texture + perimeter + area + smoothness +
##           compactness + concavity + concave.points + symmetry + compactness:texture
##
##           Df Deviance    AIC
## - smoothness      1  112.19 132.19
## - concave.points    1  113.04 133.04
## <none>              112.18 134.18
## - perimeter        1  114.45 134.46
## - area              1  115.84 135.84
## - symmetry          1  117.17 137.17
## - concavity         1  119.17 139.17
## - texture:compactness 1  121.74 141.74
## - radius            1  128.19 148.19
##
## Step:  AIC=132.19
## malignant ~ radius + texture + perimeter + area + compactness +
```



```

##      concavity + concave.points + symmetry + texture:compactness
##
##              Df Deviance    AIC
## - concave.points      1   113.05 131.04
## <none>                  112.19 132.19
## - perimeter           1   114.48 132.48
## - area                 1   115.85 133.85
## - symmetry             1   117.17 135.17
## - concavity            1   119.18 137.18
## - texture:compactness  1   121.96 139.96
## - radius               1   128.20 146.20
##
## Step:  AIC=131.05
## malignant ~ radius + texture + perimeter + area + compactness +
##      concavity + symmetry + texture:compactness
##
##              Df Deviance    AIC
## <none>                  113.05 131.04
## - perimeter           1   116.25 132.25
## - area                 1   117.07 133.07
## - symmetry             1   118.07 134.07
## - concavity            1   121.35 137.35
## - texture:compactness  1   124.06 140.06
## - radius               1   129.12 145.12

m9<-glm(malignant ~ radius + texture + perimeter + area + compactness +
      concavity + symmetry + texture:compactness, family = binomial)

backBIC<-step(m6,direction = "backward",k=log(n))

## Start:  AIC=184.22
## malignant ~ radius + texture + perimeter + area + smoothness +
##      compactness + concavity + concave.points + symmetry + compactness:texture
##
##              Df Deviance    AIC
## - smoothness           1   112.19 177.69
## - concave.points        1   113.04 178.53
## - perimeter             1   114.45 179.95
## - area                   1   115.84 181.34
## - symmetry              1   117.17 182.66
## <none>                   112.18 184.22
## - concavity             1   119.17 184.67
## - texture:compactness   1   121.74 187.24
## - radius                1   128.19 193.69
##
## Step:  AIC=177.69
## malignant ~ radius + texture + perimeter + area + compactness +
##      concavity + concave.points + symmetry + texture:compactness
##
##              Df Deviance    AIC
## - concave.points        1   113.05 171.99
## - perimeter             1   114.48 173.43
## - area                   1   115.85 174.80
## - symmetry              1   117.17 176.11

```

```

## <none>                112.19 177.69
## - concavity           1   119.18 178.13
## - texture:compactness 1   121.96 180.91
## - radius              1   128.20 187.15
##
## Step:  AIC=171.99
## malignant ~ radius + texture + perimeter + area + compactness +
##           concavity + symmetry + texture:compactness
##
##           Df Deviance    AIC
## - perimeter      1   116.25 168.65
## - area            1   117.07 169.46
## - symmetry        1   118.07 170.47
## <none>            113.05 171.99
## - concavity       1   121.35 173.74
## - texture:compactness 1   124.06 176.45
## - radius          1   129.12 181.52
##
## Step:  AIC=168.65
## malignant ~ radius + texture + area + compactness + concavity +
##           symmetry + texture:compactness
##
##           Df Deviance    AIC
## - area            1   120.92 166.77
## - symmetry         1   121.00 166.85
## <none>            116.25 168.65
## - concavity        1   126.51 172.36
## - texture:compactness 1   128.15 174.00
## - radius           1   137.93 183.77
##
## Step:  AIC=166.77
## malignant ~ radius + texture + compactness + concavity + symmetry +
##           texture:compactness
##
##           Df Deviance    AIC
## - symmetry         1   126.35 165.65
## <none>            120.92 166.77
## - concavity        1   132.82 172.12
## - texture:compactness 1   134.18 173.48
## - radius           1   139.26 178.56
##
## Step:  AIC=165.64
## malignant ~ radius + texture + compactness + concavity + texture:compactness
##
##           Df Deviance    AIC
## <none>            126.35 165.65
## - concavity       1   137.26 170.01
## - texture:compactness 1   138.19 170.94
## - radius          1   151.32 184.07

```

```

m10<-glm( malignant ~ radius + texture + compactness +
           concavity + texture:compactness, family = binomial)

```

```

summary(m9)

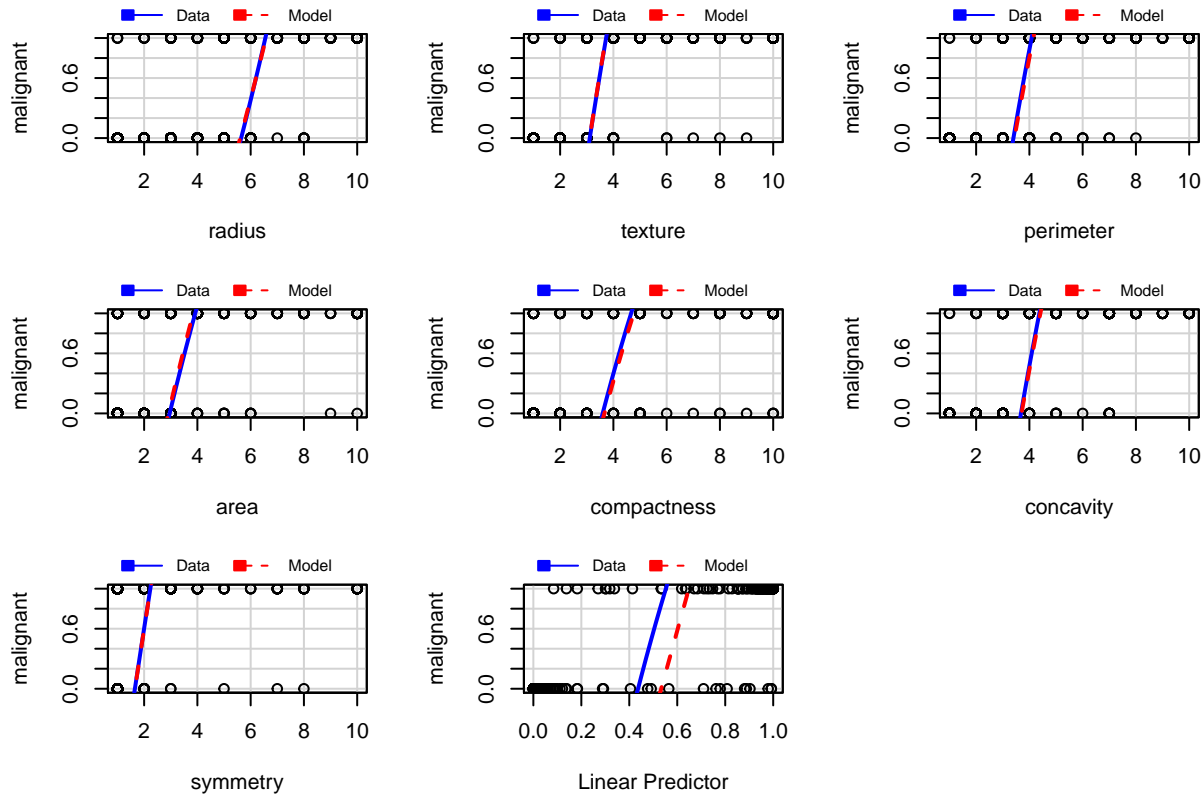
```

```
##
## Call:
## glm(formula = malignant ~ radius + texture + perimeter + area +
##      compactness + concavity + symmetry + texture:compactness,
##      family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.11536  -0.10044  -0.05326   0.05103   2.22270
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -10.90074     1.25832  -8.663 < 2e-16 ***
## radius           0.47840     0.13202   3.624 0.00029 ***
## texture          0.70535     0.28929   2.438 0.01476 *
## perimeter        0.34602     0.19112   1.810 0.07022 .
## area             0.21752     0.10920   1.992 0.04637 *
## compactness      0.80192     0.16456   4.873 1.1e-06 ***
## concavity        0.43769     0.15858   2.760 0.00578 **
## symmetry         0.59683     0.29739   2.007 0.04476 *
## texture:compactness -0.11608     0.03469  -3.346 0.00082 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 900.53  on 698  degrees of freedom
## Residual deviance: 113.05  on 690  degrees of freedom
## AIC: 131.05
##
## Number of Fisher Scoring iterations: 8
```

```
mmpls(m9)
```

```
## Warning in mmpls(m9): Interactions and/or factors skipped
```

Marginal Model Plots



```
summary(m10)
```

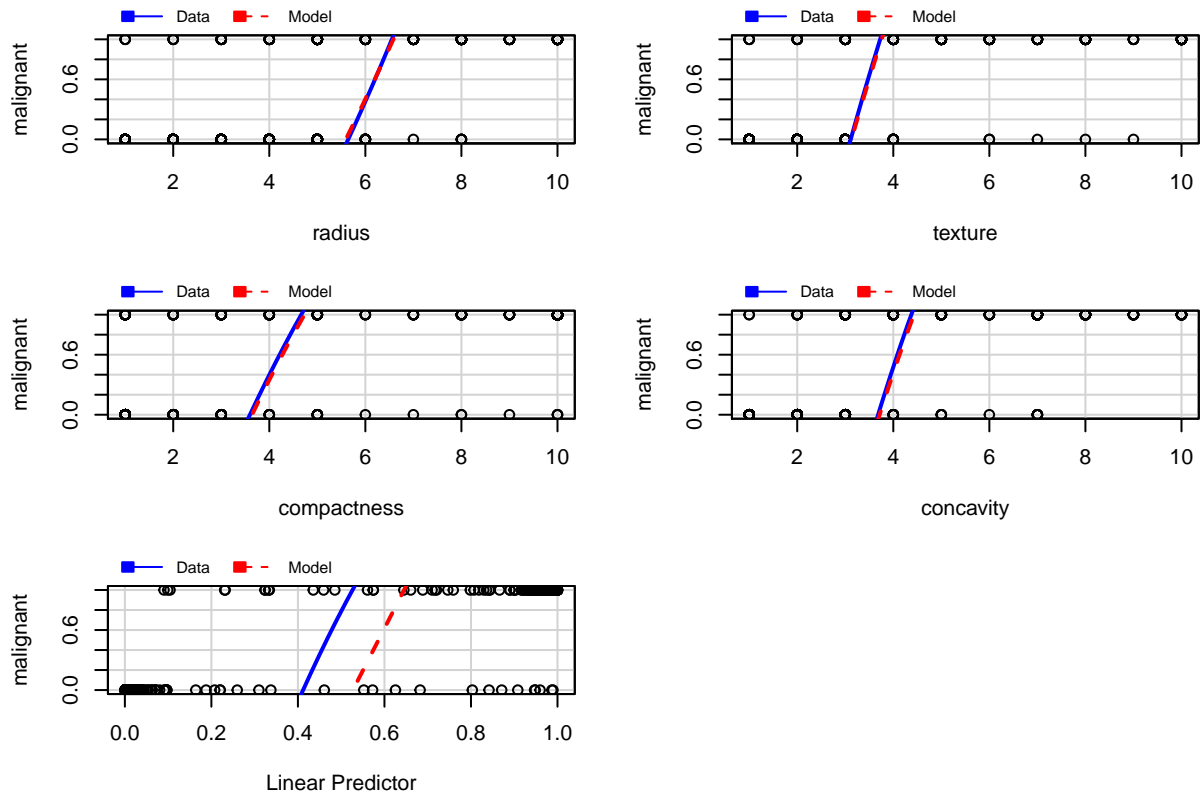
```
##
## Call:
## glm(formula = malignant ~ radius + texture + compactness + concavity +
##       texture:compactness, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.01840  -0.11010  -0.06219   0.07965   2.19198
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -10.01123    1.07420  -9.320  < 2e-16 ***
## radius         0.52069    0.11753   4.430 9.41e-06 ***
## texture        1.08470    0.25768   4.209 2.56e-05 ***
## compactness    0.86964    0.15530   5.600 2.15e-08 ***
## concavity      0.46916    0.14849   3.159 0.001581 **
## texture:compactness -0.11857    0.03387  -3.501 0.000464 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 900.53  on 698  degrees of freedom
```

```
## Residual deviance: 126.35  on 693  degrees of freedom
## AIC: 138.35
##
## Number of Fisher Scoring iterations: 8
```

```
mmps(m10)
```

```
## Warning in mmps(m10): Interactions and/or factors skipped
```

Marginal Model Plots



Final Model and Conclusion

Compiling the deviance, AIC, and BIC, we can compare our 10 models:

Model	Deviance	AIC	BIC
m1	121.7439	141.7439	180.6907
m2	120.9166	142.9166	186.4131
m3	121.6111	143.6111	187.1076
m4	121.7415	143.7415	187.2380
m5	121.6771	143.6771	187.1736
m6	112.1774	134.1774	184.2235
m7	116.7057	138.7057	188.7519
m8	115.0463	137.0463	187.0924
m9	113.0452	131.0452	171.9920
m10	126.3468	138.3468	165.6447

The model we decide on is **m9**, due to the lowest Deviance and AIC. We can formally interpret our model as

```
m9$coefficients
```

```
##      (Intercept)      radius      texture
##      -10.9007404      0.4784014      0.7053489
##      perimeter      area      compactness
##      0.3460187      0.2175191      0.8019175
##      concavity      symmetry texture:compactness
##      0.4376866      0.5968298      -0.1160840
```

$$P(\text{Malignant}) = \frac{\exp(-10.90 + 0.48X_1 + 0.71X_2 + 0.35X_3 + 0.22X_4 + 0.80X_6 + 0.44X_7 + 0.60X_9 - 0.12X_2 : X_6)}{1 + \exp(-10.90 + 0.48X_1 + 0.71X_2 + 0.35X_3 + 0.22X_4 + 0.80X_6 + 0.44X_7 + 0.60X_9 - 0.12X_2 : X_6)}$$

or the log odds, which is

$$\log\left(\frac{P(\text{Malignant})}{1 - P(\text{Malignant})}\right) = -10.90 + 0.48X_1 + 0.71X_2 + 0.35X_3 + 0.22X_4 + 0.80X_6 + 0.44X_7 + 0.60X_9 - 0.12X_2 : X_6$$

If the scale for texture increased by 1, the odds in favour of the tumor being malignant will change by a multiplicative factor of $\exp(0.71) = 2.034$, with other variables being held constant.

For the sake of curiosity, let's consider the mean of our predictors,

```
mean(radius)
```

```
## [1] 4.41774
```

```
mean(texture)
```

```
## [1] 3.134478
```

```
mean(perimeter)
```

```
## [1] 3.207439
```

```
mean(area)
```

```
## [1] 2.806867
```

```
mean(compactness)
```

```
## [1] 3.526466
```

```
mean(concavity)
```

```
## [1] 3.437768
```

```
mean(symmetry)
```

```
## [1] 1.589413
```

```
x<-c( 1, mean(radius), mean(texture), mean(perimeter), mean(area), mean(compactness),  
      mean(concavity), mean(symmetry), mean(compactness)*mean(texture) )
```

```
#Probability of the means:
```

```
probability<-exp( sum(coef(m9)*x) ) / ( 1 + exp( sum(coef(m9)*x) ) )  
probability
```

```
## [1] 0.2977685
```

```
#Log Odds of the means:
```

```
logodds<-sum(coef(m9)*x)  
logodds
```

```
## [1] -0.8579466
```