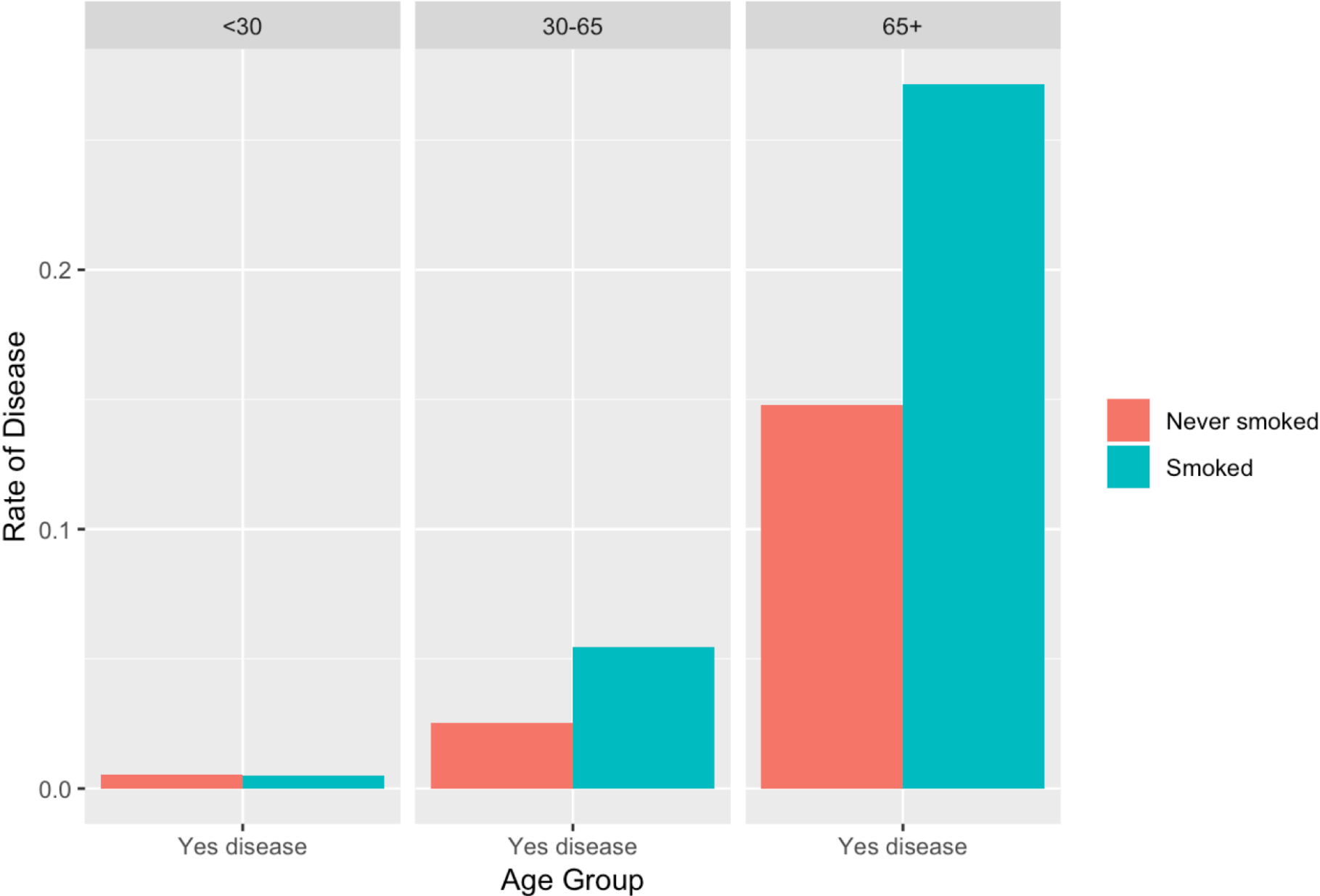


Rate of Disease for Smokers and Non-Smokers by Age Group



```
modell1 <- glm(nmes.data$disease ~ nmes.data$eversmk, family = binomial(link="logit"))
summary(modell1)
```

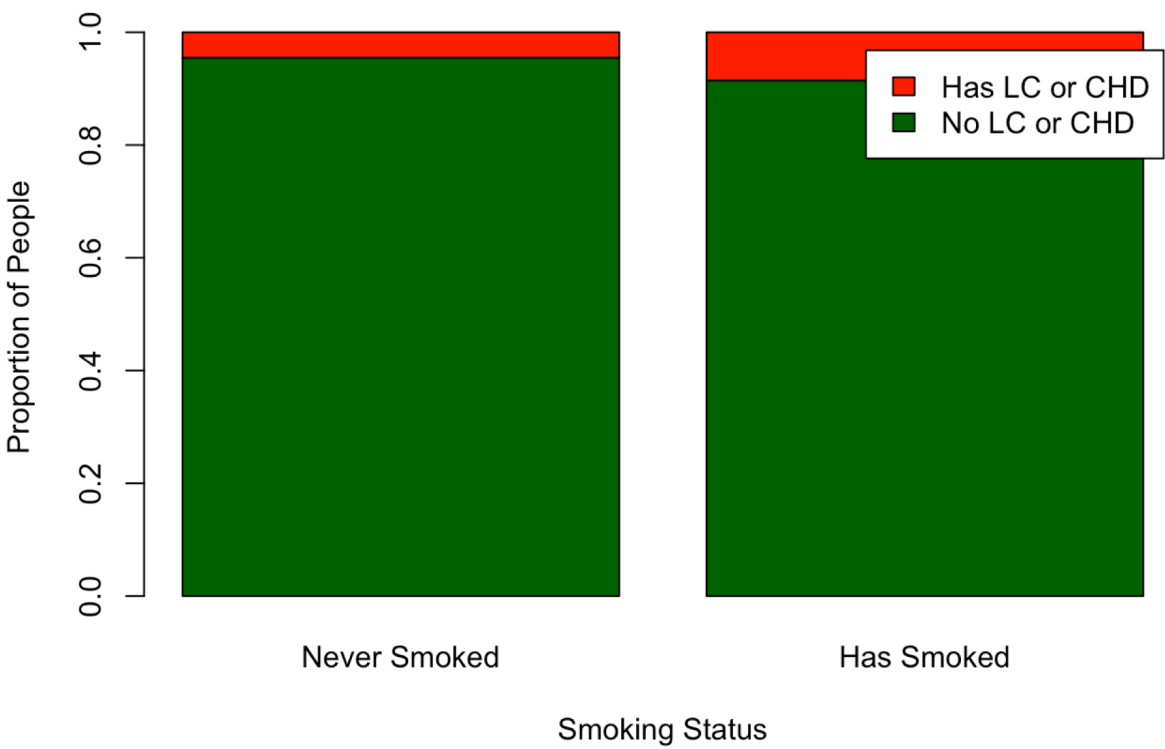
```
##
## Call:
## glm(formula = nmes.data$disease ~ nmes.data$eversmk, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4312  -0.4312  -0.3136  -0.3136   2.4645
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.9877     0.1025  -29.153   < 2e-16 ***
## nmes.data$eversmkSmoked  0.6589     0.1292   5.098 3.43e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
exp(0.6589)
```

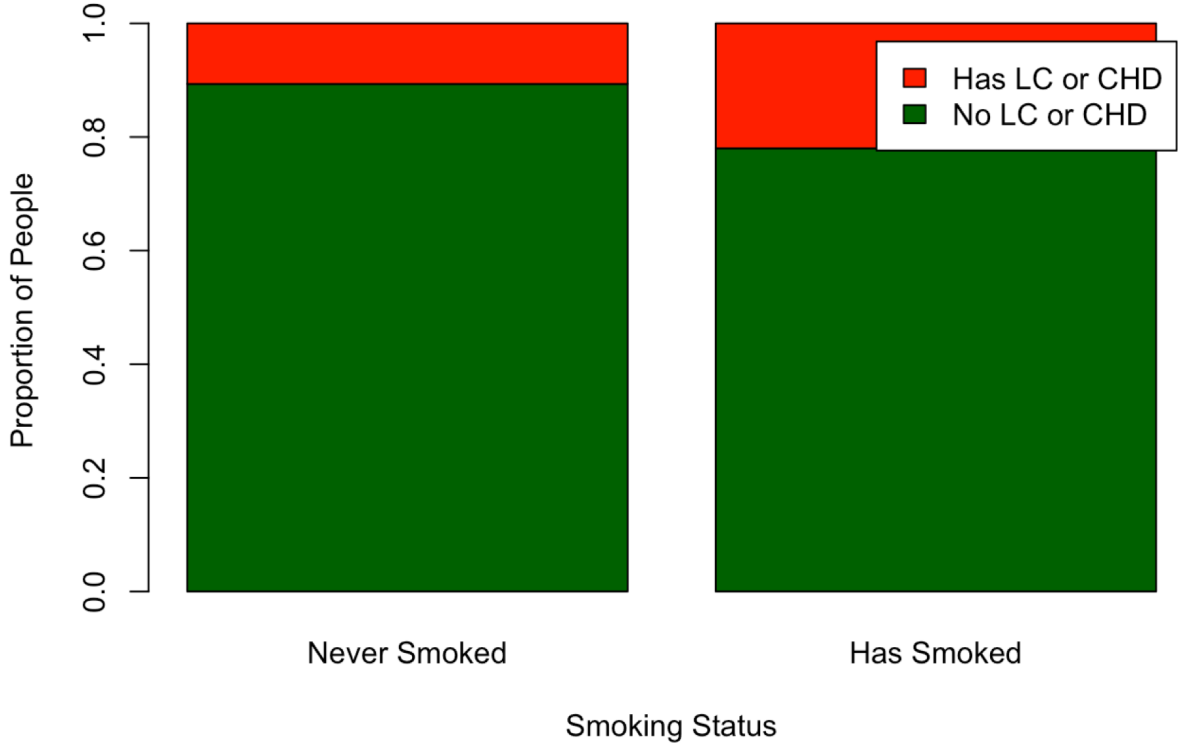
```
## [1] 1.932665
```

The odds of having chd or lung cancer for someone who has ever smoked is 1.932665 times higher than for someone who has never smoked. In other words, smokers have a 93% higher odds of having chd or lung cancer than those who have never smoked.

Disease Outcome by Smoking Status in Obese Individuals



Disease Outcome by Smoking Status in Impoverished Individuals



```
modell = glm(Obese$"Disease Outcome" ~ Obese$eversmk, family=binomial(link="logit"))
summary(modell)
```

```
##
## Call:
## glm(formula = Obese$"Disease Outcome" ~ Obese$eversmk, family = binomial(link = "logit"
##
## Deviance Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.4226	-0.4226	-0.3036	-0.3036	2.4900

```
##
## Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.0540	0.2642	-11.560	<2e-16 ***
Obese\$eversmkHas Smoked	0.6830	0.3396	2.011	0.0443 *

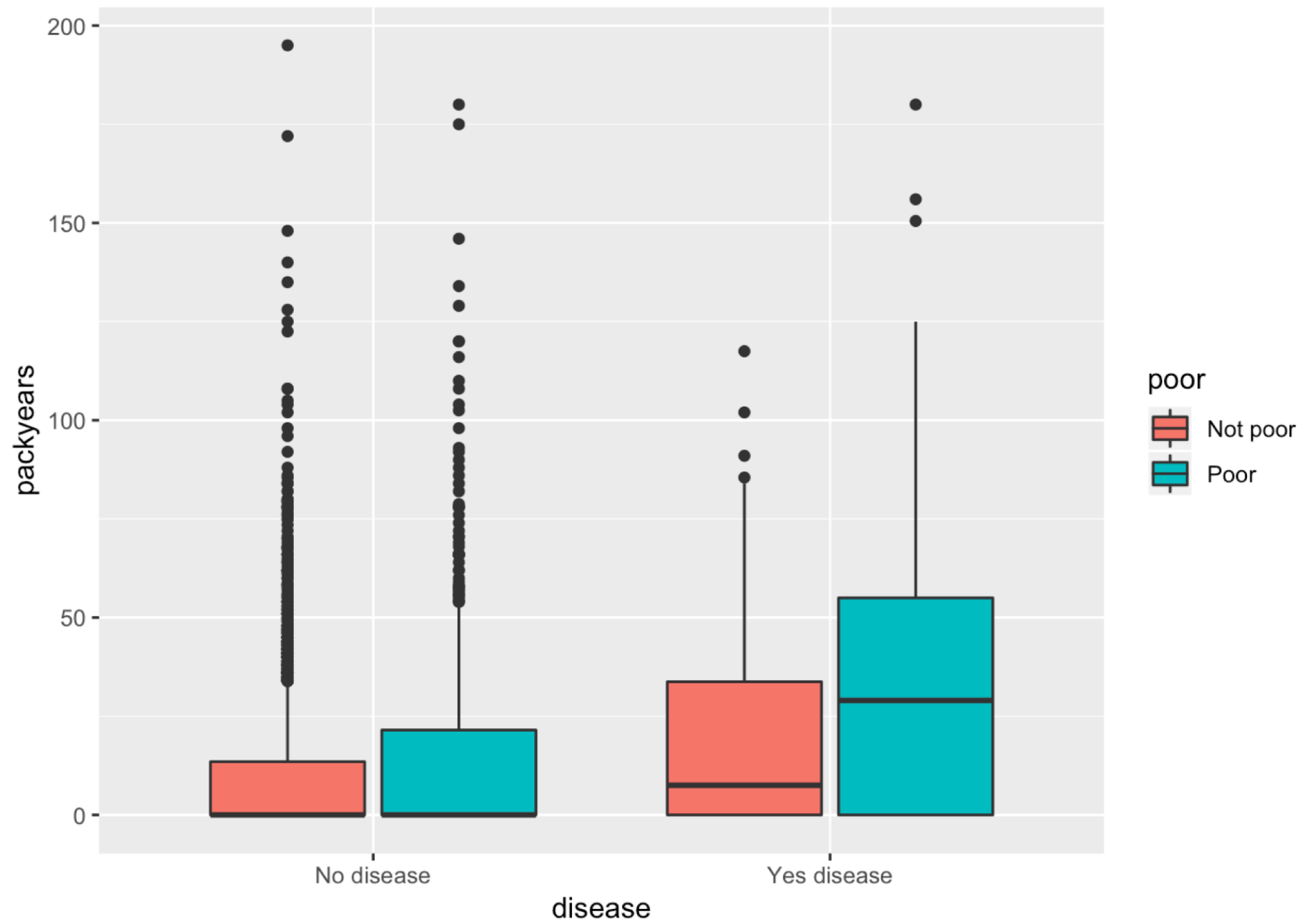
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the outcome of this logistic regression shows, with a β value of 0.6830, the log odds of having a positive disease outcome among smoking obese individuals vs. non-smoking obese individuals is 0.6830. A more practical way of interpreting this statistic would be as e^{β} which yields a value of 1.980 meaning the odds of having a positive disease outcome is 1.980 times greater in smoking obese individuals vs. non-smoking obese individuals. Though this effect is not as large as expected, it is still significant enough to conclude that smoking contributes to the risk of development of LC or CHD independent of obesity.

```
model2 = glm(Poverty$"Disease Outcome" ~ Poverty$eversmk, family=binomial(link="logit"))
summary(model2)
```

```
##
## Call:
## glm(formula = Poverty$"Disease Outcome" ~ Poverty$eversmk, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7046  -0.7046  -0.4748  -0.4748   2.1159
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.1258     0.1496 -14.209  < 2e-16 ***
## Poverty$eversmkHas Smoked   0.8590     0.1842   4.663 3.11e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

As the outcome of this second logistic regression shows, the log odds of having a positive disease outcome among smoking impoverished individuals vs. non-smoking impoverished individuals is 0.859. A more practical way of interpreting this statistic would be as $e^{0.859}$ which yields a value of 2.361 meaning the odds of having a positive disease outcome is 2.361 times greater in smoking impoverished individuals vs. non-smoking impoverished individuals. The baseline (non-smoking) amount of LC and CHD observed was surprisingly high relative to obese



```
modell1<- glm(disease~ eversmk + poor + female, family = binomial(link="logit"), data=nmes.
data)
summary(modell1)
```

```
##
## Call:
## glm(formula = disease ~ eversmk + poor + female, family = binomial(link = "logit"),
##      data = nmes.data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7414  -0.3466  -0.2840  -0.2182   2.7387
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.3182     0.1508  -22.007  < 2e-16 ***
## eversmkSmoked     0.5360     0.1359   3.946 7.96e-05 ***
## poorPoor         1.6311     0.1290  12.648  < 2e-16 ***
## femaleFemale    -0.4081     0.1320  -3.092  0.00199 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

#Below is the selection model attained by backwards elimination of covariates.

```
logmodelsmoke <- glm(nmesUNPROC$lc5~nmesUNPROC$packyears+nmesUNPROC$poor+nmesUNPROC$female
+nmesUNPROC$age+nmesUNPROC$totalexp, family = binomial(link = "logit"))
summary(logmodelsmoke)
```

```
##
## Call:
## glm(formula = nmesUNPROC$lc5 ~ nmesUNPROC$packyears + nmesUNPROC$poor +
##      nmesUNPROC$female + nmesUNPROC$age + nmesUNPROC$totalexp,
##      family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6188  -0.1097  -0.0566  -0.0343   3.4911
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -8.442e+00  8.570e-01  -9.850  < 2e-16 ***
## nmesUNPROC$packyears  2.327e-02  3.613e-03   6.440 1.19e-10 ***
## nmesUNPROC$poor      1.516e+00  3.902e-01   3.885 0.000102 ***
## nmesUNPROC$female    -1.052e+00  3.585e-01  -2.934 0.003342 **
## nmesUNPROC$age       4.866e-02  1.274e-02   3.820 0.000133 ***
## nmesUNPROC$totalexp  2.055e-05  7.469e-06   2.751 0.005948 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```


#At baseline, the odds of lung cancer is nearly 0 (0.000002), meaning that a large proportion of the risk associated with lung cancer was explained by the inclusion of the covariates. For every packyear, the odds of lung cancer increases by about 2.4%. Being poor increases the odds of lung cancer by about 355%. Being female decreases the odds of lung cancer by about 65%. Every additional year of age increases the odds of lung cancer by about 5%. Every dollar spent on health expenditures increases the odds of lung cancer by about 0.01%. The latter of course, is largely an association, albeit a very strong one. For predictive purposes, I would leave it in the model. For inferential purposes, I would exclude it, as health expenditure will almost always succeed lung cancer diagnosis in otherwise similar individuals.