

Final Project

Victor Huang & Xander Roti

12 November 2021

```
### Importing the data set
```

```
heartoriginal <- read.csv("~/Stat230/heart.csv")
```

```
### Uploading the dataset and removing incomplete cases.
```

```
### Dropping N/A values that would interfere with the dataset
```

```
heartoriginal <- heartoriginal %>% drop_na()
```

```
heart <- filter(heartoriginal, Cholesterol > 0)
```

```
### Fitting the full model with all the predictors on the entire Dataset.
```

```
### Creating the full glm model for the dataset with all variables
```

```
heart_glm <- glm(HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol + FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak + ST_Slope, family = binomial, data = heart)
```

Call:

```
glm(formula = HeartDisease ~ Age + Sex + ChestPainType + RestingBP +  
    Cholesterol + FastingBS + RestingECG + MaxHR + ExerciseAngina +  
    Oldpeak + ST_Slope, family = binomial, data = heart)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6697	-0.3841	-0.1085	0.4465	2.7371

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-5.4373046	1.7625169	-3.085	0.002036	**
Age	0.0313784	0.0148105	2.119	0.034119	*
SexM	1.8655490	0.3134065	5.952	2.64e-09	***
ChestPainTypeATA	-1.6731804	0.3544226	-4.721	2.35e-06	***
ChestPainTypeNAP	-1.5730121	0.3029404	-5.192	2.08e-07	***
ChestPainTypeTA	-1.6332529	0.4838117	-3.376	0.000736	***
RestingBP	0.0117792	0.0072988	1.614	0.106557	
Cholesterol	0.0024955	0.0019773	1.262	0.206928	
FastingBS	0.2923999	0.3311265	0.883	0.377212	
RestingECGNormal	-0.2297888	0.2842091	-0.809	0.418791	
RestingECGST	-0.1746017	0.3941671	-0.443	0.657792	
MaxHR	0.0005807	0.0057810	0.100	0.919991	
ExerciseAnginaY	0.9073515	0.2671360	3.397	0.000682	***
Oldpeak	0.4108355	0.1406671	2.921	0.003493	**
ST_SlopeFlat	1.3038217	0.5197574	2.509	0.012124	*

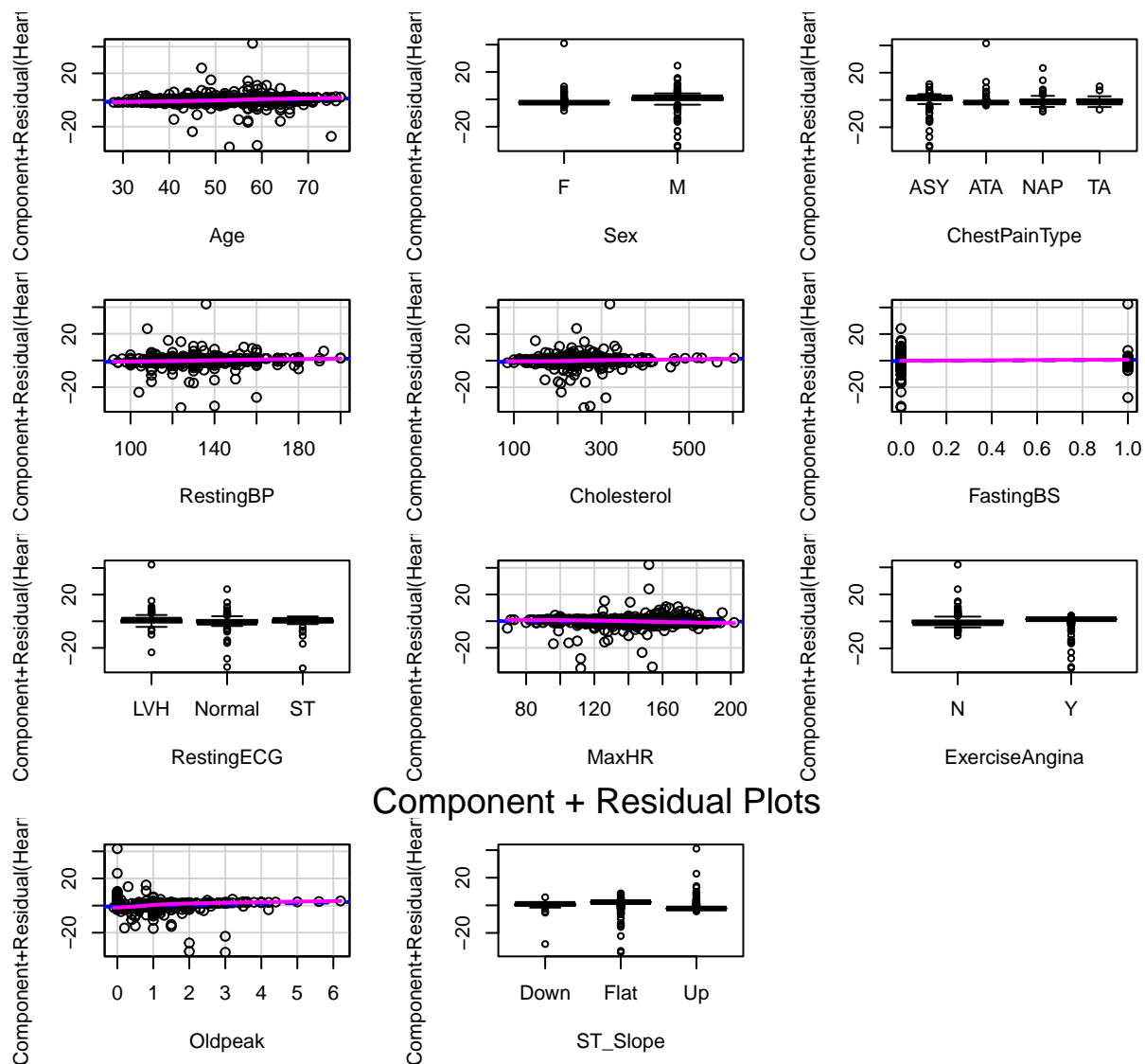
```
ST_SlopeUp      -1.2100372  0.5655279  -2.140 0.032382 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

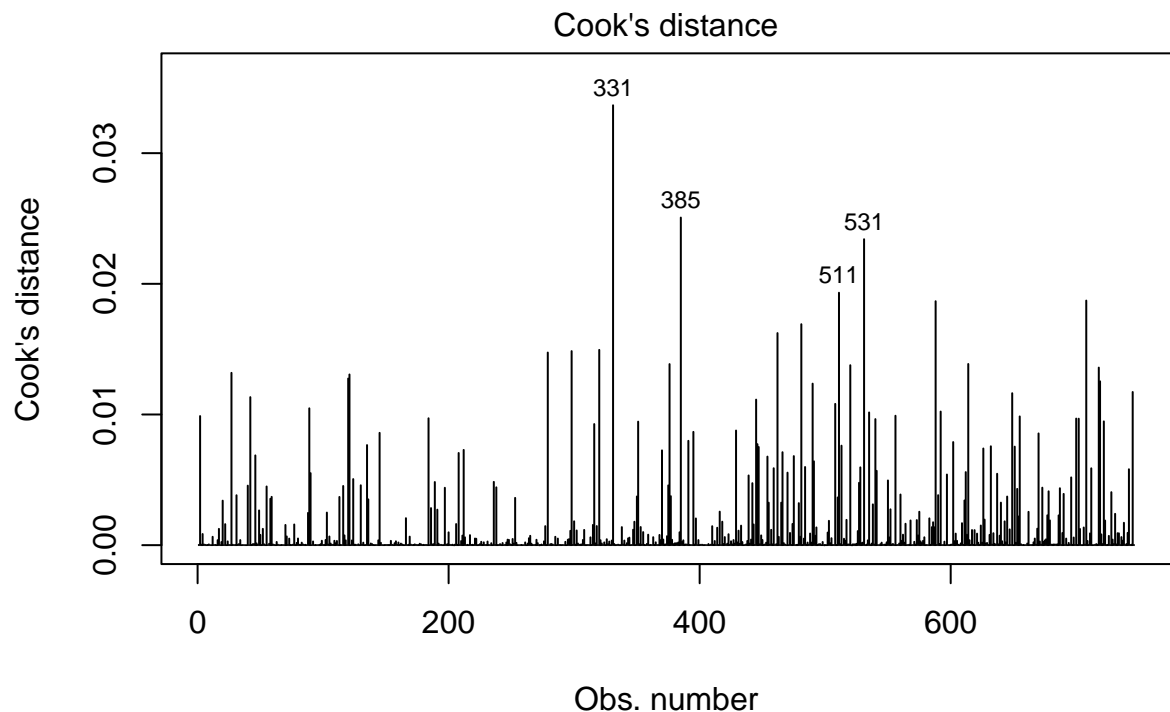
```
Null deviance: 1032.63 on 745 degrees of freedom
Residual deviance: 483.58 on 730 degrees of freedom
AIC: 515.58
```

```
Number of Fisher Scoring iterations: 6
```

```
###Augmenting code for easier viewing experience
crp(heart_glm)
```

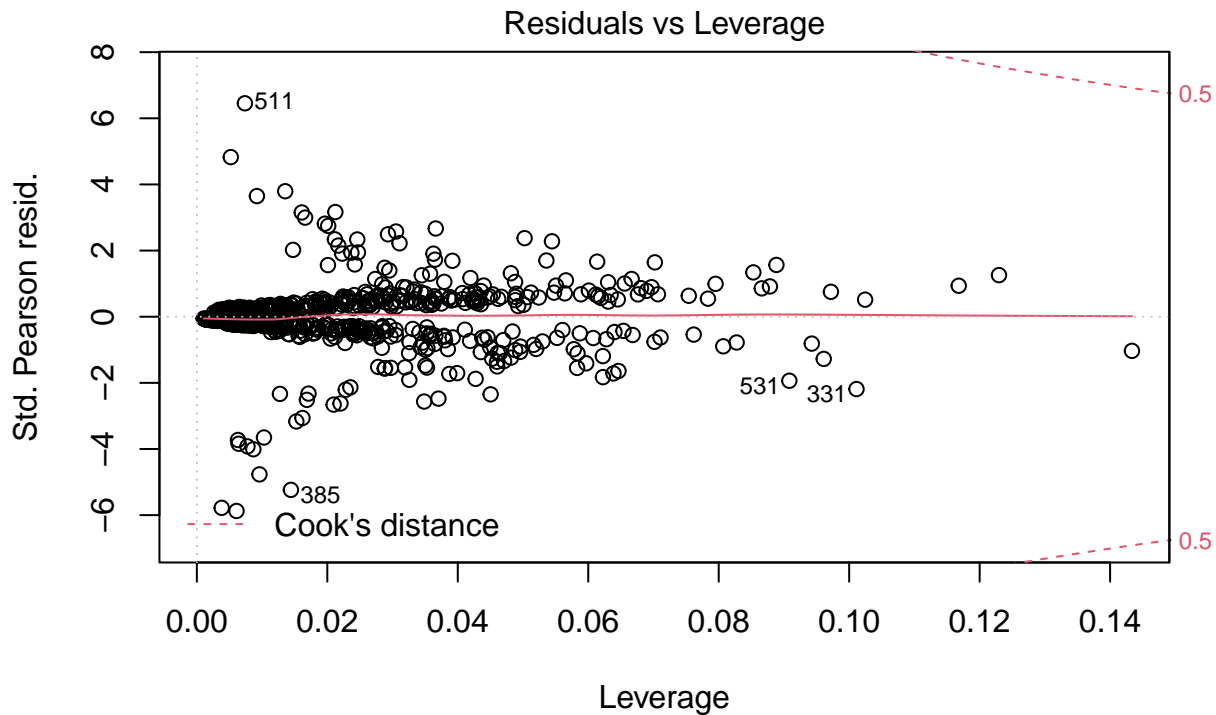


```
heart_aug <- augment(heart_glm)
heart_aug %>%
  slice_max(.hat, n = 1)
# A tibble: 1 x 18
  HeartDisease Age Sex ChestPainType RestingBP Cholesterol FastingBS
    <int> <int> <chr> <chr>          <int>      <int>      <int>
1         0   67 F     NAP             115        564         0
# ... with 11 more variables: RestingECG <chr>, MaxHR <int>,
# ExerciseAngina <chr>, Oldpeak <dbl>, ST_Slope <chr>, .fitted <dbl>,
# .resid <dbl>, .std.resid <dbl>, .hat <dbl>, .sigma <dbl>, .cooks_d <dbl>
plot(heart_glm, which=4, id.n = 4)
```



glm(HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol + Fa ..

```
plot(heart_glm, which=5, id.n = 4)
```



glm(HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol + Fa ..

```
### Finding the dimensions for our dataset
```

```
dim(heart)
```

```
[1] 746 12
```

```
heart <- mutate(heart, Age_grps = ntile(Age, n = 20))
```

```
table(heart$Age_grps)
```

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
38 38 38 38 38 38 37 37 37 37 37 37 37 37 37 37 37 37 37
```

```
### Finding important values for our dataset (median, proportion, and log odds)
```

```
heart_empL0 <- heart %>%
```

```
  group_by(Age_grps) %>%
```

```
    summarize(Age_grps_med = median(Age),      # median Age of groups
```

```
              pi_emp = mean(HeartDisease),    # proportion died
```

```
              log_odds_emp = log(pi_emp/(1-pi_emp))) # log odds
```

```
heart_empL0
```

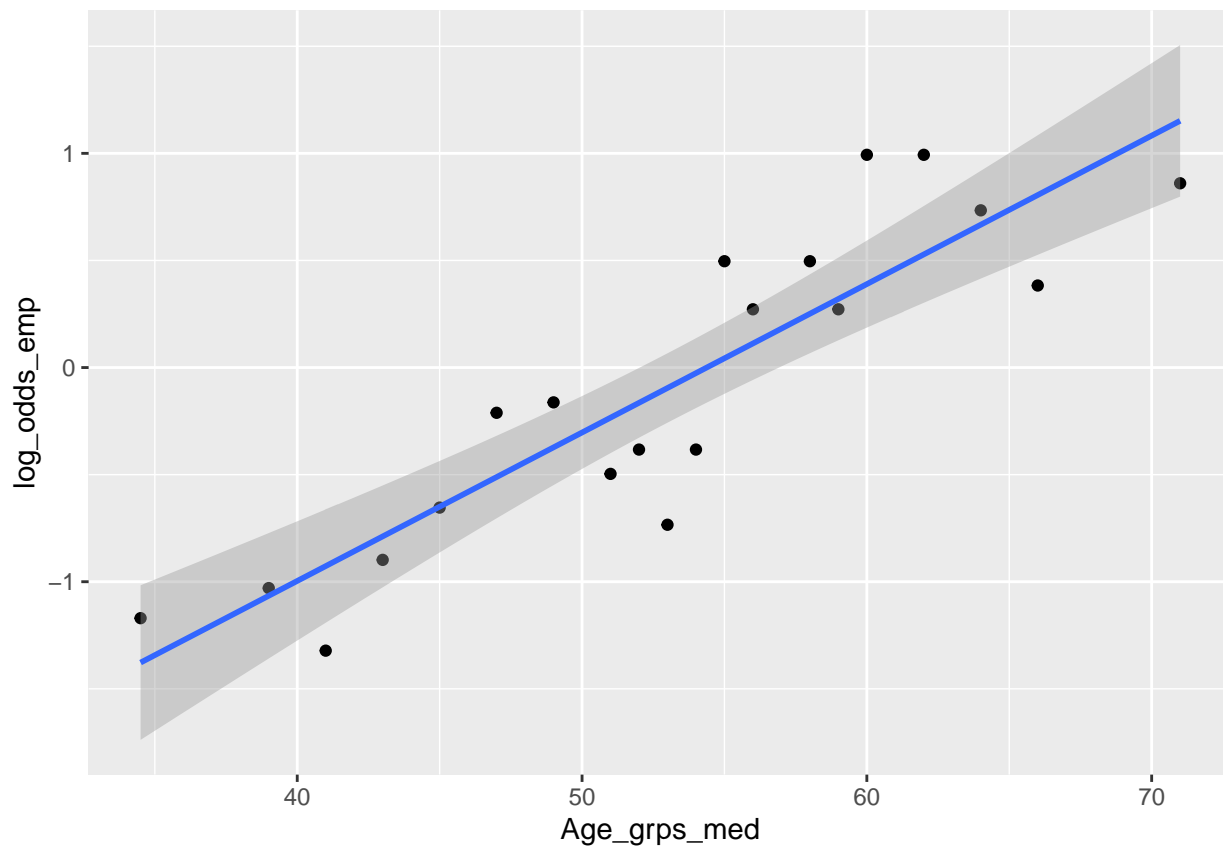
```
# A tibble: 20 x 4
```

	Age_grps	Age_grps_med	pi_emp	log_odds_emp
	<int>	<dbl>	<dbl>	<dbl>
1	1	34.5	0.237	-1.17
2	2	39	0.263	-1.03
3	3	41	0.211	-1.32
4	4	43	0.289	-0.898
5	5	45	0.342	-0.654
6	6	47	0.447	-0.211
7	7	49	0.459	-0.163
8	8	51	0.378	-0.496
9	9	52	0.405	-0.383

10	10	53	0.324	-0.734
11	11	54	0.405	-0.383
12	12	55	0.622	0.496
13	13	56	0.568	0.272
14	14	58	0.622	0.496
15	15	59	0.568	0.272
16	16	60	0.730	0.993
17	17	62	0.730	0.993
18	18	64	0.676	0.734
19	19	66	0.595	0.383
20	20	71	0.703	0.860

Plotting the model

```
ggplot(heart_empL0, aes(x=Age_grps_med, y=log_odds_emp)) +
  geom_point() +
  geom_smooth(method = lm)
`geom_smooth()` using formula 'y ~ x'
```



Mutating the dataset and creating a table

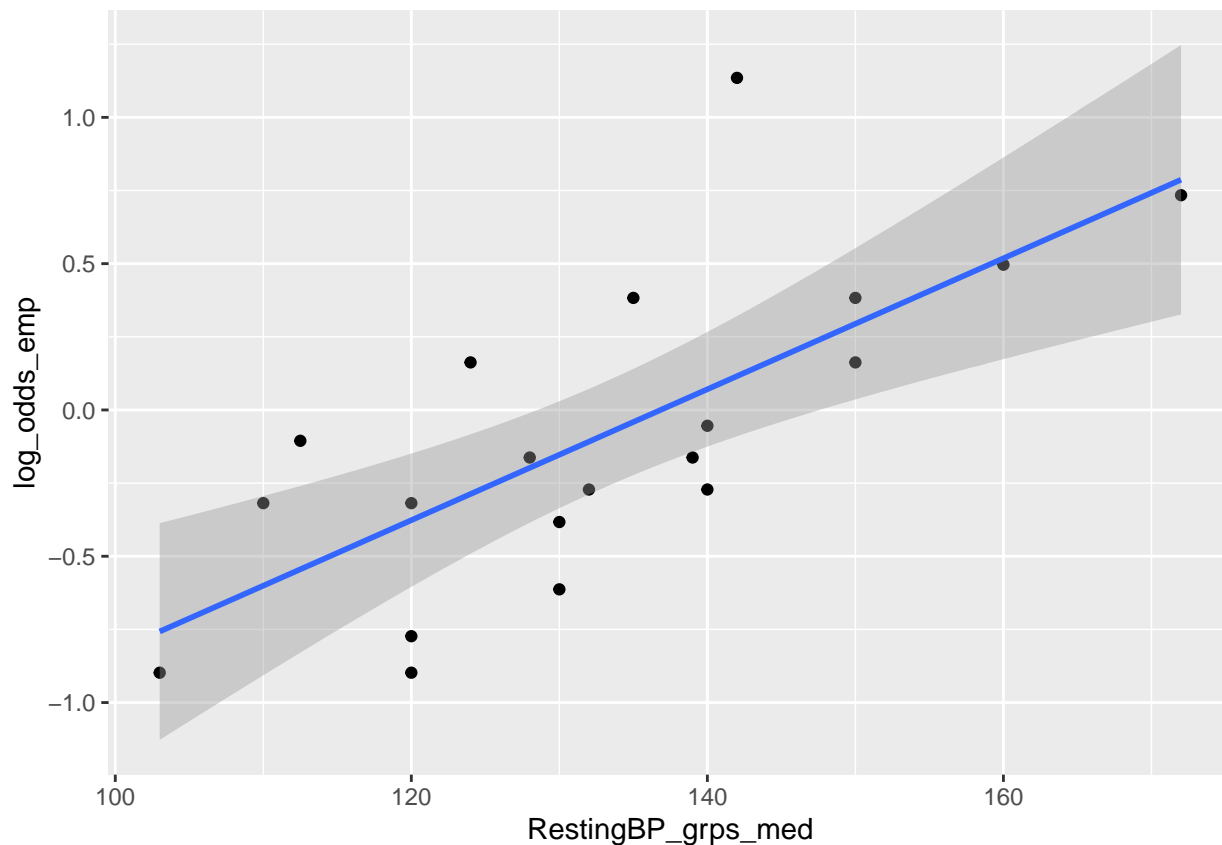
```
dim(heart)
[1] 746 13
heart <- mutate(heart, RestingBP_grps = ntile(RestingBP, n = 20))
table(heart$RestingBP_grps)
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
38	38	38	38	38	38	37	37	37	37	37	37	37	37	37	37	37	37	37	37

```
###Repeating the process above but for a different variable
heart_empL02 <-heart %>%
  group_by(RestingBP_grps) %>%
    summarize(RestingBP_grps_med = median(RestingBP), # median BP of groups
              pi_emp = mean(HeartDisease),           # proportion died
              log_odds_emp = log(pi_emp/(1-pi_emp)))  # log odds
```

```
heart_empL02
# A tibble: 20 x 4
  RestingBP_grps RestingBP_grps_med pi_emp log_odds_emp
    <int>          <dbl>    <dbl>    <dbl>
1         1         103    0.289    -0.898
2         2         110    0.421    -0.318
3         3         112    0.474    -0.105
4         4         120    0.289    -0.898
5         5         120    0.316    -0.773
6         6         120    0.421    -0.318
7         7         124    0.541     0.163
8         8         128    0.459    -0.163
9         9         130    0.351    -0.613
10        10         130    0.405    -0.383
11        11         132    0.432    -0.272
12        12         135    0.595     0.383
13        13         139    0.459    -0.163
14        14         140    0.432    -0.272
15        15         140    0.486    -0.0541
16        16         142    0.757     1.13
17        17         150    0.541     0.163
18        18         150    0.595     0.383
19        19         160    0.622     0.496
20        20         172    0.676     0.734
```

```
ggplot(heart_empL02, aes(x=RestingBP_grps_med, y=log_odds_emp)) +
  geom_point() +
  geom_smooth(method = lm)
`geom_smooth()` using formula 'y ~ x'
```



```
dim(heart)
[1] 746 14
heart <- mutate(heart, Cholesterol_grps = ntile(Cholesterol, n = 20))
table(heart$Cholesterol_grps)
```

```
1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
38 38 38 38 38 38 37 37 37 37 37 37 37 37 37 37 37 37 37
```

```
###Repeating the process above but for a different variable
heart_empL03 <-heart %>%
  group_by(Cholesterol_grps) %>%
  summarize(Cholesterol_grps_med = median(Cholesterol),
            pi_emp = mean(HeartDisease),           # proportion died
            log_odds_emp = log(pi_emp/(1-pi_emp))) # log odds
```

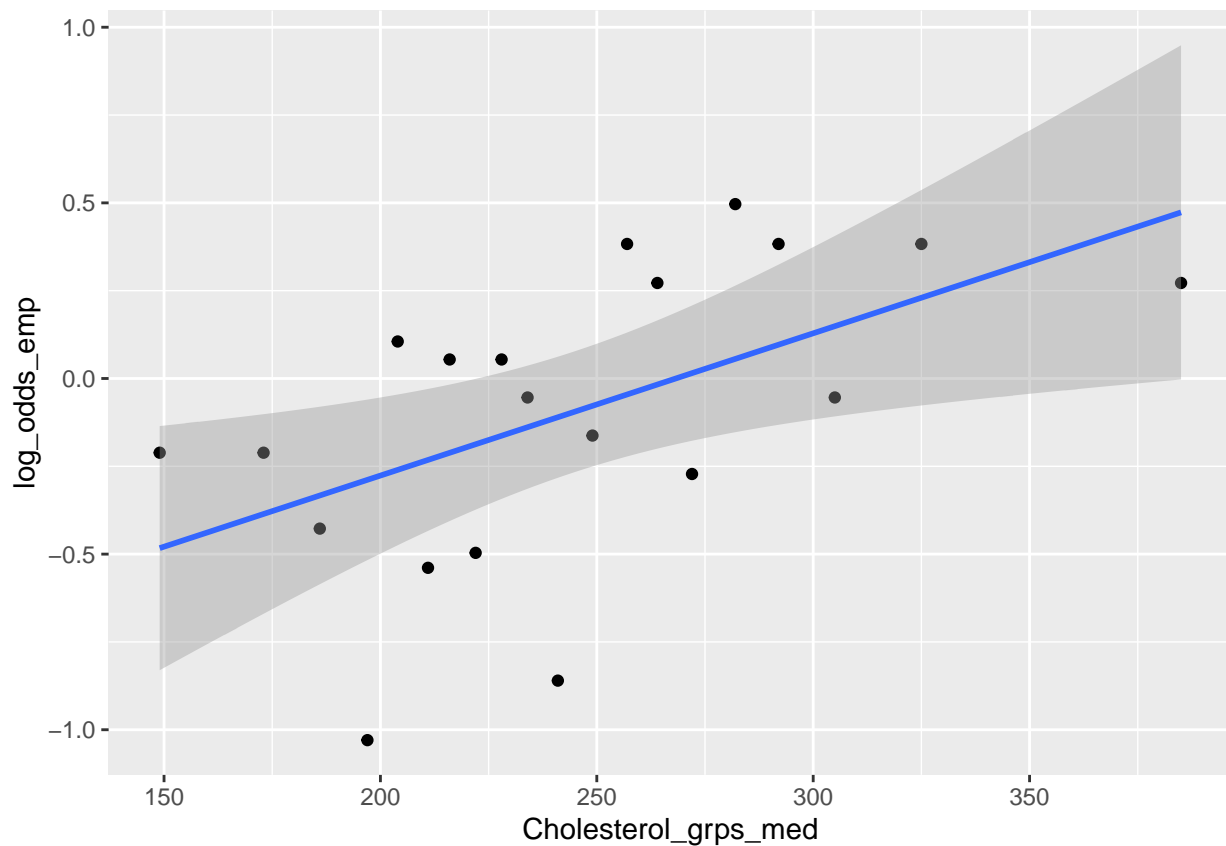
```
heart_empL03
```

```
# A tibble: 20 x 4
```

	Cholesterol_grps	Cholesterol_grps_med	pi_emp	log_odds_emp
	<int>	<dbl>	<dbl>	<dbl>
1	1	149	0.447	-0.211
2	2	173	0.447	-0.211
3	3	186	0.395	-0.427
4	4	197	0.263	-1.03
5	5	204	0.526	0.105
6	6	211	0.368	-0.539
7	7	216	0.514	0.0541
8	8	222	0.378	-0.496
9	9	228	0.514	0.0541

10	10	234	0.486	-0.0541
11	11	241	0.297	-0.860
12	12	249	0.459	-0.163
13	13	257	0.595	0.383
14	14	264	0.568	0.272
15	15	272	0.432	-0.272
16	16	282	0.622	0.496
17	17	292	0.595	0.383
18	18	305	0.486	-0.0541
19	19	325	0.595	0.383
20	20	385	0.568	0.272

```
ggplot(heart_empL03, aes(x=Cholesterol_grps_med, y=log_odds_emp)) +
  geom_point() +
  geom_smooth(method = lm)
`geom_smooth()` using formula 'y ~ x'
```



```
dim(heart)
[1] 746 15
heart <- mutate(heart, MaxHR_grps = ntile(MaxHR, n = 20))
table(heart$MaxHR_grps)
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
38	38	38	38	38	38	37	37	37	37	37	37	37	37	37	37	37	37	37	37

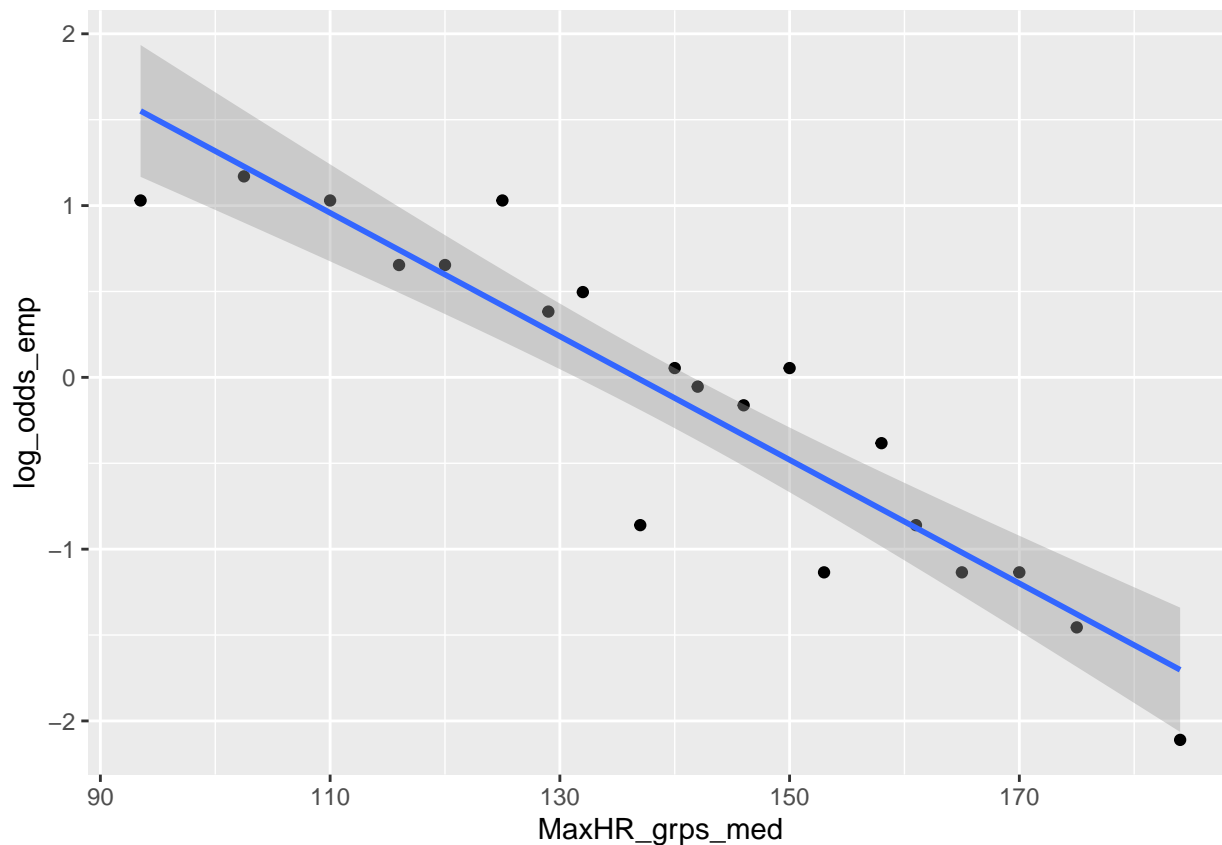

```
heart_empL04 <-heart %>%
  group_by(MaxHR_grps) %>%
  summarize(MaxHR_grps_med = median(MaxHR),      # median MaxHR of groups
            pi_emp = mean(HeartDisease),        # proportion died
            log_odds_emp = log(pi_emp/(1-pi_emp))) # log odds
```

```
heart_empL04
```

```
# A tibble: 20 x 4
```

	MaxHR_grps	MaxHR_grps_med	pi_emp	log_odds_emp
	<int>	<dbl>	<dbl>	<dbl>
1	1	93.5	0.737	1.03
2	2	102.	0.763	1.17
3	3	110	0.737	1.03
4	4	116	0.658	0.654
5	5	120	0.658	0.654
6	6	125	0.737	1.03
7	7	129	0.595	0.383
8	8	132	0.622	0.496
9	9	137	0.297	-0.860
10	10	140	0.514	0.0541
11	11	142	0.486	-0.0541
12	12	146	0.459	-0.163
13	13	150	0.514	0.0541
14	14	153	0.243	-1.13
15	15	158	0.405	-0.383
16	16	161	0.297	-0.860
17	17	165	0.243	-1.13
18	18	170	0.243	-1.13
19	19	175	0.189	-1.46
20	20	184	0.108	-2.11

```
ggplot(heart_empL04, aes(x=MaxHR_grps_med, y=log_odds_emp)) +
  geom_point() +
  geom_smooth(method = lm)
`geom_smooth()` using formula 'y ~ x'
```



```
heart_glm <- glm(HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol + FastingBS + RestingECGNormal + RestingECGST + MaxHR + ExerciseAngina + Oldpeak + ST_Slope, family = binomial, data = heart)
summary(heart_glm)
```

Call:

```
glm(formula = HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol + FastingBS + RestingECGNormal + RestingECGST + MaxHR + ExerciseAngina + Oldpeak + ST_Slope, family = binomial, data = heart)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6697	-0.3841	-0.1085	0.4465	2.7371

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-5.4373046	1.7625169	-3.085	0.002036	**
Age	0.0313784	0.0148105	2.119	0.034119	*
SexM	1.8655490	0.3134065	5.952	2.64e-09	***
ChestPainTypeATA	-1.6731804	0.3544226	-4.721	2.35e-06	***
ChestPainTypeNAP	-1.5730121	0.3029404	-5.192	2.08e-07	***
ChestPainTypeTA	-1.6332529	0.4838117	-3.376	0.000736	***
RestingBP	0.0117792	0.0072988	1.614	0.106557	
Cholesterol	0.0024955	0.0019773	1.262	0.206928	
FastingBS	0.2923999	0.3311265	0.883	0.377212	
RestingECGNormal	-0.2297888	0.2842091	-0.809	0.418791	
RestingECGST	-0.1746017	0.3941671	-0.443	0.657792	
MaxHR	0.0005807	0.0057810	0.100	0.919991	

```
ExerciseAnginaY    0.9073515  0.2671360   3.397 0.000682 ***
Oldpeak           0.4108355  0.1406671   2.921 0.003493 **
ST_SlopeFlat      1.3038217  0.5197574   2.509 0.012124 *
ST_SlopeUp        -1.2100372  0.5655279  -2.140 0.032382 *
```

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

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 1032.63 on 745 degrees of freedom
Residual deviance: 483.58 on 730 degrees of freedom
AIC: 515.58
```

Number of Fisher Scoring iterations: 6

Doing drop-in-deviance tests to test the significance of the variable

```
heart_glm2 <- glm(HeartDisease ~ Sex + ChestPainType + RestingBP + Cholesterol + FastingBS + RestingECG
```

```
anova(heart_glm2, heart_glm, test = "Chisq")
```

Analysis of Deviance Table

```
Model 1: HeartDisease ~ Sex + ChestPainType + RestingBP + Cholesterol +
  FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak +
  ST_Slope
```

```
Model 2: HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol +
  FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak +
  ST_Slope
```

```
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      731      488.14
2      730      483.58 1      4.559  0.03275 *
```

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

```
1-pchisq(483.58, 730)
```

```
[1] 1
```

```
488.14 - 483.58 ##LRT
```

```
[1] 4.56
```

```
0.016550/0.2087463
```

```
[1] 0.07928284
```

```
2*pnorm(0.07928284)
```

```
[1] 1.063192
```

The estimated coefficient for Age is 0.0313784, and the standard error is 0.0148105. The p-value that we get for the goodness of fit test is 1. The p-value that we get for the Wald's test is 0.034119. The p-value that we get for the drop-in-deviance test is 0.03275. This means that Age is significant and that we should keep this in our model.

Doing drop-in-deviance tests to test the significance of the variable

```
heart3_glm <- glm(HeartDisease ~ Age + ChestPainType + RestingBP + Cholesterol + FastingBS + RestingECG
```

```
anova(heart3_glm, heart_glm, test = "Chisq")
```

Analysis of Deviance Table

```
Model 1: HeartDisease ~ Age + ChestPainType + RestingBP + Cholesterol +
  FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak +
  ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol +
  FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak +
  ST_Slope
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      731      523.59
2      730      483.58  1    40.005 2.533e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this model, the estimated coefficient for Sex is 1.8655490 with a standard error of 0.3134065. The p-value that we get for the Wald's test is 2.64e-09. In this drop-in-deviance test, the p-value is significant here, showing that we do not take Sex out of the model.

Doing drop-in-deviance tests to test the significance of the variable

```
heart4_glm <- glm(HeartDisease ~ Age + Sex + RestingBP + Cholesterol + FastingBS + RestingECG + MaxHR +
  ST_Slope, data = heart)
anova(heart4_glm, heart_glm, test = "Chisq")
Analysis of Deviance Table

Model 1: HeartDisease ~ Age + Sex + RestingBP + Cholesterol + FastingBS +
  RestingECG + MaxHR + ExerciseAngina + Oldpeak + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol +
  FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak +
  ST_Slope
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      733      525.90
2      730      483.58  3    42.319 3.433e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value that we get for the Wald's test for each ChestPainType is 2.64e-09. In this drop-in-deviance test, the p-value is significant here, showing that we do not take ChestPainType out of the model.

Doing drop-in-deviance tests to test the significance of the variable

```
heart_glm5 <- glm(HeartDisease ~ Age + Sex + ChestPainType + Cholesterol + FastingBS + RestingECG + MaxHR +
  ST_Slope, data = heart)
summary(heart_glm5)

Call:
glm(formula = HeartDisease ~ Age + Sex + ChestPainType + Cholesterol +
  FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak +
  ST_Slope, family = binomial, data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7222  -0.3890  -0.1173   0.4445   2.7024

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.8655490    0.3134065  5.95177 2.64e-09 ***
Age          0.0000000    0.0000000  0.00000 1.0000000
Sex          1.8655490    0.3134065  5.95177 2.64e-09 ***
ChestPainType
  atypical    2.64e-09 ***
  major       2.64e-09 ***
  non-anginal 2.64e-09 ***
Cholesterol  0.0000000    0.0000000  0.00000 1.0000000
FastingBS    0.0000000    0.0000000  0.00000 1.0000000
RestingECG   0.0000000    0.0000000  0.00000 1.0000000
MaxHR        0.0000000    0.0000000  0.00000 1.0000000
ExerciseAngina
  no          2.64e-09 ***
  yes         2.64e-09 ***
Oldpeak      0.0000000    0.0000000  0.00000 1.0000000
ST_Slope     0.0000000    0.0000000  0.00000 1.0000000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

(Intercept)      -4.2217418   1.5882242  -2.658 0.007857 **
Age               0.0352513   0.0146593   2.405 0.016186 *
SexM             1.8280371   0.3086990   5.922 3.19e-09 ***
ChestPainTypeATA -1.6288593   0.3514200  -4.635 3.57e-06 ***
ChestPainTypeNAP -1.5713404   0.3031520  -5.183 2.18e-07 ***
ChestPainTypeTA  -1.5549800   0.4795066  -3.243 0.001183 **
Cholesterol       0.0028097   0.0019746   1.423 0.154747
FastingBS        0.3602864   0.3265471   1.103 0.269888
RestingECGNormal -0.2195001   0.2831660  -0.775 0.438243
RestingECGST      -0.1345451   0.3926224  -0.343 0.731837
MaxHR            0.0006743   0.0057665   0.117 0.906910
ExerciseAnginaY   0.9451880   0.2657907   3.556 0.000376 ***
Oldpeak          0.4295708   0.1388026   3.095 0.001969 **
ST_SlopeFlat     1.3003924   0.5166849   2.517 0.011843 *
ST_SlopeUp       -1.1782705   0.5612939  -2.099 0.035799 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 1032.6 on 745 degrees of freedom
Residual deviance: 486.2 on 731 degrees of freedom
AIC: 516.2

```

Number of Fisher Scoring iterations: 6

```

anova(heart_glm5, heart_glm, test = "Chisq")
Analysis of Deviance Table

```

```

Model 1: HeartDisease ~ Age + Sex + ChestPainType + Cholesterol + FastingBS +
  RestingECG + MaxHR + ExerciseAngina + Oldpeak + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol +
  FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak +
  ST_Slope
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      731      486.20
2      730      483.58 1    2.6213   0.1054

```

In this ANOVA test, when analyzing the term for RestingBP, the drop-in-deviance p-value is 0.1054, which shows that RestingBP is not significant to the model.

Doing drop-in-deviance tests to test the significance of the variable

```

heart_glm6 <- glm(HeartDisease ~ Age + Sex + ChestPainType + FastingBS + RestingECG + MaxHR + ExerciseAngina,
  data = heart, family = binomial)
summary(heart_glm6)

```

Call:

```

glm(formula = HeartDisease ~ Age + Sex + ChestPainType + FastingBS +
  RestingECG + MaxHR + ExerciseAngina + Oldpeak + ST_Slope,
  family = binomial, data = heart)

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-2.7022  -0.3823  -0.1218   0.4420   2.7468

```

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.413099   1.4742536  -2.316 0.020575 *
Age           0.0340911   0.0146149   2.333 0.019668 *
SexM          1.7649533   0.3038883   5.808 6.33e-09 ***
ChestPainTypeATA -1.6422280   0.3505312  -4.685 2.80e-06 ***
ChestPainTypeNAP -1.5993727   0.3016820  -5.302 1.15e-07 ***
ChestPainTypeTA -1.5801583   0.4771498  -3.312 0.000927 ***
FastingBS     0.3929287   0.3243364   1.211 0.225710
RestingECGNormal -0.2503346   0.2821015  -0.887 0.374868
RestingECGST   -0.1721090   0.3900980  -0.441 0.659072
MaxHR         0.0007638   0.0057227   0.133 0.893820
ExerciseAnginaY 0.9538006   0.2654084   3.594 0.000326 ***
Oldpeak       0.4243075   0.1388137   3.057 0.002238 **
ST_SlopeFlat   1.3266653   0.5125717   2.588 0.009646 **
ST_SlopeUp     -1.1793038   0.5574743  -2.115 0.034392 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 1032.63 on 745 degrees of freedom
Residual deviance: 488.28 on 732 degrees of freedom
AIC: 516.28

```

Number of Fisher Scoring iterations: 5

```

anova(heart_glm6, heart_glm5, test = "Chisq")
Analysis of Deviance Table

```

```

Model 1: HeartDisease ~ Age + Sex + ChestPainType + FastingBS + RestingECG +
  MaxHR + ExerciseAngina + Oldpeak + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + Cholesterol + FastingBS +
  RestingECG + MaxHR + ExerciseAngina + Oldpeak + ST_Slope
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      732    488.28
2      731    486.20 1    2.0758  0.1497

```

The p-value for this drop-in-deviance test for the term Cholesterol is 0.1497, which shows that the smaller model is better, as this term is not significant.

Doing drop-in-deviance tests to test the significance of the variable

```

heart_glm7 <- glm(HeartDisease ~ Age + Sex + ChestPainType + RestingECG + MaxHR + ExerciseAngina + Oldpeak,
  data = heart)
summary(heart_glm7)

```

Call:

```

glm(formula = HeartDisease ~ Age + Sex + ChestPainType + RestingECG +
  MaxHR + ExerciseAngina + Oldpeak + ST_Slope, family = binomial,
  data = heart)

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-2.7372  -0.3905  -0.1207   0.4492   2.8723

```

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.4534928  1.4666683  -2.355 0.018540 *
Age             0.0367148  0.0144116   2.548 0.010847 *
SexM            1.8054616  0.3020732   5.977 2.27e-09 ***
ChestPainTypeATA -1.6400974  0.3496293  -4.691 2.72e-06 ***
ChestPainTypeNAP -1.5838085  0.3008390  -5.265 1.40e-07 ***
ChestPainTypeTA  -1.5368769  0.4732512  -3.247 0.001164 **
RestingECGNormal -0.2648430  0.2808551  -0.943 0.345687
RestingECGST     -0.1301001  0.3874668  -0.336 0.737044
MaxHR            0.0007577  0.0056988   0.133 0.894224
ExerciseAnginaY   0.9543314  0.2648203   3.604 0.000314 ***
Oldpeak         0.4147557  0.1388159   2.988 0.002810 **
ST_SlopeFlat     1.2711619  0.5106020   2.490 0.012791 *
ST_SlopeUp       -1.2599518  0.5547673  -2.271 0.023139 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 1032.63 on 745 degrees of freedom
Residual deviance: 489.76 on 733 degrees of freedom
AIC: 515.76

```

Number of Fisher Scoring iterations: 5

```

anova(heart_glm7, heart_glm6, test = "Chisq")
Analysis of Deviance Table

```

```

Model 1: HeartDisease ~ Age + Sex + ChestPainType + RestingECG + MaxHR +
  ExerciseAngina + Oldpeak + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + FastingBS + RestingECG +
  MaxHR + ExerciseAngina + Oldpeak + ST_Slope
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      733      489.76
2      732      488.28 1    1.4843   0.2231

```

For testing the term FastingBS, we get a drop-in-deviance p-value of 0.2231. This shows that this term is not significant to the model.

```

heart_glm8 <- glm(HeartDisease ~ Age + Sex + ChestPainType + MaxHR + ExerciseAngina + Oldpeak + ST_Slope,
summary(heart_glm8)

```

Call:

```

glm(formula = HeartDisease ~ Age + Sex + ChestPainType + MaxHR +
  ExerciseAngina + Oldpeak + ST_Slope, family = binomial, data = heart)

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-2.7170  -0.3892  -0.1217   0.4592   2.9284

```

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.867397  1.390712  -2.781 0.005421 **

```

```

Age            0.040159  0.013997  2.869 0.004116 **
SexM           1.798305  0.300876  5.977 2.27e-09 ***
ChestPainTypeATA -1.666637  0.347196 -4.800 1.58e-06 ***
ChestPainTypeNAP -1.592914  0.300528 -5.300 1.16e-07 ***
ChestPainTypeTA -1.530705  0.471104 -3.249 0.001157 **
MaxHR          0.001452  0.005544  0.262 0.793353
ExerciseAnginaY 0.946540  0.263994  3.585 0.000336 ***
Oldpeak        0.419059  0.138048  3.036 0.002401 **
ST_SlopeFlat   1.240071  0.509260  2.435 0.014890 *
ST_SlopeUp     -1.289846  0.553381 -2.331 0.019762 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 1032.63 on 745 degrees of freedom
Residual deviance: 490.68 on 735 degrees of freedom
AIC: 512.68

```

Number of Fisher Scoring iterations: 5

```

anova(heart_glm8, heart_glm7, test = "Chisq")
Analysis of Deviance Table

```

```

Model 1: HeartDisease ~ Age + Sex + ChestPainType + MaxHR + ExerciseAngina +
  Oldpeak + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + RestingECG + MaxHR +
  ExerciseAngina + Oldpeak + ST_Slope
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      735      490.68
2      733      489.76 2  0.91379  0.6332

```

For testing the term RestingECG, we get a drop-in-deviance p-value of 0.6332. This shows that this term is not significant to the model.

Doing drop-in-deviance tests to test the significance of the variable

```

heart_glm9 <- glm(HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak + ST_Slope, family = binomial)
summary(heart_glm9)

```

Call:

```

glm(formula = HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina +
  Oldpeak + ST_Slope, family = binomial, data = heart)

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-2.7265  -0.3928  -0.1227   0.4556   2.9284

```

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.60085    0.94673  -3.803 0.000143 ***
Age           0.03891    0.01316   2.958 0.003098 **
SexM          1.79156    0.29957   5.980 2.22e-09 ***
ChestPainTypeATA -1.65550    0.34450  -4.806 1.54e-06 ***
ChestPainTypeNAP -1.57932    0.29584  -5.338 9.38e-08 ***

```



```

ChestPainTypeTA -1.51087    0.46339   -3.260 0.001112 **
ExerciseAnginaY  0.93541    0.26059    3.590 0.000331 ***
Oldpeak         0.42148    0.13782    3.058 0.002227 **
ST_SlopeFlat    1.23923    0.51047    2.428 0.015198 *
ST_SlopeUp      -1.27465    0.55154   -2.311 0.020828 *
---

```

```

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

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 1032.63 on 745 degrees of freedom
Residual deviance: 490.74 on 736 degrees of freedom
AIC: 510.74

```

Number of Fisher Scoring iterations: 5

```

anova(heart_glm9, heart_glm8, test = "Chisq")
Analysis of Deviance Table

```

```

Model 1: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak +
  ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + MaxHR + ExerciseAngina +
  Oldpeak + ST_Slope
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      736      490.74
2      735      490.68 1  0.068734  0.7932

```

For testing the term MaxHR, we get a drop-in-deviance p-value of 0.7932. This shows that this term is not significant to the model.

Doing drop-in-deviance tests to test the significance of the variable

```
heart_glm10 <- glm(HeartDisease ~ Age + Sex + ChestPainType + Oldpeak + ST_Slope, family = binomial, data = heart)
```

```

anova(heart_glm10, heart_glm9, test = "Chisq")
Analysis of Deviance Table

```

```

Model 1: HeartDisease ~ Age + Sex + ChestPainType + Oldpeak + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak +
  ST_Slope
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      737      503.45
2      736      490.74 1  12.709 0.000364 ***
---

```

```

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

```

For testing the term ExerciseAngina, we get a drop-in-deviance p-value of 0.000364. This shows that this term is significant to the model.

Doing drop-in-deviance tests to test the significance of the variable

```
heart_glm11 <- glm(HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + ST_Slope, family = binomial, data = heart)
```

```

anova(heart_glm11, heart_glm9, test = "Chisq")
Analysis of Deviance Table

```

```

Model 1: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak +
  ST_Slope
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      737      500.64
2      736      490.74  1    9.8999 0.001653 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

For testing the term OldPeak, we get a drop-in-deviance p-value of 0.001653. This shows that this term is significant to the model.

```

### Doing drop-in-deviance tests to test the significance of the variable
heart_glm12 <- glm(HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak, family = binomial)

anova(heart_glm12, heart_glm9, test = "Chisq")
Analysis of Deviance Table

Model 1: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak
Model 2: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak +
  ST_Slope
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      738      592.42
2      736      490.74  2   101.68 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

For testing the term ST_Slope, we get a drop-in-deviance p-value of 2.2e-163. This shows that this term is significant to the model.

```

### Doing drop-in-deviance tests to test the significance of the variable
heart_glm <- glm(HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol + FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak + ST_Slope, family = binomial, data = heart)
summary(heart_glm)

Call:
glm(formula = HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol + FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak + ST_Slope, family = binomial, data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6697  -0.3841  -0.1085   0.4465   2.7371

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.4373046  1.7625169  -3.085 0.002036 **
Age           0.0313784  0.0148105   2.119 0.034119 *
SexM         1.8655490  0.3134065   5.952 2.64e-09 ***
ChestPainTypeATA -1.6731804  0.3544226  -4.721 2.35e-06 ***
ChestPainTypeNAP -1.5730121  0.3029404  -5.192 2.08e-07 ***
ChestPainTypeTA -1.6332529  0.4838117  -3.376 0.000736 ***
RestingBP     0.0117792  0.0072988   1.614 0.106557


```

```

Cholesterol      0.0024955  0.0019773   1.262 0.206928
FastingBS        0.2923999  0.3311265   0.883 0.377212
RestingECGNormal -0.2297888  0.2842091  -0.809 0.418791
RestingECGST     -0.1746017  0.3941671  -0.443 0.657792
MaxHR            0.0005807  0.0057810   0.100 0.919991
ExerciseAnginaY   0.9073515  0.2671360   3.397 0.000682 ***
Oldpeak          0.4108355  0.1406671   2.921 0.003493 **
ST_SlopeFlat     1.3038217  0.5197574   2.509 0.012124 *
ST_SlopeUp       -1.2100372  0.5655279  -2.140 0.032382 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 1032.63 on 745 degrees of freedom
Residual deviance: 483.58 on 730 degrees of freedom
AIC: 515.58

```

Number of Fisher Scoring iterations: 6

```
heart_glm13 <- glm(HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak + ST_Slope, family = binomial)
```

```
anova(heart_glm13, heart_glm, test = "Chisq")
Analysis of Deviance Table
```

```

Model 1: HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak + ST_Slope
Model 2: HeartDisease ~ Age + Sex + ChestPainType + RestingBP + Cholesterol + FastingBS + RestingECG + MaxHR + ExerciseAngina + Oldpeak + ST_Slope
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      736      490.74
2      730      483.58 6    7.1638    0.306

```

Taking out all of the insignificant terms gives us a drop-in-deviance p-value of 0.306.

```
#Summarizing the final model for the dataset
summary(heart_glm13)
```

Call:

```
glm(formula = HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina + Oldpeak + ST_Slope, family = binomial, data = heart)
```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-2.7265  -0.3928  -0.1227   0.4556   2.9284

```

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.60085    0.94673  -3.803 0.000143 ***
Age           0.03891    0.01316   2.958 0.003098 **
SexM          1.79156    0.29957   5.980 2.22e-09 ***
ChestPainTypeATA -1.65550    0.34450  -4.806 1.54e-06 ***
ChestPainTypeNAP -1.57932    0.29584  -5.338 9.38e-08 ***

```

```

ChestPainTypeTA -1.51087 0.46339 -3.260 0.001112 **
ExerciseAnginaY 0.93541 0.26059 3.590 0.000331 ***
Oldpeak 0.42148 0.13782 3.058 0.002227 **
ST_SlopeFlat 1.23923 0.51047 2.428 0.015198 *
ST_SlopeUp -1.27465 0.55154 -2.311 0.020828 *
---

```

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1032.63 on 745 degrees of freedom
Residual deviance: 490.74 on 736 degrees of freedom
AIC: 510.74

Number of Fisher Scoring iterations: 5

95% Confidence term for all coefficient

confint(heart_glm13)

Waiting for profiling to be done...

	2.5 %	97.5 %
(Intercept)	-5.48437757	-1.76505093
Age	0.01334485	0.06503033
SexM	1.21625973	2.39270126
ChestPainTypeATA	-2.34842497	-0.99380289
ChestPainTypeNAP	-2.16919520	-1.00697143
ChestPainTypeTA	-2.43603526	-0.61385476
ExerciseAnginaY	0.42347136	1.44700022
Oldpeak	0.15650326	0.69790711
ST_SlopeFlat	0.22458553	2.23291298
ST_SlopeUp	-2.37035853	-0.20214809

Wald-Z Tests for all parameters

summary(heart_glm13)

Call:

```

glm(formula = HeartDisease ~ Age + Sex + ChestPainType + ExerciseAngina +
    Oldpeak + ST_Slope, family = binomial, data = heart)

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7265	-0.3928	-0.1227	0.4556	2.9284

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.60085	0.94673	-3.803	0.000143 ***
Age	0.03891	0.01316	2.958	0.003098 **
SexM	1.79156	0.29957	5.980	2.22e-09 ***
ChestPainTypeATA	-1.65550	0.34450	-4.806	1.54e-06 ***
ChestPainTypeNAP	-1.57932	0.29584	-5.338	9.38e-08 ***
ChestPainTypeTA	-1.51087	0.46339	-3.260	0.001112 **
ExerciseAnginaY	0.93541	0.26059	3.590	0.000331 ***
Oldpeak	0.42148	0.13782	3.058	0.002227 **
ST_SlopeFlat	1.23923	0.51047	2.428	0.015198 *

```
ST_SlopeUp      -1.27465    0.55154  -2.311 0.020828 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1032.63  on 745  degrees of freedom
Residual deviance:  490.74  on 736  degrees of freedom
AIC: 510.74

Number of Fisher Scoring iterations: 5
```