

Doctor Ratings Analysis

Danny Moncada (monca016), MSBA

```
## Load libraries
suppressWarnings(suppressPackageStartupMessages({
  library(readr)
  library(lsp)
}))
```

```
## Load data file
doctorRatings <- read_csv("doctorRatings.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
str(doctorRatings)
```

```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 1425 obs. of  21 variables:
## $ israted      : num  1 1 0 1 0 1 0 0 0 0 ...
## $ webrating     : num  5 5 NA 5 NA 5 NA NA NA NA ...
## $ ofratings     : num  4 6 0 1 0 2 0 0 0 0 ...
## $ surveyrating  : num  92 78 95 85 79 83 85 72 85 77 ...
## $ ratingBefore  : num  5 5 NA NA NA 4 NA NA NA NA ...
## $ ratingAfter   : num  5 NA NA NA NA ...
## $ isRatedBefore : num  1 1 0 0 0 1 0 0 0 0 ...
## $ isRatedAfter  : num  1 0 0 0 0 1 0 0 0 0 ...
## $ board         : num  1 1 0 1 1 1 1 1 0 1 ...
## $ gender        : num  1 1 1 1 1 1 1 1 0 1 ...
## $ experience    : num  31 30 44 30 31 37 38 38 41 30 ...
## $ peerrating    : num  4 0 0 0 0 3 0 0 0 0 ...
## $ howmanypatients: num  78 52 21 44 61 62 61 41 49 48 ...
## $ population    : num  176 527 291 291 488 ...
## $ rawzero       : num  7 7 9 12 14 14 12 5 7 15 ...
## $ ratedzero     : num  5 1 5 9 8 7 9 1 4 7 ...
## $ urban         : num  0 0 1 1 1 1 1 0 1 1 ...
## $ largeurban    : num  0 0 1 1 1 1 1 0 1 1 ...
## $ median        : num  82.9 57.3 55.9 55.9 53.6 ...
## $ denver        : num  1 1 1 1 1 1 1 1 1 1 ...
## $ memphis       : num  0 0 0 0 0 0 0 0 0 0 ...
## - attr(*, "spec")=
## .. cols(
## ..   israted = col_double(),
## ..   webrating = col_double(),
## ..   ofratings = col_double(),
## ..   surveyrating = col_double(),
## ..   ratingBefore = col_double(),
```

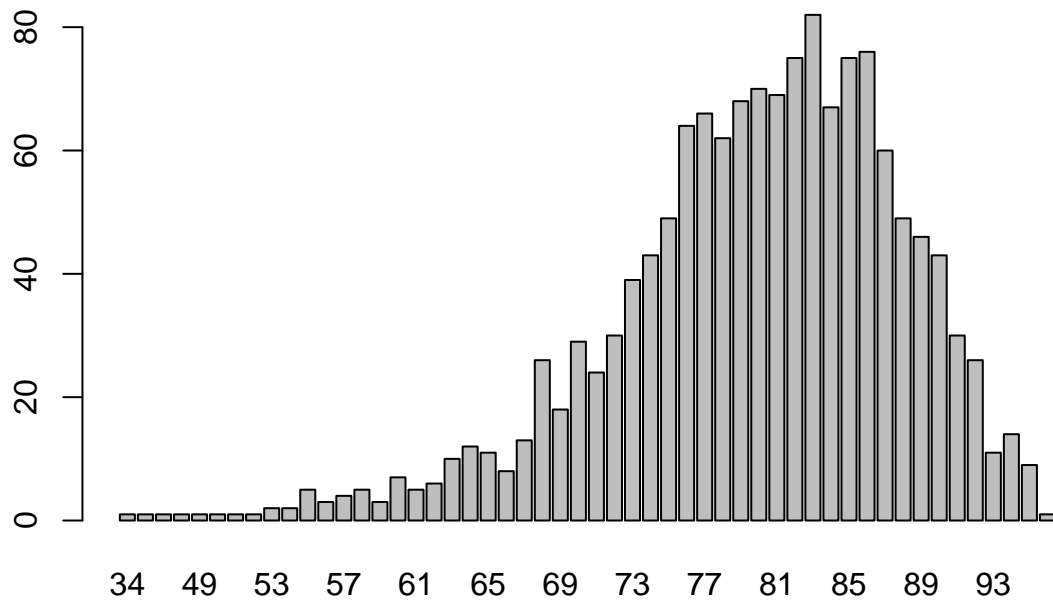
```
## .. ratingAfter = col_double(),
## .. isRatedBefore = col_double(),
## .. isRatedAfter = col_double(),
## .. board = col_double(),
## .. gender = col_double(),
## .. experience = col_double(),
## .. peerrating = col_double(),
## .. howmanypatients = col_double(),
## .. population = col_double(),
## .. rawzero = col_double(),
## .. ratedzero = col_double(),
## .. urban = col_double(),
## .. largeurban = col_double(),
## .. median = col_double(),
## .. denver = col_double(),
## .. memphis = col_double()
## .. )
```

```
summary(doctorRatings)
```

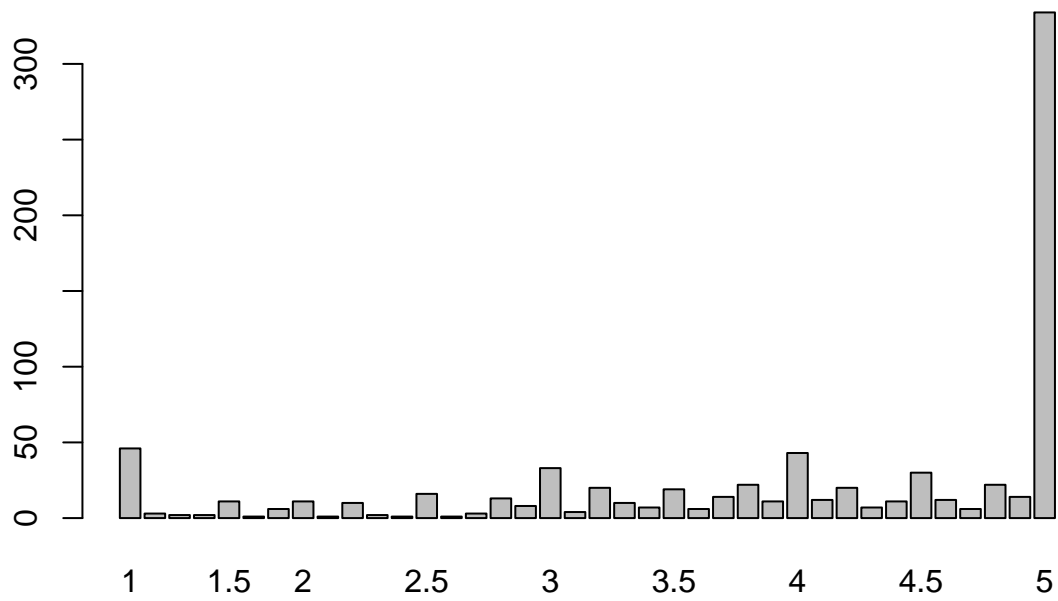
```
##      israted      webrating      ofratings      surveyrating
## Min.   :0.0000   Min.   :1.000   Min.    : 0.000   Min.    :34.00
## 1st Qu.:0.0000   1st Qu.:3.300   1st Qu.: 0.000   1st Qu.:75.00
## Median :1.0000   Median :4.500   Median : 1.000   Median :81.00
## Mean   :0.5572   Mean   :4.021   Mean    : 1.772   Mean   :79.78
## 3rd Qu.:1.0000   3rd Qu.:5.000   3rd Qu.: 2.000   3rd Qu.:86.00
## Max.   :1.0000   Max.   :5.000   Max.    :24.000   Max.    :96.00
##                      NA's    :631   NA's    :13
##      ratingBefore      ratingAfter      isRatedBefore      isRatedAfter
## Min.   :1.000   Min.   :1.000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:3.000   1st Qu.:2.500   1st Qu.:0.0000   1st Qu.:0.0000
## Median :5.000   Median :5.000   Median :0.0000   Median :0.0000
## Mean   :4.017   Mean   :3.761   Mean    :0.3326   Mean    :0.2365
## 3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:1.0000   3rd Qu.:0.0000
## Max.   :5.000   Max.   :5.000   Max.    :1.0000   Max.    :1.0000
## NA's    :951   NA's    :1088
##      board      gender      experience      peerrating
## Min.   :0.0000   Min.   :0.0000   Min.    : 5.00   Min.    : 0.0000
## 1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:16.00   1st Qu.: 0.0000
## Median :1.0000   Median :1.0000   Median :24.00   Median : 0.0000
## Mean   :0.7986   Mean   :0.6807   Mean    :23.97   Mean    : 0.1635
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:31.00   3rd Qu.: 0.0000
## Max.   :1.0000   Max.   :1.0000   Max.    :60.00   Max.    :12.0000
##
##      howmanypatients      population      rawzero      ratedzero
## Min.   : 10.00   Min.   : 8.969   Min.    : 0.00   Min.    : 0.000
## 1st Qu.: 44.00   1st Qu.:291.288   1st Qu.: 6.00   1st Qu.: 2.000
## Median : 54.00   Median :527.056   Median :12.00   Median : 6.000
## Mean   : 55.23   Mean   :518.122   Mean    :15.59   Mean    : 7.161
## 3rd Qu.: 66.00   3rd Qu.:654.880   3rd Qu.:21.00   3rd Qu.: 9.000
## Max.   :105.00   Max.   :897.472   Max.    :57.00   Max.    :27.000
##
##      urban      largeurban      median      denver
## Min.   :0.0000   Min.   :0.0000   Min.    :23.27   Min.    :0.0000
```

```
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:39.50 1st Qu.:0.0000
## Median :1.0000 Median :0.0000 Median :39.59 Median :0.0000
## Mean :0.5347 Mean :0.4028 Mean :47.21 Mean :0.2947
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:55.86 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :82.93 Max. :1.0000
##
## memphis
## Min. :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean :0.2688
## 3rd Qu.:1.0000
## Max. :1.0000
##
```

```
survey.table <- table(doctorRatings$surveyrating)
webrating.table <-table(doctorRatings$webrating)
barplot(survey.table)
```

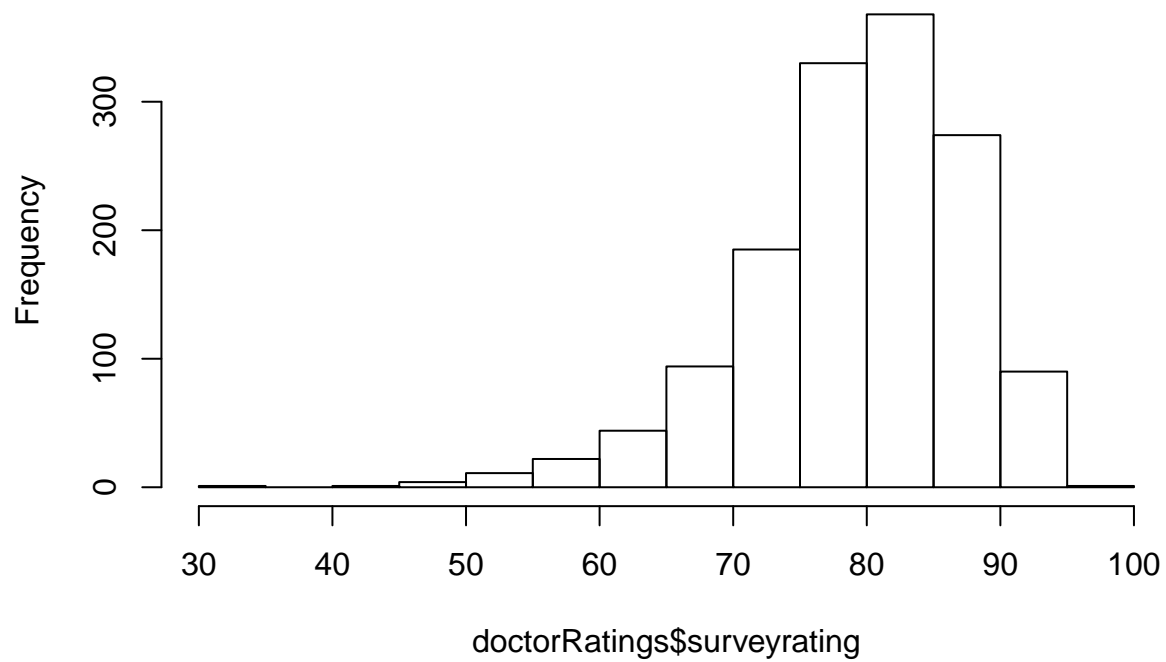


```
barplot(webrating.table)
```

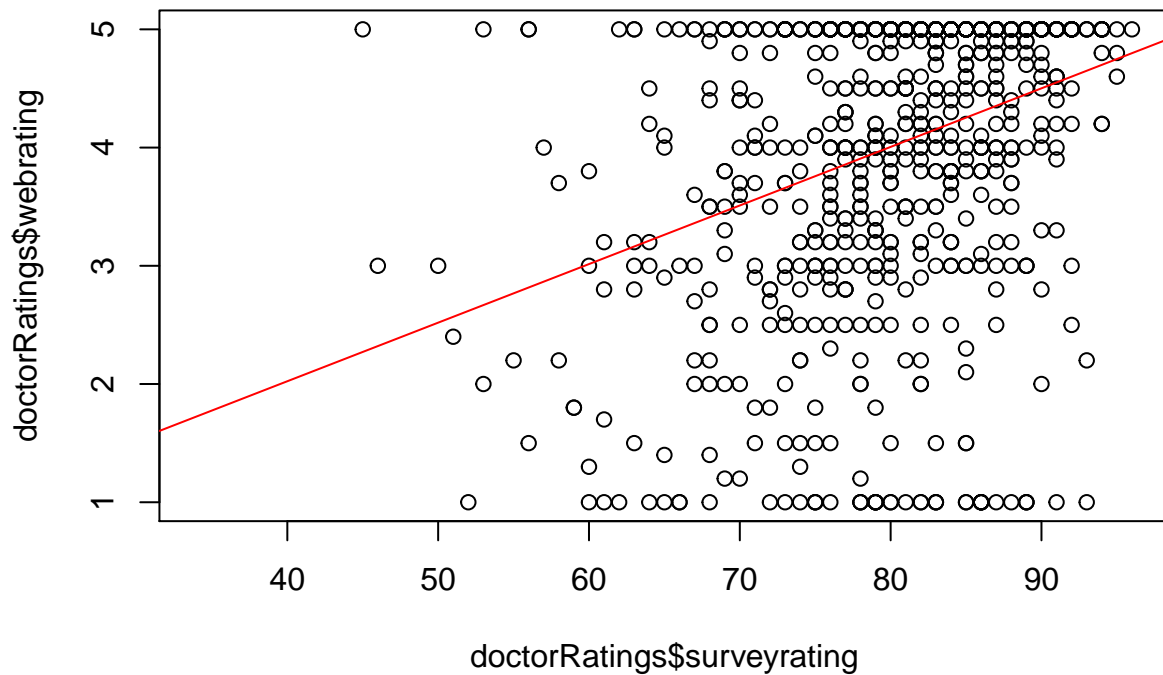


```
hist(doctorRatings$surveyrating)
```

Histogram of doctorRatings\$surveyrating



```
plot(doctorRatings$webrating ~ doctorRatings$surveyrating)
m1 <- lm(doctorRatings$webrating ~ doctorRatings$surveyrating)
coeff <- coefficients(m1)
abline(m1,col = 'red')
```



```
a1 <- aov(doctorRatings$surveyrating ~ as.factor(doctorRatings$israted))
anova(a1)
```

```
## Analysis of Variance Table
##
## Response: doctorRatings$surveyrating
##              Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(doctorRatings$israted)  1    483   483.46    7.223 0.007282 **
## Residuals                1423   95248    66.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## the following line gives the rating of a standard doctor (Intercept = 79.12837) and the rating
## if the doctor is rated (1.17)
a1$coefficients
```

```
##              (Intercept) as.factor(doctorRatings$israted)1
##              79.12837                                1.17264
```

```
TukeyHSD(a1)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
```

```
## Fit: aov(formula = doctorRatings$surveyrating ~ as.factor(doctorRatings$israted))
##
## $`as.factor(doctorRatings$israted)`
##      diff      lwr      upr      p adj
## 1-0 1.17264 0.3167363 2.028543 0.0072816
```

```
m2 <- lm(doctorRatings$israted ~ doctorRatings$surveyrating)
summary(m2)
```

```
##
## Call:
## lm(formula = doctorRatings$israted ~ doctorRatings$surveyrating)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6227 -0.5495  0.3988  0.4419  0.5926
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.213595   0.128521   1.662  0.09674 .
## doctorRatings$surveyrating 0.004307   0.001602   2.688  0.00728 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4958 on 1423 degrees of freedom
## Multiple R-squared:  0.00505,    Adjusted R-squared:  0.004351
## F-statistic: 7.223 on 1 and 1423 DF,  p-value: 0.007282
```

```
m3 <- lm(doctorRatings$israted ~ doctorRatings$surveyrating + doctorRatings$denver + doctorRatings$memp
      + doctorRatings$urban + doctorRatings$largeurban + doctorRatings$population + doctorRatings$me
      + doctorRatings$rawzero + doctorRatings$ratedzero + doctorRatings$experience + doctorRatings$ge
      + doctorRatings$board)
summary(m3)
```

```
##
## Call:
## lm(formula = doctorRatings$israted ~ doctorRatings$surveyrating +
##      doctorRatings$denver + doctorRatings$memphis + doctorRatings$urban +
##      doctorRatings$largeurban + doctorRatings$population + doctorRatings$median +
##      doctorRatings$rawzero + doctorRatings$ratedzero + doctorRatings$experience +
##      doctorRatings$gender + doctorRatings$board)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8047 -0.5313  0.3191  0.4389  0.7114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5.224e-02  1.589e-01  -0.329  0.74244
## doctorRatings$surveyrating  4.423e-03  1.661e-03   2.663  0.00783 **
## doctorRatings$denver       5.681e-02  3.396e-02   1.673  0.09454 .
## doctorRatings$memphis     -2.821e-02  4.116e-02  -0.685  0.49319
## doctorRatings$urban       9.139e-02  4.416e-02   2.069  0.03869 *
```

```
## doctorRatings$largeurban -4.829e-02 4.363e-02 -1.107 0.26867
## doctorRatings$population 5.183e-05 6.726e-05 0.771 0.44103
## doctorRatings$median 4.982e-03 1.585e-03 3.142 0.00171 **
## doctorRatings$rawzero 4.488e-03 2.100e-03 2.137 0.03277 *
## doctorRatings$ratedzero -3.552e-03 4.458e-03 -0.797 0.42573
## doctorRatings$experience -1.976e-03 1.582e-03 -1.250 0.21163
## doctorRatings$gender -3.503e-03 3.025e-02 -0.116 0.90783
## doctorRatings$board -4.865e-02 3.463e-02 -1.405 0.16027
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4915 on 1412 degrees of freedom
## Multiple R-squared: 0.02985, Adjusted R-squared: 0.02161
## F-statistic: 3.621 on 12 and 1412 DF, p-value: 2.305e-05

m4 <- glm(doctorRatings$israted ~ doctorRatings$surveyrating, family = binomial(link = 'logit'))
summary(m4)
```

```
##
## Call:
## glm(formula = doctorRatings$israted ~ doctorRatings$surveyrating,
##      family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.394  -1.263   1.009   1.080   1.341
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.162106    0.523299  -2.221  0.02637 *
## doctorRatings$surveyrating  0.017458    0.006533   2.672  0.00753 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1956.8  on 1424  degrees of freedom
## Residual deviance: 1949.6  on 1423  degrees of freedom
## AIC: 1953.6
##
## Number of Fisher Scoring iterations: 4
```

```
m5 <- glm(doctorRatings$israted ~ doctorRatings$surveyrating + doctorRatings$denver + doctorRatings$memph
      + doctorRatings$urban + doctorRatings$largeurban + doctorRatings$population + doctorRatings$me
      + doctorRatings$rawzero + doctorRatings$ratedzero + doctorRatings$experience + doctorRatings$ge
      + doctorRatings$board, family = binomial(link = 'logit'))
summary(m5)
```

```
##
## Call:
## glm(formula = doctorRatings$israted ~ doctorRatings$surveyrating +
##      doctorRatings$denver + doctorRatings$memphis + doctorRatings$urban +
##      doctorRatings$largeurban + doctorRatings$population + doctorRatings$median +
```



```
##      doctorRatings$rawzero + doctorRatings$ratedzero + doctorRatings$experience +
##      doctorRatings$gender + doctorRatings$board, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7468  -1.2323   0.8781   1.0723   1.5633
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.3064632   0.6665604  -3.460  0.00054 ***
## doctorRatings$surveyrating  0.0184403   0.0069220   2.664  0.00772 **
## doctorRatings$denver      0.2356938   0.1415558   1.665  0.09591 .
## doctorRatings$memphis    -0.1104209   0.1705518  -0.647  0.51735
## doctorRatings$urban      0.3846628   0.1859669   2.068  0.03860 *
## doctorRatings$largeurban -0.2049457   0.1834764  -1.117  0.26399
## doctorRatings$population  0.0002032   0.0002784   0.730  0.46540
## doctorRatings$median      0.0210115   0.0067232   3.125  0.00178 **
## doctorRatings$rawzero      0.0187282   0.0088590   2.114  0.03451 *
## doctorRatings$ratedzero   -0.0150779   0.0187325  -0.805  0.42088
## doctorRatings$experience  -0.0082872   0.0065821  -1.259  0.20801
## doctorRatings$gender      -0.0139505   0.1260866  -0.111  0.91190
## doctorRatings$board       -0.2059552   0.1447579  -1.423  0.15481
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1956.8  on 1424  degrees of freedom
## Residual deviance: 1913.7  on 1412  degrees of freedom
## AIC: 1939.7
##
## Number of Fisher Scoring iterations: 4
```

```
quantile(doctorRatings$surveyrating)
```

```
##      0%   25%   50%   75%  100%
##      34    75    81    86    96
```

```
##bins the bad doctors into sr1 and good doctors into sr4
```

```
doctorRatings$sr_1 <- ifelse(doctorRatings$surveyrating<= 75,1,0)
doctorRatings$sr_4 <- ifelse(doctorRatings$surveyrating>85, 1,0)
```

```
## shows that bad doctors are less likely to be rated, and not that good doctors are more likely to be
```

```
m6 <- glm(doctorRatings$israted ~ doctorRatings$sr_1 + doctorRatings$sr_4 + doctorRatings$denver + doctorRatings$urban + doctorRatings$largeurban + doctorRatings$population + doctorRatings$rawzero + doctorRatings$ratedzero + doctorRatings$experience + doctorRatings$gender + doctorRatings$board, family = binomial(link = 'logit'))
summary(m6)
```

```
##
## Call:
```

```
## glm(formula = doctorRatings$israted ~ doctorRatings$sr_1 + doctorRatings$sr_4 +
##   doctorRatings$denver + doctorRatings$memphis + doctorRatings$urban +
##   doctorRatings$largeurban + doctorRatings$population + doctorRatings$median +
##   doctorRatings$rawzero + doctorRatings$ratedzero + doctorRatings$experience +
##   doctorRatings$gender + doctorRatings$board, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7281  -1.2349   0.8753   1.0682   1.5855
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.7542669   0.4136374  -1.823  0.06823 .
## doctorRatings$sr_1    -0.3930728   0.1340157  -2.933  0.00336 **
## doctorRatings$sr_4     0.0514938   0.1348515   0.382  0.70257
## doctorRatings$denver    0.2418689   0.1418693   1.705  0.08822 .
## doctorRatings$memphis  -0.1151720   0.1706789  -0.675  0.49981
## doctorRatings$urban     0.3863104   0.1861787   2.075  0.03799 *
## doctorRatings$largeurban -0.2182108   0.1839323  -1.186  0.23548
## doctorRatings$population  0.0002160   0.0002785   0.776  0.43796
## doctorRatings$median     0.0208492   0.0067228   3.101  0.00193 **
## doctorRatings$rawzero    0.0187347   0.0088602   2.114  0.03447 *
## doctorRatings$ratedzero -0.0145250   0.0187155  -0.776  0.43769
## doctorRatings$experience -0.0083875   0.0065940  -1.272  0.20338
## doctorRatings$gender    -0.0166727   0.1262012  -0.132  0.89490
## doctorRatings$board     -0.1919796   0.1447263  -1.327  0.18467
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1956.8  on 1424  degrees of freedom
## Residual deviance: 1910.3  on 1411  degrees of freedom
## AIC: 1938.3
##
## Number of Fisher Scoring iterations: 4
```

```
## are web ratings dependent on survey ratings?
```

```
## subsets doctors into only those who are rated
```

```
doctorsRated <- doctorRatings[!is.na(doctorRatings$webrating),]
doctorsRated
```

```
## # A tibble: 794 x 23
```

```
##   israted webrating ofratings surveyrating ratingBefore ratingAfter
##   <dbl>     <dbl>     <dbl>         <dbl>         <dbl>         <dbl>
## 1      1      5      4      92      5      5
## 2      1      5      6      78      5     NA
## 3      1      5      1      85     NA     NA
## 4      1      5      2      83      4    1.36
## 5      1    4.20      2      76      5     NA
## 6      1      5      2      87     NA      5
## 7      1      5      1      77      5      1
## 8      1      5      1      80     NA     NA
```

```
## 9      1      4      2      86      3.61      2.19
## 10     1      1      2      78      5        NA
## # ... with 784 more rows, and 17 more variables: isRatedBefore <dbl>,
## #   isRatedAfter <dbl>, board <dbl>, gender <dbl>, experience <dbl>,
## #   peerrating <dbl>, howmanypatients <dbl>, population <dbl>,
## #   rawzero <dbl>, ratedzero <dbl>, urban <dbl>, largeurban <dbl>,
## #   median <dbl>, denver <dbl>, memphis <dbl>, sr_1 <dbl>, sr_4 <dbl>
```

```
reg1 <- lm(doctorsRated$webrating ~ doctorsRated$surveyrating)
summary(reg1)
```

```
##
## Call:
## lm(formula = doctorsRated$webrating ~ doctorsRated$surveyrating)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6498 -0.5552  0.3989  0.7962  2.7287
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.041485   0.408543   0.102   0.919
## doctorsRated$surveyrating 0.049552   0.005062   9.788 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.148 on 792 degrees of freedom
## Multiple R-squared:  0.1079, Adjusted R-squared:  0.1068
## F-statistic: 95.81 on 1 and 792 DF, p-value: < 2.2e-16
```

```
##splines?
```

```
## breaks the regression line into two distinct buckets
```

```
## ex. MPG vs. age
```

```
reg1 <- lm(doctorsRated$webrating ~ lspline(doctorsRated$surveyrating, c(76,83)))
summary(reg1)
```

```
##
## Call:
## lm(formula = doctorsRated$webrating ~ lspline(doctorsRated$surveyrating,
##      c(76, 83)))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4916 -0.5458  0.4500  0.7166  2.5940
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)    0.50880   0.81535   0.624
## lspline(doctorsRated$surveyrating, c(76, 83))1  0.04216   0.01142   3.693
## lspline(doctorsRated$surveyrating, c(76, 83))2  0.08149   0.01909   4.269
## lspline(doctorsRated$surveyrating, c(76, 83))3  0.02083   0.01745   1.194
##              Pr(>|t|)
```

```
## (Intercept) 0.532788
## lspline(doctorsRated$surveyrating, c(76, 83))1 0.000237 ***
## lspline(doctorsRated$surveyrating, c(76, 83))2 2.2e-05 ***
## lspline(doctorsRated$surveyrating, c(76, 83))3 0.232920
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.147 on 790 degrees of freedom
## Multiple R-squared:  0.1119, Adjusted R-squared:  0.1085
## F-statistic: 33.19 on 3 and 790 DF, p-value: < 2.2e-16
```

*## the interpretation of this could be that web rating isn't valuable from good doctors
good for differentiating doctors on the low end or in the intermediary*

```
reg1 <- glm(ofratings ~ surveyrating, family = 'poisson', data = doctorRatings)
summary(reg1)
```

```
##
## Call:
## glm(formula = ofratings ~ surveyrating, family = "poisson", data = doctorRatings)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8899  -1.8823  -0.6326   0.1700   8.9702
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.5857160  0.1951550   3.001  0.00269 **
## surveyrating -0.0001709  0.0024337  -0.070  0.94402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 4552  on 1411  degrees of freedom
## Residual deviance: 4552  on 1410  degrees of freedom
## (13 observations deleted due to missingness)
## AIC: 6747
##
## Number of Fisher Scoring iterations: 6
```