world_happiness 2019/2018

Nirmal Sai Swaroop Janapaneedi

INTRODUCTION

Leading experts across fields describe how measurements of well-being can be used effectively to assess the progress of nations.

This project review the state of happiness in the world in 2019 vs 2018 and show how the new science of happiness explains personal and national variations in happiness.

—— GOAL ——

In this project, i will be asking the following questions:

1. To be able to make a prediction model that will help us predict happiness and score based on multiple variables?

2. What countries or regions rank the highest in overall happiness and each of the six factors contributing to happiness?

3. How did country ranks or scores change between the 2015 and 2016 as well as the 2016 and 2017 reports?

4. Did any country experience a significant increase or decrease in happiness?

import libraries

—— Import dataset ——

```
data <- read.csv("2019.csv")
```

—— Explore and visualize data ——

This shows an organized section of data frame

glimpse(data)

```
## Rows: 156
## Columns: 9
## $ Overall.rank             <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,...
## $ Country.or.region        <fct> Finland, Denmark, Norway, Iceland, Net...
## $ Score                    <dbl> 7.769, 7.600, 7.554, 7.494, 7.488, 7.4...
## $ GDP.per.capita           <dbl> 1.340, 1.383, 1.488, 1.380, 1.396, 1.4...
## $ Social.support           <dbl> 1.587, 1.573, 1.582, 1.624, 1.522, 1.5...
## $ Healthy.life.expectancy  <dbl> 0.986, 0.996, 1.028, 1.026, 0.999, 1.0...
## $ Freedom.to.make.life.choices <dbl> 0.596, 0.592, 0.603, 0.591, 0.557, 0.5...
## $ Generosity               <dbl> 0.153, 0.252, 0.271, 0.354, 0.322, 0.2...
## $ Perceptions.of.corruption <dbl> 0.393, 0.410, 0.341, 0.118, 0.298, 0.3...
```

structure of the data frame

str(data)

```
## 'data.frame':        156 obs. of    9 variables:
##  $ Overall.rank              : int   12345      6 7  8910...
##  $ Country.or.region         : Factor w/   156 levels "Afghanistan",..: 44 37 106 58 99 134 133 100
##  $ Score                     : num   7.77  7.6  7.55  7.49 7.49 ...
##  $ GDP.per.capita            : num   1.34  1.38  1.49   1.38 1.4 ...
##  $ Social.support            : num   1.59  1.57  1.58   1.62 1.52 ...
##  $ Healthy.life.expectancy   : num   0.986 0.996 1.028 1.026 0.999 ...
##  $ Freedom.to.make.life.choices: num  0.596 0.592 0.603 0.591 0.557 0.572 0.574 0.585 0.584 0.532 ..
##  $ Generosity                : num  0.153 0.252 0.271 0.354 0.322 0.263 0.267 0.33 0.285 0.244 ...
##  $ Perceptions.of.corruption : num  0.393 0.41 0.341 0.118         0.298 0.343 0.373 0.38 0.308 0.226 ...
```

summary(data)

```
##    Overall.rank           Country.or.region        Score          GDP.per.capita
## Min.      : 1.00   Afghanistan:    1     Min.      :2.853   Min.      :0.0000
## 1st Qu.: 39.75     Albania    :    1      1st Qu.:4.545      1st Qu.:0.6028
## Median : 78.50     Algeria    :    1     Median :5.380     Median :0.9600
## Mean      : 78.50   Argentina  :    1     Mean      :5.407   Mean      :0.9051
##    3rd Qu.:117.25   Armenia    :    1      3rd Qu.:6.184      3rd Qu.:1.2325
## Max.      :156.00   Australia  :    1     Max.      :7.769   Max.      :1.6840
##                     (Other)    :150
## Social.support  Healthy.life.expectancy Freedom.to.make.life.choices
 ## Min.  :0.000     Min.  :0.0000           Min.  :0.0000
 ## 1st Qu.:1.056    1st Qu.:0.5477          1st Qu.:0.3080
 ## Median :1.272    Median :0.7890          Median :0.4170
```

```
##  Mean      :1.209   Mean      :0.7252        Mean    :0.3926
##   3rd Qu.:1.452     3rd Qu.:0.8818           3rd Qu.:0.5072
##  Max.      :1.624   Max.      :1.1410        Max.    :0.6310
##
##     Generosity      Perceptions.of.corruption
##   Min.  :0.0000   Min. :0.0000
##   1st Qu.:0.1087  1st Qu.:0.0470
##   Median :0.1775  Median :0.0855
##   Mean   :0.1848     Mean   :0.1106
##   3rd Qu.:0.2482     3rd Qu.:0.1412
##   Max.   :0.5660  Max.   :0.4530
##
```

shows highest 10 scores

```
head(data, n=10)
```

```
##     Overall.rank Country.or.region Score GDP.per.capita Social.support
## 1              1           Finland 7.769         1.340          1.587
## 2              2           Denmark 7.600         1.383          1.573
## 3              3            Norway 7.554         1.488          1.582
## 4              4           Iceland 7.494         1.380          1.624
## 5              5       Netherlands 7.488         1.396          1.522
## 6              6       Switzerland 7.480         1.452          1.526
## 7              7            Sweden 7.343         1.387          1.487
## 8              8       New Zealand 7.307         1.303          1.557
## 9              9            Canada 7.278         1.365          1.505
## 10            10           Austria 7.246         1.376          1.475
##     Healthy.life.expectancy Freedom.to.make.life.choices Generosity
## 1                     0.986                        0.596      0.153
## 2                     0.996                        0.592      0.252
## 3                     1.028                        0.603      0.271
## 4                     1.026                        0.591      0.354
## 5                     0.999                        0.557      0.322
## 6                     1.052                        0.572      0.263
## 7                     1.009                        0.574      0.267
## 8                     1.026                        0.585      0.330
## 9                     1.039                        0.584      0.285
## 10                    1.016                        0.532      0.244
##     Perceptions.of.corruption
## 1                       0.393
## 2                       0.410
## 3                       0.341
## 4                       0.118
## 5                       0.298
## 6                       0.343
## 7                       0.373
## 8                       0.380
## 9                       0.308
## 10                      0.226
```

shows lowest 10 scores

```
tail(data, n=10)
```

```
##         Overall.rank              Country.or.region Score GDP.per.capita Social.support
## 147           147                           Haiti 3.597         0.323          0.688
## 148           148                        Botswana 3.488         1.041          1.145
## 149           149                           Syria 3.462         0.619          0.378
## 150           150                          Malawi 3.410         0.191          0.560
## 151           151                           Yemen 3.380         0.287          1.163
## 152           152                          Rwanda 3.334         0.359          0.711
## 153           153                        Tanzania 3.231         0.476          0.885
## 154           154                     Afghanistan 3.203         0.350          0.517
## 155           155        Central African Republic 3.083         0.026          0.000
## 156           156                     South Sudan 2.853         0.306          0.575
##         Healthy.life.expectancy Freedom.to.make.life.choices Generosity
## 147                       0.449                        0.026      0.419
## 148                       0.538                        0.455      0.025
## 149                       0.440                        0.013      0.331
## 150                       0.495                        0.443      0.218
## 151                       0.463                        0.143      0.108
## 152                       0.614                        0.555      0.217
## 153                       0.499                        0.417      0.276
## 154                       0.361                        0.000      0.158
## 155                       0.105                        0.225      0.235
## 156                       0.295                        0.010      0.202
##         Perceptions.of.corruption
## 147                         0.110
## 148                         0.100
## 149                         0.141
## 150                         0.089
## 151                         0.077
## 152                         0.411
## 153                         0.147
## 154                         0.025
## 155                         0.035
## 156                         0.091
```

histogram of scores

"Rich countries are definitely happier than poor countries," he said. "It's no joy to be poor." But unlike the United States, where a loss of faith in institutions has dovetailed with the drop in happiness reported by Americans, people in Scandinavian countries believe in one another and their governments, Professor Sachs said.In Finland, 91 percent of survey respondents reported being satisfied with their president and 86 percent said they trust the police.
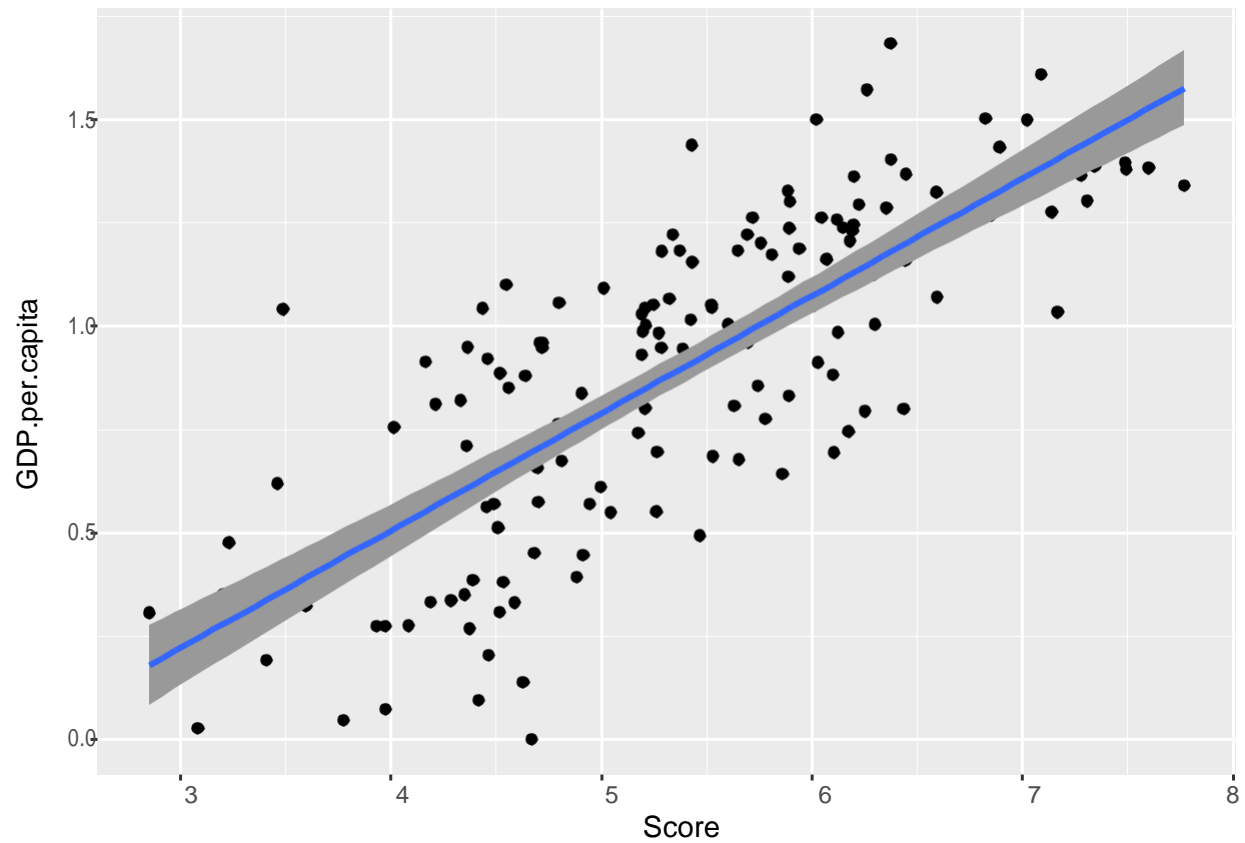
```
hist(data$Score, freq=TRUE, col="black", border="white", main="2019
    Happiness Scores", xlab="Score", ylab="Count")
```

2019 Happiness Scores

plot gdp per cap vs score

```
ggplot(data = data, aes(x = Score, y = GDP.per.capita)) +
  geom_point(color='black') +
  geom_smooth(method = "lm", se = TRUE)
```
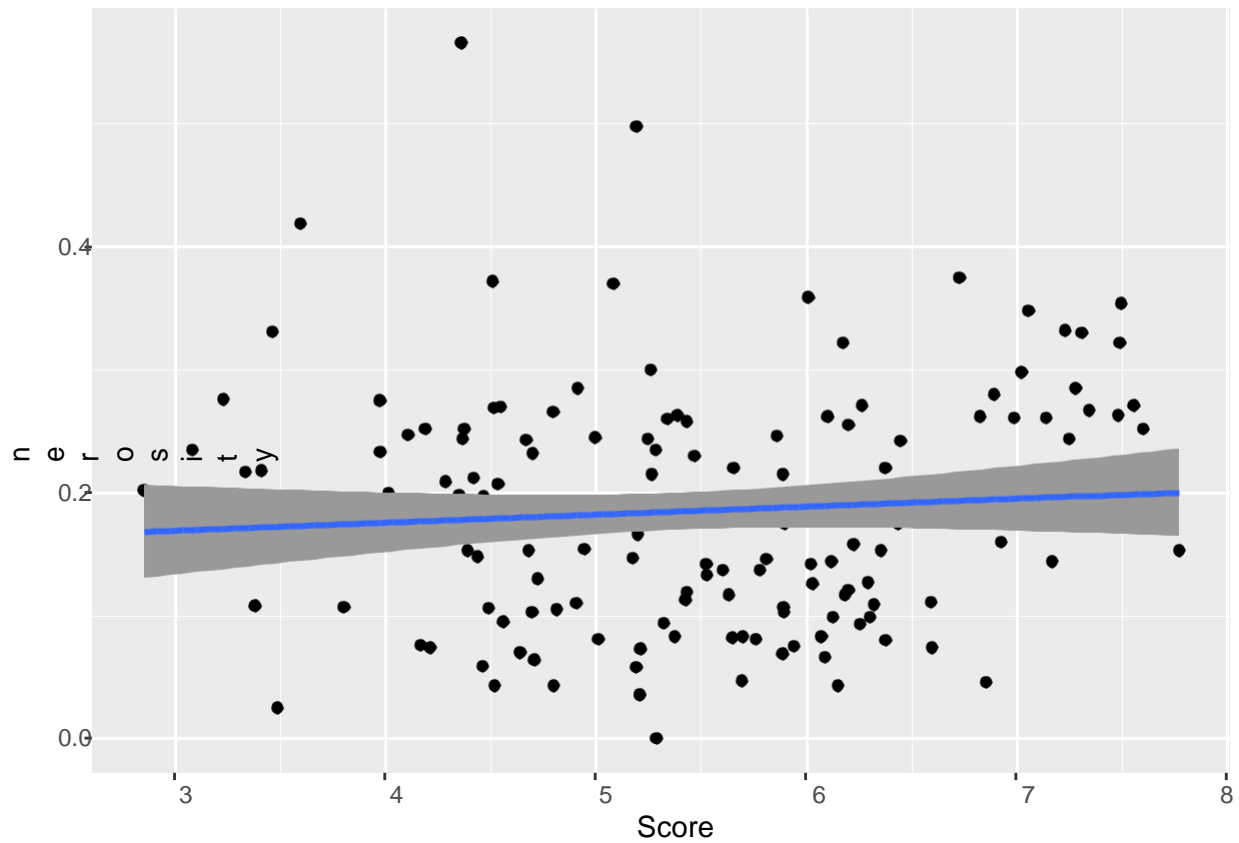
`geom_smooth()` using formula 'y ~ x'

plot generosity vs score

```
ggplot(data = data, aes(x = Score, Generosity)) +
  geom_point(color='black') +
  geom_smooth(method = "lm", se = TRUE)
```
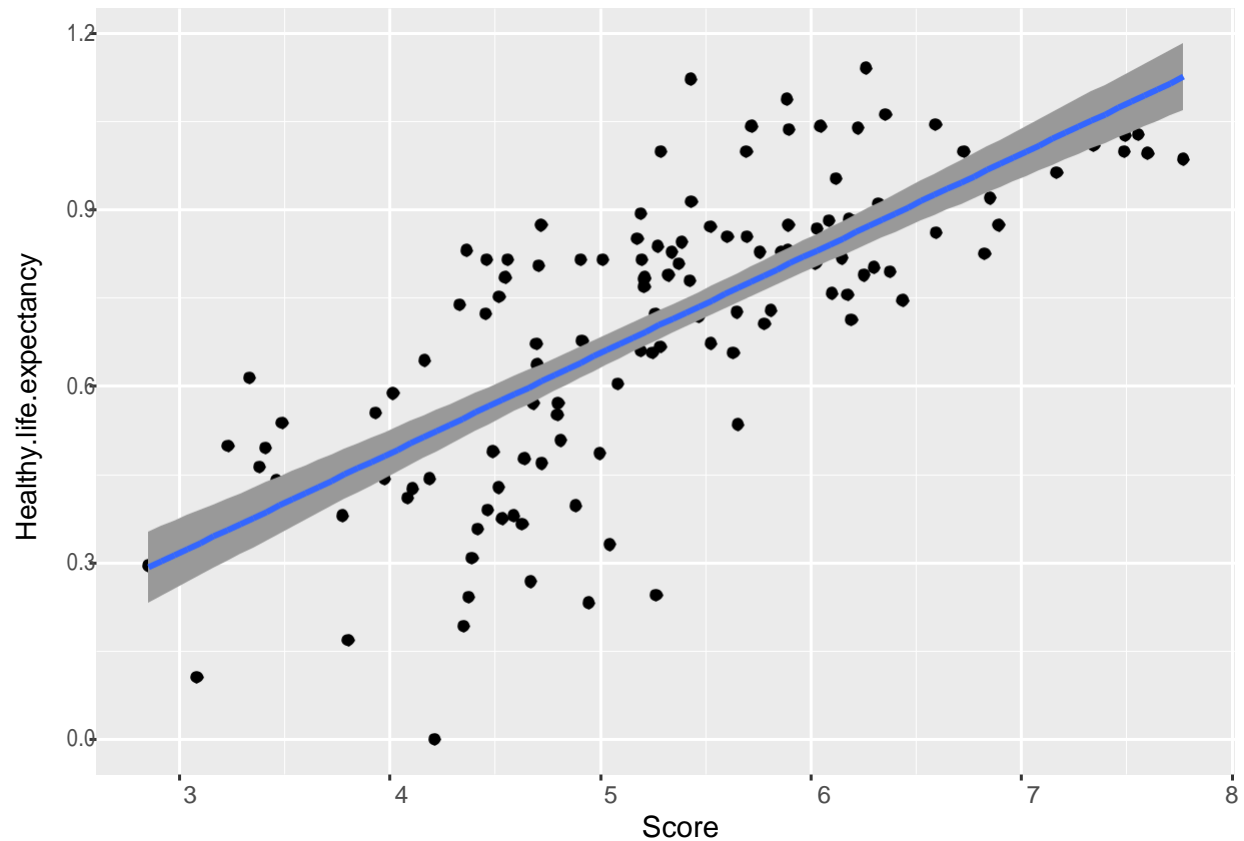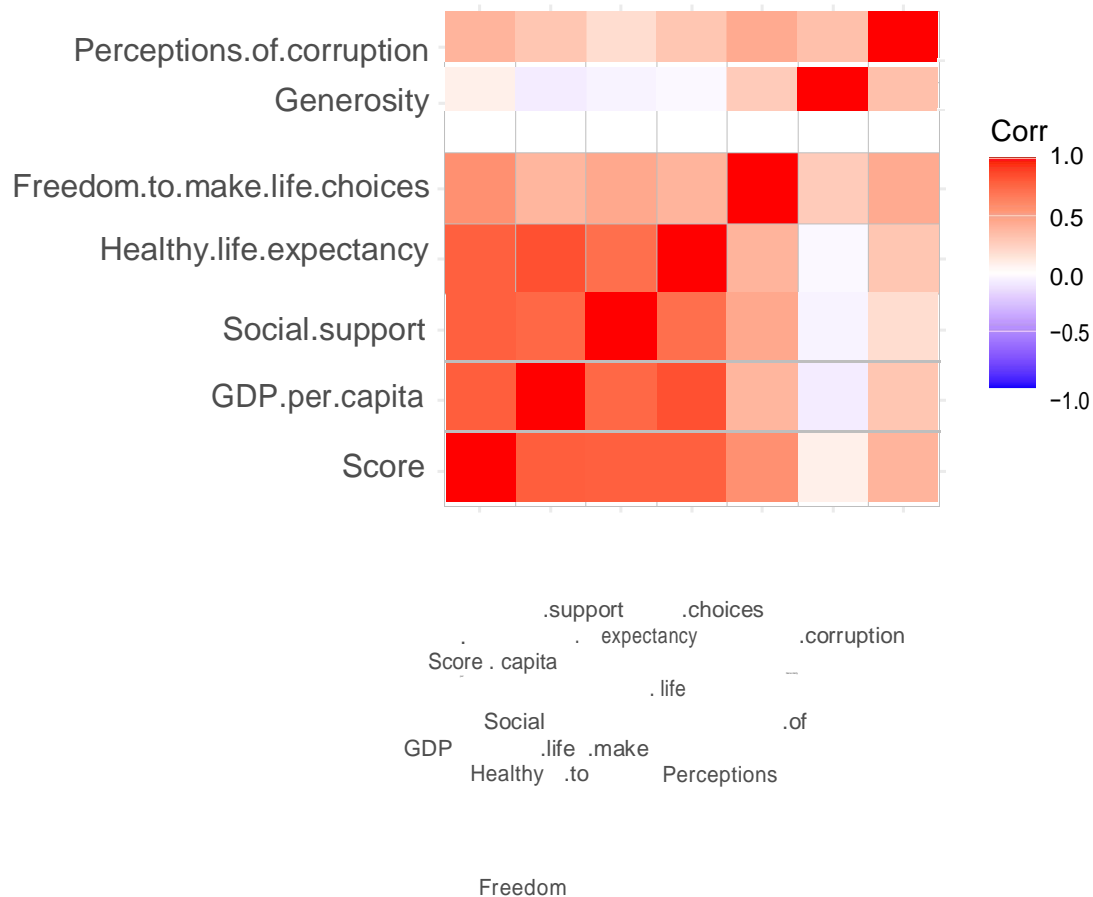
`geom_smooth()` using formula 'y ~ x'

plot life expectancy vs score

```
ggplot(data = data, aes(x = Score, Healthy.life.expectancy)) +
  geom_point(color='black') +
  geom_smooth(method = "lm", se = TRUE)
```

`geom_smooth()` using formula 'y ~ x'

Thus, moving forward, let us explore the correlation between the other predictors (economy, family, trust, etc.) and happiness score

```
temp <- data[, c(3,4,5,6,7,8,9)]
cormat <- signif(cor(temp), 2)
ggcorrplot(cormat)
```

Perceptions.of.corruption

Generosity

Freedom.to.make.life.choices

Healthy.life.expectancy

Social.support

GDP.per.capita

Score

Corr
1.0
0.5
0.0
−0.5
−1.0

.support   .choices
        .   expectancy   .corruption
Score . capita
                              . life
          Social   .of
GDP   .life .make
    Healthy .to   Perceptions

Freedom

—— Summation model ——

find predicted score by sum method and calculate the corresponding RMSE

```
sum_model <- data %>% mutate(pred_score = GDP.per.capita +
                          Social.support +
                          Healthy.life.expectancy +
                          Freedom.to.make.life.choices +
                          Generosity +
                          Perceptions.of.corruption +
                          1.85,
                       RMSE = RMSE(Score, pred_score))
```

show top results of the summation model

```
sum_model %>%
  filter(Overall.rank <= 5) %>%
  select(Overall.rank, Country.or.region, Score, pred_score, RMSE)
```

```
##     Overall.rank Country.or.region Score pred_score     RMSE
## 1             1            Finland 7.769       6.905 0.5280065
## 2             2            Denmark 7.600       7.056 0.5280065
## 3             3             Norway 7.554       7.163 0.5280065
## 4             4            Iceland 7.494       6.943 0.5280065
## 5             5        Netherlands 7.488       6.944 0.5280065
```

——— RESULT ———

——— Develop generalized linear model ———

— test for an appropriate probability in data partitioning

```
ps <- seq(from=.30, to=.90, by=.01)

rmses <- sapply(ps, function(p){
  train_index <- createDataPartition(data$Score,
                                     times=1,
                                     p=p,
                                     list=FALSE)
  train <- data[train_index,]
  test <- data[-train_index,]
  fit <- glm(Score ~ GDP.per.capita +
                 Social.support +
                 Healthy.life.expectancy +
                 Freedom.to.make.life.choices +
                 Generosity +
                 Perceptions.of.corruption,
             data = train)
  test <- test %>%
    mutate(pred_score = predict.glm(fit, newdata=test))
  RMSE(test$Score, test$pred_score)
})
```

no real clear winner in terms of best accuracy in probabilities

```
plot(ps, rmses)
```
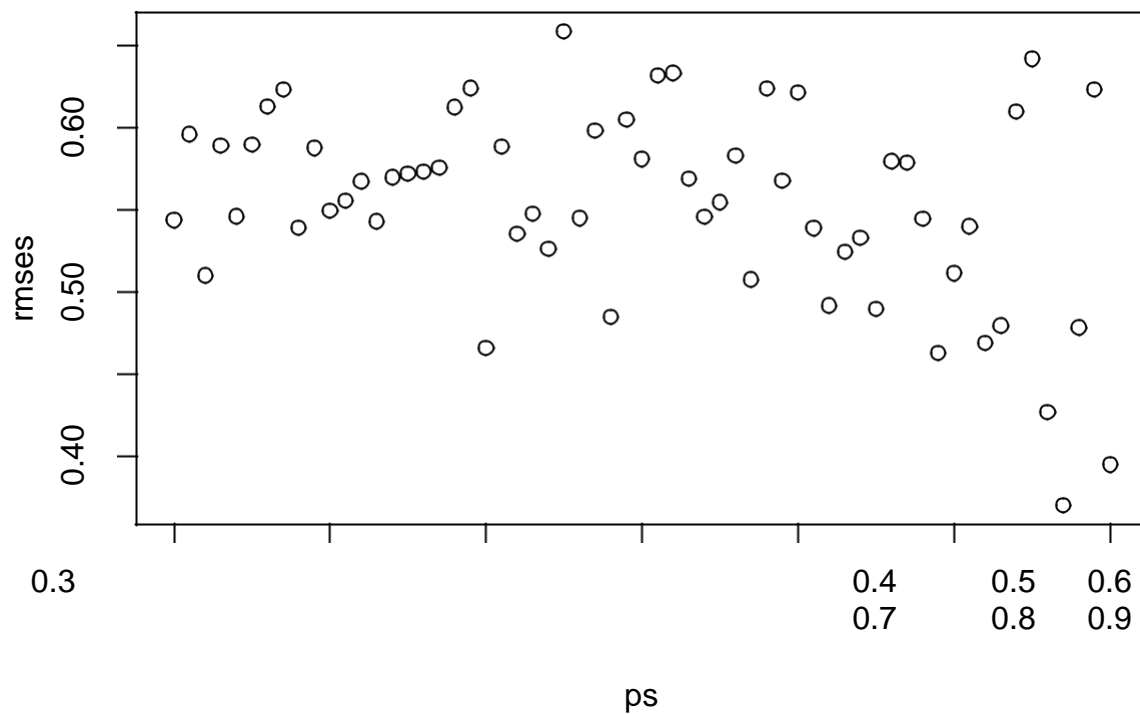
```
ps[which.min(rmses)]
```

```
## [1] 0.87
```

```
min(rmses)
```

```
## [1] 0.3698972
```

```
rm(ps, rmses)
```

—— Data partitioning ——

just using p=0.70

```
train_index <- createDataPartition(data$Score, times=1, p=0.70, list=FALSE)
train <- data[train_index,]
test <- data[-train_index,]
```

— back to our model, p used above in data separation

```
fit <- glm(Score ~ GDP.per.capita +
                Social.support +
                Healthy.life.expectancy +
                Freedom.to.make.life.choices +
                Generosity +
                Perceptions.of.corruption,
            data = train)
```

add predicted scores to our 'data' data frame

```
results <- test %>%
  mutate(pred_score = predict.glm(fit, newdata=test))
```

plot predicted scores vs actual scores

also plot 1 to 1 line

```
ggplot(data = results, aes(Score, pred_score)) +
  geom_point(color='black') +
  geom_smooth(method = "lm", se = TRUE) +
  geom_abline(color='red')
```

```
## `geom_smooth()` using formula 'y ~ x'
```
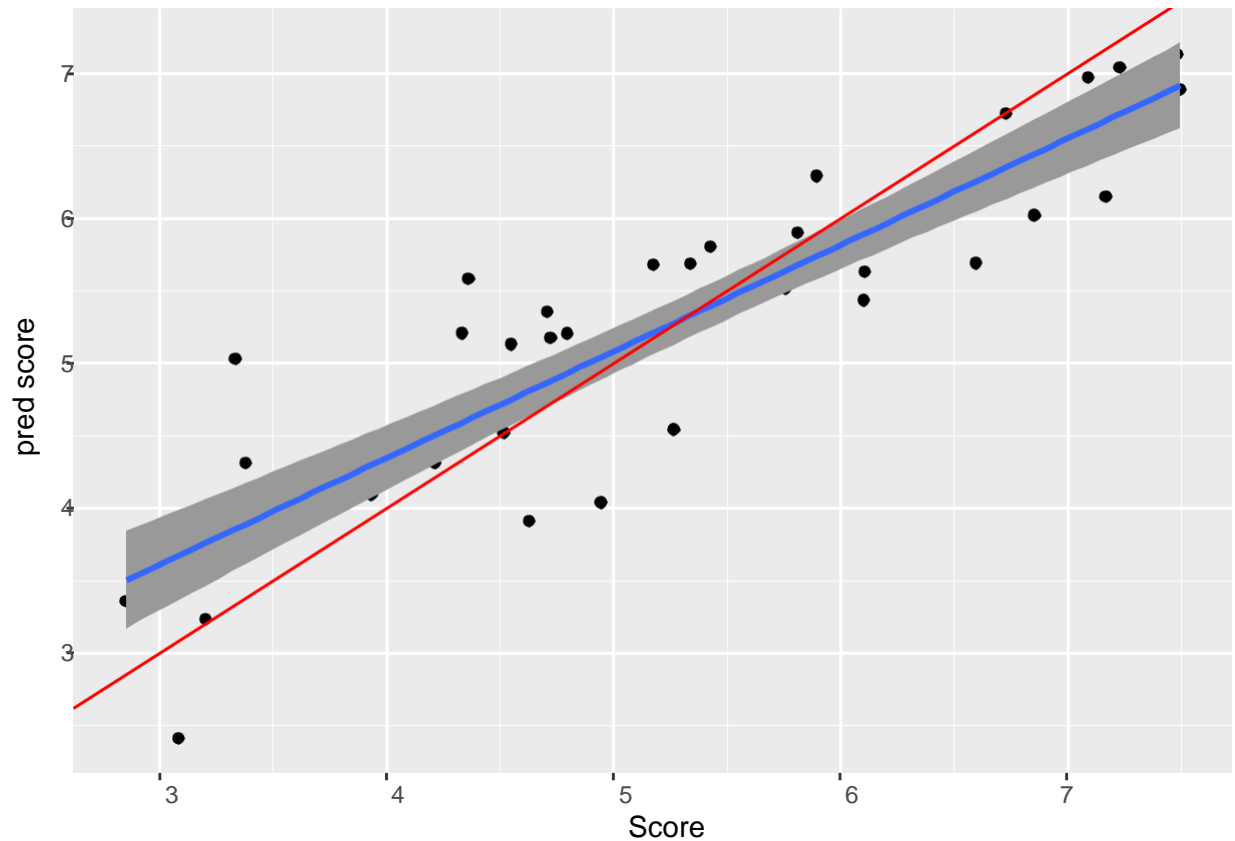
print rmse

RMSE(results$Score, results$pred_score)

[1] 0.5865818

print coefficients of fitted model

fit$coefficients

```
##                  (Intercept)              GDP.per.capita
##                    1.8080559                   0.5891217
##               Social.support      Healthy.life.expectancy
##                    1.2405204                   1.1535704
## Freedom.to.make.life.choices                   Generosity
##                    1.0131395                   0.7612858
##      Perceptions.of.corruption
##                    1.6900847
```

——— try fitting ALL year data. Country/region and rank is not used anyways, just focus on the scores and other components.

cannot use beyond 2019/2019 because the parameters were changed that year

——— Import 2018 dataset and keep only used columns

```
data18 <- read.csv("2018.csv")
data18$Overall.rank <- NULL
data18$Country.or.region <- NULL
```

remove unused columns in 19 data frame for merging

```
data$Overall.rank <- NULL
data$Country.or.region <- NULL
```

turn corruption column from factor to numeric, turn NAs to 0

```
data18$Perceptions.of.corruption <- as.numeric(as.character(data18$Perceptions.of.corruption))
```

```
## Warning: NAs introduced by coercion
```

```
data18[is.na(data18)] <- 0
```

full data set of both 2019, 2018 data

```
full_data <- rbind(data, data18)
```

——— full data fit ——–

partition full dataset

```r
train_index <- createDataPartition(full_data$Score,
                                    times=1,
                                    p=0.70,
                                    list=FALSE)
train <- full_data[train_index,]
test <- full_data[-train_index,]
```

fit a model on train data

```r
fit_full <- glm(Score ~ GDP.per.capita +
                   Social.support +
                   Healthy.life.expectancy +
                   Freedom.to.make.life.choices +
                   Generosity +
                   Perceptions.of.corruption,
                data = train)
```
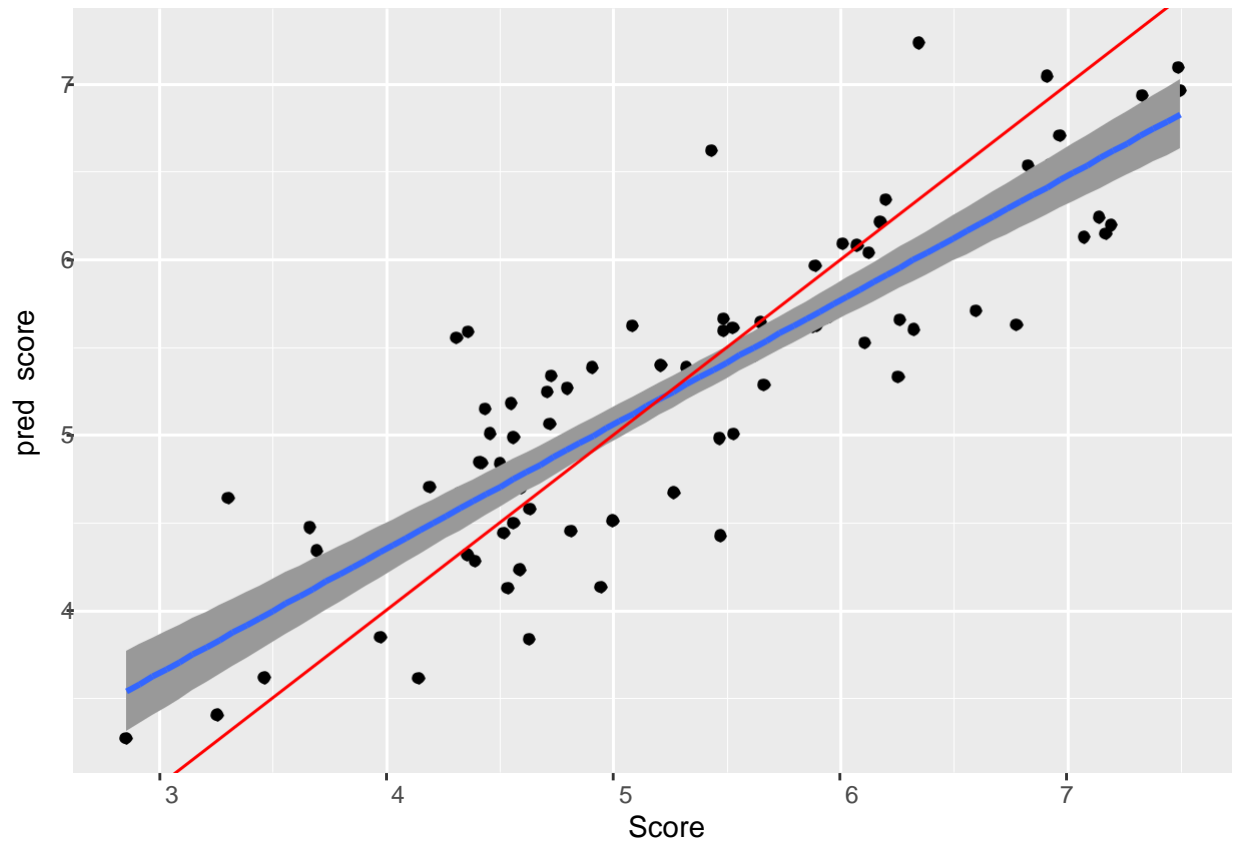
add predicted scores to our 'data' data frame

```r
results <- test %>%
  mutate(pred_score = predict.glm(fit_full, newdata=test))
```

plot predicted scores vs actual scores

also plot 1 to 1 line

```r
ggplot(data = results, aes(Score, pred_score)) +
  geom_point(color='black') +
  geom_smooth(method = "lm", se = TRUE) +
  geom_abline(color='red')
```

```
## `geom_smooth()` using formula 'y ~ x'
```

print rmse

```
RMSE(results$Score, results$pred_score)
```

[1] 0.541093

print coefficients of fitted model

```
fit_full$coefficients
```

```
##                   (Intercept)              GDP.per.capita
##                     1.8603015                   0.8618429
##                Social.support        Healthy.life.expectancy
##                     1.0683850                   0.8883216
## Freedom.to.make.life.choices                    Generosity
##                     1.4075004                   0.7519630
##     Perceptions.of.corruption
##                     1.1536190
```

—— try fit again without generosity

```
fit_full_2 <- glm(Score ~ GDP.per.capita +
                    Social.support +
                    Healthy.life.expectancy +
                    Freedom.to.make.life.choices +
                    Perceptions.of.corruption,
                data = train)
```
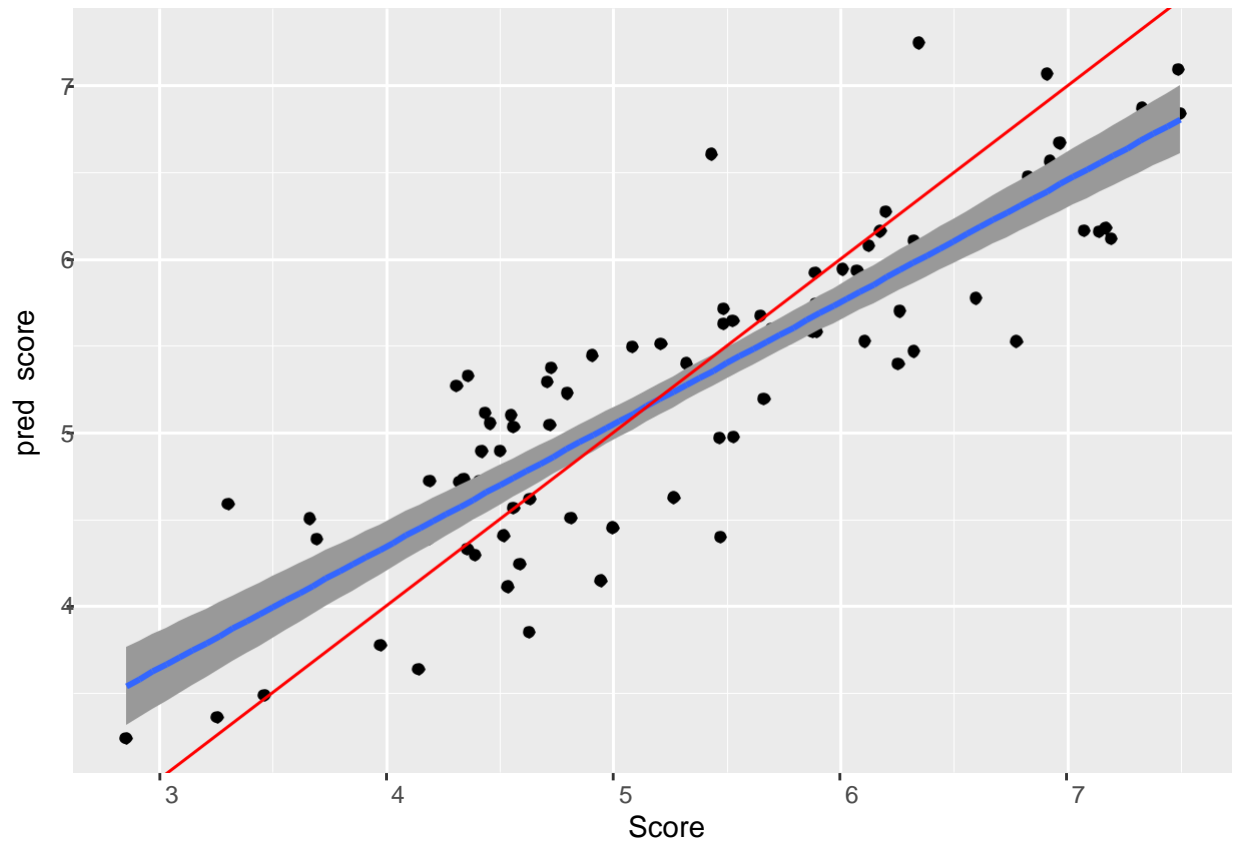
add predicted scores to our 'data' data frame

```
results <- test %>%
  mutate(pred_score = predict.glm(fit_full_2, newdata=test))
```

plot predicted scores vs actual scores

also plot 1 to 1 line

```
ggplot(data = results, aes(Score, pred_score)) +
  geom_point(color='black') +
  geom_smooth(method = "lm", se = TRUE) +
  geom_abline(color='red')
```

```
## `geom_smooth()` using formula 'y ~ x'
```

—— using rmse instead of cor: https://www.r-bloggers.com/dont-use-correlation-to-track-prediction-performance/

print residual sum of squares

```
RMSE(results$Score, results$pred_score)
```

[1] 0.5392935

—— CONCLUSION ——

The study grew out of a 2011 resolution passed by the U.N. General Assembly that called on governments to "give more importance to happiness and well-being in determining how to achieve and measure social and economic development." Over the years, the study has examined how social media, migration, differences in well-being inequality and other factors affect levels of happiness. This year, the study ranked cities around the world by their reported sense of well-being.

Not surprisingly, the happiest city in the world is Helsinki, the capital of Finland.

print coefficients of fitted model

```
fit_full_2$coefficients
```

```
##                  (Intercept)             GDP.per.capita
##                    1.9668121                  0.8098116
##               Social.support     Healthy.life.expectancy
##                    1.0633173                  0.9187529
## Freedom.to.make.life.choices     Perceptions.of.corruption
##                    1.4874223                  1.3898656
```

—— REFRENCES ——

https://www.nytimes.com/2020/03/20/world/europe/world-happiness-report.html