

## Initial Results and Code – CKME 136

NAME:	Sung Mo Cheong
Student #:	501001285
Email Address:	sung.cheong@ryerson.ca
Repository	<a href="https://github.com/scheong5/data_analysis.git">https://github.com/scheong5/data_analysis.git</a>
Link to the dataset	<a href="https://www.kaggle.com/drgilermo/home-advantage-in-soccer-and-basketball">https://www.kaggle.com/drgilermo/home-advantage-in-soccer-and-basketball</a>

### Code\_Home-Advantage-In-Sports.R

sung.cheong

2020-07-07

#### *#Load the three datasets*

```
soccer <- read.csv("C:/Users/sung.cheong/Desktop/SoccerLEagues.csv")
NBA <- read.csv("C:/Users/sung.cheong/Desktop/NBA.csv")
Country_fact <- read.csv("C:/Users/sung.cheong/Desktop/Country_facts.csv")
```

#### *#View the top / bottom of all dataset*

```
head (soccer, n = 5)
```

```
##      Sport Country      League Year Season      Team Games
HomeWins
## 1 Football Algeria Algeria-Ligue-1 2010      -      MC Oran      34
10
## 2 Football Algeria Algeria-Ligue-1 2010      -      NA Hussein Dey      34
3
## 3 Football Algeria Algeria-Ligue-1 2011      -      MC El Eulma      29
9
## 4 Football Algeria Algeria-Ligue-1 2011      -      Annaba      29
10
## 5 Football Algeria Algeria-Ligue-1 2012      -      Khroub      30
7
##      HomeDraw HomeLoss HomeRatio AwayWins AwayDraw AwayLoss AwayGoalsDiff X
## 1          5         2         11         0         6         11         -20 NA
## 2          5         9         -8         0         5         12         -19 NA
## 3          3         2         13         0         6         9         -19 NA
## 4          5         0         14         0         1         13         -23 NA
## 5          5         3         0         0         5         10         -23 NA
```

```
head (NBA, n = 5)
```

```
##      Year      Team Wins Loss Home.Wins Home.Loss Away.Wins Away.Loss
## 1 1968 Atlanta Hawks  48  34      28      12      18      21
## 2 1969 Atlanta Hawks  48  34      25      13      18      16
## 3 1970 Atlanta Hawks  36  46      21      20      14      26
## 4 1971 Atlanta Hawks  36  46      22      19      13      26
```

```
## 5 1972 Atlanta Hawks 46 36 28 13 17 23
## HomePCT AwayPCT TotalPCT
## 1 0.7000000 0.4615385 0.5853659
## 2 0.6578947 0.5294118 0.5853659
## 3 0.5121951 0.3500000 0.4390244
## 4 0.5365854 0.3333333 0.4390244
## 5 0.6829268 0.4250000 0.5609756
```

**head** (Country\_fact, n = 5)

```
## Country Region Population Area
PopDensity
## 1 Nigeria SUB-SAHARAN AFRICA 131859731 923768
142.7
## 2 Guatemala LATIN AMER. & CARIB 12293545 108890
112.9
## 3 Indonesia ASIA (EX. NEAR EAST) 245452739 1919440
127.9
## 4 Ghana SUB-SAHARAN AFRICA 22409572 239460
93.6
## 5 Uzbekistan C.W. OF IND. STATES 27307134 447400
61.0
## Coastline Net.migration Infant_mortality GDP Literacy Phones Arable
Crops
## 1 0.09 0.26 98.80 900 68.0 9.3 31.29
2.96
## 2 0.37 -1.67 35.93 4100 70.6 92.1 12.54
5.03
## 3 2.85 0.00 35.60 3200 87.9 52.0 11.32
7.23
## 4 0.23 -0.64 51.43 2200 74.8 14.4 16.26
9.67
## 5 0.00 -1.72 71.10 1700 99.3 62.9 10.83
0.83
## Other Climate Birthrate Deathrate Agriculture Industry Service FIFA_Rank
## 1 65.75 1.5 40.43 16.94 0.269 0.487 0.244 62
## 2 82.43 2.0 29.88 5.20 0.227 0.188 0.585 95
## 3 81.45 2.0 20.34 6.25 0.134 0.458 0.408 178
## 4 74.07 2.0 30.52 9.72 0.366 0.246 0.387 41
## 5 88.34 1.0 26.36 7.84 0.342 0.229 0.430 74
## UEFA_Rank Attendance Home_Away_Contrast
## 1 NA NA 0.8366959
## 2 NA NA 0.6416290
## 3 NA 9368 0.5861931
## 4 NA NA 0.5580267
## 5 NA 6911 0.4860084
```

**tail** (soccer, n = 5)

```
##      Sport  Country          League Year  Season
Team
## 8361 Football Zimbabwe Zimbabwe-Premier-Soccer-League 2011 apertura
Gunners
## 8362 Football Zimbabwe Zimbabwe-Premier-Soccer-League 2010 apertura
Gunners
## 8363 Football Zimbabwe Zimbabwe-Premier-Soccer-League 2010 apertura CAPS
United
## 8364 Football Zimbabwe Zimbabwe-Premier-Soccer-League 2009 apertura
Eagles
## 8365 Football Zimbabwe Zimbabwe-Premier-Soccer-League 2015 apertura
Tsholotsho
##      Games HomeWins HomeDraw HomeLoss HomeRatio AwayWins AwayDraw AwayLoss
## 8361    23      4      5      4      -1      4      1      5
## 8362    30      6      6      3      8      8      5      2
## 8363    30      5      6      4      7      8      3      4
## 8364    30      3     10      2      1      5      4      6
## 8365    29      2      9      4     -4      4      4      6
##      AwayGoalsDiff  X
## 8361              0 NA
## 8362             11 NA
## 8363              6 NA
## 8364             -9 NA
## 8365             -4 NA
```

**tail** (NBA, n = 5)

```
##      Year          Team Wins Loss Home.Wins Home.Loss Away.Wins
Away.Loss
## 1212 2006 Washington Wizards  41  41      26      15      15
26
## 1213 2007 Washington Wizards  43  39      25      16      18
23
## 1214 2008 Washington Wizards  19  63      13      28      6
35
## 1215 2009 Washington Wizards  26  56      15      26      11
30
## 1216 2010 Washington Wizards  23  59      20      21      3
38
##      HomePCT  AwayPCT  TotalPCT
## 1212 0.6341463 0.36585366 0.5000000
## 1213 0.6097561 0.43902439 0.5243902
## 1214 0.3170732 0.14634146 0.2317073
## 1215 0.3658537 0.26829268 0.3170732
## 1216 0.4878049 0.07317073 0.2804878
```

**tail** (Country\_fact, n = 5)

```
##      Country          Region Population  Area
## 84 Puerto Rico  LATIN AMER. & CARIB  3927188 13790
```

```

## 85      Qatar  NEAR EAST                885359    11437
## 86      Mali  SUB-SAHARAN AFRICA        11716829 1240000
## 87    Scotland                5000000    77900
## 88    Lebanon  NEAR EAST                3874050    10400
##      PopDensity Coastline Net.migration Infant_mortality    GDP Literacy
Phones
## 84      284.8      3.63      -1.46      8.24 16800      94.1
283.1
## 85      77.4      4.92      16.29      18.61 21500      82.5
232.0
## 86      9.5      0.00      -0.33      116.79    900      46.4
6.4
## 87      NA      NA      NA      NA    NA    NA    NA
NA
## 88      372.5      2.16      0.00      24.52 4800      87.4
255.6
##      Arable Crops Other Climate Birthrate Deathrate Agriculture Industry
Service
## 84    3.95    5.52 90.53      2    12.77      7.65      0.010    0.450
0.540
## 85    1.64    0.27 98.09      1    15.56      4.72      0.002    0.801
0.197
## 86    3.82    0.03 96.15      2    49.82      16.89      0.450    0.170
0.380
## 87      NA      NA      NA      NA      NA      NA      NA      NA
NA
## 88   16.62  13.98 69.40      NA    18.52      6.21      0.120    0.210
0.670
##      FIFA_Rank UEFA_Rank Attendance Home_Away_Contrast
## 84          7      NA      NA      0.060580596
## 85         80      NA    5073      0.046824819
## 86         73      NA      NA      0.036594879
## 87         45      25      NA     -0.004060382
## 88        145      NA      NA     -0.011744120

```

*#View its class*

```
class(soccer)
```

```
## [1] "data.frame"
```

```
class(NBA)
```

```
## [1] "data.frame"
```

```
class(Country_fact)
```

```
## [1] "data.frame"
```

*#View its dimensions (It shows row and column respectively)*

```
dim(soccer)
```

```
## [1] 8365 16

dim(NBA)

## [1] 1216 11

dim(Country_fact)

## [1] 88 24

#Look at the column names
names(soccer)

## [1] "Sport"          "Country"         "League"          "Year"
## [5] "Season"         "Team"            "Games"           "HomeWins"
## [9] "HomeDraw"       "HomeLoss"        "HomeRatio"       "AwayWins"
## [13] "AwayDraw"       "AwayLoss"        "AwayGoalsDiff"  "X"

names(NBA)

## [1] "Year"          "Team"            "Wins"            "Loss"            "Home.Wins"
## [5] "Home.Loss"
## [7] "Away.Wins"     "Away.Loss"       "HomePCT"         "AwayPCT"         "TotalPCT"

names(Country_fact)

## [1] "Country"          "Region"          "Population"
## [4] "Area"             "PopDensity"      "Coastline"
## [7] "Net.migration"    "Infant_mortality" "GDP"
## [10] "Literacy"         "Phones"          "Arable"
## [13] "Crops"            "Other"           "Climate"
## [16] "Birthrate"        "Deathrate"       "Agriculture"
## [19] "Industry"         "Service"         "FIFA_Rank"
## [22] "UEFA_Rank"        "Attendance"       "Home_Away_Contrast"

#View the structures of the dataset
str(soccer)

## 'data.frame': 8365 obs. of 16 variables:
## $ Sport : Factor w/ 1 level "Football": 1 1 1 1 1 1 1 1 1 1 ...
## $ Country : Factor w/ 89 levels "Algeria","Angola",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ League : Factor w/ 91 levels "Algeria-Ligue-1",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Year : int 2010 2010 2011 2011 2012 2012 2012 2014 2016 2016 ...
## $ Season : Factor w/ 3 levels "-","apertura",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Team : Factor w/ 1975 levels " de Febrero",...: 1145 1221 1144 165 1007 1221 1539 27 1852 209 ...
## $ Games : int 34 34 29 29 30 29 30 29 22 22 ...
## $ HomeWins : int 10 3 9 10 7 5 6 5 5 5 ...
```

```
## $ HomeDraw      : int  5 5 3 5 5 5 6 1 5 2 ...
## $ HomeLoss      : int  2 9 2 0 3 4 3 8 1 4 ...
## $ HomeRatio     : int 11 -8 13 14 0 5 10 -7 4 1 ...
## $ AwayWins      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ AwayDraw      : int  6 5 6 1 5 5 1 4 6 1 ...
## $ AwayLoss      : int 11 12 9 13 10 10 14 11 5 10 ...
## $ AwayGoalsDiff: int -20 -19 -19 -23 -23 -15 -27 -13 -12 -17 ...
## $ X              : num  NA NA NA NA NA NA NA NA NA NA ...
```

`str(NBA)`

```
## 'data.frame': 1216 obs. of 11 variables:
## $ Year      : int 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 ...
## $ Team      : Factor w/ 56 levels "Atlanta Hawks",...: 1 1 1 1 1 1 1 1 1 1
...
## $ Wins      : int 48 48 36 36 46 35 31 29 31 41 ...
## $ Loss      : int 34 34 46 46 36 47 51 53 51 41 ...
## $ Home.Wins : int 28 25 21 22 28 23 22 20 19 29 ...
## $ Home.Loss : int 12 13 20 19 13 18 19 21 22 12 ...
## $ Away.Wins : int 18 18 14 13 17 12 9 9 12 12 ...
## $ Away.Loss : int 21 16 26 26 23 25 32 32 29 29 ...
## $ HomePCT   : num 0.7 0.658 0.512 0.537 0.683 ...
## $ AwayPCT   : num 0.462 0.529 0.35 0.333 0.425 ...
## $ TotalPCT  : num 0.585 0.585 0.439 0.439 0.561 ...
```

`str(Country_fact)`

```
## 'data.frame': 88 obs. of 24 variables:
## $ Country      : Factor w/ 88 levels "Angola ","Argentina ",...: 55
32 36 30 84 8 60 58 79 15 ...
## $ Region       : Factor w/ 12 levels "", "ASIA (EX. NEAR EAST)
",...: 11 6 2 11 4 6 6 10 8 6 ...
## $ Population   : int 131859731 12293545 245452739 22409572 27307134
8989046 28302603 5670544 10175014 43593035 ...
## $ Area         : int 923768 108890 1919440 239460 447400 1098580
1285220 462840 163610 1138910 ...
## $ PopDensity   : num 142.7 112.9 127.9 93.6 61 ...
## $ Coastline    : num 0.09 0.37 2.85 0.23 0 0 0.19 1.11 0.7 0.28 ...
## $ Net.migration : num 0.26 -1.67 0 -0.64 -1.72 -1.32 -1.05 0 -0.57 -
0.31 ...
## $ Infant_mortality : num 98.8 35.9 35.6 51.4 71.1 ...
## $ GDP          : int 900 4100 3200 2200 1700 2400 5100 2200 6900
6300 ...
## $ Literacy     : num 68 70.6 87.9 74.8 99.3 87.2 90.9 64.6 74.2
92.5 ...
## $ Phones       : num 9.3 92.1 52 14.4 62.9 ...
## $ Arable       : num 31.3 12.5 11.3 16.3 10.8 ...
## $ Crops        : num 2.96 5.03 7.23 9.67 0.83 ...
## $ Other        : num 65.8 82.4 81.5 74.1 88.3 ...
## $ Climate      : num 1.5 2 2 2 1 1.5 1.5 2 3 2 ...
```

```
## $ Birthrate      : num  40.4 29.9 20.3 30.5 26.4 ...
## $ Deathrate      : num  16.94 5.2 6.25 9.72 7.84 ...
## $ Agriculture     : num  0.269 0.227 0.134 0.366 0.342 0.128 0.08 0.353
0.132 0.125 ...
## $ Industry        : num  0.487 0.188 0.458 0.246 0.229 0.352 0.27 0.381
0.318 0.342 ...
## $ Service         : num  0.244 0.585 0.408 0.387 0.43 0.52 0.65 0.266
0.55 0.533 ...
## $ FIFA_Rank       : int   62 95 178 41 74 72 42 55 47 8 ...
## $ UEFA_Rank       : int   NA NA NA NA NA NA NA NA NA NA ...
## $ Attendance      : int   NA NA 9368 NA 6911 6148 5178 NA NA 8594 ...
## $ Home_Away_Contrast: num  0.837 0.642 0.586 0.558 0.486 ...
```

*#View the summary of the dataset*  
summary(soccer)

```
##      Sport      Country      League
## Football:8365 Mexico : 216 Mexico-Liga-MX : 216
##      United : 202 Russia-Premier-Liga : 192
##      South : 196 Guatemala-Liga-Nacional : 156
##      Russia : 192 Paraguay-Division-Profesional: 156
##      Guatemala: 156 Japan-J.-League : 144
##      Paraguay : 156 Moldova-National-Division : 144
##      (Other) :7247 (Other) :7357
##      Year      Season      Team      Games
## Min. :2009 - :2179 Olimpia : 26 Min. : 2.00
## 1st Qu.:2011 apertura:3608 Victoria : 14 1st Qu.:22.00
## Median :2012 clausura:2578 ?guila : 13 Median :30.00
## Mean :2012 Alianza FC : 13 Mean :26.99
## 3rd Qu.:2014 Atl?tico Marte: 13 3rd Qu.:34.00
## Max. :2016 CD FAS : 13 Max. :57.00
##      (Other) :8273
##      HomeWins      HomeDraw      HomeLoss      HomeRatio
## Min. : 0.00 Min. : 0.000 Min. : 0.000 Min. : -83.000
## 1st Qu.: 4.00 1st Qu.: 2.000 1st Qu.: 2.000 1st Qu.: -1.000
## Median : 6.00 Median : 3.000 Median : 3.000 Median : 4.000
## Mean : 6.23 Mean : 3.539 Mean : 3.693 Mean : 5.218
## 3rd Qu.: 8.00 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.: 11.000
## Max. :19.00 Max. :12.000 Max. :17.000 Max. : 62.000
##
##      AwayWins      AwayDraw      AwayLoss      AwayGoalsDiff
## Min. : 0.000 Min. : 0.000 Min. : 0.000 Min. : -98.000
## 1st Qu.: 2.000 1st Qu.: 2.000 1st Qu.: 4.000 1st Qu.: -12.000
## Median : 3.000 Median : 3.000 Median : 6.000 Median : -5.000
## Mean : 3.738 Mean : 3.583 Mean : 6.333 Mean : -5.336
## 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.: 9.000 3rd Qu.: 1.000
## Max. :16.000 Max. :12.000 Max. :19.000 Max. : 50.000
##
##      X
```

```
## Min. :0.307
## 1st Qu.:0.307
## Median :0.307
## Mean :0.307
## 3rd Qu.:0.307
## Max. :0.307
## NA's :8364
```

**summary(NBA)**

```
##      Year      Team      Wins      Loss
## Min. :1950 Boston Celtics : 61 Min. : 8.00 Min. :10.00
## 1st Qu.:1976 NewYork Knicks : 61 1st Qu.:31.00 1st Qu.:31.00
## Median :1989 Detroit Pistons : 54 Median :41.00 Median :40.00
## Mean :1987 LosAngeles Lakers : 51 Mean :40.18 Mean :40.18
## 3rd Qu.:2000 Philadelphia 76ers: 48 3rd Qu.:49.00 3rd Qu.:49.00
## Max. :2010 Chicago Bulls : 45 Max. :72.00 Max. :73.00
##      (Other)      :896
##      Home.Wins      Home.Loss      Away.Wins      Away.Loss
## Min. : 2.0 Min. : 1.00 Min. : 0.00 Min. : 7.00
## 1st Qu.:19.0 1st Qu.:10.00 1st Qu.:10.00 1st Qu.:20.00
## Median :25.0 Median :14.00 Median :14.00 Median :24.00
## Mean :24.4 Mean :14.67 Mean :14.67 Mean :24.41
## 3rd Qu.:30.0 3rd Qu.:19.00 3rd Qu.:19.00 3rd Qu.:29.00
## Max. :40.0 Max. :35.00 Max. :33.00 Max. :40.00
##
##      HomePCT      AwayPCT      TotalPCT
## Min. :0.07143 Min. :0.0000 Min. :0.1098
## 1st Qu.:0.51220 1st Qu.:0.2439 1st Qu.:0.3902
## Median :0.63415 Median :0.3659 Median :0.5122
## Mean :0.62477 Mean :0.3733 Mean :0.4999
## 3rd Qu.:0.75610 3rd Qu.:0.4878 3rd Qu.:0.6098
## Max. :0.97561 Max. :0.8158 Max. :0.8780
##
```

**summary(Country\_fact)**

```
##      Country      Region      Population
## Angola : 1 LATIN AMER. & CARIB :18 Min.
:7.843e+05
## Argentina : 1 WESTERN EUROPE :15 1st
Qu.:5.387e+06
## Australia : 1 EASTERN EUROPE :11 Median
:1.110e+07
## Austria : 1 SUB-SAHARAN AFRICA :10 Mean
:5.872e+07
## Azerbaijan : 1 C.W. OF IND. STATES : 8 3rd
Qu.:4.061e+07
## Belarus : 1 ASIA (EX. NEAR EAST) : 6 Max.
:1.314e+09
```



```

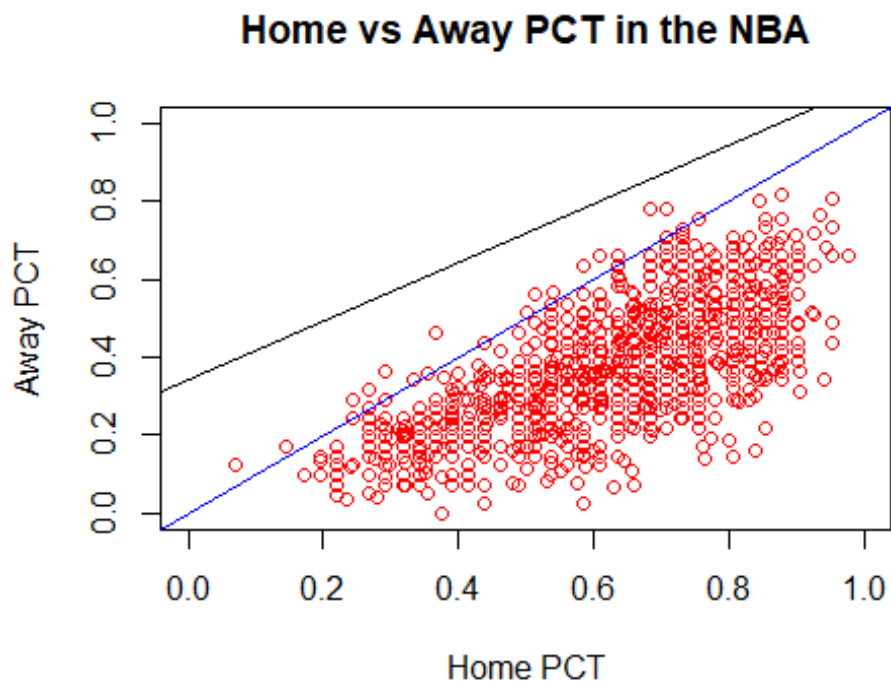
## (Other) :82 (Other) :20
## Area PopDensity Coastline Net.migration
## Min. : 5128 Min. : 2.60 Min. : 0.0000 Min. : -10.83000
## 1st Qu.: 70135 1st Qu.: 27.82 1st Qu.: 0.0875 1st Qu.: -0.92750
## Median : 292395 Median : 70.65 Median : 0.3150 Median : 0.00000
## Mean : 1106413 Mean : 93.73 Mean : 1.5040 Mean : -0.03143
## 3rd Qu.: 914980 3rd Qu.:113.38 3rd Qu.: 1.0075 3rd Qu.: 0.97500
## Max. :17075200 Max. :397.10 Max. :16.9700 Max. : 16.29000
## NA's :4 NA's :4 NA's :4
## Infant_mortality GDP Literacy Phones
## Min. : 2.770 Min. : 700 Min. : 42.00 Min. : 0.2
## 1st Qu.: 6.755 1st Qu.: 3850 1st Qu.: 86.40 1st Qu.:108.0
## Median : 20.405 Median : 7500 Median : 96.20 Median :225.3
## Mean : 27.520 Mean :11767 Mean : 89.27 Mean :258.6
## 3rd Qu.: 33.343 3rd Qu.:19050 3rd Qu.: 99.00 3rd Qu.:402.2
## Max. :191.190 Max. :37800 Max. :100.00 Max. :898.0
## NA's :4 NA's :4 NA's :7 NA's :5
## Arable Crops Other Climate
## Min. : 0.460 Min. : 0.000 Min. :33.91 Min. :1.000
## 1st Qu.: 7.162 1st Qu.: 0.415 1st Qu.:69.15 1st Qu.:2.000
## Median :13.295 Median : 1.415 Median :83.47 Median :3.000
## Mean :18.000 Mean : 2.717 Mean :79.28 Mean :2.426
## 3rd Qu.:26.980 3rd Qu.: 3.205 3rd Qu.:92.16 3rd Qu.:3.000
## Max. :56.210 Max. :13.980 Max. :98.24 Max. :4.000
## NA's :4 NA's :4 NA's :4 NA's :14
## Birthrate Deathrate Agriculture Industry
## Min. : 8.25 Min. : 2.580 Min. :0.00200 Min. :0.1100
## 1st Qu.:10.39 1st Qu.: 6.220 1st Qu.:0.03475 1st Qu.:0.2587
## Median :14.45 Median : 9.050 Median :0.07500 Median :0.3040
## Mean :17.81 Mean : 9.447 Mean :0.11149 Mean :0.3251
## 3rd Qu.:22.00 3rd Qu.:10.580 3rd Qu.:0.14125 3rd Qu.:0.3695
## Max. :49.82 Max. :24.200 Max. :0.55000 Max. :0.8010
## NA's :5 NA's :5 NA's :4 NA's :4
## Service FIFA_Rank UEFA_Rank Attendance
## Min. :0.1970 Min. : 1.00 Min. : 1.00 Min. : 316
## 1st Qu.:0.4898 1st Qu.:18.75 1st Qu.:10.25 1st Qu.: 3132
## Median :0.5740 Median : 53.00 Median :20.50 Median : 6323
## Mean :0.5634 Mean : 62.03 Mean :21.26 Mean : 9728
## 3rd Qu.:0.6703 3rd Qu.: 82.75 3rd Qu.:29.75 3rd Qu.:11070
## Max. :0.7870 Max. :204.00 Max. :51.00 Max. :43534
## NA's :4 NA's :50 NA's :36
## Home_Away_Contrast
## Min. : -0.01174
## 1st Qu.: 0.16049
## Median : 0.26528
## Mean : 0.26068
## 3rd Qu.: 0.33081
## Max. : 0.83670
##

```

```
##### NBA #####
#plot of two variables
plot(NBA$HomePCT, NBA$AwayPCT, main= "Home vs Away PCT in the NBA", xlab =
"Home PCT", ylab = "Away PCT", col = "Red", xlim = c(0,1),ylim = c(0,1))
lm(NBA$HomePCT~NBA$AwayPCT)

##
## Call:
## lm(formula = NBA$HomePCT ~ NBA$AwayPCT)
##
## Coefficients:
## (Intercept)  NBA$AwayPCT
##      0.3442      0.7516

abline(a=0.000, b=1, col = "Blue")
abline (a =0.3442, b=0.7516, col = 'Black')
```



```
#the dataset with linear regression and symmetric diagonal line to see the
comparison.
#Clearly, we can see that home win percentage is much higher than away win
percentage.
#Approximately, around 13-16 games won in away game among 1216 NBA games.

#updated dataset
soccer1 <- read.csv("C:/Users/sung.cheong/Desktop/SoccerLEagues1.csv")
NBA1 <- read.csv("C:/Users/sung.cheong/Desktop/NBA1.csv")
```

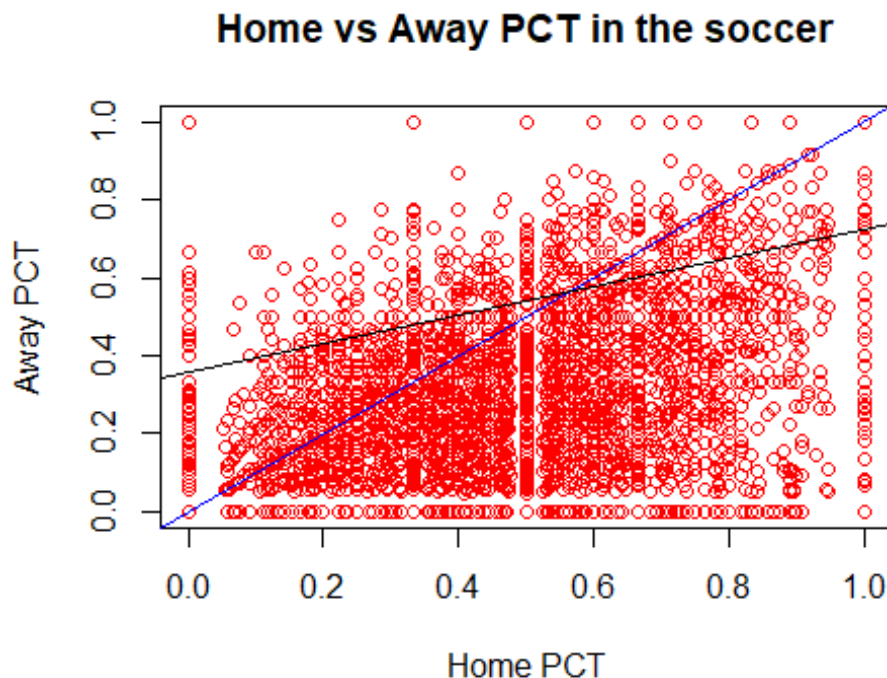
```
##### Soccer #####
names(soccer1)

## [1] "Sport"      "Country"    "League"     "Year"       "Team"       "Games"
## [7] "HomeWins"   "HomeDraw"   "HomeLoss"    "AwayWins"    "AwayDraw"    "AwayLoss"
## [13] "X"          "Home.PCT"   "Away.PCT"

plot(soccer1$Home.PCT, soccer1$Away.PCT, main= "Home vs Away PCT in the
soccer", xlab = "Home PCT", ylab = "Away PCT", col = "Red", xlim =
c(0,1),ylim = c(0,1))
abline(a=0.000, b=1, col = "Blue")
lm(soccer1$Home.PCT~soccer1$Away.PCT)

##
## Call:
## lm(formula = soccer1$Home.PCT ~ soccer1$Away.PCT)
##
## Coefficients:
##      (Intercept)  soccer1$Away.PCT
##           0.3599           0.3633

abline(a=0.3599, b=0.3633, col = "Black")
```



*#data that under the both lines are more than above the lines.  
 #Not as much clear as NBA, however home win percentage is more crowded than  
 away win percentage.*

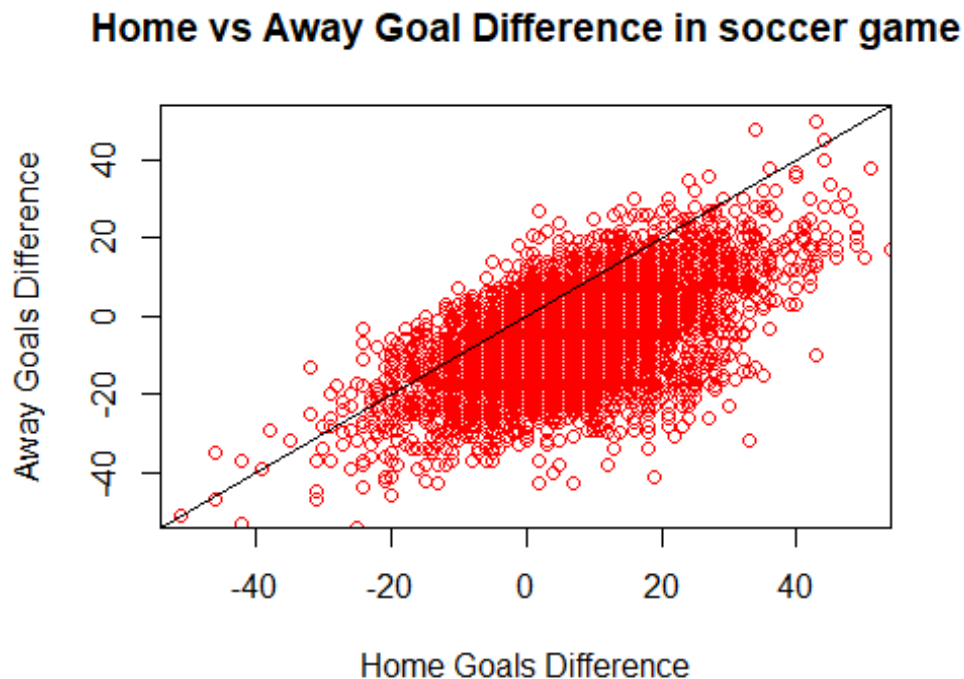
*#Not sufficient to conclude that there is home advantage in soccer yet.*

*#according to the soccer dataset, we could say that the teams are scoring more goals at home than away games.*

```
names(soccer)
```

```
## [1] "Sport"      "Country"    "League"     "Year"
## [5] "Season"     "Team"       "Games"      "HomeWins"
## [9] "HomeDraw"   "HomeLoss"   "HomeRatio"   "AwayWins"
## [13] "AwayDraw"   "AwayLoss"   "AwayGoalsDiff" "X"
```

```
plot(soccer$HomeRatio, soccer$AwayGoalsDiff, main= "Home vs Away Goal  
Difference in soccer game", xlab = "Home Goals Difference", ylab = "Away  
Goals Difference", col = "Red", xlim = c(-50,50),ylim = c(-50,50))  
abline(a=0, b=1, col = "Black")
```



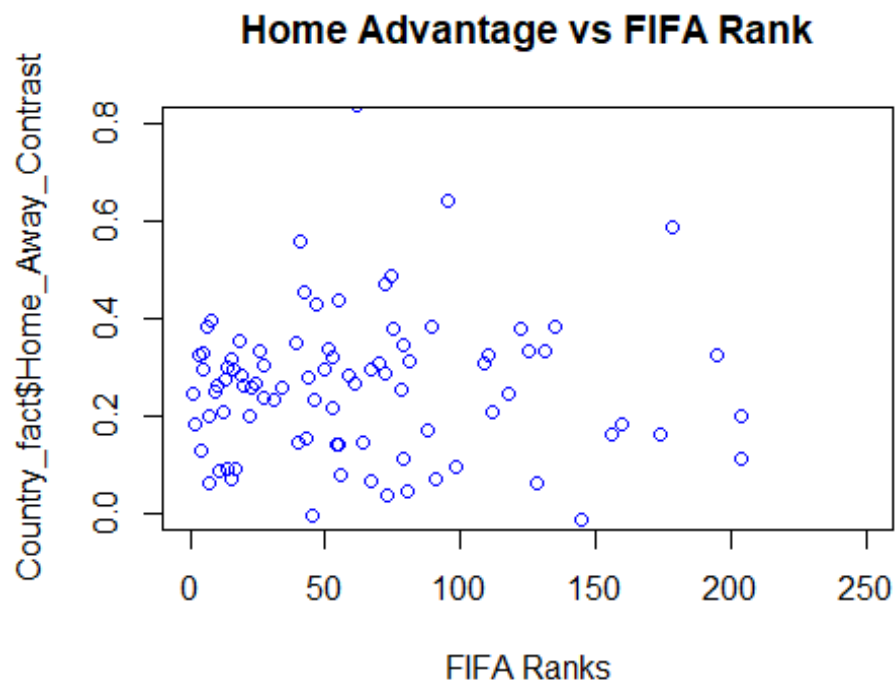
*#Let's see other factors from Country\_fact dataset*

```
names(Country_fact)
```

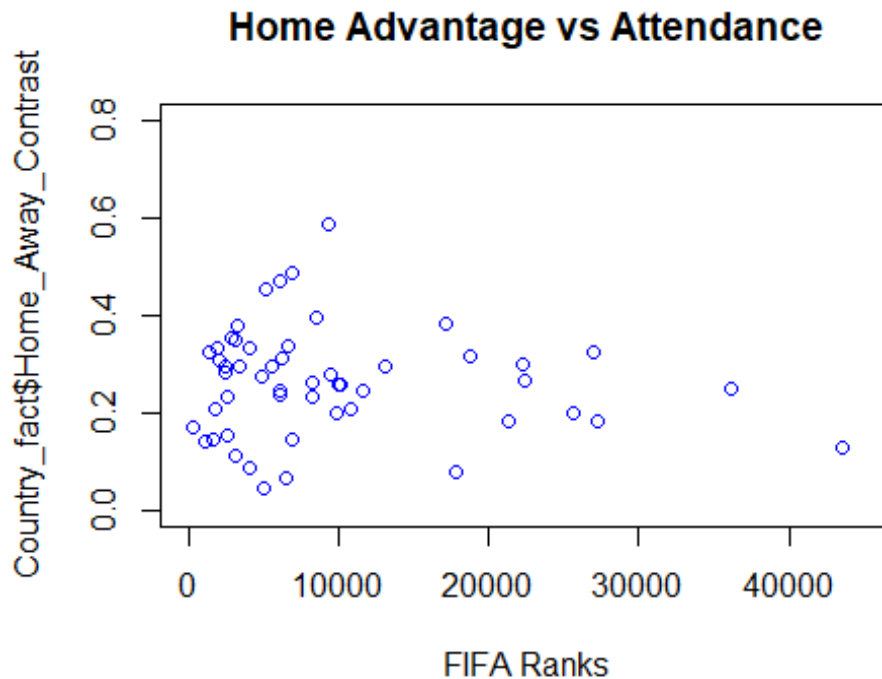
```
## [1] "Country"      "Region"     "Population"
## [4] "Area"         "PopDensity" "Coastline"
## [7] "Net.migration" "Infant_mortality" "GDP"
## [10] "Literacy"     "Phones"     "Arable"
## [13] "Crops"        "Other"      "Climate"
## [16] "Birthrate"    "Deathrate"  "Agriculture"
```

```
## [19] "Industry"          "Service"          "FIFA_Rank"
## [22] "UEFA_Rank"         "Attendance"        "Home_Away_Contrast"

plot(Country_fact$FIFA_Rank, Country_fact$Home_Away_Contrast, main= "Home
Advantage vs FIFA Rank", xlab = "FIFA Ranks", col = "blue", xlim =
c(0,250),ylim = c(0,0.8))
```



```
plot(Country_fact$Attendance, Country_fact$Home_Away_Contrast, main= "Home
Advantage vs Attendance", xlab = "FIFA Ranks", col = "blue", xlim =
c(0,45000),ylim = c(0,0.8))
```



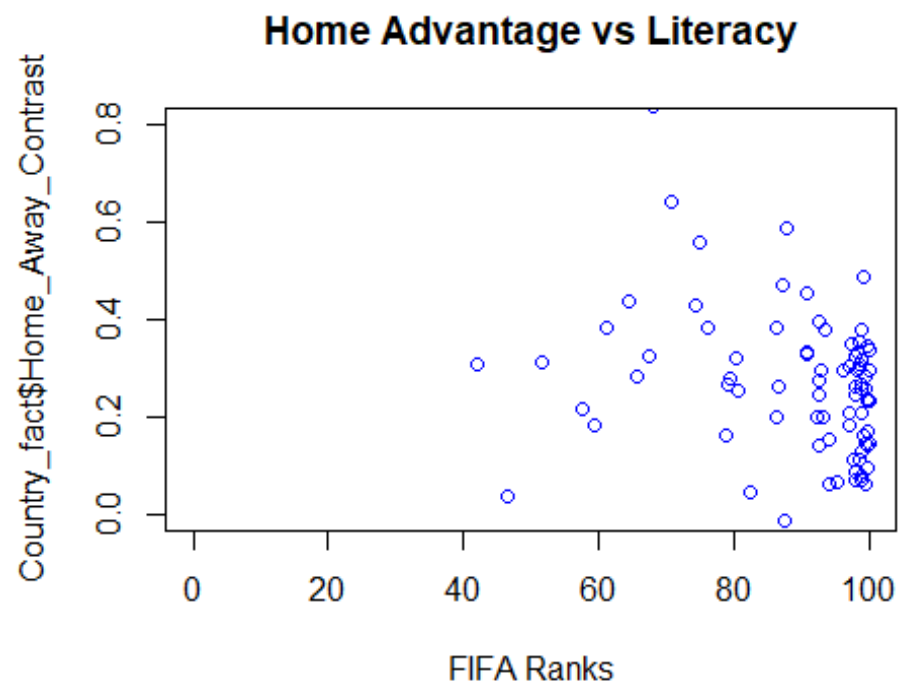
```
reg <- lm(Country_fact$Attendance~Country_fact$Home_Away_Contrast)
#not much changes in home advantage over average attendance. Actually it
decreased. FIFA Rank also does not effect the home advantage that much
according to the data that we plot.
```

```
#Home Advantages vs Literacy rate
#Home Advantages vs Infant mortality
#Home Advantages vs Industry
#Home Advantages vs # of phone
```

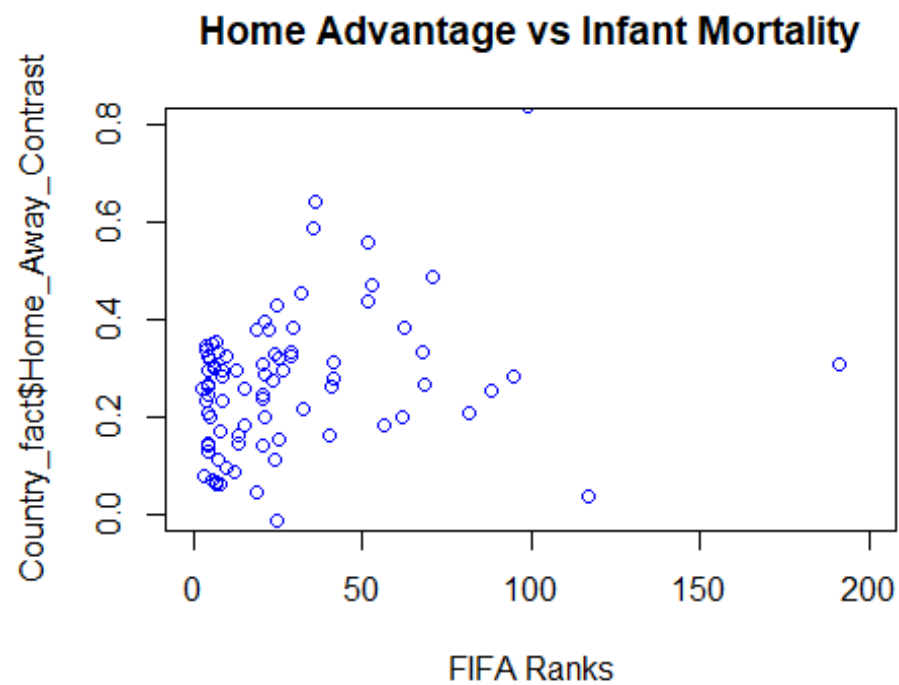
```
names (Country_fact)
```

```
## [1] "Country"          "Region"           "Population"
## [4] "Area"             "PopDensity"       "Coastline"
## [7] "Net.migration"    "Infant_mortality" "GDP"
## [10] "Literacy"         "Phones"           "Arable"
## [13] "Crops"            "Other"            "Climate"
## [16] "Birthrate"        "Deathrate"        "Agriculture"
## [19] "Industry"         "Service"          "FIFA_Rank"
## [22] "UEFA_Rank"        "Attendance"       "Home_Away_Contrast"
```

```
plot(Country_fact$Literacy, Country_fact$Home_Away_Contrast, main= "Home
Advantage vs Literacy", xlab = "FIFA Ranks", col = "blue", xlim =
c(0,100),ylim = c(0,0.8))
```

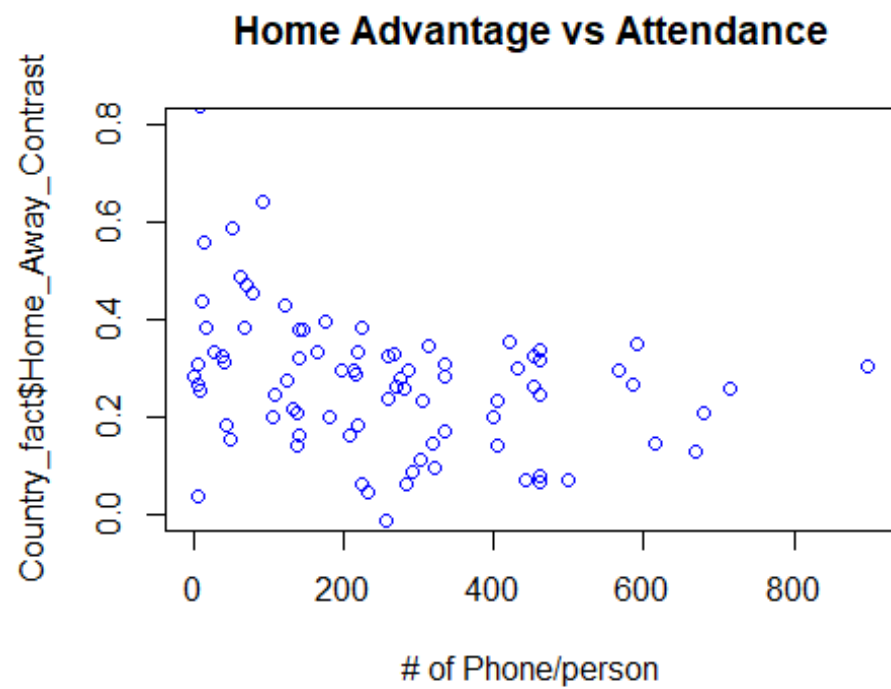


```
plot(Country_fact$Infant_mortality, Country_fact$Home_Away_Contrast, main="Home Advantage vs Infant Mortality", xlab = "FIFA Ranks", col = "blue", xlim = c(0,200),ylim = c(0,0.8))
```

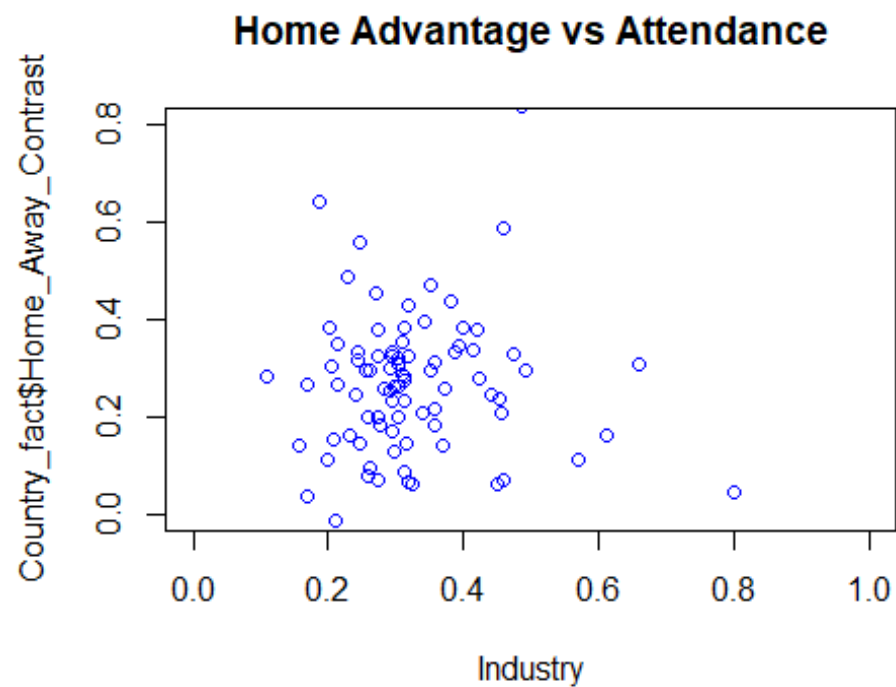


```
plot(Country_fact$Phones, Country_fact$Home_Away_Contrast, main= "Home
Advantage vs Attendance", xlab = "# of Phone/person", col = "blue", xlim =
c(0,900),ylim = c(0,0.8))
```

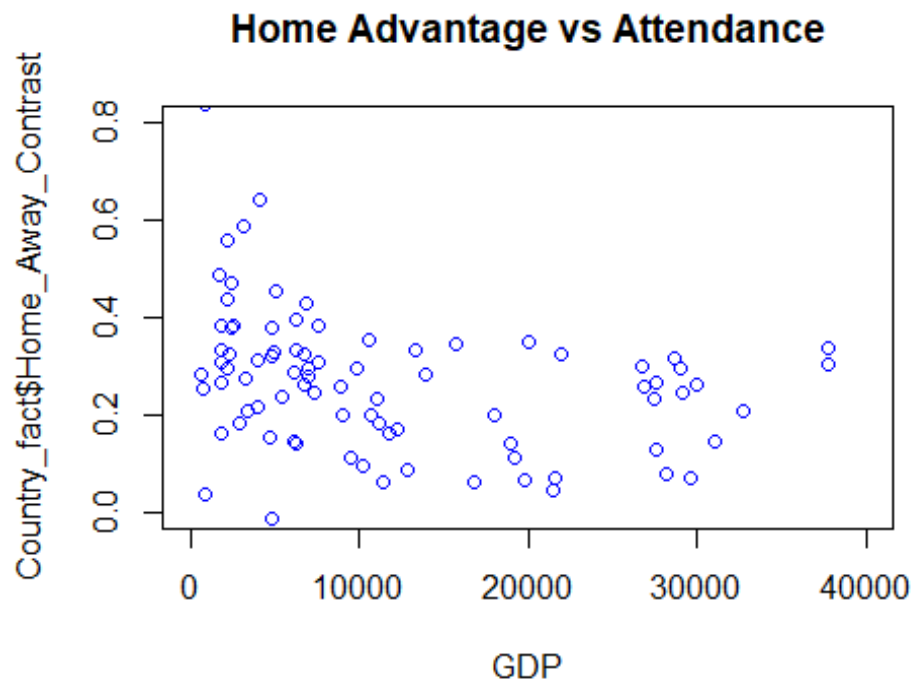




```
plot(Country_fact$Industry, Country_fact$Home_Away_Contrast, main= "Home  
Advantage vs Attendance", xlab = "Industry", col = "blue", xlim = c(0,1),ylim  
= c(0,0.8))
```



```
plot(Country_fact$GDP, Country_fact$Home_Away_Contrast, main= "Home Advantage  
vs Attendance", xlab = "GDP", col = "blue", xlim = c(0,40000),ylim =  
c(0,0.8))
```



*#Interestingly enough, it seems that the strongest predictors are indicators of general development such as # of phones/person literacy, GDP, Industry and Mortality rate for infant.*

*#The more developed the country, it shows less significant the home advantage is (in soccer).*