Who's Gonna Win The English Premier League?

Pythagorean Expectation

Retrieve The Data (The 20-21 Season)

```
library(worldfootballR)

EPL_Season <- get_season_team_stats(country="ENG", gender="M", season_end_year="2021", tier="1st", stat_

EPL_Season
```

```
##    Competition_Name Gender Country Season_End_Year          Squad
## 1   Premier League      M     ENG            2021         Arsenal
## 2   Premier League      M     ENG            2021     Aston Villa
## 3   Premier League      M     ENG            2021        Brighton
## 4   Premier League      M     ENG            2021         Burnley
## 5   Premier League      M     ENG            2021         Chelsea
## 6   Premier League      M     ENG            2021   Crystal Palace
## 7   Premier League      M     ENG            2021         Everton
## 8   Premier League      M     ENG            2021          Fulham
## 9   Premier League      M     ENG            2021     Leeds United
## 10  Premier League      M     ENG            2021   Leicester City
## 11  Premier League      M     ENG            2021       Liverpool
## 12  Premier League      M     ENG            2021  Manchester City
## 13  Premier League      M     ENG            2021   Manchester Utd
## 14  Premier League      M     ENG            2021     Newcastle Utd
## 15  Premier League      M     ENG            2021    Sheffield Utd
## 16  Premier League      M     ENG            2021      Southampton
## 17  Premier League      M     ENG            2021        Tottenham
## 18  Premier League      M     ENG            2021        West Brom
## 19  Premier League      M     ENG            2021         West Ham
## 20  Premier League      M     ENG            2021           Wolves
##    Team_or_Opponent Rk MP_Home W_Home D_Home L_Home GF_Home GA_Home GD_Home
## 1              team  8      19      8      4      7      24      21       3
## 2              team 11      19      7      4      8      29      27       2
## 3              team 16      19      4      9      6      22      22       0
## 4              team 17      19      4      6      9      14      27     -13
## 5              team  4      19      9      6      4      31      18      13
## 6              team 14      19      6      5      8      20      32     -12
## 7              team 10      19      6      4      9      24      28      -4
## 8              team 18      19      2      4     13       9      28     -19
## 9              team  9      19      8      5      6      28      21       7
## 10             team  5      19      9      1      9      34      30       4
## 11             team  3      19     10      3      6      29      20       9
```

```
## 12            team  1     19   13    2    4   43   17   26
## 13            team  2     19    9    4    6   38   28   10
## 14            team 12     19    6    5    8   26   33   -7
## 15            team 20     19    5    1   13   12   27  -15
## 16            team 15     19    8    3    8   28   25    3
## 17            team  7     19   10    3    6   35   20   15
## 18            team 19     19    3    6   10   15   39  -24
## 19            team  6     19   10    4    5   32   22   10
## 20            team 13     19    7    4    8   21   25   -4
##    Pts_Home Pts_per_G_Home xG_Home xGA_Home xGD_Home xGD_per_90_Home MP_Away
## 1        28           1.47    24.6     22.3      2.3            0.12      19
## 2        25           1.32    27.7     27.8      0.0            0.00      19
## 3        21           1.11    31.5     14.6     16.9            0.89      19
## 4        18           0.95    20.6     27.6     -7.0           -0.37      19
## 5        33           1.74    35.6     17.2     18.4            0.97      19
## 6        23           1.21    13.8     26.2    -12.4           -0.65      19
## 7        22           1.16    25.3     27.2     -1.9           -0.10      19
## 8        10           0.53    18.8     24.9     -6.1           -0.32      19
## 9        29           1.53    30.6     26.0      4.6            0.24      19
## 10       28           1.47    28.9     26.2      2.7            0.14      19
## 11       33           1.74    35.7     21.7     14.0            0.74      19
## 12       41           2.16    39.7     15.5     24.2            1.27      19
## 13       31           1.63    31.7     21.4     10.3            0.54      19
## 14       23           1.21    24.1     23.1      1.0            0.05      19
## 15       16           0.84    16.9     30.5    -13.6           -0.72      19
## 16       27           1.42    20.8     22.4     -1.6           -0.08      19
## 17       33           1.74    29.2     22.3      6.9            0.36      19
## 18       15           0.79    17.2     36.1    -18.9           -0.99      19
## 19       34           1.79    24.2     25.7     -1.5           -0.08      19
## 20       25           1.32    23.3     20.3      3.0            0.16      19
##    W_Away D_Away L_Away GF_Away GA_Away GD_Away Pts_Away Pts_per_G_Away xG_Away
## 1      10      3      6      31      18      13       33           1.74    28.9
## 2       9      3      7      26      19       7       30           1.58    25.2
## 3       5      5      9      18      24      -6       20           1.05    20.1
## 4       6      3     10      19      28      -9       21           1.11    19.3
## 5      10      4      5      27      18       9       34           1.79    28.4
## 6       6      3     10      21      34     -13       21           1.11    18.6
## 7      11      4      4      23      20       3       37           1.95    21.8
## 8       3      9      7      18      25      -7       18           0.95    22.5
## 9      10      0      9      34      33       1       30           1.58    26.9
## 10     11      5      3      34      20      14       38           2.00    27.1
## 11     10      6      3      39      22      17       36           1.89    36.9
## 12     14      3      2      40      15      25       45           2.37    33.6
## 13     12      7      0      35      16      19       43           2.26    28.6
## 14      6      4      9      20      29      -9       22           1.16    16.9
## 15      2      1     16       8      36     -28        7           0.37    14.5
## 16      4      4     11      19      43     -24       16           0.84    21.6
## 17      8      5      6      33      25       8       29           1.53    25.3
## 18      2      5     12      20      37     -17       11           0.58    16.6
## 19      9      4      6      30      25       5       31           1.63    29.7
## 20      5      5      9      15      27     -12       20           1.05    16.5
##    xGA_Away xGD_Away xGD_per_90_Away
## 1      22.1      6.9            0.36
## 2      25.1      0.1            0.01
```

```
## 3       23.0      -2.9        -0.15
## 4       30.0     -10.7        -0.57
## 5       15.6      12.8         0.67
## 6       31.3     -12.7        -0.67
## 7       24.8      -3.0        -0.16
## 8       28.1      -5.6        -0.29
## 9       36.9     -10.0        -0.53
## 10      21.4       5.7         0.30
## 11      23.6      13.3         0.70
## 12      15.8      17.8         0.93
## 13      20.9       7.7         0.41
## 14      30.9     -14.0        -0.74
## 15      31.9     -17.4        -0.92
## 16      31.8     -10.2        -0.54
## 17      27.3      -1.9        -0.10
## 18      31.7     -15.1        -0.79
## 19      22.6       7.1         0.37
## 20      25.5      -9.0        -0.48
```

Prepare The Data

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
EPL_Season <- EPL_Season %>%
  select(Squad, MP_Home, MP_Away, GF_Home, GF_Away, GA_Home, GA_Away, Pts_Home, Pts_Away)

EPL_Season <- EPL_Season %>%
  mutate(Total_Matches=MP_Home + MP_Away) %>%
  mutate(Total_points=Pts_Home + Pts_Away) %>%
  mutate(Goals_for=GF_Home + GF_Away) %>%
  mutate(Goals_against=GA_Home + GA_Away) %>%
  select(Squad, Total_Matches, Total_points, Goals_for, Goals_against)
```

```
EPL_Season
```

```
##          Squad Total_Matches Total_points Goals_for Goals_against
## 1       Arsenal            38           61        55            39
## 2   Aston Villa            38           55        55            46
## 3      Brighton            38           41        40            46
```

```
## 4           Burnley                38        39        33        55
## 5           Chelsea                38        67        58        36
## 6    Crystal Palace                38        44        41        66
## 7           Everton                38        59        47        48
## 8            Fulham                38        28        27        53
## 9      Leeds United                38        59        62        54
## 10    Leicester City               38        66        68        50
## 11        Liverpool                38        69        68        42
## 12  Manchester City                38        86        83        32
## 13   Manchester Utd                38        74        73        44
## 14    Newcastle Utd                38        45        46        62
## 15    Sheffield Utd                38        23        20        63
## 16       Southampton               38        43        47        68
## 17        Tottenham                38        62        68        45
## 18         West Brom                38        26        35        76
## 19          West Ham                38        65        62        47
## 20            Wolves               38        45        36        52
```

Calculate The Pythagorean Expectation

From Soccer Stats

```
EPL_Season$SocStatsPyt <- (1.7 * (EPL_Season$Goals_for - EPL_Season$Goals_against) / (EPL_Season$Goals_
```

From Penal.ty

```
a <- 1.22777 # exponent of goals for
b <- 1.072388 # exponent for goals for in sum of GF and GA
c <- 1.127248 # exponent for GA
n <- 15 # number of matches played

EPL_Season$PenaltyPytExp <- (EPL_Season$Goals_for ** a / (EPL_Season$Goals_for ** b + EPL_Season$Goals_a
```

From StatsBomb

```
EPL_Season$StatsBombPyt <- 0.677 * (EPL_Season$Goals_for - EPL_Season$Goals_against) + 52.29
```

```
EPL_Season <- EPL_Season %>%
  select(Squad, Total_points, SocStatsPyt, PenaltyPytExp, StatsBombPyt) %>%
  arrange(desc(Total_points))

EPL_Season
```

```
##              Squad Total_points SocStatsPyt PenaltyPytExp StatsBombPyt
## 1  Manchester City           86     79.94870     131.52220       86.817
## 2   Manchester Utd           74     67.31197     107.88310       71.923
## 3        Liverpool           69     66.56909     105.64617       69.892
```
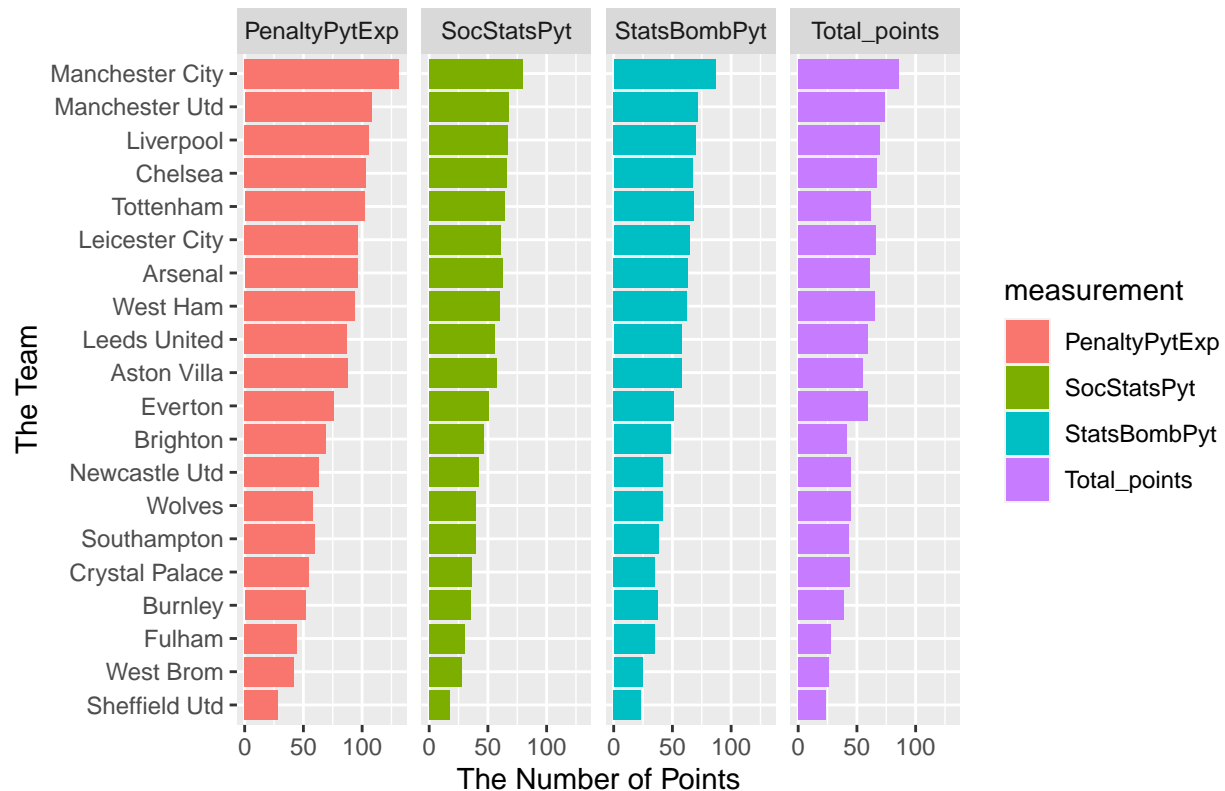
```
## 4          Chelsea              67      66.41915     103.20586       67.184
## 5    Leicester City             66      61.15424      96.76285       64.476
## 6         West Ham              65      60.18991      94.06538       62.445
## 7        Tottenham              62      64.44867     102.15384       67.861
## 8          Arsenal              61      62.29574      95.93903       63.122
## 9          Everton              59      50.62000      76.29461       51.613
## 10    Leeds United             59      55.75517      87.01099       57.706
## 11     Aston Villa             55      57.05644      87.71894       58.383
## 12    Newcastle Utd            45      41.72963      63.14819       41.458
## 13          Wolves             45      39.55455      58.32354       41.458
## 14   Crystal Palace            44      36.20654      54.62733       35.365
## 15      Southampton            43      39.50348      60.14318       38.073
## 16         Brighton            41      46.79302      69.25804       48.228
## 17          Burnley            39      35.15000      51.84565       37.396
## 18           Fulham            28      30.30500      44.49791       34.688
## 19        West Brom            26      27.43874      42.18491       24.533
## 20      Sheffield Utd          23      17.83253      28.56955       23.179
```

Visualize The Result

```
EPL_Season %>%
  pivot_longer(2:(length(colnames(EPL_Season))), names_to='measurement', values_to='total') %>%
  arrange(desc(total)) %>%
  ggplot(aes(x=total, y=reorder(Squad, total, sum), fill=measurement)) +
  geom_bar(position='stack', stat='identity') +
  facet_grid(~measurement) +
  ggtitle("The Number of Points at the End of EPL Season 20-21") +
  xlab("The Number of Points") +
  ylab("The Team")
```

The Number of Points at the End of EPL Season 20–21

Find The Best Equation

```
EPL_Season %>%
  mutate(ptsSoc = (Total_points - SocStatsPyt)^2,
         ptsPen = (Total_points - PenaltyPytExp)^2,
         ptsStats = (Total_points - StatsBombPyt)^2) %>%
  summarise(meanSoc = mean(ptsSoc) ** 0.5, meanPen = mean(ptsPen) ** 0.5, meanStats = mean(ptsStats) **
```

```
##    meanSoc meanPen meanStats
## 1 4.597008 27.4727  4.164511
```

Test it on Current Season of EPL (21-22)

```
EPL_Season <- get_season_team_stats(country="ENG", gender="M", season_end_year="2022", tier="1st", stat_

EPL_Season <- EPL_Season %>%
  select(Squad, MP_Home, MP_Away, GF_Home, GF_Away, GA_Home, GA_Away, Pts_Home, Pts_Away)

EPL_Season <- EPL_Season %>%
  mutate(Total_Matches=MP_Home + MP_Away) %>%
  mutate(Total_points=Pts_Home + Pts_Away) %>%
```

```
  mutate(Goals_for=GF_Home + GF_Away) %>%
  mutate(Goals_against=GA_Home + GA_Away) %>%
  select(Squad, Total_Matches, Total_points, Goals_for, Goals_against)

EPL_Season
```

```
##                Squad Total_Matches Total_points Goals_for Goals_against
## 1           Arsenal            12           20        13            17
## 2       Aston Villa            12           13        16            20
## 3         Brentford            12           13        16            17
## 4          Brighton            12           17        12            14
## 5           Burnley            12            9        14            20
## 6           Chelsea            12           29        30             4
## 7    Crystal Palace            12           16        18            17
## 8           Everton            11           15        16            16
## 9      Leeds United            11           11        11            18
## 10   Leicester City            12           15        16            21
## 11        Liverpool            12           25        35            11
## 12  Manchester City            11           23        22             6
## 13   Manchester Utd            12           17        20            21
## 14     Newcastle Utd           12            6        15            27
## 15     Norwich City            12            8         7            27
## 16      Southampton            12           14        11            14
## 17        Tottenham            11           16         9            16
## 18          Watford            12           13        16            20
## 19         West Ham            12           23        23            14
## 20           Wolves            12           19        12            12
```

```
EPL_Season$StatsBombPyt <- 0.677 * (EPL_Season$Goals_for - EPL_Season$Goals_against) + 52.29

EPL_Season %>%
  select(Squad, Total_points, StatsBombPyt) %>%
  arrange(desc(StatsBombPyt))
```

```
##                Squad Total_points StatsBombPyt
## 1           Chelsea           29       69.892
## 2         Liverpool           25       68.538
## 3   Manchester City           23       63.122
## 4          West Ham           23       58.383
## 5    Crystal Palace           16       52.967
## 6           Everton           15       52.290
## 7            Wolves           19       52.290
## 8         Brentford           13       51.613
## 9    Manchester Utd           17       51.613
## 10         Brighton           17       50.936
## 11      Southampton           14       50.259
## 12          Arsenal           20       49.582
## 13      Aston Villa           13       49.582
## 14          Watford           13       49.582
## 15   Leicester City           15       48.905
## 16          Burnley            9       48.228
## 17     Leeds United           11       47.551
```

```
## 18      Tottenham         16      47.551
## 19   Newcastle Utd          6      44.166
## 20    Norwich City          8      38.750
```

```
EPL_Season %>%
  select(Squad, Total_points, StatsBombPyt) %>%
  ggplot(aes(x=StatsBombPyt, y=reorder(Squad, StatsBombPyt, sum))) +
  geom_bar(stat='identity', fill='steelblue')
```