Week 1 Assignment

26 July 2020

Step 1

```r
# As usual, we begin by importing the packages we will need

library("readxl",quietly = TRUE)
library("tidyverse",quietly = TRUE)
```

```r
# Load the data.
# EPL results for the 2017/18 season

EPL18 <- read_excel('EPL2017-18.xlsx')
names(EPL18)
```

```
## [1] "Date"     "HomeTeam" "AwayTeam" "FTHG"     "FTAG"     "FTR"
```

Step 2

```r
# Our data is in the form of game results.
# We first identify whether the result was a win for the home team (H),
# the away team (A) or a draw (D).
# We also create the counting variable.


EPL18[,'hwinvalue'] <- ifelse(EPL18$FTR=='H',1,ifelse(EPL18$FTR=='D',.5,0))

EPL18[,'awinvalue'] <- ifelse(EPL18$FTR=='A',1,ifelse(EPL18$FTR=='D',.5,0))

EPL18[,'count']=1
```

Step 3

```r
# divide data into two halves- games before Jan 1 2018 (call them EPL2017)
# and games after (call them EPL2018)

EPL2017 <- EPL18 %>% filter(Date < 20180000)
head(EPL2017)
```

```
## # A tibble: 6 x 9
##       Date HomeTeam    AwayTeam   FTHG  FTAG FTR   hwinvalue awinvalue count
```

```
##      <dbl> <chr>       <chr>      <dbl> <dbl> <chr>      <dbl>     <dbl> <dbl>
## 1  2.02e7 Arsenal     Leicester      4     3 H              1         0     1
## 2  2.02e7 Brighton    Man City       0     2 A              0         1     1
## 3  2.02e7 Chelsea     Burnley        2     3 A              0         1     1
## 4  2.02e7 Crystal Pa~ Huddersf~      0     3 A              0         1     1
## 5  2.02e7 Everton     Stoke          1     0 H              1         0     1
## 6  2.02e7 Southampton Swansea        0     0 D            0.5       0.5     1
```

```r
tail(EPL2017)
```

```
## # A tibble: 6 x 9
##       Date HomeTeam    AwayTeam    FTHG  FTAG FTR    hwinvalue awinvalue count
##      <dbl> <chr>       <chr>      <dbl> <dbl> <chr>      <dbl>     <dbl> <dbl>
## 1  2.02e7 Liverpool   Leicester      2     1 H              1         0     1
## 2  2.02e7 Man United  Southamp~      0     0 D            0.5       0.5     1
## 3  2.02e7 Newcastle   Brighton       0     0 D            0.5       0.5     1
## 4  2.02e7 Watford     Swansea        1     2 A              0         1     1
## 5  2.02e7 Crystal Pa~ Man City       0     0 D            0.5       0.5     1
## 6  2.02e7 West Brom   Arsenal        1     1 D            0.5       0.5     1
```

```r
EPL2018 = EPL18%>% filter(Date > 20180000)
head(EPL2018)
```

```
## # A tibble: 6 x 9
##       Date HomeTeam    AwayTeam    FTHG  FTAG FTR    hwinvalue awinvalue count
##      <dbl> <chr>       <chr>      <dbl> <dbl> <chr>      <dbl>     <dbl> <dbl>
## 1 20180101 Brighton    Bournemou~     2     2 D            0.5       0.5     1
## 2 20180101 Burnley     Liverpool      1     2 A              0         1     1
## 3 20180101 Everton     Man United     0     2 A              0         1     1
## 4 20180101 Leicester   Huddersfi~     3     0 H              1         0     1
## 5 20180101 Stoke       Newcastle      0     1 A              0         1     1
## 6 20180102 Man City    Watford        3     1 H              1         0     1
```

```r
tail(EPL2018)
```

```
## # A tibble: 6 x 9
##       Date HomeTeam    AwayTeam    FTHG  FTAG FTR    hwinvalue awinvalue count
##      <dbl> <chr>       <chr>      <dbl> <dbl> <chr>      <dbl>     <dbl> <dbl>
## 1 20180513 Man United  Watford        1     0 H              1         0     1
## 2 20180513 Newcastle   Chelsea        3     0 H              1         0     1
## 3 20180513 Southampton Man City       0     1 A              0         1     1
## 4 20180513 Swansea     Stoke          1     2 A              0         1     1
## 5 20180513 Tottenham   Leicest~       5     4 H              1         0     1
## 6 20180513 West Ham    Everton        3     1 H              1         0     1
```

Step 4 (home team)

```r
# First half the season

# We have to create separate dfs to calculate home team and away team performance.

# Here is the home team df, including only the variables we need.
EPL2017home <- EPL2017 %>%  group_by(HomeTeam)%>%
  dplyr::summarise(count = sum(count),hwinvalue = sum(hwinvalue),
FTHG = sum(FTHG),FTAG = sum(FTAG)) %>%
  dplyr::rename( team = HomeTeam, Ph = count, FTHGh = FTHG, FTAGh = FTAG)

head(EPL2017home)
```

```
## # A tibble: 6 x 5
##   team              Ph hwinvalue FTHGh FTAGh
##   <chr>          <dbl>     <dbl> <dbl> <dbl>
## 1 Arsenal           10       8.5    25    10
## 2 Bournemouth       11       4.5    14    17
## 3 Brighton          10       5.5    10    12
## 4 Burnley           10       6       7     6
## 5 Chelsea           11       8.5    21     7
## 6 Crystal Palace    11       5      14    18
```

```r
tail(EPL2017home)
```

```
## # A tibble: 6 x 5
##   team          Ph hwinvalue FTHGh FTAGh
##   <chr>      <dbl>     <dbl> <dbl> <dbl>
## 1 Stoke         10       5      13    19
## 2 Swansea       10       3       6    15
## 3 Tottenham     10       7.5    21     8
## 4 Watford       11       4.5    14    23
## 5 West Brom     11       4.5    10    15
## 6 West Ham       9       4      10    14
```

```r
# Now we create the mirror image df for the away team results.

EPL2017away <- EPL2017 %>%  group_by(AwayTeam)%>%
  dplyr::summarise(count = sum(count),awinvalue = sum(awinvalue),
FTHG = sum(FTHG),FTAG = sum(FTAG)) %>%
  dplyr::rename( team = AwayTeam, Pa = count, FTHGa = FTHG, FTAGa = FTAG)

head(EPL2017away)
```

```
## # A tibble: 6 x 5
##   team              Pa awinvalue FTHGa FTAGa
```

```
##    <chr>          <dbl>    <dbl> <dbl> <dbl>
## 1 Arsenal           11        5    16    13
## 2 Bournemouth       10        3    15     6
## 3 Brighton          11        3    13     5
## 4 Burnley           11      6.5    11    11
## 5 Chelsea           10        7     7    18
## 6 Crystal Palace    10      2.5    14     4
```

```
tail(EPL2017away)
```

```
## # A tibble: 6 x 5
##    team          Pa awinvalue FTHGa FTAGa
##    <chr>      <dbl>     <dbl> <dbl> <dbl>
## 1 Stoke         11       2.5    27    10
## 2 Swansea       11         3    17     7
## 3 Tottenham     10       5.5    12    18
## 4 Watford       10       4.5    14    16
## 5 West Brom     10       2.5    13     5
## 6 West Ham      11         3    24    12
```

Step 5 (home team)

```
# Merge the home team and away team results.

EPL17 <- merge(EPL2017home, EPL2017away, by = 'team')
head(EPL17)
```

```
##              team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa
## 1         Arsenal 10       8.5    25    10 11       5.0    16    13
## 2     Bournemouth 11       4.5    14    17 10       3.0    15     6
## 3        Brighton 10       5.5    10    12 11       3.0    13     5
## 4         Burnley 10       6.0     7     6 11       6.5    11    11
## 5         Chelsea 11       8.5    21     7 10       7.0     7    18
## 6  Crystal Palace 11       5.0    14    18 10       2.5    14     4
```

```
tail(EPL17)
```

```
##             team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa
## 15         Stoke 10       5.0    13    19 11       2.5    27    10
## 16       Swansea 10       3.0     6    15 11       3.0    17     7
## 17     Tottenham 10       7.5    21     8 10       5.5    12    18
## 18       Watford 11       4.5    14    23 10       4.5    14    16
## 19     West Brom 11       4.5    10    15 10       2.5    13     5
## 20      West Ham  9       4.0    10    14 11       3.0    24    12
```

4

Step 6 (home team)

```
# Sum the results by home and away measures to get the team overall performance
# for the season

EPL17[,'W'] = EPL17[,'hwinvalue']+EPL17[,'awinvalue']
EPL17[,'G'] = EPL17[,'Ph']+EPL17[,'Pa']
EPL17[,'GF'] = EPL17[,'FTHGh']+EPL17[,'FTAGa']
EPL17[,'GA'] = EPL17[,'FTAGh']+EPL17[,'FTHGa']
head(EPL17)
```

```
##              team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W    G
## 1         Arsenal 10       8.5    25    10 11       5.0    16    13 13.5 21
## 2     Bournemouth 11       4.5    14    17 10       3.0    15     6  7.5 21
## 3        Brighton 10       5.5    10    12 11       3.0    13     5  8.5 21
## 4         Burnley 10       6.0     7     6 11       6.5    11    11 12.5 21
## 5         Chelsea 11       8.5    21     7 10       7.0     7    18 15.5 21
## 6   Crystal Palace 11       5.0    14    18 10       2.5    14     4  7.5 21
##    GF GA
## 1 38 26
## 2 20 32
## 3 15 25
## 4 18 17
## 5 39 14
## 6 18 32
```

```
tail(EPL17)
```

```
##            team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W    G GF
## 15        Stoke 10       5.0    13    19 11       2.5    27    10  7.5 21 23
## 16      Swansea 10       3.0     6    15 11       3.0    17     7  6.0 21 13
## 17    Tottenham 10       7.5    21     8 10       5.5    12    18 13.0 20 39
## 18      Watford 11       4.5    14    23 10       4.5    14    16  9.0 21 30
## 19    West Brom 11       4.5    10    15 10       2.5    13     5  7.0 21 15
## 20     West Ham  9       4.0    10    14 11       3.0    24    12  7.0 20 22
##     GA
## 15 46
## 16 32
## 17 20
## 18 37
## 19 28
## 20 38
```

```
# Create the win percentage and Pythagorean Expectation

EPL17[,'wpc17'] = EPL17[,'W']/EPL17[,'G']
```

```
EPL17[,'pyth17'] = EPL17[,'GF']**2/(EPL17[,'GF']**2 + EPL17[,'GA']**2)
head(EPL17)
```

```
##                team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W  G
## 1           Arsenal 10       8.5    25    10 11       5.0    16    13 13.5 21
## 2       Bournemouth 11       4.5    14    17 10       3.0    15     6  7.5 21
## 3          Brighton 10       5.5    10    12 11       3.0    13     5  8.5 21
## 4           Burnley 10       6.0     7     6 11       6.5    11    11 12.5 21
## 5           Chelsea 11       8.5    21     7 10       7.0     7    18 15.5 21
## 6     Crystal Palace 11      5.0    14    18 10       2.5    14     4  7.5 21
##    GF GA     wpc17    pyth17
## 1 38 26 0.6428571 0.6811321
## 2 20 32 0.3571429 0.2808989
## 3 15 25 0.4047619 0.2647059
## 4 18 17 0.5952381 0.5285481
## 5 39 14 0.7380952 0.8858474
## 6 18 32 0.3571429 0.2403561
```
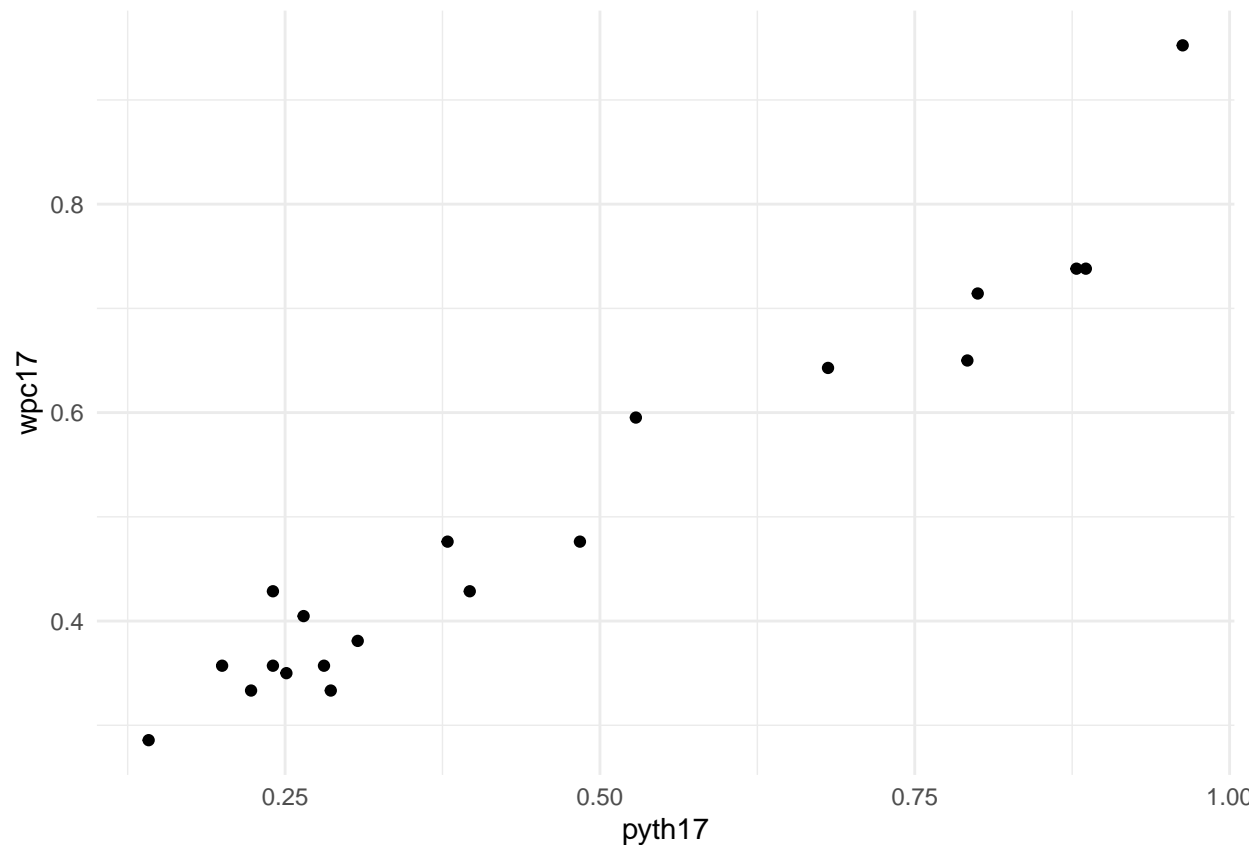
```
tail(EPL17)
```

```
##            team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W  G GF
## 15        Stoke 10       5.0    13    19 11       2.5    27    10  7.5 21 23
## 16      Swansea 10       3.0     6    15 11       3.0    17     7  6.0 21 13
## 17    Tottenham 10       7.5    21     8 10       5.5    12    18 13.0 20 39
## 18      Watford 11       4.5    14    23 10       4.5    14    16  9.0 21 30
## 19    West Brom 11       4.5    10    15 10       2.5    13     5  7.0 21 15
## 20     West Ham  9       4.0    10    14 11       3.0    24    12  7.0 20 22
##    GA     wpc17    pyth17
## 15 46 0.3571429 0.2000000
## 16 32 0.2857143 0.1416597
## 17 20 0.6500000 0.7917751
## 18 37 0.4285714 0.3966505
## 19 28 0.3333333 0.2229931
## 20 38 0.3500000 0.2510373
```

Optional steps, not required for Assessment

```
# Plot the data

ggplot(data =EPL17)+
  geom_point(aes(x=pyth17, y=wpc17))+theme_minimal()
```

```r
# Run the regression

pyth_lm = lm(formula = wpc17 ~ pyth17, data=EPL17)
pyth_lm %>% summary()
```

```
##
## Call:
## lm(formula = wpc17 ~ pyth17, data = EPL17)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.066568 -0.030653 -0.009135  0.028715  0.123805
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.19784    0.02119   9.335 2.55e-08 ***
## pyth17       0.65514    0.03990  16.420 2.81e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04701 on 18 degrees of freedom
## Multiple R-squared:  0.9374, Adjusted R-squared:  0.9339
```

```
## F-statistic: 269.6 on 1 and 18 DF,  p-value: 2.811e-12
```

Step 7 (=Step 4 (away team))

```r
EPL2018home <- EPL2018 %>%  group_by(HomeTeam)%>%
  dplyr::summarise(count = sum(count),hwinvalue = sum(hwinvalue),
FTHG = sum(FTHG),FTAG = sum(FTAG)) %>%
  dplyr::rename( team = HomeTeam, Ph = count, FTHGh = FTHG, FTAGh = FTAG)

head(EPL2018home)
```

```
## # A tibble: 6 x 5
##    team             Ph hwinvalue FTHGh FTAGh
##    <chr>         <dbl>     <dbl> <dbl> <dbl>
## 1 Arsenal           9       7.5    29    10
## 2 Bournemouth       8       5      12    13
## 3 Brighton          9       5.5    14    13
## 4 Burnley           9       3.5     9    11
## 5 Chelsea           8       4.5     9     9
## 6 Crystal Palace    8       4.5    15     9
```

```r
tail(EPL2018home)
```

```
## # A tibble: 6 x 5
##    team          Ph hwinvalue FTHGh FTAGh
##    <chr>      <dbl>     <dbl> <dbl> <dbl>
## 1 Stoke          9       2.5     7    11
## 2 Swansea        9       4.5    11     9
## 3 Tottenham      9       7.5    19     8
## 4 Watford        8       5.5    13     8
## 5 West Brom      8       3      11    14
## 6 West Ham      10       6      14    12
```

```r
EPL2018away <- EPL2018 %>%  group_by(AwayTeam)%>%
  dplyr::summarise(count = sum(count),awinvalue = sum(awinvalue),
FTHG = sum(FTHG),FTAG = sum(FTAG)) %>%
  dplyr::rename( team = AwayTeam, Pa = count, FTHGa = FTHG, FTAGa = FTAG)

head(EPL2018away)
```

```
## # A tibble: 6 x 5
##    team             Pa awinvalue FTHGa FTAGa
##    <chr>         <dbl>     <dbl> <dbl> <dbl>
## 1 Arsenal           8       1      15     7
## 2 Bournemouth       9       4      16    13
## 3 Brighton          8       1.5    16     5
```

```
## 4 Burnley              8        4       11       9
## 5 Chelsea              9      4.5       15      14
## 6 Crystal Palace       9      4.5       14      12
```

tail(EPL2018away)

```
## # A tibble: 6 x 5
##   team         Pa awinvalue FTHGa FTAGa
##   <chr>     <dbl>     <dbl> <dbl> <dbl>
## 1 Stoke         8         3    11     5
## 2 Swansea       8         2    15     4
## 3 Tottenham     9       6.5     8    16
## 4 Watford       9       0.5    19     1
## 5 West Brom     9       2.5    14     5
## 6 West Ham      8         3    18    12
```

Step 7 (=Step 5 (away team))

```
EPL2018 <- merge(EPL2018home, EPL2018away, by = 'team')
head(EPL2018)
```

```
##             team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa
## 1        Arsenal  9       7.5    29    10  8       1.0    15     7
## 2    Bournemouth  8       5.0    12    13  9       4.0    16    13
## 3       Brighton  9       5.5    14    13  8       1.5    16     5
## 4        Burnley  9       3.5     9    11  8       4.0    11     9
## 5        Chelsea  8       4.5     9     9  9       4.5    15    14
## 6 Crystal Palace  8       4.5    15     9  9       4.5    14    12
```

tail(EPL2018)

```
##           team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa
## 15       Stoke  9       2.5     7    11  8       3.0    11     5
## 16     Swansea  9       4.5    11     9  8       2.0    15     4
## 17   Tottenham  9       7.5    19     8  9       6.5     8    16
## 18     Watford  8       5.5    13     8  9       0.5    19     1
## 19   West Brom  8       3.0    11    14  9       2.5    14     5
## 20    West Ham 10       6.0    14    12  8       3.0    18    12
```

Step 7 (=Step 6 (away team))

```
EPL2018[,'W'] = EPL2018[,'hwinvalue']+EPL2018[,'awinvalue']
EPL2018[,'G'] = EPL2018[,'Ph']+EPL2018[,'Pa']
EPL2018[,'GF'] = EPL2018[,'FTHGh']+EPL2018[,'FTAGa']
EPL2018[,'GA'] = EPL2018[,'FTAGh']+EPL2018[,'FTHGa']
head(EPL2018)
```

```
##              team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W  G
## 1        Arsenal  9       7.5    29    10  8       1.0    15     7 8.5 17
## 2    Bournemouth  8       5.0    12    13  9       4.0    16    13 9.0 17
## 3       Brighton  9       5.5    14    13  8       1.5    16     5 7.0 17
## 4        Burnley  9       3.5     9    11  8       4.0    11     9 7.5 17
## 5        Chelsea  8       4.5     9     9  9       4.5    15    14 9.0 17
## 6 Crystal Palace  8       4.5    15     9  9       4.5    14    12 9.0 17
##    GF GA
## 1 36 25
## 2 25 29
## 3 19 29
## 4 18 22
## 5 23 24
## 6 27 23
```

`tail(EPL2018)`

```
##           team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W  G GF
## 15       Stoke  9       2.5     7    11  8       3.0    11     5  5.5 17 12
## 16     Swansea  9       4.5    11     9  8       2.0    15     4  6.5 17 15
## 17   Tottenham  9       7.5    19     8  9       6.5     8    16 14.0 18 35
## 18     Watford  8       5.5    13     8  9       0.5    19     1  6.0 17 14
## 19   West Brom  8       3.0    11    14  9       2.5    14     5  5.5 17 16
## 20    West Ham 10       6.0    14    12  8       3.0    18    12  9.0 18 26
##    GA
## 15 22
## 16 24
## 17 16
## 18 27
## 19 28
## 20 30
```

```r
# Create the win percentage and Pythagorean Expectation

EPL2018[,'wpc18'] = EPL2018[,'W']/EPL2018[,'G']
EPL2018[,'pyth18'] = EPL2018[,'GF']**2/(EPL2018[,'GF']**2 + EPL2018[,'GA']**2)
head(EPL2018)
```

```
##              team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W  G
## 1        Arsenal  9       7.5    29    10  8       1.0    15     7 8.5 17
## 2    Bournemouth  8       5.0    12    13  9       4.0    16    13 9.0 17
## 3       Brighton  9       5.5    14    13  8       1.5    16     5 7.0 17
## 4        Burnley  9       3.5     9    11  8       4.0    11     9 7.5 17
## 5        Chelsea  8       4.5     9     9  9       4.5    15    14 9.0 17
## 6 Crystal Palace  8       4.5    15     9  9       4.5    14    12 9.0 17
##    GF GA     wpc18    pyth18
```

```
## 1 36 25 0.5000000 0.6746486
## 2 25 29 0.5294118 0.4263302
## 3 19 29 0.4117647 0.3003328
## 4 18 22 0.4411765 0.4009901
## 5 23 24 0.5294118 0.4787330
## 6 27 23 0.5294118 0.5794913
```

```r
tail(EPL2018)
```
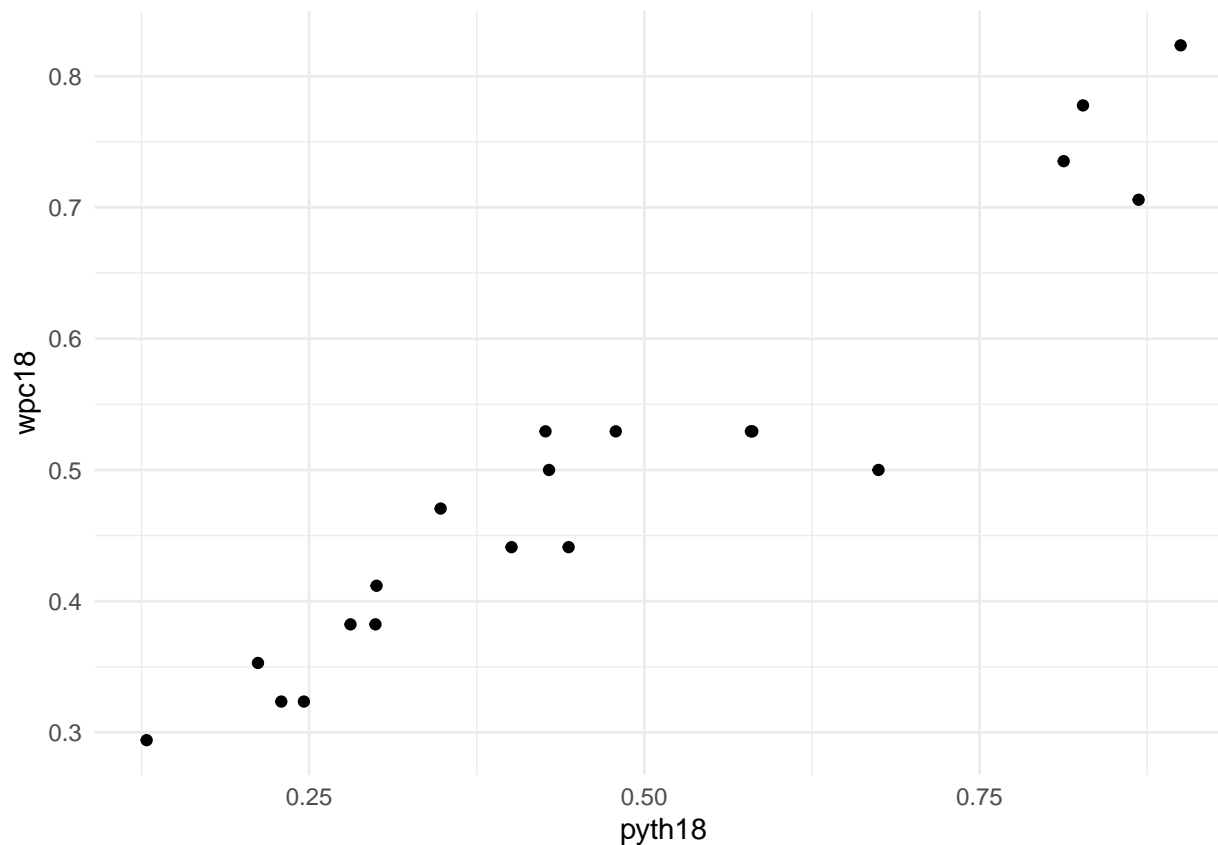
```
##           team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W    G GF
## 15       Stoke  9       2.5     7    11  8       3.0    11     5  5.5 17 12
## 16     Swansea  9       4.5    11     9  8       2.0    15     4  6.5 17 15
## 17   Tottenham  9       7.5    19     8  9       6.5     8    16 14.0 18 35
## 18     Watford  8       5.5    13     8  9       0.5    19     1  6.0 17 14
## 19   West Brom  8       3.0    11    14  9       2.5    14     5  5.5 17 16
## 20    West Ham 10       6.0    14    12  8       3.0    18    12  9.0 18 26
##    GA     wpc18     pyth18
## 15 22 0.3235294 0.2292994
## 16 24 0.3823529 0.2808989
## 17 16 0.7777778 0.8271438
## 18 27 0.3529412 0.2118919
## 19 28 0.3235294 0.2461538
## 20 30 0.5000000 0.4289340
```

Optional steps, not required for Assessment

```r
# Plot the data

ggplot(data =EPL2018)+
  geom_point(aes(x=pyth18, y=wpc18))+theme_minimal()
```

```
# Run the regression

pyth_lm = lm(formula = wpc18 ~ pyth18, data=EPL2018)
pyth_lm %>% summary()
```

```
## 
## Call:
## lm(formula = wpc18 ~ pyth18, data = EPL2018)
## 
## Residuals:
##       Min        1Q    Median        3Q       Max 
## -0.124735 -0.034465  0.006485  0.027278  0.059548 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  0.20397    0.02376   8.586 8.83e-08 ***
## pyth18       0.62368    0.04512  13.823 5.02e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.04653 on 18 degrees of freedom
## Multiple R-squared:  0.9139, Adjusted R-squared:  0.9091
```

```
## F-statistic: 191.1 on 1 and 18 DF,  p-value: 5.017e-11
```

Step 8

```
Half2predictor <- merge(EPL17,EPL2018[,c('team','wpc18','pyth18')], by ='team')
head(Half2predictor)
```

```
##                team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W  G
## 1          Arsenal 10       8.5    25    10 11       5.0    16    13 13.5 21
## 2      Bournemouth 11       4.5    14    17 10       3.0    15     6  7.5 21
## 3         Brighton 10       5.5    10    12 11       3.0    13     5  8.5 21
## 4          Burnley 10       6.0     7     6 11       6.5    11    11 12.5 21
## 5          Chelsea 11       8.5    21     7 10       7.0     7    18 15.5 21
## 6   Crystal Palace 11       5.0    14    18 10       2.5    14     4  7.5 21
##    GF GA      wpc17     pyth17     wpc18     pyth18
## 1 38 26 0.6428571 0.6811321 0.5000000 0.6746486
## 2 20 32 0.3571429 0.2808989 0.5294118 0.4263302
## 3 15 25 0.4047619 0.2647059 0.4117647 0.3003328
## 4 18 17 0.5952381 0.5285481 0.4411765 0.4009901
## 5 39 14 0.7380952 0.8858474 0.5294118 0.4787330
## 6 18 32 0.3571429 0.2403561 0.5294118 0.5794913
```

```
tail(Half2predictor)
```

```
##             team Ph hwinvalue FTHGh FTAGh Pa awinvalue FTHGa FTAGa    W  G GF
## 15        Stoke 10       5.0    13    19 11       2.5    27    10  7.5 21 23
## 16      Swansea 10       3.0     6    15 11       3.0    17     7  6.0 21 13
## 17    Tottenham 10       7.5    21     8 10       5.5    12    18 13.0 20 39
## 18      Watford 11       4.5    14    23 10       4.5    14    16  9.0 21 30
## 19    West Brom 11       4.5    10    15 10       2.5    13     5  7.0 21 15
## 20     West Ham  9       4.0    10    14 11       3.0    24    12  7.0 20 22
##     GA      wpc17     pyth17     wpc18     pyth18
## 15 46 0.3571429 0.2000000 0.3235294 0.2292994
## 16 32 0.2857143 0.1416597 0.3823529 0.2808989
## 17 20 0.6500000 0.7917751 0.7777778 0.8271438
## 18 37 0.4285714 0.3966505 0.3529412 0.2118919
## 19 28 0.3333333 0.2229931 0.3235294 0.2461538
## 20 38 0.3500000 0.2510373 0.5000000 0.4289340
```
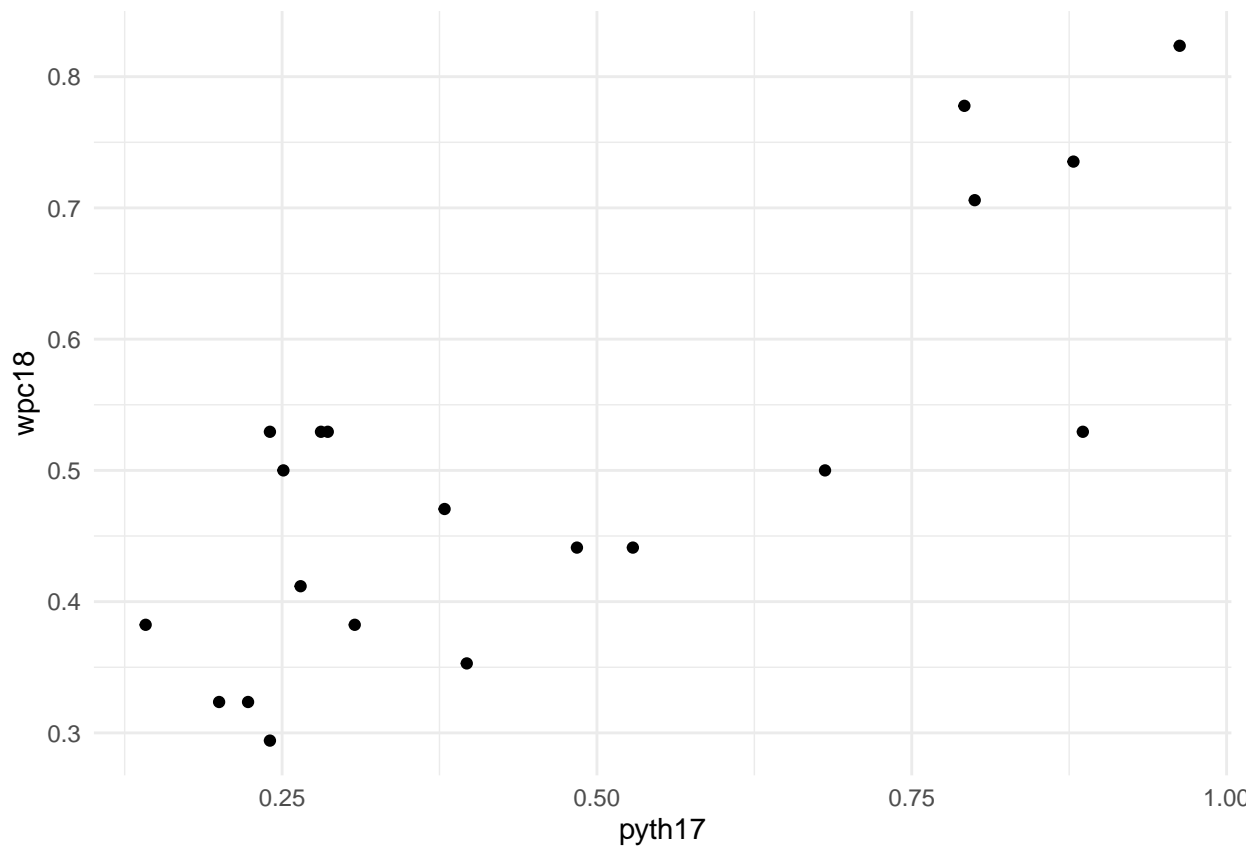
Step 9

```
keyvars <- Half2predictor[,c('team','wpc17','wpc18','pyth17','pyth18')]
keyvars[,-1] %>% cor()
```

```
##           wpc17     wpc18     pyth17     pyth18
```
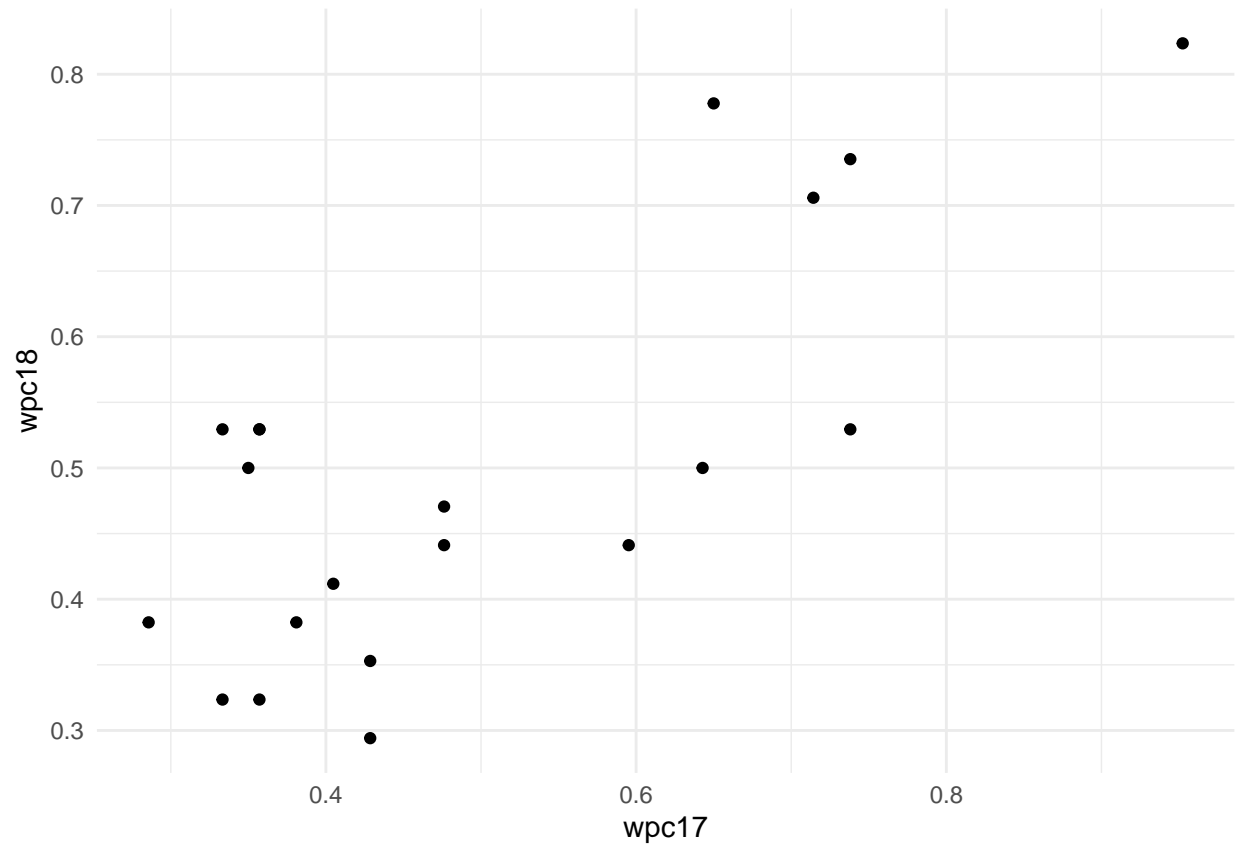
```
## wpc17   1.0000000 0.7565726 0.9682037 0.7458319
## wpc18   0.7565726 1.0000000 0.7956926 0.9559855
## pyth17 0.9682037 0.7956926 1.0000000 0.7953312
## pyth18 0.7458319 0.9559855 0.7953312 1.0000000
```

Optional steps, not required for Assessment

```
ggplot(data = Half2predictor)+
  geom_point(aes(x=pyth17, y=wpc18))+theme_minimal()
```



```
ggplot(data = Half2predictor)+
  geom_point(aes(x=wpc17, y=wpc18))+theme_minimal()
```

Now you have completed the assignment, are these results consistent with those we found for Major League Baseball?