MOOC 1 Week 5 assignment

As usual, we begin by loading the packages we will need

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

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

Now we load the data

```
IPL = read_xlsx("IPL (assignment) data.xlsx")
```

```
summary(IPL)
```

```
##       year          team              played           won        
##  Min.   :2008   Length:92          Min.   :14.00   Min.   : 2.000  
##  1st Qu.:2010   Class :character   1st Qu.:14.00   1st Qu.: 6.000  
##  Median :2013   Mode  :character   Median :14.00   Median : 7.000  
##  Mean   :2013                      Mean   :14.39   Mean   : 7.076  
##  3rd Qu.:2016                      3rd Qu.:14.00   3rd Qu.: 9.000  
##  Max.   :2018                      Max.   :16.00   Max.   :11.000  
##       lost           noresult          points        netrunrate       
##  Min.   : 3.000   Min.   :0.0000   Min.   : 4.00   Min.   :-1.436000  
##  1st Qu.: 5.000   1st Qu.:0.0000   1st Qu.:12.00   1st Qu.:-0.343750  
##  Median : 7.000   Median :0.0000   Median :14.00   Median : 0.018000  
##  Mean   : 7.076   Mean   :0.2391   Mean   :14.39   Mean   : 0.001772  
##  3rd Qu.: 8.000   3rd Qu.:0.0000   3rd Qu.:18.00   3rd Qu.: 0.323000  
##  Max.   :13.000   Max.   :2.0000   Max.   :22.00   Max.   : 1.084000  
##    champions        runnersup          third             fourth        
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000  
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000  
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.00000  
##  Mean   :0.1196   Mean   :0.1196   Mean   :0.09783   Mean   :0.08696  
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000  
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000  
##     salaries        
##  Min.   : 1725000  
##  1st Qu.: 5178750  
##  Median : 6809583  
##  Mean   : 7110445  
##  3rd Qu.: 9286749  
##  Max.   :13345000
```

```
str(IPL)
```

```
## tibble [92 x 13] (S3: tbl_df/tbl/data.frame)
```

```
##  $ year      : num [1:92] 2008 2009 2010 2011 2012 ...
##  $ team      : chr [1:92] "Chennai Super Kings" "Chennai Super Kings" "Chennai Super Kings" "Chennai
##  $ played    : num [1:92] 14 14 14 14 16 16 14 14 14 14 ...
##  $ won       : num [1:92] 8 8 7 9 8 11 9 9 9 2 ...
##  $ lost      : num [1:92] 6 5 7 5 7 5 5 5 5 12 ...
##  $ noresult  : num [1:92] 0 1 0 0 1 0 0 0 0 0 ...
##  $ points    : num [1:92] 16 17 14 18 17 22 18 18 18 4 ...
##  $ netrunrate: num [1:92] -0.192 0.951 0.274 0.443 0.1 0.53 0.385 0.709 0.253 -0.467 ...
##  $ champions : num [1:92] 0 0 1 1 0 0 0 0 1 0 ...
##  $ runnersup : num [1:92] 1 0 0 0 1 1 0 1 0 0 ...
##  $ third     : num [1:92] 0 0 0 0 0 0 1 0 0 0 ...
##  $ fourth    : num [1:92] 0 0 0 0 0 0 0 0 0 0 ...
##  $ salaries  : num [1:92] 5825000 6765000 4890000 6330000 7900000 ...
```

```
Sumsal = IPL %>% group_by(year) %>%
  summarise(allsal = sum(salaries))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(Sumsal)
```

```
## # A tibble: 6 x 2
##    year   allsal
##   <dbl>    <dbl>
## 1  2008 34105000
## 2  2009 33445000
## 3  2010 33075000
## 4  2011 62210000
## 5  2012 59706250
## 6  2013 81535000
```

```
tail(Sumsal)
```

```
## # A tibble: 6 x 2
##    year   allsal
##   <dbl>    <dbl>
## 1  2013 81535000
## 2  2014 73973332
## 3  2015 65242665
## 4  2016 63483666
## 5  2017 63014833
## 6  2018 84370160
```

```
IPL = left_join(IPL, Sumsal, by='year')
head(IPL)
```

```
## # A tibble: 6 x 14
##    year team  played   won  lost noresult points netrunrate champions runnersup
##   <dbl> <chr>  <dbl> <dbl> <dbl>    <dbl>  <dbl>      <dbl>     <dbl>     <dbl>
## 1  2008 Chen~     14     8     6        0     16     -0.192         0         1
## 2  2009 Chen~     14     8     5        1     17      0.951         0         0
## 3  2010 Chen~     14     7     7        0     14      0.274         1         0
## 4  2011 Chen~     14     9     5        0     18      0.443         1         0
## 5  2012 Chen~     16     8     7        1     17      0.1           0         1
## 6  2013 Chen~     16    11     5        0     22      0.53          0         1
## # ... with 4 more variables: third <dbl>, fourth <dbl>, salaries <dbl>,
## #   allsal <dbl>
```

```
tail(IPL)
```

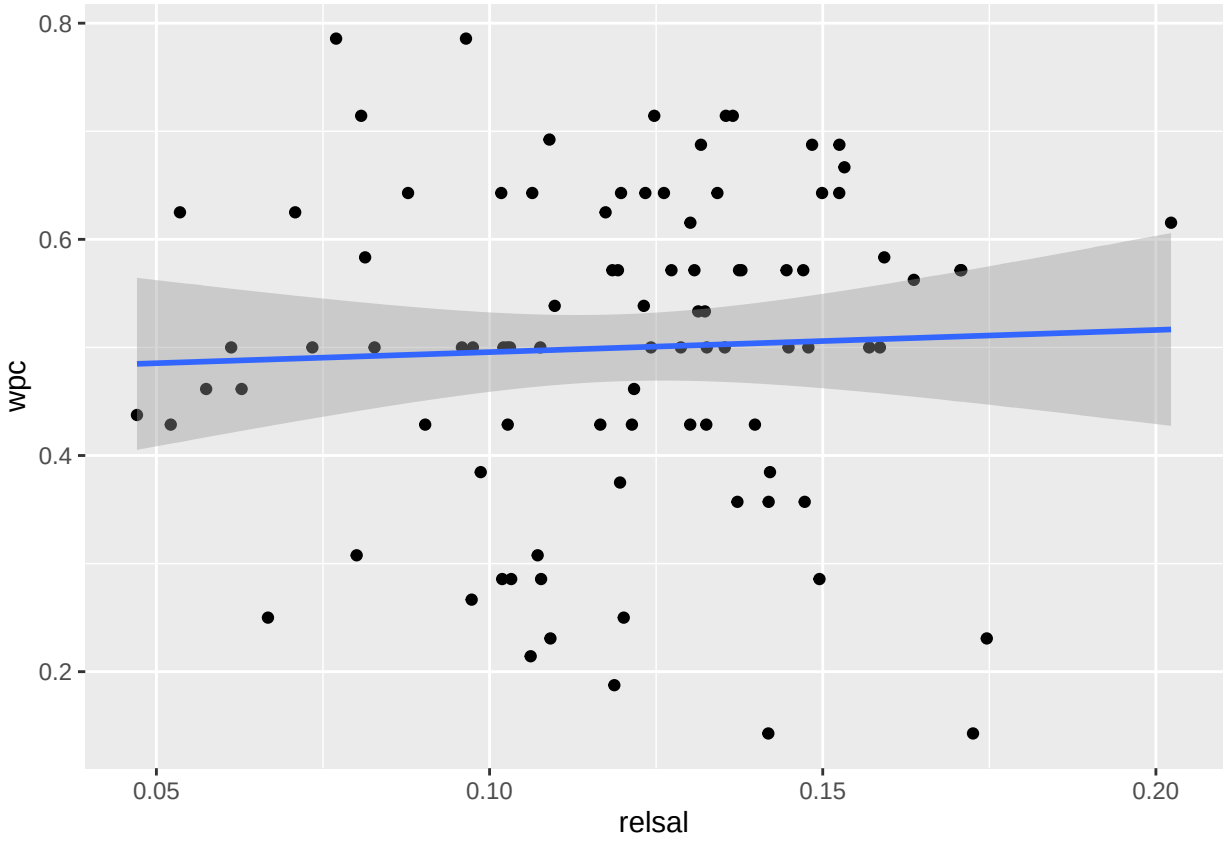
```
## # A tibble: 6 x 14
##    year team  played   won  lost noresult points netrunrate champions runnersup
##   <dbl> <chr>  <dbl> <dbl> <dbl>    <dbl>  <dbl>      <dbl>      <dbl>      <dbl>
## 1  2013 Sunr~     16    10     6        0     20      0.003          0          0
## 2  2014 Sunr~     14     6     8        0     12     -0.399          0          0
## 3  2015 Sunr~     14     7     7        0     14     -0.239          0          0
## 4  2016 Sunr~     14     8     6        0     16      0.245          1          0
## 5  2017 Sunr~     14     8     5        1     17      0.599          0          0
## 6  2018 Sunr~     14     9     5        0     18      0.284          0          1
## # ... with 4 more variables: third <dbl>, fourth <dbl>, salaries <dbl>,
## #   allsal <dbl>
```

```
IPL$relsal= IPL$salaries/IPL$allsal
```

```
IPL$wpc = IPL$won/(IPL$played - IPL$noresult)
```

```
ggplot(IPL, aes(x=relsal, y=wpc)) + geom_point() +
  geom_smooth(method=lm)
```

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




```
wpcsal1_lm = lm(wpc ~ relsal, data = IPL)
summary(wpcsal1_lm)
```

```
##
## Call:
## lm(formula = wpc ~ relsal, data = IPL)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36767 -0.07338  0.00503  0.11661  0.29478
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.47515    0.06283   7.563 3.21e-11 ***
## relsal       0.20498    0.50901   0.403    0.688
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1496 on 90 degrees of freedom
## Multiple R-squared:  0.001799,   Adjusted R-squared:  -0.009292
## F-statistic: 0.1622 on 1 and 90 DF,  p-value: 0.6881
```

```
IPL = IPL %>% arrange(team, year) %>% group_by(team) %>%
  mutate(wpc_lag = lag(wpc)) %>% ungroup()
head(IPL)
```

```
## # A tibble: 6 x 17
##    year team  played   won  lost noresult points netrunrate champions runnersup
##   <dbl> <chr>  <dbl> <dbl> <dbl>    <dbl>  <dbl>      <dbl>     <dbl>     <dbl>
## 1  2008 Chen~     14     8     6        0     16     -0.192         0         1
## 2  2009 Chen~     14     8     5        1     17      0.951         0         0
## 3  2010 Chen~     14     7     7        0     14      0.274         1         0
## 4  2011 Chen~     14     9     5        0     18      0.443         1         0
## 5  2012 Chen~     16     8     7        1     17      0.1           0         1
## 6  2013 Chen~     16    11     5        0     22      0.53          0         1
## # ... with 7 more variables: third <dbl>, fourth <dbl>, salaries <dbl>,
## #   allsal <dbl>, relsal <dbl>, wpc <dbl>, wpc_lag <dbl>
```

```
tail(IPL)
```

```
## # A tibble: 6 x 17
##    year team  played   won  lost noresult points netrunrate champions runnersup
##   <dbl> <chr>  <dbl> <dbl> <dbl>    <dbl>  <dbl>      <dbl>     <dbl>     <dbl>
## 1  2013 Sunr~     16    10     6        0     20      0.003         0         0
## 2  2014 Sunr~     14     6     8        0     12     -0.399         0         0
## 3  2015 Sunr~     14     7     7        0     14     -0.239         0         0
## 4  2016 Sunr~     14     8     6        0     16      0.245         1         0
## 5  2017 Sunr~     14     8     5        1     17      0.599         0         0
## 6  2018 Sunr~     14     9     5        0     18      0.284         0         1
## # ... with 7 more variables: third <dbl>, fourth <dbl>, salaries <dbl>,
## #   allsal <dbl>, relsal <dbl>, wpc <dbl>, wpc_lag <dbl>
```

```
wpcsal2_lm = lm(wpc ~wpc_lag + relsal, data=IPL)
summary(wpcsal2_lm)
```

```
##
## Call:
## lm(formula = wpc ~ wpc_lag + relsal, data = IPL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33858 -0.07134  0.01342  0.09588  0.30110
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.34887    0.08714   4.003 0.000144 ***
## wpc_lag      0.14059    0.10819   1.300 0.197699
## relsal       0.67849    0.52305   1.297 0.198495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1426 on 76 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.04018,    Adjusted R-squared:  0.01492
## F-statistic: 1.591 on 2 and 76 DF,  p-value: 0.2105
```

```
wpcsal3_lm = lm(wpc ~ wpc_lag + relsal + factor(team), data=IPL)
summary(wpcsal3_lm)
```

```
##
## Call:
## lm(formula = wpc ~ wpc_lag + relsal + factor(team), data = IPL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29812 -0.06911  0.01763  0.06681  0.33547
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                            0.66652    0.13810   4.826 8.78e-06
## wpc_lag                               -0.06362    0.11488  -0.554  0.58164
## relsal                                -0.10137    0.68951  -0.147  0.88357
## factor(team)Deccan Chargers           -0.18566    0.08488  -2.187  0.03233
## factor(team)Delhi Daredevils          -0.20591    0.06928  -2.972  0.00414
## factor(team)Gujarat Lions             -0.32944    0.14263  -2.310  0.02409
## factor(team)Kings XI Punjab           -0.17468    0.06858  -2.547  0.01324
## factor(team)Kolkata Knight Riders     -0.09720    0.06503  -1.495  0.13985
## factor(team)Mumbai Indians            -0.04046    0.06508  -0.622  0.53628
## factor(team)Pune Warriors India       -0.38931    0.11794  -3.301  0.00157
## factor(team)Rajasthan Royals          -0.12593    0.08563  -1.471  0.14621
## factor(team)Rising Pune Supergiants    0.01185    0.14368   0.082  0.93453
## factor(team)Royal Challengers Bangalore -0.11754    0.06393  -1.838  0.07056
## factor(team)Sunrisers Hyderabad       -0.06694    0.07634  -0.877  0.38380
##
## (Intercept)                           ***
## wpc_lag
## relsal
## factor(team)Deccan Chargers           *
## factor(team)Delhi Daredevils          **
## factor(team)Gujarat Lions             *
## factor(team)Kings XI Punjab           *
## factor(team)Kolkata Knight Riders
## factor(team)Mumbai Indians
## factor(team)Pune Warriors India       **
## factor(team)Rajasthan Royals
## factor(team)Rising Pune Supergiants
## factor(team)Royal Challengers Bangalore .
```

```
## factor(team)Sunrisers Hyderabad
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1319 on 65 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.2974, Adjusted R-squared:  0.1569
## F-statistic: 2.117 on 13 and 65 DF,  p-value: 0.02454
```