

```

> library(tidyverse)
> library(broom)
> library(Lahman)
> Teams_small <- Teams %>%
+   filter(yearID %in% 1961:2001) %>%
+   mutate(avg_attendance = attendance/G)
>
> dat
# A tibble: 1,011 x 49
# Groups:   W [6]
   yearID lgID teamID franchID divID Rank    G Ghome    W    L DivWin WCWin
   <int> <fct> <fct>   <fct>   <chr> <int> <int> <dbl> <int> <chr> <chr>
1    1961 AL   BAL    BAL    NA      3   163    82    10    67 NA    NA
2    1961 AL   BOS    BOS    NA      6   163    82     8    86 NA    NA
3    1961 AL   CHA    CHW    NA      4   163    81     9    76 NA    NA
4    1961 NL   CHN    CHC    NA      7   156    78     6    90 NA    NA
5    1961 NL   CIN    CIN    NA      1   154    77     9    61 NA    NA
6    1961 AL   CLE    CLE    NA      5   161    81     8    83 NA    NA
7    1961 AL   DET    DET    NA      2   163    82    10    61 NA    NA
8    1961 AL   KC1    OAK    NA      9   162    80     6   100 NA    NA
9    1961 AL   LAA    ANA    NA      8   162    82     7    91 NA    NA
10   1961 NL   LAN    LAD    NA      2   154    77     9    65 NA    NA
# ... with 1,001 more rows, and 37 more variables: LgWin <chr>, WSWin <chr>,
# R <int>, AB <int>, H <int>, X2B <int>, X3B <int>, HR <int>, BB <dbl>,
# SO <int>, SB <dbl>, CS <dbl>, HBP <dbl>, SF <int>, RA <int>, ER <int>,
# ERA <dbl>, CG <int>, SHO <int>, SV <int>, IPouts <int>, HA <int>,
# HRA <int>, BBA <int>, SOA <int>, E <int>, DP <int>, FP <dbl>, name <chr>,
# park <chr>, attendance <int>, BPF <int>, PPF <int>, teamIDBR <chr>,
# teamIDlahman45 <chr>, teamIDretro <chr>, avg_attendance <dbl>
>
>
>
> dat <- Teams_small %>% mutate(W = round(W/10)) %>% group_by(W) %>% filter(W %in% 5:10 & n() >= 20)
> sum(dat$W == 8)
[1] 338
>
>
>
> dat %>% mutate(R_per_game = R/G) %>% group_by(W) %>% do(tidy(lm(avg_attendance ~ R_per_game, data = .),
conf.int = TRUE)) %>% filter(term == 'R_per_game')
# A tibble: 6 x 8
# Groups:   W [6]
   W term      estimate std.error statistic  p.value conf.low conf.high
<dbl> <chr>      <dbl>      <dbl>      <dbl>    <dbl>    <dbl>    <dbl>
1     5 R_per_game  4362.      1112.      3.92 4.20e- 4    2099.    6624.
2     6 R_per_game  4343.       903.      4.81 5.05e- 6    2552.    6133.
3     7 R_per_game  3888.       464.      8.38 1.08e-14    2973.    4803.
4     8 R_per_game  3128.       380.      8.23 4.06e-15    2381.    3875.
5     9 R_per_game  3701.       607.      6.09 4.75e- 9    2504.    4898.
6    10 R_per_game  3107.       827.      3.76 2.80e- 4    1468.    4746.
>
>
>
> dat %>% mutate(HR_per_game = HR/G) %>% group_by(W) %>% do(tidy(lm(avg_attendance ~ HR_per_game, data = .),
), conf.int = TRUE)) %>% filter(term == 'HR_per_game')
# A tibble: 6 x 8
# Groups:   W [6]
   W term      estimate std.error statistic  p.value conf.low conf.high
<dbl> <chr>      <dbl>      <dbl>      <dbl>    <dbl>    <dbl>    <dbl>
1     5 HR_per_game 10192.     3423.      2.98 5.41e- 3    3227.    17156.
2     6 HR_per_game  7032.     2444.      2.88 4.85e- 3    2187.    11878.
3     7 HR_per_game  8931.     1126.      7.93 1.70e-13    6710.    11151.
4     8 HR_per_game  6301.      886.      7.11 7.05e-12    4557.     8044.
5     9 HR_per_game  5863.     1279.      4.58 7.58e- 6    3342.     8383.
6    10 HR_per_game  4917.     1976.      2.49 1.44e- 2     999.     8835.

```

```

>
>
>
> dat %>% mutate(R_per_game = R/G) %>% group_by(W) %>% summarize(cor = cor(R_per_game, avg_attendance))
# A tibble: 6 x 2
      W    cor
<dbl> <dbl>
1     5 0.564
2     6 0.423
3     7 0.517
4     8 0.410
5     9 0.376
6    10 0.343
> dat %>% mutate(HR_per_game = HR/G) %>% group_by(W) %>% summarize(cor = cor(HR_per_game, avg_attendance))
# A tibble: 6 x 2
      W    cor
<dbl> <dbl>
1     5 0.460
2     6 0.269
3     7 0.496
4     8 0.362
5     9 0.292
6    10 0.235
>

```