MinneMUDAC

Tom Bareket

2023-03-23

```r
#average temp per month
dat2 = add_dist[add_dist$year > 2018,]

months = numeric(length(dat2$Date))
for (d in 1:length(dat2$Date)) {
  #indices 6 and 7 are the date, e.g. the "05" in "2021/05/14"
  months[d] = substr(dat2$Date[d], 6, 7)
}
dat2['murica_temp'] = 9/5 * dat2['HistoricalAvgHrlyTemp'] + 32

mar = dat2[months == '03',]
apr = dat2[months == '04',]
may1 = dat2[months == '05',]
jun = dat2[months == '06',]
jul = dat2[months == '07',]
aug = dat2[months == '08',]
sep = dat2[months == '09',]
oct = dat2[months == '10',]


march = mean(mar$Attendance)
april = mean(apr$Attendance)
may = mean(na.omit(may1$Attendance))
june = mean(jun$Attendance)
july = mean(jul$Attendance)
august =mean(aug$Attendance)
september = mean(sep$Attendance)
october = mean(oct$Attendance)


avg_temp_by_month = c(mean(mar$murica_temp),mean(apr$murica_temp),mean(may1$murica_temp),mean(jun$murica

avg_temp_by_month.t = c(mean(mar$murica_temp),mean(apr$murica_temp),mean(may1$murica_temp),mean(jun$muri
```

```r
#linear regression with basically all vars

dat2 = add_dist[add_dist$year > 2018,]

mod = lm(Attendance ~ DayofWeek + elo + DistanceBetweenStadiums + fixed_roof + retractable_roof + Histor
summary(mod)
```

```
##
```

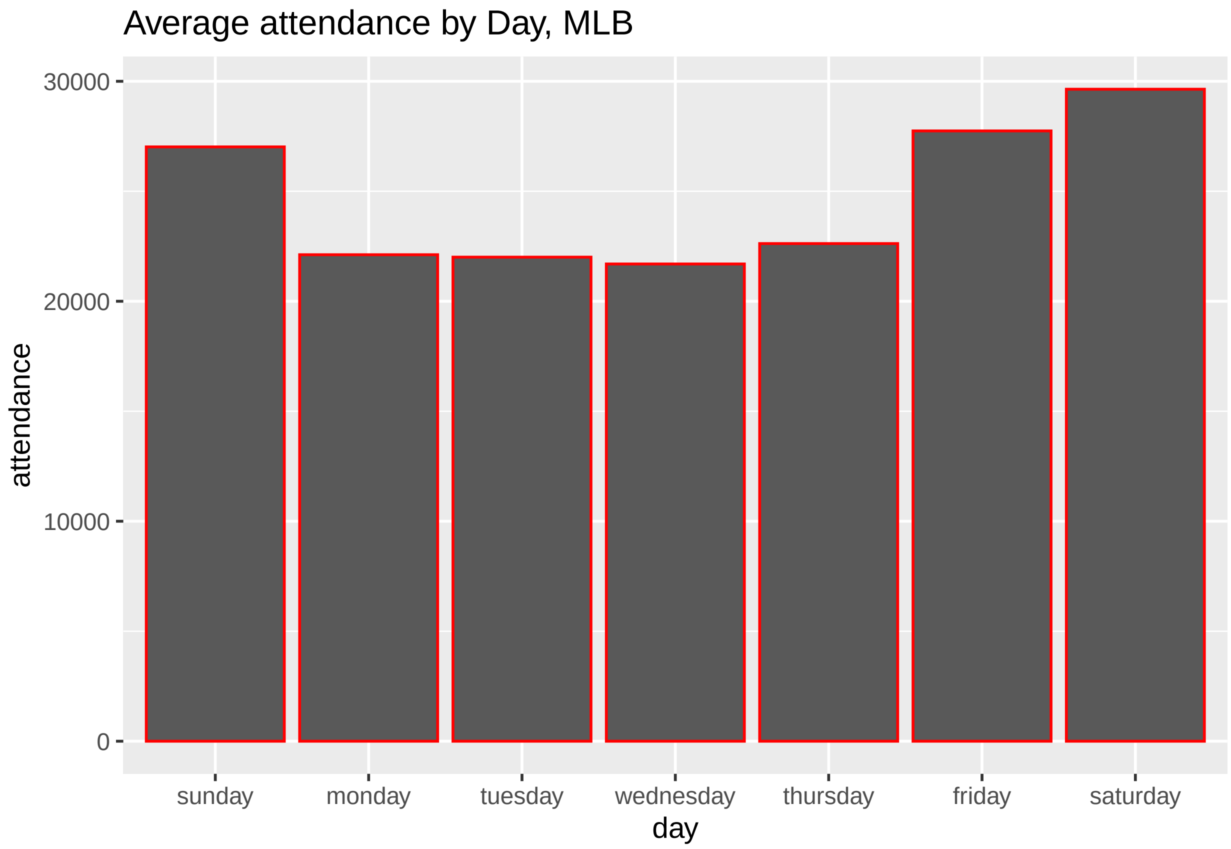
```
## Call:
## lm(formula = Attendance ~ DayofWeek + elo + DistanceBetweenStadiums +
##     fixed_roof + retractable_roof + HistoricalAvgHrlyTemp + DayNight,
##     data = dat2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32401  -6925    226   7402  34839
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -2.169e+05  6.448e+03 -33.632  < 2e-16 ***
## DayofWeekMon            -5.328e+03  5.155e+02 -10.336  < 2e-16 ***
## DayofWeekSat             2.668e+03  4.613e+02   5.783 7.66e-09 ***
## DayofWeekSun             8.602e+02  5.382e+02   1.598    0.110
## DayofWeekThu            -4.361e+03  5.133e+02  -8.496  < 2e-16 ***
## DayofWeekTue            -5.690e+03  4.523e+02 -12.580  < 2e-16 ***
## DayofWeekWed            -5.418e+03  4.690e+02 -11.551  < 2e-16 ***
## elo                      1.597e+02  4.279e+00  37.315  < 2e-16 ***
## DistanceBetweenStadiums -1.301e+01  9.863e+00  -1.319    0.187
## fixed_roof              -1.960e+04  6.942e+02 -28.236  < 2e-16 ***
## retractable_roof        -2.442e+03  3.288e+02  -7.428 1.24e-13 ***
## HistoricalAvgHrlyTemp    2.033e+02  1.774e+01  11.457  < 2e-16 ***
## DayNightN                1.681e+03  3.375e+02   4.981 6.50e-07 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 10200 on 6543 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2822, Adjusted R-squared:  0.2808
## F-statistic: 214.3 on 12 and 6543 DF,  p-value: < 2.2e-16
```

```
#mean for each day

dat2 = add_dist[add_dist$year > 2018,]

sun = dat2[dat2$DayofWeek == "Sun",]
mon = dat2[dat2$DayofWeek == "Mon",]
tue = dat2[dat2$DayofWeek == "Tue",]
wed = dat2[dat2$DayofWeek == "Wed",]
thu = dat2[dat2$DayofWeek == "Thu",]
fri = dat2[dat2$DayofWeek == "Fri",]
sat = dat2[dat2$DayofWeek == "Sat",]

sunday = mean(sun$Attendance)
monday = mean(mon$Attendance)
tuesday = mean(tue$Attendance)
#wednesdays had an NA
wed.att = na.omit(wed$Attendance)
wednesday = mean(wed.att)
thursday = mean(thu$Attendance)
friday = mean(fri$Attendance)
saturday = mean(sat$Attendance)
```

```
#rbind(sunday,monday,tuesday,wednesday,thursday,friday,saturday)

#plotting
library(ggplot2)
```

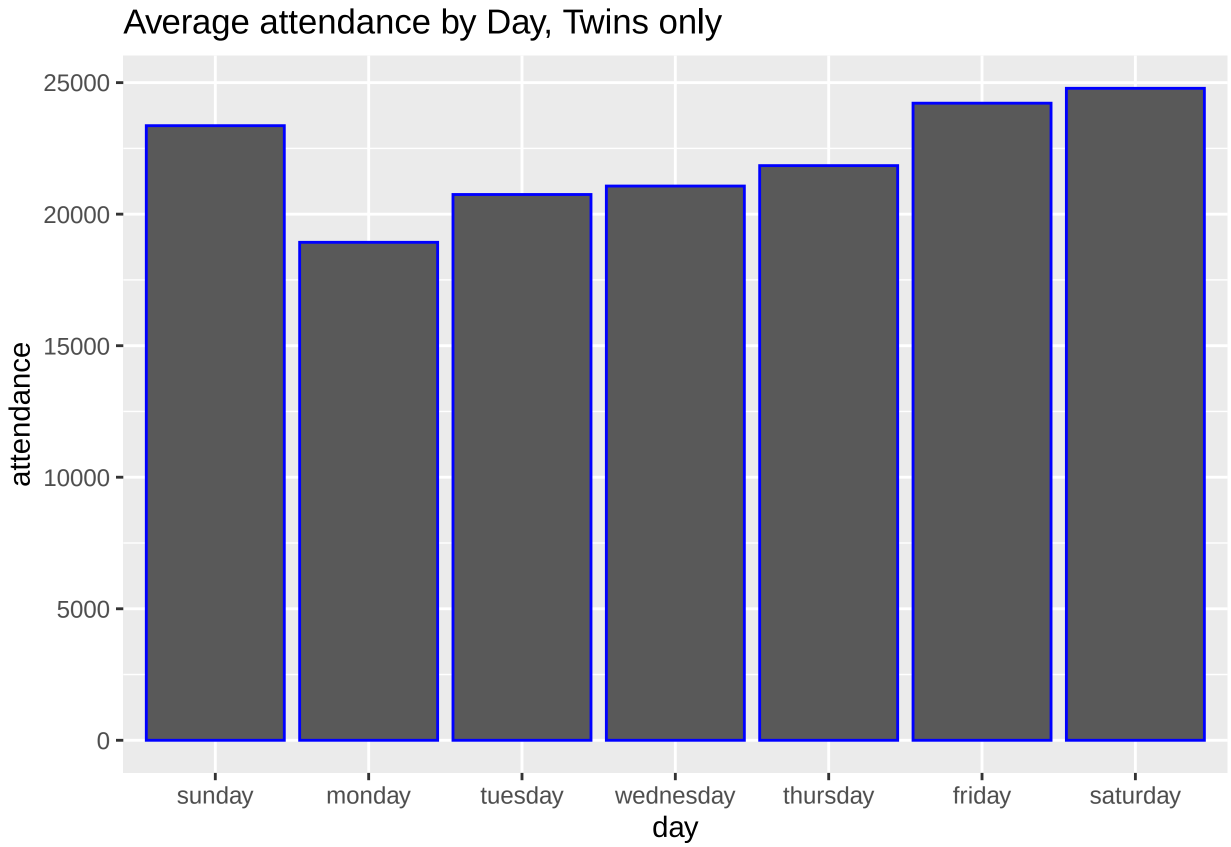
```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
days = c('sunday', 'monday', 'tuesday', 'wednesday', 'thursday', 'friday', 'saturday')
value = c(sunday,monday,tuesday,wednesday,thursday,friday,saturday)
dat = data.frame(day = factor(days, levels = c('sunday', 'monday', 'tuesday', 'wednesday', 'thursday',
ggplot(dat, aes(x = day, y = attendance)) +
  geom_bar(color = "red", stat = "identity", position = "dodge") + ggtitle("Average attendance by Day, M
```



```
#mean for each day, only twins

dat2 = add_dist[add_dist$year > 2018,]
dat3 = dat2[dat2$hometeam == 'MIN',]

sun.t = dat3[dat3$DayofWeek == "Sun",]
mon.t = dat3[dat3$DayofWeek == "Mon",]
tue.t = dat3[dat3$DayofWeek == "Tue",]
wed.t = dat3[dat3$DayofWeek == "Wed",]
thu.t = dat3[dat3$DayofWeek == "Thu",]
fri.t = dat3[dat3$DayofWeek == "Fri",]
```

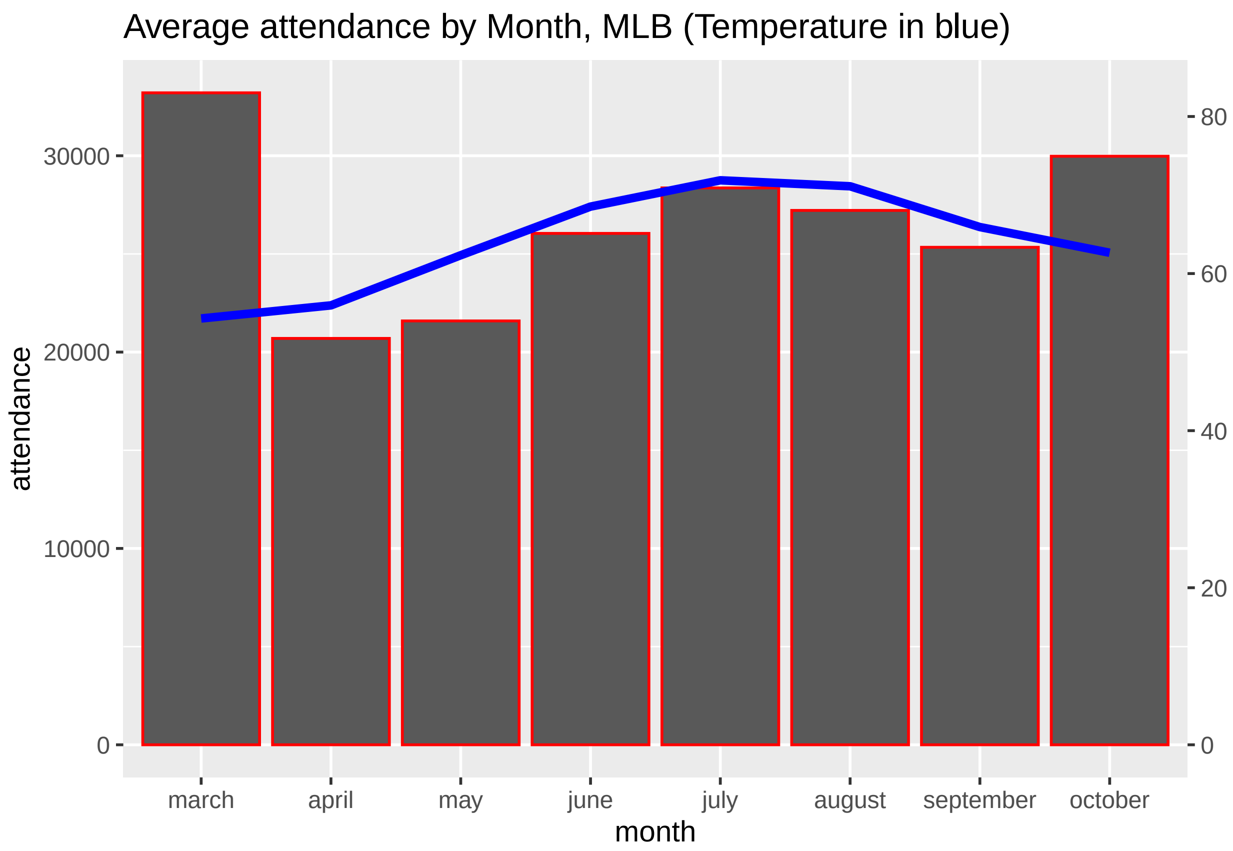
```r
sat.t = dat3[dat3$DayofWeek == "Sat",]
sunday = mean(sun.t$Attendance)
monday = mean(mon.t$Attendance)
tuesday = mean(tue.t$Attendance)
#wednesdays had an NA
wed.att = na.omit(wed.t$Attendance)
wednesday = mean(wed.att)
thursday = mean(thu.t$Attendance)
friday = mean(fri.t$Attendance)
saturday = mean(sat.t$Attendance)

#rbind(sunday,monday,tuesday,wednesday,thursday,friday,saturday)

#plotting
library(ggplot2)
days = c('sunday', 'monday', 'tuesday', 'wednesday', 'thursday', 'friday', 'saturday')
value = c(sunday,monday,tuesday,wednesday,thursday,friday,saturday)
dat = data.frame(day = factor(days, levels = c('sunday', 'monday', 'tuesday', 'wednesday', 'thursday',
ggplot(dat, aes(x = day, y = attendance)) +
  geom_bar(color = "blue", stat = "identity", position = "dodge") + ggtitle("Average attendance by Day,
```



```r
#mean for each month

dat2 = add_dist[add_dist$year > 2018,]
```

```
months = numeric(length(dat2$Date))
for (d in 1:length(dat2$Date)) {
  #indices 6 and 7 are the date, e.g. the "05" in "2021/05/14"
  months[d] = substr(dat2$Date[d], 6, 7)
}

mar = dat2[months == '03',]
apr = dat2[months == '04',]
may1 = dat2[months == '05',]
jun = dat2[months == '06',]
jul = dat2[months == '07',]
aug = dat2[months == '08',]
sep = dat2[months == '09',]
oct = dat2[months == '10',]


march = mean(mar$Attendance)
april = mean(apr$Attendance)
may = mean(na.omit(may1$Attendance))
june = mean(jun$Attendance)
july = mean(jul$Attendance)
august =mean(aug$Attendance)
september = mean(sep$Attendance)
october = mean(oct$Attendance)

#rbind(march, april, may, june, july, august, september, october)


#plotting
library(ggplot2)
months = c('march', 'april', 'may', 'june', 'july', 'august', 'september', 'october')
value = c(march, april, may, june, july, august, september, october)
dat = data.frame(month = factor(months, levels = c('march', 'april', 'may', 'june', 'july', 'august', '
ggplot(dat, aes(x = month, y = attendance, group = 1)) + ylim(0, 50000) +
  geom_bar(color = "red", stat = "identity", position = "dodge") +
  geom_line(aes(y= temperature*400), linewidth = 1.5, color="blue") +
  scale_y_continuous(sec.axis = sec_axis(~./400)) +
  ggtitle("Average attendance by Month, MLB (Temperature in blue)")
```

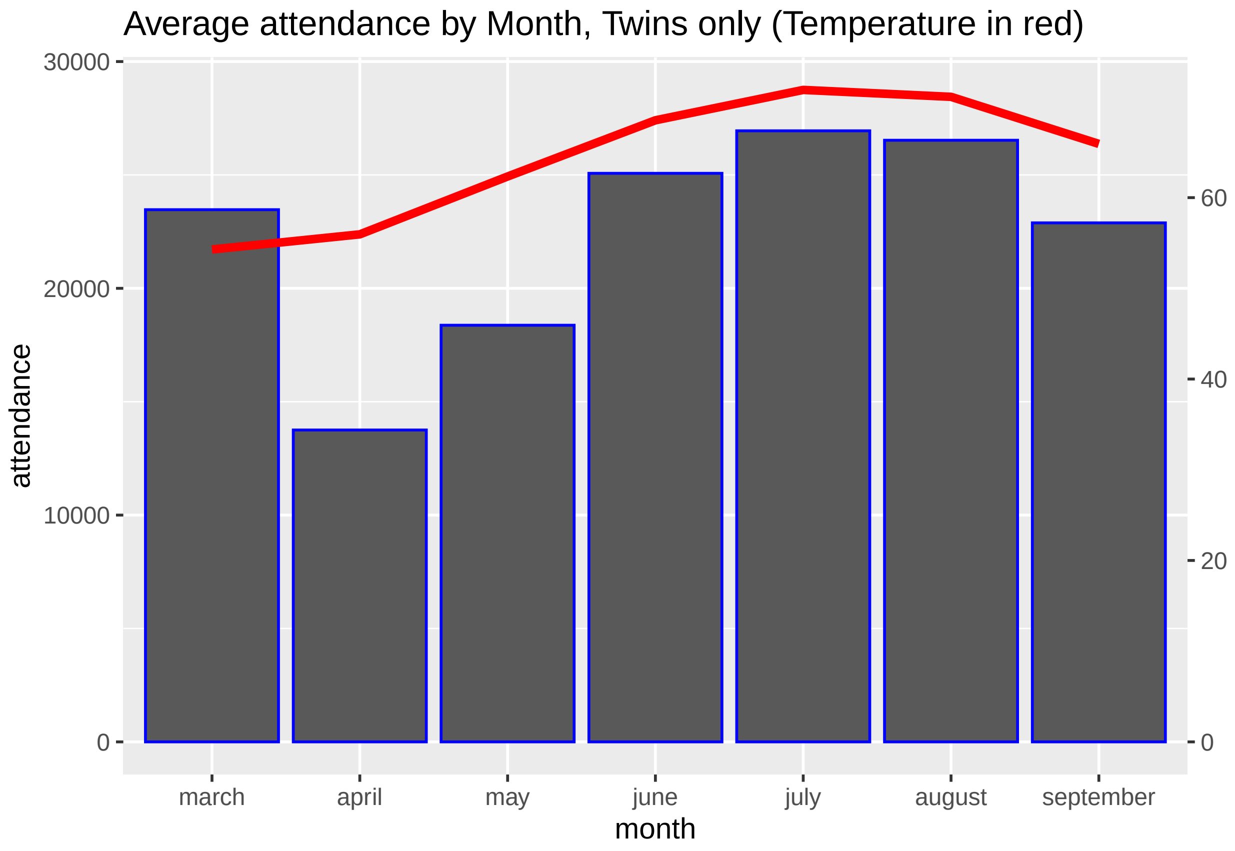
```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```



```
#mean each month, only twins

dat2 = add_dist[add_dist$year > 2018,]
dat3 = dat2[dat2$hometeam == 'MIN',]

months = numeric(length(dat3$Date))
for (d in 1:length(dat3$Date)) {
  months[d] = substr(dat3$Date[d], 6, 7)
}

mar.t = dat3[months == '03',]
apr.t = dat3[months == '04',]
may1.t = dat3[months == '05',]
jun.t = dat3[months == '06',]
jul.t = dat3[months == '07',]
aug.t = dat3[months == '08',]
sep.t = dat3[months == '09',]
oct.t = dat3[months == '10',]

march = mean(mar.t$Attendance)
april = mean(apr.t$Attendance)
may = mean(na.omit(may1.t$Attendance))
june = mean(jun.t$Attendance)
july = mean(jul.t$Attendance)
august =mean(aug.t$Attendance)
september = mean(sep.t$Attendance)
```

```
october = mean(na.omit(oct.t$Attendance))
#No October because Twins didn't play in last 5 years of playoffs

#rbind(march, april, may, june, july, august, september)

#plotting
library(ggplot2)
months = c('march', 'april', 'may', 'june', 'july', 'august', 'september')
value = c(march, april, may, june, july, august, september)
dat = data.frame(month = factor(months, levels = c('march', 'april', 'may', 'june', 'july', 'august', 's
ggplot(dat, aes(x = month, y = attendance, group = 1)) + ylim(0, 50000) +
  geom_bar(color = "blue", stat = "identity", position = "dodge") +
  geom_line(aes(y= temperature*400), linewidth = 1.5, color="red") +
  scale_y_continuous(sec.axis = sec_axis(~./400)) +
  ggtitle("Average attendance by Month, Twins only (Temperature in red)")
```

```
## Scale for y is already present.
## Adding another scale for y, which will replace the existing scale.
```


Average attendance by Month, Twins only (Temperature in red)
attendance
march
april
may
june
july
august
september
month