da420\_PROJECT2\_GRAHN

Jason Grahn

1/17/2019

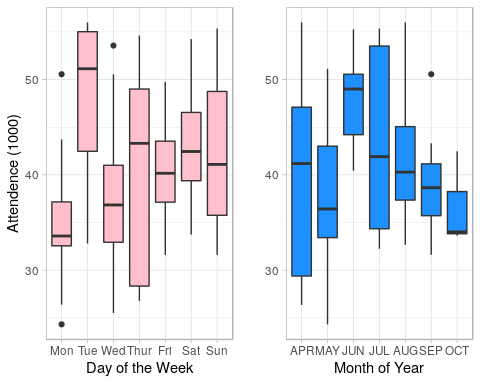
Download Dodger.csv

dodgers <- read\_csv("dodgers.csv",   
 col\_types = cols(#day\_of\_week = col\_factor(levels = c("Monday", "Tuesday",   
 # "Wednesday", "Thursday",   
 # "Friday", "Saturday",   
 # "Sunday")),  
 month = col\_factor(levels = c("JAN", "FEB", "MAR", "APR",  
 "MAY", "JUN", "JUL", "AUG",  
 "SEP", "OCT", "NOV", "DEC"))))  
#View(dodgers)  
glimpse(dodgers)

## Observations: 81  
## Variables: 12  
## $ month <fct> APR, APR, APR, APR, APR, APR, APR, APR, APR, APR, AP…  
## $ day <dbl> 10, 11, 12, 13, 14, 15, 23, 24, 25, 27, 28, 29, 7, 8…  
## $ attend <dbl> 56000, 29729, 28328, 31601, 46549, 38359, 26376, 440…  
## $ day\_of\_week <chr> "Tuesday", "Wednesday", "Thursday", "Friday", "Satur…  
## $ opponent <chr> "Pirates", "Pirates", "Pirates", "Padres", "Padres",…  
## $ temp <dbl> 67, 58, 57, 54, 57, 65, 60, 63, 64, 66, 71, 74, 67, …  
## $ skies <chr> "Clear", "Cloudy", "Cloudy", "Cloudy", "Cloudy", "Cl…  
## $ day\_night <chr> "Day", "Night", "Night", "Night", "Night", "Day", "N…  
## $ cap <chr> "NO", "NO", "NO", "NO", "NO", "NO", "NO", "NO", "NO"…  
## $ shirt <chr> "NO", "NO", "NO", "NO", "NO", "NO", "NO", "NO", "NO"…  
## $ fireworks <chr> "NO", "NO", "NO", "YES", "NO", "NO", "NO", "NO", "NO…  
## $ bobblehead <chr> "NO", "NO", "NO", "NO", "NO", "NO", "NO", "NO", "NO"…

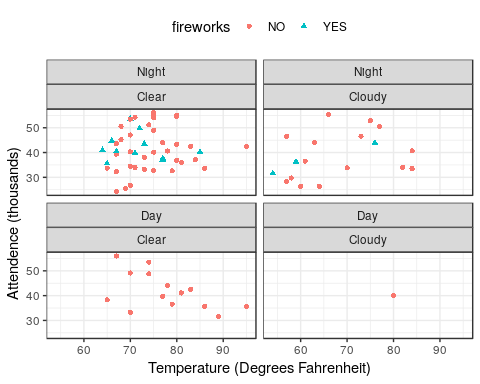
use R to develop two side-by-side boxplots, as seen in Figure 2.1 and Figure 2.2

#need to transform full name days to abbreviated; ie "monday" to "mon"  
dodgers <- dodgers %>% mutate("Attend/1000" = attend/1000,  
 day\_of\_week = case\_when(day\_of\_week == "Monday" ~ "Mon",  
 day\_of\_week == "Tuesday" ~ "Tue",  
 day\_of\_week == "Wednesday" ~ "Wed",  
 day\_of\_week == "Thursday" ~ "Thur",  
 day\_of\_week == "Friday" ~ "Fri",  
 day\_of\_week == "Saturday" ~ "Sat",  
 day\_of\_week == "Sunday" ~ "Sun"),  
 #and need to assign them factor levels  
 day\_of\_week\_f = factor(day\_of\_week, levels = c("Mon", "Tue", "Wed",  
 "Thur", "Fri", "Sat", "Sun")),  
 day\_night\_f = factor(day\_night, levels = c("Night", "Day")))  
  
# boxplots with ggplot using the dodger data  
## day of week on X  
## attendence in K's on Y  
dow\_box <- dodgers %>%   
 ggplot(aes(x = day\_of\_week\_f, y = (attend/1000))) +  
 geom\_boxplot(fill = "pink") +  
 theme\_light() +  
 theme(legend.position="none") +  
 labs(y = "Attendence (1000)",  
 x = "Day of the Week")  
## month on X  
## attendence in K's on Y  
mon\_box <- dodgers %>%   
 ggplot(aes(x = month, y = (attend/1000))) +  
 geom\_boxplot(fill = "dodgerblue1") +  
 theme\_light() +  
 theme(legend.position="none") +  
 labs(y = "", x = "Month of Year")  
  
#put these side by side  
cowplot::plot\_grid(dow\_box, mon\_box)



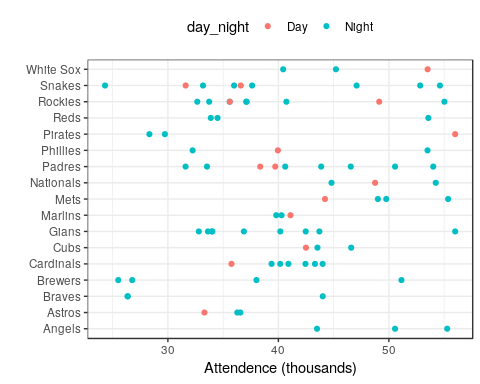
and two scatter plots, as seen in Figure 2.3

#we're analyzing days with fireworks against day / night, clear / cloudy, AND temperature  
#fireworks by shape and color  
#attendence (K's) on Y  
#temp on X  
dodgers %>%   
 ggplot(aes(x = temp,   
 y = attend/1000,  
 shape = fireworks,  
 color = fireworks)) +  
 geom\_point() +  
 theme\_bw() +  
 theme(legend.position="top") +  
 labs(y = "Attendence (thousands)",  
 x = "Temperature (Degrees Fahrenheit)") +  
 facet\_wrap(day\_night\_f ~ skies)



and Figure 2.4

# scatterplot teams against attendence   
# y axis is opposing teams  
# x axis is attendence (K's)  
# fill by day/night  
dodgers %>%   
 ggplot(aes(x = (attend/1000), y = opponent)) +   
 geom\_point(aes(color = day\_night)) +  
 theme\_bw() +  
 theme(legend.position="top") +  
 labs(y = "",  
 x = "Attendence (thousands)")



Use R to develop a Regression Model Performance, as seen in Figure 2.5. Examine regression diagnostics for the fitted model. Hint: Look at the Exhibit 2.1 in page 25-27.

#dodgers\_lm <-