Model Building Homework

We will use the hflights dataset for this homework.

1. (2 pts) Create a dataset from hflights that has the Date, Day of the Week, and the number of flights for that date. Visualize the data both as a line graph (# of flights vs Date) and as a boxplot for each day of the week.
2. (1 pt) Construct a model using day of the week as the predictor. What does this model tell us? Visualize the residuals.
3. (2 pts) Add a variable to account for seasonal variation. You can adjust the breaks something like this (feel free to change the dates)

season <- function(date) {

cut(date,

breaks = ymd(20110101, 20110301, 20110605, 201130905, 20120101),

labels = c("winter","spring", "summer", "fall")

)

}

mod1 <- lm(n ~ wday \* season, data = daily)

daily\_res <- daily %>%

add\_residuals(mod1, "resid")

1. (3 1/2 pts)
   1. (1/2 pt) Identify the dates with the largest residual values. What do you think is the cause for the days with the highest and lowest residual values?
   2. (1 pt) Add a variable to identify dates fitting this criterion.
   3. (2 pts) Build a model to explain your data using the variables you now have. Visualize the residuals.
2. (1 1/2 pts) Use what you have learned above to predict the number of flights for 2020 per day. Print a graph that overlays the number of flights in 2011 with your number of predicted flights in 2020. How many flights do you predict for each day June 20 - July 10 of 2020?