hw05\_DevinAdams

Filtering and Arranging

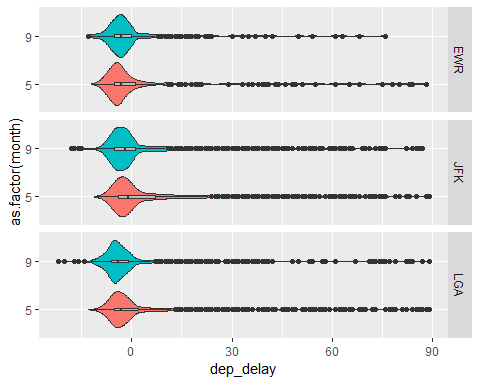
## Concepts

1. Explain the differences among “=”, “==”, and “<-”. Give an example of how you would use each of these symbols.  
   “=” is used as for assignment similar to the assignment operator “<-”. However, “=” is used also as an operator that specifies the parameter for a function. “==” is used to compare to values to find out if they are equal to each other.  
   Examples: - x <- 1 means that x has the value of 1  
   - month == 9 checks to see if the month is equal to 9  
   - x = 1 assignms the value of 1 to x - lm(formula=y~x)
2. Compare and contrast the logical operators “x & y” and “xor(x, y).  
   ”x & y" and “xor(x, y)” are both boolean logical operatores. “x & y” is true for values that satisfy both x and y, but not just x or just y. “xor(x, y)” is true for values that are only exclusively x or y and is false for values that satisfy x and y. Essentially, these two logical operators are opposites.

## Interpretation of code

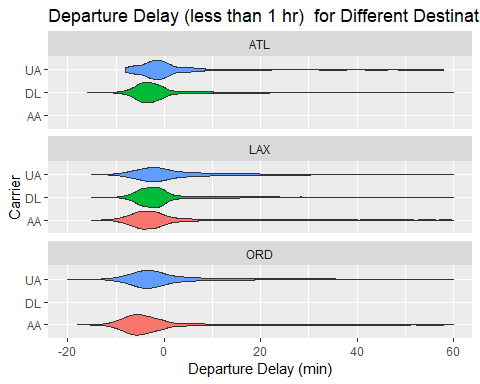
Interpret the following R code:

delta\_spring\_fall <- filter(flights, month == 5 | month == 9, carrier == "DL")  
ggplot(data = filter(delta\_spring\_fall, dep\_delay < 90),  
 mapping = aes(x = as.factor(month),  
 y = dep\_delay,  
 fill = as.factor(month))) +  
 geom\_violin(show.legend = FALSE) +  
 geom\_boxplot(width = 0.05, fill = "grey", show.legend = FALSE) +  
 coord\_flip() +  
 facet\_grid(rows = vars(origin))

 Line 31: Creates a variable “delta\_spring\_fall” and assigns to it all the flights that occurred in May and September with the carrier DL.  
Line 32: Creates a ggplot with the data as the flights by Delta in May and September for flights that departed with a delay of less than 90 minutes.  
Line 33: Maps the month as a factor to the x axis  
Line 34: Maps the departure delay time to the y axis  
Line 35: Maps the month to the fill color  
Line 36: Creates a violin plot of the filtered data “delta\_spring\_fall” and hides the legend.  
Line 37: Creates a boxplot of the filtered data “delta\_spring\_fall” and sets the width of the box to 0.05, the fill color is superceded to grey, and the legend is hidden.  
Line 38: Flips the x and the y axes  
Line 39: Breaks the plots into separate facets based on the airport of origin

## Coding Problem 1: Use violin plots and facets to explore the distribution of departure delays for the American, Delta, and United (AA, DL, and UA) for flights going to Atlanta, Chicago, and Los Angeles (ATL, ORD, and LAX).

delays\_to\_west <- filter(flights, carrier %in% c("AA","DL","UA"), dest %in% c("ATL","ORD","LAX"), dep\_delay <=60)  
 ggplot(delays\_to\_west, mapping = aes(x = carrier, y = dep\_delay, fill = carrier))+  
 geom\_violin(show.legend = FALSE)+  
 coord\_flip()+  
 facet\_wrap(~dest, nrow = 3) +  
 labs(title = "Departure Delay (less than 1 hr) for Different Destinations for Each Carrier",  
 y = "Departure Delay (min)",  
 x = "Carrier")



## Coding Problem 2: Air time and distance travelled are undoubtedly positively related. Investigate this relationship for all April flights from JFK.

* Use jittering and transparency to help with the overplotting.
* What might explain the clusters of points in the scatterplot?  
  The clusters of points are most likely due to the fact that the same route is being flown multiple times during the month of April. There is going to be a distribution of times for just that individual route due to differences in pilot, weather, etc. which is why some times the time is greater and some times the time is less.
* Is the relationship linear?  
  The relationship appears to be linear, as shown by the blue line that is a linear fit to the data. This makes sense since I imagine that most planes travel at roughly the same speed and have the same take off & landing durations.

april\_flights <- filter(flights, month == 4, origin == "JFK")  
ggplot(april\_flights, mapping = aes(x = distance, y = air\_time))+  
geom\_point(position = "jitter", alpha = 0.3)+  
 geom\_smooth(method = "lm")+  
 labs(title = "Air Time vs Distance for April Flights from JFK",  
 x = "Distance (miles)",  
 y = "Air Time (min)")

