Homework #5

Filtering and Arranging - Derrik Adams

## Concepts

1. Explain the differences among “=”, “==”, and “<-”. Give an example of how you would use each of these symbols.
2. Compare and contrast the logical operators “x & y” and "xor(x, y).

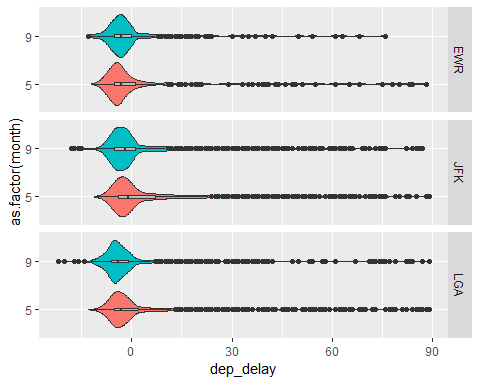
## Concept Answers

1. “=” is an assignment operator often used to pass values into functions and for assigning values. The “==” is used to compare/show equality between values and not assign values to anything. “<-” is the assignment operator where we assign values or something to a variable. Examples:  
   *ggplot(data, mapping = aes(x = xdata, y = ydata)) or x = 3 could work, but not great* x <- filter(data, x==3) \*x <- 3
2. “x & y” will only select the data that satisfies both condition x and condition y. It is essentially the intersection portion of a venn diagram. “xor(x,y)” basically means the opposite, so it selects the data that satisfies either just x or just y, but not the data that satisfies both x and y.

## Interpretation of code

Interpret the following R code:

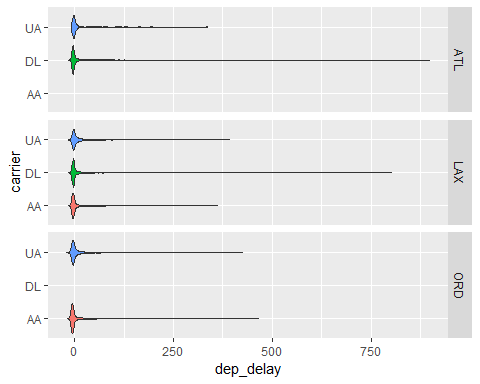
delta\_spring\_fall <- filter(flights, month == 5 | month == 9, carrier == "DL")  
ggplot(data = filter(delta\_spring\_fall, dep\_delay < 90),  
 mapping = aes(x = as.factor(month),  
 y = dep\_delay,  
 fill = as.factor(month))) +  
 geom\_violin(show.legend = FALSE) +  
 geom\_boxplot(width = 0.05,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

 ### Code Line Interpretation

1. Flight data is filtered to show all flight information for flights in May (spring) and September (fall) for Delta
2. Create a plot using the filtered the data where the departure delay is less than 90 min
3. Map the month on the x-axis aesthetic
4. Map the departure delay to the y-axis aesthetic
5. Make the fill aesthetic based on the month, so the color of each violin will be based on months
6. Create a violin plot using the previous mapping and do not show the legend
7. Create a boxplot within the violin plot with width .05, colored grey, and do not show the legend
8. Flip the coordinates so that the y- and x-axis are opposite the typical cartesian format
9. Facet the plots in a grid where the rows are the flight 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).

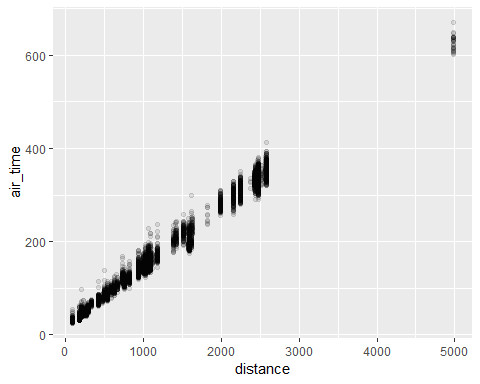
dep\_delay\_filt <- filter(flights, carrier %in% c("AA", "DL", "UA"), dest %in% c("ATL","ORD","LAX"))  
ggplot(data = dep\_delay\_filt, mapping = aes(x = carrier, y = dep\_delay, fill = carrier))+  
 geom\_violin(show.legend = FALSE)+  
 coord\_flip()+  
 facet\_grid(rows = vars(dest))



## 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?
* Is the relationship linear?

JFK\_Apr <- filter(flights, month == 4, origin == "JFK")  
ggplot(data = JFK\_Apr, mapping = aes(x = distance, y = air\_time))+  
 geom\_point(alpha = .1, position = "jitter")



The clusters of points are due to the flight distances being the same since flights from JFK go to the same airports multiple times. The time taken to get there though changes, meaning you get a small variation in time taken to get to the same airport. The transparent dots show that the extremely different times don’t happen often though.  
The relationship is definitely linear, you can tell just by looking at it.