LC2.1
Compared to the Alaska_flights, there are far more data entries for the flights. Unlike the Alaska_flights data, where the data entries are from 3 separate months, the flight entries are only based on the month of January.

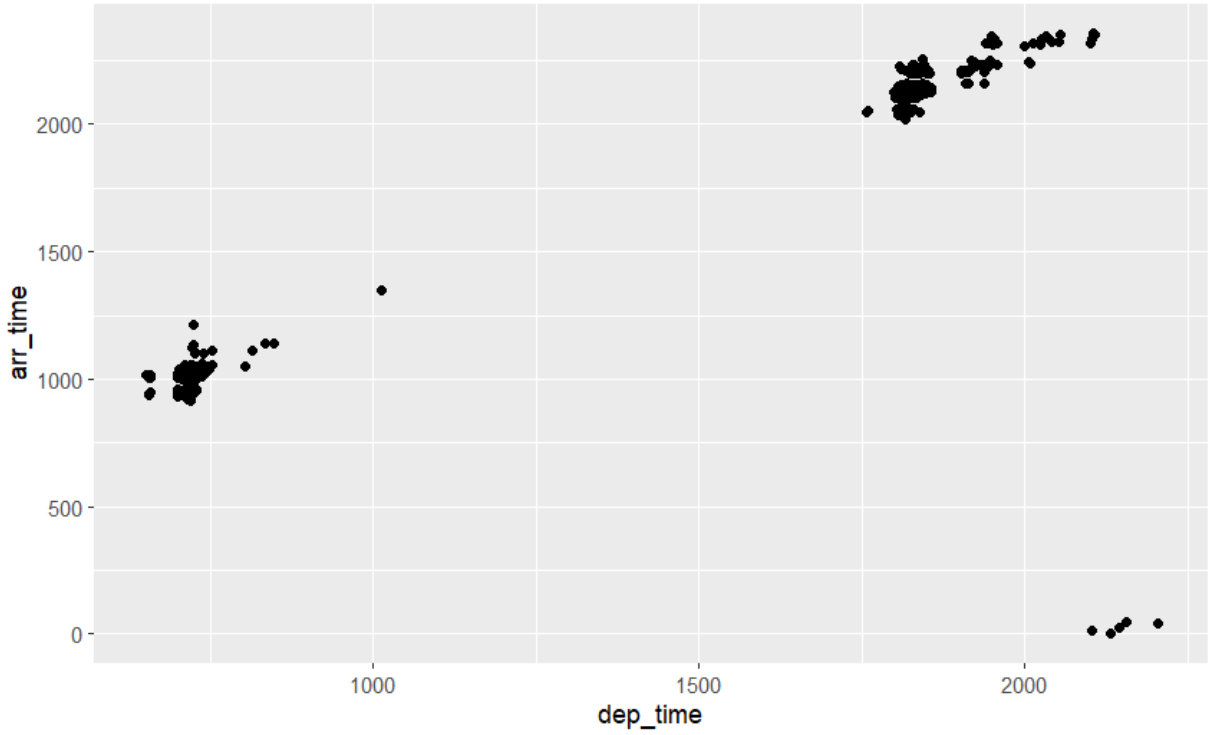
LC2.2
Because of scheduling and air traffic protocol requirements the connection, between departure delays (dep_delay) and arrival delays (arr_delay) is advantageous. Each flight must follow a predetermined schedule. Adhere to landing and takeoff procedures. These processes cannot be. Require time. Consequently, if a plane experiences a delay, during takeoff or landing it is highly probable that it will also arrive later at its intended destination.

LC2.3
Dep_delay would be negatively correlated with visibility. This is due to the possibility that the dep_delay will grow as the visibility in miles diminishes.

LC2.4
Due to the fact that they are the planes that are arriving and departing on schedule, I think there is a cluster of points close to (0,0). All of these points indicate flights that experienced minimal to no delays.

LC2.5
The fact that most airplanes take off and land in under an hour is one reason.

LC2.6



LC2.7
It helps in the overplotting or point clustering in scatterplots, making it valuable. Additionally, it aids in the allocation of points.

LC2.8
According to figure 2.4, arrival delays often range from on time to 12 minutes late, whereas departure delays typically occur between 50 minutes early and on time. Although the grid is larger, we can see that the density is greater there even if I don't think the plot altered all that much.

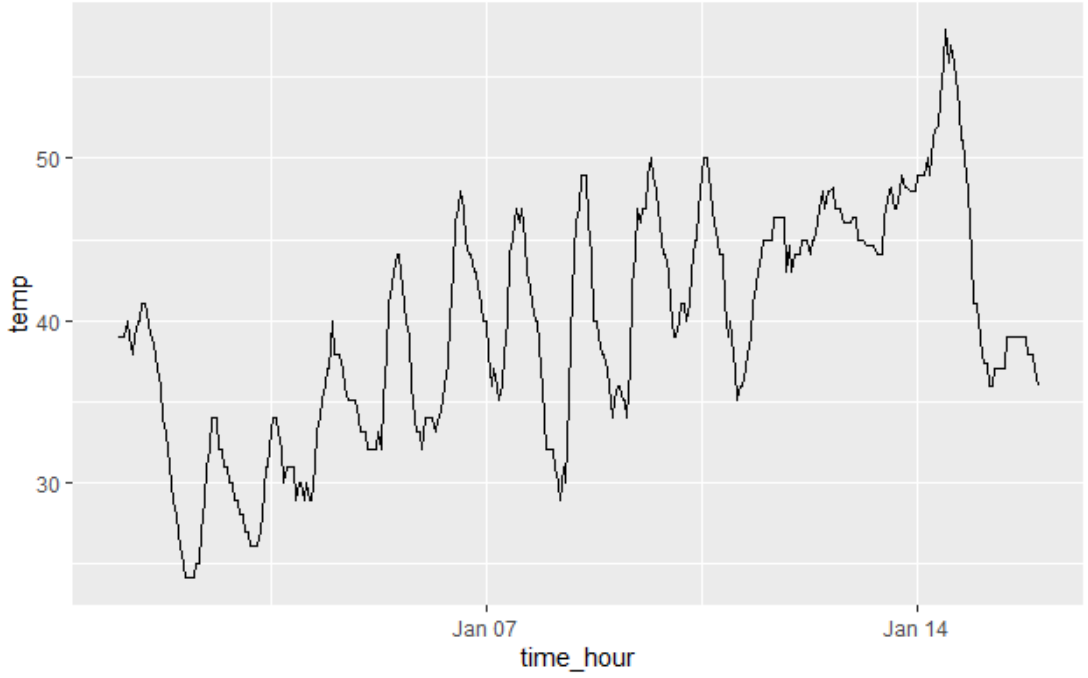
LC2.9
The fact that early_january_weather is only a subset of weather, the weather data frame has significantly more items.

LC2.10
A more precise time is available in the time_hour variable. which, as opposed to the variable hour, which only displays the number of hours, indicates the date and time of the trip.

LC2.11
This is so that the data cannot be ordered without the horizontal axis, which would only result in an erroneous depiction of the data.

LC2.12
This is due to the fact that since time is sequential, utilising a line graph with the time variable as the x-axis might help to depict the change over time.

LC2.13



LC2.14
The graph's distribution is a normal distribution, and it is clear that the majority of the temperatures are concentrated in the graph's centre.

LC2.15
The distribution of temperatures appears to be more symmetrical than not, in my opinion.

LC2.16
I believe that 55 is the distribution's centre value. This is due to the fact that it seems to be the middle number.

LC2.17
The data appears to be quite close. This is due to the fact that, after analysing the data, the majority of the points are sort of centred between the temps of 35 and 80.

LC2.18
We may add more variables to the study in a faceted plot, as is the case when we examine the variable temperature across time.

LC2.19
The number 1 through 12 represents the months of the year since we are examining the temperature across time. The temperatures for the moths are represented by the digits 25, 50, 75, and 100.

LC2.20
For variables that would have a lot of possible values, faceted charts would not be useful. especially if each value is distinct. As a result, a subset would have to be made for each of the distinctive values.

LC2.21
Given that you can observe the many seasons in accordance with the yearly temperatures, I'd say it has.

LC2.22
This outliner, in my opinion, might have resulted from a data entering error.

LC2.23
The biggest temperature fluctuation was between November and January. These boxes, which are the tallest, show that this month's IQR is the highest.

LC2.24
Similar to faceted plots, a boxplot of a variable with more than 400 entries, such as pressure, would not be practical since too many charts would be generated, which would be impossible to comprehend or read.

LC2.25
If there is an outliner in a boxplot, it is obvious.

LC2.26
Because histograms concentrate on assessing numerical variables, they cannot clearly depict categorical data.

LC2.27
As was already said, histograms emphasise numerical variables whereas barplots would primarily emphasise data that is more appropriately categorised.
LC2.28
Almost 26000 flights

LC2.29
A US airline Using a bar plot, this may be sorted by these.

LC2.30
This is so because pie charts are dependent on percentages, which are challenging for people to understand from a simple visual representation. This would result in the pie charts being interpreted incorrectly.

LC2.31
They still utilize these, in my opinion, since creating them is common knowledge.

LC2.32
I would argue it is difficult to compare the performance of different carriers over airports since they are not aligned.

LC2.33
Most, if not all, of the departures for some airlines come from particular airports. JetBlue is a JFK carrier, for instance.

LC2.34
Aligning these would make it simpler to compare the bars to one another.

LC2.35
Dodge bagplots have limits when processing complicated data, such as negative values and those data sets that could have too many subcategories, which is one of their drawbacks.

LC2.36
This is due to the fact that by just looking at the charts, it is much simpler to compare the various categories.

LC2.37
We can quickly determine which airline serves each airport.