This is an individual milestone for the project. You will use your group project data to create visualizations yourself. They can be used in the project but they don’t have to be.

1) Create a visualization using one of the techniques from the latter half of the class (after the midterm). For example, from Week 7 you could use one of the techniques from the Categorical unit, like a mosaic plot or Bertin matrix, or you could try applying an interactivity example to your project data. Week 8 offers uncertainty visualizations and contours and 2D binning to apply to numerical variable relationships. Despite being covered earlier, a cartogram is also allowed. Each group member’s visualizations must be distinct. The group’s visualizations can use the same technique if they cover different aspects of the data or use the technique in different ways.

I’d like to know if ,the more flights of an airline, the more flight delays could happen.
From the following graphs, we could know that the correlation between the number of flights and the number of delayed flight in a day is positive,which means that if an airline operate more flights in the a day , the more number of delayed flights could be there

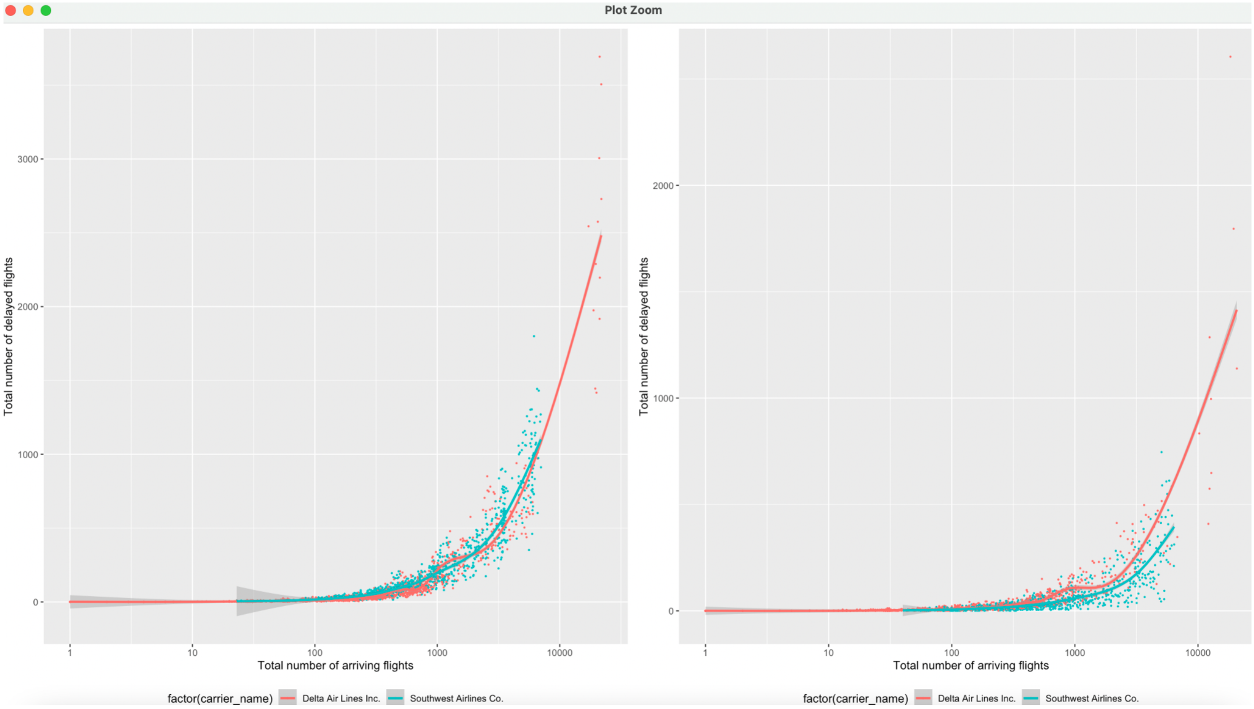
Carrier_name : Airline
Arr_del15 : Total number of delayed flights in the observation
arr_flights:Total number of arriving flights in the observation

Curve Fit

```
Time_2019<-Airline_Delay_Cause._2019 %>%
          select(year,month, arr_flights, arr_del15,carrier_name) %>% drop_na() %>%
          group_by(month) %>%
          filter(carrier_name == as.character("Delta Air Lines Inc.") |carrier_name ==
as.character("Southwest Airlines Co."))%>%
           ggplot( aes(x=arr_flights, y=arr_del15, colour = factor(carrier_name) ))+
            geom_point(size= .3) +theme(legend.position='bottom')+labs(y = "Total number of
delayed flights", x = "Total number of arriving flights",fill ="Airline")+
          scale_y_continuous()  + scale_x_continuous(trans = "log10") +geom_smooth()
```

```
Time_2020<-Airline_Delay_Cause._2020 %>%
           select(year,month, arr_flights, arr_del15,carrier_name) %>% drop_na() %>%
            group_by(month) %>%
           filter(carrier_name == as.character("Delta Air Lines Inc.") |carrier_name ==
as.character("Southwest Airlines Co."))%>%
           ggplot( aes(x=arr_flights, y=arr_del15, colour = factor(carrier_name) ))+
          geom_point(size= .3) +theme(legend.position='bottom')+labs(y = "Total number of
          delayed flights", x = "Total number of arriving flights",fill ="Airline")+
        scale_y_continuous()  + scale_x_continuous(trans = "log10") +geom_smooth()

grid.arrange(Time_2019,Time_2020,ncol=2)
```



2) Do the same as in item 1 but for another type of visualization. For this, you may use visualizations going back to Week 4, which includes geographical, statistical and special time series plots (e.g. tile plots or line graphs with smoothing). You may use the same type of visualization as item 1 if the two cover different aspects of the data or use the technique in different ways.

Assuming that I was the person who is going to buy a flight ticket for a business trip in 2021, since punctuality is important for a business trip, I would try to decide which airline to buy the flight ticket by comparing the airline performance in 2019 and 2020.

I create a Heatmap by tile plot,using the Percentage of delay time of each airline as the factor of fill color and I choose the top 5 busiest airports to make the comparison.

From the heatmap below, I could know that Delta Airline &Southwest Airline have relatively good performance on punctuality in 2019, on the contrast, ExpressJet and Skywest have relatively good performance.

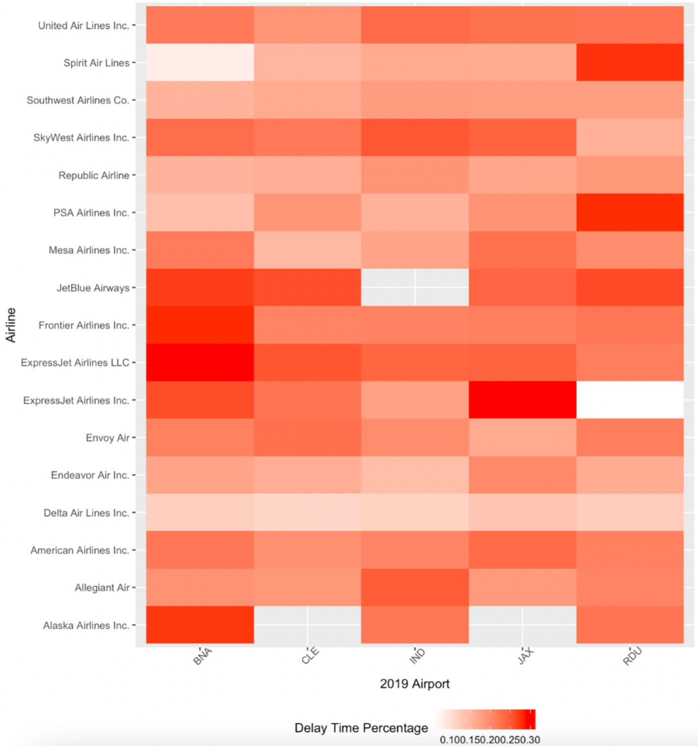
In 2020, the percentage of delay time of All of the Airline has decreased a lot, the reason for that maybe is because fewer people took the flights, the fewer passengers in each flight, so the

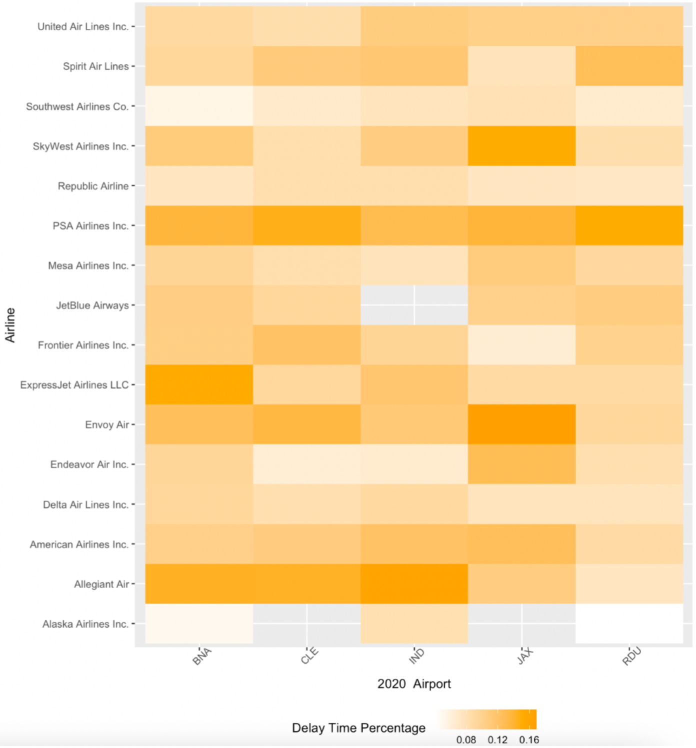
flight could take off more punctually. Under this condition, Delta Airline &Southwest Airline still performed relatively well.
In conclusion, if I were a person who try to buy the flight ticket, Delta Airline &Southwest Airline would be my top 2 choice.

Carrier_name : Airline
Arr_del15 : Total number of delayed flights in the observation
arr_flights:Total number of arriving flights in the observation

```
p1<-Airline_Delay_Cause._2019 %>%
          select(carrier_name, airport, arr_del15, arr_flights) %>% drop_na() %>%
           filter(airport == as.character("RDU") |airport == as.character("JAX")|airport ==
          as.character("BNA")|airport == as.character("CLE")|airport == as.character("IND"))%>%
           group_by(airport, carrier_name) %>% dplyr::summarize_all(funs(sum)) %>%
           mutate(del_pct = arr_del15/arr_flights) %>%
           ggplot(aes(x=factor(airport), y= factor(carrier_name), fill=del_pct),na.rm=TRUE) +
          geom_tile() +
           theme(axis.text.x=element_text(angle=45)) +
           scale_fill_gradient(low = "white", high = "red")+labs(y = "Airline", x = "2019 Airport",fill
          ="Delay Time Percentage")+theme(legend.position = "bottom")


p2<-Airline_Delay_Cause._2020 %>%
          select(carrier_name, airport, arr_del15, arr_flights) %>% drop_na() %>%
            filter(airport == as.character("RDU") |airport == as.character("JAX")|airport ==
          as.character("BNA")|airport == as.character("CLE")|airport == as.character("IND"))%>%
            group_by(airport, carrier_name) %>% dplyr::summarize_all(funs(sum)) %>%
            mutate(del_pct = arr_del15/arr_flights) %>%
            ggplot(aes(x=factor(airport), y= factor(carrier_name), fill=del_pct),na.rm=TRUE) +
          geom_tile() +
            theme(axis.text.x=element_text(angle=45)) +
            scale_fill_gradient(low = "white", high = "orange")+labs(y = "2020 Airline", x =
          "Airport",fill ="Delay Time Percentage ") +theme(legend.position = "bottom")
grid.arrange(p1,p2,ncol=2)
```


Airline
United Air Lines Inc.
Spirit Air Lines
Southwest Airlines Co.
SkyWest Airlines Inc.
Republic Airline
PSA Airlines Inc.
Mesa Airlines Inc.
JetBlue Airways
Frontier Airlines Inc.
ExpressJet Airlines LLC
ExpressJet Airlines Inc.
Envoy Air
Endeavor Air Inc.
Delta Air Lines Inc.
American Airlines Inc.
Allegiant Air
Alaska Airlines Inc.
BNA
CLE
IND
JAX
RDU
2019 Airport
Delay Time Percentage
0.10
0.15
0.20
0.25
0.30


Airline
United Air Lines Inc.
Spirit Air Lines
Southwest Airlines Co.
SkyWest Airlines Inc.
Republic Airline
PSA Airlines Inc.
Mesa Airlines Inc.
JetBlue Airways
Frontier Airlines Inc.
ExpressJet Airlines LLC
Envoy Air
Endeavor Air Inc.
Delta Air Lines Inc.
American Airlines Inc.
Allegiant Air
Alaska Airlines Inc.
BNA
CLE
IND
JAX
RDU
2020 Airport
Delay Time Percentage
0.08
0.12
0.16