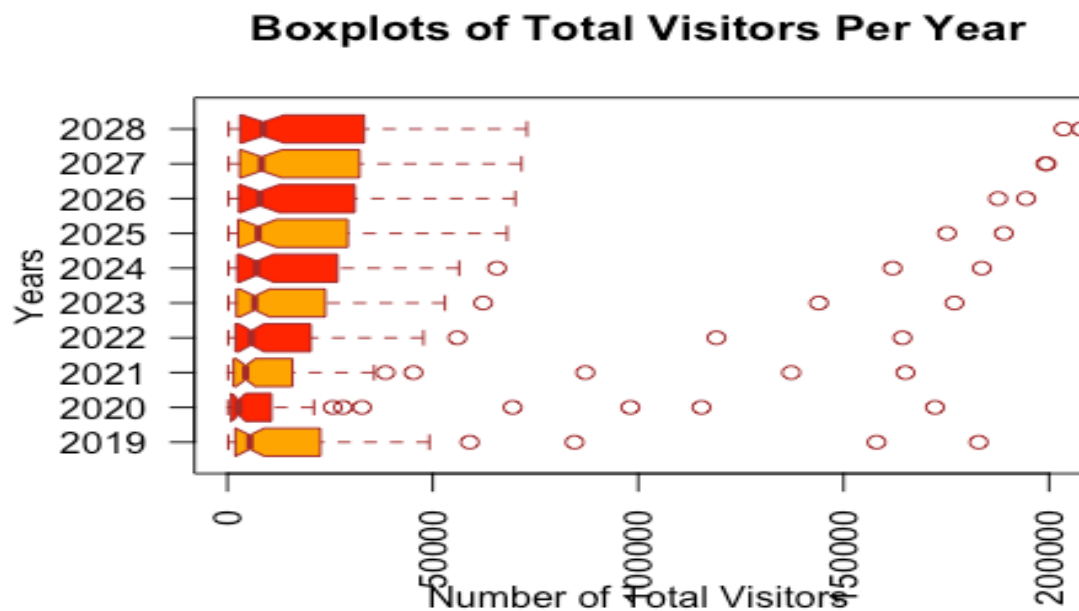# Global World Cities

Aaron Angeles

1/7/2021

After doing an initial eye-test of the dataset and brainstorming potential areas of interest, I thought the best first step would be to sort each city into its corresponding country, combine the number of domestic and international visitors.

**Boxplots of Total Visitors Per Year**

You can see that

- there exists an upward trend of total visitors to all countries
    - **indicative that travel and tourism industry is growing after a steep decline in 2020 due to the COVID-19 pandemic.**
- an ancitipated rise from 2019 in the years preceding 2022
    - **the 75th percentile of each year seems to grow gradually to 2028**
    - **the mean of all the countries within the year seems to stay around the same range after about 2022**
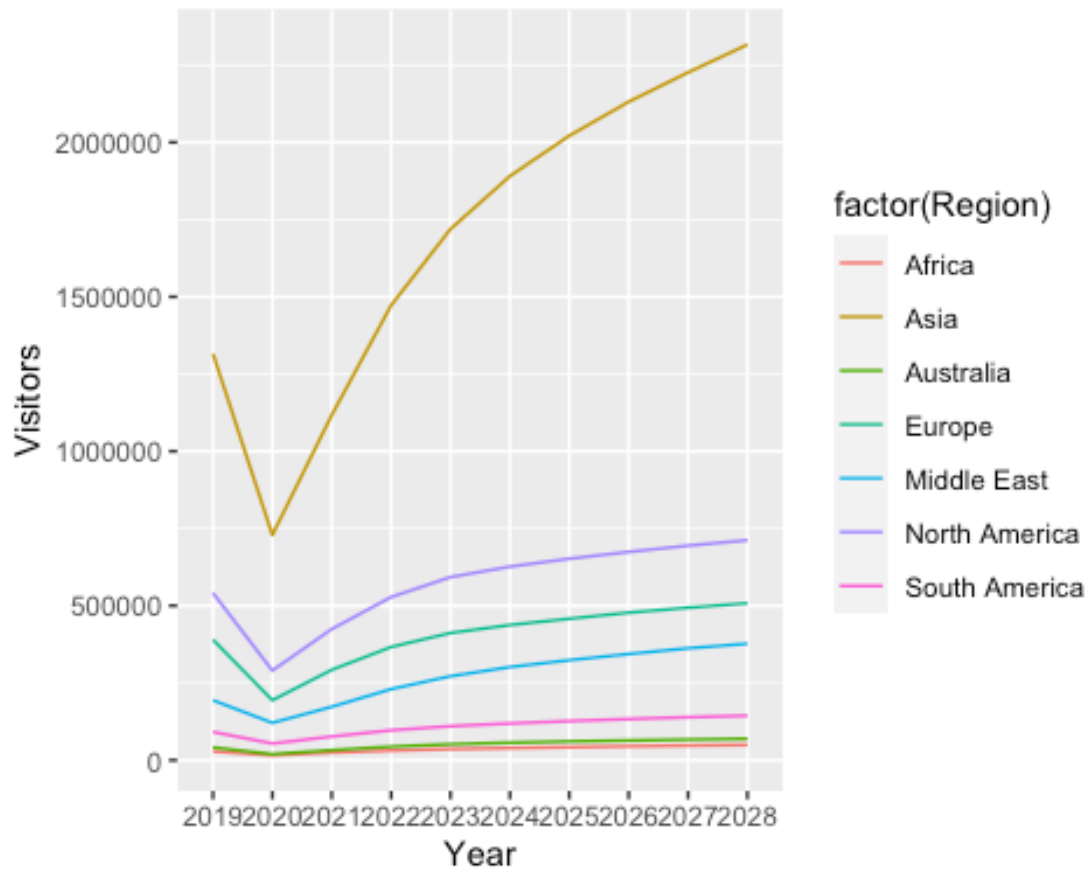
Sorting Countries into Regions, and Analyzing Number of Visitors from 2019-2028

South America
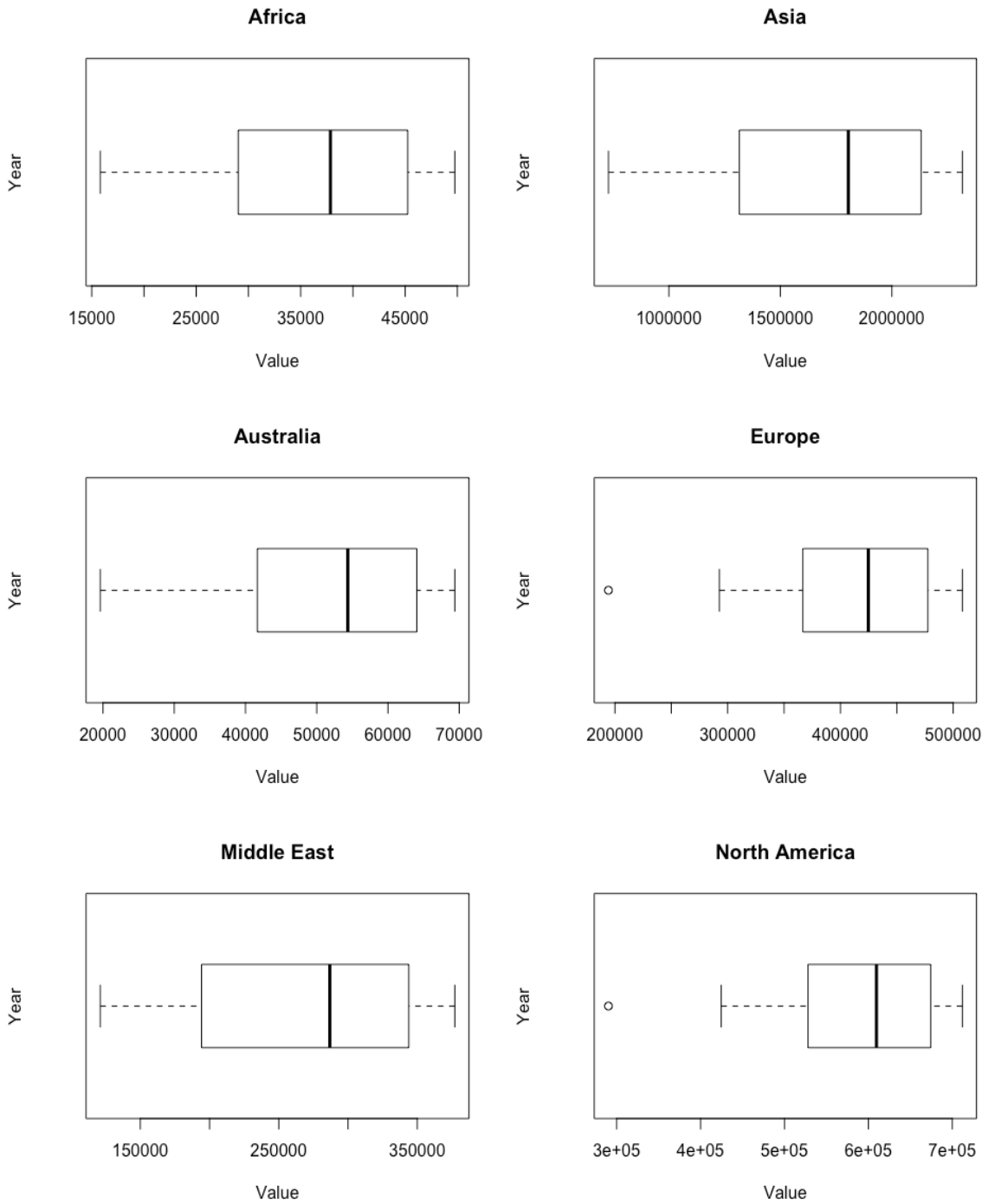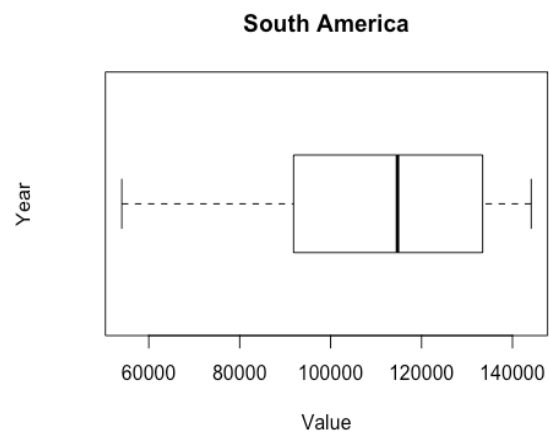
The various plots show the distribution of visitors over the course of the 2019-2028 range within each region.

- Value is the total number of visitors

- Each region has an instance of total number of visitors for each of the years within 2019-2028

The line-graph is showing the number of total visitors over the course of of the 2019-2028 range within each region.

- For most regions, there exists a sharp decrease from 2019 to 2020 as a result of the COVID-19 pandemic.

- After that point, most regions begin to steadily increase the number of total visitors. Asia appears to have the most visitors and the sharpest increase post-pandemic.

- *Australia, South America and Africa* don't appear to experience as sharp pandemic-related decreases, nor post-pandemic increases.

Finding Percent Increase from Year to Year per Region
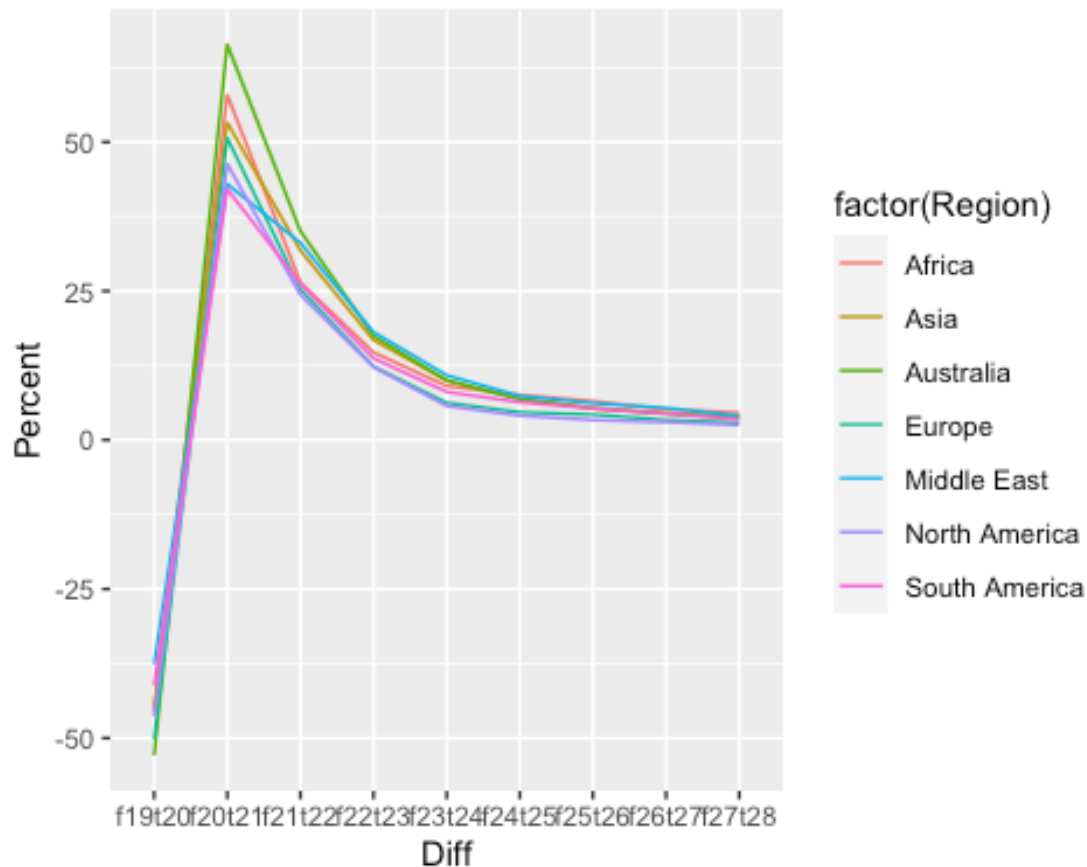
This figure is graphing the percent increase/decrease for each year-year increment.

- From 2019-2020, the percent decrease for most regions is around -50% to -25%, indicative of the effect that the COVID-19 pandemic has had on the tourism industry.

- The percent increase from year to year rapidly increases from 2020-2021 and continues to remain positive in each year-year increment.

- The curve does flatten out and the percent increase approaches 0 as the tourism numbers appear to stabilize.

Essentially, you can expect a surge in tourism around the world from 2020 to 2021, when we can expect the pandemic to end. This trend is consistent for almost all the regions included in the analysis.

However, some things to note are…

- Middle East appears to experience less steep of stabilization of tourism between 2021 and 2023, where other regions experience the same surge but their year to year percent increase stabilizes "quicker."
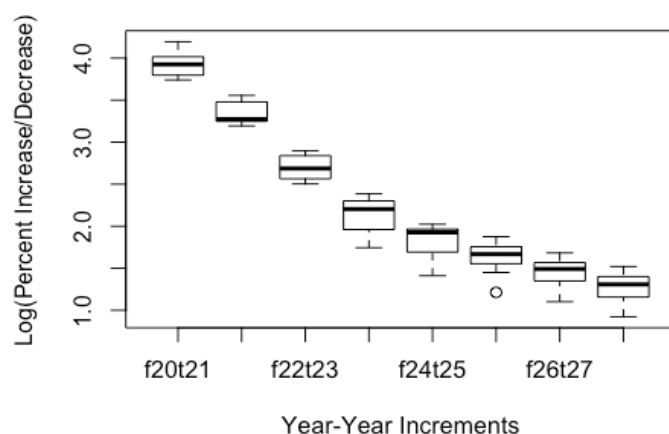
Modeling the Yera-Year Percent Increase Change with a Transformed SLR

ercent Increase from Year-Year Increments from 2020
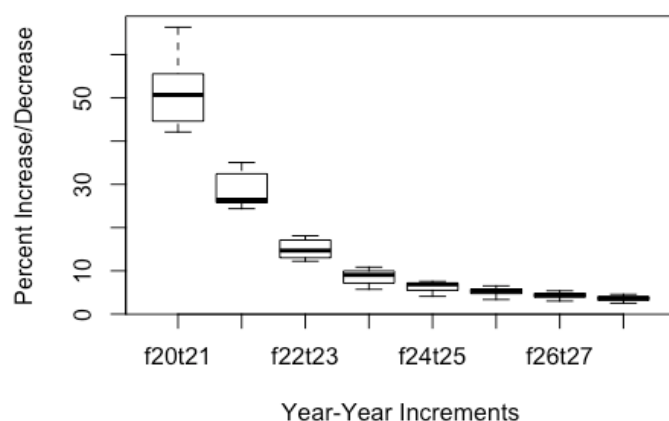
This figure is the percent increases in year-year increments from 2020-2028

ercent Increase from Year-Year Increments from 2020

This figure is the logarithmic function of the percent increases in year-year increments from 2020-2028.

```
##
## Call:
## lm(formula = log(longer_data_reg_mod$Percent) ~ longer_data_reg_mod$Diff)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
```

```
## -0.41205 -0.16231  0.04211  0.14872  0.26792
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      3.92690    0.07713   50.92  < 2e-16 ***
## longer_data_reg_mod$Difff21t22  -0.57263    0.10907   -5.25 3.42e-06 ***
## longer_data_reg_mod$Difff22t23  -1.22728    0.10907  -11.25 4.63e-15 ***
## longer_data_reg_mod$Difff23t24  -1.80452    0.10907  -16.54  < 2e-16 ***
## longer_data_reg_mod$Difff24t25  -2.11520    0.10907  -19.39  < 2e-16 ***
## longer_data_reg_mod$Difff25t26  -2.30161    0.10907  -21.10  < 2e-16 ***
## longer_data_reg_mod$Difff26t27  -2.48249    0.10907  -22.76  < 2e-16 ***
## longer_data_reg_mod$Difff27t28  -2.66002    0.10907  -24.39  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2041 on 48 degrees of freedom
## Multiple R-squared:  0.9575, Adjusted R-squared:  0.9514
## F-statistic: 154.7 on 7 and 48 DF,  p-value: < 2.2e-16
```
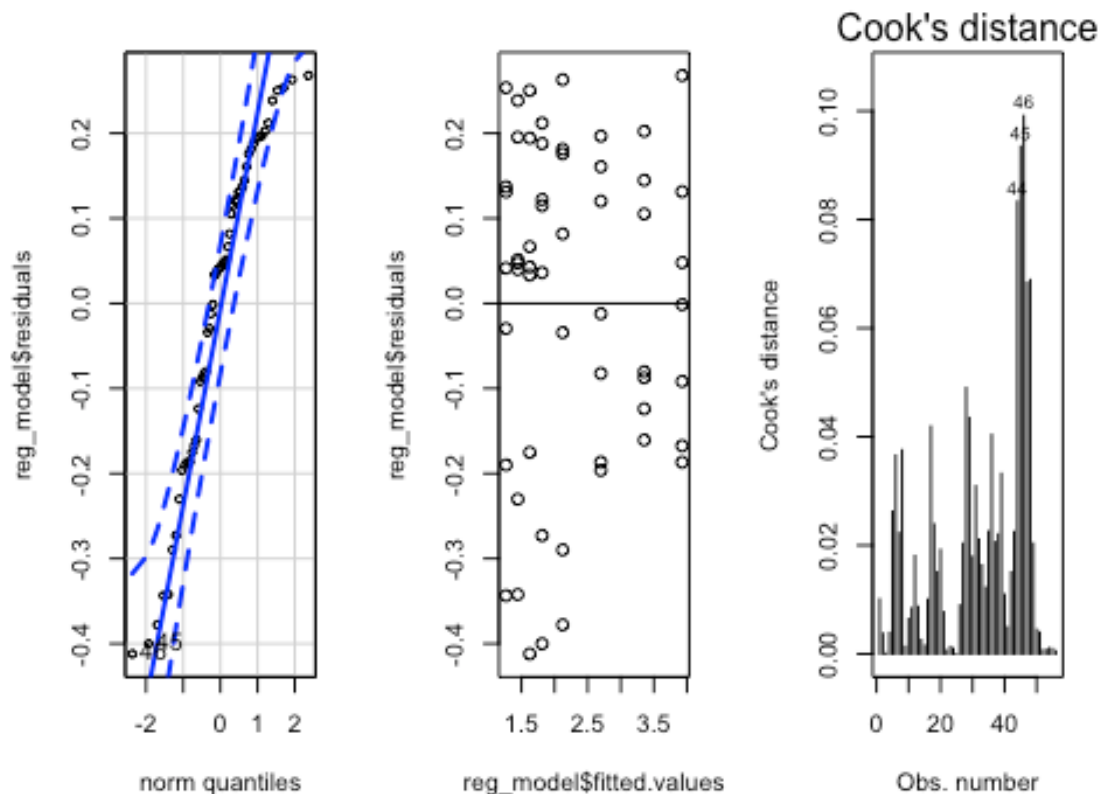

This model is another piece of strong evidence to indicate the percent change between year-year trends in the tourism industry.

From the initial line graph...

- The year-year percent change seemed to stabilize closer to 0 after a huge spike at the completion of 2020/2021 (and the COVID-19) pandemic.

- The intial graph of year-year percent change and percent increase took the shape of a logarithmic curve.

After transformation of the year-year percent changes to become

- log(Percent Change) ~ Year-Year Difference

  - graph appears to linearize.

- reported adjusted r^2 was 0.9514

- all year-year increments predictors significant in th predction of the log(Percent Change).

Given this information...

- we can expect the log(Percent Change) and the general percent change to stabilize towards 0 from 2028 and beyond

- cannot extrapolate beyond the bounds of our data set.

Additionally, all assumptions for SLR were met.

- Error points appear to follow a normal distribution

- There is no pattern within the residual plot, indicative that there is constant variance around the line.

- In our Cook's Distance, there are no outliers that appear to be influential points which are driving conclusions. No point has a cook's distance exceeding 0.5.
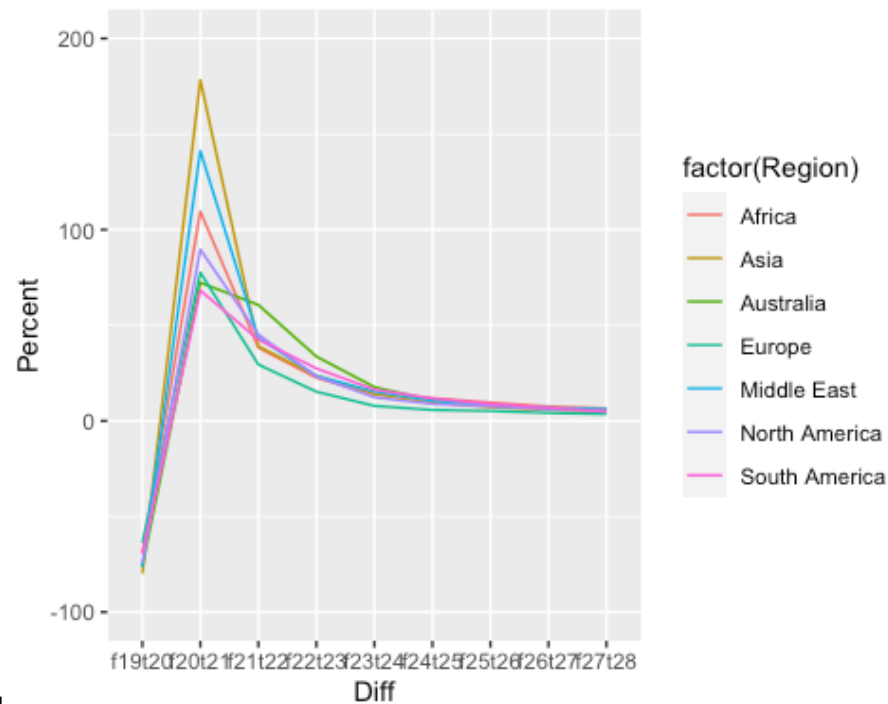
Analyzing Domestic v.s. International Travel Metrics

Finding Percent Increase for Each Country from Year-Year Increments and Comparing the Domestic and International Increase
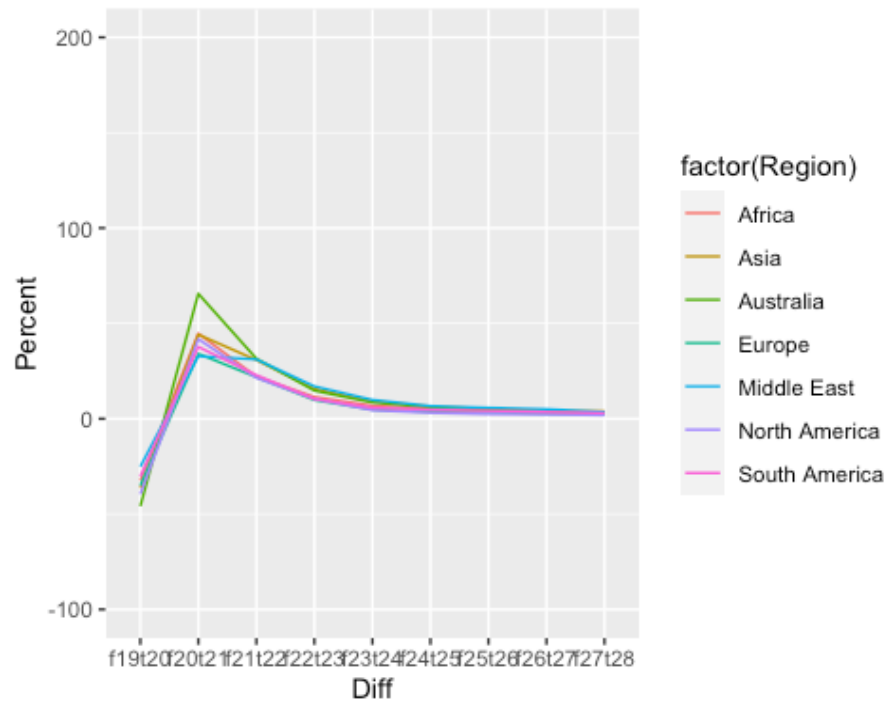
Domestic.

International

After writing some queries to sort the data into countries and then regions, I was able to compute a year-year percent increase increment like I did above. Instead of doing a full region summation, I left it to regional data per the varying indicators, domestic and international. In doing so, I wanted to compare the percent increase for domestic travel versus international travel, within each varying region and within the larger scope of tourism.

- The left-most figure is year-year percent increase of domestic overnight visits and the right-most figure is the year-year percent increase of international overnight visits.

- After the 2019-2020 interval with the percent decrease due to COVID-19, overall tourism increases from year to year, and the rate of increase stabilizes (displayed earlier).

From a comparison standpoint, (after the 2019-2020 COVID-19 affected interval)

- the expected international overnight travel experiences larger yearly percent increases after the pandemic compared to domestic overnight travel

  o could be a result of limited international travel abilities in 2020 due to pandemic restrictions, where domestic travel isn't as restricted.

- It appears that the yearly percent changes are consistent with the region, and the indicator discrepancy is present for almost all the regions.

  o Australia doesn't experience as drastic of a difference between domestic and international visits

  o The other regions all experience similar differences between domestic and international.

Summary Statistics & Insights

2019 Pre-Pandemic Metrics
- Cities in the 99th Percentile of Domestic Visitors in 2019 were **Mexico City (~54,608), Karachi (~58,099), Shanghai (~65,242) and Chongqing(~145,578)**

- Cities in the bottom 1 Percentile of Domestic Visitors in 2019 were **Malacca (~10), Dubrovnik (~57), Liege (~107) and Surat (~121).**

- Cities in the 99th Percentile for International Visitors in 2019 were **Dubai (~16,730), Bangkok (~22,523), Macao (~20,353) and Hong Kong (~18,922).**

- Cities in the bottom 1 Percentile for International Visitors in 2019 were **Surat (~2), Malacca (~19), Tianan (~39) and Leicester (~77).**

- The median number of domestic overnight visitors for all cities included within the database was 2,462. The minimum and maximum were ~10 (Malacca) and ~145,578 (Chongqing).

- The median number of international overnight visitors for all cities included within the database was ~836. The minimum and maximum were ~2 (Surat), and ~22,523 (Bangcok).

2028 Post-Pandemic Metrics

- Cities in the 99th Percentile of Domestic Visitors in 2028 were **Wuhan (~91,136), Karachi (~143,228),Shanghai (~125,674) and Chongqing(~288,924)**

- Cities in the bottom 1 Percentile of Domestic Visitors in 2028 were **Malacca (~18), Dubrovnik (~89), Liege (~129) and Reykjavik (~208).**

- Cities in the 99th Percentile for International Visitors in 2028 were **Dubai (~25,699), Bangcok (~27,116), Macao (~27,846) and Hong Kong (~32,457).**

- Cities in the bottom 1 Percentile for International Visitors in 2028 were **Surat (~3), Malacca (~37), Tianan (~50) and Leicester (~79).**

- The median number of domestic overnight visitors for all cities included within the database was 2,462. The minimum and maximum were ~10 (Malacca) and ~145,578 (Chongqing).

- The median number of international overnight visitors for all cities included within the database was ~1,246. The minimum and maximum were ~3 (Surat), and ~32,457 (Hong Kong).

2019 US Cities Metrics

- Cities in the 95th Percentile for Domestic Travel in 2019 were **Las Vegas (~22,947), New York City (~16,309) and Chicago (~14,070).**

- Cities in the bottom 5th Percentile for Domestic Travel in 2019 were **Providence (~290), El Paso (~434) and Buffalo (~1,337).**

- Cities in the 95th Percentile for International Travel in 2019 were **New York City (~13,523), Orlando (~6,344) and Los Angeles (~7,422)**

- Cities in the bottom 5th Percentile for International Travel in 2019 were **Providence (~108), Oklahoma City (~119) and Kansas City (~142)**.

- The median number of domestic travel in the US in 2019 was 3,803. The minimum and maximum domestic visitors was ~290 (Providence) and ~22,947 (Las Vegas.)

- The median number of domestic travel in the US in 2019 was 3,803. The minimum and maximum domestic visitors was ~290 (Providence) and ~22,947 (Las Vegas.)

2028 US Cities Metrics

- Cities in the 95th Percentile for Domestic Travel in 2028 were **Las Vegas (~31,050), New York City (~18,664) and Chicago (~17,465).**

- Cities in the bottom 5th Percentile for Domestic Travel in 2028 were **Providence (~375), El Paso (~577) and Buffalo (~1,794).**

- Cities in the 95th Percentile for International Travel in 2028 were **New York City (~18,713), Orlando (~7,836) and Los Angeles (~9,168)**

- Cities in the bottom 5th Percentile for International Travel in 2028 were **Providence (~137), Oklahoma City (~150) and Kansas City (~174).**

- The median number of domestic travel in the US in 2028 was ~5,224. The minimum and maximum domestic visitors was ~375 (Providence) and ~31,050 (Las Vegas.)

- The median number of domestic travel in the US in 2028 was 3,803. The minimum and maximum domestic visitors was ~137 (Providence) and ~18,713 (New York City).