

24-1 Data Science and R Final Project

Analysis of Incheon International Airport Transfers

Group1

Nam Hee Yeon

Son Chae Ri

Kyung Dong Heon

Contents



01 Introduction

02 Data Import & Pre-processing

03 Insights from Basic Investigation

04 Hypotheses and Verifications

05 Challenges

06 Conclusions



▶ Data Introduction : Arrival and Departure Transfer Dataset from Incheon International Airport

▶ Data Name

- Incheon International Airport's Transfer Passengers Information (January 2023 - June 2023)

▶ Data Source

- Incheon International Airport Corporation's Open Data Platform

▶ Data Size & Column Names

- Rows: About 61,000
- Columns: 8
- Column names:

Arrival/Departure Country, Airport,

Airport Code, Flight Date, Scheduled Time, Actual Time, Flight Name, Number of Transfer Passengers

▷ Arrival Dataset: Passenger's final destinations from Incheon Airport

| 도착국가 <chr> | 도착공항 <chr> | 공항코드 <chr> | 운항일자 <date> | 계획시간 <S3: hms> | 실제시간 <S3: hms> | 운항편명 <chr> | 환승승객 <dbl> |
|---------------|---------------|---------------|----------------|-------------------|-------------------|---------------|---------------|
| 베트남 | 두옹 당(푸퍽) | PQC | 2023-01-01 | 00:30:00 | 01:01:00 | VJ978 | 0 |
| 필리핀 | 마닐라 | MNL | 2023-01-01 | 04:10:00 | 03:36:00 | 7C2306 | 5 |
| 태국 | 방콕/수완나품 | BKK | 2023-01-01 | 04:20:00 | 04:05:00 | KE658 | 15 |
| 베트남 | 나트랑 | CXR | 2023-01-01 | 04:30:00 | 04:24:00 | VJ836 | 0 |
| 일본 | 도쿄/하네다 | HND | 2023-01-01 | 04:35:00 | 04:47:00 | MM809 | 0 |
| 필리핀 | 마닐라 | MNL | 2023-01-01 | 04:45:00 | 03:59:00 | KE624 | 99 |

▷ Departure Dataset: Where did the passengers leave before arriving at Incheon Airport?

| 출발국가 <chr> | 출발공항 <chr> | 공항코드 <chr> | 운항일자 <date> | 계획시간 <S3: hms> | 실제시간 <S3: hms> | 운항편명 <chr> | 환승승객 <dbl> |
|---------------|---------------|---------------|----------------|-------------------|-------------------|---------------|---------------|
| 터키 | 이스탄불 | IST | 2023-01-01 | 00:15:00 | 00:19:00 | TK091 | 0 |
| 카타르 | 도하 | DOH | 2023-01-01 | 00:25:00 | 00:40:00 | QR859 | 2 |
| 에티오피아 | 볼레(아디스아바바) | ADD | 2023-01-01 | 00:30:00 | 00:38:00 | ET673 | 4 |
| 필리핀 | 마닐라 | MNL | 2023-01-01 | 00:40:00 | 00:57:00 | 5J187 | 0 |
| 네덜란드 | 암스테르담 | AMS | 2023-01-01 | 01:25:00 | 01:46:00 | KL862 | 9 |
| 베트남 | 두옹 당(푸퍽) | PQC | 2023-01-01 | 01:45:00 | 01:48:00 | VJ975 | 0 |

▶ Project Goal: What do we want? Why do we want?

▶ Project Goal

- Analyzing the airport transfer of striking country/continent over time
- Estimating and suggesting the flight path related to them

▶ Why do we want it?

- Support travelers to make them understand trends over time and establish efficient travel plans.
- From Incheon airport's view, they can get an idea of profitable/popular routes to develop and operate more intensively.

▶ So we focused on the following columns:

- Arrival/Departure Country, Airport, Flight Date, Number of Transfer Passengers (throughout the analysis)
- Scheduled Time (additionally in estimating the routes)

▶ Importing dataset & Pre-Processing

▶ Import two dataset as 'arrival' and 'departure'

▶ 1) Check missing data

```
sum(is.na(arrival))
sum(is.na(departure))
```

```
[1] 0
[1] 0
```

```
# Both dataset have no missing data.
```

▶ 2) Rename Korean column names into English

| Arrival_Country <chr> | Arrival_Airport <chr> | A_Airport_Code <chr> | A_Flight_Date <date> | A_Scheduled_Time <S3: hms> | A_Actual_Time <S3: hms> | A_Flight_Name <chr> | A_Transfer_Passengers <dbl> |
|----------------------------|----------------------------|-------------------------|-------------------------|-------------------------------|----------------------------|------------------------|--------------------------------|
| Departure_Country <chr> | Departure_Airport <chr> | D_Airport_Code <chr> | D_Flight_Date <date> | D_Scheduled_Time <S3: hms> | D_Actual_Time <S3: hms> | D_Flight_Name <chr> | D_Transfer_Passengers <dbl> |

▶ 3) Create new columns from existing columns

i) 'A/D_Continent' from 'Arrival/Departure_Country'

ii) 'A/D_Day of Week' from 'A/D_Flight_Date'

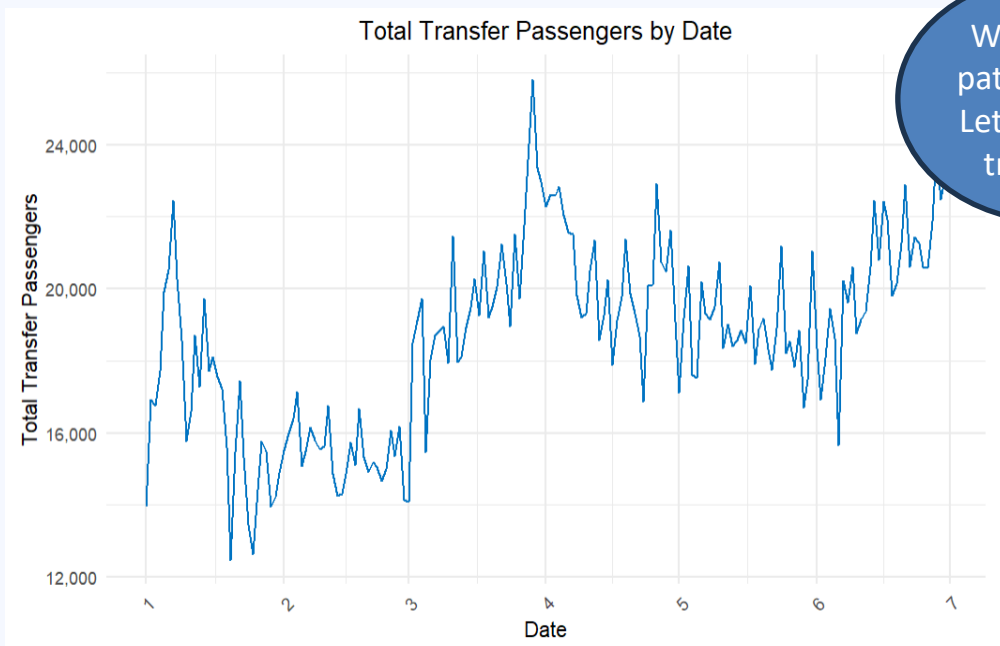
| A_Continent <chr> | A_Day_of_Week <chr> | D_Continent <chr> | D_Day_of_Week <chr> |
|----------------------|------------------------|----------------------|------------------------|
| SouthEast Asia | 일 | West/South Asia | 일 |
| SouthEast Asia | 일 | West/South Asia | 일 |
| SouthEast Asia | 일 | Africa | 일 |
| SouthEast Asia | 일 | SouthEast Asia | 일 |
| East Asia | 일 | Europe | 일 |
| SouthEast Asia | 일 | SouthEast Asia | 일 |
| SouthEast Asia | 일 | East Asia | 일 |
| SouthEast Asia | 일 | SouthEast Asia | 일 |
| SouthEast Asia | 일 | SouthEast Asia | 일 |
| East Asia | 일 | SouthEast Asia | 일 |

► Glimpse of the Number of Total Transfer Passengers Over Time

► Why do we show only the 'Total' result now?

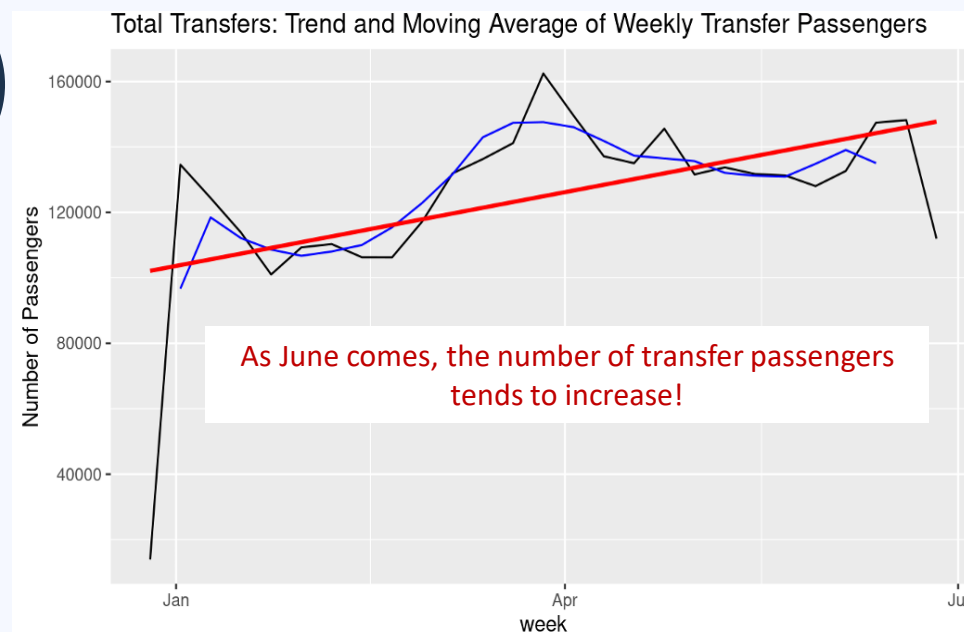
: The pattern of both datasets == The pattern of total dataset (will be proved in report)

► The Number of Transfer Passengers Over Time



What's the pattern here?
Let's see the trendline

► Trend and Moving Average of Weekly Transfer Passengers



03

Insights from Basic Investigation

Now, we are moving to countries and continents information: Visualization by World Map

Total Transfer Passengers by Country



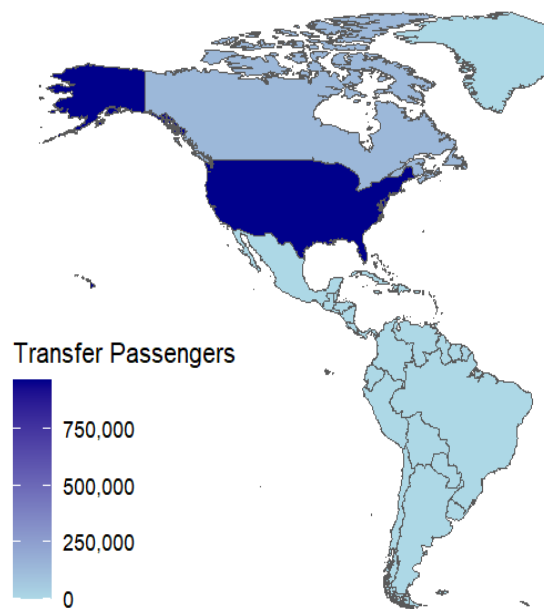
► North America & Asia seem to be significant in our dataset!

Insights from Basic Investigation

► More specific maps of America and Asia

► 3 noticeable continents to focus on

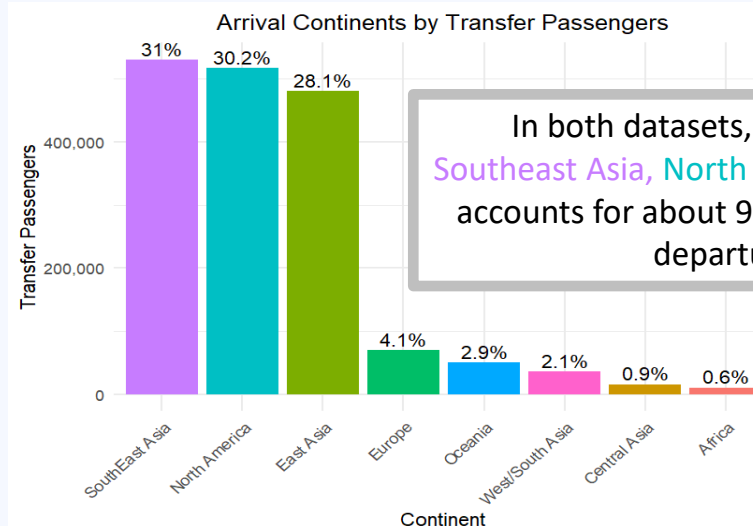
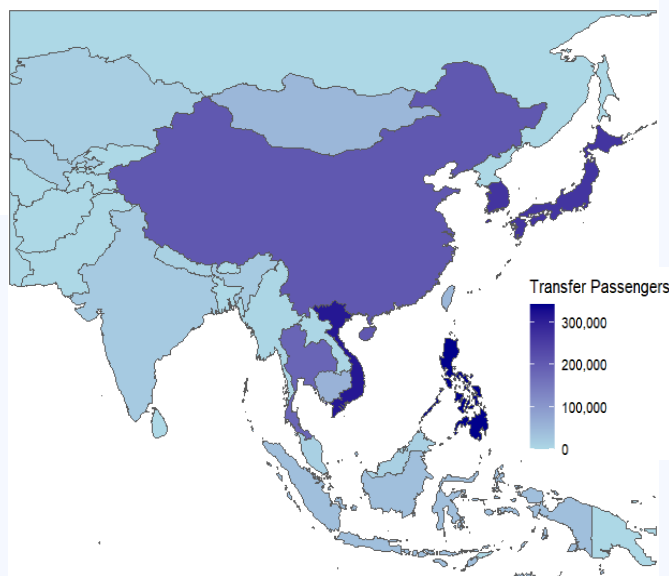
Total Transfer Passengers by Country (America)



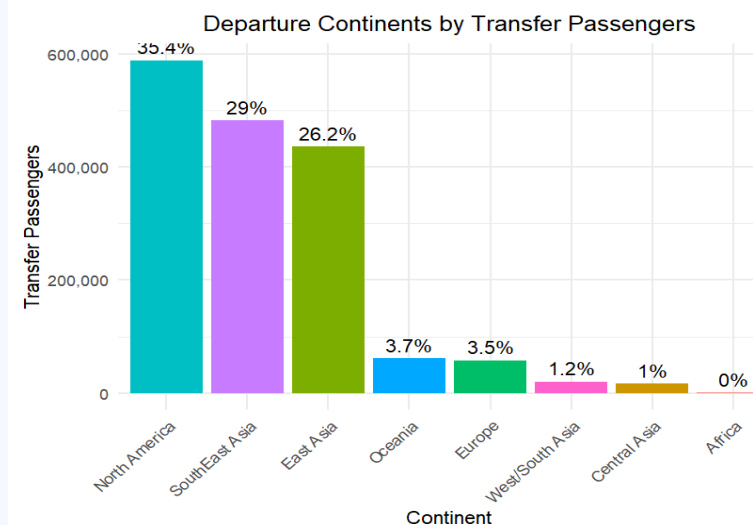
◀ For America, the United States seems to have the highest number of passengers.

► For Asia, especially East and Southeast Asian countries shows high numbers.

Total Transfer Passengers by Country (Asia)

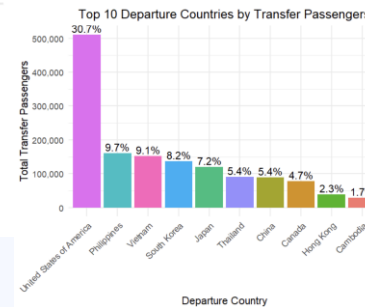
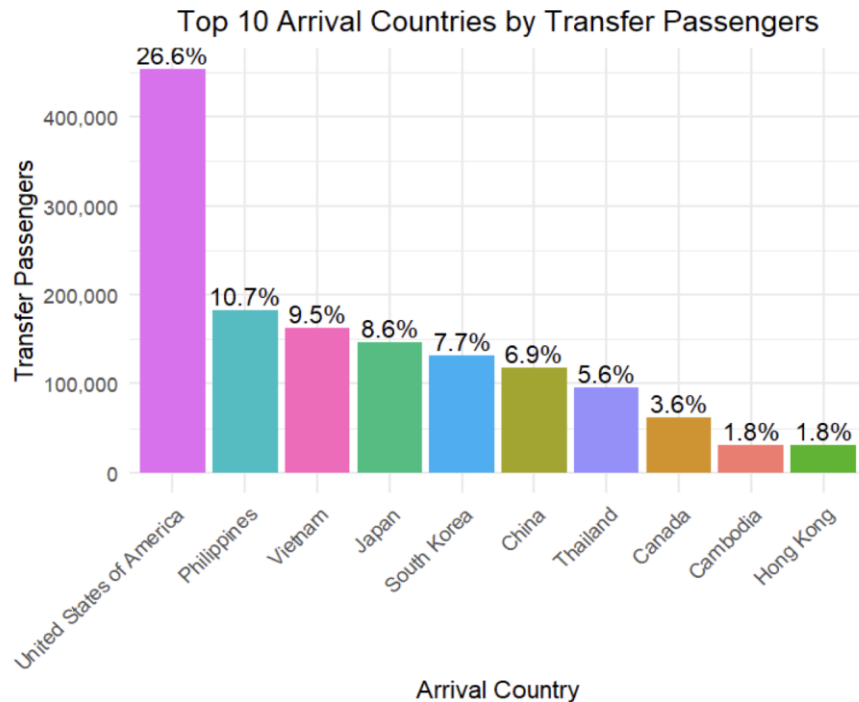


In both datasets, 3 continents
 Southeast Asia, North America, East Asia
 accounts for about 90% of arrival and
 departure

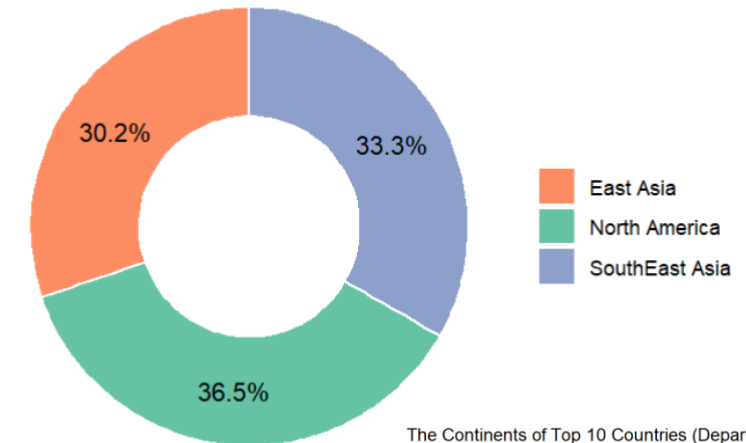


Insights from Basic Investigation

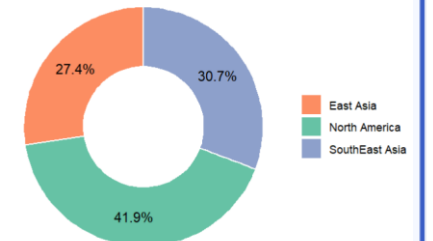
► Top 10 Countries by Transfer Passengers in Arrival and Departure



The Continents of Top 10 Countries (Arrival)



The Continents of Top 10 Countries (Departure)



- 1) the US 2) Southeast Asian countries 3) East Asian countries
- Both have the same countries with a little difference in rank
- All the top 10 countries in both datasets are the countries in the top 3 continents!

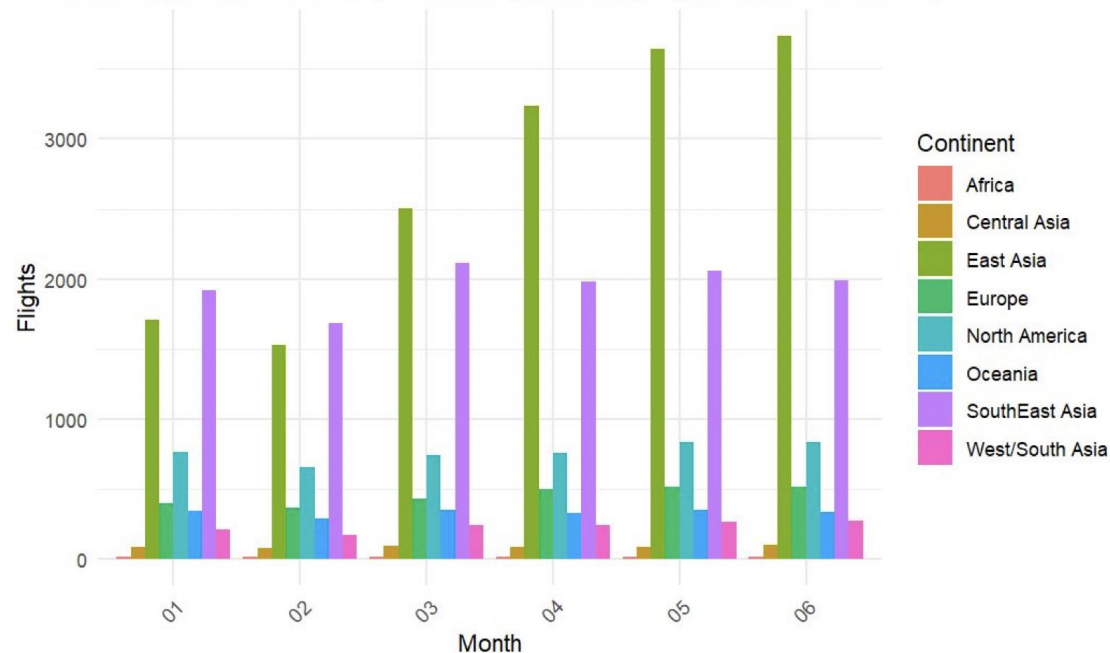
- 1) North America 2) Southeast Asia 3) East Asia

03

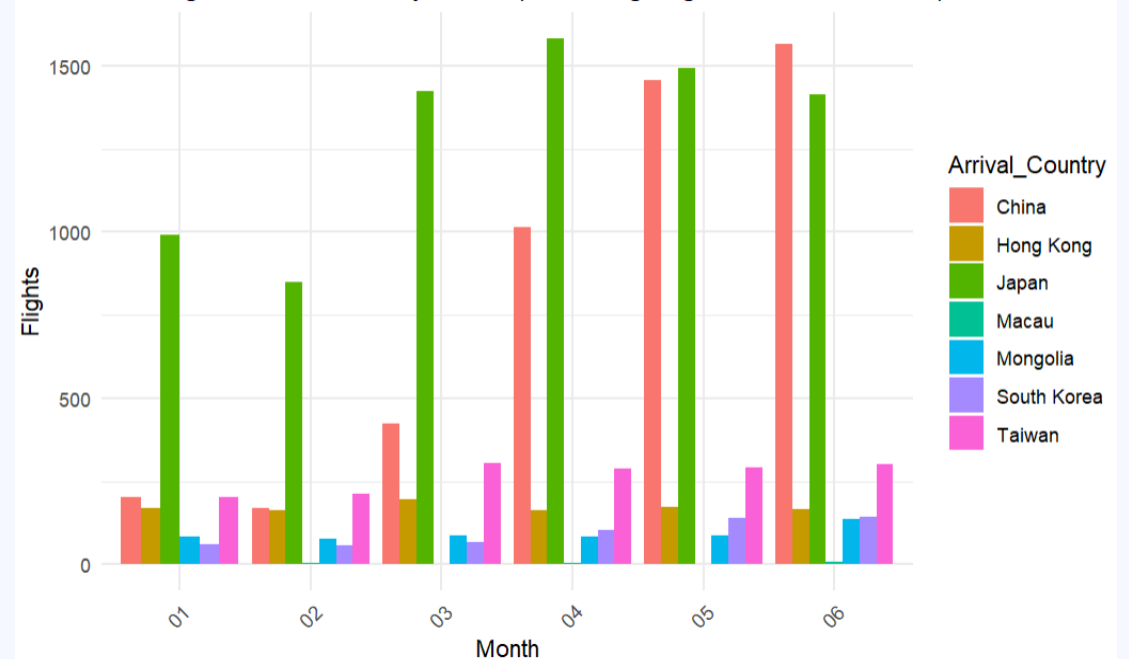
Insights from Basic Investigation

► How about the number of flights to the three continents over time?

Arrival Flights by Continent per Month (Excluding Flights with 0 Transfers)



Arrival Flights to East Asia by Month (Excluding Flights with 0 Transfers)



► Initial Findings

- From January to June, the number of flights arriving in **SouthEast Asia** and **North America** seems consistent.
- Rather a notable increase in the arrival flights to **East Asia**

► Which country would have contributed to East Asia's increase?

- Most of the flights to East Asia arrived in **Japan**.
- Arrivals to **China** have increased rapidly.

04 Hypotheses

► Insights so far and Ideas for hypotheses

► Insights from previous visualizations

- i) The **number of transfer passengers** at Incheon Airport **has increased over time**.
- ii) Our dataset's most crowded/frequent countries and continents were **North America, South Asia, and East Asia**.
- iii) However, the number of flights showed a distinct change only in East Asia.

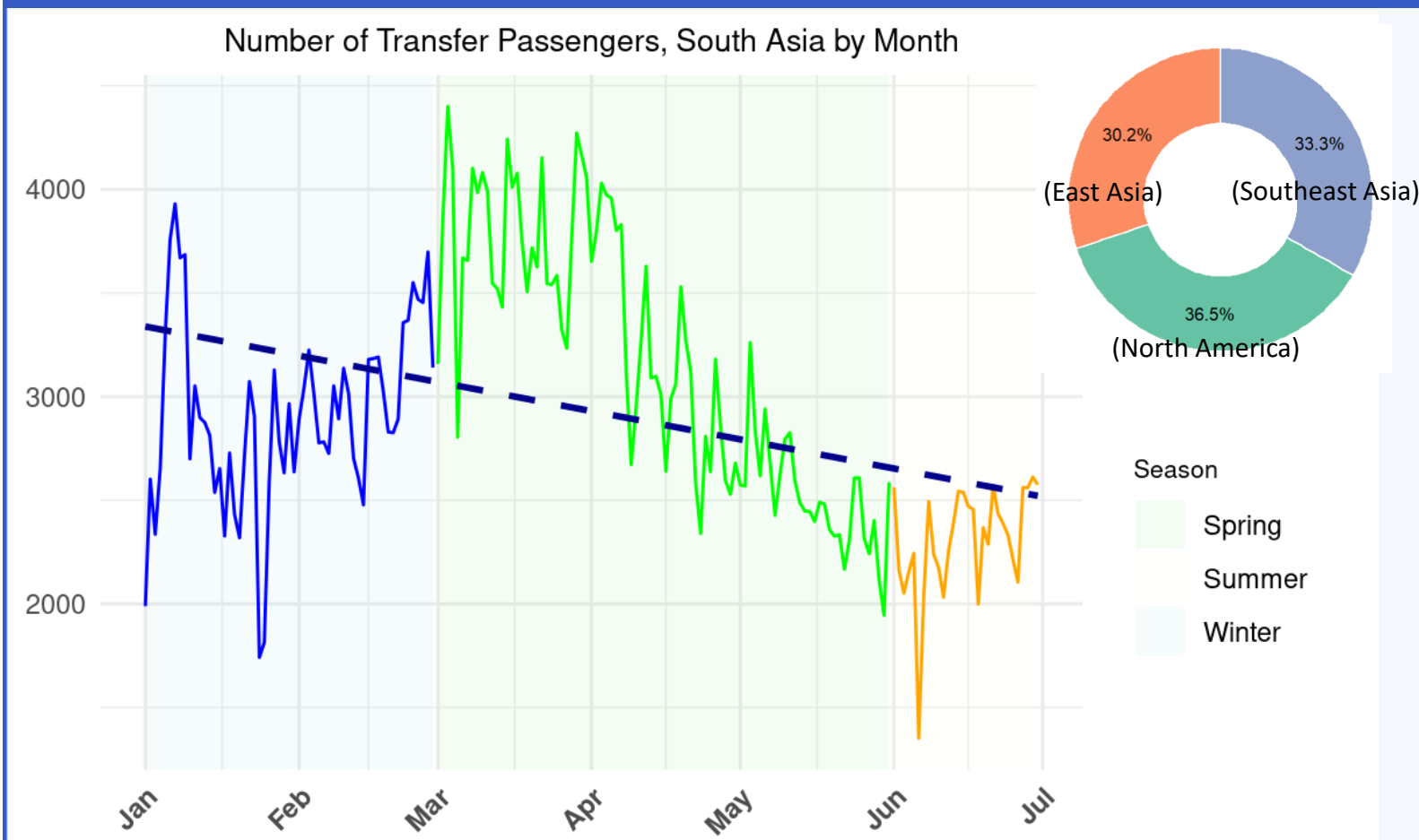
► We got ideas for our hypotheses!

1. **The United States** showed very high numbers of passengers. However, it is quite different from our common sense that the number of passengers will decrease as the distance between the country and the Incheon becomes further. Then, **how will the number of transfer passengers change as the distance increases?** Will it be less or more? (Hypothesis 3)
2. **Southeast Asia** was also remarkable. Considering **the increase of transfer passengers as summer comes, we can assume that the number of transfer passengers going to Southeast Asia is expected to increase**, potentially due to the summer resorts and closeness to Incheon. (Hypothesis 1)
3. **East Asia** was also ranked in the top continents. We can imagine that **the number of transfer passengers to East Asia will be highest in winter** due to the influence of the Chinese New Year. Will it be true? (Hypothesis 2)

04

EDA Hypothesis 1 Verification

► Hypothesis 1: As summer comes, the number of transfer passengers arriving in Southeast Asia will increase.



Linear Regression Analysis:

•**Idea:** The pie chart shows a significant portion of South Asia, which is bigger than East Asia.

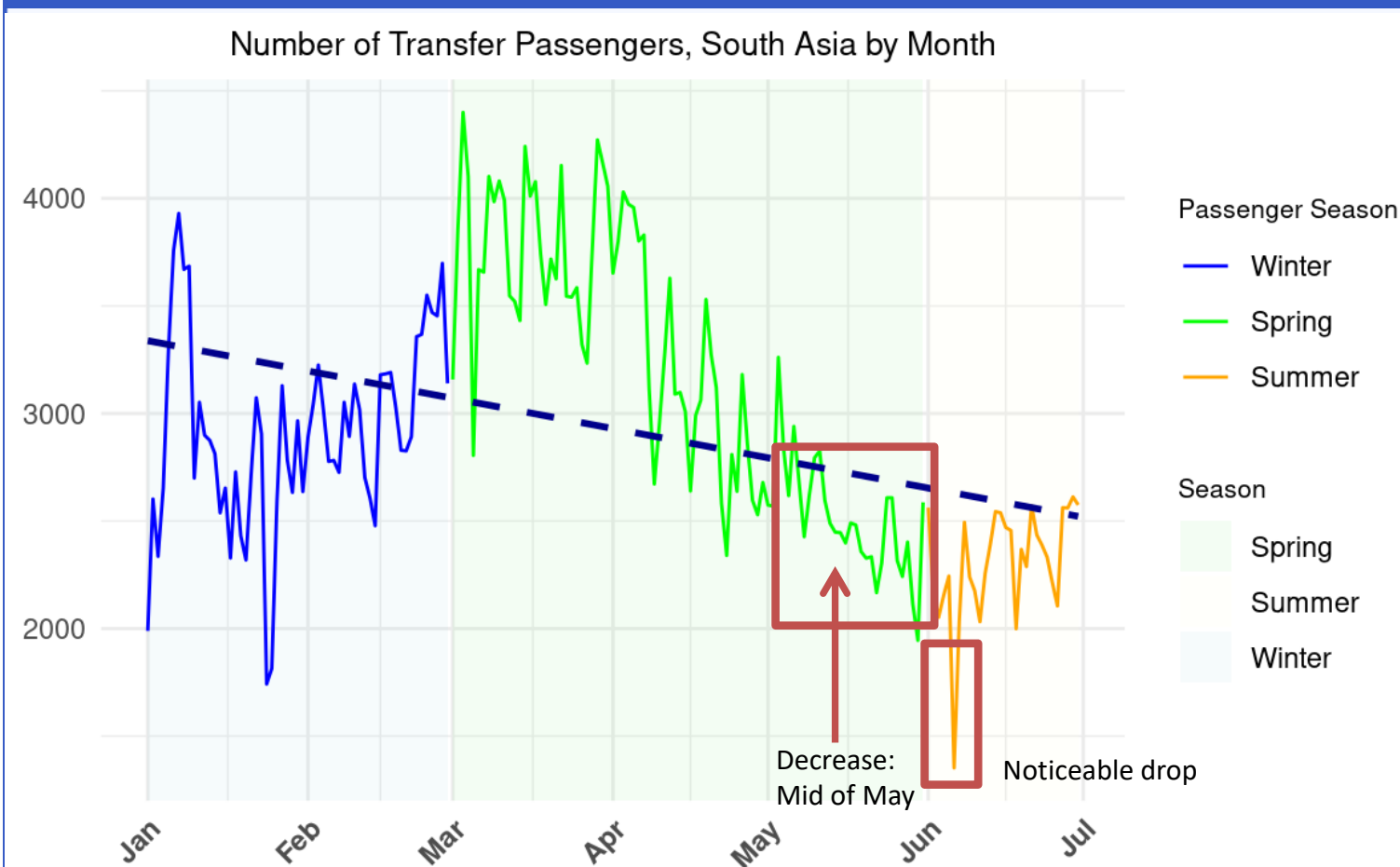
•**Negative Slope:** Indicates a downward trend in transfer passengers over the observed period.

•**Overall Trend:** Transfer passengers to South Asia generally decreases from winter through summer.

⇒ Then the hypothesis should be...
(ACCEPTED?/REJECTED?)

EDA Hypothesis 1 Verification

► Hypothesis 1 Answer: As summer comes, the number of transfer passengers arriving in Southeast Asia will increase. => Rejected

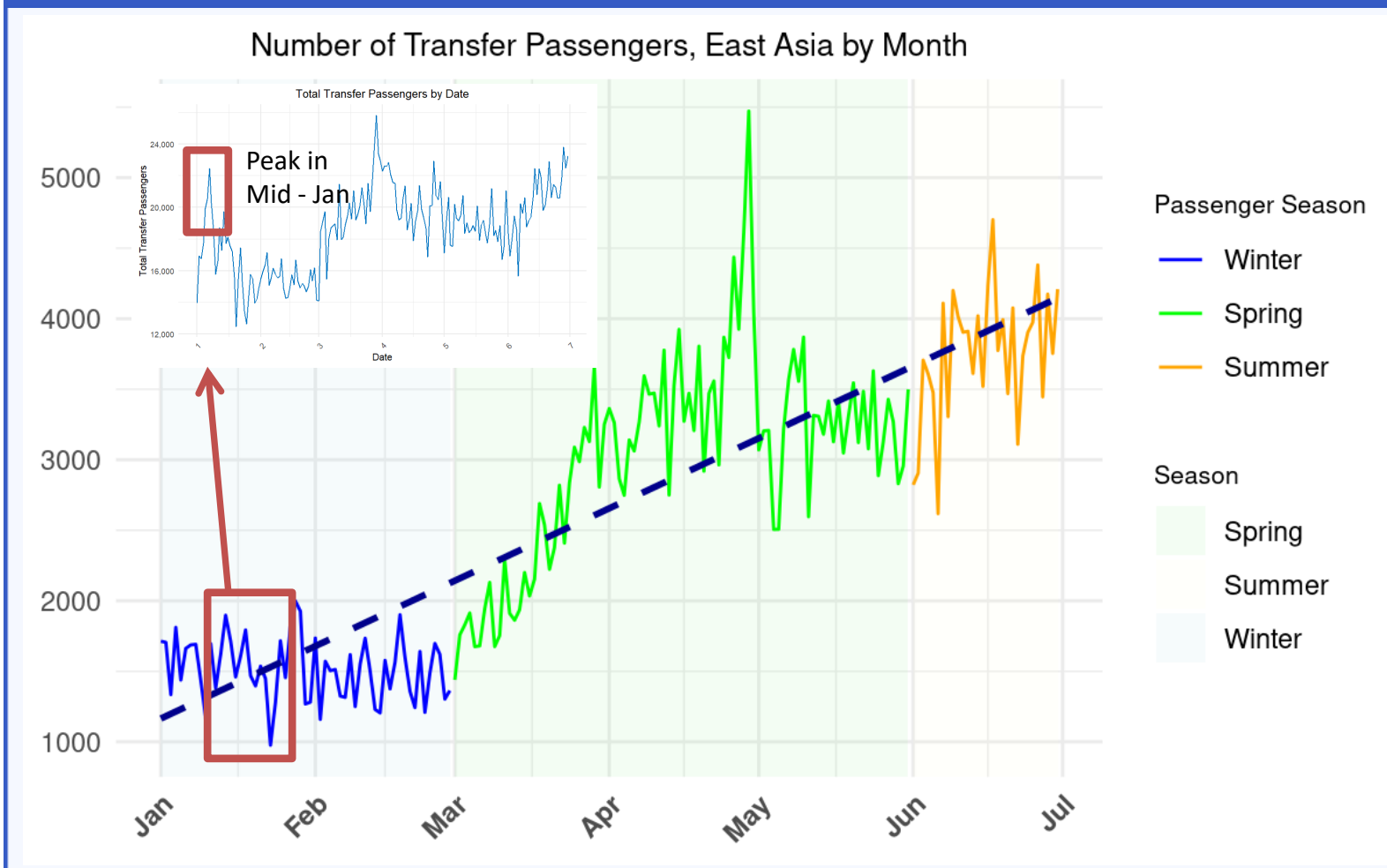


- **Answer: Rejected !!!**
- **Verification:** Noticeable decrease in transfer passengers from the middle of May. (Monsoon season in South Asia begins = heavy rainfall and potential disruptions may deter travelers from planning vacations)

Seasonal Trends Summary

- **Winter Trends:** Significant fluctuations with peaks at the beginning and several troughs.
- **Spring Variability:** High variability with peaks in early March and April.
- **Summer Trends:** Continued decline in transfer passengers with a noticeable drop.

► Hypothesis 2 : For the high number of transfer passengers in January, East Asia contributed significantly due to China's New Year holiday.



► Linear Regression Analysis

Hypothesis2 :

- When we look at the trend of total transfer passengers by date, we can see a peak in mid-January.

Positive Slope:

- Indicates an overall upward trend

Overall Trend:

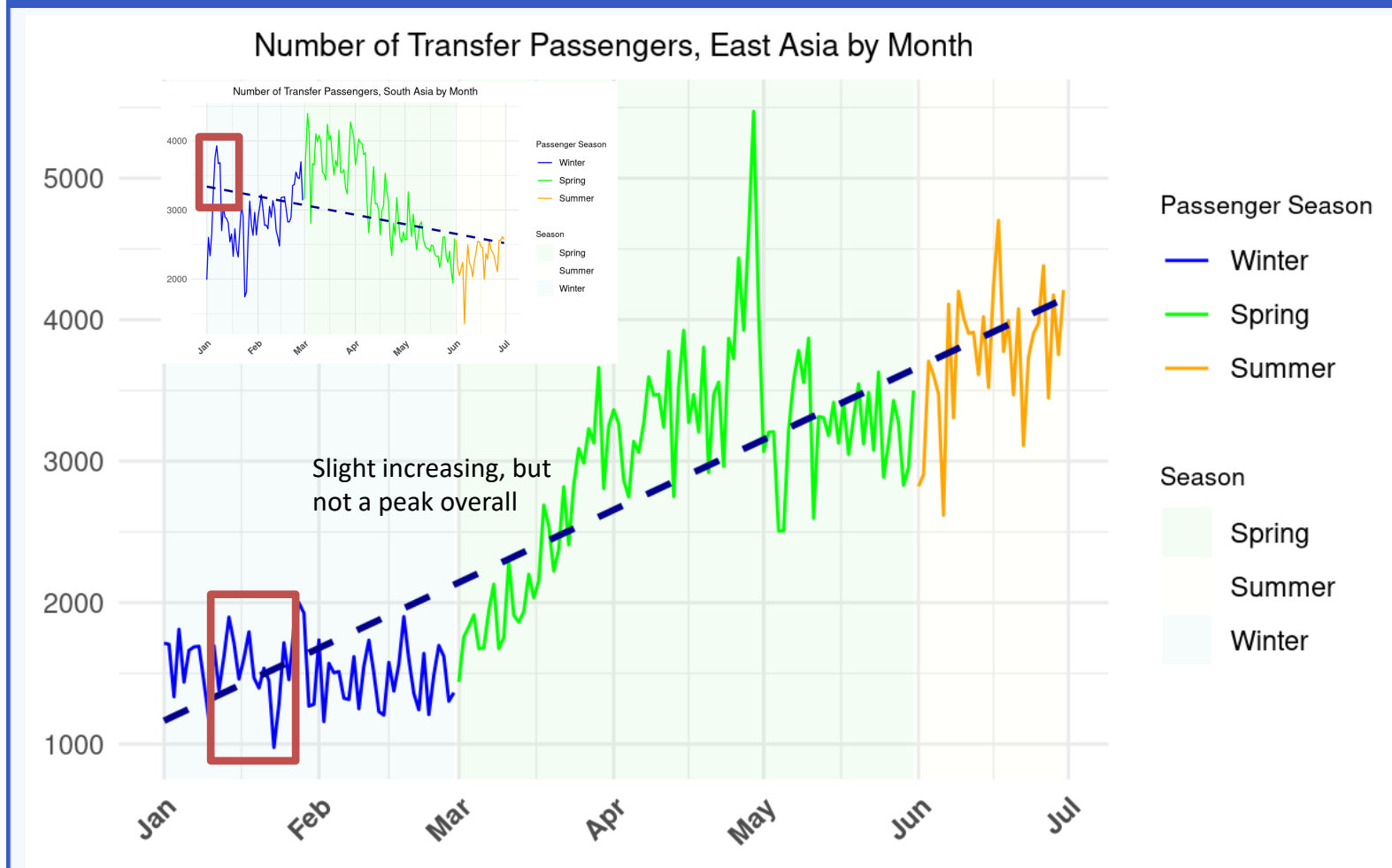
- The number of transfer passengers related to East Asia generally increases from winter to summer.

=> Then the hypothesis should be..
(ACCEPTED?/REJECTED?)

04

EDA Hypothesis 2 Verification

► Hypothesis 2 Answer: For the high number of transfer passengers in January, East Asia contributed significantly due to China's New Year holiday => Rejected



► **Answer: Rejected**

• **Verification:** We can't see any noticeable peak in the mid-January in East Asia. (However, South Asia might contribute to those trends)

► **Seasonal Trends Summary**

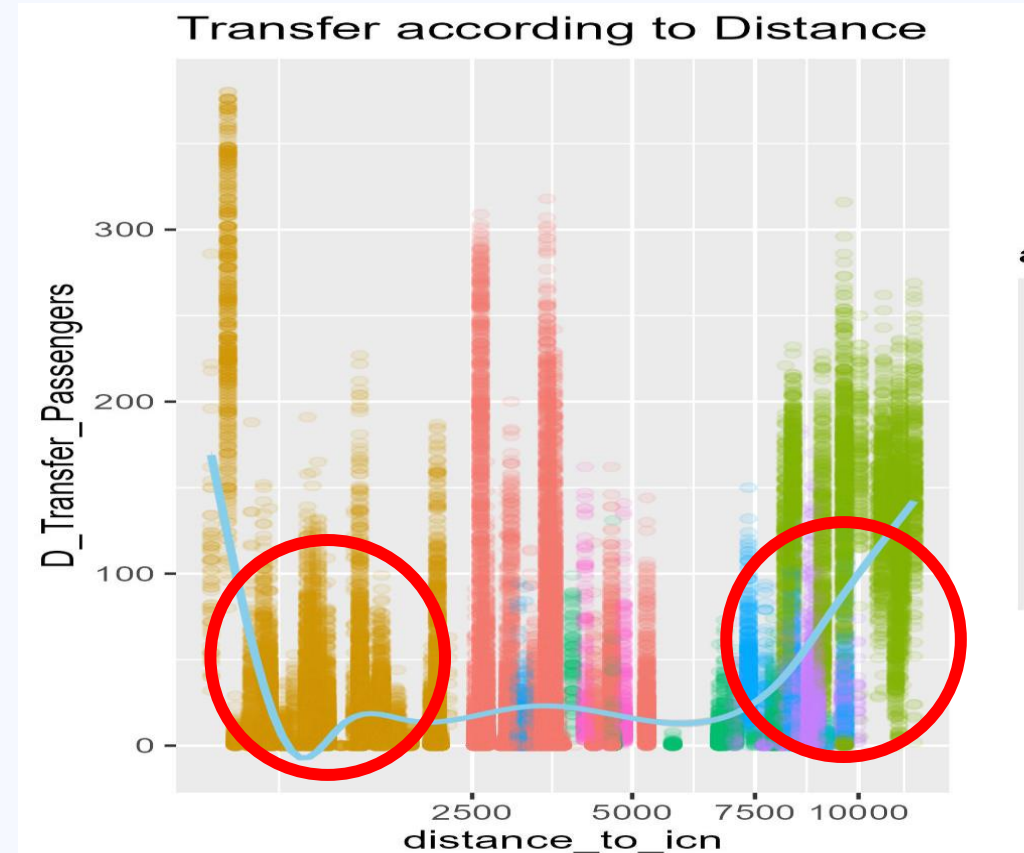
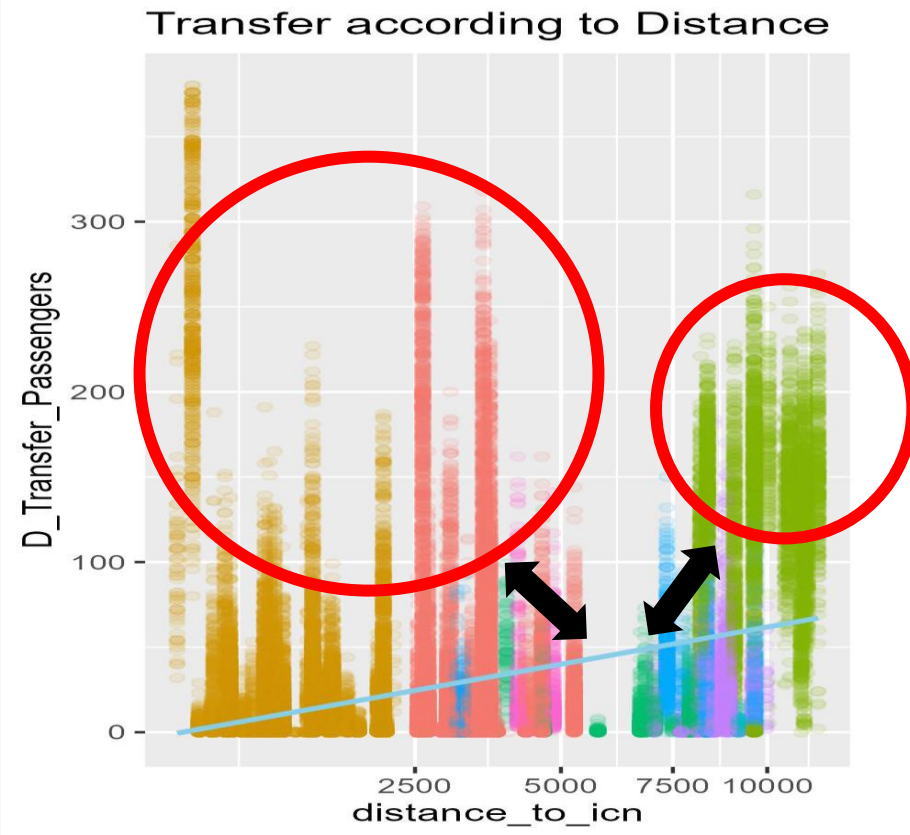
- **Winter Trends:** Fluctuations with a general increase in transfer passengers.
- **Spring Variability:** Significant increase with noticeable peaks in April and May. Consistently upward, indicating growing transfer passenger numbers.
- **Summer Trends:** Continued increase in transfer passengers.

04

EDA Hypothesis 3 Verification

► Hypothesis 3 : The further the distance from the transfer airport, the greater the number of transfer passengers will be there.

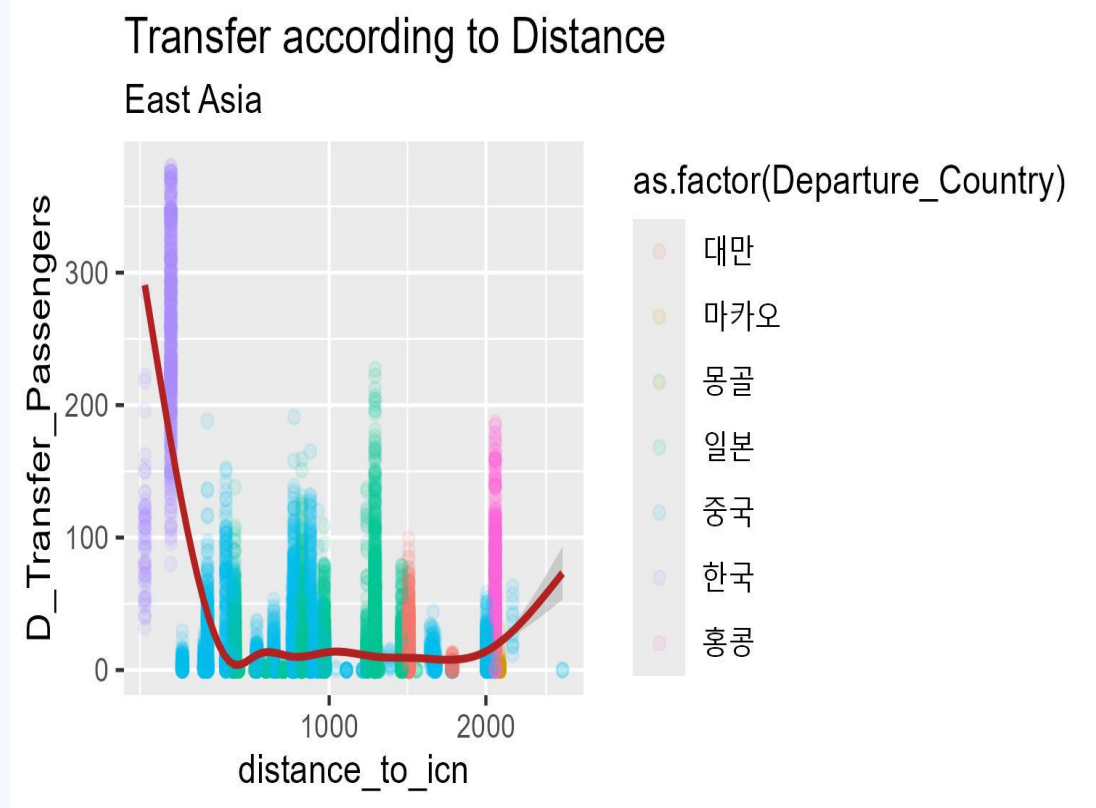
- When we use Linear Regression, it seems to exist in relation, but looking at the data distribution, it is REJECTED.



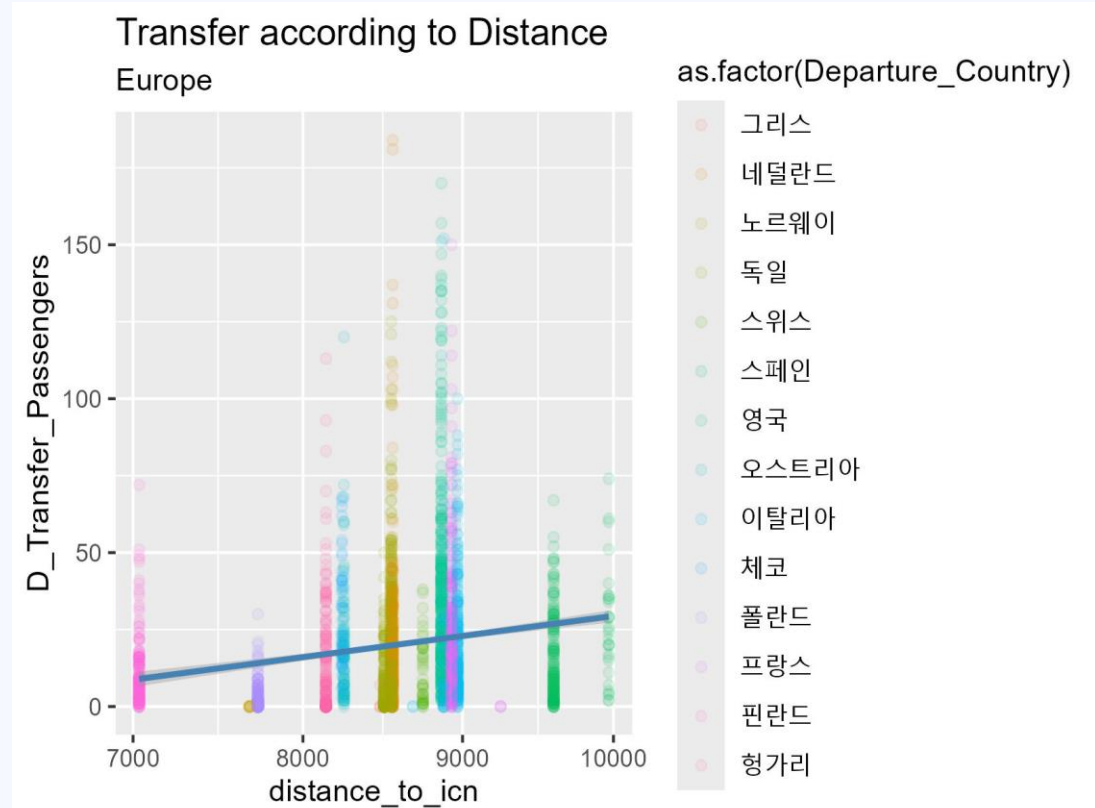
04

EDA Hypothesis 3 Verification

However, meaningful characteristics show that the influence of East Asia and the West on the transfer at Incheon Airport is large.



**East Asia Similar to Transfer Patterns
Around the World**



**Western world similar to transfer relationships and patterns
around the world**

▶ Purpose of suggesting the flight route

▶ Incheon International Airport has a status as a **hub airport** that transfers a lot. Incheon International Airport's transfer airliner supply was only skewed to Southeast Asia and the United States.

<https://www.traveltimes.co.kr/news/articleView.html?idxno=405885>

However, We believe that the diversity of transit passengers should be guaranteed at the hub airport. Additionally, data analysis showed that East Asia had a noticeable number of transit passengers and that Busan acted as a transit route, whether it was the destination or the departure point.

Although it is a popular place, airports that have not developed much as transfer routes were also seen in East Asia and Europe. Therefore, we wanted to propose **airports that have not developed much as transfer routes as new transfer routes**, and **existing transfer routes as new direct routes**.

▶ Why we use that method :

The hypothesis about the condition explained about the inner join will be explained later.

Also, we think that the total average passenger can be **calculated average of an average of both departure and arrival airports**. Therefore, we use that kind of formula.

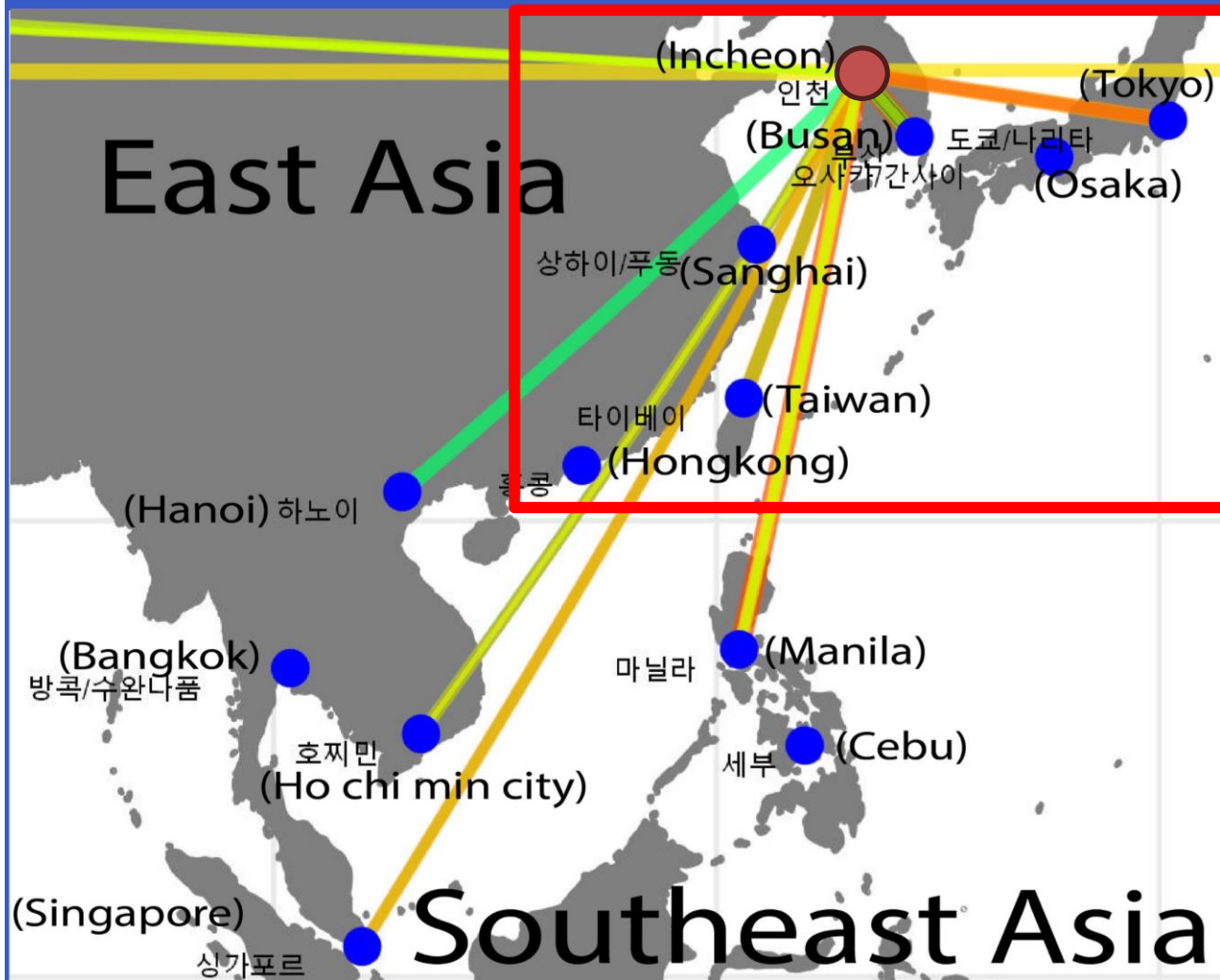
The same color of the line means countries are included in one route (it doesn't matter if it's a departure or arrival).

The more routes, the more colors can overlap, and the more passengers the thicker the lines.

For example, map for less than 30 minutes, Busan has more than 3 routes(Busan-Manila, Busan-Osaka, Busan-Ho-chi-min city)

04 Route Proposal

▶ <30



▶ **Preconditions:** When one transfers to Incheon International Airport, one will transfer on the same day and transfer from the departing passenger to the arriving passenger within an hour and a half.

▶ **Methods:** Subset by top 25 transfer airports -> Inner Join by day -> Subset by time required for transfer

```
total_transfer_passengers =  
(sum(A_Transfer_Passengers, na.rm = TRUE) + sum(D_Transfer_Passengers, na.rm = TRUE)) / (2 * n())
```

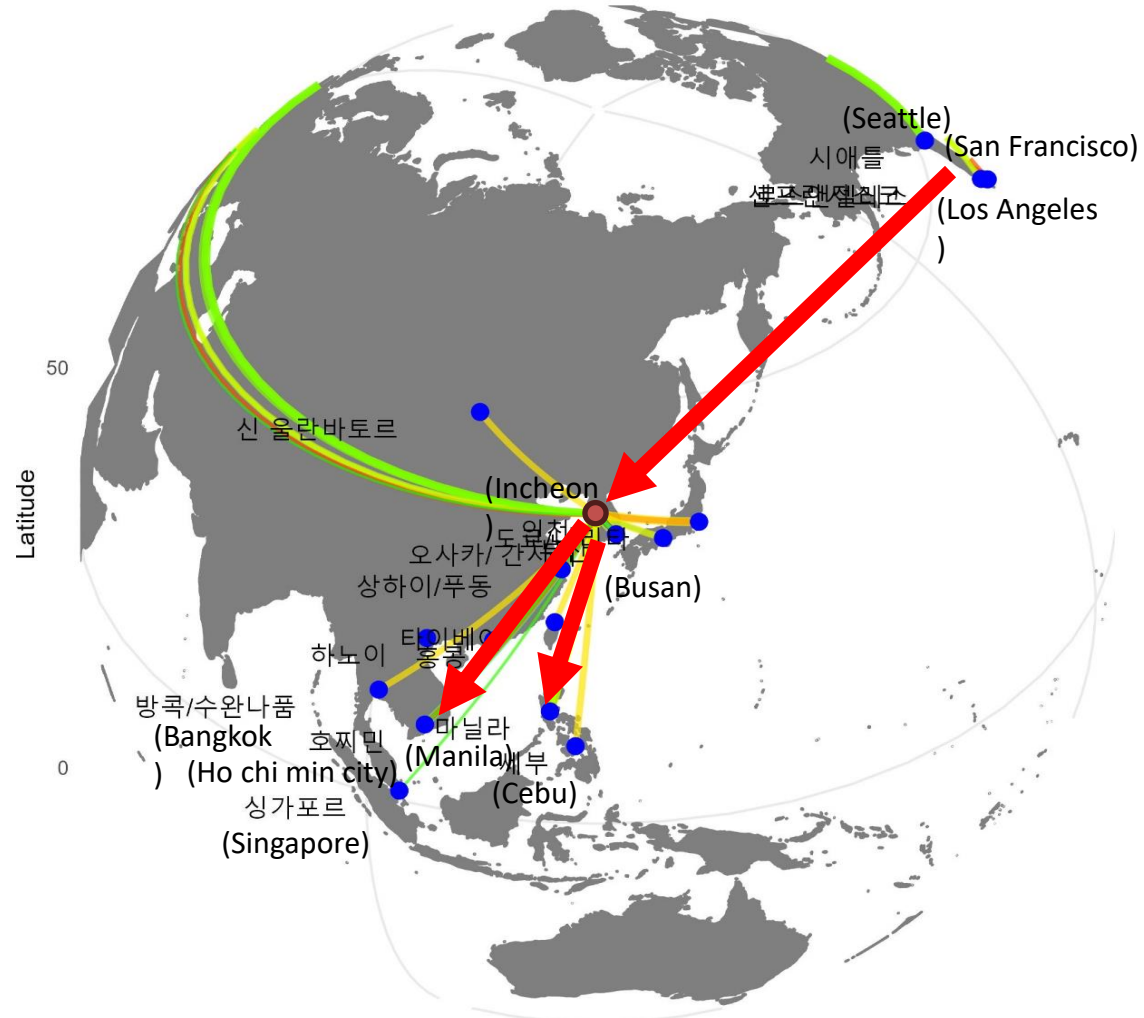
▶ **East Asia** make up the majority (Tokyo, Osaka, Shanghai, Taipei)

▶ **Busan** is mainly included in routes.

04

Route Proposal

▶ 30<= & <60

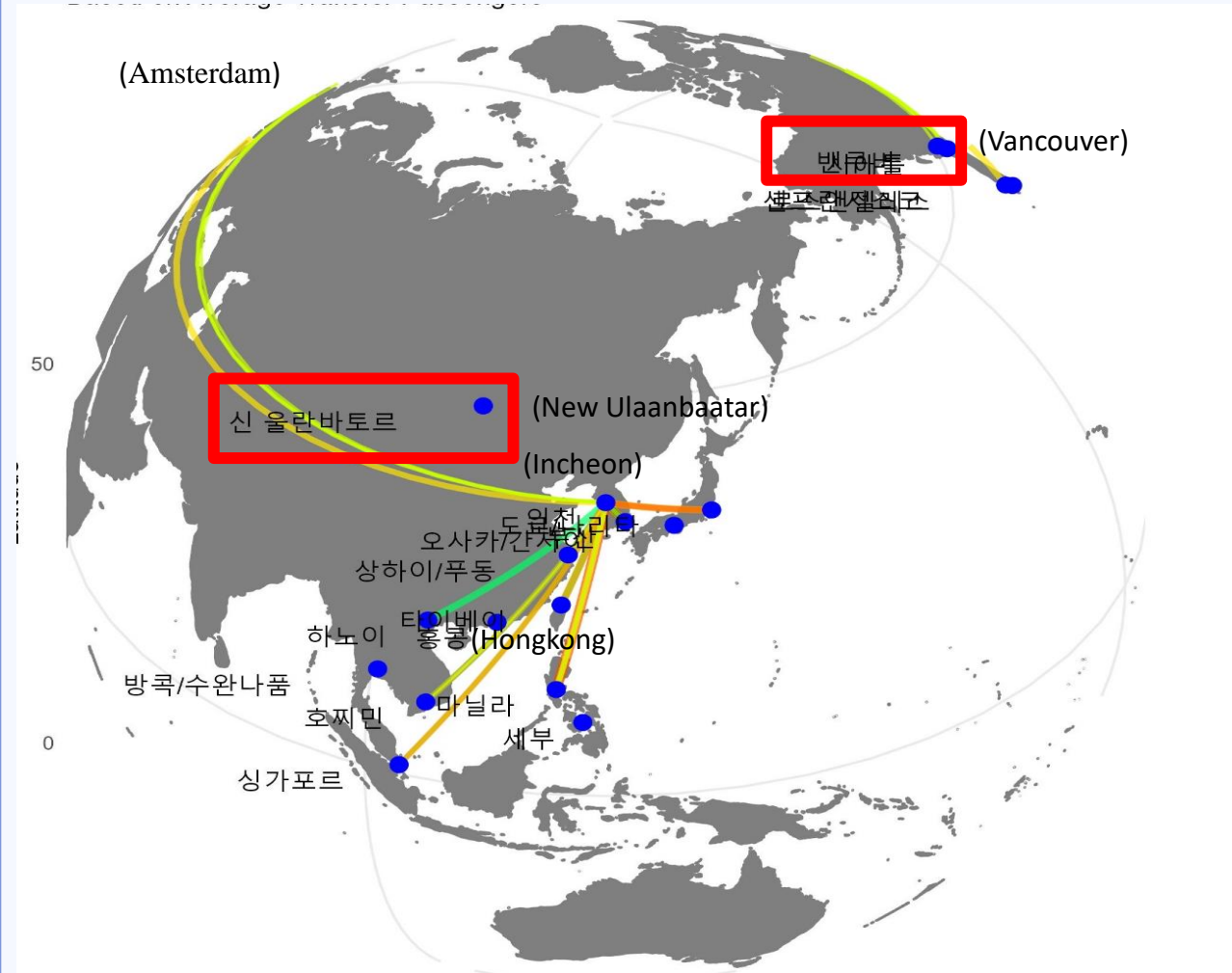


- **Still, Busan** is mainly included in routes.
- Far East Asia and Southeast Asia increase
- Routes that include **both Western and Southeast Asia** exploded
Example :
1) Seattle, San Francisco - Manila
2) Ho Chi Min City – Los Angeles
- **Strengthen** the consistency of existing analysis that Western and East Asia account for the majority of transfers

04

Route Proposal

▶ 60<=&<90



- Western countries increased more
- More diverse Airports
- Routes that include **both Western and Southeast Asia dominated**
- **New Routes Suggestion**
- EX) New Ulaanbaatar, Hong Kong in East Asia
- EX) Amsterdam, Vancouver

► Challenges that we went through

1. Non-Technical Challenges

P) As we worked most of the time separately, our analysis varied, and some findings became less significant. It was hard to merge our work into one coherent plot.

S) We discussed more and more to make our project the best. We defined **clear objectives** for the analysis to ensure everyone worked towards the same goals over time. Also, we **focused on the works that had been worked well** and developed them to support our later hypotheses and suggestions through our previous work.

2. Technical Challenges

P1) Since the data in CSV was almost non-numeric, meaningful numerical data conversion to find a relation was the biggest challenge. To derive the relation between distance and the number of transit passengers, it was necessary to obtain the distance from numerous airports to Incheon Airport.

S1) **Searching and applying packages/functions** (apply each function to the extent that the computer is turned off once while applying to obtain the distance one by one function)

P2) Regarding suggesting a new route through modeling, the dataset was not very suitable to use modeling because the data survey period was very short (about 6 months)

S2) Assuming that transfer time **is limited to less than 90 minutes** & existence of a **good basis** for proposing a transfer route

► Conclusions for each hypothesis and proposal

- (1) Hypothesis 1: As summer approached, the number of transfer passengers to **South Asia** didn't increase, likely due to the monsoon season. In the mid of January, while East Asia didn't significantly contribute to a peak in total passengers, South Asia might have, and we plan to explore this further.
- (2) Hypothesis 2: In mid-January, while **East Asia** didn't significantly contribute to a peak in total passengers, South Asia might have, and we plan to explore this further.
- (3) Hypothesis 3: From the **United States** case, <The further the distance from the transfer airport, the greater the number of transit passengers will be there>, is rejected. Rather, we found that regions like East Asia and Western countries play a significant role.
- (4) **New Routes** can be proposed. Such as...
- i) Direct Flights to Busan
 - : Limited to only East Asia such as Tokyo, Osaka, Shanghai, Taipei, Hongkong
 - ii) New Connections (East Asia and Southeast to Western)
 - : Such as Bangkok to Amsterdam, Vancouver to New Ulaanbaatar

24-1 Data Science and R Final Project

Thank You