

Data Mining Practice-team2

❶ 생성자	❷ 벼리 문
❸ 생성 일시	@2024년 9월 7일 오후 4:15
☰ 태그	

22000283	박예준
22000245	문벼리
22001030	신혜

활용 데이터셋

```
data <- read.csv("C:/Users/silkj/Desktop/한동대학교/5학기/데이터 마이닝 실습/Data-Mining-Practicum/myR/public_emdat_project.csv")
```

Task1: 탐색적 데이터 시각화

▼ Cross-Sectional Approach

데이터 전처리

Step 1

해당 함수를 활용해 데이터의 결측치 비율, unique value 등을 점검

```
def analysis_df(data):  
    analysis_data = pd.DataFrame()  
    analysis_data['Columns'] = data.columns  
    analysis_data['Data Type'] = data.dtypes.values  
    analysis_data['Unique Values'] = data.nunique().values  
    analysis_data['Missing Values'] = data.isna().sum().values  
    analysis_data['Missing Value Percentage'] = data.isna().sum() / data.shape[0]  
    return analysis_data
```

```

anaylsis_info = analysis_df(data).sort_values(by='Missing Value Percentage', ascending=False)
anaylsis_info

```

✓ 0.0s

	Columns	Data Type	Unique Values	Missing Values	Missing Value Percentage
37	Reconstruction Costs, Adjusted ('000 US\$)	float64	32	15751	99.790928
36	Reconstruction Costs ('000 US\$)	float64	31	15751	99.790928
19	AID Contribution ('000 US\$)	float64	429	15294	96.895590
39	Insured Damage, Adjusted ('000 US\$)	float64	564	15090	95.603142
38	Insured Damage ('000 US\$)	float64	281	15089	95.596807
24	River Basin	object	1147	14572	92.321338
34	No. Homeless	float64	578	14460	91.611759
23	Longitude	float64	1610	13969	88.501014
22	Latitude	float64	1615	13969	88.501014
7	External IDs	object	1826	13379	84.763051
41	Total Damage, Adjusted ('000 US\$)	float64	2250	12673	80.290167
40	Total Damage ('000 US\$)	float64	1046	12658	80.195134
15	Associated Types	object	123	12488	79.118094
20	Magnitude	float64	1629	12406	78.598581
14	Origin	object	788	11829	74.942980
8	Event Name	object	2101	10829	68.607451
32	No. Injured	float64	655	9994	63.317283
33	No. Affected	float64	3047	8612	54.561581
43	Admin Units	object	6971	7286	46.160669
21	Magnitude Scale	object	6	5892	37.328941
35	Total Affected	float64	4153	4102	25.988343
31	Total Deaths	float64	547	3129	19.823872
27	Start Day	float64	31	1509	9.560314
30	End Day	float64	31	1442	9.135834
13	Location	object	13593	648	4.105423
42	CPI	float64	24	163	1.032691
29	End Month	float64	12	162	1.026356
26	Start Month	float64	12	69	0.437152
44	Entry Date	datetime64[ns]	2443	0	0.000000

Step2

가용가능한 column이 많기에, 우선적으로 결측값이 없는 column들을 활용하는 방향으로 가설을 설정

```

data_crossectional = data[['Country', 'ISO', 'Region','Disaster Group','Disaster Type', 'Disaster Subtype','Disaster Subgroup']]
data_crossectional

```

✓ 0.0s

	Country	ISO	Region	Disaster Group	Disaster Type	Disaster Subtype	Disaster Subgroup
0	Djibouti	DJI	Africa	Natural	Drought	Drought	Climatological
1	Sudan	SDN	Africa	Natural	Drought	Drought	Climatological
2	Somalia	SOM	Africa	Natural	Drought	Drought	Climatological
3	Angola	AGO	Africa	Technological	Road	Road	Transport
4	Angola	AGO	Africa	Natural	Flood	Riverine flood	Hydrological
...
15779	Colombia	COL	Americas	Natural	Drought	Drought	Climatological
15780	Zambia	ZMB	Africa	Natural	Drought	Drought	Climatological
15781	Marshall Islands	MHL	Oceania	Natural	Drought	Drought	Climatological
15782	Malawi	MWI	Africa	Natural	Drought	Drought	Climatological
15783	Zimbabwe	ZWE	Africa	Natural	Drought	Drought	Climatological

15784 rows × 7 columns

가설: 지리적 특성이 Disaster type에 영향을 미칠 것이다.

가설을 검증하기 위한 질문 리스트업

1. 어떤 Unit을 기준으로 를 분석하는게 좋을까?

→ Country ,Continent, Location

- Location은 지나치게 세분화되어 부적합
 - 지나치게 작은 identifier별 데이터 포인트 수 (data total rows = 15784)
 - 비교분석이 어려움

```
data.Location.nunique()
✓ 0.0s
13593

data.Country.nunique()
✓ 0.0s
222

data.Region.nunique()
✓ 0.0s
5
```

- 따라서 Continent(Region) 과 Country를 기준으로 분석하기로 함.
2. 대륙별로 발견될 수 있는 재해의 차이점은 어떤 것들이 있을까?
 3. 국가별로 발견될 수 있는 재해 특징은 어떤 것들이 있을까? 어떤 요인들이 이런 차이점을 나타내는가?

분석 시 유의해야할 점

지역별 빈도수의 총량을 비교하는 것은 위험하다. Asia의 인구 수는 America의 인구 수보다 많다. 대륙의 크기도 차이가 있다. 특히 인재(technological Disaster)의 경우는 인구수의 영향을 많이 받을 수밖에 없다. 따라서 이를 정량적으로 비교하려면 per area로 나누어줘서 발생빈도를 비교하는 것이 옳다.

결론적으로 총 value_count만을 활용하는 현재의 상황에서는 한 지역 내에서의 상대적인 발생 빈도수를 비교하는 게 좋다.

Case Division

Identifier\variable	Disaster Type	Disaster Subtype	Disaster Subgroup
Continent	Macro Case1	Macro Case2	Macro Case3
Country	Micro Case1	Micro Case2	Micro Case3

Continent : Africa, Americas, Asia, Europe

Country: Korea, Japan, China, USA

Macro to Micro view로 접근

1. Region 별로 발생하는 Disaster의 특이사항이 있을까?

1-1 Disaster_type by Region

공통 특징

대륙별 Disaster type 발생빈도는 Flood가 모두 1위. 수해가 가장 큰 빈도를 차지함

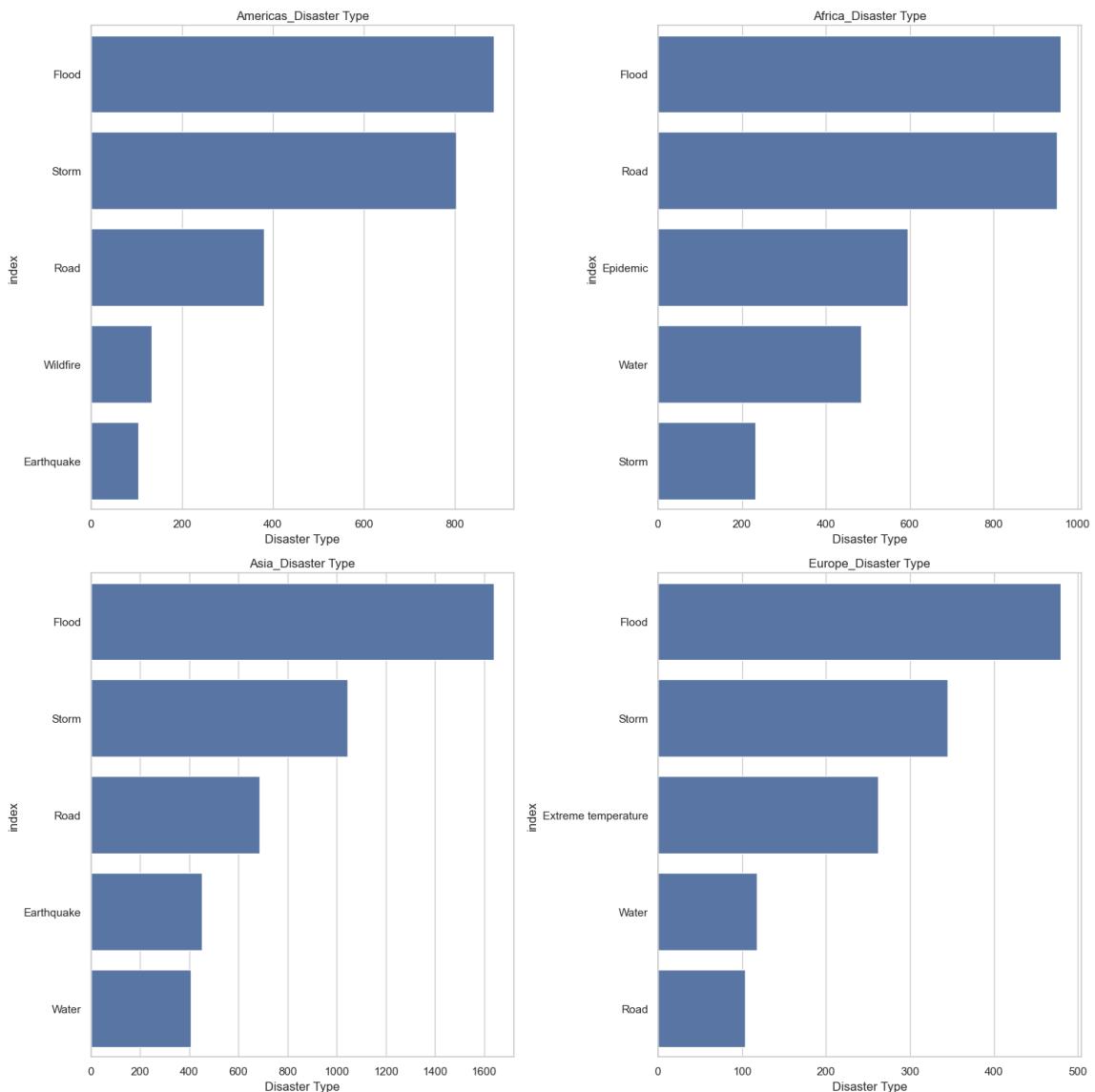
재미있는 점은 Road, 즉 집계된 교통사고 빈도가 대륙별로 2~5위로 잡힌다는 것이다.

대륙별

Africa → epidemics가 많이 발생

Americas → Wildfire가 많이 발생

Asia → Road, Earthquake가 많이 발생



question

water와 flood의 차이는 뭘까?

-> Water의 경우 해상 운송과 관련된 사고들을 담고 있음

-> 전원이 천재지변이 아닌 technological issues

```
data[data['Disaster Type']=='Water'].head(5)
```

✓ 0.0s

	DisNo.	Historic	Classification Key	Disaster Group	Disaster Subgroup	Disaster Type	Disaster Subtype
60	2000-0064-UGA	No	tec-tra-wat-wat	Technological	Transport	Water	Water
65	2000-0070-ITA	No	tec-tra-wat-wat	Technological	Transport	Water	Water
134	2000-0145-ZAF	No	tec-tra-wat-wat	Technological	Transport	Water	Water
187	2000-0203-BMU	No	tec-tra-wat-wat	Technological	Transport	Water	Water
202	2000-0220-NGA	No	tec-tra-wat-wat	Technological	Transport	Water	Water

5 rows x 46 columns

```
data[data['Disaster Type']=='Water']['Disaster Group'].value_counts()
```

#Water는 해상 운송에서 발생한 사고들을 counting.
#Flood가 수해.

✓ 0.0s

```
Technological    1111
Name: Disaster Group, dtype: int64
```

1-2 Disaster_subtype by Region

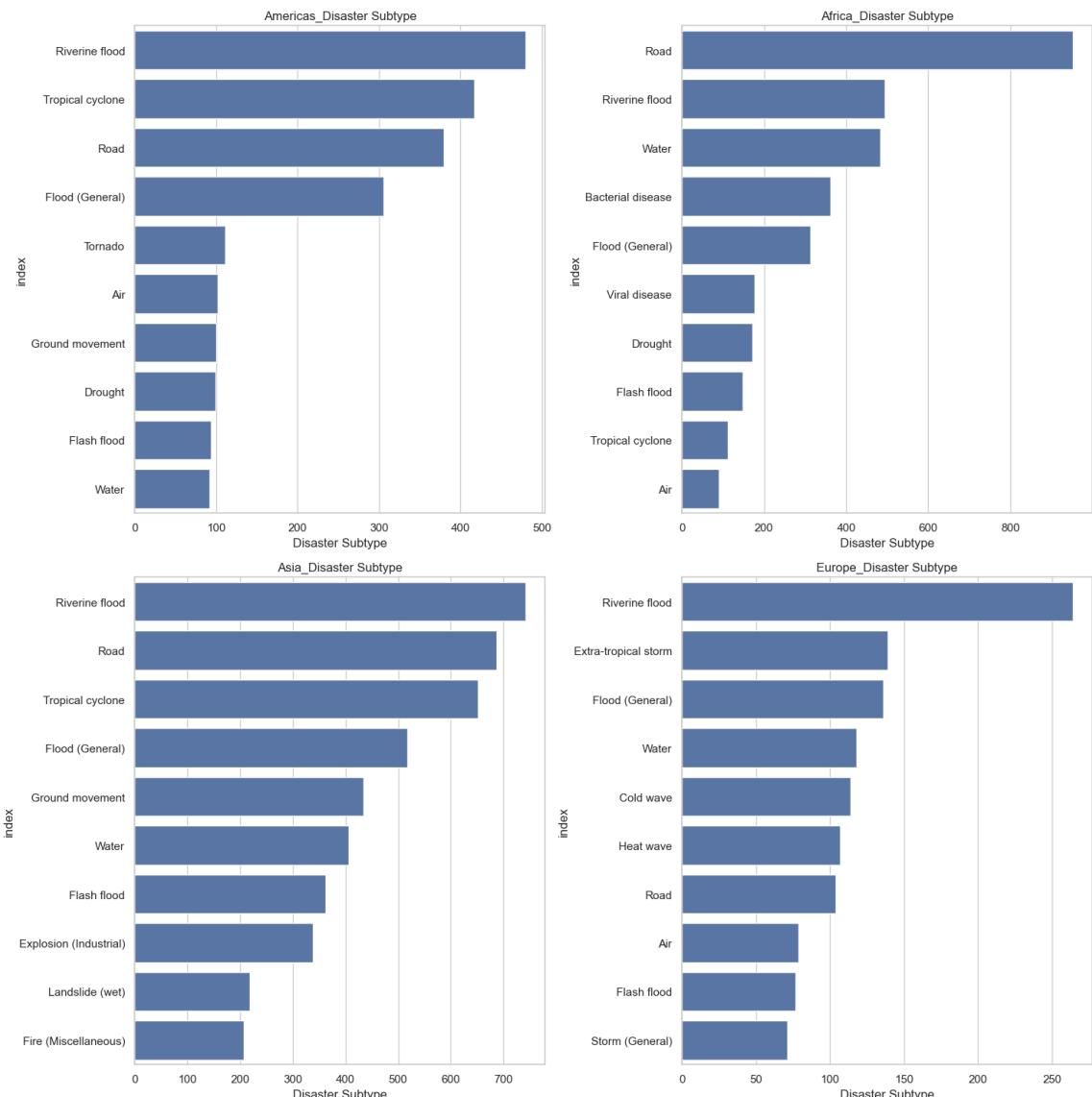
좀 더 세분화하여 분석해보자.

- America의 경우 Reverine Flood, Tropical Cyclone, flood 등 수해가 가장 높은 빈도로 발생되며 Road(교통사고)가 그 뒤를 잇는다.
- Africa의 경우 Road(교통사고) 가 단연 압도적인 비율을 차지하며, Water(수상사고)도 높은 빈도를 보인다. Bacterial/viral 등 epidemic의 빈도가 타 대륙에 비해 높다.
- Asia 역시 교통사고의 빈도가 높다. 흥미로운 점은 Industrial Explosion이 8위에 위치해있다는 것이다.

- Europe은 Riverin flood의 발생 비율이 압도적으로 높다. 타 대륙에서 보이지 않던 Cold Wave, Heat Wave라는 attribute가 보인다.

Insight1:

- 이상한 점은 지난 24년간의 데이터라고 하기에는 value count가 너무 적다. 집계 방식 및 데이터의 validity에 대하여 추가적인 검증이 필요할 듯 함.
- Asia와 Africa의 경우 대륙이 넓어 대부분의 이동/수송이 육로로 이루어진다. 차량 노후화 및 도로 인프라, 신호체계 등의 열악함 때문에 미국에 비해 교통사고 빈도가 높은 것으로 이해할 수 있다.
- Africa에서의 bacterial disease, Viral disease의 빈도는 저소득으로 인한 열악한 영양 및 위생상태 때문임으로 보인다.
- Asia의 Industrial Explosion은 공장이 많고, 생산직 노동자들을 많이 고용하는 중국, 인도네시아 등의 국가에서 많이 발생할 것이라고 추론해볼 수 있다.
- Europe은 상대적으로 고소득 국가들이 많다. 바다를 끼고 있기 때문에 해상 운송이 용이하고, 대중교통이 발달해있기 때문에 교통사고의 빈도가 적다고 볼 수 있다.



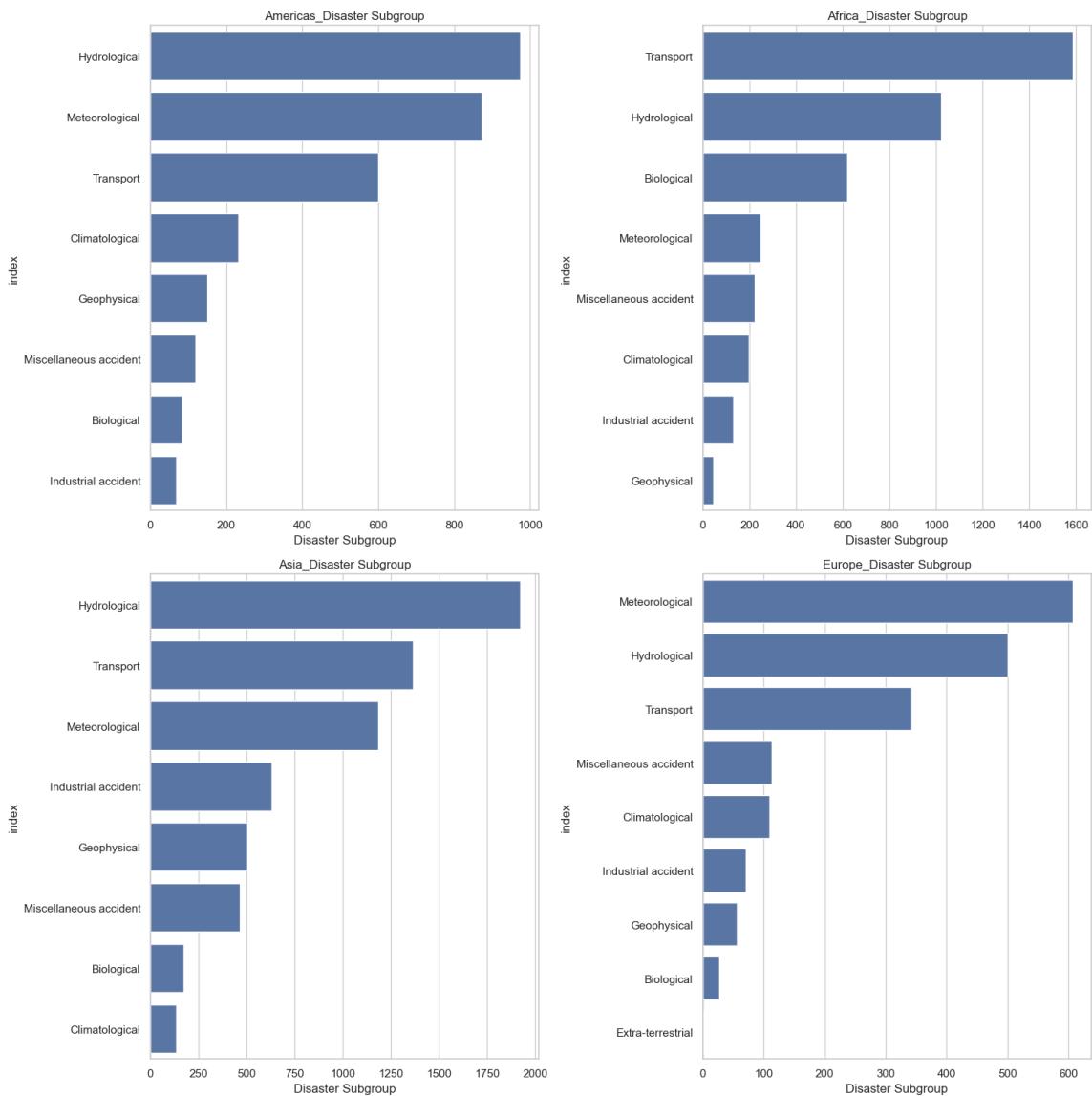
1-3 Disaster_subgroup by Region

Hydrological, Meteorlogical, Transport가 모든 대륙에서 가장 높은 빈도를 차지한다.

Hydrological은 공통적으로 모든 대륙에 큰 영향을 미친다.

Meteorlogical은 특히 Europe에서의 발생빈도가 높다.

Technological에 분류되는 Transport는 대륙의 크기, 전반적인 소득 수준 등과 연관이 있어보인다.



Region 단위로 얻어낸 Insight를 기반으로 Country 단위로 들어가서 더 세부적으로 분석해보자.

Region별로 대표적인 국가들을 뽑아서 분석한다면 더 좋겠지만, 현재 나는 동아시아 3국, Korea, Japan, China를 중점으로 보고 싶다. 따라서 Korea, Japan, China, USA를 중점으로 Asia와 America에 대해서 더 들여다 보자.

국가별 가설

Korea: 한국에서 살다보면 태풍 피해가 가장 많고, 교통사고도 많이나는데 이 둘의 빈도수가 가장 높지 않을까?

China: Asia 내에서의 industrial 인재는 중국에서 발생하지 않았을까? 규모도 크고 노동인구수도 많으니.

Japan: 환태평양에 위치해있으니, 당연히 지진의 발생빈도가 높겠지?

USA: 육로로 운송 많이하니 교통사고, cyclone이 자주 오니 폭풍에 대한 영향이 크지 않을까?

2. Country/ISO 별로 발생하는 Disaster의 특이사항이 있을까?

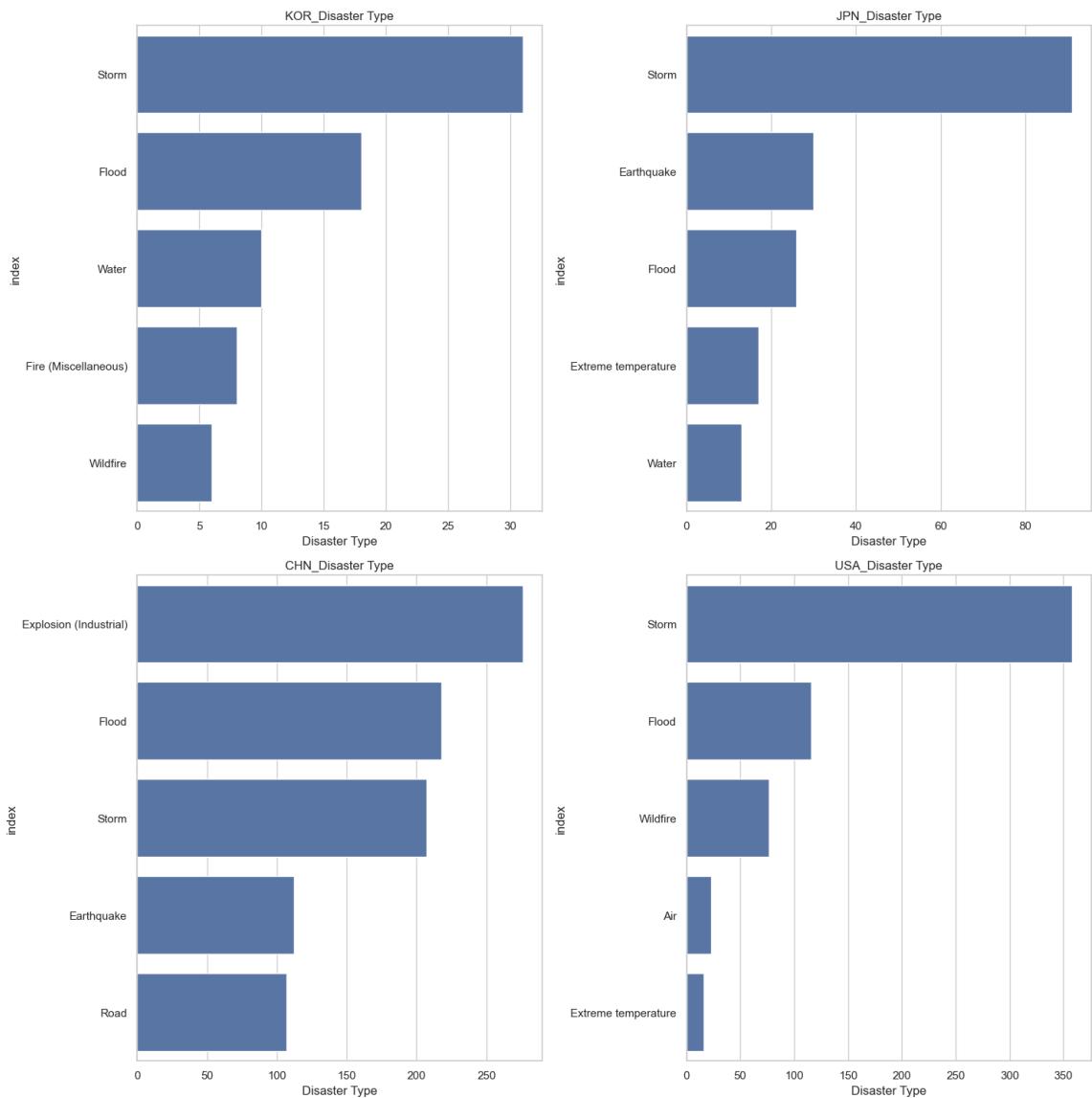
2-1. Disaster_type by Country

Korea: 가설대로 태풍, 수해가 많다. Storm, Flood, Wildfire는 언론에서도 많이 보도되어 예상가능했지만, 실제 집계로는 Water(선박 사고)가 3위로 예상외로 많음을 확인

China: 예상보다 Explosion(Industrial)의 비중이 높다. 공장 폭발 사고는 1회 별 피해 규모도 클 것으로 예상된다. 발생 빈도가 높으니, 피해자/피해액도 크지 않을까 추론해볼 수 있음. 천재지변보다 인재의 발생빈도가 높은 특이한 현상을 보이는 국가다.

Japan: 예상대로 Earthquake의 빈도가 높고, Extreme Temperature의 영향을 크게 받고 있음을 확인할 수 있다. 다만 많은 타 국가에서 Extreme Temperature 이 집계되지 않는 국가들이 있음을 고려해야한다.

USA: 가설대로 Storm, Flood가 가장 높은 비중을 차지함.



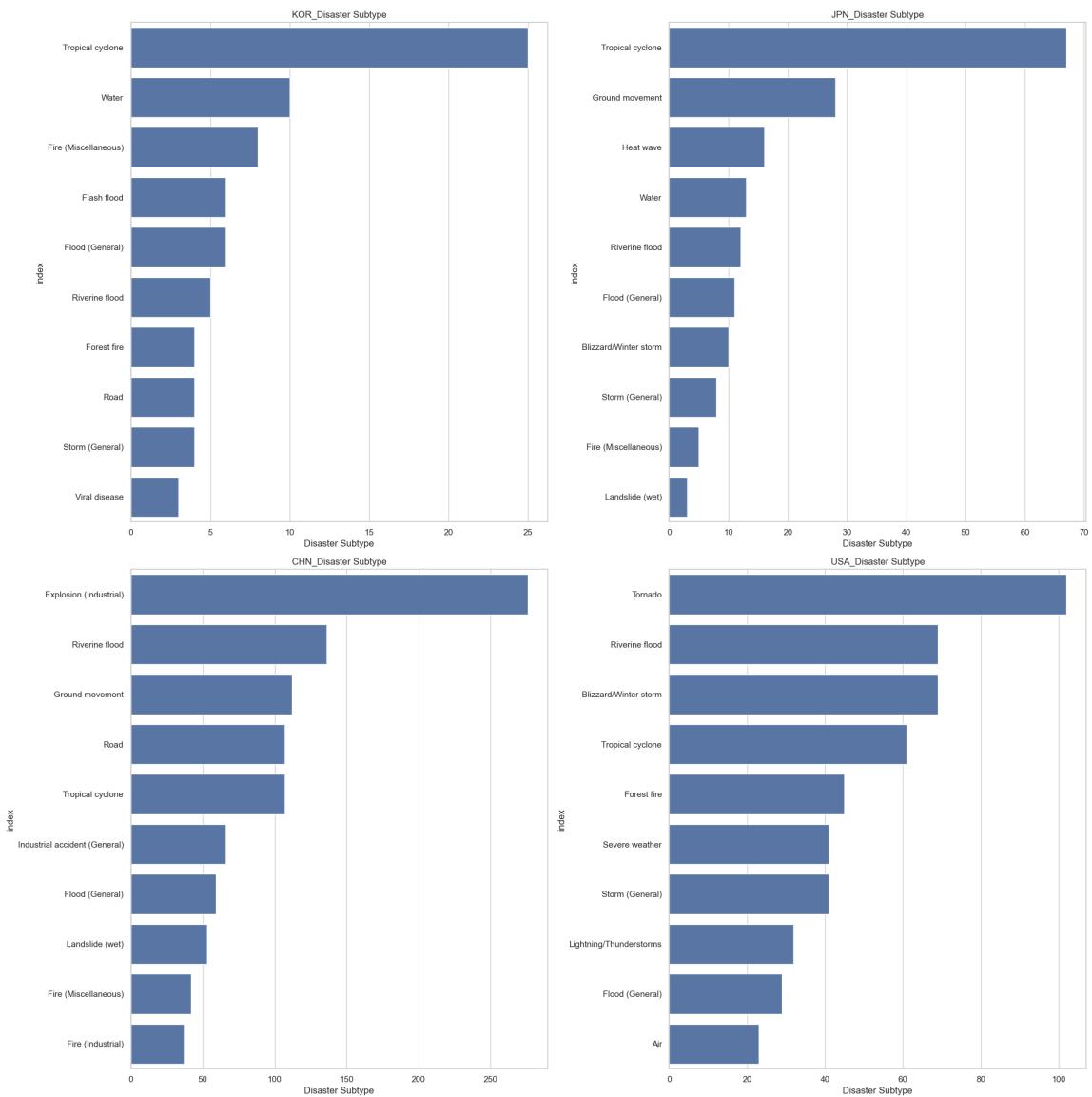
2-2. Disaster_Subtype by Country

KOR: 자연재해로는 태풍의 영향과 산불의 발생빈도가 높고, 선박사고의 발생비율이 높다.

JPN: 자연재해로는 태풍과 지각변동(지진)의 발생빈도가 높다.

CHN: 인재인 Industrial Explosion, Fire(Industrial)이 높은 비중을 차지한다. 자연재해로는 홍수, 지진이 가장 높은 영향을 미친다.

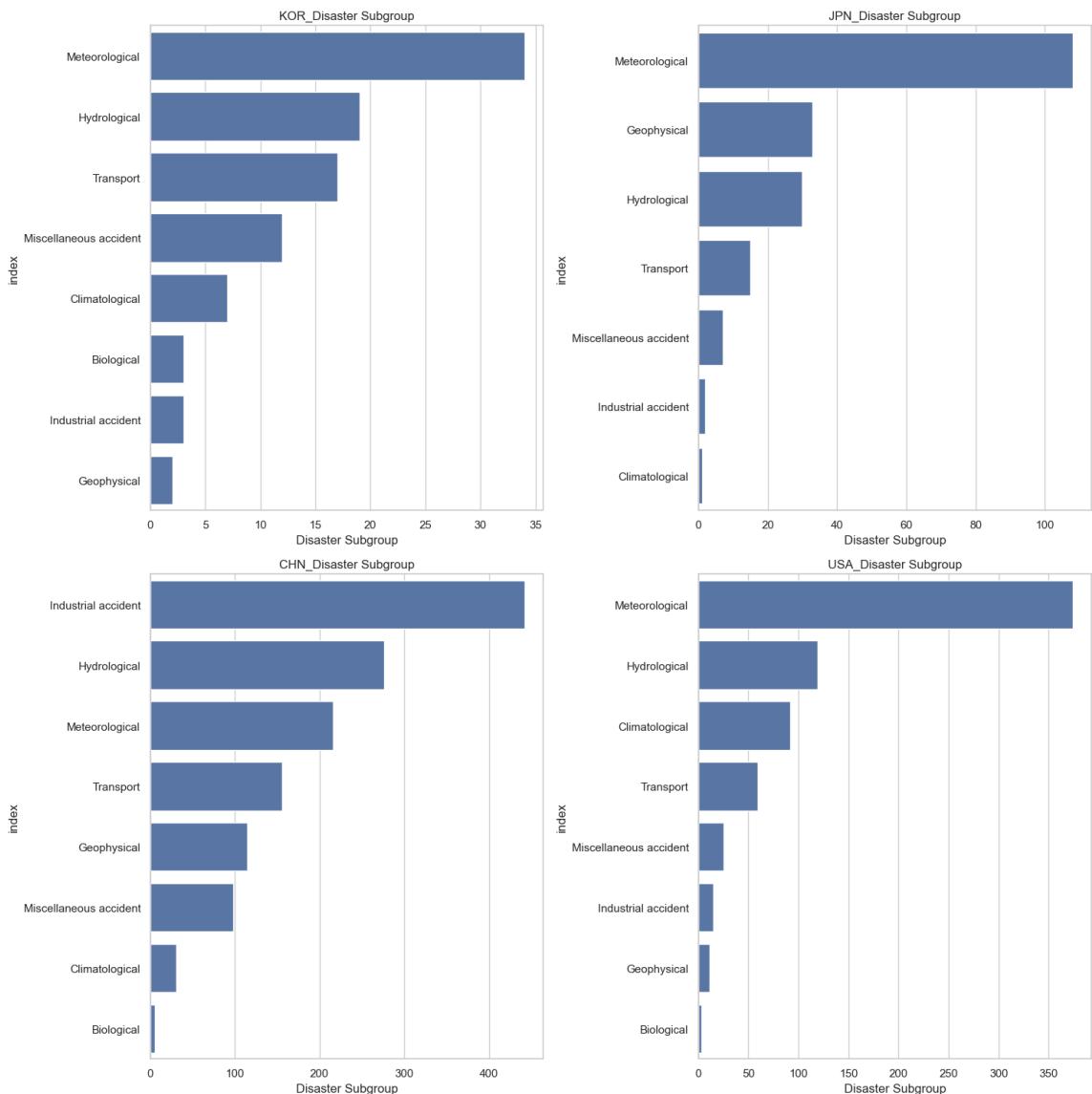
USA: 토네이도, 블리자드, 사이클론, 홍수 등 수속성 재해들, 그리고 산불이 가장 큰 영향을 미침



2-3. Disaster_subgroup by Country

KOR, JPN, USA 모두 Meteorlogical이 Subgroup 2위의 2배~4배의 발생빈도를 보인다.

그러나 CHN의 경우는 Industrial accident가 2위인 Hydrological의 2배를 보인다. 홍수피해의 경우, 많은 경우에 댐에 물을 가두고 물을 방류하는 과정에서 발생한다. 정확한 원인은 추가적인 분석이 필요하나 Hydrological도 완전한 자연재해라기보다는 어느 정도 수자원관리라는 시스템이 반영되기 때문에 technological로도 볼 수 있지 않을까란 생각이 든다.



▼ Longitudinal Approach

재앙의 종류 Group, subGroup, Type, subType별로 시간에 따른 빈도수의 추이를 확인해보자.

각 재앙별로 자주 발생하는 시기(계절 혹은 월) 이 있는가? 혹은 연도별로 추이는 어떻게 되는가?

1.1 어떤 시간 변수를 활용함이 적절한가?

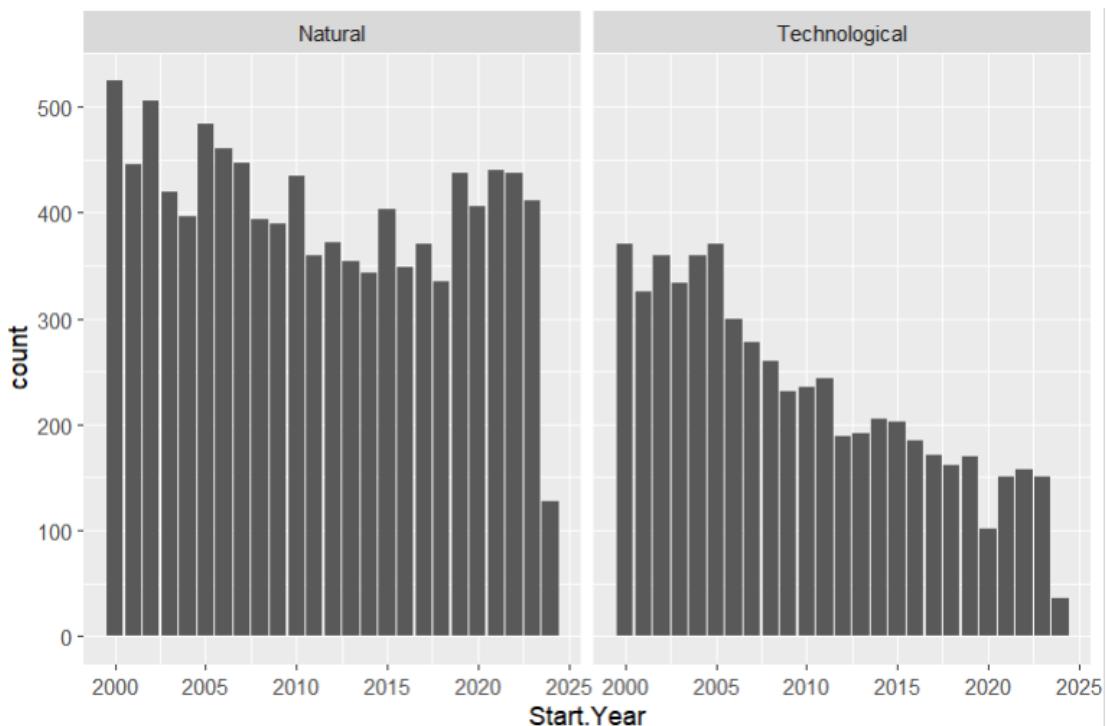
시간에 따른 빈도수를 disaster 별로 시각화해보자.

→ 재앙의 종류, 분류 등등에 따라 비교

Disaster.group → 2가지 종류가 있으며, 자연재해와 기술적 사고가 있다.

기술적 재해는 관련한 제도 및 법규에 따라 그 빈도수가 년도별로 감소할 것이라고 생각하며, 월별로 일정한 편균 빈도를 가질 것이라 예상한다.

```
data%>%
  group_by(Disaster.Group)%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Group)
```

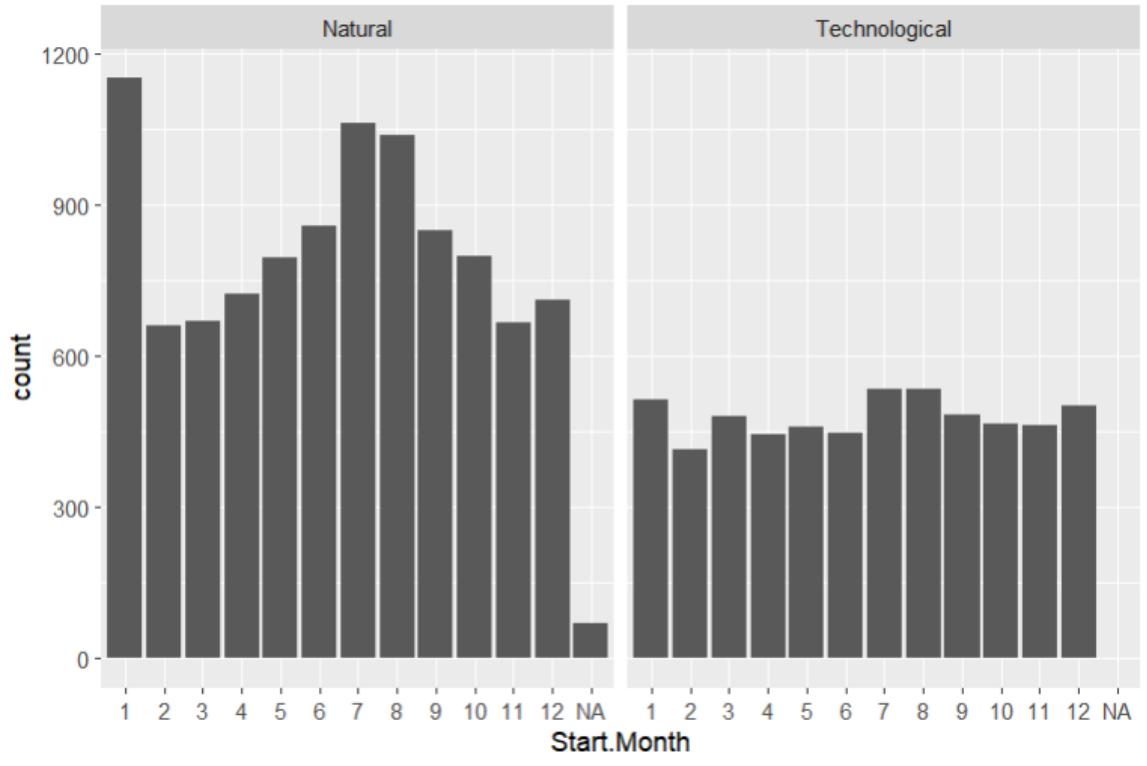


년도별로 빈도수를 시각화함. 기본적으로 자연재해가 더 높게 나타남을 확인.

자연재해의 경우 대부분 비슷하게 나타나지만 기술재해의 경우 점점 감소하는 추세를 보임

```
data%>%
  group_by(Disaster.Group)%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
```

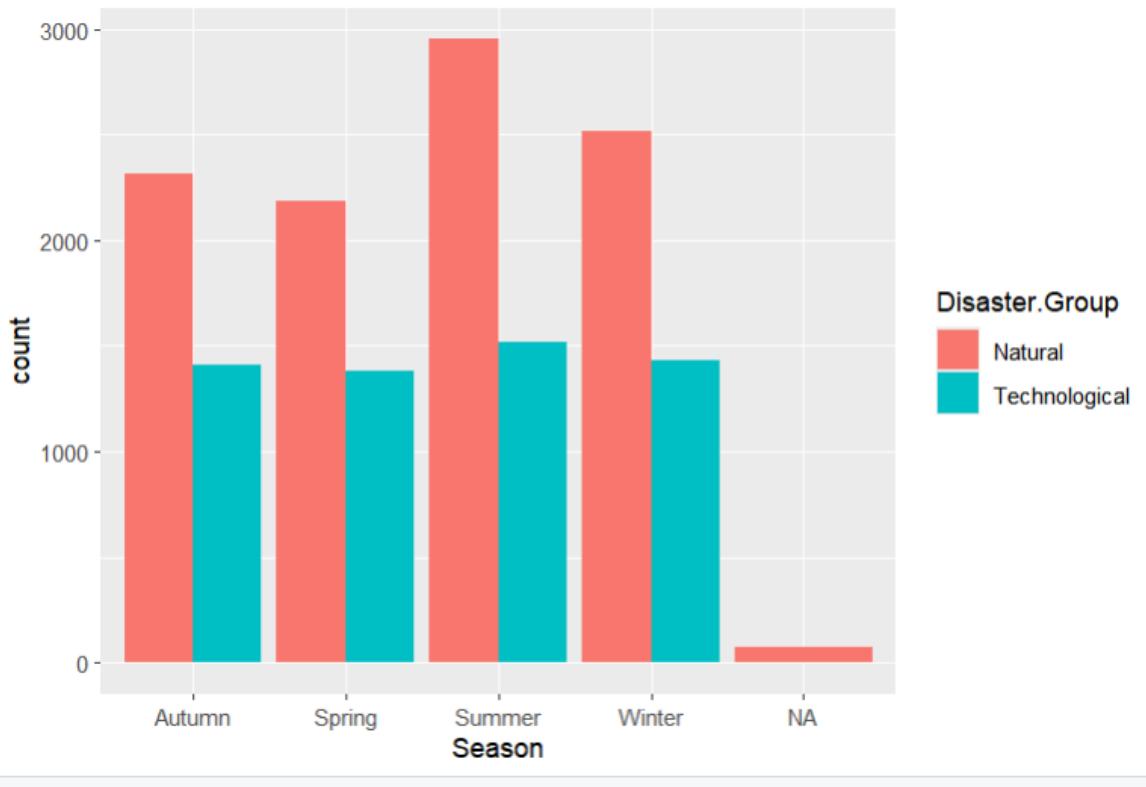
```
geom_bar() +  
facet_wrap(~Disaster.Group)
```



이번에는 월별로 자연재해 및 기술재해의 빈도를 시각화함.

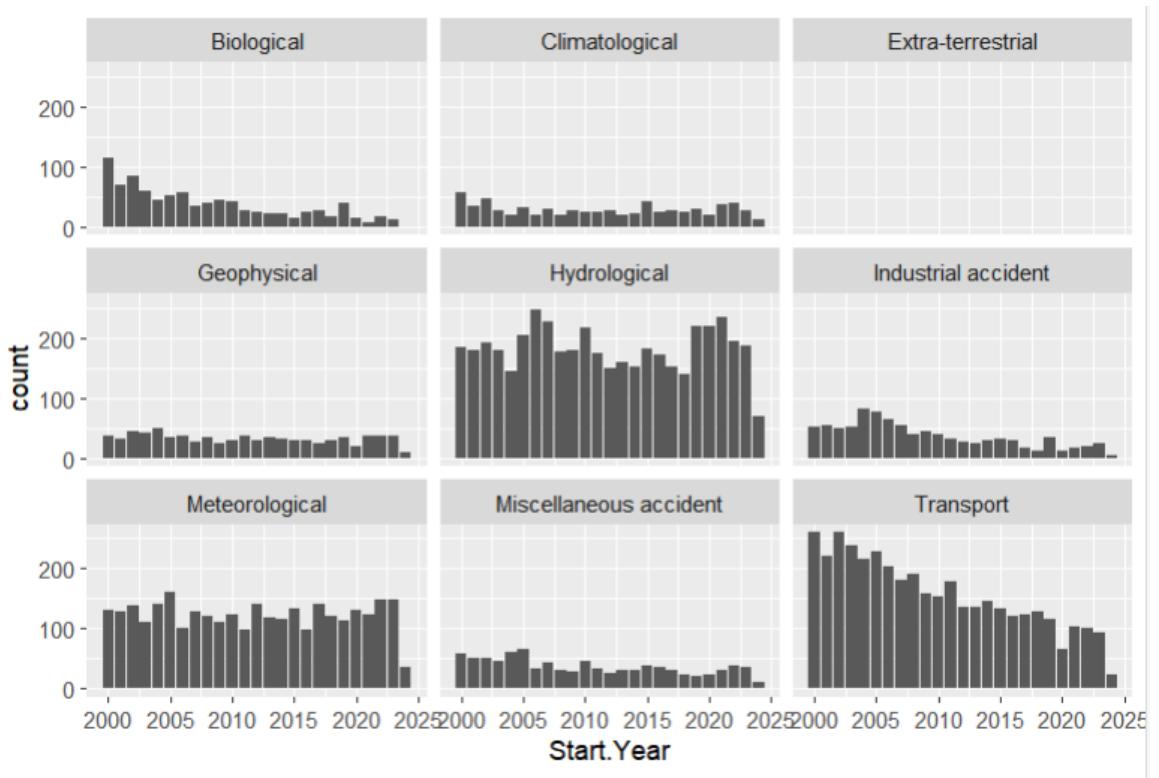
기술적 재해의 경우 자연환경의 영향을 받지 않으므로 거의 모든 달이 일정하게 나왔으나, 자연재해의 경우 특정 월에 더 치우쳐져 있음.

```
data %>%  
  mutate(Season = case_when(  
    Start.Month %in% c(12, 1, 2) ~ "Winter",  
    Start.Month %in% c(3, 4, 5) ~ "Spring",  
    Start.Month %in% c(6, 7, 8) ~ "Summer",  
    Start.Month %in% c(9, 10, 11) ~ "Autumn"  
  )) %>%  
  ggplot(aes(x = Season, fill = Disaster.Group)) +  
  geom_bar(position = "dodge")
```



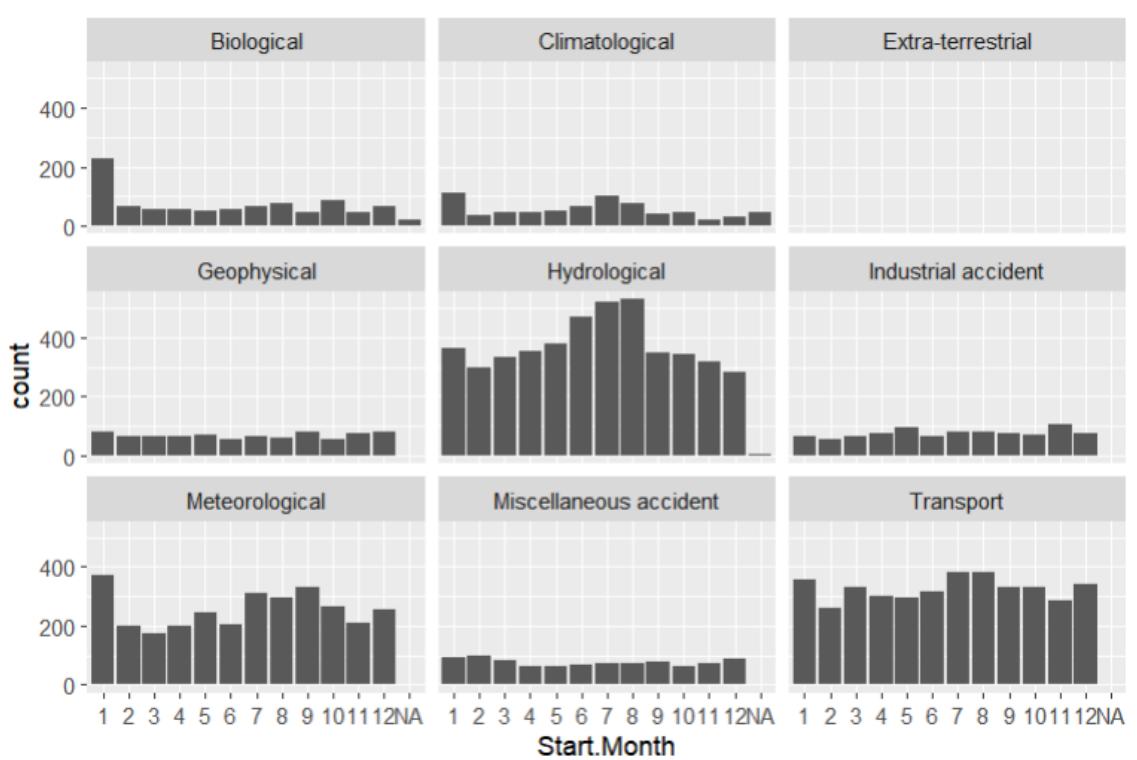
조금더 보기 쉽도록 계절별로 그룹화를 한 후 비교함. 여름과 겨울에 비교적 자연재해가 많이 발생함

```
data%>%
  group_by(Disaster.Subgroup)%>%
  ggplot(aes(x = Start.Year)) +
  geom_bar() +
  facet_wrap(~Disaster.Subgroup)
```



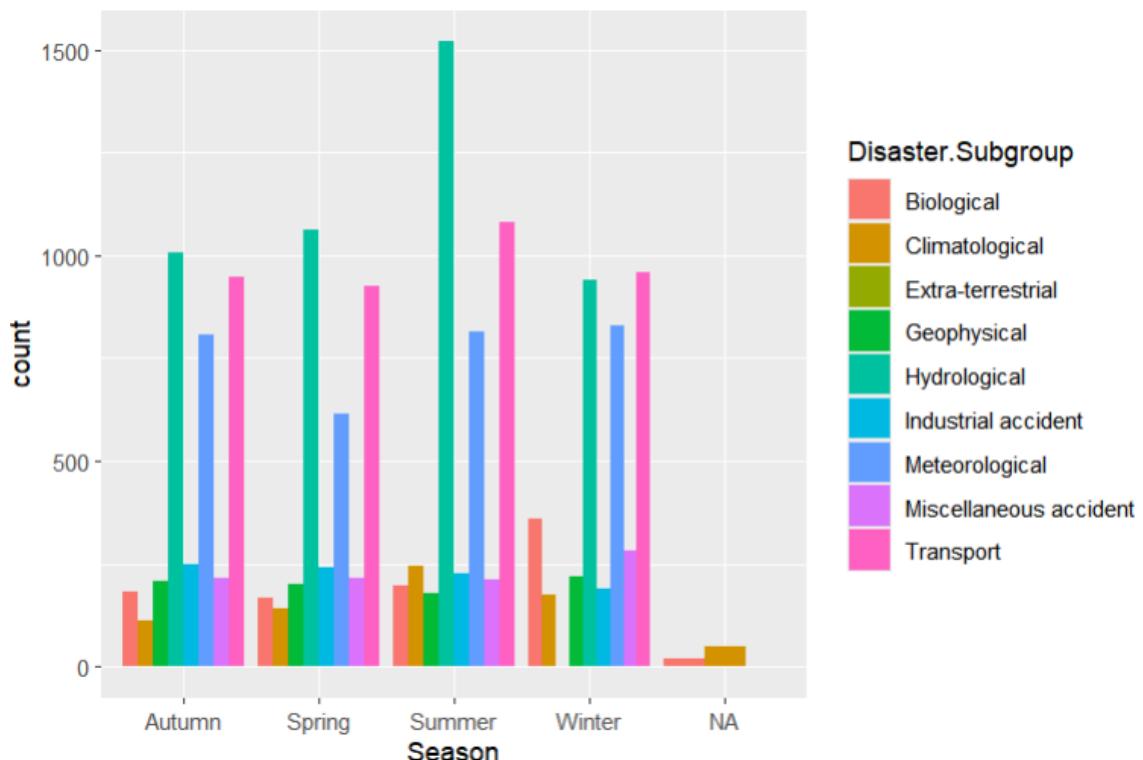
이번에는 년도별 재해 종류를 시각화함. 가장 빈도가 높아보이는 재해는 Hydrological과 Meterorological, Transport이다. 이중 Meterorological과 Hydrological은 년도 별로 고른 빈도를 보였고, Transport는 2000년대 초반에는 높았으나 점점 해가 지날 수록 감소하고 있음.

```
data%>%
  group_by(Disaster.Subgroup)%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Subgroup)
```



동일하게 월별로 다시 시각화를 진행함, 홍수의 경우 여름철 장마와 함께 높은 빈도를 보이며, 지진의 경우에는 그래도 각 월별로 고르긴 하지만 1월에 조금 더 많은 평균 빈도를 보임

```
data %>%
  mutate(Season = case_when(
    Start.Month %in% c(12, 1, 2) ~ "Winter",
    Start.Month %in% c(3, 4, 5) ~ "Spring",
    Start.Month %in% c(6, 7, 8) ~ "Summer",
    Start.Month %in% c(9, 10, 11) ~ "Autumn"
  )) %>%
  ggplot(aes(x = Season, fill = Disaster.Subgroup)) +
  geom_bar(position = "dodge")
```



계절별로 나누어 표현함

이번에는 disaster subGroup과 그 하위 분류인 Type, subType에 대해 시각화를 진행함

▼ Climatological기후 관련 재해

기후와 관련된 재해는 연도별 빈도수에서는 지구 온난화와 연관지어 그 빈도수가 늘어날 것이라고 예상한다.

```
data %>%
  select(Disaster.Type, Disaster.Subgroup) %>%
  filter(Disaster.Subgroup == "Climatological") %>%
  distinct()
```

```

Disaster.Type Disaster.Subgroup
1           Drought   climatological
2        Wildfire   climatological
3 Glacial lake outburst flood   climatological
> |

```

Drought (가뭄)

Wildfire (산불)

Glacial lake outburst flood (빙하호 폭발 홍수)

subGroup인 climatological에는 3개의 Type이 있음

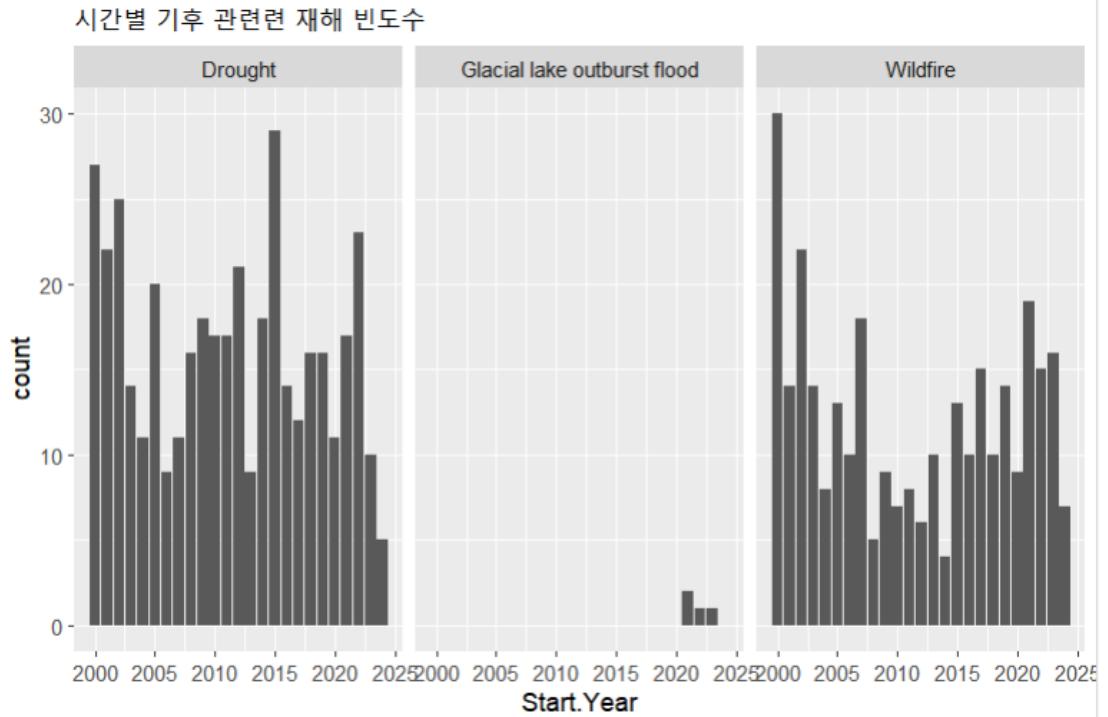
시간별 기후 관련 재해 정리

년도-빈도

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Climatological")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "시간별 기후 관련 재해 빈도수")

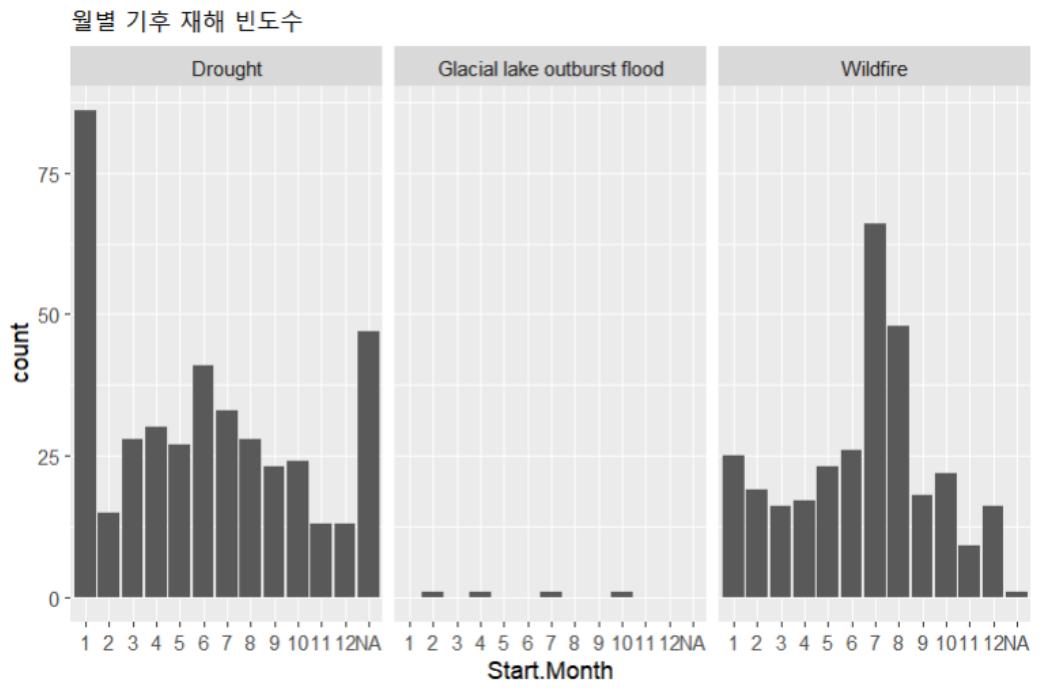
```



년도별 기후 관련 재해 빈도를 보면, 가뭄과 산불의 빈도수가 높음을 알 수 있음. 또한 빙하호 폭발 호수는 근래 2020년대 이후로 등장하여 기후변화와 연관성을 찾을 필요가 있어보임.

이번에는 월별로 특이사항을 확인.

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Climatological")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "월별 기후 재해 빈도수")
```



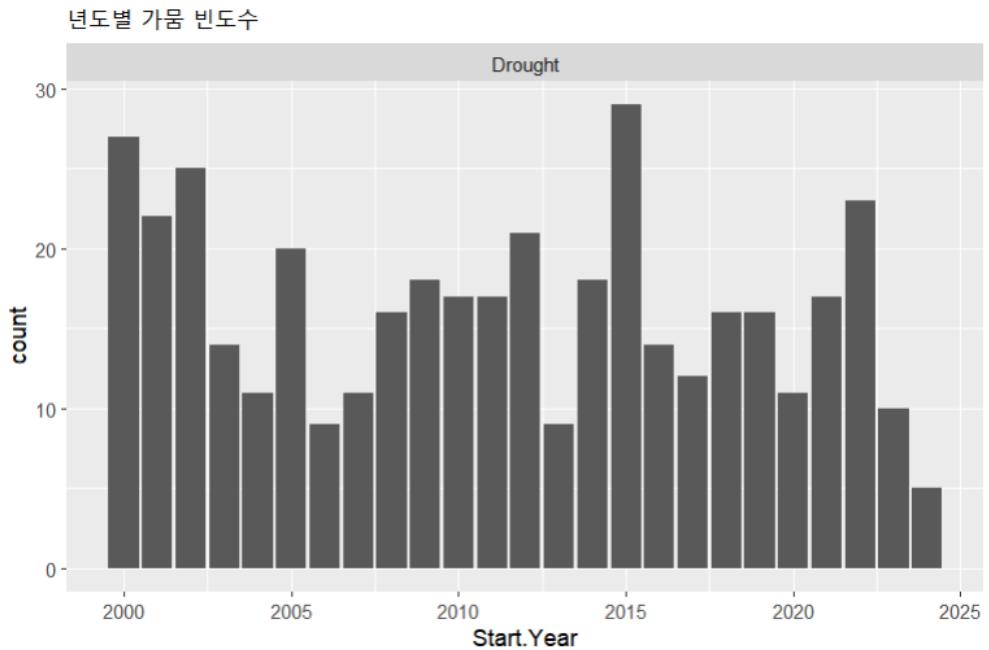
특이한 점은 산불에 대한 재해가 건조한 가을, 겨울이 아닌 여름에 많이 나타났고, 가뭄 피해가 1월에 많이 나타남.

이번에는 Type안에서도 세부 Type별 빈도수를 체크함

▼ Drought

먼저 가뭄의 경우에는 세부 항목 역시 가뭄으로 하나뿐이라 위에서 진행한 시각화 자료와 동일하다.

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Drought")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 가뭄 빈도수")
```



▼ Wildfire

다음으로 산불과 관련된 세부 항목과 빈도수를 확인

산불의 세부 항목으로는

Disaster.Type	Disaster.Subtyp
e	
1	Wildfire Land fire (Brush, Bush, Pastur
e)	
2	Wildfire Forest fir
e	
3	Wildfire (Genera
l)	

Land fire (Brush, Bush, Pasture) - 덤불, 관목지, 초지 화재

Forest fire - 산림 화재

Wildfire (General) - 일반적인 산불

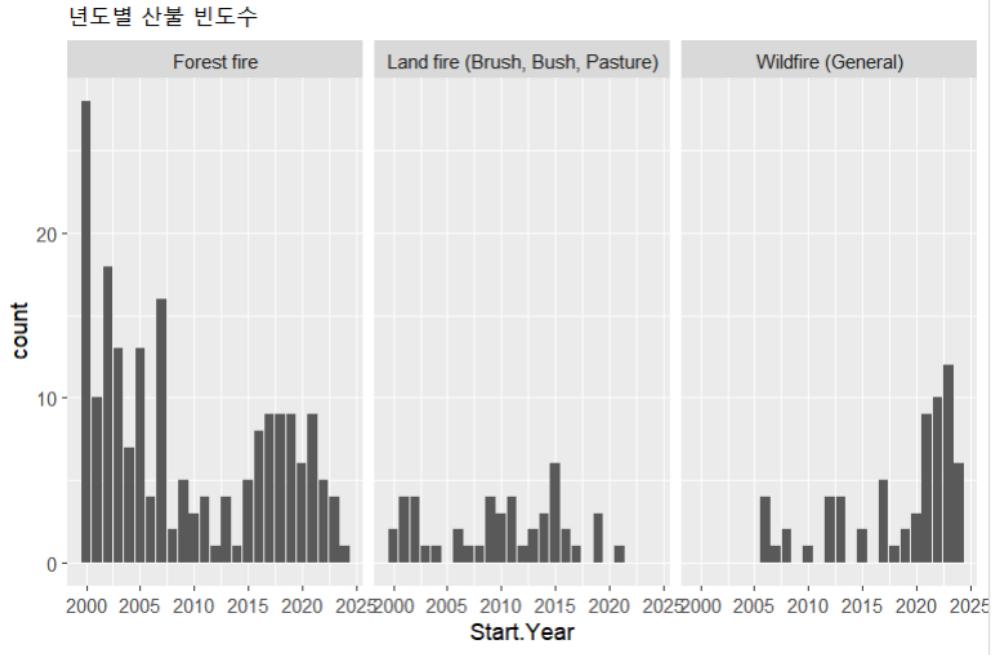
이 있으며 이를 년도별로 시각화하면

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Wildfire")%>%
```

```

ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 산불 빈도수")

```



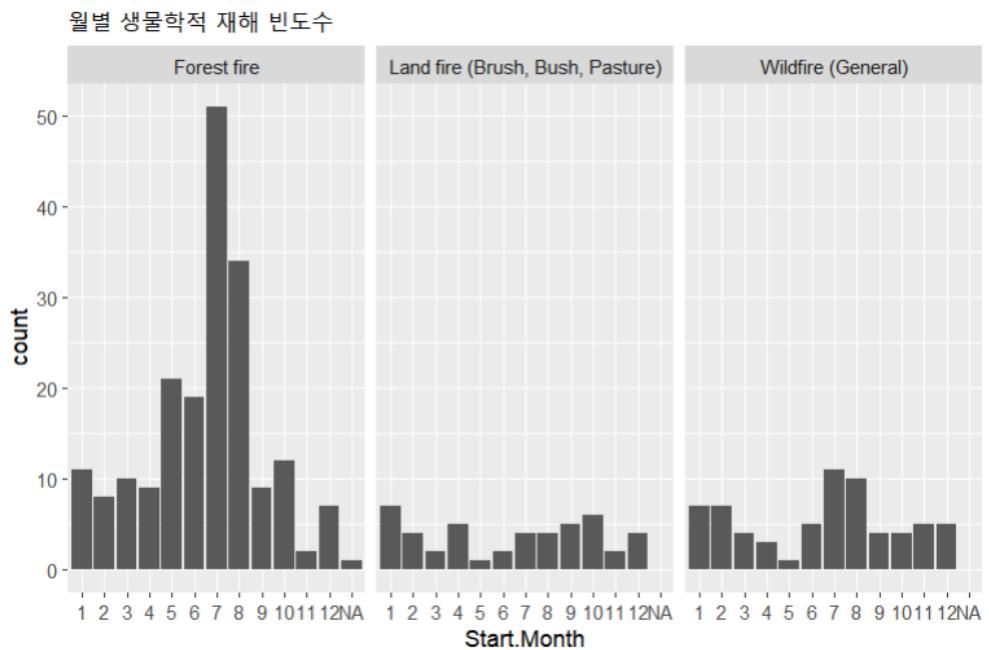
산림 산불의 경우 가장 빈도수가 많았지만 최근들어 빈도가 줄었고, 일반적인 산불이 오히려 빈도가 증가함을 확인할 수 있음.

월별로도 체크해본다면

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Wildfire")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "월별 산불 재해 빈도수")

```



다음과 같이 산림 산불의 경우 여름철에 주로 몰려있음을 확인할 수 있음.

▼ Transport 교통 관련 재해

교통 관련 재해에서는

```
data %>%
  select(Disaster.Type, Disaster.Subgroup) %>%
  filter(Disaster.Subgroup == "Transport") %>%
  distinct()
```

	Disaster.Type	Disaster.Subgroup
1	Road	Transport
2	Rail	Transport
3	Air	Transport
4	Water	Transport
5		

Road (도로 사고)

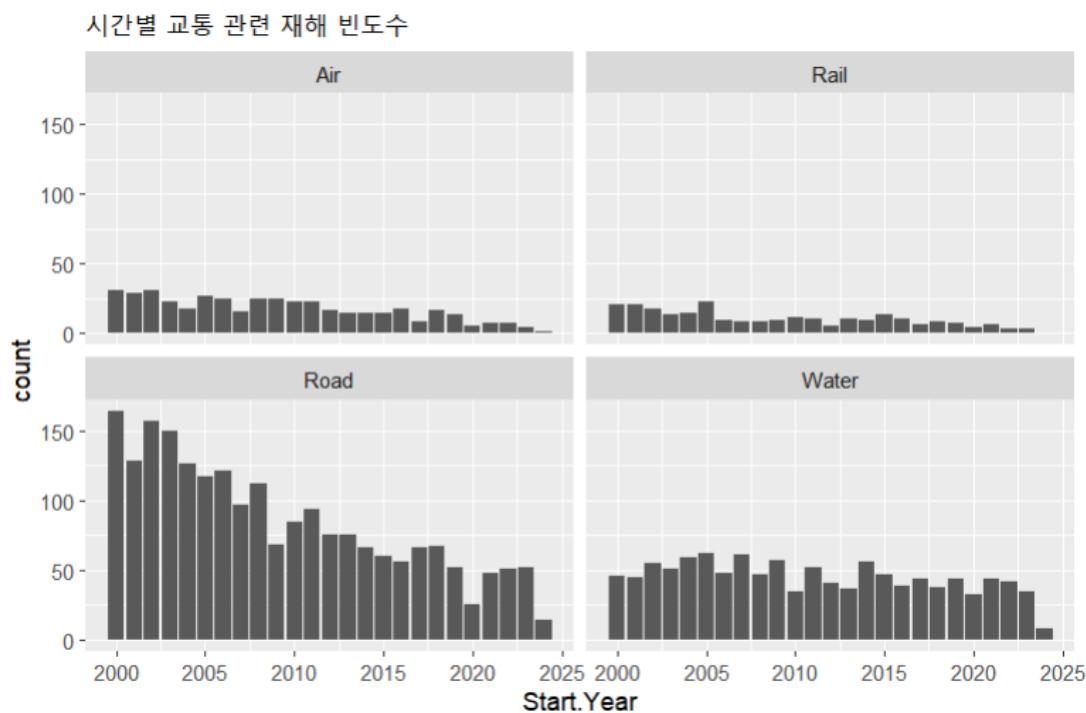
Rail (철도 사고)

Air (항공 사고)

Water (수상 사고)

이렇게 4가지 유형이 있음

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Transport")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Type) +
  labs(title = "시간별 교통 관련 재해 빈도수")
```



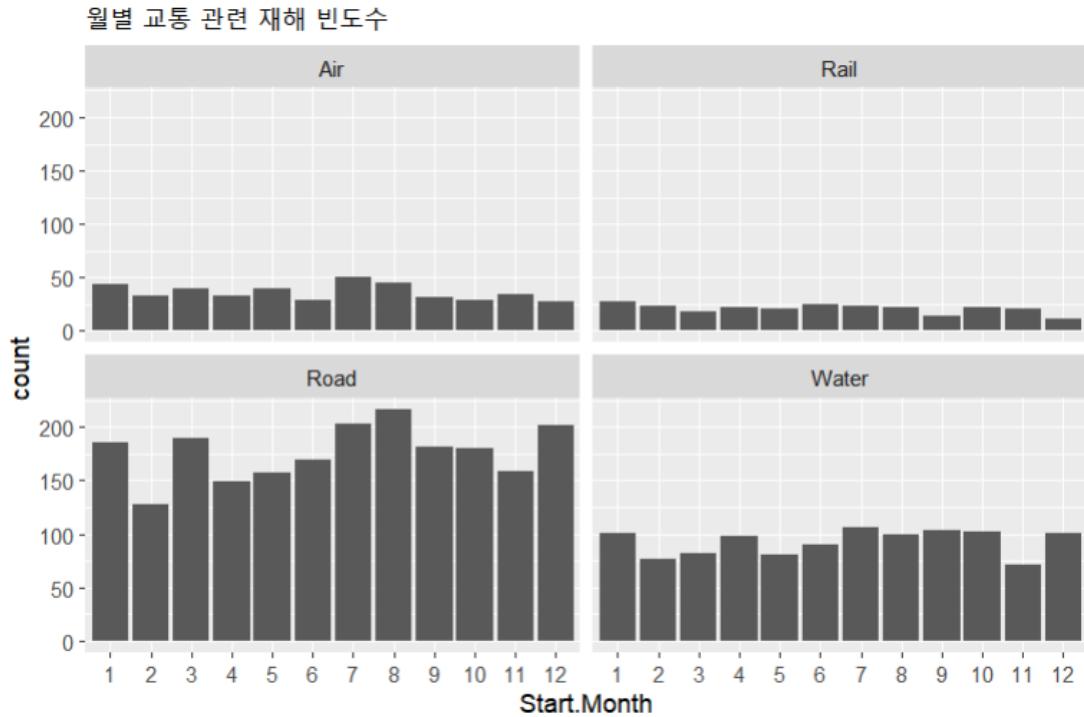
교통 관련 재해의 경우 자연재해와는 달리 환경과 기후 등의 조건에 따라 빈도수가 몰리고 하지는 않음. 그러나 4가지 유형 모두 전반적으로 빈도수가 갈수록 감소하는 것에는 관련된 법규나 제약이 생기며 빈도수를 줄인다고 할 수 있음

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Transport")%>%
  mutate(Start.Month = factor(Start.Month))%>%
```

```

ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "월별 교통 관련 재해 빈도수")

```



월별로 빈도수를 보았을때는 그 빈도에서 큰 차이를 보이지는 않음 대신 도로사고와 수상사고의 빈도가 항공이나 철도 사고에 비해 큰 것으로 보아 이용자 수와 연관시키는 것이 가능함

이번에는 subType별 빈도수를 확인

```

Disaster.Type Disaster.Subtype
1           Air          Air

```

```

Disaster.Type Disaster.Subtype
1           Rail         Rail

```

```
Disaster.Type Disaster.Subtype  
1 Road Road
```

```
Disaster.Type Disaster.Subtype  
1 Water Water
```

교통사고의 경우에는 세부 Type들이 기존 Type과 동일하므로 그 빈도수는 Type의 빈도와 같음.

▼ Hydrological (수문학적 재해)

```
data %>%  
  select(Disaster.Type, Disaster.Subgroup) %>%  
  filter(Disaster.Subgroup == "Hydrological") %>%  
  distinct()
```

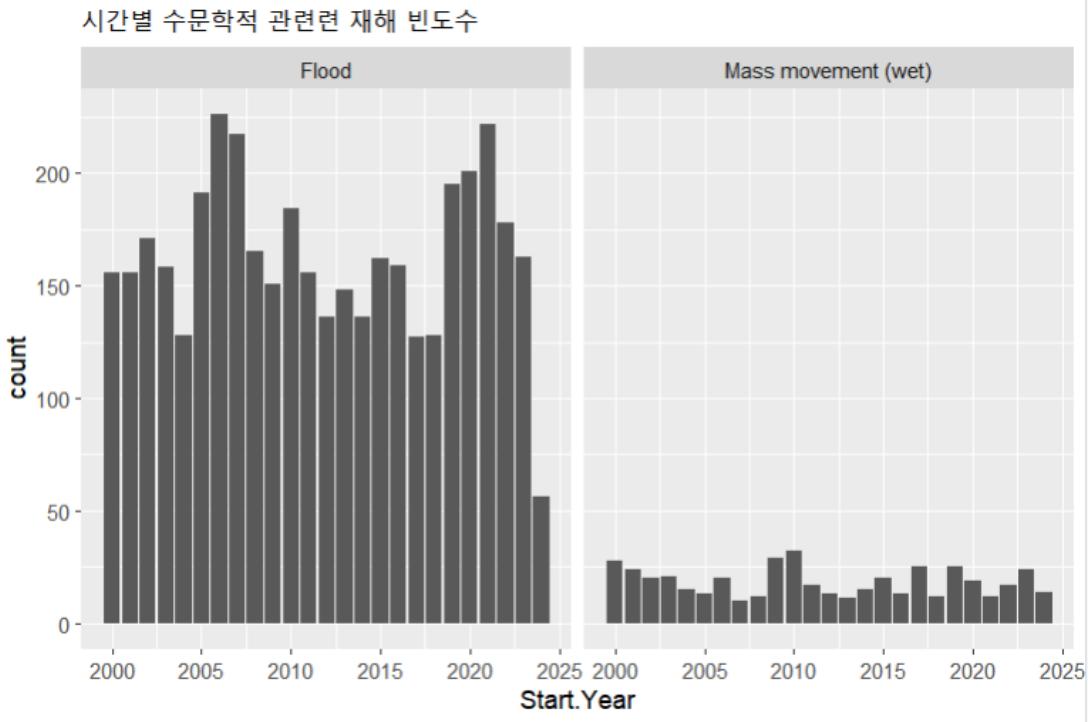
```
Disaster.Type Disaster.Subgroup  
1 Flood Hydrological  
2 Mass movement (wet) Hydrological  
> |
```

Flood (홍수)

Mass movement (wet) (대규모 이동 - 습한 조건)

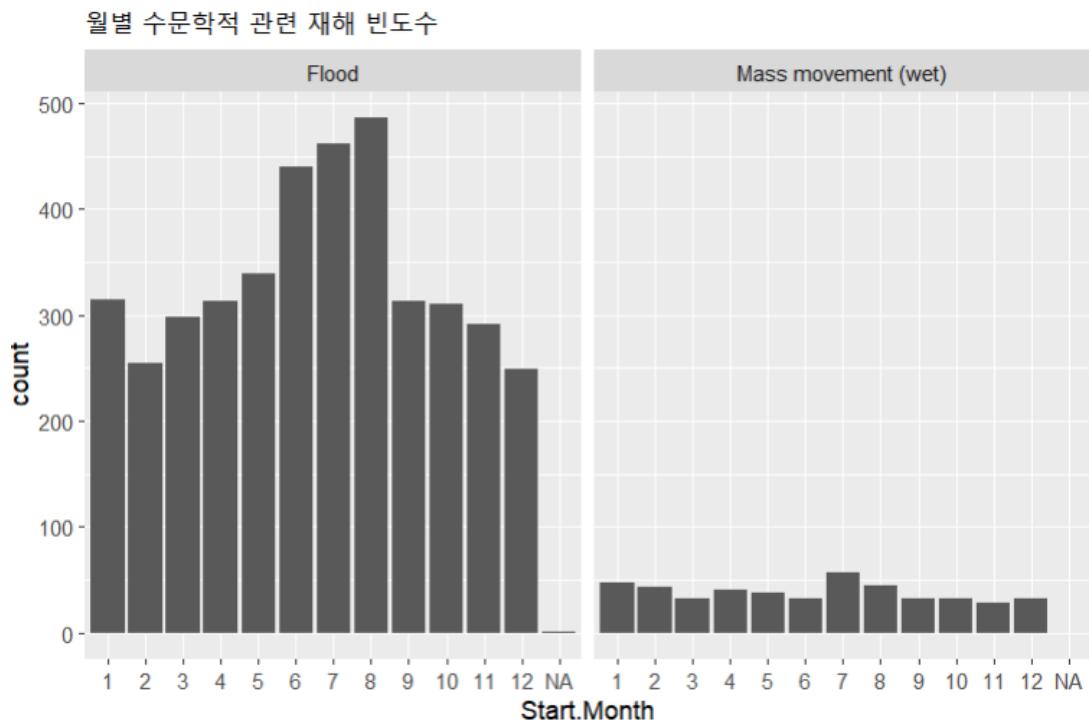
수해에는 2가지 유형이 있음.

```
data%>%  
  group_by(Disaster.Group)%>%  
  filter(Disaster.Subgroup == "Hydrological")%>%  
  ggplot(aes(x = Start.Year)) +  
  geom_bar() +  
  facet_wrap(~Disaster.Type) +  
  labs(title = "시간별 수문학적 관련된 재해 빈도수")
```



년도별로 빈도를 표시했을 때 홍수의 빈도가 굉장히 많음을 확인할 수 있음

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Hydrological")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "월별 수문학적 관련 재해 빈도수")
```



월별 수문학적 재해를 확인했을 때, 주로 6,7,8월 장마철과 함께 빈도가 올라감을 확인할 수 있음

수문학적 재해 역시 세부Type별로 종류를 찾고 빈도를 조사

▼ Flood

Disaster.Type	Disaster.Subtype
1	Flood Riverine flood
2	Flood Flash flood
3	Flood Coastal flood
4	Flood Flood (General)

가장 먼저 홍수에는 4 종류가 있으며,

Riverine flood-하천 홍수

Flash flood- 돌발 홍수

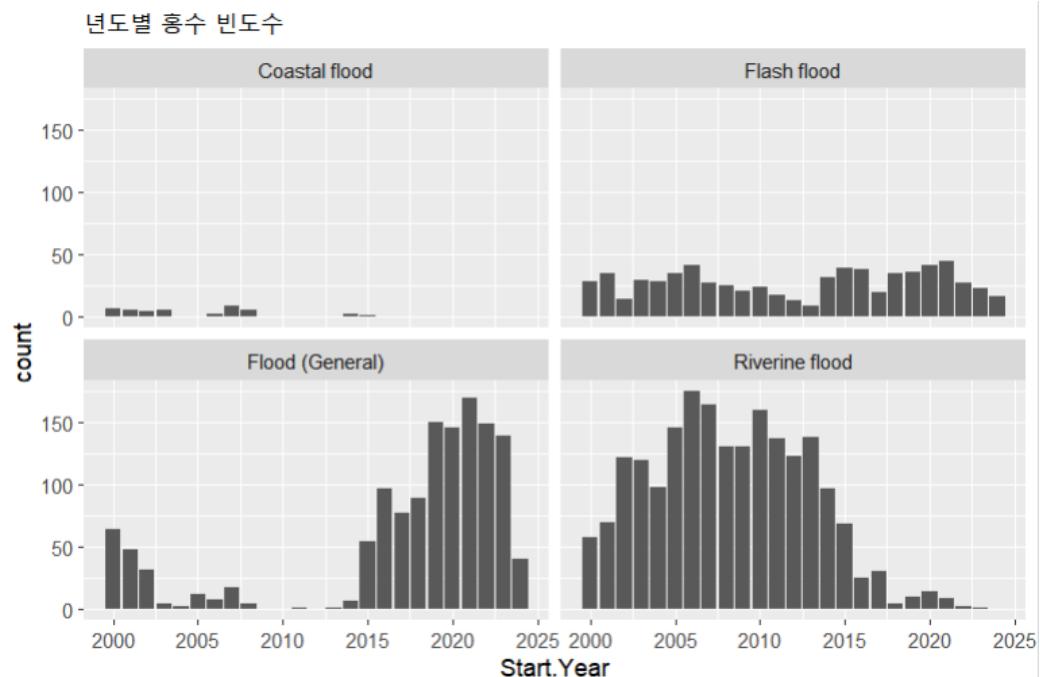
Coastal flood-해안 홍수

Flood (General)-일반적인 홍수로 구분됨

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Flood")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 홍수 빈도수")

```



년도별 홍수 빈도수를 비교해 보았을 때 하천홍수의 경우 2000년대 초반부터 2010년도 중반까지 가장 많은 빈도를 가졌으며 최근에는 그 수가 줄었지만 일반적인 홍수의 경우 반대로 최근들어 급격히 그 빈도가 증가함

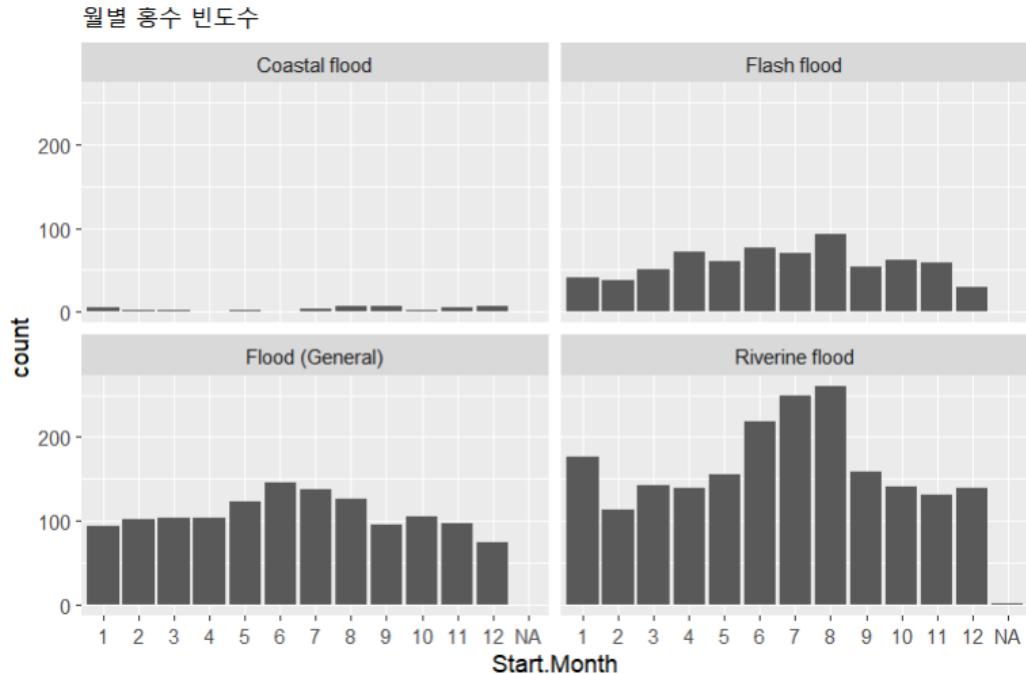
이번에는 월별로 확인

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Flood")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()

```

```
facet_wrap(~Disaster.Subtype)+  
  labs(title = "월별 홍수 빈도수")
```



▼ Mass movement (wet)

다음으로는 Mass movement (wet) 산사에 대한 종류로는

Disaster.Type	Disaster.Subtype
1 Mass movement (wet)	Landslide (wet)
2 Mass movement (wet)	Mudslide
3 Mass movement (wet)	Avalanche (wet)
4 Mass movement (wet)	Sudden Subsidence (wet)
5 Mass movement (wet)	Rockfall (wet)

Landslide (wet) - 습한 조건에서의 산사태

Mudslide - 진흙사태

Avalanche (wet) - 습한 조건에서의 눈사태

Sudden Subsidence (wet) - 습한 조건에서의 급격한 침하

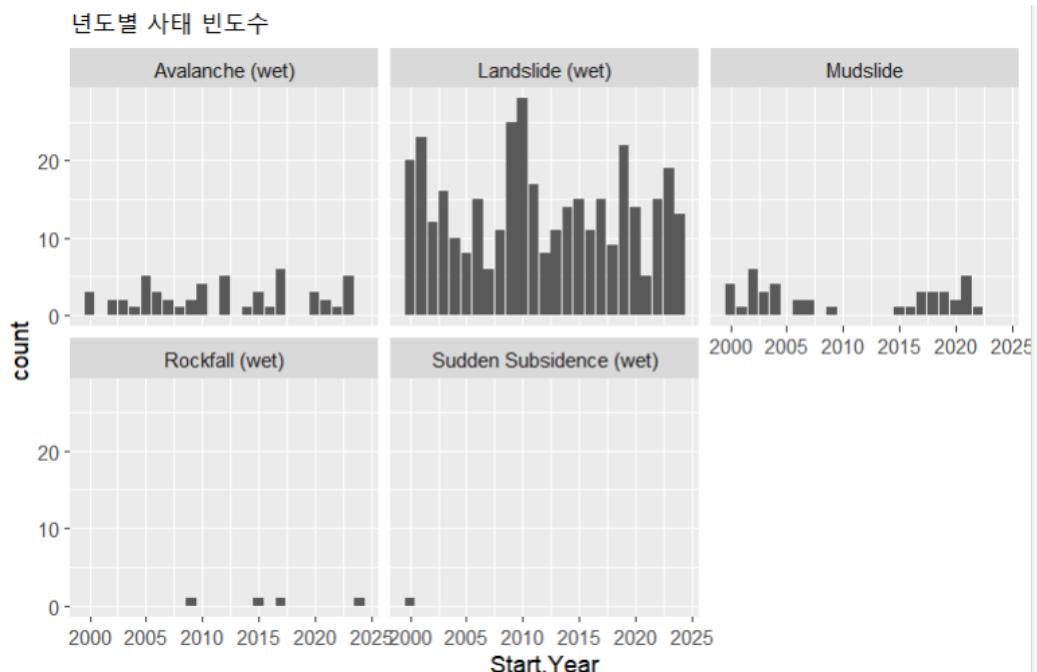
Rockfall (wet) - 습한 조건에서의 낙석

가 있으며 이번에도 년도별, 월별 빈도수를 시각화

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Mass movement (wet)")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 사태 빈도수")

```

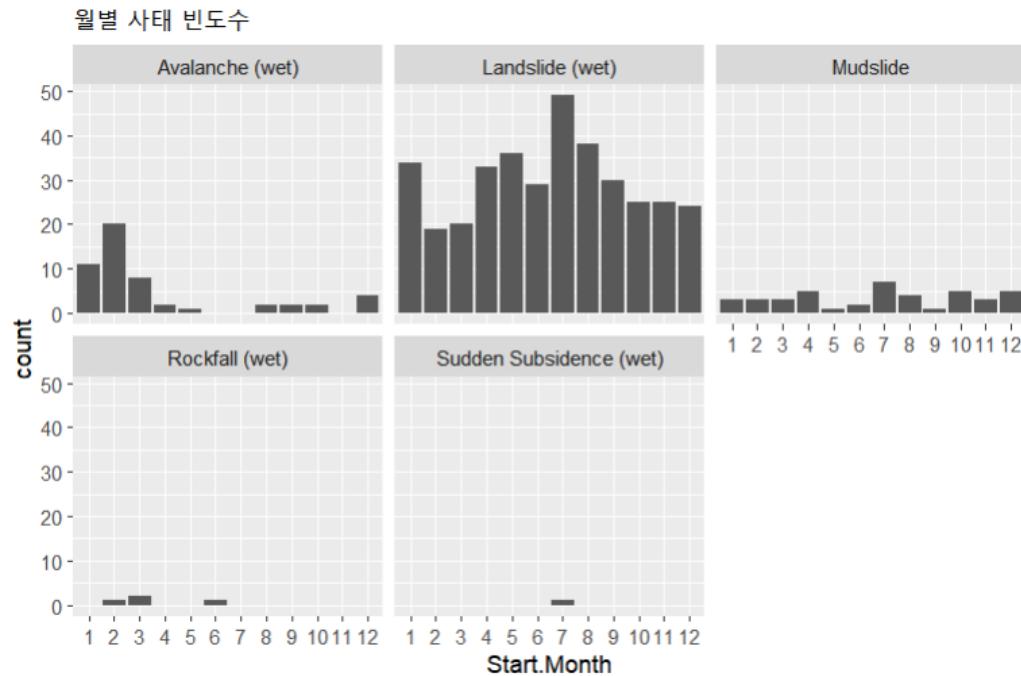


년도별 사태 종류에서는 산사태의 빈도가 가장 많았음을 확인 가능함.

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Mass movement (wet)")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "월별 사태 빈도수")

```



▼ Meteorological (기상 관련 재해)

```
data %>%
  select(Disaster.Type, Disaster.Subgroup) %>%
  filter(Disaster.Subgroup == "Meteorological") %>%
  distinct()
```

	Disaster.Type	Disaster.Subgroup
1	Extreme temperature	Meteorological
2	Storm	Meteorological

Storm (폭풍)

Extreme temperature (극한 온도)

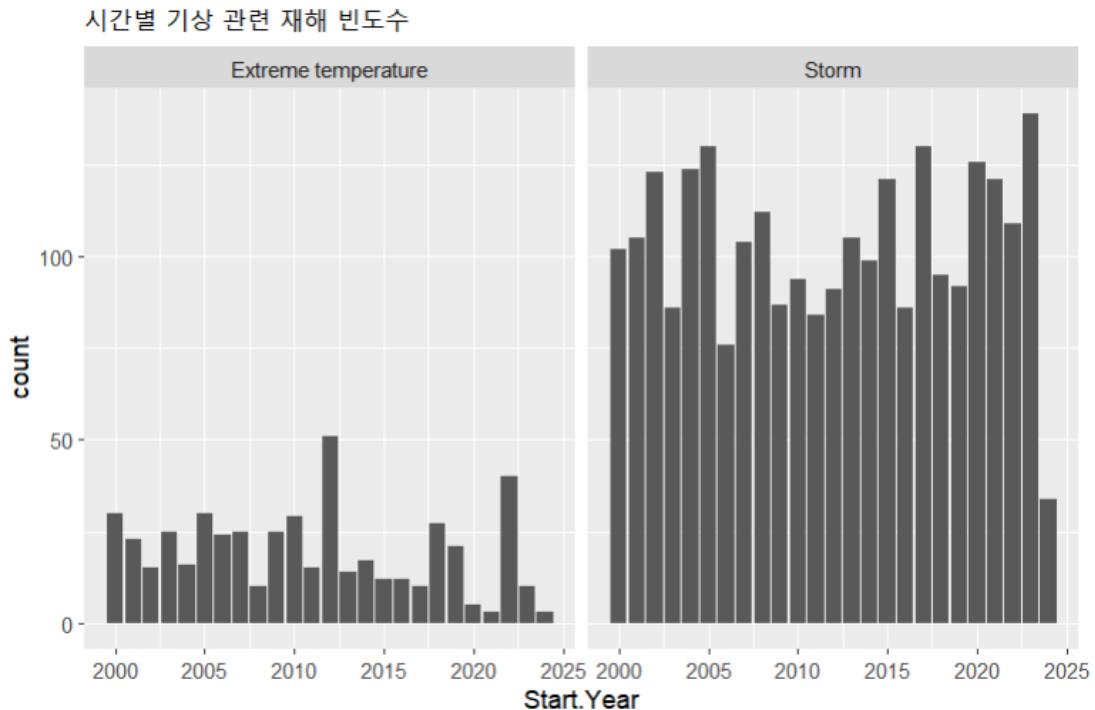
기상 관련 재해로는 폭풍과 극한 온도가 있음

```
data%>%
  group_by(Disaster.Group)%>%
```

```

filter(Disaster.Subgroup == "Meteorological")%>%
ggplot(aes(x = Start.Year))+
geom_bar()+
facet_wrap(~Disaster.Type) +
labs(title = "시간별 기상 관련 재해 빈도수")

```

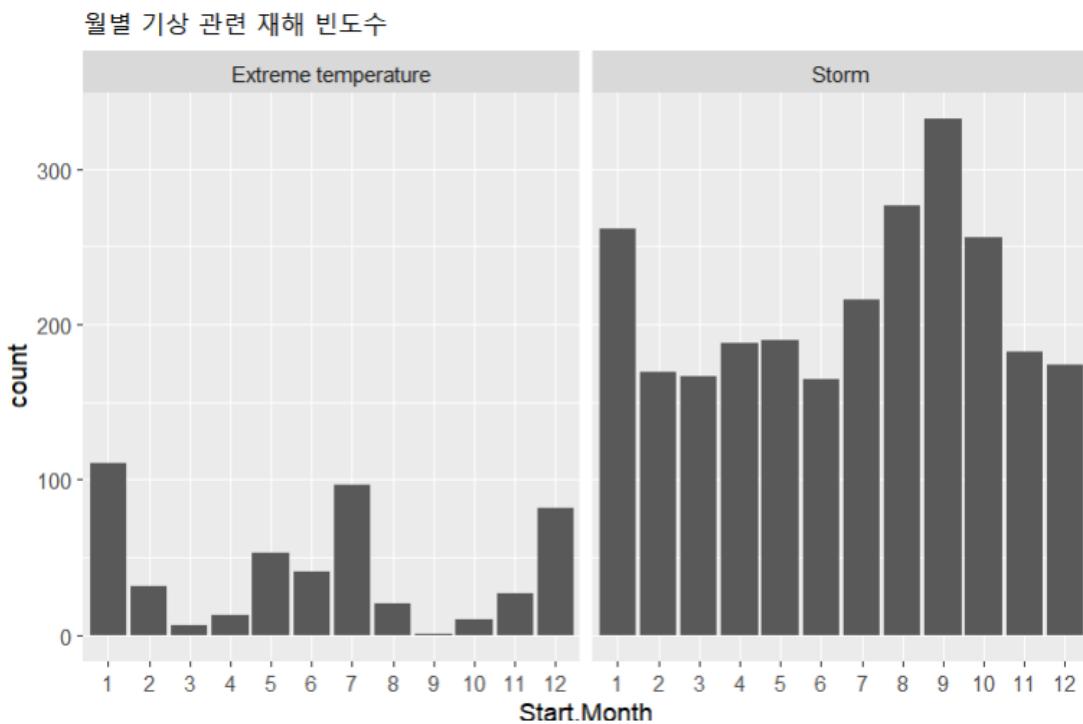


각각의 빈도수는 폭풍의 빈도가 굉장히 높았고, 폭풍의 경우 조금씩 증가하는 추세를 보임

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Meteorological")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Type) +
  labs(title = "월별 기상 관련 재해 빈도수")

```



극한 온도의 경우 겨울과 여름에 많이 발생하며 12월~1월에 발생하는 극한 온도와 7월에 많이 발생하는 극한 온도의 조건이 다를 것임.

이번에도 Type별 세부 항목을 시각화

▼ Extreme temperature

Disaster.Type	Disaster.Subtype
1 Extreme temperature	Cold wave
2 Extreme temperature	Heat wave
3 Extreme temperature	Severe winter conditions

Cold wave - 한파

Heat wave - 폭염

Severe winter conditions - 혹독한 겨울 조건

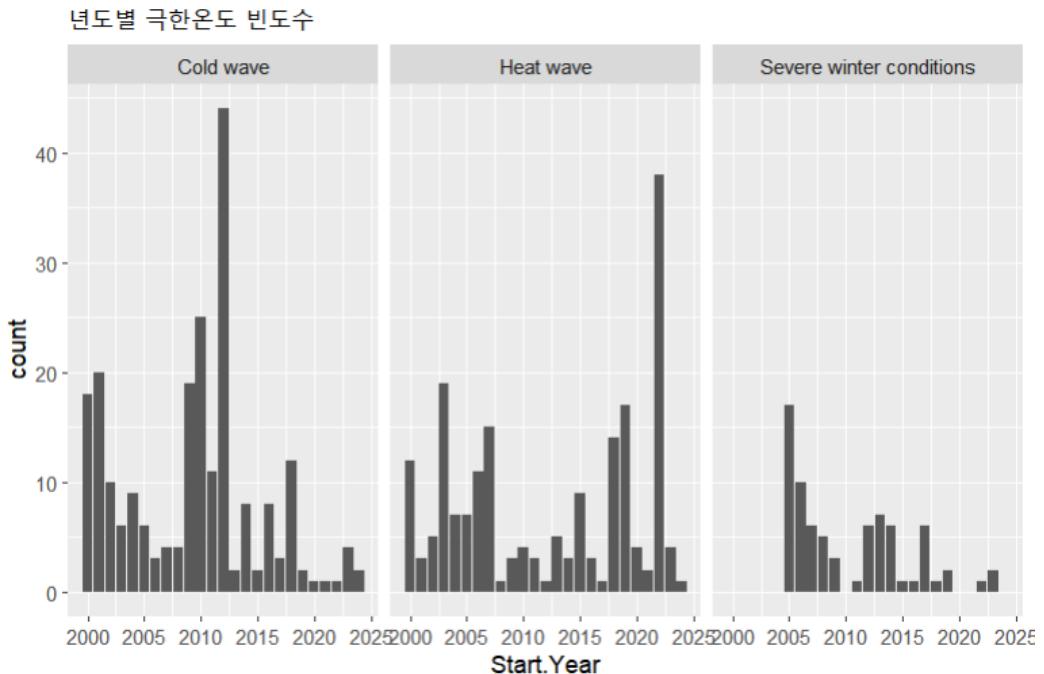
이상 세 개의 세부Type이 있음

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Extreme temperature")%
>%
```

```

ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 극한온도 빈도수")

```

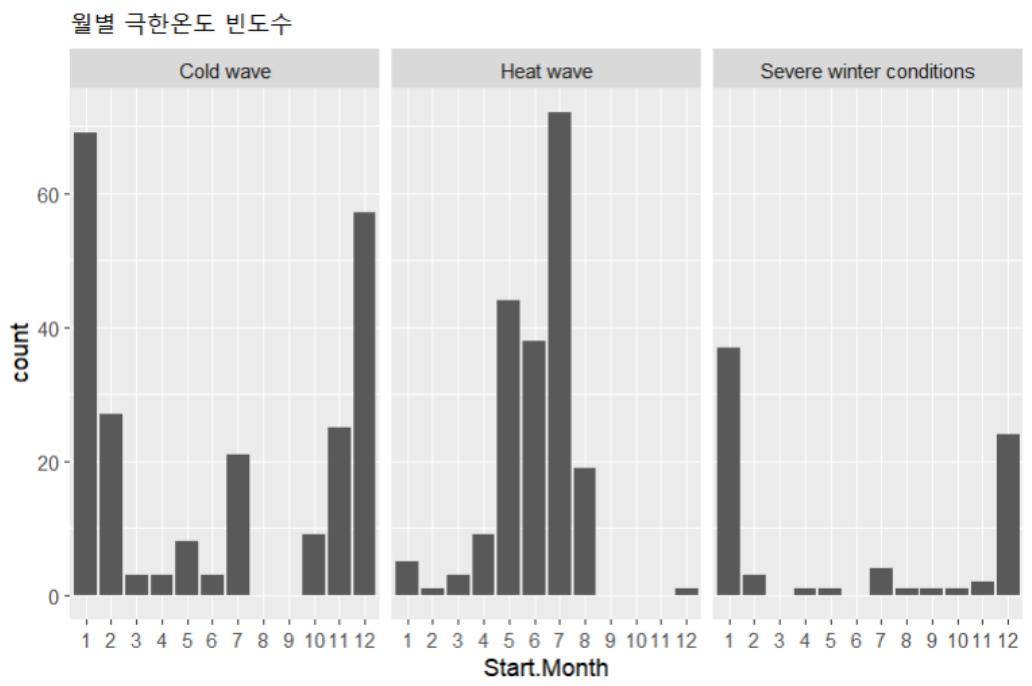


한파의 경우 2010년도, 2011년도에 매우 그 빈도가 잣았고, 그 외에는 감소하는 추세를 보이는 반면, 폭염은 최근들어 점점 그 빈도가 증가하여 기후변화와 연관성을 찾을 수 있음.

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Extreme temperature")%
>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "월별 극한온도 빈도수")

```



▼ Storm

이번에는 폭풍과 관련하여 세부 type을 찾아보자

Disaster.Type	Disaster.Subtype
1	Storm (General)
2	Tornado
3	Blizzard/Winter storm
4	Tropical cyclone
5	Hail
6	Severe weather
7	Lightning/Thunderstorms
8	Sand/Dust storm
9	Extra-tropical storm
10	Storm surge
11	Derecho

11종류의 폭풍 종류가 있으며,

Storm (General) - 일반적인 폭풍

Tornado - 토네이도

Blizzard/Winter storm - 눈보라/겨울 폭풍

Tropical cyclone - 열대성 저기압

Hail - 우박

Severe weather - 악천후

Lightning/Thunderstorms - 번개/뇌우

Sand/Dust storm - 모래/먼지 폭풍

Extra-tropical storm - 온대성 저기압

Storm surge - 폭풍 해일

Derecho - 데레초

종류가 많아 하나의 그래프로 시각화를 하기에는 굉장히 복잡하고 효율적이지 못하기에 조금 그룹화를 진행함.

Severe Weather (악천후)

→ **Storm (General), Severe weather, Lightning/Thunderstorms**

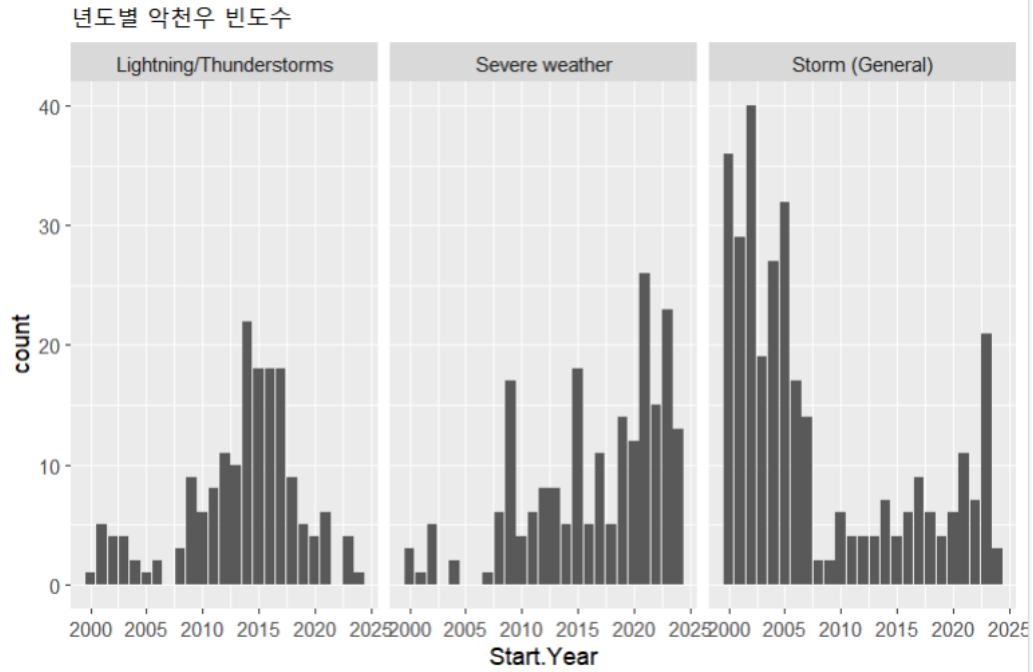
일반적인 폭풍, 악천후, 번개/뇌우

```
data %>%
  mutate(Storm.Group = case_when(
    Disaster.Subtype %in% c("Storm (General)", "Severe weather", "Lightning/Thunderstorms") ~ "Severe Weather",
    Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",
    Disaster.Subtype %in% c("Blizzard/Winter storm", "Severe winter conditions") ~ "Winter and Cold Weather",
    Disaster.Subtype %in% c("Tornado", "Derecho", "Sand/Dust storm") ~ "Wind and Dust",
    Disaster.Subtype %in% c("Storm surge", "Hail") ~ "Water-Related Events",
    TRUE ~ "Other"
  )) %>%
  filter(Storm.Group == "Severe Weather") %>%
  ggplot(aes(x = Start.Year)) +
```

```

geom_bar()+
facet_wrap(~Disaster.Subtype)+
labs(title = "년도별 악천우 빈도수")

```



악천후를 확인해볼때, 번개와 놀오는 2015년도에 자주 발생했고, 악천후의 경우 점점 증가하는 추세를 보이며, 단순 폭풍의 경우 2000년대 초반에 굉장히 빈도가 많았던 것 이후로 확 꺽였다가 최근 다시 증가하고 있음

```

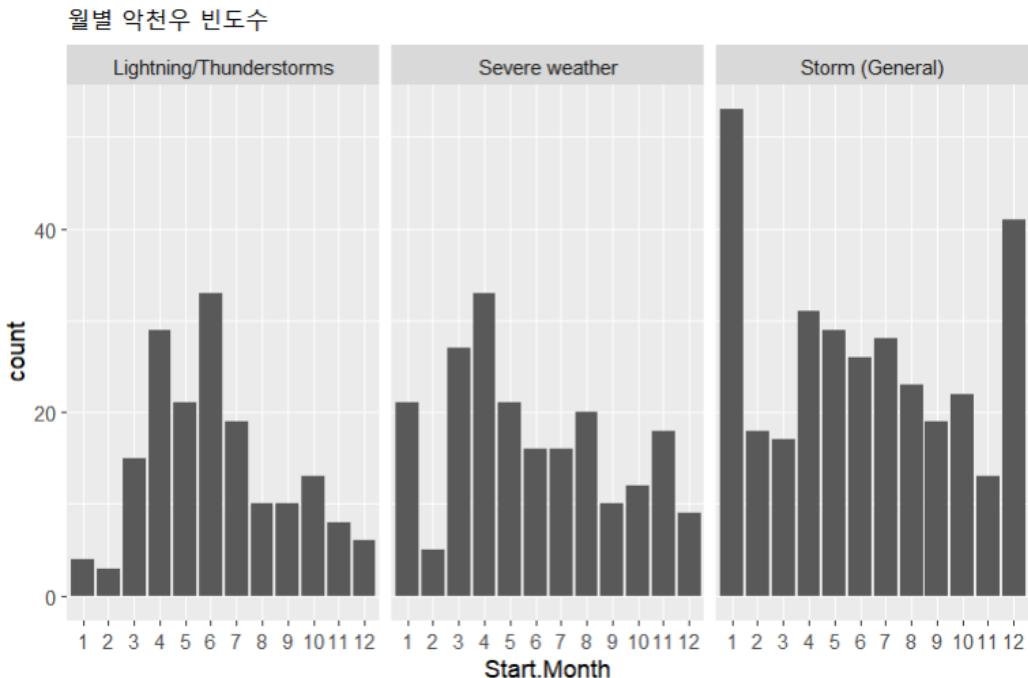
data %>%
  mutate(Storm.Group = case_when(
    Disaster.Subtype %in% c("Storm (General)", "Severe weather", "Lightning/Thunderstorms") ~ "Severe Weather",
    Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",
    Disaster.Subtype %in% c("Blizzard/Winter storm", "Severe winter conditions") ~ "Winter and Cold Weather",
    Disaster.Subtype %in% c("Tornado", "Derecho",

```

```

    "Sand/Dust storm") ~ "Wind and Dust",
    Disaster.Subtype %in% c("Storm surge", "Hail")
~ "Water-Related Events",
TRUE ~ "Other"
))%>%
filter(Storm.Group == "Severe Weather") %>%
mutate(Start.Month = factor(Start.Month))%>%
ggplot(aes(x = Start.Month)) +
geom_bar()+
facet_wrap(~Disaster.Subtype)+
labs(title = "월별 악천우 빈도수")

```



월별 빈도를 보았을때, 놀라울 동반한 악천후는 주로 3,4,5,6월에 많이 발생하고, 일반적인 폭풍은 12,1월에 많이 발생하고 있다.

Tropical and Extra-tropical Cyclones (열대성 및 온대성 저기압)

→ **Tropical cyclone, Extra-tropical storm**

열대성 저기압, 온대성 저기압

```

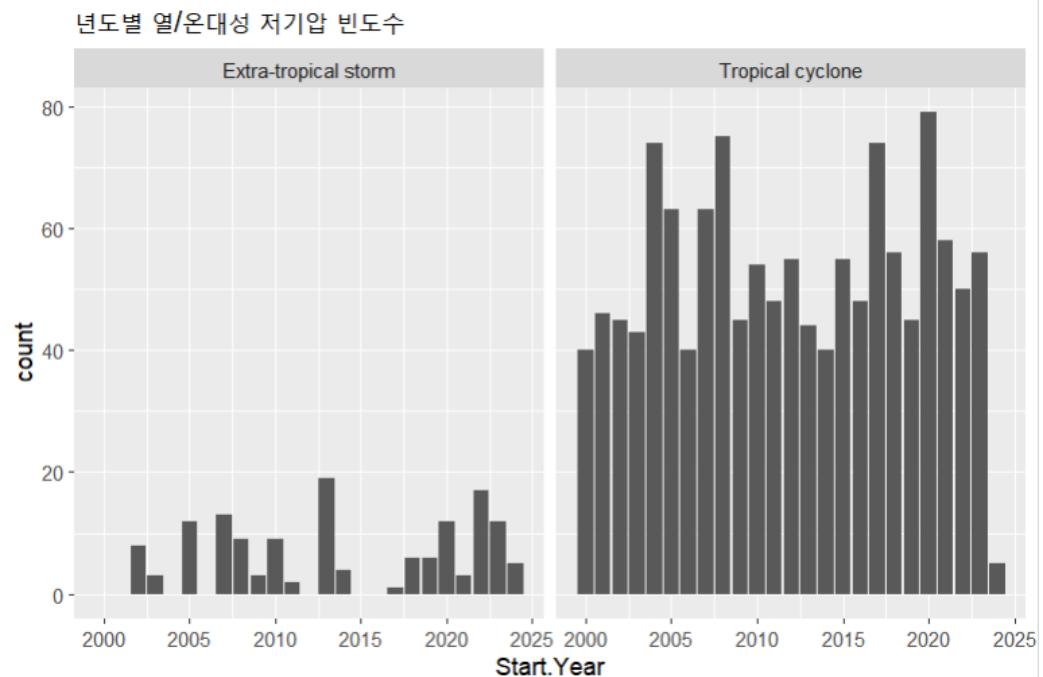
data %>%
mutate(Storm.Group = case_when(

```

```

Disaster.Subtype %in% c("Storm (General)", "Severe weather", "Lightning/Thunderstorms") ~ "Severe Weather",
Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",
Disaster.Subtype %in% c("Blizzard/Winter storm", "Severe winter conditions") ~ "Winter and Cold Weather",
Disaster.Subtype %in% c("Tornado", "Derecho", "Sand/Dust storm") ~ "Wind and Dust",
Disaster.Subtype %in% c("Storm surge", "Hail") ~ "Water-Related Events",
TRUE ~ "Other"
))%>%
filter(Storm.Group == "Tropical and Extra-tropical Cyclones") %>%
ggplot(aes(x = Start.Year)) +
geom_bar() +
facet_wrap(~Disaster.Subtype) +
labs(title = "년도별 열/온대성 저기압 빈도수")

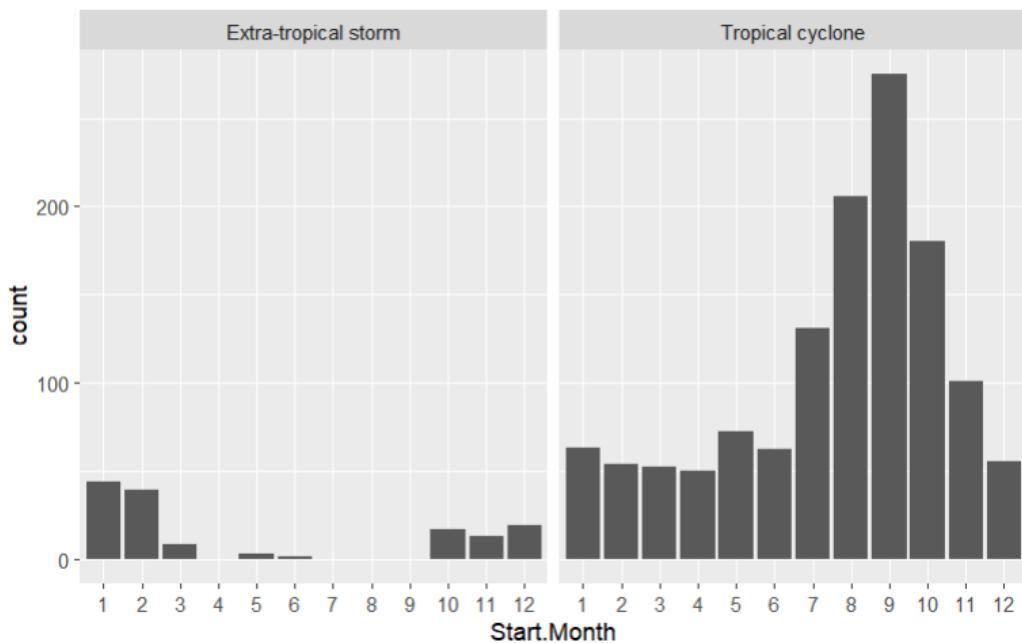
```



년도별 데이터상으로는 온대성 저기압보다 열대성 저기압의 비중이 더 높았으며

```
data %>%
  mutate(Storm.Group = case_when(
    Disaster.Subtype %in% c("Storm (General)", "Severe weather", "Lightning/Thunderstorms") ~ "Severe Weather",
    Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",
    Disaster.Subtype %in% c("Blizzard/Winter storm", "Severe winter conditions") ~ "Winter and Cold Weather",
    Disaster.Subtype %in% c("Tornado", "Derecho", "Sand/Dust storm") ~ "Wind and Dust",
    Disaster.Subtype %in% c("Storm surge", "Hail") ~ "Water-Related Events",
    TRUE ~ "Other"
  ))%>%
  filter(Storm.Group == "Tropical and Extra-tropical Cyclones")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month)) +
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "월별 열/온대성 저기압 빈도수")
```

월별 열/온대성 저기압 빈도수



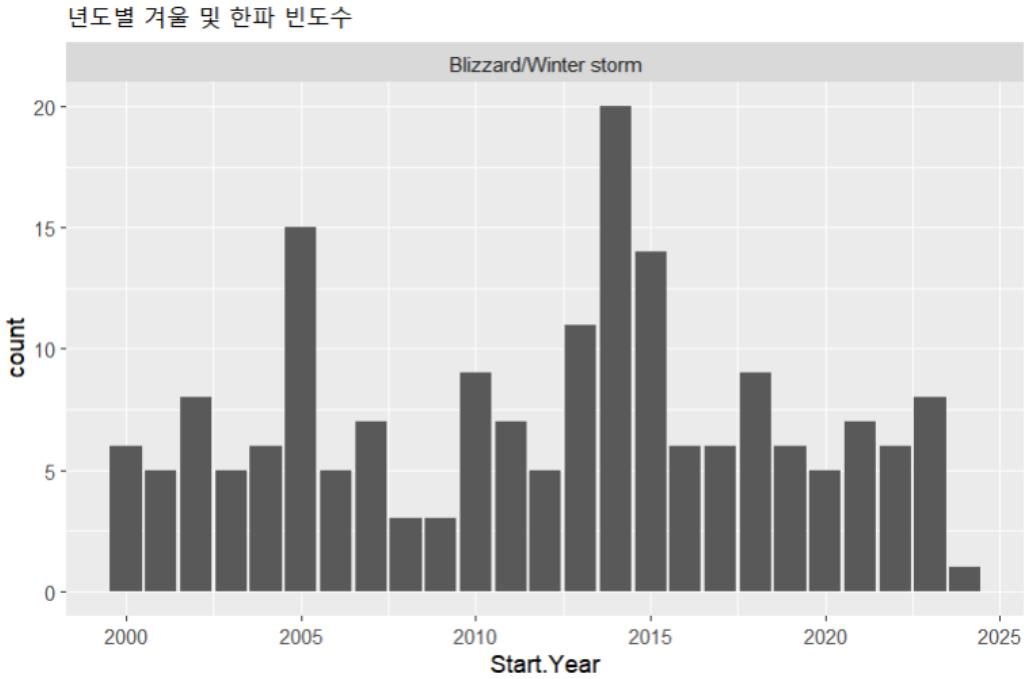
월별 데이터상에서 열대성 저기압의 경우 9월에 가장 많이 나타났음

Winter and Cold Weather (겨울 및 한파)

→ Blizzard/Winter storm, Severe winter conditions

```
data %>%
  mutate(Storm.Group = case_when(
    Disaster.Subtype %in% c("Storm (General)", "Severe weather", "Lightning/Thunderstorms") ~ "Severe Weather",
    Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",
    Disaster.Subtype %in% c("Blizzard/Winter storm") ~ "Winter and Cold Weather",
    Disaster.Subtype %in% c("Tornado", "Derecho", "Sand/Dust storm") ~ "Wind and Dust",
    Disaster.Subtype %in% c("Storm surge", "Hail") ~ "Water-Related Events",
    TRUE ~ "Other"
  )) %>%
  filter(Storm.Group == "Winter and Cold Weather")
```

```
%>%  
  ggplot(aes(x = Start.Year)) +  
  geom_bar() +  
  facet_wrap(~Disaster.Subtype) +  
  labs(title = "년도별 겨울 및 한파 빈도수")
```



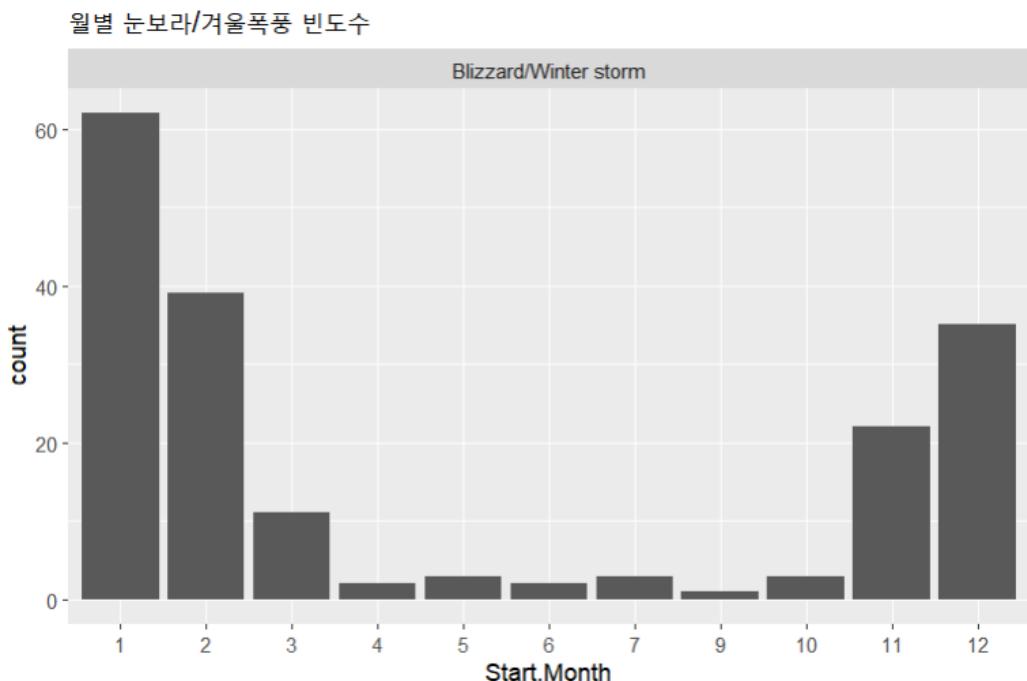
겨울 폭풍의 경우 2015년 전후로 빈도가 높았음

```
data %>%  
  mutate(Storm.Group = case_when(  
    Disaster.Subtype %in% c("Storm (General)", "Severe weather", "Lightning/Thunderstorms") ~ "Severe Weather",  
    Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",  
    Disaster.Subtype %in% c("Blizzard/Winter storm") ~ "Winter and Cold Weather",  
    Disaster.Subtype %in% c("Tornado", "Derecho", "Sand/Dust storm") ~ "Wind and Dust",  
    Disaster.Subtype %in% c("Storm surge", "Hail") ~ "Water-Related Events",
```

```

TRUE ~ "Other"
))%>%
filter(Storm.Group == "Winter and Cold Weather")
%>%
mutate(Start.Month = factor(Start.Month))%>%
ggplot(aes(x = Start.Month)) +
geom_bar()+
facet_wrap(~Disaster.Subtype)+
labs(title = "월별 눈보라/겨울폭풍 빈도수")

```



월별 빈도 역시 겨울에 집중되어있음

Wind and Dust (바람과 먼지 관련 재해)

→ Tornado, Derecho, Sand/Dust storm

```

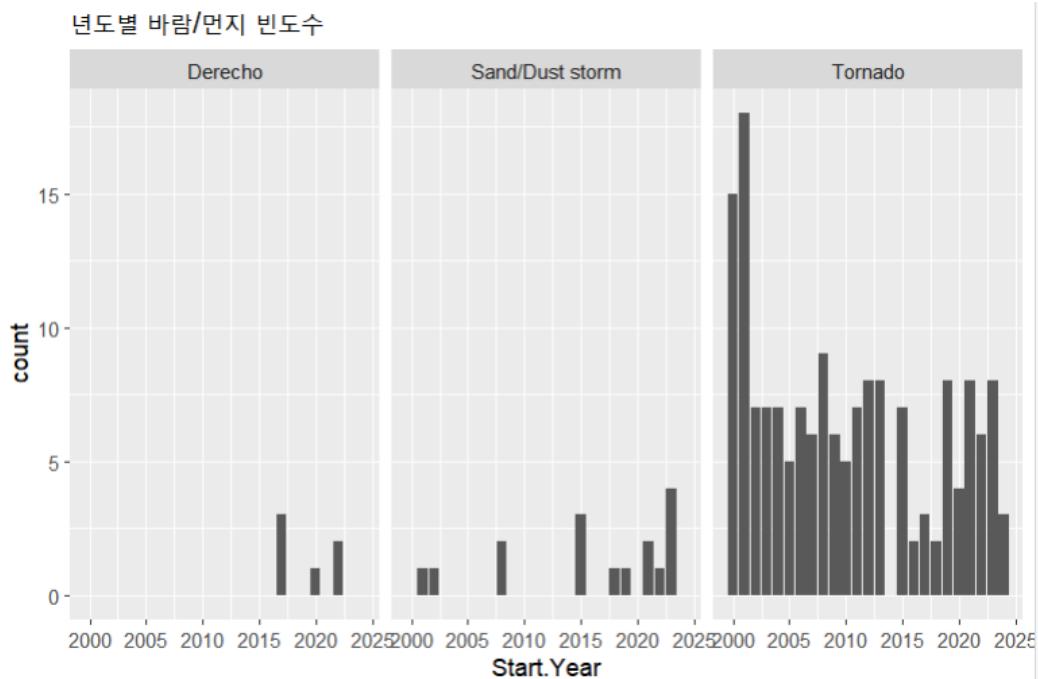
data %>%
  mutate(Storm.Group = case_when(
    Disaster.Subtype %in% c("Storm (General)", "Se
    vere weather", "Lightning/Thunderstorms") ~ "Sever
    e Weather",
    Disaster.Subtype %in% c("Tropical cyclone", "E

```

```

xtra-tropical storm") ~ "Tropical and Extra-tropic
al Cyclones",
Disaster.Subtype %in% c("Blizzard/Winter stor
m", "Severe winter conditions") ~ "Winter and Cold
Weather",
Disaster.Subtype %in% c("Tornado", "Derecho",
"Sand/Dust storm") ~ "Wind and Dust",
Disaster.Subtype %in% c("Storm surge", "Hail")
~ "Water-Related Events",
TRUE ~ "Other"
))%>%
filter(Storm.Group == "Wind and Dust") %>%
ggplot(aes(x = Start.Year)) +
geom_bar()+
facet_wrap(~Disaster.Subtype)+
labs(title = "년도별 바람/먼지 빈도수")

```



```

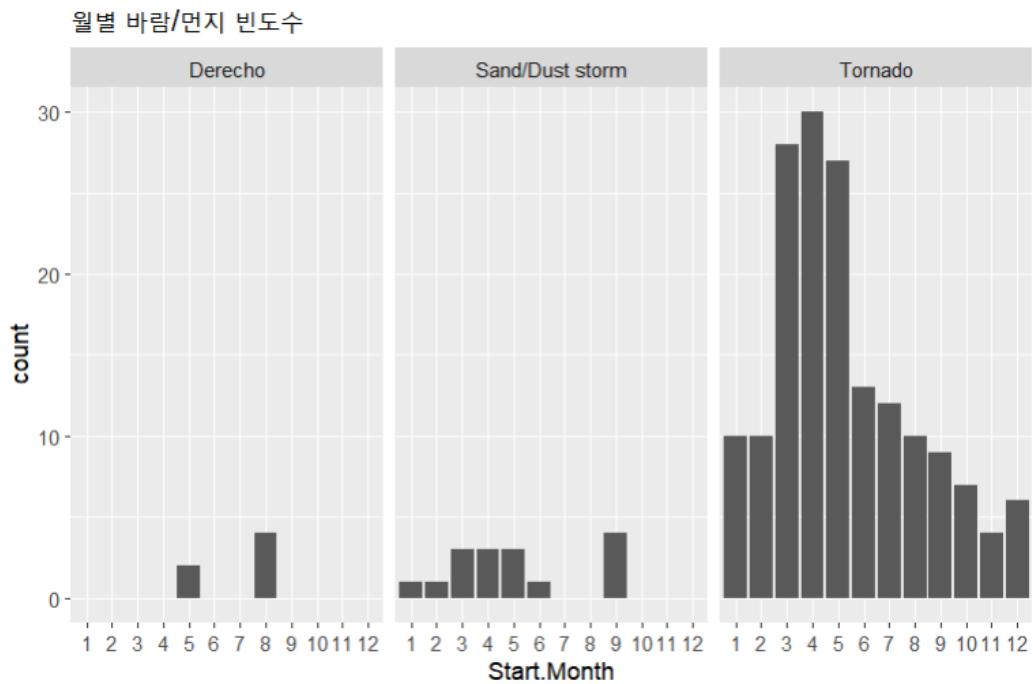
data %>%
mutate(Storm.Group = case_when(
Disaster.Subtype %in% c("Storm (General)", "Se
vere weather", "Lightning/Thunderstorms") ~ "Sever
e Weather",

```

```

Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",
Disaster.Subtype %in% c("Blizzard/Winter storm", "Severe winter conditions") ~ "Winter and Cold Weather",
Disaster.Subtype %in% c("Tornado", "Derecho", "Sand/Dust storm") ~ "Wind and Dust",
Disaster.Subtype %in% c("Storm surge", "Hail") ~ "Water-Related Events",
TRUE ~ "Other"
))%>%
filter(Storm.Group == "Wind and Dust") %>%
mutate(Start.Month = factor(Start.Month))%>%
ggplot(aes(x = Start.Month)) +
geom_bar()+
facet_wrap(~Disaster.Subtype)+
labs(title = "월별 바람/먼지 빈도수")

```



Water-Related Events (물 관련 현상)

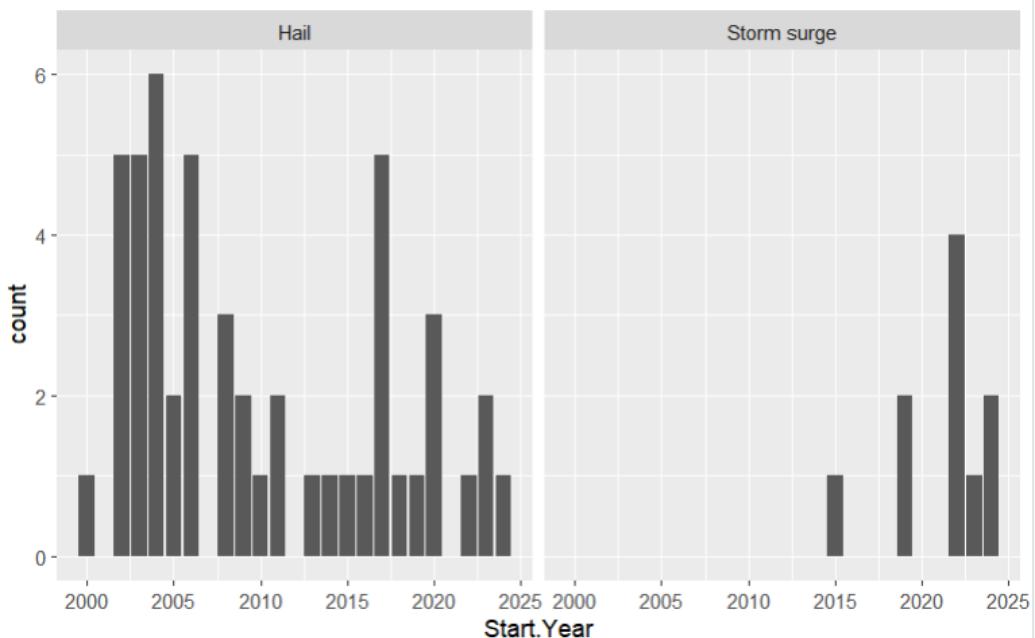
→ Storm surge, Hail

```

data %>%
  mutate(Storm.Group = case_when(
    Disaster.Subtype %in% c("Storm (General)", "Severe weather", "Lightning/Thunderstorms") ~ "Severe Weather",
    Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",
    Disaster.Subtype %in% c("Blizzard/Winter storm", "Severe winter conditions") ~ "Winter and Cold Weather",
    Disaster.Subtype %in% c("Tornado", "Derecho", "Sand/Dust storm") ~ "Wind and Dust",
    Disaster.Subtype %in% c("Storm surge", "Hail") ~ "Water-Related Events",
    TRUE ~ "Other"
  ))%>%
  filter(Storm.Group == "Water-Related Events") %>%
  ggplot(aes(x = Start.Year)) +
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 물 관련 빈도수")

```

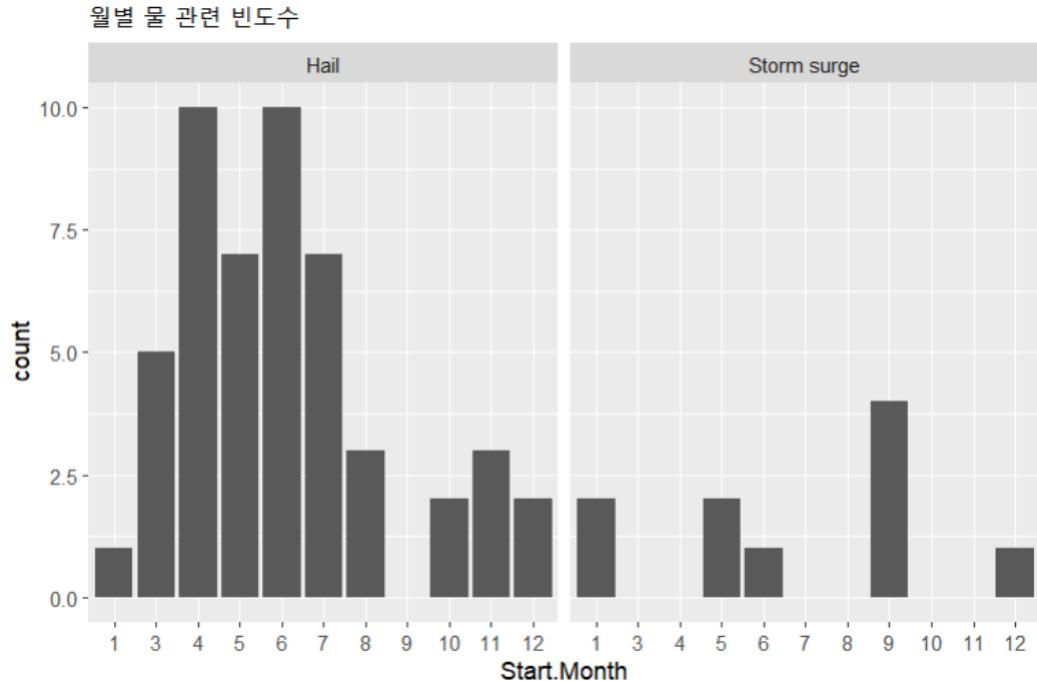
년도별 물 관련 빈도수



```

data %>%
  mutate(Storm.Group = case_when(
    Disaster.Subtype %in% c("Storm (General)", "Severe weather", "Lightning/Thunderstorms") ~ "Severe Weather",
    Disaster.Subtype %in% c("Tropical cyclone", "Extra-tropical storm") ~ "Tropical and Extra-tropical Cyclones",
    Disaster.Subtype %in% c("Blizzard/Winter storm", "Severe winter conditions") ~ "Winter and Cold Weather",
    Disaster.Subtype %in% c("Tornado", "Derecho", "Sand/Dust storm") ~ "Wind and Dust",
    Disaster.Subtype %in% c("Storm surge", "Hail") ~ "Water-Related Events",
    TRUE ~ "Other"
  )) %>%
  filter(Storm.Group == "Water-Related Events") %>%
  mutate(Start.Month = factor(Start.Month)) %>%
  ggplot(aes(x = Start.Month)) +
  geom_bar() +
  
```

```
facet_wrap(~Disaster.Subtype) +
  labs(title = "월별 물 관련 빈도수")
```



▼ Miscellaneous accident (기타 사고)

```
data %>%
  select(Disaster.Type, Disaster.Subgroup) %>%
  filter(Disaster.Subgroup == "Miscellaneous accident") %>%
  distinct()
```

	Disaster.Type	Disaster.Subgroup
1	Fire (Miscellaneous)	Miscellaneous accident
2	Collapse (Miscellaneous)	Miscellaneous accident
3	Explosion (Miscellaneous)	Miscellaneous accident
4	Miscellaneous accident (General)	Miscellaneous accident

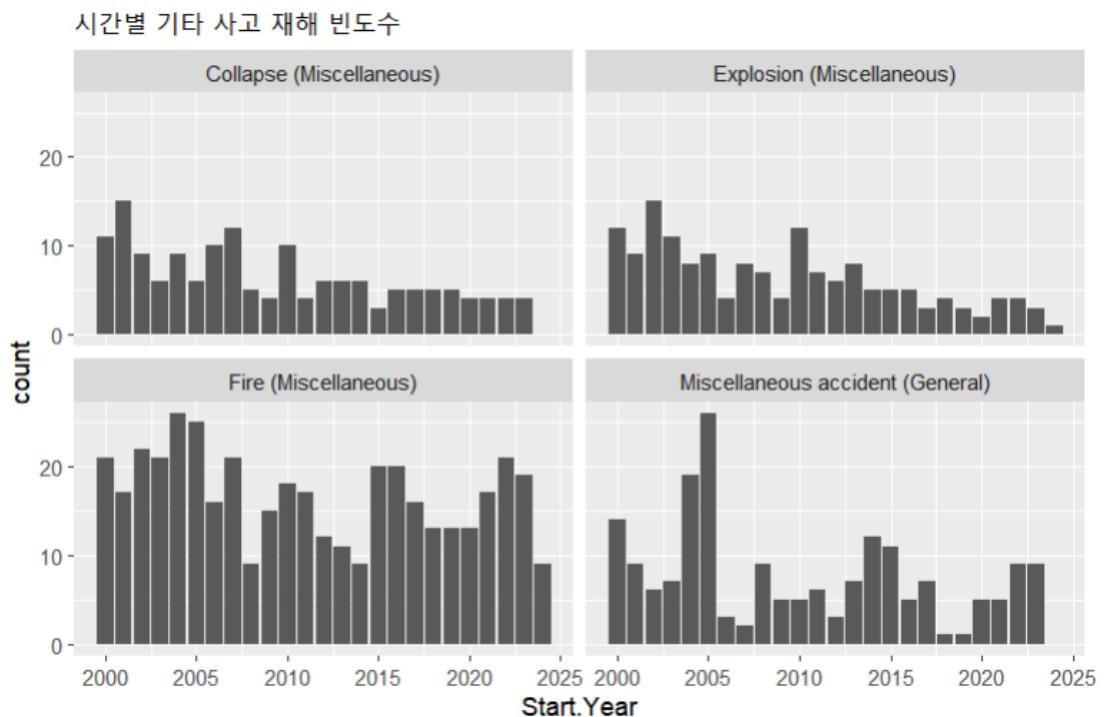
Collapse (Miscellaneous) (붕괴 - 기타)

Fire (Miscellaneous) (화재-기타)

Explosion (Miscellaneous) (폭발 - 기타)

Miscellaneous accident (General) (기타 사고 - 일반)

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Miscellaneous accident")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "시간별 기타 사고 재해 빈도수")
```



기타 사고중 폭발과 붕괴는 해마다 수가 감소하는 추세로 관련 법률이나 제도의 변경으로 사고가 줄어들었다고 생각할 수 있음

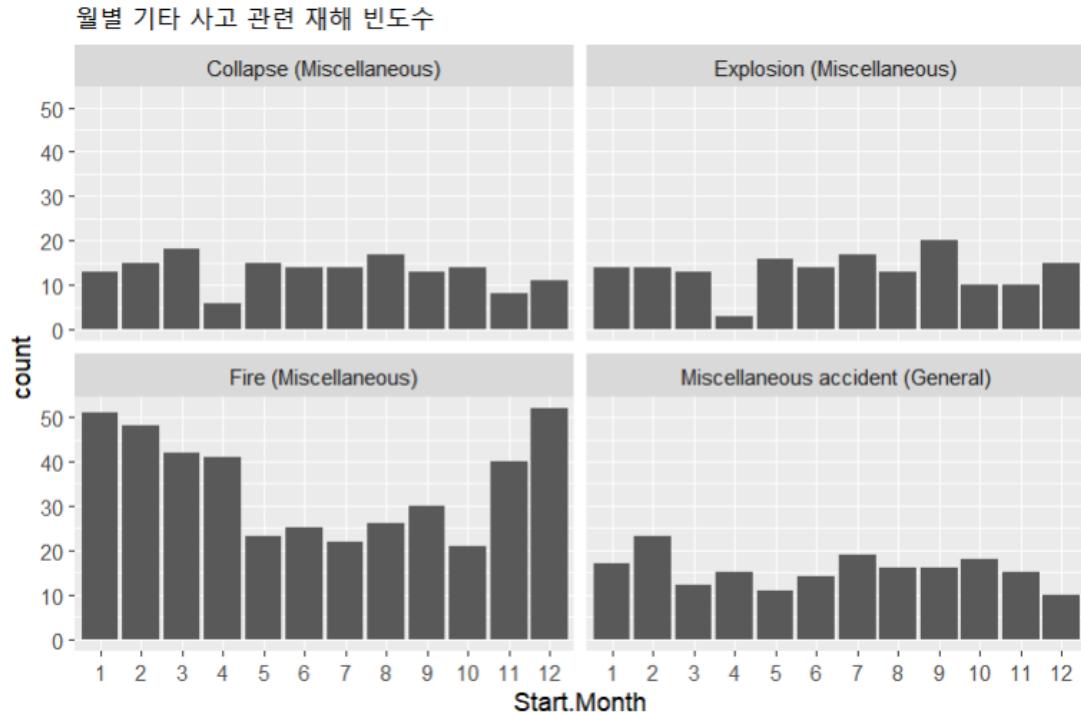
```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Miscellaneous accident")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()
```

```

facet_wrap(~Disaster.Type)+  

labs(title = "월별 기타 사고 관련 재해 빈도수")

```



기타 사고 역시 세부 type별 분류를 하여 시각화를 시작한다

Disaster.Type	Disaster.Subtype
1 Collapse (Miscellaneous)	Collapse (Miscellaneous)

Disaster.Type	Disaster.Subtype
1 Fire (Miscellaneous)	Fire (Miscellaneous)

Disaster.Type	Disaster.Subtype
1 Explosion (Miscellaneous)	Explosion (Miscellaneous)

Disaster.Type	Disaster.Subtype
1 Miscellaneous accident (General)	Miscellaneous accident (General)

기타 사고의 세부 type은 모두 하나씩 동일하므로 Type별 시각화 자료와 동일하다

▼ Geophysical (지질학적 재해)

```
data %>%
  select(Disaster.Type, Disaster.Subgroup) %>%
  filter(Disaster.Subgroup == "Geophysical") %>%
  distinct()
```

	Disaster.Type	Disaster.Subgroup
1	Volcanic activity	Geophysical
2	Earthquake	Geophysical
3	Mass movement (dry)	Geophysical

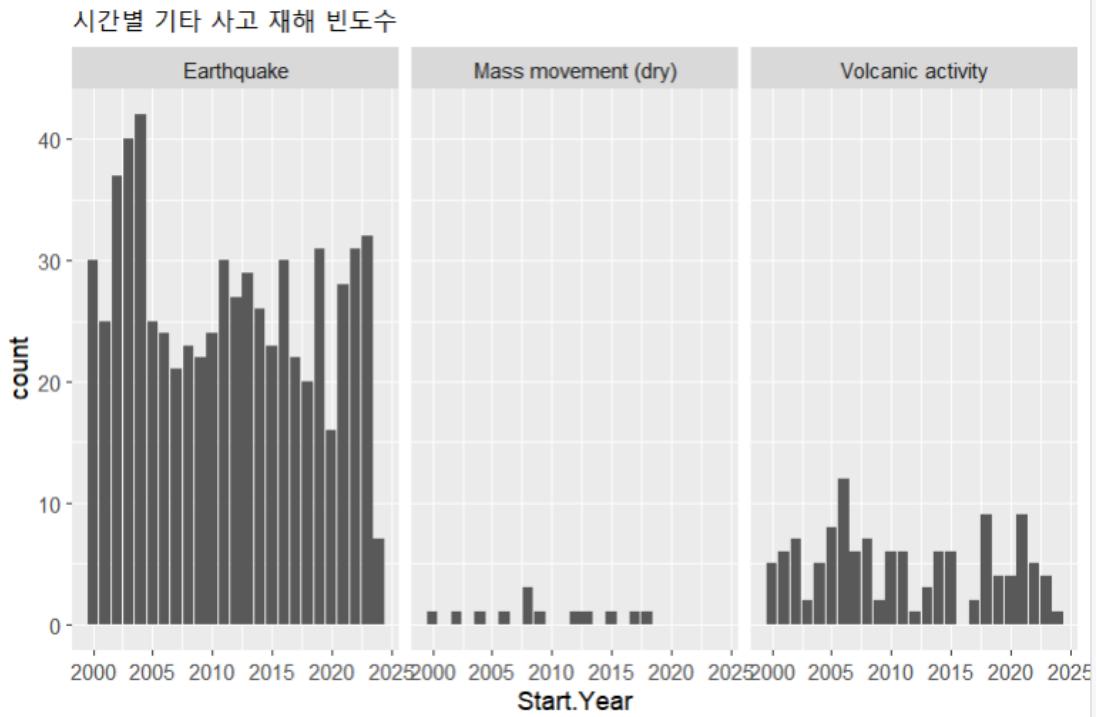
> |

Earthquake (지진)

Volcanic activity (화산 활동)

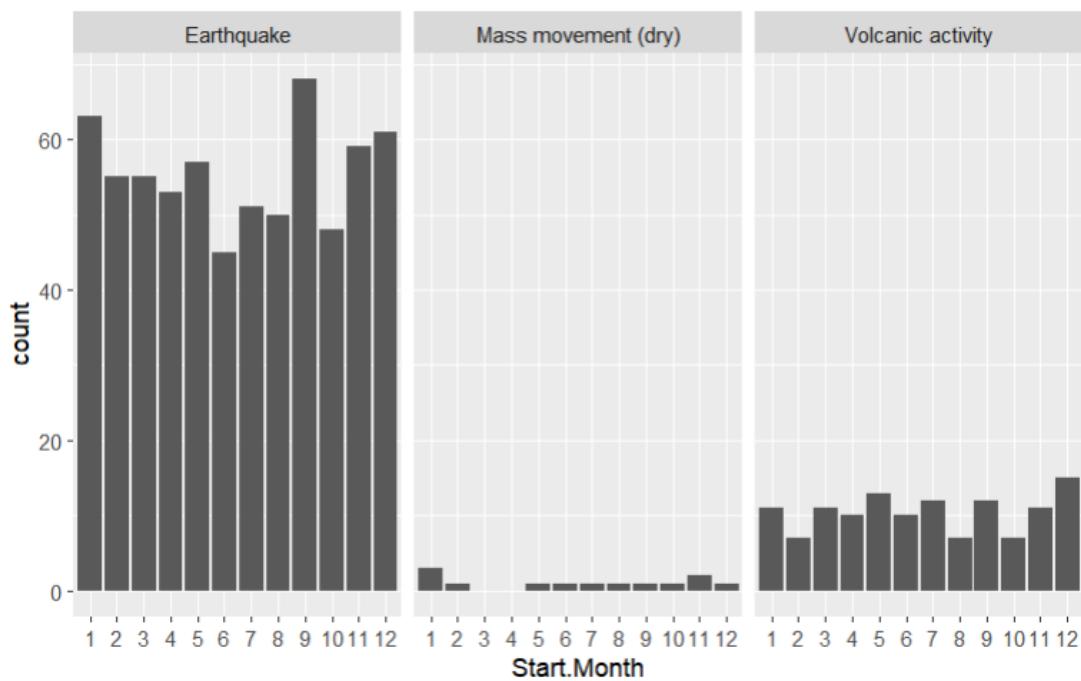
Mass movement (dry) (대규모 이동 - 건조한 조건)

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Geophysical")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "시간별 지질학적 재해 빈도수")
```



```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Geophysical")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "월별 지질학적 재해 빈도수")
```

월별 지질학적 재해 빈도수



▼ Earthquake

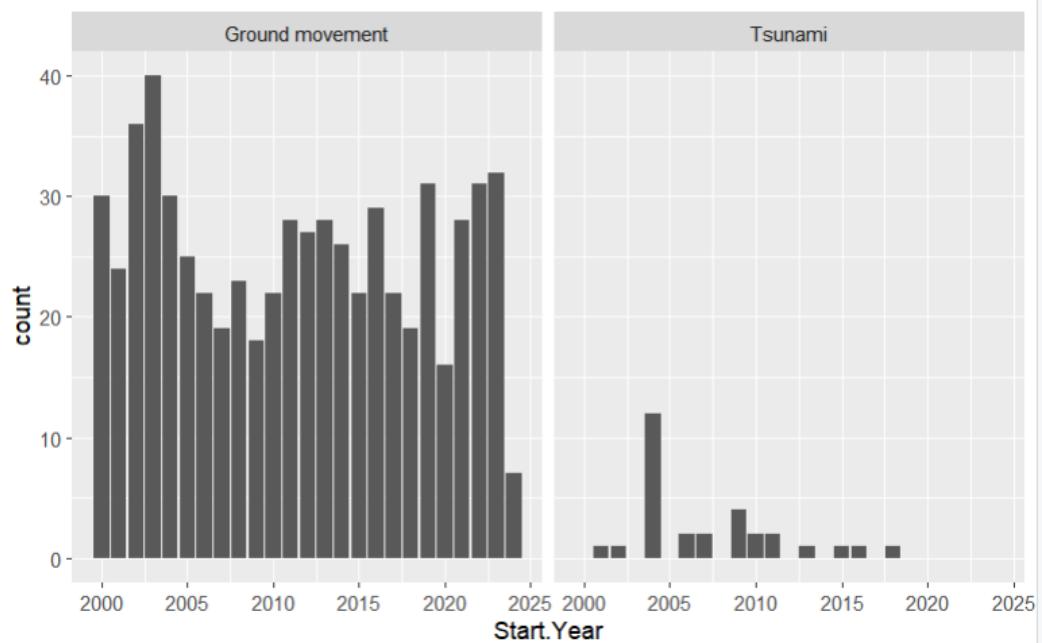
```
Disaster.Type Disaster.Subtype
1   Earthquake   Ground movement
2   Earthquake       Tsunami
```

Earthquake (Ground movement) - 지진 (지반 이동)

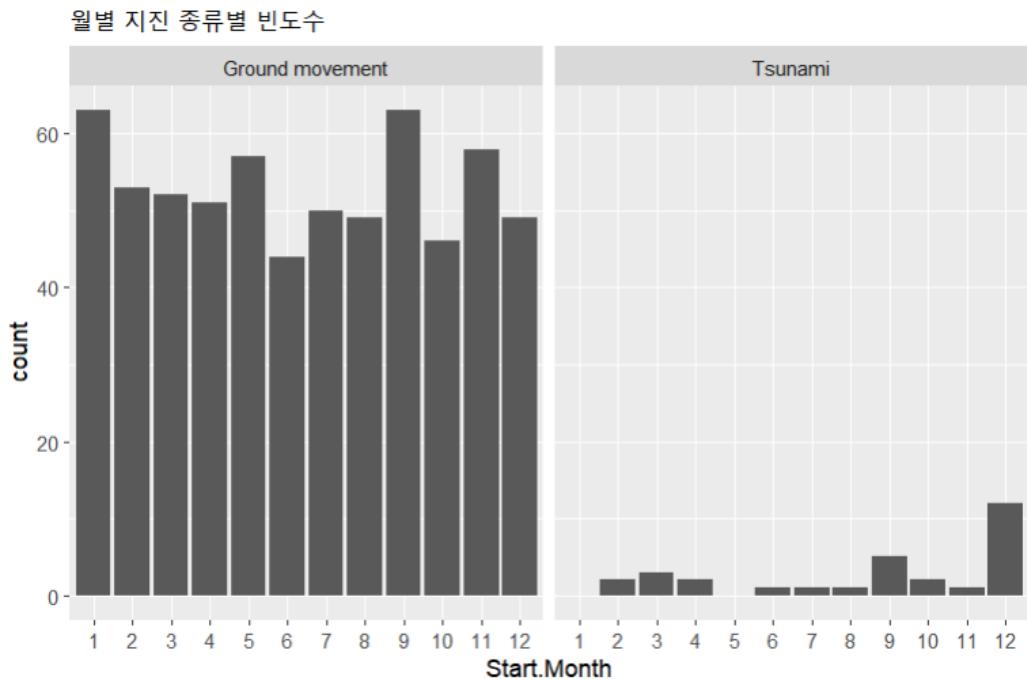
Earthquake (Tsunami) - 지진 (쓰나미)

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Earthquake")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 지진 종류별 빈도수")
```

년도별 지진 종류별 빈도수



```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Earthquake")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "월별 지진 종류별 빈도수")
```



▼ Volcanic activity

화산 활동과 관련해서는

Disaster.Type	Disaster.Subtype
1 Volcanic activity	Ash fall
2 Volcanic activity	Volcanic activity (General)
3 Volcanic activity	Lava flow
4 Volcanic activity	Pyroclastic flow

4개의 세부 type이 있음

Volcanic activity (Ash fall) - 화산 활동 (화산재 낙하)

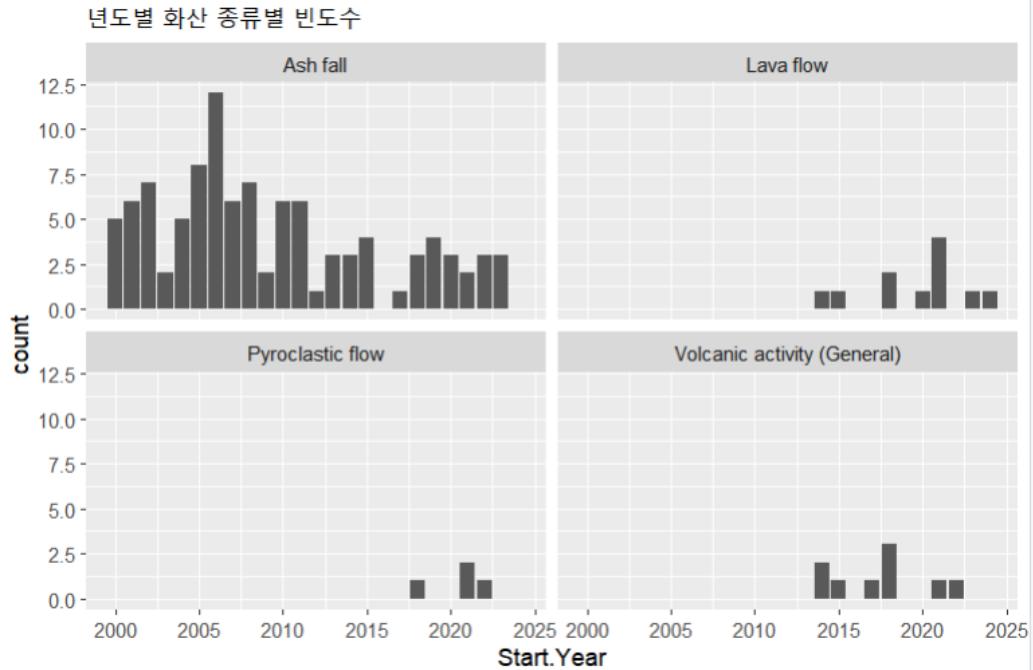
Volcanic activity (Volcanic activity - General) (일반)

Volcanic activity (Lava flow) - 화산 활동 (용암류)

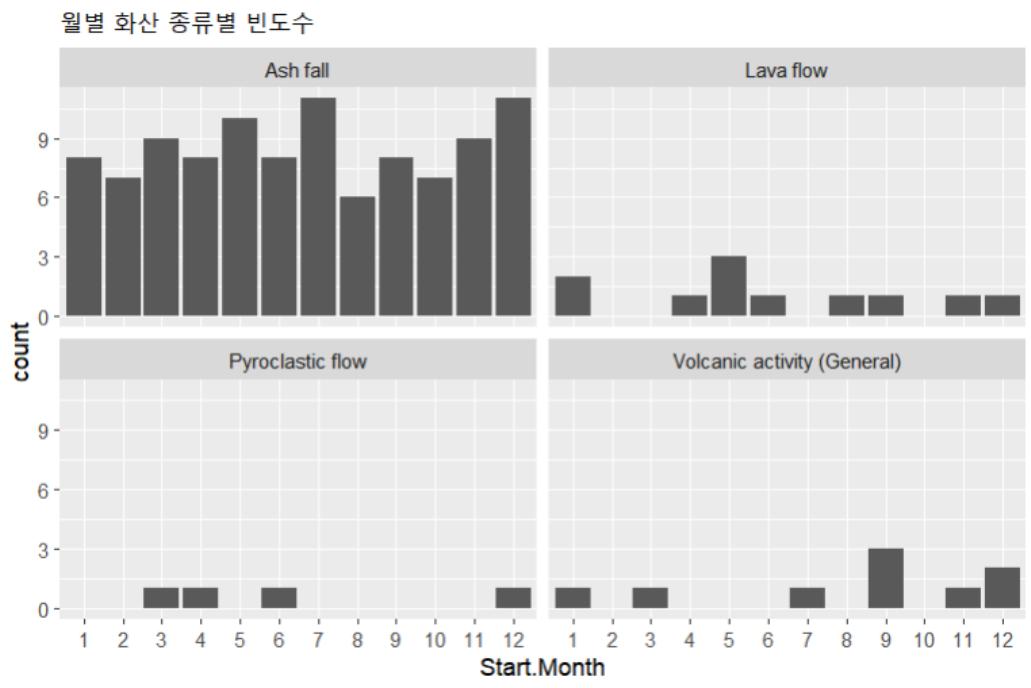
Volcanic activity (Pyroclastic flow) - 화산 활동 (화쇄류)

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Volcanic activity")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()
```

```
facet_wrap(~Disaster.Subtype)+  
  labs(title = "년도별 화산 종류별 빈도수")
```



```
data%>%  
  group_by(Disaster.Group)%>%  
  filter(Disaster.Type == "Volcanic activity")%>%  
  mutate(Start.Month = factor(Start.Month))%>%  
  ggplot(aes(x = Start.Month))+  
  geom_bar() +  
  facet_wrap(~Disaster.Subtype)+  
  labs(title = "월별 화산 종류별 빈도수")
```



▼ Mass movement (dry)

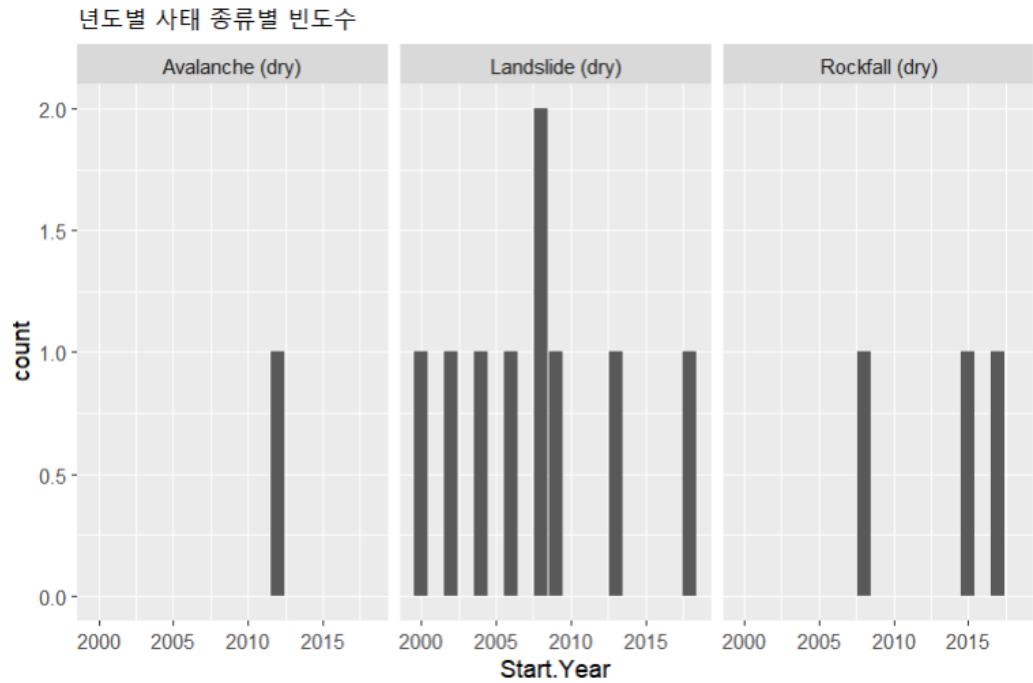
```
Disaster.Type Disaster.Subtype
1 Mass movement (dry) Landslide (dry)
2 Mass movement (dry) Rockfall (dry)
3 Mass movement (dry) Avalanche (dry)
```

Mass movement (dry) (Landslide - dry) 산사태

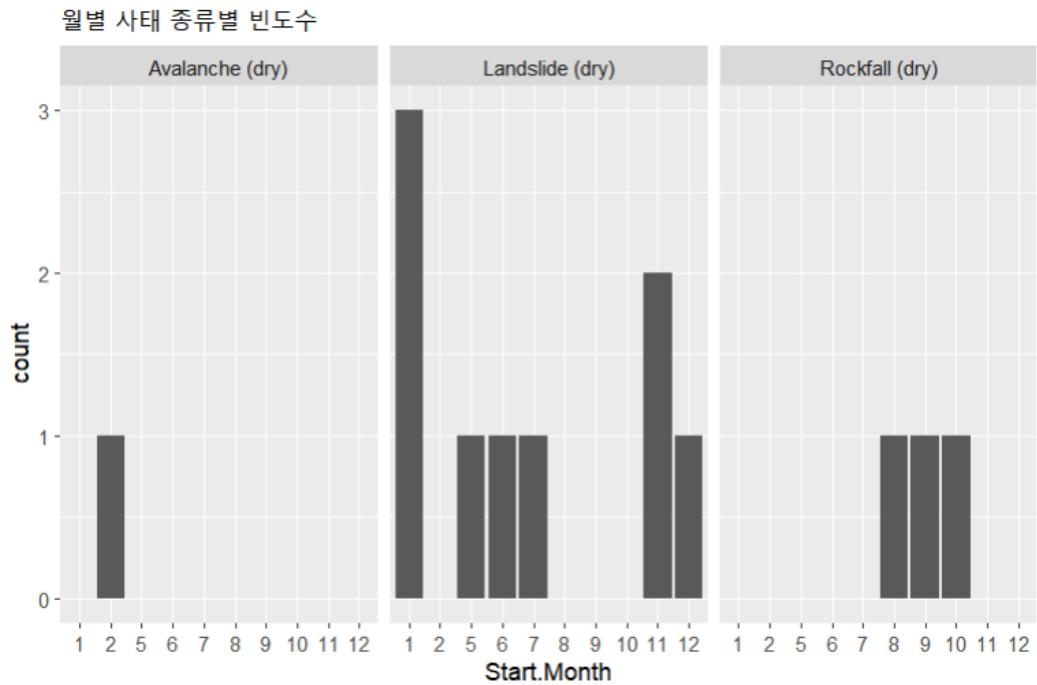
Mass movement (dry) (Rockfall - dry) 낙석

Mass movement (dry) (Avalanche - dry) 눈사태

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Mass movement (dry)")%
>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 사태 종류별 빈도수")
```



```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Mass movement (dry)")%
>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "월별 사태 종류별 빈도수")
```



▼ Industrial accident (산업 재해)

```
data %>%
  select(Disaster.Type, Disaster.Subgroup) %>%
  filter(Disaster.Subgroup == "Industrial accident")
%>%
  distinct()
```

	Disaster.Type	Disaster.Subgroup
1	Collapse (Industrial)	Industrial accident
2	Fire (Industrial)	Industrial accident
3	Explosion (Industrial)	Industrial accident
4	Chemical spill	Industrial accident
5	Gas leak	Industrial accident
6	Poisoning	Industrial accident
7	Industrial accident (General)	Industrial accident
8	Radiation	Industrial accident
9	Oil spill	Industrial accident

Fire (Industrial) (화재 - 산업)

Explosion (Industrial) (폭발 - 산업): 산업 현장에서 발생한 폭발.

Chemical spill (화학물질 유출): 화학 물질이 유출되어 발생하는 사고.

Gas leak (가스 누출): 산업 현장에서 가스가 누출되어 발생하는 재해.

Radiation (방사능 누출): 방사능이 누출되어 환경이나 인체에 영향을 미치는 재해.

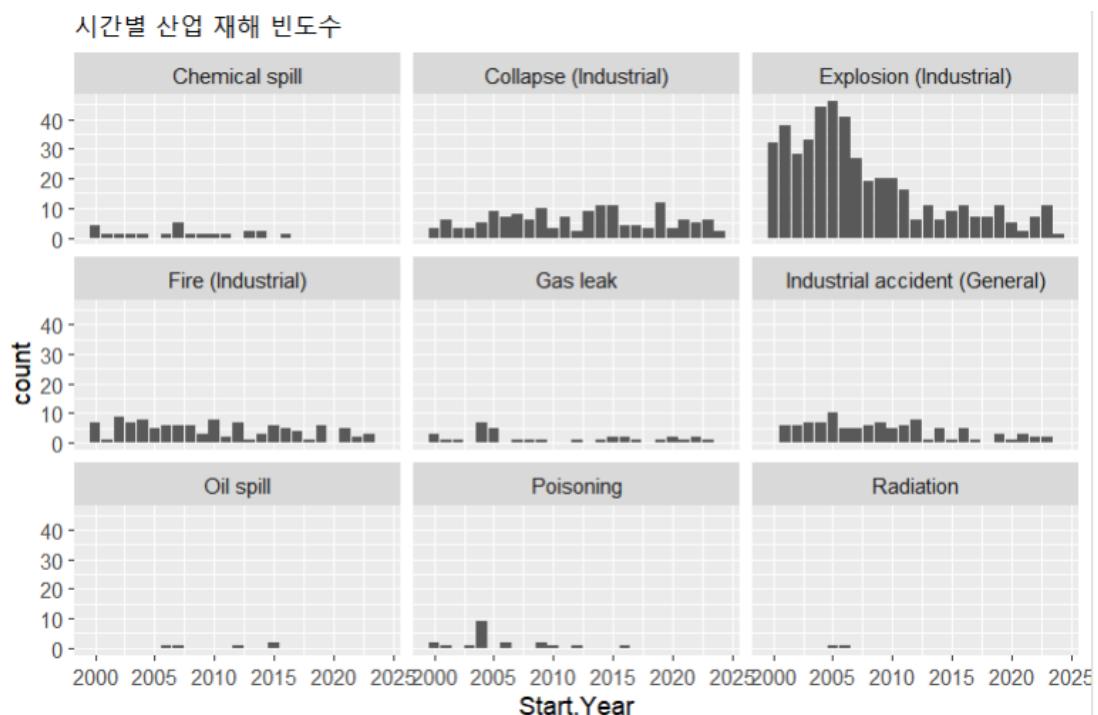
Industrial accident (General) (산업 사고 - 일반): 산업 전반에서 발생하는 기타 사고.

Oil spill (기름 유출)

Poisoning

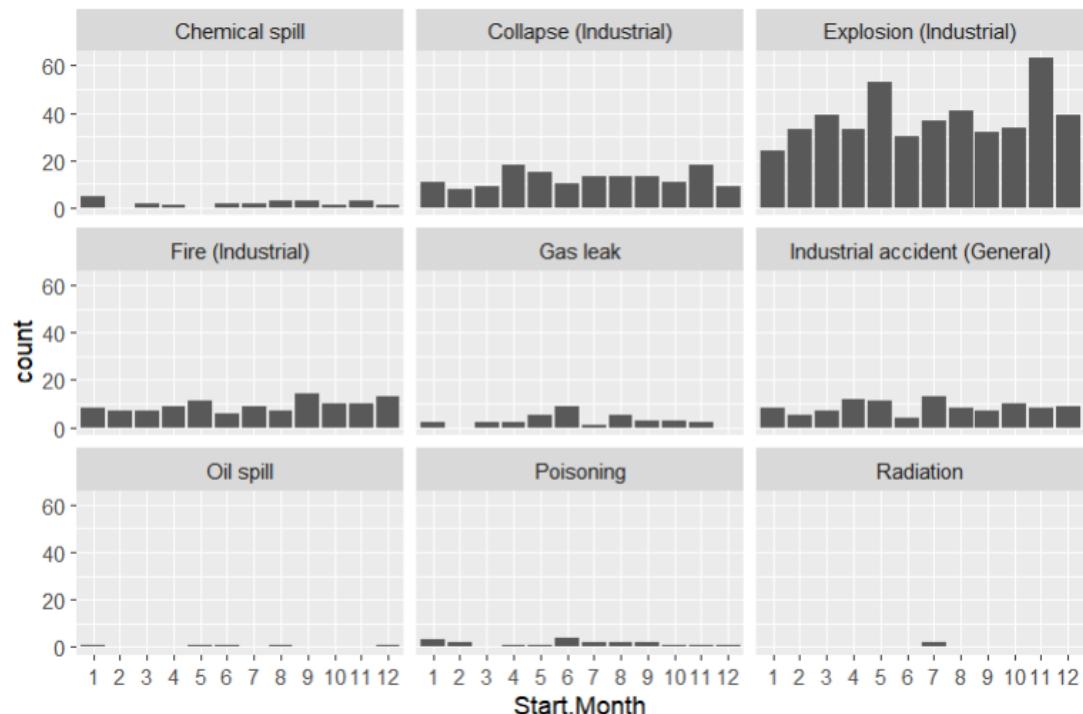
Collapse (Industrial)

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "시간별 산업 재해 빈도수")
```



월별 빈도수 추이

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Type)
```



type별 세부 유형 확인

```
Disaster.Type    Disaster.Subtype
1 Fire (Industrial) Fire (Industrial)
```

```
Disaster.Type    Disaster.Subtype
1 Explosion (Industrial) Explosion (Industrial)
```

```
Disaster.Type Disaster.Subtype  
1 Chemical spill    Chemical spill
```

```
Disaster.Type Disaster.Subtype  
1      Gas leak        Gas leak
```

```
Disaster.Type Disaster.Subtype  
1      Radiation       Radiation
```

```
Disaster.Type          Disaster.Subtype  
r.Subtype  
1 Industrial accident (General) Industrial accident  
(General)
```

```
Disaster.Type Disaster.Subtype  
1      Oil spill        Oil spill
```

```
Disaster.Type Disaster.Subtype  
1      Poisoning        Poisoning
```

```
Disaster.Type          Disaster.Subtype  
1 Collapse (Industrial) Collapse (Industrial)
```

산업재해 역시 type별 세부 type이 동일해 년도, 월별 빈도는 동일하다

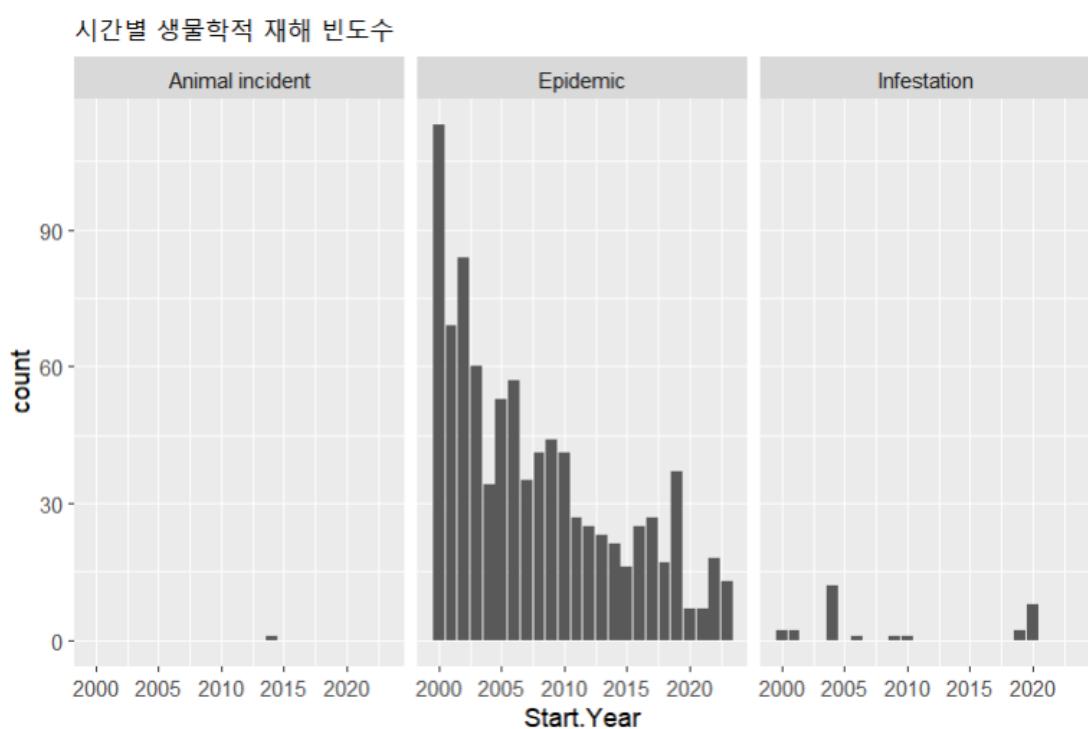
▼ Biological (생물학적 재해)

```
data %>%  
  select(Disaster.Type, Disaster.Subgroup) %>%  
  filter(Disaster.Subgroup == "Biological") %>%  
  distinct()
```

	Disaster.Type	Disaster.Subgroup
1	Epidemic	Biological
2	Infestation	Biological
3	Animal incident	Biological

- **Epidemic (전염병)**
- **Infestation (해충 침입)**
- **Animal incident (동물 사고)**

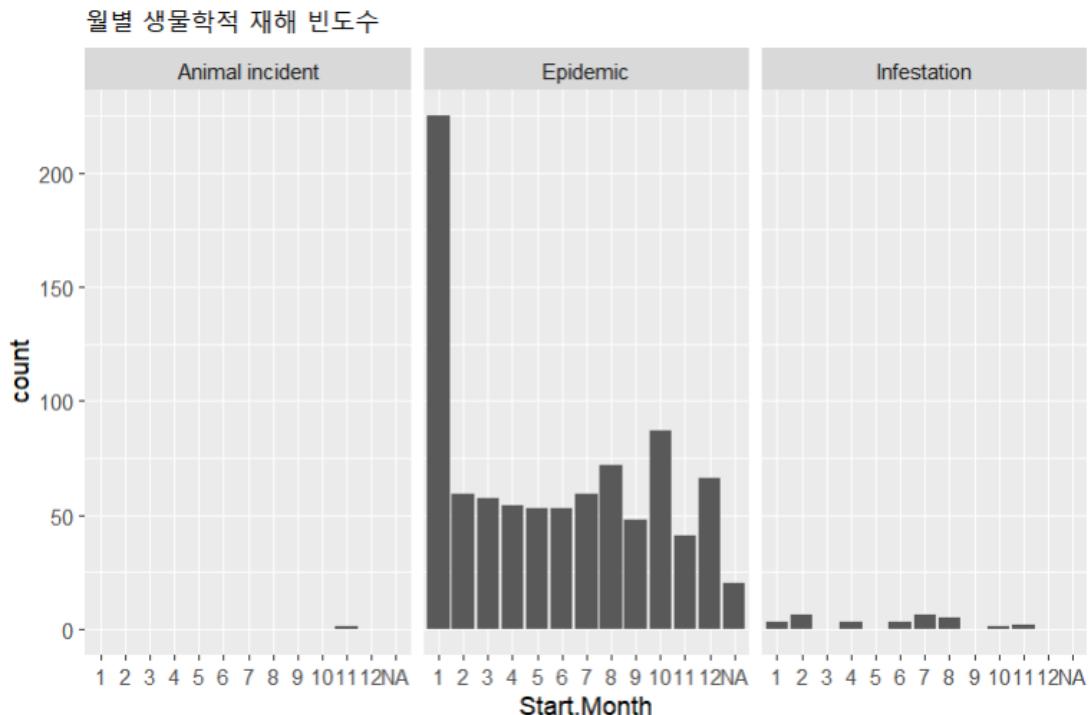
```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Biological")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "시간별 생물학적 재해 빈도수")
```



```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Biological")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Type)+
  labs(title = "월별 생물학적 재해 빈도수")

```



생물학 재해에서 유형별 세부 type

▼ Epidemic

Disaster.Type	Disaster.Subtype
1 Epidemic	Viral disease
2 Epidemic	Infectious disease (General)
3 Epidemic	Bacterial disease
4 Epidemic	Parasitic disease

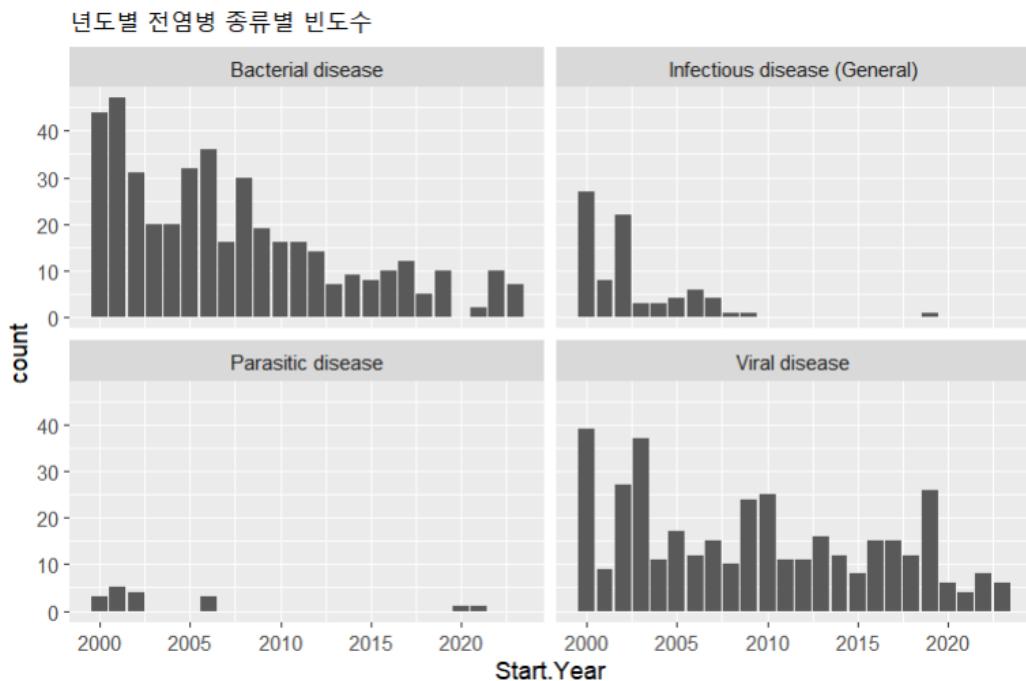
Epidemic (Viral disease) - 전염병 (바이러스 질병)

Epidemic (Infectious disease - General) - 전염병 (일반적인 감염병)

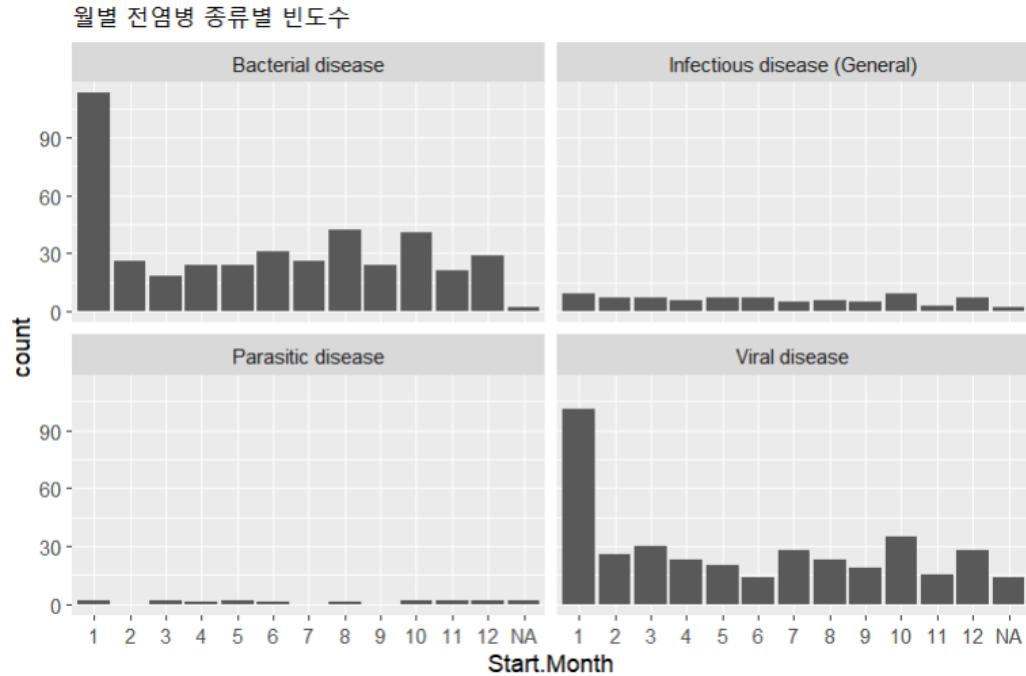
Epidemic (Bacterial disease) - 전염병 (세균성 질병)

Epidemic (Parasitic disease) - 전염병 (기생충 질병)

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Epidemic")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 전염병 종류별 빈도수")
```



```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Epidemic")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "월별 전염병 종류별 빈도수")
```



▼ Infestation

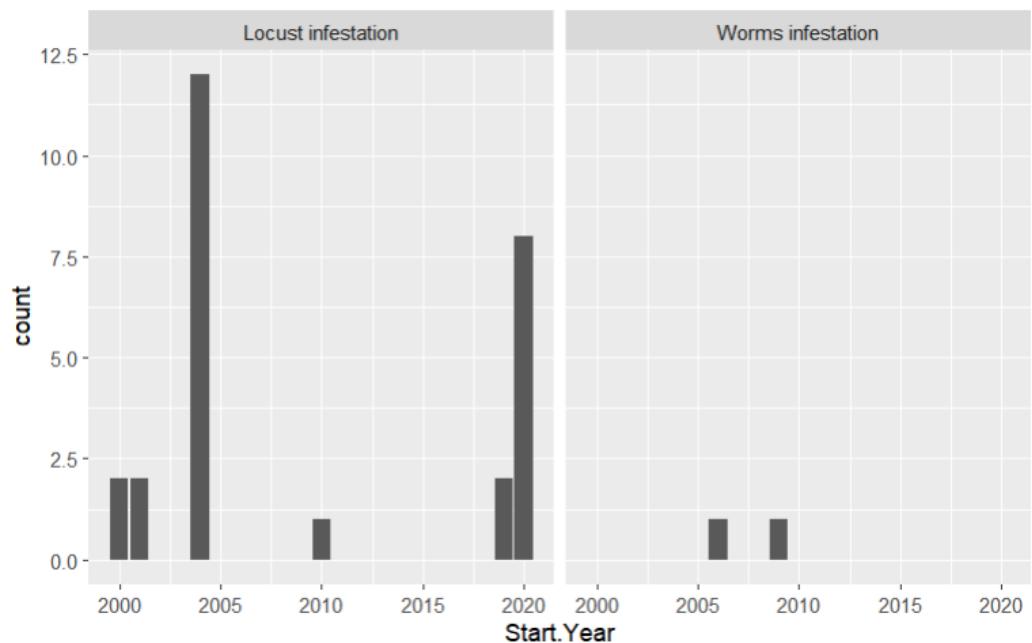
```
Disaster.Type    Disaster.Subtype
1   Infestation Locust infestation
2   Infestation Worms infestation
```

Infestation (Locust infestation) - 해충 침입 (메뚜기 침입)

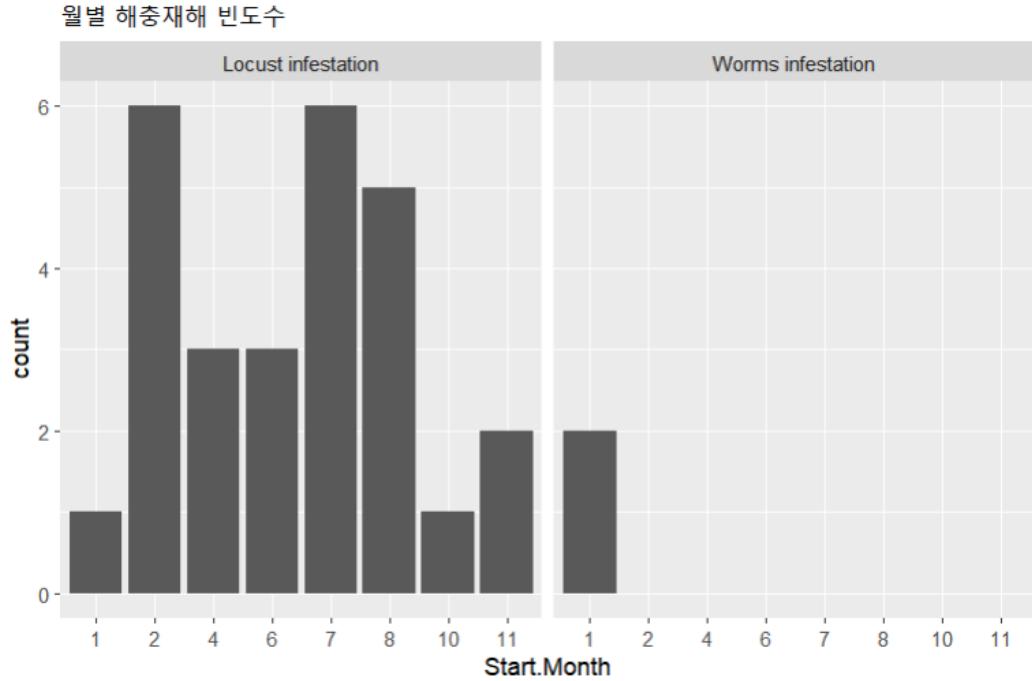
Infestation (Worms infestation) - 해충 침입 (벌레 침입)

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Infestation")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "년도별 해충재해 빈도수")
```

년도별 해충재해 빈도수



```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Infestation")%>%
  mutate(Start.Month = factor(Start.Month))%>%
  ggplot(aes(x = Start.Month))+
  geom_bar()+
  facet_wrap(~Disaster.Subtype)+
  labs(title = "월별 해충재해 빈도수")
```



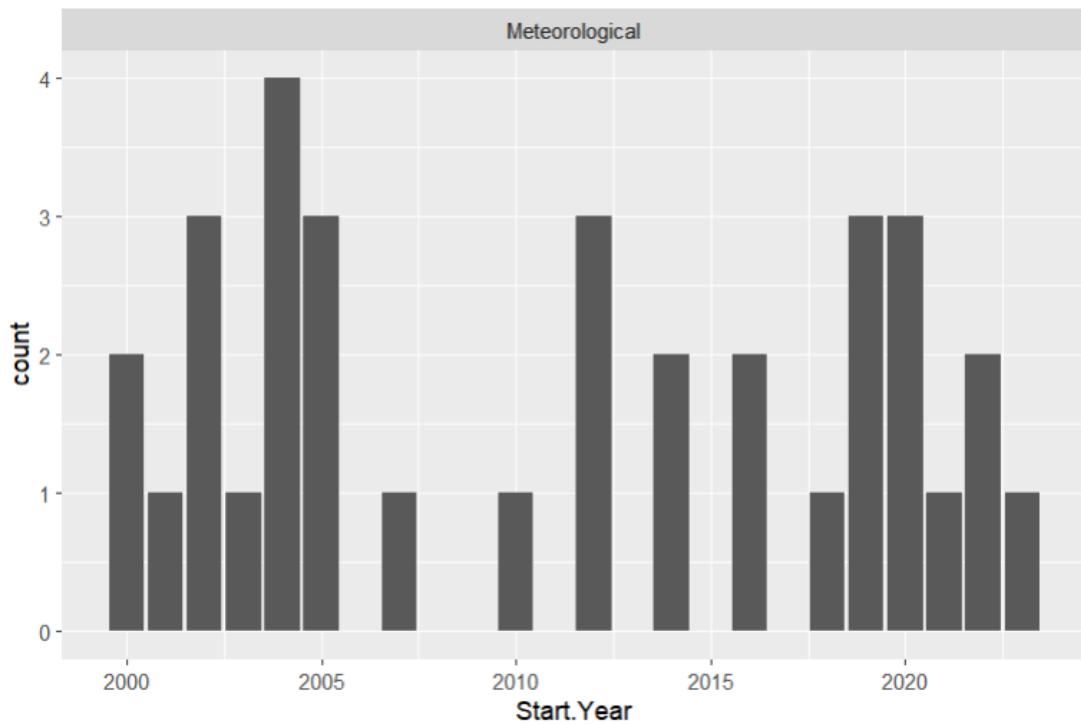
```
Disaster.Type Disaster.Subtype
1 Animal incident Animal incident
```

동물 사고의 경우 그 유형이 한가지로, type과 빈도수가 같다

▼ 한국의 연도별 기상관련 재해와 대응

초기 가설: 한국에서 발생하는 기상 재해는 주로 폭풍이 있다. 그렇다면 이 폭풍과 관련해서 우리나라의 피해는 어떤 추세를 보일까? 기술이 발전하면서 폭풍에 대한 대처가 잘 되고 있을까? 피해 규모가 점점 해마다 줄어드는지 확인해보자

```
data%>%
  group_by(Disaster.Subgroup)%>%
  filter(Disaster.Subgroup == "Meteorological")%>%
  filter(Country == "Republic of Korea") %>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  facet_wrap(~Disaster.Subgroup)
```



2000년도부터 한국의 기상관련 재해 빈도수를 체크

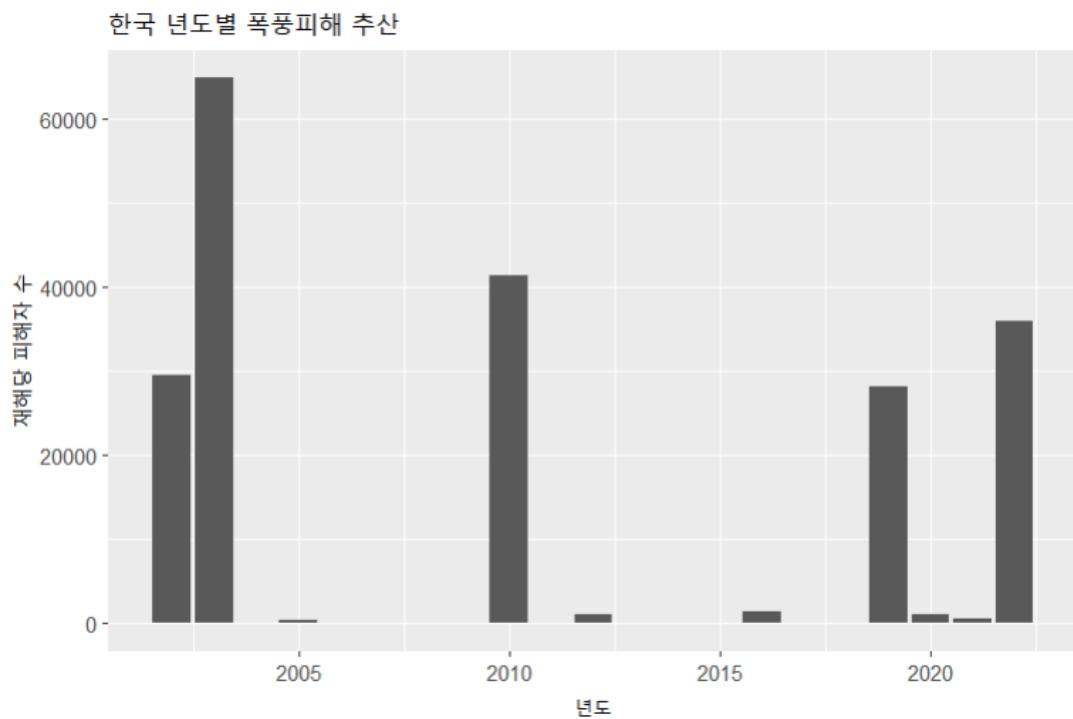
빈도는 년 평균 2회 정도 되는 것 같다. 그렇다면 연도별 회당 피해상황을 확인해보고, 기상관련 재해에 관하여 특별한 점이 있는지 체크

```
data%>%
  filter(Disaster.Type == "Extreme temperature")%>%
  filter(Country == "Republic of Korea") %>%
  group_by(Start.Year) %>%
  summarise(
    Total_Affected = sum(No..Affected, na.rm = TRUE),
    # 피해자 수 총합
    Frequency = n(),
    # 재해 발생 빈도 (건수)
  )
```

	Start.Year	Total_Affected	Frequency
1	2016	0	1
2	2018	0	1
3	2022	0	1

우선 한국의 극한 온도에 대해서는 총 3건이 존재했으며 각각에 대한 피해자는 0명 이었고, 이 외 기상재해는 전부 폭풍으로 인한 피해라고 볼 수 있다.

```
data%>%
  filter(Disaster.Type == "Storm")%>%
  filter(Country == "Republic of Korea") %>%
  group_by(Start.Year) %>%
  summarise(
    Total_Affected = sum(No..Affected, na.rm = TRUE),
    # 피해자 수 총합
    Frequency = n()
    # 재해 발생 빈도 (건수)
  ) %>%
  mutate(Affected_per_Event = Total_Affected / Frequency)%>%
  filter(Affected_per_Event > 1)%>%
  ggplot(aes(x = Start.Year, y = Affected_per_Event))+
  geom_col()+
  labs(title = "한국 년도별 폭풍피해 추산", x = "년도", y =
  "재해당 피해자 수")
```



같은 폭풍이라고 해도 그 규모나 발생한 위치, 시간대에 따라 피해의 규모가 천차만별이므로 단순하게 재해당 피해자 수로는 대처를 잘 했는지 아닌지 확인하기가 어렵다. 이는 폭풍 뿐 아니라 자연재해의 특성이 아닐까 한다.

▼ 아프리카의 전염병 발생과 전 세계 발생의 추이

아프리카의 경우 가장 낙후된 지역으로, 위생과 전염병에 취약하며 그로 인해 가장 전염병의 발생 빈도가 높을 것이라 예상한다. 또한, 최근 뉴스에 나왔던 엠포스의 경우 그 발원지가 아프리카인것으로 밝혀져 아프리카의 세균 혹은 바이러스성 전염병이 아프리카에서 많이 발생하고 세계적으로 확산되는지 확인이 필요하다

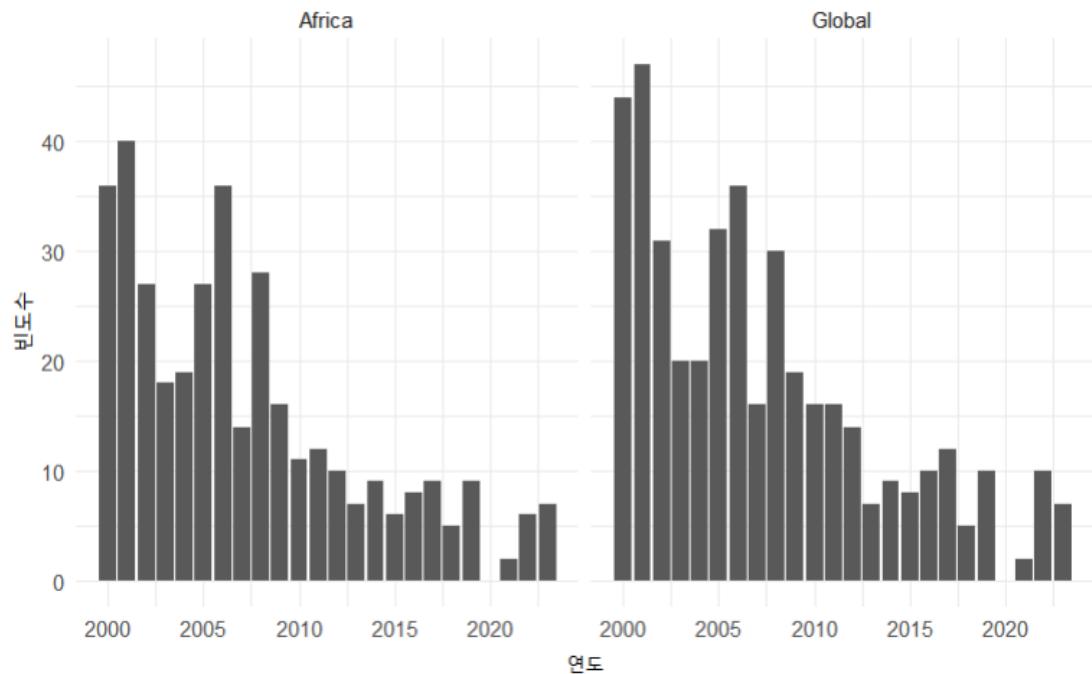
```
data_africa <- data %>%
  filter(Disaster.Subtype == "Bacterial disease") %>%
  filter(Region == "Africa") %>%
  mutate(Region = "Africa") # 지역을 "Africa"로 명시

# 두 번째 데이터에 "Global" 명시
data_global <- data %>%
  filter(Disaster.Subtype == "Bacterial disease") %>%
  mutate(Region = "Global") # 전 세계 데이터를 "Global"로 명시

# 두 데이터를 결합
combined_data <- bind_rows(data_africa, data_global)

# 결합된 데이터로 시각화
combined_data %>%
  ggplot(aes(x = Start.Year)) +
  geom_bar() +
  facet_wrap(~Region) + # Region별로 시각화
  labs(title = "아프리카 및 전 세계 연도별 박테리아 질환 빈도수",
       x = "연도",
       y = "빈도수") +
  theme_minimal()
```

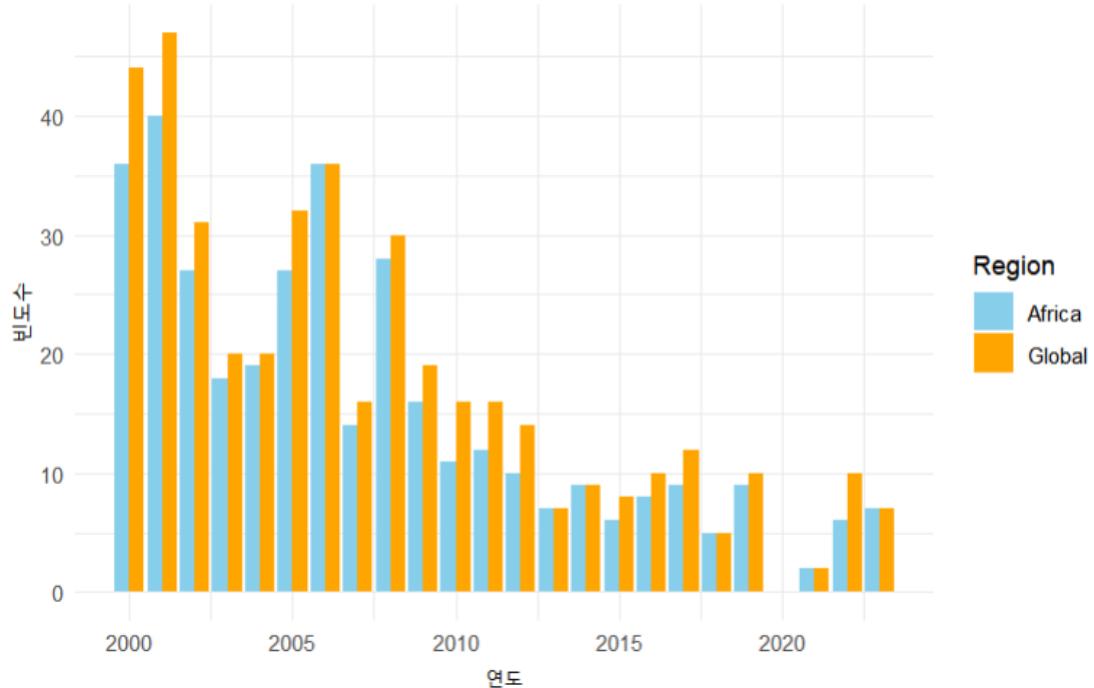
아프리카 및 전 세계 연도별 박테리아 질환 빈도수



조금 더 보기 편하도록 두 시각화 그래프를 합쳐보면

```
combined_data %>%
  ggplot(aes(x = Start.Year, fill = Region)) +
  geom_bar(position = "dodge") + # Region별로 나란히 표시
  labs(title = "아프리카 및 전 세계 연도별 바이러스성 질환 빈도수",
       x = "연도",
       y = "빈도수") +
  scale_fill_manual(values = c("Africa" = "skyblue",
                               "Global" = "orange")) + # 색상 설정
  theme_minimal()
```

아프리카 및 전 세계 연도별 바이러스성 질환 빈도수



아프리카의 박테리아 질환 빈도수는 전 세계 빈도수의 추이와 동일하다는 것을 확인할 수 있으며 대부분의 박테리아 질환이 아프리카에서 많이 발생함을 확인할 수 있다.

그렇다면 세균성 질환은 어떨까?

```
data_africa1 <- data %>%
  filter(Disaster.Subtype == "Viral disease") %>%
  filter(Region == "Africa") %>%
  mutate(Region = "Africa") # 지역을 "Africa"로 명시

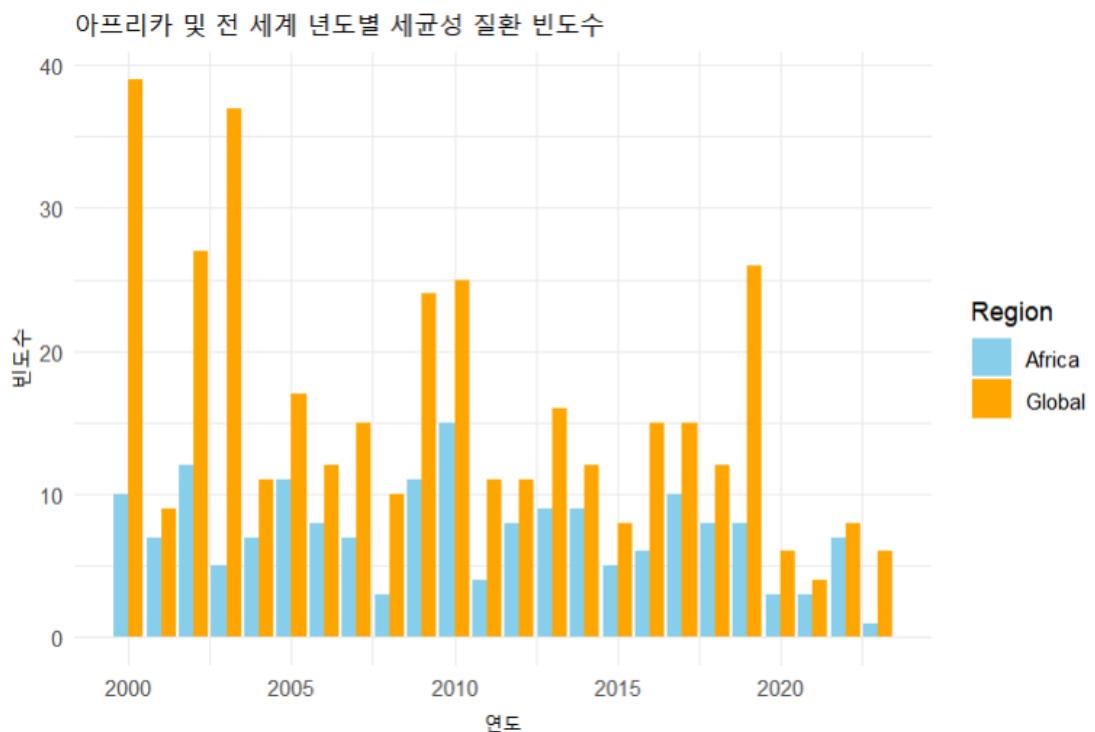
# 두 번째 데이터에 "Global" 명시
data_global1 <- data %>%
  filter(Disaster.Subtype == "Viral disease") %>%
  mutate(Region = "Global") # 전 세계 데이터를 "Global"로 명시

# 두 데이터를 결합
combined_data1 <- bind_rows(data_africa1, data_global1)
```

```

# 결합된 데이터로 시각화
combined_data1 %>%
  ggplot(aes(x = Start.Year, fill = Region)) +
  geom_bar(position = "dodge") + # Region별로 나란히 표시
  labs(title = "아프리카 및 전 세계 연도별 세균성 질환 빈도수",
       x = "연도",
       y = "빈도수") +
  scale_fill_manual(values = c("Africa" = "skyblue",
                               "Global" = "orange")) + # 색상 설정
  theme_minimal()

```



세균성 질환은 바이러스성 질환에 비해서는 전 세계적 비중보다는 적으나 그래도 큰 부분을 차지하고 있다.

다음으로는 피해자 수를 비교해보자

```

data_africa2 <- data %>%
  filter(Disaster.Subtype == "Bacterial disease") %>%
  filter(Region == "Africa") %>%
  group_by(Start.Year) %>%

```

```

    summarise(
      Total_Affected = sum(No..Affected, na.rm = TRUE)
    ) %>%
    filter(Total_Affected > 1) %>%
    mutate(Region = "Africa") # 지역을 "Africa"로 명시

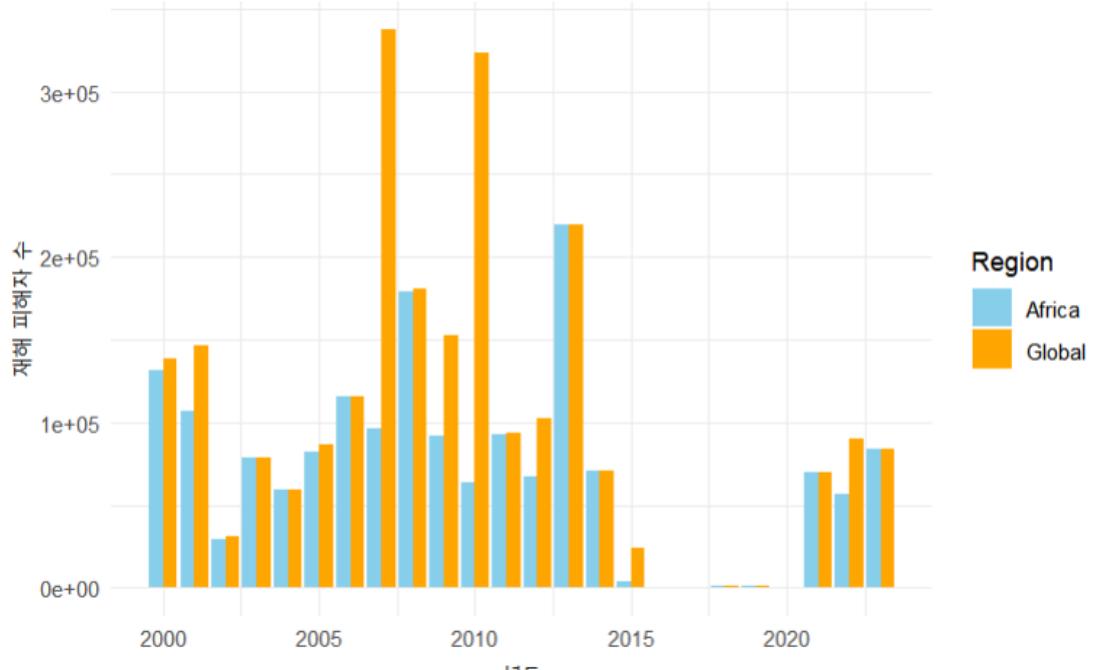
# 두 번째 데이터(Global) 처리
data_global2 <- data %>%
  filter(Disaster.Subtype == "Bacterial disease") %>%
  group_by(Start.Year) %>%
  summarise(
    Total_Affected = sum(No..Affected, na.rm = TRUE)
  ) %>%
  filter(Total_Affected > 1) %>%
  mutate(Region = "Global") # 지역을 "Global"로 명시

# 두 데이터를 결합
combined_data2 <- bind_rows(data_africa2, data_global2)

# 결합된 데이터로 시각화
combined_data2 %>%
  ggplot(aes(x = Start.Year, y = Total_Affected, fill = Region)) +
  geom_col(position = "dodge") + # 두 데이터를 나란히 비교
  labs(title = "년도별 세균성 질환 피해 추산 (아프리카 및 전 세계)",
       x = "년도",
       y = "재해 피해자 수") +
  scale_fill_manual(values = c("Africa" = "skyblue",
                               "Global" = "orange")) + # 색상 설정
  theme_minimal()

```

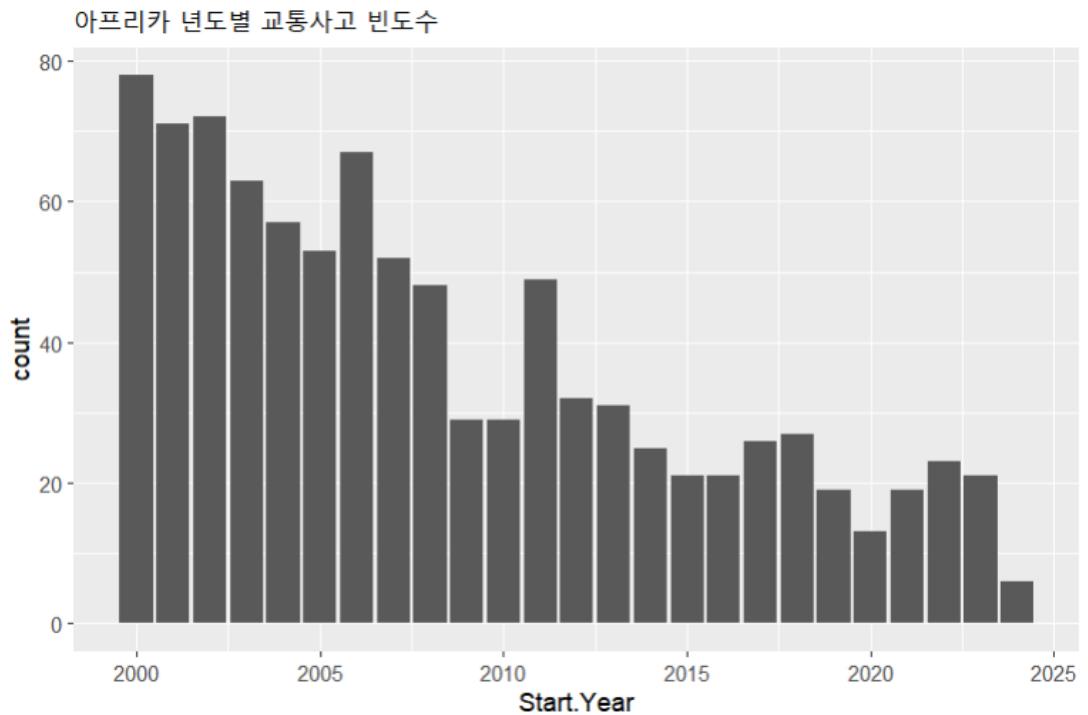
년도별 바이러스성 질환 피해 추산 (아프리카 및 전 세계)



▼ 유럽과 아프리카 교통사고 빈도수 및 피해자 추산

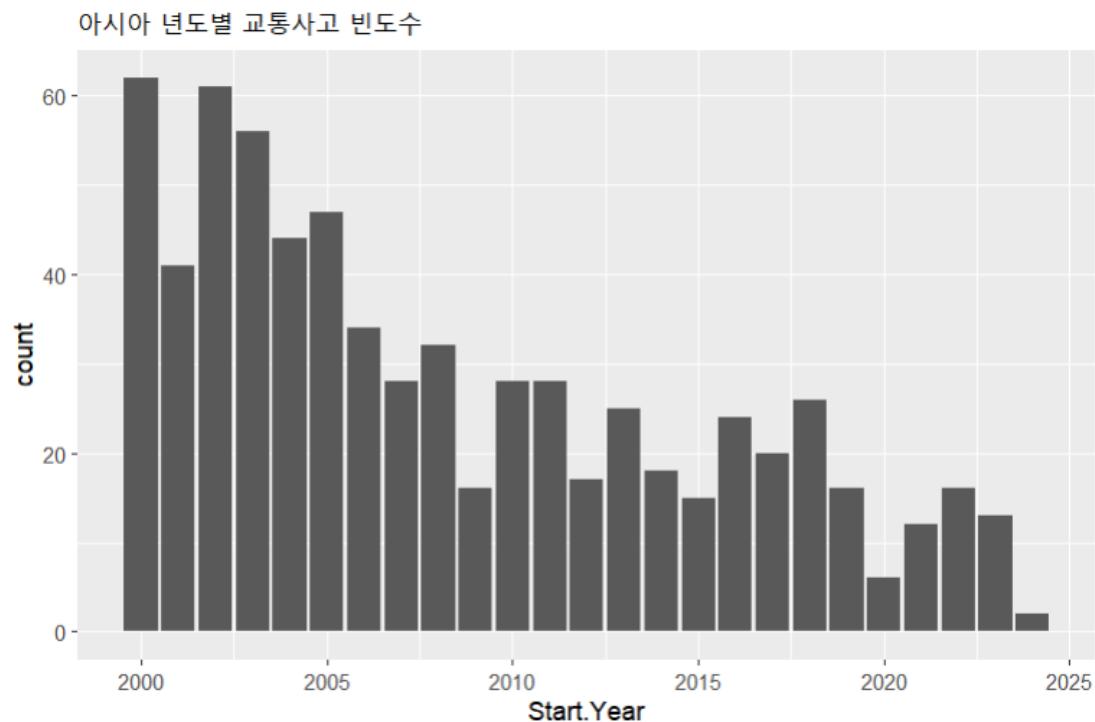
교통사고 road의 경우 유럽과 같이 선진화된 대륙은 인프라가 잘 갖추어져 사고의 빈도가 적을 것이고, 반대로 아프리카는 도로가 깔리지 않아 교통사고의 빈도가 높을 것이라 예상한다.

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Road")%>%
  filter(Region == "Africa") %>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  labs(title = "아프리카 연도별 교통사고 빈도수")
```



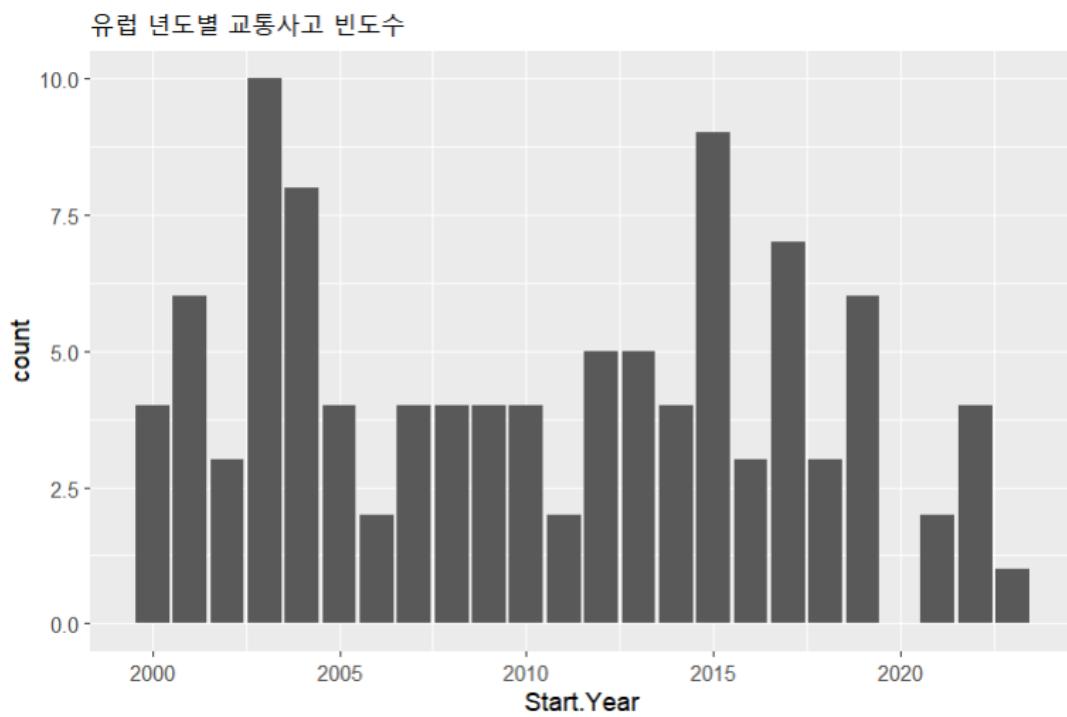
아프리카의 연도별 교통사고 빈도수는 해마다 감소하는 추세로 볼 수 있다.

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Road")%>%
  filter(Region == "Asia") %>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  labs(title = "아시아 연도별 교통사고 빈도수")
```



아시아의 경우에도 마찬가지로 줄어드는 추세이며 아프리카와 비교했을 때에도 조금 더 빈도가 적다 그렇다면 유럽은 어떨까?

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Road")%>%
  filter(Region == "Europe") %>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  labs(title = "유럽 년도별 교통사고 빈도수")
```



유럽은 이미 교통 인프라가 갖춰졌기에 안정화 단계에 왔다고 볼 수 있을 것 같다.
그런데 이후 피해자를 추산하는 과정에서 이상한 점을 발견했다.

```
data %>%
  filter(Disaster.Type == "Road") %>%
  filter(Region == "Asia") %>%
  select(Start.Year, Country, Disaster.Type, No..Affected) %>%
  filter(!is.na(No..Affected)) %>%
  group_by(Start.Year)
```

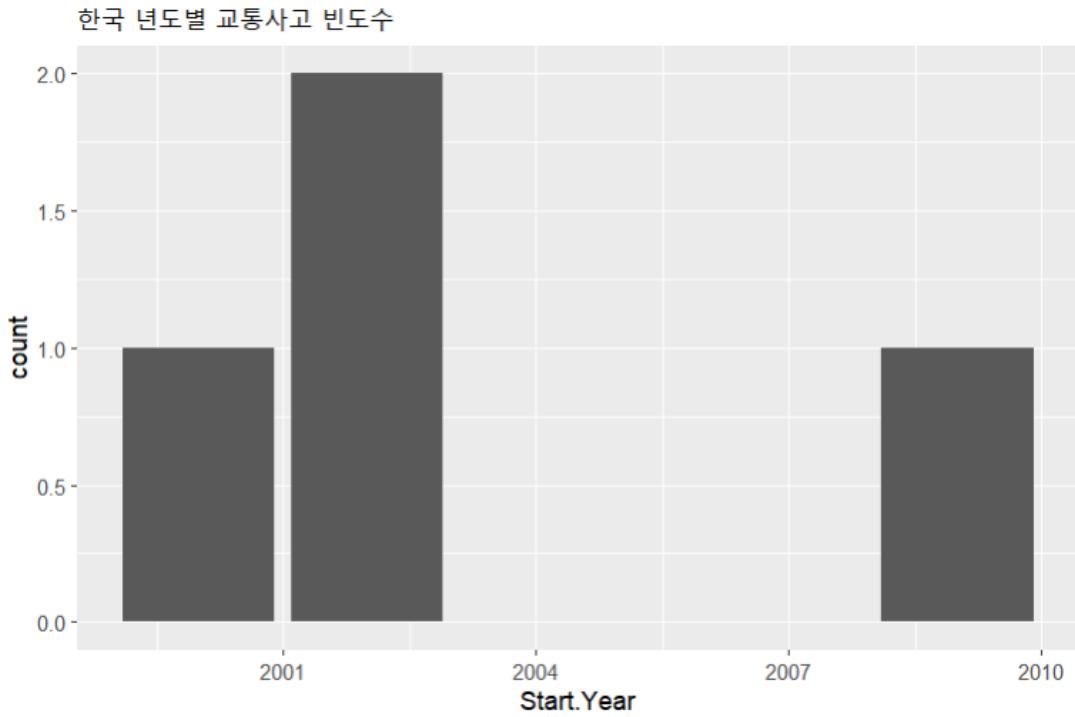
Start.Year	Country	Disaster.Type	No..Affected
1	2003 India	Road	7
2	2005 Viet Nam	Road	1
3	2005 India	Road	42
4	2005 India	Road	27
5	2006 Bangladesh	Road	10
6	2006 Nepal	Road	9
7	2006 India	Road	5
8	2006 India	Road	4

9	2010	Pakistan	Road	10
10	2010	India	Road	40
11	2010	Pakistan	Road	5
12	2011	India	Road	10
13	2011	Philippines	Road	1
14	2020	Pakistan	Road	7
15	2021	India	Road	7

아시아의 국가별 교통사고 피해자 수를 확인한 것인데, NA값을 제외하고 보면 피해자 확인이 된 것이 15건밖에 되지 않는다. 우리가 생각하는 일반적인 차량 교통사고가 아닌 것 같다.

이후 다시 우리나라의 교통사고 빈도수를 확인하니

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Type == "Road")%>%
  filter(Region == "Asia") %>%
  filter(Country == "Republic of Korea")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  labs(title = "한국 년도별 교통사고 빈도수")
```

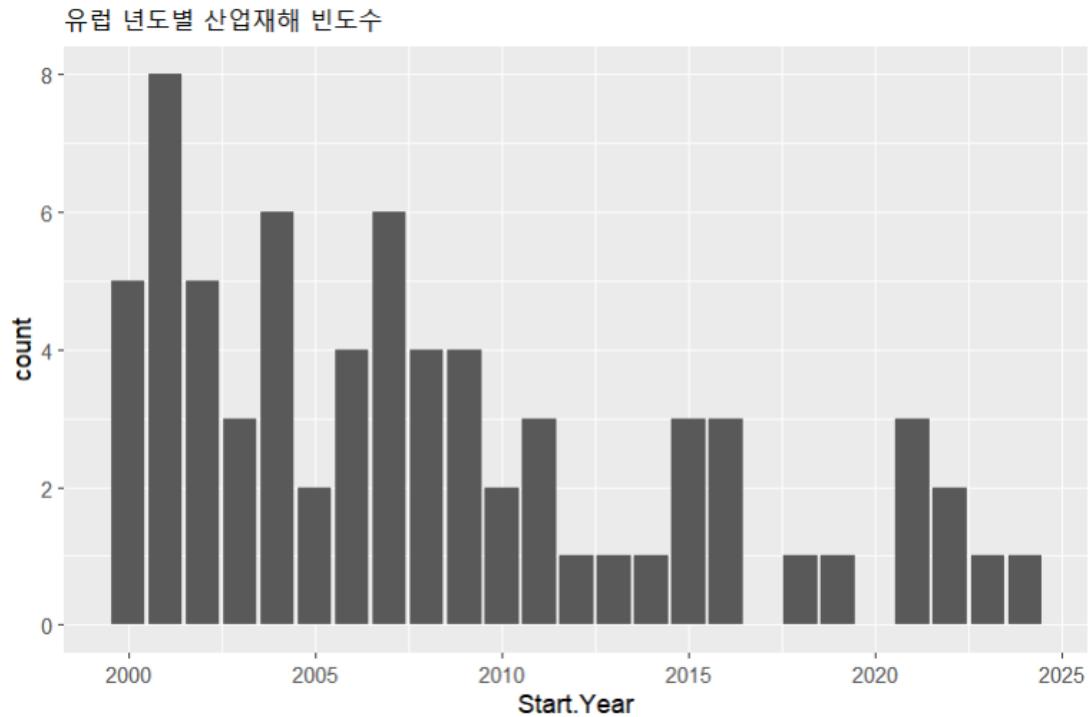


확인된 것이 총 4건으로 데이터가 담기지 않았거나, 우리가 생각하는 일반적인 교통사고가 아닌 것이라고 생각한다.

▼ 산업 재해의 빈도와 피해자

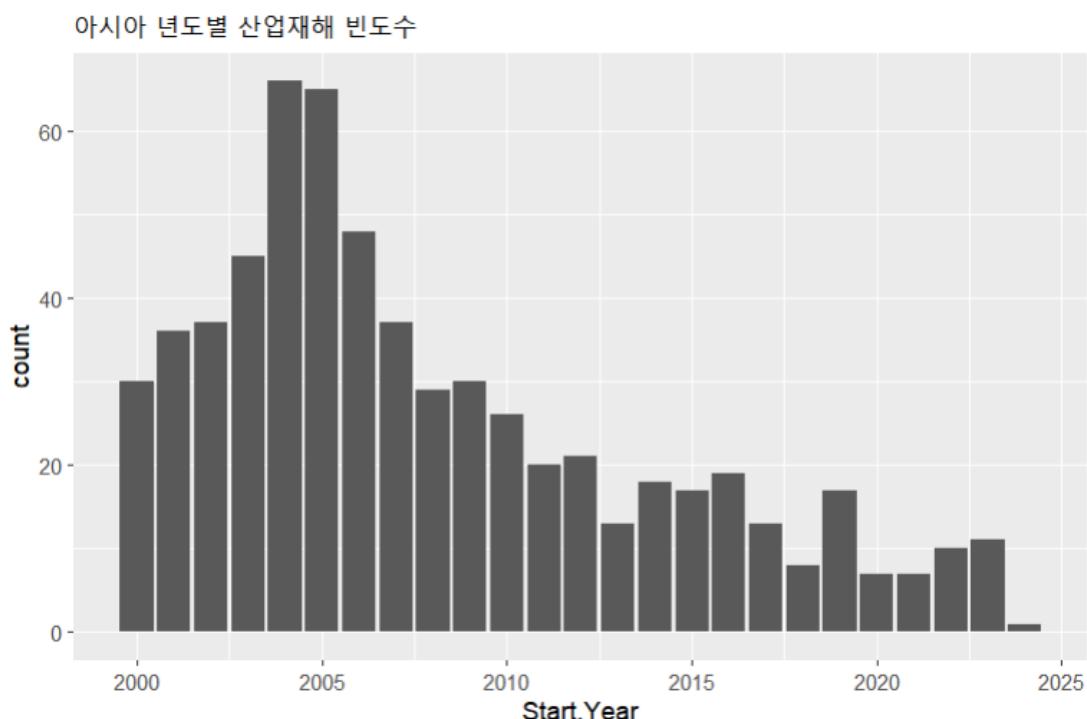
가장 선진화된 유럽의 년도별 산업재해 빈도와 피해자를 확인한 뒤, 개발 도상국이라고 할 수 있는 아시아 지역과 후진국이라고 할 수 있는 아프리카의 산업재해 빈도와 피해자를 확인해 보겠다.

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  filter(Region == "Europe") %>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  labs(title = "유럽 년도별 산업재해 빈도수")
```



가장 먼저 유럽의 산업재해 역시 점점 감소하며 2000년대에 이미 산업재해가 상당히 적음을 확인할 수 있다.

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  filter(Region == "Asia") %>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  labs(title = "아시아 년도별 산업재해 빈도수")
```

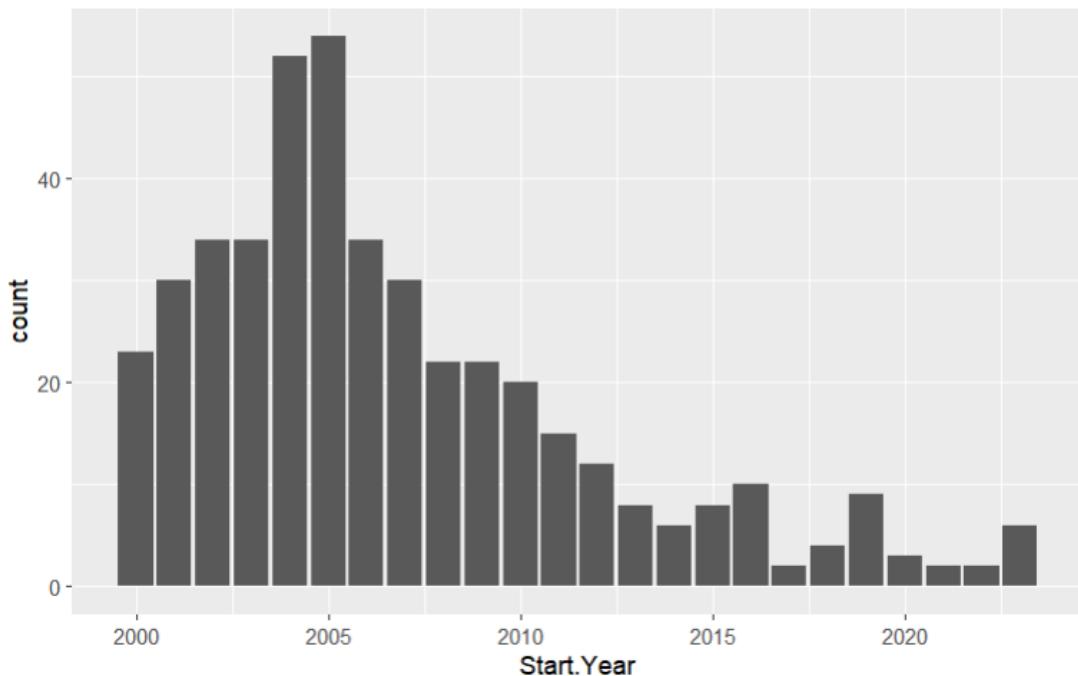


다음으로 아시아의 경우 역시 점점 감소하고 있지만 2005년도에 급격히 산업재해가 올라감을 확인할 수 있다.

그중에서 중국에 대한 내용을 확인

```
data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  filter(Region == "Asia") %>%
  filter(Country == "China")%>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  labs(title = "중국 연도별 산업재해 빈도수")
```

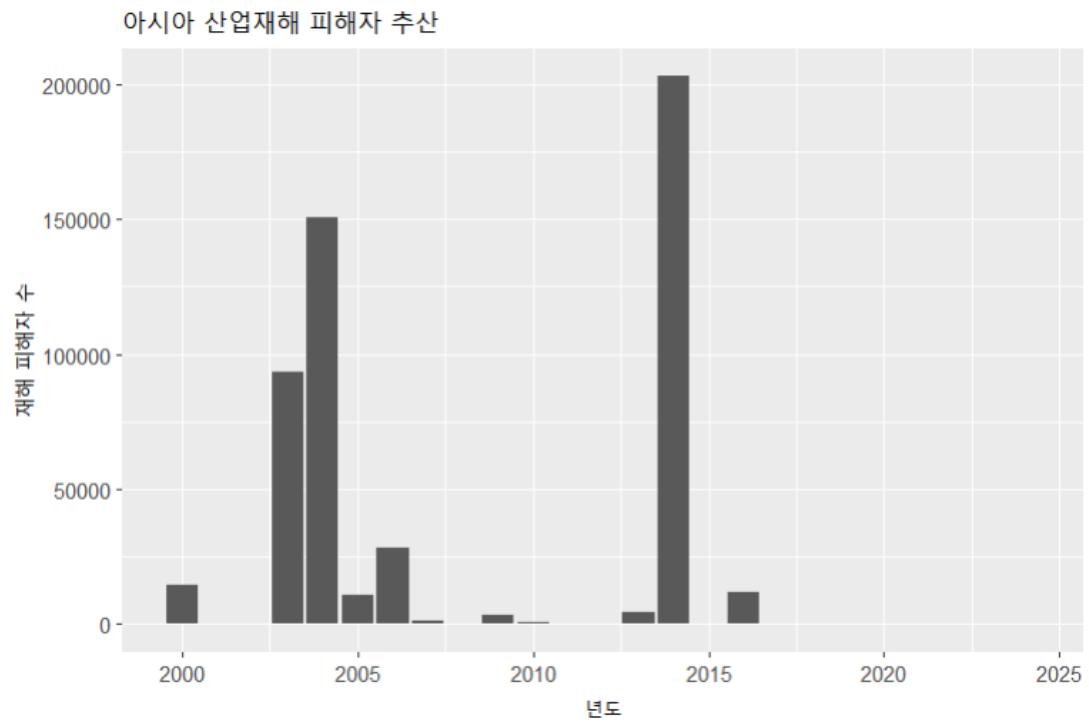
중국 년도별 산업재해 빈도수



중국의 산업재해는 아시아 전체의 산업재해와 매우 비슷한 비율이다.

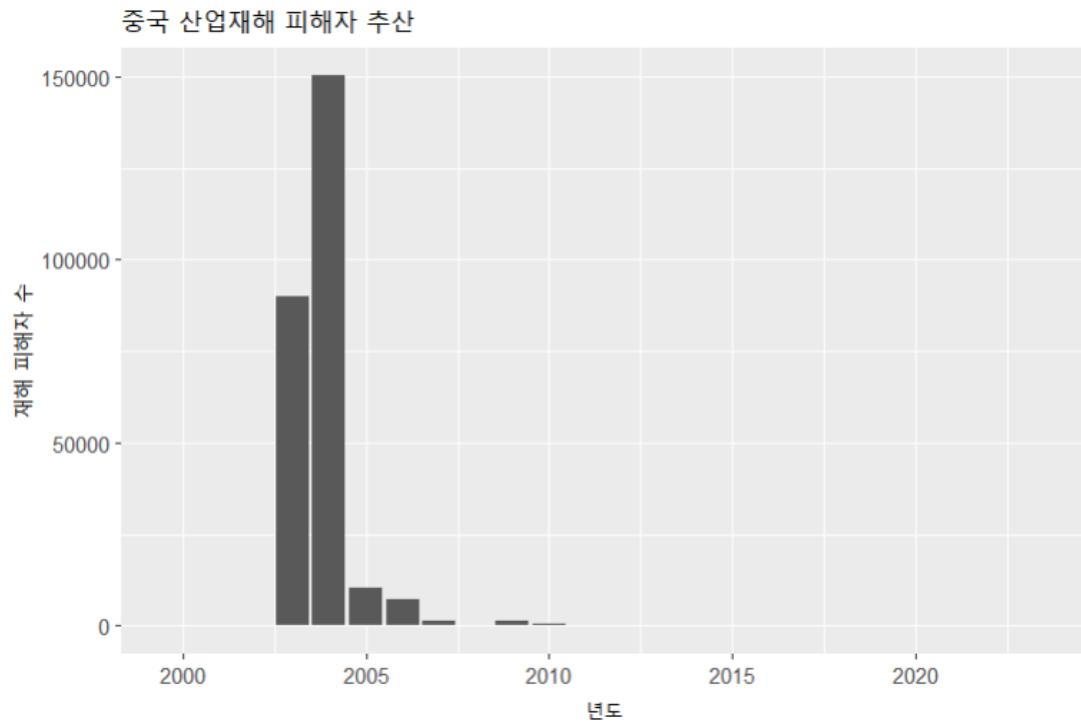
이번엔 산업피해자 수를 확인해본다

```
data%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  filter(Region == "Asia")%>%
  group_by(Start.Year) %>%
  summarise(
    Total_Affected = sum(No..Affected, na.rm = TRUE),
    # 피해자 수 총합
    # 재해 발생 빈도 (건수)
  )%>%
  ggplot(aes(x = Start.Year, y = Total_Affected))+
  geom_col()+
  labs(title = "아시아 산업재해 피해자 추산", x = "년도", y = "재해 피해자 수")
```



```

data%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  filter(Country == "China")%>%
  group_by(Start.Year) %>%
  summarise(
    Total_Affected = sum(No..Affected, na.rm = TRUE),
    # 피해자 수 총합
    # 재해 발생 빈도 (건수)
  )%>%
  ggplot(aes(x = Start.Year, y = Total_Affected))+
  geom_col()+
  labs(title = "중국 산업재해 피해자 추산", x = "년도", y =
  "재해 피해자 수")
  
```



아시아와 중국의 산업재해 피해자 수를 확인해본 결과 2004년 많이 발생했던 산업 재해에서는 중국의 피해자 수와 일치하고, 2013년에도 많은 피해자가 발생한 사고가 있었는데,

```
data%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  filter(Region == "Asia")%>%
  filter(Start.Year == 2014)%>%
  select(Country, No..Affected)%>%
  filter(!is.na(No..Affected))
```

Country	No..Affected
T\xfcrkiye	1
Maldives	203000

몰디브에서 20만명이 피해를 입은 산업재해가 있었는데, 하나는 하수 처리 회사의 해수 담수화 기술에 문제가 있어 20만명 이상이 물 부족 사태를 겪은 사건이 있다.

<https://thediplomat.com/2014/12/maldives-faces-drinking-water-crisis/>

Maldives Faces Drinking Water Crisis

A fire at the water sewage treatment plant in the country's capital has left at least 100,000 without drinking water.

D <https://thediplomat.com/2014/12/maldives-faces-drinking-water-crisis/>



```
# 첫 번째 데이터 (Asia) 처리
data_asia4 <- data %>%
  filter(Disaster.Subgroup == "Industrial accident") %>%
  filter(Region == "Asia") %>%
  group_by(Start.Year) %>%
  summarise(Frequency = n()) %>%
  mutate(Region = "Asia") # 지역을 "Asia"로 명시

# 두 번째 데이터 (China) 처리
data_china4 <- data %>%
  filter(Disaster.Subgroup == "Industrial accident") %>%
  filter(Region == "Asia") %>%
  filter(Country == "China") %>%
  group_by(Start.Year) %>%
  summarise(Frequency = n()) %>%
  mutate(Region = "China") # 지역을 "China"로 명시

# 두 데이터를 결합
combined_data4 <- bind_rows(data_asia4, data_china4)

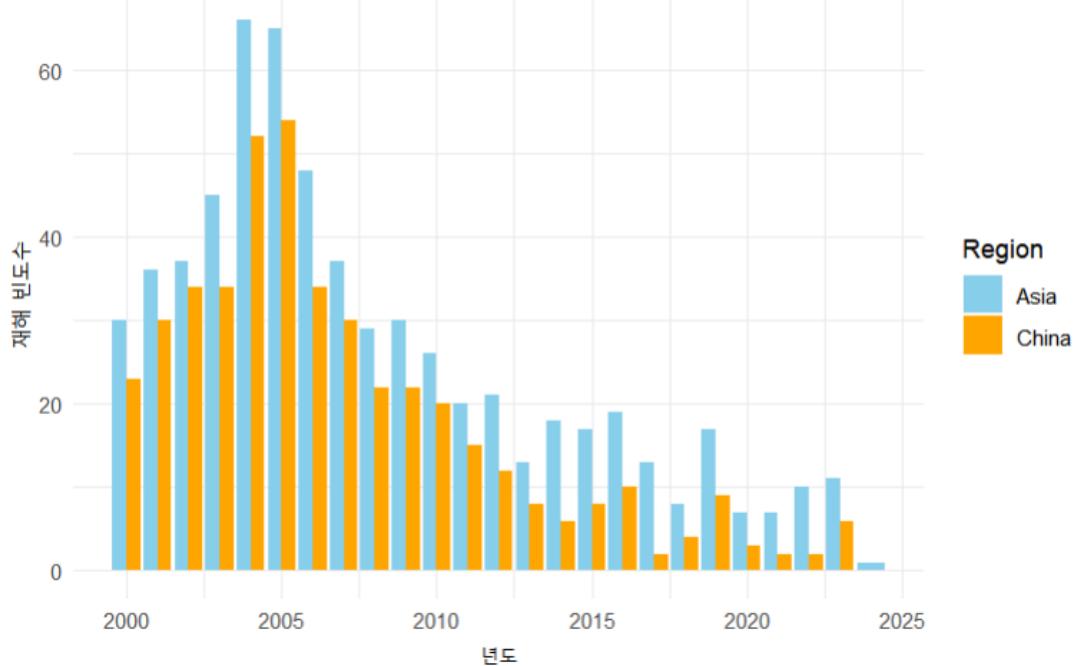
# 결합된 데이터로 시각화
combined_data4 %>%
  ggplot(aes(x = Start.Year, y = Frequency, fill = Region)) +
  geom_bar(stat = "identity", position = "dodge") +
  # 막대 나란히 표시
  labs(title = "아시아 및 중국의 연도별 산업재해 빈도수",
```

```

x = "년도",
y = "재해 빈도수") +
scale_fill_manual(values = c("Asia" = "skyblue", "China" = "orange")) + # 색상 설정
theme_minimal()

```

아시아 및 중국의 연도별 산업재해 빈도수

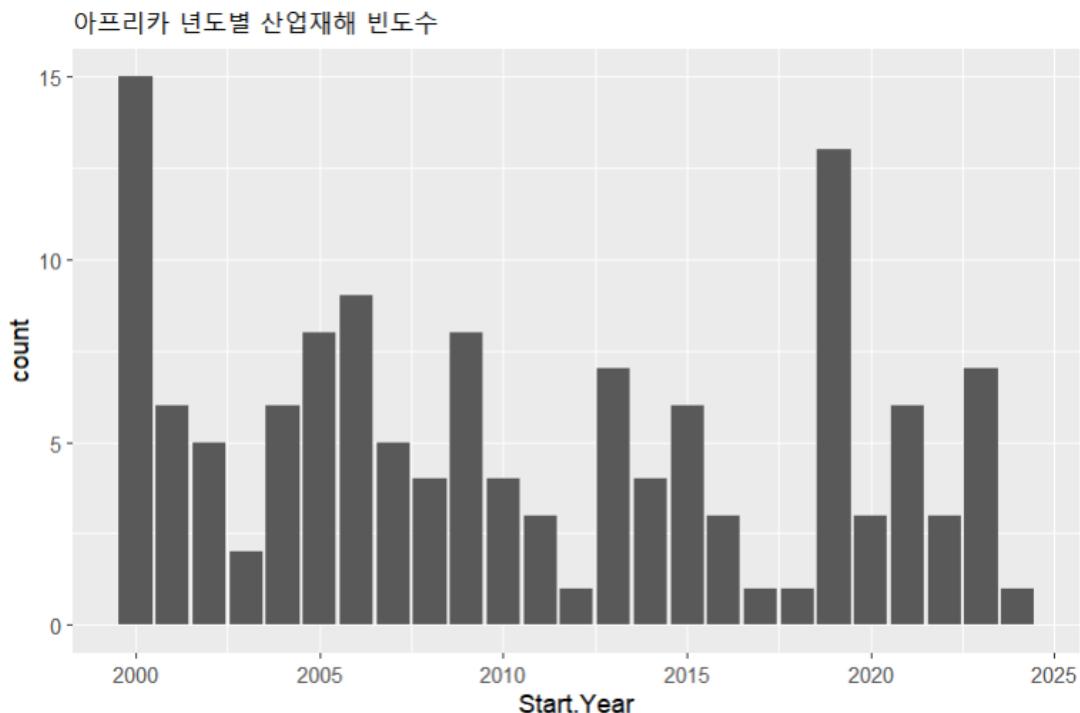


중국의 산업재해가 아시아 전체의 산업재해의 많은 비중을 가짐을 알 수 있다.

```

data%>%
  group_by(Disaster.Group)%>%
  filter(Disaster.Subgroup == "Industrial accident")%
>%
  filter(Region == "Africa") %>%
  ggplot(aes(x = Start.Year))+
  geom_bar()+
  labs(title = "아프리카 연도별 산업재해 빈도수")

```



아프리카의 경우 오히려 산업재해가 아시아보다 적음을 확인할 수 있음.

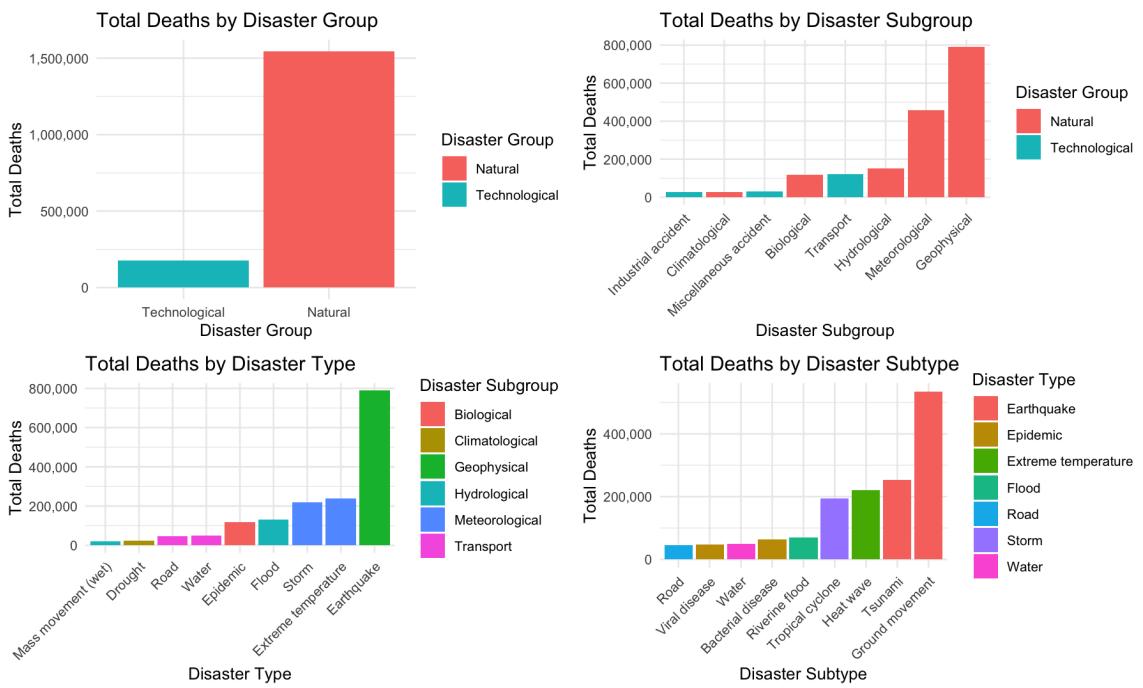
▼ Analysis of the Effect of Disaster

```
data <- read.csv("public_emdat_project.csv")
```

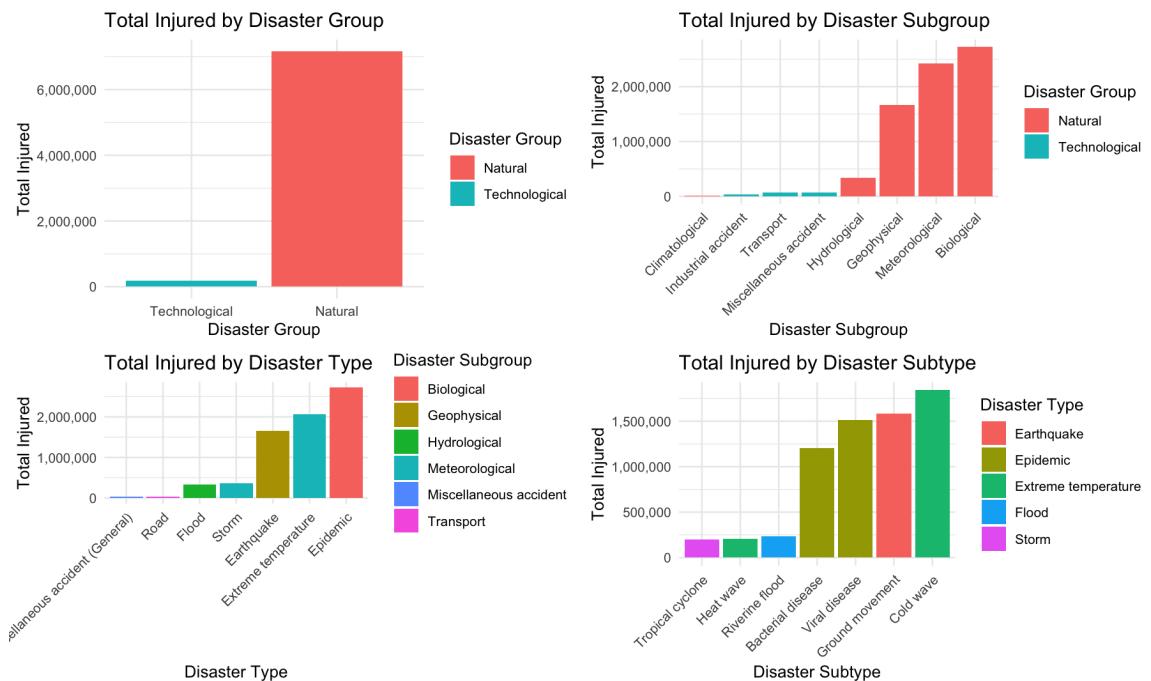
disaster로 인해서 어떤 재난이 죽음을 가장 많이 초래하였을지 분석하고자 하였다.

disaster를 Group, Subgroup, Type, Subtype의 네가지로 나누어 분류해놓았기에 모든 카테고리에서 사망자 수의 발생 추이를 확인하고자 한다.

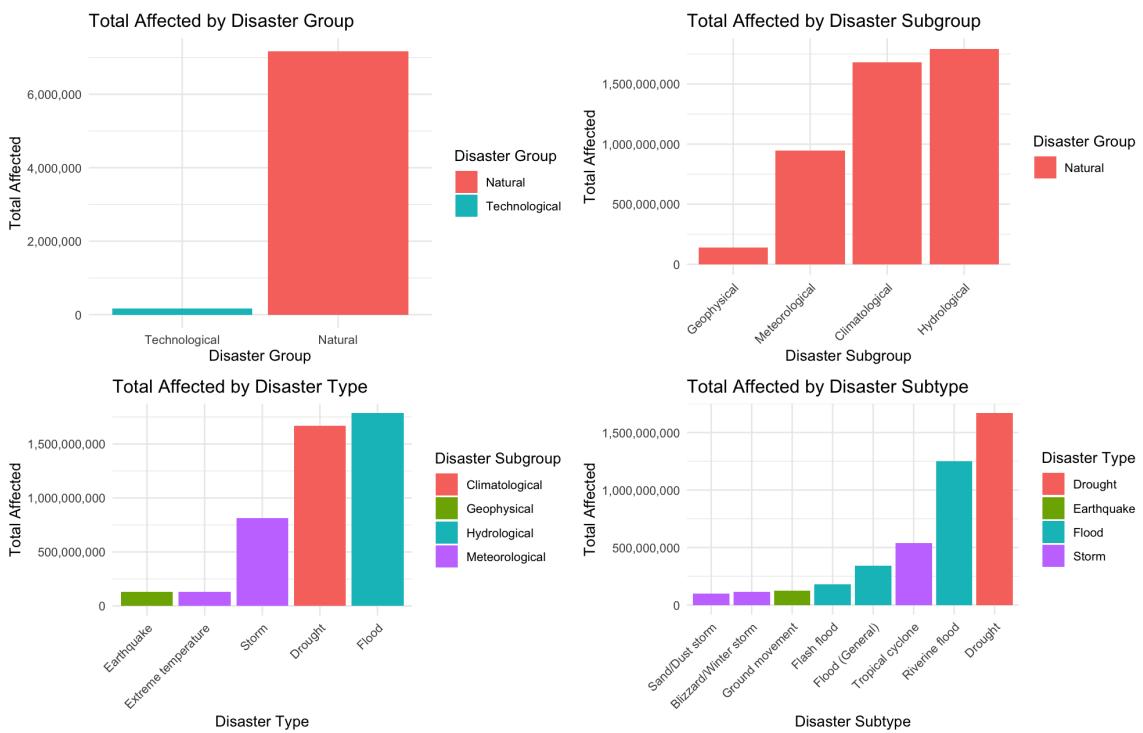
가설 : 지구온난화가 지속적으로 악화되는 상황이기에 자연재해가 가장 큰 원인일 것이고, 지진이 그 중 가장 심각한 원인일 것이다. 다양한 재해가 있지만, 지진은 발생하면 지진 발생 지역과 주변 지역까지 광범위하게 피해를 일으키며, 지각 변동과 밀접한 위치에서는 사망률이 높을 것이라 생각한다.



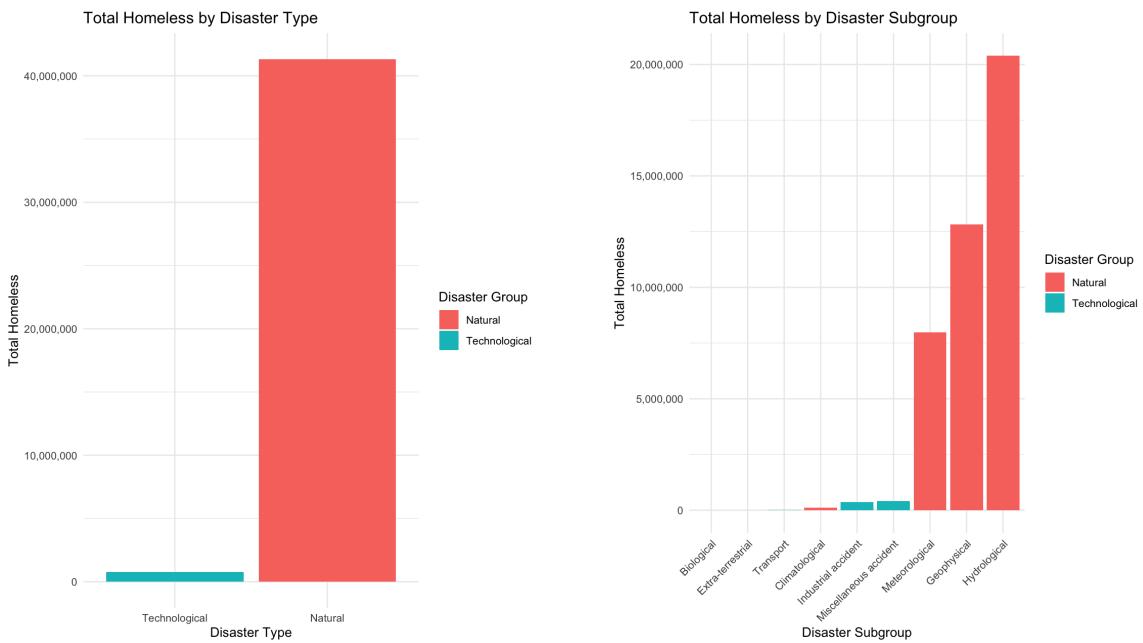
다음은 동일한 방식으로 부상자를 확인하고자 하였다.

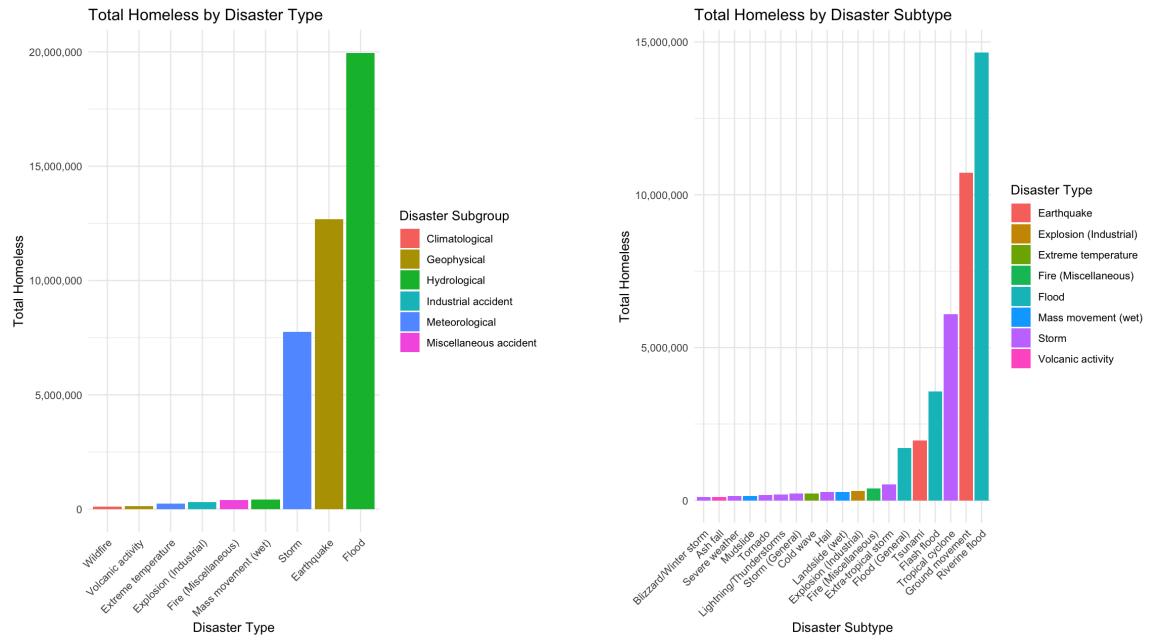


다음은 피해자 수이다.



다음은 노숙자를 확인하였다.





사망자가 가장 많이 발생한 원인은 지진이 압도적으로 우세했다.

부상자는 전염병, 이상기온, 지진 순서대로 많이 발생하였다.

피해자는 홍수, 가뭄, 폭풍 순으로 많이 발생하였고,

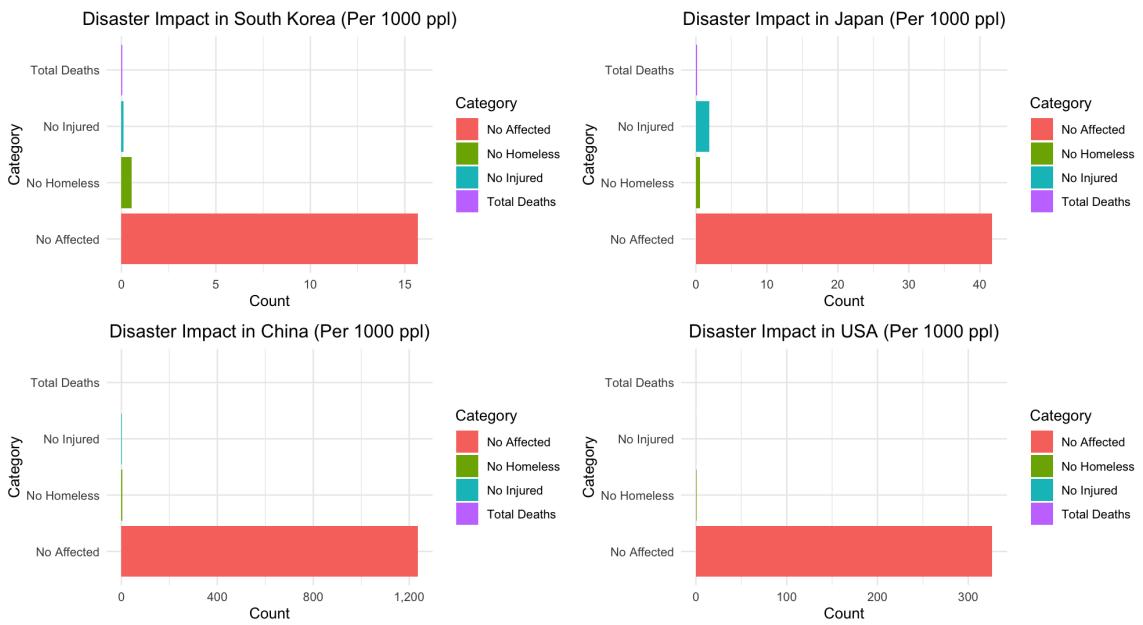
노숙자는 피해자와 비슷하게 홍수, 지진, 폭풍 순으로 주로 발생하였다.

- no.affected, total.affected의 차이점

- no.affected는 disaster로 인한 직접적인 피해를 받은 사람들의 수를 확인하고, total.affected는 injured, affected, 그리고 disaster로 인해 homeless가 된 사람들의 총합을 의미한다. 사망자를 제외한 모든 직간접적으로 피해를 입은 사람들을 포함한다고 볼 수 있다.

국가 - 한국, 일본, 중국, 미국 피해 조사

여러 국가들 중 한국, 일본, 중국, 미국에 대하여 분석해보았다. 국가마다 인구수가 상이하기에 인구수를 나누고 1000명을 기준으로 하여 1,000명당 어떤 결과를 가지는지 확 인해보았다.

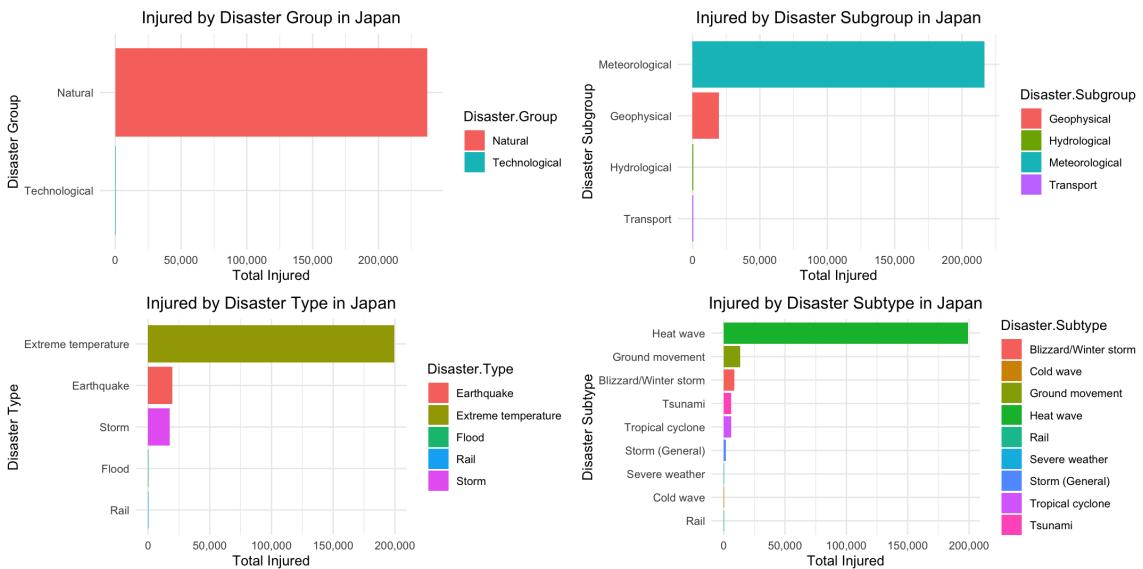


결과를 보았을 때, 사망자, 피해자, 부상자, 그리고 노숙자 중에서 네 국가 모두 피해자가 가장 많은 것으로 보인다.

중국의 피해자가 약 1236명으로 가장 많고, 그 다음은 미국이 326명, 일본 41명, 그리고 한국이 15명인 것으로 나타났다.

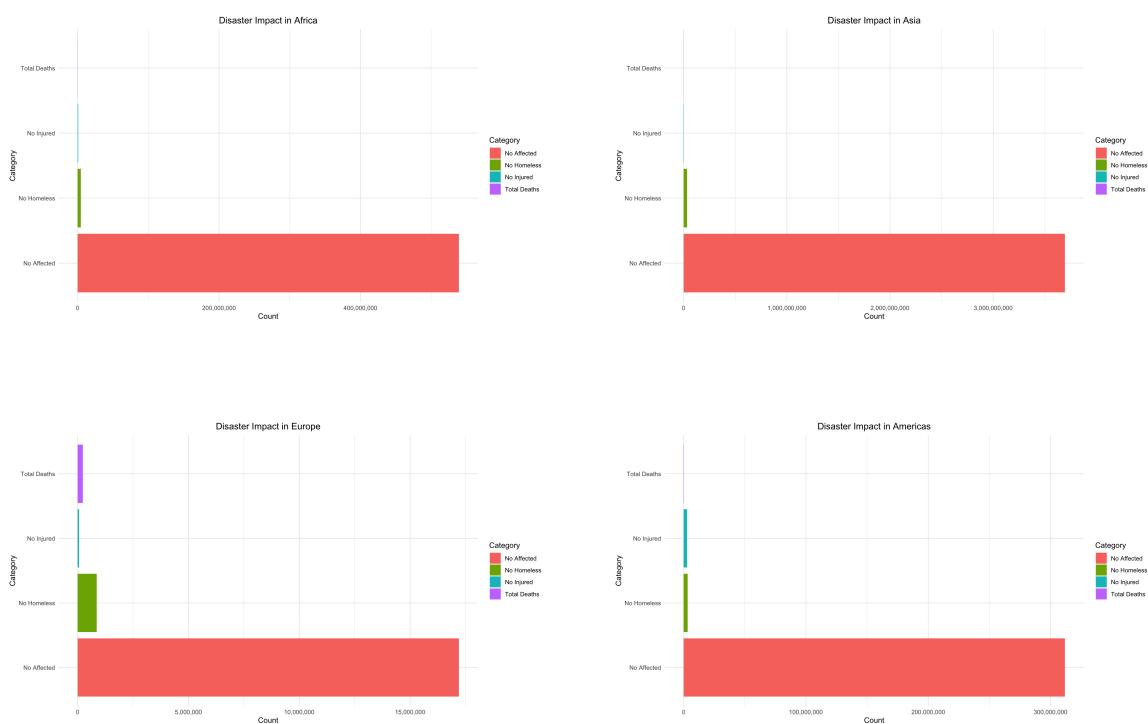
한국과 중국, 미국의 경우, disaster로 인해 노숙자가 된 인구 수가 피해자 수 다음으로 많았다.

반면 일본은 부상자의 수가 두번째로 많은 비율을 차지하는 것으로 보인다.



일본의 injured 가 어떤 disaster로 인한 것인지 알기 위해 일본의 disaster 원인을 Group, Subgroup, Type, Subtype 별로 출력하였다. 결과를 분석하였을 때, natural disaster로 인한 injured가 가장 많았다. 세부적으로는 **이상 기온으로 인한 폭염**이 198,904명으로 가장 많았고, 그 다음은 **지진**으로 인한 injured가 13,305명으로 많았다.

대륙 - 아프리카, 아시아, 유럽, 아메리카 피해 조사

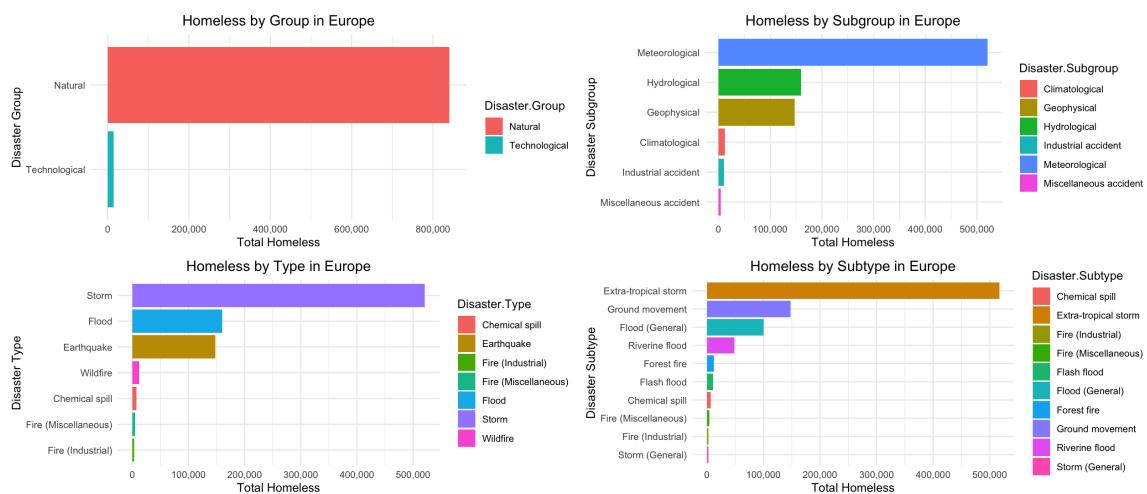


	Region	Total_Deaths	No_Injured	No_Affected	No_Homeless
1	Asia	953831	3381791	3693925006	33459028
2	Africa	231059	981144	539187134	4438049
3	Americas	299746	2898399	311693027	3254094
4	Europe	232255	65517	17200388	855628

대륙의 결과를 보았을 때, 나라별 결과와도 비슷한 양상을 띠었다. Death, Injured, Affected, Homeless 중에서 Affected의 비율이 네 대륙 모두에서 가장 높았다.

	Region	Total_Deaths	No_Injured	No_Affected	No_Homeless
1	Asia	953831	3381791	3693925006	33459028
2	Africa	231059	981144	539187134	4438049
3	Americas	299746	2898399	311693027	3254094
4	Europe	232255	65517	17200388	855628

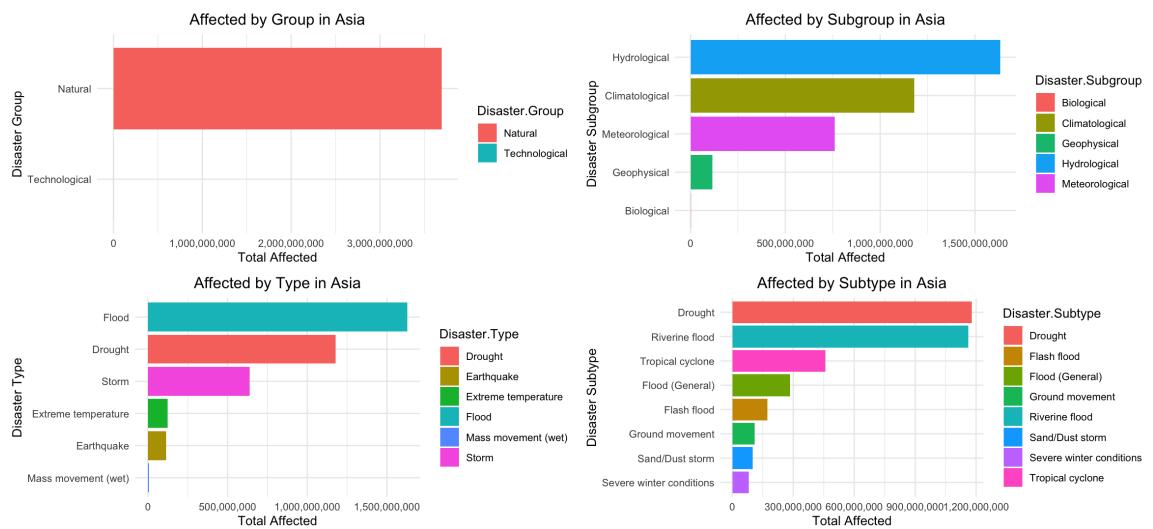
그래프를 보면, 유럽에서 두번째로 많은 비율은 homeless의 수였지만, 비율상으로만 많아보일 뿐, 실제로는 가장 적은 값을 가지고 있다.



	Disaster.Subtype	Total_Homeless
1	Extra-tropical storm	517000
2	Ground movement	147520
3	Flood (General)	100000
4	Riverine flood	48988
5	Forest fire	12059
6	Flash flood	11021
7	Chemical spill	7120
8	Fire (Miscellaneous)	4500
9	Fire (Industrial)	3082
10	Storm (General)	2377

유럽의 homeless를 그룹별로 확인하였을 때, 온대 저기압이 517,000명으로 압도적으로 높은 비율을 차지하는 것을 볼 수 있다.

네 개의 모든 대륙에서 Affected의 수가 가장 많았고, 그 중 아시아가 가장 큰 인구를 포함하고 있었다. 아시아에서 어떤 이유로 피해를 입은 사람들이 많은지 알아보기 위하여 아시아로만 분석을 돌렸다.



subtype은 다양한 종류의 flood와 drought를 세분화하였기에, Type을 확인하였고 Flood와 Drought의 피해가 가장 심각한 것으로 나타났다.

	Disaster.Type	Total_Affected
1	Flood	1626526910
2	Drought	1178290850
3	Storm	637244726
4	Extreme temperature	124317393
5	Earthquake	114372193
6	Mass movement (wet)	7740843

결과적으로, 자연재해로 인한 피해가 한국, 일본, 중국, 미국의 네 국가와 아프리카, 아시아, 유럽, 아메리카 네 대륙 모두에서 가장 컸다. 지구온난화로 인한 갑작스러운 가뭄, 홍수, 온대 저기압, 폭염, 지진 등 다양한 재해가 발생하고 있고 이로 인한 피해가 극심한 것으로 보인다.

Task2: 시각화 개선

▼ Cross-Sectional Approach

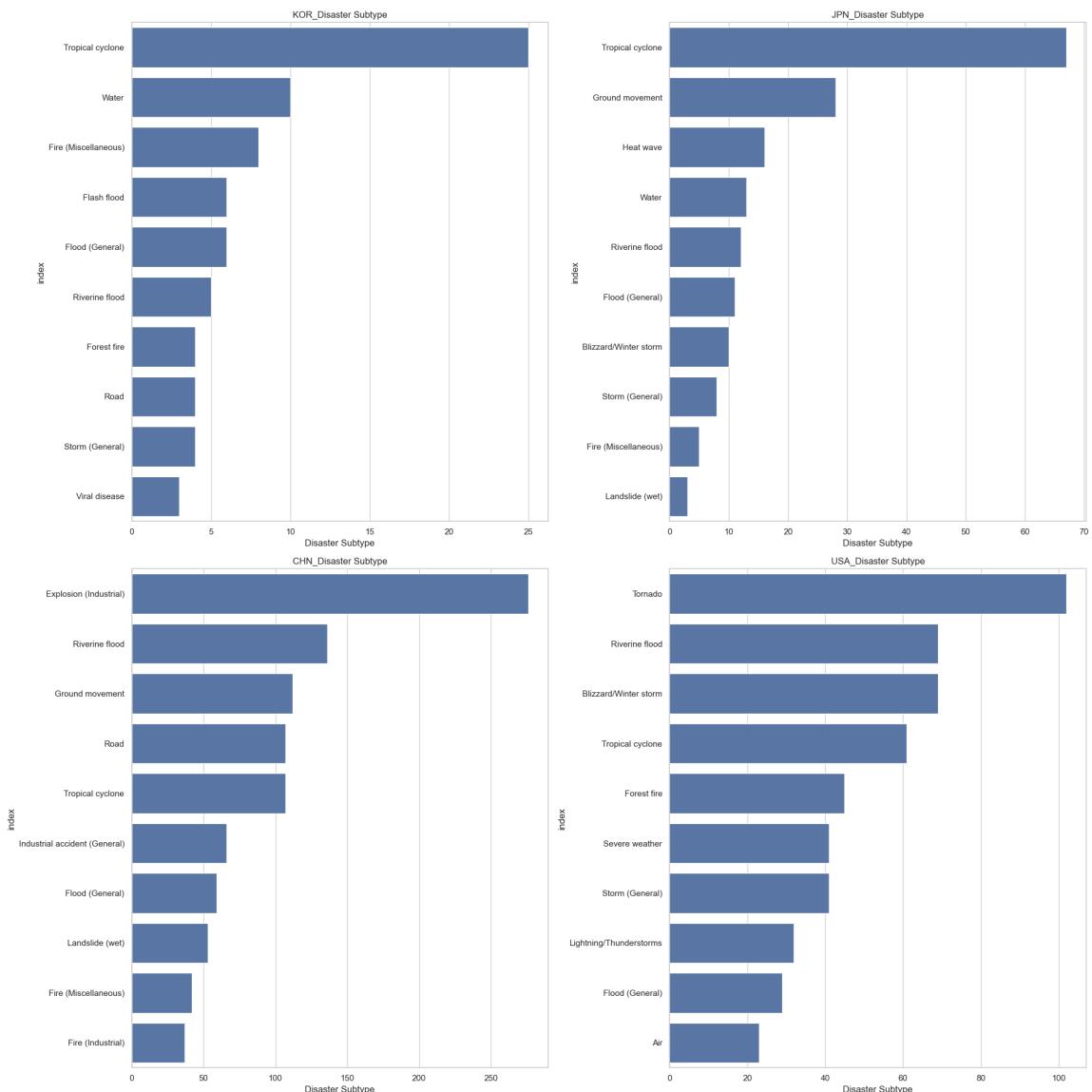
Feedback: 너무 그래프가 많고, 색이 모두 동일하여서 한눈에 정보 전달이 어렵다.

기존 시각화 코드는 국가 내의 발생 빈도를 비교하기는 쉬웠으나, 여러 국가들의 종합적인 비교는 어려웠다.

국가별 Disaster subtype을 heatmap으로 시각화 → 국가별 상위 15개 빈도만 추려내서 비교

색으로 높은 빈도들을 비교하기가 쉬워질 것이라고 생각하였음.

AS IS



TO BE

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

def plot_alternative(data1, data2, data3, data4, names=['Country 1', 'Country 2', 'Country 3', 'Country 4'], plot_type='heatmap'):
    data1['Country'] = names[0]
    data2['Country'] = names[1]
    data3['Country'] = names[2]
    data4['Country'] = names[3]
    print(data1)
    n = 15
    combined_data = pd.concat([data1.head(n), data2.head(n), data3.head(n), data4.head(n)])
    colname = combined_data.columns[1]
    plt.figure(figsize=(14, 8))

    if plot_type == 'grouped_bar':
        sns.barplot(data=combined_data, x=f'{combined_data.columns[1]}', y='index', hue='Country')
    elif plot_type == 'heatmap':
        heatmap_data = combined_data.pivot_table(index='index', columns='Country', values=f'{combined_data.columns[1]}', aggfunc='sum')
        sns.heatmap(heatmap_data, annot=True, fmt=".0f", cmap='viridis')
    elif plot_type == 'box':
        sns.boxplot(data=combined_data, x=f'{combined_data.columns[1]}', y='index', hue='Country')

    plt.title(f'{plot_type.capitalize()} Comparison by Country')
    plt.ylabel(f'{combined_data.columns[1]}')
    plt.xlabel('Country')

    plt.show()

```

해당 함수를 구현하여 모든 disaster group에 대하여 country/continent 단위로 분석 할 수 있다.

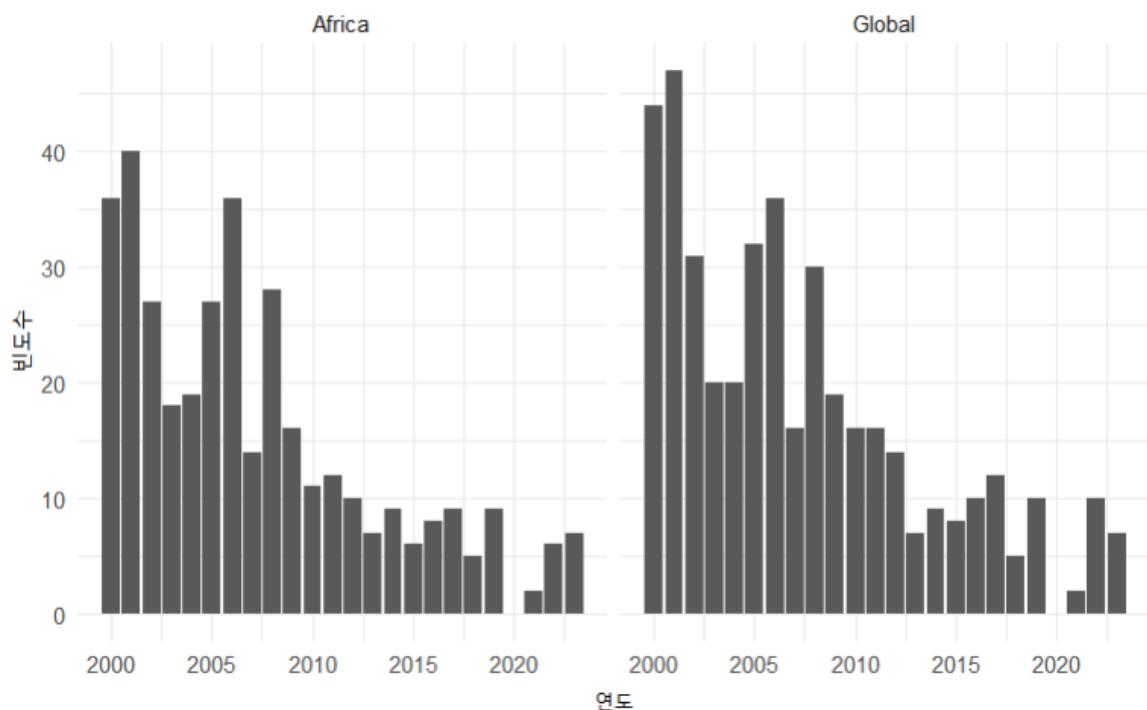
Heatmap을 통해서 일단 인구/면적이 큰 CHN, USA의 절대적인 재해빈도수가 높음을 한눈에 확인할 수 있다. 또한 CHN의 경우 Explosion(industrial)이 노란색으로 표시되어서, CHN의 산업재해가 높은 수치를 나타냄이 더욱 효과적으로 전달되었다.



▼ longitudinal approach

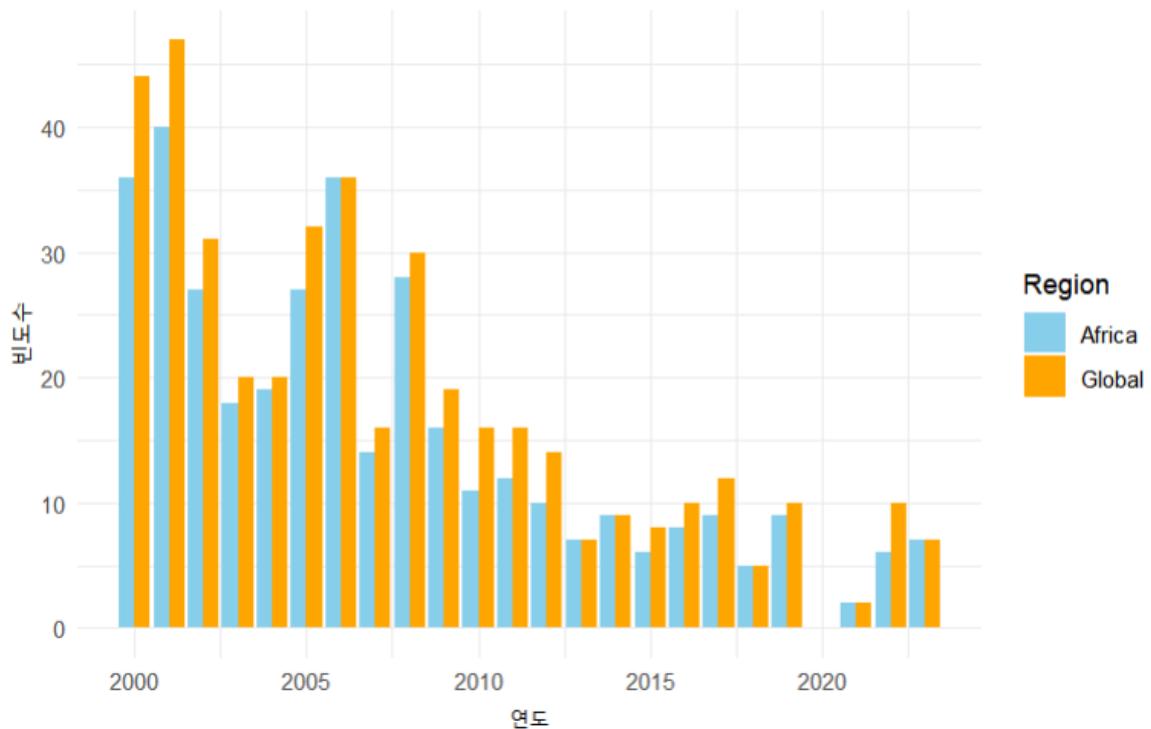
Feedback1: 데이터가 서로 다른 그래프로 그려져 비교가 어렵다

아프리카 및 전 세계 연도별 박테리아 질환 빈도수



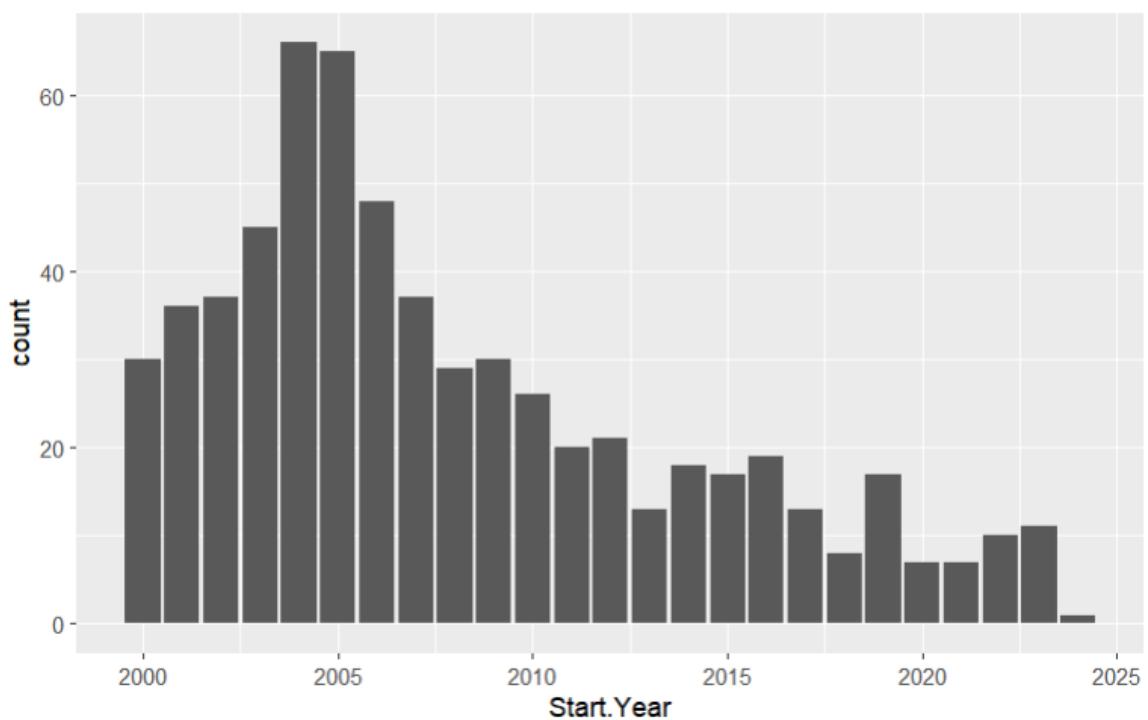
질환의 빈도수를 아프리카와 전 세계, 두가지로 나누어서 시각화를 하였는데, 정확한 비교를 한눈에 알기 어렵다.

아프리카 및 전 세계 연도별 바이러스성 질환 빈도수

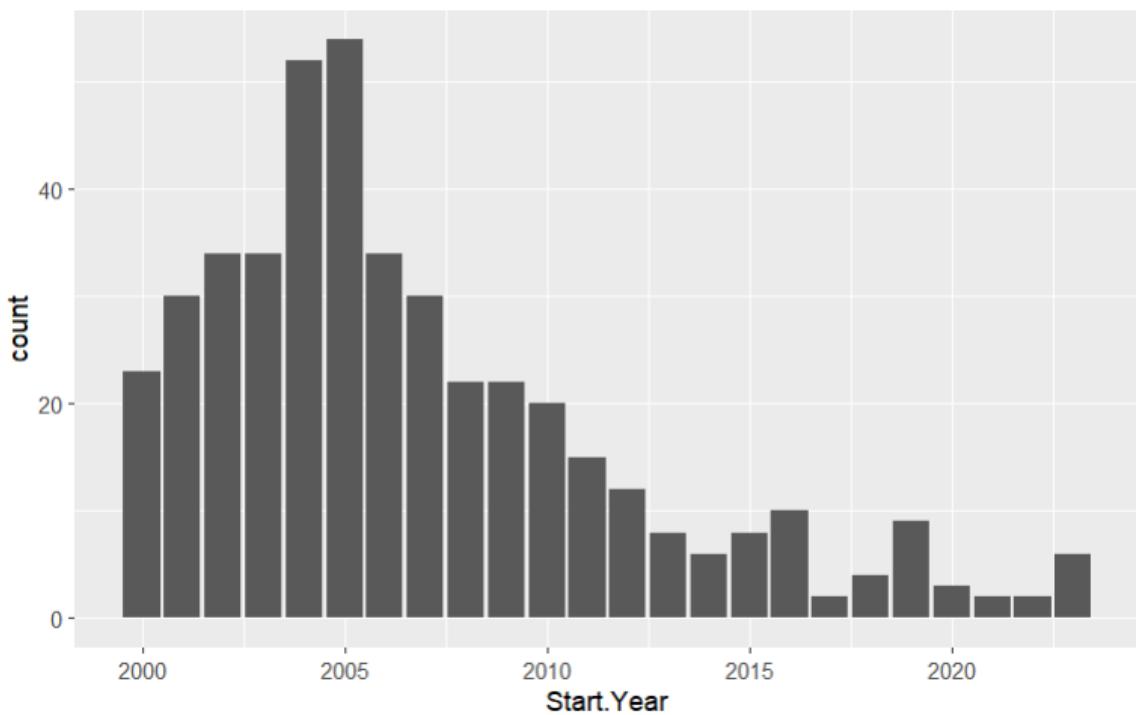


하나의 그래프 안에 서로 다른 색깔로 표시하니 비교가 한눈에 되었다.

아시아 연도별 산업재해 빈도수

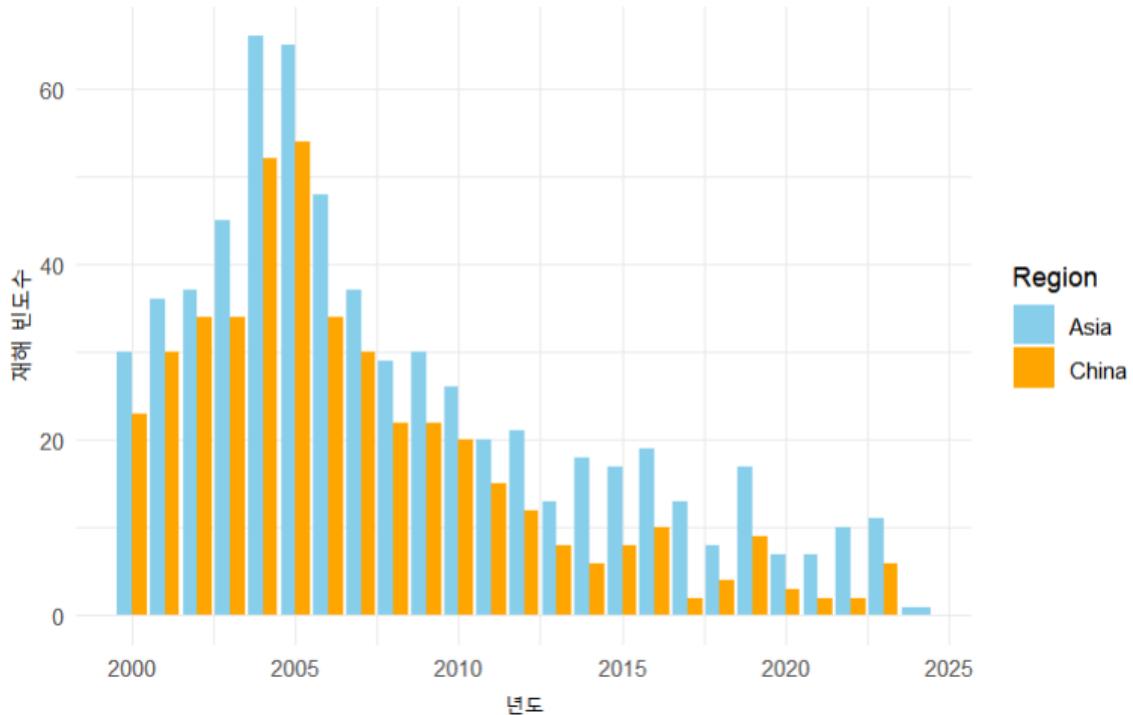


중국 연도별 산업재해 빈도수



아시아와 중국의 산업재해 빈도수를 확인할 때, 각각의 그래프로 표현하니 비슷한 추세를 가진다는 것은 알겠으나 정확한 비교분석이 어렵다

아시아 및 중국의 년도별 산업재해 빈도수

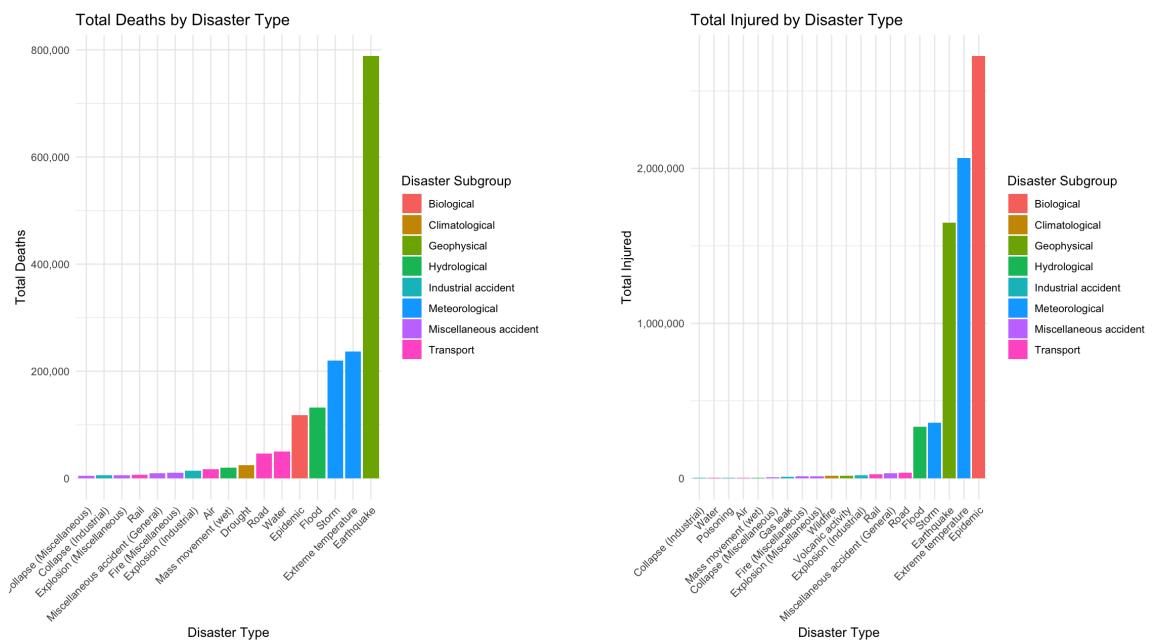


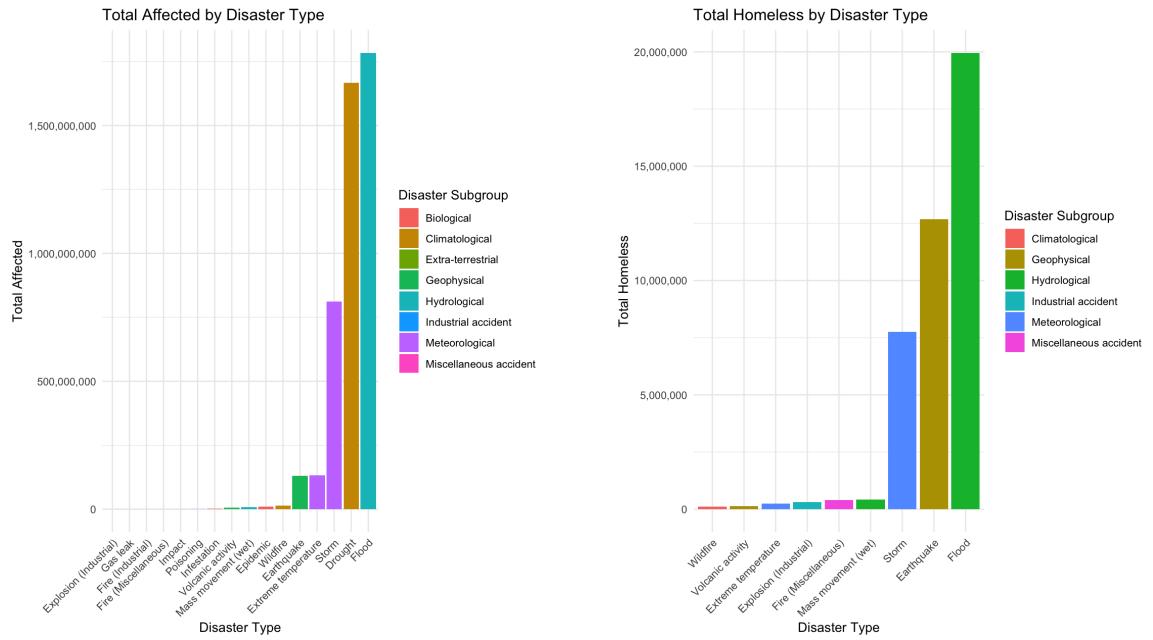
이를 하나의 그래프 안에서 아시아와 중국의 빈도수를 구분하여 시각화가 편하였다.

▼ Analysis of the Effect of Disaster

disaster로 인해서 어떤 재난이 죽음을 가장 많이 초래하였을지 분석하고자 하였다.

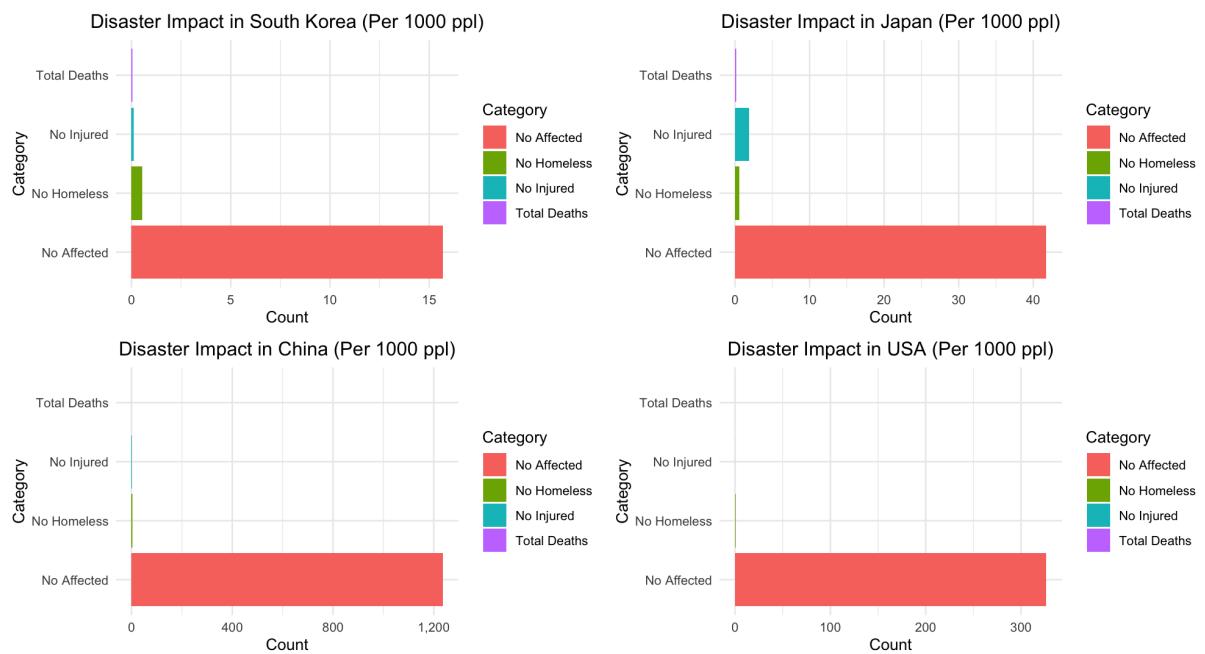
가설 : 지구온난화가 지속적으로 악화되는 상황이기에 자연재해가 가장 큰 원인일 것이고, 그중 지진으로 인한 사망률이 가장 높을 것이라 생각한다.



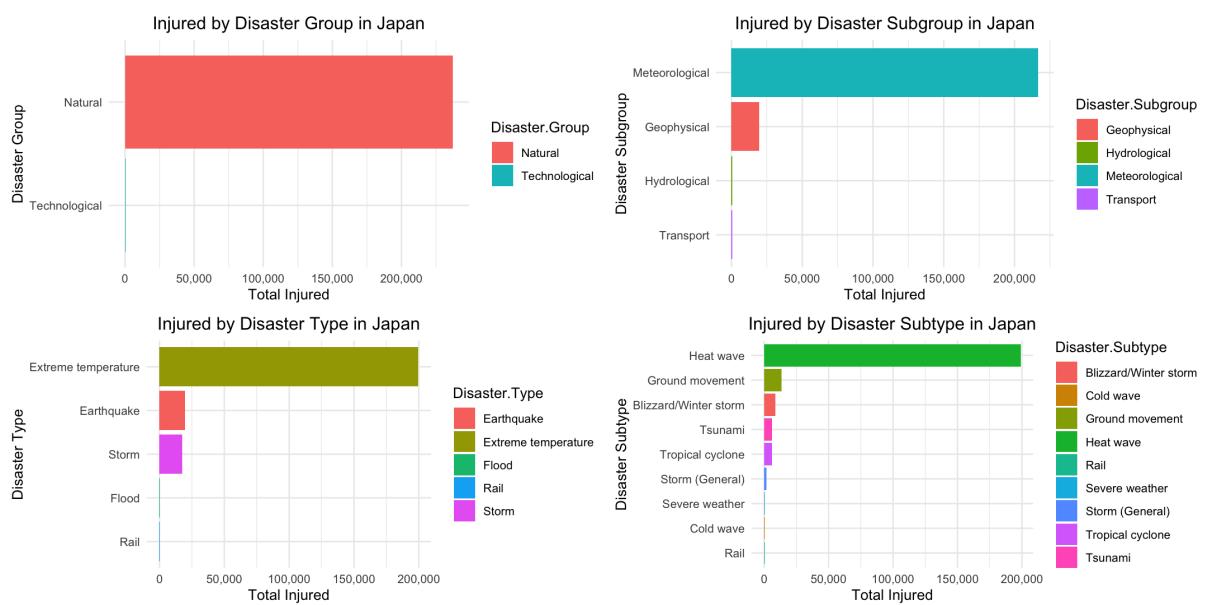


공통적으로 많이 발생한 것은 지진이었고 이상기온으로 인한 홍수, 지진, 가뭄, 폭풍 그리고 전염병으로 인한 피해로 집계되었다. 사전에 어느정도 막을 수는 있겠지만 자연재해이기 때문에 결과적인 피해는 막을 수 없었고 지구온난화가 점차 심화되고 있기에 점차 피해자는 많아질 것이다.

여러 국가들 중 한국, 일본, 중국, 미국에 대하여 분석해보았다. 국가마다 인구수가 상이하기에 인구수를 나누고 1000명을 기준으로 하여 1,000명당 어떤 결과를 가지는지 확인해보았다.



결과를 보았을 때, 사망자, 피해자, 부상자, 그리고 노숙자 중에서 네 국가 모두 피해자가 가장 많은 것으로 보인다.



특별히 일본의 경우, 부상자의 수가 두번째로 많았는데 이는 이상 기온으로 인한 폭염과 자연이 가장 주된 원인임을 확인하였다.