

대륙 및 국가별 covid-19 사망자, 확진자 분석

library 셋팅하기

분석에 활용할 library 는 아래와 같이 dplyr, ggplot2, RColorBrewer 을 선택해 R studio 에 부착해 준비합니다.

데이터 준비하기

kaggle 에서 가져온 세계 covid-19 확진자 data 가 들어있는 csv 파일을 활용해 데이터를 준비합니다.

```
covid_19 <- read.csv("country_wise_latest.csv")

head(covid_19, 10)
```

##	Country.Region	Confirmed	Deaths	Recovered	Active	New.cases	New.deaths
## 1	Afghanistan	36263	1269	25198	9796	106	
## 2	Albania	4880	144	2745	1991	117	
## 3	Algeria	27973	1163	18837	7973	616	
## 4	Andorra	907	52	803	52	10	
## 5	Angola	950	41	242	667	18	
## 6	Antigua and Barbuda	86	3	65	18	4	
## 7	Argentina	167416	3059	72575	91782	4890	
## 8	Armenia	37390	711	26665	10014	73	
## 9	Australia	15303	167	9311	5825	368	
## 10	Austria	20558	713	18246	1599	86	
##	New.recovered	Deaths...100.Cases	Recovered...100.Cases				
## 1	18	3.50	69.49				
## 2	63	2.95	56.25				
## 3	749	4.16	67.34				
## 4	0	5.73	88.53				

## 5	0	4.32	25.47
## 6	5	3.49	75.58
## 7	2057	1.83	43.35
## 8	187	1.90	71.32
## 9	137	1.09	60.84
## 10	37	3.47	88.75
##	Deaths...100.Recovered	Confirmed.last.week	X1.week.change
rease			X1.week...inc
## 1	5.04	35526	737
2.07			
## 2	5.25	4171	709
17.00			
## 3	6.17	23691	4282
18.07			
## 4	6.48	884	23
2.60			
## 5	16.94	749	201
26.84			
## 6	4.62	76	10
13.16			
## 7	4.21	130774	36642
28.02			
## 8	2.67	34981	2409
6.89			
## 9	1.79	12428	2875
23.13			
## 10	3.91	19743	815
4.13			
##	WHO.Region		
## 1	Eastern Mediterranean		
## 2	Europe		
## 3	Africa		
## 4	Europe		
## 5	Africa		
## 6	Americas		
## 7	Americas		
## 8	Europe		
## 9	Western Pacific		
## 10	Europe		

조건 없이 상위 10 개의 data 를 출력한 모습입니다. 이를 보면 각 국가별로 순서대로 검사자, 사망자, 격리해제, 신규 확진자, 신규 사망자, 신규 격리해제, 각 속성에 대한 백분율과 일주일 간 증가율, 변화율 %, 대륙을 알아 볼 수 있습니다.

세계 covid-19 사망자수 분석하기

현재 covid-19 로 인해 전 세계적으로 확진자 및 사망자가 급증하고 있습니다. 이에 data 를 활용하여 각 대륙별, 나라별 사망자 수를 분석하고자 합니다.

대륙 별 사망자 수 분석하기

```
covid_19 %>%
  mutate(tot = (Deaths + New.deaths)/2) %>%
  group_by(WHO.Region) %>%
  summarise(mean_tot = mean(tot)) %>%
  arrange(desc(mean_tot))

## # A tibble: 6 x 2
##   WHO.Region      mean_tot
##   <chr>          <dbl>
## 1 Americas      4947.
## 2 South-East Asia 2104.
## 3 Europe        1888.
## 4 Eastern Mediterranean 881.
## 5 Western Pacific 259.
## 6 Africa        131
```

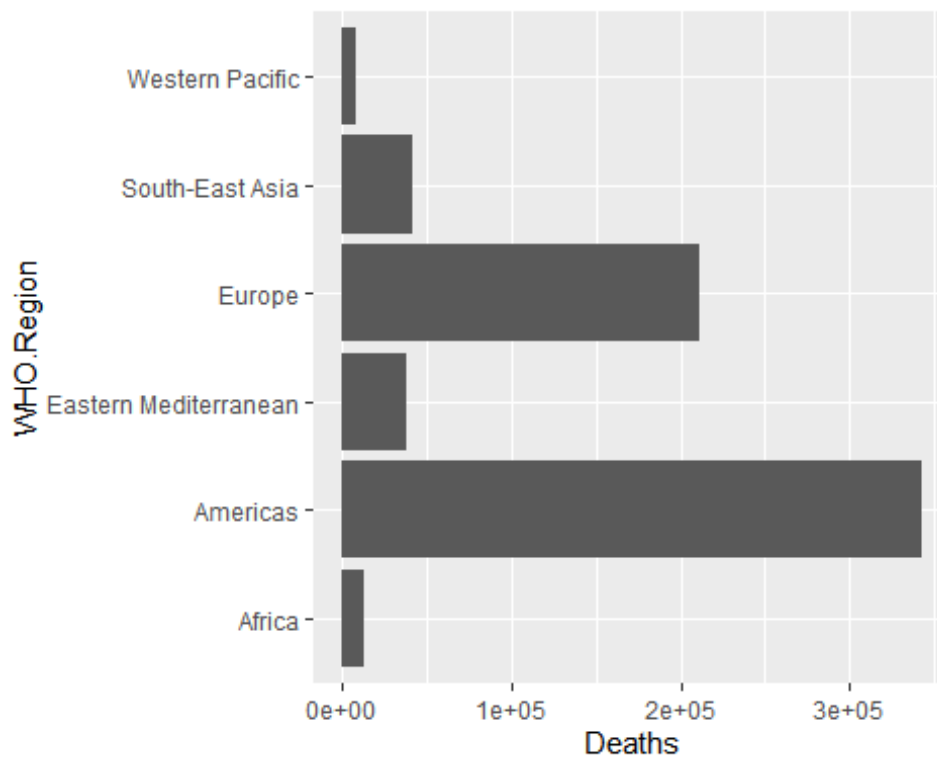
위는 대륙별 기존의 사망자와 신규 사망자의 평균을 나타낸 표입니다. 이를 보면 Americas 에서 가장 많은 사망자 수를 보였고 Africa 에서는 가장 적은 사망자 수를 보였습니다.

또한, America 에서는 2 위인 South-East Asia 보다 2 배 이상의 사망자 수를 보이는 것을 확인했습니다.

이는 해당 대륙의 인구의 밀도와 검사 횟수와 밀접한 관련이 있을 것으로 예상되지만 Ameicas 는 타 대륙보다 월등히 많은 수의 사망자가 발생한 것이라고 판단할 수 있습니다.

대륙별 사망자 수 그래프 제작하기

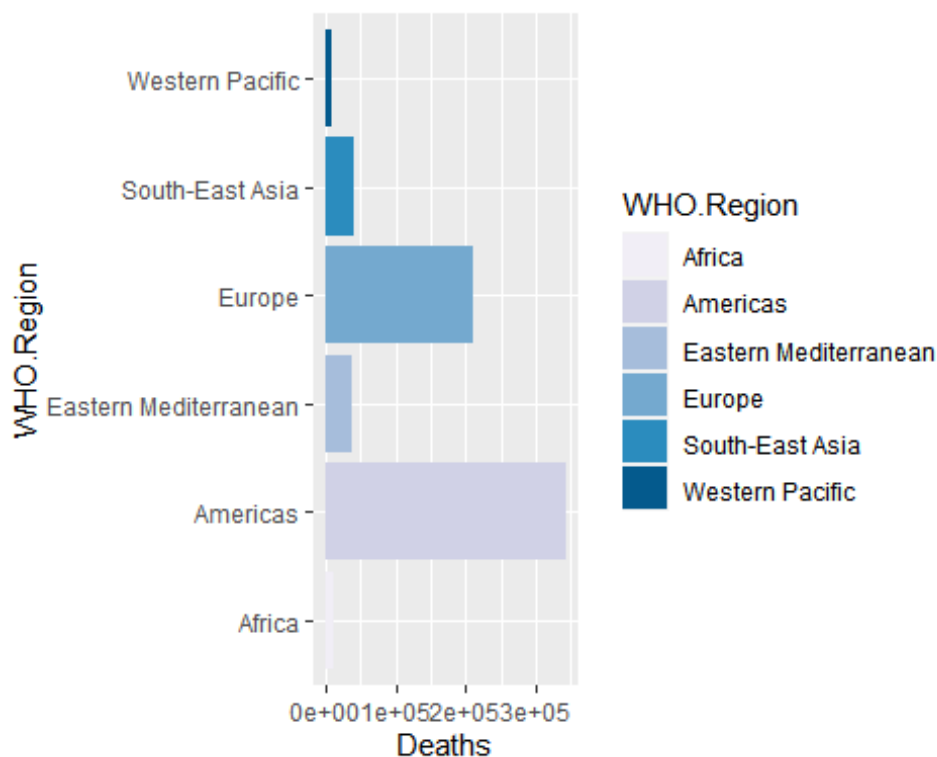
```
ggplot(data= covid_19, aes(x=WHO.Region, y=Deaths))+
  geom_col()+
  coord_flip()
```



이 그래프는 앞에서 나온 분석 자료를 시각화 한 그래프 입니다. 막대 그래프로 각 속성의 값을 한눈에 확인할 수 있도록 제작했습니다.

또한, 속성 간의 차이를 명확히 하고자 가로에 x 값을 두도록 **coord_flip()**을 활용해 표현하였습니다.

```
covid_19 %>%
  group_by(WHO.Region) %>%
  ggplot(aes(x=WHO.Region, y=Deaths))+
  geom_bar(stat = "identity", aes(fill=WHO.Region))+
  scale_fill_brewer(palette = 9) + coord_flip()
```



하지만 한가지

색상으로만 이루어져 있으면 구별하기가 어렵기 때문에 수업시간에 배웠던

`scale_fill_brewer()`를 활용해 그래프에 적용했습니다.

각 나라 별 사망자 수 분석하기

대륙별로 사망자의 추이가 어떻게 되는지를 확인했습니다. 이에 대륙이 아닌 각 나라별 사망자 수 분포를 확인하겠습니다.

covid_19 사망자가 10 만명 이상인 데이터 추출

```
covid_19[covid_19$Deaths > 100000, ]
```

```
##      Country.Region Confirmed Deaths Recovered Active New.cases New.deaths
## 174              US  4290259 148011  1325804 2816444    56336      1076
##      New.recovered Deaths...100.Cases Recovered...100.Cases
## 174          27941              3.45              30.9
##      Deaths...100.Recovered Confirmed.last.week X1.week.change
## 174              11.16              3834677      455582
##      X1.week...increase WHO.Region
## 174              11.88      Americas
```

사망자가 10 만명 이상인 국가의 총 data 를 추출했습니다. 이는 미국이고 사망자는 약 14 만 8 천명의 사망자 수를 기록함을 알 수 있습니다.

covid_19 사망자 상위 10 개국 데이터 추출

```
country_10 <- covid_19 %>%
  mutate(tot = (Deaths + New.deaths)) %>%
  arrange(desc(tot)) %>%
  head(10)
```

country_10

##	Country.Region	Confirmed	Deaths	Recovered	Active	New.cases	New.deaths
## 1	US	4290259	148011	1325804	2816444	56336	1076
## 2	Brazil	2442375	87618	1846641	508116	23284	614
## 3	United Kingdom	301708	45844	1437	254427	688	7
## 4	Mexico	395489	44022	303810	47657	4973	342
## 5	Italy	246286	35112	198593	12581	168	5
## 6	India	1480073	33408	951166	495499	44457	637
## 7	France	220352	30212	81212	108928	2551	17
## 8	Spain	272421	28432	150376	93613	0	0
## 9	Peru	389717	18418	272547	98752	13756	575
## 10	Iran	293606	15912	255144	22550	2434	212
##	New.recovered	Deaths...100.Cases	Recovered...100.Cases				
## 1	27941	3.45	30.90				
## 2	33728	3.59	75.61				
## 3	3	15.19	0.48				
## 4	8588	11.13	76.82				
## 5	147	14.26	80.64				
## 6	33598	2.26	64.26				
## 7	267	13.71	36.86				
## 8	0	10.44	55.20				
## 9	4697	4.73	69.93				
## 10	1931	5.42	86.90				
##	Deaths...100.Recovered	Confirmed.last.week	X1.week.change	X1.week...inc			
## 1	11.16	3834677	455582				
## 2	4.74	2118646	323729				
## 3	3190.26	296944	4764				
## 4	14.49	349396	46093				
## 5	17.68	244624	1662				
## 6	3.51	1155338	324735				
## 7	37.20	214023	6329				
## 8	18.91	264836	7585				

```
## 9          6.76          357681          32036
8.96
## 10         6.24          276202          17404
6.30
##          WHO.Region    tot
## 1          Americas 149087
## 2          Americas  88232
## 3           Europe  45851
## 4          Americas  44364
## 5           Europe  35117
## 6    South-East Asia  34045
## 7           Europe  30229
## 8           Europe  28432
## 9          Americas  18993
## 10 Eastern Mediterranean 16124
```

대륙 별이 아닌 총 국가 별로 가장 사망자 수가 높은 국가 10 개국을 뽑아 data 표로 제작했습니다.

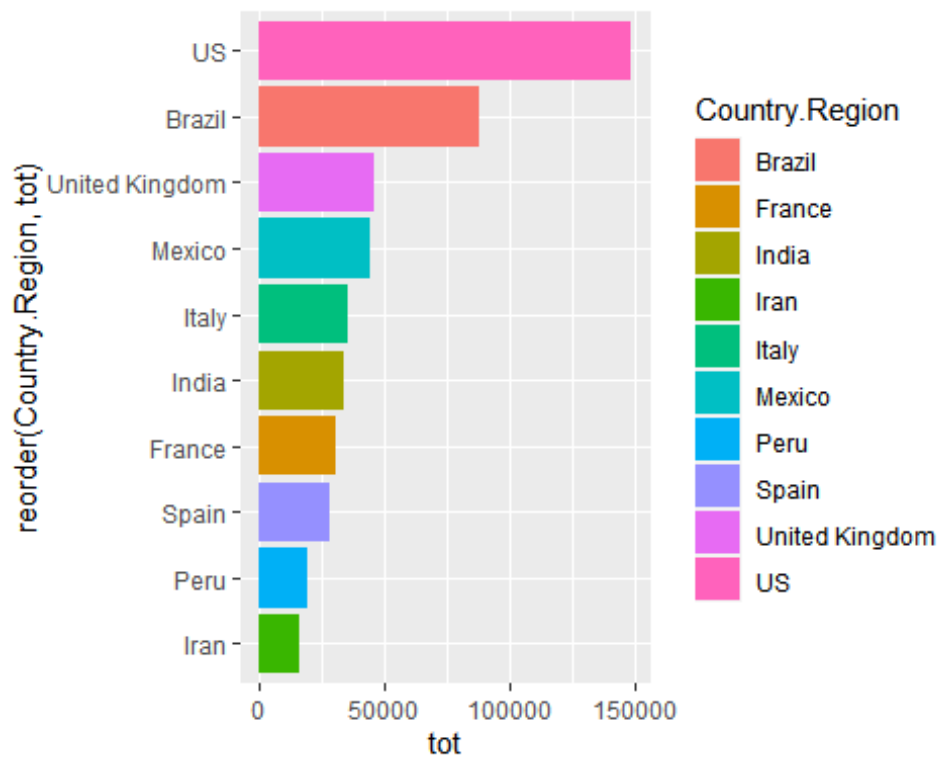
국가별 사망자 순, 미국, 브라질, 영국, 멕시코, 이탈리아, 인도, 프랑스, 스페인, 페루, 이란 순으로 나타납니다.

위에서 확인했듯이 가장 높은 국가는 US 이고 그 다음으로는 브라질, 영국이 위치해있습니다. 하지만 2 위는 1 위의 절반보다 조금 더 많은 사망자 수를 기록하고 3 위는 2 위의 절반의 사망자 수를 보이는 것을 알 수 있습니다. 이 표를 통해 사망자 수 상위 3 개국의 차이가 굉장히 큰 것을 확인할 수 있습니다.

각 나라별 사망자 수 그래프 제작하기

```
ggplot(country_10, aes(x=reorder(Country.Region, tot), y=tot, fill=Country.Region))+
  coord_cartesian(expand = F)+
  geom_bar(stat = "identity")+
  coord_flip()

## Coordinate system already present. Adding new coordinate system, which will
1 replace the existing one.
```



가장 직관적인 막대

그래프를 통해 사망자 수 상위 10 개국의 그래프를 제작했습니다. 이를 보면 미국의 사망자가와 브라질 의 사망자가 월등하고 그 이후 3 위인 영국 부터는 사망자 수가 4-5 위와 비슷한 것을 확인할 수 있습니다.

covid_19 사망자 하위 10 개국 데이터 추출

```
country_b10 <- covid_19 %>%
  group_by(Country.Region) %>%
  arrange(desc(Deaths)) %>%
  tail(10)
```

country_b10

A tibble: 10 x 15

Groups: Country.Region [10]

	Country.Region	Confirmed	Deaths	Recovered	Active	New.cases	New.deaths
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>
## 1	Holy See	12	0	12	0	0	0
## 2	Laos	20	0	19	1	0	0
## 3	Mongolia	289	0	222	67	1	0
## 4	Papua New Gui~	62	0	11	51	0	0
## 5	Saint Kitts a~	17	0	15	2	0	0
## 6	Saint Lucia	24	0	22	2	0	0
## 7	Saint Vincent~	52	0	39	13	0	0
## 8	Seychelles	114	0	39	75	0	0


```
## 9 Timor-Leste      24      0      0      24      0      0
## 10 Vietnam        431      0     365     66     11      0
## # ... with 8 more variables: New.recovered <int>, Deaths...100.Cases <dbl>,
## #   Recovered...100.Cases <dbl>, Deaths...100.Recovered <dbl>,
## #   Confirmed.last.week <int>, X1.week.change <int>, X1.week...increase <dbl>,
## #   WHO.Region <chr>
```

covid-19 사망자 하위 10 개국을 뽑아 data 표로 작성했습니다. 이는 사망자 수가 전부 0 으로 나오는 것을 알 수 있습니다.

이에 사망자 수가 0 인 국가 리스트를 알아보겠습니다.

```
covid_19[covid_19$Deaths == 0, ]
```

##	Country.Region	Confirmed	Deaths	Recovered	Active
## 20	Bhutan	99	0	86	13
## 31	Cambodia	226	0	147	79
## 50	Dominica	18	0	18	0
## 56	Eritrea	265	0	191	74
## 60	Fiji	27	0	18	9
## 69	Greenland	14	0	13	1
## 70	Grenada	23	0	23	0
## 76	Holy See	12	0	12	0
## 95	Laos	20	0	19	1
## 115	Mongolia	289	0	222	67
## 131	Papua New Guinea	62	0	11	51
## 141	Saint Kitts and Nevis	17	0	15	2
## 142	Saint Lucia	24	0	22	2
## 143	Saint Vincent and the Grenadines	52	0	39	13
## 149	Seychelles	114	0	39	75
## 169	Timor-Leste	24	0	0	24
## 182	Vietnam	431	0	365	66

##	New.cases	New.deaths	New.recovered	Deaths...100.Cases	Recovered...100.Cases
## 20	4	0	1	0	
## 31	1	0	4	0	
## 50	0	0	0	0	1
## 56	2	0	2	0	
## 60	0	0	0	0	
## 69	1	0	0	0	
## 70	0	0	0	0	1

00.00					
## 76	0	0	0	0	1
00.00					
## 95	0	0	0	0	
95.00					
## 115	1	0	4	0	
76.82					
## 131	0	0	0	0	
17.74					
## 141	0	0	0	0	
88.24					
## 142	0	0	0	0	
91.67					
## 143	0	0	0	0	
75.00					
## 149	0	0	0	0	
34.21					
## 169	0	0	0	0	
0.00					
## 182	11	0	0	0	
84.69					
##	Deaths...100.Recovered Confirmed.last.week X1.week.change				
## 20		0	90	9	
## 31		0	171	55	
## 50		0	18	0	
## 56		0	251	14	
## 60		0	27	0	
## 69		0	13	1	
## 70		0	23	0	
## 76		0	12	0	
## 95		0	19	1	
## 115		0	287	2	
## 131		0	19	43	
## 141		0	17	0	
## 142		0	23	1	
## 143		0	50	2	
## 149		0	108	6	
## 169		0	24	0	
## 182		0	384	47	
##	X1.week...increase WHO.Region				
## 20	10.00	South-East Asia			
## 31	32.16	Western Pacific			
## 50	0.00	Americas			
## 56	5.58	Africa			
## 60	0.00	Western Pacific			
## 69	7.69	Europe			
## 70	0.00	Americas			
## 76	0.00	Europe			
## 95	5.26	Western Pacific			

```
## 115          0.70 Western Pacific
## 131        226.32 Western Pacific
## 141          0.00      Americas
## 142          4.35      Americas
## 143          4.00      Americas
## 149          5.56      Africa
## 169          0.00 South-East Asia
## 182        12.24 Western Pacific
```

총 17 개국으로 South-East Asia 2 개, Western Pacific 6 개, Ameicas 5 개, Africa 2 개, Europe 2 개로 이루어져 있는 것을 확인할 수 있습니다.

세계 covid-19 확진자 수 분석하기

앞에서는 사망자 수를 분석했으니 이번에는 각 대륙별, 나라별 확진자 수를 분석하겠습니다. 먼저 대륙별 확진자 수를 분석해 그래프로 시각화하여 각 대륙의 추이를 명확하게 나타내겠습니다.

각 대륙 별 확진자 수 분석하기

각 대륙별 확진자 표 제작하기

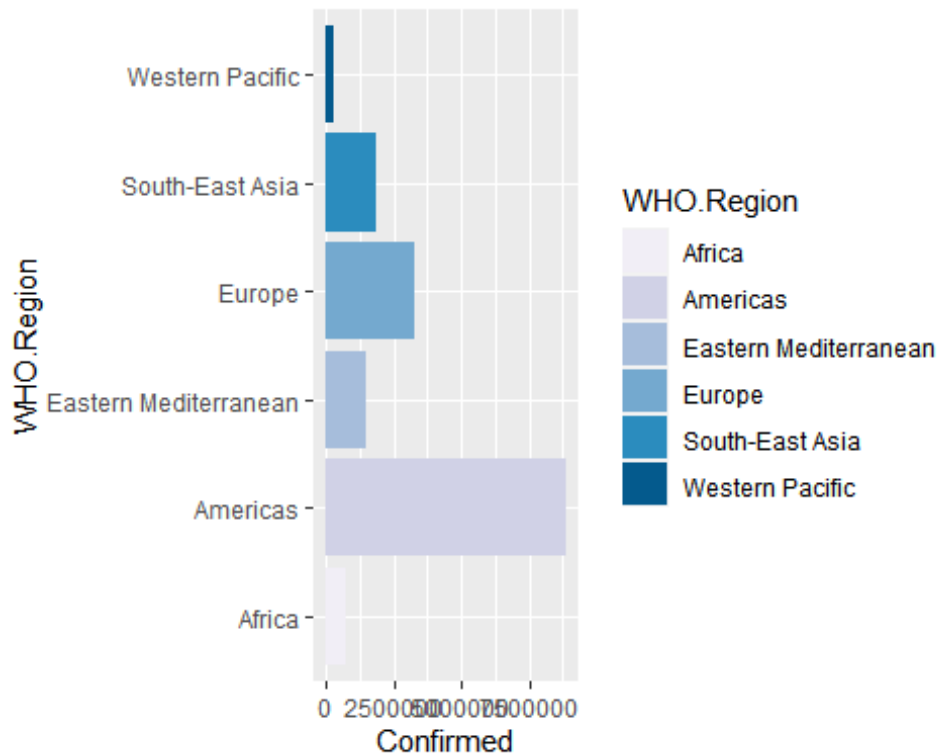
```
covid_19 %>%
  mutate(tot = (Confirmed + Active)/2) %>%
  group_by(WHO.Region) %>%
  summarise(mean_tot = mean(tot)) %>%
  arrange(desc(mean_tot))

## # A tibble: 6 x 2
##   WHO.Region      mean_tot
##   <chr>          <dbl>
## 1 Americas      183817.
## 2 South-East Asia 123616.
## 3 Eastern Mediterranean 39585.
## 4 Europe        39234.
## 5 Western Pacific 11557.
## 6 Africa        10349.
```

각 대륙 별 확진자 수 를 분석한 결과 Americas 와 South-East Asia 가 압도적으로 높은 확진자 수를 보이는 것을 확인할 수 있었습니다. 또한, Eastern Mediterranean, Europe 가 비슷한 확진자 수를 보이고 Western Pacific, Africa 가 비슷한 확진자 수를 보이는 것을 확인할 수 있습니다.

각 대륙별 확진자 수 그래프

```
covid_19 %>%
  group_by(WHO.Region) %>%
  ggplot(aes(x=WHO.Region, y=Confirmed))+
  geom_bar(stat = "identity", aes(fill=WHO.Region))+
  scale_fill_brewer(palette = 9) + coord_flip()
```



이 그래프는 앞에서 나온 분석 자료를 시각화 한 그래프 입니다. 막대 그래프로 각 속성의 값을 한눈에 확인할 수 있도록 제작했습니다.

또한, 속성 간의 차이를 명확히 하고자 가로에 x 값을 두도록 **coord_flip()**을 활용해 표현하였습니다. 그리고 한가지 색상으로 그래프를 작성하면 구별하기가 어렵기 때문에 수업시간에 배웠던 **scale_fill_brewer()**를 활용해 그래프에 적용했습니다.

나라별 확진자 수 분석하기

나라별 확진자 수 상위 10 개국 표 제작하기

```
dcountry_10 <- covid_19 %>%
  mutate(tot = (Confirmed + Active)) %>%
  arrange(desc(tot)) %>%
  head(10)
```

dcountry_10

##	Country.Region	Confirmed	Deaths	Recovered	Active	New.cases	New.deaths
## 1	US	4290259	148011	1325804	2816444	56336	1076
## 2	Brazil	2442375	87618	1846641	508116	23284	614
## 3	India	1480073	33408	951166	495499	44457	637
## 4	Russia	816680	13334	602249	201097	5607	85
## 5	South Africa	452529	7067	274925	170537	7096	298
## 6	United Kingdom	301708	45844	1437	254427	688	7
## 7	Peru	389717	18418	272547	98752	13756	575
## 8	Mexico	395489	44022	303810	47657	4973	342
## 9	Colombia	257101	8777	131161	117163	16306	508
## 10	Chile	347923	9187	319954	18782	2133	75
##	New.recovered	Deaths...100.Cases	Recovered...100.Cases				
## 1	27941	3.45	30.90				
## 2	33728	3.59	75.61				
## 3	33598	2.26	64.26				
## 4	3077	1.63	73.74				
## 5	9848	1.56	60.75				
## 6	3	15.19	0.48				
## 7	4697	4.73	69.93				
## 8	8588	11.13	76.82				
## 9	11494	3.41	51.02				
## 10	1859	2.64	91.96				
##	Deaths...100.Recovered	Confirmed.last.week	X1.week.change	X1.week...inc			
## 1	11.16	3834677	455582				
## 2	4.74	2118646	323729				
## 3	3.51	1155338	324735				
## 4	2.21	776212	40468				
## 5	2.57	373628	78901				
## 6	3190.26	296944	4764				
## 7	6.76	357681	32036				
## 8	14.49	349396	46093				
## 9	6.69	204005	53096				
## 10	2.87	333029	14894				
##	WHO.Region	tot					
## 1	Americas	7106703					

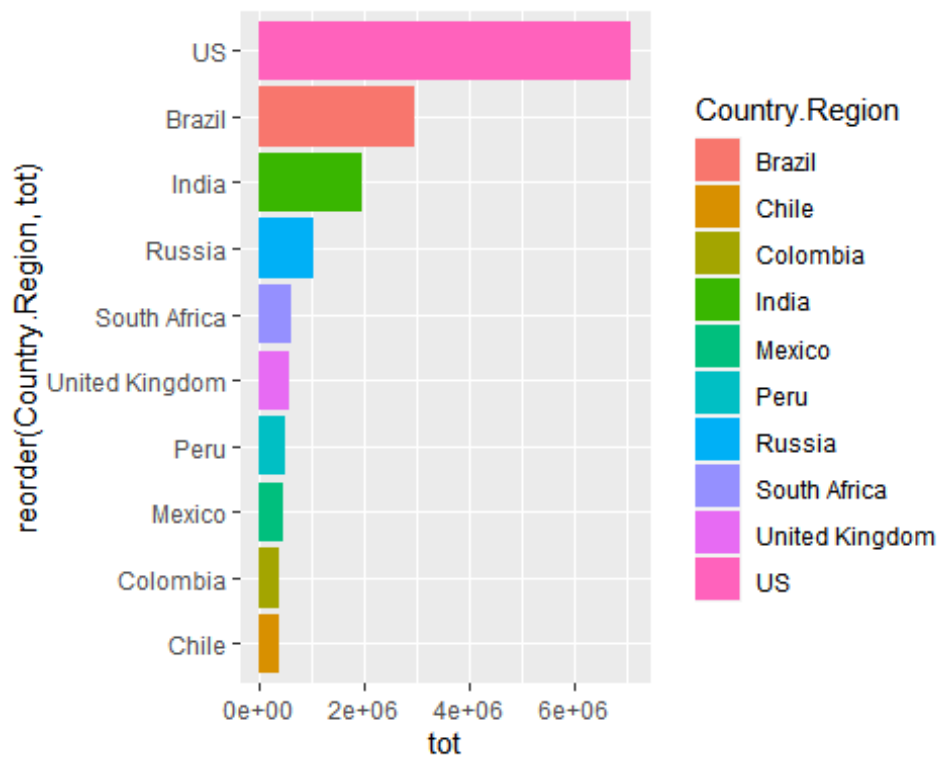
```
## 2      Americas 2950491
## 3 South-East Asia 1975572
## 4      Europe 1017777
## 5      Africa 623066
## 6      Europe 556135
## 7      Americas 488469
## 8      Americas 443146
## 9      Americas 374264
## 10     Americas 366705
```

확진자는 사망자와 같이 미국이 압도적으로 높은 수를 차지 하는 것을 볼 수 있습니다. 2 위는 브라질이 1 위의 절반의 확진자 수를 차지했고 3 위는 인도가 브라질의 절반의 확진자 수를 차지했습니다.

나라별 확진자 수 상위 10 개국 그래프 제작

```
ggplot(dcountry_10, aes(x=reorder(Country.Region, tot), y=tot, fill=Country.Region))+
  coord_cartesian(expand = F) +
  geom_bar(stat = "identity")+
  coord_flip()
```

Coordinate system already present. Adding new coordinate system, which will replace the existing one.



막대 그래프를

활용해 직관적으로 나라 별 확진자 수의 추이를 나타냈습니다. 확실하게 미국이 가장 높은 확진자 수를 보임을 한눈에 알 수 있습니다.

위의 그래프는 사망자수와 비슷하게 미국과 브라질이 1 위와 2 위를 차지했습니다. 사망자 3 위는 영국이었던 것과 달리 확진자 3 위는 인도가 이를 차지했습니다.

대륙 내 상위10 개국 확진자 분석

전체 대륙을 전부 조사했으나 보고서가 너무 길어지는 관계로 아시아의 분석 결과만 넣었습니다.

동남 아시아 내 확진자 수 상위10 개국 분석

```
South_East_Asia_Confirmed <- covid_19 %>%
  filter(WHO.Region == "South-East Asia") %>%
  group_by(Country.Region) %>%
  summarise(Confirmed) %>%
  arrange(desc(Confirmed)) %>%
  head(10)
```

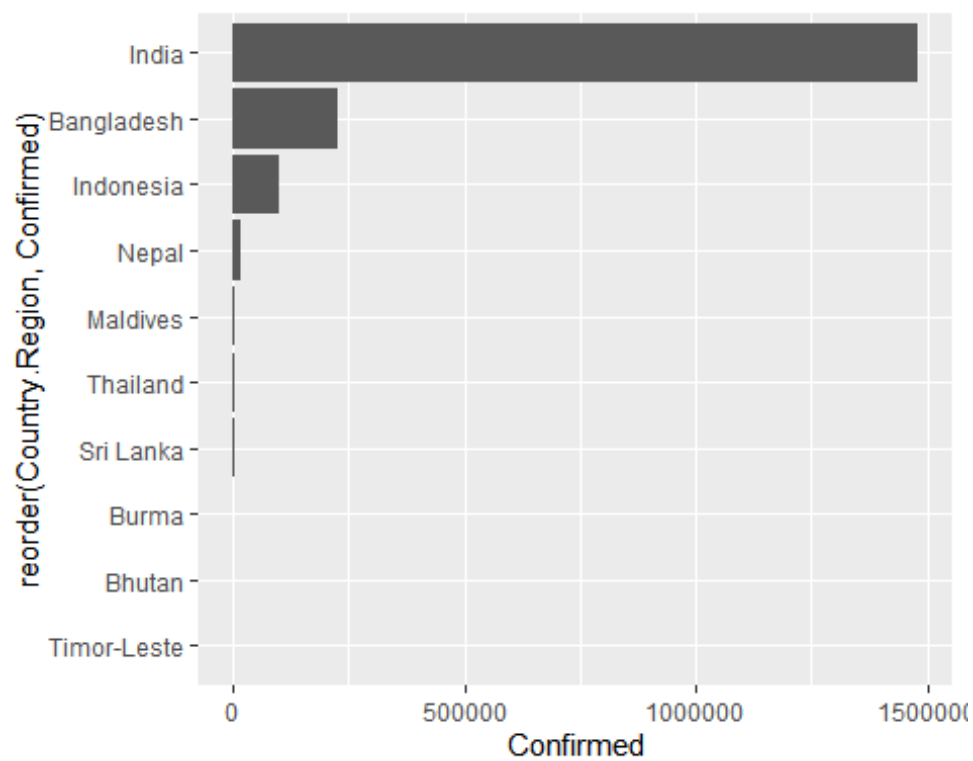
South_East_Asia_Confirmed

```
## # A tibble: 10 x 2
##   Country.Region Confirmed
##   <chr>          <int>
## 1 India          1480073
## 2 Bangladesh     226225
## 3 Indonesia      100303
## 4 Nepal          18752
## 5 Maldives        3369
## 6 Thailand        3297
## 7 Sri Lanka       2805
## 8 Burma           350
## 9 Bhutan           99
## 10 Timor-Leste     24
```

동남 아시아 내에서는 인도가 가장 높은 추이를 보임을 알 수 있습니다. 2 위인 방글라데시의 약 7 배에 달하는 확진자 수를 보이고 있습니다. 또한, 방글라데시는 약 20 만의 확진자 수를 보이고 이의 절반의 확진자 수인 10 만의 확진자 수를 가진 인도네시아가 3 위를 차지했습니다.

동남 아시아 내 확진자 수 상위 10 개국 그래프 제작

```
ggplot(data=South_East_Asia_Confirmed, aes(x=reorder(Country.Region, Confirmed), y=Confirmed))+
  geom_col()+
  coord_flip()
```

이를 그래프로 표현하면 나머지 1 위에서 4 위를 제외한 나머지는 거의 그래프에 표시되지 않을 만큼 작은 추이를 가지고 있습니다.