출석수업 과제물(평가결과물) 표지(온라인제출용)

**교과목명 : 데이터시각화**

**학 번 : 202135-368864**

**성 명 : 홍 원 표**

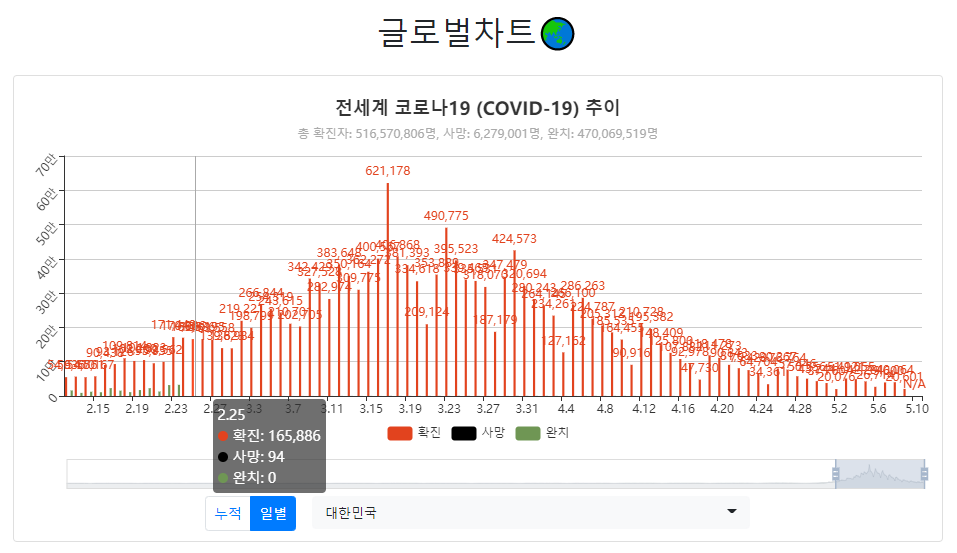
**강 의 실 : 경기(성남) 지역대학 호**

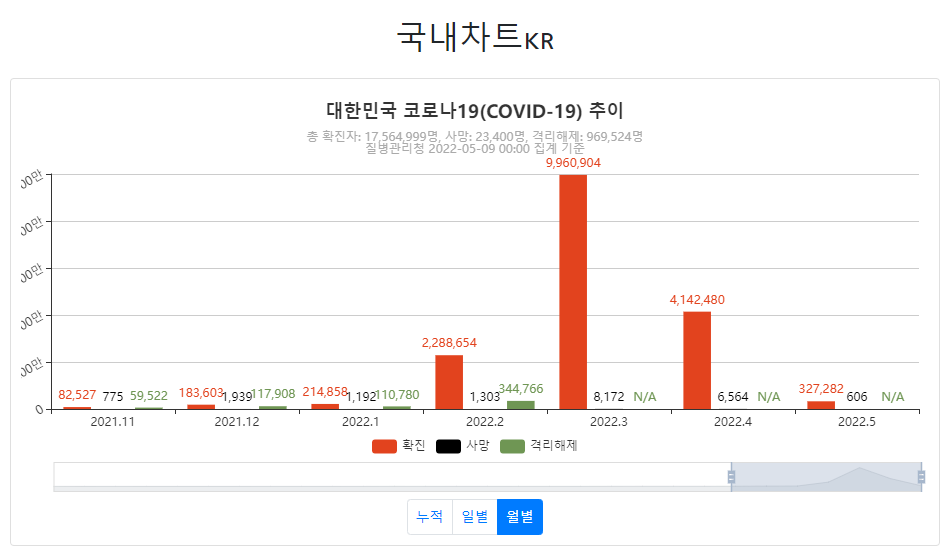
**연 락 처 : 010-5343-4341**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

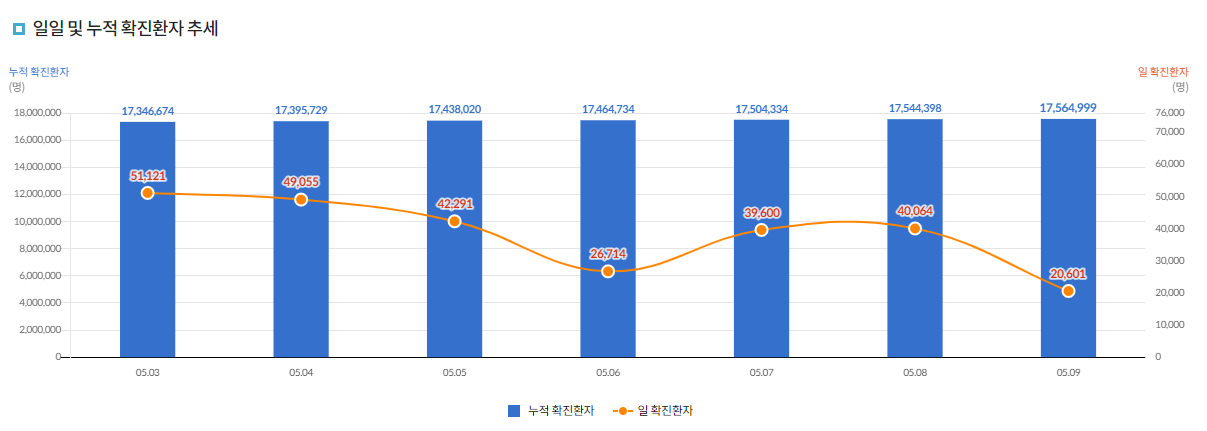
- 이하 과제 작성

1. COVID-19 관련 데이터 시각화의 국내외 사례 3개를 찾고 비교하여 정리하시오. (반드시 2022년 데이터가 포 함된 사례를 사용할 것. 이미지를 캡처하여 한글이나 워드 파일에 첨부할 것. 이미지를 별도의 파일로 제출하 지 말 것) (6점)

coronaboard.kr 사이트의 글로벌 차트에서 대한민국의 일별 코로나 추이 막대 그래프입니다. 해당 그래프는 누적과 일별 전세계코로나19 추이를 선택해서 볼 수 있으며, 전세계 또는 원하는 국가의 그래프를 볼 수 있습니다. 그리고 오른쪽 하단에 기간을 조정하여 볼 수 있는 기능이 있습니다. 한국에서는 3월 24일부터는 완치자에 대한 수치가 모두 0인것으로 보아 완치자에 대한 집계가 이뤄지지 않고 있다는 것을 알 수 있습니다. 사망자의 경우 확진자 대비 작은 수치라 그래프를 크게 해서 볼 수는 있으나 일별로 비교하기는 어렵습니다. 그리고 일일 확진자의 수치표시가 일정범위 이상 되면 겹치게 되어 사실상 확인하기 어려운 경우가 많게 되어 표시를 하지 않는 것이 더 좋을 것 같다는 생각이 듭니다.



coronaboard.kr 사이트의 국내 차트입니다. 유저가 누적/일별/월별 차트를 선택하여 볼 수 있고 기간별로 선택하여 그래프를 볼 수 있습니다. 국내는 2월24일 완치 판정을 하지 않아 집계가 이뤄지지 않았습니다. 따라서 2월까지의 격리해제 데이터는 있으나 3월부터는 데이터가 없음(N/A)으로 표기하였습니다. 확진자 대비 사망자의 비율이 적어 막대그래프 상으로는 보이지 않고 수치표시한 것으로 확인이 가능합니다. 2월에 감염자가 증가하기 시작하여 3월에 감염자 발생이 최고조를 이루고 4월에 감소하여 5월에는 안정세가 되가는 상황임을 알 수 있습니다.



ncov.mohw.go.kr 사이트의 국내발생 현황의 일일 누적 확진 환자 추세 그래프입니다. 데이터는 7일로 고정되어 있고 왼쪽편의 Y축은 누적확진자 단위를 나타내고 오른쪽에는 일일 확진자 단위를 나타내고 있습니다. 표는 이해하기 쉬우나 사용자가 원하는 기간을 설정하는 기능이 없기 때문에 원하는 정보를 얻기에는 충분하지 않습니다.

1. 한스 로즈링의 TED 강의(아래의 URL 이용)를 보고 데이터 시각화의 역할 등 느낀 점을 1페이지 이내로 정리 하시오.(6점) <https://www.ted.com/talks/hans_rosling_let_my_dataset_change_your_mindset> (오른쪽 아래 메뉴에서 한글 자막 설정 가능)

[스웨덴 카롤린스카 연구소](http://ki.se/ki/jsp/polopoly.jsp?d=130&l=en) (Karolinska Institute)의 세계 보건 교수인 한스 로즈링(Hans Rosling) 교수는 대부분의 사람들이 편견에 의해 사고방식(mindset)이 자료들(dataset)과 일치하지 않는다는 것을 이해하기 쉽게 표현해 주셨습니다. 개발도상국과 제3세계의 국가들은 서구와 동떨어진 세계가 아니고 건강과 번영을 서구와 동일한 궤도에 맞추기 위해 서구보다 2배빠른 속도로 움직이는 국가들도 있다고 합니다.

전세계의 건강과 빈곤 추세를 추적하기위한 데이터 분석 방법은 매우 지루한 작업일 것으로 여겨지는데 교수님의 프레젠테이션을 보고 있으면 지루해지지 않고 다른 항목의 데이터로 추세를 추적해보면 어떨까 하는 생각에 흥미로웠습니다. 대부분의 사람들은 관점에 편견이 있는 것을 알게 되었습니다.

아동사망률을 보면 21세기 달성목표(MDG)가 단순히 경제나 사회적으로 경제나 사회발전에 영향이 높을 거라는 막연한 생각에서 수질개선과 구호식량 지원, 말라리아퇴치와 같은 국제사회의 지원이 중요한 역할을 하게 된다는 것도 알게 되었습니다.

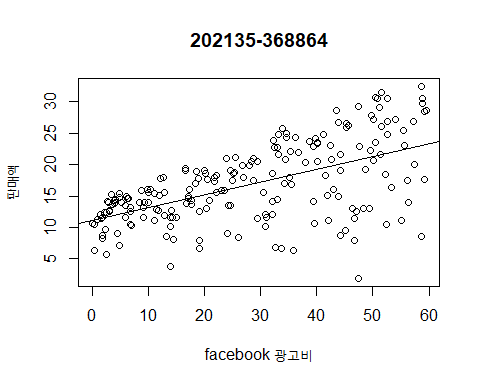
HIV의 경우 남아프리카 지역을 보면 비교적 고소득인 경우임에도 불구하고 성인인구의 약 20%정도가 감염될 정도로 높습니다. 탄자니아의 경우 상위 20% 국민들이 가장 가난한 국민들보다 HIV 감염율이 높다고 합니다. 즉 HIV의 감염 요인이 극심한 가난이 아니라 일부 인구집단 내에서의 반복적인 성관계가 원인일 것이라고 합니다.

아프리카 지역의 콘조(Konzo)라는 질병은 전세계에 걸쳐 장애를 유발하고 있는 에볼라 바이러스보다 심각하다고 합니다. 콘조는 모잠비크와 탄자니아 북부 지방에서 유행하는 질병인데 기근 상태에서 카사바 뿌리의 독성을 충분히 제거하지 못해서 생기는 병이라고 합니다. 극심한 가난에 의해 발생하는 병이라서 부유한 사람은 절대 걸릴 일이 없다고 합니다.

세계는 점점 수렴하고 있지만 아직도 하위 십억 인구는 언제나 그랬듯이 가난에 시달리고 있습니다. 이것이 지속되지 않도록 많은 초강대국들의 힘을 합쳐 자국만을 위해서가 아닌 전세계를 위해 필요한 역할을 수행해야 합니다. 선진국과 개발도상국이라는 사고방식을 뛰어 넘는 새로운 태도가 필요하다는 것에 공감합니다.

1. R의 datarium 패키지에 내장된 marketing 데이터셋은 광고 미디어에 사용한 비용과 판매액의 데이터이다. facebook 컬럼은 facebook 광고비로 사용한 금액이고, sales 컬럼은 판매액이다. facebook을 x축, sales를 y 축으로 하는 산점도를 그리시오. facebook을 독립변수(설명변수), sales를 종속변수(반응변수, 결과변수)로 하 는 회귀직선을 산점도 위에 그리시오. 산점도의 제목으로 본인의 학번을 출력하시오. (9점) (힌트: datarium 패 키지를 설치, 로드한 후 콘솔에 dat<-marketing을 입력하면 marketing 데이터셋이 dat에 저장된다)

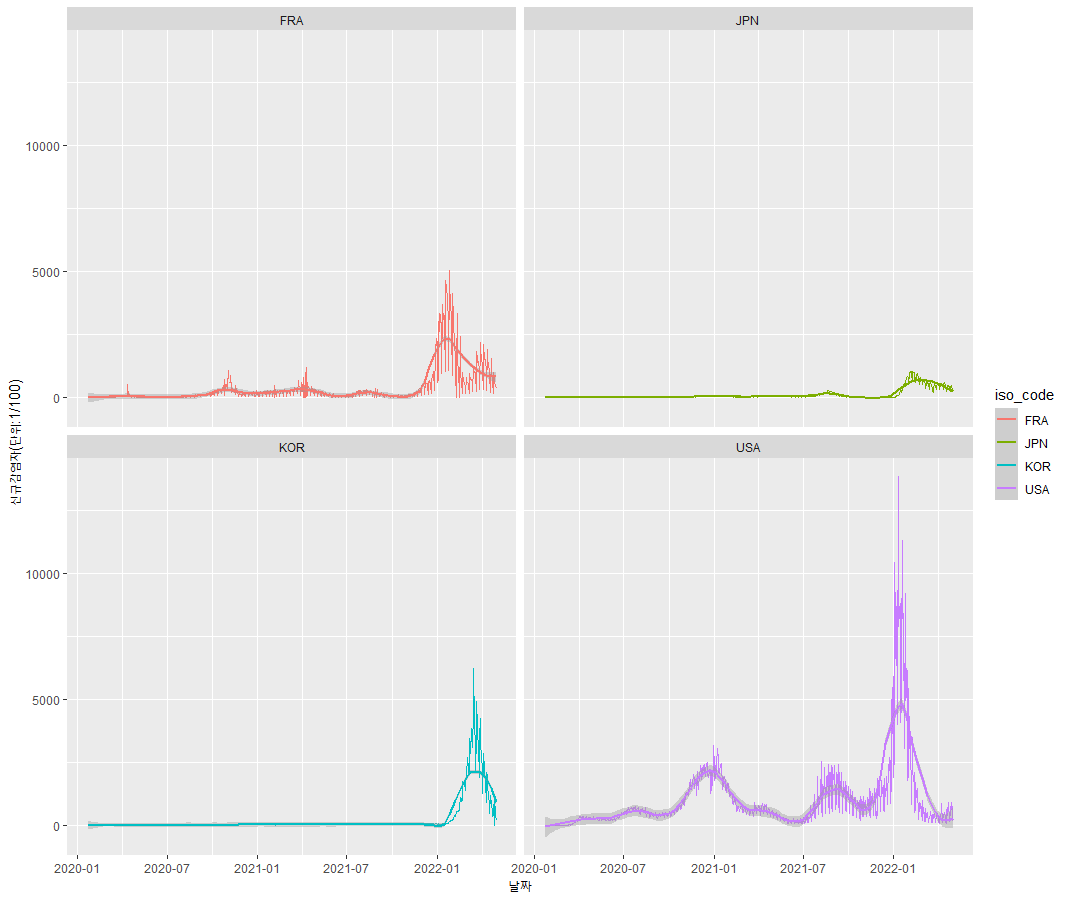
# install.packages("datarium")  
# datarium 라이브러리 로드  
library(datarium)  
# marketing 데이터를 이용하여 dat 객체 생성  
dat<-marketing  
# facebook과 sales 컬럼을 이용하여 산점도를 그린다.  
plot(dat$facebook, dat$sales, xlab="facebook 광고비", ylab="판매액",main="202135-368864")  
# Linear Model을 obj에 저장한다.  
obj<-lm(sales~facebook, data=dat)  
# 회귀직선을 그린다.  
abline(obj)



1. 한국, 미국, 프랑스, 일본의 COVID-19 신규 확진자 수의 시간에 따른 추이를 데이터 시각화로 비교하고 향후 추이에 대해 의견 기술하라(데이터는 과제 작성일까지 올라와 있는 것을 이용하면 되며 학생별로 동일할 필요 는 없음).(9점) *데이터 소스:* [*https://ourworldindata.org/covid-deaths*](https://ourworldindata.org/covid-deaths) *에서 “Our work belongs to everyone“이라 쓰여있 는 네모 안의 .csv(아래 그림에서 빨간 네모)를 클릭하여 데이터를 다운로드 받아서 사용할 것.* 변수이름: iso\_code가 각 국가를 나타낸다. (한국: KOR, 미국:USA, 프랑스: FRA, 일본: JPN) new\_cases가 신규 확진자 수를 나타낸다.

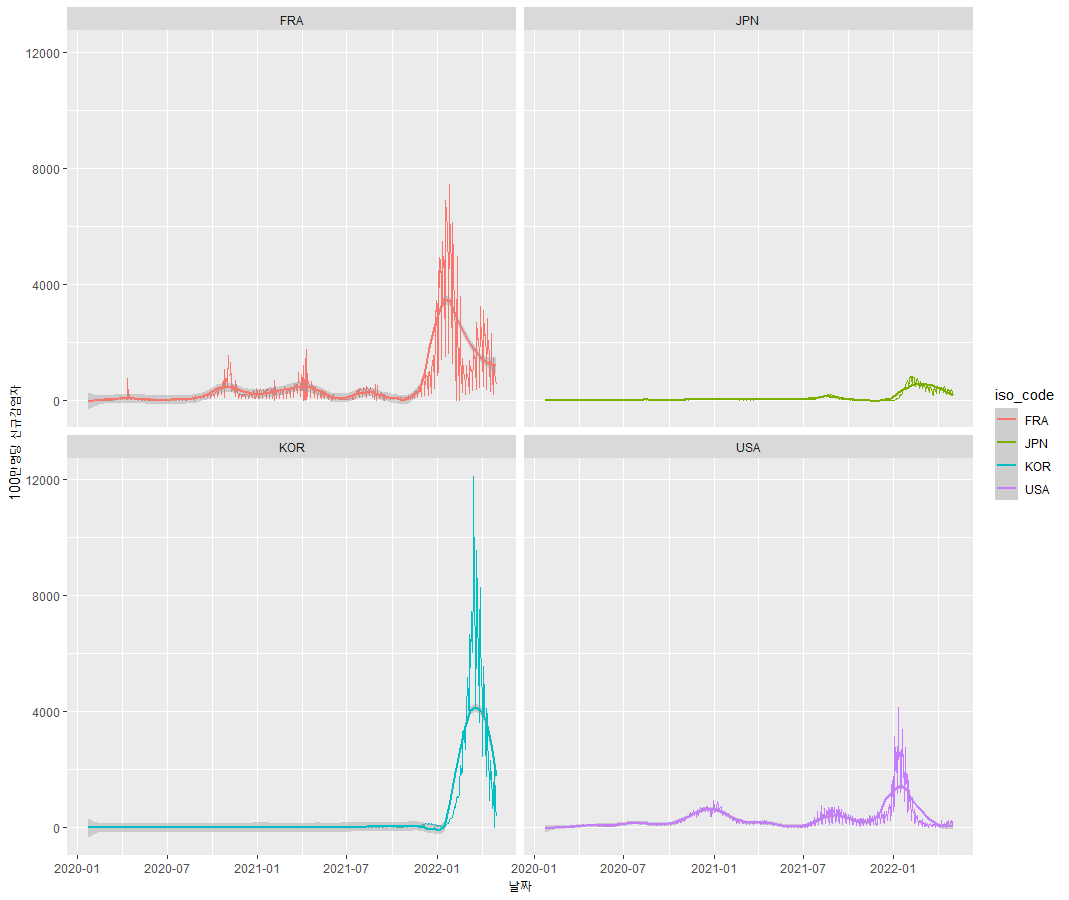
library(ggplot2)  
library(dplyr)

dat0 = read.csv("./data/owid-covid-data.csv", header = T)  
dat1 = dat0 %>% filter(iso\_code == "KOR" |  
 iso\_code == "JPN" |  
 iso\_code == "USA" |  
 iso\_code == "FRA") %>%  
 mutate(date = as.Date(date))  
  
ggplot(dat=dat1, aes(x=date, y=new\_cases/100, color=iso\_code)) +  
 labs(x="날짜", y="신규감염자(단위:1/100)") +  
 geom\_line() + geom\_smooth(span=0.2) + facet\_wrap(~iso\_code)



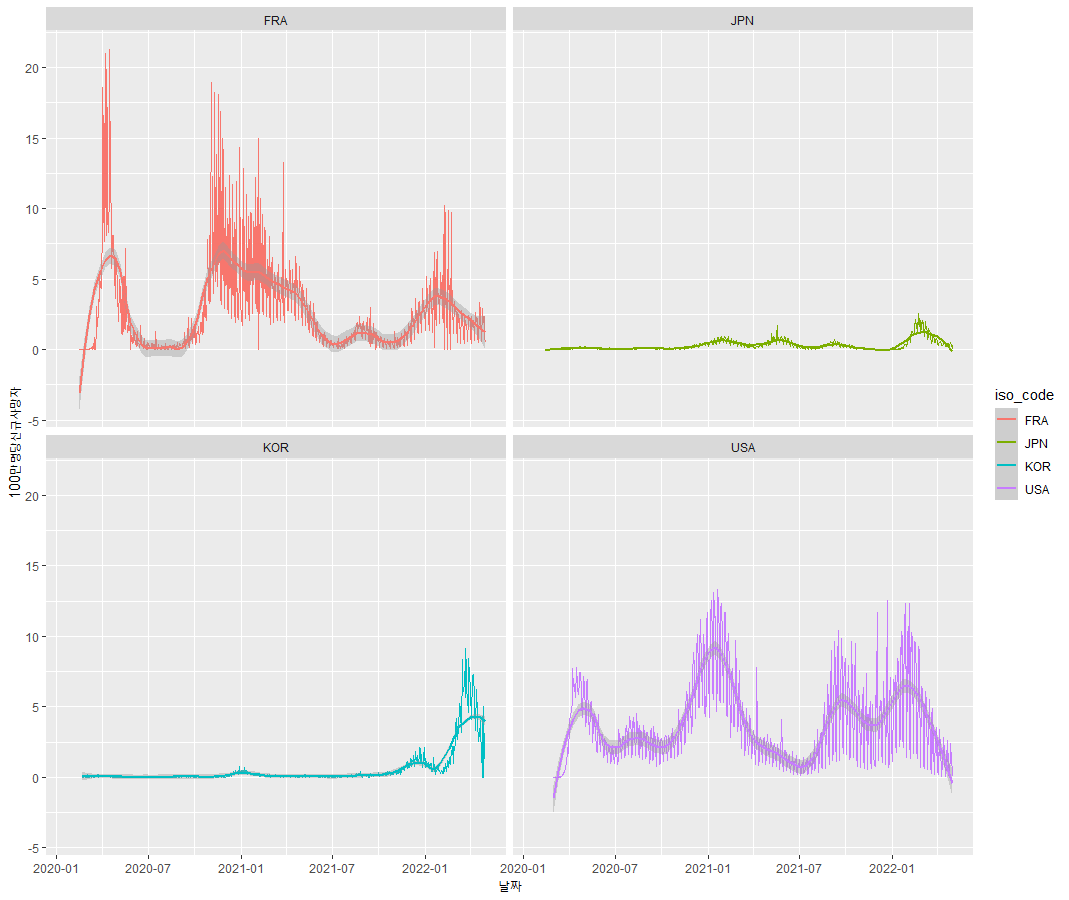
신규감염자 그래프를 보면 오미크론의 전파력은 매우 심각하다고 볼 수 있습니다. 2022년 1월 미국을 시작으로 프랑스, 일본, 대한민국 순으로 정점을 이루고 있습니다. 각 국가별 인구가 다르기 때문에 신규감염자 그래프로 보면 미국이 제일 감염이 심각했다고 볼 수 있습니다.

ggplot(dat=dat1, aes(x=date, y=new\_cases\_per\_million, color=iso\_code)) +  
 labs(x="날짜", y="100만명당신규감염자") +  
 geom\_line() + geom\_smooth(span=0.2) + facet\_wrap(~iso\_code)



하지만 100만명당 신규감염자 그래프를 보면 대한민국이 제일 감염이 심했고 프랑스, 미국, 일본 순으로 볼 수 있습니다. 이렇게 차이가 나는 것은 각 국가별로 의료체계와 방역정책, 국민들의 개인방역 등 차이점이 있기 때문이라고 생각합니다.

ggplot(dat=dat1, aes(x=date, y=new\_deaths\_per\_million, color=iso\_code)) +  
 labs(x="날짜", y="100만명당신규사망자") +  
 geom\_line() + geom\_smooth(span=0.2) + facet\_wrap(~iso\_code)



100만명당 COVID-19에 의한 신규사망자 그래프를 보면 프랑스와 미국은 신규감염자 그래프와 다른 그래프형태고 대한민국과 일본은 신규감염자 그래프와 비슷합니다. 프랑스와 미국은 오미크론 이전 변이에 의해 사망한 사람들이 많았다는 것을 알 수 있습니다.

오미크론 신규감염자는 대한민국, 미국, 프랑스, 일본 모두 정점이 지나 가파르게 감소하였지만 향후에는 좀 더 완만하게 감소할 것으로 예상됩니다.

-- 데이터시각화 출석수업 과제 끝 --