パンデミックを個人的に記録し一般向けに情報発信するために書いたRコード

神戸大学大学院保健学研究科
中澤　港

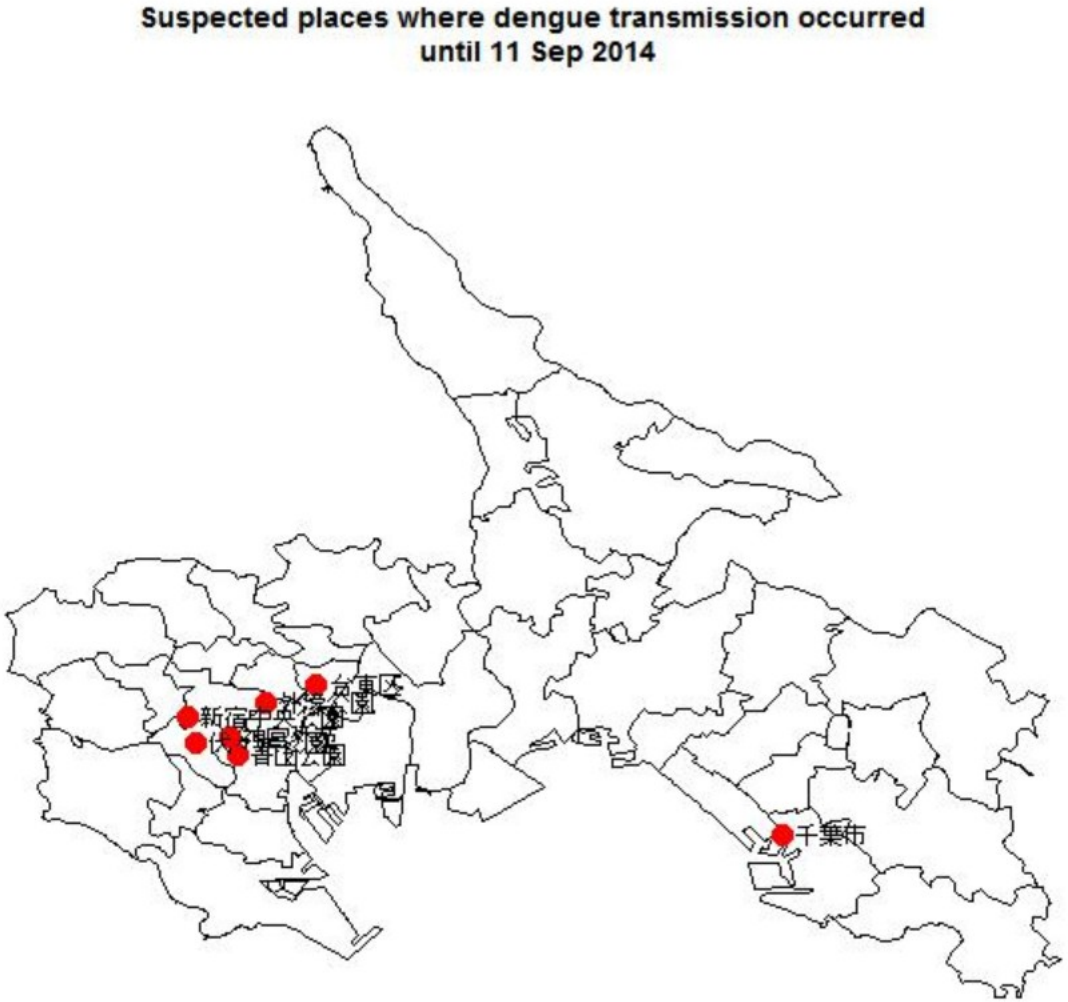
自分がやってきたこと

- 教育補助や研究成果公開、著書のサポートなどの目的で個人ウェブサイトを25年以上(3回移転したが)運営している
- 人類生態学、国際保健学、人口学の専門家として、1997年5月から記名でウェブログを書いて公開している
- 2009年から2011年までのパンデミックインフルエンザ、2014年のデング熱、2020年からのCOVID-19については、ウェブログから抜粋し加筆修正して別ページにまとめてきた
- わかりやすく情報提示するため、Rコードを書いて公開してきた

2014年のデング熱

2014年9月11日（当日の鵜記）

■岩田健太郎先生のtweetを拝見し，簡単な地図へのプロットを描いてみた（Rコード。ただし，#コメントで書いたようにESRI JapanからshpをダウンロードしてepiMap（EpiInfoに入っているので無料で使えるけれども，ESRIからライセンスされているソフトで，とても操作がわかりやすい）などで加工しないと，このコードは実行できない。あと，地名から緯度経度情報への変換は，Geocodingを使って得た数値をハードコードしているが，GeocodingにはXMLを返すAPIがあるので，それを使うと情報更新が楽になると思う）。



https://minato.sip21c.org/denguehs.R

```
# plotting areas suspected as dengue virus transmission occurred
# 11 September 2014
# coded by Minato Nakazawa <minato-nakazawa@umin.net>
library(maptools)
tokyo23chiba <- readShapeSpatial("./tokyo23chibasub.shp")
# shape file is created using EpiMap (in EpiInfo)
# データソース:
# 国土数値情報「行政区域(面)データ　国土交通省」
# ESRIジャパン「全国市区町村界データ」
# (c) Esri Japan
# 緯度経度情報はGeocodingより入手
# http://www.geocoding.jp
denguehs <- data.frame(
 pn = c("代々木公園", "新宿中央公園", "神宮外苑", "外濠公園", "千葉市", "青山公園", "台東区"),
 lat = c(35.671736, 35.689763, 35.675706, 35.699271, 35.607267, 35.663637, 35.712607),
 lon = c(139.694945, 139.689649, 139.718789, 139.744682, 140.106291, 139.724634, 139.779996)
)
jpeg("denguehs.jpg", width=800, height=600, units="px")
par(mar=c(5,2,4,1))
plot(tokyo23chiba, col=NA, xlab="", ylab="", axes=FALSE)
print(par()$usr)
points(denguehs$lon, denguehs$lat, pch=16, col="red", cex=2)
text(denguehs$lon, denguehs$lat, denguehs$pn, col="black", pos=4, cex=1)
title(main="Suspected places where dengue transmission occurred¥n until 11 Sep 2014",
 sub="Geographical data: (c) Esri Japan")
dev.off()
```

2019-nCoVについてのメモとリンク

- ブログ(鐵人三國誌 https://minato.sip21c.org/im3r.html)に情報源へのリンク付メモ
  →抜粋して形式を整え保存(パンデミックインフルエンザ2009やデング熱2014と同様に)
  https://minato.sip21c.org/im3r/20200106.html
  →https://minato.sip21c.org/2019-nCoV-im3r.html#nCoV
- 体裁としては、2020年1月6日から**目次付時系列メモ**開始、後に**カテゴリ別リンク集**追加
- ナショナル・ジオグラフィックスの「研究室に行ってみた:特別編」にもなった
  https://natgeo.nikkeibp.co.jp/atcl/web/19/050800015/

新型コロナウイルス

日本ではあまり報道されていないように思うが，中国での原因不明の肺炎（
りかねないので，注意しておく必要があると思う。

大晦日の初発報告から1週間ちょっと経って，SARSともMERSとも異なる新

時系列のメモ

以下は，ところどころ筆者のspeculationも入った，未整理の情報です。

新型コロナウイルス（2020年1月6日，11日追記 - 当該鐵人三國誌）

日本ではあまり報道されていないように思うが，中国での原因不明の肺炎（リンク先はWHOのリリース）は，この段階での封じ込めに失敗するとPHEICになりかねないので，注意しておく必要があると思う。

大晦日の初発報告から1週間ちょっと経って，SARSともMERSとも異なる新型コロナウイルスが原因と確定し，とうとう死者が出た（1月11日初報）。大変心配。

変異株関係

時系列メモ目次

新型コロナウイルス（2020
インペリグループによる患者
患者数急増，西浦さんたちの
WHOはPHEIC宣言せず（2
絶対リスクと相対リスク（2
研究ラッシュが起こるかも（

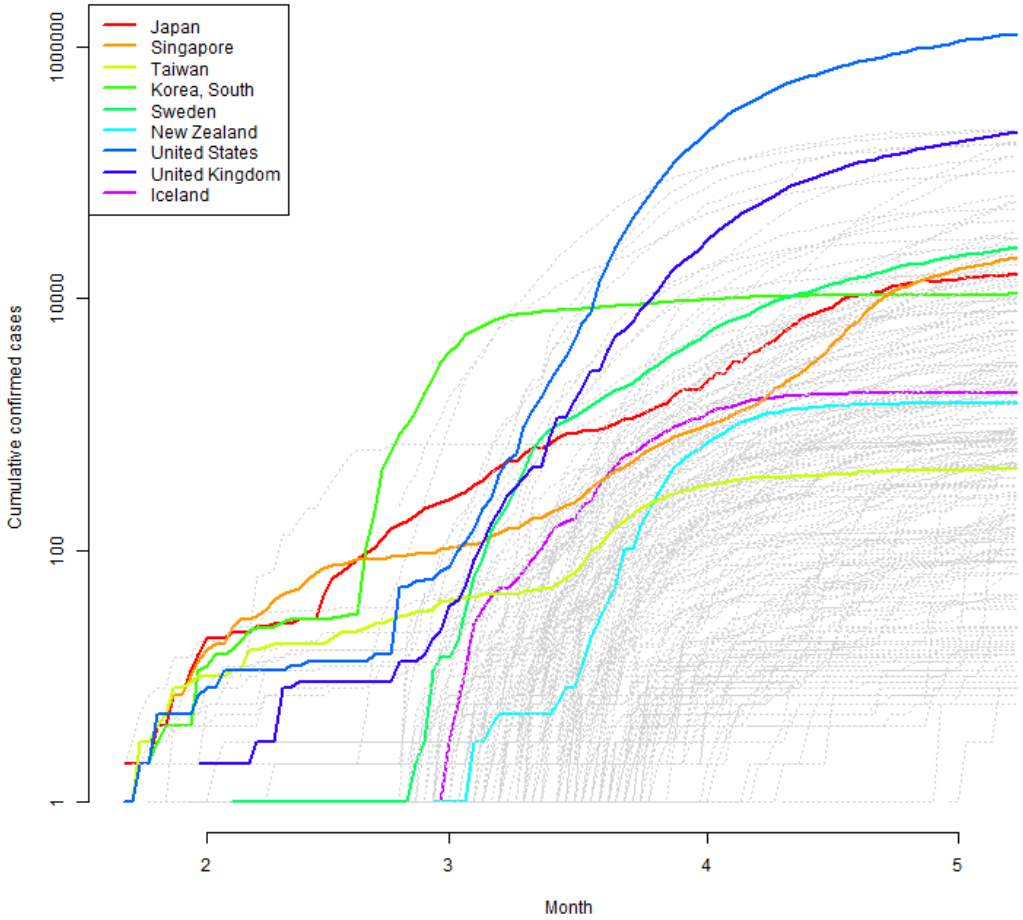
https://minato.sip21c.org/testgraph.R

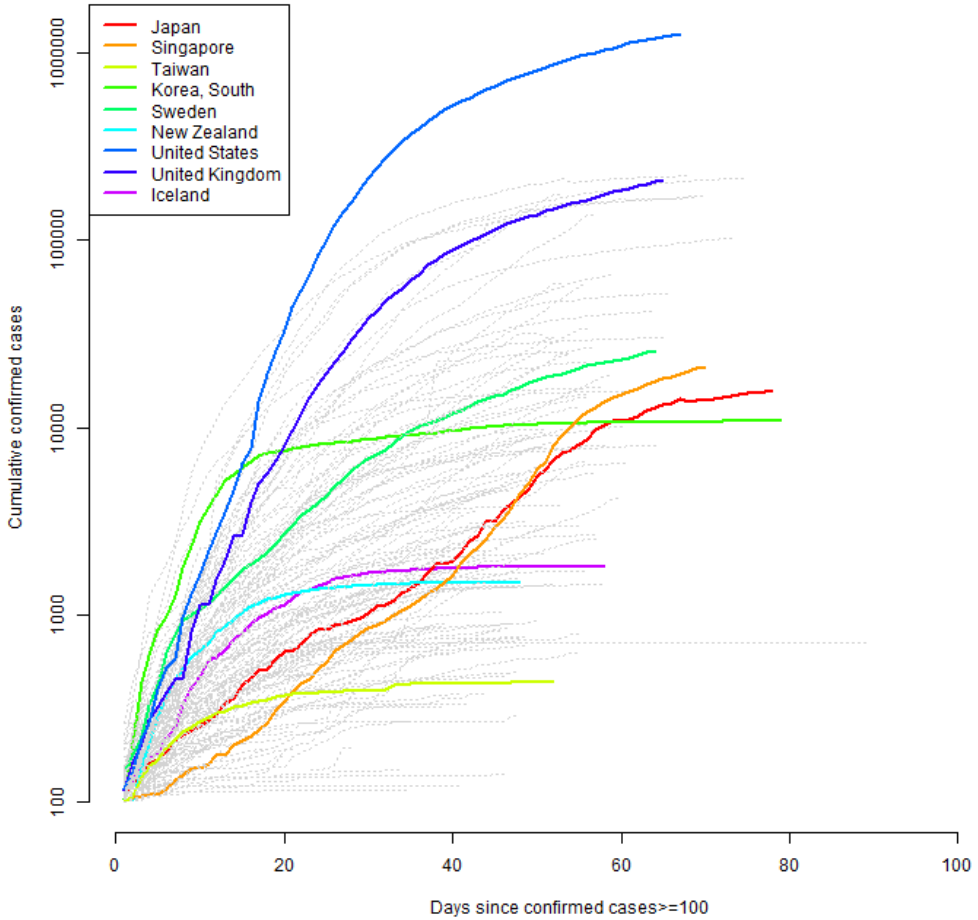
片対数グラフを描くコード（2020年5月9日 - 当該鐵人三國誌）

5月5日にリンクしたが，R関係のリソースは使いやすくて，国際比較のデータを読んで片対数グラフを描くコードくらいなら，このように（リンク先は先に示したコードの実行結果PNG画像）簡単にできる。ちなみに，このコードで工夫したのは，ローカルにデータファイルがあればそれを読み，無ければダウンロードしてから読んで描画するようにした点。

```
# Plotting confirmed cases against date in semilog scale.
# 8 May 2020, <minato-nakazawa@umin.net>
# Data taken from https://github.com/covid19datahub/COVID19
# https://covid19datahub.io/articles/data.html
# Reference
# Guidotti E, Ardia D (2020) COVID-19 Data Hub.
# Working paper https://doi.org/10.13140/RG.2.2.11649.81763
# For Natgeo's graph, see https://minato.sip21c.org/graphJu.R
if (!file.exists("data-1.csv")) {
  download.file("https://storage.covid19datahub.io/data-1.csv",
"data-1.csv")
}
x <- read.csv("./data-1.csv", stringsAsFactors=TRUE)
x$date <- as.Date(x$date)
x$country <- x$administrative_area_level_1 # countryという変数が消滅したため

options(scipen=7)
png("worldsemilog.png", width=1500, height=700)
layout(t(1:2))
xx <- subset(x, confirmed>0)
plot(confirmed ~ date, xx, log="y", type="n", frame=FALSE,
     xlab="Month", ylab="Cumulative confirmed cases")
cnts <- c("Japan", "Singapore", "Taiwan", "Korea, South",
"Sweden", "New Zealand", "United States", "United Kingdom",
"Iceland")
cols <- rainbow(10)
for (i in levels(x$country)) {
 dd <- subset(x, country == i)
 if (i %in% cnts) {
   lines(dd$date, dd$confirmed, col=cols[grep(i, cnts)], lwd=2)
 } else {
   lines(dd$date, dd$confirmed, col="lightgray", lwd=0.5, lty=3)
 }
}
legend("topleft", col=cols, lwd=2,
legend=cnts)

days <- 1:difftime(max(x$date),
min(x$date))
cfd <- x$confirmed[x$confirmed>=100]
plot(c(1,max(days)), c(min(cfd),
max(cfd)),
     log="y", type="n", frame=FALSE,
     xlab="Days since confirmed
cases>=100",
     ylab="Cumulative confirmed cases")
for (i in levels(x$country)) {
  dd <- subset(x, country == i)
  confirmed <-
dd$confirmed[dd$confirmed>=100]
  daysi <- days[1:length(confirmed)]
  if (length(daysi)>1) {
    if (i %in% cnts) {
      lines(daysi, confirmed, col=cols[grep(i,
cnts)], lwd=2)
    } else {
      lines(daysi, confirmed, col="lightgray",
lwd=0.5, lty=3)
    }
  }
}
legend("topleft", col=cols, lwd=2,
legend=cnts)
dev.off()
```

https://minato.sip21c.org/testgraph.R





COVID19パッケージの紹介(未使用)

COVID19パッケージ（2020年5月11日 - 当該鐵人三國誌）

一昨日は確定患者数について国際比較するコードを書いたが，ふと思い立って死者数について書いてみた。

deaths.Rとworlddeathssemilog.pngをリンクしておく。Rコードにコメントしてあるが，データは，COVID19パッケージのgithubページからリンクされているcsvファイルを使った（Guidotti E, Ardia D (2020) COVID-19 Data Hub. Working paper https://doi.org/10.13140/RG.2.2.11649.81763）。

ちなみに，使ったcsvファイル(data-1.csv)はレベル1，つまり国レベルの情報で，国名(country)，日付(date)，累積確定患者数(confirmed)，累積死者数(deaths)だけではなく，以下の変数も含まれている。

- 累積検査数(tests),
- 累積治癒数(recovered),
- 各日における入院患者数(hosp),
- 各日におけるICU入院患者数(icu),
- 各日における侵襲のある人工呼吸器を必要とする患者数(vent),
- 学校閉鎖状況（0: 対策無し，1: 閉鎖推奨，2: 部分的閉鎖要請，3: 全面的閉鎖要請）(school_closing),
- 職場閉鎖状況（0: 対策無し，1: 閉鎖または在宅勤務推奨，2: 部署あるいは職種により閉鎖要請，3: 社会維持に必須な職場を除き閉鎖または在宅勤務要請）(workplace_closing),
- イベントキャンセル（0: 対策無し，1: イベントキャンセル推奨，2: イベントキャンセル要請）(cancel_events),
- 集会制限（0: 制限なし，1: 1000人以上のマスギャザリング制限，2: 100人以上の集会制限，3: 10人以上の集会制限，4: 10人未満でも集会制限）(gatherings_restrictions),
- 交通停止（0: 対策無し，1: 交通停止あるいは減便推奨，2: 交通停止あるいは大部分の市民について利用禁止要請）(transport_closing),
- 外出制限（0: 対策無し，1: 外出非推奨，2: エクササイズや生活必需品購入のためを除き在宅要請，3: 2-3日に1度一人だけという最小限の例外を除き在宅要請）(stay_home_restrictions),
- 国内移動制限（0: 対策無し，1: 閉鎖推奨，2: 閉鎖要請）(internal_movement_restrictions),
- 国際移動制限（0: 対策無し，1: スクリーニング，2: ハイリスク地域からの到着者について検疫，3: ハイリスク地域について出入国禁止，4: 全面的国境閉鎖）(international_movement_restrictions),
- 情報キャンペーン（0: なし，1: 公的注意喚起，2: 複数セクタが協力した公的情報キャンペーン）(information_campaigns),
- 検査政策（0: なし，1: 症状がありかつ特定基準を満たす人だけ検査，2: 症状があれば誰でも検査，3: 無症状でも誰でも検査）(testing_policy),
- 接触者追跡（0: なし，1: 限定的接触者追跡，2: 全ての症例について包括的な接触者追跡）

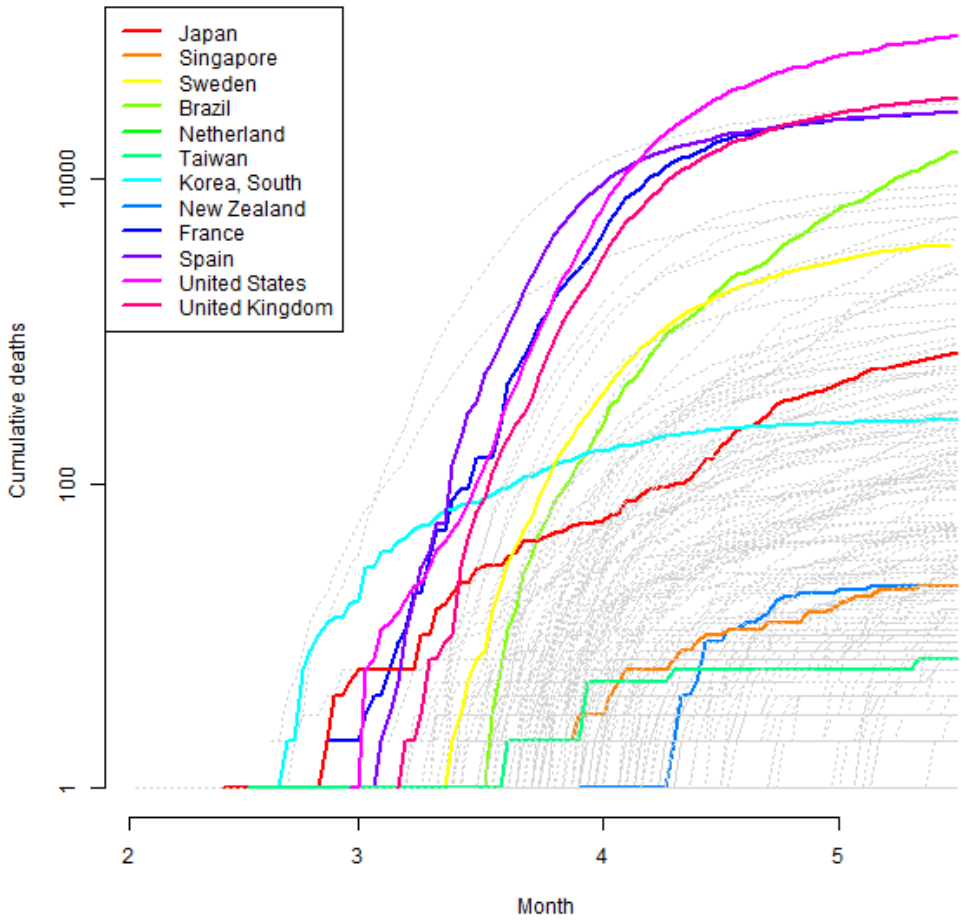
さらに，人口の情報（popが総人口，pop_femaleが女性人口割合，pop_14が年少人口割合，pop_15_64が生産年齢人口割合，pop_65が老年人口割合）なども含まれている。しかし，social distancingの程度とか三密予防をしているかといった情報は含まれていない。

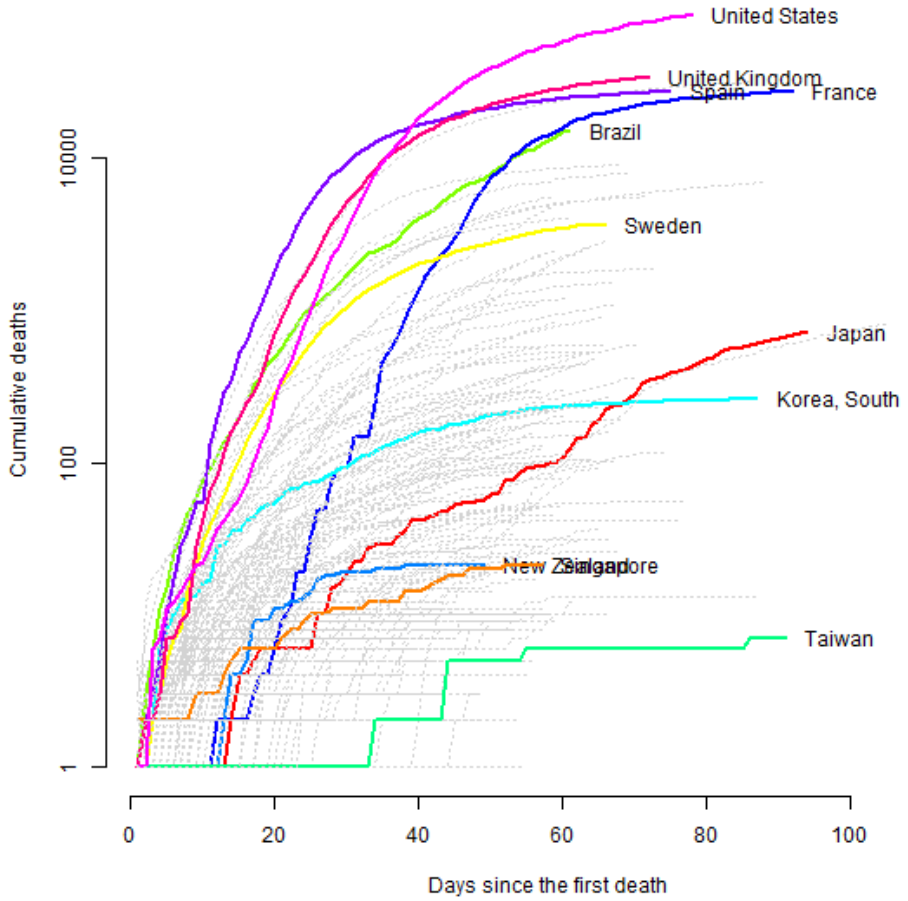
レベル1の他にレベル2のstateレベルのファイルとレベル3の市町村レベルのファイルがあり，COVID19パッケージを使うと，例えば，covid19(c("ITA","USA"), level = 3)でイタリアとUSAの市町村別データが得られるなど，どのレベルの情報でも読めると書かれている。

https://minato.sip21c.org/deaths.R

```
# Plotting deaths against date in semilog scale.
# 11 May 2020, <minato-nakazawa@umin.net>
# Data taken from https://github.com/covid19datahub/COVID19
# Reference
# Guidotti E, Ardia D (2020) COVID-19 Data Hub.
# Working paper https://doi.org/10.13140/RG.2.2.11649.81763
if (!file.exists("data-1.csv")) {
  download.file("https://storage.covid19datahub.io/data-1.csv",
"data-1.csv")
}
x <- read.csv("./data-1.csv", stringsAsFactors=TRUE)
x$date <- as.Date(x$date)
x$country <- x$administrative_area_level_1 # The variable
"country" disappeared!
cnts <- c("Japan", "Singapore",
 "Sweden", "Brazil", "Netherland",
 "Taiwan", "Korea, South", "New Zealand",
 "France", "Spain", "United States", "United Kingdom")
cols <- rainbow(length(cnts))

options(scipen=7)
png("worlddeathssemilog.png", width=1200, height=600)
layout(t(1:2))
xx <- subset(x, deaths>0)
plot(deaths ~ date, xx, log="y", type="n", frame=FALSE,
     xlab="Month", ylab="Cumulative deaths")
for (i in levels(x$country)) {
 dd <- subset(x, country == i)
 if (i %in% cnts) {
   lines(dd$date, dd$deaths, col=cols[grep(i, cnts)], lwd=2)
 } else {
   lines(dd$date, dd$deaths, col="lightgray", lwd=0.5, lty=3)
 }
}
legend("topleft", col=cols, lwd=2,
legend=cnts)

days <- 1:difftime(max(x$date),
min(x$date))
cfd <- x$deaths[x$deaths>0]
plot(c(1,max(days)), c(min(cfd),
max(cfd)),
     log="y", type="n", frame=FALSE,
     xlab="Days since the first death",
     ylab="Cumulative deaths")
for (i in levels(x$country)) {
  dd <- subset(x, country == i)
  death <- dd$deaths[dd$deaths>0]
  daysi <- days[1:length(death)]
  if (length(daysi)>1) {
    if (i %in% cnts) {
      lines(daysi, death, col=cols[grep(i,
cnts)], lwd=2)
      text(max(daysi)+1, max(death), i,
pos=4)
    } else {
      lines(daysi, death, col="lightgray",
lwd=0.5, lty=3)
    }
  }
}
dev.off()
```

出力：https://minato.sip21c.org/worlddeathssemilog.png

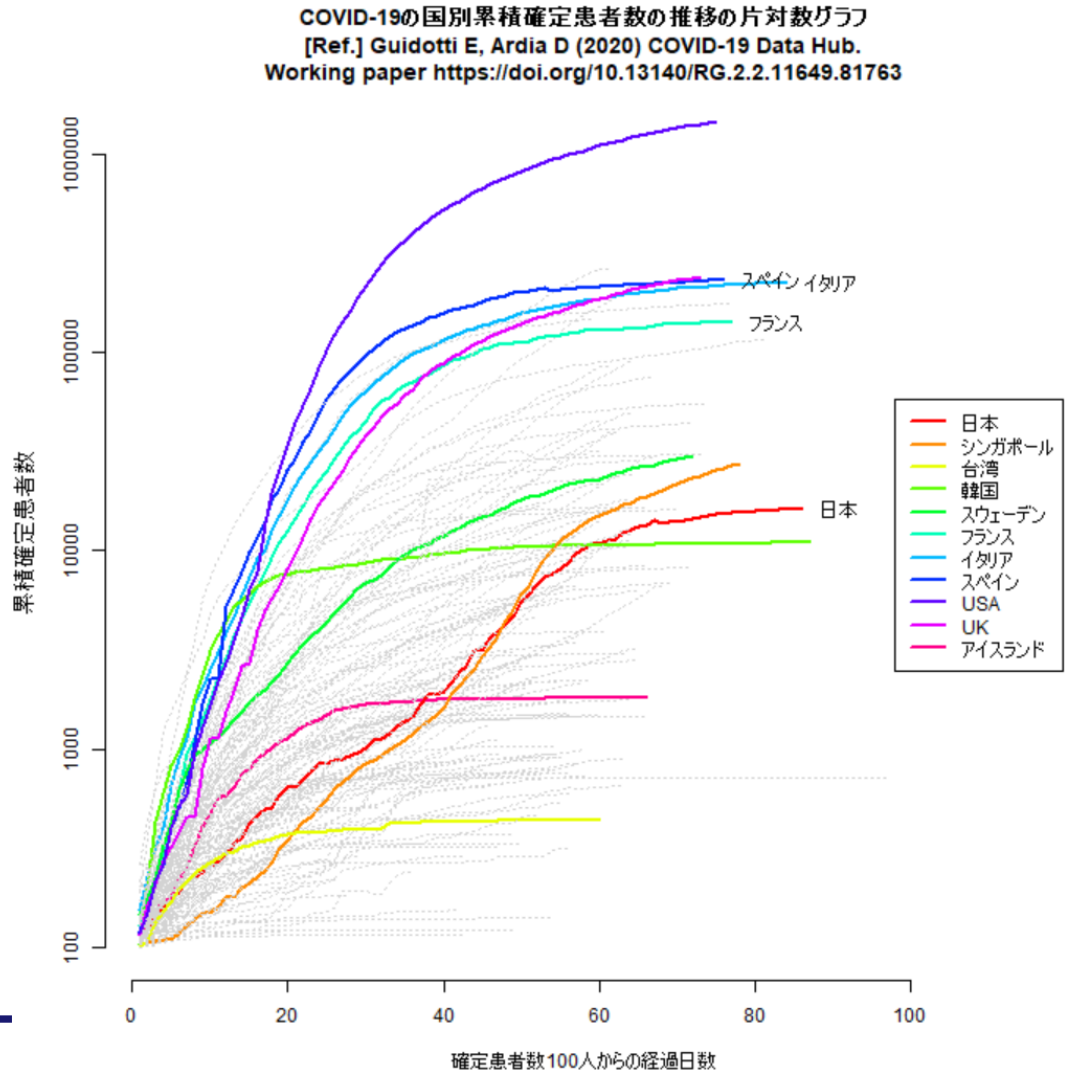




ナショジオ連載第6回(2020年5月17日 - 当該鐵人三國誌)

ナショジオ連載の第6回が公開された。今日の話は，検査性能（2月25日）と傾きが緩い意味（3月10日）関連。以前も書いたが，https://www.datacat.cc/covid/で作ったグラフはメディアへの掲載が許されないらしく，Rでグラフを作った。記事にURLが載っているコードだと英語表示なのと，いつの間にかデータファイルからcountryという変数が消えてしまったのでadministrative_area_level_1をcountryに付値し，メッセージも日本語化したコードを載せておく（Shift-JIS版とUTF-8版）。国名を変えたり増減させるだけで他の国も灰色でない表示にできるはずなので，例えばアイスランドよりブラジルを表示したかったら"Iceland"を"Brazil"に，"アイスランド"を"ブラジル"に変えるだけで，自動的にアイスランドの線が灰色になり，ブラジルの線にマゼンタ色が付く。

LancetのKaplan HS et al. "Voluntary collective isolation as a best response to COVID-19 for indigenous populations? A case study and protocol from the Bolivian Amazon" https://doi.org/10.1016/S0140-6736(20)31104-1（2020年5月15日掲載）は，ボリビアアマゾンに住むTsimaneという採集・焼畑農耕民でのCOVID-19の拡散を防ぐため，人類学者，医師，部族の指導者，地区当局が共同で，多段階のCOVID-19予防と封じ込めの計画を開発し実装したという話。第1段階では教育，アウトリーチ，準備，第2段階では封じ込め，患者管理，隔離に焦点を当てるという戦略は，Tsumane同様にCOVID-19が流行してしまったら高い死亡リスクが見込まれる他の先住民にも適用可能だろうと論じている。



波長222 nmのUV-C（2020年5月18日 - 当該鐵人三國誌）

ナショジオ連載の第7回が公開された。今日の話は　まとめスライドrev.2のスラ

https://minato.sip21c.org/graphJ.R

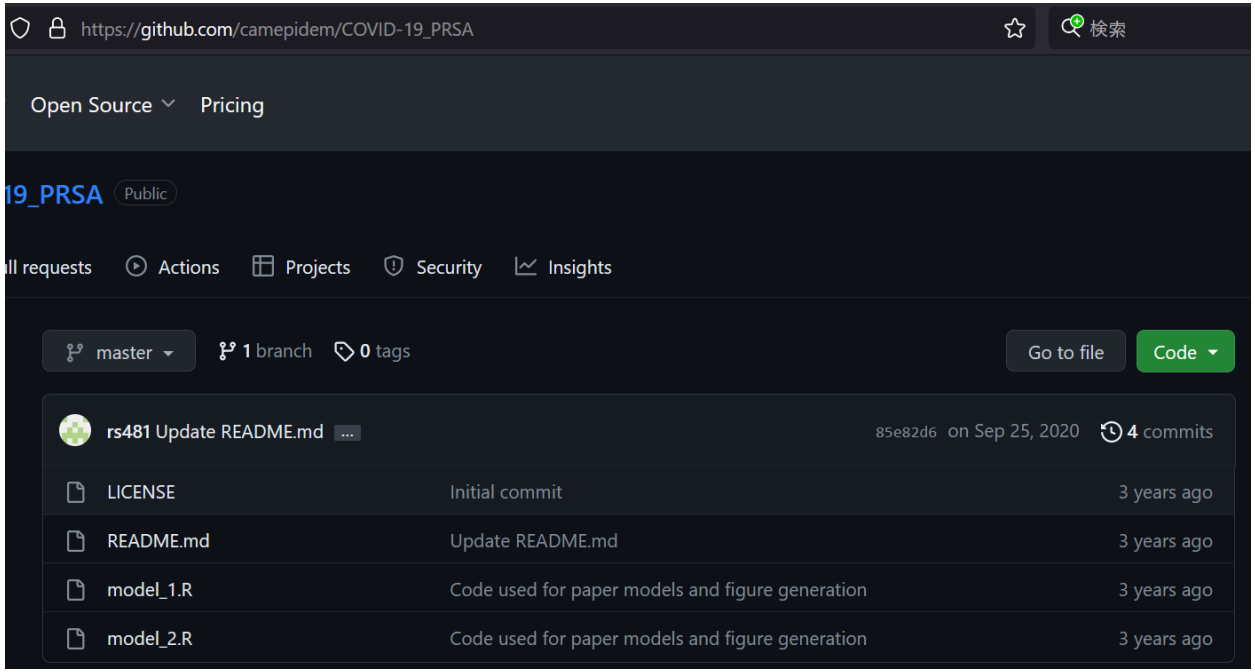
```
# Plotting confirmed cases against date in semilog scale.
# 8 May 2020, <minato-nakazawa@umin.net>
# Data taken from https://github.com/covid19datahub/COVID19
# Reference
# Guidotti E, Ardia D (2020) COVID-19 Data Hub.
# Working paper https://doi.org/10.13140/RG.2.2.11649.81763
if (!file.exists("data-1.csv")) {
download.file("https://storage.covid19datahub.io/data-1.csv", "data-1.csv")
}
x <- read.csv("./data-1.csv", stringsAsFactors=TRUE)
x$date <- as.Date(x$date)
x$country <- x$administrative_area_level_1 # countryという変数が消滅したため

options(scipen=7)
png("worldsemilogJ2.png", width=700, height=700)

cnts <- c("Japan", "Singapore", "Taiwan", "Korea, South", "Sweden", "France", "Italy", "Spain", "United States", "United Kingdom", "Iceland")
cntsJ <- c("日本","シンガポール","台湾","韓国","スウェーデン","フランス","イタリア","スペイン","USA","UK","アイスランド")
cntsJ2 <- c("日本","","","","","","フランス","イタリア","スペイン","","","","")
cols <- rainbow(length(cnts))
days <- 1:difftime(max(x$date), min(x$date))
cfd <- x$confirmed[x$confirmed>=100]
plot(c(1,max(days)), c(min(cfd), max(cfd)),
     log="y", type="n", frame=FALSE,
     xlab="確定患者数100人からの経過日数",
     ylab="累積確定患者数",
     main="COVID-19の国別累積確定患者数の推移の片対数グラフ
[Ref.] Guidotti E, Ardia D (2020) COVID-19 Data Hub.
Working paper https://doi.org/10.13140/RG.2.2.11649.81763")
for (i in levels(x$country)) {
  dd <- subset(x, country == i)
  confirmed <- dd$confirmed[dd$confirmed>=100]
  daysi <- days[1:length(confirmed)]
  if (length(daysi)>1) {
    if (i %in% cnts) {
      lines(daysi, confirmed, col=cols[grep(i, cnts)], lwd=2)
      text(max(daysi)+1, max(confirmed), cntsJ2[grep(i, cnts)], pos=4)
    } else {
      lines(daysi, confirmed, col="lightgray", lwd=0.5, lty=3)
    }
  }
}
legend("right", col=cols, lwd=2, legend=cntsJ)
dev.off()
```

感染モデルの論文は、多数Rコードが公開された

マスクの効果のモデル論文（2020年6月12日 - 当該鐵人三國誌）

Furuse Y, Sando E, Tsuchiya N, Miyahara R, Yasuda I, Ko YK, et al. Clusters of coronavirus disease in communities, Japan, January–April 2020. Emerg Infect Dis. 2020 Sep [cited on 12 June 2020]. https://doi.org/10.3201/eid2609.202272（2020年6月10日掲載）は，たぶんクラスター対策班の活動の一環で手分けして集めたデータを記述的に解析した結果と思われる，日本の2020年1月から4月までのクラスターデータの速報。東北大学から京都大学の白眉プロジェクトに採用されたウイルス学者，古瀬祐気さんが筆頭著者。

奥村先生のtweetで知ったが，Stutt ROJH et al. "A modelling framework to assess the likely effectiveness of facemasks in combination with 'lock-down' in managing the COVID-19 pandemic." Proc. Royal Soc. A（2020年6月10日）は，マスク着用の効果を評価しているモデルで，症状がなくても一般公衆が常にマスクをしていれば実効再生産数を1未満にできて感染拡大を緩和できるとabstractに書かれている。ざっと目を通した感じでは，マスク着用するかどうかは日々刻々と変わると思うので，シミュレーションするなら個人ベースで，その振り分けを二項乱数でやった方が良いと思うし（ぼくが長袖長ズボンの着用によるマラリア防御モデルでやったように），デフォルト値がmF=1，mA=0.5，mS=0.5，mD=0.5というのは非現実的で，ウイルスが付着したものを触った手で顔を触るのを防ぐわけだから，mF=0.1とかmF=0がデフォルトであるべきだし，普通の布マスクやサージカルマスクならば，着用している人の唾液飛沫が出て行くのを防ぐ効果の方が飛沫やマイクロ飛沫が入ってくるのを防ぐ効果よりずっと大きいはずだから，mDは0.5より大きく，mAやmSは0.5よりずっと小さくなると考えるのが自然ではないだろうか。コードが公開されている（MITライセンス，いくつかの依存パッケージはあるが，ピュアなRのコード2つからなる）ので，後で動かしてみよう。

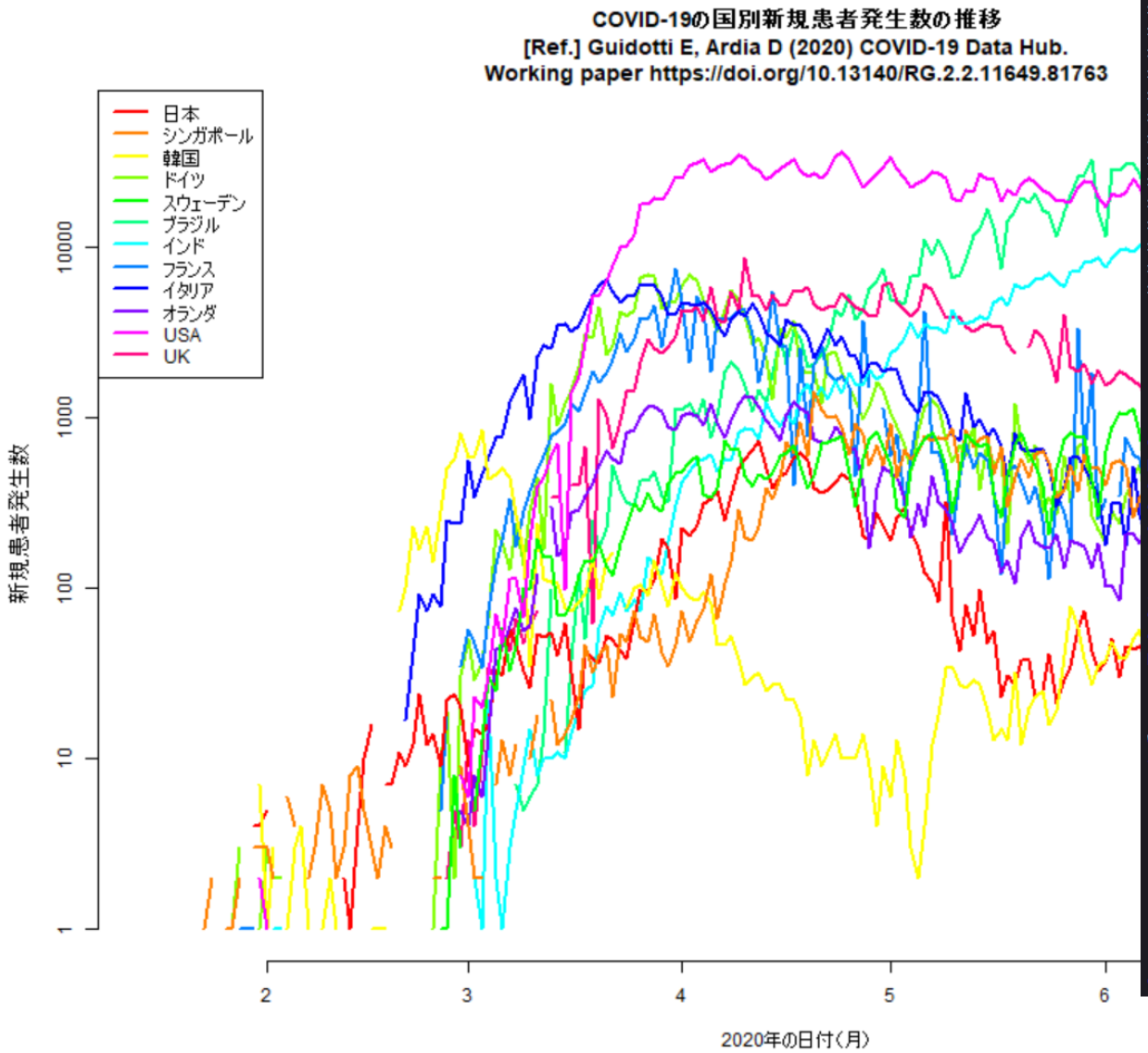


COVID-19パッケージを使用(2020.6.27-8)

https://minato.sip21c.org/libuse.R

3文字国名コードは https://www-yousei.jsps.go.jp/yousei1/kuniList.do

ちなみに，RのCOVID19パッケージを使って，日付別新規患者発生数をいくつかの国についてと都道府県について描かせてみると（コード），下図が描か
増加傾向にある。



https://minato.sip21c.org/libuse.R

```
# Plotting new confirmed cases against date in semilog scale.
# 8 May 2020, <minato-nakazawa@umin.net>
# Data taken from https://github.com/covid19datahub/COVID19
# 19 June 2020, 新規患者数に
# 29 June 2020, パッケージ利用＋都道府県レベル
# > library(COVID19)
# > res <- covid19("JPN", level=2)
# > levels(as.factor(res$administrative_area_level_2))
#  [1] "Aichi"     "Akita"     "Aomori"    "Chiba"     "Ehime"     "Fukui"     "Fukuoka"   "Fukushima" "Gifu"
# [10] "Gunma"     "Hiroshima" "Hokkaido"  "Hyogo"     "Ibaraki"   "Ishikawa"  "Iwate"     "Kagawa"    "Kagoshima"
# [19] "Kanagawa"  "Kochi"     "Kumamoto"  "Kyoto"     "Mie"       "Miyagi"    "Miyazaki"  "Nagano"    "Nagasaki"
# [28] "Nara"      "Niigata"   "Oita"      "Okayama"   "Okinawa"   "Osaka"     "Saga"      "Saitama"   "Shiga"
# [37] "Shimane"   "Shizuoka"  "Tochigi"   "Tokushima" "Tokyo"     "Tottori"   "Toyama"    "Wakayama"  "Yamagata"
# [46] "Yamaguchi" "Yamanashi"

if (!require(COVID19)) {
 install.packages("COVID19", dep=TRUE)
 library(COVID19)
}

options(scipen=7)
png("incidenceworldnew.png", width=1000, height=700)
cnts <- c("JPN", "SGP", "KOR", "DEU", "SWE", "BRA", "IND", "FRA", "ITA", "NLD", "USA", "GBR")
cntsJ <- c("日本", "シンガポール", "韓国", "ドイツ", "スウェーデン", "ブラジル", "インド", "フランス", "イタリア", "
cols <- rainbow(length(cnts))
x <- covid19(country=cnts, level=1, start="2020-01-01", end=Sys.Date())
str(x)
x$id <- as.factor(x$id)
cfd <- c(x$confirmed[1], x$confirmed[-1]-x$confirmed[-length(x$confirmed)])
cfd <- ifelse(cfd<0, 0, cfd)
plot(cfd ~ date, data=x, xlim=c(min(x$date), max(x$date)+20),
     log="y", type="n", frame=FALSE, xlab="2020年の日付（月）",
     ylab="新規患者発生数", main="COVID-19の国別新規患者発生数の推移
[Ref.] Guidotti E, Ardia D (2020) COVID-19 Data Hub.
Working paper https://doi.org/10.13140/RG.2.2.11649.81763")
for (i in levels(x$id)) {
  dd <- subset(x, id == i)
  dd$confirmedpp <- c(dd$confirmed[1], dd$confirmed[-1]-dd$confirmed[-length(dd$confirmed)])
  dd$confirmedpp <- ifelse(dd$confirmedpp<0, 0, dd$confirmedpp)
  if (i %in% cnts) {
    lines(dd$date, dd$confirmedpp, col=cols[grep(i, cnts)], lwd=2)
    text(max(dd$date)+1, dd$confirmedpp[length(dd$confirmedpp)], cntsJ[grep(i, cnts)], pos=4)
  }
}
legend("topleft", col=cols, lwd=2, legend=cntsJ)
dev.off()
```

パッケージ紹介

EpiEstimパッケージ（2020年7月25日 - 当該鐵人三國誌）

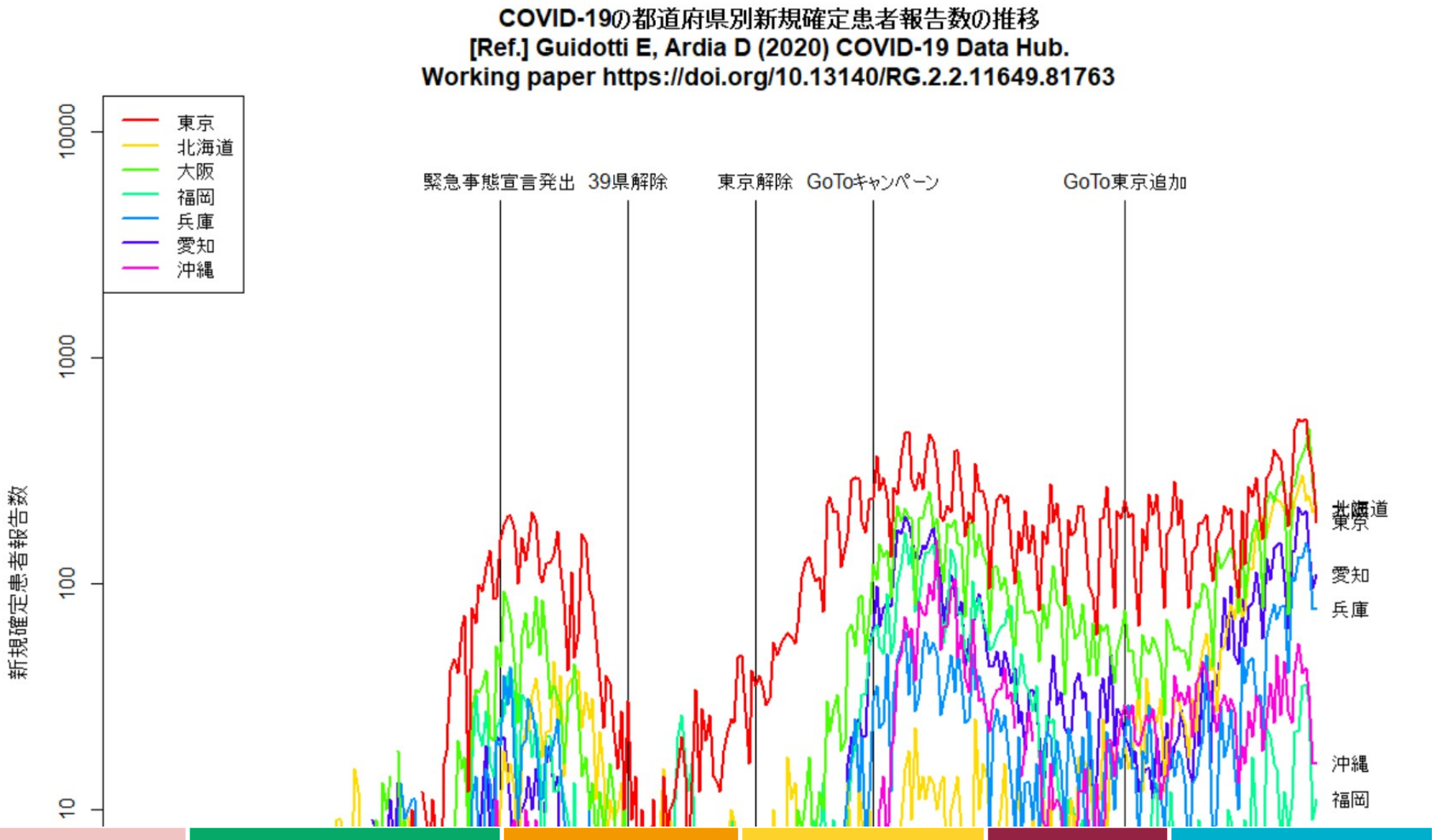
このtweetで紹介されている論文はプレプリント（2020年7月3日）だが，その中で計算に使われているRのEpiEstimというパッケージ（vignette）は，既にいくつかの査読を通った論文でも使われている。例えば，Ruiz-Patiño A et al. "Mortality and Advanced Support Requirement for Patients With Cancer With COVID-19: A Mathematical Dynamic Model for Latin America"（JCO Global Oncology，2020年5月29日）やNajafi F et al. "Serial interval and time-varying reproduction number estimation for COVID-19 in western Iran"（New microbes and new infections，2020年6月14日）でも使われていて，MCMCを使った再生産数リアルタイム推定の常套手段になりつつあると言って良いかもしれない。暇を見つけて試してみたい。最初に触れたプレプリント論文の著者たちは，EpiEstimを使ってCOVID-19の再生産数推定を実装するためのパッケージとして，EpiNow，EpiSoon，NCoVUtilsという3つを開発しているようだ。

日本の新規感染者数推移グラフ

https://minato.sip21c.org/libusejapan.R

11月下旬から12月上旬のメモ（2020年12月9日）

日本の東京や大阪などのCOVID-19の新規感染者数推移を縦軸を対数軸にして描画するRコードは，libusejapan.Rとしてダウンロードできる。COVID19パる度にデータをダウンロードし，実行日の前日までをプロットするようになっている。(2020.11.24)

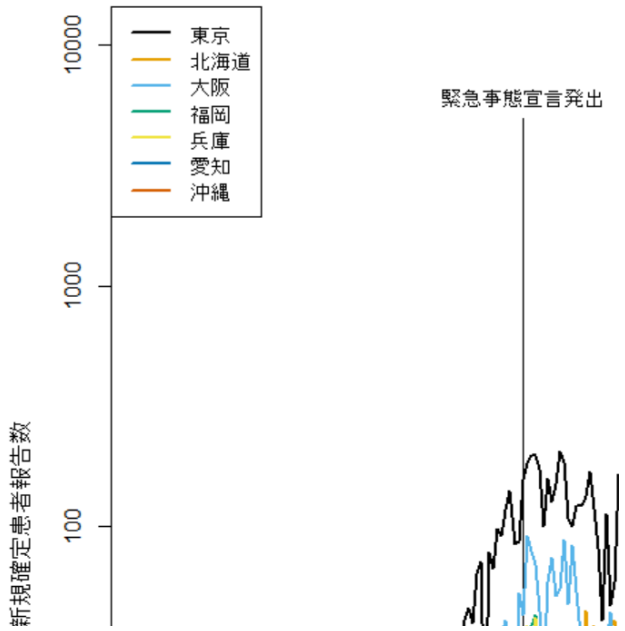


グラフの色使いをCUDに変更

最近のBMJニュースから（2020年12月25E

看護専攻の公衆衛生学の11回目は感染症とその予防。最後の
言っておいた。(2020.12.21)

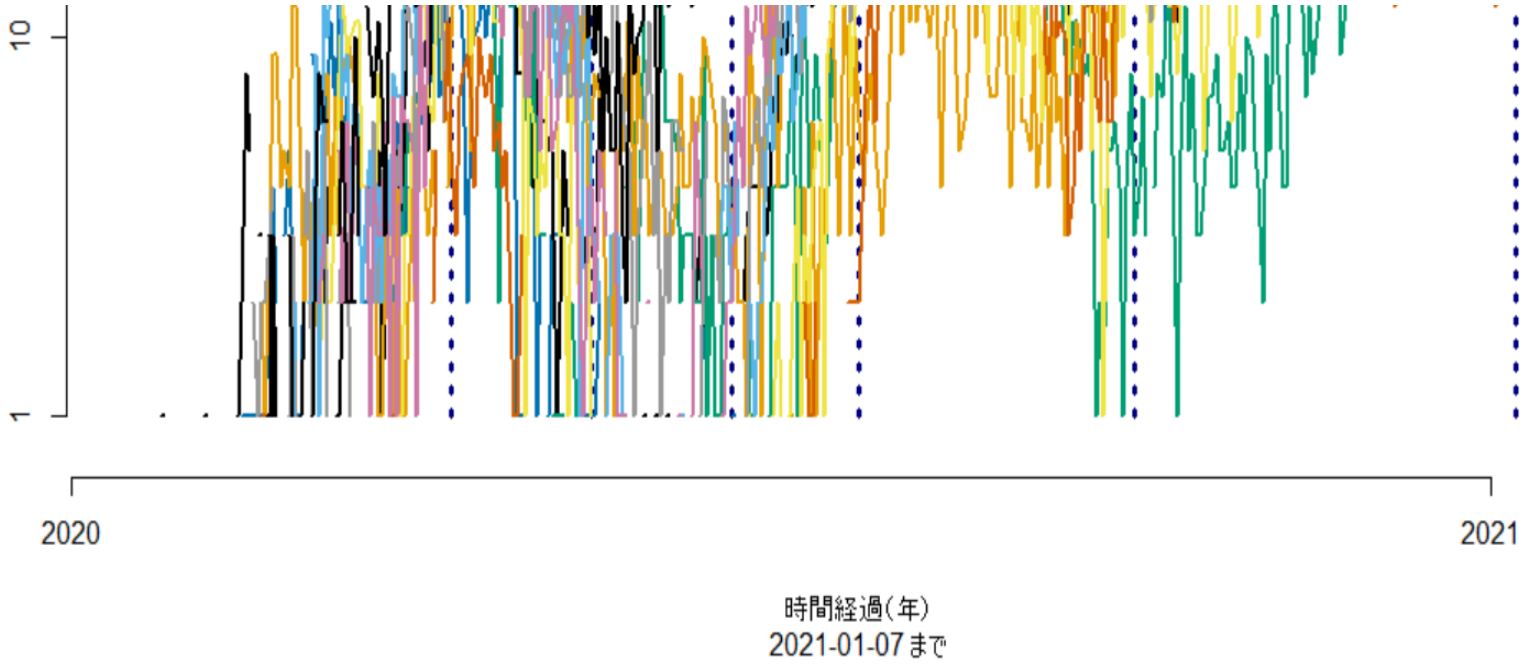
libusejapan.Rをpalette("Okabe-Ito")を使って色をsafe col



https://minato.sip21c.org/libusejapan.R

```
if (!require(COVID19)) {
 install.packages("COVID19", dep=TRUE)
 library(COVID19)
}

options(scipen=7)
prefs <- c("Tokyo", "Hokkaido", "Osaka", "Fukuoka", "Hyogo", "Aichi", "Okinawa",
"Saitama", "Kanagawa", "Miyagi")
prefsJ <- c("東京", "北海道", "大阪", "福岡", "兵庫", "愛知", "沖縄", "埼玉", "神奈川",
"宮城")
# cols <- rainbow(length(prefs))
cols <- 1:length(prefs)
palette("Okabe-Ito")
y <- covid19(country="JPN", level=2, start="2020-01-01", end=Sys.Date())
y <- subset(y, date<max(date))
y$prefs <- as.factor(y$administrative_area_level_2)
cfd <- c(y$confirmed[1], y$confirmed[-1]-y$confirmed[-length(y$confirmed)])
cfd <- ifelse(cfd<0, 0, cfd)
plot(cfd ~ as.Date(date, "%d/%m/%y"), data=y, xlim=c(min(y$date), max(y$date)+20),
ylim=c(1, 10000),
     axes=FALSE, log="y", type="n", frame=FALSE, xlab="時間経過（年）",
     ylab="新規確定患者報告数", main="COVID-19の都道府県別新規確定患者報告数の推移
[Ref.] Guidotti E, Ardia D (2020) COVID-19 Data Hub.
Working paper https://doi.org/10.13140/RG.2.2.11649.81763",
sub=sprintf("%s まで", max(y$date)))
axis(1, c(as.Date(sprintf("2020-%02d-01", 1:12)), as.Date(sprintf("2021-%02d-01", 1:4))),
 c(sprintf("2020-%02d-01", 1:12), sprintf("2021-%02d-01", 1:4)))
axis(2, c(1, 10, 100, 1000, 10000))
segments(as.Date("2020-04-08"), 1, as.Date("2020-04-08"), 8000, lty=3, lwd=3, col="navy")
text(as.Date("2020-04-08"), 5000, "緊急事態宣言¥n発出", pos=3)
# (以下略)
```

CRANの新しいパッケージとか（2021年1月10日）

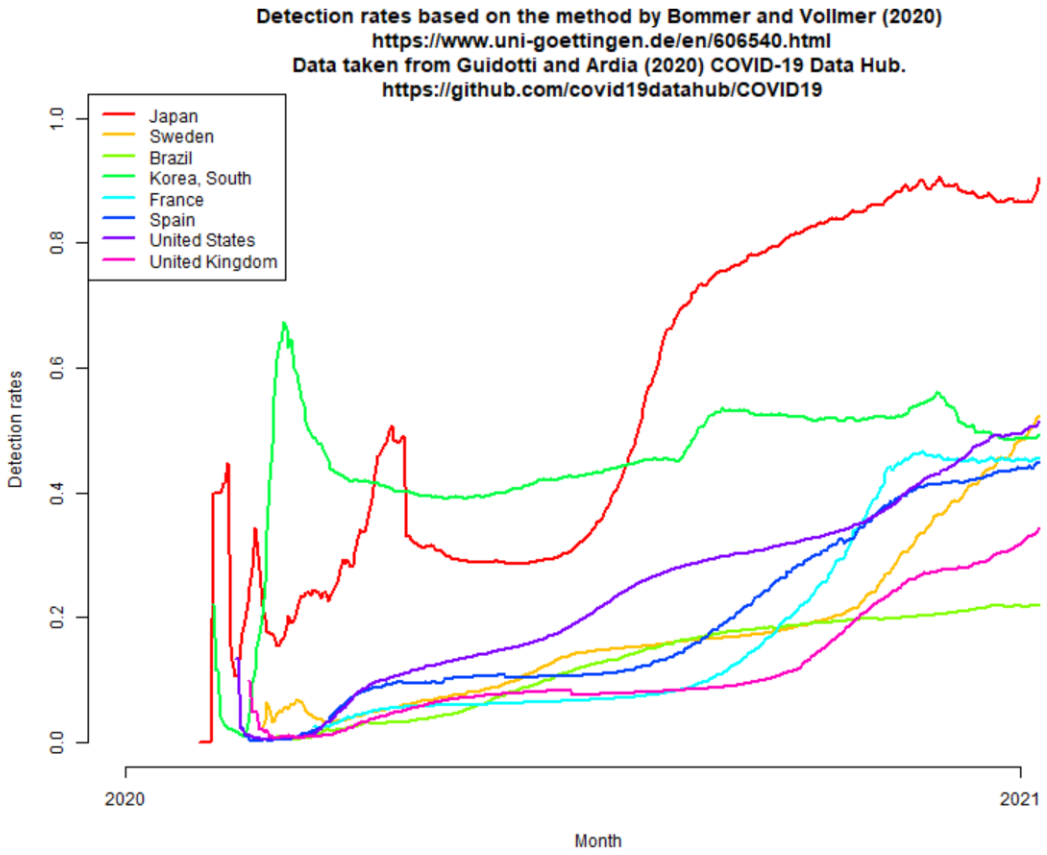


昨日付けで，covid19jpというパッケージがcranに入っていた。GitHubに書かれている説明によると，データソースは，大変有名な東洋経済の荻原和樹さんが公開しているものとのこと。荻原さんの公開データはCSV形式なので，covid19jpパッケージは，そこにカスタマイズされたインターフェースを付け加えたということだろう。cranのCOVID19関係のパッケージも随分増えて，ぼくはこれまでも作図のためCOVID19パッケージをよく使ってきたが，日本のデータに特化した上述covid19jpパッケージやブラジルのマイクロデータを提供しているCOVIDIBGEパッケージもできてきたし，移動の傾向をさまざまなソースから引っ張ってくるcovid19mobilityパッケージ（紹介記事）もあるし，モデルについてもSPARSMODrパッケージが入院率が変化したときの感染動態の理解に使えることを示した文書が発表されたりしているし，7月にも触れたEpiEstimパッケージを使った論文はますます増えて，Cheng Q et al. "Heterogeneity and effectiveness analysis of COVID-19 prevention and control in major cities in China through time-varying reproduction number estimation"（Scientific Reports，2020年12月15日）でも使われている。(2021.1.10)

https://minato.sip21c.org/detectionrates-1.1.R

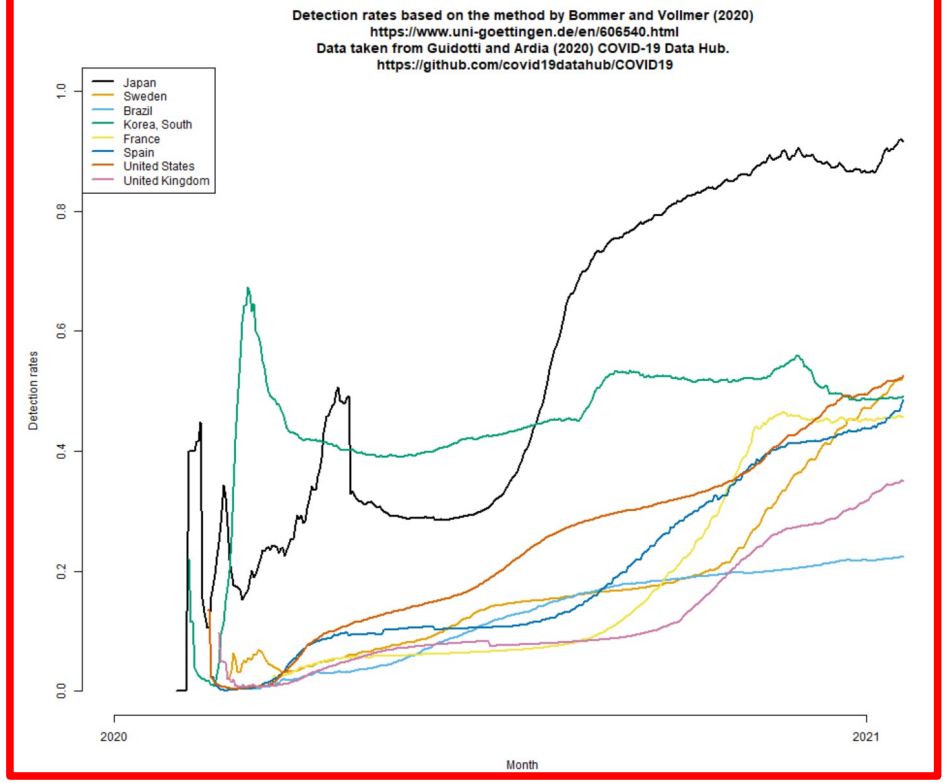
検出率？（2021年1月24日）

昨年4月，Gigazineで紹介されていたVollmer教授らの年齢調整IFRで死亡数を割って求めた感染者数と確定感染者数から検出割合を計算する方法について簡単に説明したが，最近までのデータについて，この方法を適用してみると（Rコード），下図のように，検出率が異様に高くなってしまう。こんなに捕捉されているとは考えにくいので，可能性としては，(1)昨年4月に想定されていたよりIFRが下がっている，(2)感染確定から死亡までの平均期間が長くなっている――Prof. Vollmerの方法では2週間という仮定なので，確かに短すぎるだろう，(3)見過ごされる死亡が増えている，といったことが考えられる。



検出率再び（2021年2月3日）

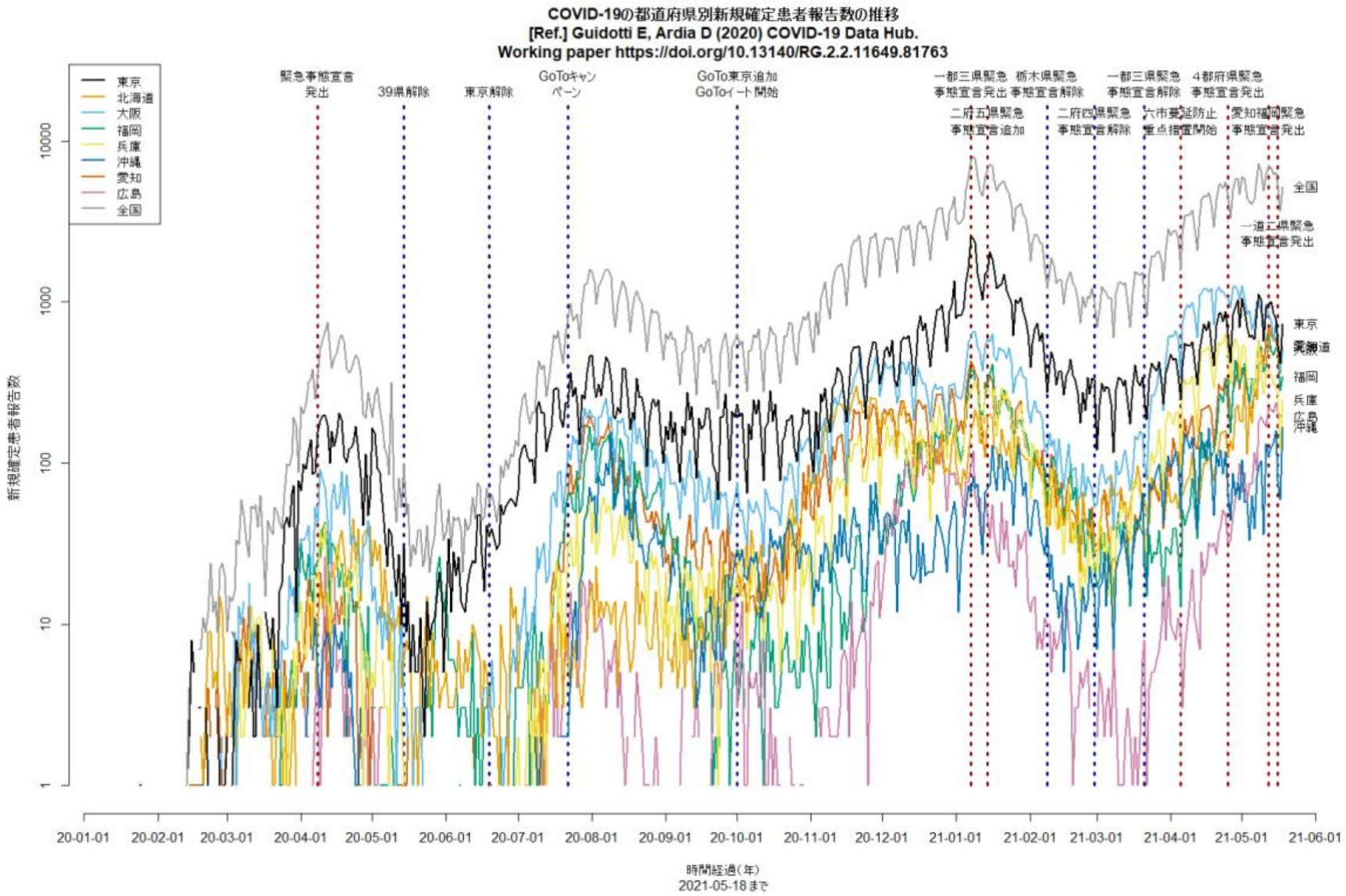
この検討は簡単にできるので続けているが，Vollmer教授の方法で推定される捕捉率が異様に高いままなので（現在までのデータの"Okabe-Ito"に変えて描画したのが下図），おそらく実はもっと捕捉率が低い可能性(1)～(3)のどれかが現実だと思われる。



（追記）仮に上記3つの可能性がすべて否定されたら，日本の感染者はこれまで8割方捕捉されてきたことになる。それが意味するのは，(1)保健所の方々を中心とする接触者追跡により，感染者と濃厚接触して二次感染した人の（前向きの）リンクはほぼ辿れていたこと，(2)しかし時間遅れなどのため，その先の感染を防ぐことができていなかったこと，その結果が(3)年末年始の急激な感染拡大になってしまったこと，そして，最近いくつかの自治体が始めてしまったように接触者追跡を諦めたら，捕捉率（検出率）が急激に低下し，さらなる感染

縦線segments()の色分け開始

また報告数推移グラフを作ってみた。(2021.5.19)

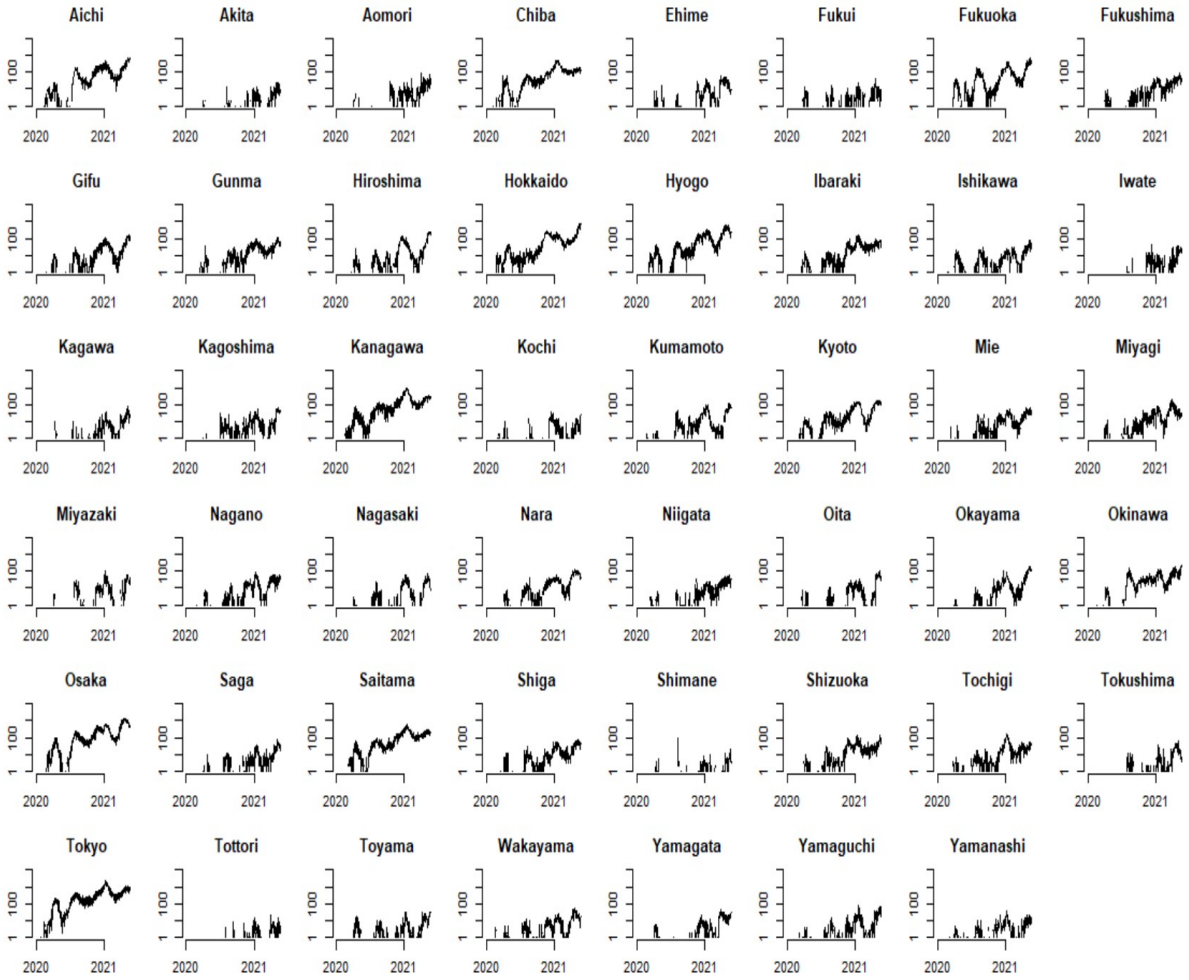


都道府県別画面分割描画

https://minato.sip21c.org/libusejapanbypref.R

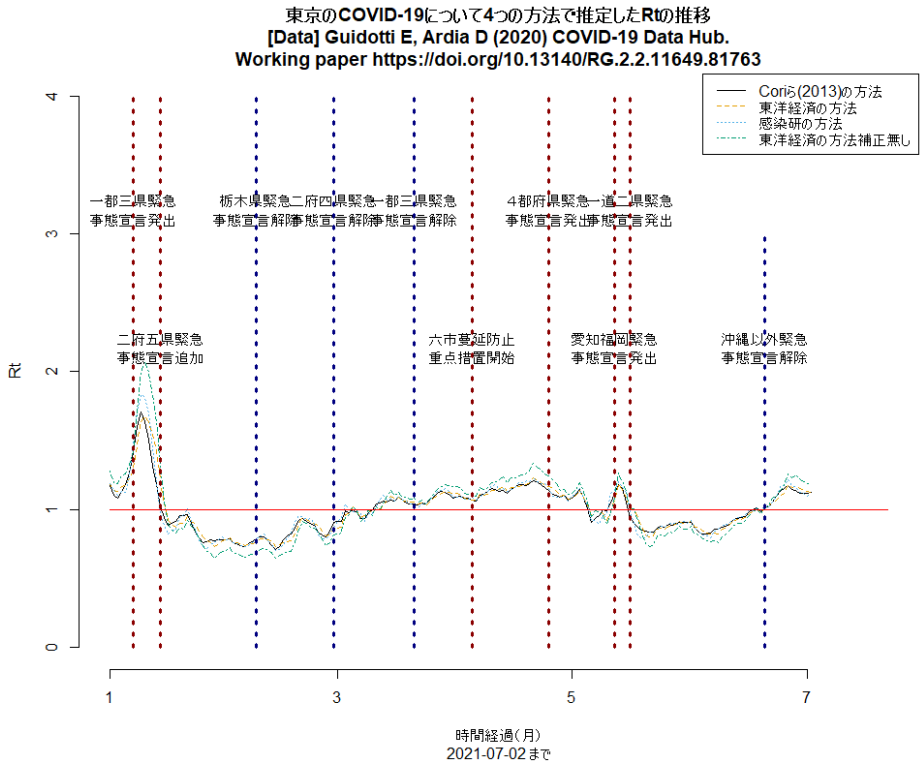
性差についての仮説（2021年5月23日）

画面を6行8列に分割して都道府県別covid-19の新規報告数推移の折れ線を描くコードを動かすと下図が得られる。色分けして重ね描きするより，ある意味
い。




```
if (!require(COVID19)) {
 install.packages("COVID19", dep=TRUE)
 library(COVID19)
}

x <- covid19(country="JPN", level=1, start="2020-01-01",
end=Sys.Date())
x <- subset(x, date<max(date))
x$cfd <- c(x$confirmed[1], x$confirmed[-1]-x$confirmed[-
length(x$confirmed)])
x$cfd <- ifelse(x$cfd<0, 0, x$cfd)

options(scipen=7)
y <- covid19(country="JPN", level=2, start="2020-01-01",
end=Sys.Date())
y <- subset(y, date<max(date))
y$prefs <- as.factor(y$administrative_area_level_2)
prefnames <- names(table(y$prefs))
y$cfd <- c(y$confirmed[1], y$confirmed[-1]-y$confirmed[-
length(y$confirmed)])
y$cfd <- ifelse(y$cfd<0, 0, y$cfd)
par(mar=c(3, 2, 2, 1))
layout(matrix(1:48, 6, 8, byrow=TRUE))
XAXIS <- pretty(c(as.Date("2020-01-01"), Sys.Date()))
for (i in prefnames) {
  dd <- subset(y, prefs==i)
  plot(cfd ~ as.Date(date, "%d/%m/%y"), data=dd,
       xlim=c(min(XAXIS), max(XAXIS)), ylim=c(1, 20000),
       log="y", type="l", frame=FALSE, col=1,
main=sprintf("%s", i))
}
```

https://minato.sip21c.org/libusejapanforRt.R

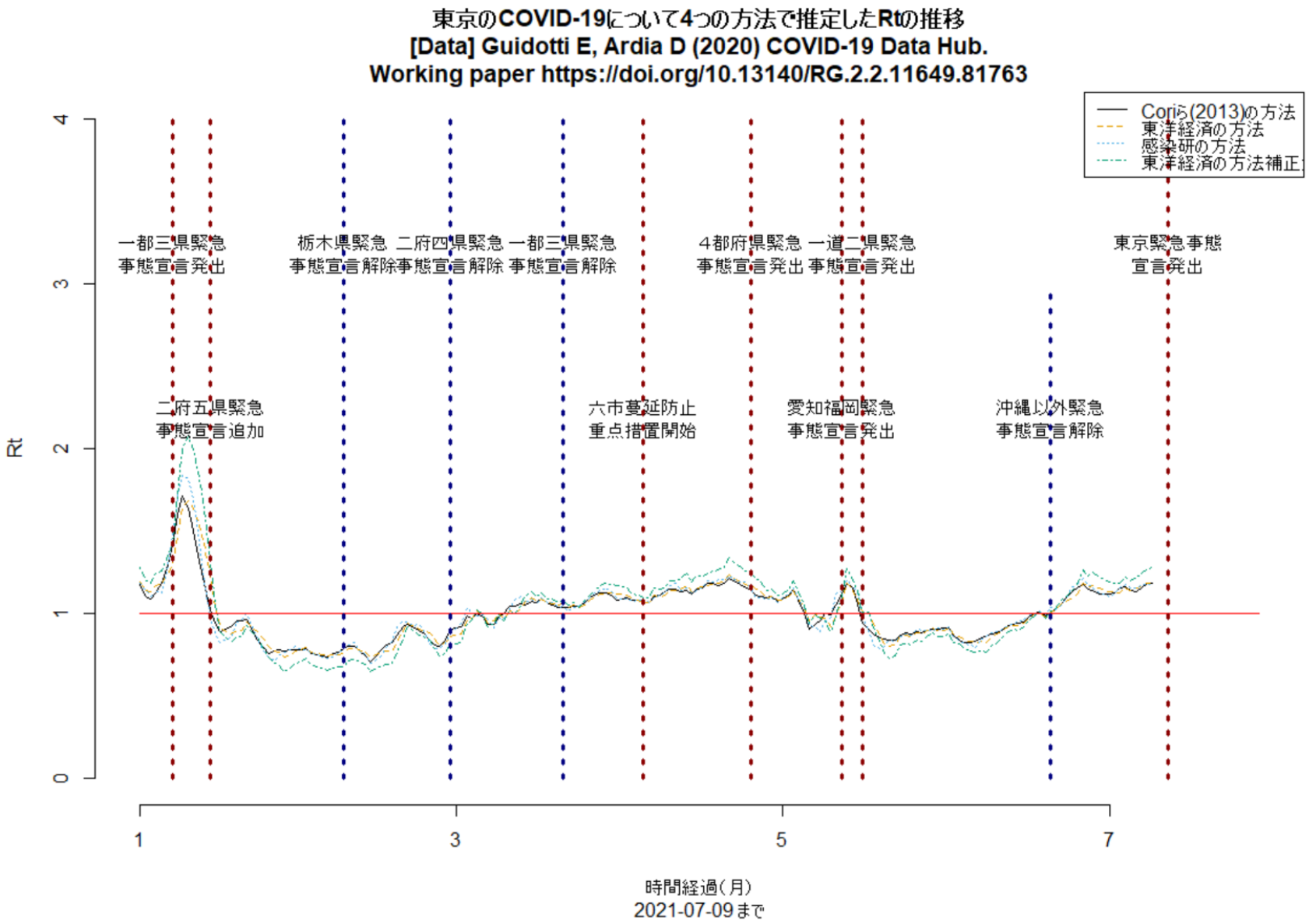


- 感染研「直近7日間の新規陽性報告者数／（世代時間）日前7日間の新規陽性報告者数」，世代時間を5日
  →Cori A et al. ”A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics”（AJE, 2013年9月15日）とはほぼ一致
- 2020年5月12日西浦さん（直近7日間の新規報告患者数／その前7日間の新規報告患者数）が近似的に使えるかもしれない
- 東洋経済（直近7日間の新規陽性者数／その前の7日間の新規陽性者数）^（平均世代時間／報告間隔）；**平均世代時間は5日間，報告間隔は7日間**
- 西浦さんが厚労省に提示している資料や論文：世代時間は4.8±2.3日，潜伏期間は5.6±3.9日とし，右側打ち切りを考慮し，Rのsurveillanceパッケージを使って確定診断日から発症日を逆推定した上で発症日から感染日を逆推定し，感染日から二次感染日までの間隔である世代時間について再生方程式を立ててRtを計算。報告遅れも世代時間のばらつきも考慮できるが，コンピュータなしでは計算できないし，何日も前の値までしか推定できない

```
if(!require(EpiEstim)) {
  install.packages("EpiEstim", dep=TRUE)
  library(EpiEstim)
}
# use estimate_R() funtion, which includes Cori et al (2013)'s method

# moving average from https://stackoverflow.com/questions/743812/calculating-moving-average
ma <- function(x, n=7, s=1) {filter(x, rep(1 / n, n), sides=s)}

x <- covid19(country="JPN", level=2, start="2020-01-01", end=Sys.Date())
y <- subset(x, date<max(date))
z <- subset(y, administrative_area_level_2 == "Tokyo")
z$cfd <- c(z$confirmed[1], z$confirmed[-1]-z$confirmed[-length(z$confirmed)])
z$cfd <- ifelse(z$cfd<0, 0, z$cfd)

# Cori's method
res <- estimate_R(incid = z$cfd, method="parametric_si",
                  config=make_config(list(mean_si=4.8, std_si=2.3)))
z$Rt1 <- c(rep(NA, 7), as.vector(res$R["Mean(R)"][[1]]))

# Toyo-Keizai method
TK <- function(dat, gt=5, rl=7) {
    dx <- c(rep(NA, 7), dat[1:(length(dat)-7)])
    Rt <- (ma(dat)/ma(dx))^(gt/rl)
    return(Rt)
}
z$Rt2 <- TK(z$cfd)

# NIID method (29 June 2021)
NIID <- function(dat, gt=5) {
  dx <- c(rep(NA, 5), dat[1:(length(dat)-5)])
  Rt <- ma(dat)/ma(dx)
  return(Rt)
}
z$Rt3 <- NIID(z$cfd)

# Nishiura Seminar method (12 May 2020)
HNS <- function(dat) {
  dx <- c(rep(NA, 7), dat[1:(length(dat)-7)])
  Rt <- (ma(dat)/ma(dx))
  return(Rt)
}
z$Rt4 <- HNS(z$cfd)
```

東京のRtの推定（続）

明日からの東京緊急事態宣言の線を入れるコードを追加して（東京のRt推移，日本の都道府県別新規感染者数推移），実行させた結果をtweetした。
照）。

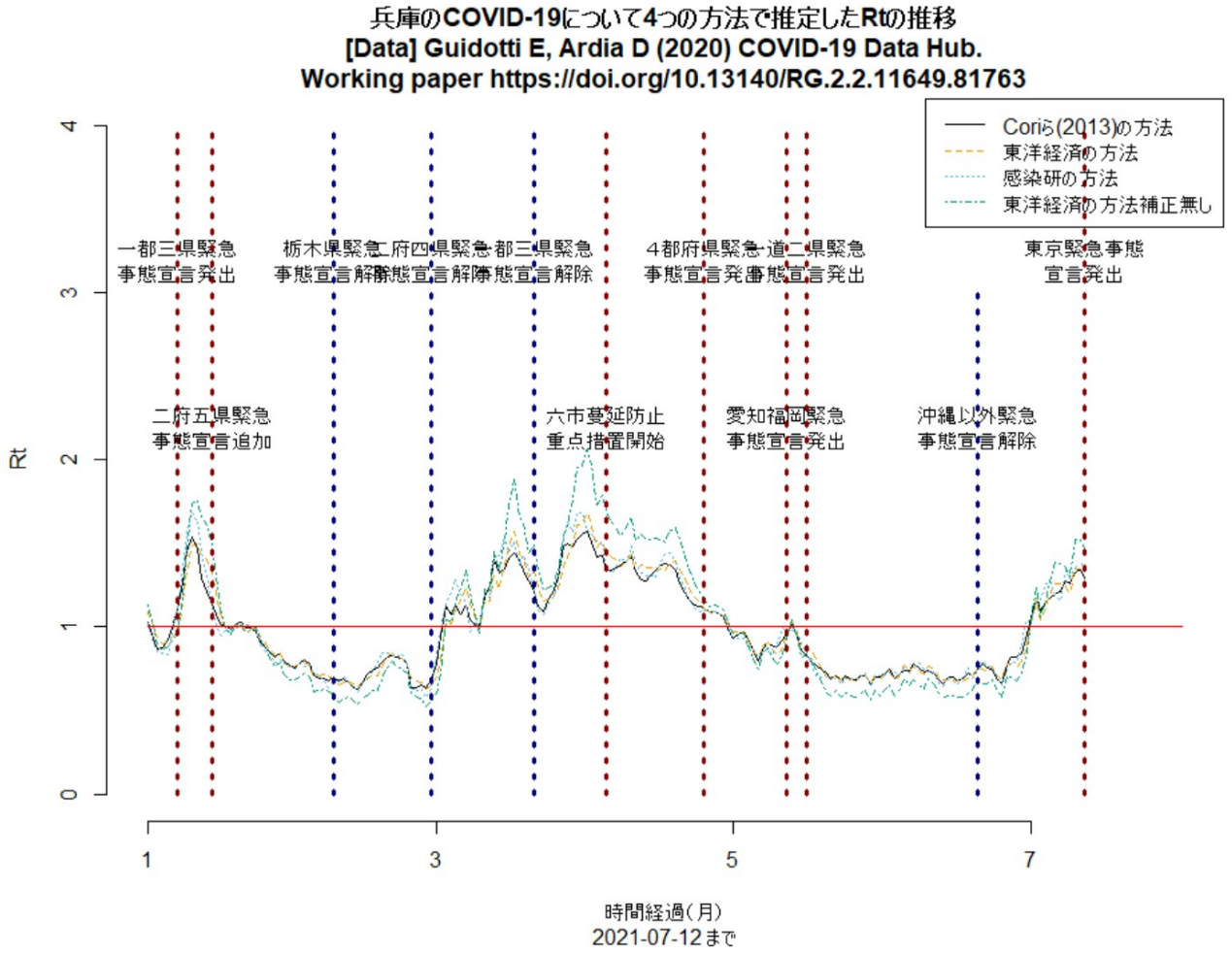


兵庫のRtの推定

https://minato.sip21c.org/libusejapanforRt-Hyogo.R

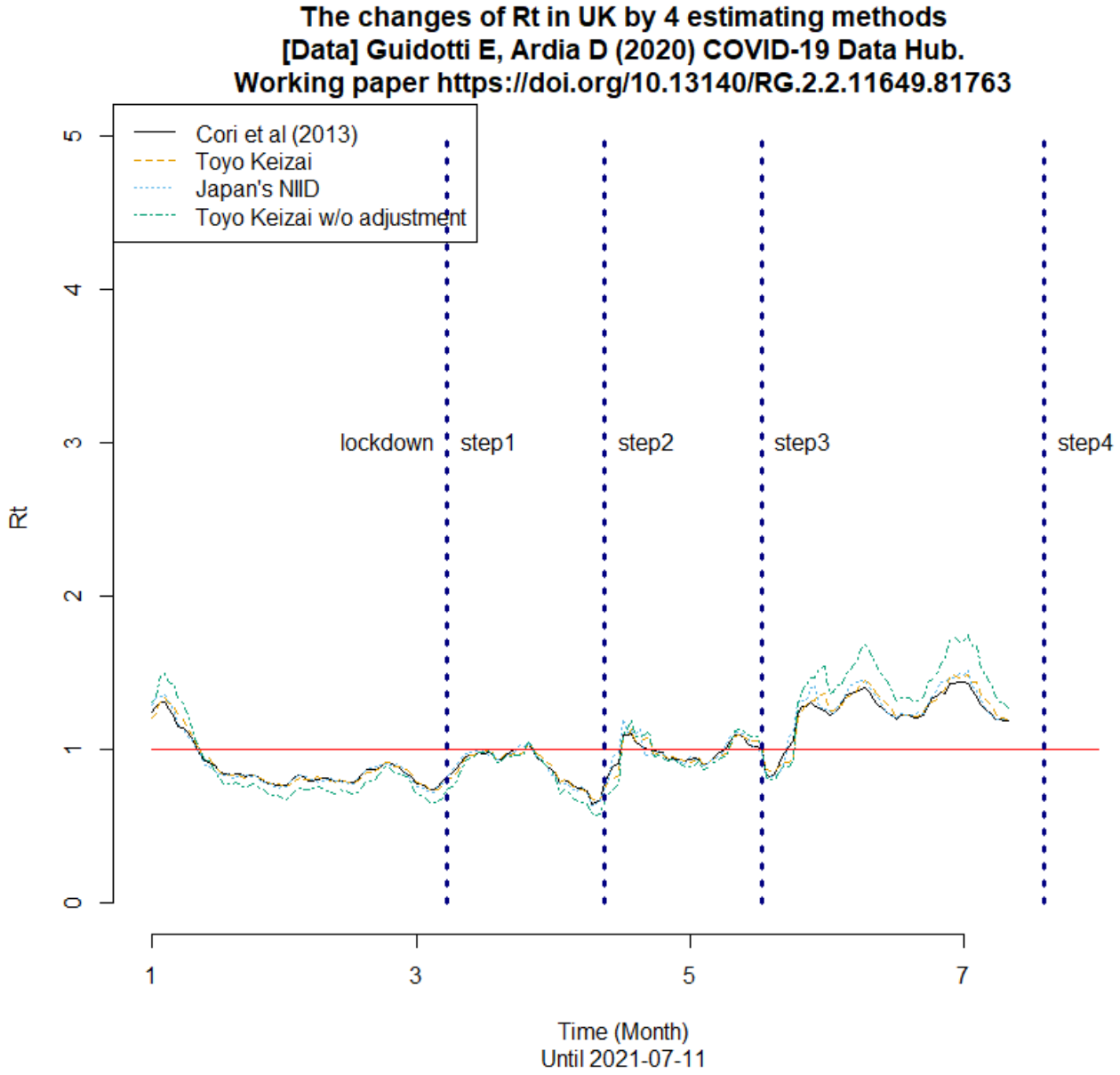
兵庫県は「まん延防止」を解除して良かったのか（2021年7月13日）

神戸市を含む兵庫県の5区域で昨日から「まん延防止」が解除され，「感染リバウンド防止」という，より緩い制限に移ったが，宣言解除以降，大阪同様にRtは上昇中で，既に1を超えている（このRコードで描ける下図参照。このコードは都道府県名を変えるだ
すぐに再び3桁／日になりそう。



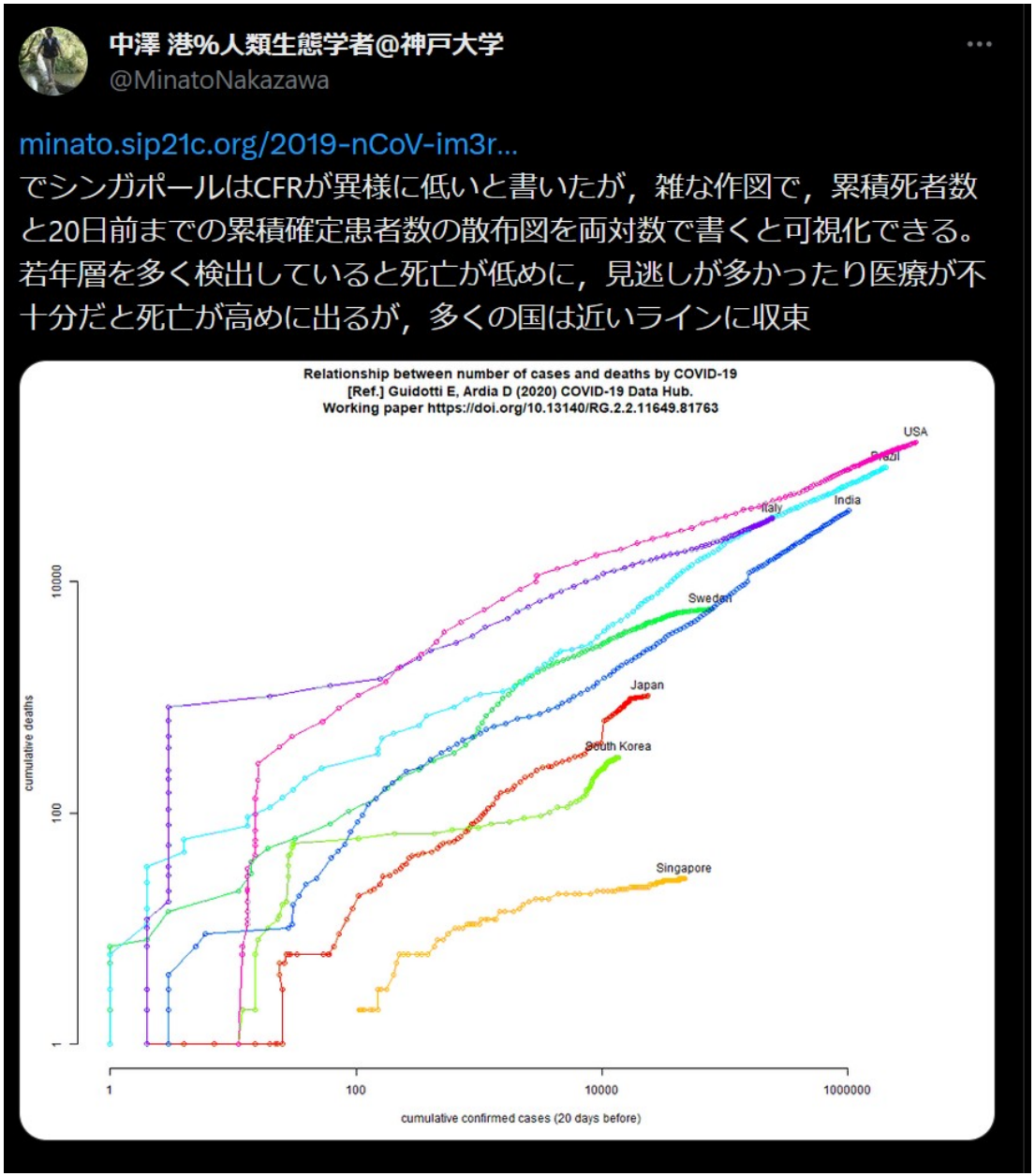
UKのRtの推定

https://minato.sip21c.org/libuseforRt-UK.R



ラフなCFR推定と作図(2020.8.7)

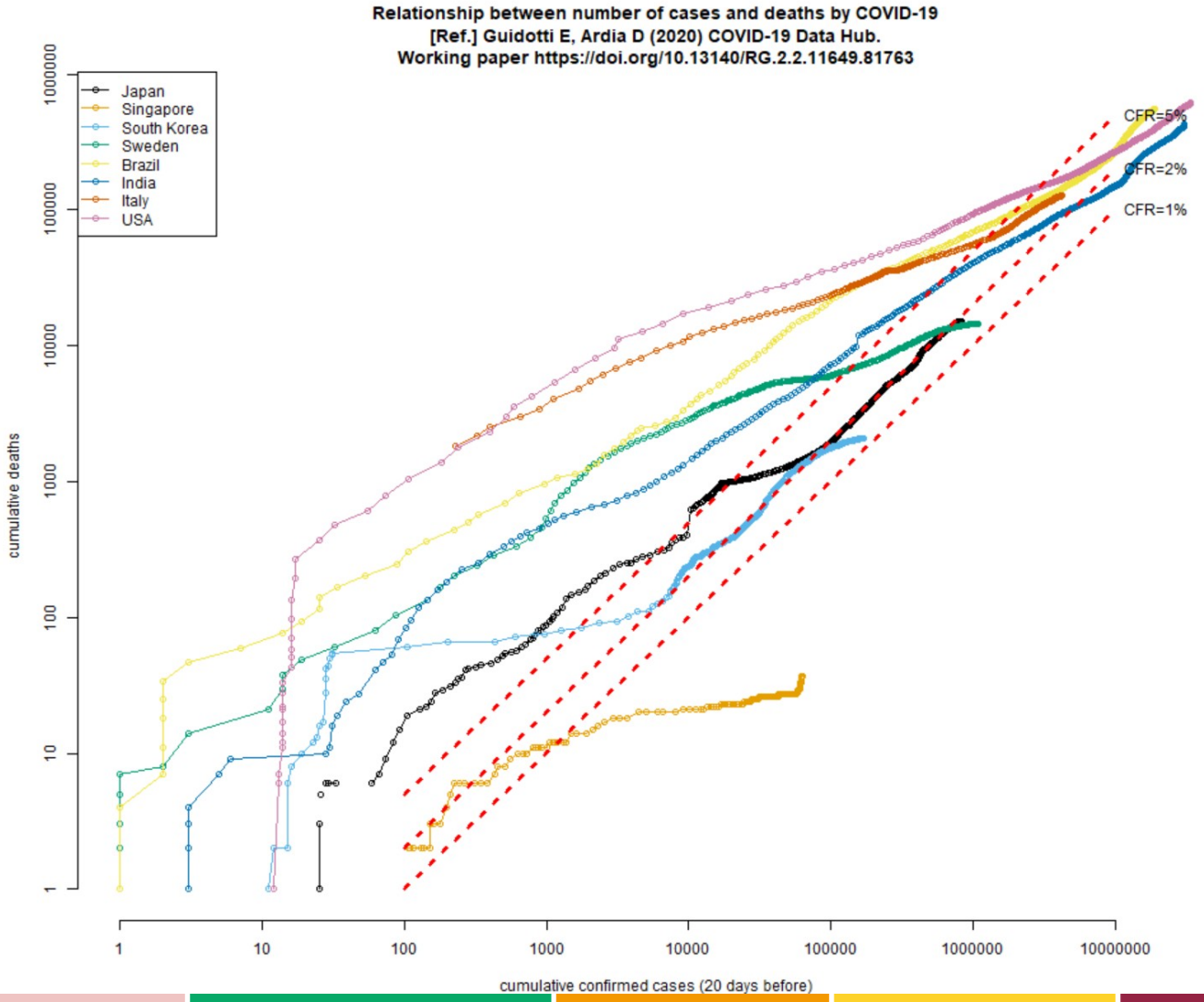
https://twitter.com/MinatoNakazawa/status/1291552240549810176



続・大雑把なCFR見積もり(2021.8.2)

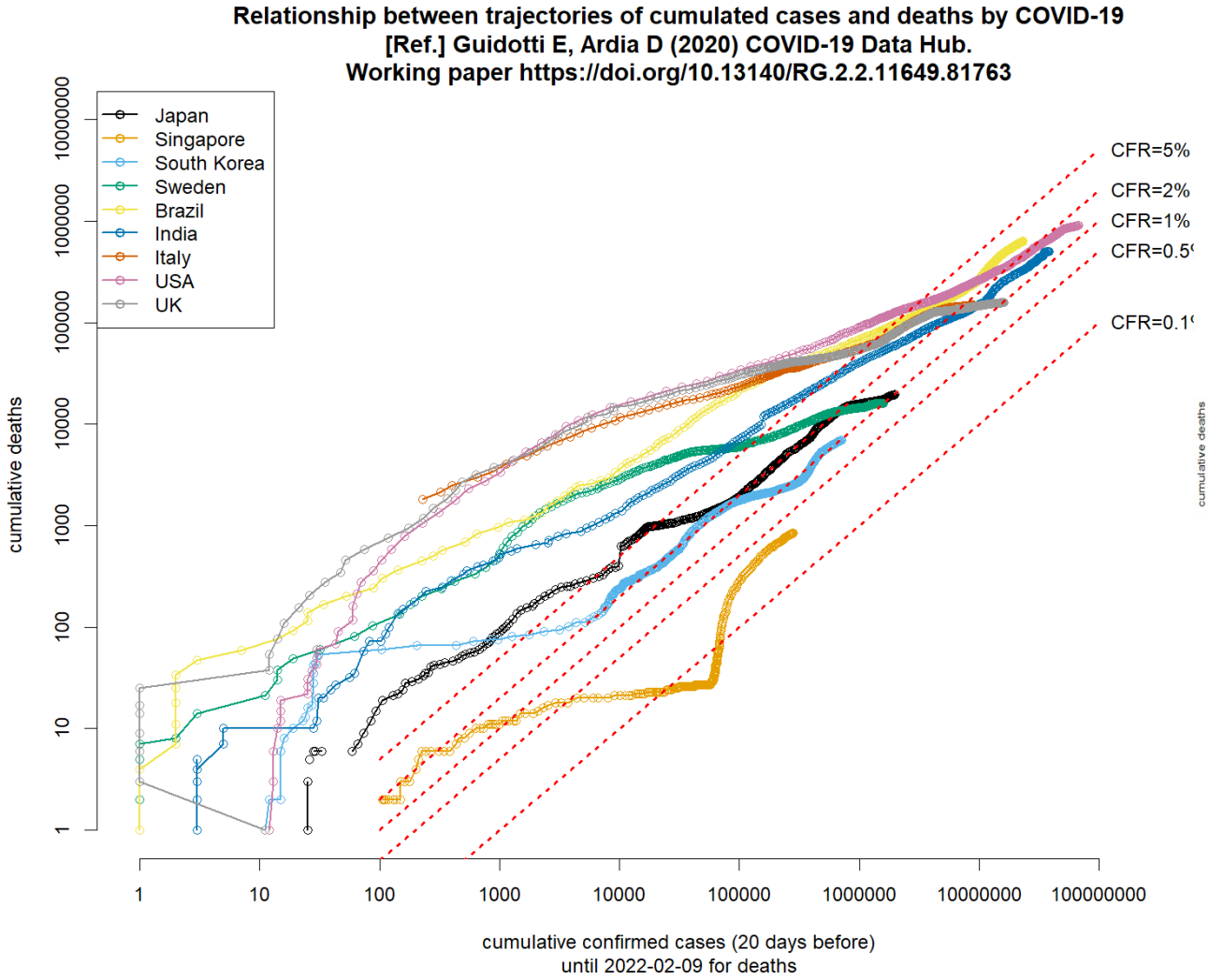
https://minato.sip21c.org/libuse-cumdeath-cumincidence.R

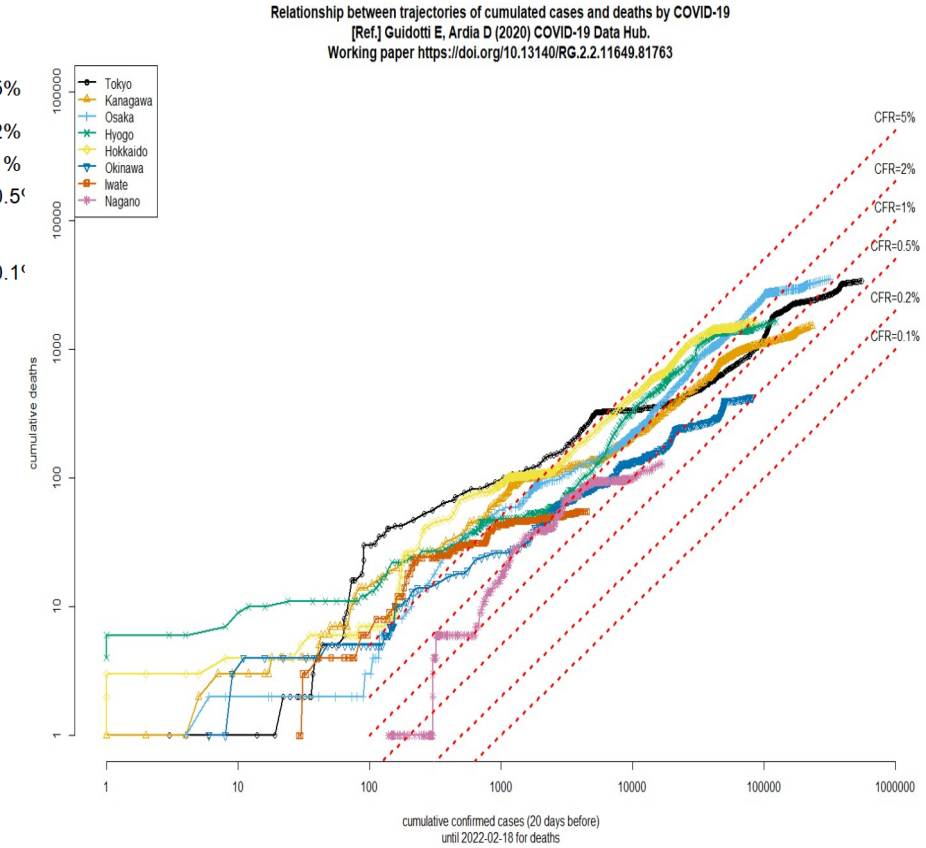
以前も示したことがあるが，このコードで累積死亡数と20日前までの累積確定感染者報告数の関係を両対数グラフにプロットすると，ごく大雑把なCFRを見積もれる（下図）。シンガポールを除けばどの国でもだいたい1%から5%くらいの幅に収束しそうで，日本はこのところ2%くらい。まだ死者が急増していなくても甘く見てはいけない。



オミクロン株後にCFR見積もり再び

https://minato.sip21c.org/libuse-cumdeath-cumincidence-for-Japan.R

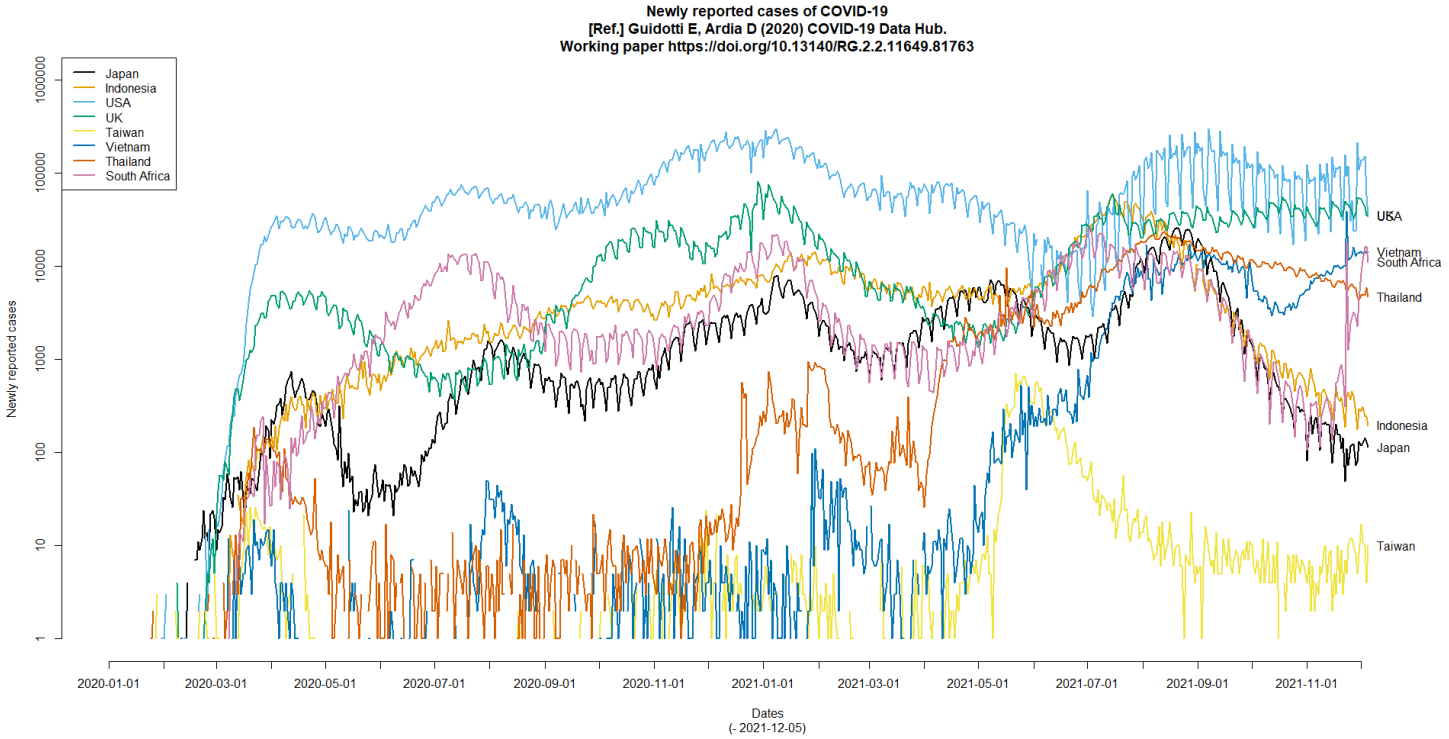




COVID19パッケージ仕様変更

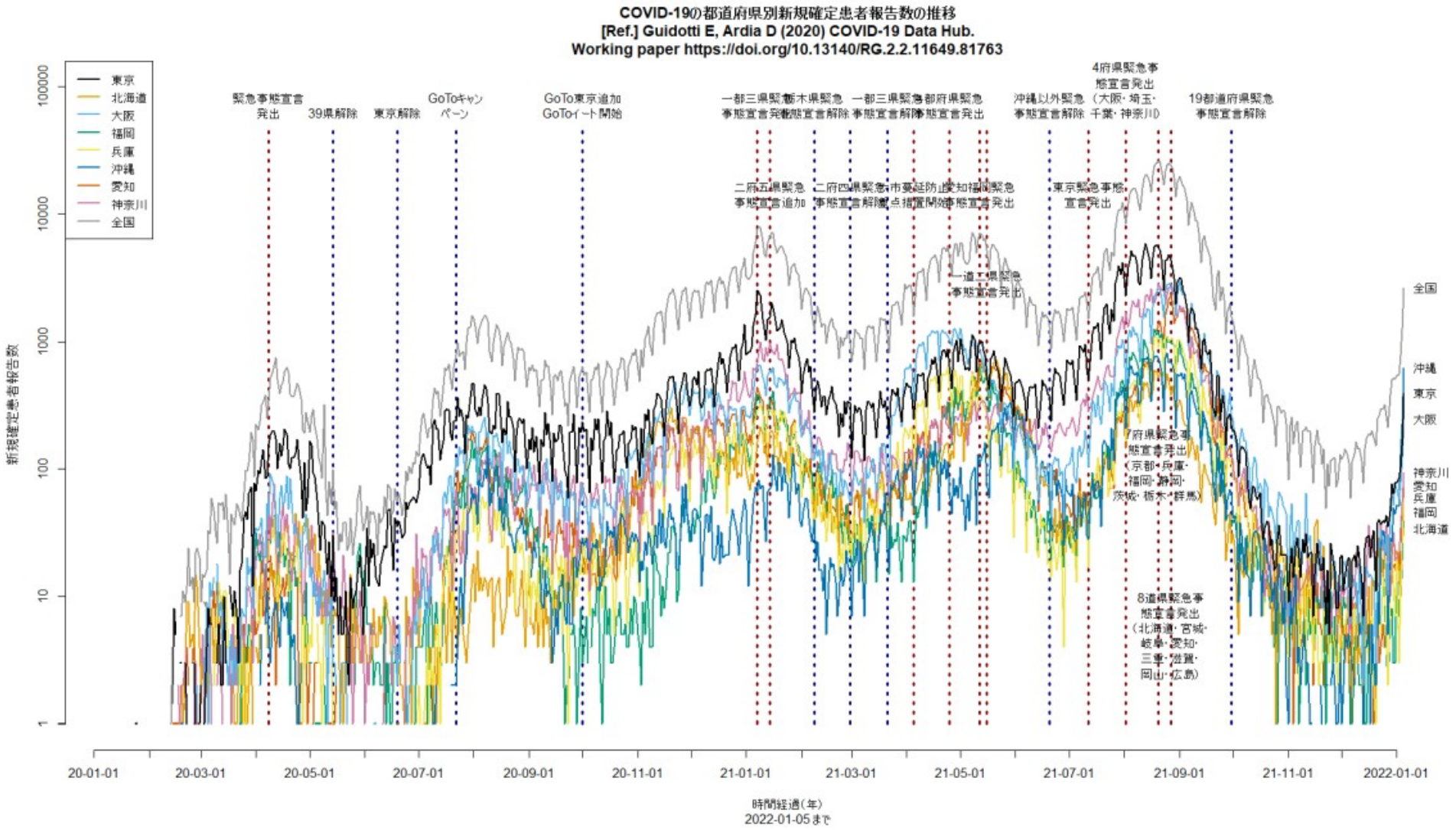
https://minato.sip21c.org/libuseworld20211208.R

- COVID19パッケージのアルファベット3文字国名コードを示す変数名がidからiso_alpha_3という変数名に変わっていたので(変数名idはまだ存在しているが,意味不明の文字列になっていた), かつて作ったCOVID-19の毎日の感染確定報告数の推移を示す折れ線グラフを描くコードが動作しなくなっていた。変数名を書き換え, 南アフリカを追加したコードに書き換えてグラフを作ってみた。8月下旬から11月上旬くらいまで日本とインドネシアと南アフリカがほぼ同じような動きなのに, 11月中旬以降の南アフリカの報告数急増が凄い。B.1.1.529変異株が南アフリカからWHOに報告されたのは11月24日だったが, それ以前から増えていたということだろう。(2021.12.8)

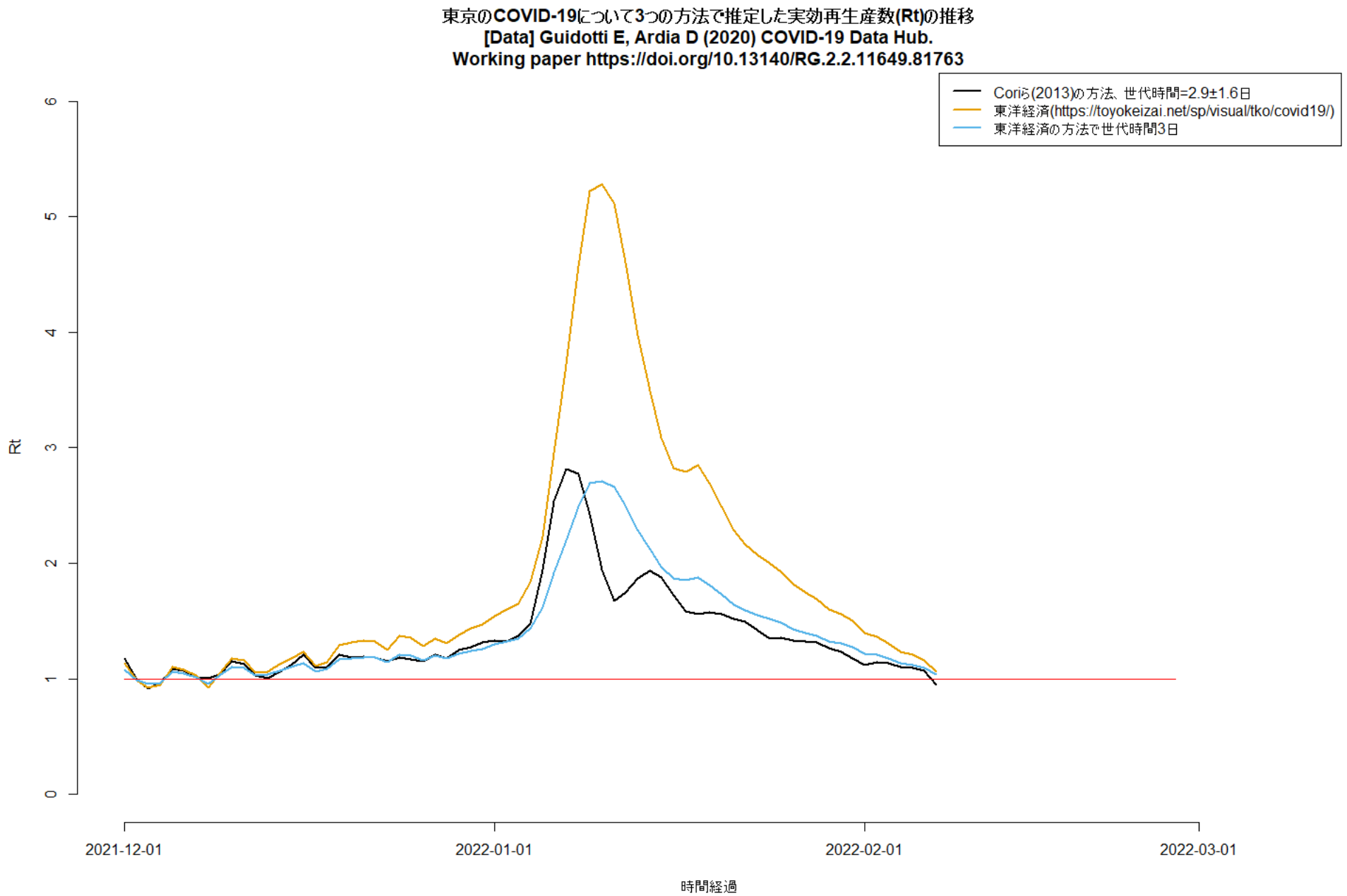


オミクロン株急増

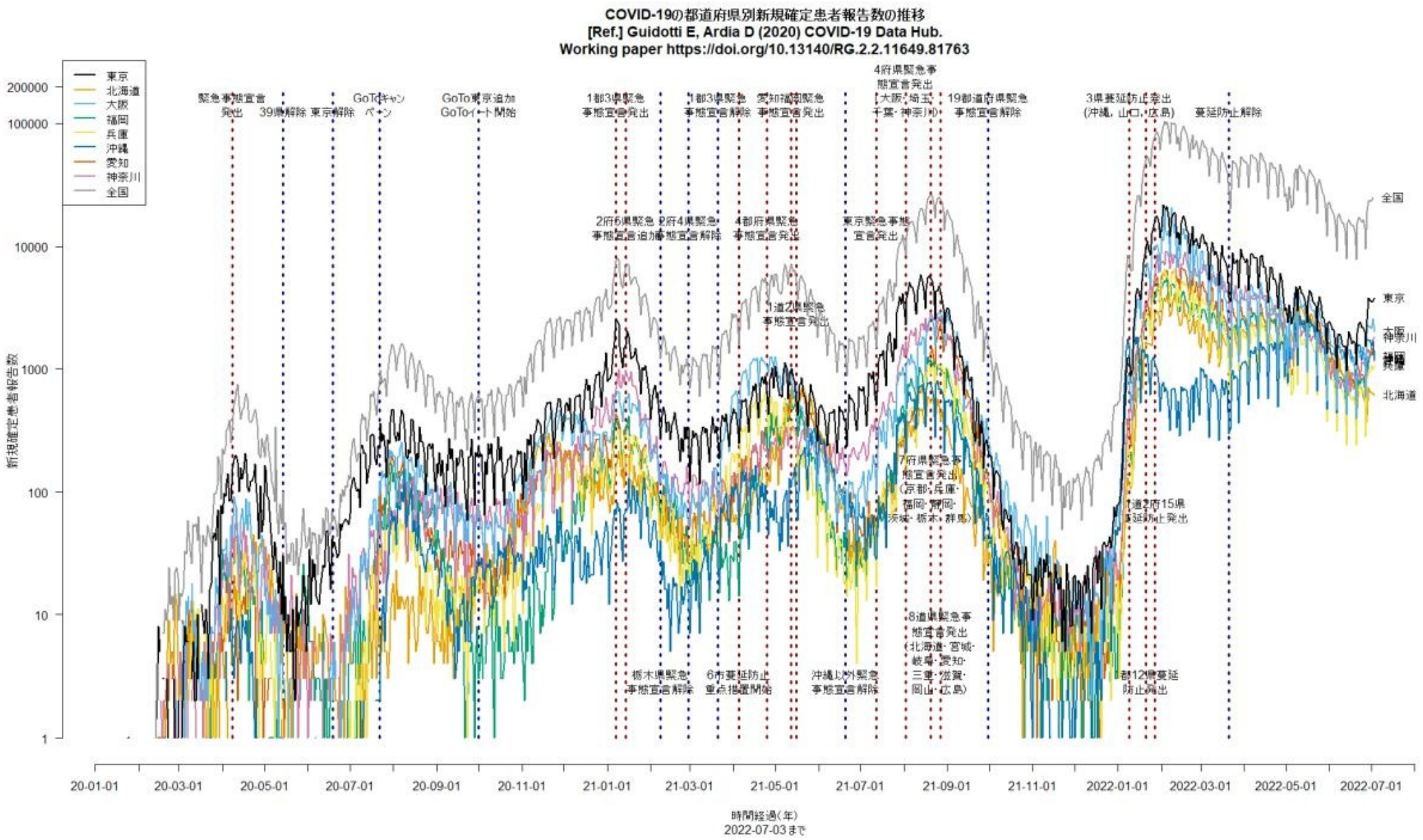
https://minato.sip21c.org/libusejapant-20220105.R



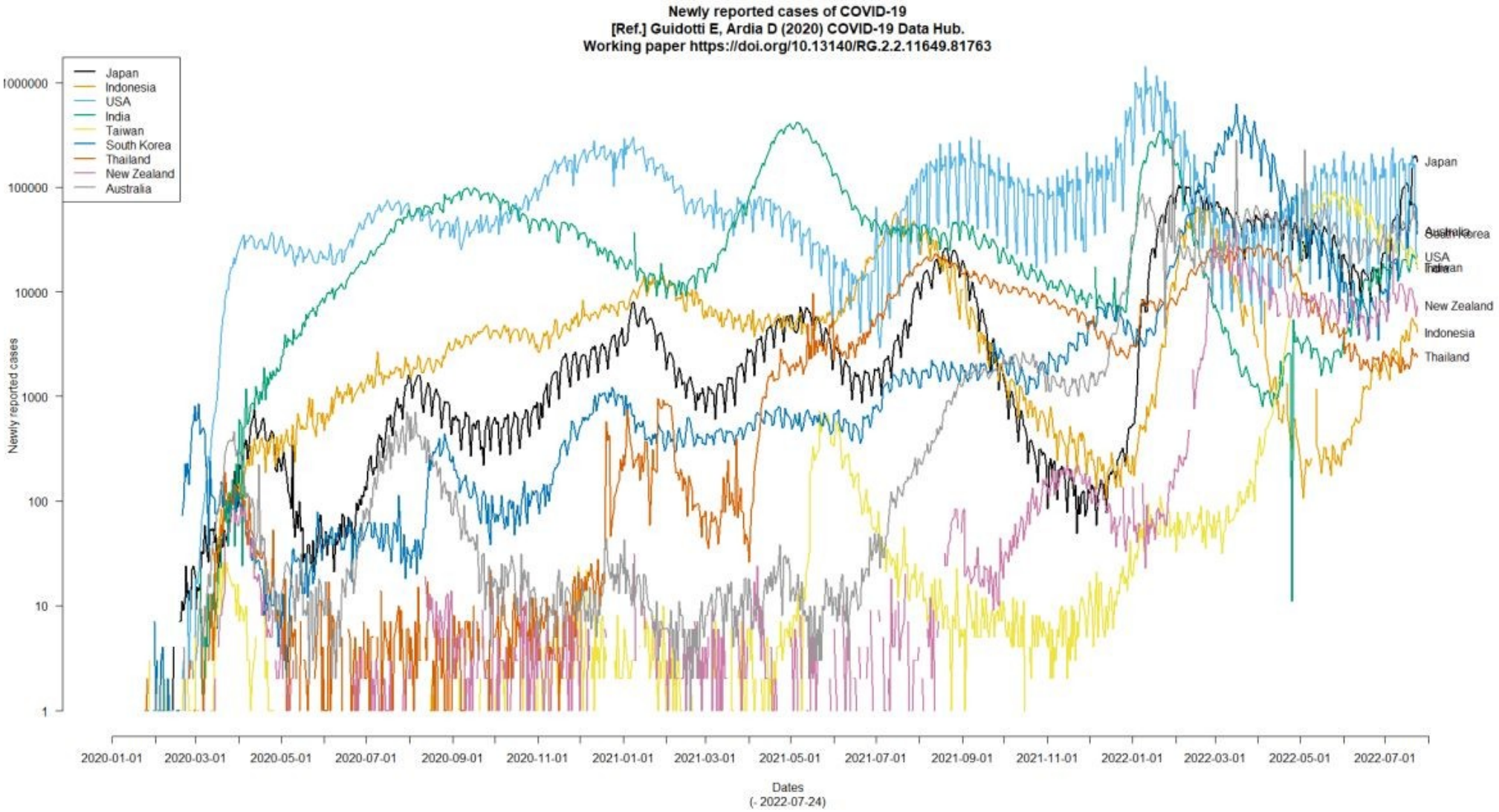
オミクロン株は世代時間が短い



常在化(2022.7.5)

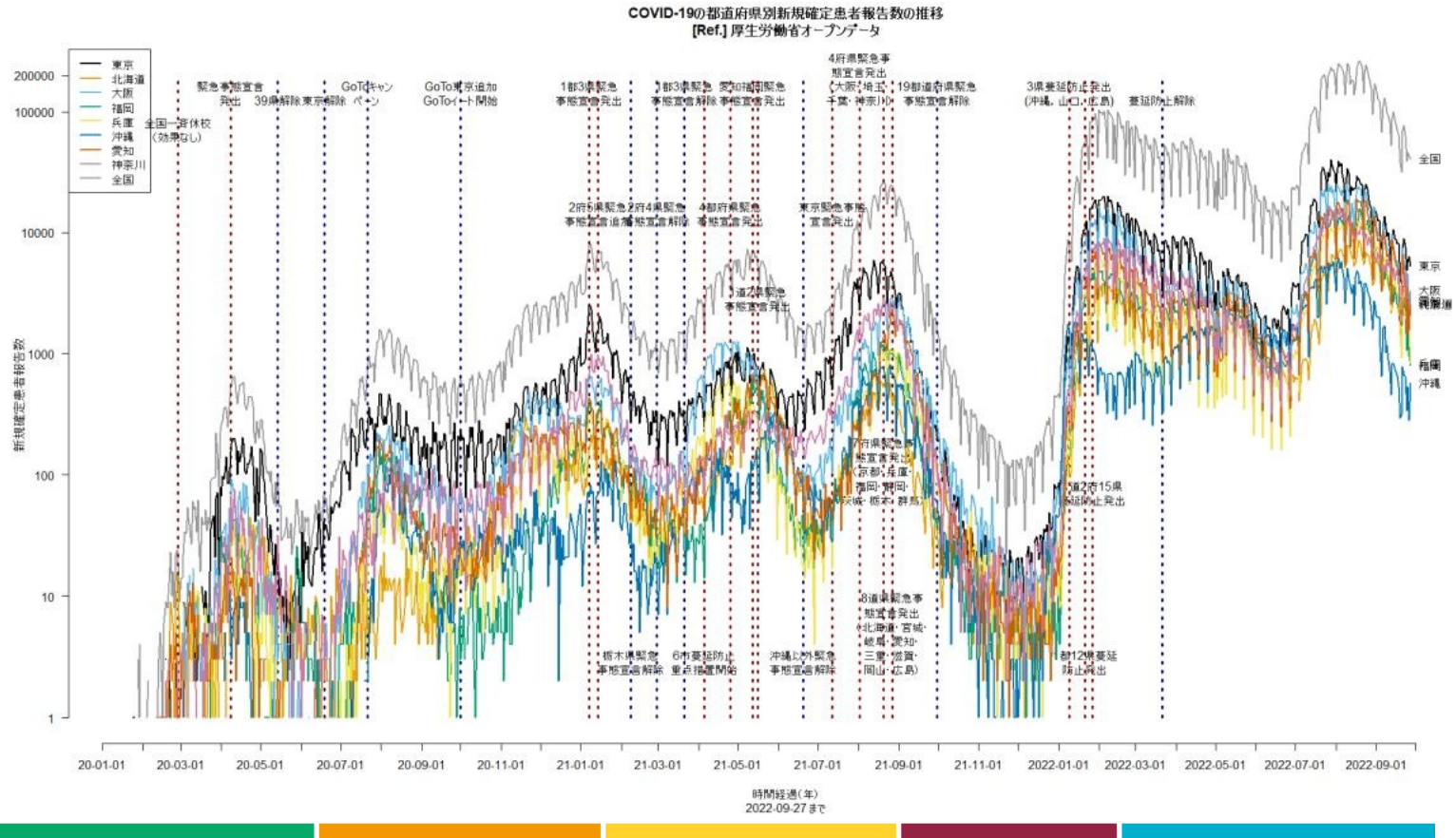

COVID-19の都道府県別新規確定患者報告数の推移
[Ref.] Guidotti E, Ardia D (2020) COVID-19 Data Hub.
Working paper https://doi.org/10.13140/RG.2.2.11649.81763
東京
北海道
大阪
福岡
兵庫
沖縄
愛知
神奈川
全国
新規確定患者報告数
200000
100000
10000
1000
100
10
1
緊急事態宣言発出
39県解除
東京解除
GoTo キャンペーン
GoTo 東京追加
GoTo イート開始
1都3県緊急事態宣言発出
1都3県緊急事態宣言解除
愛知福岡緊急事態宣言発出
4府県緊急事態宣言発出（大阪・埼玉・千葉・神奈川）
19都道府県緊急事態宣言解除
3県蔓延防止要出（沖縄，山口，広島）
蔓延防止解除
2府6県緊急事態宣言追加
2府4県緊急事態宣言解除
4都府県緊急事態宣言発出
東京緊急事態宣言発出
1道2県緊急事態宣言発出
7府県緊急事態宣言発出（京都・兵庫・福岡・茨城・栃木・群馬）
1道2府15県蔓延防止発出
8道県緊急事態宣言発出（北海道・宮城・岐阜・愛知・三重・滋賀・岡山・広島）
8都12県蔓延防止発出
栃木県緊急事態宣言解除
6市蔓延防止重点措置開始
沖縄以外緊急事態宣言解除
全国
東京
大阪
神奈川
福岡
兵庫
北海道
20-01-01
20-03-01
20-05-01
20-07-01
20-09-01
20-11-01
21-01-01
21-03-01
21-05-01
21-07-01
21-09-01
21-11-01
2022-01-01
2022-03-01
2022-05-01
2022-07-01
時間経過（年）
2022-07-03まで

データ更新が止まり始める
（2022.8、COVID19パッケージからUKデータが不完全に）



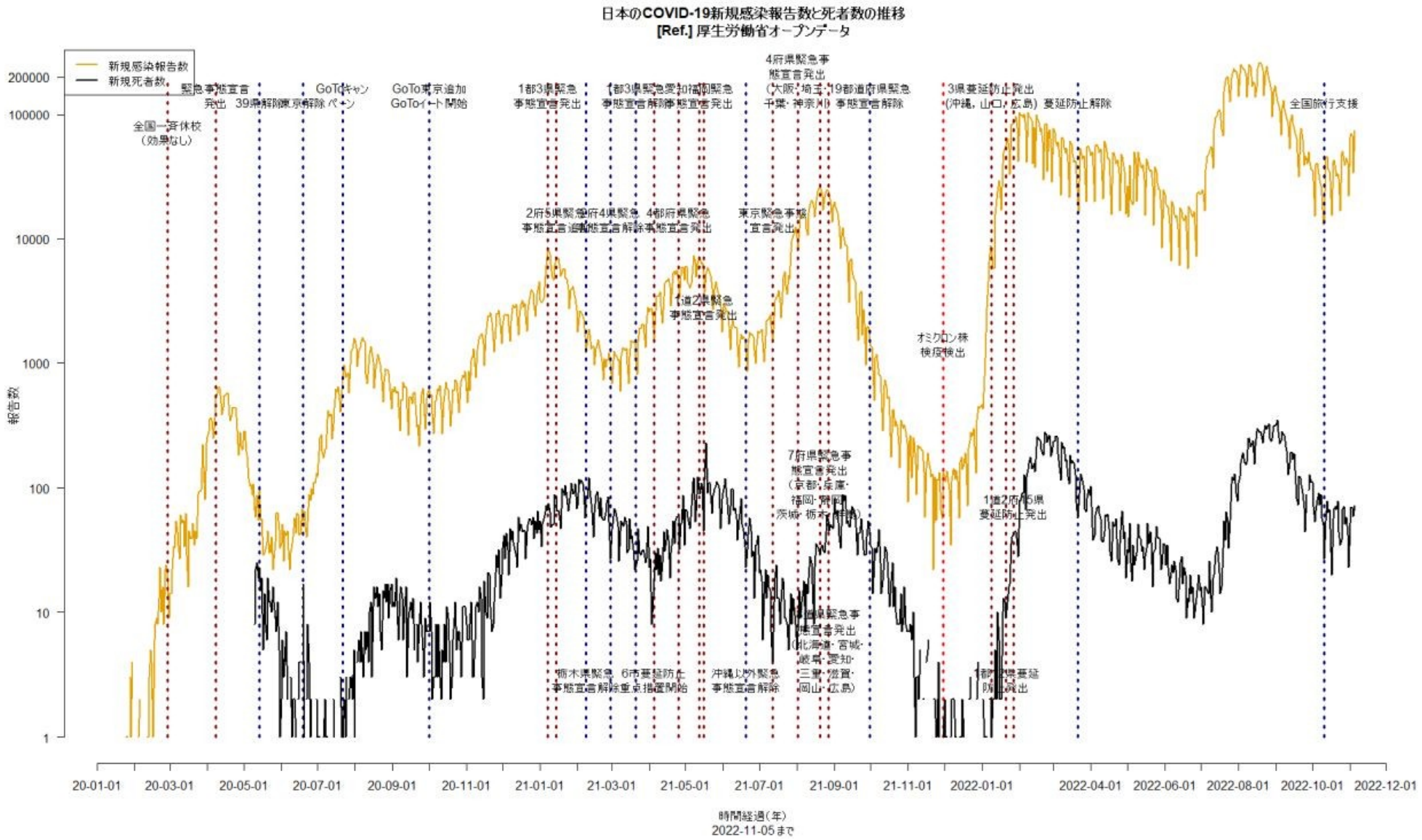
新規感染者数推移片対数グラフのデータソース切り替え
https://minato.sip21c.org/japanpref-from-mhlwopendata.R

- 厚労省令で全数報告が緩められ、データの信頼性低下→メタコビ開始
- RのCOVID19パッケージの日本の都道府県データの元のcsvファイルが2022年9月5日が最終更新→厚労省のデータがそれ以降取れなくなっている→厚労省オープンデータ自体は更新されているので、COVID19パッケージで厚労省データが取れなくなった原因は不明
- COVID-19の新規確定感染者数の推移描画Rコードを厚労省オープンデータを直接読むバージョンに書き換え(2022.9.27)



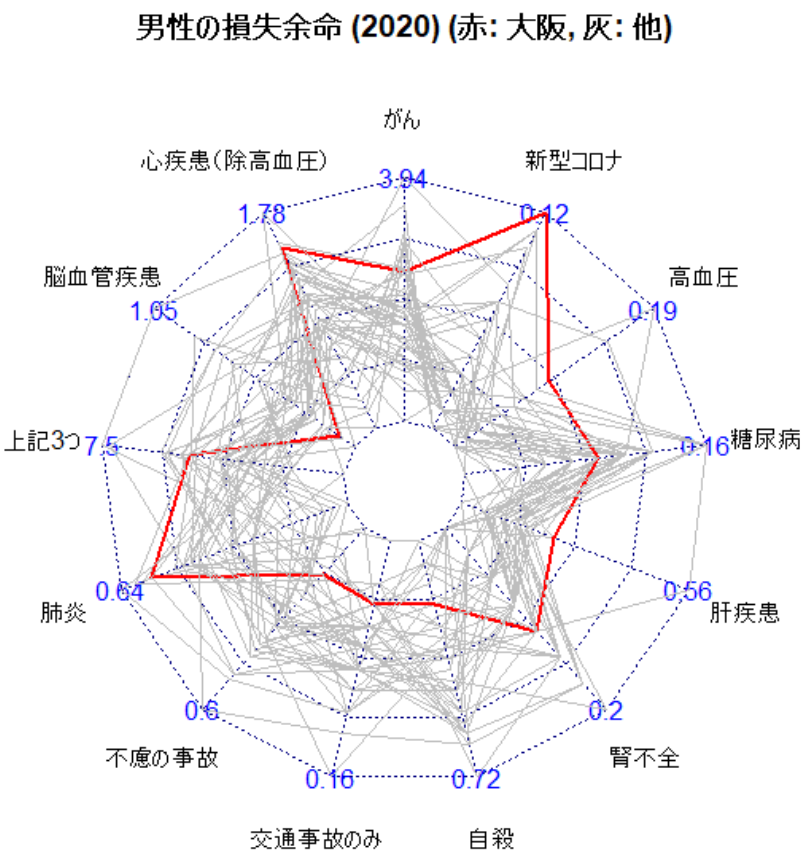
新規感染者数と死者数の推移を重ね描き

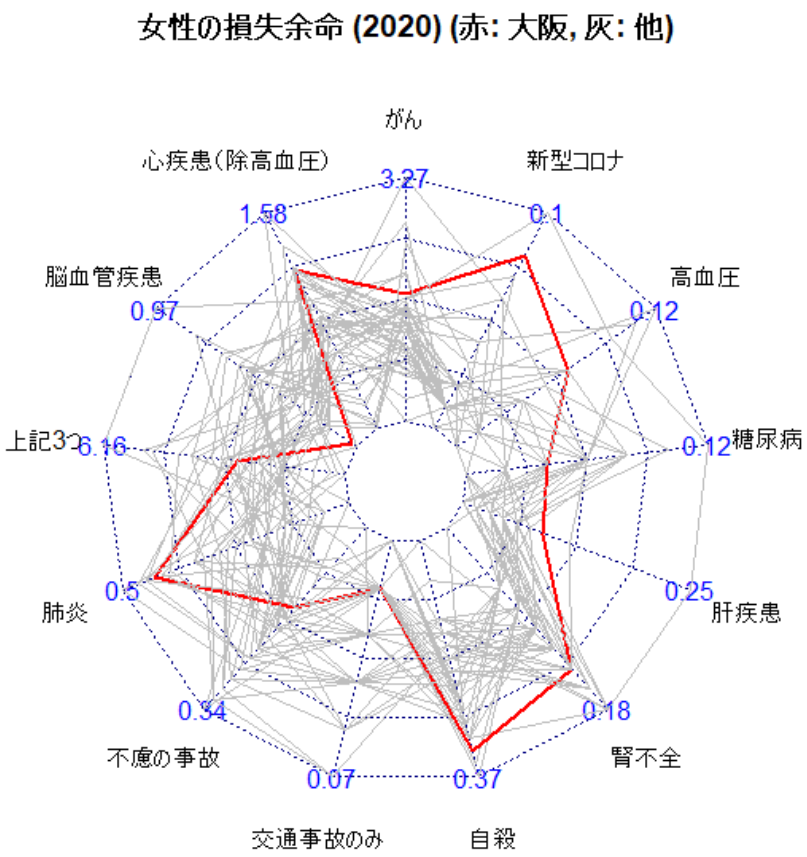
https://minato.sip21c.org/japanncnd-from-mhlwopendata.R



都道府県別の死因別損失余命レーダーチャートで見える2020年大阪の医療崩壊

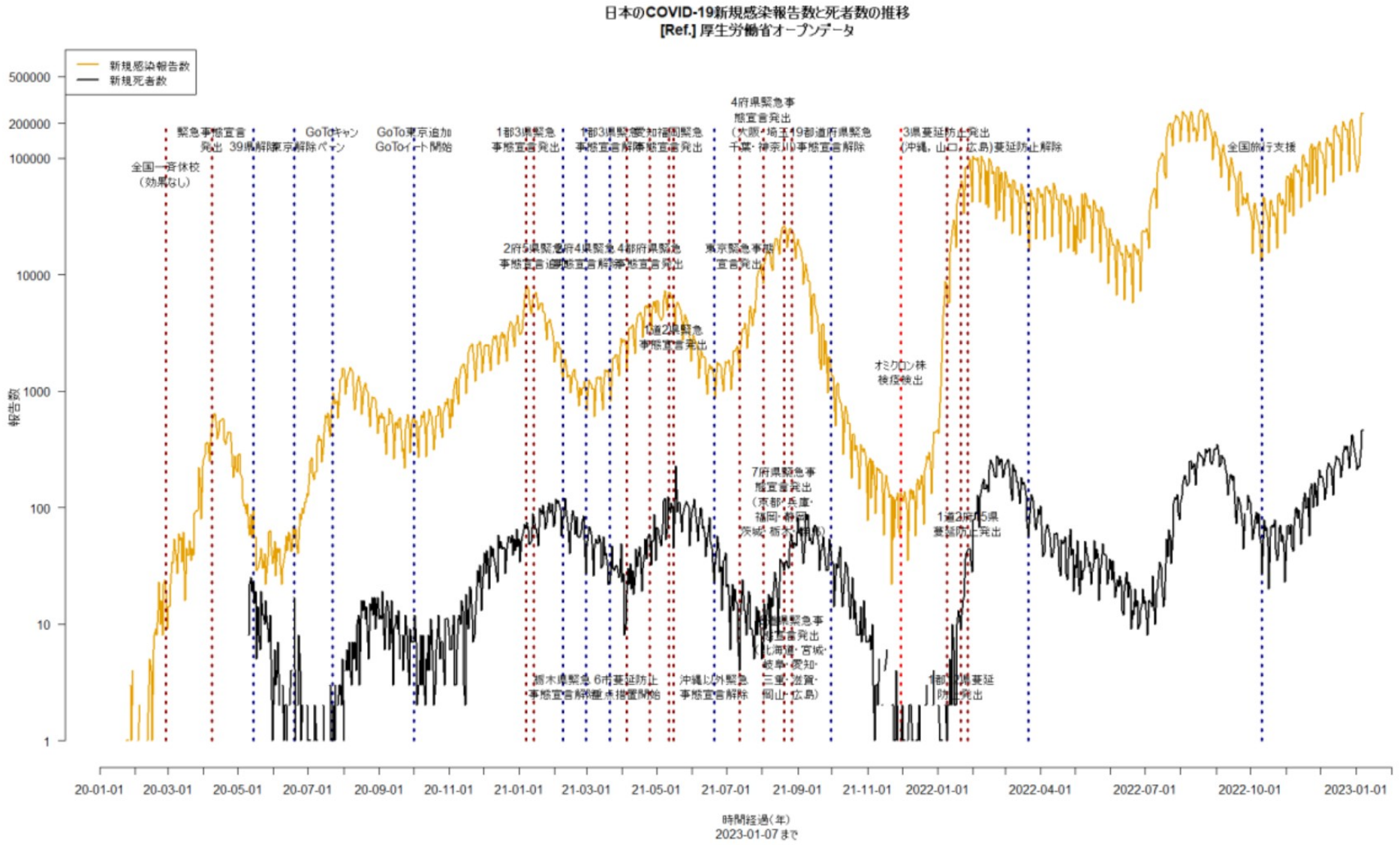
https://minato.sip21c.org/demography/osakaYLL2020.R





新規感染者数が過少報告に

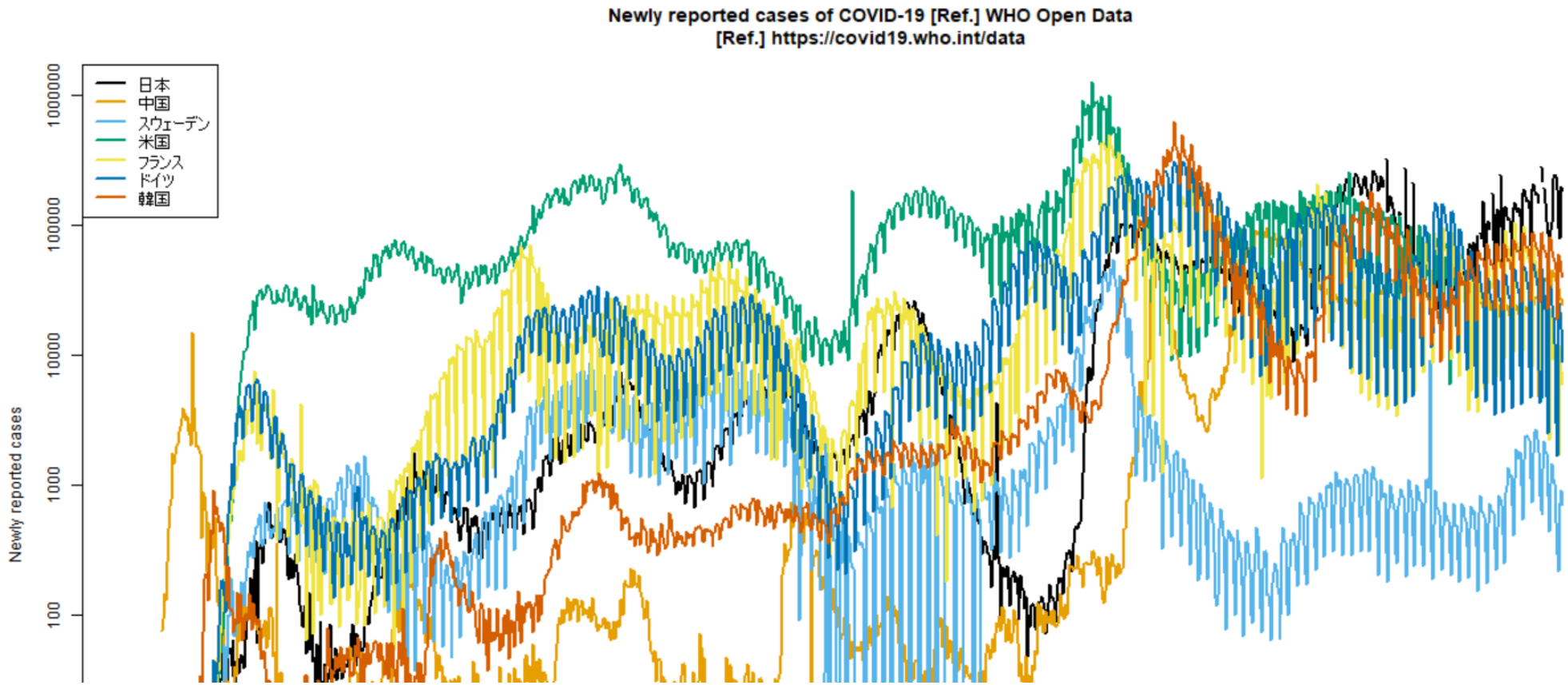
新規死者数が過去最高を更新したというニュースを見たので、グラフを描いてみた。病原性が強いというよりは新規感染者数が過少報告なのだろう。(2023.1.7)



国際比較はWHOオープンデータで

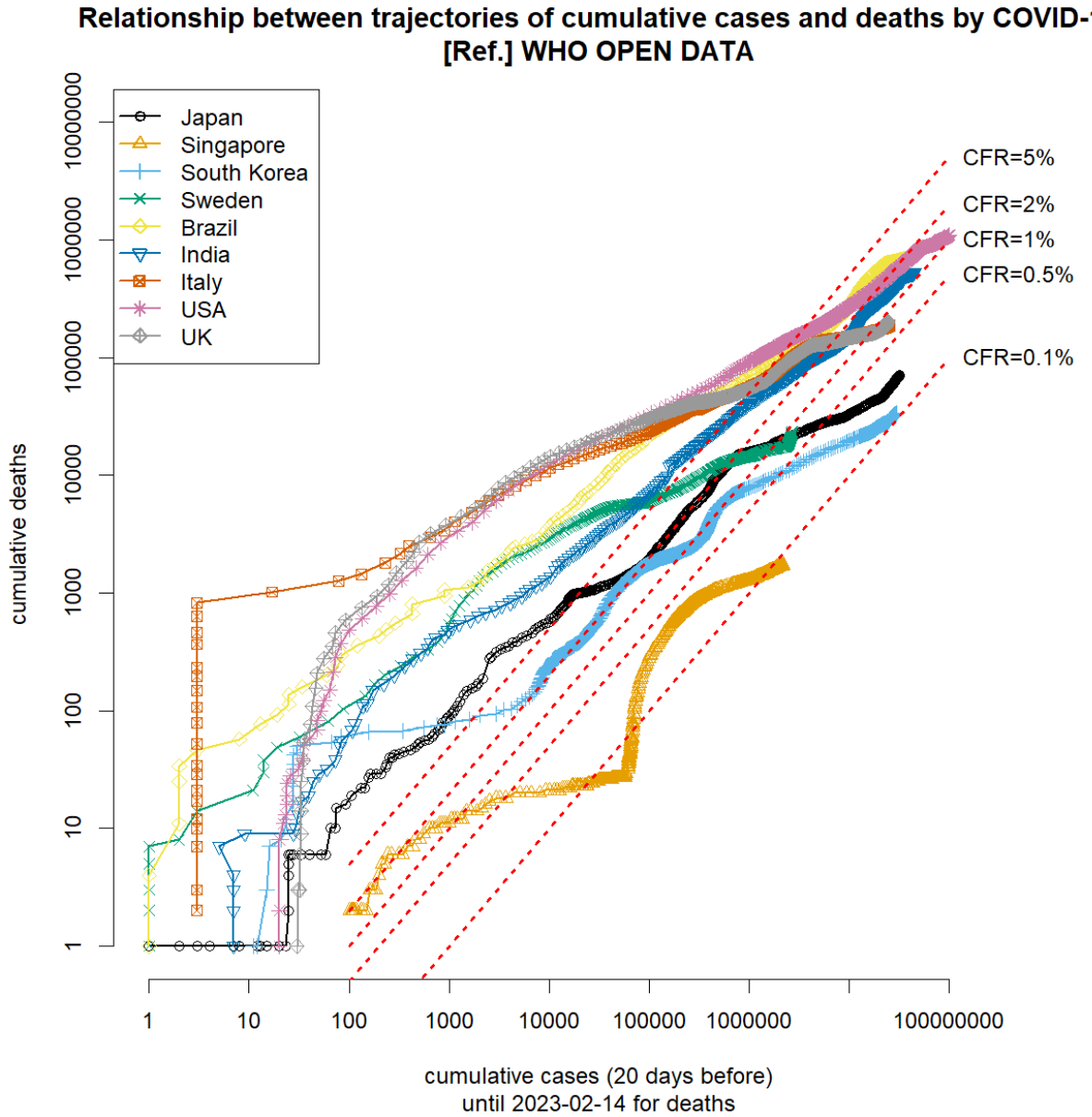
https://minato.sip21c.org/covid19-world-whodata.R

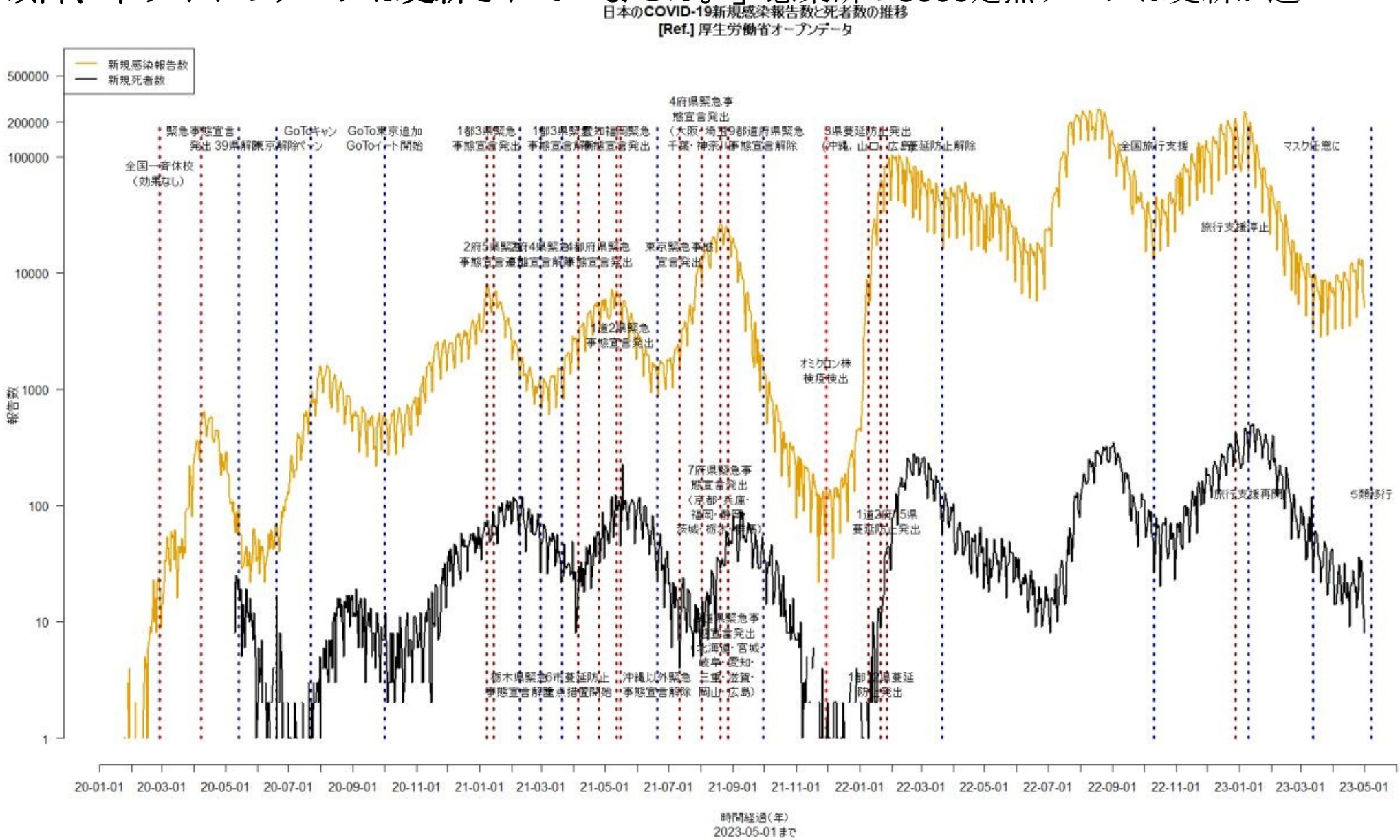
COVID-19 Data Hubの更新が止まってから、COVID19パッケージを使うコードが使えなくなって、COVID-19の感染者数や死者数の国際比較はOurWorld in Dataで作図していたが、IHR2005に基づいて加盟国から毎日の感染者数と死者数がWHOに報告されるという仕組みは生きていて（検査が十分にされていないなどの理由で報告数自体が過少報告になっている可能性は高いが、この定義に基づく一律の症例定義で報告されているはず）、WHOのサイトからCSVをダウンロードできるので、そのデータを使うコードに書き換えて（変数名が違うし、国名コードがCOVID19パッケージではISO3桁コードだったのが、WHO Open DataではISO2桁コードになっていたため、その辺りを直した）作図してみた。やはり2022年2月以降は全世界でエンデミック化して「大炎上」状態と見做さざるをえない。(2023.1.15)



WHOオープンデータを使う簡易CFR

https://minato.sip21c.org/cumdeath-cumincidence-20230215.R



最後のグラフ作成(2023.5.2)
→以降はモデルナの4000定点に頼るしかない

厚労省オープンデータ**「新型コロナウイルス感染症の感染症法上の位置づけ変更に伴い、2023年5月7日分のデータが、本サイトの最終集計値となっております。最終更新日（2023年5月9日16時）以降、本サイトのデータは更新されていません。」**感染研の5000定点データは更新が遅い