

# ビジネスに役立つデータ分析 (入門編)

瓜生真也 (徳島大学デザイン型AI教育研究センター)

# 諸注意

資料置き場: [https://github.com/uribo/cue2022aw\\_r104](https://github.com/uribo/cue2022aw_r104)

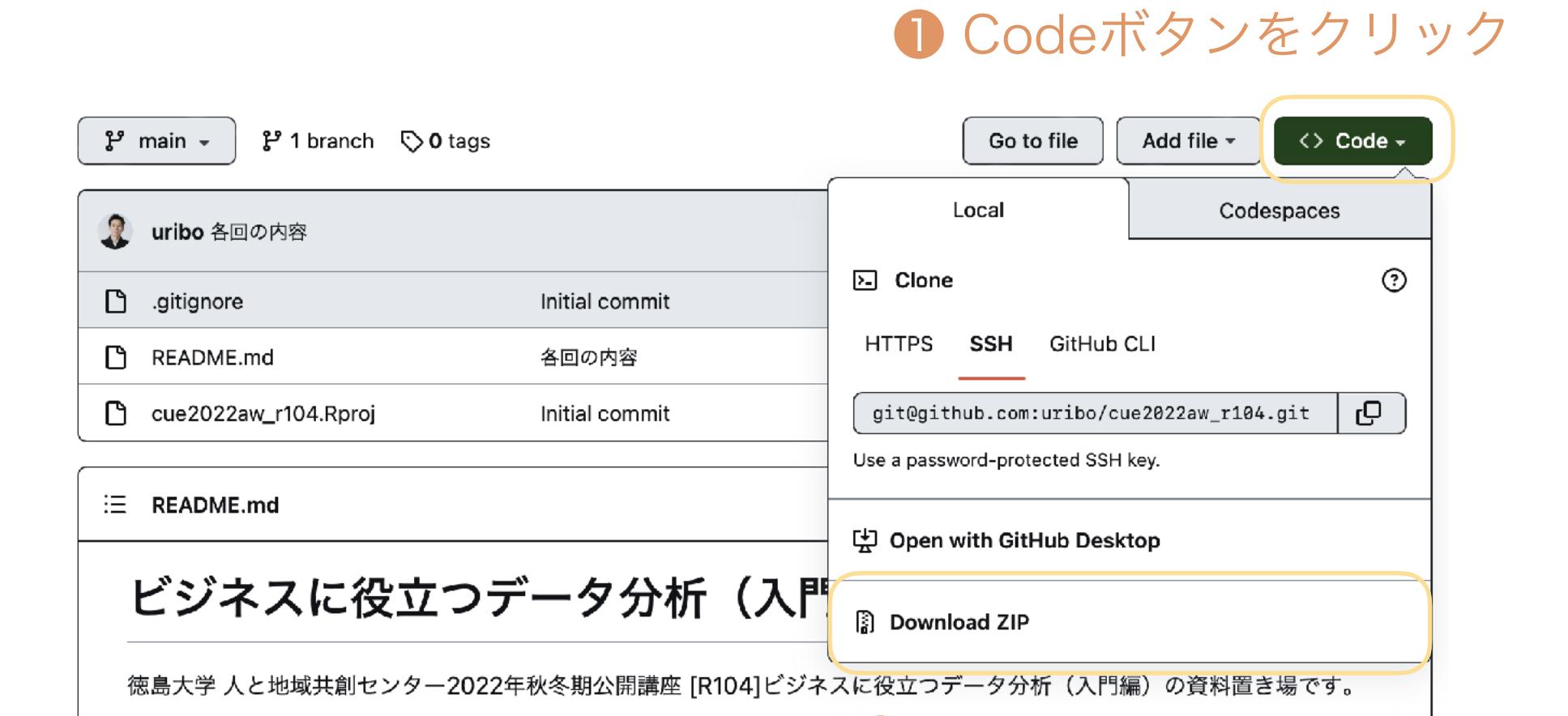
投影するプレゼンテーション、ソースコードを置いています  
(来週分は来週更新)

## Rコードと実行環境

```
# 気温とアイスの相関係数を求める
cor(df_icecream_temperature$temperature_average_c,
     df_icecream_temperature$value)
#> [1] 0.9144466
```

[https://mybinder.org/v2/gh/uribo/cue2022aw\\_r104/main?urlpath=rstudio](https://mybinder.org/v2/gh/uribo/cue2022aw_r104/main?urlpath=rstudio)

から必要なパッケージ、データ、ソースコードを含んだRStudioが起動します



① Codeボタンをクリック

② Download ZIP をクリック

# 第五週: さまざまなデータの扱い

瓜生真也（徳島大学デザイン型AI教育研究センター）

# 講座の内容

## 第一週

情報化社会におけるデータの利活用

## 第二週

データを扱うためのリテラシー

## 第三週

データサイエンス入門 (1)

## 第四週

データサイエンス入門 (2)

## 第五週

さまざまなデータの扱い  
時系列データ  
地理空間データ

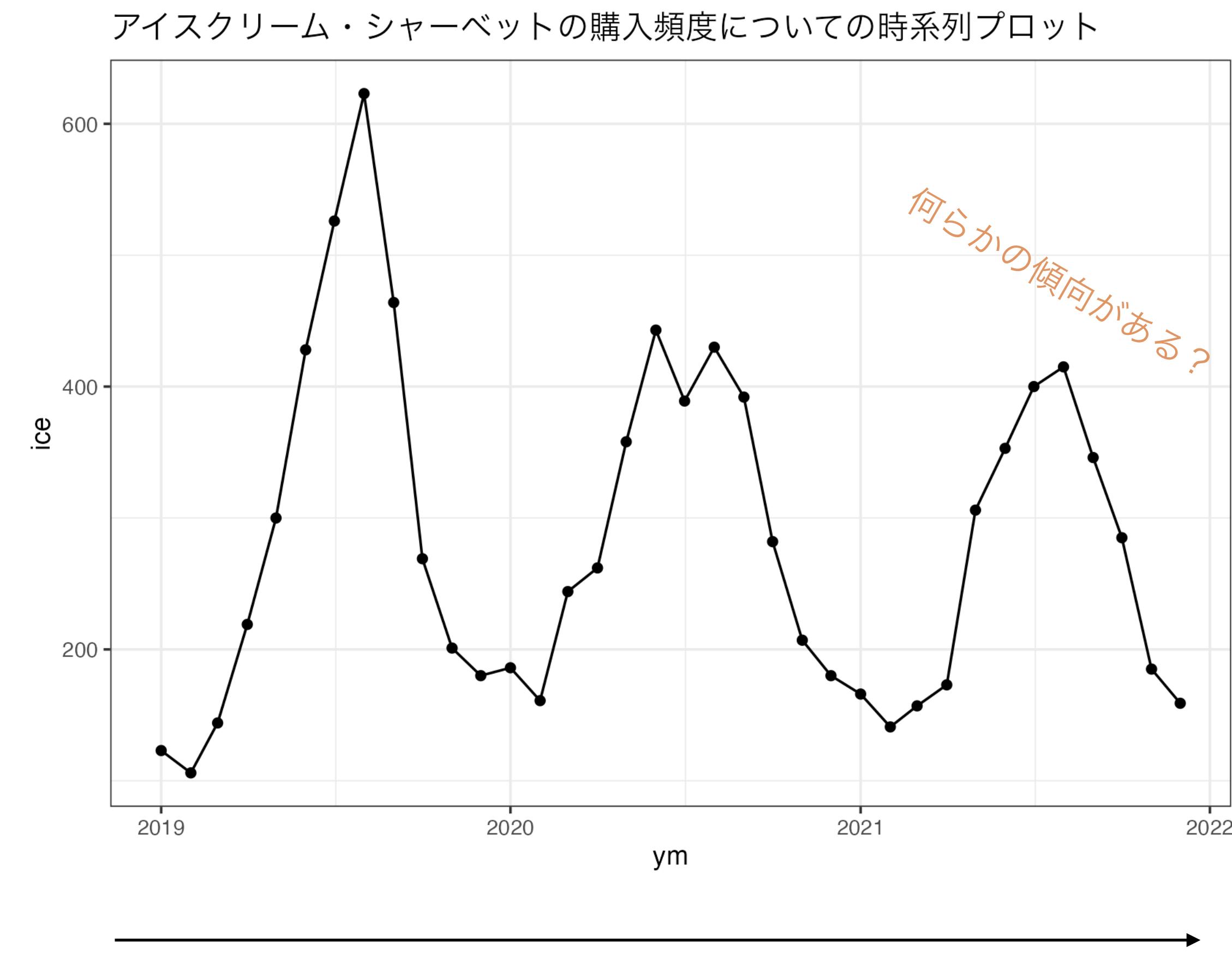
# 時系列データ

# 時間とともに観測されるデータ

何はともあれ可視化してみる（時系列プロット）

データを時系列の順にならべ、値を示す点を線で結ぶ

年月	購入頻度
201901	123
201902	106
201903	144
201904	219
201905	300
201906	428
201907	526
201908	623
201909	464



# 観測頻度

## 規則的なもの

	国勢調査（5年に1度）
年次	地価
半期	
四半期	財務、業績情報
月次	家計調査
週次	給油所小売価格
日次	株価
時間	セッション数

間隔…短い（高頻度）→長い（低頻度）

## 不規則な間隔

地震
動物の行動
ラジオ番組中の音楽再生

# 暦年と年度の違いに注意

## 暦年

1月から12月までを1年間とする

## 年度 (Fiscal Year: FY)

会計年度開始の月から12ヶ月を1年間とする

日本の場合、4月から翌年3月までを1年間とする

## 半期ごとに観測されるデータの場合…

暦年 1~6月 (上半期) 7~12月 (下半期)

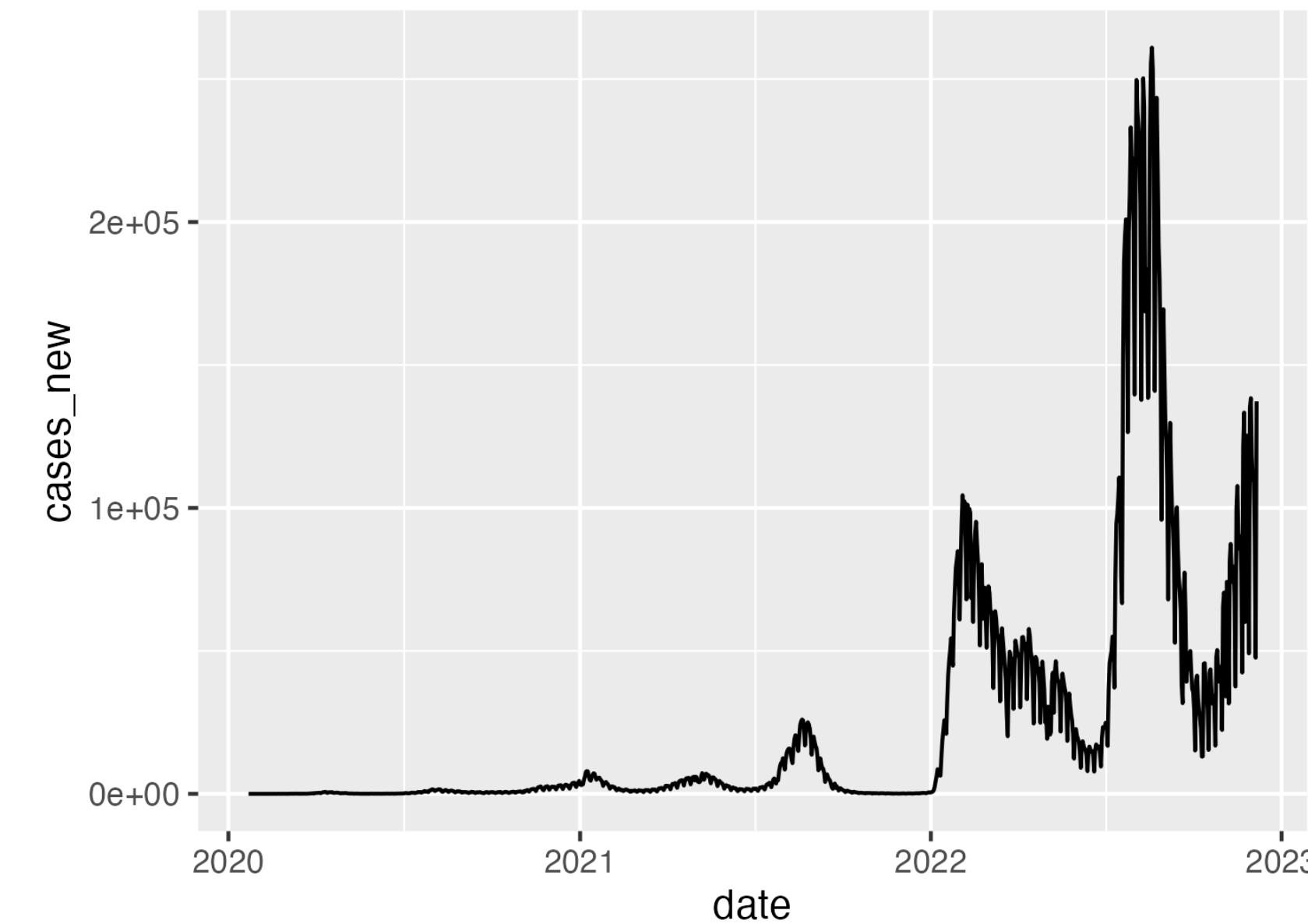
年度 4~9月 (上半期) 10~翌3月 (下半期)

# 同じデータでも観測頻度の違いで見た目が異なる

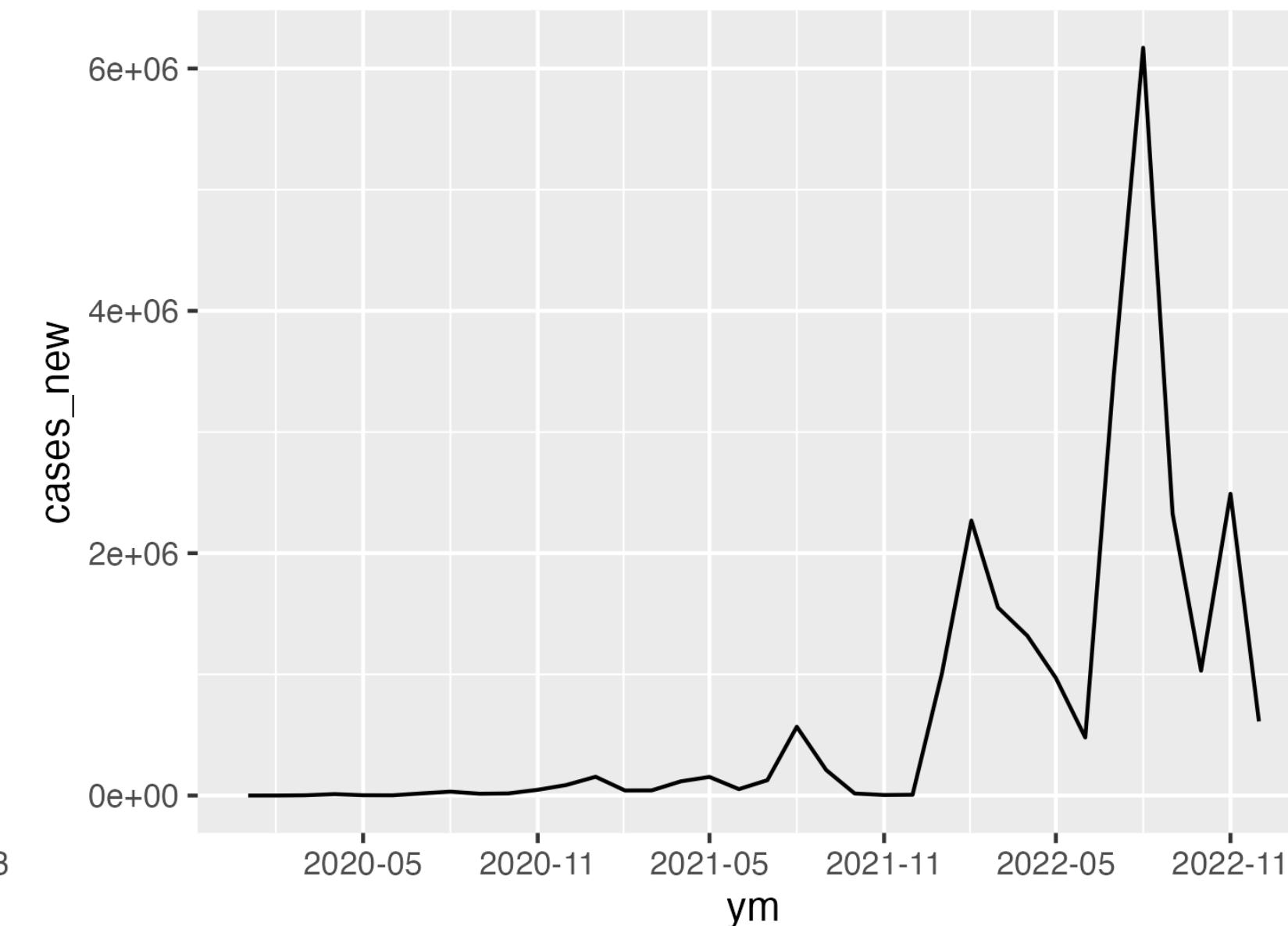
## 国内での新型コロナウィルス感染症感染者数

値はジョンズ・ホプキンズ大学による集計値

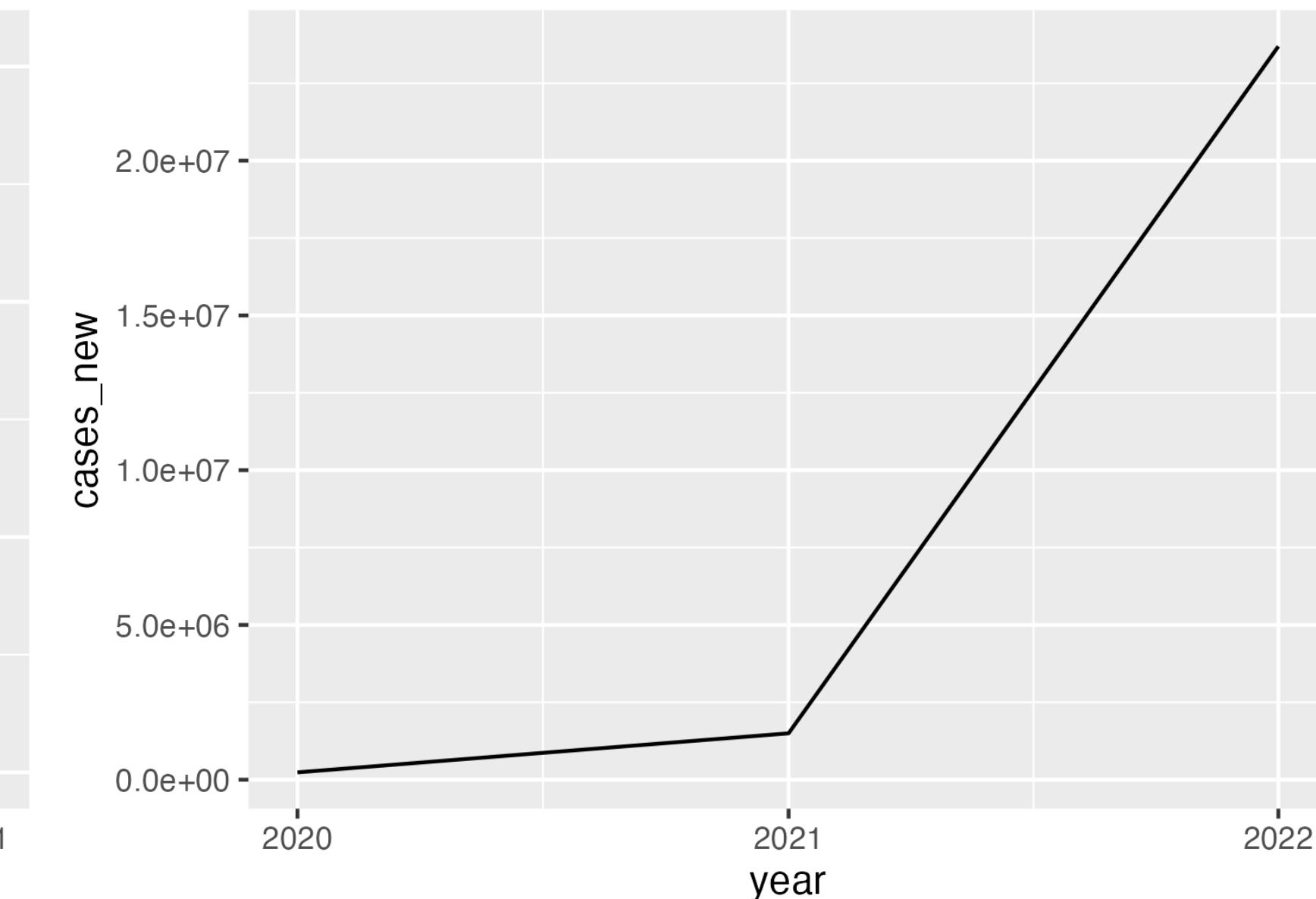
日次



月次



年次



# 時系列解析

時系列データから何らかの知見を得るための分析

**記述** 視覚的に時系列の特徴を表現する。プロット、コレログラム、自己相関

**モデリング** 時系列の確率的構造を解析

**予測** 現在までに得られた情報から将来の変動を予測。状態空間モデリング

# 変動を分解する

4つの変動に着目すると時系列の構造を読み取るのに役立つ

**傾向変動 (Trend)** 長期変動とも。長期間にわたる上昇または下降の持続的な変化

**循環変動 (Cycle)** 周期的に繰り返される変化

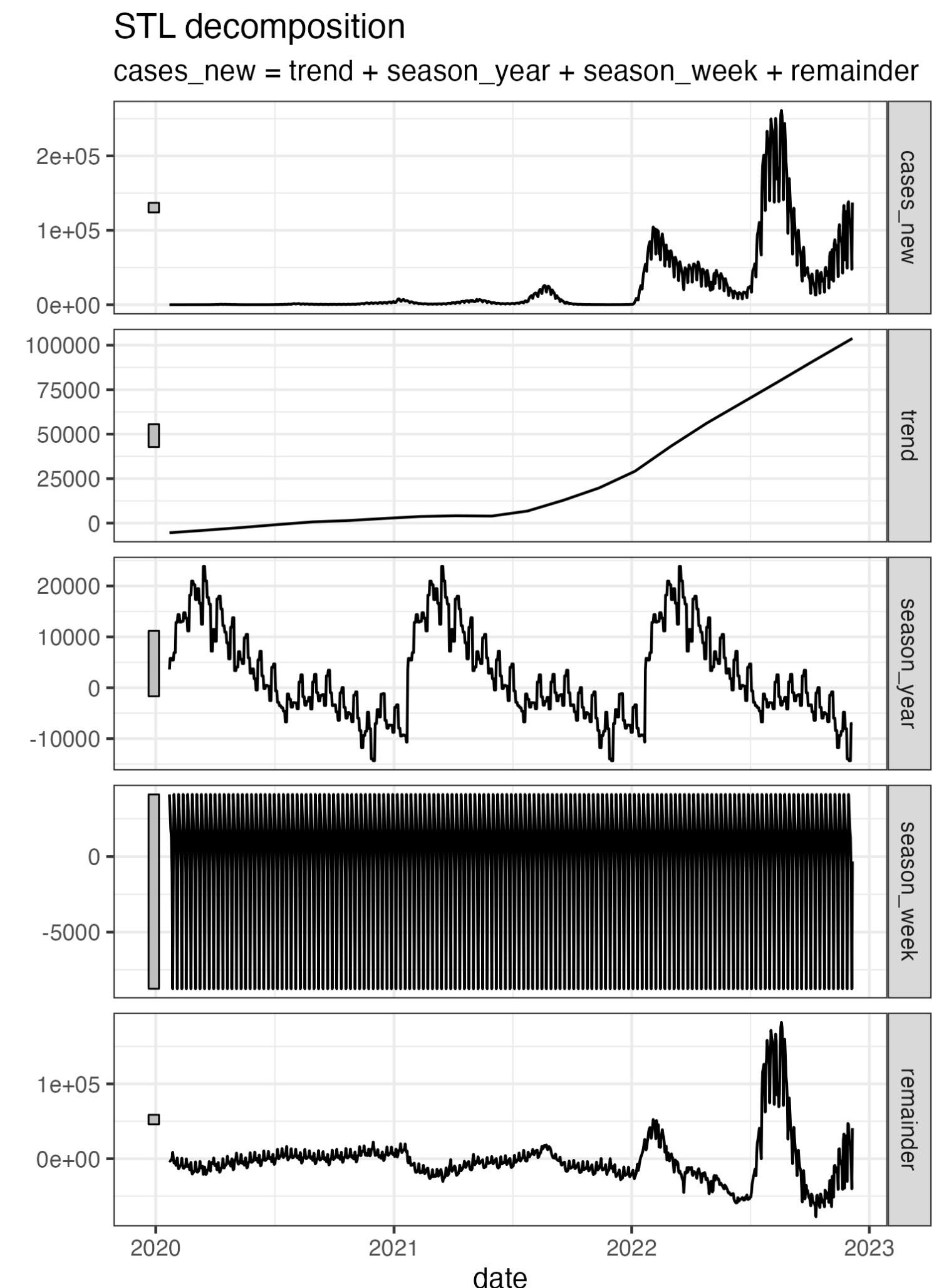
**季節変動 (Seasonal)** 1年の間（季節ごと）に繰り返される変化

**不規則変動 (Noise, Irregular)** 上記以外の説明がつかない変化  
突発的な変動、不規則かつ短期間に起こる変動

# STL分解: 変動を分解する

データ = トレンド + 季節性 + 残差

```
dcmp <-  
  as_tsibble(df_covid19_japan,  
             index = date) |>  
  model(STL(cases_new ~ season(window = Inf)))  
  components(dcmp) |>  
  autoplot()
```



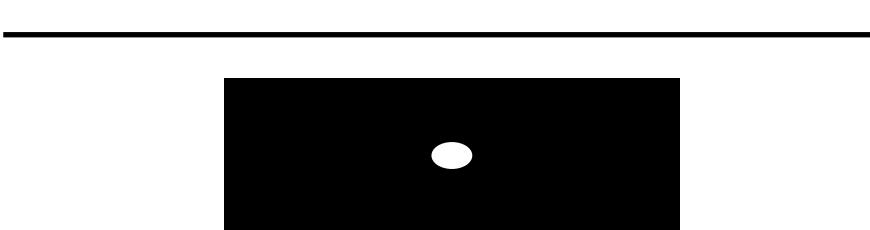
# 地理空間データ

# ベクタとラスタ

## ベクタデータ

地球上に存在する地物をデータ上に示したもの

特に点、線、面を扱う



徳島駅の場所を示すとき… 点

線路を示すとき… 線

領域を示すとき… 面

地物は座標と属性情報をもち、地図上で表現できる

## ラスタデータ

一定領域を格子状のグリッドで囲んだデータ

リモートセンシング

ラスタを構成する各グリッドに値が記録される

ラスタは時系列で重ねることができる

# Rでベクタデータを扱う: sfパッケージ

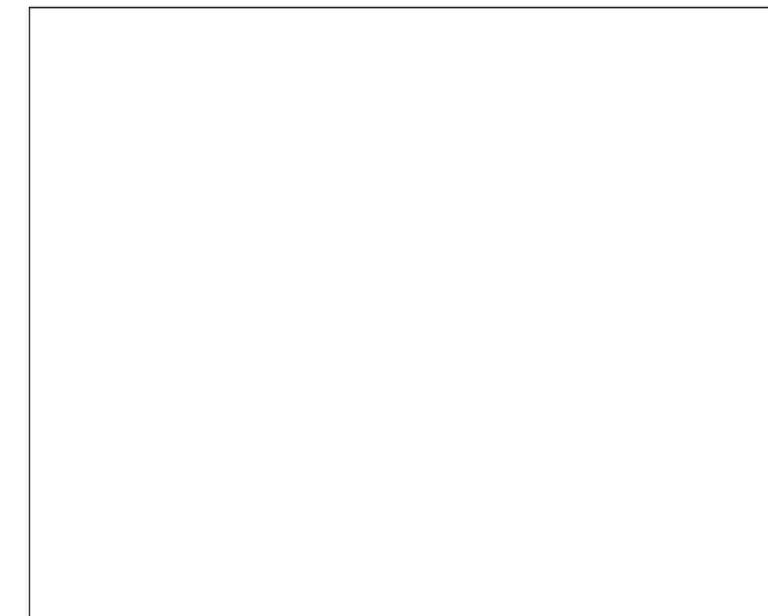
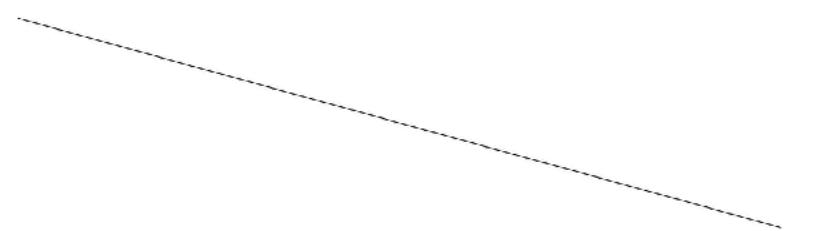
```
# 徳島駅の座標  
tksm_st_p <-  
  st_point(c(134.55129, 34.07470))
```



```
# 徳島駅と佐古駅の間の線（直線）  
tksm_st_line <-  
  st_linestring(  
    matrix(c(134.55129, 34.07470,  
            134.53868, 34.08056),  
    nrow = 2,  
    byrow = TRUE))
```



```
# 北方領土を含めた国土  
jp_area <-  
  matrix(  
    c(122.932502747, 20.42527771,  
      153.986663818, 20.42527771,  
      153.986663818, 45.557220459,  
      122.932502747, 45.557220459,  
      122.932502747, 20.42527771),  
    byrow = TRUE,  
    ncol = 2) |>  
  list() |>  
  st_polygon()
```

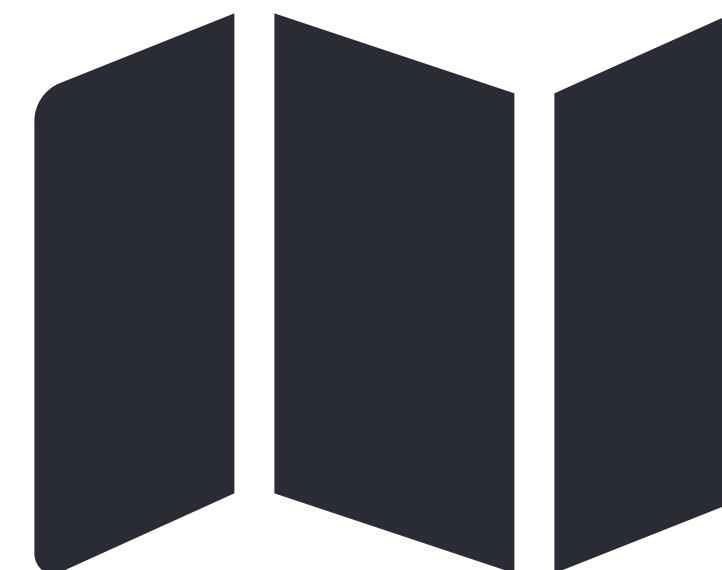


# 地球(3次元)を平面(2次元)に表現するには?

球体



平面



歪みが発生



# 座標参照系

空間データ座標が地球上のどの位置にあるかを特定するための仕組み

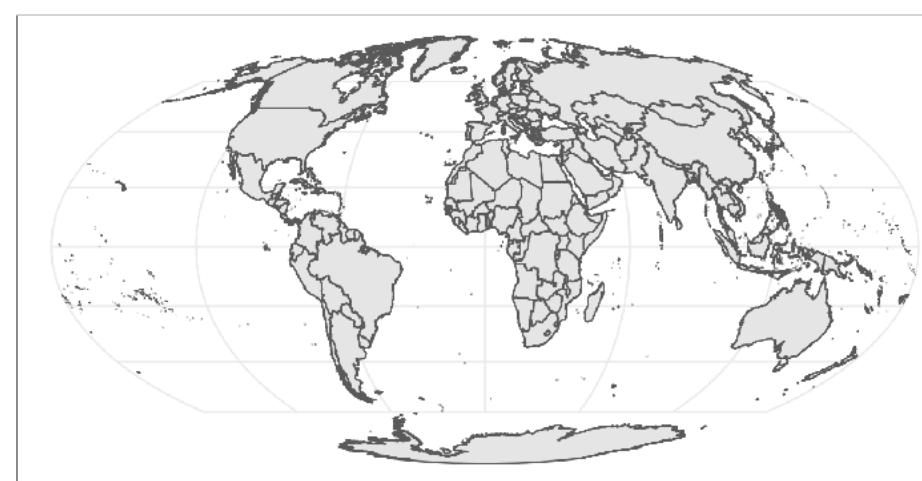
## 地理座標系

緯線と経線で構成される。東西方向を経度、南北方向を緯度で表現する。

多くの場合、北半球は正の緯度を持ち、南半球は負の緯度となる。経度の原点として本初子午線が使用される。

## 投影座標系

2次元のXY座標で表現。面積・距離・角度のいずれかに対する歪みを補正できる。



モルワイデ図法



メルカトル図法



グード図法

- 日本周辺を表示する投影法
- ・メルカトル図法
  - ・UTM座標系
  - ・平面直角座標系

# 座標参照系

## 地理座標系

SRID: 4326(WGS84)

世界測地系

SRID: 6668(JDG2011)

日本測地系2011

## 投影座標系

SRID: 6669~6687(JDG2011)

平面直角座標系

距離や面積の計算など、地理空間的な演算を行う際は地理座標系から投影座標系への変換を行う

# 座標参照系の指定

```
# 徳島駅と佐古駅のポイントに地理座標系（世界測地系）を指定する
```

```
sfc_tksm_st_p <-  
  st_sf(tksm_st_p, crs = 4326)  
sfc_sako_st_p <-  
  st_sf(sako_st_p, crs = 4326)
```

```
mapview(sfc_tksm_st_p) +  
  mapview(sfc_sako_st_p)
```



# 座標参照系の違いによる影響と変換

## 距離の計算

### 地理座標系

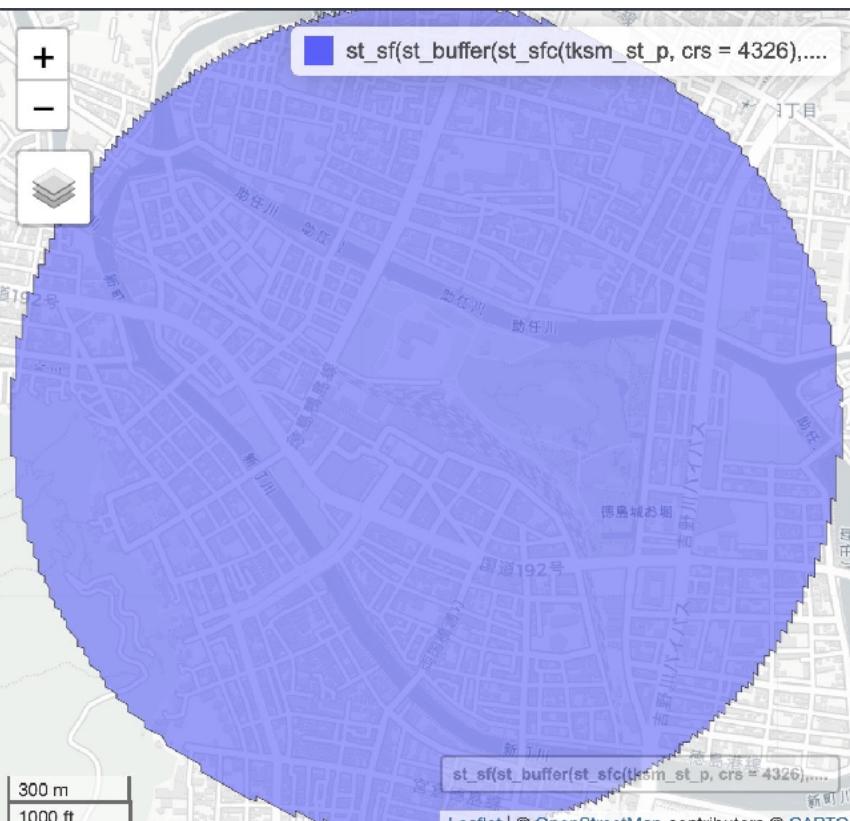
```
st_distance(  
  sfc_tksm_st_p,  
  sfc_sako_st_p)  
#> Units: [m]  
#> [,1]  
#> [1,] 1331.694
```

### 投影座標系

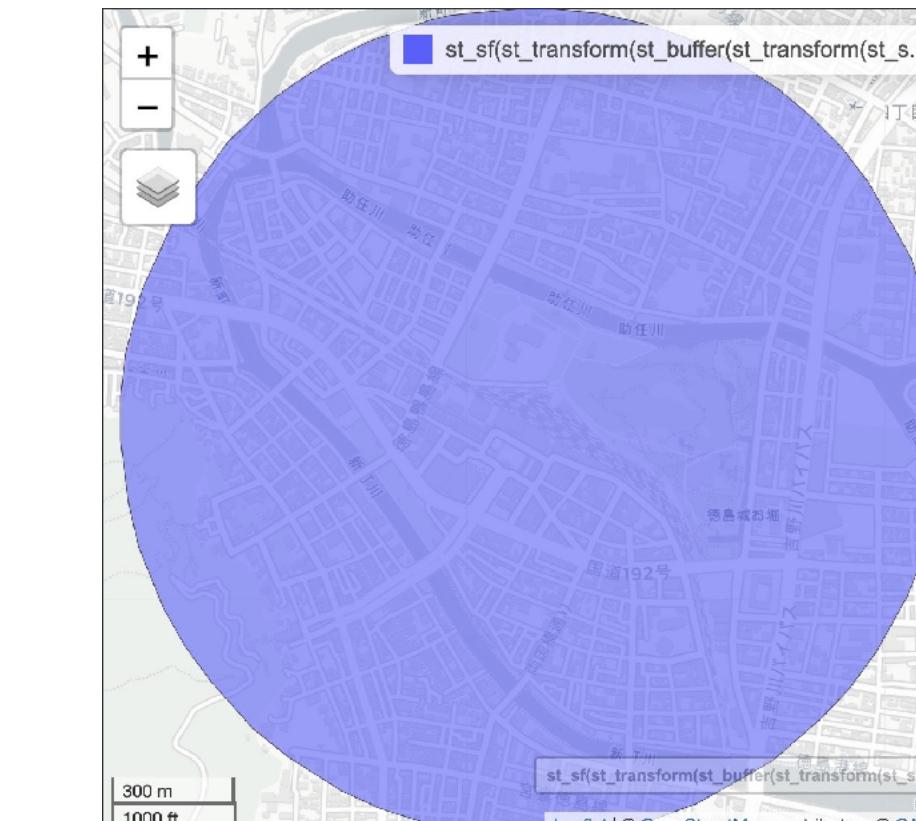
```
st_distance(  
  st_transform(sfc_tksm_st_p, crs = 6672),  
  st_transform(sfc_sako_st_p, crs = 6672))  
#> Units: [m]  
#> [,1]  
#> [1,] 1333.139
```

## バッファの生成

### 地理座標系



### 投影座標系



# 参考文献・URL

白石博 (2022). 時系列データ解析 (森北出版) ISBN: 978-4-627-08251-9