

情報分析

データ分析の目的

■ 正しい情報を得ても、分析しないと理解できないことがある

東京都_新型コロナウイルス陽性患者発表詳細

[リソースへ行く](#)

URL: https://stopcovid19.metro.tokyo.lg.jp/data/130001_tokyo_covid19_patients.csv

東京都 新型コロナウイルス対策サイト <https://stopcovid19.metro.tokyo.lg.jp/> で公開している都内 新型コロナウイルス陽性患者の詳細データです。（注）退院済フラグには、死亡退院含む

※ システムの仕様上、下記では1000recordsと表示されておりますが、実際には1000件以上となっております。

データエクスプローラー

埋めこみ

グリッド グラフ マップ 1000 records

« 1 - 100 »

Q Search data ... Go »

フィルター

No	全国地方...	都道府県名	市区町村名	公表_年月日	曜日	発症_年...	患者_居...	患者_年代	患者_性別	患者_属性	患者_状態	患者
1	130001	東京都		2020-01-24T00:00:00	金		湖北省武...	40代	男性			
2	130001	東京都		2020-01-25T00:00:00	土		湖北省武...	30代	女性			
3	130001	東京都		2020-01-30T00:00:00	木		湖南省長...	30代	女性			
4	130001	東京都		2020-02-13T00:00:00	木		都内	70代	男性			
5	130001	東京都		2020-02-14T00:00:00	金		都内	50代	女性			
6	130001	東京都		2020-02-14T00:00:00	金		都内	70代	男性			
12	130001	東京都		2020-02-15T00:00:00	土		都内	40代	男性			
14	130001	東京都		2020-02-15T00:00:00	土		都内	40代	男性			
8	130001	東京都		2020-02-15T00:00:00	土		都内	50代	女性			

なんのデータだとおもいますか？なにがわかりますか？

東京都オープンデータカタログサイトより引用

引用URL : <https://catalog.data.metro.tokyo.lg.jp/dataset/t000010d00000000068/resource/c2d997db-1450-43fa-8037-ebb11ec28d4c>

情報科学

2

データ分析の目的

■ 適切な指標を使って、適切な加工を行うことで分析・理解が進む



5月に入って陽性者数は減っているが、6月に入って増加している

東京都新型コロナウイルス感染症対策サイトより引用

引用URL : <https://stopcovid19.metro.tokyo.lg.jp/>

オープンデータ

■ 誰でも自由に使える形で公開されているデータ

- オープンなライセンス（著作権）
- オープンなアクセス（入手方法）
- オープンな形式（加工方法）

■ オープンデータの目的

- データをオープンにすることで、新しい価値を見出してもらう
- 公開団体のイメージ向上
 - 行政：透明性、信頼性
 - 企業：新規価値の発掘、ブランド向上

オープンデータのライセンス

■ クリエイティブ・コモンズ

- <https://creativecommons.org/licenses/by/4.0/deed.ja>



表示 4.0 国際 (CC BY 4.0)

これは人が読んでわかりやすいようにした[ライセンス](#)の要約です。(ライセンスの代わりになるものではありません。) [免責条項](#)。

あなたは以下の条件に従う限り、自由に：

共有 — どのようなメディアやフォーマットでも資料を複製したり、再配布できます。

翻案 — マテリアルをリミックスしたり、改変したり、別の作品のベースにしたりできます。
営利目的も含め、どのような目的でも。

あなたがライセンスの条件に従っている限り、許諾者がこれらの自由を取り消すことはできません。



あなたの従うべき条件は以下の通りです。



表示 — あなたは [適切なクレジット](#) を表示し、ライセンスへのリンクを提供し、[変更があったらその旨を示さ](#) なければなりません。これらは合理的であればどのような方法で行っても構いませんが、許諾者があなたやあなたの利用行為を支持していると示唆するような方法は除きます。

オープンデータのライセンス

■ 質問：どういう意味ですか？



表示 4.0 国際 (CC BY 4.0)

これは人が読んでわかりやすいようにした[ライセンス](#)の要約です。(ライセンスの代わりになるものではありません。) [免責条項](#)。

あなたは以下の条件に従う限り、自由に：

共有 — どのようなメディアやフォーマットでも資料を複製したり、再配布できます。

翻案 — マテリアルをリミックスしたり、改変したり、別の作品のベースにしたりできます。
営利目的も含め、どのような目的でも。

あなたがライセンスの条件に従っている限り、許諾者がこれらの自由を取り消すことはできません。



あなたの従うべき条件は以下の通りです。




表示 — あなたは [適切なクレジット](#) を表示し、ライセンスへのリンクを提供し、[変更があったらその旨を示さ](#) なければなりません。これらは合理的であればどのような方法で行っても構いませんが、許諾者があなたやあなたの利用行為を支持していると示唆するような方法は除きます。

オープンデータのライセンス

■ クリエイティブ・コモンズ

- ルールを守れば、コピー、再配布、加工OK



表示 4.0 国際 (CC BY 4.0)


これは人が読んでわかりやすいようにした [ライセンス](#) の要約です。(ライセンスの代わりになるものではありません。) [免責条項](#).

あなたは以下の条件に従う限り、自由に：

共有 — どのようなメディアやフォーマットでも資料を複製したり、再配布できます。

翻案 — マテリアルをリミックスしたり、改変したり、別の作品のベースにしたりできます。
営利目的も含め、どのような目的でも。

あなたがライセンスの条件に従っている限り、許諾者がこれらの自由を取り消すことはできません。



オープンデータのライセンス

■ クリエイティブ・コモンズ

● ルール

- クレジット、引用元を明記
- 加工した場合は、「加工した」と書く

あなたの従うべき条件は以下の通りです。



表示 — あなたは 適切なクレジット を表示し、ライセンスへのリンクを提供し、変更があったらその旨を示さ なければなりません。これらは合理的であればどのような方法で行っても構いませんが、許諾者があなたやあなたの利用行為を支持していると示唆するような方法は除きます。

オープンデータのアクセス

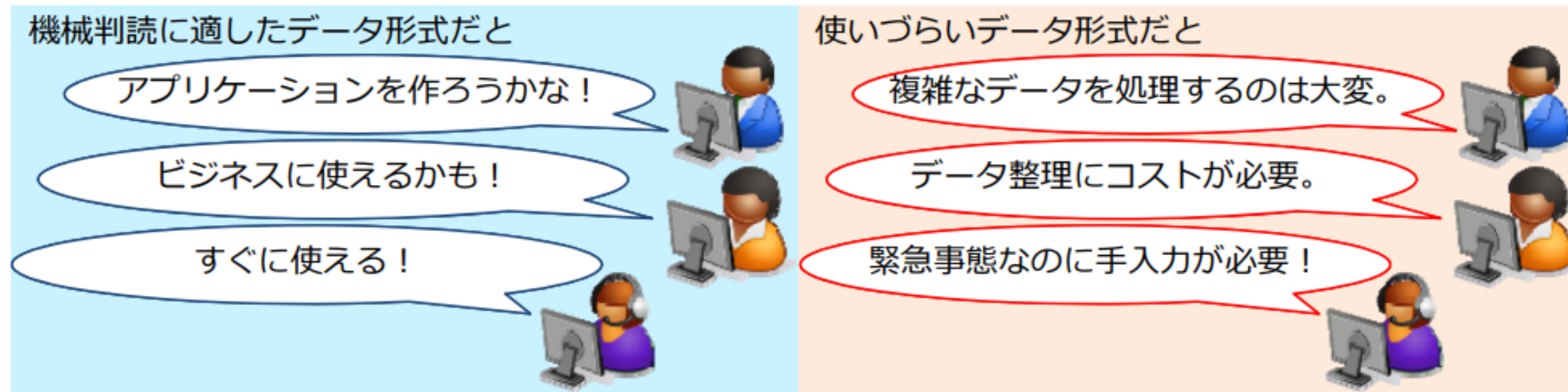
■ 誰でもアクセスできること

- 以下はオープンデータではありません
 - お金を払わないとアクセスできない
 - 申請しないとアクセスできない
 - 特定の団体でないとアクセスできない

オープンデータの形式

■ 使いやすい形式であること

- ダウンロードしたらすぐに使える
- データの整理不要
- データを開くのに特殊なソフトが不要



オープンデータの形式

■ ファイルの形式

- できるだけ全員が使えるような形式がよい
 - 無料のソフト、多くのソフトで開ける形式（テキスト、PDFなど）
 - 有料のソフトが必要な場合、できるだけたくさんの人が使っている形式（Officeなど）

国際的な機関が制定したファイル形式の例

規格名	ファイルの拡張子	規格
Office Open XML	.docx, .xlsxなど	ECMA-376, ISO/IEC 29500
PDF (Portable Document Format)	.pdf	ISO 32000-1
CSV (Comma-Separated Values)	.csv	RFC 4180
GML (Geography Markup Language)	.gml	ISO 19136

オープンデータの例

■ 東京都オープンデータカタログサイト

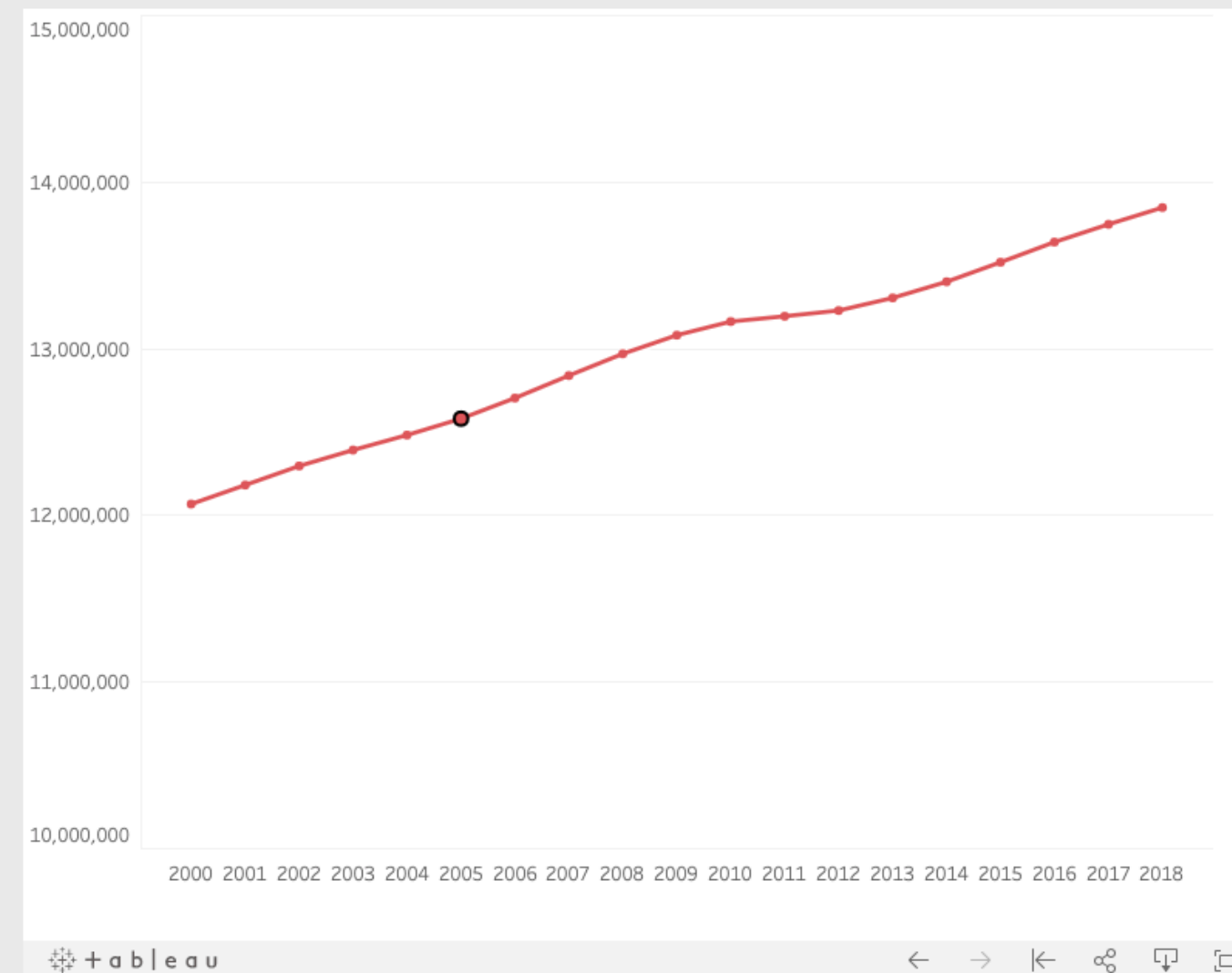
- <https://portal.data.metro.tokyo.lg.jp/>



オープンデータの例

東京都のオープンデータを使うとどんなことがわかるでしょうか？

総人口：13,843,403人 (2018年8月1日時点)



データソース

明治33年～平成29年 東京都全体 出生数・死亡・死産・婚姻・離婚・合計特殊出生率・平均初婚年齢

東京都オープンデータカタログサイト（東京都）より引用

引用： <https://portal.data.metro.tokyo.lg.jp/>

オープンデータの例

2 知的財産権の取扱い

利用者は、本サイトで掲載されている情報等に関する以下の事項について理解した上で、第三者の知的財産権を尊重するものとし、情報等の取扱いについては慎重な配慮を行うようにしてください。

本サイトに存在する著作物（掲載されている情報等を含みます。）の著作権は、「注」があるものを除いて、クリエイティブ・コモンズ・ライセンス表示 4.0 国際のもとでライセンスされています。

従って、コンテンツは、本利用規約以外の別の利用ルールが個別に適用されるコンテンツを除き、どなたでも本利用規約に従って複製、公衆送信、翻訳・変形等の翻案等、自由に利用できます。商用利用も可能です。

コンテンツ利用に際しては、以下を参考にクレジット等を記載してください。また、【 】内の部分は、利用者において記載してください。

オープンデータの例

(1) クリエイティブ・コモンズ・ライセンス 表示4.0国際のもとでライセンスされているコンテンツを使用する場合

ア ライセンスされている著作物を改変せず、複製して利用する場合は、以下のクレジットを記載してください。

【ライセンスされている著作物のタイトル】、東京都・【その他の著作権者】、クリエイティブ・コモンズ・ライセンス 表示4.0国際
(<https://creativecommons.org/licenses/by/4.0/deed.ja>)

イ ライセンスされている著作物を改変して利用する場合は、上記クレジットに加えて編集・加工等を行った旨を記載してください。なお、改変して利用する場合、編集・加工をした情報を、東京都又は都内区市町村（以下「コンテンツ提供者」といいます。）が作成したかのような態様で公表・利用することを禁止します。

この【作品・アプリ・データベース等】は、以下の著作物を改変して利用しています。

【ライセンスされている著作物のタイトル】、東京都・【その他の著作権者】、クリエイティブ・コモンズ・ライセンス 表示4.0国際
(<https://creativecommons.org/licenses/by/4.0/deed.ja>)

ライセンスの表示は、「クリエイティブ・コモンズ・ライセンス 表示4.0国際」の文字部分にハイパーリンクを貼る方法や、クリエイティブ・コモンズの指定するマーク
(<https://kommt.files.wordpress.com/2015/04/by.png?w=150&h=52>) を表示する方法で行うことも可能です。

質問：データ収集と価値

- コンビニエンスストアのレジで、どのようなデータを集められるか？
- 集めたデータを何に活用できるか？

発展：機械学習・ビッグデータ分析

■ ASL指文字のデータセット Sign-Language-MNIST(SL-MNIST)

- 数万枚の指文字画像データ



発展：機械学習・ビッグデータ分析

■ ASL指文字のデータセット Sign-Language-MNIST(SL-MNIST)

- 機械学習、ビッグデータ分析により指文字を認識

**CNN using Keras(100% Accuracy)**
Python notebook using data from [Sign Language MNIST](#) · 4,203 views · 1mo ago · 

Description of the Sign-Language MNIST Problem (American Sign Language)

American Sign Language (ASL) is a complete, natural language that has the same linguistic properties as spoken languages, with grammar that differs from English. ASL is expressed by movements of the hands and face. It is the primary language of many North Americans who are deaf and hard of hearing, and is used by many hearing people as well. The dataset format is patterned to match closely with the classic MNIST. Each training and test case represents a label (0-25) as a one-to-one map for each alphabetic letter A-Z (and no cases for 9=J or 25=Z because of gesture motions). The training data (27,455 cases) and test data (7172 cases) are approximately half the size of the standard MNIST but otherwise similar with a header row of label, pixel1,pixel2....pixel784 which represent a single 28×28 pixel image with grayscale values between 0-255. The original hand gesture image data represented multiple users repeating the gesture against different backgrounds. The Sign Language MNIST data came from greatly extending the small number (1704) of the color images included as not cropped around the hand region of interest. To create new data, an image pipeline was used based on ImageMagick and included cropping to hands-only, gray-scaling, resizing, and then creating at least 50+ variations to enlarge the quantity.



A B C D E F

Version 5 of 5

Quick Version

Notebook

Description Of The Sign-Language MNIS...

Loading The ASL Dataset

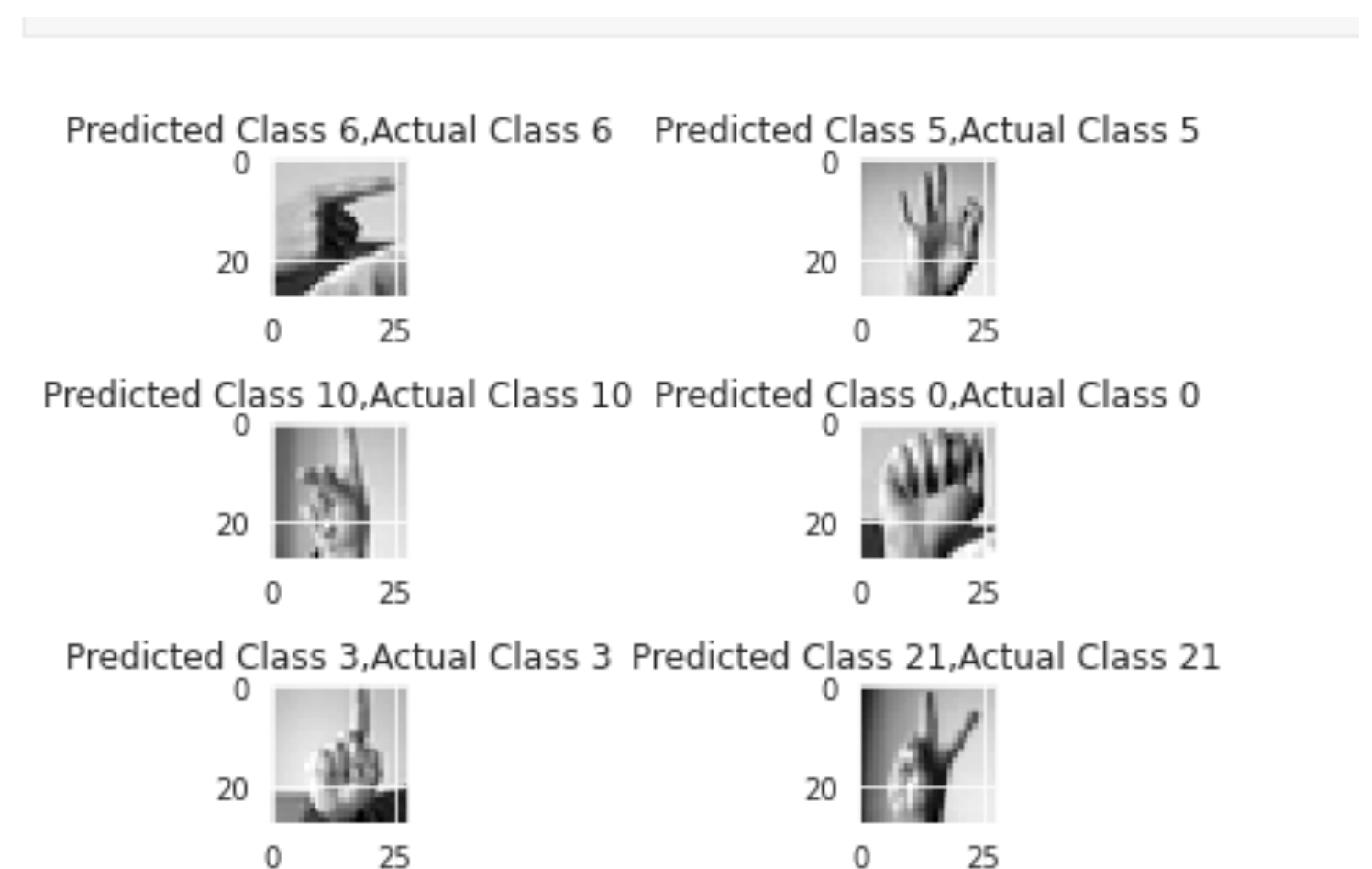
Data Visualization And Preprocessing

Data Augmentation

CONVOLUTIONAL NEURAL NETWORKS...

Training The Model

Analysis After Model Training



ここまでのまとめ

■ データ分析の目的

- データはそのままでは利用できない
- 加工・分析することで、理解を深める

■ オープンデータ

- 国、企業などが公開した「誰でもアクセス・加工」ができるデータ

課題

■ 東京都新型コロナウイルス対策サイトからオープンデータを入手し、以下と同様のグラフを作成してください。

- 日ごとに、「陽性者数」「7日間移動平均」を算出してください
- 同じ傾向であることがわかれば、色や形式などなど同じにする必要はありません
- Web版だと棒グラフと折れ線グラフを同時に表示できないため

ここは不要

モニタリング項目(1)
新規陽性者数

152.4 人

7/11 の数値 (前日比: +10.7 人)



課題

■ 注意

- データの出典を明記すること（引用のルール）
- クレジットを明記すること（ライセンスのルール）

課題

■ 東京都新型コロナウイルス対策サイト

- <https://stopcovid19.metro.tokyo.lg.jp/>

■ オープンデータ

- <https://catalog.data.metro.tokyo.lg.jp/dataset/t000010d00000000068/resource/c2d997db-1450-43fa-8037-ebb11ec28d4c>
- 「リソースへ行く」を選択してDL
- ライセンスもみておくこと

東京都_新型コロナウイルス陽性患者発表詳細

URL: https://stopcovid19.metro.tokyo.lg.jp/data/130001_tokyo_covid19_patients.csv

東京都 新型コロナウイルス対策サイト <https://stopcovid19.metro.tokyo.lg.jp/> で公開している都内 新型コロナウイルス陽性患者の詳細データです。(注) 退院済フラグには、死亡退院含む

※ システムの仕様上、下記では1000recordsと表示されておりますが、実際には1000件以上となっております。

リソースへ行く

リソース

東京都_新型コロナウイルス陽性患者発表詳細

ソーシャル

Twitter

Facebook

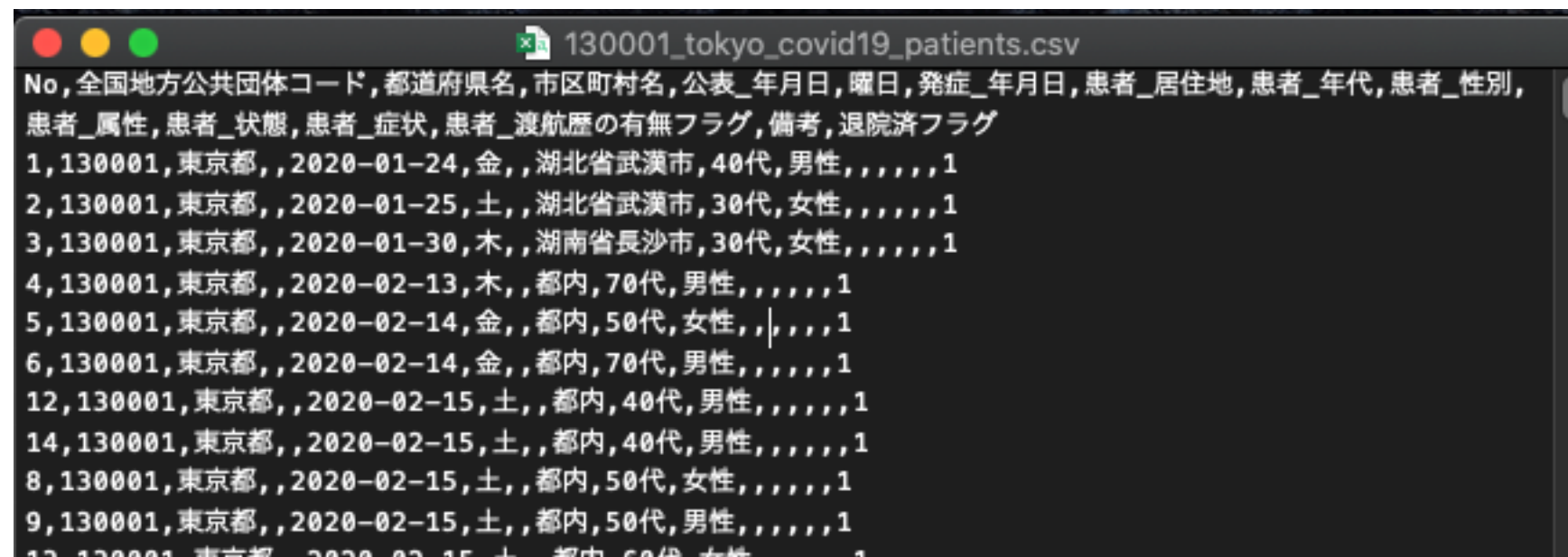
追加情報

フィールド	値
最終更新日	2020年7月17日, 午前8時30分 (UTC+09:00)
メタデータ最終更新日時	2020年3月5日, 午後4時0分 (UTC+09:00)
作成日	2020年3月5日, 午後4時0分 (UTC+09:00)
データ形式	CSV
ライセンス	クリエイティブ・コモンズ 表示 (CC BY)

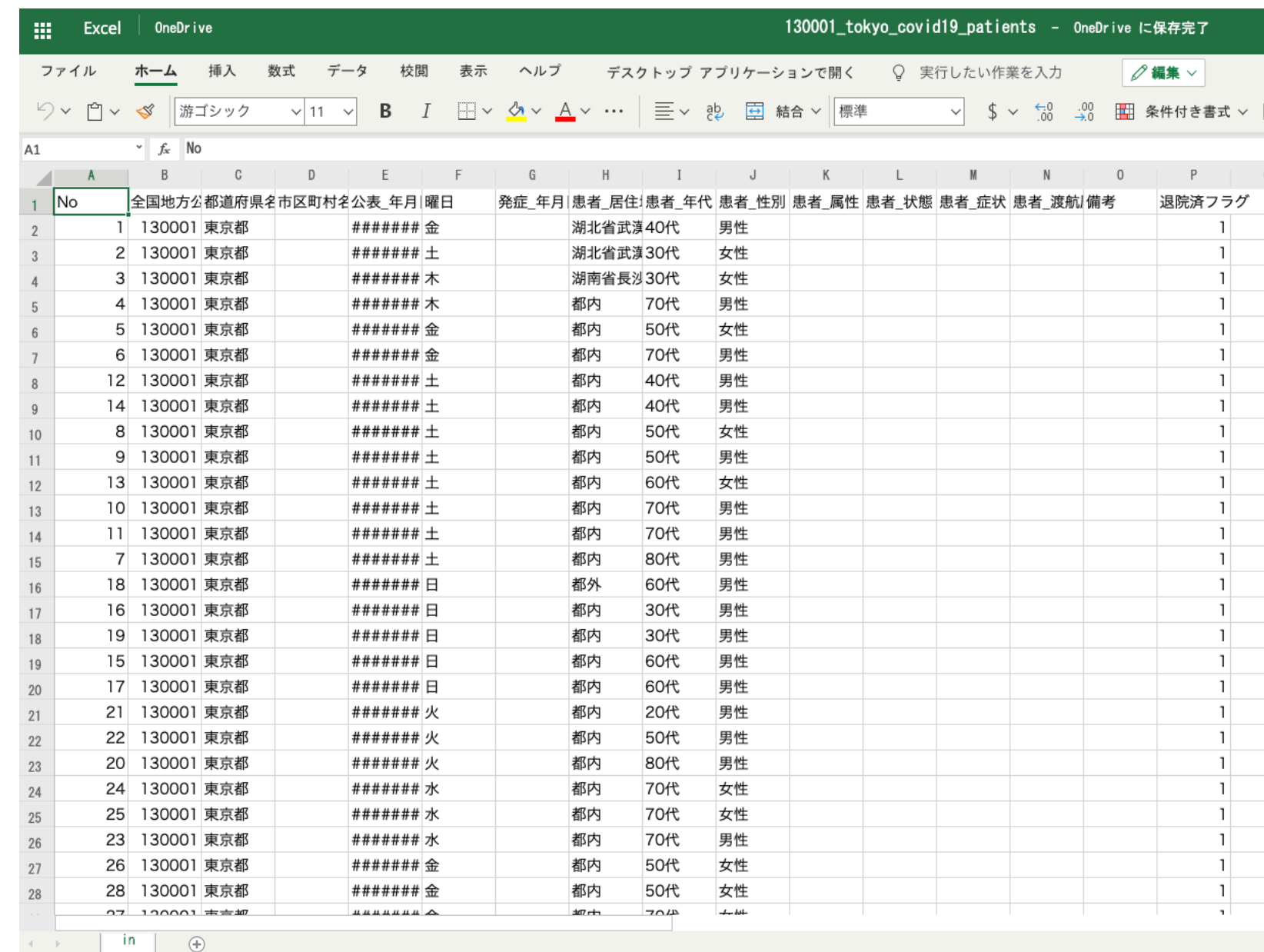
課題

■ オープンデータをExcelで開く

- オープンデータはCSV形式（"/,"で区切られた形式。プログラムで扱いやすいため。）
- 一度Office 365にアップロードすると、Excelでひらける（デスクトップアプリなら、直接Excelで開けば良い）



```
No,全国地方公共団体コード,都道府県名,市区町村名,公表_年月日,曜日,発症_年月日,患者_居住地,患者_年代,患者_性別,患者_属性,患者_状態,患者_症状,患者_渡航歴の有無フラグ,備考,退院済フラグ
1,130001,東京都,,2020-01-24,金,,湖北省武漢市,40代,男性,,,,,1
2,130001,東京都,,2020-01-25,土,,湖北省武漢市,30代,女性,,,,,1
3,130001,東京都,,2020-01-30,木,,湖南省長沙市,30代,女性,,,,,1
4,130001,東京都,,2020-02-13,木,,都内,70代,男性,,,,,1
5,130001,東京都,,2020-02-14,金,,都内,50代,女性,,|,,,1
6,130001,東京都,,2020-02-14,金,,都内,70代,男性,,,,,1
12,130001,東京都,,2020-02-15,土,,都内,40代,男性,,,,,1
14,130001,東京都,,2020-02-15,土,,都内,40代,男性,,,,,1
8,130001,東京都,,2020-02-15,土,,都内,50代,女性,,,,,1
9,130001,東京都,,2020-02-15,土,,都内,50代,男性,,,,,1
13,130001,東京都,,2020-02-15,土,,都内,60代,女性,,,,,1
```



130001_tokyo_covid19_patients - OneDrive に保存完了															
Excel OneDrive															
ファイル ホーム 挿入 数式 データ 校閲 表示 ヘルプ デスクトップ アプリケーションで開く 実行したい作業を入力 編集															
A1 No															
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
No	全国地方公	都道府県名	市区町村名	公表_年月	曜日	発症_年月	患者_居住	患者_年代	患者_性別	患者_属性	患者_状態	患者_症状	患者_渡航	備考	退院済フラグ
1	130001	東京都		2020-01-24	金		湖北省武漢市	40代	男性						1
2	130001	東京都		2020-01-25	土		湖北省武漢市	30代	女性						1
3	130001	東京都		2020-01-30	木		湖南省長沙市	30代	女性						1
4	130001	東京都		2020-02-13	木		都内	70代	男性						1
5	130001	東京都		2020-02-14	金		都内	50代	女性						1
6	130001	東京都		2020-02-14	金		都内	70代	男性						1
7	130001	東京都		2020-02-15	土		都内	40代	男性						1
8	130001	東京都		2020-02-15	土		都内	40代	男性						1
9	130001	東京都		2020-02-15	土		都内	50代	女性						1
10	130001	東京都		2020-02-15	土		都内	50代	男性						1
11	130001	東京都		2020-02-15	土		都内	60代	女性						1
12	130001	東京都		2020-02-15	土		都内	70代	男性						1
13	130001	東京都		2020-02-15	土		都内	70代	男性						1
14	130001	東京都		2020-02-15	土		都内	80代	男性						1
15	130001	東京都		2020-02-15	土		都外	60代	男性						1
16	130001	東京都		2020-02-15	土		都内	30代	男性						1
17	130001	東京都		2020-02-15	土		都内	30代	男性						1
18	130001	東京都		2020-02-15	土		都内	60代	男性						1
19	130001	東京都		2020-02-15	土		都内	60代	男性						1
20	130001	東京都		2020-02-15	土		都内	20代	男性						1
21	130001	東京都		2020-02-15	土		都内	50代	男性						1
22	130001	東京都		2020-02-15	土		都内	80代	男性						1
23	130001	東京都		2020-02-15	土		都内	70代	女性						1
24	130001	東京都		2020-02-15	土		都内	70代	女性						1
25	130001	東京都		2020-02-15	土		都内	70代	男性						1
26	130001	東京都		2020-02-15	土		都内	50代	女性						1
27	130001	東京都		2020-02-15	土		都内	50代	女性						1
28	130001	東京都		2020-02-15	土		都内	70代	女性						1

課題

■ ヒント

- 各日の感染者数をカウントする
- オートフィルで日にちを自動入力できる

	A	B	C	D	E	F	G	H
1								
2		2020/4/1						
3		2020/4/2						
4		2020/4/3						
5		2020/4/4						
6		2020/4/5						
7		2020/4/6						
8		2020/4/7						
9		2020/4/8						
10		2020/4/9						
11		2020/4/10						
12		2020/4/11						
13		2020/4/12						
14		2020/4/13						
15		2020/4/14						
16								
17								
18								

課題

■ ヒント

- 各日の感染者数をカウントする
- COUNTIFで以下を指定
 - 範囲：「公表_年月日」
 - 検索条件：「B2」

SUM						
	A	B	C	D	E	F
1						
2		2020/4/1	Sheet1!B2)			
3		2020/4/2				
4		2020/4/3				
5		2020/4/4				
6		2020/4/5				
7		2020/4/6				
8		2020/4/7				
9		2020/4/8				
0		2020/4/9				
1		2020/4/10				
2		2020/4/11				
3		2020/4/12				
4		2020/4/13				
5		2020/4/14				
6						
7						

130001_tokyo_covid19_patients						
	A	B	C	D	E	F
1	No	全国地方公共	都道府県名	市区町村名	公表_年月日	曜日
2	1	130001	東京都		2020/1/24	金
3	2	130001	東京都		2020/1/25	土
4	3	130001	東京都		2020/1/30	木
5	4	130001	東京都		2020/2/13	木
6	5	130001	東京都		2020/2/14	金
7	6	130001	東京都		2020/2/14	金
8	12	130001	東京都		2020/2/15	土
9	14	130001	東京都		2020/2/15	土
10	8	130001	東京都		2020/2/15	土
11	9	130001	東京都		2020/2/15	土
12	13	130001	東京都		2020/2/15	土
13	10	130001	東京都		2020/2/15	土
14	11	130001	東京都		2020/2/15	土
15	7	130001	東京都		2020/2/15	土
16	18	130001	東京都		2020/2/16	日
17	16	130001	東京都		2020/2/16	日
18	19	130001	東京都		2020/2/16	日
19	15	130001	東京都		2020/2/16	日
20	17	130001	東京都		2020/2/16	日
21	21	130001	東京都		2020/2/18	火
22	22	130001	東京都		2020/2/18	火
23	20	130001	東京都		2020/2/18	火
24	24	130001	東京都		2020/2/19	水
25	25	130001	東京都		2020/2/19	水
26	23	130001	東京都		2020/2/19	水
27	26	130001	東京都		2020/2/21	金
28	28	130001	東京都		2020/2/21	金
29	27	130001	東京都		2020/2/21	金
30	29	130001	東京都		2020/2/22	土

あとはオートフィル

2020/4/1	67
2020/4/2	98
2020/4/3	92
2020/4/4	118
2020/4/5	141
2020/4/6	85
2020/4/7	87
2020/4/8	156
2020/4/9	183
2020/4/10	199
2020/4/11	198
2020/4/12	174
2020/4/13	100
2020/4/14	159

課題

■ ヒント：7日移動平均

- 基準日から7日分を合計し、7で割った値（7日間の平均値）
 - 7日目：1日目～7日目
 - 8日目：2日目～8日目
 - 9日目：3日目～9日目
- 1日だけ急増、急減するような値（スパイク値）の影響を防ぎ、全体を俯瞰することができる

課題

■ ヒント：7日移動平均

- 基準日から7日分を合計し、7で割った値（7日間の平均値）

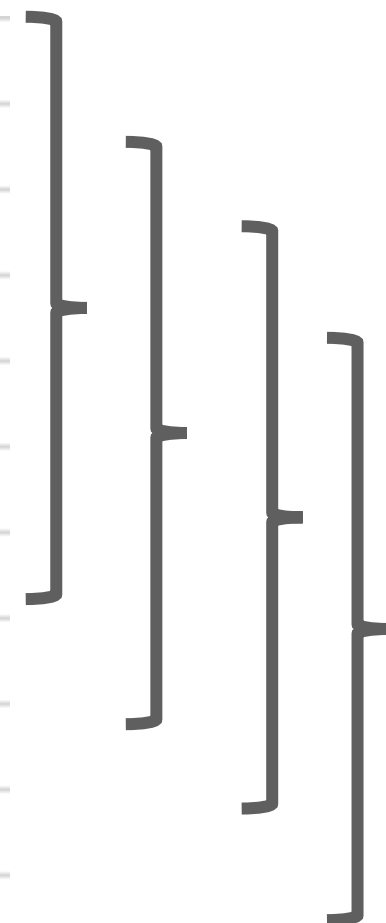
2020/4/1	67	} 平均値をとるだけ
2020/4/2	98	
2020/4/3	92	
2020/4/4	118	
2020/4/5	141	
2020/4/6	85	
2020/4/7	87	
2020/4/8	156	
2020/4/9	183	
2020/4/10	199	
2020/4/11	198	
2020/4/12	174	
2020/4/13	100	
2020/4/14	159	

課題

■ ヒント：7日移動平均

- 基準日から7日分を合計し、7で割った値（7日間の平均値）

2020/4/1	67
2020/4/2	98
2020/4/3	92
2020/4/4	118
2020/4/5	141
2020/4/6	85
2020/4/7	87
2020/4/8	156
2020/4/9	183
2020/4/10	199
2020/4/11	198
2020/4/12	174
2020/4/13	100
2020/4/14	159



1日ずらして平均をとるので、「移動」平均

課題

■ ヒント：7日移動平均

- 基準日から7日分を合計し、7で割った値（7日間の平均値）

2020/4/1	67
2020/4/2	98
2020/4/3	92
2020/4/4	118
2020/4/5	141
2020/4/6	85
2020/4/7	87
2020/4/8	156
2020/4/9	183
2020/4/10	199
2020/4/11	198
2020/4/12	174
2020/4/13	100
2020/4/14	159

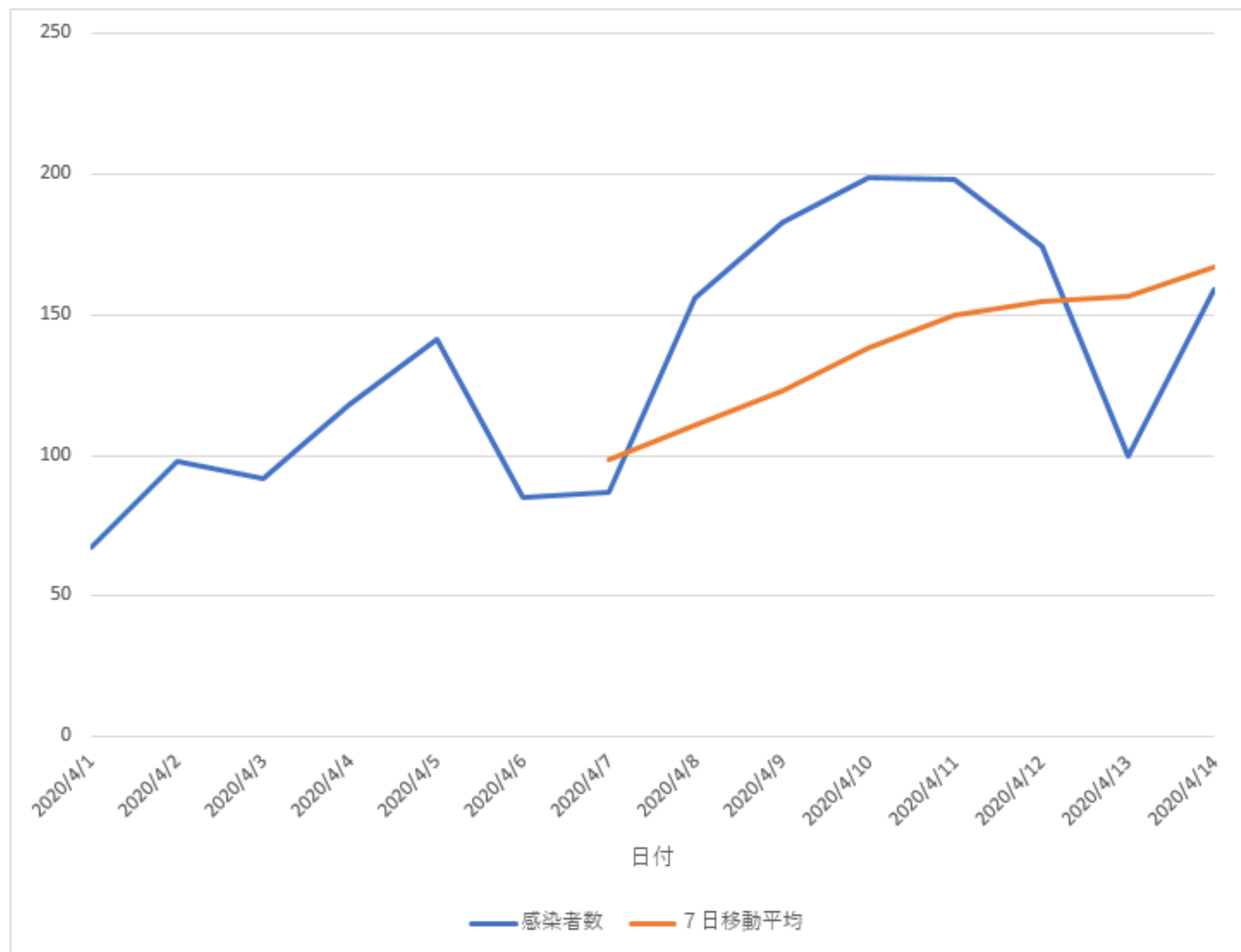
日付	感染者数	7日移動平均
2020/4/1	67	
2020/4/2	98	
2020/4/3	92	
2020/4/4	118	
2020/4/5	141	
2020/4/6	85	
2020/4/7	87	98.2857143
2020/4/8	156	111
2020/4/9	183	123.142857
2020/4/10	199	138.428571
2020/4/11	198	149.857143
2020/4/12	174	154.571429
2020/4/13	100	156.714286
2020/4/14	159	167

7日分のデータがないので
空っぽになる
(これでOK)

課題

■ ヒント

- あとは体裁をととのえるだけ！



課題

■ ヒント

● ライセンス

この【作品・アプリ・データベース等】は、以下の著作物を改変して利用しています。

【ライセンスされている著作物のタイトル】、東京都・【その他の著作権者】、クリエイティブ・コモンズ・ライセンス 表示4.0国際

(<https://creativecommons.org/licenses/by/4.0/deed.ja>)

● 今回は以下のように書くと良いでしょう

この図は、以下の著作物を改変して利用しています。

東京都_新型コロナウイルス陽性患者発表詳細、東京都、クリエイティブ・コモンズ・ライセンス 表示4.0
国際 (<https://creativecommons.org/licenses/by/4.0/deed.ja>)