

アメダス気象データ分析 チャレンジ！入門

4. データ整理・可視化・分析

主催：気象ビジネス推進コンソーシアム

共催：岐阜大学工学部附属応用気象研究センター

資料作成：吉野 純（岐阜大学）



本教材について

WXBC 2022年度テクノロジー研修

「アメダス気象データ分析チャレンジ！入門」講習テキスト

Copyright (c) 気象データ×IT 勉強会, 吉野純, 岐阜大学工学部附属応用気象研究センター 2022 All rights reserved.

<利用条件>

本書は、本書に記載した要件・技術・方式に関する内容が変更されないこと、および出典を明示した上で、本書を前提に、無償でその全部または一部を複製、翻案、翻訳、転記、引用、公衆送信等して利用できます。なお、全体または一部を複製、翻案、翻訳された場合は、本書にある著作権表示および利用条件を明示してください。

<免責事項>

本書の著作権者は、本書の記載内容に関して、その正確性、商品性、利用目的への適合性等に関して保証するものではなく、特許権、著作権、その他の権利を侵害していないことを保証するものでもありません。本書の利用により生じた損害について、本書の著作権者は、法律上のいかなる責任も負いません。

気象データをExcelで見てください



ダウンロードした気象データ「amedas-2017.xlsx」を開いてみましょう。

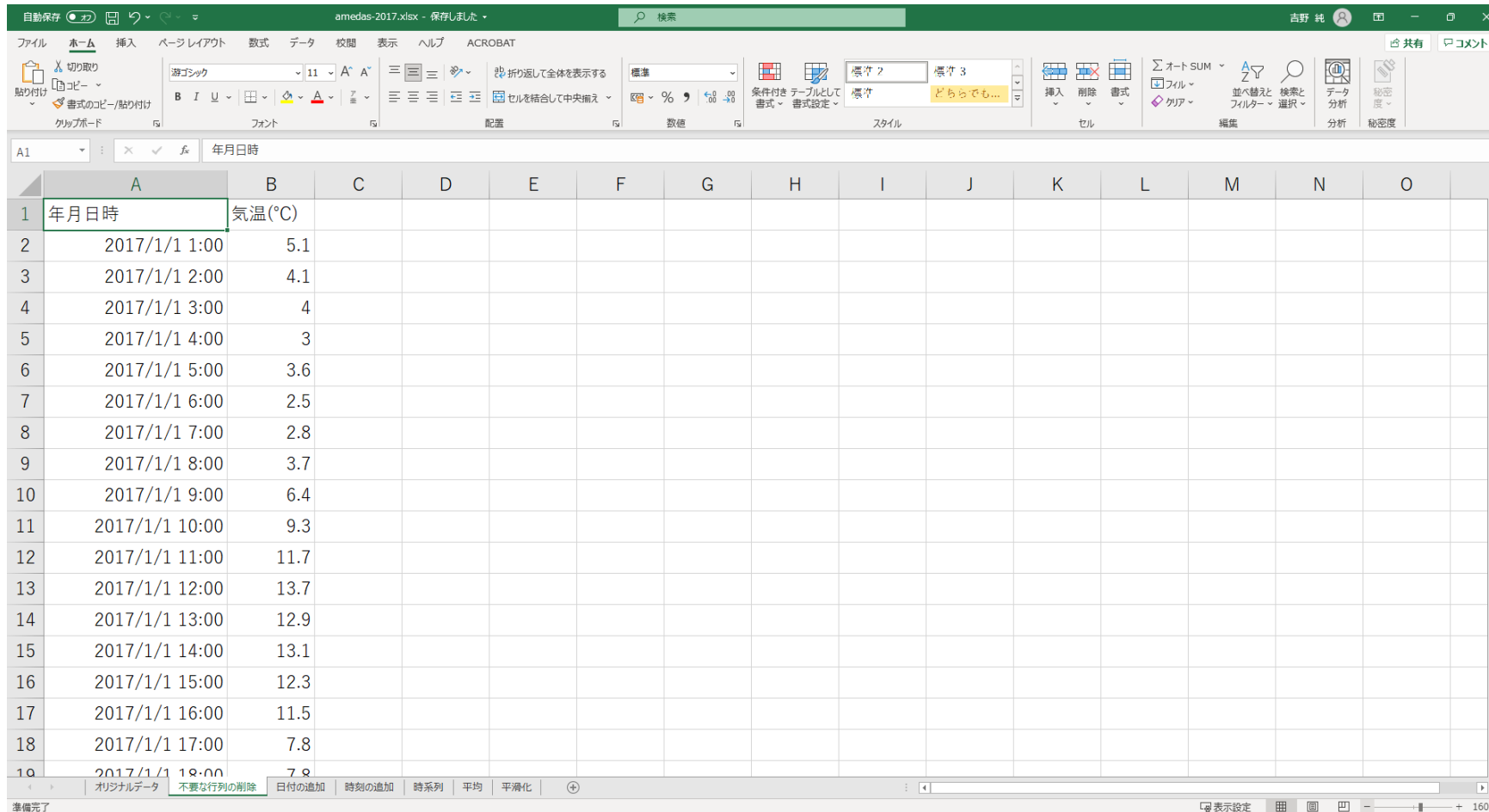
年月日時	気温(°C)	品質情報	均質番号
2017/1/1 1:00	5.1	8	1
2017/1/1 2:00	4.1	8	1
2017/1/1 3:00	4	8	1
2017/1/1 4:00	3	8	1
2017/1/1 5:00	3.6	8	1
2017/1/1 6:00	2.5	8	1
2017/1/1 7:00	2.8	8	1
2017/1/1 8:00	3.7	8	1
2017/1/1 9:00	6.4	8	1
2017/1/1 10:00	9.3	8	1
2017/1/1 11:00	11.7	8	1
2017/1/1 12:00	13.7	8	1
2017/1/1 13:00	12.9	8	1
2017/1/1 14:00	12.1	8	1

4_data/amedas-2017.xlsxに一連の処理が記録されています

不要な行や列は削除しましょう

データ分析に不要な行は削除しましょう。

不要な列や行の先頭（灰色のところ）で右クリックで削除（D）を選びます。



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	年月日時	気温(°C)													
2	2017/1/1 1:00	5.1													
3	2017/1/1 2:00	4.1													
4	2017/1/1 3:00	4													
5	2017/1/1 4:00	3													
6	2017/1/1 5:00	3.6													
7	2017/1/1 6:00	2.5													
8	2017/1/1 7:00	2.8													
9	2017/1/1 8:00	3.7													
10	2017/1/1 9:00	6.4													
11	2017/1/1 10:00	9.3													
12	2017/1/1 11:00	11.7													
13	2017/1/1 12:00	13.7													
14	2017/1/1 13:00	12.9													
15	2017/1/1 14:00	13.1													
16	2017/1/1 15:00	12.3													
17	2017/1/1 16:00	11.5													
18	2017/1/1 17:00	7.8													
19	2017/1/1 18:00	7.8													

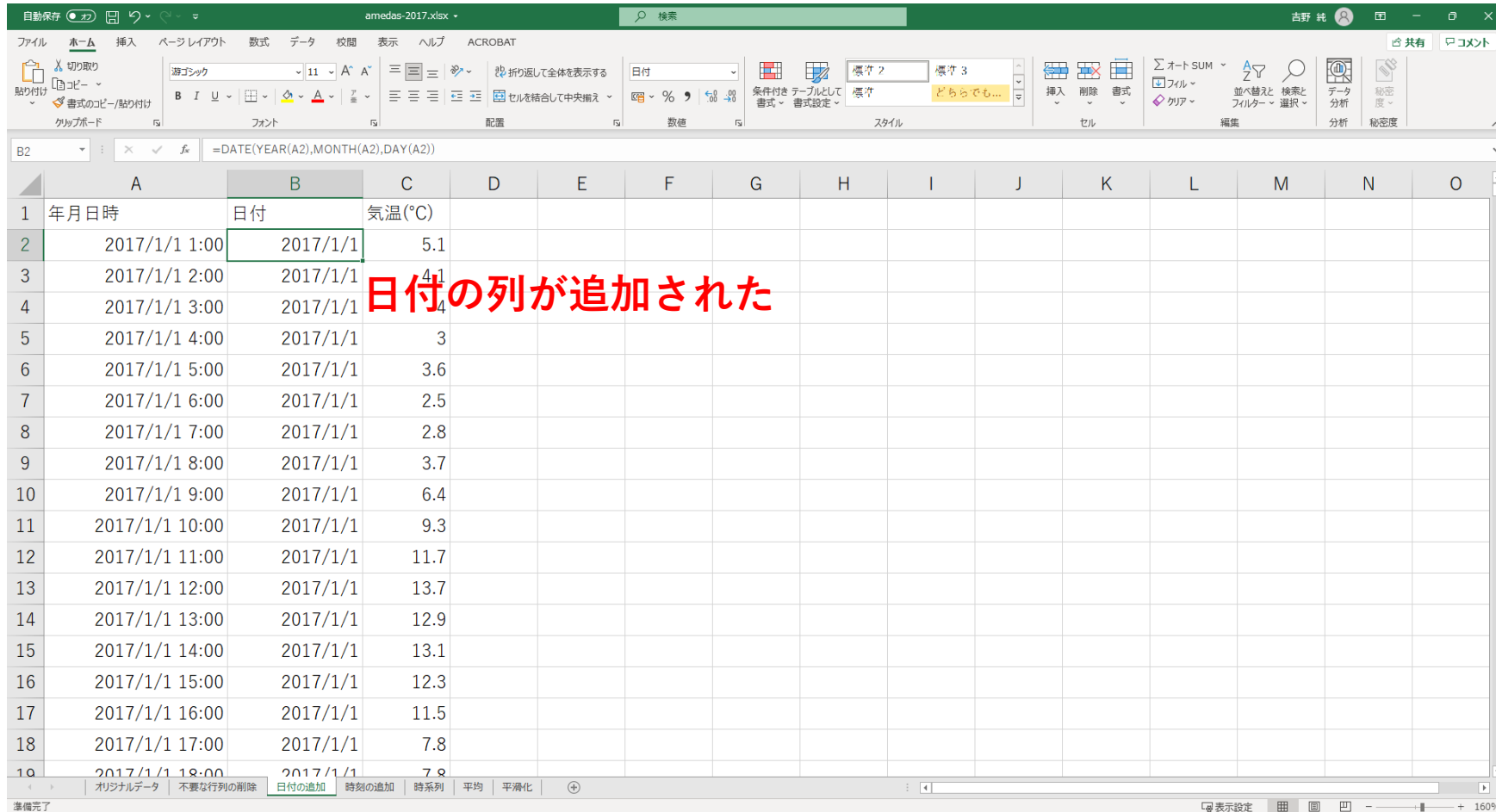
4_data/amedas-2017.xlsxに一連の処理が記録されています

日付を追加しましょう

元々、年月日時分のデータがA列にありますが、DATE関数で日付のデータをB列に追加しましょう。

=DATE(YEAR(A2),MONTH(A2),DAY(A2))

年 月 日にち



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	年月日時	日付	気温(°C)												
2	2017/1/1 1:00	2017/1/1	5.1												
3	2017/1/1 2:00	2017/1/1	4.1												
4	2017/1/1 3:00	2017/1/1	4.1												
5	2017/1/1 4:00	2017/1/1	3												
6	2017/1/1 5:00	2017/1/1	3.6												
7	2017/1/1 6:00	2017/1/1	2.5												
8	2017/1/1 7:00	2017/1/1	2.8												
9	2017/1/1 8:00	2017/1/1	3.7												
10	2017/1/1 9:00	2017/1/1	6.4												
11	2017/1/1 10:00	2017/1/1	9.3												
12	2017/1/1 11:00	2017/1/1	11.7												
13	2017/1/1 12:00	2017/1/1	13.7												
14	2017/1/1 13:00	2017/1/1	12.9												
15	2017/1/1 14:00	2017/1/1	13.1												
16	2017/1/1 15:00	2017/1/1	12.3												
17	2017/1/1 16:00	2017/1/1	11.5												
18	2017/1/1 17:00	2017/1/1	7.8												
19	2017/1/1 18:00	2017/1/1	7.8												

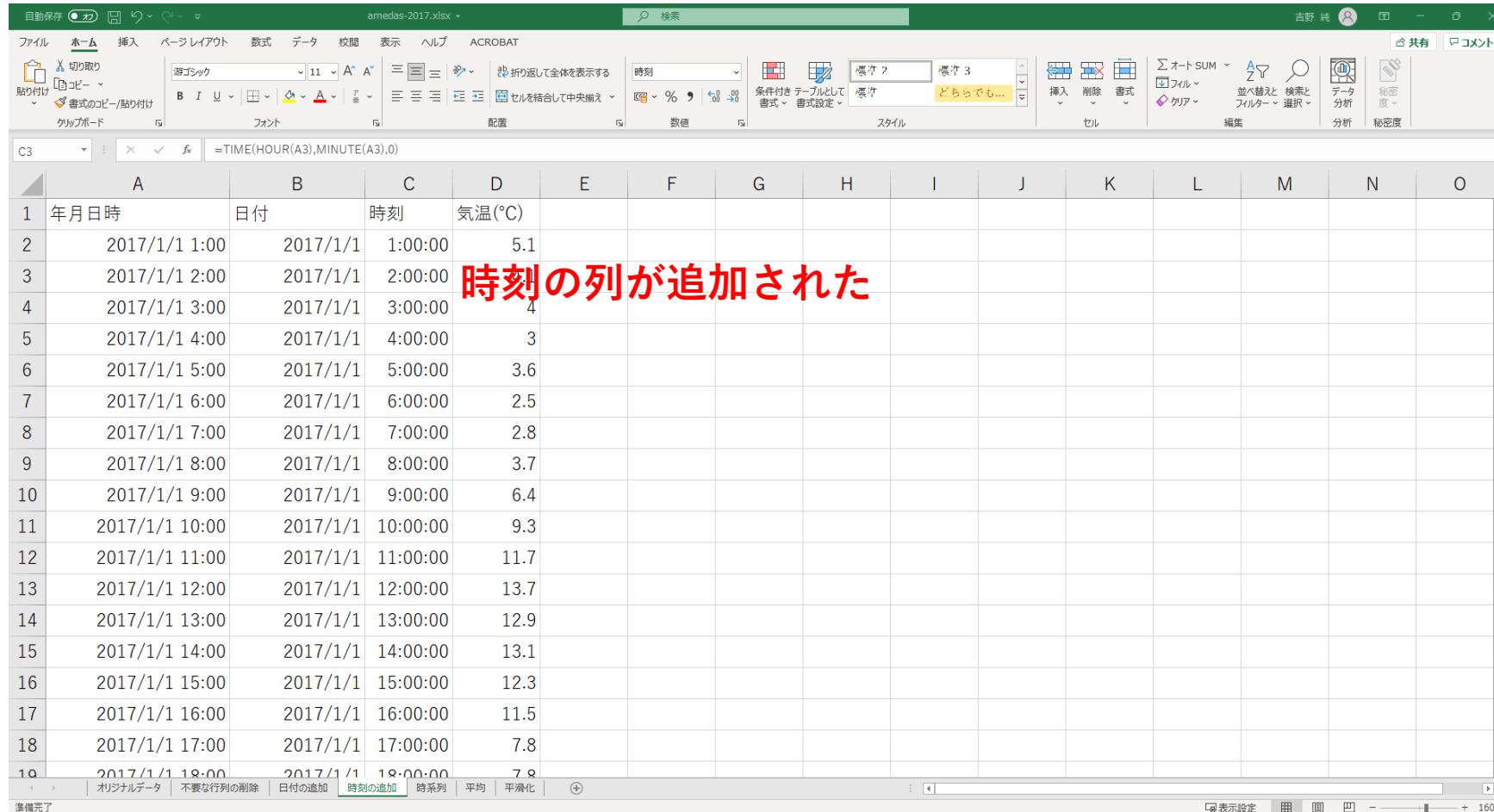
4_data/amedas-2017.xlsxに一連の処理が記録されています

時刻を追加しましょう

元々、年月日時分のデータがA列にありますが、TIME関数で時刻のデータをC列に追加しましょう。

=TIME(HOUR(A2),MINUTE(A2),0)

時間 分 秒

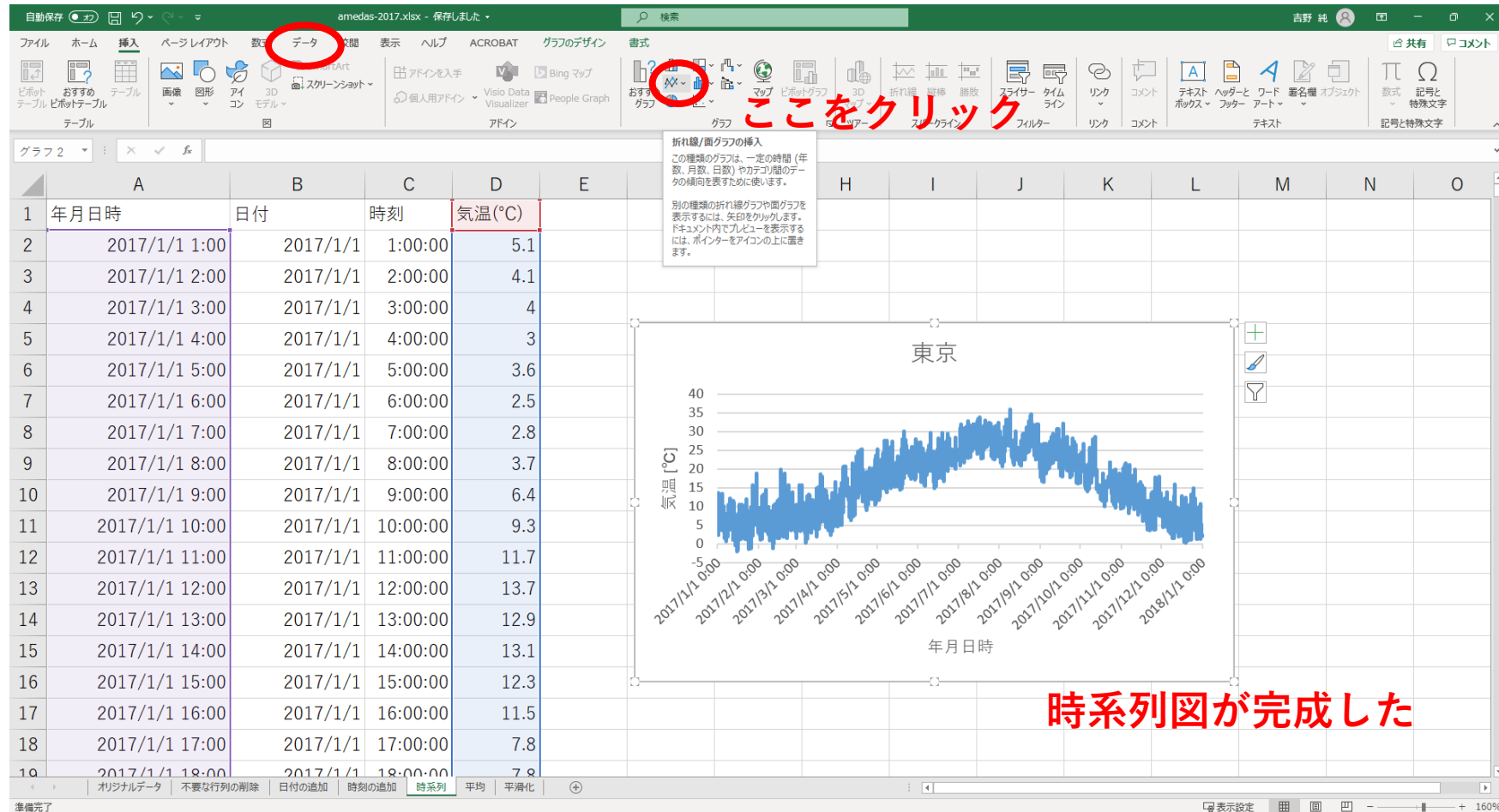


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	年月日時	日付	時刻	気温(°C)											
2	2017/1/1 1:00	2017/1/1	1:00:00	5.1											
3	2017/1/1 2:00	2017/1/1	2:00:00												
4	2017/1/1 3:00	2017/1/1	3:00:00	4											
5	2017/1/1 4:00	2017/1/1	4:00:00	3											
6	2017/1/1 5:00	2017/1/1	5:00:00	3.6											
7	2017/1/1 6:00	2017/1/1	6:00:00	2.5											
8	2017/1/1 7:00	2017/1/1	7:00:00	2.8											
9	2017/1/1 8:00	2017/1/1	8:00:00	3.7											
10	2017/1/1 9:00	2017/1/1	9:00:00	6.4											
11	2017/1/1 10:00	2017/1/1	10:00:00	9.3											
12	2017/1/1 11:00	2017/1/1	11:00:00	11.7											
13	2017/1/1 12:00	2017/1/1	12:00:00	13.7											
14	2017/1/1 13:00	2017/1/1	13:00:00	12.9											
15	2017/1/1 14:00	2017/1/1	14:00:00	13.1											
16	2017/1/1 15:00	2017/1/1	15:00:00	12.3											
17	2017/1/1 16:00	2017/1/1	16:00:00	11.5											
18	2017/1/1 17:00	2017/1/1	17:00:00	7.8											
19	2017/1/1 18:00	2017/1/1	18:00:00	7.8											

4_data/amedas-2017.xlsxに一連の処理が記録されています

時系列図を作成しましょう

年月日時（A列）と気温（D列）を選択して、時系列図を作成しましょう。



4_data/amedas-2017.xlsxに一連の処理が記録されています

日平均値を作成しましょう

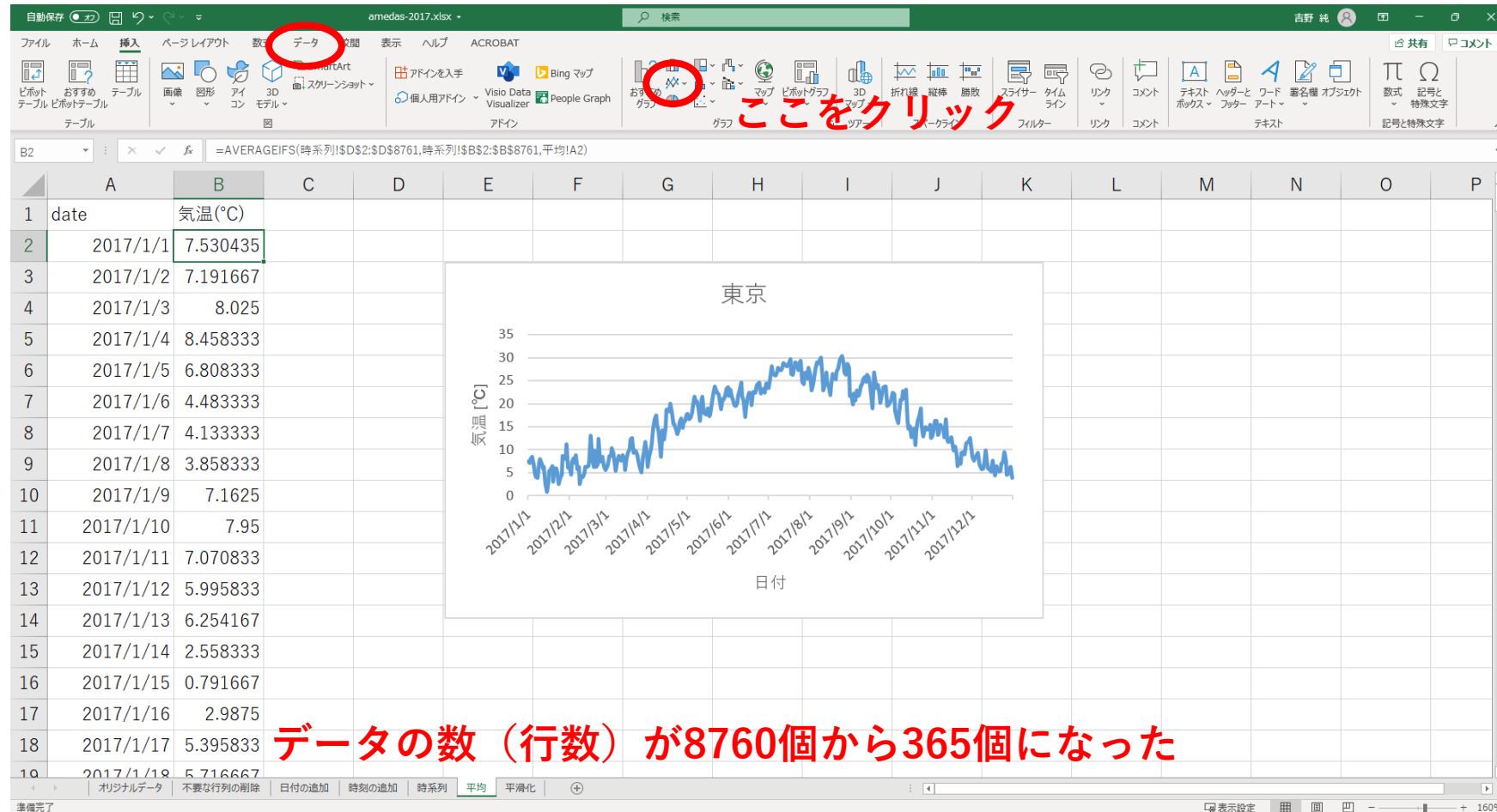
先のシート（'時系列'）の中の気温（D列）に対して、AVERAGEIFS関数で日平均値を計算しましょう。

=AVERAGEIFS(時系列!\$D\$2:\$D\$8761,時系列!\$B\$2:\$B\$8761,平均!A2)

平均対象範囲

条件範囲

条件

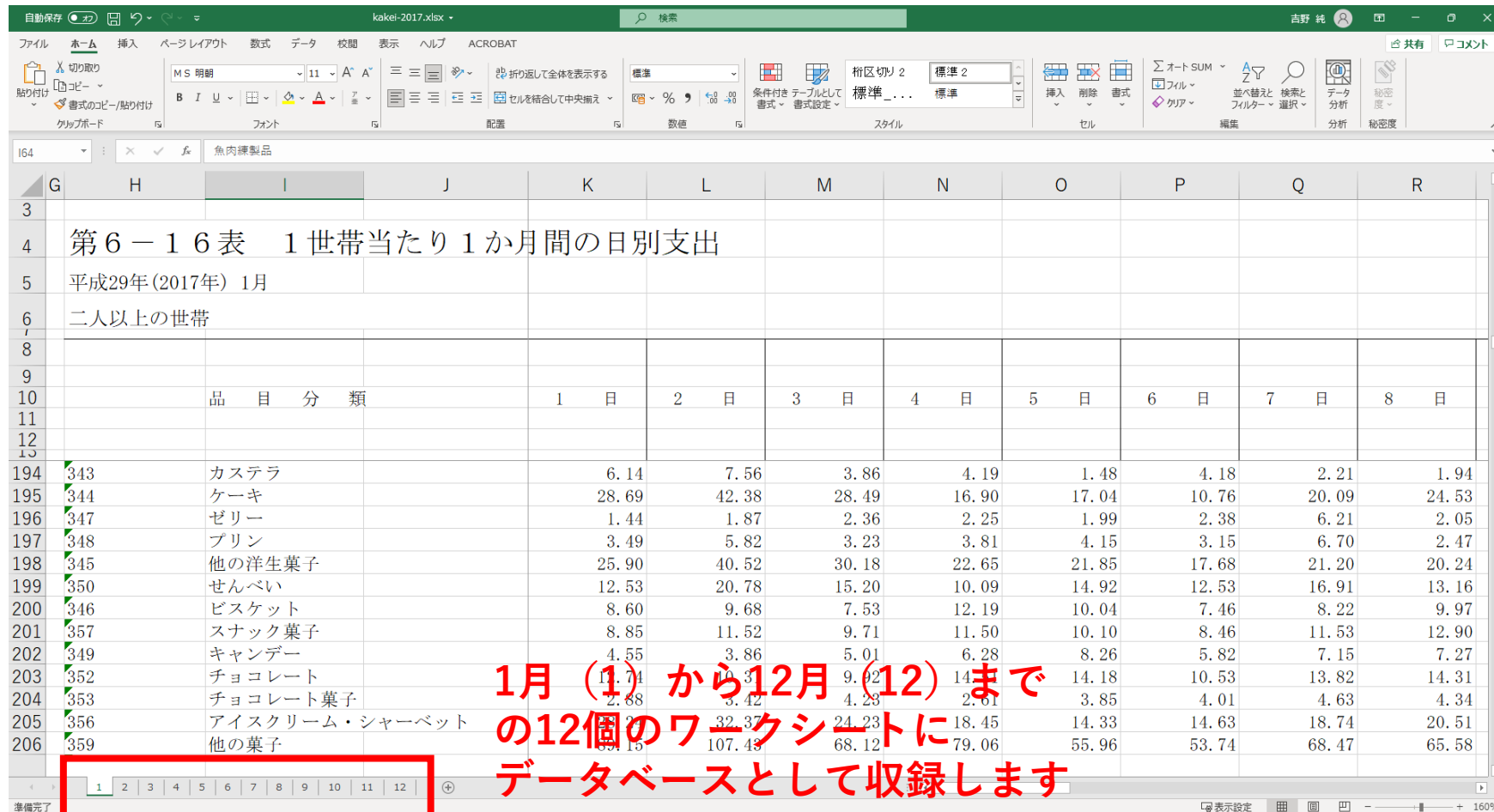


データの数（行数）が8760個から365個になった

4_data/amedas-2017.xlsxに一連の処理が記録されています

家計調査データをExcelで見てください

ダウンロードした家計調査データ「kakei-2017.xlsx」を開いてみましょう。
たくさんの品目があります。キーワード全680項目！



			1 日	2 日	3 日	4 日	5 日	6 日	7 日	8 日
343	カステラ		6.14	7.56	3.86	4.19	1.48	4.18	2.21	1.94
344	ケーキ		28.69	42.38	28.49	16.90	17.04	10.76	20.09	24.53
347	ゼリー		1.44	1.87	2.36	2.25	1.99	2.38	6.21	2.05
348	プリン		3.49	5.82	3.23	3.81	4.15	3.15	6.70	2.47
345	他の洋生菓子		25.90	40.52	30.18	22.65	21.85	17.68	21.20	20.24
350	せんべい		12.53	20.78	15.20	10.09	14.92	12.53	16.91	13.16
346	ビスケット		8.60	9.68	7.53	12.19	10.04	7.46	8.22	9.97
357	スナック菓子		8.85	11.52	9.71	11.50	10.10	8.46	11.53	12.90
349	キャンデー		4.55	3.86	5.01	6.28	8.26	5.82	7.15	7.27
352	チョコレート		11.74	10.31	9.92	14.18	10.53	13.82	14.31	
353	チョコレート菓子		2.88	3.42	4.23	2.01	3.85	4.01	4.63	4.34
356	アイスクリーム・シャーベット		32.37	24.23	18.45	14.33	14.63	18.74	20.51	
359	他の菓子		107.45	68.12	79.06	55.96	53.74	68.47	65.58	

4_data/kakei-2017.xlsxに一連の処理が記録されています

検索により必要なデータを抽出しましょう



「アイスクリーム・シャーベット」に該当する日別値を1月～12月のデータベースの中から抽出します。
=VLOOKUP(\$B\$1,INDIRECT(MONTH(A3)&"!\$I\$14:\$AO\$692"), 2+DAY(A3), FALSE)
検索値 検索範囲 抽出する列番号 完全一致検索

ワークシート「抽出」

セルA3から月の値（この場合1）を取り出して下の「1」のタブのデータベースの中から、

「アイスクリーム・シャーベット」の支出額を1年分抽出できた

date	支出（二人以上の世帯）
2017/1/1	28.24
2017/1/2	32.37
2017/1/3	24.23
2017/1/4	18.45
2017/1/5	14.33
2017/1/6	14.63
2017/1/7	18.74
2017/1/8	20.51
2017/1/9	17.82
2017/1/10	12.17
2017/1/11	10.46
2017/1/12	9.78
2017/1/13	15.07
2017/1/14	16.09
2017/1/15	15.82
2017/1/16	10.55
2017/1/17	11.75

ワークシート「1」

第6-16表 1世帯当たり1か月間の日別支出
平成29年(2017年) 1月
二人以上の世帯

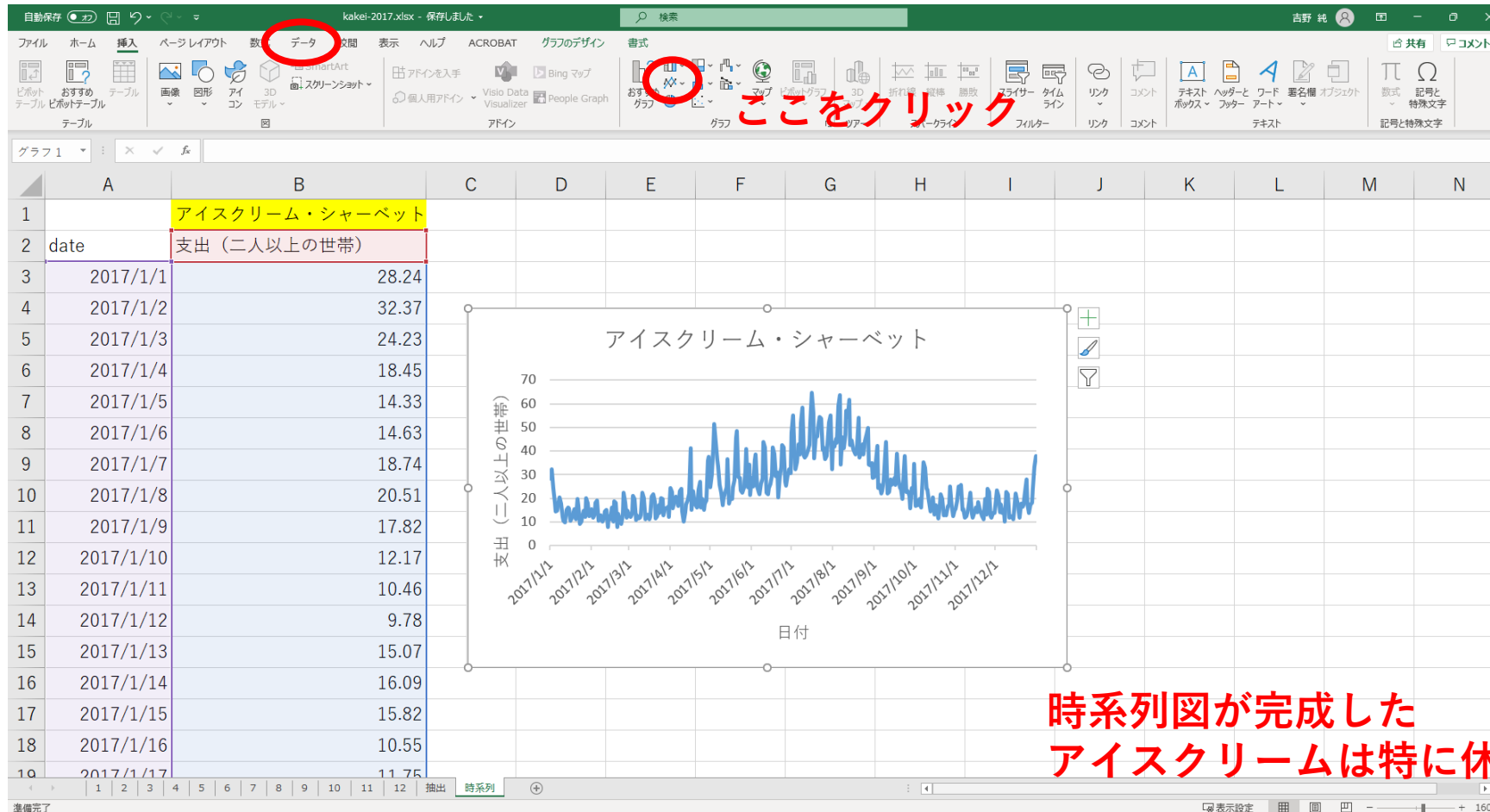
品目分類	1日	2日	3日
アイス・シャーベット	8.60	9.68	7.53
スナック菓子	8.85	11.52	9.71
キャンデー	4.55	3.86	5.01
チョコレート	12.74	10.31	9.92
アイス・シャーベット	28.24	32.37	24.23
他の菓子	69.15	107.43	68.12
調理食品	79.89	59.1	29.99
主食的調理食品	6.58	20.73	17.73
弁当	15.21	24.69	29.91
すし(弁当)	10.99	14.4	8.44
おにぎり・その他	7.98	9.81	11.68

セルB1のキーワードと合致する行を検索して、セルA3の日付に該当する列の値を抽出する

4_data/kakei-2017.xlsxに一連の処理が記録されています

時系列図を作成しましょう

日付（A列）と支出（B列）を選択して、時系列図を作成しましょう。



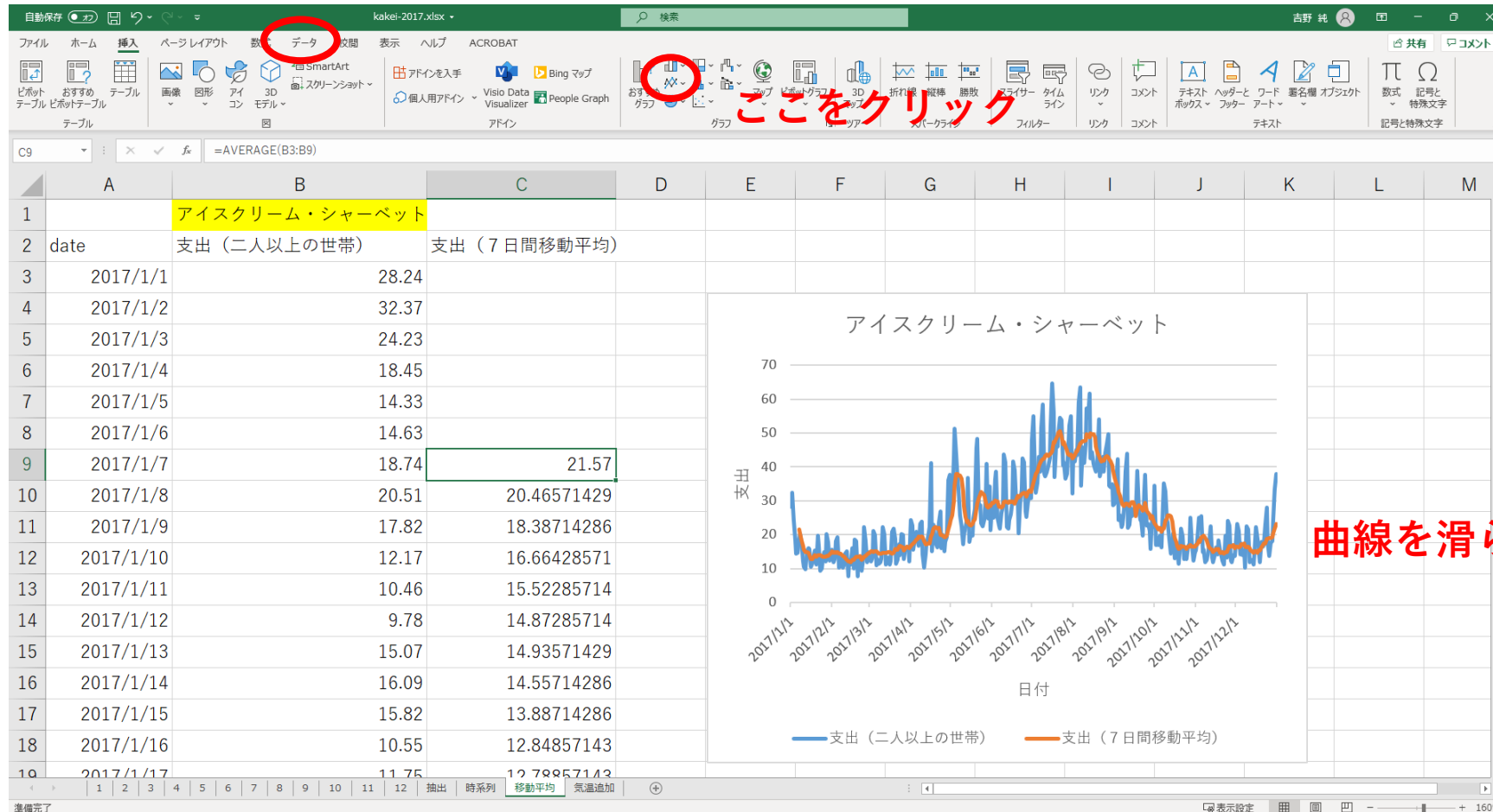
4_data/kakei-2017.xlsxに一連の処理が記録されています

データを平滑化しましょう

支出（B列）を選択して、AVERAGE関数で7日間移動平均を計算しましょう。

=AVERAGE(B3:B9)

平均対象範囲(7日間)

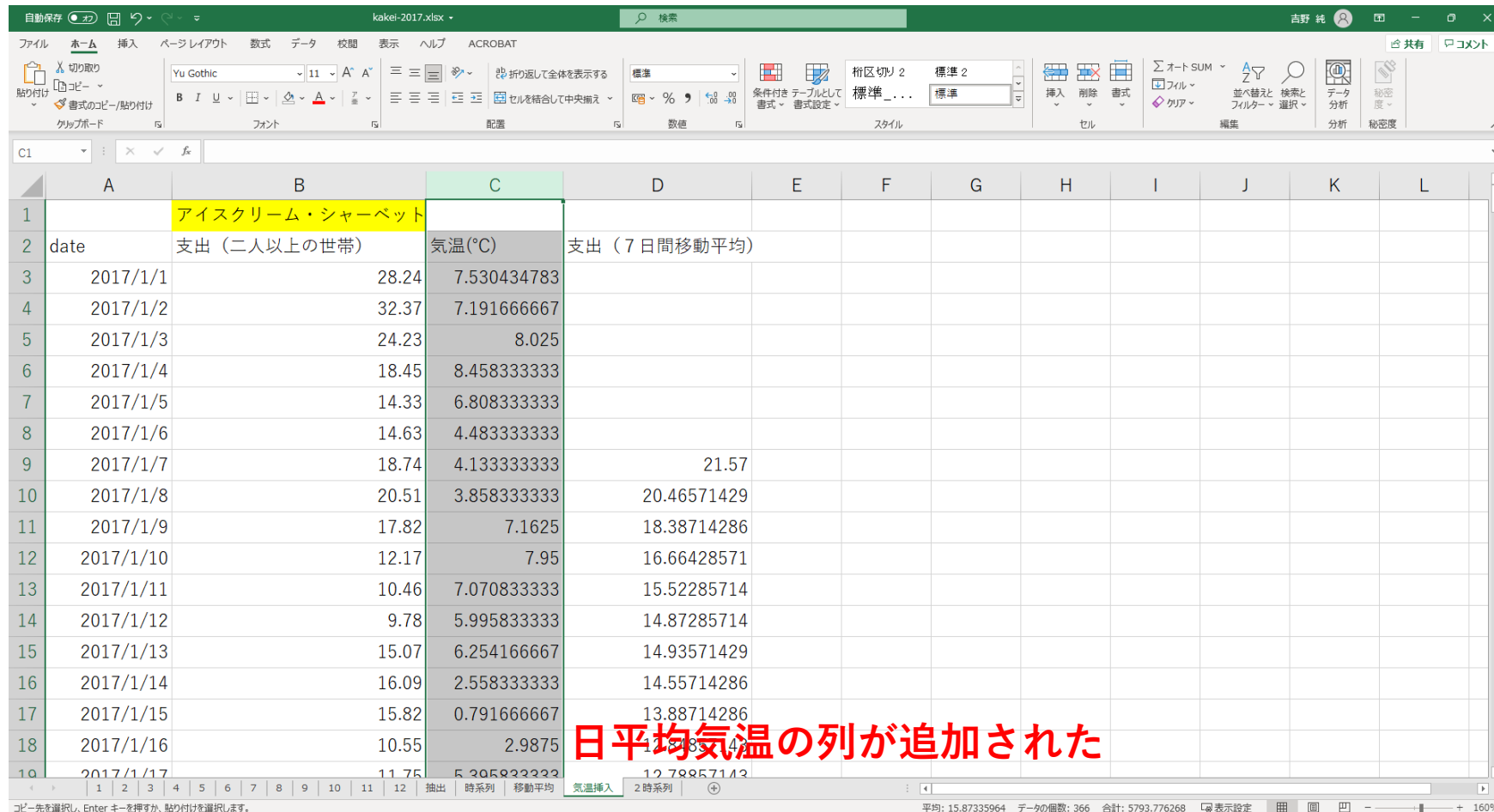


4_data/kakei-2017.xlsxに一連の処理が記録されています

日平均気温のデータを並べましょう

先に作成した日平均気温のデータをコピーして、C列に挿入しましょう。

アイスクリーム・シャーベットの購入量と日平均気温との間に関係性があると仮説を立てて比較してみましょう。



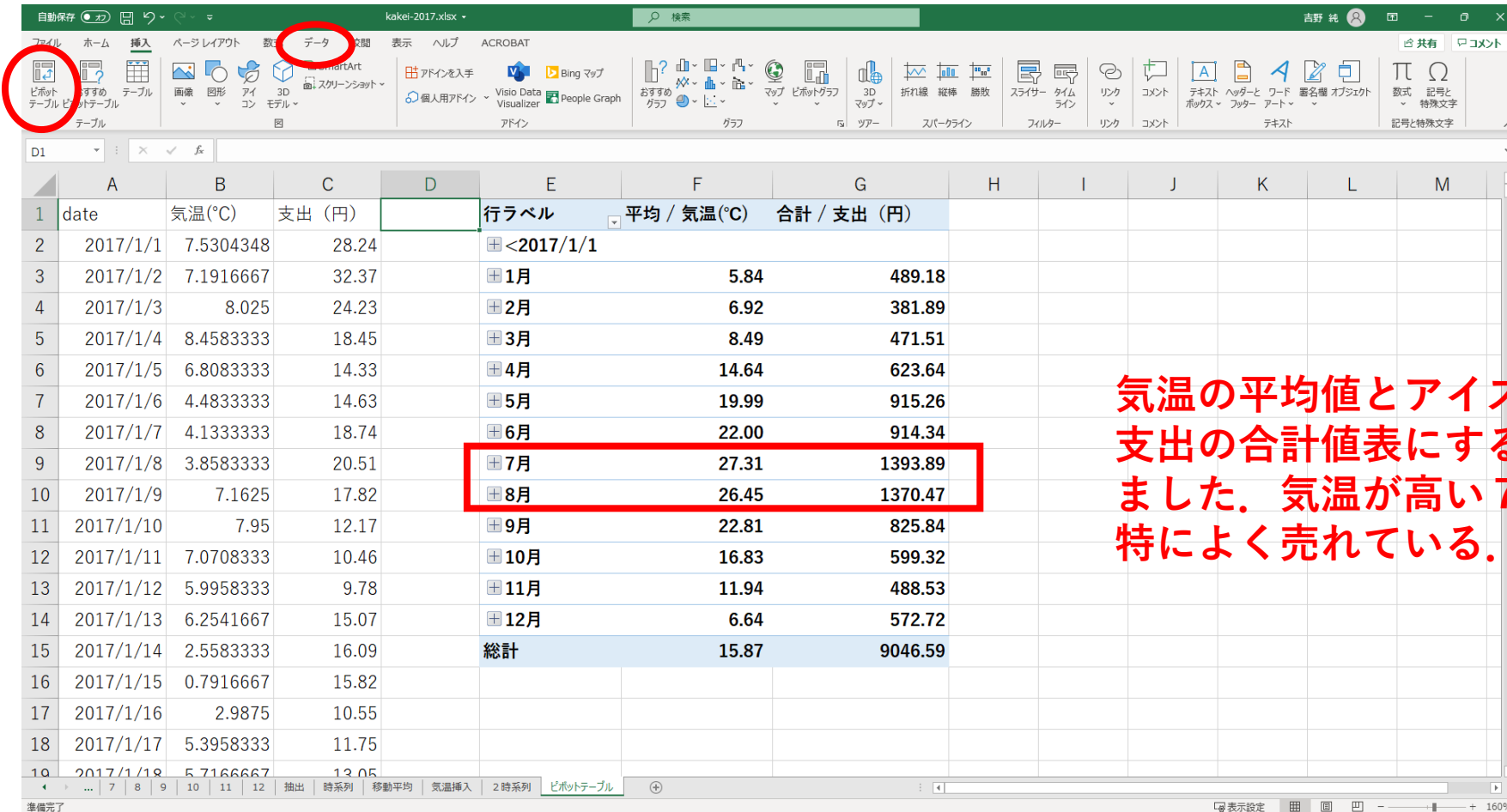
	A	B	C	D	E	F	G	H	I	J	K	L
1		アイスクリーム・シャーベット										
2	date	支出（二人以上の世帯）	気温(°C)	支出（7日間移動平均）								
3	2017/1/1	28.24	7.530434783									
4	2017/1/2	32.37	7.191666667									
5	2017/1/3	24.23	8.025									
6	2017/1/4	18.45	8.458333333									
7	2017/1/5	14.33	6.808333333									
8	2017/1/6	14.63	4.483333333									
9	2017/1/7	18.74	4.133333333	21.57								
10	2017/1/8	20.51	3.858333333	20.46571429								
11	2017/1/9	17.82	7.1625	18.38714286								
12	2017/1/10	12.17	7.95	16.66428571								
13	2017/1/11	10.46	7.070833333	15.52285714								
14	2017/1/12	9.78	5.995833333	14.87285714								
15	2017/1/13	15.07	6.254166667	14.93571429								
16	2017/1/14	16.09	2.558333333	14.55714286								
17	2017/1/15	15.82	0.791666667	13.88714286								
18	2017/1/16	10.55	2.9875	12.97857143								
19	2017/1/17	11.75	5.305833333	12.78857143								

日平均気温の列が追加された

4_data/kakei-2017.xlsxに一連の処理が記録されています

データを集計しましょう

日付（A列）気温（B列）支出（C列）を選択して、ピボットテーブルによりデータを集計しましょう。
1月から12月までの各月の気温の平均値やアイスクリームの支出の合計値を表示します。

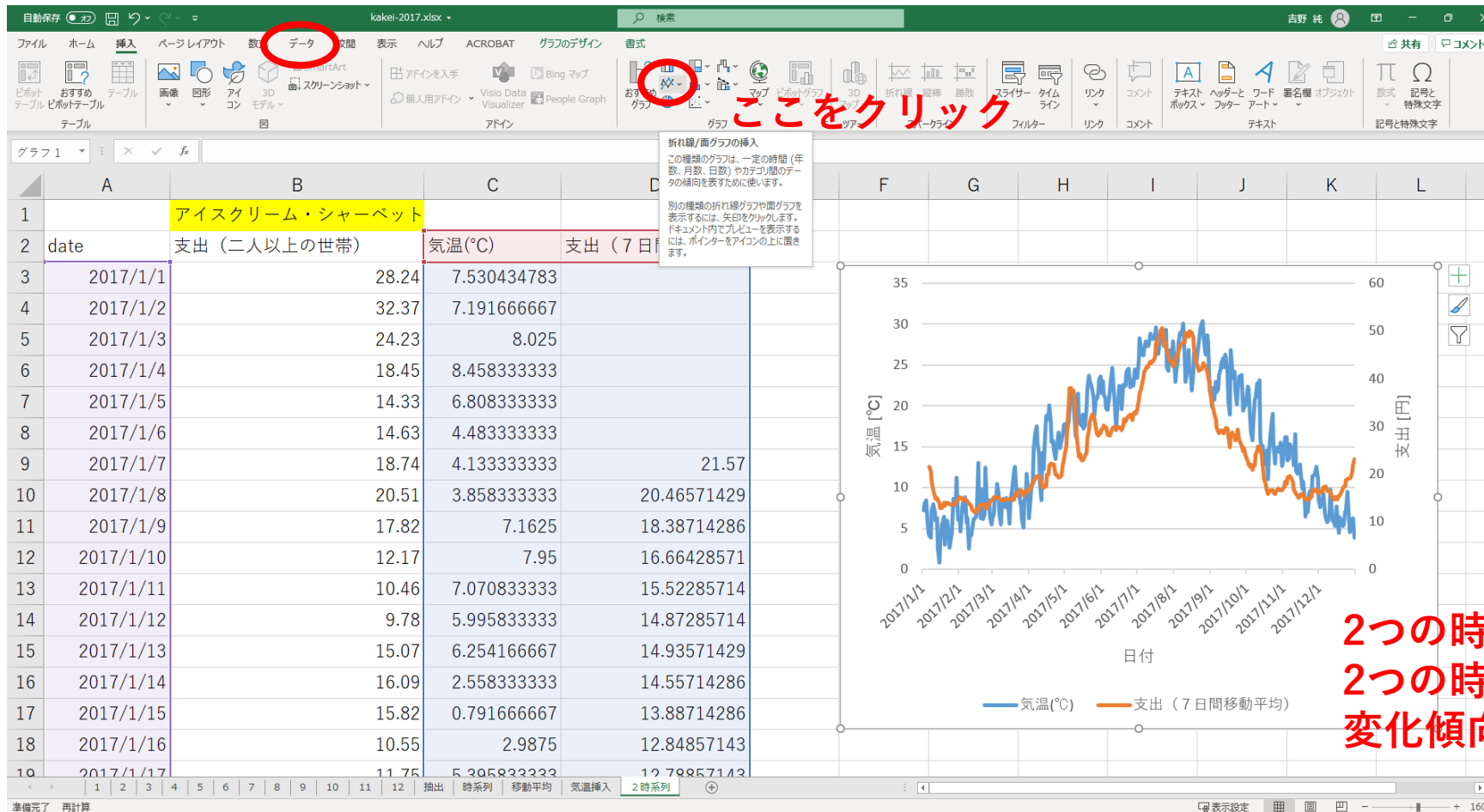


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	date	気温(°C)	支出(円)		行ラベル	平均 / 気温(°C)	合計 / 支出(円)						
2	2017/1/1	7.5304348	28.24		<2017/1/1								
3	2017/1/2	7.1916667	32.37		1月	5.84	489.18						
4	2017/1/3	8.025	24.23		2月	6.92	381.89						
5	2017/1/4	8.4583333	18.45		3月	8.49	471.51						
6	2017/1/5	6.8083333	14.33		4月	14.64	623.64						
7	2017/1/6	4.4833333	14.63		5月	19.99	915.26						
8	2017/1/7	4.1333333	18.74		6月	22.00	914.34						
9	2017/1/8	3.8583333	20.51		7月	27.31	1393.89						
10	2017/1/9	7.1625	17.82		8月	26.45	1370.47						
11	2017/1/10	7.95	12.17		9月	22.81	825.84						
12	2017/1/11	7.0708333	10.46		10月	16.83	599.32						
13	2017/1/12	5.9958333	9.78		11月	11.94	488.53						
14	2017/1/13	6.2541667	15.07		12月	6.64	572.72						
15	2017/1/14	2.5583333	16.09		総計	15.87	9046.59						
16	2017/1/15	0.7916667	15.82										
17	2017/1/16	2.9875	10.55										
18	2017/1/17	5.3958333	11.75										
19	2017/1/18	5.7166667	13.05										

気温の平均値とアイスクリームの支出の合計値表にすることができました。気温が高い7月や8月に特によく売れている。

時系列図を作成しましょう

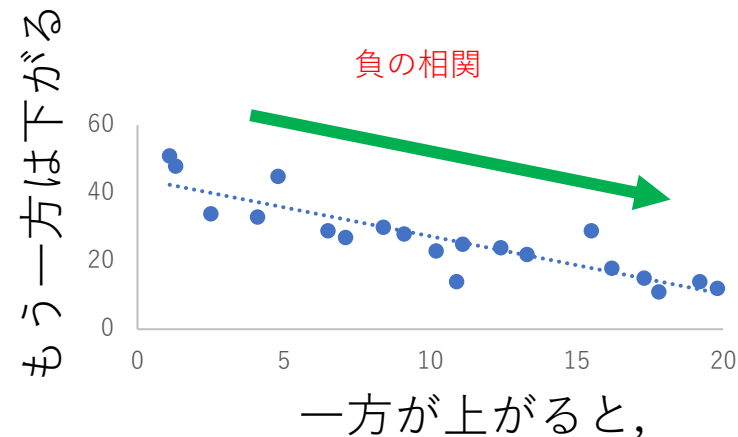
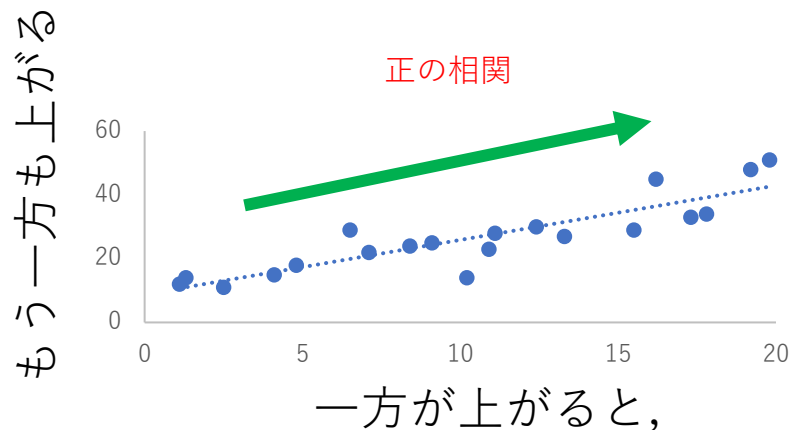
日付（A列）気温（C列）支出（D列）を選択して、2つの時系列を重ねて表示しましょう。
支出の時系列図には第2軸を表示させましょう。



4_data/kakei-2017.xlsxに一連の処理が記録されています

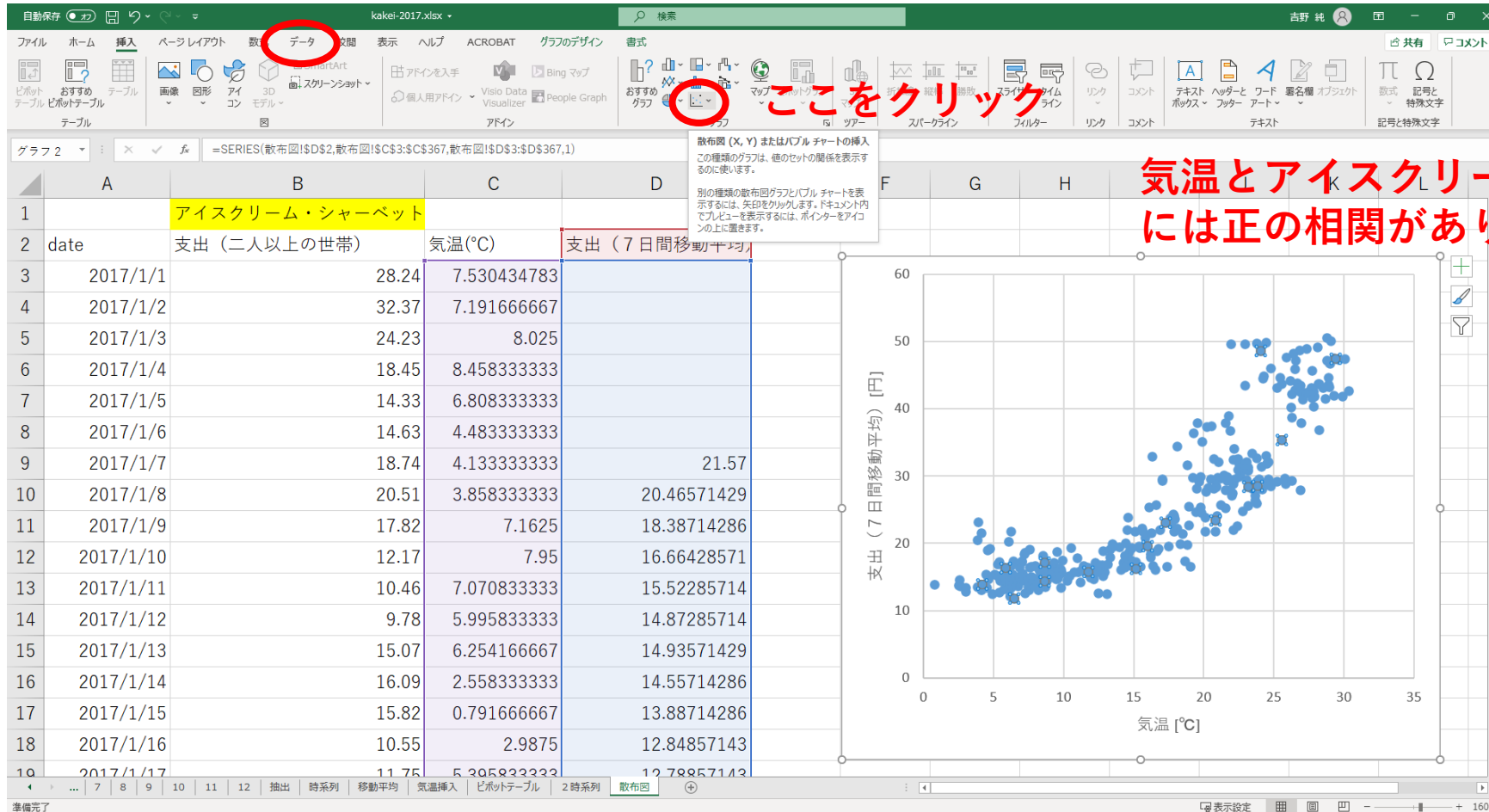
散布図を作成しましょう

- 時系列図やピボットテーブルから、気温とアイスクリームの支出の間から関係性を想像することができました。「一方が上がれば、もう一方も上がる（**正の相関**）」という関係や「一方が上がれば、もう一方は下がる（**負の相関**）」といった関係を「**相関関係がある**」（あるいは単に「**相関がある**」）と言います。
- 先の時系列図から「気温とアイスクリームの支出には正の相関がある」と推測されます。
- 相関があるかどうかを確かめたいときには、まず、**散布図**を作成します。



散布図を作成しましょう

気温（C列）支出（D列）を選択して、散布図を作成しましょう。

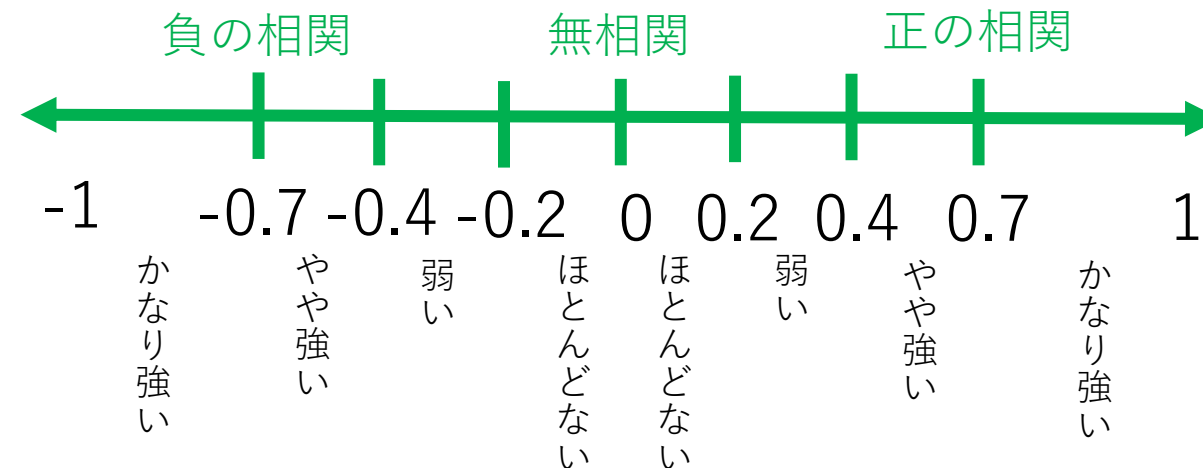


気温とアイスクリームの支出との間には正の相関がありそうだ。

4_data/kakei-2017.xlsxに一連の処理が記録されています

相関係数を計算しましょう

- 散布図により正の相関があることが読み取れたが、それを数字などで示せないとなかなか人にはなかなか納得してもらえません。
- 相関関係の強さを表す指標に、「**相関係数**」があり、 $-1 \sim 1$ の数字で表します。相関係数が1に近いほど正の相関が強く、相関係数が -1 に近いほど負の相関が強くなります。
- 相関の強さを表す相関係数は、エクセルの関数「CORREL」で計算できます。

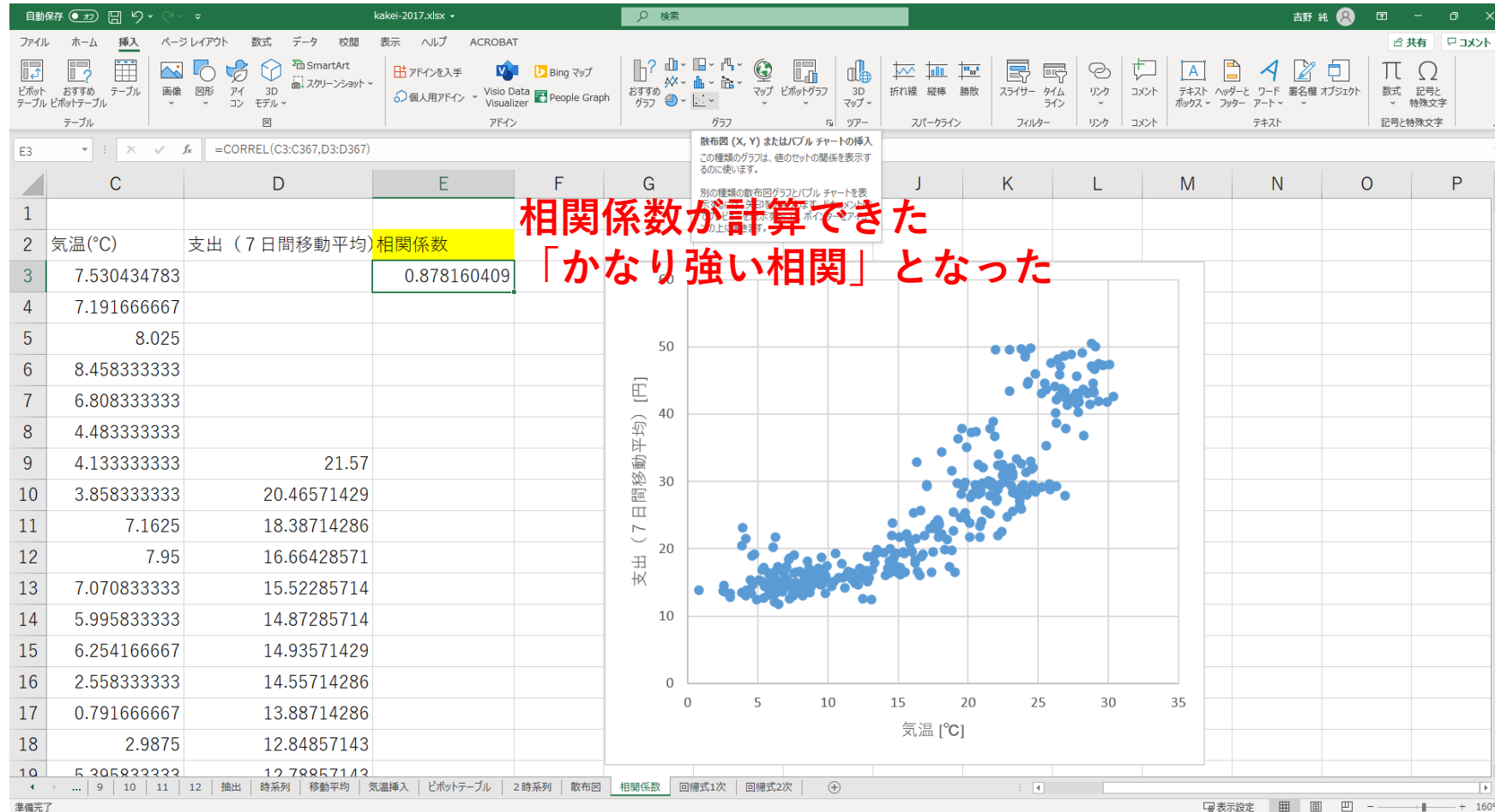


相関係数を計算しましょう

気温（C列）と支出（D列）を選択して、CORREL関数で相関係数を計算しましょう。

=CORREL(C3:C367,D3:D367)

気温 支出



4_data/kakei-2017.xlsxに一連の処理が記録されています

回帰分析しましょう

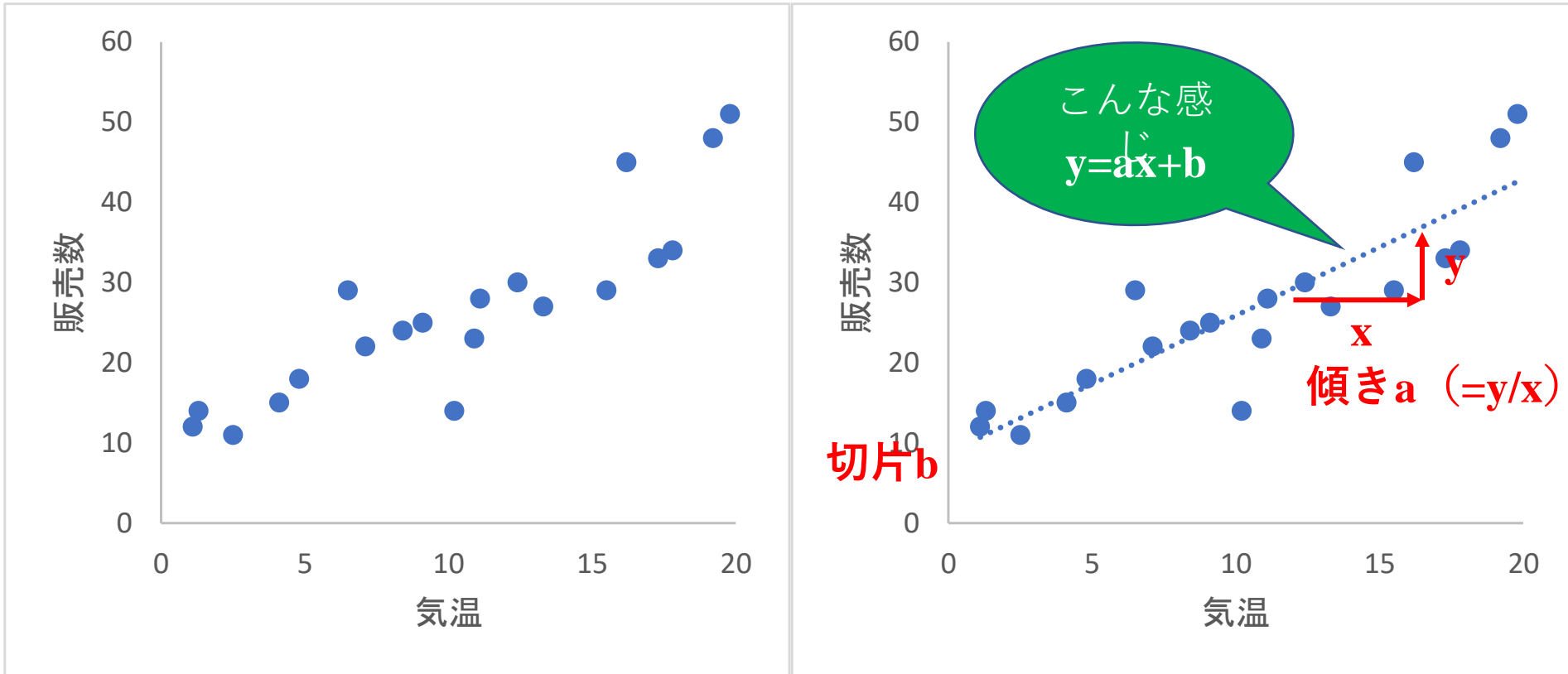
- 気温とアイスクリームの支出に正の相関があることが分かったが、実務的には気温から販売数を推定できれば、仕入れや在庫の適正化を図れる。
- このような推定を行うには「回帰分析」という方法が使える。

$$\begin{array}{|c|} \hline \text{アイスク} \\ \text{リームの} \\ \text{支出} \\ \hline \end{array} \mathbf{y} = \mathbf{a} \begin{array}{|c|} \hline \text{気温} \\ \hline \end{array} \mathbf{x} + \mathbf{b}$$

例えば線形単回帰分析の場合、気温 X （説明変数）を代入したらアイスクリームの支出 y （目的変数）を推定できるような式のパラメータ a や b を知りたい！

回帰分析しましょう

- 回帰分析では、散布図中に誤差が最も小さくなるような1本の直線や曲線で表わします。



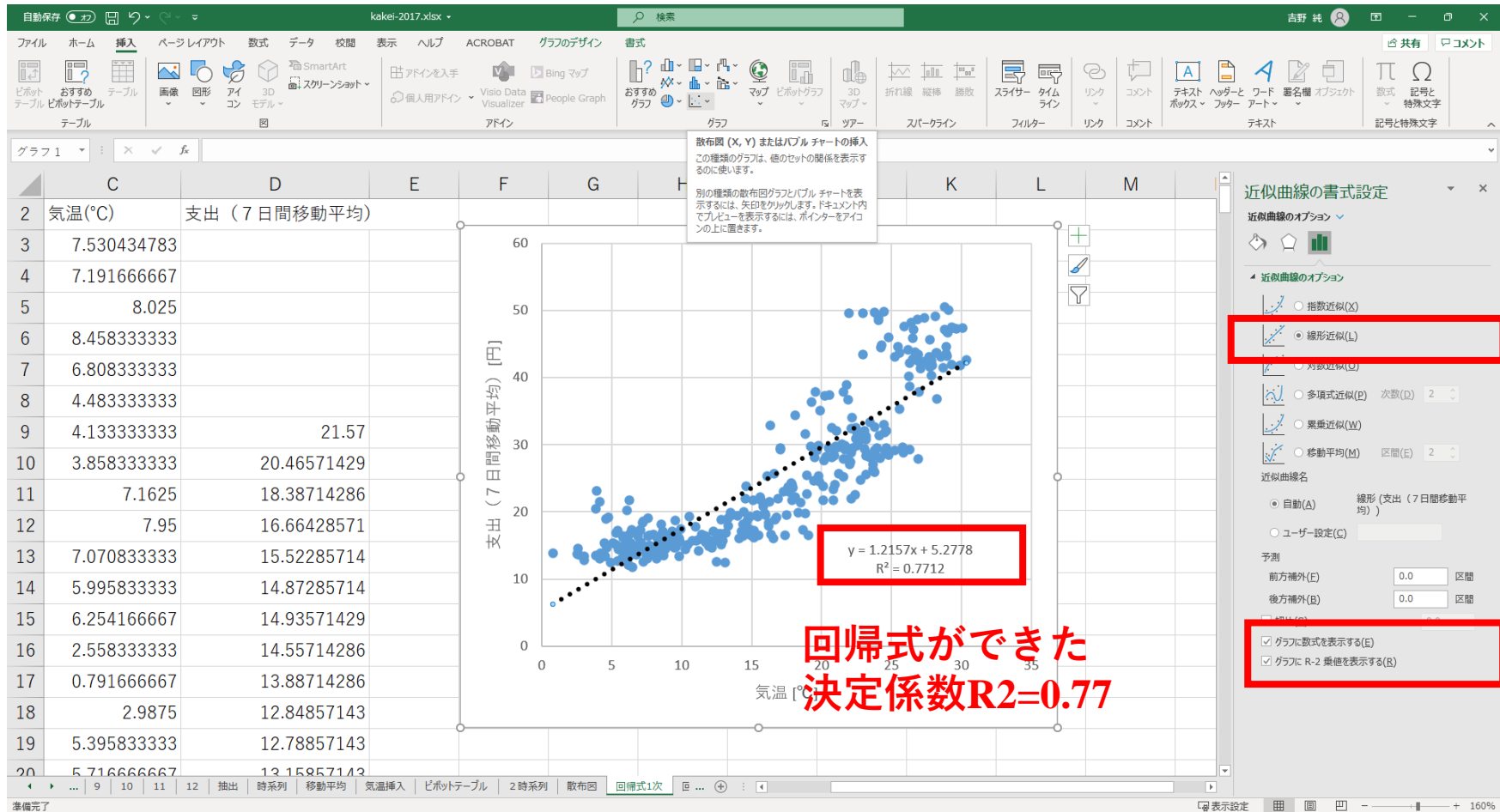
この直線のことを**回帰直線**といい、回帰直線を表す数式を**回帰式**といいます。

回帰分析しましょう

1 次関数（線形近似）による回帰曲線を引いてみましょう。

散布図中の点を右クリックして「近似曲線の追加（R）」を選びます。

アイスクリームの支出(y)と気温(x)の回帰式： $y=1.2157x+5.2778$ が得られました。



**線形近似に
チェック**

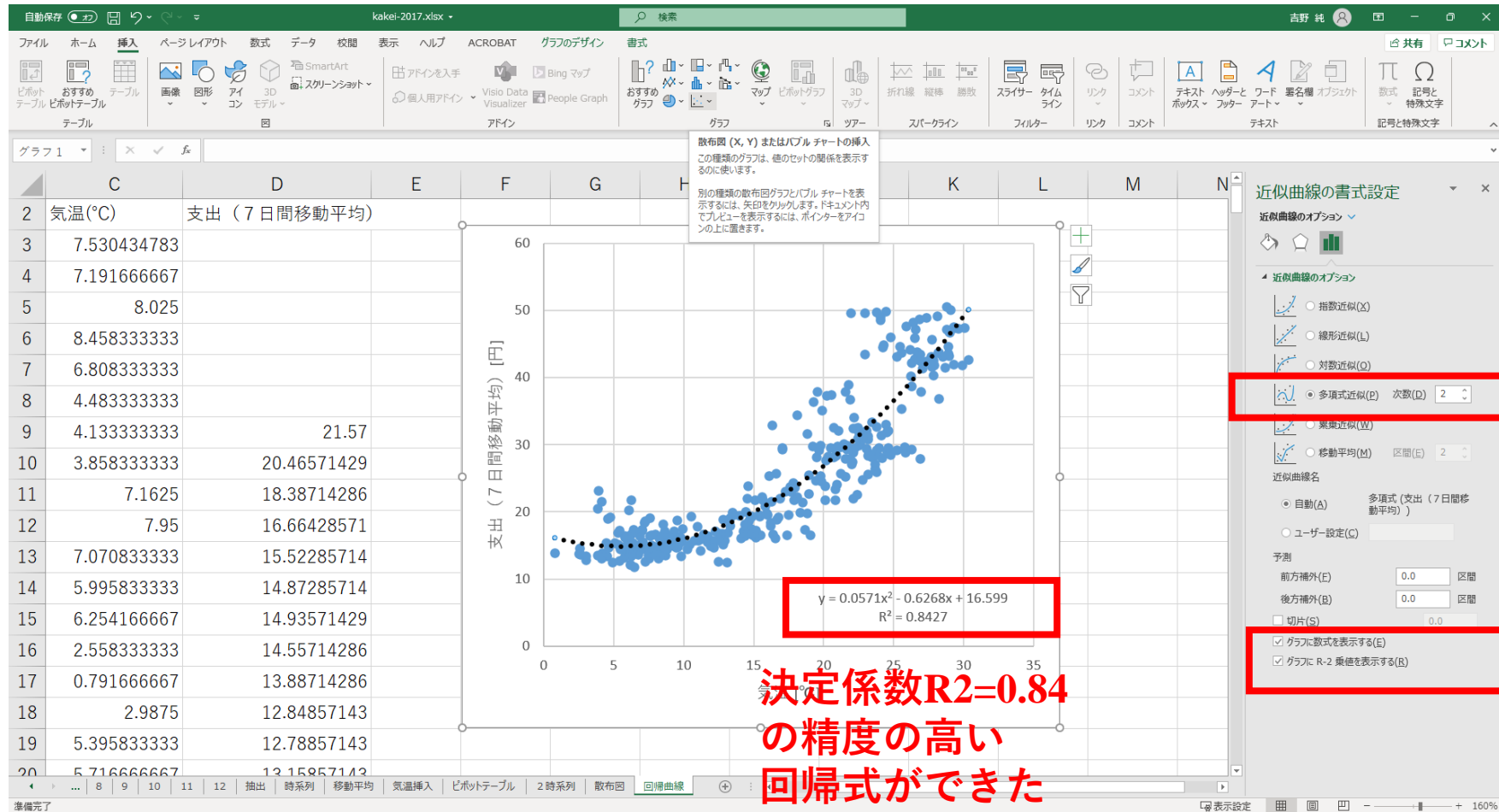
**回帰式ができた
決定係数R2=0.77**

**数式とR-2乗値
を表示する**

4_data/kakei-2017.xlsxに一連の処理が記録されています

回帰分析しましょう

もっと当てはまりの良さそうな2次関数（多項式近似）による回帰曲線を引いてみましょう。
 散布図中の点を右クリックして「近似曲線の追加（R）」を選びます。
 アイスcreamの支出(y)と気温(x)の回帰式： $y=0.0571x^2-0.6268x+16.599$ が得られました。



多項式近似に
 チェック
 次数は2

数式とR-2乗値
 を表示する

4_data/kakei-2017.xlsxに一連の処理が記録されています

やってみよう 1

- 気温とアイスクリームの支出との間には関係があることが分かりました.
- アイスクリームの支出に影響を与える要素は、気温以外に何かあるでしょうか？時系列図や散布図を見ながら考えてみましょう.
- アイスクリーム以外の品目で気温データと関連性のあるものを探してみましょう. また、その回帰式を作ってみましょう.
- 気温以外の気象データ（降水量，日照時間，風速，相対湿度，全天日射量など）と関連性のある品目を探してみましょう. また，その回帰式を作ってみましょう.

