

Bobino Environmental Data Best Practices Module

June 5, 2018 Taylor Brockman

Abstract

This educational module is written for educators and students who are ready to learn best practices for analyzing, visualizing, and interpreting environmental data sets using modern technology tools. This module is designed to extend the Bobino lesson plan. Completion of the module requires either an internet connection and web browser if online, or a spreadsheet application (such as Microsoft Excel) and copy of the NOAA data if offline. If possible, each team of students should have a dedicated computer or laptop to follow along with the instructor and complete each of the steps in order to gain experience working with real data sets and analysis tools.

1. Questions to Ask Your Data

(EV to help fill in content!)

? Are you comparing or contrasting two things?

? Are you trending changes over time?

? Do you know how your data was collected?

? Is your data pre-filtered or pre-edited or averaged or are they raw data points?

? How could measurement error be introduced into your data set and is it enough to possibly affect your conclusions?

2. Acquire Data Sets

NOAA National Centers for Environmental Information

<https://www.ncdc.noaa.gov/data-access/marineocean-data>

Secure | https://www.ncdc.noaa.gov/data-access/marinocean-data


NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
 

Formerly the National Climatic Data Center (NCDC)... [more about NCEI](#)

[Home](#) [Climate Information](#) [Data Access](#) [Customer Support](#) [Contact](#) [About](#) [Search](#) 

Home > Data Access > Marine/Ocean Data

Marine/Ocean Data

Quick Links

- Land-Based Station
- Satellite
- Radar
- Model
- Weather Balloon
- Marine / Ocean**
 - [Global Marine Data](#)
 - [ICOADS](#)
 - [Marine Data Map Access](#)
 - [Extended Reconstructed Sea Surface Temperature \(ERSST\)](#)
 - [NOAA Merged Land Ocean Global Surface Temperature Analysis \(NOAAGlobalTemp\)](#)

NCEI receives and archives meteorological data from ships at sea, moored and drifting buoys, coastal stations, rigs, and platforms. The temporal frequency of the observations range from sub-hourly to six-hourly synoptic and are global in spatial coverage.

- [Global Marine Data](#)
Historical and current marine observations from various national and international sources.
- [International Comprehensive Ocean-Atmosphere Data Set \(ICOADS\)](#)
Largest collection of verified surface marine observations in the world.
- [NOAA Merged Land Ocean Global Surface Temperature Analysis \(NOAAGlobalTemp\)](#)
Spatially gridded ($5^{\circ} \times 5^{\circ}$) global surface temperature dataset, with monthly resolution from 1880 to present



Scientists repair a moored ocean buoy

? What is Temporal Frequency?

NOAA Merged Land Ocean Global Surface Temperature Analysis

<https://www.ncdc.noaa.gov/data-access/marinocean-data/noaa-global-surface-temperature-noaaglobaltemp>

Secure | https://www.ncdc.noaa.gov/data-access/marinocean-data/noaa-global-surface-temperature-noaaglobaltemp


NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
 

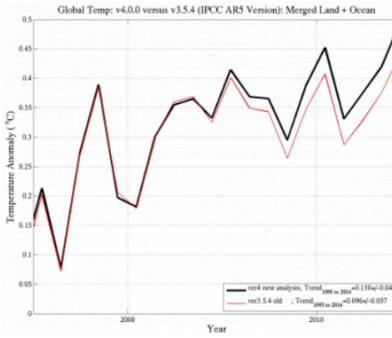
Formerly the National Climatic Data Center (NCDC)... [more about NCEI](#)

[Home](#) [Climate Information](#) [Data Access](#) [Customer Support](#) [Contact](#) [About](#) [Search](#) 

Home > Data Access > Marine / Ocean > NOAA Merged Land Ocean Global Surface Temperature Analysis (NOAAGlobalTemp)

NOAA Merged Land Ocean Global Surface Temperature Analysis (NOAAGlobalTemp)

The NOAA Merged Land Ocean Global Surface Temperature Analysis Dataset (NOAAGlobalTemp) is a merged land-ocean surface temperature analysis (formerly known as MLOST). It is a spatially gridded ($5^{\circ} \times 5^{\circ}$) global surface temperature dataset, with monthly resolution from January 1880 to present. We combine a global sea surface (water) temperature (SST) dataset with a global land surface air temperature dataset into this merged dataset of both the Earth's land and ocean surface temperatures, currently as version v4.0.1. The SST dataset is the Extended Reconstructed Sea Surface Temperature (ERSST) version



Year	ver4 new analysis (°C)	ver3.5.4 old (°C)
2000	0.15	0.15
2002	0.35	0.35
2004	0.25	0.25
2006	0.35	0.35
2008	0.30	0.30
2010	0.45	0.45

Data Access

- [Global Gridded 5° × 5° Data](#)
Monthly data in both FORTRAN IEEE binary and ASCII text formats
- [ASCII Time Series](#)
Monthly and annual land-ocean temperature time series are available from 1880 to present for several zonal bands
- [Climate at a Glance](#)
This interactive tool provides global charts and graphics using NOAA GlobalTemp data.

Data Access: ASCII Time Series

<ftp://ftp.ncdc.noaa.gov/pub/data/noaaglobaltemp/operational/>

? What is Time Series?

? What is Spatial?

Index of /pub/data/noaaglobaltemp/operational/

<ftp://ftp.ncdc.noaa.gov/pub/data/noaaglobaltemp/operational/>



Index of /pub/data/noaaglobaltemp/operational/

[\[parent directory\]](#)

Name	Size	Date Modified
ancillarydata/		5/17/18, 4:46:00 AM
gridded/		5/17/18, 4:46:00 AM
Readme.status.changes	3.6 kB	5/17/18, 4:46:00 AM
timeseries/		5/17/18, 4:46:00 AM

? What is FTP?

Index of /pub/data/noaaglobaltemp/operational/timeseries/

 [parent directory]

	Name	Size	Date Modified
	aravg.ann.land.00N.30N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.00N.90N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.20N.90N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.20S.20N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.30N.60N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.30S.00N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.60N.90N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.60S.30S.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.60S.60N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.90S.00N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.90S.20S.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.90S.60S.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land.90S.90N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land_ocean.00N.30N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land_ocean.00N.90N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land_ocean.20N.90N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land_ocean.20S.20N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land_ocean.30N.60N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM
	aravg.ann.land_ocean.30S.00N.v4.0.1.201804.asc	8.8 kB	5/17/18, 4:46:00 AM

? What do the different filenames mean?

Readme file!

<ftp://ftp.ncdc.noaa.gov/pub/data/noaaglobaltemp/operational/timeseries/readme.timeseries>

← → C ⌂ ⓘ Not Secure | <ftp://ftp.ncdc.noaa.gov/pub/data/noaaglobaltemp/operational/timeseries/readme.timeseries>

```

file name convention for areal average (aravg) time series:
ann=annual average
mon=monthly average
land_ocean=merged land-ocean surface temperature
land=land surface temperature
ocean=ocean surface temperature
latitudes=southern and northern limits of areal average
v=version number
yyyymm=date for the latest data

Annual data (aravg.ann.*):
1st column = year
2nd column = anomaly of temperature (K)
3rd column = total error variance (K**2)
4th column = high-frequency error variance (K**2)
5th column = low-frequency error variance (K**2)
6th column = bias error variance (K**2)

Monthly data (aravg.mon.*):
1st column = year
2nd column = month
3rd column = anomaly of temperature (K)
4th column = total error variance (K**2)
5th column = high-frequency error variance (K**2)
6th column = low-frequency error variance (K**2)
7th column = bias error variance (K**2)
8th column = diagnostic variable
9th column = diagnostic variable
10th column= diagnostic variable

NOTE: anomalies are based on the climatology from 1971 to 2000

```

? What do the columns mean?

<ftp://ftp.ncdc.noaa.gov/pub/data/noaaglobaltemp/operational/timeseries/aravg.mon.ocean.30N.60N.v4.0.1.201804.asc>

← → C ⌂ ⓘ Not Secure | <ftp://ftp.ncdc.noaa.gov/pub/data/noaaglobaltemp/operational/timeseries/aravg.mon.ocean.30N.60N.v4.0.1.201804.asc> ⌂

1880	1	-0.064775	0.021532	0.008957	0.000009	0.012566	0.135365	0.267226	0.131861
1880	2	-0.096933	0.017173	0.004511	0.000009	0.012652	0.124812	0.203302	0.078490
1880	3	-0.250461	0.011067	0.000672	0.000009	0.010386	0.148818	0.167809	0.018991
1880	4	-0.107491	0.003794	0.000225	0.000009	0.003559	0.110001	0.091892	0.018108
1880	5	0.165908	0.002049	0.000225	0.000009	0.001815	0.130682	0.116295	0.015000
1880	6	0.169995	0.001355	0.000225	0.000009	0.001121	0.158259	0.150145	0.015000
1880	7	0.024341	0.002136	0.000851	0.000009	0.001276	0.217238	0.207464	0.015000
1880	8	-0.029892	0.003943	0.002316	0.000009	0.001618	0.235666	0.212304	0.023363
1880	9	-0.096522	0.002902	0.000993	0.000009	0.001900	0.211037	0.286397	0.075360
1880	10	-0.246165	0.006827	0.000248	0.000009	0.006570	0.199034	0.233178	0.034144
1880	11	-0.177689	0.019216	0.003352	0.000009	0.015855	0.172035	0.225828	0.053793
1880	12	0.179403	0.015451	0.000225	0.000009	0.015217	0.141714	0.130979	0.015000
1881	1	0.318140	0.021117	0.008957	0.000011	0.012149	0.135365	0.251950	0.116584

? Which datasets should I download locally?

3. Data Set First Impressions

? What are the boundaries of my data? What is in the ‘domain’ and what ‘ranges’ of values are available?

? What format is the data in? CSV, Tab Separated, Column Headers, Binary?

? Do you see any gaps in time?

? Do you see any obvious outliers?

? Is your data “rolled up”, pre-analyzed or are you working with raw data points?

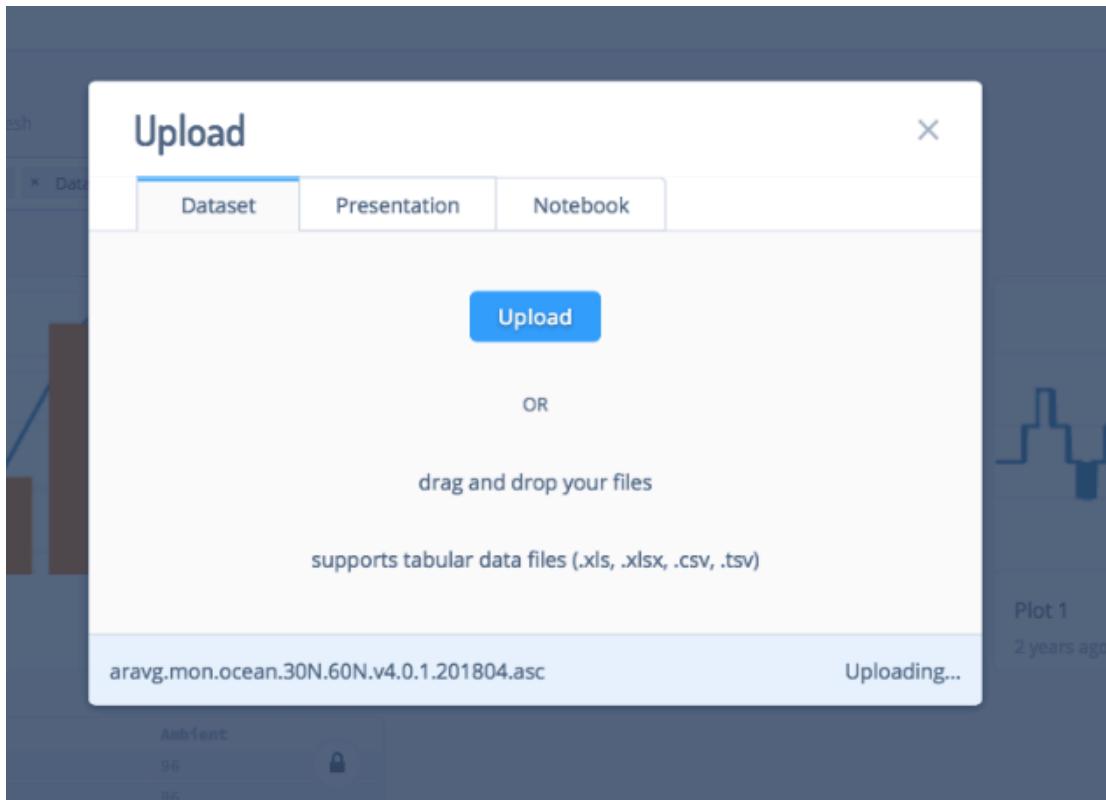
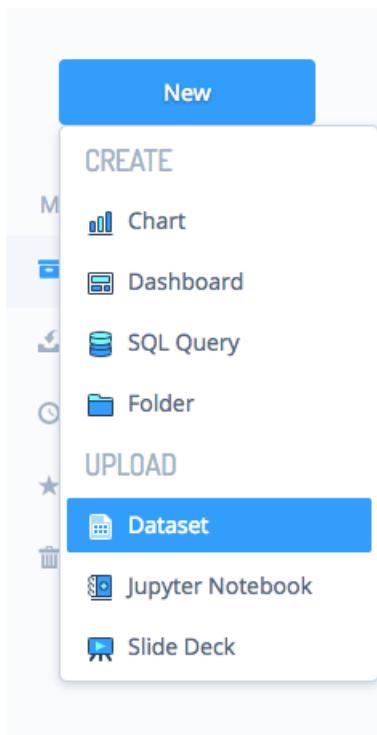
4. Online Data Visualization

<https://plot.ly/>

The screenshot shows the Plotly dashboard interface. At the top, there's a navigation bar with links for Pricing, Products, My Files, Chart Studio, and a '+ Create' button. The main area features a large, decorative background image of various data visualizations. Below this, on the left, is a sidebar with a user profile picture and the name 'taylor.brockman'. The sidebar includes sections for 'New', 'My Files' (with 'Home' selected), 'Shared with me', 'Recent', and 'Favorite'. In the center, there's a 'Home' section with a 'Refresh' button and a 'Show' dropdown menu containing options like 'Charts', 'Datasets', 'Folders', etc. To the right of the sidebar, there's a table titled 'aravg.mon.ocean.30N.60N.v4.0.1.20' with columns 'Col4', 'Col5', and 'Col12'. Next to it is a bar chart titled 'Plot 3' with orange bars. Further right is a table titled 'Grid 2' with columns 'A', 'C', and 'B'. At the bottom right, there's a 'Sort by' dropdown set to 'Modified'.

[Sign up for Free Account](#)

[Upload](#)



Uploaded aravg.mon.ocean.30N.60N.v4.0.1.201804.asc with no modifications

Dataset Appears on your Home Screen

The screenshot shows the Plotly home interface. At the top, there's a navigation bar with links for Pricing, Products, My Files, and Chart Studio, along with a '+ Create' button and a user profile for 'taylor.brockman'. Below the navigation is a sidebar titled 'My Files' containing sections for Home, Shared with me, Recent, and Favorite. A 'New' button is also in the sidebar. The main area features a chart titled 'Plot 3' showing a bar chart with orange bars and a blue line graph above it. To the right of the chart is a table titled 'A' with columns 'C' and 'B'. Below the chart and table are two smaller preview cards for 'Plot 3' and 'Grid 2'.

Browsing the Data Set in Detail

This screenshot shows the 'Data' tab of a specific dataset page. The top navigation bar includes a back arrow, a file icon, and a user profile for '@taylor.brockman'. Below the navigation are three tabs: 'Data' (which is active), 'Forking History', and 'SQL Query'. The main content area displays a large table with 16 columns labeled Col1 through Col18. The first few rows of data are visible, showing numerical values for each column across multiple rows. The table has a light gray background with white borders between cells.

Commenting on the Data Set

1654	2817	11	0.782873	0.011623	0.005125	0.000025	0.005873	0.172835	0.089299	0.082737
1655	2817	12	0.678581	0.008378	0.002355	0.000025	0.005998	0.141714	0.095131	0.046583
1656	2818	1	0.68975	0.007484	0.001884	0.000025	0.005575	0.135365	0.09078	0.044585
1657	2818	2	0.64717	0.005855	0.00095	0.000025	0.00488	0.124812	0.07587	0.048943
1658	2818	3	0.647769	0.009671	0.004306	0.000025	0.00534	0.148618	0.067261	0.081617
1659	2818	4	0.715468	0.004749	0.000225	0.000025	0.00498	0.110001	0.072322	0.037679

★ f g+ t ↗ ↘ </>

 **taylor.brockman** a few seconds ago 
Found at <ftp://ftp.ncdc.noaa.gov/pub/data/noaaglobaltemp/operational/timeseries/>

B I S H₁ H₂ H₃ 66 ≡ ≢ % Comment

Create Visualization

Secure | <https://plot.ly/create/?fid=taylor.brockman%3A4>

Graph

Create

Filter

Group

Aggregate

Style

Analysis

JSON

Export

Save

Share

Upgrade

+ Trace

Grid 1

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	K	L	M	N	O	P	Q
1	1880	1	-0.064775	0.021532	0.008957	0.000009	0.012566	0.135365	0.267226	0.131861							
2	1880	2	-0.096933	0.017173	0.004511	0.000009	0.012652	0.124812	0.203302	0.07849							
3	1880	3	-0.250461	0.011067	0.000672	0.000009	0.010386	0.148818	0.167809	0.018991							
4	1880	4	-0.107491	0.003794	0.000225	0.000009	0.003559	0.110001	0.091892	0.018108							
5	1880	5	0.465909	0.002046	0.000235	0.000009	0.004045	0.132623	0.116205	0.016145							

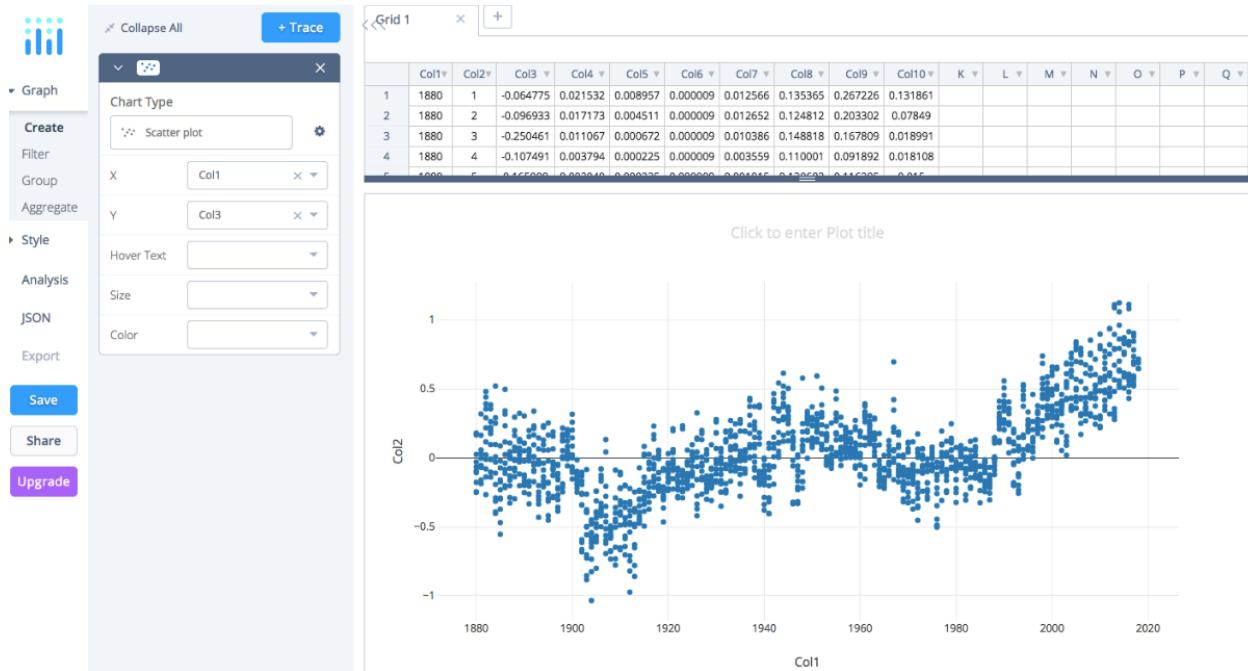
Click to enter Plot title

Click to enter Y axis title

0 0.2 0.4 0.6 0.8 1

Click to enter X axis title

Set Columns into X and Y



? Could you add a 2nd File as another data series in a different color to serve as a comparison?

? How do we add the month level detail into the X domain?

? What other visualization types help you understand the shape and the trend of this dataset?

5. Offline Data Visualization

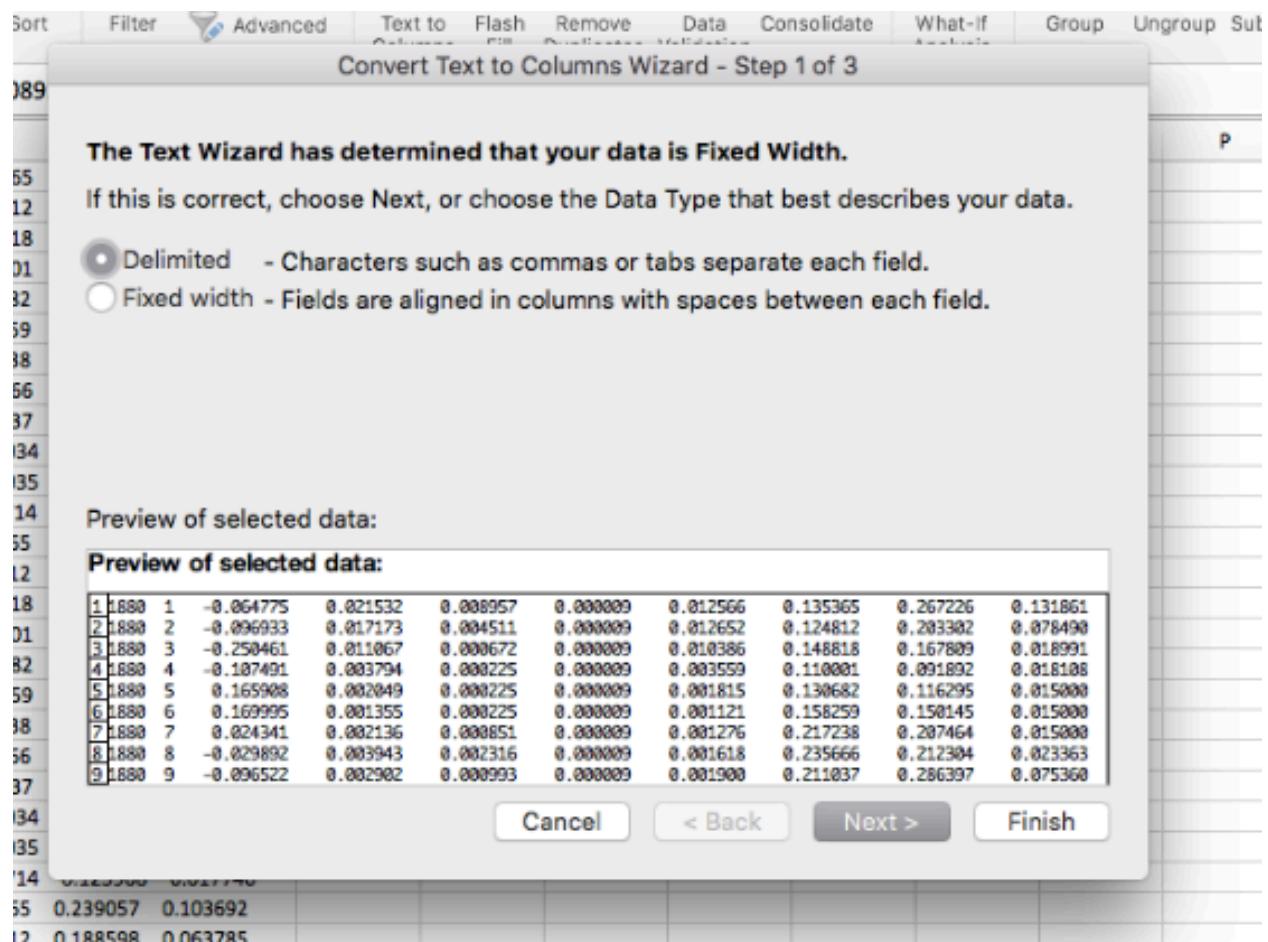
Using Microsoft Excel, the File opens as a single column per row:

Excel screenshot showing the Data tab selected. The ribbon tabs are Home, Insert, Page Layout, Formulas, Data, Review, View. The Data tab has its own set of icons for Connections, Properties, Sort, Filter, Advanced, Text to Columns, Flash Fill, Remove Duplicates, and Data Validation.

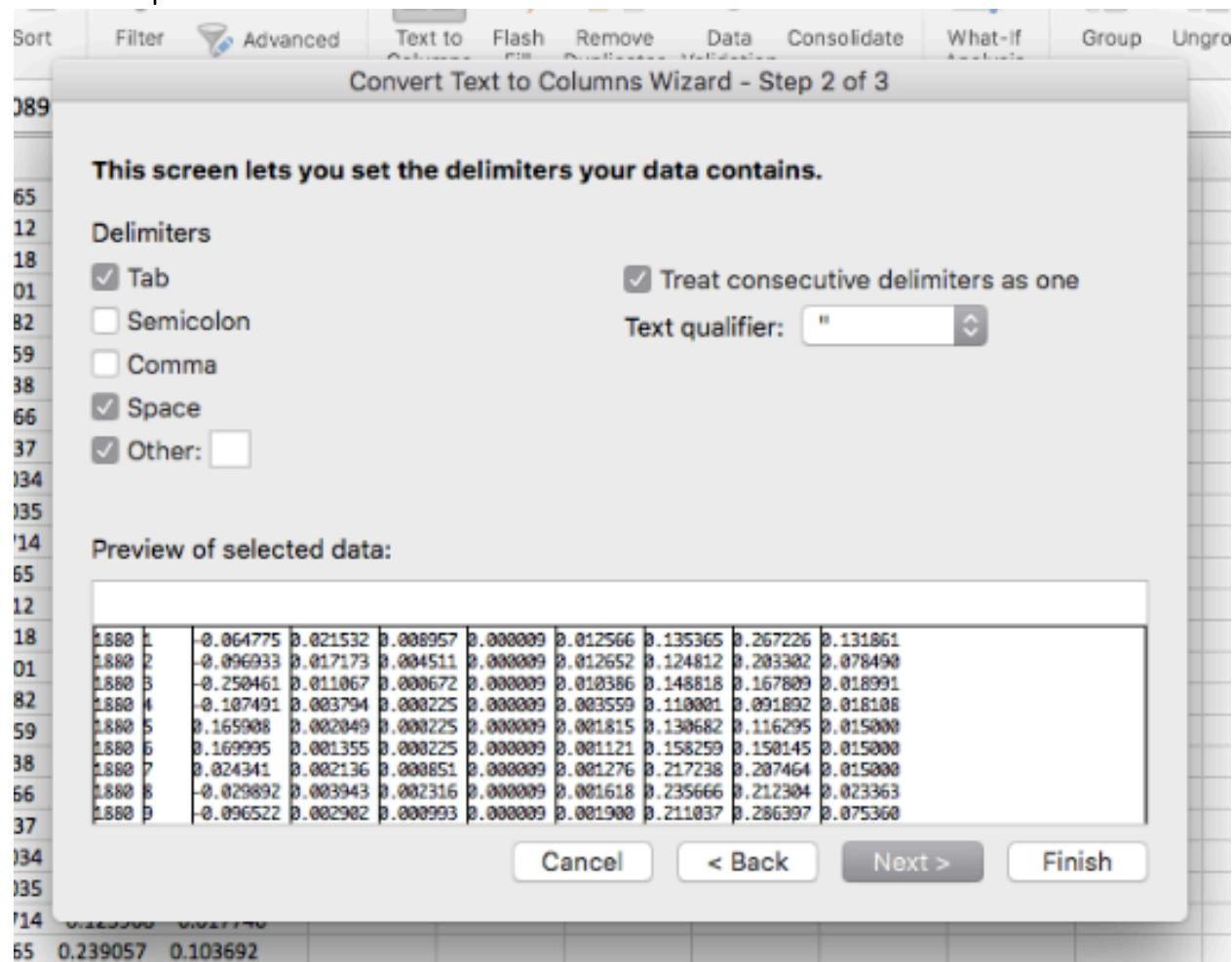
The main area shows a table with 12 rows and 12 columns. Row 1 contains headers: A1, fx, 1880, 1, -0.064775, 0.021532, 0.008957, 0.000009, 0.012566, 0.135365, 0.267226, 0.131861. Rows 2 through 12 contain numerical data.

	A	B	C	D	E	F	G	H	I	J	K	L
1	1880	1	-0.064775	0.021532	0.008957	0.000009	0.012566	0.135365	0.267226	0.131861		
2	1880	2	-0.096933	0.017173	0.004511	0.000009	0.012652	0.124812	0.203302	0.078490		
3	1880	3	-0.150461	0.011067	0.000672	0.000009	0.010386	0.148818	0.167809	0.018991		
4	1880	4	-0.107491	0.003794	0.000225	0.000009	0.003559	0.110001	0.091892	0.018108		
5	1880	5	0.165908	0.002049	0.000225	0.000009	0.001815	0.130682	0.116295	0.015000		
6	1880	6	0.169995	0.001355	0.000225	0.000009	0.001121	0.158259	0.150145	0.015000		
7	1880	7	0.024341	0.002136	0.000851	0.000009	0.001276	0.217238	0.207464	0.015000		
8	1880	8	-0.029892	0.003943	0.002316	0.000009	0.001618	0.235666	0.212304	0.023363		
9	1880	9	-0.096522	0.002902	0.000993	0.000009	0.001900	0.211037	0.286397	0.075360		
10	1880	10	-0.246165	0.006827	0.000248	0.000009	0.006570	0.199034	0.233178	0.034144		
11	1880	11	-0.177689	0.019216	0.003352	0.000009	0.015855	0.172035	0.225828	0.053793		
12	1880	12	0.179403	0.015451	0.000225	0.000009	0.015217	0.141714	0.130979	0.015000		

Convert Text To Columns Wizard



Delimiter Space:



Add Column Headers To Label Data

Excel File Edit View Insert Format Tools Data Window Help

aravg.mon.oce

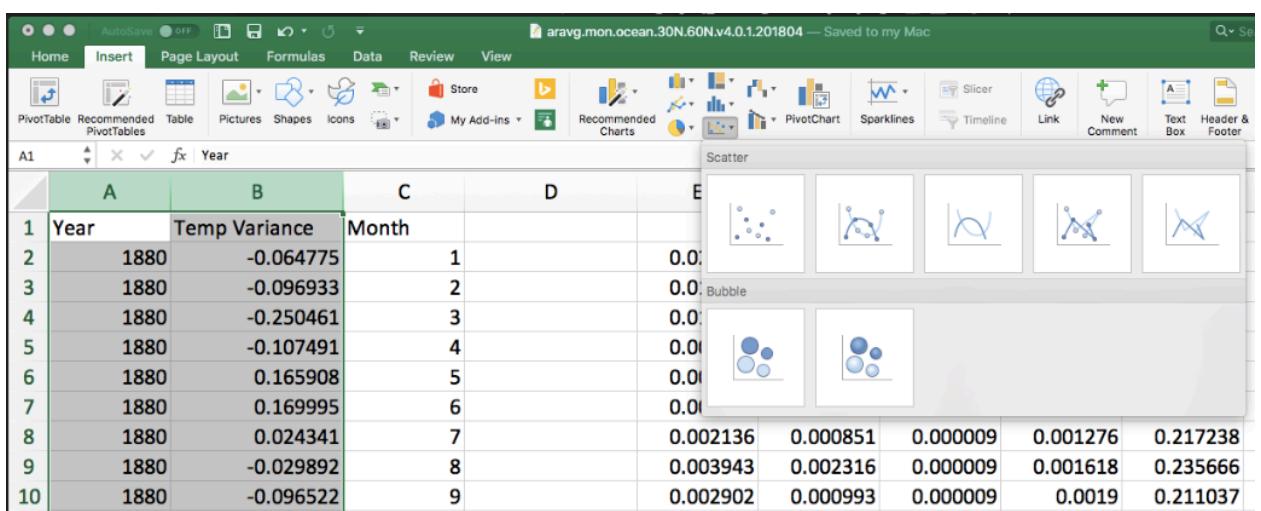
Home Insert Page Layout Formulas Data Review View

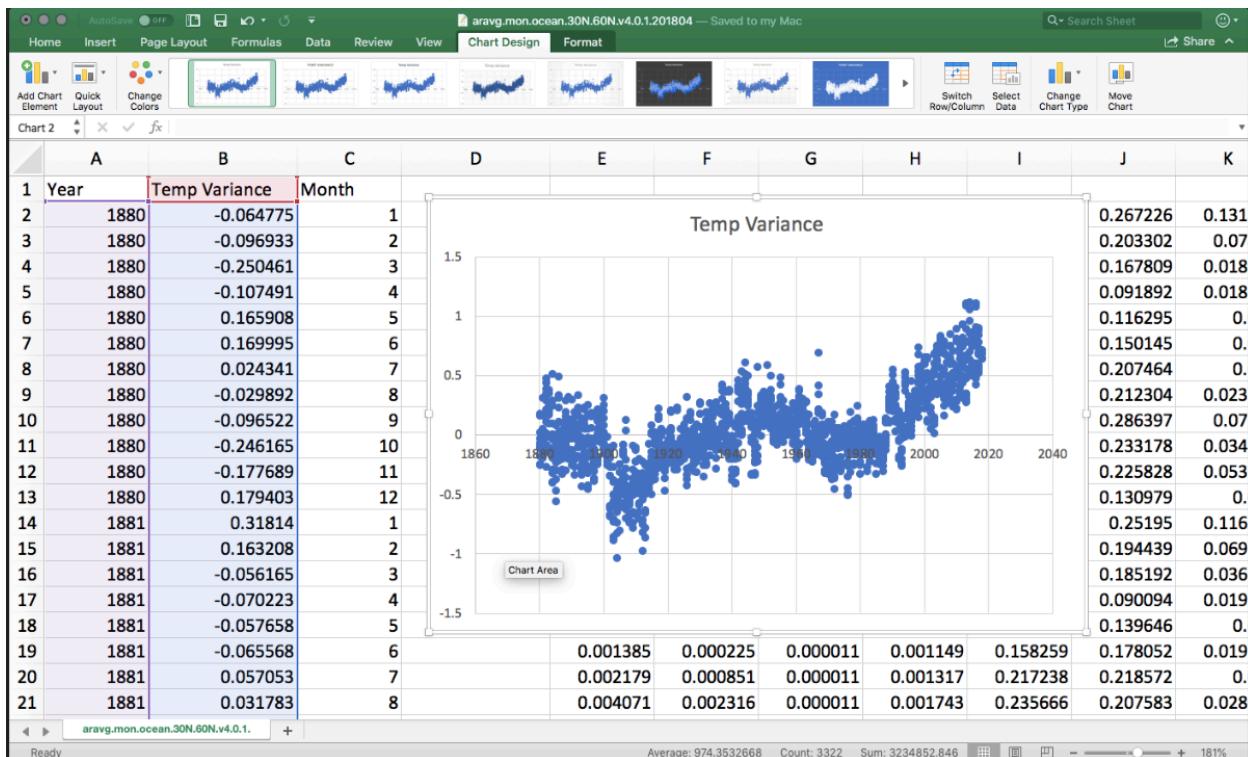
PivotTable Recommended PivotTables Table Pictures Shapes Icons My Add-ins Recommended Charts

A1 fx Year

	A	B	C	D
1	Year	Temp Variance	Month	
2	1880	-0.064775	1	
3	1880	-0.096933	2	
4	1880	-0.250461	3	
5	1880	-0.107491	4	
6	1880	0.165908	5	
7	1880	0.169995	6	
8	1880	0.024341	7	
9	1880	-0.029892	8	

Create new Scatter Plot





6. Communicating Your Conclusions

- ? Who is your audience?
- ? Will your conclusion inform a specific decision? Do you understand the implications of that decision?
- ? Will you be presenting the conclusions using an in person presentation, conference call, screen share, email or printed paper?
- ? Can you predict the challenging questions your audience may ask about your conclusions or analysis techniques?
- ? What are you list of data sources to list as references?