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?

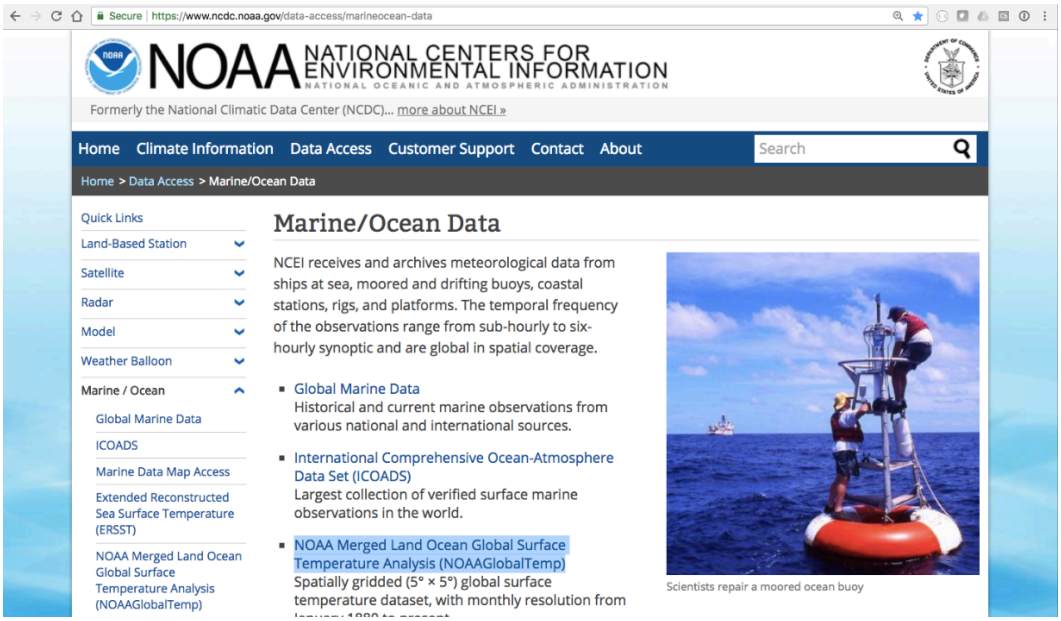
? Do you know how your data was collected?

? 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 conclustions?

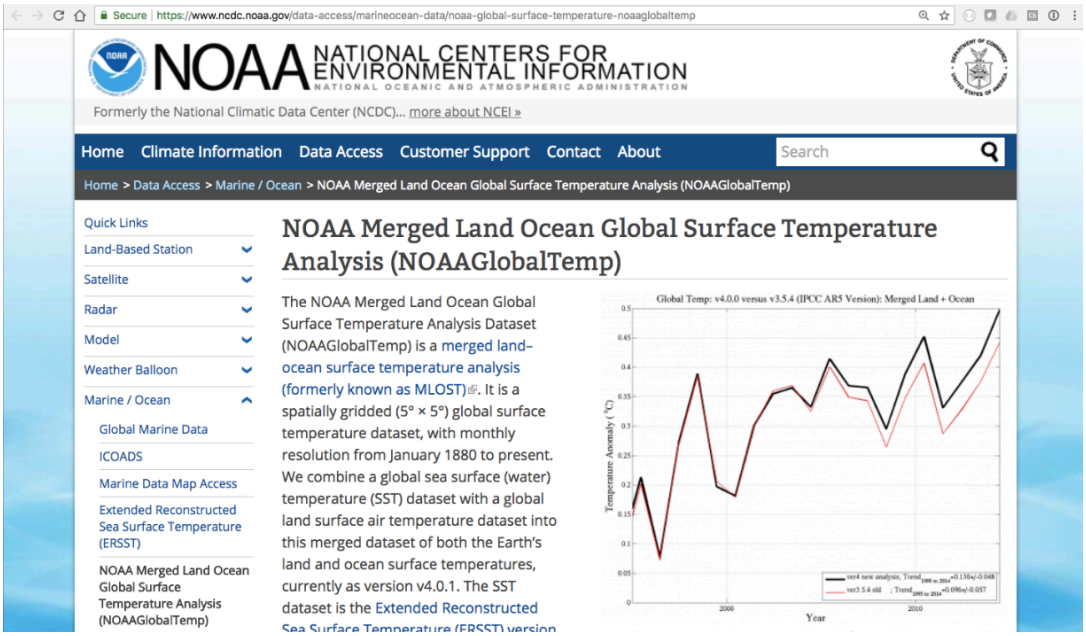
2. Acquire Data Sets

NOAA National Centers for Environmental Information
https://www.ncdc.noaa.gov/data-access/marineocean-data



? What is Temporal Frequency?

NOAA Merged Land Ocean Global Surface Temperature Analysis
https://www.ncdc.noaa.gov/data-access/marineocean-data/noaa-global-surface-temperature-noaaglobaltemp



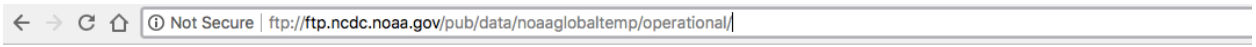
Data Access

- 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 NOAAGlobalTemp 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/

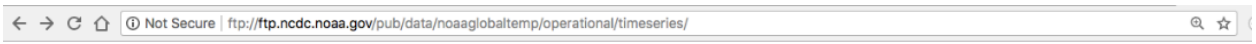


Index of /pub/data/noaaglobaltemp/operational/

[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

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

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?

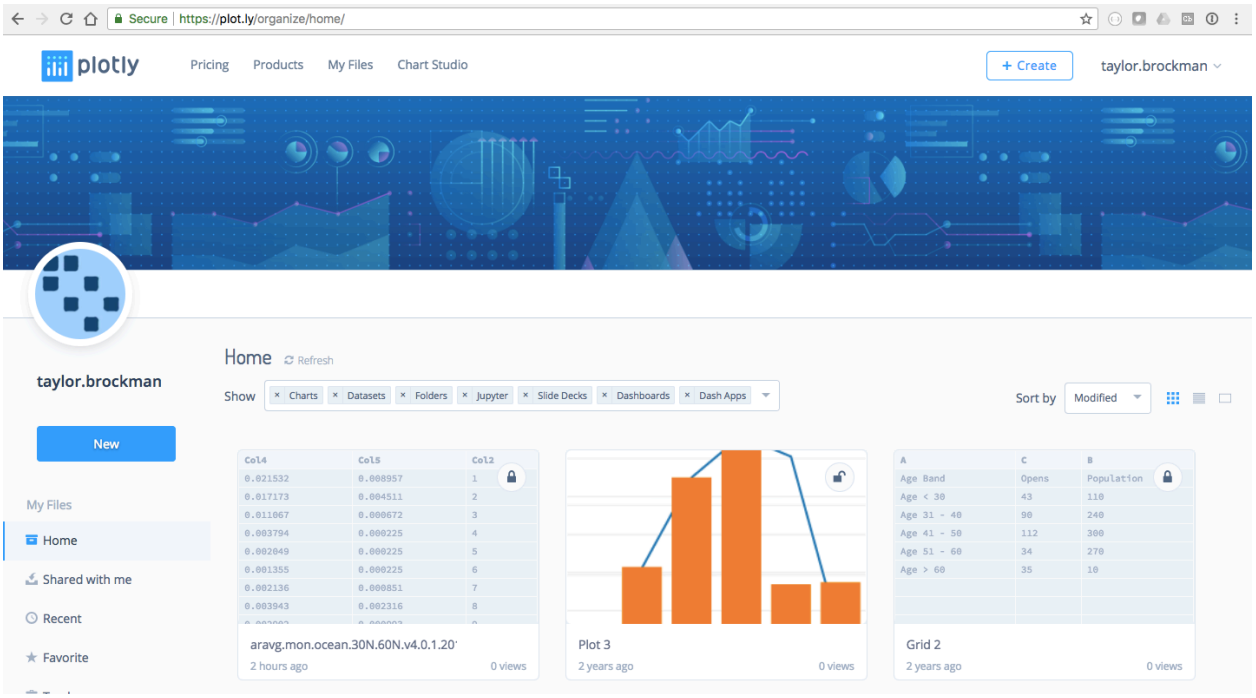
? Do you see any gaps in time?

? Do you see any obvious outliers?

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

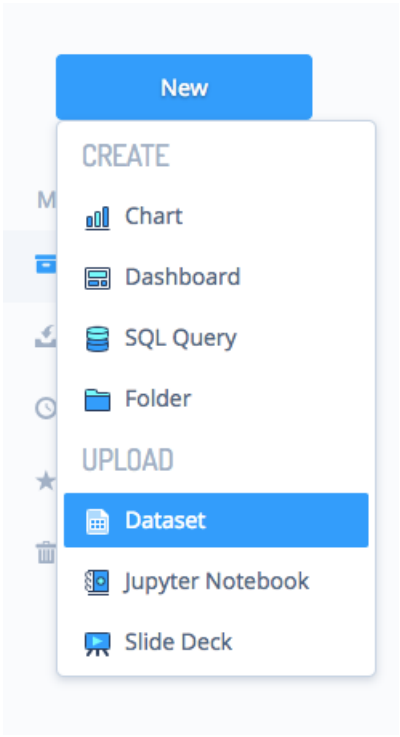
4. Online Data Visualization

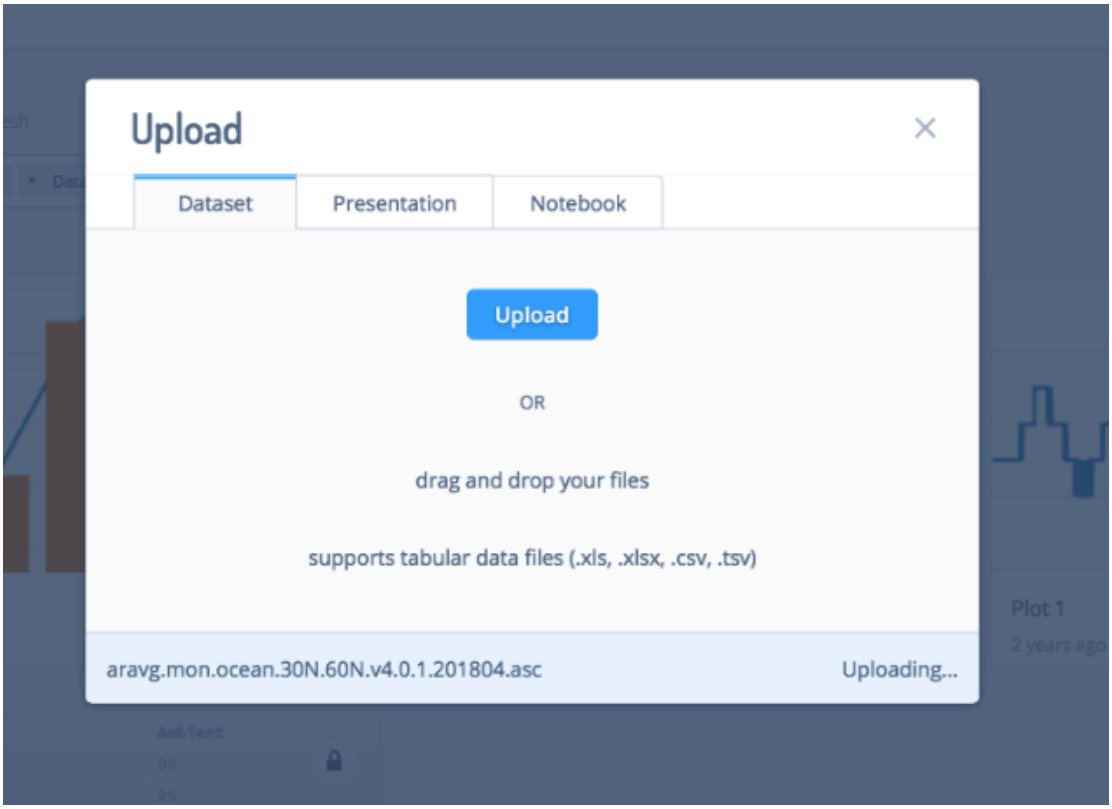
https://plot.ly/



Sign up for Free Account

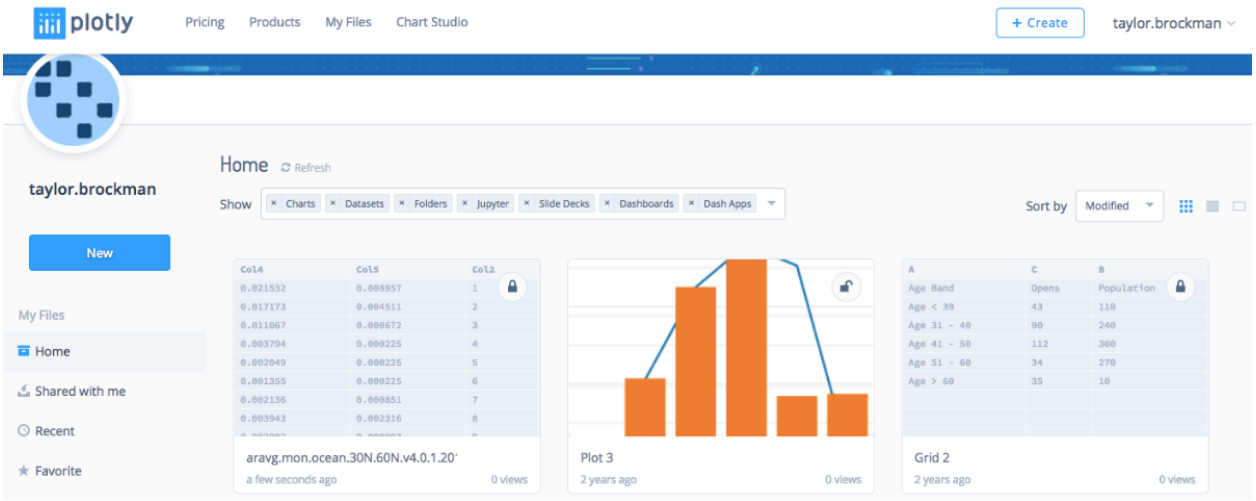
Upload



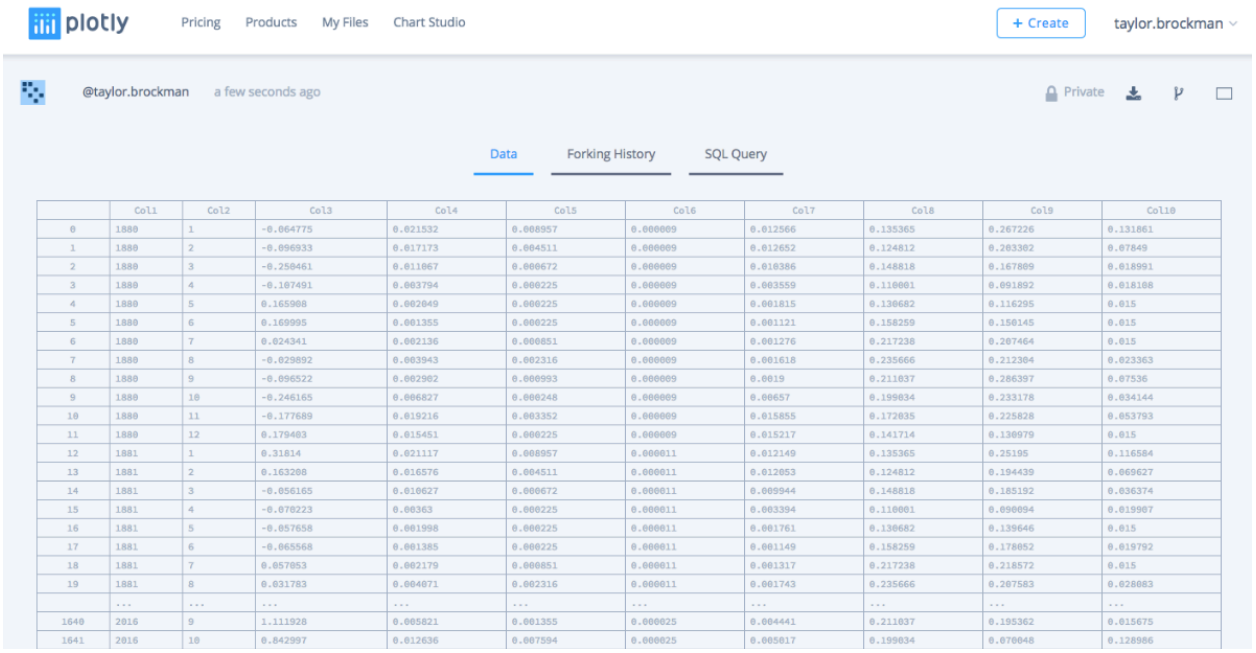


Uploaded aravg.mon.ocean.30N.60N.v4.0.1.201804.asc with no modifications

Dataset Appears on your Home Screen

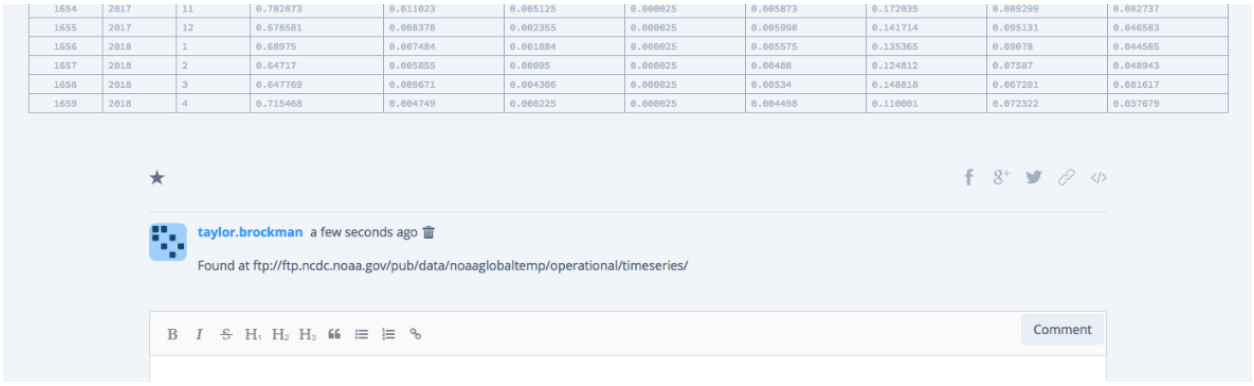


Browsing the Data Set in Detail

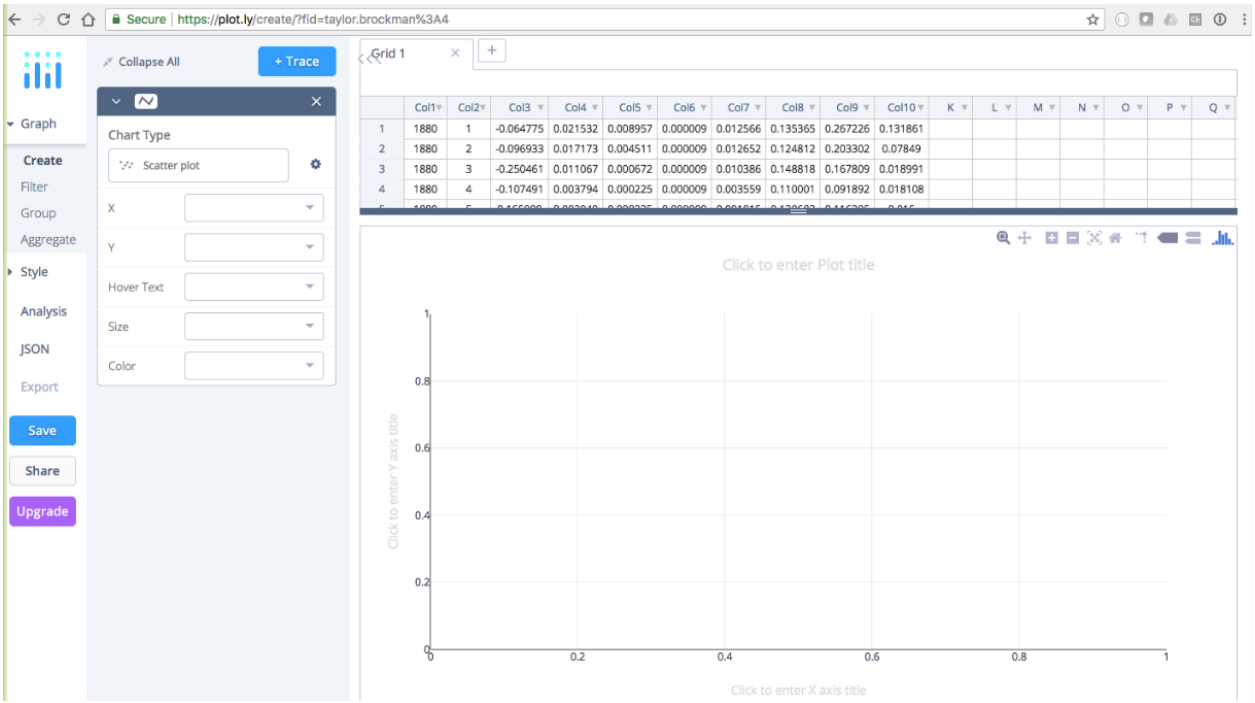


	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10
0	1880	1	-0.064775	0.021532	0.008957	0.000009	0.012566	0.135365	0.267226	0.131861
1	1880	2	-0.096933	0.017173	0.004511	0.000009	0.012652	0.124812	0.263382	0.07849
2	1880	3	-0.250461	0.011067	0.000672	0.000009	0.016386	0.148818	0.167809	0.018991
3	1880	4	-0.107491	0.003794	0.000225	0.000009	0.003559	0.110001	0.091892	0.018108
4	1880	5	0.165908	0.002049	0.000225	0.000009	0.001815	0.130682	0.116295	0.015
5	1880	6	0.169995	0.001355	0.000225	0.000009	0.001121	0.158259	0.150145	0.015
6	1880	7	0.024341	0.002136	0.000851	0.000009	0.001276	0.217238	0.207464	0.015
7	1880	8	-0.029892	0.003943	0.002316	0.000009	0.001618	0.235666	0.212304	0.023363
8	1880	9	-0.096522	0.002902	0.000993	0.000009	0.0019	0.211037	0.286397	0.07536
9	1880	10	-0.246165	0.006827	0.000248	0.000009	0.00657	0.199034	0.233178	0.034144
10	1880	11	-0.177689	0.019216	0.003352	0.000009	0.015855	0.172035	0.225828	0.053793
11	1880	12	0.179403	0.015451	0.000225	0.000009	0.015217	0.141714	0.130979	0.015
12	1881	1	0.31814	0.021117	0.008957	0.000011	0.012149	0.135365	0.25195	0.116584
13	1881	2	0.163208	0.016576	0.004511	0.000011	0.012053	0.124812	0.194439	0.069627
14	1881	3	-0.056165	0.010627	0.000672	0.000011	0.009944	0.148818	0.185192	0.036374
15	1881	4	-0.070223	0.00363	0.000225	0.000011	0.003394	0.110001	0.090094	0.019907
16	1881	5	-0.057658	0.001998	0.000225	0.000011	0.001761	0.130682	0.139646	0.015
17	1881	6	-0.065568	0.001385	0.000225	0.000011	0.001149	0.158259	0.178052	0.019792
18	1881	7	0.057053	0.002179	0.000851	0.000011	0.001317	0.217238	0.218572	0.015
19	1881	8	0.031783	0.004071	0.002316	0.000011	0.001743	0.235666	0.207583	0.028083
	...	...	...	...	...	...	...	...	...	...
1640	2016	9	1.111928	0.005821	0.001355	0.000025	0.004441	0.211037	0.195362	0.015675
1641	2016	10	0.842997	0.012636	0.007584	0.000025	0.005017	0.199034	0.670048	0.128986

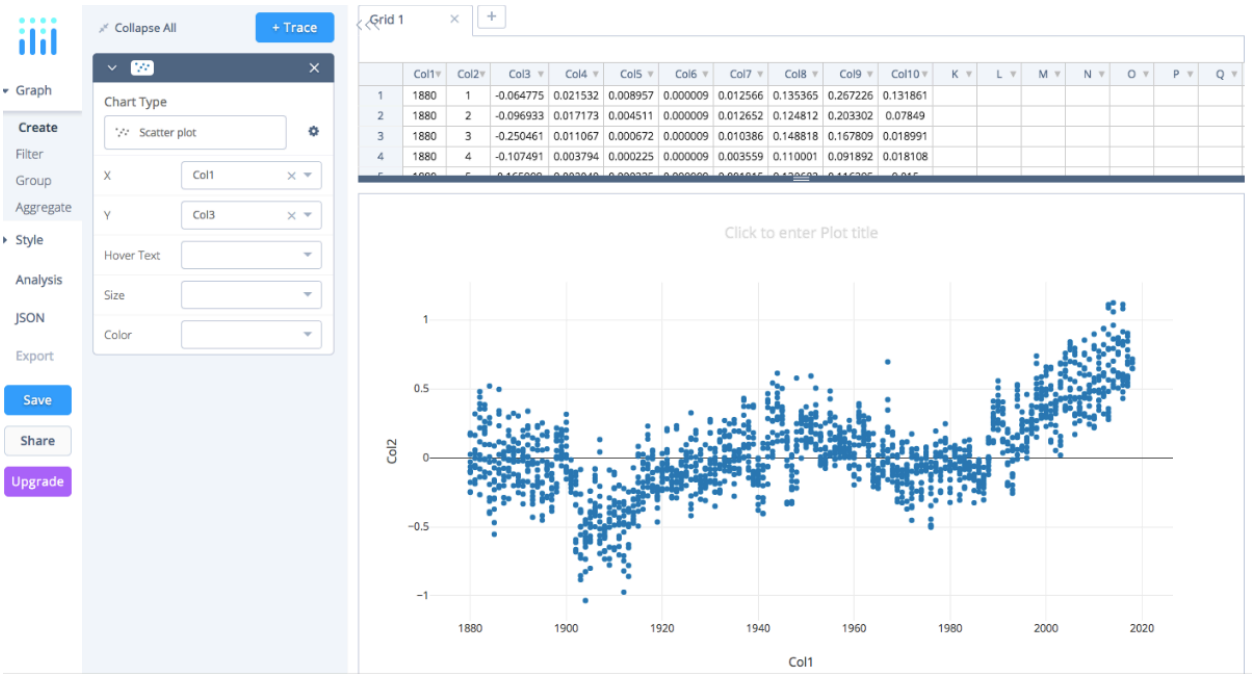
Commenting on the Data Set



Create Visualization



Set Columns into X and Y



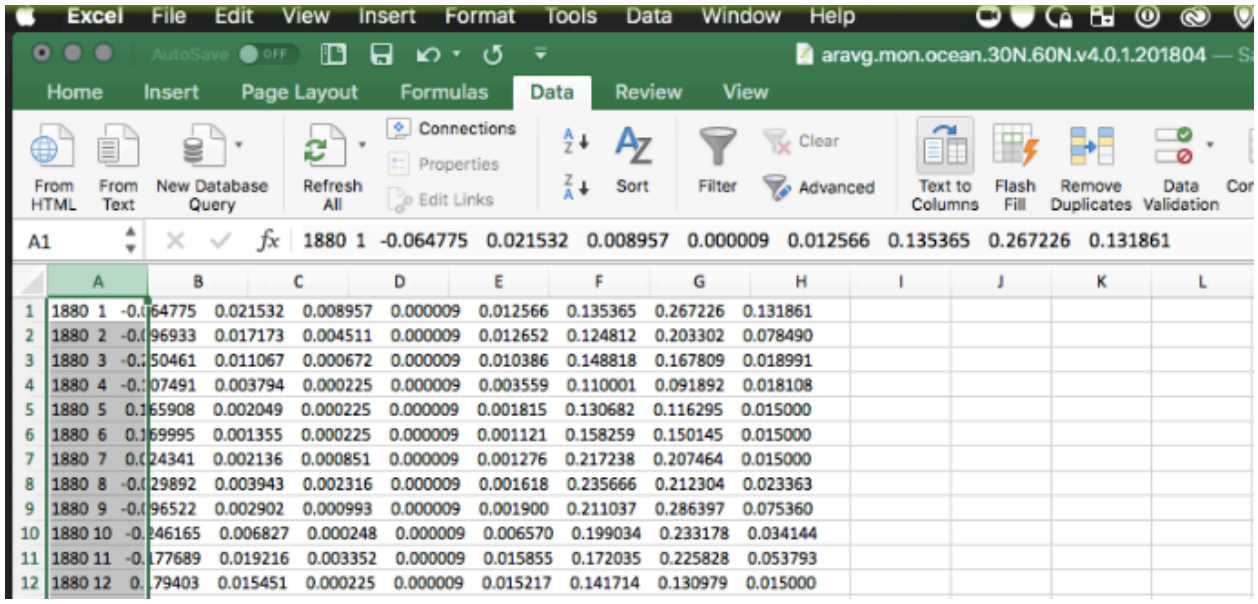
? Could you add a 2nd File as another data series in a different color to serve as a comparison?

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

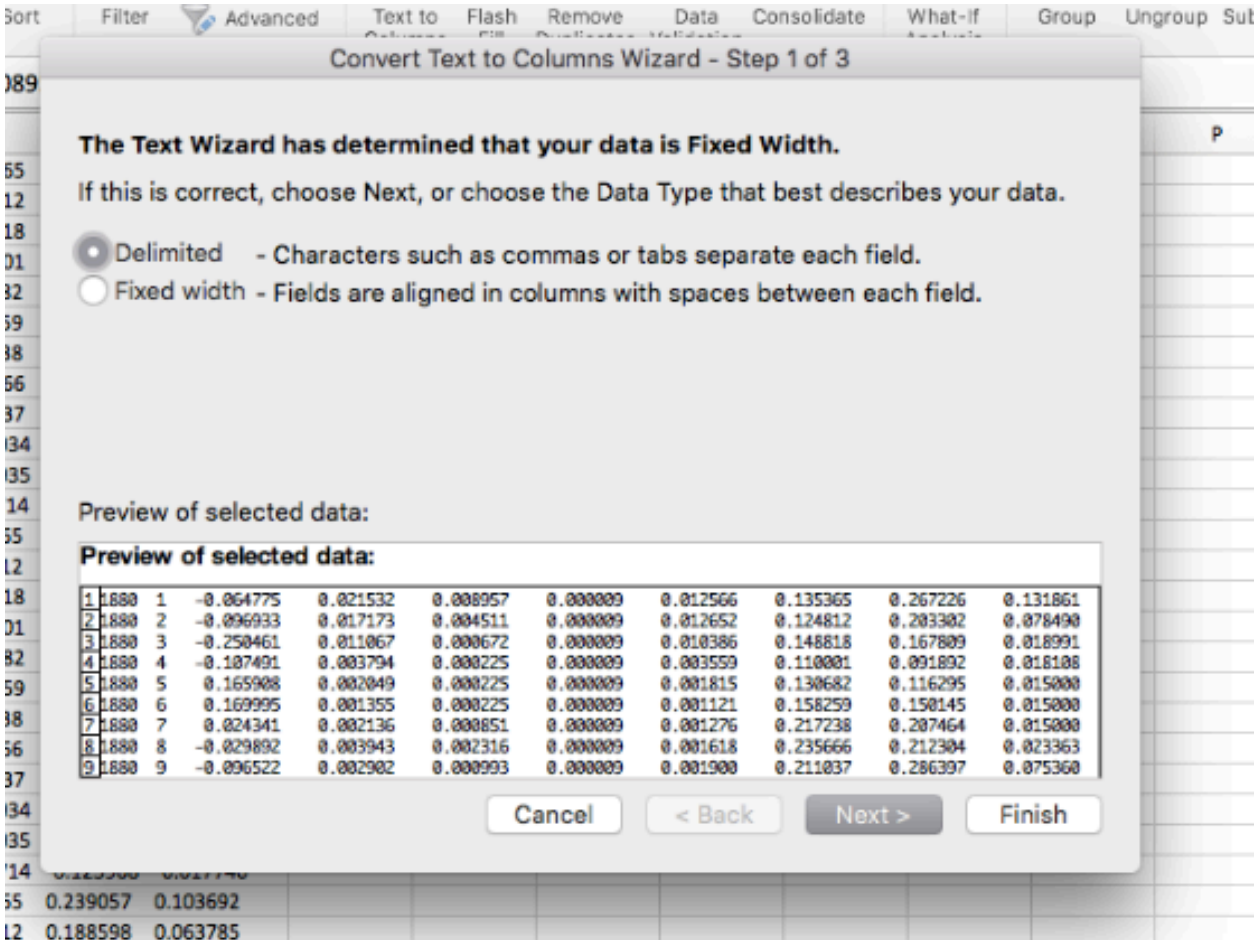
? What other visualization types help you understand the shape and the trend of this dataset?

5. Offline Data Visualization

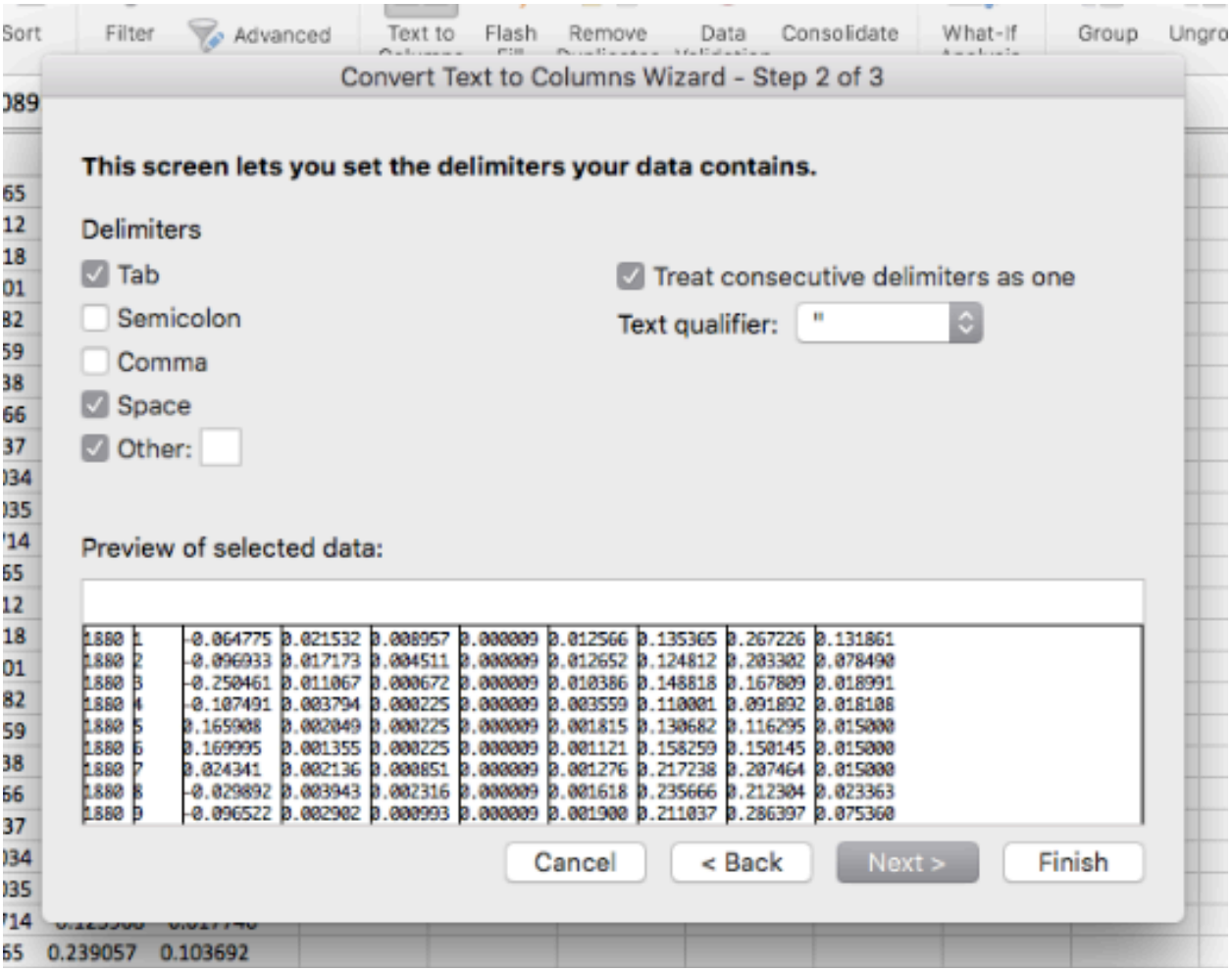
Using Microsoft Excel, the File opens as a single column per row:



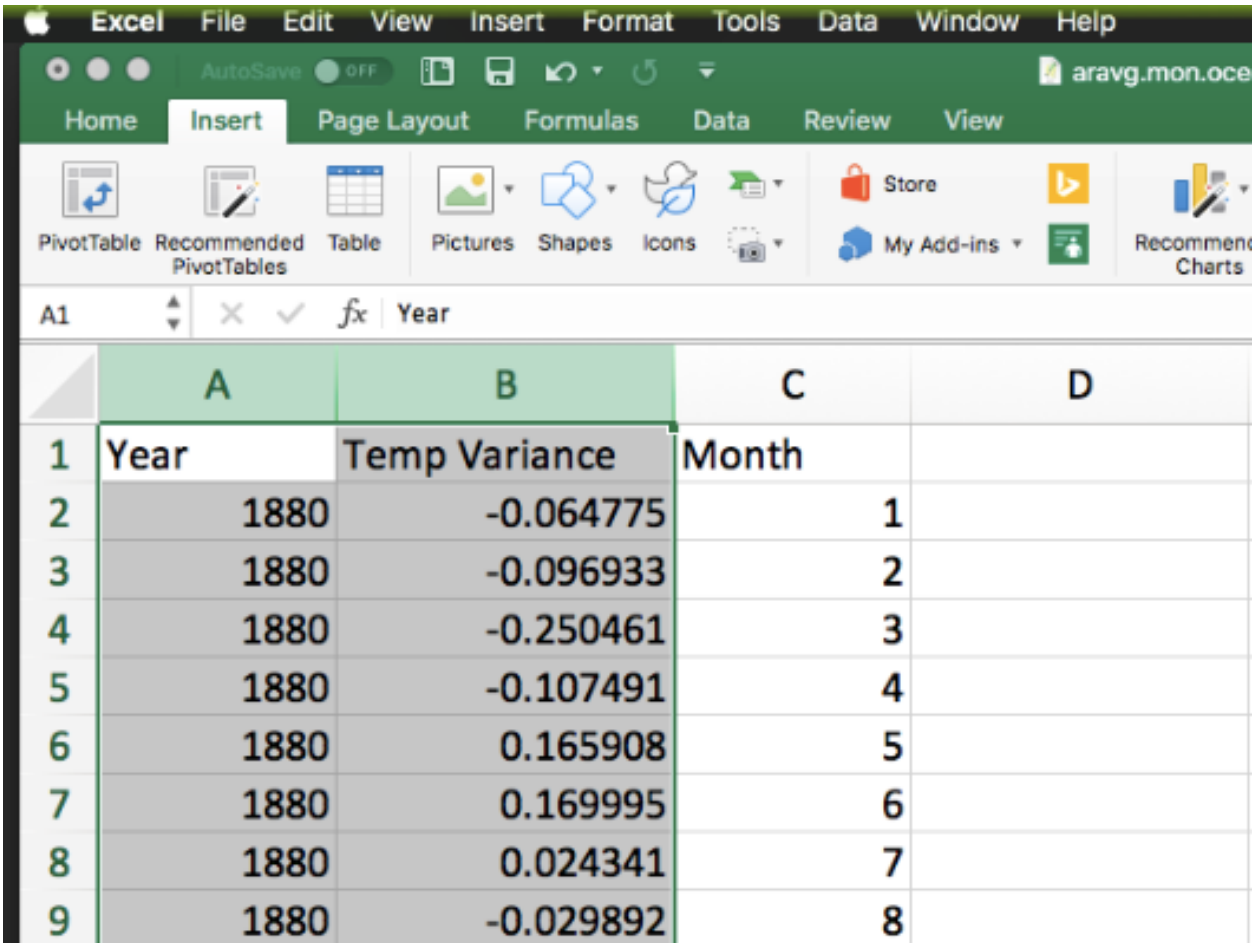
Convert Text To Columns Wizard



Delimiter Space:

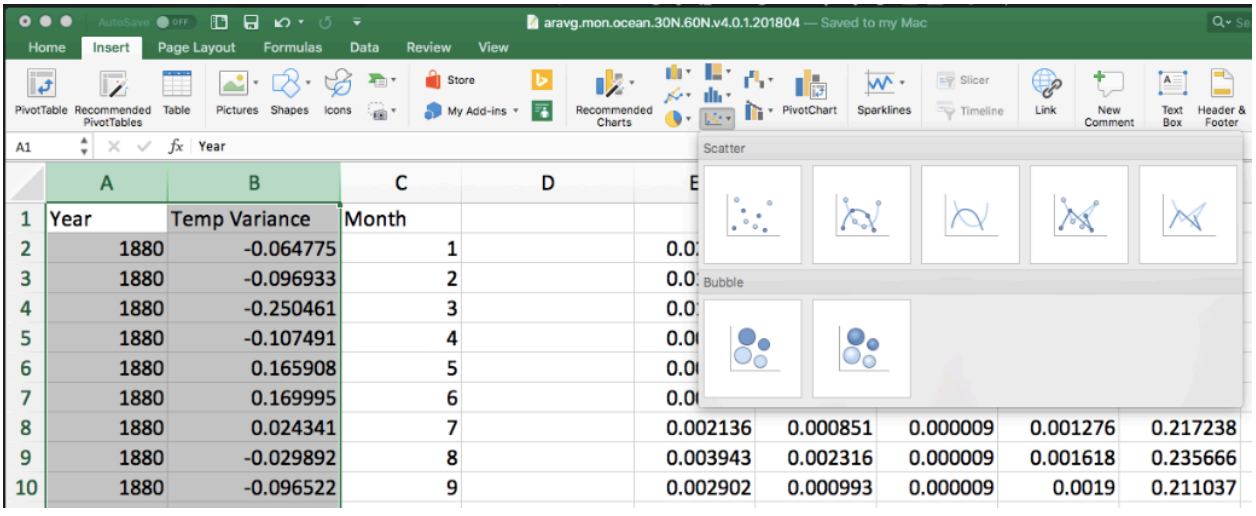


Add Column Headers To Label Data

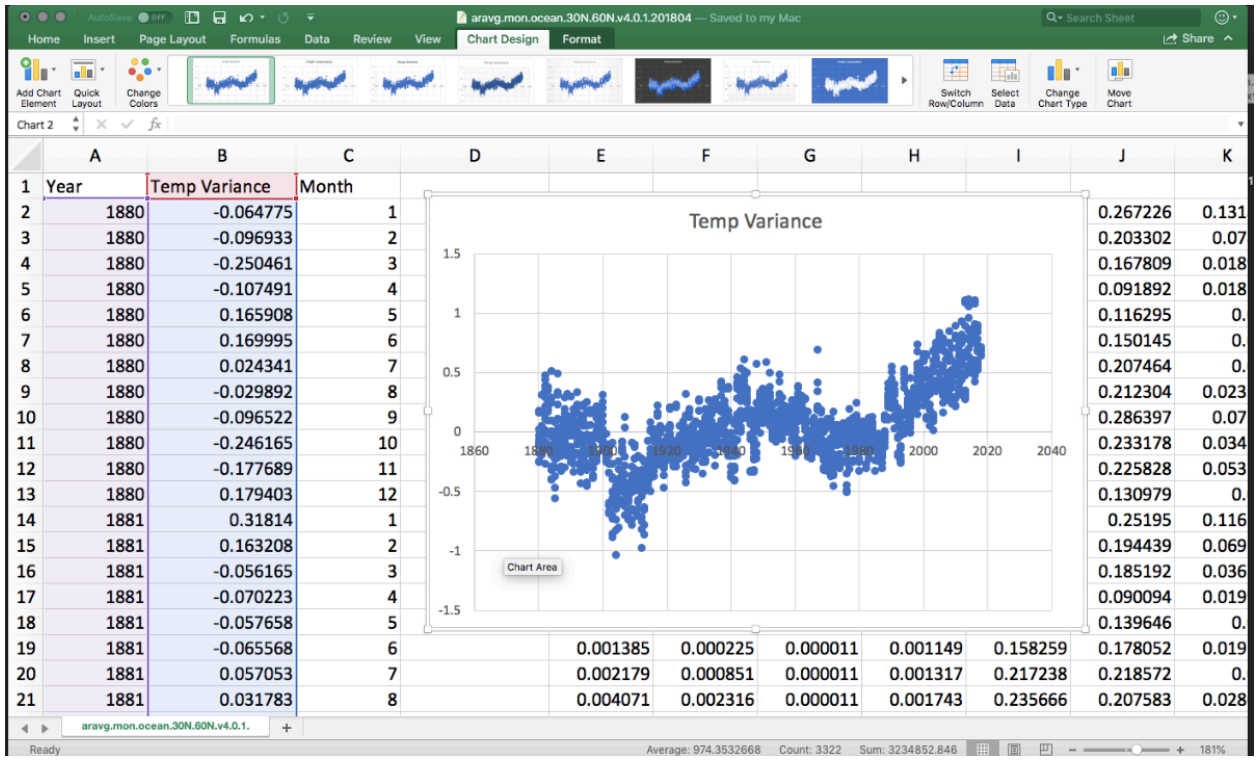


	A	B	C	D
1	Year	Temp Variance	Month	
2	1880	-0.064775	1	
3	1880	-0.096933	2	
4	1880	-0.250461	3	
5	1880	-0.107491	4	
6	1880	0.165908	5	
7	1880	0.169995	6	
8	1880	0.024341	7	
9	1880	-0.029892	8	

Create new Scatter Plot



	A	B	C	D	E				
1	Year	Temp Variance	Month						
2	1880	-0.064775	1		0.0				
3	1880	-0.096933	2		0.0				
4	1880	-0.250461	3		0.0				
5	1880	-0.107491	4		0.0				
6	1880	0.165908	5		0.0				
7	1880	0.169995	6		0.0				
8	1880	0.024341	7		0.002136	0.000851	0.000009	0.001276	0.217238
9	1880	-0.029892	8		0.003943	0.002316	0.000009	0.001618	0.235666
10	1880	-0.096522	9		0.002902	0.000993	0.000009	0.0019	0.211037



	A	B	C	D	E	F	G	H	I	J	K
1	Year	Temp Variance	Month								
2	1880	-0.064775	1							0.267226	0.131
3	1880	-0.096933	2							0.203302	0.07
4	1880	-0.250461	3							0.167809	0.018
5	1880	-0.107491	4							0.091892	0.018
6	1880	0.165908	5							0.116295	0.
7	1880	0.169995	6							0.150145	0.
8	1880	0.024341	7							0.207464	0.
9	1880	-0.029892	8							0.212304	0.023
10	1880	-0.096522	9							0.286397	0.07
11	1880	-0.246165	10							0.233178	0.034
12	1880	-0.177689	11							0.225828	0.053
13	1880	0.179403	12							0.130979	0.
14	1881	0.31814	1							0.25195	0.116
15	1881	0.163208	2							0.194439	0.069
16	1881	-0.056165	3							0.185192	0.036
17	1881	-0.070223	4							0.090094	0.019
18	1881	-0.057658	5							0.139646	0.
19	1881	-0.065568	6		0.001385	0.000225	0.000011	0.001149	0.158259	0.178052	0.019
20	1881	0.057053	7		0.002179	0.000851	0.000011	0.001317	0.217238	0.218572	0.
21	1881	0.031783	8		0.004071	0.002316	0.000011	0.001743	0.235666	0.207583	0.028

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?