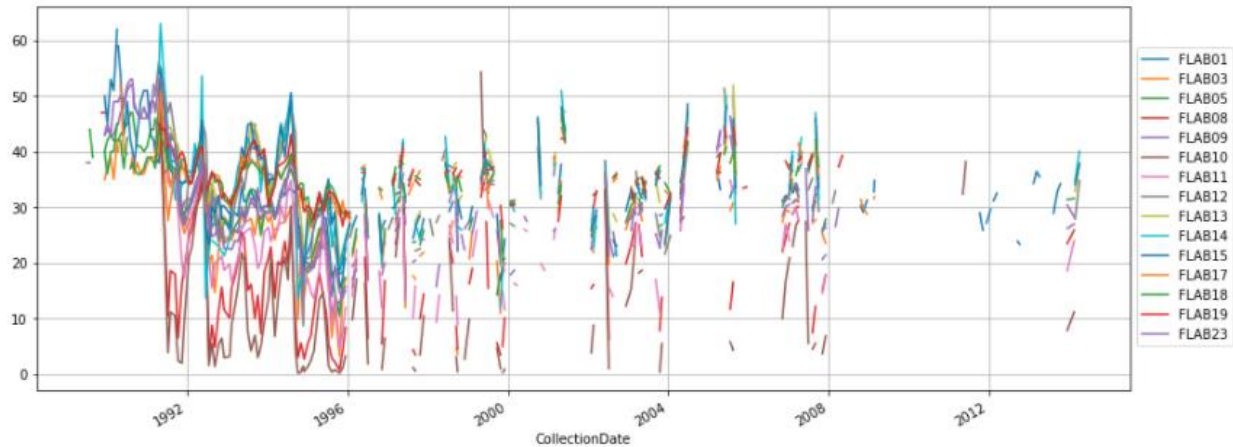


## DSC 521 Assignment 1

Nolan Stelzner & David Jackson

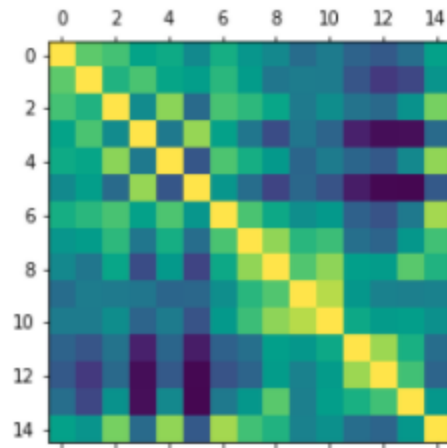
The initial shape of the salt data frame is 596x15. We went ahead and plotted the original salt data frame as shown below.



As we resample the data by monthly mean it reduced the data set shape from 596x15 to 299x15. This was even further reduced to 229x15 as we dropped all NaN values using `.dropna()`.

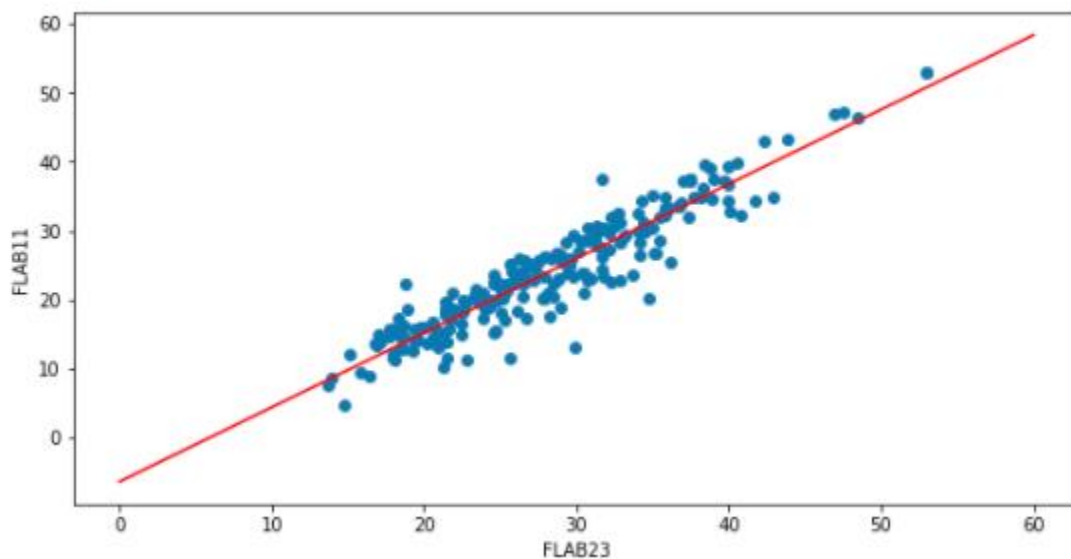
The correlation matrix on this refined data frame showed us that the biggest correlation for FLAB23 was with FLAB11 at .933973.

Station ID	FLAB01	FLAB03	FLAB05	FLAB08	FLAB09	FLAB10	FLAB11	FLAB12	FLAB13	FLAB14	FLAB15	FLAB17	FLAB18	FLAB19	FLAB23
Station ID															
FLAB01	1.000000	0.883028	0.862493	0.795271	0.810663	0.742812	0.818645	0.768269	0.737579	0.683248	0.713135	0.662425	0.644609	0.685580	0.791610
FLAB03	0.883028	1.000000	0.828934	0.868321	0.799161	0.785553	0.839682	0.780842	0.700965	0.715403	0.710584	0.636467	0.588895	0.611719	0.763983
FLAB05	0.862493	0.828934	1.000000	0.746958	0.917825	0.678962	0.865440	0.840474	0.800942	0.709890	0.750571	0.696531	0.675976	0.759694	0.901380
FLAB08	0.795271	0.868321	0.746958	1.000000	0.707413	0.923879	0.790884	0.700566	0.621342	0.701135	0.665343	0.551729	0.522526	0.523906	0.679145
FLAB09	0.810663	0.799161	0.917825	0.707413	1.000000	0.638216	0.869082	0.817841	0.774058	0.669406	0.711365	0.668250	0.643680	0.740352	0.920050
FLAB10	0.742812	0.785553	0.678962	0.923879	0.638216	1.000000	0.768566	0.684160	0.605748	0.672421	0.632811	0.543933	0.514607	0.511723	0.643345
FLAB11	0.818645	0.839682	0.865440	0.790884	0.869082	0.768566	1.000000	0.866499	0.806646	0.752488	0.773767	0.659428	0.632102	0.707303	0.933973
FLAB12	0.768269	0.780842	0.840474	0.700566	0.817841	0.684160	0.866499	1.000000	0.919997	0.836417	0.854892	0.688537	0.664813	0.769295	0.858985
FLAB13	0.737579	0.700965	0.800942	0.621342	0.774058	0.605748	0.806646	0.919997	1.000000	0.872260	0.918248	0.786990	0.780700	0.878922	0.830024
FLAB14	0.683248	0.715403	0.709890	0.701135	0.669406	0.672421	0.752488	0.836417	0.872260	1.000000	0.948016	0.770903	0.724940	0.721388	0.724790
FLAB15	0.713135	0.710584	0.750571	0.665343	0.711365	0.632811	0.773767	0.854892	0.918248	0.948016	1.000000	0.806554	0.779301	0.784576	0.757381
FLAB17	0.662425	0.636467	0.696531	0.551729	0.668250	0.543933	0.659428	0.688537	0.786990	0.770903	0.806554	1.000000	0.927633	0.820824	0.685178
FLAB18	0.644609	0.588895	0.675976	0.522526	0.643680	0.514607	0.632102	0.664813	0.780700	0.724940	0.779301	0.927633	1.000000	0.859080	0.667389
FLAB19	0.685580	0.611719	0.759694	0.523906	0.740352	0.511723	0.707303	0.769295	0.878922	0.721388	0.784576	0.820824	0.859080	1.000000	0.770941
FLAB23	0.791610	0.763983	0.901380	0.679145	0.920050	0.643345	0.933973	0.858985	0.830024	0.724790	0.757381	0.685178	0.667389	0.770941	1.000000



As we take those two variables and turn them into a linear regression model, below is the plot & fitted line that contains an r squared value of .8723 – which is a pretty good initial value to tell us that the model will work.

-6.443013617727573 1.0808615154334005



Lastly, we used this linear regression model in order to predict the first 21 months of the data set for FLAB11. Per your request we started by inserting values 40 and 41 into the 4<sup>th</sup> and 5<sup>th</sup> month of FLAB23.

Station ID	FLAB01	FLAB03	FLAB05	FLAB08	FLAB09	FLAB10	FLAB11	FLAB12	FLAB13	FLAB14	FLAB15	FLAB17	FLAB18	FLAB19	FLAB23
CollectionDate															
1989-06-30	NaN	NaN	34.0	36.0	38.0	NaN	NaN	NaN	52.0	NaN	57.0	42.0	44.0	46.0	38.0
1989-07-31	NaN	NaN	38.0	26.0	38.0	NaN	NaN	NaN	54.0	NaN	50.0	40.0	44.0	44.0	38.0
1989-08-31	NaN	NaN	38.0	18.0	39.0	NaN	NaN	NaN	52.0	NaN	44.0	37.0	39.0	44.0	39.0
1989-09-30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	40.0
1989-10-31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	41.0
1989-11-30	NaN	NaN	39.0	27.0	42.0	NaN	NaN	NaN	53.0	NaN	46.0	37.0	40.0	47.0	42.0
1989-12-31	NaN	NaN	40.0	32.0	43.0	NaN	NaN	NaN	56.0	NaN	50.0	35.0	40.0	47.0	43.0
1990-01-31	NaN	NaN	42.0	NaN	44.0	NaN	NaN	NaN	NaN	NaN	45.0	37.0	36.0	NaN	45.0
1990-02-28	NaN	NaN	43.0	NaN	43.0	NaN	NaN	NaN	NaN	NaN	53.0	39.0	40.0	NaN	43.0
1990-03-31	NaN	NaN	42.0	NaN	43.0	NaN	NaN	NaN	NaN	NaN	51.0	35.0	40.0	NaN	49.0

After doing this we took this information to predict the values in FLAB11 and plot them.

