

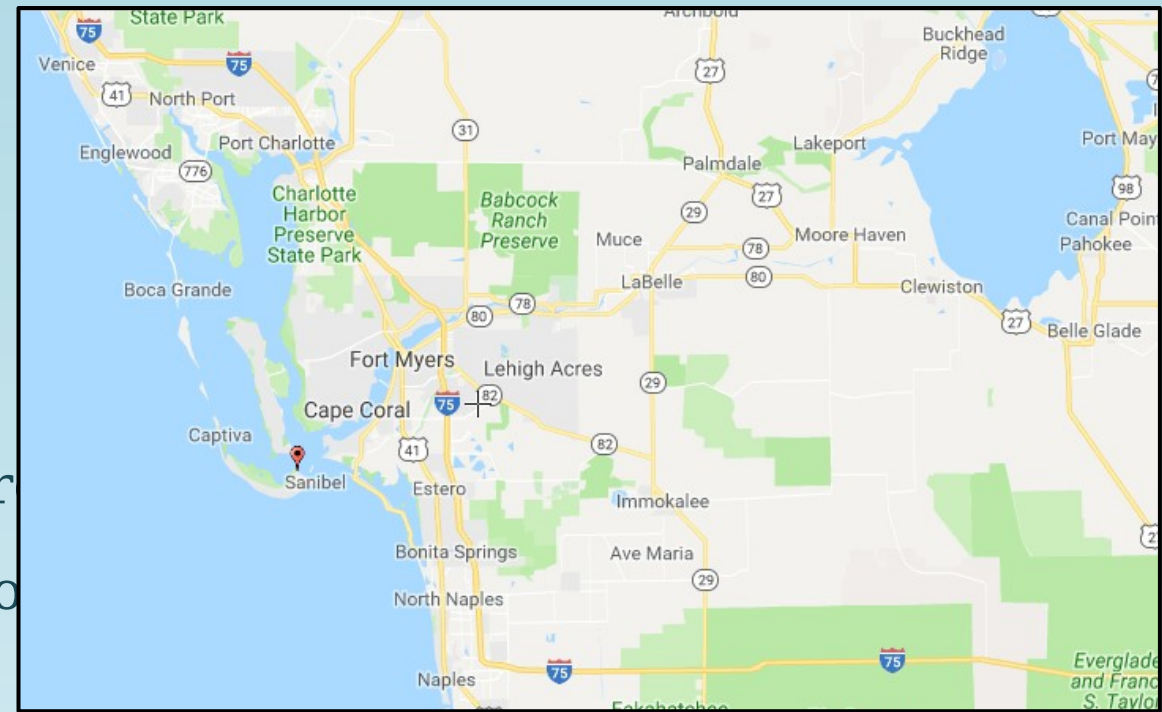
Principal Component Analysis: Pine Island Sound Monitoring Station

Introduction: Pine Island Sound

- Location: Southwest Florida
 - Lies between Sanibel Island/Pine Island
 - Connected to other bodies of water
 - Gulf of Mexico/Estero Bay/Caloosahatchee River
- Abundant clean water is the foundation of the region's
Environment Quality of Life Economy
- Algal blooms/Red tide
 - More frequent/More intense
 - Rising temperatures/Lake Okeechobee discharges
- Algae feed nutrients
 - Overpopulate -> DO reduced in water
 - Toxins -> fish poisoned -> respiratory issues

No solution

- Use PCA



PCA you say?

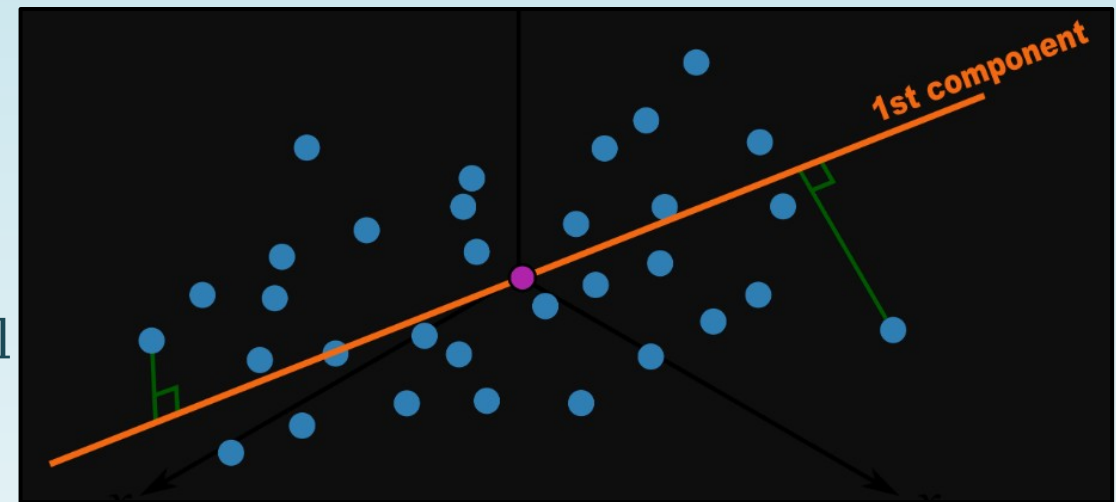
- PCA – Statistical method used to identify patterns in data with many dimensions

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n - 1)}$$

- Calculate covariance matrix
 - Establish eigenvectors/principal components
 - New axis to highlight variance between each dimension
 - PC1 = most variation PC2 = 2nd most variation PC3 = 3rd most variation

$$C = \begin{pmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{pmatrix}$$

- Obtain eigenvalues for each eigenvector
 - Eigenvector with largest eigenvalue = PC
 - Remaining values – which ones to retain?
- Develop feature/transpose/multiply by original
 - Expresses original data in terms of new vectors
 - Tells how far data point deviates

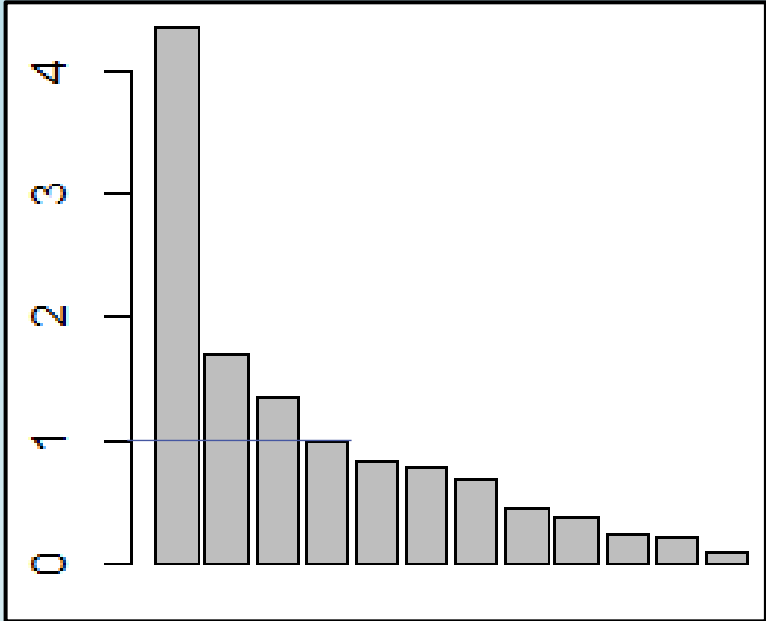
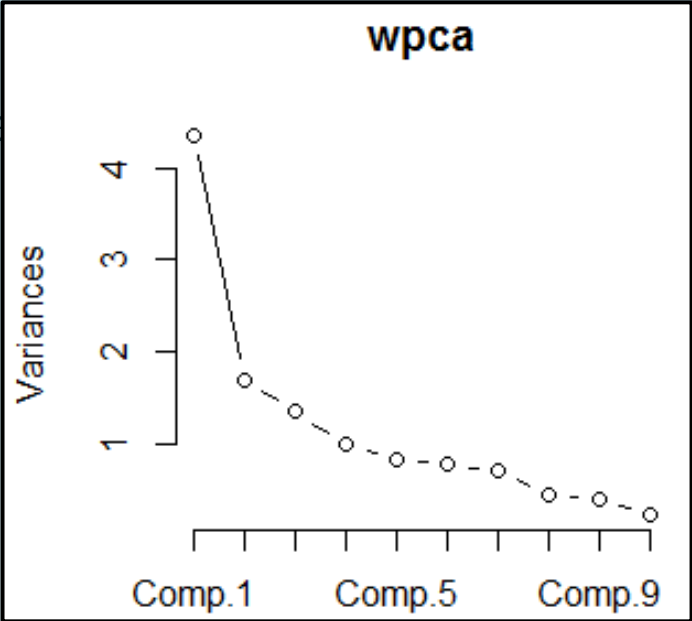


Methods:

- Lee county's environmental lab
 - 12 water quality variables over 10 years (01/01/2009 - 12/31/2018)
 - Monthly averages
 - Dissolved
Oxygen/Ammonia/Nitrite/Nitrate/pH/Salinity/Silica/Phosphorous/Temperature/
Nitrogen/Total Organic Carbon/Total Suspended Solids
- Standardize
 - Normally distributed (Mean - 0/Unit Variance)
 - Different measurement units/Variable naturally have different ranges
- Correlation matrix/PCA
 - R -> princomp function

Results: Reducing Dimensionality

- Summary/What components to retain?
 - St.dev - measurement of spread (PC1)
 - Proportion of variance (PC1)
 - Scree Plot - shows change in variance
 - PC5 - Changes less drastic
 - Eigenvalues - Keep those above 1
 - PC4 - Nope

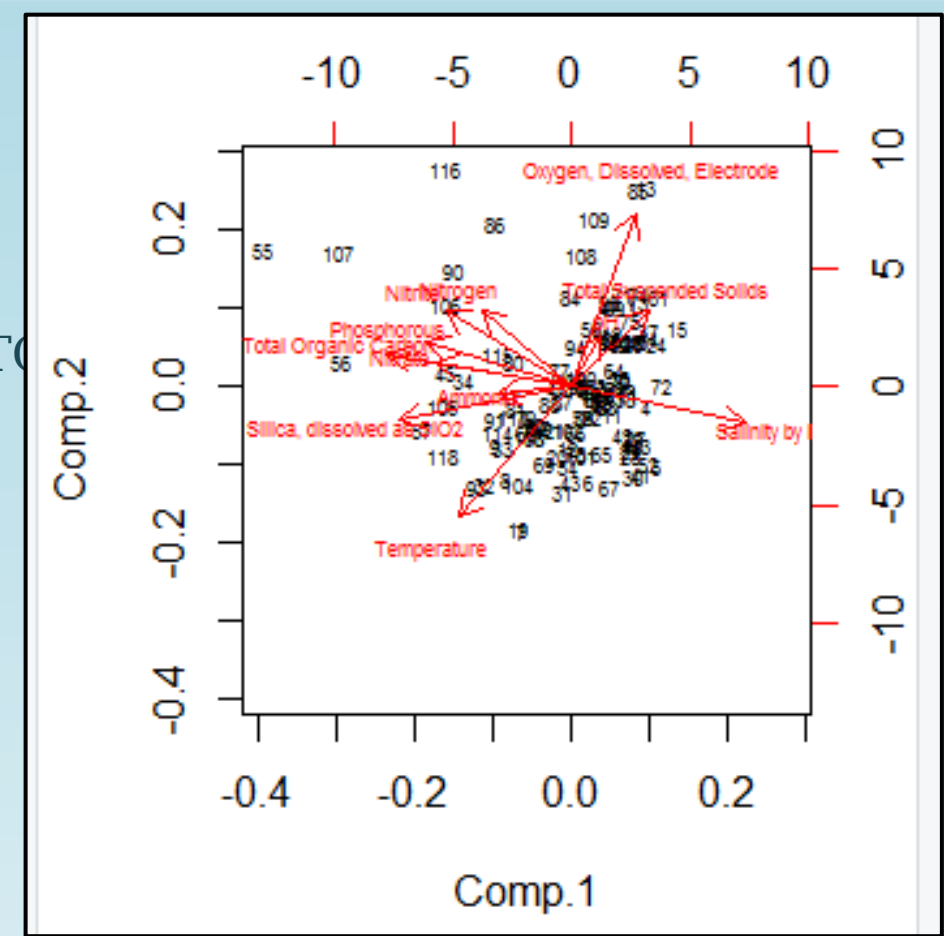


Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	2.0853162	1.3018011	1.1574089	0.99193454	0.90724136	0.87537218
Proportion of Variance	0.3623787	0.1412238	0.1116329	0.08199451	0.06859057	0.06385637
Cumulative Proportion	0.3623787	0.5036025	0.6152354	0.69722995	0.76582052	0.82967689
	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12
Standard deviation	0.83178579	0.66485065	0.61891855	0.47607180	0.46095241	0.296313375
Proportion of Variance	0.05765563	0.03683553	0.03192168	0.01888703	0.01770643	0.007316801
Cumulative Proportion	0.88733253	0.92416806	0.95608974	0.97497677	0.99268320	1.000000000

Loadings

- PC1
 - Greatest positive value – Salinity
 - Greatest negative value – Nitrate/Silica/Phosphorous/Temperature
 - Positively correlated
- PC2
 - Greatest positive value – Dissolved Oxygen
 - Greatest negative value – Temperature
- PC3
 - Greatest positive value – pH/Temperature
 - Greatest negative value – Ammonia/Nitrate

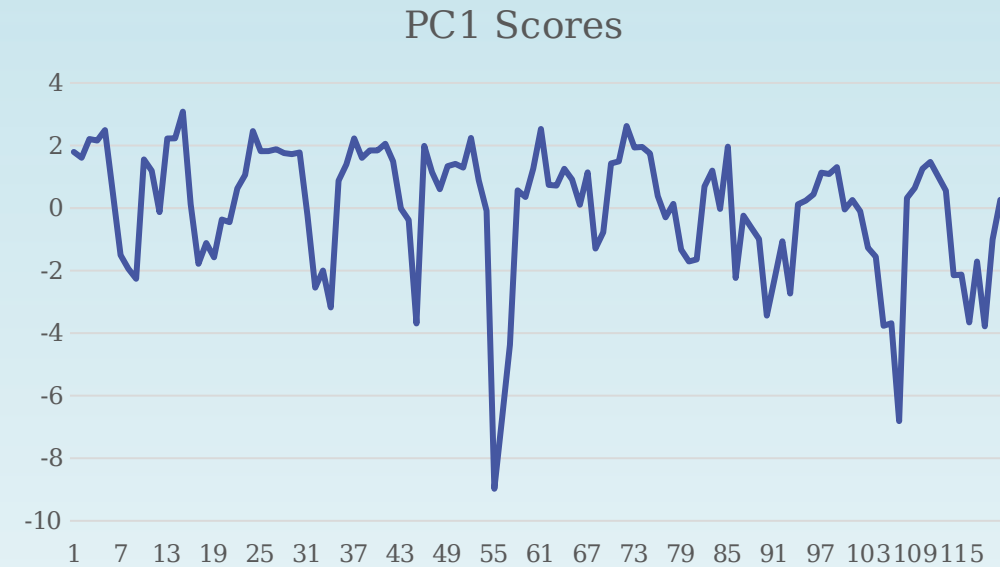
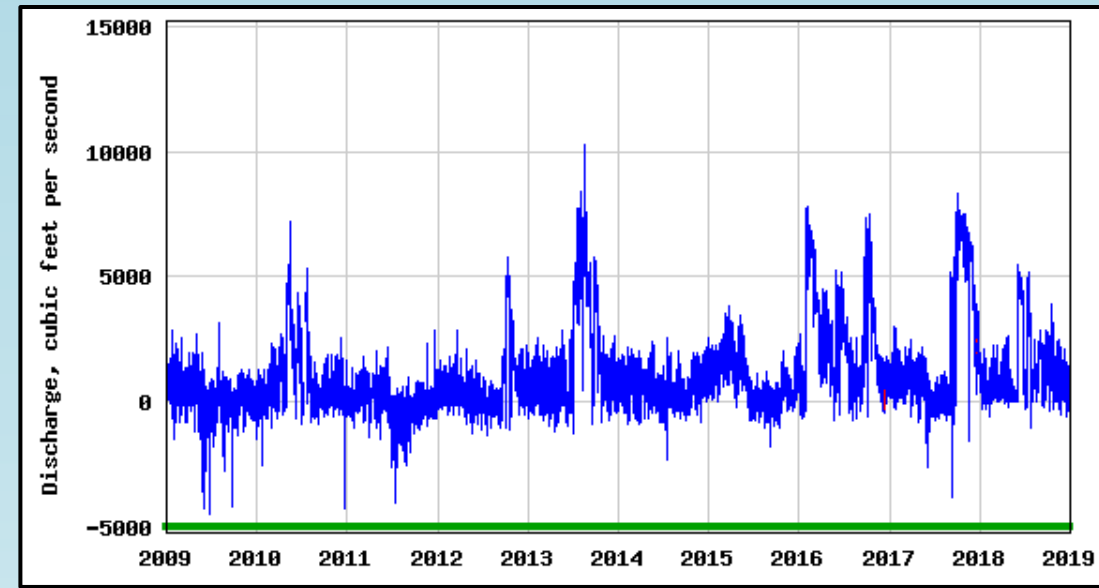


Loadings:												
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12
Oxygen, Dissolved, Electrode	0.147	0.639		0.195	0.316						0.616	0.203
Ammonia	-0.168		-0.510	0.406	-0.297	0.393	0.457	-0.196		0.185	0.131	
Nitrite	-0.289	0.282	-0.268	-0.188		-0.179	-0.473	-0.545	0.316	0.250		
Nitrate	-0.315		-0.399	0.114	-0.201		-0.419	0.330	-0.491	-0.325	0.118	0.170
pH		0.191	0.484	0.586	-0.521	-0.117	-0.173			0.121	-0.213	
Salinity by Meter	0.404	-0.136			-0.146	0.194	-0.357		-0.302	0.394	0.328	-0.521
Silica, dissolved as sio2	-0.399	-0.121				-0.245		0.508	0.101	0.687		
Phosphorous	-0.338	0.171	0.295	-0.153	0.158	0.135	0.205	-0.357	-0.686	0.211	-0.134	
Temperature	-0.259	-0.478	0.317	-0.112	-0.254			-0.256		-0.130	0.606	0.259
Nitrogen	-0.208	0.286	0.214	-0.367	-0.254	0.680		0.310	0.233			
Total Organic Carbon	-0.429	0.119	0.166	0.131		-0.196	0.124		0.106	-0.301	0.149	-0.757
Total Suspended Solids	0.179	0.284		-0.460	-0.552	-0.421	0.401				0.109	
SS loadings												
Proportion Var	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
Cumulative Var	0.083	0.167	0.250	0.333	0.417	0.500	0.583	0.667	0.750	0.833	0.917	1.000

	Oxygen, Dissolved, Electrode	Ammonia	Nitrite	Nitrate	pH	Salinity by Meter	Silica, dissolved as SiO2	Phosphorous	Temperature	Nitrogen	Total Organic Carbon	Total Suspended Solids
Oxygen, Dissolved, Electrode	1.00000000	-0.15697691	0.01544822	-0.1766764	0.244694903	0.08457504	-0.3370151	-0.01441407	-0.655004692	0.06881966	-0.08452484	0.19427653
Ammonia	-0.15697691	1.00000000	0.18656296	0.3914665	-0.113453934	-0.30283511	0.1931546	0.06932780	0.031586851	0.05652537	0.20705930	-0.13874760
Nitrite	0.01544822	0.18656296	1.00000000	0.5632786	-0.145136439	-0.47170079	0.3685789	0.34922034	0.091715108	0.29941513	0.47554430	-0.03623924
Nitrate	-0.17667636	0.39146647	0.56327860	1.00000000	-0.123749228	-0.42455111	0.5048909	0.28058842	0.129377898	0.20943620	0.50020925	-0.18090451
pH	0.24469490	-0.11345393	-0.14513644	-0.1237492	1.000000000	0.09116998	-0.1424832	-0.02262022	0.001693865	0.01900293	0.06185490	0.04395041
Salinity by Meter	0.08457504	-0.30283511	-0.47170079	-0.4245511	0.091169982	1.00000000	-0.6821614	-0.59525766	-0.296046982	-0.28829117	-0.85546214	0.21132099
Silica, dissolved as SiO2	-0.33701509	0.19315459	0.36857894	0.5048909	-0.142483230	-0.68216139	1.00000000	0.47878438	0.472506818	0.22564941	0.73826067	-0.31047673
Phosphorous	-0.01441407	0.06932780	0.34922034	0.2805884	-0.022620218	-0.59525766	0.4787844	1.00000000	0.340197177	0.44189030	0.65683448	-0.19166918
Temperature	-0.65500469	0.03158685	0.09171511	0.1293779	0.001693865	-0.29604698	0.4725068	0.34019718	1.000000000	0.15762071	0.43474983	-0.31264924
Nitrogen	0.06881966	0.05652537	0.29941513	0.2094362	0.019002931	-0.28829117	0.2256494	0.44189030	0.157620711	1.00000000	0.34292812	-0.01461618
Total Organic Carbon	-0.08452484	0.20705930	0.47554430	0.5002092	0.061854895	-0.85546214	0.7382607	0.65683448	0.434749832	0.34292812	1.00000000	-0.27918044
Total Suspended Solids	0.19427653	-0.13874760	-0.03623924	-0.1809045	0.043950412	0.21132099	-0.3104767	-0.19166918	-0.312649240	-0.01461618	-0.27918044	1.00000000

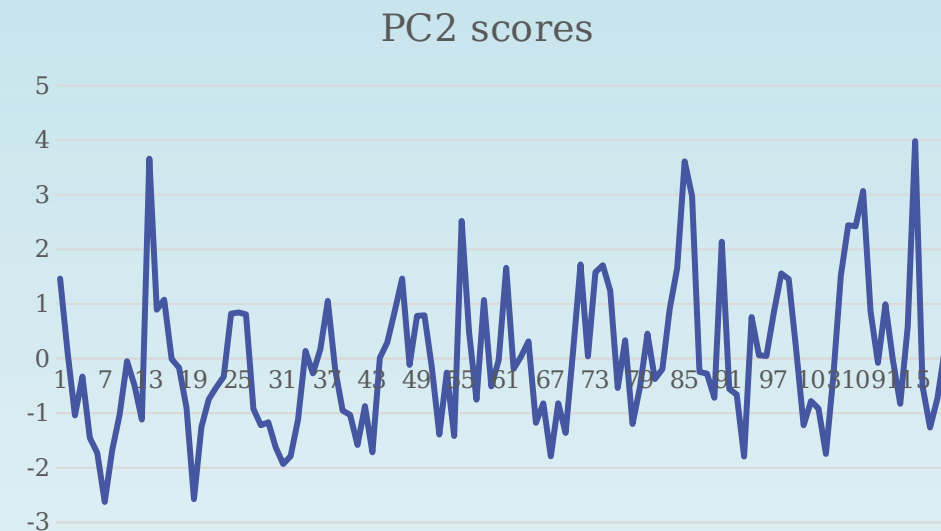
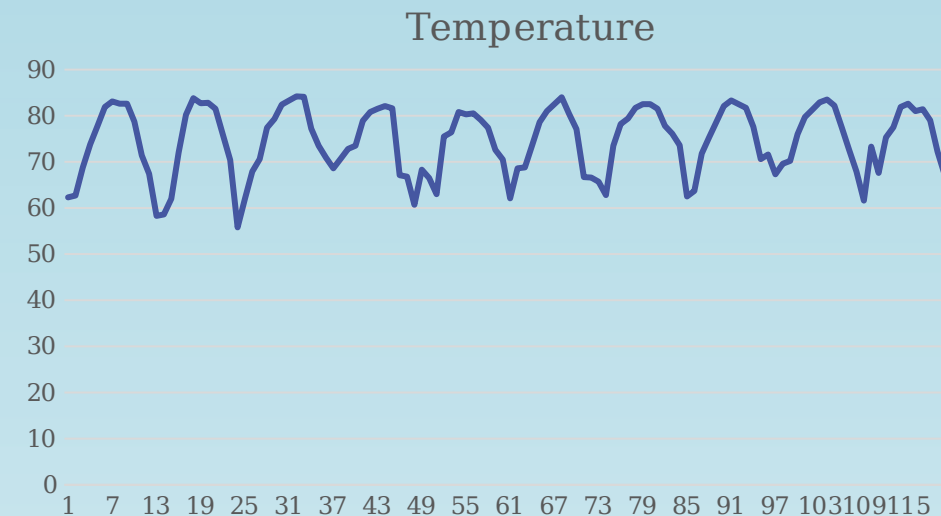
Discussion: PC1

- Salinity vs. Nitrate/Silica/Phosphorous/Total Organic Carbon
- Geography of the station
 - Located in seawater -> Caloosahatchee River Drainage
 - High nutrient loads deposited from Lake O -> Caloosahatchee
 - More freshwater deposits -> salinity decreases
 - When effects combined
 - Impacts on native species
 - Promote growth of invasive predators/Overpopulating species (algae)
- Lake O. Discharges/PC1 scores time-series
 - Same patterns/inverse
 - High negative loading values associated with N, Si, P, & TOC
- PC1 – agricultural runoff explains most of the



Discussion: PC2

- Dissolve Oxygen vs. Temperature
 - DO decreases as temperature rises
 - Oxygen less soluble at higher temperatures
 - Fish & other aquatic organisms cannot survive
- Lee county temperature/PC2 scores time-series
 - Temperature highest -> PC2 scores lowest
 - DO showed most variance in loadings with high negative values



Discussion.

PC3

- pH and Temperature vs Ammonia and Nitrates
 - pH/Ammonia showed greatest difference in variance
- Ammonia (basic pH)
 - Processes that create ammonia in water produce enough hydrogen ions to overcome/lower pH
 - pH and ammonia have a stronger interaction than all other variables in this component
- More ammonia -> more acidic

Conclusion:

- Principal component analysis provided very useful information to help understand the patterns occurring in the dataset
- Few calculations
 - Reduced multiple dimensions of the dataset to only 3 components
 - Identified most influential variable affecting water quality
- Found patterns that agree with other datasets
- The 3 PCs retained highlight interaction taking place between multiple variables
 - Salinity, Temperature, pH, and several chemical components influence water quality