<u>ATOC5860 - Homework 5 - due April 7, 2022</u>

Please send your homework to Jen/Prof. Kay on Slack as a direct message.

Please Name Your Homework Files: "ATOC5860_HW5_LastName.pdf, .html, .ipynb"

Your submissions should include: 1) A .pdf document with responses to the questions below, 2) Your code in both .ipynb and .html format.

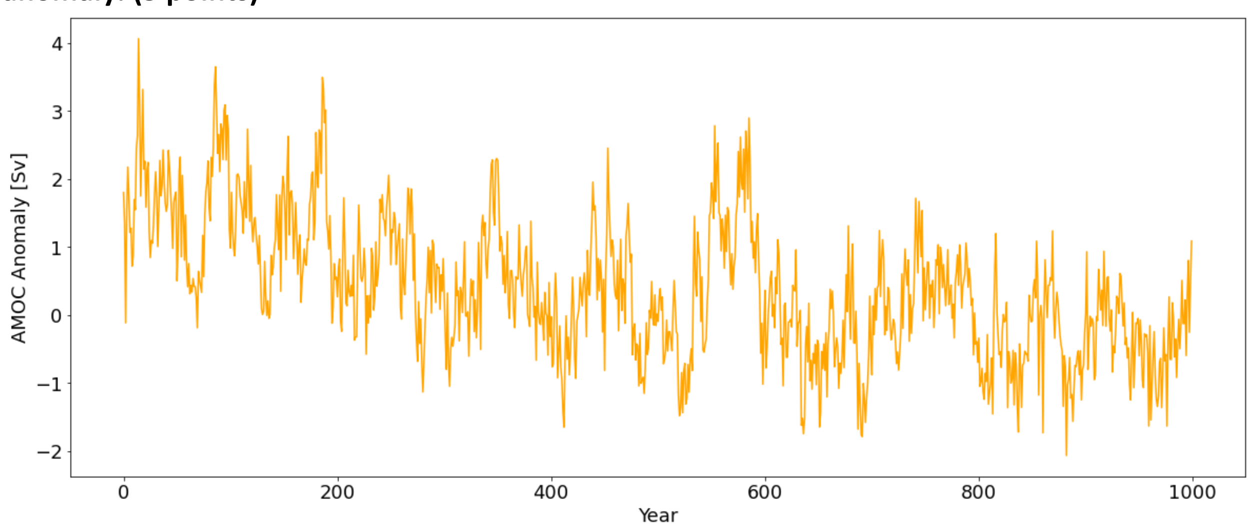
Show all work including the equations used (e.g., by referring to the Barnes Notes).
Write in complete, clear, and concise sentences.
Eliminate spelling/grammar mistakes.
Label all graph axes. Include units.
Report values using appropriate rounding.

Problem I) Filter out high frequency variability from a timeseries. (40 points total)

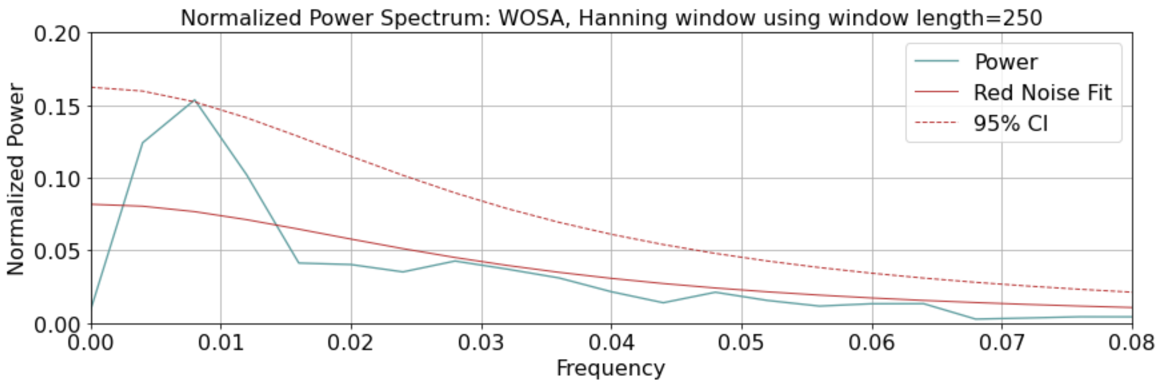
Your data: You will be analyzing a timeseries of the annual mean Atlantic Meridional Overturning Circulation (AMOC) in a climate model. The file containing the data is in netcdf4 format: CESM1_LENS_Coupled_Control.cvdp_data.401-2200.nc The variable you will analyze is "amoc_timeseries_ann". Use the first 1000 years.

```
float amoc_timeseries_ann(TIME) ;
    amoc_timeseries_ann:_FillValue = 1.e+20f ;
    amoc_timeseries_ann:long_name = "AMOC pc1 timeseries (annual)" ;
    amoc_timeseries_ann:missing_value = 1.e+20f ;
    amoc_timeseries_ann:units = "Sv" ;
```

a) Read in the data. Calculate the anomaly (remove the mean). Subset to the first 1000 years. Plot your data – a timeseries of the first 1000 years of the annual mean AMOC strength anomaly. (5 points)



b) Calculate the power spectra of the data. Use Welch's method (WOSA!) with a Hanning window and a window length of 250 years. Make a plot of normalized spectral power vs. frequency. Add the red noise fit and the 95% confidence intervals for testing statistical significance. Describe what you find. Is there any statistically significant spectral power, i.e., power at frequencies that cannot be explained using a red noise null hypothesis? (10 points)



The autocorrelation of the AMOC timeseries is calculated at a lag of 1 timestep. This step is performed following Barnes Chapter 2 Equation 68:

$$\gamma(\tau) = \overline{x'(t)x'(t+\tau)}$$

where $\gamma(\tau)$ is the autocovariance, τ is the lag of 1 timestep, and prime values indicate a perturbation from the mean. The autocovariance is normalized to the autocovariance at a lag of 0, or the population standard deviation $\overline{x'^2}$, to retrieve the autocorrelation. Next we calculate the e-folding time for a red noise timeseries with the same autocorrelation following Hartman Chapter 6 Equation 62:

$$T_e = \frac{-1}{\ln(\gamma(\tau))}$$

The autocorrelation of the AMOC data and the and e-folding time of the red noise process with the same autocorrelation are 0.82 and 5 years, respectively. We create a power spectrum for this red noise for comparison following Hartmann Chapter 6 Equation 64:

$$\Phi(\omega) = \frac{2T_e}{1 + \omega^2 T_e^2}$$

Next, we compute the critical f-score using a 95% confidence level. The critical f-score can be compared with the power spectra of the AMOC data to determine if peaks in power are statistically significant. When computing the f-score, we assume a null hypothesis that the data is red noise and not periodic. This value is computed from Barnes Chapter 4 Equation 34:

$$f = \frac{S_1^2}{S_2^2}$$

where S_1^2 and S_0^2 are the variances of the original data and the red noise fit, respectively. In order to reject the null hypothesis, the power must exceed the critical value.

Next we compute the power spectrum for the AMOC data using the Welch's method and an overlapping Hanning window of length 250 years. We calculate the degrees of freedom from Barnes Chapter 4 Equation 59:

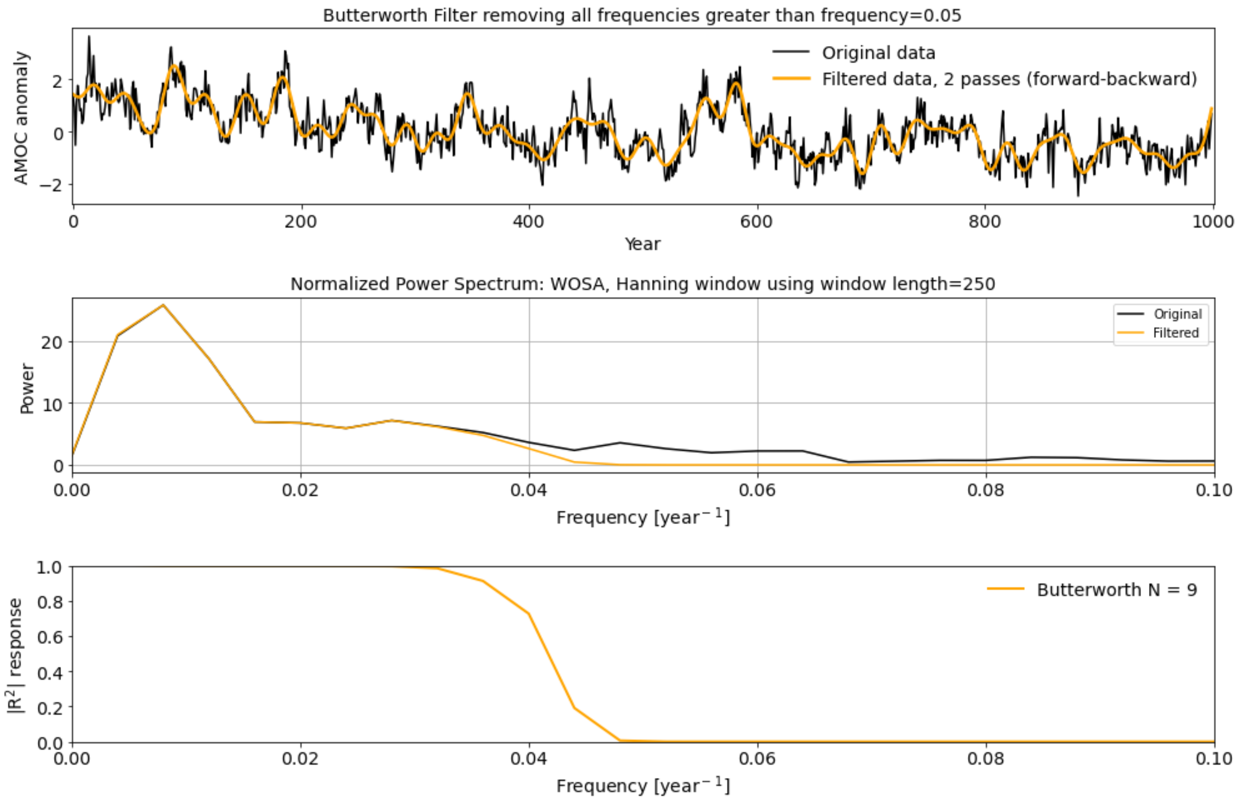
$$DOF = \frac{N}{(250/2)} f_w$$

where N is the timeseries length of 1000 years, 250 is the window length, and f_w is here assumed to be 1. This leaves a DOF of 8 for both datasets. Finally, we calculate the power using a built in function at each frequency which represents Barnes Chapter 4 Equation 42:

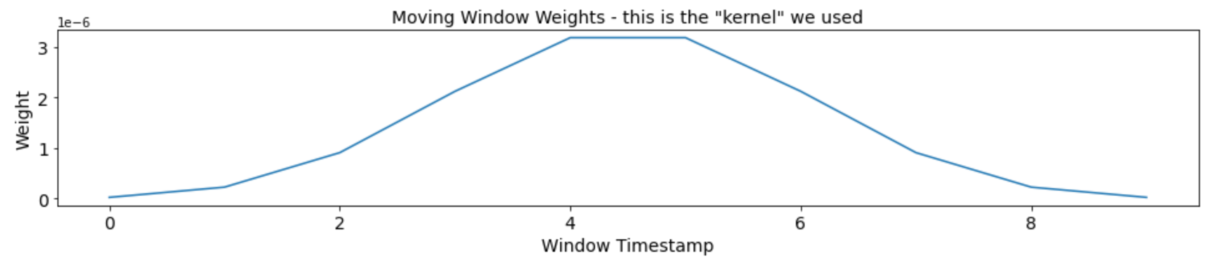
$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} d\omega$$

At the 95% confidence level, only one frequency exists where power is statistically significant. This power at occurs at a frequency of .008 yr^{-1} and has a power of 0.154. The power reduces at both higher and lower frequencies. This oscillation of anomalous transport corresponds with a cycle of 125 years. However, we note that at the 99% confidence level, there is no statistically significant power.

c) Apply a Butterworth filter with N=9 (number of weights equal to 9) to remove frequencies greater than 0.05 per year (i.e., variations in time shorter than 20 years). Make a three panel plot. The top plot should have the original data and the filtered data in the time domain. The middle plot should have the power spectra of the original and filtered data. Note: Do not normalize the power spectra so that they are easier to compare. Do not forget to include units. The bottom plot should show the response function in frequency space for the Butterworth filter. (15 points)



d) Write a paragraph summarizing the results shown in each of the panels. Explain how you calculated the values in all three of the panels, including equations if needed. Did the Butterworth filter do the job of removing the high frequency noise? (10 points)



We apply a Butterworth low-pass filter to remove the high frequency oscillations from the dataset. Different filtering weights are applied using moving windows of length 9. Weighting is applied using a weighted average, shown above. The new timeseries is calculated from Barnes Chapter 4 Equation 114:

$$y(t) = \sum_{\tau=-J}^{0} b_\tau \cdot x(t+\tau) + \sum_{\tau=-J}^{-1} a_\tau \cdot y(t+\tau)$$

where $x(t)$ is the original data, $y(t)$ is the new filtered dataset, a_τ and b_τ are the weights for each lag, and τ represents the lags from 1 to 9. Note the Butterworth filter is non-recursive and includes the filtered dataset. As before, we calculate the Fourier Transforms of both the filtered and unfiltered datasets from Barnes Chapter 4 Equation 42:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}d\omega$$

where $F(\omega)$ is the Fourier Transform in frequency space of the timeseries $f(t)$. Finally, we calculate the response function for a Butterworth filter from Barnes Chapter 4 Equation 115:

$$|R(\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2N}}$$

where ω is each frequency, ω_c is the cutoff frequency of 0.05 yr^{-1}, and N is the number of weights, here chosen to be 9. Finally, to verify that the filtering has removed high frequency oscillations, we calculate the autocorrelation of both the original and filtered timeseries at a lag of 1 using Barnes Chpt. 2 Eq. 67:

$$\gamma(\tau) = \frac{x'(t) \cdot x'(t+\tau)}{(N-\tau)\gamma(0)}$$

where primes indicate perturbations from the mean, N is the timeseries length, τ is a lag of 1, the dot product incorporates the sum, and the resulting value is normalized by $\gamma(0) = \overline{x'^2}$, the variance of the population without lag.

This method has removed the high-frequency oscillations from the timeseries in the filtered line plot (top). We can verify this because the timeseries has increased in "redness" moving between unfiltered to filtered from 0.82 to 0.99. Due to this filtering, the power at frequencies greater than the cutoff frequency decrease in relation to the original power (middle). However, the power does not reduce instantaneously at the cutoff. Since the Butterworth method applies low-pass filtering, power at frequencies less than the cutoff remain mostly unaffected. This is because the weighting function has low tangency, i.e. not rectangular, by ramping up to the maximum weight and then ramping back down to reduce lobe effects. Thus, the reduction in power begins somewhat before the cutoff frequency as can be seen by the decrease in response before 0.05 yr^{-1} and continuing after the cutoff (bottom). In the case that boxcar weighting were applied, we would observe that the response function immediately decrease to 0 at the cutoff frequency.

Problem II) Apply filtering to a dataset of your choice (60 points)

a) **Identify a data time series - X(t) – that is of interest to you. X(t) can be anything…. just be sure your data are evenly spaced in time and do not have any missing values. Provide a thorough description of your data including the reference, the variable (including names and units), the sampling frequency, etc. Describe why you think filtering may be useful to apply to your data. (5 points)**

Data for this analysis is acquired by running a Weather Research and Forecasting (WRF) model simulation using the Wind Farm Parameterization (WFP) (Fitch, et al., 2012) for the course of 1 month. The data timeseries is total power production in MW from the planned Vineyard wind farm offshore of the U.S. East Coast. This data has a constant output frequency of 10 minutes.

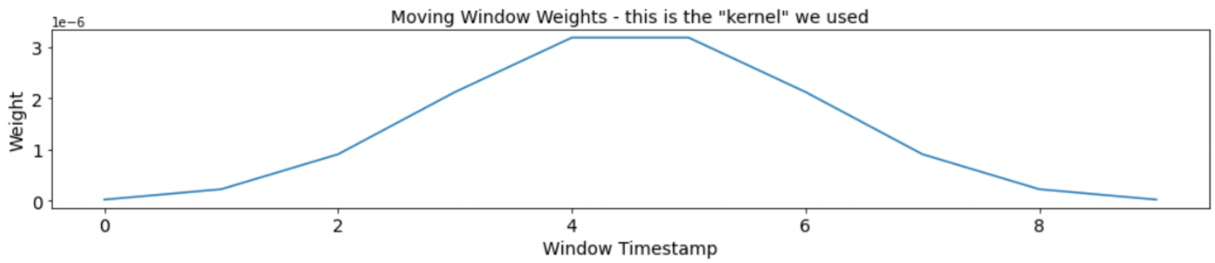
Wind farms are being sited in this region to supplement power from renewable energy to the New England utility grid. However, some planners may be less concerned with the high-frequency oscillations in power output from wind turbines. This is because most of the power production along the U.S. grid is supplied by synchronous generators (coal and gas power plants). Because synchronous generators rotate and have angular momentum, they have high resiliency to short temporal fluctuations. Thus, the high frequency fluctuations in power output from wind farms, those that could be responsible for brown-outs if wind energy operated alone, may not need to be considered when analyzing power output for planning purposes. Furthermore, WRF models are susceptible to numerical noise. This noise can exist as high frequency oscillations in wind speed that cause incorrect power output from the model and thus need to be removed.

Fitch, A.C., Local and Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model. Monthly Weather Review 140, 3017–3038 (2012).

b) Identify frequency variations within your data that you wish to remove. Explain what you expect to see when you apply filtering to your dataset. Do this **before you do the analysis. (There is not any penalty for it being wrong…!) Commonly removed frequencies include the seasonal cycle, the daily cycle, or noise at high or low frequencies (5 points)**

The frequency variations to be removed from this dataset are those greater than an hourly cycle. This corresponds to a frequency of 0.16667 since the data is at a 10-min output frequency. I expect that applying this type of filtering will have a rather small impact on the data. However, I am hoping to see that the high-frequency variations in power output are smoothed out.

c) Describe the method including equations that you plan to use to remove frequencies from your time series. Why did you pick this method? (15 points)



We apply a Butterworth low-pass filter to remove the high frequency oscillations from the power output. Different filtering weights are applied using a window length of 6. Weighting is applied using a moving weighted average, shown above. The new timeseries is calculated from Barnes Chapter 4 Equation 114:

$$y(t) = \sum_{\tau=-J}^{0} b_\tau \cdot x(t+\tau) + \sum_{\tau=-J}^{-1} a_\tau \cdot y(t+\tau)$$

where $x(t)$ is the original data, $y(t)$ is the new filtered dataset, a_τ and b_τ are the weights for each lag, and τ represents the lags from 1 to 6. Note the Butterworth filter is non-recursive and includes the filtered dataset, $y(t)$. As before, we calculate the Fourier Transforms of both the filtered and unfiltered datasets from Barnes Chapter 4 Equation 42:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} d\omega$$

where $F(\omega)$ is the Fourier Transform in frequency space of the timeseries $f(t)$. Finally, we calculate the response function for a Butterworth filter from Barnes Chapter 4 Equation 115:

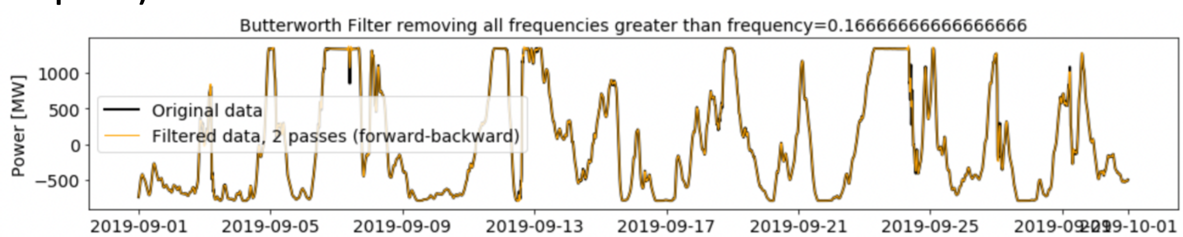
$$|R(\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2N}}$$

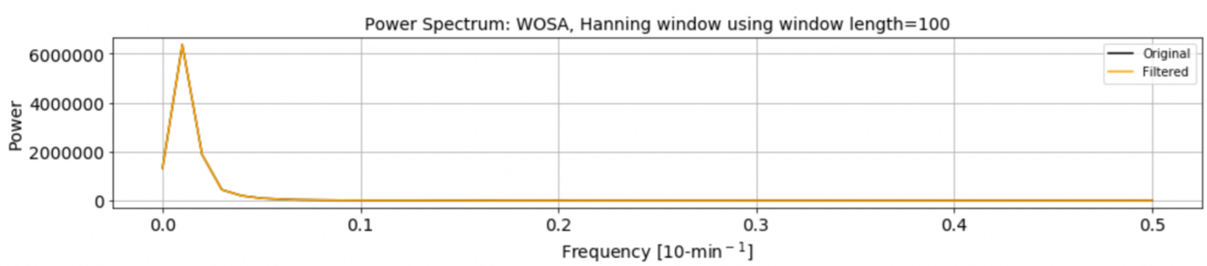
where ω is each frequency, ω_c is the cutoff frequency of 0.1667 yr^{-1}, and N is the number of weights, here chosen to be 6. Finally, to verify if the filtering has removed high frequency oscillations, we calculate the autocorrelation of both the original and filtered timeseries at a lag of 1 using Barnes Chpt. 2 Eq. 67:

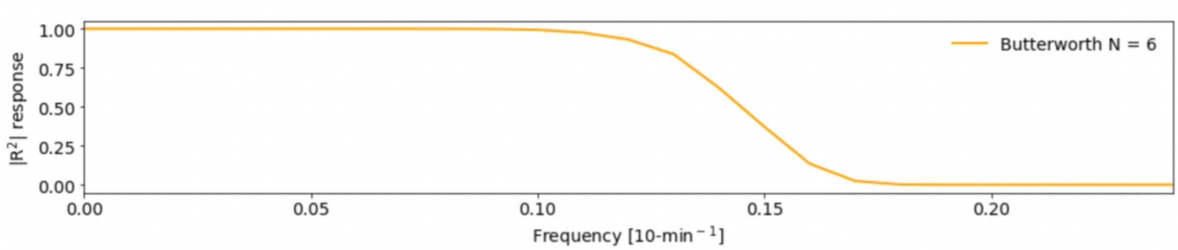
$$\gamma(\tau) = \frac{x'(t) \cdot x'(t+\tau)}{(N-\tau)\gamma(0)}$$

where primes indicate perturbations from the mean, N is the timeseries length, τ is a lag of 1, the dot product incorporates the sum, and the resulting value is normalized by $\gamma(0) = \overline{x'^2}$, the variance of the population without lag.

d) Filter your data. Include relevant figures (e.g., those requested in part problem Ic). (25 points)







e) Discuss your results. Did your filtering work? What were the issues that you found as you were crafting the ideal filter for your data? (10 points)

Filtering did not provide benefit for this application. The timeseries (top) shows that the fluctuations in power output have remained largely unchanged. Significant high-frequency changes in power output, those that do not need to be considered in a scenario with resilient power supplement to the grid, have been smoothed out somewhat. This is apparent during 2019-09-07 where power output remains relatively constant before a considerable reduction. However, the filtered timeseries output still reduces and the impact from filtering is. The power spectrum has also not changed. I assumed that there would be two major peaks in power; a daily cycle that corresponds with diurnal wind flow and a high-frequency peak due to numerical noise. However, only one peak is evident for both the filtered and unfiltered datasets. Furthermore, the autocorrelations of the filtered and unfiltered datasets are both 0.998, indicating that no major changes have occurred. One hurdle in crafting the ideal filter was choosing the correct window length. Here, most lengths had a negligible effect on the resulting timeseries and power spectra. Furthermore, numerical noise in the model typically occurs due to sub-grid scale processes. From this, the noise appears at the shortest timestep of 10-min due to leakage. Unfortunately, choosing window lengths between the hourly cycle and the shortest possible cycle of 10 minutes had no effect. For future analysis, I would probably use simpler methods, such as hourly averaging to reduce the effects of noise.

Note for grading Problem II. You are analyzing your own data. Since only you know the "right answer", you will be largely graded on how well I can follow your description of the data, the methods, the results, and the conclusions. Keep your code and your explanations simple, clear, and easy to follow. Spend the time to make your code concise, clear, and well documented. Look at your code as an opportunity to re-enforce the understanding you have gained in class as expressed through analyzing your own data.