This course provides the basic tools and skills for data analysis within ocean sciences. The focus is on the *quantitative* aspect, and really on the *tools*, though the examples and context will be largely from the ocean science context. Familiarity with the relevant problems in ocean science to be tackled would be beneficial but not completely necessary; if anything, one can potentially learn a bit more about the topic concerned by actively looking through the data itself rather than passively taking on content. A particular focus of this course is the active/experimental/hands-on type learning, provided through guided computer workshops, and supplemented with lectures (not the ordering, that it is not the computer workshops supplementing the lectures).

Some misc. notes

- The proposed schedule is once a week three hour sessions, first 45 mins being a lecture and the remaining being a computer workshop.
- The course requires you to be not afraid of computer programming and generally experimenting; prior experience extremely beneficial but not strictly necessary.
- The programming language of choice is going to be Python, partly because the argument that if you can do Python you can probably use Excel and/or R. Python will be used through Jupyter notebooks.
- It is extremely beneficial if you use your own laptop for this, as you then in principle have complete control over what Python/Jupyter packages and data you have on your machine (highly recommend the package management is through Anaconda; installation details to be provided). Failing that, computer barns will have versions of Python/Jupyter on there, though you are potentially stick with doing stuff at the computer barn. Even failing that, you could probably use Google Colab on the cloud, but then you have to live with potentially downloading packages every time you refresh.
- You can work in groups but there are no group assessments.

Assessment

25% attendance, $3 \times 25\%$ assessment

NOTE: The threshold for passing will be higher for this course, at around 70%.

Learning outcomes

- basic understanding of data relevant to ocean sciences
- basic and standard analysis techniques for data using a programming language
- awareness of data repositories, formats and possible issues with quality within ocean sciences
- experiment data exploration as an avenue for scientific understanding
- data presentation, complemented accordingly by science and prose

Supplementary material list

(Keyword being "supplementary")

- Google + stack exchange
  - seriously, the main resource for coding, syntax and bug fixing
- Spiegelhalter (2019), "The Art of Statistics: Learning from Data", Pelican Books
  - excellent book about statistics, its use as a tool, interpretation, and how statistics is / can be sabused
- Takahashi (2008), "The Manga Guide to Statistics", No Starch Press
  - actually a very solid and careful presentation of useful statistics
  - has examples in Excel
- Abernathey (continuous), Python for Scientific Computing
  - another course on related concepts
  - `https://github.com/rabernat/python_teaching`
- many others online, why not try Google (which itself is probably going to be more updated than books)?

Proposed Syllabus

[Some slack/flexibility in timing and pace of lectures]

L01 Basics of Jupyter, Python, and data structures

L02 More Python (simple data reading, plotting, indexing, data manipulation), basic statistics (simple data reading, mean, standard deviation, more plotting, using packages)

L03 Linear regression (line of best fit, plotting, correlation coefficients, polyfit, under/overfitting, extrapolation, trends)

L04 Multi-linear regression (plotting, correlations, principal component analysis, more plotting)

*** assignment 1 (2 weeks to do) ***

L05 Statistical tests (elementary probability, probability distribution functions, historgram, hypothesis testing, χ^2 tests)

L06 Statistical tests (t-tests, F-tests, types of errors, interpretation, common pitfalls, more plotting)

L07 Time-series analysis (reading data, averaging windows, high/low pass filters, auto-correlation, trend, and regression)

L08 Time-series analysis (lag correlation, Fourier analysis and Fourier transforms, de-trending, power spectrum, more plotting)

*** assignment 2 (2 weeks to do) ***

L09 Fun with maps (NetCDF data, data acquisition, basics of map plotting, map projections, animations, inter/extrpolation)

L10 Spatio-temporal data (processing into e.g. time-series data, more analysis, more plotting, empirical orthogonal functions)

*** assignment 3 (3 weeks to do) ***

L11 Optional topic: BIG data analysis (xarray, Dask etc.)

L12 Optional topic (machine learning) or slack session

L13 Slack session