Boring but important disclaimers:

- If you are not getting this from the GitHub repository or the associated Canvas page (e.g. CourseHero, Chegg etc.), you are probably getting the substandard version of these slides Don't pay money for those, because you can get the most updated version for free at

  `https://github.com/julianmak/academic-notes`

  The repository principally contains the compiled products rather than the source for size reasons.
- Associated Python code (as Jupyter notebooks mostly) will be held on the same repository. The source data however might be big, so I am going to be naughty and possibly just refer you to where you might get the data if that is the case (e.g. JRA-55 data). I know I should make properly reproducible binders etc., but I didn't...
- I do not claim the compiled products and/or code are completely mistake free (e.g. I know I don't write Pythonic code). Use the material however you like, but use it at your own risk.
- As said on the repository, I have tried to honestly use content that is self made, open source or explicitly open for fair use, and citations should be there. If however you are the copyright holder and you want the material taken down, please flag up the issue accordingly and I will happily try and swap out the relevant material.

OCES 3301 :

basic Data Analysis in ocean sciences

Session 1: logistics, Jupyter notebooks and Python

Outline

- canvas, Zoom and admin things
- approach of the course
  $\rightarrow$ lecture, workshop, assessment
- demonstration

Content of course is somewhat ocean science motivated, but skills are entirely generic

Practicalities

Instructors:

I Julian Mak (`jclmak@ust.hk`)

TA Jonathan Lee (`hcleear@connect.ust.hk`)

Course grade breakdown:

method	%
~~participation~~	~~20~~
assignments	$25 \times 4 = 100$

- pass mark is going to be 60%
- ”A” boundary will be around 85-90%

  (I don’t grade to a curve)

- assignments with plagiarism will get a minimum of **zero**

Practicalities

- weekly **Fri 1330 to 1630**

  $\rightarrow$ first 40-60 mins is lecture / outline

  $\rightarrow$ rest of the time is computer workshop, which is the **main learning part**

- F2F in class

  $\rightarrow$ it's more for you: us trying to help you code/debug without control/access to your computer is going to be incredibly irritating

Approach of course

What this course is:

- a hands-on intro to basic data analysis tools and Python

  $\rightarrow$ on understanding principles, learning a few tools well, and a tackle a few problems but thoroughly (less is more in this case...)

  $\rightarrow$ transferable skills

- **you have to try do stuff**

  $\rightarrow$ the lectures are entirely supplementary to the computer workshops

- **focus on teaching you to teach yourself, and thinking algorithmically**

  $\rightarrow$ you will probably fail if you try and focus on cramming over learning

Approach of course

This course is <u>not</u>:

- a computer science and/or Python course as such

  $\rightarrow$ we use Python but the focus is the analysis tools

Approach of course

This course is not:

- a computer science and/or Python course as such

  $\rightarrow$ we use Python but the focus is the analysis tools
- a statistics/maths course

  $\rightarrow$ you will be dealing with numbers quite a bit

Approach of course

This course is not:

- a computer science and/or Python course as such

  $\rightarrow$ we use Python but the focus is the analysis tools
- a statistics/maths course

  $\rightarrow$ you will be dealing with numbers quite a bit

After this course you will not be:

- a computer wizard

  $\rightarrow$ it will hopefully give you some mindset and tools to work towards being one (if you want to)

Approach of course

This course is not:

- a computer science and/or Python course as such

  → we use Python but the focus is the analysis tools
- a statistics/maths course

  → you will be dealing with numbers quite a bit

After this course you will not be:

- a computer wizard

  → it will hopefully give you some mindset and tools to work towards being one (if you want to)
- know how to solve every data problem in ocean science

  → but it should teach you how to look up how you might get started on other problems

Material

- most of the material is in the Jupyter notebooks

  $\rightarrow$ most things in slides are actually taken from the notebooks

  $\rightarrow$ if you don't want to listen to me you could always go straight to the notebooks

Material

- most of the material is in the Jupyter notebooks

  $\rightarrow$ most things in slides are actually taken from the notebooks

  $\rightarrow$ if you don't want to listen to me you could always go straight to the notebooks

- for extra reading, try Google (search the topic)

Material

- most of the material is in the Jupyter notebooks

  $\rightarrow$ most things in slides are actually taken from the notebooks

  $\rightarrow$ if you don't want to listen to me you could always go straight to the notebooks

- for extra reading, try Google (search the topic)
- for Python syntax queries, try Google (stack overflow / stack exchange)

Material

- most of the material is in the Jupyter notebooks

  $\rightarrow$ most things in slides are actually taken from the notebooks

  $\rightarrow$ if you don't want to listen to me you could always go straight to the notebooks

- for extra reading, try Google (search the topic)
- for Python syntax queries, try Google (stack overflow / stack exchange)
- for help with debugging, try Google (stack overflow / stack exchange)

Material

- most of the material is in the Jupyter notebooks

  → most things in slides are actually taken from the notebooks

  → if you don't want to listen to me you could always go straight to the notebooks

- for extra reading, try Google (search the topic)
- for Python syntax queries, try Google (stack overflow / stack exchange)
- for help with debugging, try Google (stack overflow / stack exchange)
- for help with science queries, try Google

- you could try me but I would probably tell you to try Google first

Syllabus

S01 Jupyter + Python

S02 reading data + basic stats

S03 regression

S04 regression

S05 statistical testing

S06 statistical testing

S07 time series

S08 time series

S09 fun with maps

S10 fun with maps

- lecture material is **cumulative**
- assignments at S02, S05, S08, and S10

  $\rightarrow$ model assignment and marking criteria will be provided when assignment 1 is released, and applicable to all subsequent assignments

Demonstration

(going to skip the spiel about why is data analysis important etc.)

- GitHub repository
- Jupyter notebooks
- Python (maybe some Anaconda and/or Google CoLab)
- data types (strings, floats, integers, boolean)
- lists, arrays, dictionaries
- indexing, basic manipulations
- loops, conditionals, subroutines (no modules/classes here)

Moving to a Jupyter notebook $\rightarrow$