

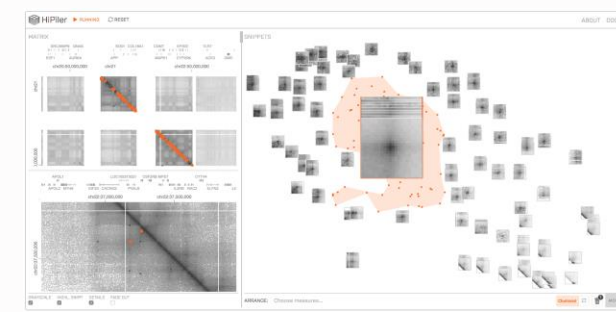


IM339: Data science Across Disciplines: Principles, Practice and Critique

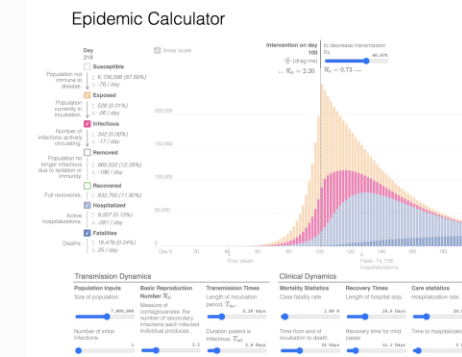
Term 1

Dr Kevin Narasimhan (Module Convenor)
Professor Cagatay Turkey
Dr Ching Jin
Esha Nasir

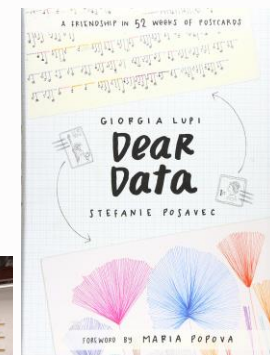
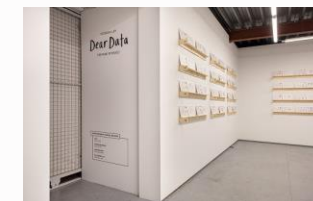
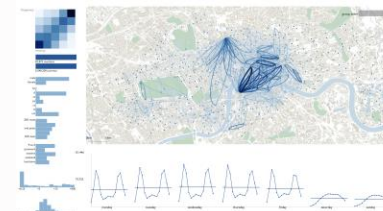
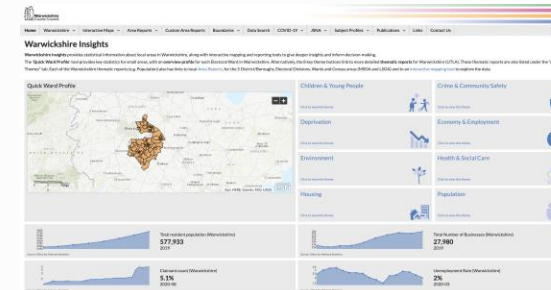
UNIVERSITY
OF WARWICK



Lekschas et al., 2017



<https://gabgoh.github.io/COVID/>



Beecham and Wood, 2014

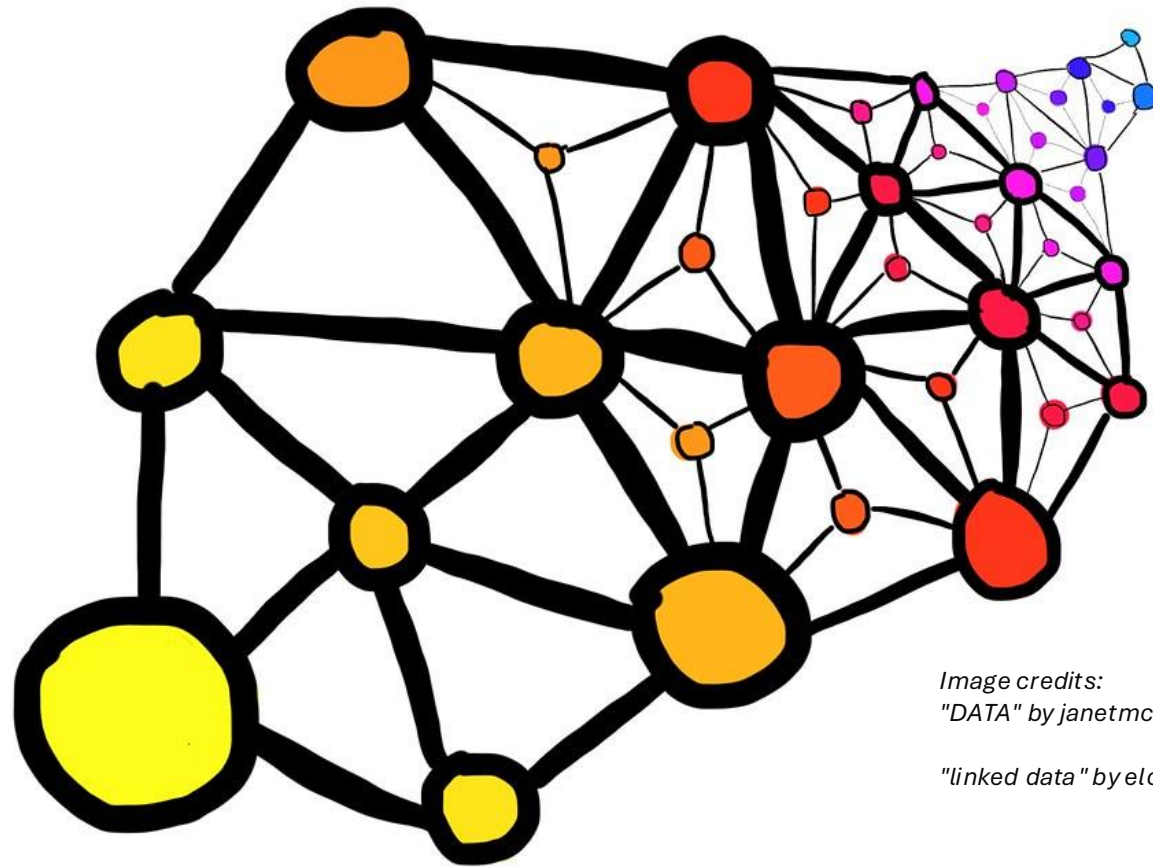


Image credits:

"DATA" by janetmck is licensed under CC BY 2.0

"linked data" by elcovs is licensed under CC BY-SA 2.0

IM339: Data science Across Disciplines: Principles, Practice and Critique

Dr Kavin Narasimhan (Module Convenor), Professor Cagatay Turkey,
Dr Ching Jin and Esha Nasir

Centre for Interdisciplinary Methodologies



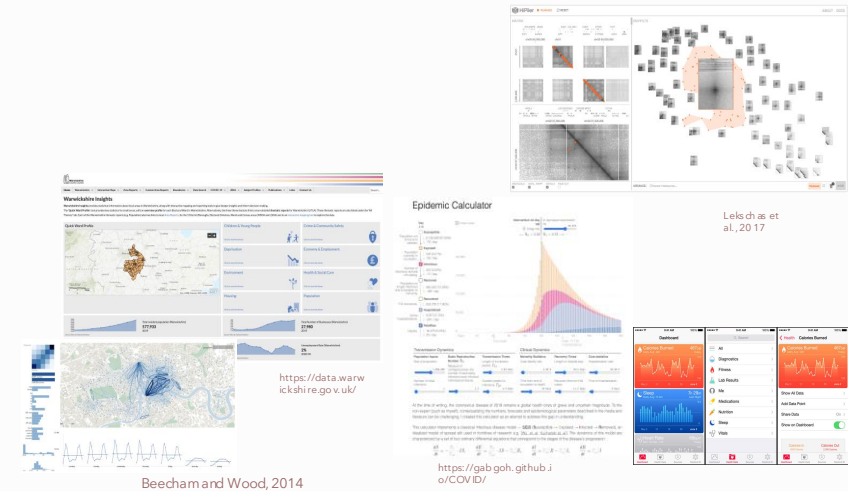
UNIVERSITY
OF WARWICK

IM339: Data Science Across Disciplines

... introduces fundamental **techniques, concepts, and contemporary discussions** across the broad field of data science

We will:

- discuss the **theoretical & scientific frameworks** of data science
- introduce the **fundamental techniques** for working with data through appropriate procedures, algorithms, and visualisation
- engage with **contemporary discussions** in data science
- apply data science techniques to **diverse problems**
- **critically approach** data and data-driven artefacts



IM339: Data Science Across Disciplines

Indicative Syllabus*

Session 1: Introduction, Historical Perspectives & Basic Concepts

Session 2: Thinking Data: Theoretical and Practical Concerns

Session 3: Abstractions & Models

Session 4: Structures and Spaces

Session 5: Multi-Model Thinking and Rigour in Data Science

Session 6: Recognising and Avoiding Traps

Session 7: Data Science & Society

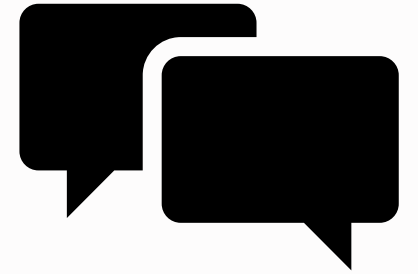
Sessions 8 & 9: Data Science Workshops

* There could be minor changes/adaptations

IM339: Data Science Across Disciplines

A pæddiŋ & active style

- Lectures and content will be released before the sessions
- We get together for discussions, questions, and reflections
- We start each week with a brief recap before diving into questions
- Discussion questions/prompts each week



IM339: Data Science Across Disciplines

Lectures/Labs/Workshops

Weeks 1–5 , 7 & 8:

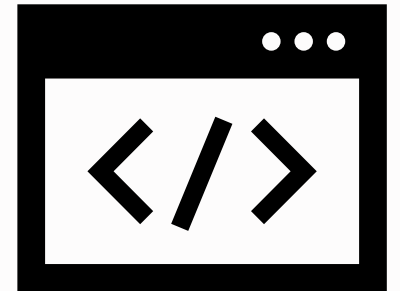
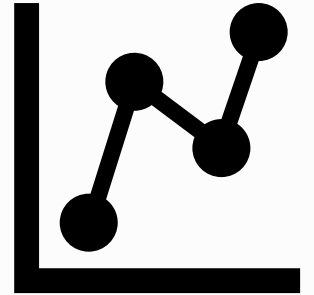
- Video recordings of lecture content & weekly readings
- 1-hour online discussion and Q&A session
- 2 hours of hands-on coding labs using Python and Jupyter Notebooks
- Working with data and algorithms on applied examples

Weeks 9 & 10:

- Workshops where we go through the full Data Science process

Week 6

- Reading Week



IM339: Data Science Across Disciplines

Labs and Workshops

We will use Python as the main programming environment and Anaconda as the installation medium

We will work through Jupyter notebooks with annotations

We will use a range of libraries and packages

- Pandas, Numpy, Scipy, Altair, Seaborn...

We will work with data sets, conduct analyses, and share results

Lab work can be done at your own pace, self-driven, and released beforehand

Exercises each week accompanied by formative feedback



IM339: Data Science Across Disciplines

Coding?

Don't worry if you are new to coding!

We'll support you as best as we can:

- Introductory materials in the first few weeks
- Python skills workshop

Make use of online resources and tutorials

- Some links on Moodle:
 - Beginners in Python : <http://www.codecademy.com/en/tracks/python>
 - Python for Non-programmers: http://en.wikibooks.org/wiki/Non-Programmer%27s_Tutorial_for_Python_3
 - <http://www.learnpython.org/>
 - <https://wiki.python.org/moin/IntroductoryBooks>

```
import pandas as pd
import numpy as np
import re
import networkx as nx
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

from os import listdir
from os.path import isfile, join

""" This is a processing script that contains a set of functions to read in
an infection map file and generate an internal networkx graph which can then
be exported to viz-friendly formats such as graphml. """

"""
Compared to the previous version where the code was just reading the infection map, this version
reads through the several files, computes statistics, and integrates these within the graph data
structure. So contains a number of functions to help with these.
"""

# A series of global variables to hold the data loaded from the auxillary files
inputContacts = None
eventsDF = None
nodeAges = pd.DataFrame()

virusEventsDF = None
statesPerNode_vir = None
timeSeriesPerNode_vir = None
eventTypesPerNode_vir = None

def loadInputDataFilesToDataFrames():
    global inputContacts;
    global nodeAges;

    inputContacts = pd.read_csv('../data/2020_07_01_newSet/contacts_covid_model.csv')

    # Read in ages
    nodeAges = pd.read_csv('../data/2020_07_01_newSet/ids_Paul.csv')

def loadEventDataFilesToDataFrames(eventFileName):
    global eventsDF;
    global virusEventsDF;
```


IM339: Data Science Across Disciplines

Assessments

- **A critical review** of a selected Data Science project, examined through a critical lens covered during the lectures.
 - Weighting: 40%
 - Word count: 1000 words (15 CATs); 1500 words (30 CATs)
- **A data-intensive essay** in which students report on a data science project they carried out using a chosen question and an appropriate data set. The essay will be reporting on the data science process from initiation to evaluation to reflection while engaging with the relevant literature in the domain.
 - Weighting: 60%
 - Word count: 1500 words (15 CATs); 2500 words (30 CATs)

Plus, Formative Feedback elements:

- Regular feedback on lab exercises
- End of workshop presentation from working groups