CSC 233 Intro to Data Analytics – Week 01

Today (105')

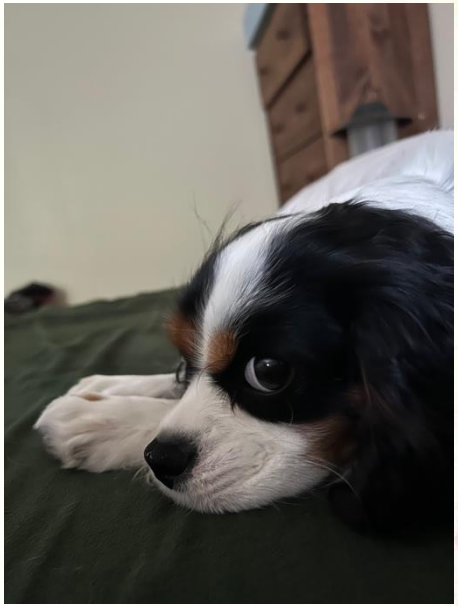
- Intros (30')
- Fermi questions (15')
- What is DA? (30')
- Syllabus etc. (15')
- Notebook 1 (15')

Introductions

Me:

- Georgia Doing, PhD
- she/her/hers
- My interests: computational microbiology, small cells & big data

My dog Ginny



* Please email me ahead of office hours if you would like me to make sure Ginny won't be there, not a problem at all!

You:

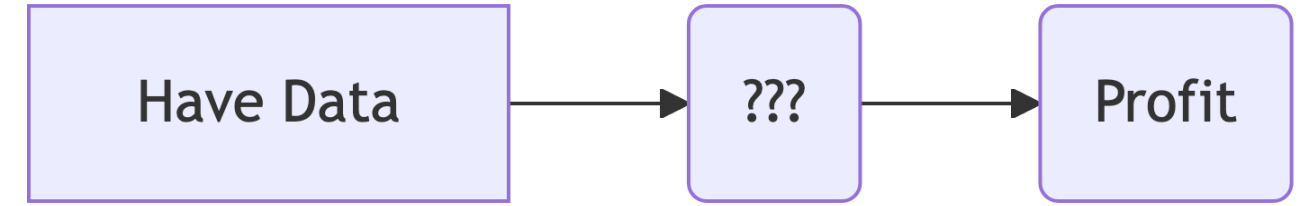
- Name and pronouns
- Help me find you on my roster
- What persepctive do you bring to this class? Could be from a life experience, memory, etc. or your major, minor, research, internship, etc.
- Is there a dataset you've wrestled with in the past that you wished you could have analyzed more deeply?

Fermi questions

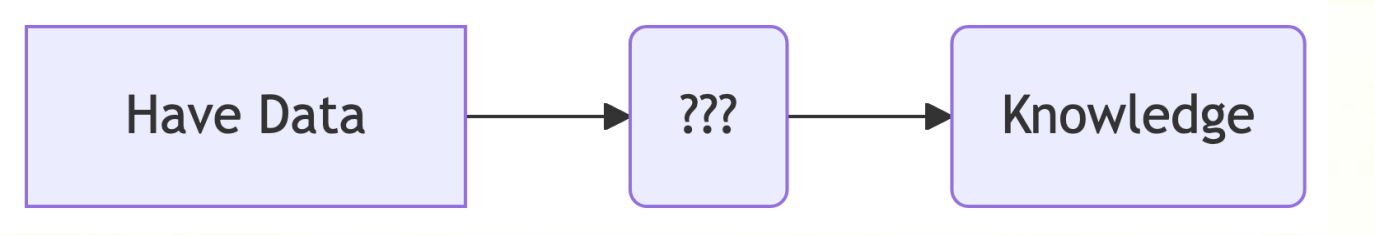
https://fermi-questions.andrechek.com/

- in groups of 3
- if you want a different question, refresh page
- 10 mins
- answers are in orders of magnitude

What is ‘Data Analytics’?

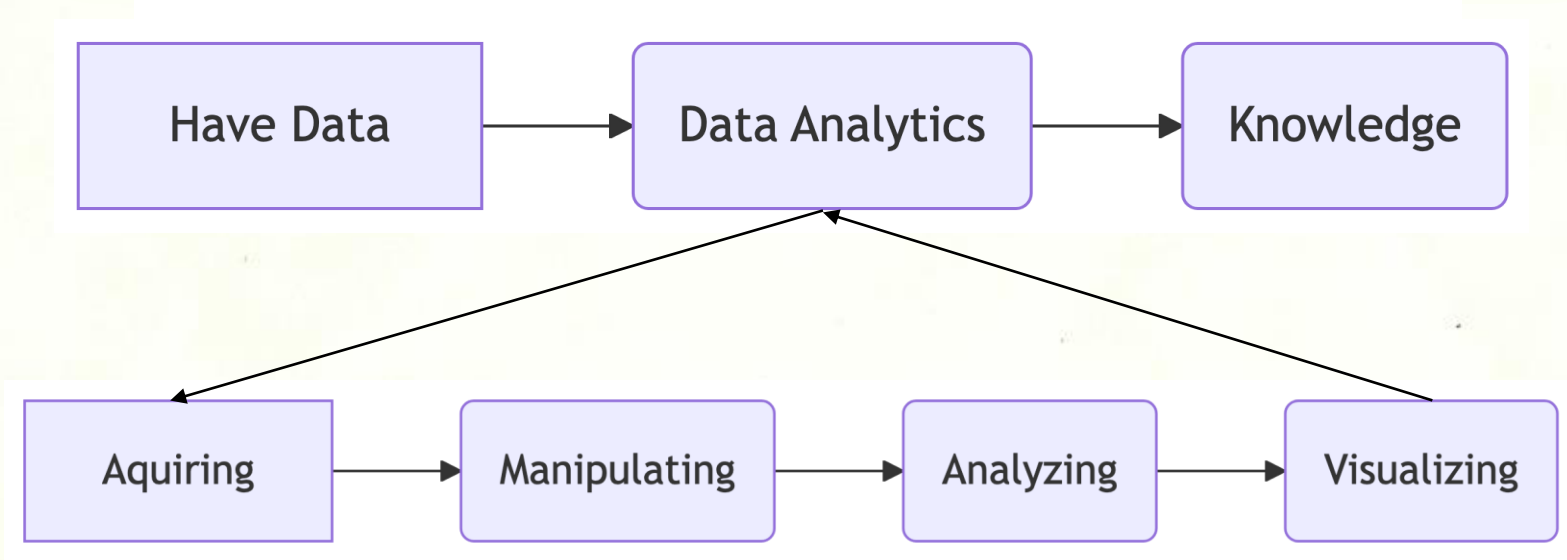


What is ‘Data Analytics’?



What is ‘Data Analytics’?

What is ‘Data Analytics’?



An Example

Screenshot of Smithsonian database

Smithsonian
National Museum of Natural History

Donate

Plan Your Visit | Exhibitions | Education | Research & Collections | About Us | Calendar | Search

NMNH Home › Research & Collections › Botany ›

Flora of the West Indies

Catalogue of the Seed Plants of the West Indies

Return to Map of Antilles
Return to Detailed Query

Results of Query (endemic genera in red)
6714 Records

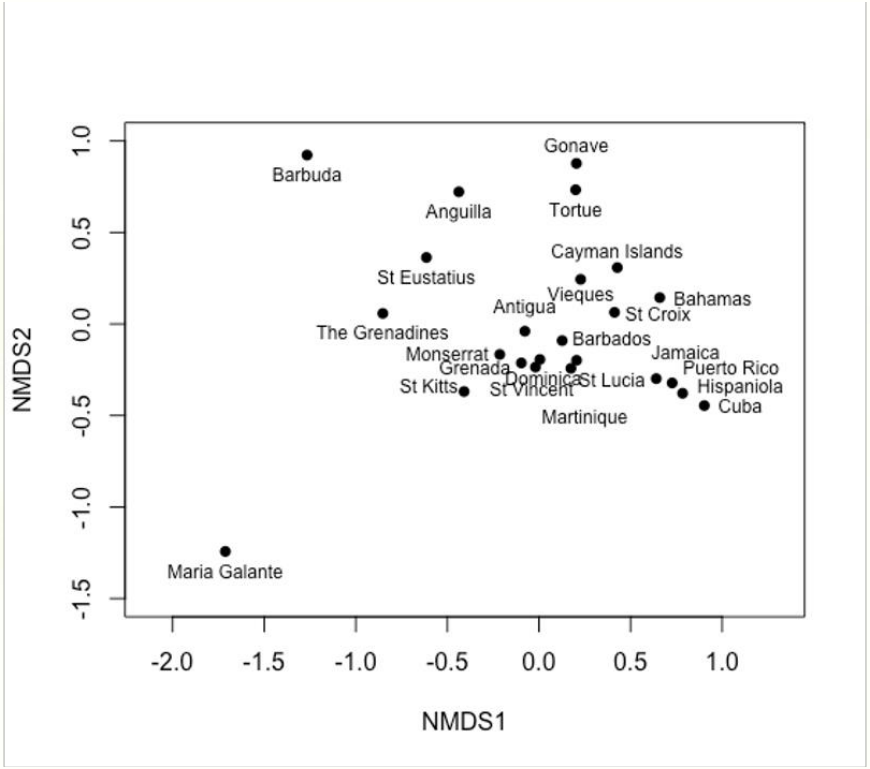
Plants of Cuba

Click on species for full record display. indicates image(s) for the species.

Acanthaceae *(contributed by P. Acevedo-Rodriguez & M.T. Strong)*

Ancistranthus harpochiloides (Griseb.) Lindau
Andrographis paniculata (Burm. f.) Wall. ex Nees
Apassalus cubensis (Urb.) Kobuski
Apassalus humistratus (Michx.) Kobuski
Apassalus parvulus Alain & Leonard
Aphelandra sinclairiana Nees ex Benth.
Aphelandra tetragona (Vahl) Nees
Asystasia gangetica (L.) T. Anderson
Avicennia germinans (L.) L.
Barleria cristata L.
Barleria lancifolia T. Anderson
Barleriola saturejoides (Griseb.) M. Gómez subsp. *acunae* Borhidi & O. Muñiz
Barleriola saturejoides (Griseb.) M. Gómez subsp. *hirsuta* Borhidi & O. Muñiz
Barleriola saturejoides (Griseb.) M. Gómez subsp. *saturejoides*
Barleriola solanifolia (L.) Oerst. ex Lindau
Bravaisia berlandieriana (Nees) T.F. Daniel
Crossandra infundibuliformis (L.) Nees
Dasytropis fragilis Urb.
Dicliptera sexangularis (L.) Juss.
Dyschoriste bayatensis (Urb.) Urb.
Elytraria bissei H. Dietr.
Elytraria cubana Alain

The result

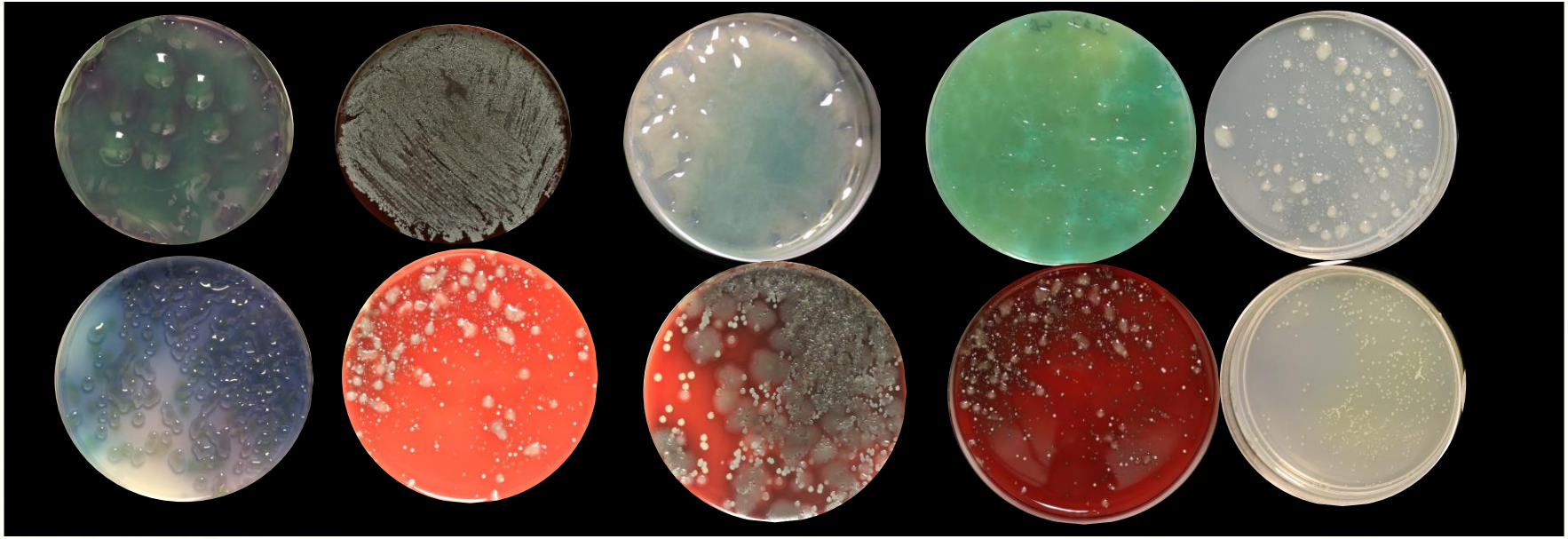

Barbuda
Gonave
Anguilla
Tortue
Cayman Islands
St Eustatius
Vieques
Bahamas
Antigua
St Croix
The Grenadines
Barbados
Monserrat
Jamaica
Grenada
Puerto Rico
Dominica
St Lucia
St Kitts
St Vincent
Hispaniola
Cuba
Martinique
Maria Galante
NMDS1
NMDS2
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

Why shoud you believe me?

- BA in Biology & CS, PhD, Postdoc
- research
  - microbiomes
  - computational biology
  - methods in big data/ML

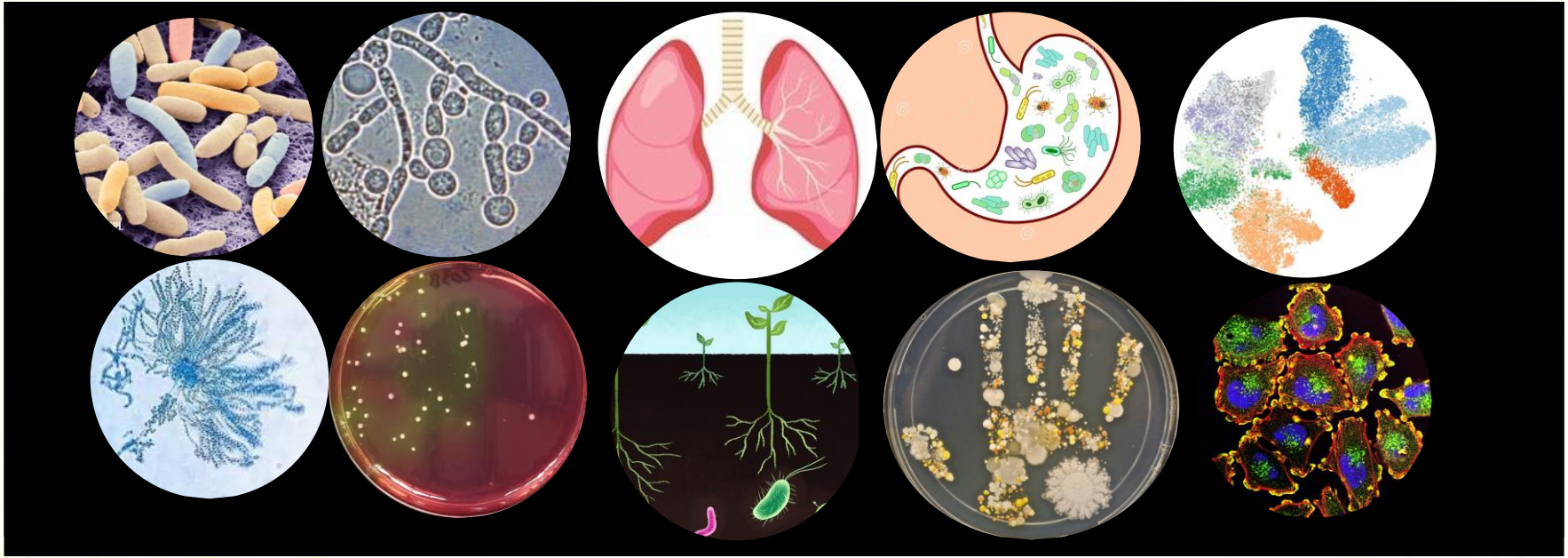
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



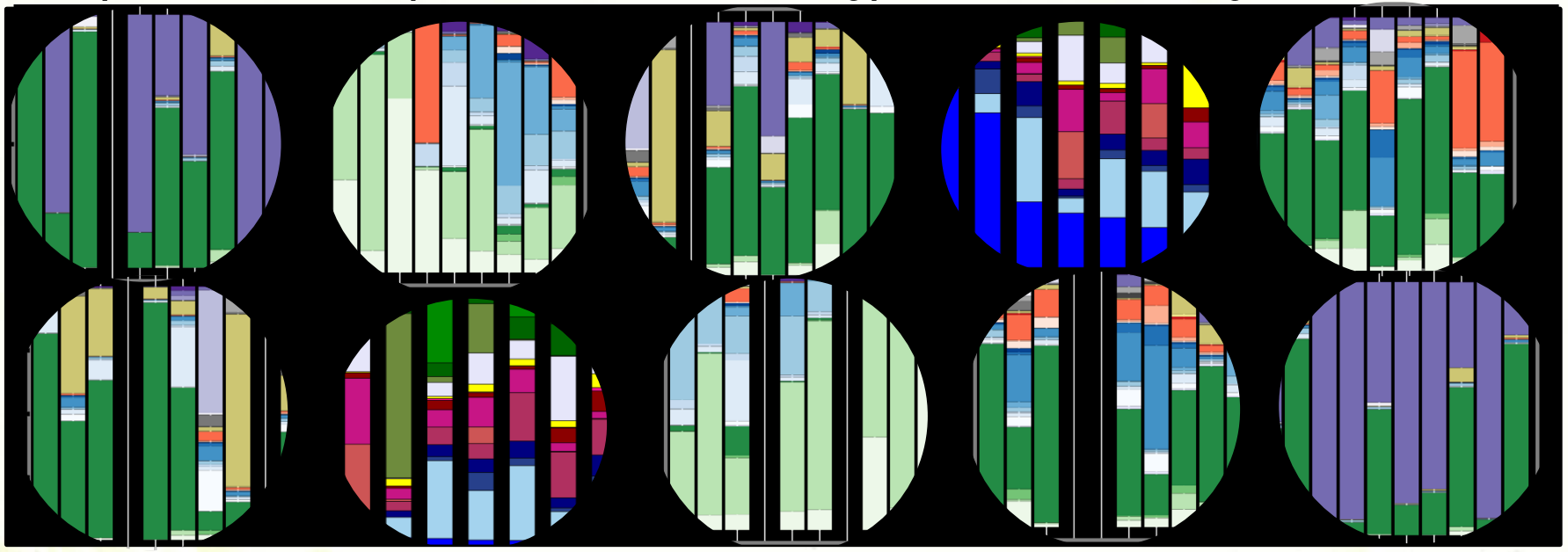
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



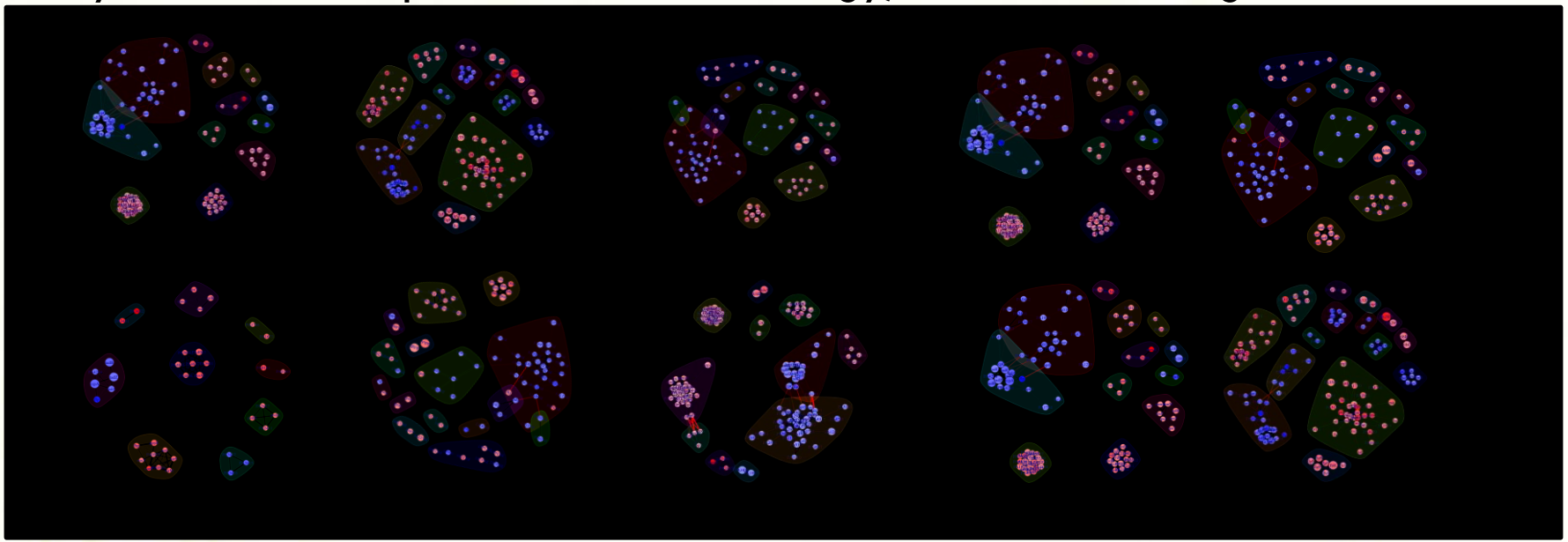
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



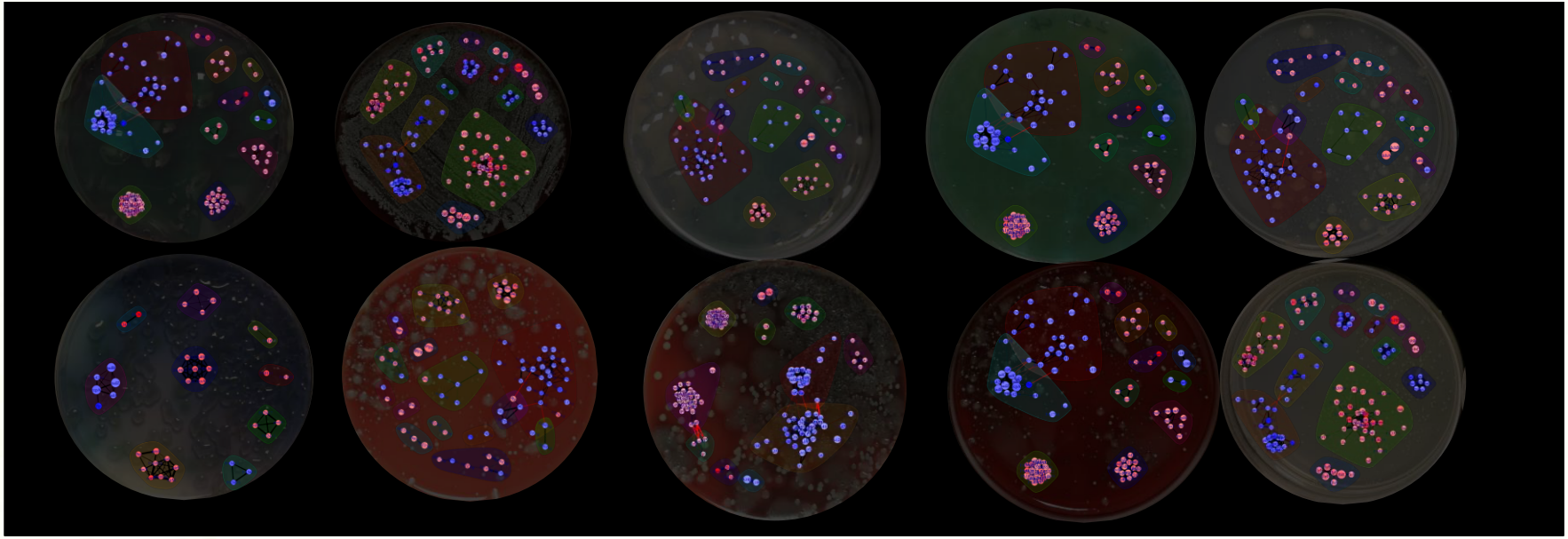
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



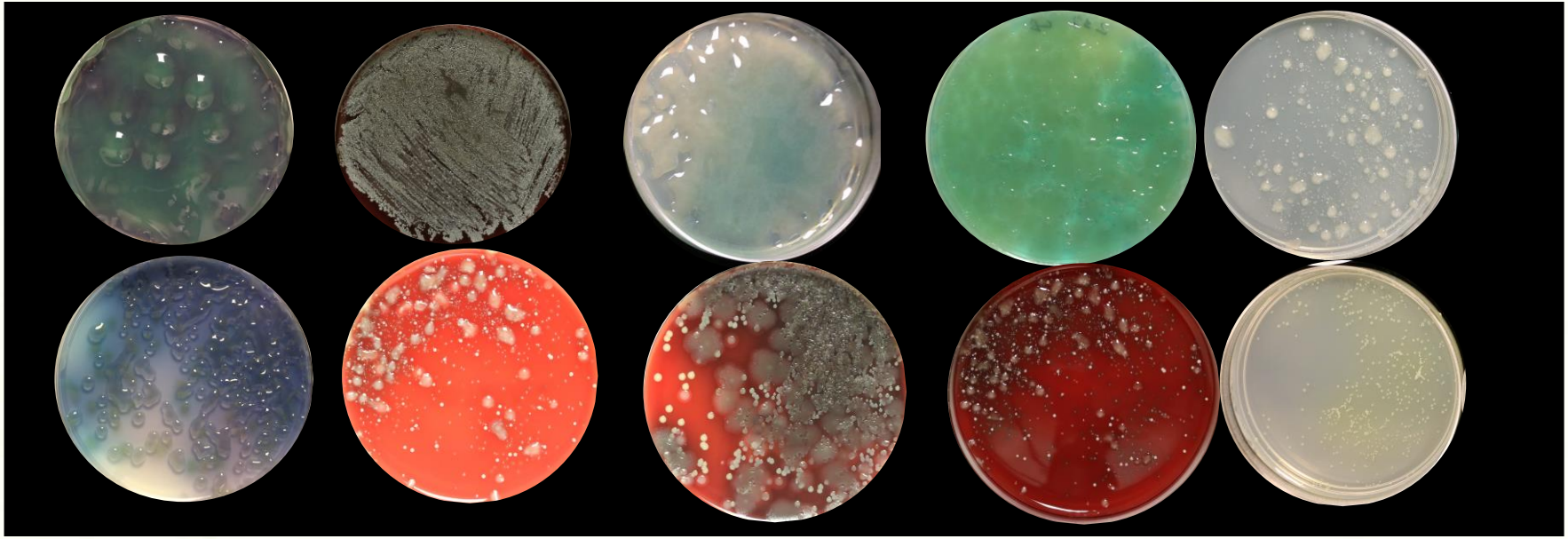
Me: Georgia Doing

My interests: computational microbiology, small cells & big data

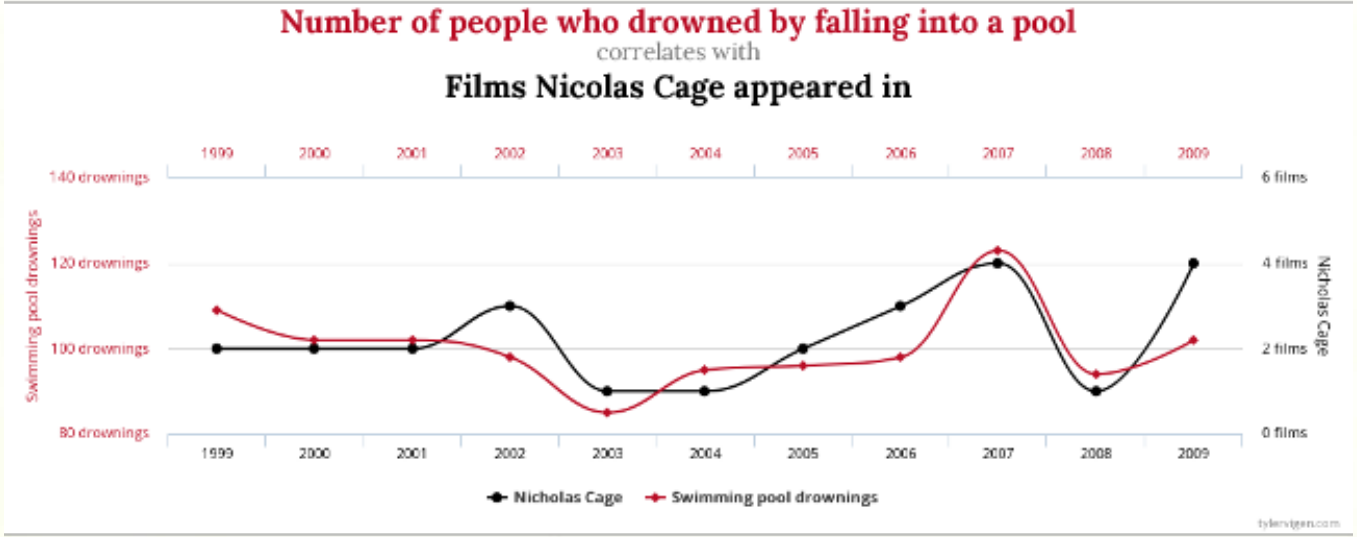


Me: Georgia Doing

My interests: computational microbiology, small cells & big data



Correlation can be…



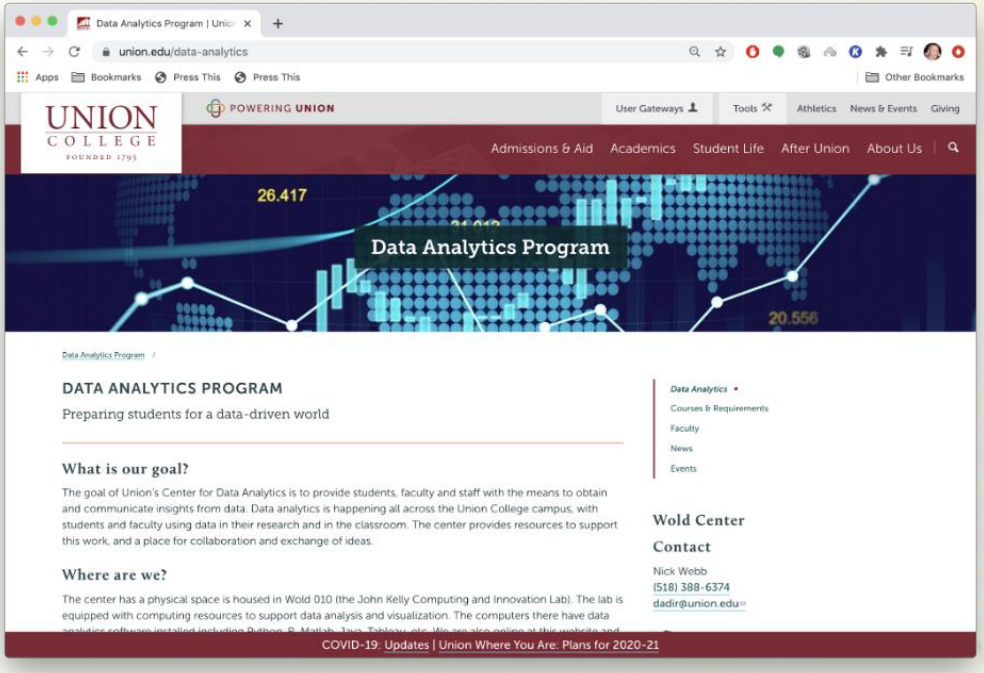
https://www.tylervigen.com/spurious-correlations

Do we have courses that teach that?

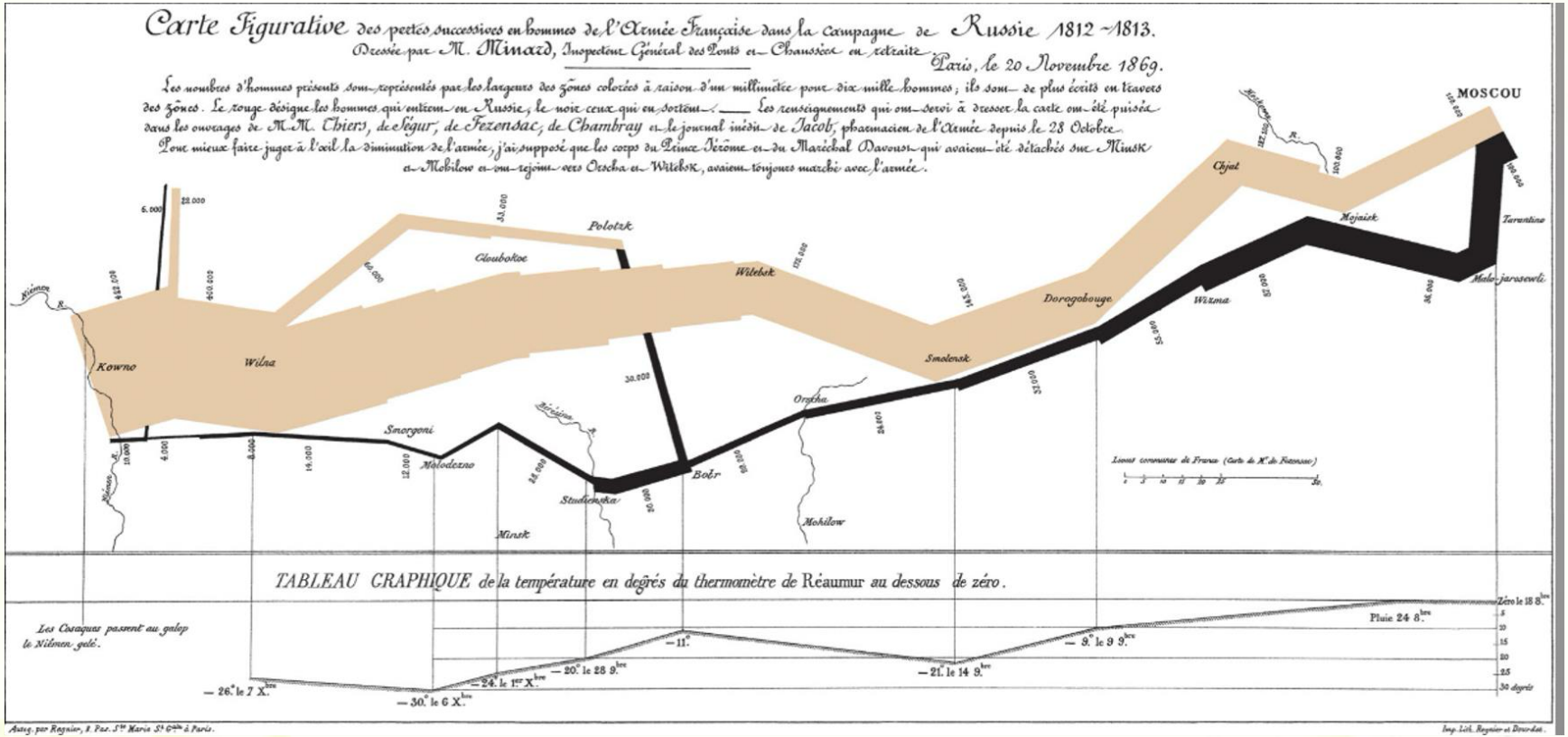
- Analyze Data
  - CSC 233 Data Analytics
  - ENS 215 Exploring Environmental Data
  - ECO 364 Business Analytics
- Visualize Data
  - CSC 234 Data Visualization
  - ECO 134 Data Visualization
- https://www.union.edu/academic/majors-minors/data-analytics

Center for Data Analytics

- Physical & Virtual Space
  - WOLD 010
  - https://www.union.edu/data-analytics
  - analyticscenter@union.edu
- Workshops
- Seminars
- Tutorials
- Helpdesk



Data Visualization



Lab machines: getting started

- Log on:
  - Your Union usernames
  - Default password: the word union followed by your ID number; e.g., union12345

Always log off before leaving the lab