

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:

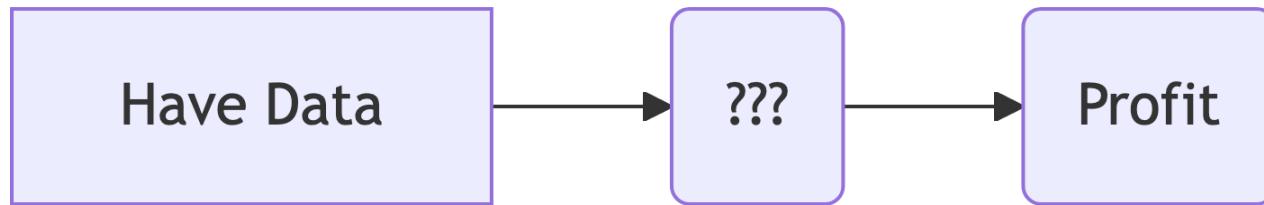
- Name and pronouns
- Help me find you on my roster
- What perspective do you bring to this class? Could be from a life experience, memory, etc. or your major, minor, research, internship, etc.
- Is there a dataset you've wrestled with in the past that you wished you could have analyzed more deeply?

Fermi questions

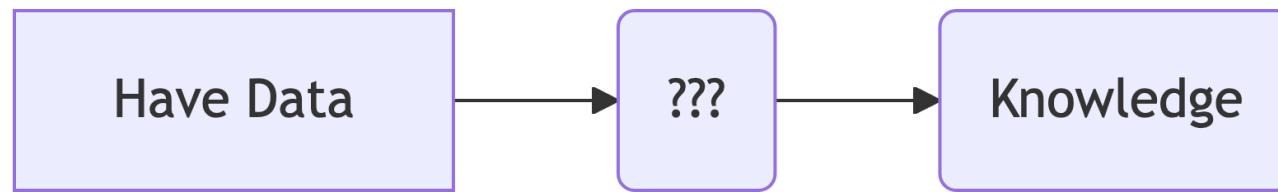
<https://fermi-questions.andrechek.com/>

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

What is ‘Data Analytics’?



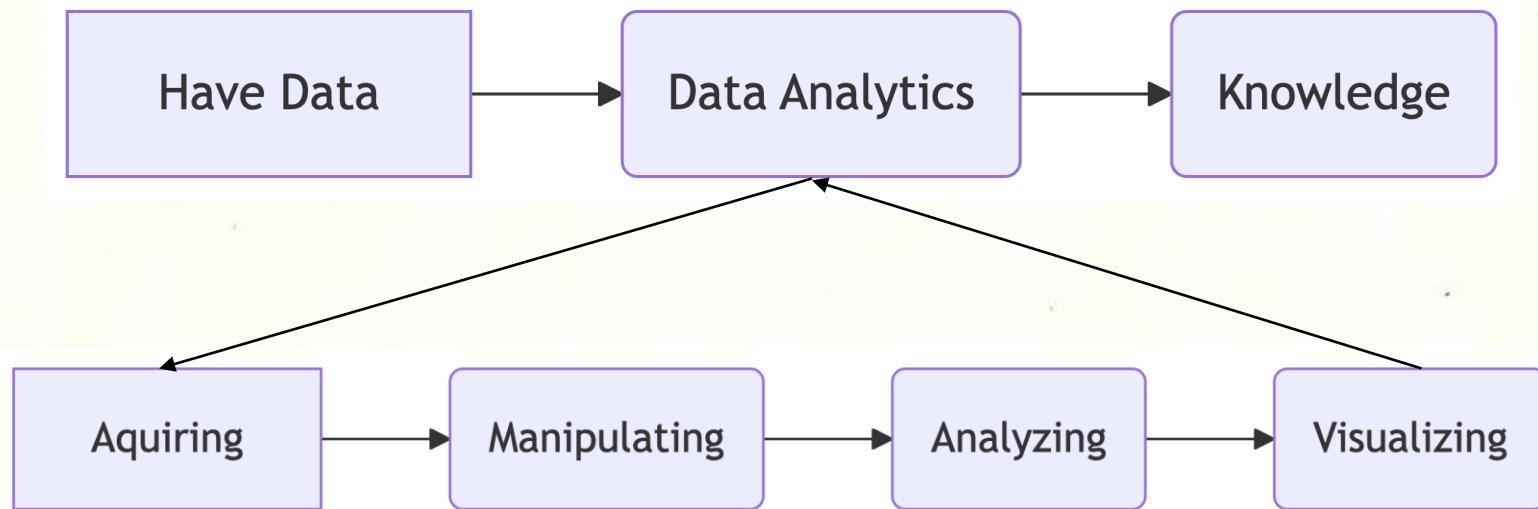
What is ‘Data Analytics’?



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

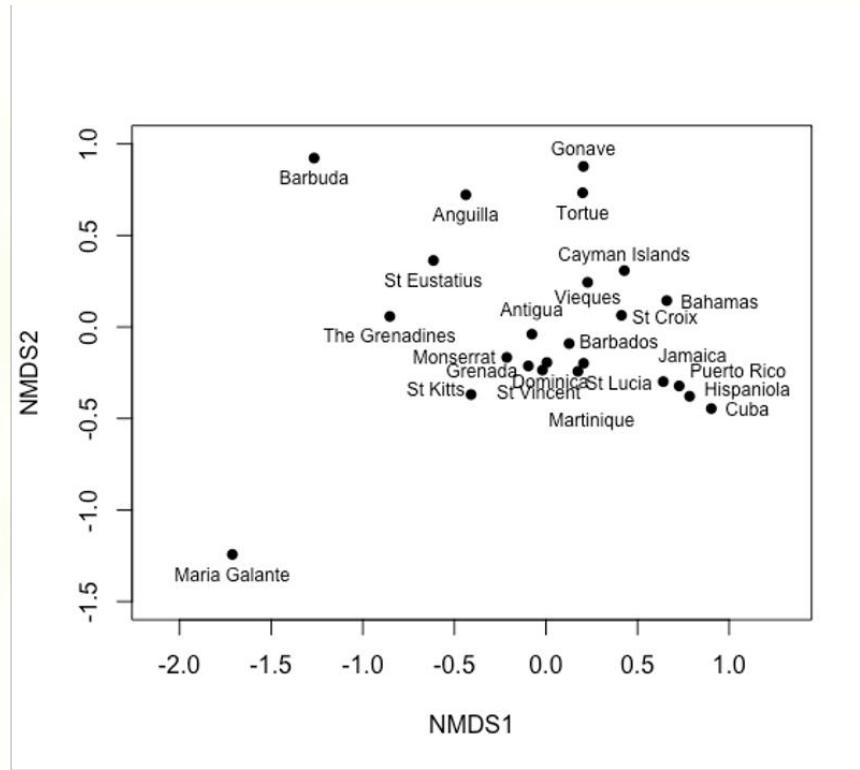
Plants of Cuba

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

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

[*Ancistranthus harpochilooides* \(Griseb.\) 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 berlandieriiana* \(Nees\) 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

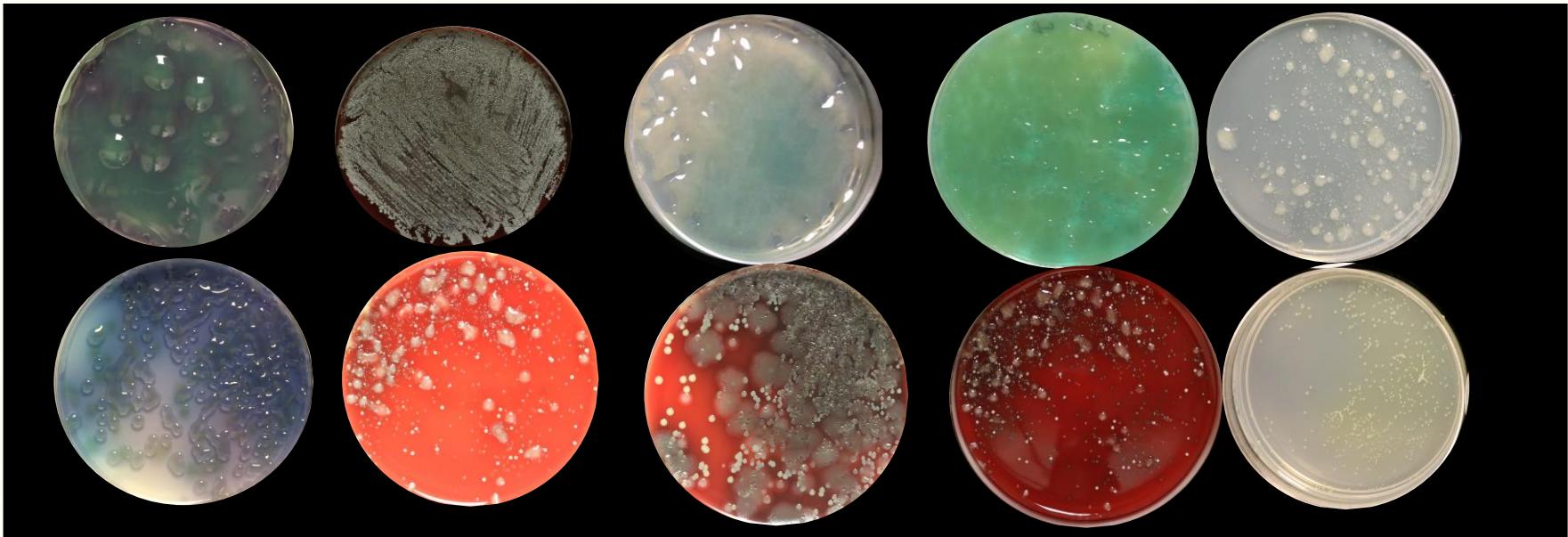


Why shoud you believe me?

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

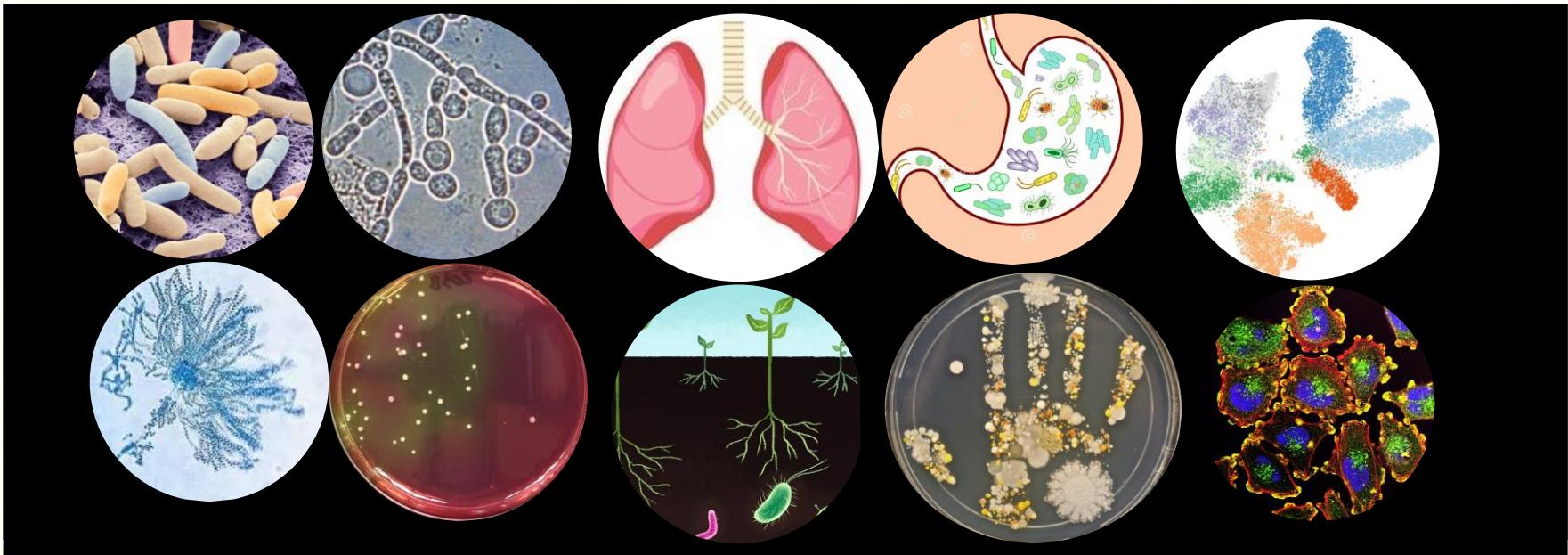
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



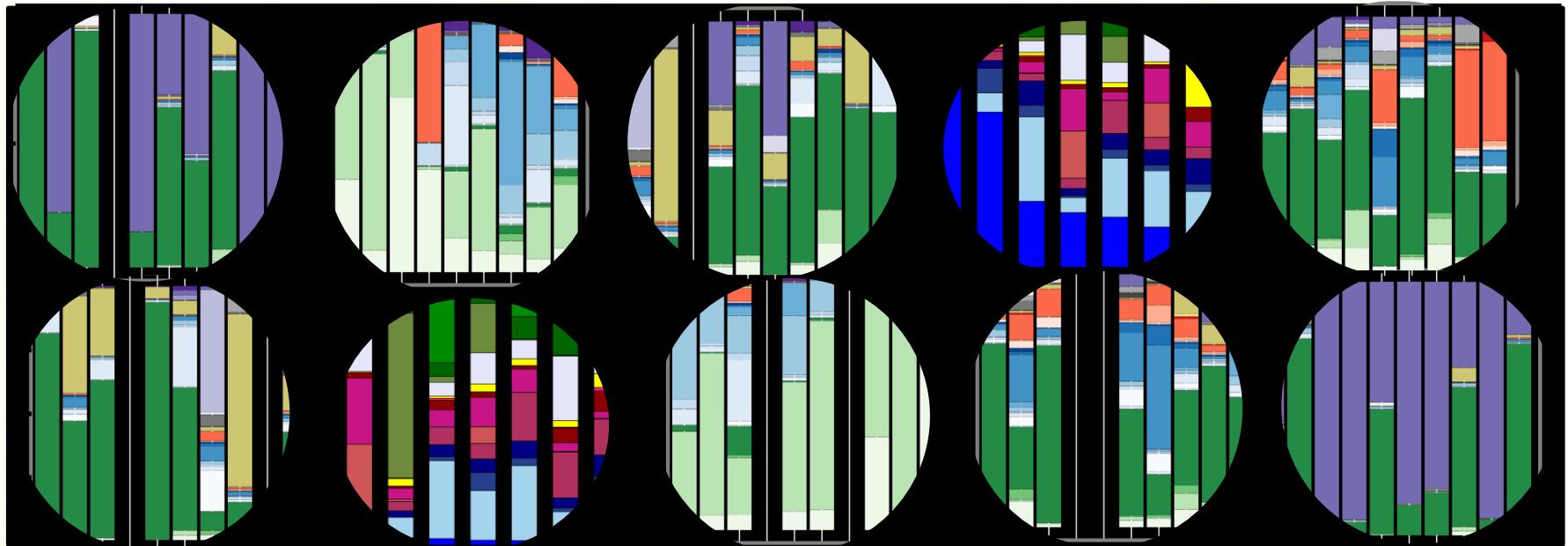
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



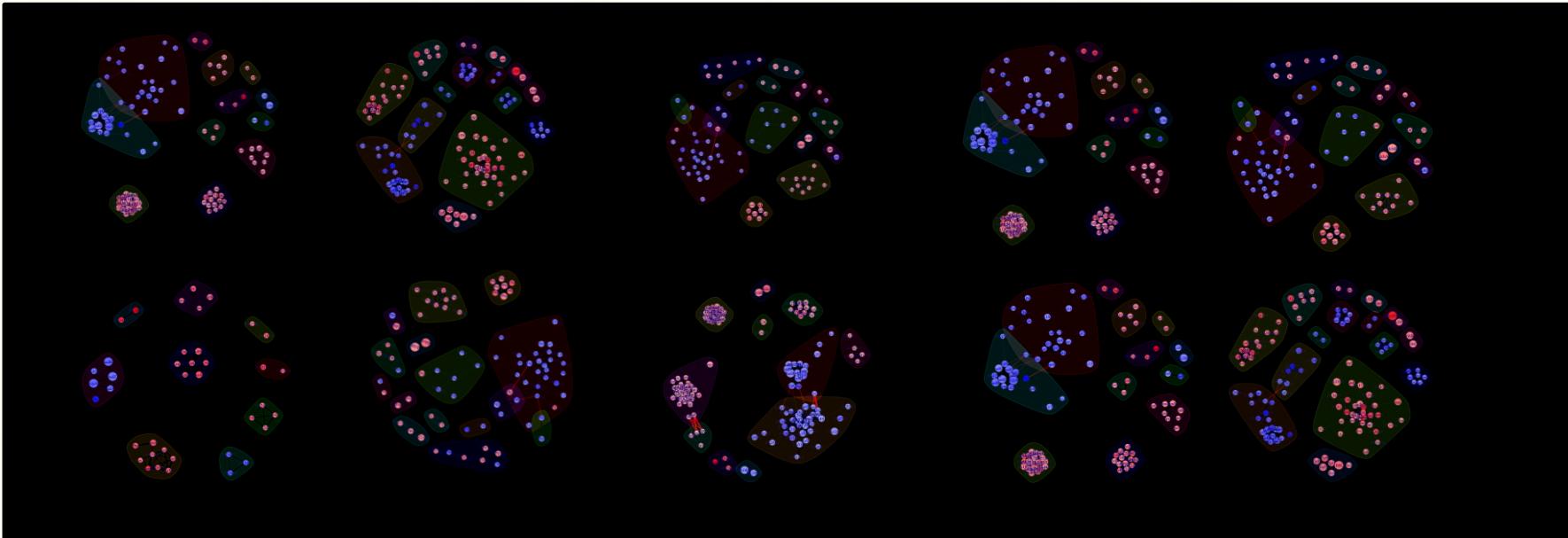
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



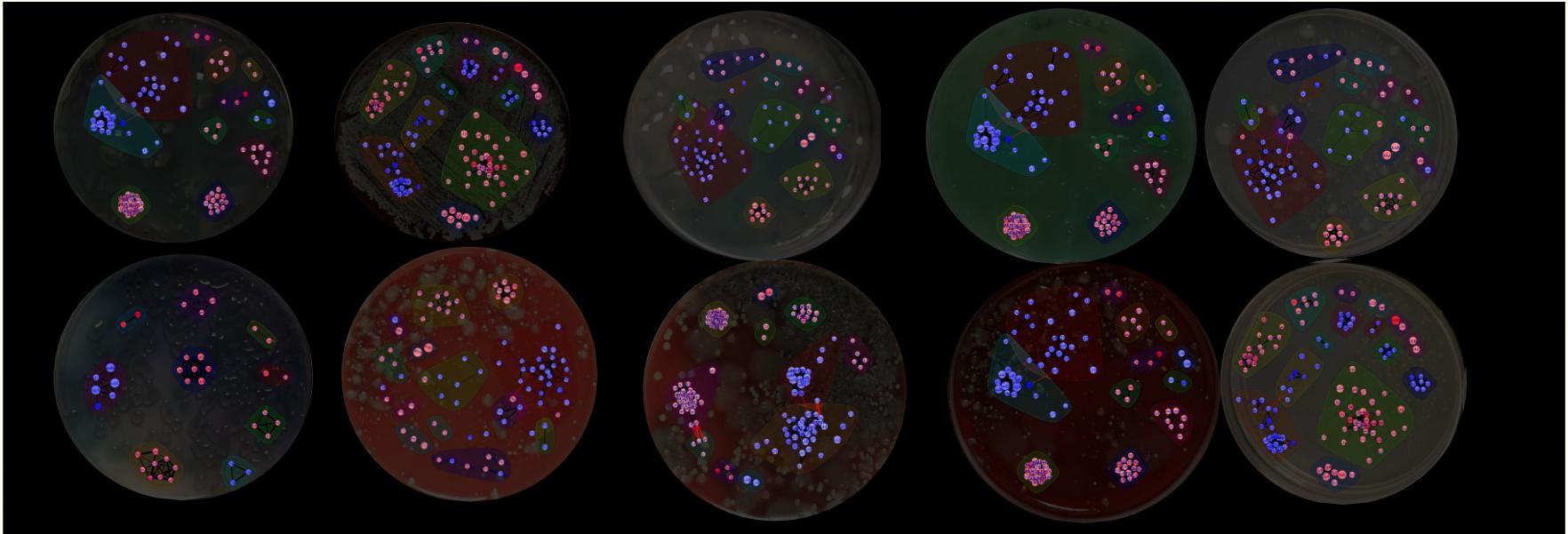
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



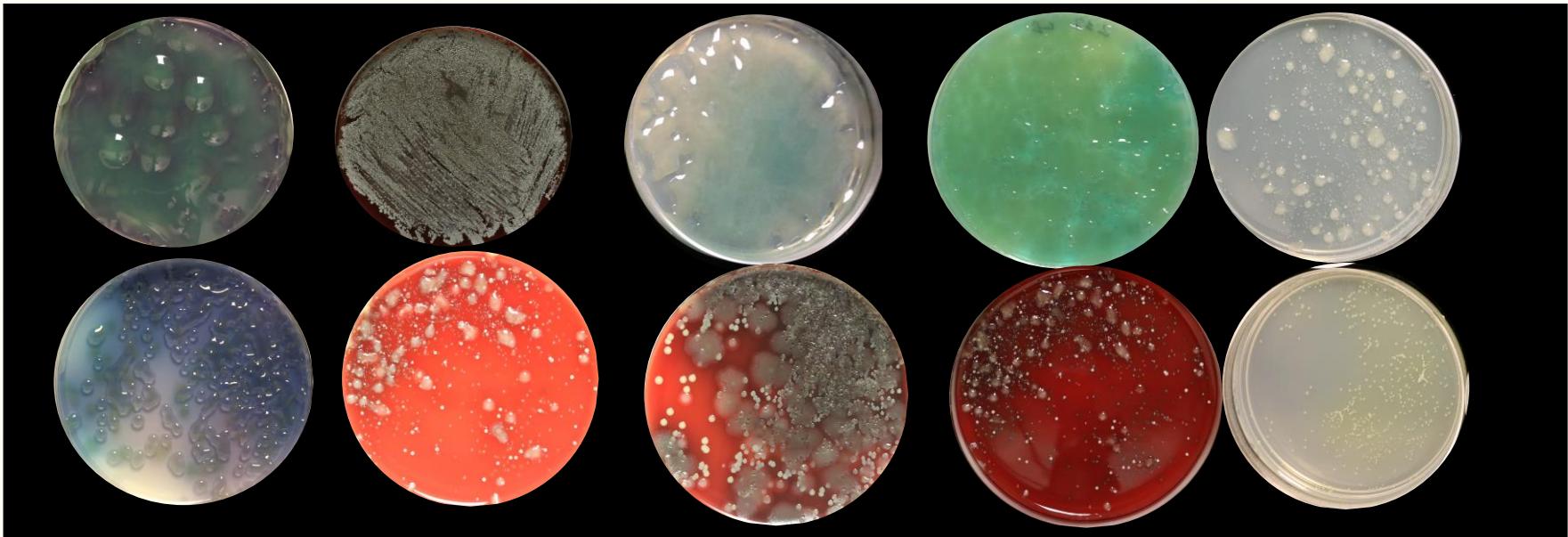
Me: Georgia Doing

My interests: computational microbiology, small cells & big data

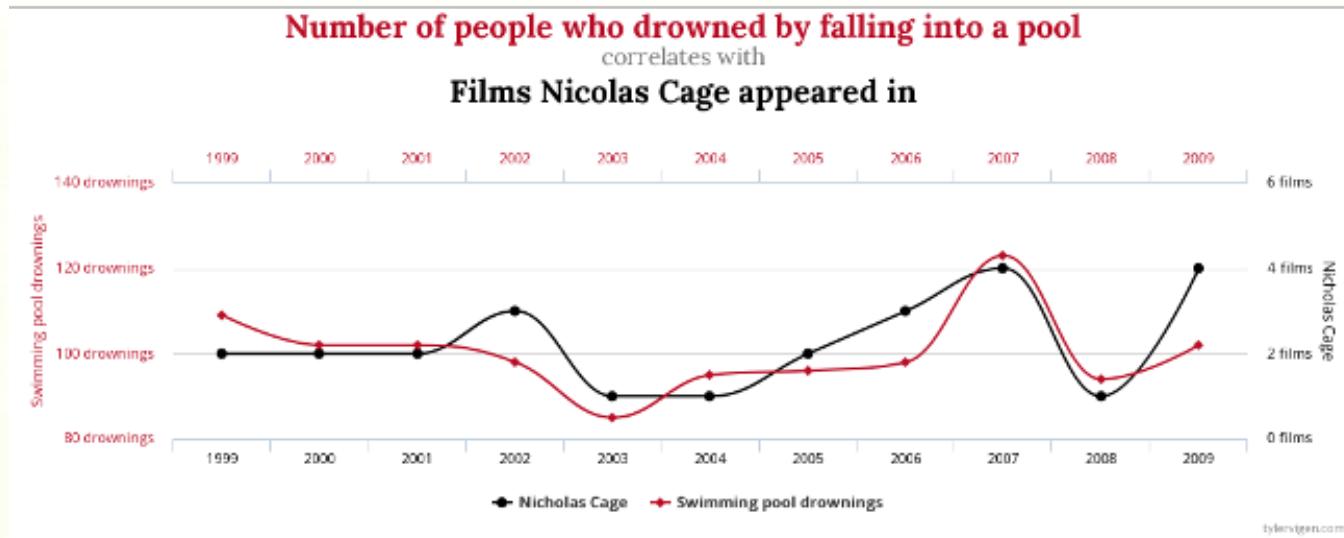


Me: Georgia Doing

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

The screenshot shows a web browser displaying the Union College Data Analytics Program website. The header features the Union College logo and navigation links for User Gateways, Tools, Athletics, News & Events, and Giving. The main content area has a dark background with a blue and white data visualization graphic showing a line graph and bar charts. The text "Data Analytics Program" is overlaid on the graphic. Below the graphic, the page title is "DATA ANALYTICS PROGRAM" with the subtitle "Preparing students for a data-driven world". A section titled "What is our goal?" explains the center's purpose in providing tools and resources for data analysis across the campus. Another section, "Where are we?", describes the physical location in Wold 010 and the software tools available. The footer includes COVID-19 updates and contact information for Nick Webb.

Data Analytics Program | Union

union.edu/data-analytics

UNION COLLEGE FOUNDED 1795

POWERING UNION

Admissions & Aid Academics Student Life After Union About Us

26.417

26.017

20.556

Data Analytics Program

Data Analytics Program /

DATA ANALYTICS PROGRAM

Preparing students for a data-driven world

What is our goal?

The goal of Union's Center for Data Analytics is to provide students, faculty and staff with the means to obtain and communicate insights from data. Data analytics is happening all across the Union College campus, with students and faculty using data in their research and in the classroom. The center provides resources to support this work, and a place for collaboration and exchange of ideas.

Where are we?

The center has a physical space is housed in Wold 010 (the John Kelly Computing and Innovation Lab). The lab is equipped with computing resources to support data analysis and visualization. The computers there have data analysis software installed including Python, R, Matlab, Java, Tableau, etc. We are also online at this website and dadir@union.edu.

COVID-19: Updates | Union Where You Are: Plans for 2020-21

Data Analytics Courses & Requirements Faculty News Events

Wold Center Contact Nick Webb (518) 388-6374 dadir@union.edu

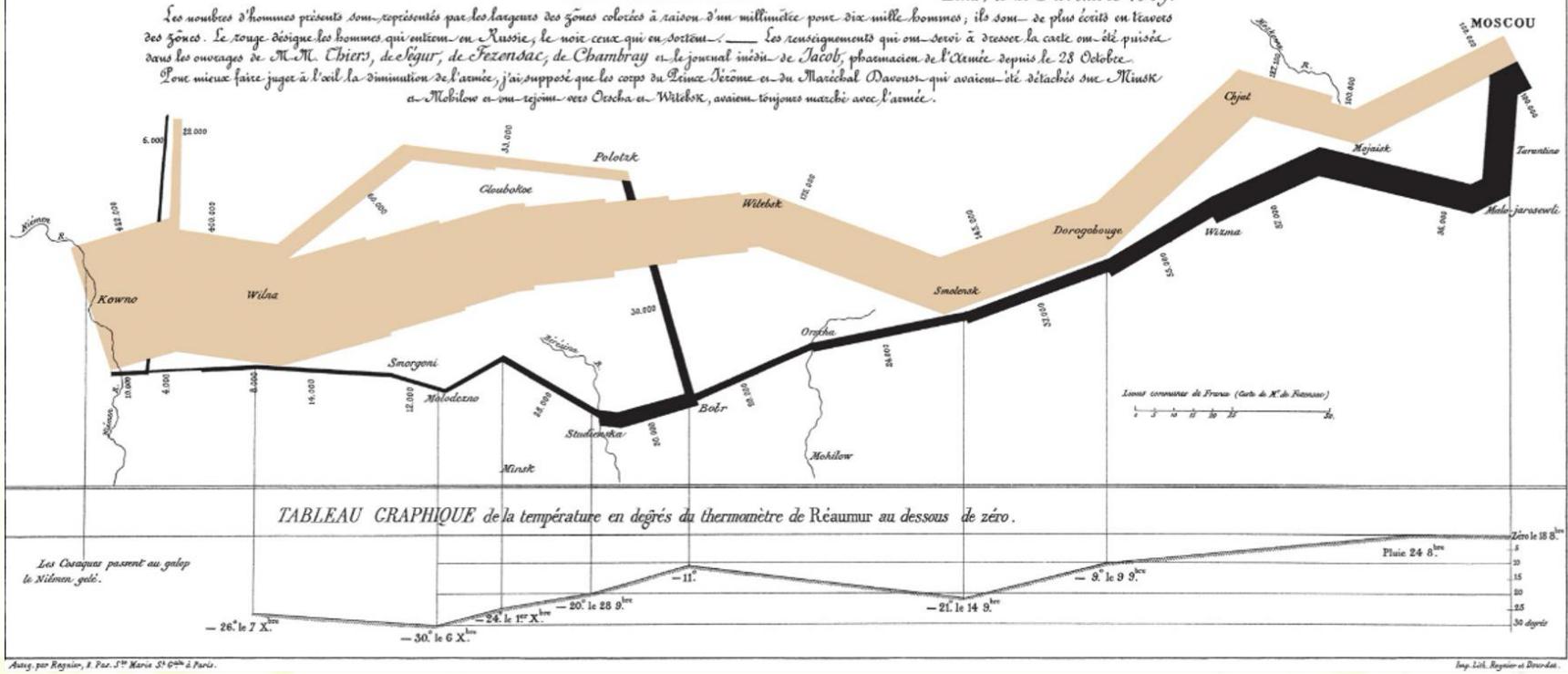
Data Visualization

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812 ~ 1813.
Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes perdus sont représentés par les largeurs des zones colorées à raison d'un millimètre pour six-mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dessiner la carte ont été prisée dans les ouvrages de M. M. Chiers, de Segur, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent les Osses et Wilobek, avaient toujours marché avec l'armée.



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