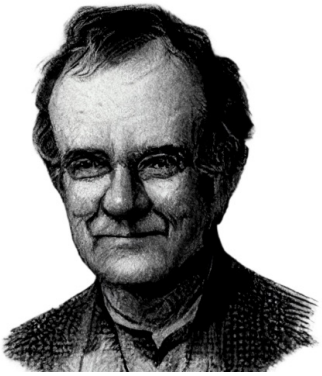
DOINGG@UNION.EDU

COURSE: CSC233: INTRO TO DATA ANALYTICS DOCUMENTATION

SELF

Contents

Home



The course website for CSC233, part of the Union College CS curriculum.
Here you can find our weekly schedule, assignmnets and resources.

Announcements

Upcoming due dates

Next assignment (HW 01) is due Tuesday Jan 20th **before** class, at 10 am..

Office Hours

- Student/Office Hours:
  - Steinmetz 108B

- Wednesday 2:00 pm - 3:30 pm
- Thursday 4:00 pm - 5:30 pm
- subject to change, check course website for most up-tp-date schedule
- drop-in or schedule a 15 minute slot: https://calendar.app.google/8bus6pfDvyphR9ar5
- by appointment for another time or over zoom

Course Description

Data analytics, the process of analyzing, revealing, interpreting and visualizing information concealed inside big data is revolutionizing daily life, as used by companies such as Amazon, Google and Facebook, for the diagnosis of medical conditions or the way medical claims are handled, for investment strategies and real estate pricing, and in academia, with the analysis of historical texts, understanding the deliberations of the Supreme Court or the European Commission, or processing large amounts of genomics data.

In this class, students will be introduced to techniques to acquire data from the web, manipulate and pre-process data into manageable forms, perform analyses from a descriptive and predictive standpoint, and learn the basics of visualization of the result, all with a focus on storytelling through data, enhancing data literacy.

In this course you will use the Python programming language to scrape web data, prepare data for analysis, analyze to produce explainable results, and visualize those results. Throughout we will discuss the pitfalls, ethics and challenges of working with data. A 1-page schedule-at-a-glance is included at the end of the syllabus.

This course requires the pre-approval of the department. For more information, please visit: https://union.edu/advising-registration/pre-approval.

Schedule

Weekly Schedule

Schedule

W.	TOPIC	Due	Runestone
1	Introduction to Data Analytics	Introductory Survey	httlads ch. 1 & 4
2	Using Descriptive Statistics to Talk About Data	HW 0	
3	Python Tools for Data Analytics	HW 1	
4	PANDAS for Exploratory Data Analysis	HW 2, project data	httlads ch. 2 & 7
5	Collecting Data from the Web	HW 3, project question	httlads ch. 5
6	Data Wrangling and Cleaning	HW 4	
7	Merging, selecting and transforming data	HW 5, project outline	httlads ch. 6
8	Simple Machine Learning Methods to Explore Data	HW 6, Midterm Exam	httlads ch. 9
9	Evaluating performance	Project Draft	
10	Project Presentation	Final Project & Presentation	

Resources

Runestone (Online Textbook)

- https://runestone.academy/runestone/default/user/register
- your Union College email
- unioncollege_httlads_winter26 as the course name
- for review, you may also use the Python course: unioncollege_py4e-int_winter26_da

Google Colab Pro

- Pro account sign-up: https://colab.research.google.com/signup
- Use your Union email
- A pro account is not needed for this course, just an option Union pays for
- Notes template: https://colab.research.google.com/drive/15rPmKyQKCCbwsdDCpC39o8hczWZ-B5zV?usp=sharing

Late Token Form

- These are for project-related deadlines, not Rosalind problems
- These can be used for any reason
- These do not apply to your final deadline
- Form link: https://forms.gle/6W3LrkK4SdhWGSzL7

Python Style Guide

- PEP 8
- should be used in all code cells of Colab notebooks
- Link: https://peps.python.org/pep-0008/

Links to Google Drive folders

- uploading for peer review: https://drive.google.com/drive/folders/1W-FplVDW2b9qG9BDEaPMNEMWRsFrJabE?usp=drive_link

Syllabus

Syllabus

CSC233 - Introduction to Data Analytics Syllabus

1 CSC 233: Introduction to Data Analytics

- Union College, Winter 2026
- Georgia Doing, PhD
- email: doingg@union.edu,
- office: Steinmetz 108B

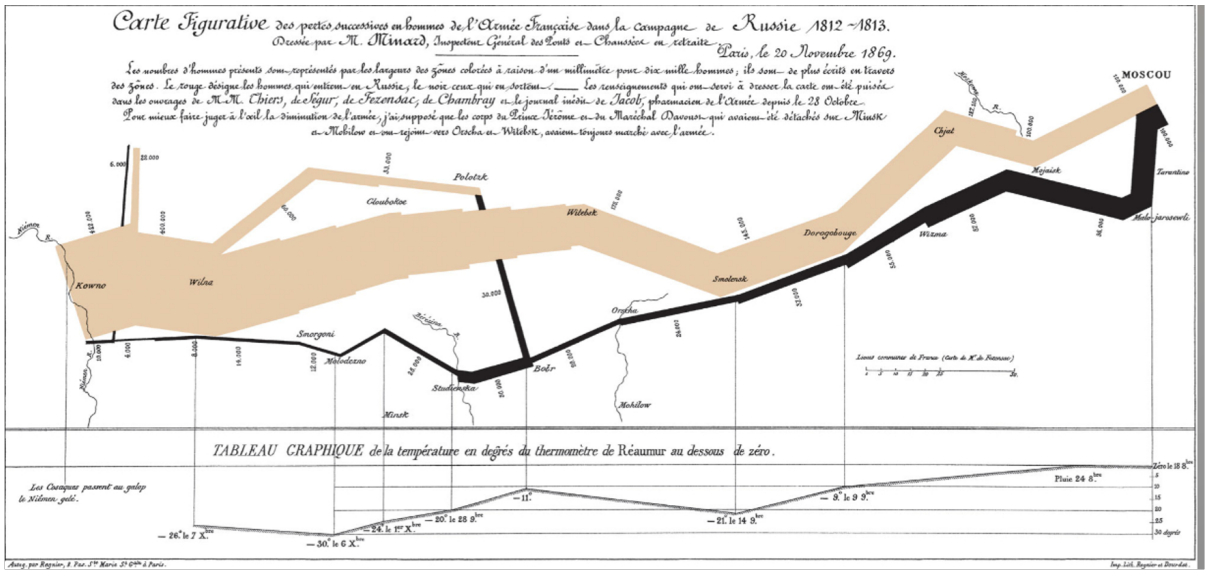


Figure 1: Map of Napolean's Russian campaign on 1812, drawn by Charles Joseph Minard, 1869. The thick band shows the size of the army at each position. The path of Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales.

2 COURSE BASICS

- Lecture: Tuesday/Thursday 10:55 am - 12:40 pm
- Location: ISEC 070
- Student/Office Hours:
  - Wednesday 2:00 pm - 3:30 pm
  - Thursday 4:00 pm - 5:30 pm
  - subject to change, check course website for most up-to-date schedule
  - drop-in or schedule a 15 minute slot: https://calendar.app.google/8bus6pfDvyphR9ar5
  - by appointment for another time or over zoom

3 OVERVIEW

Data analytics, the process of analyzing, revealing, interpreting and visualizing information concealed inside big data is revolutionizing daily life, as used by companies such as Amazon, Google and Facebook, for the diagnosis of medical conditions or the way medical claims are handled, for investment strategies and real estate pricing, and in academia, with the analysis of historical texts, understanding the deliberations of the Supreme Court or the European Commission, or processing large amounts of genomics data.

In this class, students will be introduced to techniques to acquire data from the web, manipulate and pre-process data into manageable forms, perform analyses from a descriptive and predictive standpoint, and learn the basics of visualization of the result, all with a focus on storytelling through data, enhancing data literacy.

In this course you will use the Python programming language to scrape web data, prepare data for analysis, analyze to produce explainable results, and visualize those results. Throughout we will discuss the pitfalls, ethics and challenges of working with data. A 1-page schedule-at-a-glance is included at the end of the syllabus.

This course requires the pre-approval of the department. For more information, please visit: https://union.edu/advising-registration/pre-approval.

4 LEARNING OBJECTIVES

By the end of the course, you should be able to *answer* the following questions:

- What are the fundamental steps of a data analytics workflow?
- What are the quality control checks involved in the responsible use of data?

- What elements facilitate storytelling with data?

By the end of the course, you should be able to *do* the following:

- Aquire and prepare data for analysis.
- Conduct descriptive data exploration.
- Communicate findings of an independent project.

5 CLASS POLICIES AND GUIDELINES

The Computer Science Department as a whole welcomes all people, regardless of age, background, beliefs, ethnicity, gender identity, gender expression, national origin, religious affiliation, sexual orientation and any other differences, be they visible or non-visible. It's also important to recognize that institutional racism has prevented members of marginalized groups - especially black, indigenous and other people of color (BIPOC) - from fully participating in the field of Computer Science.

You. Yes you, the one reading this syllabus. You belong here.

As an instructor, I will do my utmost to uphold these principles and to treat everyone with respect. As students in this class, I expect you to do the same since one person is not a community unto themselves. We're all in this together, so let's treat each other well. Lastly, I welcome feedback on issues we might discuss or ways that our class can be a more just one for all people.

6 COURSE MATERIALS

Online Textbook: Runestone (free).

We are going to use an online, interactive, free and open source textbook. This textbook has a lot of embedded interactive exercises. Reading this book always also means doing the activities. To get access to the book, please register on Runestone using the following informationm which can also be found on Nexus:

- https://runestone.academy/runestone/default/user/register
- your Union College email
- unioncollege_py4e-int_fall25 as the course name

We'll use the following websites and software throughout the course:

- The course website: https://georgiadoing.github.io/tufte-quarto/

- Gradescope: used to submit programming assignments
- Google Colab: the could-based platform to write and run Jupyter Notebooks with python code

7 COURSE RESOURCES

The course website has a list and links to the readings and software resources that we are going to use. In class, we are going to use the Linux machines in ISEC 070. Outside of class you have a choice of hardware. You can access Google Colab via a browser on your own machine, install the software on your own computer (Python is freely available for Windows, Mac, and Linux) or you can work in one of the Computer Science labs. We have three spaces that you can use 24/7 using your ID card, except when classes are being held in them:

- Olin 107
- PASTA Lab (ISEC 051+)
- Computer Science Resource room (Steinmetz Hall, 209A)

8 LIBRARY RESOURCES

The library is available to help you with your research needs! Librarians can help you develop research questions, search for and select the best sources for your projects, identify research strategies, evaluate sources, and assist you with creating citations. There are multiple ways for you to contact a librarian. For more information, please see our Ask A Librarian page: https://www.union.edu/schaffer-library/ask-a-librarian.

9 ACCOMODATIONS

It is the policy of Union College to make reasonable accommodations for qualified individuals with disabilities. If you are a person with a disability and wish to request accommodations to complete your course requirements, please make an appointment with me or stop by during my office hours as soon as possible, all discussions will remain confidential. You must provide reasonable notice and be in touch with Accomodative Services on the 2nd floor of Schaffer Library, if you have not already, if you wish to take advantage of extra time on exams.

10 GRADING

Each week I will give you some *problem exercises*. You will complete these and submit them online. There will be a *final project*, involving the acquisition, manipulation and analysis of data. Finally, class attendance, participation and engagement are critical components of the course.

- Problem sets: 50%
- Midterm exam: 25%
- Final project: 25%

Letter Grade Scale:

- A: 93-100; A-: 90-92
- B+: 87-89; B: 83-86; B-: 80-82
- C+: 77-79; C: 73-76; C-: 70-72
- D: 60-69; F: 0-59

11 Basic Course Requirement

In order to pass the class you must earn a passing grade. In addition, you must meet the following basic requirement. You must achieve a passing grade (e.g. 60% or higher) overall in both the problem sets and the final project. In other words, you cannot blow off the final project and pass the class! Note that this basic requirement is necessary but not sufficient to pass the class.

12 LATE POLICIES

Homework assignments will usually be discussed in class on the day that they are due and thus, without prior agreement, there is NO late submission of assignments allowed. You have three "extension tokens" that can be used for due dates related to your final project for this class. Each token will extend a single project deadline by 24 hours. You can use each token on a different deadline, or use two or even all three on a single deadline. Once all three tokens have been used, no late submissions will be accepted (barring exceptional circumstances).

https://forms.gle/6W3LrkK4SdhWGSzL7

13 ATTENDANCE

Class participation is a critical component of the course and attendance is mandatory. I realize that sometimes other things come up (interviews, athletic events, illness, etc.). In those cases, just let me know (in advance, if at all possible) that you are going to be absent. You may not receive credit for make-up material if you did not discuss your absence with me prior to class. No matter how much class you have missed, please do not come to class if you have an illness that is likley contageous, please email me to work out a reasonable solution.

If you do miss class, it is your responsibility to make up any material that you missed. Get notes from a classmate, make sure to complete all homework assignments due or assigned during the class that you missed, and come see me if you have any questions on the material. Unless you have made an arrangement with me ahead of time or you had an emergency, I will expect you to hand in all assignments by the due date, regardless of whether you were in class. You will not be able to make up missed exams or in-class questions and assignments.

Attendence in class constitutes being present, respectful and engaged. Accessing off-topic websites or software, checking email or phones, *utilizing large language models* or otherwise attending to things not pertinant to the course is distracting to yourself and others and will result in a reduction of your overall grade, up to a 0.5% reduction per class.

14 PARTICIPATION AND ENGAGEMENT

Following these guidelines will constitute postiive engagement:

- Respect. This course is a space for rigorous and respectful debate. When confronting ideas and people different from you, lead with curiosity rather than judgement. This commitment extends to all our face to face and digital interactions. We will critique ideas, not people. To protect our shared learning environment, you may not record, photograph, or share any part of our class sessions outside of our class.
- Show up to class on time and prepared. Show up ready to learn by completing the readings and exercises. When you are late or unprepared, you are disrespecting the learning experience for the group.
- Embrace intellectual humility. Recognize that there are limitations to your knowledge and that some of your beliefs could actually be wrong. Be curious about your thinking and open to learning from others. This is hard, but so vital to your learning in this course—we'll work on developing this throughout the term!

- Get help when you need it. If you are stuck or confused or lost, be proactive and get help. The best learners and thinkers can figure out when they are stuck and make a plan to get unstuck. Come to Office Hours (Wed/Thurs 2:00-3:30pm Steinmetz 108B), the CS Helpdesk (Sun-Thurs 7:00-9:00pm Olin 107) or ask another student. Often the best person to explain something is the person who just figured it out. Try reaching out to another student in the course who seems like they get it—they will probably be flattered you asked!

15 ACADEMIC INTEGRITY AND "ARTIFICAL INTELLIGENCE" USE

Union College recognizes the need to create an environment of mutual trust as part of its educational mission. Responsible participation in an academic community requires respect for and acknowledgement of the thoughts and work of others, whether expressed in the present or in some distant time and place. Matriculation at the College is taken to signify implicit agreement with the Academic Honor Code, available at: http://muse.union.edu/honorcode

It is each student's responsibility to ensure that submitted work is their own and does not involve any form of academic misconduct. Students are expected to ask their course instructors for clarification regarding, but not limited to, collaboration, citations, and plagiarism. Ignorance is not an excuse for breaching academic integrity. Students are also required to affix the full Honor Code Affirmation, or the following shortened version, on each item of coursework submitted for grading:

Union College has an honor code as follows. That material will appear in every notebook you submit to me for grading.

> As a student at Union College, I am part of a community that values intellectual effort, curiosity and discovery. I understand that in order to truly claim my educational and academic achievements, I am obligated to act with academic integrity. Therefore, I affirm that I will carry out my academic endeavors with full academic honesty, and I rely on my fellow students to do the same.

Scholastic dishonesty is misrepresenting someone else's work as your own and will not be tolerated.

For this class in particular, we encourage working together during class time. If you missed something, talk to me, talk to your classmates or reach out via piazza or helpdesk. However ALL HOMEWORK must be completed individually. You are encouraged to:

- Talk about concepts in solutions
- Discuss ideas
- Look up online documentation, or examples

You MAY NOT:

- Share code
- Look at another student's code
- Look up solutions to specific problems on the internet
- Copy or paste any text or code to or from a large language model

Ultimately, I may choose to call you into office hours to explain the choices you made in solving a problem. You should be comfortable with explaining the choices you made, how that choice was implemented and why.

Your goal for discussions about any assignment should be that you come away with a better understanding of the problem and of possible ways to approach it so that you can then try out these approaches on your own. You should never leave a discussion with just an answer, without an understanding of how to arrive at that answer. A good general guideline for any discussion, or interaction with an AI model, is that you should not leave the discussion with anything written or typed.

16 CONTACTING ME OUTSIDE OF CLASS

The best method for contacting me outside of class is to stop by my office during Office Hours or whenever my office door is open. You can also schedule a time with me if my regular office hours don't work. Please contact me via email (doingg@union.edu) and we can arrange a phone or zoom call. I respond to emails as soon as possible, but it may take up to one day to get a response during the term, and longer between terms.

17 ADDITIONAL RESOURCES

Mental Health and Campus Resources

Union College is committed to supporting and advancing the mental health and well-being of our students. During the course of their academic careers, students often experience personal challenges that contribute to barriers in learning, such as drug/alcohol problems, strained relationships, chronic worrying, persistent sadness or loss of interest in enjoyable activities, family conflict, grief and loss, domestic violence, difficulty concentrating, problems with organization, procrastination and/or lack of motivation. Students also sometimes come to college with a history of learning difficulties (e.g., any form of special education), experience difficulties succeeding in a particular subject (e.g., math, reading), or have experienced some form of trauma be it emotional or physical (e.g., head injury). These mental health concerns can lead to diminished academic performance and can interfere with daily life activities. If you or someone you know has a history of mental health concerns or if you are unsure and

would like a consultation, a variety of confidential services are available. The Eppler-Wolff Counseling Center provides free counseling and therapy services (including psychiatry) to all enrolled students. Please call (518) 388-6161 or visit the Wicker Wellness Center in person any weekday between 8:30 and 5:00 to schedule an initial contact appointment. Visit the Counseling Center website for more information.

In a crisis situation, or after hours, contact Campus Safety at (518) 388-6911. The National Suicide Prevention hotline also offers a 24-hour hotline at (800) 273-8255.

Any student who faces challenges securing their food or housing and believes this may affect their performance in the course is urged to contact the Dean of Students (dos_office@union.edu) for support or drop into office hours Monday - Friday 8:30 am - 5:00 pm in the Reamer Campus Center room 306. Furthermore, please notify me if you are confortable doing so and I will provide any resources I can. The Union College Persistance Fund can provide financial support in times of unexpected harship to cover needs such as emergency medical expenses not covered by insurance and basic living expenses. Additional sources of supoprt are outlines on the Dean of Students webpage: https://www.union.edu/dean-students .

17.1 ADDENDA

Any community agreed upon additions:

18 TENTATIVE SCHEDULE

W	TOPIC	Due	Runestone
1	Introduction to Data Analytics	Introductory Survey	
2	Using Descriptive Statistics to Talk About Data	HW 0	
3	Python Tools for Data Analytics	HW 1	
4	PANDAS for Exploratory Data Analysis	HW 2	
5	Collecting Data from the Web	HW 3	
6	Data Wrangling and Cleaning	HW 4	
7	Merging, selecting and transforming data	HW 5	
8	Using Simple Machine Learning Methods to Explore Data	HW 6, Midterm Exam	
9	Evaluating performance	Project Draft	
10	Project Presentation	Final Project & Presentation	

19 PROGRESS TRACKER

Learning Goals	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
Fundamental steps of a data analytics workflow										
Responsible use of data										
Storytelling with data										
Aquire and, prepare data for analysis.										
Descriptive data analysis.										
Final project										

Grading

Table 2: Grading Scheme

Course Element	Grade Point Contribution	Notes
Problem Sets	50%	coding
Midterm Exam	25%	written
Final Project	25%	
total	100%	

Table 3: Letter Grade Scale

Letter Grade	Percent range
A	93 – 100
A-	90 – 92
B+	87 – 89
B	83 – 86
B-	80 – 82
C+	77 – 79
C	73 – 76
C-	70 – 72
D	60 – 69
F	0 – 59

Assignments

CSC233 HW 00

CSC 233: Assignment 0, Due Jan 13, 2026

1. Setup Colab Notebook (1 pts)

1. Make a copy of the Colab Notebook template:

https://drive.google.com/file/d/1Ih-HblyS0zAduHUlQfBw4VoeCRx1jyNt/view?usp=sharing

2. Rename it with your own name: [Name]_CSC233_Assignment0.ipynb
3. Follow the instructions in the Colab Notebook.

president_heights.csv: https://drive.google.com/file/d/1PC1EDXtvzMqKFuqVHwC6VbVfvG4h6WsQ/view?usp=sharing

4. Share your notebook with me (doingg@union.edu) as your submission.

2. Runsestone Reading & Activities (1 pts)

1. Enroll in our Runestone course "How to Think Like a Data Scientist"

- https://runestone.academy/runestone/default/user/register
- your Union College email

- unioncollege_httlads_winter26 as the course name
- for review, you may also use the Python course: unioncollege_py4e-int_winter26_da

2. Read and complete the activities for chapters 1 & 4

- you do not need to install python or anaconda, you can use Colab.

3. Setup Check-in Notebook (1 pts)

1. Make a copy of the Check-in Notebook template:

https://colab.research.google.com/drive/18tnFEURw_emX64WbeCRSaA9nyIsNvVsv?usp=sharing

2. Rename it with your own name: [Name]_CSC233_check-in_template.ipynb
3. Follow the instructions in the Colab Notebook.
4. Share your notebook with me (doingg@union.edu) as your submission.

4. Office Hours (1 pts)

1. After completing your check-in notebook, come to office hours sometime in the next 2 weeks.

- Location: ISEC 070
- Student/Office Hours:
  - Wednesday 2:00 pm - 3:30 pm
  - Thursday 4:00 pm - 5:30 pm
  - subject to change, check course website for most up-to-date schedule
  - drop-in or schedule a 15 minute slot: https://calendar.app.google/8bus6pfDvyphR9ar5
  - by appointment for another time or over zoom
- Be careful not to procrastinate, 2 classes worth of students share the office hours.

CSC233 HW 01

CSC 233: Assignment 1, Due Jan 20, 2026, before class

1. Setup Colab Notebook (1 pts)

1. Make a copy of the Colab Notebook template:

https://drive.google.com/file/d/1_FH32gu7P_SzzLX7yeNejaxZ9s5cpCid/view?usp=sharing

2. Rename it with your own name: [Name]_CSC233_W26_Assignment1.ipynb
3. Follow the instructions in the Colab Notebook.

assignment1.csv: https://drive.google.com/file/d/1FMtPBR3YvZaONBNrk-87qoTFhlIT4j1x/view?usp=sharing

4. Share your notebook with me (doingg@union.edu) as your submission.

2. Project Proposal (1 pts)

1. Make a copy of the Project Proposal Notebook template: https://colab.research.google.com/drive/1jSCGEu13RRWJyG-cMt5KmSWa2lVmhVgv?usp=sharing
2. Explore **3** data sets from 3 different data sources you might use

- record your data sets/sources, justifications and interest in Markdown cells
- be sure to include relevent links
- include considerations as to the quality and trustworthiness of your sources
- note that it is OK if your final project does not end up using these exact data sets, but you will at least have these as options

3. Update Check-in Notebook (1 pts)

1. Review your check-in notebook and scores for last week's homework assignment
2. Include a Markdown chunk that attests to my notes or requests a correction/update

CSC233 HW 02

CSC 233: Assignment 2, Due Jan 27 2026, by 10 am

1. Setup Colab Notebook (2 pts)

1. Make a copy of the Colab Notebook template:

https://drive.google.com/file/d/1qDLg_vi5607bz371OWhqv1ely7NBqaWz/view?usp=sharing

2. Rename it with your own name: [Name]_CSC233_W26_Assignment2.ipynb
3. Follow the instructions in the Colab Notebook. Be sure to comment your code and add Markdown cells to document your work.
4. Share your notebook with me (doingg@union.edu) as your submission.

2. Project Data (1 pts)

1. Make a copy of the Project Proposal Notebook template: https://colab.research.google.com/drive/1SEylu5xoSB5CgBtghQYB-qL4Xcu_u_sR#scrollTo=6rqIQUhHRUNR
2. Save a version of your chosen dataset to your google drive.
3. Follow the instructions in the Colab Notebook to load and introduce your data.

3. Runsestone Reading & Activities (4 pts)

1. Read and complete the activities for chapters 2 & 7

Final Project

Project Details

The goal of the final project is to show off your analytic skills. That means you need to tell me a story using data. There are VERY loose requirements, listed below. Ultimately, I need to see:

- your analytic skills (using Python to get, manipulate, analyze and visualize)
- your story telling skills (how do you use the above, to tell a coherent story using data)

BOTH are equally valued.

Here are some tips / thoughts about specifically what I need to see:

1. I need to see you use Pandas to do things. That means choosing data (or sets of data) that are NOT already pre-processed, and perfectly packaged. I need to see some (but not all) of merging, joining, creating new columns, selecting columns, changing types of data.
2. I need to see visualizations. Don't just graph everything because you can. Show me things that assist with your story.
3. I would like some specific analytics - discussion in combination with analytic approaches - i.e. what features are important for a specific

problem, what trends exist in the data and how do you explain them? I'm not looking for you to be right, I'm looking for your story to be well-motivated by data

4. Notice I keep saying 'story'. To tell a story with data REQUIRES that you have a good understanding of the data. To that end I encourage you to choose data, or focus on a problem, about which you ALREADY have some knowledge, or intuition

5. You can find data in all kinds of places on the web. I would focus on CSV data (but you do not have to). You can find useful repositories at data.gov, and kaggle.com, among MANY others. Be cautious about kaggle. I will check to see if others have already performed exactly the analysis you attempt. You certainly don't have to be completely novel, but I would stay away from extremely popular, heavily used data.

6. It is impossible to give minimum or maximum requirements on data size. However, too little data (less than 500 instances) leaves little chance for finding useful patterns. Too few attributes (columns, less than 5) means that you either need to merge in complementary data, or create your own additional columns. Conversely too much data (in rows or columns) means that you almost certainly have to START reducing the size of your data.

7. I cannot tell you how long your notebook should be. HOWEVER consider the size of the grade which this notebook is worth. It should be CONSIDERABLY more in depth than a weekly assignment.

8. I am looking to learn. Part of your grade will be in how well YOU take ME on a journey through data which honestly, at the end, you should understand significantly better than I do. Telling that story is a combination of code (show me results, do not comment out the good stuff) and the write up. This IS part essay, part presentation, part coding exercise.

Weekly Notes

Week 1

Slides from Week 1

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')

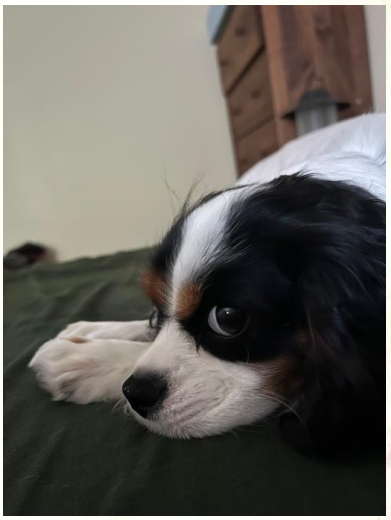
Introductions

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:

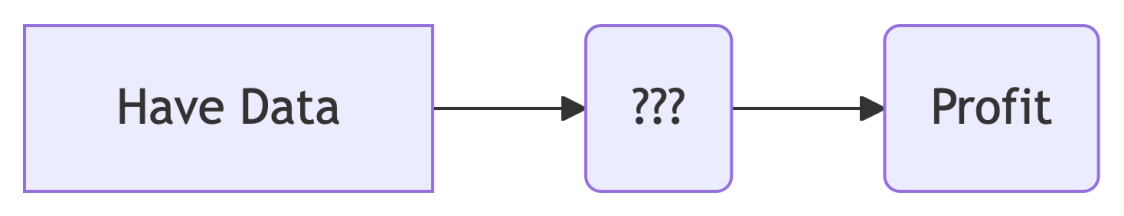
- Name and pronouns
- Help me find you on my roster
- What persepctive do you bring to this class? Could be from a life experience, memory, etc. or your major, minor, research, internship, etc.
- Is there a dataset you've wrestled with in the past that you wished you could have analyzed more deeply?

Fermi questions

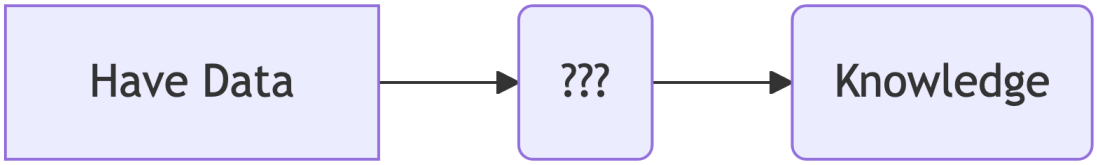
https://fermi-questions.andrechek.com/

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

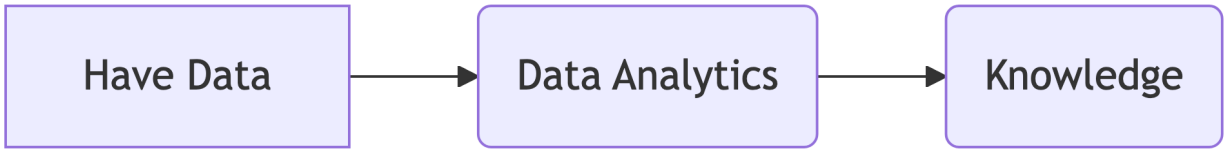
What is ‘Data Analytics’?



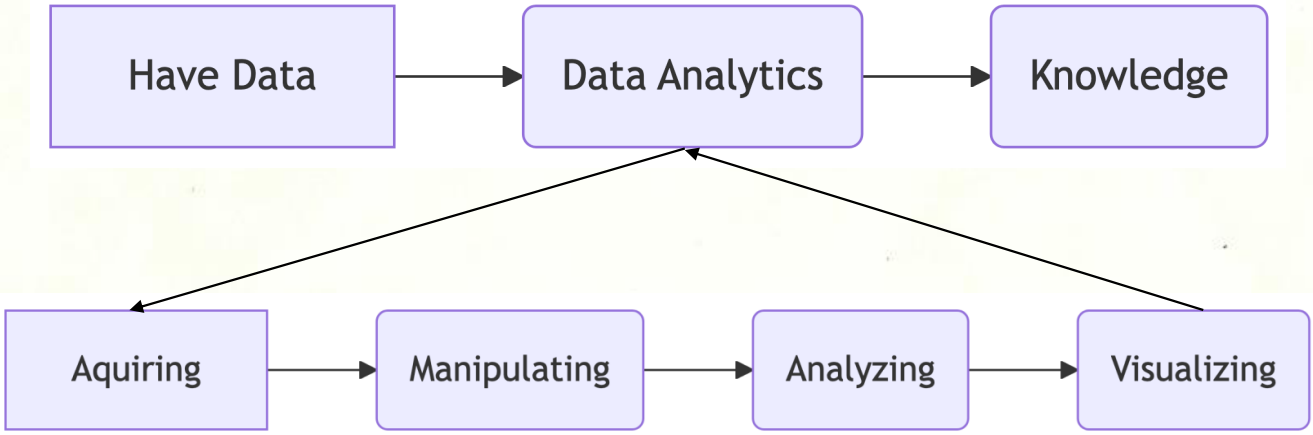
What is ‘Data Analytics’?



What is ‘Data Analytics’?

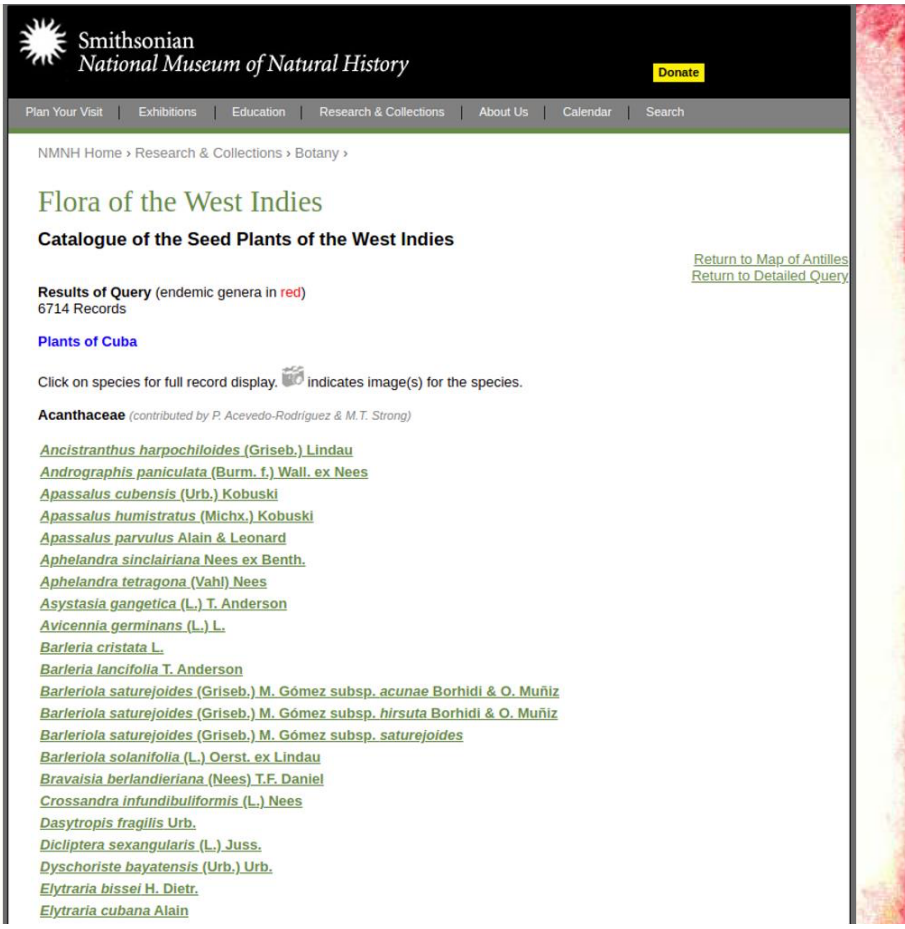


What is ‘Data Analytics’?



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

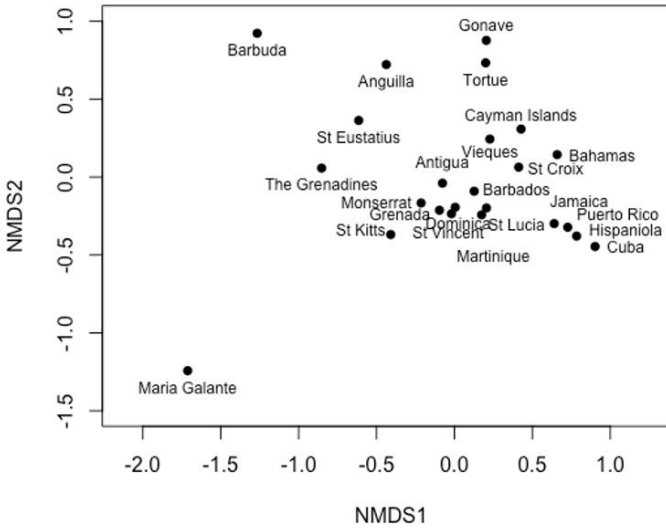
Plants of Cuba

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

Acanthaceae *(contributed by P. Acevedo-Rodríguez & 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

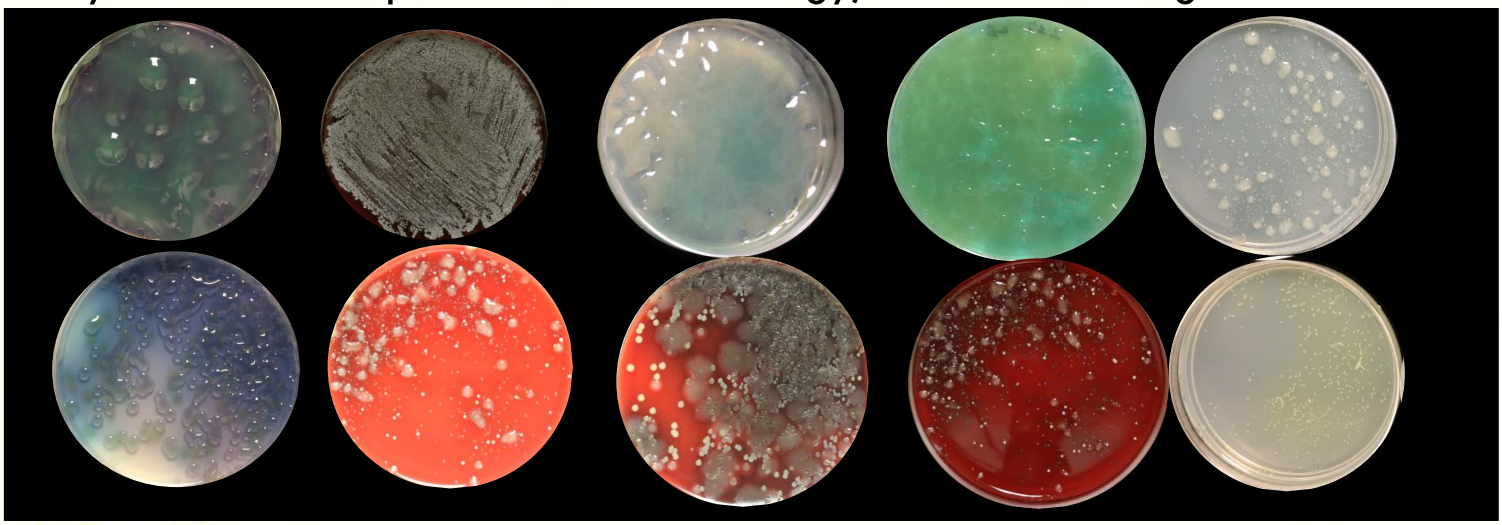

Gonave
Barbuda
Anguilla
Tortue
Cayman Islands
St Eustatius
Vieques
Bahamas
Antigua
St Croix
The Grenadines
Barbados
Monserrat
Jamaica
Grenada
Puerto Rico
Dominica
St Lucia
Hispaniola
St Kitts
St Vincent
Cuba
Martinique
Maria Galante
NMDS1
NMDS2
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
-1.5
-1.0
-0.5
0.0
0.5
1.0

Why shoud you believe me?

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

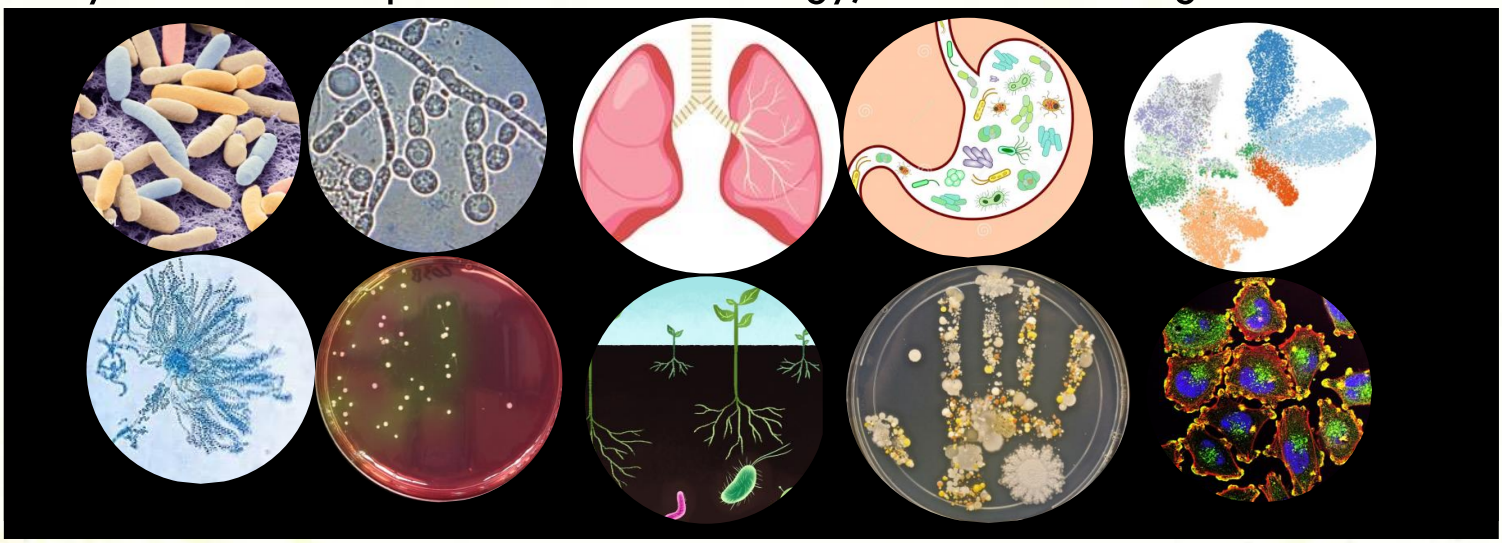
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



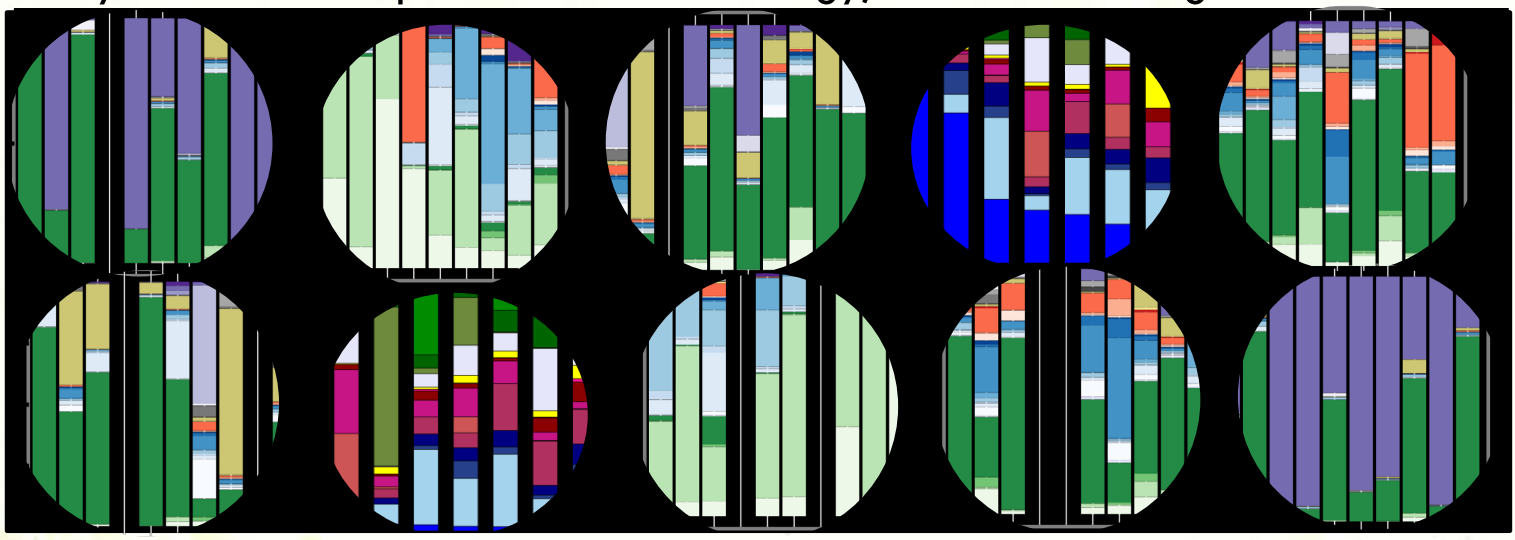
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



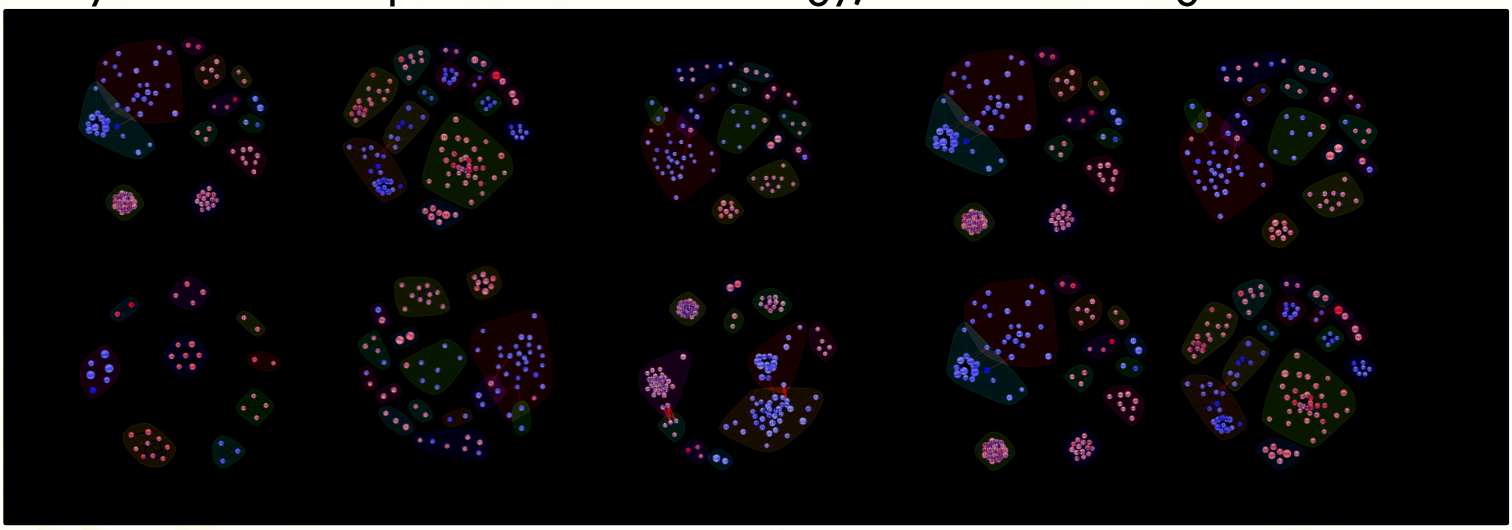
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



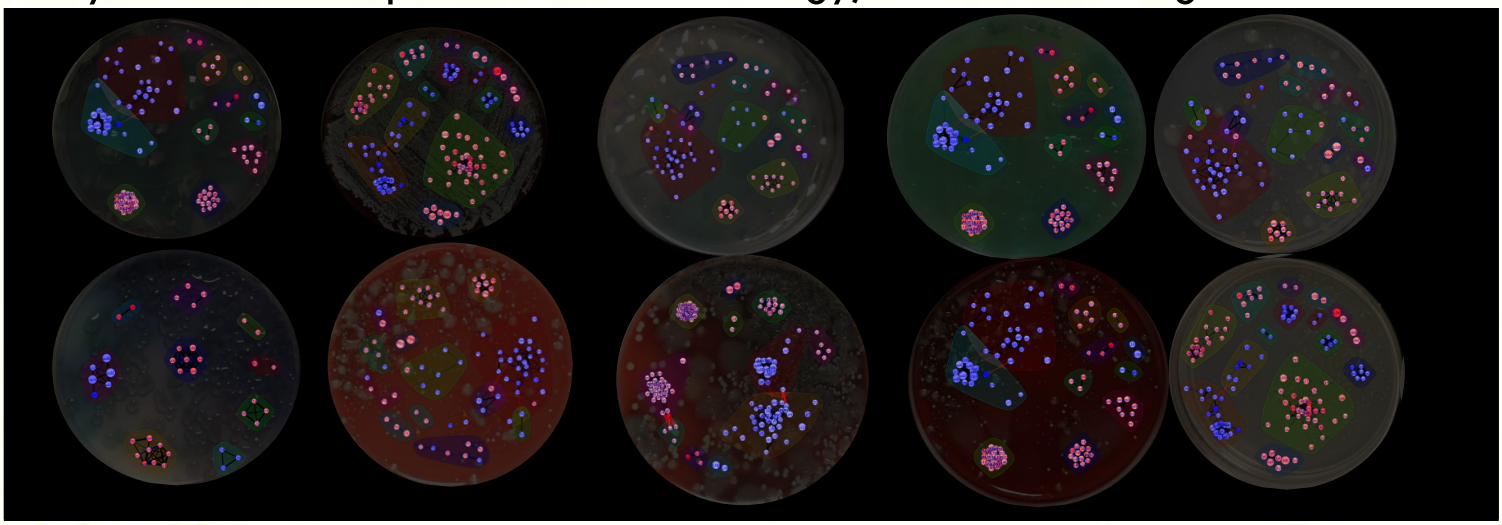
Me: Georgia Doing

My interests: computational microbiology, small cells & big data



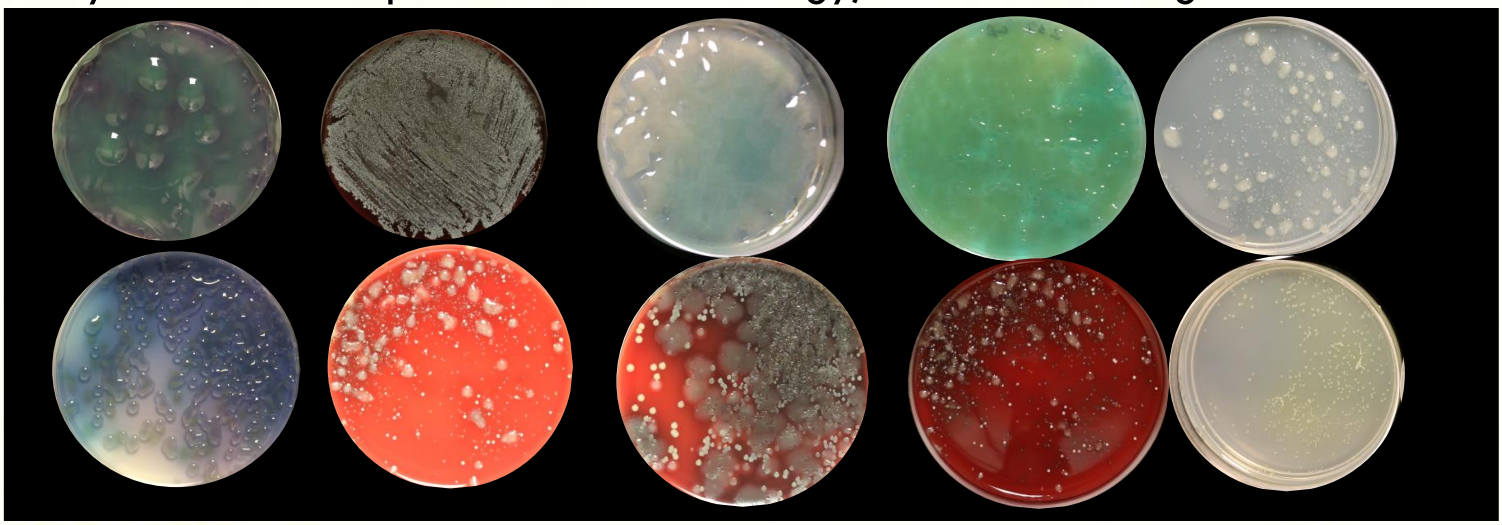
Me: Georgia Doing

My interests: computational microbiology, small cells & big data

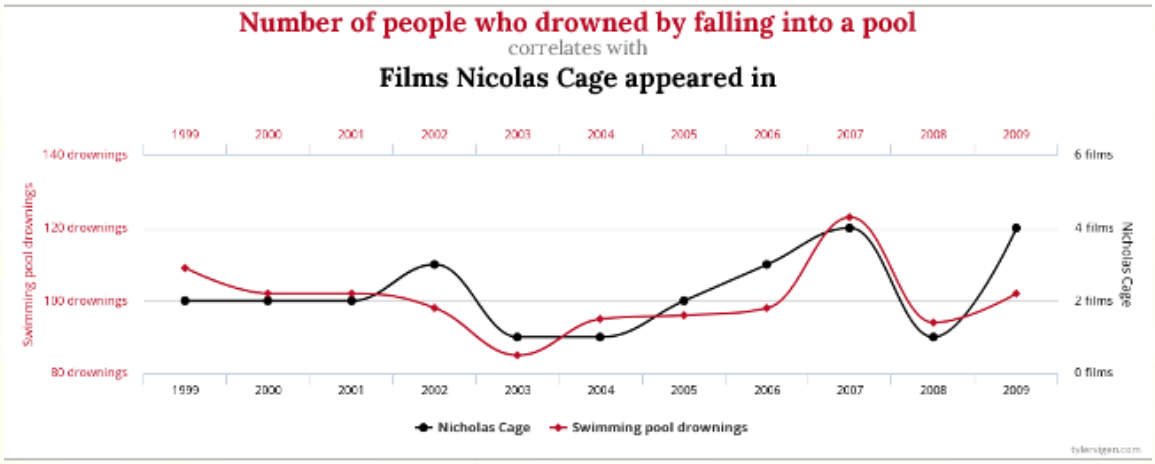


Me: Georgia Doing

My interests: computational microbiology, small cells & big data



Correlation can be…



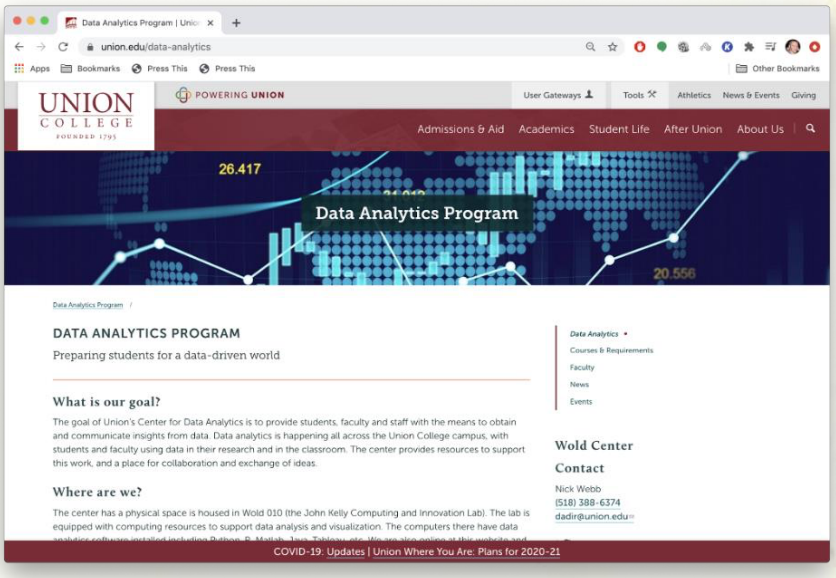
https://www.tylervigen.com/spurious-correlations

Do we have courses that teach that?

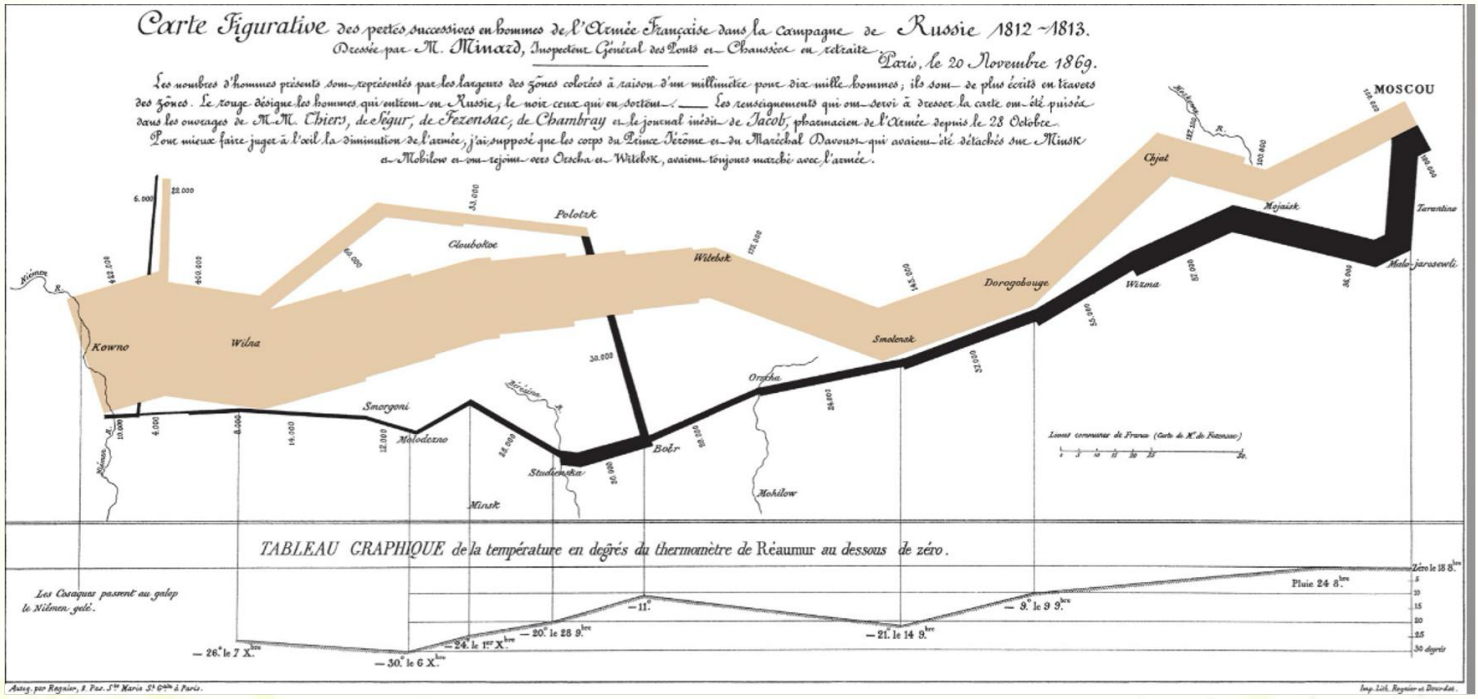
- Analyze Data
  - CSC 233 Data Analytics
  - ENS 215 Exploring Environmental Data
  - ECO 364 Business Analytics
- Visualize Data
  - CSC 234 Data Visualization
  - ECO 134 Data Visualization
- https://www.union.edu/academic/majors-minors/data-analytics

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

Week 2

In-class Colab Notes for Weeks 1 & 2

Week 3

In-class Colab Notes for Week 3

Week 4

Data Links

- course website data path prefix: https://georgiadoing.github.io/CSC233-W26/Data/
  - available datasets:
  - president_heights.csv
  - Seattle2014.csv
  - state_example.csv
  - Employee_ID_NAME_One.csv
  - Employee_ID_NAME_Two.csv
  - Employee_ID_SPECIALTY.csv
  - titanic_train.csv
- state_example.csv: https://georgiadoing.github.io/CSC233-W26/Data/state_example.csv

About

References