DALHOUSIE UNIVERSITY

Inspiring Minds

CSCI 2203 (Section 1):
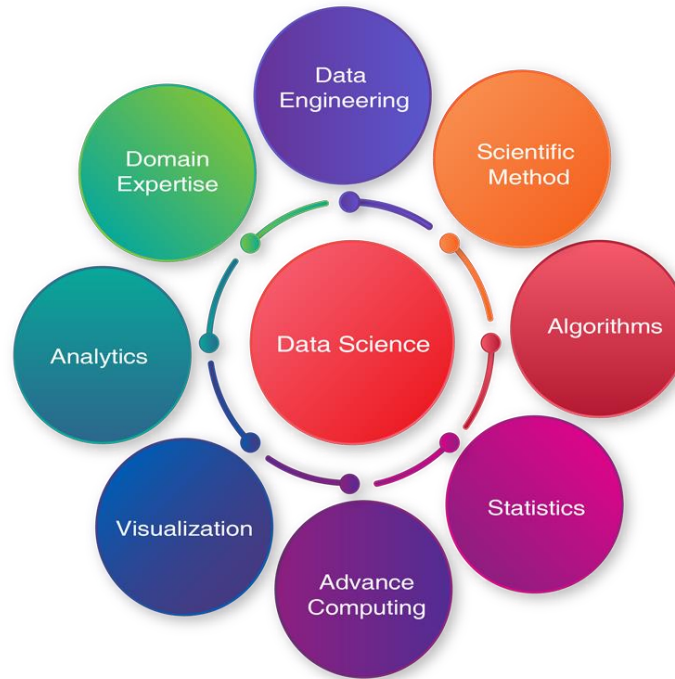# Data Science for Everyone

Yannick Marchand

marchand@cs.dal.ca

Lectures on Monday and Wednesday between

14:35 and **15:55 (**Carleton TUPPER BLDG THTR B**).**

A broad definition

Data Science

A set of scientific methods to extract knowledge and insights from datasets.

Requirements & Grading

- **Quizzes on Articles/Videos:**
  → **30% of final grade**

- **Assignments:**
  → **46% of final grade**

- **Final Exam:**
  → **24% of final grade**

Quizzes on articles/videos

- There will be 7 quizzes during the course:
  - done in class;
  - 2 quizzes on videos;
  - 4 quizzes on research articles;
  - 1 biography article from Wikipedia.

- Each Quiz will have 5 multiple choice questions.

- The mark will be based on your 6 best quizzes (i.e. your lowest mark will not count).

Quizzes on articles/videos

- The duration for each quiz will be 5 minutes.
- The quizzes MUST be taken between 14:45 and 14:55 otherwise you will receive a zero to the quizzes.
- In person (if the sanitary situation permit it): **Wednesday at 14:45**
  - **Closed book;**
  - **No dictionaries, notes, calculators, cell phones, PDAs, talking slide rulers, or other electronic aids allowed;**
  - **I will give you a password to access to the quiz;**
  - **do not ask passwords from other students!;**
  - **do not provide/give passwords to other students!;**
  - **Make sure to run TestQuiz to install Respondus LockDown on your laptop.**

Quizzes: videos/articles

Quiz 1: January 24
Video

YouTube video entitled:
"Deb Roy: The birth of a word"

Please copy and paste the following URL into your browser:

https://www.youtube.com/watch?v=RE4ce4mexrU
https://www.ted.com/talks/deb_roy_the_birth_of_a_word/transcript (transcript)

DURATION: 19 min. 52 sec.

Quizzes: videos/articles

Quizzes: videos/articles

Quiz 2: January 31
Journal article

The Small-World Problem
By Stanley Milgram

Psychology Today, vol. 1, no. 1, May 1967, pp61-67

Quizzes: videos/articles

Quiz 3: February 7
Journal article

Gender, social norms, and survival in maritime disasters

Mikael Elinder[a,b,1] and Oscar Erixson[a]

[a]Department of Economics, Uppsala University, SE-751 20 Uppsala, Sweden; and [b]The Research Institute of Industrial Economics, SE-102 15 Stockholm, Sweden

Proceedings of the National Academy of Sciences (2012), 109 (33), 13220-13224.

Quizzes: videos/articles

Quiz 4: February 14
Biography article from Wikipedia

1912 - 1954

Quizzes: videos/articles

Quiz 5: February 28
Journal article

PLoS one

Positivity of the English Language

Isabel M. Kloumann[1,4], Christopher M. Danforth[1,2,3]*, Kameron Decker Harris[1], Catherine A. Bliss[1], Peter Sheridan Dodds[1,2,3]*

1 Department of Mathematics and Statistics, University of Vermont, Burlington, Vermont, United States of America, 2 Center for Complex Systems, University of Vermont, Burlington, Vermont, United States of America, 3 Vermont Advanced Computing Center, University of Vermont, Burlington, Vermont, United States of America, 4 Department of Physics, University of Vermont, Burlington, Vermont, United States of America

Abstract

Over the last million years, human language has emerged and evolved as a fundamental instrument of social communication and semiotic representation. People use language in part to convey emotional information, leading to the central and contingent questions: (1) What is the emotional spectrum of natural language? and (2) Are natural languages neutrally, positively, or negatively biased? Here, we report that the human-perceived positivity of over 10,000 of the most frequently used English words exhibits a clear positive bias. More deeply, we characterize and quantify distributions of word positivity for four large and distinct corpora, demonstrating that their form is broadly invariant with respect to frequency of word use.

Quizzes: videos/articles

Quiz 6: March 6
Journal article

Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures

Scott A. Golder* and Michael W. Macy

Quizzes: videos/articles

Quiz 7: March 13
Video

YouTube video entitled:
"The World's Most Mysterious Manuscript"

Please copy and paste the following URL
into your browser:

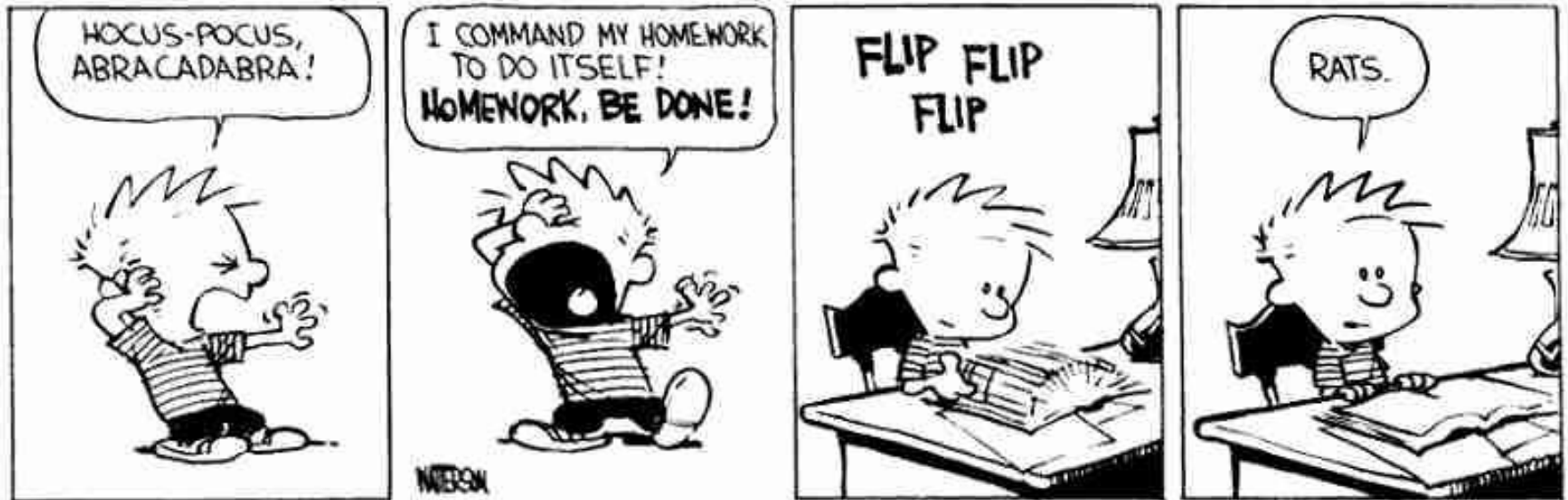https://www.youtube.com/watch?v=awGN5NApDy4

DURATION: 50 min. 20 sec.

Quizzes: example of question

- **Question: Which language is not represented in the corpora of this study:**
  - A. English
  - B. French
  - C. German
  - D. Arabic
  - E. Spanish

Homework assignments

- 3 assignments
  - 15 points for assignments 1 and 2; 16 points for assignment 3;
  - "Open quizzes" for 3 weeks on BrightSpace.

Homework assignments

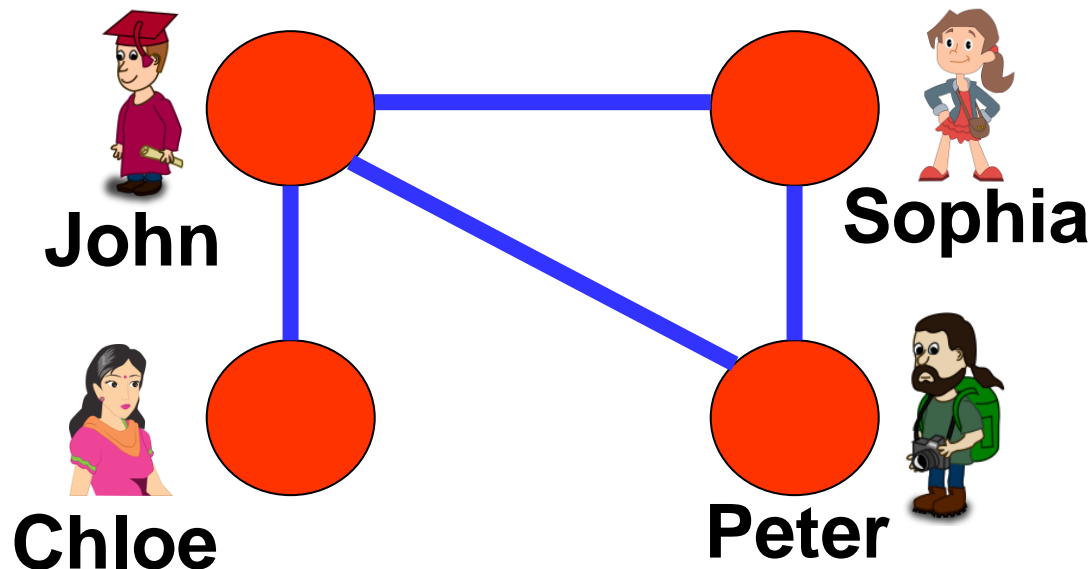| Assignment	When is it given/posted?	Due date
#1	January 1	January 31
#2	January 31	February 28
#3	February 28	March 27

Assignment #1

- Calculate statistics from real Titanic data using a BrightSpace quiz (e.g. **What is the percentage of female passengers?**)

- Excel file:

pclass	survived	name	sex	age	sibsp	parch	ticket
1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160
1	1	Allison, Master. Hudson Trevor	male	0.917	1	2	113781
1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781
1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781
1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781
1	1	Anderson, Mr. Harry	male	48	0	0	19952
1	1	Andrews, Miss. Kornelia Theodosia	female	63	1	0	13502
1	0	Andrews, Mr. Thomas Jr	male	39	0	0	112050
1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53	2	0	11769
1	0	Artagaveytia, Mr. Ramon	male	71	0	0	PC 17609
1	0	Astor, Col. John Jacob	male	47	1	0	PC 17757
1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	18	1	0	PC 17757
1	1	Aubart, Mme. Leontine Pauline	female	24	0	0	PC 17477
1	1	Barber, Miss. Ellen "Nellie"	female	26	0	0	19877
1	1	Barkworth, Mr. Algernon Henry Wilson	male	80	0	0	27042
1	0	Baumann, Mr. John D	male		0	0	PC 17318
1	0	Baxter, Mr. Quigg Edmond	male	24	0	1	PC 17558
1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50	0	1	PC 17558
1	1	Bazzani, Miss. Albina	female	32	0	0	11813
1	0	Beattie, Mr. Thomson	male	36	0	0	13050

Assignment #2

- You will be provided with data (names) to build a social network of friends.

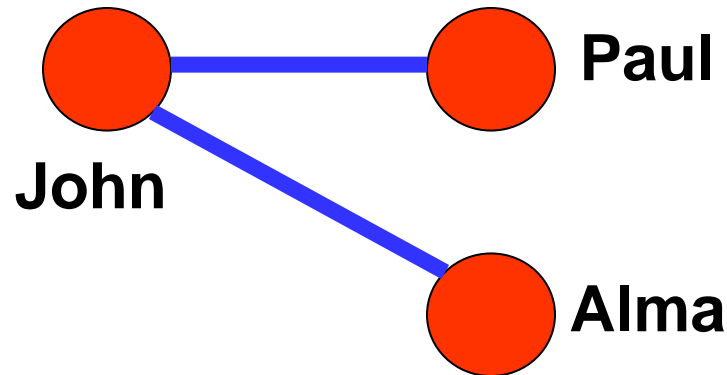- You will be asked to provide/calculate some features on this social network.

John   Sophia

Chloe   Peter

Assignment #2

<u>Simplistic example</u>:

John knows **Paul**

John knows **Alma**
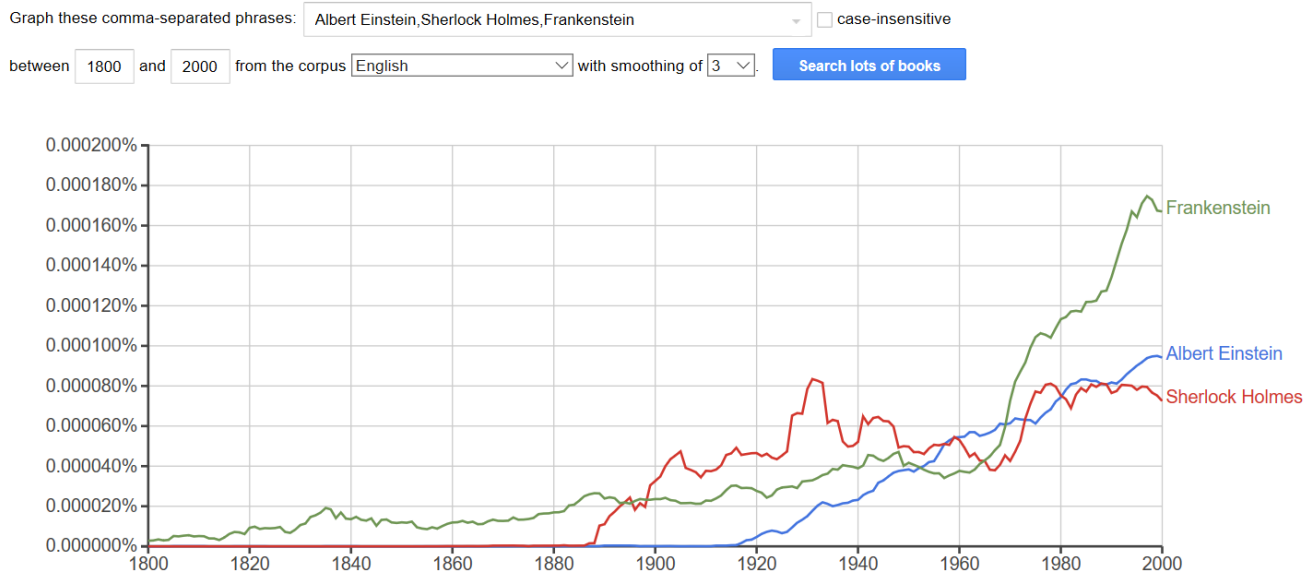
Step 1: Build the social network of friends.

Question: How many friends does John have?

Answer: Two

Assignment #3

- Based on **Google Books Ngram Viewer** (https://books.google.com/ngrams).

- Details of the assignment provided later on BrightSpace.

Homework assignments

- NO plagiarism!!!
- The assignment papers **MUST** be submitted on time.
- NO submissions will be accepted after the deadline (i.e. a zero will be given to the assignment).
- Assignments will be posted **at least 3 weeks** before the submission deadlines.

Homework assignments

- ZERO Tolerance!

- I will not reply to the following kinds of emails:

"Unfortunately, my Wi-Fi was very spotty last night and when I attempted to submit the assignment, it was cutting out."

"I somehow fell asleep at 8:30pm yesterday on the couch after dinner. I forgot to upload my homework."

"I have a bad cold."

Final exam

- 40 min on **<u>Wednesday 3 April 2023 starting at 2:45pm.</u>**

- About 60 questions based on lecture content throughout the course.

- A variety of question types:
  - Short answer
  - Multiple choice
  - Exercises (e.g. finding the shortest path in a network)

Final exam

- Closed book;
- No dictionaries, notes, calculators, cell phones, talking slide rulers, or other electronic aids allowed;
- I will give you a password to access to the quiz;
  - do not ask passwords from other students!;
  - do not provide/give passwords to other students!
- Make sure to run TestQuiz to install Respondus LockDown on your laptop.

Success tips

- Attend and understand the lectures.

- Carefully read the syllabus to make sure not to miss any deadlines!

- Refer to BrightSpace on a regular basis.

- Take notes when reading papers.

- Complete and submit assignments on time (the time for these assignments should only take 1 or 2 hours).

Students with a Disability

Please see/contact me **with at least two weeks notice** if you have a disability that may require some modification of the seating, testing or other class requirements, so that appropriate arrangements may be made

Tentative schedule (lectures)

- **Syllabus**

- **Topic 1:** **Data Science: What Is Data? What Is Science?**

- **Topic 2:** **Social Network Analysis: We Are All Connected**

Tentative schedule (lectures)

- **Topic 3:** **Culturomics: What Can We Learn From 5 Million Digitized Books?**

- **Topic 4:** **Machine Learning: Would You Survive the Titanic?**

BrightSpace

CSCI 2203 (section 1): Key dates

QUIZZES On Wednesday at 2:45pm	<u>Quiz 1:</u> January 24 <u>Quiz 2:</u> January 31 <u>Quiz 3:</u> February 7 <u>Quiz 4:</u> February 14 <u>Quiz 5:</u> February 28 <u>Quiz 6:</u> March 6 <u>Quiz 7:</u> March 13
ASSIGNMENTS On Wednesday (deadline: 11:59pm)	<u>A1:</u> Posted on January 1 (<u>due:</u> January 31) <u>A2:</u> Posted on January 31 (<u>due:</u> February 28) <u>A3:</u> Posted on February 28 (<u>due:</u> March 27)
FINAL EXAM	Wednesday April 3 at 2:45pm

Lectures on <u>Monday</u> and <u>Wednesday</u> between
2:35pm and **3:55pm** (Carleton TUPPER BLDG THTR B).