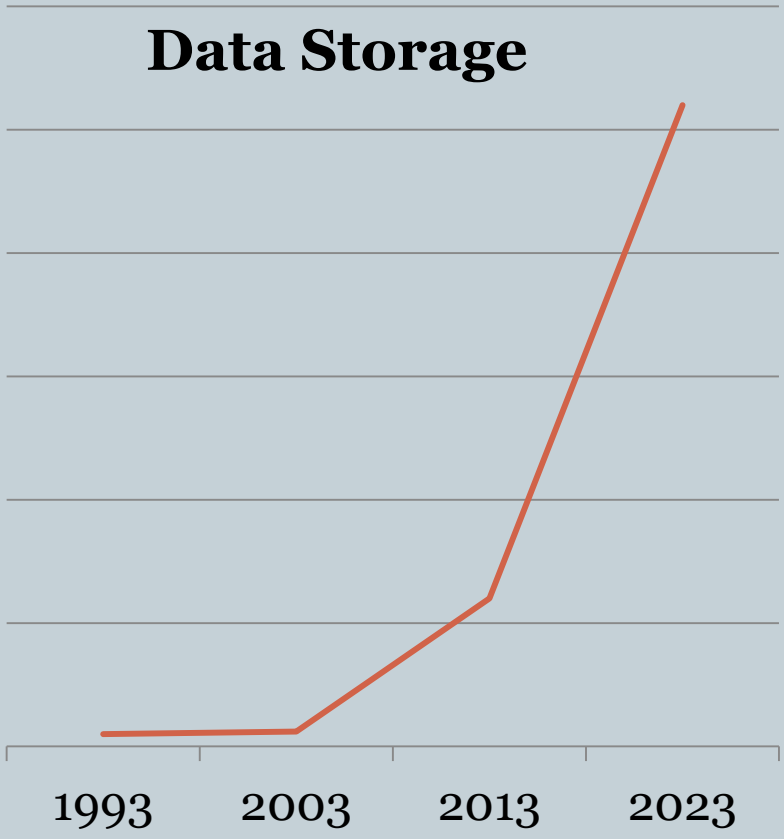
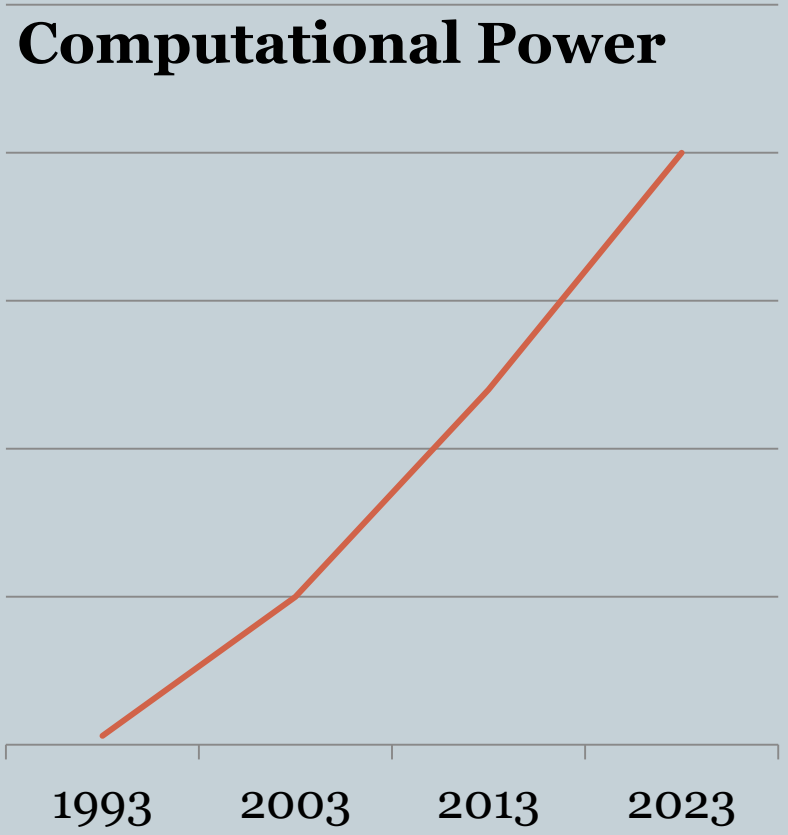
Computational Tools and Techniques

ANUSTUBH AGNIHOTRI
THE UNIVERSITY OF CALIFORNIA AT BERKELEY
WEDNESDAY 2-4 PM

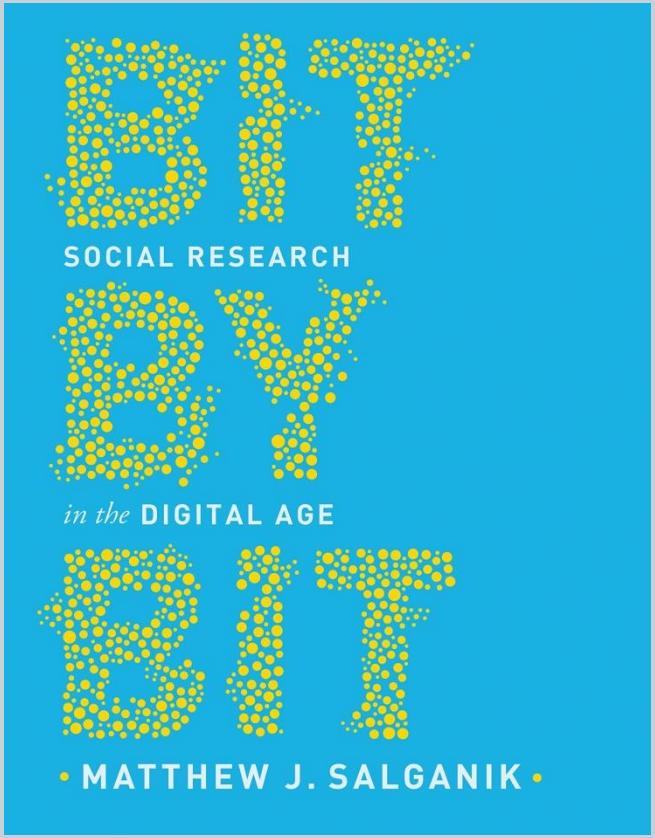
Motivation for this class





Research Examples

- King, Pan, Roberts (2013): How Censorship in China Allows Government Criticism but Silences Collective Expression
- Chetty et al (2014): Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States
- Blumenstock et al (2015): Predicting poverty and wealth frommobile phone metadata



Goals of this Class

- Making commonly used software programs and tools more accessible
- Making collaboration with other researchers easier
- Introducing you to diverse techniques for collecting and analyzing digital information

Making software programs/tools accessible

- By the end of the course, you should know how to
  - Find solutions to new programming challenges online
    - Organize your projects
    - Find and collect data using the internet
  - Clean data and present your results in R
    - Use tidyverse!
    - Use functions!

Collaboration

- Using Version Control Systems (Github)
- Sharing and editing scripts/programs across multiple team members

Previewing advance techniques

- Basics of a Web scrapping
  - Beautiful Soup
  - Selenium
- APIs
- Text as Data
- Machine Learning

Where can I find materials for this course?

- Github
  - Course content will be posted on Github
  - We will learn how to use Github towards the end of the class
- bCourses
  - Assignments submitted on bCourses.
  - Ask and provide help to classmates using discussion threads
- Datacamp
  - Everyone in the class will receive access to Datacamp's online tutorials

Syllabus

- Main topics
  - R/Python
  - Webscrapping + HTML
  - Data Cleaning/ Visualization
  - Text as Data
  - Machine Learning
- Grading
  - Assignments
  - Participation
  - Final Project

Assignments

- Datacamp Tutorials
- Webscraping assignment
- API assignment
- Data cleaning+visualization project

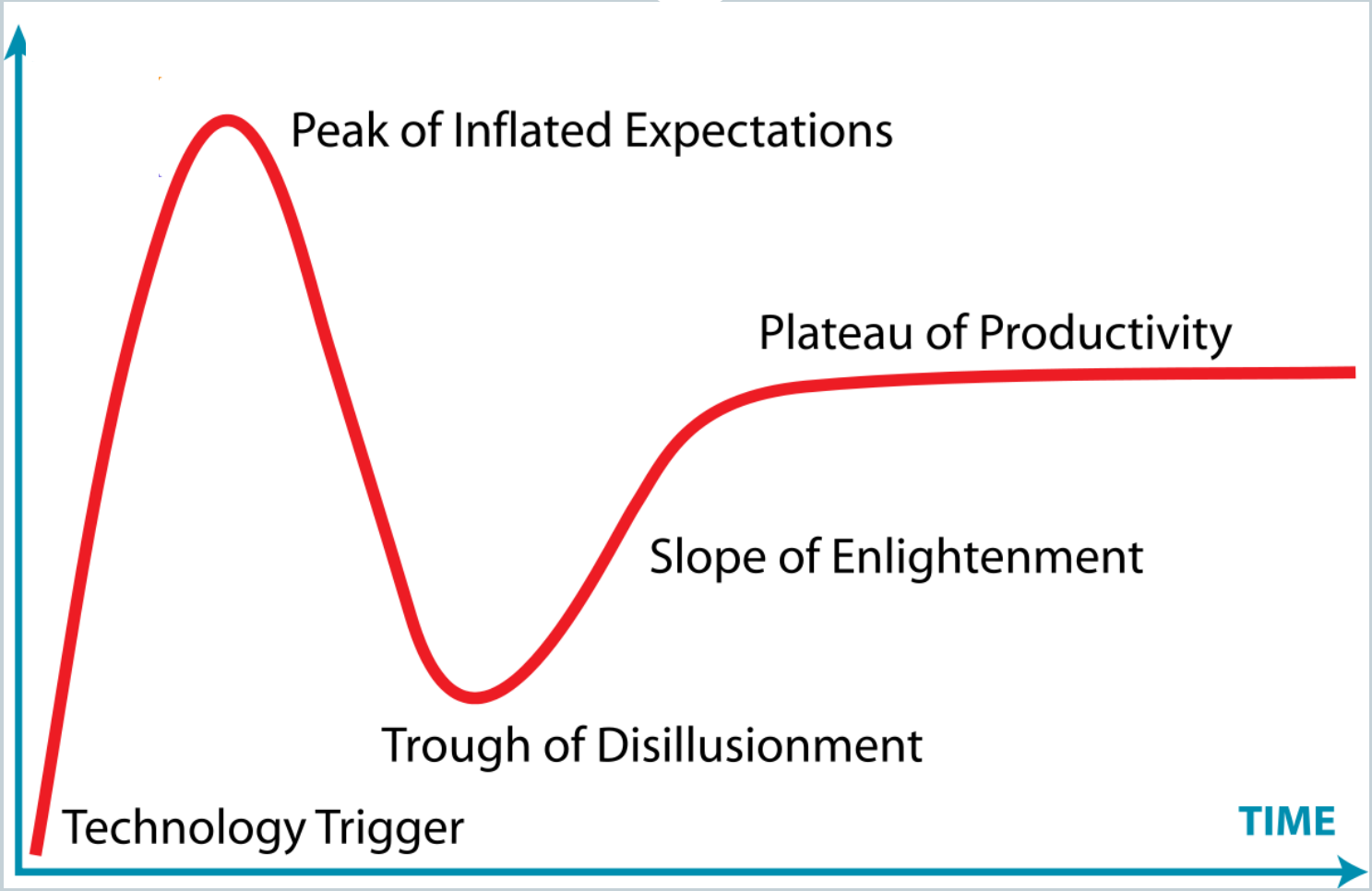
Participation

- Coming to lectures (section?)
- Asking questions
- Seeking feedback on projects/assignments

Final Project

- Final Project Examples
  - Collecting and plotting Facebook API data on reactions to political candidates
  - Scraping data on court cases in China and predicting duration of punishment
  - Creating and newspaper dataset and doing a topic model

Learning new programming tools



Common Obstacles to Learning

- Some stuff is hard
  - Learning to think like a computer
  - Developing intuition about dataset structure
- Some stuff takes work
  - Becoming impatient while learning basic concepts (most errors lie here)
  - Infrequent practice (fluency doesn't come from weekends)
  - Copy-and-pasting instead of typing (ctrl+c is not your friend)*
  - Disorganized coding (would you write an essay without an outline?)
  - Failing to comment code as you go (## what does this do?)
- But other stuff is a gatekeeping problem
  - Unspoken/tacit knowledge among coders
  - Inconsistent and/or technical language used to describe concepts

Github Introduction and Installation

- Use github desktop (GUI)
- Clone the repository on your desktop
- Folders/files will be updated before each class