Introduction: Social Data Analytics

INF1005/6, Prof. Alex Hanna

March 1, 2018

Class Description

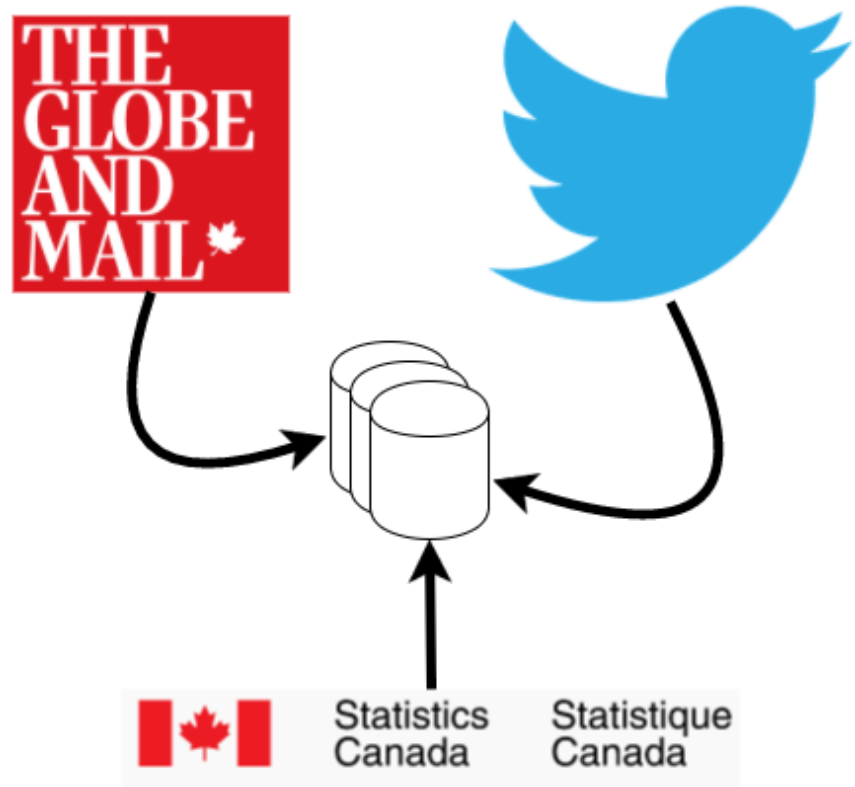
- Data Science
- Python programming
- Social data

Class Goals

- Identify
- Collect
- Transform
- Analyze
- Visualise

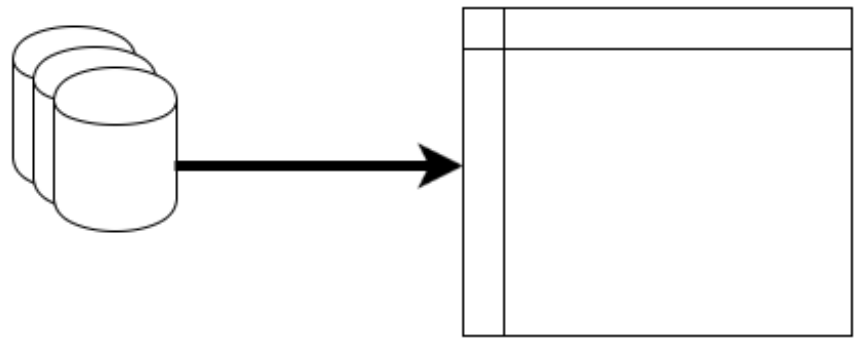
Class Goals

- Identify
- Collect
- Transform
- Analyze
- Visualise



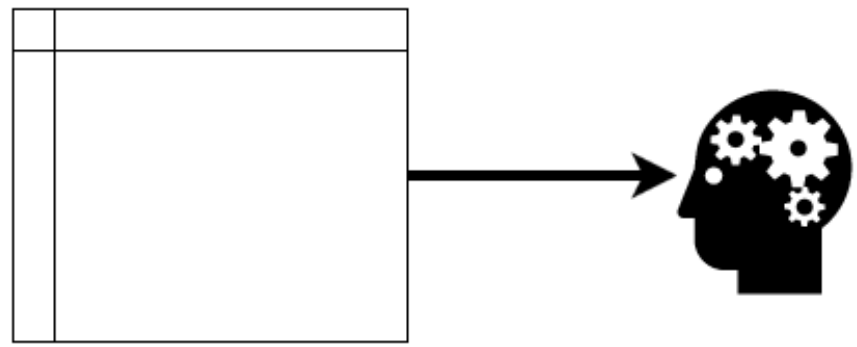
Class Goals

- Identify
- Collect
- Transform
- Analyze
- Visualise



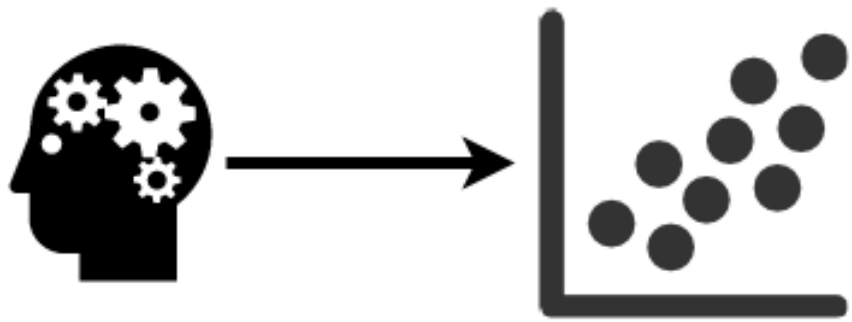
Class Goals

- Identify
- Collect
- Transform
- Analyze
- Visualise



Class Goals

- Identify
- Collect
- Transform
- Analyze
- Visualise



Class Requirements

- Attendance (20%)
- Assignment 1 (40%)
- Assignment 2 (40%)

Class Requirements

- Attendance (20%)
- Assignment 1 (40%)
- Assignment 2 (40%)

Class Requirements

- Attendance (20%)
- Assignment 1 (40%)
- Assignment 2 (40%)

Procedures and Rules

- E-Culture
- Academic integrity

Procedures and Rules

- E-Culture
- Academic integrity

About Me

- Professor Alex Hanna (Prof. Hanna, or Dr. Hanna)
- PhD and MS in Sociology
- BS in Computer Science
- Transgender and queer
- Pronouns: she/her/hers

About Me

- Professor Alex Hanna (Prof. Hanna, or Dr. Hanna)
- PhD and MS in Sociology
- BS in Computer Science
- Transgender and queer
- Pronouns: she/her/hers

About Me

- Professor Alex Hanna (Prof. Hanna, or Dr. Hanna)
- PhD and MS in Sociology
- BS in Computer Science
- Transgender and queer
- Pronouns: she/her/hers

About Me

- Professor Alex Hanna (Prof. Hanna, or Dr. Hanna)
- PhD and MS in Sociology
- BS in Computer Science
- Transgender and queer
- Pronouns: she/her/hers

About Me

- Professor Alex Hanna (Prof. Hanna, or Dr. Hanna)
- PhD and MS in Sociology
- BS in Computer Science
- Transgender and queer
- Pronouns: she/her/hers

About Me

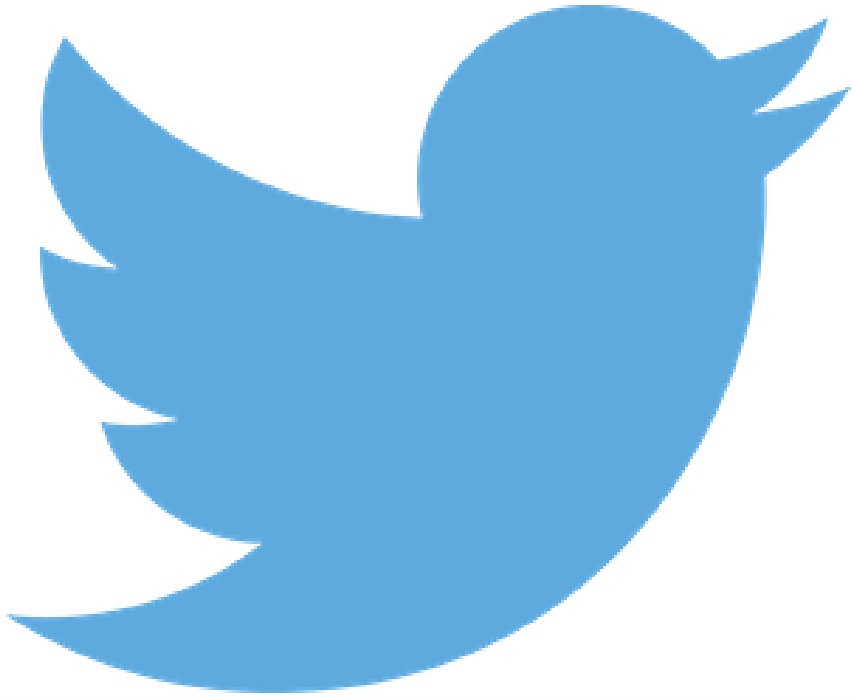
- Professor Alex Hanna (Prof. Hanna, or Dr. Hanna)
- PhD and MS in Sociology
- BS in Computer Science
- Transgender and queer
- Pronouns: she/her/hers
- Roller derby skater

About You!

http://bit.ly/inf1006-1

What are Social Data Analytics?

500M+	300M+	80%+	4.2B	95M+
Monthly Actives	Daily Actives	Outside The U.S.	Likes Daily	Photos/Videos Per Day



313M

Monthly active users

1B

Unique visits monthly to sites with embedded Tweets

82%

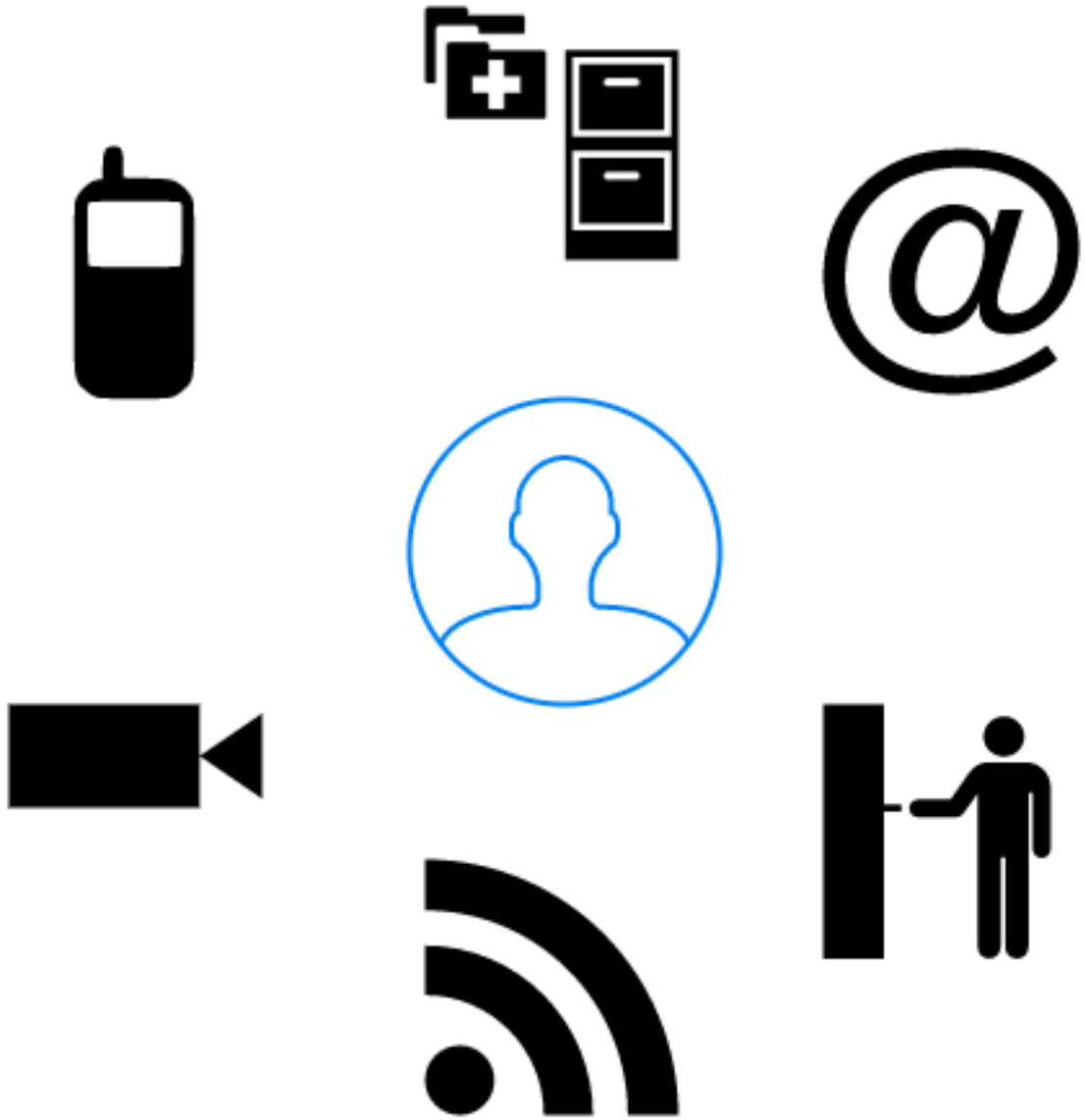
Active users on mobile

3,860

Employees around the world

Five-year Census of every Canadian (>35 million)
350+ active surveys on Canadian life

Big Data is “Trace” Data



Big Data Fields

"Data Science"
"Business intellgence and analytics"
"Computational social science"

Computational Social Science (Shah et al. 2015)

- the use of large, complex datasets, often... measured in terabytes or petabytes
- the frequent involvement of 'naturally occurring' social and digital media sources and other electronic databases
- the use of computational or algorithmic solutions to generate patterns and inferences from these data
- the applicability to social theory in a variety of domains from the study of mass opinion to public health, from examinations of political events to social movements

Computational Social Science (Shah et al. 2015)

- the use of large, complex datasets, often... measured in terabytes or petabytes
- the frequent involvement of 'naturally occurring' social and digital media sources and other electronic databases
- the use of computational or algorithmic solutions to generate patterns and inferences from these data
- the applicability to social theory in a variety of domains from the study of mass opinion to public health, from examinations of political events to social movements

Computational Social Science (Shah et al. 2015)

- the use of large, complex datasets, often... measured in terabytes or petabytes
- the frequent involvement of 'naturally occurring' social and digital media sources and other electronic databases
- the use of computational or algorithmic solutions to generate patterns and inferences from these data
- the applicability to social theory in a variety of domains from the study of mass opinion to public health, from examinations of political events to social movements

Computational Social Science (Shah et al. 2015)

- the use of large, complex datasets, often... measured in terabytes or petabytes
- the frequent involvement of 'naturally occurring' social and digital media sources and other electronic databases
- the use of computational or algorithmic solutions to generate patterns and inferences from these data
- the applicability to social theory in a variety of domains from the study of mass opinion to public health, from examinations of political events to social movements

Computational Social Science

- Network analysis
- Automated content (e.g. text and image) analysis
- Massive Online Experiments

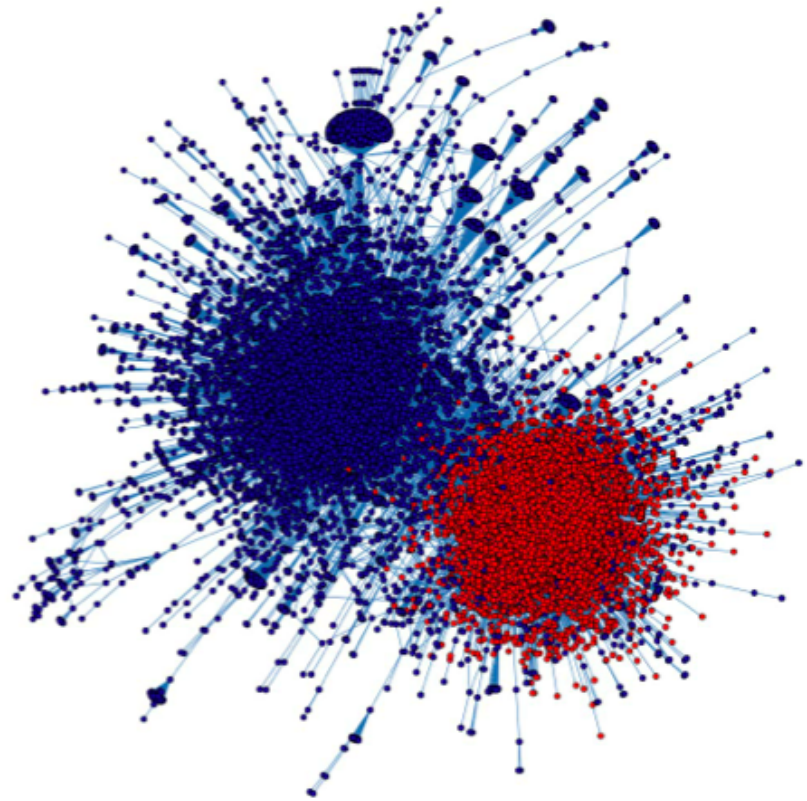
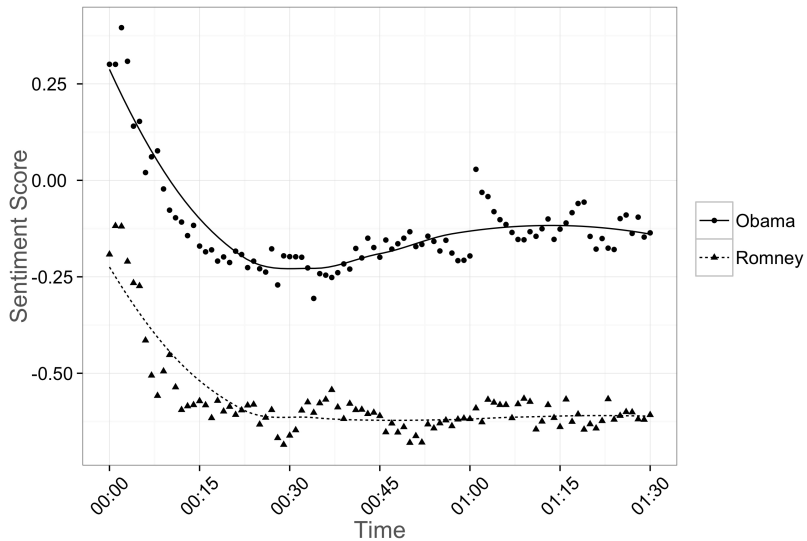


Fig. 2. The political retweet network, laid out using a force-directed algorithm. Node colors reflect cluster assignments (see text).

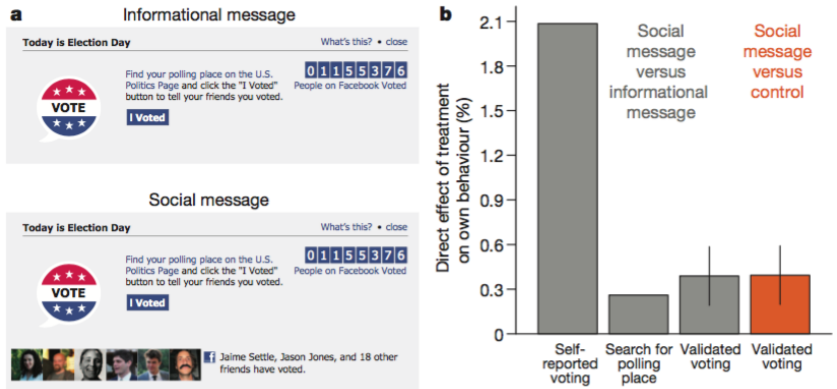
Computational Social Science

- Network analysis
- Automated content (e.g. text and image) analysis
- Massive Online Experiments

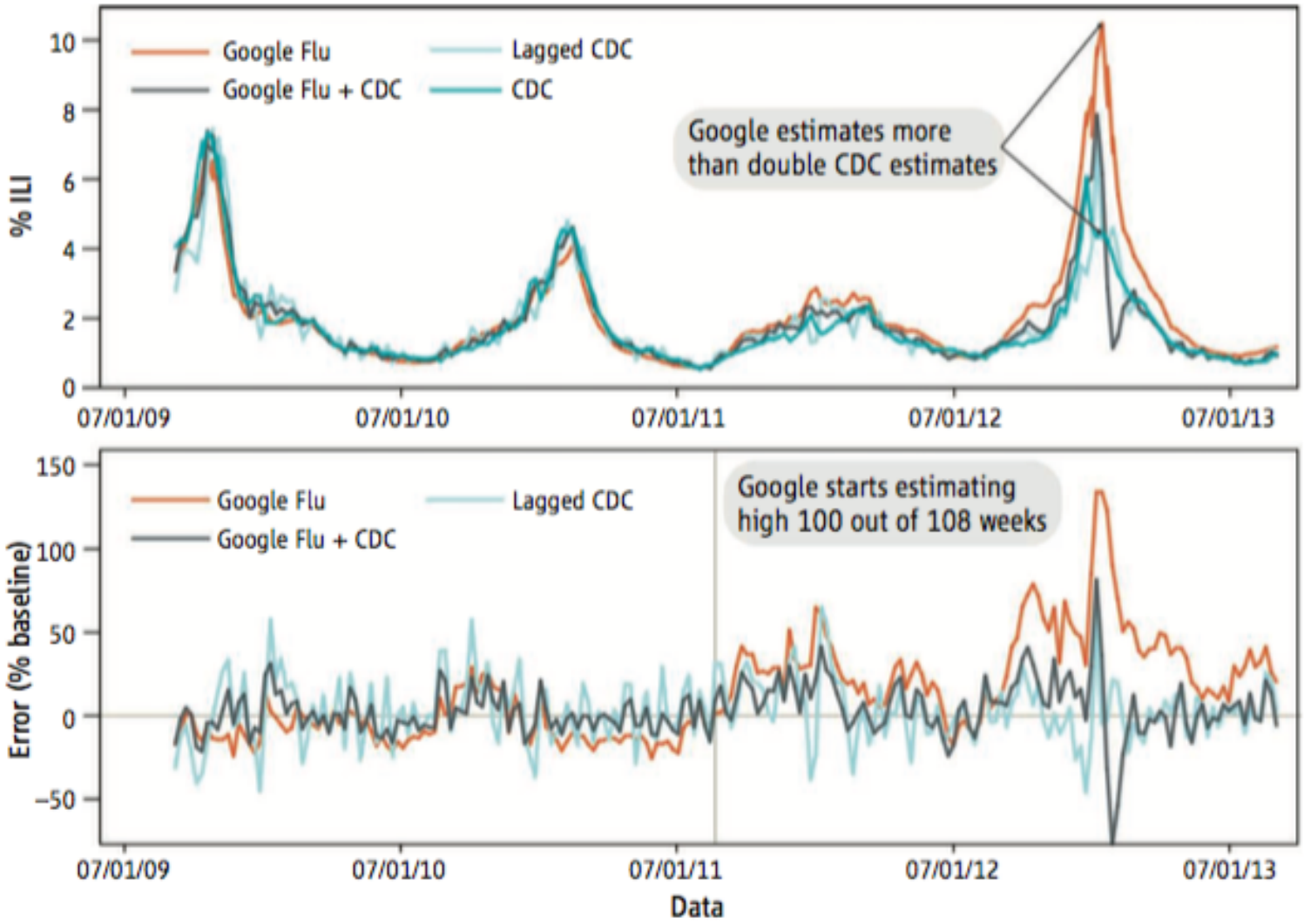


Computational Social Science

- Network analysis
- Automated content (e.g. text and image) analysis
- Massive Online Experiments



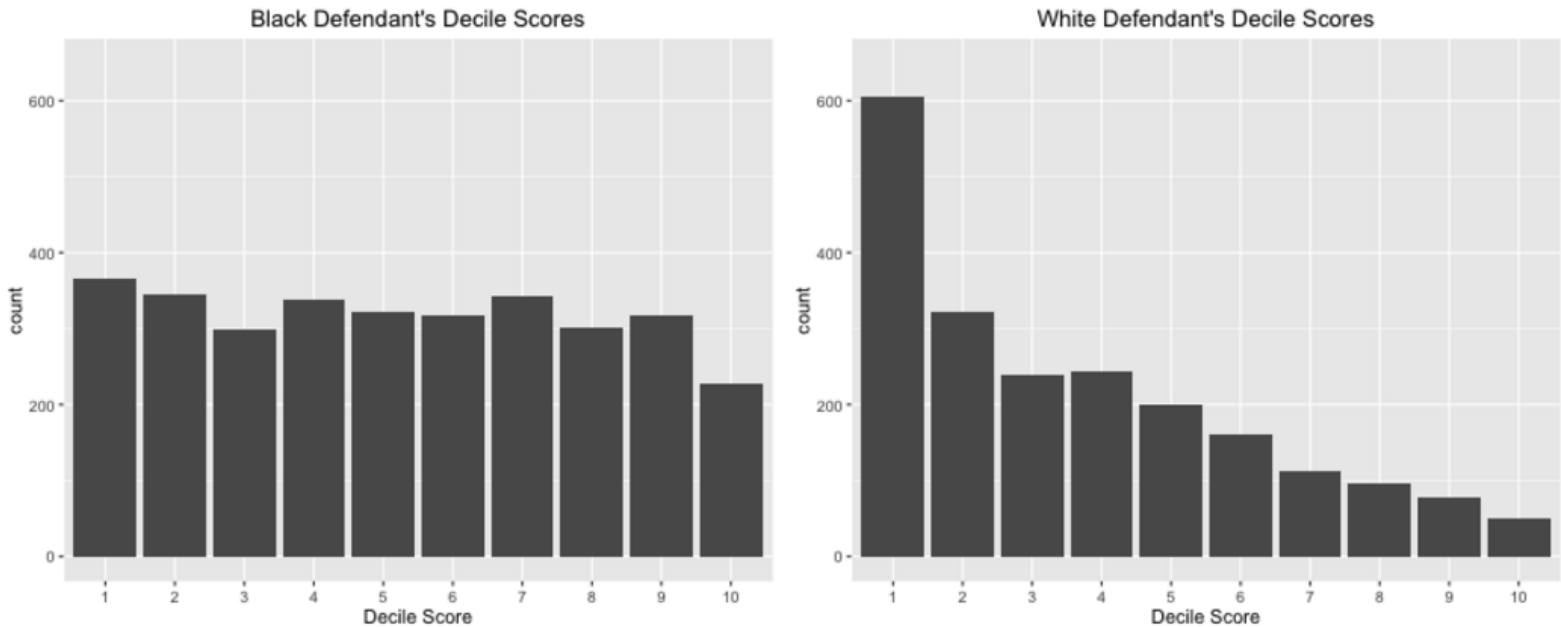
Prediction and Platform Politics



Data and Informed Consent

Data and Criminal Justice

We began by looking at the risk of recidivism score. Our initial analysis looked at the simple distribution of the COMPAS decile scores among whites and blacks. We plotted the distribution of these scores for 6,172 defendants who had not been arrested for a new offense or who had recidivated within two years.



These histograms show that scores for white defendants were skewed toward lower-risk categories, while black defendants were evenly distributed across scores. In our two-year sample, there were 3,175 black defendants and 2,103 white defendants, with 1,175 female defendants and 4,997 male defendants. There were 2,809 defendants who recidivated

Link: Machine Bias in Pre-trial Risk Assessment

Coding as Thinking

- Programming languages get old, thinking does not
- When you program, you open doors
  - DIY
  - Troubleshooting
  - Career

Tools

DataCamp