Ethics

Prof. Bisbee

Vanderbilt University

Lecture Date: 2023/04/19

Slides Updated: 2023-04-18

Agenda

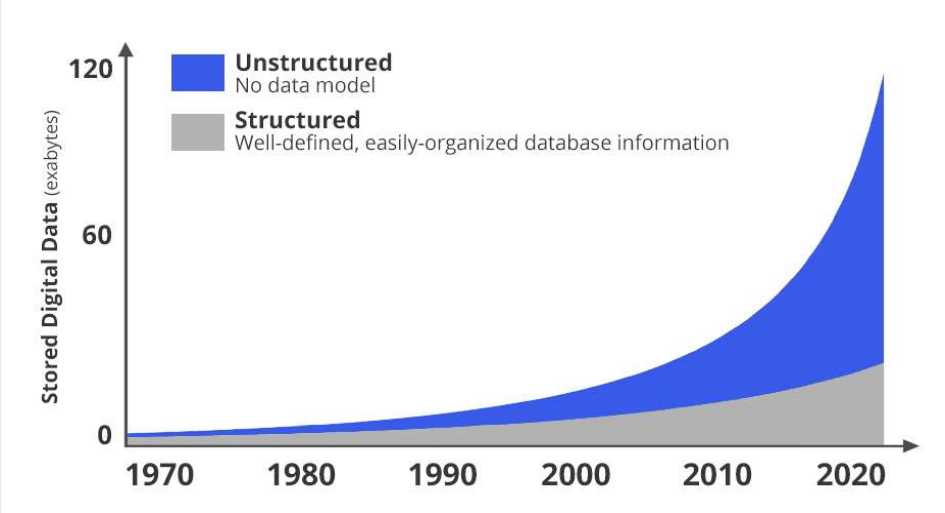
- Ethics
    1. Ethics & data science
    2. Ethics & machine learning
        - NB: some of this could be triggering (and it is certainly vulgar)
- Brief review for pset 9

Why should we focus on ethics?

- First, why should anyone?
- Second, why should researchers?
    - Researchers contribute to **human knowledge**

Why should we focus on ethics?

- First, why should anyone?
- Second, why should researchers?
  - Researchers contribute to **human knowledge**
  - This is power, and with it comes responsibility
- Third, why should data scientists in particular?

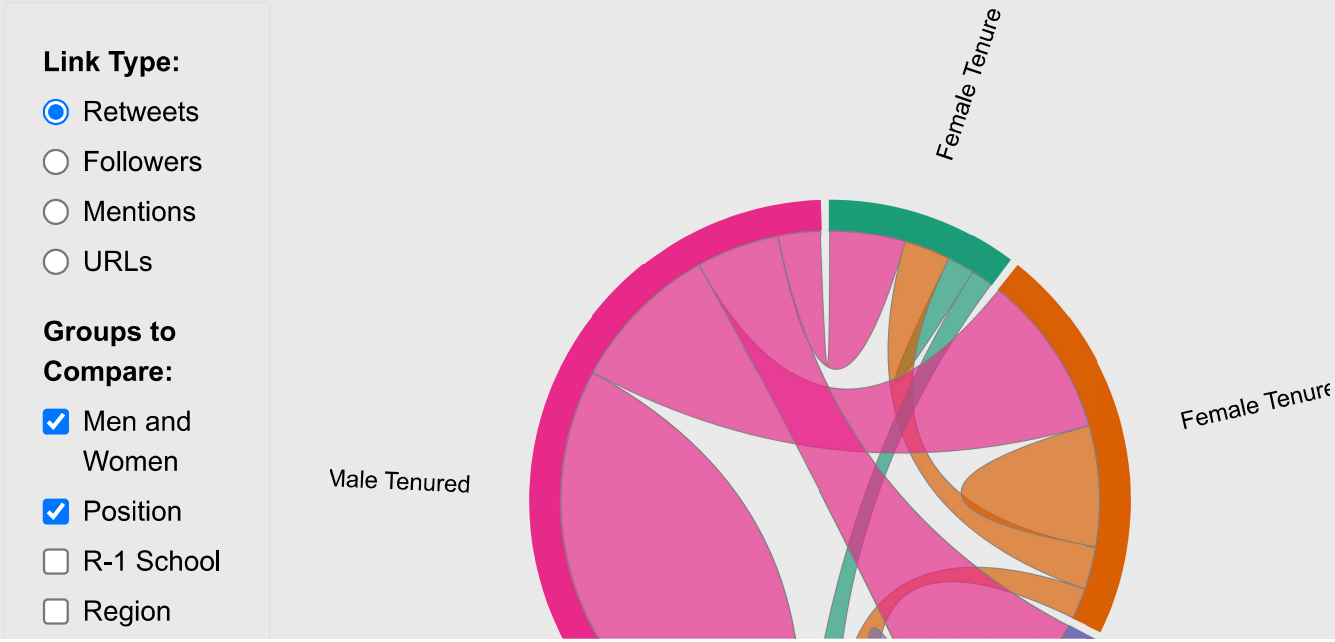


> [J]ust as the invention of the telescope revolutionized the study of the heavens, so too by **rendering the unmeasurable measurable**, the technological revolution in mobile, Web, and Internet communications has the potential to revolutionize our understanding of ourselves and how we interact.
>
> -- Duncan Watts (2011, p. 266)

Readymade Data

- Data generated by human behaviors **without their knowledge**
  - Data Science is the study of **readymade** data
- An example:

#polisci Twitter Link Flows



Readymade Data

- Twitter accounts are public
  - If anyone can read this tweet, why can't I study it?
  - "They should know better."

Readymade Data

- Privacy as a human right
    1. Right to solitude
    2. Right to intimacy
    3. Right to anonymity
    4. Right to reserve

    Frank T. McAndrew (2019) "On the Nature of Privacy"

- Human capacity vs technological capacity for change
- "Should know" is not grounds for an ethical argument

How is ethics practiced in data science today?

How is ethics practiced in data science today?

- In academia
  - Primarily **rules-based** approach
- Elsewhere
  - **Ad hoc** approach
- Increasing push for a **Principles-based** approach

Rules-Based Approach

- IRBs + federal law (in the U.S.)
- Limitations
  - Slow to change
  - Hard to explain

Ad Hoc Approach

- People thinking about ethics independently
- Limitations
  - No community feedback
  - Hard to do!

Principles-Based Approach

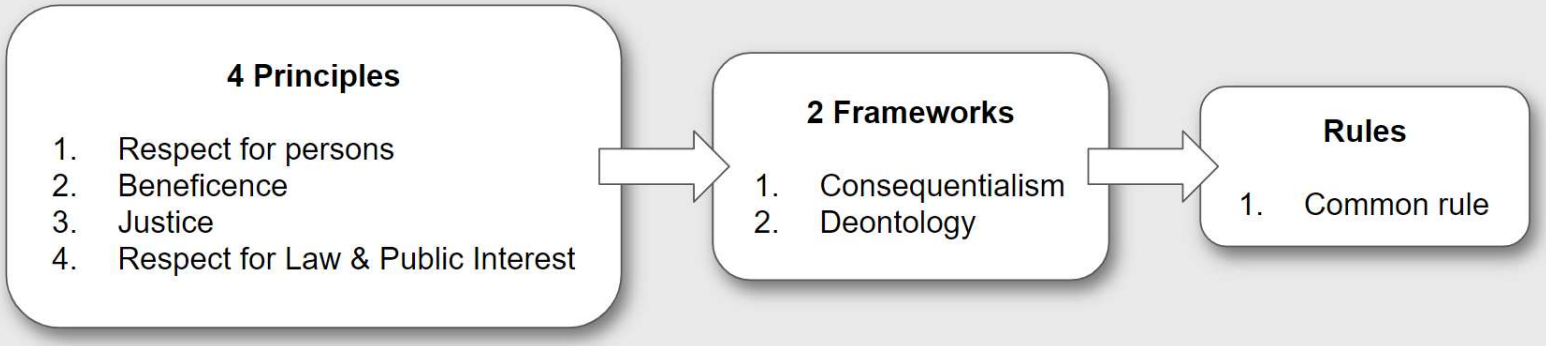
- Implicit in both **rules-based** and **ad hoc** approaches
- Puts the principles first
    - Principles developed by ethics community

Principles-Based Approach

1. Respect for persons
    - People's autonomy should be respected
2. Beneficence
    - Define / measure costs and benefits
    - Evaluate whether benefits outweight costs
3. Justice
    - Distribution of costs and benefits
4. Respect for Law & Public Interest
    - Compliance with laws, terms of service
    - Transparency about methods, procedures to ensure accountability

Principles-Based Approach

- In practice, ensuring all 4 principles is **very difficult**
  - I.e., sharing replication materials creates tension between #4 (transparency) and #1 & #2 (respect for persons and costs)
- Ethical Frameworks
  1. Deontology (focus on means)
  2. Consequentialism (focus on ends)



The Importance of Explanation

- Balancing principles with Ethical Frameworks requires explanation
  - Elicits feedback → new **perspectives**
  - Consistent with Principle #4: transparency & accountability
- Prevents **dogmatism** & **extremism**
  - Extreme consequentialism: organ donor example
  - Extreme deontology: public safety example

Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

1. Informed Consent

- **Simple Idea:** Informed consent from all participants

Counter Example

The Mark of a Criminal Record[1]

Devah Pager
Northwestern University

With over 2 million individuals currently incarcerated, and over half a million prisoners released each year, the large and growing number of men being processed through the criminal justice system raises important questions about the consequences of this massive institutional intervention. This article focuses on the consequences of incarceration for the employment outcomes of black and white job seekers. The present study adopts an experimental audit approach—in which matched pairs of individuals applied for real entry-level jobs—to formally test the degree to which a criminal record affects subsequent employment opportunities. The findings of this study reveal an important, and much underrecognized, mechanism of stratification. A criminal record presents a major barrier to employment, with important implications for racial disparities.

Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

1. Informed Consent

- **Simple Idea:** Informed consent from all participants
- **Principles-Based Defense**:
  - Limited harm to employers
  - Great social benefit
  - Lack of alternative methods
  - Consistent with setting norms

Counter Example

The Mark of a Criminal Record[1]

Devah Pager
Northwestern University

With over 2 million individuals currently incarcerated, and over half a million prisoners released each year, the large and growing number of men being processed through the criminal justice system raises important questions about the consequences of this massive institutional intervention. This article focuses on the consequences of incarceration for the employment outcomes of black and white job seekers. The present study adopts an experimental audit approach—in which matched pairs of individuals applied for real entry-level jobs—to formally test the degree to which a criminal record affects subsequent employment opportunities. The findings of this study reveal an important, and much underrecognized, mechanism of stratification. A criminal record presents a major barrier to employment, with important implications for racial disparities.

Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

1. Informed Consent

- ~~**Simple Idea:** Informed consent from all participants~~
- **Improved Idea:** Some form of consent for most research

Counter Example

The Mark of a Criminal Record[1]

Devah Pager
Northwestern University

With over 2 million individuals currently incarcerated, and over half a million prisoners released each year, the large and growing number of men being processed through the criminal justice system raises important questions about the consequences of this massive institutional intervention. This article focuses on the consequences of incarceration for the employment outcomes of black and white job seekers. The present study adopts an experimental audit approach—in which matched pairs of individuals applied for real entry-level jobs—to formally test the degree to which a criminal record affects subsequent employment opportunities. The findings of this study reveal an important, and much underrecognized, mechanism of stratification. A criminal record presents a major barrier to employment, with important implications for racial disparities.

Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

2. Informational Risk

   1. Economic
   2. Social
   3. Psychological
   4. Criminal

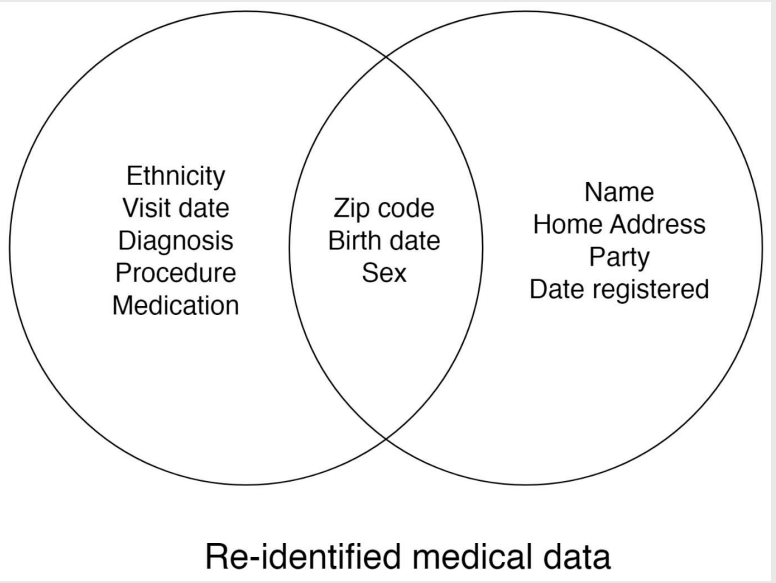
Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

2. Informational Risk

- **Simple Idea:** Data can be anonymized

Counter Example



- Sweeney (2002)

Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

2. Informational Risk

- **Simple Idea:** Data can be anonymized
- **Simple Idea:** We can tell which data is sensitive

Counter Example

RYAN SINGEL SECURITY 12.17.09 04:29 PM

NETFLIX SPILLED YOUR BROKEBACK MOUNTAIN SECRET, LAWSUIT CLAIMS

"[M]ovie and rating data contains information of a more highly personal and sensitive nature [sic]. The member's movie data exposes a Netflix member's personal interest and/or struggles with various highly personal issues, including sexuality, mental illness, recovery from alcoholism, and victimization from incest, physical abuse, domestic violence, adultery, and rape." *(Singel, 2009)*

Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

2. Informational Risk

- ~~**Simple Idea:** Data can be anonymized~~
- ~~**Simple Idea:** We can tell which data is sensitive~~
- **Improved Idea:** All data are potentially identifiable & sensitive

- **Data Protection Plans** can minimize Informational Risk
- "Five Safes" *(Desai et al 2016)*
    1. Safe Projects
    2. Safe People
    3. Safe Data
    4. Safe Settings
    5. Safe Output

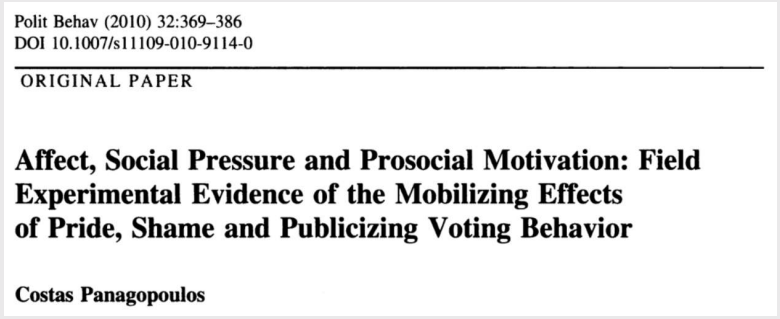
Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

3. Privacy

Simple Idea: Privacy is defined by public/private dichotomy

Counter Example



Polit Behav (2010) 32:369–386
DOI 10.1007/s11109-010-9114-0

ORIGINAL PAPER

Affect, Social Pressure and Prosocial Motivation: Field Experimental Evidence of the Mobilizing Effects of Pride, Shame and Publicizing Voting Behavior

Costas Panagopoulos

Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

3. Privacy

Simple Idea: Privacy is defined by public/private dichotomy

Counter Example

WHO VOTES IS PUBLIC INFORMATION!

Dear registered voter:

On November 6, 2007, an election to select local leaders will be held in Ely, IA.

As a registered voter, you are eligible to vote in this election. We urge you to exercise your civic duty and vote on November 6th.

We also remind you that who votes is a matter of public record.

To promote participation in the election, we will obtain a complete list of registered voters who cast ballots on Election Day from local election officials. Shortly after the November 2007 election, we will publish in the local newspaper a complete list of all Ely registered voters who did not vote.

The names of those who took the time to vote will not appear on this list.

DO YOUR CIVIC DUTY! VOTE ON ELECTION DAY!

Applying Principles is Hard

- Simple idea + counter example(s) → improved idea

3. Privacy

~~**Simple Idea:** Privacy is defined by public/private dichotomy~~

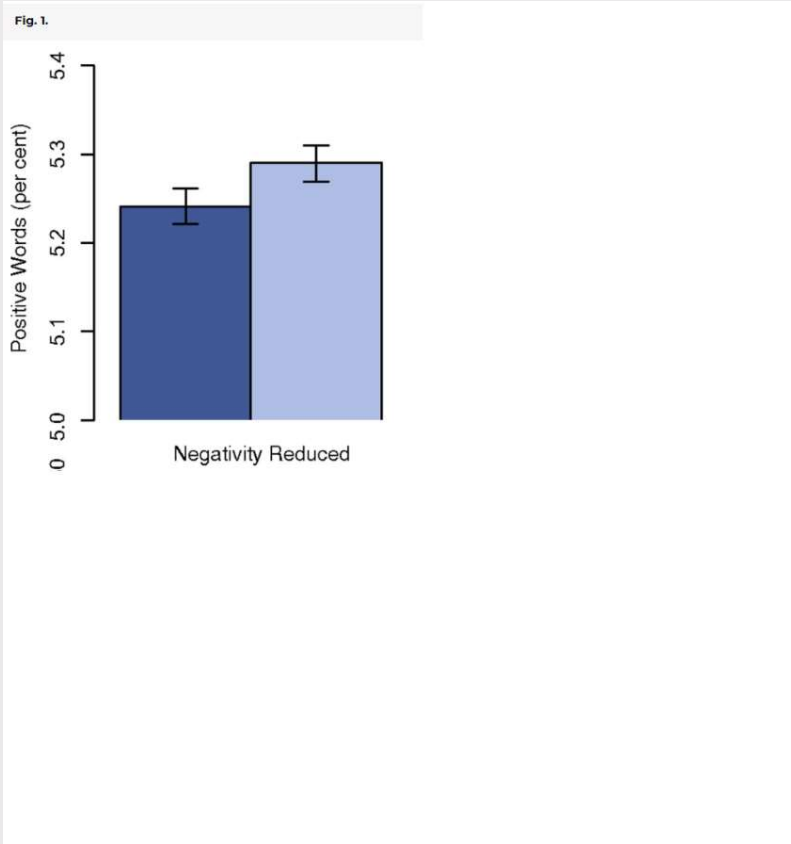
Improved Idea: Contextual integrity *(Nissenbaum)*

Contextual Information Norms

- Actors
- Attributes
- Transmissions

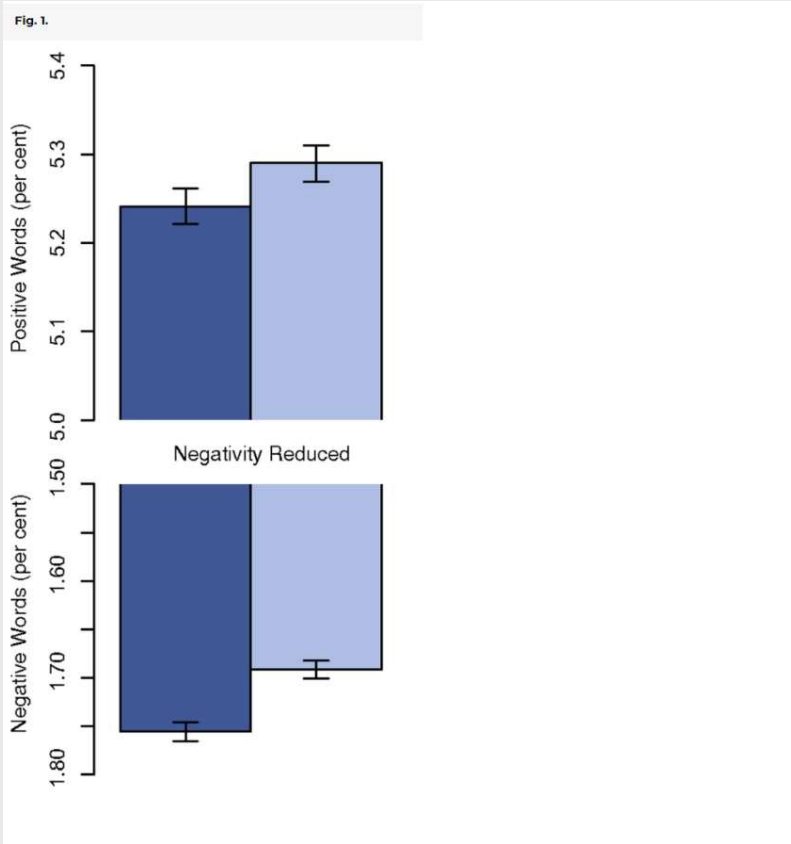
Some examples

- From academia
  1. Emotional Contagion (Kramer et al, 2014)



Some examples

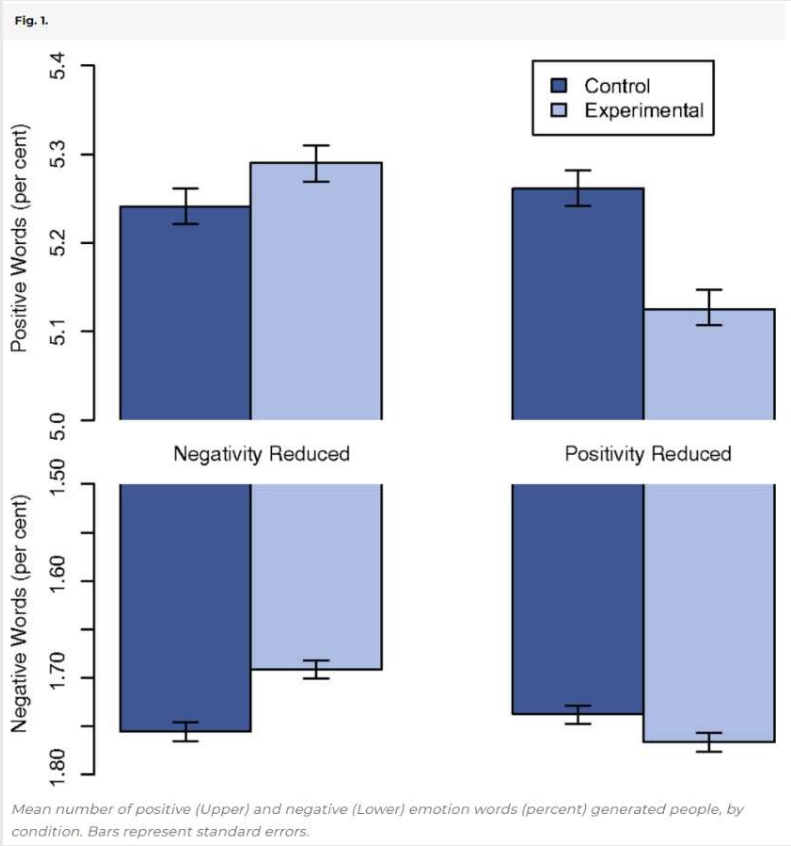
- From academia
  1. Emotional Contagion (Kramer et al, 2014)



Some examples

- From academia

1. Emotional Contagion (Kramer et al, 2014)



Some examples

- From academia

  2. Tastes, Ties & Time (Lewis et al, 2008)

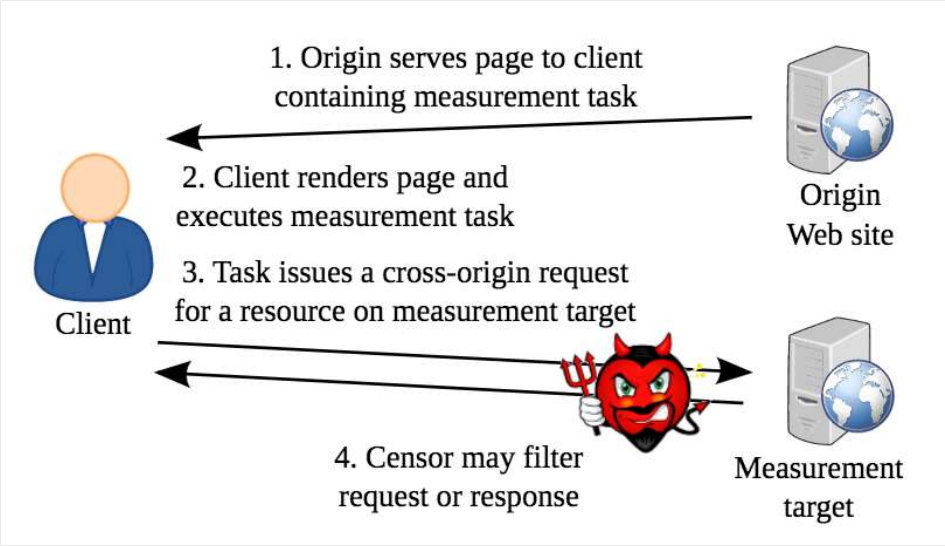
Table 5. Taste preferences of students

	Movies	Music	Books
Dominant form	Title	Artist	Author/title
# Respondents	1086	1107	1076
Mean # tastes listed	9.775	14.771	6.619
S.D. # tastes listed	7.456	15.563	4.576
Min # tastes listed	1	1	1
Max # tastes listed	63	175	34
# Unique taste listings	1927	3451	1613
Most popular (N)	The Lord of the Rings (144)	The Beatles (250)	J.K. Rowling (290)
2nd	Wedding Crashers (131)	Coldplay (238)	F. Scott Fitzgerald (167)
3rd	Star Wars (119)	Dave Matthews Band (159)	Jane Austen (142)
4th	Gladiator (116)	Green Day (143)	J.D. Salinger (137)
5th	Fight Club (112)	Jack Johnson (140)	Dan Brown (120)

Some examples

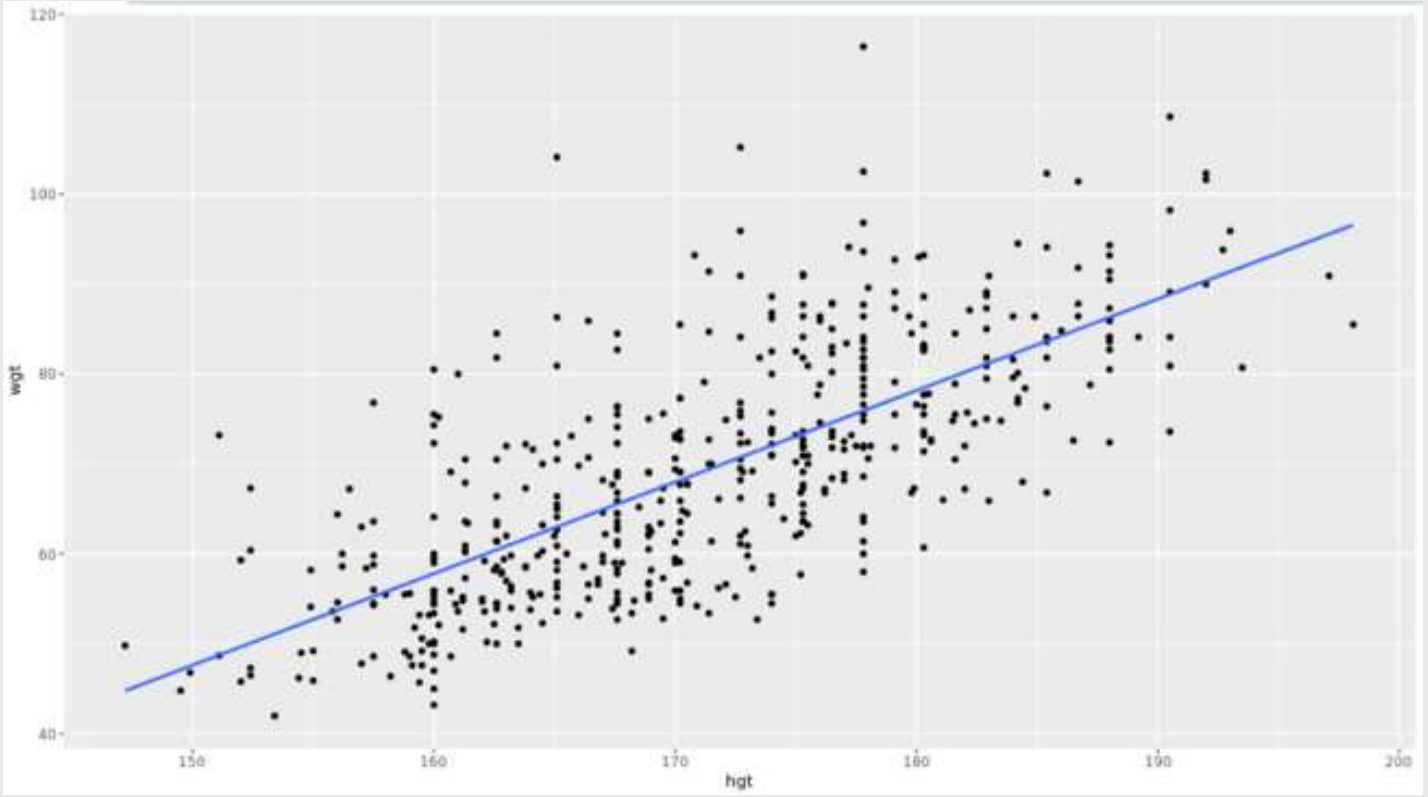
- From academia

  3. Encore (Burnett & Feamster, 2015)



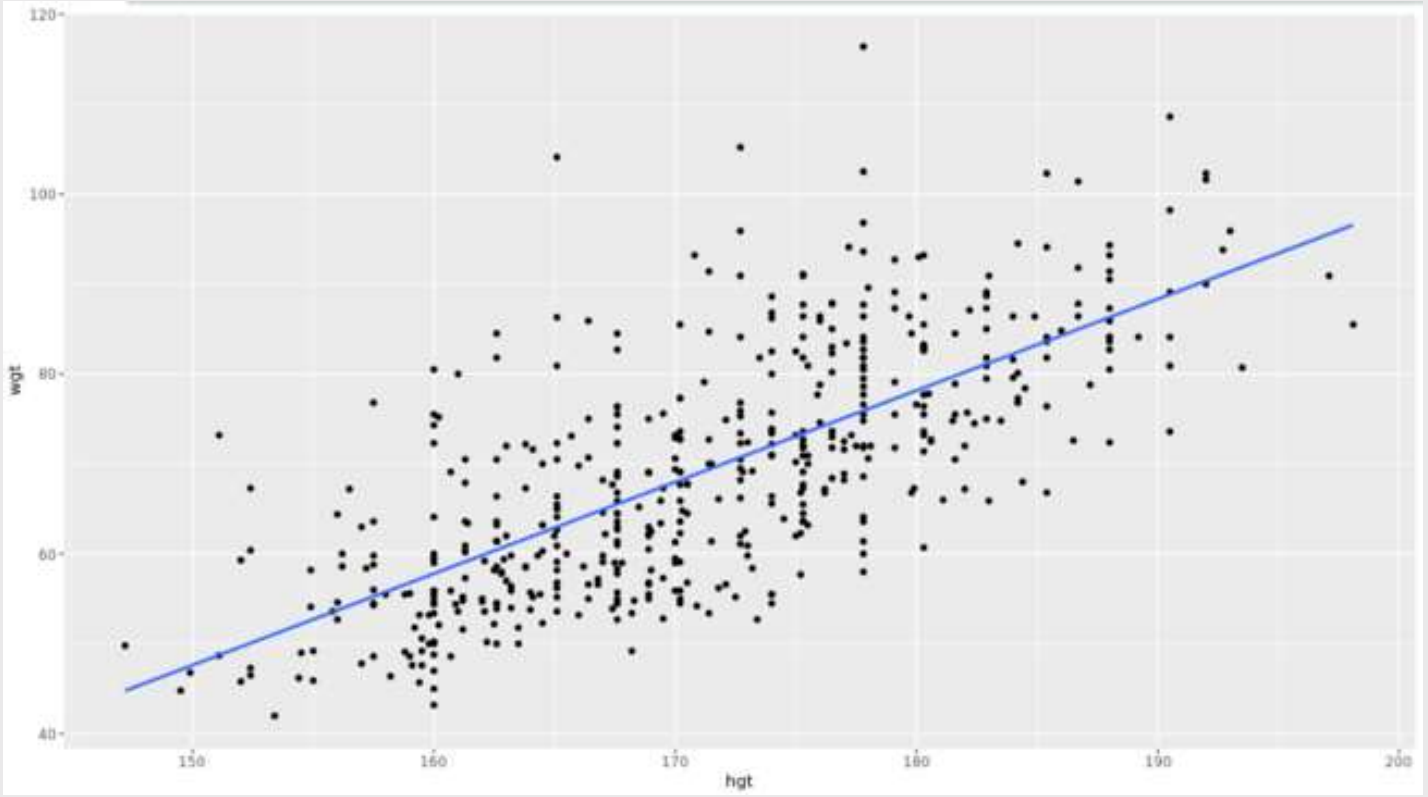
Outside Academia

- "Machine Learning" + prediction
  - "Learn" patterns from A, then use to predict values for B



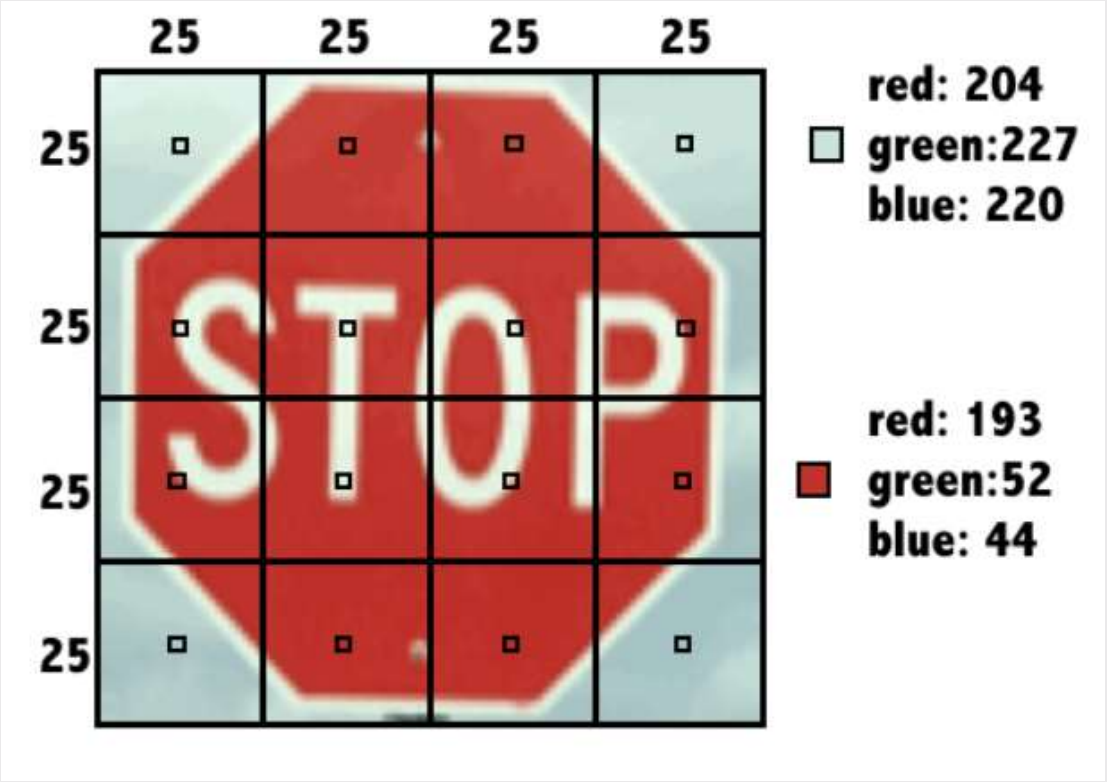
Outside Academia

- "Machine Learning" + prediction
  - Linear regression does this



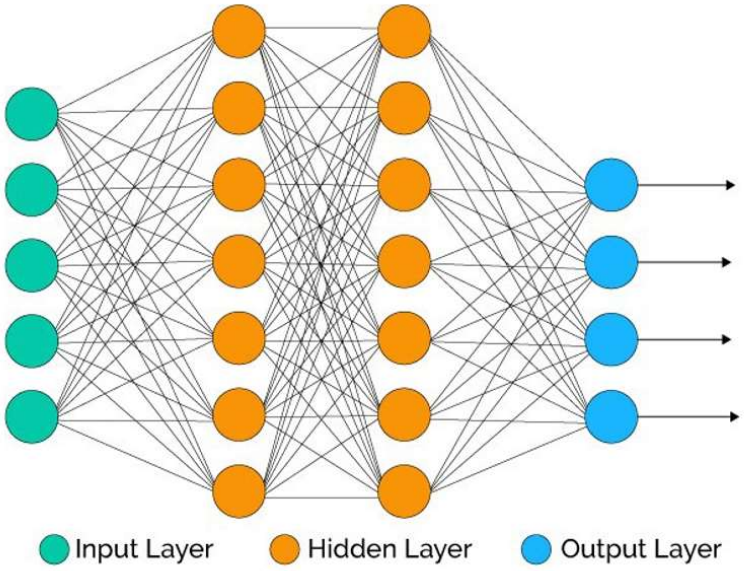
Outside Academia

- "Machine Learning" + prediction
  - But the intuition maps to fancier examples



Outside Academia

- "Machine Learning" + prediction
  - Make a prediction → Evaluate → Adjust
  - Repeat as many times as it takes!

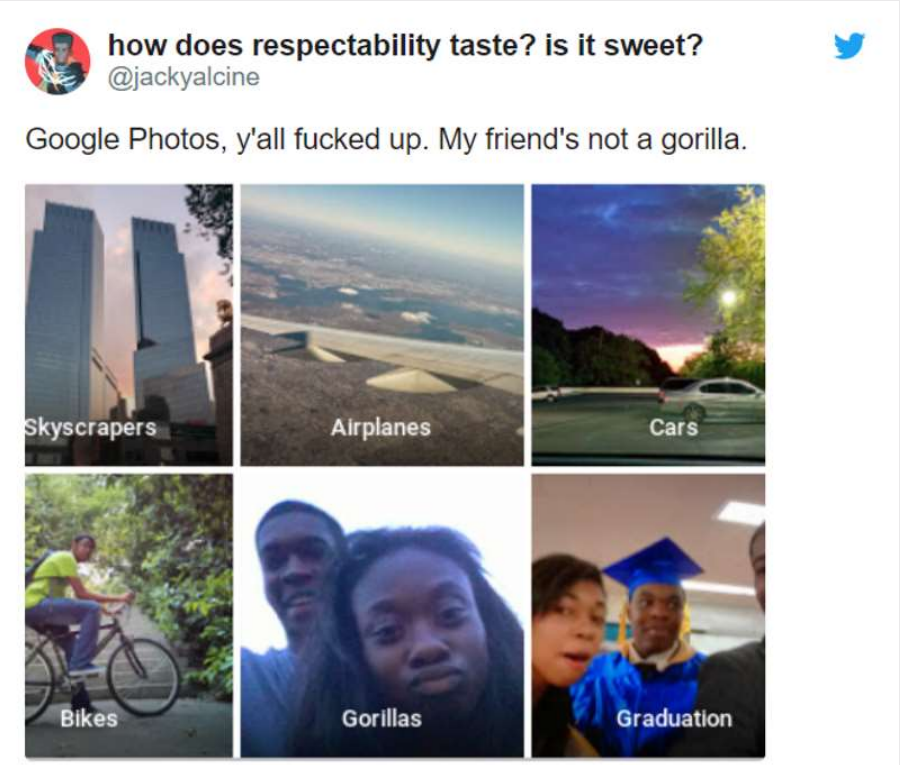


Outside Academia

- "Evaluate" is an ethical concept
- ...that we delegate to an algorithm
- What happens if the algorithm's evaluation is disconnected from ethics?

Outside Academia

- "Evaluate" is an ethical concept
- ...that we delegate to an algorithm
- What happens if the algorithm's evaluation is disconnected from ethics?

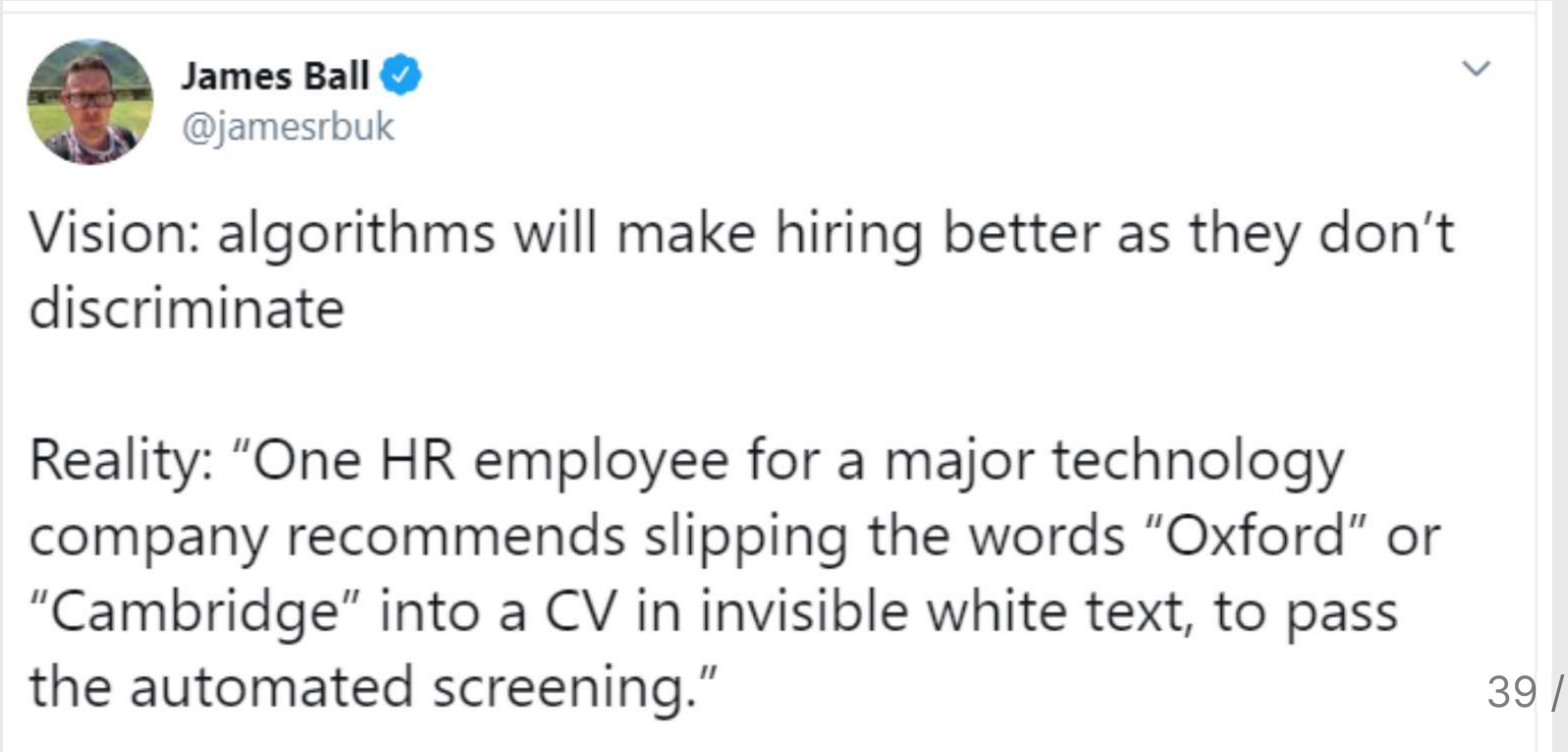


Outside Academia

- "Evaluate" is an ethical concept
- ...that we delegate to an algorithm
- What happens if the algorithm's evaluation is disconnected from ethics?

Outside Academia

- "Evaluate" is an ethical concept
- ...that we delegate to an algorithm
- What happens if the algorithm's evaluation is disconnected from ethics?



Outside Academia

- "Evaluate" is an ethical concept
- ...that we delegate to an algorithm
- What happens if the algorithm's evaluation is disconnected from ethics?

And Within Academia

Os Keyes
@farbandish

TIL there's a shitton of computer vision literature in 2017-2018 that COINCIDENTALLY tries to build facial recognition for Uyghur people. How. Curious.

WILEY

FOCUS ARTICLE

Facial feature discovery for ethnicity recognition

Cunrui Wang[1,2] | Qingling Zhang[2] | Wanquan Liu[3] | Yu Liu[1] | Lixin Miao[1]

[1]Dalian Key Lab of Digital Technology for National Culture & Institute of System Science, Northeastern University, Dalian Nationalities University, Dalian, China

[2]Institute of System Science, Northeastern University, Shenyang, China

[3]Department of Computing, Curtin University, Perth, Western Australia, Australia

Correspondence
Cunrui Wang, Dalian Key Lab of Digital Technology for National Culture & Institute of System Science, Northeastern University, Dalian Nationalities University, China.
Email: cunruiwang@qq.com

Funding information
National Natural Science Foundation of China, Grant Number: 61562093, 61772575; China

The salient facial feature discovery is one of the important research tasks in ethnical group face recognition. In this paper, we first construct an ethnical group face dataset including Chinese Uyghur, Tibetan, and Korean. Then, we show that the effective sparse sensing approach to general face recognition is not working anymore for ethnical group facial recognition if the features based on whole face image are used. This is partially due to a fact that each ethnical group may have its own characteristics manifesting only in specified face regions. Therefore, we will analyze the particularity of three ethnical groups and aim to find the common characterizations in some local regions for the three ethnical groups. For this purpose, we first use the facial landmark detector STASM to find some important landmarks in a face image, then, we use the well-known data mining technique, the mRMR algorithm, to select the salient geometric length features based on all possible lines connected by any two landmarks. Second, based on these selected salient features, we construct three "T" regions in a face image for ethnical feature representation and

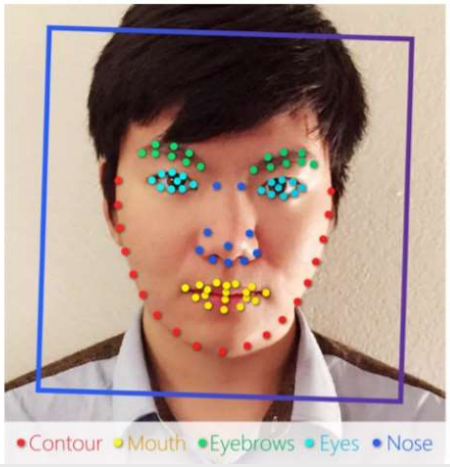
And Within Academia

Deep neural networks are more accurate than humans at detecting sexual orientation from facial images

Yilun Wang, Michal Kosinski

Graduate School of Business, Stanford University, Stanford, CA94305, USA

michalk@stanford.edu



Excitement Blinds Us

"Identifying links between facial features and psychological traits by employing methodology similar to the one used here could boost our understanding of the origins and nature of a broad range of psychological traits, preferences, and psychological processes" -- Wang & Kosinski, 2018

Foppington's Law [edit]

> Once bigotry or self-loathing permeate a given community, it is only a matter of time before deep metaphysical significance is assigned to the shape of human skulls.
>
> —Natalie Wynn

Excitement Blinds Us

"Identifying links between facial features and psychological traits by employing methodology similar to the one used here could boost our understanding of the origins and nature of a broad range of psychological traits, preferences, and psychological processes" -- Wang & Kosinski, 2018

Quick tip on pset 9

- Do Trump's joyful or sad tweets get more engagement?
- Need to:
    1. Calculate number of words by sentiment and by tweet
    2. `spread()` to get sentiments in their own columns
    3. `mutate()` to calculate difference
    4. `lm()` to run regression (or visualize or whatever)

Quick tip on pset 9

```
require(tidyverse)
require(tidytext)

tweet_words <-
read_rds(file="https://github.com/jbisbee1/DS1000_S2023/blob/main/Lectu
raw=true")
nrc <- get_sentiments("nrc")

# Merging with inner join
tweet_sentiment <- tweet_words %>%
    inner_join(nrc, by = "word")
```

Quick tip on pset 9

```
# Summing retweets
toplotRT_sum <- tweet_sentiment %>%
  filter(sentiment %in% c('sadness','joy')) %>% # Filter to the
sentiments you are interested in
  count(document,retweets,sentiment) %>% # Count number of words
(keep retweets for later)
  spread(sentiment,n,fill = 0) %>% # Spread the sentiments to own
columns
  mutate(sentiment = ifelse(joy > sadness,'Joyful', # calcualte net
sentiment
                            ifelse(sadness > joy,'Sad','Neutral')))
%>%
  group_by(sentiment) %>%
  summarise(totTweets = n(),
            retweets = sum(retweets,na.rm=T)) %>%
  ungroup()
```

Quick tip on pset 9

```
# For regression
sentimentAnalysis <- tweet_sentiment %>%
  filter(sentiment %in% c('sadness','joy')) %>%
  count(document,retweets,Tweeting.year,sentiment) %>%
  spread(sentiment,n,fill = 0) %>%
  mutate(sentiment = ifelse(joy > sadness,'Joyful',
                            ifelse(sadness > joy,'Sad','Neutral')),
         net_joy = joy - sadness,
         year = as.numeric(as.character(Tweeting.year))) %>%
  ungroup()
```

Quick tip on pset 9

```
summary(lm(log(retweets+1) ~ net_joy,data = sentimentAnalysis))
```

```
##
## Call:
## lm(formula = log(retweets + 1) ~ net_joy, data =
sentimentAnalysis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8825 -2.5852 -0.1687  2.7888  7.0068
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.41751    0.02042  314.35   <2e-16 ***
## net_joy     -0.46497    0.01512  -30.75   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.838 on 19833 degrees of freedom
## Multiple R-squared:  0.0455,    Adjusted R-squared:  0.04545
## F-statistic: 945.3 on 1 and 19833 DF,  p-value: < 2.2e-16
```