

Strategic Data Science (SDS)

Web data

Karl Ho

School of Economic, Political and Policy Sciences

University of Texas at Dallas

Web data

Web data

Accumulation of data begins with transaction and interaction between humans. The advent of new internet technology transcends the data accumulation using high speed computation, large storage and caching systems.

Web data

Web data

Big data exploration starts with information on people or on a group of people. The internet becomes the open platform and media to facilitate a rich source of human generated information. It can also act as pointers to new sources of information, allowing a pivot from one perspective to another, from one kind of query to another.

Web data

How do we take advantage of the web data?

Web data

How do we take advantage of the web data?

1. Purpose of web data

Web data

How do we take advantage of the web data?

1. Purpose of web data
2. Generation process of web data

Web data

How do we take advantage of the web data?

1. Purpose of web data
2. Generation process of web data
3. What is data of data?

Web data

How do we take advantage of the web data?

1. Purpose of web data
2. Generation process of web data
3. What is data of data?
4. Why social scientists need to collect web data?

Web data: Technical side

Web data: Technical side

Web scraping

- obtaining information directly from web pages

Web data: Technical side

Web scraping

- obtaining information directly from web pages

APIs (Application program interface)

- web services that allow an interaction with, and retrieval of, structured data.

Web data: Tools

Web data: Tools

1. **SAS**
2. **R**
3. **Python**
4. **Tableau**
5. **Data Mining Packages**

Web data technologies

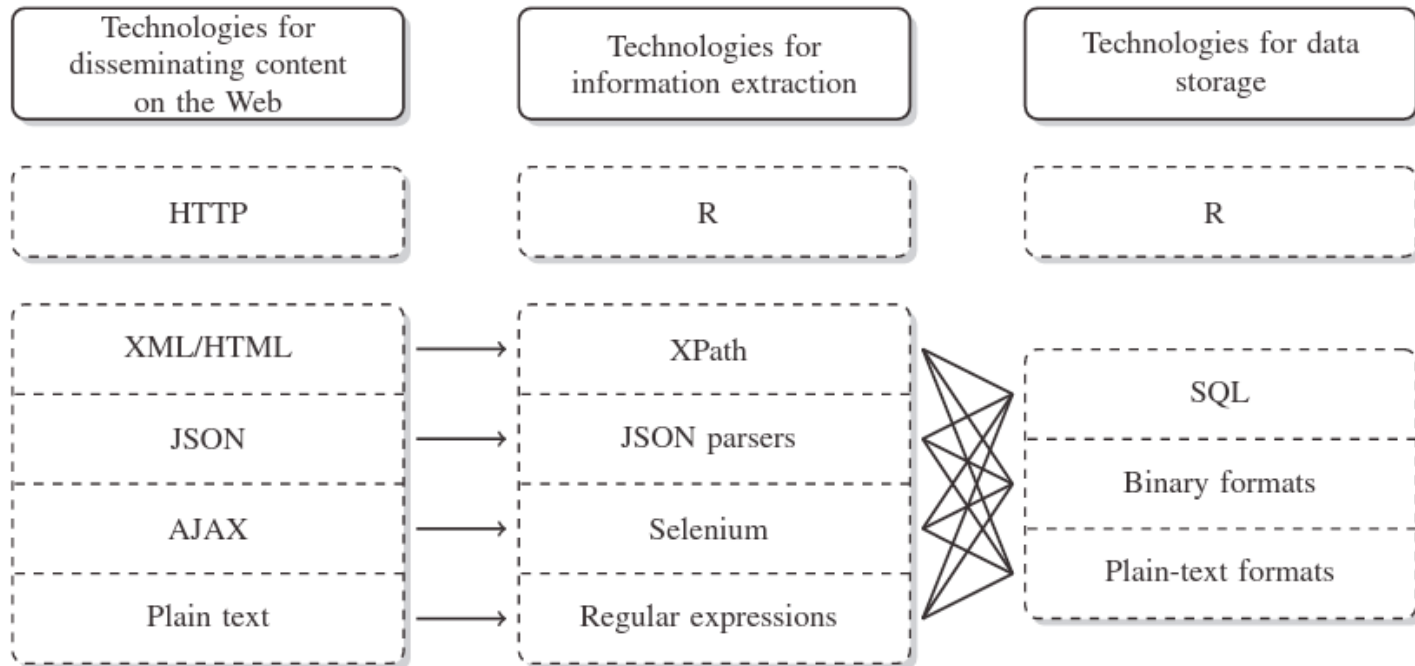


Figure 1.4 Technologies for disseminating, extracting, and storing web data

Web data: API's (data source)

Web data: API's (data source)

1. Social Media

Web data: API's (data source)

1. Social Media

1. Facebook

Web data: API's (data source)

1. Social Media

1. Facebook

2. Twitter

Web data: API's (data source)

1. Social Media

1. Facebook

2. Twitter

3. Instagram

Web data: API's (data source)

1. Social Media

1. Facebook

2. Twitter

3. Instagram

2. News websites

Web data: API's (data source)

1. Social Media

- 1. Facebook

- 2. Twitter

- 3. Instagram

2. News websites

3. Government websites

Web data: API's (data source)

1. Social Media

1. Facebook

2. Twitter

3. Instagram

2. News websites

3. Government websites

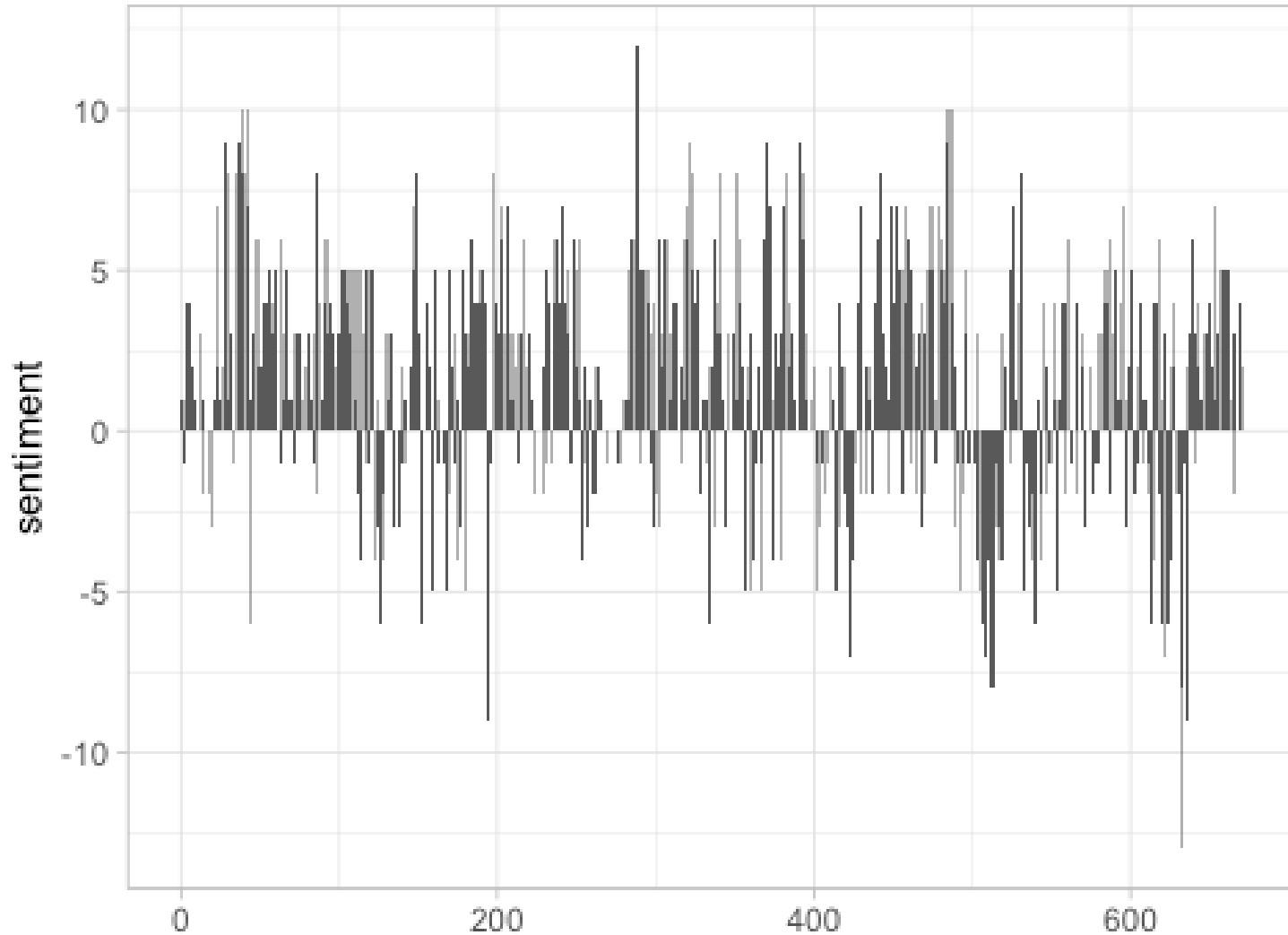
4. NGOs

Sentiment Analysis

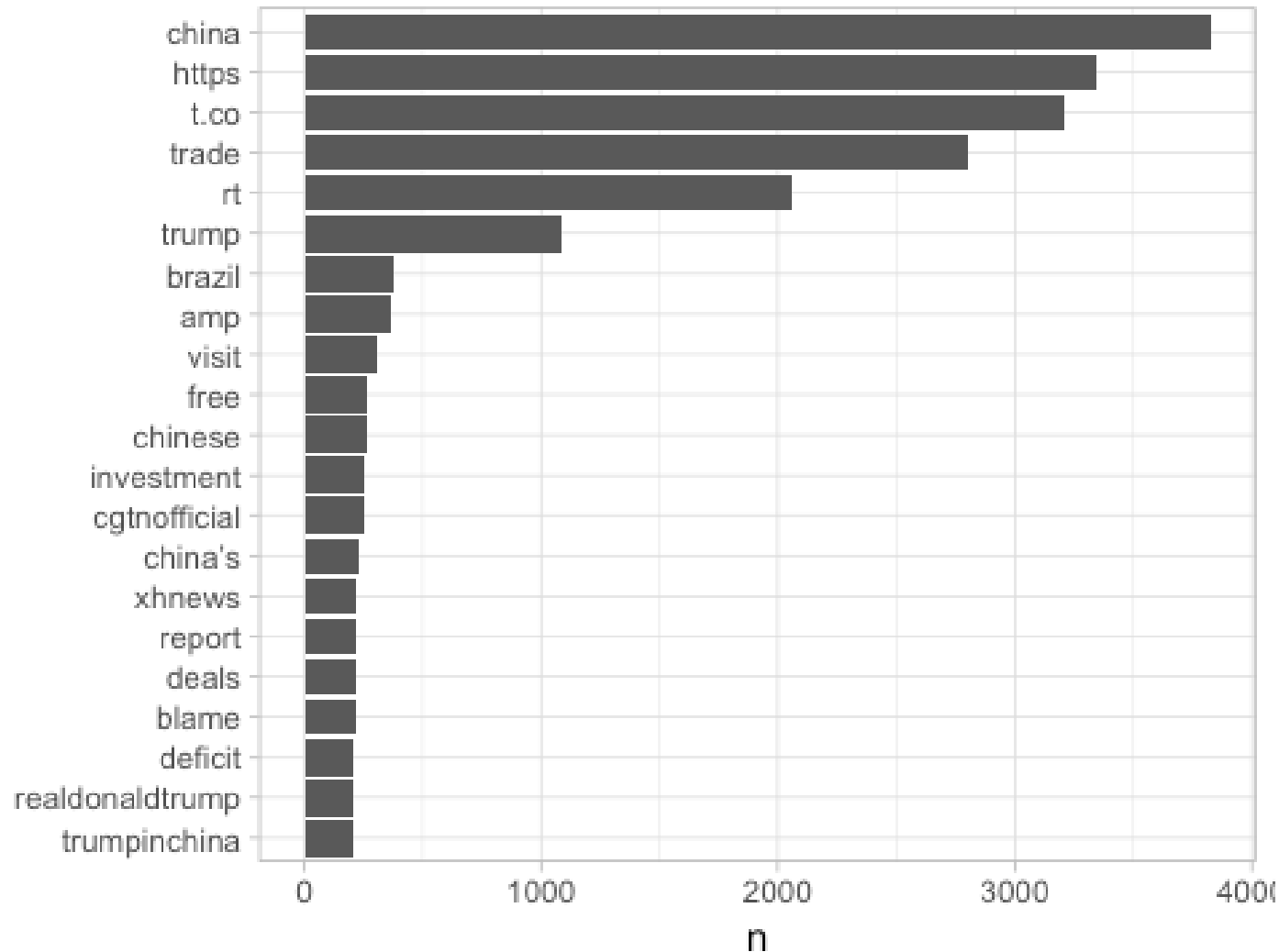
When human readers approach a text, we use our understanding of the emotional intent of words to infer whether a section of text is positive or negative, or perhaps characterized by some other more nuanced emotion like surprise or disgust.

- Silge and Robinson 2017

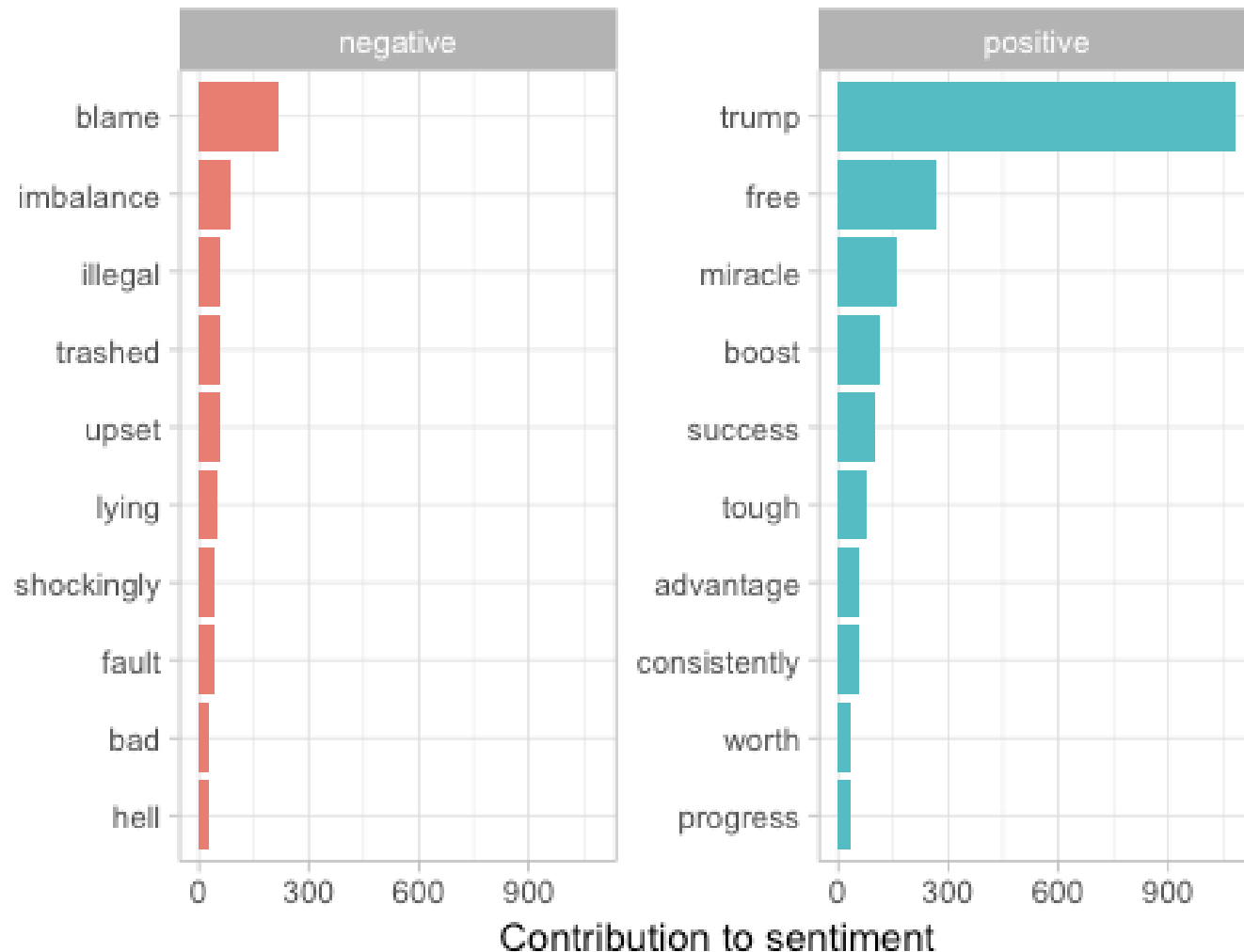
Workshop demo: sentiment analysis



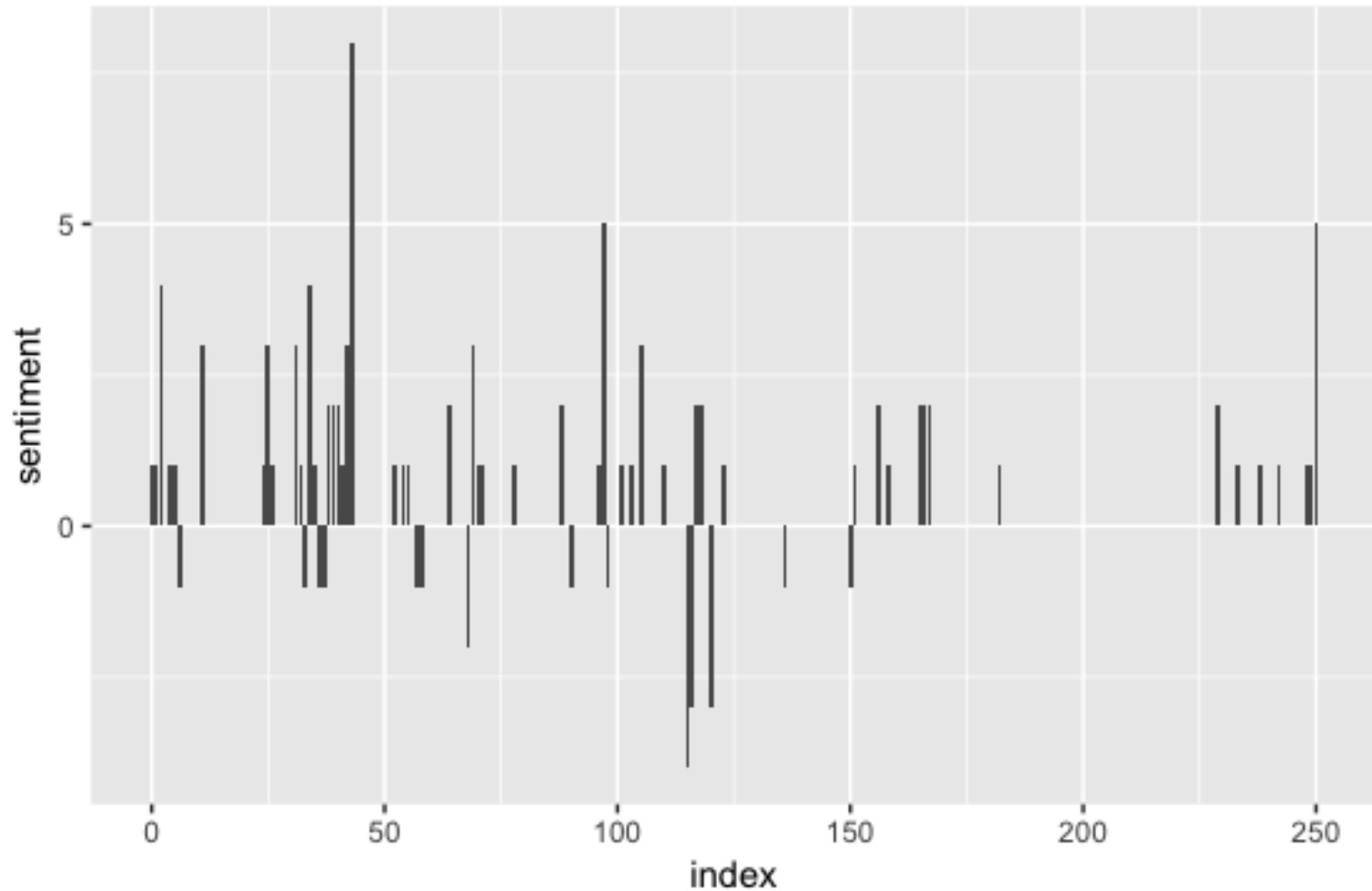
Workshop demo: sentiment analysis



Workshop demo: sentiment analysis



Social Media Impression of "Taiwan"



Social Media Impression of "Taiwan"

