Data Mining in Education

Social Media + Text

Qiang Hao

neohao@uga.edu

http://tobeneo.com

Goals

- **What is Data Mining?**
- **What tools / knowledge do you need to do Data Mining?**
- **What is the basic process of Data Mining?**

Questions Answered by Data Mining

- Can we predict whether the coming email is a spam?

UCI

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Spambase Data Set

Download: Data Folder, Data Set Description

Abstract: Classifying Email as Spam or Non-Spam

Data Set Characteristics:	Multivariate	**Number of Instances:**	4601	**Area:**	Computer
Attribute Characteristics:	Integer, Real	**Number of Attributes:**	57	**Date Donated**	1999-07-01
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	143804

Questions Answered by Data Mining

- Can we predict whether the coming email is a spam?

Questions Answered by Data Mining

- Can we predict whether the coming email is a spam?

Questions Answered by Data Mining

- What is the attitude of people on Twitter towards the presidential candidate ***Donald Trump***?

Questions Answered by Data Mining

- What is the attitude of people on Twitter towards the presidential candidate ***Donald Trump***?

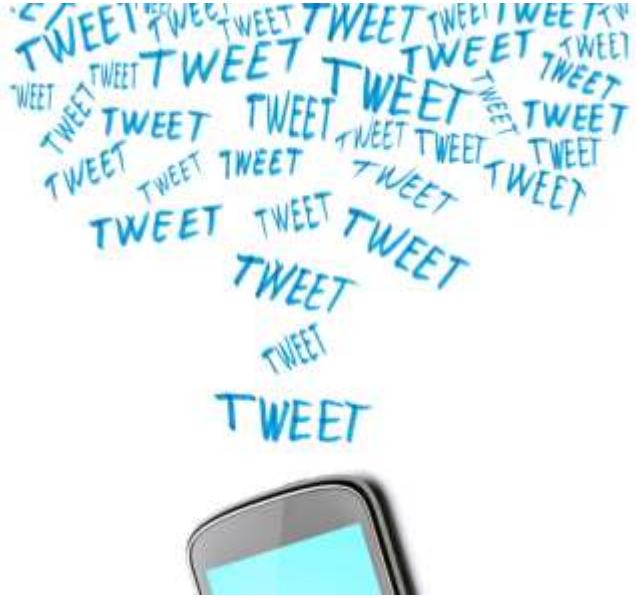
#Trump

#DonaldTrump

#GOPTrump

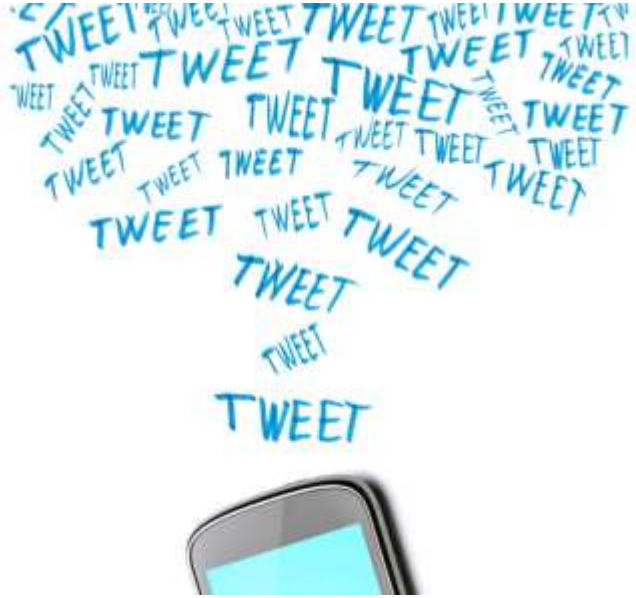
Questions Answered by Data Mining

- What is the attitude of people on Twitter towards the presidential candidate ***Donald Trump***?



Questions Answered by Data Mining

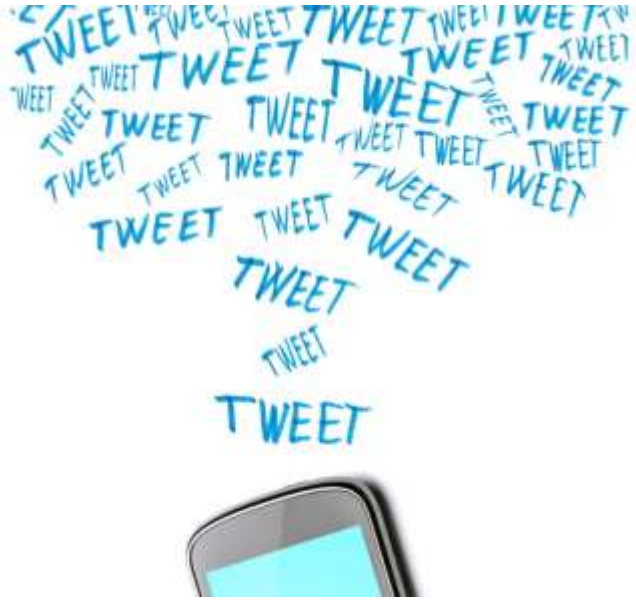
- What is the attitude of people on Twitter towards the presidential candidate ***Donald Trump***?

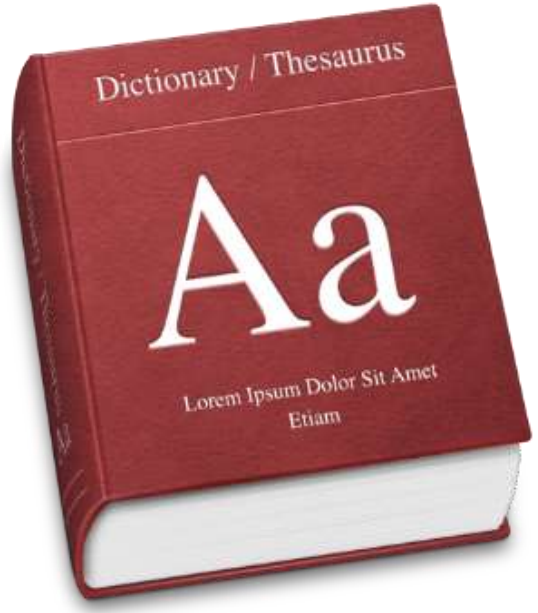


a, an, the, is, are, was, were, …

Questions Answered by Data Mining

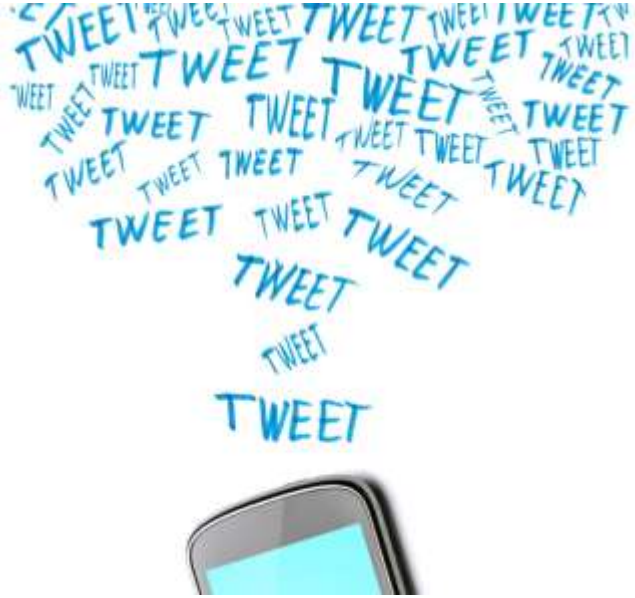
- What is the attitude of people on Twitter towards the presidential candidate ***Donald Trump***?

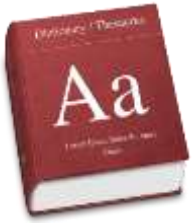




Questions Answered by Data Mining

- What is the attitude of people on Twitter towards the presidential candidate ***Donald Trump***?





Negative

Neutral

Positive

Educational Questions to Answer by Data Mining

Educational Questions to Answer by Data Mining

- What algorithm can score essays as teachers do?

Educational Questions to Answer by Data Mining

- What courses should we recommend to students based on their online activities?

Educational Questions to Answer by Data Mining

- Does the intervention improve students' lexical variety in their writing?

Educational Questions to Answer by Data Mining

- Are there different patterns in students' questions; if so, are the patterns related to their academic performance?

Educational Questions to Answer by Data Mining

- What sub-topics do students tend to cover when discussing this topic?

Educational Questions to Answer by Data Mining

- What predictor is the most important one for whether college students seek help online in their learning?

Goals

- **What is Data Mining?**

  Replicable

  Reproducible

  Automatic

Goals

- **What is Data Mining?**
- **What tools / knowledge do you need to do Data Mining?**

Tools / Knowledge

Tools / Knowledge

Carmen Reinhart

Kenneth Rogoff

Thomas Herndon

Goals

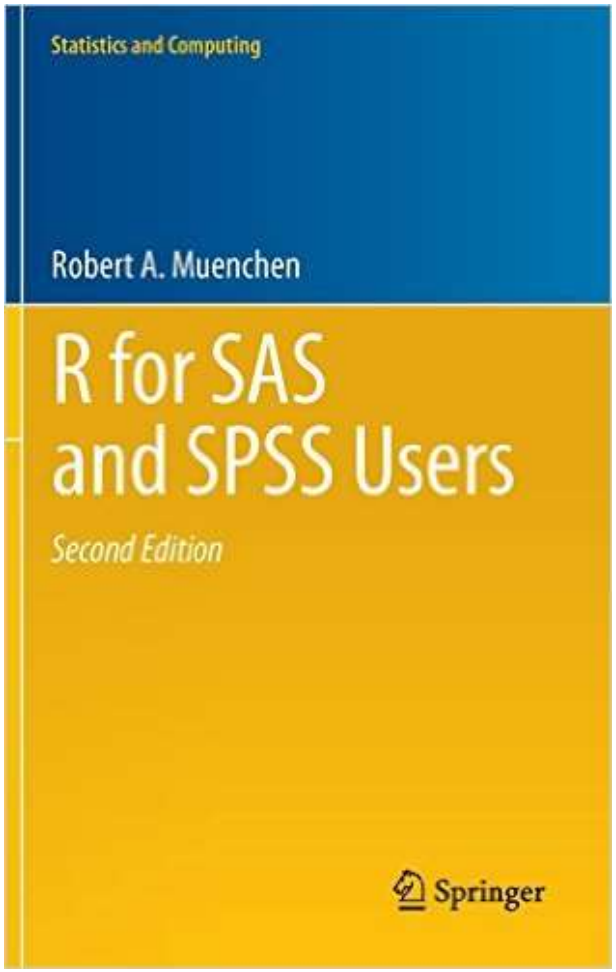
- What tools / knowledge do you need to do Data Mining?

  Expert level of knowledge in statistics

  Intermediate level of knowledge in programming

  Familiarity with R/Python

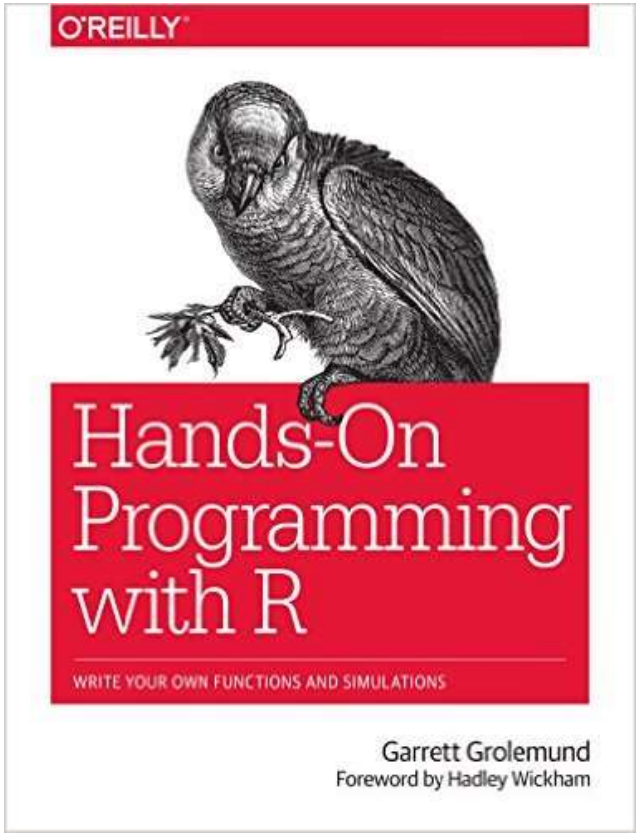
Goals



R for SAS and SPSS Users

Robert A. Muenchen

Goals



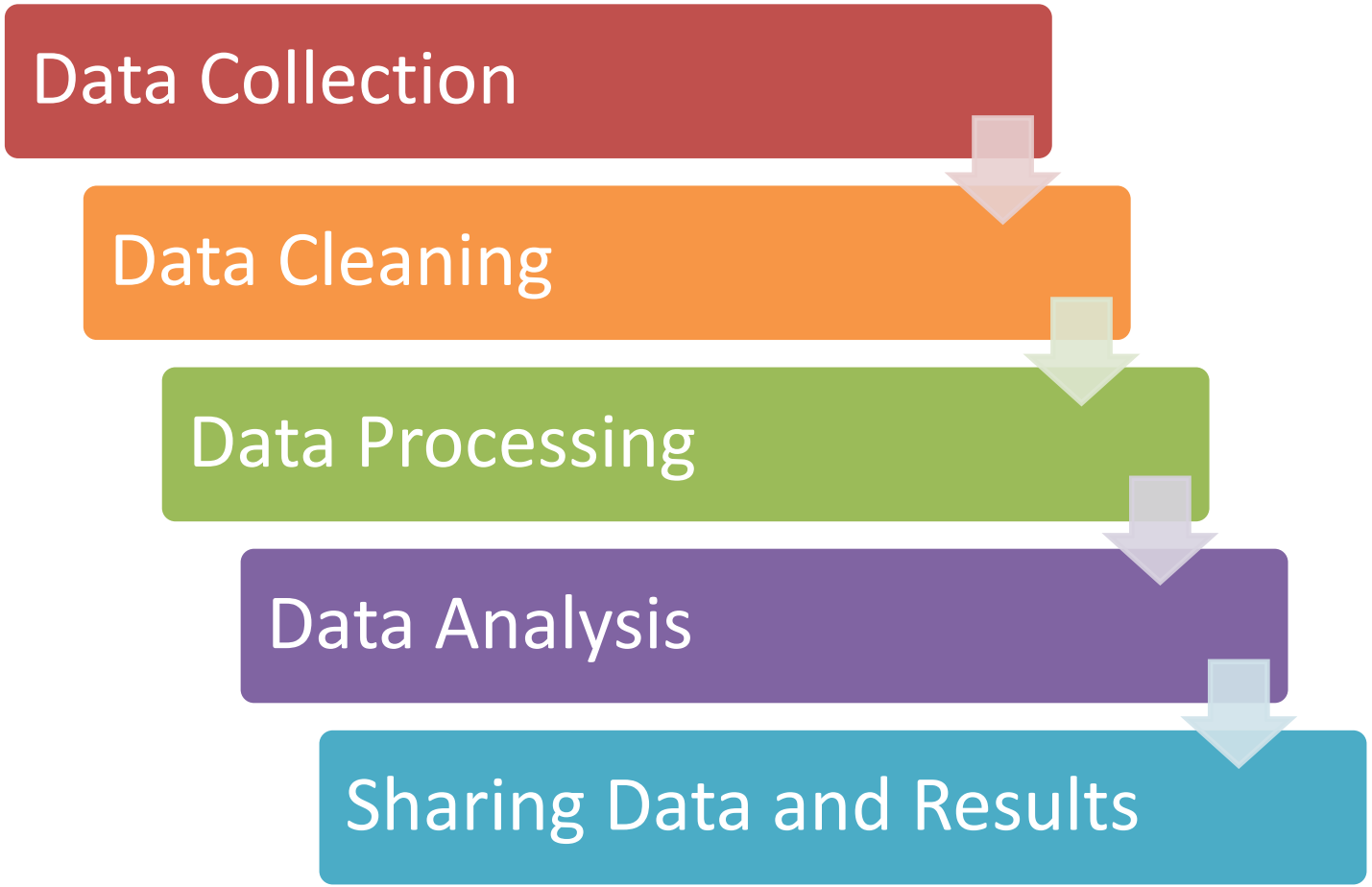
Hands-On Programming with R

Garrett Grolemund

Goals

- **What is Data Mining?**
- **What tools / knowledge do you need to do Data Mining?**
- **What is the basic process of Data Mining?**

Research Pipeline



Data Collection

Data Collection

- **XML**

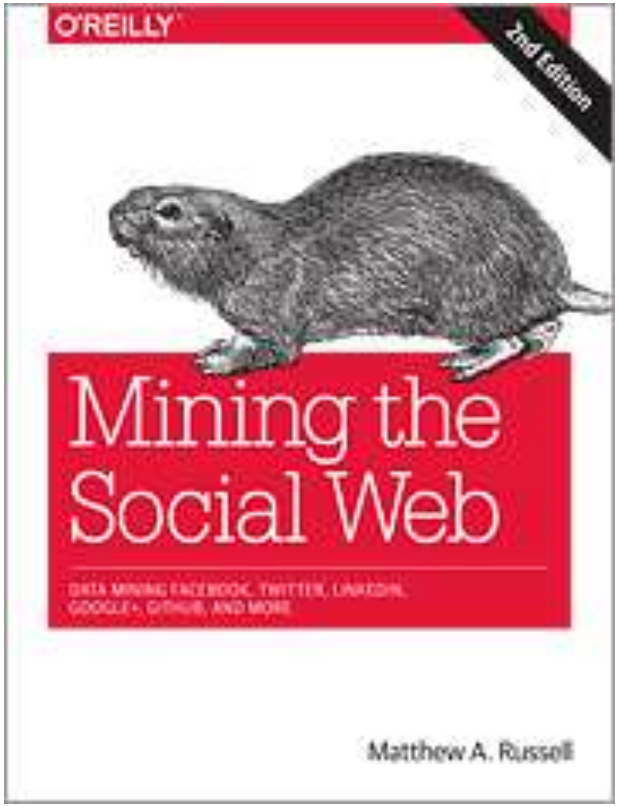
```
− <change-log type="array">
  − <change-log>
      <when type="datetime">2015-05-26T17:42:37Z</when>
      <data>ia5m23j5hbx5ms</data>
      <type>create</type>
      <anon>no</anon>
      <uid>gd6v7134AUa</uid>
    </change-log>
  </change-log>
  <folders type="array"/>
  <children type="array"/>
  <no_answer_followup>0</no_answer_followup>
```

Data Collection

- **JSON**

```
{
    hey: "guy",
    anumber: 243,
  - anobject: {
        whoa: "nuts",
      - anarray: [
            1,
            2,
            "thr<h1>ee"
        ],
        more: "stuff"
    },
    awesome: true,
    bogus: false,
    meaning: null,
    japanese: "明日がある。",
    link: http://jsonview.com,
    notLink: "http://jsonview.com is great"
}
```

Data Collection



Mining the Social Web 2nd Edition

Matthew A. Russell

Python

Data Cleaning

	text	favorited	favoriteC	replyToSN	created	truncated	replyToSI	id	replyToUI	statusSou
1	@mesterman @Ec	FALSE	0	mesterman	2015/4/15 23:52	FALSE	5.88E+17	5.88E+17	14906194	<a href="(
2	#monopolistic	FALSE	0	NA	2015/4/15 23:44	FALSE	NA	5.88E+17	NA	<a href="l
3	RT @heosat: Ar	FALSE	0	NA	2015/4/15 23:35	FALSE	NA	5.88E+17	NA	<a href="
4	RT @heosat: Ar	FALSE	0	NA	2015/4/15 23:35	FALSE	NA	5.88E+17	NA	<a href="l
5	RT @heosat: Ar	FALSE	0	NA	2015/4/15 23:35	FALSE	NA	5.88E+17	NA	<a href="
6	Another new re	FALSE	0	NA	2015/4/15 23:35	FALSE	NA	5.88E+17	NA	<a href="l
7	#Teachers shou	FALSE	0	NA	2015/4/15 23:01	FALSE	NA	5.88E+17	NA	<a href="l
8	RT @CirrusAsse	FALSE	0	NA	2015/4/15 22:44	FALSE	NA	5.88E+17	NA	<a href="
9	Teachers: get	FALSE	0	NA	2015/4/15 22:32	FALSE	NA	5.88E+17	NA	<a href="
10	How 2 Put Meta	FALSE	0	NA	2015/4/15 22:02	FALSE	NA	5.88E+17	NA	<a href="
11	RT @CanvasPenr	FALSE	0	NA	2015/4/15 21:11	FALSE	NA	5.88E+17	NA	<a href="l
12	Great tool for	FALSE	0	NA	2015/4/15 20:38	FALSE	NA	5.88E+17	NA	<a href="
13	Be the change	FALSE	0	NA	2015/4/15 20:23	FALSE	NA	5.88E+17	NA	<a href="
14	DYSLEXIC WHO,,	FALSE	0	NA	2015/4/15 20:02	FALSE	NA	5.88E+17	NA	<a href="
15	7 Cyberlearnir	FALSE	0	NA	2015/4/15 20:01	FALSE	NA	5.88E+17	NA	<a href="
16	RT @grahamlfox	FALSE	0	NA	2015/4/15 19:54	FALSE	NA	5.88E+17	NA	<a href="r
17	RT @Spencer_GG	FALSE	0	NA	2015/4/15 19:47	FALSE	NA	5.88E+17	NA	<a href="r
18	RT @bsarte: #M	FALSE	0	NA	2015/4/15 19:45	FALSE	NA	5.88E+17	NA	<a href="(
19	#GoogleClassrc	FALSE	2	NA	2015/4/15 19:43	FALSE	NA	5.88E+17	NA	<a href="
20	#MDM: Mobile d	FALSE	1	NA	2015/4/15 19:35	FALSE	NA	5.88E+17	NA	<a href="l
21	bsarte: #MDM:	FALSE	1	NA	2015/4/15 19:32	FALSE	NA	5.88E+17	NA	<a href=""
22	#MDM: Mobile d	FALSE	1	NA	2015/4/15 19:31	FALSE	NA	5.88E+17	NA	<a href="l
23	#MDM: Mobile d	FALSE	1	NA	2015/4/15 19:25	FALSE	NA	5.88E+17	NA	<a href="l
24	#MDM: Mobile d	FALSE	1	NA	2015/4/15 19:20	FALSE	NA	5.88E+17	NA	<a href="l
25	El impacto de	FALSE	0	NA	2015/4/15 19:13	FALSE	NA	5.88E+17	NA	<a href="

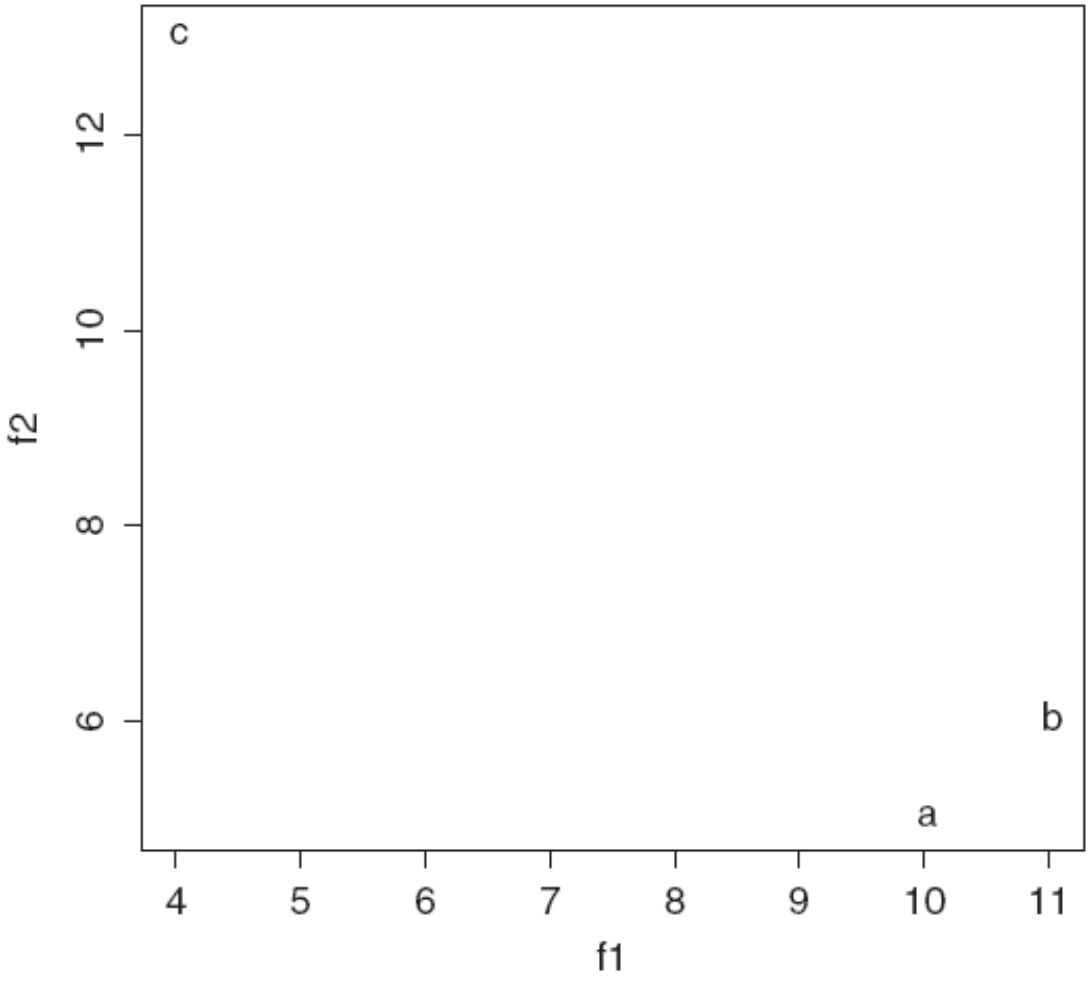
Data Processing

	apply	associatior	lecture	exciting	popular	Trump	please	stop	…
Person 1	0	0	0	0	0	1	0	0	…
Person 2	0	0	0	0	2	0	0	0	…
Person 3	0	0	0	0	0	0	0	1	…
Person 4	0	0	0	0	0	0	0	3	…
Person 5	0	0	1	0	0	0	0	0	…
Person 6	1	0	0	0	0	0	0	0	…
Person 7	0	0	0	0	0	0	0	0	…
Person 8	0	0	0	0	0	0	2	0	…
Person 9	0	1	1	0	0	0	0	0	…
Person 10	0	0	0	0	0	0	0	0	…
Person 11	0	0	0	0	0	0	0	0	…
Person 12	2	0	0	0	0	0	0	1	…
…	…	…	…	…	…	…	…	…	…

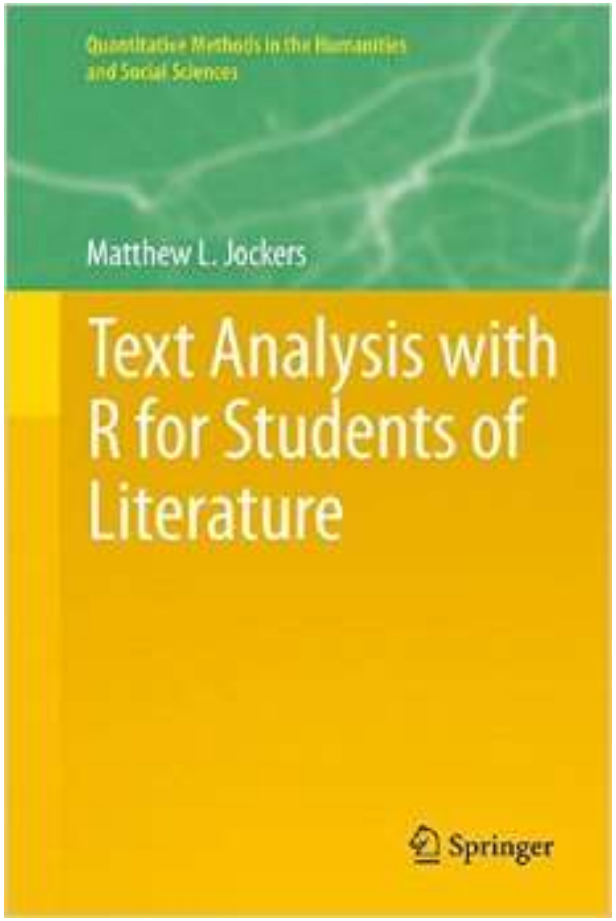
Data Processing

	f1	f2
a	10	5
b	11	6
c	4	13

Data Processing



Data Processing



Text Analysis with R for Students of Literature

Matthew L. Jockers

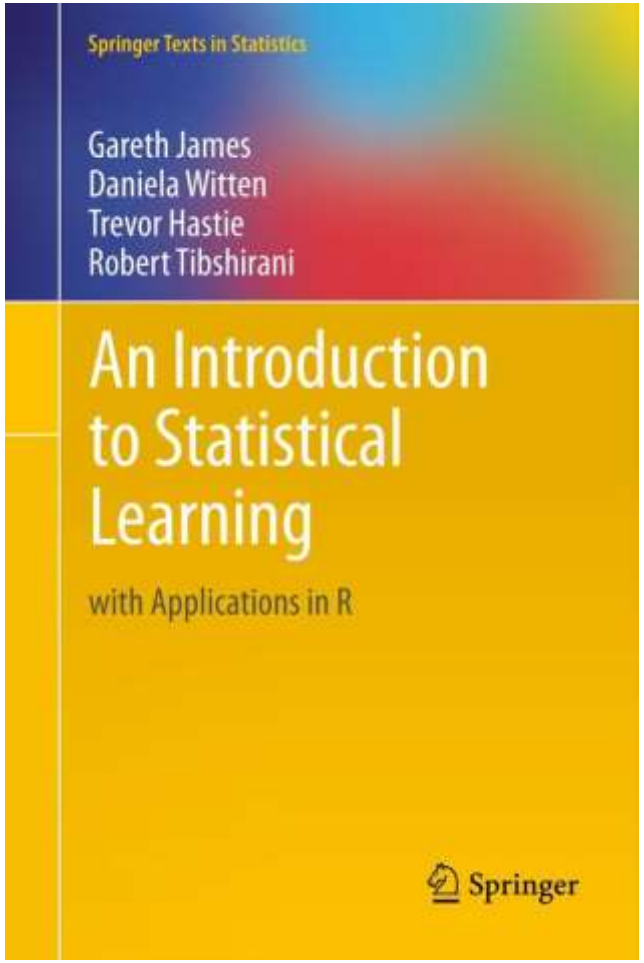
Data Analysis

- Lexical Variety
- Classification
    - Clustering Analysis
    - Latent Semantic Analysis
    - Support Vector Machine
    - Sentimental Analysis
- Topic Modeling

Data Analysis

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive science*, *21*(1), 1-29.

Data Analysis



An Introduction to Statistical Learning

Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani

Sharing Data and Results

- **R + KnitR + RPub**

- **GitHub**

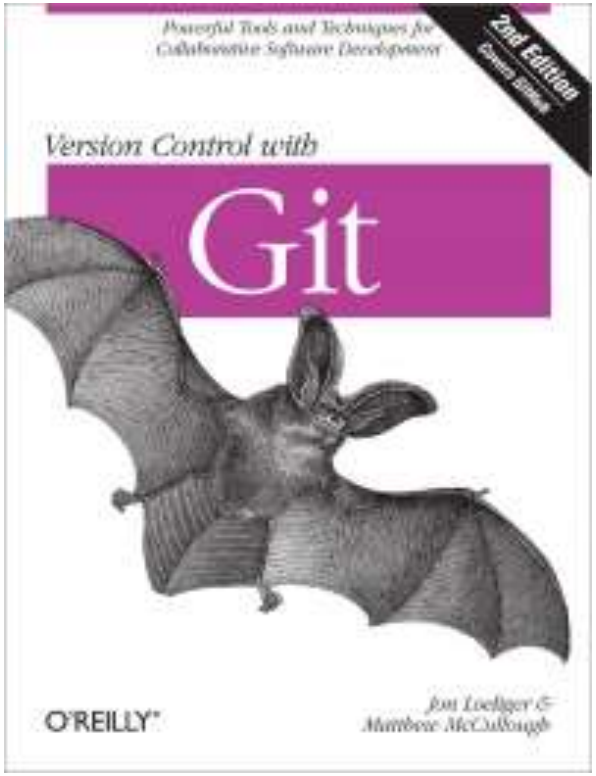
Sharing Data and Results

- **R + KnitR + RPub: http://rpubs.com/neohao/online-help-seeking**

Sharing Data and Results

- **GitHub: https://github.com/Neo-Hao/TwitterHashtagR**

Sharing Data and Results



Version control with Git

Jon Loeliger

Educational Text Mining

Making Educational Research Transparent

Workshop Title: **Online learning analytics on social networking sites: how to tap the potential of data mining in research of educational technology**

Workshop Time: **9:00 am - 12:00 am, 4th Novermber, 2015**

Location: **AECT 2015 Conference, Hyatt Regency-Indianapolis, Indiana**

Thanks!