

Introduction to Text Mining in Educational Research

Neo Qiang Hao

Learning, Design and Technology & Computer Science
University of Georgia

Stephanie Panke

University of North Carolina at Chapel Hill

Text Mining

Problems:

- Prediction / Classification

Is the coming email a spam?

- Qualitative Understanding

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

Text Mining

- **Document 1:**
- **Document 2:**
- **Document 3:**

“THAI TRADE DEFICIT WIDENS IN FIRST QUARTER Thailand's trade deficit widened to 4.5billion baht in the first quarter of 1987 from 2.1 billion year ago, the Business Economics Department said. It said Janunary/March imports rose to 65.1 billion bahtfrom 58.7 billion. Thailand's improved business climate this year resulted in a 27 pct increase in imports of raw material sand semi-finished products. The country's oil import bill, however, fell 23 pct in the first quarter due to lower oil prices.”

Text Mining

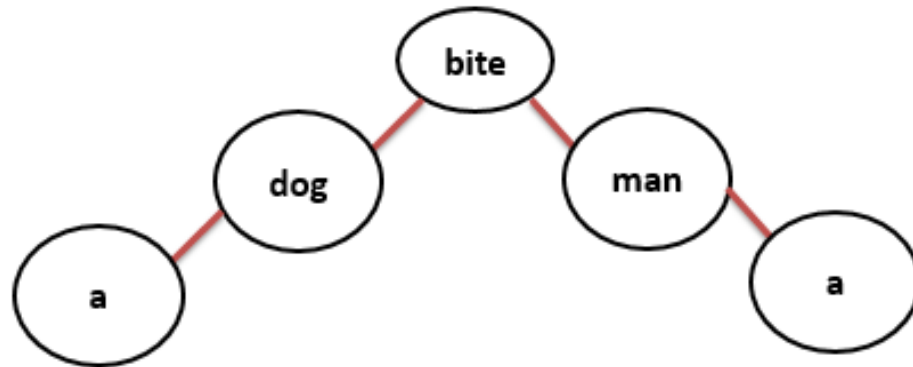
- **A dog bites a man.**
- **A man bites a dog.**

Text Mining

Occam's razor:

Among competing hypotheses, the one with the fewest assumptions should be selected.

“a”, “man”, “dog”, “bites”



Text Mining

[illegible]

Text Mining

- Tf-IDF (www.tfidf.com)
- $TF\text{-}IDF = TF * IDF$

$TF = (\text{Number of times term } t \text{ appears in a document}) / (\text{Total number of terms in the document})$

$IDF = \log_e(\text{Total number of documents} / \text{Number of documents with term } t \text{ in it})$

Text Mining

men	entered	bank	charlotte	missiles	masks	aryan	guns	witnesses	reported
0.239441	0	0.153457	0.195243	0	0.237029	0	0.195243	0.237029	0.140004
0	0	0	0	0	0	0	0	0	0
0	0.192197	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0.158432	0	0	0	0	0	0	0	0
0	0	0	0.197255	0	0	0	0	0	0.141447
0	0	0	0	0	0	0	0	0	0
0	0.234323	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.139629	0	0.127389	0	0	0
0	0	0	0	0	0	0	0	0	0.180656
0	0	0	0	0	0	0.117966	0	0	0.117966
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0.235418	0	0	0	0.214781	0	0	0
0	0	0	0	0.151753	0	0	0	0	0

Questions Answered by Text Mining

- Is the coming email a spam?



Questions Answered by Text Mining

- What aspects do the product reviews for ***Fig Newtons*** on Amazon cover?



Questions to Answer by Text Mining in Education

- What letter score should be given to this essay automatically?



Questions to Answer by Text Mining in Education

- Is that a question relevant to learning?
- Is that question an effective learning question?



Research Methods

- Supervised Learning
 - Logistics Regression
 - Support Vector Machine
 - Decision Tree
 -
- Unsupervised Learning
 - Clustering analysis
 - Manifold learning
 -

Tools



Tools



Carmen Reinhart



Kenneth Rogoff

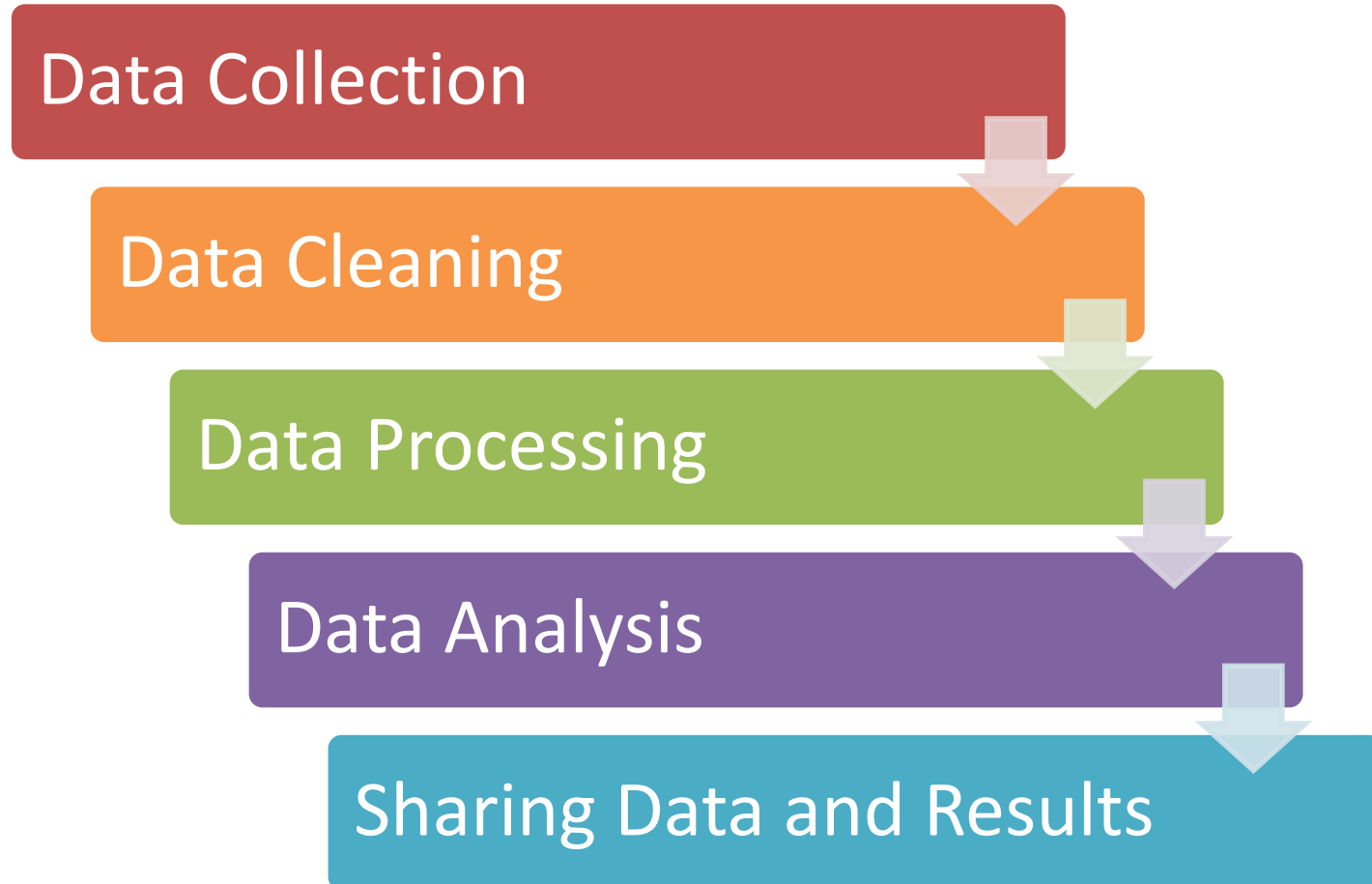


Thomas Herndon

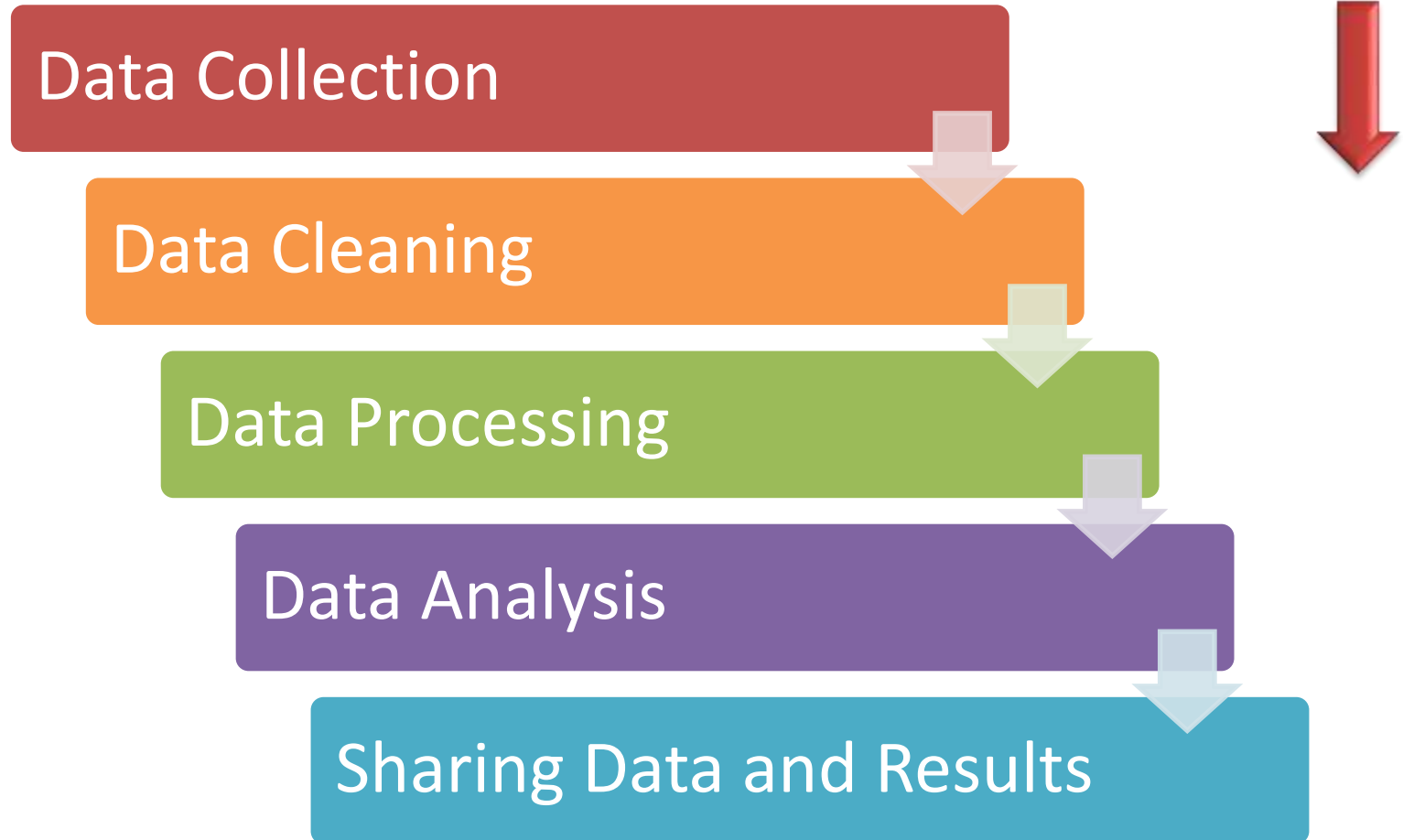
Tools



Research Pipeline



Research Pipeline



Data Cleaning

	text	favorite	favoriteC	replyToSM	created	truncated	replyToSlid	replyToUI	statusSou
1	@mesterman @Ed	FALSE	0	mestermar	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=
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=
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=
7	#Teachers shou	FALSE	0	NA	2015/4/15 23:01	FALSE	NA	5.88E+17	NA <a href=
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 @CanvasPenn	FALSE	0	NA	2015/4/15 21:11	FALSE	NA	5.88E+17	NA <a href=
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 Cyberlearnin	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=
17	RT @Spencer_GG	FALSE	0	NA	2015/4/15 19:47	FALSE	NA	5.88E+17	NA <a href=
18	RT @bsarte: #M	FALSE	0	NA	2015/4/15 19:45	FALSE	NA	5.88E+17	NA <a href=
19	#GoogleClassro	FALSE	2	NA	2015/4/15 19:43	FALSE	NA	5.88E+17	NA <a href=
20	#MDM: Mobile c	FALSE	1	NA	2015/4/15 19:35	FALSE	NA	5.88E+17	NA <a href=
21	bsarte: #MDM:	FALSE	1	NA	2015/4/15 19:32	FALSE	NA	5.88E+17	NA <a href=
22	#MDM: Mobile c	FALSE	1	NA	2015/4/15 19:31	FALSE	NA	5.88E+17	NA <a href=
23	#MDM: Mobile c	FALSE	1	NA	2015/4/15 19:25	FALSE	NA	5.88E+17	NA <a href=
24	#MDM: Mobile c	FALSE	1	NA	2015/4/15 19:20	FALSE	NA	5.88E+17	NA <a href=
25	El impacto de	FALSE	0	NA	2015/4/15 19:13	FALSE	NA	5.88E+17	NA <a href=

Data Cleaning

	text	favorite	favorite	replyToSM	created	truncated	replyToSlid	replyToUI	statusSou
1	@mesterman @Ed	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=
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=
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=
7	#Teachers shou	FALSE	0	NA	2015/4/15 23:01	FALSE	NA	5.88E+17	NA <a href=
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 @CanvasPenn	FALSE	0	NA	2015/4/15 21:11	FALSE	NA	5.88E+17	NA <a href=
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 Cyberlearnin	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=
17	RT @Spencer_GG	FALSE	0	NA	2015/4/15 19:47	FALSE	NA	5.88E+17	NA <a href=
18	RT @bsarte: #M	FALSE	0	NA	2015/4/15 19:45	FALSE	NA	5.88E+17	NA <a href=
19	#GoogleClassro	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=
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=
23	#MDM: Mobile d	FALSE	1	NA	2015/4/15 19:25	FALSE	NA	5.88E+17	NA <a href=
24	#MDM: Mobile d	FALSE	1	NA	2015/4/15 19:20	FALSE	NA	5.88E+17	NA <a href=
25	El impacto de	FALSE	0	NA	2015/4/15 19:13	FALSE	NA	5.88E+17	NA <a href=

Data Cleaning



Regular Expression

madam, baad, dad, gooffoog

Data Cleaning



Regular Expression

```
reg <- "([a-zA-Z0-9]+://)?([a-zA-Z0-9-  
9_]+:[a-zA-Z0-9_]+@)?([a-zA-Z0-9.-  
]+\\.[A-Za-z]{2,4})(:[0-9]+)?(/.*)?«
```

Data Cleaning

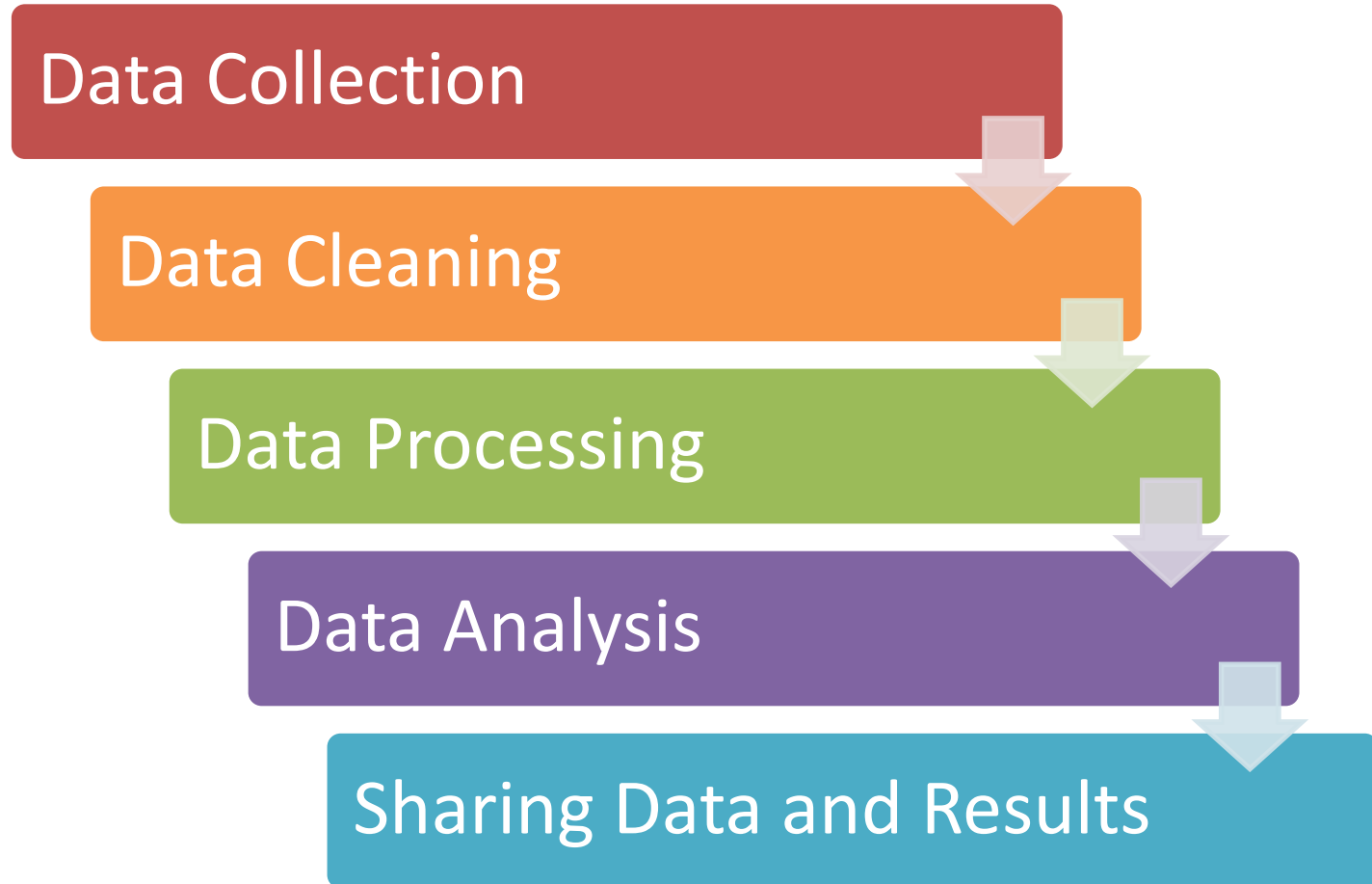


Regular Expression

```
reg <- "([a-zA-Z0-9]+://)?([a-zA-Z0-9-  
9_]+:[a-zA-Z0-9_]+@)?([a-zA-Z0-9.-  
]+\\.[A-Za-z]{2,4})(:[0-9]+)?(/.*)?«
```

www.regular-expressions.info

Research Pipeline



Data Processing

Basic Procedures:

- 1. Remove punctuation**

Data Processing

Basic Procedures:

- 1. Remove punctuation**
- 2. Remove other non-characters**

Data Processing

Basic Procedures:

1. Remove punctuation
2. Remove other non-characters

!@#\$%^&*()_+-~|\\/<>

Data Processing

Basic Procedures:

1. Remove punctuation
2. Remove other non-characters
3. Remove stop words

Data Processing

Basic Procedures:

1. Remove punctuation
2. Remove other non-characters
3. Remove stop words

a, an, the, he, him, I, me, ...

Data Processing

Basic Procedures:

- 1. Remove punctuation**
- 2. Remove other non-characters**
- 3. Remove stop words**
- 4. Lowercases**

Data Processing

Basic Procedures:

1. Remove punctuation
2. Remove other non-characters
3. Remove stop words
4. Lowercases
5. Stem

Data Processing

Basic Procedures:

1. Remove punctuation
2. Remove other non-characters
3. Remove stop words
4. Lowercases
5. Stem

play
played

Data Processing

Basic Procedures:

1. Remove punctuation
2. Remove other non-characters
3. Remove stop words
4. Lowercases
5. Stem

do
does
did

Data Processing

Basic Procedures:

1. Remove punctuation
2. Remove other non-characters
3. Remove stop words
4. Lowercases
5. Stem

go
goes
went

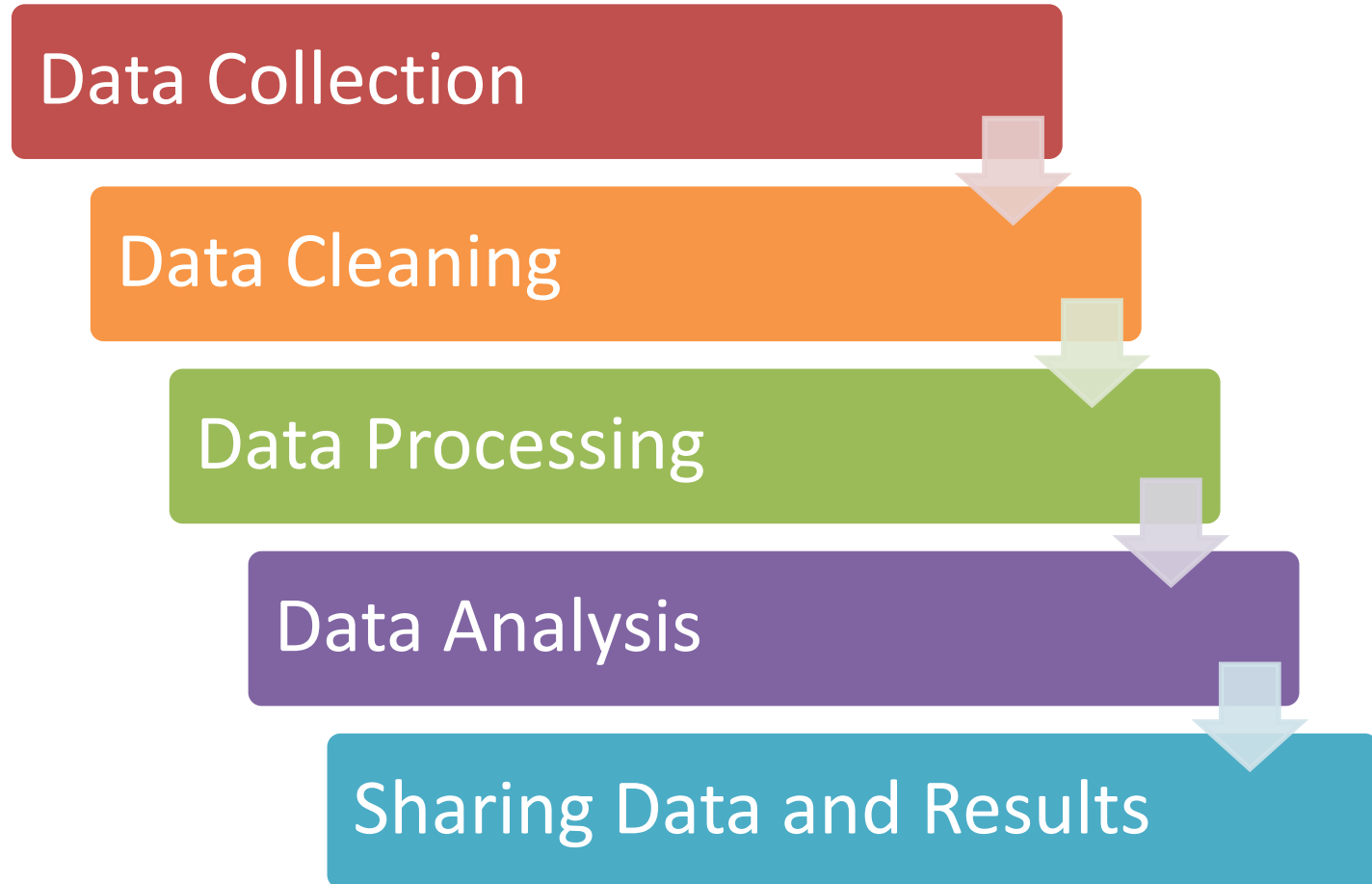
Data Processing

Basic Procedures:

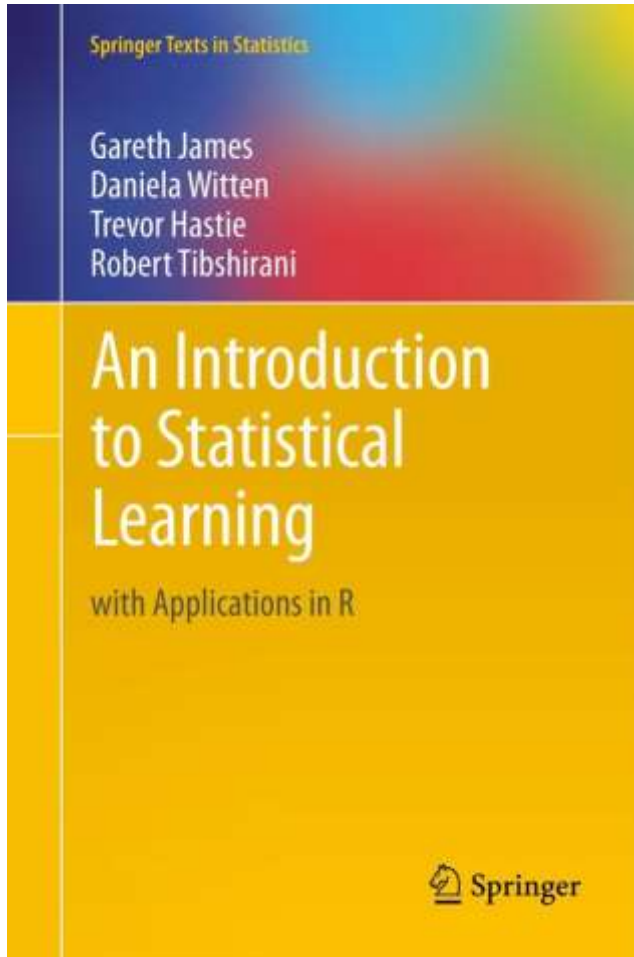
1. Remove punctuation
2. Remove other non-characters
3. Remove stop words
4. Lowercases
5. Stem

lie
lay
laid

Research Pipeline

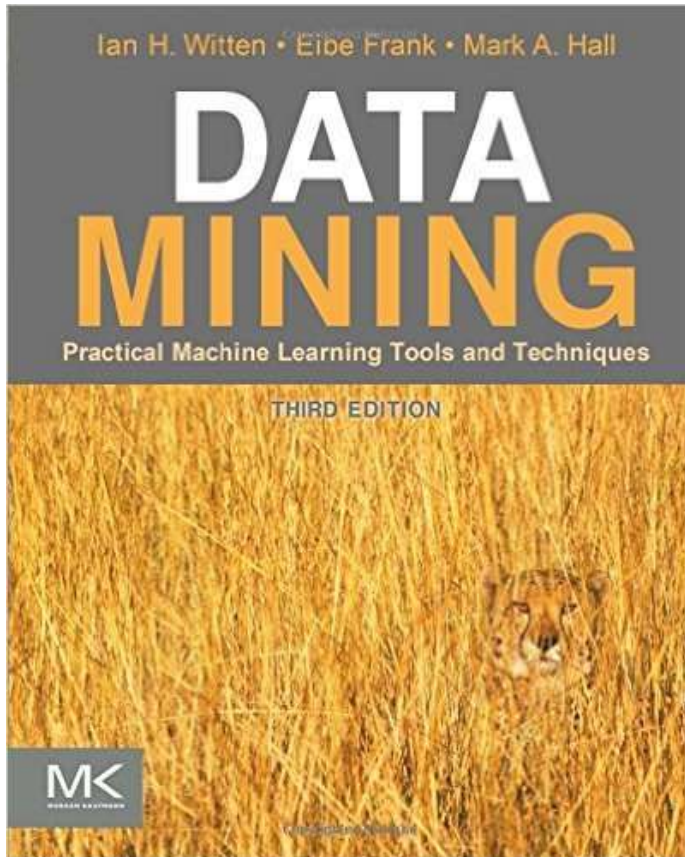


Data Analysis



An Introduction to Statistical Learning with Application in R

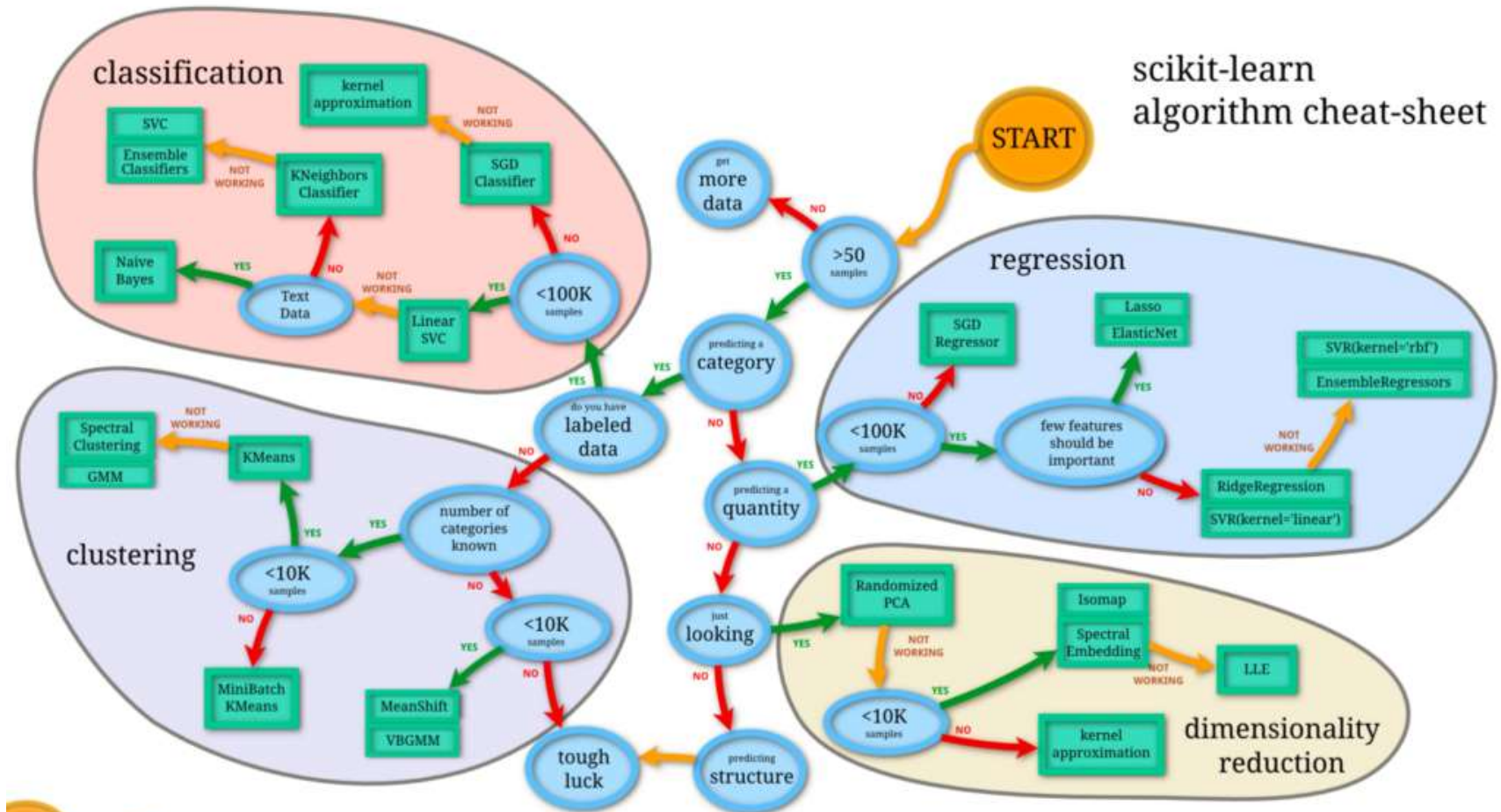
Data Analysis



Data Mining: Practical Machine Learning Tools and Techniques

Data Analysis

scikit-learn
algorithm cheat-sheet

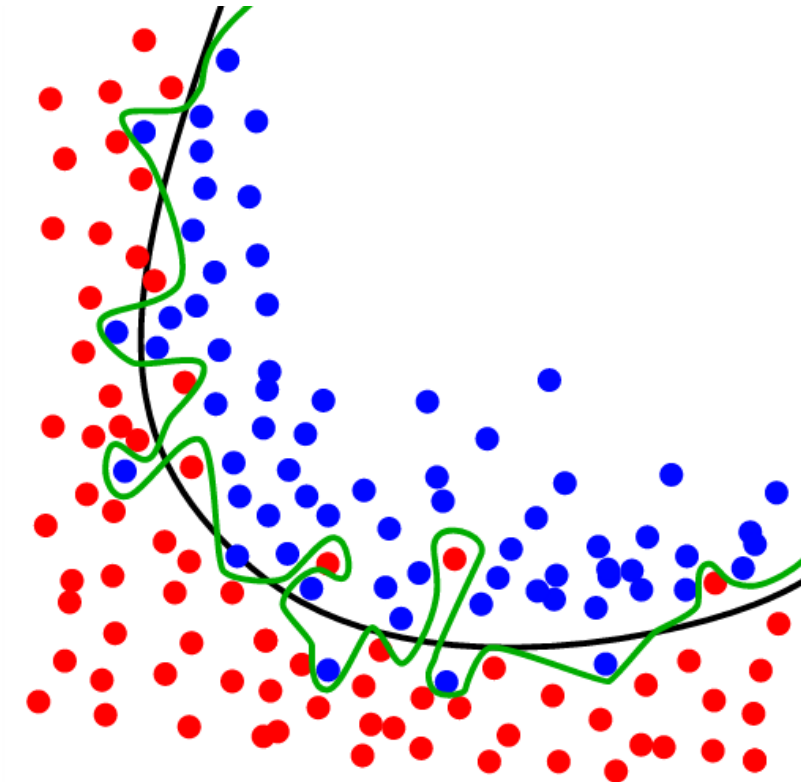


Data Analysis

- **Supervised Learning**
 - **Naïve Bayes Multinomial**
 - **Support Vector Machine**
 -
- **Unsupervised Learning**
 - **Clustering Analysis**
 -

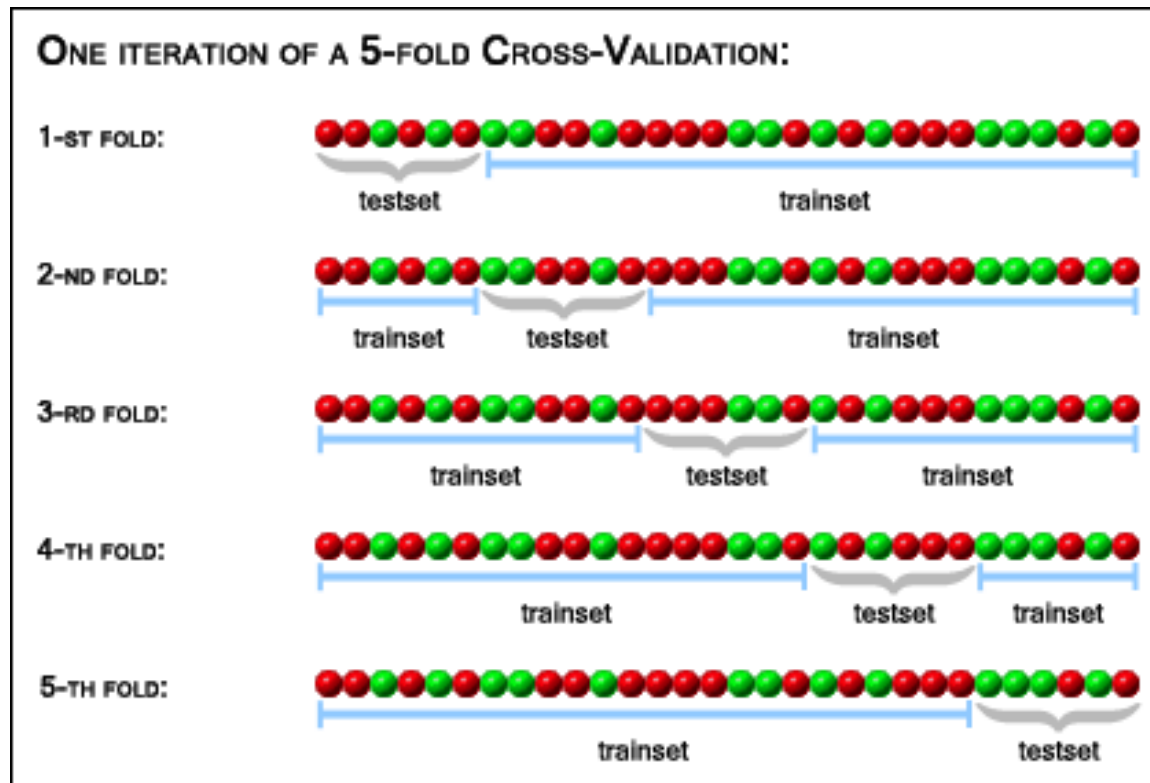
Data Analysis

- **Overfitting**



Data Analysis

- Cross-validation



Data Analysis

- **Naïve Bayes Multinomial**

$$P(y \mid x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n \mid y)}{P(x_1, \dots, x_n)}$$

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Chinese Chinese Tokyo Japan	?

Data Analysis

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Chinese Chinese Tokyo Japan	?

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w | c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|}$$

Data Analysis

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Chinese Chinese Tokyo Japan	?

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w | c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|}$$

Priors:

$$P(c) = \frac{3}{4}$$

$$P(j) = \frac{1}{4}$$

Conditional Probabilities:

$$P(\text{Chinese} | c) = (5+1) / (8+6) = 6/14 = 3/7$$

$$P(\text{Tokyo} | c) = (0+1) / (8+6) = 1/14$$

$$P(\text{Japan} | c) = (0+1) / (8+6) = 1/14$$

$$P(\text{Chinese} | j) = (1+1) / (3+6) = 2/9$$

$$P(\text{Tokyo} | j) = (1+1) / (3+6) = 2/9$$

$$P(\text{Japan} | j) = (1+1) / (3+6) = 2/9$$

Data Analysis

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Chinese Chinese Tokyo Japan	?

$$\hat{P}(c) = \frac{N_c}{N}$$

Priors:

$$P(c) = \frac{3}{4}$$

$$P(j) = \frac{1}{4}$$

$$\hat{P}(w|c) = \frac{\text{count}(w,c) + 1}{\text{count}(c) + |V|}$$

Conditional Probabilities:

$$P(\text{Chinese}|c) = (5+1) / (8+6) = 6/14 = 3/7$$

$$P(\text{Tokyo}|c) = (0+1) / (8+6) = 1/14$$

$$P(\text{Japan}|c) = (0+1) / (8+6) = 1/14$$

$$P(\text{Chinese}|j) = (1+1) / (3+6) = 2/9$$

$$P(\text{Tokyo}|j) = (1+1) / (3+6) = 2/9$$

$$P(\text{Japan}|j) = (1+1) / (3+6) = 2/9$$

Choosing a class:

$$P(c|d5) \propto 3/4 * (3/7)^3 * 1/14 * 1/14$$

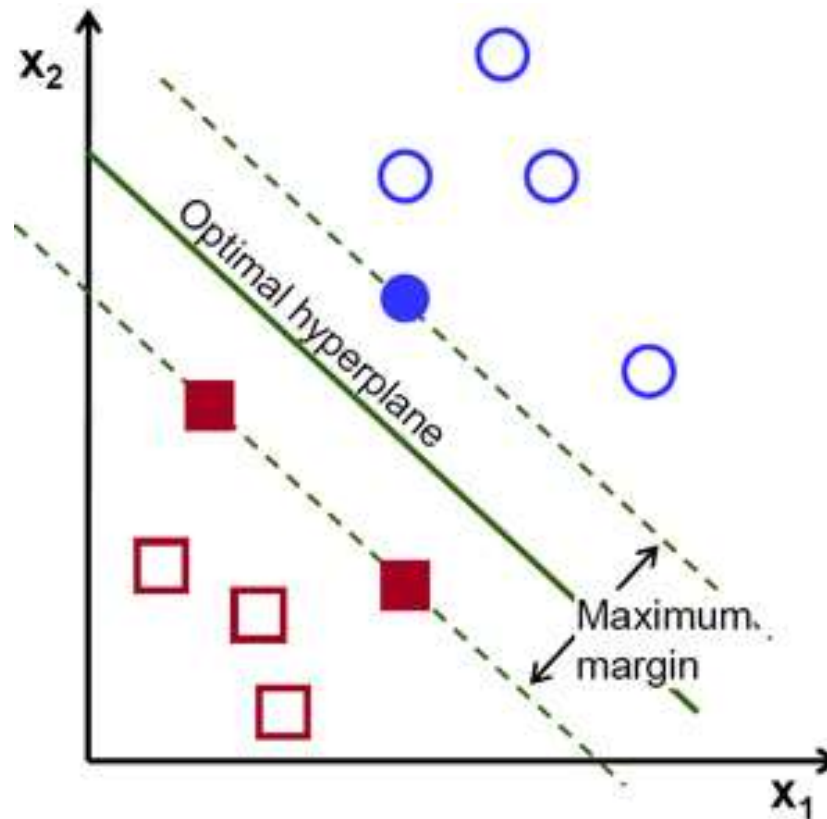
$$\approx 0.0003$$

$$P(j|d5) \propto 1/4 * (2/9)^3 * 2/9 * 2/9$$

$$\approx 0.0001$$

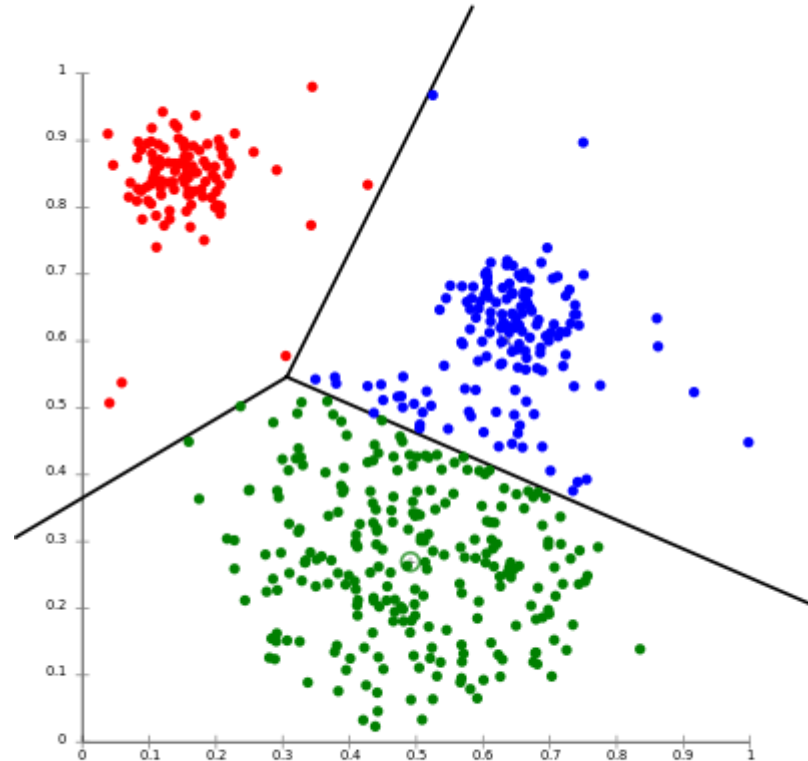
Data Analysis

- **Support Vector Machine**

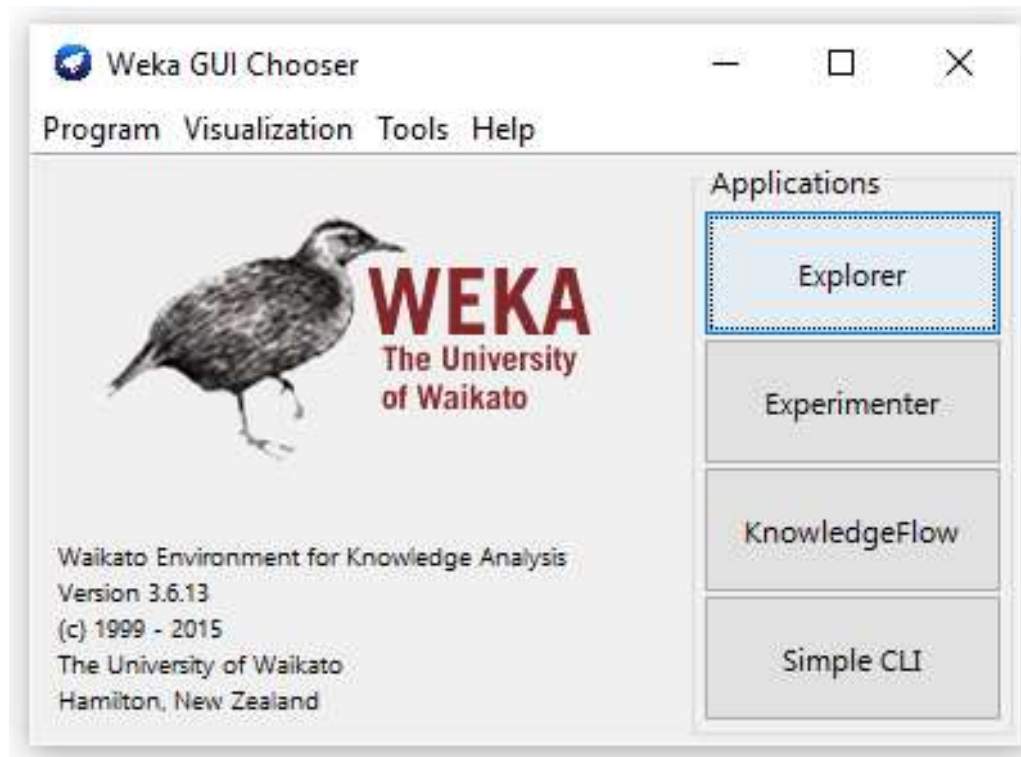


Data Analysis

- **Cluster Analysis**



Data Analysis



Data Analysis



Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** [Apply]

Current relation
Relation: None
Instances: None Attributes: None

Attributes
[All] [None] [Invert] [Pattern]

[Remove]

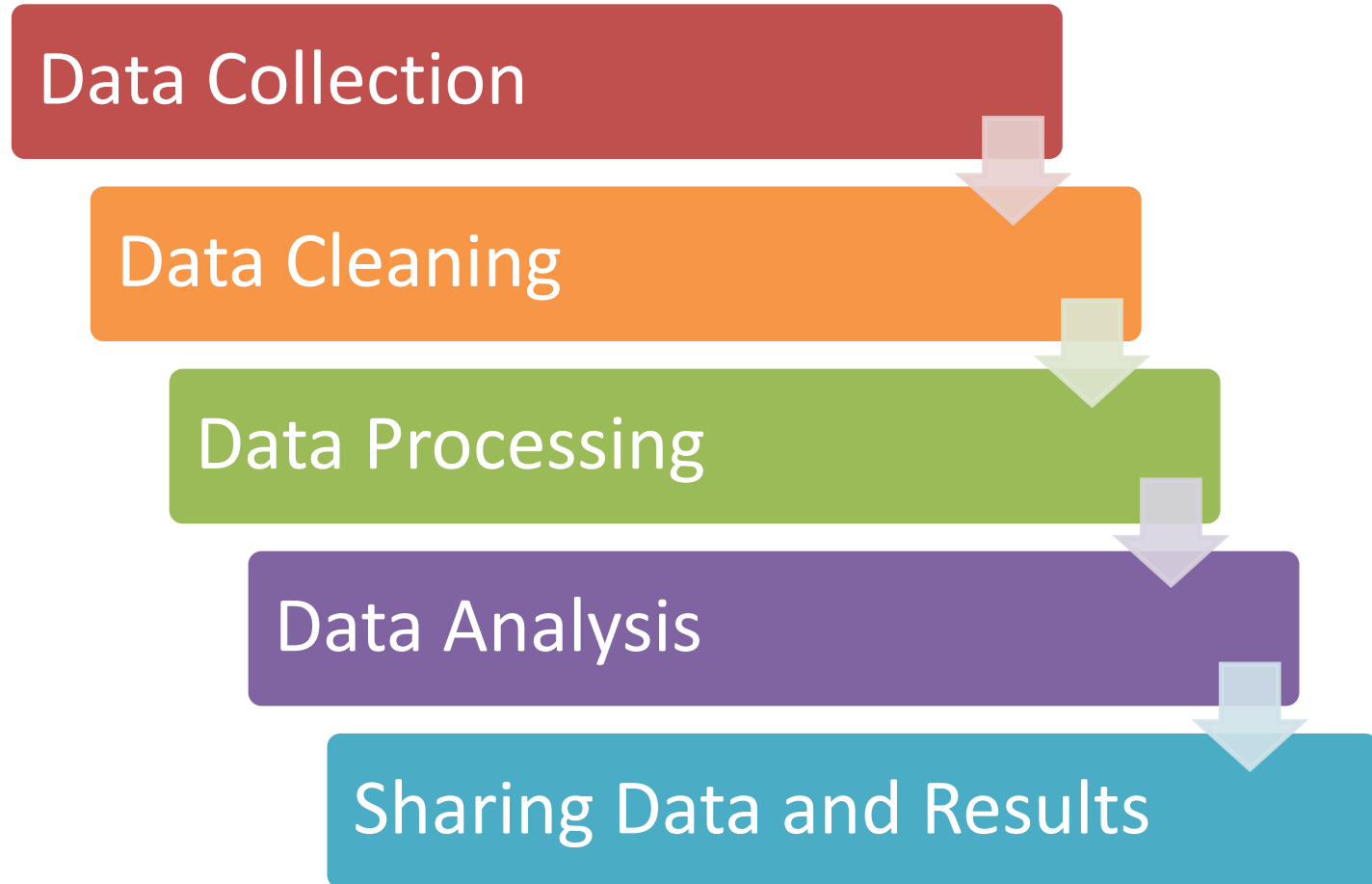
Selected attribute
Name: None Type: None
Missing: None Distinct: None Unique: None

[Visualize All]

Status
Welcome to the Weka Explorer

Log [kiwi icon] x 0

Research Pipeline



Sharing Data and Results

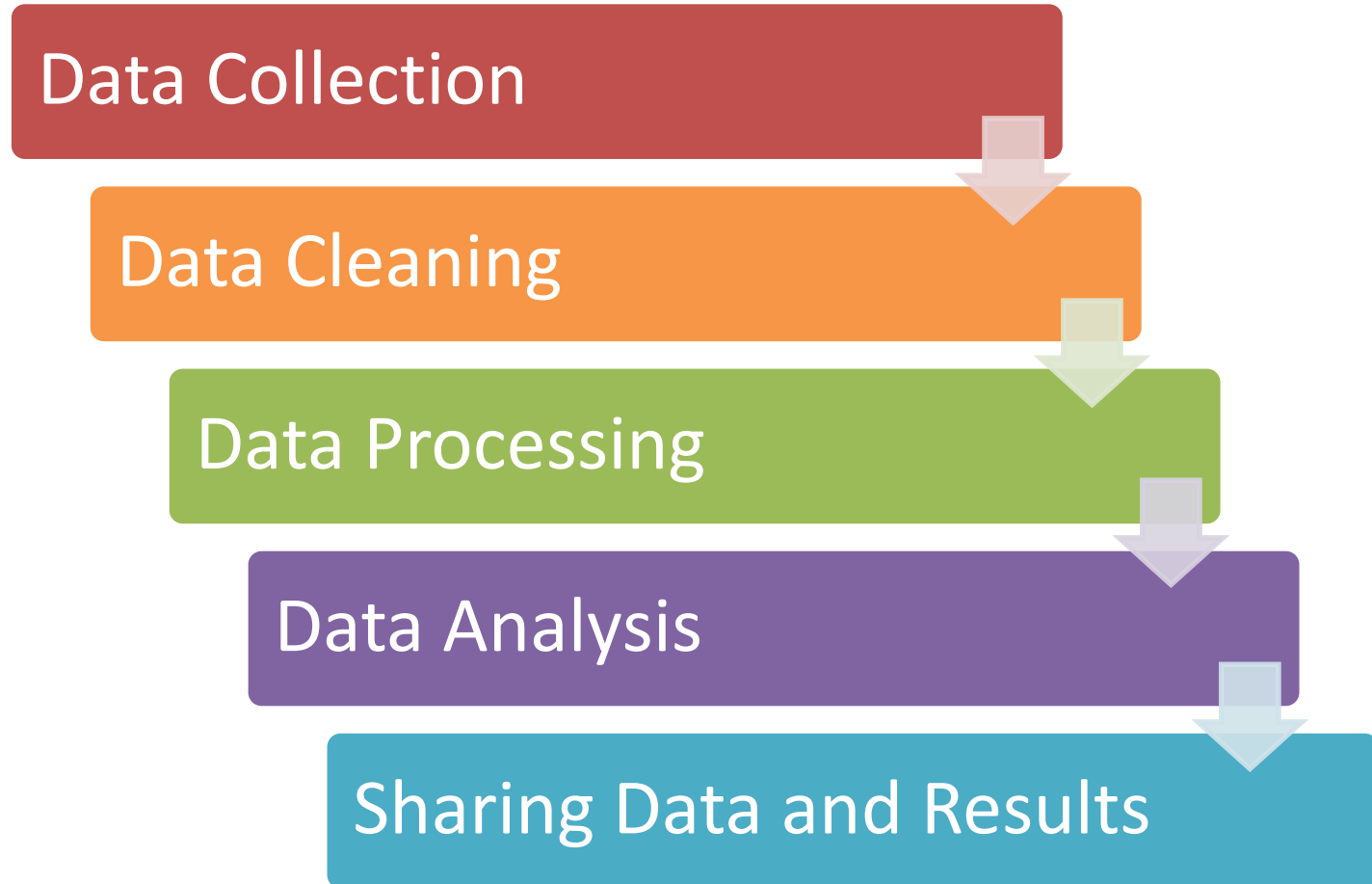
- **Git + GitHub**

- **Git:** <https://git-scm.com/downloads>
-  <https://github.com/Neo-Hao>

- **KnitR + Rpubs**

- **Example:** rpubs.com/neohao/online-help-seeking

Research Pipeline



```
graph BT; A[Discussion Forums or Q&A Platforms] --> D[Data Collection]; B[Facebook] --> D; C[Twitter] --> D; E[Static Pages] --> D;
```

Data Collection

**Discussion Forums or
Q&A Platforms**

Facebook

Twitter

Static Pages

Data Collection



<http://www.reginfo.gov/public/do/eAgendaViewRule?pubId=200210&RIN=1125-AA38>

Data Collection



OFFICE of INFORMATION and REGULATORY AFFAIRS
OFFICE of MANAGEMENT and BUDGET
EXECUTIVE OFFICE of the PRESIDENT

Reginfo.gov

U.S. General
Services
Administration 

Search: ☐ Agenda ☐ Reg Review ☐ ICR

Go

Home | Unified Agenda | Regulatory Review | Information Collection Review | FAQs / Resources | Contact Us

View Rule

[View EO 12866 Meetings](#)[Printer-Friendly Version](#)[Download RIN Data in XML](#)

DOJ/EOIR

RIN: 1125-AA38

Publication ID: Fall 2002

Title: •Protective Orders in Immigration Administration Proceedings

Abstract: This rule amends regulations governing the Executive Office for Immigration Review (EOIR) by authorizing immigration judges to issue protective orders to limit public disclosure of sensitive law enforcement or national defense information during immigration proceedings. The rule is applicable in all proceedings before immigration judges, but involves only a small number of cases.

Agency: Department of Justice(DOJ)

RIN Status: First time published in the Unified Agenda

Major: No

CFR Citation: [8 CFR 3](#)

Legal Authority: [5 USC 301](#), [8 USC 1101 note, 1103, 1231, 1252 note, 1252b, 1324b, 1253, 1362](#), [28 USC 509, 510, 1746](#), sec 2, Reorg Plan No 2 of 1950, [3 CFR 1949-1953 Comp. p 1002](#), [section 203 of PL 105-100](#), [111 Stat 2196-200](#), [sections 1506 and 1510 of PL 106-386](#), [114 Stat 1527-29, 1531-32](#), [section 1505 of PL 106-554](#), [114 Stat 2763A-326 to 2763A-328](#)

Legal Deadline: None

Timetable:

Action	Date	FR Cite
NPRM	05/28/2002	67 FR 36799
NPRM Comment Period End	07/29/2002	
Final Action	12/00/2002	

Regulatory Flexibility Analysis Required: No

Small Entities Affected: No

Included in the Regulatory Plan: No

Agency Contact:
Robin M. Stutman
General Counsel, Office of the General Counsel, Executive Office for Immigration Review

Priority: Substantive, Nonsignificant

Agenda Stage of Rulemaking: Final Rule Stage

Unfunded Mandates: No

Government Levels Affected: None

Federalism: No

Data Collection



<http://www.reginfo.gov/public/do/eAgendaViewRule?pubId=200210&RIN=1125-AA38>

Timetable:		
	Action	Date
NPRM		05/28/2002
NPRM Comment Period End		07/29/2002
Final Action		12/00/2002

Data Collection



<https://github.com/Neo-Hao/Web-Scraping-from-USGSA>

Data Collection



<http://www.reginfo.gov/public/do/eAgendaViewRule?pubId=200210&RIN=1125-AA38>

Timetable:		
	Action	Date
NPRM		05/28/2002
NPRM Comment Period End		07/29/2002
Final Action		12/00/2002

Data Collection

Scrapping data form static web pages:

- 1. A good understanding of HTML & CSS**
- 2. A good understanding of XML & JSON**



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

- 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

- 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

Scrapping data form static web pages:


- 1. A good understanding of HTML & CSS**
- 2. A good understanding of XML & JSON**



Data Collection

Scrapping data form static web pages:

1. A good understanding of HTML & CSS
2. A good understanding of XML & JSON
3. Familiar with Development Tools of Browsers



Chrome DevTools

The Chrome DevTools are a set of web authoring and debugging tools built into Google Chrome. Use the DevTools to iterate, debug and profile your site. [Chrome Canary](#) always has the latest DevTools.

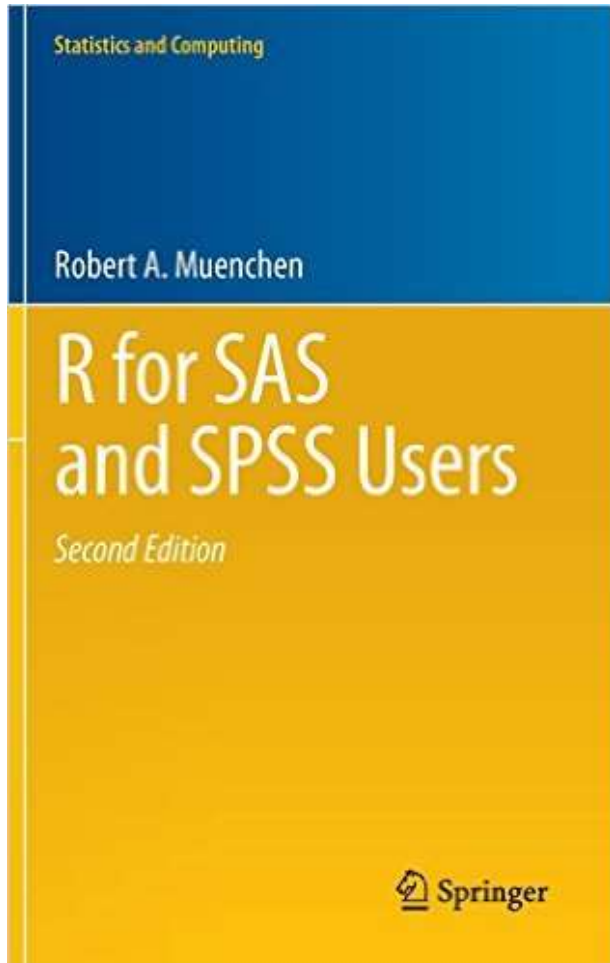
- Select **More Tools > Developer Tools** from the Chrome Menu.
- Right-click on a page element and select Inspect
- Use `Ctrl/Cmd + Shift + I` ([more shortcuts](#))

Data Collection

Scrapping data form static web pages:

- 1. A good understanding of HTML & CSS**
- 2. A good understanding of XML & JSON**
- 3. Familiar with Development Tools of Browsers**
- 4. Familiar with R and package “XML”**

Data Collection

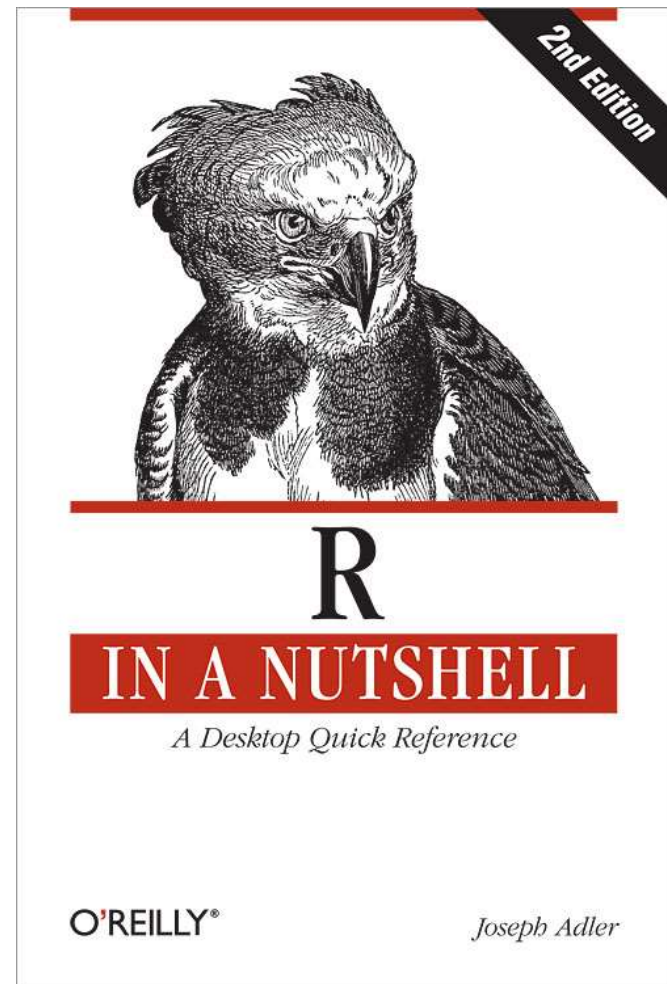


R for SAS and SPSS Users

**Google *r xml package*
*filetype:pdf***

Data Collection

R in a Nutshell



Data Collection



```
getwd()
```

```
setwd("C:/Users/John/Analysis")
```

```
setwd("/home/Analysis")
```

```
setwd("XXX/TwitterHashtag  
R/data")
```

Data Collection



Authentication

1. Register your own app
2. Keep your consumer keys and secrets
3. Go to *Data Collection/Authentication.R*
4. Replace consumer keys and secrets with yours
5. Run lines 1-42

Data Collection



Collect User Info

1. Go to *Data Collection/collectUsers.R*
2. Run lines 1-33

Data Collection



Collect User Info

1. Go to *Data Collection/collectUsers.R*
2. Run lines 1-33
3. **Practice:** Find 5 twitter accounts that you would like to collect information about, and collect their basic information in a .csv file

Data Collection



Collect tweets of particular users

1. Go to *Data Collection/getTweetsByUser.R*
2. Run lines 1-24

Data Collection



Collect tweets of particular users

1. Go to *Data Collection/getTweetsByAllUser.R*
2. Run lines 1-68

Note: Make sure you have a file named “three_conferences.csv” in the current directory.

Data Collection



Collect tweets of particular users

1. Go to *Data Collection/getTweetsByAllUser.R*
2. Run lines 1-68
3. **Practice:** Get tweets from 2 different twitter accounts

Data Collection



Collect tweets by Hashtag

1. Go to *Data Collection/hashtagSearch.R*
2. Run lines 1-22

Data Collection



Collect tweets by Hashtag

1. Go to *Data Collection/hashtagSearch.R*
2. Run lines 1-22
3. **Practice:** Get tweets with one hashtag you like

Thanks!

neohao@uga.edu