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

Qiang (Neo) Hao, Learning, Design and Technology & Computer Science, University of Georgia

Robert Maribe Branch, Learning, Design and Technology, University of Georgia

Questions Answered by Text Mining

- Is the coming email a spam?

Classification Problem

Questions Answered by Text Mining

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

Classification Problem

Questions to Answer by Text Mining

- What algorithm can score essays as teachers do?

Classification Problem

Questions Answered by Text Mining

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

Clustering Problem

Questions to Answer by Text Mining

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

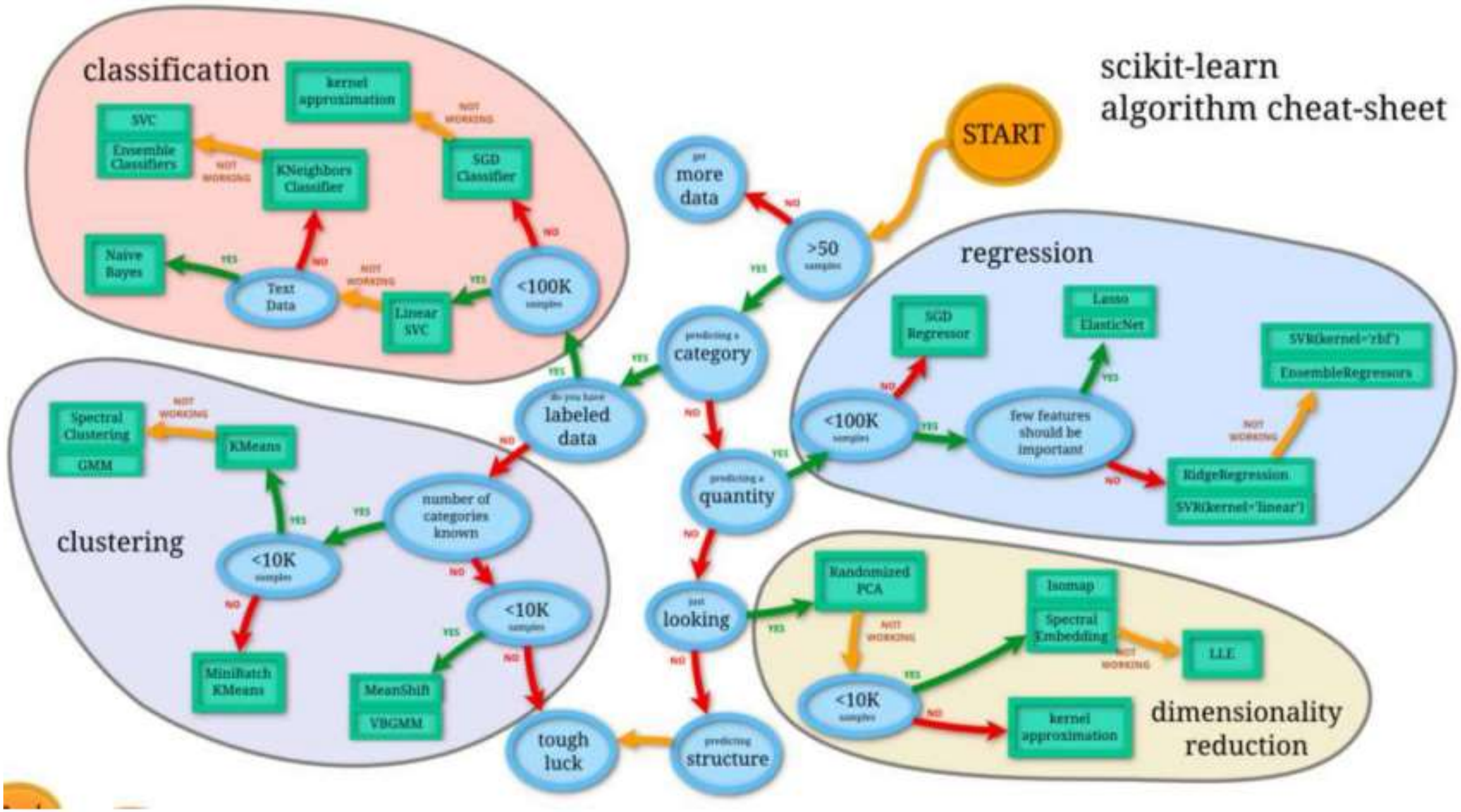
Clustering Problem

Questions to Answer by Text Mining

- What courses should we recommend students' based on their course reviews and engagement levels of their enrolled courses?

Recommendation Problem

Road Map



scikit-learn
algorithm cheat-sheet
START
>50 samples
get more data
predicting a category
do you have labeled data
predicting a quantity
just looking
predicting structure
tough luck
classification
kernel approximation
SVC
Ensemble Classifiers
KNeighbors Classifier
SGD Classifier
Naive Bayes
Text Data
Linear SVC
<100K samples
clustering
Spectral Clustering
GMM
KMeans
number of categories known
<10K samples
MiniBatch KMeans
MeanShift
VBGMM
regression
SGD Regressor
Lasso
ElasticNet
SVR(kernel='rbf')
EnsembleRegressors
few features should be important
RidgeRegression
SVR(kernel='linear')
dimensionality reduction
Randomized PCA
Isomap
Spectral Embedding
LLE
YES
NO
NOT WORKING

Road Map

1. Whether there are targets
   – Are targets necessary?
   – Classification, Prediction / Clustering Analysis
2. Whether the targets are continuous
   – Would you prefer continuous targets?
   – Classification / Prediction
3. Noisy attributes?
   – Attribute Selection
   – Dimension Deduction

Road Map

1. Prediction
   – Regression
   – Neuro Network
2. Classification
   – Support Vector Machine
   – Naïve Bayes
3. Clustering
   – Cluster Analysis
4. Attribute Collapsing
   – Greedy Algorithm
   – Principle Component Analysis

Tools

hadoop

APACHE Spark

dmlc XGBoost

amazon web services EC2

R

python

WEKA

X

SPSS

M

Tools

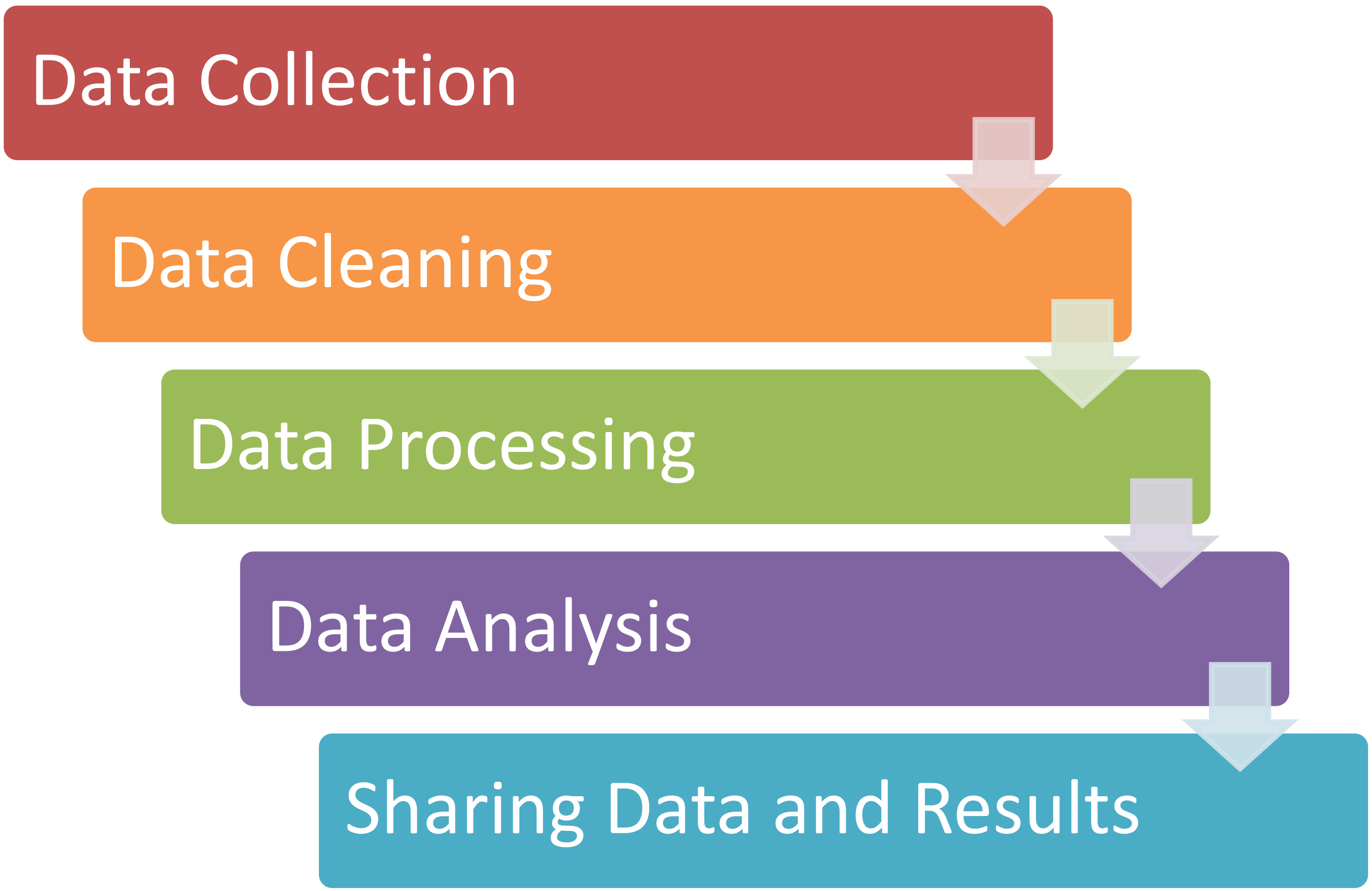
Carmen Reinhart

Kenneth Rogoff

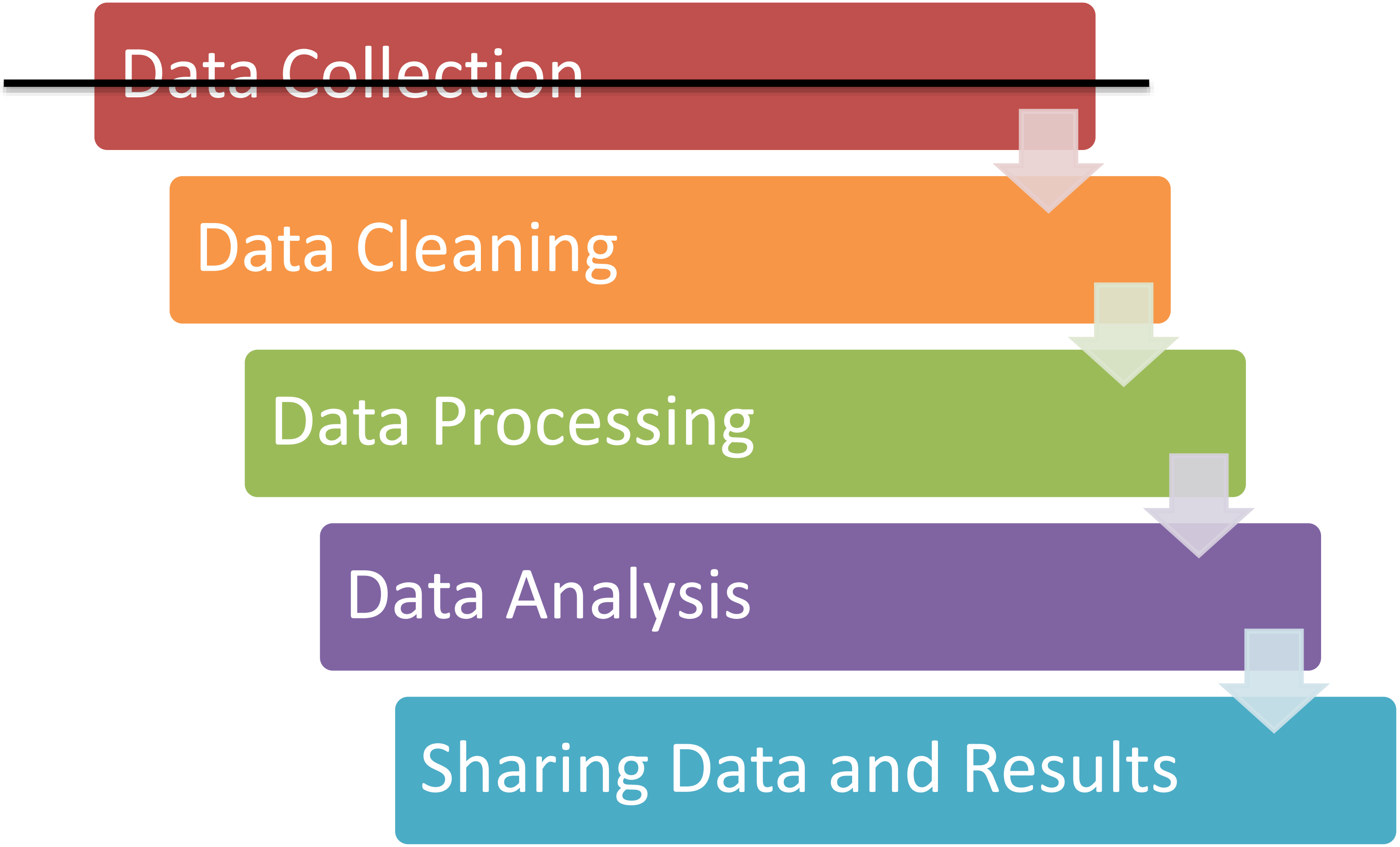
Thomas Herndon

Tools

Research Pipeline



Research Pipeline



Data Cleaning

	text	favoritec	favoriteC	replyToSN	created	truncatec	replyToSI	id	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 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

	text	favoritec	favoriteC	replyToSN	created	truncatec	replyToSI	id	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 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="
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	#GoogleClassrc	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

Regular Expression

`otter, otters, Otter, OTTER, OTTERS`

Data Cleaning

Regular Expression

`madam, baad, dad, gooffoog`

Palindrome

Data Cleaning

Regular Expression

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

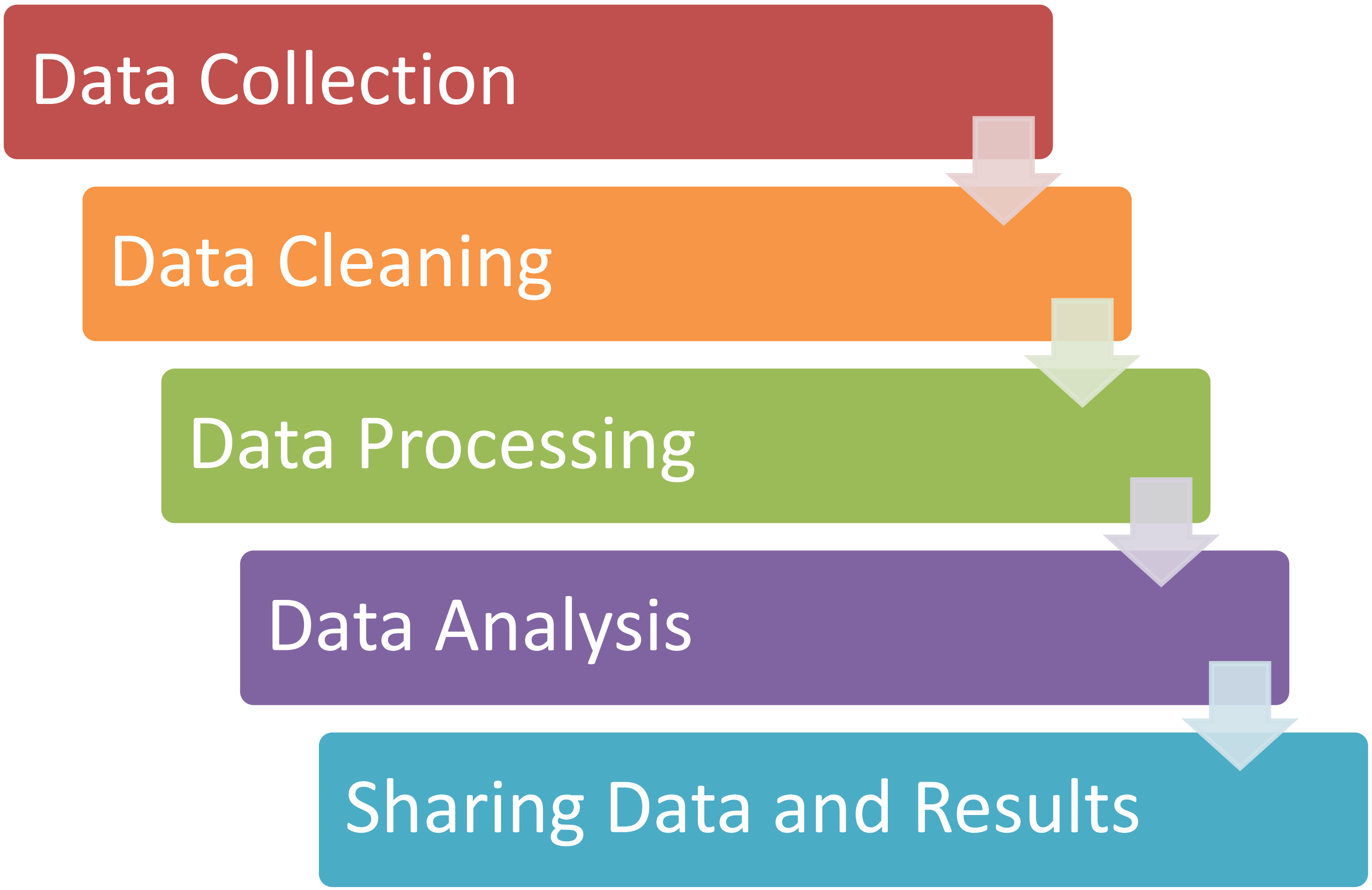
Data Cleaning

Regular Expression

```
reg <- "([a-zA-Z0-9]+://)?([a-zA-Z0-9_]+:[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
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**

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

Data Processing

Basic Procedures:

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

try

tries

tried

Data Processing

Assumption:

1. **Bag of words**

A dog bites a man.
A man bites a dog.

“a”, “man”, “dog”, “bites”

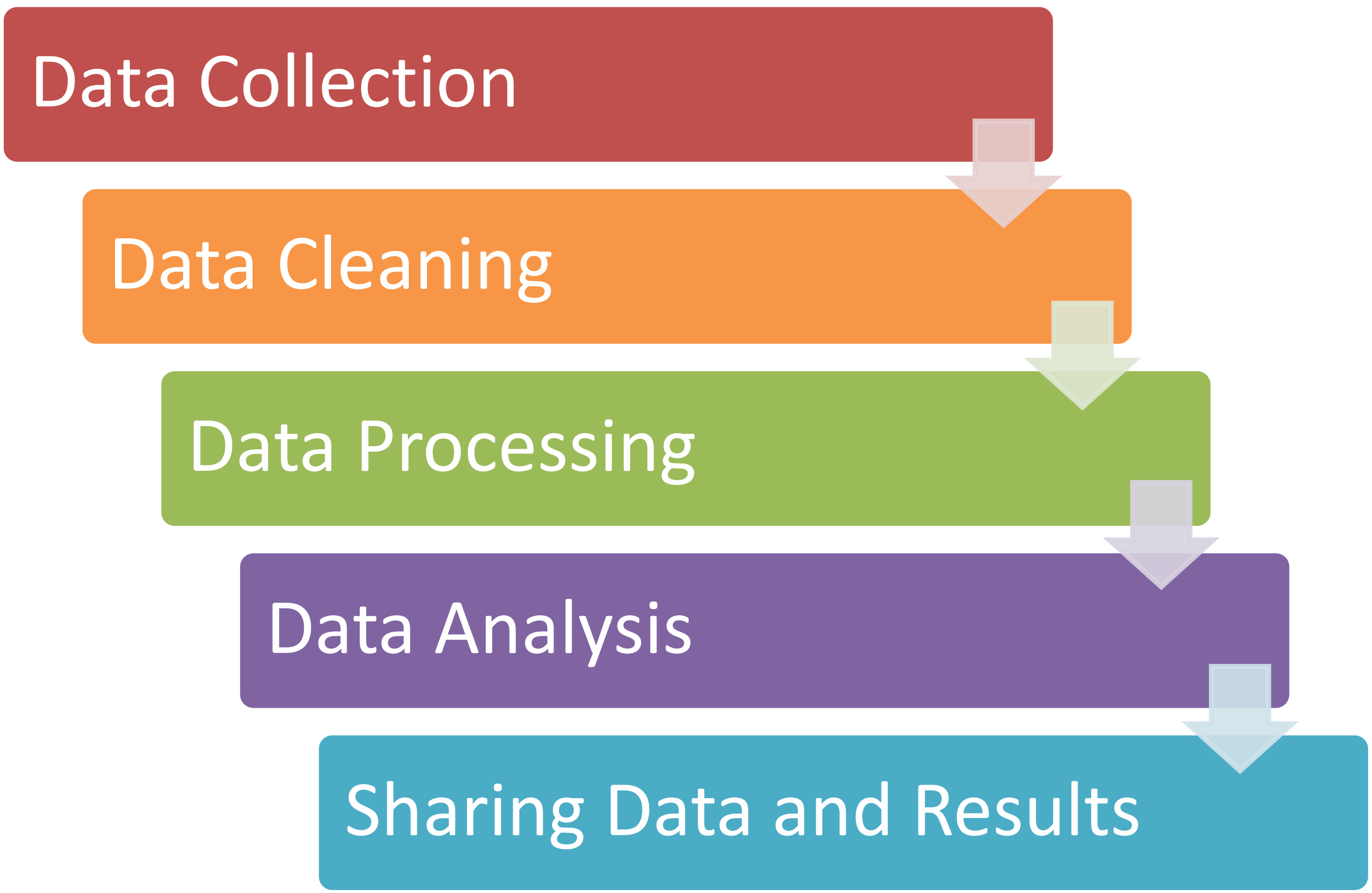
Data Processing

Assumption:

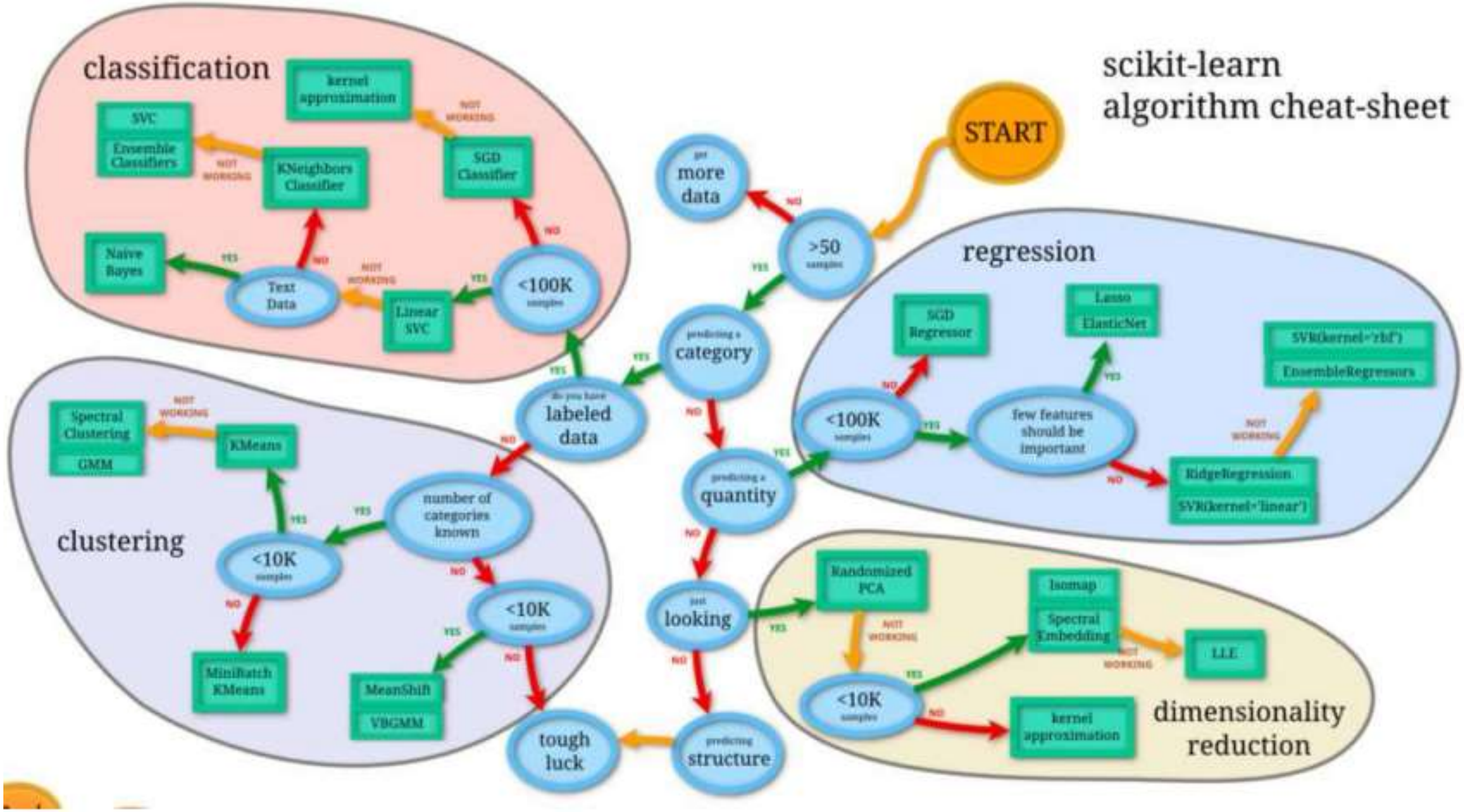
1. **Bag of words**
2. **Words as features**

	word1	word2	word3	word4	word5	word6	word7	word8	...
doc1	0	0	0	0	0	0	0	0	...
doc2	0	0	0	0	0	0	0	2	...
doc3	0	0	0	0	0	0	0	0	...
doc4	0	3	0	0	0	6	0	0	...
doc5	0	0	0	8	0	0	0	0	...
doc6	0	0	0	0	0	0	0	0	...
doc7	0	0	0	0	0	0	4	0	...
doc8	0	0	0	0	0	0	0	0	...
doc9	0	0	5	0	0	0	0	0	...
doc10	0	0	0	0	0	0	0	0	...
doc11	0	0	0	0	0	1	0	0	...
doc12	0	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...

Research Pipeline



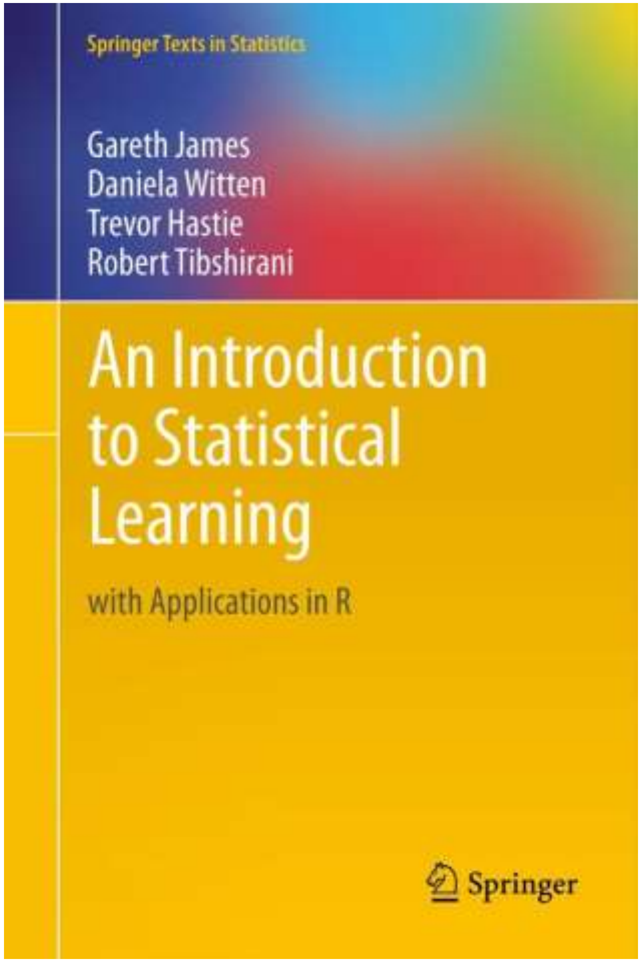
Data Analysis



Data Analysis

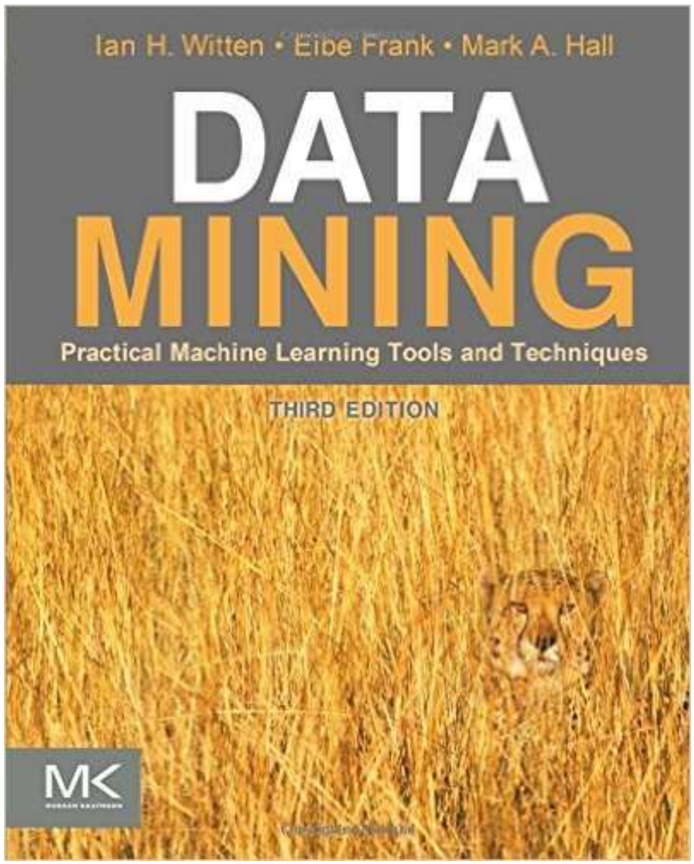
1. Whether there are targets
   - Are targets necessary?
   - Classification, Prediction / Clustering Analysis
2. Whether the targets are continuous
   - Would you prefer continuous targets?
   - Classification / Prediction
3. Noisy attributes?
   - Attribute Selection
   - Dimension Deduction

Data Analysis



An Introduction to Statistical Learning with Application in R

Data Analysis



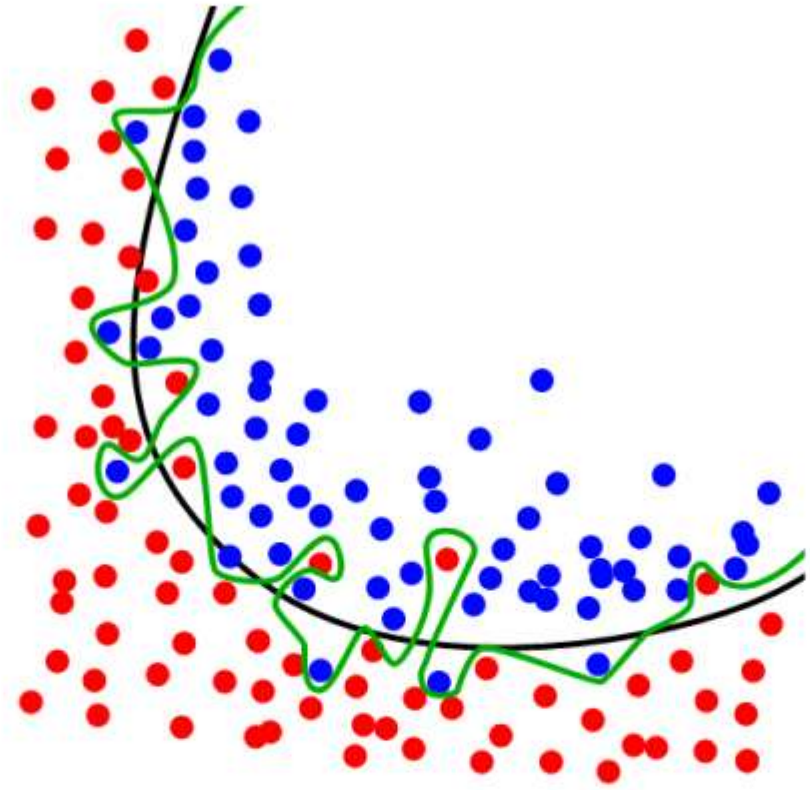
Data Mining: Practical Machine Learning Tools and Techniques

Data Analysis

- Overfitting
- Cross validation
- Naïve Bayes
- K-means
- Support Vector Machine

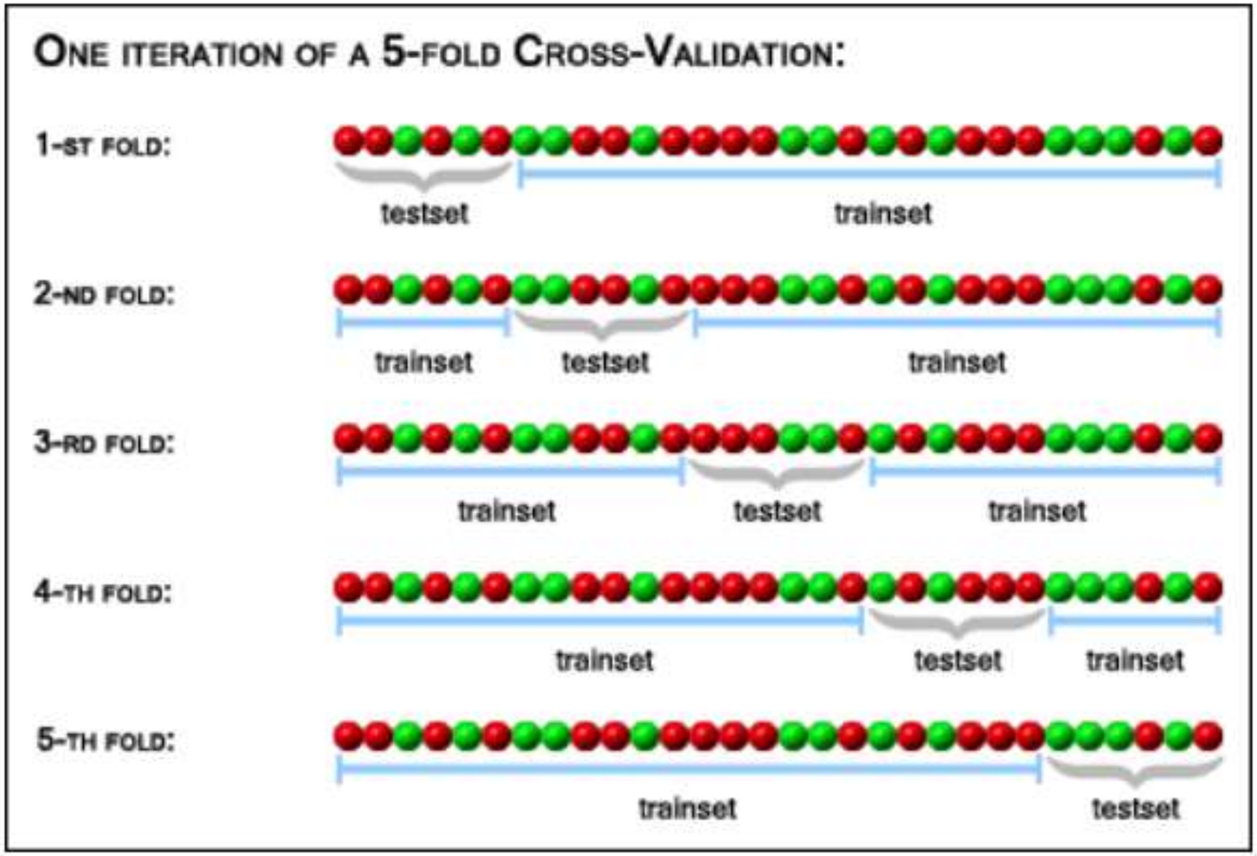
Data Analysis

- **Overfitting**



Data Analysis

- **Cross Validation**



Data Analysis

- **Naïve Bayes**

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	C
	2	Chinese Chinese Shanghai	C
	3	Chinese Maco	C
	4	Japan Tokyo Chinese	J
Test	5	Chinese Chinese Chinese Japan Tkyo	?

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

Data Analysis

- **Naïve Bayes**

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	C
	2	Chinese Chinese Shanghai	C
	3	Chinese Maco	C
	4	Japan Tokyo Chinese	J
Test	5	Chinese Chinese Chinese Japan Tkyo	?

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

Data Analysis

- **Naïve Bayes**

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	C
	2	Chinese Chinese Shanghai	C
	3	Chinese Maco	C
	4	Japan Tokyo Chinese	J
Test	5	Chinese Chinese Chinese Japan Tkyo	?

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

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

Data Analysis

- **Naïve Bayes**

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	C
	2	Chinese Chinese Shanghai	C
	3	Chinese Maco	C
	4	Japan Tokyo Chinese	J
Test	5	Chinese Chinese Chinese Japan Tkyo	?

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

- **Naïve Bayes**

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	C
	2	Chinese Chinese Shanghai	C
	3	Chinese Maco	C
	4	Japan Tokyo Chinese	J
Test	5	Chinese Chinese Chinese Japan Tkyo	?

Priors:

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

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

Conditional Probabilities:

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

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

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

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

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

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

$$P(c|d5) \propto 3/4 * (3/7)^3 * 1/14 * 1/14 \approx 0.0003$$

$$P(j|d5) \propto 1/4 * (2/9)^3 * 2/9 * 2/9 \approx 0.0001$$

Data Analysis

- **K-means Algorithm**

- **K-means Algorithm**

Input:

- K (number of clusters)
- Training set $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

Data Analysis

- **K-means Algorithm**

Randomly initialize K cluster centroids $\mu_1, \mu_2, \ldots, \mu_K \in \mathbb{R}^n$

Repeat {

  for i = 1 to m

    $c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$

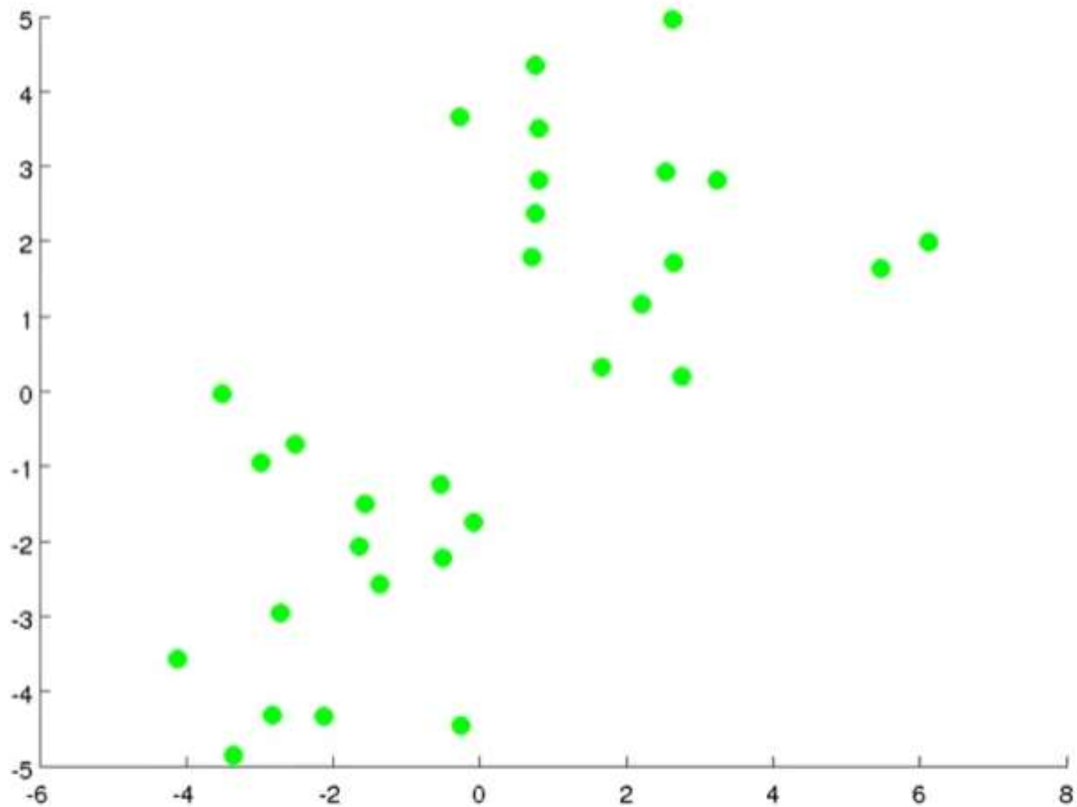
  for k = 1 to K

    μ_k := average (mean) of points assigned to cluster k

}

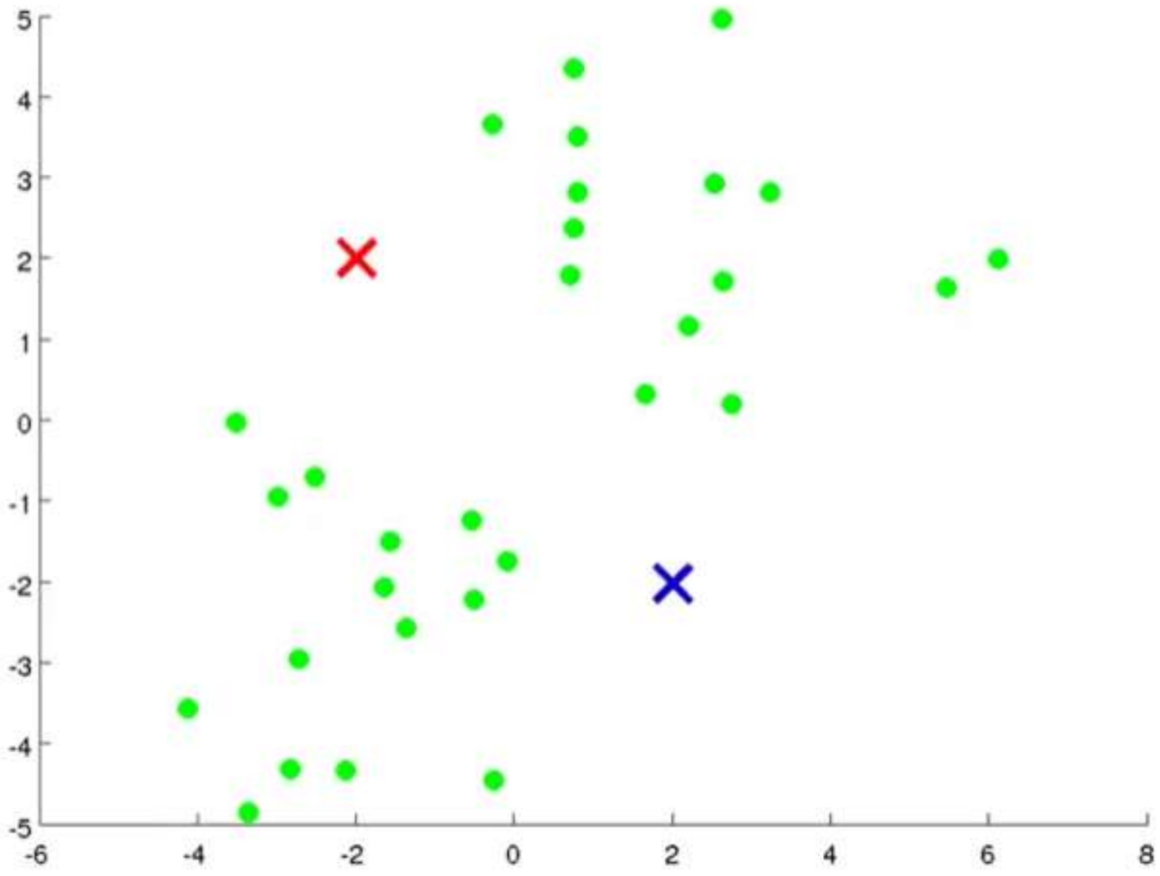
Data Analysis

- **K-means Algorithm**



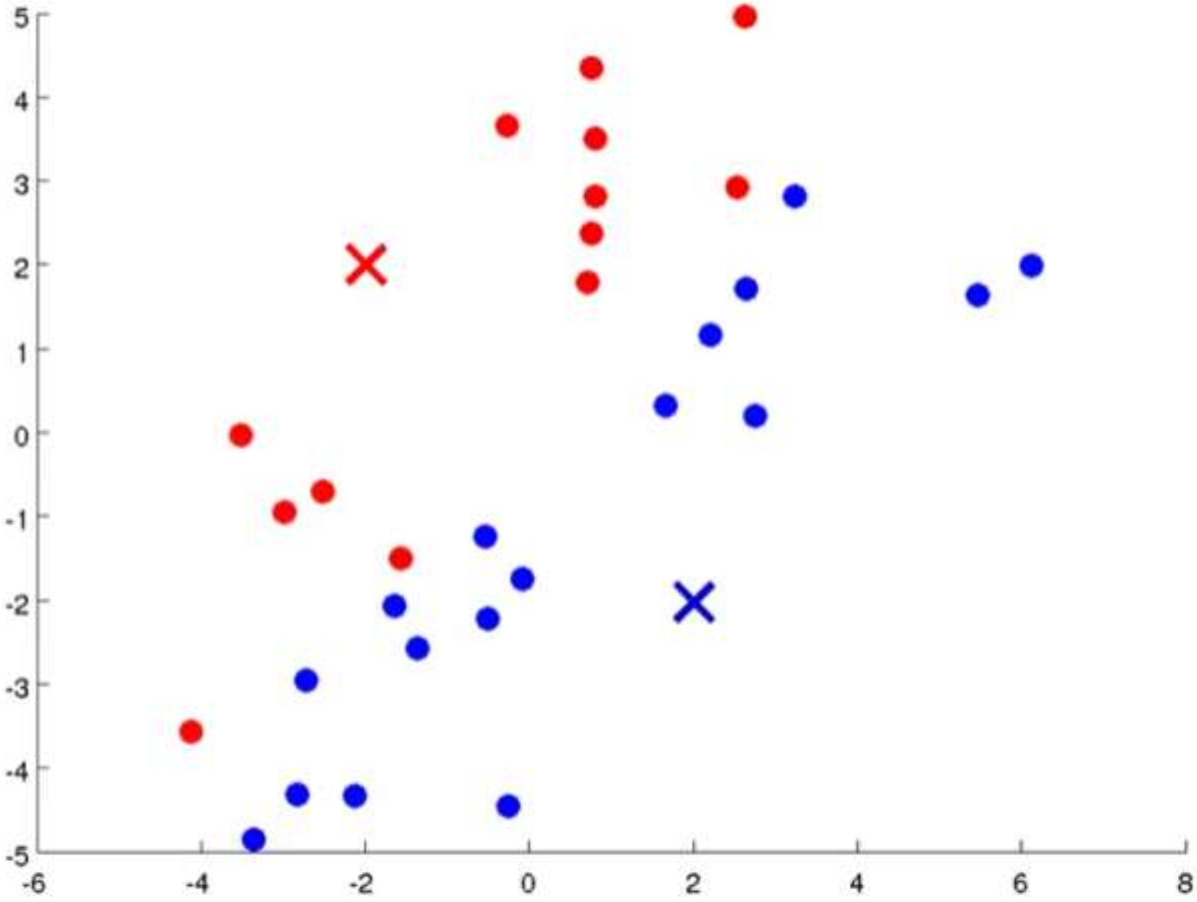
Data Analysis

- **K-means Algorithm**



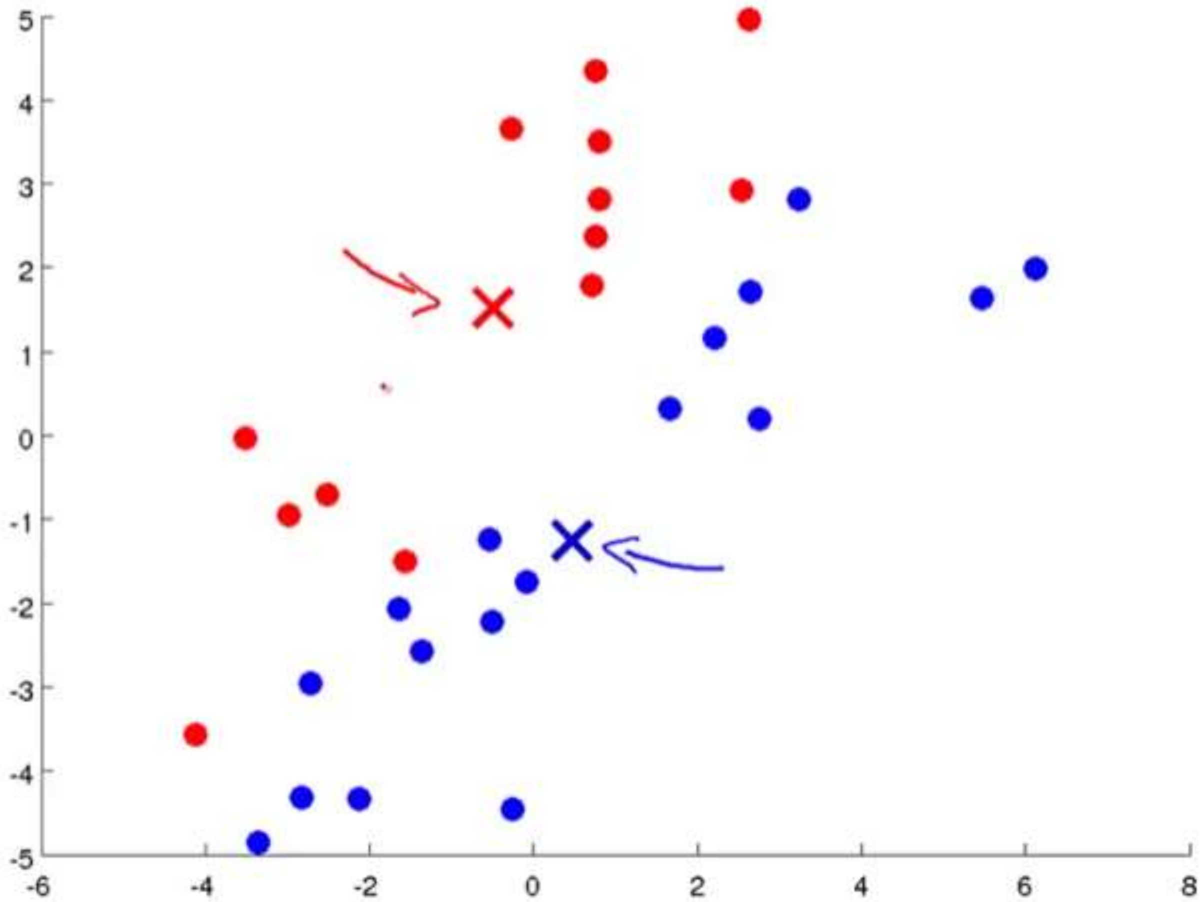
Data Analysis

- **K-means Algorithm**



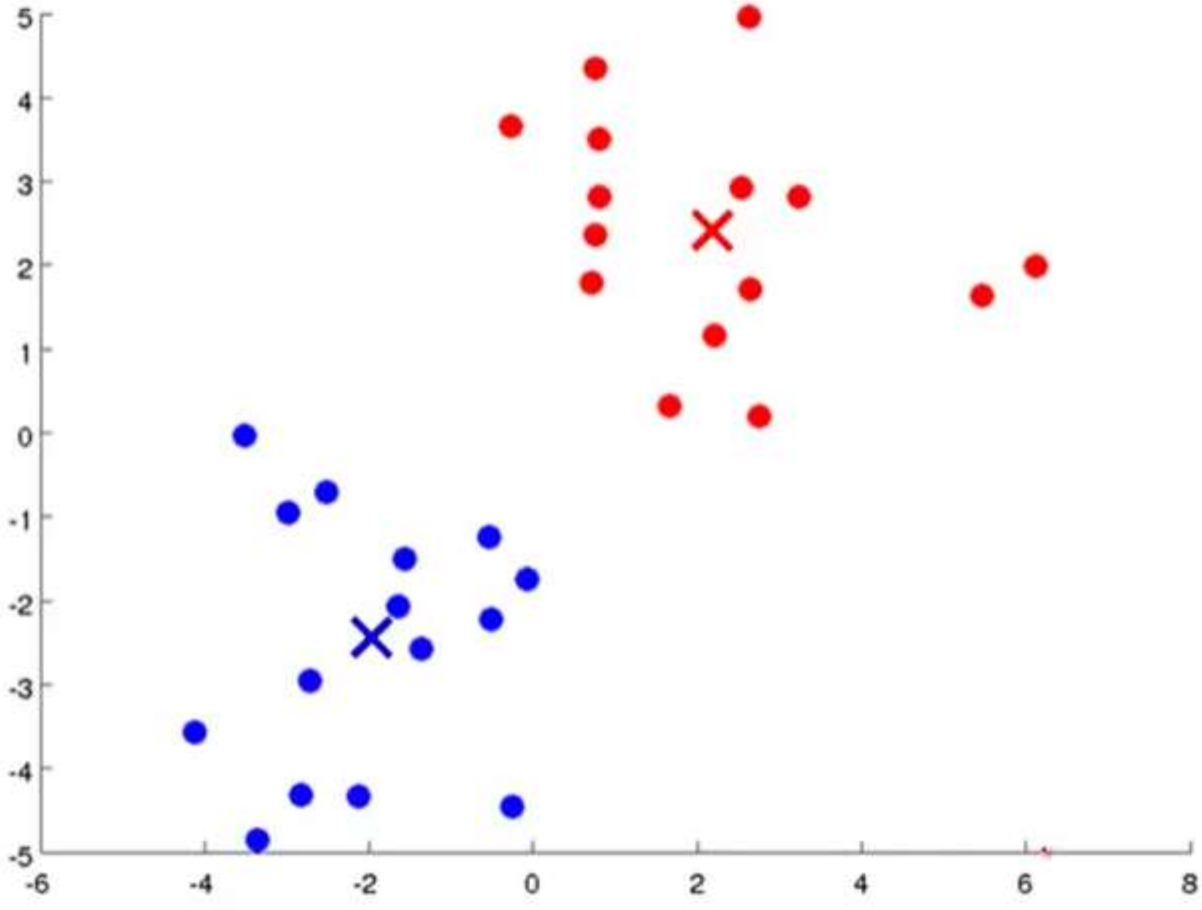
Data Analysis

- **K-means Algorithm**



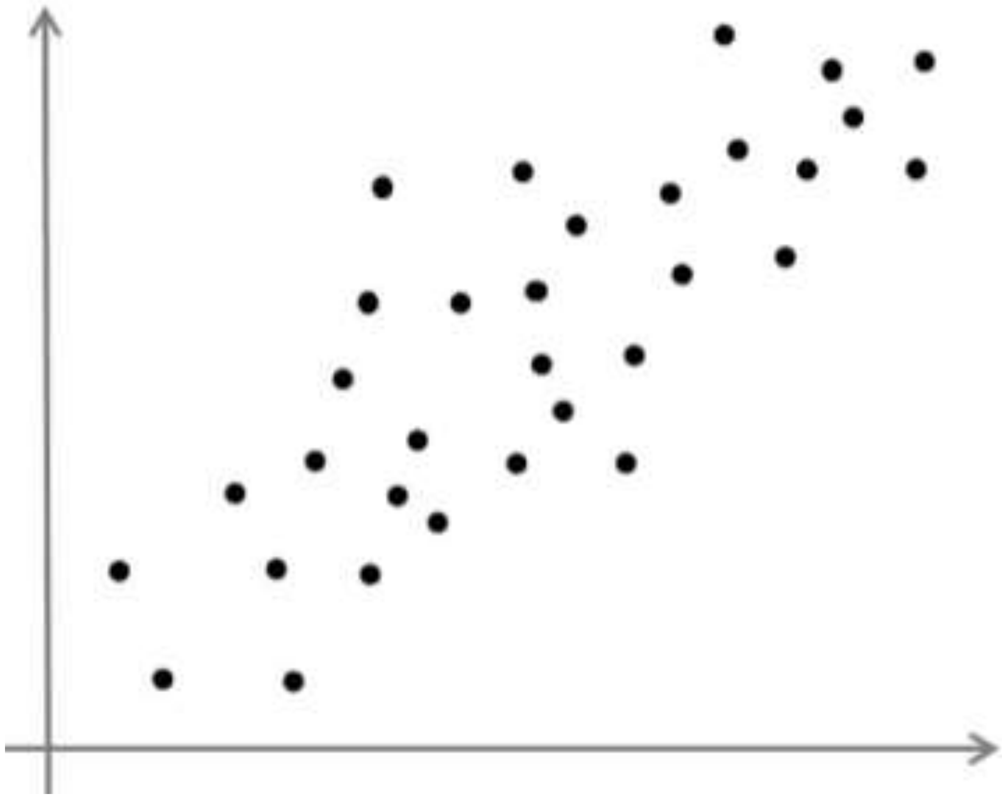
Data Analysis

- **K-means Algorithm**



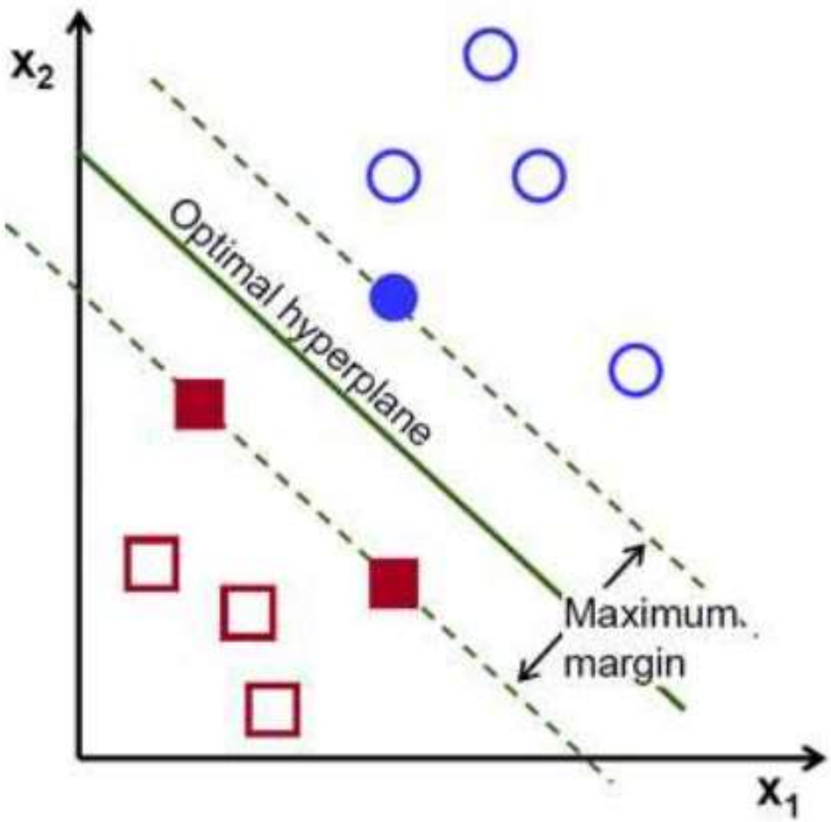
Data Analysis

- **K-means Algorithm**

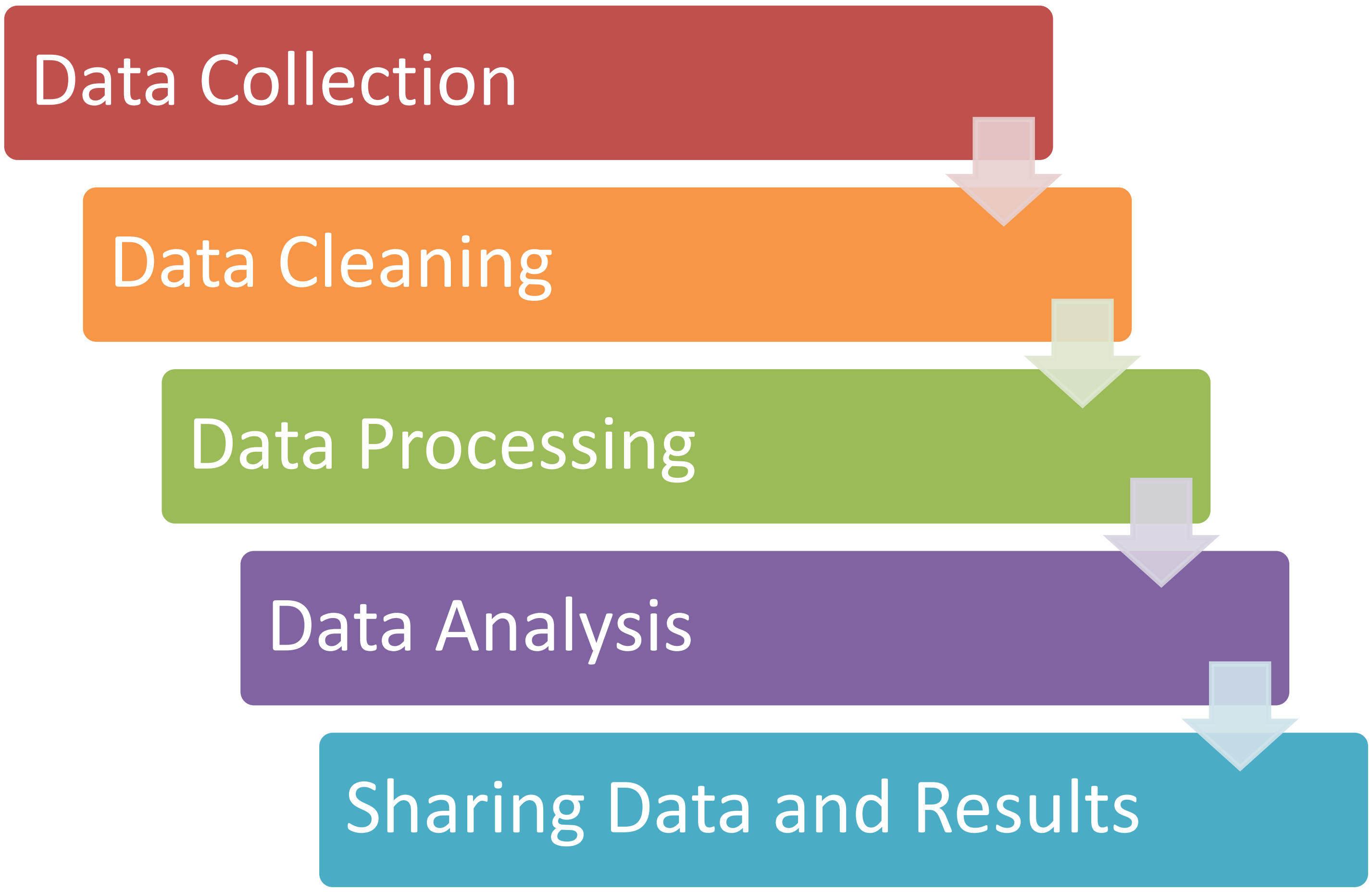


Data Analysis

- **Support Vector Machine**



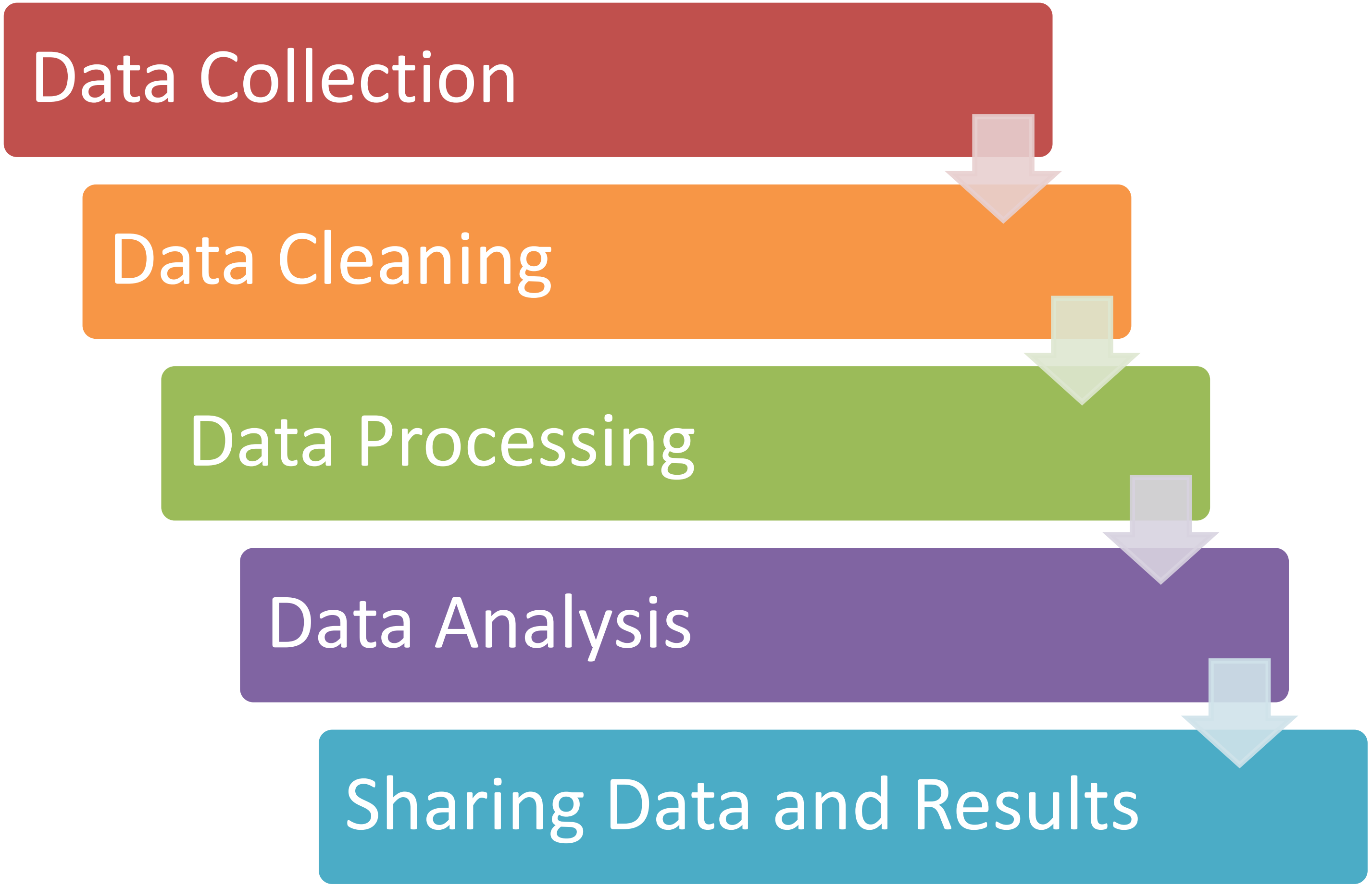
Research Pipeline



Sharing Data and Results

- Git + GitHub
  - Git: https://git-scm.com/downloads
  - https://github.com/Neo-Hao

Research Pipeline



Data Collection

Use Public Data Whenever Possible

U.S. DEPARTMENT OF EDUCATION

College Scorecard Data

Data Documentation College Scorecard

Join the Conversation: StackExchange

Data Insights

While there is variation in the amount of debt and fraction of students borrowing by sector, on average, **students at private for-profit two-year and four-year institutions have high rates of borrowing and their graduates often have large amounts of debt.** While debt per se may not be problematic where students are able to repay their loans, it should be paired with other data, such as completion rates and post-school earnings, to provide a more comprehensive picture of student outcomes.

Learn more in the Technical Paper

Introduction to the Data

Welcome to the College Scorecard Data site.

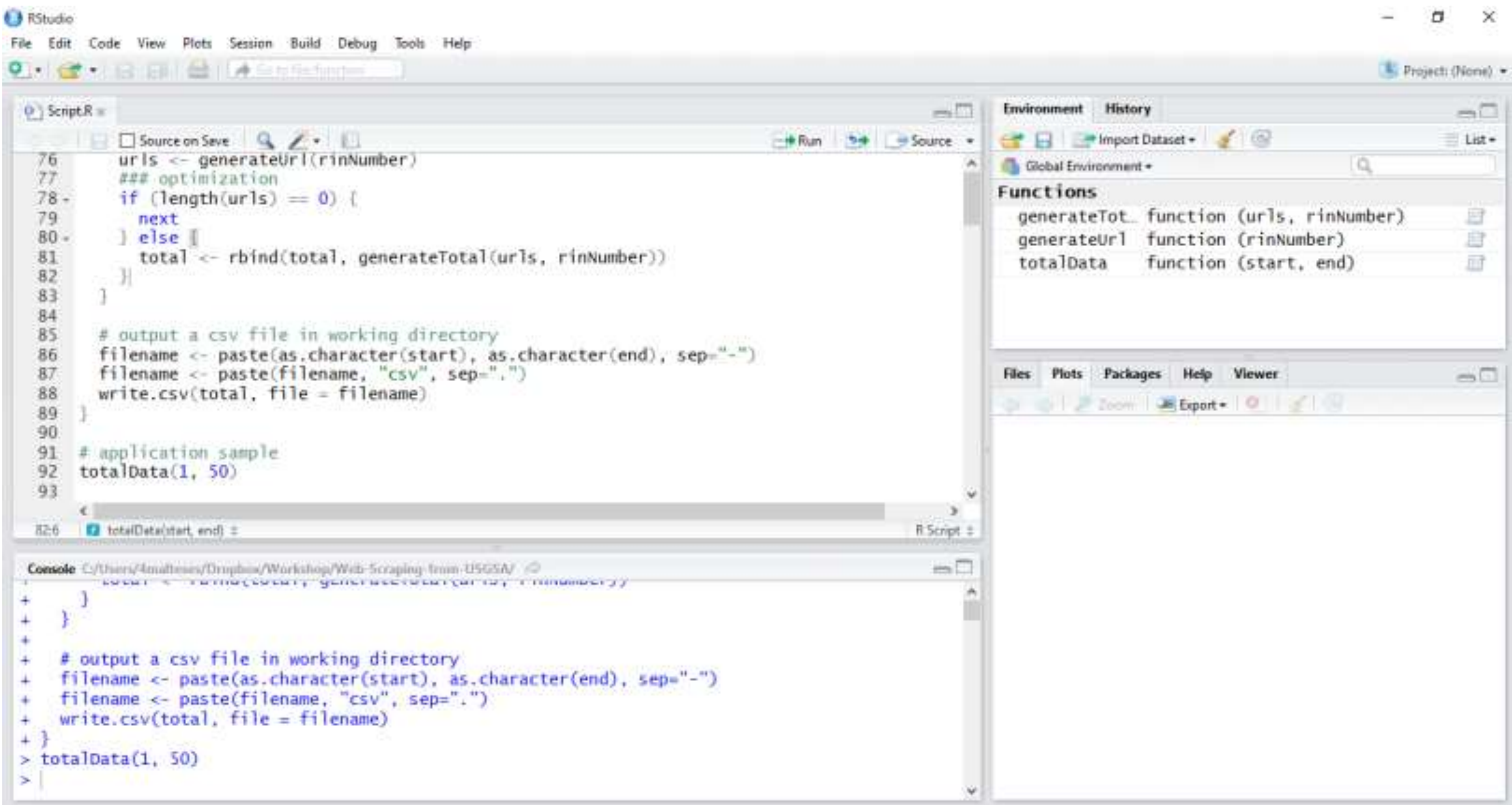
Here, you can get the data behind the College Scorecard, as well as other data on federal financial aid and earnings information. These data provide insights into the performance of schools eligible to receive federal financial aid, and offer a look at the outcomes of students at those schools.

Data Collection

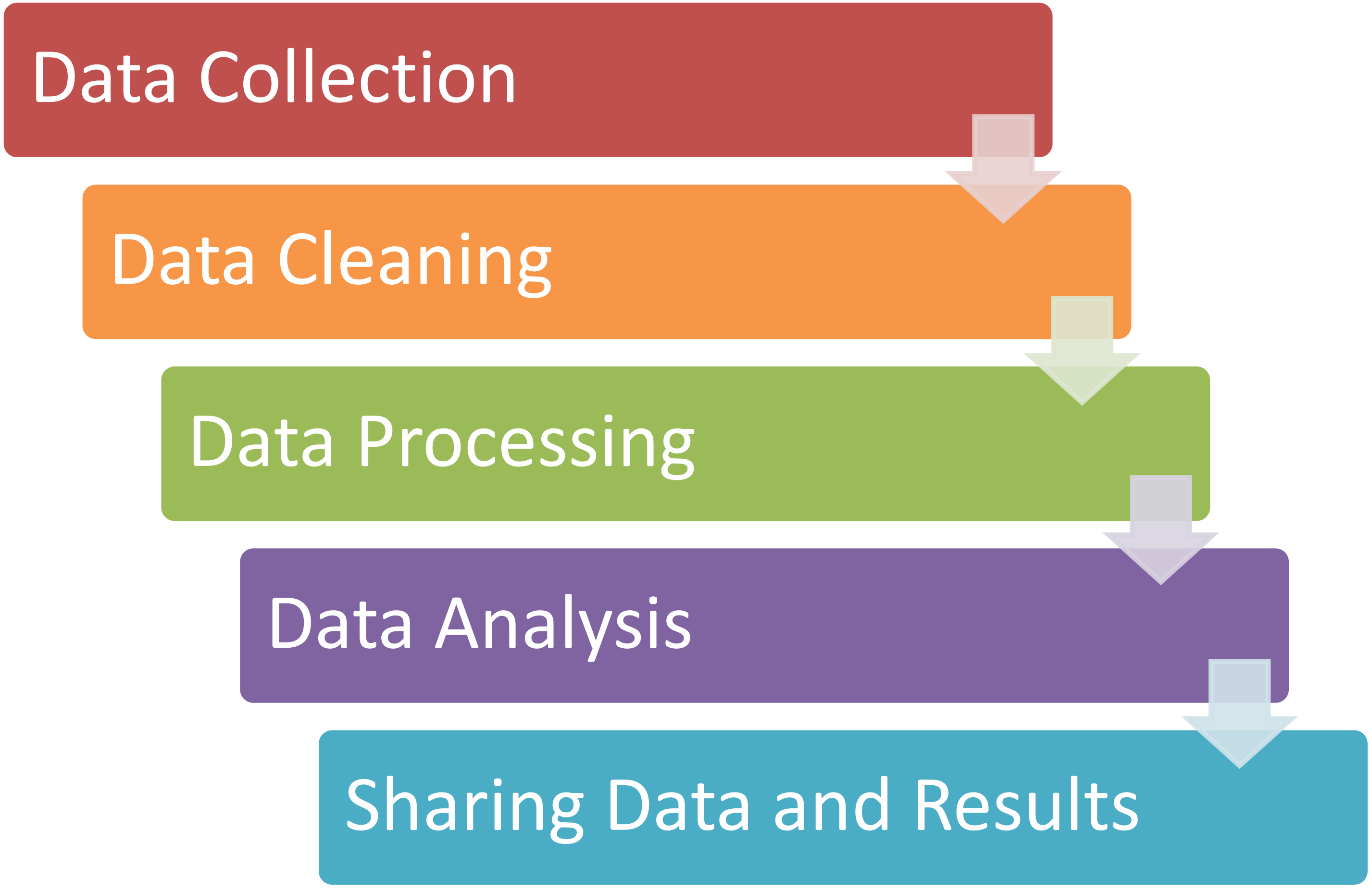
http://home.tobeneo.com/edutextmining/

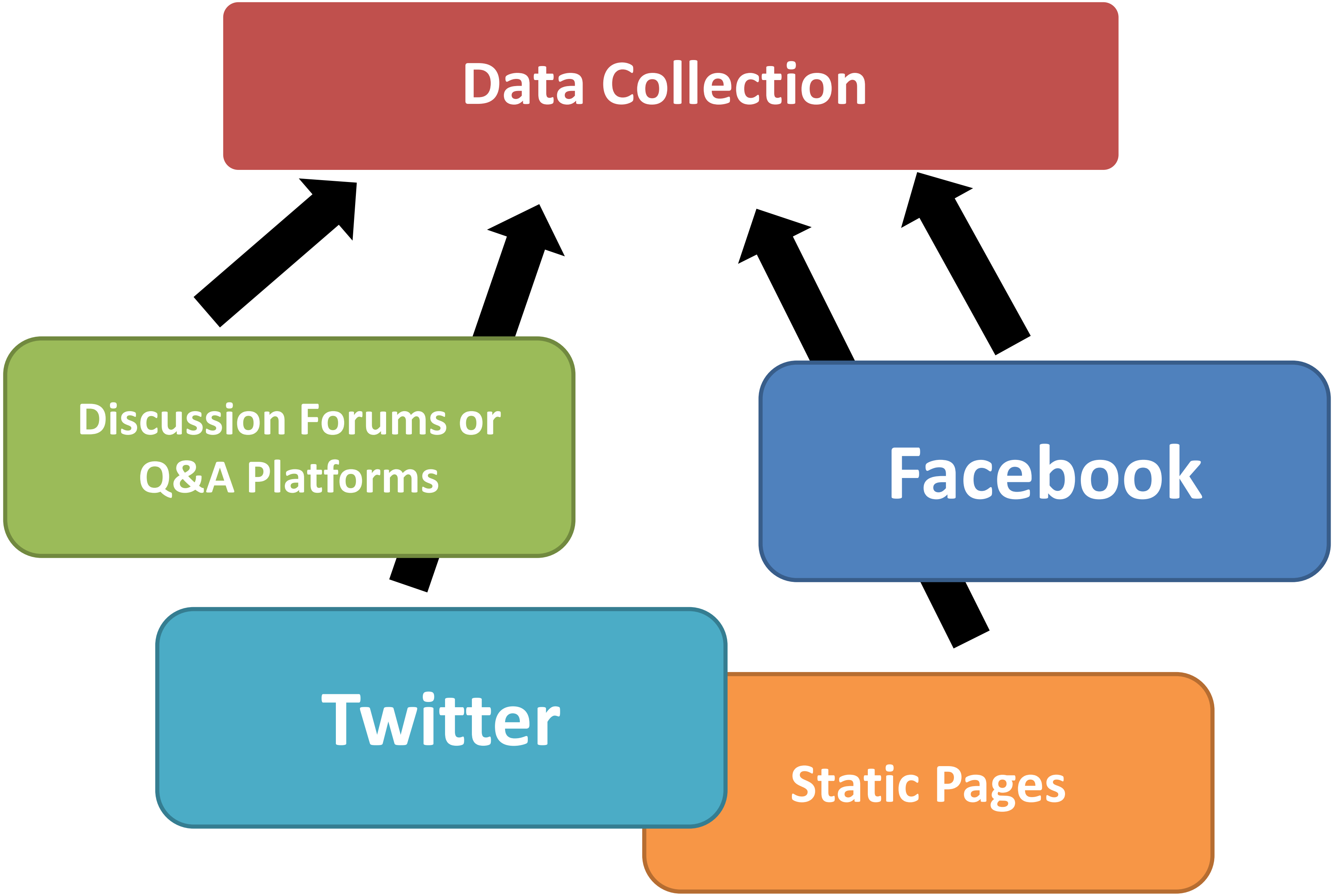
/// Download R and RStudio

Data Collection



Research Pipeline




Data Collection
Discussion Forums or Q&A Platforms
Facebook
Twitter
Static Pages

Data Collection

OFFICE of INFORMATION and REGULATORY AFFAIRS
OFFICE of MANAGEMENT and BUDGET
EXECUTIVE OFFICE OF THE PRESIDENT

Reginfo.gov

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

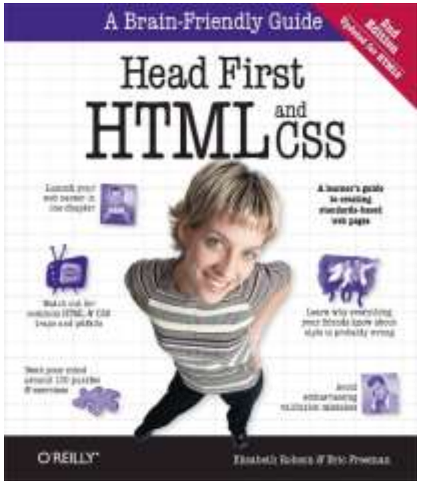
Data Collection

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

Data Collection

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

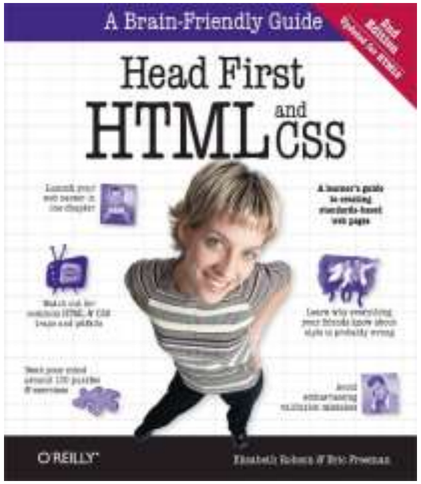
- **JSON**

```
{
    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:

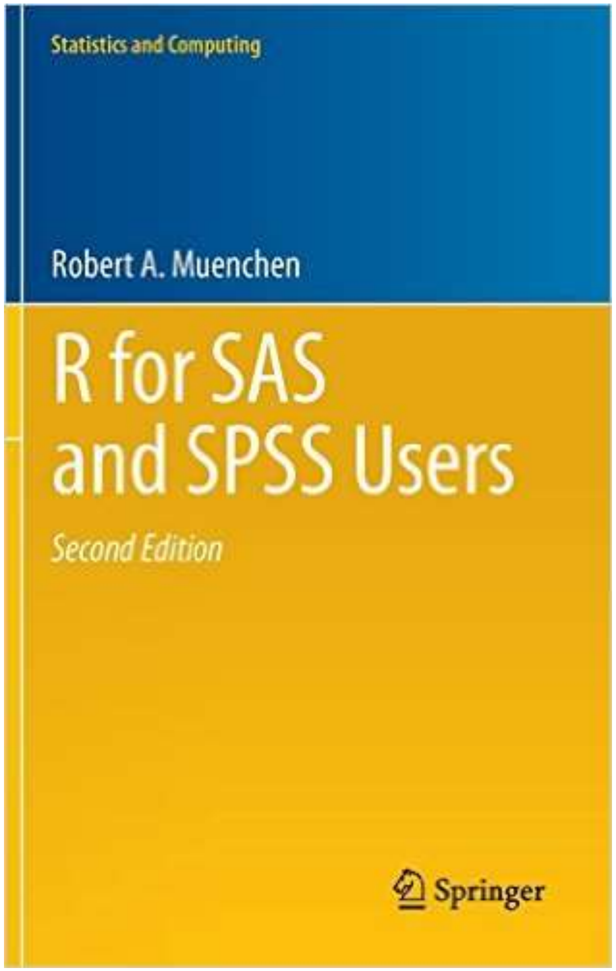
1. **A good understanding of HTML & CSS**
2. **A good understanding of XML & JSON**
3. **Familiar with Development Tools of Browsers**

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”**

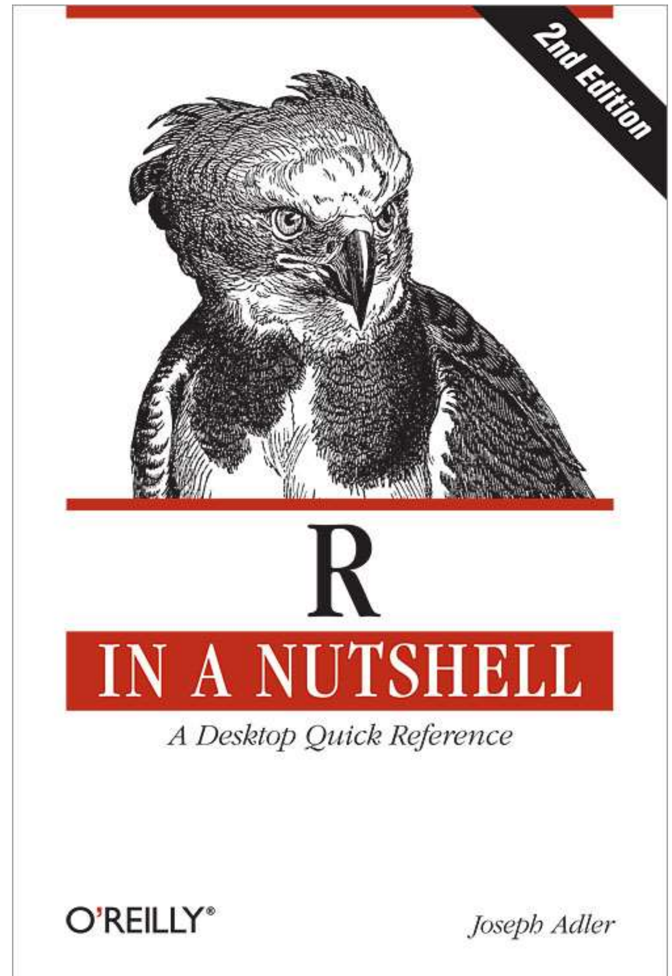
Data Collection



R for SAS and SPSS Users

Data Collection

R in a Nutshell



Data Collection

Data Collection

```r
getwd()

setwd(“C:/Users/John/Analysis”)
setwd(“/home/Analysis”)

setwd(“XXX/TwitterHashtagR/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**

Data Collection

Collect tweets by Web Scrapping

1. **Go to *Data Collection/parse_Tweets.R***
2. **Run lines 1-34, 76-77**

Data Collection

Collect tweets by Web Scrapping

1. **Go to *Data Collection/parse_Tweets.R***
2. **Run lines 1-34, 76-77**
3. **Practice: Do one web scrapping yourself**
   1. ***Search a hashtag using Twitter; keep scrolling down until you have all or enough number of tweets***
   2. ***Download the HTML page***
   3. ***……***

Thanks!

neohao@uga.edu