Data, Models, and First Steps

Digging into Data: Jordan Boyd-Graber

University of Maryland

January 28, 2013



COLLEGE OF INFORMATION STUDIES

Slides adapted from Dave Blei and Lauren Hannah

- The goals and ideas of the course
- Administrivia
- Getting started with Rattle and R

3

э.

3 1 4

A B A B A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

Outline



- How this course is organized
- 3 Administrivia and Introductions
- Introducing R and Rattle
- 5 Finding and using data
- 6 Showing off Rattle

🔵 Wrapup

(本間) (本語) (本語)

Data are everywhere.

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

3

(日) (同) (三) (三)

User ratings

<u>Ikiru</u> (1952)	UR	Foreign	0 ది ది ది ది ది
Junebug (2005)	R	Independent	이 <mark>슈 슈 슈</mark> ☆ ☆
La Cage aux Folles (1979)	R	Comedy	이 <mark>슈 슈 슈</mark> 슈
The Life Aquatic with Steve Zissou (2004)	R	Comedy	이 <mark>슈 슈 슈</mark> 슈
Lock, Stock and Two Smoking Barrels (1998)	R	Action & Adventure	0☆☆☆☆ ☆
Lost in Translation (2003)	R	Drama	0☆☆☆☆ ☆
Love and Death (1975)	PG	Comedy	이 <mark>슈 슈 슈</mark> 슈
The Manchurian Candidate (1962)	PG-13	Classics	이 <mark>슈 슈 슈</mark> 슈
Memento (2000)	R	Thrillers	⊘☆☆☆☆ ☆
Midnight Cowboy (1969)	R	Classics	0 **** *

- 2

・ロト ・四ト ・ヨト ・ヨト

Purchase histories

	Circese			
0.5/0.51 lb	Cabot Vermont Cheddar	0.51 lb	\$7.99/lb	\$4.07
	Dairy			
1/1	Friendship Lowfat Cottage Cheese (16oz)		\$2.89/ea	\$2.89
1/1	Nature's Yoke Grade A Jumbo Brown Eggs (1 dozen)		\$1.49/ea	\$1.49
1/1	Santa Barbara Hot Salsa, Fresh (16oz)		\$2.69/ea	\$2.69
1/1	Stonyfield Farm Organic Lowfat Plain Yogurt (32oz)		\$3.59/ea	\$3.59
	Fruit			
3/3	Anjou Pears (Farm Fresh, Med)	1.76 lb	\$2.49/lb	\$4.38
2/2	Cantaloupe (Farm Fresh, Med)		\$2.00/ea	\$4.00 S
	Grocery			
1/1	Fantastic World Foods Organic Whole Wheat Couscous (12oz)		\$1.99/ea	\$1.99
1/1	Garden of Eatin' Blue Corn Chips (9oz)		\$2.49/ea	\$2.49
1/1	Goya Low Sodium Chickpeas (15.5oz)		\$0.89/ea	\$0.89
2/2	Marcal 2-Ply Paper Towels, 90ct (lea)		\$1.09/ea	\$2.18 T
1/1	Muir Glen Organic Tomato Paste (6oz)		\$0.99/ea	\$0.99
1/1	Starkist Solid White Albacore Tuna in Spring Water (6oz)		\$1.89/ea	\$1.89

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 6 / 57

(日) (四) (三) (三) (三)

Document collections

SCIENCE: VERE INTERIOR NUMBER Martine	Science		
₩	SCIENCE Addres de la registration de la construction de la constructio	SCIENCE	
	SCIENCE	SCIENCE	Science

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

イロト イポト イヨト イヨト

Genomics



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

Neuroscience



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

Social networks



Finance

Dow Jones Composite Average (^DJA) - DJI 4,372.78 + 20.68 (0.47%) 1:48PM EDT



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ● □ ● ● ● ●

Add to Portfolio

Data can help us solve problems.

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 12 / 57

3

- - E - N

Image: A math a math

Will NetFlix user 493234 like Transformers?



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 13 / 57

Will NetFlix user 493234 like Transformers?







(日) (周) (三) (三)

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 14 / 57

3

How do you know?



Digging into Data: Jordan Boyd-Graber (UM

January 28, 2013 15 / 57

Group many images and determine the number of groups



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 16 / 57

Which genes are associated with a disease? How can expression values be used to predict survival?



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

Is it likely that this stock was traded based on illegal insider information?



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 18 / 57

Who will vote and for whom?



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 19 / 57

э

Subject: CHARITY. Date: February 4, 2008 10:22:25 AM EST To: undisclosed-recipients:; Reply-To: s.polla@yahoo.fr

Dear Beloved,

My name is Mrs. Susan Polla, from ITALY. If you are a christian and interested in charity please reply me at : (s.polla@yahoo.fr) for insight. Respectfully, Mrs Susan Polla.

くほと くほと くほと

From: [snipped] Subject: Superbowl? Date: January 28, 2013 8:09:00 PM EST To: jbg@umiacs.umd.edu, [snipped]

Anyone interested in coming by to watch the game? Beer and pizza, I'd imagine. Should be an exciting game!

Where are the faces?



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 22 / 57

Data contain patterns that can help us solve problems.

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 23 / 57

This Course (Digging into Data)

We will study algorithms that find and exploit patterns in data.

- These algorithms draw on ideas from statistics and machine learning.
- Applications include
 - natural science (e.g., genomics, neuroscience)
 - web technology (e.g., Google, NetFlix)
 - finance (e.g., stock prediction)
 - policy (e.g., predicting what intervention X will do)
 - and many others

不同下 不至下 不至下

This Course (Digging into Data)

We will study algorithms that find and exploit patterns in data.

- Goal: fluency in thinking about modern data analysis problems.
- We will learn about a suite of tools in modern data analysis.
 - When to use them
 - The assumptions they make about data
 - Their capabilities, and their limitations
- We will learn a language and process for of solving data analysis problems. On completing the course, you will be able to learn about a new tool, apply it data, and understand the meaning of the result.

Basic idea behind everything we will study

- Collect or happen upon data.
- Analyze it to find patterns.
- **③** Use those patterns to do something.



Outline



2 How this course is organized

- 3 Administrivia and Introductions
- Introducing R and Rattle
- 5 Finding and using data
- 6 Showing off Rattle

🔵 Wrapup

(本部) (本語) (本語)

Of course, there is no one way to organize such a broad subject. These concepts will recur through the course:

- Probabilistic foundations
- Supervised learning (more of this)
- Unsupervised learning (less of this)
- Methods that operate on discrete data (more of this)
- Methods that operate on continuous data (less of this)
- Representing data / feature engineering
- Evaluating models
- Understanding the assumptions behind the methods

Supervised vs. unsupervised methods



- **Supervised methods** find patterns in **fully observed** data and then try to predict something from **partially observed** data.
- For example, we might observe a collection of emails that are categorized into *spam* and *not spam*.
- After learning something about them, we want to take new email and automatically categorize it.

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

Supervised vs. unsupervised methods



- **Unsupervised methods** find **hidden structure** in data, structure that we can never formally observe.
- E.g., a museum has images of their collection that they want grouped by similarity into 15 groups.
- Unsupervised learning is more difficult to evaluate than supervised learning. But, these kinds of methods are widely used.

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

Discrete vs. continuous methods



- Discrete methods manipulate a finite set of objects
 - e.g., classification into one of 5 categories.
- Continuous methods manipulate continuous values
 - e.g.,prediction of the change of a stock price.

	discrete	continuous
supervised	classification	regression
unsupervised	clustering	dimensionality reduction

2

<ロ> (日) (日) (日) (日) (日)

Data representation



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 32 / 57

Understanding assumptions



- The methods we'll study make **assumptions** about the data on which they are applied. E.g.,
 - Documents can be analyzed as a sequence of words;
 - or, as a "bag" of words.
 - Independent of each other;
 - or, as connected to each other
- What are the assumptions behind the methods?
- When/why are they appropriate?
- Much of this is an art

Outline

What can we do with data?

2 How this course is organized

3 Administrivia and Introductions

- Introducing R and Rattle
- 5 Finding and using data
- 6 Showing off Rattle

🔵 Wrapup

3

(日) (同) (三) (三)

What you need for this course

- You need to use R and Rattle
- Helps to have a laptop to bring to class
- Math background
 - Not a machine learning course
 - Won't ask you to: prove anything, do integrals
 - > You do need to be comfortable with some notation (sums, variables)
 - Will ask you to: add, divide, count, take logs
- Computer / programming skills
 - Don't need to know how to program (might help)
 - But you do need to be comfortable with assigning objects to variables
 - Need to be comfortable with the concept of functions (variables, return, etc.)
 - We'll use the command line (but you don't need to be a ninja)

- 4 目 ト - 4 日 ト - 4 日 ト

- Sign up on Piazza (use a photo)
- Keep track of course webpage
- 5 late days
- Let me know about special needs

-

< 🗗 🕨

3

Course reading



- We will provide reading materials, mostly from the book.
- The reading will cover more than we cover in class.

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

・ ▲ 重 ▶ ▲ 重 ▶ 重 少 Q (? January 28, 2013 37 / 57 We will use Piazza to manage all communication

http://piazza.com/class#spring2012/cos424

- Questions answered within 1 day (hopefully sooner)
- Hosts discussions among yourselves
- Use for any kind of technical question
- Use for **most** administrative questions
- Can use to send us private questions too

- Explain what you're trying to do
- Give a minimal example
 - Someone else should be able to replicate the problem easily
 - Shouldn't require any data / information that only you have
- Explain what you think should happen
- Explain what you get instead (copy / paste or screenshot if you can)
- Explain what else you've tried

Me

- Third year assistant professor
 - iSchool and UMIACS
 - Offices: 2118C Hornbake / 3126 AV Williams
- First time teaching the class
- Born in Colorado (where all my family live)
- Grew up in Iowa (hometown: Keokuk, Iowa)
- Went to high school in Arkansas
- Undergrad in California
- Grad school in New Jersey
- Brief jobs in between:
 - Working on electronic dictionary in Berlin
 - Worked on Google Books in New York
- ying / jbg / jordan / boyd-graber

3

Outline

What can we do with data?

- 2 How this course is organized
- 3 Administrivia and Introductions

Introducing R and Rattle

- 5 Finding and using data
- 6 Showing off Rattle

🔵 Wrapup

< 回 > < 三 > < 三 >

- It's Free
- Standard for statistical data science
- Used by major corporations (Facebook and Google)
- You can go very deep (if you need to)

3

- - E - N

A (1) > A (1) > A

- It's easy
- Introduces the power of R through a GUI
- Does 90% of what most users need
- \bullet Slowly eases you in to the other 10%

3

-

< 🗗 🕨

Installing R



Download Installation File

http://watson.nci.nih.gov/cran_mirror/

- In Linux, many package managers already have it
- It is also installed on http://libcc.org

Installing Rattle

- Start R
- You'll see a command line



- This tells it too look for the package "rattle" and install it
- It will ask you to choose a mirror to download the file from; choose an MD one (it's in Bethesda)

Running Rattle for the First Time

```
R Console
> library(rattle)
Rattle: A free graphical interface for data mining with R.
Version 2.6.18 Copyright (c) 2006-2011 Togaware Pty Ltd.
Type 'rattle()' to shake, rattle, and roll your data.
> rattle()
```

- It will ask you to install a bunch of things
- Just say "yes"
- If you have problems, try exiting R and trying again

http://rattle.togaware.com/rattle-install-troubleshooting.html

Homework 0 (not for credit)

Install R and Rattle to try it out!

- 4 @ > - 4 @ > - 4 @ >

Outline

What can we do with data?

- 2 How this course is organized
- 3 Administrivia and Introductions
- Introducing R and Rattle
- 5 Finding and using data
 - 6 Showing off Rattle

🔵 Wrapup

3

(日) (同) (三) (三)

Where to get data?

- data.gov Obama initiative to get all government data in one place
- gapminder.org/data/ Global development data
- infochimps.org Pointers to interesting data
- http://bitly.com/bundles/hmason/1 A set of links to data
- http://www.ncbi.nlm.nih.gov/ National Center for Biotechnology Information
- http://www.ldc.upenn.edu/ Linguistic Data Consortium
- Wild, Wild, Web
- Devices
- Research

Where to get data?

- data.gov Obama initiative to get all government data in one place
- gapminder.org/data/ Global development data
- infochimps.org Pointers to interesting data
- http://bitly.com/bundles/hmason/1 A set of links to data
- http://www.ncbi.nlm.nih.gov/ National Center for Biotechnology Information
- http://www.ldc.upenn.edu/ Linguistic Data Consortium
- Wild, Wild, Web
- Devices
- Research
- First Homework: Find some data and describe it

Let's get some data

Download data from a weather station http://goo.gl/X6EpSOpen it in a text application

"Dates, "Location", "MinTemp", "MaxTemp", "Painfall", "Evaporation", "Sunshine", "WindGustDir", "WindGustSpeed", "WindDir9am", "Wi ", Temp3am", "Temp3am", "MainToday", "Bist Mem", PaintEmorrow" 2007-11-01, "Canherra", 1, 24:3,0:3,4,6:3, "NM", 30, "SW", "NN", 6,20, 66;29,1019,7,1015,7,7,14,423.6, "Ne", 3,6, "Yea"

2007-11-03, "Camberra", 13, 7, 23, 43, 65, 56, 0, 33, "NW", 55, "N", "NNE", 65, 62, 69, 1009, 55, 1007, 2, 6, 7, 15, 42, 02, "Yes", 39, 6, "Yes" 2007-11-04, "Camberra", 13, 7, 23, 43, 65, 56, 03, 3, "NW", 54, "NNE", "M, 30, 24, 62, 56, 1005, 51, 007, 2, 7, 13, 51, 14, 1, "Yes", 2.6, "Yes" 2007-11-04, "Camberra", 13, 15, 5, 39, 6, 7, 2, 9, 1, "NN", 54, "NNE", "G, 20, 26, 66, 49, 1018, 3, 1018, 3, 7, 11, 11, 15, 4, "Yes", 2.6, "Yes"

• Open it up in Excel or your favorite Spreadsheet

诸 we	🚡 weather.csv - OpenOffice.org Calc								
<u>D</u> atei	Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster Hilfe								
i 🗟 • 😕 🗔 🖙 i 🛃 🖴 🕒 i 💖 🜉 i 📈 🛍 🛍 • 🛷 i 🗐 • 🥙 • i 💩 🐉 👪 i 🏨									
💀 Arial 🔽 10 💌 F K U E E E 🗏 🎝 % 🐝 🗯									
A1	A1 \checkmark $\gamma_{\rm X} \Sigma = Date$								
	A	В	С	D	E	F	G	Н	
1	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	
2	11/01/07	Canberra	8	24.3	0	3.4	6.3	NW	
3	11/02/07	Canberra	14	26.9	3.6	4.4	9.7	ENE	
4	11/03/07	Canberra	13.7	23.4	3.6	5.8	3.3	NW	
5	11/04/07	Canberra	13.3	15.5	39.8	7.2	9.1	NW	
6	11/05/07	Canberra	7.6	16.1	2.8	5.6	10.6	SSE	
7	11/06/07	Canberra	6.2	16.9	0	5.8	8.2	SE	

Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 49 / 57

- Carryover from punchcards (easier to type)
- Each data item is separated by comma (or another character)
- Just about everything can use it (lowest common denominator)
 - Libraries in programming languages (starting with Fortran)
 - Spreadsheet
 - Exports from applications / devices

- 4 同 6 4 日 6 4 日 6

Outline

What can we do with data?

- 2 How this course is organized
- 3 Administrivia and Introductions
- Introducing R and Rattle
- 5 Finding and using data

6 Showing off Rattle

7) Wrapup

(本部) (本語) (本語)

- We'll only be showing off "coolness"
- Explanations later
- Goal: Get a sense of the data
- Goal: Predict when it will rain

3

Play with the weather data

Ex	ecute New	Dpen	Save	Report	Export	(X) Stop	Quit		
Data	Explore Test 1	ransform 0	Cluster As	sociate M	odel Evalu	uate Log			
Sou	rce: Spreads	heet 🔿 Al		DBC O	R Dataset	O RData	a File 🔘 I	_ibrary 🔿	Corpus 🔾 Script
Filer	name: 📄 weather	r.csv []	Separate	or: , Dec	imal: [. [✓ Header			
	Partition [70/15/1	5 500	0: [42	÷	View	Edit			
₽ ₽	Input elgnore	Weight Ca	d: 42	1	View	Target D	ata Type o () Categ	goric () N	umeric 🔿 Survival
⊡ F	Input Ignore	Weight Ca	a: 142 Iculator: Input	Target	Risk	Target D • Auto	ata Type o () Categ Ignore	goric () N	umeric O Survival
✓ F No. 1	Partition 70/15/1 Input Ignore Variable Date	Weight Ca Data Type Ident	a: 142 Iculator: Input	Target	Risk	Target D. • Auto Ident	ata Type	goric () N Weight ()	umeric O Survival Comment Unique: 366
 ✓ F No. 1 2 	Partition 70/15/1 Input Ignore Variable Date Location	Weight Ca Data Type Ident Constant	a: 42 Iculator: Input O	Target	Risk O	Target D. O Auto	ata Type	goric ON	umeric O Survival Comment Anique: 366 Unique: 1
✓ F No. 1 2 3	Artition 70/15/1 Input Ignore Variable Date Location MinTemp	Weight Ca Data Type Ident Constant Numeric	a: 42 culator: [Input 0 0	Target	Risk O O	Edit Target D	ata Type	yoric ON	Unique: 180

-2

<ロ> (日) (日) (日) (日) (日)

Ex	ecute New	Dpen	Save	Report E	kport :	Stop Quit
Data	Explore Test 1	Fransform	Cluster As	sociate Mo	del Evalua	te Log
Туре	e: 🔿 Summary	 Distrit 	outions O	Correlation	O Princip	al Components O Interactive
Num	neric: <u>≜</u> <u>C</u> lear	Plots per P	Page: 4	🕈 🗌 Annot	ate	Target: RainTomorrow
	Benford	d Bars B	enford Digit:	1≞⊙ ab	s 🔿 +ve 🔿) -ve
No.	Variable	Box Plot	Histogram	Cumulative	Benford	Min; Median/Mean; Max
3	MinTemp					-5.30; 7.45/7.27; 20.90
4	MaxTemp					7.60; 19.65/20.55; 35.80
5	Rainfall					0.00; 0.00/1.43; 39.80
6	Evaporation					0.20; 4.20/4.52; 13.80
7						0.00; 8.60/7.91; 13.60
9	WindGustSpeed					13.00; 39.00/39.84; 98.00
12	WindSpeed9am					0.00; 7.00/9.65; 41.00

- 2

<ロ> (日) (日) (日) (日) (日)

Finding connections ...

Distribution of Sunshine (sample) by RainTomorrow



Digging into Data: Jordan Boyd-Graber (UM

Data, Models, and First Steps

January 28, 2013 53 / 57

3

Making predictions ...

Execute New Open Save Report Export Stop Quit
Data Explore Test Transform Cluster Associate Model Evaluate Log
Type: O Tree O Forest O Boost O SVM O Linear O Neural Net O Survival O All
Target: RainTomorrow Algorithm: Traditional O Conditional Model Builder: rpart
Min Split: 20 12 Max Depth: 30 2 Priors: Include Missing
Min Bucket: 7 2 Complexity: 0.0100 2 Loss Matrix: Rules Draw
Summary of the Decision Tree model for Classi%cation (built using 'rpart'):
n= 256
node), split, n, loss, yval, (yprob) * denotes terminal node
 roct 256 41 No (0.83984375 0.16015625) Pressure3pm>=011.9 204 16 No (0.92156863 0.07843137) Cloud3pm<7.5 195 10 No (0.94871795 0.05128205) * Cloud3pm>=7.5 9 3 Yes (0.3333333 0.66666667) * Pressure3pm<1011.9 52 25 No (0.51923077 0.48076923) Sunshine~8.85 25 5 No (0.8000000 0.20000000) * Sunshine<8.85 27 7 Yes (0.25925926 0.74074074) *

(日) (四) (王) (王) (王)

Making predictions ...

Decision Tree weather.csv \$ RainTomorrow



Digging into Data: Jordan Boyd-Graber (UM

54 / 57

Outline

- 3 Administrivia and Introductions
- Introducing R and Rattle
- Finding and using data



Wrapup

3

イロト イヨト イヨト イヨト

A statistician's manifesto (From T. Hastie, via J. McAuliffe)

- Understand the ideas behind the statistical methods, so you know how to use them, when to use them, when *not* to use them.
- Complicated methods build on simple methods. Understand simple methods first.
- The results of a method are of little use without an assessment of how well or poorly it is doing.

- What are probability distributions
- How to compute probabilities
- Properties of distributions

3

A (10) A (10) A (10)