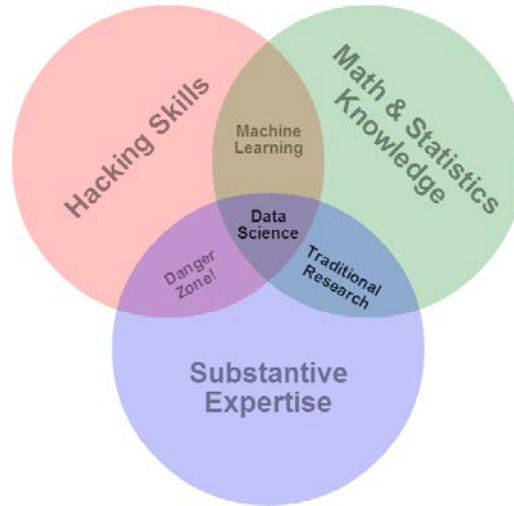


google: conway data science aws
And explore the areas of the below Venn Diagram



https://s3.amazonaws.com/aws.drewconway.com/viz/venn_diagram/data_science.html

1.1 Mon 7/13

Course Introduction

Dr. Raja Sooriamurthi
Information Systems Program
Carnegie Mellon University
raja@cmu.edu

Applied Data Science Program
2015

Daily Schedule + Course Structure

- 9:30 – 10:30 Data Science / Lab
- 10:30 – 10:45 Break
- 10:45 – 12:00 Data Science / Lab
- 12:00 – 1:15 Lunch
- 1:30 Catch Shuttle to UPMC
- 2:00 – 4:40 UPMC

Course Themes

- Problem Domain
- Questions / Pain Points
- Data
- Data Science Techniques
- Communicate Actionable Insights

Now your turn ... an Information Sleuth

On the card, write
a little known fact about yourself
that you would like to share with the
group

e.g.,

- ✓ As a child growing up in India I had 3 parrots
- ✓ The names of my two boys (11 and 9) are synonyms that mean *happiness* in my native language

Learning Objectives

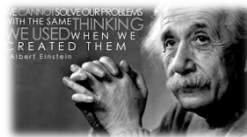
- What is data science?
- What are the takeaways from the Big Data documentary?
- What is Data Driven Decision Making?
- Why now the Data Revolution?
- What can we do with (lots of) data?

Doing Science ...

Experimental



Theoretical



Computational



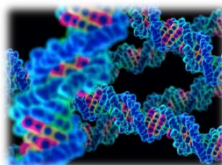
Tianhe-2
33.86 petaflops

Data Driven



BBC: The Age of Big Data

- Predictive Policing
- Healthcare
- Finance
- Personalization
- Astronomy



Quiz: The Age of Big Data

- Predictive Policing
- Healthcare
- Finance
- **Personalization**
- Astronomy



- What was the problem?
- How was data used to address that problem?

Activity: Data Driven Decision Making

Problem

- Fatal Traffic Accidents

Data

Analysis

Policy

Data Driven Decision Making

Problem

- Fatal Traffic Accidents

Data

- Fatalities, Location, time of day, weather, Vehicle details, Driver details (age, BAC), Speed, Seat-belts?

Analysis

- Location?
- Speed?

Policy

- Education
- Penalty

Activity: Data Driven Decision Making

Problem

- What societal problem would you like to address?

Data

- What data would you need to collect?

Analysis

- What would you look for?

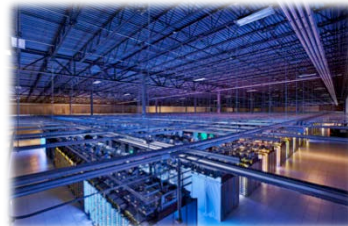
Policy

- ?

The Data Revolution

The Hardware Revolution

- Mainframes
- Minis
- Super Computers
- Personal Computers
 - Handheld
 - Wearable
- Warehouse Computing



The Software Revolution

- Paradigms
 - Imperative
 - Object-oriented
 - Functional
 - Concurrent
- Methodologies
 - Agile
 - Lean
- Models
 - Shrink wrapped
 - Licensed
 - SaaS

Agile Method



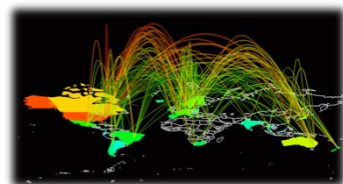
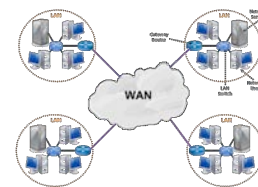
The Communication Revolution



OSI Model

Application		Application
Presentation		
Session		
Transport		Transport
Network		Internetwork
Data Link		
Physical		Link and Physical

TCP / IP



The Data Revolution

- We are generating It
 - Automation + Internet
- We are collecting it
- We are storing it
 - 4 TB @ \$140 (¢3.5 / GB)
- We are processing it

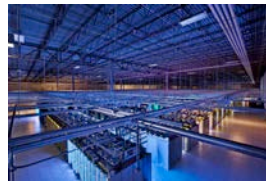


CDC 7600

\$5M in 1976 (\$32)
vs.
\$150



iPhone4



Seagate Desktop ST4000DM000 HDD
4 TB SATA 6Gb/s NCQ 64MB Cache
3.5-Inch Internal Ba

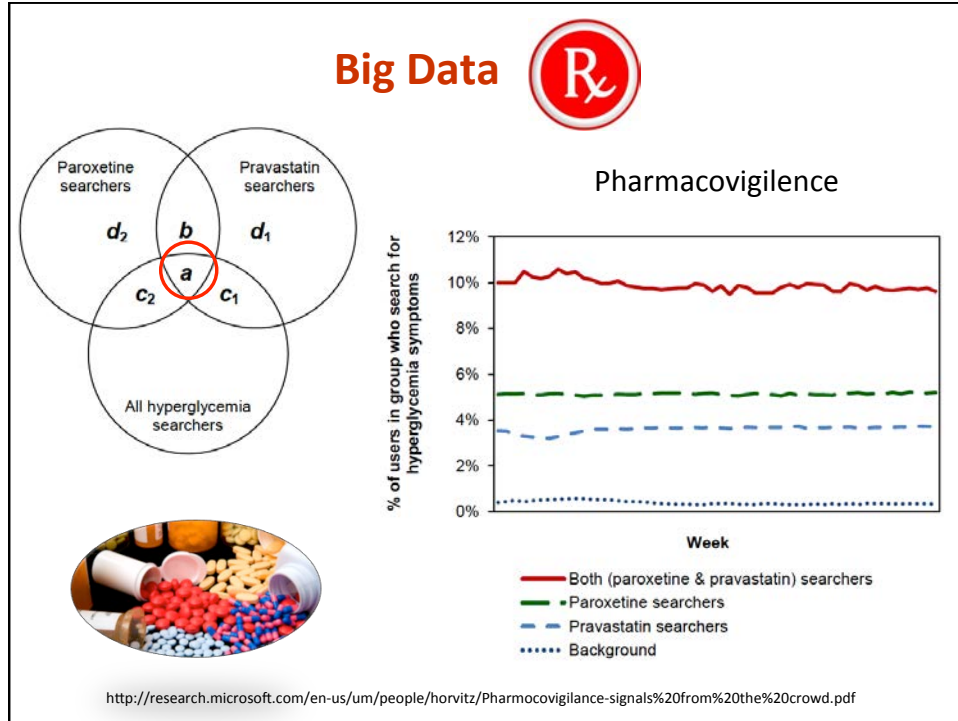
by Seagate
★★★★☆ 4,108 custom
| 51 answered questions

List Price: \$209.99
Price: \$139.99 Prime



A Google
server warehouse

More finer pixels (data) better information



Can you guess?

நான்கு ஓட்டங்கள் மற்றும் எங்கள் தந்தைகள் முன்பு ஏழு ஆண்டுகள் படைப்பில் அனைவரும் சமம் என்று சுதந்திரம் உணரப்படுகின்றது, மற்றும் முன்மொழிவாகும் அர்ப்பணிக்கப்பட்ட, இந்த கண்டத்திலும் ஒரு புதிய தேசத்தை மாறி மாறி வந்தது. இப்போது நாம் அந்த தேசத்தை, அல்லது எந்த நாட்டின் மிகவும் சிந்தித்து மிகவும் அர்ப்பணிக்கப்பட்ட, நீண்ட பொறுத்துக்கொள்ள முடியும் என்பதை பரிசோதித்து, ஒரு பெரிய உள்நாட்டு போரில் ஈடுபட்டு வருகின்றனர். அந்த போர் ஒரு பெரும் போர்-களத்தில் சந்தித்தார். நாம் இங்கே அந்த நாடு வாழ வேண்டும் என்று அவர்கள் உயிர் நீத்த அந்த ஒரு இறுதி நின்றுபோகும் இடத்தில் போல், அந்த துறையில் ஒரு பகுதியை ஒதுக்கி வந்திருக்கிறேன். அதை நாம் செய்ய வேண்டும் என்று முற்றிலும் பொருத்தமான சரியான உள்ளது.

ஆனால், ஒரு பெரிய பொருளில், நாம் இந்த நிலத்தில் உள்ளீடில் முடியாது, நாம் புனிதமாக்கு முடியாது, அர்ப்பணிக்கவும் முடியாது. இங்கே போராடியது யார் நாடு மற்றும் இறந்த தைரியமாக ஆண்கள்,, இதுவரை சேர்க்க அல்லது குறை நமது ஏழை சக்தி மேலே, அதை கும்பாபிஷேகம். உலக கொஞ்சம் கவனிக்க, அல்லது நீண்ட நாம் இங்கே என்ன ரூபகம், ஆனால் அவர்கள் இங்கு என்ன செய்தார்கள் மறக்க முடியாது. அதை இங்கே போராடிய அவர்கள் இதுவரை அப்படி nobly மேம்பட்ட கொண்ட முடிக்கப்படாத வேலை இங்கே அர்ப்பணிக்கப்பட்ட வேண்டும், மாறாக, எங்களுக்கு வாழ்க்கை இல்லை. அது இங்கே நமக்கு முன் மீதமுள்ள பெரும் பணியை அர்ப்பணிக்க வேண்டும் நமக்கு மாறாக இந்த சிறப்பு இறந்த இருந்து நாங்கள் அவர்களுக்கு நாம் இங்கு மிகவும் சரி பக்தி, அந்த கடைசி முழு நடவடிக்கை கொடுத்த அந்த காரணம் அதிகரித்துள்ளது பக்தி எடுத்து இந்த இறந்த செய்யாதிருப்பாயாக இந்த நாடு, கடவுள் கீழ், மக்கள் சுதந்திரம் மற்றும் அரசாங்கம் ஒரு புதிய பிறப்பு வேண்டும் என்ற வீண் என்று இறந்ததாக, மக்கள், மக்கள், பூமியில் இருந்து அழிந்து.

How about this?

5,87歲以前，我們的祖先在這片大陸帶來了一個新國家，它孕育於自由，並且獻身給一種理念，所有的人都生來平等。現在我們正從事一場偉大的內戰，這個國家，或者任何一個國家如此構思和專用的測試是否能夠長久存在下去。我們今天相聚在這場戰爭中的一個偉大戰場上。我們來到這裡，奉獻的那場部分，對於那些誰在這裡為使這個國家能夠生存下去而獻出自己的生命作為最後的安息之地。這是完全適當的，我們應該這樣做。

但是，從更廣泛的意義上來說，我們不能奉獻，我們不能夠聖化，不能夠神化這片土地。的勇士們，活著的和死去的，在這裡戰鬥過的人，都奉獻了，遠非我們微薄的力量所能增減的。全世界將很少注意到，也不會長久記住我們在這裡所講的話，但它永遠不會忘記他們在這裡做的。相反，這是對我們的生活，獻身未完成的工作，他們戰鬥過的人迄今有如此高尚地推進。倒是我們應該在這裡獻身的偉大任務擺在我們面前的，其餘要從這些光榮的死者身上，我們需要更多的獻身精神，因為他們曾作最後全部的奉獻，我們在這裡下定最大的決心，這些死不得徒勞的，這個國家在上帝的保佑下，應當有一個新誕生的自由和政府的人已經死亡，由人民，為人民，不得從地球上消失。

This?

Quatre scores et il ya sept ans nos pères ont sur ce continent une nouvelle nation, conçue dans la liberté, et dédié à la proposition selon laquelle tous les hommes sont créés égaux. Maintenant, nous sommes engagés dans une grande guerre civile, tester si cette nation, ou toute nation ainsi conçue et vouée, ne peut durer longtemps. Nous sommes réunis sur un champ de bataille beaucoup de cette guerre. Nous sommes venus pour consacrer une partie de ce domaine, comme un lieu de repos final pour ceux qui ont donné leur vie ici que cette nation puisse vivre. Il est tout à fait approprié et opportun que nous devrions le faire.

Mais, dans un sens plus large, nous ne pouvons pas consacrer, nous ne pouvons pas consacrer, nous ne pouvons pas sanctifier cette terre. Les braves, vivants et morts, qui ont lutté ici, l'ont consacré, au-dessus de notre faible pouvoir d'ajouter ou de retrancher. Le monde va peu à noter, ni souviendra longtemps de ce que nous disons ici, mais il ne pourra jamais oublier ce qu'ils ont fait ici. C'est pour nous, les vivants, plutôt, être consacrée ici à l'œuvre inachevée que ceux qui ont combattu ici ont jusqu'ici si noblement. Il s'agit plutôt pour nous d'être ici dédié à la grande tâche qui reste devant nous, que ces morts honorés nous prenons la dévotion portée à cette cause pour laquelle ils ont donné la pleine mesure de leur dévotion que nous sommes ici très à résoudre que ces morts ne doivent pas sont morts en vain, que cette nation, sous Dieu, aura une nouvelle naissance de la liberté, et que le gouvernement du peuple, par le peuple, pour le peuple, ne disparaîtra pas de la terre.

This?

Hace ochenta y siete años nuestros padres fundaron en este continente una nueva nación, concebida en libertad y dedicada a la proposición de que todos los hombres son creados iguales. Ahora estamos empeñados en una gran guerra civil, probando si esta nación, o cualquier nación así concebida y consagrada, puede perdurar. Estamos reunidos en un gran campo de batalla de esa guerra. Hemos venido a dedicar una porción de ese campo, como un lugar de descanso final para aquellos que aquí dieron sus vidas para que esta nación pudiera vivir. Es perfectamente justo y propio que deberíamos hacer esto.

Pero, en un sentido más amplio, no podemos dedicar, no podemos consagrar, no podemos santificar este suelo. Los valientes hombres, vivos y muertos, que lucharon aquí lo han consagrado, muy por encima nuestro pobre poder de añadir o restar. El mundo notará poco, ni mucho tiempo recordará lo que decimos aquí, pero nunca podrá olvidar lo que ellos hicieron aquí. Es para nosotros, los vivos, más bien, a dedicarnos ahora a la obra inacabada que quienes aquí lucharon dejaron tan noblemente avanzado. Es más bien para nosotros estar aquí dedicado a la gran tarea que nos queda por delante—que de estos muertos honrados tomamos devoción mayor a la causa por la que ellos dieron hasta la última y de devoción que aquí decidamos que estos muertos no han muerto en vano, que esta nación, bajo Dios, tendrá un nuevo nacimiento de la libertad y el gobierno que el pueblo, por el pueblo y para el pueblo, no perecerá de la tierra.

Lincoln's Gettysburg address

Four score and seven years ago our fathers brought forth on this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle-field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we can not dedicate, we can not consecrate, we can not hallow this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us—that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion—that we here highly resolve that these dead shall not have died in vain—that this nation, under God, shall have a new birth of freedom—and that government of the people, by the people, for the people, shall not perish from the earth.

Data Science

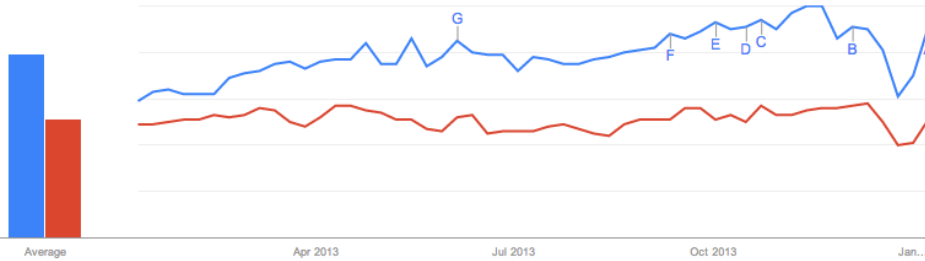


Simple models and a lot of data trump more elaborate models based on less data.
- Peter Norvig

Four score and seven years ago our fathers brought forth on this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle-field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this ...

நான்கு ஓட்டங்கள் மற்றும் எங்கள் தந்தைகள் முன்பு ஏழு ஆண்டுகள் படைப்பில் அனைவரும் சமம் என்று சுதந்திரம் உணரப்படுகின்றது, மற்றும் முன்மொழிவாகும் அர்ப்பணிக்கப்பட்ட, இந்த கண்டத்திலும் ஒரு புதிய தேசத்தை மாறி மாறி வந்தது. இப்போது நாம் அந்த தேசத்தை, அல்லது எந்த நாட்டின் மிகவும் சிந்தித்து மிகவும் அர்ப்பணிக்கப்பட்ட, நீண்ட பொறுத்துக்கொள்ள முடியும் என்பதை பரிசோதித்து, ஒரு பெரிய உள்நாட்டு போரில் ஈடுபட்டு வருகின்றனர். அந்த போர் ஒரு பெரும் போர்-களத்தில் சந்தித்தார். நாம் இங்கே அந்த நாடு வாழ வேண்டும் என்று அவர்கள் உயிர் நீத்த அந்த ஒரு இறுதி நின்றுபோகும் இடத்தில் போல், அந்த துறையில் ஒரு பகுதியை ஒதுக்கி வந்திருக்கிறேன். அதை நாம் செய்ய வேண்டும் என்று முற்றிலும் பொருத்தமான சரியான உள்ளது. ...

Big Data



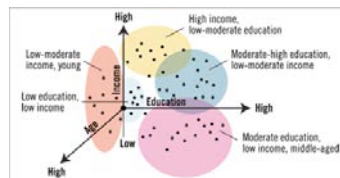
- Ngram viewer
- Google Flu trends
- Other trends

Activity: Google Trends

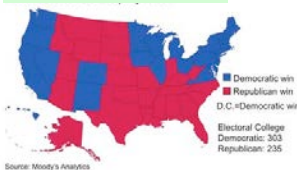
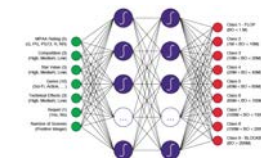
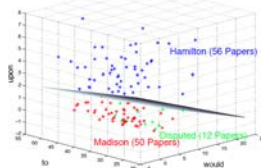
- Select two categories from the Oscar awards
- Look up the nominees and do a Google trends analysis. Did the trends predict the ultimate winner? Try to explain what you observe, even if your results are not quite right on.
- What is the trend in interest in Samsung mobile phones? What is the interest in Nokia phones? In iPhones? Where is interest the greatest or least? Can you explain why?
- Select a favorite consumer product, note trend in interest over time and try to explain what you observe.

What are some classical analytics tasks?

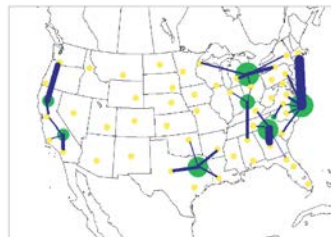
- Association
- Classification
- Clustering
- Design
- Optimization
- Prediction
- Recommendation
- Search/ranking



Separating Plane for the Federalists Papers - 1788 (Fung)



Source: Moody's Analytics



Resources to Learn Python

- Codecademy
<http://www.codecademy.com/en/tracks/python>
- Learn Python
<http://www.learnpython.org/>
- Learn Python the Hard way
<http://learnpythonthehardway.org/>

Next Time

- Introduction to Python and the Ipython notebook environment
- Prep: Work through some of the learning Python resources. Codecademy is a useful starting point.
- <http://www.codecademy.com/en/tracks/python>
 - Python Syntax
 - Strings & Console Output

Setting up your Python Environment

1. google: miniconda
2. Download and install miniconda3
3. % conda update conda
4. % conda install ipython
5. % conda install ipython-notebook

The % refers to the (Windows or MacOS) prompt at the command line