

# Data Revolution

A journey from Red Bluff to Big Data

Clark Fitzgerald - @clarkfitzg

Red Bluff High School

21 August 2014

# Goals Today

- What I wish I would've known...
- A sense of possibility around math and data

Personal Journey

Big Data

Opportunity

Wrapping up

# Outline

Personal Journey

Big Data

Opportunity

Wrapping up

# The Army



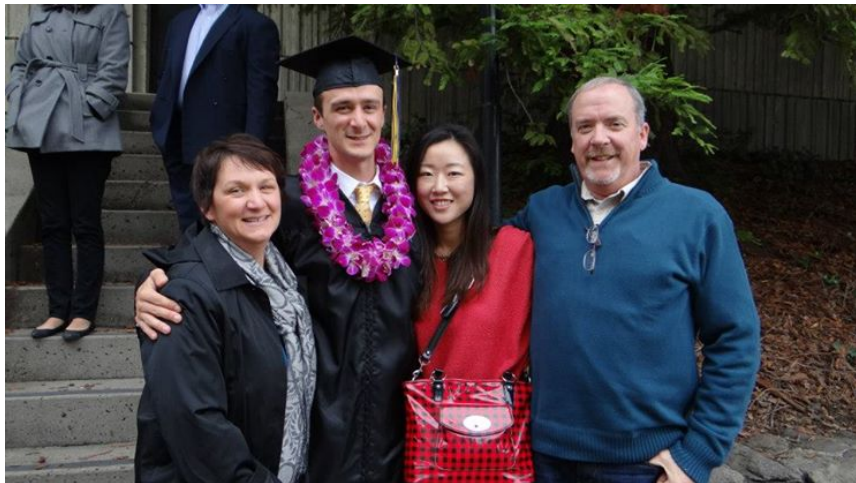
# Mountain Guide



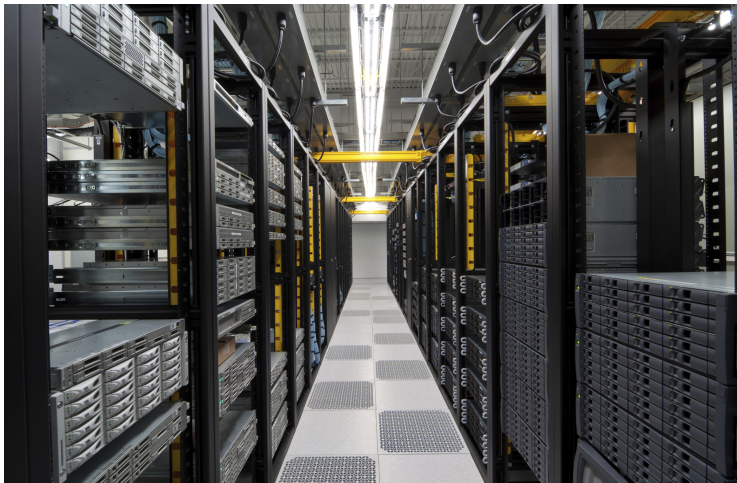
## Marriage to Yeji



# Math at Shasta then UC Berkeley



# Cisco Systems





## UC Davis - Phd in Statistics



# Outline

Personal Journey

Big Data

Opportunity

Wrapping up

# What is Data?

Data is information.

- Facebook friends
- Amazon product pricing
- bank account balance
- text messages
- traffic sensors
- daily sales at WalMart
- words in a book
- pixels in an image

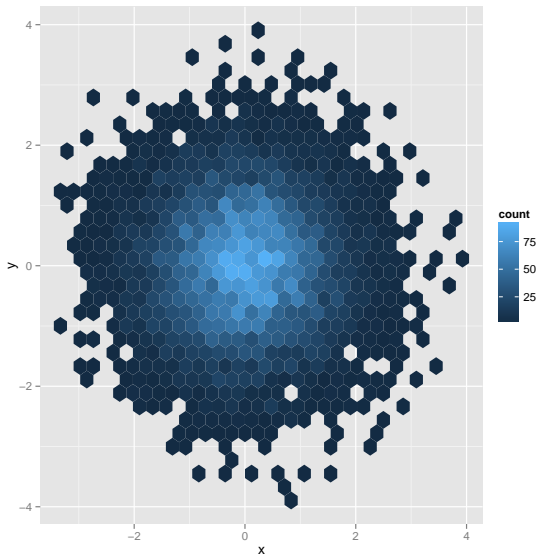
## An example of data

Pixels in an image, transformed into a table:

```
$ head python_logo.csv
x,y,blue,green,red
0,0,255,255,255
1,0,255,255,255
2,0,255,255,255
3,0,255,255,255
4,0,255,255,255
5,0,255,255,255
6,0,255,255,255
7,0,255,255,255
8,0,255,255,255
```

# Normal Data

10,000 normally distributed points



# Outline

Personal Journey

Big Data

Opportunity

Wrapping up

# It's like the Gold Rush!

**A NEW AND MAGNIFICENT CLIPPER FOR SAN FRANCISCO.**  
**MERCHANTS' EXPRESS LINE OF CLIPPER SHIPS!**  
Loading none but First-Class Vessels and Regularly Dispatching the greatest number.  
**THE SPLENDID NEW OUT-AND-OUT CLIPPER SHIP**



**CALIFORNIA**  
HENRY BARBER, Commander, AT PIER 13 EAST RIVER.

This elegant Clipper Ship was built expressly for this trade by Samuel Hall, Esq., of East Boston, the builder of the celebrated Clippers "SURPRISE," "GAMECOCK," "JOHN GILPIN," and others. **She will fully equal them in speed!** Unusually prompt dispatch and a very quick trip may be relied upon. Engagements should be completed at once.

Agents in San Francisco,  
Messrs. DE WITT KITTLE & CO. }

**RANDOLPH M. COOLEY, 88 Wall Street, Tontine Building.**

NEEDHAM & CO., PRINTERS.

Why is it like the Gold Rush?



Because few have the skills.

- Statistics
- Programming
- Mathematics

## Code to process a large data set

```
import csv
import collections

count = collections.Counter()

with open('Train.csv') as Train:
    for row in csv.DictReader(Train):
        # split tag column on spaces
        tags = row['Tags'].split(' ')
        count.update(tags)
```

You can start learning this stuff *today!*  
Everything is free and open access.



# Outline

Personal Journey

Big Data

Opportunity

Wrapping up

# Goals Today Revisited

- There is so much to do!  
Find something you like.
- This will help you in nearly *any* field.

# Questions?