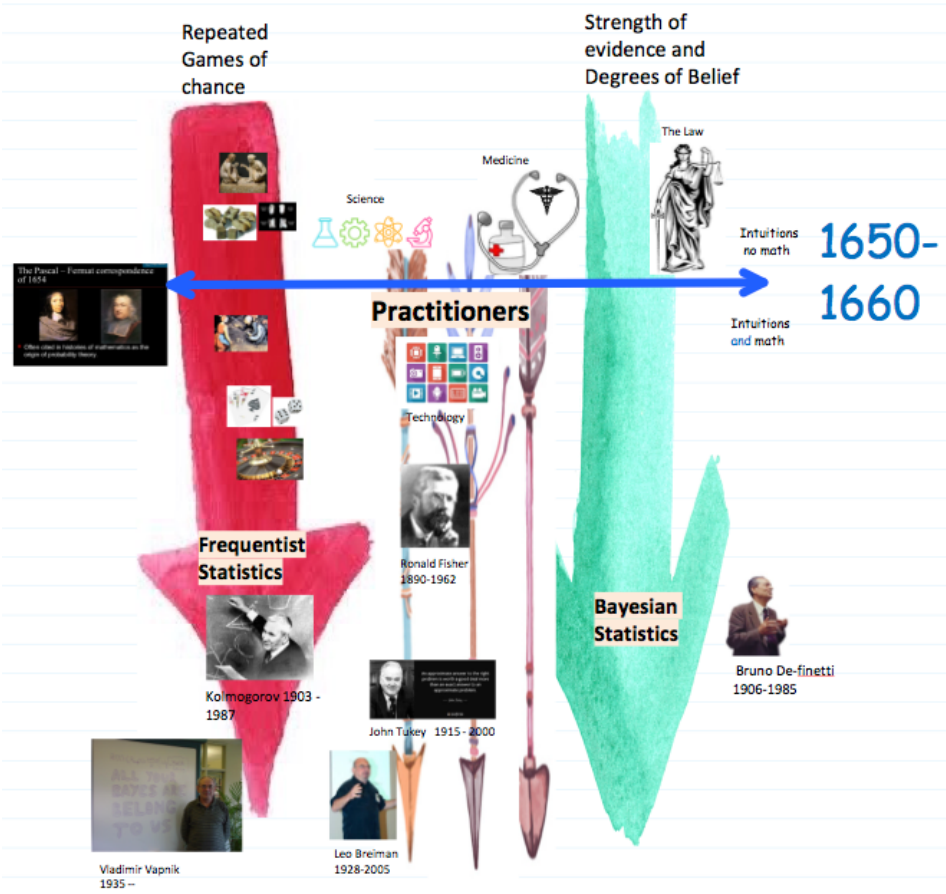
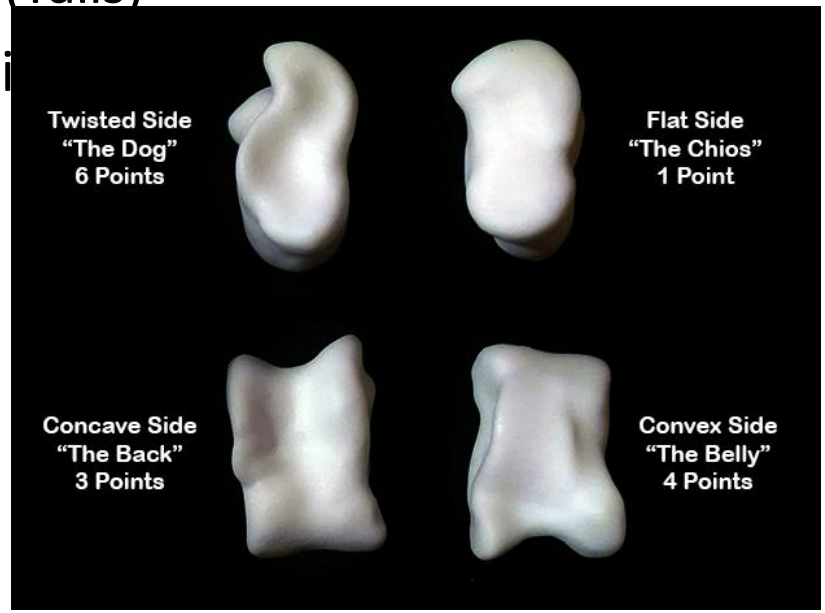
A short history of probability And Statistics

Games of chance VS. Strength of evidence



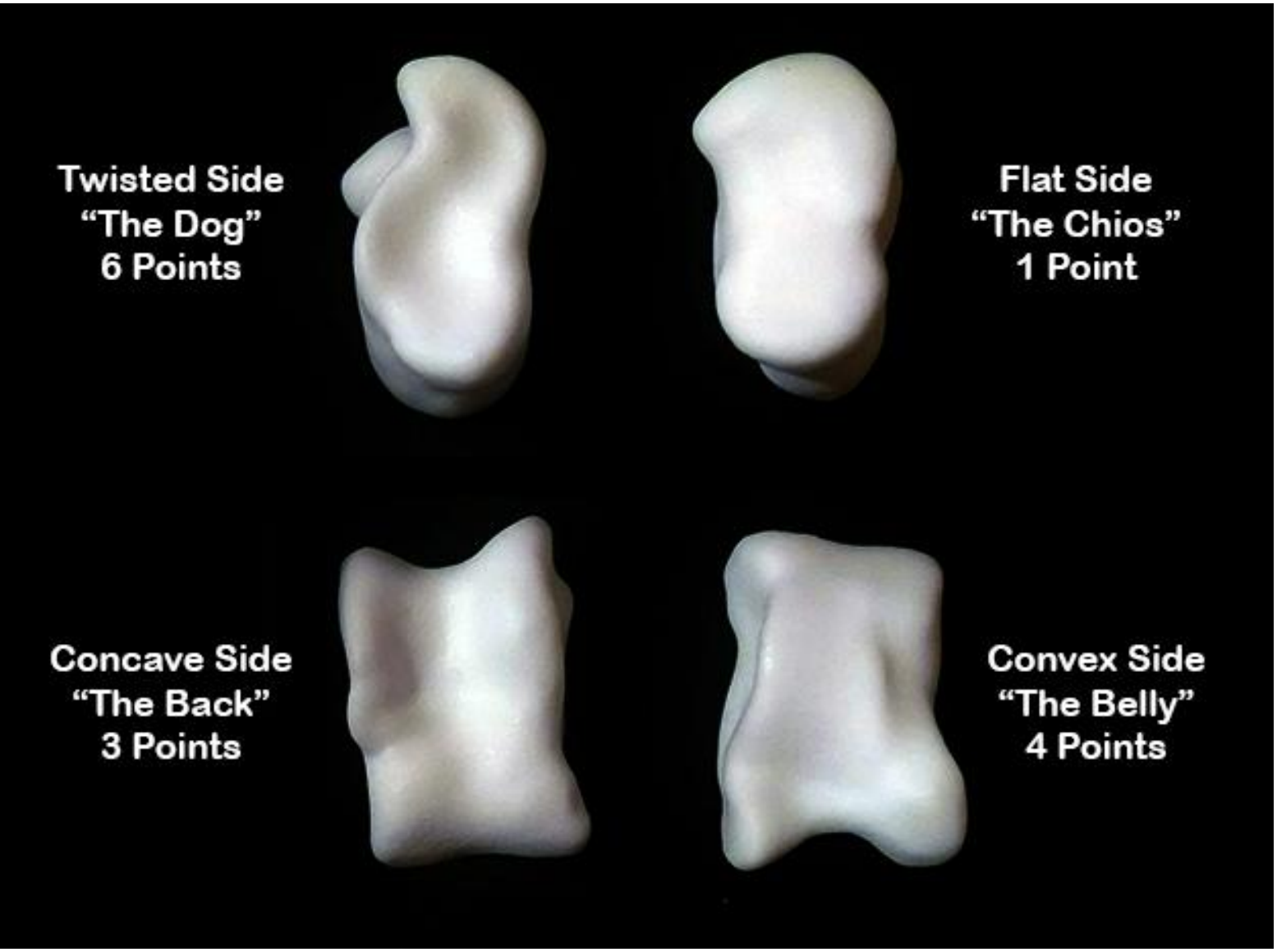
Games of chance

- Sumeria, Assyria, ancient Greece, ancient Rome
- Knuckle Bones (Talis)
- Repeat the basi



From knuckle bones to dice and cards

- Winning or losing is up to chance, luck, or god.
- **Equal probability Assumption**: all outcomes have the same probability.
- True for dice and roulette
- Not true for knuckle bones.


Twisted Side
"The Dog"
6 Points
Flat Side
"The Chios"
1 Point
Concave Side
"The Back"
3 Points
Convex Side
"The Belly"
4 Points

Long Term Frequencies

- The probability that a knucklebone lands on a narrow face is smaller than it lands on a wide face.
- Each knucklebone is different, the probabilities are different.
- Suppose we have P(**6**)=0.1, P(**1**)=0.2,p(**3**)=0.3, p(**4**)=0.4
- Flip 1000 times:

Long Term Frequencies

- The probability of landing on a narrow face is smaller than that of landing on a wide face.
- Each knucklebone is different, the probabilities are different.
- Suppose we have P(**6**)=0.1, P(**1**)=0.2,p(**3**)=0.3, p(**4**)=0.4
- Flip 100 times:

```
6 probability=0.10 frequency= 12/100 = 0.12
1 probability=0.20 frequency= 21/100 = 0.21
3 probability=0.30 frequency= 29/100 = 0.29
4 probability=0.40 frequency= 38/100 = 0.38
```

Long Term Frequencies

- The probability of landing on a narrow face is smaller than that of landing on a wide face.
- Each knucklebone is different, the probabilities are different.
- Suppose we have P(**6**)=0.1, P(**1**)=0.2,p(**3**)=0.3, p(**4**)=0.4
- Flip 10 times:

```
6414114444

6 probability=0.10 frequency= 1/10 = 0.10
1 probability=0.20 frequency= 3/10 = 0.30
3 probability=0.30 frequency= 0/10 = 0.00
4 probability=0.40 frequency= 6/10 = 0.60
```

Stopping a game in the middle

- Simplified version of problem in famous letter from Pascal to Fermat in 1654
- Suppose a card game of pure chance is played until one side wins.
- Both players put in 1$.
- The winner takes the 2$
- Suppose the game is **stopped** before either side wins.
- How should the 2$ be split?
- What is the probability that player 1 will win given the cards currently held?

The frequentist point of view

- To assign a probabilities to the outcomes of a game/experiment is the same as saying that if we repeat the game many times, the long term frequencies of the outcomes converge to the probabilities.
- Provides a solid foundation on which probability theory is built.
- Makes sense in games and other situations where one can repeat the same random choice many times.
- Not always possible ….

Situations where repetition is hard

1. A meteorologist says that the probability of rain tomorrow is 10%.
   - What does that mean?
   - It will either rain or not rain.
   - Tomorrow happens only once.
2. Suppose a surgeon says that there is a 2% chance of complications with a particular surgery.
   - It might mean that 2% of the patients that underwent the surgery had complications.
   - What does it mean for you ?
   - Maybe most of the complications where with patients older than 90 (and you are 35) …

The colloquial meaning of probability

- The word “probable” was in use before 1650. But it’s meaning was not quantitative
- Even today the words “probable” and “probably” have common use meanings that is qualitative, not quantitative.

Definition of PROBABLY Merriam Webster Dictionary

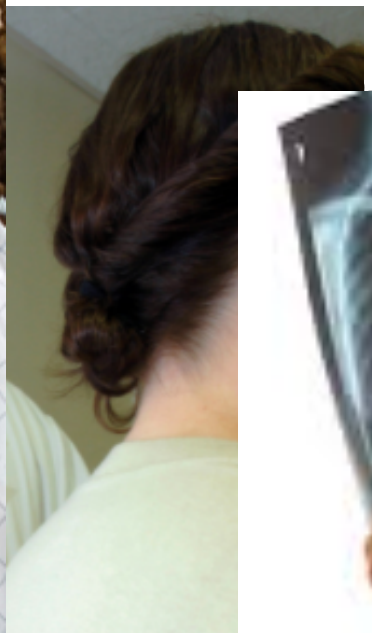
: insofar as seems reasonably true, factual, or to be expected : without much doubt • is *probably* happy • it will *probably* rain

A probable doctor

- Before 1660 it was common to say that someone is a “probable doctor”.
- It meant that the doctor was **approved** by some authority.
- At the time, in Europe, the authority was usually the church.
- Today MDs are approved by a board, after passing the board exams.

Combining evidence for Diagnosis

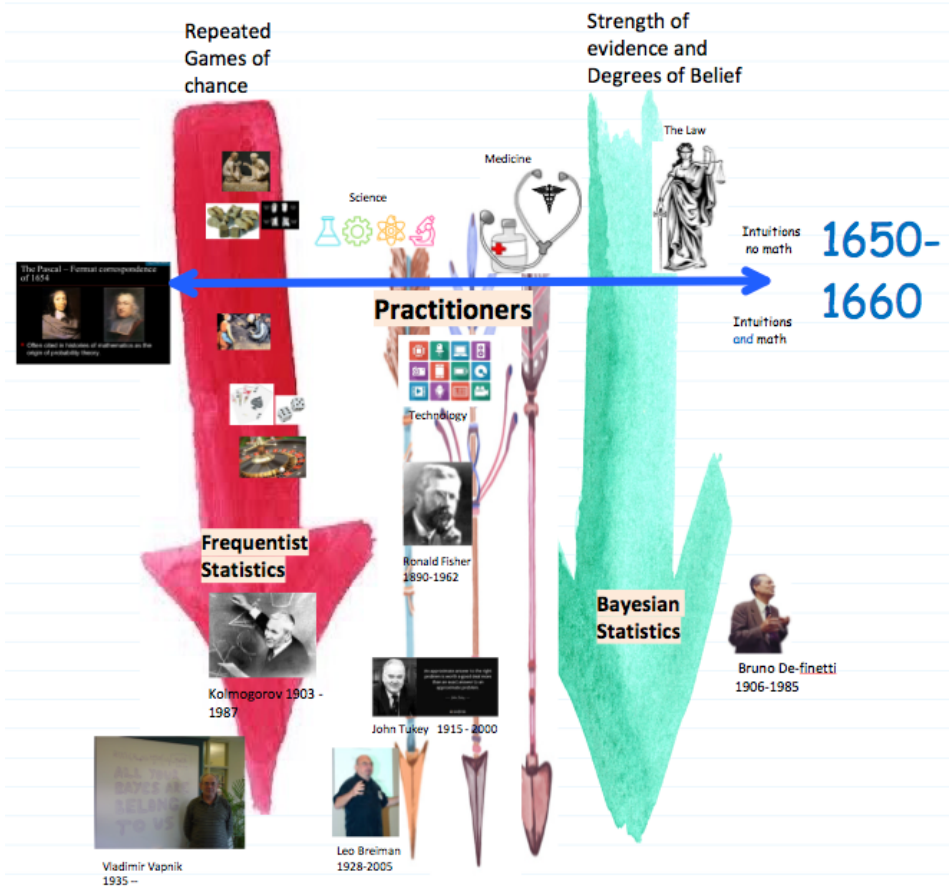
- Diagnosing a patient requires combining pieces of information.
- Most information is uncertain (measurement error)
- different relevance.



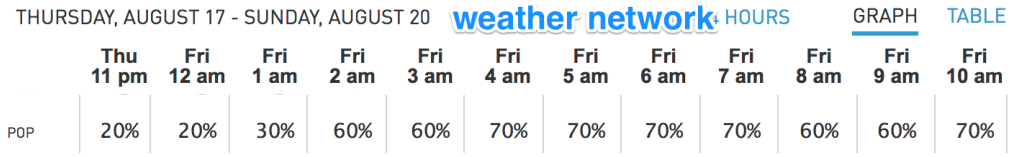
Combining evidence

- Central to many fields: Medicine, economics, investment, Law, Science, Technology …..
- Typically, you don’t repeat an experiment many times.
- The math used is probability theory, but much of the discussion is not mathematical.
- Closely related concepts: Fairness, pricing.
- A popular approach: Bayesian Statistics.

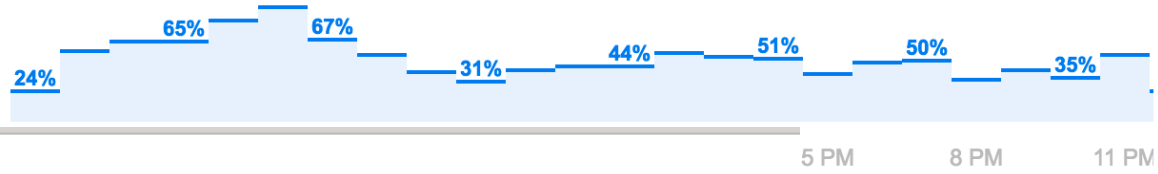
Next video: an exploration of duality



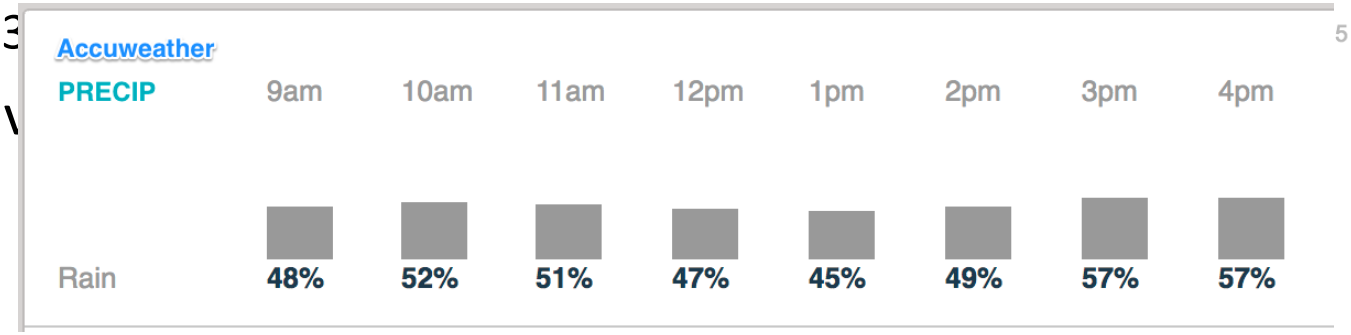
Combining e

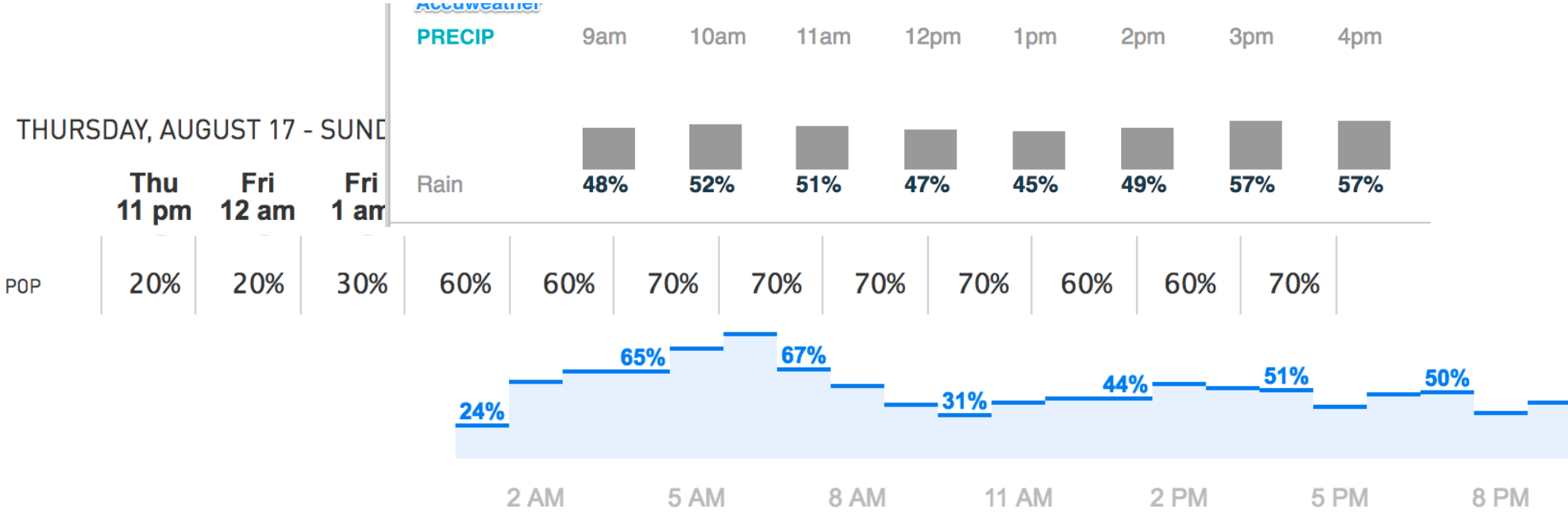


- Each morning you need to choose one of several routes to your work.
  1. You consult several w
  2. You choose a route.
  3



- Ov





	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed
Weather 1									
Weather 2									
Weather 3									
Rain?									

	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed
forecaster1									
forecaster2									
forecaster3									
Rain?									

Making rational decisions

- You want to know whether it will rain tomorrow.
- You consult different forecasters:
- Chance of rain 20%

Making probabilistic inferences

- You consult several weather prediction channels.
  - Each channel predict rain with a different probability.
  - What is your prediction?
- You consult several surgeons
  - You have access to their past performance
  - Who should you trust:
    - A young doctor who has done 10 surgeries and all were successful
    - Or
    - An oldr doctor who has done 100 surgeries and 95 were successful

The problem of stopping a game in the middle

Introduction to the

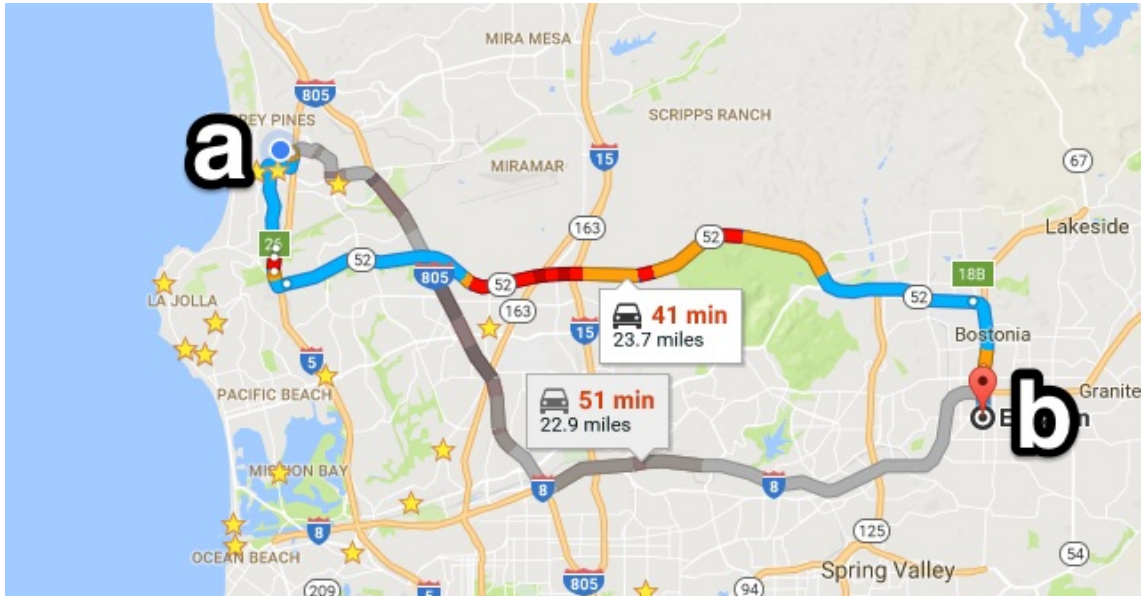
Introduction to Probability and Statistics

Probability

Statistics

Why should you care about prob&stat?
I

- Navigation software:
  - **Certainty**: Find the shortest route from a to b.
  - **Uncertainty**: Find the fastest route from from a to b.



Why should you care about prob&stat? II

- Search Engine:
  - **Certainty**: Find **all** web pages that contain the words "Trump","Hillary" and "debate"
  - **Uncertainty**: Find the 10 **most relevant** pages for the query "Trump, Hillary debate"

Why should you care about prob&stat? III

- Insurance Company:
  - **Certainty**: If a person with life insurance dies, the insurance company has to pay the family $X
  - **Uncertainty**: What is the minimal life insurance premium such that the probability that the life insurance company will be bankrupt in 10 years is smaller than 1% ?

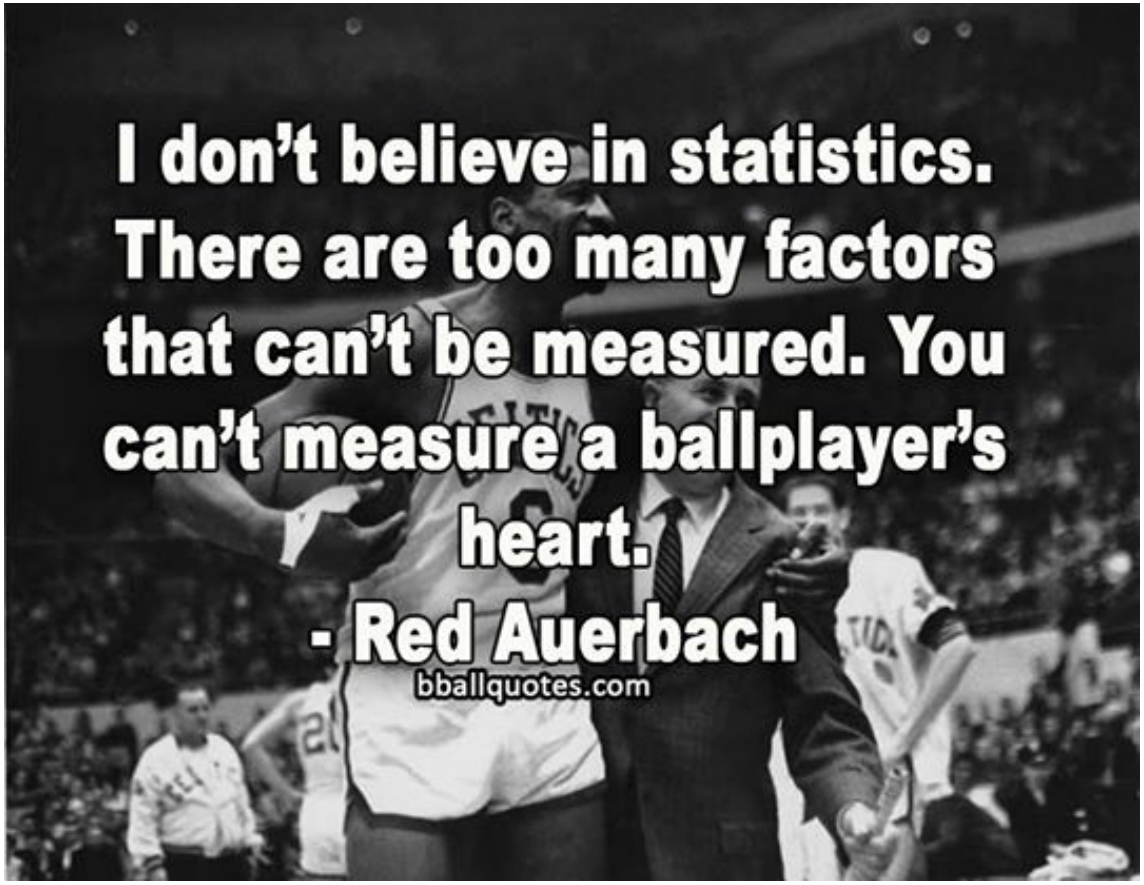
What will you learn in this course?

- Navigation and search engine problems are advanced, in this class you will learn the foundations.
- Solve basic problems of reasoning under uncertainty:
- Examples:
  - If you flip a coin 100 times, what is the probability of getting at most 10 "heads" ?
  - What is the probability of getting a "4 of a kind" hand in poker.

Computer science examples

- If you want to hash 1,000,000 elements and can allow more than 5 indirections for only 10 elements, how big does the table need to be?
- Suppose that the expected time between failures for a router is one year. What is the probability that the router will fail during the first month?

Some don't believe in statistics



Many do


Sort On
Clear
45 32 31 22 11 10
25
TurboStats
34
21 20 1
Clear
Hide
GAMES 2
15-29 52% FG
7-13 54% 3P
8-11 73% FT
#03 45 PTS
Stop
HOME
T30
T60
Eastside High
76
FOULS 7 B 1
TOL
1:02
4 Period
15-4 0 1 F
Poss
Westside High
84
TOL 1 D 11
FOULS
F 1 0
VISITOR
JMP
TKC
Clock
GAMES 2
5-7 71% FG
0-0 % 3P
2-2 100% FT
#05 12 PTS
Eastside High
Live
Stats
Shot Chart
Large Chart
Info
Summary
+/-
Events
Scout
BoxScore
Lineups
Game Factors
Westside High
> Mark Grant
25 Gary Adams
> Reggie Jones
> Jake Scott
> Bobby Brown
> Pete Thomas
31 Ken Jones
45 Jason Gamble
22 Donovan Korn
10 Roger Aken
Opponents
Eastside High
Def
Mark Grant
Def
Westside High
Jake Scott
PTS 10 PF 4 +/- -7 Run 5
Bobby Brown
PTS 9 PF 4 +/- 0 Run 7
Reggie Jones
PTS 12 PF 3 +/- -7 Run 7
Pete Thomas
PTS 9 PF 4 +/- 6 Run 2
Mark Grant
PTS 21 PF 3 +/- -4 Run 7
01
15
42
30
03
Ast TO Stl OF Foul Tec Blk Off Def
UA Made P FT P Miss LS
02
12
03
30
05
Nathan Smith
PF 3 +/- 11 Run 1 PTS 18
Ryan Sidney
PF 4 +/- 4 Run 7 PTS 14
Tim Thomas
PF 4 +/- 8 Run 7 PTS 9
Steve Smith
PF 4 +/- 0 Run 2 PTS 10
Martin Mason
PF 3 +/- -1 Run 0 PTS 12
Play..
15 Bobby B... oul
Clear
Other
DPTS
T
T
Plays
03
Post Online
02
Defense
Bench
Add
32 Erik Smith
11 Ryan Long
TEAM
Bench
Add
+ Add Player
01 Kevin Ian
Game Statistics for Eastside High
+ Video
Scout
Lrg Display
Factors
Events
Update
Name GS Min FG-A FG% 3P-A 3P% OR DR REB Ast Stl Blk TO PF Pts Eff NET
03 Mark Grant 1 28.2 8-14 57.1 4-8 50.0 4 0 0 0 2 3 21 17 95.5
25 Gary Adams 21.2 2-7 28.6 1-3 33.3 2 1 0 0 4 5 8 73.6
End Live
< Score >
Games
Save Game

Summary

- Uncertainty is all around us.
- Probability and Statistics provide a rational way to deal with uncertainty.

- Next:
  - What is probability?
  - What is statistics?