Under Used Statistical Pedagogical Ideas...

Some junk I’ve stumbled on that may help teaching stats and data analysis…

James (JD) Long

Here’s my pitch...

Goal is to Kick Ass

Student retention is highest when a concept helps them do something **they** feel is useful. They don’t want a linear progression of ideas that build incrementally. They want super powers to do something useful.

Best Data

Fully controlled simulation

OR

Data students care about

Nothing in between

Teach Toy Models

Students who learn a theorem know one theorem. Students who can simulate a problem can back into **MANY** theorem’s and gain better understanding and intuition.

Teach Meta Skills

“Learn to learn” … Learn to embrace not knowing because you know the solution to ignorance.

Ignorance is a solved problem.

So Who Am I?

GlobalAgRisk INC.

KPMG

Genworth

COLLINS

GUY CARPENTER

RenaissanceRe

→

Morgan Stanley

KOREAN AIR

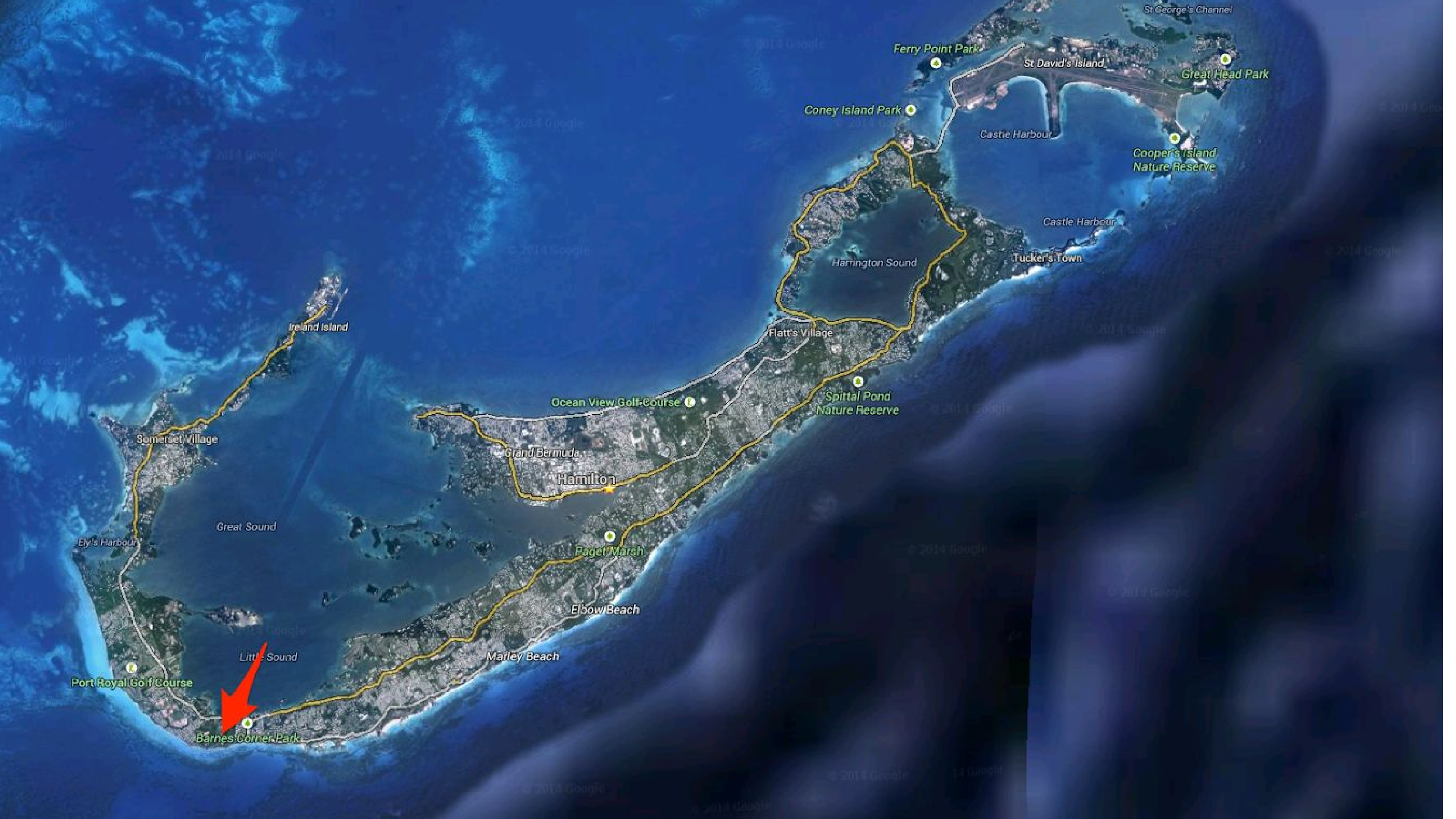
Tyson

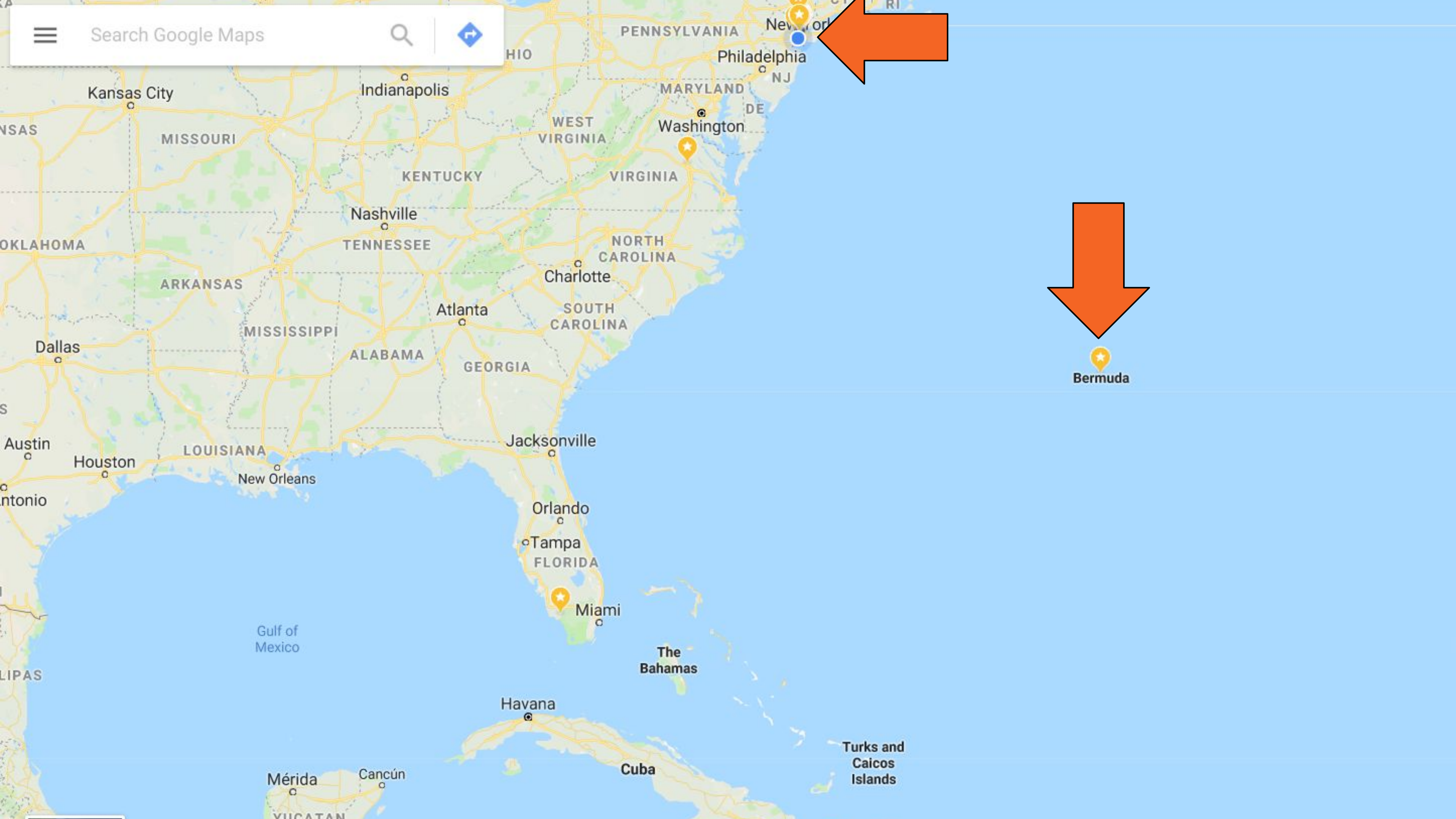
Tyson Foods, Inc.

MIRANT

PSEG

Swiss Re


St George's Channel
Ferry Point Park
St David's Island
Great Head Park
Coney Island Park
Castle Harbour
Cooper's Island Nature Reserve
Castle Harbour
Harrington Sound
Tucker's Town
Ireland Island
Flatt's Village
Ocean View Golf Course
Spittal Pond Nature Reserve
Somerset Village
Grand Bermuda
Hamilton
Great Sound
Ely's Harbour
Paget Marsh
Elbow Beach
Marley Beach
Little Sound
Port Royal Golf Course
Barnes Corner Park
© 2014 Google


Search Google Maps
PENNSYLVANIA
New York
RI
HIO
Philadelphia
NJ
Kansas City
Indianapolis
MARYLAND
DE
NSAS
MISSOURI
WEST VIRGINIA
Washington
KENTUCKY
VIRGINIA
Nashville
OKLAHOMA
TENNESSEE
NORTH CAROLINA
ARKANSAS
Charlotte
Atlanta
SOUTH CAROLINA
MISSISSIPPI
Dallas
ALABAMA
GEORGIA
Bermuda
Austin
Houston
LOUISIANA
New Orleans
Jacksonville
ntonio
Orlando
Tampa
FLORIDA
Miami
Gulf of Mexico
The Bahamas
LIPAS
Havana
Turks and Caicos Islands
Cuba
Mérida
Cancún
YUCATAN

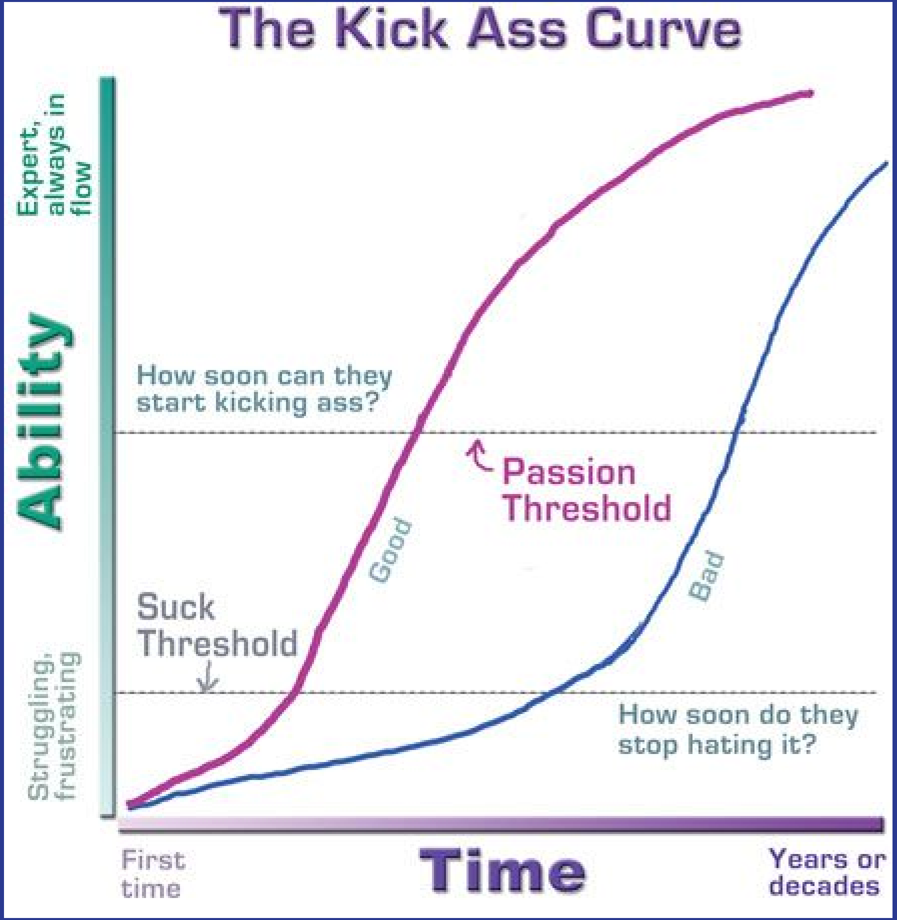
Nobody wants to learn analytics. They want to learn to kick ass.

“The more knowledge and skill someone has, the more passionate they become, and the more passionate they become, the more they try to improve their knowledge and skills.”
- Kathy Sierra

http://headrush.typepad.com/creating_passionate_users/2005/06/kicking_ass_is_.html

The Kick Ass Curve

Also borrowed from Kathy Sierra



http://headrush.typepad.com/creating_passionate_users/2005/10/getting_users_p.html

How to Raid Fort Kickass

Build Motivation

Toddlers...

Toddlers learn to walk not because it’s the next syllabus item. They learn because they want to carry two toys at once.

Get to the Good Stuff

Where are we going?

Illustrate what a concept will allow the student to **DO** in the future… DOing is more motivating than knowing

Make the Hard Bits Easier

Put the bowling bumpers up sometimes

It’s OK to skip some ugly bits… then come back to them later if it builds motivation.

e.g. use ‘Tidyverse’ code in R to make R easier to use

Let’s talk about example data...

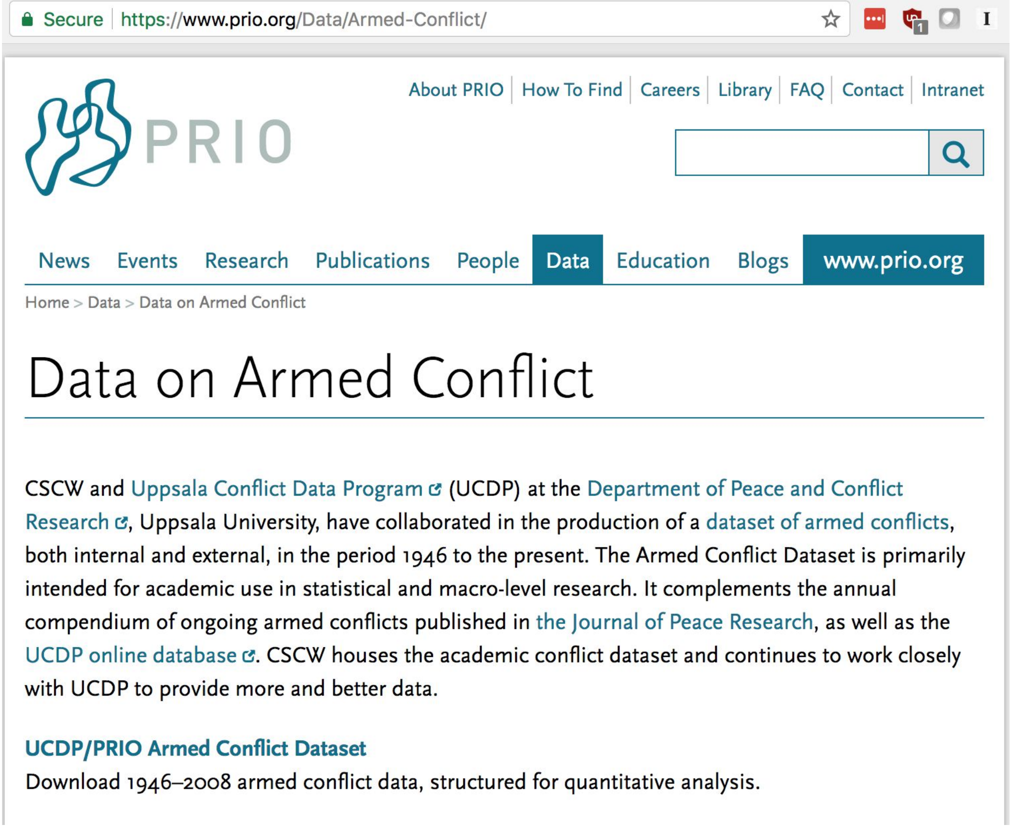
Fully Simulated Data

Data Students Care About

What might learners care about?

Actual business data that lead to a story...

Conflict data?
The Peace Research Institute Oslo (PRIO)



Data on Armed Conflict

CSCW and Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University, have collaborated in the production of a dataset of armed conflicts, both internal and external, in the period 1946 to the present. The Armed Conflict Dataset is primarily intended for academic use in statistical and macro-level research. It complements the annual compendium of ongoing armed conflicts published in the Journal of Peace Research, as well as the UCDP online database. CSCW houses the academic conflict dataset and continues to work closely with UCDP to provide more and better data.

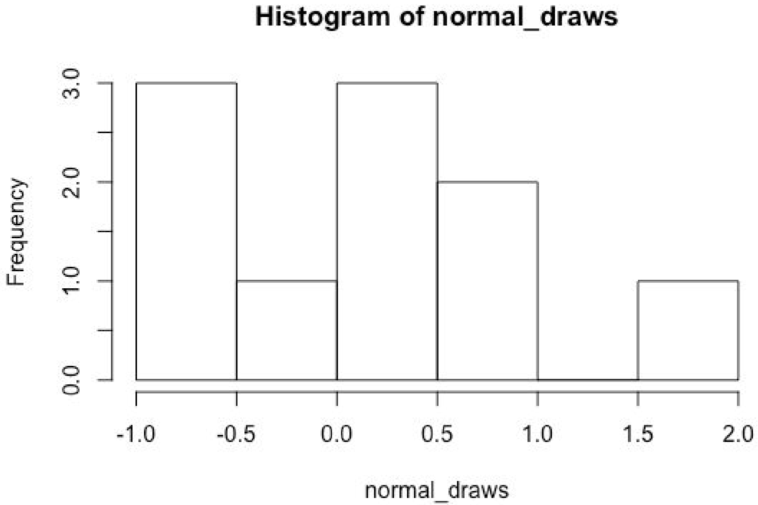
UCDP/PRIO Armed Conflict Dataset

Download 1946–2008 armed conflict data, structured for quantitative analysis.

Fully Simulated Data...

```
sample_size <- 10
# simple illustration of random draws
normal_draws <-  rnorm(sample_size, mean=0,
sd=1)
mean(normal_draws)
hist(normal_draws)
```

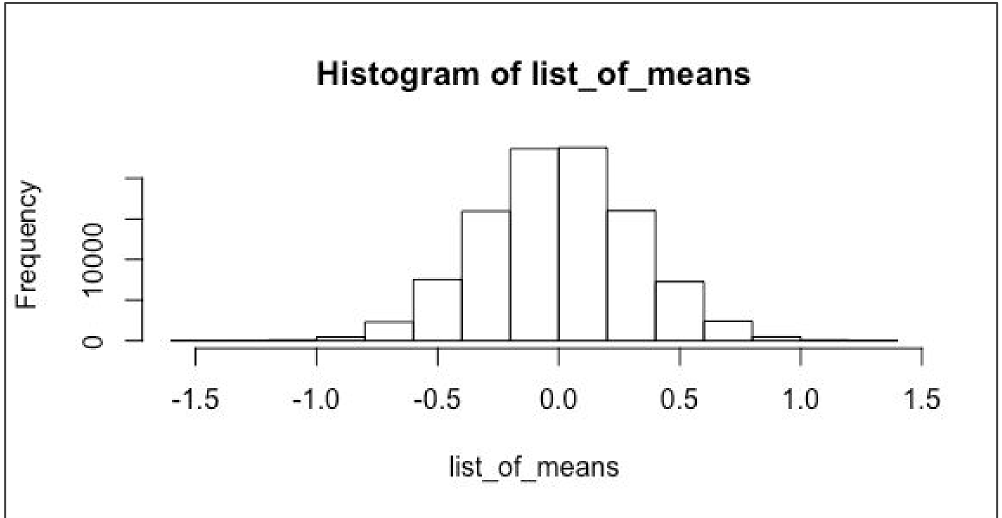
```
[1] 0.1322028
```



```
# what's the distribution of the mean measurements as we
do this over and over?
list_of_means <- array()
times_to_loop <- 100000

for (i in 1:times_to_loop){
  normal_draws <-  rnorm(sample_size, mean=0, sd=1)
  list_of_means[i] <- mean(normal_draws)
}

hist(list_of_means)
```



7 Lines of code… And well on our way to backing into Student T test, sqrt(n) intuition

Building Toy Models:
Expansion of Simulated Data

```
## let's play with a regression now
draws <- 1000
set.seed(2)

# create a DF with 3 columns, 1000 rows of random standard normal draws
random_regression <- data.frame(replicate(3,rnorm(draws)))

#calculate the dependent variable Y
random_regression %>% mutate(
    e = rnorm(draws, 0, 1), ## better add some error noise
    Y = 2 * X1 + 3 * X2 + 4 * X3 + e
) -> random_regression

# build a linear regression
model <- lm( Y ~ X1 + X2 + X3, data=random_regression )
summary(model)
```

```
Call:
lm(formula = Y ~ X1 + X2 + X3, data = random_regression)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2144 -0.6782  0.0100  0.6499  3.1942 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.02382    0.03135    0.76    0.448    
X1           1.93678    0.03079   62.90   <2e-16 ***
X2           3.04803    0.03144   96.95   <2e-16 ***
X3           4.04283    0.03066  131.88   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9872 on 996 degrees of freedom
Multiple R-squared:  0.9689,	Adjusted R-squared:  0.9688 
F-statistic: 1.036e+04 on 3 and 996 DF,  p-value: < 2.2e-16
```

Other Toy Models?

Actual Experience:

If we have 30 observations from a lognormal distribution, what’s our confidence around the 90% percentile tail measurement? What about the 50% percentile?

Meta Skills:

Ultimately we need only teach one skill:

How to learn something we don’t already know.

Top Technical Meta Skills

- How to create a reproducible example
- How to ask a question
- How to explain a problem
- How to query Google
- How to RTFM
- How to document a process
- How to pick the right tool
- How good is good enough
- Learn that “design patterns” exist

James (JD) Long

jdlong@gmail.com

Twitter: @cmastication
GitHub:
https://github.com/CerebralMastication/WestPointPresentation