一起学

What do you expect/want to learn?

A few misunderstandings about Data Science

- Data science = Coding
- Data science = Math
- Data science = Fitting a machine learning model
- Data science = Deep learning

My understanding of key areas to learn in Data Science

- Learn programming skills
- Learn modeling skills
- Learn problem solving skills
- Learn skills working in a data science team
- Learn presentation skills
- Learn data analysis planning skills

Programming: learn SQL + R/Python

/

- Language to query from databases
- SQL like thinking is critical in programming

- Very good statistics packages
- Very good visualization packages
- Dplyr!!

- Accepted by CS
- Closest language to C++/Java
- Decent amount of packages

Key areas planed to cover in R

- Basic syntax
- How to write all codes with data frame (dplyr)
- Best tool for visualization (ggplot2)
- How to avoid for loop
- How to validate

```
msleep %>%
  group_by(vore) %>%
  summarise_at(vars(contains("sleep")), mean, na.rm=TRUE) %>%
  rename_at(vars(contains("sleep")), ~paste0("avg_", .))

## # A tibble: 5 x 4
##   vore    avg_sleep_total avg_sleep_rem avg_sleep_cycle
##   <chr>             <dbl>         <dbl>           <dbl>
## 1 carni              10.4          2.29           0.373
## 2 herbi               9.51         1.37           0.418
## 3 insecti            14.9          3.52           0.161
## 4 omni               10.9          1.96           0.592
## 5 <NA>               10.2          1.88           0.183
```

Goal of a good programmer: 指哪打哪

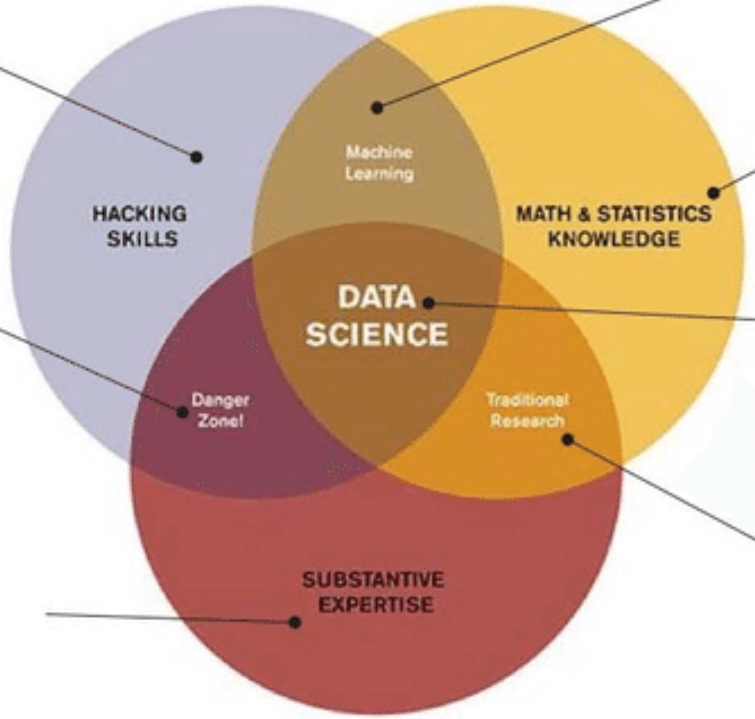
- It is fine to Google
- You need to write short code
- Code as you think

Learn enough analytical/modeling skills

Hacking Skills are necessary for working with massive amount of electronic data that must be acquired, cleared and manipulated

Danger zone: Hacking skills combined with substantive expertise without rigorous method can beget incorrect analysis

Substantive Expertise in a scientific field is crucial for generating motivating question and hypotheses and interpreting result



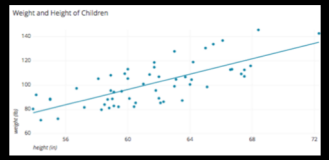
Machine Learning stems from combining skills with math and statistics knowledge, but does not require scientfic motivation

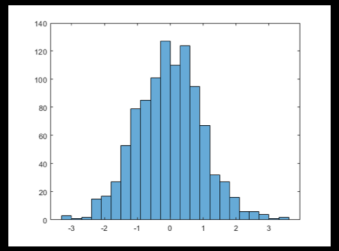
Math and Statistics Knowledge allow a data scientist. to chose and apply appropriate methods and tools in order to extract insight from data.

Data science, due to its interdisciplinary nature, requires an intersection of abilities: hacking skills, math and statistics knowledge, and a susbstantive expertise in a field of science

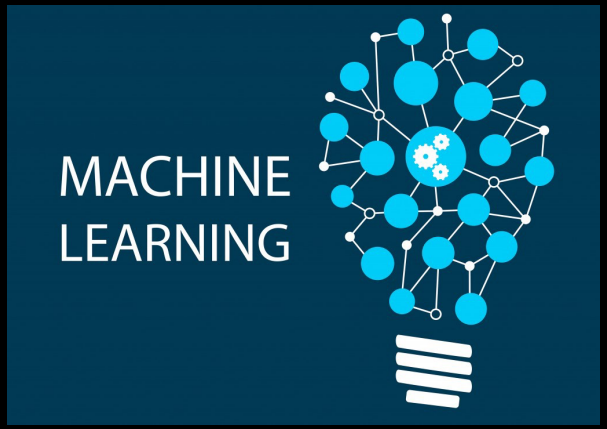
Traditional Research lies at the intersection of knowledge ok math and statistics with substantive expertise in a scientific field.

Will cover several key modeling techniques



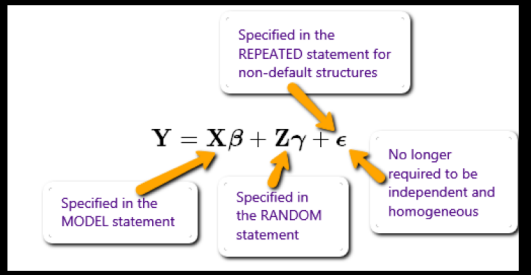


Basics



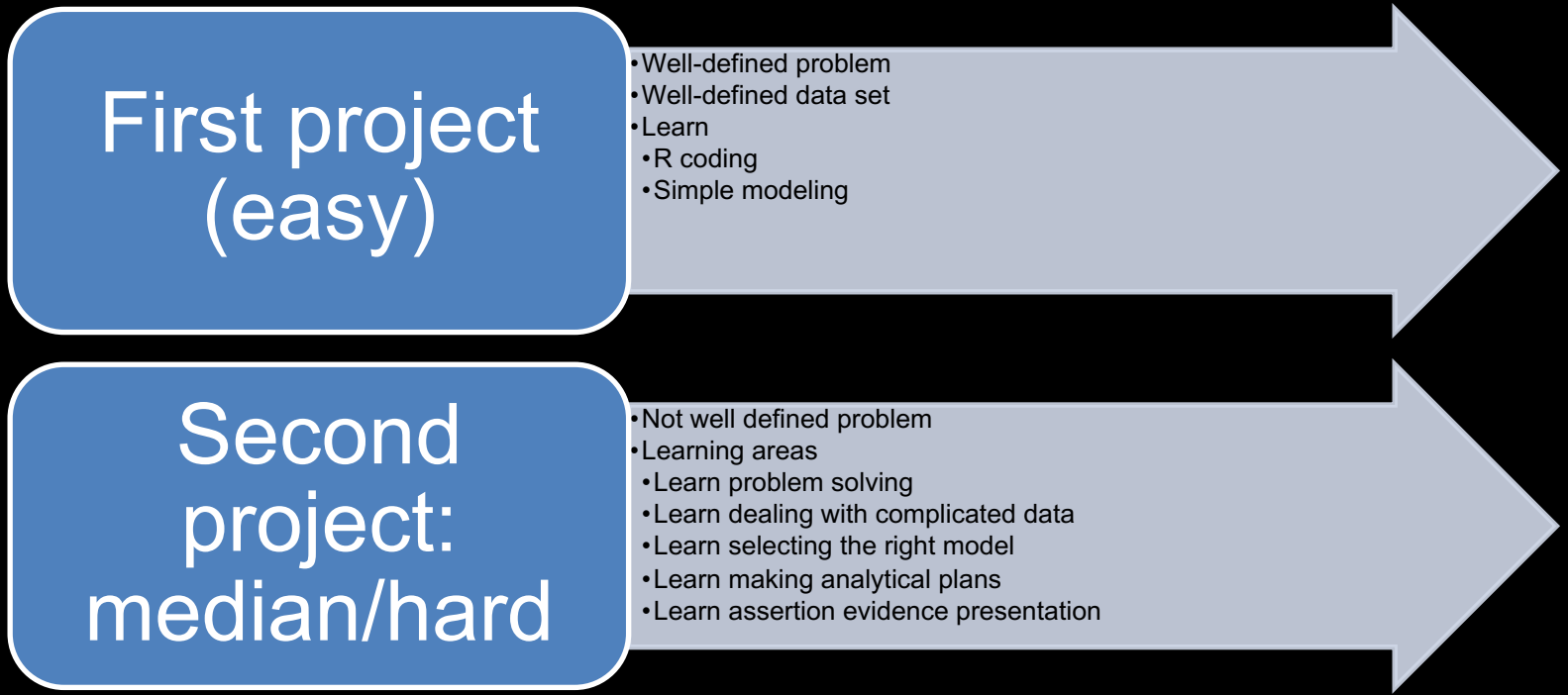
Machine learning (regression, tree based, neural network)

Operations Research



Variance decomposition

Designing two projects to help you learn key data science skills

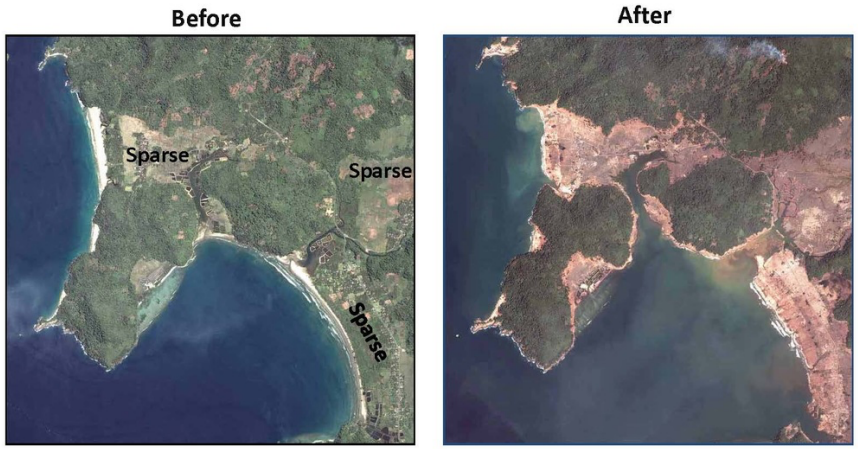


Science is not finished until it is communicated.

- Mark Walport

Learning Assertion Evidence is as important as learning all others

Tsunamis cause devastating destruction, especially to sparsely vegetated areas

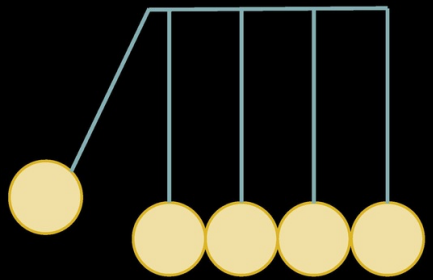


2004 Indian Ocean Tsunami: Gleebruk Village, Sri Lanka

[Alley, 2013] [homepage.mac.com/demark/]

Momentum is equal to the mass of an object multiplied by its velocity

$p = mv$



p = momentum

https://www.assertion-evidence.com/tutorial.html

Improve yourself in a data science team

- How to work with your manager?
- What are the data science tasks you will be assigned and how to work on them?
- How to prioritize your daily work?
- When is the time you need to reach out and ask for questions?
- How to improve yourself at work technically and personally?

敬畏耶和华是智慧的开端，
认识至圣者便是聪明。

Planning for this course