

# BUF DATA SCIENCE

OR

YET ANOTHER BUFFALO THING

# NOT THAT KIND OF BUF

- Summarize Survey
- Logistics
- Data Storytelling



# Data Scientist's Job

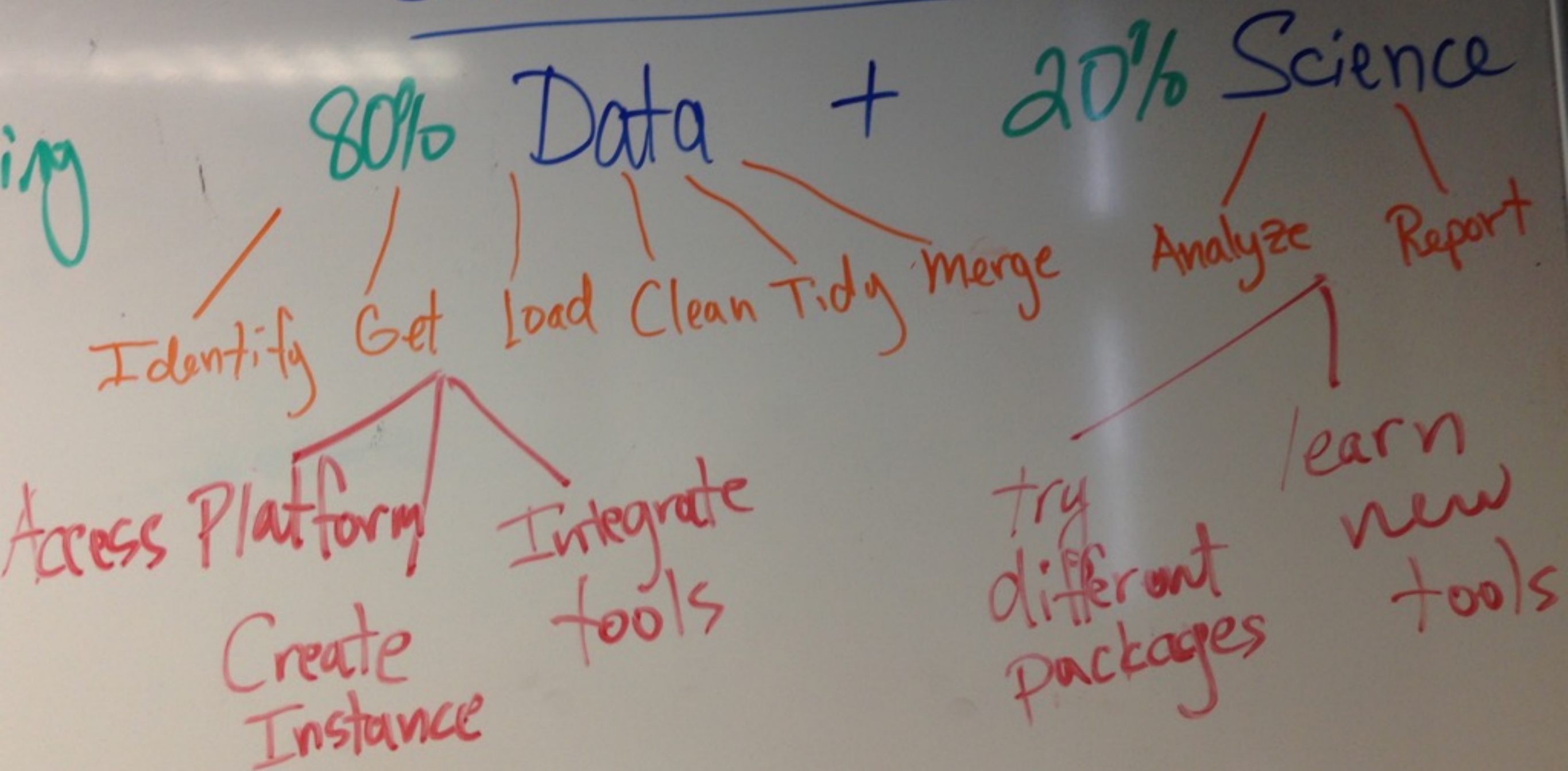
80% Data + 20% Science

```
graph TD; Title["Data Scientist's Job"]; Title --- SubTitle["80% Data + 20% Science"]; SubTitle --- Data["80% Data"]; SubTitle --- Science["20% Science"]; Data --- Identify; Data --- Get; Data --- Load; Data --- Clean; Data --- Tidy; Data --- Merge; Science --- Analyze; Science --- Report;
```

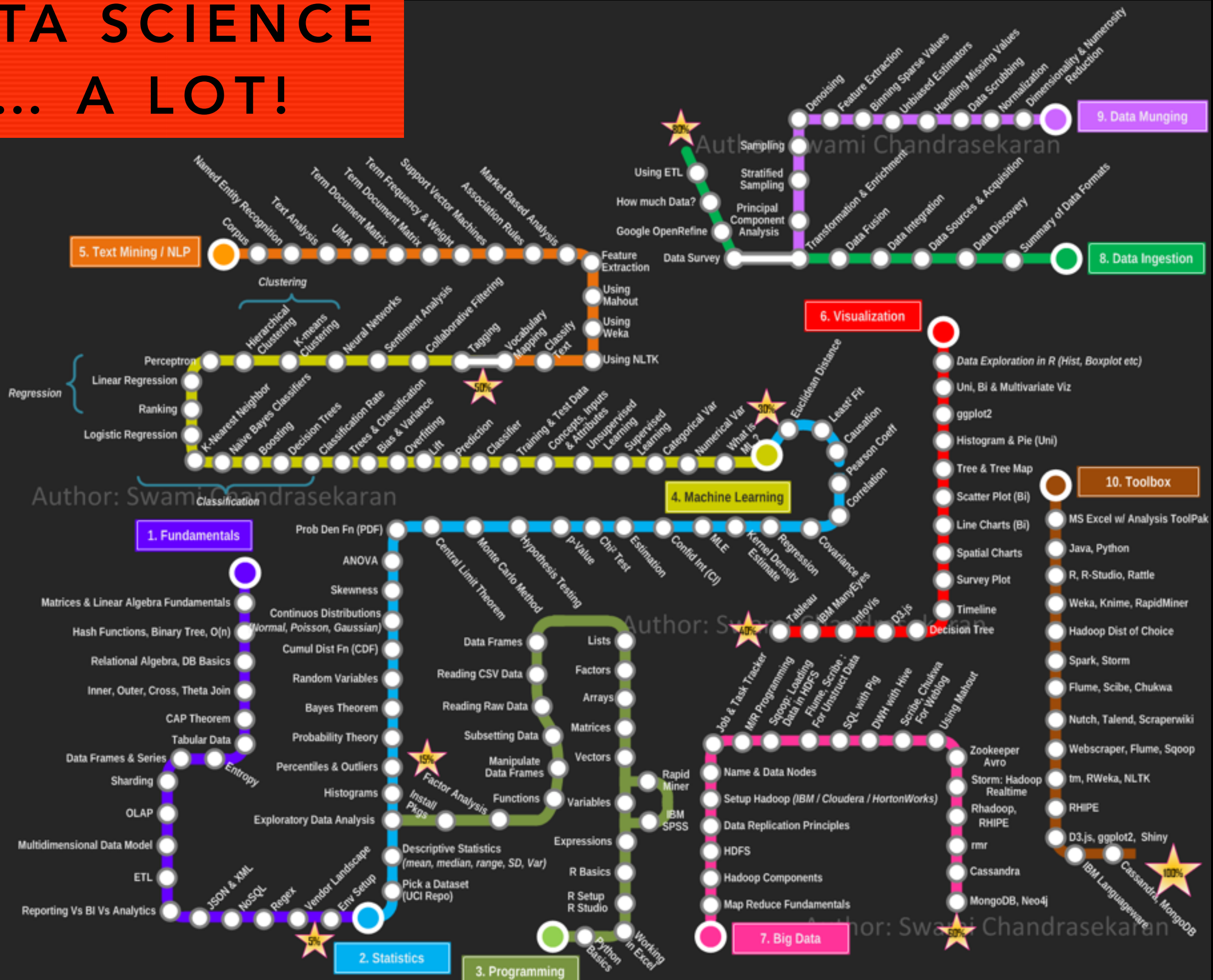
Identify Get Load Clean Tidy Merge Analyze Report



# Data Scientist's Job



# DATA SCIENCE IS ... A LOT!





# ?-SHAPED PEOPLE

FROM

T



TO

$\pi$

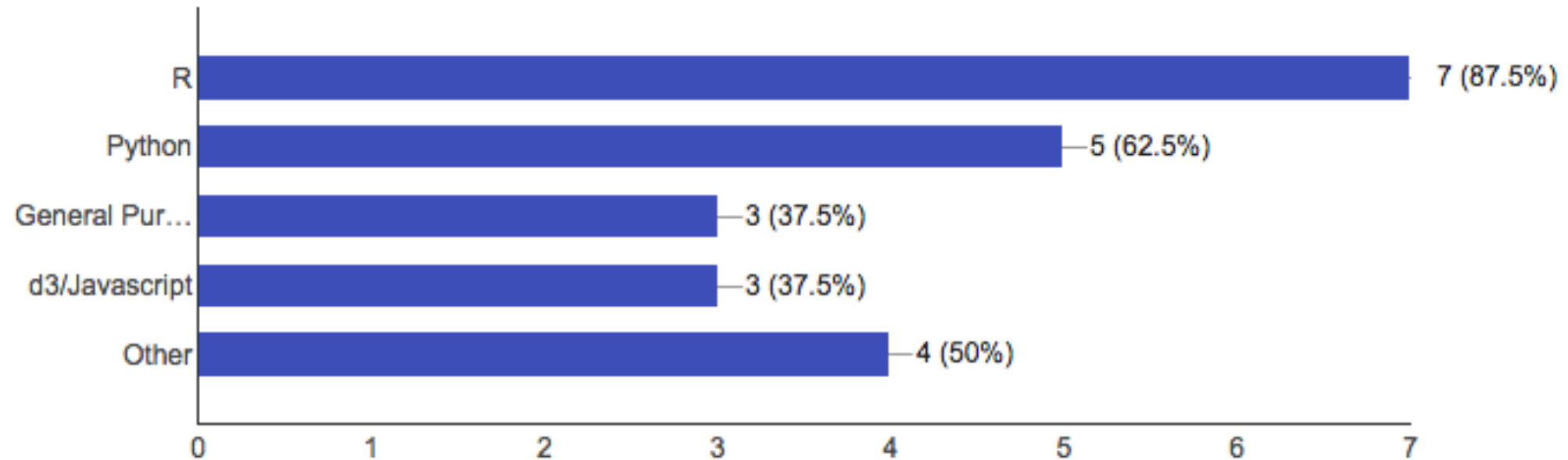


# SUPER POWERS

- Data
- Science



Which open-source data science languages do you use? (8 responses)

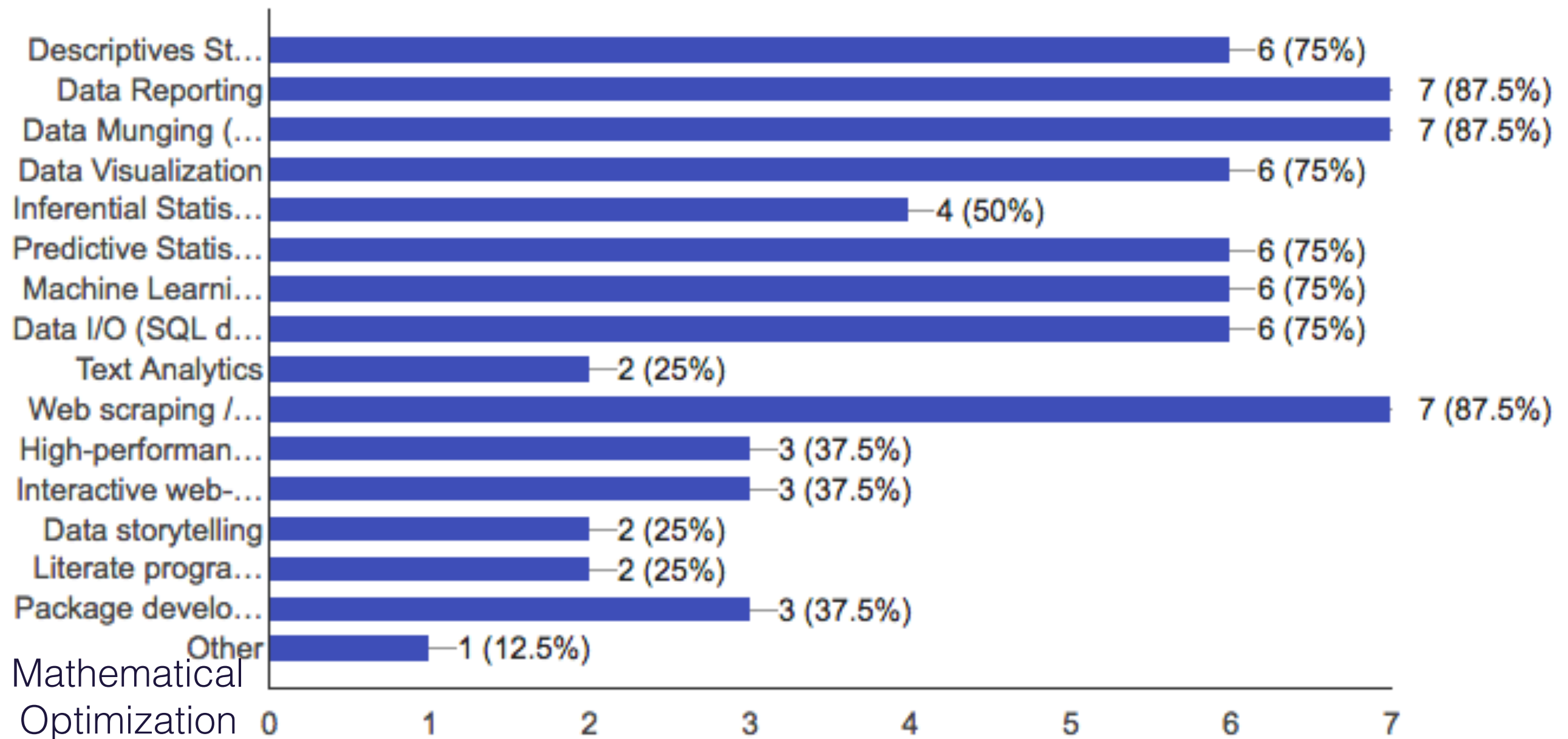


# No language war here

others = Perl, SAS, Stata, bash/shell



## Which data science skills have experience in? (8 responses)



interesting assortment of skills ; what did we leave out?

# Topics

- Take a survey?

# Logistics

- how frequently we meet (when next?)
- is email really our best option (Slack , Telegram, WhatsApp?)
- different formats
  - coffee meet up every other week during day
  - big presentations
  - tutorials, advanced topic, chatting
- do we want to WoM or meetup.com?
- food / bevs



# With these coding challenges, what next?

- FizzBuzz      vs.      Palindrome tester

# Getting to know you

- Name
- Role
- Data Science thing that
  - is victory or excites you
  - drives you mad or feels overwhelming

# Data Storytelling

- Touted requirements: creativity / communication
- Storytelling ingredients
- It's not the size of the data in the fight but size of fight in the data



# Taunted Requirements

- creativity / communication
  - - actually one of the biggest hurdles
  - - to be internally disruptive is hard
  - - most people want change without change

# Storytelling ingredients

- actual process of interest
- data generation process (how sensors are attached / what data comes out)
- what problem (within the process) is being solved

# Not size of data in fight, but size of fight in data

- Big data is not a requirement, and usually is actual a technical hindrance to many problems
- value of small data
- most valuable is asking good questions, and understanding data availability, provenance, quality
- - 5 why's
- - being allowed to disagree
    - play dumb