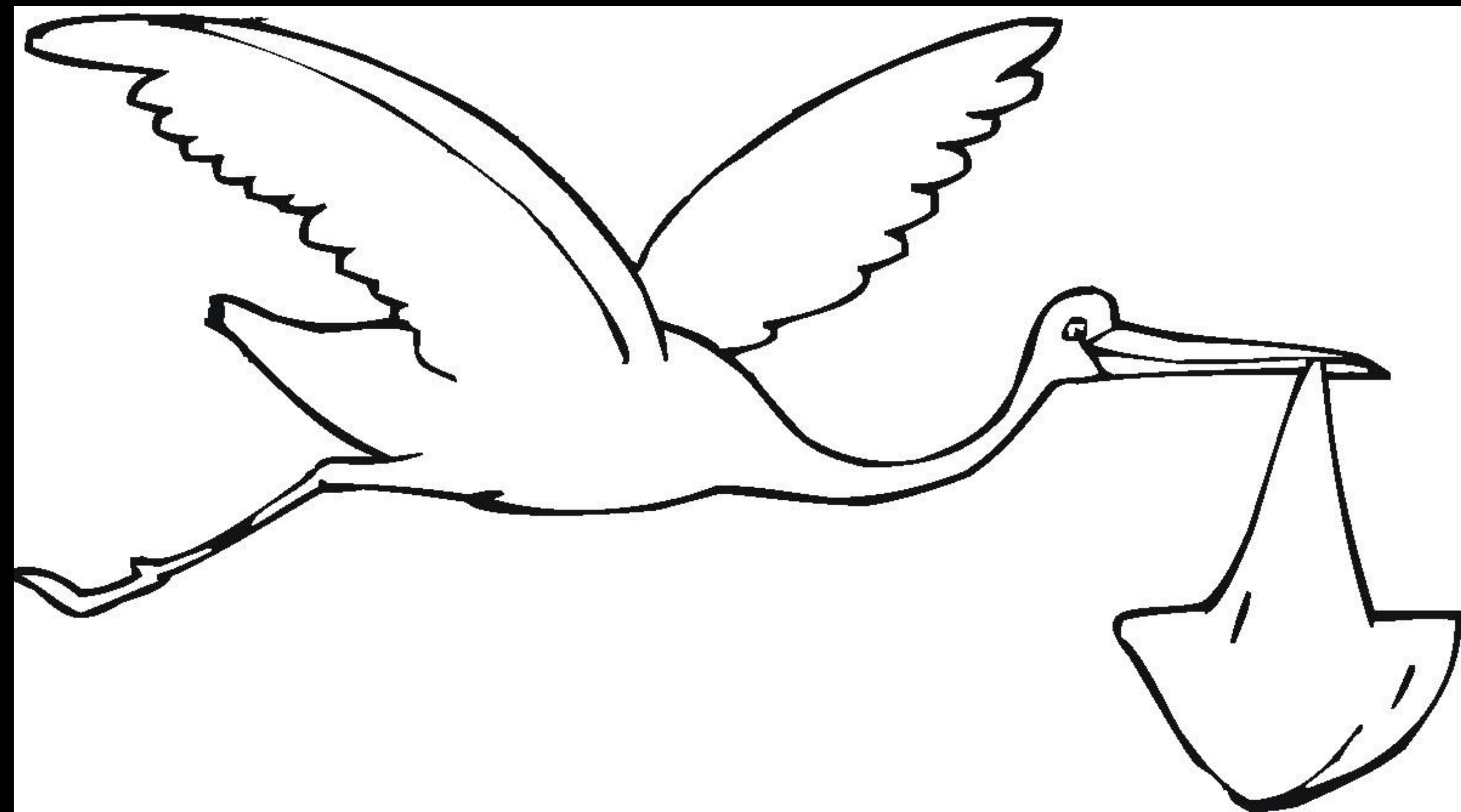


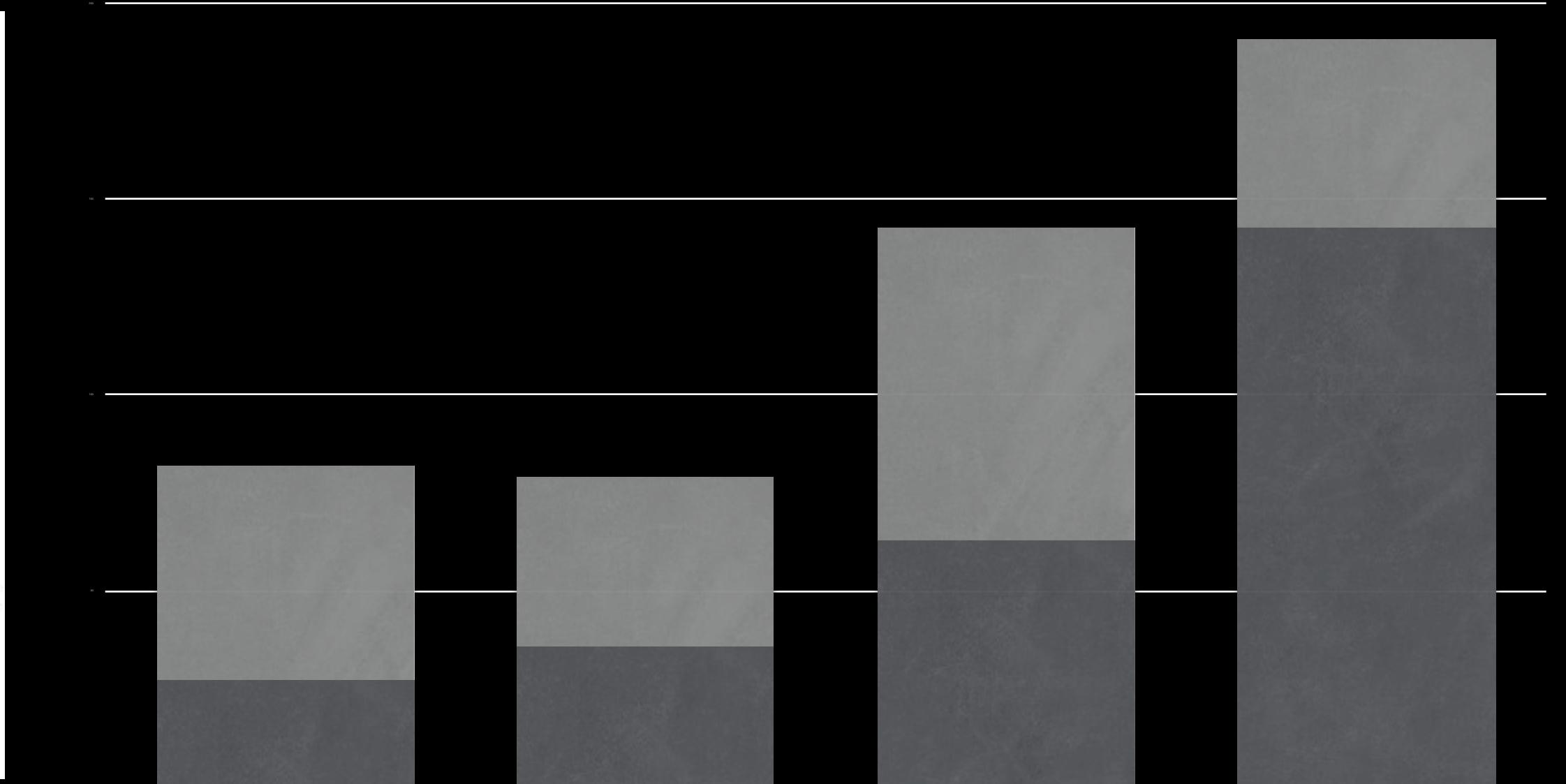
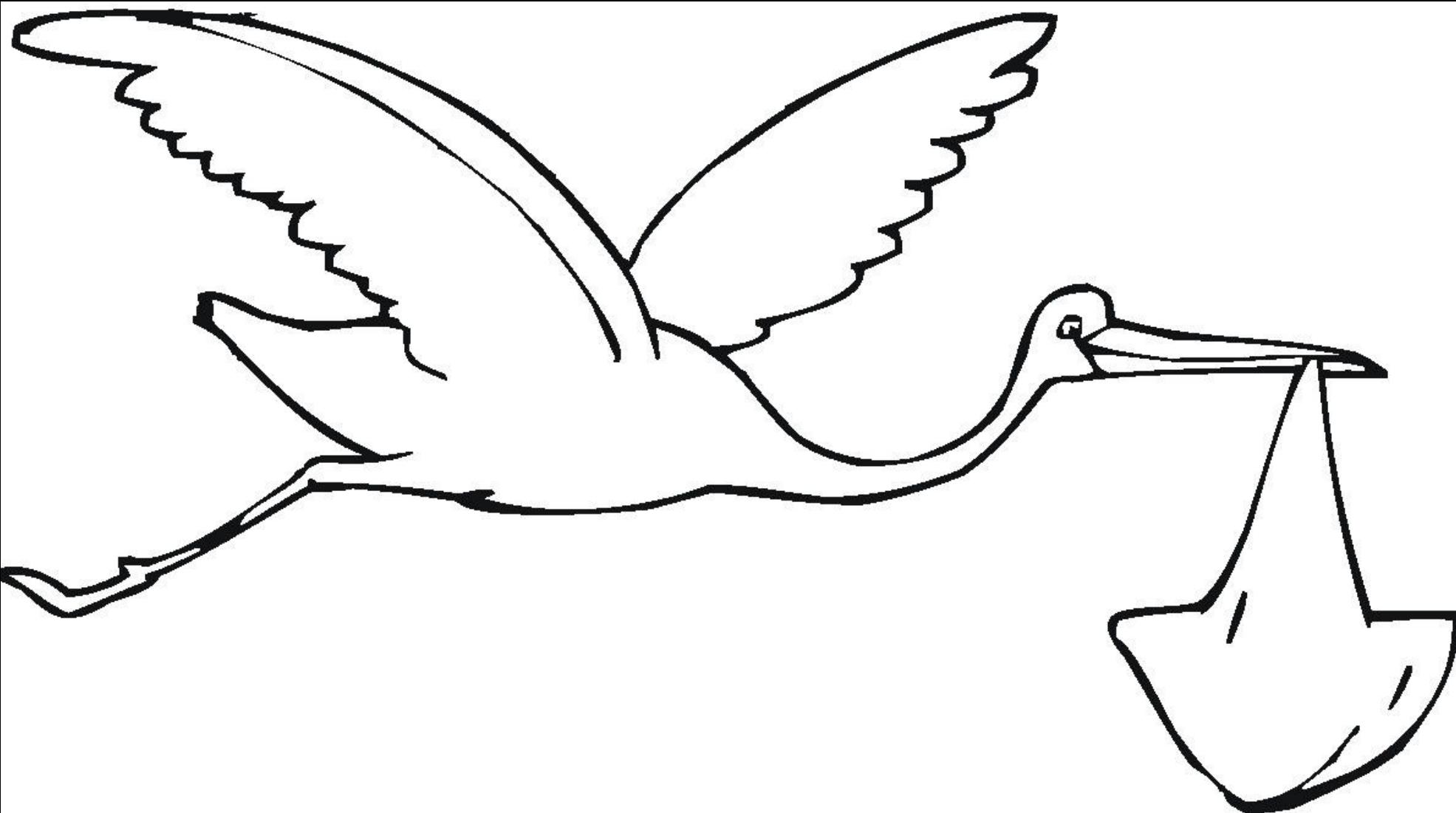
Visualizing Models and Methods for Decision-makers

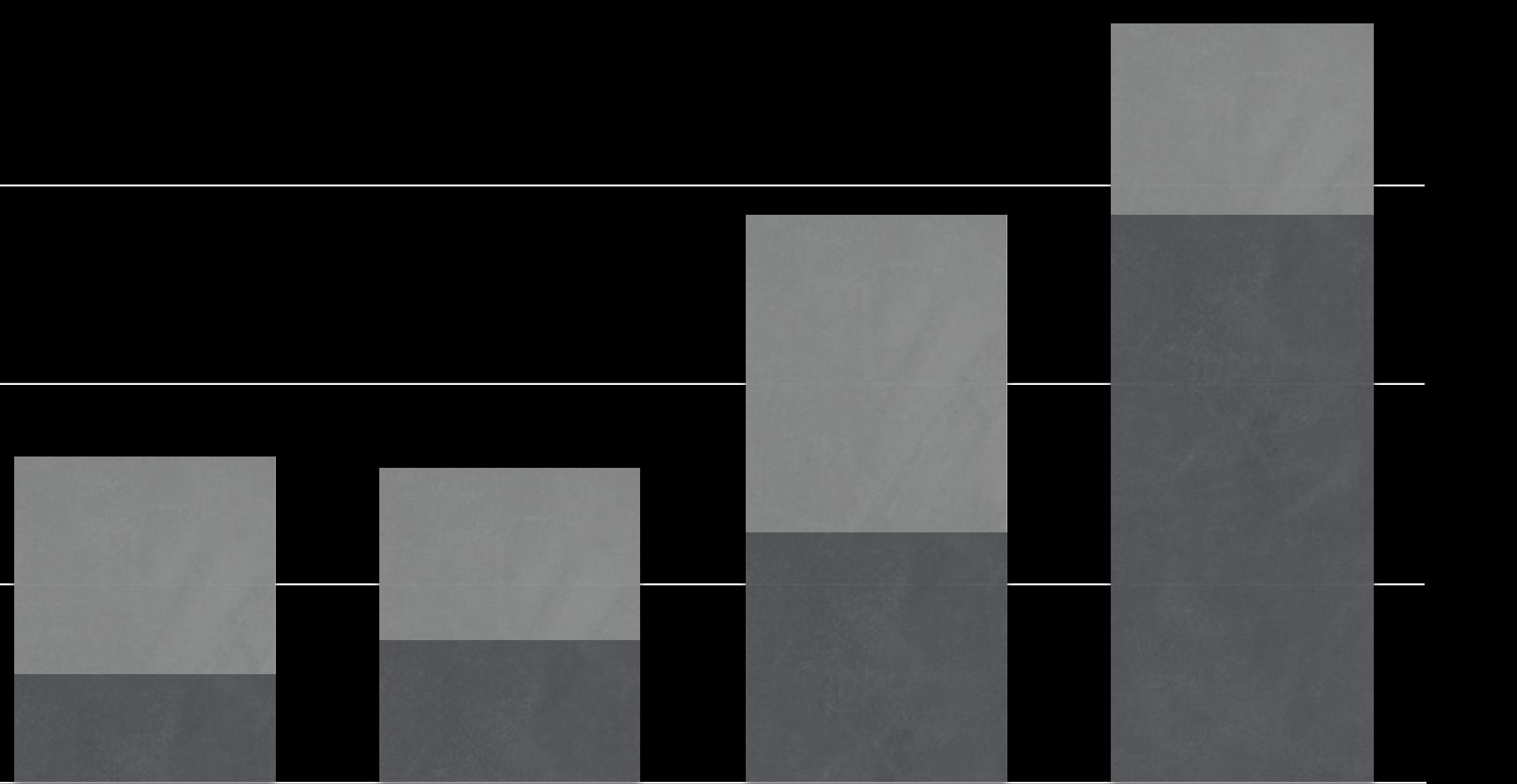
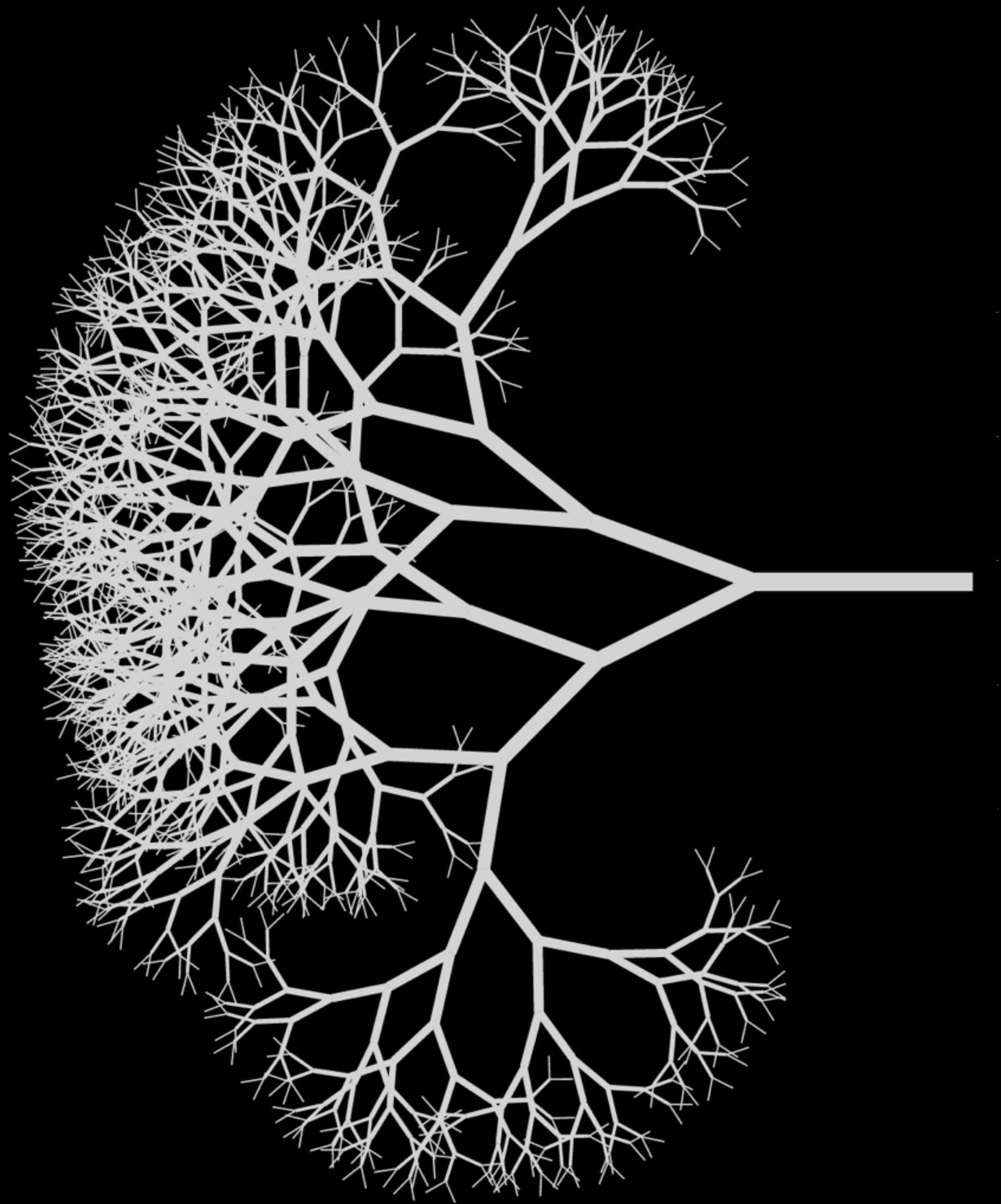
Michael Freeman
University of Washington
[@mf_viz](https://twitter.com/mf_viz)

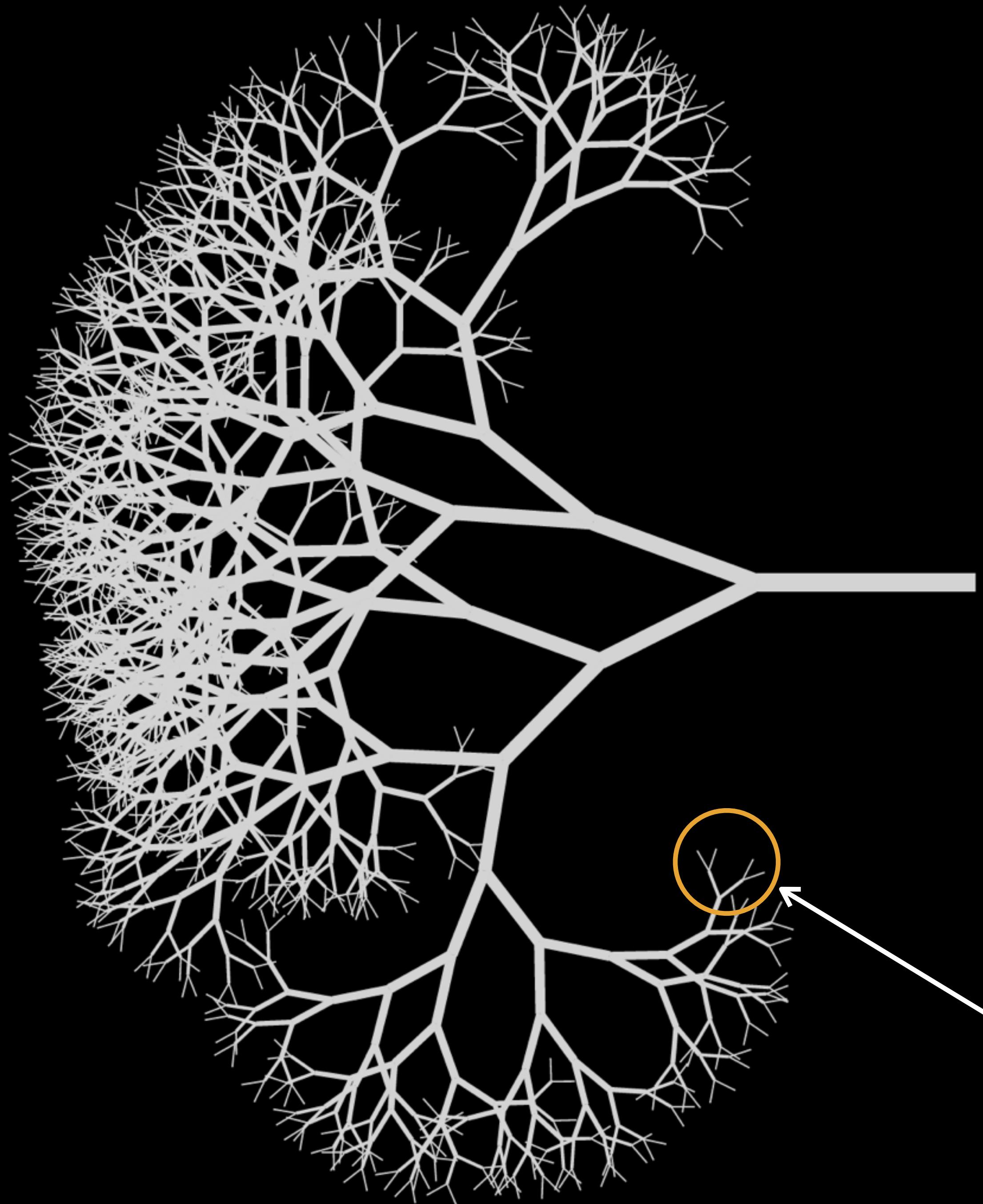
{a plea...}

{Where does **data** come from?}





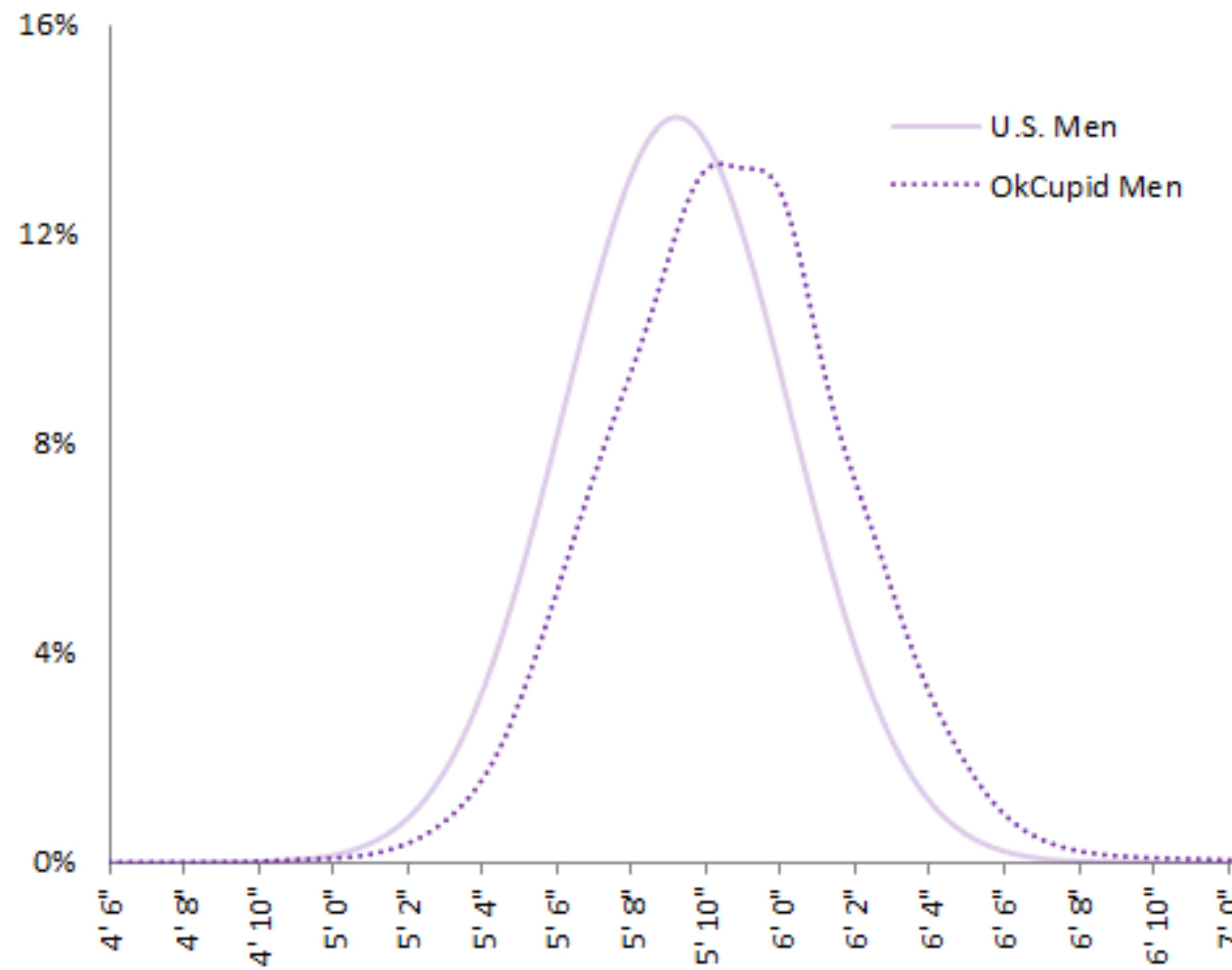




These **sources** can
be pretty important

{they can be biased}

Male Height Distribution On OkCupid



{there may be assumptions}

Street Bump

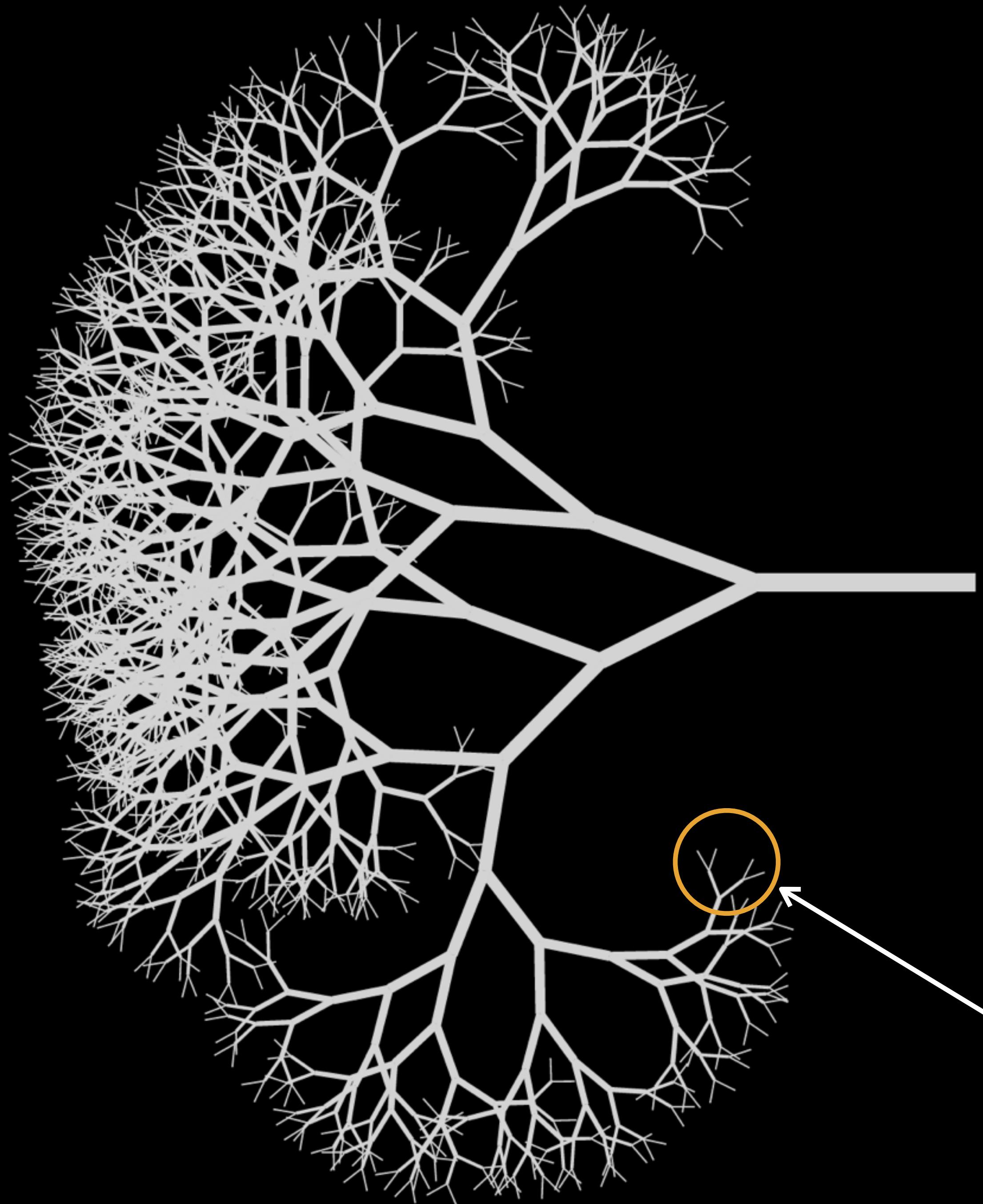
About



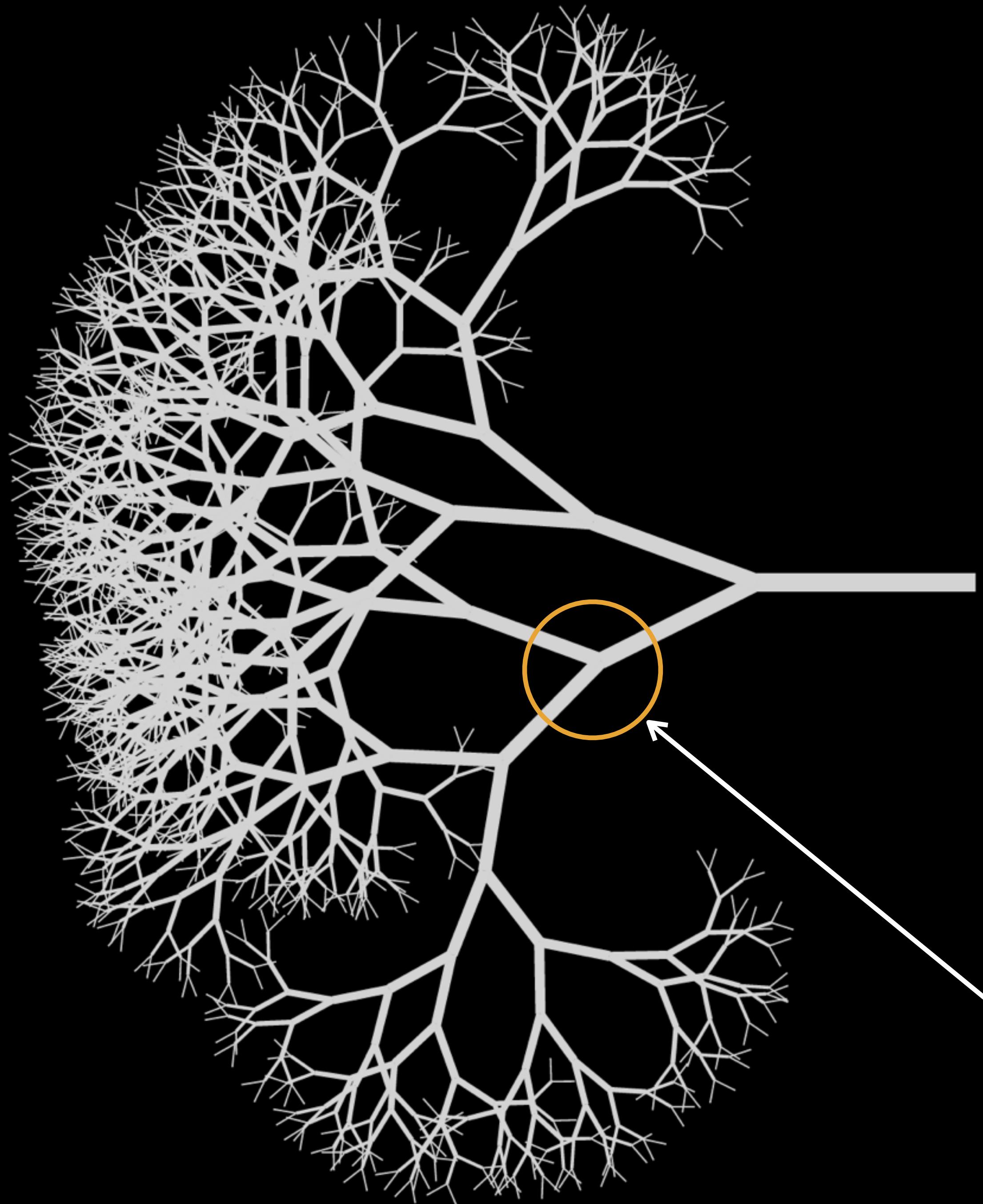
Help keep your streets smooth

RECORD A TRIP

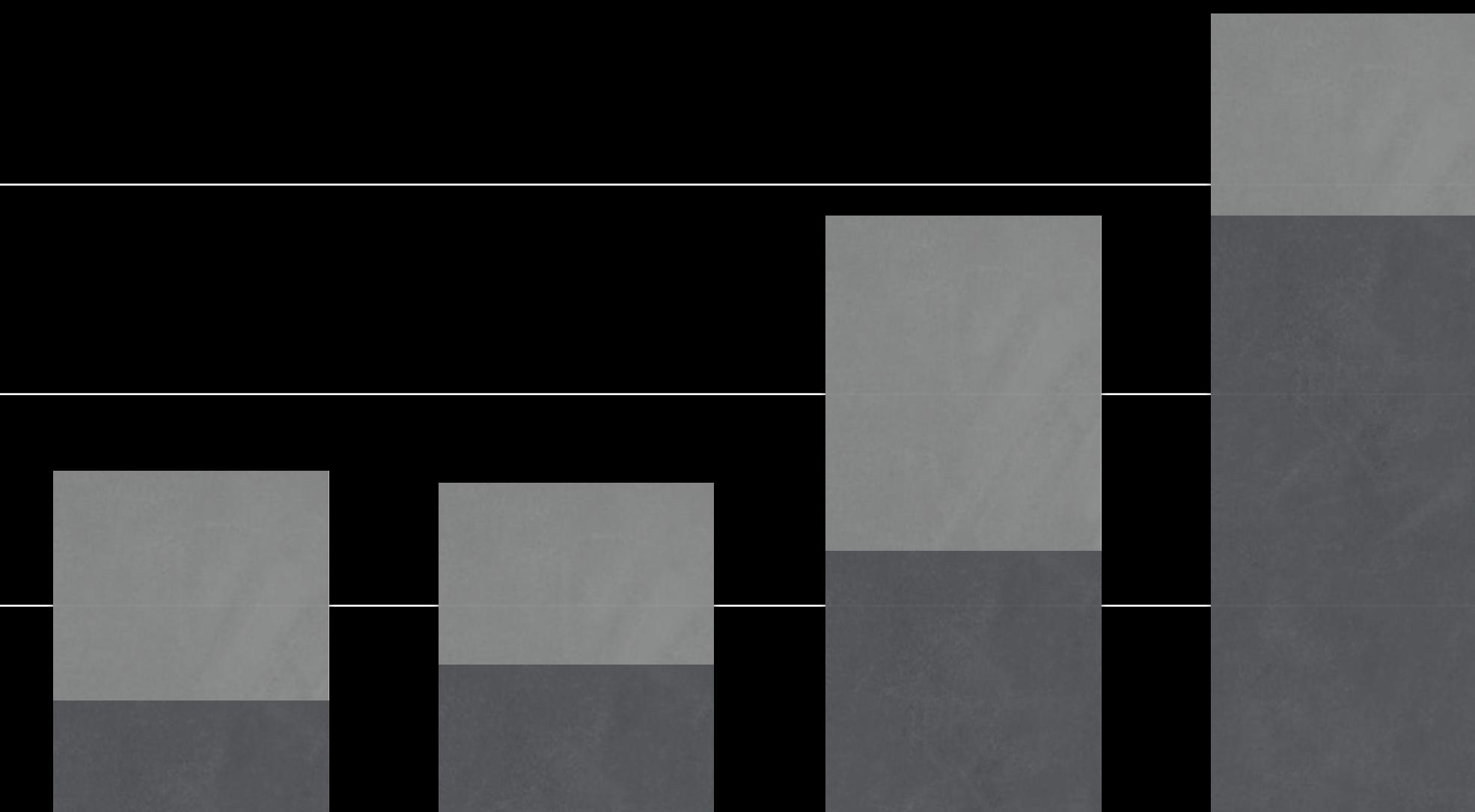
MY TRIPS

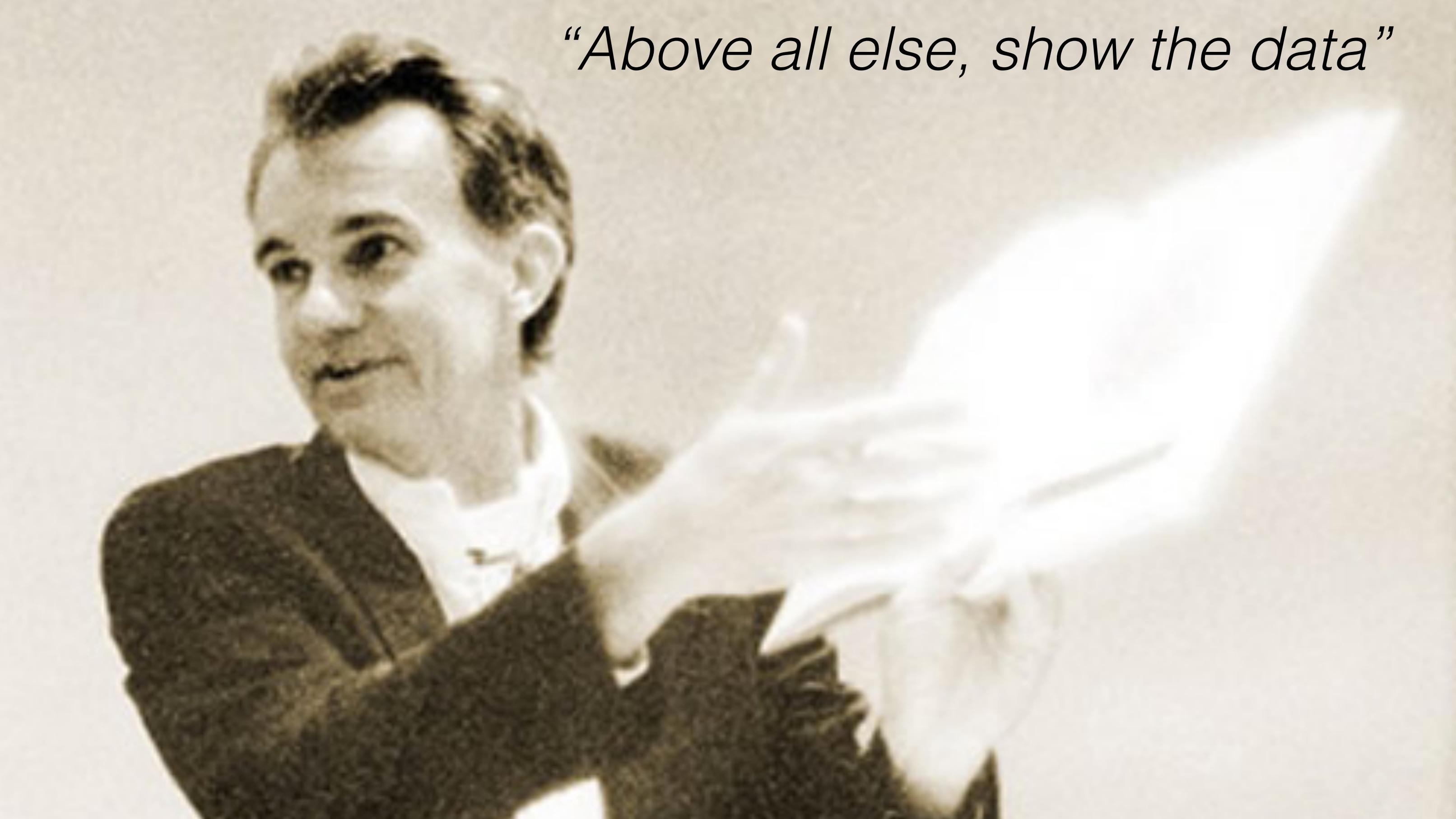


Decision-makers should
understand these sources



What happens here?





“Above all else, show the data”

{nothing}

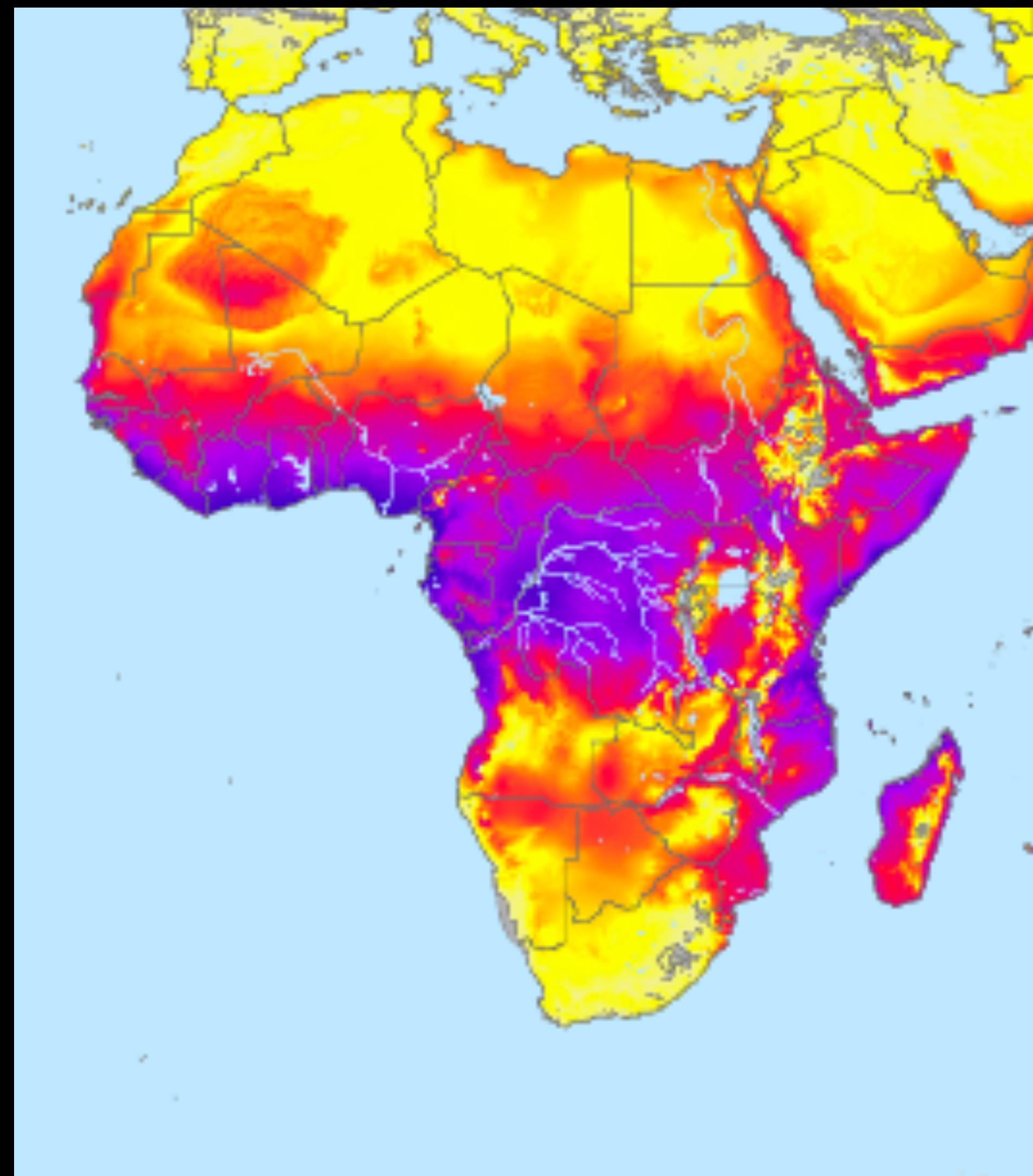


{clean}

the terrors of

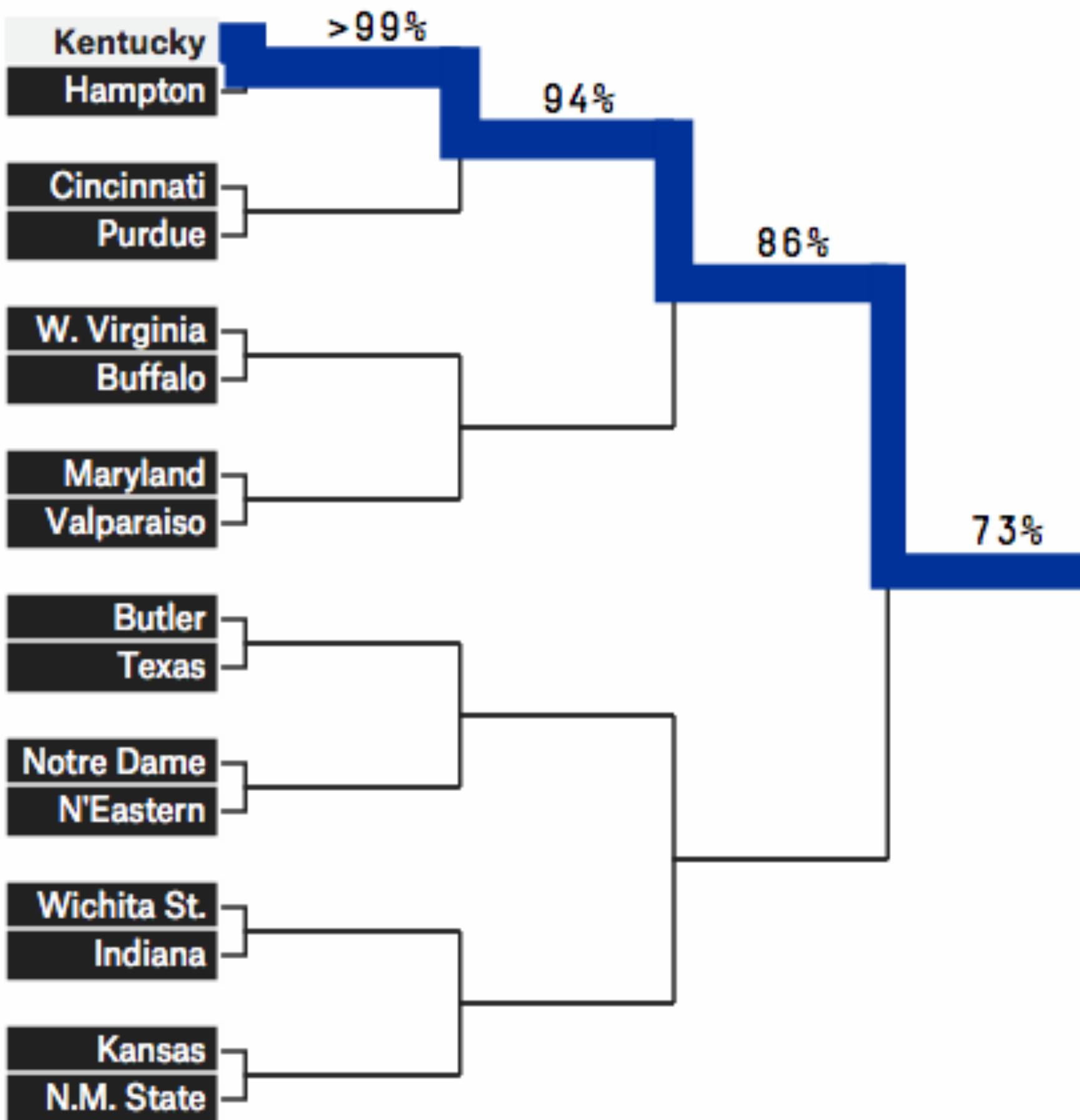
DIRTY DATA





{model}

MIDWEST



MEN'S
CURRENT PREDICTIONS

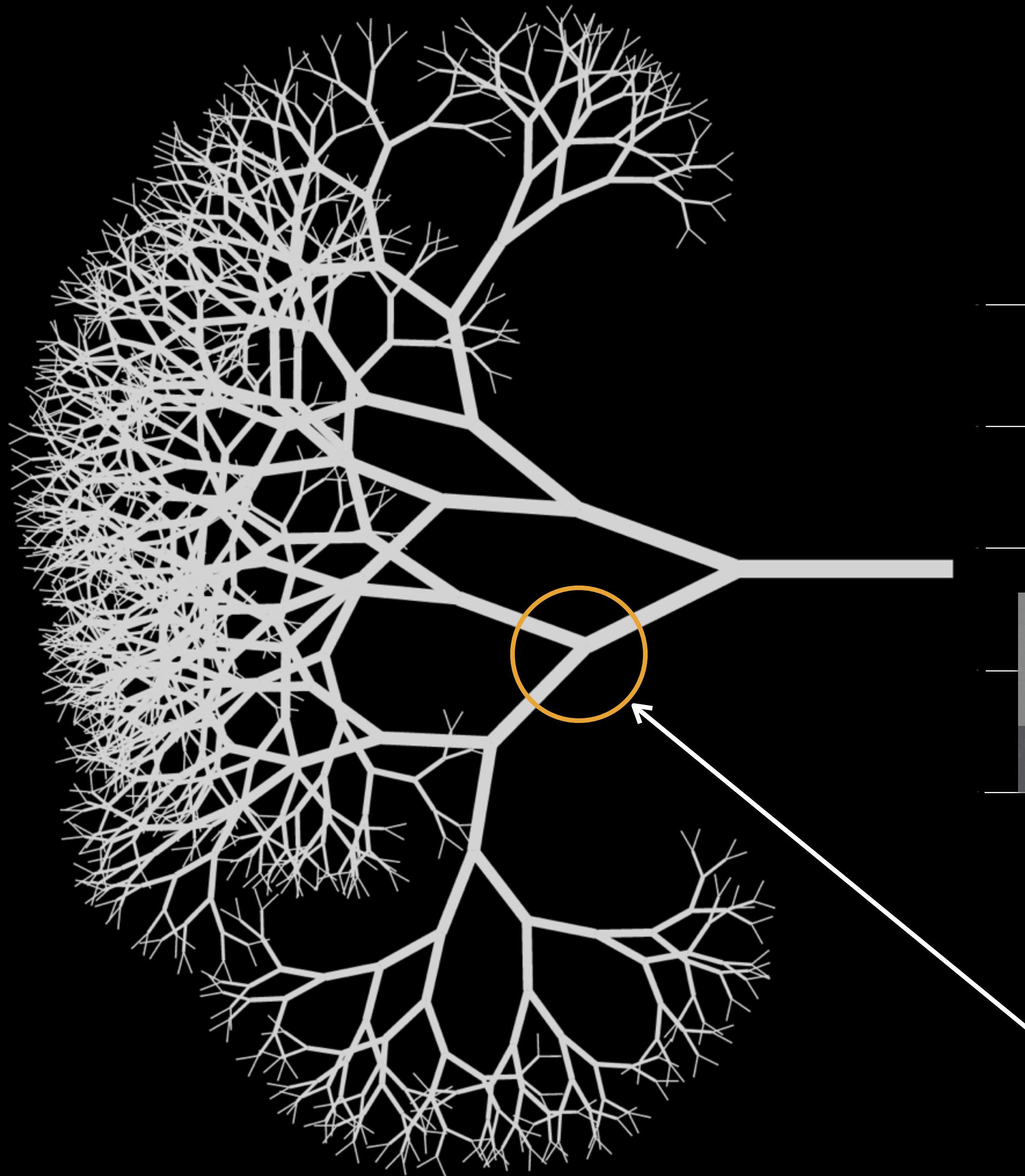
CHANCE OF
WINNING
TOURNAMENT

41%

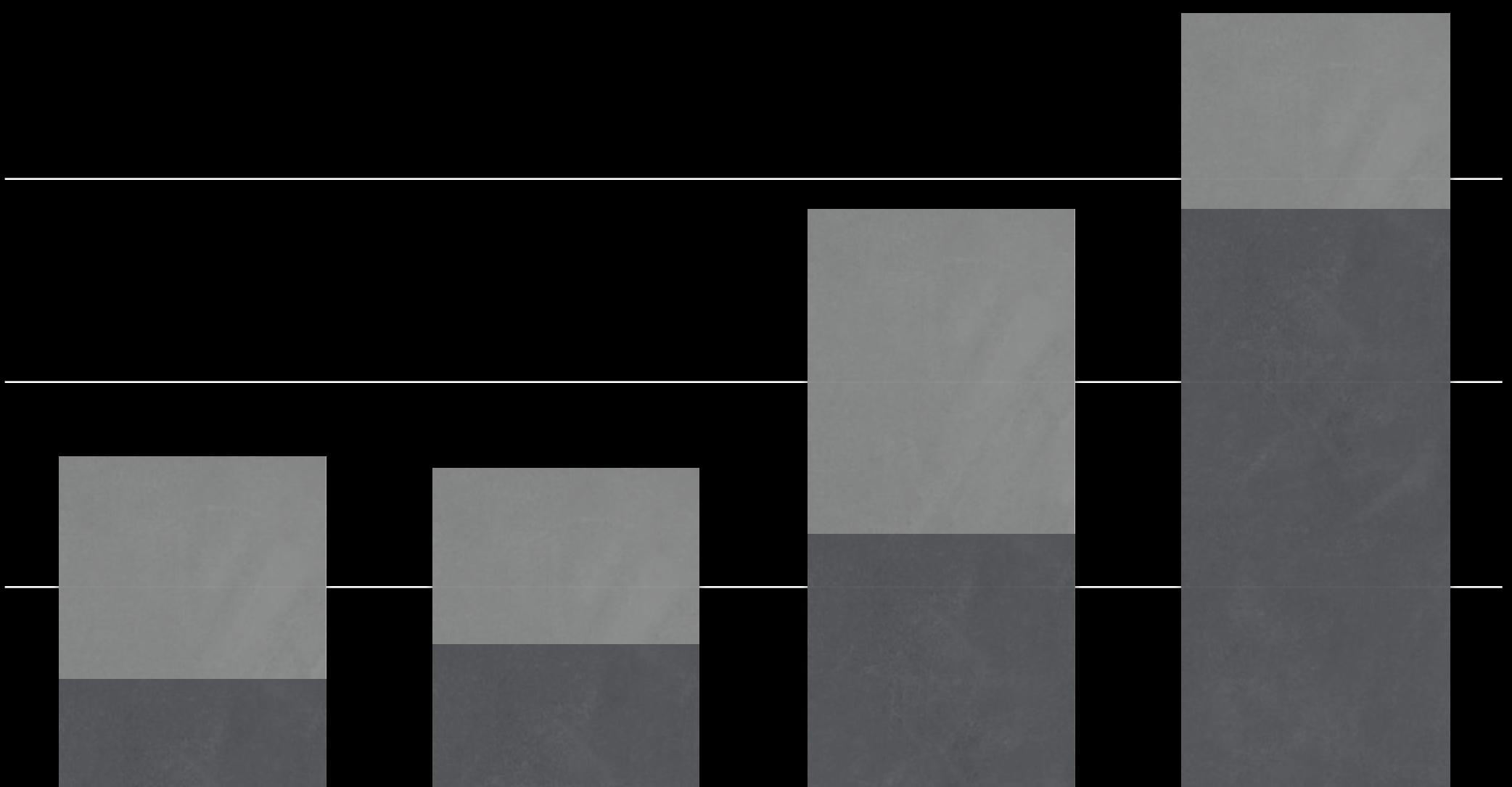


54%

{predict}



These **actions** can
be pretty important



{we make mistakes}

INVESTMENT BANKING | LEGAL/REGULATORY

Bank of America Finds a Mistake: \$4 Billion Less Capital

By PETER EAVIS and MICHAEL CORKERY APRIL 28, 2014 9:18 AM 92 Comments



Following the accounting error, Brian Moynihan, Bank of America's chief, will have to again rebuild trust in the bank. Drew Angerer/Getty Images

{we make assumptions}

How we amplify privilege with supervised machine learning

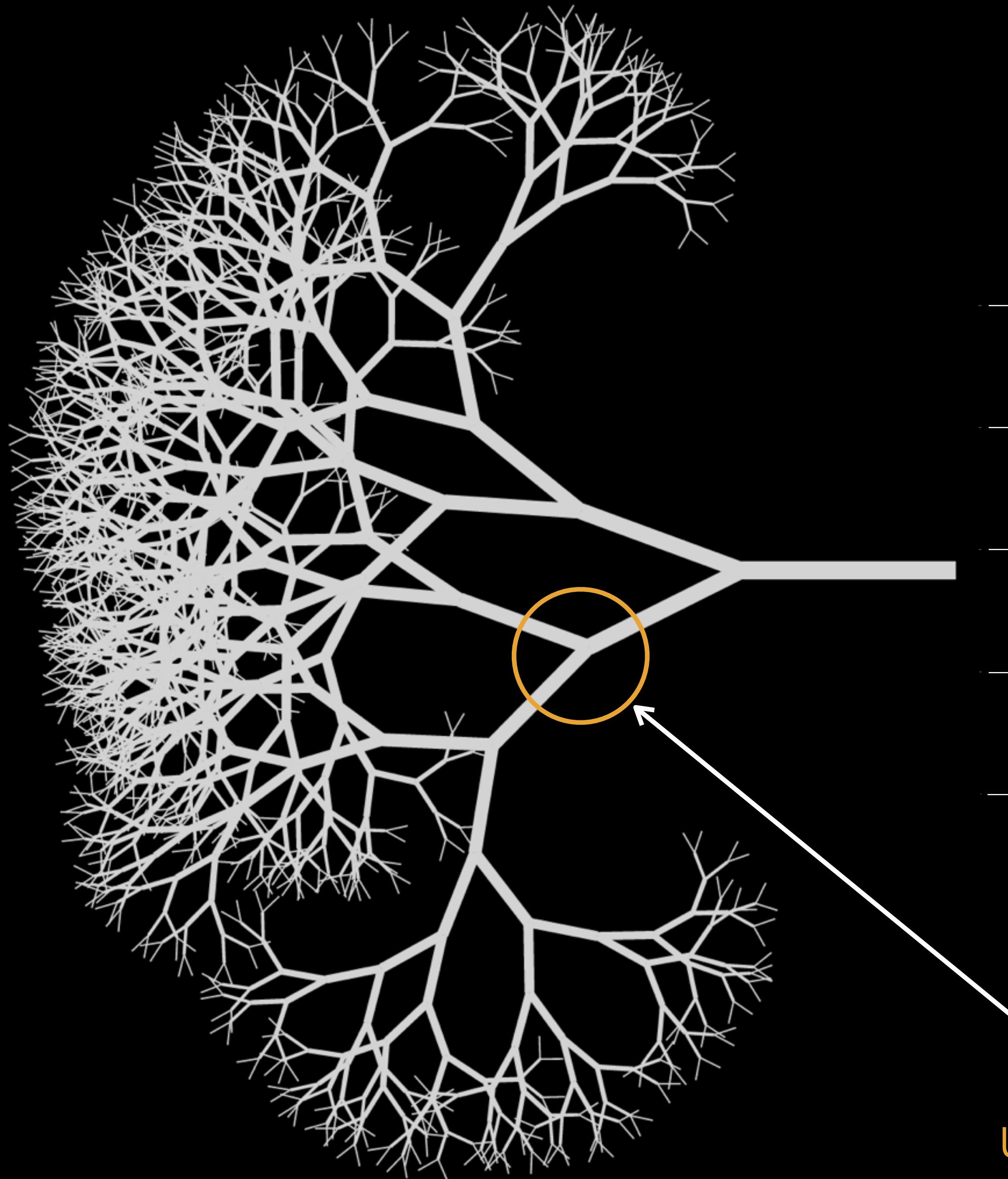
Michael Williams // @williamsmj_

Attempt to
identify
negative sentiment

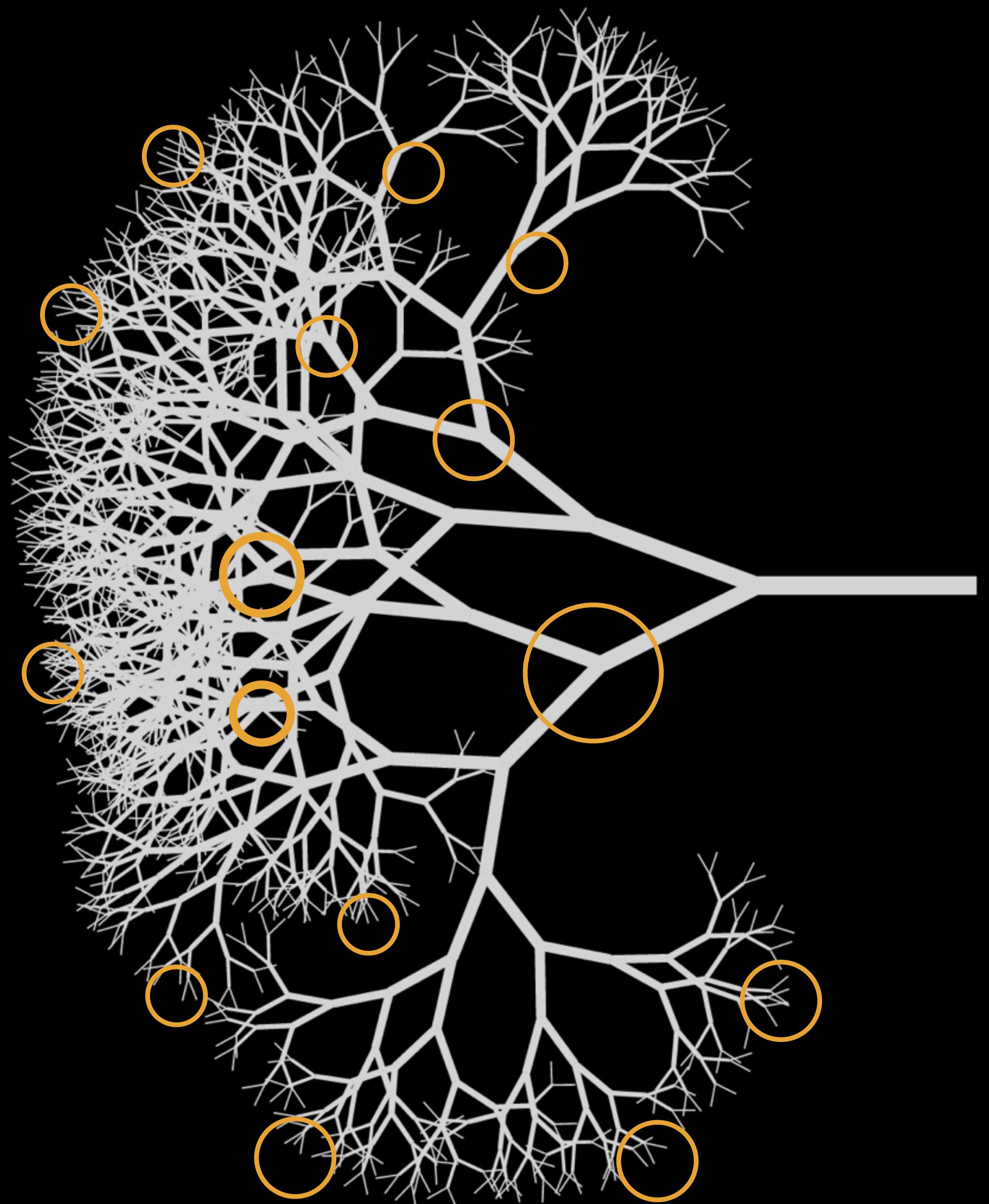
Train a tool that
identifies
harsh language

People using severe
language
get attention

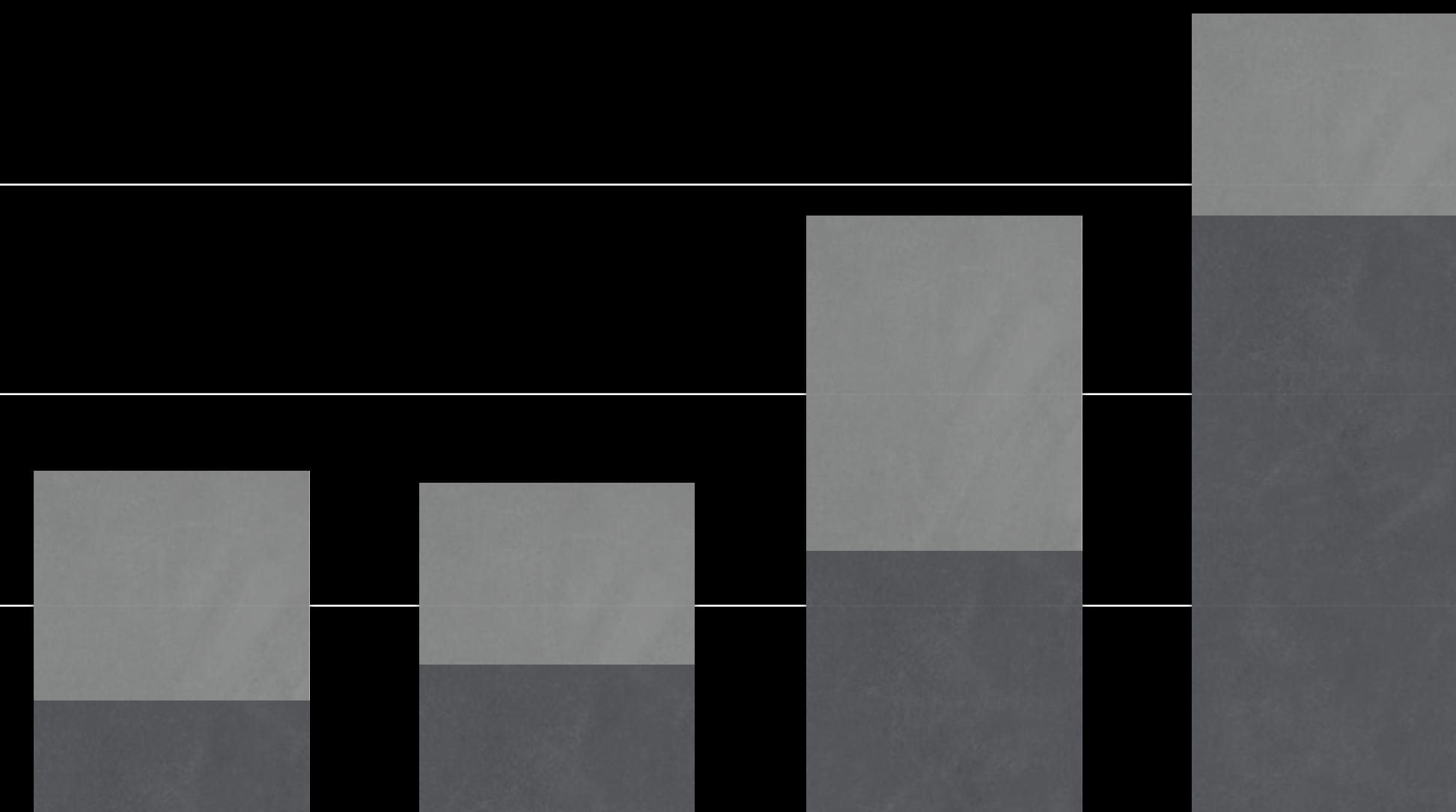
{assumes men and women use strident language at the same rate}



Decision-makers should
understand these processes



What do we talk about?



{user-centered design}

User-centered design

Who is your audience?

What is their question?

What is their level of familiarity?

What is the impact of their decisions?

{visualize}

Choose the graphical encodings...
... that are best visually decoded.

Can we do this with methods and models?

{which values do we encode?}

Visually encoding methods

Express the **structure** of the data

Encode the **process** employed

Display any **impact** on the data

{expressing structure}

Global Health Funding

Source

(where money comes from)

Governments

Gates foundation

Global Health Funding

Source

(where money comes from)

Governments

Gates foundation

Channel

(body that money is spent through)

United Nations

Global Fund

Governments

Global Health Funding

Source

(where money comes from)

Governments

Gates foundation

Channel

(body that money is spent through)

United Nations

Global Fund

Governments

Recipient

(who/what gets money)

Health focus area

Region

<http://vizhub.healthdata.org/fgh/>

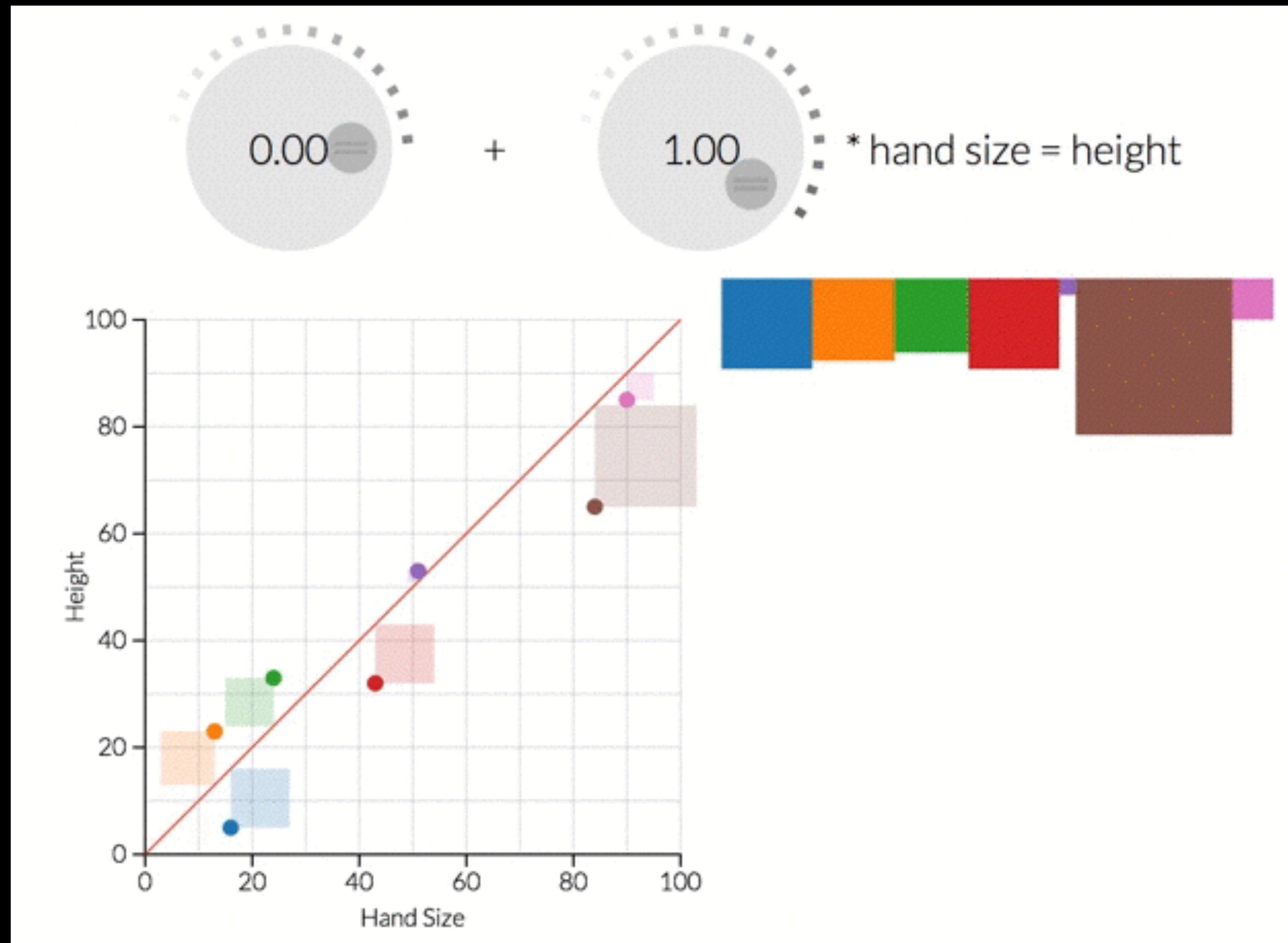
{encoding process}

Linear regression

Find a function that describes the relationship between two+ variables

Done by minimizing the squared distance between data and function

Linear regression



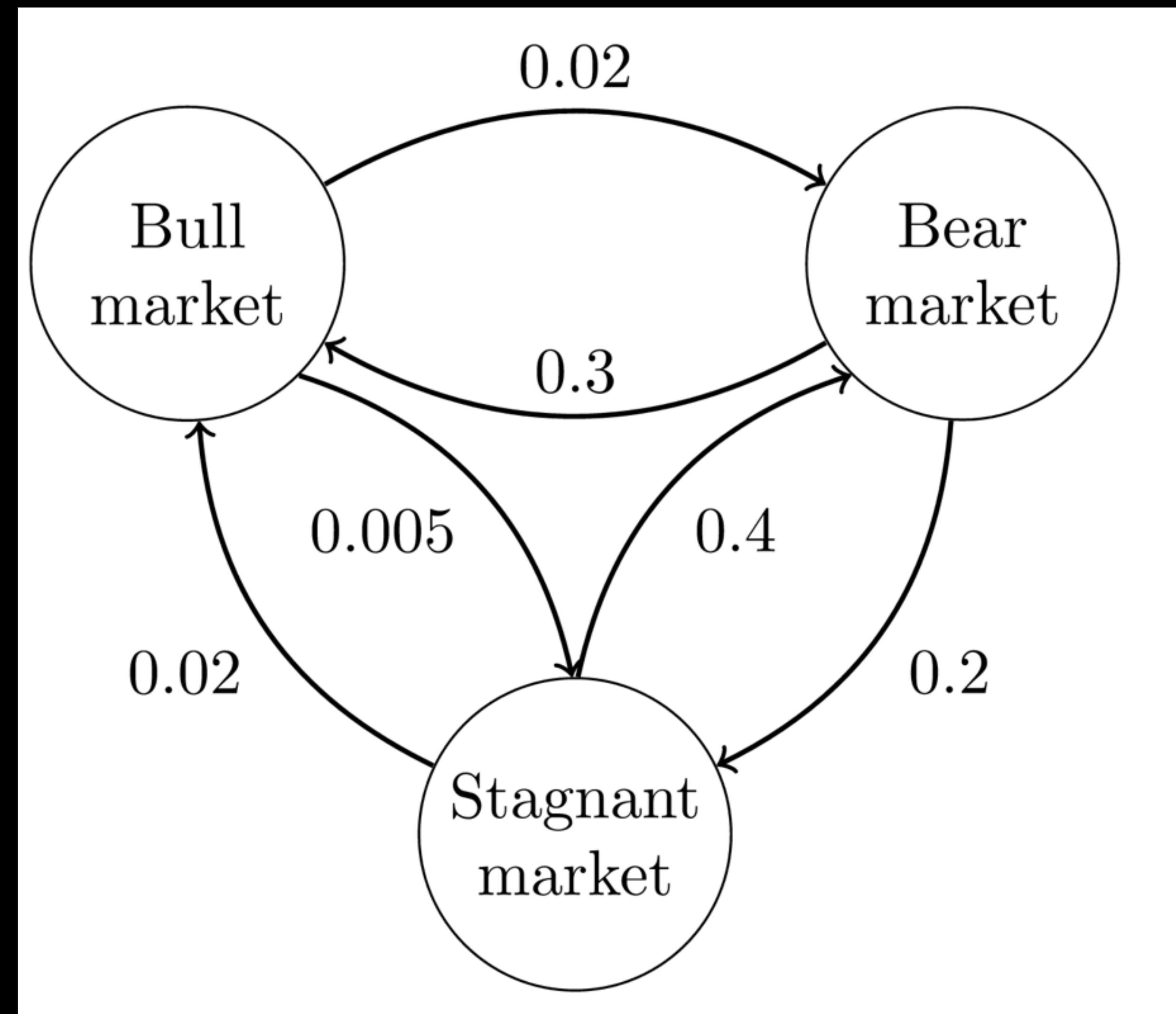
credit: setosa.io

Markov chains

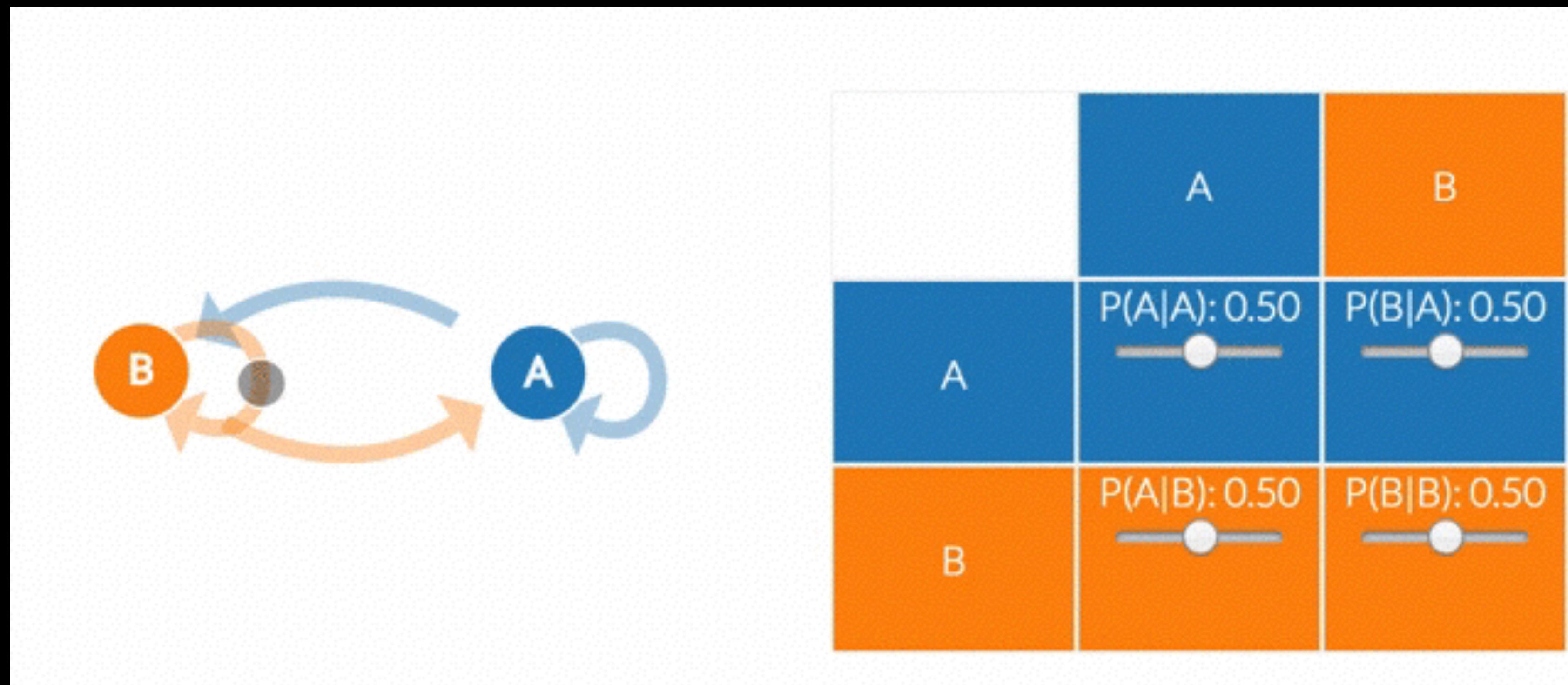
Used in simulations as a random process for state transitions

Probability distribution of the next state depends only on the current state

Markov chains



Markov chains

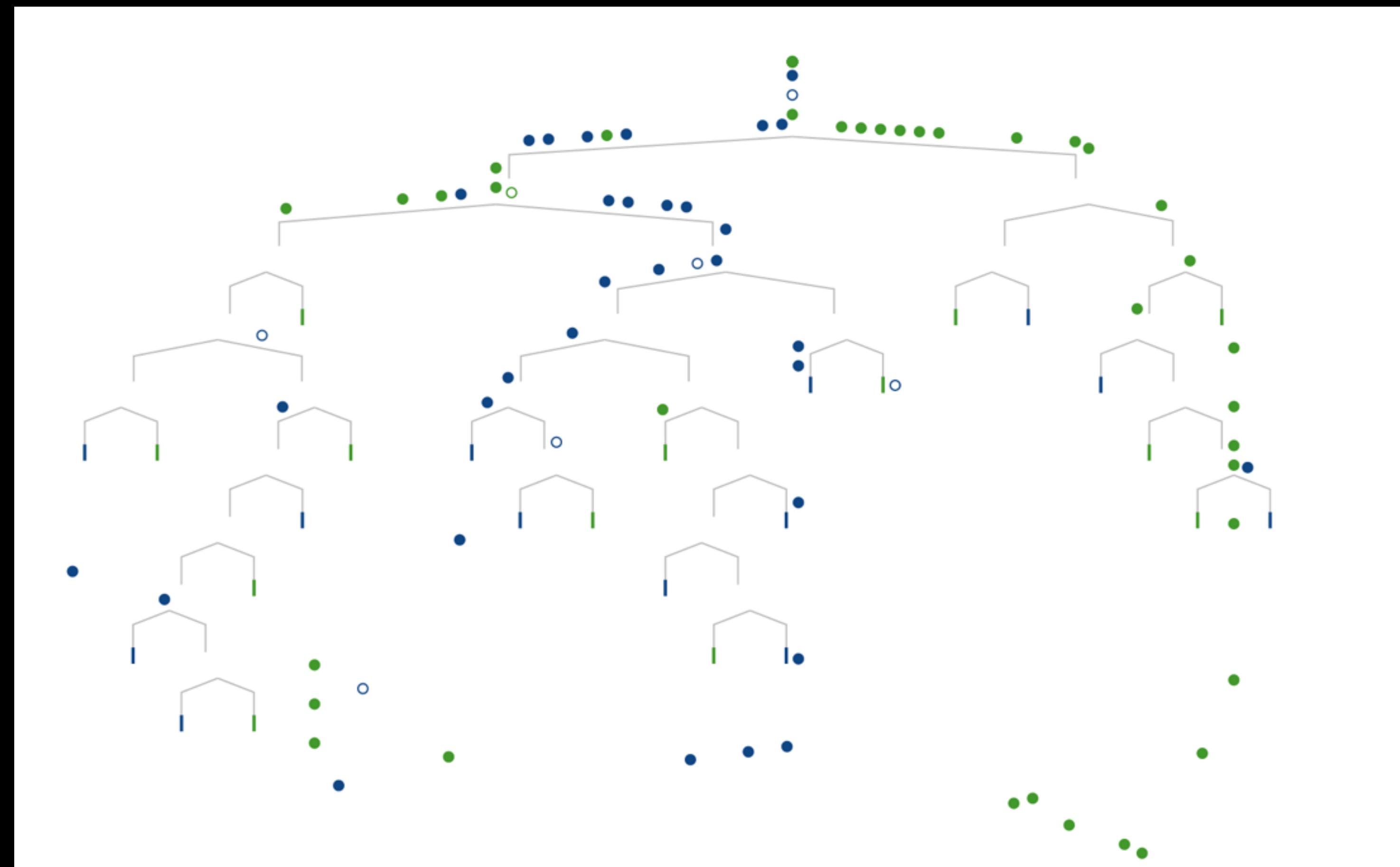


credit: setosa.io

Machine learning

{let's just look at the visual}

Machine learning

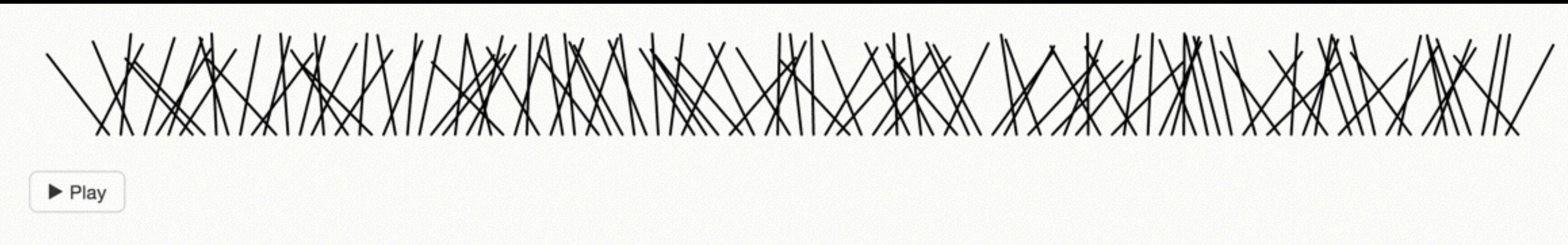


credit: r2d3.us

{displaying impact}



```
function shuffle(array) {  
  var n = array.length, t, i;  
  while (n) {  
    i = Math.random() * n-- | 0; // 0 ≤ i < n  
    t = array[n];  
    array[n] = array[i];  
    array[i] = t;  
  }  
  return array;  
}
```



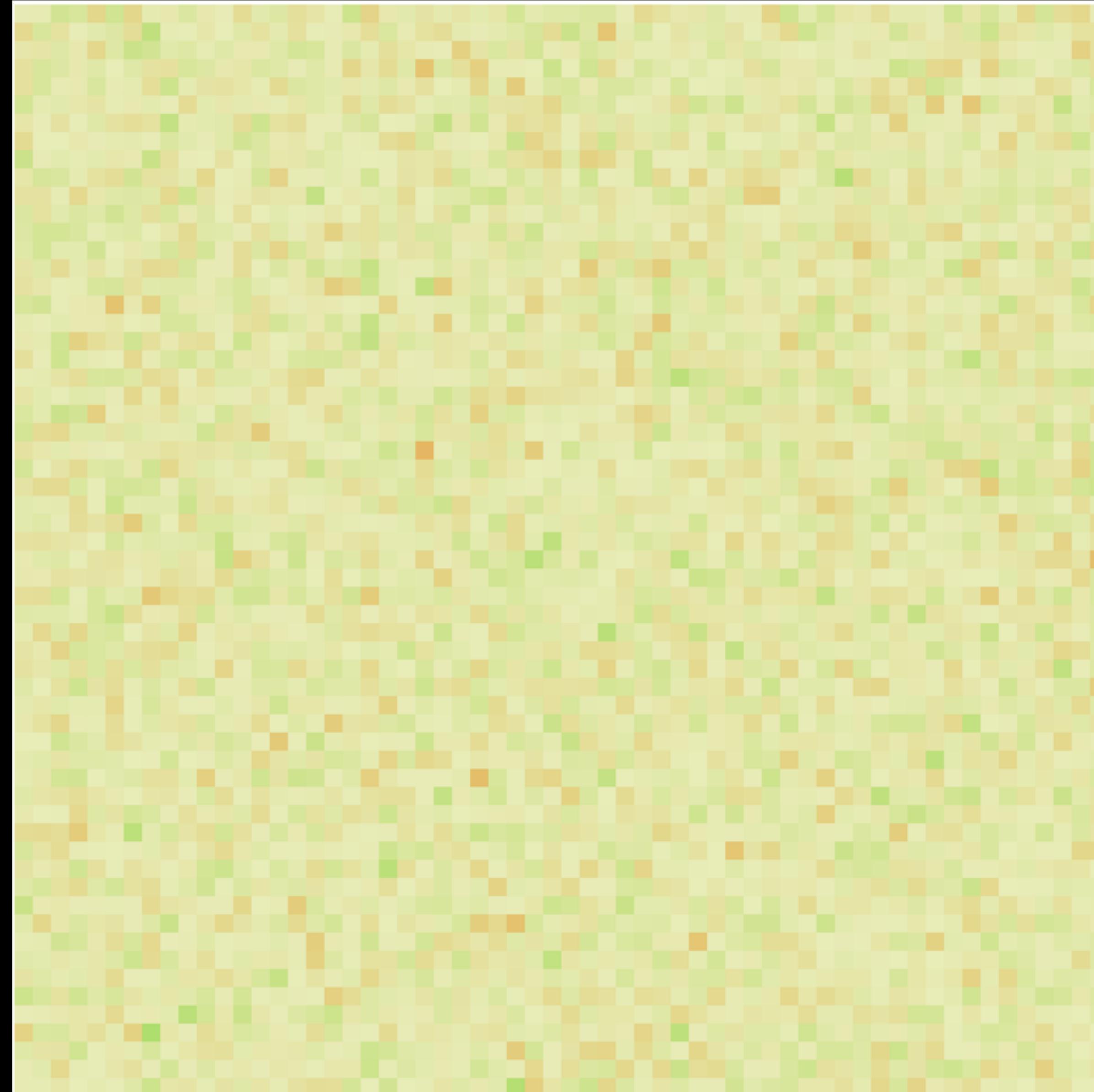
credit: mike bostock

```
function shuffle(array) {  
    var n = array.length, t, i;  
    while (n) {  
        i = Math.random() * n-- | 0; // 0 ≤ i < n  
        t = array[n];  
        array[n] = array[i];  
        array[i] = t;  
    }  
    return array;  
}
```

```
// DON'T DO THIS!  
function shuffle(array) {  
    return array.sort(function(a, b) {  
        return Math.random() - .5; // σ_σ  
    });  
}
```

shuffled position

initial position ↑

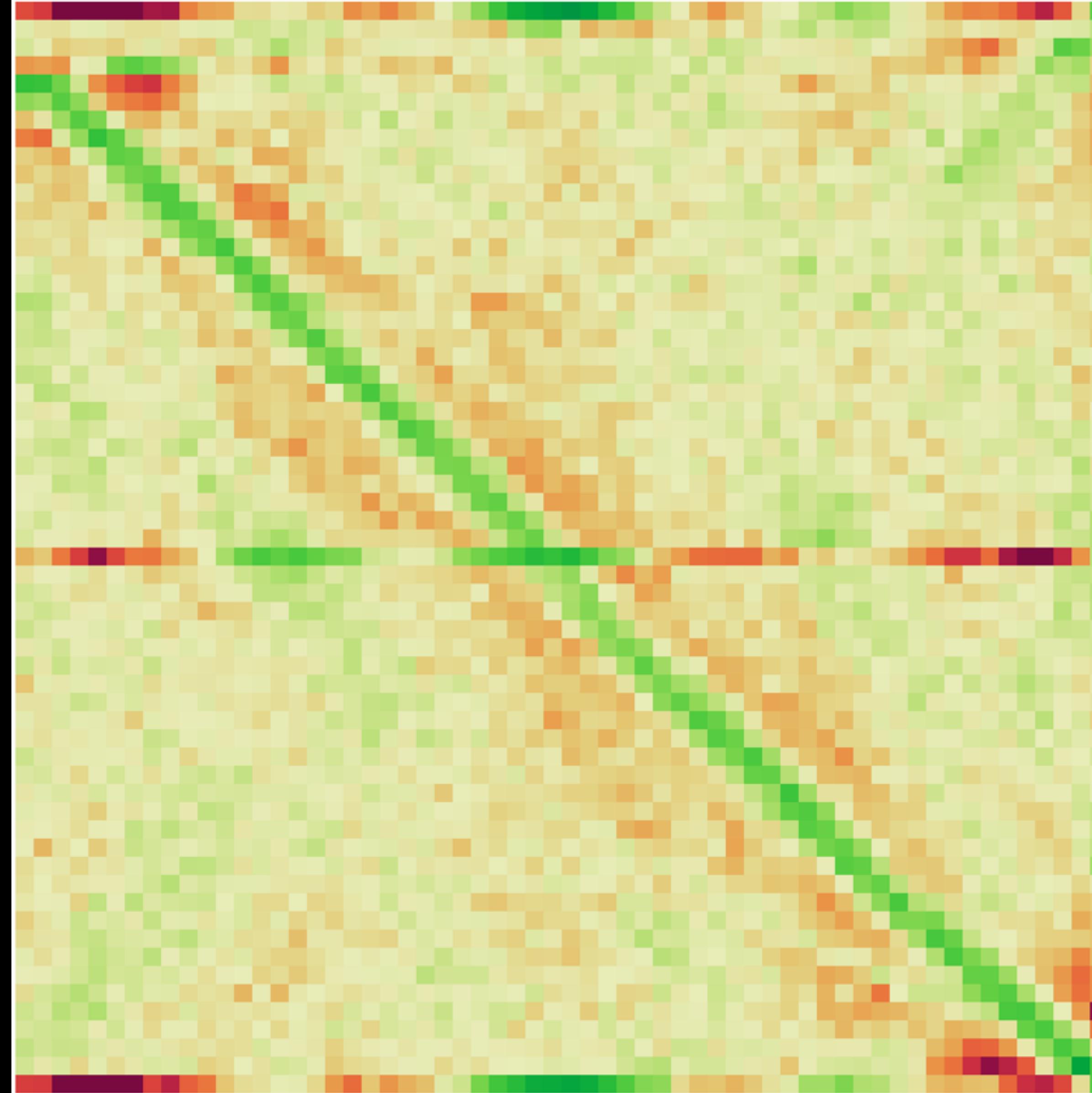


This is **good!**

credit: mike bostock

shuffled position

initial position ↑

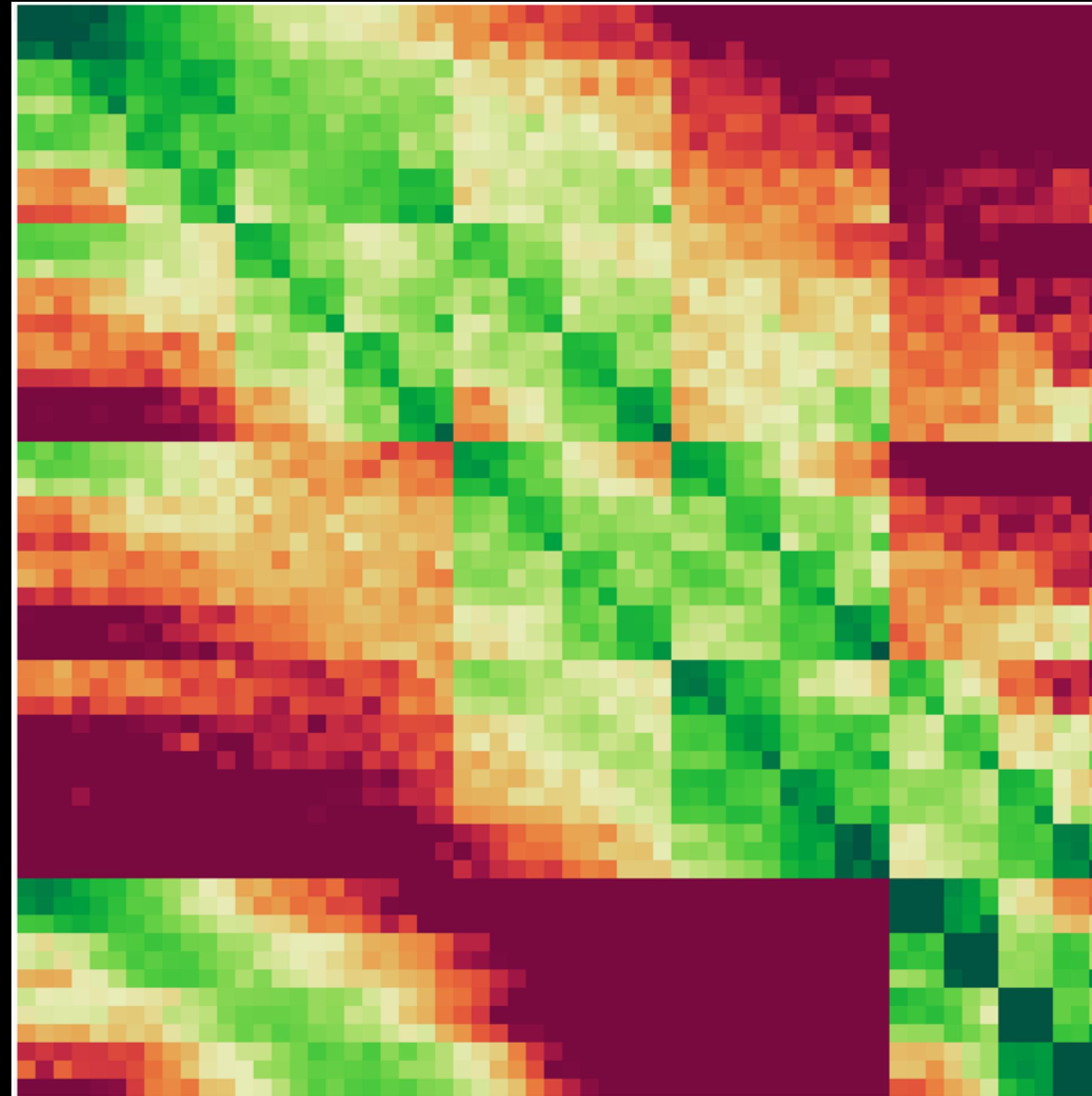


credit: mike bostock

This is **bad!**

shuffled position

initial position ↑



This is **firefox!**

credit: mike bostock

{so what should you do?}

Understand
Users

Express
Structure

Encode
Process

Display
Impact

Credits

Financing Global Health (IHME)

<http://vizhub.healthdata.org/fgh/>

Linear Regression (setosa.io)

<http://setosa.io/ev/ordinary-least-squares-regression/>

Markov Chains (setosa.io)

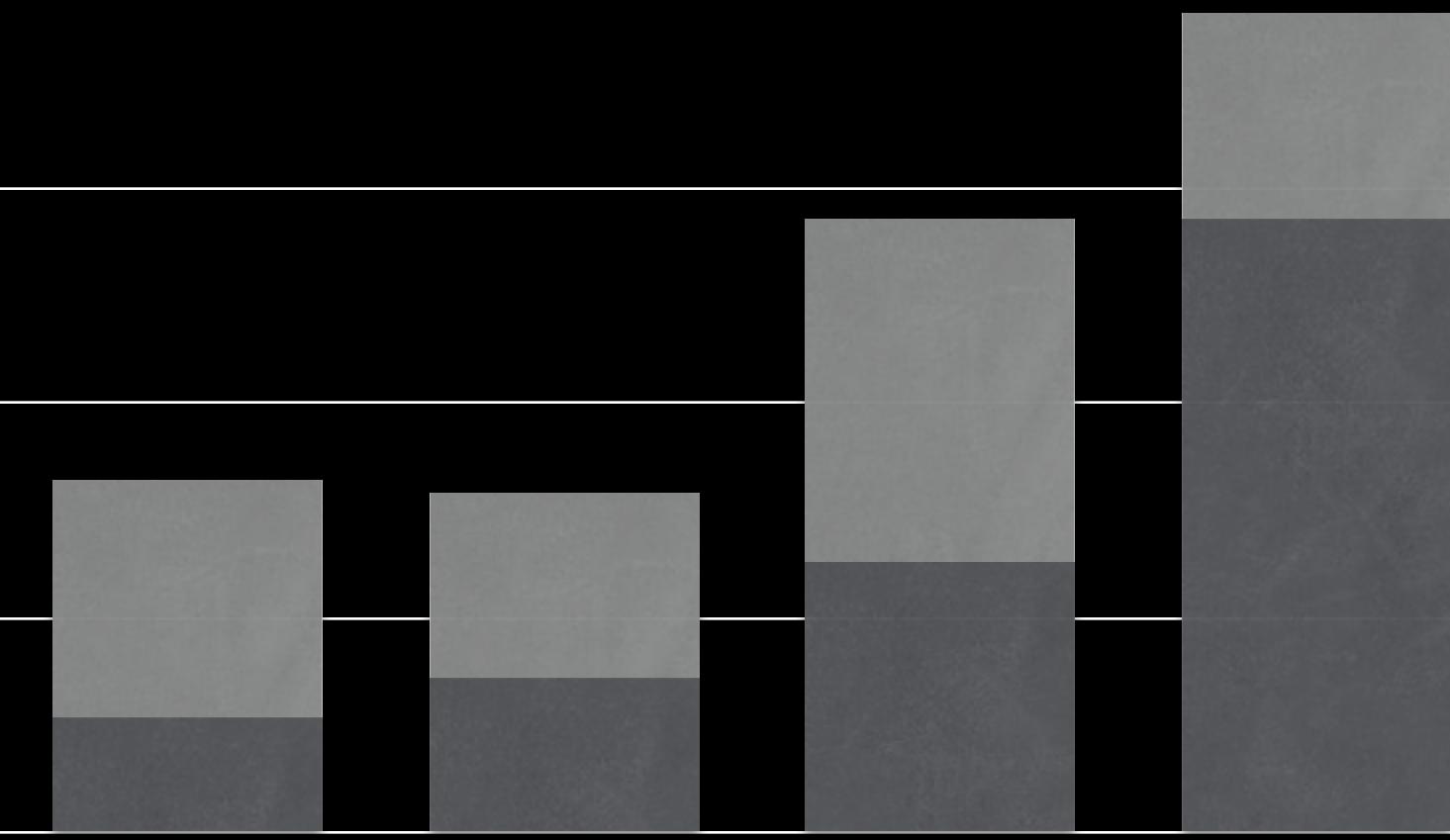
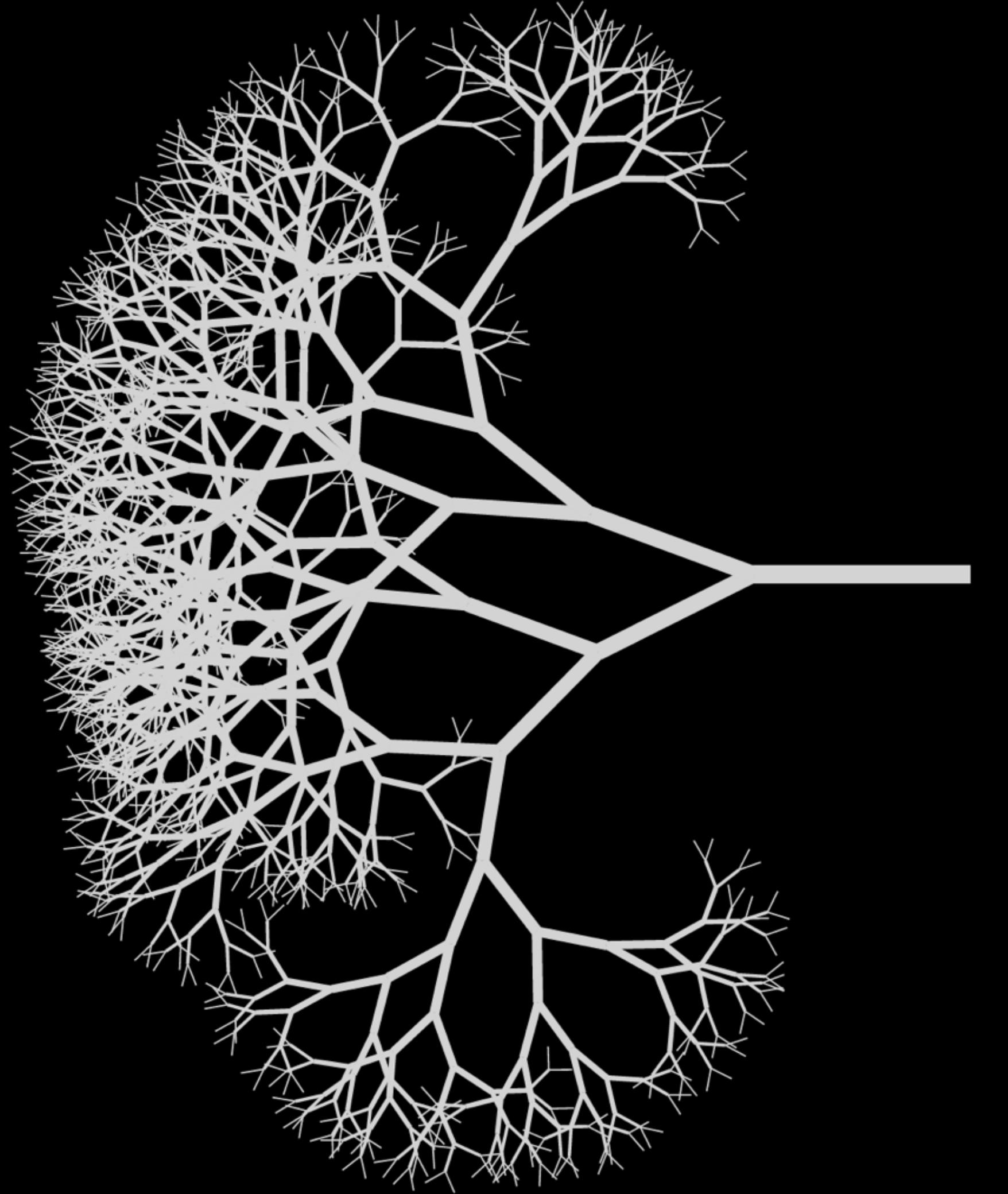
<http://setosa.io/ev/markov-chains/>

Machine Learning (r2d3)

<http://www.r2d3.us/visual-intro-to-machine-learning-part-1/>

Shuffling (Mike Bostock)

<http://bost.ocks.org/mike/algorithms/#shuffling>



In order to understand data...

... you need to understand what you did to it.

Thank you

Michael Freeman

mikefree@uw.edu

mfviz.com

@mf_viz

mfviz.com/strata-2015