

# Machine Learning Workshop

2019-12-30

Paul Kaefer

Instructor, Data Scientist, etc.

M.S., B.S.

```
print(model.summary())
```

```
Optimization terminated successfully.
Current function value: 0.000000
Iterations: 7
```

```
=====
Dep. Variable:
Model:
```



# Who am I?

Senior Analytics Engineer at



Former Data Scientist II at [UnitedHealthcare](#)

Instructor, [The Coding Bootcamp](#)  
Data Analytics & Visualization

M.S., Computational Sciences

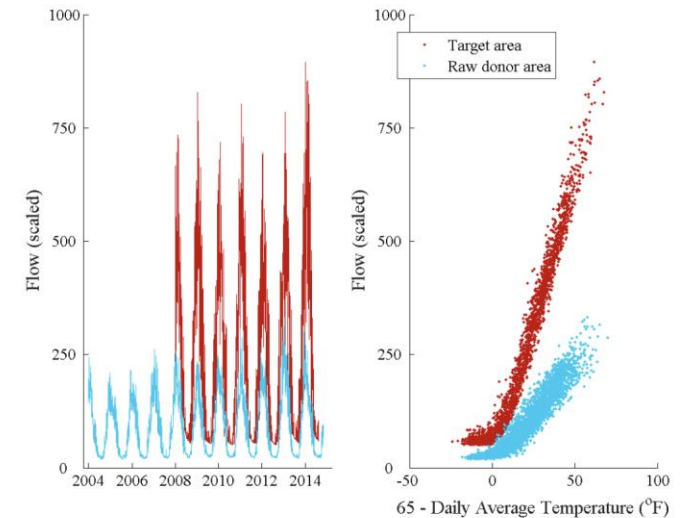


Figure 4.1: Target area from the East Coast and potential donor area from the Southwest.

# Who am I?

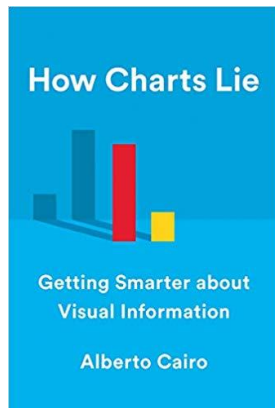


Co-author of *Introduction to Python for Business and the Social Sciences*  
(Expected publication Summer 2020)

Wikipedian



Reader, world traveler



# AI vs. ML vs. data science vs. ...

There are [many different definitions](#).<sup>1</sup>

1. See also definitions on Wikipedia (I'd start with [en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)), blog articles/opinion pieces like [this one](#), and books like *Data Analysis from Scratch with Python*. Peters Morgan. CreateSpace Independent Publishing Platform: 2016.

# AI vs. ML vs. data science vs. ...

There are [many different definitions](#).<sup>1</sup>

Machine learning refers to statistical techniques used to detect patterns and trends and train models.

1. See also definitions on Wikipedia (I'd start with [en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)), blog articles/opinion pieces like [this one](#), and books like *Data Analysis from Scratch with Python*. Peters Morgan. CreateSpace Independent Publishing Platform: 2016.

# AI vs. ML vs. data science vs. ...

There are [many different definitions](#).<sup>1</sup>

Machine learning refers to statistical techniques used to detect patterns and trends and train models.

Let's skip the philosophical discussion & get to methods.

1. See also definitions on Wikipedia (I'd start with [en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)), blog articles/opinion pieces like [this one](#), and books like *Data Analysis from Scratch with Python*. Peters Morgan. CreateSpace Independent Publishing Platform: 2016.



# My approach

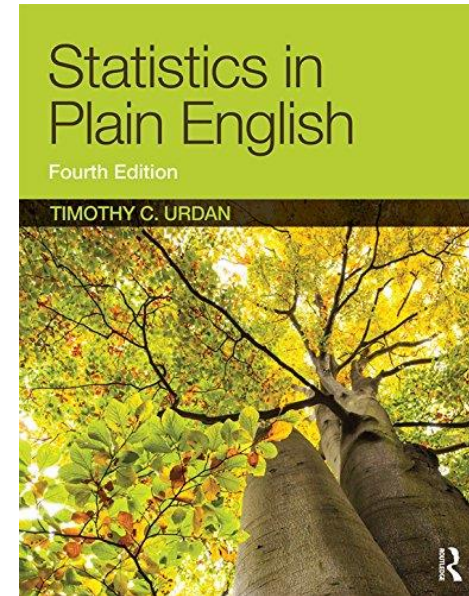


I've also led statistics workshops in peri-urban Tanzania.

## Focus on the basics!

- deep understanding of the basics is vital to properly applying and communicating “advanced” methods
- Linear models can outperform NNs

Please stop me at any point with questions.





# Let's do an example!

From my in-progress book

Data from the General Social Survey (GSS)

Has been given since 1972; annual data

[gss.norc.org](http://gss.norc.umd.edu)

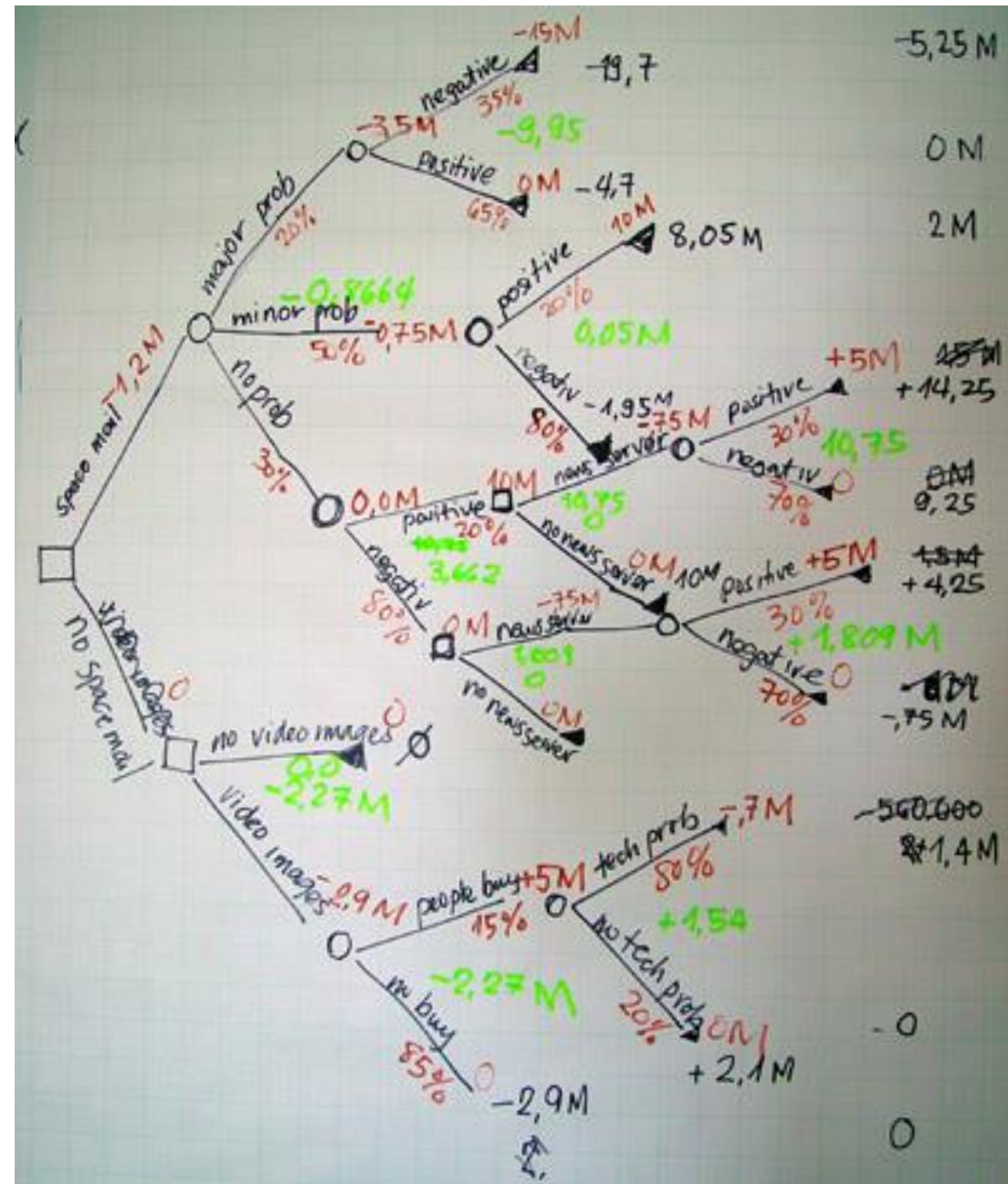
Let's check out the notebook!



# Géron example #1

Decision trees

See notebook!



As time permits...

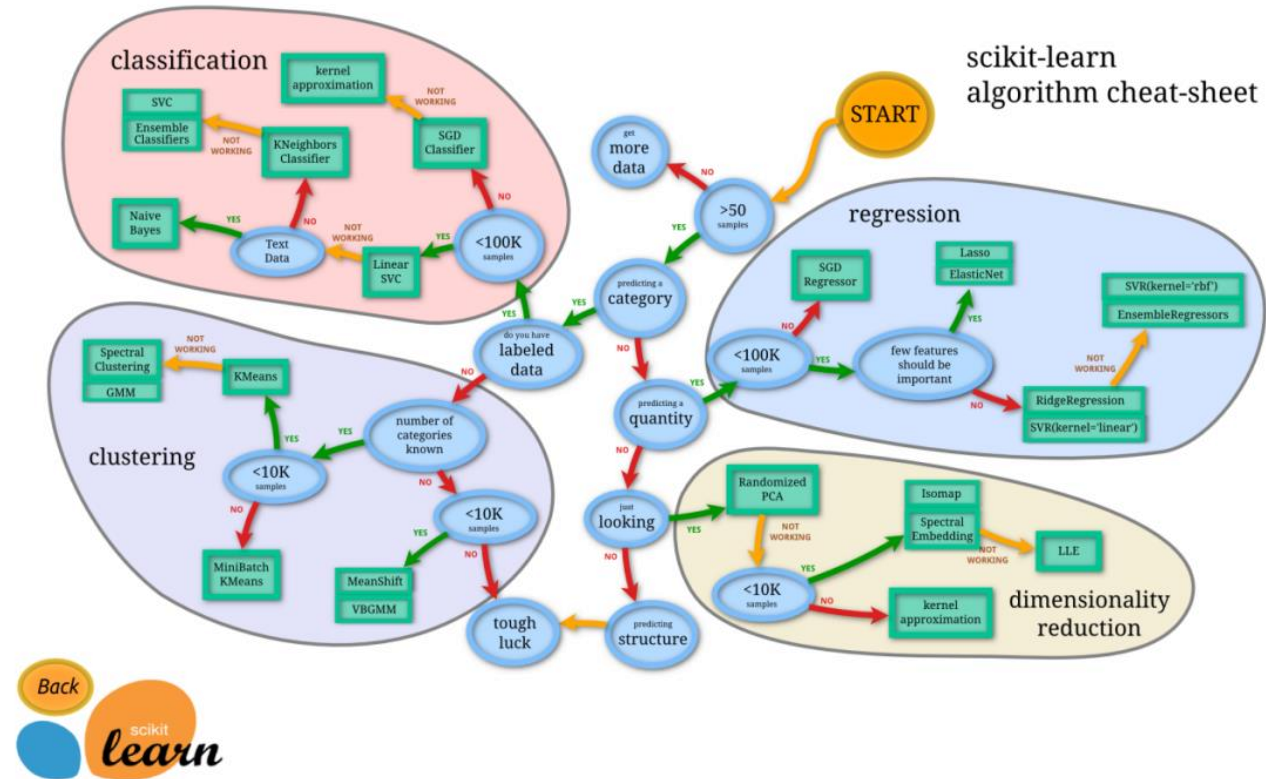
Géron Example #2: Random forrests, option of bagging.

# Where to go from here?

*Hands-On Machine Learning with Scikit-Learn and TensorFlow*, by Aurélien Géron

Click on the image! →

My book (to be released 2020)!



# Where to go from here?

Reference and further learning:

*Statistics in Plain English*, by Timothy C. Urdan

*An Introduction to Statistical Methods and Data Analysis*,  
by R. Lyman Ott and Michael T. Longnecker

Reading/thought-provoking:

*Statistics Done Wrong: The Woefully Complete Guide*,  
by Alex Reinhart

*How Charts Lie*, by Alberto Cairo

# What comes next?

Are you looking for a job? [carrothealth.com/jobs](https://carrothealth.com/jobs)

...or a job? [trilogyed.com/about/careers](https://trilogyed.com/about/careers)

Let's chat!

Let's connect either way: [linkedin.com/in/paulkaefer](https://linkedin.com/in/paulkaefer)