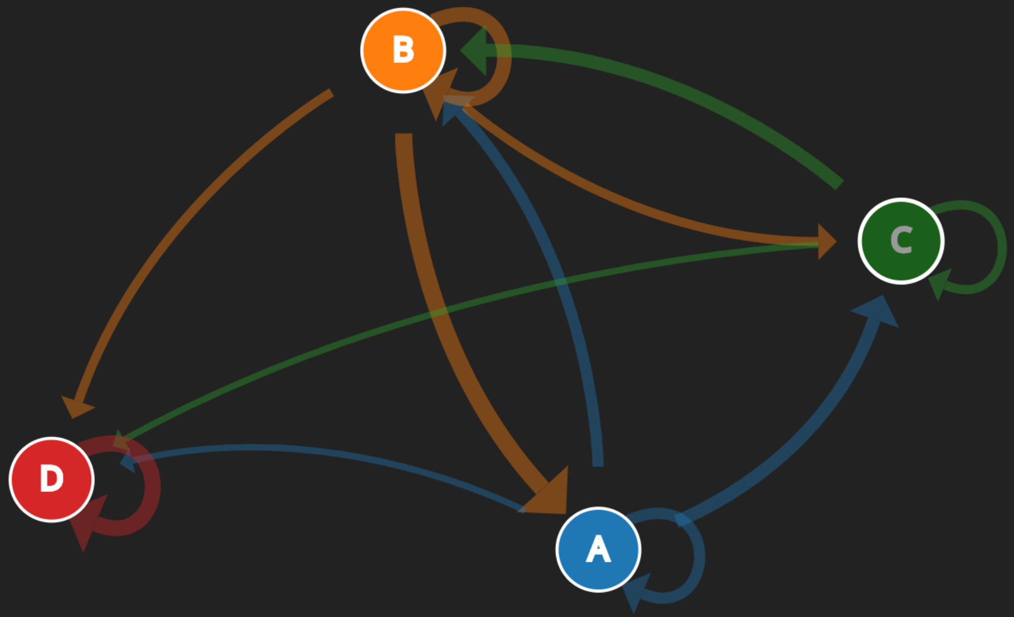


CVEN 6833 - Advanced Data Analysis

November 16, 2017

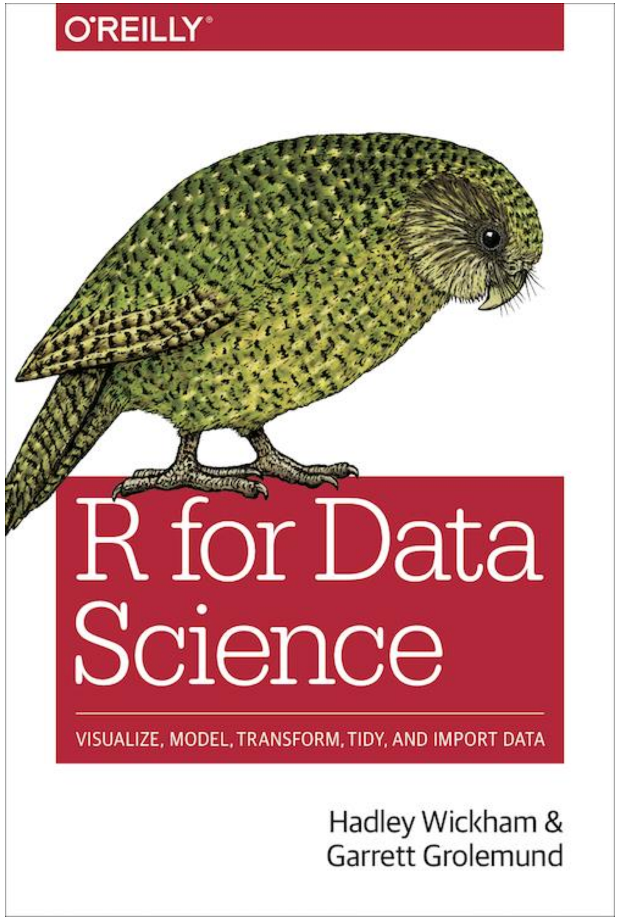
Billy Raseman

Hidden Markov Models

An aside…helpful R resources

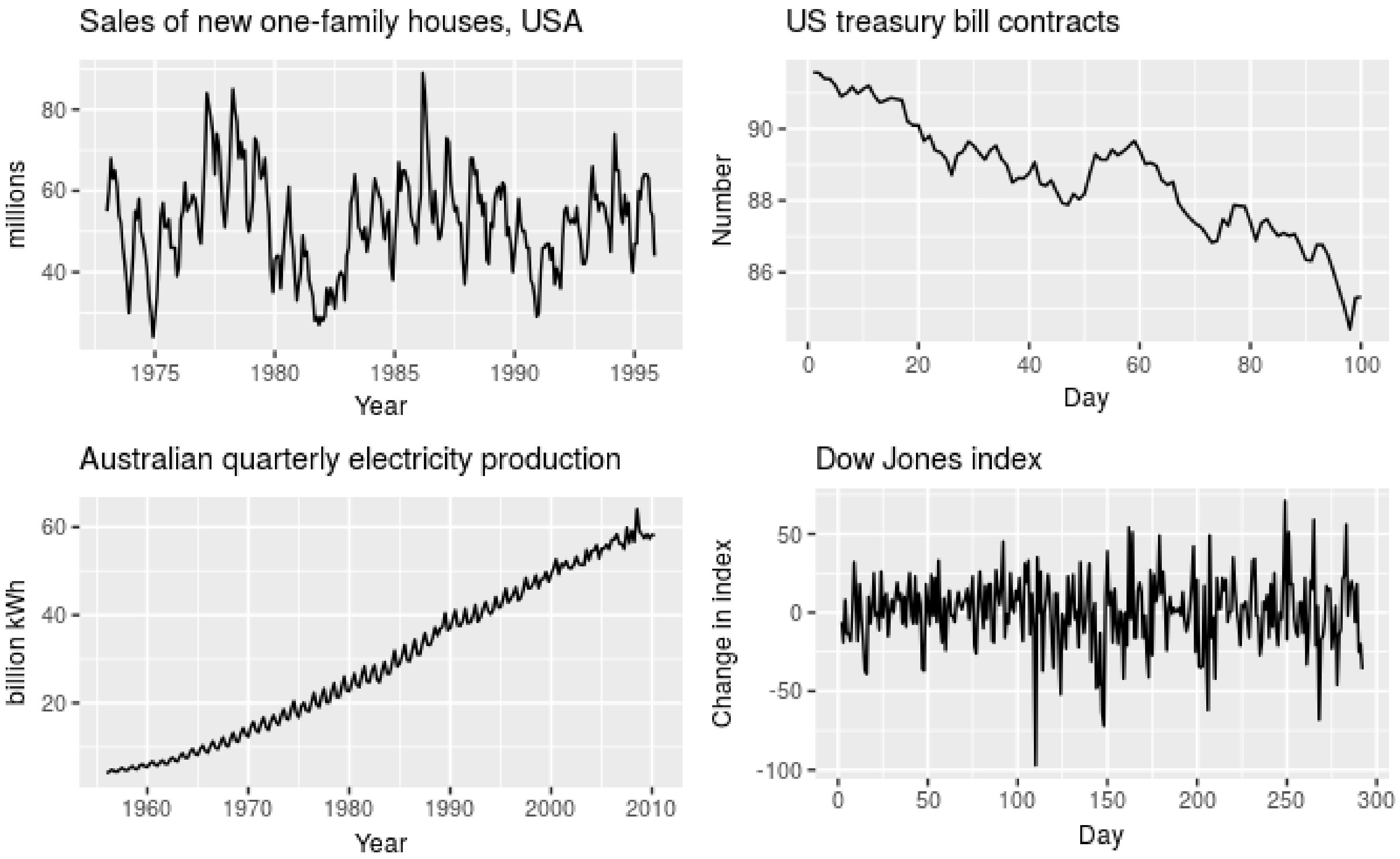
R for Data Science

- free online book: http://r4ds.had.co.nz/
- content: steps through visualizing, modeling, data wrangling and importation
- uses popular modern R packages known collectively as the ‘tidyverse’: ggplot2, dplyr, readr, etc.



Time Series Analysis and Forecasting in R

- free online book: otexts.org/fpp2/
- content: a modern R implementation of forecasting, time series analysis and visualization, with example datasets built-in



Lesson Goals

Following this lecture, you should be able to answer the following questions:

- What is a Markov chain?
- What is the difference between a Markov model and a Hidden Markov model (HMM)?
- What is the purpose of the following algorithms: Viterbi and Baum-Welch?
- When should you try to apply HMMs?

Following this lecture, you should be able to fit an HMM to a given timeseries dataset and simulate from that model.

Motivation: persistence in Nile River flow

source: http://kos2013.org/img/HEH_formal_portrait.jpeg

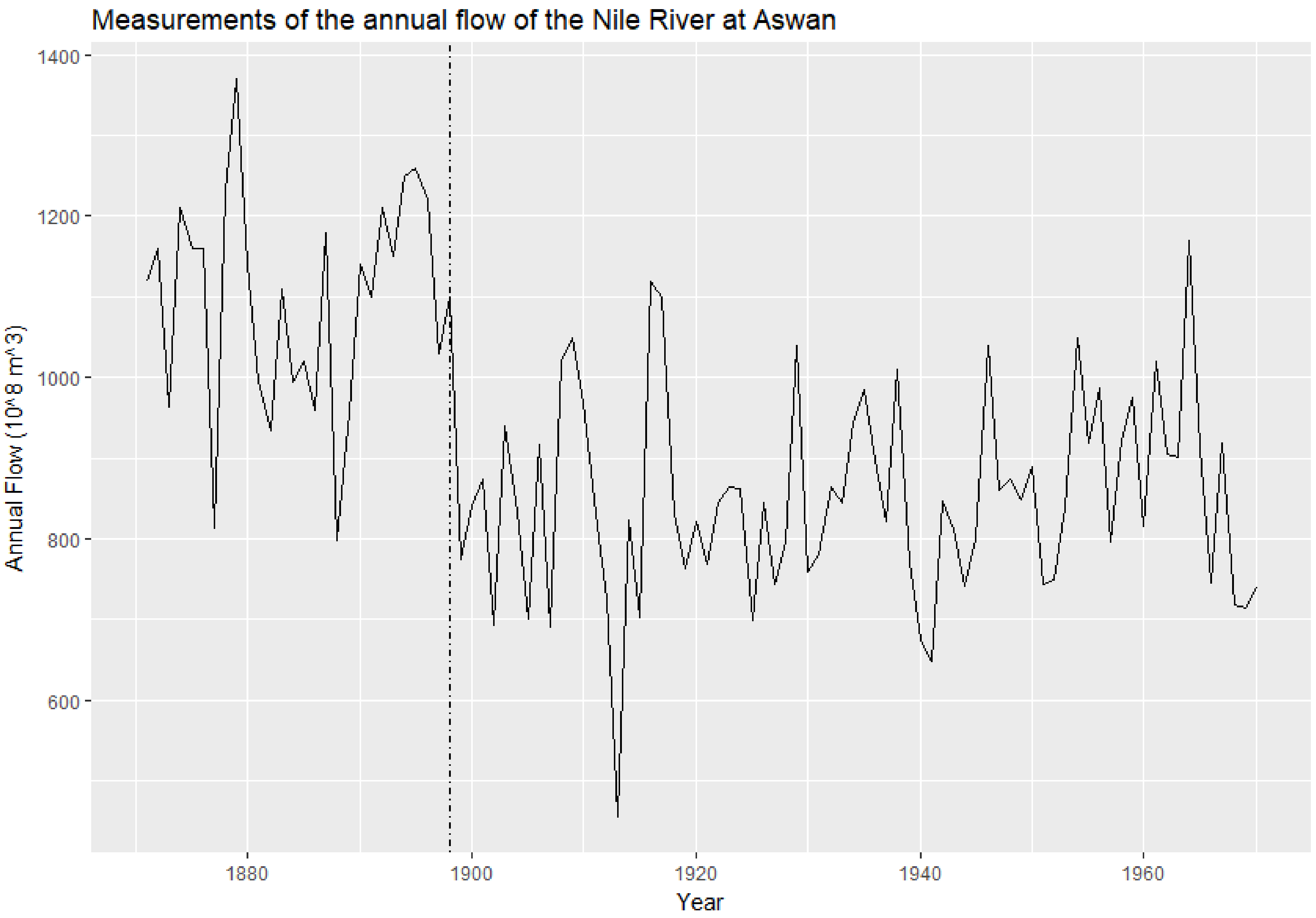
H.E. Hurst: British Hydrologist and "Father of the Nile"

- Spent decades measuring and studying the river
- Hurst noticed that annual streamflow was persistent (i.e. wet periods and dry periods)
- Heavy floods tended to be followed by heavier than average flood years and vice versa.

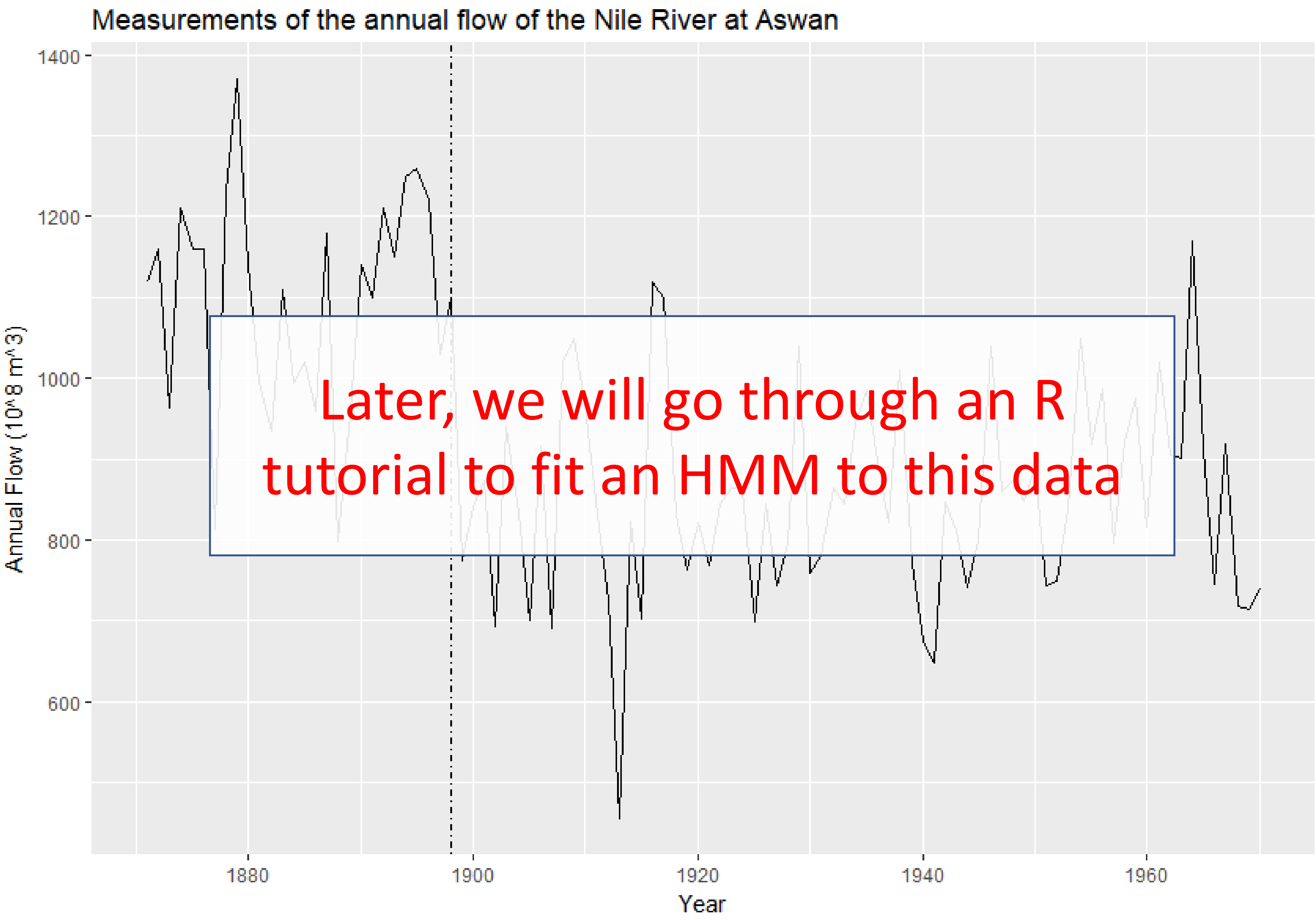
Motivation: persistence in Nile River flow

source: http://emerald.tufts.edu/alumni/magazine/spring2007/images/features/A4763_NS.jpg

- Hurst number which shows persistence in a given state
  - 0.5 – 1: time series with long-term positive autocorrelation
  - 0 – 0.5: time series with long-term switching between high and low values in adjacent pairs
- Hidden Markov models (HMM) can do a nice job of modeling persistence



Measurements of the annual flow of the Nile River at Aswan
Annual Flow (10^8 m^3)
1400
1200
1000
800
600
1880
1900
1920
1940
1960
Year


Measurements of the annual flow of the Nile River at Aswan
Annual Flow (10^8 m^3)
Year
1400
1200
1000
800
600
1880
1900
1920
1940
1960
Later, we will go through an R tutorial to fit an HMM to this data

Motivation: other applications

Google search PageRank: a way of measuring the importance of website pages.

International Journal of Computer Applications (0975 – 8887)
Volume 138 – No.9, March 2016

Google PageRank Algorithm: Markov Chain Model and Hidden Markov Model

source: http://www.ijcaonline.org/research/volume138/number9/rai-2016-ijca-908942.pdf

What is a Hidden Markov model?

A statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e. hidden) states. [source: Wikipedia]

What is a Markov process?

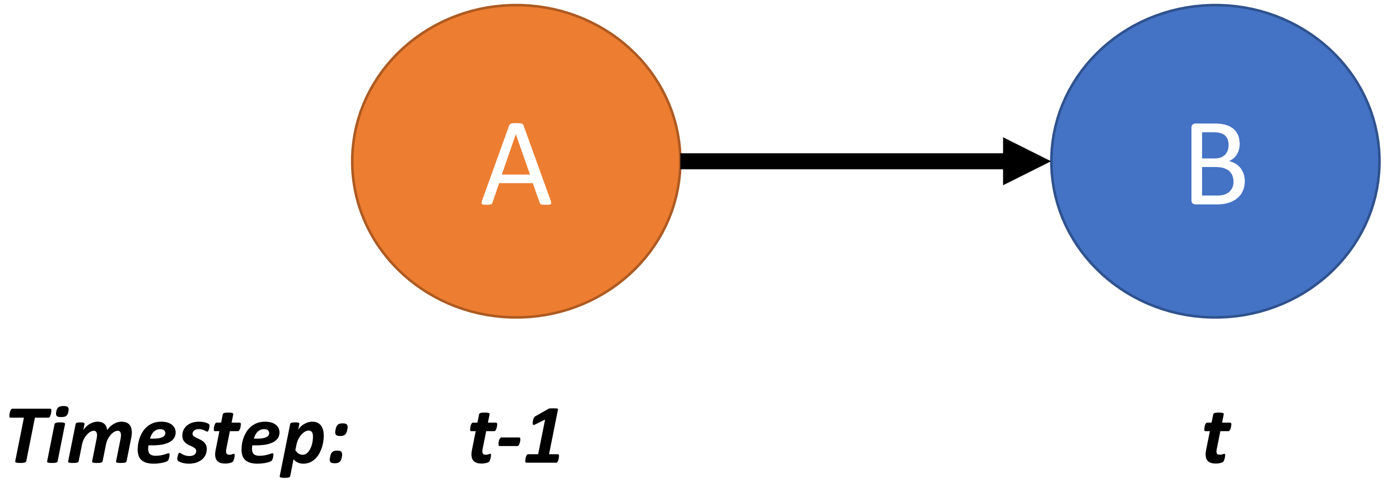
What is a Hidden Markov model?

*A statistical **Markov model** in which the system being modeled is assumed to be a **Markov process** with unobserved (i.e. hidden) states.* [source: Wikipedia]

What is a Markov process?

Markov Process

A phenomenon in which the outcome of a given event is only dependent on the previous timestep.

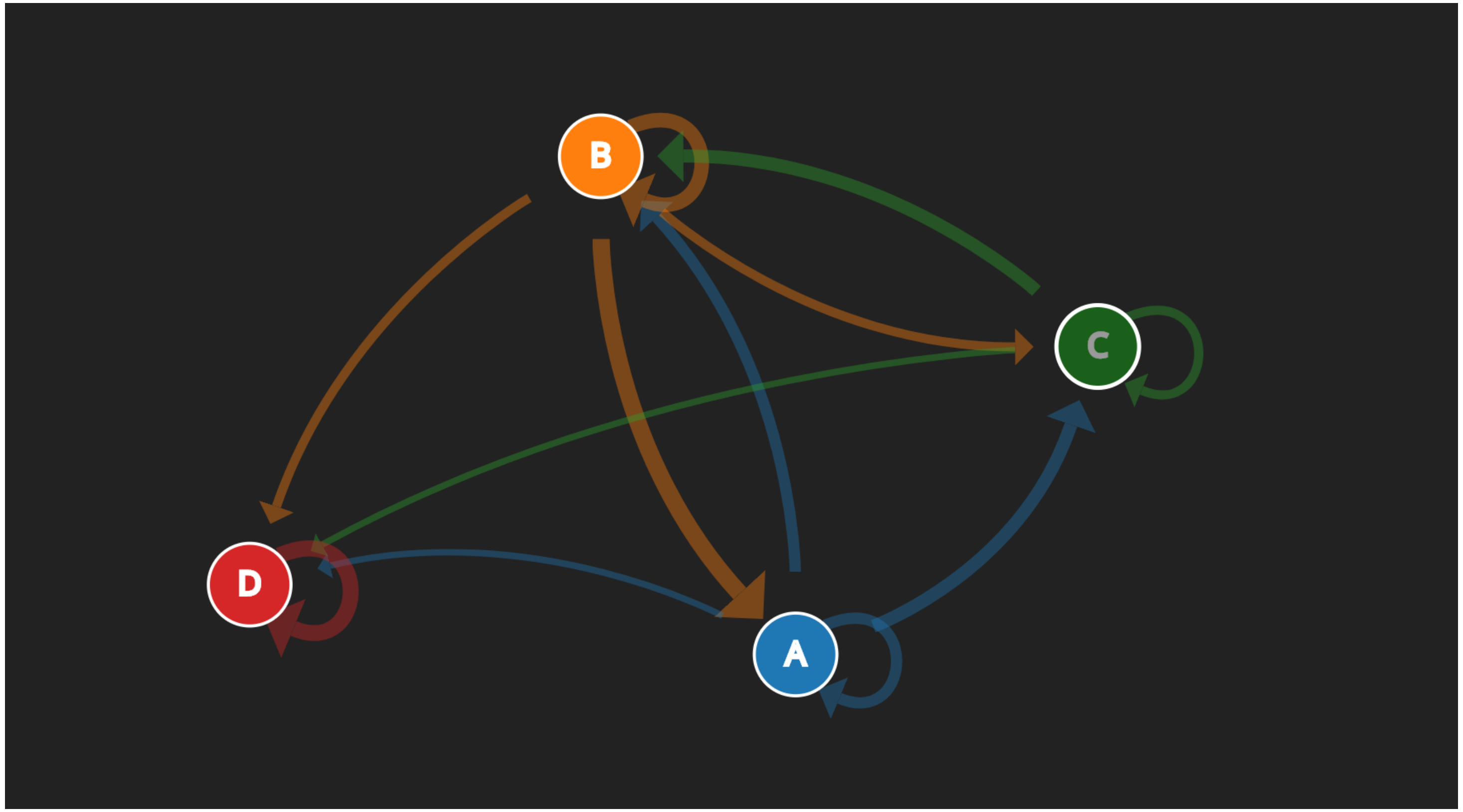


This is often represented using what are known as acyclic graphs.

Markov Chains

Markov chains, named after Andrey Markov, are mathematical systems that hop from one "state" (a situation or set of values) to another.

source: http://setosa.io/ev/markov-chains/



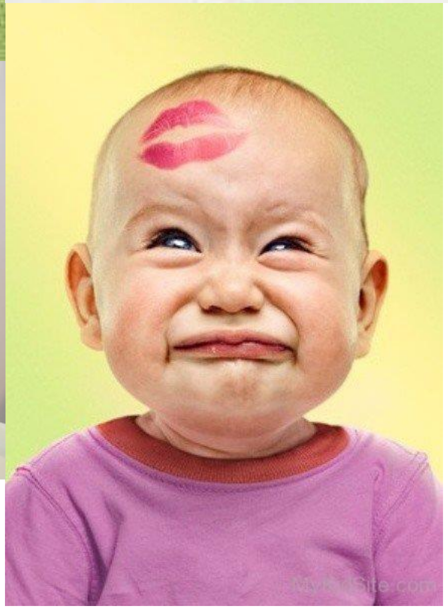
setosa.io/markov

For example, a baby could be in very different states…

<u>Baby state space</u>

- Laughing
- Sleeping
- Eating
- Crying

Each state has some relationship with the others.



Markov Chains: weather example

Imagine you have the following daily timeseries in which the weather on that day can be classified as rainy (R) or sunny (S):

RRRRRSSSSRRRRRRRRRRRRSSSSSSSSRRRRRRRRRRRSSSSSSSSSSSSRRRRR
RRRRRRRRSSSSSSSSSSSSSSSSSSSSSSSSSRRRRRRRRRRRRSSSSSSSSSSSSSS

This would represent a two-state Markov chain.

Markov Chains: weather example

You might notice that exactly half of these days are rainy and the other half are sunny:

RRRRRSSSSRRRRRRRRRRRRSSSSSSSSRRRRRRRRRRRSSSSSSSSSSSSRRRRR
RRRRRRRRSSSSSSSSSSSSSSSSSSSSSSSSSRRRRRRRRRRRRSSSSSSSSSSSSSS

This equates to a total probability of **P(R) = 0.5 and P(S) = 0.5**

Markov Chains: weather example

Now let's generate a new sequence of weather events. Since it's a 50-50 chance it'll be just like flipping a coin, right?

RSRSSRRSRRSSSRRSRSRSRRSRSSRRSRSRSRSSRRSRRSSSRRSRSRSRRSRSSR
RSRRSRSSRRSRRSSSRRSRSRSRRRSRSRRSRRSRSSRRSRRSSSRRRSRSRRSRS

Markov Chains: weather example

How'd we do? **Not well…**

Observed record:

RRRRRSSSSRRRRRRRRRRRRSSSSSSSSRRRRRRRRRRRSSSSSSSSSSSSRRRRR
RRRRRRRRSSSSSSSSSSSSSSSSSSSSSSSSSRRRRRRRRRRRRSSSSSSSSSSSSSS

Our model:

RSRSSRRSRRSSSRRSRSRSRRSRSSRRSRSRSRSSRRSRRSSSRRSRSRSRRSRSSR
RSRRSRSSRRSRRSSSRRSRSRSRRRSRSRRSRRSRSSRRSRRSSSRRRSRSRRSRS

Markov Chains: weather example

We looked at the *total* probability of each event occurring but we haven't considered the conditional probabilities. Let's say the data looked like this for n=80 observed samples:

Count	Result	
Condition	R_t	S_t
R_{t-1}	36	8
S_{t-1}	4	32

$$Total\ Probability: P(A) = \frac{\#\ outcomes\ of\ A}{total\ outcomes}$$

$$P(R_t) = \frac{36+4}{80} = \frac{1}{2}$$

$$P(S_t) = \frac{32+8}{80} = \frac{1}{2}$$

Markov Chains: weather example

We looked at the *total* probability of each event occurring but we haven't considered the conditional probabilities. Let's say the data looked like this for n=80 observed samples:

Count	Result	
Condition	R_t	S_t
R_{t-1}	36	8
S_{t-1}	4	32

$$Total\ Probability: P(A) = \frac{\#\ outcomes\ of\ A}{total\ outcomes}$$

$$P(R_t) = \frac{36+4}{80} = \frac{1}{2}$$

$$P(S_t) = \frac{32+8}{80} = \frac{1}{2}$$

Probability review: what are the relevant conditional probabilities for this problem?

$P(R_t | R_{t-1}) = ?$

$P(S_t | R_{t-1}) = ?$

$P(R_t | S_{t-1}) = ?$

$P(S_t | S_{t-1}) = ?$

Markov Chains: weather example

We looked at the *total* probability of each event occurring but we haven't considered the conditional probabilities. Let's say the data looked like this for n=80 observed samples:

Count	Result	
Condition	R_t	S_t
R_{t-1}	36	8
S_{t-1}	4	32

$$Intersection\ (A \cap B) = \frac{all\ outcomes\ resulting\ in\ A\ and\ B}{total\ outcomes}$$

$$P(R_t \cap R_{t-1}) = 36/80 = 0.45$$

$$P(R_t \cap S_{t-1}) = 4/80 = 0.05$$

$$P(S_t \cap R_{t-1}) = 8/80 = 0.10$$

$$P(S_t \cap S_{t-1}) = 32/80 = 0.40$$

Markov Chains: weather example

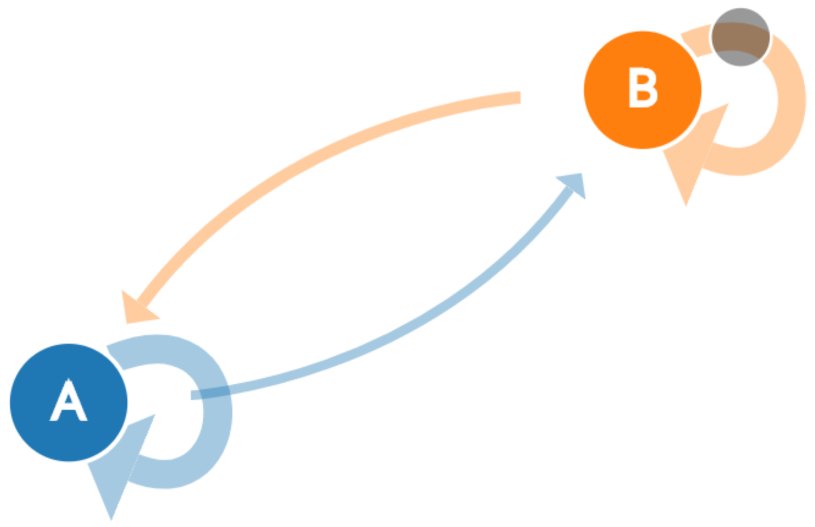
We looked at the *total* probability of each event occurring but we haven't considered the conditional probabilities. Let's say the data looked like this for n=80 observed samples:

$$Conditional\ Probability\ P(B|A) = P(A \cap B)/P(A)$$

Probability	Result	
Condition	R_t	S_t
R_{t-1}	$P(R_t\|R_{t-1}) = 0.45/0.5 = 0.90$	$P(S_t\|R_{t-1}) = 0.10/0.5 = 0.20$
S_{t-1}	$P(R_t\|S_{t-1}) = 0.05/0.5 = 0.10$	$P(S_t\|S_{t-1}) = 0.40/0.5 = 0.80$

Markov Chains

The thickness of the lines represents the conditional probability.

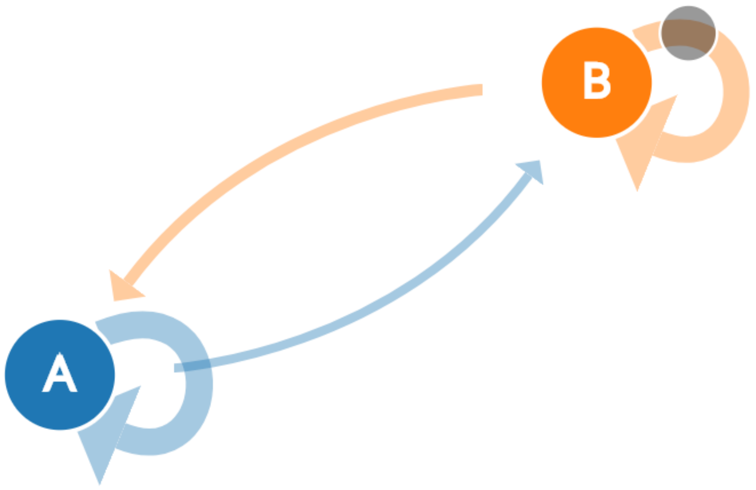


For Markov chains, this is known as the **transition matrix:**

Probability	Result	
Condition	R_t	S_t
R_{t-1}	$P(R_t\|R_{t-1}) = 0.45/0.5 = 0.90$	$P(S_t\|R_{t-1}) = 0.10/0.5 = 0.20$
S_{t-1}	$P(R_t\|S_{t-1}) = 0.05/0.5 = 0.10$	$P(S_t\|S_{t-1}) = 0.40/0.5 = 0.80$

Markov Chains

This is all shown interactively at the following website:
http://setosa.io/ev/markov-chains/



	A	B
A	P(A|A): 0.90	P(B|A): 0.10
B	P(A|B): 0.16	P(B|B): 0.84

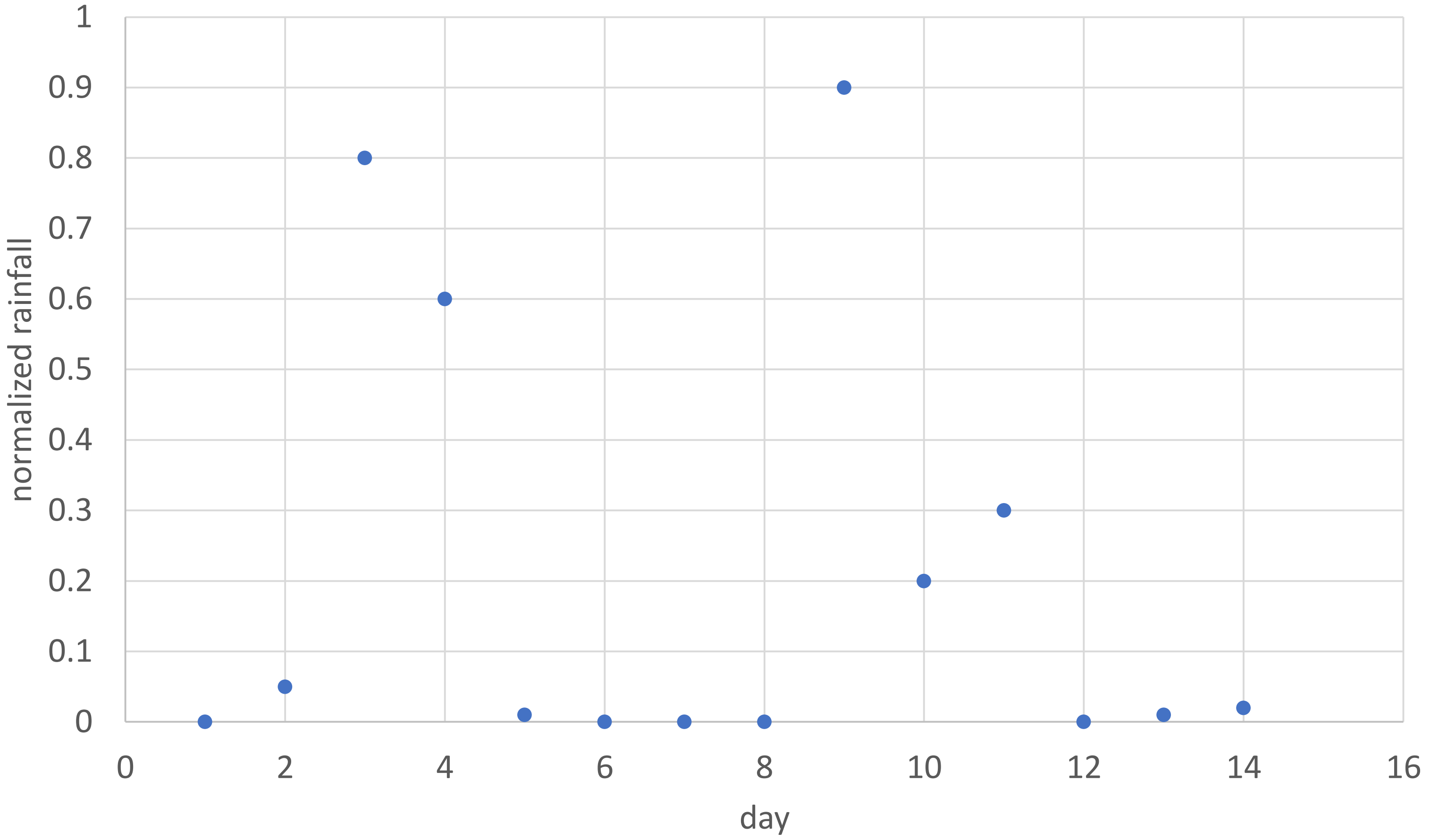
What is a hidden Markov model?

Wikipedia: a statistical Markov model in which the system being modeled is assumed to be a Markov process with *unobserved (i.e. hidden) states.*

Weather example: where the states were the observed data (this is a Markov model)

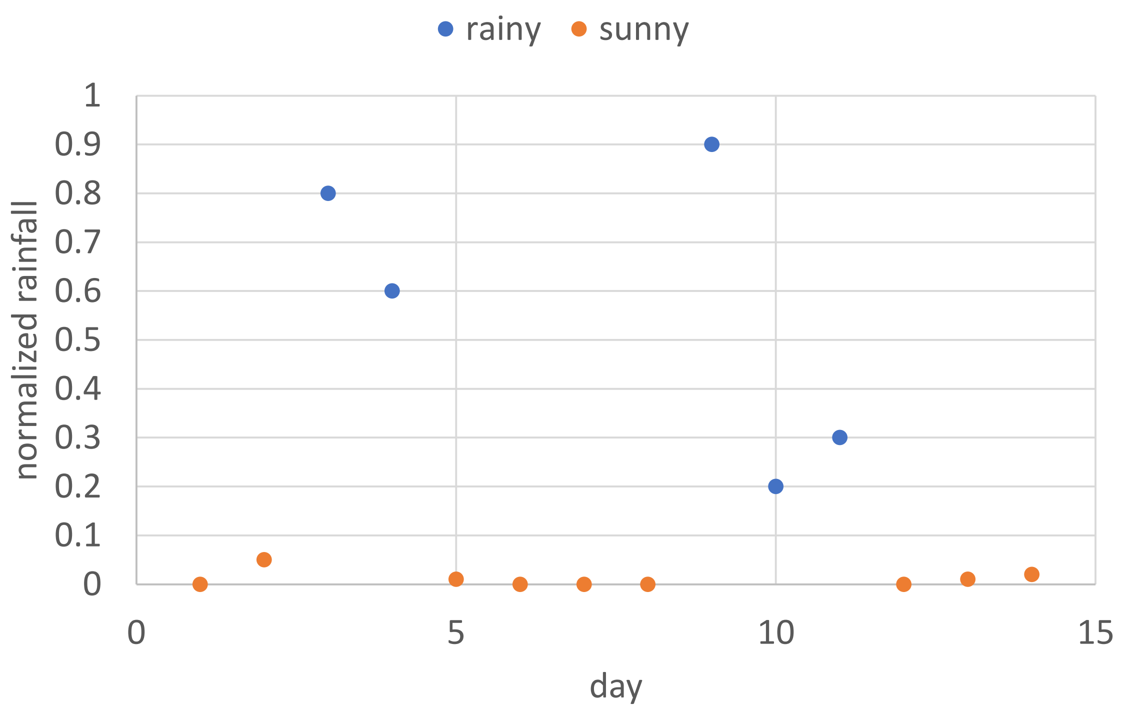
What if we want to predict how much precipitation is going to occur rather than whether it will by rainy or sunny?

How many states are present in this data?



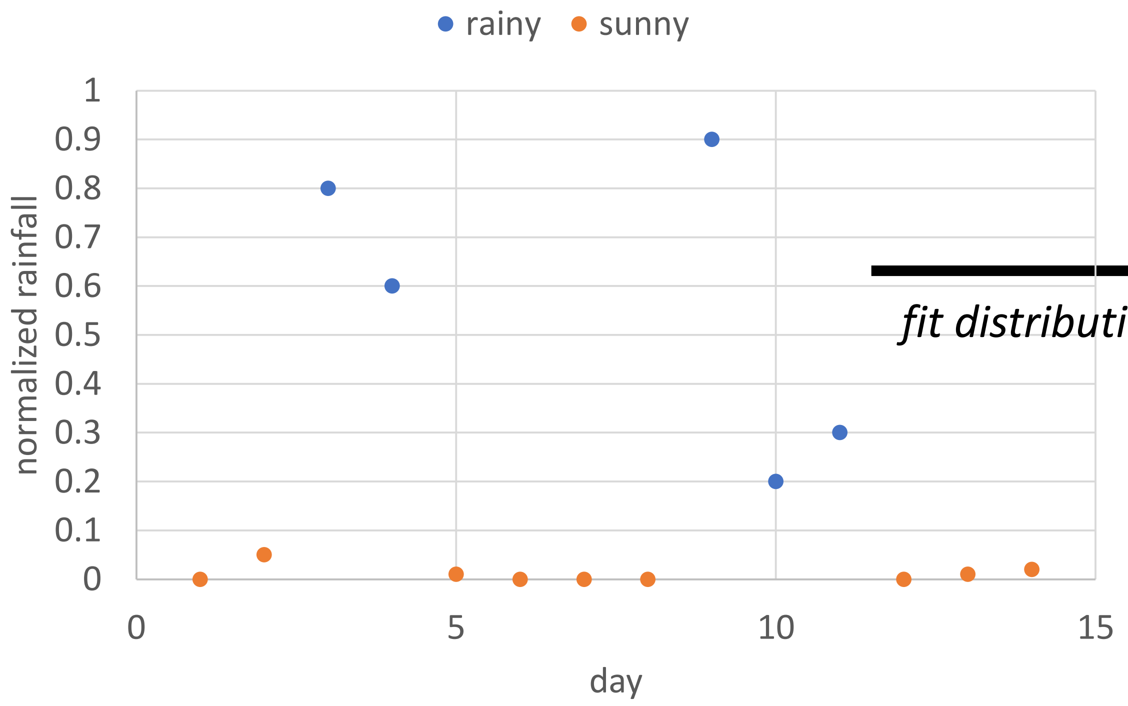
HMM: 2-state precipitation example

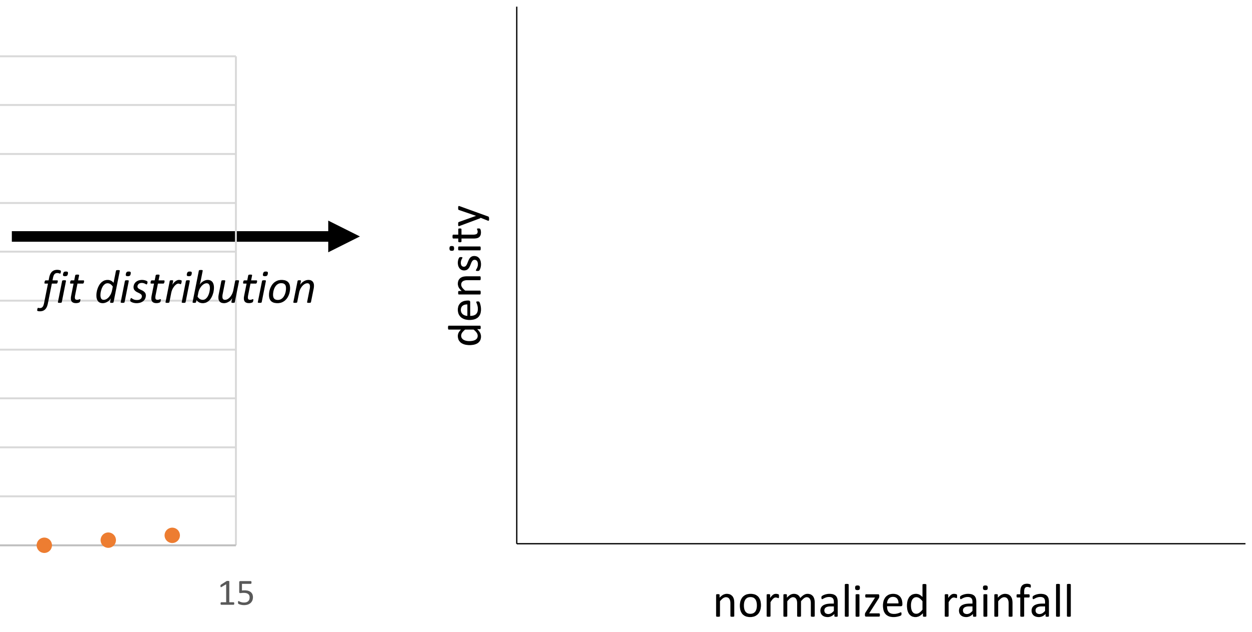
NEW weather example: where we want to predict precipitation and we notice that there seem to be different states.



HMM: 2-state precipitation example

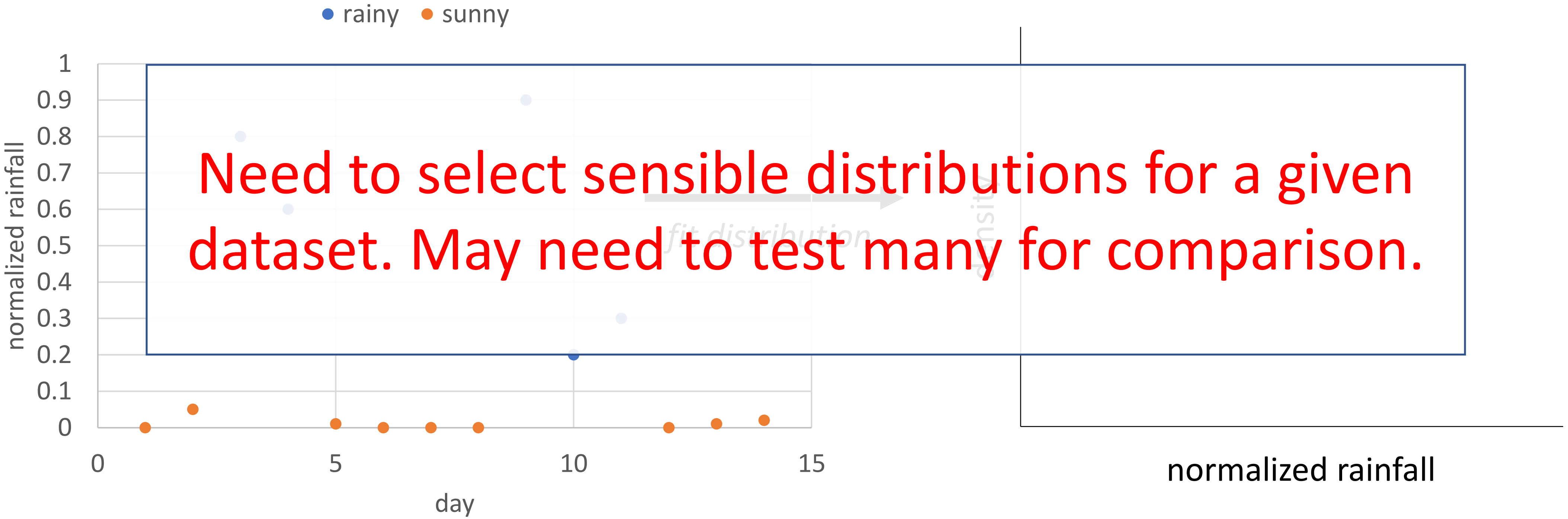
NEW weather example: where we want to predict precipitation and we notice that there seem to be different states.





HMM: 2-state precipitation example

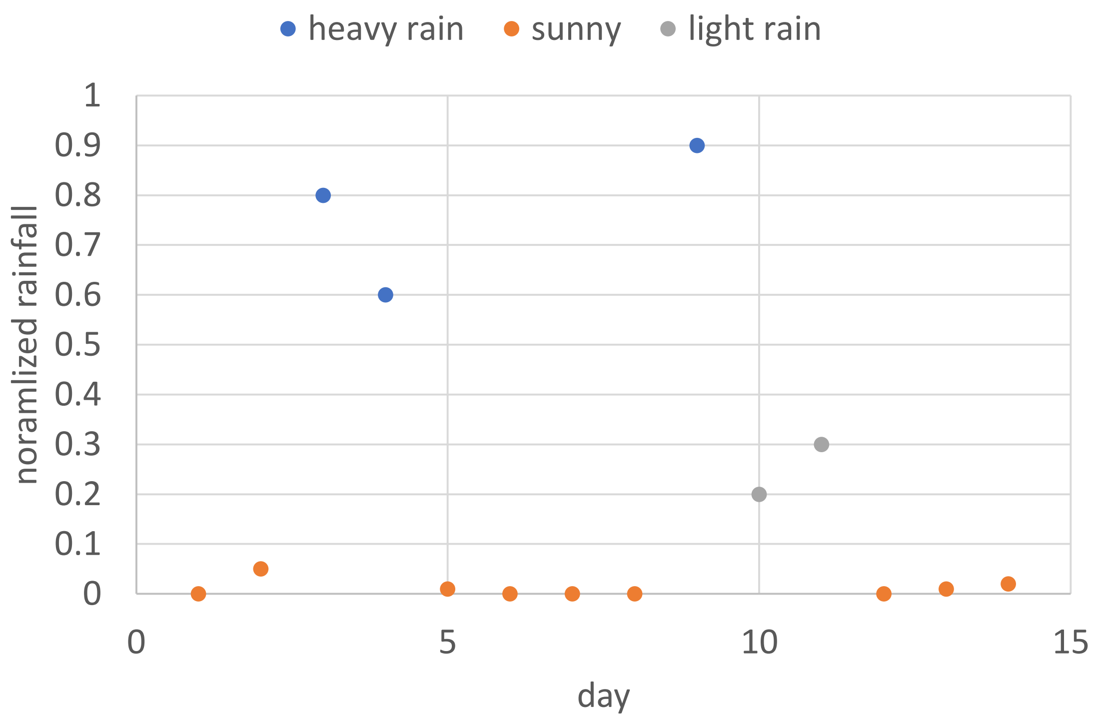
NEW weather example: where we want to predict precipitation and we notice that there seem to be different states.



Need to select sensible distributions for a given dataset. May need to test many for comparison.

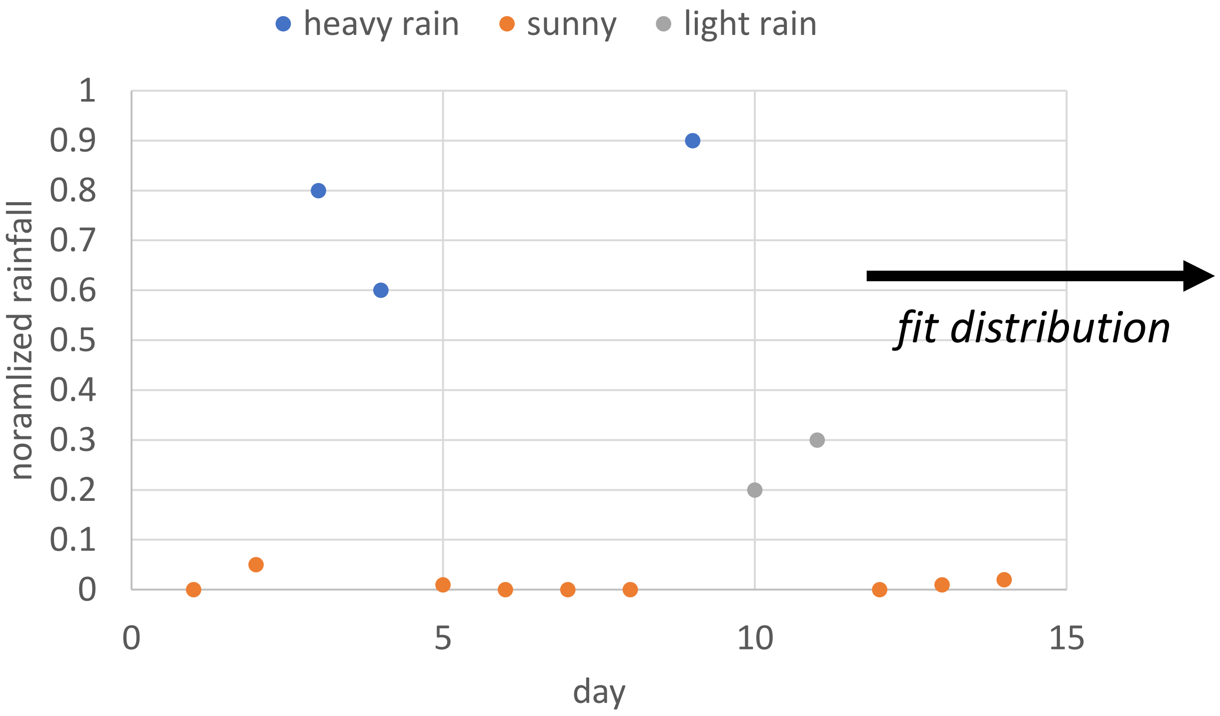
HMM: 3-state precipitation example

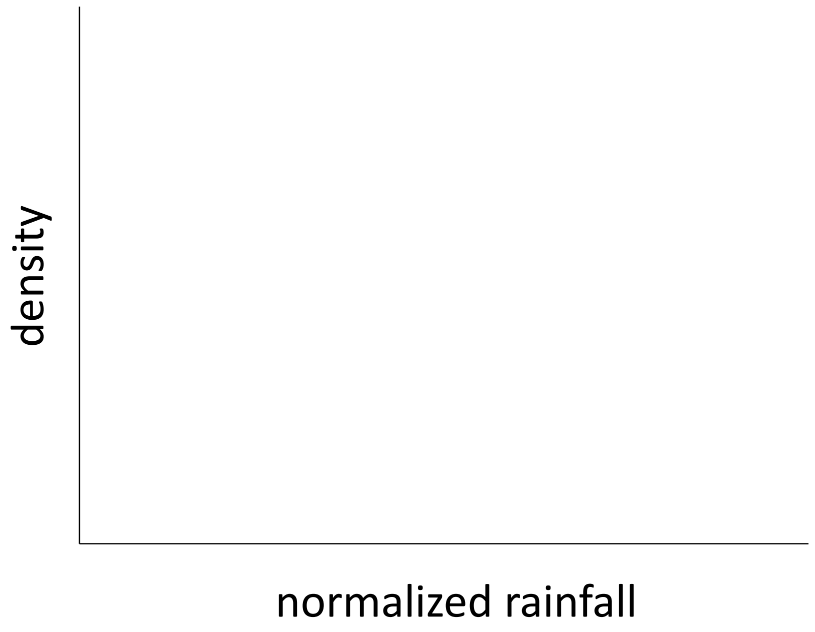
Why do we think it is two states? Could there be a third?



HMM: 3-state precipitation example

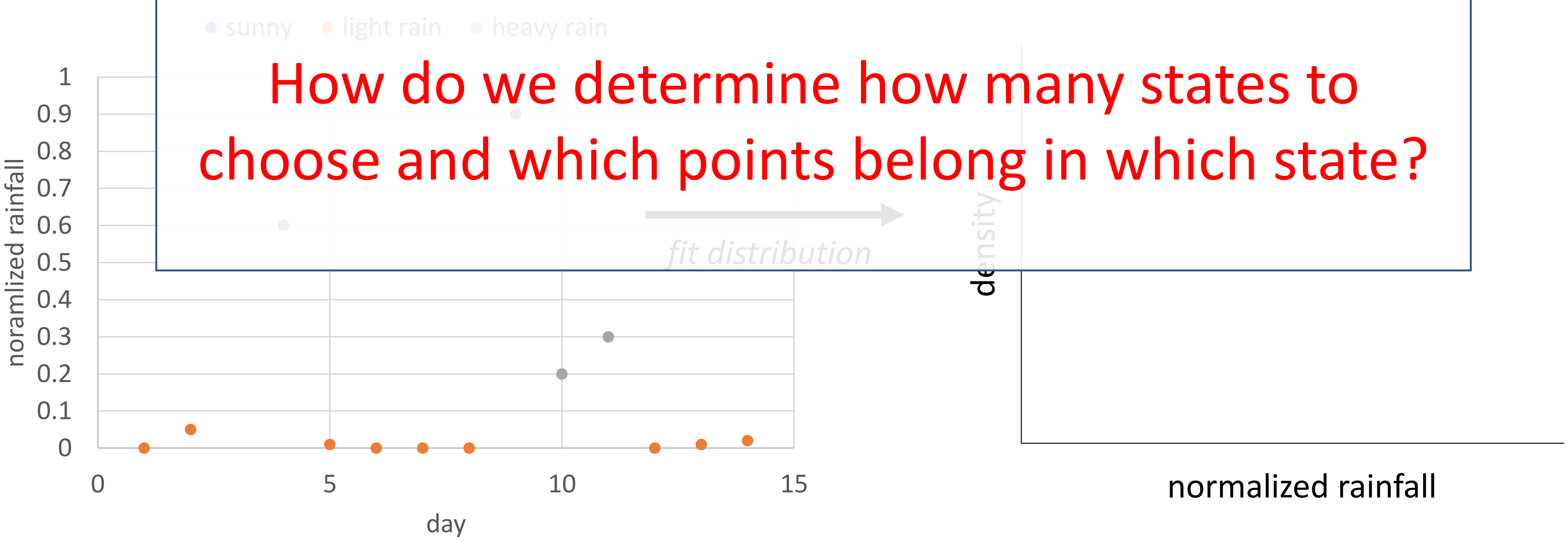
Why do we think it is two states? Could there be a third?





HMM: 3-state precipitation example

Why do we think it is two states? Could there be a third?



How do we determine how many states to choose and which points belong in which state?

Important HMM algorithms

- Baum-Welch (aka the Forward-Backward algorithm): automatically estimates parameters of HMM

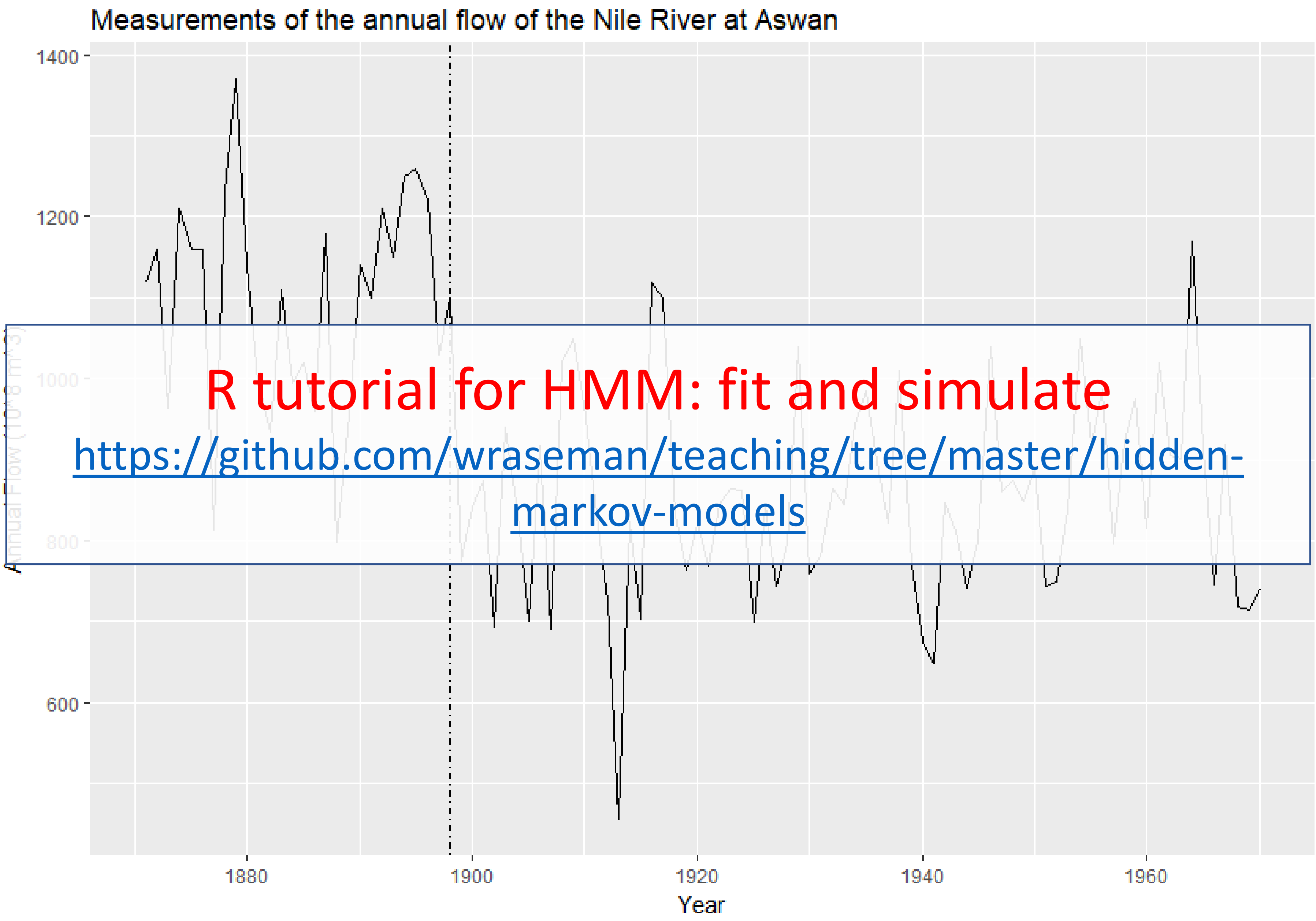
```
HiddenMarkov package function: BaumWelch()
```

- Viterbi: predicts the most likely sequence of Markov states given the observed dataset

```
HiddenMarkov package function: Viterbi()
```

R tutorial for HMM: fit and simulate

https://github.com/wraseman/teaching/tree/master/hidden-markov-models



Problem I noticed with Nile River example…

```
> fit.hmm$Pi
                 [,1]       [,2]
[1,] 9.640019e-01 0.03599814
[2,] 1.841490e-08 0.99999998
```

Pi is the name for the transition matrix.

What's potentially unrealistic about this model?

Concluding remarks…

Be able to answer these questions!

- What is a Markov chain?
- What is the difference between a Markov model and a Hidden Markov model (HMM)?
- What is the purpose of the following algorithms: Viterbi and Baum-Welch?
- When should you try to apply HMMs?