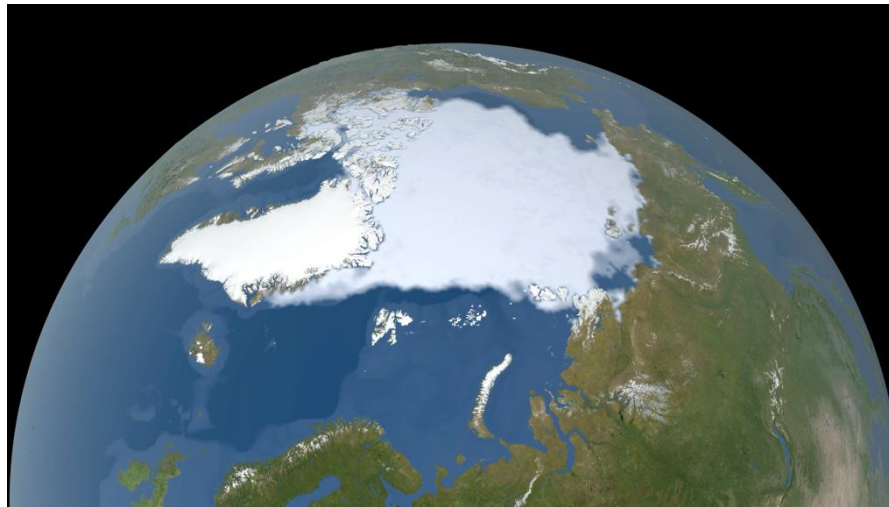
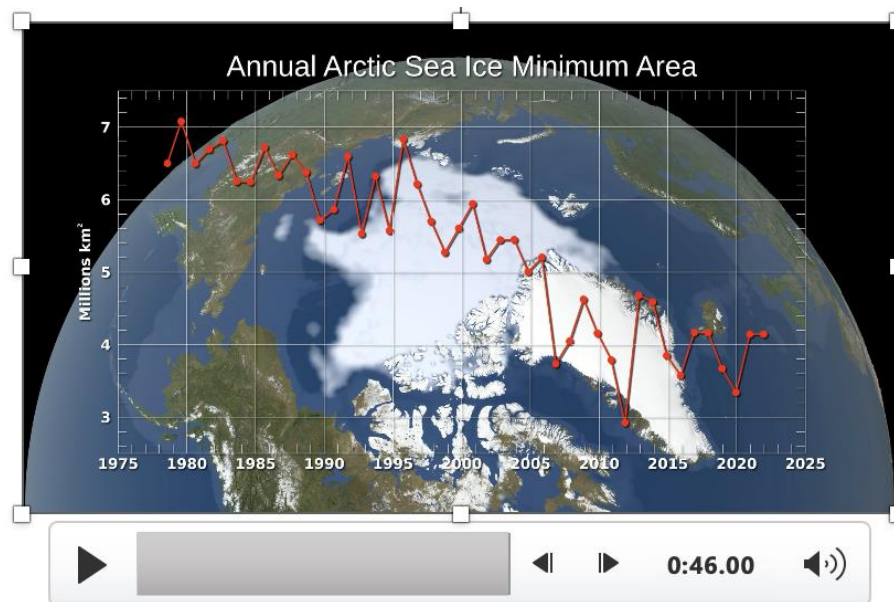


# Finding Structure in Randomness



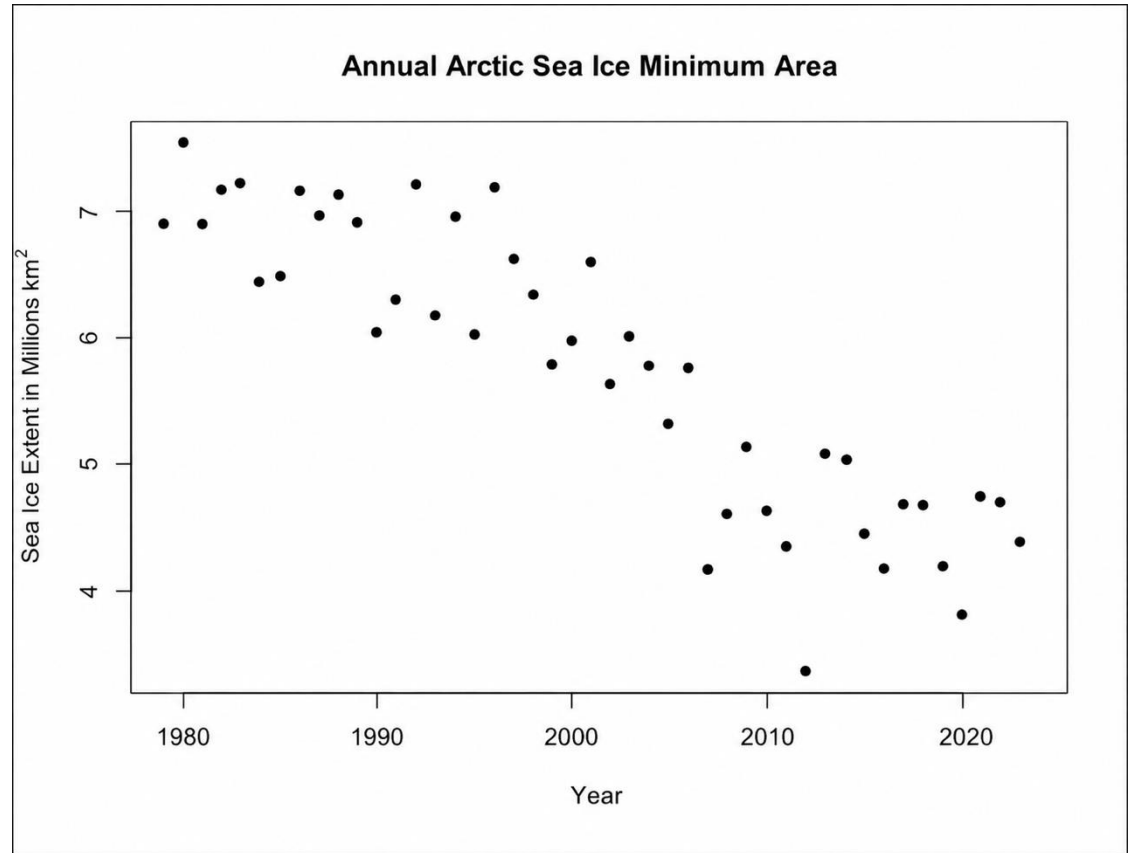
Visualisation by Trent L. Schindler (USRA), NASA/Goddard Space Flight Center Scientific Visualization Studio, The Blue Marble data is courtesy of Reto Stockli (NASA/GSFC).

# Finding Structure in Randomness



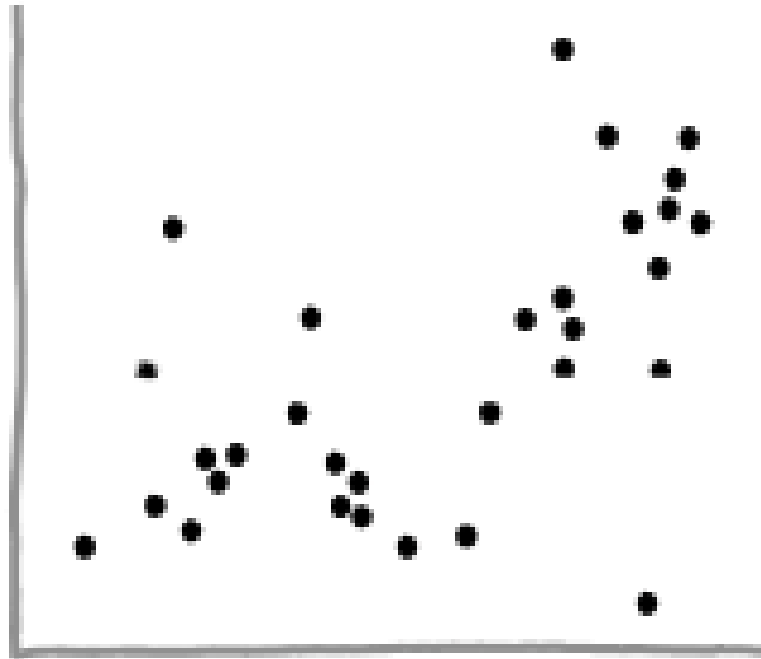
Visualisation by Trent L. Schindler (USRA), NASA/Goddard Space Flight Center Scientific Visualization Studio, The Blue Marble data is courtesy of Reto Stockli (NASA/GSFC).

# Task: Fit a curve to Arctic Sea Ice data

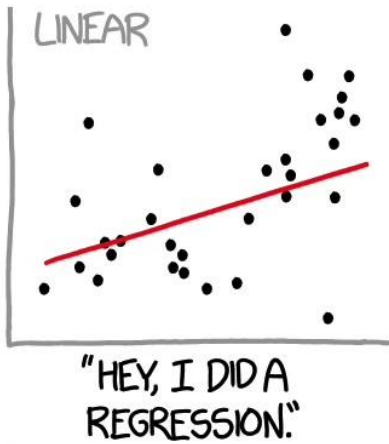
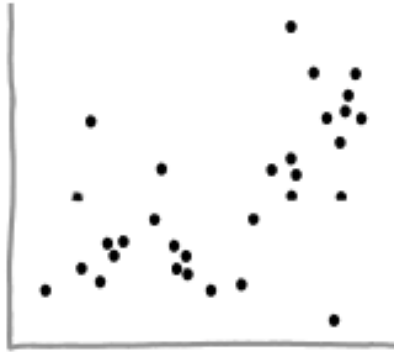


# General question: How do we fit a curve to data?

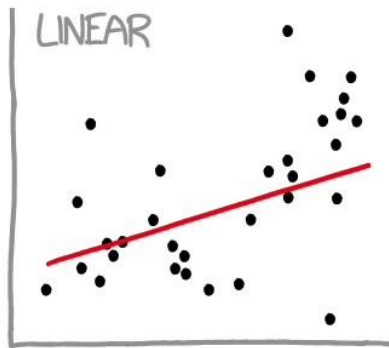
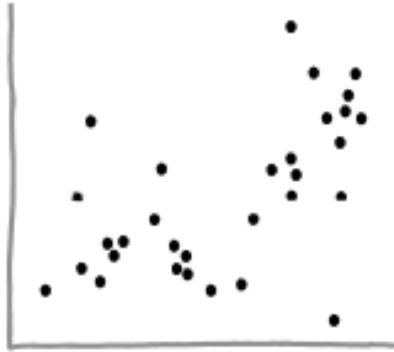
A simple example...



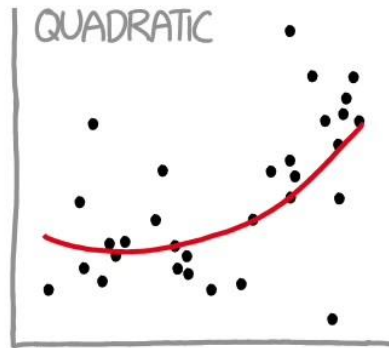
# How do we fit a curve to data? Linear...



# How do we fit a curve to data? ... Quadratic...

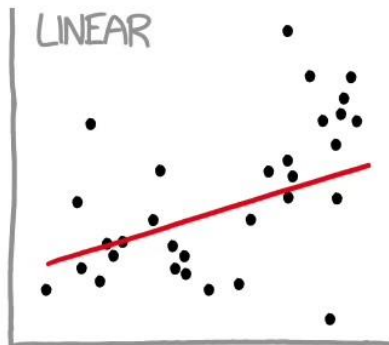
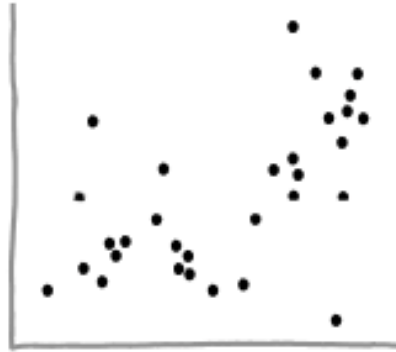


"HEY, I DID A  
REGRESSION."

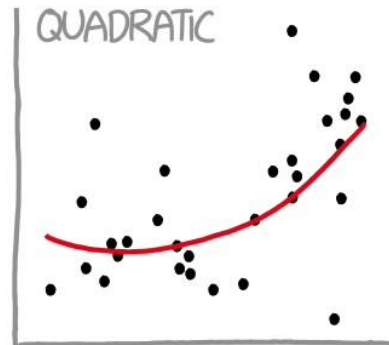


"I WANTED A CURVED  
LINE, SO I MADE ONE  
WITH MATH."

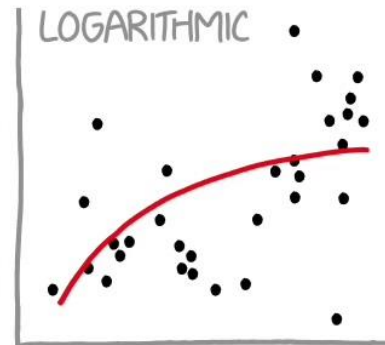
# How do we fit a curve to data? ... Logarithmic...



"HEY, I DID A  
REGRESSION."

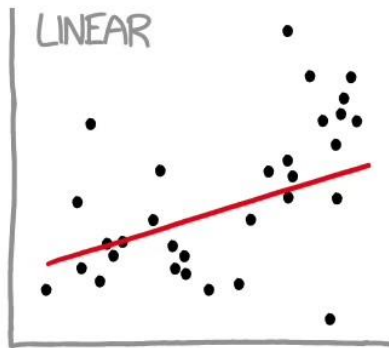
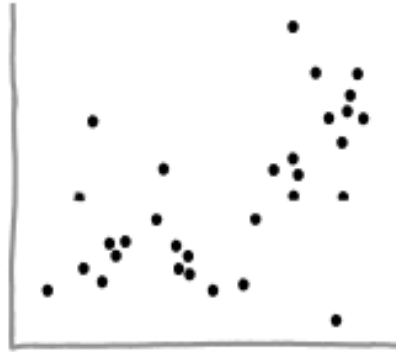


"I WANTED A CURVED  
LINE, SO I MADE ONE  
WITH MATH."

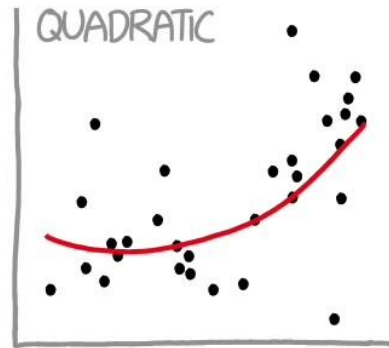


"LOOK, IT'S  
TAPERING OFF!"

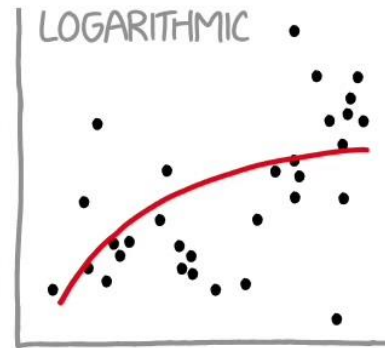
# How do we fit a curve to data? ... Exponential...



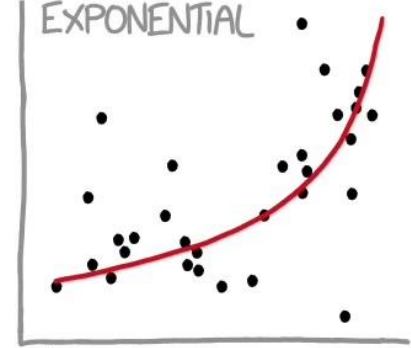
"HEY, I DID A  
REGRESSION."



"I WANTED A CURVED  
LINE, SO I MADE ONE  
WITH MATH."

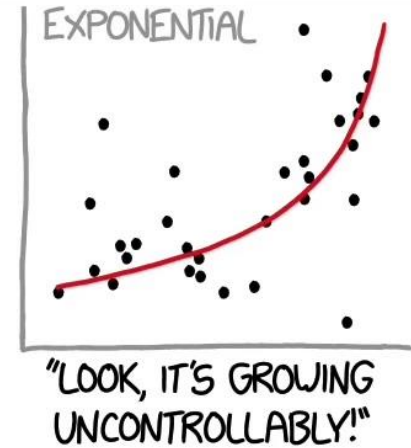
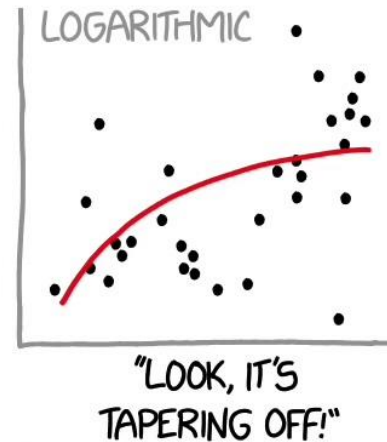
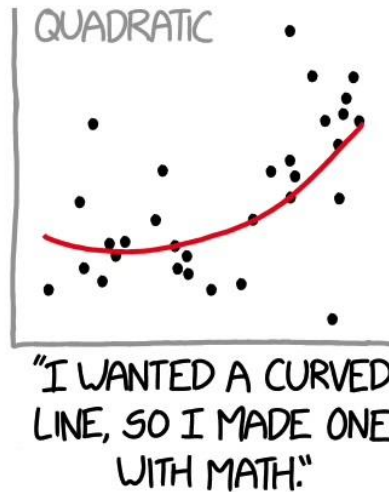
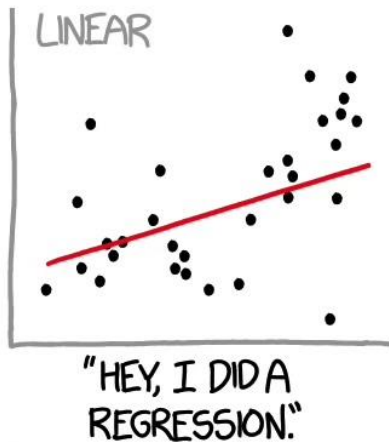
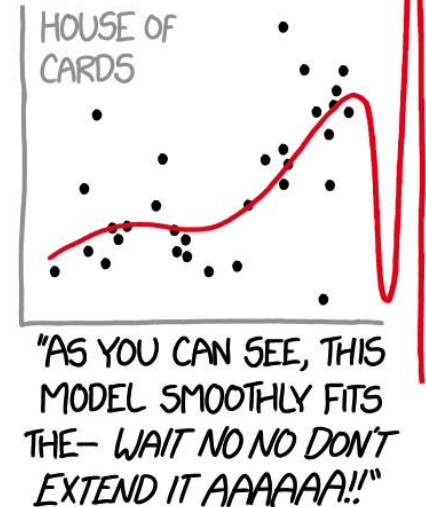
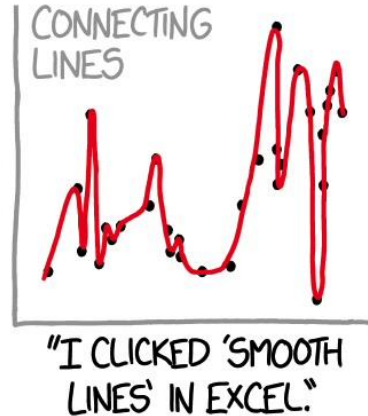
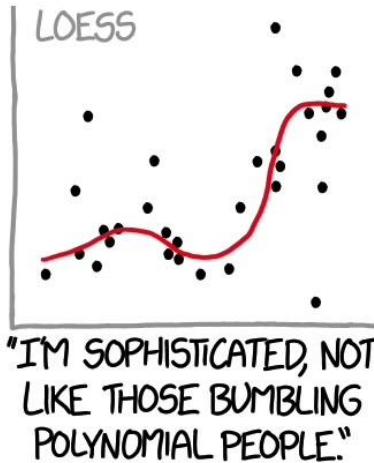


"LOOK, IT'S  
TAPERING OFF!"

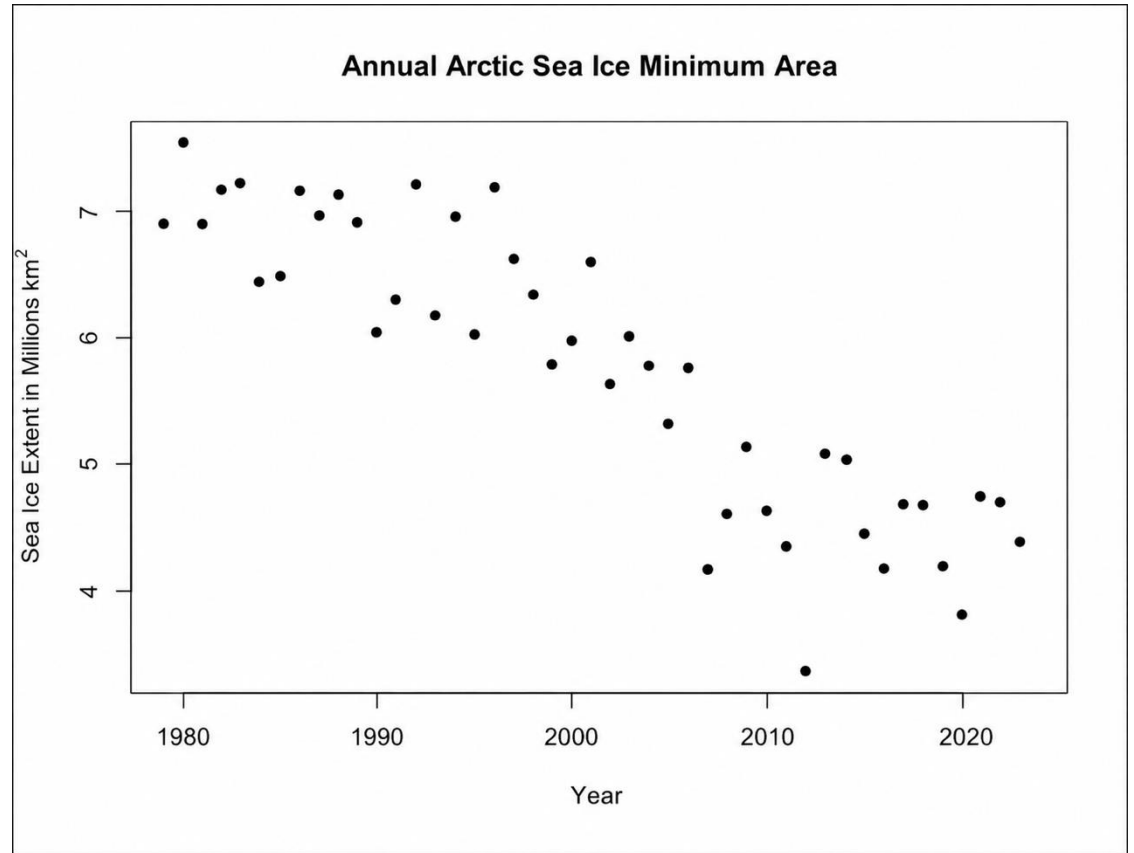


"LOOK, IT'S GROWING  
UNCONTROLLABLY!"

# Many ways to fit a curve to data... But which one is best?

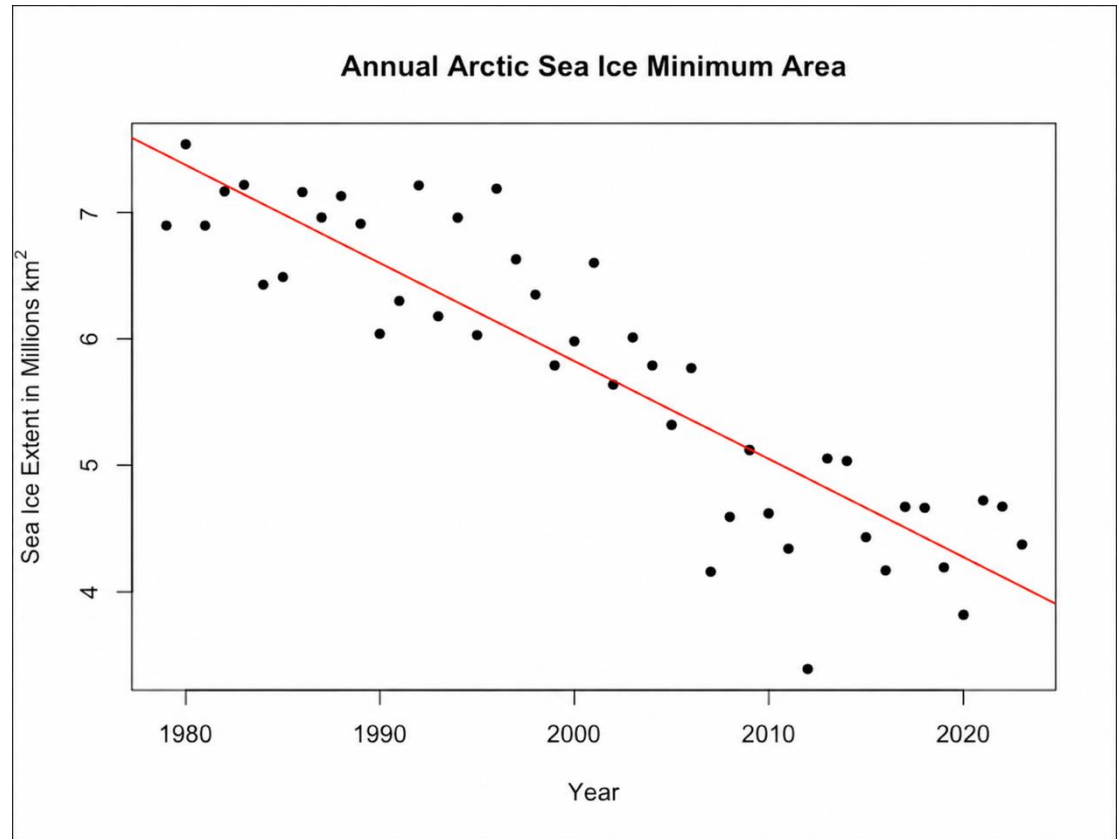


# Task: Fit a curve to the Arctic Sea Ice data



# Fit a curve to the Arctic Sea Ice data: Attempt 1

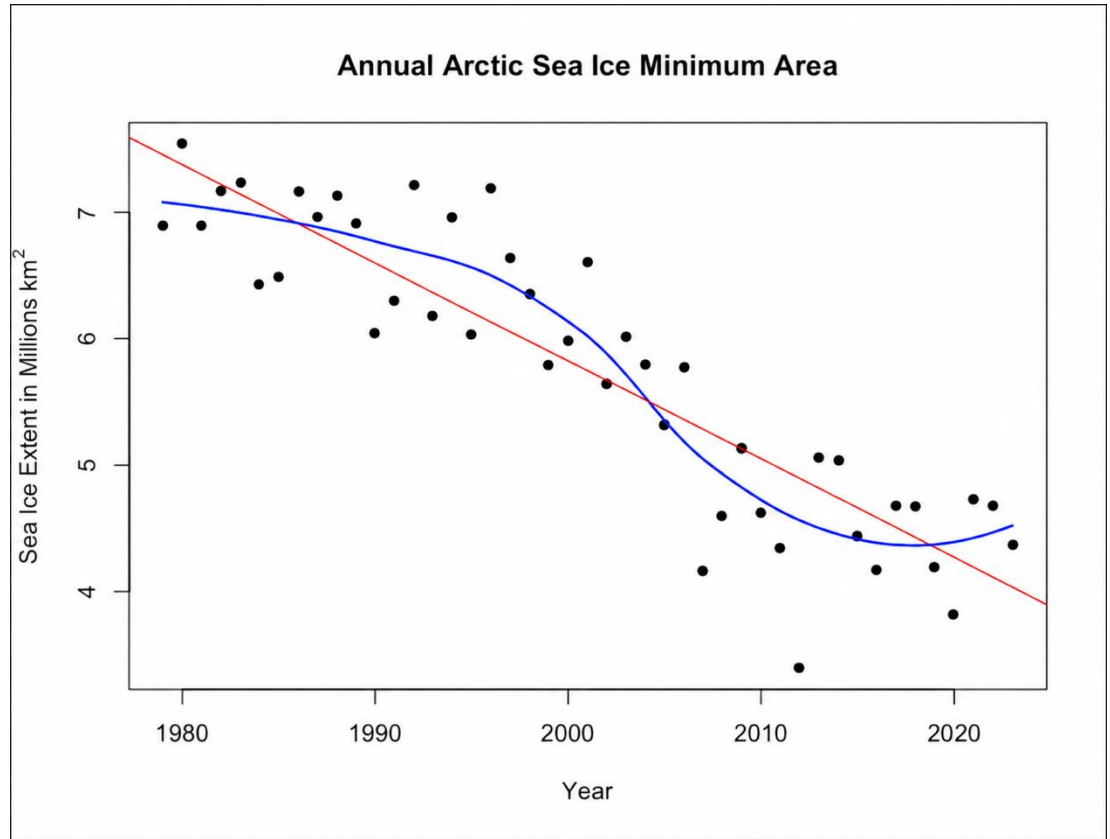
— Linear fit



# Fit a curve to the Arctic Sea Ice data: Attempt 2

- Linear fit
- Quadratic fit

Which *model* is best?



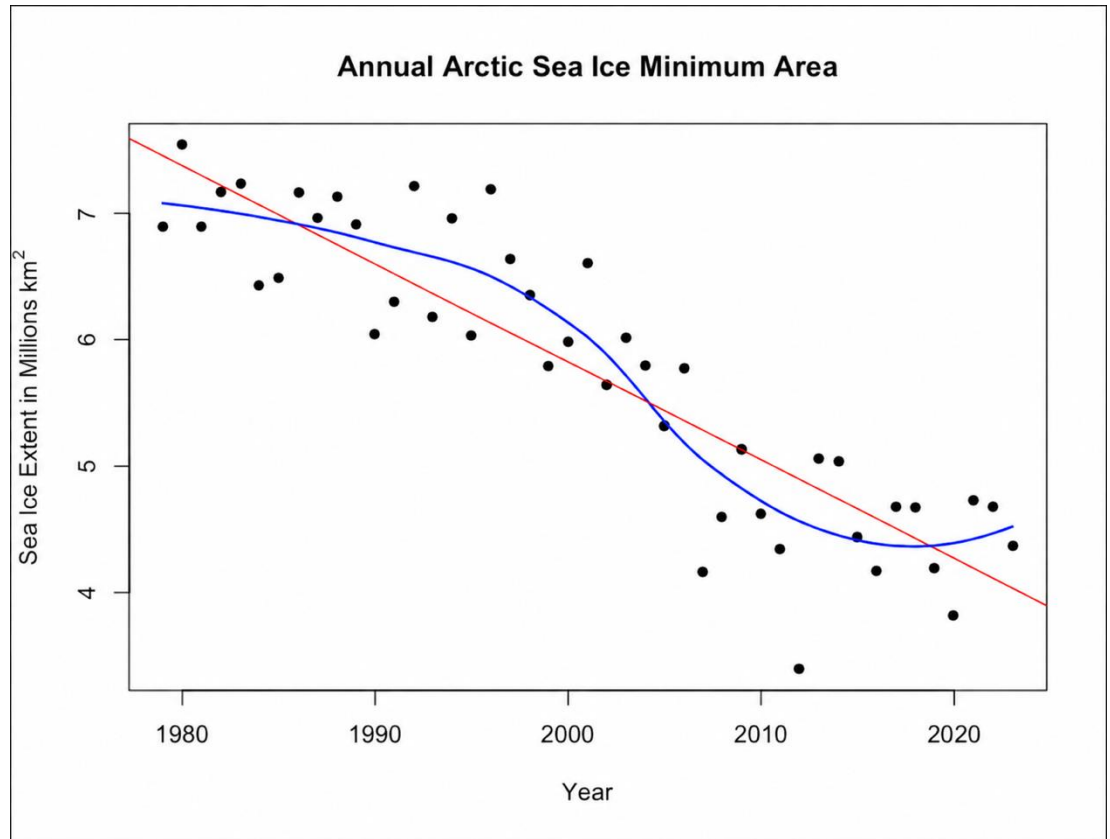
# Fit a curve to the Arctic Sea Ice data: Attempt 2

- Linear fit
- Quadratic fit

Which *model* is best?

Model captures the *explainable* aspects of the relationship.

The deviations should be *random*.



# What is randomness?

## Definitions:

- *Lawlessness*: absence of a rule
- *Maximal difficulty of prediction*: no available prediction method does better than chance



# What is randomness?

## Definitions:

- *Lawlessness*: absence of a rule
- *Maximal difficulty of prediction*: no available prediction method does better than chance

## Examples: sequences of coin tosses, rolling dice, spinner

In theory, knowing all initial and environmental conditions coin tosses would follow the rules of mechanics and be predictable, but in practice this is not feasible, hence they are random (our epistemic position).



# Can we distinguish “true” from “fake” randomness?

Which of the two sequences was produced by tossing a coin 100 times and which by someone pretending to do so?

## Sequence #1:

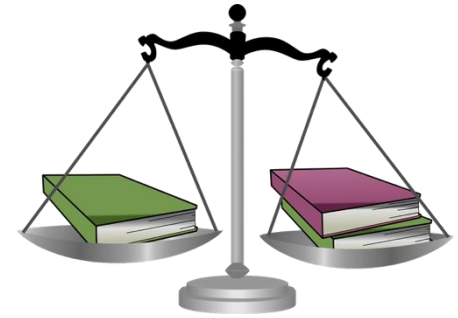
T H T T H H T H T H T T H T H T T H H H T H H T T  
H T H H H T H T T H T T T H T T H H T H H H T T T  
H T H H T H T H H T T T H T T T H T T H T H T T T  
H H T H H T T H T T T H H T T H T T H T H T T H H

## Sequence #2:

T H H T H H H T T H H H T H H H H T H T H T T H T T  
T H H H H H T H H H H T H H T H T T T H T T T H H  
H H T T H H T H T H T H T T T H H H T T H T H H  
H T T H H H T H H H T T H T H H T T H H H T



# A statistical court room



## Null hypothesis:

**The sequence stems from genuine coin flips.**

Defense position: assumption of innocence.

## Alternative hypothesis:

**The sequence does not stem from genuine coin flips.**

Prosecutors position: evidence of guilt.

**We reject the null hypothesis if there is *sufficient evidence against it*.**

# What characteristics might we use to establish “guilt”?

Measurable evidence against the Null?

## Sequence #1:

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | H | T | T | H | H | T | H | T | H | T | T | H | T | H | T | T | H | H | H | T | H | H | T | T |
| H | T | H | H | H | T | H | T | T | H | T | T | T | H | T | T | H | H | T | H | H | H | T | T | T |
| H | T | H | H | T | H | T | H | H | T | T | T | H | T | T | T | H | T | T | H | T | H | T | T | T |
| H | H | T | H | H | T | T | H | T | T | T | H | H | T | T | H | T | T | H | T | H | T | T | H | H |



# What characteristics might we use to establish “guilt”?

Measurable evidence against the Null, e.g.:

- ▶ Number of heads
- ▶ Number of switches
- ▶ Number of runs
- ▶ Longest run

## Sequence #1:

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | H | T | T | H | H | T | H | T | H | T | T | H | T | H | T | T | H | H | H | T | H | H | T | T |
| H | T | H | H | H | T | H | T | T | H | T | T | T | H | T | T | H | H | T | H | H | H | T | T | T |
| H | T | H | H | T | H | T | H | H | T | T | T | H | T | T | T | H | T | T | H | T | H | T | T | T |
| H | H | T | H | H | T | T | H | T | T | T | H | H | T | T | H | T | T | H | T | H | T | T | H | H |



# Number of heads

*Assuming the sequence is a genuine (coin flips),*  
how many heads do we expect to see on average?

## A mathematical model:

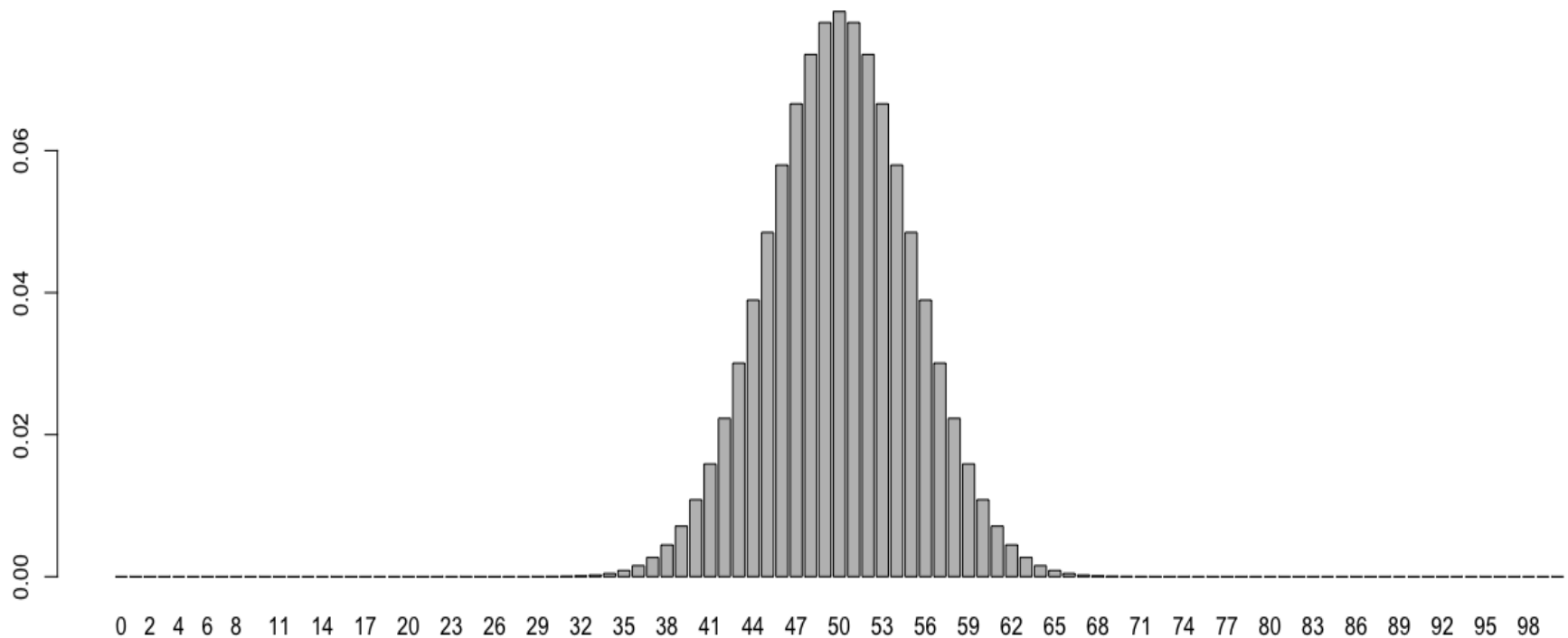
- 100 independent coin flips: Number of trials  $N = 100$ .
- In each coin flip, the probability of heads occurring is  $\frac{1}{2}$ .  
Success probability  $p = \frac{1}{2}$ .
- Number of heads  $X$  has a Binomial distribution:  
 $X \sim \text{Bin}(100, \frac{1}{2})$ .



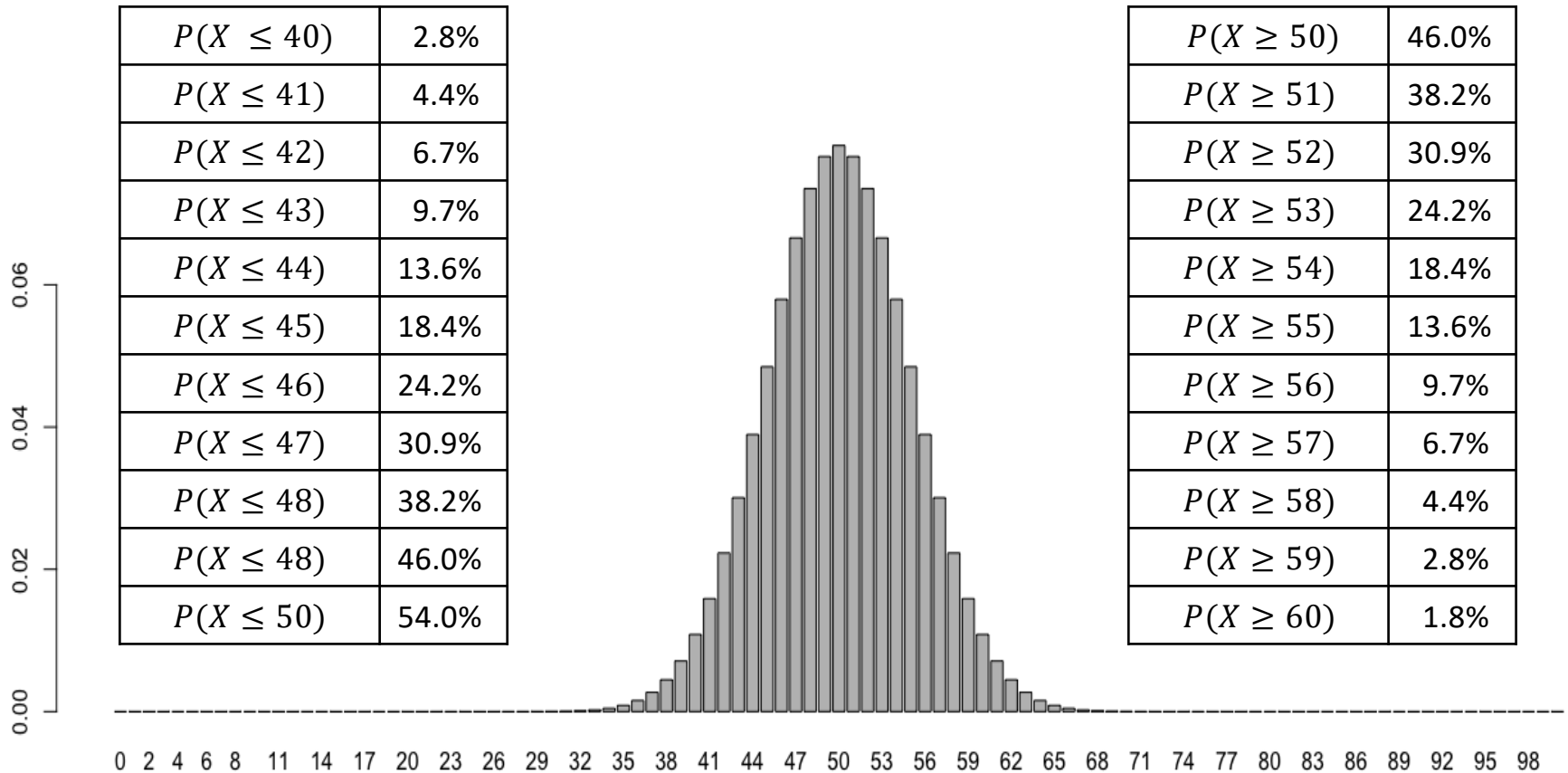
# $X$ = Number of heads

## Probability mass function:

Shows the probability of observing a specific number of heads in 100 independent fair coin flips.

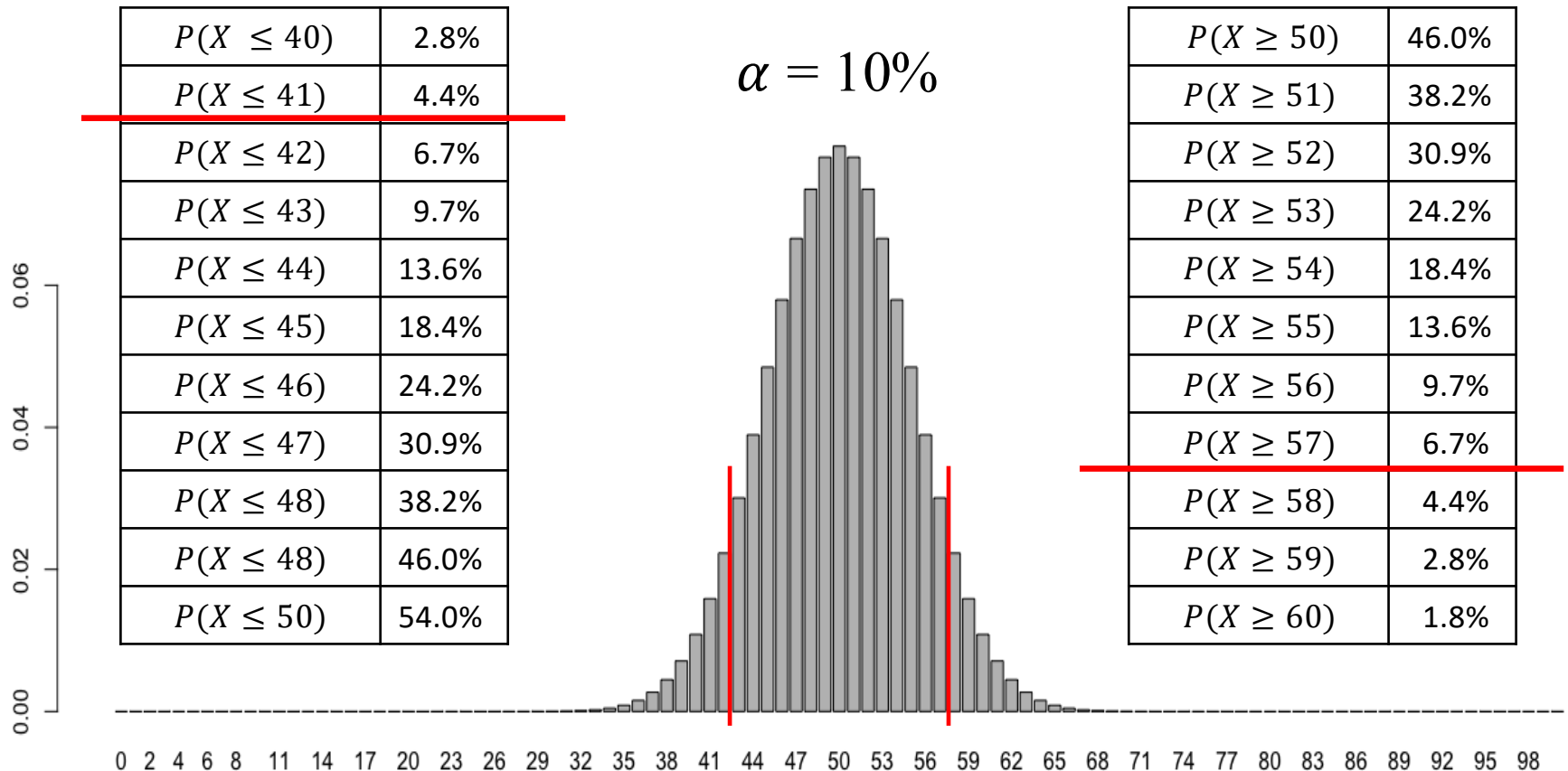


# X = Number of heads



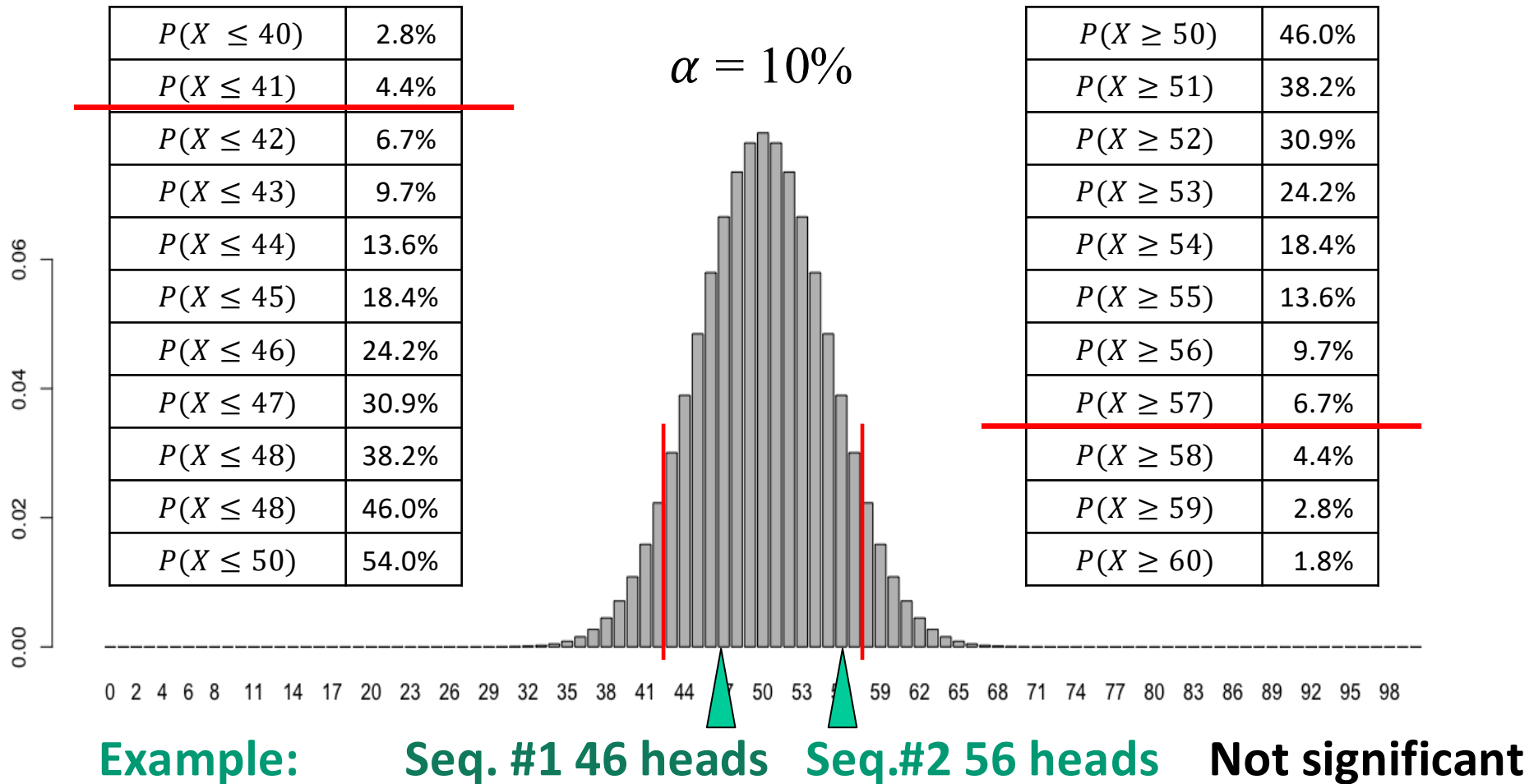
# X = Number of heads

Significance level  $\alpha$  = probability of falsely accusing the innocent of guilt



# X = Number of heads

Significance level  $\alpha$  = probability of falsely accusing the innocent of guilt



# Run lengths

- ▶ A run is a subsequence of equal outcomes

## Sequence #1:

T H T T H H T H T H T T H T H T T H H H T H H T T H  
T H H H T H T

1

2

3

T H T T T H T T H H T H H H T T T H T H H T H T H H T T T  
H T T T H T T H T H T T T H H T H H T T H T T T H H T T  
H T T H T H T T H H

# Number of runs

- ▶ A run is a subsequence of equal outcomes

**Sequence #1:**

|                       |    |    |   |   |   |   |   |
|-----------------------|----|----|---|---|---|---|---|
| <b>Length of run</b>  | 1  | 2  | 3 | 4 | 5 | 6 | 7 |
| <b>Number of runs</b> | 33 | 20 | 9 | 0 | 0 | 0 | 0 |

**Sequence #2:**

|                       |    |    |   |   |   |   |   |
|-----------------------|----|----|---|---|---|---|---|
| <b>Length of run</b>  | 1  | 2  | 3 | 4 | 5 | 6 | 7 |
| <b>Number of runs</b> | 25 | 11 | 9 | 4 | 2 | 0 | 0 |

**Expected from 100 independent coin flips, on average:**

|                       |       |       |      |      |      |      |      |
|-----------------------|-------|-------|------|------|------|------|------|
| <b>Length of run</b>  | 1     | 2     | 3    | 4    | 5    | 6    | 7    |
| <b>Number of runs</b> | 25.50 | 12.62 | 6.25 | 3.09 | 1.53 | 0.76 | 0.38 |



# Number of runs

- ▶ A run is a subsequence of equal outcomes

Sequence #1:

| Length of run  | 1  | 2  | 3 | 4 | 5 | 6 | 7 |
|----------------|----|----|---|---|---|---|---|
| Number of runs | 33 | 20 | 9 | 0 | 0 | 0 | 0 |

Sequence #2:

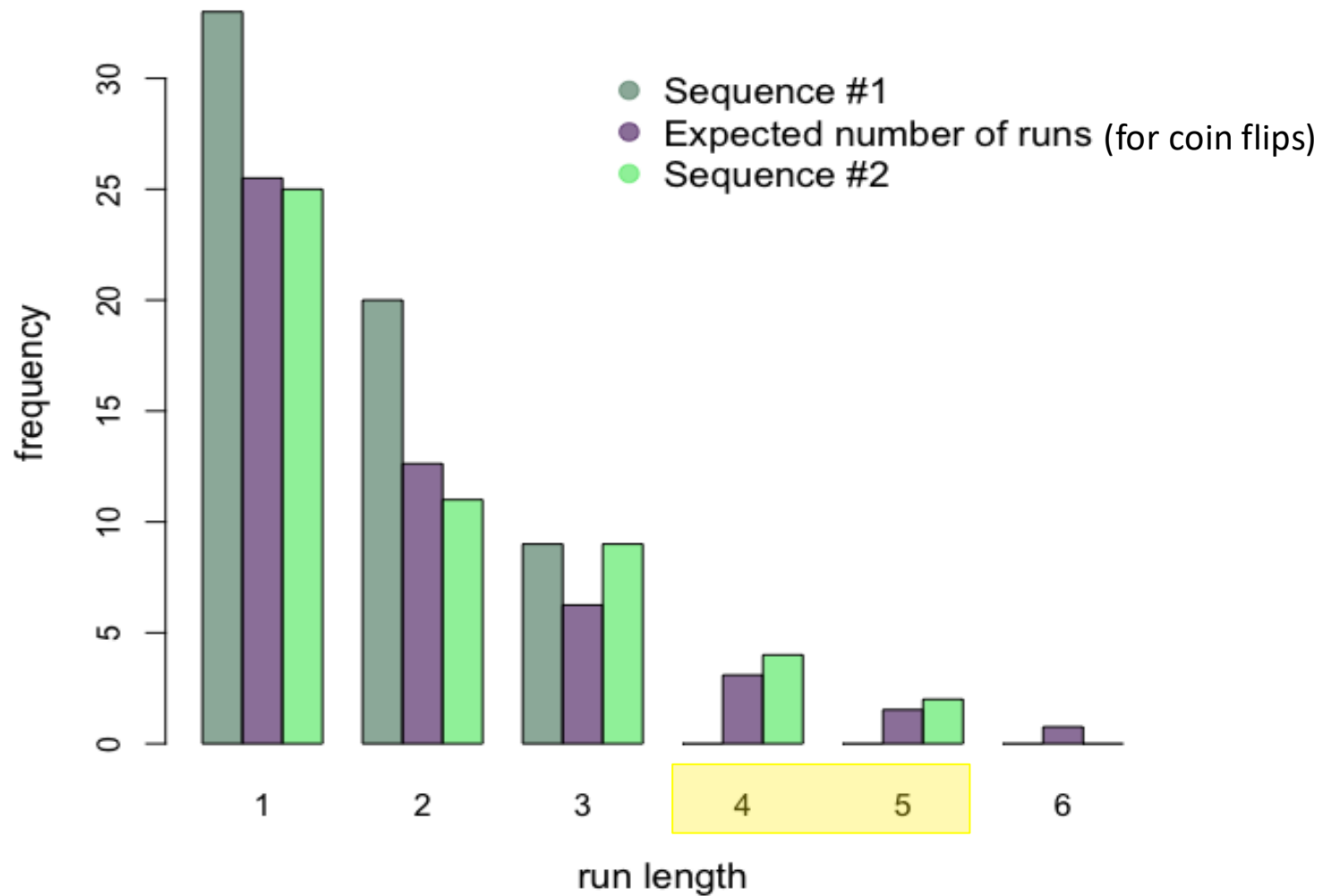
| Length of run  | 1  | 2  | 3 | 4 | 5 | 6 | 7 |
|----------------|----|----|---|---|---|---|---|
| Number of runs | 25 | 11 | 9 | 4 | 2 | 0 | 0 |

Expected from 100 independent coin flips, on average:

| Length of run  | 1     | 2     | 3    | 4    | 5    | 6    | 7    |
|----------------|-------|-------|------|------|------|------|------|
| Number of runs | 25.50 | 12.62 | 6.25 | 3.09 | 1.53 | 0.76 | 0.38 |



# Number of runs



# Number of runs: deriving the distribution

A mathematical exercise

*(doable e.g. with content from ST118 Introduction to Probability Theory)*

Model a sequence of  $n$  fair coin tosses by a sequence of independent random variables  $X_i$  ( $i = 1, \dots, n$ ) with values in  $\{0, 1\}$  and with  $P(X_i = 1) = p$  and  $P(X_i = 0) = 1 - p = q$  for all  $i = 1, \dots, n$ .

Let  $Z_r$  be the number of runs of heads or of tails of length  $r$  in  $n$  tosses of the coin. Calculate the expectation of  $Z_r$ .



# Longest run and number of switches

## Sequence #1:

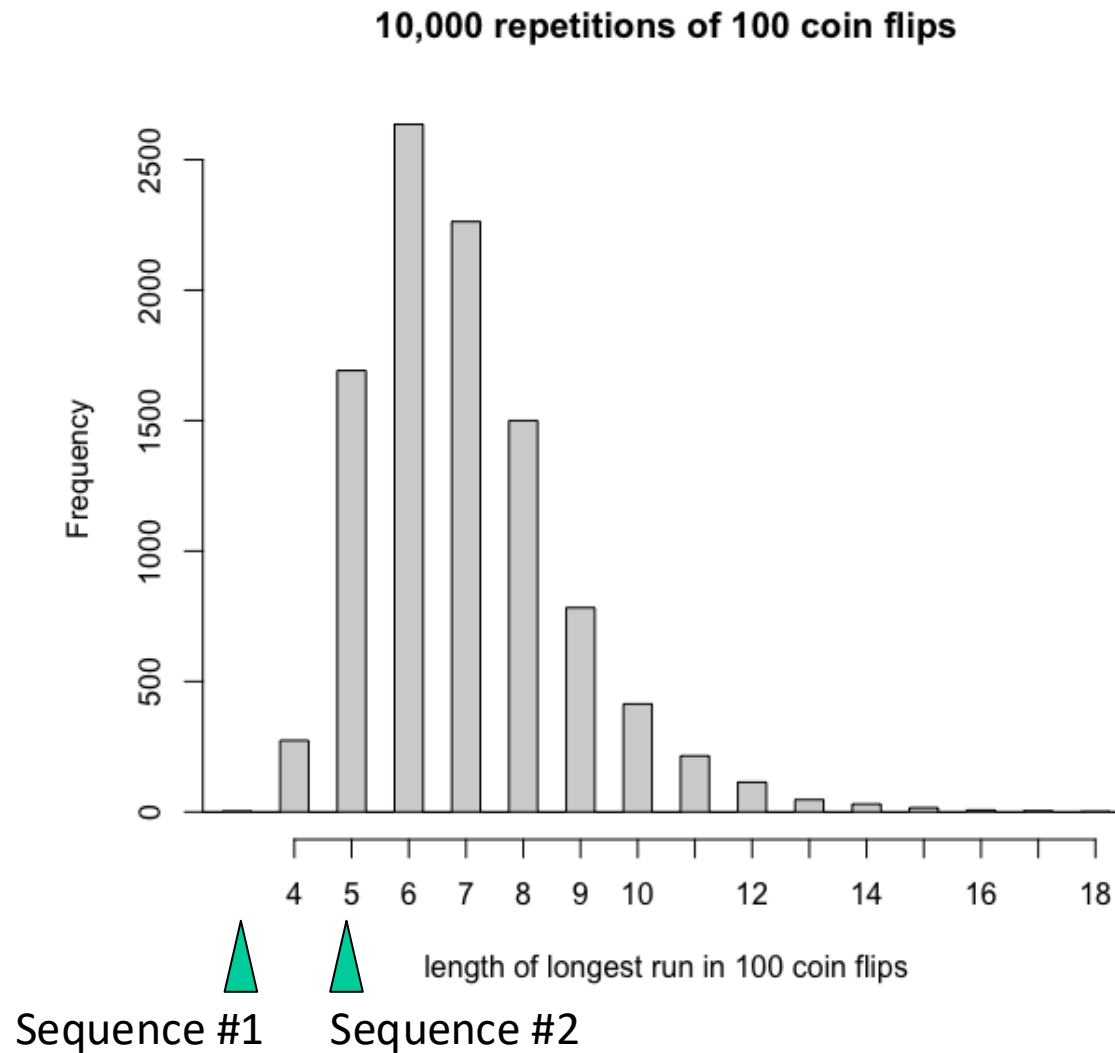
- Longest run is of length 3
- Number of switches is 61

## Sequence #2:

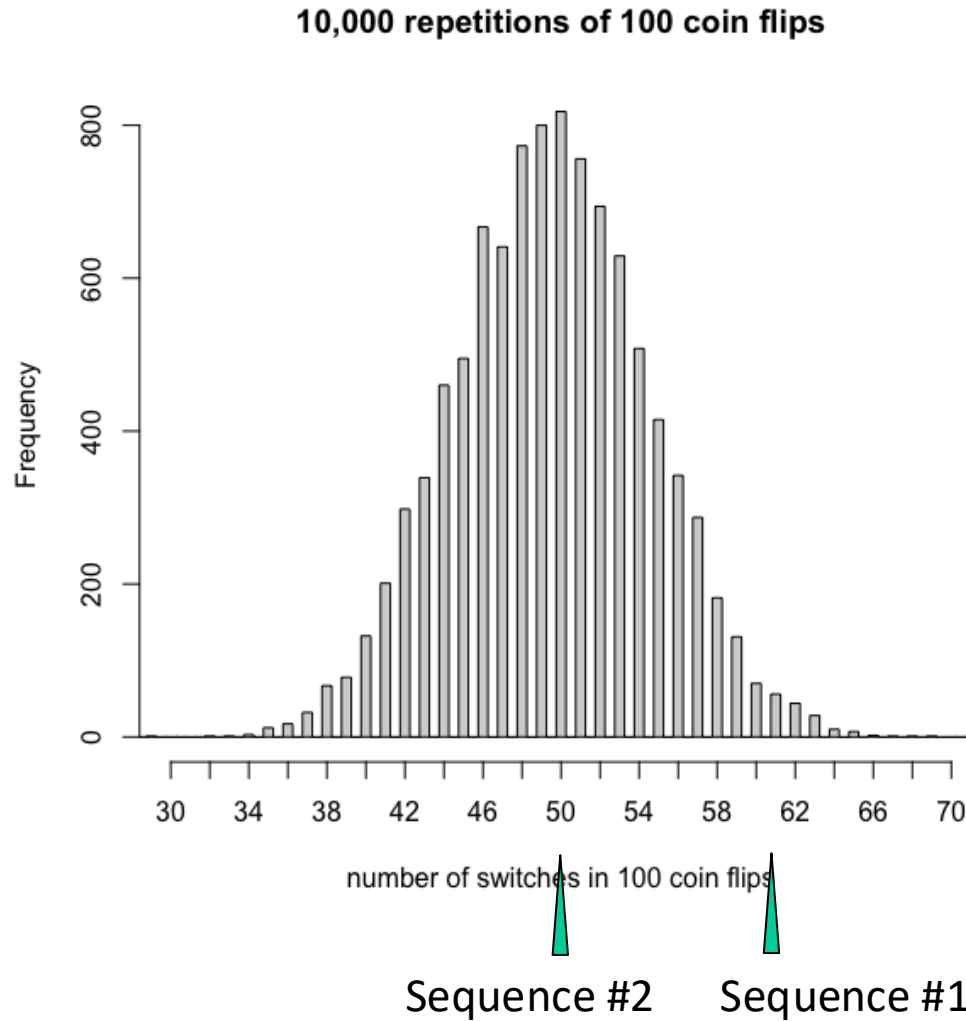
- Longest run is of length 5
- Number of switches is 50



# Longest run frequency distribution (via simulation)

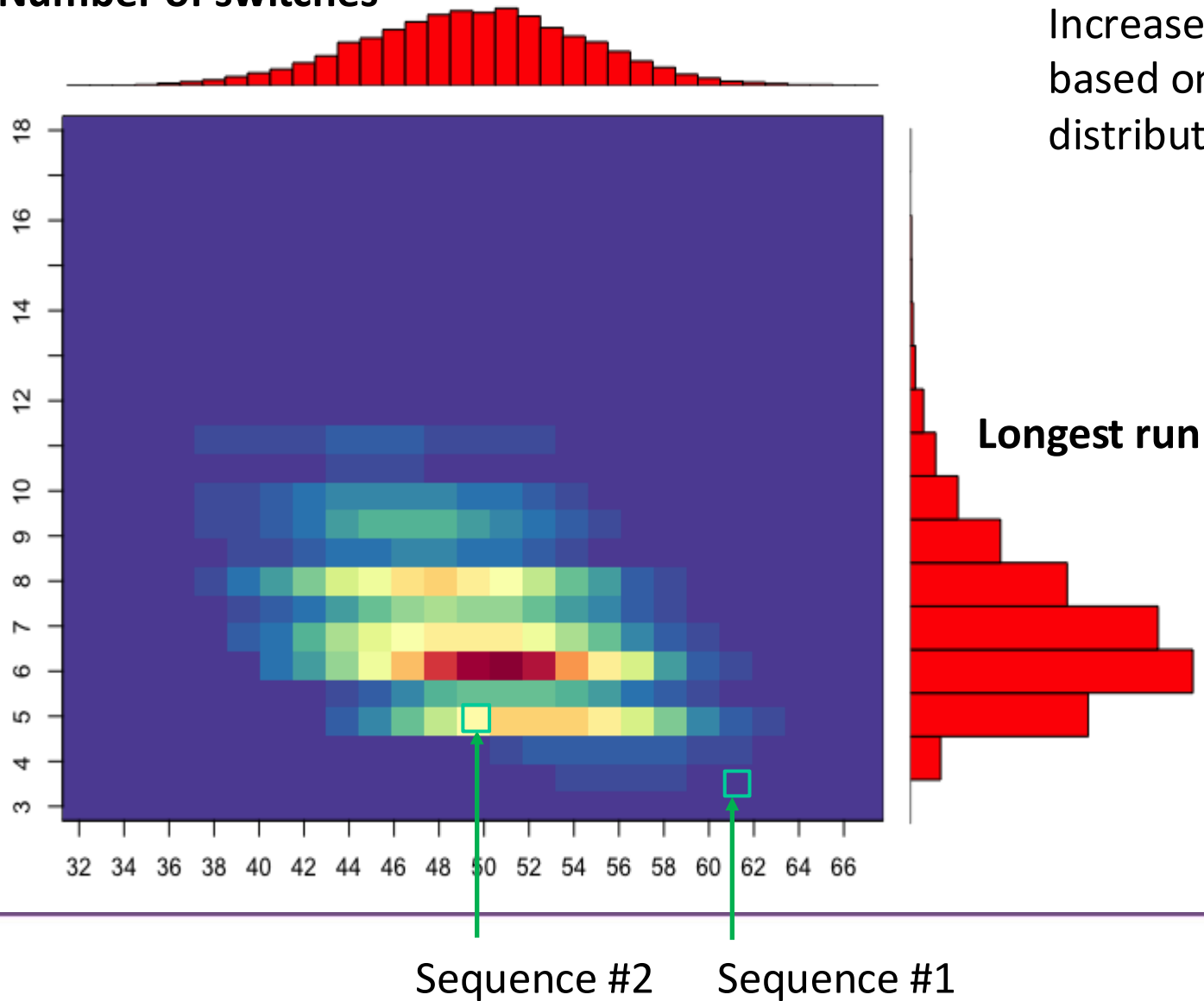


# Number of switches frequency distribution (via simulation)



# Both criteria

Number of switches



# Where is this being used?

- ▶ Random number generators
- ▶ Random sampling from probability distributions (“Monte Carlo simulations”)

Applications include:

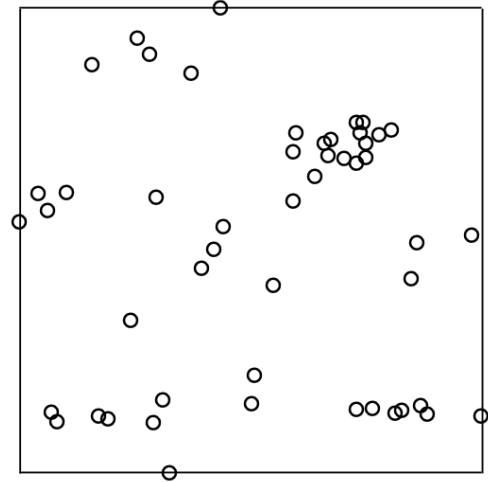
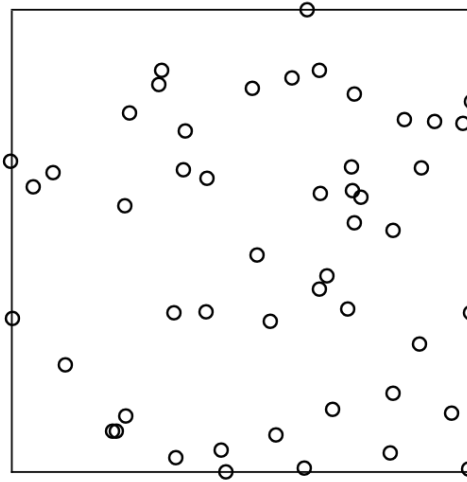
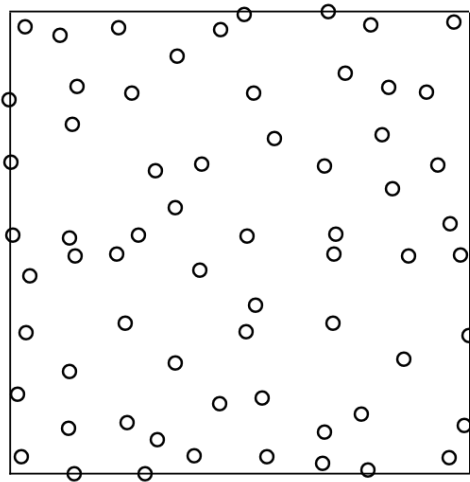
- weather modelling, climate change predictions
- option pricing, default risk assessment
- radiation therapy dosimetry
- protein folding
- structural engineering (reliability and risk)
- search and rescue operations
- policy evaluation and decision support



# Spotting true randomness: a spatial example

Complete spatial randomness (CSR)

= a point pattern that has *no underlying systematic structure*



Which of the above patterns is CSR?

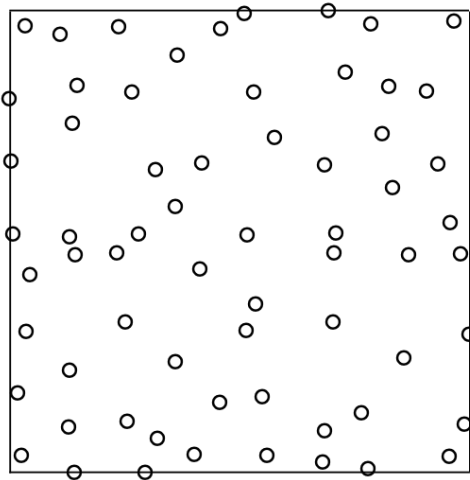
What about the others?



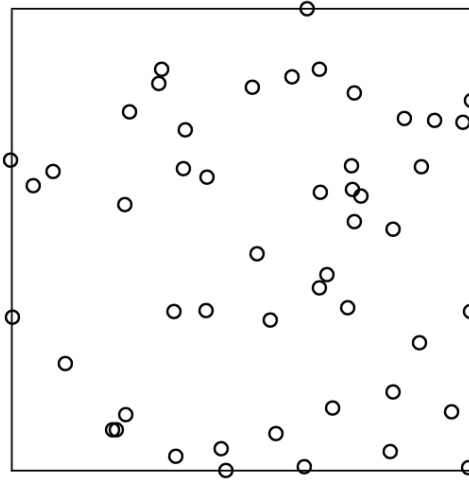
# Spotting true randomness: a spatial example

Complete spatial randomness (CSR)

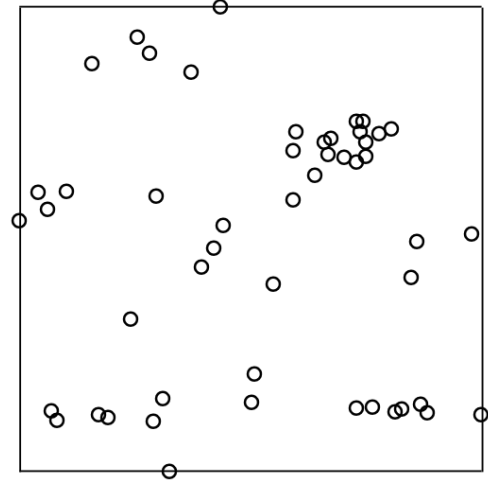
= a point pattern that has *no underlying systematic structure*



Regular (nearly)



CSR



Clustering



**Thank you** - You are welcome to talk to me about randomness or other aspects of this presentation or the courses by the Department of Statistics.

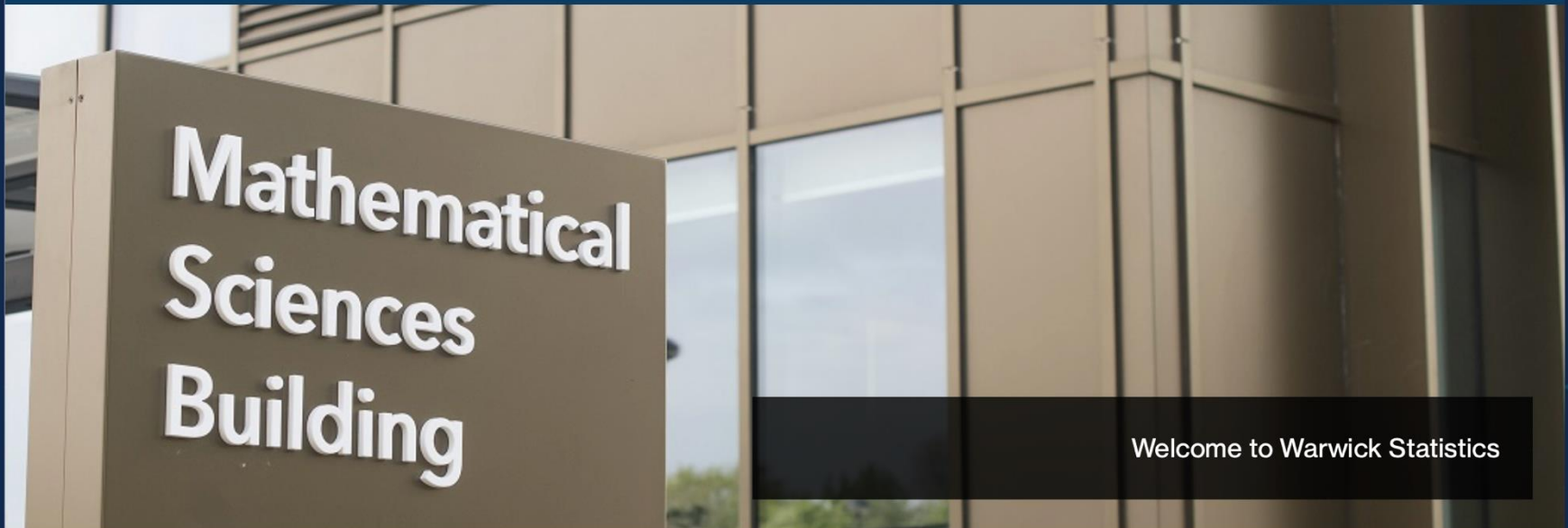
**UNIVERSITY  
OF WARWICK**

[Study](#) [Research](#) [Business](#) [Alumni](#) [News](#) [Engagement](#)

Search Warwick 

## ≡ Department of Statistics

[Study Here](#) ▾ | [Current Students](#) ▾ | [Postgraduate](#) ▾ | [Research](#) ▾ | [Statistics Alumni](#) | [Working with Us](#) ▾ | [News](#) ▾ | [Events](#) | [People](#) ▾ | [Contact](#) | [Intranet](#)  ▾



Statistics and probability theory are key to understanding data and underlying processes involving risk and uncertainty. Founded in 1972, the Department of Statistics is internationally renowned for wide-ranging research and innovative undergraduate and postgraduate degree programmes.