foundations of data science for everyone

II: probability and statistics

dr.federica bianco | fbb.space | fedhere | fedhere

this slide deck

https://slides.com/federicabianco/fds_02

0

Recap

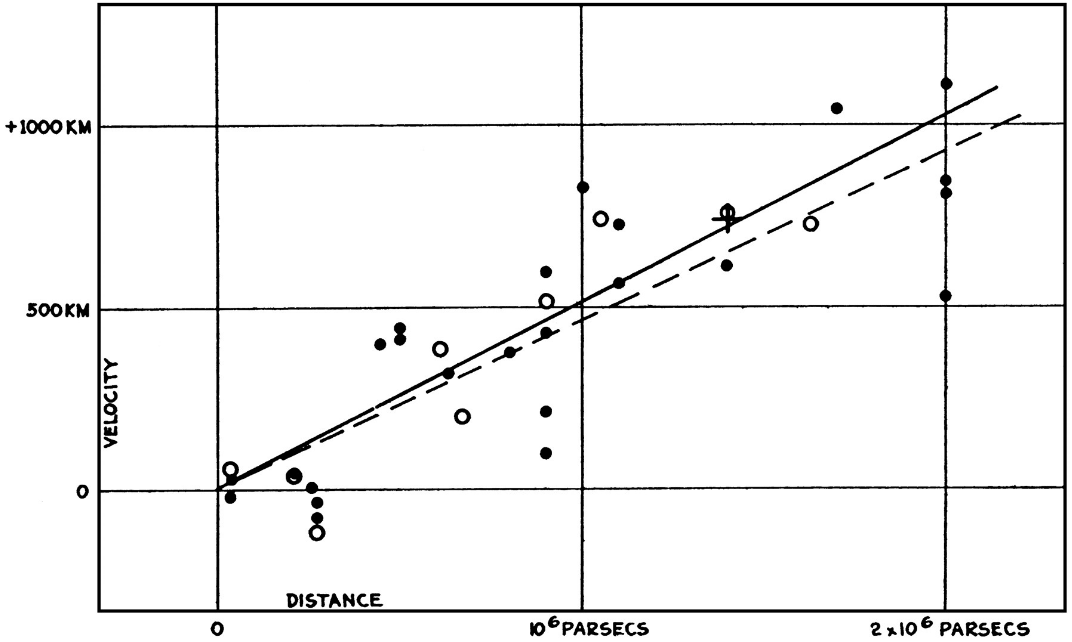
Improving Decision Making through Data

https://books.google.com/ngrams/interactive_chart?
content=%22+data+science+%22+%2B+%22+Data+Science+%22%2C%22+statistics+%22+%2B+%22+Statistics+%22%2C%
22+Informatics+%22%2B+%22+informatics+%22&year_start=1800&year_end=2019&case_insensitive=on&corpus=26&smooth
ing=3&direct_url=t1%3B%2C%28%22+data+science+%22+%2B+%22+Data+Science+%22%29%3B%2Cc0%3B.t1%3B%2C%
28%22+statistics+%22+%2B+%22+Statistics+%22%29%3B%2Cc0%3B.t1%3B%2C%28%22+Informatics+%22+%2B+%22+inf
ormatics+%22%29%3B%2Cc0

It is estimated that the amount of data collected by humans from the beginning of history through 2003 is 5 exabytes.

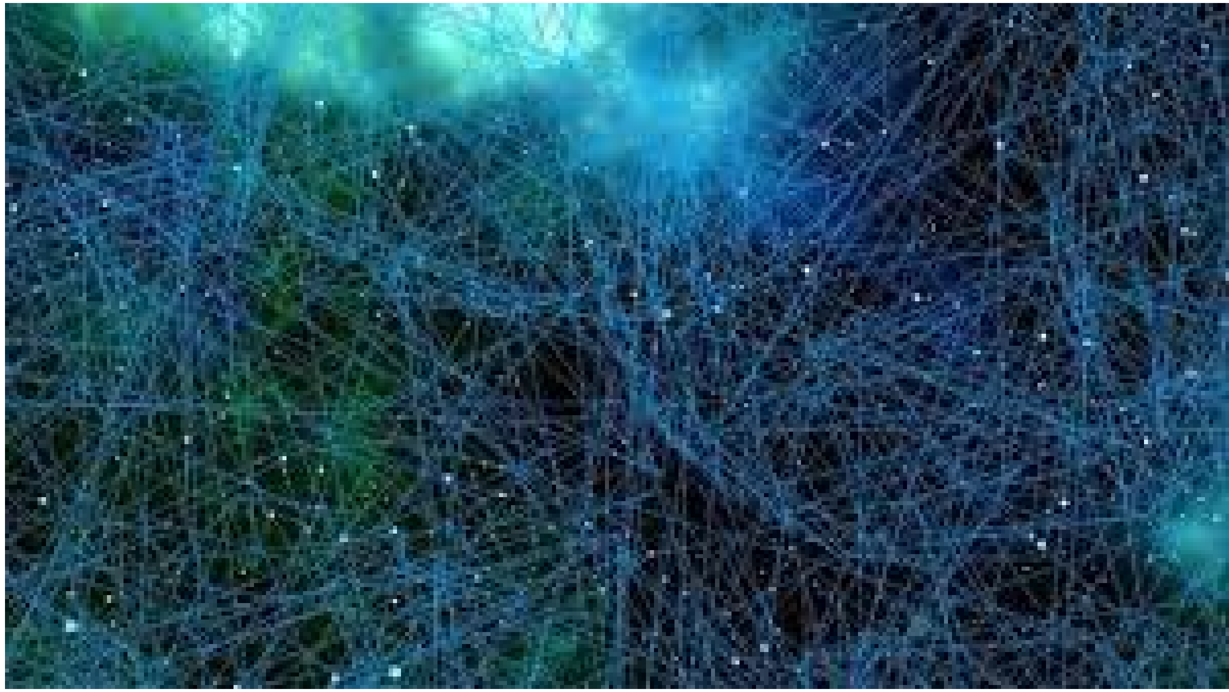
Since 2013 humans generate and store ~5 exabytes of data *every day*

Improving Decision Making through Data



how do i maximize information from few datapoints

It is estimated that the amount of data collected by humans from the beginning of history through 2003 is 5 exabytes.



how do i extract the critical informatoin and throw away unnecessary content from big data

Since 2013 humans generate and store ~5 exabytes of data *every day*

Improving Decision Making through Data

https://books.google.com/ngrams/interactive_chart?
content=%22+data+science+%22+%2B+%22+Data+Science+%22%2C%22+statistics+%22+%2B+%22+Statistics+%22%2C%
22+Informatics+%22%2B+%22+informatics+%22&year_start=1800&year_end=2019&case_insensitive=on&corpus=26&smootl
ing=3&direct_url=t1%3B%2C%28%22+data+science+%22+%2B+%22+Data+Science+%22%29%3B%2Cc0%3B.t1%3B%2C%
28%22+statistics+%22+%2B+%22+Statistics+%22%29%3B%2Cc0%3B.t1%3B%2C%28%22+Informatics+%22+%2B+%22+in
ormatics+%22%29%3B%2Cc0

Data Collection

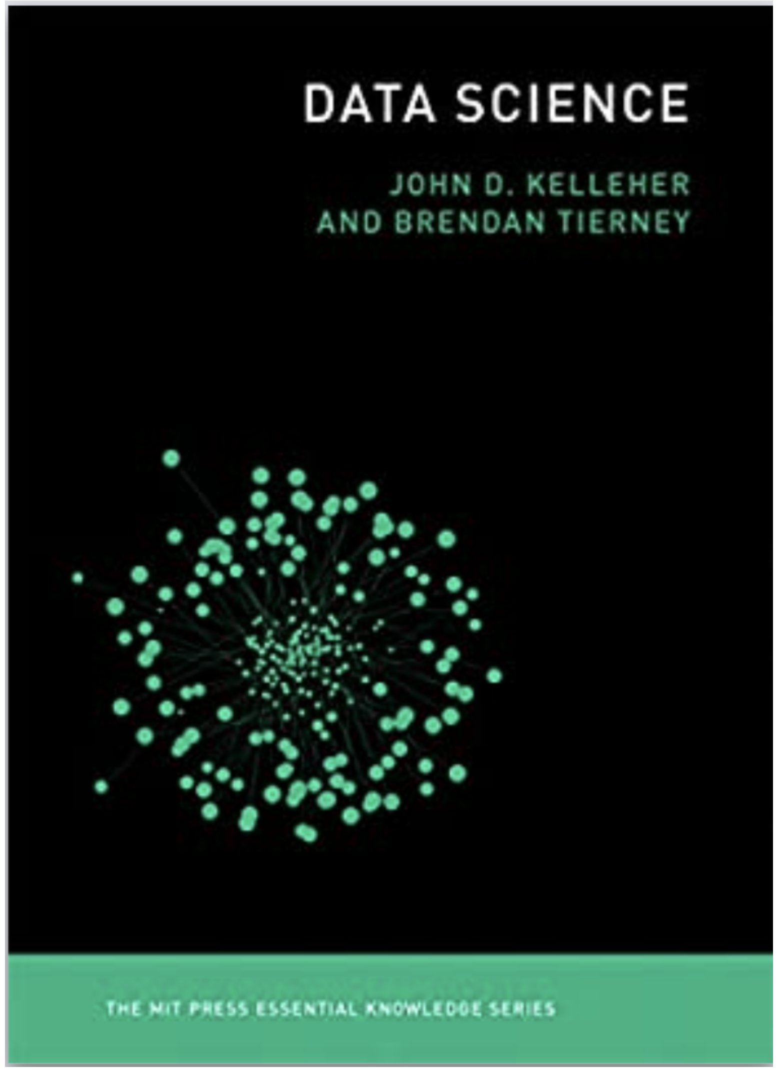
Pattern Extraction

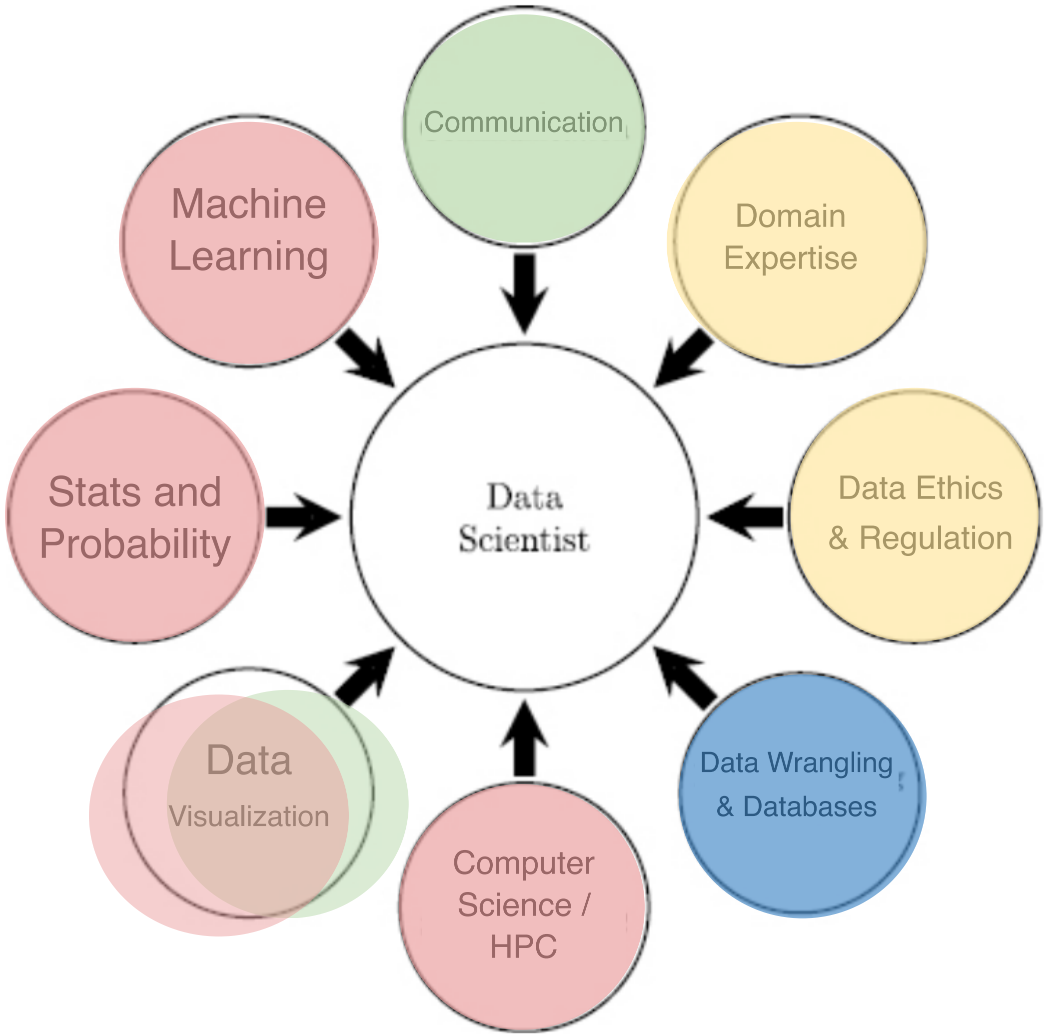
Interpretation

tech skills

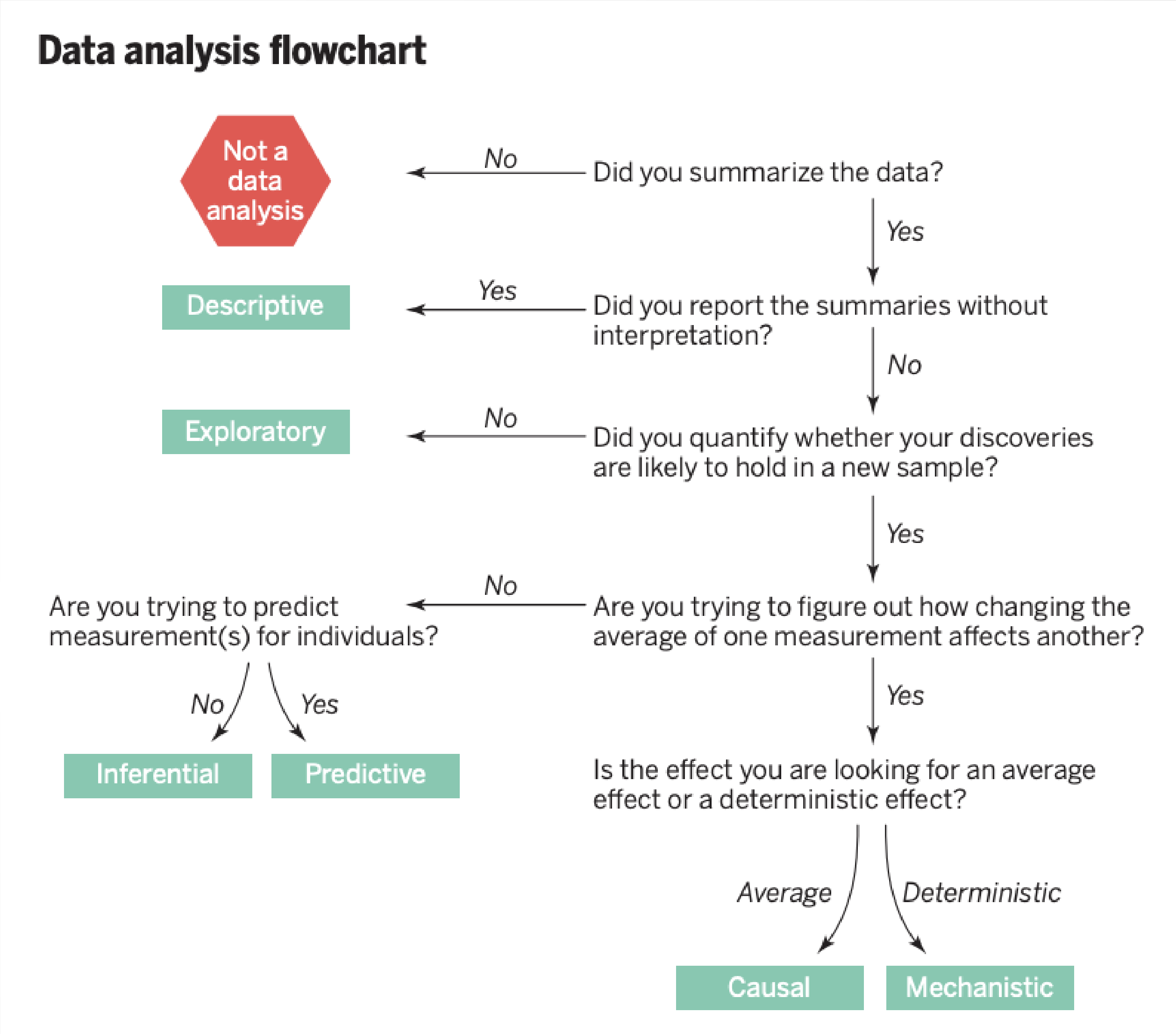
comm skills

additional skills





Data Science - Kelleher & Tierney



the steps of data-analysis and inference:

descriptive and exploratory analysis

- how is data organized
- is data complete?
- what are the statistical properties of the data

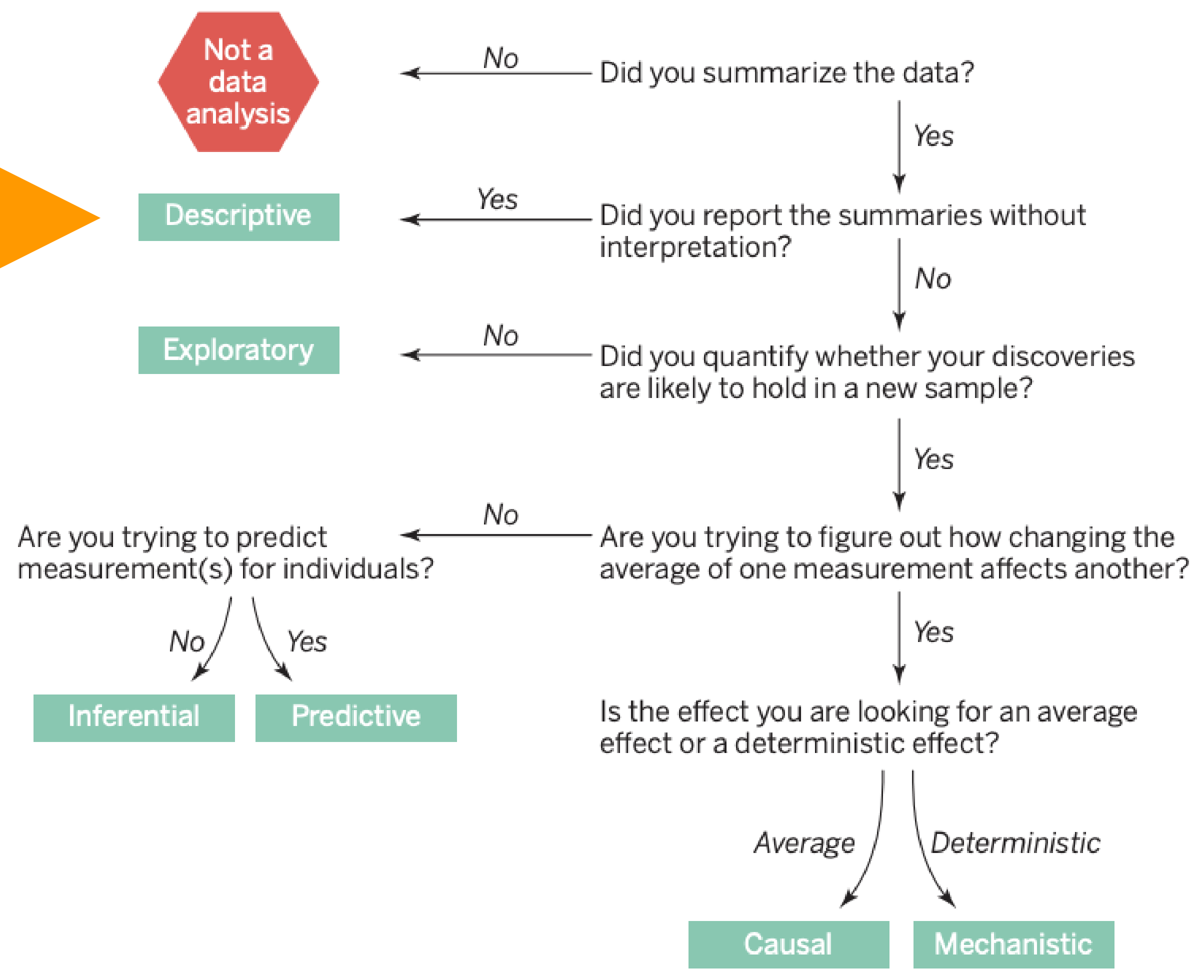
```python
import pandas as pd
df = pd.read_csv(file_name)
df.describe()
```

```python
videos.describe()
```

	category_id	views	likes	dislikes	comment_count
count	40949.000000	4.094900e+04	4.094900e+04	4.094900e+04	4.094900e+04
mean	19.972429	2.360785e+06	7.426670e+04	3.711401e+03	8.446804e+03
std	7.568327	7.394114e+06	2.288853e+05	2.902971e+04	3.743049e+04
min	1.000000	5.490000e+02	0.000000e+00	0.000000e+00	0.000000e+00
25%	17.000000	2.423290e+05	5.424000e+03	2.020000e+02	6.140000e+02
50%	24.000000	6.818610e+05	1.809100e+04	6.310000e+02	1.856000e+03
75%	25.000000	1.823157e+06	5.541700e+04	1.938000e+03	5.755000e+03
max	43.000000	2.252119e+08	5.613827e+06	1.674420e+06	1.361580e+06

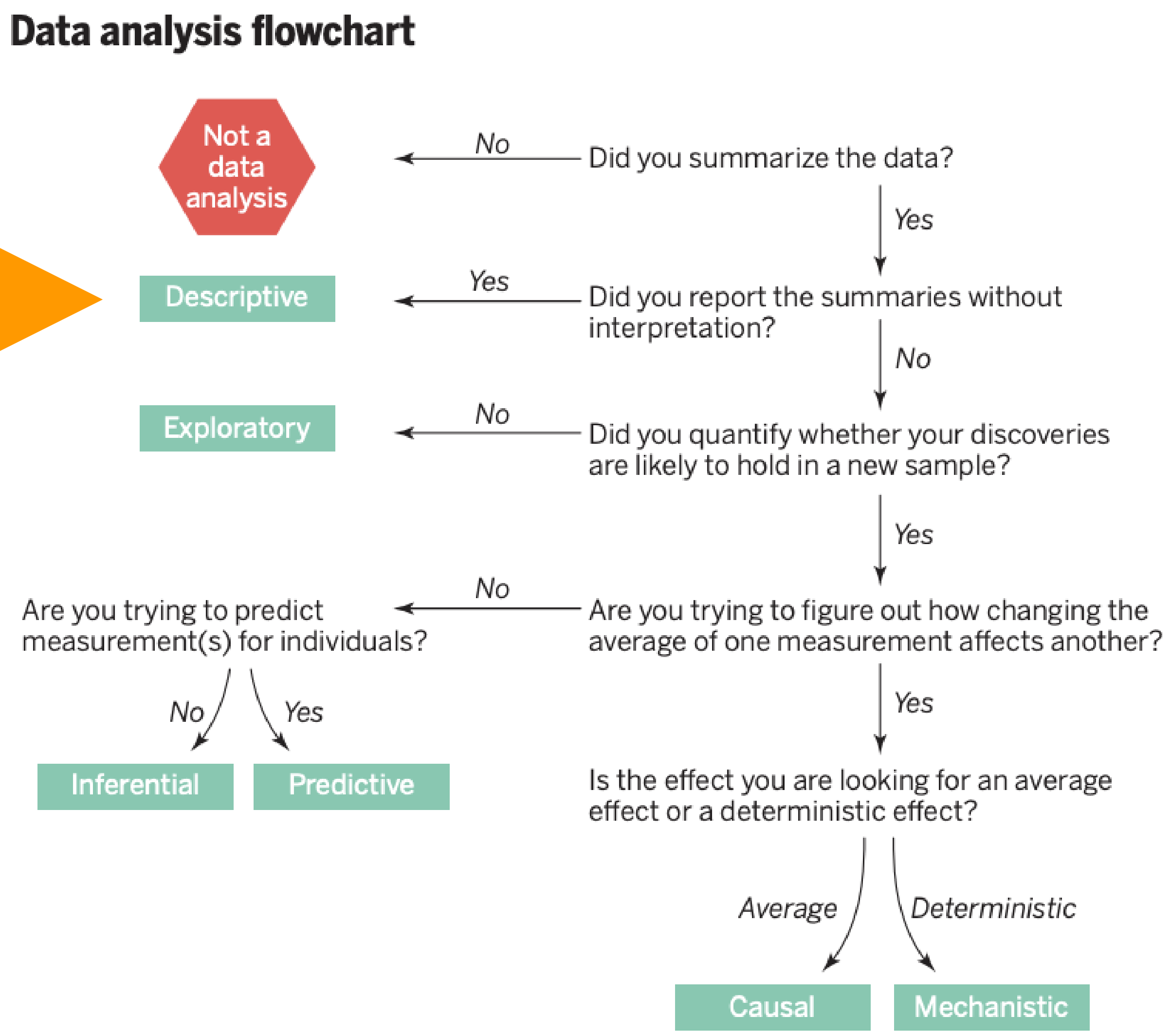
we will look at the statistical properties: mean, standard deviation, median, quantiles...

Data analysis flowchart



- how is data organized
- is data complete?
- what are the statistical properties of the data

```python
import pandas as pd
df = pd.read_csv(file_name)
df.describe()
```

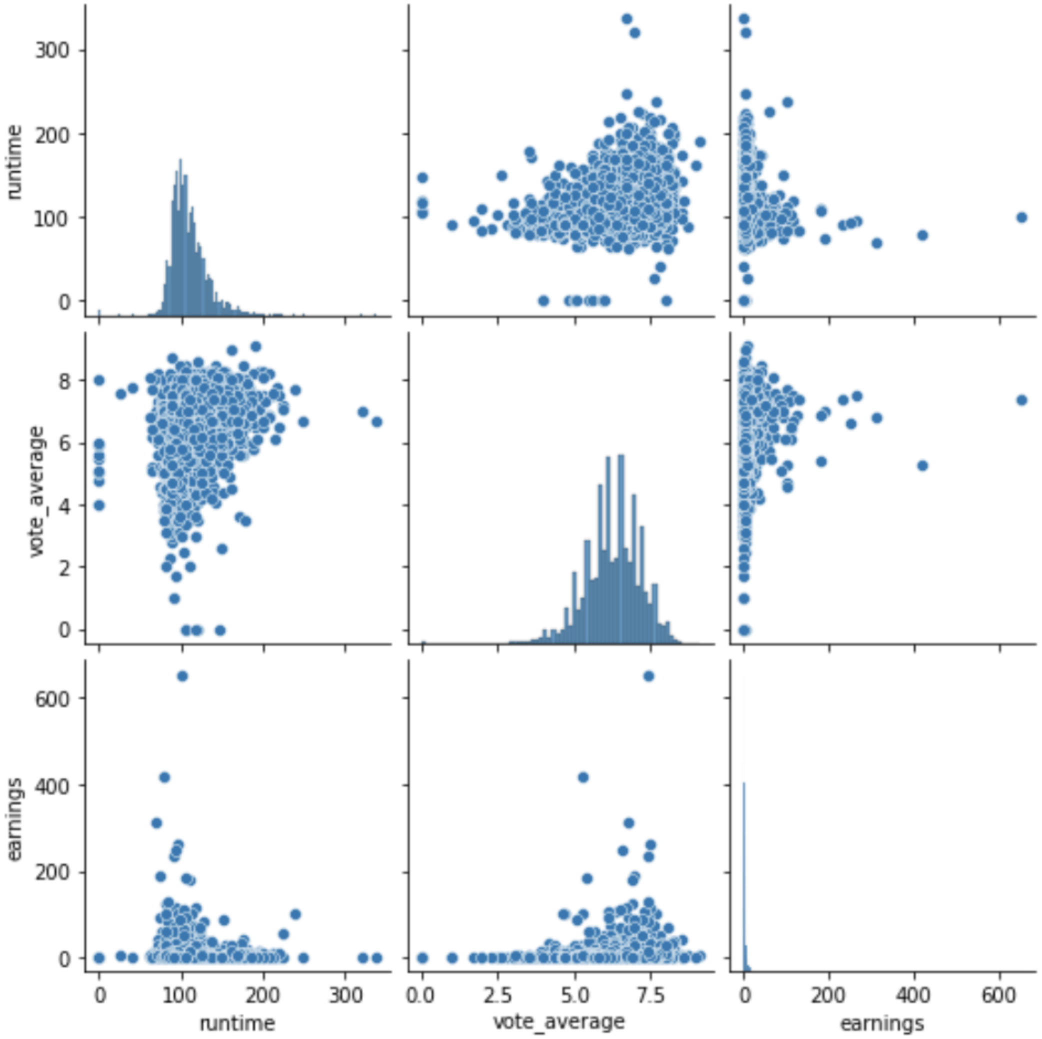


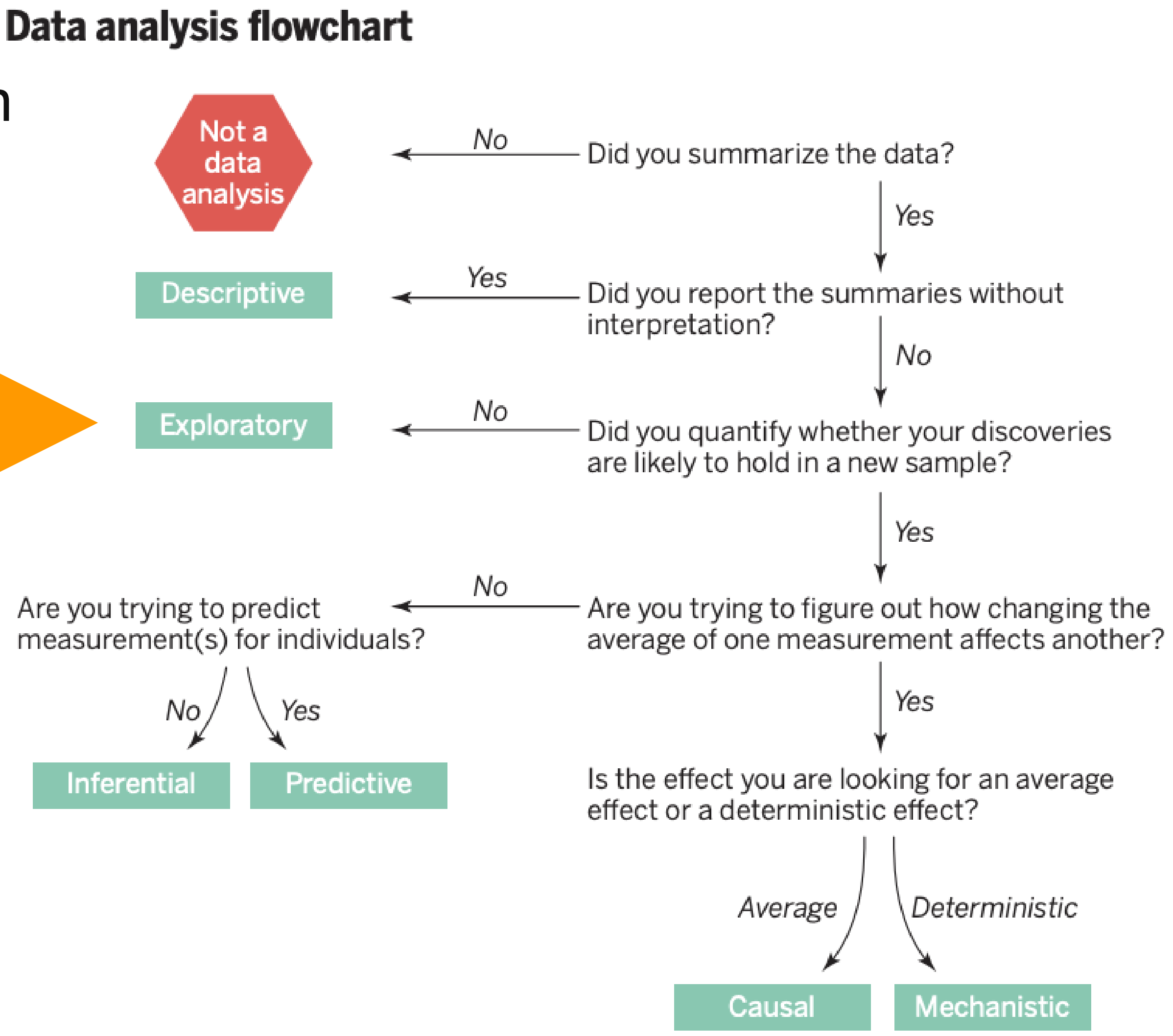

```python
videos.describe()
```

	category_id	views	likes	dislikes	comment_count
count	40949.000000	4.094900e+04	4.094900e+04	4.094900e+04	4.094900e+04
mean	19.972429	2.360785e+06	7.426670e+04	3.711401e+03	8.446804e+03
std	7.568327	7.394114e+06	2.288853e+05	2.902971e+04	3.743049e+04
min	1.000000	5.490000e+02	0.000000e+00	0.000000e+00	0.000000e+00
25%	17.000000	2.423290e+05	5.424000e+03	2.020000e+02	6.140000e+02
50%	24.000000	6.818610e+05	1.809100e+04	6.310000e+02	1.856000e+03
75%	25.000000	1.823157e+06	5.541700e+04	1.938000e+03	5.755000e+03
max	43.000000	2.252119e+08	5.613827e+06	1.674420e+06	1.361580e+06

we will look at the statistical properties: mean, standard deviation, median, quantiles...

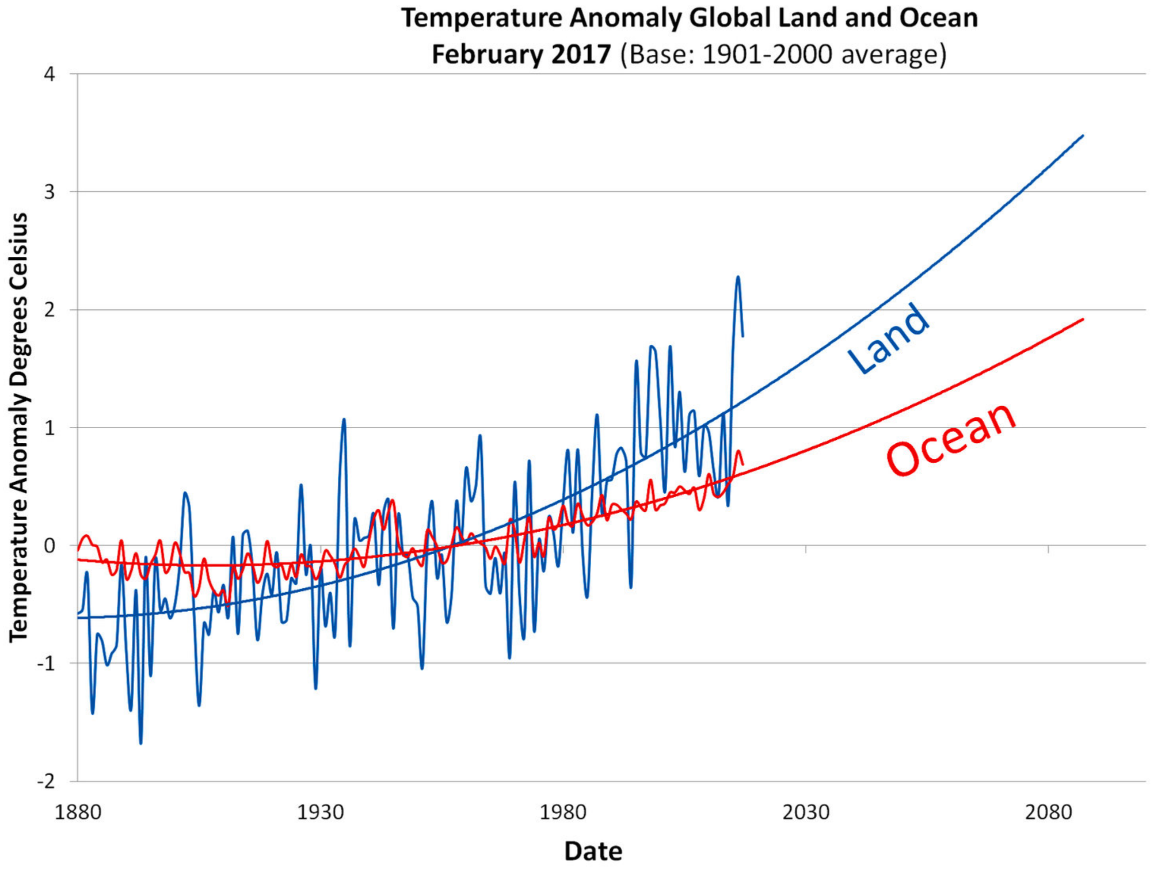
- searching for anomalies, trends
- searching for relationships between the measurements (correlation)

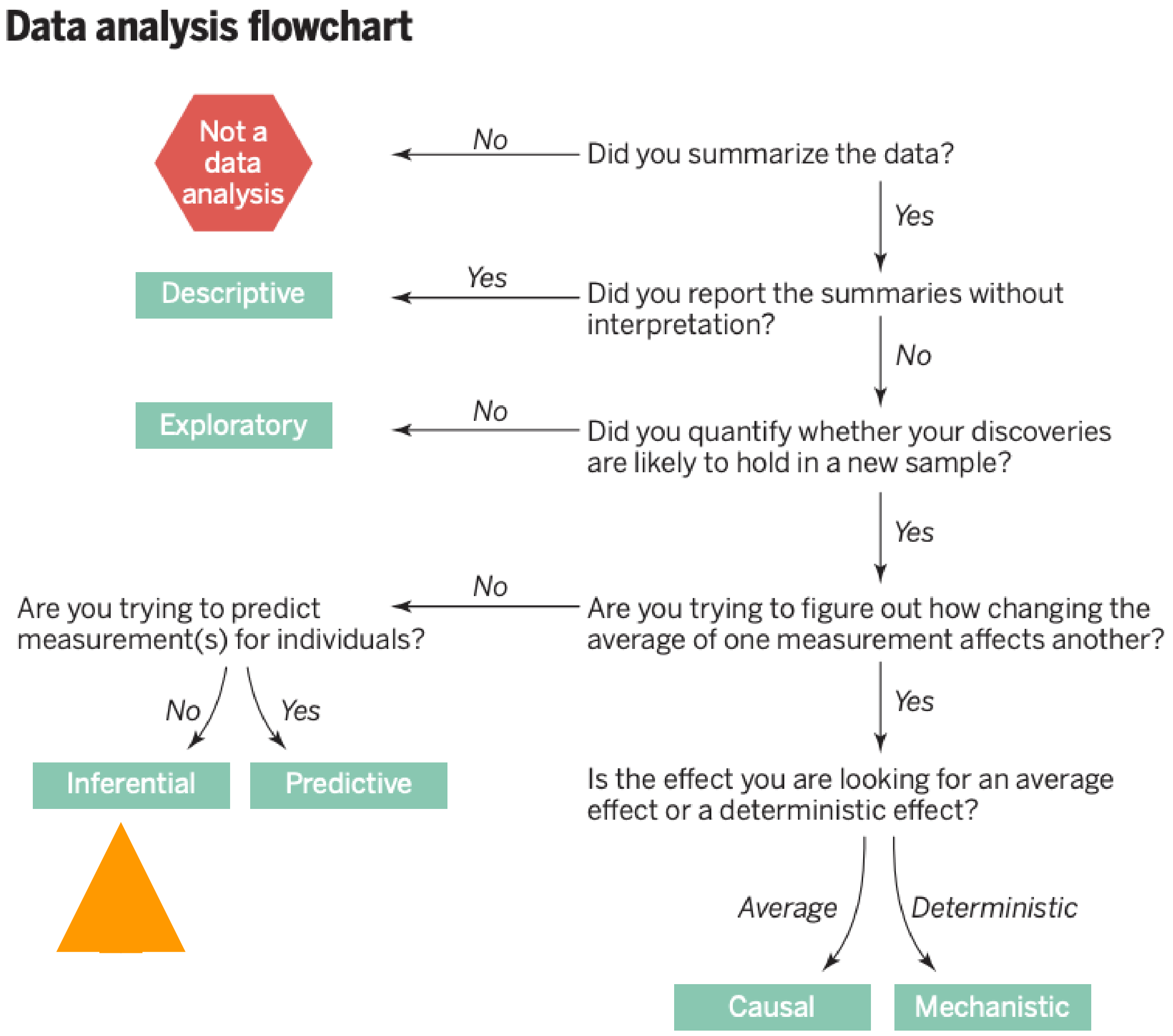




An exploratory data analysis builds on a descriptive analysis by searching for discoveries, trends, correlations, or relationships between the measurements to generate ideas or hypotheses.

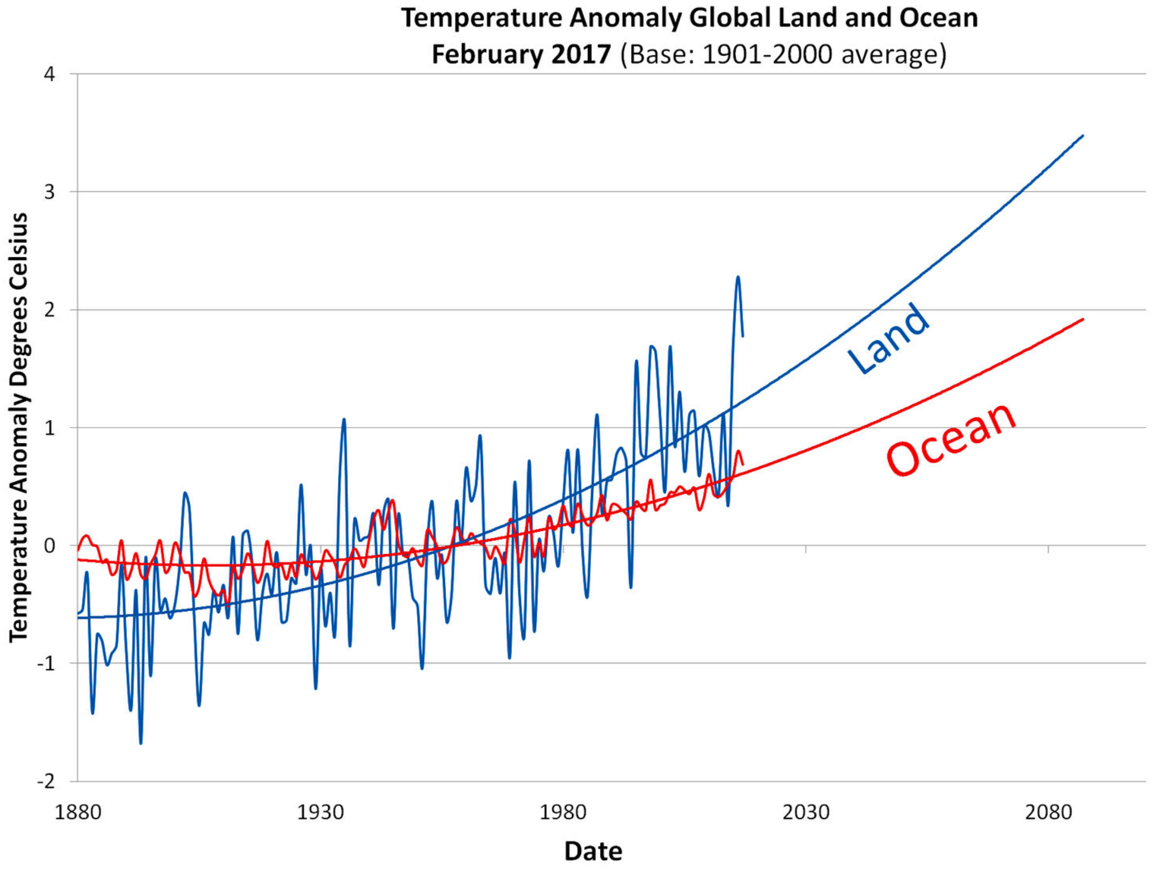
Inferential: will patterns hold?

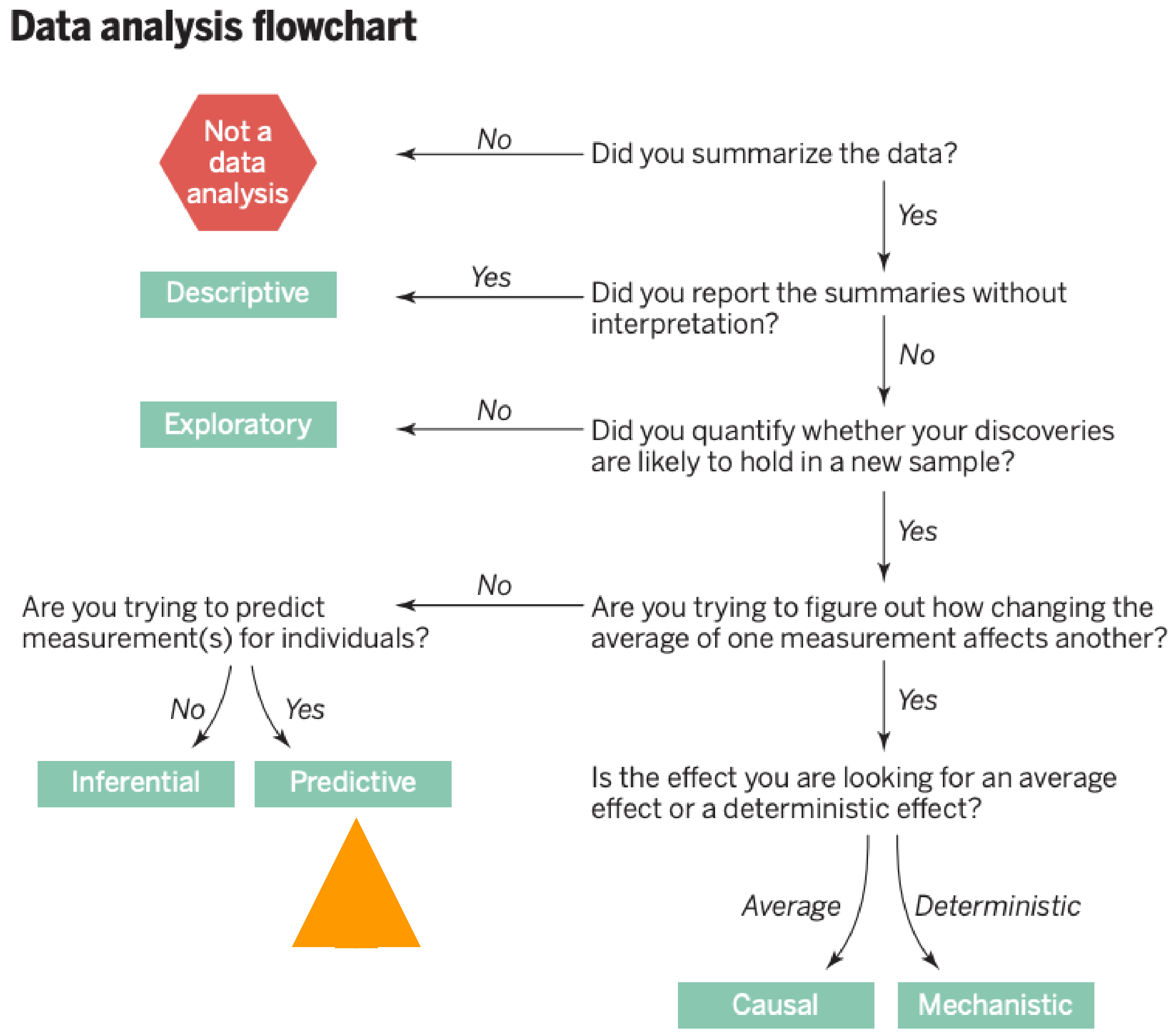




An inferential data analysis quantifies whether an observed pattern will likely hold beyond the data set in hand.

Predictive: what will be the temperature in 2030?

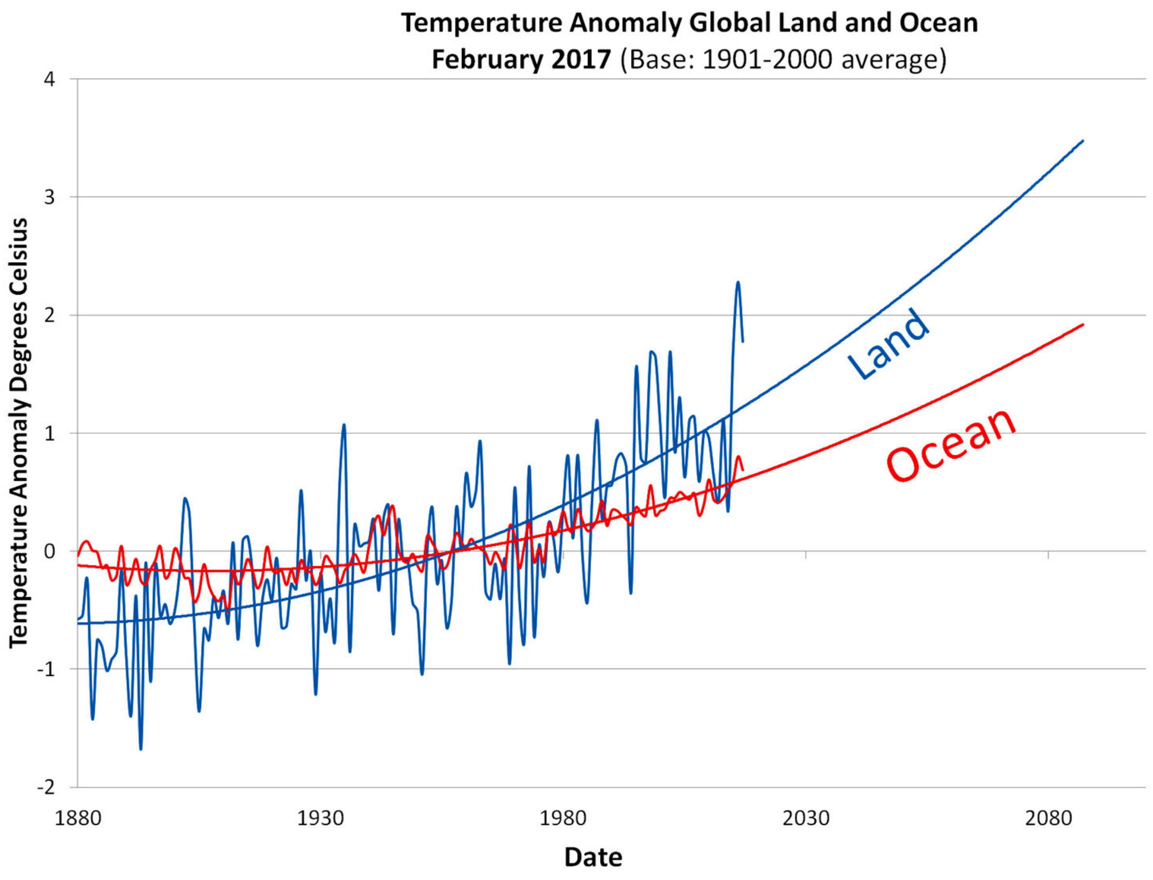


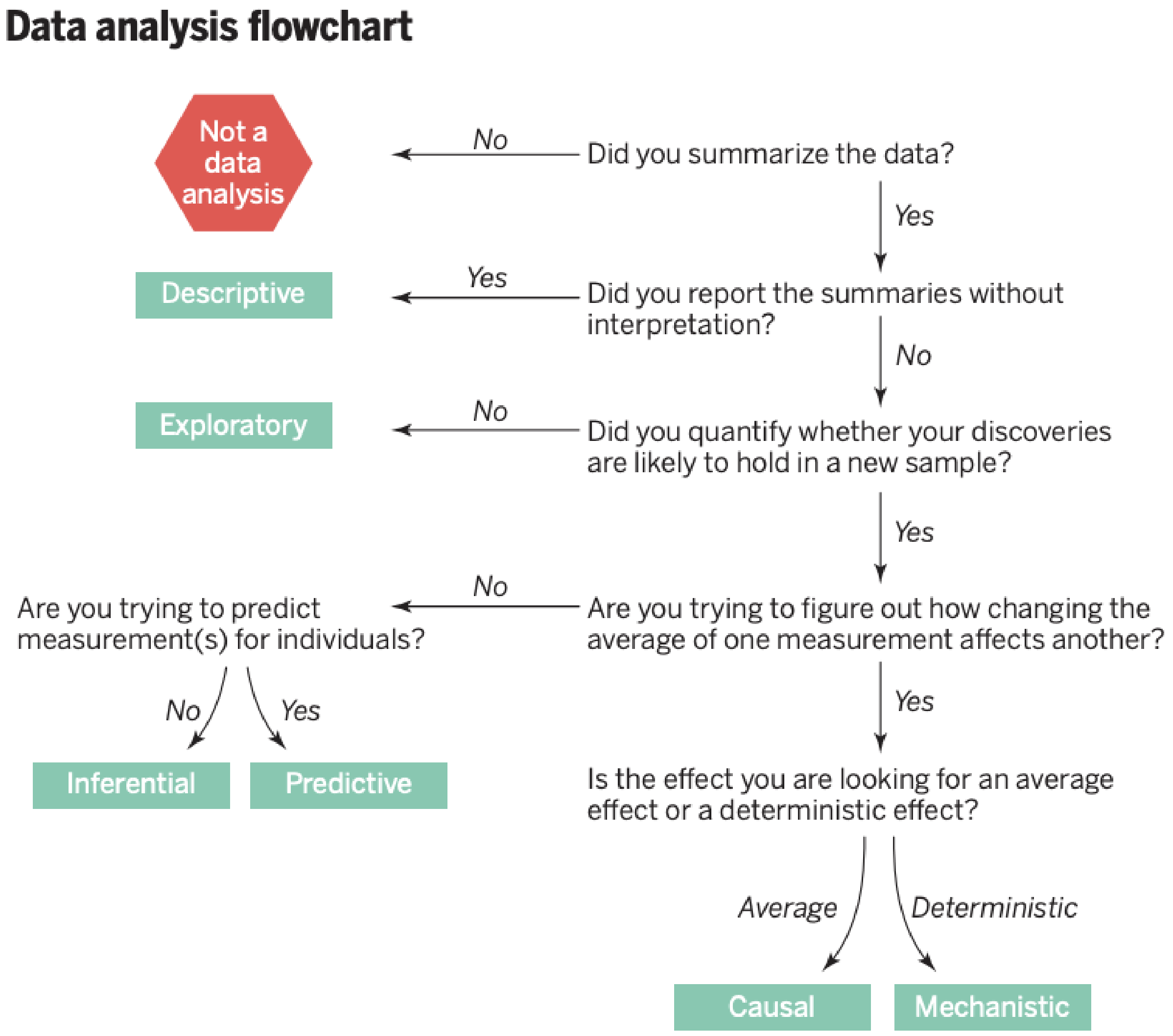


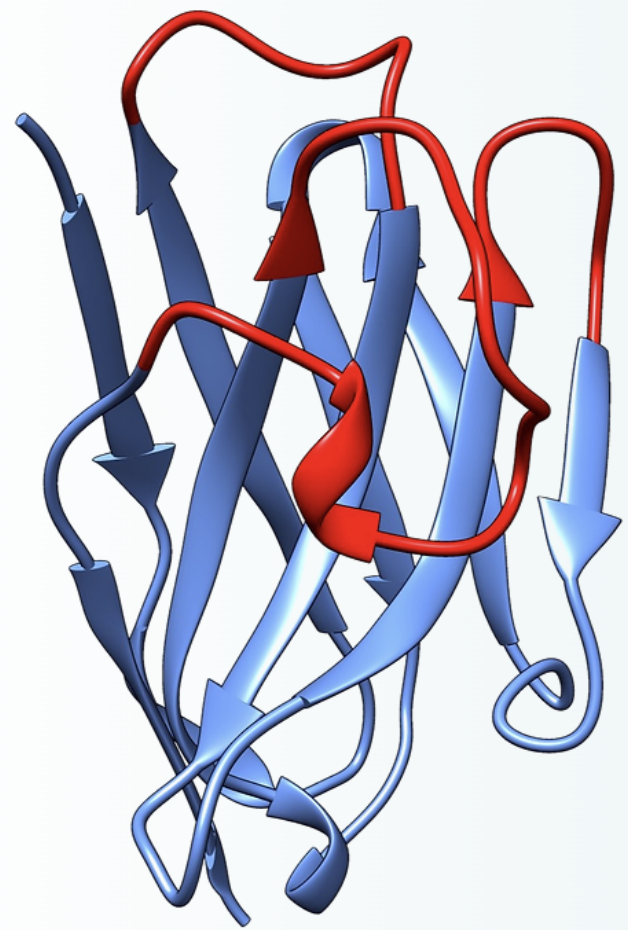
a predictive data analysis uses

a subset of measurements (the features) to predict another measurement (the outcome) on a single person or unit. Web sites

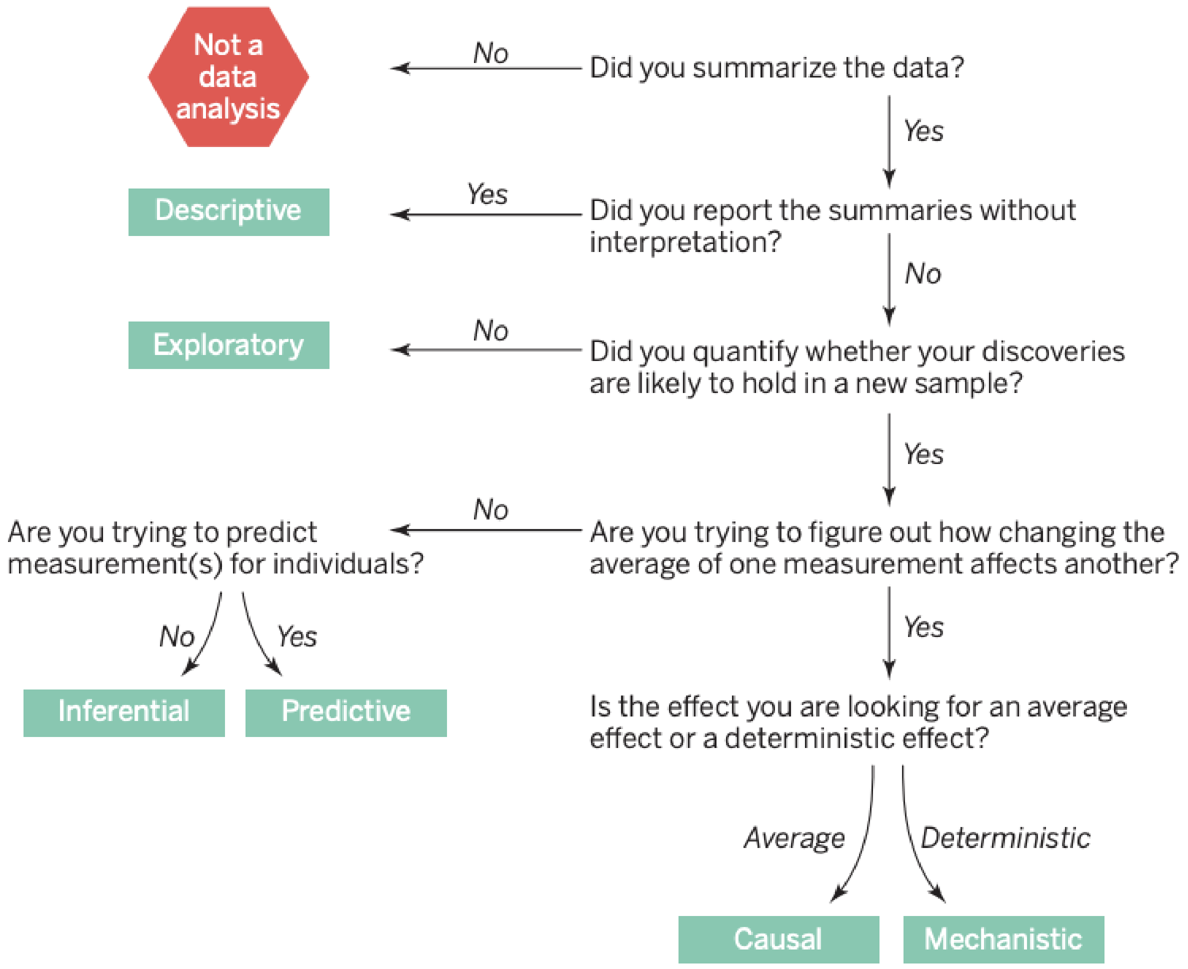
Causal: is the cause of the heating of the earth? CO2? Solar Cycles?







Data analysis flowchart



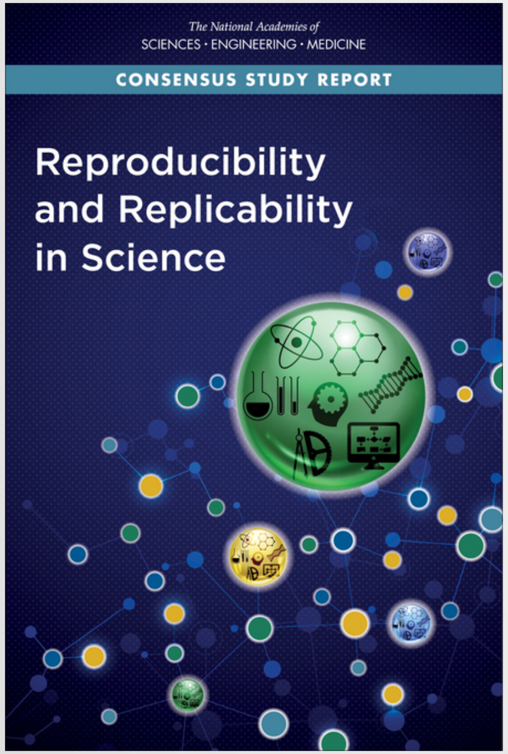
Engineering cells with nanobodies

1

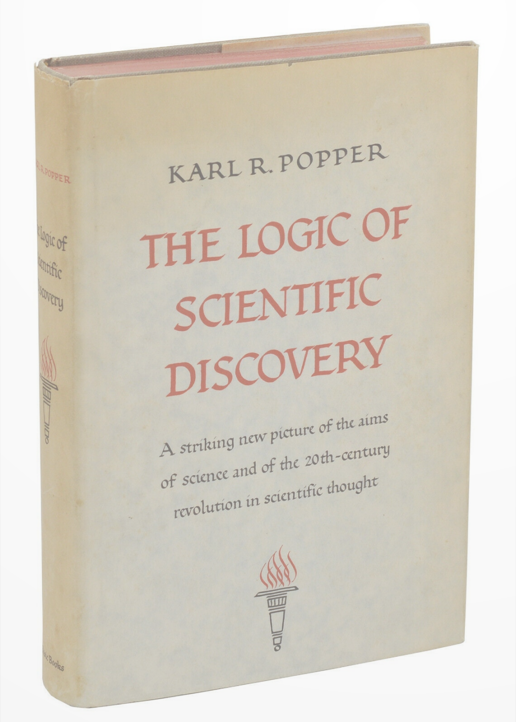
Data Science Practices

3 General principles of "good" science

Reproducibility



Falisifiability



Parsimony

github

reproducibility

allows reproducibility through code distribution

https://github.com

Reproducible research means:

all numbers in a data analysis can be recalculated exactly (down to stochastic variables!) using the **code** and **raw data** provided by the analyst.

Claerbout, J. 1990,

Active Documents and Reproducible Results, Stanford Exploration Project Report, 67, 139

Reproducibility

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

Reproducibility

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

why?

1

assures a result is grounded in evidence

#openscience

#opendata

Reproducibility

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

why?

2

facilitates scientific progress by avoiding the need to duplicate unoriginal research

Reproducibility

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

why?

3

facilitate collaboration and teamwork

Reproducibility

see slides week 1:

https://slides.com/d/d7YarW8/live#/6

Reproducible research means:

the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.

https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf

Reproducible research in practice:

all numbers in a data analysis can be recalculated exactly (down to stochastic variables!)

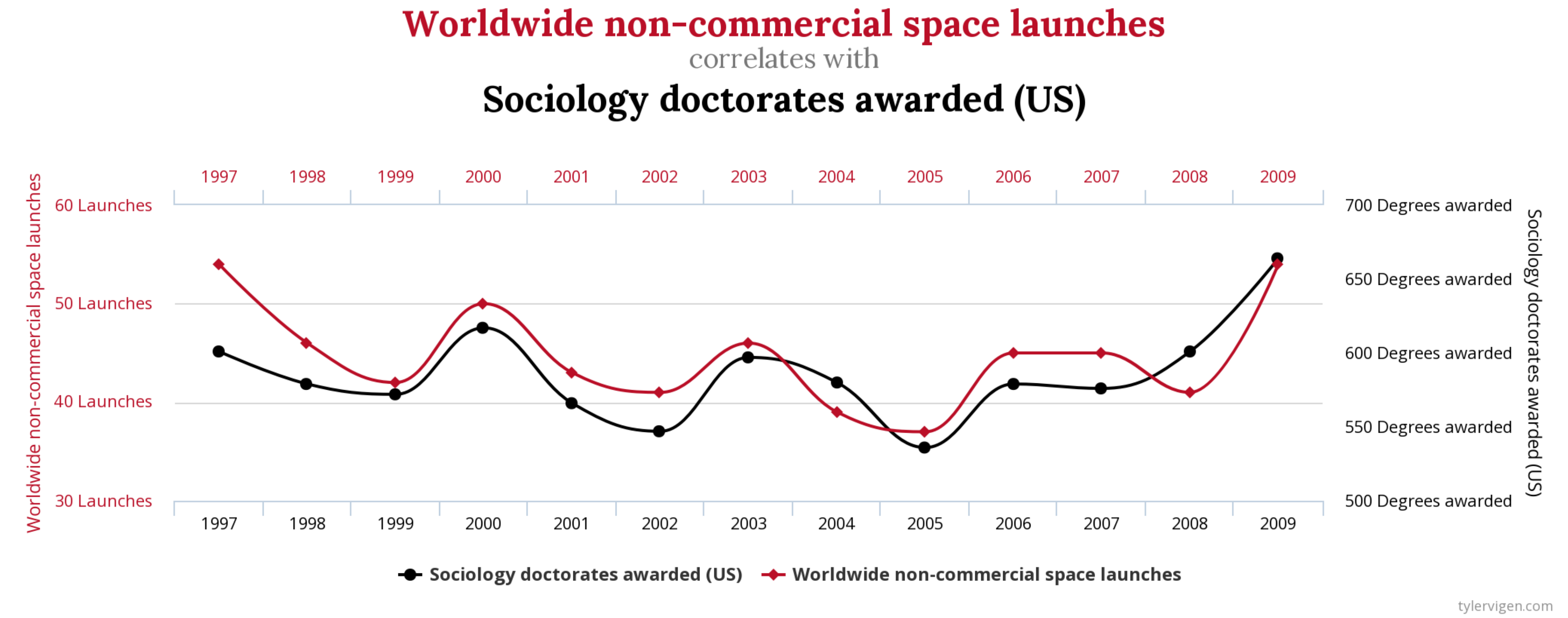
using the **code** and **raw data** provided by the analyst.

- provide raw data and code to reduce it to all stages needed to get outputs
- provide code to reproduce all figures
- provide code to reproduce all number outcomes

2 EXPLORATORY DATA ANALYSIS

correlation

Correlation does not imply causality!!



2 things may be related because they share a cause but not cause each other:

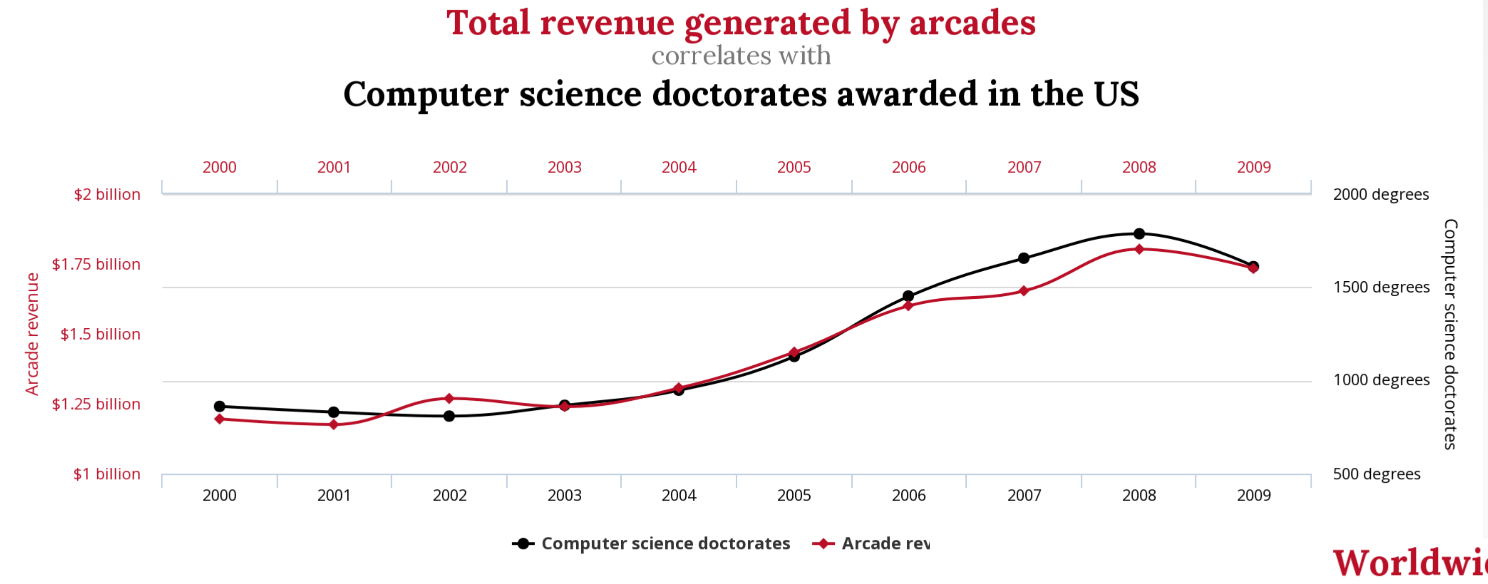
icecream sales with temperature /death by drowning with temperature

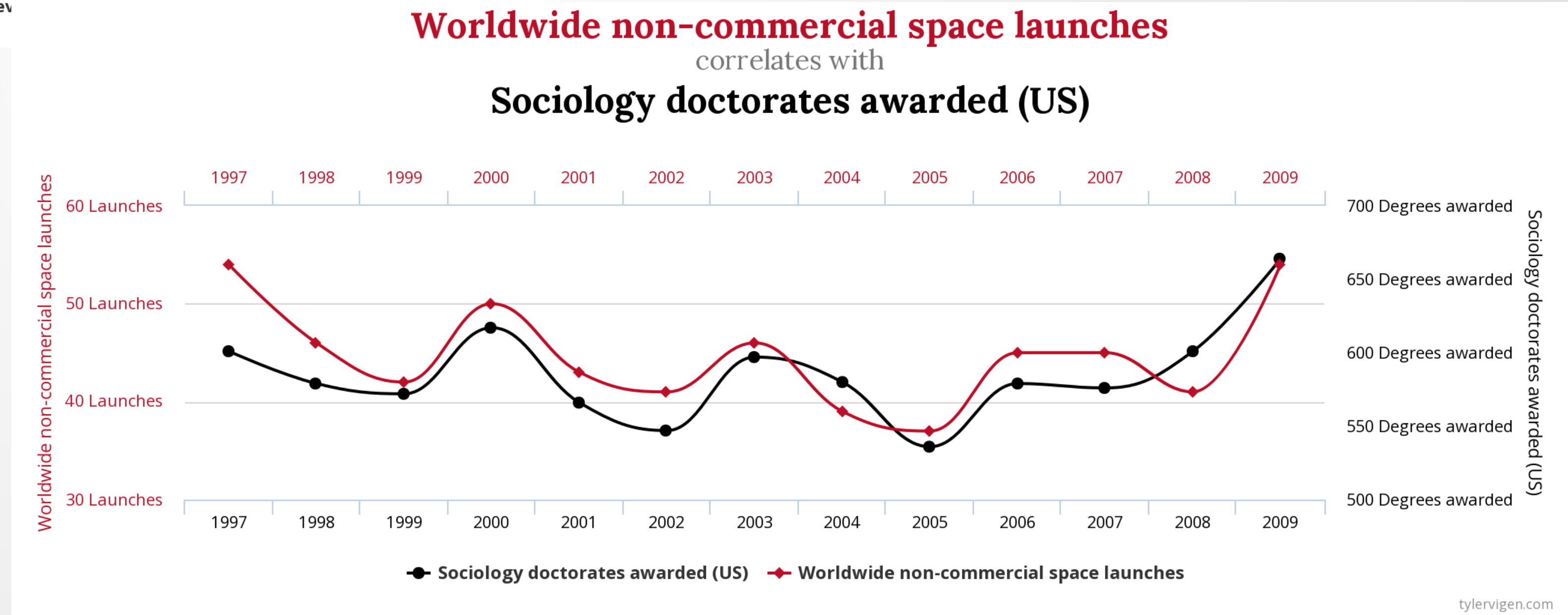
In the era of big data you may encounter truly spurious correlations

divorce rate in Maine / consumption of Margarine

http://www.tylervigen.com/spurious-correlations

correlation





http://www.tylervigen.com/spurious-correlations

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x}\right) \left(\frac{y_i - \bar{y}}{s_y}\right)$$

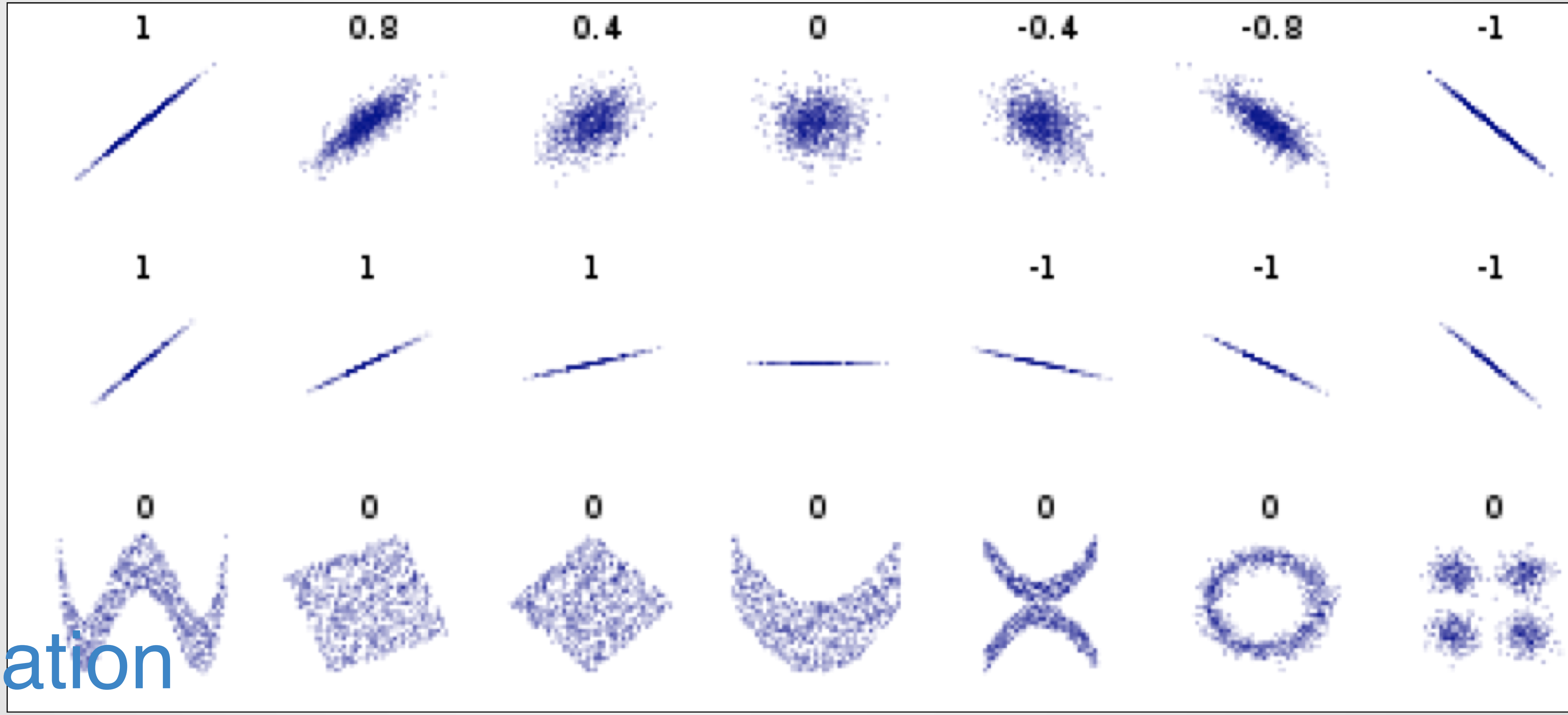
$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Pearson's correlation measures *linear* correlation



correlation

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x}\right) \left(\frac{y_i - \bar{y}}{s_y}\right)$$

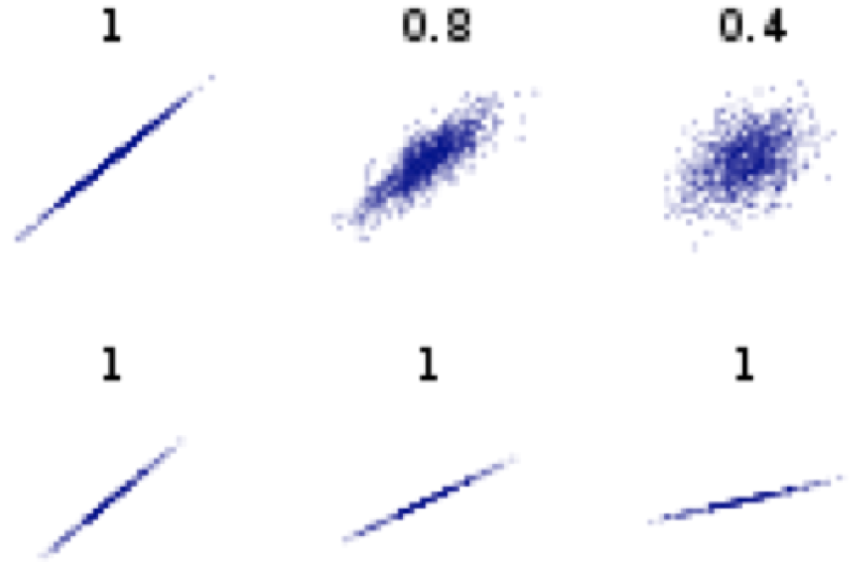
$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Pearson's correlation measures *linear* correlation



correlated

"positively" correlated

correlation

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

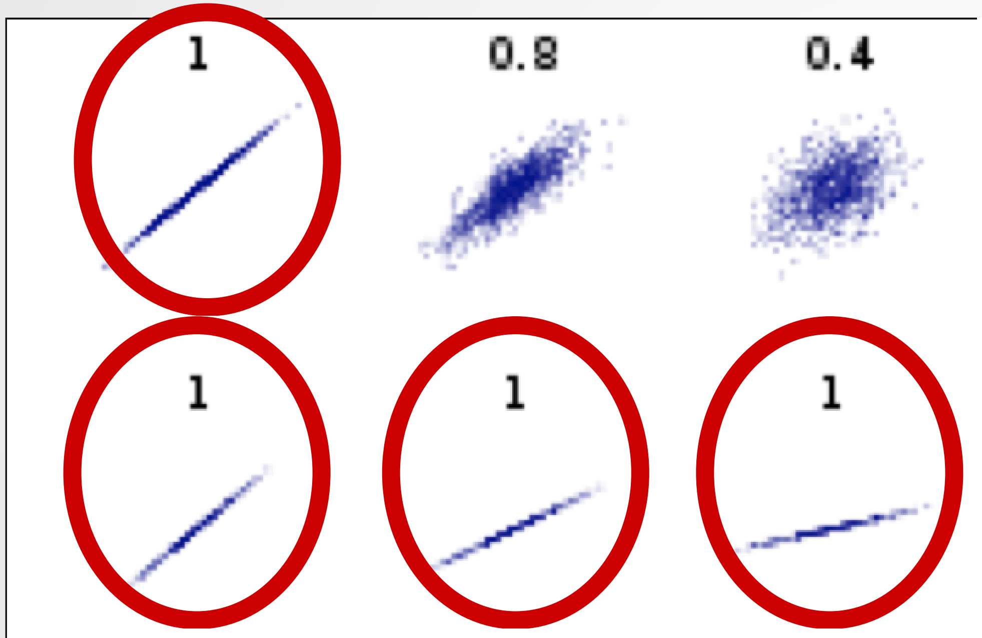
Pearson's correlation measures *linear* correlation

$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$



correlated

"positively" correlated

$$r_{xy} = 1 \text{ iff } y = ax$$

maximally correlated

correlation

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

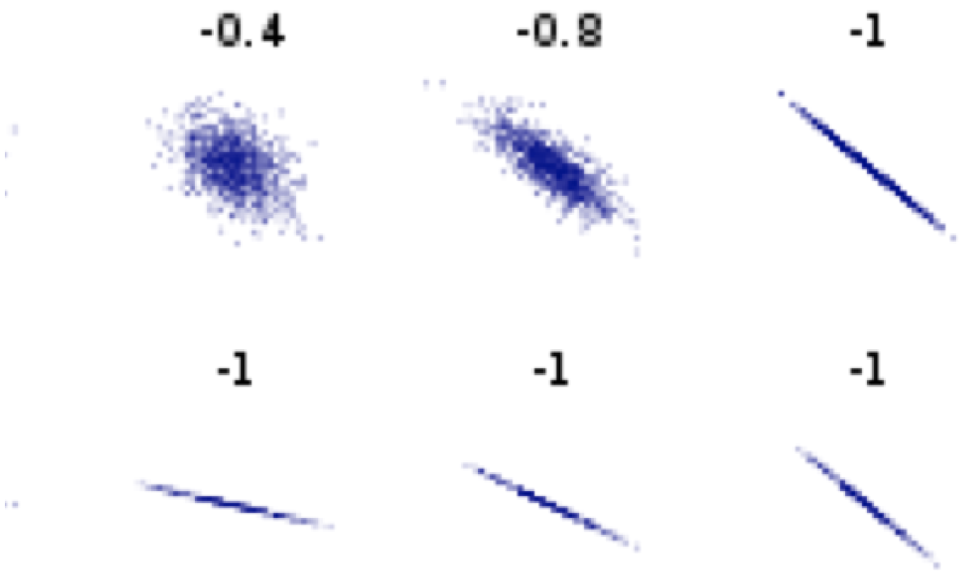
n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Pearson's correlation measures *linear* correlation

anticorrelated

"negatively" correlated



correlation

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x}\right) \left(\frac{y_i - \bar{y}}{s_y}\right)$$

$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

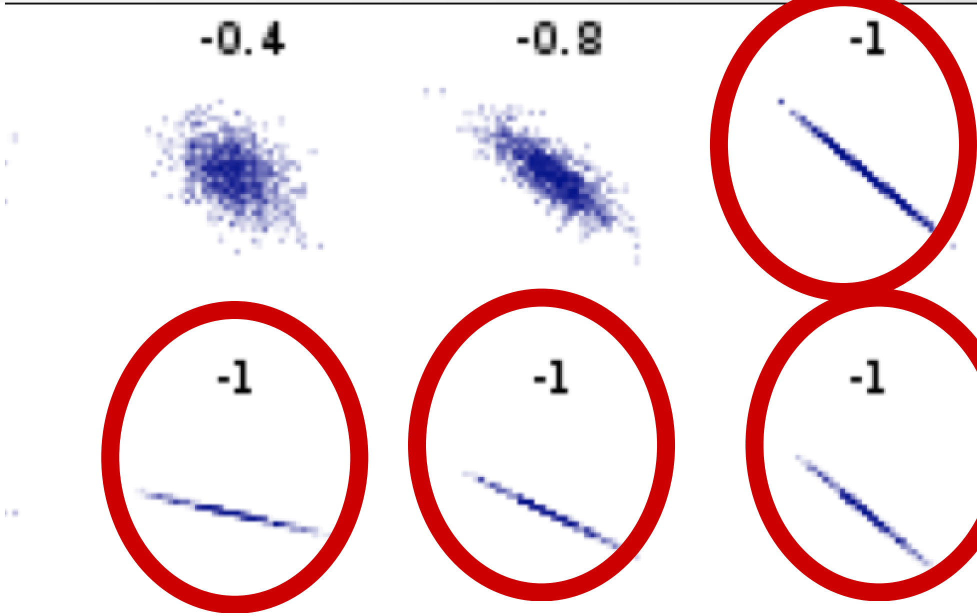
Pearson's correlation measures *linear* correlation

anticorrelated

"negatively" correlated

$$r_{xy} = 1 \text{ iff } y = -ax$$

maximally anticorrelated



correlation

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x}\right) \left(\frac{y_i - \bar{y}}{s_y}\right)$$

$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Pearson's correlation measures *linear* correlation

not *linearly* correlated

Pearson's coefficient = 0

does not mean that *x* and *y* are independent!

correlation

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Spearman's test

(Pearson's for ranked values)

$$\rho_{xy} = 1 - \frac{6 \sum_{i=1}^{N} (x_i - y_i)^2}{n(n^2 - 1)}$$

How many dichotomous[+] (binary) variables?				
0	Both variables interval or ratio?			
	Y	Measures are linear? (No = monotonic[*])		
		Y	Pearson correlation	
		N	Spearman correlation	
	N	Both variables are ordinal?		
		Y	Kendall correlation	
		N	Both variables can be ranked?	
			Y	Kendall correlation
			N	Convert to frequency data and use Chi-square test for independence
1	Biserial Correlation Coefficient			
2	2 x 2 table?			
	Y	Phi		
	N	Cramer's V		

Save the figure

correlation

Pearson's correlation

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{s_x}\right) \left(\frac{y_i - \bar{y}}{s_y}\right)$$

$\bar{x}$: mean value of x

$\bar{y}$: mean value of y

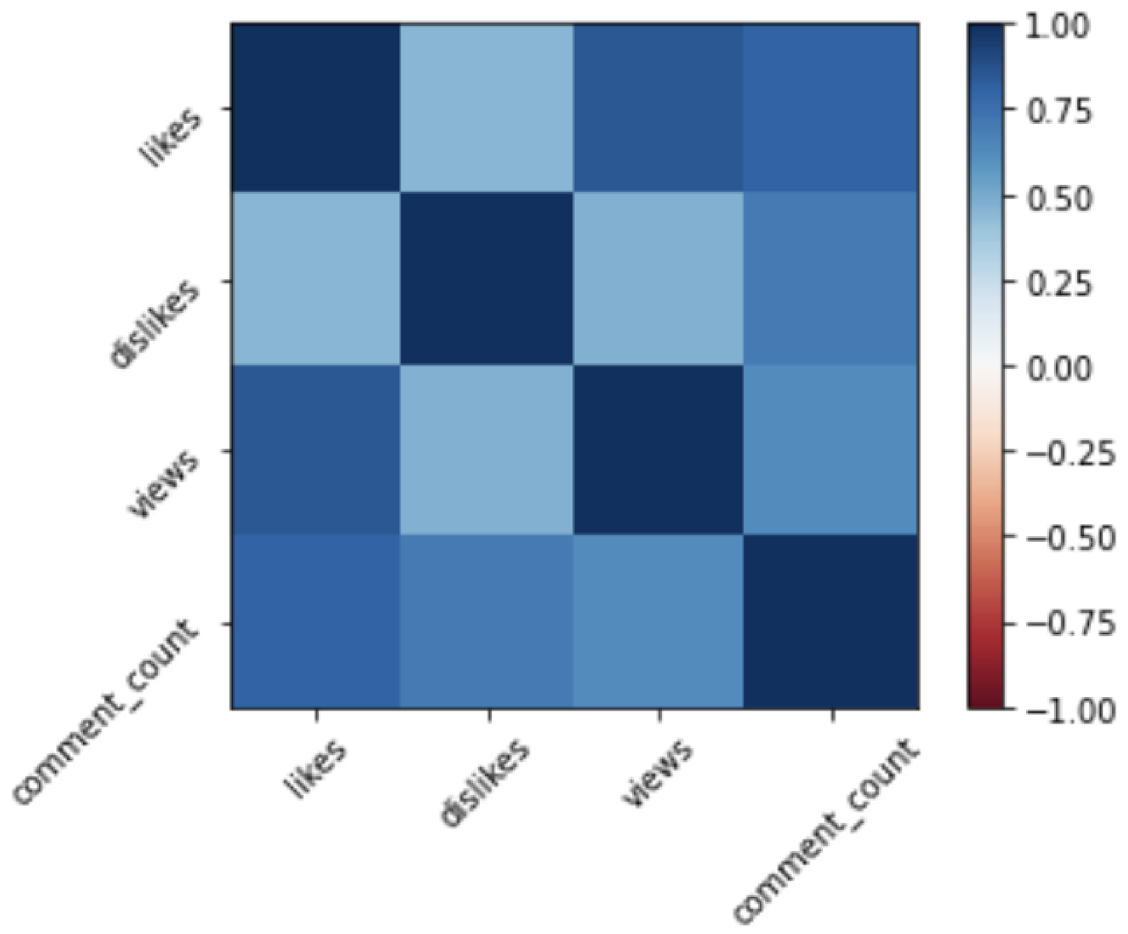
n : number of datapoints

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

```python
import pandas as pd
df = pd.read_csv(file_name)
df.corr()
```

	likes	dislikes	views	comment_count
likes	1.000000	0.447186	0.849177	0.803057
dislikes	0.447186	1.000000	0.472213	0.700184
views	0.849177	0.472213	1.000000	0.617621
comment_count	0.803057	0.700184	0.617621	1.000000

```python
pl.imshow(vdf.corr(), clim=(-1,1),  cmap='RdBu')
pl.xticks(list(range(len(df.corr()))),
                    df.columns, rotation=45)
pl.yticks(list(range(len(df.corr()))),
                    df.columns, rotation=45)
pl.colorbar();
```

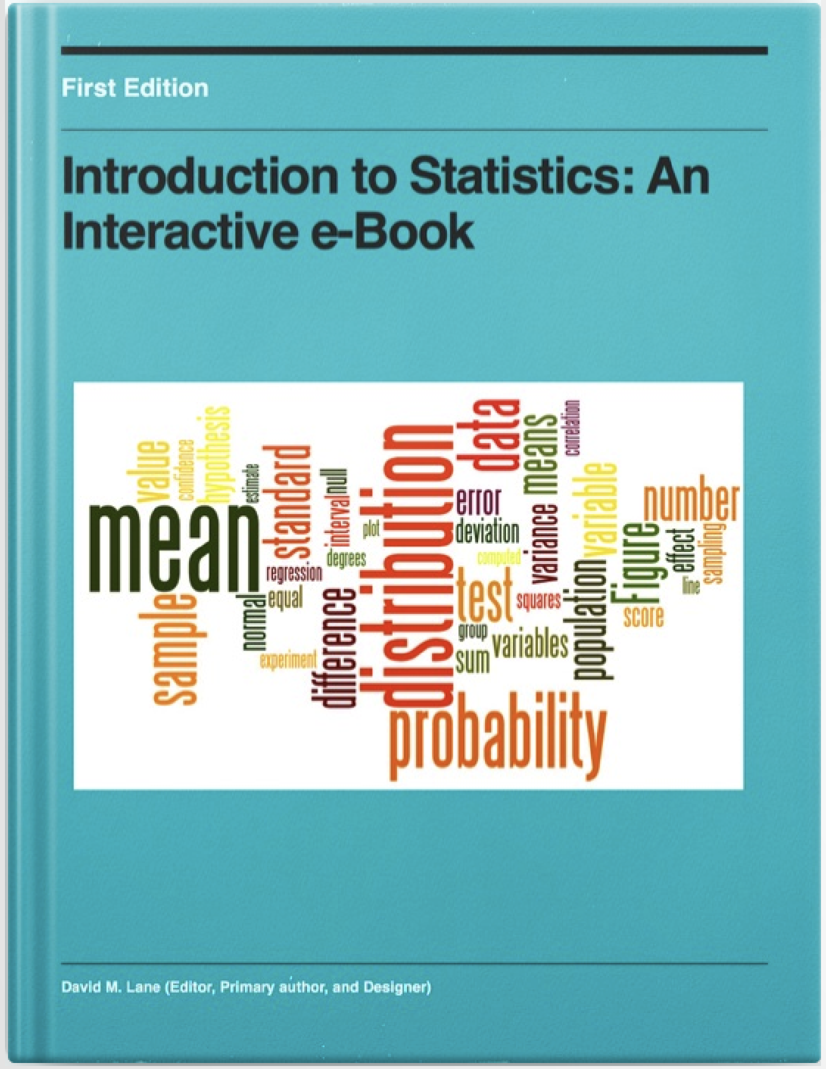


<- anticorrelated | correlated ->

correlation

21

probability



Introduction to Statistics: An Interactive e-Book

David M. Lane

Crush Course in Statistics

freee statistics book: http://onlinestatbook.com/

what are probability and statistics?

Basic Probability ***Frequentist*** interpretation

fraction of times something happens

$\Updownarrow$

probability of it happening

Basic Probability Bayesian interpretation

represents a level of certainty relating to a potential outcome or idea:

if I believe the coin is unfair (tricked) then even if I get a head and a tail I will still believe I am more likely to get heads than tails

Basic Probability
Frequentist interpretation

$$P(E) = \text{frequency of } E$$

$$P(\text{coin} = \text{head}) = 6/11 = 0.55$$

fraction of times something happens

$\Updownarrow$

probability of it happening

Basic Probability ***Frequentist*** interpretation

$$P(E) = \text{frequency of } E$$

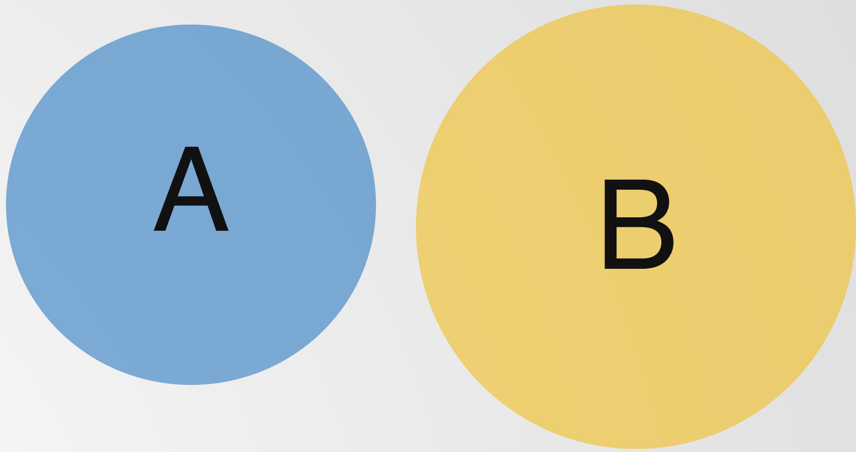
$$P(\text{coin} = \text{head}) = 6/11 = 0.55$$

$$P(\text{coin} = \text{head}) = 51/101 = 0.504$$

fraction of times something happens

$\Updownarrow$

probability of it happening



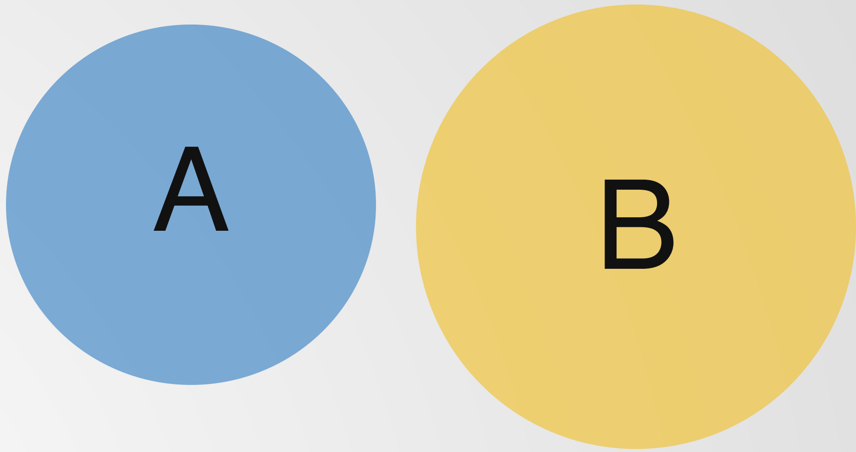
Basic probability arithmetics

Probability Arithmetic

Rules:

$$0 <= \; P(A) \; <= 1$$

Basic probability arithmetics



Probability Arithmetic

Rules:

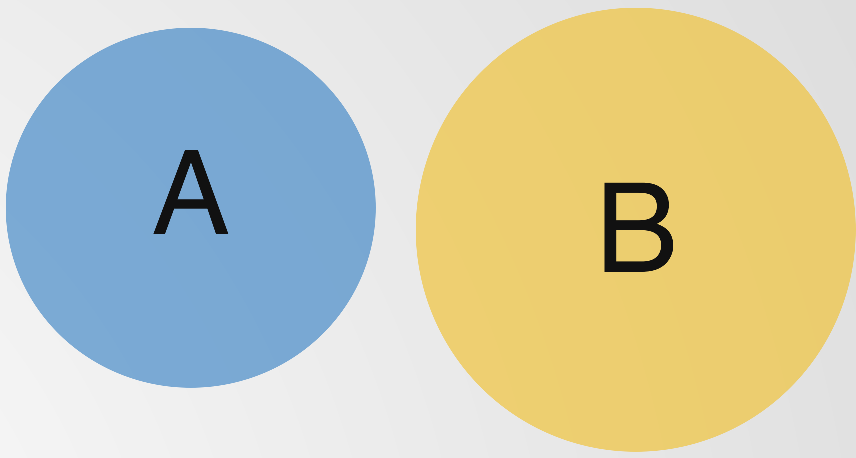
$$0 <= \ P(A) \ <= 1$$

if $\bar{A}$ is the complement of A

(everything but A)

$$P(A) + P(\bar{A}) = 1$$

Basic probability arithmetics



Probability Arithmetic

disjoint events:

$$\text{if}\, P(A) \cap P(B) = 0 \text{ then :}$$

$$P(A \text{or} B) = P(A) + P(B)$$

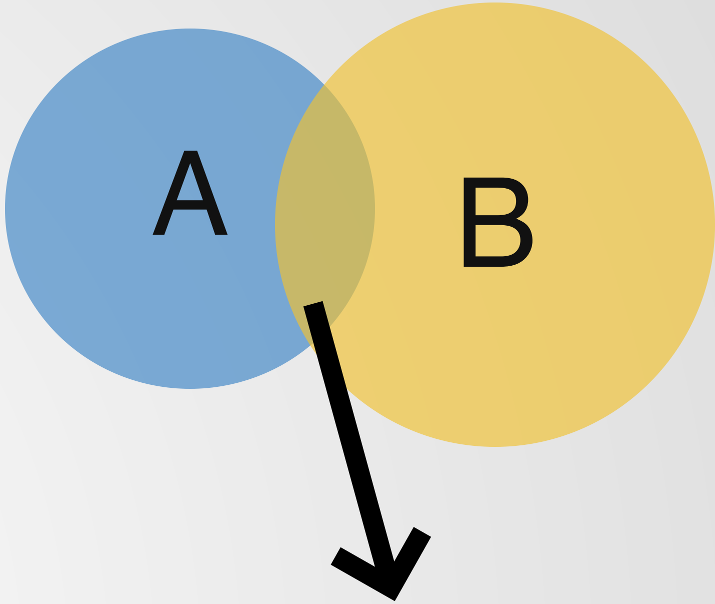
independent probabilities

$$P(A \text{and} B) = P(A) * P(B)$$

$$P(A|B) = P(A)$$

Basic probability arithmetics

Probability Arithmetic



$$P(A \cap B)$$

$$\text{if } P(A) \cap P(B) > 0 \text{ then :}$$

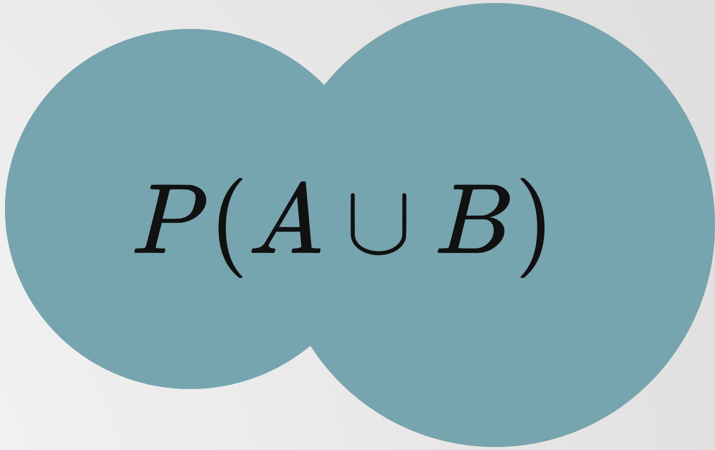
related events:

dependent probabilities

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) < P(A)$$

$$P(A \cap B) = P(A)P(B|A)$$



Basic probability arithmetics

Probability Arithmetic

$$if\, P(A) \cap P(B) > 0 \text{ then :}$$

dependent probabilities

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) < P(A)$$

$$P(A \cap B) = P(A)P(B|A)$$

22 statistics

statistics

takes us from observing a limited number of samples to infer on the population

TAXONOMY

Distribution: a formula (a model describing outcomes of measurements)

Population: all of the elements of a "family" or class

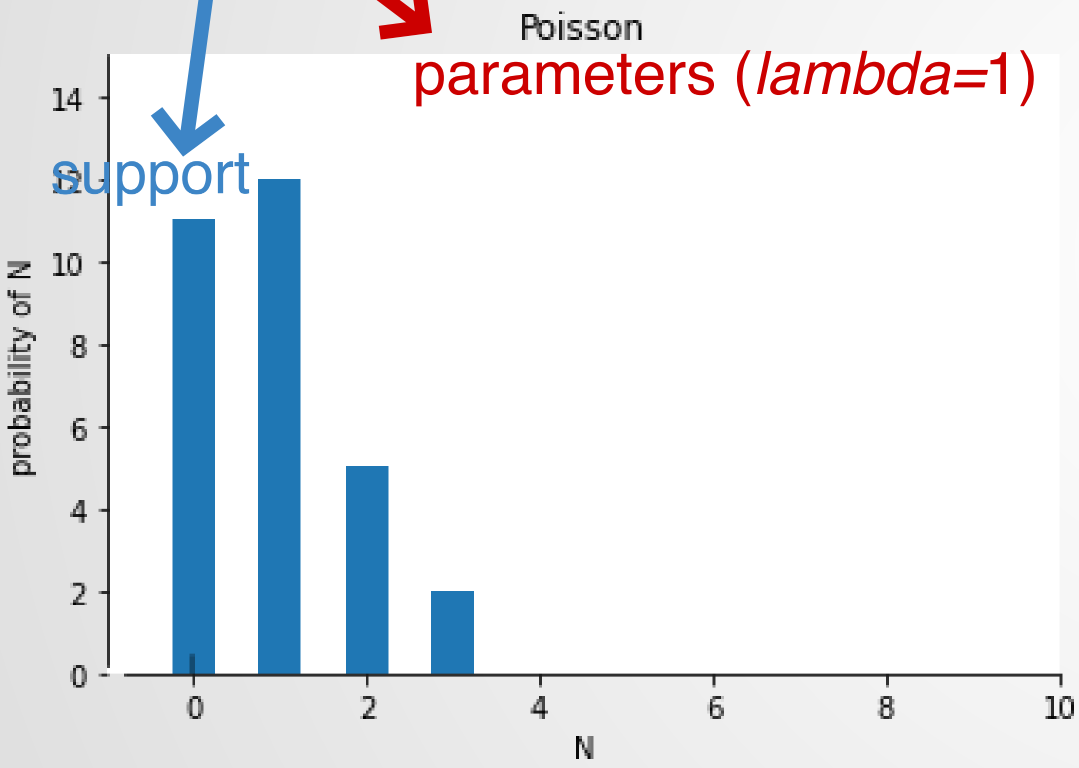
Sample: a finite subset of the population that you observe

coding time!

https://colab.research.google.com/

distributions

$$P(k|\lambda) \sim \frac{\lambda^k \, e^{-\lambda}}{!k}$$



A ***distribution is*** a collection of datapoints whose frequency in the sample corresponds to a known formula

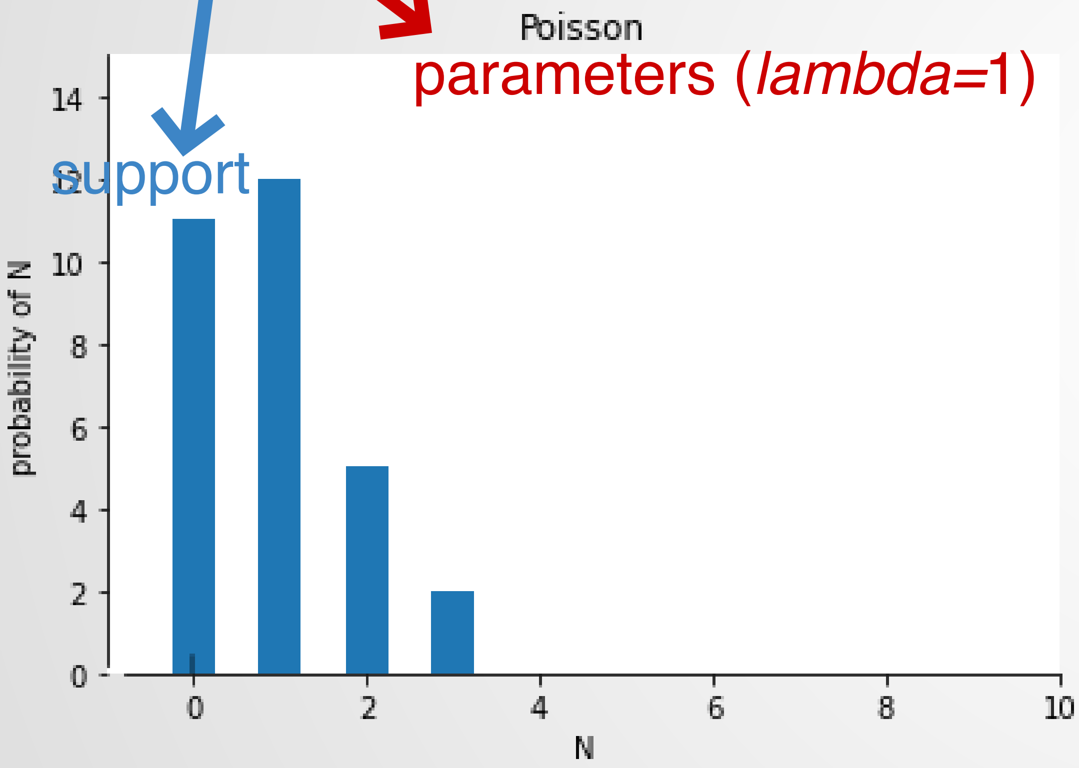
A number *k* (e.g. 1) has some probability of being drawn

The probability depents on the parameters of the distribution λ

If I draw *N* numbers and plot a histogram of them the histogram will have a specific shape

distributions

$$P(k|\lambda) \sim \frac{\lambda^k\, e^{-\lambda}}{!k}$$



A ***distribution is*** a collection of datapoints whose frequency in the sample corresponds to a known formula

A number *k* (e.g. 1) has some probability of being drawn

The probability depents on the parameters of the distribution λ

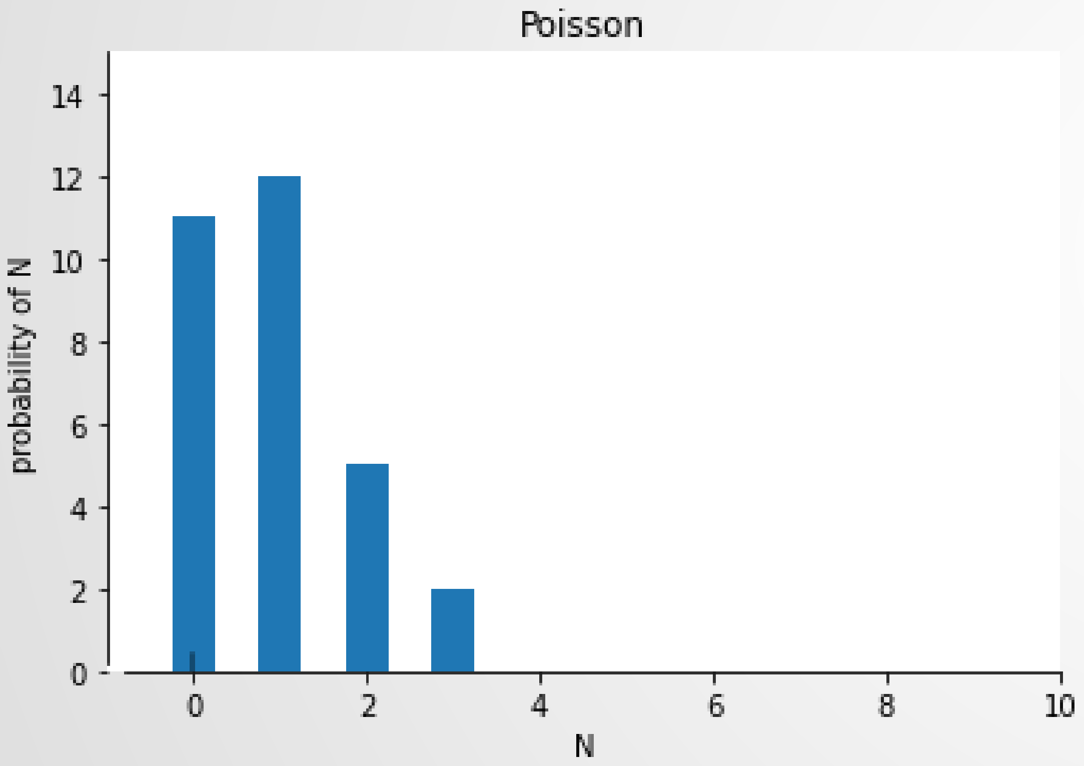
If I draw *N* numbers and plot a histogram of them the histogram will have a specific shape

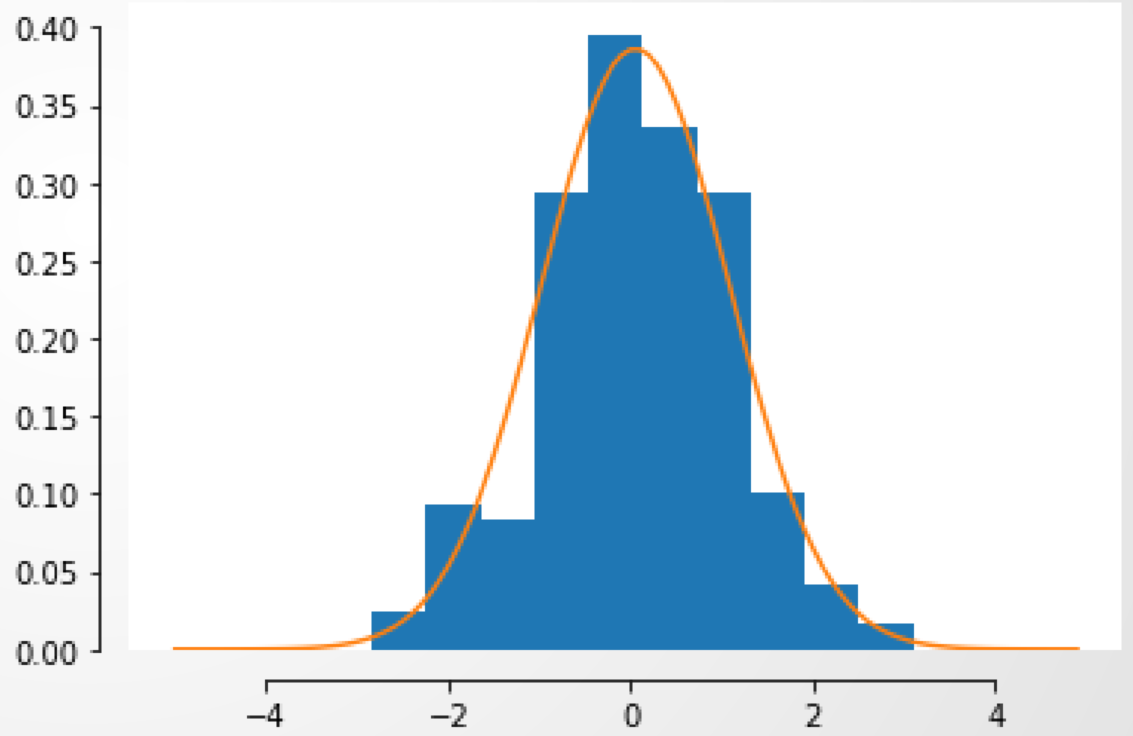
distributions

support parameters (-0.1, 0.9)

$$P(k|\lambda) \sim \frac{\lambda^k \, e^{-\lambda}}{!k}$$

$$N(r|\mu,\sigma) \sim \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r-\mu)^2}{2\sigma^2}}$$





Poisson

discrete support (only ints) $(1, +\inf]$

normal or Gaussian

continuous support $[-\inf, +\inf]$

Moments and frequentist probability

a distribution's moments summarize its properties:

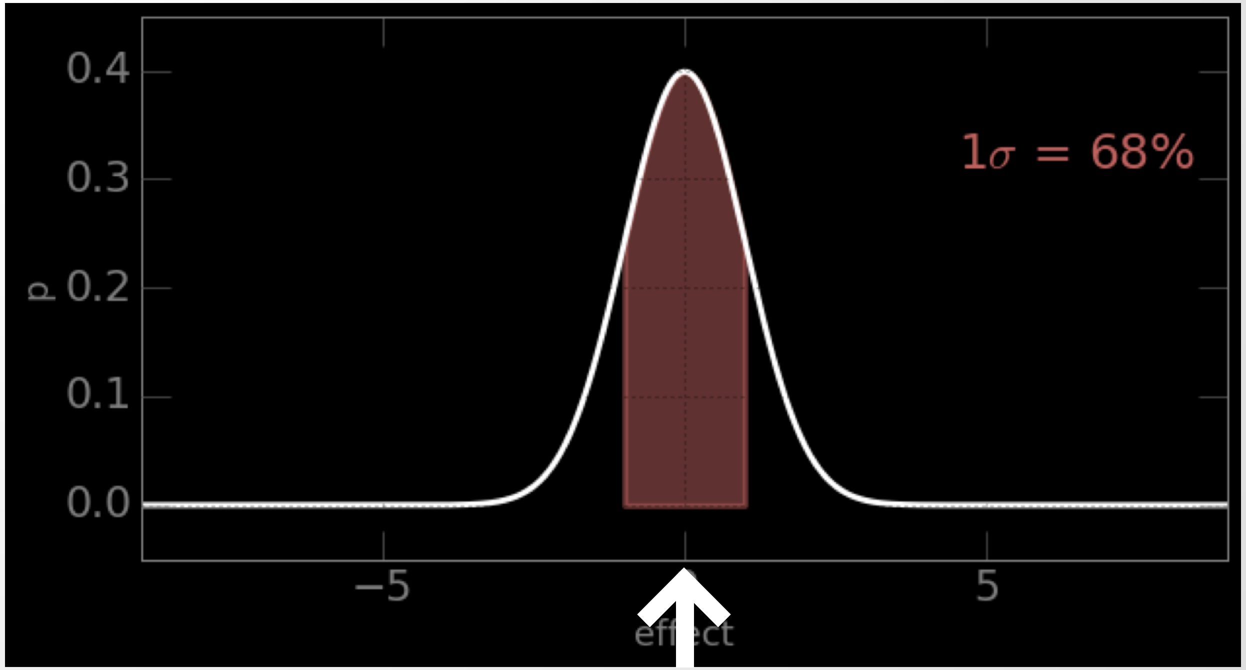
$$m_n = \int_{-\inf}^{\inf} (x - c)^n f(X) dx$$

central tendency: mean (n=1), median, mode

spread: standard deviation/variance (n=2), quartiles range

symmetry: skewness (n=3)

cuspiness: kurtosis (n=4)



Moments and frequentist probability

a distribution's moments summarize its properties:

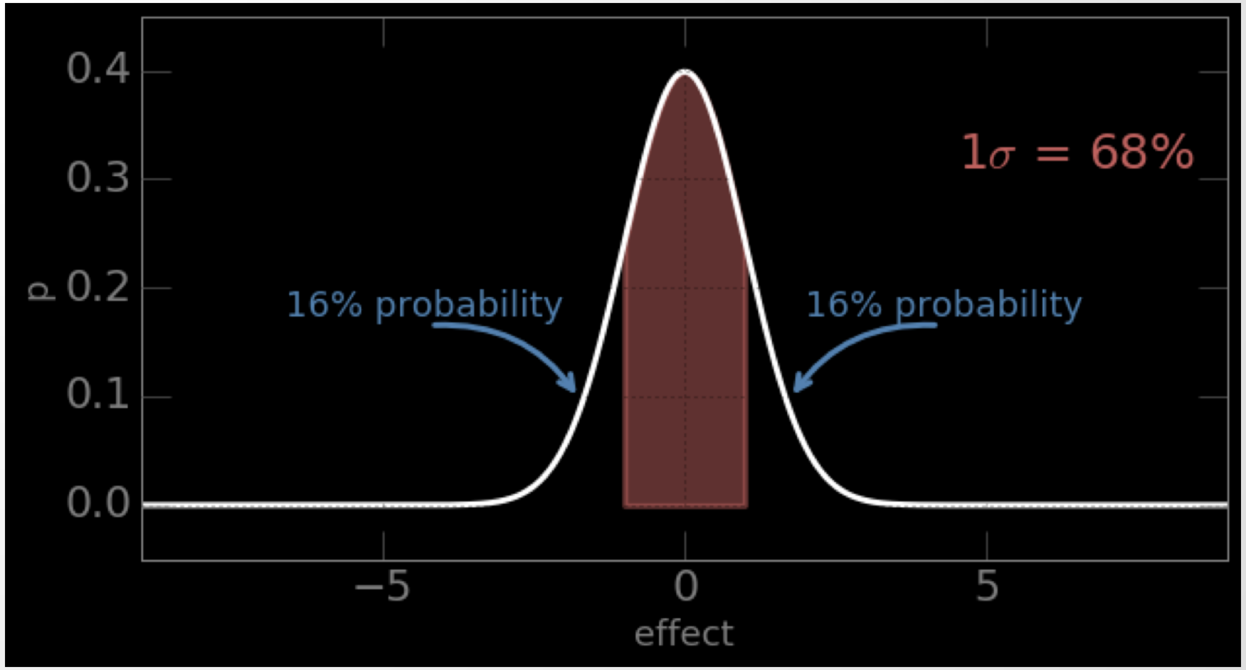
$$m_n = \int_{-\inf}^{\inf} (x - c)^n f(X) dx$$

central tendency: mean (n=1), median, mode

spread: standard deviation/variance (n=2), quartiles range

symmetry: skewness (n=3)

cuspiness: kurtosis (n=4)



Moments and frequentist probability

a distribution's moments summarize its properties:

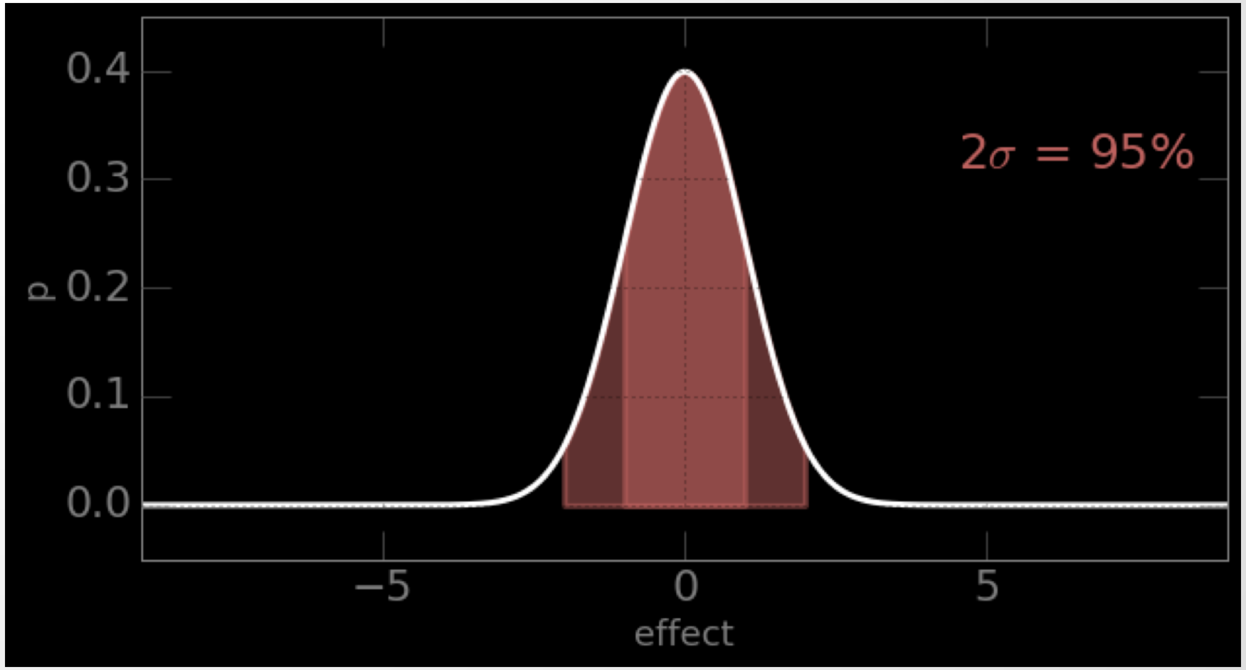
$$m_n = \int_{-\inf}^{\inf} (x - c)^n f(X) dx$$

central tendency: mean (n=1), median, mode

spread: standard deviation/variance (n=2), quartiles range

symmetry: skewness (n=3)

cuspiness: kurtosis (n=4)



Moments and frequentist probability

a distribution's moments summarize its properties:

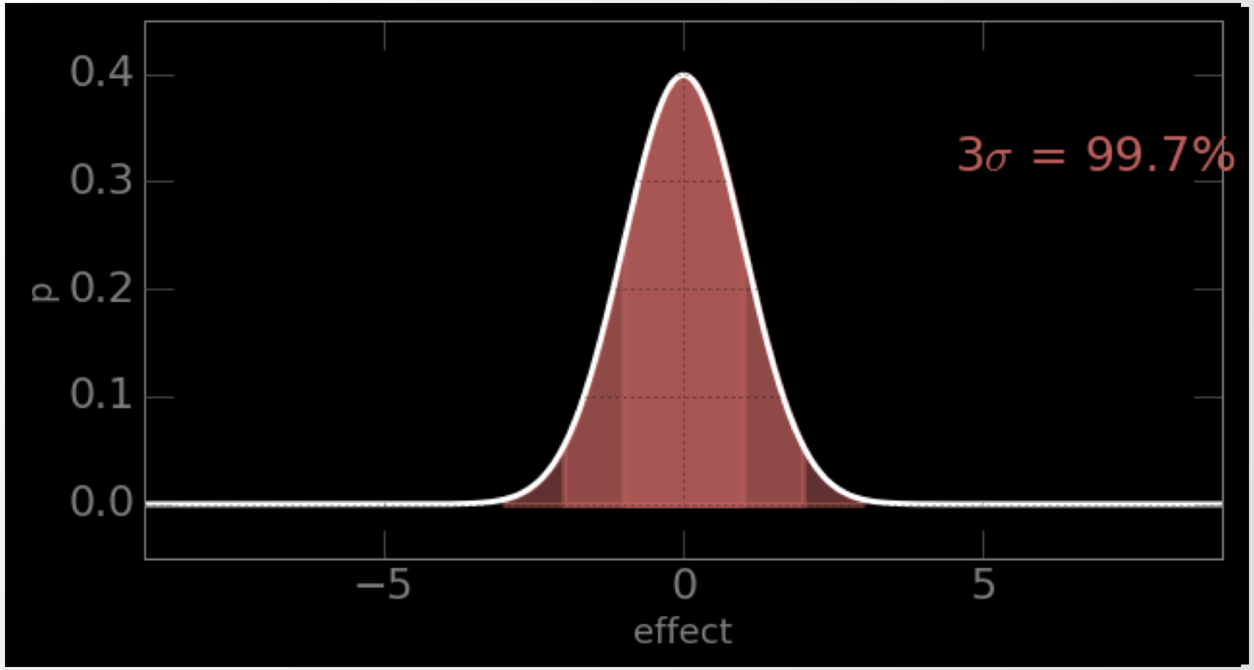
$$m_n = \int_{-\inf}^{\inf} (x-c)^n f(X) dx$$

central tendency: mean (n=1), median, mode

spread: standard deviation/variance (n=2), quartiles range

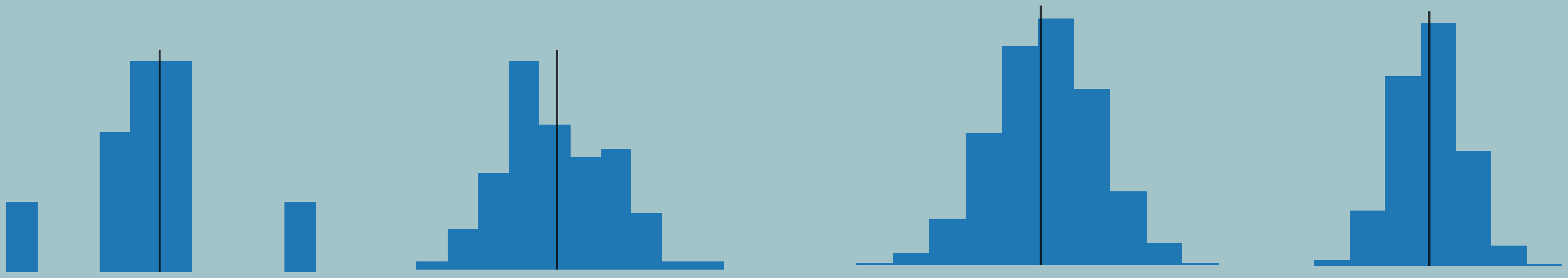
symmetry: skewness (n=3)

cuspiness: kurtosis (n=4)



Law of Large Numbers

As the size of a _______________ tends to infinity the mean of the sample tends to the mean of the _________________



Central Limit Theorem

HW extra credits

Laplace (1700s) but also: Poisson, Bessel, Dirichlet, Cauchy, Ellis

Let $x_1 \ldots x_N$ be an N-elements sample from a population whose distribution has

mean **μ** and standard deviation **σ**

In the limit of N -> infinity

the sample mean $\bar{x}$ approaches a Normal (Gaussian) distribution with mean μ and standard deviation σ

regardless of the distribution of X

$$\bar{x} \sim N\left(\mu, \sigma/\sqrt{N}\right)$$

Probability distributions

Binomial

I bet heads:

head = success

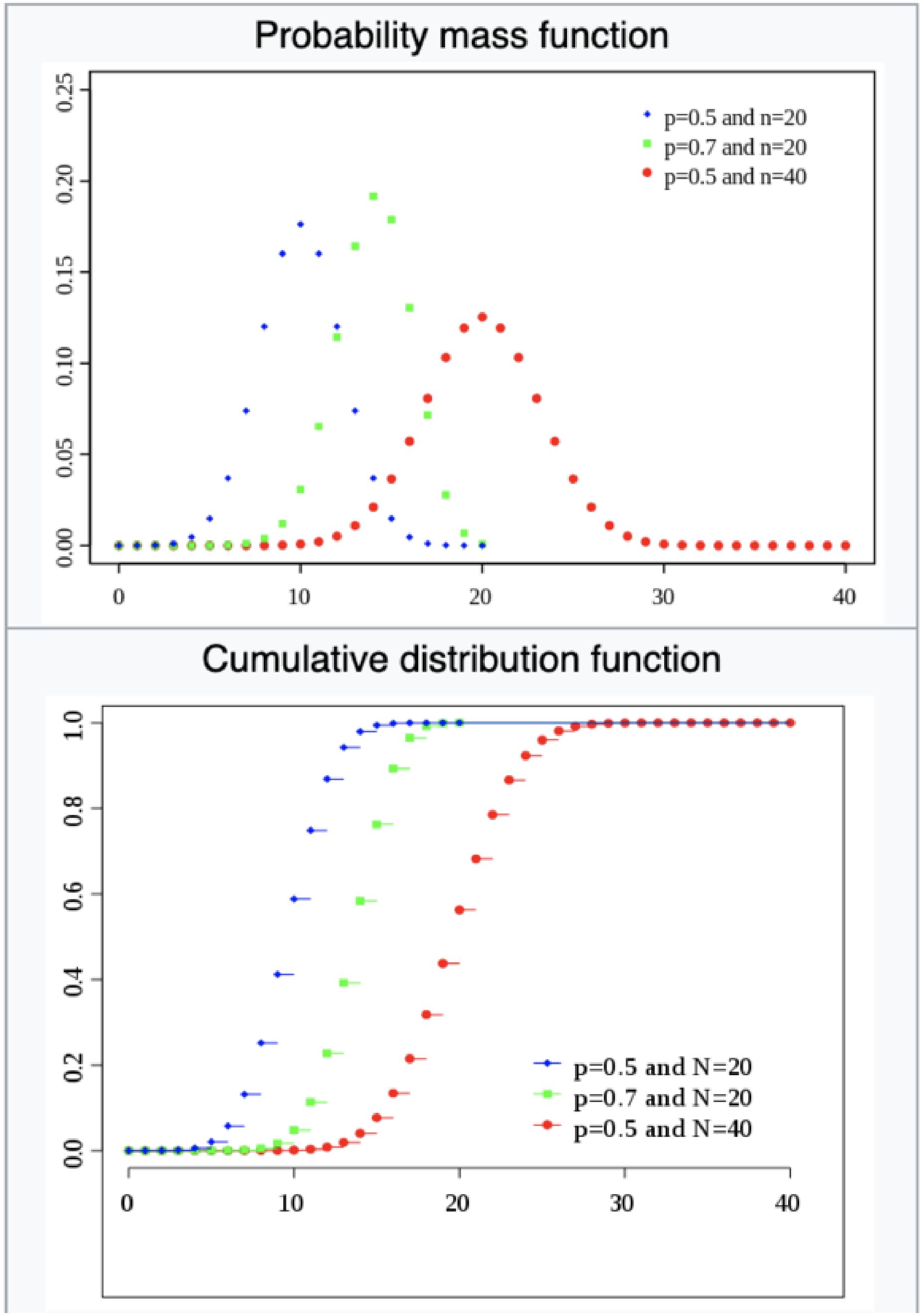
"given n tosses, each with a probability of 0.5 to get head"

Coin toss:

fair coin: p=0.5 n=1

Vegas coin: p≠0.5 n=1

Binomial distribution



Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi enp(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

Binomial

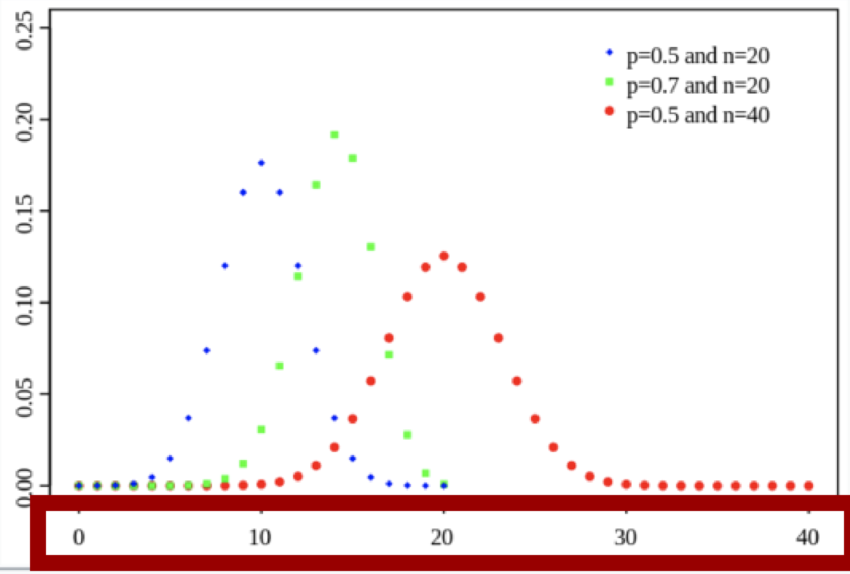
Coin toss:

fair coin: p=0.5 n=1

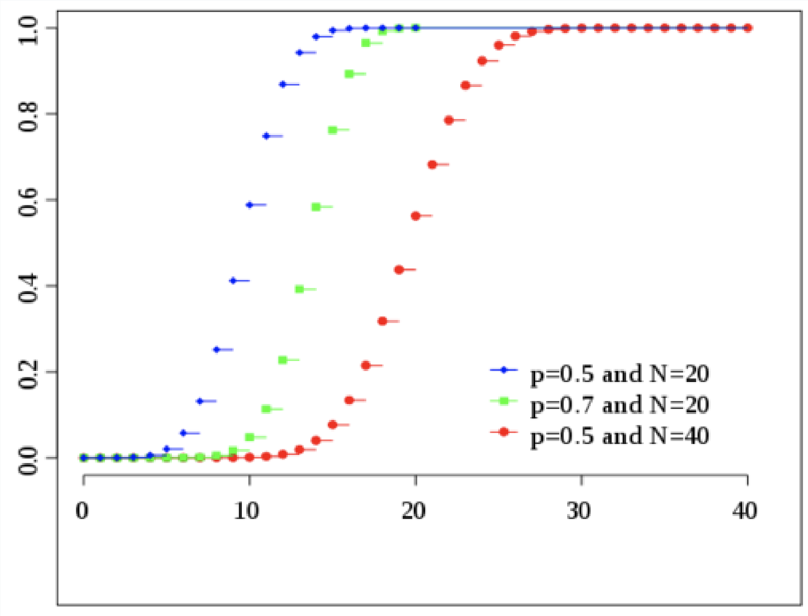
Vegas coin: p≠0.5 n=1

Binomial distribution

Probability mass function



Cumulative distribution function



Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi e np(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

Binomial

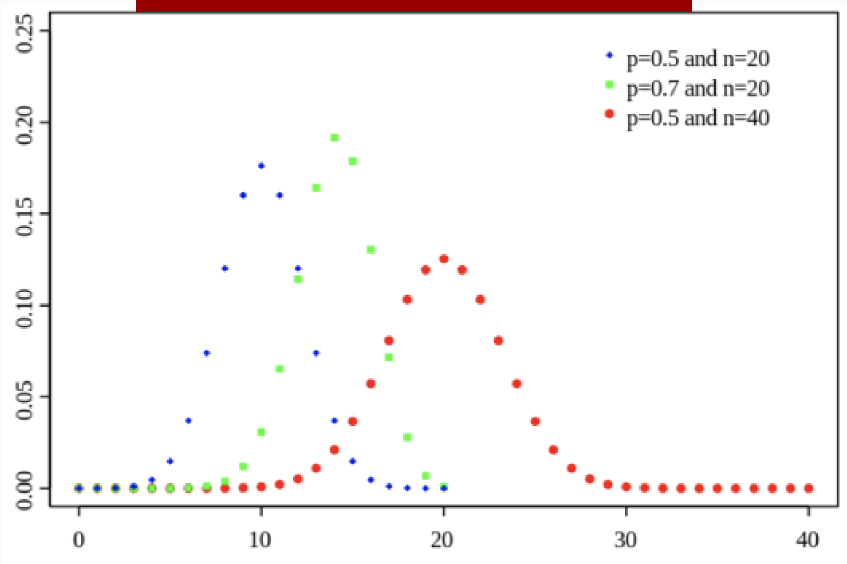
Coin toss:

fair coin: p=0.5 n=1

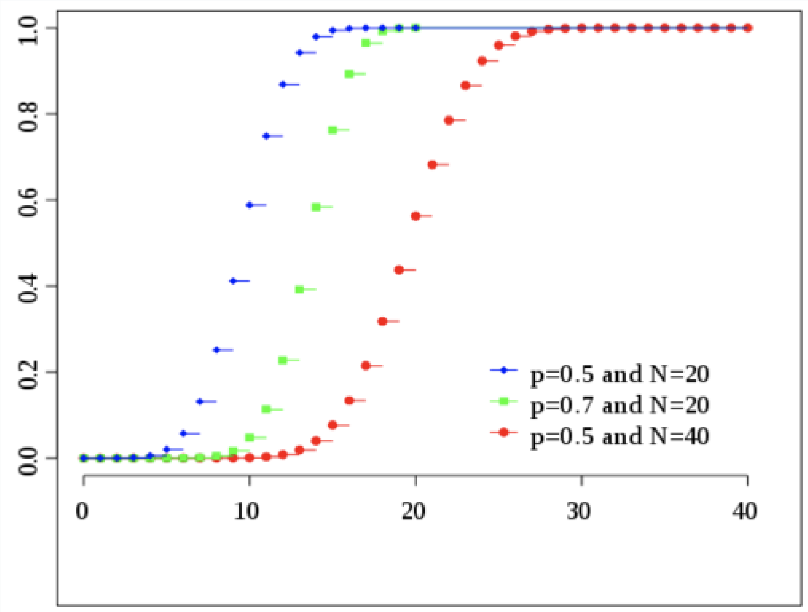
Vegas coin: p≠0.5 n=1

Binomial distribution

Probability mass function



Cumulative distribution function



Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi e np(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

Binomial

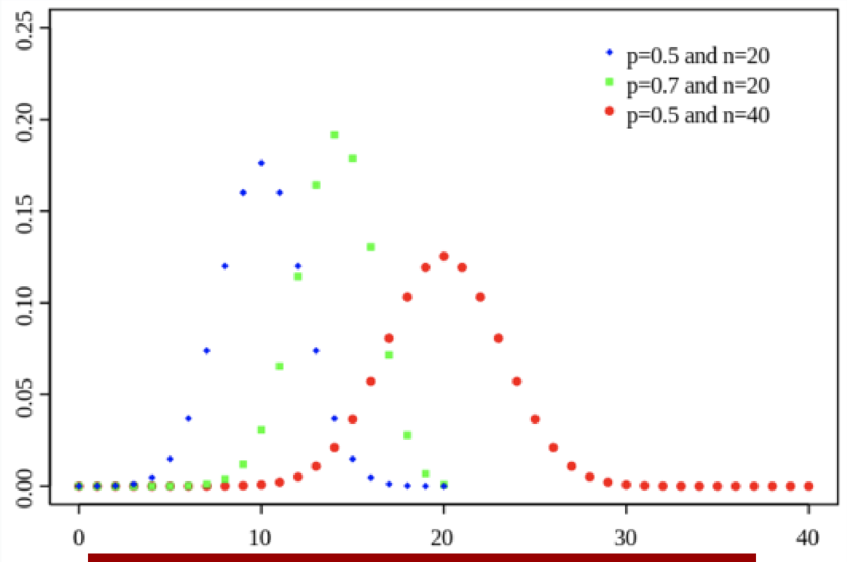
Coin toss:

fair coin: p=0.5 n=1

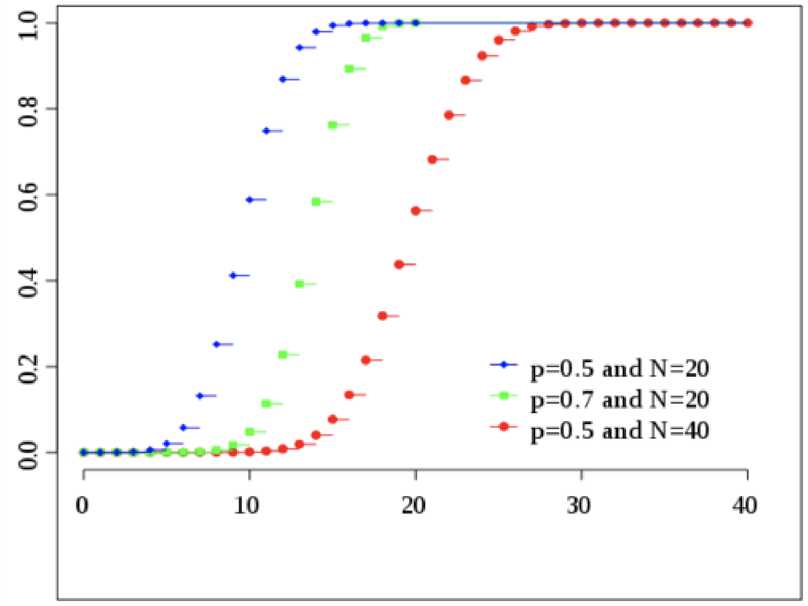
Vegas coin: p≠0.5 n=1

Binomial distribution

Probability mass function



Cumulative distribution function



Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi e n p(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

Binomial

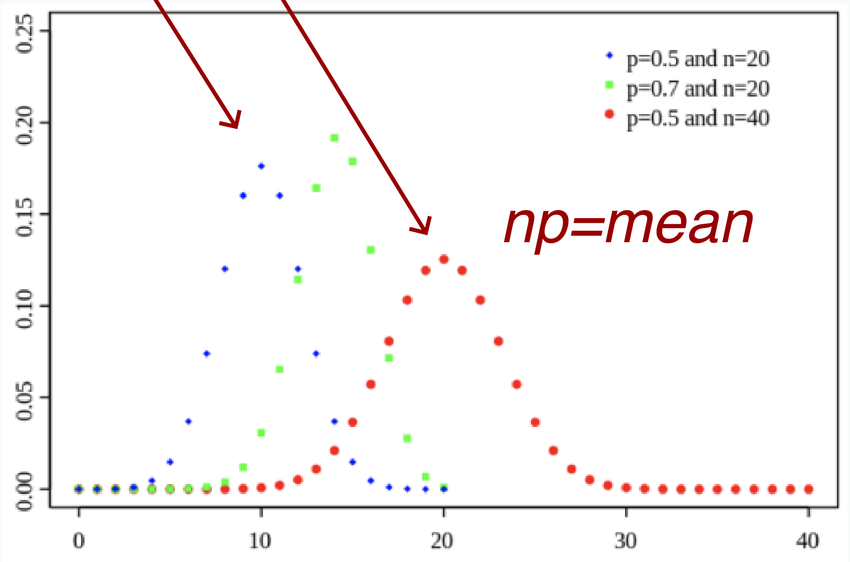
Coin toss:

fair coin: p=0.5 n=1

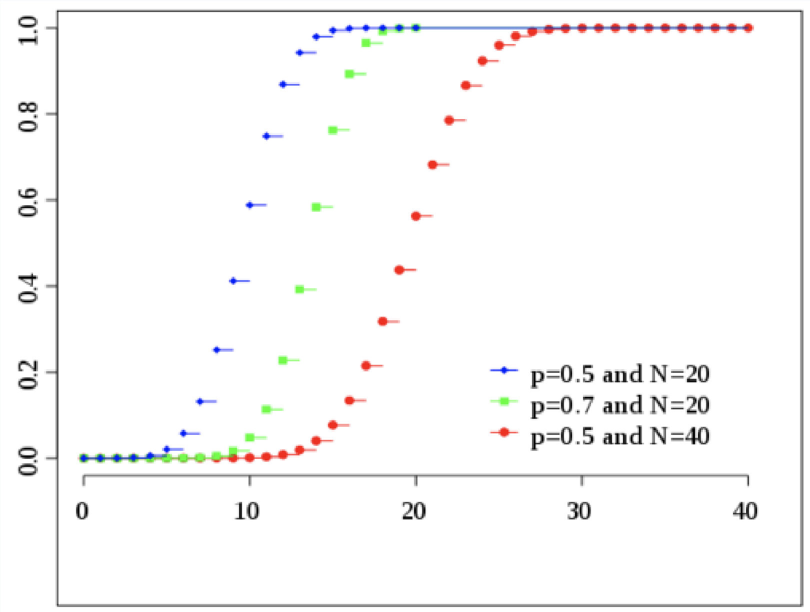
Vegas coin: p≠0.5 n=1

Binomial distribution

Probability mass function



Cumulative distribution function



central tendency

Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi e n p(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Probability distributions

Binomial

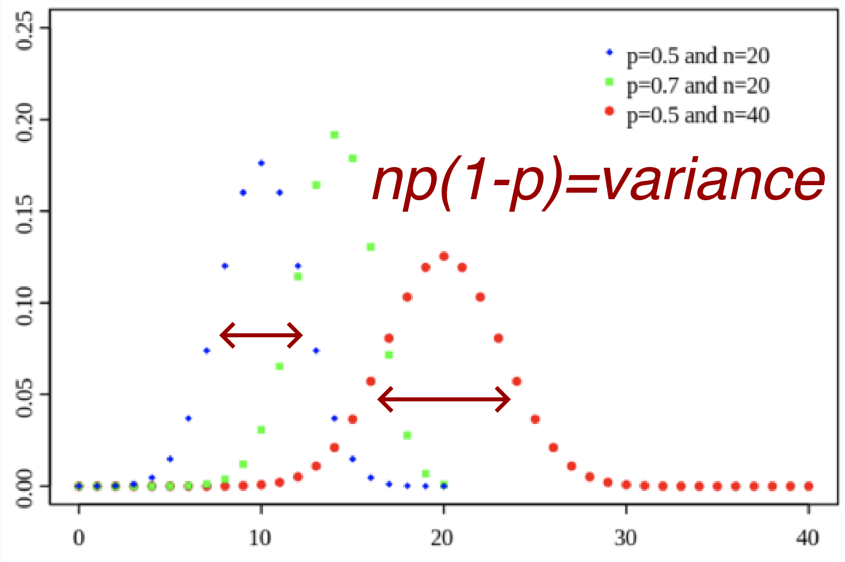
Coin toss:

fair coin: p=0.5 n=1

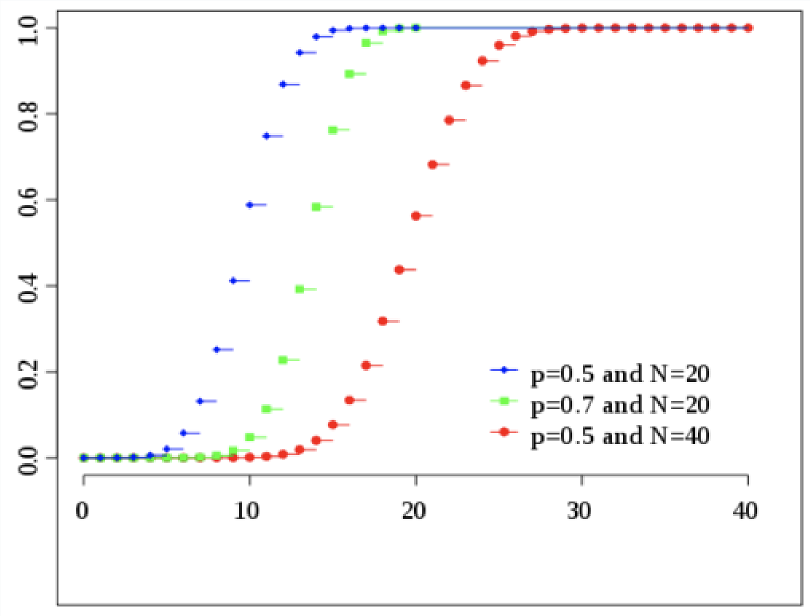
Vegas coin: p≠0.5 n=1

Binomial distribution

Probability mass function



Cumulative distribution function



Notation	$B(n,p)$
Parameters	$n \in \{0,1,2,\ldots\}$ – number of trials $p \in [0,1]$ – success probability for each trial
Support	$k \in \{0,1,\ldots,n\}$ – number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\dfrac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\dfrac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2(2\pi e np(1-p)) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$
Fisher information	$g_n(p) = \dfrac{n}{p(1-p)}$ (for fixed n)

Probability distributions

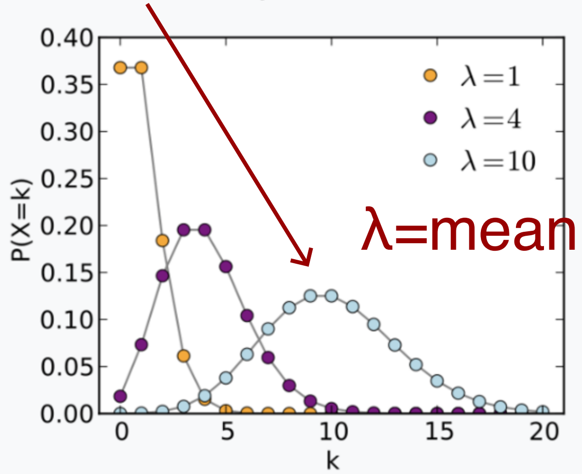
Poisson

Shut noise/count noise

The innate noise in natural steady state processes (star flux, rain drops...)

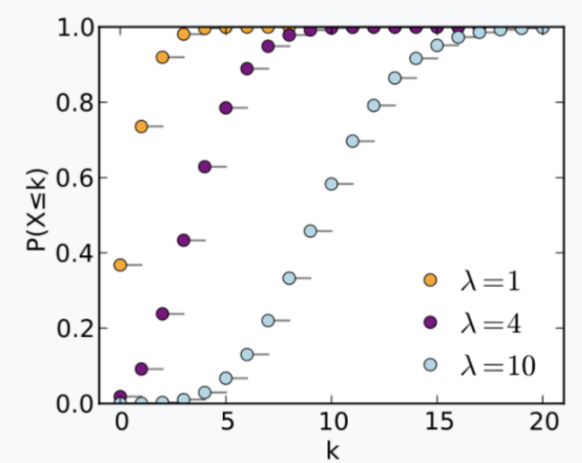
Poisson

Probability mass function



The horizontal axis is the index k, the number of occurrences. λ is the expected number of occurrences, which need not be an integer. The vertical axis is the probability of k occurrences given λ. The function is defined only at integer values of k. The connecting lines are only guides for the eye.

Cumulative distribution function



The horizontal axis is the index k, the number of occurrences. The CDF is discontinuous at the integers of k and flat everywhere else because a variable that is Poisson distributed takes on only integer values.

Notation	$\mathrm{Pois}(\lambda)$
Parameters	$\lambda > 0$, (real) — rate
Support	$k \in \{0, 1, 2, \ldots\}$
pmf	$\frac{\lambda^k e^{-\lambda}}{k!}$
CDF	$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor !}$, or $e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$, or $Q(\lfloor k+1 \rfloor, \lambda)$ (for $k \geq 0$, where $\Gamma(x, y)$ is the upper incomplete gamma function, $\lfloor k \rfloor$ is the floor function, and Q is the regularized gamma function)
Mean	λ
Median	$\approx \lfloor \lambda + 1/3 - 0.02/\lambda \rfloor$
Mode	$\lceil \lambda \rceil - 1, \lfloor \lambda \rfloor$
Variance	λ
Skewness	$\lambda^{-1/2}$
Ex. kurtosis	λ^{-1}
Entropy	$\lambda[1 - \log(\lambda)] + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \log(k!)}{k!}$ (for large λ) $\frac{1}{2}\log(2\pi e \lambda) - \frac{1}{12\lambda} - \frac{1}{24\lambda^2} - \frac{19}{360\lambda^3} + O\left(\frac{1}{\lambda^4}\right)$
MGF	$\exp(\lambda(e^t - 1))$
CF	$\exp(\lambda(e^{it} - 1))$
PGF	$\exp(\lambda(z - 1))$
Fisher information	$\frac{1}{\lambda}$

Probability distributions

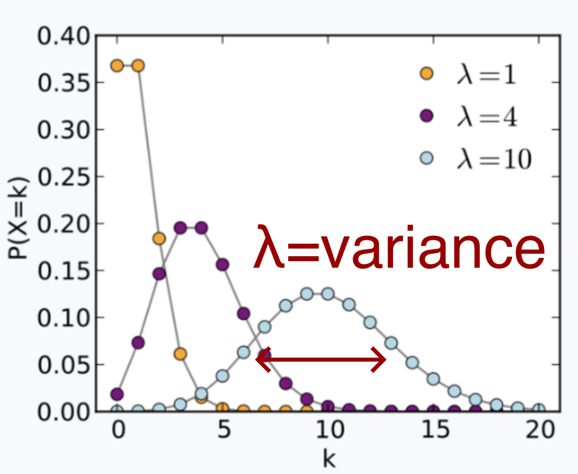
Poisson

Shut noise/count noise

The innate noise in natural steady state processes (star flux, rain drops...)

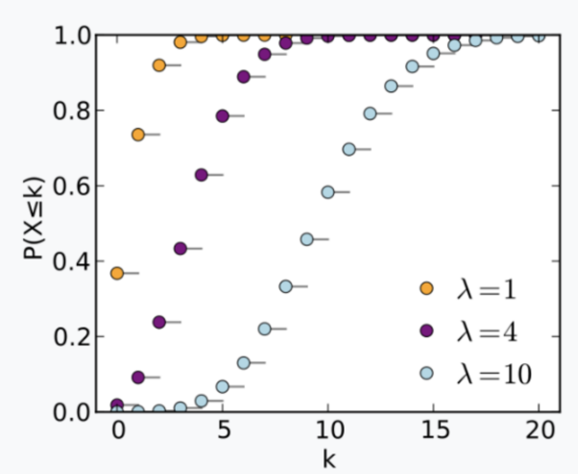
Poisson

Probability mass function



The horizontal axis is the index k, the number of occurrences. λ is the expected number of occurrences, which need not be an integer. The vertical axis is the probability of k occurrences given λ. The function is defined only at integer values of k. The connecting lines are only guides for the eye.

Cumulative distribution function



The horizontal axis is the index k, the number of occurrences. The CDF is discontinuous at the integers of k and flat everywhere else because a variable that is Poisson distributed takes on only integer values.

Notation	$\mathrm{Pois}(\lambda)$
Parameters	$\lambda > 0$, (real) — rate
Support	$k \in \{0, 1, 2, \ldots\}$
pmf	$\frac{\lambda^k e^{-\lambda}}{k!}$
CDF	$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor!}$, or $e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$, or $Q(\lfloor k+1 \rfloor, \lambda)$ (for $k \geq 0$, where $\Gamma(x, y)$ is the upper incomplete gamma function, $\lfloor k \rfloor$ is the floor function, and Q is the regularized gamma function)
Mean	λ
Median	$\approx \lfloor \lambda + 1/3 - 0.02/\lambda \rfloor$
Mode	$\lceil \lambda \rceil - 1, \lfloor \lambda \rfloor$
Variance	λ
Skewness	$\lambda^{-1/2}$
Ex. kurtosis	λ^{-1}
Entropy	$\lambda[1 - \log(\lambda)] + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \log(k!)}{k!}$ (for large λ) $\frac{1}{2}\log(2\pi e\lambda) - \frac{1}{12\lambda} - \frac{1}{24\lambda^2} - \frac{19}{360\lambda^3} + O\left(\frac{1}{\lambda^4}\right)$
MGF	$\exp(\lambda(e^t - 1))$
CF	$\exp(\lambda(e^{it} - 1))$
PGF	$\exp(\lambda(z - 1))$
Fisher information	$\frac{1}{\lambda}$

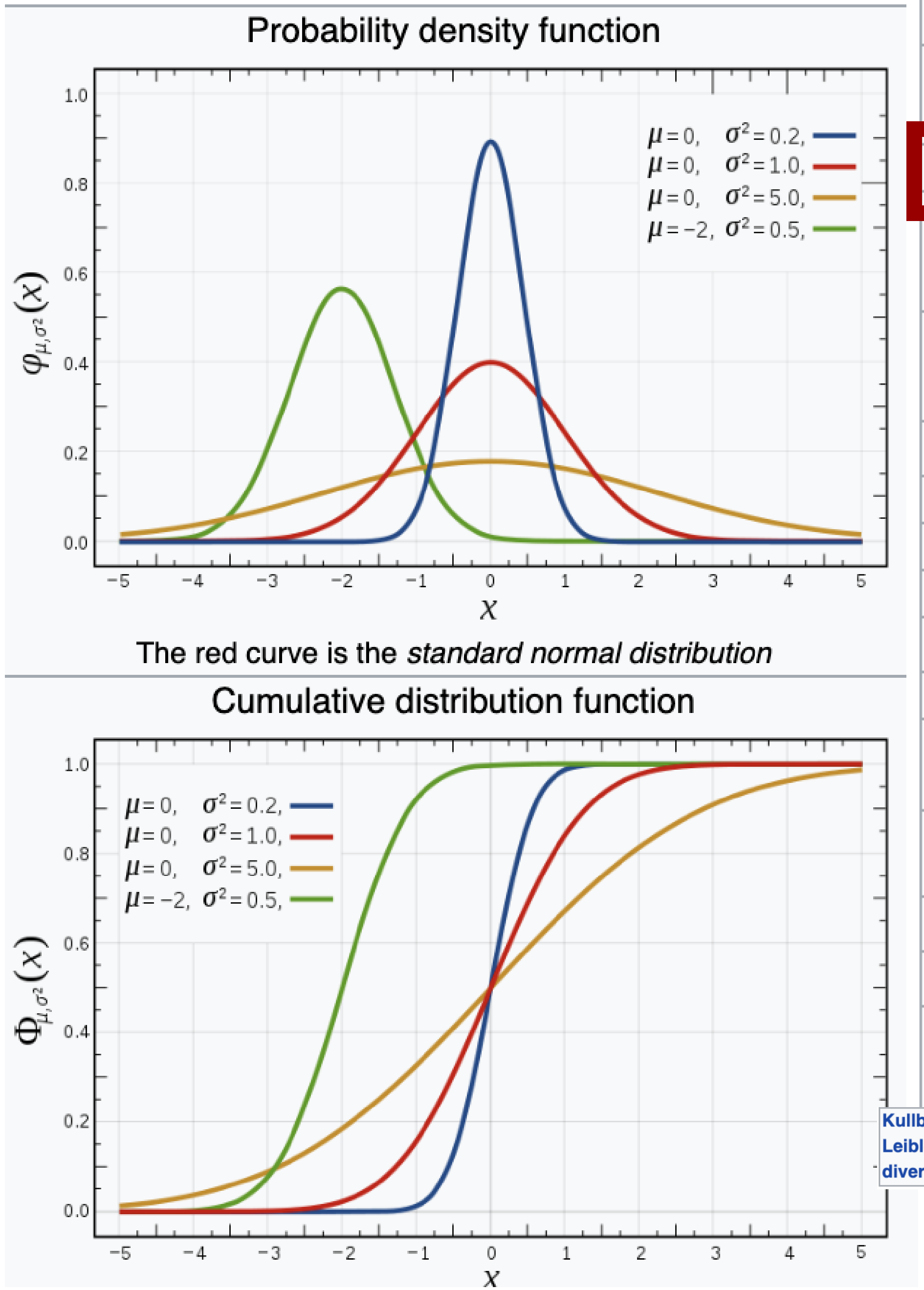
Probability distributions

Gaussian

most common noise:

well behaved mathematically, symmetric, when we can we will assume our uncertainties are Gaussian distributed

Normal Distribution



Notation	$\mathcal{N}(\mu,\sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared sc
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu+\sigma\sqrt{2}\,\operatorname{erf}^{-1}(2F-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
Skewness	0
Ex. kurtosis	0
Entropy	$\frac{1}{2}\log(2\pi e\sigma^2)$
MGF	$\exp(\mu t+\sigma^2 t^2/2)$
CF	$\exp(i\mu t-\sigma^2 t^2/2)$
Fisher information	$\mathcal{I}(\mu,\sigma)=\begin{pmatrix}1/\sigma^2 & 0\\ 0 & 2/\sigma^2\end{pmatrix}$
Kullback-Leibler divergence	$D_{\mathrm{KL}}(\mathcal{N}_0\Vert\mathcal{N}_1)=\frac{1}{2}\{(\sigma_0/\sigma_1)^2+\frac{(\mu_1-\mu_0)^2}{\sigma_1^2}-1+2\ln\frac{\sigma_1}{\sigma_0}\}$

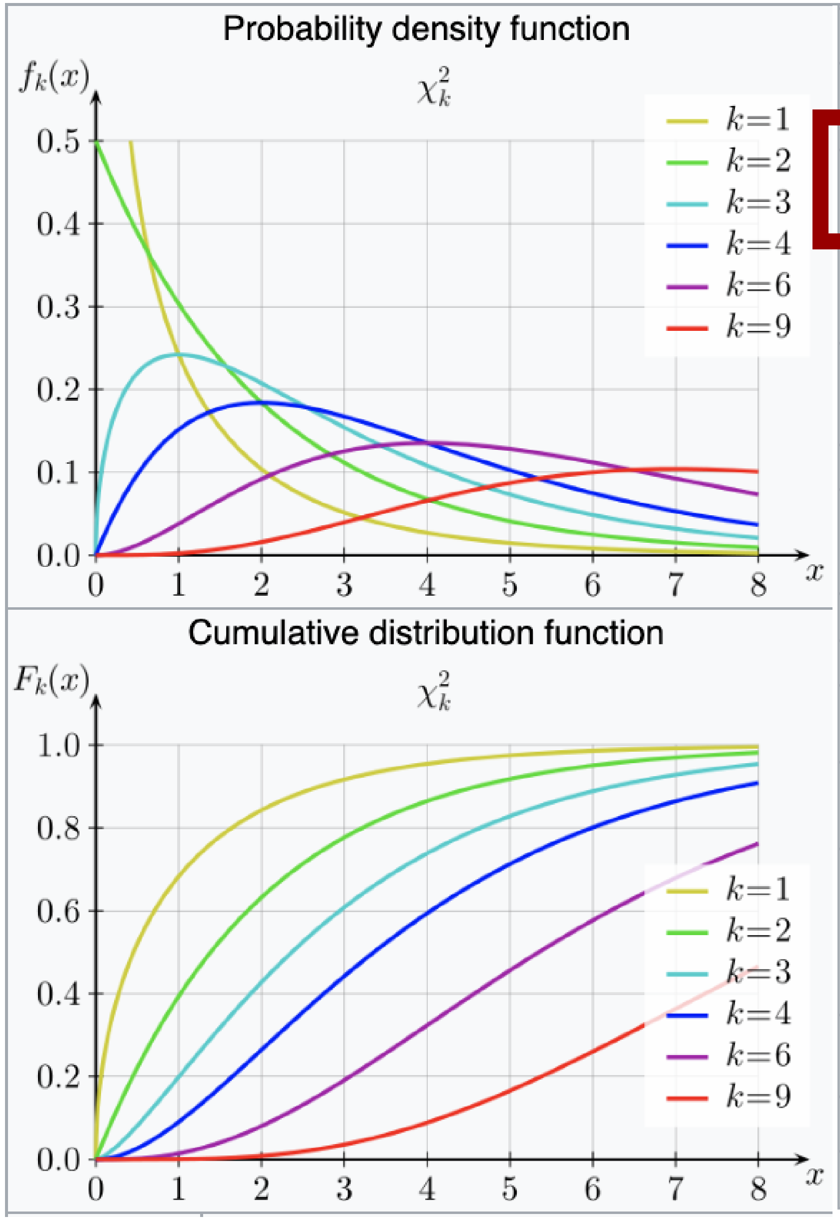
Probability distributions

Chi-square (χ2)

turns out its extremely common

many pivotal quantities follow this distribution and thus many tests are based on this

chi-square



Notation	$\chi^2(k)$ or χ^2_k
Parameters	$k \in \mathbb{N}_{>0}$ (known as "degrees of freedom")
Support	$x \in (0, +\infty)$ if $k = 1$, otherwise $x \in [0, +\infty)$
PDF	$\frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$
CDF	$\frac{1}{\Gamma(k/2)} \gamma\left(\frac{k}{2}, \frac{x}{2}\right)$
Mean	k
Median	$\approx k\left(1 - \frac{2}{9k}\right)^3$
Mode	$\max(k-2, 0)$
Variance	$2k$
Skewness	$\sqrt{8/k}$
Ex. kurtosis	$\frac{12}{k}$
Entropy	$\frac{k}{2} + \ln(2\Gamma(\frac{k}{2})) + (1 - \frac{k}{2})\psi(\frac{k}{2})$
MGF	$(1-2t)^{-k/2}$ for $t < \frac{1}{2}$
CF	$(1-2it)^{-k/2}$ [1]
PGF	$(1-2\ln t)^{-k/2}$ for $0 < t < \sqrt{e}$

coding time!

https://colab.research.google.com/

https://github.com/fedhere/FDSfE_FBianco/blob/master/statistics/distributionParametersDemo.ipynb