Urban Computing Skills Lab
Probabilities
Summer, 2016

Prof. Stanislav Sobolevsky

What is probability?

Frequentist vs Bayesian

Frequentist approach

P(E) is a frequency of E

P(C=heads)? 100 51 heads

P(C=heads)=51/100=0.51

Not have enough observations?

Bayesian approach

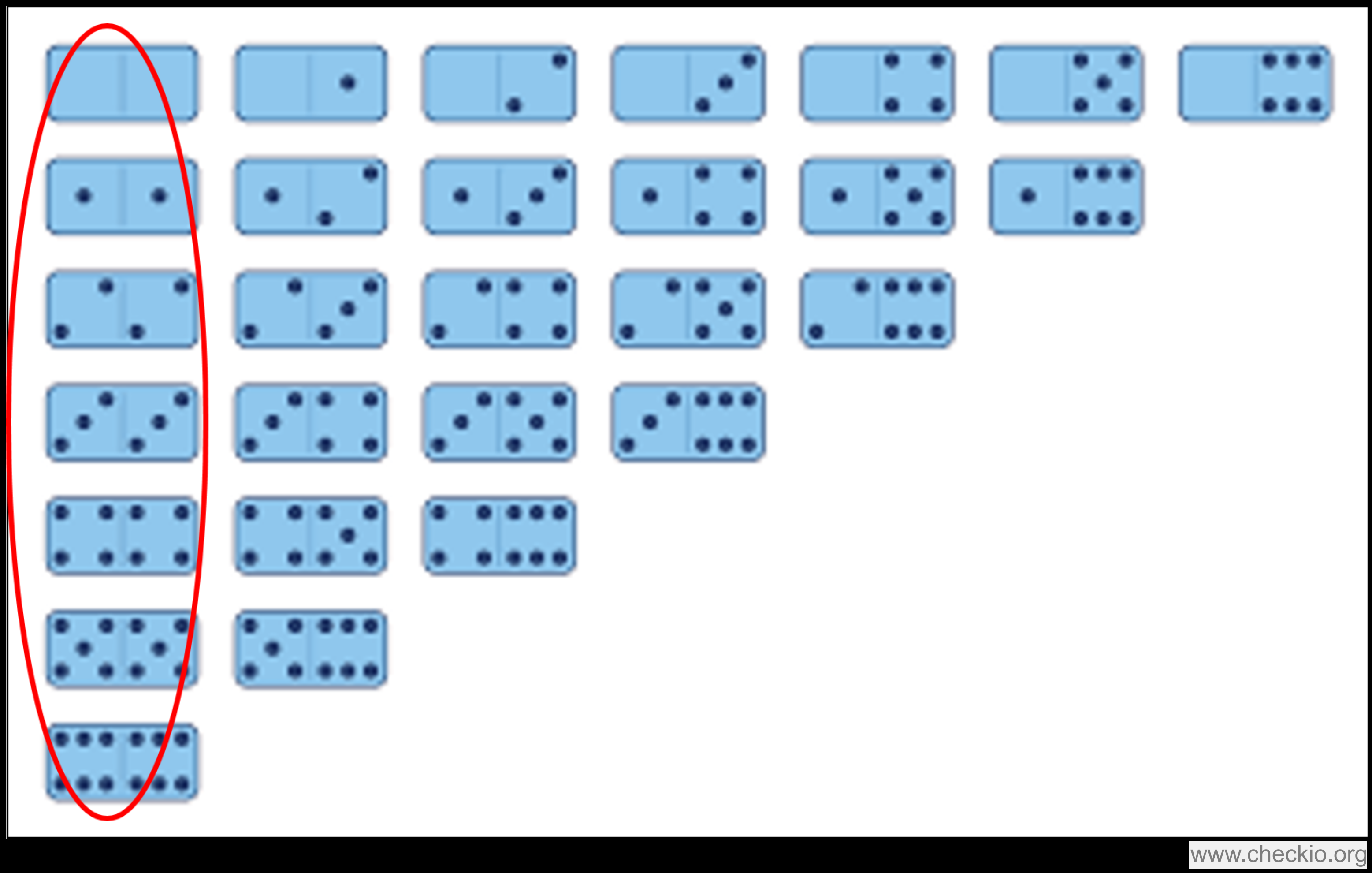
Probability P(E) - degree of confidence E will happen

P(Event|Beliefs)

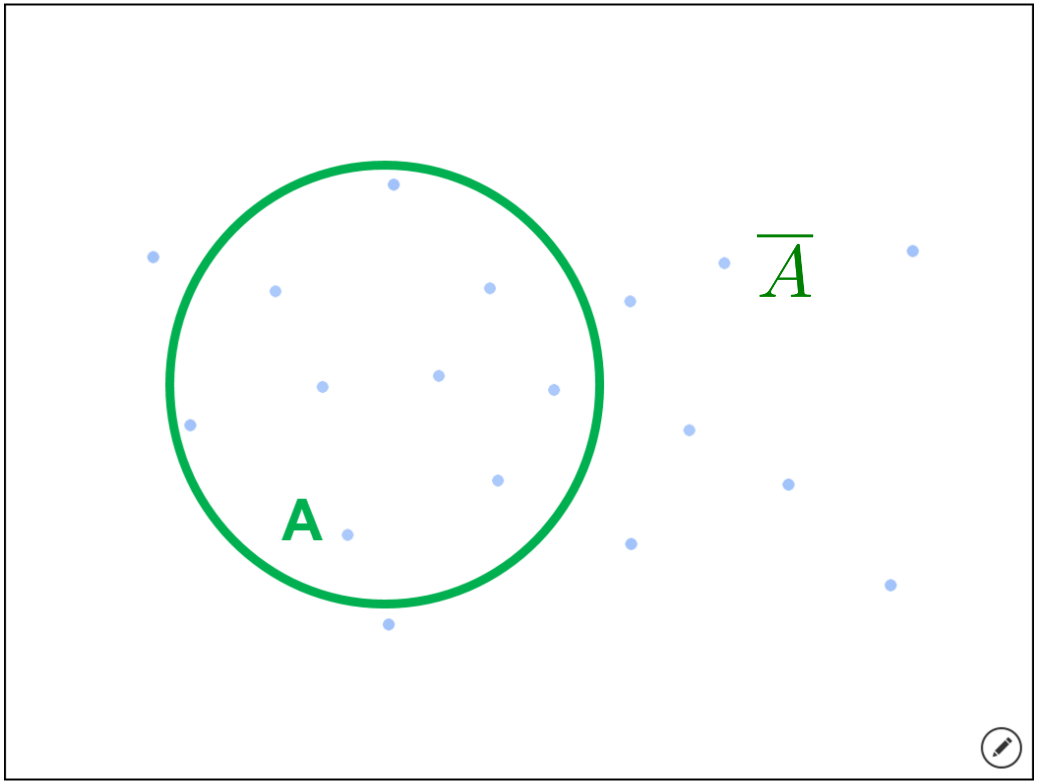
P(C=heads)=0.5

www.3dartistonline.com

Complex event



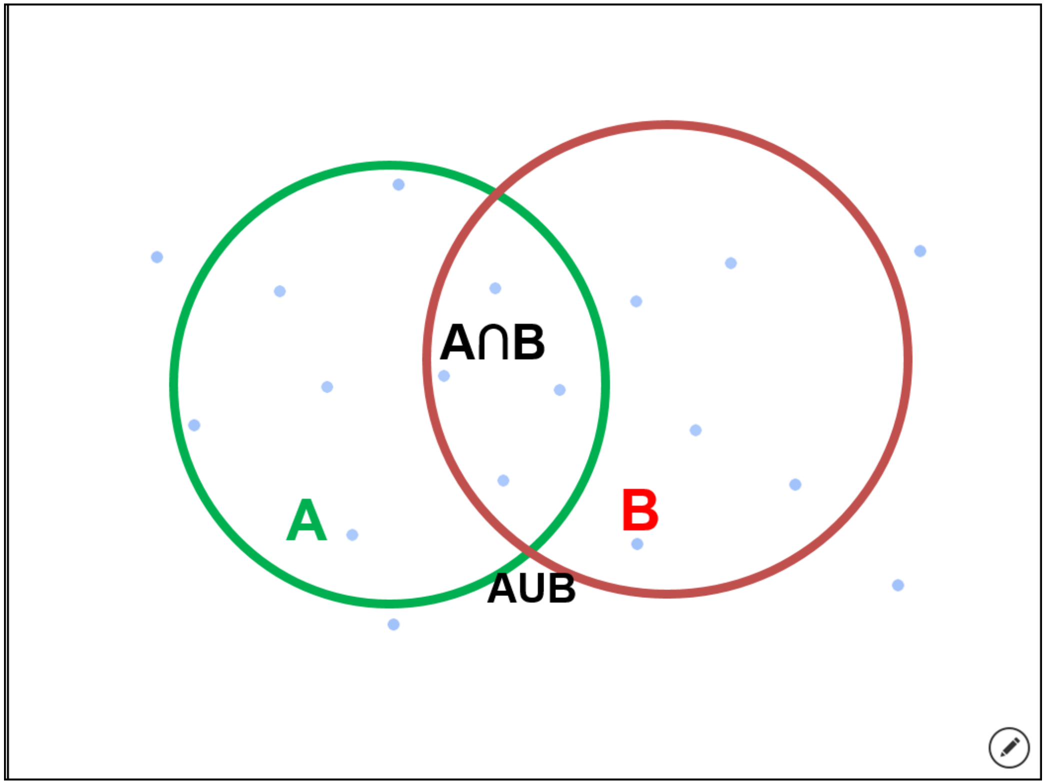
Complex events



P(A)=9/18=1/2

P(not A)=1−P(A)=1−1/2=1/2

Complex events - union/intersection

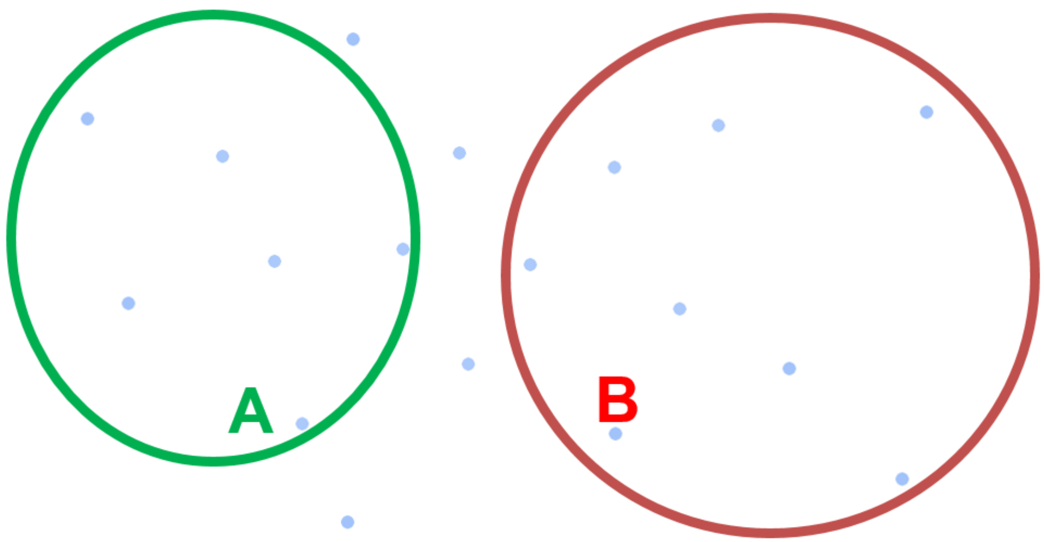


$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(\overline{A \cup B}) = P(\overline{A} \cap \overline{B})$$

Mutually exclusive complex events

$$P(A \cap B) = 0$$



$$P(A \cup B) = P(A) + P(B)$$

Independent events

$$P(AB) = P(A)P(B)$$

Discrete random variables

X - variable which we are uncertain about

Event as a binary random variable:

$X=1$ if event happened

$X=0$ if not

$X \in S$

$S=\{1,2,\ldots,M\}$

$P(X=k)$ - pmf

1.3.1 Discrete random variable

P(S=k), k=0..12

P(S=0)=1/28

P(S=1)=1/28

P(S=2)=2/28=1/14

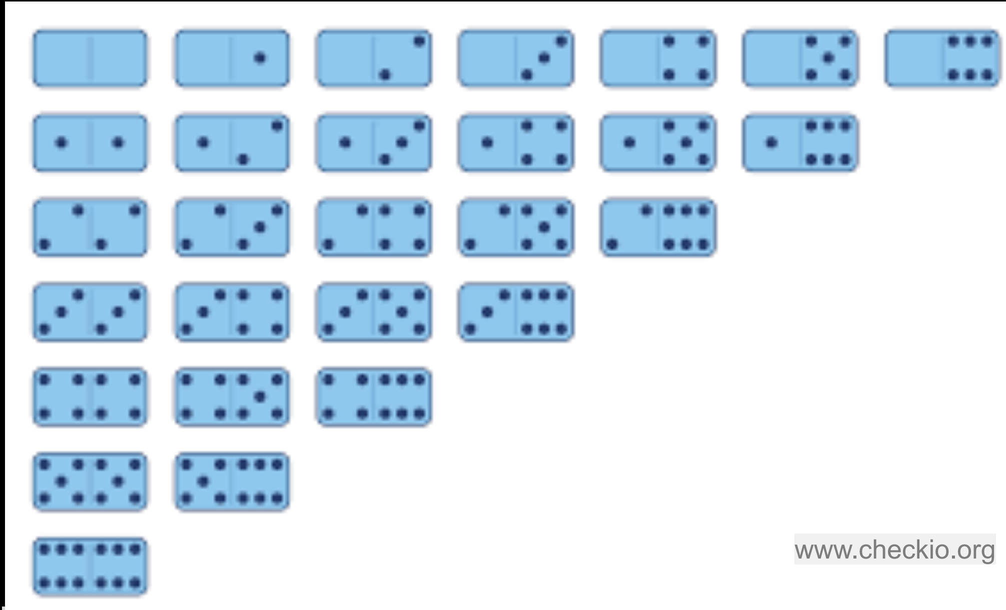
P(S=3)=2/28=1/14

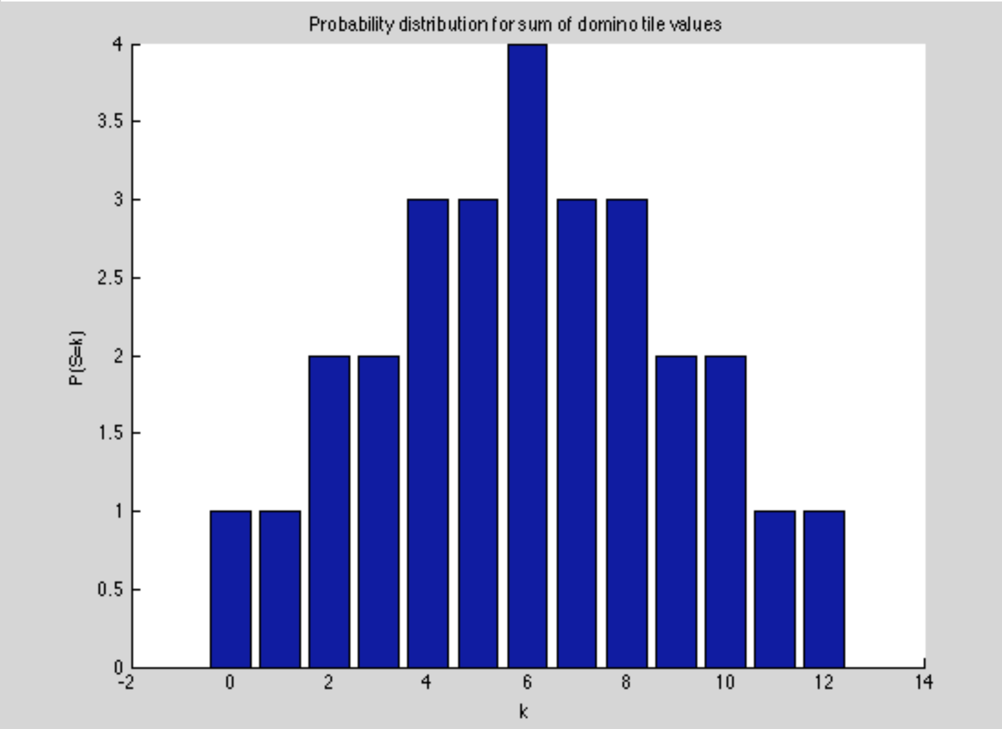
P(S=4)=3/28

P(S=5)=3/28

P(S=6)=4/28=1/7

P(S=6+k)=P(S=6-k)





Mean and variance

$$\mu = E[X] = \sum_k P(X = k)k$$

$$\sigma^2 = var[X] = E[(X - E[X])^2] = \sum_k P(X = k)\,(k - \mu)^2$$

$$\sigma = std[X] = \sqrt{var[X]}$$

$$E[X^2] = \mu^2 + \sigma^2$$

Classical discrete distributions - Bernoulli

Bernoulli: flipping a coin

$$Bern(X = 1|p) = p$$

$$Bern(X = 0|p) = 1 - p$$

$$\mu = p * 1 + (1 - p) * 0 = p$$

$$\sigma = \sqrt{p * (1 - p)^2 + (1 - p) * p^2} = \sqrt{p(1 - p)}$$

Classical discrete distributions - Binomial

Binomial

$$Bin(X = k|n, p) = C_n^k p^k (1 - p)^{n-k}$$

$$C_n^k = \frac{n!}{(n-k)!k!}$$

$$\mu = E[X] = pn$$

$$\sigma = std[X] = \sqrt{np(1-p)}$$

Classical discrete distributions - Poisson

Poisson

$$Poi(X = k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$$

$$E[X] = var[X] = \lambda = \mu = \sigma^2$$

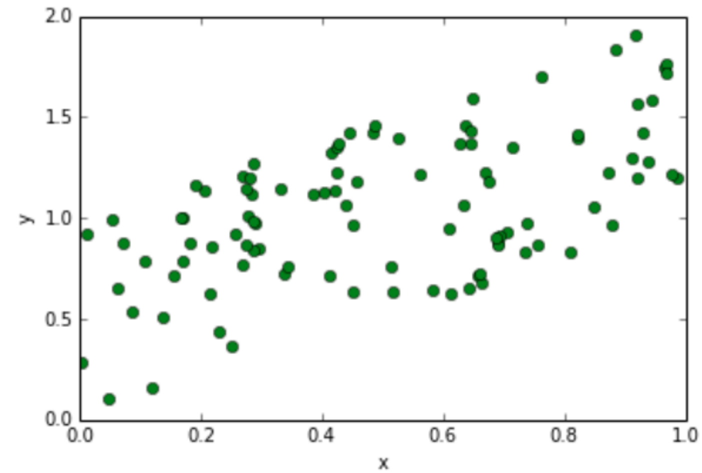
Correlation

Covariance:

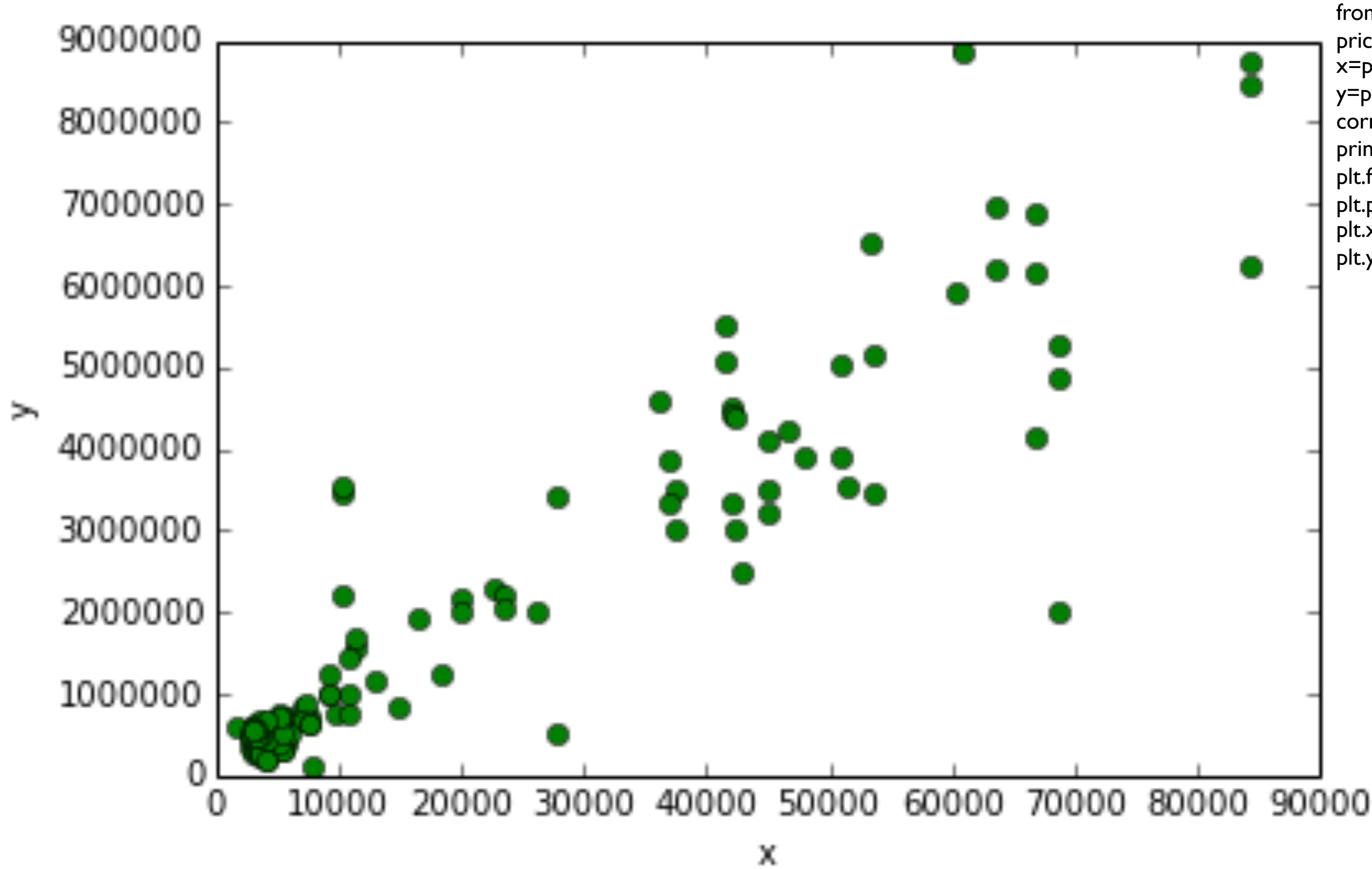
$$cov(X,Y) = E\left[(X - E[X])\,(Y - E[Y])\right]$$

Pearson’s correlation coefficient:

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$$



Correlation - house price vs size




```
import numpy as np
from scipy.stats.stats import pearsonr
prices = np.loadtxt("NYC_RE_10466_multi.csv",delimiter=",")
x=prices[:,0]
y=prices[:,1]
corr=pearsonr(x,y)[0]
print('Correlation={0}'.format(corr))
plt.figure()
plt.plot(x,y,'og')
plt.xlabel('x')
plt.ylabel('y')
```

Correlation=0.926477987l4

Correlation and causality

logical fallacy: cum hoc ergo propter hoc