$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

$$var(x) = \int (x-\mu)^2\, p(x)\, dx = E([x-\mu]^2)$$

variance can be interpreted as **mean of $[x-\mu]^2$**

$$= E(x^2 - 2x\mu + \mu^2)$$

$$= \int [x^2 - 2x\mu + \mu^2]\, p(x)\, dx$$

$$= \int x^2\, p(x)\, dx - 2\mu \int x\, p(x)\, dx + \mu^2 \int p(x)\, dx$$

$$= E(x^2) - 2\mu\, E(x) + \mu^2\, E(1)$$

$$\int p(x)\, dx = 1$$

$$= E(x^2) - 2\mu\, E(x) + \mu^2$$

$$\mu = E(x)$$

$$\boxed{\boldsymbol{\sigma^2 = E(x^2) - E(x)^2}}$$

Refresher:

$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

recap:

a, b = const $\boxed{\sigma^2 = E(x^2) - E(x)^2}$

$$var([a\,x_1 + b\,x_2]) = E([a\,x_1 + b\,x_2]^2) - E(a\,x_1 + b\,x_2)^2$$

$$= E(a^2\,x_1^2 + 2ab\,x_1x_2 + b^2\,x_2^2) - E(a\,x_1 + b\,x_2)^2$$

$$= a^2E(x_1^2) + 2ab\,E(x_1x_2) + b^2\,E(x_2^2) - E(a\,x_1 + b\,x_2)^2$$

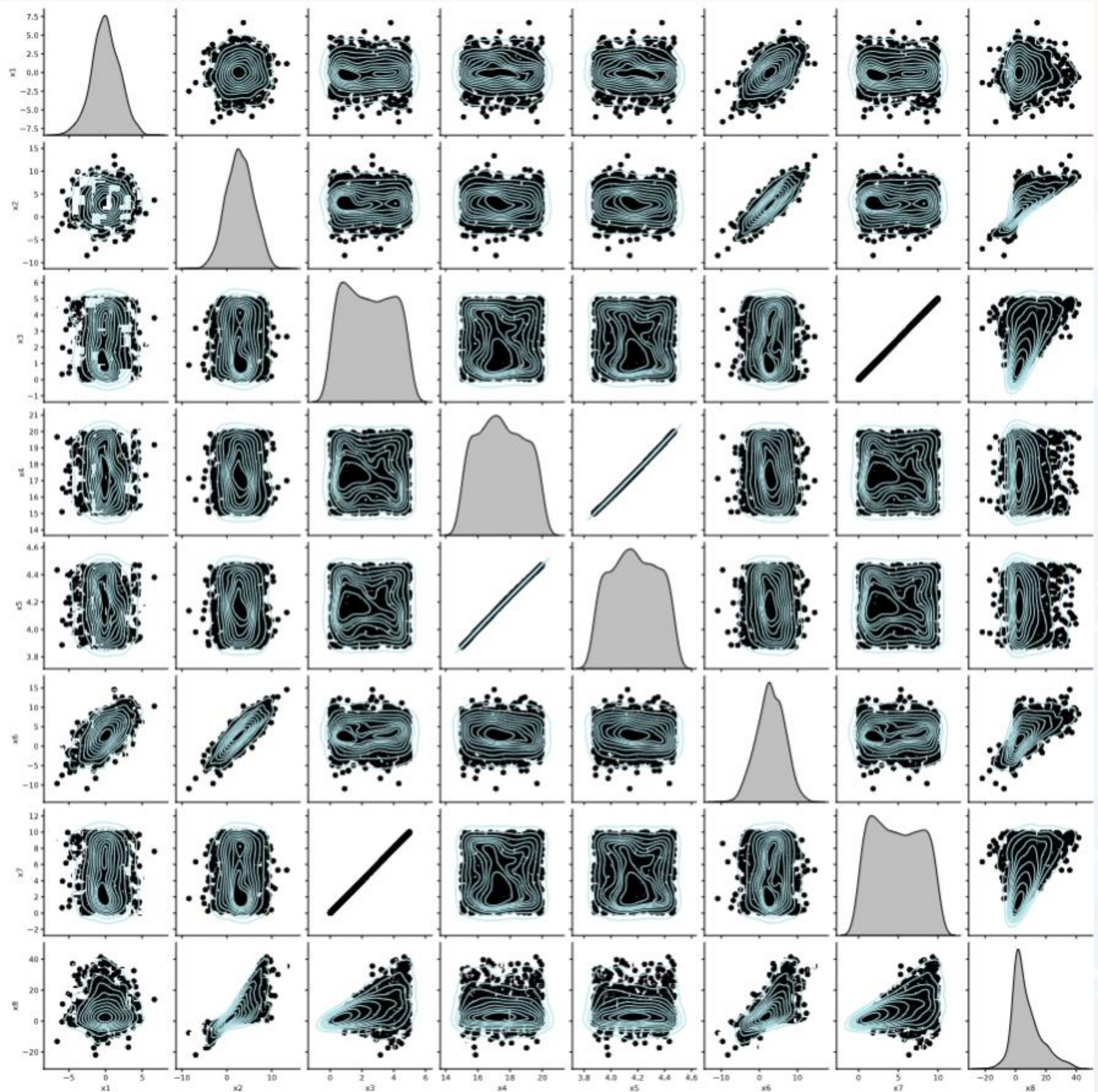
$$= a^2E(x_1^2) + 2ab\,E(x_1x_2) + b^2\,E(x_2^2) - [aE(x_1) + b\,E(x_2)]^2$$

$$= \underbrace{a^2E(x_1^2) - a^2E(x_1)^2}_{a^2\,var(x_1)} + \underbrace{b^2\,E(x_2^2) - b^2E(x_2)^2}_{b^2\,var(x_2)} + \underbrace{2ab\,E(x_1x_2) - 2abE(x_1)E(x_2)}_{2ab\,cov(x_1,x_2)}$$

$$= a^2\,var(x_1) + b^2\,var(x_2) + 2ab\,cov(x_1,x_2)$$

$\boxed{cov(x_1,x_2) = cov(x_2,x_1) = E(x_1x_2) - E(x_1)E(x_2)}$ **covariance**

Refresher: variance, covariance, correlation



$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

based on the shape of the data cloud

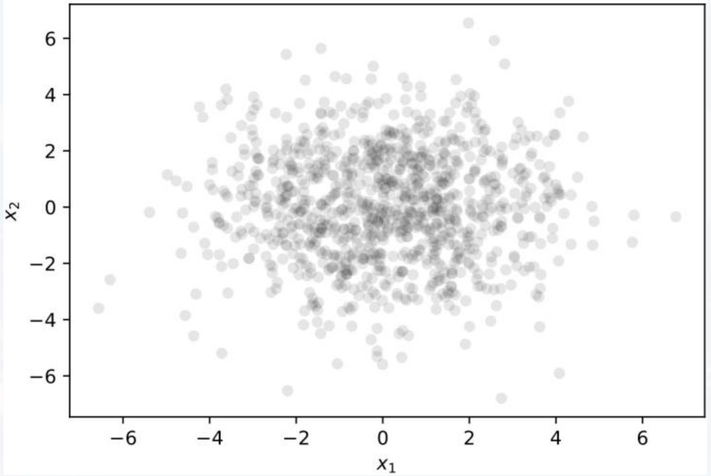
→ prediction how x_1 and x_2 are related, i. e. how they **correlate**

→ how to quantify?

$$\boldsymbol{cov(x_1, x_2) = cov(x_2, x_1)} = E(x_1 x_2) - E(x_1)E(x_2)$$

$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x - \mu)^2\, p(x)\, dx$$



a) x_1 and x_2 are independent

$$E(x_1 x_2) - E(x_1)E(x_2)$$

$$= \iint x_1 x_2\, p(x_1)\, p(x_2)\, dx_1 dx_2 - \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2$$

x_1 and x_2 are independent: x_1 is not a function of x_2 and vise verse
x_1 cannot be predicted by x_2 and vise verse

$$= \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2 - \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2 = 0$$

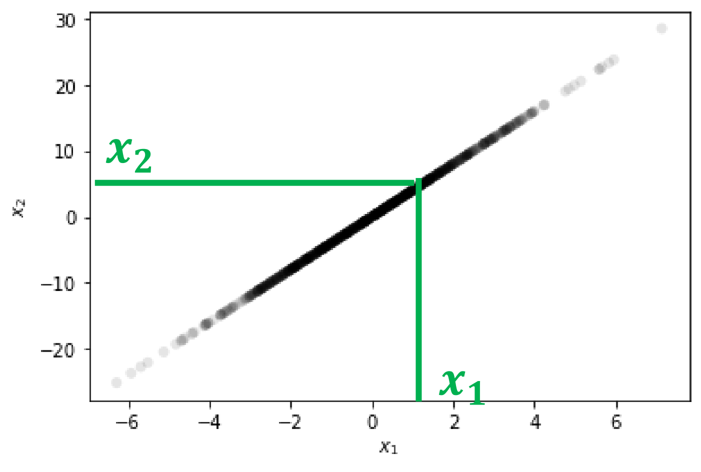
covariance equals **zero**
if samples are **independent**!

Refresher: variance, covariance, correlation

$$\boldsymbol{cov(x_1, x_2) = cov(x_2, x_1)} = E(x_1 x_2) - E(x_1)E(x_2)$$

$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$



b) x_1 and x_2 are **not** independent

$$E(x_1 x_2) - E(x_1)E(x_2)$$

$$= \iint x_1 x_2\, p(x_1)\, p(x_2)\, dx_1 dx_2 - \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2$$

x_1 and x_2 are **not** independent: x_1 ***is*** a function of x_2 and vise verse

x_1 ***can*** be predicted by x_2 to certain degree and vise verse

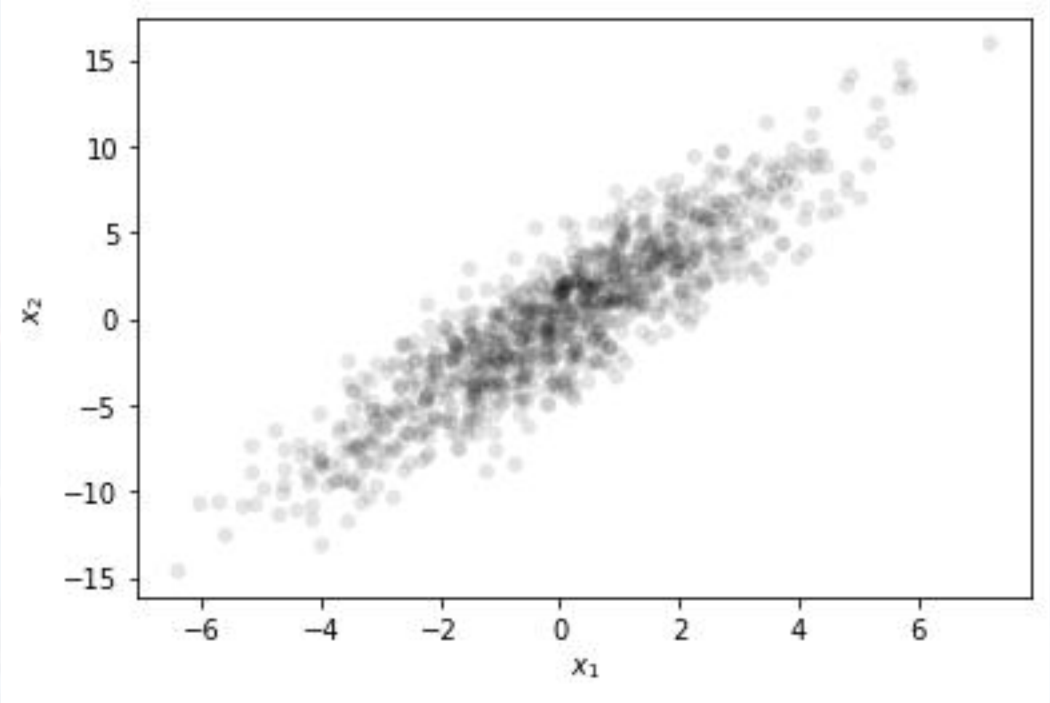
$$= \iint x_1\, p(x_1)\;\; x_2(x_1)\, p(x_2(x_1))\, dx_1 dx_2(x_1) - \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2$$

covariance **does not** equal **zero**!

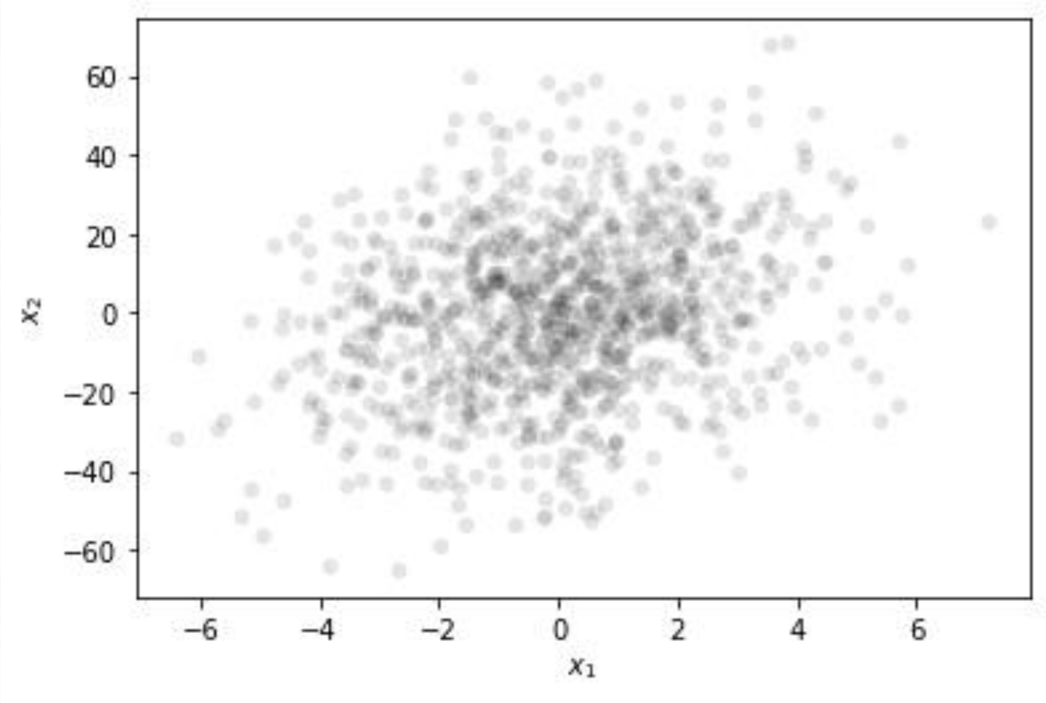
$$\boxed{cov(x_1, x_2) = cov(x_2, x_1) = E(x_1 x_2) - E(x_1)E(x_2)}$$

covariance

```
x1 = np.random.normal(0, 2, (1000,))
x2 = 2*x1 + np.random.normal(0, 2, (1000,))
```




```
x1 = np.random.normal(0, 2, (1000,))
x2 = 2*x1 + np.random.normal(0, 20, (1000,))
```

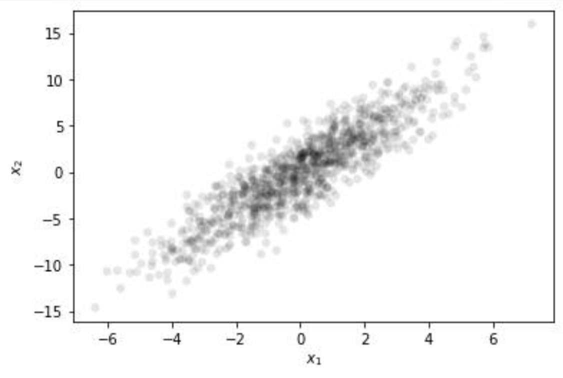


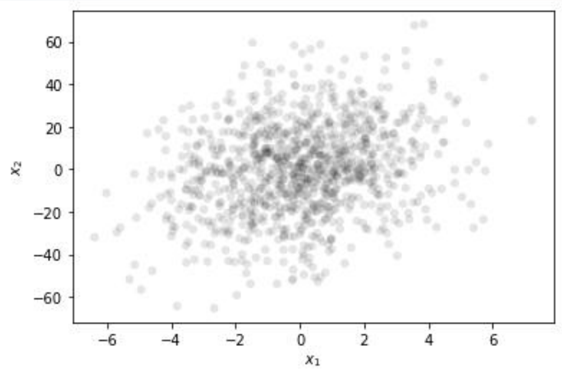
Same dependency, but different variance!

Refresher:

$$\boxed{cov(x_1, x_2) = cov(x_2, x_1) = E(x_1 x_2) - E(x_1)E(x_2)}$$

covariance





Same dependency, but different variance!

Need to scale for the variance!

Pearson's correlation coefficient

$$\boxed{\rho(x_1, x_2) = \frac{cov(x_1, x_2)}{\sqrt{\sigma_1^2 \sigma_2^2}}}$$

$\rho(x_1, x_2)$:

- ranges from -1 to +1
- zero: no correlation

  (completely independent)
- -1: max anti correlation
- +1: max correlation

Refresher: variance, covariance, correlation

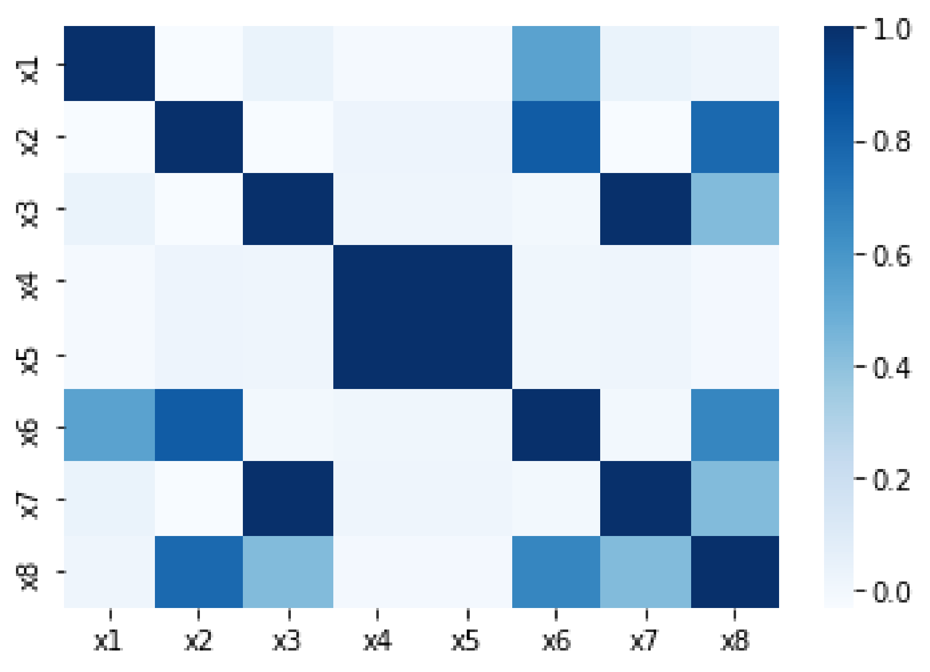
$$\boxed{cov(x_1, x_2) = cov(x_2, x_1) = E(x_1 x_2) - E(x_1)E(x_2)}$$

covariance

$$\boxed{\rho(x_1, x_2) = \frac{cov(x_1, x_2)}{\sqrt{\sigma_1^2 \sigma_2^2}}}$$

Pearson's correlation coefficient

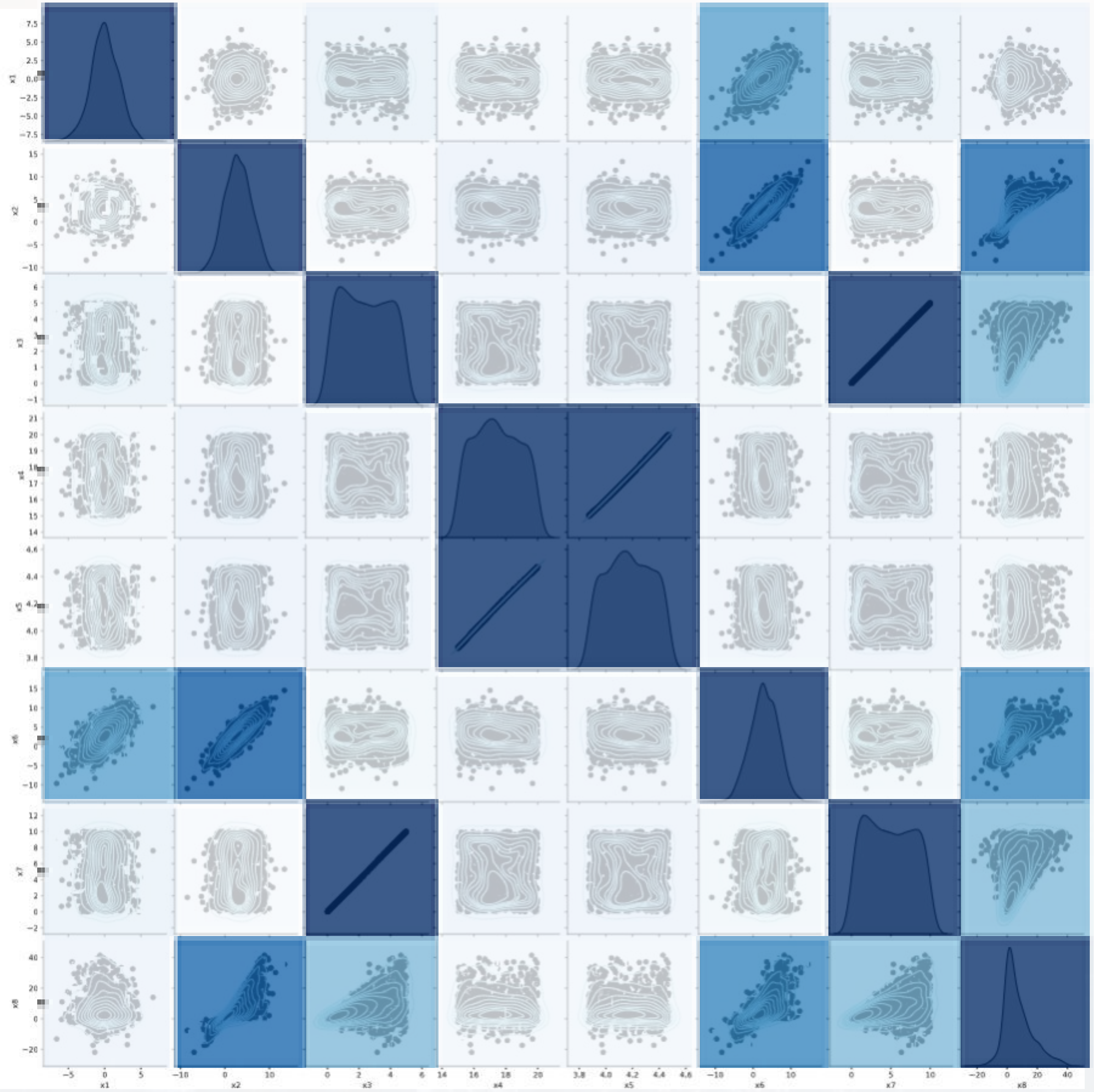
```
sns.heatmap(data.corr(), cmap = "Blues")
```



$\rho(x_1, x_2)$:

- ranges from -1 to +1
- zero: no correlation
  (completely independent)
- -1: max anti correlation
- +1: max correlation

Refresher: variance, covariance, correlation



$\rho(x_1, x_2)$:

- ranges from -1 to +1
- zero: no correlation (completely independent)
- -1: max anti correlation
- +1: max correlation

Important quantities you should know:

mean $$\mu = E(x) = \int x\, p(x)\, dx$$

median *m* $$\int_a^{m} p(x)\, dx = \frac{1}{2}$$

variance $$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

$$\sigma^2 = E(x^2) - E(x)^2$$

$$\sigma_{tot}^2 = \sigma_{11}^2 + \sigma_{22}^2 + 2\, cov(x_1, x_2)$$

covariance $$cov(x_1, x_2) = E(x_1 x_2) - E(x_1)E(x_2)$$

correlation coefficient $$\rho(x_1, x_2) = \frac{cov(x_1, x_2)}{\sqrt{\sigma_{11}^2 \sigma_{22}^2}}$$

note:

$$\int (x-\mu)^n\, p(x)\, dx$$

called n-th *moment* of a pdf

Questions?