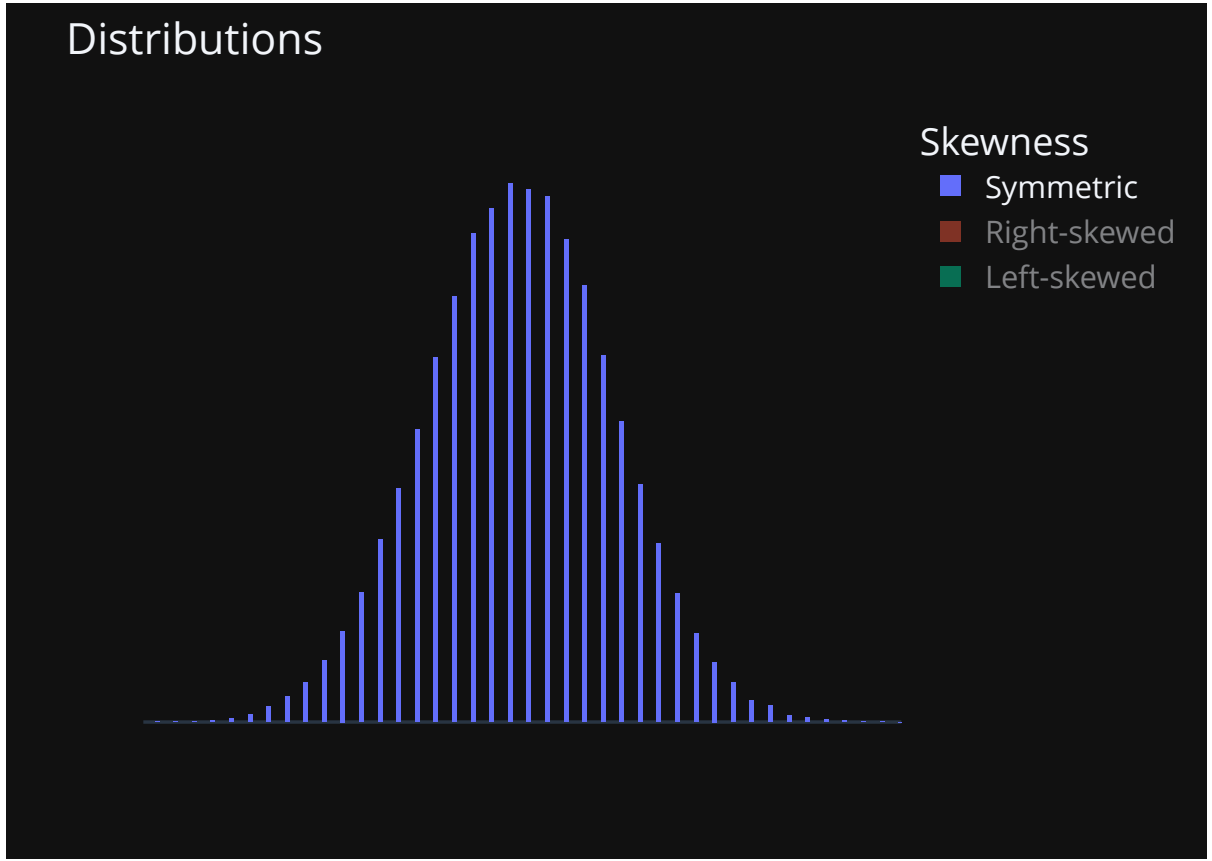
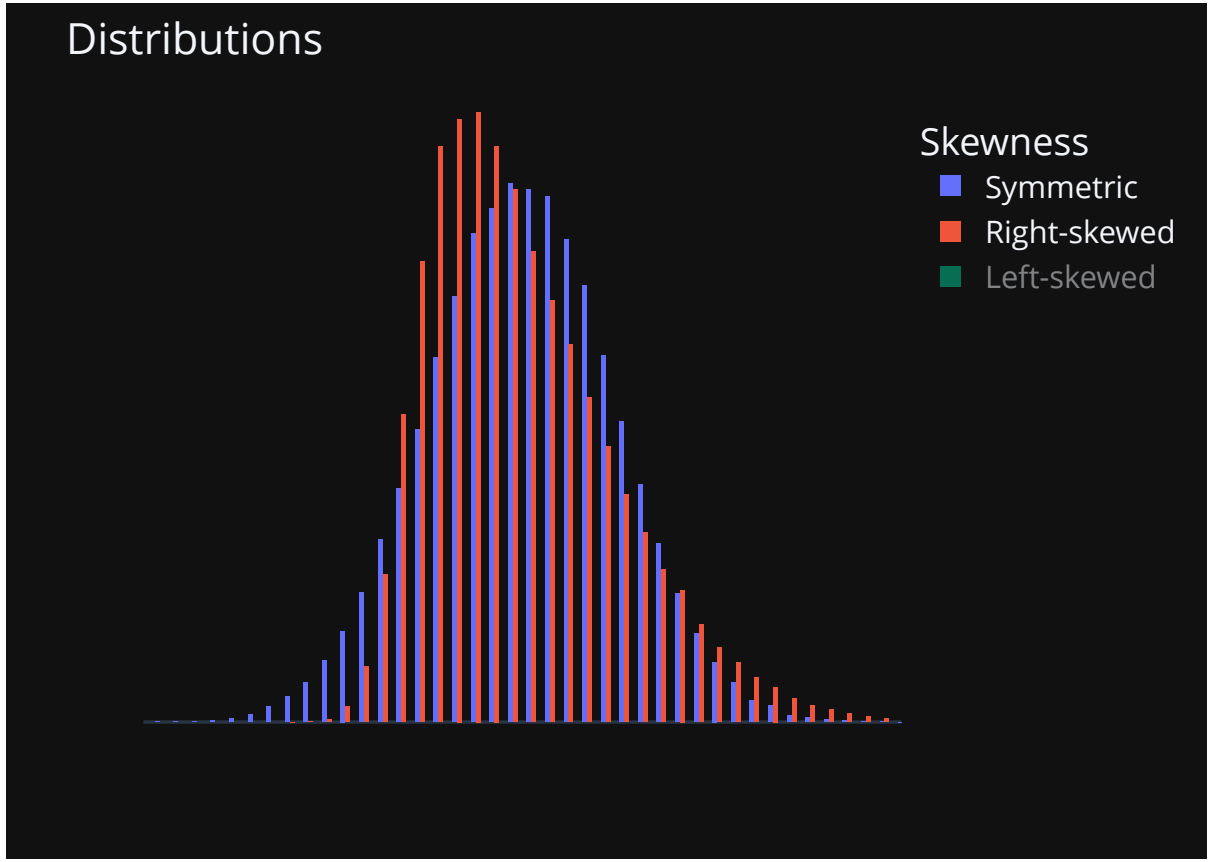
Applied Data Analytics

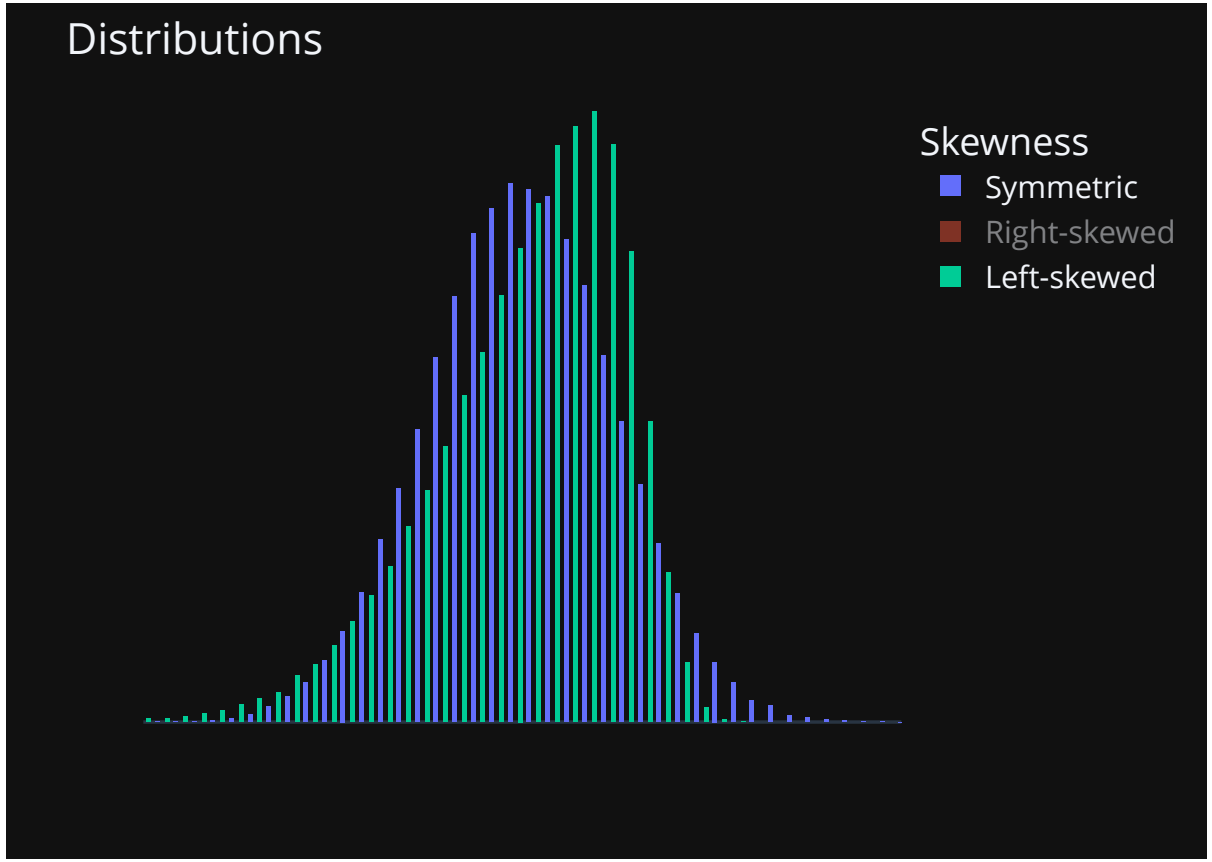
Statistics — Dispersion & concentration

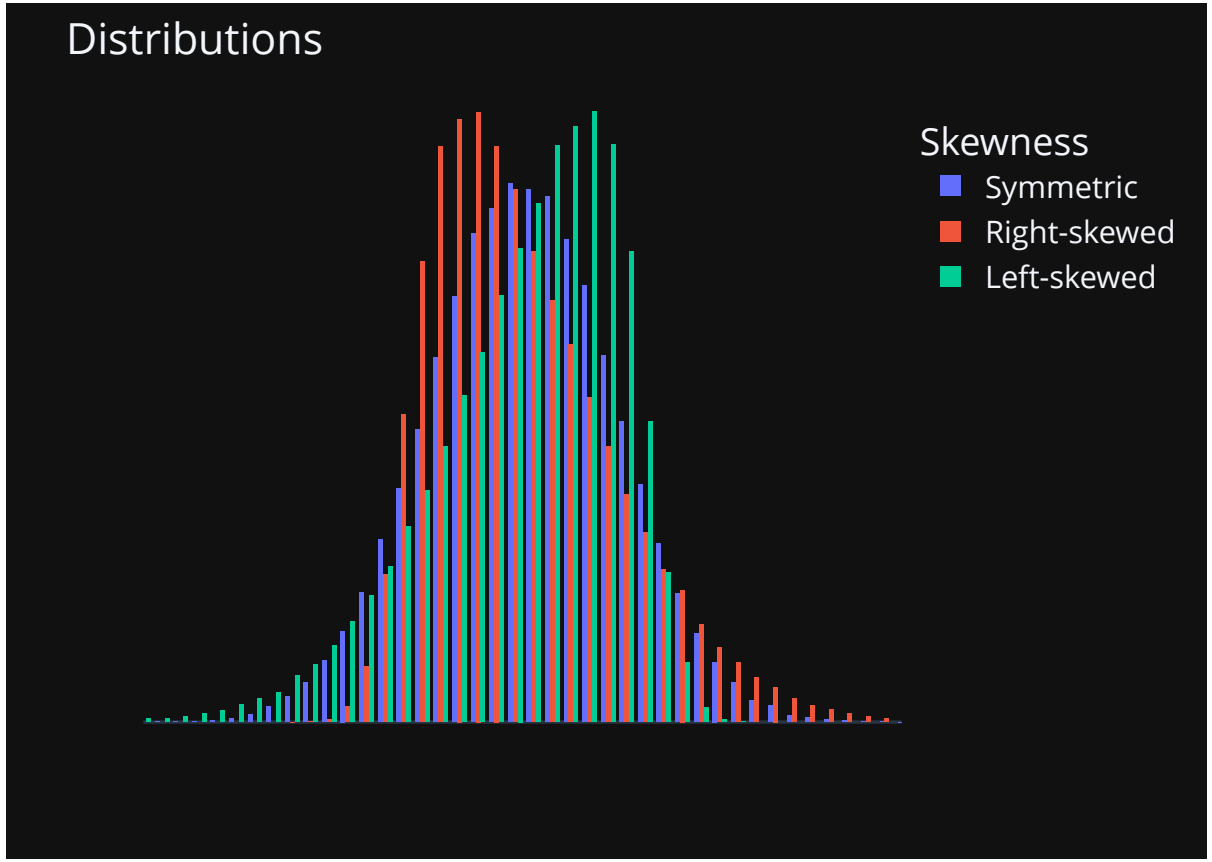
Skewness

Hans-Martin von Gaudecker and Aapo Stenhammar


Distributions
Skewness
Symmetric
Right-skewed
Left-skewed


Distributions
Skewness
Symmetric
Right-skewed
Left-skewed


Distributions
Skewness
Symmetric
Right-skewed
Left-skewed


Distributions
Skewness
Symmetric
Right-skewed
Left-skewed

Describe a DataFrame

```
df.describe().round(2)
```

Example	count	mean	std	min	25%	50%	75%	max
Symmetric	100000	-0	1	-4.54	-0.67	0	0.68	4.35
Right-skewed	100000	-0	1	-2.46	-0.75	-0.18	0.59	5.77
Left-skewed	100000	-0	1	-5.72	-0.59	0.18	0.76	2.38

Describe a DataFrame

```
df.skew().round(1)
```

Example	Value
Symmetric	0.0
Right-skewed	0.9
Left-skewed	-0.9

Skewness

$$\frac{n}{(n-1)(n-2)} \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

- The only important bit is the second fraction
- Pandas actually uses a slightly different df normalisation

Skewness

A	$(A - 4)^3$	B	$(B - 4)^3$
2	-8	1	-27
4	0	3	-1
6	8	8	64
SCD	0	SCD	36
Skewness	0	Skewness	0.84