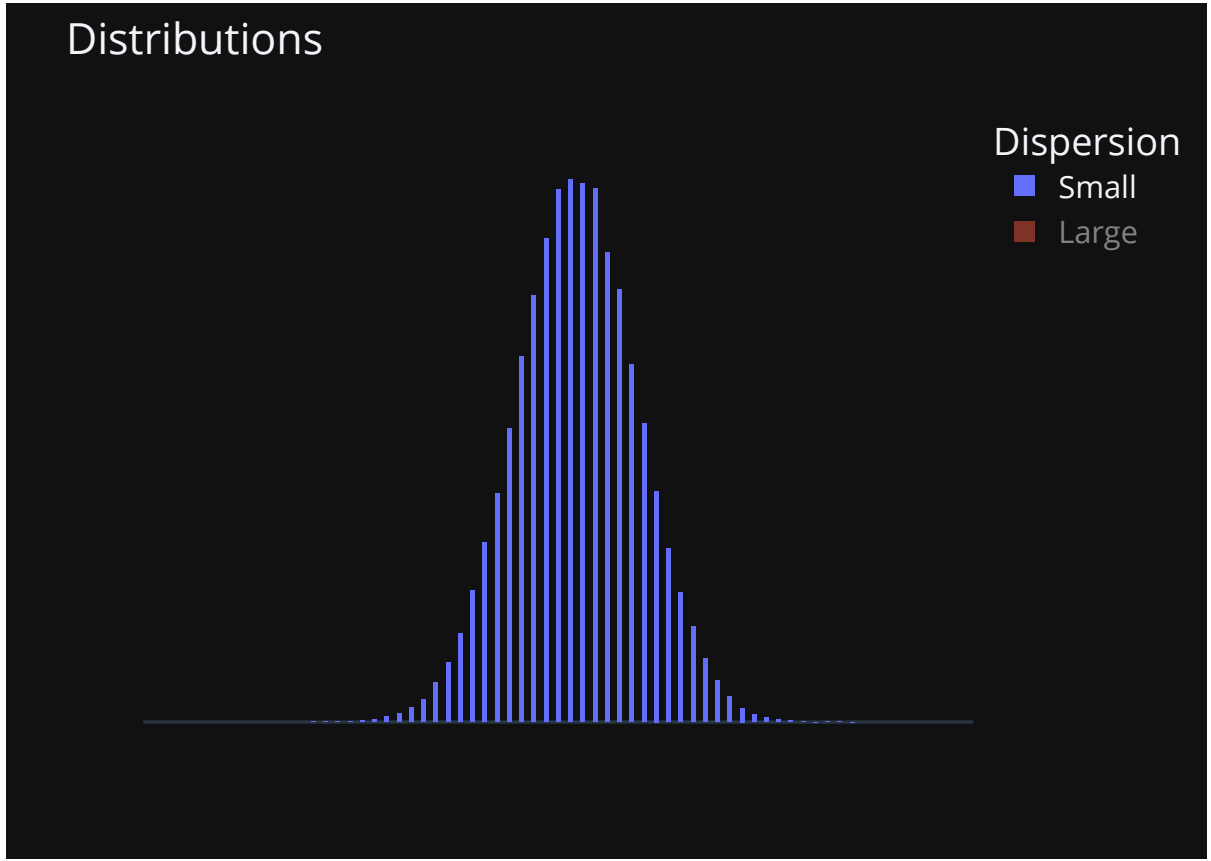
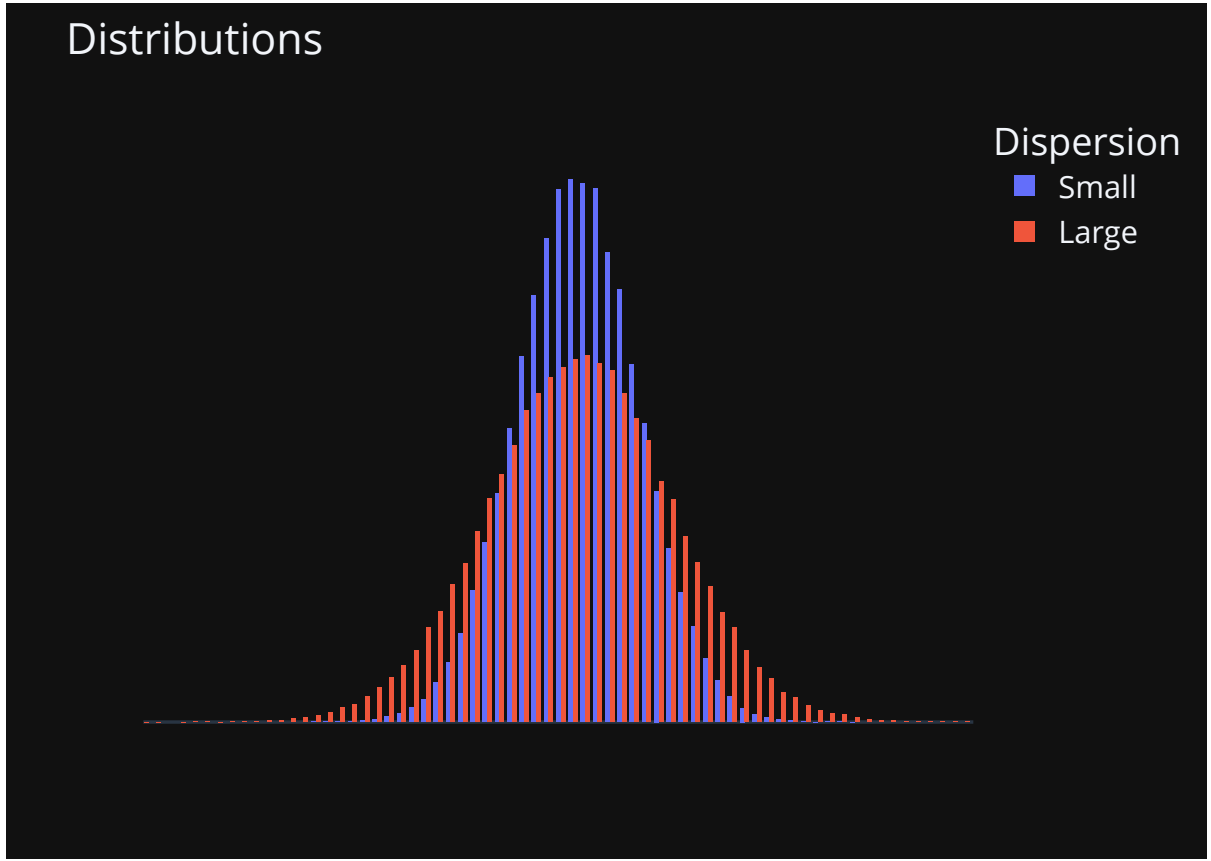
Applied Data Analytics

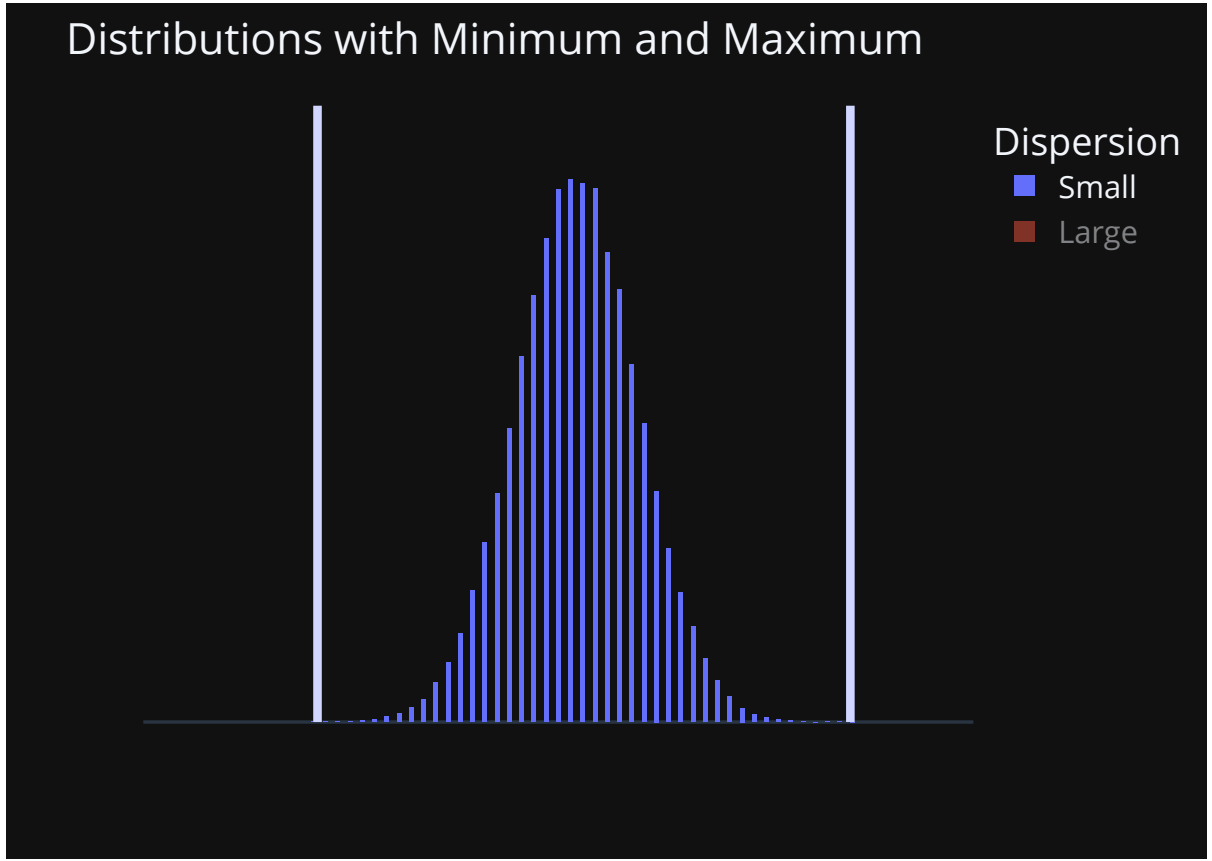
Statistics — Dispersion & concentration

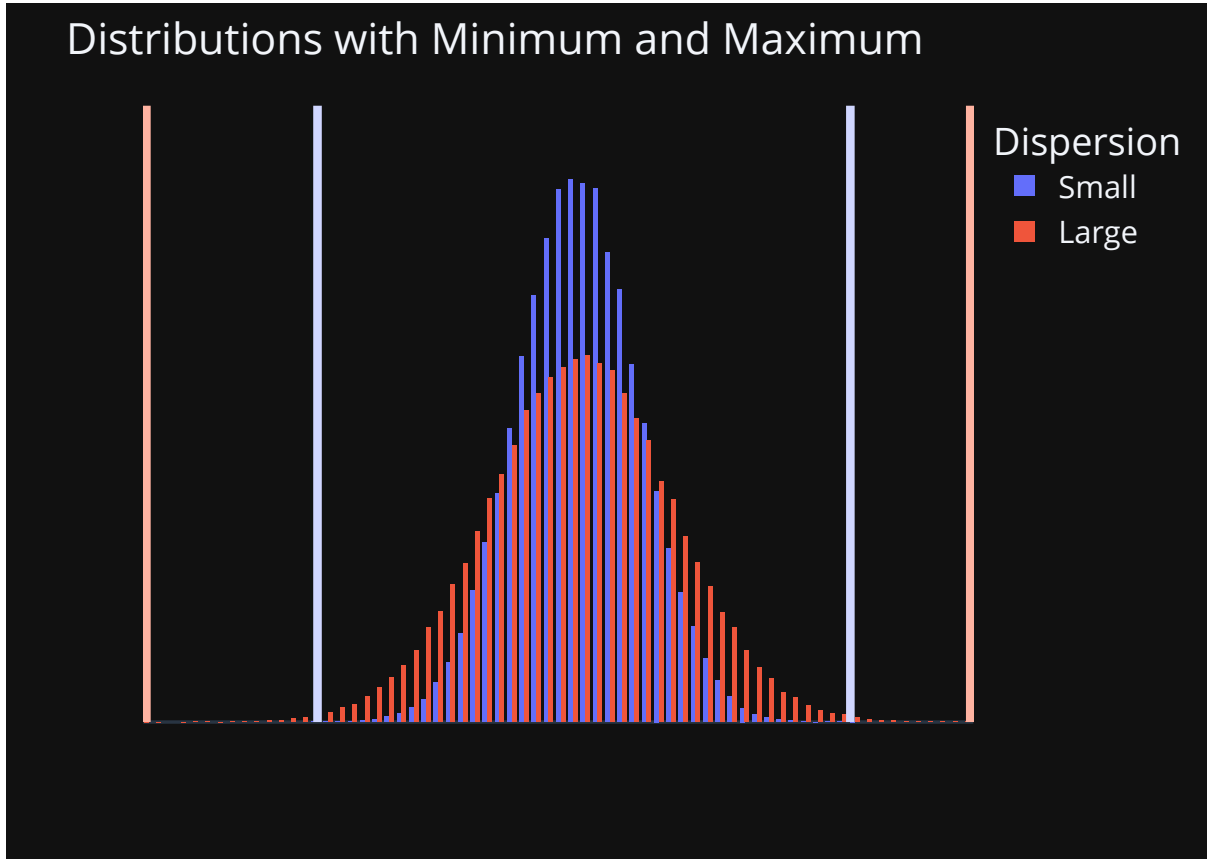
Dispersion based on absolute deviations

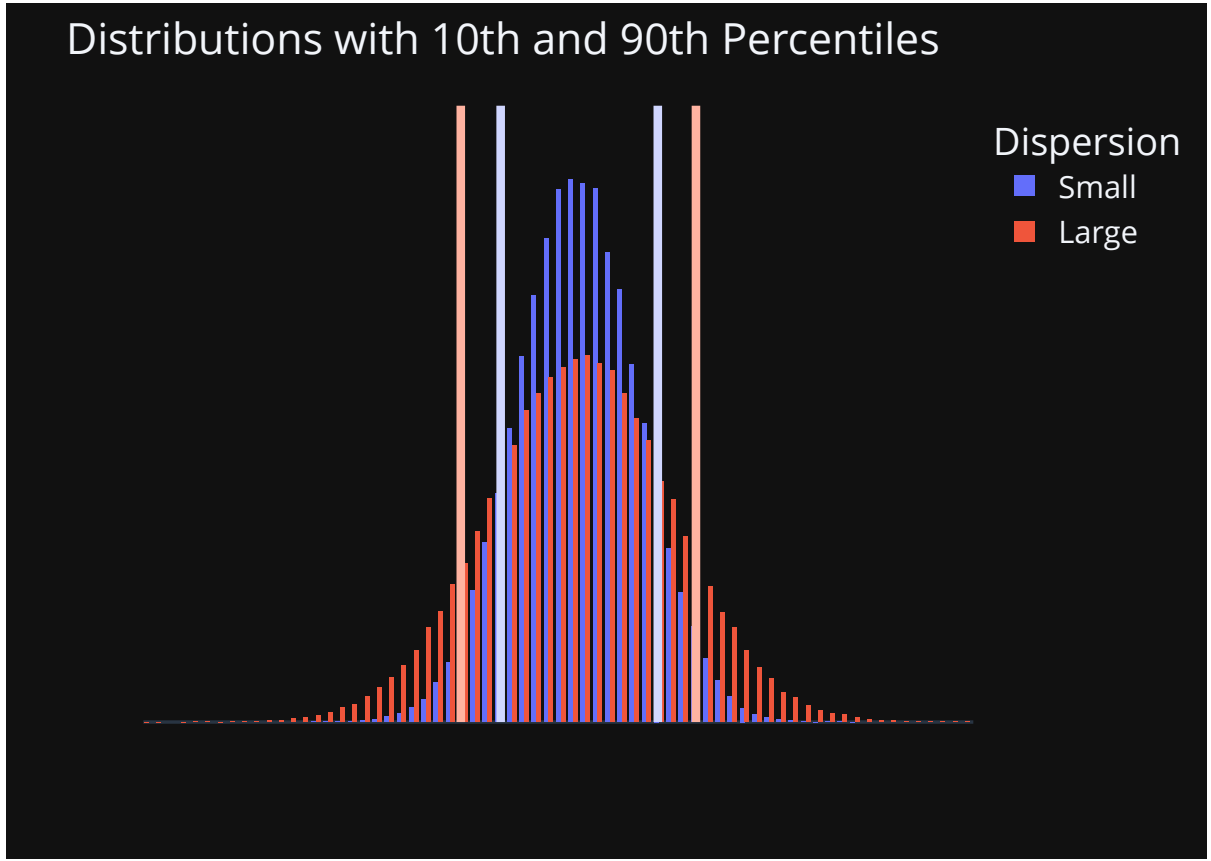
Hans-Martin von Gaudecker and Aapo Stenhammar

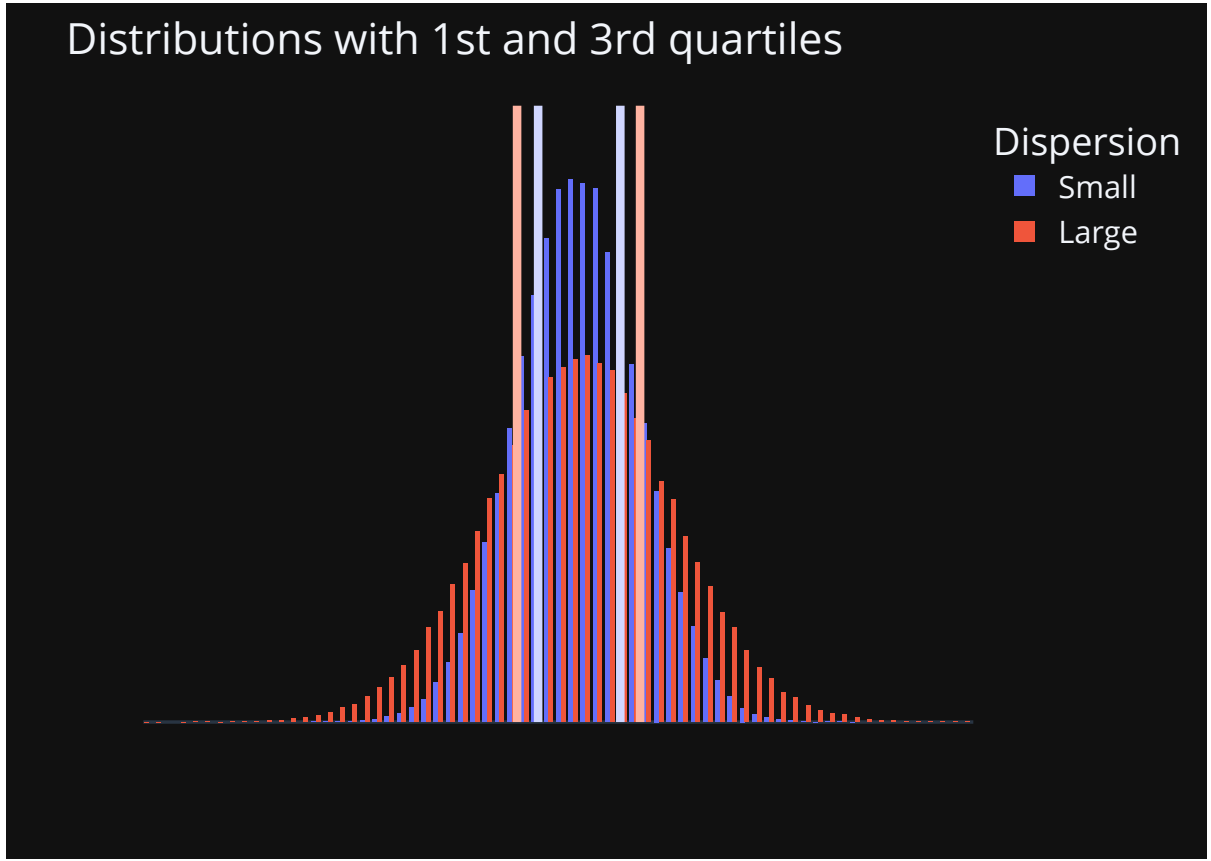

Distributions
Dispersion
Small
Large


Distributions
Dispersion
Small
Large


Distributions with Minimum and Maximum
Dispersion
Small
Large


Distributions with Minimum and Maximum
Dispersion
Small
Large


Distributions with 10th and 90th Percentiles
Dispersion
Small
Large


Distributions with 1st and 3rd quartiles
Dispersion
Small
Large

Ranges

- **Range**: difference between the maximum and minimum values
- **Interdecile range**: difference between the 90th and 10th percentiles
- **Interquartile range**: difference between the 75th and 25th percentiles
- For ratio scales with strictly positive values, ratios between quantiles can be useful statistics, too

Describe a DataFrame

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

Dispersion	count	mean	std	min	25%	50%	75%	max
Small	100000	0	1	-4.27	-0.67	-0	0.67	4.43
Large	100000	-0.01	1.5	-7.07	-1.01	-0	1	6.39

Mean absolute deviation

$$s_{\mathrm{MAD}} = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - x_{\mathrm{MED}}\right|$$

Mean absolute deviation

A	|A - 4|	B	|B - 3|
2	2	1	2
4	0	3	0
6	2	8	5
MAD	1.33	MAD	2.33