FOR NULL VALUES

- To decide what we'll do with the missing values.
- Drop them?
- Fill them? If we decide to fill them, what will be use as fill value?

For example: we can use the previous value and just assume the price stayed the same!

TO IDENTIFY OUTLIERS OR SKEWED VALUES

- Visualizations helps make sense of the data and let us judge if our analysis and work is on the right track. But we need a more powerful method to handle our data. That's what we call "analysis". We'll use *analytical* methods to identify these outliers or these skewed values.

- Central Tendency: Use a set of common indicators of to measure central tendency and identify outliers:

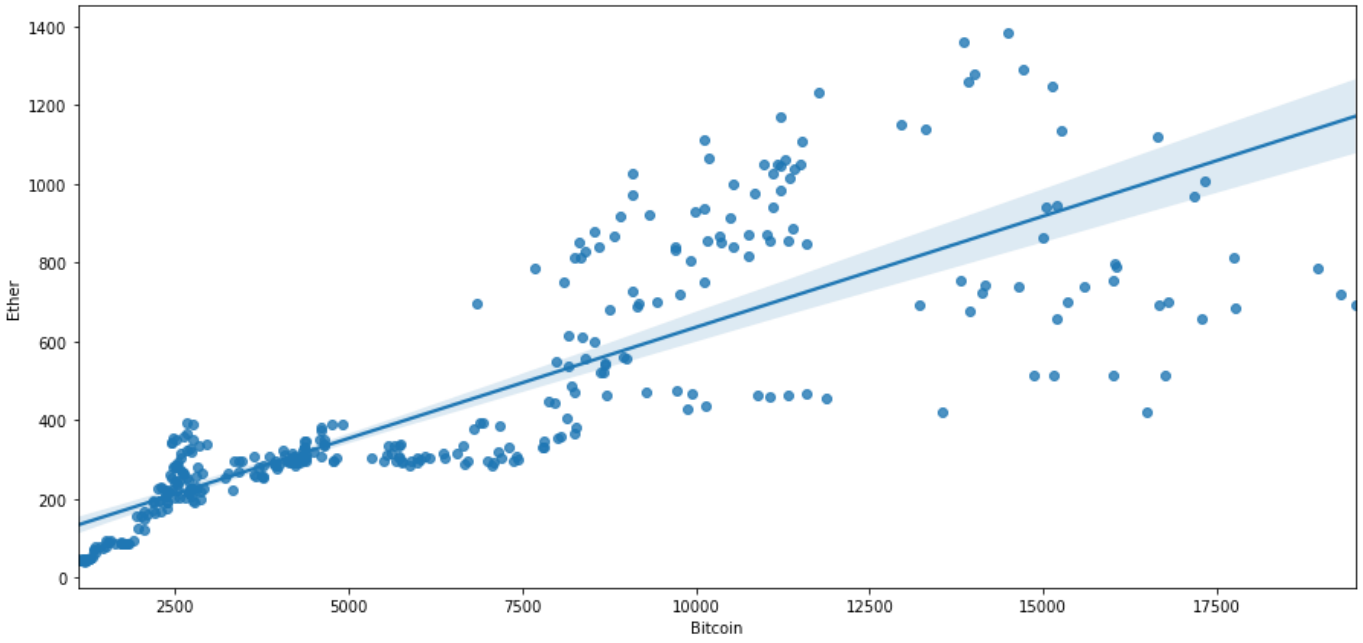
- Mean
- Median
- Mode
- Histogram

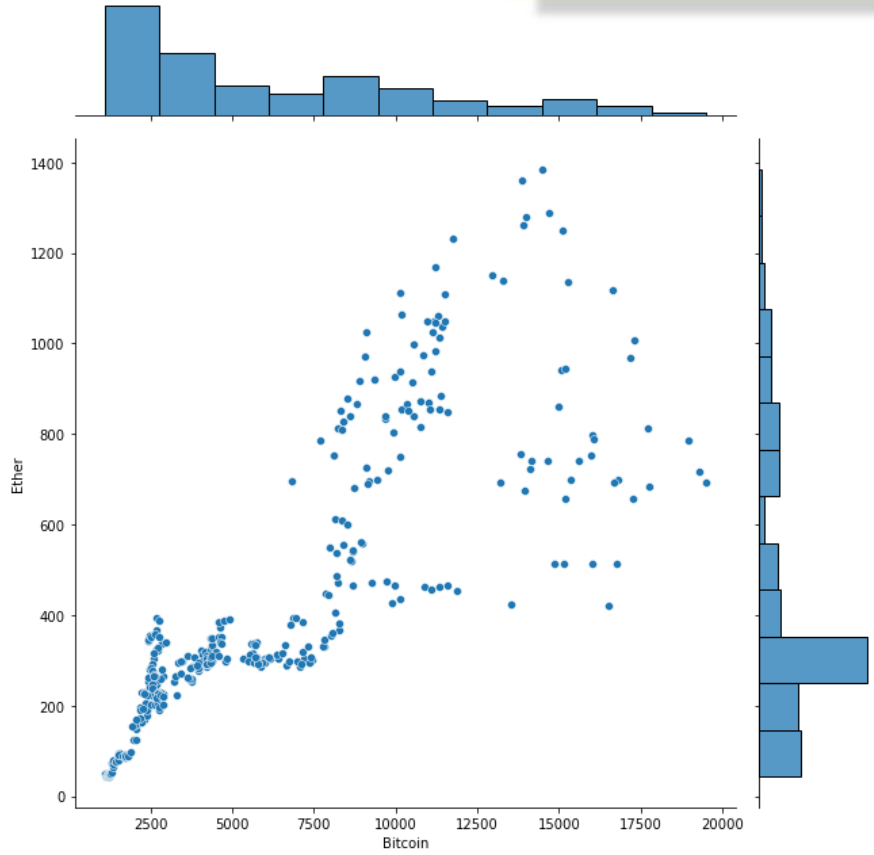
VISUALIZING BIVARIATE DISTRIBUTIONS

Reminder!

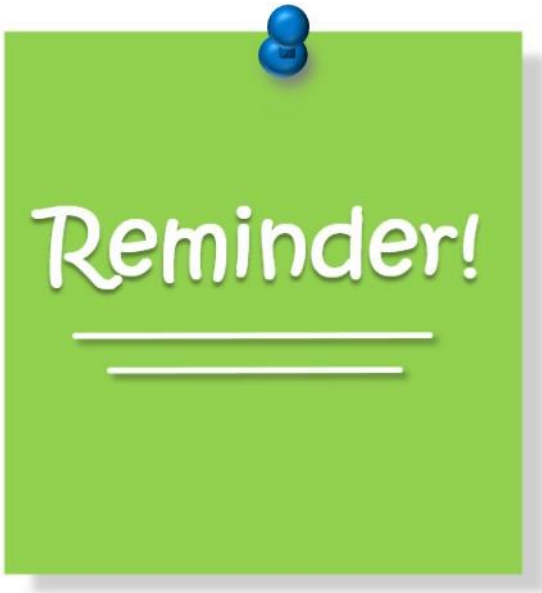
- The most common way to observe a bivariate distribution is a scatterplot,

the `jointplot` will also include the distribution of the variables:





DISPERSION



Use a few methods to measure dispersion in our dataset, most of them well known:

- Range: Range is **really** sensitive to outliers (range = max-min)
- Variance and Standard Deviation: Both variance and std are sensible to outliers as well
- IQR: The Interquartile range is a good measure of "centered" dispersion, and is calculated as `Q3 - Q1` (3rd quartile - 1st quartile). The IQR is more robust than std or range, because it's not so sensitive to outliers.

Note: These methods are sensitive to outliers, so it's recommended to clean the dataset before applying them.

MISSING VALUES WITH `NA_VALUES` PARAMETER

Reminder!

- We can define a `na_values` parameter with the values we want to be recognized as NA/NaN. In this case empty strings `''`, `?` and `-` will be recognized as null values.

	0	1
0	2/4/17 0:00	1099.169125
1	3/4/17 0:00	1141.813
2	4/4/17 0:00	?
3	5/4/17 0:00	1133.079314
4	6/4/17 0:00	-

```
df = pd.read_csv('dataset_name.csv',
                 header=None,
                 na_values=[' ', '?', '-'])
```

	0	1
0	2/4/17 0:00	1099.169125
1	3/4/17 0:00	1141.813000
2	4/4/17 0:00	NaN
3	5/4/17 0:00	1133.079314
4	6/4/17 0:00	NaN