

# Distributions

intro

# Descriptors for distribution

range

quartiles

variance

standard deviation

# Visuals for distributions

Histograms or frequency distributions

box and whisker plots

# Histograms

- bar chart with binned variable on the x axis and counts on the y axis
- sounds a bit like computing the mode!

```
data = np.array([5, 1, 4, 2, 5, 3, 6, 4, 5, 6, 7, 8, 9, 7])
```

```
# find the unique items and counts
```

# Histograms

- bar chart with binned variable on the x axis and counts on the y axis
- sounds a bit like computing the mode!

```
data = np.array([5, 1, 4, 2, 5, 3, 6, 4, 5, 6, 7, 8, 9, 7])
```

```
# find the unique items and counts
```

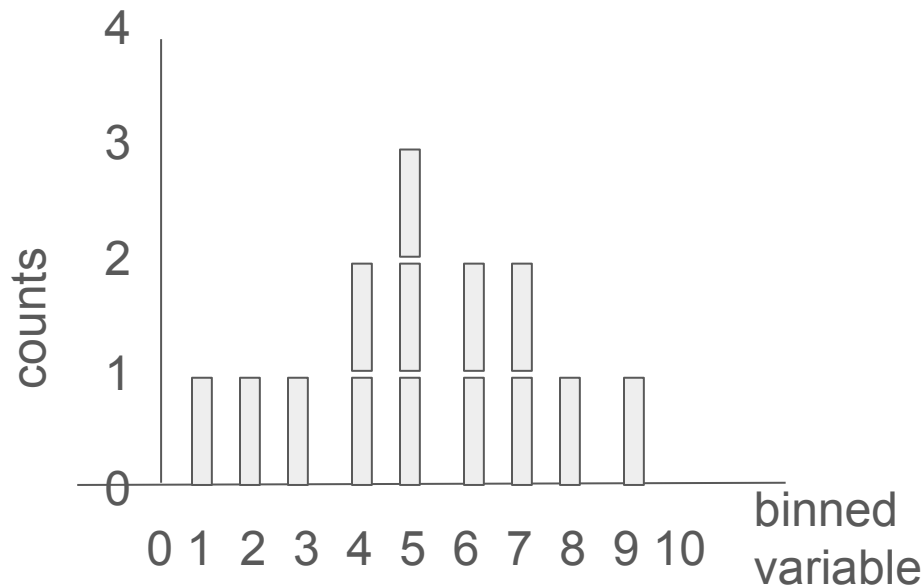
```
unique_items, counts = np.unique(data, return_counts = True)
```

# Histograms

- bar chart with binned variable on the x axis and counts on the y axis
- sounds a bit like computing the mode!

unique\_items: [1, 2, 3, 4, 5, 6, 7, 8, 9]

counts: [1, 1, 1, 2, 3, 2, 2, 1, 1]



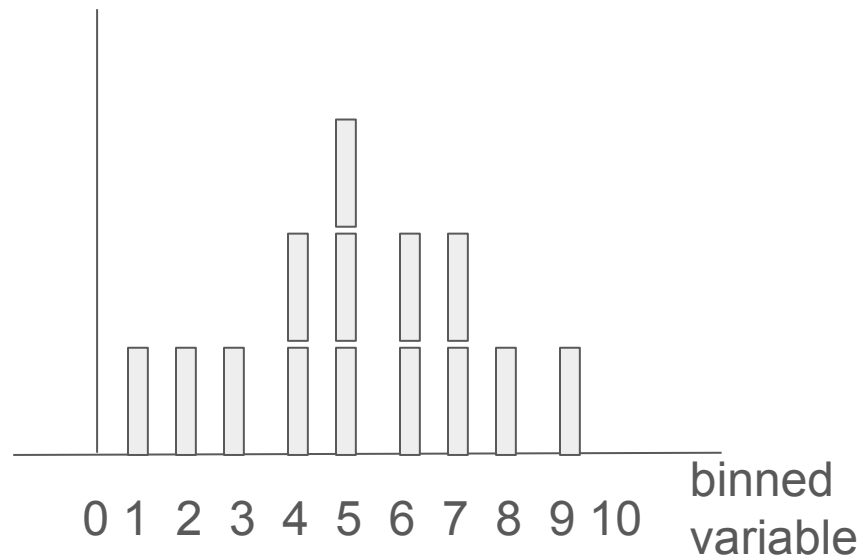
# change the y axis

unique\_items: [1, 2, 3, 4, 5, 6, 7, 8, 9]

counts: [1, 1, 1, 2, 3, 2, 2, 1, 1]

proportion:

percentage:



# change the y axis

unique\_items: [1, 2, 3, 4, 5, 6, 7, 8, 9]

counts: [1, 1, 1, 2, 3, 2, 2, 1, 1]

proportion:

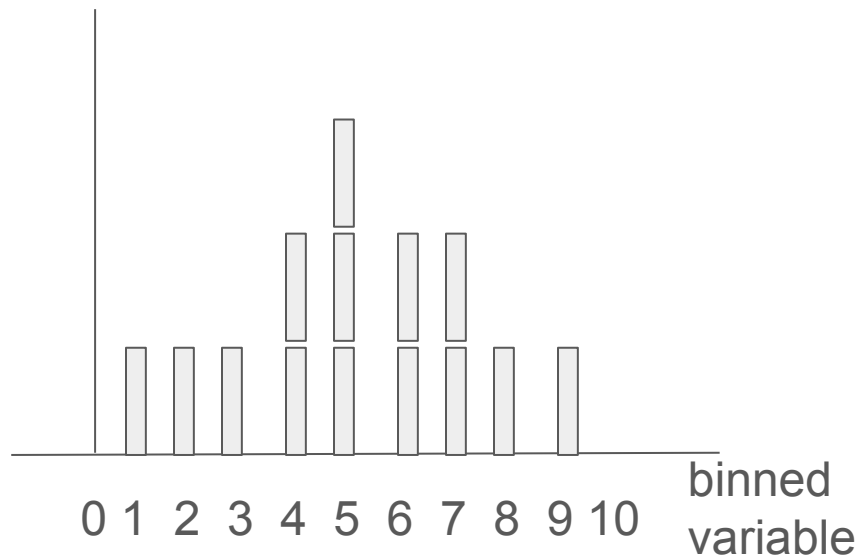
percentage:

# total number of counts

`total = np.sum(counts)`

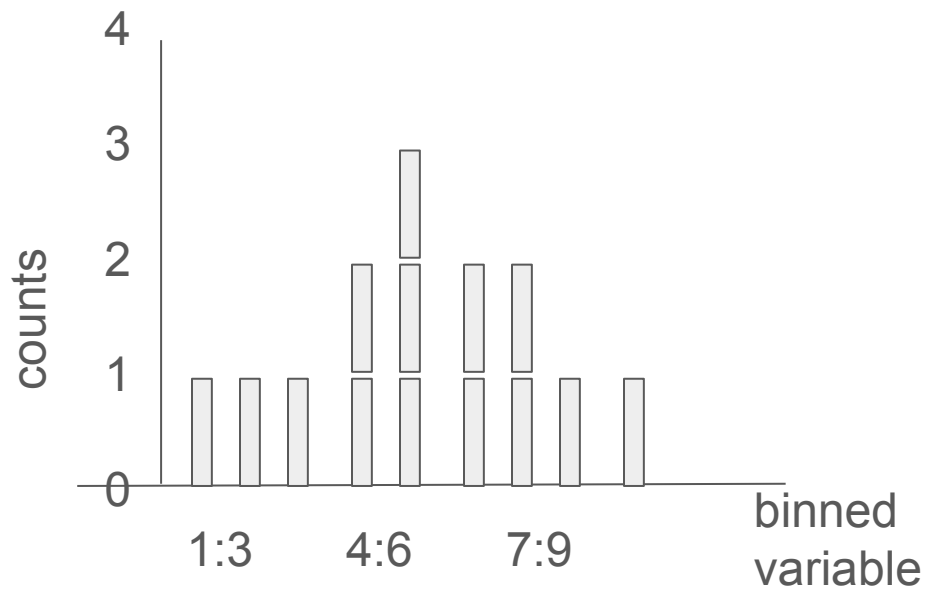
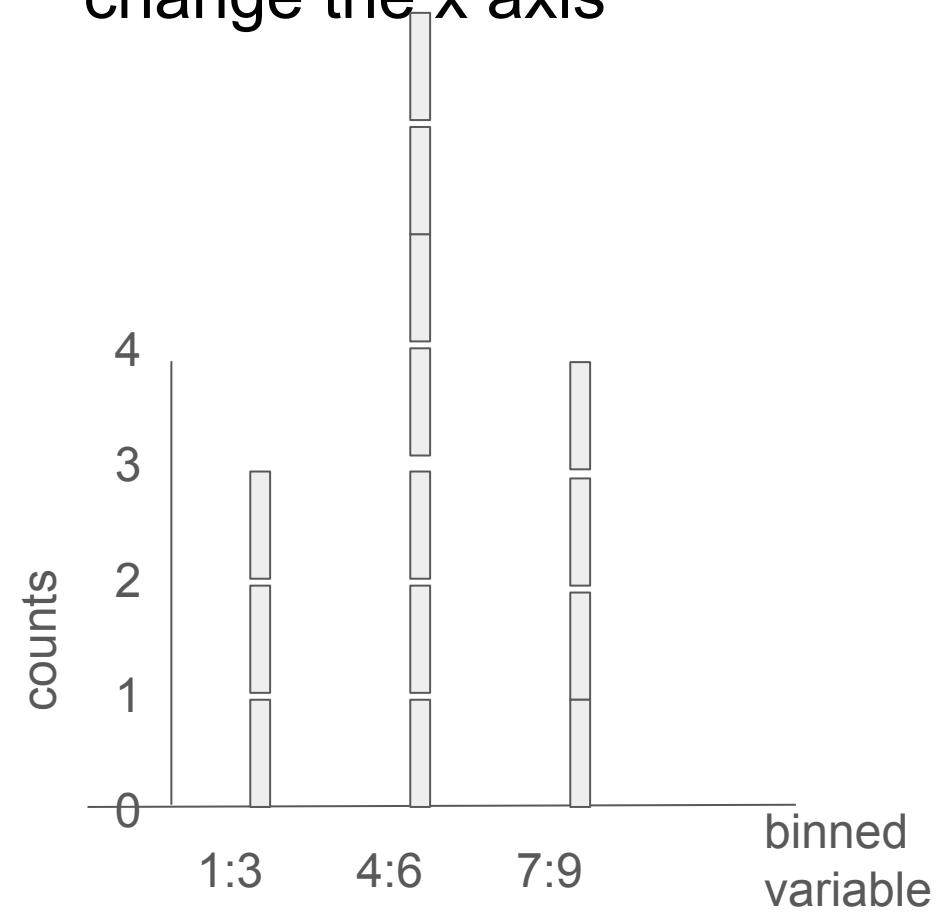
`proportion = counts/total`

`percentage = proportion * 100`

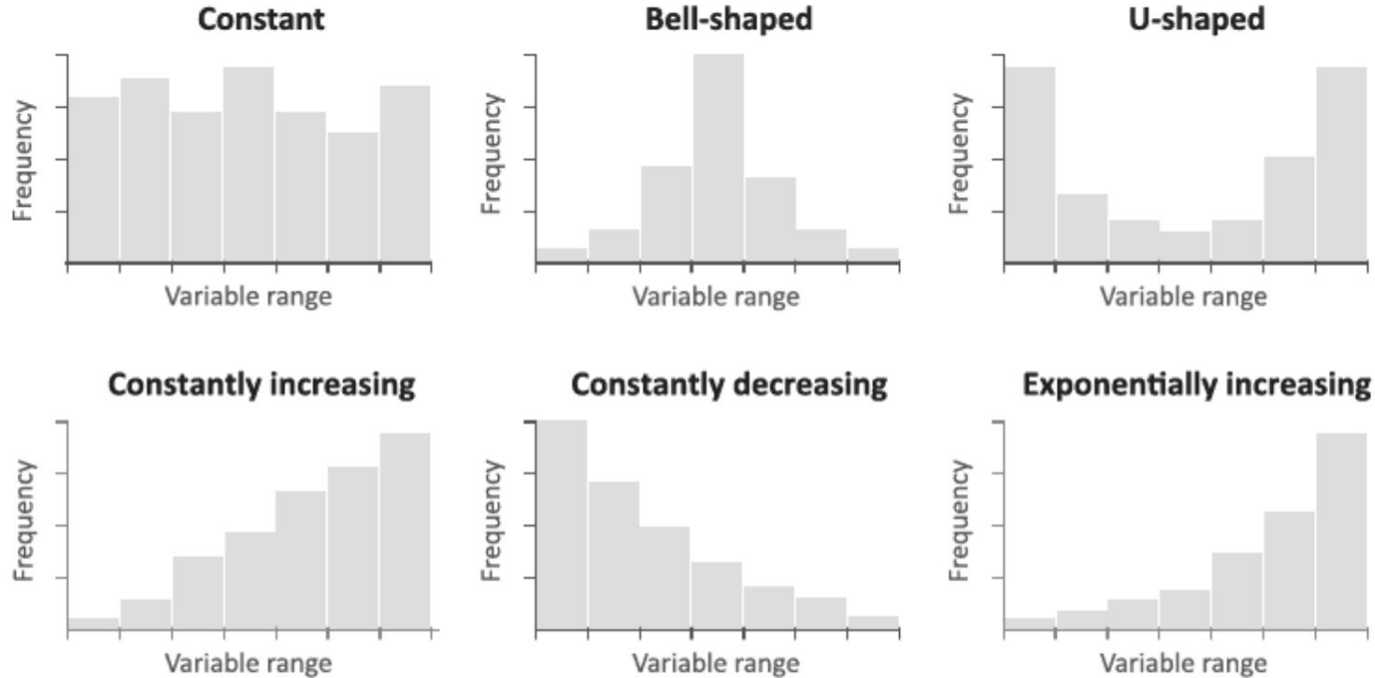




change the x axis



# Names of distribution shapes

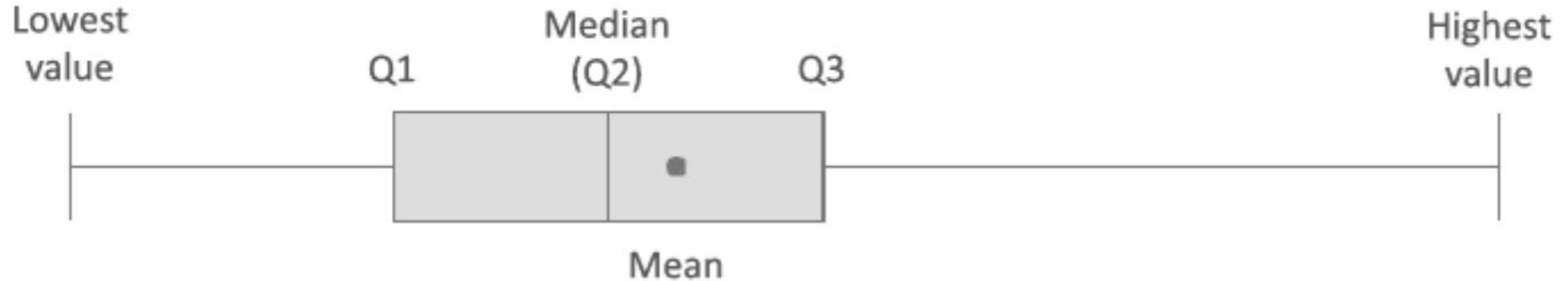


**FIGURE 2.6** Examples of frequency distributions.

# Box plots - quartiles and range

`np.quantile()`

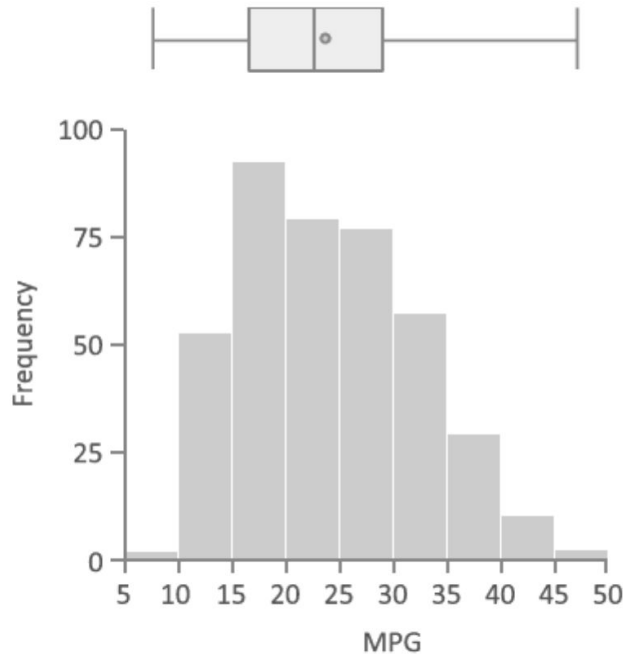
- `np.quantile(a, 0.5)`



**FIGURE 2.8** Overview of elements of a box plot.

# converting distribution into box plot (quartiles and range)

- Let's write the code!



# Confidence Intervals

"Information derived from a sample of observations can only be an approximation of the entire population. To make a definitive statement about an entire population , every member of that population would need to be measured."

Myatt, Glenn J., and Wayne P. Johnson. *Making Sense of Data I : A Practical Guide to Exploratory Data Analysis and Data Mining*, John Wiley & Sons, Incorporated, 2014. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/allegHENY-ebooks/detail.action?docID=1729064>.

- Not everything can be measured
- We need to make estimates based on existing data
- This leads to ability to say something is "statistically significant"

# Hypothesis Testing

Future!