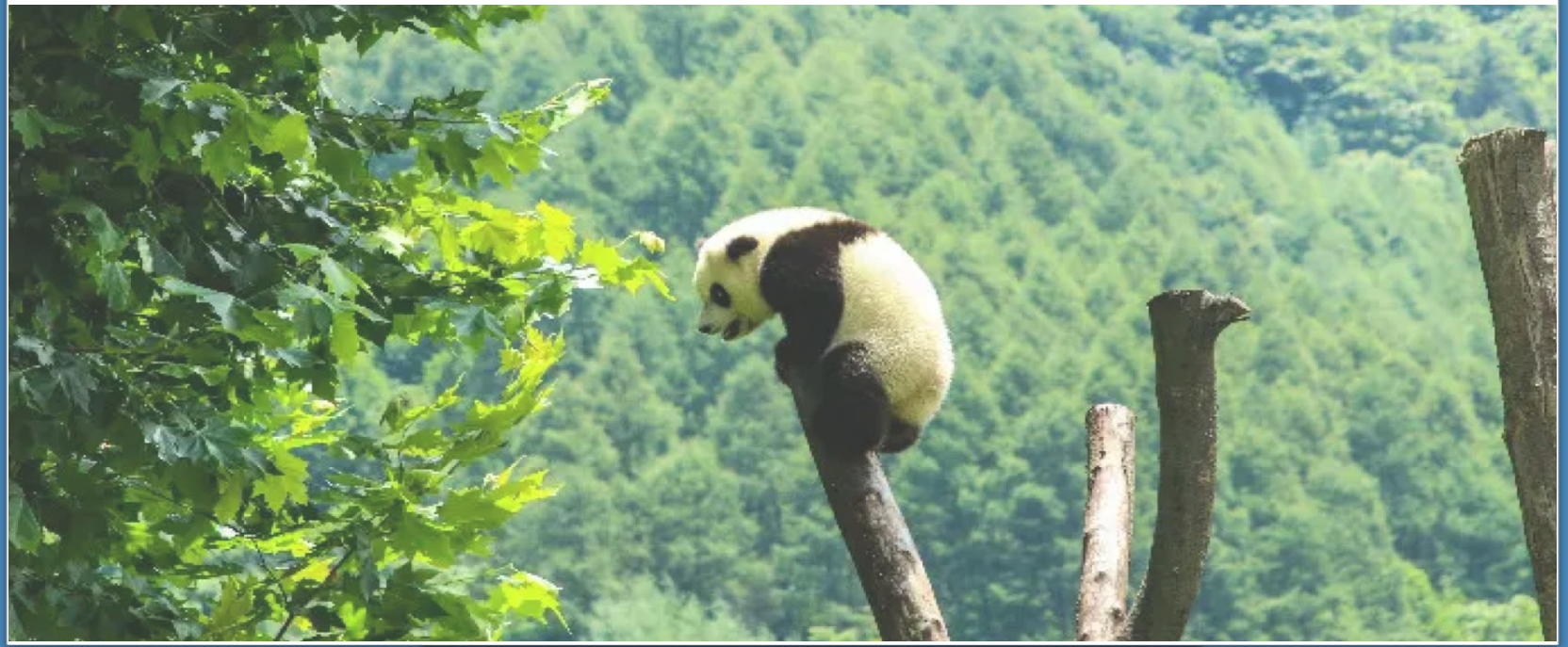
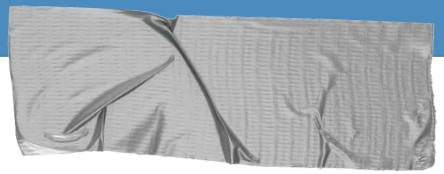
pandas

FLATIRON SCHOOL

# Guided Practice

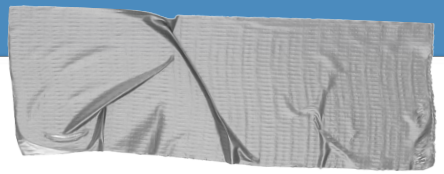




# Pre Quiz

# Agenda

- **pandas Intro**
- **Variables Intro**
- **Quantitative Variables**
  - Continuous & Discrete
- **Categorical Variables**
  - Ordinal, Nominal, and Binary
- **Variables with Pandas**
  - Explore variables in python
- **Q&A**
  - Opportunity to delve into questions



# Pandas Resource: https://pandas.pydata.org/

pandas

About us | Getting started | Documentation | Community | Contribute

## pandas

**pandas** is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Install pandas now!

### Getting started

- Install pandas
- Getting started

### Documentation

- User guide
- API reference
- Contributing to pandas
- Release notes

### Community

- About pandas
- Ask a question
- Ecosystem

With the support of:

NUMFOCUS OPEN CODE = BETTER SCIENCE | TWO SIGMA | URSA LABS | d-fine | TIDELIFT | Chan Zuckerberg Initiative

The full list of companies supporting *pandas* is available in the sponsors page.

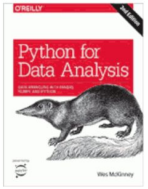
### Latest version: 1.4.2

- What's new in 1.4.2
- Release date: Apr 02, 2022
- Documentation (web)
- Documentation (pdf)
- Download source code

### Follow us

Follow @pandas_dev

### Get the book




### Previous versions

- 1.4.1 (Feb 12, 2022) changelog | docs | pdf | code
- 1.4.0 (Jan 22, 2022) changelog | docs | pdf | code
- 1.3.5 (Dec 12, 2021)

# pandas Intro

Pandas is a NumFOCUS sponsored library built on NumPy.


We can import pandas by calling:

```
In [1]: import numpy as np

In [2]: import pandas as pd
```

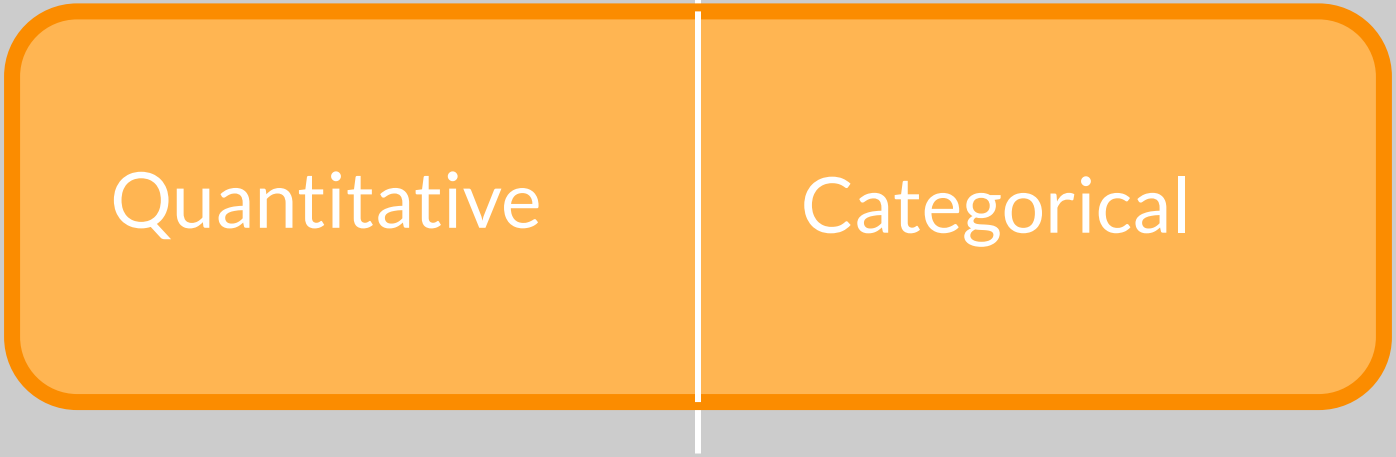
# Pandas DataFrame

```
titanic.head()
```

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare |
|---|---|---|---|---|---|---|---|---|---|---|
| **12** | 904 | 1 | 1 | Snyder, Mrs. John Pillsbury (Nelle Stevenson) | female | 23.0 | 1 | 0 | 21228 | 82.2667 |
| **14** | 906 | 1 | 1 | Chaffee, Mrs. Herbert Fuller (Carrie Constance... | female | 47.0 | 1 | 0 | W.E.P. 5734 | 61.1750 |
| **24** | 916 | 1 | 1 | Ryerson, Mrs. Arthur Larned (Emily Maria Borie) | female | 48.0 | 1 | 3 | PC 17608 | 262.3750 |
| **26** | 918 | 1 | 1 | Ostby, Miss. Helene Ragnhild | female | 22.0 | 0 | 1 | 113509 | 61.9792 |
| **28** | 920 | 0 | 1 | Brady, Mr. John Bertram | male | 41.0 | 0 | 0 | 113054 | 30.5000 |

# Variables

**parameters that can hold a set of values.**



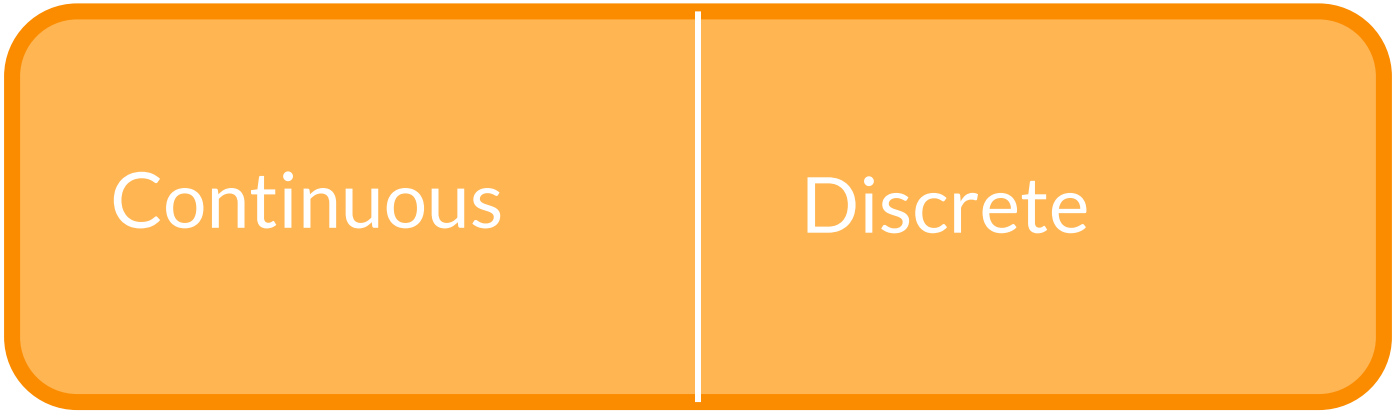
# Variables in a Dataset

```
titanic.head()
```

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare |
|---|---|---|---|---|---|---|---|---|---|---|
| **12** | 904 | 1 | 1 | Snyder, Mrs. John Pillsbury (Nelle Stevenson) | female | 23.0 | 1 | 0 | 21228 | 82.2667 |
| **14** | 906 | 1 | 1 | Chaffee, Mrs. Herbert Fuller (Carrie Constance... | female | 47.0 | 1 | 0 | W.E.P. 5734 | 61.1750 |
| **24** | 916 | 1 | 1 | Ryerson, Mrs. Arthur Larned (Emily Maria Borie) | female | 48.0 | 1 | 3 | PC 17608 | 262.3750 |
| **26** | 918 | 1 | 1 | Ostby, Miss. Helene Ragnhild | female | 22.0 | 0 | 1 | 113509 | 61.9792 |
| **28** | 920 | 0 | 1 | Brady, Mr. John Bertram | male | 41.0 | 0 | 0 | 113054 | 30.5000 |

# Quantitative Variables

**amounts or counts; for example income, age, and household size.**



# Continuous Variables

**can contain measurements with infinite decimal precision, for example the height of a person.**

# Discrete Variables

**can contain counts that must be whole integer values, such as the number of members in a family.**

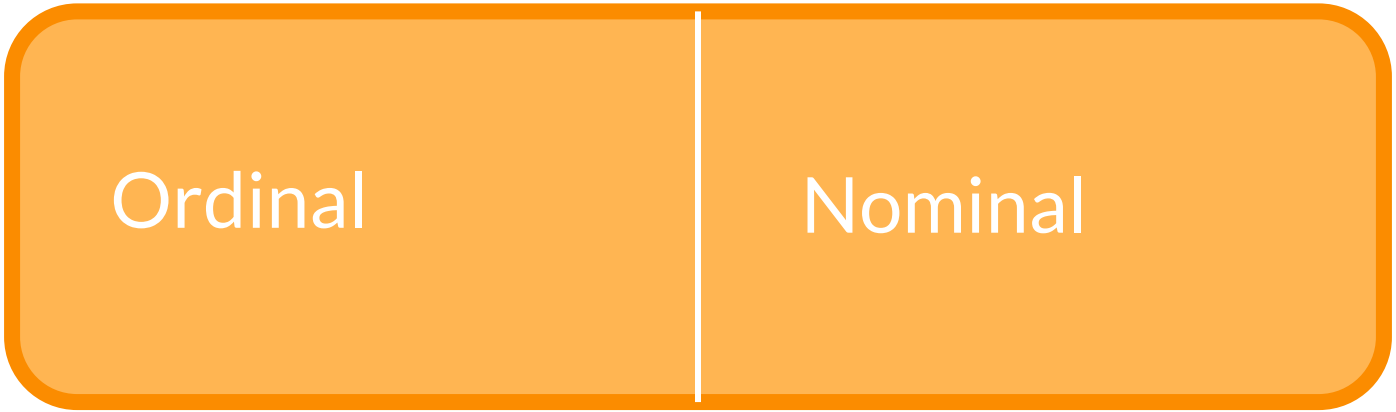
# Quantitative Variables

```
titanic.head()
```

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare |
|---|---|---|---|---|---|---|---|---|---|---|
| **12** | 904 | 1 | 1 | Snyder, Mrs. John Pillsbury (Nelle Stevenson) | female | 23.0 | 1 | 0 | 21228 | 82.2667 |
| **14** | 906 | 1 | 1 | Chaffee, Mrs. Herbert Fuller (Carrie Constance... | female | 47.0 | 1 | 0 | W.E.P. 5734 | 61.1750 |
| **24** | 916 | 1 | 1 | Ryerson, Mrs. Arthur Larned (Emily Maria Borie) | female | 48.0 | 1 | 3 | PC 17608 | 262.3750 |
| **26** | 918 | 1 | 1 | Ostby, Miss. Helene Ragnhild | female | 22.0 | 0 | 1 | 113509 | 61.9792 |
| **28** | 920 | 0 | 1 | Brady, Mr. John Bertram | male | 41.0 | 0 | 0 | 113054 | 30.5000 |

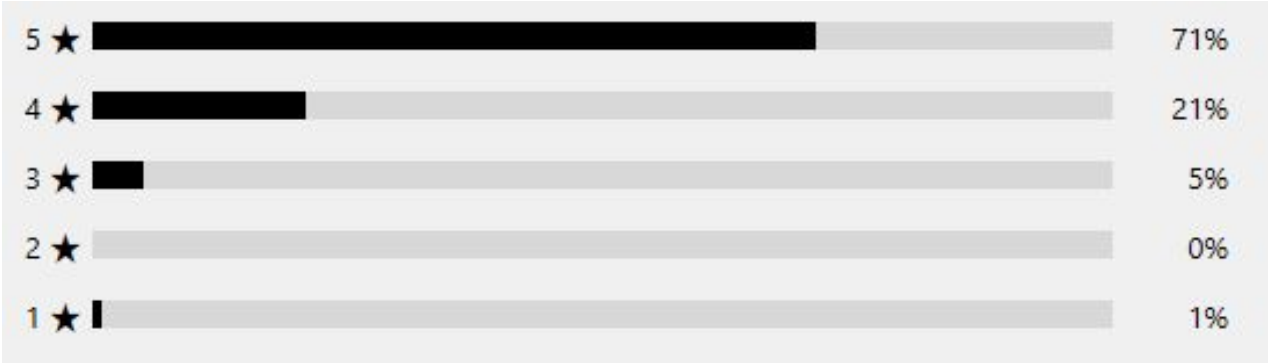
# Categorical Variables

**data that can be grouped into discrete categories.**



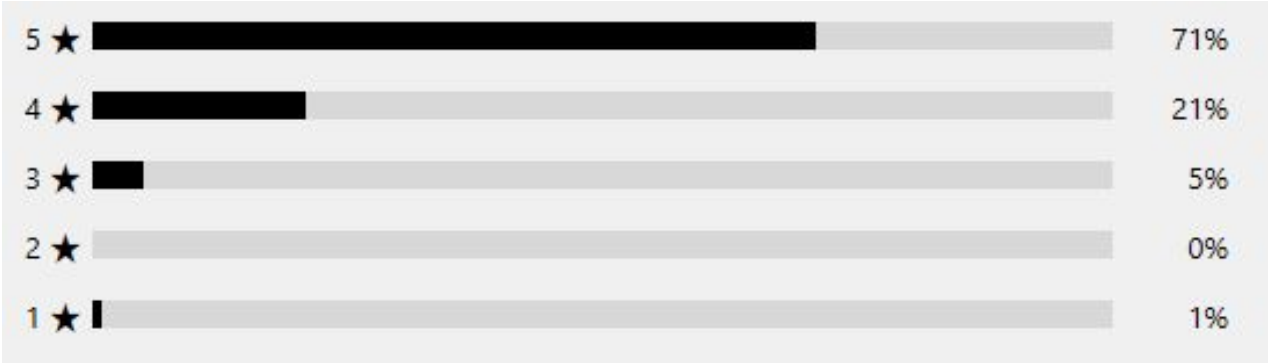
# Ordinal Variables

**groups that contain an inherent ranking, such as ratings or a point scale.**



# Ordinal vs. Discrete

**there is a uniform difference between discrete variables not necessarily present with ordinal.**



# Nominal Variables

made of categories without an inherent order, such as species of penguin or people's hair color

# Binary Variables

**nominal variables that contain only two mutually exclusive categories.**

# Categorical Variables

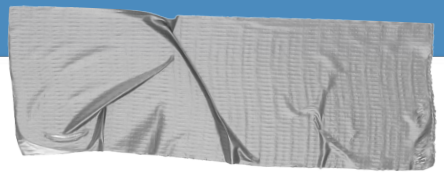
```
titanic.head()
```

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare |
|---|---|---|---|---|---|---|---|---|---|---|
| **12** | 904 | 1 | 1 | Snyder, Mrs. John Pillsbury (Nelle Stevenson) | female | 23.0 | 1 | 0 | 21228 | 82.2667 |
| **14** | 906 | 1 | 1 | Chaffee, Mrs. Herbert Fuller (Carrie Constance... | female | 47.0 | 1 | 0 | W.E.P. 5734 | 61.1750 |
| **24** | 916 | 1 | 1 | Ryerson, Mrs. Arthur Larned (Emily Maria Borie) | female | 48.0 | 1 | 3 | PC 17608 | 262.3750 |
| **26** | 918 | 1 | 1 | Ostby, Miss. Helene Ragnhild | female | 22.0 | 0 | 1 | 113509 | 61.9792 |
| **28** | 920 | 0 | 1 | Brady, Mr. John Bertram | male | 41.0 | 0 | 0 | 113054 | 30.5000 |

# Now let's practice!

## Join on GitHub.

https://github.com/mojo-flat/pandas_gp_mojo



# Q&A