

# What's new in pandas?

PyData 2014

<https://github.com/jreback/pydata2014-pandas>

# Jeff Reback

*@jreback*

- former quant
- currently working on projects at Continuum
- core committer to pandas for last 3 years
- manage pandas since mid 2013

# So, what's new in pandas?

- Categorical Data
- Timedeltas
- MultiIndex Slicers
- **.dt** accessor
- Grouping Enhancements

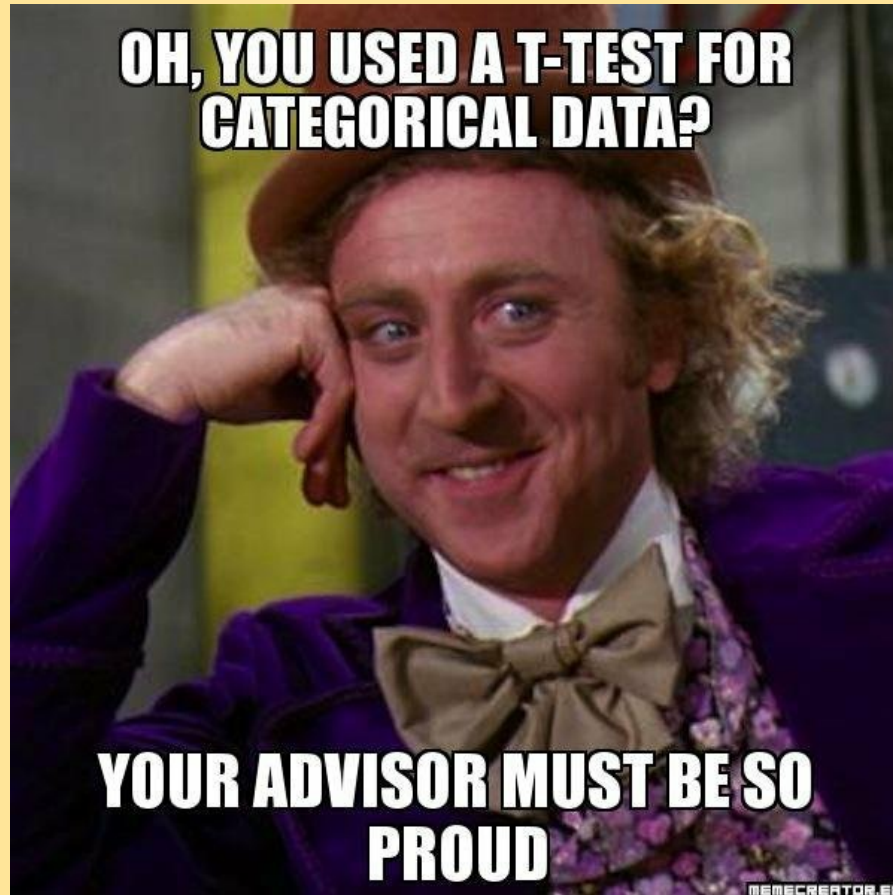
How are these changes even possible?

- ***Series*** Refactor in 0.13
- ***Index*** Refactor in 0.15
- Internal Blocking

# Categorical Data

- Discrete & Limited Set of Values
  - Gender: Male, Female
  - Meal: Breakfast, Lunch, Dinner
- Possibly ordered by something
  - non-lexical
- Representation saves memory
- Codes have meaning (export)

# Categorical Data



# Categorical Data



# Timedeltas

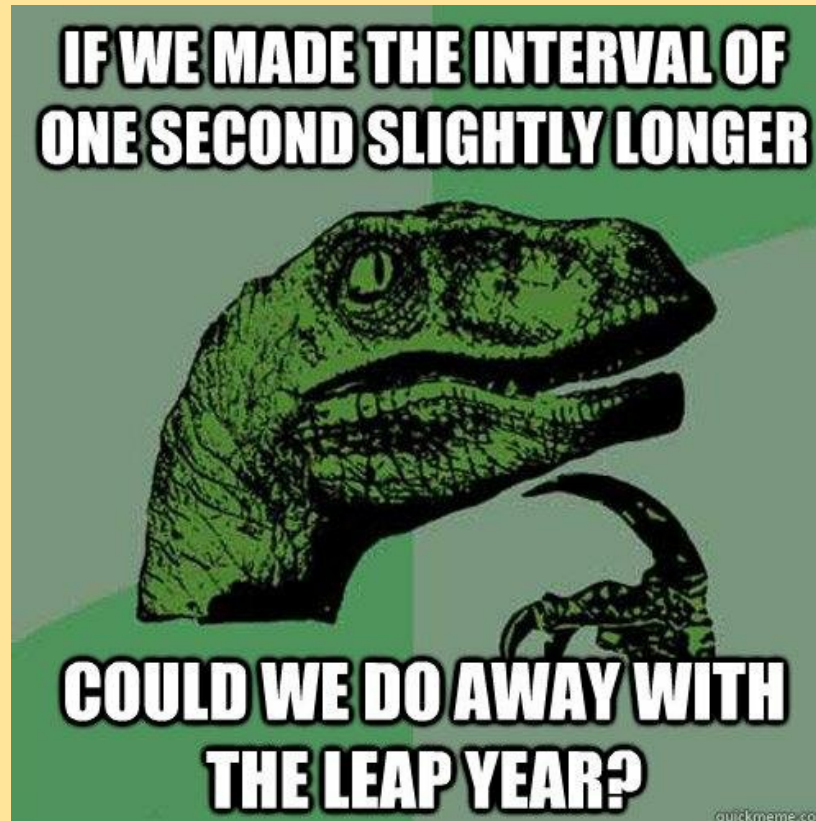
- Track Intervals of Time
- Diff Datetimes / DatetimeIndexes
- 0-based time-like data



# Timedeltas



# Timedeltas



# Timedeltas



# MultiIndex Slicers

- Introduced in 0.14
- Allow multi-dimensional index slicing 'easily'
- Replace the need for **.xs**
- Allow setting
- Future Thoughts

# MultIndex Slicers



## **.dt** accessors

- disambiguate values vs index
- allow the select attribute based on the type of object
- selective tab completion

```
In [2]:
```

```
idx = pd.date_range('20130101 01:23:45.678', periods=3, freq='12H')
s = Series(idx, index=idx+pd.Timedelta('1d'))
s
```

```
Out[2]:
```

```
2013-01-02 01:23:45.678000    2013-01-01 01:23:45.678000
2013-01-02 13:23:45.678000    2013-01-01 13:23:45.678000
2013-01-03 01:23:45.678000    2013-01-02 01:23:45.678000
Freq: 12H, dtype: datetime64[ns]
```

# Groupby Enhancements

- `pd.Grouper`
- chained resample

# What's in the future?

- IntervalIndex
- CategoricalIndex
- pd.String
- TZ aware blocks
- dynd
- bcolz / virtualized DataFrames

<https://github.com/ContinuumIO/dynd-python>

<https://github.com/Blosc/bcolz>