

Thursday, November 19, 2020 | **Class #14**

# SciPy (linear regression, 1-D interpolation)

OCEAN 215 | Autumn 2020

**Ethan Campbell** and Katy Christensen

# Linear regression in SciPy

---

Correlation coefficient ( $r$ )

Standard error

Two-sided  $p$ -value

`slope, intercept, rvalue, pvalue, stderr`  
`= stats.linregress(x, y)`

1-D NumPy arrays of the same length

# What if your x-values are `datetime` objects?

---

```
1 import matplotlib.dates as mdates
2
3 t = np.array([datetime(2020,1,1),
4               datetime(2020,2,1),
5               datetime(2020,3,1)])
6
7 t_as_numbers = mdates.date2num(t)
8
9 print(t_as_numbers)
```

← `linregress()` can't handle an array of `datetime` objects as x-values



← This converts `datetime` objects to numbers representing “days since 0001-01-01 plus one”, which `linregress()` can handle

```
[737425. 737456. 737485.]
```



# 1-D interpolation in SciPy is a two-step process

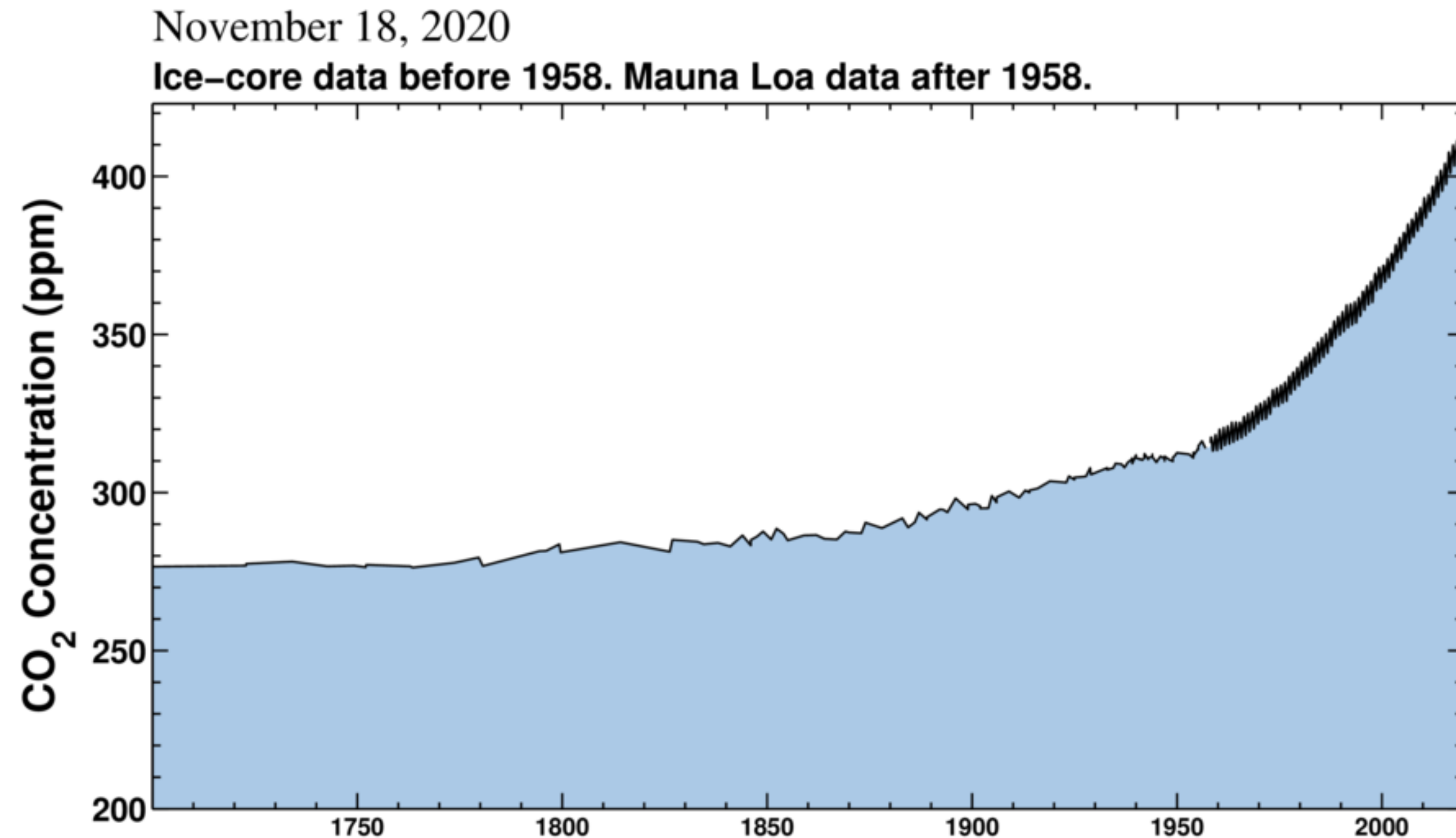
---

```
interp_func = interpolate.interp1d(x, y,  
                                   kind='linear',  
                                   bounds_error=False,  
                                   fill_value=np.NaN)
```

```
y_new = interp_func(x_new)
```

# Activity: global temperature measurements

---



Google Doc with activities (also accessible from Canvas Modules or Google Drive folder):

**<https://tinyurl.com/OCEAN215-Class14>**

# Think/pair/share, then start cleaning and analyzing final project data

---

- **Round 1 (5 minutes):** Try to answer the following questions:
  1. What's the format of your data, e.g. CSV, netCDF, other? What function(s) have you used to load the data, or how are you planning to load the data? Are there any problems (or do you anticipate any problems) with the tidiness or structure of the data?
  2. How big is your data, in MB or GB? What is its shape, e.g. columns/rows or shape of variable arrays?
  3. What NumPy/SciPy quantitative analyses or Matplotlib/Cartopy visualizations are you hoping/planning to make?
- **Round 2 (5 minutes):** As a pair (or with another pair), share with each other what data you're using and your answers to these questions.
- **Round 3 (10 minutes):** As a class, we'll ask a few of you to share your answers to these questions.
- **Round 4 (rest of class):** Work time on cleaning and analyzing your data!