

1 Python 101: Homework

1.1 By Evelyn J. Boettcher

1.2 Week 3 Lesson 1: Linear Regression

1.2.1 Problem 1

- Explain what `np.polynomial` does in this script. ** And what are my options besides `Polynomial`

```
import numpy as np
import pandas as pd

df = pd.read_csv('../data/co2_weekly_mlo.txt', skiprows=49,
                 names=['yr', 'mon', 'day', 'decimal', 'ppm', '#days', '1 yr ago', '10 yr ago', 's'],
                 delim_whitespace=True)
clean_df = df[df.ppm != -999.99]
pp = np.polynomial.Polynomial(np.polyfit(clean_df.decimal, clean_df.ppm, 1))
plt.scatter(clean_df.decimal, clean_df.ppm)
plt.plot(clean_df.decimal, pp(clean_df), color='red')
plt.show()
```

1.2.2 Problem 2

- `print(pp)`

1.2.3 Problem 3

- [] What happens if you try to run this

```
import scipy as sp
sp.stats.linregress(clean_df.decimal, clean_df.ppm)
```

1.2.4 Problem 4

- [] For the Seaborn example
 - [] Change the order to: 2 or 3

```
import seaborn as sns; sns.set_theme(color_codes=True)
import pandas as pd
import matplotlib.pyplot as plt

# Get Data
df = pd.read_csv('co2_weekly_mlo.txt', skiprows=49,
                 names=['yr', 'mon', 'day', 'decimal', 'ppm', '#days', '1 yr ago', '10 yr ago', 'since 1880'],
                 delim_whitespace=True)
clean_df = df[df.ppm != -999.99]

# Set the plots
ax = sns.regplot(x='decimal', y='ppm', data=clean_df, ci=95, order=1,
                 line_kws={'label': 'Linear regression line: $Y(X)=-3827 + 1.8 X$', 'color': 'm'},
                 label="CO2 Weekly average")
```

```
ax.set_ylabel("CO2 (ppm)")
ax.set_xlabel("Year")
ax.legend(loc="upper left")
plt.show()
```

1.2.5 Problem 5 (a,b)

- [] Add x'bins = 100
 - [] Add x'bins = 50
-

1.3 Solutions

1.3.1 Problem 1

WOW: I can do a Chebyshev and a Laguerre or Legendre

1.3.2 Problem 2

Note this is different from print(pp)

```
pp
Polynomial([ 1.80476055e+00, -3.23755696e+03], domain=[-1, 1], window=[-1, 1])
```

LOOK my coefficients!!

```
print(pp)
# poly([ 1.80476055e+00 -3.23755696e+03])
```

Look: No domain or window information

1.3.3 Problem 3

It doesn't work. This is because the stats function is really large. Python is being smart and only loading the basics.

1.3.4 Problem 4

```
ax = sns.regplot(x='decimal', y='ppm', data=clean_df, ci=95, order=2)
```

It does a quadratic that easily!

1.3.5 Problem 5

```
ax = sns.regplot(x='decimal', y='ppm', data=clean_df, order=1, x_bins=50)
```

x'bins 150 with second order

x'bins 150 with 1st order LOOK: It gives me the "error" / range of the bins automatically

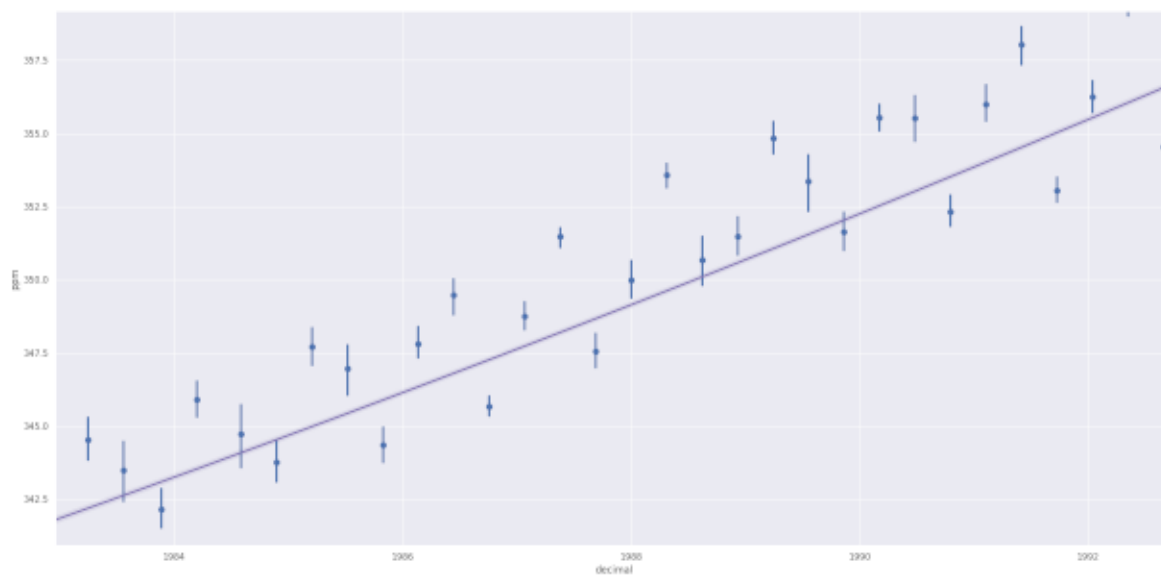


Figure 1: x'bins

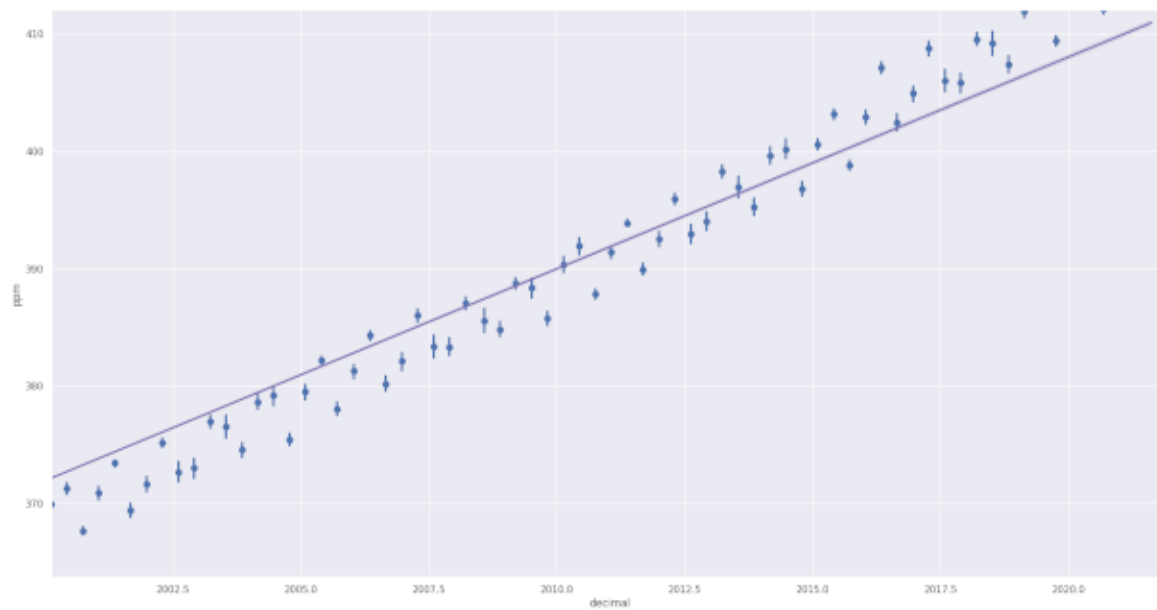


Figure 2: x'bins