CP.4.2.9, Sauer3

The file `scrippsm.txt`, available from the textbook website, is a list of 180 numbers which represent the concentration of atmospheric carbon dioxide, in parts per million by volume (ppv), recorded monthly at Mauna Loa from Jan. 1996 to Dec. 2010, taken from the same Scripps study as Computer Problem 8.

a. Carry out a least squares fit of the CO_2 data using the model

$$f(t) = c_1 + c_2 t + c_3 \cos(2\pi t) + c_4 \sin(2\pi t)$$

where t is measured in months. Report the best fit coefficients c_i and the RMSE of the fit. Plot the continuous curve from Jan. 1989 to the end of this year, including the 180 data points in the plot.

b. Use your model to predict the CO_2 concentration in May 2004, Sept. 2004, May 2005, and Sept. 2005. These months tend to contain the yearly maxima and mimima of the CO_2 cycle. The actual recorded values are 380.63, 374.06, 382.45, and 376.73 ppv, respectively. Report the model error at these four points.

c. Add the extra term $c_5 \cos(4\pi t)$ and redo part (a) and (b). Compare the new RMSE and the four model errors.

d. Repeat part (c) using the extra term $c_5 t^2$. Which terms lead to more improvement in the model, part(c) or (d)?

e. Add both terms from (c) and (d) and redo parts (a) and (b). Prepare a table summarizing your results from all parts of the problem, and try to provide an explanation for the results.

See the website `http://scrippsco2.ucsd.edu` for much more data and analysis of the Scripps carbon dioxide study.

Do not worry about the questions of Sauer. This problem is (more or less) solved in the following notebook. You have one Colab question and one Handwritten question to answer. See below.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Jan.	362.05	363.18	365.32	368.15	369.14	370.28	372.43	374.68	376.78	378.34	381.37	382.49	385.02	386.71	388.56
Feb.	363.25	364.00	366.15	368.87	369.46	371.50	373.09	375.63	377.36	379.66	382.02	383.72	385.87	387.17	389.98
Mar.	364.02	364.57	367.31	369.59	370.52	372.12	373.52	376.11	378.40	380.38	382.63	384.33	385.85	388.62	390.99
Apr.	364.72	366.35	368.61	371.14	371.66	372.87	374.86	377.65	380.50	382.14	384.40	386.23	386.77	389.51	392.52
May	365.41	366.80	369.30	371.00	371.82	374.02	375.55	378.35	380.61	382.24	384.94	386.43	388.49	390.18	393.22
Jun.	364.97	365.62	368.87	370.35	371.70	373.30	375.40	378.13	379.55	382.10	384.08	385.87	387.92	389.60	392.29
Jul.	363.65	364.47	367.64	369.27	370.12	371.62	374.02	376.61	377.77	380.66	382.37	384.44	386.32	388.01	390.49
Aug.	361.48	362.51	365.78	366.93	368.12	369.55	371.49	374.49	375.84	378.68	380.48	381.84	384.17	386.07	388.55
Sep.	359.45	360.19	363.90	364.63	366.62	367.96	370.71	372.98	374.05	376.40	378.78	380.86	383.00	384.61	386.54
Oct.	359.60	360.77	364.23	365.13	366.73	368.09	370.25	373.00	374.22	376.79	379.07	380.88	382.81	384.34	386.21
Nov.	360.76	362.43	365.46	366.68	368.29	369.68	372.08	374.35	375.84	378.32	380.17	382.40	384.06	386.02	388.61
Dec.	362.33	364.28	366.97	368.00	369.53	371.24	373.78	375.69	377.44	380.02	381.67	383.72	385.15	387.36	389.83

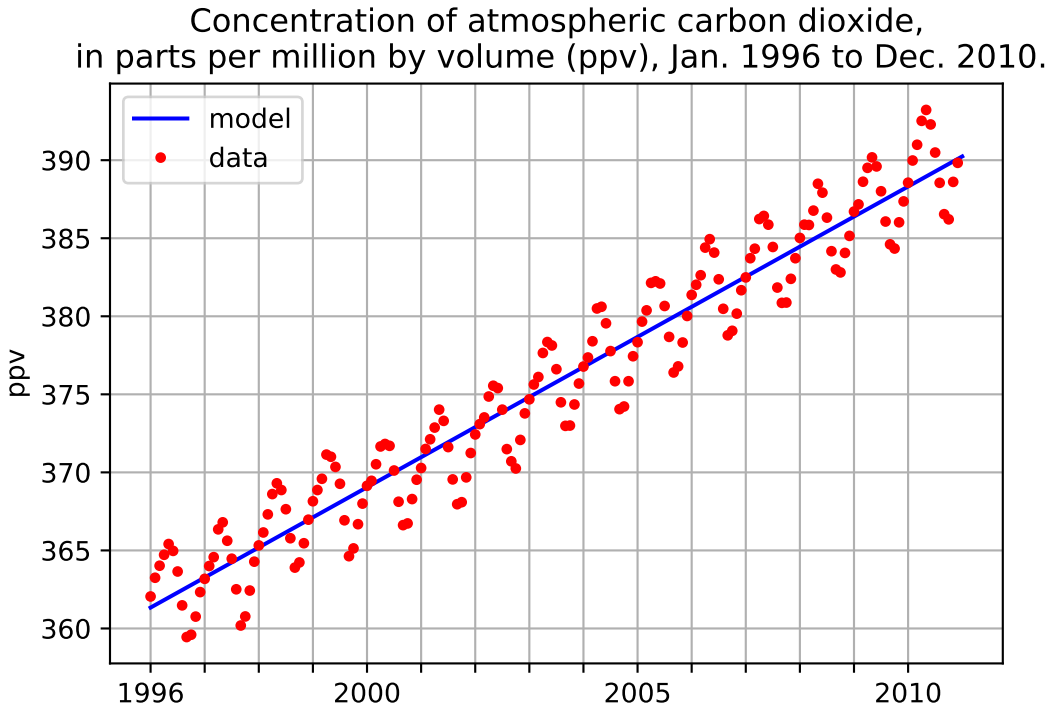
Concentration of atmospheric carbon dioxide, in parts per million by volume (ppv), recorded monthly from Jan. 1996 to Dec. 2010

Getting started: `https://colab.research.google.com/drive/1juTBIlQWf06NjXEzScfy2CNc4MPUJOS-`

Colab Question: The jupyter notebook below does the best fit with a function of the type

$$c_0 + c_1 t$$

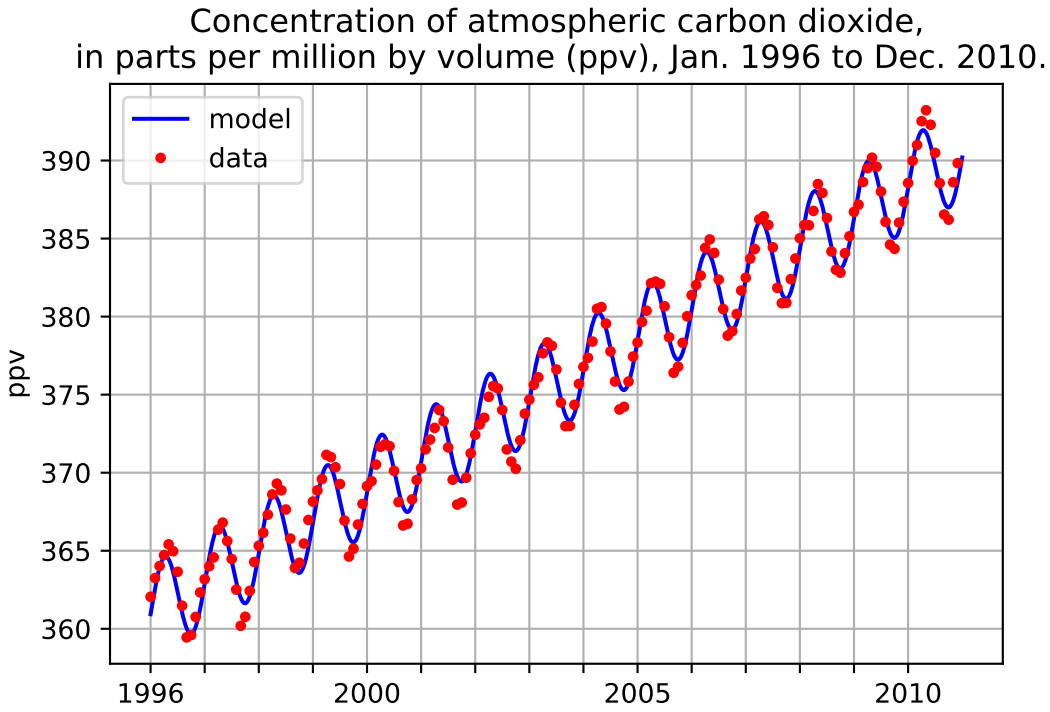
We get the following picture:



Change the model so that we get a best fit with a function of the type

$$c_0 + c_1 t + c_2 \cos(2\pi t) + c_3 \sin(2\pi t)$$

You should get the following picture:



You only need to change two lines of code:

```
A = np.array([ np.ones( t.shape ), t ]).T
```

and

```
yyy = 279. + x[0] + x[1] * xxx
```

Handwritten Question: Give the RMSEs of the two models $c_0 + c_1 t$ and of $c_0 + c_1 t + c_2 \cos(2\pi t) + c_3 \sin(2\pi t)$. Compare the RMSEs. Which model is better? Why?