

# How do I design a class in Python?

Asked 14 years, 1 month ago    Modified 6 years, 1 month ago

Viewed 74k times

---



**150**

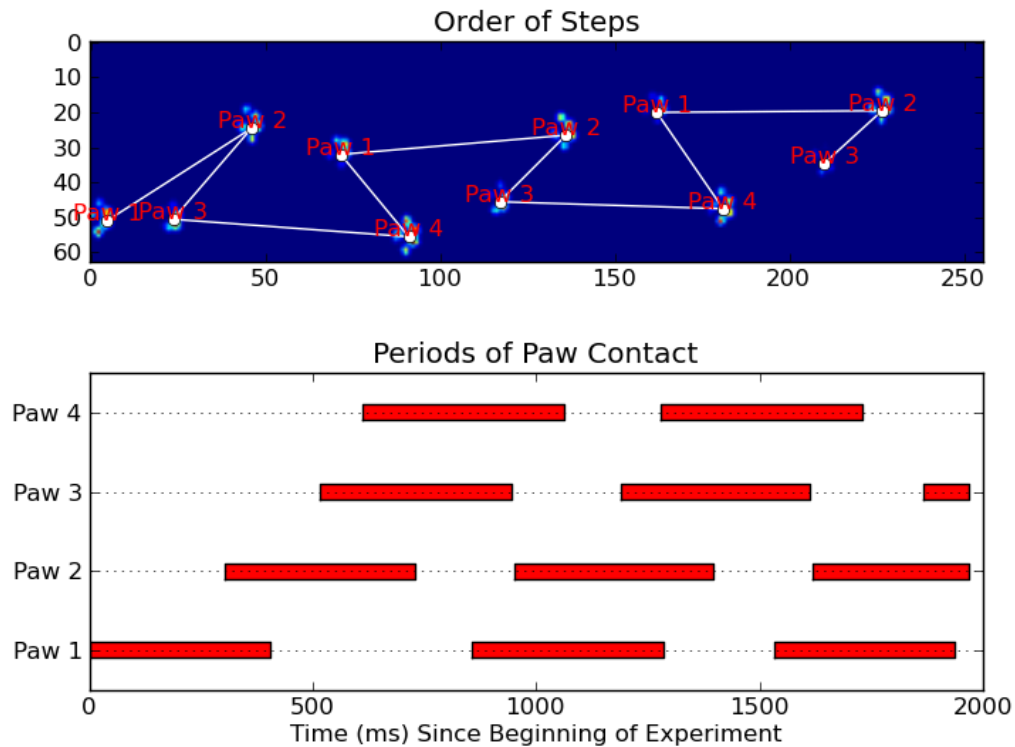


I've had some really awesome help on my previous questions [for detecting paws](#) and [toes within a paw](#), but all these solutions only work for one measurement at a time.

[Now I have data](#) that consists off:



- about 30 dogs;
- each has 24 measurements (divided into several subgroups);
- each measurement has at least 4 contacts (one for each paw) and
  - each contact is divided into 5 parts and
  - has several parameters, like contact time, location, total force etc.



Obviously sticking everything into one big object isn't going to cut it, so I figured I needed to use classes instead of the current slew of functions. But even though I've read Learning Python's chapter about classes, I fail to apply it to my own code ([GitHub link](#))

I also feel like it's rather strange to process all the data every time I want to get out some information. Once I know the locations of each paw, there's no reason for me to calculate this again. Furthermore, I want to compare all the paws of the same dog to determine which contact belongs to which paw (front/hind, left/right). This would become a mess if I continue using only functions.

So now I'm looking for advice on how to create classes that will let me process my data ([link to the zipped data of one dog](#)) in a sensible fashion.

python

oop

class-design

Share

Improve this question

Follow

edited May 23, 2017 at 10:31



Community Bot

1 • 1

asked Nov 17, 2010 at 9:43



Ivo Flipse

10.3k • 18 • 51 • 63

- 
- 4 You might also want to consider using a database (like sqlite: [docs.python.org/library/sqlite3.html](https://docs.python.org/library/sqlite3.html)). You could write a program which reads your huge data files and converts it into rows in database tables. Then as a second stage you can write programs which draws data out of the database to do further analysis. – [unutbu](#) Nov 17, 2010 at 11:27
- 

[You mean something like I asked here](#) @unutbu? I'm planning on making it do that, but first I'd like to be able to process all the data in a more organized fashion

– [Ivo Flipse](#) Nov 17, 2010 at 11:33

---

6 Answers

Sorted by:

Highest score (default)



449



How to design a class.

1. Write down the words. You started to do this. Some people don't and wonder why they have problems.
2. Expand your set of words into simple statements about what these objects will be doing. That is to say,





write down the various calculations you'll be doing on these things. Your short list of 30 dogs, 24 measurements, 4 contacts, and several "parameters" per contact is interesting, but only part of the story. Your "locations of each paw" and "compare all the paws of the same dog to determine which contact belongs to which paw" are the next step in object design.

3. Underline the nouns. Seriously. Some folks debate the value of this, but I find that for first-time OO developers it helps. Underline the nouns.
4. Review the nouns. Generic nouns like "parameter" and "measurement" need to be replaced with specific, concrete nouns that apply to your problem in your problem domain. Specifics help clarify the problem. Generics simply elide details.
5. For each noun ("contact", "paw", "dog", etc.) write down the attributes of that noun and the actions in which that object engages. Don't short-cut this. Every attribute. "Data Set contains 30 Dogs" for example is important.
6. For each attribute, identify if this is a relationship to a defined noun, or some other kind of "primitive" or "atomic" data like a string or a float or something irreducible.
7. For each action or operation, you have to identify which noun has the responsibility, and which nouns merely participate. It's a question of "mutability". Some objects get updated, others don't. Mutable

objects must own total responsibility for their mutations.

8. At this point, you can start to transform nouns into class definitions. Some collective nouns are lists, dictionaries, tuples, sets or namedtuples, and you don't need to do very much work. Other classes are more complex, either because of complex derived data or because of some update/mutation which is performed.

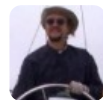
Don't forget to test each class in isolation using unittest.

Also, there's no law that says classes must be mutable. In your case, for example, you have almost no mutable data. What you have is derived data, created by transformation functions from the source dataset.

Share Improve this answer

answered Nov 17, 2010 at 11:21

Follow



**S.Lott**

391k ● 82 ● 517 ● 788



29

The following advices (similar to @S.Lott's advice) are from the book, [Beginning Python: From Novice to Professional](#)



1. Write down a description of your problem (what should the problem do?). Underline all the nouns, verbs, and adjectives.
2. Go through the nouns, looking for potential classes.

3. Go through the verbs, looking for potential methods.
4. Go through the adjectives, looking for potential attributes
5. Allocate methods and attributes to your classes

To refine the class, the book also advises we can do the following:

6. Write down (or dream up) a set of *use cases* —scenarios of how your program may be used. Try to cover all the functionally.
7. Think through every use case step by step, making sure that everything we need is covered.

Share Improve this answer

Follow

edited Nov 18, 2018 at 16:40



endolith

26.7k ● 35 ● 135 ● 202

answered Nov 22, 2013 at 18:09



mitchellc

1,657 ● 4 ● 20 ● 25

---

It would be good to have some examples of the kind of sentences we're supposed to write. – endolith Nov 18, 2018 at 17:27

---



15



I like the TDD approach... So start by writing tests for what you want the behaviour to be. And write code that passes. At this point, don't worry too much about design, just get a test suite and software that passes. Don't worry if you end up with a single big ugly class, with complex methods.

Sometimes, during this initial process, you'll find a behaviour that is hard to test and needs to be decomposed, just for testability. This may be a hint that a separate class is warranted.

Then the fun part... refactoring. After you have working software you can see the complex pieces. Often little pockets of behaviour will become apparent, suggesting a new class, but if not, just look for ways to simplify the code. Extract service objects and value objects. Simplify your methods.

If you're using git properly (you are using git, aren't you?), you can very quickly experiment with some particular decomposition during refactoring, and then abandon it and revert back if it doesn't simplify things.

By writing tested working code first you should gain an intimate insight into the problem domain that you couldn't easily get with the design-first approach. Writing tests and code push you past that "where do I begin" paralysis.

Share Improve this answer

answered Nov 22, 2013 at 18:42

Follow



- 
- 1 I too agree with this answer, although breaking down the problem and identifying possible classes (i.e. doing "just enough" software architecture) can be very useful if the problem is going to be worked on in parallel by several team members. – [Ben Smith](#) Nov 22, 2013 at 22:05
- 



4

The whole idea of OO design is to make your code map to your problem, so when, for example, you want the first footprint of a dog, you do something like:



```
dog.footstep(0)
```



Now, it may be that for your case you need to read in your raw data file and compute the footprint locations. All this could be hidden in the `footstep()` function so that it only happens once. Something like:

```
class Dog:
    def __init__(self):
        self._footsteps=None
    def footstep(self,n):
        if not self._footsteps:
            self.readInFootsteps(...)
        return self._footsteps[n]
```

[This is now a sort of caching pattern. The first time it goes and reads the footprint data, subsequent times it just gets it from `self._footsteps`.]



But yes, getting OO design right can be tricky. Think more about the things you want to do to your data, and that will inform what methods you'll need to apply to what classes.

Share Improve this answer

answered Nov 17, 2010 at 10:07

Follow



Spacedman

94.1k ● 12 ● 147 ● 229



3



After skimming your linked code, it seems to me that you are better off *not* designing a Dog class at this point.

Rather, you should use [Pandas](#) and [dataframes](#). A dataframe is a table with columns. Your dataframe would have columns such as: `dog_id`, `contact_part`, `contact_time`, `contact_location`, etc. Pandas uses Numpy arrays behind the scenes, and it has many convenience methods for you:

- Select a dog by e.g. :

```
my_measurements['dog_id']=='Charly'
```

- save the data:

```
my_measurements.save('filename.pickle')
```

- Consider using [pandas.read\\_csv\(\)](#) instead of manually reading the text files.

Share Improve this answer

edited Nov 24, 2013 at 0:24

Follow



cyborg

10.1k ● 4 ● 39 ● 56

answered Nov 24, 2013 at 0:08



2

Writing out your nouns, verbs, adjectives is a great approach, but I prefer to think of class design as asking the question *what data should be hidden?*



Imagine you had a `Query` object and a `Database` object:



The `Query` object will help you create and store a query - store, is the key here, as a function could help you create one just as easily. Maybe you could say:

```
Query().select('Country').from_table('User').where('Country == "Brazil"').
```

It doesn't matter exactly the syntax -- that is your job! -- the key is the object is helping you *hide something*, in this case the data necessary to store and output a query. The power of the object comes from the syntax of using it (in this case some clever chaining) and not needing to know what it stores to make it work. If done right the `Query` object could output queries for more than one database. It internally would store a specific format but could easily convert to other formats when outputting (Postgres, MySQL, MongoDB).

Now let's think through the `Database` object. What does this hide and store? Well clearly it can't store the full contents of the database, since that is why we have a database! So what is the point? The goal is to *hide how the database works* from people who use the `Database` object. Good classes will simplify reasoning when manipulating internal state. For this `Database` object you

could hide how the networking calls work, or batch queries or updates, or provide a caching layer.

The problem is this `Database` object is HUGE. It represents how to access a database, so under the covers it could do anything and everything. Clearly networking, caching, and batching are quite hard to deal with depending on your system, so hiding them away would be very helpful. But, as many people will note, a database is insanely complex, and the further from the raw DB calls you get, the harder it is to tune for performance and understand how things work.

This is the fundamental tradeoff of OOP. If you pick the right abstraction it makes coding simpler (`String`, `Array`, `Dictionary`), if you pick an abstraction that is too big (`Database`, `EmailManager`, `NetworkingManager`), it may become too complex to really understand how it works, or what to expect. The goal is to *hide complexity*, but some complexity is necessary. A good rule of thumb is to start out avoiding `Manager` objects, and instead create classes that are like `structs` -- all they do is hold data, with some helper methods to create/manipulate the data to make your life easier. For example, in the case of `EmailManager` start with a function called `sendEmail` that takes an `Email` object. This is a simple starting point and the code is very easy to understand.

As for your example, think about what data needs to be together to calculate what you are looking for. If you wanted to know how far an animal was walking, for

example, you could have `AnimalStep` and `AnimalTrip` (collection of `AnimalSteps`) classes. Now that each `Trip` has all the `Step` data, then it should be able to figure stuff out about it, perhaps `AnimalTrip.calculateDistance()` makes sense.

Share Improve this answer

answered Nov 23, 2013 at 1:08

Follow



[Evan Moran](#)

4,173 ● 36 ● 22



**Highly active question.** Earn 10 reputation (not counting the [association bonus](#)) in order to answer this question. The reputation requirement helps protect this question from spam and non-answer activity.