

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

Working with Data

Assignment Project Exam Help

<https://powcoder.com>

Data Science for Design Week 2

Add WeChat powcoder

Overview

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

- What is data?
- Data types
- Data formats
- Data shapes
- Operations on data

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Data?

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

- Data - plural of *datum*

<https://powcoder.com>

- Latin: *dare* - that which is given

Add WeChat powcoder

CHAPTER 3

<https://powcoder.com>
Assignment Project Exam Help
Add WeChat powcoder

Data, capta, information and knowledge

Assignment Project Exam Help
Peter Checkland and Sue Holwell
<https://powcoder.com>

Add WeChat powcoder

[The concept of information is a subtle one and one on which there is by no means complete agreement. However, this concept is needed if we are to understand how information supports purposeful action within organizations.]

There is at present no well-defined definition of such terms as 'data' and 'information' upon which there is general agreement. It is noteworthy that a current encyclopaedia of software engineering (Morris and Tamm, 1993) contains no entries for either 'data' or 'information'. Indeed one entry asserts that

Data?

Avison and Fitzgerald (1995)

<https://powcoder.com>

Data represent unstructured facts (p. 12)

Clare and Loucopoulos (1987)

[Assignment Project](#) [Exam Help](#)

Facts collected from observations or recordings about events, objects, or people
(p. 2)

Galland (1982)

[Add WeChat](#) [powcoder](#)

Facts, concepts or derivatives in a form that can be communicated and interpreted (p. 57)

Hicks (1993)

[Assignment Project](#) [Exam Help](#)

A representation of facts, concepts or instructions in a formalized manner suitable for communication, interpretation, or processing by humans or by automatic means (p. 668)

<https://powcoder.com>

Knight and Silk (1990)

Numbers representing an observable object or event (fact) (p. 22)
[Add WeChat](#) [powcoder](#)

Laudon and Laudon (1991)

Raw facts that can be shaped and formed to create information

Maddison (ed.) (1989)

Natural language: facts given, from which others may be deduced, inferred.
Info. processing and computer science: signs or symbols, especially as for transmission in communication systems and for processing in computer systems; usually but not always representing information (sic), agreed facts or assumed knowledge; and represented using agreed characters, codes, syntax and structure

Martin and Powell (1992)

The raw material of organizational life; it consists of disconnected numbers, words, symbols and syllables relating to the events and processes of the business (p. 10)

Information?

Avison and Fitzgerald (1995)	Information has a meaning ... [it] comes from selecting data, summarizing it and presenting it in such a way that it is useful to the recipient (p. 12) https://powcoder.com
Clare and Loucopoulos (1987)	A pre-requisite for a decision to be taken. Information is the product of the meaningful processing of data (p. 2) https://powcoder.com
Galland (1982)	Information is that which results when some human mental activity (observation, analysis) is successfully applied to data to reveal its meaning or significance (p. 127) https://powcoder.com
Hicks (1993)	Data that has been processed so that it is meaningful to a decision maker to use in a particular decision (p. 675) https://powcoder.com
Knight and Silk (1990)	Human significance associated with an observable object or event (p. 22) https://powcoder.com
Laudon and Laudon (1991)	Data that have been shaped or formed by humans into a meaningful and useful form (p. 14)
Maddison (ed.) (1989)	Understandable useful relevant communication at an appropriate time; any kind of knowledge about things and concepts in a universe of discourse that is exchangeable between users; it is the meaning that matters, not the representation (p. 174)
Martin and Powell (1992)	Information comes from data that has been processed to make it useful in management decision making (p. 10)

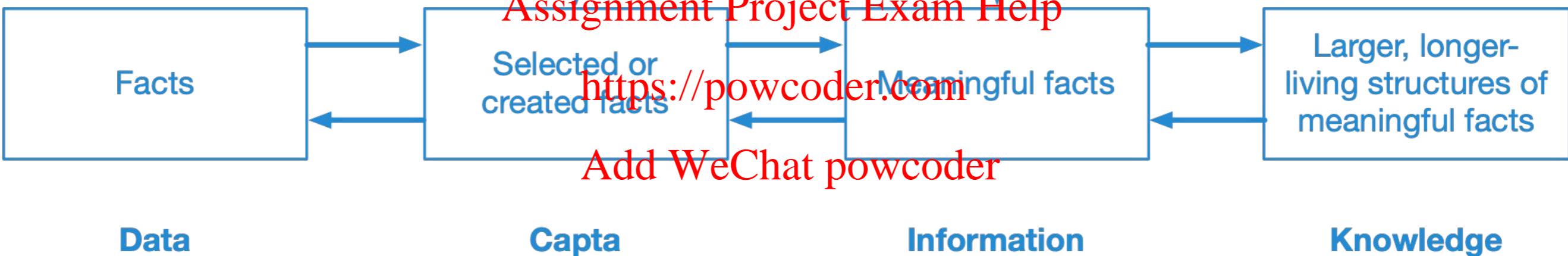
<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

Cognitive settings

Context, interests



<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

Data Details

<https://powcoder.com>
Add WeChat powcoder

Quick Glossary

Fields: individual pieces of data
<https://powcoder.com>

[Assignment](#) [Project](#) [Exam](#) [Help](#)

Types: what kind of thing is an individual piece of data?
[Add WeChat](#) [powcoder](#)

Records: set of data about one thing

[Assignment](#) [Project](#) [Exam](#) [Help](#)

Attributes: fields belonging to a record
<https://powcoder.com>

Schema: what are the types and values the data can take
[Add WeChat](#) [powcoder](#)

Formats: how data is structured so that fields are intelligible

Forms: how the data is contained or transmitted

Metadata: Data that *describes* the data

Binary: Bits and Bytes

<https://powcoder.com>

- 01010010101011000101010
Assignment Project Exam Help
Add WeChat powcoder
- Each binary value is a "bit" - a fundamental unit of information (Shannon)
Assignment Project Exam Help
<https://powcoder.com>
- Often grouped together into "bytes" - 8 bits
Add WeChat powcoder
- One byte can have 256 values, often used to represent numbers from 0 to 255, or 00 to FF in hex

Counting in binary

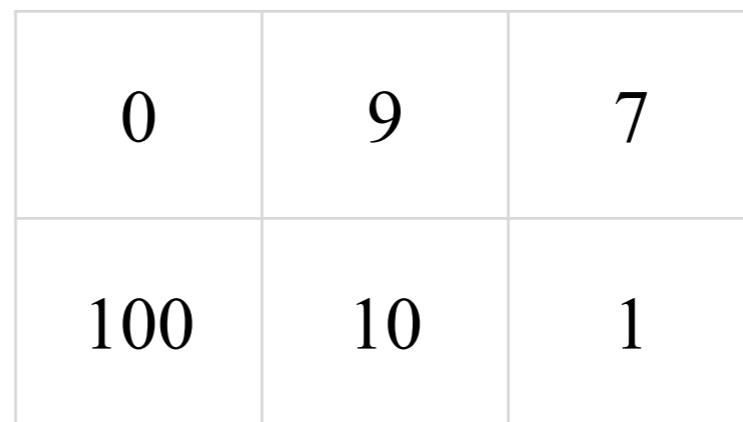
<https://powcoder.com>

Assignment Project Exam Help

Binary Base 10

Add WeChat powcoder

Binary	Base 10
0000	0
0001	1
0010	2
0011	3
0100	4
0101	5
0110	6
0111	7
1000	8



(base 10
version)

$$(0 * 100) + (9 * 10) + (7 * 1) = 97$$

Binary interpretations

<https://powcoder.com>

Assignment Project Exam Help
01100001

Add WeChat powcoder

The same "bit pattern" is:

[Assignment Project Exam Help](https://powcoder.com)

The character 'a' in ASCII

97 in decimal

Add WeChat powcoder

61 in hexadecimal

sometimes written 0x61

<http://www.rapidtables.com/convert/number/ascii-hex-bin-dec-converter.htm>

Number delimiter: Space

ASCII text: a

Hex: 61

Binary: 01100001

Decimal: 97

Base64: YQ==

Length: 1 chars/bytes

Checksum: 61

8-bit Sum

Text (ASCII / ANSI)

I gave a cry of astonishment. I saw and thought nothing of the other four Martian monsters; my attention was riveted upon the nearer incident.

Simultaneously two other shells burst in the air near the body as the hood twisted suddenly to receive, but not in time to dodge, the fourth shell.

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

Convert

Highlight Text

Binary

```
01001001 00100000 01100111 01100001 01110110  
01100101 00100000 01100001 00100000 01100011  
01110010 01111001 00100000 01101111 01100110  
00100000 01100001 01110011 01110100 01101111  
01101110 01101001 01110011 01101000 01101101  
01100101 01101110 01110100 00101110 00100000  
01001001 00100000 01110011 01100001 01110111  
00100000 01100001 01101110 01100100 00100000
```

Convert

Highlight Text

Hexadecimal

```
49 20 67 61 76 65 20 61 20 63 72 79 20 6f 66 20 61  
73 74 6f 6e 69 73 68 6d 65 6e 74 2e 20 49 20 73 61  
77 20 61 6e 64 20 74 68 6f 75 67 68 74 20 6e 6f 74  
68 69 6e 67 20 6f 66 20 74 68 65 20 6f 74 68 65 72  
20 66 6f 75 72 20 4d 61 72 74 69 61 6e 20 6d 6f 6e  
73 74 65 72 73 3b 20 6d 79 20 61 74 74 65 6e 74 69  
6f 6e 20 77 61 73 20 72 69 76 65 74 65 64 20 75 70  
6f 6e 20 74 68 65 20 6e 65 61 72 65 72 20 69 6e 63
```

Convert

Highlight Text

BASE64

<https://powcoder.com>

Add WeChat powcoder

```
SSBnYXZlIC EgY3J5IG9mIGFzdG9uaXNobWVudC4gSS  
BzYXcgYW5kIHRob3VnaHQgbm90aGluZyBvZiB0aGUg  
b3RoZXIgZm91cBNxj0aWFUIG1VdnN0ZXIzOjEtcB  
hdHRIbnRpb24gd2FzIHJpdmV0ZWQgdXBvbiB0aGUg  
bmVhcmVYIGluY2lkZW50LiBTaW1bHRhbhVvdXNse  
SB0d28gb3RoZXIgc2hlbGxzIGJ1cnN0IGluIHRoZSBha  
XIgbmVhciB0aGUgYm9keSBhcYB0aGUgaG9vZCB0d2I  
zdGVkIHJvdW5kIGluIHRpbWUgdG8gcmVjZWI2ZSwgY
```

Convert

Highlight Text

Decimal

```
73 32 103 97 118 101 32 97 32 99 114 121 32 111  
102 32 97 115 116 111 110 105 115 104 109 101 110  
116 46 32 73 32 115 97 119 32 97 110 100 32 116  
104 111 117 103 104 116 32 110 111 116 104 105  
110 103 32 111 102 32 116 104 101 32 111 116 104  
101 114 32 102 111 117 114 32 77 97 114 116 105  
97 110 32 109 111 110 115 116 101 114 115 59 32  
109 121 32 97 116 116 101 110 116 105 111 110 32
```

Convert

Highlight Text

ROT13

V tnir n pel bs nfgbavfuzrag. V fnj naq gubhtug abguvat bs gur bgure sbhe Znegvna zbafgref; zl nggragvba jnf evirgrq hcba gur arnere vapvrag.

URL Encoded

I+gave+a+cry+of+astonishment.+I+saw+and+thought+nothing+of+the+other+four+Martian+monsters%3B+my+attention+was+riveted+upon+the+nearer+incident

Data Types

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

- Data stored within an individual field

<https://powcoder.com>

- What do they look like, what operations might we do?

Quick Glossary

<https://powcoder.com>

Fields: individual pieces of data

Assignment Project Exam Help

- **Types: what kind of thing is an individual piece of data?**
[Add WeChat powcoder](#)

Records: set of data about one thing

Assignment Project Exam Help

Attributes: fields belonging to a record

<https://powcoder.com>

Schema: what are the types and values the data can take
[Add WeChat powcoder](#)

Formats: how data is structured so that fields are intelligible

Forms: how the data is contained or transmitted

Metadata: Data that describes the data

Boolean

<https://powcoder.com>
Values that are "true" or "false", 1
or 0
[Assignment](#) [Project](#) [Exam](#) [Help](#)

Used heavily in logic and
conditionals
[Add WeChat](#) [powcoder](#)

Examples: has something been
processed or checked? does a
person have an attribute?
<https://powcoder.com>
[Add WeChat](#) [powcoder](#)

Operations: **AND**, **OR**, e.g.

people who have been
vaccinated **AND** are showing
symptoms

people who have a UK Visa **OR**
are EU Citizens

```
>>> True and False
False
>>> True or False
True
>>> False or False
False
>>> type(False)
<class 'bool'>
>>> |
```

Booleans

<https://powcoder.com>

True or False
Assignment Project Exam Help

Add WeChat powcoder
Combine with

"or", "l"

"and", "&"

"Negated" with "not"
(sometimes written "!" or "^")

"True" and "False" in Python

Use "and" / "or" in Python

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder

```
>>> type( True )
<type 'bool'>
>>> not True
False
>>> True and True
True
>>> True and False
False
>>> True or False
True
>>> True and not False
True
>>> █
```

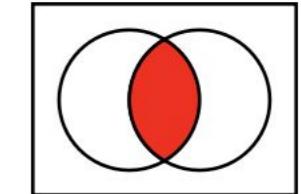
Truth Tables / Truth Functions

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

<https://powcoder.com>

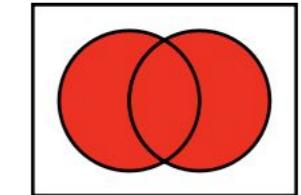
Assignment Project Exam Help

Add WeChat powcoder

Conjunction															
Notation	Equivalent formulas	Truth table	Venn diagram												
$P \wedge Q$ $P \& Q$ $P \cdot Q$ $P \text{ AND } Q$	$P \rightarrow \neg Q$ $\neg P \leftarrow Q$ $\neg P \downarrow \neg Q$ Kpq	<table border="1"> <tr> <td></td> <td></td> <td style="text-align: center;">Q</td> </tr> <tr> <td></td> <td>0</td> <td>1</td> </tr> <tr> <td style="text-align: center;">P</td> <td>0</td> <td>0</td> </tr> <tr> <td>1</td> <td>0</td> <td>1</td> </tr> </table>			Q		0	1	P	0	0	1	0	1	
		Q													
	0	1													
P	0	0													
1	0	1													

<https://powcoder.com>

Add WeChat powcoder Disjunction

Notation	Equivalent formulas	Truth table	Venn diagram												
$P \vee Q$ $P \text{ OR } Q$	$P \leftarrow \neg Q$ $\neg P \rightarrow Q$ $\neg P \uparrow \neg Q$ $\neg(\neg P \wedge \neg Q)$ Apq	<table border="1"> <tr> <td></td> <td></td> <td style="text-align: center;">Q</td> </tr> <tr> <td></td> <td>0</td> <td>1</td> </tr> <tr> <td style="text-align: center;">P</td> <td>0</td> <td>0</td> </tr> <tr> <td>1</td> <td>1</td> <td>1</td> </tr> </table>			Q		0	1	P	0	0	1	1	1	
		Q													
	0	1													
P	0	0													
1	1	1													

Example - student registers

id	registered	seen	registered & seen	! registered	reg'd & !seenPT
	Assignment	Project	Exam	P	Help
1	1	1	1	0	0
2	1	0	0	0	1
3	0	0	0	1	0
4	0	1	0	1	0

Numbers

<https://powcoder.com>

Integers: counting numbers / whole numbers

[Assignment](#) [Project](#) [Exam Help](#)

"int" in Python

[Add WeChat powcoder](#)

Can: add, subtract, multiply, divide

But division may be funny!

- $7/2 = 3$ (Python 2) <https://powcoder.com>
[Add WeChat powcoder](#)
- $7/2 = 3.5$ (Python 3)

Remainder: $7 \% 2 = 1$

Floating Point: rational numbers

"float" in Python

Add, subtract, divide etc.

Division and remainder work

```
>>> type(7)
<type 'int'>
>>> 7 / 2
3
>>> 7 % 2
1
```

```
>>> type(3.2)
<type 'float'>
>>> 7.0 / 2
3.5
>>> 7.0 % 2
1.0
```

Enumerations

<https://powcoder.com>

Also called categorical variables, controlled vocabulary
Assignment Project Exam Help
Add WeChat powcoder

When you have a precise set of values that something can take

Possible values come from
the data schema

Examples: country, marital status, eye color, sex, lots of status variables, e.g. waiting, processing, complete

Add WeChat powcoder

```
>>> from enum import Enum  
am>>> class Fruit(Enum):  
...:  
om:  
...:  
coder:  
...  
>>> print(Fruit.APPLE)  
Fruit.APPLE  
>>> Fruit(2)  
<Fruit.BANANA: 2>  
>>> █
```

Strings

<https://powcoder.com>

Assignment Project Exam Help

Text – a sequence of characters

Add WeChat powcoder

Used for many things – names,
descriptions, user feedback,
tweets

Assignment Project Exam Help

Operations: joining
(concatenating), splitting
parsing (extracting structure),
modifying, regular expressions

<https://powcoder.com>

Add WeChat powcoder

Transformed (e.g. lowercase)

Different representations (ASCII,
UTF8)

Can have structure (e.g. dates)

```
>>> type("test string")
<type 'str'>
>>> "test" + "string"
'teststring'
>>> "test, string".split(",")
['test', ' string']
>>> "teststring".upper()
'TESTSTRING'
>>> 
```

Compound Types

<https://powcoder.com>

Dates/times
Assignment Project Exam Help

Add WeChat powcoder
many different
formats, cause Assignment Project Exam Help
horrible problems!

```
>>> import datetime
>>> datetime.datetime.now()
datetime.datetime(2017, 9, 26, 18, 28, 52, 437370)
>>> str(datetime.datetime.now())
'2017-09-26 18:29:02.357087'
```

Location (Lat/Long)
Assignment Project Exam Help

<https://powcoder.com>

```
>>> pos = (53.2, 14.5)
>>> pos
(53.2, 14.5)
>>> lat, lon = pos
>>> lat
53.2
>>> lon
14.5
>>> █
```

Tuples: ordered
collections of fields

Colours: **#FF0000**,
#00FF00, **#FFFF00**
#55C0D1

```
Color("red")          ## human, web compatible representation
Color(red=1)          ## default amount of blue and green is 0.0
Color("blue", hue=0)   ## hue of blue is 0.66, hue of red is 0.0
Color("#f00")          ## standard 3 hex digit web compatible representation
Color("#ff0000")       ## standard 6 hex digit web compatible representation
Color(hue=0, saturation=1, luminance=0.5)
Color(hsl=(0, 1, 0.5)) ## full 3-uple HSL specification
Color(rgb=(1, 0, 0))   ## full 3-uple RGB specification
```

Falsehoods Programmers Believe About Names

People have exactly one canonical full name.
<https://powcoder.com>

People have exactly one full name which they go by.

People have, at this point in time, exactly one canonical full name.

[Assignment](#) [Project](#) [Exam](#) [Help](#)

People have, at this point in time, one full name which they go by.

People have exactly N names, for any value of N.

People's names fit within a certain defined amount of space.

People's names do not change.

People's names change, but only at a certain enumerated set of events.

People's names are written in ASCII.

People's names are written in any single character set.

People's names are all mapped in Unicode code points.

People's names are case sensitive.

People's names are case insensitive.

People's names sometimes have prefixes or suffixes, but you can safely ignore those.

People's names do not contain numbers.

People's names are not written in ALL CAPS.

People's names are not written in all lower case letters.

People's names have an order to them. Picking any ordering scheme will automatically result in consistent ordering among all systems, as long as both use the same ordering scheme for the same name.

People's first names and last names are, by necessity, different.

People have last names, family names, or anything else which is shared by folks recognized as their relatives.

[Assignment](#) [Project](#) [Exam](#) [Help](#)

<https://powcoder.com>

[Add](#) [WeChat](#) [powcoder](#)

Names...

Scunthorpe problem

<https://powcoder.com>

Assignment Project Exam Help

special page
Add WeChat powcoder
Log in / create account

From Wikipedia, the free encyclopedia

Assignment Project Exam Help

Login error

The user name "ScunthorpeM181" has been blacklisted from creation.

<https://powcoder.com>

Add WeChat powcoder

Create account

Already have an account? [Log in](#).

Registering a free account takes only a few seconds and has many benefits.

Simply choose a [username](#) (not the same as your e-mail address) and a unique password and click [Create account](#).

- For your own security, please choose a password that would be **difficult to guess**. ([Help](#))
- All usernames *will* begin with a capitalized letter, regardless of what you type.
- All underscores *will* be converted to spaces, regardless of what you type.
- For technical reasons, the #, / and \ symbols cannot be used in a username.
- You do not have to provide an e-mail address, but if you forget your password, you will need it to reset it.

(Your e-mail address is never given to anyone, with one exception: if you e-mail another user, your the e-mail address is given to them.)

URI / URLs

Unique Resource Identifier – URI

<https://powcoder.com>

Slightly magic – there is only one "thing" in the universe that a URI "points" to

[Assignment](#) [Project](#) [Exam](#) [Help](#)

Different types:

[Add WeChat](#) powcoder

https://en.wikipedia.org/wiki/Uniform_Resource_Identifier

[Assignment](#) [Project](#) [Exam](#) [Help](#)

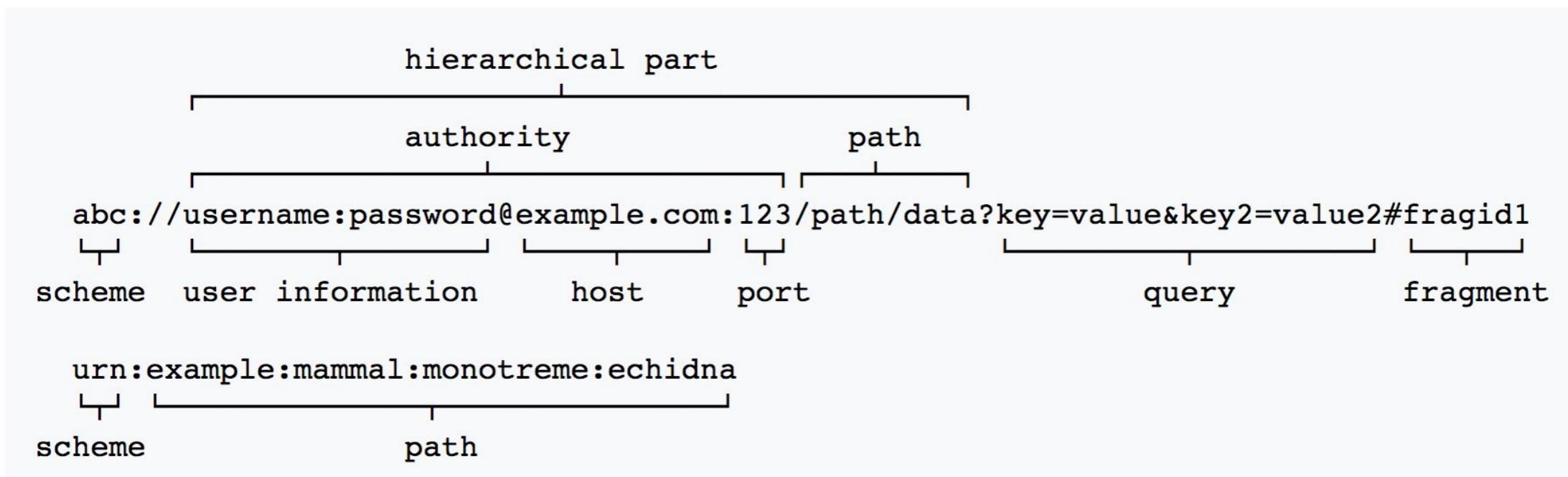
ISBN 0-486-27557-4 (Romeo and Juliet)

<https://powcoder.com>

A URL is simply a URI that happens to point to a resource over a network

[Add WeChat](#) powcoder

Structured



Ontologies

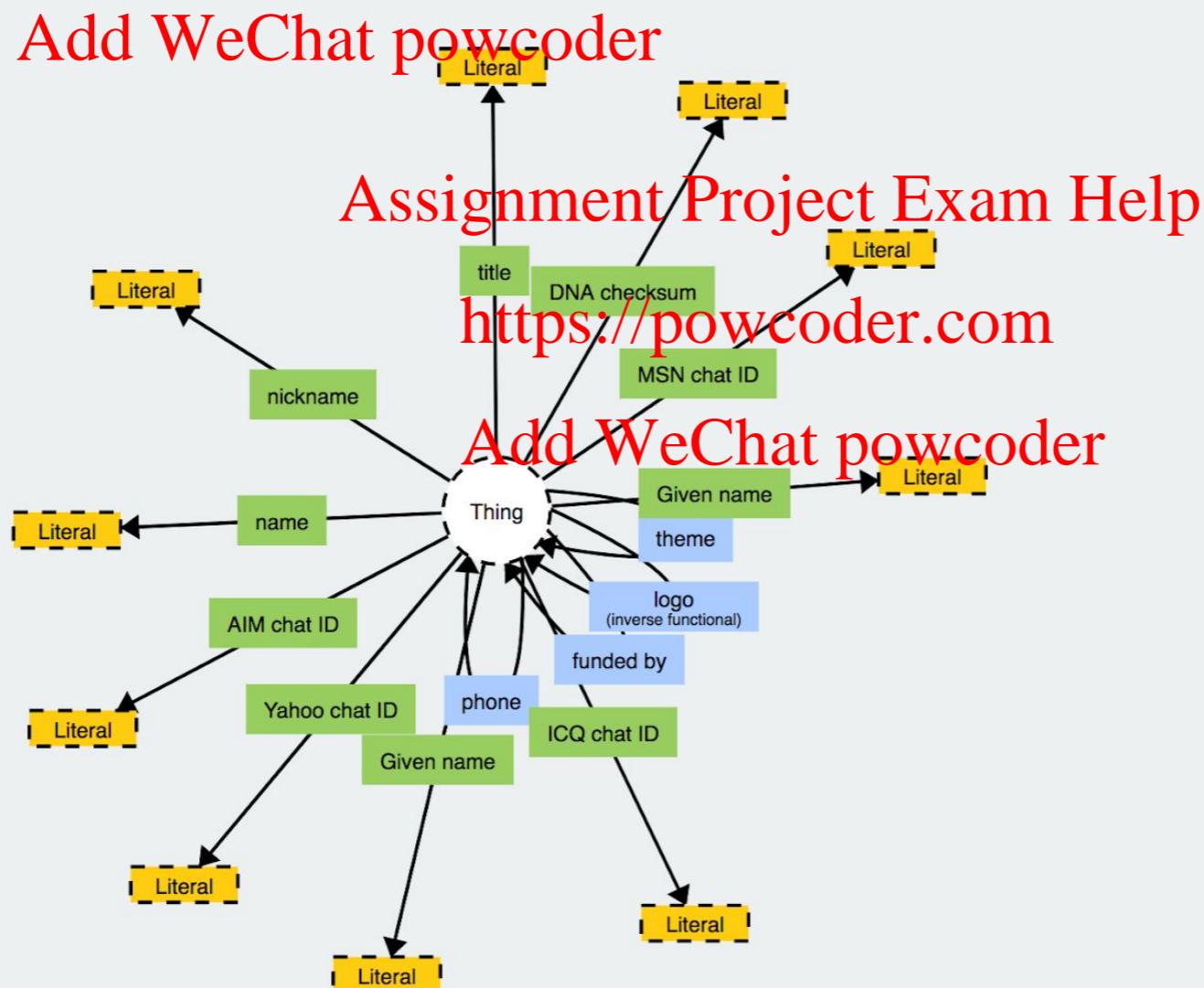
Structured dictionaries of terms, and relations between them

<https://powcoder.com>

Define the kinds of data, their meanings and their structure

Assignment Project Exam Help

WebVOWL
1.0.4



Friend of a Friend (FOAF) vocabulary

<http://xmlns.com/foaf/0.1/>

Version: --
Author(s): --
Language:

▼ Description

The Friend of a Friend (FOAF) RDF vocabulary, described using W3C RDF Schema and the Web Ontology Language.

► Metadata

► Statistics

► Selection Details

<http://visualdataweb.de/webowl/>

Conversions

- Text to numbers
<https://powcoder.com>
Assignment Project Exam Help
- Numbers to text
[Add WeChat powcoder](#)
Assignment Project Exam Help
- Strings to dates (and back again)
<https://powcoder.com>
- Text to codes/enumerations
[Add WeChat powcoder](#)
- URLs to documents
- Terms in one ontology to another

Fields and Formats

<https://powcoder.com>

Assignment Project Exam Help

- Fields: individual pieces of data
[Add WeChat powcoder](#)
- Formats: how data is structured so that fields are intelligible
<https://powcoder.com>
[Assignment Project Exam Help](#)
[Add WeChat powcoder](#)
- Forms: how the data is contained or transmitted
- Metadata: Data that *describes* the data

Quick Glossary

- **Fields:** individual pieces of data
<https://powcoder.com>
Assignment Project Exam Help

Types: what kind of thing is an individual piece of data?
[Add WeChat powcoder](#)

- **Records:** set of a data about one thing
<https://powcoder.com>
Assignment Project Exam Help

- **Attributes:** fields belonging to a record
<https://powcoder.com>

Schema: what are the types and values the data can take
[Add WeChat powcoder](#)

- **Formats:** how data is structured so that fields are intelligible
- **Forms:** how the data is contained or transmitted

Metadata: Data that describes the data

Form (file) **Format** **Fields** **Schema**

productsales.csv

TOP HIT https://powcoder.com

DOCUMENTS Assignment Project Exam Help

SOURCES.txt Add WeChat powcoder

OTHER

files

DEVELOPER

test_sas7bdat.py

Show all in Finder...

Record →

ACTUAL	PREDICTION	COUNTRY	REGION	DIVISION	PRODTYPE	PRODNAME
923	850	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
999	297	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
608	846	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
643	533	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
656	646	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
948	486	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
612	717	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
114	564	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
685	230	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
657	494	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
608	903	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
353	266	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
107	190	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
354	139	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE
101	217	CANADA	EAST	EDUCATION	FURNITURE	SOFWARE

productsales.csv (~/anaconda/pkgs/pandas-0.20.3-py27_0/lib/python2

1 ACTUAL, PREDICT, COUNTRY, REGION, DIVISION, PRODTYPE, PRODUCT, QUARTER, YEAR, MONTH
2 925, 850, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1993, 12054
3 999, 297, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1993, 12085
4 608, 846, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1993, 12113
5 642, 533, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1993, 12144
6 656, 646, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1993, 12174
7 948, 486, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1993, 12205
8 612, 717, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1993, 12235
9 114, 564, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1993, 12266
10 685, 230, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1993, 12297
11 657, 494, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1993, 12327
12 608, 903, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1993, 12358
13 353, 266, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1993, 12388
14 107, 190, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1994, 12419
15 354, 139, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1994, 12450
16 101, 217, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1994, 12478
17 553, 560, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1994, 12509
18 877, 148, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1994, 12539
19 431, 762, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1994, 12570
20 511, 457, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1994, 12600
21 157, 532, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1994, 12631
22 520, 629, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1994, 12662
23 114, 491, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1994, 12692
24 277, 0, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1994, 12723

productsales.csv (~/anaconda/pkgs/pandas-0.20.3-py27_0/lib/python2

1 ACTUAL, PREDICT, COUNTRY, REGION, DIVISION, PRODTYPE, PRODUCT, QUARTER, YEAR, MONTH
2 925, 850, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1993, 12054
3 999, 297, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1993, 12085
4 608, 846, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1993, 12113
5 642, 533, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1993, 12144
6 656, 646, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1993, 12174
7 948, 486, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1993, 12205
8 612, 717, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1993, 12235
9 114, 564, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1993, 12266
10 685, 230, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1993, 12297
11 657, 494, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1993, 12327
12 608, 903, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1993, 12358
13 353, 266, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1993, 12388
14 107, 190, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1994, 12419
15 354, 139, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1994, 12450
16 101, 217, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 1, 1994, 12478
17 553, 560, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1994, 12509
18 877, 148, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1994, 12539
19 431, 762, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 2, 1994, 12570
20 511, 457, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1994, 12600
21 157, 532, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1994, 12631
22 520, 629, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 3, 1994, 12662
23 114, 491, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1994, 12692
24 277, 0, CANADA, EAST, EDUCATION, FURNITURE, SOFA, 4, 1994, 12723

Assignment Project Exam Help
Add WeChat powcoder
<https://powcoder.com>

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

info.json

```
1 [{"id":104912,"name":"Dave's Test Channel","description":"General stuff for my testing","latitude":"0.0","longitude":"0.0"}]
```

Add WeChat powcoder

info.json

Settings

Core

Editor

Keybindings

Packages

Themes

Updates

Install

Open Config Folder

+ Assignment Project Exam Help

Packages are published to atom.io and are installed to /Users/dmrust/.atom/packages
Add WeChat powcoder

json

Packages

Themes

Assignment Project Exam Help

pretty-json 1.6.4

373,155

Format JSON

https://powcoder.com



federomero

Settings

Uninstall

Disable

Add WeChat powcoder

language-json 0.19.1

303,699

JSON language support in Atom



atom

Settings

Disable

autocomplete-json 5.5.0

60,046

Atom autocomplete+ plugin for semantic JSON completions

```
1 [ List
2 { Entry
3     "id": 104912, Field
4     "name": "Dave's Test Channel",
5     "description": "General Stuff for my testing",
6     Field name "latitude": "0.0", Value
7         "longitude": "0.0",
8         "created_at": "2016-04-01T13:14:23Z",
9             "elevation": "",
10            "last_entry_id": 412,
11            "public_flag": true,
12            "url": "",
13            "ranking": 50,
14            "metadata": "",
15            "license_id": 0,
16            "tags": Add WeChat powcoder
17            "api_keys": [
18                Nested
19                Object
20                {
21                    "api_key": "L2QAYXVDZANGUYRT",
22                    "write_flag": true
23                },
24                {
25                    "api_key": "XBCMSYSRV9SIMM04",
26                    "write_flag": false
27                }
28            ]
29 }
```



<https://powcoder.com>

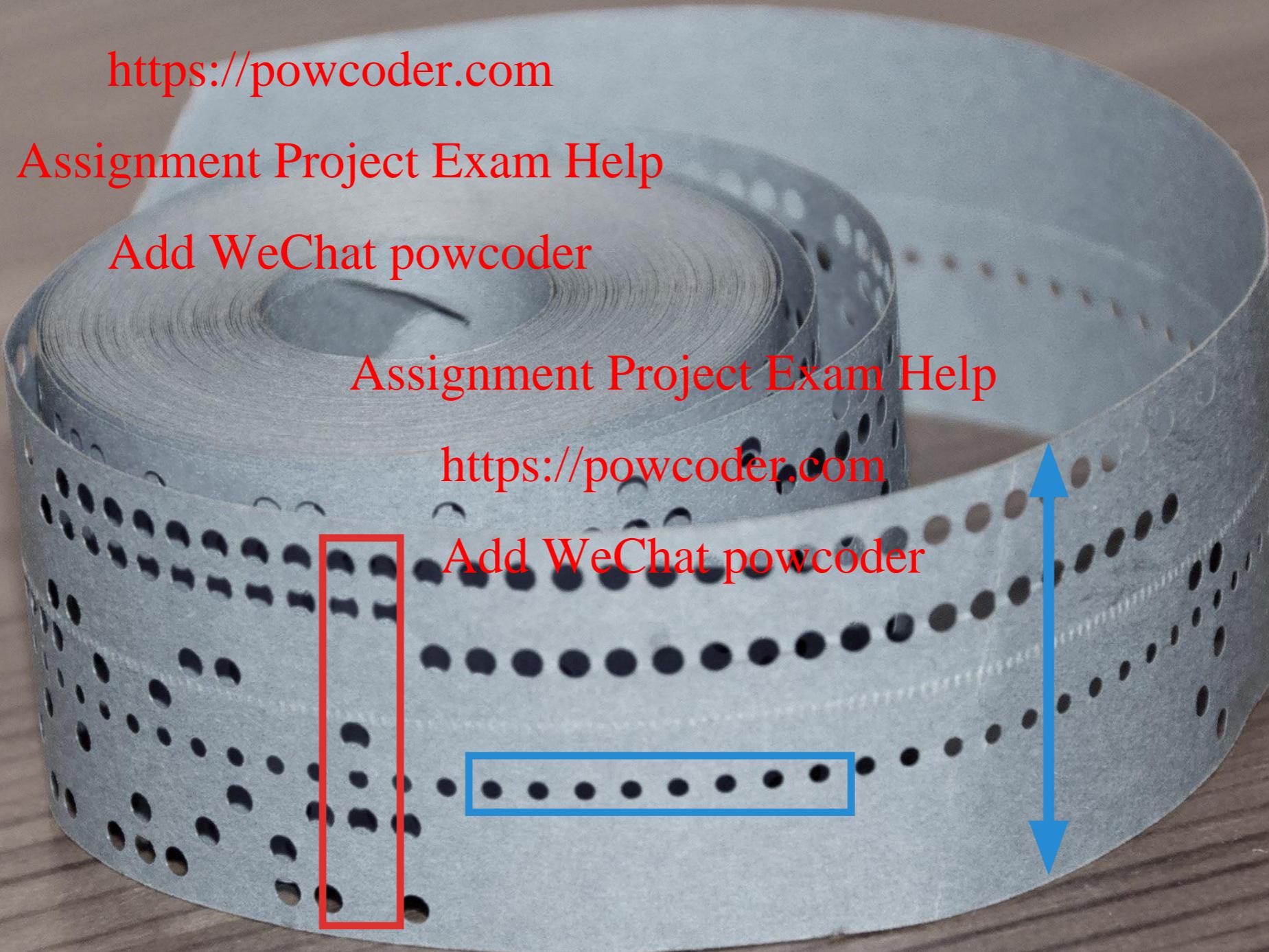
Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Data Formats

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

- How the data is organised and made accessible
[Assignment Project Exam Help](https://powcoder.com)
- Long history of different kinds of format
[Add WeChat powcoder](https://powcoder.com)
- Relates to data representation, and "shape"

CSV, TSV

<https://powcoder.com>
Table of fields, separated
by commas
[Assignment](#) ([CSV](#)) [Project](#) [Jobs](#) [Exam](#) [Help](#)
(TSV)

Add WeChat powcoder

Easy to use, easy to parse

[Assignment](#) [Project](#) [Exam](#) [Help](#)

Read and written by many
programs

<https://powcoder.com>

Add WeChat powcoder

Only handles simple tables

Complex values are often
quoted

Doesn't say *what* the data
is

Represents *tabular* data

	ACTUAL, PREDICT, COUNTRY, REGION, DIVISION	productsales
1	925,850, CANADA, EAST, EDUCATION, FURNITURE,	
2	999,297, CANADA, EAST, EDUCATION, FURNITURE,	
3	608,846, CANADA, EAST, EDUCATION, FURNITURE,	
4	642,533, CANADA, EAST, EDUCATION, FURNITURE,	
5	656,646, CANADA, EAST, EDUCATION, FURNITURE,	
6	948,486, CANADA, EAST, EDUCATION, FURNITURE,	
7	612,717, CANADA, EAST, EDUCATION, FURNITURE,	
8	114,584, CANADA, EAST, EDUCATION, FURNITURE,	
9	685,230, CANADA, EAST, EDUCATION, FURNITURE,	
10	657,191, CANADA, EAST, EDUCATION, FURNITURE,	
11	608,903, CANADA, EAST, EDUCATION, FURNITURE,	
12	353,266, CANADA, EAST, EDUCATION, FURNITURE,	
13	107,190, CANADA, EAST, EDUCATION, FURNITURE,	
14	354,139, CANADA, EAST, EDUCATION, FURNITURE,	
15	101,217, CANADA, EAST, EDUCATION, FURNITURE,	
16	553,560, CANADA, EAST, EDUCATION, FURNITURE,	
17	877,148, CANADA, EAST, EDUCATION, FURNITURE,	
18	431,762, CANADA, EAST, EDUCATION, FURNITURE,	
19	511,457, CANADA, EAST, EDUCATION, FURNITURE,	
20	157,532, CANADA, EAST, EDUCATION, FURNITURE,	
21	520,629, CANADA, EAST, EDUCATION, FURNITURE,	
22	114,491, CANADA, EAST, EDUCATION, FURNITURE,	
23	277,0, CANADA, EAST, EDUCATION, FURNITURE,	
24		

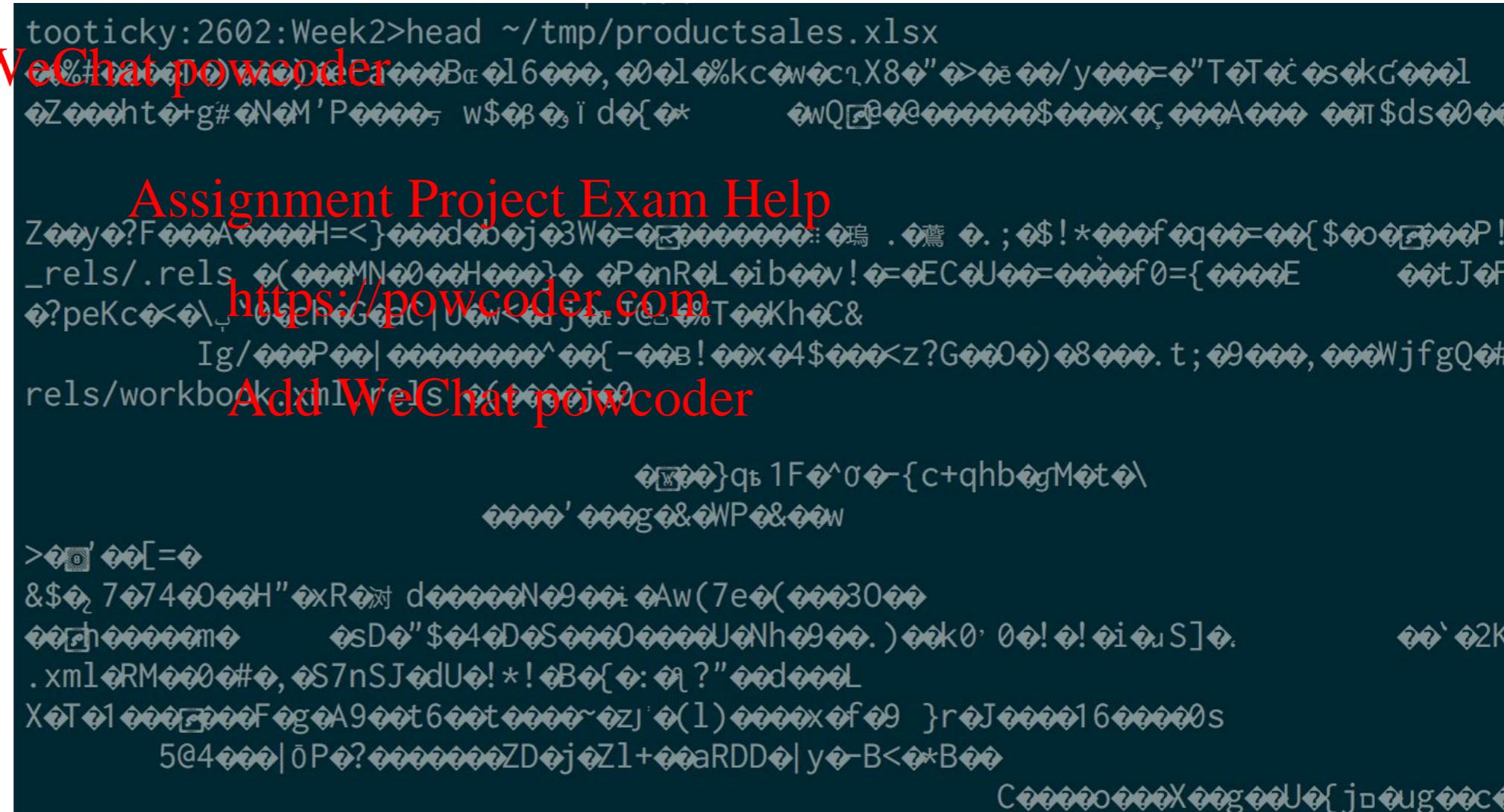
Excel

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

- Common
- Horrible
- (can
export
CSV...)

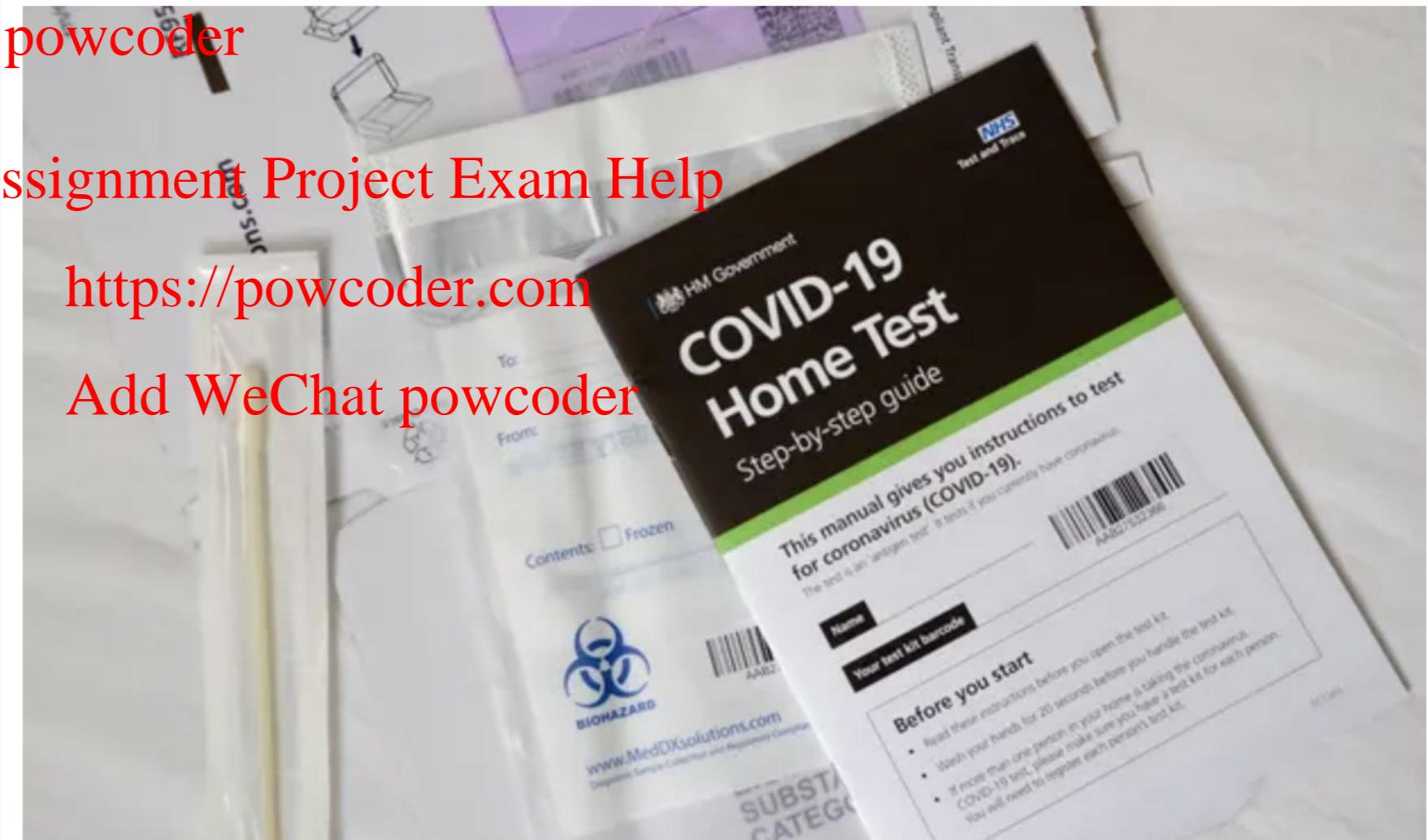


Covid: how Excel may have caused loss of 16,000 test results in England

<https://powcoder.com>

Assignment Project Exam Help
See all our coronavirus coverage

Add WeChat powcoder



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

▲ More than 50,000 potentially infectious people may have been missed by contact tracers after 15,841 positive tests were left off the daily figures. Photograph: Simon Leigh/Alamy

A million-row limit on Microsoft's Excel spreadsheet software may have led to Public Health England misplacing nearly 16,000 Covid test results, it is understood.

<https://www.theguardian.com/politics/2020/oct/05/ministers-accused-of-putting-lives-at-risk-with-covid-data-error>

XML

	<p>https://powcoder.com</p>	
Looks a bit like HTML	<p>Assignment Project Exam Help</p> <p>Add WeChat powcoder</p>	<pre><?xml version="1.0" encoding="UTF-8"?> <breakfast_menu></pre>
Stricter		<pre><food></pre>
Includes metadata		<pre><name>Belgian Waffles</name> <price>\$5.95</price> <description></pre>
Matched tags		<pre>Two of our famous Belgian Waffles with plenty of real maple syrup </description> <calories>650</calories></pre>
Tags can have attributes		<pre></food> <food></pre>
Represents <i>heirarchical</i> data		<pre><name>Strawberry Belgian Waffles</name> <price>\$7.95</price> <description></pre>
		<pre>Light Belgian waffles covered with strawberries and whipped cream </description> <calories>900</calories></pre>
		<pre></food> <food></pre>
		<pre><name>Berry-Berry Belgian Waffles</name> <price>\$8.95</price> <description></pre>
		<pre>Belgian waffles covered with assorted fresh berries and whipped cream</pre>

JSON

<https://powcoder.com>

Assignment Project Exam Help
Grew out of Javascript
Add WeChat powcoder

Widespread

Hierarchical

Similar structure to XML
Add WeChat powcoder

(but more 'readable', although no attributes)

Returned by a lot of web services

Close to computing data structures – arrays, dictionaries

```
{  
    "firstName": "John",  
    "lastName": "Smith",  
    "isAlive": true,  
    "age": 25,  
    "address": {  
        "streetAddress": "21 2nd Street",  
        "city": "New York",  
        "state": "NY",  
        "postalCode": "10021-3100"  
    },  
    "phoneNumbers": [  
        {  
            "type": "home",  
            "number": "212 555-1234"  
        },  
        {  
            "type": "office",  
            "number": "646 555-4567"  
        },  
        {  
            "type": "mobile",  
            "number": "123 456-7890"  
        }  
    "children": [],  
    "spouse": null  
}
```

people.json

```
1 [          https://powcoder.com
2 {           Assignment Project Exam Help
3   "name": "Kal",      Add WeChat powcoder
4   "age": 27,
5   "phone": "0777 777777"
6 },
7 {
8   "name": "Jan",      Assignment Project Exam Help
9   "age": 34,
10  "phone": "0131 555 5555"
11 }
12 ]
```

people.csv

```
1 Name, Age, Phone Number
2 Kal, 27, 07777 777777
3 Jan, 34, 0131 555 5555
```

people.xml

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <people>
3   <person>
4     <name>Kal</name>
5     <age>27</age>
6     <phone>0777 777777</phone>
7   </person>
8   <person>
9     <name>Jan</name>
10    <age>34</age>
11    <phone>0131 555 5555</phone>
12  </person>
13 </people>
```

Other formats

<https://powcoder.com>

Many, many, many more
Assignment, Project, Exam Help

Good questions:
[Add WeChat powcoder](#)

can I read it? [Assignment Project Exam Help](#)

what kind of data can it hold? Datatypes? Tabular?
Hierarchical? [Add WeChat powcoder](#)

can other people read it?

how many different programs can read it?

do all programs read it the same?

how well does it map to my data?

how efficient is it? (Time, Space)

Data Shapes

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

- We've heard about ~~Assignment Project Exam Help~~ **tabular** and **hierarchical** data - what does that mean? <https://powcoder.com>
Add WeChat powcoder
- Are there other types?

Tabular Data

<https://powcoder.com>

Assignment Project Exam Help
Data is in a
table, with rows
and columns

All the entries in
a column have
the same type

All the entries in
a row relate to
the same thing

Name	Age	Phone Number
Jan	35	3545
Jim	25	22232
Kev	53	22353

Relational Data

Data doesn't fit in a single table
[Assignment Project Exam Help](https://powcoder.com)

Adds relations between tables – one to many, many to one etc.

[Add WeChat powcoder](https://powcoder.com)

- **"Primary Key"** – id of thing in this table – unique within table

"Foreign Key" – id of a thing in another table

id	Name	Age	Phone Number
1	Jan	35	3545
2	Jim	25	22232
3	Kev	53	22353

Primary model of Relational Databases

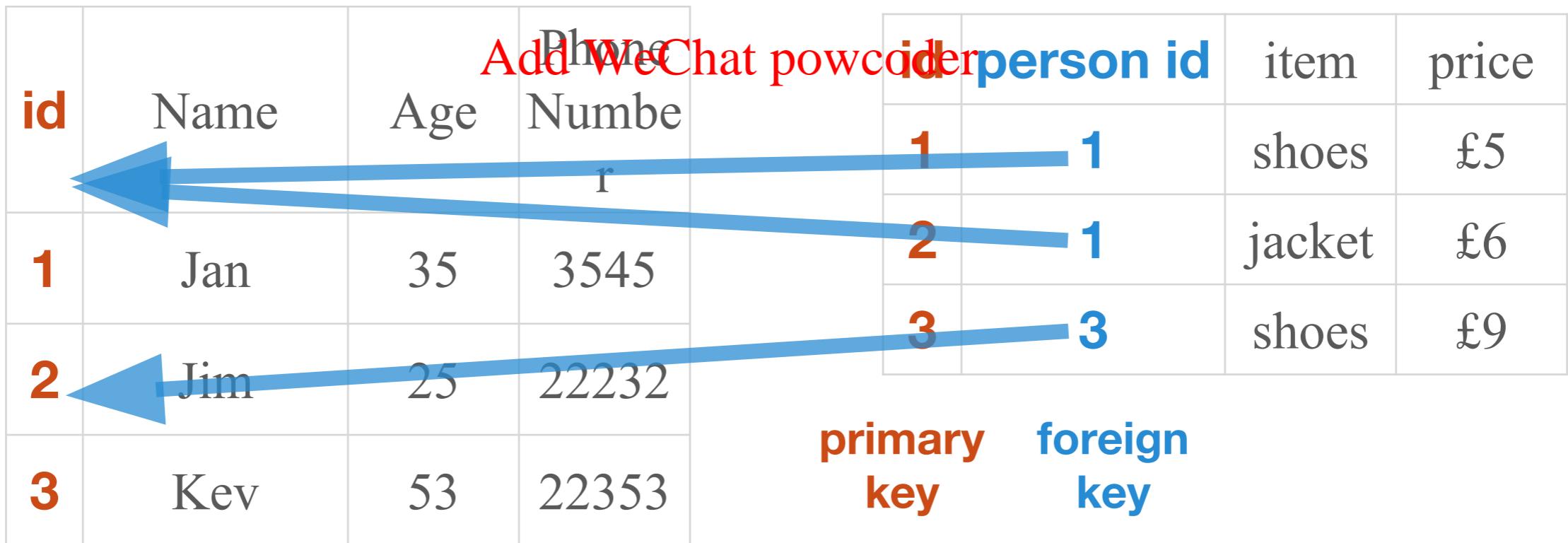
One to Many

In an online store, each person might have bought *many* items, but
each transaction is made by *one* person

Assignment Project Exam Help
Can't add columns to the database - don't know how many there might be
Add WeChat powcoder

Add a table for transactions, that, references the person who bought it
Assignment Project Exam Help

"Foreign Key" - id of a thing in another table



Many to Many

On Facebook, you can "poke" people. Each person can poke *many* people and be poked by *many* people

<https://powcoder.com>

Add a table for Pokes, that points to the poker and the pokee

[Assignment](#) [Project](#) [Exam](#) [Help](#) [Add WeChat powcoder](#)

id	Name	Age	Phone Number	People	Assignment Project Exam Help
1	Jan	35	3545		https://powcoder.com
2	Jim	25	22232		Add WeChat powcoder
3	Kev	53	22353		

Pokes

id	poker	pokee	date
1	1	2	shoes
2	1	3	
3	2	3	jacket
4	3	1	shoes

Many to Many

At Uni, you are enrolled on *many* courses, and each course has *many* people on it
<https://powcoder.com>

Create a table for Enrollments, that points to both tables
Assignment Project Exam Help

Add WeChat powcoder

id	Name	Age	Phone Number
1	Jan	35	3545
2	Jim	25	22232
3	Kev	53	22353

Phone

Number

Assignment Project Exam Help

<https://powcoder.com>

id	Name	Credits
DESI11100	Data Science for Design	20
DESI11073	Histories and Futures	20
INFR11094	Case Studies 1	20

id

Name

Credits

Data Science for Design

20

Histories and Futures

20

Case Studies 1

20

Add WeChat powcoder

id	Student	Course	Mark
1	1	DESI11100	NULL
2	1	DESI11073	NULL
3	2	DESI11100	NULL
4	3	DESI11100	NULL

id

Student

Course

Mark

DESI11100

NULL

DESI11073

NULL

DESI11100

NULL

DESI11100

NULL

(Joins)

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder
Enrollments Joined with People and Courses

Assignment Project Exam Help

id	Student	Course	https://powcoder.com	Mark	Person.Name	P.Age	Course.Credits
1	1	DESI1100	NULL	Mark	John	35	20
2	1	DESI11073	NULL	Jan		35	20
3	2	DESI1100	NULL	Jim		25	20
4	3	DESI1100	NULL	Kev		53	20

More info

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

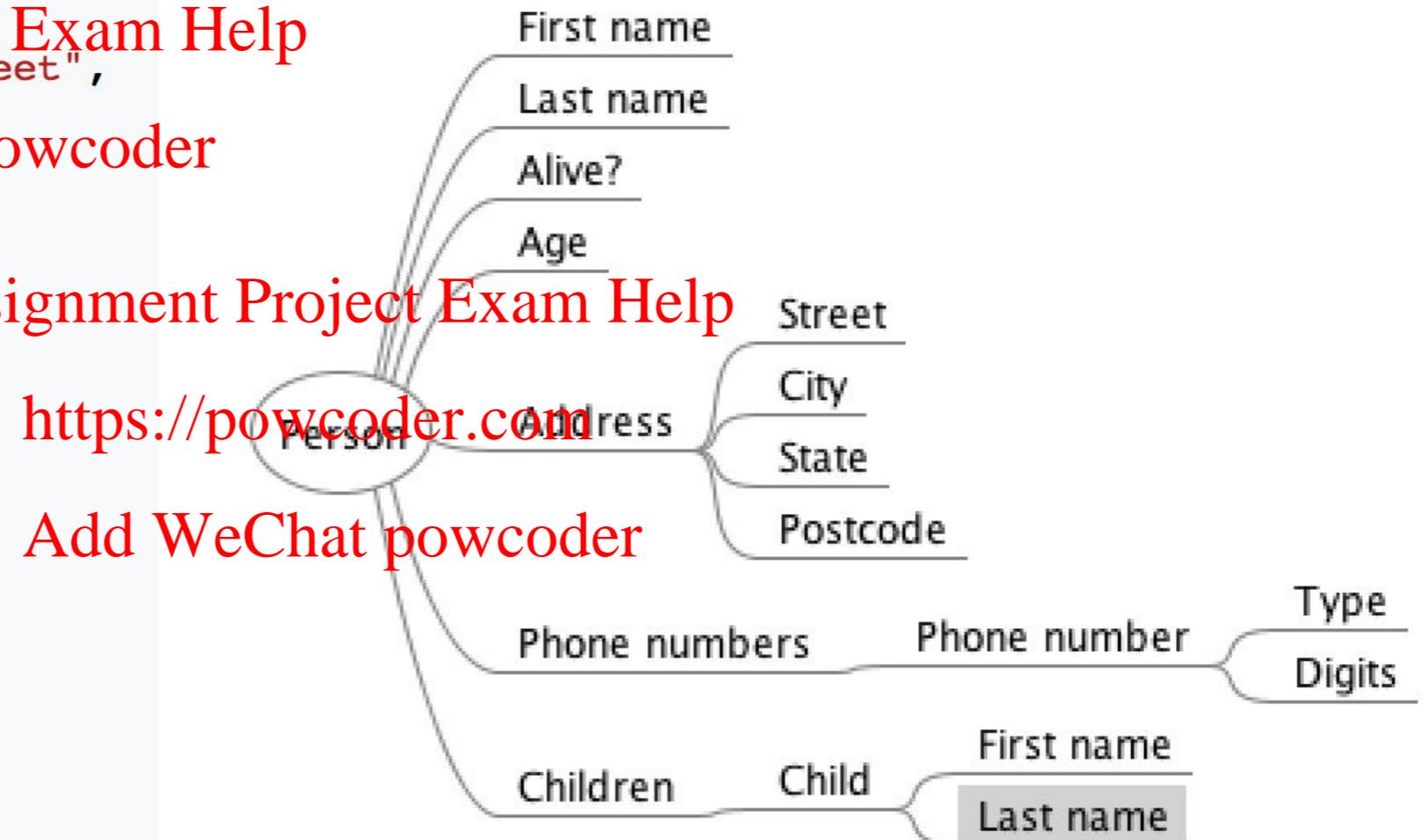
Assignment Project Exam Help

- <http://www.ucl.ac.uk/archaeology/cisp/database/manual/node1.html> Add WeChat powcoder

Hierarchical Data

- Data that contains other data
[Assignment Project Exam Help](https://powcoder.com)
- Lists of items, items with variable numbers of fields, fields that are items etc.
[Assignment Project Exam Help](https://powcoder.com)
- Can be more or less defined
[Add WeChat powcoder](https://powcoder.com)
 - JSON on its own can have a huge range of structures
 - An application's JSON probably has a particular structure, so people can make use of it

```
{
  "firstName": "John",
  "lastName": "Smith",
  "isAlive": true,
  "age": 25,
  "address": {
    "streetAddress": "21 2nd Street",
    "city": "New York",
    "state": "NY",
    "postalCode": "10021-3100"
  },
  "phoneNumbers": [
    {
      "type": "home",
      "number": "212 555-1234"
    },
    {
      "type": "office",
      "number": "646 555-4567"
    },
    {
      "type": "mobile",
      "number": "123 456-7890"
    }
  ],
  "children": [],
  "spouse": null
}
```



Translating

The image shows a code editor interface with three tabs:

- people.json**: A JSON file containing two objects. The first object has a "name" field with a red placeholder "Assignment Project Exam Help". The second object has a "name" field with a red placeholder "Assignment Project Exam Help".
- people.xml**: An XML file with a root element <people>. It contains two <person> elements. Each <person> element has a <name> field with a red placeholder "Assignment Project Exam Help".
- people.csv**: A CSV file with three rows. The first row is "Name, Age, Phone Number". The second row is "Kal, 27, 07777 777777". The third row is "Jan, 34, 0131 555 5555".

The code editor has a dark theme with syntax highlighting. Red placeholder text is used to indicate where user input or generated code should be placed.

```
people.json
[{"name": "Assignment Project Exam Help", "age": 27, "phone": "0777 777777"}, {"name": "Assignment Project Exam Help", "age": 34, "phone": "0131 555 5555"}]
```

```
people.xml
<?xml version="1.0" encoding="UTF-8"?>
<people>
  <person>
    <name>Assignment Project Exam Help</name>
    <age>27</age>
    <phone>0777 777777</phone>
  </person>
  <person>
    <name>Assignment Project Exam Help</name>
    <age>34</age>
    <phone>0131 555 5555</phone>
  </person>
</people>
```

```
people.csv
Name, Age, Phone Number
Kal, 27, 07777 777777
Jan, 34, 0131 555 5555
```

Graph Data

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

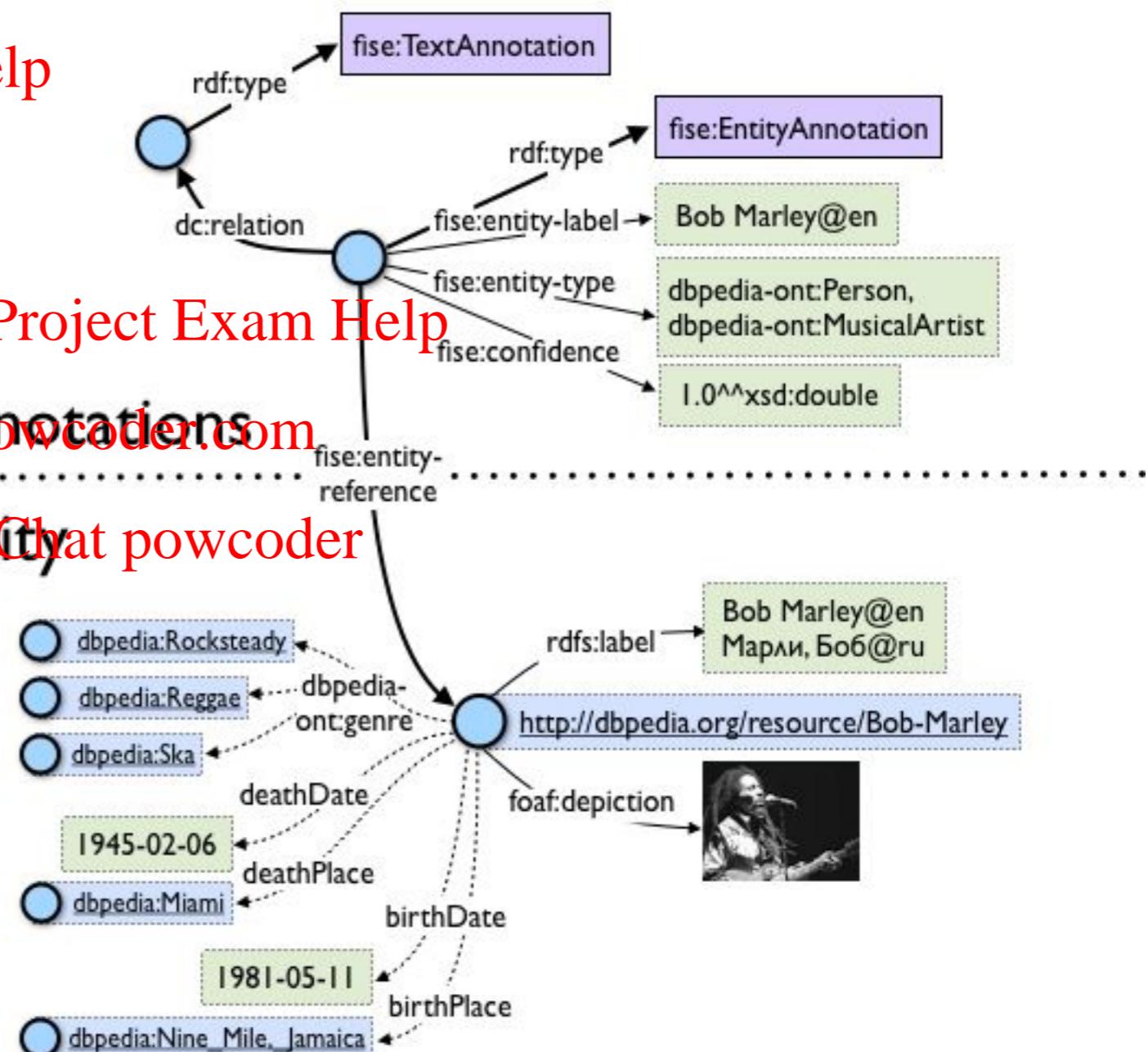
- Nodes with links between them

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- Nodes are *things*, links are *relations*



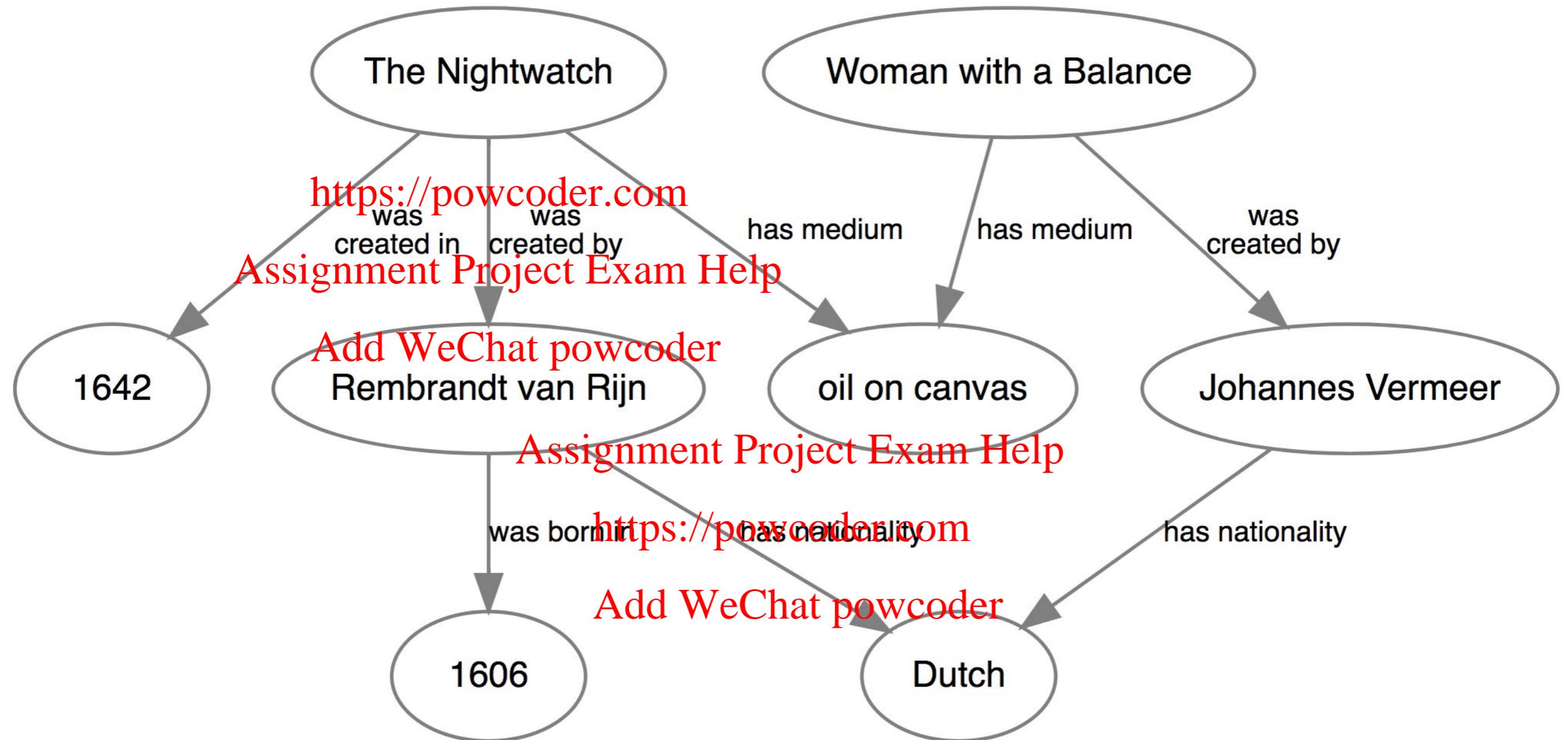


Image from:

<https://programminghistorian.org/lessons/graph-databases-and-SPARQL>

(lots of good information about graph queries there)

Graph Data

<https://powcoder.com>

Often stored
Assignment Project Exam Help

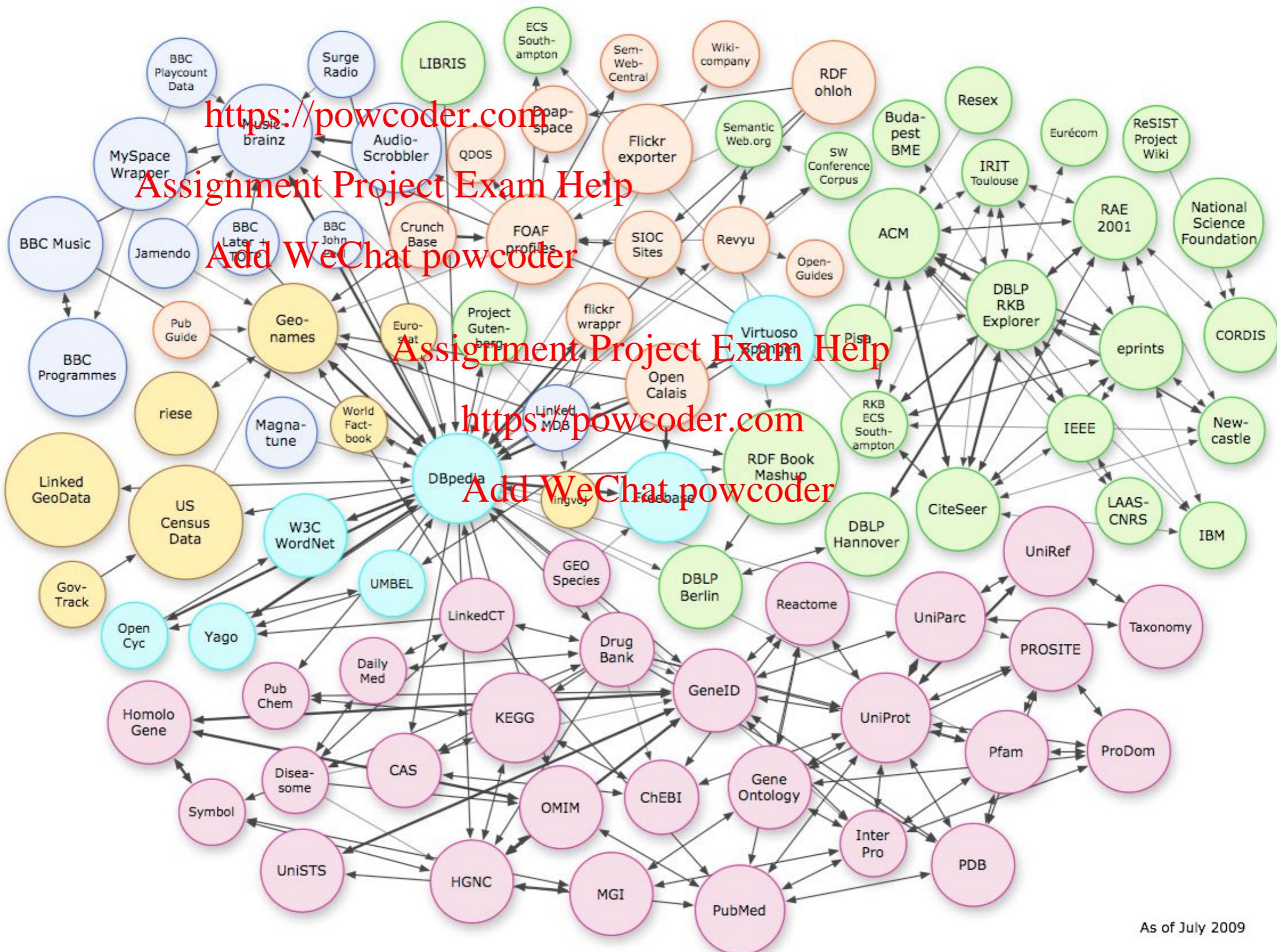
as "triples"
Add WeChat powcoder

Subject,
Predicate,
Object

Using URIs
relates your
data to
other
peoples

Subject	Predicate	Object
Romke van Rijn	hasNationality	Dutch
https://powcoder.com	likes	ice cream
d.murray-rust@e.d.ac.uk	food:likes	http://dbpedia.org/resource/Ice_cream

Semantic Web / Linked Open Data



Summary

<https://powcoder.com>

Tabular: simple, easy to work with and author, maps
Assignment Project Exam Help
onto **DataFrames**

Add WeChat powcoder

Relational: adds links between tables, can represent
more structure. Use SQL or other query languages to
access

<https://powcoder.com>

Add WeChat powcoder

Hierarchical: create complex nested structures. Maps
onto programming concepts, new generation of
databases (object- or document-databases)

Graph: good for linking

Data Wrangling

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

- Read data - get it available within your system
Assignment Project Exam Help
- Clean data - missing values, incorrect fields, outliers - filtering, harmonising
Add WeChat powcoder
- Bring data together

Tools

Many many tools

<https://powcoder.com>

Excel can do many things, but hit limitations

[Assignment](#) [Project](#) [Exam](#) [Help](#)

Programming languages vs Toolkits (some overlap)

R is a reasonable language, but its built in data structures make it popular, and good graphing tools

[Assignment](#) [Project](#) [Exam](#) [Help](#)
<https://powcoder.com>

Python is a good all purpose language, but needs a good toolkit for serious data processing

[Add](#) [WeChat](#) [powcoder](#)

(but there are several)

Javascript (with D3) can be great for visualisation

Processing can be good for interactives

Other languages (C++, Perl, Scala, Java, Haskell etc.) get used

Text editors – try Atom if you don't have a favourite

Command line: if you get good, you can do complicated things really quickly

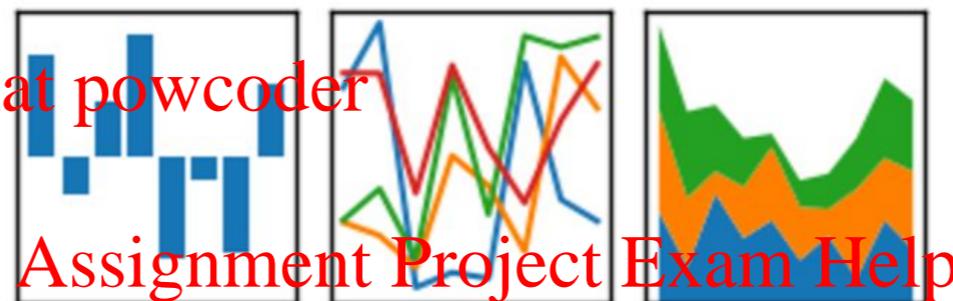
Pandas

<https://powcoder.com>

Assignment Project Exam Help

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



<https://powcoder.com>

[home](#) // [about](#) // [get pandas](#) // [documentation](#) // [community](#) // [talks](#) // [donate](#)

Python Data Analysis Library

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](#) programming language.

pandas is a [NUMFocus](#) sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to [donate](#) to the project.

A Fiscally Sponsored Project of
NUMFOCUS
OPEN CODE = BETTER SCIENCE

VERSIONS

Release

0.20.3 - July 2017

[download](#) // [docs](#) // [pdf](#)

Development

0.21.0 - 2017

[github](#) // [docs](#)

Previous Releases

0.19.2 - [download](#) // [docs](#) // [pdf](#)

0.18.1 - [download](#) // [docs](#) // [pdf](#)

0.17.1 - [download](#) // [docs](#) // [pdf](#)

0.16.2 - [download](#) // [docs](#) // [pdf](#)

Pandas

pandas is well suited for many different kinds of data:

Assignment Project Exam Help
Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet
<https://powcoder.com>
[Add WeChat powcoder](#)

Ordered and unordered (not necessarily fixed-frequency) time series data.

Assignment Project Exam Help
<https://powcoder.com>
Arbitrary matrix data (homogeneously typed or heterogeneous) with row and column labels
[Add WeChat powcoder](#)

Any other form of observational / statistical data sets. The data actually need not be labeled at all to be placed into a pandas data structure

Two primary data structures: Series (1-dimensional) and DataFrame (2-dimensional)

Pandas Features

Easy handling of missing data (represented as NaN) in floating point as well as non-floating point data
<https://powcoder.com>

Size mutability: columns can be inserted and deleted from DataFrame and higher dimensional objects
[Assignment](#) [Project](#) [Exam](#) [Help](#)

Automatic and explicit data alignment: objects can be explicitly aligned to a set of labels, or the user can simply ignore the labels and let Series, DataFrame, etc. automatically align the data for you in computations
[Add WeChat](#) [powcoder](#)

Powerful, flexible group by functionality to perform split-apply-combine operations on data sets, for both aggregating and transforming data
<https://powcoder.com>

Make it easy to convert ragged, differently-indexed data in other Python and NumPy data structures into DataFrame objects
[Add WeChat](#) [powcoder](#)

Intelligent label-based slicing, fancy indexing, and subsetting of large data sets

Intuitive merging and joining data sets

Flexible reshaping and pivoting of data sets

Hierarchical labeling of axes (possible to have multiple labels per tick)

Robust IO tools for loading data from flat files (CSV and delimited), Excel files, databases, and saving / loading data from the ultrafast HDF5 format

Time series-specific functionality: date range generation and frequency conversion, moving window statistics, moving window linear regressions, date shifting and lagging, etc.

Data Frames

<https://powcoder.com>

Assignment Project Exam Help
Data structures with named columns of specific types
Add WeChat powcoder

```
In [3]: df2 = pd.DataFrame({'Id': ["id5", "id6"],  
                           'Age': [11, 23],  
                           'Eyes': ['Yellow', 'Slate'],  
                           'Hair': ['Green', 'Purple'],  
                           'Name': ['Jo', 'Jun']})  
df2
```

Out[3]:

	Id	Age	Eyes	Hair	Name
0	id5	11	Yellow	Green	Jo
1	id6	23	Slate	Purple	Jun

10 Minutes to pandas:

[http://pandas.pydata.org/pandas-docs/stable/10min.html](https://pandas.pydata.org/pandas-docs/stable/10min.html)

Assignment Project Exam Help

Intro to data structures:

[http://pandas.pydata.org/pandas-docs/stable/dsintro.html](https://pandas.pydata.org/pandas-docs/stable/dsintro.html)

Assignment Project Exam Help
https://powcoder.com

All of their tutorials:

[http://pandas.pydata.org/pandas-docs/stable/tutorials.html](https://pandas.pydata.org/pandas-docs/stable/tutorials.html)

Pandas cookbook:

<https://github.com/jvns/pandas-cookbook>),
chapters on slicing (Chapter 2,3)

Bringing data together

<https://powcoder.com>

The `concat` function (in the main pandas namespace) does all of the heavy lifting of performing concatenation operations along an axis while performing optional set logic (union or intersection) of the indexes (if any) on the other axes. Note that I say "if any" because there is only a single possible axis of concatenation for Series.

<https://pandas.pydata.org/pandas-docs/stable/merging.html>

Vertical Concatenation

<https://powcoder.com>

	Age	Eyes	Hair	Id	Name
0	35	Blue	Black	id1	Jon
1	67	Blue	Black	id2	Jack
2	12	Grey	Brown	id3	Jim
3	24	Brown	Pink	id4	Jen

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

	Age	Eyes	Hair	Id	Name
0	11	Yellow	Green	id5	Jo
1	23	Slate	Purple	id6	Jun

	Age	Eyes	Hair	Id	Name
0	35	Blue	Black	id1	Jon
1	67	Blue	Black	id2	Jack
2	12	Grey	Brown	id3	Jim
3	24	Brown	Pink	id4	Jen
4	11	Yellow	Green	id5	Jo
5	23	Slate	Purple	id6	Jun

Same Fields, Same Order

Horizontal Concatenation

Assignment Project Exam Help
Add WeChat powcoder

	Age	Eyes	Hair	Id	Name		FavFood	NickName
0	35	Blue	Black	id1	Jon		Chips	Jonny
1	67	Blue	Black	id2	Jack		Cheese	Jackie
2	12	Grey	Brown	id3	Jim		Broccoli	Jimmy
3	24	Brown	Pink	id4	Jen		Candyfloss	Jenny

Same Rows, Same Order

Add WeChat powcoder

	Age	Eyes	Hair	Id	Name	FavFood	NickName
0	35	Blue	Black	id1	Jon	Chips	Jonny
1	67	Blue	Black	id2	Jack	Cheese	Jackie
2	12	Grey	Brown	id3	Jim	Broccoli	Jimmy
3	24	Brown	Pink	id4	Jen	Candyfloss	Jenny

Joining

<https://powcoder.com>

	Age	Eyes	Hair	Id	Name
0	35	Blue	Black	id1	Jon
1	67	Blue	Black	id2	Jack
2	12	Grey	Brown	id3	Jim
3	24	Brown	Pink	id4	Jen

Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help

<https://powcoder.com>

	FavFood	NickName
0	Candyfloss	Jenny
1	Chips	Jonny
2	Bananas	Ray

Add WeChat powcoder

Joining

<https://powcoder.com>

	Age	Eyes	Hair	Id	Name
0	35	Blue	Black	id1	Jon
1	67	Blue	Black	id2	Jack
2	12	Grey	Brown	id3	Jim
3	24	Brown	Pink	id4	Jen

	FavFood	Ident	NickName
0	Candyfloss	id4	Jenny
1	Chips	id1	Jonny
2	Bananas	id12	Ray

pd.merge(df1, df4b, left_on="Id", right_on="Ident")

	Age	Eyes	Hair	Id	Name	FavFood	Ident	NickName
0	35	Blue	Black	id1	Jon	Chips	id1	Jonny
1	24	Brown	Pink	id4	Jen	Candyfloss	id4	Jenny

Use 'how' parameter to specify:

Join Types

inner (default): only keep rows that are in both tables

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

outer: keep rows that are in either table

	Age	Eyes	Hair	Id	Name
0	35	Blue	Black	id1	Jon
1	67	Blue	Black	id2	Jack
2	12	Grey	Brown	id3	Jim
3	24	Brown	Pink	id4	Jen

	FavFood	Ident	NickName
0	Candyfloss	id4	Jenny
1	Chips	id1	Jonny
2	Bananas	id12	Ray

Assignment Project Exam Help

left / right: keep all the rows in the left or right table

<https://powcoder.com>

`pd.merge(df1,df4b, left_on="Id", right_on="Ident", how="outer")`

Add WeChat powcoder

	Age	Eyes	Hair	Id	Name	FavFood	Ident	NickName
0	35.0	Blue	Black	id1	Jon	Chips	id1	Jonny
1	67.0	Blue	Black	id2	Jack	NaN	NaN	NaN
2	12.0	Grey	Brown	id3	Jim	NaN	NaN	NaN
3	24.0	Brown	Pink	id4	Jen	Candyfloss	id4	Jenny
4	NaN	NaN	NaN	NaN	NaN	Bananas	id12	Ray

Summary

<https://powcoder.com>
Philosophical/historical introduction to data
Assignment Project Exam Help

Records, Fields, Schema, Format etc.

Data types: booleans, numbers, strings, compound types

<https://powcoder.com>
Data formats: CSV, JSON, XML
Add WeChat powcoder

Data shapes: Tabular, Relational, Hierarchical, Graph

Pandas

Joining data together

Homework!

<https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder

- Reading Assignment Project Exam Help
 - Coding <https://powcoder.com>
 - git Add WeChat powcoder
 - Start with Assignment 1

Reading

<https://powcoder.com>

- Read: Thatcher, Jim, David O'Sullivan, and Dillon Mahmoudi. "Data colonialism through accumulation by dispossession: New metaphors for daily data." *Environment and Planning D: Society and Space* 34.6 (2016): 990-1006.
- Work through some of the Pandas cookbook (<https://github.com/jvns/pandas-cookbook>) at least Chapters 1, 2, 3 possibly 4

Article

Data colonialism through accumulation by dispossession: New metaphors for daily data

Jim Thatcher

University of Washington—Tacoma, USA

David O'Sullivan

University of California, Berkeley, USA

Dillon Mahmoudi

Portland State University, USA

Abstract

In recent years, much has been written on 'big data' in the hubristic declaration of the 'end of theory' suggesting that increasingly pervasive data collection implications for the social sciences, even if the social, behind big data are less new than they are often portrayed. Much of its press, academic critiques of big data have continued importance of more traditional forms of many academic responses to big data enthusiastically claim and the potential for new insights and perspectives these critiques is a lack of attention to the role of technology labor process, the continued extension of labor relations and the commoditization of more and more aspects variety of big data definitions to argue that it is only when or more are linked together algorithmically that 'big data do not occur in a vacuum but as part of an asymmetric

Coding

- Data Visualisation Notebooks

<https://powcoder.com>

- **essential for Assignment 1**

Assignment Project Exam Help

- Defensive Programming Notebook

[Add WeChat powcoder](#)

- **also essential for Assignment 1**

Assignment Project Exam Help

- Join / Merge Notebooks

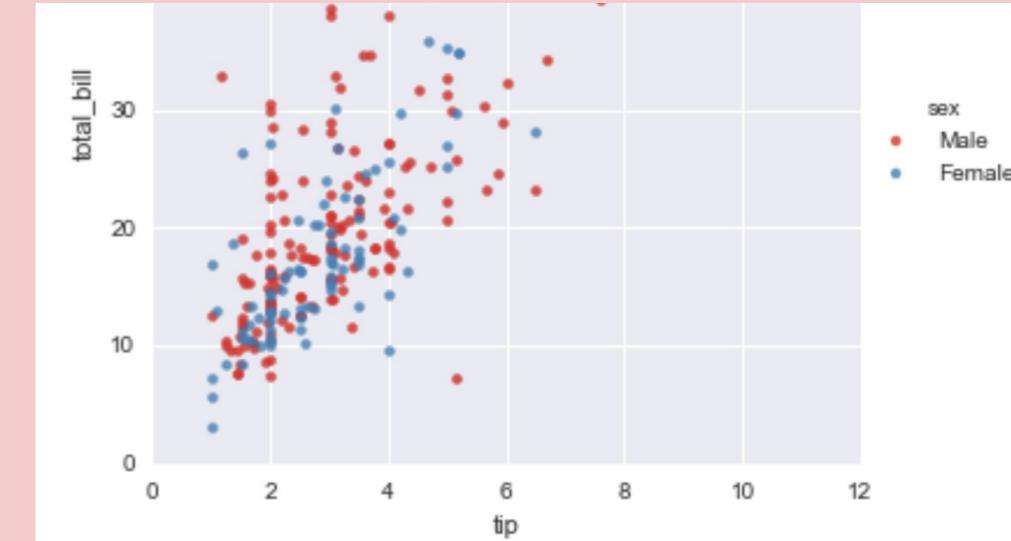
<https://powcoder.com>

- Less essential, but very useful. Try the [Add WeChat powcoder](#) exercises, and also check out the example joins

- Extra Learning: Pandas Cookbook

- <https://github.com/jvns/pandas-cookbook>

not essential right now, but if you want to get good, work through some of the Pandas cookbook – might mean you know the right things to do when you have data. Look at Chapters 1, 2, 3 possibly 4.



```
assert range_overlap([
```

```
assert range_overlap([
```

```
assert range_overlap([
```

	Age	Eyes	Hair	Id	Name	FavFood	Ident	NickName
0	35.0	Blue	Black	id1	Jon	Chips	id1	Jonny
1	67.0	Blue	Black	id2	Jack	NaN	NaN	NaN
2	12.0	Grey	Brown	id3	Jim	NaN	NaN	NaN
3	24.0	Brown	Pink	id4	Jen	Candyfloss	id4	Jenny
4	NaN	NaN	NaN	NaN	NaN	Bananas	id12	Ray

[jvns / pandas-cookbook](#)

Code

Issues 11

Pull requests 4

Projects

Recipes for using Python's pandas library

To get set up, download the whole thing as a zip file from their github, upload individual notebooks and data to your server as necessary

git - version control

- Git will save your life (or at least your code)

<https://powcoder.com>

- Code Repository: stores every change you make, so when you break something, you always got a previous version that works

[Assignment Project Exam Help](#)
[Add WeChat powcoder](#)

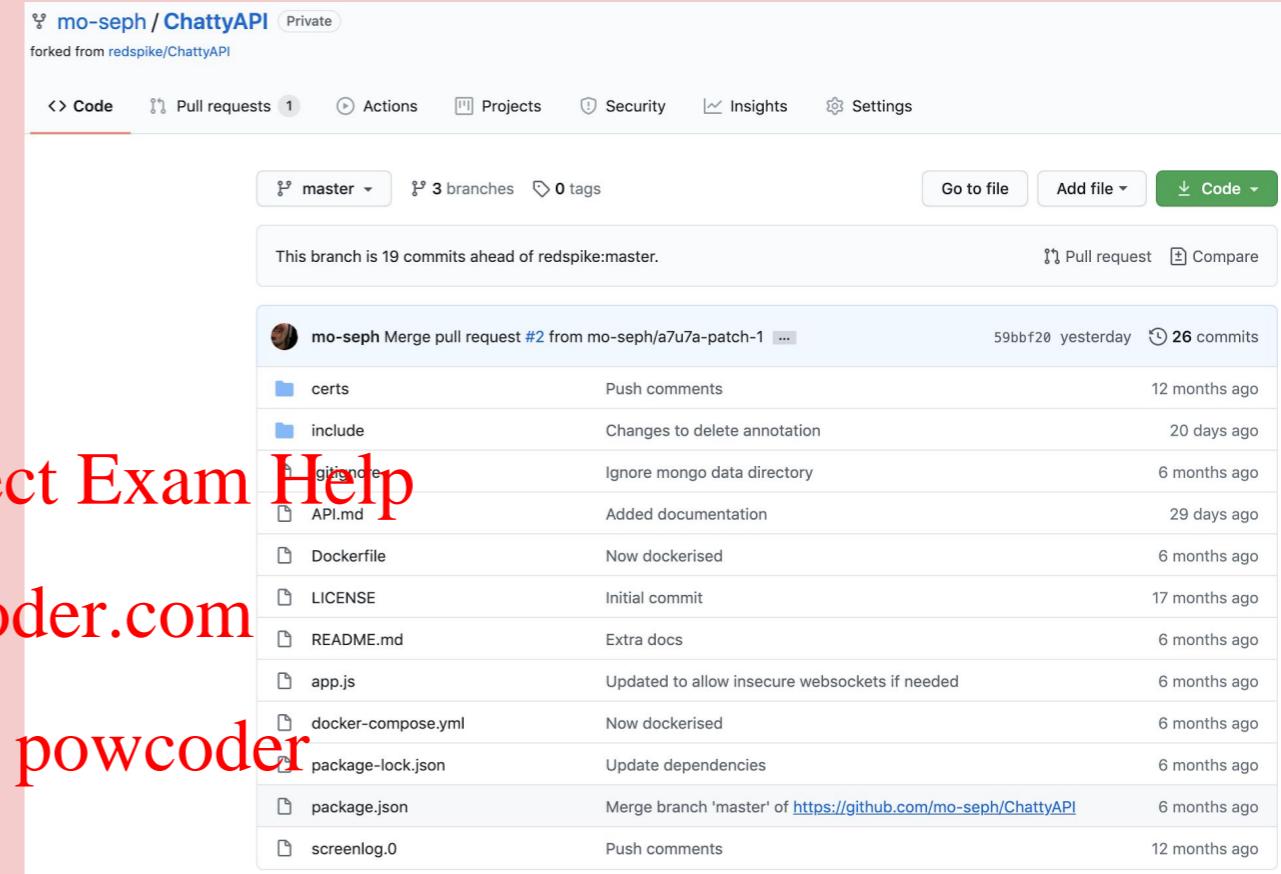
- ... and you don't have 100 files all called *real_final_version_1.4_finished.py*

[Assignment Project Exam Help](#)

- Github is how proper software projects are managed

[Add WeChat powcoder](#)

- You will be assessed on your use of git – start early!



- Step 1: go through Evan Morgan's Intro to Git – short videos and tutorial sheet
- Step 2: make a private github repository for your group
- Step 3: make sure that everyone in your group has cloned it, added a file and pushed the results