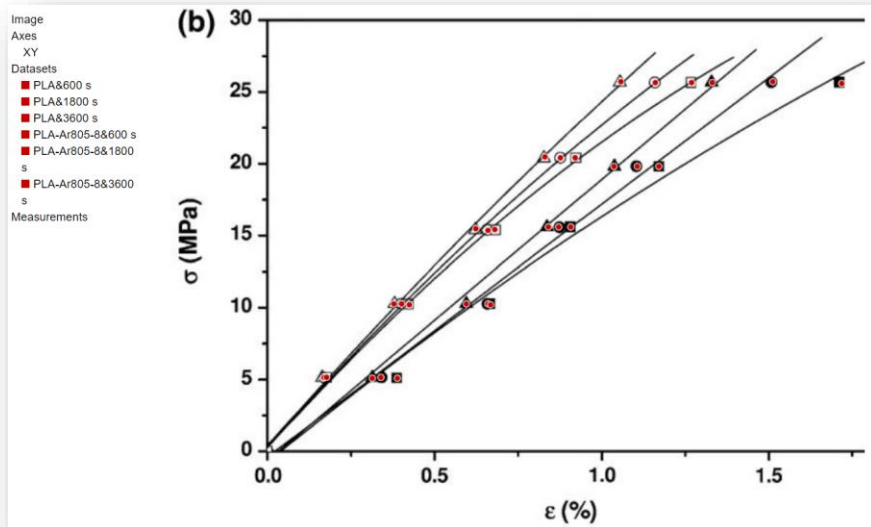


# Getting Started with TidyCatWPD

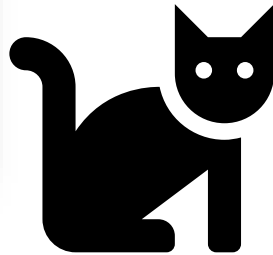
1. [What is the TidyCatWPD App?](#)
2. [What “Tidy” Data Means](#)
3. [Creating the Configuration File](#)
4. [Naming Data Series in WebPlotDigitizer](#)
5. [Exporting Data from WebPlotDigitizer](#)
6. [Overview of the Grammar of Graphics](#)
7. [Making Tidy Data Visual with Vega-Lite](#)

“Tidy!”

# What is the TidyCatWPD App?



I have one job... I turn  
[WebPlotDigitizer](#) data  
into a tidy format!



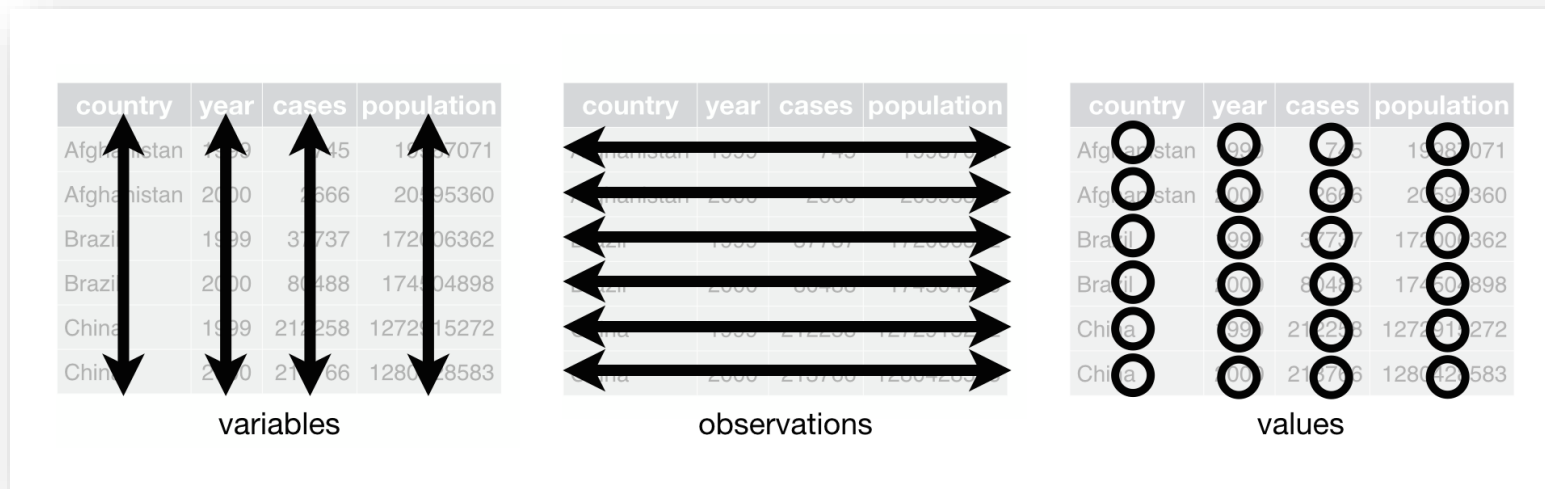
	A	B	C	D
1	X	Y	Sample	Time
2	0.314262	5.024497	PLA	600 s
3	0.59749	10.17213	PLA	600 s
4	0.844558	15.50303	PLA	600 s
5	1.042194	19.6956	PLA	600 s
6	1.338523	25.50351	PLA	600 s
7	0.340139	5.082543	PLA	1800 s
8	0.662143	10.16695	PLA	1800 s
9	0.875592	15.50054	PLA	1800 s
10	1.11202	19.69001	PLA	1800 s
11	1.51957	25.54912	PLA	1800 s
12	0.38926	5.018485	PLA	3600 s
13	0.669886	10.10621	PLA	3600 s
14	0.911798	15.49764	PLA	3600 s
15	1.176674	19.68482	PLA	3600 s
16	1.726431	25.4123	PLA	3600 s
17	0.166851	5.036313	PLA-Ar805	600 s
18	0.380253	10.18955	PLA-Ar805	600 s
19	0.62729	15.40021	PLA-Ar805	600 s
20	0.835473	20.3735	PLA-Ar805	600 s
21	1.064407	25.58561	PLA-Ar805	600 s
22	0.169453	5.096225	PLA-Ar805	1800 s
23	0.403528	10.18768	PLA-Ar805	1800 s
24	0.663465	15.27707	PLA-Ar805	1800 s
25	0.882008	20.30965	PLA-Ar805	1800 s
26	1.167837	25.5172	PLA-Ar805	1800 s
27	0.177211	5.095603	PLA-Ar805	3600 s
28	0.426788	10.1257	PLA-Ar805	3600 s
29	0.68417	15.33553	PLA-Ar805	3600 s
30	0.925972	20.30612	PLA-Ar805	3600 s
31	1.276456	25.50849	PLA-Ar805	3600 s

Not “tidy!” (variables X and Y repeat, unique labels in first row)

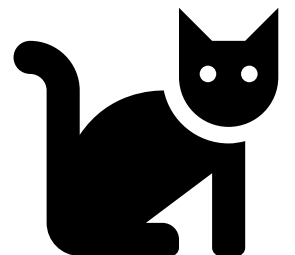
	A	B	C	D	E	F	G	H	I	J	K	L
1	PLA&600 s		PLA&1800 s		PLA&3600 s		PLA-Ar805-8&600 s		PLA-Ar805-8&1800 s		PLA-Ar805-8&3600 s	
2	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
3	0.314262	5.024497	0.340139	5.082543	0.38926	5.018485	0.166851	5.036313	0.169453	5.096225	0.177211	5.095603
4	0.59749	10.17213	0.662143	10.16695	0.669886	10.10621	0.380253	10.18955	0.403528	10.18768	0.426788	10.1257
5	0.844558	15.50303	0.875592	15.50054	0.911798	15.49764	0.62729	15.40021	0.663465	15.27707	0.68417	15.33553
6	1.042194	19.6956	1.11202	19.69001	1.176674	19.68482	0.835473	20.3735	0.882008	20.30965	0.925972	20.30612
7	1.338523	25.50351	1.51957	25.54912	1.726431	25.4123	1.064407	25.58561	1.167837	25.5172	1.276456	25.50849

# What “Tidy” Data Means

- Each **variable** has its own column
- Each row fully describes a single **observation**
- Each cell contains the **value** of a variable for a given observation



Tidy data  
makes analysis  
much easier!



# Creating the Configuration File

- Create a new file (e.g., tidyconfig.csv) in a spreadsheet program
  - In Excel, use **Save As** > Save as type > CSV (Comma delimited) (\*.csv)

File name: tidyconfig.csv  
Save as type: CSV (Comma delimited) (\*.csv)

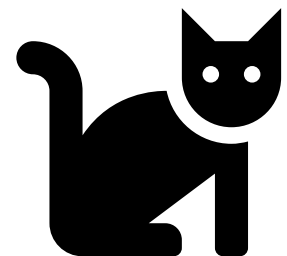
- Config file has 2 columns: *filename* and *series\_decode*

filename	series_decode
Figure2.csv	Label
Figure3a.csv	Label
Figure3b.csv	Label
Figure3c.csv	Label
Figure4a.csv	Label
Figure4b.csv	Label
Figure4c.csv	Label
Figure4d.csv	Label
Figure4e.csv	Label
Figure5.csv	Label
Figure6.csv	Label
Figure7a.csv	Label
Figure7b.csv	Label
Figure8a.csv	Label
Figure8b.csv	Sample&Time
Figure11.csv	Label

I can tidy as many files  
as you wish to upload!

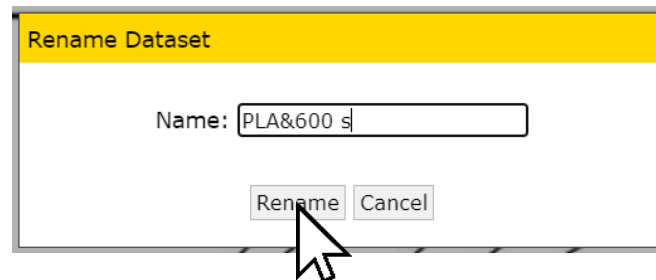
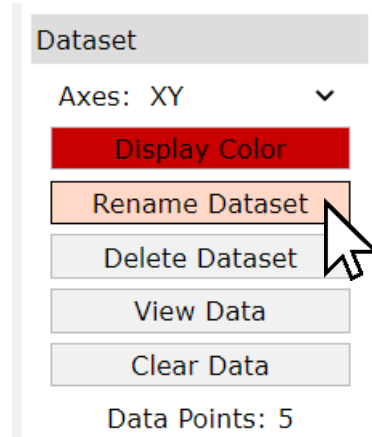
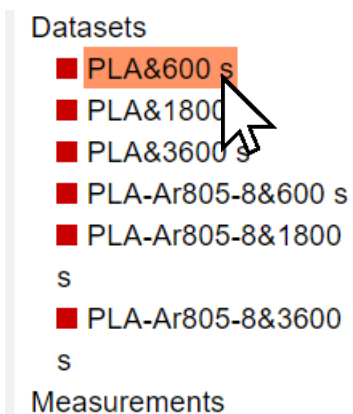
***series\_decode*** tells TidyCat the **names** of the columns (in addition to X and Y) that your tidy files will contain.

Typically, this is just one column (e.g., Label), but if you have **multiple variables** describing each data series, separate them with the '&' symbol (e.g., Sample&Time).

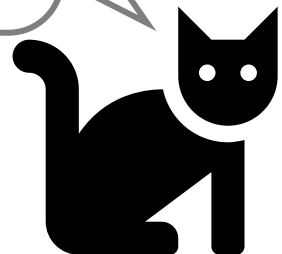


# Naming Data Series in WebPlotDigitizer

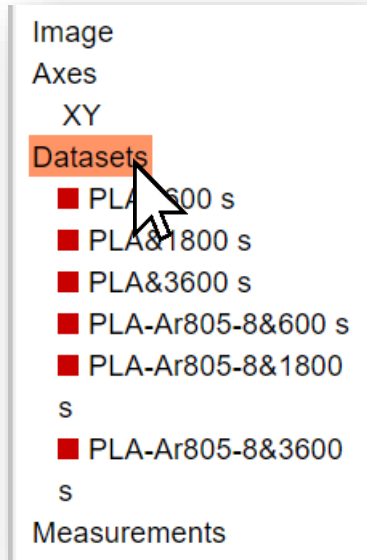
- Variable **names** are stored in the Configuration File, and variable **values** are assigned by naming data series within WebPlotDigitizer
- If you only have one extra variable (besides X and Y), then the new variable simply be the name of the dataset
- For additional variable values, separate them with the ‘&’ symbol



For variables values separated by ‘&’, use the **same order** used for variable names in the Configuration File!

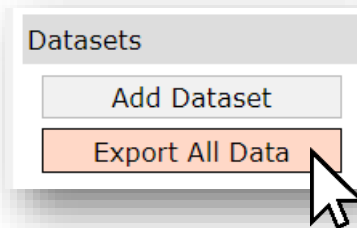


# Exporting Data from WebPlotDigitizer

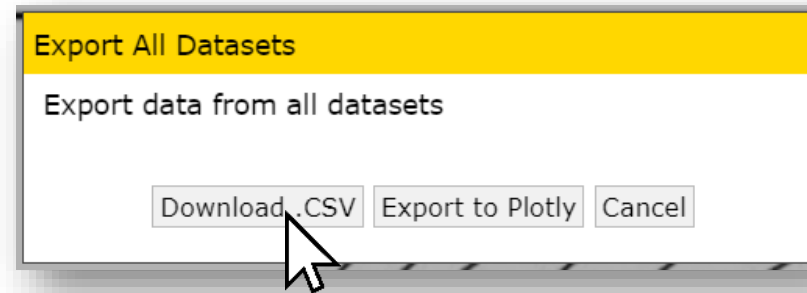


1) Click **Datasets**

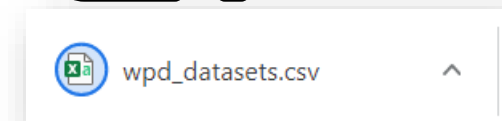
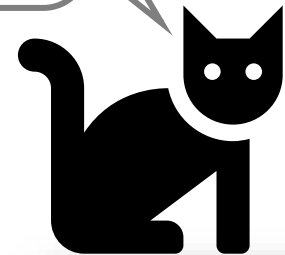
2) Click **Export All Data**



3) Click **Download .CSV**



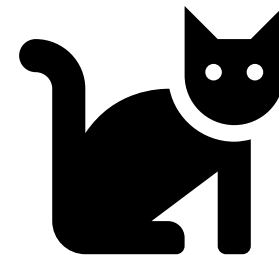
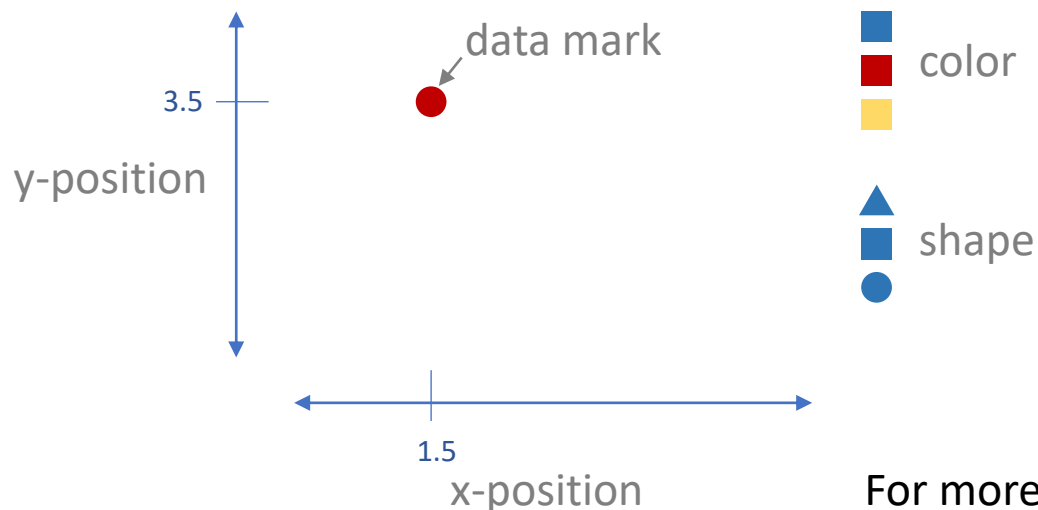
Don't forget to save your WPD project (.tar) in case you need to come back later!



4) **Rename** the dataset according to the filename in your Configuration File

# Overview of the Grammar of Graphics

- The grammar of graphics describes the *deeper structure* of charts
- Using visual *encodings*, variable values are mapped onto categorical or continuous visual scales (position, color, shape, size, etc.)
- Data are represented by *marks*, whose properties can be described by the encodings (e.g.,  $x = 1.5$ ,  $y = 3.5$ , color = “red”, shape = “circle”)



There are many ways to represent the same data visually!

For more info, check out: Wilkinson, Leland. *The Grammar of Graphics*. (1999)

# Making Tidy Data Visual with [Vega-Lite](#)

- A Vega-Lite *chart specification* defines how to transform data visually
- The chart spec is independent from the data itself, so you can swap out datasets as long as the data follow the same basic structure

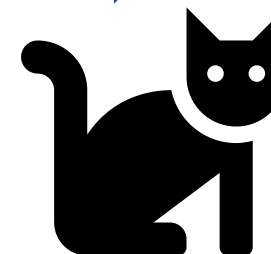
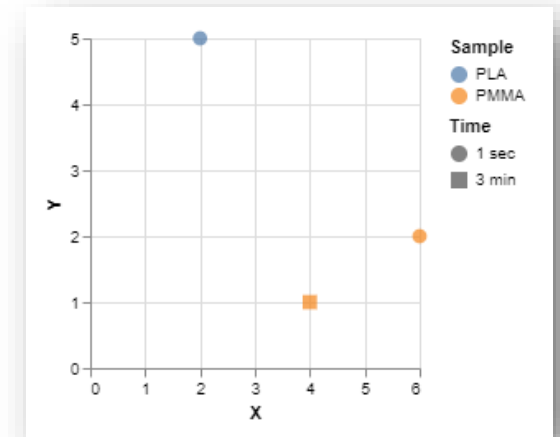
Example Vega-Lite chart spec (JSON format):

```
{
  "data": {"url": "mydata.csv"},
  "mark": {"type": "point", "filled": true, "size": 75},
  "encoding": {
    "x": {"field": "X", "type": "quantitative"},
    "y": {"field": "Y", "type": "quantitative"},
    "color": {"field": "Sample", "type": "nominal"},
    "shape": {"field": "Time", "type": "nominal"}
  }
}
```

“**field**” references correspond to the **column headers** in your data file

X	Y	Sample	Time
2	5	PLA	1 sec
6	2	PMMA	1 sec
4	1	PMMA	3 min

Vega-Lite renderer uses smart defaults to generate encoding scales based on the values in the provided dataset



Vega-Lite charts can always be customized further within the chart spec!



**TidyCatWPD** is an application written with [R Shiny](#).

Author: Michael Deagen

Date: 25 Feb. 2021

Code: <https://github.com/mdeagen/TidyCatWPD>

Tutorial revision history:

- 2021-02-25: Created Tutorial

