

# class\_\_design

Nathan Hughes

March 4, 2018

## Contents

<b>1</b>	<b><u>init.py</u></b>	<b>2</b>
<b>2</b>	<b>ct<sub>data</sub>.py</b>	<b>2</b>
2.1	CTData . . . . .	2
2.1.1	<u>init</u> . . . . .	2
2.1.2	get <sub>data</sub> . . . . .	2
2.1.3	gather <sub>data</sub> . . . . .	2
2.1.4	make <sub>dataframe</sub> . . . . .	2
2.1.5	clean <sub>data</sub> . . . . .	2
2.1.6	get <sub>files</sub> . . . . .	2
2.1.7	fix <sub>colnames</sub> . . . . .	3
2.1.8	join <sub>spikesbyrachis</sub> . . . . .	3
2.1.9	remove <sub>percentile</sub> . . . . .	3
2.1.10	get <sub>spikeinfo</sub> . . . . .	3
2.1.11	lookup . . . . .	3
2.1.12	gather <sub>data</sub> . . . . .	3
2.1.13	aggregate <sub>spikeaverages</sub> . . . . .	3
2.1.14	make <sub>plot</sub> . . . . .	4
<b>3</b>	<b>graphing.py</b>	<b>4</b>
3.1	Error . . . . .	4
3.2	InvalidPlot . . . . .	4
3.2.1	percentile <sub>grid</sub> . . . . .	4
3.2.2	qq <sub>grid</sub> . . . . .	4
3.2.3	plot <sub>boxplots</sub> . . . . .	4
3.2.4	plot <sub>histogram</sub> . . . . .	4
3.2.5	check <sub>varargs</sub> . . . . .	4

<b>4</b>	<b>scratch.py</b>	<b>4</b>
<b>5</b>	<b>statistical<sub>tests</sub>.py</b>	<b>4</b>
5.0.1	qqplot . . . . .	4
5.0.2	test <sub>normality</sub> . . . . .	4

## **1 init.py**

## **2 ct<sub>data</sub>.py**

### **2.1 CTData**

#### **2.1.1 init**

#### **2.1.2 get<sub>data</sub>**

Returns the dataframe used in this class

#### **2.1.3 gather<sub>data</sub>**

this function gathers together all the data of interest @param folder is a starting folder @returns tuple of (seed files, rachis files)

#### **2.1.4 make<sub>dataframe</sub>**

this function returns a dataframe of grain parameters and optionally of the rachis top and bottom @param grain<sub>files</sub> is the output from gather<sub>data</sub> @param rachis<sub>files</sub> is an optional output from gather<sub>data</sub> also @returns a dataframe of the information pre-joining

#### **2.1.5 clean<sub>data</sub>**

Following parameters outlined in the CT software documentation I remove outliers which are known to be errors

#### **2.1.6 get<sub>files</sub>**

Returns a tuple of grain files and rachis files

### 2.1.7 `fix_colnames`

Because Biologists like to give data which are not normalised to any degree this function exists to attempt to correct the grouping columns, after standardisation [https://github.com/SirSharpest/CT\\_Analysing\\_Library/issues/2](https://github.com/SirSharpest/CT_Analysing_Library/issues/2) this shouldn't be needed anymore, but kept for legacy issues that could arise!

### 2.1.8 `join_spikesbyrachis`

So important part of this function is that we accept that the data is what it is that is to say: rtop, rbot and Z are all orientated in the proper direction

It's main purpose is to join split spikes by rachis nodes identified in the analysis process

@param grain<sub>df</sub> is the grain dataframe to take on-board

### 2.1.9 `remove_percentile`

This function is targeted at removing a percentile of a dataframe it uses a column to decide which to measure against. By default this will remove everything above the percentile value

@param df is the dataframe to manipulate @param column is the attribute column to base the removal of @param target<sub>percent</sub> is the percentage to aim for @param bool<sub>below</sub> is a default param which if set to True will remove values below rather than above percentage

### 2.1.10 `get_spikeinfo`

This function should do something akin to adding additional information to the data frame

@note there is some confusion in the NPPC about whether to use folder name or file name as the unique id when this is made into end-user software, a toggle should be added to allow this

### 2.1.11 `lookup`

### 2.1.12 `gather_data`

### 2.1.13 `aggregate_spikeaverages`

This will aggregate features (specified by attributes) into their medians on a per-spike basis.

Makes direct changes to the dataframe (self.df)

@param attributes list of features to average

#### **2.1.14   make<sub>plot</sub>**

Returns false if plot could not be created for invalid parameters

### **3   graphing.py**

#### **3.1   Error**

Base class for other exceptions

#### **3.2   InvalidPlot**

Except to trigger when a graph is given wrong args

##### **3.2.1   percentile<sub>grid</sub>**

##### **3.2.2   qq<sub>grid</sub>**

##### **3.2.3   plot<sub>boxplots</sub>**

##### **3.2.4   plot<sub>histogram</sub>**

Simple histogram function

returns a plot axes

##### **3.2.5   check<sub>varargs</sub>**

Helper function to fix bad arguments before they get used in evaluations

### **4   scratch.py**

### **5   statistical<sub>tests</sub>.py**

#### **5.0.1   qqplot**

#### **5.0.2   test<sub>normality</sub>**

<https://stackoverflow.com/a/12839537>

Null Hypothesis is that X came from a normal distribution

which means: If the p-val is very small, it means it is unlikely that the data came from a normal distribution

As for chi-square: <https://biology.stackexchange.com/questions/13486/deciding-between-chi-square-and-t-test>