**Introduction**

Here is a study on the nocturnal flight-calling behavior and vulnerability to artificial light in migratory birds: <https://royalsocietypublishing.org/doi/10.1098/rspb.2019.0364#d3e550>

The dataset used in the study provides 40 years of data on bird collision data collected in Chicago, Illinois. We will be using a slightly modified version of the dataset to do some data cleaning and processing exercises.

Here is the original dataset (and references) used in the study: <https://github.com/rfordatascience/tidytuesday/tree/47567cb80846739c8543d158c1f3ff226c7e5a5f/data/2019/2019-04-30/raw>

**Practice Dataset**

We will use a slightly modified version of the dataset in this exercise. The dataset has 3 JSON files that you can find in the separate zip folder.

1. **light\_levels.json**This dataset provides the light levels (integer value) as measured on various dates.
2. **flight\_call.json**This dataset provides the information of whether or not a particular bird family uses *flight call* to communicate with each other. Per the paper, this is an important indicator that affects bird-building collisions.
3. **chicago\_collision\_data.json**This dataset has information on the dates and the locality where the light measurements were taken.

As is usual, real world datasets are messy and the above datasets are no different. In particular, the `**flight\_call.json**` dataset has incorrect column names. In order to link the datasets the the following renaming is necessary:

Species -> Genus

Family -> Species

Collisions -> Family

Call -> Flight Call

For instance, here is the wikipedia page on *Zonotrichia*, a *Genus* of *American Sparrows*

<https://en.wikipedia.org/wiki/Zonotrichia>

Looking at the wikipedia page and the scientific classification and species for *Zonotrichia*, and comparing with the provided data, it is clear that the column names are incorrect in the flight\_call data.

**Problem Statement**

We would like to create a summary tabular dataset that can be used to analyze the bird collision data.

Your task is to create a data processing script in *Python* that takes the given *JSON* files as input, cleans the data and then combines them to produce a summary dataset.

**Deliverables**

1. A well commented python script that takes two arguments from user:

* path to input json files
* path to write the generated summary output file

1. A READMEfile that provides instructions on installing required dependencies and running the script.
2. A brief writeup of the processing done to generate the summary dataset and the choice of file format chosen for the processed summary dataset. Pls. include a short justification for any data transformations. For instance, see the justification for choosing the column name mapping in the dataset description section.
3. [optional] Any interesting observations/analysis of the data and/or critique of the data.

**Guidelines**

Please note that, in addition to correctness, we will also be evaluating the deliverables for style, modularity, adherence to good coding practices and effective use of python and the tools.

We expect you to use **Pandas** library to solve this problem. Feel free to use additional libraries but ensure that they are specified in the installation instructions in the README.