

Data Visualization Concepts



BINF4234

Dr. Alexandra Diehl

Exercise and Homework Completion Requirements

1. Exercises and reading assignments are **mandatory** and they must be completed successfully to finish the class and get a sufficient passing final grade.
2. Exercises are graded coarsely into categories **incomplete** or **complete**.
 - An **incomplete** is given to failed submissions and partial solutions, but no bonus points are awarded.
 - A **complete** indicates that the exercise is sufficiently good to receive the bonus points.
3. A **minimum of 5 points** from the exercises 1, 2, and 4 must be achieved to pass the module, whereas exercises 3 and 5 are optional and can offer you bonus points as described in Point 4.
 - Failure to achieve this minimum will result in a failing grade for the entire module.
 - Hence at least the first two, or the fourth exercise has to be fully solved.
4. The five exercises give rise to the following point distribution: 2 – 3 – 2 (**optional bonus points**) – 5 – 3 (**optional bonus points**).
 - Only the bonus points can and will be added directly to the final grade.
5. Do not copy assignments, tools to detect copying and plagiarism will be used.
 - The exercise results are an integral part of the final course grade and therefore the handed in attempts and solutions to the exercises **must be your personal work**.

Submission Rules

- Submitted code must compile and run without errors using the indicated Python environment, using the included libraries, packages and frameworks. If additional libraries/packages are needed, please specify in your 'readme.txt' file.
- The whole project source code must be zipped and submitted before the given deadline, including the output results (saved in .html file or a screenshot picture).
- Submit your .zip archive named `dvc_ex1_MATRIKELNUMBER.zip` (e.g. `dvc_ex1_01234567.zip`) through the OLAT course page.
- **Deadline is Wednesday, 9 October 2019 at 23:59h**

Exercise 1

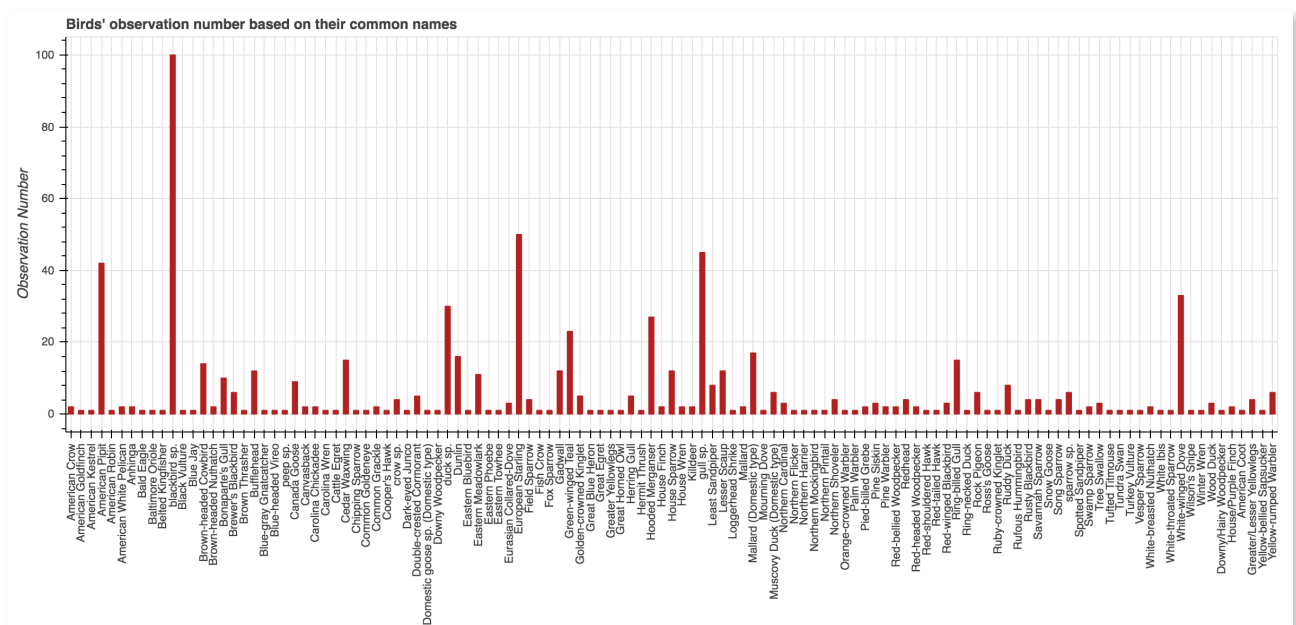
The aim of this exercise is to get familiarized with the basic data manipulation techniques using Python, including data cleaning and basic data visualization. In this exercise, you will need to first download the eBird raw dataset (there are several files, the one with all the data we will use is named “**ebd_US-AL-101_201801_201801_relMay-2018.txt**”) from [this link](#) after signing in and agreeing with the terms and conditions, then complete the following tasks:

Task1: Data preprocessing.

T1.1: Data reading: read the data using pandas and store it into a data frame.

T1.2: Data cleaning: follow the task descriptions in the code skeleton, learn and apply different functionalities provided by pandas ([documentation](#)).

Task2: Data visualization. Specifically, plot the histogram of **observation count** based on birds' **scientific name** using bokeh. Before starting the task, read the reference tutorial from the skeleton which is sufficient to help you finish this task. The final visualization result should look **similar** as the picture shown below:



Feel free to learn more about the eBird project from it's website: <https://ebird.org/about>

The deliverables of this exercise will be a clean version of your code with proper comments, any additional files necessary for executing it, a “readme.txt” file for your comments or remarks (if necessary), as well as an export of the final output result in .html or .jpg format. The absence of any required deliverable files will automatically lead to a **FAIL**.