Data Visualization and Analysis

BINF4245

Prof. Dr. Renato Pajarola

Exercise and Homework Completion Requirements

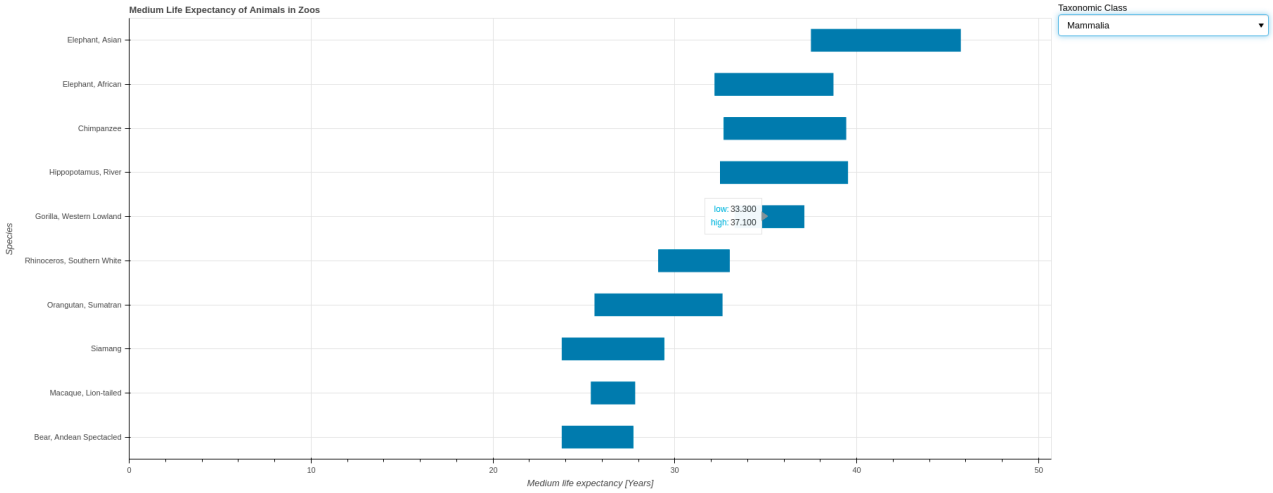
- Exercises and reading assignments are **mandatory** and they must be completed successfully to finish the course and get a sufficient passing final grade.
- Exercises are graded coarsely into categories **pass** or **fail**.
  - *A **fail** is given to failed submissions and partial solutions, and no points are awarded.*
  - *A **pass** indicates that the exercise is sufficiently good to receive the corresponding points.*
  - *Late submissions (up to one day) will result in "-1" point.*
- The four exercises are allocated to the following point distribution: 2 – 3 – 5 – 5
  - A **minimum of 7 points** must be achieved to pass the module.
  - Thus, at least two exercises must be solved correctly, including at least one from the advanced ones.
  - *Failure to achieve this minimum will result in a failing grade for the entire module*
- We award **bonus points** for students who have collected more than 8 points from all the exercises.
  - Thus, **7 points** from the exercises is required, **8 points** is still normal passing, **9 and above** would give 1 or more extra bonus points.
  - Only the bonus points can and will be added directly to the final grade.
- 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 run without errors using the indicated Python environment, included libraries, packages and frameworks. If additional libraries/packages are needed, please specify these in a *readme.txt* file together with your submission.
- The whole project source code must be zipped and submitted before the given deadline, including exactly the files indicated in the exercise.
- Submit your .zip archive named *dva_ex1_Firstname_Lastname_MATRIKELNUMBER.zip* (e.g. dva_ex1_Hans_Muster_01234567.zip) through the OLAT course page.
- **Deadline is Monday, 19 October 2020 at 23:59h**

Exercise 1

The aim of this exercise is to get familiarized with pandas, bokeh and some basic visualization methods for the purpose of data analysis. Your task will be to read data from a file, process and filter it and present it in a dashboard with different types of visualization methods. The arrangement of the widgets in your dashboard should look like the figure below. Exact tasks and hints how to solve them are provided in the code skeleton.



Task 1: Preprocessing:

The input file includes, among others, the names, taxon classes and median life expectancies of different animal species. In a first step, you must split and clean the data and remove outliers. You have to split the data into three different sets based on the taxon classes ‘Mammalia’, ‘Aves’ and ‘Reptilia’ and remove all entries that have either deficient male or female data or both.

Task 2: Visualization

We want to plot a vertical bar chart of the animals with the highest median life expectancy for each taxon class. Which taxon class is displayed can be chosen by the user with a select widget to the right of the plot. The limits of the horizontal bars are determined by ‘Overall ci – lower’ and ‘ Overall ci – upper’. Additionally, there is a hover tool showing the limits of the bars. Detailed information and guidance is provided in the skeleton Python file.

Important: All deliverables of this exercise must be submitted before the deadline. The absence of any required files will automatically lead to a **FAIL**.

Submission:

- clean version of your code file with proper comments

- .csv file with the data needed for the execution

- readme.txt – Use this file for your comments or remarks (if necessary). If you used any additional libraries that are not imported in the skeleton explain why and for what in this file. This file may be empty if you have no comments and only used the provided libraries.

- An export of the final dashboard in .pdf or .jpg format. (A screenshot is also accepted.)

- Put all required files into a .zip archive using the naming scheme detailed on the first page of this document. Put the files directly into the archive and do not package a folder.