

Due: 11:59pm, Feb. 2, 2024

Learning Objectives

The goal of this assignment is to become acquainted with Matplotlib customization.

Data

Download the `sleep.csv` data set from Brightspace.

This data contains information on the daily number of hours of sleep that are recommended, appropriate, and not recommended by various age groups. This data was compiled by the National Sleep Foundation in the USA from a panel comprised of members and sleep experts from various American medical associations.

You can read more about sleep recommendations and the data in the following articles:

- Journey into sleep: An illustrated tour of why scientists are finding that sleep matters more than we know, *Reuters*, <https://www.reuters.com/graphics/HEALTH-SLEEP/mopakyjmnpa/>
- National Sleep Foundation's sleep time duration recommendations: methodology and results summary, *Sleep Health*, <https://doi.org/10.1016/j.sleh.2014.12.010>

Instructions

Using the provided data set, create a Jupyter notebook to answer the following questions.

The only libraries you are allowed to import to complete this assignment are: `pandas`, `Matplotlib`, and `math` or `NumPy`.

There are many levels of customization that can be accomplished with `Matplotlib`. You are free to use all parts of the `Matplotlib` API. Two hints for this assignment:

- One helpful method is `text()`, which will write a string of text in a specified location.
- The borders around the frame of the plot are called 'spines'.
- The colours needed for this assignment are '#051b2c', '#4789be', '#0c3759', '#f3e2c0', '#6ebeec', and '#95cef2'.

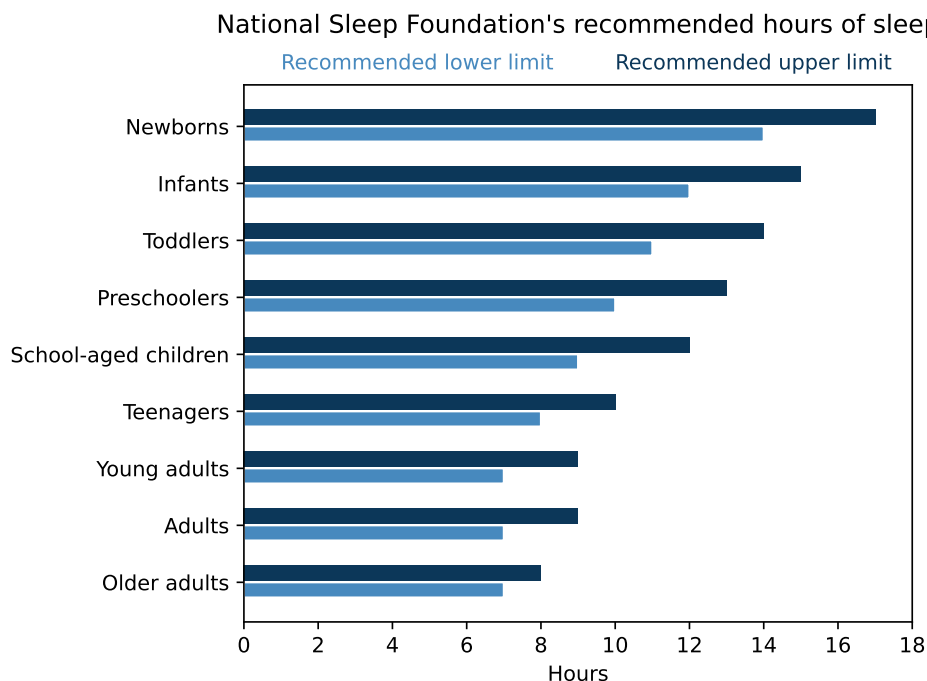
Question 1 – Grouped Bar Plot: (25 pts)

A grouped bar plot is a bar plot where each bar is separated into multiple sub-categories. This allows for comparison of values within a single item, as well as across multiple items. As the number of sub-categories increases, it can become difficult to represent them without creating too much visual complexity.

The plot below shows the recommended lower and upper limit on daily hours of sleep per age category. People tend to require fewer hours of sleep as they age. There is usually a 2–3 hour variance on the number of recommended hours of sleep across all age categories.

Re-create the plot below using Matplotlib. All elements should be replicated as shown. In particular, ensure that,

- the bars are coloured appropriately,
- there is a space between the two bars within a group,
- the text below the title is present and coloured correctly,
- the axes limits are correct,
- labelling for the title and x-axis are present, and
- the category labels along the y-axis are ordered and positioned correctly.

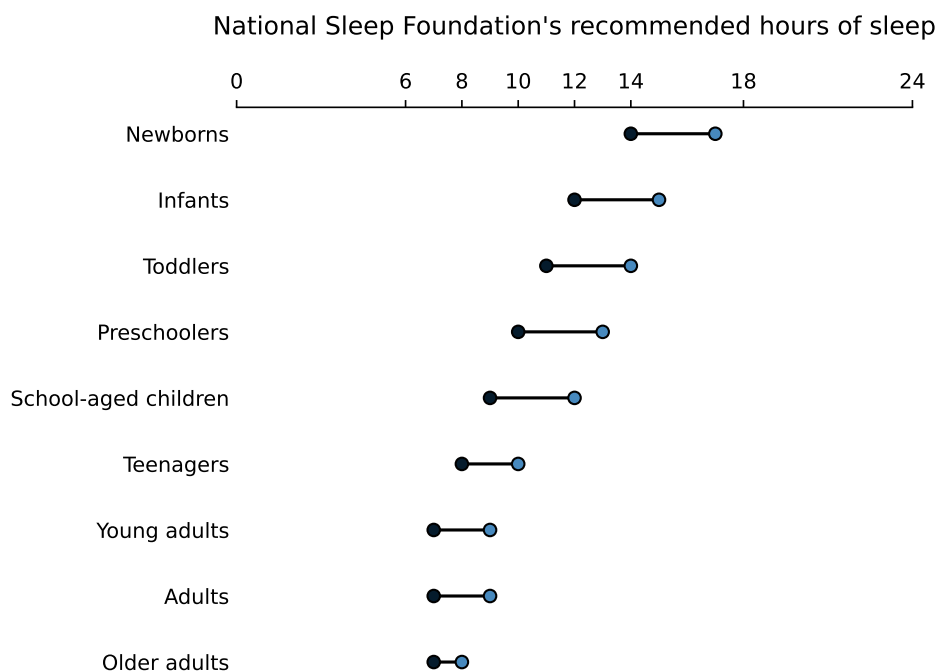


Question 2 – Connected Dot Plot: (35 pts)

A connected dot plot shows quantitative values for different categories of items, with each item containing a secondary categorical breakdown. Each dot represents a data point, and the dots are connected by a line. This helps enable comparison of values within a single item by displaying the difference (delta) between data points, in addition to comparison of values across multiple items. A connected dot plot is useful to represent, for example, high and low values or targetted and realized values, though more than two sub-categories is possible as well.

Re-create the plot below using Matplotlib. All elements should be replicated as shown. In particular, ensure that,

- the dots are coloured correctly,
- the category labels along the y-axis are ordered and positioned correctly.
- the x-axis tick marks and labels are correct,
- the x-axis tick marks and labels are above the plot (instead of below),
- the spines along the left, bottom and right side are removed, and
- the title is present.



6934 Students Only – Question 3 - Marimekko Chart: (40 pts)

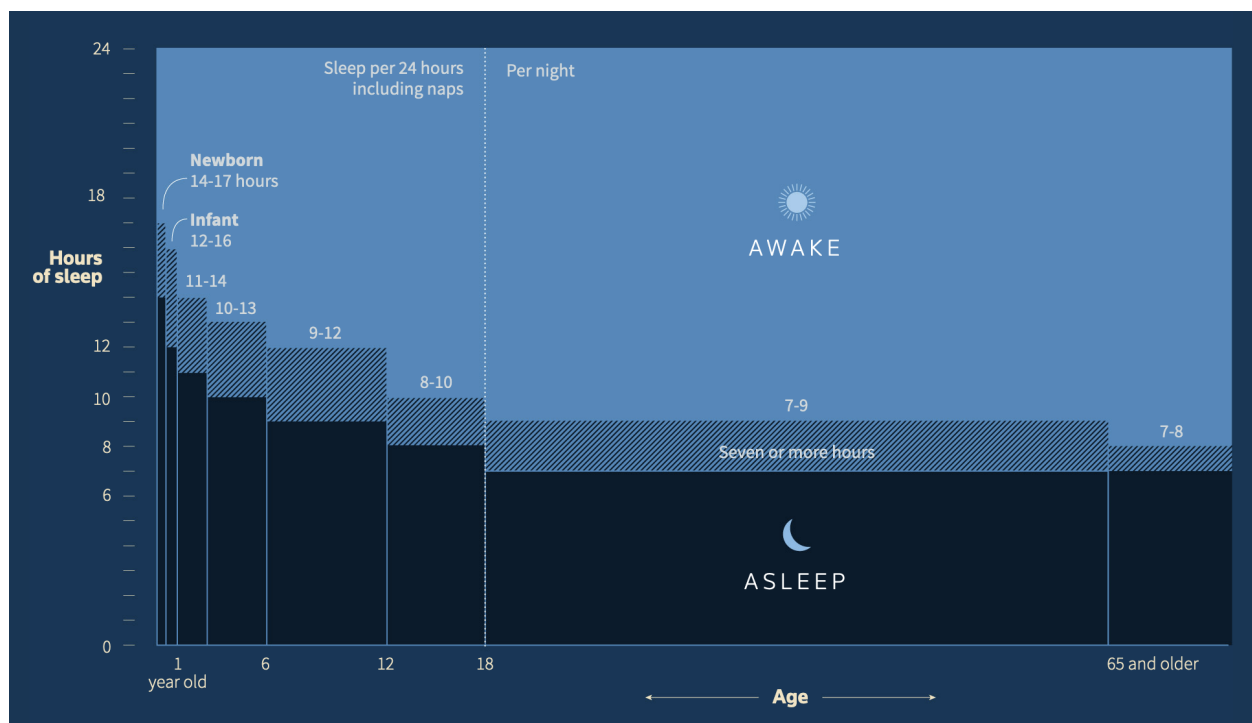
A marimekko chart is a bar plot where the width of the bars encode information about the data set, making for a two-dimensional comparison along both the x- and y-axes. Each bar can be

subdivided into sub-categories. This type of plot is used for a parts-to-whole comparison.

Re-create the Marimekko chart displayed below, which comes from the Reuters article. The x-axis represents age, with the y-axis the number of hours of sleep. The region with the diagonal hatching is the recommended number of hours of sleep for that age category.

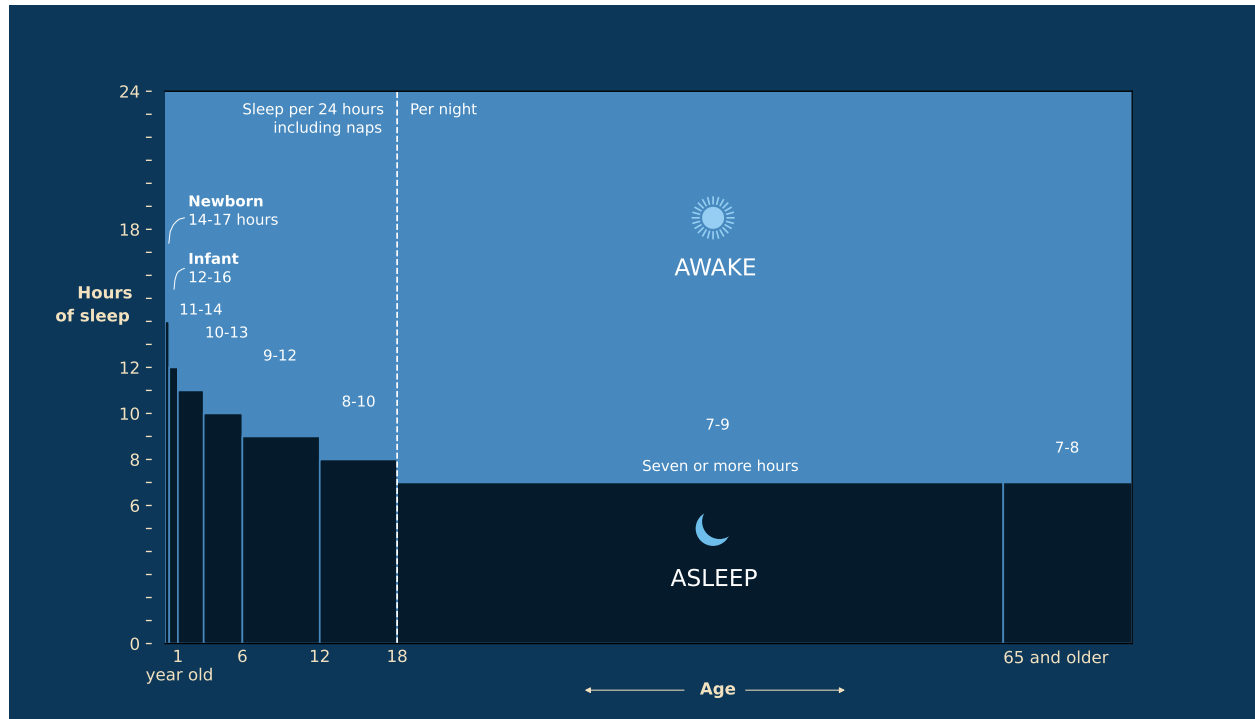
Elements that should be included in your final representation include:

- the width of each bar is sized according to the age,
- there is a small gap between adjacent bars,
- the diagonal hatching for the recommended number of hours is present with similar density of lines,
- the x/y-axes labels and tick mark locations / labels are correct,
- the background colour outside the main frame of the plot is present,
- all colours of all elements should be accurate (see assignment instructions for list of colours),
- there is a space between the y-tick marks and the frame of the plot,
- the vertical white, dashed line is present to denote 18 years of age,
- all textual elements inside the plot are present and in the correct locations,
- the moon and sun icons are represented, and
- the arrows on either side of the “Age” label are present.



For reference, here is what I was able to achieve in Matplotlib using only what was taught in the lectures, along with `text()` and the methods to set colours and remove spines.

Note that the original image does not evenly space the tick marks along the x-axis, whereas I have used consistent spacing between tick values. See the difference in bar widths between the 6–12 and 12–18 age groups, for example. The original image uses larger width bars for younger age brackets so that they are easier to visually distinguish. I have intentionally not followed their spacing in my representation (which means you do not have to either).



Submission

Submit your Jupyter notebook (.ipynb) through Brightspace.

Late submissions will be subject to a 10% penalty for each hour past the deadline.

Attribution

Submissions should include an attribution section indicating any sources of material, ideas or contribution of others to the submission.

Submissions must represent your independent work.

You are encouraged to use any resources to help with your solution, but your solution must represent independent work. If your submitted work includes unacknowledged collaboration, code materials, ideas or other elements that are not your original work, it may be considered plagiarism or some

other form of cheating under MUN general regulations 6.12.4.2 (4.12.4.2 for graduate students) and academic penalties will be applied accordingly.

Avoid academic penalties by properly attributing any contribution to your submission by others, including internet sources and classmates. This will also help distinguish what elements of the submission are original. You may not receive full credit if your original elements are insufficient, but you can avoid penalties for plagiarism or copying if you acknowledge your sources.

Github

I encourage you to store and version your work on GitHub. It is good practice to do so as everyone uses git in the real world.

However, **it is a requirement that git repositories containing assignment material be private.** University regulations (undergraduate 6.12.4.2 and graduate 4.12.4.2) consider it cheating if you allow your work to be copied. There will be zero tolerance for this.