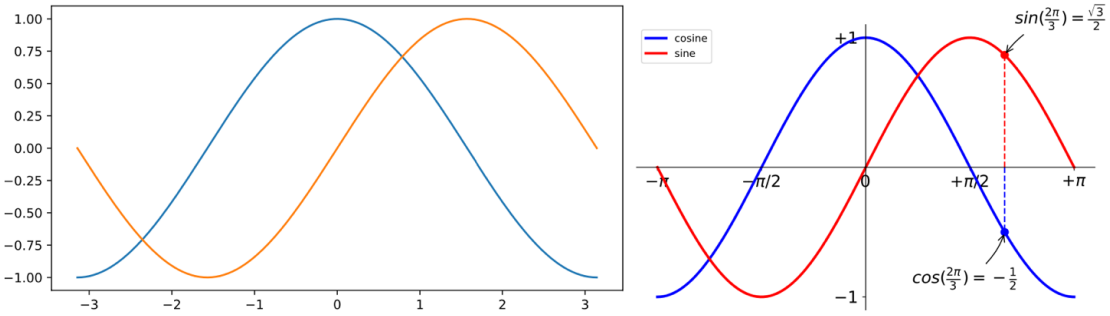
Data Science II - Data Visualization

Lab 3 - Summer Semester 2024
https://moodle.haw-landshut.de/course/view.php?id=10969

1. Visualizing functions with matplotlib

Study the examples at `https://matplotlib.org/stable/gallery/index.html` and reproduce the plot on the right



using matplotlib. You don't need to include the two annotations $\sin(\frac{2\pi}{3}) = \frac{\sqrt{3}}{2}$ and $\cos(\frac{2\pi}{3}) = -\frac{1}{2}$.

Maybe useful: There are additional matplotlib exercises at `https://github.com/matplotlib/AnatomyOfMatplotlib` if you are interested in better understanding the `matplotlib` library.

2. Visualizing Amounts

The datasets

- *2020 median US annual household income vs age group*
- *poverty rates by age and sex 2020*
- *female and male passengers on the Titanic*

are provided to you in our Moodle course. Use the data sets and recreate the bar plots, grouped and stacked bar plots from the lecture on "Visualizing Amounts and Distributions" using the `matplotlib` library.

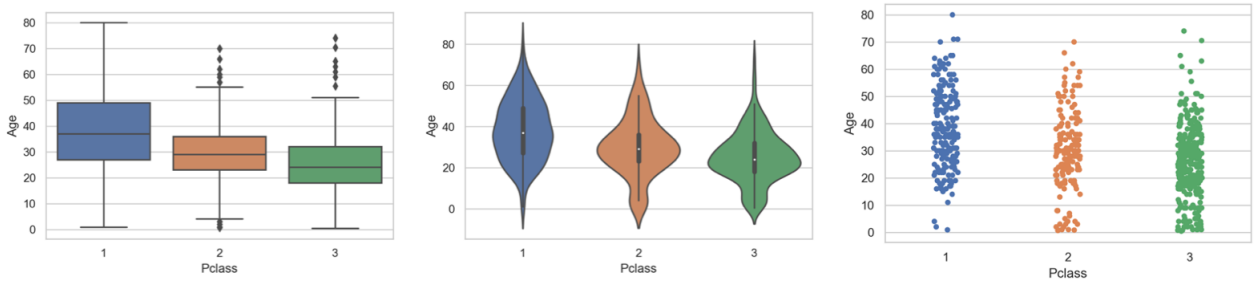
3. Visualizing Distributions

Use column "Age" from the titanic data set (see ./data/titanic.csv), divide the data into bins of width 5 (years) and count the number of points that fall in each bin. You may use `pandas` or `numpy` for counting the frequencies (e.g. read the documentation of `np.where(...)` or `np.histogram(...)`.

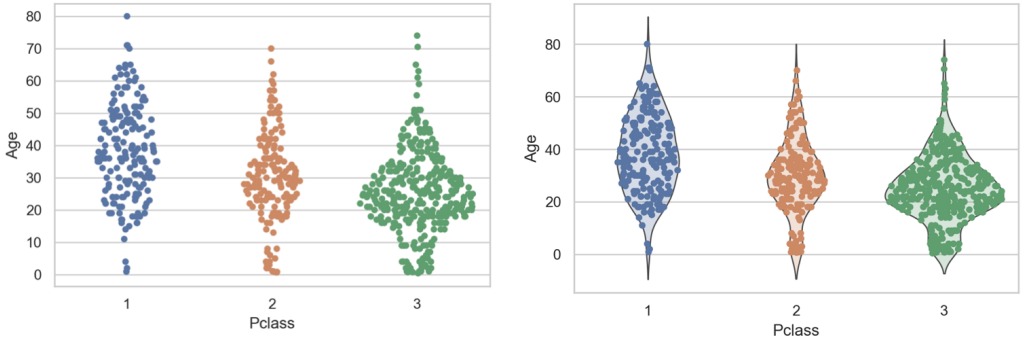
- draw the corresponding histogram with pen and paper
- confirm the correctness of your drawing by using the `histogram` function of `matplotlib.pyplot`; change the bin width to 1 year and then to 10 years and compare the histograms
- How does your histogram change if you normalize it?
- Can you verify (using `numpy`) that the sum of heights of the rectangles of the normalized histogram equals 1?

4. Visualizing Multiple Distributions Simultaneously

(a) Read the documentation of seaborn.boxplot, seaborn.violinplot and seaborn.stripplot and generate the corresponding visualizations for the Titanic data set. Use "Pclass" as grouping variable (x-axis) and "Age" as response variable (y-axis).



(b) Check out the Github repository `https://github.com/mparker2/seaborn_sinaplot` for a sina plot function and generate a sina plot according to the specification in (a).



5. Nightingale Charts

Use the data from ./data/war.csv to recreate the original Nightingale Chart created by Florence Nightingale. You can use e.g. `https://plotly.com/python/wind-rose-charts/`.