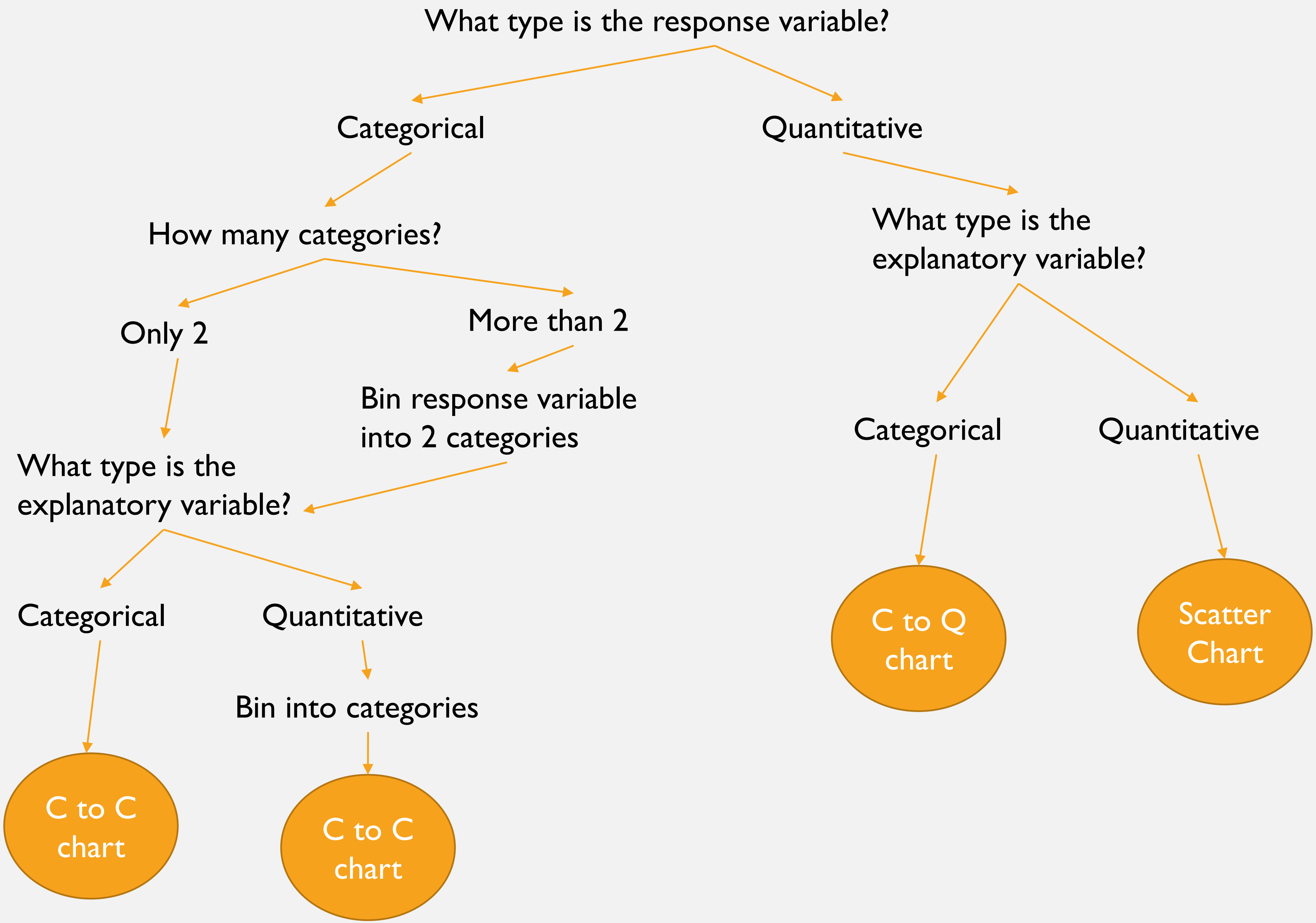
DATA VISUALISATION

DATA VISUALISATION

- Quantitative response variable
- Lab5



What type is the response variable?

Categorical

How many categories?

Only 2

What type is the explanatory variable?

Categorical

C to C chart

Quantitative

Bin into categories

C to C chart

More than 2

Bin response variable into 2 categories

Quantitative

What type is the explanatory variable?

Categorical

C to Q chart

Quantitative

Scatter Chart

DATA VISUALISATION

- The relationship between two quantitative variables can be investigated using a scatterplot.
- Python provides scatterplots through the use of the seaborn package.
- The regplot() function is used, the X axis and Y axis are named in the function call, along with the data frame name.
- By default the best fit line is included in a scatterplot, this is not always needed and can be suppressed.
- Again the labels can be set for the X and Y axis as well as the chart title.

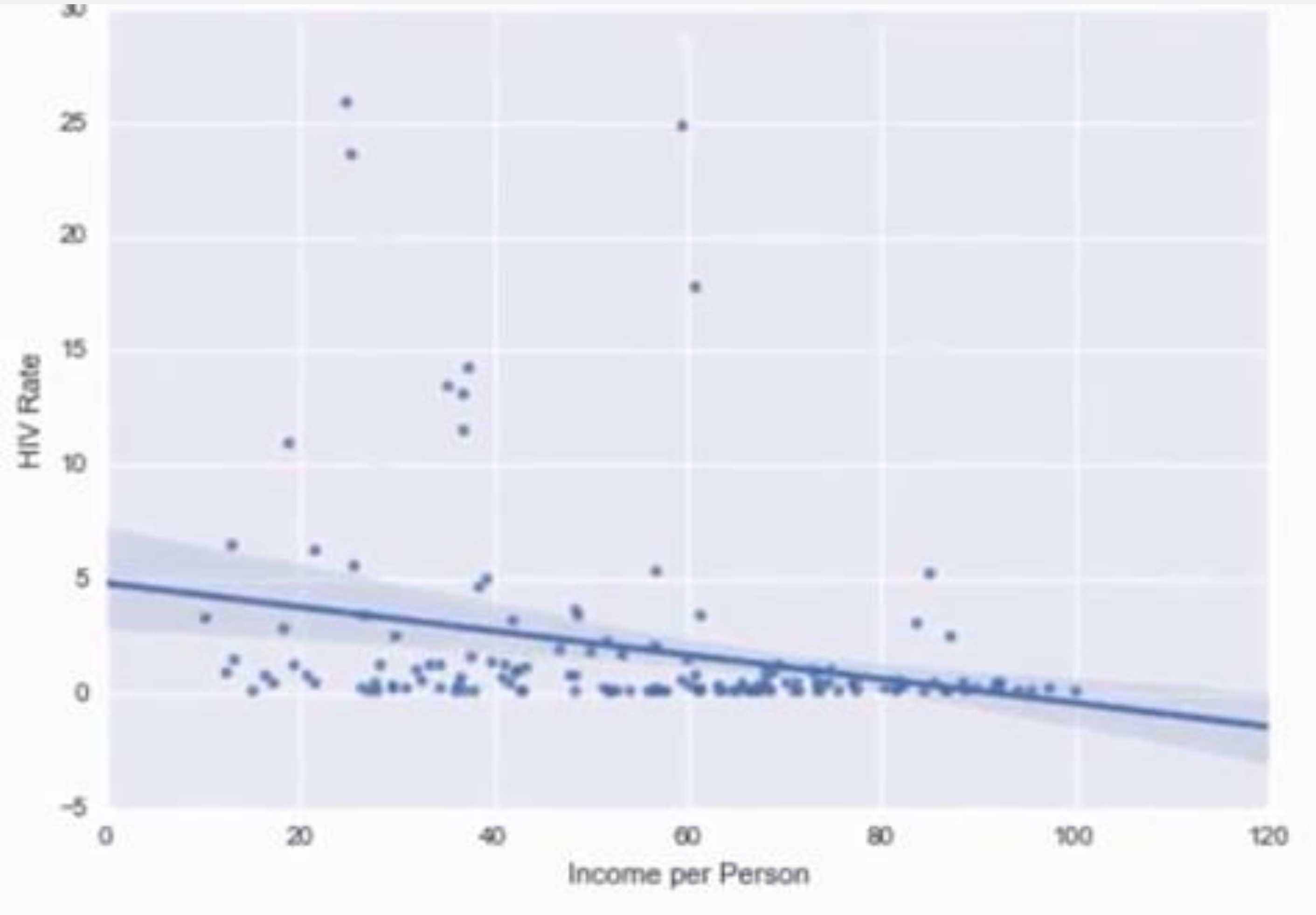
DATA VISUALISATION

- Positive relationships show as an increasing slope. (higher values in one variable indicate higher values in another variable)
- Negative relationships show as a decreasing slope. (higher values in one variable indicate lower values in another variable)
- The strength of a relationship is shown by the density of the dots. How closely the data points for the linear slope.
- Not all relationships can be classified as positive or negative.
- If you can not put a line through the dots then there may be no relationship.

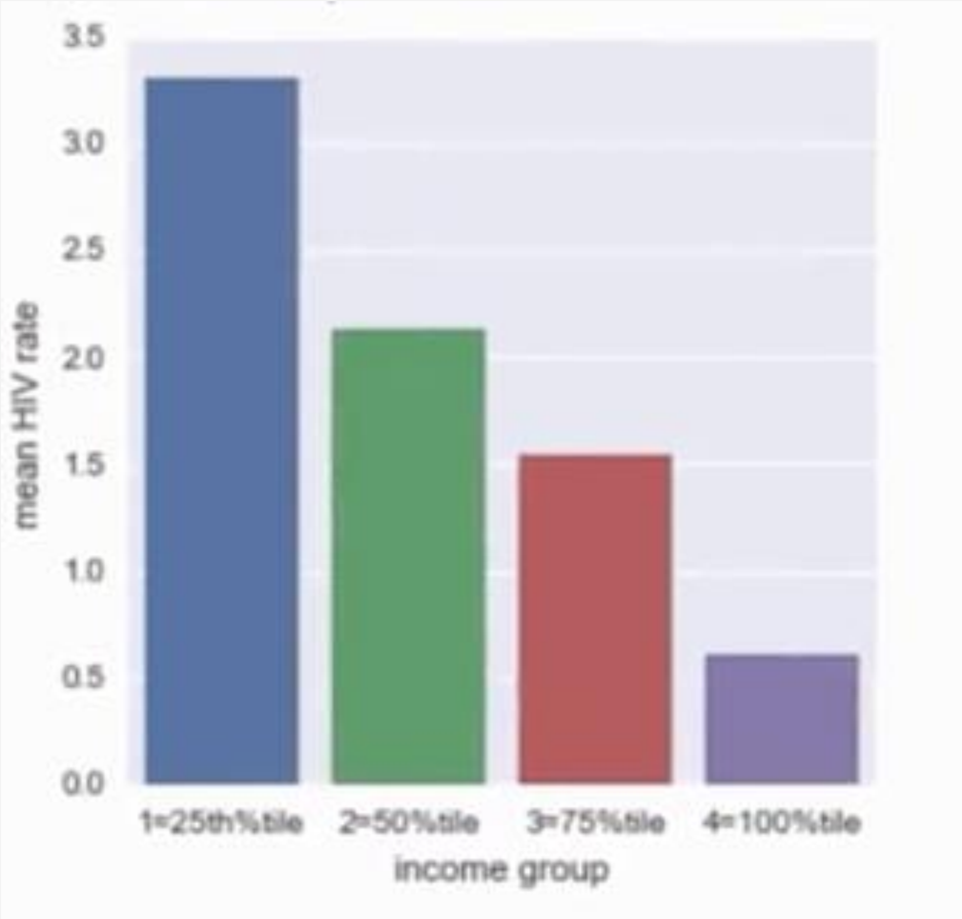
DATA VISUALISATION

- Sometimes a scatterplot is limited in its ability to allow us to evaluate a relationship visually.
- It may be necessary to change the explanatory variable to a categorical variable.
- For example plotting the income per person (quantitative) with the rate of HIV (quantitative) may not show an obvious relationship between the two variables.
- By grouping the income per person into multiple categories and plotting a bar chart may show more clearly the relationship.,

DATA VISUALISATION



DATA VISUALISATION

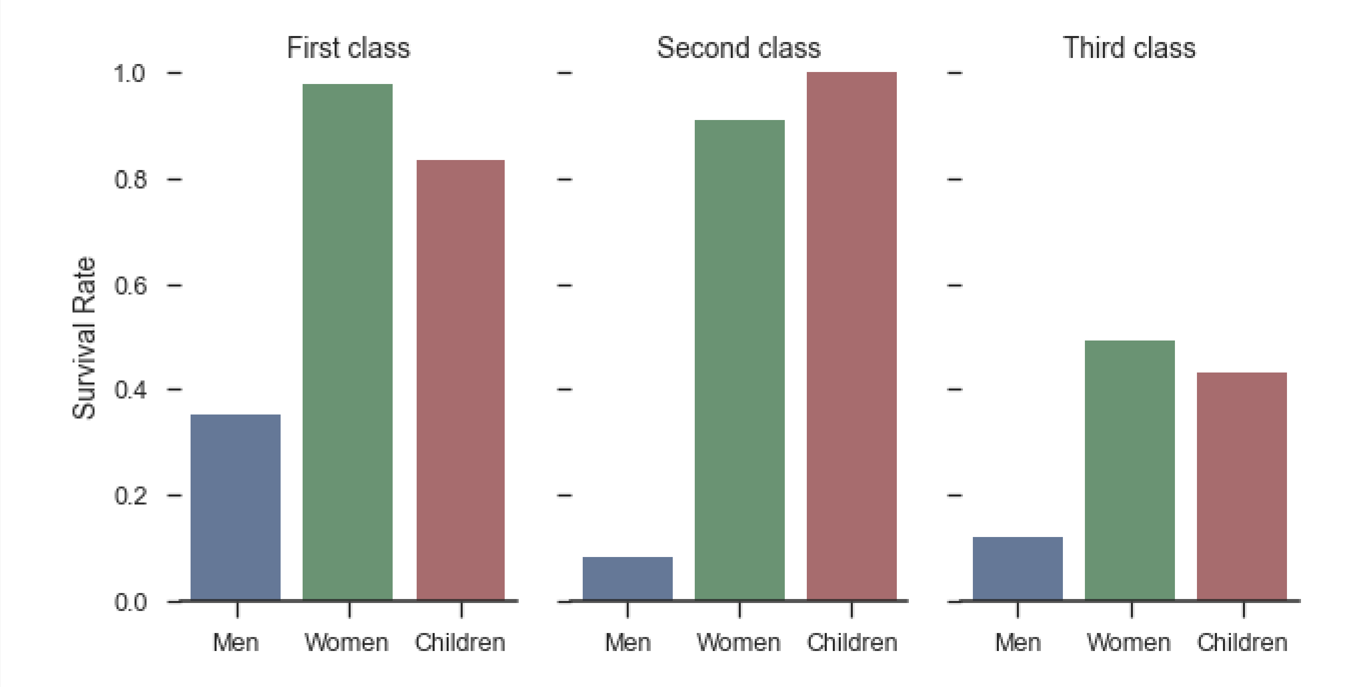


DATA VISUALISATION

- Seaborn has standard settings that you can adjust such as palette, legend, and kind.
- You have seen the use of kind="bar" for the catplot.
- legend can be set = False
- palette can be set = "muted"
- Within each function such as countplot() or catplot etc there are parameters you can pass.
- In the labs we concentrate on two variables, one for the X axis and one for the Y axis

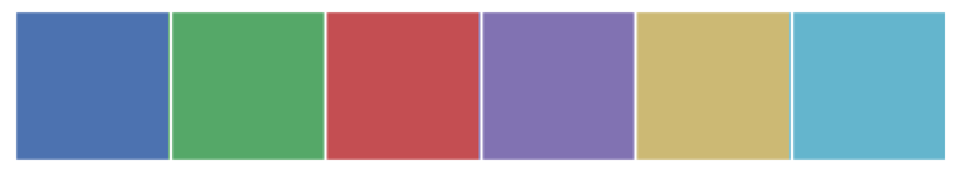
DATA VISUALISATION

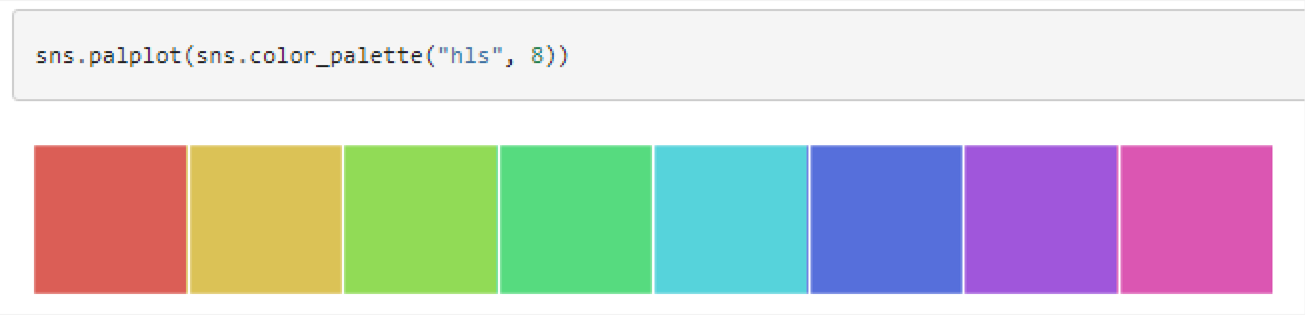
- In some charts you may have a third dimension you wish to plot.



DATA VISUALISATION

- colour palettes available are deep, muted, pastel, bright, dark and colorblind.
- Six default colours are set when seaborn is imported, if you need more than six colours on your charts the easiest is to use hls colour space which creates hues of colours between the default 6.

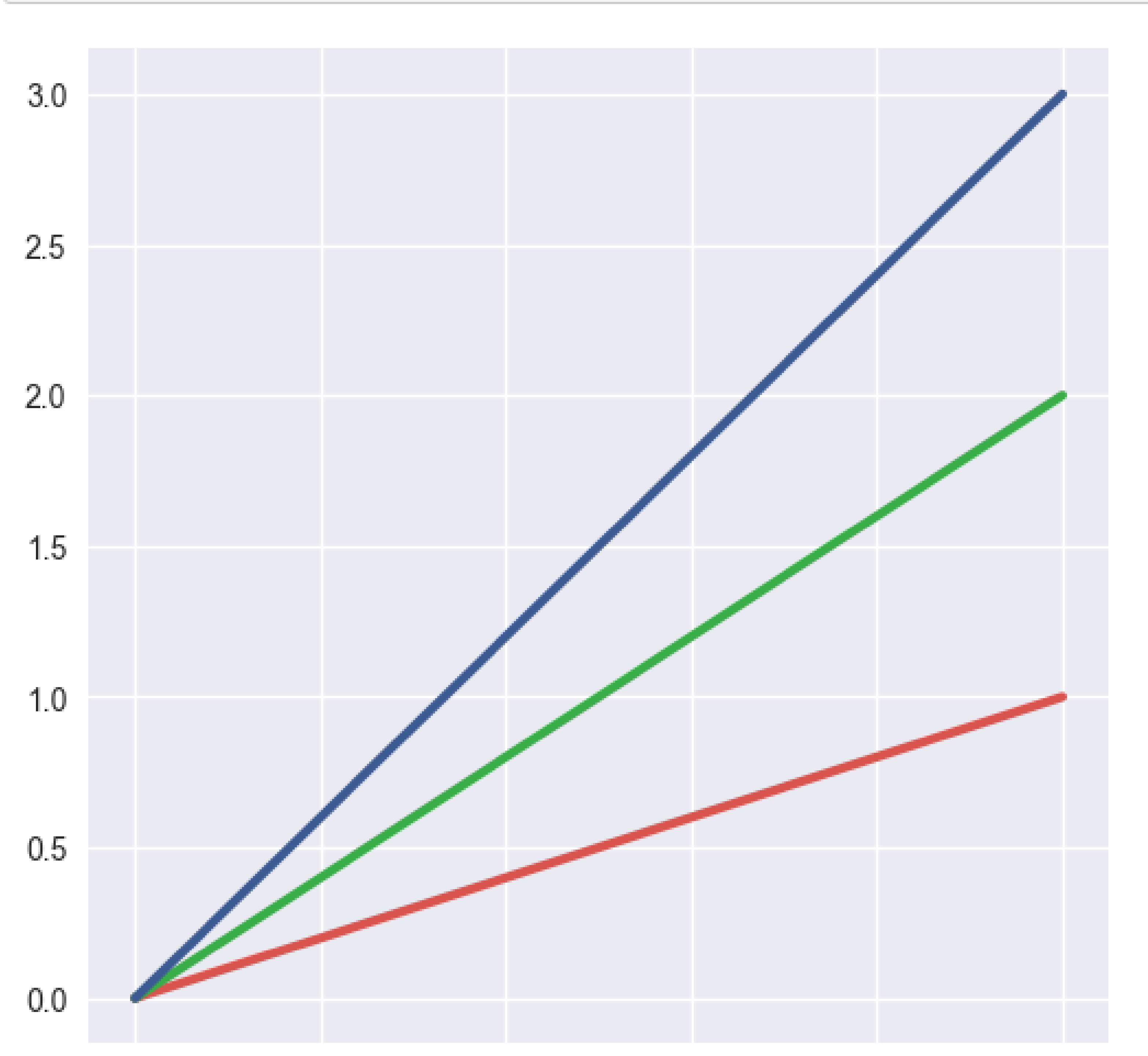




```
sns.palplot(sns.color_palette("hls", 8))
```

DATA VISUALISATION

- It is also possible to name some colours, there are 954 named colours to pick from.
- Use the xkcd_rgb function in seaborn to set the colours of your choice.



DATA VISUALISATION

A good place to start looking at the charts you can create using seaborn is the following url:

https://seaborn.pydata.org/examples/index.html

Quick ref API guide:

https://seaborn.pydata.org/api.html