Day 2 Morning: Data Visualization

Gabriella Gerlach

December 10, 2020

1 Univariate Plots

Associated screen cast: link Load the tips dataset (tips.csv)

1. Plot a density plot of the total bill

2. Plot a frequency (bar) plot of which day people came to dinner

3. Plot a frequency (bar) plot of the size of the parties

2 Bivariate Plots

Associated screen cast: link Load the tips dataset (tips.csv)

1. Plot a scatter plot of tips vs total bill

2. Plot a box plot of the total bill by day of the week

3. Compute the percentage tipped and plot it against family size

3 Adding a 3rd Variable with Color

Associated screen cast: link Load the tips dataset (tips.csv)

1. Plot a scatter plot of tip against total bill and color the points by day of the week

2. Plot a scatter plot of tip against total bill and color the points by size of the party

4 Facets

Associated screen cast: link Load the tips dataset (tips.csv)

1. Plot a barplot (categorical vs continuous) of percentage tip against time of meal and facet it by the day of the week

2. Plot a frequency barplot (countplot) of party size and facet by the day of the week as columns and time of meal as rows

5 Pair Plots

Associated screen cast: link

1. load the diamonds dataset

2. Create and save a reduced dataset using the code:`diamonds.sample(n=1000, random_state=3)`

3. Use this dataset to render a pairs plot with scatter plots above the diagonal, contour plots (kdeplot) below the diagonal and distplot on the diagonal.

4. hint: we have not gone over a distplot, so use the documentation or internet to find out what it is.

6 Matplotlib Customization

Associated screen cast: link Take any one of the graphs you made in the earlier assignments (or load a data set you are working with outside of this workshop) and customize at least three things about it.

For example you could take a scatter plot and change the shape of the points, change the theme, and change the xticks.

The goal of this is to get some experience with customization because in bioinformatics generally the default plotting will not be sufficient for what you need to do. If there is something specific you want to do try looking at the matplotlib documentation or googling. No one remembers all the syntax so learning how to find what you are looking for is a very important skill.