Data Visualization

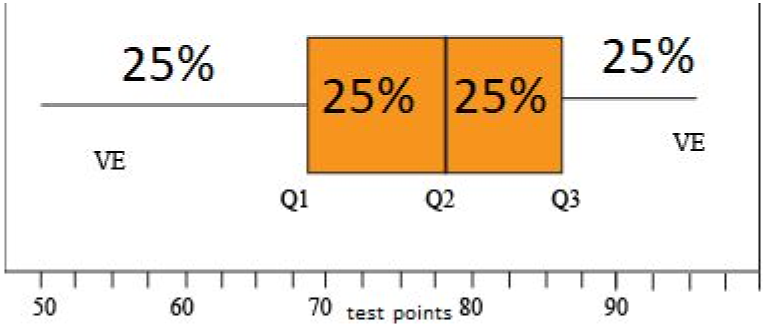
How to visualize data. Graphs, heatmap, pychart,...

How to call data.

Python Package named as **seaborn** (seaborn.pydata.org) statistical data visualization, based on **matplotlib**. You can install with pip install seaborn (documentation->Installing and getting started. It requires **numpy,scipy, matplotlib, pandas** package.

On gallery section of seaborn web we have a lot of different visualizations:

❑ Boxplots: source is sort the data and get the minimum, maximum, quartiles Q1(25%, lower median half), Q2(50% all data median) and Q3(75%, upper median half) and interquartile range (IQR= Q3 - Q1). Are a useful way to graph data divided into four quartiles, each with the same amount of values. The box diagram does not graph frequency or show individual statistics, but in them we can clearly see where half the data is. It is a good diagram to analyze the asymmetry in the data. Values below Q1–1.5 · IQR or higher than Q3 + 1.5 · IQR are considered outliers



❑ Heatmaps (anotated): color represents values, eg. pivot table. Color represents values of the data.

❑ Join Plot: Hexagonal bin plot with marginal distributions.

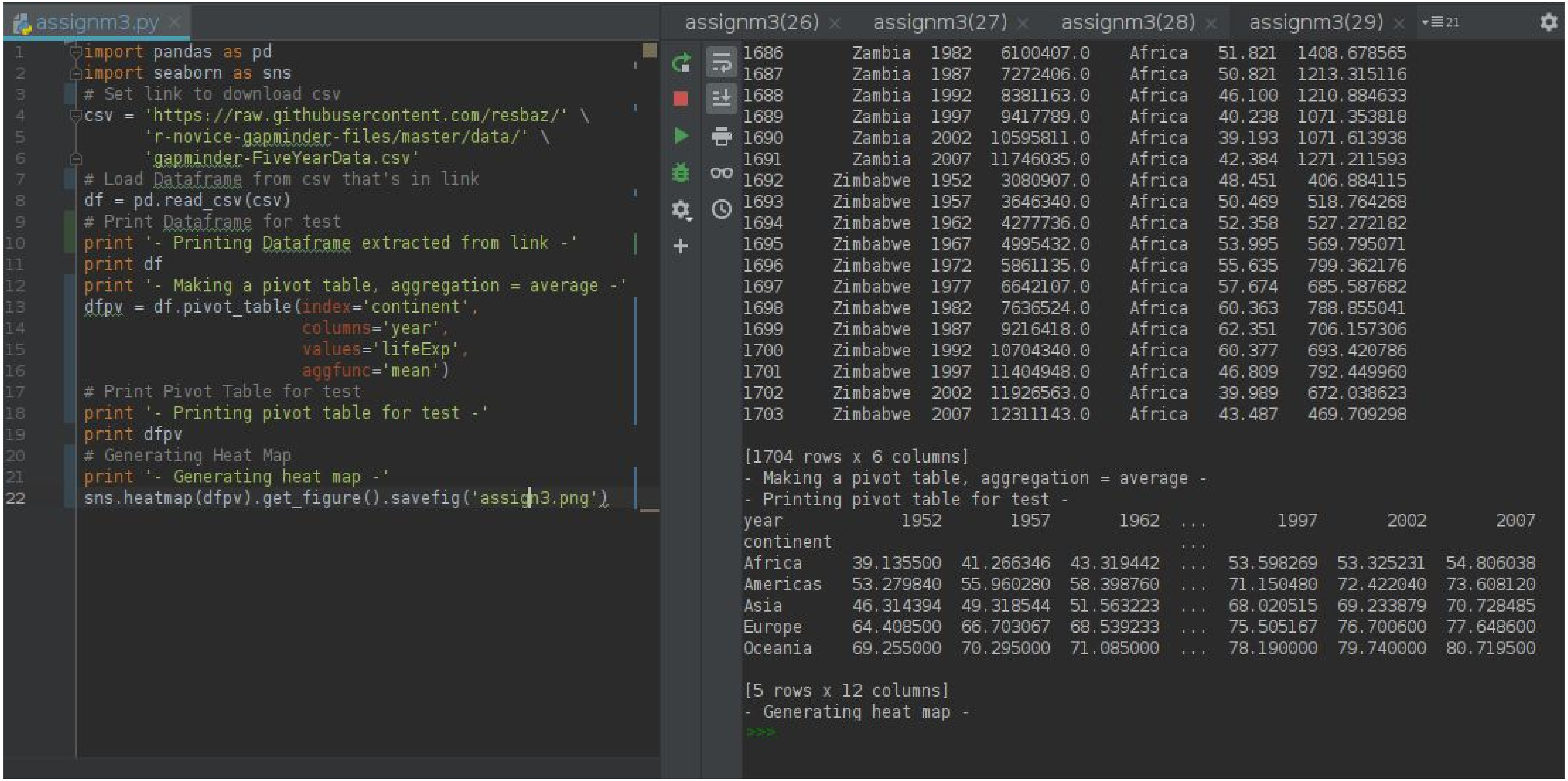
❑ Histograms: accurate representation of the distribution of numerical data, is an estimate of the probability distribution of a continuous variable. First step is to "bin" or "bucket" the range of values, that is, divide the entire range of values into a series of non overlapping intervals, that must be adjadcent, but not required of equal size; then count how many values fall into each interval. Histogram can be normalized to display relative frequencies; it shows the proportion of cases that fall into each of several categories, with the sum of the heights equaling 1. When bins are not equal witdh, vertical axis is not a frequency, is a frequency density.

❑ connell plots

❑

Assignment #3

It keeps simple: Download csv as dataframe, and make pivot table, after render this data with seaborn.



```python
import pandas as pd
import seaborn as sns
# Set link to download csv
csv = 'https://raw.githubusercontent.com/resbaz/' \
      'r-novice-gapminder-files/master/data/' \
      'gapminder-FiveYearData.csv'
# Load Dataframe from csv that's in link
df = pd.read_csv(csv)
# Print Dataframe for test
print '- Printing Dataframe extracted from link -'
print df
print '- Making a pivot table, aggregation = average -'
dfpv = df.pivot_table(index='continent',
                      columns='year',
                      values='lifeExp',
                      aggfunc='mean')
# Print Pivot Table for test
print '- Printing pivot table for test -'
print dfpv
# Generating Heat Map
print '- Generating heat map -'
sns.heatmap(dfpv).get_figure().savefig('assign3.png')
```

```
1686      Zambia  1982   6100407.0    Africa  51.821  1408.678565
1687      Zambia  1987   7272406.0    Africa  50.821  1213.315116
1688      Zambia  1992   8381163.0    Africa  46.100  1210.884633
1689      Zambia  1997   9417789.0    Africa  40.238  1071.353818
1690      Zambia  2002  10595811.0    Africa  39.193  1071.613938
1691      Zambia  2007  11746035.0    Africa  42.384  1271.211593
1692    Zimbabwe  1952   3080907.0    Africa  48.451   406.884115
1693    Zimbabwe  1957   3646340.0    Africa  50.469   518.764268
1694    Zimbabwe  1962   4277736.0    Africa  52.358   527.272182
1695    Zimbabwe  1967   4995432.0    Africa  53.995   569.795071
1696    Zimbabwe  1972   5861135.0    Africa  55.635   799.362176
1697    Zimbabwe  1977   6642107.0    Africa  57.674   685.587682
1698    Zimbabwe  1982   7636524.0    Africa  60.363   788.855041
1699    Zimbabwe  1987   9216418.0    Africa  62.351   706.157306
1700    Zimbabwe  1992  10704340.0    Africa  60.377   693.420786
1701    Zimbabwe  1997  11404948.0    Africa  46.809   792.449960
1702    Zimbabwe  2002  11926563.0    Africa  39.989   672.038623
1703    Zimbabwe  2007  12311143.0    Africa  43.487   469.709298

[1704 rows x 6 columns]
- Making a pivot table, aggregation = average -
- Printing pivot table for test -
year            1952       1957       1962  ...       1997       2002       2007
continent                                   ...
Africa     39.135500  41.266346  43.319442  ...  53.598269  53.325231  54.806038
Americas   53.279840  55.960280  58.398760  ...  71.150480  72.422040  73.608120
Asia       46.314394  49.318544  51.563223  ...  68.020515  69.233879  70.728485
Europe     64.408500  66.703067  68.539233  ...  75.505167  76.700600  77.648600
Oceania    69.255000  70.295000  71.085000  ...  78.190000  79.740000  80.719500

[5 rows x 12 columns]
- Generating heat map -
>>>
```

A pivot table is an aggregation method to summarize info from several dimensions as rows and columns.