DA2I
Data Visualization Methods

Leandro Nunes de Castro, Ph.D.

ldecastrosilva@fgcu.edu

DA2I

Data Visualization Methods

Summary

- Chapter Datasets
- Distributions
- Associations
- Amounts
- Proportions
- Evolution and Flow
- Geospatial

Data Visualization Types

Characteristic	Description
Distribution	Shows the spread of data
Associations	Reveals relationships between variables
Amounts	Represents quantities of different items
Proportions	Displays parts of a whole dataset
Evolution and Flow	Illustrates changes over time
Geospatial	Presents data on a map
AI-Enhanced Dataviz	Uses AI to improve visualizations
Design Principles for Dataviz	Guidelines for effective data presentation

Chapter Datasets

Iris Dataset of Fischer

Table 7: Iris dataset sample: first and last five objects in the dataset.

Object	Sepal length	Sepal width	Petal length	Petal width	Class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

Auto MPG Dataset

Table 8: Auto MPG dataset sample: first and last five objects in the dataset. Cyl: cylinders; Disp: displacement; HP: horsepower; WG: weight; Acc: acceleration; MY: model year; OG: origin.

Obj	mpg	Cyl	Disp	HP	weight	Acc	MY	OG	car name
0	18	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15	8	350	165	3693	11.5	70	1	buick skylark 320
2	18	8	318	150	3436	11.0	70	1	plymouth satellite
3	16	8	304	150	3433	12.0	70	1	amc rebel sst
4	17	8	302	140	3449	10.5	70	1	ford torino
..	...	...	...	...	...	...	...	...	...
393	27	4	140	86	2790	15.6	82	1	ford mustang gl
394	44	4	97	52	2130	24.6	82	2	vw pickup
395	32	4	135	84	2295	11.6	82	1	dodge rampage
396	28	4	120	79	2625	18.6	82	1	ford ranger
397	31	4	119	82	2720	19.4	82	1	chevy s-10

Gapminder Dataset

Table 9: Gapminder dataset sample available on Kaggle: first and last five objects in the dataset. The NaN in the table represents missing values.

Obj	country	continent	year	life_exp	hdi_index	co2_consump	gdp	services
0	Afghanistan	Asia	1998	53.3	0.344	0.052	NaN	24.4
1	Afghanistan	Asia	1999	54.7	0.348	0.040	NaN	24.6
2	Afghanistan	Asia	2000	54.7	0.350	0.037	NaN	24.7
3	Afghanistan	Asia	2001	54.8	0.353	0.038	NaN	24.7
4	Afghanistan	Asia	2002	55.5	0.384	0.047	333	25.6
...	...	...	...	...	...	...	...	...
3670	Zimbabwe	Africa	2014	58	0.547	0.881	1440	25.4
3671	Zimbabwe	Africa	2015	58.6	0.553	0.881	1450	25.7
3672	Zimbabwe	Africa	2016	59.2	0.558	0.771	1430	26.1
3673	Zimbabwe	Africa	2017	59.9	0.563	0.845	1480	26.6
3674	Zimbabwe	Africa	2018	60.6	0.569	0.850	1510	27.2

Naturalearth_lowres

Table 10: Naturalearth_lowres dataset sample available at the `Geopandas` library: first and last five objects in the dataset.

Obj	pop_est	continent	name	iso_a3	gdp_md_est	geometry*
0	889,953	Oceania	Fiji	FJI	5496	MULTIPOLYGON
1	58,005,463	Africa	Tanzania	TZA	63177	POLYGON
2	603,253	Africa	W. Sahara	ESH	907	POLYGON
3	37,589,262	America	Canada	CAN	1736425	MULTIPOLYGON
4	328,239,523	America	USA	USA	21433226	MULTIPOLYGON
...	...	...	...	...	...	...
172	6,944,975	Europe	Serbia	SRB	51475	POLYGON
173	622,137	Europe	Montenegro	MNE	5542	POLYGON
174	1,794,248	Europe	Kosovo	–99	7926	POLYGON
175	1,394,973	America	Trinidad and Tobago	TTO	24269	POLYGON
176	11,062,113	Africa	S. Sudan	SSD	11998	POLYGON

* The coordinates were suppressed in order not to overload the table.

Data Visualization Methods

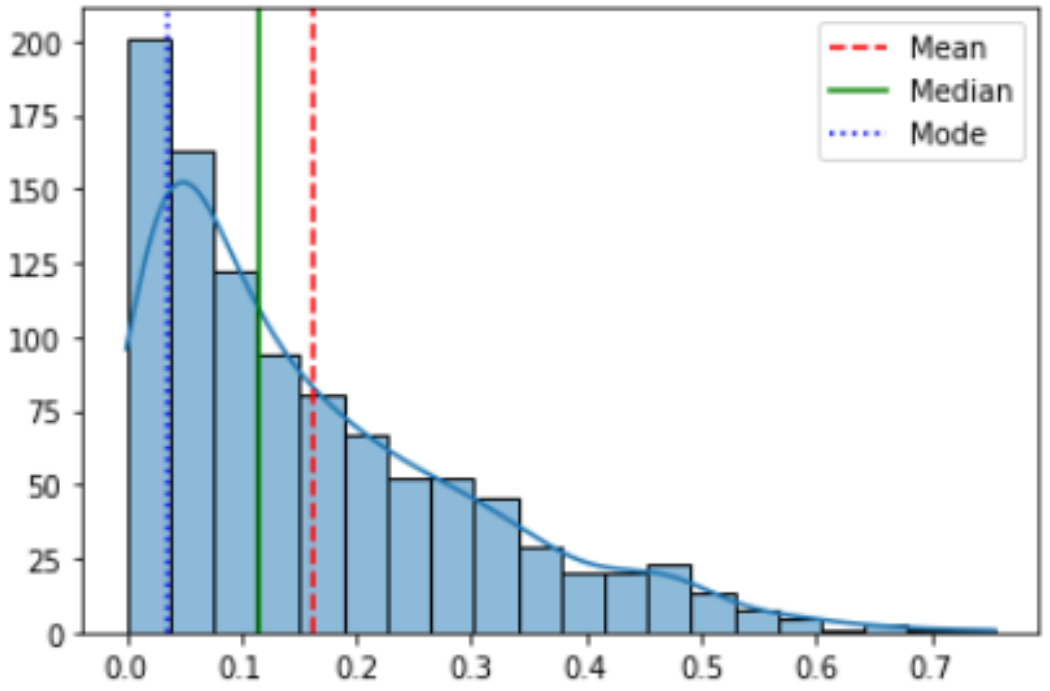
Distribution	Associations	Amounts	Proportions	Evolution and Flow	Geospatial
• Histogram • Boxplot • Violin	• Scatterplot • Scatterplot matrix • Heatmaps • Bubble chart	• Bar chart • Radar chart • Word clouds*	• Pie charts • Treemaps • Doughnut chart • Sunburst* • Dendrogram*	• Sankey chart • Gantt chart	• Choropleth map • Bubble map

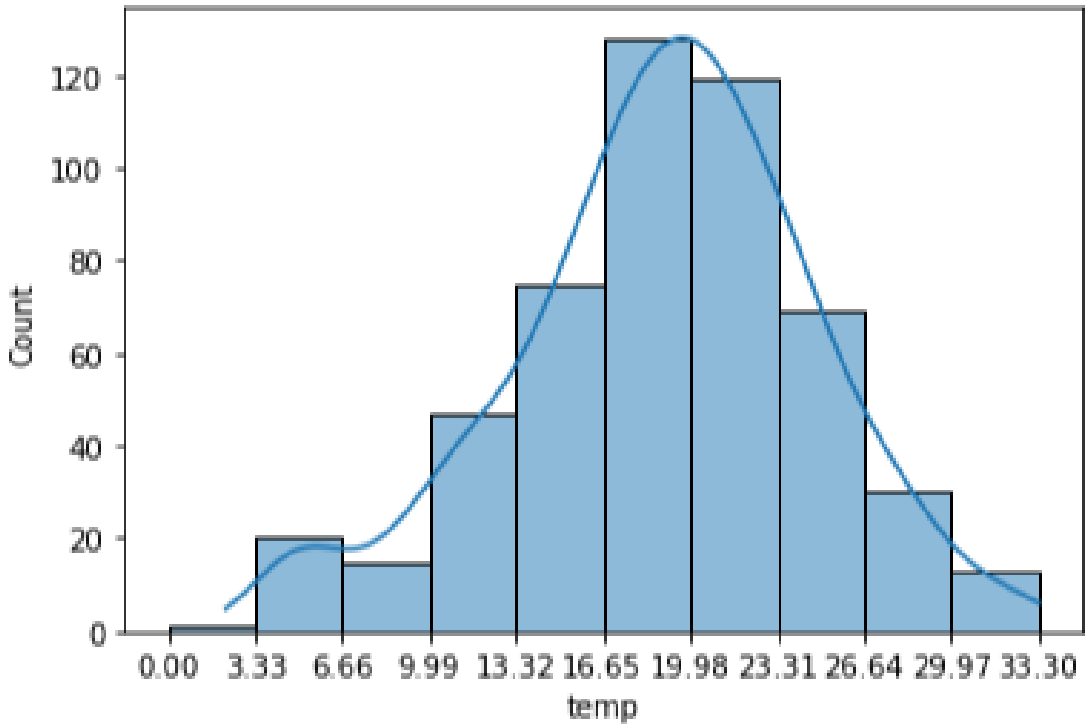
Distributions

Histogram

- **Purpose**: to explore and visualize the distribution of a variable.
- **Common type of data**: continuous data.
- **Interpretation**: the histogram divides the range of data into bins and presents the frequency of occurrence (count) of each bin in its height, providing a general overview of the variable distribution. Thus, the histogram interpretation involves assessing its overall pattern, its center, spread, shape, outliers, gaps, modes, skewness, and kurtosis.
- **Examples of applications**: to visualize any type of (continuous) variable, such as height, distance, weight, temperature, blood pressure, exam scores, income levels, etc.

Histogram

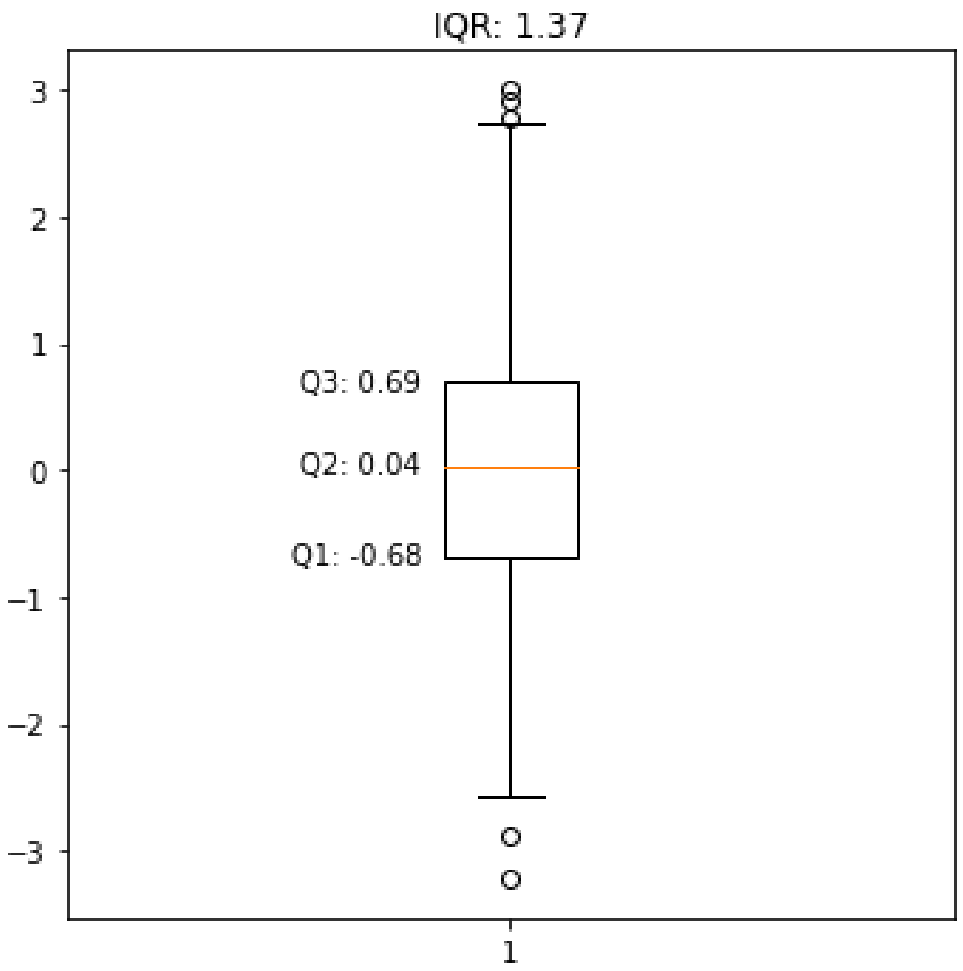

Mean
Median
Mode
200
175
150
125
100
75
50
25
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

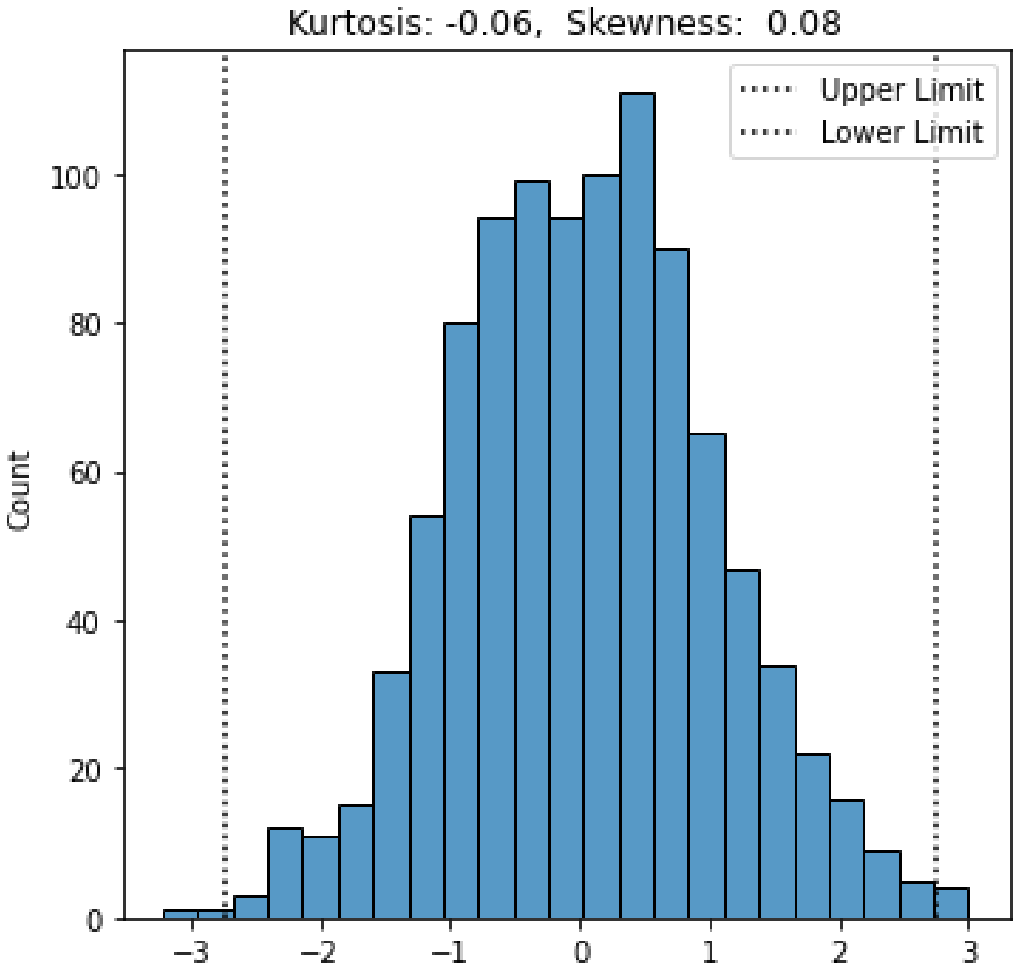

Count
120
100
80
60
40
20
0
0.00
3.33
6.66
9.99
13.32
16.65
19.98
23.31
26.64
29.97
33.30
temp

Boxplot (Box and Whisker Plot)

- **Purpose**: to explore and visualize the distribution of a variable, or to compare the distribution of different variables, using the five number summary: min, Q1, Q2, Q3, max.
- **Common type of data**: continuous data.
- **Interpretation**: allows us to observe the data distribution, its shape, and the presence of outliers.
- **Examples of applications**: to visualize the distribution of any type of (continuous) variable, such as height, distance, weight, temperature, blood pressure, exam scores, income levels, etc.

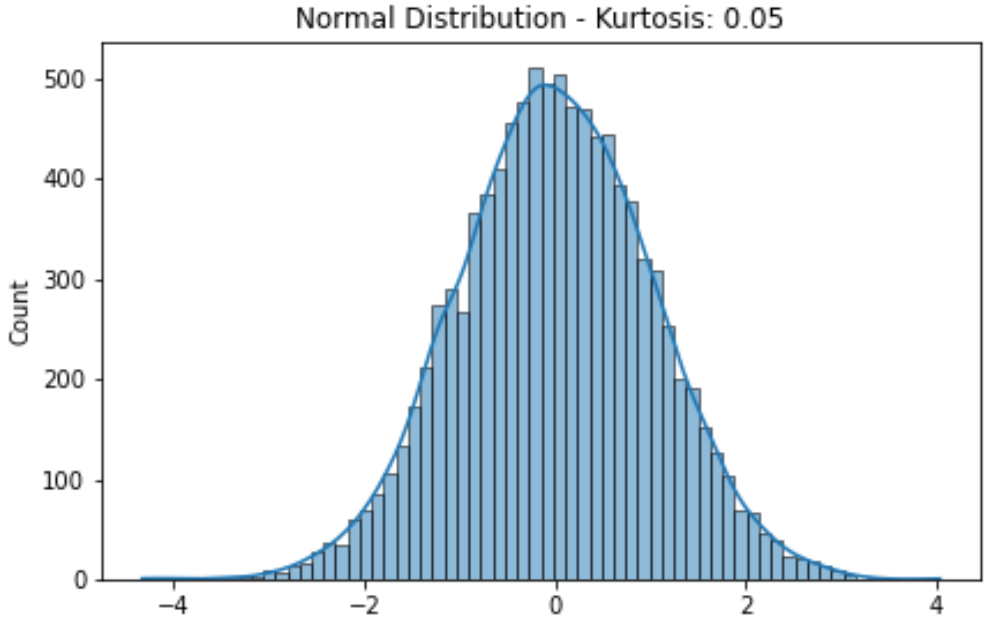
Boxplot (Box and Whisker Plot)

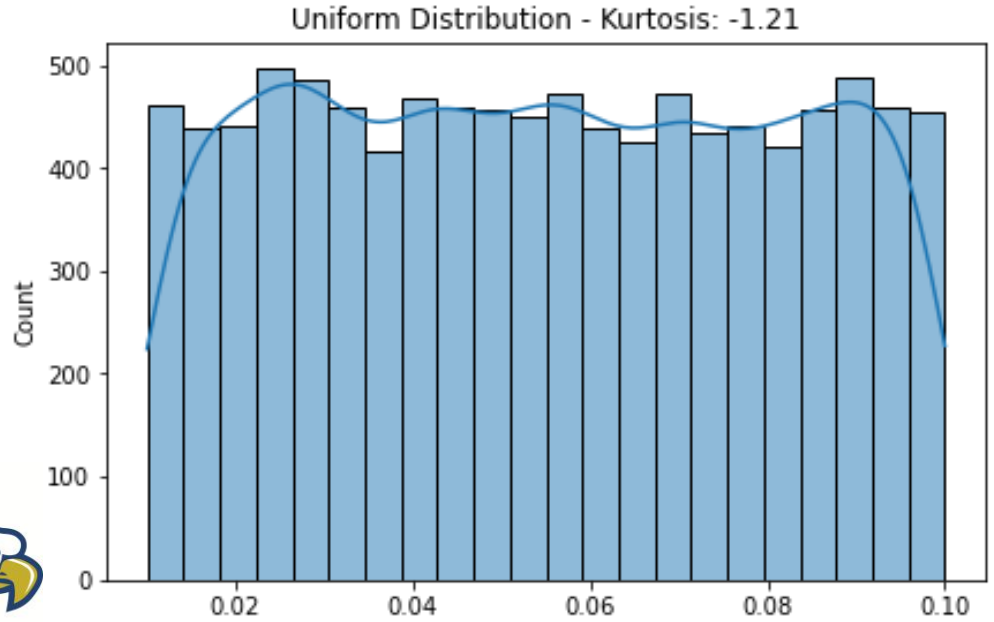


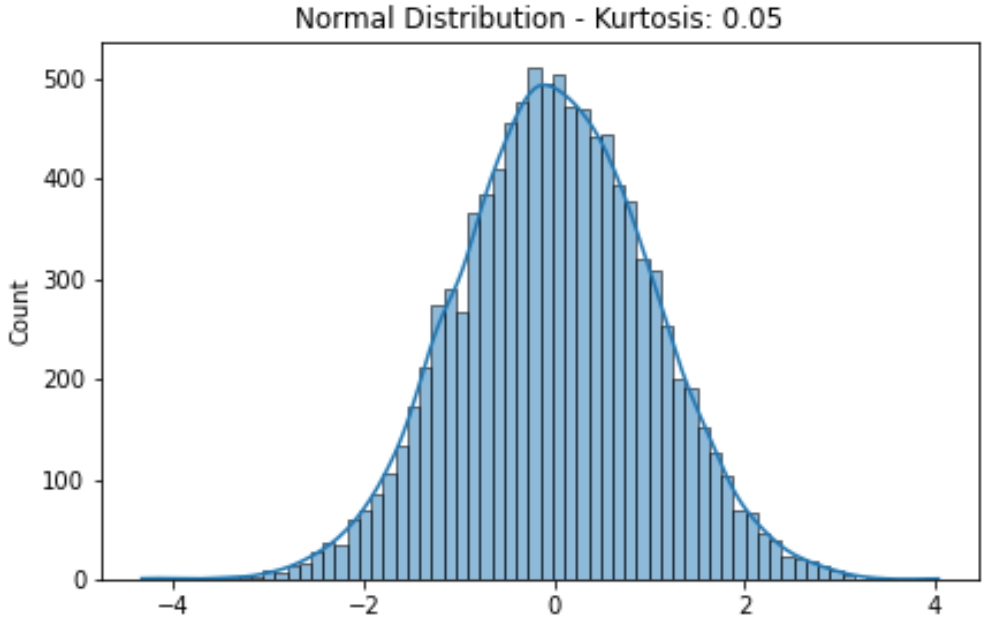


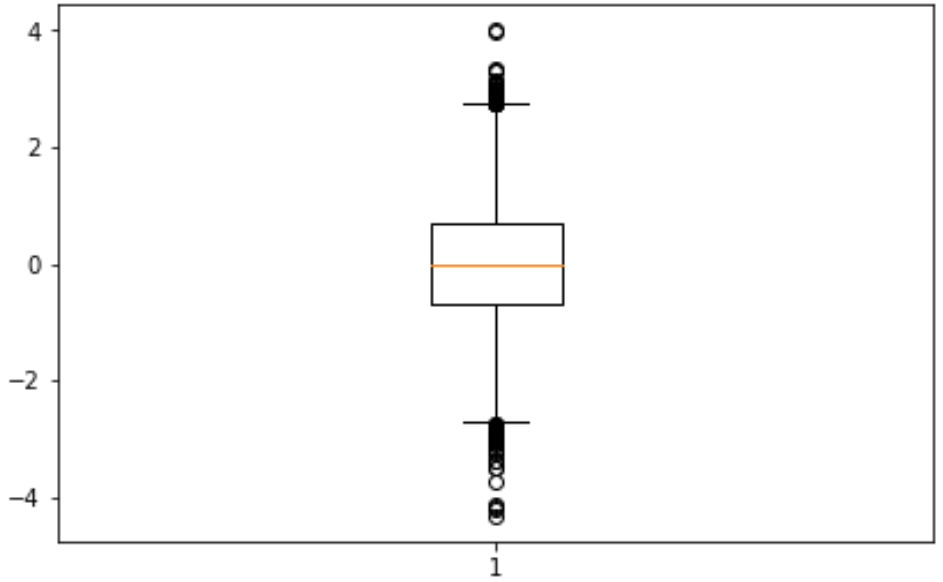
Using Boxplot to Detect Outliers

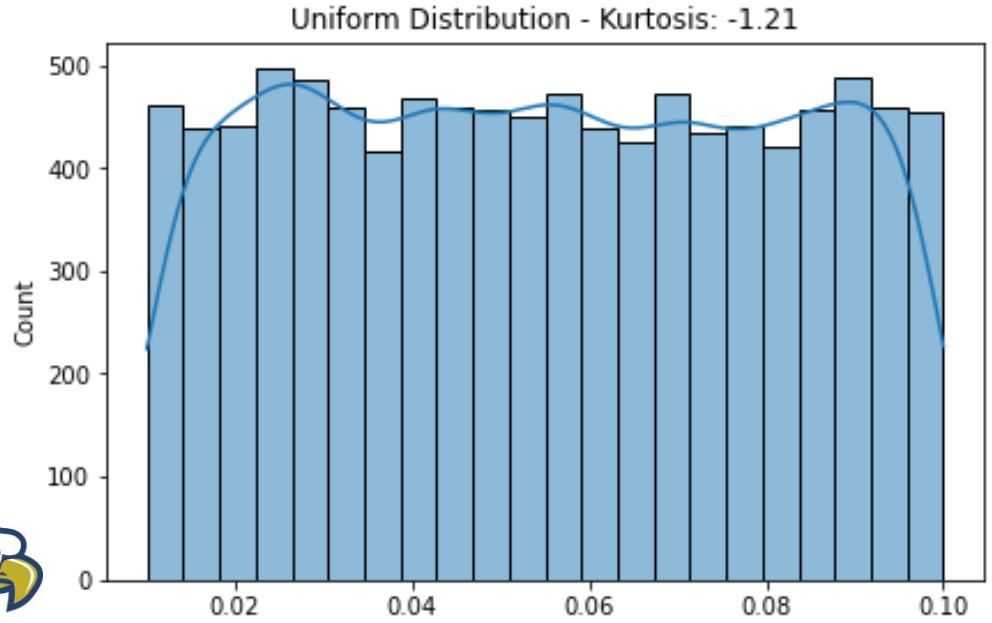
- Boxplots are a powerful tool to identify outliers in a single variable.
- Values greater than (Q_3 + γ*IQR) or smaller than (Q_1 – γ*IQR), where γ is a pre-defined IQR multiple, are considered anomalies and are plot outside the whiskers.
- Most software packages use γ = 1.5 as a standard value to identify outliers.

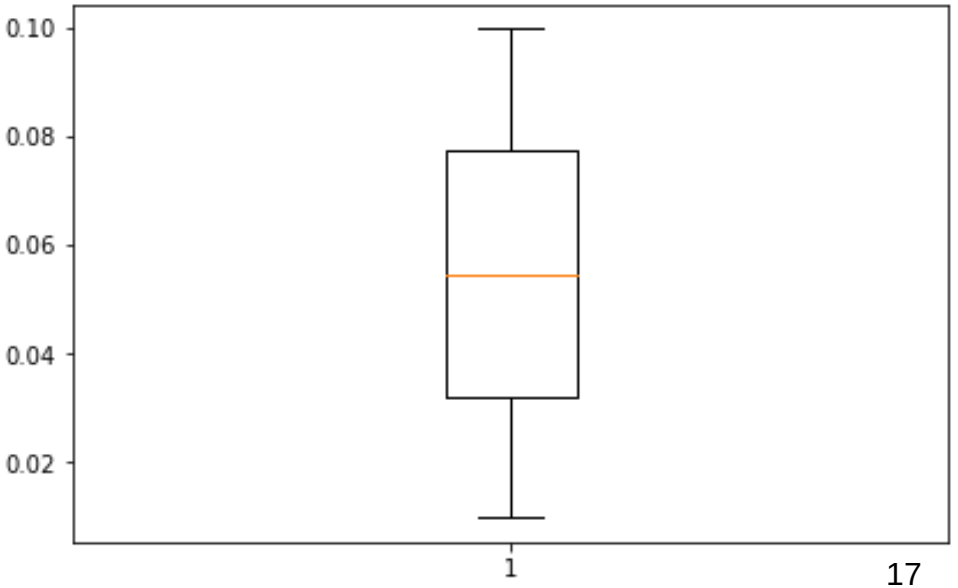

Normal Distribution - Kurtosis: 0.05
Count
500
400
300
200
100
0
−4
−2
0
2
4

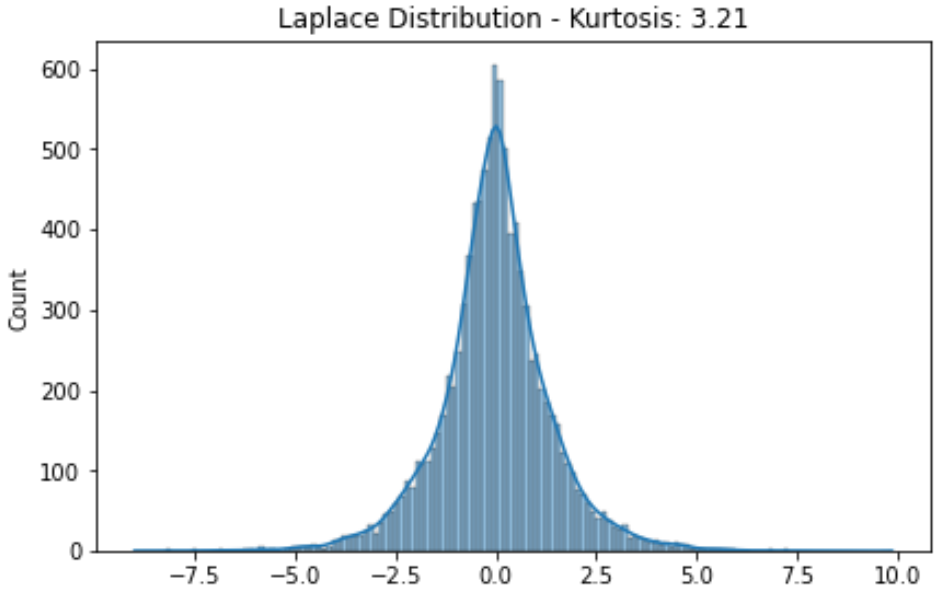

Uniform Distribution - Kurtosis: -1.21
Count
500
400
300
200
100
0
0.02
0.04
0.06
0.08
0.10

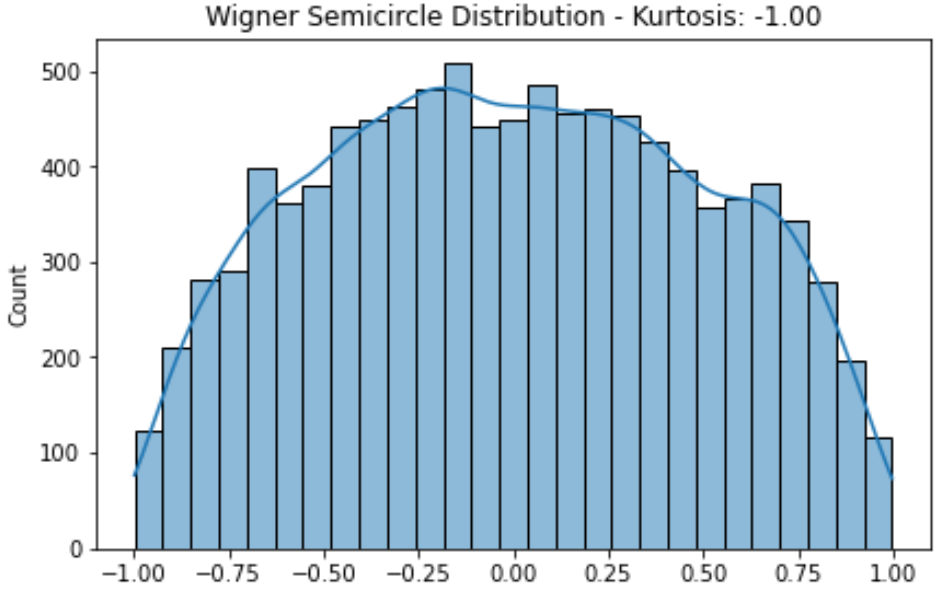

Normal Distribution - Kurtosis: 0.05
Count
500
400
300
200
100
0
−4
−2
0
2
4

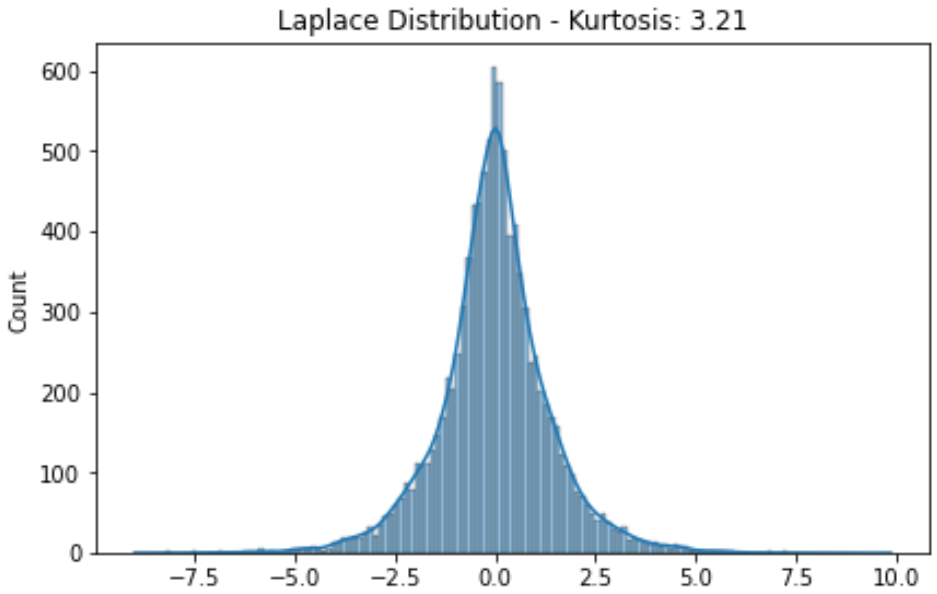

4
2
0
−2
−4
1

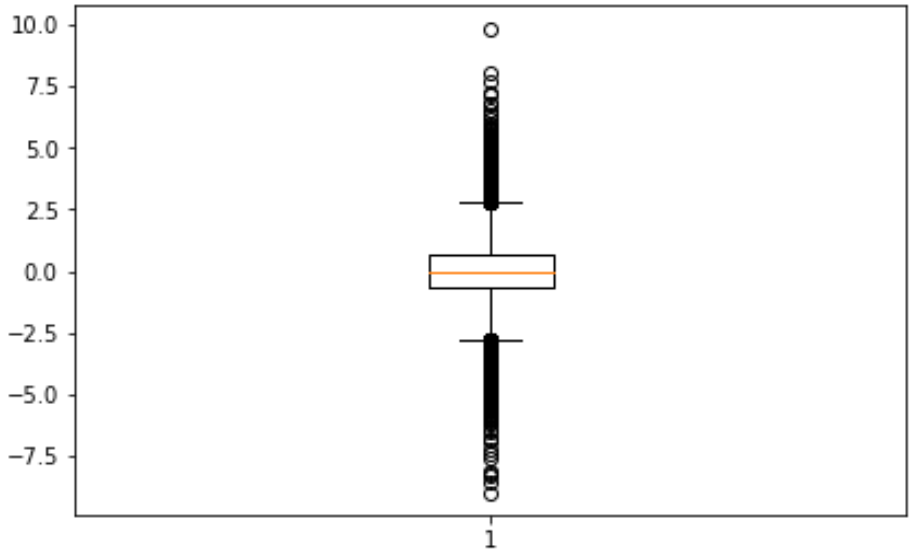

Uniform Distribution - Kurtosis: -1.21
Count
500
400
300
200
100
0
0.02
0.04
0.06
0.08
0.10

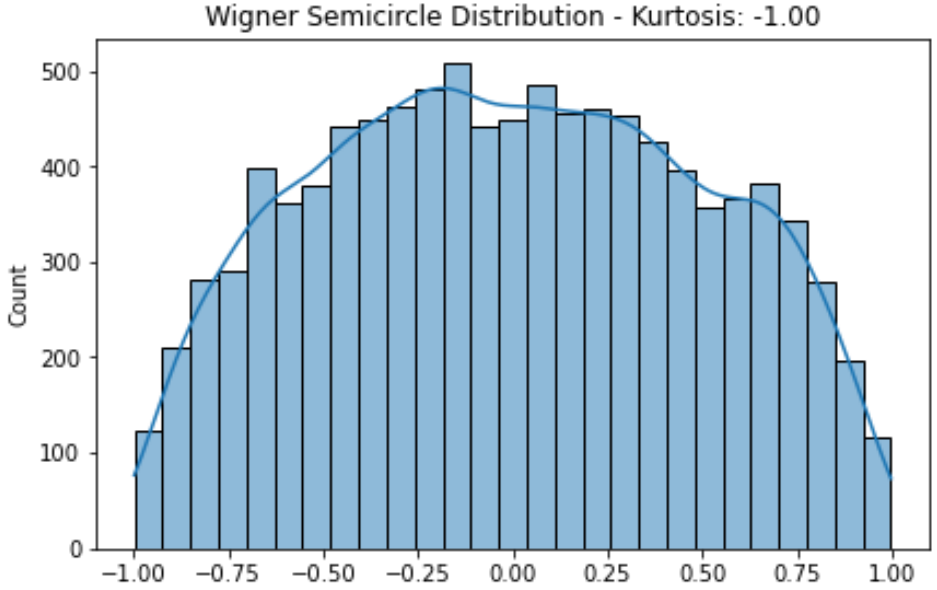

0.10
0.08
0.06
0.04
0.02
1

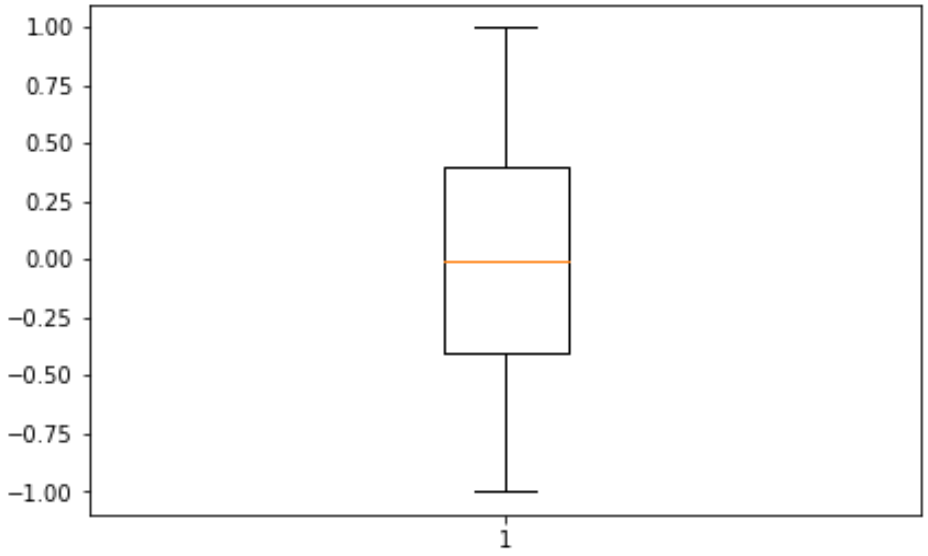

Laplace Distribution - Kurtosis: 3.21
Count
600
500
400
300
200
100
0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0

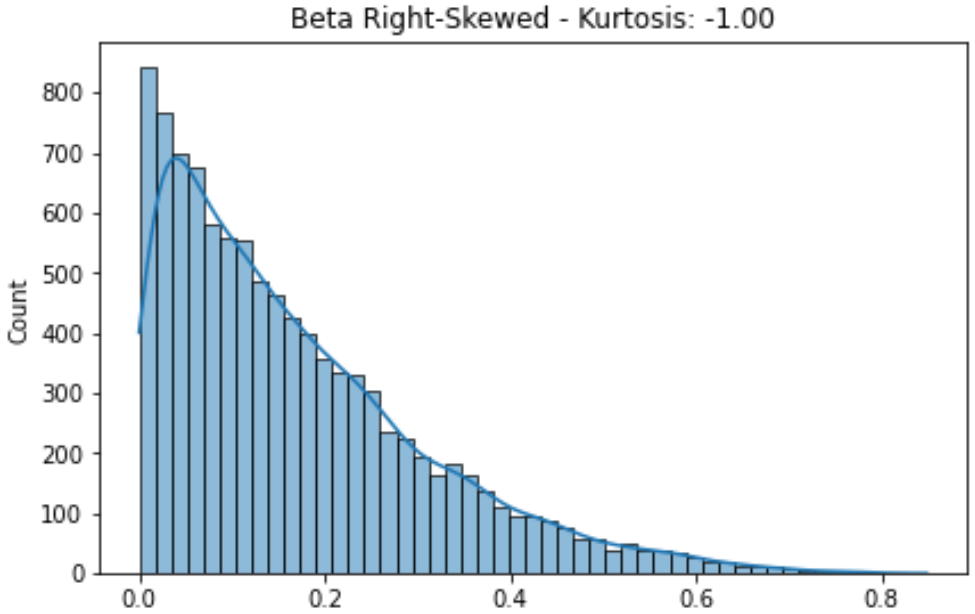

Wigner Semicircle Distribution - Kurtosis: -1.00
Count
500
400
300
200
100
0
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

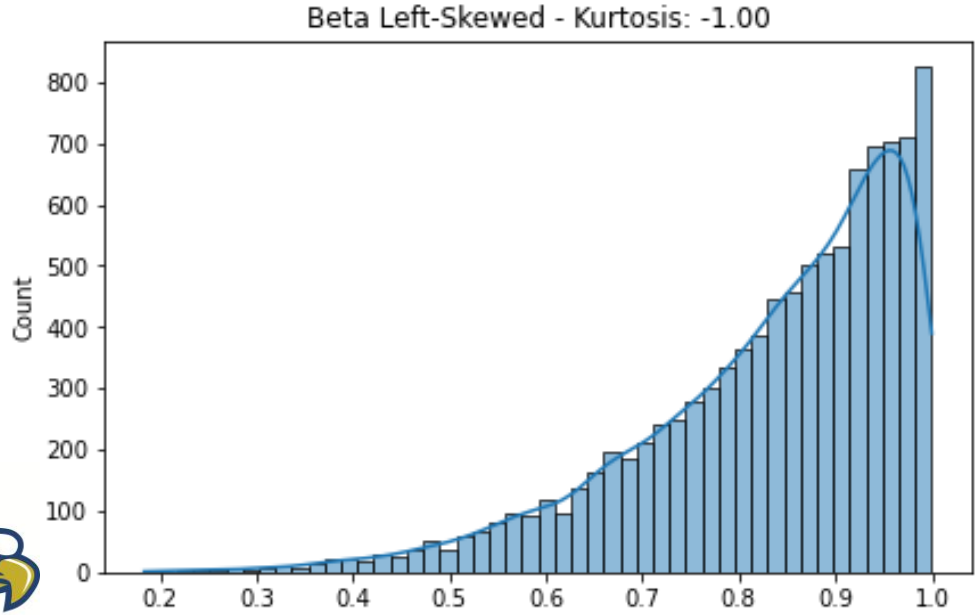

Laplace Distribution - Kurtosis: 3.21
Count
600
500
400
300
200
100
0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0

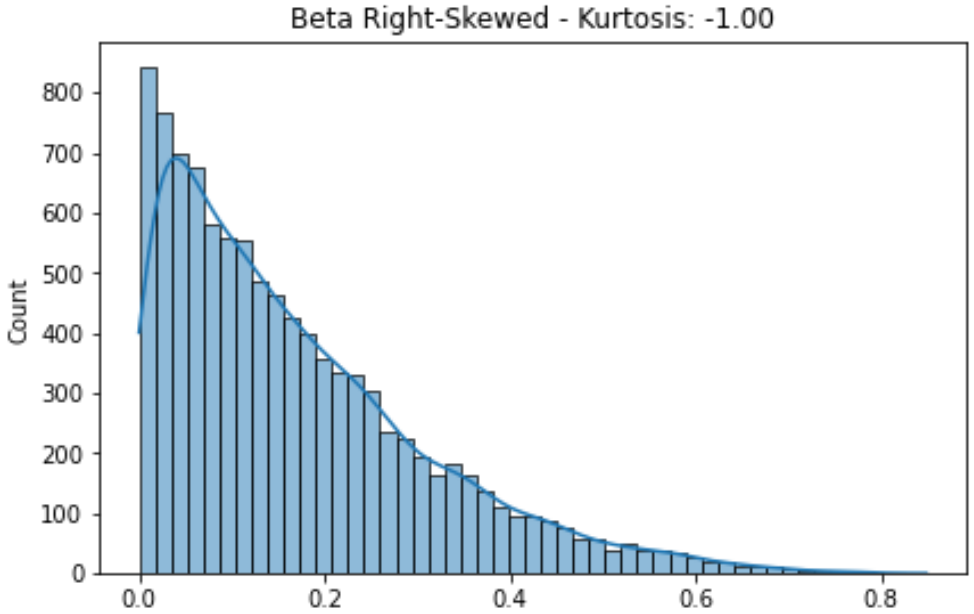

10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
1

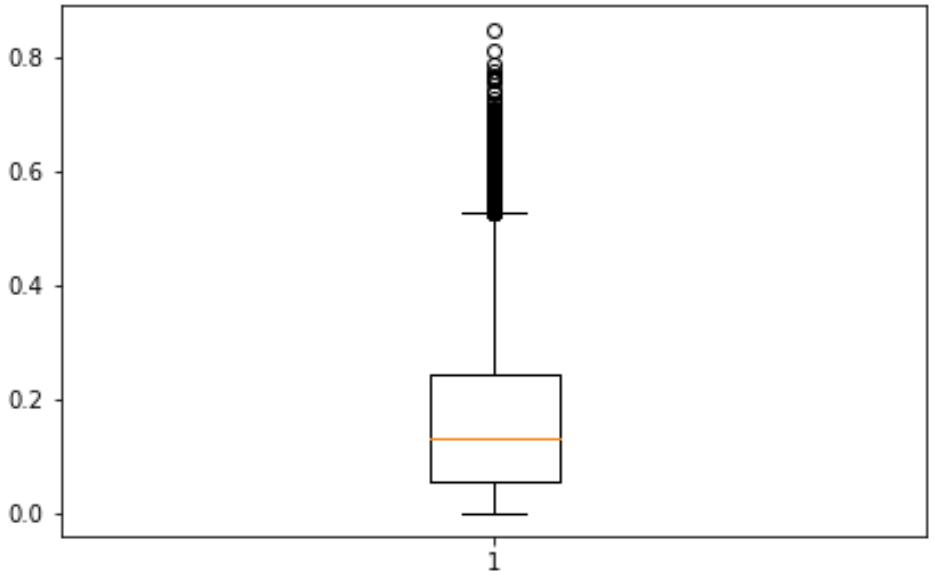

Wigner Semicircle Distribution - Kurtosis: -1.00
Count
500
400
300
200
100
0
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

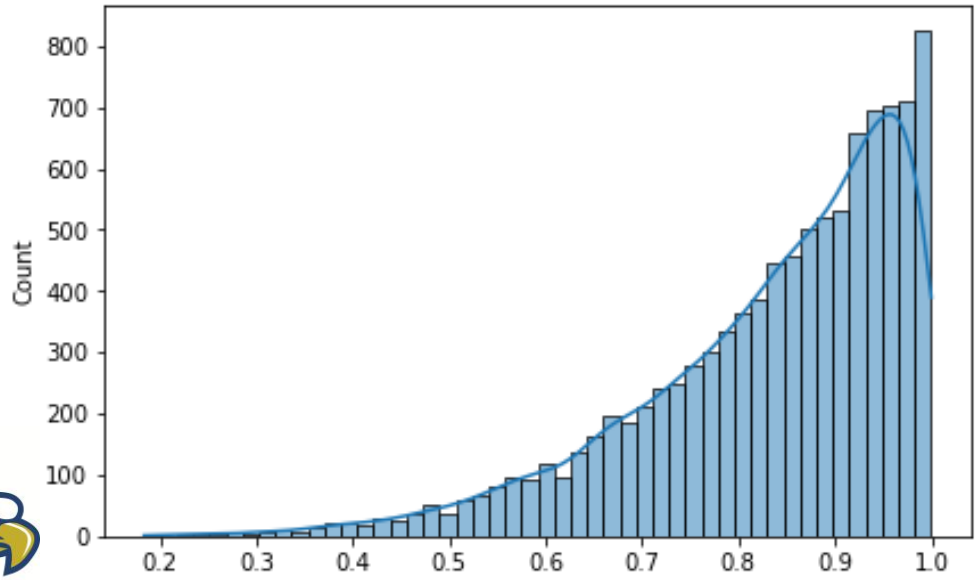

1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
1

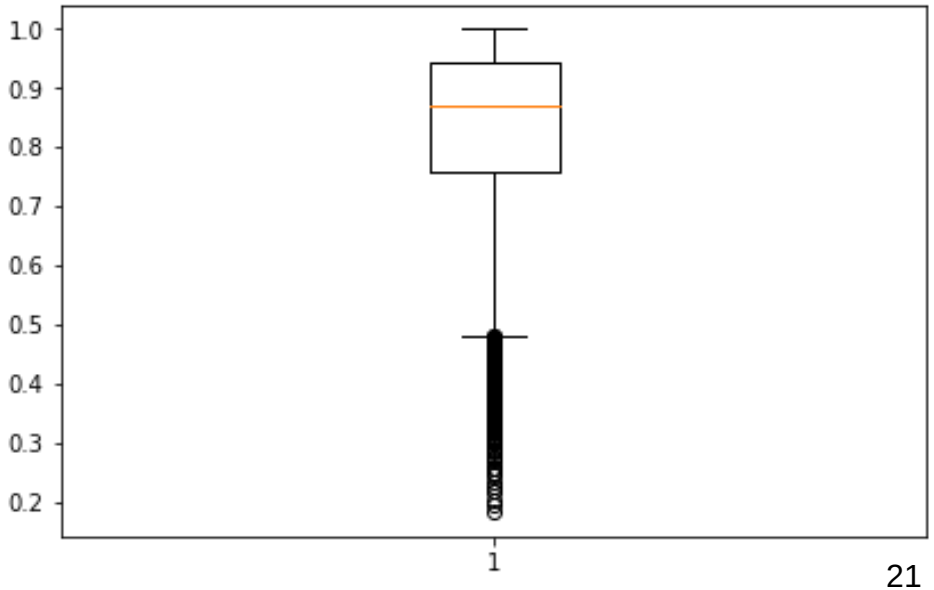

Beta Right-Skewed - Kurtosis: -1.00
Count
0
100
200
300
400
500
600
700
800
0.0
0.2
0.4
0.6
0.8


Beta Left-Skewed - Kurtosis: -1.00
Count
0
100
200
300
400
500
600
700
800
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Beta Right-Skewed - Kurtosis: -1.00
Count
0
100
200
300
400
500
600
700
800
0.0
0.2
0.4
0.6
0.8


0.0
0.2
0.4
0.6
0.8
1

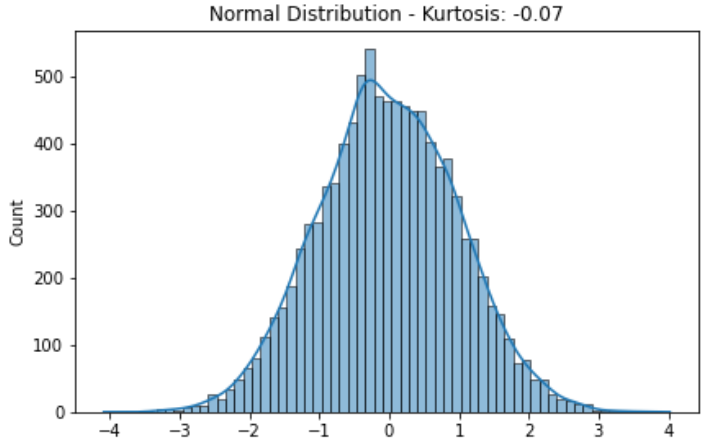

Beta Left-Skewed - Kurtosis: -1.00
Count
0
100
200
300
400
500
600
700
800
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

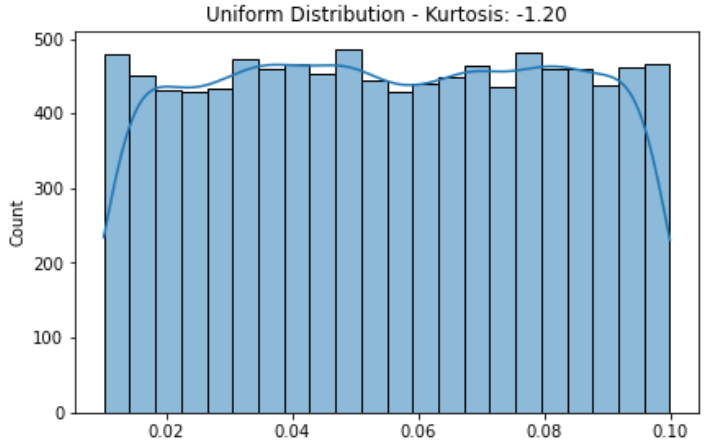

0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1

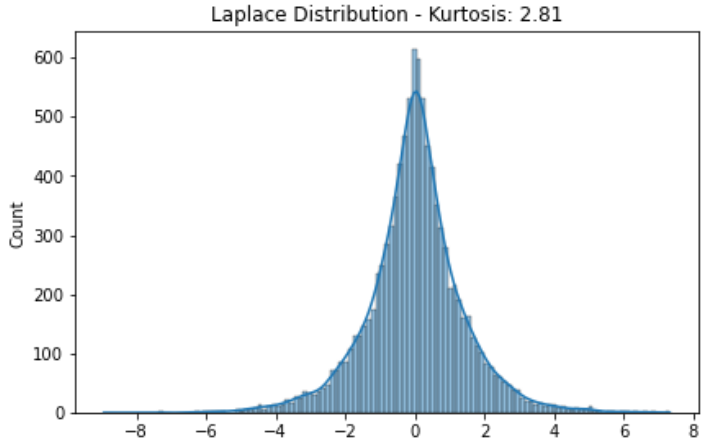
Violin Plot

- **Purpose**: to explore and visualize the distribution of a variable, or to compare the distribution of different variables, using a combination of the boxplot with a kernel density plot.
- **Common type of data**: continuous data.
- **Interpretation**: wider areas represent a larger density of data points, narrower areas mean less data points, and outliers appear out of the violin body. Thus, the violin plots allow us to observe the data distribution, its shape, and the presence of outliers.
- **Examples of applications**: to visualize the distribution of any type of (continuous) variable, such as height, distance, weight, temperature, blood pressure, exam scores, income levels, etc.

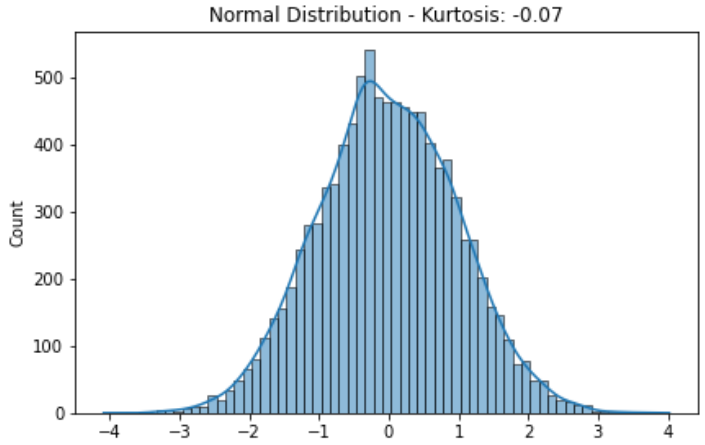
Violin Plot

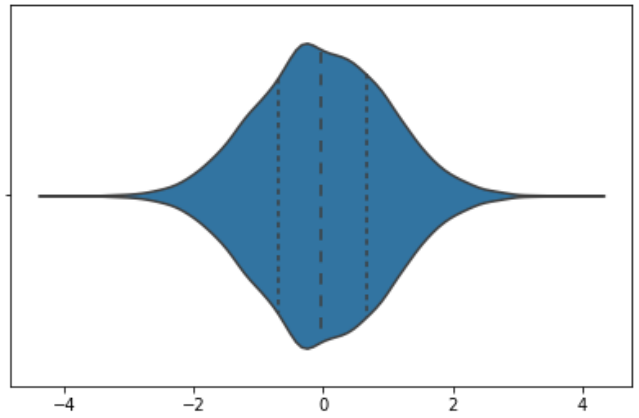

Normal Distribution - Kurtosis: -0.07
Count
0
100
200
300
400
500
−4
−3
−2
−1
0
1
2
3
4

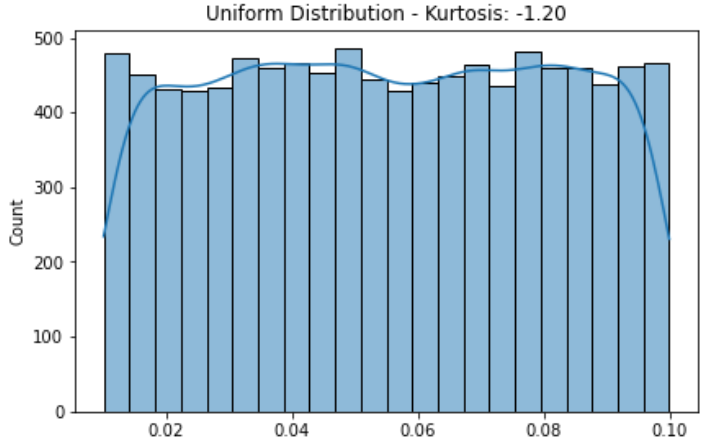

Uniform Distribution - Kurtosis: -1.20
Count
0
100
200
300
400
500
0.02
0.04
0.06
0.08
0.10

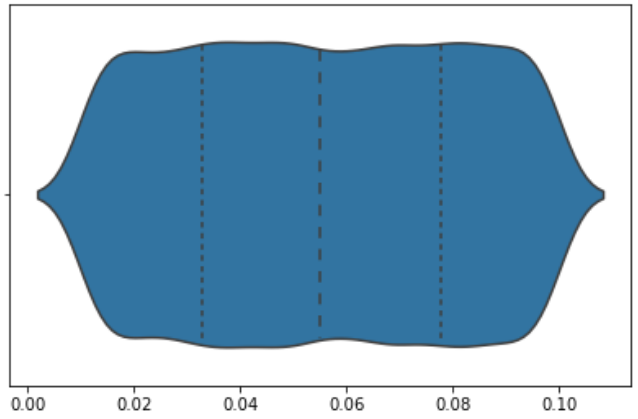

Laplace Distribution - Kurtosis: 2.81
Count
0
100
200
300
400
500
600
−8
−6
−4
−2
0
2
4
6
8

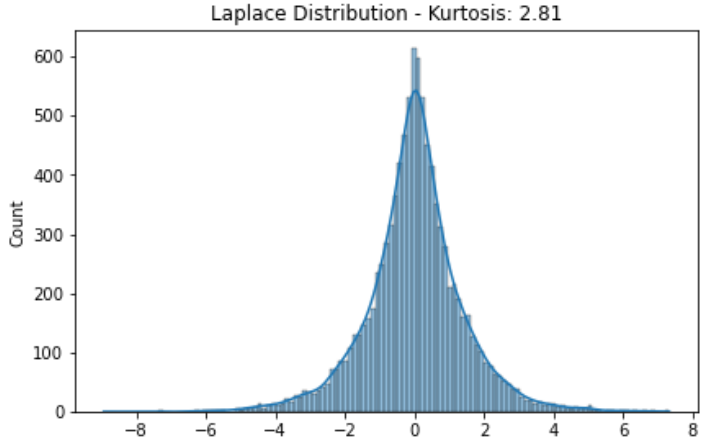
Violin Plot

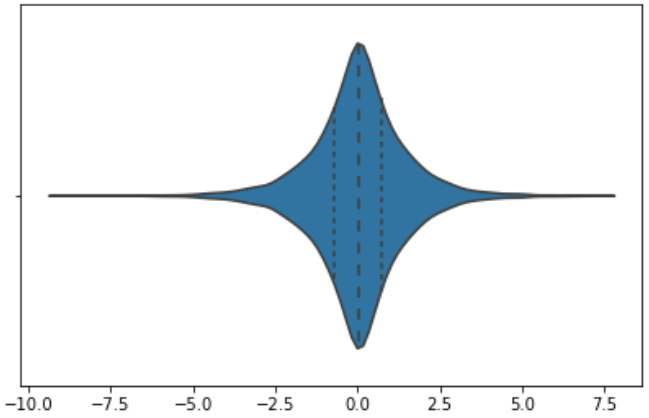

Normal Distribution - Kurtosis: -0.07
Count



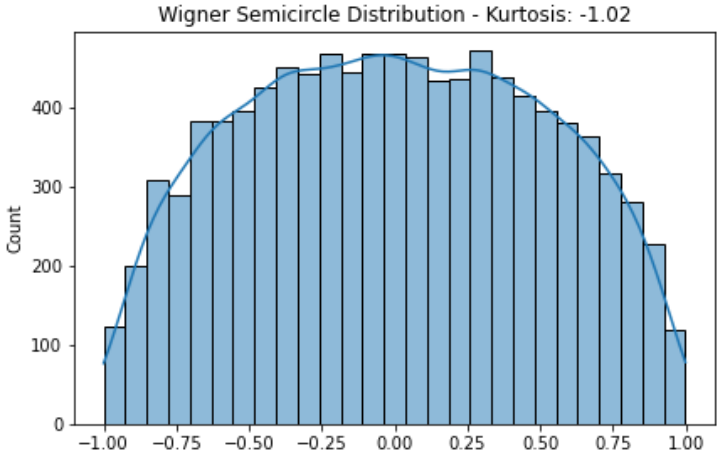

Uniform Distribution - Kurtosis: -1.20
Count

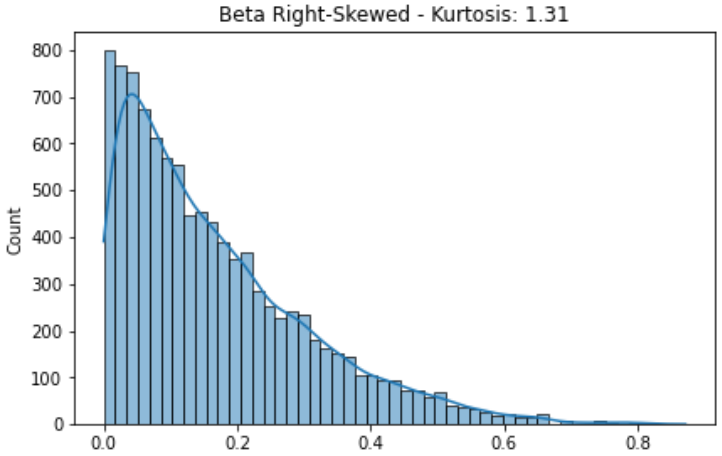


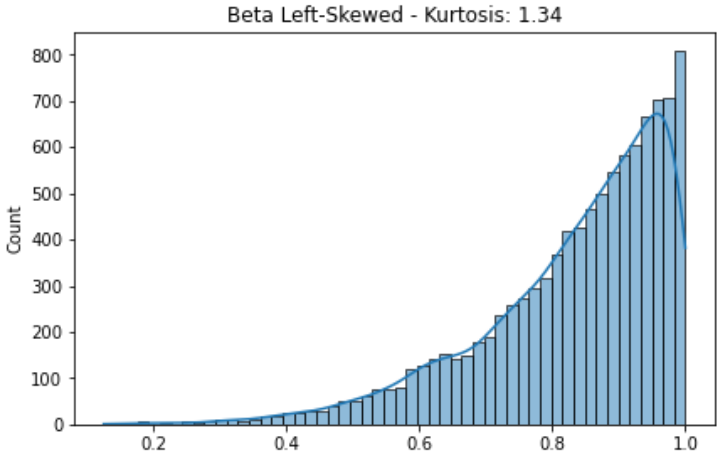

Laplace Distribution - Kurtosis: 2.81
Count



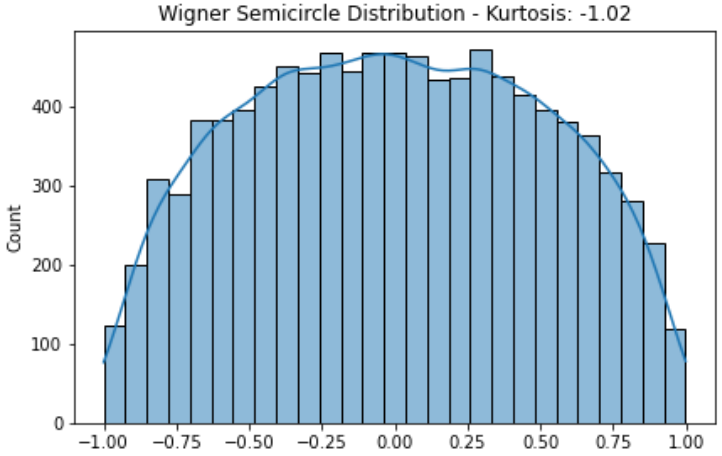
Violin Plot

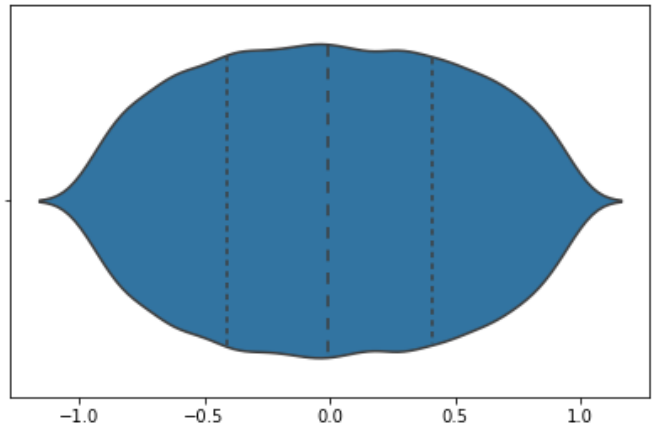

Wigner Semicircle Distribution - Kurtosis: -1.02
Count
0
100
200
300
400
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

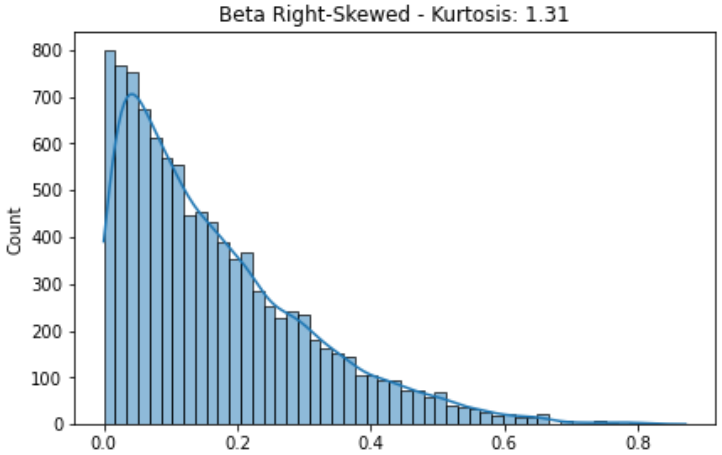

Beta Right-Skewed - Kurtosis: 1.31
Count
0
100
200
300
400
500
600
700
800
0.0
0.2
0.4
0.6
0.8

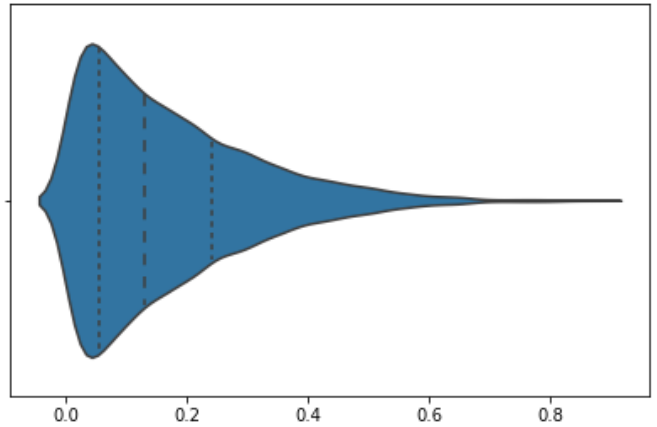

Beta Left-Skewed - Kurtosis: 1.34
Count
0
100
200
300
400
500
600
700
800
0.2
0.4
0.6
0.8
1.0

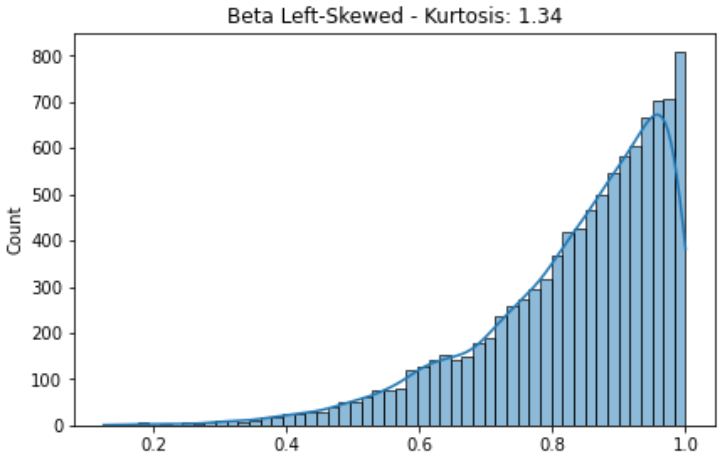
Violin Plot

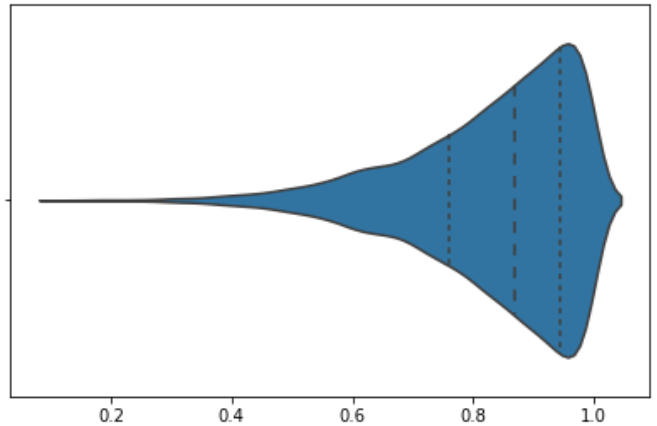

Wigner Semicircle Distribution - Kurtosis: -1.02
Count
0
100
200
300
400
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00


−1.0
−0.5
0.0
0.5
1.0

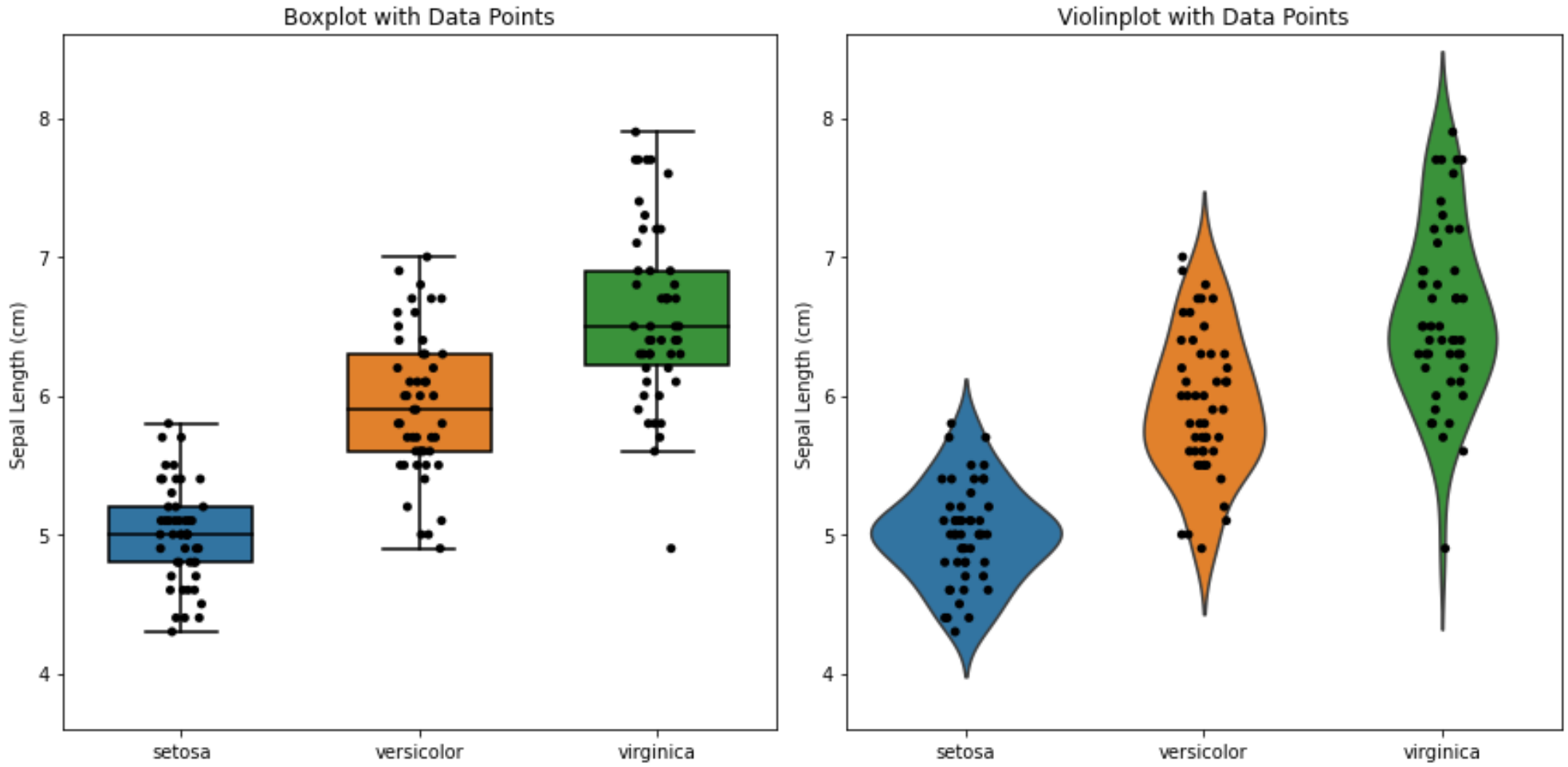

Beta Right-Skewed - Kurtosis: 1.31
Count
0
100
200
300
400
500
600
700
800
0.0
0.2
0.4
0.6
0.8


0.0
0.2
0.4
0.6
0.8


Beta Left-Skewed - Kurtosis: 1.34
Count
0
100
200
300
400
500
600
700
800
0.2
0.4
0.6
0.8
1.0


0.2
0.4
0.6
0.8
1.0

Boxplot and Violin Plot



Boxplot and Violin Plot with AI

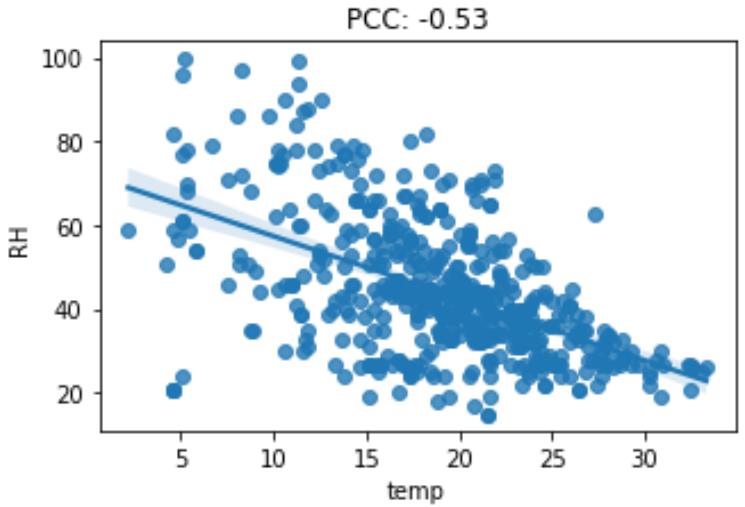
“*Generate the boxplot and the violin plot for the iris dataset of Fisher.*”

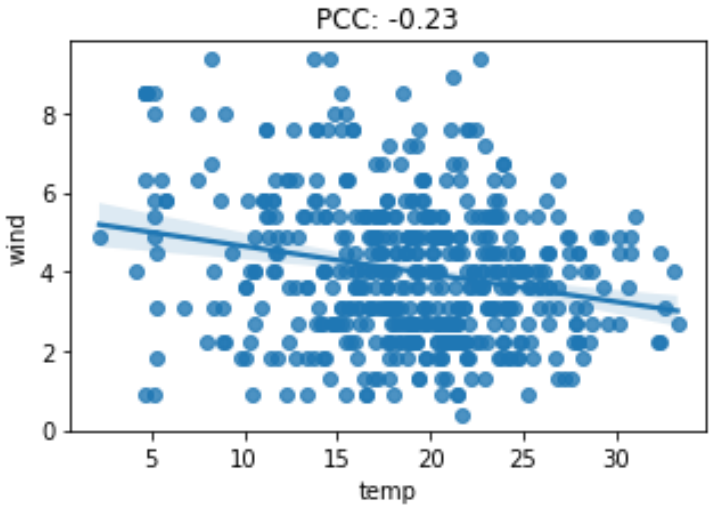
Associations

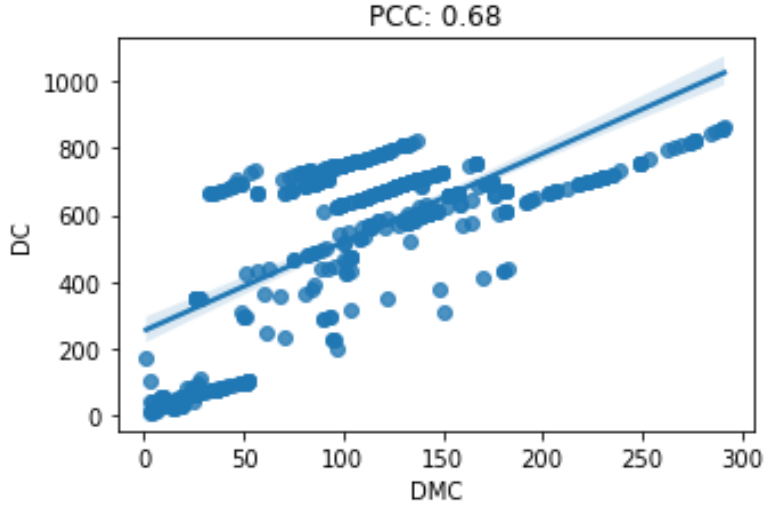
Scatter Plot

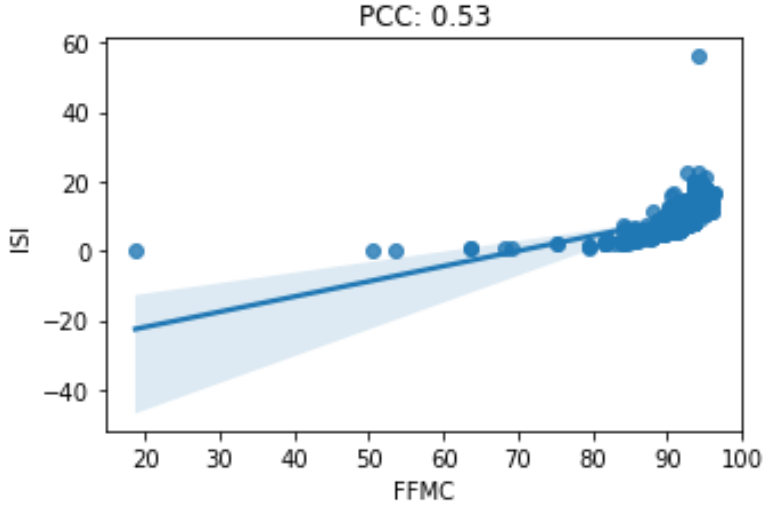
- **Purpose**: to visualize the relationship between two or three variables.
- **Common type of data**: continuous data on both axes, and categorical data in the colors.
- **Interpretation**: to analyze the type and strength of the relationship between variables (e.g., correlation) and to identify trends, patterns, and changes. It is also useful to identify clusters and outliers.
- **Examples of applications**: to visualize the relationship between weight and height, income and education level, inflation and tax rates, advertisement expenditure and sales revenue, etc.

Scatter Plot





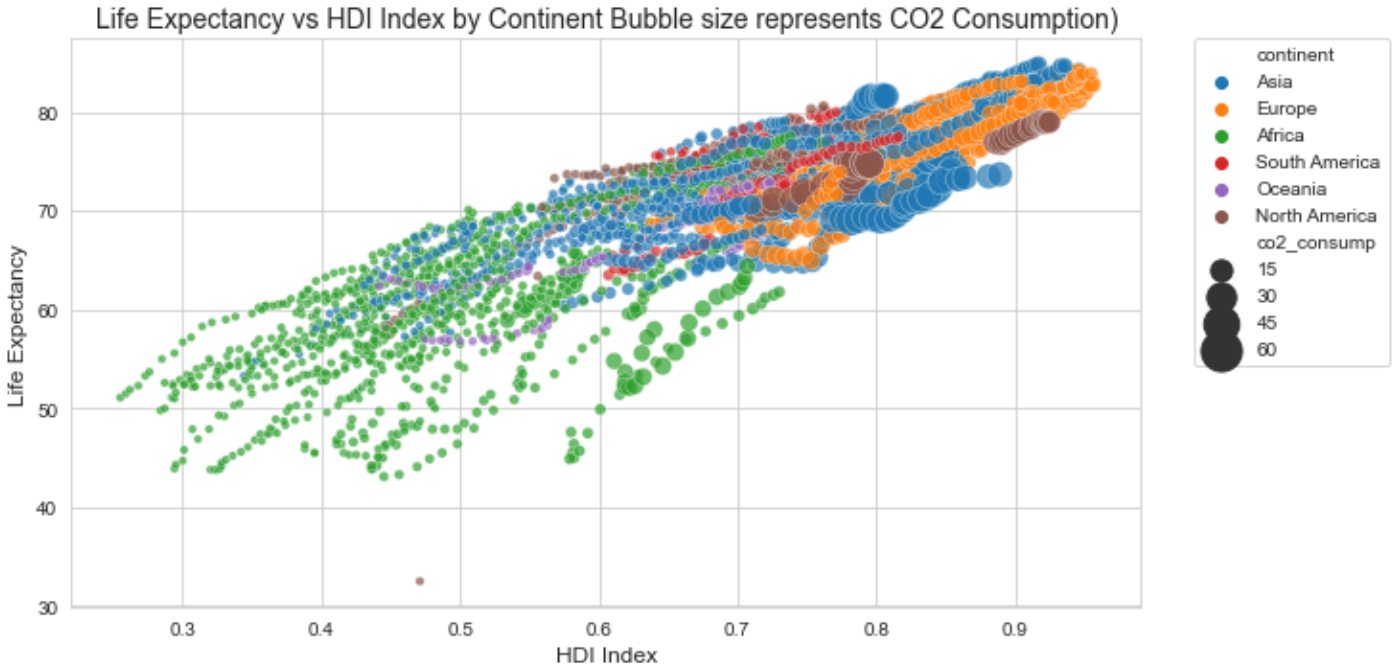


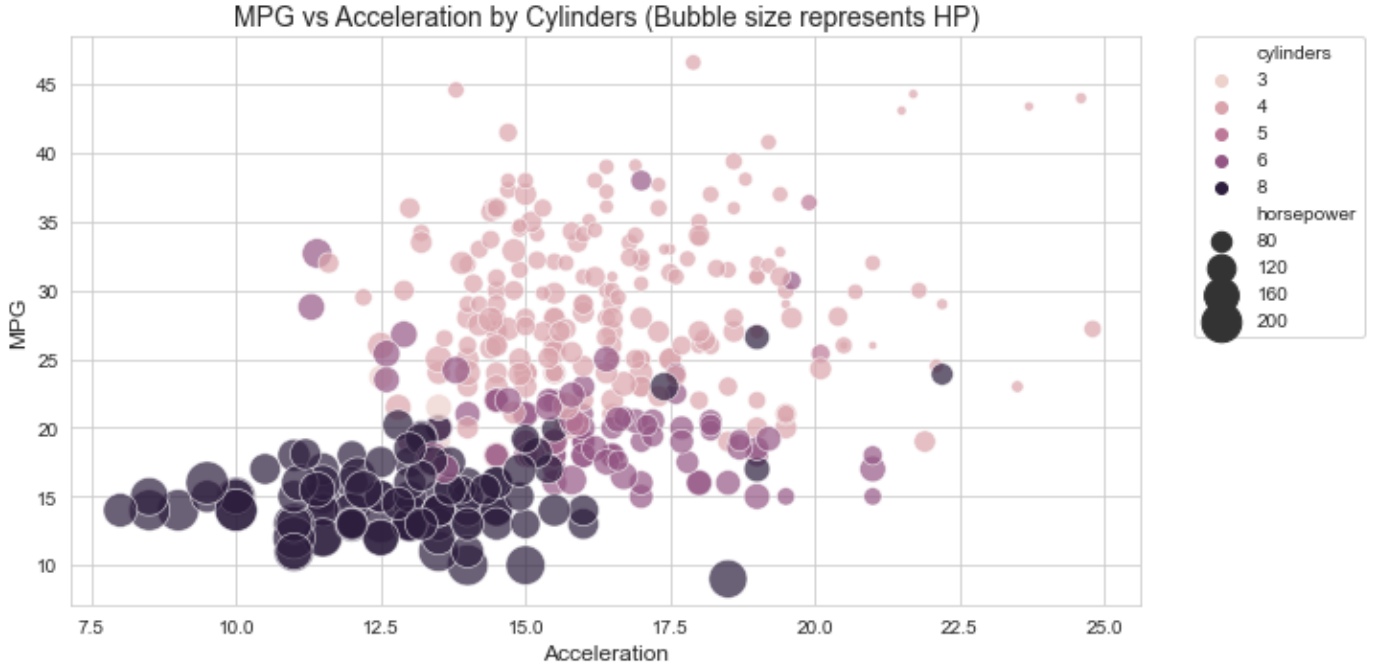


Bubble Chart

- **Purpose**: to visualize the relationship between three or four variables.
- **Common type of data**: continuous data on both axes and bubble size, and categorical data in the colors.
- **Interpretation**: an extension of the scatterplot with bubbles of varying sizes as the value of the third or fourth variable.
- **Examples of applications**: to visualize the relationship between three or four variables, like the price versus quality and market share of a given product; the relationship between income, education, and age; the fuel consumption versus acceleration, the number of cylinders and horsepower of a vehicle; and the life expectancy versus the human development index, the continent and CO_2 consumption, etc.

Bubble Chart





Bubble Chart with AI

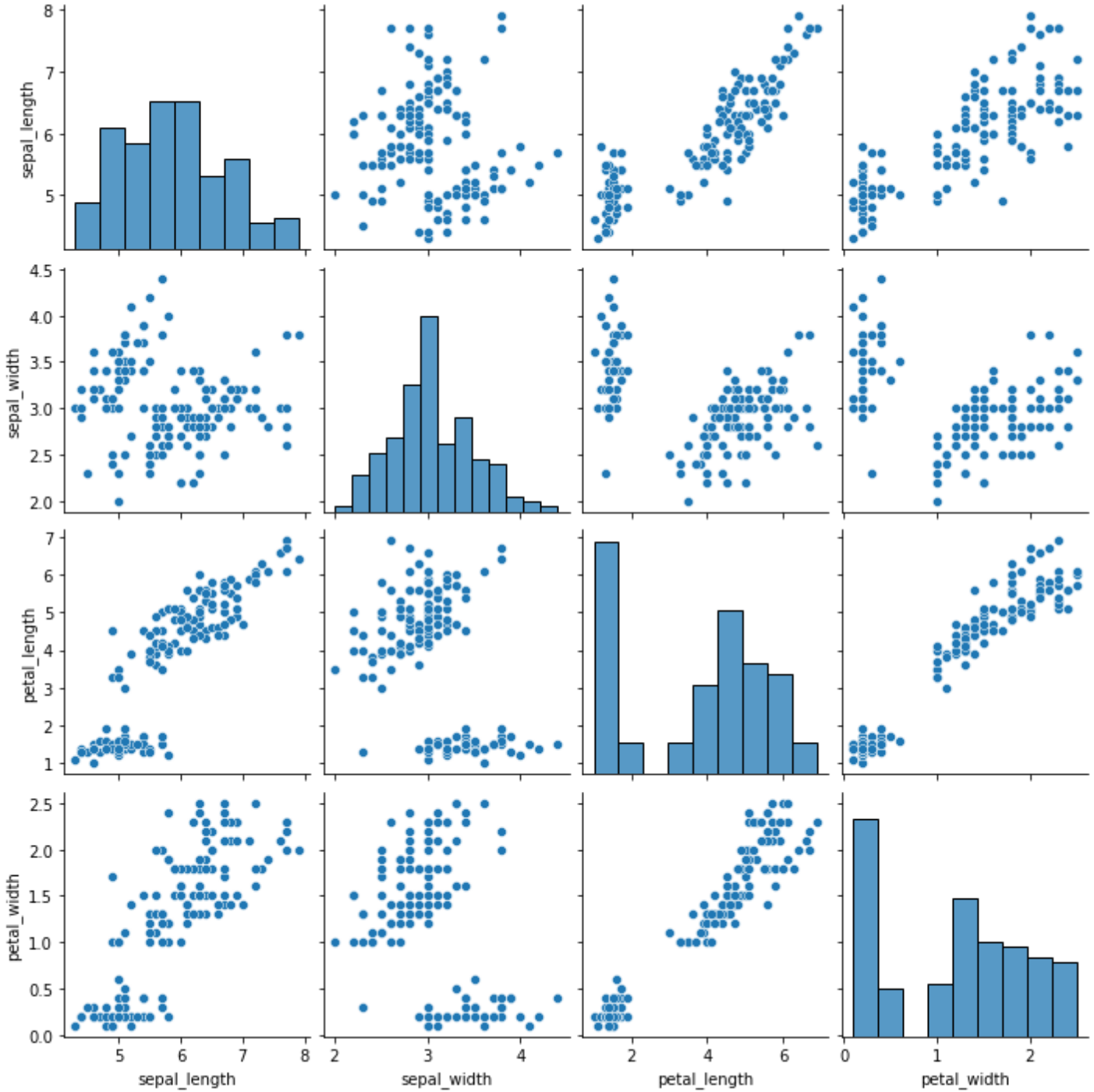
“Generate a bubble chart for the gapminder dataset attached having variable ‘HDI Index’ in the x-axis, variable 'Life Expectancy' on the y-axis, variable ‘co2_consump’ as the bubble size, and variable 'continent' as the bubble color.”

‘Generate a bubble chart for the mpg dataset attached having variable 'Acceleration' in the x-axis, variable 'MPG' on the y-axis, variable 'horsepower' as the bubble size, and variable 'cylinders' as the bubble color.’

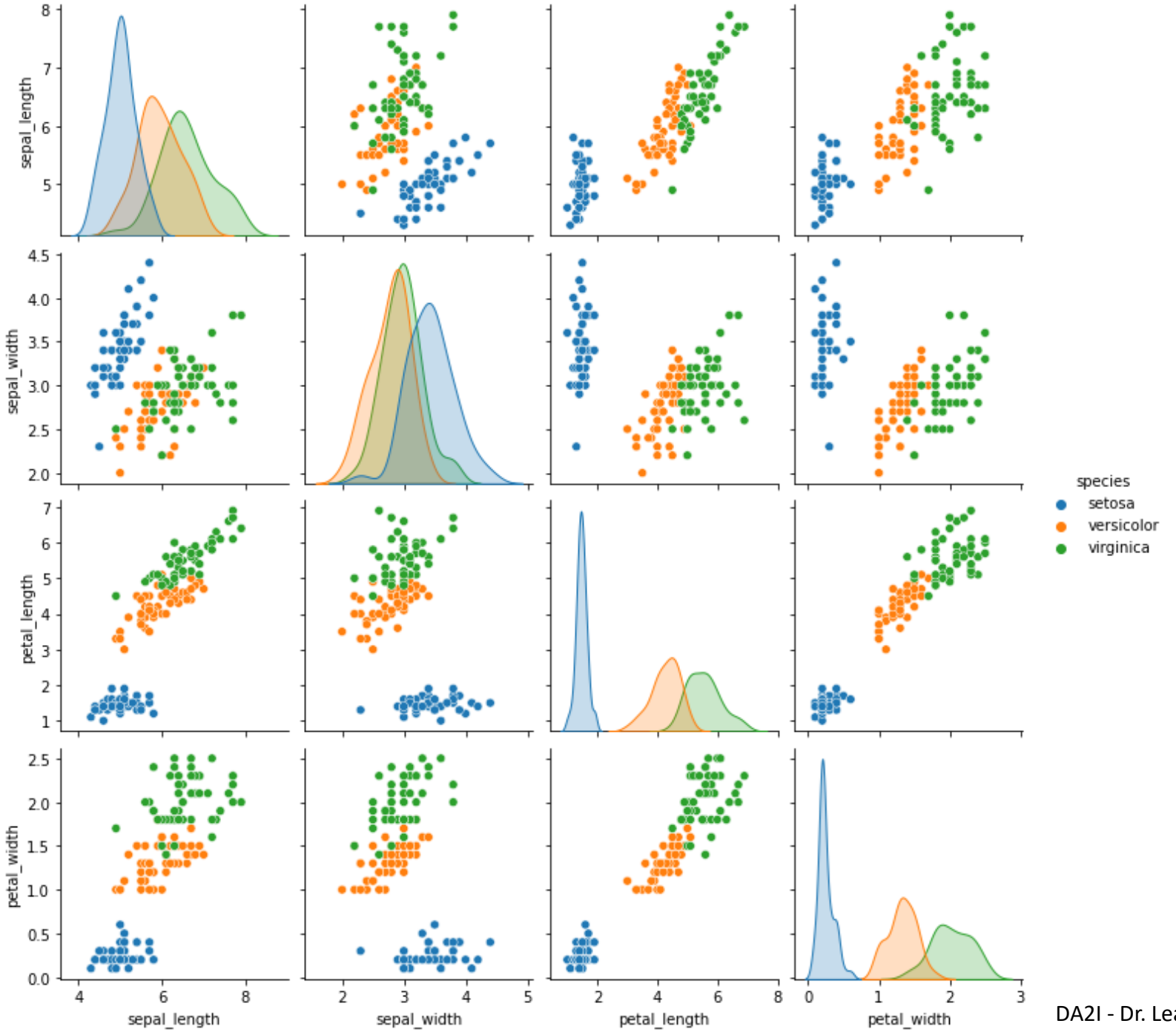
Scatterplot Matrix

- **Purpose**: to explore and visualize the relationship (association) between multiple variables simultaneously.
- **Common type of data**: continuous data.
- **Interpretation**: the diagonal plots are histograms that show the distribution of each variable, and the off-diagonal plots are pairwise scatterplots showing the relationship between each pair of variables.
- **Examples of applications**: to visualize the relationship between stock prices and financial indicators, physiological measurements and life quality, socioeconomic indicators and education level, etc.

Scatterplot Matrix



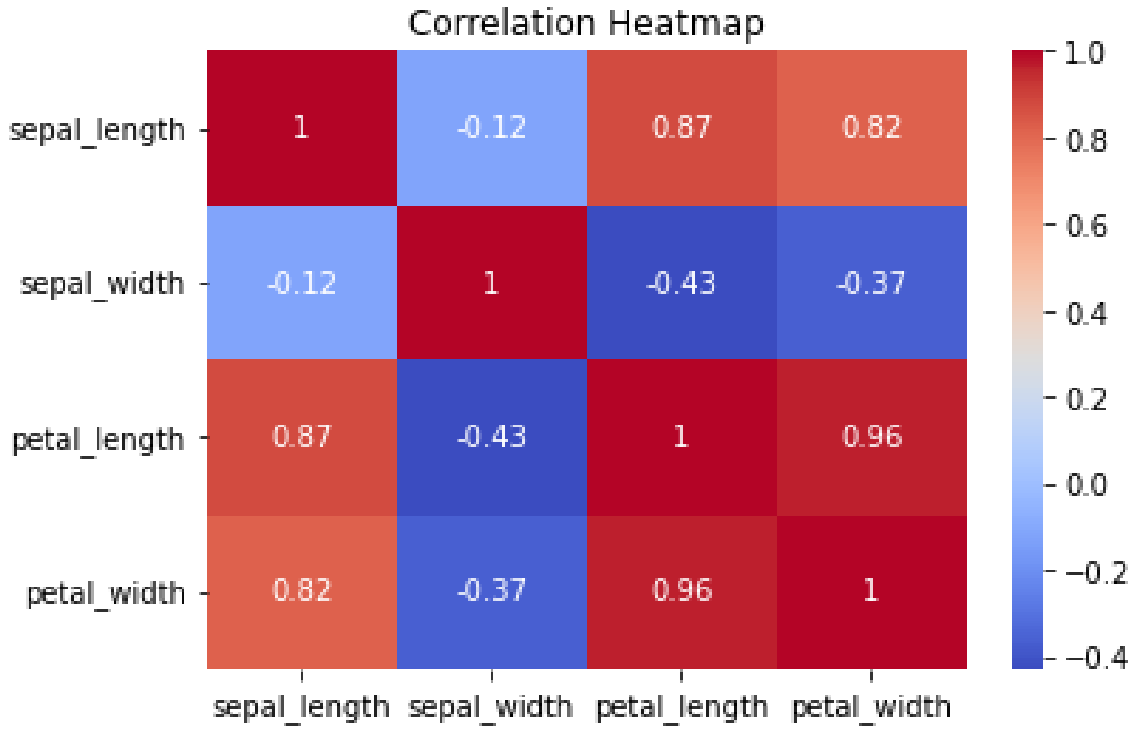
Scatterplot Matrix



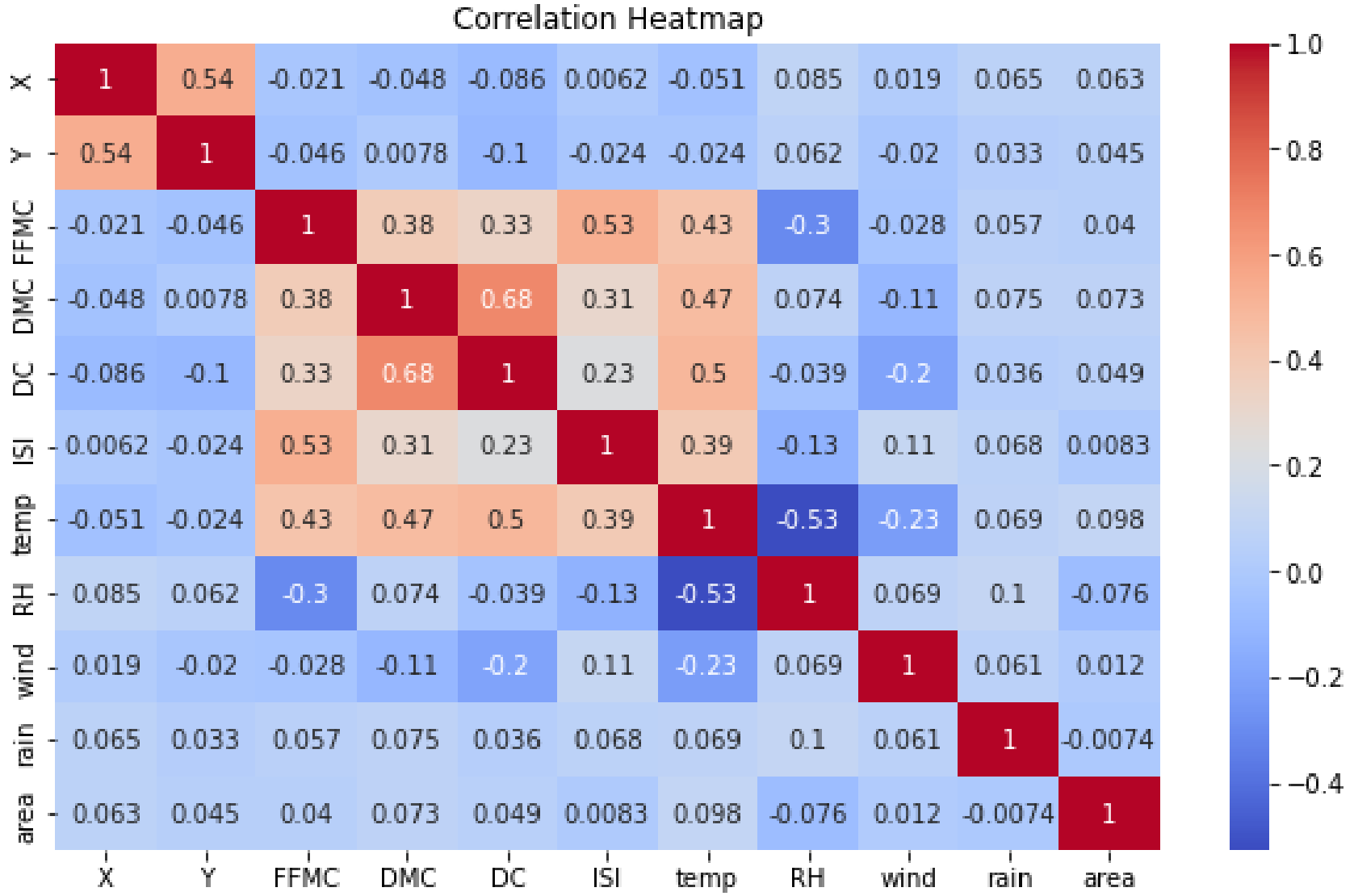
Heatmaps and Correlograms

- **Purpose**: to use colors to explore and visualize the magnitude (e.g., correlations) of variables.
- **Common type of data**: numerical (discrete or continuous) data.
- **Interpretation**: typically use a color gradient to represent the values of the data, where darker colors represent higher values whilst lighter colors represent lower values.
- **Examples of applications**: to visualize the relationship among stock prices and financial indicators, physiological measurements and life quality, relationship of gene expression and dis-eases, socioeconomic indicators and education level, customer behavior and preferences, etc.

Heatmaps and Correlograms



Heatmaps and Correlograms



Scatterplot Matrix and Heatmap with AI

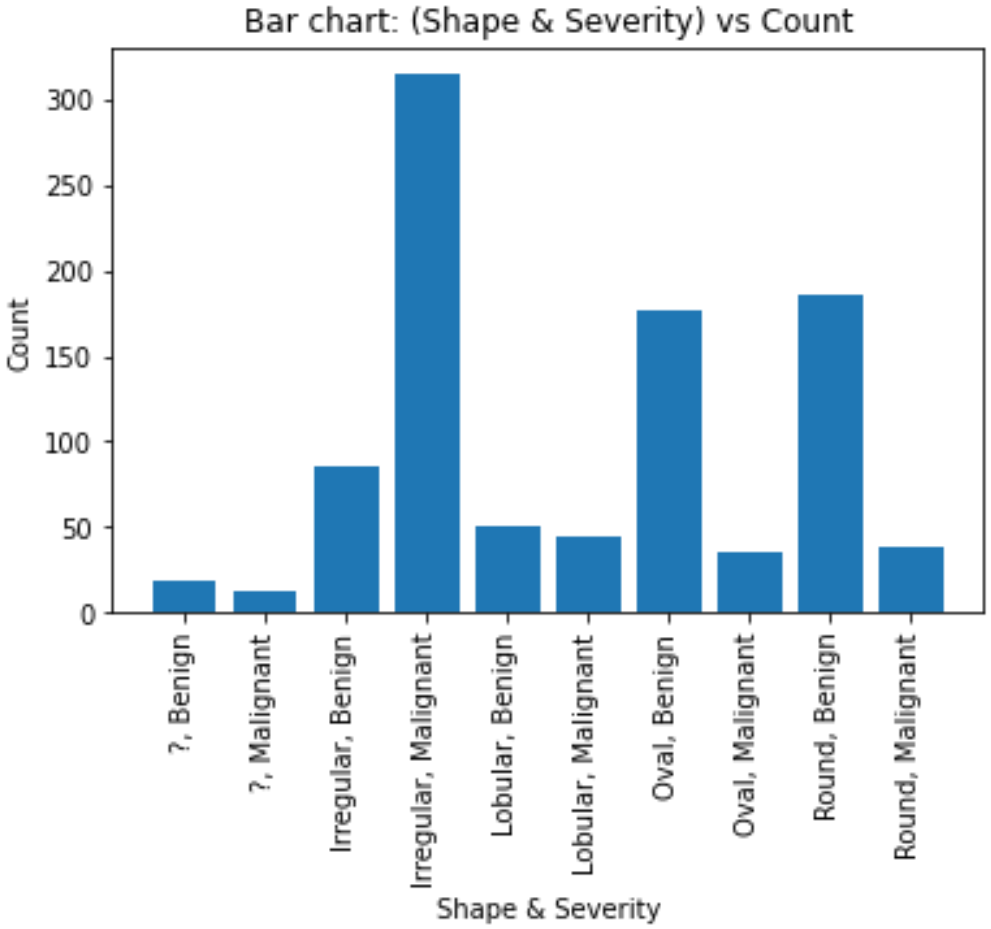
“*Generate the scatterplot matrix and correlation heatmap for the attached Iris dataset of Fisher*”

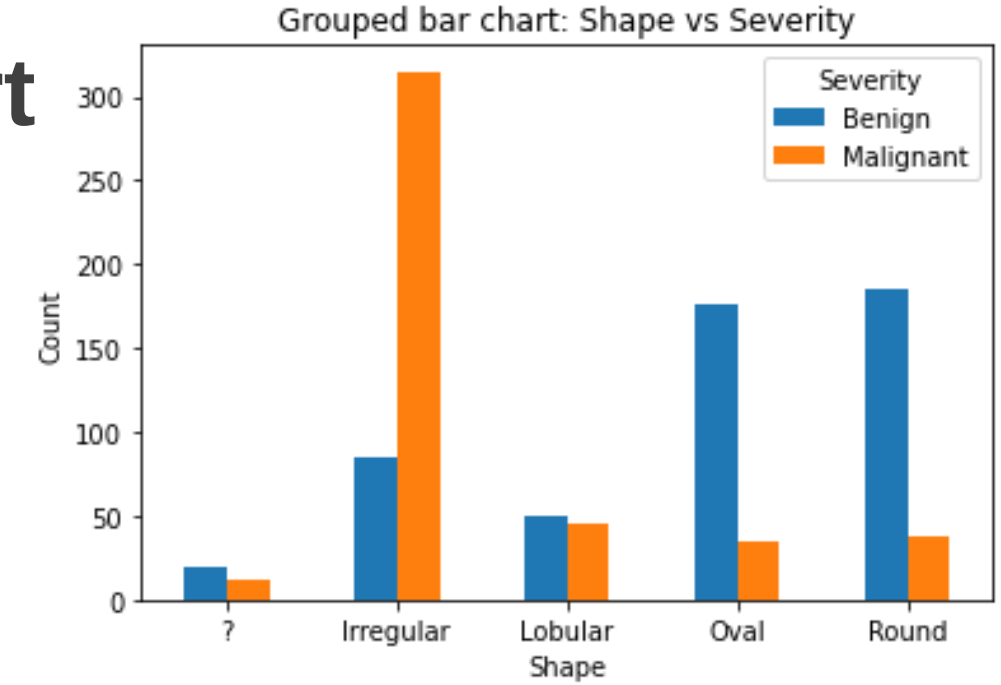
Amounts

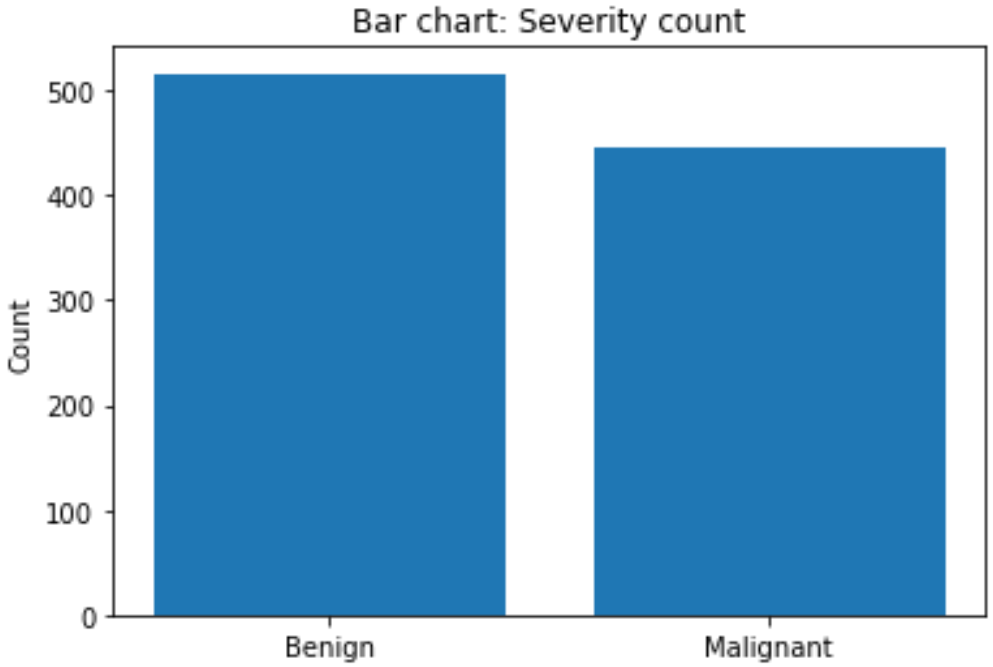
Bar Chart

- **Purpose**: to visualize data that can be clustered into discrete categories or groups, and to compare the magnitude of different variables or categories.
- **Common type of data**: categorical or discrete data.
- **Interpretation**: to analyze the relative magnitudes of the different categories or groups, to identify trends, patterns, and changes in the data, and to compare multiple variables or data series.
- **Examples of applications**: to visualize the sales levels of different product categories, survey results, demographic data, financial data, etc.

Bar Chart


Bar chart: (Shape & Severity) vs Count
Count
0
50
100
150
200
250
300
?, Benign
?, Malignant
Irregular, Benign
Irregular, Malignant
Lobular, Benign
Lobular, Malignant
Oval, Benign
Oval, Malignant
Round, Benign
Round, Malignant
Shape & Severity


Grouped bar chart: Shape vs Severity
Severity
Benign
Malignant
Count
0
50
100
150
200
250
300
?
Irregular
Lobular
Oval
Round
Shape


Bar chart: Severity count
Count
0
100
200
300
400
500
Benign
Malignant

Bar Charts with AI

Generate the following visuals for the mammographic mass dataset:

1) A bar chart having the bars combining variables 'shape' and 'severity'.

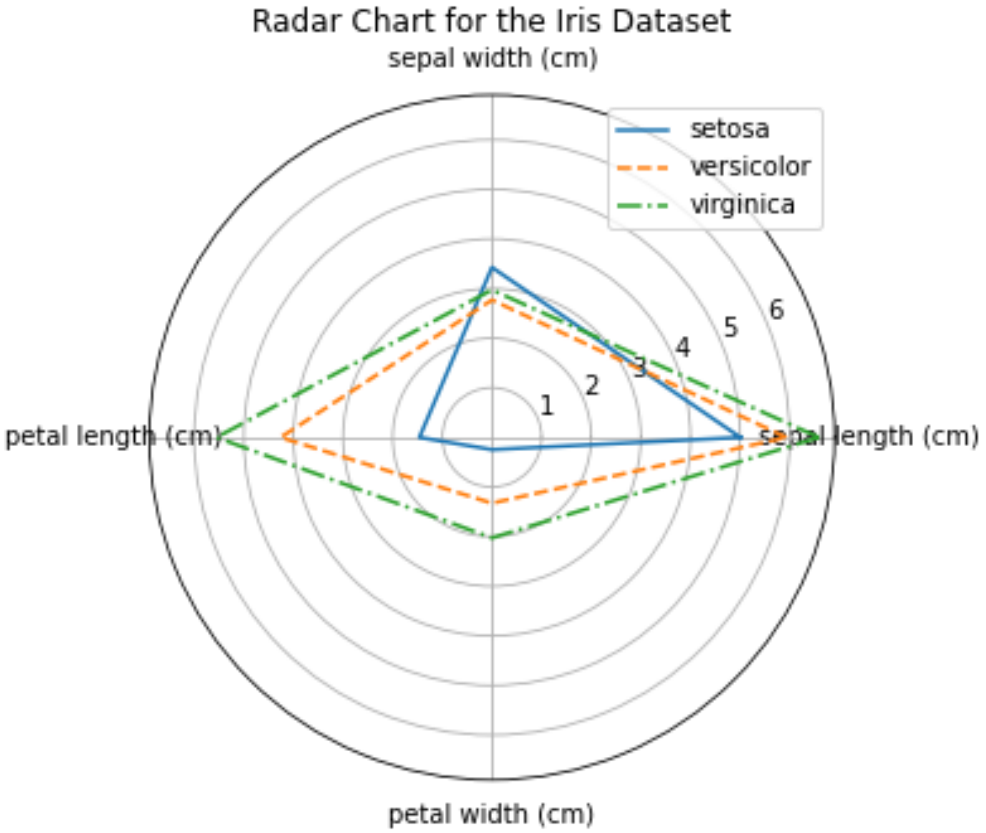
2) A grouped bar chart for variable 'shape' by 'severity'.

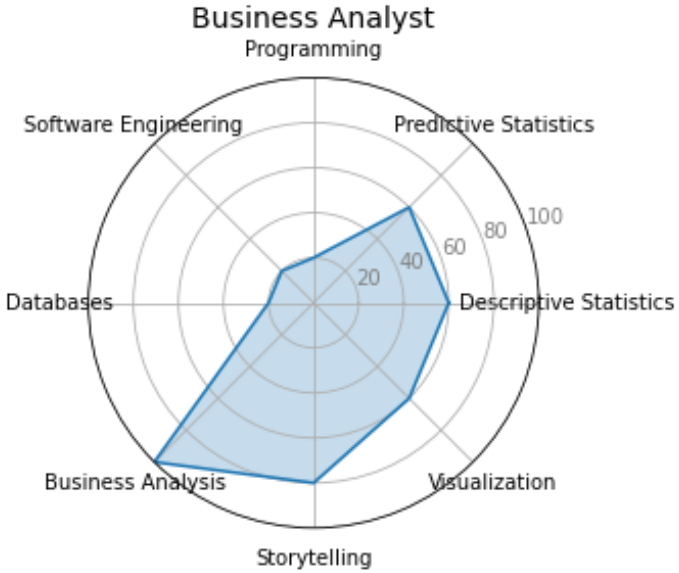
3) A bar chart with the count of each category for variable 'severity'.

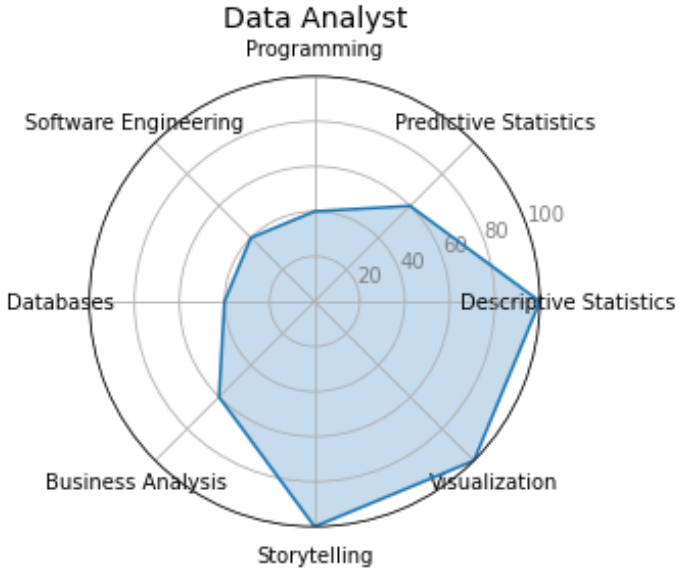
Radar Chart

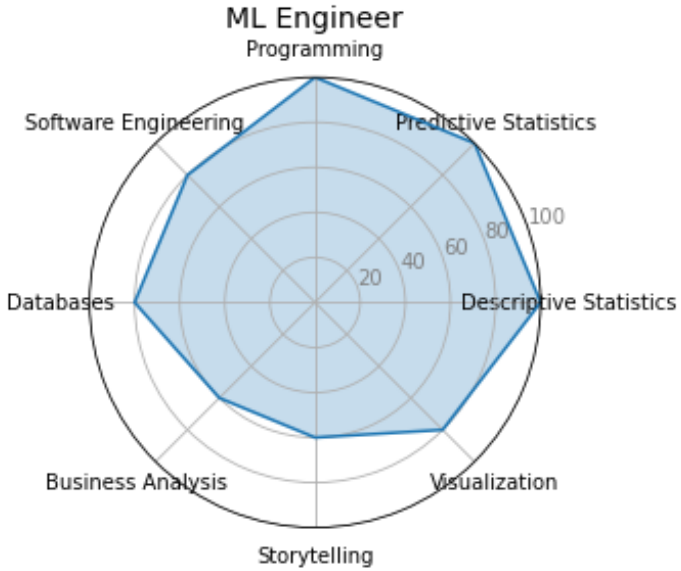
- **Purpose**: to compare and visualize multiple series of continuous variables in a single radial chart.
- **Common type of data**: continuous data.
- **Interpretation**: the radar chart plots one or more series of values considering numerical variables over a radial graph, forming a polygon. Larger values are farther from the radar center, allowing the comparison of different series in terms of amplitude. The shape and size of the polygon can be used to compare the different series.
- **Examples of applications**: to compare brands considering their features, stocks in relation to their (fundamental) indicators, students in relation to their grades in each subject, animals in relation to their characteristics, etc.

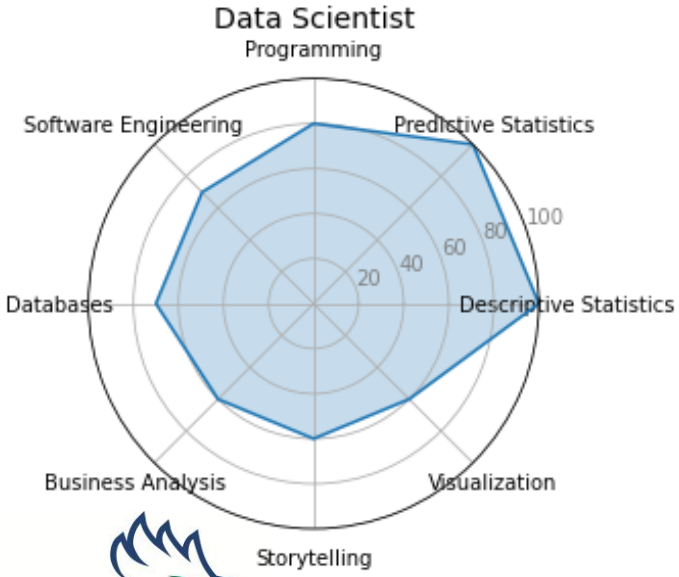
Radar Chart

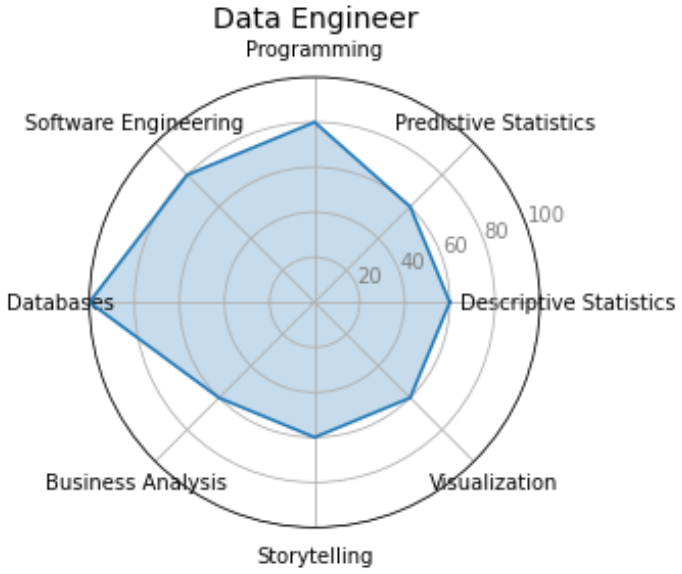

Radar Chart for the Iris Dataset
sepal width (cm)
setosa
versicolor
virginica
1
2
3
4
5
6
petal length (cm)
sepal length (cm)
petal width (cm)

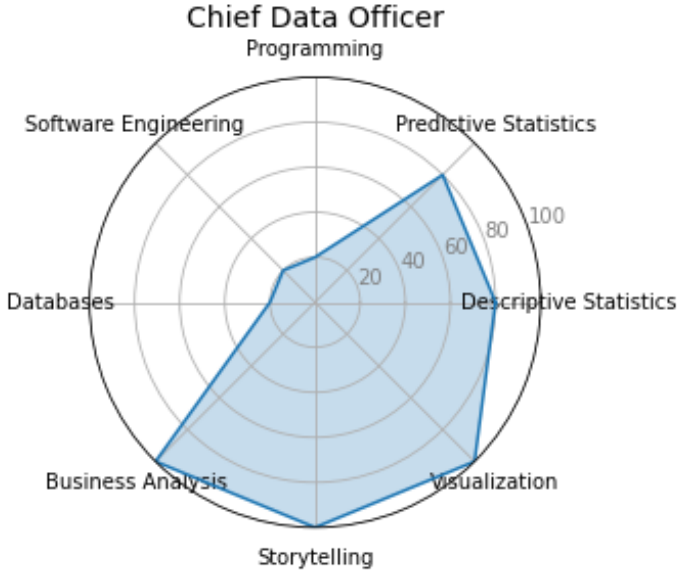

Business Analyst
Programming
Predictive Statistics
Descriptive Statistics
Visualization
Storytelling
Business Analysis
Databases
Software Engineering
20
40
60
80
100


Data Analyst
Programming
Predictive Statistics
Descriptive Statistics
Visualization
Storytelling
Business Analysis
Databases
Software Engineering
20
40
60
80
100


ML Engineer
Programming
Predictive Statistics
Descriptive Statistics
Visualization
Storytelling
Business Analysis
Databases
Software Engineering
20
40
60
80
100


Data Scientist
Programming
Predictive Statistics
Descriptive Statistics
Visualization
Storytelling
Business Analysis
Databases
Software Engineering
20
40
60
80
100


Data Engineer
Programming
Predictive Statistics
Descriptive Statistics
Visualization
Storytelling
Business Analysis
Databases
Software Engineering
20
40
60
80
100

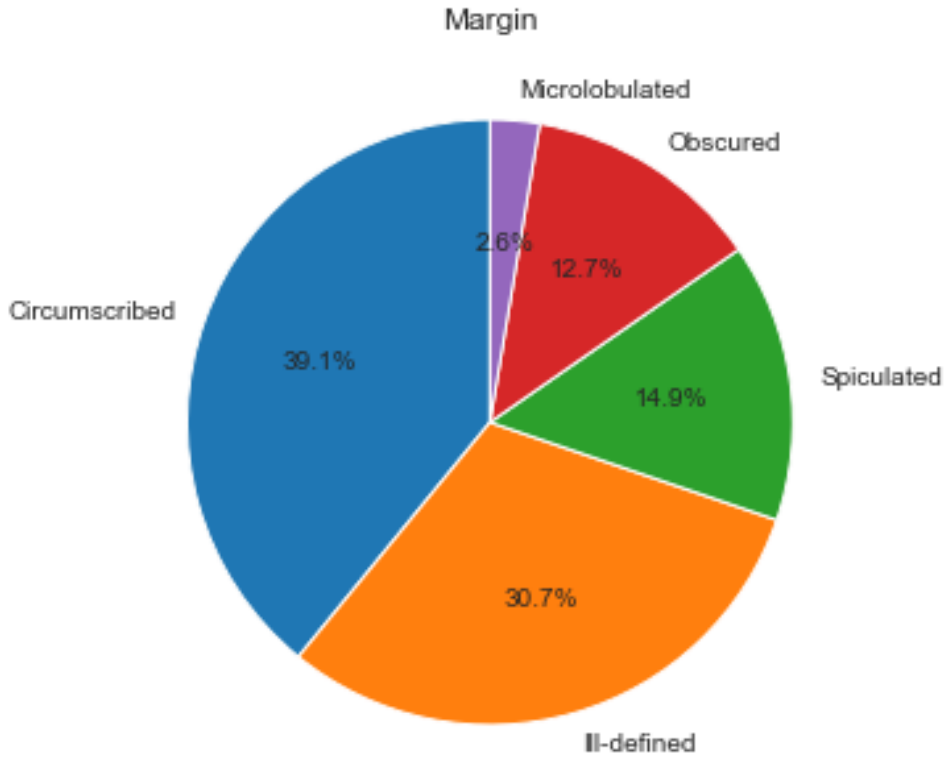

Chief Data Officer
Programming
Predictive Statistics
Descriptive Statistics
Visualization
Storytelling
Business Analysis
Databases
Software Engineering
20
40
60
80
100

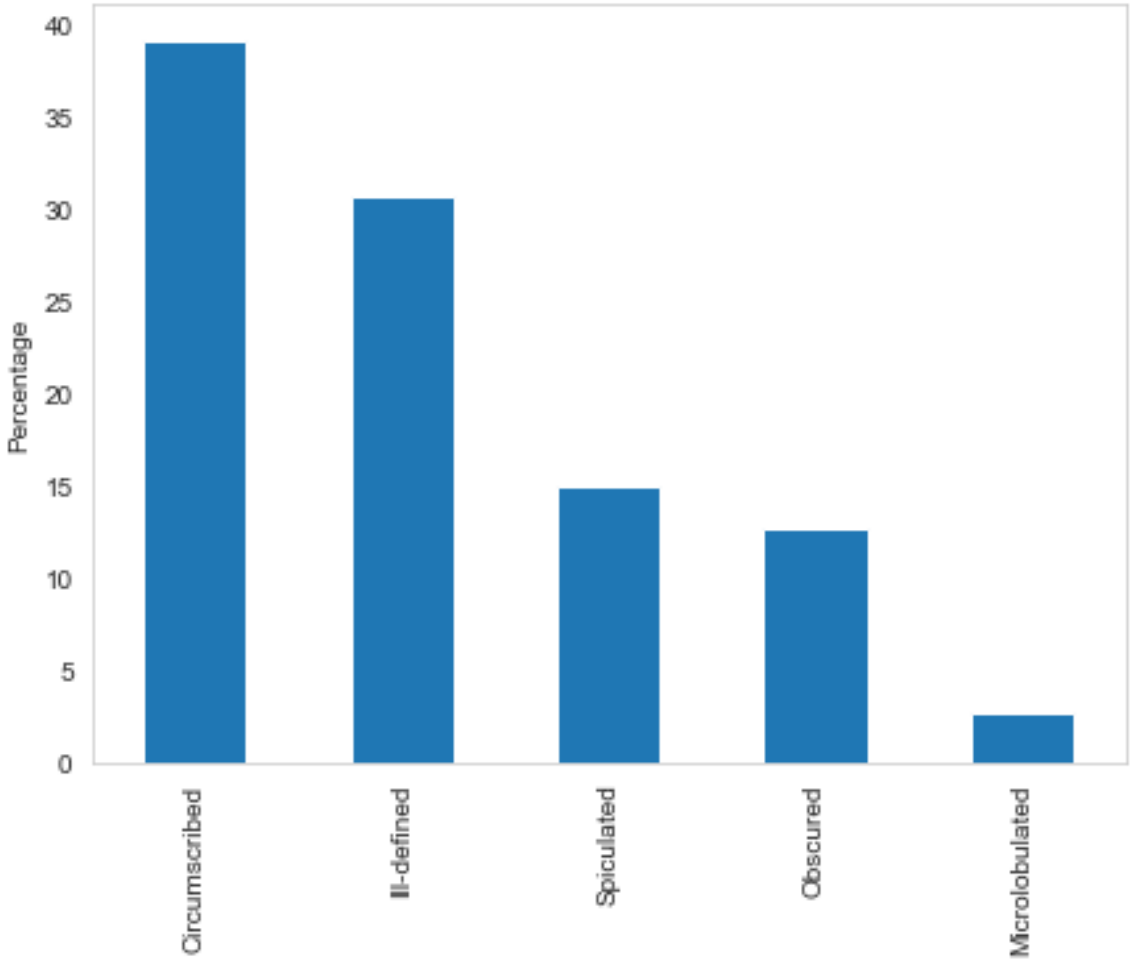
Proportions

Pie Chart

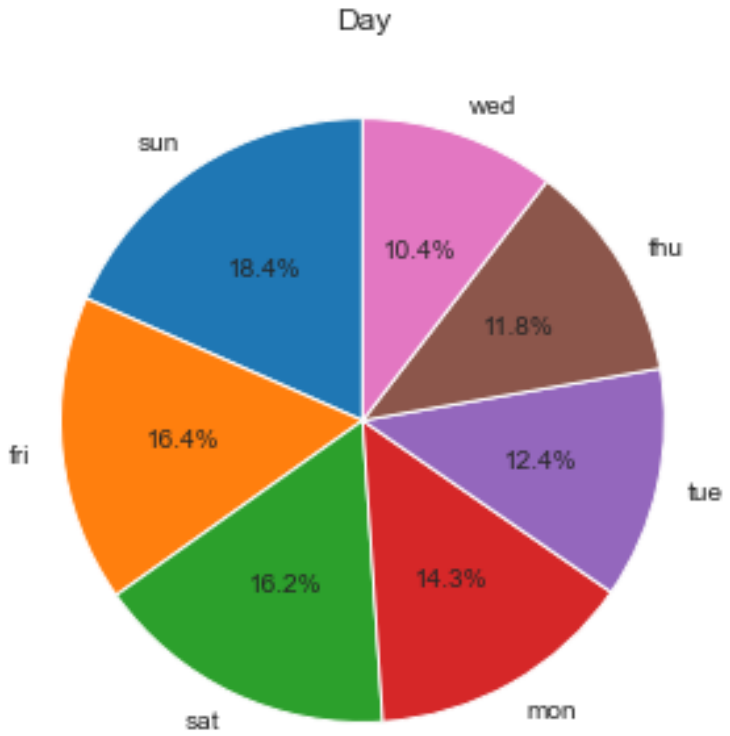
- **Purpose**: to explore and visualize the distribution of a variable that can be divided into a small number of different and mutually exclusive categories, where each category is a fraction (slice) of the whole.
- **Common type of data**: categorical data.
- **Interpretation**: each pie chart slice represents a category within the whole, and the larger the slice, the greater its value (proportion).
- **Examples of applications**: to visualize the proportion of any categorical variable, such as colors in a painting, customers who buy each product type, marital status in a given area, etc.

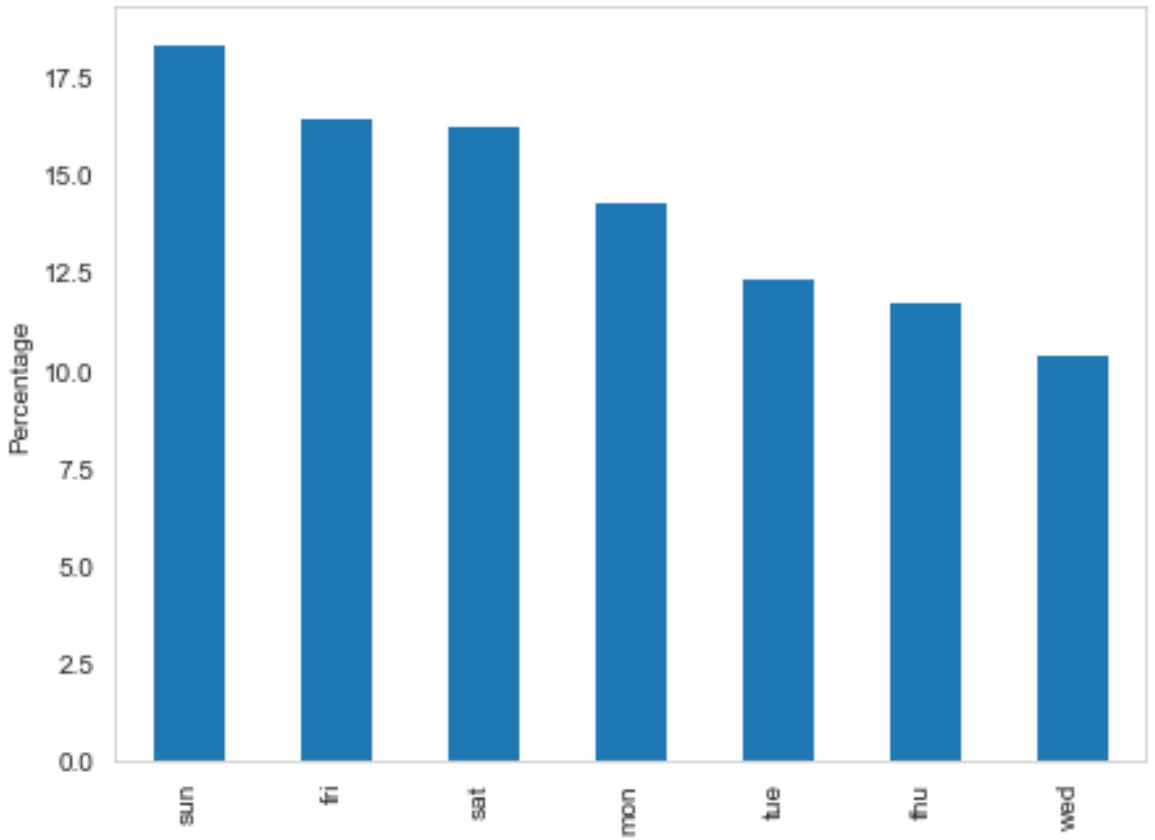
Pie Chart


Margin
Microlobulated
Obscured
Circumscribed
Spiculated
Ill-defined
2.6%
12.7%
39.1%
14.9%
30.7%

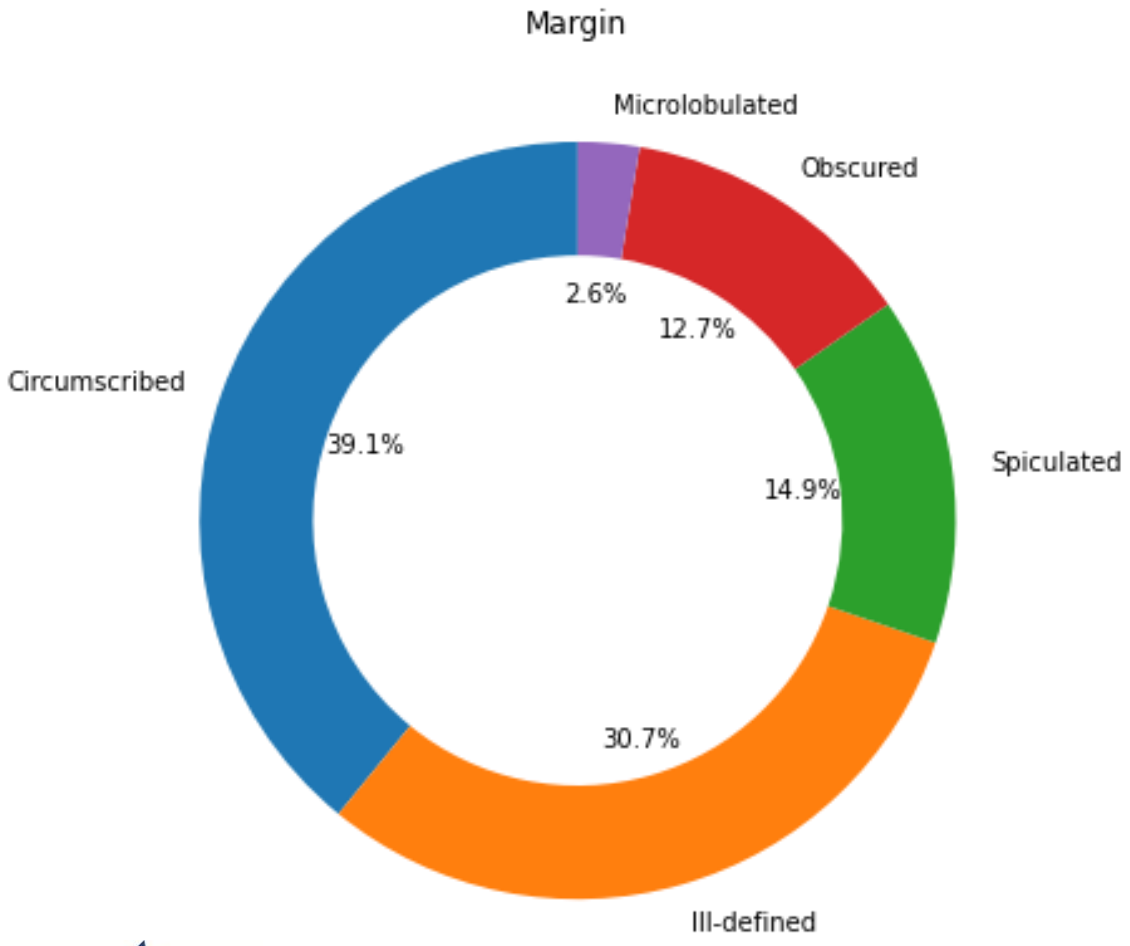

Percentage
0
5
10
15
20
25
30
35
40
Circumscribed
Ill-defined
Spiculated
Obscured
Microlobulated

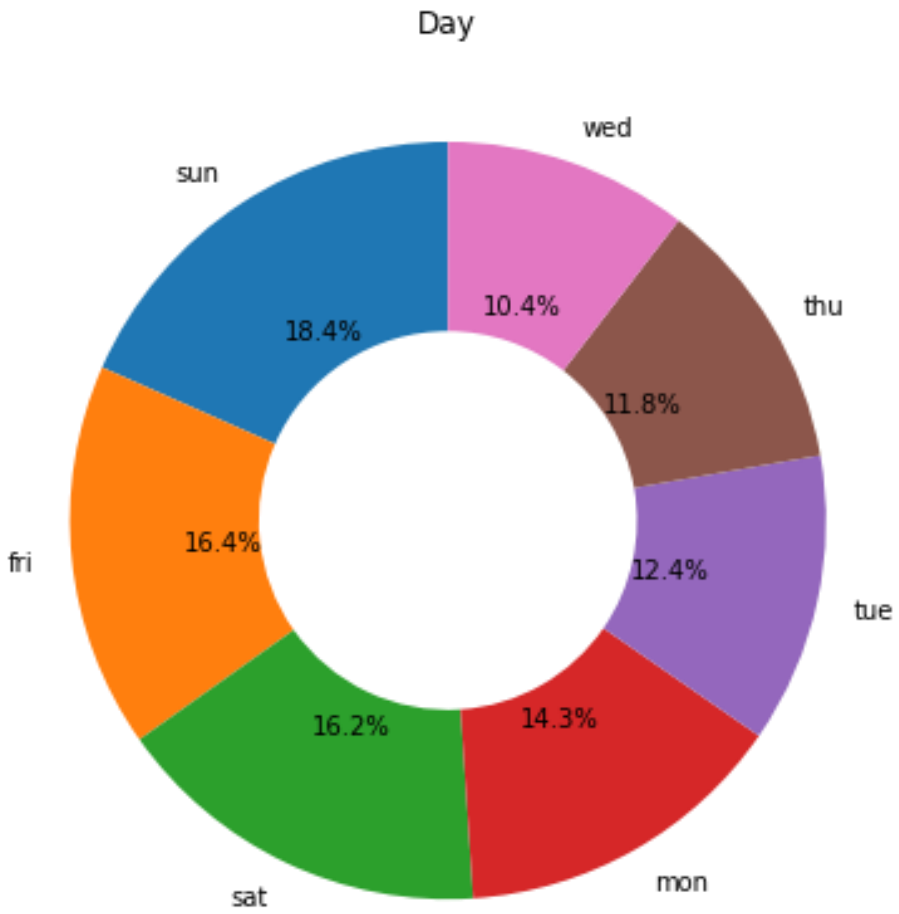
Pie Chart


Day
wed
sun
thu
18.4%
10.4%
11.8%
fri
16.4%
12.4%
tue
16.2%
14.3%
sat
mon


Percentage
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
sun
fri
sat
mon
tue
thu
wed

Doughnut Chart

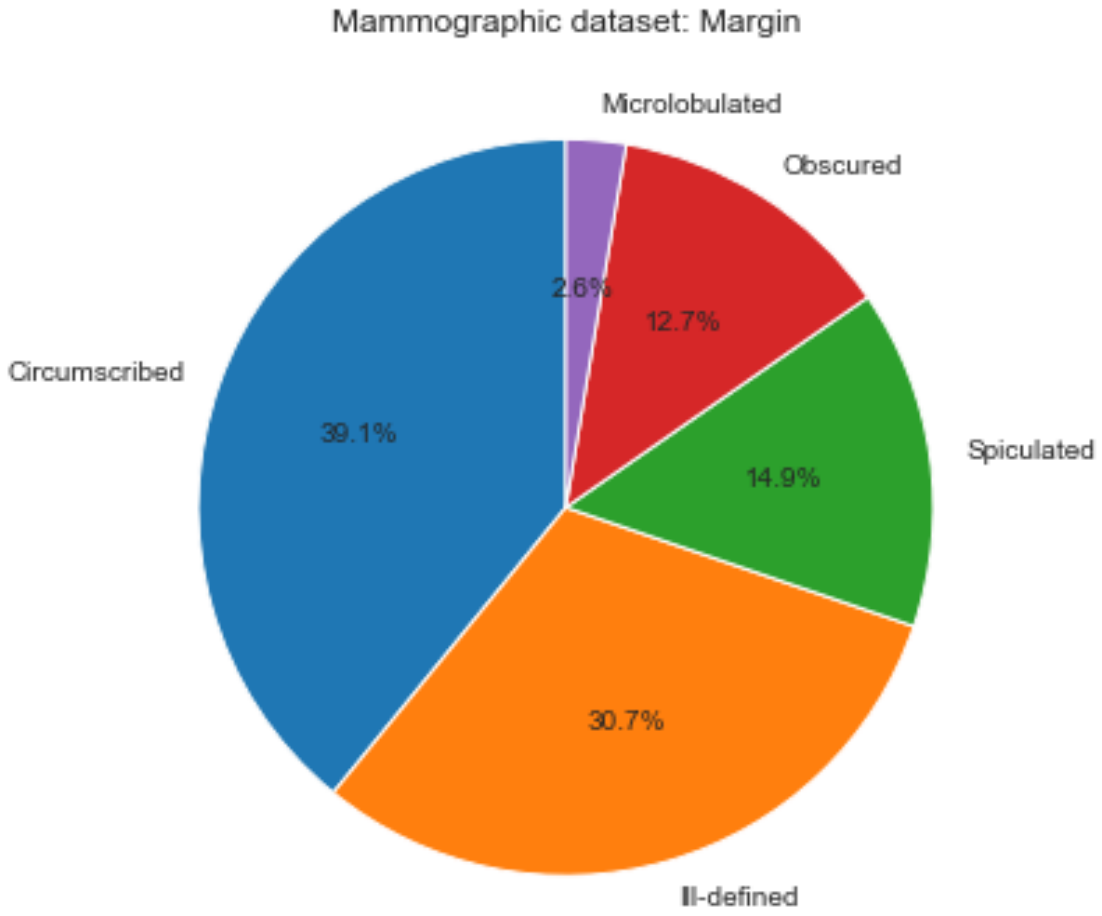


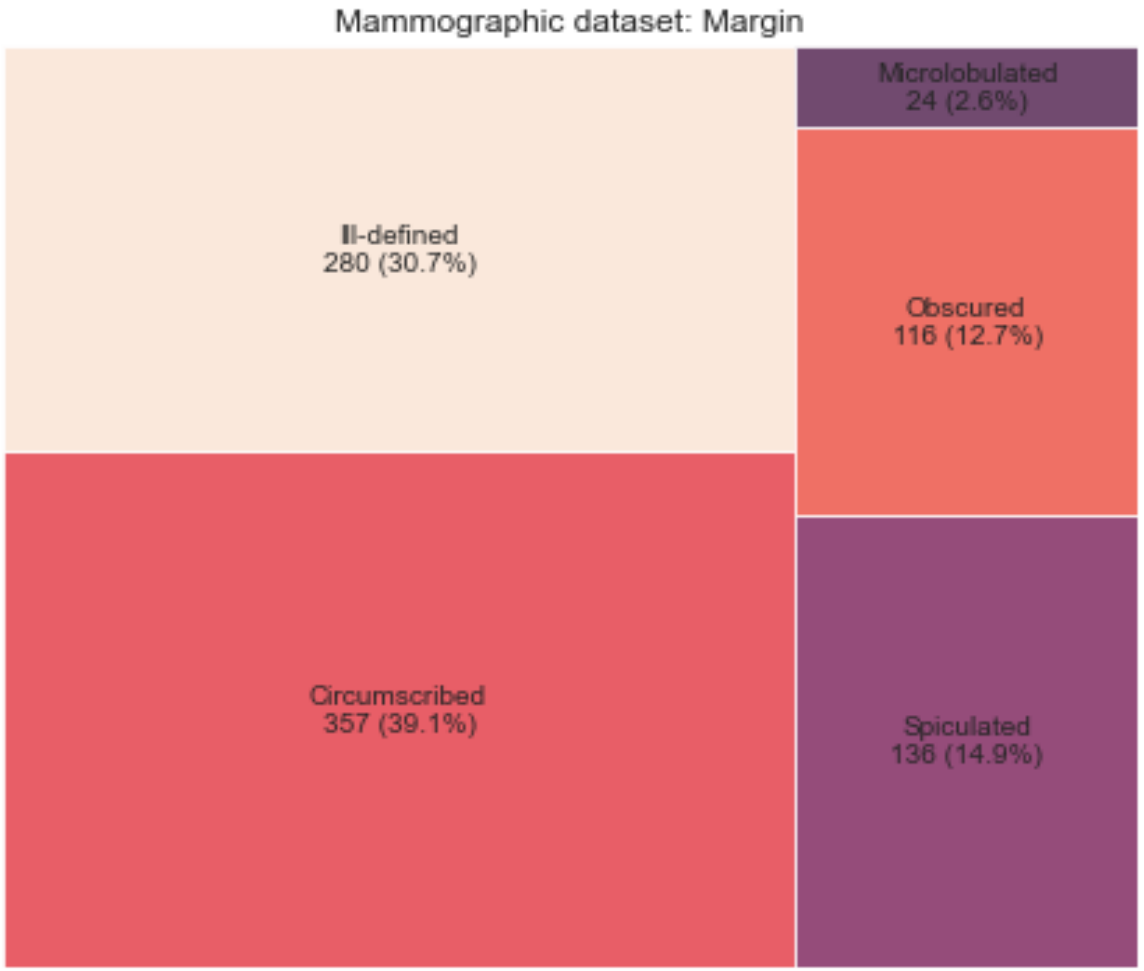


Treemap

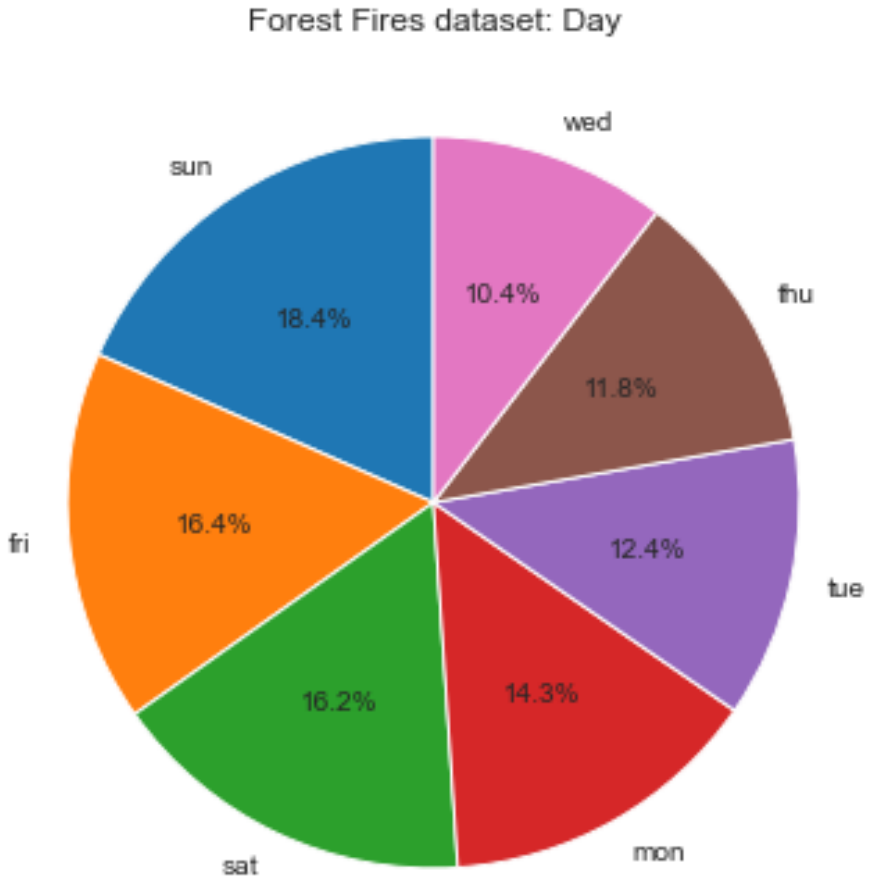
- **Purpose**: to explore and visualize hierarchical data in nested and mutually exclusive rectangles of varying sizes, where each category is a fraction (slice) of the whole.
- **Common type of data**: data with a hierarchical structure and categories that can be divided into subcategories.
- **Interpretation**: each rectangle represents a category within the whole, and the larger the rectangle, the greater its value (proportion). Color can be used to represent other values or variables.
- **Examples of applications**: to visualize the proportion of hierarchical variables, such as continents, countries, and population; month, day, and rain volume; etc.

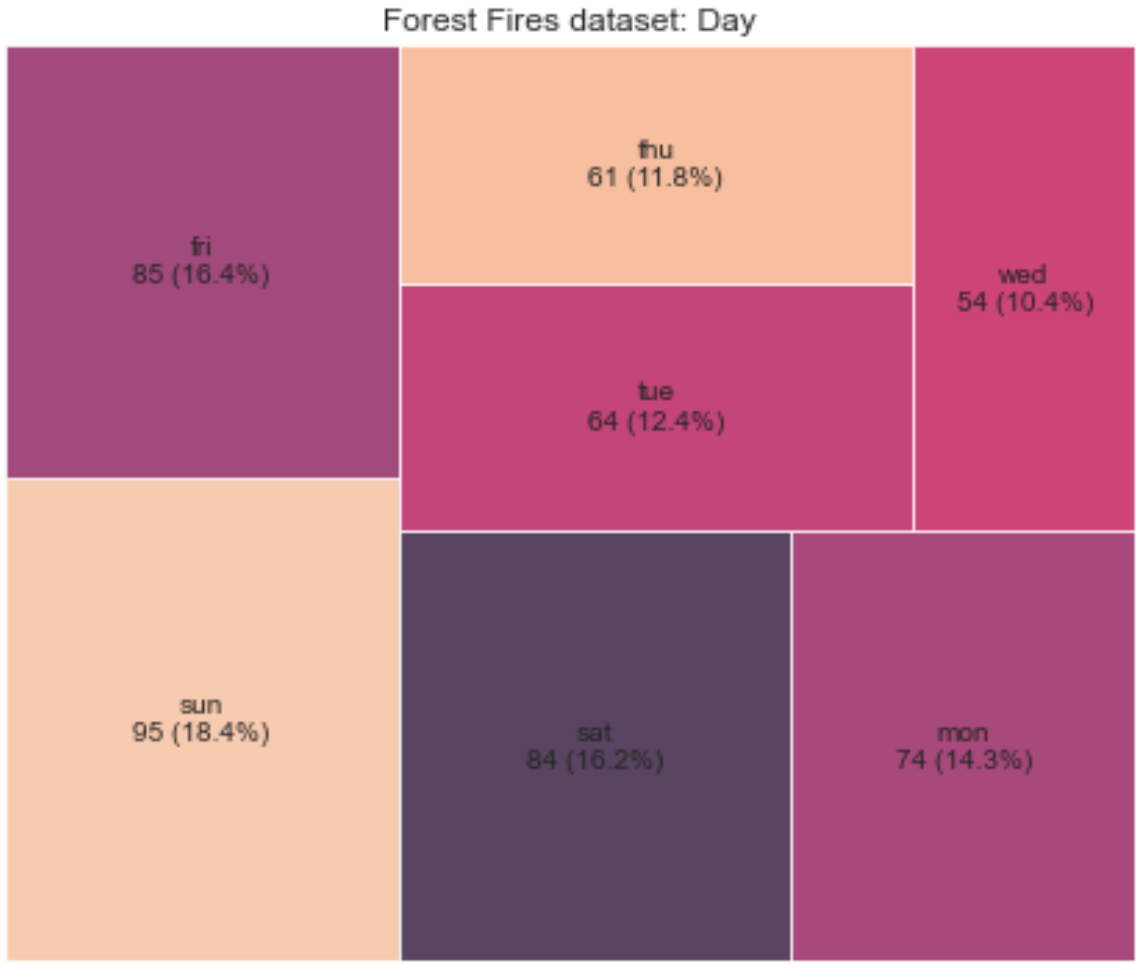
Treemap


Mammographic dataset: Margin
Microlobulated
Obscured
Circumscribed
Spiculated
Ill-defined
2.6%
12.7%
39.1%
14.9%
30.7%

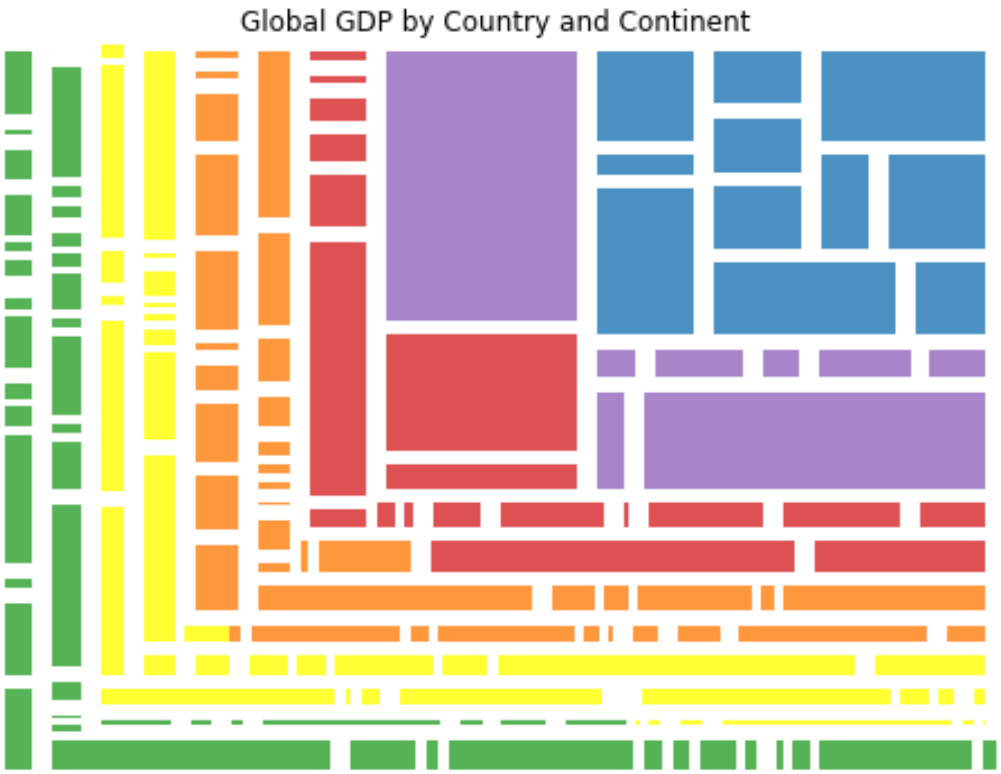

Mammographic dataset: Margin
Ill-defined
280 (30.7%)
Microlobulated
24 (2.6%)
Obscured
116 (12.7%)
Circumscribed
357 (39.1%)
Spiculated
136 (14.9%)

Treemap

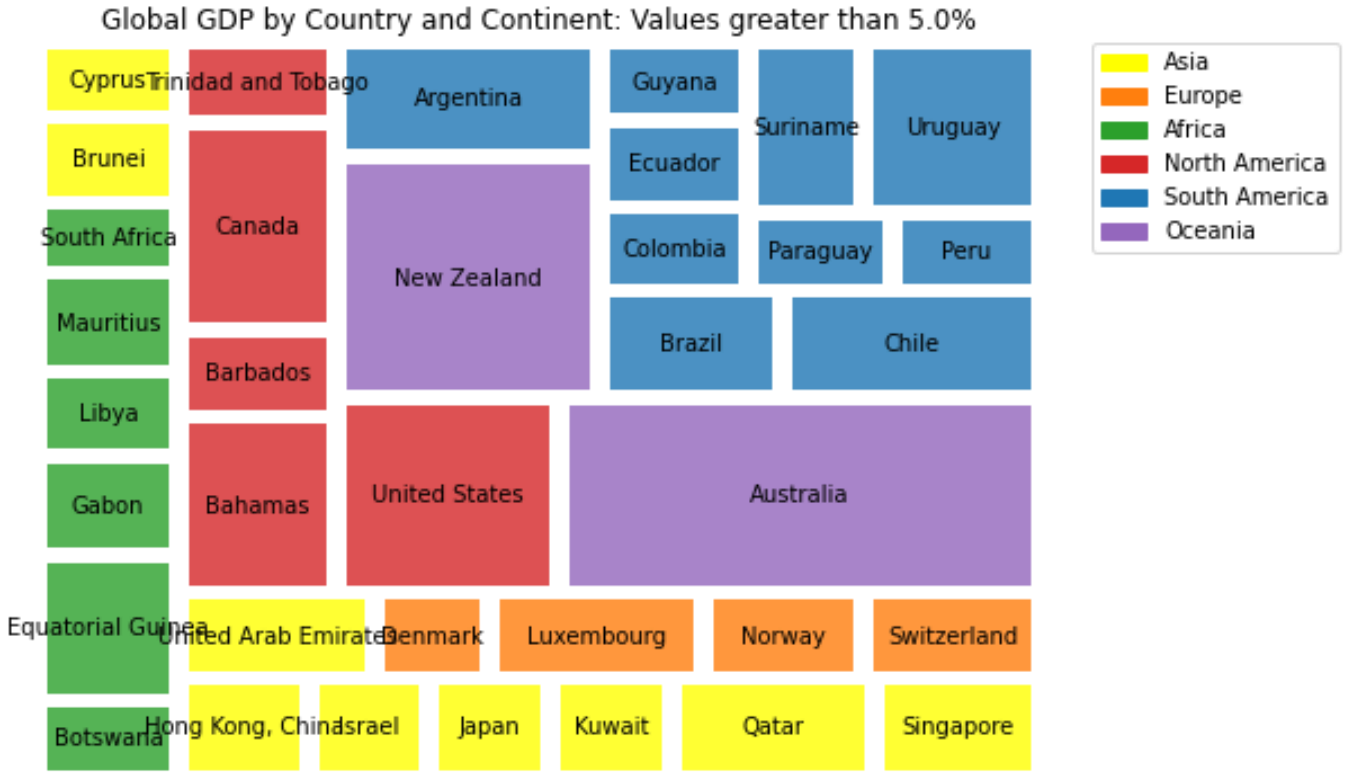

Forest Fires dataset: Day
wed
sun
thu
10.4%
18.4%
11.8%
fri
16.4%
12.4%
tue
16.2%
14.3%
sat
mon


Forest Fires dataset: Day
fri
85 (16.4%)
thu
61 (11.8%)
wed
54 (10.4%)
tue
64 (12.4%)
sun
95 (18.4%)
sat
84 (16.2%)
mon
74 (14.3%)

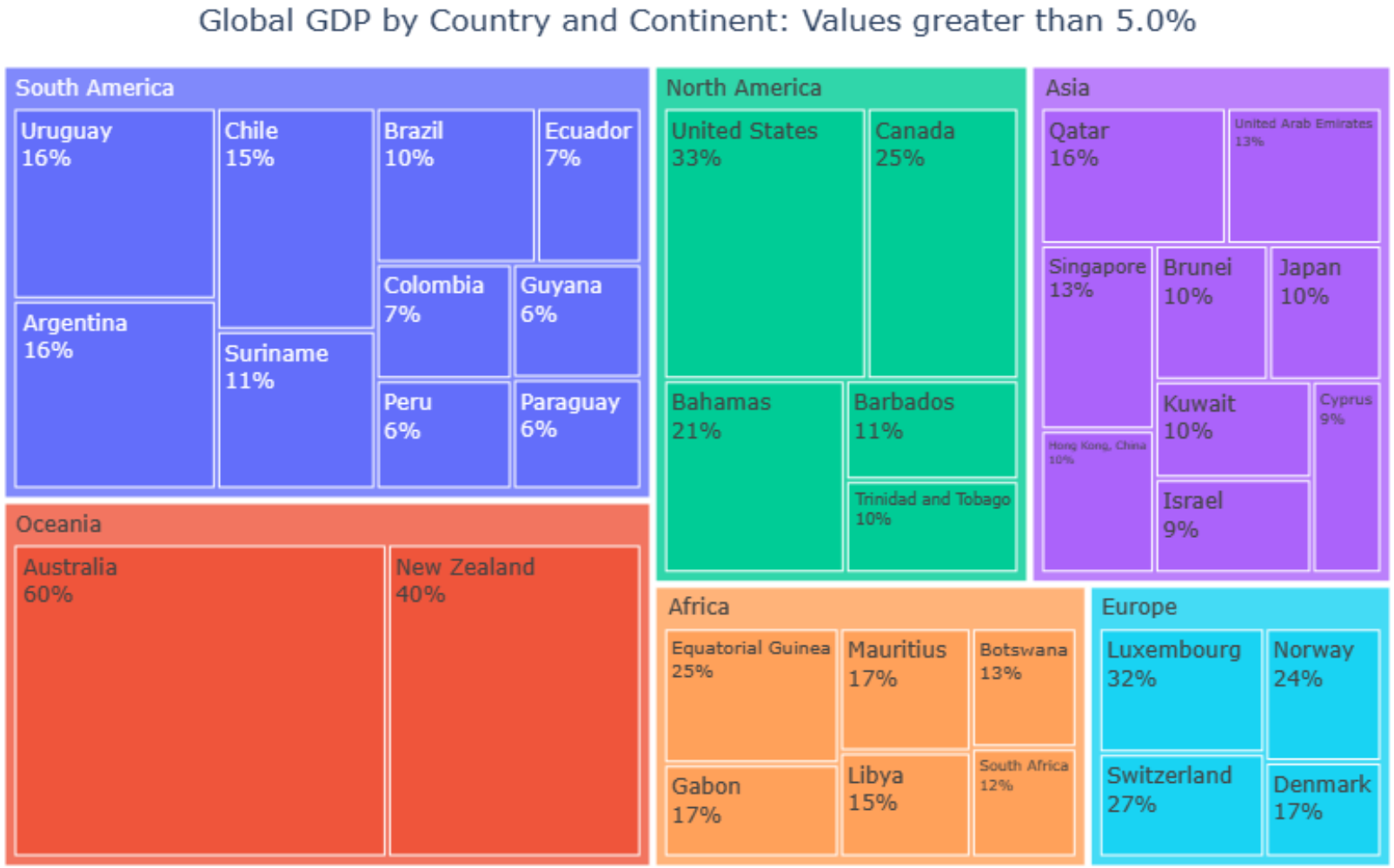
Treemap


Global GDP by Country and Continent


Asia
Europe
Africa
North America
South America
Oceania


Global GDP by Country and Continent: Values greater than 5.0%
Cyprus
Trinidad and Tobago
Argentina
Guyana
Suriname
Uruguay
Brunei
Ecuador
South Africa
Canada
New Zealand
Colombia
Paraguay
Peru
Mauritius
Brazil
Chile
Barbados
Libya
Gabon
Bahamas
United States
Australia
Equatorial Guinea
United Arab Emirates
Denmark
Luxembourg
Norway
Switzerland
Botswana
Hong Kong, China
Israel
Japan
Kuwait
Qatar
Singapore
Asia
Europe
Africa
North America
South America
Oceania

Treemap


Global GDP by Country and Continent: Values greater than 5.0%
South America
Uruguay 16%
Chile 15%
Brazil 10%
Ecuador 7%
Argentina 16%
Suriname 11%
Colombia 7%
Guyana 6%
Peru 6%
Paraguay 6%
North America
United States 33%
Canada 25%
Bahamas 21%
Barbados 11%
Trinidad and Tobago 10%
Asia
Qatar 16%
United Arab Emirates 13%
Singapore 13%
Brunei 10%
Japan 10%
Hong Kong, China 10%
Kuwait 10%
Cyprus 9%
Israel 9%
Oceania
Australia 60%
New Zealand 40%
Africa
Equatorial Guinea 25%
Mauritius 17%
Botswana 13%
Gabon 17%
Libya 15%
South Africa 12%
Europe
Luxembourg 32%
Norway 24%
Switzerland 27%
Denmark 17%

Pie and Doughnut Charts, and Treemap

“Plot a pie chart, a doughnut chart, and a bar chart for variable 'margin' of the mammographic dataset.”

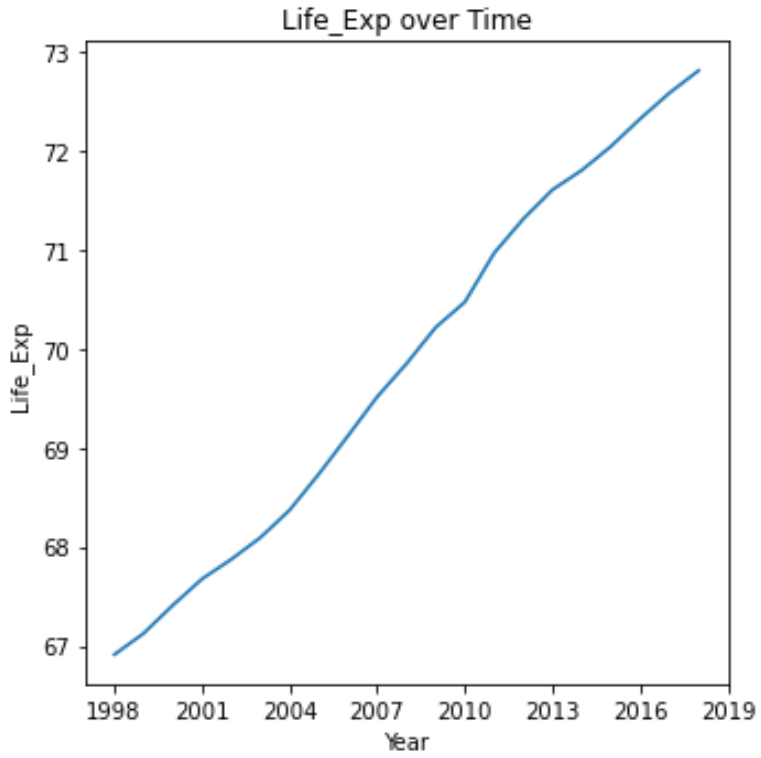
“For the attached gapminder_data_graphs dataset plot a treemap of the global GDP by Country and Continent for values greater than 5% per continent.”

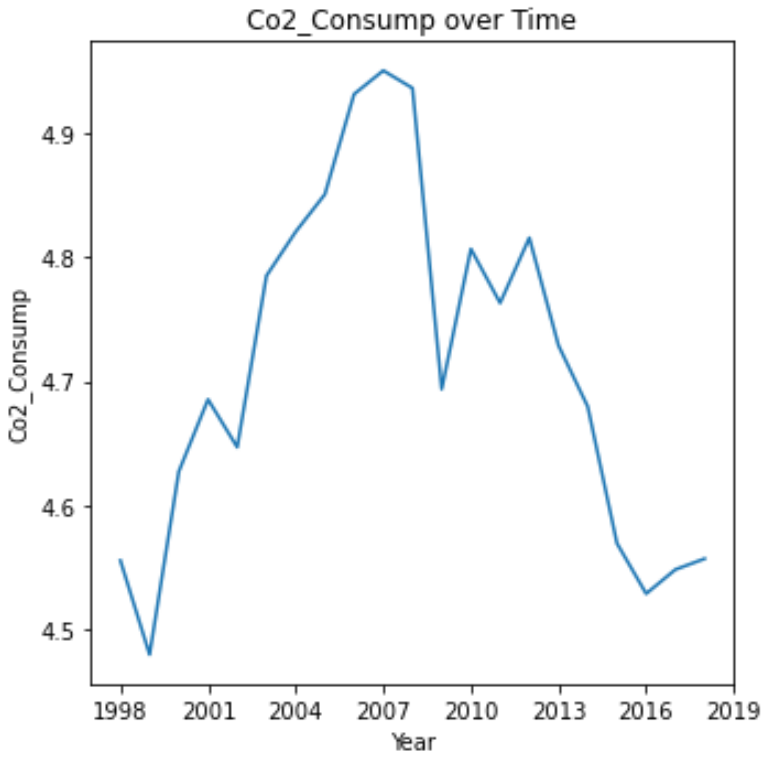
Evolution and Flow

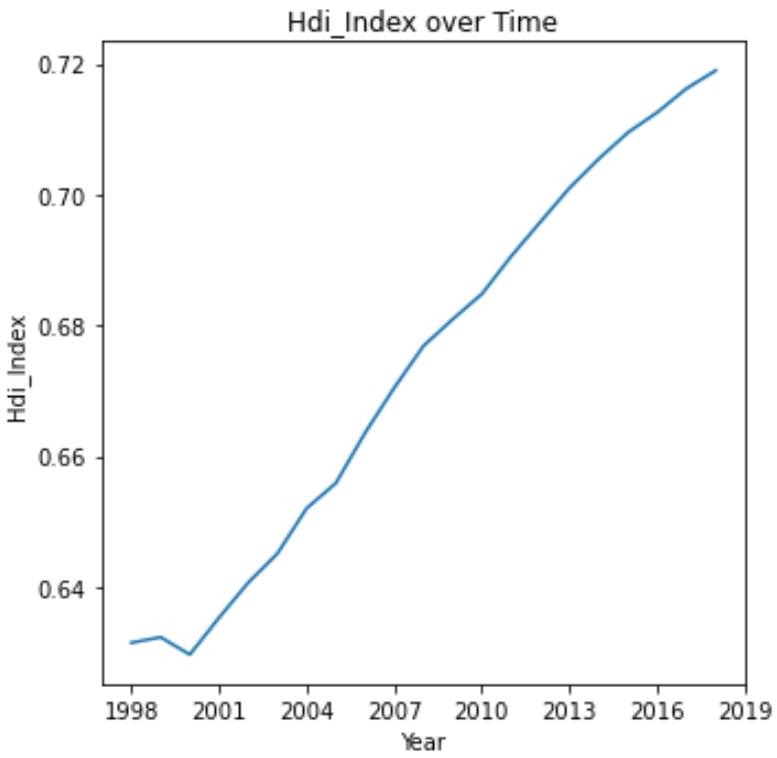
Line Chart

- **Purpose**: to show patterns, trends, changes, and anomalies in data over time or a strictly increasing variable (quantity).
- **Common type of data**: continuous one-dimensional sequences.
- **Interpretation**: to identify the trend, changes in direction or magnitude of the trend, looking for spikes or dips, identifying repeating patterns or cycles, etc.
- **Examples of applications**: to visualize stock prices, weather data, population trends, disease spread, etc.

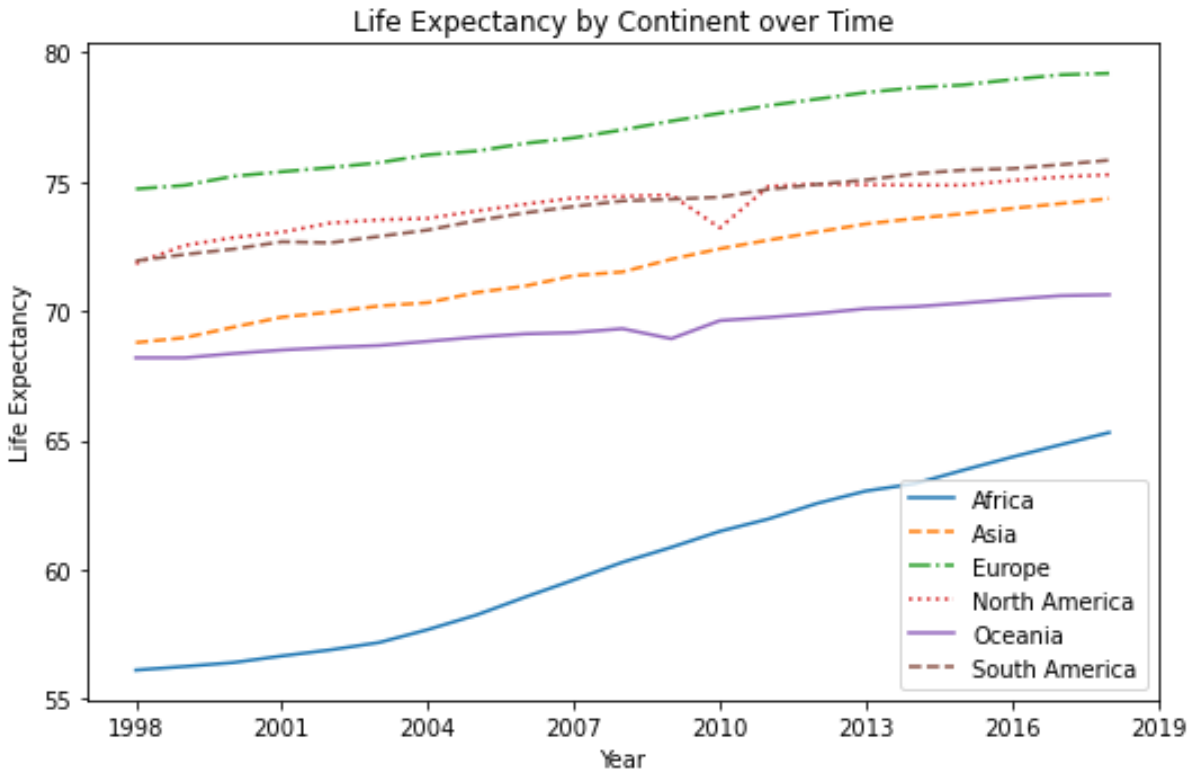
Line Chart

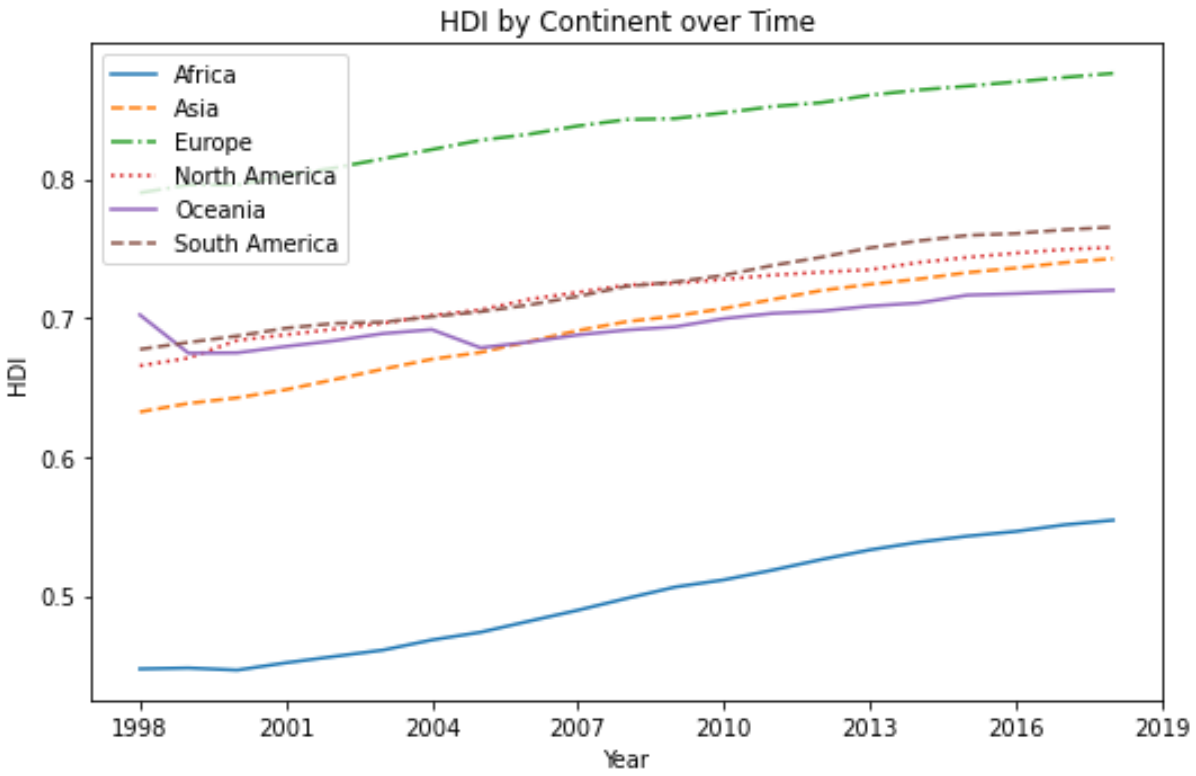

Life_Exp over Time
Life_Exp
Year
73
72
71
70
69
68
67
1998
2001
2004
2007
2010
2013
2016
2019


Co2_Consump over Time
Co2_Consump
Year
4.9
4.8
4.7
4.6
4.5
1998
2001
2004
2007
2010
2013
2016
2019


Hdi_Index over Time
Hdi_Index
Year
0.72
0.70
0.68
0.66
0.64
1998
2001
2004
2007
2010
2013
2016
2019

Line Chart

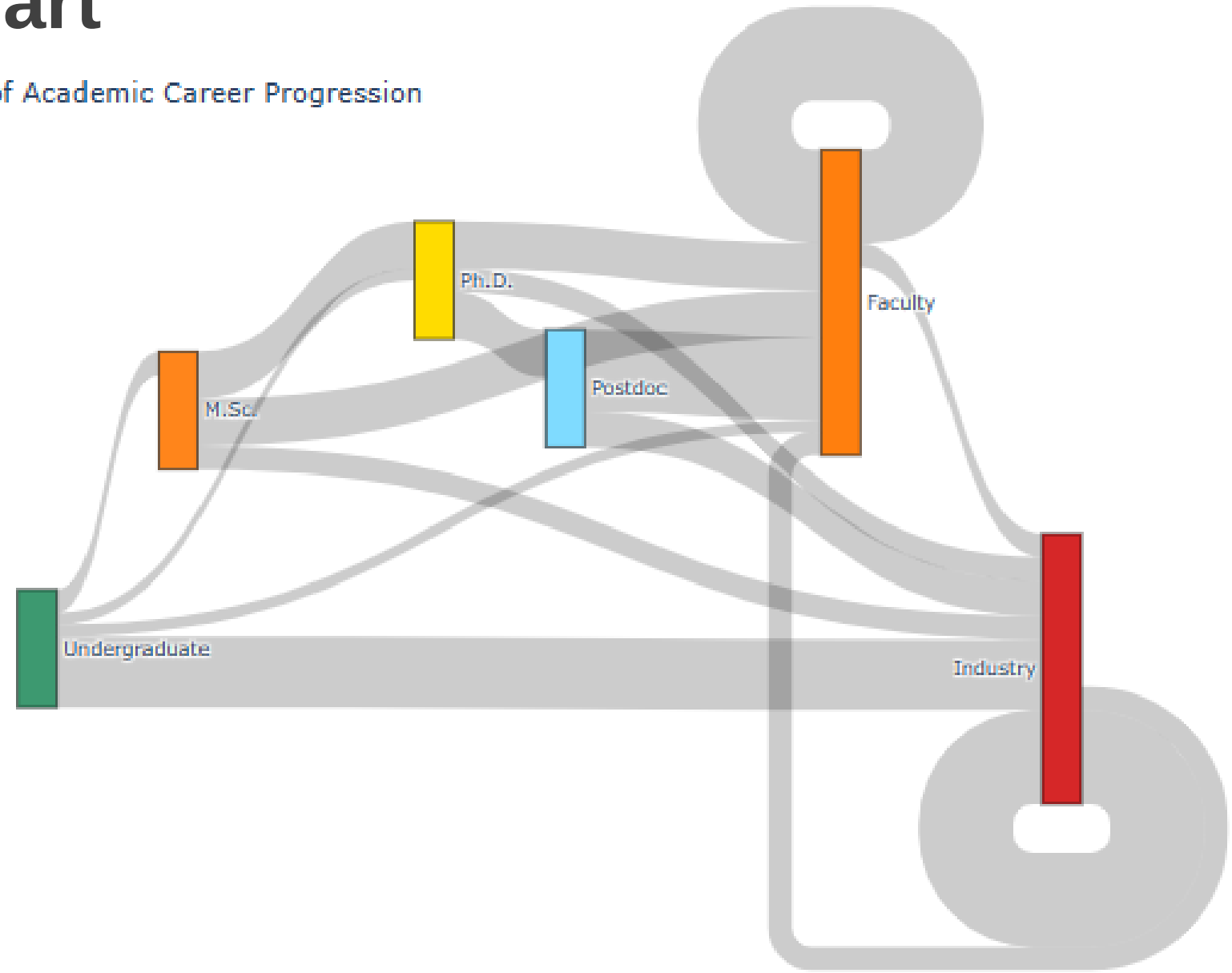

Life Expectancy by Continent over Time
Life Expectancy
Year
80
75
70
65
60
55
1998
2001
2004
2007
2010
2013
2016
2019
Africa
Asia
Europe
North America
Oceania
South America


HDI by Continent over Time
HDI
Year
0.8
0.7
0.6
0.5
1998
2001
2004
2007
2010
2013
2016
2019
Africa
Asia
Europe
North America
Oceania
South America

Sankey Chart

- **Purpose**: to compare and visualize the flow of data through various stages or transitions of a process.
- **Common type of data**: categorical and numerical (ordinal and continuous) data.
- **Interpretation**: the Sankey chart involves nodes and flows that correspond to the data and its amount being moved over the flow. Larger (wider) flows mean larger amounts.
- **Examples of applications**: the flow of energy in a power grid, the flow of people over airports or countries, the flow of goods and services in a supply chain, the flow of cars on a highway, the flow of students over the years, etc.

Sankey Chart



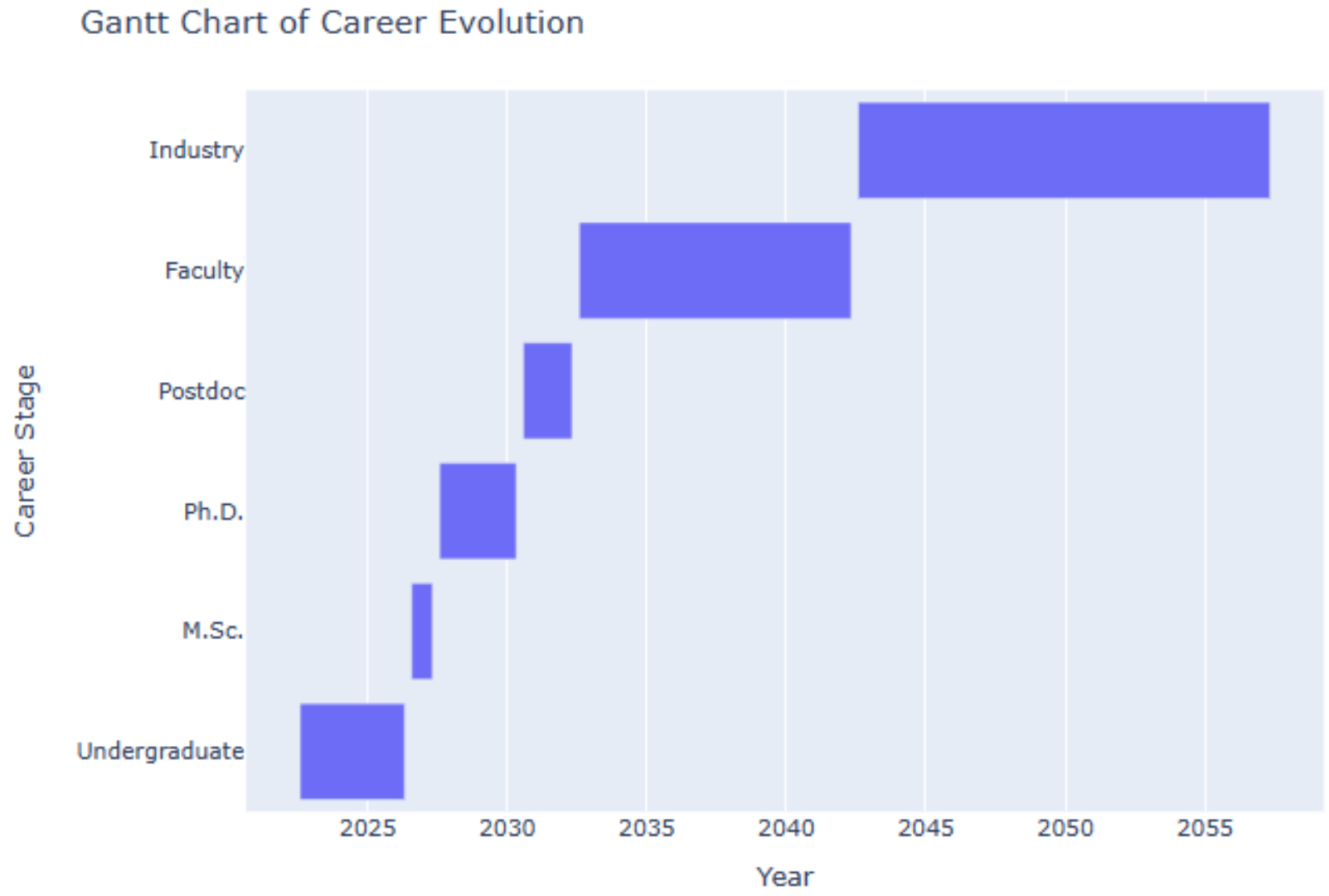
Gantt Chart

- **Purpose**: to visualize and explore the timeline of tasks, their durations, and dependencies over time.
- **Common type of data**: the tasks are normally described by nominal variables, whilst start dates, end dates, and durations can be described in hours, days, months, years, or any other time-related numeric variable. Other qualitative variables can be added, such as dependencies, milestones, resources, etc.
- **Interpretation**: horizontal bars represent the tasks, their lengths correspond to the duration of the tasks, their initial position are the initial time of the task, and their end position are the end time of the task. Connecting lines or arrows can be used to show dependencies among tasks, and aggregated bars represent subtasks.

Gantt Chart

- **Examples of applications**: used mainly to visualize project schedules for project management purposes. It allows progress monitoring, the identification of bottlenecks and delays, resource allocation, and decision making.

Gantt Charts



Line, Sankey, and Gantt Charts with AI

“Generate a line chart with the 'life expectancy' by 'continent' over time for the Gapminder dataset.”

“*Create a Sankey chart for a potential career path at an american university*”

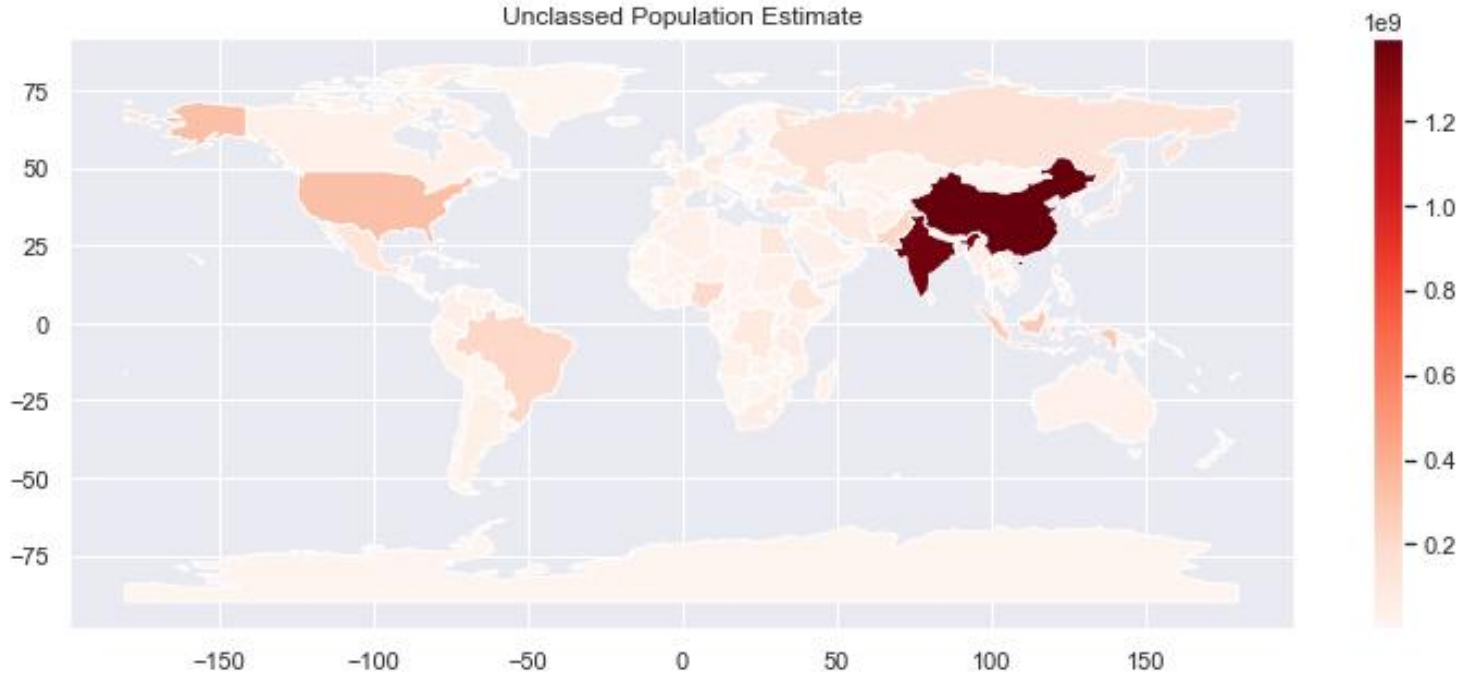
“*Create a Gantt chart for a hypothetical project on the building of a house*”

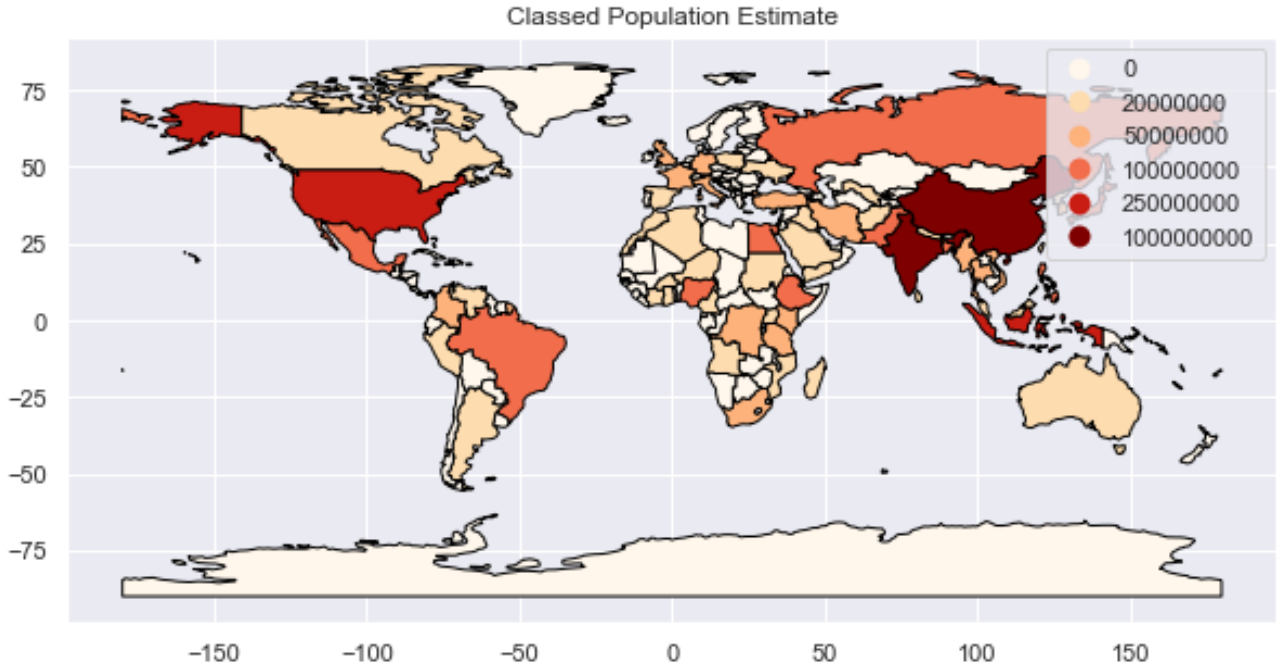
Geospatial

Choropleth Map

- **Purpose**: to explore and visualize spatial data in an area map.
- **Common type of data**: spatial data, that is, data containing different values or categories associated with specific (geographic) regions.
- **Interpretation**: the map involves coloring or shading the regions such that darker colors usually indicate higher values, whilst lighter colors indicate smaller values.
- **Examples of applications**: to visualize spatial data, such as population density, disease spread, electricity consumption, weather data, political preferences, employment rates, economic indices (e.g., GDP and HDI), etc.

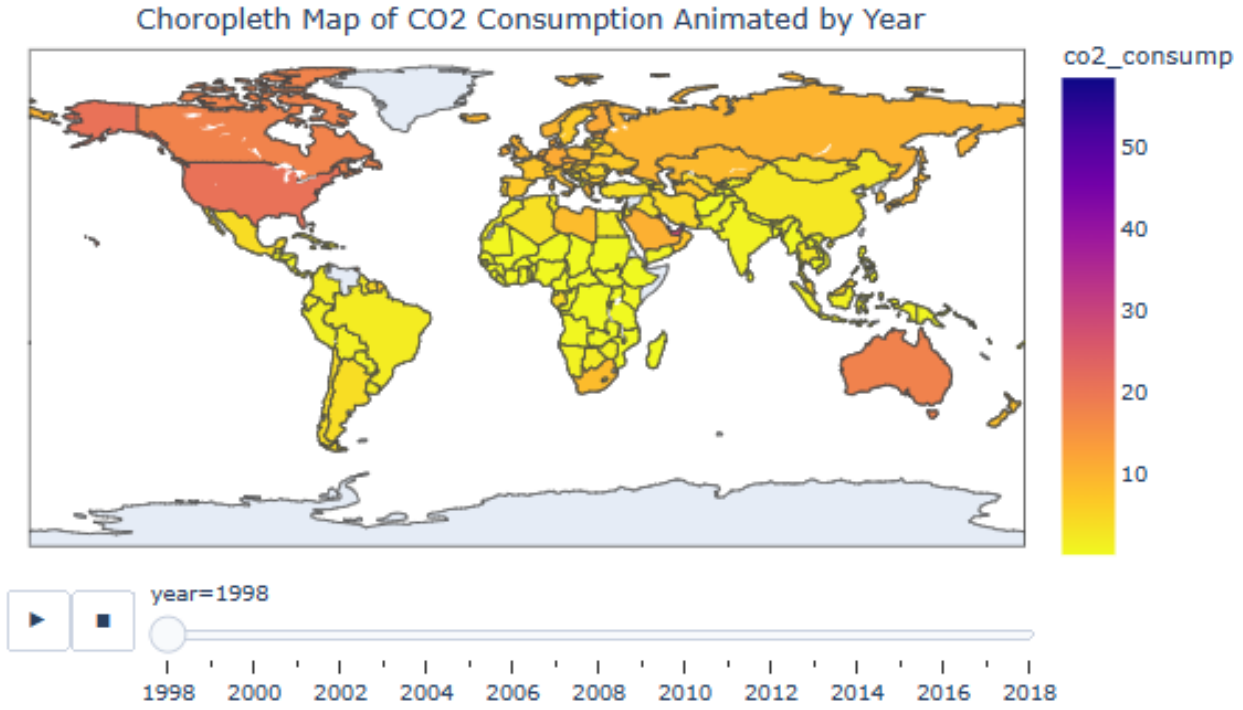
Choropleth Map

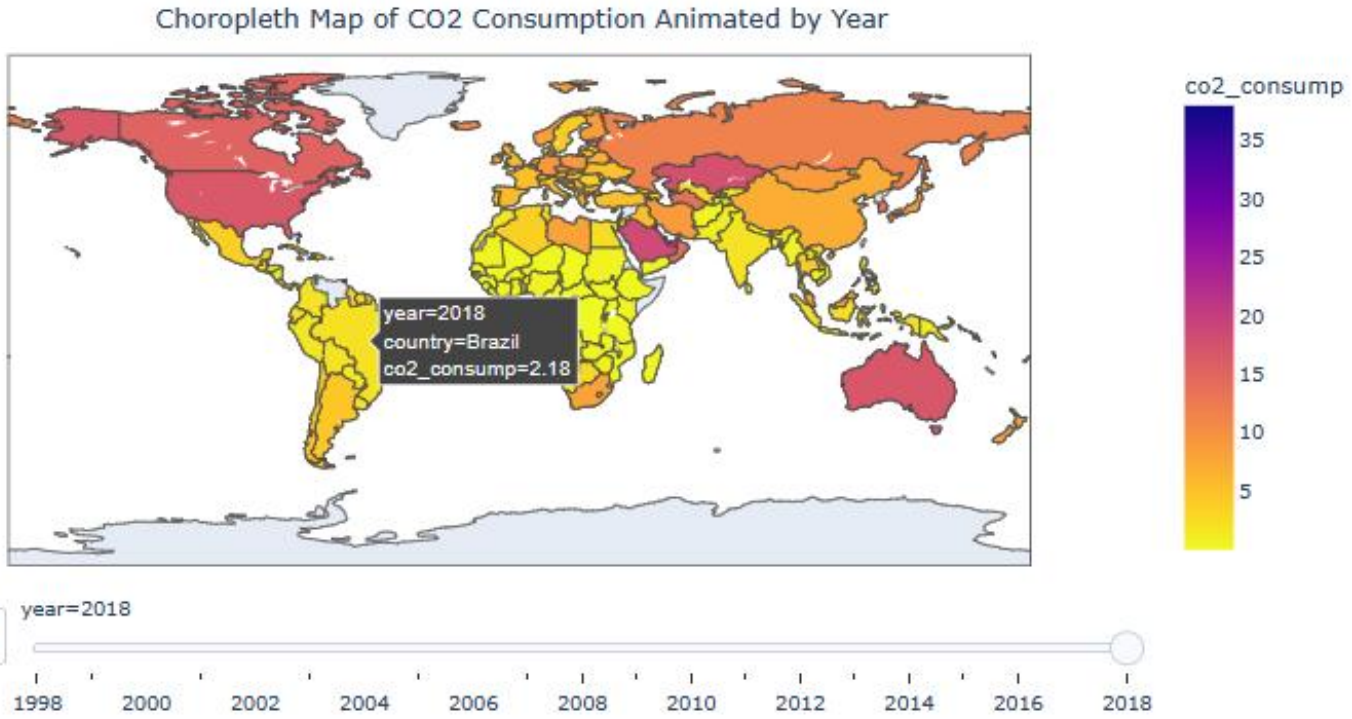




FGCU

Choropleth Map



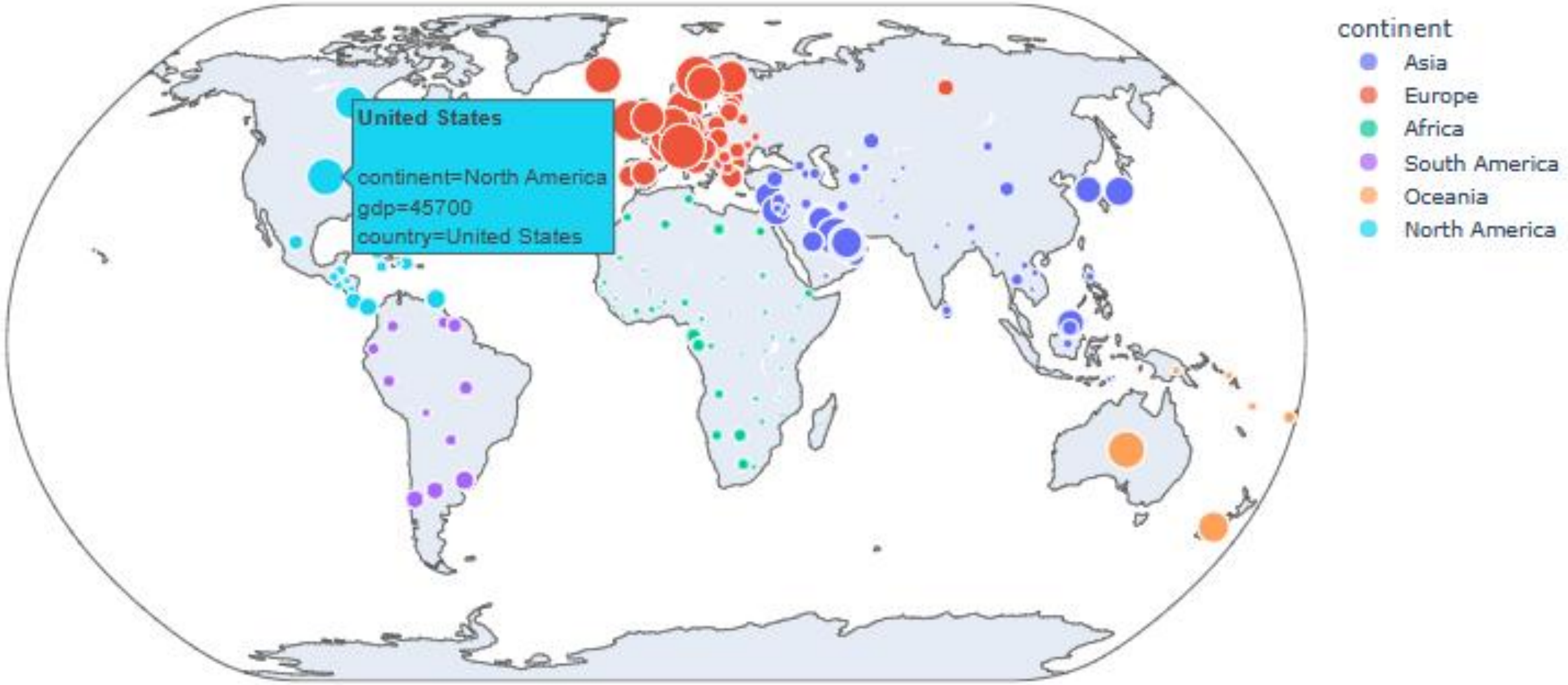


FGCU

Bubble Map

- **Purpose**: to explore and visualize spatial data represented by bubbles or circles of varying sizes in an area map.
- **Common type of data**: spatial data, that is, data containing different values or categories associated with specific (geographic) regions.
- **Interpretation**: the bubble map uses dots (bubbles) of varying sizes in the map to indicate the variable's value.
- **Examples of applications**: to visualize spatial data, such as population density, disease spread, electricity consumption, weather data, employment rates, economic indices (e.g., GDP and HDI), etc.

Bubble Map of GDP by Country and Continent

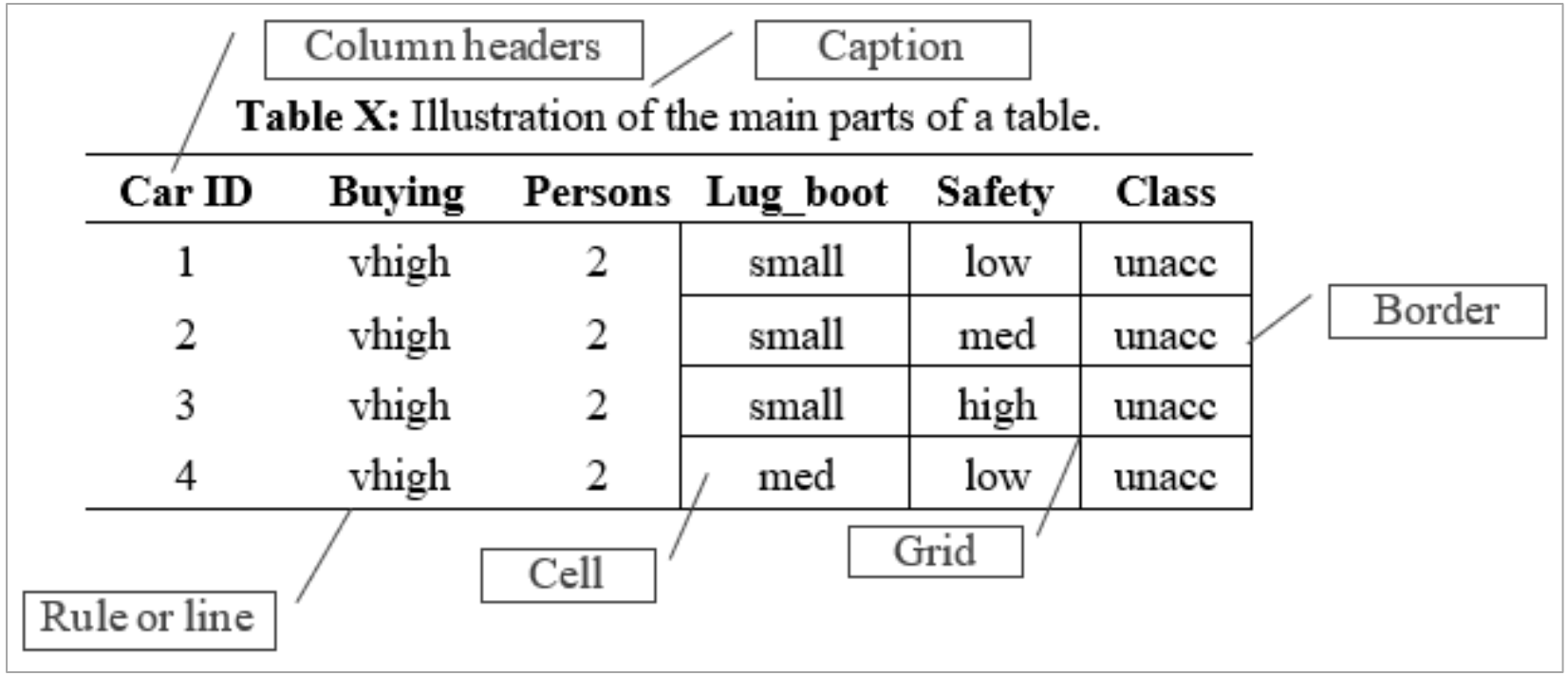


Choropleth and Bubble Maps with AI

“Create some choropleth and bubble maps with the Gapminder dataset.”

Design Principles for Data Visualization

Tables


Column headers
Caption
Table X: Illustration of the main parts of a table.
Car ID
Buying
Persons
Lug_boot
Safety
Class
1
vhigh
2
small
low
unacc
2
vhigh
2
small
med
unacc
3
vhigh
2
small
high
unacc
4
vhigh
2
med
low
unacc
Border
Cell
Grid
Rule or line

When to Use Tables?

- When specific values must be identified with precision.
- When pairs of related values must be compared.
- When look-up and one-to-one comparisons must be made.
- When objects must be compared over various characteristics with different units of measure.
- When a single display must be used to present unit and summary data.

Graphs

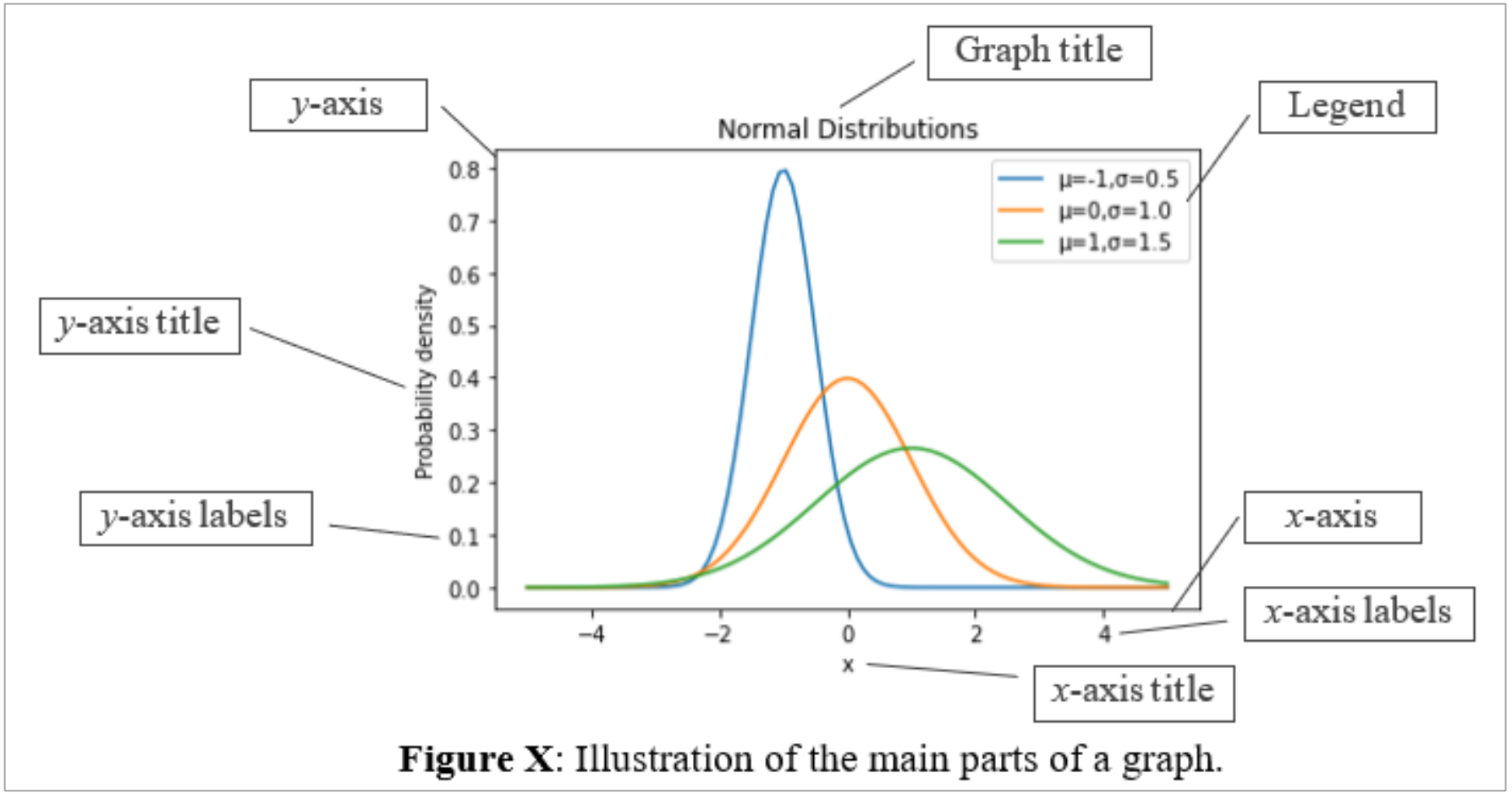


Figure X: Illustration of the main parts of a graph.

When to Use Graphs?

- When a set of quantitative values must be examined to find its general shape, pattern, or trend.
- When a single unit of measure is sufficient to compare the different values.
- When specific patterns are sought, such as anomalous values and clusters.
- When the volume of data is too large for a table but can be summarized in a graph.

Table 21: Comparison of the suitability of tables and graphs for data visualization based on some of their characteristics.

Characteristic	Tables	Graphs
Accuracy in identifying values	Yes	No
Easier identification of trends in data	No	Yes
Look-up and one-to-one comparisons	Yes	No
Comparison over multiple characteristics simultaneously	Yes	No
Display of unitary and summary information	Yes	No
Large volumes of data	No	Yes
Fast identification of patterns, trends, and outliers	No	Yes

Guidelines for Table Visualization

1. Use lines only as table frames and to separate the title row
2. Offset the headers from body
3. Use subtle dividers instead of heavy gridlines
4. Right-align numbers
5. Left-align text
6. Center-align columns with single characters

Guidelines for Table Visualization

7. Align the headers with their data
8. Select the appropriate level of precision
9. Guide the reader with space between rows and columns
10. Remove unit repetition
11. Highlight specific values
12. Group similar data and detach them

Version 1

pop_est	continent	name	iso_a3	gdp_md_est	geometry
889,953	Oceania	Fiji	FJI	5496	MULTIPOLYGON
58,005,463	Africa	Tanzania	TZA	63177	POLYGON
603,253	Africa	W. Sahara	ESH	907	POLYGON
37,589,262	America	Canada	CAN	1736425	MULTIPOLYGON
328,239,523	America	USA	USA	21433226	MULTIPOLYGON
6,944,975	Europe	Serbia	SRB	51475	POLYGON
622,137	Europe	Montenegro	MNE	5542	POLYGON
1,794,248	Europe	Kosovo	–99	7926	POLYGON
1,394,973	America	Trinidad and Tobago	TTO	24269	POLYGON
11,062,113	Africa	S. Sudan	SSD	11998	POLYGON

Version 2

pop_est	continent	name	iso_a3	gdp_md_est	geometry
889,953	Oceania	Fiji	FJI	5496	Multipolygon
58,005,463	Africa	Tanzania	TZA	63177	Polygon
603,253	Africa	W. Sahara	ESH	907	Polygon
37,589,262	America	Canada	CAN	1736425	Multipolygon
328,239,523	America	USA	USA	21433226	Multipolygon
6,944,975	Europe	Serbia	SRB	51475	Polygon
622,137	Europe	Montenegro	MNE	5542	Polygon
1,794,248	Europe	Kosovo	–99	7926	Polygon
1,394,973	America	Trinidad and Tobago	TTO	24269	Polygon
11,062,113	Africa	S. Sudan	SSD	11998	Polygon

Version 3

continent	name	pop_est	iso_a3	gdp_md_est	geometry
Oceania	Fiji	889,953	FJI	5496	Multipolygon
Africa	Tanzania	58,005,463	TZA	63177	Polygon
	W. Sahara	603,253	ESH	907	Polygon
	S. Sudan	11,062,113	SSD	11998	Polygon
America	Canada	37,589,262	CAN	1736425	Multipolygon
	USA	328,239,523	USA	21433226	Multipolygon
	Trinidad and Tobago	1,394,973	TTO	24269	Polygon
Europe	Serbia	6,944,975	SRB	51475	Polygon
	Montenegro	622,137	MNE	5542	Polygon
	Kosovo	1,794,248	–99	7926	Polygon

Version 1

pop_est	continent	name	iso_a3	gdp_md_est	geometry
889,953	Oceania	Fiji	FJI	5496	MULTIPOLYGON
58,005,463	Africa	Tanzania	TZA	63177	POLYGON
603,253	Africa	W. Sahara	ESH	907	POLYGON
37,589,262	America	Canada	CAN	1736425	MULTIPOLYGON
328,239,523	America	USA	USA	21433226	MULTIPOLYGON
6,944,975	Europe	Serbia	SRB	51475	POLYGON
622,137	Europe	Montenegro	MNE	5542	POLYGON
1,794,248	Europe	Kosovo	–99	7926	POLYGON
1,394,973	America	Trinidad and Tobago	TTO	24269	POLYGON
11,062,113	Africa	S. Sudan	SSD	11998	POLYGON

Version 2

pop_est	continent	name	iso_a3	gdp_md_est	geometry
889,953	Oceania	Fiji	FJI	5496	Multipolygon
58,005,463	Africa	Tanzania	TZA	63177	Polygon
603,253	Africa	W. Sahara	ESH	907	Polygon
37,589,262	America	Canada	CAN	1736425	Multipolygon
328,239,523	America	USA	USA	21433226	Multipolygon
6,944,975	Europe	Serbia	SRB	51475	Polygon
622,137	Europe	Montenegro	MNE	5542	Polygon
1,794,248	Europe	Kosovo	–99	7926	Polygon
1,394,973	America	Trinidad and Tobago	TTO	24269	Polygon
11,062,113	Africa	S. Sudan	SSD	11998	Polygon

Version 3

continent	name	pop_est	iso_a3	gdp_md_est	geometry
Oceania	Fiji	889,953	FJI	5496	Multipolygon
Africa	Tanzania	58,005,463	TZA	63177	Polygon
	W. Sahara	603,253	ESH	907	Polygon
	S. Sudan	11,062,113	SSD	11998	Polygon
America	Canada	37,589,262	CAN	1736425	Multipolygon
	USA	328,239,523	USA	21433226	Multipolygon
	Trinidad and Tobago	1,394,973	TTO	24269	Polygon
Europe	Serbia	6,944,975	SRB	51475	Polygon
	Montenegro	622,137	MNE	5542	Polygon
	Kosovo	1,794,248	–99	7926	Polygon

Guidelines to Display Graphs

- **Show the data**: detach (only) those values that are of central importance for the story being told;
- **Reduce the clutter**: focus on the necessary and sufficient visual elements to convey a specific message (result), and be parsimonious on colors, number precision, labels, annotations, etc.
- **Integrate the graphics and text**: titles, labels, legends, and annotations are often as important as the graph itself, for example, instead of using legends the data can be directly labeled.

Guidelines to Display Graphs

- **Avoid the spaghetti chart**: the excess use of colors, icons, bars, lines, texture, annotations, etc., all can make the reading and understanding of a chart more complicated, resulting in the so-called *spaghetti chart*.
- **Start with gray**: starting with gray forces you to be more careful and conscious in the planning of colors and other visual elements.

Leandro Nunes de Castro

ldecastrosilva@fgcu.edu