

1921
— 2021

UN SECOLO
DI STORIA
D'AVANTI A NOI



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

Time Series Analysis and Forecasting

MGO962

Lab 2: Data visualization in Python

General Structure

The general structure of the code that is used to plot figures is as follows:

```
import matplotlib.pyplot as plt
# Plot
f, ax = plt.subplots(1,1)
# Legend
plt.title('plot')
plt.xlabel('day')
```

We note that the `plt.subplots()` command enables to specify the figure size.

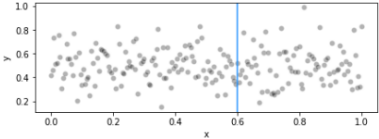
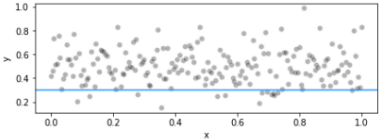
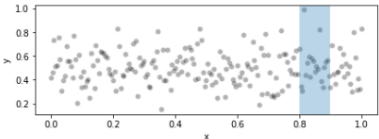
Basic Plots

Type	Command and parameters	Illustration
Scatter plot	<pre>sns.scatterplot(x, y, hue, size)</pre>	 <p>A scatter plot with x-axis from 0.0 to 1.0 and y-axis from 0.2 to 1.0. It displays numerous blue circular data points scattered across the plot area, showing a general positive trend.</p>
Line plot	<pre>sns.lineplot(x, y, hue, size)</pre>	 <p>A line plot with x-axis from 0.0 to 1.0 and y-axis from 0.0 to 1.0. It shows a single blue line oscillating in a regular, periodic wave pattern between y=0 and y=1.</p>
Bar chart Histogram	<pre>sns.barplot(x, y, hue)</pre>	 <p>A bar chart with x-axis from 0.0 to 1.0 and y-axis from 0.0 to 1.0. It features approximately 20 blue bars of varying heights, representing the frequency of data points within specific bins.</p>
Box plot	<pre>sns.boxplot(x, y, hue)</pre>	 <p>A box plot with x-axis labels 'group_1', 'group_2', 'group_3', 'group_4', and 'group_5' and y-axis from 0.0 to 1.0. Each group has a blue box plot showing the median, quartiles, and range of the data.</p>

Basic Plot Parameters

Command	Description	Use case
hue	Color of a line / point / border	'red'
fill	Color of an area	'red'
size	Size of a line / point	4
linetype	Shape of a line	'dashed'
alpha	Transparency, between 0 and 1	0.3

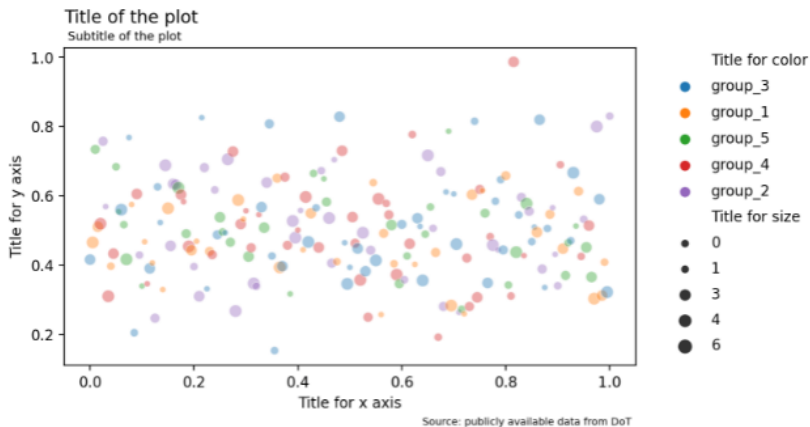
Additional Plot Elements

Type	Command	Illustration
Line	<pre>ax.axvline(x, ymin, ymax, color, linewidth, linestyle)</pre>	 A scatter plot showing a distribution of grey data points on a coordinate system with x and y axes ranging from 0.0 to 1.0. A solid blue vertical line is drawn at x = 0.6, extending from the bottom to the top of the plot area.
	<pre>ax.axhline(y, xmin, xmax, color, linewidth, linestyle)</pre>	 A scatter plot showing a distribution of grey data points on a coordinate system with x and y axes ranging from 0.0 to 1.0. A solid blue horizontal line is drawn at y = 0.3, extending from the left to the right of the plot area.
Rectangle	<pre>ax.axvspan(xmin, xmax, ymin, ymax, color, fill, alpha)</pre>	 A scatter plot showing a distribution of grey data points on a coordinate system with x and y axes ranging from 0.0 to 1.0. A light blue shaded vertical rectangle is drawn, spanning from x = 0.8 to x = 0.9 and from y = 0.2 to y = 1.0.

Plot Legend

Element	Command
Title / subtitle of the plot	<code>ax.set_title('text', loc, pad)</code>
	<code>plt.suptitle('text', x, y, size, ha)</code>
Title of the x / y axis	<code>ax.set_xlabel('text') / ax.set_ylabel('text')</code>
Title of the size / color	via <code>ax.get_legend_handles_labels()</code>
Caption of the plot	<code>ax.text('text', x, y, fontsize)</code>

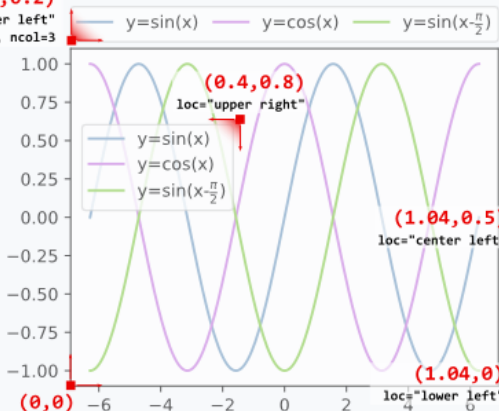
Plot Legend



Plot Legend Position

$(0, 1.02, 1, 0.2)$

loc="lower left"
mode="expand", ncol=3



$(1.04, 1)$

y=sin(x)
y=cos(x)
y=sin(x- $\frac{\pi}{2}$)

y=sin(x)
y=cos(x)
y=sin(x- $\frac{\pi}{2}$)

y=sin(x)
y=cos(x)
y=sin(x- $\frac{\pi}{2}$)

$(1, 0)$ lloc="lower right",
bbox_transform=fig.transFigure

y=sin(x) y=cos(x) y=sin(x- $\frac{\pi}{2}$)

$(0, 0)$

Plot Legend Position

```
import matplotlib.pyplot as plt
```

```
f, ax = plt.subplots(1,1)
```

```
l1 = plt.legend(bbox_to_anchor=(1.04,1), borderaxespad=0)
```

```
l2 = plt.legend(bbox_to_anchor=(1.04,0), loc="lower left", borderaxespad=0)
```

```
l3 = plt.legend(bbox_to_anchor=(1.04,0.5), loc="center left", borderaxespad=0)
```

```
l4 = plt.legend(bbox_to_anchor=(0,1.02,1,0.2), loc="lower left",  
                mode="expand", borderaxespad=0, ncol=3)
```

```
l5 = plt.legend(bbox_to_anchor=(1,0), loc="lower right",  
                bbox_transform=f.transFigure, ncol=3)
```

```
l6 = plt.legend(bbox_to_anchor=(0.4,0.8), loc="upper right")
```

Australian Beer Production Dataset

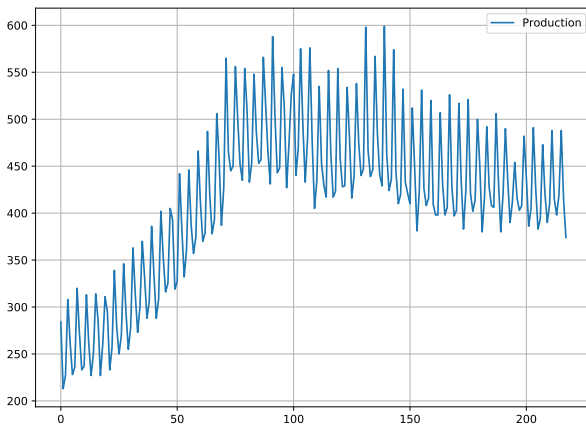
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/ausbeer.csv")
print(df.head())
```

##	Time	Production
## 0	1956-01-01	284
## 1	1956-04-01	213
## 2	1956-07-01	227
## 3	1956-10-01	308
## 4	1957-01-01	262

Australian Beer Production Plot

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/ausbeer.csv")
df.plot()
plt.show()
```

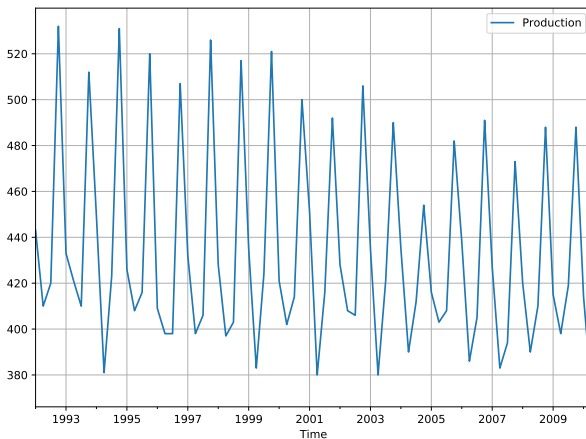
Australian Beer Production Plot



Australian Beer Production Plot > 1992

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/ausbeer.csv")
df.plot()
plt.show()
```

Australian Beer Production Plot > 1992



Electricity Production Dataset

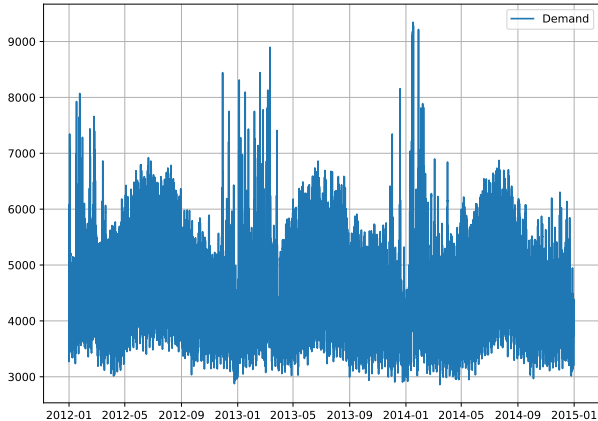
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/vic_elec.csv", header=0,
                 parse_dates=True, index_col=0)
print(df.head())
```

##		Demand	Temperature	Date	Holiday
##	Time				
##	2011-12-31 13:00:00+00:00	4382.825174	21.40	2012-01-01	True
##	2011-12-31 13:30:00+00:00	4263.365526	21.05	2012-01-01	True
##	2011-12-31 14:00:00+00:00	4048.966046	20.70	2012-01-01	True
##	2011-12-31 14:30:00+00:00	3877.563330	20.55	2012-01-01	True
##	2011-12-31 15:00:00+00:00	4036.229746	20.40	2012-01-01	True

Electricity Production Plot

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/vic_elec.csv", header=0,
                 parse_dates=True, index_col=0)
plt.plot('Demand', data=df)
plt.grid()
plt.show()
```


Electricity Production Plot



Australian Tourism Dataset

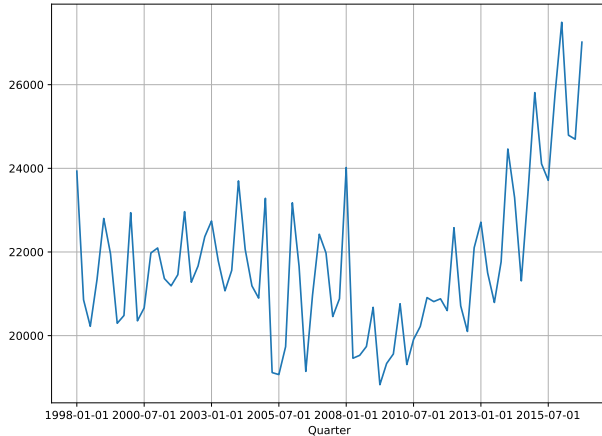
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/tourism.csv").drop(columns=["Unnamed: 0"])
print(df.head())
```

##	Quarter	Region	State	Purpose	Trips
## 0	1998-01-01	Adelaide	South Australia	Business	135.077690
## 1	1998-04-01	Adelaide	South Australia	Business	109.987316
## 2	1998-07-01	Adelaide	South Australia	Business	166.034687
## 3	1998-10-01	Adelaide	South Australia	Business	127.160464
## 4	1999-01-01	Adelaide	South Australia	Business	137.448533

Australian Tourism Total Trips

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/tourism.csv").drop(columns=["Unnamed: 0"])
df = df.groupby("Quarter")
df = df['Trips'].agg(sum)
df.plot(label='Total')
```

Australian Tourism Total Trips



Australian Tourism Trips by Purpose

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/tourism.csv").drop(columns=["Unnamed: 0"])
purposes = df["Purpose"].unique()
for purpose in purposes:
    dfp = df.query('Purpose == "' + purpose + '"')
    dfp = dfp.groupby("Quarter")
    dfp = dfp['Trips'].agg(sum)
    dfp.plot(label=purpose)
```

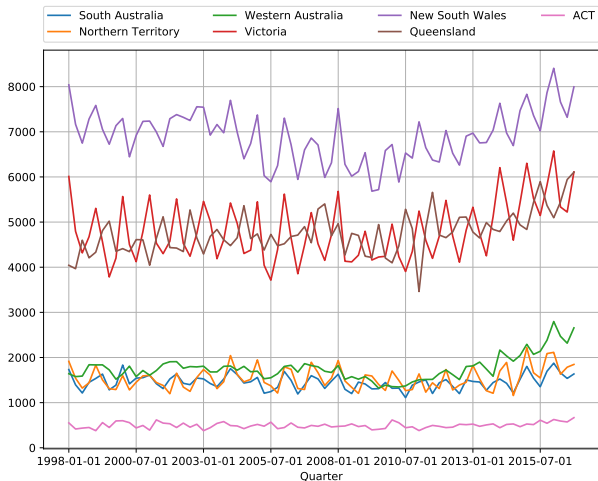
Australian Tourism Trips by Purpose



Australian Tourism Trips by State

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/tourism.csv").drop(columns=["Unnamed: 0"])
states = df["State"].unique()
for state in states:
    dfp = df.query('State == "' + state + '"')
    dfp = dfp.groupby("Quarter")
    dfp = dfp['Trips'].agg(sum)
    dfp.plot(label=state)
```

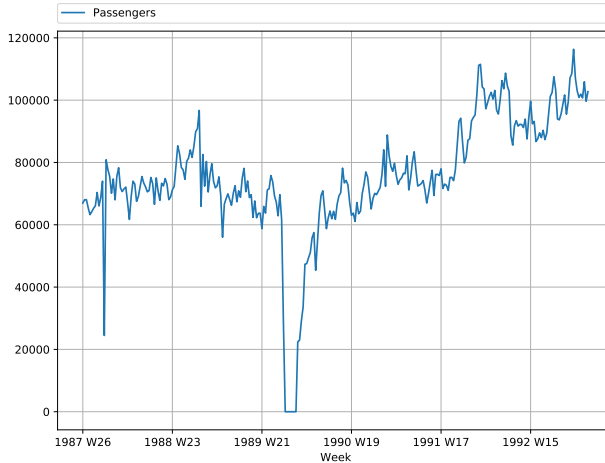
Australian Tourism Trips by State



Ansett Airlines Total Trips

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/ansettDate.csv")
dft = df.groupby("Week")
dft = dft['Passengers'].agg(sum)
dft.plot()
```

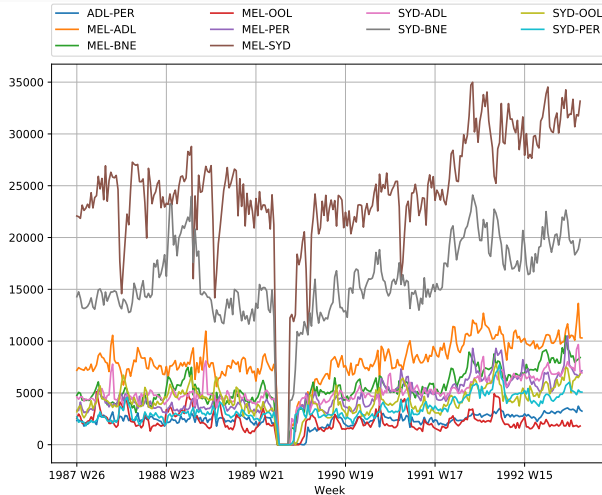
Ansett Airlines Total Trips



Ansett Airlines Trips by Airport

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/ansettDate.csv")
airports = df["Airports"].unique()
for airport in airports:
    dfs = df.query('Airports == "' + airport + '"')
    dfs = dfs.groupby("Week")
    dfs = dfs['Passengers'].agg(sum)
    dfs.plot(label=airport)
```

Ansett Airlines Trips by Airport



Us Employment Dataset

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/usemployment.csv",
                  header=0, parse_dates=True, index_col=0)
print(df.head())
```

##	Series_ID	Title	Employed
## Month			
## 1939-01-01	CEU0500000001	Total Private	25338.0
## 1939-02-01	CEU0500000001	Total Private	25447.0
## 1939-03-01	CEU0500000001	Total Private	25833.0
## 1939-04-01	CEU0500000001	Total Private	25801.0
## 1939-05-01	CEU0500000001	Total Private	26113.0

Us Employment Dataset Retail Plot

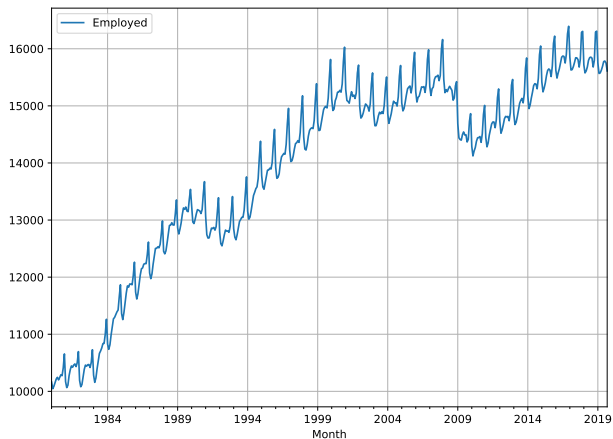
```
import pandas as pd
import matplotlib.pyplot as plt

# load the data
df = pd.read_csv("data/usemployment.csv",
                  header=0, parse_dates=True, index_col=0)
dfs = df.query('Title == "Retail Trade"')
dfs = dfs.loc['1980-01-01':'2020-01-01']

dfs.plot()

plt.grid()
plt.show()
```

Us Employment Dataset Retail Plot



Australian Production Dataset

```
import pandas as pd
df = pd.read_csv("data/aus_production.csv",
                 header=0, parse_dates=['Quarter'], index_col=0)
print(df.head())
```

##		Beer	Tobacco	Bricks	Cement	Electricity	Gas
##	Quarter						
##	1956 Q1	284	5225.0	189.0	465	3923	5
##	1956 Q2	213	5178.0	204.0	532	4436	6
##	1956 Q3	227	5297.0	208.0	561	4806	7
##	1956 Q4	308	5681.0	197.0	570	4418	6
##	1957 Q1	262	5577.0	187.0	529	4339	5

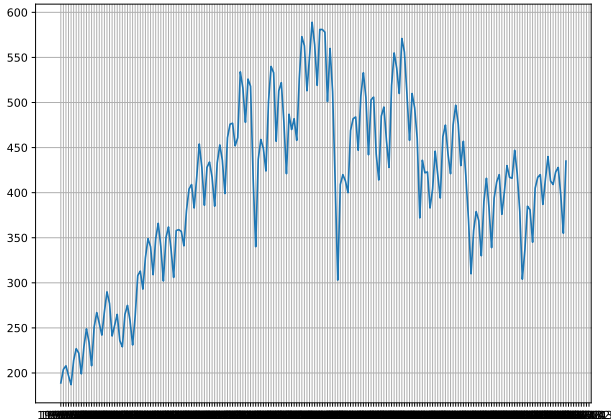
Australian Bricks Production Plot

```
import pandas as pd
import matplotlib.pyplot as plt

# load the data
df = pd.read_csv("data/aus_production.csv",
                 header=0, parse_dates=['Quarter'], index_col=0)
dfs = df.loc['1956-01-01':'2005-01-01']

# Total
plt.plot('Bricks', data=df)
plt.grid()
plt.show()
```

Australian Bricks Production Plot



Australian Arrivals Dataset

```
import pandas as pd
df = pd.read_csv("data/arrivals.csv", parse_dates=True)
df.head()
```

##		index	key	value
## 0		1981-01-01	Japan	14.763
## 1		1981-04-01	Japan	9.321
## 2		1981-07-01	Japan	10.166
## 3		1981-10-01	Japan	19.509
## 4		1982-01-01	Japan	17.117

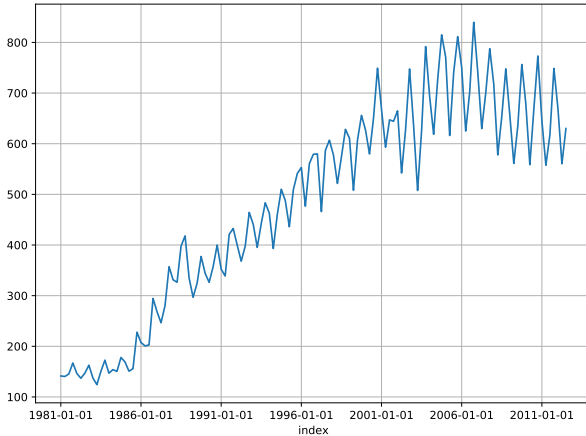
Australian Arrivals Total Plot

```
import pandas as pd
import matplotlib.pyplot as plt

# load the data
df = pd.read_csv("data/arrivals.csv", parse_dates=True)

# Total
dft = df.groupby("index")
dft = dft['value'].agg(sum)
dft.plot()
plt.grid()
plt.show()
```

Australian Arrivals Total Plot



Australian Arrivals Plot by State

```
import pandas as pd
import matplotlib.pyplot as plt

# load the data
df = pd.read_csv("data/arrivals.csv", parse_dates=True)

# States
states = df["key"].unique()
for state in states:
    dfs = df.query('key == "' + state + '"')
    dfs = dfs.groupby("index")
    dfs = dfs['value'].agg(sum)
    dfs.plot(label=state)

plt.legend()
plt.grid()
plt.show()
```

Australian Arrivals Plot by State

