

# IMD0033 - Probabilidade

## Aula 09 - Análise Exploratória de Dados

### Introdução

Ivanovitch Silva  
Março, 2019



# Agenda

---

- Motivação
- Estudo de caso: taxa de desemprego
- Representação tabular vs visual
- Matplotlib
- Gráficos em linhas
- Multiplot
- Personalização

# Atualizar o repositório

---

```
git clone https://github.com/ivanovitchm/imd0033_2019_1
```

Ou ....

```
git pull
```



# Motivação

---



# Estudo de caso: taxa de desemprego (US)



Source: U.S. Bureau of Labor Statistics

[fred.stlouisfed.org](https://fred.stlouisfed.org)

[myf.red/g/eCMW](https://myf.red/g/eCMW)

# Analizando a base de dados

---

<b>DATE</b> Ano-Mês-Dia	<b>VALUE</b>
1948-01-01	3.4
1948-02-01	3.8
1948-03-01	4.0
1948-04-01	3.9
1948-05-01	3.5

Conversão de tipos (Object to Datetime)

```
import pandas as pd  
df['col'] = pd.to_datetime(df['col'])
```

DATE	VALUE
1948-01-01	3.4
1948-02-01	3.8
1948-03-01	4.0
1948-04-01	3.9
1948-05-01	3.5
1948-06-01	3.6
1948-07-01	3.6
1948-08-01	3.9
1948-09-01	3.8
1948-10-01	3.7
1948-11-01	3.8
1948-12-01	4.0

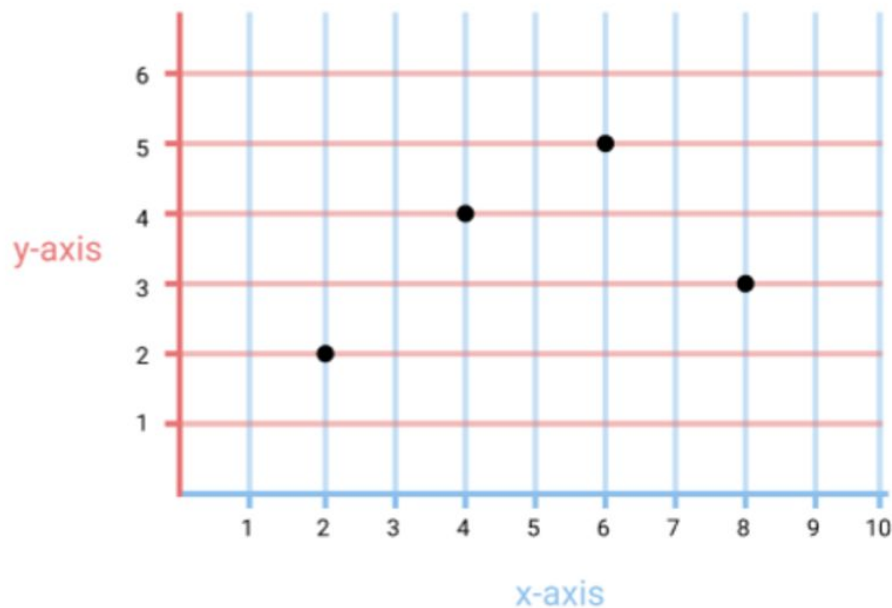
## Quais conclusões podem tirar?

- Qual o valor mínimo?
- Qual o valor máximo?
- Existe uma tendência/sazonalidade?
- Quais intervalos são de crescimento?
- Quais os intervalos são de decrescimento?
- A representação em forma de tabela é útil?

# Representação Visual

---

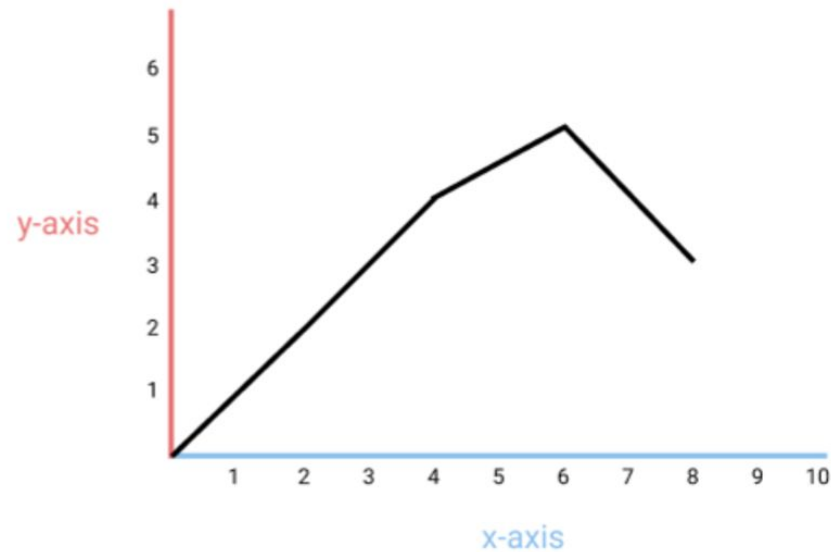
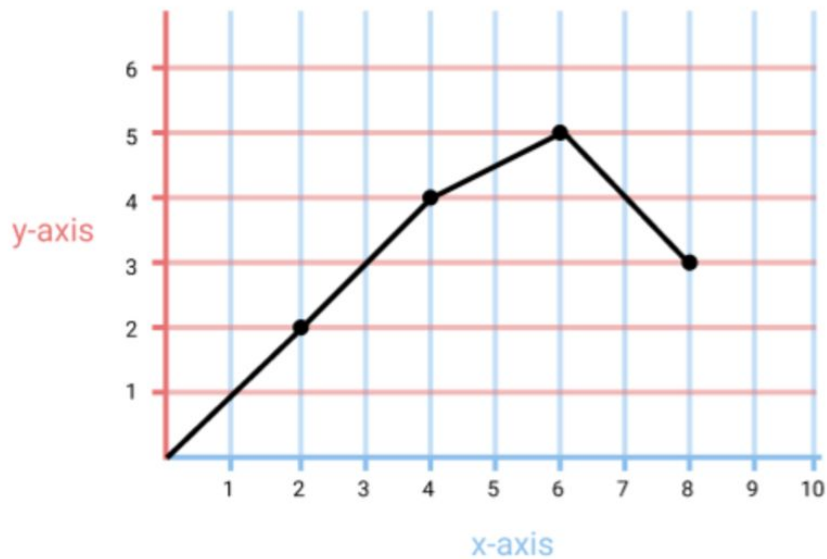
x	y
2	2
4	4
6	5
8	3



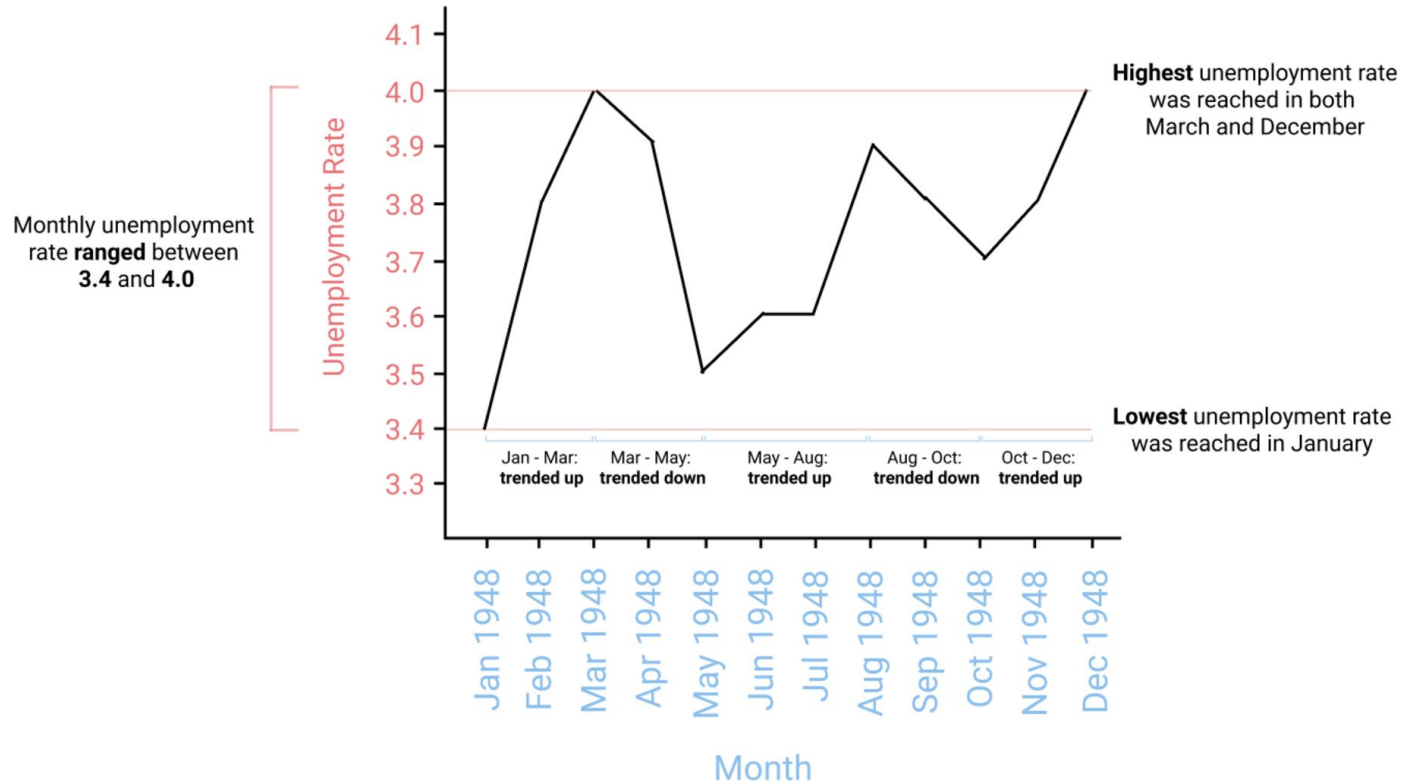


# Representação Visual

---

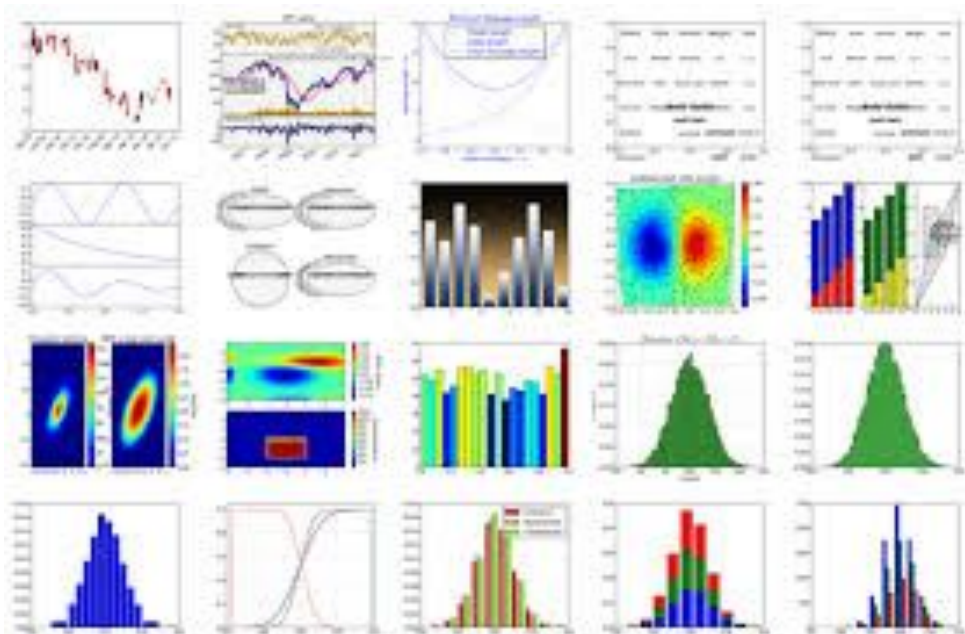


# Representação Visual



# matplotlib

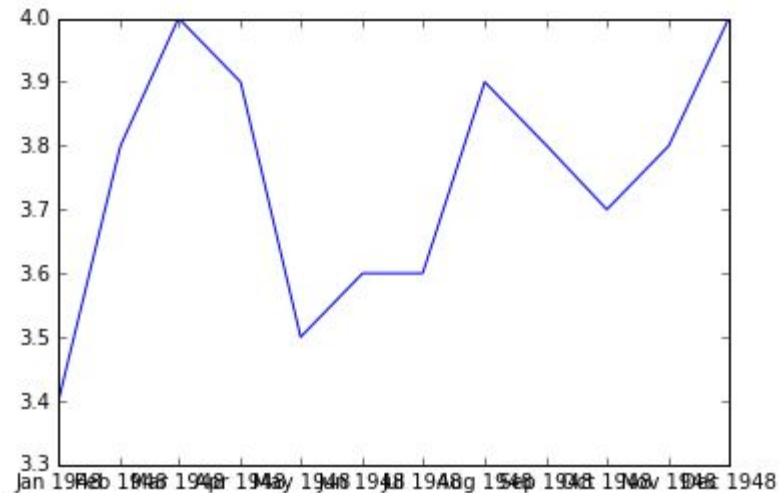
```
import matplotlib.pyplot as plt  
plt.plot()  
plt.show()
```



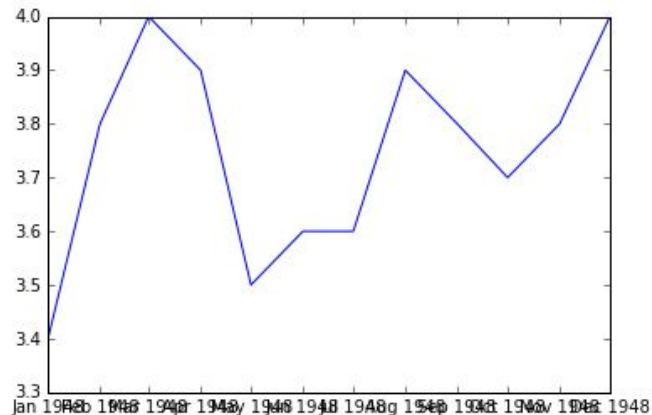
# Adicionando dados

---

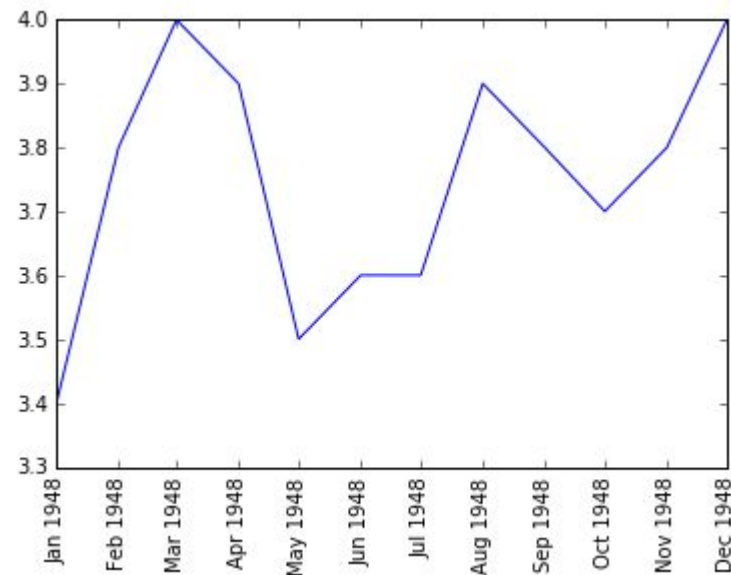
```
plt.plot(x_values, y_values)
```



# Ajustando os eixos



```
x = unrte.loc[:, "DATE"]
y = unrte.loc[:, "VALUE"]
plt.plot(range(0, 12), y)
plt.xticks(range(0, 12),
            x.dt.strftime("%b %Y"),
            rotation=90)
plt.show()
```



# Informações adicionais



```
plt.xlabel("Month")  
plt.ylabel("Unemployment Rate")  
plt.title("Monthly Unemployment Trends, 1948")
```

# EDA - Line Charts.ipynb

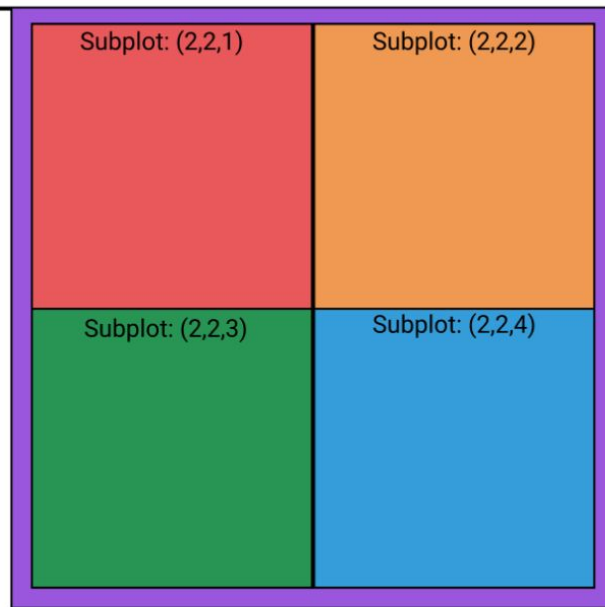


# Múltiplos gráficos

---

```
import matplotlib.pyplot as plt  
fig = plt.figure()  
ax1 = fig.add_subplot(2,2,1)  
ax2 = fig.add_subplot(2,2,2)  
ax3 = fig.add_subplot(2,2,3)  
ax4 = fig.add_subplot(2,2,4)
```

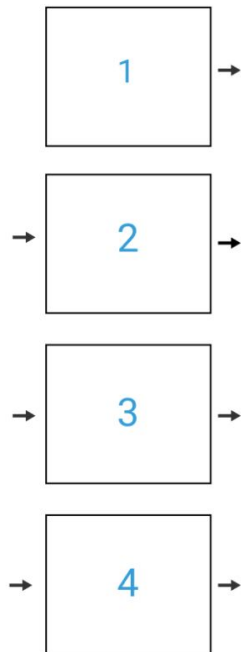
Figure



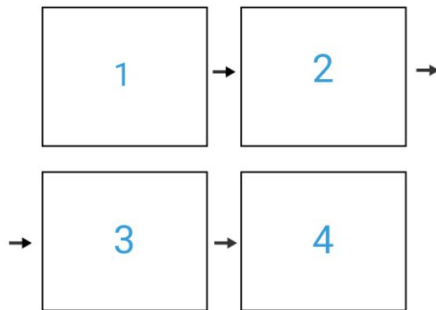


# Múltiplos gráficos

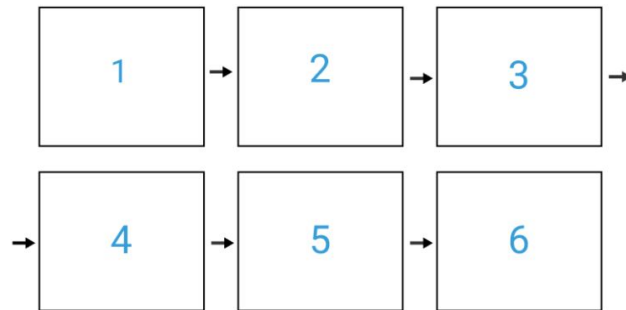
`fig.add_subplot(4, 1, x)`



`fig.add_subplot(2, 2, x)`



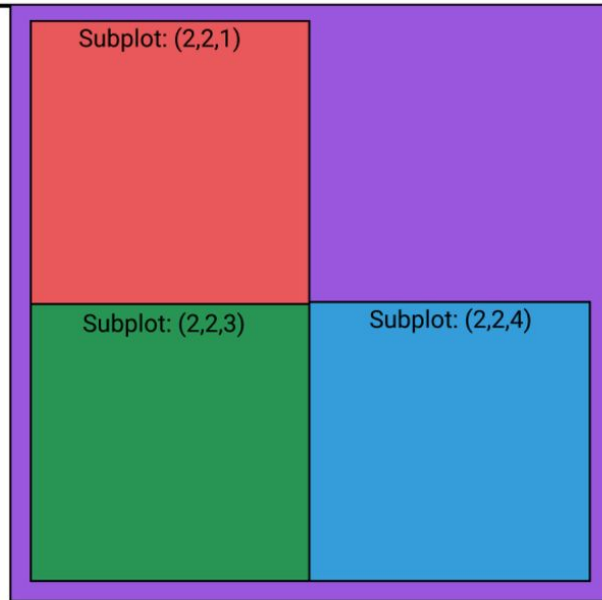
`fig.add_subplot(2, 3, x)`



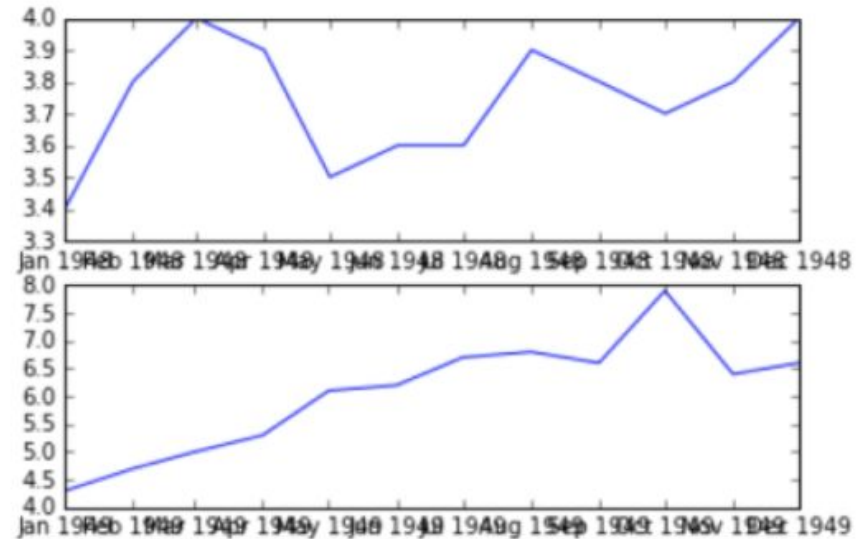
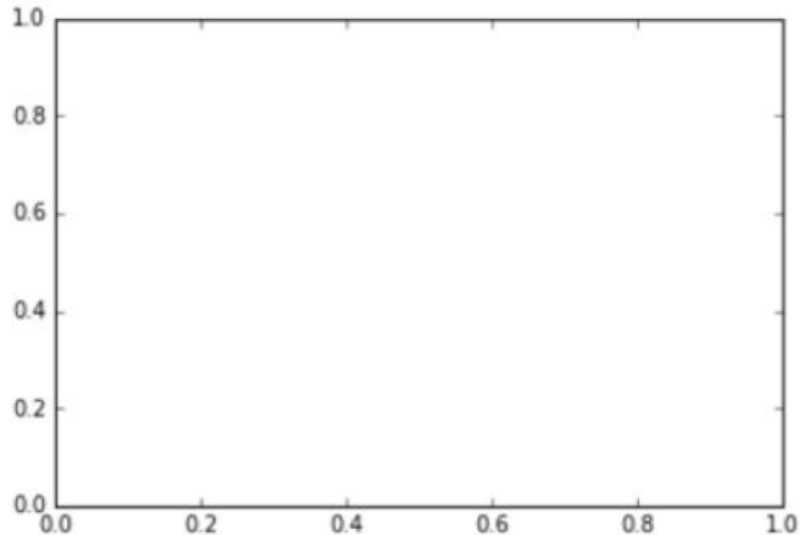
# Múltiplos gráficos

```
import matplotlib.pyplot as plt  
fig = plt.figure()  
ax1 = fig.add_subplot(2,2,1)  
ax3 = fig.add_subplot(2,2,3)  
ax4 = fig.add_subplot(2,2,4)
```

Figure

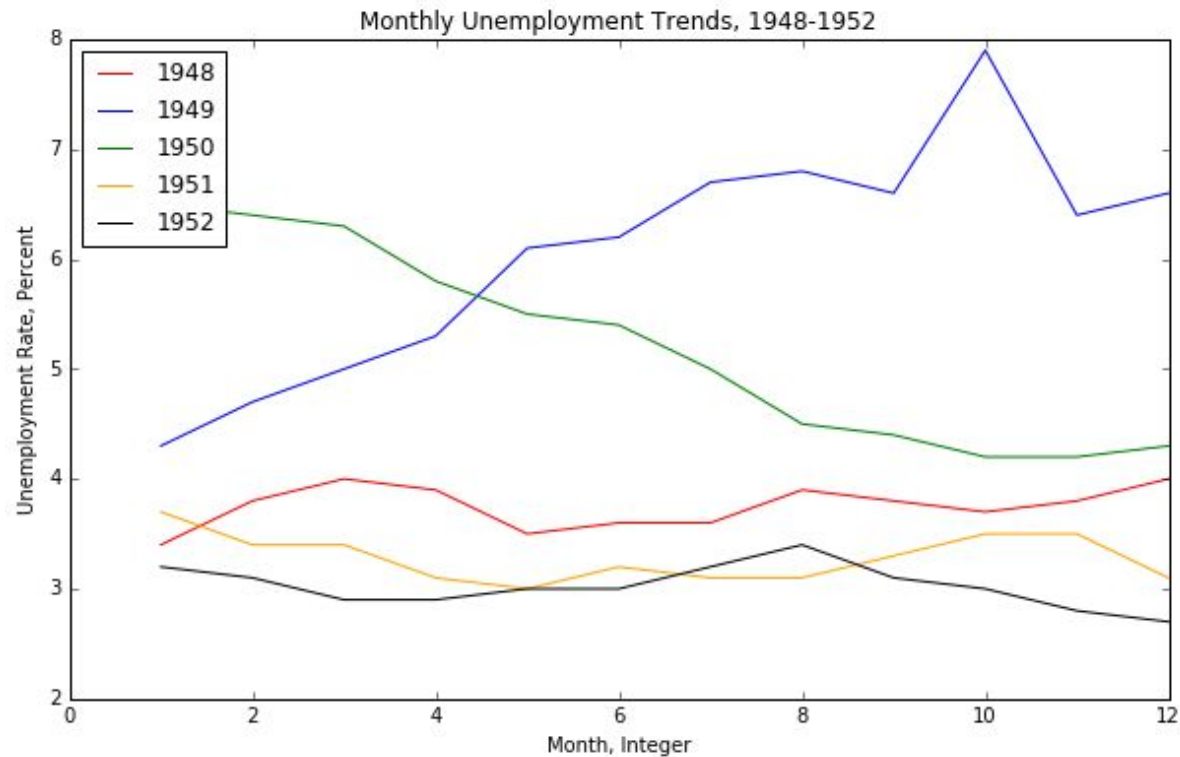


# Formatando as dimensões



```
fig = plt.figure(figsize=(width, height))
```

# Sobreposição de gráficos, legendas e rótulos



# EDA - Multiple Charts.ipynb

