

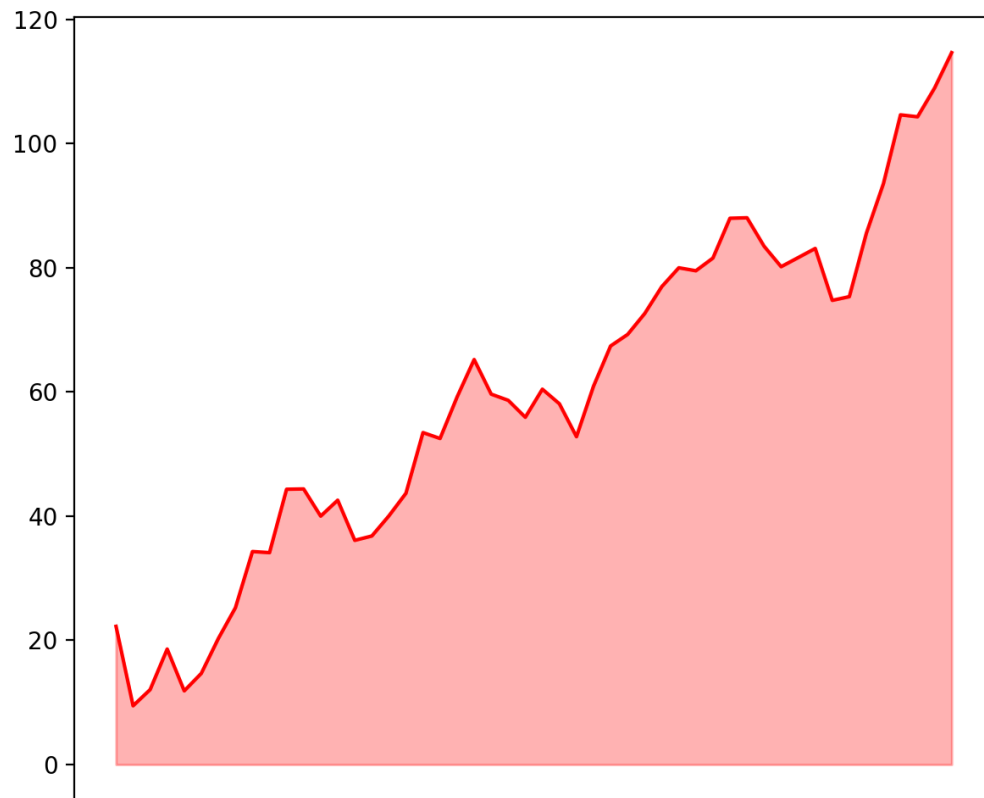
# Data Visualizations

Why do we need  
visualizations

# Before

```
array([ 22.2545198 ,  9.46306667, 12.06767132, 18.59783811,  
       11.86490354, 14.68040278, 20.30153772, 25.24777714,  
       34.3022338 , 34.12490434, 44.33391473, 44.38379237,  
       40.00574845, 42.57340636, 36.10801652, 36.80541831,  
       40.04538794, 43.69025546, 53.46028177, 52.50945039,  
       59.19988263, 65.21990689, 59.65118444, 58.65185448,  
       55.92723599, 60.44817943, 58.09343653, 52.79842096,  
       60.93714419, 67.40567495, 69.26647731, 72.62978286,  
       76.95759959, 80.0000368 , 79.51964481, 81.56353416,  
       87.97679347, 88.05404069, 83.47695913, 80.17622344,  
       81.63942456, 83.11399608, 74.75389511, 75.35131548,  
       85.5736879 , 93.56250189, 104.63174345, 104.31686973,  
      108.96186346, 114.64848866])
```

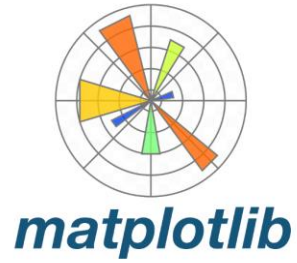
# After



# Main libraries

# Main libraries

Fundamental



# Main libraries

Fundamental



*matplotlib*



seaborn

# Main libraries

Fundamental



*matplotlib*



seaborn

Interactive





# Main libraries

Fundamental



Interactive



Geospatial



# Main libraries

Fundamental



*matplotlib*

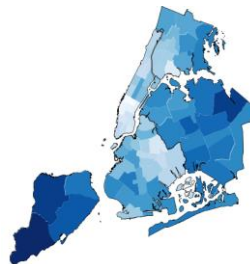


seaborn

Interactive



Geospatial



**Geoplot**



**Folium**

Other visualization apps

# Other visualization apps



# How to visualize data distribution



BRENDA N · UPDATED 3 YEARS AGO



1027

New Notebook



Download (12 kB)



# Titanic dataset

Gender submission and test file merged



Data Card

Code (394)

Discussion (2)

Suggestions (1)

## About Dataset



### Usability ⓘ

10.00

### License

CC0: Public Domain

### Expected update frequency

Never

### Tags

Beginner

Data Visualization

**HISTORICAL ACCURACY**



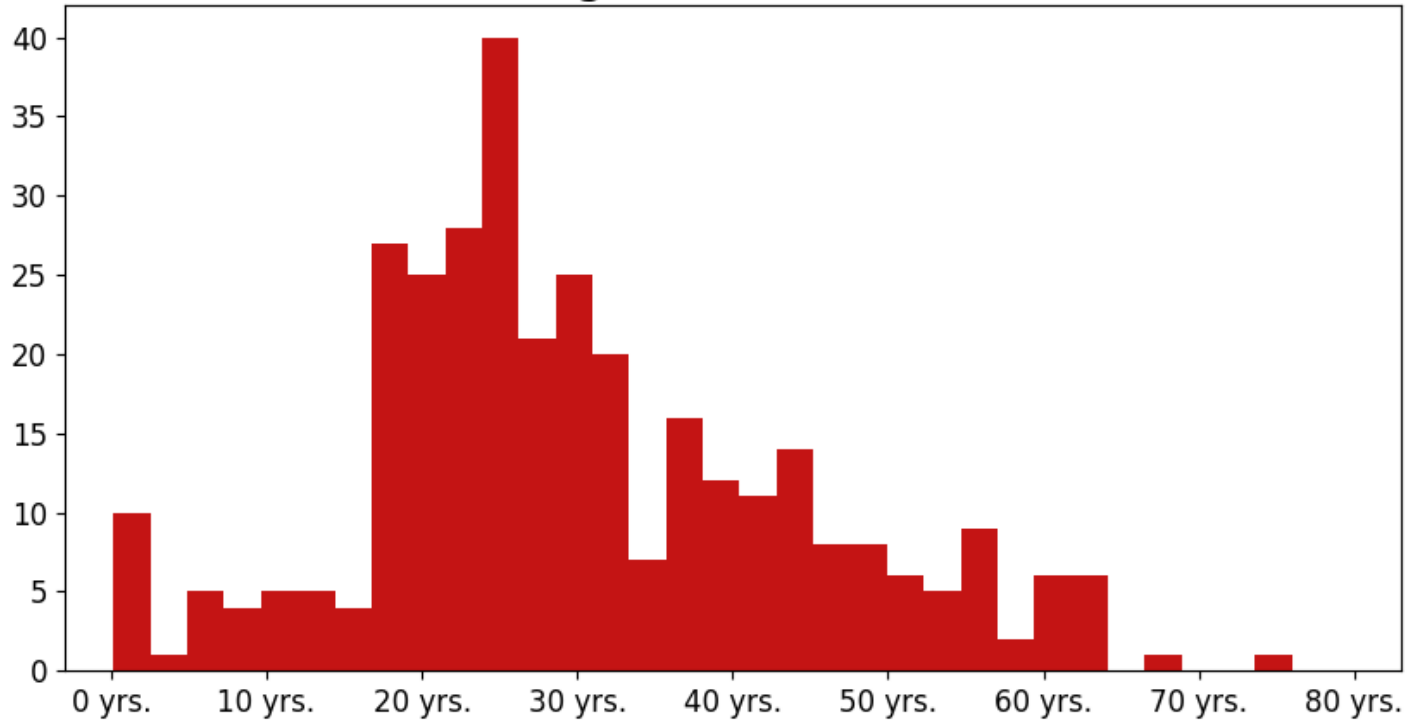
**KAGGLE COMPETITORS**

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	0	Third	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	1	Third	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	0	Second	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	0	Third	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	1	Third	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S
5	897	0	Third	Svensson, Mr. Johan Cervin	male	14.0	0	0	7538	9.2250	NaN	S
6	898	1	Third	Connolly, Miss. Kate	female	30.0	0	0	330972	7.6292	NaN	Q
7	899	0	Second	Caldwell, Mr. Albert Francis	male	26.0	1	1	248738	29.0000	NaN	S
8	900	1	Third	Abraham, Mrs. Joseph (Sophie Halaut Easu)	female	18.0	0	0	2657	7.2292	NaN	C
9	901	0	Third	Davies, Mr. John Samuel	male	21.0	2	0	A/4 48871	24.1500	NaN	S

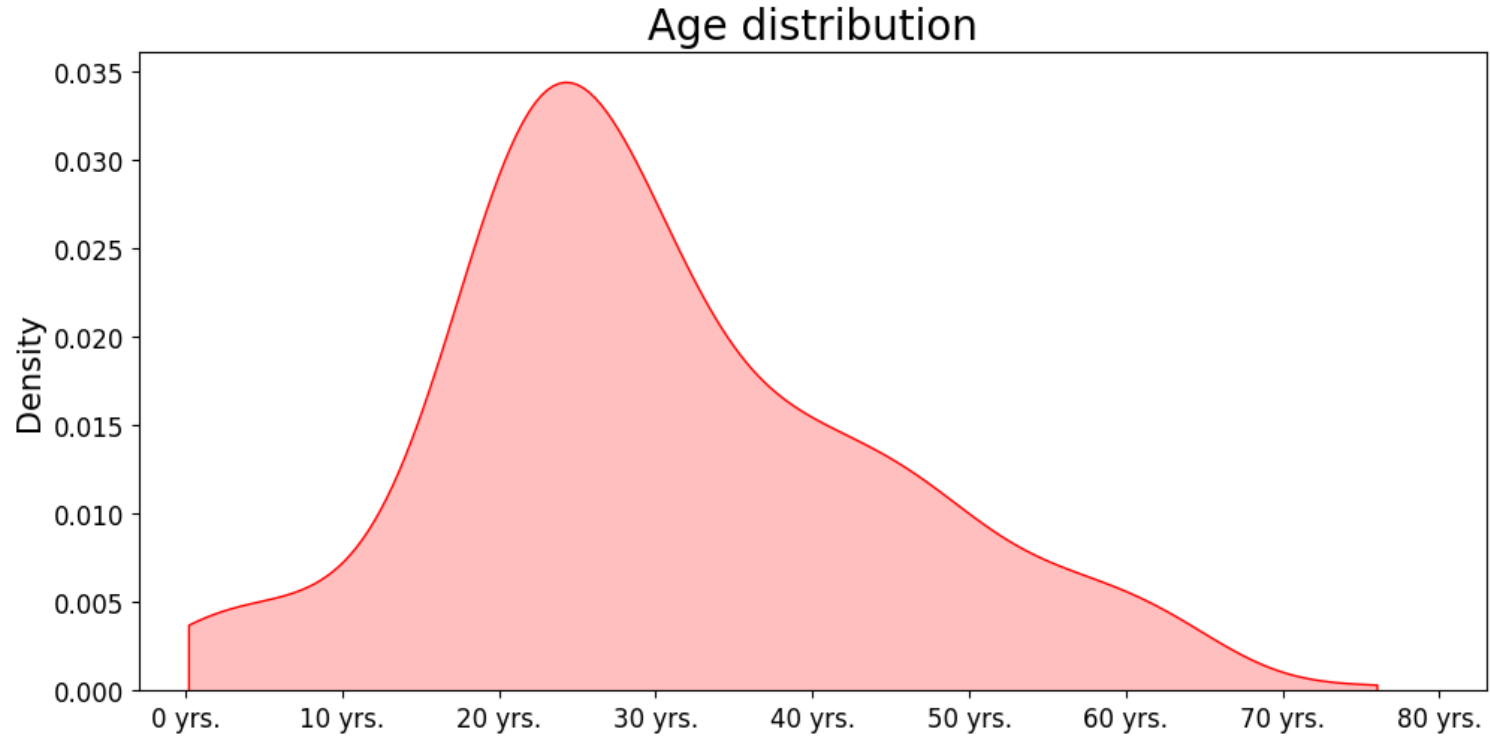


# Histogram

Age distribution



# Density plot

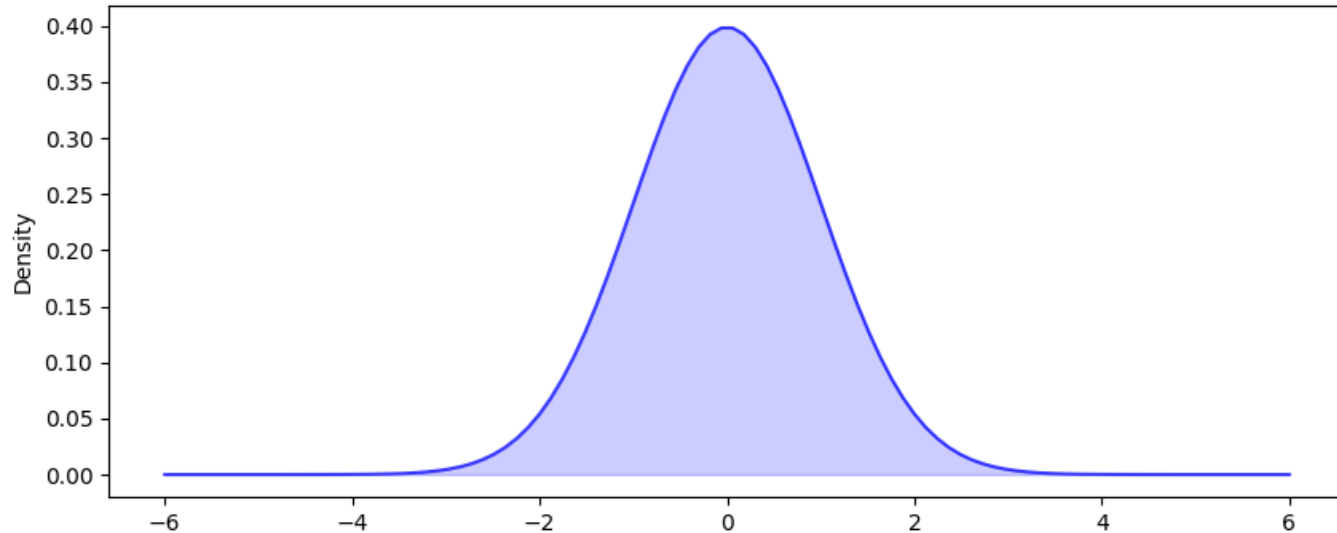


KDE

KDE - Kernel Density Estimator

# Density plot

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{x^2}{2} \right]$$



# Density plot

$$K(x - x_i)$$

# Density plot

$$K(x - x_i)$$

$$K\left(\frac{x - x_i}{h}\right)$$

# Density plot

h - kernel bandwidth

$$K(x - x_i)$$

$$K\left(\frac{x - x_i}{h}\right)$$



h - kernel bandwidth

# Density plot

$$K(x - x_i)$$

$$K\left(\frac{x - x_i}{h}\right)$$

$$\frac{1}{h}K\left(\frac{x - x_i}{h}\right)$$

# Density plot

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

# Density plot

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\frac{1}{h} K \left( \frac{x - x_1}{h} \right)$$

# Density plot

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\frac{1}{h}K\left(\frac{x - x_1}{h}\right) + \frac{1}{h}K\left(\frac{x - x_2}{h}\right)$$

## Density plot

$$f(x) = \frac{1}{2} \left[ \frac{1}{h} K \left( \frac{x-x_1}{h} \right) + \frac{1}{h} K \left( \frac{x-x_2}{h} \right) \right]$$

## Density plot

$$\begin{aligned} f(x) &= \frac{1}{2} \left[ \frac{1}{h} K \left( \frac{x-x_1}{h} \right) + \frac{1}{h} K \left( \frac{x-x_2}{h} \right) \right] = \\ &= \frac{1}{2h} \left[ K \left( \frac{x-x_1}{h} \right) + K \left( \frac{x-x_2}{h} \right) \right] \end{aligned}$$

## Density plot

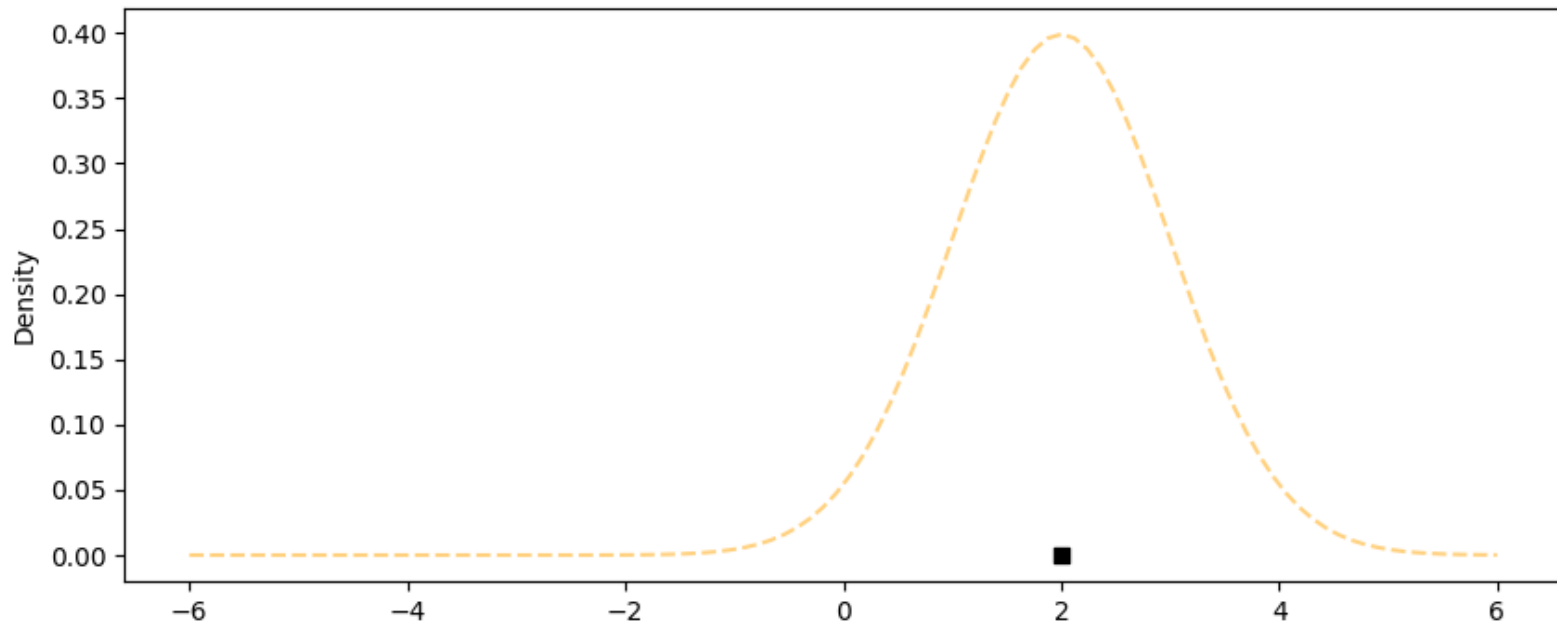
$$\begin{aligned} f(x) &= \frac{1}{2} \left[ \frac{1}{h} K \left( \frac{x-x_1}{h} \right) + \frac{1}{h} K \left( \frac{x-x_2}{h} \right) \right] = \\ &= \frac{1}{2h} \left[ K \left( \frac{x-x_1}{h} \right) + K \left( \frac{x-x_2}{h} \right) \right] = \\ &= \frac{1}{2h} \sum_{i=1}^2 K \left( \frac{x-x_i}{h} \right) \end{aligned}$$

# Density plot

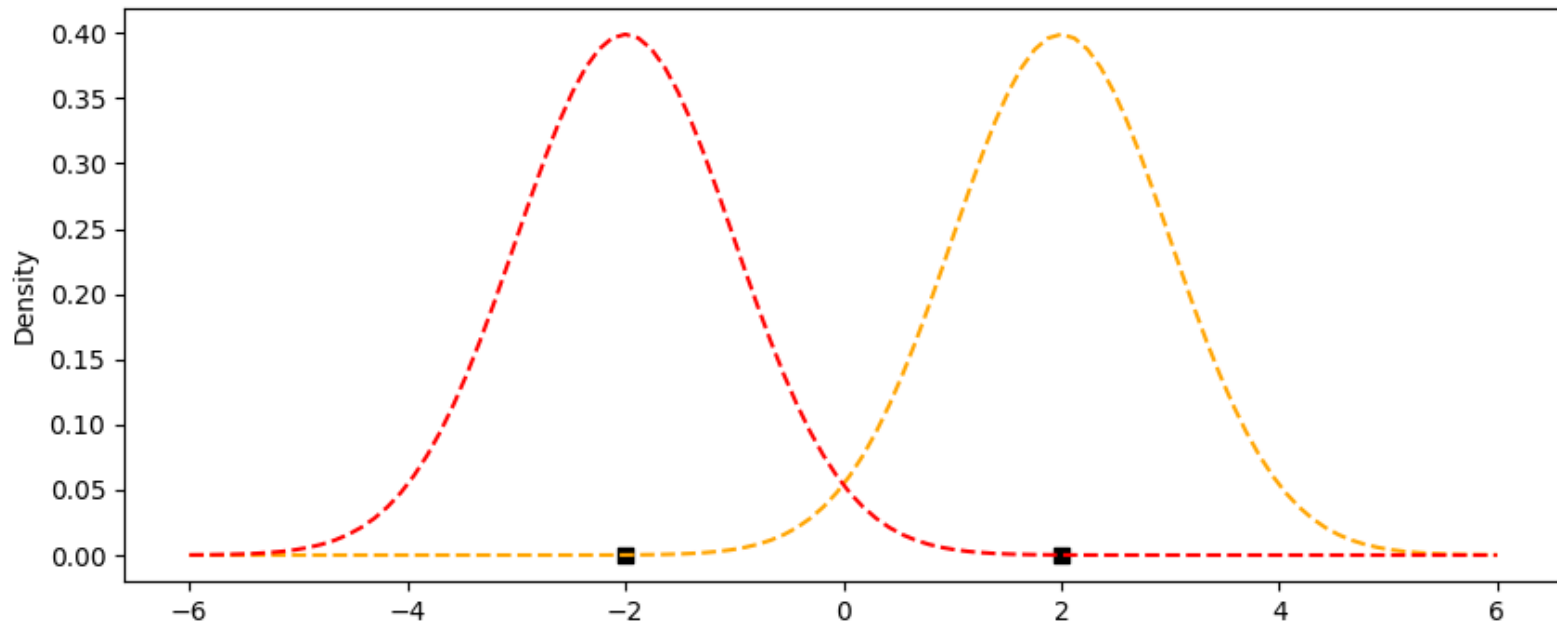
$$f(x) = \frac{1}{nh} \sum_{i=1}^n K \left( \frac{x - x_i}{h} \right)$$



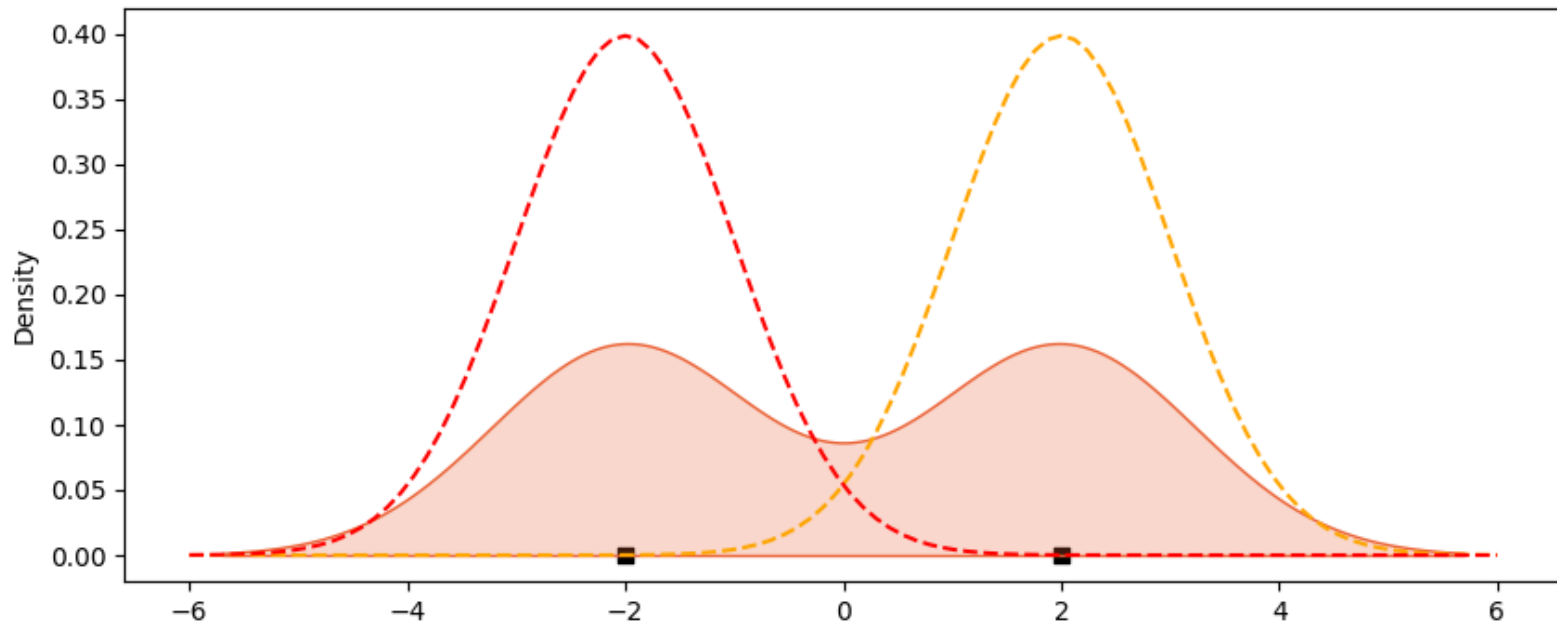
# Density plot



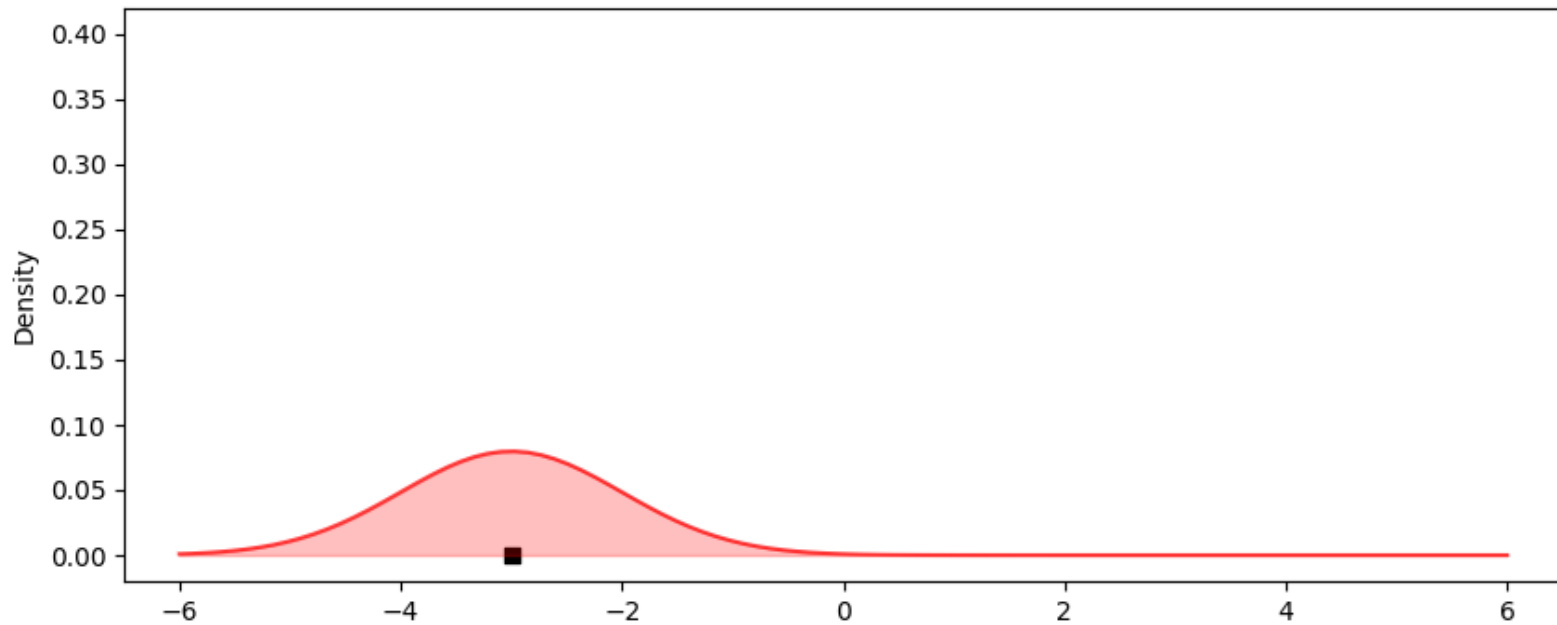
# Density plot



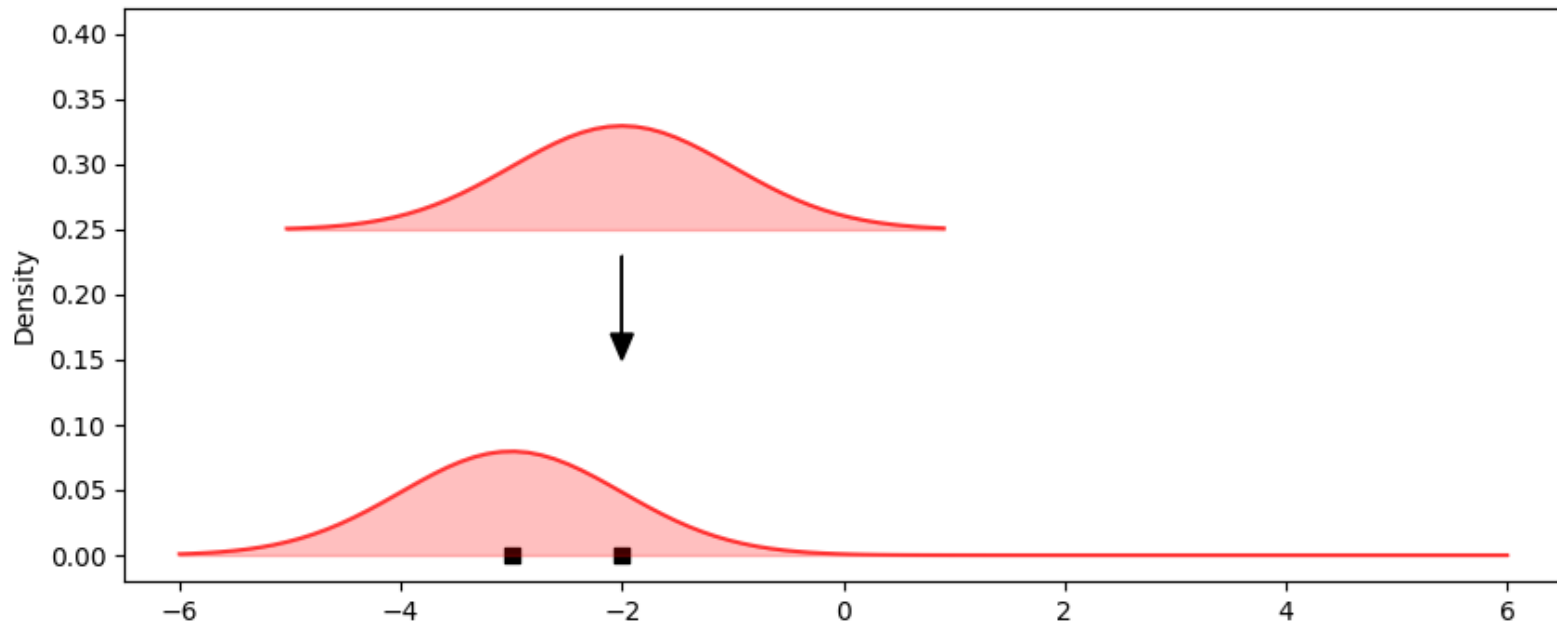
# Density plot



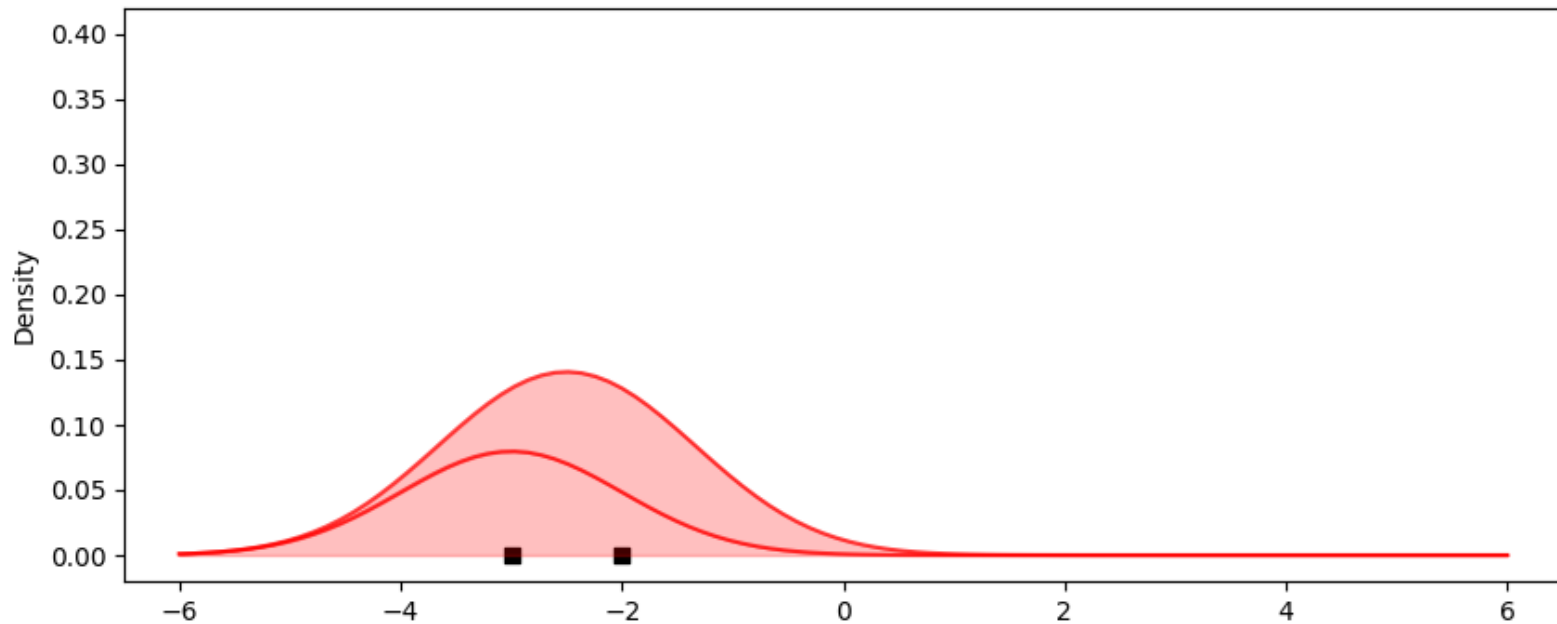
# Density plot



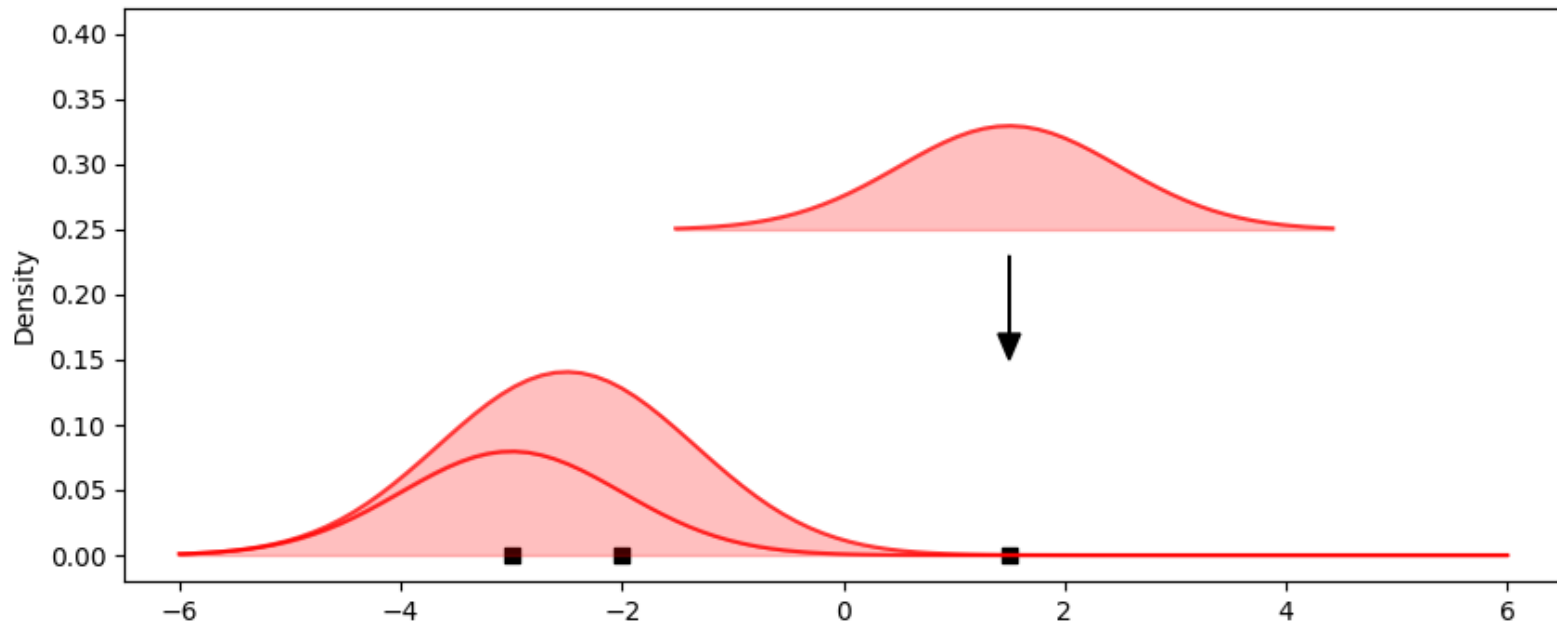
# Density plot



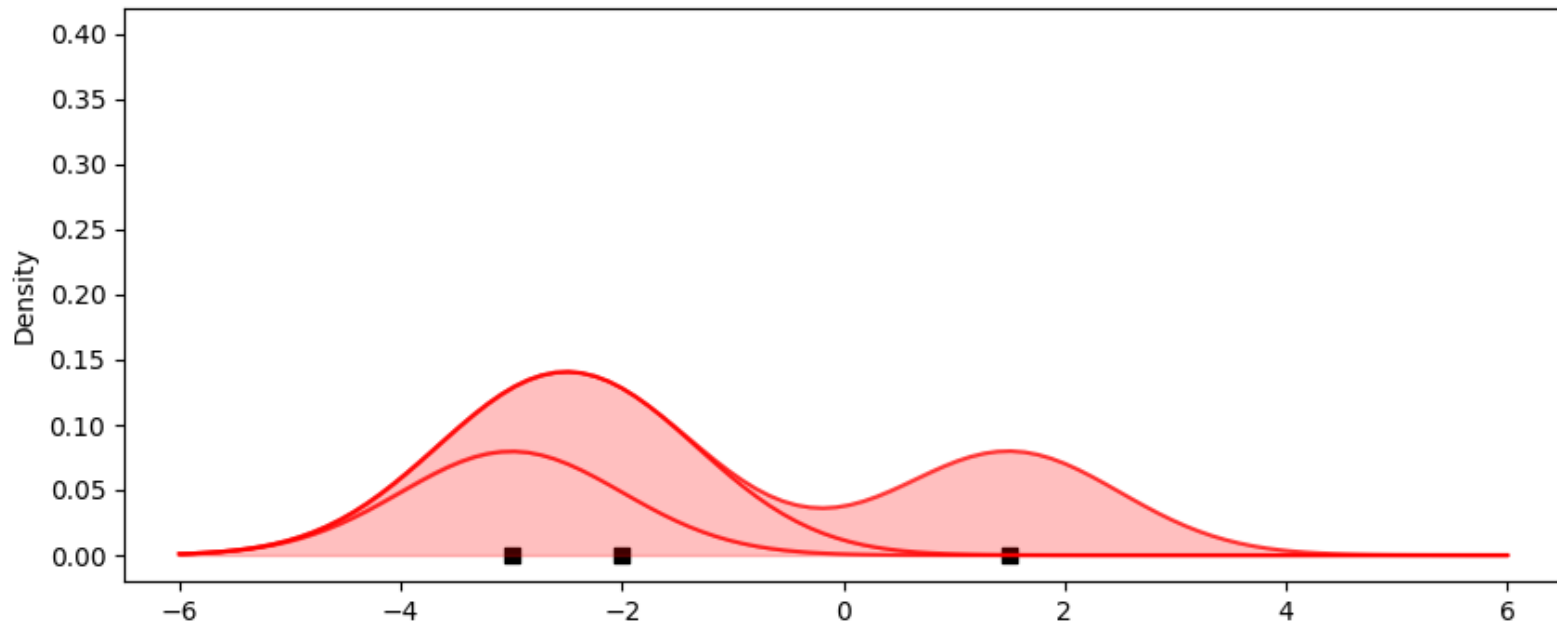
# Density plot



# Density plot

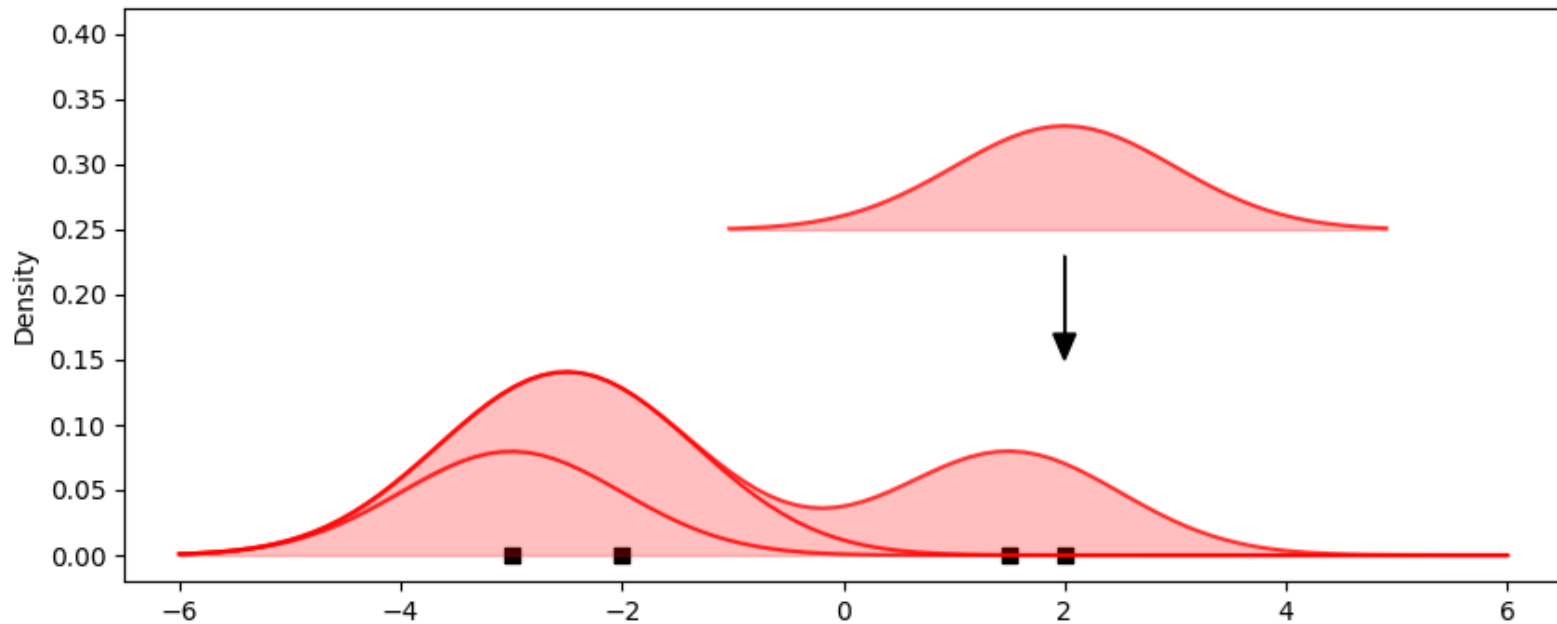


# Density plot

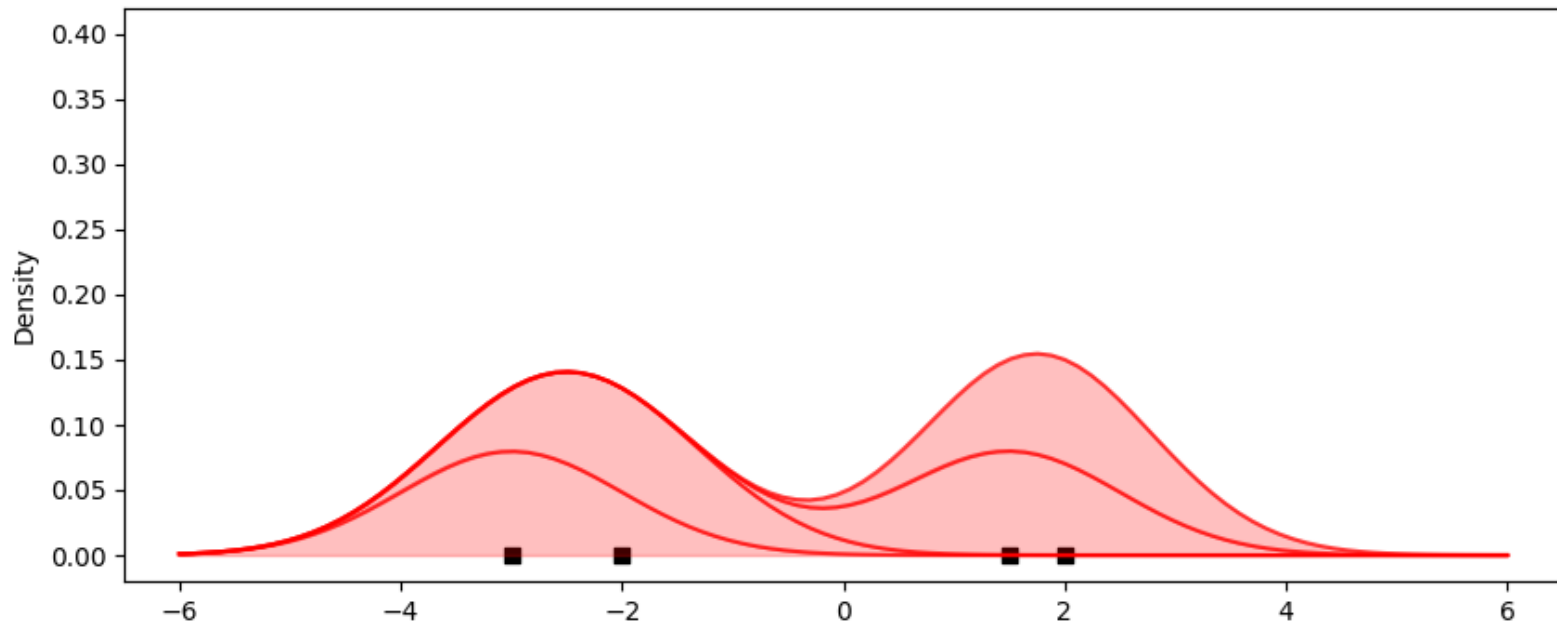




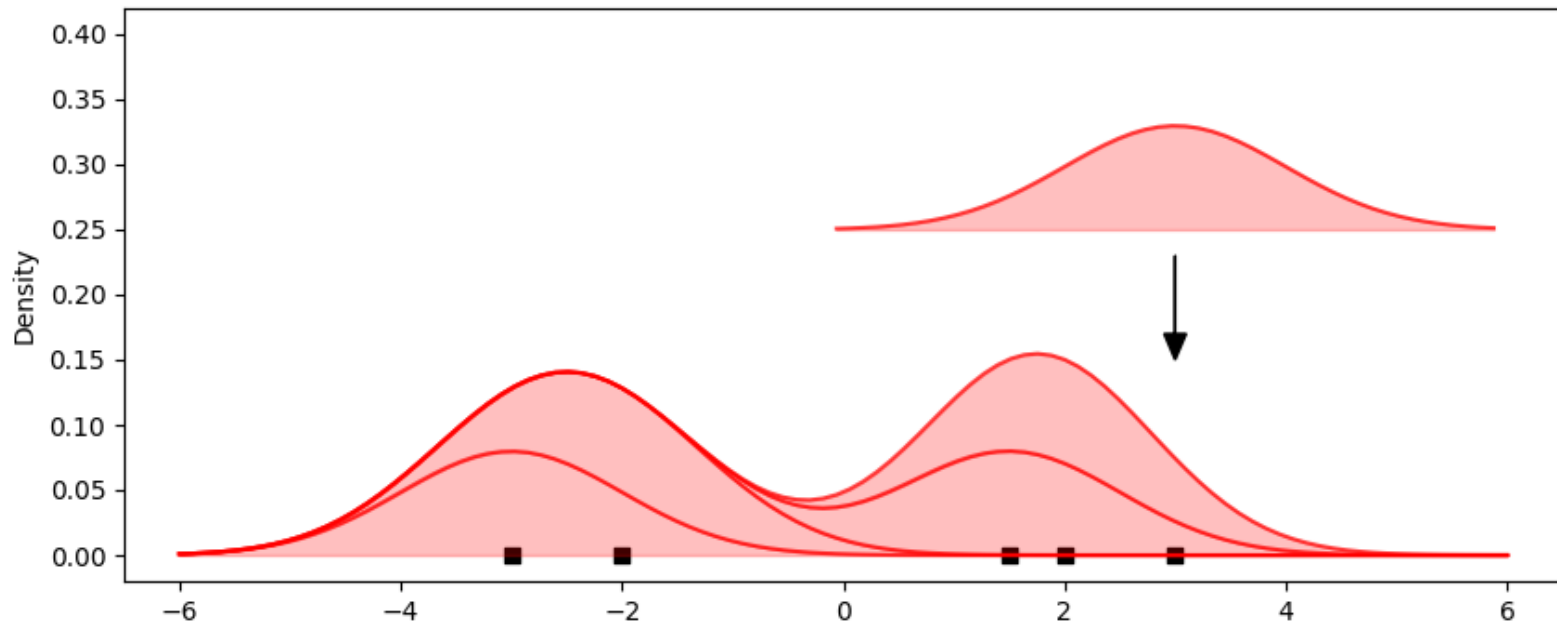
# Density plot



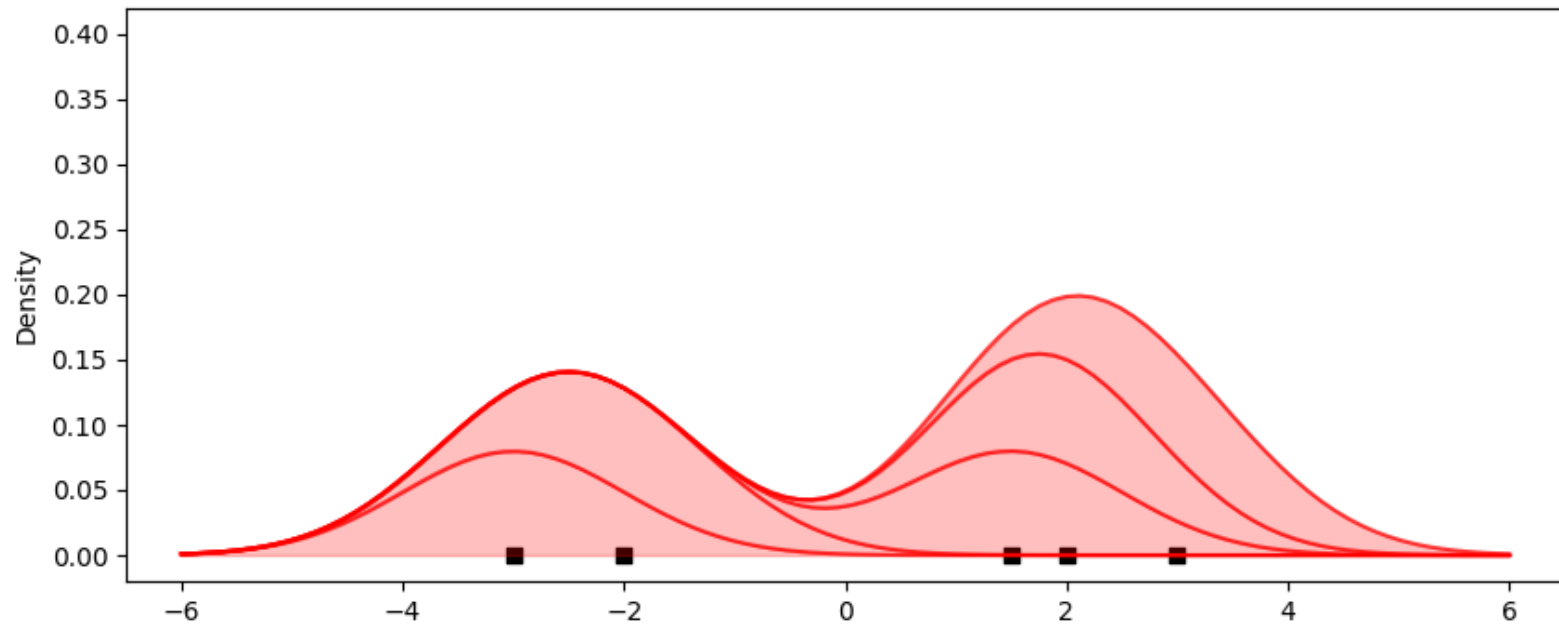
# Density plot



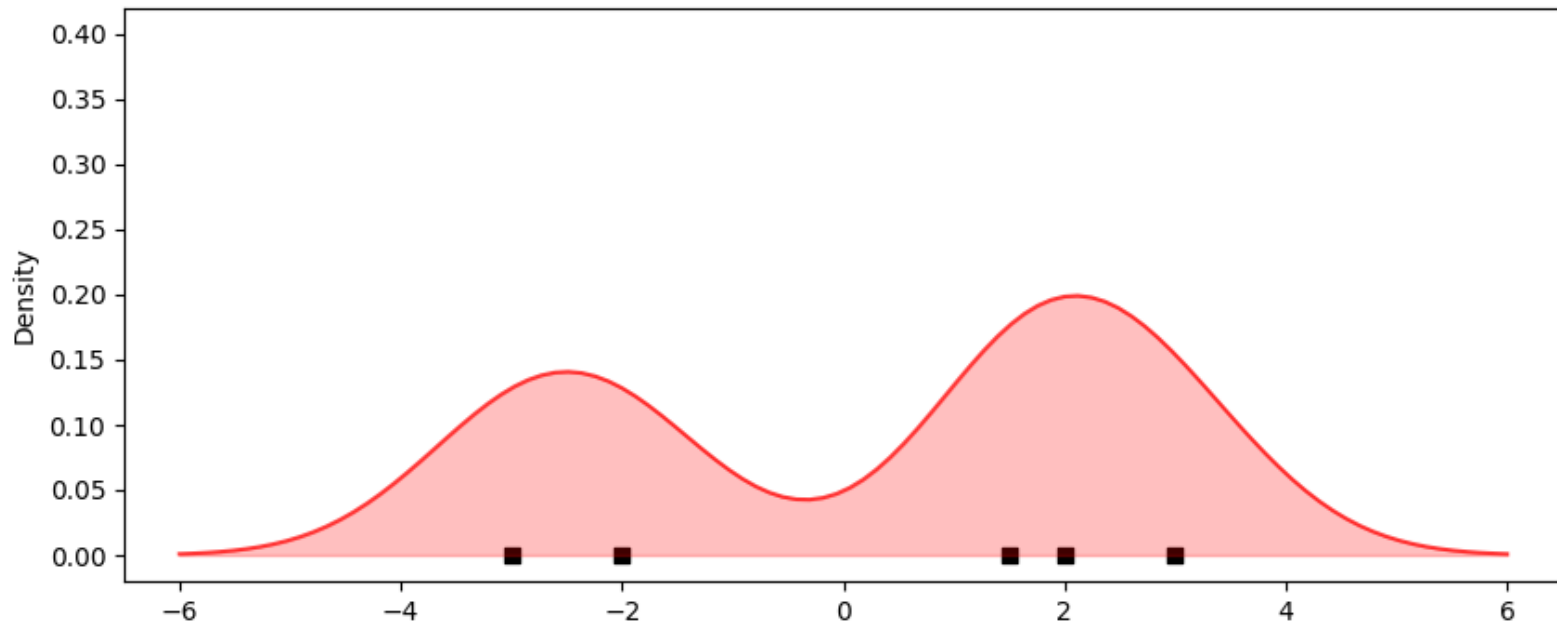
# Density plot



# Density plot

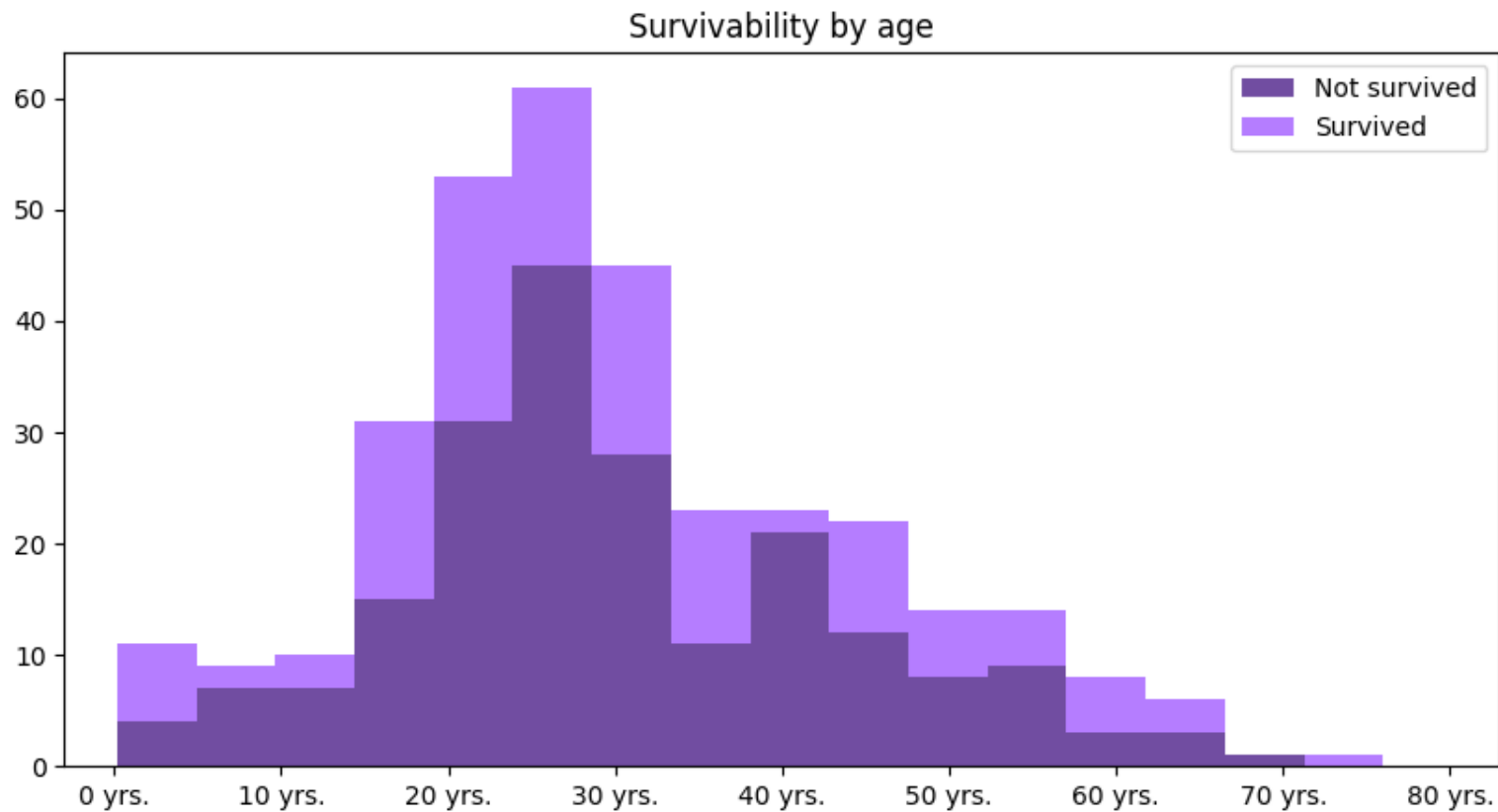


# Density plot

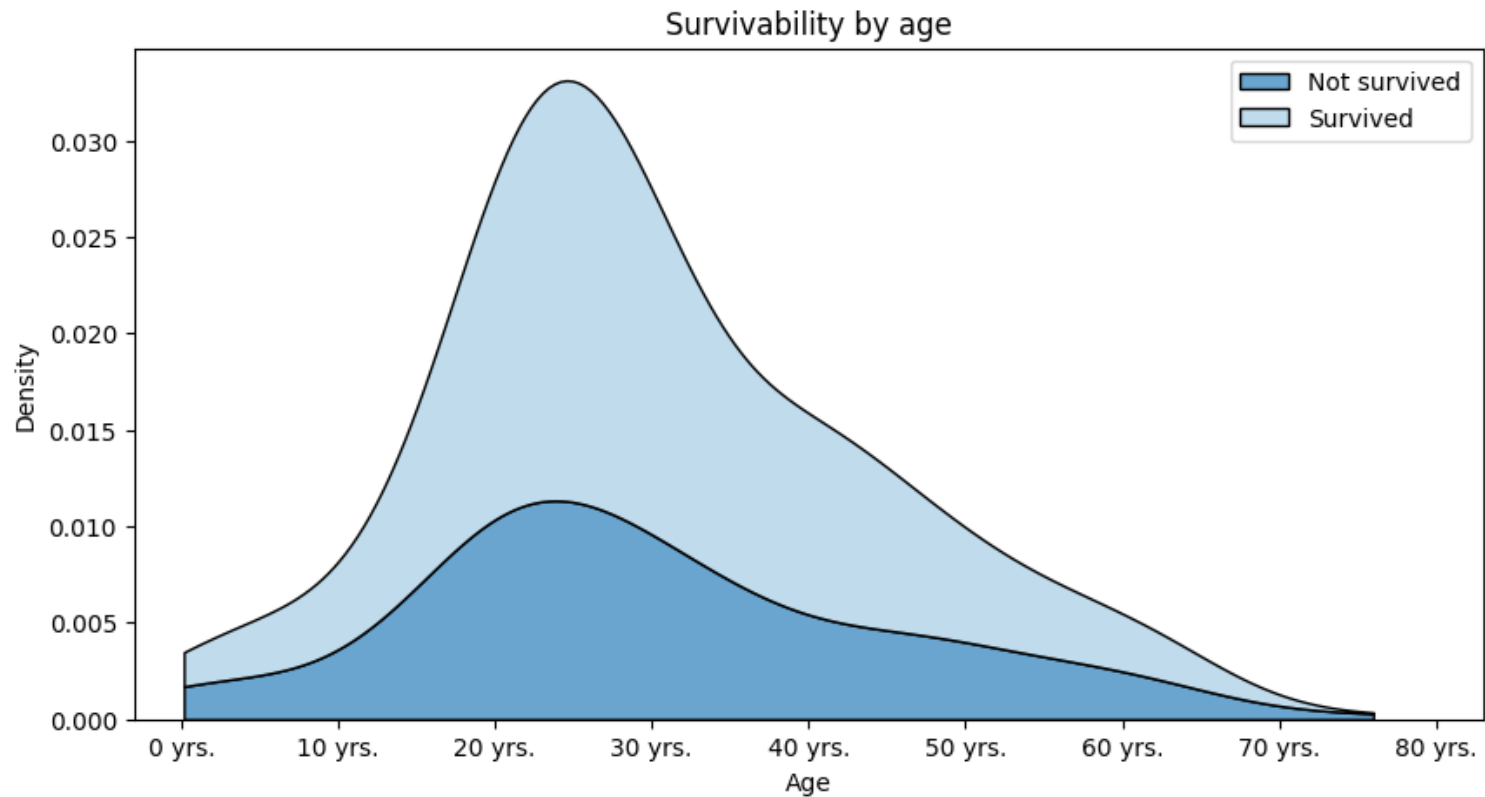


# Comparing multiple distributions

# Stacked histogram

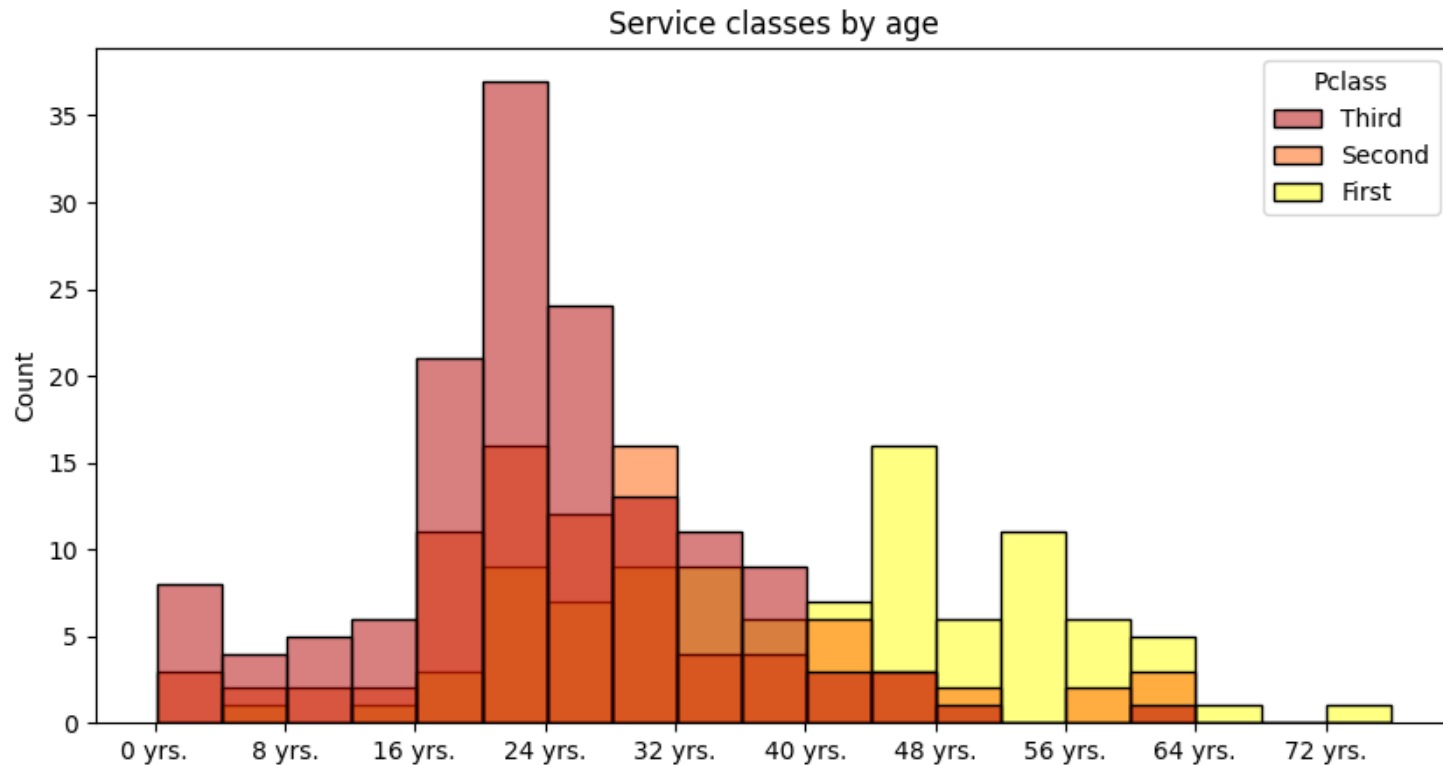


# Stacked density plot

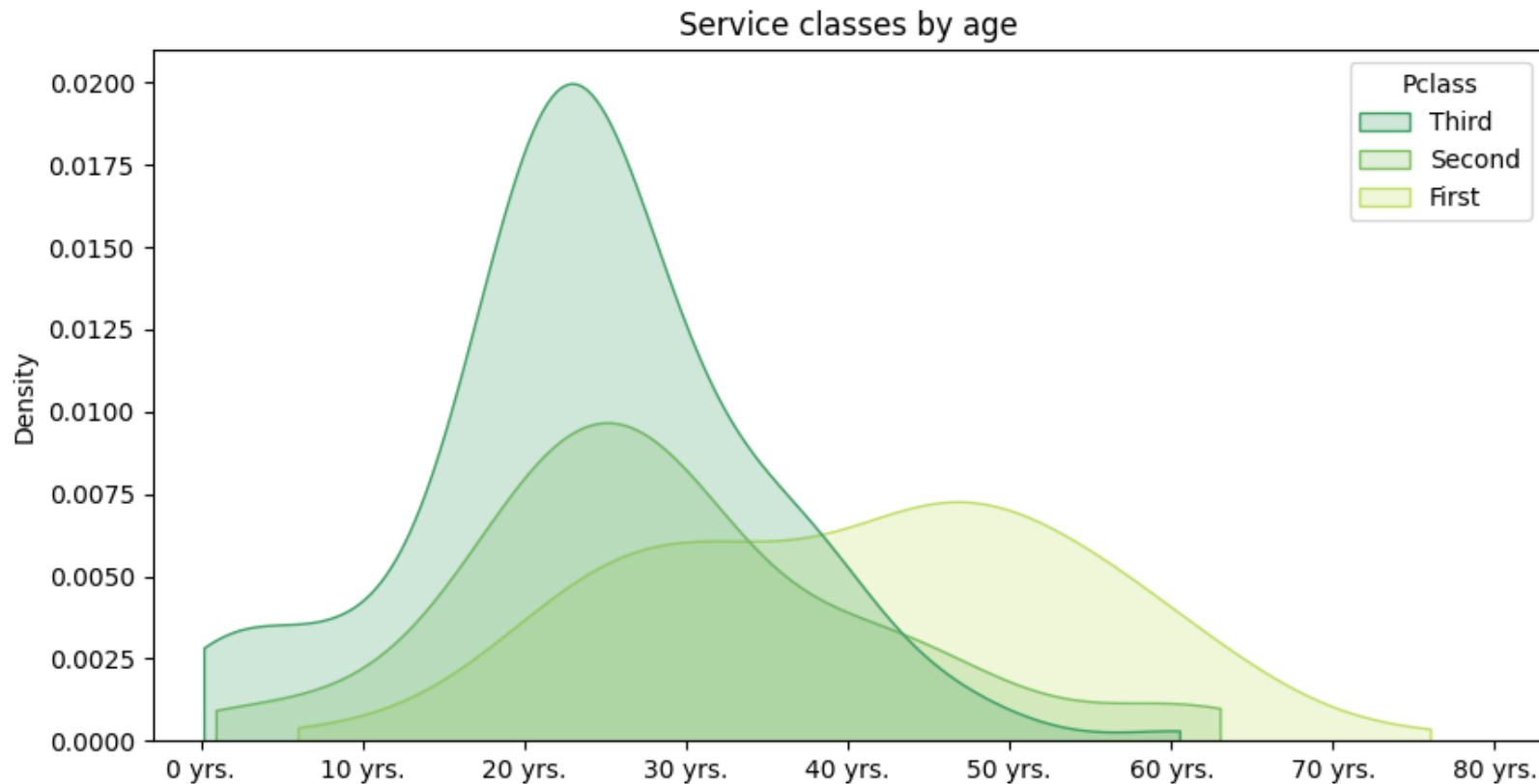




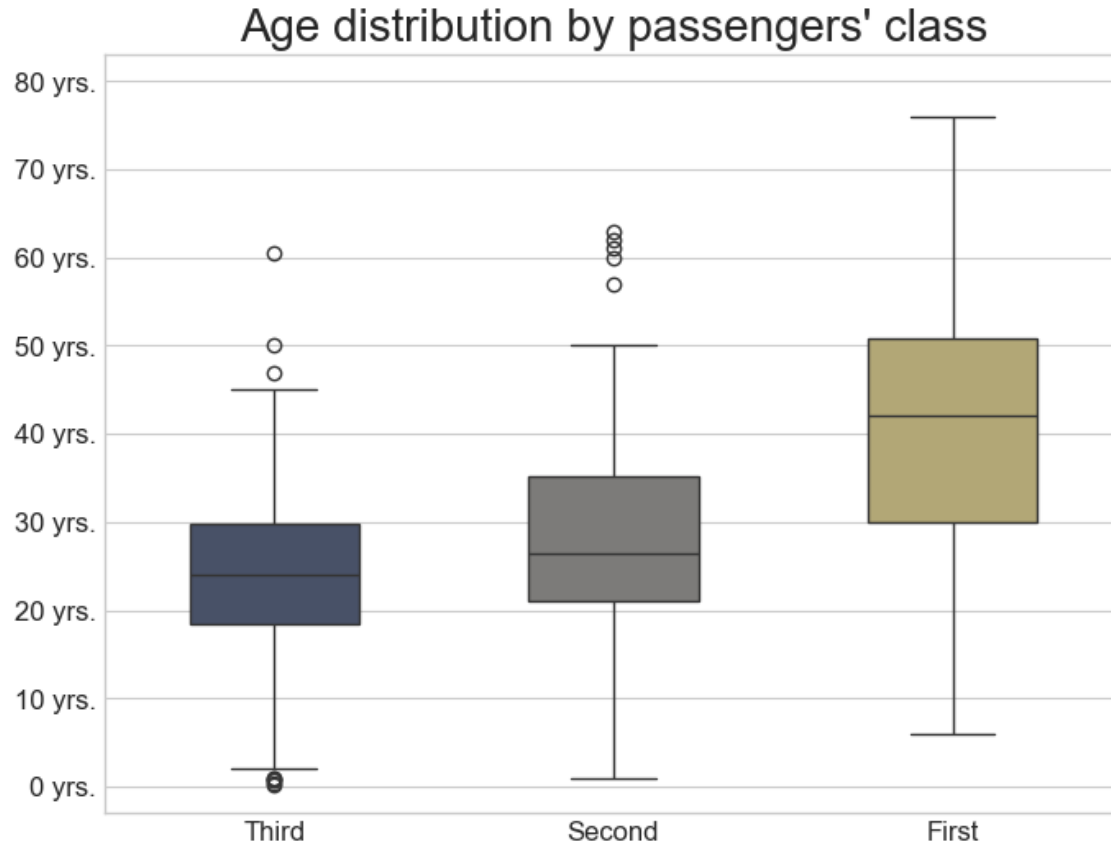
# Overlapping density plot



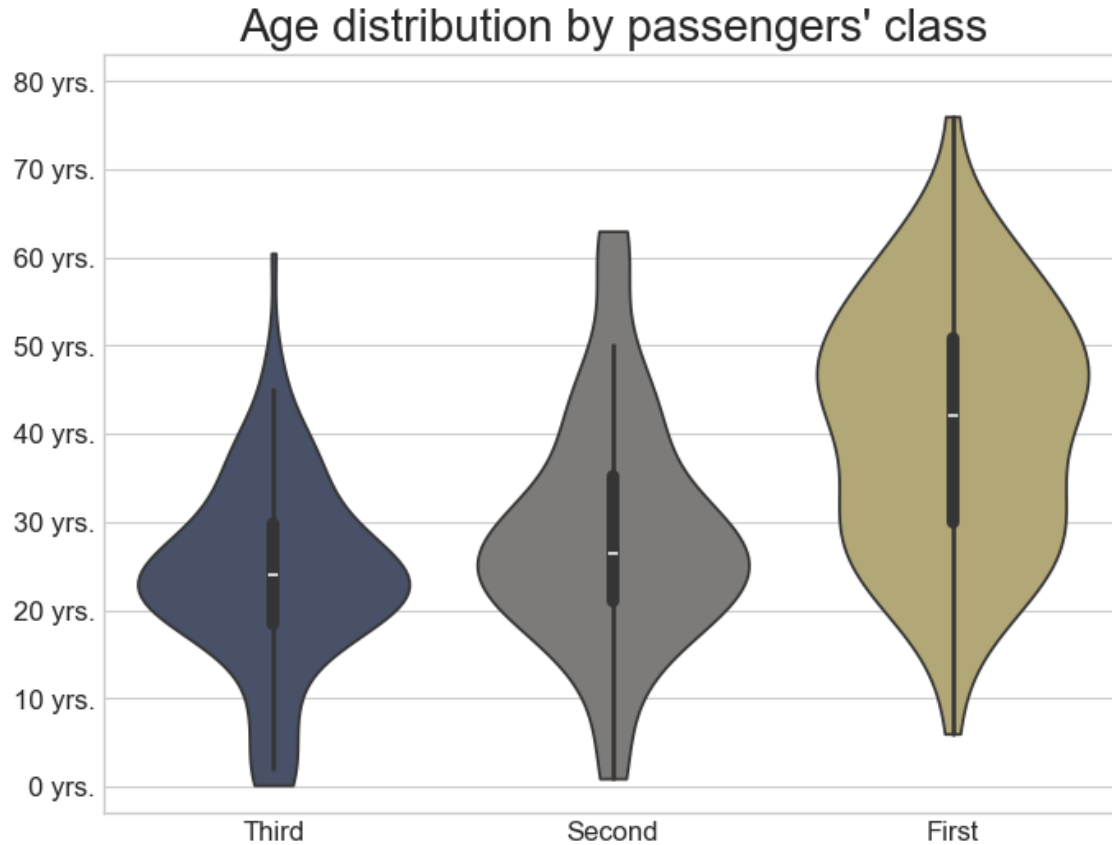
# Overlapping density plot



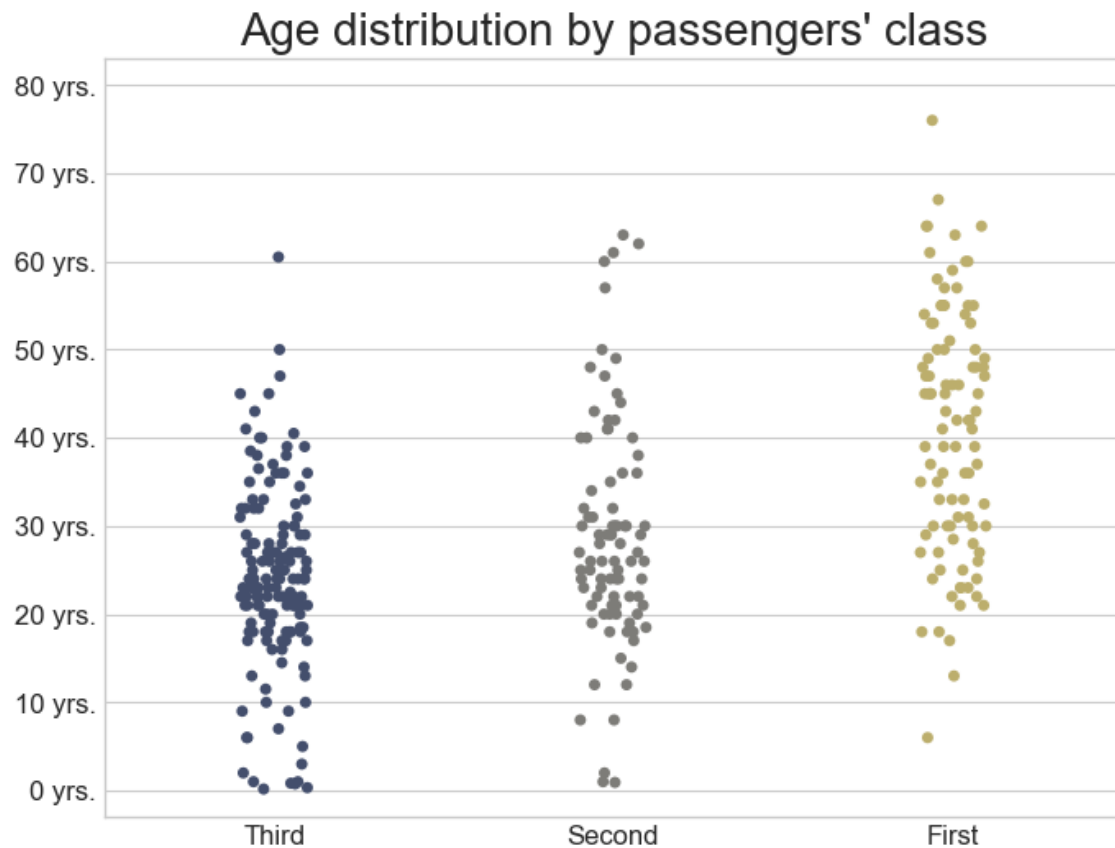
# Box plots



# Violin plot

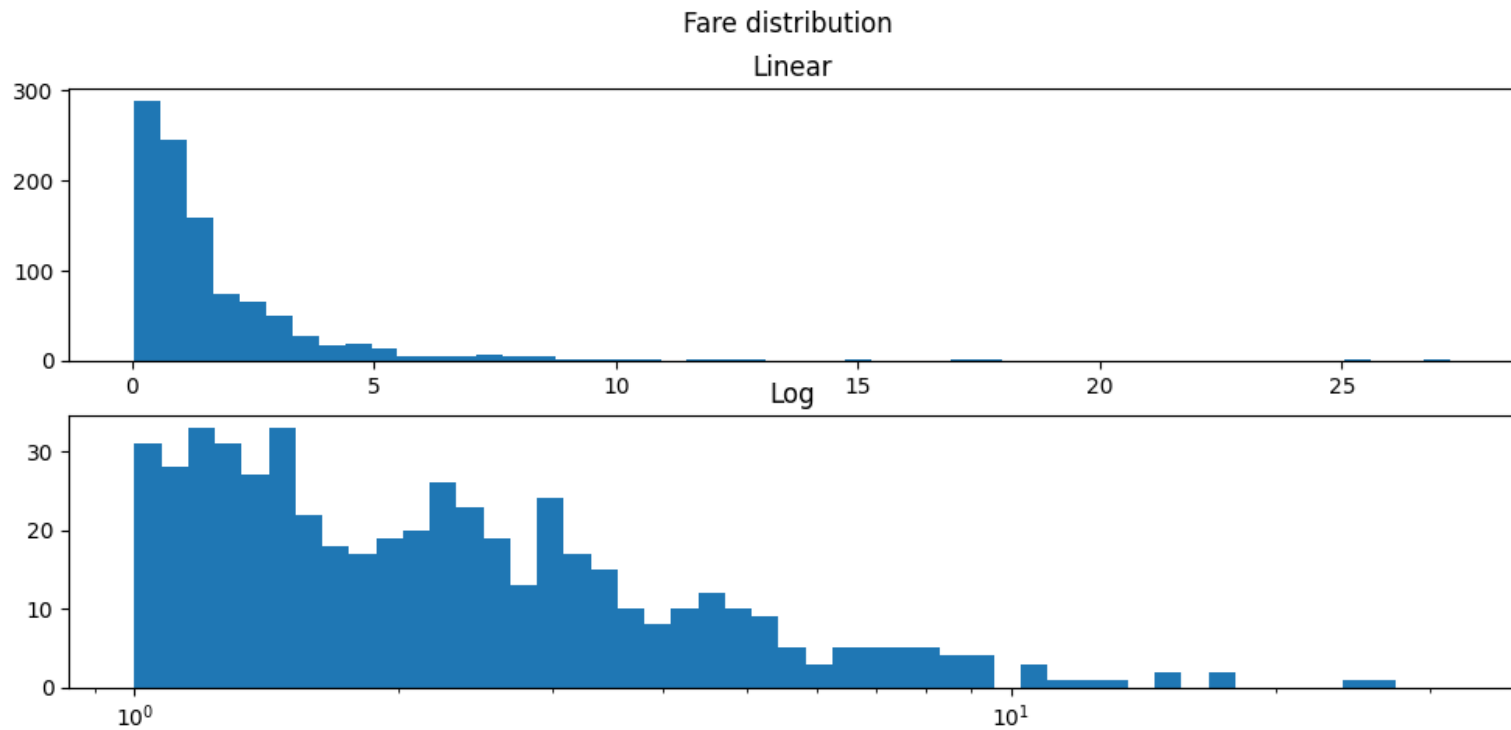


# Strip plot



Logarithmic scale

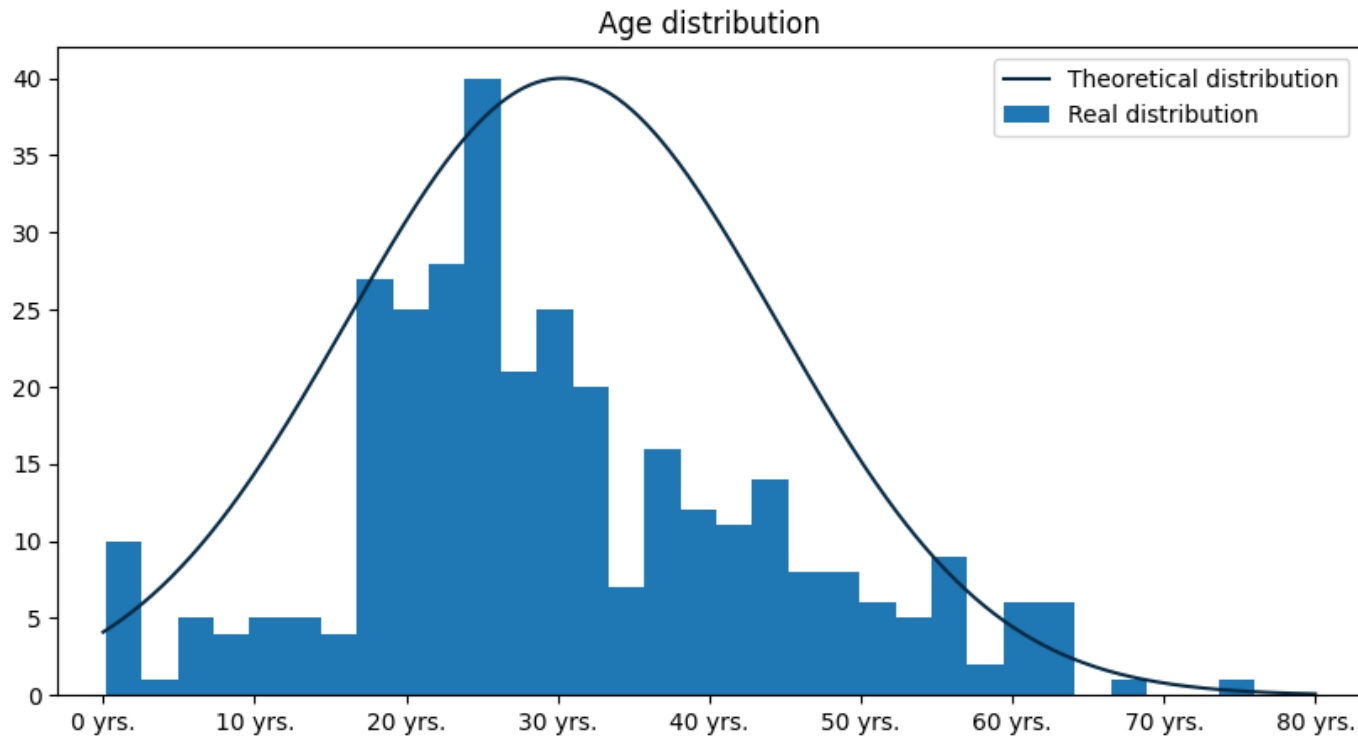
# Logarithmic scale



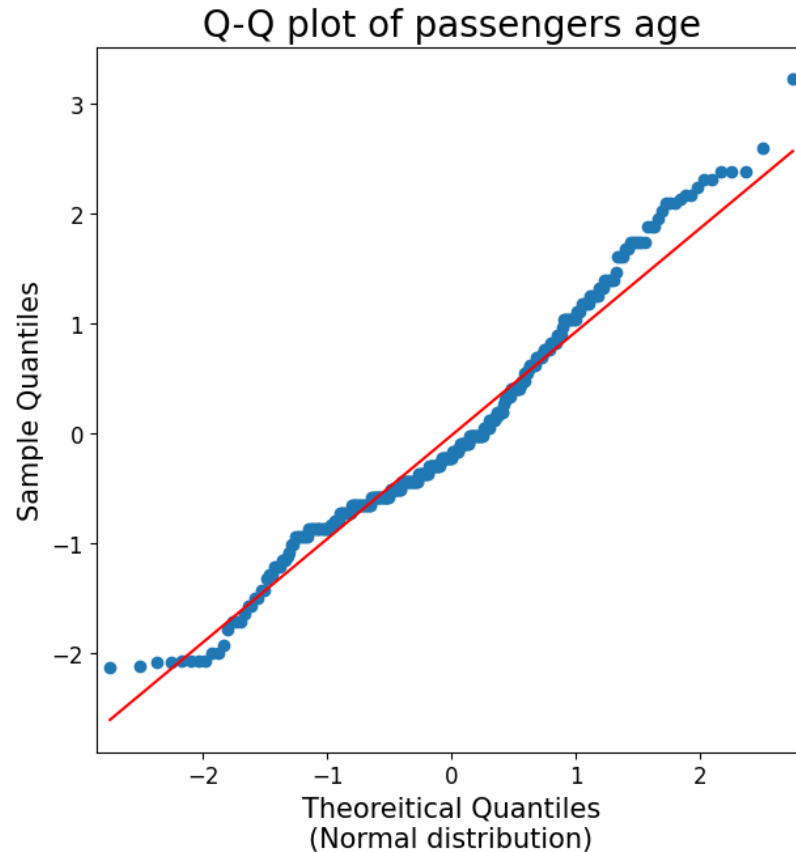
Real vs theoretical distribution



# Real vs theoretical distribution



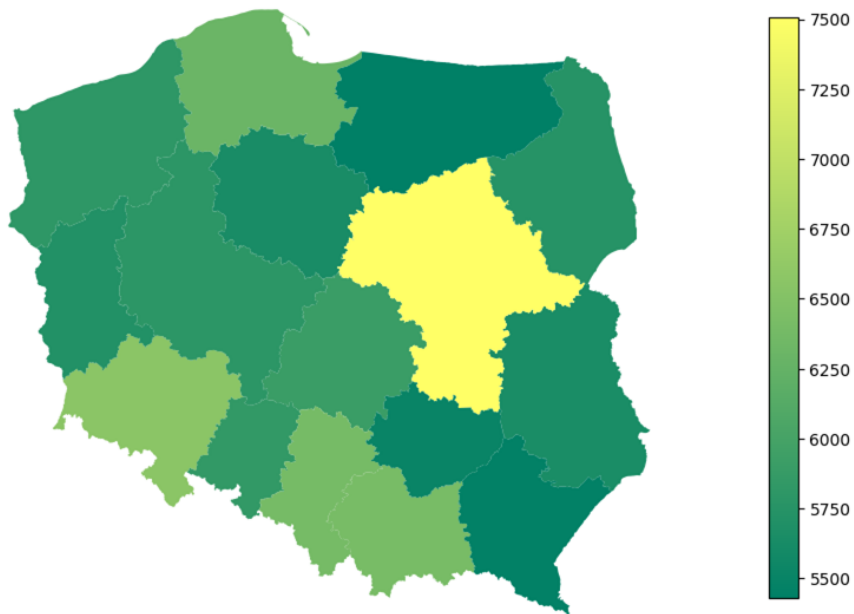
# Real vs theoretical distribution



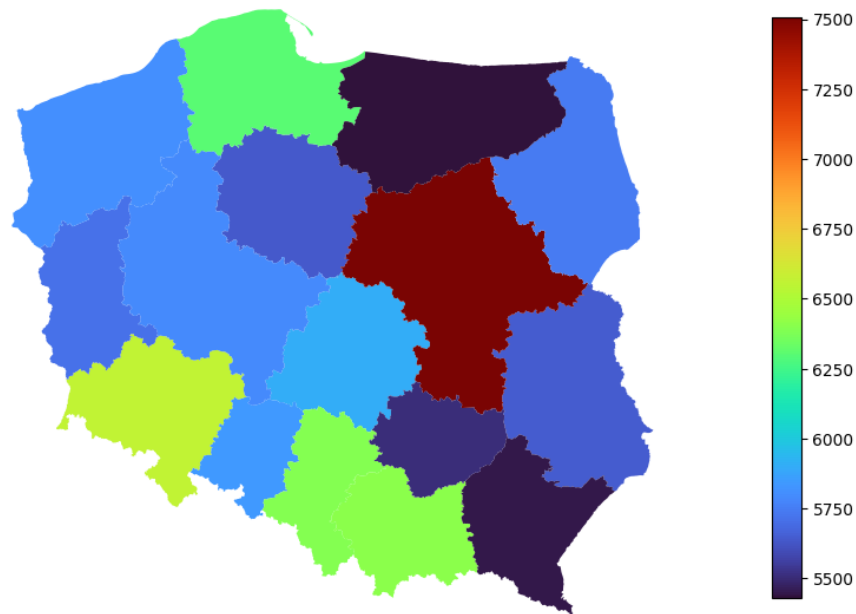
Frequently made mistakes

# Incorrect color scale

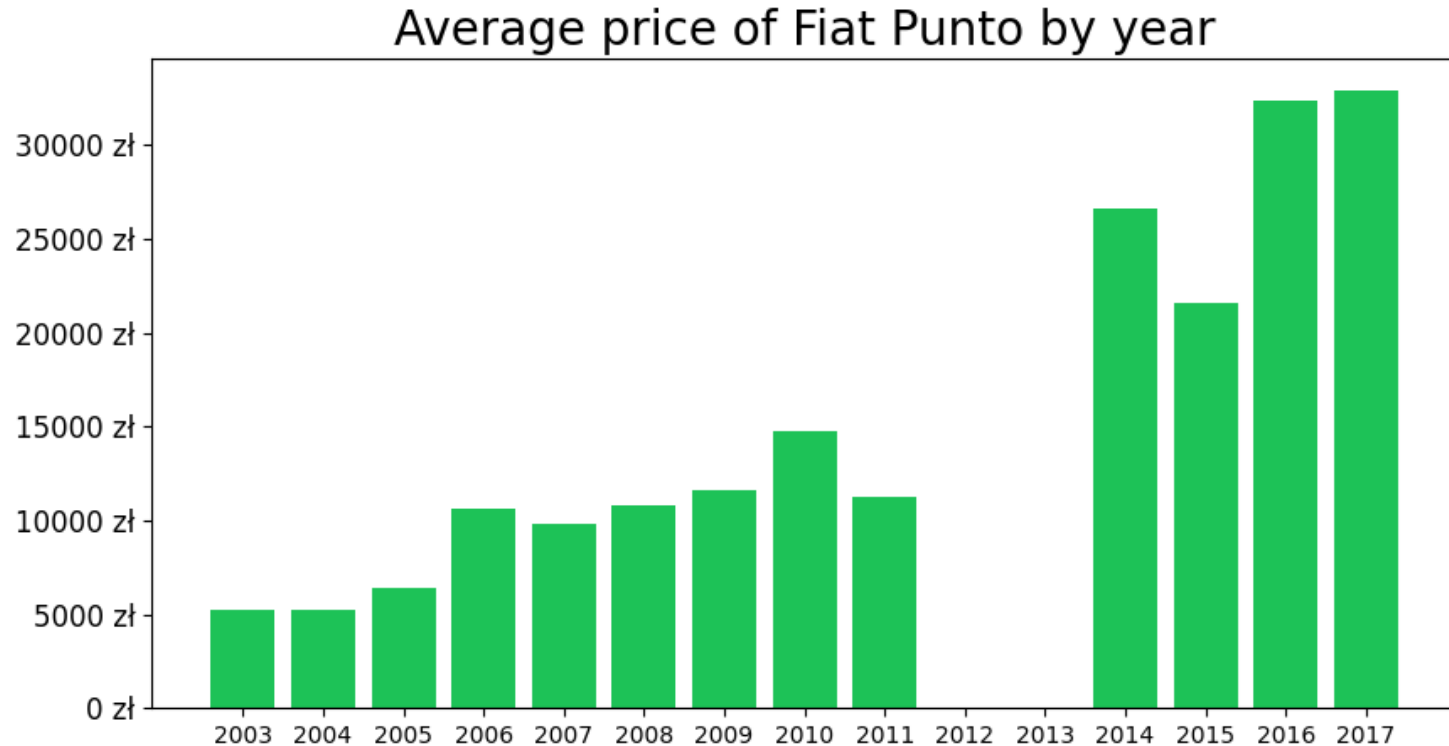
Salary by viovodership



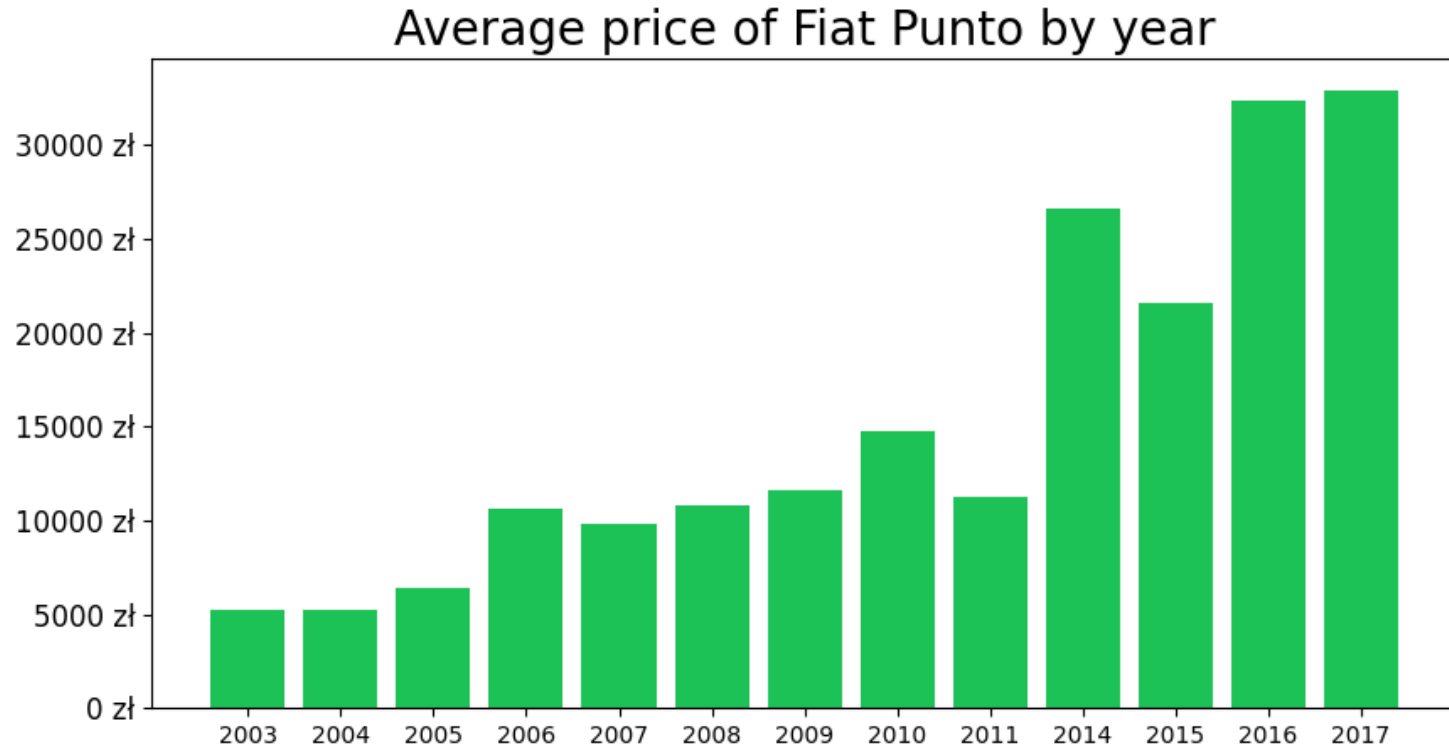
Salary by viovodership



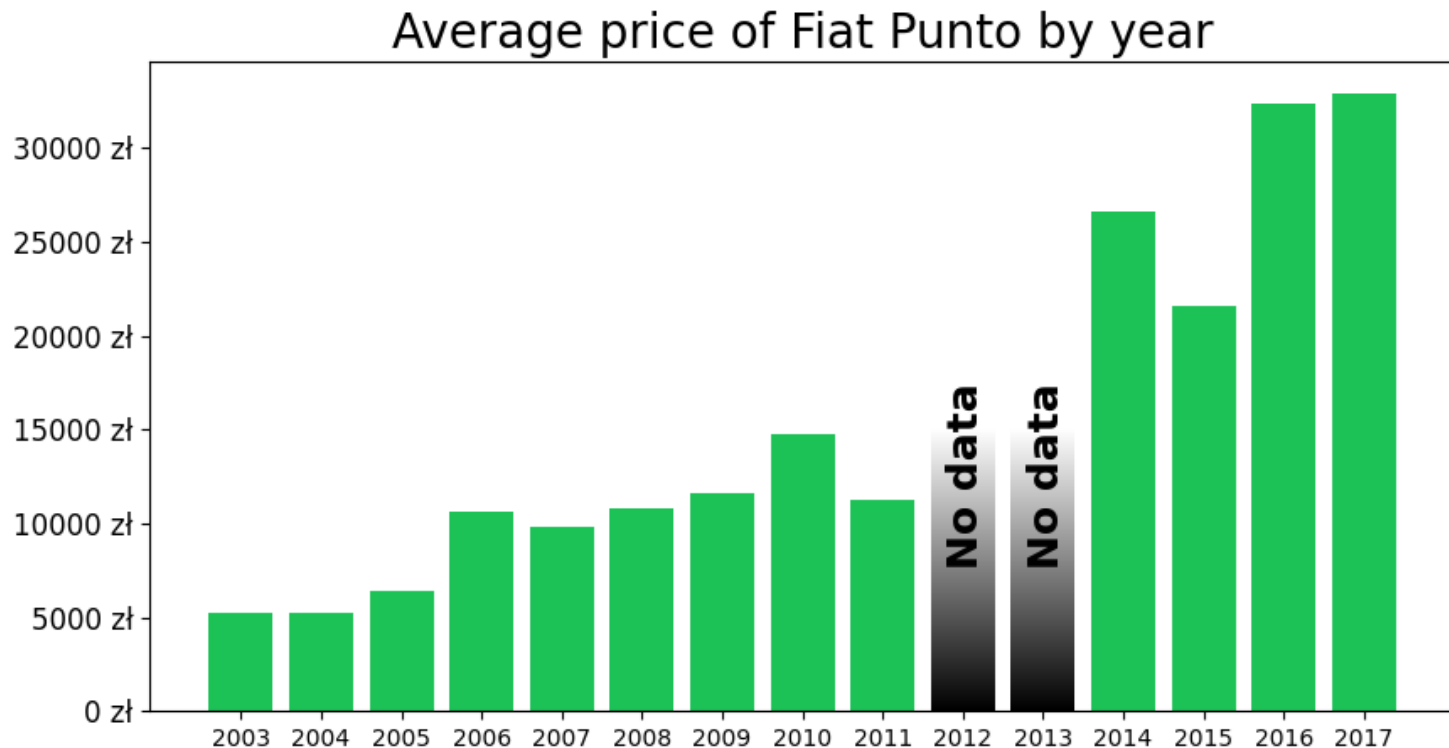
# Informing about missing data



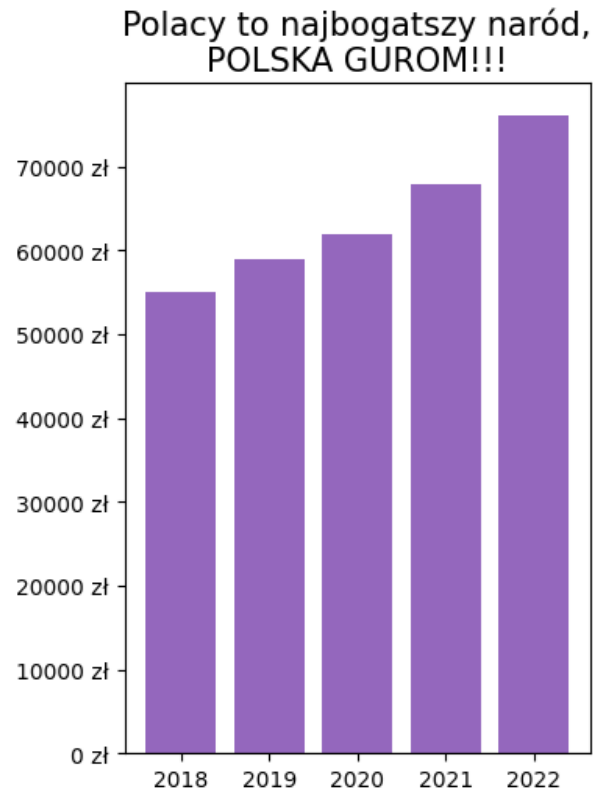
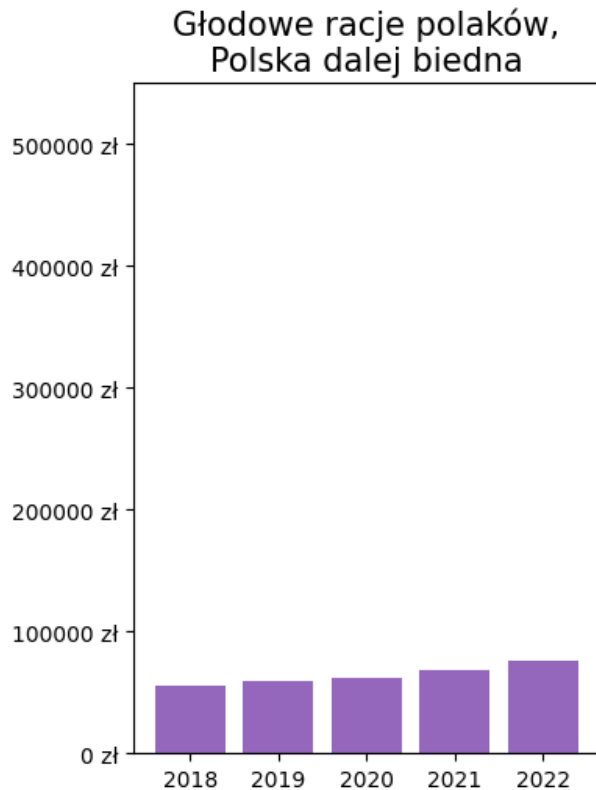
# Informing about missing data



# Informing about missing data

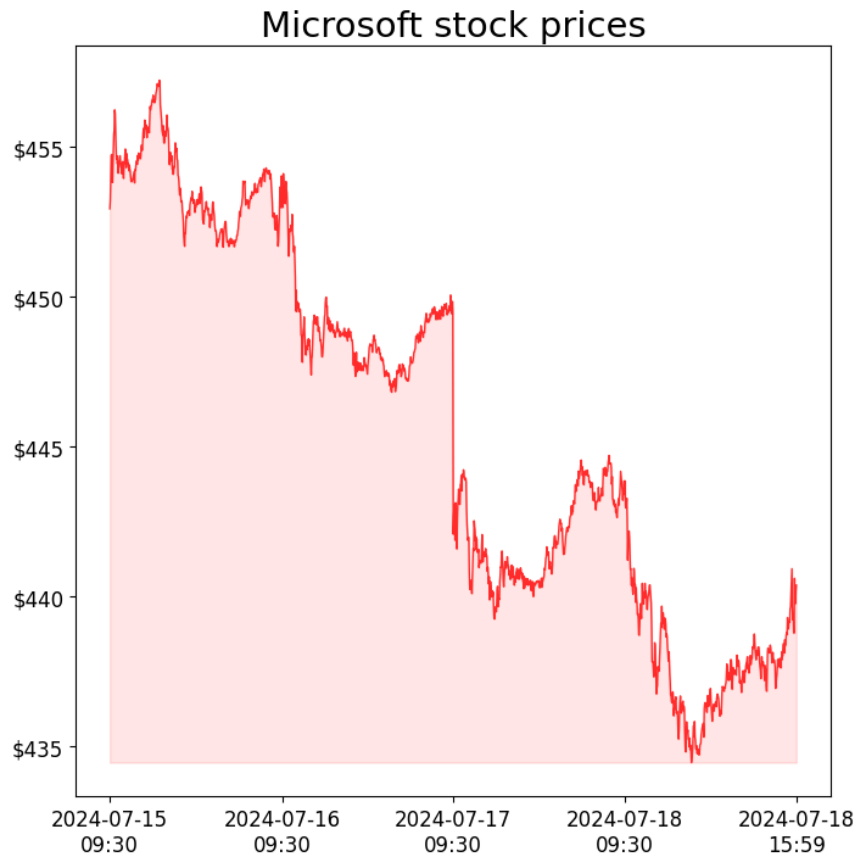


# Manipulating scale

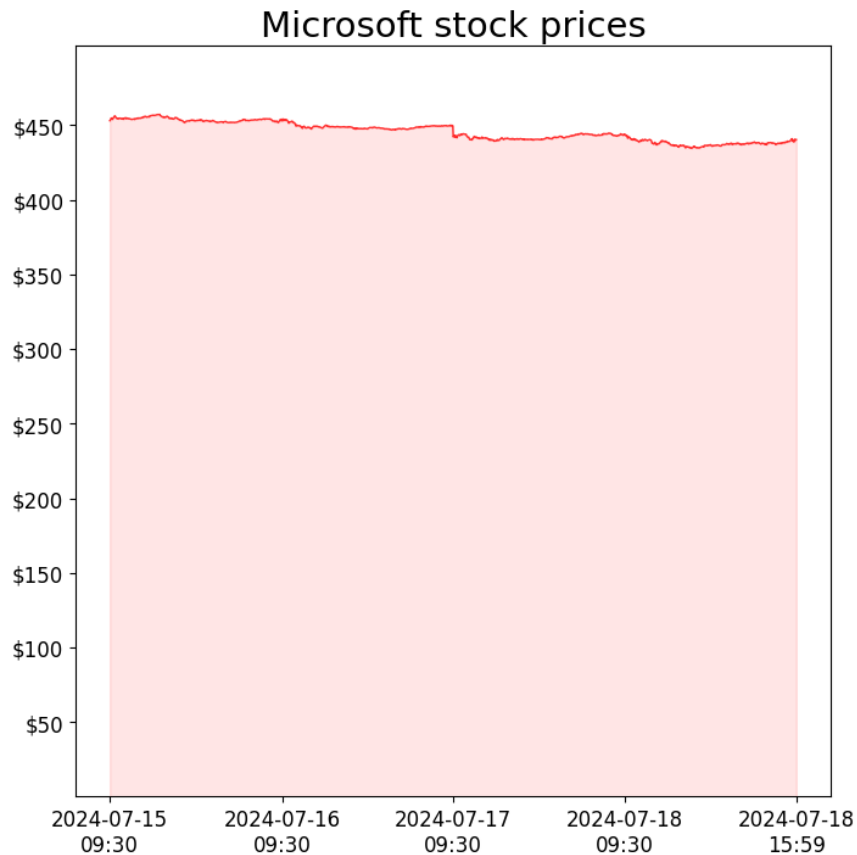




# Manipulating scale



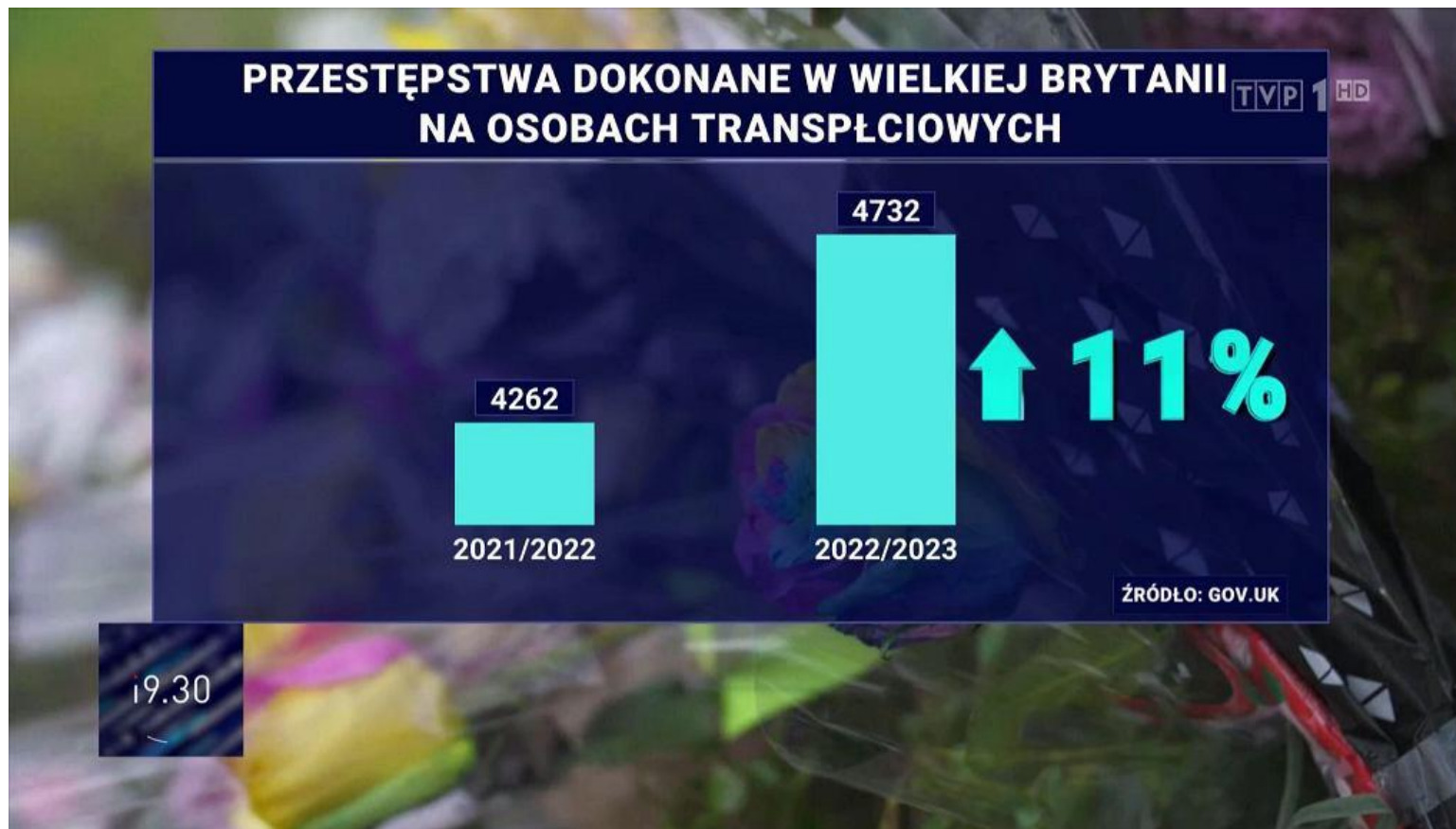
# Manipulating scale



# Manipulating scale

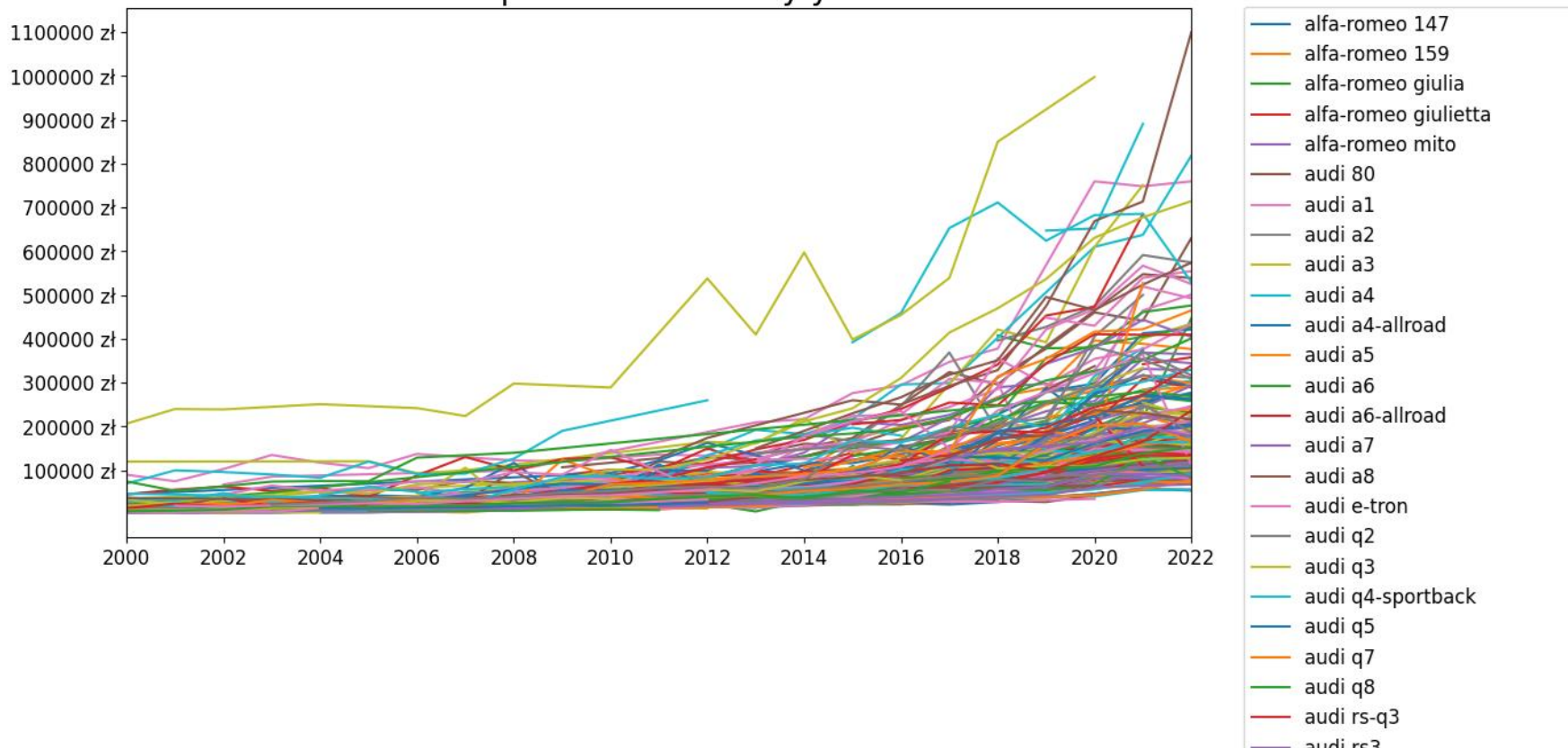


# Manipulating scale

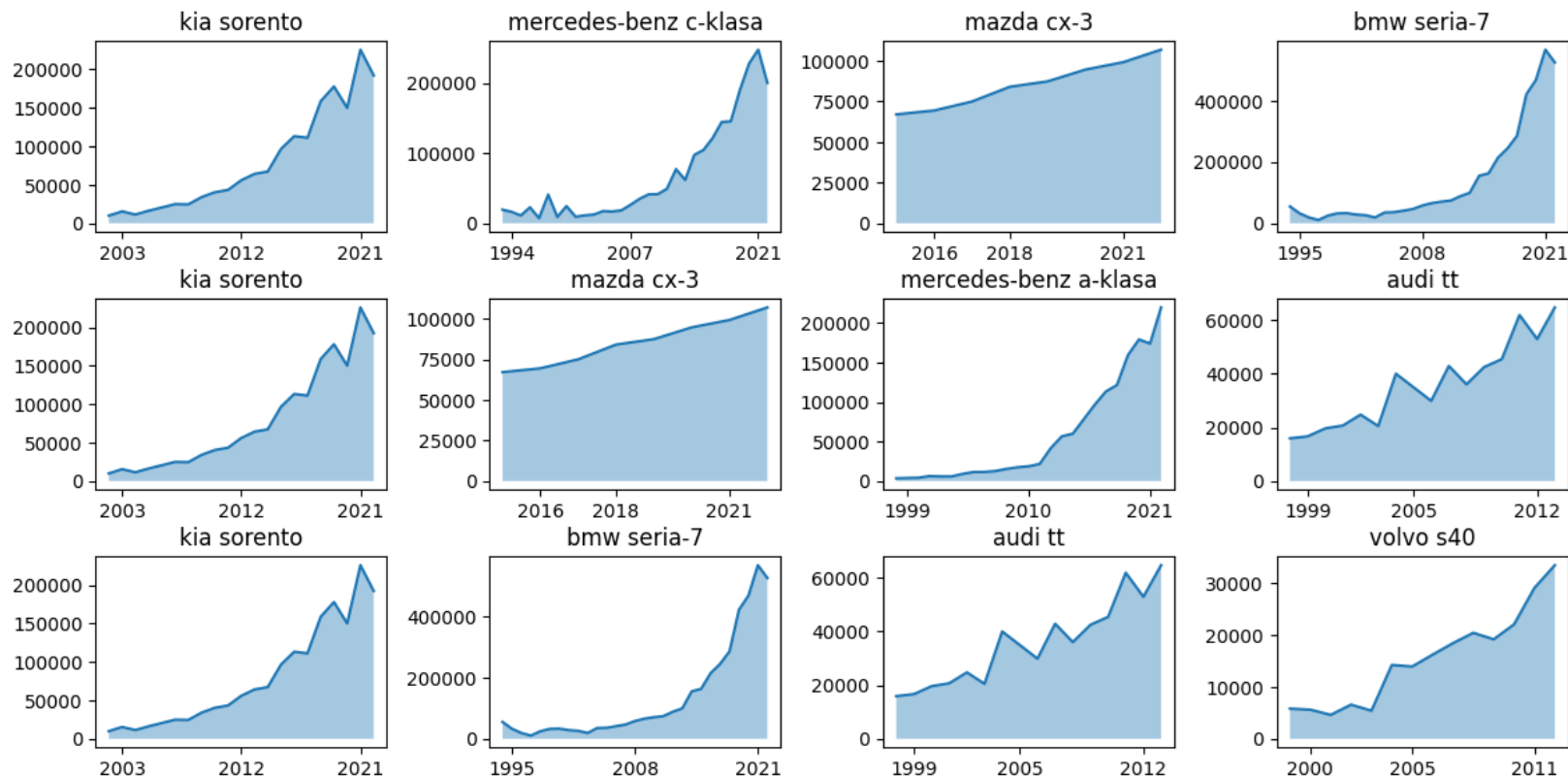


# Too many informations

Car prices in Poland by year



# Too many informations



# References

- <https://github.com/Skamlo/Visualization-Techniques-in-Python>
- <https://towardsdatascience.com/kernel-density-estimation-explained-step-by-step-7cc5b5bc4517>
- <https://www.kaggle.com/datasets/brendan45774/test-file>
- <https://stat.gov.pl/sygnalne/komunikaty-i-obwieszczenia/lista-komunikatow-i-obwieszczen/obwieszczenie-w-sprawie-wysokosci-przecietnego-miesiecznego-wynagrodzenia-brutto-w-gospodarce-narodowej-w-wojewodztwach-w-2022-roku,295,9.html>
- <https://www.zus.pl/baza-wiedzy/skladki-wskazniki-odsetki/wskazniki/przecietne-wynagrodzenie-w-latach>
- <https://www.kaggle.com/datasets/aleksandrglotov/car-prices-poland>
- <https://towardsdatascience.com/kernel-density-estimation-explained-step-by-step-7cc5b5bc4517>
- <https://www.youtube.com/watch?v=cy8r7WSuT1I>