

Exercise Sheet 3 – Data Science

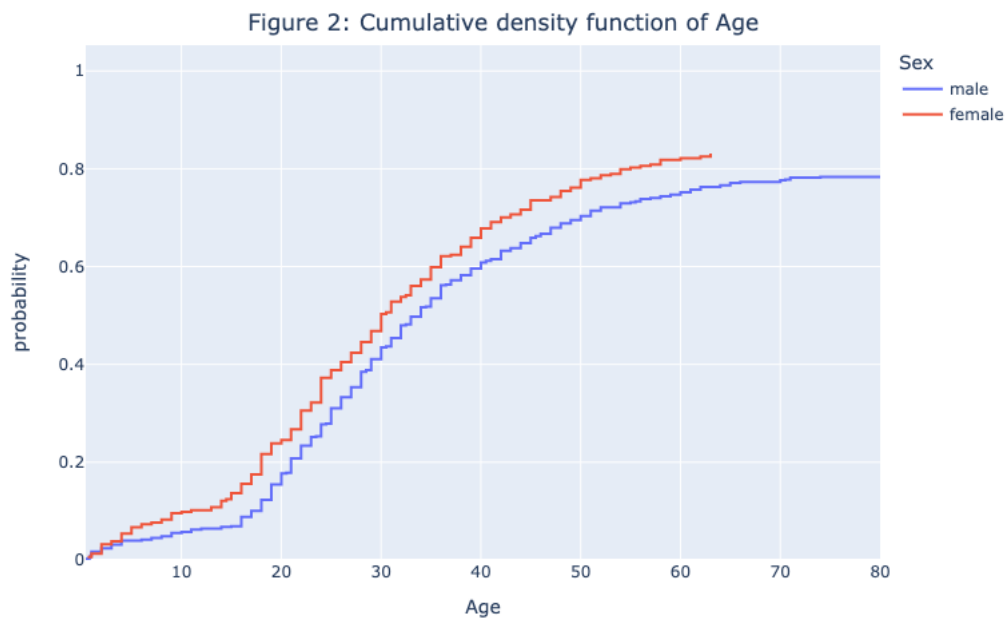
INF161 Autumn Semester 2021

Exercise 1 Re-create the following figure using the *salary-Industry.csv* data set. To this end, proceed as follows:

- Set axis text as you see in the diagram.
- Select markers, marker size, market line and marker color as you see in the diagram.
- Align the title properly(in the middle) as you see in the diagram.
- Find the trend line using linear regression.

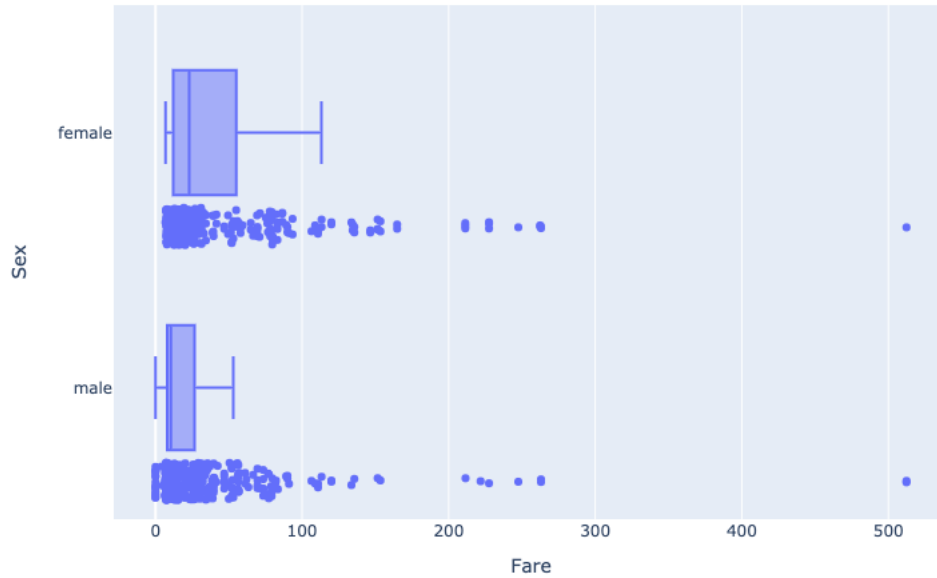


Exercise 2 Re-create the following figure using the *titanic.csv* data set by setting the axis text and aligning the title properly as you see in diagram.



Exercise 3 Re-create the following figure using the `titanic.csv` data set by setting the axis text and aligning the title properly as you see in diagram. Read off (and write in your answer to this problem) the median and the 25% quantile for male, female passengers.

Figure 3: Box plot with displaying the underlying data



Exercise 4 When the Titanic sunk, the probability of surviving was different for people with cheap tickets (i.e., with a low ticket fare) and for those with more expensive tickets (i.e., with a high fare value in the data set). Create visualization(s) to figure out:

- Whether having a cheaper ticket increases or decreases the probability of survival
- How strong the effect is(i.e. how much does it increase or decrease the probability)

Motivate your answers using the visualization(s) you have created.