Exercise Sheet 2 – Visualization

INF161 Autumn Semester 2020

Exercise 1 *Re-create the following figure using the* `salary.csv` *data set.*

- *Select markers, line type, and colors yourself*
- *Format axis tick marks and text as you see fit*
- *Find the trend line using linear regression*

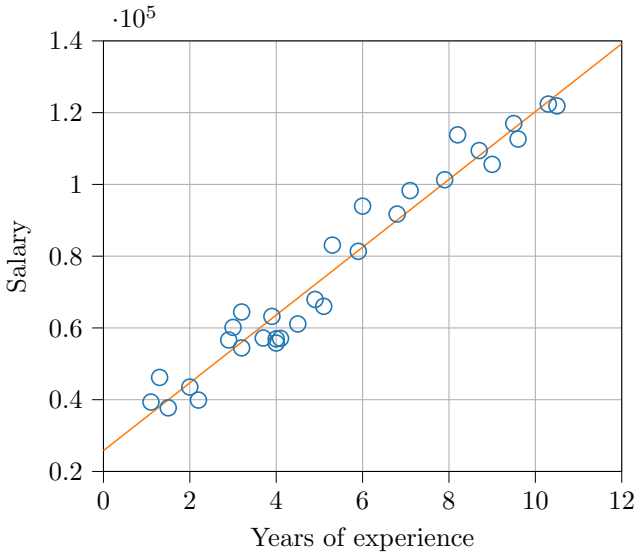

Figure 1: Years of experience versus salary.

Exercise 2 *Re-create the following figure using the* `titanic.csv` *data set. Select line type, color yourself, and format axis tick marks and text as you see fit.*

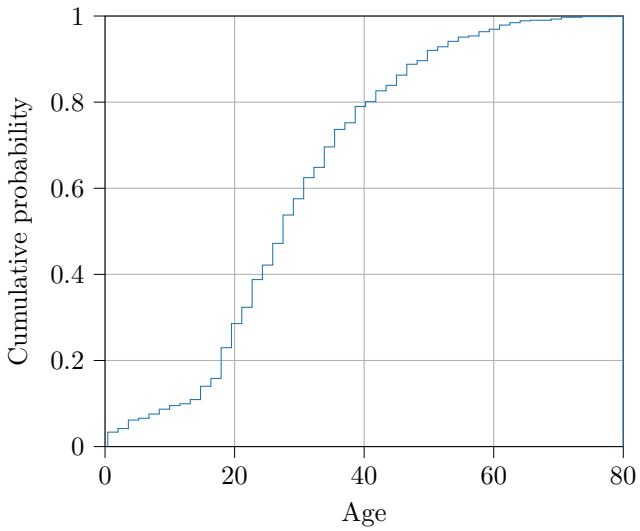

Figure 2: Titanic passenger age cumulative distribution function

Exercise 3 *Re-create the following figure using the* `titanic.csv` *data set and then read off (and write in your answer to this problem) the median and the* 25% *quantile for male, female, and all passengers.*

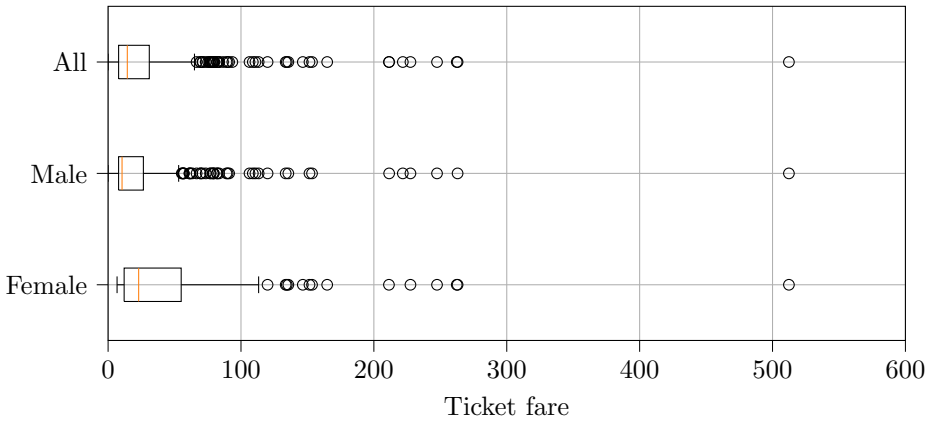

Figure 3: Titanic ticket fare distribution for male, female, and all passengers.

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*

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

Exercise 5 *Create a Jupyter notebook to visualize the Coronavirus, with data from:*

- *https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset*

For each visualization you create you should write down

- *Why you think this visualization is interesting*
- *What conclusions and insights can be drawn from the visualization*

You should at least add (use the `COVID19_line_list_data.csv` *file for these):*

- *A histogram over the age of the patients*
- *A plot showing the fraction of male and female patients*

Beyond that it is up to you how far you want to go. The sky is the limit here; there is no end to the number of ways to visualize this data. See this page for some examples:

- *https://informationisbeautiful.net/visualizations/covid-19-coronavirus-infographic-datapack/*