|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ **Информатика, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ и системы  
 управления**

КАФЕДРА **Компьютерные системы и сети (ИУ6)**

НАПРАВЛЕНИЕ ПОДГОТОВКИ **09.04.01 Информатика и вычислительная техника**

МАГИСТЕРСКАЯ ПРОГРАММА **09.04.01/07 Интеллектуальные системы анализа,**

**обработки и интерпретации больших данных**

**Отчет**

**по лабораторной работе №2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | ИУ6-32М |  |  | А.М. Панфилкин |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | М.А. Захаров |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

**Задание**

1. Постройте лестницу из случайного полученного количества ступенек при помощи Scatter Plots, при том, что длина каждой ступени увеличивается в 2 раза по отношению к предыдущей.
2. Круговая диаграмма по числу жертв по городам и поддиаграмма по расе
3. Боксплот распределения по возрастам среди женщин для разных рас

**Выполнение**

Листинг 1 – Выполнение задания

|  |
| --- |
| import numpy as np  import matplotlib.pyplot as plt  import pandas as pd  import seaborn as sns  homicide\_data = pd.read\_csv('data/homicide-data.csv', encoding='ISO-8859-1')  # Постройте лестницу из случайного полученного количества  # ступенек при помощи ​Scatter Plots, при том, что длина каждой ступени  # увеличивается в 2 раза по отношению к предыдущей.  # например координаты точек ступенек будут такими:  def steps(n, l=1):      p = 2 \*\* np.arange(n+1) - 1      x = np.vstack((p[:-1], p[1:])).T.ravel()\*l      y = np.repeat(np.arange(n), 2)      return x, y  x, y = steps(np.random.randint(2, 10), 1.5)  plt.scatter(x, y)  for i in range(1, len(x), 2):      plt.hlines(y[i], x[i-1], x[i], color='r', linestyles='dashed')  # Круговая диаграмма по числу жертв по городам и поддиаграмма по расе  # cols: reported\_date,victim\_last,victim\_first,victim\_race,victim\_age,victim\_sex,city,state,lat,lon,disposition  # Группировка данных по городам и расам жертв  grouped\_city\_race = homicide\_data.groupby(['city', 'victim\_race']).size().unstack(fill\_value=0)  top\_10\_cities = homicide\_data['city'].value\_counts().head(10).index  # Выбор данных только для топ 10 городов  top\_10\_grouped = grouped\_city\_race.loc[top\_10\_cities]  # Создание графика  fig, ax = plt.subplots(figsize=(10, 7))  # Внешний круг (кольцо - количество жертв по топ 10 городам)  outer\_circle = ax.pie(top\_10\_grouped.sum(axis=1), labels=top\_10\_grouped.index, autopct='%1.1f%%', pctdistance=0.85, wedgeprops=dict(width=0.3, edgecolor='w'))  # Внутренний круг (кольцо - распределение расы жертв в каждом городе)  inner\_circle = ax.pie(top\_10\_grouped.values.flatten(), autopct='%1.1f%%', pctdistance=0.65, wedgeprops=dict(width=0.3, edgecolor='w'))  # Добавление заголовка  ax.set\_title('Кольцевая диаграмма: количество жертв по городам и распределение по расам')  plt.show()  # Боксплот распределения по возрастам среди женщин для разных рас  homicide\_data = homicide\_data[homicide\_data['victim\_age'] != 'Unknown']  homicide\_data['victim\_age'] = homicide\_data['victim\_age'].astype(int)  female\_victims = homicide\_data[homicide\_data['victim\_sex'] == 'Female']  women\_race\_data = {}  for race, race\_data in female\_victims.groupby('victim\_race')['victim\_age']:      women\_race\_data[race] = race\_data  plt.figure(figsize=(8, 6))  plt.boxplot(women\_race\_data.values(), labels=women\_race\_data.keys())  plt.title('Age Distribution Among Women for Different Races')  plt.xlabel('Race')  plt.ylabel('Age')  plt.grid(True)  plt.show() |

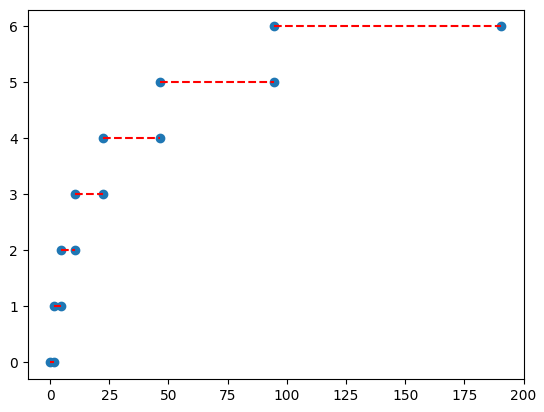


Рисунок 1 – Результат выполнение пункта 1 задания

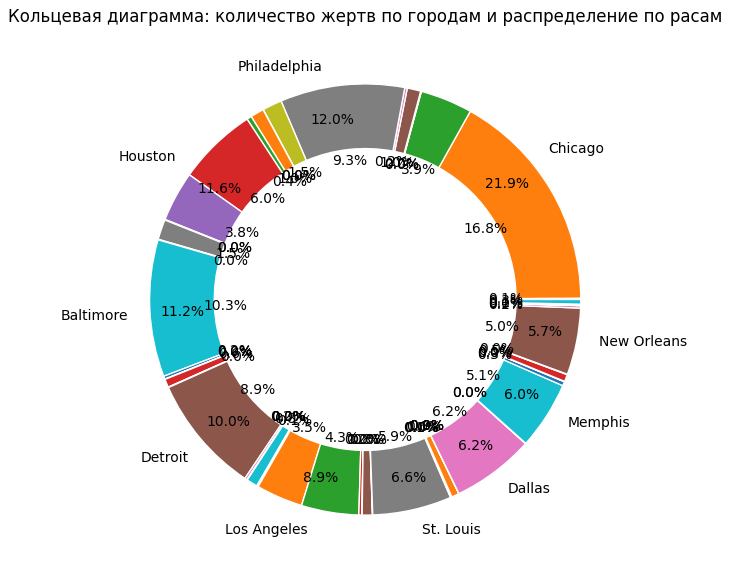


Рисунок 2 – Результат выполнения пункта 2 задания

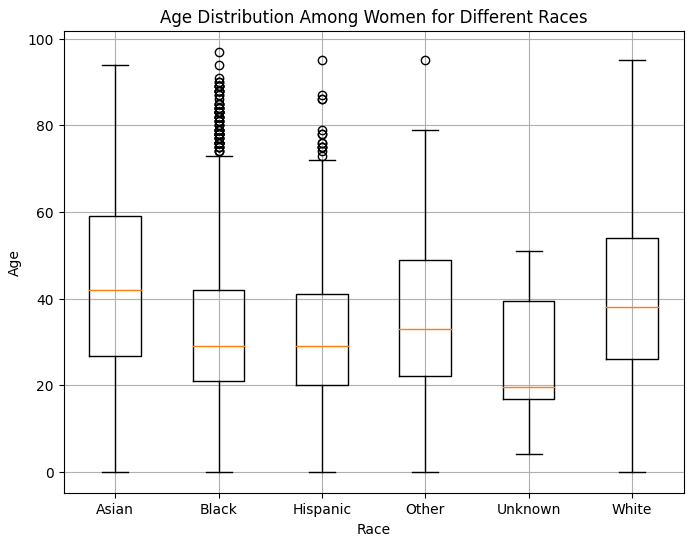


Рисунок 3 – Результат выполнения пункта 3 задания

**Вывод**

В ходе лабораторной работы мы усовершенствовали навыки визуализации данных, применяя Scatter Plots для построения лестницы и круговую диаграмму для анализа числа жертв по городам и расам. Также мы освоили создание боксплотов для распределения возрастов среди женщин в зависимости от расы. Полученные навыки позволяют нам более эффективно представлять и анализировать разнообразные аспекты данных.