



**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**

**О Т Ч Е Т**

**по лабораторной работе №2**

Студент

ИУ6-32М

(Группа)

\_\_\_\_\_

(Подпись, дата)

А.М. Панфилкин

(И.О. Фамилия)

Преподаватель

\_\_\_\_\_

(Подпись, дата)

М.А. Захаров

(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', linestyle='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()

```

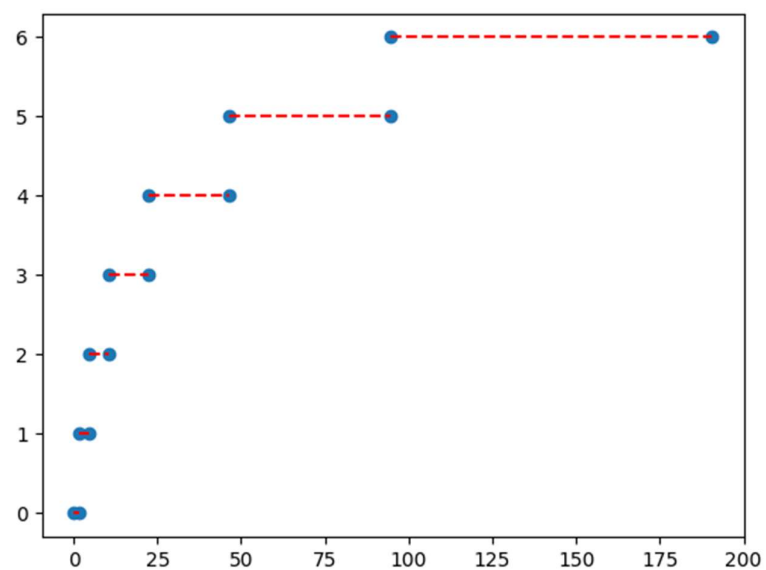


Рисунок 1 – Результат выполнение пункта 1 задания

Кольцевая диаграмма: количество жертв по городам и распределение по расам

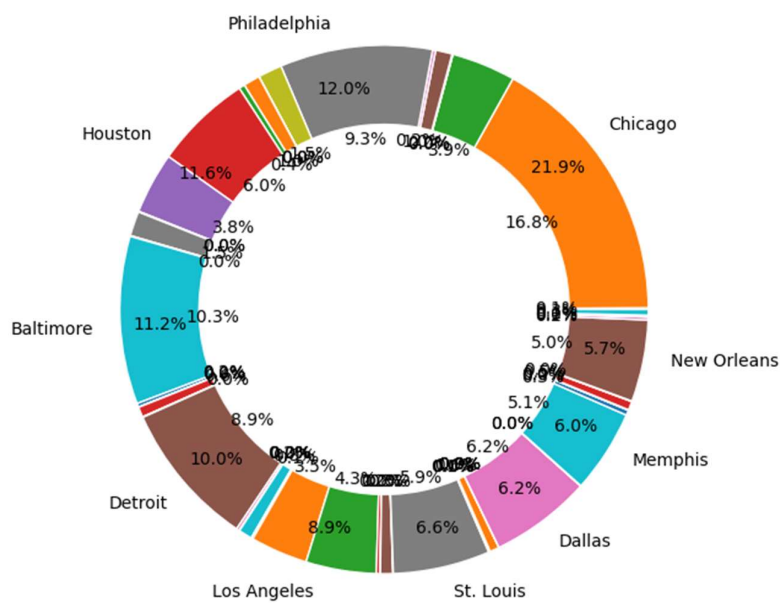


Рисунок 2 – Результат выполнения пункта 2 задания

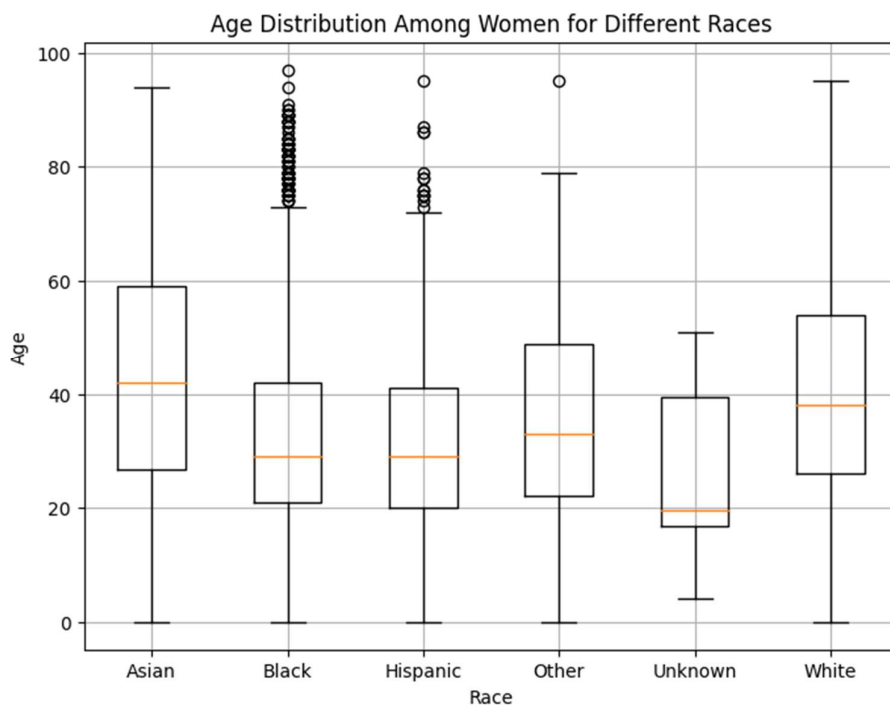


Рисунок 3 – Результат выполнения пункта 3 задания

## 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.