Task 5 Summary

**Objective**The main aim of this task was to understand my patient dataset through visual representation.  
By using different types of plots like histograms, scatter plots, pie chart, box plot, and pair plot, I wanted to observe the overall distribution and find visible patterns among Age, BP, Cholesterol, DrugR1 values.

**1. Descriptive Statistics Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | **Age** | **BP** | **Chlstrl** |
| **Mean** | 37.77 | 127.33 | 185.07 |
| **Std** | 18.80 | 22.85 | 28.46 |
| **Min** | 16.0 | 95.0 | 130.0 |
| **25% (Q1)** | 22.0 | 111.25 | 172.75 |
| **Median (Q2)** | 31.0 | 122.50 | 182.50 |
| **75% (Q3)** | 53.25 | 143.75 | 200.0 |

From this table, it’s clear that the patients are from mixed age groups (16–81 years).  
The average BP (127.33) and cholesterol level (185.07) fall in the moderate range, though a few patients have high readings.

**2. Histogram for Age, BP, and Cholesterol:**

The histograms for Age, Blood Pressure (BP), and Cholesterol (Chlstrl) clearly show how these health indicators are spread among the patients.

* **Age Distribution:**  
  Most patients are young adults, mainly between 20 to 35 years. The frequency decreases gradually as age increases, showing that there are fewer older patients. The slight right skew indicates a few elderly patients around 70–80 years.
* **Blood Pressure:**The blood pressure readings are centred mostly around 110–130 mmHg, which means most patients fall in a normal or slightly elevated BP range. However, there are a few patients with higher readings up to 180 mmHg, showing mild variation.
* **Cholesterol Levels:**  
  Cholesterol values are mostly between 170–200 mg/dL, which is a moderate level. Only a few patients have cholesterol near 250 mg/dL, indicating higher cholesterol cases are less common in this dataset.

**3. Scatter Plot for BP vs DrugR1 and Chlstrl vs DrugR1:**

* ***Scatter Plot 1: Age vs DrugR1:***

1. The data points are scattered widely, showing **no clear linear relationship** between Age and DrugR1.
2. **Younger patients (around 20–35 years)** seem to have **higher DrugR1 values**, suggesting stronger or more frequent responses.
3. As **age increases beyond 50**, DrugR1 levels appear to **decrease or vary less**, indicating age might influence how the drug works.
4. Overall, the relationship seems **non-linear.**

* ***Scatter Plot 2: BP vs DrugR1:***
  1. The distribution again appears **scattered**, meaning **BP (Blood Pressure)** doesn’t have a strong direct correlation with DrugR1.
  2. Most points are clustered between **BP 110–140**, where DrugR1 ranges from **1 to 7**, showing **mixed drug responses** across similar BP levels.
  3. A few outliers with **BP above 160** show **lower DrugR1 values**, hinting that **very high BP may reduce the drug’s effectiveness**.
* ***Scatter Plot 3: Chlstrl vs DrugR1:***
  1. The relationship between **Cholesterol and DrugR1** also appears **weak.**
  2. Most patients fall within **Cholesterol levels of 150–220**, but DrugR1 responses vary widely within that range.

**4. Boxplot for Age:**

A box plot (also called a box-and-whisker plot) visually displays the distribution of a dataset through its five-number summary:

1. Minimum (Q0)
2. First Quartile (Q1)
3. Median (Q2)
4. Third Quartile (Q3)
5. Maximum (Q4)

**From the plot:**

* **Minimum** (16 years), **First Quartile** (25 years) 25% of the patients are younger than this age, **Median** (32 years) Represents the middle value (50th percentile), **Third Quartile** (50 years) 75% of patients are younger than this age, **Maximum** (82 years).
* Most patients fall in the 25–50 years range, representing the main concentration of data.
* The median age (32 years) indicates a relatively young patient group overall.
* **Interquartile Range** (IQR): Q3 - Q1 = 50 - 25 = 25 years.

**5. Correlation Heatmap**

* **Age** shows strong positive correlation with **BP (0.54)** and **AnxtyLH (0.8) -**meaning older patients have higher blood pressure and anxiety/low happiness.
* **Age** has strong negative correlation with **Prgnt (-0.55)** - pregnancy is more common in younger patients.
* **DrugR** and **DrugR1** are highly correlated (**0.94**)- both represent similar medication patterns.
* **Chlstrl (Cholesterol)** shows very weak relation with other variables.
* Overall, **Age** and **AnxtyLH** are key influencing factors, while **DrugR1** may be redundant due to overlap with **DrugR**.