Assignment #4: Visualizing and Interpreting Hospital Patient Data   5 Points Possible

9/22/2025

Attempt 1 ⌄		◯ In Progress		⬚ Add Comment

In Progress
NEXT UP: Submit Assignment

Unlimited Attempts Allowed

9/15/2025

⌄ **Details**

Objectives

- Practice data cleaning and handling missing values in R.
- Create side-by-side boxplots and histograms for continuous measures.
- Interpret relationships between patient vitals and physician assessments.
- Communicate results clearly on your blog and GitHub.

Tasks

Enclosed is the dataset. (https://usflearn.instructure.com/courses/1984690/files/199244782?wrap=1)

⤓ **(https://usflearn.instructure.com/courses/1984690/files/199244782/download? download_frd=1)**

1. **Data Preparation and Cleaning**

   - Define each vector in R, converting categorical strings to numeric codes (bad=1, good=0; low=0, high=1) and handling `NA` appropriately:

   ```
   Frequency    <- c(0.6, 0.3, 0.4, 0.4, 0.2, 0.6, 0.3, 0.4, 0.9, 0.2)
   BloodPressure <- c(103, 87, 32, 42, 59, 109, 78, 205, 135, 176)
   FirstAssess   <- c(1, 1, 1, 1, 0, 0, 0, 0, NA, 1)    # bad=1, good=0
   SecondAssess  <- c(0, 0, 1, 1, 0, 0, 1, 1, 1, 1)    # low=0, high=1
   FinalDecision <- c(0, 1, 0, 1, 0, 1, 0, 1, 1, 1)    # low=0, high=1

   df_hosp <- data.frame(
     Frequency, BloodPressure, FirstAssess,
     SecondAssess, FinalDecision, stringsAsFactors = FALSE
   )
   # Inspect and handle NA:
   summary(df_hosp)
   df_hosp <- na.omit(df_hosp)
   ```

2. **Generate Basic Visualizations**
   - *A. Side-by-Side Boxplots*

   Plot blood pressure distributions by each assessment and final decision:

     ```
     boxplot(
       BloodPressure ~ FirstAssess,
       data = df_hosp,
       names = c("Good","Bad"),
       ylab = "Blood Pressure",
       main = "BP by First MD Assessment"
     )

     boxplot(
       BloodPressure ~ SecondAssess,
       data = df_hosp,
       names = c("Low","High"),
       ylab = "Blood Pressure",
       main = "BP by Second MD Assessment"
     )

     boxplot(
       BloodPressure ~ FinalDecision,
       data = df_hosp,
       names = c("Low","High"),
       ylab = "Blood Pressure",
     ```
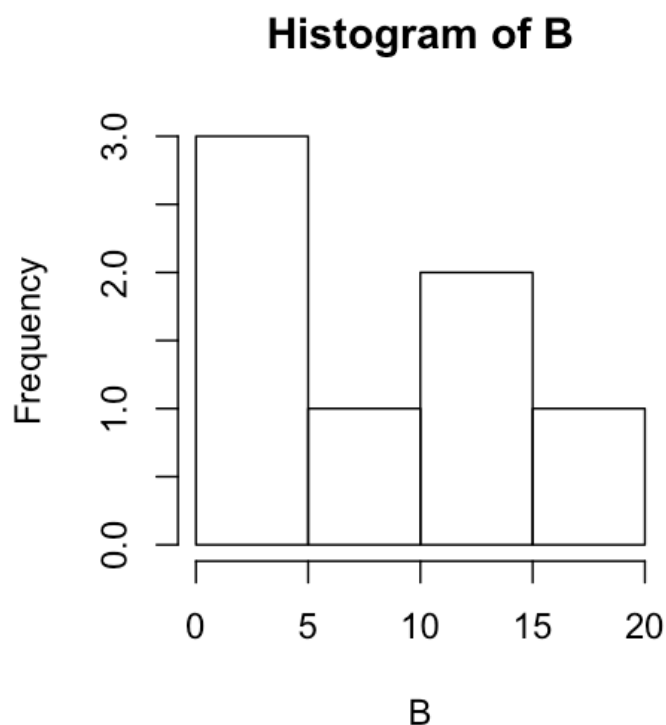
```
    main = "BP by Final Decision"
)
```
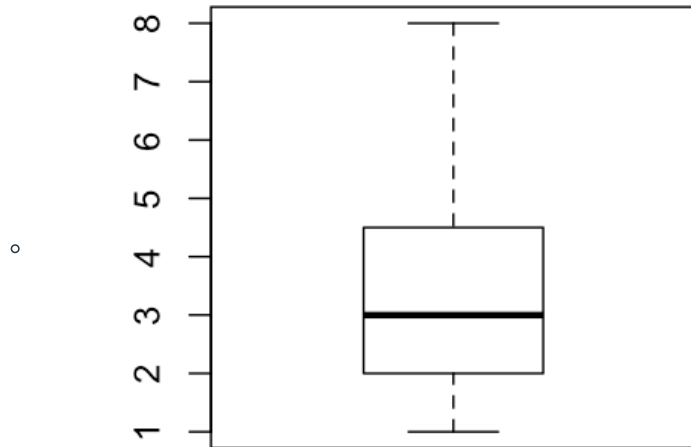

- *B. Histograms*
  Visualize overall distributions of Frequency and Blood Pressure:

```
hist(
  df_hosp$Frequency,
  breaks = seq(0, 1, by = 0.1),
  xlab = "Visit Frequency",
  main = "Histogram of Visit Frequency"
)

hist(
  df_hosp$BloodPressure,
  breaks = 8,
  xlab = "Blood Pressure",
  main = "Histogram of Blood Pressure"
)
```

Histogram of B

3. **Interpretation and Discussion**
   - On your blog, write 2–3 paragraphs addressing:
     - How blood pressure varies with each doctor's assessment and the final decision.
     - Any notable patterns or outliers in the histograms.
     - Potential clinical implications or limitations of this made-up data.
     - How handling of `NA` values affected your analysis.
4. **Submission**
   - Push your cleaned R script and generated plots to GitHub.
   - Embed code snippets, plots, and your discussion in a new blog post.
   - Share the blog URL and GitHub link in the Canvas "Assignment #4" submission area.

Submission Checklist

- R script with data cleaning, boxplots, and histograms
- PNG versions of each plot saved and embedded
- 2–3 paragraph discussion on your blog
- Links to your blog post and GitHub repository

Choose a submission type

T	🔗
Text	Web URL