

Toán rời rạc và thuật toán

Đại học Khoa học Tự nhiên

Khoa Toán - Cơ - Tin học

Khoa học dữ liệu K4

Tháng 8 năm 2022

Bài tập số 1

Nguyễn Mạnh Linh, Nguyễn Thị Đông, Triệu Hồng Thúy

1 Bài 1

Xin chào, đây là Linh

2 Bài 2

2.1 Chia để trị

2.1.1 Ý tưởng

Phương pháp chia để trị dựa trên 2 thao tác chính:

- Chia (*divide*): phân rã bài toán ban đầu thành các bài toán con có kích thước nhỏ hơn, có cùng cách giải.
- Trị (*conquer*): giải từng bài toán con (theo cách tương tự bài toán đầu - đệ qui) rồi tổng hợp các lời giải để nhận kết quả của bài toán ban đầu.

This homework answers the problem set sequentially.

1. *Download the US Presidential Elections data set `uspresidentialelections.dta` from the course ILIAS site. Load the data set in R.*

Copy your R Code to answer the question here.

2. *Describe the dataset. What variables does it contain? How many observations are there? What time span does it cover?*

Please type your answer here.

Put the right R command here.

3. *Compute measures of central tendency and variability of the variables **vote** and **growth** using R. Use the numerical measures of central tendency and variability discussed in class. Describe them in your own words and make a nice table. Plot the distribution of both variables using a boxplot and histogram. Make sure to make your plots as nice-looking as possible. Especially, include a title and label the axes.*

Your answer goes here

3 R commands here

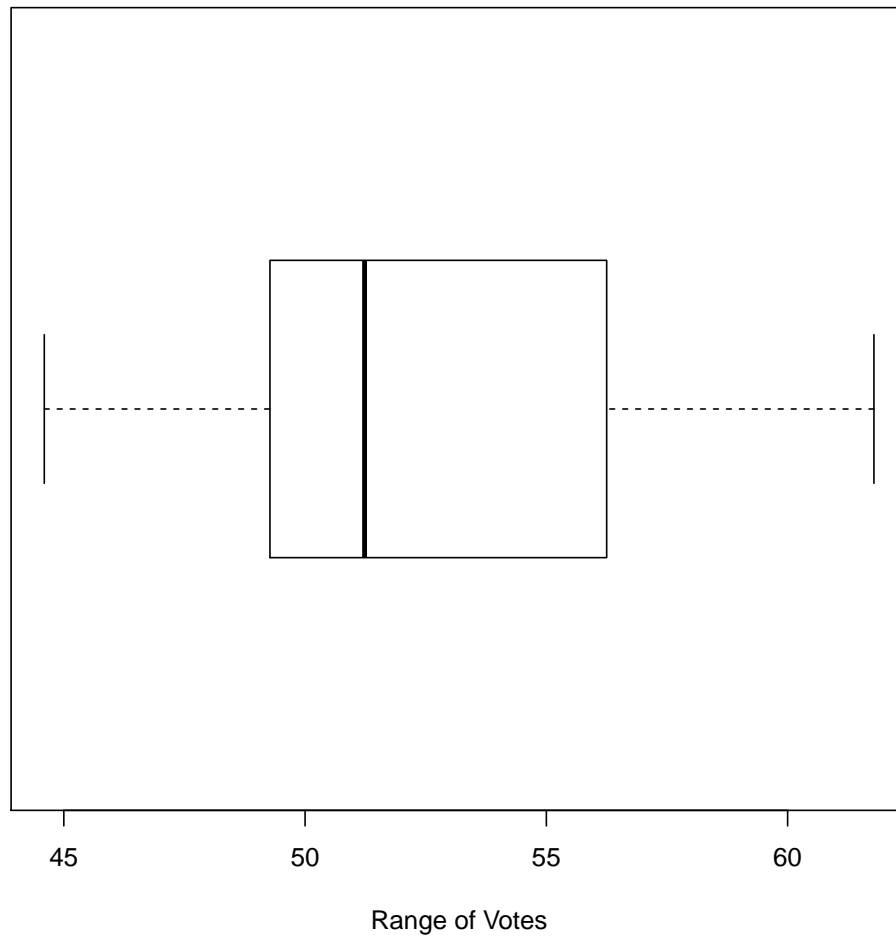
Potentially, your answer continues here.

4 R commands here

And more of your answer here.

And more space for your R commands.

```
# This is the code to produce the first boxplot.
pdf(file = "box1.pdf")
boxplot(us_data$vote, horizontal = T,
        main = "A Boxplot of the Variable Vote",
        names = "Vote",
        xlab = "Range of Votes")
dev.off()
```

A Boxplot of the Variable Vote

Hình 1: Boxplot of Incumbent Vote share

Variable	<i>Mean</i>	<i>Median</i>	<i>Mode</i>	<i>Var</i>	<i>SD</i>	<i>Range</i>	<i>IQR</i>
Vote	x	x	x	x	x	x	x
Growth	x	x	x	x	x	x	x

Bảng 1: Measures of central tendency and variability.

4. *Make a bar plot of the party affiliation of incumbent presidential candidates.*

Include the code for the bar plot and the plot here.

5. *During the presidential campaign in 1992, Bill Clinton's campaign coined the phrase "It's the economy, stupid!" Let's investigate the relationship between the economy and electoral success. Generate a nice-looking scatterplot of economic growth and vote share. Label the data points with the year of the election. Describe the pattern that you see in your own words.*

Include the code for the scatterplot as well as the plot here.

Then, describe the pattern you see. In the scatterplot we can see that...

R-Code

Finally, copy and paste the entire script here.