

Homework 2 – Due 11:59 pm CDT, 22 September 2024

The total points on this homework is 100. Three points are included for clarity of presentation, punctuation and commenting with respect to the code.

Year	Winner	Height	Opponent	Height
2008	Barack Obama	185 cm	John McCain	175 cm
2004	George W. Bush	182 cm	John Kerry	193 cm
2000	George W. Bush	182 cm	Al Gore	185 cm
1996	Bill Clinton	189 cm	Bob Dole	187 cm
1992	Bill Clinton	189 cm	George H.W. Bush	188 cm
1988	George H.W. Bush	188 cm	Michael Dukakis	173 cm
1984	Ronald Reagan	185 cm	Waiter Mondale	180 cm
1980	Ronald Reagan	185 cm	Jimmy Carter	177 cm
1976	Jimmy Carter	177 cm	Gerald Ford	183 cm
1972	Richard Nixon	182 cm	George McGovern	185 cm
1968	Richard Nixon	182 cm	Hubert Humphrey	180 cm
1964	Lyndon B. Johnson	193 cm	Barry Goldwater	180 cm
1960	John F. Kennedy	183 cm	Richard Nixon	182 cm
1956	Dwight D. Eisenhower	179 cm	Adlai Stevenson	178 cm
1952	Dwight D. Eisenhower	179 cm	Adlai Stevenson	178 cm
1948	Harry S. Truman	175 cm	Thomas Dewey	173 cm

Table 1: Height of the election winner in the Electoral College and height of the main opponent in the U.S. Presidential election of 1948 through 2008.

- Table 1 displays the heights of the presidents and their opponents in the U.S. presidential election of 1948 through 2008.
 - Enter the data in R as a data frame with 3 columns: **year**, **winner** and **opponent**. Enter the height of the winner candidate in the column **winner**, the height of the opponent candidate in the column **opponent**. [5 points]
 - Create a dataframe with the same fields as before, but including an additional column called **difference** which is the differences in height between the winner and opponent in the column **difference**. [5 points]
 - Add a new column **taller.won** to your data frame in (a) with logical values (**TRUE/FALSE**) determining whether the taller candidate won the election. [5 points]
 - Use the **table** function to display percentages of **TRUE/FALSE** in the column **taller.won**. Interpret the result. [5 points]
 - Display a bar plot of the **difference** column. Use the **rev** function to reverse the order of the differences so that the election year is increasing from left to right. [5 points]
 - Using the function **write.table()** perhaps, save the table obtained in 1(c) above in a new file called heights.csv. (Note that you do not have to turn in the created file, just the complete function that you used.) Make sure that the created file is as you would like. [5 points]
- Read the dataset **students.txt** available in the **Datasets** section of Canvas into an R object named as students. [5 points]
 - What is the mean height and shoesize of all the students? What about their standard deviations? [5 points]
 - How many female students are in the sample? How many male students? [5 points]

- (c) Recode the population variable with color names (**kuopio = blue, tampere = red**), and create a new dataset called **students.new**. [5 points]
 - (d) Make two subsets of the dataset students. Split it in two according to gender and export the two datasets to **female.txt** and **male.txt** files. [5 points]
 - (e) make two new datasets of the dataset students that containing individuals below and above the median height. Export the two new datasets to **below.csv** and **abovem.csv** files. [5 points]
3. We are given two 83×108 slices of the (same) human brain. The first slice, contained in *anat.dat*, a file available on Canvas, contains the anatomic information of the brain, revealing its structure, as per a Magnetic Resonance Image (MRI), while the second file, located at *activ.dat* also available on Canvas, is the probability that that particular location (pixel) in the slice is activated in response to the tapping of a finger. The objective of this exercise is to display the anatomic structure of the brain, overlaid with the functional image to provide context and understanding of which structures of the brain slice respond to the application of a stimulus (tapping of one's right index finger against the thumb). To do so, you may use **demo(image)** to get ideas, or the help file of the following functions: **scan**, **matrix**, **contour**, **image**. The function **rev** may also perhaps be needed.
- (a) Read in the data from each of the two files into R and store each dataset as a matrix of appropriate dimensions. Note that both files have NA's to represent where there is no structure in the brain slice (outside the brain). [8 points]
 - (b) Provide an image of the activation data using a grayscale color map. Make sure that you set axes = F, and also perhaps make sure that the lighter scale of the image map represents low values. [6 points]
 - (c) On the image plot above, overlay a contour plot of the anatomic structure of the brain. Note that there is no need to label the contour lines, and also that there are a few options to control the levels of the contour plots as well as the line width. Note that overlaying means to add a figure atop an existing plot. [6 points]
 - (d) *What to turn in:* Turn in the final plots and also the final R codes that you used for the problem.
4. *Romano-British Pottery.* Samples from Romano-British pottery were taken at four sites in the United Kingdom. The dataset is available on Canvas at **pottery.dat**. Twenty-six samples of Romano-British pottery were found at four different kiln sites in Wales, Gwent and the New Forest. The sites are Llanederyn (L), Caldicot (C), Island Thorns (I), and Ashley Rails (A). The other variables are the percentage of oxides of various metals measured by atomic absorption spectro-photometry. The data were collected to see if different sites contained pottery of different chemical compositions. A chemical analysis of the pottery was performed to measure the percentage of five metal oxides present in each sample.
- 1. Al: Percentage of aluminum oxide in sample
 - 2. Fe: Percentage of iron oxide in sample
 - 3. Mg: Percentage of magnesium oxide in sample
 - 4. Ca: Percentage of calcium oxide in sample
 - 5. Na: Percentage of sodium oxide in sample
 - 6. Site: Site where pottery sample was collected
- (a) Read in the dataset into R. [2 points]
 - (b) For each site, provide summaries of chemical content using means, medians, standard deviations and interquartile ranges (if applicable). [8 points]
 - (c) Use the function **boxplot** to provide boxplot displays of the site-wise composition of each chemical. Hand in these plots, all on one page. (To do this, use the graphical function **par** before plotting. The function `> par(mfrow=c(2,3))` divides the plotting region into 2 rows and three columns. Thus, successive calls to a plot would fill the region one by one. [10 points]
 - (d) Discuss the characteristics of the distributions of the percentages of different metal oxides, as provided by the above boxplots and summary measures. [2 points]