

## Assignment: RBase Graphics

**Purpose:** The purpose of this assignment is to reinforce the RBase graphics skills covered in lab by applying them to a specific file. A second purpose is to learn to search for appropriate packages and to google techniques that can be modified to your needs.

**Objectives:** You will review several of the following skills in R:

- Histograms
- Boxplot
- Barplot
- Heatmap (use a matrix)
- Image (use a matrix)
- Mosaicplot
- Pie chart
- Plot (simply line plot)
- Use of color (ColorBrewer, )
- `par(mfrow)`, `cex`, labels, etc.

### What you need to do:

1. Create a new project in RStudio. Name it *RBase-HW-yourname*.
2. Within your project, name the default R source file `WeightLoss.R`.
3. Set your working directory to wherever you have saved your project. This means that you will either store your dataset in that directory, or you will read in your dataset using the full path for it. (For these homeworks, I would opt to store it in the same directory, just to make things easy. But you certainly may prefer to store all of your data files in one directory.)
4. Load the dataset `WeightLoss.csv` into a variable `wl` (that's the letter "el", not the number 1). You may also want to look at the documentation, which is in `WeightLoss.docx`. These are available on Sakai.
5. Packages that you will be using: `ColorBrewer`, `data.table`
6. **Create a histogram of the variable `wl1`.** What does this tell you? Is this useful? When we looked at `VADeaths` in the lab, you could simply use the command `hist(VADeaths)`, and it created a histogram of all `VADeaths`, binned into default or specified bins. That's because `VADeaths` is treated as a matrix, with all numeric values. But one of the `WeightLoss` variables is not numeric. So you have to specify the numeric variable, or you will get an error. Alternatively, you can **convert just the numeric parts of the data.frame to a matrix, and run a histogram on that**. Do that. Is this useful? What

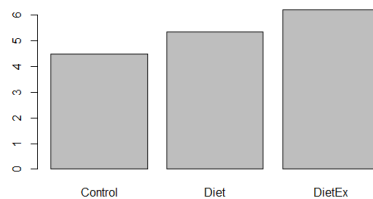
does this tell you? Recall that the first three numeric columns are the weight loss over the first three weeks? The second three columns are the self-esteem. Do you want these numbers mixed up together in a histogram? Why or why not?

- a. You can run separate histograms on each variable. This has limited usefulness, although it can be useful to compare the distribution of the different weeks of weight loss. **Create three separate histograms for the variables wl1, wl2, and wl3. Put them all in one page**, so we can see how they compare. (Recall that you can use mfrow for this.)
  - b. You can run all the weight loss numbers together, and all of the self esteem numbers together, constructing two histograms. In this case, you should **create a matrix that has the weight loss values, and a separate matrix that has the self-esteem values, and you can create a histogram for each. Display these histograms 2 to a page**. Why a matrix? A data.frame can have different types of variables for each column (vector). A matrix requires that all values are the same type. Since a histogram requires only numeric values, using a matrix of numeric values is permissible (and a good idea too!).
7. Or maybe, a histogram isn't the best representation beyond some very initial exploration. Recall that a histogram is useful for looking at the distribution of one variable. Let's compare the first week's weight loss by group. The groups are the control group, the group that just dieted, and the group that dieted and also exercised. To use it as a grouping variable, it has to be a factor? Is it? (If not, convert it to factor.) **Create a boxplot showing the distribution of wl1 by group**. (Be sure to set mfrow back to one plot per page, because this one plot will have 3 boxplots in it!) Is it easier to see who lost the most weight?
  8. Or maybe a scatter plot would show the differences better. **Plot wl1 by group**. Hey, that's not a scatter plot! Why??!
  9. Better question: **Does self-esteem increase as weight loss increases? Plot that**.
  10. **Does self-esteem differ as weight loss increases differently for each group? Plot that**.
  11. You know, maybe if I lose weight in week 1, it still has an effect on my self-esteem in later weeks. Maybe if I have good self-esteem for starters, it's correlated with good self-esteem in later weeks. . Or maybe if I lose weight in the first week, I find it motivating to lose weight in later weeks. Or maybe if I lose weight in the first week, I find it more difficult to lose weight in later weeks. Who knows what's correlated to what here! Let's **plot every variable against every other variable. And color it by group**. 3 colors, because there's 3 groups.
  12. And now for something new. I think it might be good to see a **bar chart comparing a week's weight loss across groups (a "side by side" bar chart)**. This would require, however, some type of summative measure—one measure per group for each variable. A mean would be a good measure for this purpose. Please review this video: <https://www.youtube.com/watch?v=AmE4LXPQErM>. (He is using ggplot towards the end of the video. We are NOT, for this assignment.) Note especially that it requires

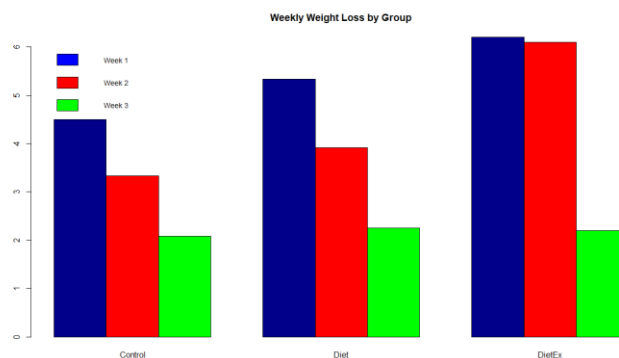
importing the data.table package; creating a data.table from your data.frame; and creating the summary measures. Please create a new variable named wIT which is a data.table version of your data.frame. Then, create another variable, and name it summarizedwIT, which contains the summary measure (in this case, the means) or the weight loss for the first three weeks. You should be able to follow the instructions from the video to construct summarizedwIT. It should look like this: ([paste your data.table into your homework document](#)).

	group	Average_wI1	Average_wI2	Average_wI3
1:	Control	4.500000	3.333333	2.083333
2:	Diet	5.333333	3.916667	2.250000
3:	DietEx	6.200000	6.100000	2.200000

13. **Construct a bar chart of just summarizedwIT\$Average\_wI1, grouped by “group”.** We constructed a similar bar chart in our lab, using the “dragons” dataset. Your bar chart should look like this:



14. But what I would really like is **a chart that compares the weekly weight loss for each week, by group.**



How did I do this???

- The bar chart works better with a table or matrix. I had some issues loading my data into a table, so I used a matrix.

- I created a variable called Values. I assigned to Values a matrix that contained the three Average values in my summarizedWIT data.table. The matrix has 3 rows and 3 columns, and be sure to set byrow = TRUE, so that it knows that it treats each row as one observation.
- If you look at the barplot parameters, you'll see that you can specify:
  - "main", which is the title of the plot
  - "names" or "names.arg" (I used "names" as we did in our lab). This is a vector with the names of the groups. We have these names in our summarized.wIT\$group variable, so just use that.
  - "col", which is a vector of colors. You can define the vector in advance, and just use the vector name here, or you can list the vector inline.
  - Beside = TRUE for a side-by-side bar plot. The default is a stacked plot.
- After your chart looks good, you can add the legend in a separate command. Look at the [legend parameters](#), and see if you can duplicate my chart.
  - Specify where the legend should be placed. You can actually specify the exact location. Or, you can use the keywords "topleft", "topright", "left", "top", etc.
  - Specify a vector which describes what each color means. In my barplot I had specified darkblue, red and green. So in my legend I want blue to be labeled "Week 1", red as "Week 2" and green as "Week 3". My data are in that order, and my colors are specified in that order, so my legend names should be specified in that order too. Similarly, the "fill" parameter should specify those colors, in that order.
  - I found that the legend was originally too big for the plot (it was covering some of the bars). So I changed the "cex" to be .85.
- You may look at the help in the help pane in RStudio. And you may google anything that helps you with the bar chart. There are quite a few other ways to generate this chart, and you may use other ways, *as long as you are using only RBase graphics. No lattice or ggplot2 at this point.*
- Oh, BTW, assume your viewers may be colorblind, so change my colors.... You don't have to, but it would be nice if you did.

### What to submit:

1. You will submit one Word (or similar) document. This file should be named RBase-hw-*yourname.docx*. For each question above, beginning with Question #3, write:

Question #3:

*Insert the code that accomplishes the task. Or, if you used the GUI, describe the steps that you took to accomplish the task.*

*If there is a plot, paste a copy of the plot. To copy plots, use the zoom button in the plots window, right-click and copy. Then you can paste directly into your word processor.*

*If there are questions to answer, answer the questions.*

Then move on to the next question.

2. Submit on Sakai.
3. You may work in teams of two, one submission per team. Each team member must submit something on Sakai. If you are submitting on behalf of your team, be sure to include a comment "I am submitting on behalf of Bob" in the assignment box on Sakai. If someone else is submitting for your team, be sure to include a comment "Bob is submitting on behalf of our team".