Econ 294 Assignment 1

Curtis Kephart

Winter 2015

Use R to answer the following questions.

Create a `.R` script that creates a `list` object containing all solutions.

Your solutions should take the form of a `list` object saved as a `.RData` file. The list should have the following naming convention `{First Name}{Last Name}Assignment2` (e.g. CurtisKephartAssignment2), and the `.RData` file should be named `{First Name}{Last Name}Assignment2.RData` (e.g. CurtisKephartAssignment2.RData).

- Due: by Jan. 22 2016 (after the next lecture).
- You will turn in your `.R` script and `.RData` files by pushing them to your public github repository, and emailing the URLs of those files to your instructor. (at `curtisk+econ294_02@ucsc.edu`)
- For the `.R` script, it is important that the instructor should be able to run the script completely (`source()`) without error, recreating your `list`. Warning messages are okay.
- Please comment your code so that it is easy to understand which code blocks answer particular questions. For the `list` object you are to create and save as an .RData file, at the end of each question I will give you the list's component tag (e.g. `CurtisKephartAssignment2$tagName`) to store your answers in. I also note the data type I expect to see.

Just to be clear, and to help you get started, here are examples that satisfy the first two sets of questions, at `https://github.com/EconomiCurtis/econ294_2015/raw/master`:

- `/Assignments/CurtisKephartAssignment2Creator.R`
- `/Assignments/CurtisKephartAssignment2.RData`

0. Identifying informaiton

- Your first name (`$firstName` character string)
- Your last name (`$lastName` character string)
- Your email (`$email` character string)
- Your student ID number (`$studentID` numeric)

1. Load the following .RData file.

   `https://github.com/EconomiCurtis/econ294_2015/raw/master/data/diamonds.RData`

- How many observations are there? (`$s1a` numeric)
- How many columns are there? (`$s1b` numeric)
- What are the header names? (`$s1c` chr vector, length 4)
- What is the `summary` of prices? (`$s1d` summary table, length 7)
- Up to this point, see solutions in `Assignments/CurtisKephartAssignment2Creator.R`

2. Load the following tab-separated file. (Be sure to handle its header correctly.)

   `https://github.com/EconomiCurtis/econ294_2015/raw/master/data/NHIS_2007_TSV.txt`

- How many observations are there? (`$s2a` numeric)
- How many columns are there? (`$s2b` numeric)
- What are the header names? (`$s2c` chr vector, length 9)
- What is the mean weight of the `weight` columm? (`$s2d`, numeric)
- What is the median weight of the `weight` columm? (`$s2e`, numeric)
- Create a histogram of these weights (e.g. `hist(df$weight)` and `table(df$weight)`). Note there is a group between 996 and 999 pounds. It turns out these are codes for various types of missing data. Use `ifelse` to create a new column, setting these weight observations to `NA`.
- What is the new mean weight of this adjusted `weight` columm? (`$s2f`, numeric)
- What is the new median weight of this adjusted `weight` columm? (`$s2g`, numeric)
- If the `SEX` column indicates mean with `1` and woman with `2`:
- What is the `summary` of weights for woman (`$s2h`, summary table, length 7) and men (`$s2i`, summary table, length 7)?

3. Extracting values from a a vector, data frame, and list.

```
vec <- c(letters,LETTERS)
```

- Extract even index values (2,4,6...) from `vec` (`$s3a` char vector, length 26)
- Use `vec` to extract to extract the first three letters of your name (e.g. for the Cur in Curtis, `paste(vec[c(29,21,18)], collapse="")`) (`$s3b`, character) (tip, use the `[...]` extracting, and there are a number of ways to do this, modulo , `rep`, `seq`.)

```
arr <- array( c(letters,LETTERS), dim = c(3,3,3)
```

- Use `arr[...]` to extract the first column from the second matrix of `arr` (`$s3c` char vector, length 3, `j,k,l`)
- Use `arr[...]` to extract the middle values from each of the three matrices in `arr` (`$s3d` , char vector, length 3, `e,n,w`)
- Extracting values from `arr`, spell the first three letters of your first name (e.g. for the cur in curtis, `paste(arr[3,1,1],arr[3,1,3],arr[3,3,2], sep = "")`) (`$s3e`, character, length 1, should match `firstName`).

4. Working with data.frames, modifying, grouping and summarizing.

- Load Alan's `org_example.dta` file.
- Find average `rw` for each `year-month-educ` group. (ignore `NA`s in the mean calculation). Place these values into a data.frame with four columns, year, month, educ, and rw. (`$s4` data.frame 420 by 4)

Tips: if you load the file as `org_example`, `sort(unique(org_example$year))` will list each unique year, `sort(unique(org_example$month))` will list each unique month, and `sort(unique(org_example$educ))` will list each unique education level for a for-loop. You might also use the `aggregate` function.