NR 592
Assignment 2

INSERT YOUR NAME HERE

INSERT THE DATE HERE

Last Updated: 2019-04-02

Instructions

Show your work. The response to each question should include all of the commands necessary to reproduce your results, and should be commented to give a naive reader an idea of the intent behind each section of code. Answers to questions should be clearly shown in the code output or spelled out in text or comments.

Before submitting your assignment, ensure that your code for each question works correctly when the lines are run in sequence, starting with an empty environment. You may submit either this R Markdown document with your code entered into the appropriate code blocks, or a regular R script with the question numbers as comments and the answers following. If you choose to Knit the R Markdown document to HTML or PDF, include those files with your submission (but it is not required that it Knit successfully).

Question 1

1.1

Find an external public data set (not included in R) that interests you and has at least 2 numeric variables. Download it and load it into R.

If you're having trouble finding something, you can look through the data sets available at OpenData Fort Collins or the Colorado Information Marketplace (and filter by "Datasets" under "View Types" on the left side).

Important: Include the data file in your submission and include a link to its source here.

```
# Source
Delete this and paste a link to the source of your data set here.
```

```
# Load data
```

1.2

Examine the dimensions, class, and structure of your data, using the appropriate R functions.

If your data set is not already in the format of a data frame, coerce it into a data frame. If any of your variables' classes are not properly assigned, coerce them into the correct classes.

Print the head of your data frame.

(Hint: Check the list of "Common functions" in Lecture 02.)

1.3

Check to see whether your data frame has any missing values. If it does, remove or replace them (using any method you choose).

1.4

Are your data "tidy"? Why or why not (according to the "three rules")?

`Delete this and enter your response here.`

If not, what would you need to do to make them tidy? You do not have to actually do it, just briefly describe the steps you would take and what the final result would look like.

If your data already are tidy, imagine that you were the one collecting the data. Briefly describe one thing you could have done differently, or a different way of recording the data, that could have resulted in an "un-tidy" data set.

`Delete this and enter your response here.`

1.5

Subset your data frame using some criterion of your choice and assign the result to a new data frame.

Question 2

2.1

Choose two numeric variables in your original or subset data frame. For each of the two variables, find its mean, median, standard deviation, and range (maximum value and minimum value).

2.2

For each of your two numeric variables, create a plot to visualize its distribution.

If the variable is discrete, create a histogram or density plot.

If the variable is continuous, create a boxplot or violin plot.

2.3

Find the correlation between your two numeric variables.

2.4

Create a scatterplot to visualize the relationship between your two numeric variables.

2.5

Based on their correlation and by "eyeballing" the plot, do the two variables seem to be related? If so, is the relationship positive (as one variable increases, the other increases) or negative (as one variable increases, the other decreases)?

`Delete this and enter your response here.`

2.6

Are there any obvious outliers in the plot, or points that could potentially be outliers?

`Delete this and enter your response here.`

2.7

Use the `boxplot.stats()` function to test each of your two numeric variables for outliers.

2.8

Create one additional plot of your choice depicting one or more of the variables in your data frame.

Comment briefly on something you learned about your data from this plot.

`Delete this and enter your response here.`

(pdf / Rmd)