Data Science Principles and Practice
CS x415.1, Fall 2016
Assignment #1
Due: Sunday, February 19, 2017 11:59 PM (no late submissions accepted)
Submit on Canvas / onlinelearning.berkeley.edu (no email submissions accepted)
See submission instructions below

Note: the due date for this assignment has been pushed back a week to give you time to brush up on ggplot and R Markdown (see Chapters 1 and 21 R4DS and the online resources included in my email).

Assignment 2, much, much larger and more substantial, will be posted next week. Get an early start on Assignment 1, post questions to the Discussion section (please use meaningful subject names and open a new thread if you are starting a new topic).

---

1. Log into https://onlinelearning.berkeley.edu/
   Follow instructions for first time login if necessary.
   Post a reply in the discussion "Introduce Yourself"

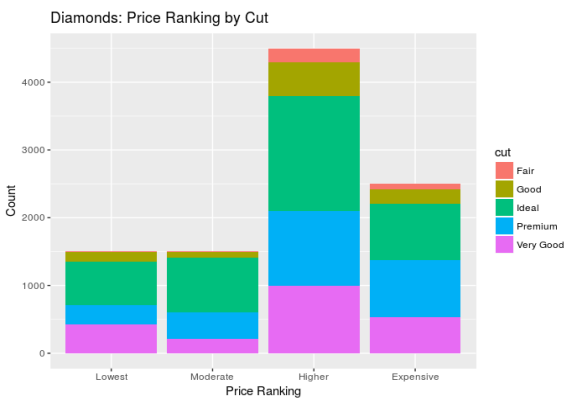
2. Post your photo on Canvas that we can use to identify you

Programming Assignment 1

Use the tab-delimited gem data set ddset.tsv posted in the Files section under Assignment 1 to do the following:

- Write an RMarkdown file and use it to generate a knitted HTML file that includes a bar plot of the count of observations in the data set showing price ranking by cut, where price ranking is defined as price:
  - up to the 15th percentile: Lowest
  - 16th to the 30th percentile: Moderate
  - 31st to the 75th percentile: Higher
  - above 75th percentile: Expensive

Your plot should look something like this:



Hand in three files:

1. Your RMarkdown file: asn1_firstname_lastname.Rmd
2. Knitted HTML with source: asn1_firstname_lastname_with_source.html (with source code, i.e. chunk option echo = TRUE)
3. Knitted HTML with no source: asn1_firstname_lastname.html (with source code, i.e. chunk option echo = FALSE)

Mandatory Specifications

1. Assume ddset.tsv is on the same directory as your .Rmd file. Your HTML file results should be 100% reproducible: I should be able to Knit it to HTML without modification
2. Include in your results the complete problem description text for Programming Assignment 1 above
3. In the YAML section of your .Rmd file, include title "Programming Assignment 1," author (your name), and date. Of course, output should be HTML
4. Use base R and ggplot2 only in your solution. Use no other packages except ggplot2 (do not use dplyr!)
5. Follow the style guidelines described by Hadley Wickham in this R style guide
6. Include a short (no more than one paragraph) description of your approach to solving the problem at the end of your HTML document

Grading:

100 points total. You will be graded on:

- Completeness

  (in this assignment, includes posting your bio and photo!)
- Follows specificcations (especially the mandatory specificcations above!)
- Implementation quality (code quality, adherence to style guidelines)