

Rmarkdown-exercise-worksheet

Reid Otsuji

1/5/2020

Instructions:

Complete the worksheet using RStudio and Rmarkdown. Some R Packages will be necessary to complete the worksheet.

Answer all questions in this document and have fun with Rmarkdown!

Before you begin: You will need to load the GGLOT package to use the `diamonds` data set. Run the code each of the following chunks once to install and load the GGLOT2 package.

1. Install and Load the ggplot2 package

(if you already have GGLOT2 installed skip this code chunk.)

```
# This is an example of a code chunk to install an R package  
# Run this code chunk manually if you need to install the GGLOT2 package  
# click the green arrow to the left to install the package.  
install.packages("ggplot2")
```

- Run this code chunk to Load the GGLOT2 library.

```
# load the GGLOT2 library  
library(ggplot2)
```

2. In the Diamonds Data Set Description text provided below, add the proper Rmarkdown formatting to each line of text. The highlighted recommendations at the end of each line.

- **optinoal:** add a link to the Diamond Data Set text below using the link formt [link text] (http://add link here) to the dataset information.
 - The 4 C's of Diamond Quality <https://4cs.gia.edu/en-us/4cs-diamond-quality/>

Prices of 50,000 round cut diamonds H1 heading

Description H2 heading

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows: bold this line

Usage h3 heading

- diamonds bullet

Format h3 heading

A data frame with 53940 rows and 10 variables: bold

-

price h4 heading, bullet

- price in US dollars (\$326–\$18,823) nested bullet

•

carat h4 heading, bullet

- weight of the diamond (0.2–5.01) nested bullet

•

cut h4 heading, bullet

- quality of the cut (Fair, Good, Very Good, Premium, Ideal) nested bullet

•

color h4 heading, bullet

- diamond colour, from D (best) to J (worst) nested bullet

•

clarity h4 heading, bullet

- a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best)) nested bullet

•

x h4 heading, bullet

- length in mm (0–10.74) nested bullet

•

y h4 heading, bullet

- width in mm (0–58.9) nested bullet

•

z h4 heading, bullet

- depth in mm (0–31.8) nested bullet

•

depth h4 heading, bullet

- total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43–79) nested bullet, inline code block

•

table h4 heading, bullet

- width of top of diamond relative to widest point (43–95) nested bullet

For the next section you will need to load the following packages:

- knitr
- ggplot2
- kable
- pander

```
# this is an example of a code chunk
#run manually by clicking the green arrow to the left if you need to load the library
library(knitr)
```

Next, you will practice working with code chunks!

3. Adding code chunks

- On the tool bar, use the **Insert** button and selecting **R** or manually create one by typing a “code fence” 3 opening and closing backticks with `{r}` e.g. ````{r}` at the beginning and closing with ````` to create a code chunk (see example above).
- In between the “code fence” add and run the R code `summary(diamonds)` to get the summary of the data set: (run the code chunk to make sure it works)
 - **optional-** hide the code chunk in your knitted document by adding `echo=FALSE` in the code chunk options

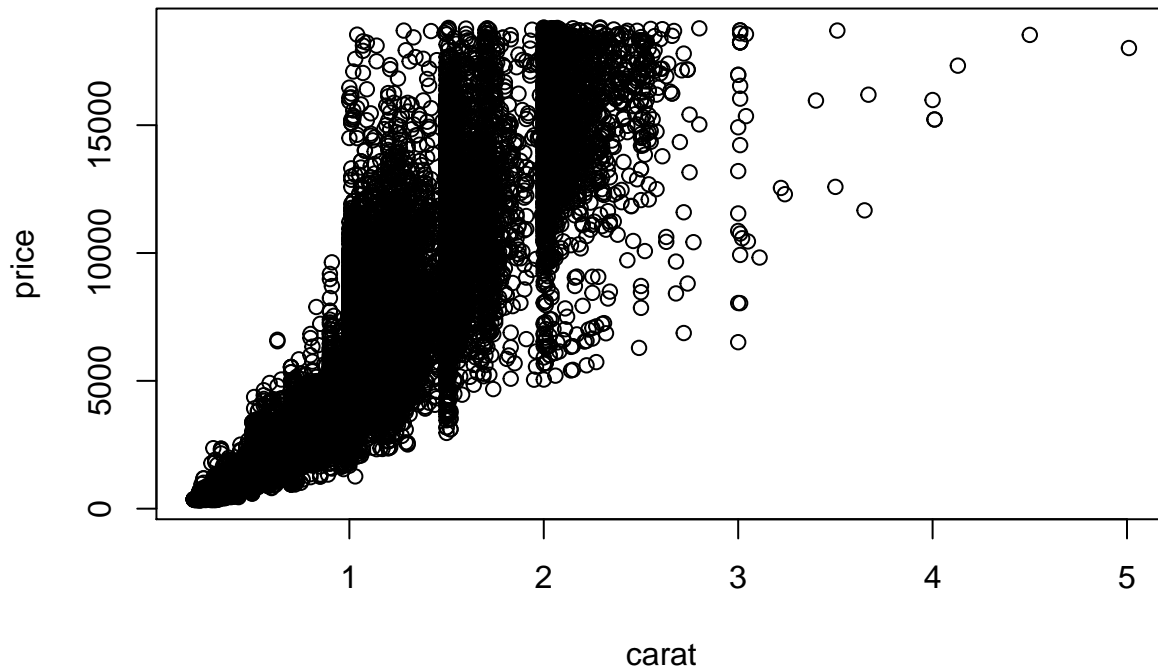
Add code chunk below:

```
##      carat      cut      color      clarity
## Min.   :0.2000 Fair      : 1610 D: 6775 SI1      :13065
## 1st Qu.:0.4000 Good      : 4906 E: 9797 VS2      :12258
## Median :0.7000 Very Good:12082 F: 9542 SI2      : 9194
## Mean   :0.7979 Premium  :13791 G:11292 VS1      : 8171
## 3rd Qu.:1.0400 Ideal     :21551 H: 8304 VVS2     : 5066
## Max.   :5.0100              I: 5422 VVS1     : 3655
##              J: 2808 (Other): 2531
##      depth      table      price      x
## Min.   :43.00 Min.   :43.00 Min.   : 326 Min.   : 0.000
## 1st Qu.:61.00 1st Qu.:56.00 1st Qu.: 950 1st Qu.: 4.710
## Median :61.80 Median :57.00 Median : 2401 Median : 5.700
## Mean   :61.75 Mean   :57.46 Mean   : 3933 Mean   : 5.731
## 3rd Qu.:62.50 3rd Qu.:59.00 3rd Qu.: 5324 3rd Qu.: 6.540
## Max.   :79.00 Max.   :95.00 Max.   :18823 Max.   :10.740
##
##      y      z
## Min.   : 0.000 Min.   : 0.000
## 1st Qu.: 4.720 1st Qu.: 2.910
## Median : 5.710 Median : 3.530
## Mean   : 5.735 Mean   : 3.539
## 3rd Qu.: 6.540 3rd Qu.: 4.040
## Max.   :58.900 Max.   :31.800
##
```

4. Create a new code chunk to add a basic scatter plot to your document that shows the price as a function of carat weight:

- Use this code `plot(price ~ carat, data=diamonds)` Add code chunk below:

```
plot(price ~ carat, data=diamonds)
```



5. Create a basic R table that shows the count for the diamonds cut and clarity following these steps:

1. Create a code chunk
2. Hide the code chunk in the knitted document output using `echo=FALSE`
3. Use this code `table(diamonds$cut, diamonds$clarity)` to generate the table

Add code chunk below:

```
##
##           I1  SI2  SI1  VS2  VS1  VVS2  VVS1  IF
## Fair       210  466  408  261  170   69   17   9
## Good       96 1081 1560  978  648  286  186  71
## Very Good  84 2100 3240 2591 1775 1235  789 268
## Premium   205 2949 3575 3357 1989  870  616 230
## Ideal     146 2598 4282 5071 3589 2606 2047 1212
```

Next, improve to formatting of this table by using the knitr package function `kable`:

1. Use `kable` to improve the look of the default table
2. Create a new code chunk with the option `echo=FALSE`
3. Use this code `kable(table(diamonds$cut, diamonds$clarity))` to generate the table

- if you receive an error message, make sure the knitr library is enabled. run `library(knitr)` before you run the kable code.

Add code chunk below:

	I1	SI2	SI1	VS2	VS1	VVS2	VVS1	IF
Fair	210	466	408	261	170	69	17	9
Good	96	1081	1560	978	648	286	186	71
Very Good	84	2100	3240	2591	1775	1235	789	268
Premium	205	2949	3575	3357	1989	870	616	230
Ideal	146	2598	4282	5071	3589	2606	2047	1212

6. Create a publication style table using the pander package

1. Install pander package using: `install.packages("pander")` run this code chunk if you need to install the pander package.
2. Load the pander package using: `library(pander)`

```
library(pander)
```

3. Run an simple anova (analysis of variance model) to model price as a function of carat.
 - Run this code to add the ANOVA to the variable model1:

```
model1 <- aov(price ~ cut, data=diamonds)
summary(model1)
```

4. Gereate the table using the pander package.
 - Create a new code chunk with the option `echo=FALSE`
 - Generate the markdown output using pander
 - Use the code `pander(model1)`

Add code chunk below:

Table 2: Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cut	4	1.104e+10	2.76e+09	175.7	8.428e-150
Residuals	53935	8.474e+11	15712087	NA	NA

7. Add a linked image to your document.

1. Use the image format `![image title](image link)` to embed a web linked image in the document.
 - Use the sample image link to add an image to this document: <https://bit.ly/2ZQA1cg>

Note: knitting a document with web linked images will only work when knitting to HTML. Web linked images will produce a LaTeX error when knitting to PDF.

Add Rmarkdown code below:

8. Knit your exercise worksheet to HTML or PDF output formats.

- Use the `knit` button in the tool bar to knit to the default setting
- Use the small chevron next to `knit` to select optional output formats.

Congratulations! You're on your way to making great documents using RStudio and Rmarkdown!